

# Glottalized and Nonglottalized Tones under Emphasis: Open Quotient Curves Remain Stable, $F_0$ Curve is Modified

Alexis Michaud<sup>1</sup> & Vu Ngoc Tuan<sup>2</sup>

<sup>1</sup>Laboratoire Phonétique et Phonologie (UMR 7018) CNRS/ Sorbonne Nouvelle, Paris

<sup>2</sup>LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur), Orsay

Alexis.Michaud@univ-paris3.fr, Tuan.Vu-Ngoc@limsi.fr

## Abstract

Tones B2 and D2 of Hanoi Vietnamese have strongly contrasting voice quality features: tone B2, which is glottally constricted, is here compared with a similar but nonconstricted tone (tone D2). Under emphasis, the slope of  $F_0$  curves is increased and/or  $F_0$  register is raised. The open quotient values (calculated from the derivative of the electroglottographic signal, which has been shown to yield very accurate values) do not vary significantly with emphasis. A description in terms of *curve amplification* is put forward. As for changes in length, they appear to be speaker-dependent.

## 1. Linguistic background: voice-register tones and "intonational accent"

- The tones of Vietnamese possess complex phonation-type features, as described by [15,4,18,11].
- Vietnamese also possesses *intonational emphasis*: as in many languages, the great variability observed in the realization of the lexical tones largely reflects the informational prominence of the various syllables in the utterance (for a detailed noninstrumental description of intonative accent in Vietnamese, see [15]; for Mandarin Chinese, which is similar in this respect, this has been known since [1] and [12]).

This typological situation raises the question of how the intonational prominence (or: "intonative accent", "emphasis") affects the various lexical tones, and their voice quality features.

The present investigation focuses on two tones which involve phenomena of glottal constriction (*pressed* phonation) vs. *relaxed* phonation: tones B2 and D2. Vietnamese has a six-tone paradigm for sonorant-final syllables, and a two-tone paradigm for obstruent-final syllables. Tone B2, belonging to the first paradigm, is glottally constricted; tone D2, which is the Low member of the second paradigm, is often described as rather close to B2 in terms of pitch, but is not glottally constricted [15]. Somewhat misleadingly, both are written the same way in the orthography, with a subscript dot (name: "tone *nặng*").

Two main parameters will be investigated: fundamental frequency ( $F_0$ ), and open quotient ( $O_q$ ).

## 2. Speech materials

In a pilot study, audio data was collected from 12 speakers in Hanoi; but because of superior recording facilities, the combined electroglottographic and audio recordings reported on here were done in France, with four (paid) native speakers, aged 19 to 24, who had left Hanoi for France less than one year before the time of the recording. The corpus includes all the combinations of vowel and final consonant with tones B2

and D2 allowed by Vietnamese, i.e. 84 syllables, recorded by the 4 speakers, under 2 reading conditions:

- condition 1: within the carrier sentence *Đây là chữ \_\_\_\_*. ("This is the word \_\_\_\_") or *Tôi đọc chữ \_\_\_\_*. ("I am reading the word \_\_\_\_.") The informants were instructed to imagine a context in which they were teaching a child or a foreigner who could not read.
- condition 2: within the carrier sentence *Đây là chữ \_\_\_\_ cơ mà!* "This is the word \_\_\_\_, can't you see?"

The final particle *cơ mà* is crucial for eliciting emphasis (in the same way as interrogation necessitates particles in Vietnamese and Chinese), and was therefore used despite the fact that it detracts from the symmetry of the two carrier sentences: the target syllable is utterance-final under condition 1 and not under condition 2. To check for possible effects of position, a third condition was tested (with one speaker only): a variant of condition 1 with carrier sentence *Đây là chữ \_\_\_\_ ạ*, also meaning "This is the word \_\_\_\_", but with a final particle (indicating respect). As there was no significant difference between conditions 1 and 3, it will be assumed in what follows that the differences between conditions 1 and 2 are due to the *emphasis* parameter and not to the *utterance-final/non-utterance-final* parameter. Condition 1 will be called "NE" (Non-Emphasized) and condition 2 "E" (Emphasized).

Lastly, the four speakers also read a newspaper article containing some tone-B2 and tone-D2 syllables. Visual inspection and calculations for individual syllables point to the same conclusions as under conditions NE and E; the facts from continuous speech, and observations on the cases where voicing takes up again briefly after the glottal interrupt of tone B2, will be presented elsewhere.

The target syllables were presented one by one in random order, alternating with items carrying another tone. The recordings took place in the sound-treated booth of the LIMSI laboratory (Orsay, France), with an EG-2 glottograph [13].

## 3. Method

The calculation of  $F_0$  and  $O_q$  is based on the derivative of the EGG signal (DEGG). The usefulness of the DEGG signal for the characterization of non-pathological phonation, and especially for a very precise determination of the instants of closing and opening of the glottis (as illustrated in figure 5), is shown on the basis of physiological data in [2,6,7].  $O_q$  is equal to the duration of the open phase (i.e. the duration between glottis-opening-instant and following glottis-closure-instant) divided by the period (defined as the duration between two glottal closures). It gives an indication on vocal fold adduction: a low  $O_q$  indicates a "pressed voice", a high  $O_q$  a "relaxed voice" [14,18].

For the male speakers (speakers 2, 3 and 4), the automatic detection of the glottis closing and opening instants was

successful in 95% of the cases. In the remaining cases, and in almost every case for speaker 1 (female speaker), some of the opening peaks did not stand out (it is a general observation that EGG recordings are more noisy for female subjects), resulting in erroneous opening peak detections. Two adjustments were therefore made in the programs: (i) the step for DEGG smoothing was set at 6 points (i.e. 0.3ms) (ii) the position of the opening peak was detected on the smoothed DEGG curve, not directly on the DEGG curve. The results of automatic detection were verified by visual inspection of every item. (In the cases where opening peak detection had been successful without smoothing, the modified program yielded identical results.)

For speaker 1, however, another method had to be used: detecting the closing instant from DEGG and then approximating the opening instant by a threshold method applied to the EGG signal [3,8]. (This method was used for speakers 2, 3 and 4 as a further check on the results.) As the results for speaker 1 are compromised values, which moreover add a *gender* variable, they are not compared with the results for speakers 2-3-4 here. The total number of syllables reported on in the present analyses is therefore 504 (84 syllables times 3 speakers times 2 reading conditions).

## 4. Results

### 4.1. Nonemphatic: A sharp contrast betw. tones B2 and D2

Typical EGG signals for tones D2 and B2 are presented in figure 1: nonglottalized vs. glottalized voice offset.  $F_0$  and  $O_q$  values are plotted in figure 2.  $O_q$  values for tone D2 are much higher;  $O_q$  increases gradually in the course of the syllable.

As for tone B2, the glottal constriction translates as a marked lengthening of the interval between glottal closures. Near the end of the syllable there is strictly speaking no *periodic* phenomenon anymore. The inverse of the duration between two glottal closures can nonetheless be calculated (following Fujimura's recommendation in [16]), still using the term " $F_0$ " for convenience. The open quotient values calculated at these points are well below the range that is found in sustained voicing, which is roughly from 30% to 70%.  $O_q$  is low from the onset of voicing: voice is "pressed", and the compression increases in the course of the syllable. Figure 2 shows that there is no exception to this pattern (this is true for all four speakers).

### 4.2. Emphatic: $O_q$ values stable, $F_0$ modified

In order to conduct statistical comparison of the two conditions, the individual curves (e.g. those plotted in figure 2) were normalized for duration and interpolated at 100 points. To summarize the data visually, a syllable simulation was also built for each 42-syllable subset (e.g. tone B2 by speaker 2 under emphasis), with averaged number of glottal cycles and averaged duration. Eight of these averaged curves are plotted in figures 3-4.

The statistical results are as follows:

#### 4.2.1. $O_q$ values under emphasis (E)

As the distribution of the data warrants the use of parametric tests, t-tests were used (two-tail) at each of the 100 equally distant points of the resampled curves. For tone D2,  $O_q$  tends to be slightly higher, a difference which only has marginal significance: it is significant at some of the 100 time points, but not on the syllable as a whole for speakers 2 and 3:

Table 1. Significance of  $O_q$  variation from NE to E

TONE D2	average diff. with NE	average p on the whole syllable
speaker 2	+5.6%	not significant (p=0.07)
speaker 3	+5.6%	not significant (p=0.09)
speaker 4	+6.0%	significant (p=0.02)
TONE B2	average diff. with NE	average p on the whole syllable
speaker 2	-0.3%	not significant (p=0.30)
speaker 3	-11%	not significant (p=0.18)
speaker 4	-2.9%	not significant (p=0.26)

#### 4.2.2. $F_0$ values under emphasis (E)

Table 2. Significance of  $F_0$  variation from NE to E

TONE D2	average diff. with NE	significant difference	on whole syllable (averaged p value)
speaker 2	+15% (1 tone ¼)	from 3 <sup>rd</sup> to 97 <sup>th</sup> points	highly significant (p=5*10 <sup>-11</sup> )
speaker 3	+21% (1 tone ¾)	at all points	significant (p=7*10 <sup>-4</sup> )
speaker 4	+2% (less than ¼ tone)	at no single point	not significant (p=0.48)
TONE B2	average diff. with NE	significant difference	on whole syllable (averaged p value)
speaker 2	+30% (close to 2 ½ tones)	at all points	highly significant: (p=8*10 <sup>-8</sup> )
speaker 3	+9.8% (¾ tone)	higher until pt 74; lower after pt 82	significant (p=0.044)
speaker 4	+2%	at no single point	not significant (p=0.4)

Average differences are only offered as a first attempt to estimate the distance between the two sets of syllables; the nonemphatic/ emphatic difference actually calls for a description in terms of *curve shapes*, as will be argued in section 5.

#### 4.2.3. Variations in other parameters

Two additional parameters were measured:

1) syllable length. No common pattern emerged, syllable lengths being higher under emphasis (E) for speakers 2 and 3, but lower for speaker 4.

Table 3. Significance of length variation from NE to E

	speaker	difference betw. NE and E	significant (s) or not significant (ns)
tone D2	2	+7%	ns (p = 0.22)
	3	+7%	ns (p = 0.09)
	4	-19%	s (p=1.8*10 <sup>-3</sup> )
tone B2	2	+8%	ns (p=0.39)
	3	+28%	s (p = 5.7*10 <sup>-8</sup> )
	4	-6%	ns (p = 0.13)

2) the amplitude of the peak on the derivative of the EGG signal at glottal closure (proposed name: DECPA), indicating the highest speed in increase of vocal fold contact surface at glottal closure. On the basis of a study on "DECPA, and its application in prosody" (presented at the present conference), the measured variable was *the DECPA maximum reached in the syllable* (rather than a mean value computed for the whole syllable). The pattern is similar to that for length: for speakers 2 and 3, the DECPA maximum is significantly higher under emphasis (p ranging from 10<sup>-4</sup> to 10<sup>-13</sup>), indicating sharper closing of the glottis; for speaker 4, the DECPA maximum is significantly lower under emphasis.

A last statistical observation is that the standard deviation of all parameters tends to be higher under emphasis (indicating greater variation across items), though in most cases the difference is not significant.

## 5. Discussion and conclusions

### 5.1. Variety across speakers, and underlying unity

The facts are not adequately captured by averaging values over the whole syllable. They call for modeling in terms of *curve amplification* under emphasis; one of its possible manifestations is  $F_0$  register raising (salient for speaker 2; figure 3), the other possible manifestation being that *the difference between maximum  $F_0$  and minimum  $F_0$  becomes greater*: speaker 4 realizes an  $F_0$  curve of increased range over a shortened syllable duration. The  $O_q$  curve is also amplified (details in 5.2). The comparison of speakers 2-3-4 helps understand which characteristics are significant across speakers: under emphasis, speaker 4 produces shorter syllables, i.e. the opposite of speakers 2 and 3; in Vietnamese, syllable lengthening therefore appears as a speaker-dependent variable, whereas a stable correlate of emphasis is curve amplification, manifested, according to the speaker, rather as an *increased slope of  $F_0$  curve* (and, less importantly, of  $O_q$  curve) or as  $F_0$  register raising.

### 5.2. $O_q$ as a stable characteristic of the tones

One salient finding from the experiment is the robustness of  $O_q$  as a correlate of the lexical tone. The very marked difference of  $O_q$  between tones B2 and D2 is retained under emphasis, with remarkable consistency across items and across speakers. The  $O_q$  contour is falling in tone B2; under emphasis,  $O_q$  values are higher early in the syllable, and lower towards the end, than under “Non-Emphasized” condition, i.e. there is a more pronounced movement of vocal fold adduction in the course of the syllable. In tone D2,  $O_q$  is high; under emphasis, it becomes still higher (in marginally significant proportions; see figures 3 and 4).

Incidentally, this study shows that unlike in Fukienese [10], Cantonese [9] and many other Sino-Tibetan languages, neither glottal closure nor glottalization accompanies Vietnamese final stops [reminder: all tone-D2 syllables have final stops; no tone-B2 syllable has final stops]. This goes against the claims of [17] (which were not based on experimental evidence, nor on first-hand data).

## 6. Perspectives

From a technical point of view, this study confirms that the open quotient calculated from the derivative of the electroglottographic signal (DEGG) offers precise information when the EGG signal is of sufficient quality (low-noise).

From a linguistic point of view, emphasis in tone languages is traditionally described in terms of *curve amplification* (e.g. [1,12,15]). It appears that this notion can be applied to voice quality parameters (here, open quotient) as well as to fundamental frequency. The results presented here offer a challenge to the recent models of tone-language speech synthesis (e.g. [5]): if the pragmatic component of intonation is to be modeled and integrated in addition to the syntactic component, synthesis of languages such as Vietnamese has to integrate the voice quality component in addition to  $F_0$ .

### Resource sharing - Hypertext link

The programs for EGG analysis can be obtained from: <http://voiceresearch.free.fr/egg/>. Data samples are available on the CD of the Proceedings; they are documented in an Appendix to the present document (5<sup>th</sup> page of electronic document).

## Acknowledgments

Many thanks to J. Vaissière, N. Henrich, K. Kohler, C. Fougeron, J.-Y. Dommergues, C. Gendrot; to F. Dell, M. Ferlus and L. Sagart; and to the Vietnamese informants and friends. *Financial support was given by the Fondation de France and the Conseil Scientifique of University Paris 3.*

## 7. References

- [1] Chao Yuen-ren, 1933. Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology* 4:3: 121-134.
- [2] Childers, D.G.; Naik, J.M.; Larar, J.N., *et al.*, 1983. Electroglottography, speech and ultra-high speed cinematography. *Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control*. I. R. Titze and R. C. Scherer. Denver Center for the Performing Arts: 202-220.
- [3] Davies, P.; Lindsey, G.A.; Fuller, H., *et al.*, 1986. Variation of glottal open and closed phases for speakers of English. *Proceedings of the Institute of Acoustics* 8: 539-46.
- [4] Ferlus, M., 2001. The Origin of Tones in Viet-Muong. Southeast Asian Linguistic society XI, Bangkok, Institute of Languages and Culture, Mahidol University.
- [5] Fujisaki, H.; Ohno, S.; Luksaneeyanawin, S., 2003. Analysis and synthesis of  $F_0$  contours of Thai utterances based on the command-response model. International Congress of Phonetic Sciences 15, Barcelona.
- [6] Henrich, N., 2001. *Etude de la source glottique en voix parlée et chantée*, Université Paris 6 Ph.D. dissertation.
- [7] Henrich, N.; d'Alessandro, C.; Castellengo, M., *et al.*, accepted. On the use of the derivative of electroglottographic signals for characterization of non-pathological voice phonation. *Journal of the Acoustical Society of America*.
- [8] Howard, D.M., 1995. Variation of electroglottographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice* 9(2): 163-72.
- [9] Iwata, R.; Sawashima, M.; Hirose, H., 1981. Laryngeal adjustments for syllable-final stops in Cantonese. *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics* 15: 45-54.
- [10] Iwata, R.; Sawashima, M.; Hirose, H., *et al.*, 1979. Laryngeal adjustments of Fukienese stops: initial plosives and final aplosives. *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics* 13: 61-81.
- [11] Pham, Andrea Hoa, 2003. *Vietnamese Tone: A New Analysis*. London/New York, Routledge.
- [12] Pike, K.L., 1948. *Tone Languages*. Ann Arbor, University of Michigan Press.
- [13] Rothenberg, M., 1992. A multichannel electroglottograph. *Journal of Voice* 6(1): 36-43.
- [14] Rothenberg, M.; Mahshie, J.J., 1988. Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research* 31: 338-51.
- [15] Thompson, L.C., 1965. *A Vietnamese Reference Grammar*, University of Washington Press.
- [16] Thongkum, T.L., 1988. Phonation types in Mon-Khmer languages. *Voice production: Mechanisms and functions*. O. Fujimura. New York, Raven Press: 319-333.
- [17] Thurgood, G., 2003. Vietnamese and tonogenesis: revising the model and the analysis. *Diachronica* 19(2): 333-363.
- [18] Vu Ngoc, T.; d'Alessandro, C.; Rosset, S., 2002. A phonetic study of Vietnamese tones: acoustic and electroglottographic measurements. ICSLP, Boulder, Colorado, USA.

Left column: tone 8 (nonglottalized)

Right column: tone 4 (glottalized)

Figure 1. Voice offset under tone 8 (left) and tone 4 (right). (Speaker 4, NE.) Vertical lines: closing instants; arrows: opening instants.

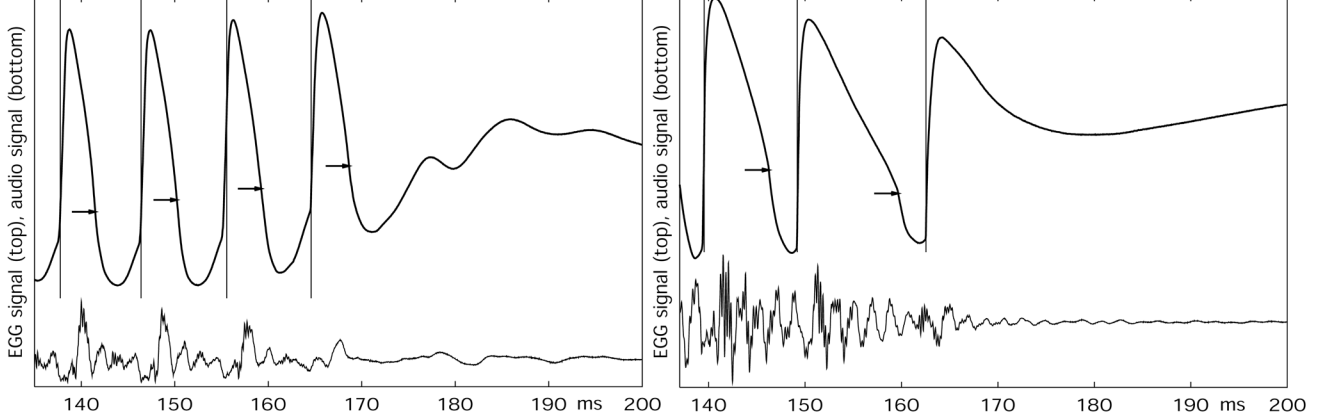


Figure 2.  $F_0$  and  $Oq$  values for tone 8 (left) and tone 4 (right). Speaker 2. Condition NE. 42 syllables are plotted in each figure.

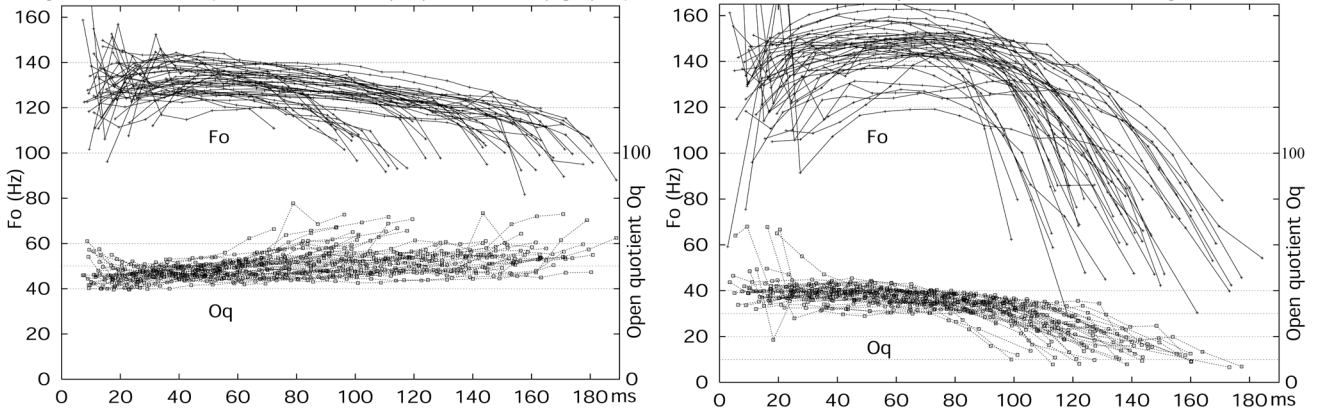


Figure 3. Averaged curves for tone 8 (left) and tone 4 (right). Speaker 2.

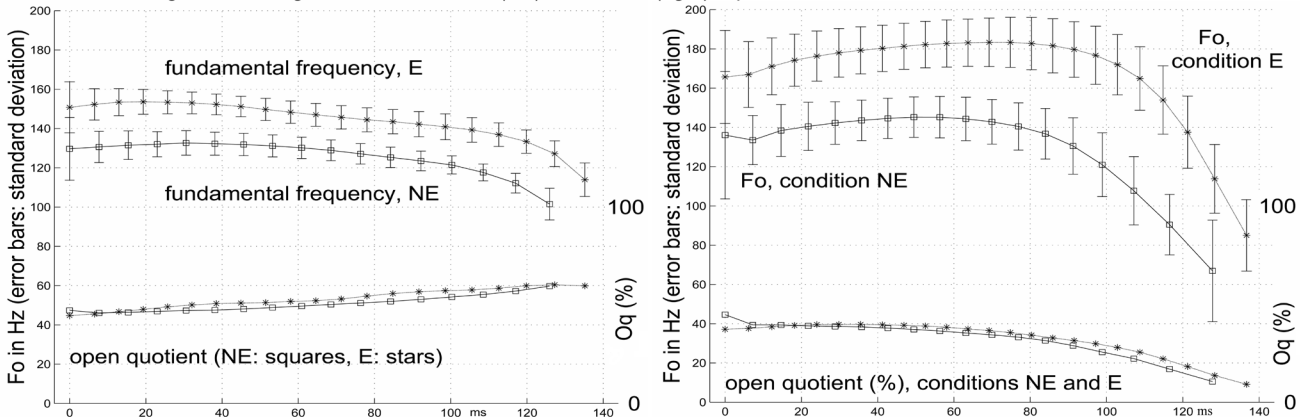


Figure 4. Averaged curves for tone 8 (left) and tone 4 (right). Speaker 4.

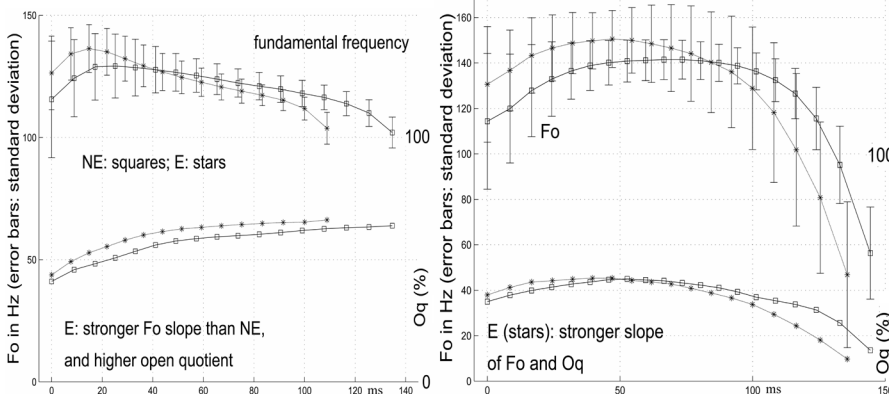
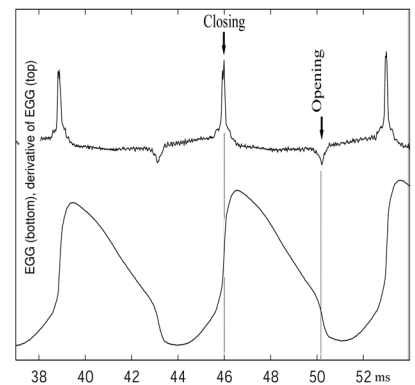


Fig. 5. Ex. of EGG signal and its derivative.



## Glottalized and Nonglottalized Tones under Emphasis: Open Quotient Curves Remain Stable, F<sub>0</sub> Curve is Modified

Alexis Michaud<sup>1</sup> & Vu Ngoc Tuan<sup>2</sup>

<sup>1</sup>Laboratoire Phonétique et Phonologie (UMR 7018) CNRS/ Sorbonne Nouvelle, Paris

<sup>2</sup>LIMSI (Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur), Orsay  
Alexis.Michaud@univ-paris3.fr, Tuan.Vu-Ngoc@limsi.fr

### 1. CONVENTIONS USED IN SOUND

#### FILE NAMES

code	meaning
s1	speaker 1
s2	speaker 2
<i>etc.</i>	
t8	tone 8 (=non-glottalized tone, D2)
t4	tone 4 (=glottalized tone, B2)
<i>etc.</i>	
ne	non-emphatic (=first reading condition)
e	emphatic (=second reading condition)
<i>etc.</i>	
A	Audio file
E	Electroglottographic signal

#### Example:

s3t4neE means :

speaker: 3  
tone: 4 (=tone B2)  
non-emphatic reading condition  
electroglottographic recording.

### 2. SEGMENTAL COMPOSITION OF THE RECORDED SYLLABLES

The item given for speaker 1 is syllable /ʃk/ (in Vietnamese spelling: *âc*), carrying tone 8 (D2).

The items given for speaker 2 are: tone 8 (D2): syllable /ãk/ (in Vietnamese spelling: *ăc*), tone 4 (B2): syllable /ɤm/ (in Vietnamese spelling: *øm*).

The items given for speaker 3 are: tone 8 (D2): syllable /ʃk/ (in Vietnamese spelling: *âc*), tone 4 (B2): syllable /ɤm/ (in Vietnamese spelling: *øm*).

The items given for speaker 4 are the recordings corresponding to figure 1 (given without the carrier sentence):

- fig1aEGG.wav, fig1aAUD.wav: recordings corresponding to the item in the left portion of figure 1: syllable carrying tone 8 (nonglottalized, D2). EGG: electroglottographic signal, AUD: audio signal.

- fig1bEGG.wav, fig1bAUD.wav: recordings corresponding to the item in the right portion of figure 1: syllable carrying tone 4 (glottalized, B2). EGG: electroglottographic signal, AUD: audio signal.

#### Note:

The tone-4 (tone-B2) syllable illustrates the difficulty there is in detecting the opening peak on the last glottal cycle. In such cases (which are frequent), Oq can only be computed until the period before last.