HAL
open science

Final consonants and glottalization: new perspectives from Hanoi Vietnamese

Alexis Michaud

To cite this version:

Alexis Michaud. Final consonants and glottalization: new perspectives from Hanoi Vietnamese. Phonetica, 2004, 61(2-3), pp. 119-146. 10.1159/000082560 . hal-00130119

HAL Id: hal-00130119

https://hal.science/hal-00130119v1

Submitted on 9 Feb 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

To: The Editor of *Phonetica*, Prof. Dr. Klaus Kohler

Institut für Phonetik und digitale Sprachverarbeitung

Universität Kiel

D-24098 Kiel

Submitted article

Title:

Final consonants and glottalization: new perspectives from Hanoi Vietnamese

Author's name:

Alexis Michaud

Institute where the work was conducted:

Laboratoire Phonétique et Phonologie (UMR 7018) CNRS/ Sorbonne Nouvelle, Paris

Short title:

Final consonants and glottalization

Full address:

Laboratoire Phonétique et Phonologie (UMR 7018) CNRS/ Sorbonne Nouvelle,

ILPGA, 19 rue des Bernardins, 75005 Paris, France

Telephone number: (00) 33 156 232 813

E-mail address: Alexis.Michaud@univ-paris3.fr

Abstract

The evolution from final obstruents to final glottal stop and then to rhyme glottalization (i.e. from /at/ to /aʔ/, then to /aˀ/) is a well-established general trend in the history of the Sino-Tibetan language family and beyond. It has further been shown by laryngoscopy that in three languages which retain the non-released syllable-final obstruents /p/, /t/ and /k/ (Standard Thai, and two Chinese dialects), these obstruents are often accompanied by a glottal stop. The present research raises the issue whether there is another typological possibility: can non-released final obstruents be accompanied consistently by modal phonation, without glottal stop? Analysis of electroglottographic recordings of 126 syllables in two carrier sentences spoken by four speakers shows that, in Hanoi Vietnamese, the final obstruents /p/, /t/ and /k/ are not accompanied by glottalization, and that the open quotient increases in the course of the syllable rhyme. Obstruent-final rhymes (which may carry either of two tones: D1 or D2) are compared with nasal-final rhymes which, under one of the tones (tone B2), are confirmed to be glottalized. Our finding is that tones D1 and D2 (i.e. obstruent-final rhymes) are both produced in modal voice, which shows that the typological paradigm of observed realizations of syllable-final obstruents must be enlarged. The discussion puts forward the hypothesis that the unusual association of segments and voice quality found in Hanoi Vietnamese is a strategy to maintain the opposition between B2-tone and D2-tone rhymes.

Introduction: the issue and its phonological background

Across language families, the presence of a glottal stop in association with a final stop, or as replacement for a final stop, is a common phenomenon [on the glottal stop occasionally accompanying voiceless final stops in various languages: Ladefoged and Maddieson 1996:73; on the "glottaling" of /t/ in British English: Przedlacka 2000 and references therein; on glottal stops and glottalization as signals of oral plosives /t/ and /k/ at word boundaries in German: Kohler 1994:38, 1999:89-90]. This phenomenon is particularly well attested in the Sino-Tibetan language family, both diachronically and synchronically. As for Hanoi Vietnamese (Austroasiatic family, Mon-Khmer branch), contradictory claims are found in the literature, some stating that final consonants are consistently glottalized [Thurgood 2002], others that they are consistently non-glottalized [e.g. Pham 2003]. The aim of the present investigation is to settle this controversy and to pursue the typological implications of the state of affairs found in Hanoi Vietnamese.

Why it is often assumed that all unreleased final stops in Asian languages tend to be accompanied by glottal stop

Diachronically, a change from syllable-final obstruents /p/ /t/ /k/ to syllable-final glottal stop took place in many Sino-Tibetan languages. The earliest documented languages in the family are assumed, on the basis of reconstruction and of evidence from the earliest writing systems, to have possessed final /p/ /t/ /k/ [Benedict 1972:13]: Old Chinese, Classical Tibetan (the Tibetan writing system was created in

the 7th century AD), and Classical Burmese (which has a writing system dating back to the 11th-12th century AD). This set of “weakly articulated, imploded lenis surds” *(ibid.)* is also postulated at the Proto-Tibeto-Burman level. The glottalization of final stops has taken place at a different pace in the various languages of the family, so that several stages can be observed across languages. In Burmese, the three final stops have merged into a single glottal stop [Armstrong and Pe Maung Tin 1925; see also Stewart 1955; Okell 1969]; a similar situation is found in neighbouring Lisu, Lahu and Akhan [Burling 1967:10; Bradley 1982:119]. In most Chinese dialects, from Late Middle Chinese (7th-10th c. A.D.) to Mandarin (10th c. A.D. to this day), final /p/ /t/ /k/ have disappeared without leaving any traces; in other dialects (Min, Hakka...), these stops have merged into a glottal stop, which can still be observed in synchrony [see, e.g., Sagart and Xu 2001:14]. An intermediate stage can be observed in Hayu (Bodic division of Tibeto-Burman), where final stops are reported impressionistically to be accompanied by simultaneous glottalization. In Hayu, each final stop has two contextual allophones: one is voiceless, non-released and accompanied by a simultaneous glottal stop, the other voiced and realized with creaky voice (laryngealization), without a complete glottal stop: [b̰] [d̰] [g̰] [Michailovsky 1988:53, 214; 2003].

The terms “glottal stop”, “glottal constriction”, “creaky voice/laryngealization” will be discussed below; in short, these three types can be characterized phonetically as follows: 1. *glottal stop* is a gesture of closure that has limited coarticulatory effects on

the voice quality of the surrounding segments; 2. *glottal constriction* (also referred to here as *glottal interrupt*) is a tense gesture of adduction of the vocal folds that extends over the whole of a syllable rhyme; 3. *laryngealization* (i.e. lapse into *creaky voice*), resulting in irregular vocal fold vibration, is not tense in itself. *Glottalization* is used as a cover term for laryngealization and glottal constriction [following Henton, Ladefoged and Maddieson 1992:73].

Laryngoscopic investigation into two of the Chinese dialects that have retained final stops /p/ /t/ /k/ showed that they are in fact usually accompanied by a glottal stop: in the Fujian (Fukien) dialect, the phenomenon can be regularly observed on isolated syllables, though not always in continuous speech [Iwata, Sawashima, Hirose and Niimi 1979]. Similar observations were made for Cantonese [Iwata, Sawashima and Hirose 1981]. In Standard Thai, the final stops /p/ /t/ /k/ are similarly reported to be accompanied by simultaneous glottal stop [impressionistic observations in Henderson 1964; evidence from laryngoscopy in Harris 2001].

In light of these converging observations, also paralleled in the Hmông-Miên language family [diachronic evidence in Niederer 1998], it is assumed by Thurgood [2002] that the general evolution of final consonants in East and Southeast Asia is toward simplification to a glottal phenomenon, which first accompanies the oral stop and later replaces it altogether: the typical pattern is that, at a first stage, a glottal stop is present, though not always (e.g. Fujian and Cantonese); at a second stage, a

glottal stop and/or glottalization is always present (e.g. Hayu); at a later stage, only glottalization of the rhyme is present.

In other Southeast Asian language families such as Austroasiatic, however, final /p/ /t/ /k/ did not turn into a glottal phenomenon in the course of history (apart from some languages in close contact with Sino-Tibetan languages); it has not been shown whether they have accompanying glottalization. Our experimental study focuses on Hanoi Vietnamese (Austroasiatic family, Mon-Khmer branch, Viet-Muong group).

The final consonants of Vietnamese in the history of the language

A brief diachronic overview of Vietnamese final consonants and tones will shed light on the status of obstruent-final syllables in this language. Vietnamese is an “isolating” language, in which each syllable usually has an independent meaning in isolation, and disyllables can be analyzed as combinations of monosyllables [see, e.g., Doan Thiên Thuât 1999]. Syllabic structure is (C1)(w)V(C2), where C1 is an initial consonant, w the semi-vowel /w/, V the vocalic nucleus (a short vowel, a long vowel, or a diphthong), and C2 a final consonant, which can be /p/, /t/, /k/, /m/, /n/, /ŋ/, or a semi-vowel /w/ or /j/; brackets indicate optional constituents. Each syllable possesses one lexical tone. The Vietnamese tone system belongs to the *pitch-plus-voice quality* type, i.e. the tone is not defined solely in terms of pitch: it is a complex bundle of pitch contour and voice quality characteristics [Maspero 1912; see also Han and Kim 1974;

Doan Thiên Thuât 1999]. The length of the vowel, and the presence or absence of a final nasal, have no influence on which tones the syllable can bear: there is no need to distinguish “heavy” syllables and “light” syllables, or to posit a division of the rhyme into morae. In contrast to the tone systems of African languages (e.g., typically, Niger-Congo family), in Vietnamese, as in many Asian languages, there are neither tone spreading and floating tones nor downstep. Unlike in some varieties of Chinese [Chen 2000], there is no tone sandhi in Vietnamese.

Vietnamese has a six-tone paradigm for sonorant-final syllables, and a two-tone paradigm for obstruent-final syllables (see table 1c). The historical developments that led to the complex tone system of present-day Vietnamese are by now a textbook example of tonogenesis, the various stages of the process having been explained since Haudricourt [1954]. “We can safely assert that all tone systems in modern Viet-Muong languages derive from a fundamental three-way contrast of Proto-Viet-Muong between –Ø (unmarked voiced ending rhyme), -ˀ (constricted voiced ending rhyme) and –h (laryngeal spirant ending rhyme). Checked syllables in -C (stop-ending rhyme) are apart. This implies that Proto-Viet-Muong was a toneless language” [Ferlus 2001:2]. (The symbol / ˀ / for glottal constriction, not used in IPA, dates back to Smalley [1963:389ff].) Table 1a sums up this state of affairs.

This three-way contrast was transphonologized into three lexical tones, commonly referred to as tones A, B and C [following Gedney 1989]. At that stage, there were

therefore three distinctive tones for sonorant-ending syllables, and obstruent-final syllables constituted a fourth set of syllables, without distinctive tone: category D (see table 1b). At a later stage, a second tonal split *(bipartition)* involving the disappearance of the opposition between the voiced and unvoiced initial consonants created the current paradigm of six tones for syllables with final sonorants (tones A1 through C2) and two tones for syllables with final obstruents (tones D1 and D2; more strictly speaking, two *architonemes*), as shown in table 1c.

Some doubts as to the validity of Haudricourt's analysis were voiced by Gage [1985], who pointed out apparent counter-examples to the reconstructions. In answer, Ferlus [2001:12-14] shows that "far from being an anomaly, these two-way correspondences are in fact a result of the first stage of Viet-Muong tonogenesis".

A schematic representation of the eight tone templates of Hanoi Vietnamese, based on data from one speaker (M1), is presented in figure 2. Phonetically, tones A1, A2 and B1 are produced with modal voice: A1 is High-level, A2 is Low-falling, B1 is High-rising. Tones B2 and C2 are glottalized: B2 has glottal constriction throughout, and is typically falling; C2 has medial glottal constriction and ends on a high F_0 value. Tone C1 is produced with comparatively relaxed voice, and varies (across speakers, and, for some speakers, across items) between a High-falling-Rising realization, and a Falling realization with final laryngealization [see impressionistic description in Thompson 1965; experimental data in Vu Ngoc, d'Alessandro and Rosset 2002; Pham Thi Ngoc Yên, Castelli and Nguyên Quôc Cuong 2002:26; Pham 2003 and

data in the present study, mainly on tones B2, D1 and D2]. As for tones D1 and D2, the orthography identifies tone D1 with tone B1 as *sắc* (written as an acute accent) and tone D2 with tone B2 as *nặng* (written as a subscripted dot); see table 1c, which indicates the names that the tones carry in present-day Vietnamese orthography [essentially dating back to de Rhodes 1651, drawing on work by close predecessors]. For instance, the name of the people and the language, *Việt* (carrying tone D2) and the word *viện* "academy, place" (carrying tone B2) are both said to have tone *nặng*. From the point of view of linguistic description, it is still an open question whether these tones are actually identical or not; contradictory views have been expressed on the phonetic realization and phonological status of tones B2 and D2, the two "*nặng* tones". During a two-year stay in Hanoi, in the process of collecting data from twelve speakers aged 20 to 65, I had the impression that tone D2 was not glottalized; this was the starting-point of the present research. The experiment reported below primarily aims to investigate the voice quality of tones D2 and B2; it will be shown that they have very different voice qualities. The discussion will hinge on the motivation of this polarization, and its implications. Incidentally, it will be confirmed that tone D1 is very similar phonetically to tone B1 (both being High-Rising, and produced with modal voice), but I will not attempt to resolve the phonological issue whether D1 and B1 are actually one and the same tone in the present state of the language.

Method

The investigation relies on electroglottography, combined with audio recordings and spectrographic analysis.

Corpus, speakers and listeners

The corpus includes all the combinations of the 14 vocalic nuclei and final consonant, with tones D1 and D2 (final consonants /p/ /t/ /k/) and tone B2 (final consonants /m/ /n/ /ŋ/), i.e. three sets of 42 syllables: (14 x 3) x 2 + (14 x 3) = a total of 126 syllables. These syllables consist in *rhymes*, of the structure V(C2), without any initial consonant. They therefore all have glottal onset, which in Hanoi Vietnamese serves as empty-onset filler in the absence of an initial consonant [Thompson 1965:4]. These syllables constitute lexical items, except for three out of the 126: the vowel /ɤ/ does not appear in combination with velar consonants /k/ and /ŋ/ (in Vietnamese spelling, including tone: *ớc, *ợc, *ợng). These syllables were nonetheless included in the corpus: reading “nonsense words” is usually accepted readily by native speakers, as noted for Vietnamese [Earle 1975:59; Gsell 1980] and Burmese [Thein Tun 1982:79]. When performing the reading task, informants are only intermittently aware (if at all) of the meaning carried by the syllables; knowledge of the writing system enables them to pronounce them easily, whether meaningful or not.

The 126 target syllables were recorded under two reading conditions:

1) within the carrier sentence *Đây là chữ* _____. (“This is the word _____.”) The informants were instructed to imagine a context in which they were teaching a child or a foreigner who could not read.

2) within the carrier sentence *Đây là chữ* _____ *cơ mà !* “This is the word _____, can’t you see?”

For brevity, condition 1 will be referred to as *Careful* reading (C), and condition 2 as *Emphatic* reading (E). The second condition was included to test for the influence of intonational parameters on tone realization. The final particle *cơ mà* is described by Nguyên Nhu Y [1999] as expressing *emphasis*, *strong affirmation*, with a suggestion of *contradiction*.

In Vietnamese (as in Chinese and many other Asian languages), there are numerous final particles; sentence mode and other intonational parameters cannot be changed without adding the appropriate final particle, as noted by Dô Thê Dung, Trân Thien Huong and Boulakia [1998:401]. Some studies overlook this important linguistic feature, either copying patterns from Germanic or Romance languages, or because adding a final particle modifies the number of syllables in the utterance, detracting from the symmetry of the reading conditions. Gårding [1987], Yuan Jiahong, Shih Chilin and Kochanski [2002] choose to instruct informants to read exactly the same sentences in affirmative and in interrogative mode, changing only the intonation; this yields “minimal pairs” of utterances which lend

themselves easily to comparison, but which are not found in natural speech. The validity of the conclusions that can be drawn from such materials needs to be questioned.

The target syllables (with final stops and with final nasals) were presented once, in random order. Each informant also read a supplementary corpus: 2934 syllables (open syllables and syllables with final nasals) arranged in tonal series: tones A1 to C2, without carrier sentences; and a newspaper article which contained some B2-, D1- and D2-tone syllables, allowing for a study of their realization in longer stretches of speech. Some of these data are used in the present study.

For technical reasons, the combined electroglottographic and audio recordings reported on here were not done in Vietnam but in France, with four native speakers (three male: M1 to M3, and one female: F1) aged 19 to 24 who had left Hanoi less than one year before the time of the recording. They were paid for the task. The recordings took place in a sound-treated booth (at LIMSI-CNRS laboratory, Orsay, France), with a GLOTTALENTERPRISES EG2 glottograph [Rothenberg 1992]. A microphone was placed about 30 cm from the mouth of the speaker. The audio and electroglottographic signals were recorded directly into a computer through a high-quality sound card at a sampling rate of 44,100 Hz.

Recent advances in the interpretation of the electroglottographic signal

The electroglottograph (hereafter EGG), also called laryngograph, was created by Fabre [1957] and developed by Fourcin [1971] and others. Some specialists caution against over-interpretation of the EGG signal [Colton and Conture 1990; Orlikoff 1998]. Yet numerous studies show that the EGG signal is a reliable measurement of the surface of vocal fold contact [Gilbert, Potter and Hoodin 1984; Titze 1990; Childers, Hicks, Moore, Eskenazi and Lalwani 1990; Baken 1992]. The method chosen here rests on the analysis of the *Derivative* of the EGG signal (hereafter DEGG), obtained by computing the difference between successive samples, i.e. calculating the slope of the EGG curve at each point. On typical DEGG signals, alternating positive peaks and negative peaks are observed (see figure 2). The interpretation of these peaks was established in light of the comparison of EGG and DEGG signals and simultaneously recorded images of the glottis [Childers, Naik, Larar, Krishnamurthy and Moore 1983; Baer, Lofqvist and McGarr 1983; Anastaplo and Karnell 1988; Karnell 1989; Hess and Ludwigs 2000]: the positive peak on the DEGG signal, which indicates the point at which the EGG signal rises most steeply, corresponds to the instant when the glottis closes over its full length (commonly called “glottis-closure-instant”); the negative peak (i.e. the point at which the EGG signal falls most steeply) usually marks the onset of glottal opening along the superior surface of the vocal folds (“opening instant”). It is therefore assumed here that the positive and negative peaks on the DEGG signal are indicators of glottis closing and opening, respectively. They allow for the measurement of the open quotient, defined as the length of the open phase (from one opening instant to the

following closing instant) divided by the length of the glottal cycle, itself defined as the time between two closing instants. The open quotient allows for the monitoring of vocal fold abduction: a low open quotient is an indicator of a *tight/pressed* voice, a high open quotient an indicator of *breathy* voice, as shown by Rothenberg and Mahshie [1988] on the basis of a comparison with airflow measurements.

How the derivative-EGG method was implemented in this study

The detection of the peaks on the derivative of the EGG signal was performed semi-automatically. The software developed for this study yields the following results: position and amplitude of the positive peaks and negative peaks, and values of F_0 and open quotient computed for each glottal cycle, without any condition on the values obtained (such as: plausible range, or continuity from one cycle to the next).

A different open quotient measurement method was also used, as a further check on the results in ambiguous cases: detecting the closing instant from DEGG and then approximating the opening instant by a threshold method applied directly to the EGG signal. This method was devised by Davies, Lindsey, Fuller and Fourcin [1986] and used by Howard, Lindsey and Allen [1990] and Howard [1995]. For a comparison between this measurement of open quotient, the DEGG method, and threshold methods [e.g. those used by Esling 1984; Rothenberg *et al.* 1988 and chosen by

Watkins 1999, 2002 in the study of register languages], see Henrich, d'Alessandro, Castellengo and Doval [2004:1329-1330].

Three minor limitations of the DEGG method

- Observation does not go beyond the last positive peak

Determining the length of the closed interval that follows the last closure does not appear possible on the basis of electroglottography: as can be seen on figures 3 and 4, no negative peak in the DEGG signal (which would indicate rapid opening of the glottis) appears after the last positive peak (corresponding to the last glottis-closure-instant).

- Two very close positive peaks are occasionally found instead of one single peak

In EGG recordings, a double closing peak may sometimes be observed. It is not an artefact: this double peak is related to the way in which the glottis closes [see discussion in Henrich *et al.* 2004:1324-1327]. No principled answer can be given as yet on which should be considered as the “true” closing peak. This happens more frequently for some speakers than for others (in our data, most often for speakers F1 and M2). Such double closing peaks can be seen on figure 3, on the two periods before last. A comparison of measurements made using the first and the second

peak showed that the incidence of the choice of the first or second peak is small: the difference in the open quotient and F_0 results is on the order of 2%. The software created for the present study selects the first peak automatically.

- Smoothing the signal over 0.3 ms was necessary because of unclear opening peaks

It is a general problem with the derivative-EGG method that some opening peaks do not stand out clearly. In 5% of the 756 syllables produced by the three male speakers (M1, M2, and M3), and in almost every syllable for the female speaker (F1), there were some sudden jumps on the automatically computed open quotient curve, and values outside the plausible range of variation. Inspection of the DEGG signal showed that this was due to the fact that some opening peaks were unclear (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 software: the DEGG signal was smoothed over 0.3 ms, for all speakers (see figure 2); and 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 4.36% cases (out of 21,700 measurements) where no significant opening peak

could be distinguished on the smoothed DEGG curve, or several opening peaks appeared, no open quotient value was assigned to the period in the results file.

Identification test

An identification test was later performed on the 126 syllables by two native listeners: one who had not taken part in the recordings (and was on a short stay in France), and speaker F1 (1 year after the recordings). The stimuli were played in randomized order through high-fidelity loudspeakers. The test was divided into two sessions. In the first session, four 126-syllable series (randomized anew each time) were presented within their carrier sentence, to evaluate the identification rate over a large sample of the data; in the second session, one 126-syllable series was presented three times (without carrier sentence) to find out whether there were any systematic misperceptions. The subjects were allowed to hear each stimulus again up to eight times at their own choice. They were instructed to write down the target syllable in Vietnamese orthography. Five-minute pauses were made between the series, which lasted about 12 minutes each. The purpose of the test was to check the distinctiveness of the hypothesized syllable contrasts.

Data analysis

The data for individual syllables (i.e. the F_0 and open quotient values calculated for each glottal cycle, as plotted in figures 5a and 6a) were interpolated linearly at 100 time points equally spaced between the first and last glottis-closure-instants, so that the results could be averaged over syllables. Phonologically short and long vowels are handled together in the calculations, and plotted together on figures 5-6, despite the high variation coefficient for length (22% on average). This was favoured over more complex interpolation procedures because it seemed better not to make assumptions on the division of the curves into sections; the choice of this procedure also rests on the belief that, in careful reading style, Vietnamese speakers use compression strategies rather than truncation strategies [on this distinction see Grønnum 1991; Grabe, Post, Nolan and Farrar 2000] when realizing the tone on a phonologically short rhyme. To illustrate the data visually, a syllable simulation was also computed for each 42-syllable subset (tone x by speaker y under reading condition z), with averaged number of glottal cycles and averaged duration. Four of these syllable simulations are plotted in figures 5b-6b. The calculations were made under the technical computing environment MATLAB.

Results

The results for the sharply contrasting tones B2 and D2 will be presented first, followed by the results for tone D1.

Under condition C (careful reading): a clear difference between glottally constricted and nonglottalized voice offset

A strong difference in voice quality between tone D2 and tone B2 is clearly evidenced by the shape of the electroglottographic signal and its derivative, and the F_0 and open quotient values.

- Qualitative data description by visual inspection of the EGG and DEGG signals

Figures 3 and 4 present typical signals showing that tone B2 has glottally constricted voice offset, and tone D2 nonglottalized voice offset. At the end of tone B2, a strong lengthening of the interval between successive glottal closures is observed over the last three or four cycles. Visual inspection of the EGG and DEGG signals (figure 3) indicates a tight adduction of the vocal folds under tone B2: the closed-glottis phase is very long, as evidenced by the fact that the glottis-opening-instant is found very shortly before the following closure. This is a point of similarity with creaky voice, for which “sharp, short pulses followed by a long closed glottal interval” were found by Childers and Lee [1991:2394] (see also Hollien [1974]; Blomgren, Chen Yang, Manwa and Gilbert [1998]), although in creaky voice, vibration is irregular, whereas at the end of tone B2 there is a monotonic lengthening from each glottal cycle to the next up to the end of voicing. The difference between creaky voice (also referred to

as *vocal fry* or *laryngealization*) and the phenomenon in tone B2, which is here called *glottal constriction*, will be further discussed below.

Voice quality is quite the opposite under tone D2, as illustrated by figure 4. The amplitude of the EGG signal increases steadily over the second half of the rhyme, a phenomenon regularly observed (in our data on Vietnamese and on other languages) at transitions into breathy voice (note that the amplitude of the EGG signal, which is indicative of the amplitude of variation in vocal fold contact area, has no direct correlation with the amplitude of the audio signal). Voicing then ends with a few oscillations of very small amplitude, which are not present for tone B2. These last oscillations seem to indicate that, at the offset of voicing, the glottis is open: as the vocal folds cease to be adducted, the changes in vocal fold contact area become very small. The offset of voicing in obstruent-ending rhymes with tone D2 is similar to that found in sonorant-ending rhymes with nonglottalized tones, e.g. rhymes such as /a/ and /am/, with tones A1, A2 or B1, which were also recorded with the same informants as part of the supplementary corpus of 2934 syllables. However, the small oscillations at the end are fewer in the obstruent-ending rhymes (tone D2): there are only two or three, versus about twice as many for sonorant-ending rhymes with the nonglottalized tones A1, A2 or B1. This can be assumed to be a consequence of the build-up of oral pressure following the oral closure: transglottal pressure falls sharply, and vocal fold vibration therefore terminates more quickly. The EGG signal shows that no closure of the glottis takes place immediately after the end of voicing. In some

cases (only 2 cases out of 84 for speaker F1, but almost always for speaker M3), there are two to three positive peaks in the DEGG signal 60 to 250 ms after the end of voicing; they are of smaller amplitude than the closing peaks during phonation. These movements of opening and closing of the glottis are also found after sonorant-ending rhymes carrying non-glottalized tones (A1, A2 and B1), which indicates that they should not be considered as a distinct property of tone D2.

- Length of glottal cycles (fundamental frequency)

Representative fundamental frequency values (speaker M1) are plotted in figures 5a and 6a. Near the end of tone B2, it is a general observation (valid across speakers and across items) that cycle length increases considerably from one cycle to the next; averaging across speakers and reading conditions, the difference between the last two cycles is 52%, i.e. cycle length increases by one half. The inverse of the duration between successive glottal closures can nonetheless be calculated [following Fujimura's recommendation in Thongkum 1988:333], still using the term "F_0" for convenience.

In tone D2, F_0 decreases more gradually in the course of the syllable, with a slightly stronger lowering towards the end. Averaging across speakers and reading conditions, this decrease amounts to 26%, i.e. 2 musical tones (*vs.* 64% for tone B2).

- Open quotient

The open quotient values of tone B2 and tone D2 are plotted at the bottom of figures 5a-b and 6a-b. They approach the opposite extremes of the range reported in the literature for sustained voicing, which is roughly from 30% to 70% [Titze 1988:236]. Tone B2 even results in final open quotient values under 30%. To researchers accustomed to measurements on modal voice, the very low open quotient values at the end of tone B2 may seem at first sight to be measurement artefacts. On B2-tone rhymes (bottom of figures 5a and 5b), open quotient is in the low part of the speaker's range throughout the syllable: voice is *pressed*, and the compression increases in the second half of the syllable, ending at values below the limit of the usual range reported by Titze *(op. cit.)*. Figure 5a shows that there is no exception to this pattern (this is true for all four speakers). On D2-tone rhymes (bottom of figures 6a and 6b), open quotient values are mid-range, increasing in the course of the syllable, indicating a decrease in the degree of vocal fold adduction. The respective ranges of open quotient values for tones B2 and D2 have almost no overlap, apart from the dispersed values early in the syllable, which are an effect of glottal onset. Under tone D2, the mean value across speakers on the last period is slightly above 60%, a value which hints at high airflow [Rothenberg *et al.* 1988 and discussion *supra*].

Comparison with emphatic reading (condition E): the voice quality characteristics of tone B2 and tone D2 are enhanced under emphasis

Recall that comparison of *Careful* and *Emphatic* reading conditions was meant to bring out the features of the tone that remain stable when the intonation and the position of the item within the utterance vary. The results confirm that, across conditions, the values of open quotient under tone D2 are consistently opposite to those found under tone B2.

- Under emphatic reading, F_0 values, rhyme length and acoustic intensity follow speaker-specific patterns, whereas open quotient values consistently become more extreme

The most striking outcome is the stability of open quotient across conditions, as opposed to the stronger variability of other parameters. This is illustrated by the curves at the bottom of figures 5b and 6b, which are very close. The tendency for open quotient values is to become slightly more marked under emphasis: even lower for tone B2 (though this is not apparent in figure 4b due to a difference in syllable length), even higher for tone D2. This tendency is generally significant over the second half of the syllable (the detailed results are found in table 4).

The parameters which change from condition C to condition E differ across speakers, and, for some speakers, across tones. As the distribution of the data warrants the use of parametric tests, statistical comparison of the two reading conditions was conducted by two-tail paired t-tests applied at each of the 100 equally distant points of the resampled curves. (Unpaired t-tests were applied in the cases where missing open quotient values resulted in groups of different size.) Tables 2 to 5 present statistics on F_0 values averaged over syllables, the difference between the highest and lowest F_0 value in each syllable (called " F_0 decrease", for short), the difference between lowest and highest open quotient, and syllable length, respectively. A summary is provided in table 6. F_0 variation exhibits various patterns: for speaker M1, it is overall F_0 register that varies most between conditions; for speaker M2, the amplitude and duration of F_0 decrease are consistently affected; for speakers F1 and M3, the parameters involved are different depending on the tone. Changes in length and intensity also appear to be speaker-dependent: syllable length is greater under emphasis for speakers F1, M1 and M2, but very significantly smaller for speaker M3's tone D2. As for overall acoustic intensity, the results must be interpreted with some caution, as the distance to the microphone may have varied somewhat during the recording session; a contrast can nonetheless be noticed between speakers F1-M1-M2, for whom a strong increase in intensity was found under emphasis (+5.8 dB on average for speaker M1), and speaker M3, for whom no significant intensity difference was found.

One last observation is that greater inter-token variation in F_0 is a regular correlate of emphasis. No such trend emerged for open quotient: in three cases out of eight, variability is lower under condition E.

- In nonfinal position, vocal fold vibration resumes in 37% of the cases after the glottal constriction of tone B2

For tone B2, in some cases, irregular vocal fold vibration takes place after the longest cycle. This depends on the position in the utterance: in nonfinal position (condition E, and continuous text reading), voicing takes up again in about two cases out of five (66% for F1, 2% for M1, 76% for M2, 41% for M3). In final position (condition C), voicing terminates abruptly, ending with lowest open quotient and longest glottal cycle: it takes up again only in less than 5% of the cases for speakers M1, M2 and M3, and in 18% of the cases for speaker F1.

This post-constriction voicing was never found in open syllables carrying tone B2 (data from the supplementary corpus), only in syllables that have a final nasal. In the statistics presented above, only the portion up to glottal constriction is included, as the focus is on the glottal constriction. In cases of such *post-constriction voicing*, there are five glottal cycles on average, which, to the foreign ear, may sound like pre-nasalization added to the syllable that follows: a short but audible voiced nasal segment is realized *after* the glottal constriction, with mid-range fundamental

frequency and open quotient values (mean open quotient on these periods is 56, standard deviation: 9), indicating that at these points of post-constriction voicing, the glottal constriction has been relaxed.

Tone D1, like tone D2, is not glottalized, but its open quotient is not as high as for tone D2

It will be recalled that tone D1 and tone D2 are the two possible tones for obstruent-final rhymes (cf. table 1c), tone D2 being low and falling, whereas tone D1 has high, rising fundamental frequency. Our experiment establishes that tone D1, like tone D2, is not glottalized, and that it is similar to tone B1, the High-Rising tone that appears on sonorant-final rhymes (leaving open the question whether they have actually become phonologically identical); they also indicate that open quotient under tone D1 is not as high as under tone D2, a fact which turns out to be of some importance to the discussion.

The EGG recordings of some D1-tone syllables contain an artefact: in 9% of the 336 syllables, vocal fold vibration ceased to be recorded before the end of the syllable (as evidenced by comparison with the audio signal). This may be due to the fact that, with the rapid rise in pitch, the vibrating portion of the vocal folds becomes smaller, making it more difficult for the electroglottograph to capture the changes in vocal fold contact area. In these cases, the whole syllable was excluded from the calculations;

but the occasional presence of this artefact casts some doubt on the accuracy of the measurement towards the end of the syllable in the 91% other cases as well. Fine-detail investigation into the changes in laryngeal configuration that take place at the end of tone D1 will require other exploratory techniques.

Even so, the data recorded appear sufficient to indicate that tone D1, like tone D2, is not glottalized. The offset of voicing is similar to that described for tone D2. Figure 7 presents average curves for tone D1. The open quotient curves have some similarity with those for tone D2: they are in the upper part of the speaker's mid-range, becoming higher in the course of the syllable. Interestingly, detailed comparison with the values for tone D2 indicates that, for speakers M1, M2 and M3, open quotient values for tone D1 are lower (this is true under both reading conditions; values across reading conditions: M1: - 12.5%; M2: - 4.80%; M3: - 13.8%). Comparison of tones D1 and D2 should take into account the fact that tone D1 has much higher F_0, as F_0 and open quotient are not entirely independent parameters: all other things being equal, open quotient tends to show a positive correlation with F_0 [data from singers in Henrich 2001:151 and references therein]. The findings for Vietnamese tones D1 and D2 are the more significant as they go against this tendency: under tone D1, F_0 is much higher than under tone D2 but open quotient significantly lower.

For speaker F1, the open quotient values under tone D1 are close to those under tone D2, and significantly higher on the last third of the syllable. This is due to the fact

that speaker F1, a female speaker, produces low pitches (e.g. tone D2) in “chest voice” (called, more technically, *laryngeal mechanism 1* in the classification of Roubeau, Chevrie-Muller and Arabia-Guidet [1987, 1991]; Henrich, Roubeau and Castellengo [2003]), and high pitches, e.g. at the end of tone D1, in “head voice” (laryngeal mechanism 2; *ibid.*), which has intrinsically higher open quotient.

The difference from non-emphatic to emphatic reading follows patterns similar to those that were found for tones B2 and D2: for speakers M1 and M2, fundamental frequency values are globally higher, and for speaker M3 syllable duration is reduced. For all speakers open quotient values are stable across the two reading conditions: the difference is on the order of 3%.

Results from the identification test: no systematic misperceptions; place of articulation is often misperceived

The results of the identification test are the following: there were no systematic misperceptions, but there were misperceptions regarding segments on 21% of the stimuli. This proportion is higher than the 15% reported in the experiment by Abramson and Tingsabadh [1999] on the perception of Thai final stops by native speakers, the difference being perhaps due to the more varied stimuli used in our test: 126 different syllables, as against 24.

The listener who had not participated as a speaker misperceived 24% of the stimuli. He did not perceive the presence of 19% of the final nasals under careful reading (condition C, i.e. utterance-final), hearing the corresponding syllables as open syllables, e.g. *ạ* (/a/, tone B2) instead of *ạn* (/an/, tone B2), and perceived one token of a B2-tone nasal-final syllable as a D2-tone obstruent-final syllable (*ẹn* /ɛn/, tone B2 heard as *ẹc* /ɛk/, tone D2). In 13% of the cases, the place of articulation of the final consonant was not perceived correctly, the most often misperceived consonant being /p/, which was often perceived as /t/ or /k/.

The listener who had also participated as a speaker (speaker F1) misperceived 18% of the stimuli. In 16% of the cases, the place of articulation was not perceived correctly; again, /p/ was often perceived as /t/ or /k/. She perceived one token of a B2-tone nasal-final syllable as a D2-tone obstruent-final syllable (*ệm* /em/ heard as *ệt* /et/), and conversely, heard one *ặt* /ăt/ as *ặn* /ăn/.

Discussion

The first part of the discussion brings out the salient finding of the experiment, briefly reviews earlier publications on this topic, and places the results in a typological perspective. The second part develops a hypothesis concerning the language-internal motivation of the typologically unusual state of affairs found in Vietnamese.

It appeared that the High-Rising tone D1 (found on obstruent-final rhymes) is not glottalized: it is produced with modal voice, like the High-Rising tone B1 (found on sonorant-final rhymes). In view of their phonetic likeness both in terms of F_0 and of open quotient, it appears possible to hold that D1 and B1 are one and the same tone: in synchrony, their identification in the orthography may well represent phonemic reality, which amounts to saying that the D1 architoneme has become identified with tone B1 in the history of the language. The main finding from the present experiment is that the same may not be said of tones D2 and B2: Earle [1975:54] argues that they are likewise allotones, but a consistent difference between the two was in fact observed.

These diverging conclusions can be traced back to differences in the method of investigation. Earle's experiment consists in measuring F_0 by autocorrelation on 20 ms frames. Three points within each syllable are then selected: onset, midpoint and endpoint. *Onset* is defined as "the F_0 reading of the first stable data point, with 'stable' defined as 'not differing from the succeeding point by more than 5 Hz' " [Earle 1975:71]. The *endpoint* value is similarly defined as the last stable data point; Earle notes that for tone B2, this endpoint is in fact "the last stable data point before the incidence of non-modal phonation", excluding the final portion of the rhyme, where glottal constriction is strongest. The result of statistical comparison over these simplified data is that there is no detectable difference between the F_0 values of

tones B2 and D2, from which Earle draws the conclusion that tone D2 is a conditioned variant of tone B2 [Earle 1975:168, 173, 186]. But in fact, as the final glottal constriction of tone B2 is observed without exception in the speech of all informants (his as well as ours), it must be part of the phonological description, as must the absence of glottal constriction of tone D2 [Pham 2003:37-54; see also Avery 1983; Nguyên Van Loi and Edmondson 1997]. A salient finding from our experiment is the robustness of voice quality as a correlate of the lexical tone. It is evidenced by the stability of open quotient across the two reading conditions, as opposed to the variation affecting other parameters (F_0, length, and acoustic intensity). The strong contrast of open quotient between tone B2 and tone D2 is retained under emphasis, and even enhanced, with remarkable consistency across items and across speakers. The *pressed* voice quality of tone B2, which terminates in tight adduction of the vocal folds, and the modal voice quality of tone D2 (with high open quotient pointing to loose vocal fold adduction towards the end), are essential characteristics of these tones. This confirms the status of Vietnamese lexical tones as contour-plus-voice-quality tones.

Taxonomy: glottal constriction vs. glottal stop and laryngealization

For the sake of cross-language comparison, it is a matter of importance to choose a nonambiguous and phonetically adequate label for glottal phenomena. In his classical description of Vietnamese tones, Thompson [1965:41] impressionistically

observes that tone B2 has “rasping voice quality”, “drops very sharply and is almost immediately cut off by a strong glottal stop.” This corresponds to our observation that the *pressed* voice in Vietnamese tone B2 (what Thompson calls “rasping voice quality”) extends over the syllable rhyme as a whole. As for the term “glottal stop”, however, it were perhaps best restricted to the phonetic description of a gesture of closure that has less coarticulatory effects on the voice quality of the surrounding segments (whether it corresponds to the stop segment /ʔ/, or to another linguistic function: empty-onset filler as in Vietnamese or Ikwere, hiatus-breaking juncture phenomenon as in English or German, or “floating” segment as in some North American languages). In the case of a Vietnamese B2-tone syllable that has a nasal ending, e.g. /ɓan/, the notation as /ɓanʔ/ would imply that the stop comes after the nasal, whereas in fact the glottal phenomenon extends over the whole rhyme, and actually interrupts the syllable, almost suppressing the nasal segment. Figure 8 presents syllable /ɓan/ carrying tone A2, with a visible nasal segment, and carrying tone B2, without a clearly distinct nasal segment, though the transition within the vowel is still visible. (Syllables with and without final nasals still contrast when carrying tone B2, but the nasal segment is difficult to perceive, even for native speakers, as evidenced by the identification test: 12% of the nasal-final syllables with tone B2 were heard as open syllables.)

The phenomenon at issue seems best described as “glottal constriction” [as used by Gedney 1986:209-212 and Ferlus 2001, among others], or “glottal interrupt”, two

labels that are used interchangeably here. The term “glottal interrupt” aims to render the rapidity and intensity of the articulatory gesture, which, to quote again Thompson’s description, *cuts off* the syllable.

Glottal constriction should be distinguished from *creaky voice*, thus recognizing a three-way distinction: glottal stop, glottal constriction, and creaky voice. “Creaky voice” [Catford 1964; 1977:95-106; Laver 1980; Ladefoged *et al.* 1996:73], also called “laryngealization”, “vocal fry” [Hollien, Moore, Wendahl and Michel 1966] or “phonation mechanism 0” [Roubeau *et al.* 1987, 1991; Henrich *et al.* 2003], is the irregular, low-frequency vibration of the vocal folds.

The difference between glottal constriction and laryngealization can actually be illustrated using Vietnamese data (from our supplementary corpus of 2934 syllables): laryngealization is not uncommon at the end of tone C1, as illustrated by figure 10, which can be compared to the glottal constriction of tone B2 in figure 3 (by the same speaker). To present very briefly the state of affairs for tone C1: our belief is that its occasional laryngealization is one out of several phonetic correlates of a /lax/ phonation feature, greater rhyme length being another of its correlates. This description would account for the large amount of variation of tone C1 across speakers and even within the speech of the same speaker. As the [High-falling, laryngealized] variant of tone C1 is not reported in early descriptions of Hanoi Vietnamese, where only its [High-falling-Rising] realization is described, I believe that

the [High-falling, laryngealized] realization is currently in the process of superseding the other realizations of tone C1 in “advanced Hanoi Vietnamese” (a phrase coined in reference to “Advanced R.P.”, “Advanced Stockolm” etc., referring to the colloquial, present-day state of a language as spoken in a major city). This would suggest that in this language, the distinction between glottal constriction and laryngealization now plays a supporting role in a linguistic opposition (a typologically uncommon situation), raising the phonetic issue of how much perceptual relevance the final laryngealization has in the present state of the language. Full treatment of this question must be deferred until another publication. For the purpose of the present investigation, I simply use the example of tone C1 represented in figure 10 to illustrate salient phonetic differences between glottal constriction and laryngealization: *duration* and *regularity*.

- Linguistically significant final laryngealization (i.e. lapse into creaky voice) tends to extend over a longer time span than final glottal interrupt: in figure 10, final laryngealization extends over some 200 ms; in the EGG data provided by Thongkum [1988:327] for four syllables of the Chong language, an Austroasiatic register language that uses a distinction between modal voice, breathy voice, and creaky voice (laryngealization) to distinguish lexical items, final laryngealization (on syllables pronounced in isolation) lasts ten to fifteen periods.

- In creaky voice, periodicity is irregular, and “complex-repetitive patterns” (patterns of more than one peak-per-cycle) are among the characteristics of creaky voice [see Blomgren *et al.* 1998:2653]. By contrast, the periodicity at the end of tone B2 is not irregular: cycle length increases monotonically from one cycle to the next, and no “complex-repetitive patterns” are found.

Opening up the range of observed realizations of nonreleased final stops

The recognition of the phonation type of Vietnamese D2-tone syllables as consistently opposite to the glottally constricted phonation of tone B2 opens up the typological paradigm of unreleased final consonants. As reviewed in the introduction, three stages were attested across languages: (i) glottalization is present, though not always (e.g. Fujian and Cantonese); (ii) glottalization is always present in conjunction with oral closure (e.g. Hayu); (iii) only glottalization is present (a glottal stop has replaced the oral stop). A fourth possibility is illustrated by Hanoi Vietnamese, where rhymes with final consonants consistently lack glottalization.

This fact is of importance to discussions of tonogenesis, where the presence of a glottal stop on obstruent-final rhymes was assumed in recent speculations. Thurgood [2002:337-341] claims that present-day Vietnamese final stops co-occur with glottal closure, that this feature dates back a very long way, and uses this hypothesis as a clinching argument in his reinterpretation of Haudricourt’s model. It is now clear that

present-day Hanoi Vietnamese cannot be adduced to argue that glottal closure accompanied final stops in earlier states of the language[1].

Incidentally, the present experiment also demonstrates that the final stops' unreleased character is not due to the presence of a glottal stop as Thurgood [2002:341] surmises: "It is not uncommon in Southeast Asia for final stops to be accompanied by glottal closure, with this indicated by their rather conspicuously unreleased, unexploded character". This is no more self-evident than Grammont's mistaken view [Grammont 1933:40-45, taken up in Heffner 1950:119] that French stops are articulated with a simultaneous glottal stop which distinguishes them from the aspirated stops of Germanic languages.

A hypothesis on the motivation of the lax phonation type of tone D2

The main result of the experimental study raises an issue: why does Hanoi Vietnamese feature consistently modal phonation on tone D2, without glottalization (and even with high open quotient, pointing to loose vocal fold adduction towards the end)? The place of obstruent-final syllables inside the tonal system, with their own

[1] Since the article is not based on first-hand observation, a minor interpretation error might have slipped in, concerning the passage from Thompson's *Vietnamese Reference Grammar* [1965] adduced by Thurgood: Thompson calls tones B2 and D2 by one and the same name, the one they carry in the orthography: *nặng* (see *supra* table 1c), resulting in a formulation that is misleading for readers unfamiliar with the complexities of Vietnamese orthography: that the tone *nặng* of the orthography "ends in a stop [i.e. tone D2] or is cut off abruptly by a glottal stop [i.e. tone B2]" [Thompson 1965:16]. Thompson did not imply that oral stop and glottal stop are freely alternating variants of each other, though his phrasing is easy to miscontrue.

tonal sub-system, seems to be the same in Vietnamese, Cantonese and Fujian. How come they show no glottalization in Hanoi Vietnamese, whereas they do in Cantonese, Fujian, and other languages? Vocal fold abduction is a possible strategy to interrupt voicing, but adducting the vocal folds tightly can also serve the same purpose, and in any case there is no phonological voicing opposition to motivate such a strategy, as Vietnamese has no voicing distinction on final stops. Besides, the decrease of vocal fold adduction under tone D2 is gradual, extending over the whole of the syllable rhyme, so that it can hardly be put down to phonetic influence from the articulatory gestures that produce the final stop [for a phonetic study of the influence of consonants on voice quality, revealing cross-language differences, see Ní Chasaide and Gobl 1993]. The voice quality of tone D2 calls for an account of its functional motivation.

The hypothesis put forward here is that B2-tone rhymes with final nasals and D2-tone rhymes with final stop of corresponding point of articulation stand in a relation of opposition, and that the voice quality contrast between tones B2 and D2 is an important means of distinguishing within rhyme pairs such as /ap/ (carrying tone D2) and /am/ (carrying tone B2), /at/ and /an/, /ak/ and /aŋ/; in Vietnamese writing: *ạp* and *ạm*; *ạt* and *ạn*; *ạc* and *ạng*. The relationship within these pairs is different from that between the two members of a “minimal pair”: both the final phoneme and the tone differ. Segmentally, however, these pairs of rhymes stand in a relation of phonetic proximity: in addition to the general fact that the syllable-final position is a weak

position [diachronic evidence in Martinet 1955; Martinet 1975:67; recent experimental results in Benkí 2003 and references therein], in the case of Vietnamese the stops have no audible release (the diacritic [̚] is implied in all the transcriptions) and the nasals almost entirely disappear acoustically as segments under tone B2, in concomitance with the presence of glottal interrupt (see spectrograms in figure 8); nasalization of the vowel does not seem to be a distinctive cue (compare, in figure 9, syllables /aŋ/ and /ak/). One indication of the weakness of final nasals is the post-constriction revoicing strategy used (in varying degrees) by the four speakers when a B2-tone nasal-final syllable is not in utterance-final position: as post-constriction revoicing (reported in the “Results” section) was observed on nasal-final syllables and not on open syllables carrying the same tone, it can be interpreted as a strategy to enhance the perceptibility of the otherwise threatened final nasals. Further evidence is provided by the proportion (12%) of nasal finals which, in the absence of revoicing, went unheard during the identification test. By contrast, confusion between nasal and obstruent took place only in 0.2% of the cases, indicating that listeners are able to retrieve sufficient cues to distinguish the two types of final consonants.

The comparison of these two sets of rhymes, as it was conducted in this study, is therefore more than a convenient experimental setup to bring out the features of tones B2 and D2: the contrast in voice quality that it brings out might well play a functional role in the language. It seems highly plausible that the presence of glottal constriction or, on the contrary, of modal phonation is an important cue to the

distinction between the two sets of rhymes. From a perceptual point of view, the differentiation threshold for open quotient, all other parameters being kept equal, was shown to be about 15% (for sustained phonation, under laboratory listening conditions) [Scherer, Arehart, Guo, Milstein and Horii 1998; Henrich 2001:171-195]. The open quotient difference between tones B2 and D2, which is over 33% on the whole syllable, and especially strong at the end (a three-to-one ratio on the last period), can therefore be considered as a very likely perceptual cue. Confirmation from perception tests can hopefully be obtained in future by using multi-parameter speech synthesis software; the ideal setup would be to lower artificially the open quotient of D2-tone rhymes, and investigate the impact on recognition scores, but there is at present no satisfactory way of modifying a recorded audio signal to simulate a change in vocal fold adduction.

Phonologically, the present analysis echoes Benedict [1948], who notes that tone B2 has glottal constriction, and re-analyzes it as a Low tone plus the added feature of /glottal accent/, tone D2 being simply described as a Low tone. Benedict's phonological analysis may not be entirely convincing in detail, but our findings definitely substantiate his view that voice quality serves as a distinctive trait to oppose tone categories B2 and D2.

Language-internal evidence supporting the hypothesis

To support the notion that the voice quality of tone D2 is motivated by a contrast with the voice quality of tone B2, data from tone D1 can be adduced. Recall that tone D1, a Rising tone found on rhymes with final stops, bears a great phonetic resemblance to tone B1, the Rising tone of the sonorant rhymes' tonal paradigm. As tone B1 is not glottalized [as established by Vu Ngoc *et al.* 2002, and verified in the present observations], the segments of B1-tone rhymes are fully realized: in particular, the nasals are clearly audible (none was misperceived in the identification test). As no phonetic proximity of the kind observed between tones B2 and D2 threatens the D1-tone syllables, it can be predicted that tone D1 does not resort to a marked voice quality. Following this line of argument, the prediction is that D1-tone rhymes will have mid-range open quotient values, not markedly high values as in tone D2. This prediction is borne out: open quotient values for tone D1 are in the middle of the speaker's range, and the statistical comparison of tones D1 and D2 shows that the open quotient is significantly lower under tone D1. (This result is not entirely conclusive, however, because the EGG recordings of some D1-tone syllables may contain artefacts, as pointed out in the Results section.)

Theoretical implication: contrasts between rhymes, rather than between segments

It has been noted by Henderson [1985:3-5] and others that a segment-by-segment approach misses both synchronic and diachronic insights, especially in the case of

Southeast Asian languages. The syllable calls for a division into *onset* and (tone-bearing) *rhyme*, as has in fact long been the case in the local linguistic traditions. The rhyme includes in Vietnamese (brackets indicate optional constituents): (w)V(C), where V is the vocalic nucleus, and C a final consonant: /p/, /t/, /k/, /m/, /n/, /ŋ/, or a semi-vowel /w/ or /j/. Within the rhyme, there can be *feature shuffling*, to use Henderson's phrase. For example, after the back vowels /u o ɔ/, velar consonants are labialized; the vowel quality of /o/ and /ɔ/ is affected: the actual pronunciation of /ok/ /ɔk/ /oŋ/ /ɔŋ/ is approximated as [ʌuk͡p̚] [ɐuk͡p̚] [ʌuŋ͡m] [ɐuŋ͡m] by Thompson [1965:12], and as [əuk͡p̚] [auk͡p̚] [əuŋ͡m] [auŋ͡m] by Henderson [1985:21]; I propose to transcribe it as [ɤok͡p̚] [ʌɔk͡p̚] [ɤoŋ͡m] [ʌɔŋ͡m], to reflect the intuition that lip rounding is moved to the right, the onset of the diphthong consisting in the unrounding of the original vowel: from [o] to [ɤo], from [ɔ] to [ʌɔ]. In addition to such phenomena of feature movement within the rhyme, the present study further illustrates the reality of the rhyme as a unit by bringing out the existence of functionally active *rhyme contrasts* which cannot be broken up into phonemic contrasts. Our data suggest that two sets of rhymes stand in a relation of opposition, motivating a strong contrast in voice quality: D2-tone obstruent-final rhymes are consistently produced with nonglottalized phonation (tending to breathiness at the end), arguably to help keep them distinct from B2-tone nasal-final rhymes. The contrast between the two sets can be described neither as purely suprasegmental, nor as purely segmental; it is a contrast between entire rhymes.

Conclusion

From an experimental point of view, the present study confirms that precise and reliable information on phonation type (voice quality) can be obtained from electroglottography.

The study also confirms that voice quality is a robust correlate of tone in Vietnamese, showing less variability than F_0 across reading conditions. The experiment warrants the conclusion that tones D1 and D2 (i.e. the tones of syllables ending in /p/, /t/ or /k/) are not glottalized, either in final or non-final position. Tone B2 is characterized by a gesture of strong constriction that is distinct from creaky voice.

This essential result complements the findings of Iwata *et al.* [1979, 1981] and Harris [2001], by showing that the state of affairs found in two Chinese dialects and in Standard Thai should not be assumed in other languages of Southeast Asia without experimental verification: unreleased final stops are not necessarily accompanied by a glottal stop. This finding illustrates the usefulness of experimental phonetic studies for typology, as well as for language diachrony.

Within the system of Vietnamese, the consistently modal quality of D2-tone, obstruent-final rhymes is probably motivated by a need to distinguish them from B2-

tone, nasal-final rhymes. This relation of opposition, neither purely segmental nor purely suprasegmental, calls for a description of the “dynamic synchrony” [Martinet 1975:5-10] of Southeast Asian languages that takes into account the contrasts found between rhymes, and pays attention to the continuing relationship of segments and tones in Southeast Asian languages.

Acknowledgments

I cannot too warmly state my gratitude to Jacqueline Vaissière for her sympathetic and incisive criticism of this paper. Many thanks to Vu Ngoc Tuân, as part of whose research programme this investigation could be conducted, and to Nathalie Henrich for scientific guidance and technical expertise on electroglottography. Many thanks to Nick Clements, François Dell, Jean-Yves Dommergues, Michel Ferlus, Cécile Fougeron, Cédric Gendrot, Klaus Kohler, Barbara Kühnert, Martine Mazaudon, Christine Mooshammer, Laurent Sagart, and two reviewers (Arthur S. Abramson and an anonymous reviewer) for the wealth of useful comments they offered.

Preliminary results were presented at the 16th *Journées de linguistique Asie Orientale*, Paris, and the Speech Prosody 2004 conference, Nara, Japan.

Last but not least, my thanks to the Vietnamese informants and friends.

The usual restrictions apply.

Resource sharing

The original recordings and data files can be obtained from the author, as well as the MATLAB routines for curve averaging. Requests for the software for extracting parameters from the electroglottographic signal should be sent to Tuan.Vu-Ngoc@limsi.fr.

References

Abramson, A.S.; Tingsabadh, K.: Thai Final Stops: Cross-Language Perception. Phonetica *56*: 111-122 (1999).
Anastaplo, S.; Karnell, M.P.: Synchronized videostroboscopic and electroglottographic examination of glottal opening. Journal of the Acoustical Society of America *83*(5): 1883-1890 (1988).
Armstrong, L.; Pe Maung Tin: A Burmese Phonetic Reader (University of London Press, London 1925).
Avery, P.: Evidence for a laryngeal tier in Vietnamese. Toronto Working Papers in Linguistics: 1-26 (1983).
Baer, T.; Lofqvist, A.; McGarr, N.S.: Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. Journal of the Acoustical Society of America *73*(4): 1304-1308 (1983).
Baken: Electroglottography. Journal of Voice *6*(2): 98-110 (1992).
Benedict, P.K.: Tonal systems of Southeast Asia. Journal of the American Oriental Society *68*: 184-191 (1948).
Benedict, P.K.: Sino-Tibetan: A Conspectus (Cambridge University Press, Cambridge 1972).
Benkí, J.R.: Analysis of English Nonsense Syllable Recognition in Noise. Phonetica *60*: 129-157 (2003).
Blomgren, M.; Chen Yang; Manwa, L.N.; Gilbert, H.R.: Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. Journal of the Acoustical Society of America *103*(5): 2649-2658 (1998).
Bradley, D. Register in Burmese. Papers in Southeast Asian Linguistics n°8: Tonation: 117-132. D. Bradley. (Pacific Linguistics, Series A, n°62 1982).
Burling, R.: Proto-Lolo-Burmese (Mouton & Co., Bloomington/The Hague 1967).
Catford, J.C. Phonation types: The classification of some laryngeal components of speech production. In Honour of Daniel Jones: 26-37. D. Abercrombie, D. B. Fry, P.

A. D. MacCarthy, N. C. Scott and J. L. M. Trim. (Longmans, Green and Co., London 1964).
Catford, J.C.: Fundamental Problems in Phonetics (Edinburgh University Press, Edinburgh 1977).
Chen, M.Y.: Tone Sandhi : Patterns Across Chinese Dialects (Cambridge University Press, Cambridge 2000).
Childers, D.G.; Hicks, D.M.; Moore, P.; Eskenazi, L.; Lalwani, A.L.: Electroglottography and vocal fold physiology. Journal of Speech and Hearing Research *33*: 245-254 (1990).
Childers, D.G.; Lee, C.K.: Vocal quality factors: Analysis, synthesis and perception. Journal of the Acoustical Society of America *90*(5): 2394-2410 (1991).
Childers, D.G.; Naik, J.M.; Larar, J.N.; Krishnamurthy, A.K.; Moore, P. Electroglottography, speech and ultra-high speed cinematography. Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control: 202-220. I. R. Titze and R. C. Scherer. (Denver Center for the Performing Arts, Denver, Colorado 1983).
Colton, R.H.; Conture, E.G.: Problems and Pitfalls of Electroglottography. Journal of Voice *4*(1): 10-24 (1990).
Davies, P.; Lindsey, G.A.; Fuller, H.; Fourcin, A.J.: Variation of glottal open and closed phases for speakers of English. Proceedings of the Institute of Acoustics *8*: 539-46 (1986).
de Rhodes, A.: Dictionarium Annamiticum Lusitanum et Latinum (Rome 1651).
Dô Thê Dung; Trân Thien Huong; Boulakia, G. Intonation in Vietnamese. Intonation Systems: A Survey of Twenty Languages: 395-416. D. Hirst and A. Di Cristo. (Cambridge University Press, Cambridge 1998).
Doan Thiên Thuât: Ngu âm tiêng Viêt [Vietnamese phonetics] (Nhà xuât ban dai hoc quôc gia hà nôi: Hanoi National University Press, Hanoi, Vietnam 1999).
Earle, M.A.: An acoustic phonetic study of Northern Vietnamese tones. Santa Barbara, Speech Communications Research Laboratory: 211 (1975).
Esling, J.: Laryngographic study of phonation type and laryngeal configuration. Journal of the International Phonetic Association *14*: 56-73 (1984).
Fabre, P.: 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 (1957).
Ferlus, M. The Origin of Tones in Viet-Muong. Text of communication at the 11th Annual Conference of the Southeast Asian Linguistic Society XI, Bangkok, Institute of Languages and Culture, Mahidol University (publication of Proceedings is currently under way) (2001).
Fourcin, A.J.: First applications of a new laryngograph. Medical and Biological Illustration *21*: 172-182 (1971).
Gage, W.W. Glottal stops and Vietnamese tonogenesis. For Gordon H. Fairbanks, Oceanic Linguistics Special Publications n°20: 21-36. V. Z. Acson and R. L. Leed. (University of Hawaii Press, Honolulu 1985).

Gårding, E.: Speech Act [Statement, Question, Focus] and Tonal Pattern in Standard Chinese : Constancy and Variation. Phonetica *44*: 13-29 (1987).
Gedney, W.J. Speculations on early Tai tones. Contributions to Sino-Tibetan Studies: J. McCoy and T. Light. (E.J. Brill, Leiden 1986).
Gedney, W.J.: Selected papers on comparative Tai studies (The University of Michigan Publications, Ann Arbor, Michigan 1989).
Gilbert, H.R.; Potter, C.R.; Hoodin, R.: Laryngograph as a measure of vocal fold contact area. Journal of Speech and Hearing Research *27*: 173-178 (1984).
Grabe, E.; Post, B.; Nolan, F.; Farrar, K.: Pitch accent realization in four varieties of British English. Journal of Phonetics *28*: 161-185 (2000).
Grammont, M.: Traité de phonétique (Delagrave, Paris 1933).
Grønnum, N.: Prosodic parameters in a
variety of regional Danish standard languages. Phonetica *47*: 188-214 (1991).
Gsell, R.: Remarques sur la structure de l'espace tonal en vietnamien du sud (parler de Saïgon). Cahier d'études vietnamiennes, Département de Langues et Civilisations de l'Asie Orientale de l'Université Paris 7 *4* (1980).
Han, M.S.; Kim, K.-O.: Phonetic variation of Vietnamese tones in disyllabic utterances. Journal of Phonetics *2*: 223-232 (1974).
Harris, J.G. States of the glottis of Thai voiceless stops and affricates. Essays in Tai linguistics: 3–11. K. Tingsabadh and A. S. Abramson. (Chulalongkorn University Press, Bangkok 2001).
Haudricourt, A.-G.: De l'origine des tons en vietnamien. Journal Asiatique *242*: 69-82 (1954).
Heffner, R.-M.S.: General Phonetics (The University of Wisconsin Press, Madison 1950).
Henderson, E.J.A. Marginalia in Siamese phonetic studies. In Honour of Daniel Jones: 415-424. D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott and J. L. M. Trim. (Longmans, Green and Co., London 1964).
Henderson, E.J.A. Feature shuffling in Southeast Asian languages. Southeast Asian Linguistic Studies presented to André-G. Haudricourt: 1-22. S. Ratanakul, D. Thomas and P. Suwilai. (Mahidol University, Bangkok 1985).
Henrich, N.: Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception (Ph. D., Université Paris 6, Paris 2001).
Henrich, N.; d'Alessandro, C.; Castellengo, M.; Doval, B.: 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 (2004).
Henrich, N.; Roubeau, B.; Castellengo, M. On the use of electroglottography for characterisation of the laryngeal mechanisms. Stockholm Music Acoustics Conference, Sweden (2003).
Henton, C.; Ladefoged, P.; Maddieson, I.: Stops in the World's Languages. Phonetica *49*: 65-101 (1992).

Hess, M.H.; Ludwigs, M.: Strobophotoglottographic transillumination as a method for the analysis of vocal fold vibration patterns. Journal of Voice *14*(2): 255-71 (2000).
Hollien, H.: On vocal registers. Journal of Phonetics *2*: 125-143 (1974).
Hollien, H.; Moore, P.; Wendahl, R.W.; Michel, J.F.: On the nature of vocal fry. Journal of Speech and Hearing Research *9*: 245-247 (1966).
Howard, D.M.: Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. Journal of Voice *9*(2): 163-72 (1995).
Howard, D.M.; Lindsey, G.A.; Allen, B.: Toward the quantification of vocal efficiency. Journal of Voice *4*(3): 205-12 (1990).
Iwata, R.; Sawashima, M.; Hirose, H.: Laryngeal adjustments for syllable-final stops in Cantonese. Annual Bulletin of the Research Institute for Logopedics and Phoniatrics *15*: 45-54 (1981).
Iwata, R.; Sawashima, M.; Hirose, H.; Niimi, S.: Laryngeal adjustments of Fukienese stops: initial plosives and final applosives. Annual Bulletin of the Research Institute for Logopedics and Phoniatrics *13*: 61-81 (1979).
Karnell, M.P.: Synchronized videostroboscopy and electroglottography. Journal of Voice *3*(1): 68-75 (1989).
Kohler, K.J.: Glottal stops and glottalization in German. Phonetica *51*: 38-51 (1994).
Kohler, K.J. Articulatory prosodies in German reduced speech. Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, University of California at Berkeley: 89-92 (1999).
Ladefoged, P.; Maddieson, I.: The Sounds of the World's Languages (Blackwell, Oxford, U.K. & Cambridge, Massachusetts 1996).
Laver, J.: The phonetic description of voice quality (Cambridge University Press, Cambridge 1980).
Martinet, A.: Economie des changements phonétiques (A. Francke AG Verlag, Bern, Switzerland 1955).
Martinet, A.: Evolution des langues et reconstruction (Presses Universitaires de France, Paris 1975).
Maspero, H.: Etude sur la phonétique historique de la langue annamite. Les initiales. Bulletin de l'Ecole Française d'Extrême-Orient *12*: 1-127 (1912).
Michailovsky, B.: La langue hayu (CNRS Editions, Paris 1988).
Michailovsky, B. Hayu. The Sino-Tibetan Languages: 518-532. G. Thurgood and R. LaPolla. (Routledge, London & New York 2003).
Nguyên Nhu Y: Dai tu diên tiêng Viêt - Comprehensive Vietnamese Dictionary (Nhà Xuât ban van hoa thông tin, Hanoi 1999).
Nguyên Van Loi; Edmondson, J.A.: Tones and voice quality in modern northern Vietnamese: instrumental case studies. Mon-Khmer Studies *28*: 1-18 (1997).
Niederer, B.: Les langues Hmong–Mjen (Miáo–Yáo), phonologie historique (Lincom Europa, München 1998).
Ní Chasaide, A.; Gobl, C.: Contextual variation of the vowel voice source as a function of adjacent consonants. Language and Speech *36*(2/3): 303-330 (1993).

Okell, J.A.: A Reference Grammar to Colloquial Burmese (School of African and Oriental Studies, London 1969).
Orlikoff, R.F.: Scrambled EGG: The Uses and Abuses of Electroglottography. Phonoscope *1*(1) (1998).
Pham, A.H.: Vietnamese Tone: A New Analysis (Routledge-Taylor and Francis, London/New York/Oslo/Singapore 2003).
Pham Thi Ngoc Yên; Castelli, E.; Nguyên Quôc Cuong. Gabarits des tons vietnamiens ["Vietnamese tone templates"; in French]. Journées d'Etude sur la Parole (XXIV), Nancy, France: 25-28 (2002).
Przedlacka, J.: Estuary English: glottaling in the Home Counties. Oxford University Working Papers in Linguistics, Philology and Phonetics *5*: 19-24 (2000).
Rothenberg, M.: A multichannel electroglottograph. Journal of Voice *6*(1): 36-43 (1992).
Rothenberg, M.; Mahshie, J.J.: Monitoring vocal fold abduction through vocal fold contact area. Journal of Speech and Hearing Research *31*: 338-51 (1988).
Roubeau, B.; Chevrie-Muller, C.; Arabia-Guidet, C.: Electroglottographic study of the changes of voice registers. Folia Phoniatrica *39*: 280-89 (1987).
Roubeau, B.; Chevrie-Muller, C.; Arabia-Guidet, C. Control of laryngeal vibration in register change. Vocal Fold Physiology: Acoustic, perceptual, and physiological aspects of voice mechanisms: 279-86. J. Gauffin and B. Hammarberg. (Singular Publishing Group, San Diego 1991).
Sagart, L.; Xu, S.: History through loanwords: the loan correspondences between Hani and Chinese. Cahiers de linguistique - Asie Orientale *30*(1): 3-54 (2001).
Scherer, R.C.; Arehart, K.H.; Guo, C.G.; Milstein, C.F.; Horii, Y.: Just noticeable differences for glottal flow waveform characteristics. Journal of Voice *12*(1): 21-30 (1998).
Smalley, W.A.: Manual of articulatory phonetics (Practical Anthropology, New York 1963).
Stewart, J.A.: Manual of Colloquial Burmese (Luzac & Co. Ltd, London 1955).
Thein Tun, U. Some acoustic properties of tones in Burmese. Papers in Southeast Asian Linguistics n°8: Tonation: 77-116. D. Bradley. (Pacific Linguistics, Series A, n°62 1982).
Thompson, L.C.: A Vietnamese Reference Grammar (University of Washington Press, 1965).
Thongkum, T.L. Phonation types in Mon-Khmer languages. Voice production: Mechanisms and functions: 319-333. O. Fujimura. (Raven Press, New York 1988).
Thurgood, G.: Vietnamese and tonogenesis: revising the model and the analysis. Diachronica *19*(2): 333-363 (2002).
Titze, I.R. Regulation of Vocal Power and Efficiency by Subglottal Pressure and Glottal Width. Vocal physiology: voice production, mechanisms and functions: 227-237. O. Fujimura. (Raven Press, New York 1988).
Titze, I.R.: Interpretation of the Electroglottographic signal. Journal of Voice *4*(1): 1-9 (1990).

Vu Ngoc, T.; d'Alessandro, C.; Rosset, S. A phonetic study of Vietnamese tones : acoustic and electroglottographic measurements. International Conference on Speech and Language Processing, Boulder, Colorado (2002).
Watkins, J. Closed quotient of laryngeal gestures and settings in Wa. XIVth International Congress of Phonetic Sciences, University of California at Berkeley: 1017-1021 (1999).
Watkins, J.: The Phonetics of Wa (Australian National University, Canberra 2002).
Yuan Jiahong; Shih Chilin; Kochanski, G.P. Comparison of Declarative and Interrogative Intonation in Chinese. Speech Prosody 2002, Aix-en-Provence: 711-714 (2002).

Tables

Table 1. The tones of Vietnamese in diachronic perspective.

Table 1a: Late Proto-Viet-Muong rhyme types:

nonglottalized open syllable; glottally

constricted ending; /h/ ending; /p/ /t/ /k/ ending

[after Ferlus 2001:2].

ta	taˀ	tah
da	daˀ	dah

tap, tat, tak
dap, dat, dak

Table 1b: the first stage of tonogenesis in

Vietnamese: three tones; no distinctive tone on

obstruent-final syllables.

ta	tone	ta	tone	ta	tone
da	A	da	B	da	C

tap, tat, tak	category D
dap, dat, dak	(no tone)

Table 1c: the tone system of present-day Hanoi Vietnamese: number and traditional name in

Vietnamese writing. Tones A1 to C2: appear only on sonorant-final syllables; tones D1 and D2: appear

only on obstruent-final syllables.

ta	tone A1 (ngang)	ta	tone B1 (sắc)	ta	tone C1 (hỏi)
ta	tone A2 (huyền)	ta	tone B2 (nặng)	ta	tone C2 (ngã)

tap, tat, tak	tone D1 (sắc)
tap, tat, tak	tone D2 (nặng)

Table 2. Statistics on fundamental frequency values. Each value is the average over one set of syllables of same tonal category (42-syllable sets).

speaker	tone/ syllable type	mean F_0 (Hz) and standard deviation, cond. C	mean F_0 (Hz) and standard deviation, cond. E	difference : E minus C, in % (left) and in musical tones (right)		part of the syllable (out of the 100 equally spaced samples) over which the difference is significant	difference of F_0 standard deviation: E compared to C
F1	tone B2 (nasal-final)	228 (12.9)	240 (25.7)	+5.4	½	38th to 97th samples	+64%
M1		129 (10.0)	168 (12.2)	+30	2 ½	1st to 100th (=all samples)	+13%
M2		115 (9.4)	126 (14.1)	+9.8	¾	27th to 98th samples	+71%
M3		128 (9.2)	131 (15.7)	+2	under ¼	3rd to 21st (E: lower) and 52nd to 100th (E: higher)	+64%
F1	tone D2 (obstruent-final)	224 (8.5)	248 (12.7)	+11	close to 1	1st to 96th	+37%
M1		126 (4.5)	144 (4.8)	+15	1 ¼	1st to 100th (=all samples)	+7%
M2		107 (6.0)	131 (9.7)	+22	1 ¾	1st to 99th	+56%
M3		121 (5.4)	124 (5.9)	+2	under ¼	72nd to 100th	-16%

Table 3. Statistics on the difference between highest and lowest fundamental frequency in the syllable (for short: "F_0 decrease"). Each value is the average over a 42-syllable set. The letter "s" stands for "significant", "ns" for "not significant".

speaker	tone/ syllable type	F_0 decrease (in musical tones)		results of t-test on F_0 decrease (in brackets: t_{41} value and p value)	Proportion of total syllable duration over which the decrease in F_0 takes place (%)		results of t-test on duration of F_0 decrease (in brackets: t_{41} value and p value)
		Careful reading	Emphatic reading		Careful reading	Emphatic reading	
F1	tone B2 (nasal-final)	2.30	1.95	s (-4.87; $2*10^{-5}$)	77	86	s (2.09; $p > 0.03$)
M1		2.26	2.30	ns (-1.50; 0.14)	67	63	ns (0.21; $p > 0.5$)
M2		1.93	2.42	s (-13.85; $< 10^{-9}$)	69	62	s (3.82; $p < 10^{-3}$)
M3		2.11	2.09	s (-3.50; 0.001)	62	75	ns (0.69; $p > 0.4$)
F1	tone D2 (obstruent-final)	5.18	6.20	ns (1.42; 0.16)	72	63	ns (0.16; $p > 0.5$)
M1		5.0	4.6	ns (-1.81; 0.07)	61	59	ns (-0.07; $p > 0.5$)
M2		4.0	7.1	s (-4.85; $2*10^{-5}$)	70	41	s (3.76; $p < 10^{-3}$)
M3		5	5.8	ns (-0.25; 0.80)	53	48	ns (0.36; $p > 0.5$)

Table 4. Statistics on open quotient values. Each value is the average over a 42-syllable set.

speaker	tone/ syllable type	mean open quotient (%), cond.C	mean open quotient (%), cond. E	difference : E minus C, in %	part of the syllable (out of the 100 equally spaced samples) over which the difference is significant	stand. dev. of E compared to C
F1	tone B2 (nasal-final)	50	47	-4.6%	69 to 100	+8%
M1		32	32	-0.3%	38 to 48	-6%
M2		43	38	-11%	57 to 100	+13%
M3		38	37	-2.9%	77 to 100	+27%
F1	tone D2 (obstruent-final)	63	64	+1.8%	80 to 100	-30%
M1		51	53	+5.6%	7 to 96	+16%

M2		53	56	+5.6%	55 to 100	+20%
M3		57	60	+6.0%	8 to 100	-18%

Table 5. Statistics on syllable length. Each value is the average over a 42-syllable set. The letter “s” stands for “significant”, “ns” for “not significant”.

speaker	tone/ syllable type	mean length (ms), cond.C	mean length (ms), cond. E	difference: E minus C, in %	results of t-test on syllable length (in brackets: t_{41} value and p value)
F1	tone B2	126	143	+13%	ns (-1.94 ; $p > 0.05$)
M1	(nasal-final)	145	149	+3%	ns (-1.26; $p > 0.2$)
M2		126	210	+67%	s (13.0; $p < 10^{-12}$)
M3		166	183	+10%	ns (-1.61; $p > 0.1$)
F1	tone D2	121	107	-12%	s (7.24; $p < 10^{-7}$)
M1	(obstruent-	136	144	+6%	s (-3.08; $p < 0.005$)
M2	final)	131	139	+6%	ns (-1.13; $p > 0.2$)
M3		145	119	-18%	s (10.4; $p < 10^{-12}$)

Table 6. A summary of differences from condition C to condition E; hs: “highly significant”, $p < 10^{-4}$; s: “significant”, $p < 0.05$; blank: not significant, $p > 0.05$.

speaker	tone	mean F_0	amplitude of F_0 decrease	duration of F_0 decrease	syllable length	open quotient: part of the syllable over which the difference is significant
F1	4			s		less than 1/3rd
	B	s	hs		hs	less than 1/3rd
M1	4	hs	s			less than 1/3rd
	B	hs			s	about 90%

M2	4	s	hs	s	hs	about 50%
	B	s	s	s		about 50%
M3	4					less than 1/3rd
	B		s		hs	about 90%

Legends to figures

Figure 1. Schematic diagram of Hanoi Vietnamese tones, based on data from speaker M1.

Figure 2. Examples of the EGG and DEGG signals used in F_0 and open quotient calculation: top: EGG, middle: DEGG, bottom: smoothed DEGG.

Figure 3. Voice offset under tone B2. Lines: closing instants, arrows: opening instants. EGG (top), DEGG (mid), audio (bottom).

Figure 4. Voice offset under tone D2. Lines: closing instants, arrows: opening instants. EGG (top), DEGG (mid), audio (bottom).

Figure 5a. Individual curves for tone B2. Speaker M1, 42 items. Careful reading.

Figure 5b. Average curves for tone B2. Speaker M1. Conditions C and E.

Figure 6a. Individual curves for tone D2; speaker M1, 42 items. Careful reading.

Figure 6b. Averaged curves for tone D2. Speaker M1. Conditions C and E.

Figure 7. Averaged curves for tone D1. Speaker M1. Conditions C and E.

Figure 8. Syllable /ɓǎn/ with tones A2 and B2, illustrating the loss of visible final nasal under tone B2.

Figure 9. Contrast between oral and nasal velar finals with tone D2 and tone B2, respectively.

Figure 10. An example of laryngealized voice offset (tone: C1).

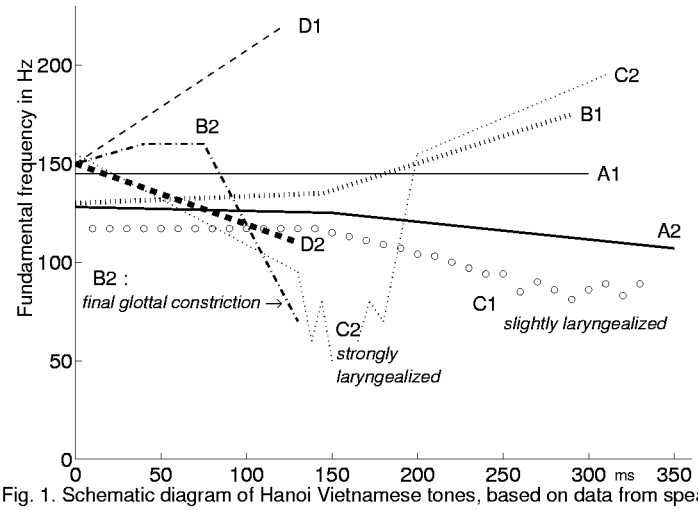

Fig. 1. Schematic diagram of Hanoi Vietnamese tones, based on data from speaker M1

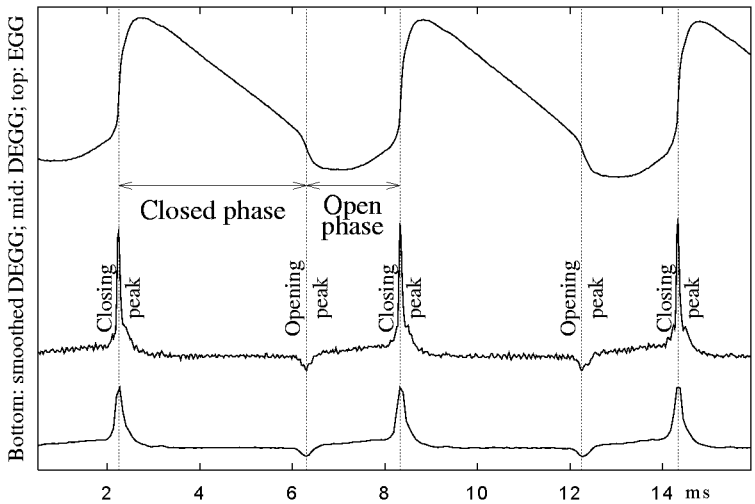

Figure 2. Examples of the EGG and DEGG signals used in Fo and Oq calculation: top: EGG, middle: DEGG, bottom: smoothed DEGG.

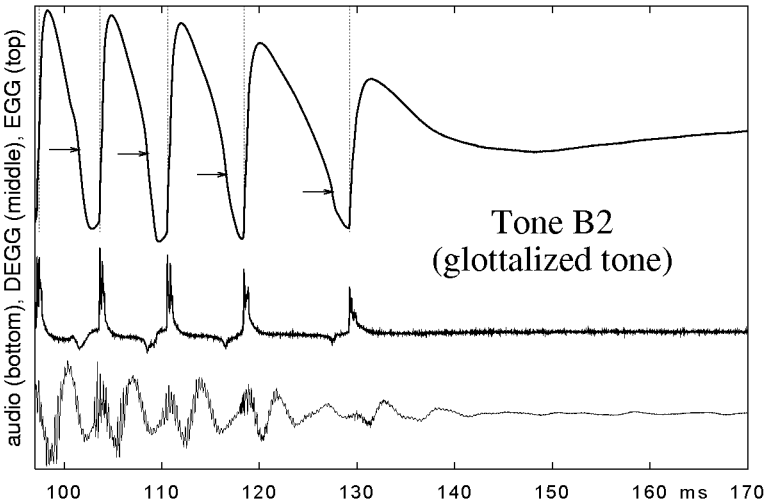

Figure 3. Voice offset under tone B2. Lines: closing instants, arrows: opening instants. EGG (top), DEGG (mid), audio (bottom).

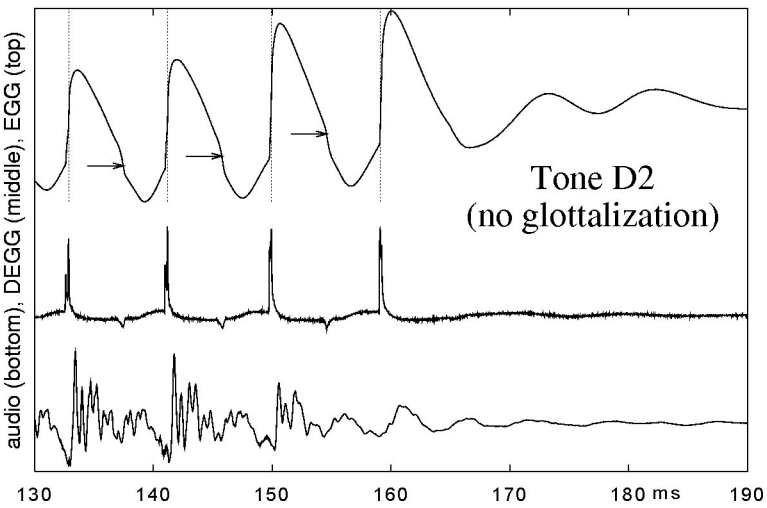

Figure 4. Voice offset under tone D2. Lines: closing instants, arrows: opening instants. EGG (top), DEGG (mid), audio (bottom).

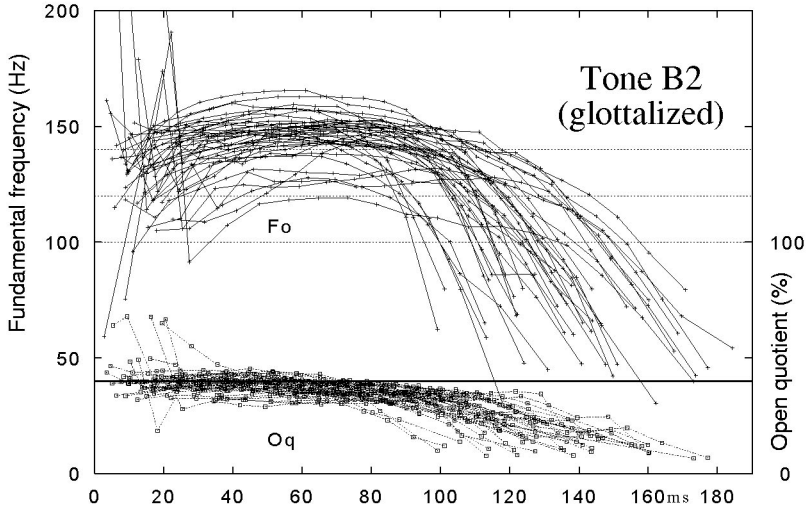

Fig. 5a. Individual curves for tone B2; speaker M1, 42 items. Careful reading.

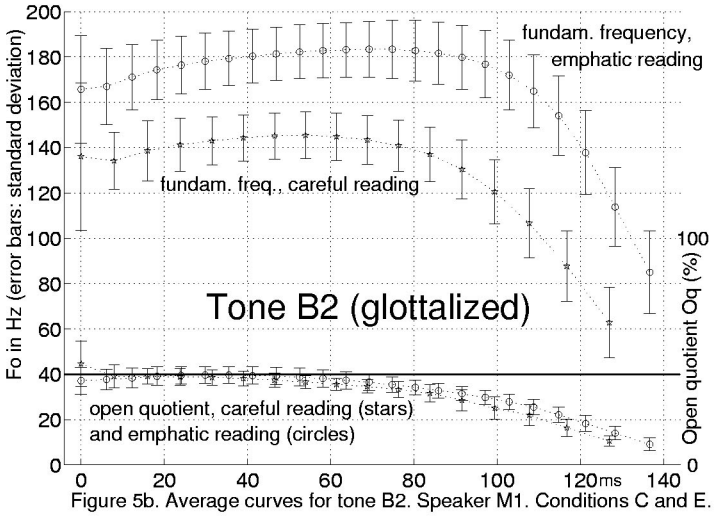

Figure 5b. Average curves for tone B2. Speaker M1. Conditions C and E.

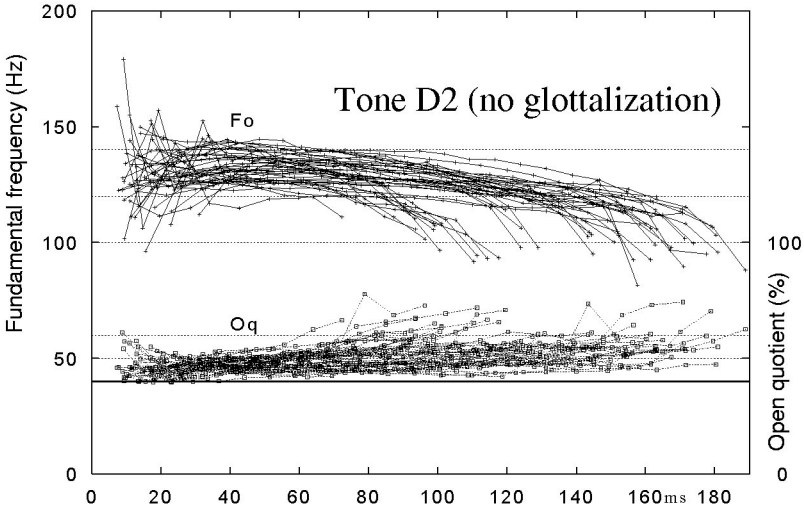

Fig. 6a. Individual curves for tone D2; speaker M1, 42 items. Careful reading.

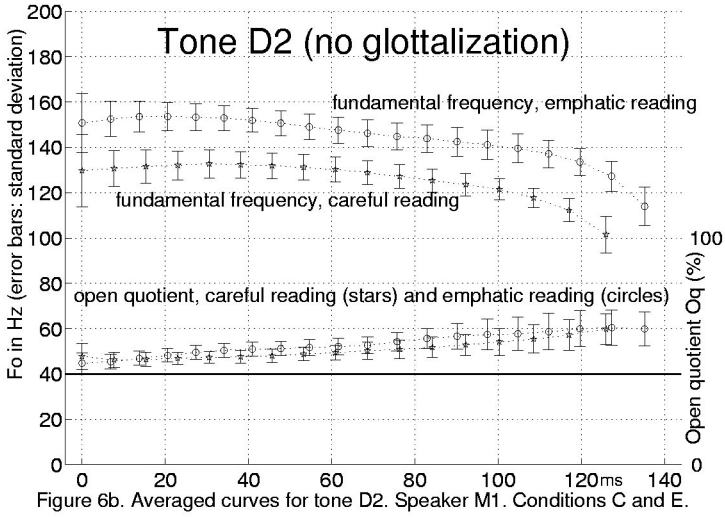

Figure 6b. Averaged curves for tone D2. Speaker M1. Conditions C and E.

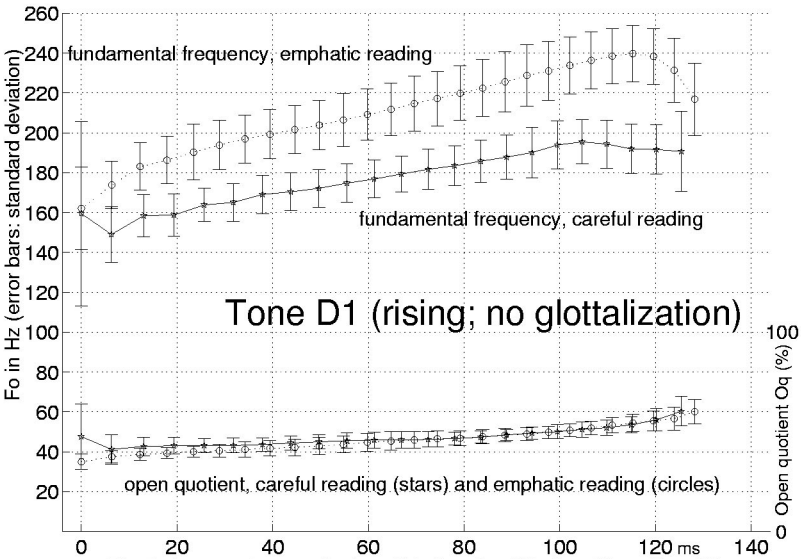

Fig. 7. Averaged curves for tone D1. Speaker M1. Conditions C and E.

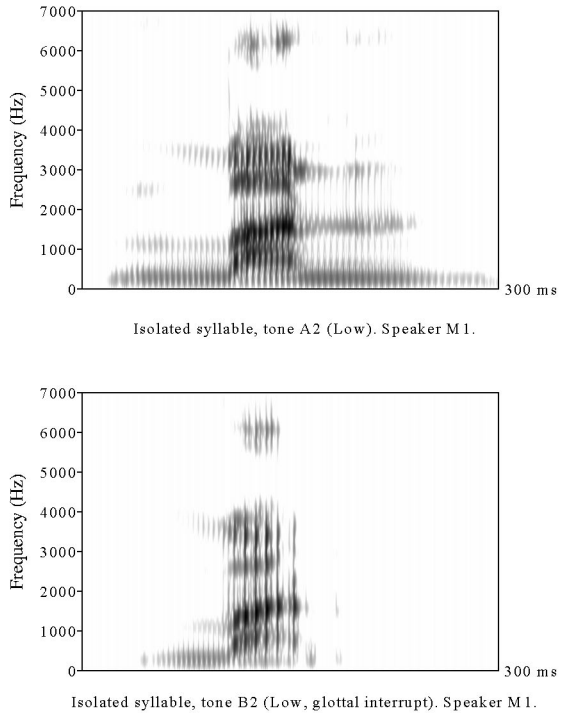

Fig. 8. Syllable /ɓan/ (short vowel a) with tones A2 and B2 illustrating the loss of visible final nasal under tone B2.

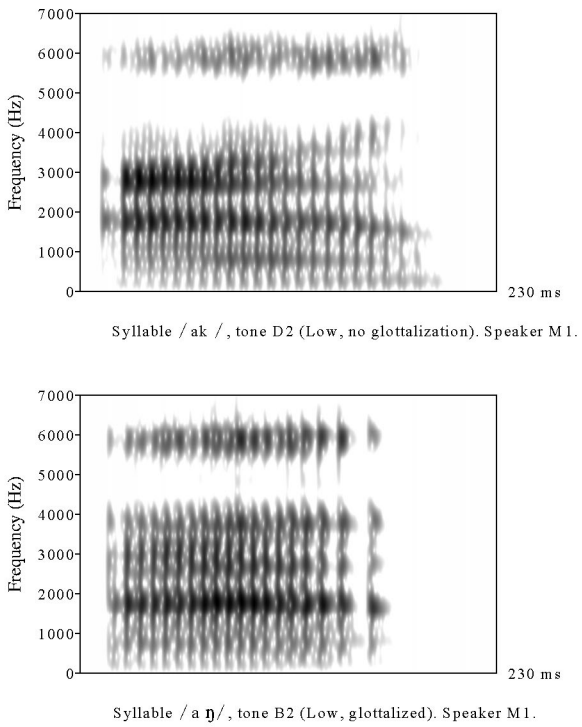

Figure 9. Contrast between oral and nasal velar finals with tone D2 and tone B2, respectively.

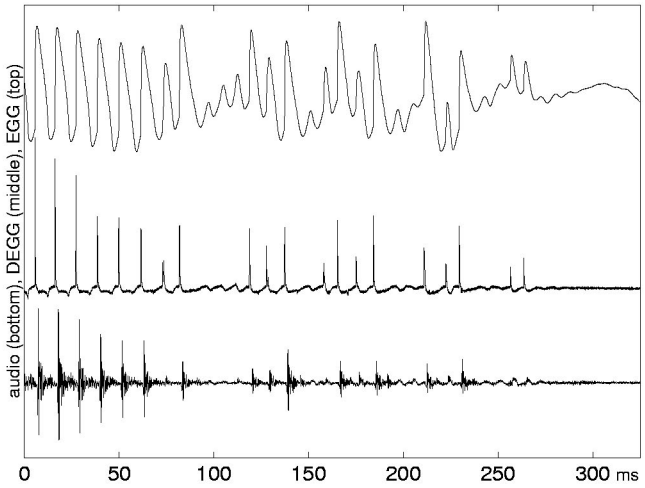

Figure 9. An example of laryngealized voice offset (tone: C1).