Using open quotient for the characterisation of Vietnamese glottalised tones

Tuân Vu-Ngoc, Christophe d'Alessandro, Alexis Michaud

Using open quotient for the characterisation of Vietnamese glottalised tones

*Vu Ngoc Tuan, Christophe d'Alessandro, Alexis Michaud**

LIMSI-CNRS, BP 133, F91403 Orsay
and *Laboratoire de Phonétique et Phonologie-CNRS, 19 rue des Bernardins, F75005 Paris
vnt@limsi.fr, cda@mlimsi.fr, alexis.michaud@univ-paris3.fr

Abstract

Vietnamese is a tone language in which the tone is a complex bundle of pitch and voice quality characteristics. The present study is restricted to Falling tones (i.e. it does not cover tone C2, called *nga* in Vietnamese spelling, which has medial glottalisation and ends on relatively high pitch), and deals mainly with tone C1 (*hoi*). This tone is generally described as falling then rising, but interestingly, some speakers realise it simply as falling. The primary aim of this paper is to investigate how these speakers maintain tone C1 distinct from two tones which are similar in terms of pitch: tones A2 (*huyen*) and B2 (*nang*). Analysis of audio and electroglottographic recordings of 7 speakers (441 syllables) confirms that there exist two types of strategies in the realisation of tone C1, and that in the falling realisation of C1, the voice quality at the end of the syllable differs from that of tones A2 and B2. It is further observed that this voice quality contrast cannot be captured by measurement of the open quotient alone, leading to general observations on the use of the open quotient in the characterisation of phenomena of glottalisation.

1. Introduction

The contrastive features of Vietnamese tones are melodic modulation and variation of voice quality [1]. How the pitch and voice quality of each of the tones is characterised in the standard variety of the language (Hanoi Vietnamese) is still a matter of debate, both phonetically and phonologically [2] [3] [4] (not to mention the considerable amount of experimental work that remains to be done on dialectal variation, and on the related Viet-Muong languages). The present study is restricted to Falling tones in Hanoi Vietnamese (and thus does not cover tone C2, called *nga* in Vietnamese spelling, which has medial glottalisation and ends on a high pitch), and deals mainly with tone C1 (*hoi*), which raises an issue concerning the voice quality contrasts used in the present-day state of Hanoi Vietnamese. Tone C1 as described by [1], and as still taught in the Vietnamese school system, is falling then rising, though interestingly, some speakers realise it as only falling, as reported by [2] and [3]. The primary aim of this paper is to investigate how these speakers maintain tone C1 distinct from two tones which are similar in terms of pitch: tones A2 (called *huyen* in Vietnamese spelling), which is low and falling, and tone B2 (called *nang*), which ends in glottal constriction. These three tones are investigated through audio and electroglottographic recordings which allow for the measurement of the open quotient (hereafter Oq; see experimental details and earlier results in [5] [6] [3]): Oq reflects the degree of adduction of the vocal folds. When the arytenoid cartilages are tightly pressed together, Oq is low.

After a short presentation of the corpus, the methods for measuring Oq are presented at some length, as this is crucial for an understanding of the usefulness and limitations of this parameter. Results for the three tones are then presented and discussed.

2. Corpus and Methods

2.1. Corpus

Three female speakers (F1-F3) and four male speakers (M1-M4) were recorded in a sound-treated booth at the LIMSI laboratory. 441 syllables were selected out of the database (which comprises about 3000 syllables recorded by each of the speakers): these syllables are of the form CV, where V is the vowel /a/, and C is one of the following consonants (in Vietnamese orthography): b, c, ch, d, đ, g, gi, h, kh, l, m, n, ng, nh, ph, r, s, t, tr, v, x; these segments are combined with tones A2, B2 and C1.

2.2. Methods

Acoustic and electroglottographic (hereafter EGG [8] [9]) signals were simultaneously recorded. The EGG signal monitors the changes in vocal fold contact area. It rises sharply when the glottis closes, reaches a maximum, then slowly decreases until the point where the vocal folds separate along their upper rim, at which point the EGG signal decreases most rapidly. The derivative (DEGG) of EGG typically has a positive peak at glottis closure and a negative peak at the opening [11]. The modulus of the closure peak is clearly greater than that of the opening peak (this is always the case in *chest voice*).

Closure peaks are easily detected by a threshold method applied to the DEGG signal (after 3-point smoothing; sampling frequency: 44100 Hz). (When there are several peaks, the peak with the most important modulus is chosen.) By contrast, opening instants are sometimes difficult to single out. One possibility in order to obtain an approximation of the opening instant in every case consists in applying a threshold method directly to the EGG signal, but the level chosen for the threshold (generally 3/7 or 1/2, see discussion in [11]) is ultimately arbitrary. In addition to Method 1 (detection of opening instants on the DEGG signal), a new method (Method 2) was attempted: a threshold method that elaborates on a proposal by [10], who used the DEGG method to detect closings and a threshold method at 3/7 of the signal to detect openings: under the present proposal ("closure-level threshold method"), the threshold is set on the basis of the position of the detected closure on the EGG signal: at the closure instant t1, the EGG signal has a value Y1; it then reaches a maximum value M1, then goes down to a minimum; at the next closure instant, t2, the EGG signal has a value Y2, then reaches a maximum M2. The threshold is set by drawing a line from the point (t1,Y1) parallel to the line M1M2; this line intersects with the EGG signal at O1, which is taken as the

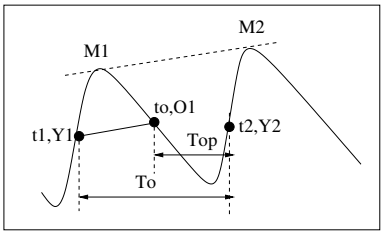

Figure 1: *A schematic illustration of Method 2 for the detection of glottis-opening instant.*

opening instant (to). (See figure 1.)

The period (To) is the duration in-between two closure-peaks. The opening duration (Top) is the time between an opening peak and the following closure peak. The open quotient (Oq) is the ratio of opening duration and To.

$$Oq = 100\frac{Top}{To} \quad (1)$$

2.3. Why both methods have some limitations

• On the whole, method 1 is to be preferred because it is based on observation of vocal fold physiology (see references in [6] [11]); but when the opening peak does not stand out clearly, Method 1, which selects the local minimum in-between two closings, yields unreliable results. This is illustrated in figure 2. The first two detected openings correspond to negative peaks in the DEGG signal, and thus can be taken as indicators of opening, whereas the last detected opening simply corresponds to a local minimum in the signal, found shortly after closing peak (positive peak); the visible opening peak (just before the next closing) goes undetected, resulting in erroneous open quotient computation.

• The value of Oq given by method 2 is only an approximation. It is highly sensitive to the exact timing of glottis closure, because at this point the EGG signal varies quickly, so that a small fluctuation in the closure instant will result in a large change in the vertical position of the threshold line; this has a heavy influence on the Oq value obtained, as shown in figure 3. This may be why [10] favours application of a threshold at a fixed height of the EGG signal (3/7) to detect openings. Also, when the EGG signal becomes irregular due to glottalisation, method 2 yields very unreliable results, as shown in figure 4 at $t > 80$ ms.

In such cases, straightforward detection of the opening instant by picking an opening peak is not possible.

Applicability of these two methods to our data is evaluated by looking at the standard deviation across all syllables.

Oq1(%) or Oq2(%) is the mean value calculated by method 1 or 2, respectively. $\sigma 1$ or $\sigma 2$ is the standard deviation.

The following ratios are also calculated:

$$r1, 2(\%) = 100\frac{\sigma_{1,2}}{Oq_{1,2}} \quad (2)$$

Tables 1 and 2 show the results for two speakers for whom the above ratios yielded opposite results:

These tables show that in tone A2 by speaker M3, $r1 < r2$; this is taken to mean that Method 1 yielded precise results, suggesting that one and a single rapid opening of the glottis is found within each glottal cycle, whereas for speaker F1, for the same tone (A2), $r1 > r2$. The standard deviation of the results yielded by method 1 is high, indicating a high proportion of

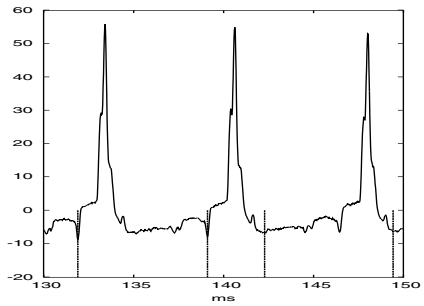

Figure 2: *An example where the opening peak does not stand out clearly, resulting in erroneous detection.*

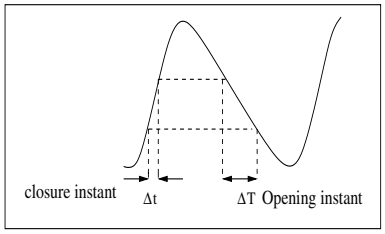

Figure 3: *A schematic illustration of the time difference in the approximation of the opening instant that can be caused by a slight change in the position of the closing peak:* $\Delta T > \Delta t$.

Table 1: Speaker *F1* .

tone	Oq1	r1	Oq2	r2
A2	62	12	60	6
C1	60	13	58	13
B2	58	23	55	16

Table 2: Speaker *M3* .

tone	Oq1	r1	Oq2	r2
A2	54	10	65	13
C1	53	13	65	12
B2	41	32	61	14

out-of-range values and thus the rarity of well-defined opening peaks. In examining all speaker/tone combinations, it is found that in approximately fifty percent of combinations, one method gives more consistent results than the other. In the next section, for each speaker/tone combination we choose the method that gives the smallest ratio *r*, on the assumption that it contains fewer out-of-range values.

3. Results

3.1. Curve fitting for F0 and Oq

The variations of F0 and Oq versus time are fitted by a line or a parabolic curve (depending on which seems appropriate for each set of data) using a least-squares method. When the fitting is done with a line, the values of F0, Oq and the slope of the line at t = 0 ms are calculated (F0L,sfL,OqL,sqL). When the fitting is done by a parabolic curve, equivalent values are calculated at 2/3 of the duration of the vowel, as a rough approximation which appears sufficient for the data at issue. This set of values

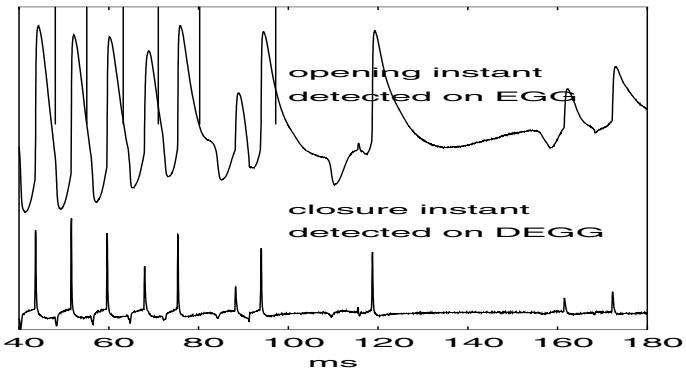

Figure 4: *An example illustrating problematic cases for opening peak detection. Top: EGG signal, bottom: DEGG signal.*

is labelled F0P,sfP,OqP,sqP.

The results are shown in the following tables. Frequencies are given in Hz and Oq in %, slopes in Hz/ms and %/ms.

Table 3: Speaker *A2* .

speaker	F0L	sfL	OqL	sqL
F1	210	-0.04	62	-0.01
F2	165	-0.08	48	0.00
F3	174	-0.10	54	+0.03
M1	120	-0.03	69	-0.01
M5	155	+0.02	59	+0.01
M2	103	-0.05	68	+0.01
M3	115	-0.02	51	-0.02

The value of speaker M5 and M3 are fitted with a parabolic curve, yielding:

M5: F0P = 128, sfP = -0.26, OqP = 63, sqP = +0.03.

M3: F0P = 104 , sfP = -0.05 ,OqP = 56 , sqP = +0.06. The results for tones *C1* and *B2* are shown in tables 4 and 5.

Table 4: *C1* .

speaker	F0L	sfL	OqL	sqL
F1	220	-0.34	67	-0.11
F2	157	-0.31	54	-0.01
F3	176	-0.52	50	+0.04
M1	124	-0.29	72	-0.16
M5	139	-0.48	56	-0.11
M2	104	-0.16	69	+0.04
M3	118	-0.18	56	+0.11
speaker	F0P	sfP	OqP	sqP
F1	176	-0.02	54	0.00
F2	158	+0.31	60	+0.05
F3	123	+0.20	59	+0.01
M1	94	+0.04	47	-0.05
M5	119	+0.32	54	+0.10

In cases of large fluctuation in F0 and Oq at the end of the syllable, curve fitting stops at the point where fluctuation increases (at about 2/3 of the duration of the vowel for speakers M2 and M3, at about 2/5 for F1). This was preferred over manual correction of individual data points: no semi-automatic corrections were done, unlike in our other studies [3] [6], in order to test the possibility of fully automatic measurement.

Table 5: *B2* .

speaker	F0L	sfL	OqL	sqL
F1	227	-0.13	59	-0.10
F2	190	-0.52	42	-0.06
F3	211	-1.52	55	-0.15
M1	138	+0.24	57	-0.08
M5	182	-0.20	50	+0.05
M2	117	-0.01	46	-0.07
M3	135	-0.13	42	-0.09
speaker	F0P	sfP	OqP	sqP
F2	120	-1.22		
F3	92	-0.19		
M1	122	-0.77		
M5	124	-0.52	63	+0.11
M2	88	-0.54		
M3	94	-0.99		

3.2. F0 at t = 0

F0 of tone *B2* is always greater than F0 of tone *A2*.

$$\Delta 1(\%) = 100\frac{F0(B2) - F0(A2)}{F0(A2)} \quad (3)$$

$\Delta 1$ varies from 8 to 22 %. (Recall that a semi-tone is a variation of about 6 %.) F0 of tone *C1* is lower than F0 of tone *A2* in some cases, higher in others.

$$\Delta 2(\%) = 100\frac{F0(C1) - F0(A2)}{F0(A2)} \quad (4)$$

$\Delta 2$ varies from -10 to +5 %. For all speakers, $|\Delta 1| > |\Delta 2|$.

3.3. Variation of F0 and Oq

• tone *A2*

Tone *A2* is slightly falling ($-0.10 < sf_1 < -0.0$) for 6 out of 7 speakers; the odd-man-out is speaker M5 ($sf_1 = +0.02$). Oq variation is small ($-0.1 < sq_1 < +0.03$).

• tone *C1*

Clearly, speakers F2, F3 and M5 realise a falling-rising F0 contour for tone *C1*. They differ in their use of voice quality. For speaker F2, Oq does not vary much in the course of the syllable (as reflected by sqL,2), whereas speaker M5 realises the falling part of tone *C1* with a voice quality that is more pressed (sqL = -0.11), and the rising part with a laxer voice quality (sqP = +0.10).

For speaker M1, the rising part of tone *C1* is less marked than for the last speakers (sfP = +0.04) and the change in voice quality is most important in the falling part (sqL = -0.16).

For speakers M2 and M3, tone *C1* is only falling.

• tone *B2*

For all speakers, tone *B2* is falling, as reported in [3] [6]. For speaker F1, the decrease is monotonous (the value of F0 are fitted with a line), for the other speakers F0 decreases slightly at onset (or even rises, in the case of speaker M1) then falls sharply.

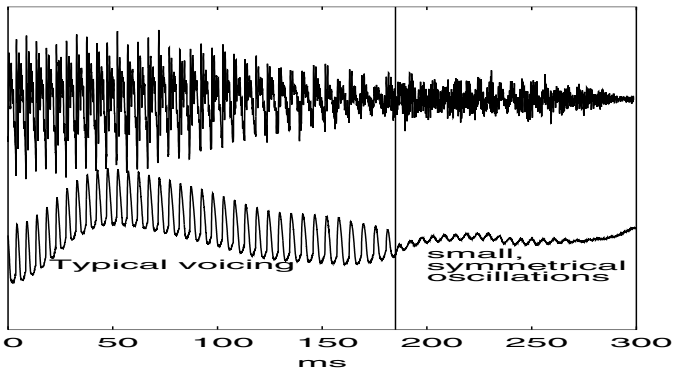

Figure 5: *Acoustic signal (top) and EGG signal (bottom).*

3.4. Tone *C1* in light of comparison across the three tones: essential observations

Speakers are distributed into two classes according to the shape of their fundamental frequency curves for tone *C1*:

- first, speakers realising tone *C1* with a falling and a rising part (F2, F3, M1 and M5). The rising part of speaker M1 is less marked than that of the other speakers, however.

- second, speakers (F1, M2, M3) realising tone *C1* as simply falling. This difference across classes does not seem to be conditioned by factors such as tempo: audio recordings of 12 speakers, made in Hanoi for a pilot study, confirm the existence of these distinct realisations of tone *C1*.

For the first class of speakers, tones *C1, A2, B2* are clearly different in terms of F0. In the second class, the slope depends on the speakers as follows (from steepest to least steep): for F1: *A2* falls less steeply than *B2*, which itself is less steep than *C1*; for M2 and M3: *C1* is less steep than *B2*. We take this as a hint that F0 is not the only cue to these tonal distinctions, otherwise greater inter-speaker consistency would be expected.

It was observed by [3] that when *C1* is simply falling, it is accompanied by laryngealisation (lapse into creaky voice). The present data, which take more speakers into account, do not show a strong similarity in voice quality for this tone across speakers. For speaker F1, tone *C1* shows no glottalisation. The open quotient curve does not point to a large change in voice quality: the values are within the range of modal voice. But in fact, the measurement of the open quotient from the DEGG signal stops at the point where the amplitude of the EGG signal decreases sharply and only symmetrical oscillations of small amplitude are found (c. 190 ms on the figure 5): a quasi-sinusoidal signal, the derivative of which shows no peaks (either closing peaks or opening peaks). We take this to mean that the glottis is open, the remaining periodic oscillations on the EGG signal being due to a (periodic) fluctuation in vocal fold contact area at the edges of the glottis. This hypothesis receives support from examination of the corresponding portion of audio signal: it shows a large amount of high-frequency noise, plausibly created by turbulence at the glottis; on first approximation, it appears similar to what happens in a voiced glottal fricative.

4. Conclusion

Measurements of open quotient were made to examine differences among the falling tones found in Hanoi Vietnamese.

- From the point of view of the measurement method, it was observed that: (i) The DEGG method needs to be complemented by other methods in numerous cases, (ii) in some cases, the open quotient does not appear as a valid measurement: it can only be meaningfully measured when each glottal cycle comprises one open phase and one phase of closure; more elaborate modelling of cycles is necessary in cases of laryngealisation, and in cases of extremely breathy voice where the glottis does not close and the fluctuation in vocal fold contact area is small. The EGG signal helps characterise these states of the glottis from a qualitative point of view, not from a quantitative point of view.

- From a linguistic point of view, the study confirms that some speakers consistently realise tone *C1* as falling; when realised as simply falling, this tone has either final laryngealisation, or extremely breathy ending. This finding complements previous studies, and has implications for modelling voice quality in Vietnamese, suggesting that the underlying voice quality feature of tone *C1* could be described as /lax/, its phonetic realisations ranging from (i) high open quotient to (ii) extreme breathiness (fricativised offset of voicing) and (iii) laryngealisation (lapse into creaky voice).

5. References

[1] Thompson, L.C 1965, *A Vietnamese Reference grammar*, University of Washington Press.

[2] Nguyen Van L., Edmondson, J.A. 1997, "Tones and voice quality in modern northern Vietnamese", *Mon-Khmer Studies* 28:1-18.

[3] Michaud, A. 2004, "Final consonants and glottalization: new perspectives from Hanoi Vietnamese", *Phonetica* 61:119-146.

[4] Pham, A.H. 2003, *Vietnamese Tone: A New Analysis*, London/New York, Routledge.

[5] Vu Ngoc T., d'Alessandro, C., Rosset, S. 2002 "A Phonetic Study of Vietnamese Tones: Acoustic and Electroglottographic Measurements", *ICSLP*, Boulder, Colorado, USA.

[6] Michaud A., Vu Ngoc T. 2004 "Glottalized and Nonglottalized Tones under Emphasis: Open Quotient Curves Remain Stable, Fo Curve is Modified", *Speech Prosody 2004*, Nara, Japan, 745-748.

[7] Laver, J. 1994 *Principles of Phonetics*, Cambridge University Press (see in particular p. 417).

[8] 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*:66-69.

[9] Childers D.G., Hicks D.M., Moore G.P., Eskenazi L. , Lalwani A.L. 1990 "Electroglottography and vocal Fold Physiology", *Journal of Speech and Hearing Research* 33: 245-245.

[10] Howard D.M. 1995 "Variation of Electrolaryngographically derived closed quotient for trained and untrained adult female singers", *Journal of Voice* 9(2): 163-72.

[11] Henrich N., d'Alessandro C., Castellengo M. and Doval B. 2004 "On the use of the derivative of electroglottographic signals for characterization of non-pathological voice phonation", *Journal of the Acoustical Society of America* 115(3): 1321-1332.