A GLOTTALIZED TONE IN MUONG (VIETIC): A PILOT STUDY BASED ON AUDIO AND ELECTROGLOTTOGRAPHIC RECORDINGS

Minh-Châu Nguyễn[1,2], Lise Buchman[2], Alexis Michaud[1,2] & Didier Demolin[2]

[1] Langues et Civilisations à Tradition Orale (UMR7107, CNRS – Sorbonne Nouvelle – INALCO)
[2] Laboratoire de Phonétique et Phonologie (UMR7018, CNRS – Sorbonne Nouvelle)
{minh-chau.nguyen; didier.demolin}@sorbonne-nouvelle.fr; lise.buchman1@gmail.com; alexis.michaud@cnrs.fr

ABSTRACT

The combination of pitch and glottalization (glottal constriction or lapse into creaky voice) as relevant phonetic/phonological dimensions of lexical tone is found in several language families in Asia. The Vietic subbranch of Austroasiatic stands out in that *all* its languages have at least one glottalized tone. Vietnamese is a well-documented example, but the others remain little-studied. The research reported here contributes experimental evidence on one of these languages: Muong *(Mường)*. Excerpts from a database of audio and electroglottographic recordings of twenty speakers allow for a characterization of this dialect's glottalized tone, as contrasted with the four other tones of this five-tone system. The ultimate goal is to determine what (sub)types of glottalized tones exist in the world's languages, bringing out typological differences in terms of (i) phonetic realizations and (ii) degree of importance of glottalization as a feature of linguistic tones.

Keywords: tone, glottalization, creaky voice, Muong language, Vietic languages.

1. INTRODUCTION

1.1. Terminological framework

Phonation types (laryngeal settings) receive sustained attention from a range of disciplines including phonetics/phonology, medicine, singing instruction, language teaching, and signal processing. Nomenclatures are highly diverse, depending on goals and approaches: in different frameworks, varying degrees of emphasis are laid on acoustics, perception and physiology, and the range of linguistic phenomena taken into account also varies.

A fundamental distinction is that proposed by [29] between four *laryngeal vibratory mechanisms*: m3 is 'whistle'-like phonation, "seldom used either in speech or in singing"; m2 and m1 roughly correspond to *head voice* and *chest voice*, respectively, and m0 to vocal *fry*, referred to here as *creak* or *creaky voice* (which we use interchangeably).

However, these laryngeal vibratory mechanisms are characterized as steady states (as found in the singing voice). In spoken language, states of the larynx can change rapidly: in particular, phonation can verge briefly on mechanism m0 (creaky voice). To characterize phonation types beyond the distinction between laryngeal vibratory mechanisms, we rely on the framework of reference formulated by Laver [21], as further refined by recent work [9]. Concerning the specific topic of the present research, we use *glottalization* as a cover term for all subtypes of creak and glottal constriction (following [15]), such as pressed voice, multiply-pulsed voice, and aperiodic creak (for further detail, see https://github.com/alexis-michaud/egg/tree/master/gallery).

1.2. Glottalized tones: an under-researched field?

"Southeast Asian languages use different combinations of pitch and phonation type to realize tonal contrasts, and nearly every imaginable combination is represented" [3, p. 193]. Specifically, the combination of pitch and glottalization as relevant phonetic/phonological dimensions of tone is reported in Sino-Tibetan (e.g. Burmese [31]), Tai-Kadai [12, pp. 305, 310]) and Hmong-Mien [1]. In this landscape, the Vietic subbranch of Austroasiatic stands out in that *all* its languages have at least one glottalized tone [11]. The glottalized tones of Northern Vietnamese can by now be considered to be well-documented phonetically [23, 2, 4, 30, 18], but beyond this textbook example, the tones of the other Vietic languages (Muong, Maleng, Arem, Chut/Ruc, Aheu, Hung, Tho) call for experimental phonetic exploration. The research reported here contributes experimental evidence on the Muong language. The ultimate goal is to determine what (sub)types of glottalized tones exist in the world's languages, bringing out typological differences in terms of (i) phonetic/phonological templates and allophonic diversity with respect to the full range of phonetic possibilities and (ii) degree of importance of glottalization as a feature of linguistic tones.

1.3. Experimental challenges

Studying glottalization on the basis of audio signals is a challenge, in that algorithms for f_0 detection based on autocorrelation give up when the signal is not quasi-periodic (for a detailed example, see [25, pp. 8-9]). The choice made here consists in using electroglottography (hereafter EGG) in addition to audio recording. The EGG signal, which offers an estimation of the variation in vocal fold contact area during phonation [10, 13, 28], provides a *linear* view of the *nonlinear* phenomena of vocal fold vibration, i.e. no more than partial information. EGG is nonetheless well suited to the exploration of glottalization: patterns of vocal fold contact area in glottalization can be discerned on the EGG signal (e.g. [8]), as will be illustrated further below by Fig. 1.

1.4. Target language

The target language is the Muong dialect of Kim Thuong, province of Phu Tho, Vietnam. It has five tones on smooth syllables (and two tones on stopped syllables, not discussed here), shown in Fig. 4. This language's *Ethnologue* code is MTQ. Seven field trips on this dialect were conducted since 2014 (total time in the field: 18 weeks).

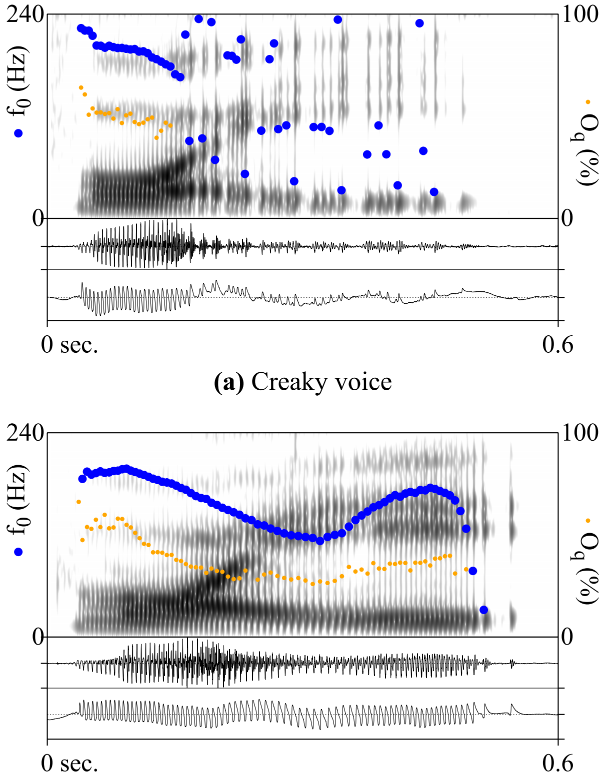
2. METHOD

2.1. Some challenges of phonetic experimental study in an unwritten language

Phonetic realizations of the glottalized tone are highly variable. Fig. 1 shows two realizations of **/vwɛj⁴/** 'salt' said in isolation by the same speaker. Fig. 1a has creaky voice, with irregular glottal cycles, starting after the first third of the syllable and lasting until the end. Fig. 1b reveals glottal constriction in the middle of syllable, and a glottalized offset of voicing. When a word is spoken in isolation, boundary effects are to be expected. Glottalization is known to be among the phonetic exponents of intonational boundaries (in tonal [20] as well as non-tonal languages [5]) and to partake in conveying attitudes [22]. Thus, there is no way to know whether the final glottalization in Fig. 1a-b is to be ascribed to the lexical tone or to intonational effects.

To guard against such unintended effects, and to stabilize the phonetic context, a standard method consists in having speakers read the target items inside a carrier sentence [26]. But Muong is an unwritten language, so reading was not an option. Vietnamese could have been used for elicitation, as speakers of Muong are also proficient in this national language, but it is so similar to Muong (somewhat like German and Dutch) that there is a high risk of interference between languages, creating experimental bias. Elicitation by means of photos appeared as the best choice.

Figure 1: Two realizations of **/vwɛj⁴/** 'salt' in isolation by speaker F1: (a) creaky voice, (b) glottal constriction. Top to bottom: spectrogram (0-5,500 Hz), acoustic signal, and EGG signal.



(a) Creaky voice

(b) Glottal constriction

Figure 2: Two minimal sets illustrating the five tones of Muong over the syllables **/paj/** and **/rɔ/**.

numbering	IPA	part of speech	Vietnamese	English
1	paj ¹	n	(quả) vải	lychee
	paj ²	n	đập bai	barrage
	paj ³	n	(cái) vại	jar
	paj ⁴	n	trái, quả	fruit
	paj ⁵	n	sải (tay)	armspan
2	rɔ ¹	adj	rảnh rỗi	idle
	rɔ ²	adj	no	to be sated
	rɔ ³	v	mò (cua)	to find crab
	rɔ ⁴	n	hoa chuối	banana flower
	rɔ ⁵	n	(con) rùa	tortoise

Two minimal sets were repeated two times by 4 speakers (one woman and three men) inside carrier sentence (1). The latest data (from fieldwork in 2018, adding 8 minimal sets said by 20 speakers) are currently being processed and should be ready for presentation in August, 2019.

(1) **/ja^2 măt^6** ____ **ţăŋ^3/**
2SG to_know *target item* INTERROG
'Do you know ____?'

2.2. Instructions to consultants

The two minimal sets shown in Fig. 2 require explanations to speakers prior to recording. Some monosyllables tend to be expanded into disyllables: for instance, 'tortoise' is **/rɔ^5/** in Muong, but it is often padded up into **/kɔn^1 rɔ^5/** by addition of a nominal classifier. Part of the preparation with speakers therefore consisted in explaining that the intended target item is always a monosyllable. A possible bias is that the task could become a sort of memory game, and the effort of remembering the intended syllable could compete with the demands of consistent, clear phonetic realization. But with just two minimal sets, we found that the procedure went well with all participants.

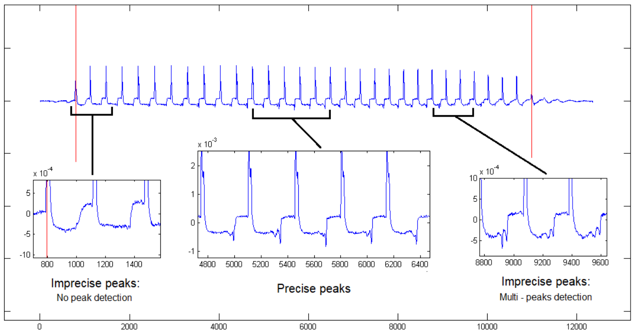
2.3. Electroglottographic analysis

Fundamental frequency (f_0) and glottal open quotient (O_q) were estimated from the *derivative* of the EGG signal, DEGG [14], using Peakdet, a script available from the COVAREP repository [6]. (An implementation in Praat is also available [19].) Peakdet is designed for semi-automatic measurement: the results for each token are verified visually. Estimating O_q requires the presence of a clear opening peak on the DEGG signal inside each cycle, a condition that is (by far) not met in all cases. 'Imprecise peaks' can mean that there are several peaks during the opening phase, or no well-marked opening peak at all. Fig. 3 illustrates examples of precise and imprecise peaks. When visual inspection shows that there is no single, clear opening peak, no estimation of O_q is provided for that cycle, as in the creaky portion in Fig. 1a. This explains the irregularity of the curves in Fig. 4b.

3. RESULTS: HOW TONE 4 STANDS OUT IN THE TONE SYSTEM

Fig. 4 shows the time course of f_0 and O_q for the Muong tone system by speaker M1. For reasons of

Figure 3: Examples of precise and imprecise opening peaks on DEGG signal during the same syllable. Abscissa: in samples (1 sample = 1/44,100 second).



space, results from speakers M5 and M6 cannot be set out here. All data and figures for these speakers are available from a Github repository: https://github.com/MinhChauNGUYEN/ICPhS-2019. Further discussion is also provided in a MA thesis available online [25].

The obtained results reveal that glottalization sets Tone 4 apart in terms of both of the acoustic properties under study. Firstly, f_0 values for Tone 4 drop from about 125 Hz to minimum values more than eight semitones lower before rising again in the second half of the rhyme, whereas the four other lexical tones are inside a range between 100 and 150 Hz.

Secondly, most tones have O_q values in the mid part of the speaker's range. In particular, Tone 2 and Tone 5 have similar values, around 50%; Tone 1 and Tone 3 are slightly lower, around 40%. Tone 4 is remarkably different. Its O_q plunges from initial mid-range values to the very bottom of range: below 30% (rising back to mid-range values at the end), which constitutes a telltale indication of glottalization.

Thus, both f_0 and O_q for Tone 4 are very different from the other tones, for the same reason: glottal constriction is strongly reflected in these two parameters, with longer cycles and longer closed phase inside each cycle.

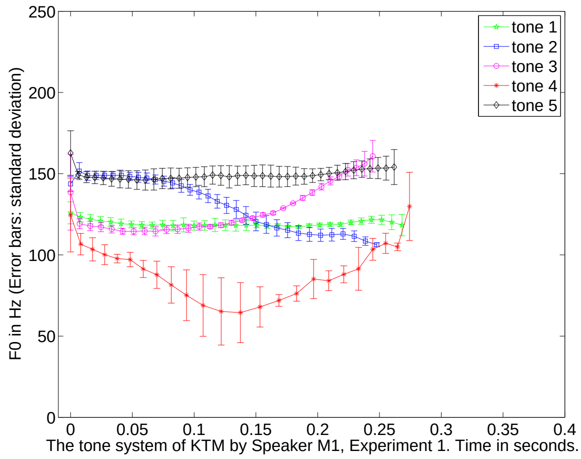
4. DISCUSSION

There have been various proposals on how to categorize phenomena variously labelled as *creak*, *vocal fry*, *laryngealization*, and *glottalization* [27, 17]. We provisionally distinguish here between creaky voice (a.k.a. phonation mechanism zero [29]), on the one hand, and on the other hand glottal constriction, which in some cases corresponds to press/tense voice as characterized in [17]. Proposals for dividing creaky voice into subtypes are offered in an online gallery: https://github.com/alexis-michaud/egg/

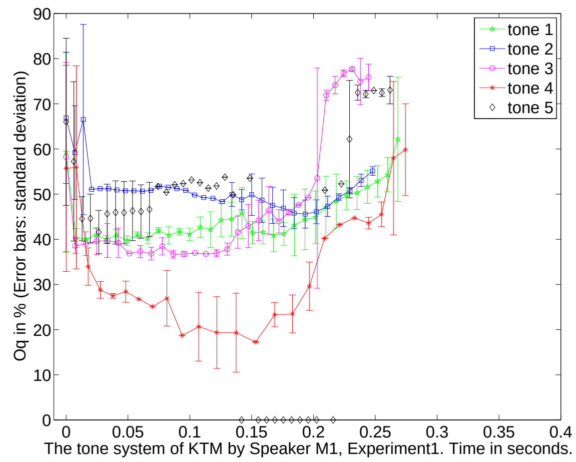
tree/master/gallery. These subtypes are all attested in realizations of Tone 4 in Muong.

For example, coming back to Fig. 1, there are striking differences between the two realizations of the same syllable **/vwɛj⁴/** 'salt' in isolation by speaker F1. The lapse into creaky voice in Fig. 1a is noticeable both in the audio and in the EGG signal by a decrease in amplitude and a salient change in waveform shape. In addition, the f_0 curve is 'saw-like', and O_q cannot be calculated for want of a clear division of the successive glottal cycles into a closed phase followed by an open phase. On the other hand, Fig. 1b has continuous curves of f_0 and O_q with average f_0 at 120 Hz and average O_q at 38%.

Figure 4: The tone system of Muong: curves of **(a)** f_0 and **(b)** O_q (with standard deviation). Speaker M1, 20 tokens.



(a) Fundamental frequency



(b) Open quotient

Concerning the distribution of allotonic variants: in deliberate speech (tending towards hyper-articulation), Tone 4 is canonically produced with creaky voice, although glottal constriction is also attested.

This helps place Muong in typological perspective. In some languages, glottalization is no more than a low-level by-product of low pitch targets [32, 20]; in others (phonation-type register languages), glottalization is a distinctive phonological property on its own [7]; and in yet others, such as Muong and Vietnamese, both pitch and glottalization matter, to varying degrees [16]. Muong is provisionally proposed as an example of a language having a lexical tone that includes *creak* as part of its phonetic/phonological template. Typologically, this canonical realization appears sufficiently distinct from the *glottal constriction* of Northern Vietnamese tones (final in Vietnamese tone B2, medial in Vietnamese tone C2 [4]) to warrant recognition as a separate type of glottalized tone.

5. CONCLUSION AND PERSPECTIVES

The pilot study reported here suggests that glottalization is central to Tone 4 in Muong, with creaky voice as its canonical realization and glottal constriction as a variant. Ongoing work on data from twenty speakers (ten women and ten men) now extends the investigation through a wider range of materials (including spontaneous speech: narratives and songs), and with statistical tools, to verify how close a match there is between phonetic creak and Tone 4 in spontaneous speech, and what patterns of allophonic variation emerge. A goal is to allow for a finer typological picture of phonological association of glottalization to tone.

In the mid run, a perspective will consist in relating observations on audio and EGG data with laryngoscopic evidence, to understand phenomena at the glottal and *epiglottal* levels, e.g. the implication of the false vocal folds during creak [24].

6. ACKNOWLEDGMENTS

Many thanks to the Muong consultants and friends. We gratefully acknowledge financial support from LACITO and *Ecole Doctorale 268*. This work is a contribution to the “Empirical Foundations of Linguistics” Labex project (ANR-10-LABX-0083).

7. REFERENCES

[1] Andruski, J. E. 2006. Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics* 34(3), 388–404.

[2] Brunelle, M. 2009. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37, 79–96.

[3] Brunelle, M., Kirby, J. 2016. Tone and phonation in Southeast Asian languages. *Language and Linguistics Compass* 10(4), 191–207.

[4] Brunelle, M., Nguyễn, K. H., Nguyễn, D. D. 2010. A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica* 67(3), 147–169.

[5] Crowhurst, M. J. 2018. The joint influence of vowel duration and creak on the perception of internal phrase boundaries. *The Journal of the Acoustical Society of America* 143(3), EL147–EL153.

[6] Degottex, G., Kane, J., Drugman, T., Raitio, T., Scherer, S. 2014. COVAREP: A collaborative voice analysis repository for speech technologies. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE 960–964.

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

[8] Esling, J. 1984. Laryngographic study of phonation type and laryngeal configuration. *Journal of the International Phonetic Association* 14, 56–73.

[9] Esling, J. H., Moisik, S. R., Benner, A., Crevier-Buchman, L. 2019. *Voice quality: the laryngeal articulator model*. Number 162 in Cambridge Studies in Linguistics. Cambridge: Cambridge University Press.

[10] Fabre, P. 1957. Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation: glottographie de haute fréquence. *Bulletin de l'Académie Nationale de Médecine* 141, 66–69.

[11] Ferlus, M. 1998. Les systèmes de tons dans les langues viet-muong. *Diachronica* 15(1), 1–27.

[12] Ferlus, M. 2008. The Tai dialects of Nghệ An, Vietnam (Tay Daeng, Tay Yo, Tay Muong). In: Diller, A., Edmondson, J. A., Yongxian, L., (eds), *The Tai-Kadai languages*. London & New York City, NY: Routledge 298–316.

[13] Fourcin, A. 1971. First applications of a new laryngograph. *Medical and Biological Illustration* 21, 172–182.

[14] Henrich, N., d'Alessandro, C., Castellengo, M., Doval, B. 2004. On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Journal of the Acoustical Society of America* 115(3), 1321–1332.

[15] Henton, C., Ladefoged, P., Maddieson, I. 1992. Stops in the world's languages. *Phonetica* 49, 65–101.

[16] Keating, P., Esposito, C., Garellek, M., Khan, S. u. D., Kuang, J. 2010. Phonation contrasts across languages. *UCLA Working Papers in Phonetics* 108, 188–202.

[17] Keating, P., Garellek, M., Kreiman, J. 2015. Acoustic properties of different kinds of creaky voice. *Proceedings of the 18th International Congress of Phonetic Sciences* Glasgow.

[18] Kirby, J. 2011. Illustrations of the IPA: Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association* 41(3), 381–392.

[19] Kirby, J. 2017. Praatdet: Praat-based tools for EGG analysis https://doi.org/10.5281/zenodo.1117189.

[20] Kuang, J. 2017. Creaky voice as a function of tonal categories and prosodic boundaries. *Proceedings of Interspeech 2017* Stockholm. 3216–3220.

[21] Laver, J. 1980. *The phonetic description of voice quality*. Cambridge, U.K.: Cambridge University Press.

[22] Mac, D.-K., Nguyen, T.-L., Michaud, A., Tran, D.-D. 2015. Influences of speaker attitudes on glottalized tones: A study of two Vietnamese sentence-final particles. *Proceedings of 18th International Congress of Phonetic Sciences* Glasgow. University of Glasgow.

[23] Michaud, A. 2004. Final consonants and glottalization: New perspectives from Hanoi Vietnamese. *Phonetica* 61(2-3), 119–146.

[24] Moisik, S. R., Esling, J. H., Crevier-Buchman, L., Amelot, A., Halimi, P. 2015. Multimodal imaging of glottal stop and creaky voice: evaluating the role of epilaryngeal constriction. *Proceedings of the 18th International Congress of Phonetic Sciences* Glasgow.

[25] Nguyen, M.-C. 2016. The tone system of Kim Thuong Muong: an experimental study of fundamental frequency, duration, and phonation types. Master's thesis Vietnam National University – Department of Linguistics Hanoi Hanoi.

[26] Niebuhr, O., Michaud, A. 2015. Speech data acquisition: The underestimated challenge. *KALIPHO - Kieler Arbeiten zur Linguistik und Phonetik* 3, 1–42.

[27] Redi, L., Shattuck-Hufnagel, S. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29(4), 407–429.

[28] Rothenberg, M. 1992. A multichannel electroglottograph. *Journal of Voice* 6(1), 36–43.

[29] Roubeau, B., Henrich, N., Castellengo, M. 2009. Laryngeal vibratory mechanisms: The notion of vocal register revisited. *Journal of Voice* 23(4), 425–38.

[30] Vu-Ngoc, T., d'Alessandro, C., Michaud, A. 2005. Using open quotient for the characterization of Vietnamese glottalized tones. *Proceedings of Eurospeech-Interspeech 2005* Lisboa. 2885–2889.

[31] Watkins, J. 2000. Notes on creaky and killed tone in Burmese. *SOAS Working Papers in Linguistics and Phonetics* 10, 139–149.

[32] Yu, K., Lam, H. W. 2014. The role of creaky voice in Cantonese tonal perception. *Journal of the Acoustical Society of America* 136(3), 1320–1333.