Tones and voice quality in modern northern Vietnamese: Instrumental case studies[1]

NGUYEN Van Loi
National Centre for Social and Human Sciences

Jerold A. EDMONDSON
University of Texas at Arlington

1. Introduction

Vietnamese was not at first recognized as a member of the Mon-Khmer language family. But the status of Vietnamese is today no longer in doubt. It has been generally accepted that Vietnamese is a MK language after the proposal describing the mechanism of tonal development in Vietnamese was outlined by the late Professor André Haudricourt (1954). Subsequent support for Haudricourt's theory of tonal development in languages related to Vietnamese, notably Thavưưng and Rục, has been provided by Michel Ferlus, Gerard Diffloth, Nguyễn Văn Lợi, and others. The stumbling block to early recognition of the Mon-Khmer heritage of Vietnamese was its rich system of six fully developed tones, traditionally called by names that themselves exhibit that tone—*ngang* (level), *huyền* (falling), *sắc* (rising), *hỏi* (fall-rise), *ngã* (rising with a glottal interrupt within the syllable), and *nặng* (falling with strong continuous constriction to stoppage)

Mon-Khmer languages have usually been remarked upon not for the linguistic category of *tone* but for the category of *register*, which includes prominent voice quality as a contrastive feature. Though Vietnamese is not a classic register language in the sense of Henderson (1952), voice quality, as well as pitch phonomena, are both clearly to be observed. One widely cited work that mentions voice quality in Vietnamese is Thompson (1965:40-1), in which these prosodic effects are described in some detail, as follows:

Sắc tone is high and rising (perhaps nearly level at the high point in rapid speech) and tense.

Ngã tone is also high and rising (in other words, the contour is roughly the same as that of *sắc*), but it is accompanied by the rasping voice quality occasioned

[1]This paper is dedicated to the memory of Professor André Georges Haudricourt. Support for a part of this research was provided to the second author by the National Endowment for the Humanities 1995 and the National Science Foundation 1996 grants for the Project, "Languages of the Vietnam-China Borderlands". The second author would also like to express appreciation to Drs. Kenneth J. Gregerson, Norris McKinney, Paul Griffith, and Mr. Greg Lyons, for comments and technical support in regard to equipment and software.

by tense glottal stricture. In careful speech such syllables are sometimes interrupted completely by a glottal stop (or a rapid series of glottal stops).

Ngang tone is lax; in contour it is nearly level in non-final syllables not accompanied by heavy stress, although even in these cases it probably trails downward slightly.

Huyền tone, also lax, starts quite low and trails downward toward the bottom of the voice range. It is often accompanied by a kind of breathy voicing, reminiscent of a sigh.

Hỏi tone is tense; it starts somewhat higher than *huyền* and drops rather abruptly. In final syllables, and especially in citation forms, this is followed by a sweeping rise at the end, and for this reason it is often called the "dipping" tone. However, non-final syllables seem only to have a brief level portion at the end and this is exceedingly elusive in rapid speech.

Nặng tone is also tense; it starts somewhat lower than *hỏi*. With syllables ending in a stop [p, t, ch, k] it drops only a little more sharply than *huyền* tone, but it is never accompanied by the breathy quality of that tone.

Thompson's description is not too different from perhaps the first documented phonetician of Vietnamese, Alexandre de Rhodes (1651), who distinguished six classes of tones and described them in the following impressionistic terms: "acute-angry" (sắc), "smooth-rising" (hỏi), "level" (ngang), "chesty-raised" (ngã), "chesty-heavy" (nặng), "grave-lowering" (huyền), cf. Gregerson (1969:46). Three and one half centuries later these characterizations still remain remarkably apt and insightful.

Among Vietnamese phoneticians, *Ngữ Âm tiếng Việt* [Vietnamese phonemes] (1977:134) by Đoàn Thiện Thuật notes that Emeneau uses the term "creaky" to describe *ngã* and *hỏi* tones, suggesting that they are similar to the pharyngealization found in Danish. The author Lê Văn Lý, also opines that in these tones there are pharyngeal stops. He describes the phenomenon as follows: when the Vietnamese speaker speaks, the vocal folds vibrate with a certain frequency. In general the frequency drops and the vocal folds close when the airflow from the lungs stops. But, the vocal folds do not stop completely, rather they continue for some time with a slower vibration. One can feel the stoppage in the pharynx and hear a small pharnygeal stop that is half voiced. Professor Thuật notes, furthermore, that for all three of the Vietnamese tones *ngã, nặng,* and *sắc* many authors suggest the presence of a sound similar to a glottal stop. He notes that the authors Hoàng Tệ and Hoàng Minh (1975) speak in this regard about "friction glottale". Another important work on Vietnamese voice quality is Phạm (1973), who studies the Vietnamese tones using sound spectograms and comments on the voice quality from spectographics evidence. Mostly this work treats pitch trajectories. In later work, Vũ Thanh Phương (1986) examined 34 speakers and found that four phonetic features distinguished the three major geographic types of Vietnamese (northern or NV, central or CV, and southern or SV): FØ, intensity, duration, and laryngealization. The first three features were studied instrumentally, but voice quality was determined solely from auditory impressions. He notes (page 74) that ...three NV informants display marked differences in the broken (ngã) tone: NF3 (had) heavy laryngealization, sharp drop in FØ and in intensity at middle; NM3 (had) no laryngealization, sharp drop in FØ and slight drop in intensity at middle and NM4 (had) no laryngealization, no sharp drop in FØ but only sharp drop in intensity at middle. For NM4 it appears that the FØ curves of the rising (sắc) and broken tones are similar; only the intensity contours differ sharply.

As one can see, descriptions of Vietnamese voice quality are diverse and possibly not all compatible. Further confounding the issue is considerable variation across speakers of the language, as we discovered. Some speakers use differing voice qualities in their phonological contrasts in a relatively more exaggerated manner, whereas others seems to rely mostly on pitch (tone) as the prime signaler of prosodic contrast.

2. Methods

Our major interest in this study is to provide greater precision in description of Vietnamese voice qualities found in its six tones by using the computer-aided instrumental technique of Inverse Filtering. We will also present our results of an analysis of tonal contrasts for both contemporary Hanoi speakers and certain other northern speakers.

2.1 The speakers

In this study we employed six speakers of northern Vietnamese, including the first author of this paper. These speakers were:

Nguyễn Văn Lợi	male	50	native of Nam Định Province
Nguyễn Thu Hải	female	24	native of Hà Nội
Nguyễn Đức Tồn	male	44	native of Nam Định Province
Phạm Hùng Việt	male	45	native of Hà Nội.
Vũ Kim Bảng,	male	42	native of Vình Phú Province
Nguyễn Thị Hòng	female	36	native of Hà Nội

All of the speakers have lived and worked for many years in Hanoi. Utterances were recorded in the congenial and familiar circumstances of the Linguistics Institute of the National Centre for Social and Human Sciences, Hanoi.

2.2 The recording

Acoustic data for tone analysis were recorded directly into the computer using an ATR-20 microphone and CECIL hardware box (JAARS Inc., Waxhaw, NC) for A-D conversion. Airflow data were collected with the Rothenberg Mask connected to a small battery-operated amplifier of our own construction, a CECIL Box, and a Toshiba Satellite Pro 486-DOS computer, cf Figure 1. The computer was running Cecil 2.1 (DOS CECIL) for the capture of the airflow wave. In later analysis we used Wincecil 2.1a (CECIL for MS-Windows 3.11) with some additional routines for digital inverse filtering developed for us by Greg Lyons. The additions to the Wincecil software allow a reconstruction of the original glottal source wave by the technique we will now describe.

2.3 *Pitch extraction*

Each speaker was asked to produce three tokens of the sample syllable *ta* in each of the six tones, i.e.: *ta* 'we'; *tà* 'fall', *tá* 'units of 12', *tả* 'describe'; *tã* 'diaper'; and *tạ* '100 kg'. The Wincecil program then extracts a pitch track from each of the recorded syllables and displays it on the computer screen. From these we made composite plots of the three recorded repetitions using some locally developed software that represents the "mean" values of the pitch tracks after they have been normalized for length and relative pitch height. The composite tracks are shown in the Results Section below.

2.4 *Inverse filtering*

Voice quality is largely, but not entirely, the result of a speaker's vocal folds muscle settings during speech production, which we cannot readily observe directly. Therefore, we must use a technique such as inverse filtering that allows us indirectly to capture the shape of the wave generated when air flows through the glottal opening producing sound. From these effects we can infer much about the settings that yielded such a wave.[2]

Inverse filtering relies on the Source-Filter Theory of Speech Production (cf. Fant 1961; Fant and Sonesson 1962; Lindqvist 1965). This theory treats the air pressure wave exiting the mouth as the coproduct of the *glottal source wave* (source), the *resonance* (filter) caused by the geometry of pharyngeal and oral cavities, and the *radiation* out into the atmosphere. By using the Rothenberg Mask and inverse filtering, one can for all intents and purposes cancel the influence of the vocal tract configuration and recover the glottal source wave. At the same time, the mask allows us to capture all the airflow discharge into the atmosphere and, consequently, leaves no additional radiation effect to be considered. Schematically, this process may be stated as follows:

Airflow Wave
- Resonance Filter
- Radiation

= Glottal Source Wave.

The instrumental setup is shown below in Figure 1.

---

[2] Inverse filtering has not been used on a wide variety of languages from Asia, because it has been regarded as somewhat complex in non-laboratory conditions. It was used by Marie Huffman to study the Hmong language.

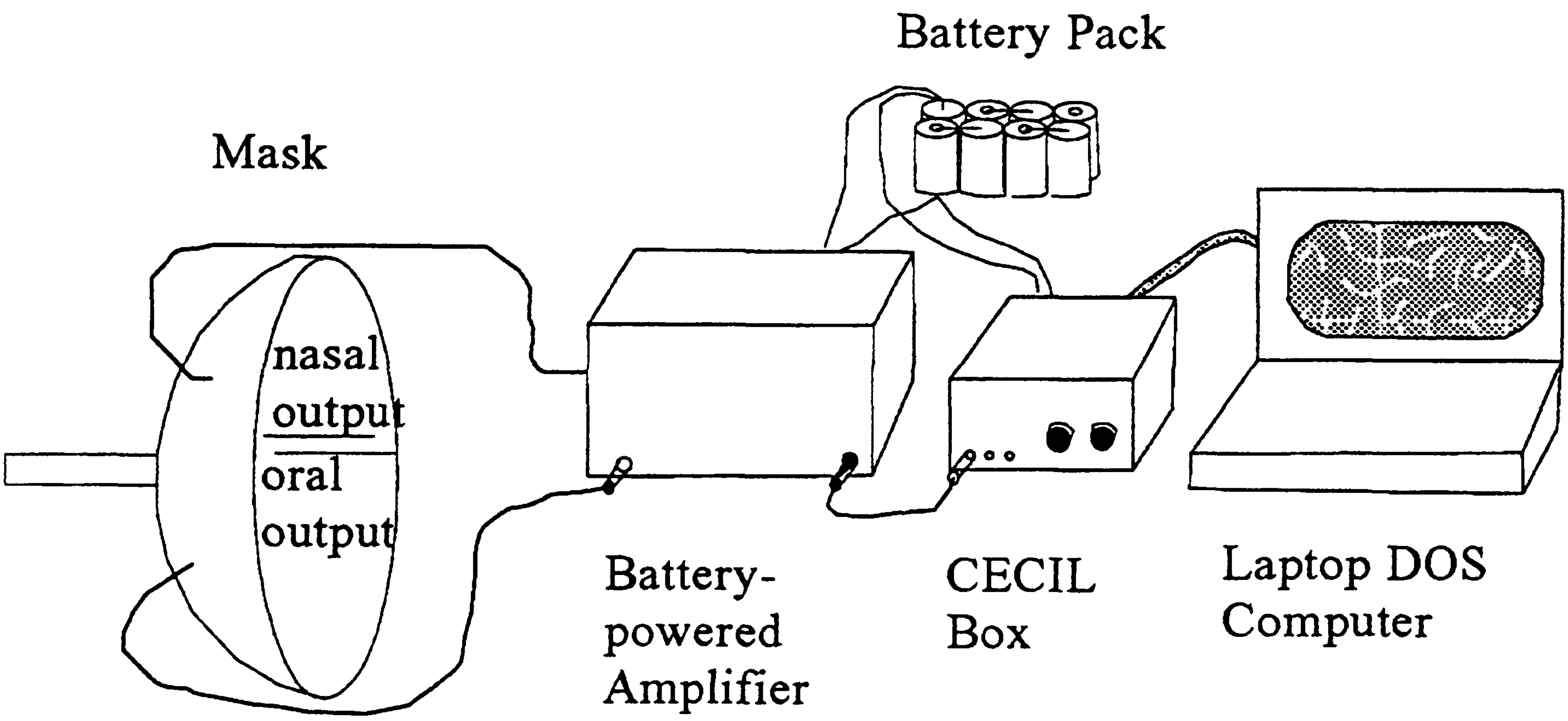

Figure 1. Equipment configuration for inverse filtering

We note further that the CECIL hardware box in Figure 1 has been modified slightly so that the signal from the amplifier enters the A-D circuitry immediately without further processing by the CECIL box. There are instructions in the hardware box manual on how this change can be effected. This adjustment is necessary due to the fact that airflow in frequency (unlike acoustic waves) may approach 0 Hz, because some airflow events happen much slower than acoustic events. The box and computer accept the signals just as any other wave coming into the input. These waves do, however, have a somewhat unusual appearance on the screen as the airflow waveform is not symmetrical about the baseline. Indeed, the asymmetry is expected as there should be net outflow of air, which will result in more of the wave above the line than below.[3] Once an airflow wave has been digitized and saved in the computer, filtering analysis can take place as described below. The waveform for modal voice has a general profile as in Figure 2.

[3]We will not address here the complex and much discussed issue of whether the Vietnamese *b* and *đ* are imploded sounds in the sense that there is ingressive airflow at the mouth.

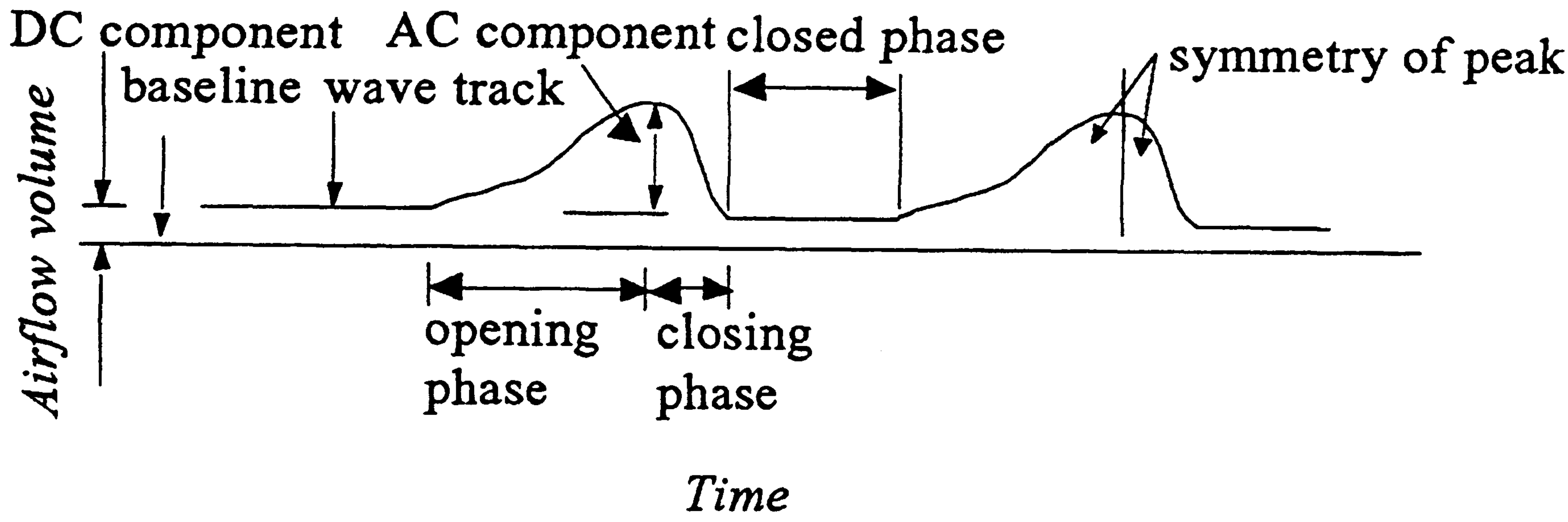

Figure 2. Idealized glottal wave for modal voice

2.5 Modal voice

Studies of modal voice (existent presumably in all languages) have found the distinctive configuration sketched in Figure 2 (cf. Laver 1980:109-18). This configuration reflects a specific pattern of glottal vibration. Laver (110) quotes Van den Berg in saying that modal voice which has large amplitude of the vocal fold muscles or vocalis at low pitches requires small passive longitudinal tension in the vocal folds which are short and thick, and an increase in active longitudinal tension in the vocalis muscle increases the pitch. Thus, the vocal fold vibration is regularly periodic, efficient in producing vibration, without audible friction from incomplete closure of the glottis. Relating what Van den Berg has said to the wave in Figure 2, we would use an electrical metaphor distinguishing DC vs. AC terminology.

Note first from left to right that there is a relatively stable steady-state *(DC component)* to the airflow. It is indicated by the constant horizontal line above the baseline that is the stable net outflow of air during production of speech sounds; we assume that modal voice does not involve ingressive airstream. In terms of airflow, the DC component corresponds to the stream of glottal air that is flowing at a more or less constant rate. It can be affected by changing the settings for *passive tension* of the cricothyroid muscles that put longitudinal tension on the vocal folds by stretching them.

The waveform then shows an *opening phase* or gradual increase in the airflow volume, as a jet of air under subglottal pressure separates the lips and the glottal folds open. At their widest aperture the process reverses, the local air pressure turns negative, sucking the lips back together, and the airflow volume shrinks to the level of the stable eggressive air in a process called the *closing phase*. We may liken the opening and closing parts of the speech wave to an electrical *AC component*. The height or amplitude of the AC component above the stable DC level is determined by the subglottal air pressure and the *active tension* in the vocalis or vocal fold muscles. It is important to note that the wave of the AC component is typically not symmetrical about its peak. The asymmetry is believed to be caused by a non-uniformity in the profile of the edges of the vocal folds that causes them to open more slowly and close more rapidly. If the vocal folds have low active tension, then the asymmetric cross section changes to a symmetric one and breathy voice quality can result.

Finally, note that in modal voice after the opening and then the closing phase there is a subinterval of closed folds *(closed phase)* with only DC airflow; the trace is a nearly invariant horizontal line. In summary, the glottis in modal voice has three distinct phases that recur in cyclical fashion, opening, closing, and closed. The manner of vocal fold vibration is what produces differences in voice quality. Different manners of vocal fold vibration will deviate from this ideal in ways that reveal a different voice quality and different glottal settings. For discussion of modal voice, tense voice, creaky voice, breathy voice in great detail, cf. Laver (1980).

As noted above, the airflow wave exiting the mouth consists of the glottal wave plus the modification caused by the tube-like features of the throat and mouth in forming vowels. The third component, the radiation into the atmosphere, is eliminated by using the Rothenberg Mask which captures all the exiting air. Inverse filtering basically amounts to reconstructing the original glottal source wave by cancelling out the effects of resonance. Thus, the item "inverse" filtering, is intended to indicate that the changes caused by the speaker's throat and mouth have been reversed. To state it somewhat jocularly, inverse filtering is like electronically cutting someone's head off to look at their vibrating glottis (do not do this at home!).

2.6 The filtering process

But before one can examine the glottal wave locked up inside the airflow wave exiting the mouth, it is necessary first to remove the resonance. That can be accomplished by putting a "zero" in all frequency locations where a resonance pole exists in the airflow wave (locations of bunching of acoustic energy in the spectrum from 0-4000 Hz or so). One must certainly put zeros in the places where there are vowel formants, F1, F2, F3, and F4 (higher formants can be treated as a group). Moreover, the mask may cause some changes in the formant values one would expect from vowels produced without the mask. Thus, one will need to determine the locations of all resonant energy concentrations (Hz). The utterances of the six speakers were subjected to inverse filtering analysis as indicated above. The summary of the results of this process is given in section 3 below.

3. Results

We were not surprised to find that Northern Vietnamese speech varieties distinguish six tones and that some of these tones often involve voice quality contrasts.

3.1 The tones

As has been traditionally stated, we found six contrastive tones in the speech of the people we examined. The figures below show the voiced part of the syllable of one speaker. Since the examples all started with *t-*, the voicing does not commence exactly at 0 msec but at a point somewhat later. In Figure 3 Tone 1 (*ngang* Tone) is shown. It has a slightly falling nature from a value of 45 semitones which falls to 44 semitones. We would assign it a value of 33 in the system of Chao's scale five-point for transcribing tones (see such a 1-5 scale for this speaker as the right vertical staff of this chart). Tone 2 (*huyền* tone) is lower than *ngang* tone, beginning at 38 semitones and falling to 36 semitones; we would assign it a Chao scale value of 21.

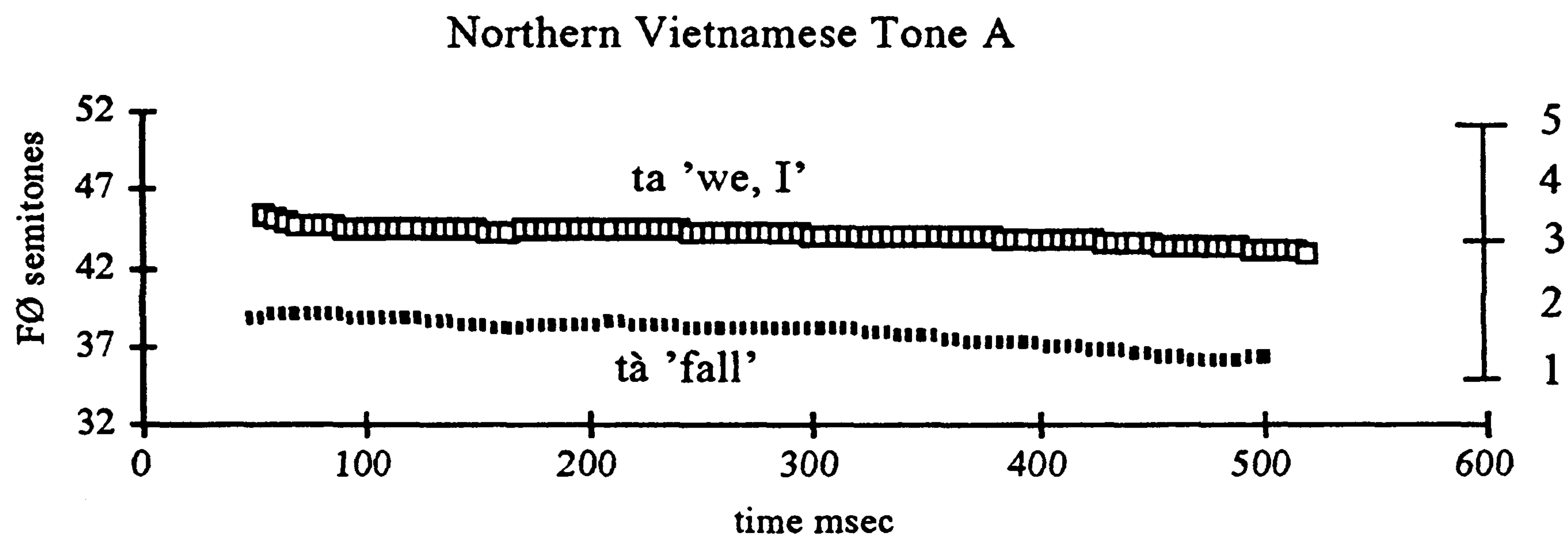

Figure 3. Tone plot of *ngang* (upper) and *huyền* (lower) in Northern Vietnamese

The B tones were also as expected. *Sắc* began at a level of 42 semitones and rose to a value of about 48, which corresponded to the top of this speaker's tone space. We would assign it a value of 35 on the Chao scale. *Nặng* began at almost the identical height of 42 semitones and fell to about 38 semitones. It was also much shorter than the other tones. Were it not for the abrupt or truncated nature of this tone, its downward trajectory could have carried it to a level of say 37 semitones. We would, therefore, have to assign it a tone value of 32, but with a tendency to go lower.

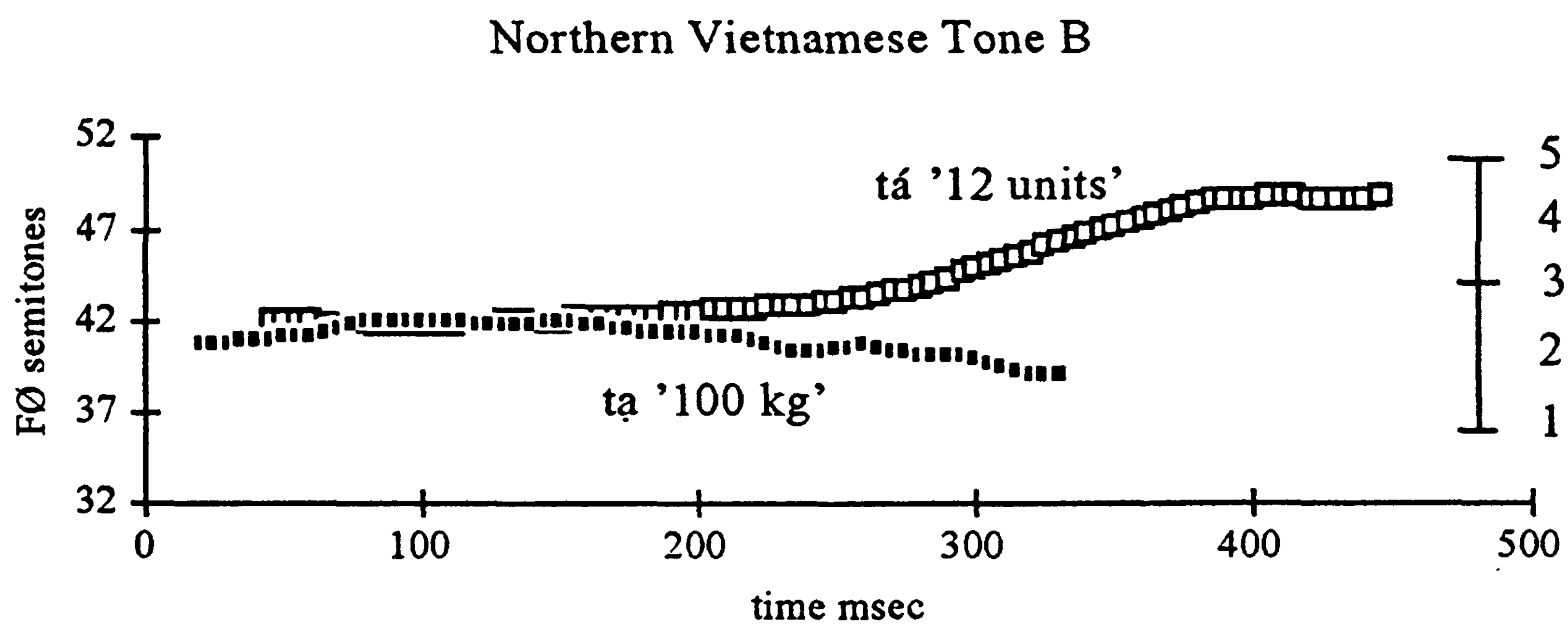

Figure 4. Tone plot of *sắc* (Upper) and *nặng* (Lower) in Northern Vietnamese

The *Ngã*, as shown in Figure 5 upper track, began at the level of 44 semitones and rose to the top of the range. As is further evident here, there is a characteristic break in the voicing at about 225 msec into the syllable. By contrast, the "dipping" feature mentioned by Thompson is evident in the *hỏi* tone (lower plot). It begins at 42 semitones and falls to 36 semitones only to rise again to about the level of the beginning.

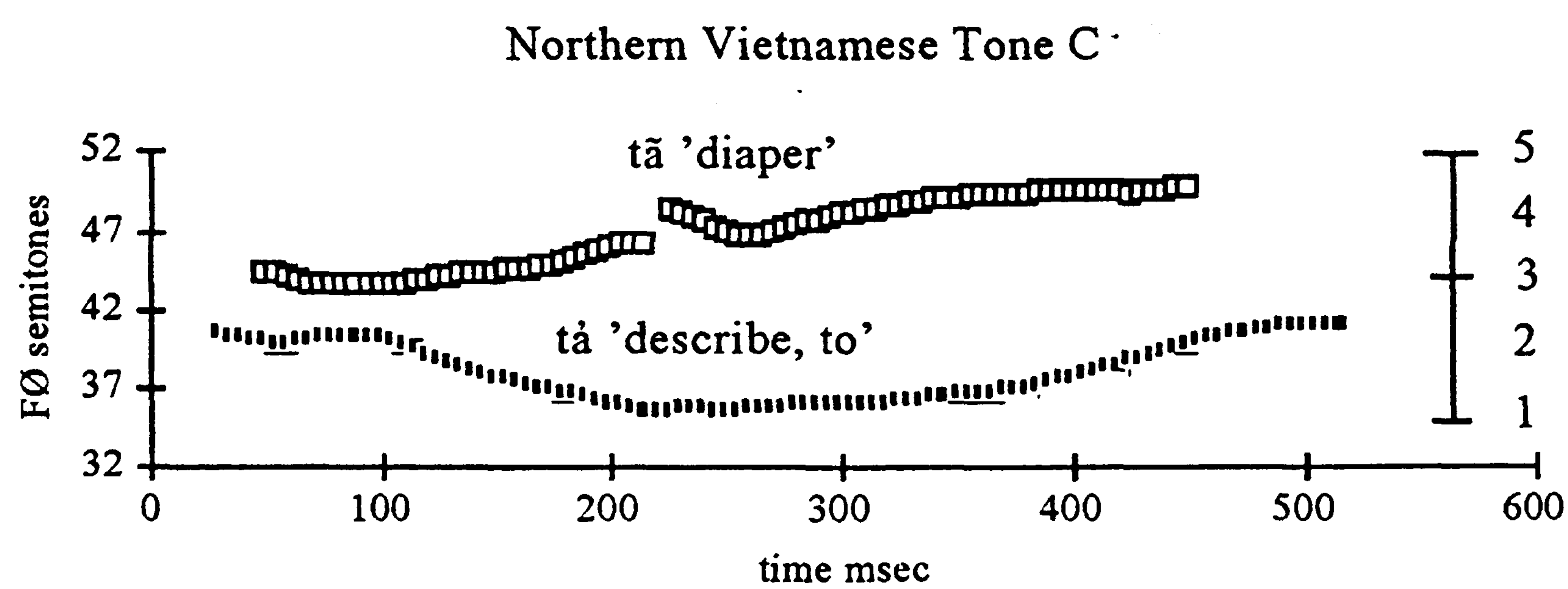

Figure 5. Tone plot of *ngã* (Upper) and *hỏi* (Lower) in Northern Vietnamese

We have presented the tones grouped into pairs with the label Tone A, B, and C. Haudricourt (1954) suggested that from a situation of no tones the three tone classes A, B, and C historically developed from pitch adjustments associated with different classes of codas at the end of the syllable. The A tone vocabulary consisted of original open or CV syllables (and modal voice) with basically level pitch prosodies. The B tone resulted from syllables that ended in a glottal stop *-ʔ* or an oral stop, i.e. today *-p, -t,* or *-k*.[4] The sharp closure of stops was accompanied by laryngeal tension and this led to rising pitch. The C tone occurred when the original syllable ended in a final *-s* or perhaps at a later time a final *-h*. Sagart (1988) revises Haudricourt a bit by suggesting that *-h* went on to develop into creaky voice quality before the tonogenetic moment occured, which then later left falling trajectories.

At a later time the original three tones evolved into six tones by means of a sound change emanating from the initial consonants called *tone splitting*. The old voiceless consonants generally yielded the set of modern Vietnamese high tones and the old voiced consonant initials yielded the modern Vietnamese low tones. Subsequently, a tonal "flip-flop" occured such that the *ngã* tone, which was once low tone but has become higher than its original counterpart high tone, which was *hỏi*.

3.2 Tone variation

There were some noticeable differences among the six speakers, which seem to correlate with their respective origins. Speaker 1, whose pitch trajectories most closely resemble those described by Thompson, was not born in Hanoi. Interestingly, speakers 2 and 6 (see the tone plots below), who deviate most from the pattern described by Thompson, are Hanoi-born speakers. They both showed in opposition to speaker 1 a very high *ngã* tone with a longer interrupt before revoicing. For speaker 2 *ngã* is considerably higher than *sắc*. Another very palpable difference was in the trajectory of the *hỏi* tone. It fails to rise in both Hanoi speakers, but rather falls only to the lowest levels and then ceases.

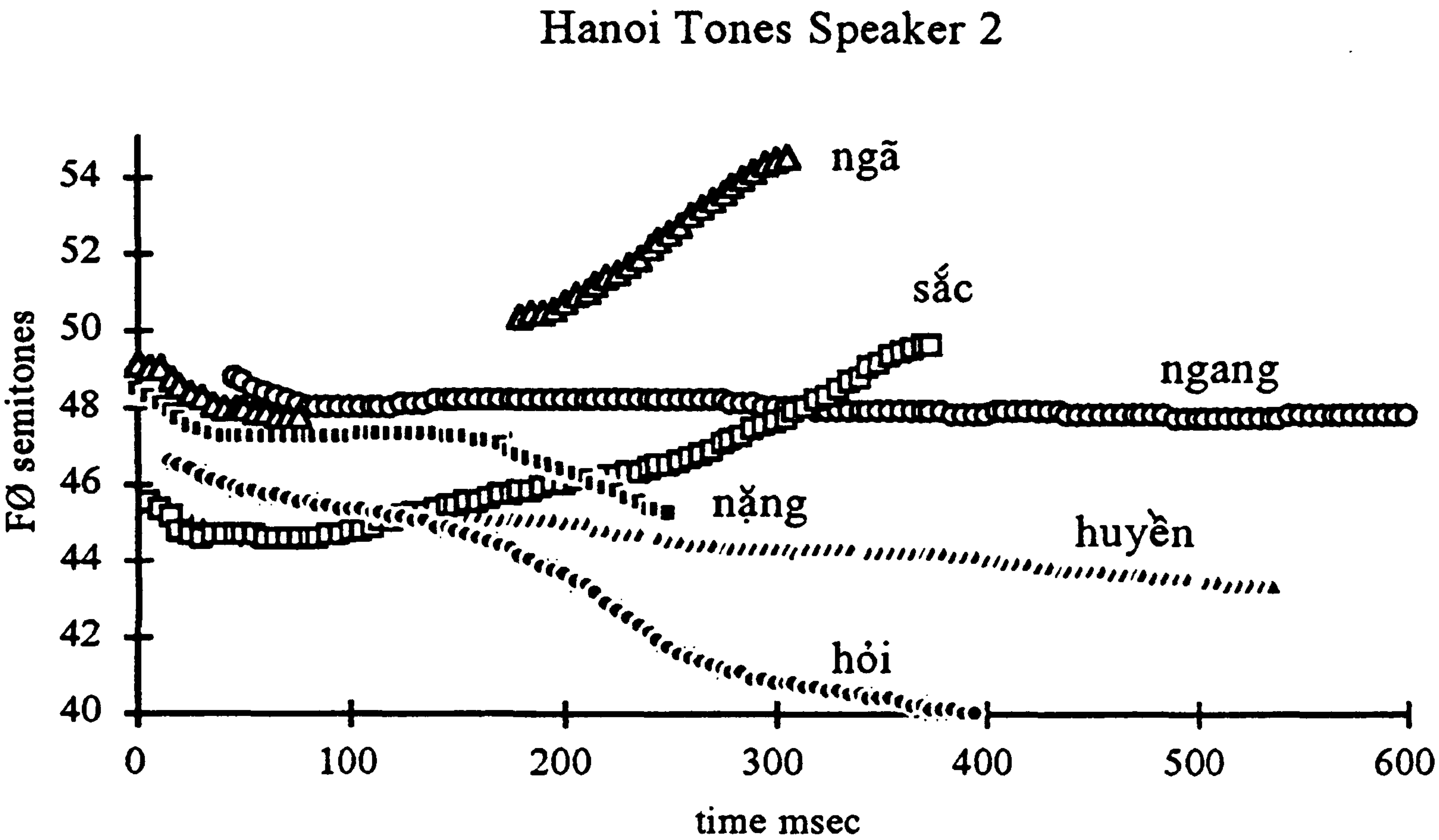

Figure 6. Hanoi Tones Speaker 2

[4]Vietnamese syllables with finals *-ʔ, -p, -t,* or *-k* merged pitch tracks unlike the Chinese, Tai, or Miao-Yao languages.

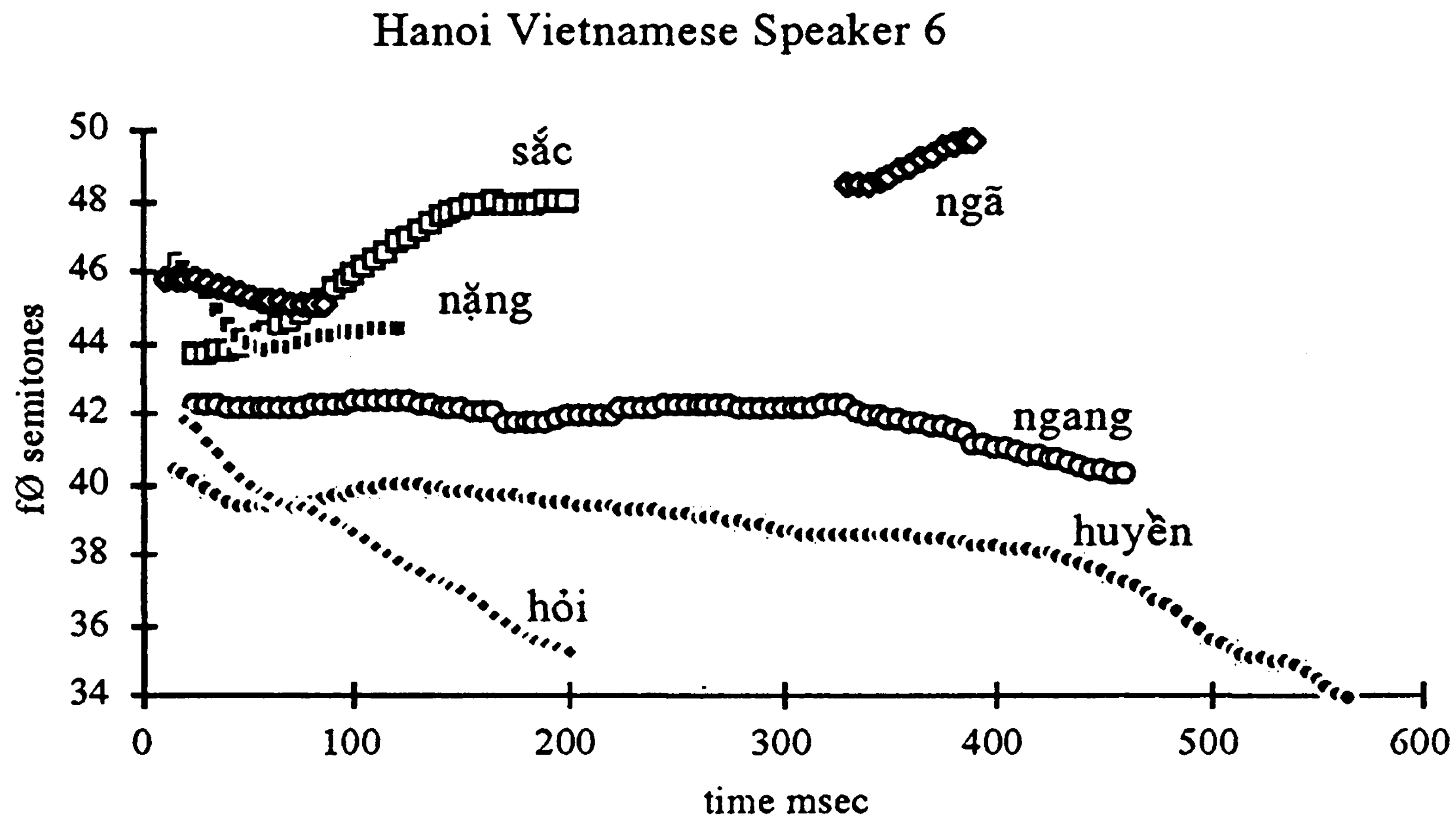

Figure 7. Hanoi Tones Speaker 6

3.2 Voice qualities

Haudricourt's theory also has implications for voice qualities of Vietnamese. B tones should be tense and C tones should be lax or have special creakiness features. We present the results of the inverse filtering study before we address the question of whether contemporary northern speech retains any residue of the tonogenetic origins.

In the following we show a 100-125 msec subinterval taken from the beginning of the syllable. In Figures 8-11 we present the syllable *ta* in the various tone categories for Speaker 6, 1, and 4 plotted from top to bottom, respectively. These three seemed to present the range of possibilities of voice quality.

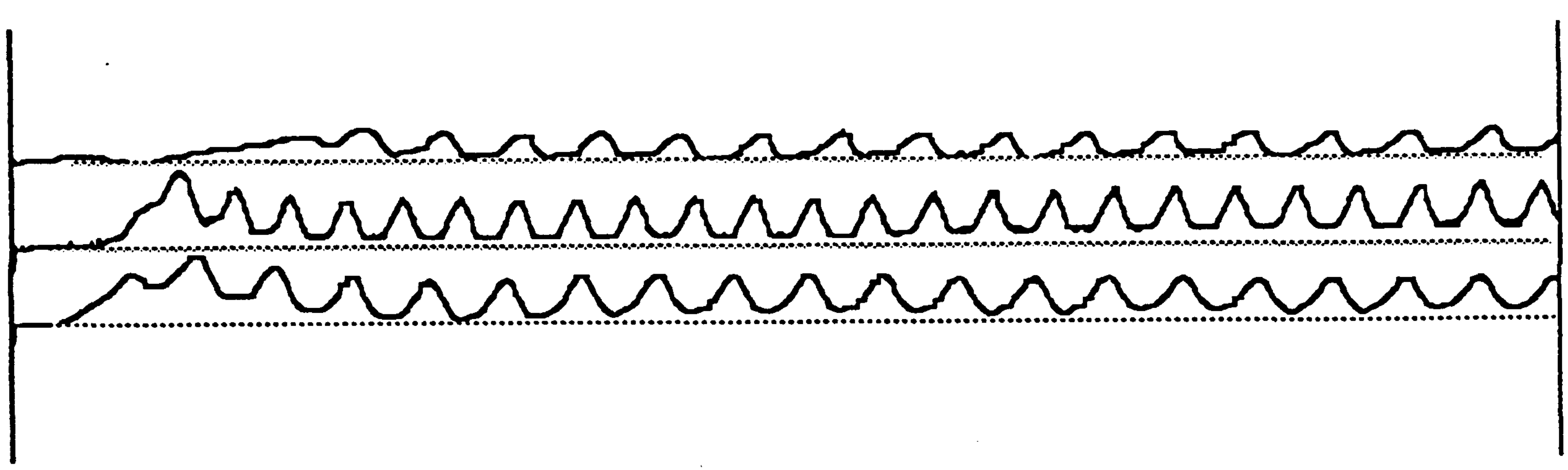

Figure 8. Filtered waveforms of a 100-125 msec subsection of *ta* 'we'

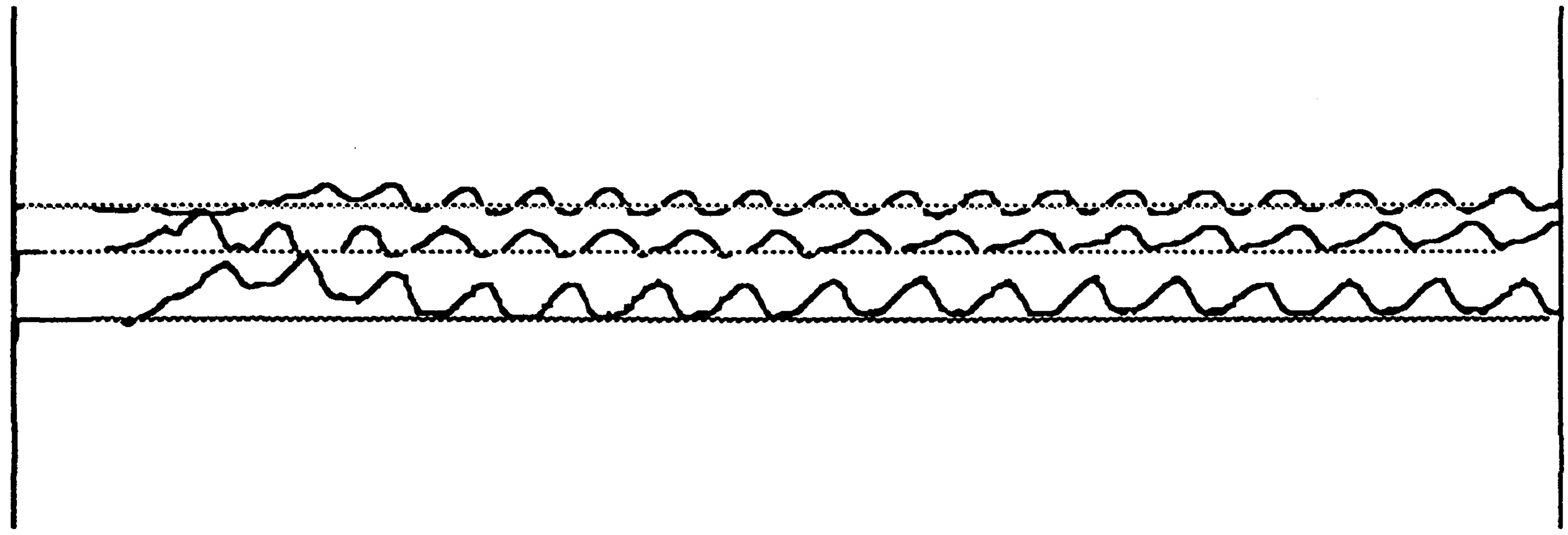

Figure 9. Filtered waveforms of the entire syllable *tà* 'fall', a subsection of 100-125 msec

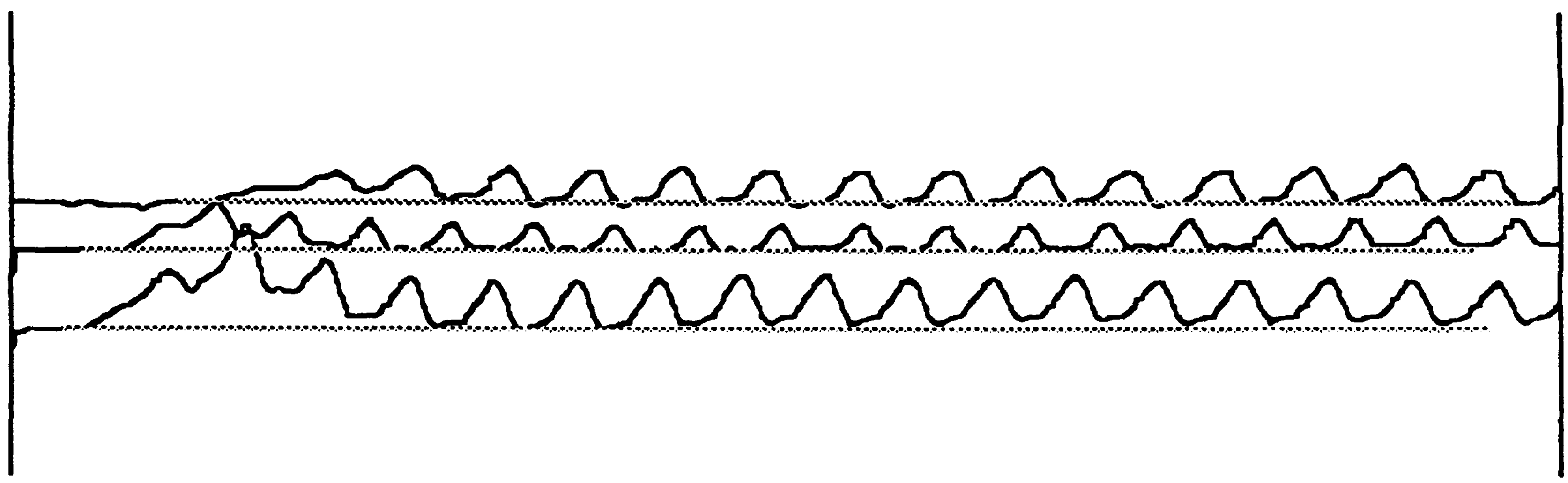

Figure 10. Filtered waveforms of the entire syllable *tá* 'group of 12', a subsection of 100-125 msec

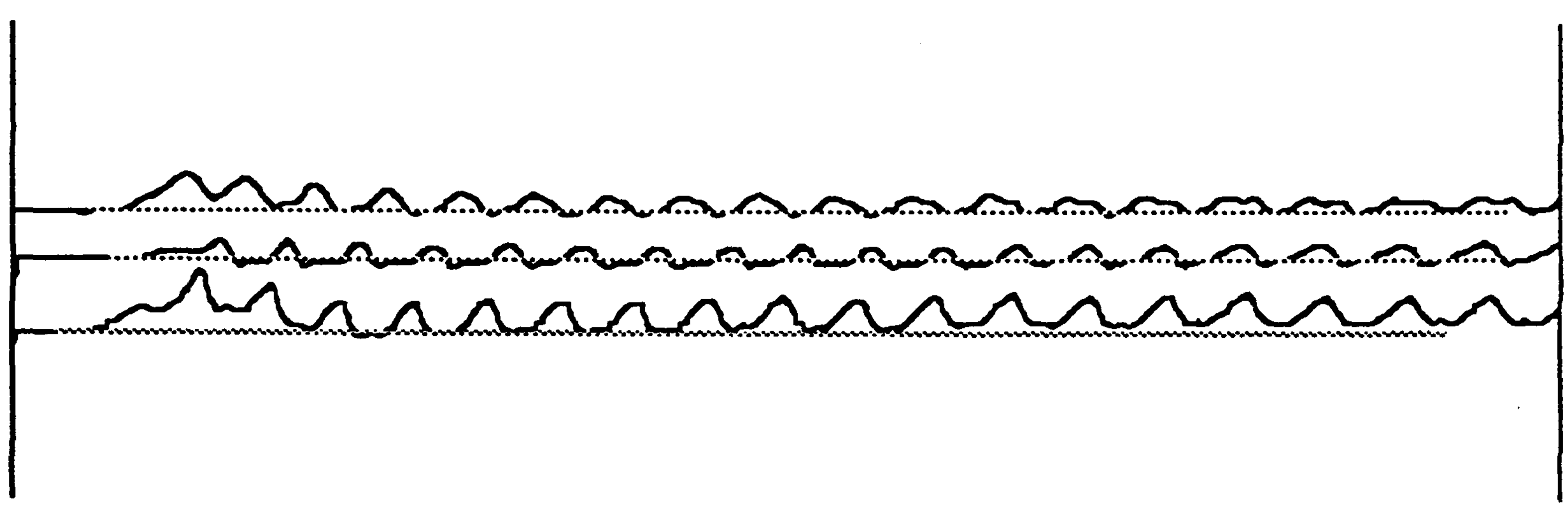

Figure 11. Filtered waveforms of the entire syllable *tả* 'describe', a subsection of 100-125 msec

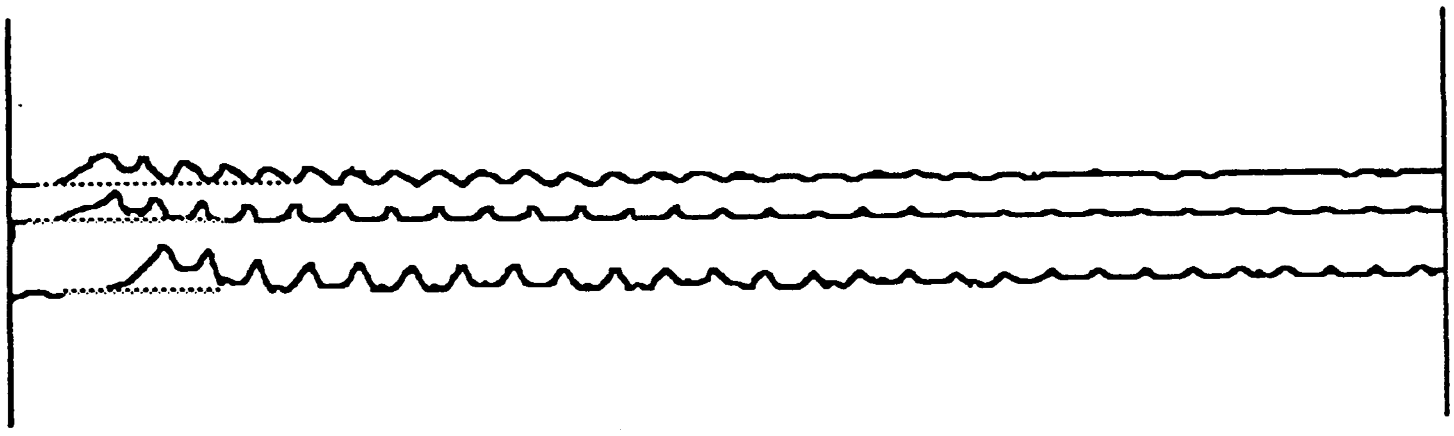

Figure 12. Filtered waveforms of the entire syllable *tã* 'diaper', a subsection of 100-125 msec

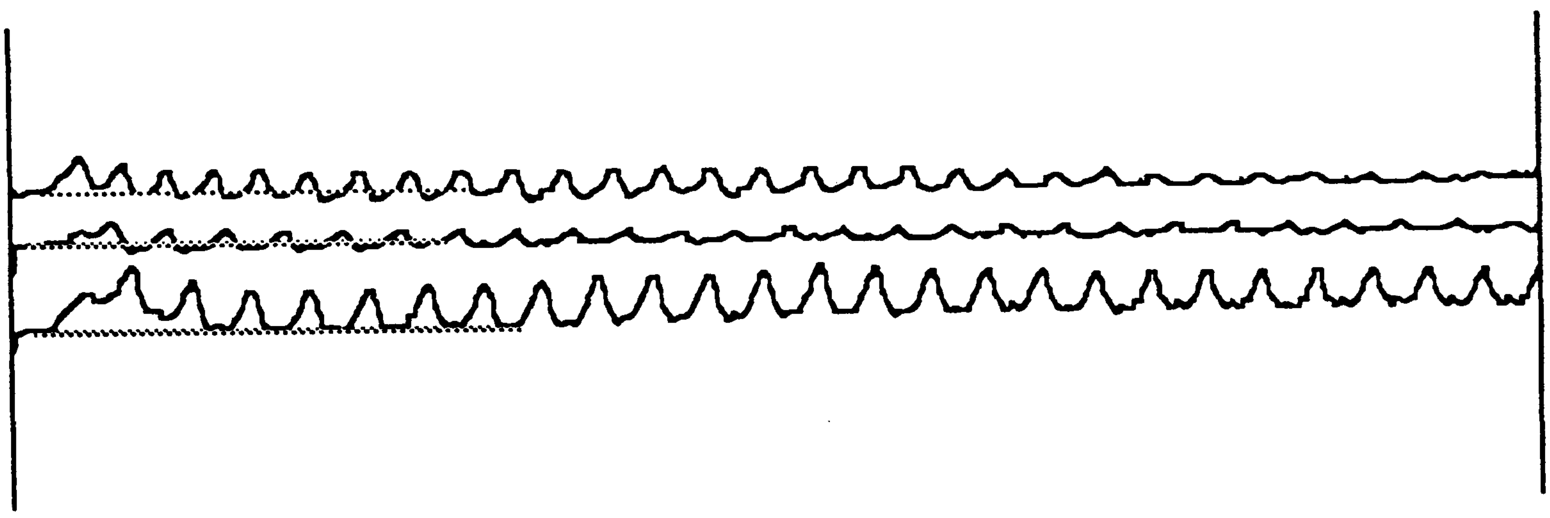

Figure 13. Filtered waveforms of the entire syllable *tạ* '100 kg', a subsection of 100-125 msec

In general all speakers changed their glottal stricture settings from tone to tone. The tones *ngã, hỏi,* and *nặng* were dramatically different from the other three in the sense that these last three, Figures 11 through 13, had lower overall airflow and that airflow changed throughout the course of the syllable. *Ngang, huyền,* and *sắc* showed static settings of the glottis across the syllable, see below for detailed discussion.

5. Discussion

We found that the tone values among contemporary northern speakers were much as described by Thompson. Some Hanoi speakers in particular failed to have a rise in the end of the syllable in the *hỏi*, even in isolation. All three Hanoi speakers failed to have the rise, whereas the three non-Hanoi northerners all had it. Also, Thompson did not mention the shortness of *hỏi* tone, but that was in fact one of its most evident features. It also seems that for some speakers in Hanoi city the *ngã* tone may have become the tone with the highest pitch trajectory. These differences may represent loss of the "dip" in all environments. If so, these developments are reminiscent of changes in Mandarin Chinese, where the dipping tone is preserved in Beijing only in isolation and phrase final and which has been replaced completely in southern forms of Mandarin by a low (falling) tone.

All speakers we studied showed, as expected, some change in their voice quality as a function of tonal pitch differences. These changes were not the same for all speakers. In this regard, we will discuss each tone in turn and compare the glottograms of the three speakers using two viewpoints: (1) the DC and AC components of the glottal wave and (2) the shape or profile of the wave during its opening and closing phases.

Ngang. According to Thompson, we would have expected a somewhat "lax" voice quality in this tone. In fact, only one speaker, number 4 (who appears as the lowest track in the diagrams), showed a somewhat laxer voice than Speaker 1 or Speaker 6 in the sense that many periods failed to have a well defined closed phase component. Nevertheless, there were many other periods in the glottograms that showed a definite closed phase. We therefore suggest that in Hanoi speech today the *ngang* tone is probably most typically characterizable as occuring with modal voice quality (in the sense of Laver 1980).

Huyền. We were expecting to see lax or breathy voice in this tone and in fact, Speakers 1 and 6 did show different profiles in their glottogram from those in *ngang* tone. Speaker 1 preserved a definite closed phase in each period, but this closure often rose at an angle of 20-30° from the horizontal, whereas Speaker 6 had no closed phase and showed complete symmetry around the peaks of the glottal wave. Both Speaker 1 and 6 had a reduced DC component, as their glottograms were somewhat closer to the baseline value. These two voice qualities might then be described as reduced tension setting from modal voice for Speaker 1 and lax and breathy voice for Speaker 6. Speaker 1 had a lower AC component than in *ngang* tone, suggesting that there was also a lowered subglottal air pressure. As for Speaker 4, there appeared to be no differences between the voice quality of *ngang* and *huyền*. The contrast was apparently being signalled mostly by pitch difference. Thus, laxness as breathiness may be viewed as optional enhancing features, with low pitch the obligatory constant in northern varieties of Vietnamese *huyền* tones.

Sắc. All three speakers had glottal wave patterns with definite closed phase elements, even Speaker 4, who typically evidenced a laxer voice than the other two. The duration of the closed phase for Speaker 1 was slightly longer in *sắc* than in *ngang* (3 msec vs 2 msec, though this difference is at the lower limit of our ability to measure), but for Speakers 4 and 6 there was no measureable differnce in duration of the closed phase. Speaker 1 was appreciably tenser than the others in this tone in terms of duration of closed phase and also showed a reduced DC component by being slightly closer to the baseline airflow value.

Hỏi. While the first three tones had airflow values that remained stable over the course of a syllable, *hỏi, ngã,* and *nặng* showed for all speakers both dynamically changing airflow quantity and quality over the course of a syllable. All speakers began the syllable with modal voice-like phonation. The glottal wave for Speaker 6 changed from modal voice with typical closed phase and asymmetrical wave shapes to a pattern with a smaller AC component and a growing tendency toward wave shapes that had two peaks, a pattern we have found before to indicate some harshness from the involvement of the ventricular folds (false vocal cords) in the phonation. From the glottogram it appears that the two fold systems are oscillating at nearly the same frequency so the degree of harshness is not as great as is found, for example, in the Bai language of Yunnan, which, in contrast to Vietnamese, uses harshness as a distinctive feature in its phonology.

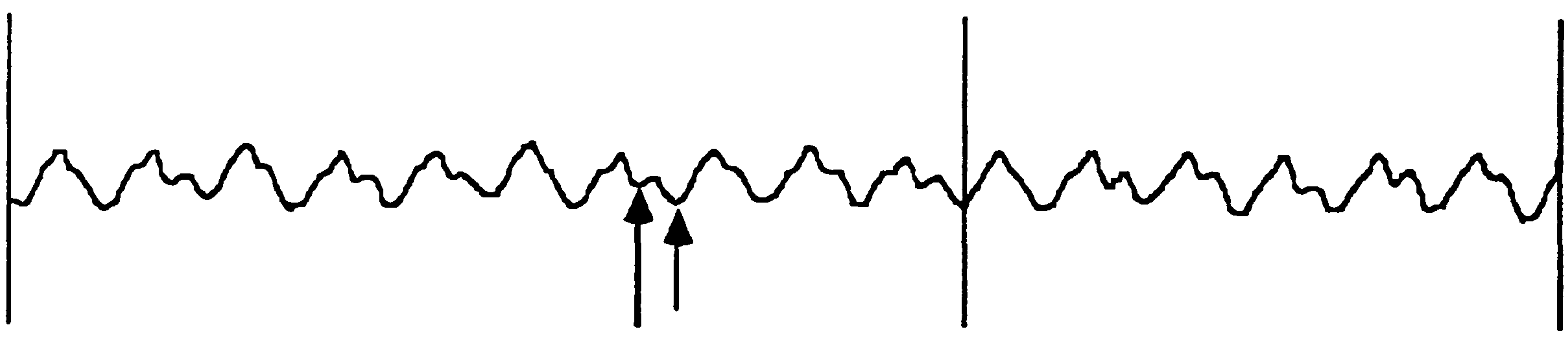

Figure 14. Glottogram of the Bai syllable *ba̰*^{31} 'to move' with the dual peaks of harsh voice quality as a linguistic category, *indicated here with a tilde subscript.*

In the speech of Speakers 1 and 6 there are distinct phases: (1) moderate tension with closed phase and asymmetrical peaks; (2) laxing of the tension with corresponding decay of the closed phase, and (3) double peaking wave profile of harsh voice. Speaker 1 uses a sequence of the first two strategies, whereas Speaker 6 used all three stages. Speaker 4 shows an increase in the duration of the closed phase with time, but none of the more dramatic effects of other contemporaries.

Ngã. All three speakers begin in the modal voice mode and throttle the amount of air to much smaller values midsyllable. The general shape of the airflow is that of an hour glass. Speaker 4 extends the duration of the closed phase much more than in the case of *hỏi*, the height of the AC component decreases to a value slightly less than half that at the beginning of the syllable, but regular periodic voicing is still maintained. But Speaker 4, just as before, is the most restricted in the use of voice quality in comparison with Speakers 1 and 6. Speaker 1 also dramatically throttles the quantity of air flowing while maintaining the general wave shape with asymmetric peak and increasing closed phase. Speaker 6, again the most extravagant user of voice quality, throttles the airflow at the end of the first quarter to the end of the third quarter of the syllable to near perfect closure with only noise of irregular frequency in evidence. The end of the syllable has airflow in volume about 1/3 that of the beginning of the syllable. The airflow wave is very close to the baseline value.

Nặng. All speakers start with modal voice and show progressively increasing closed-phase duration and reduction of airflow quantity over the course of the syllable, so that the general shape of the airflow in a syllable is triangular. The increasing tension finally brings the glottal folds to stasis. In this tone the airflow wave is very close to the baseline value, which shows a high degree of passive tension.

Our instrumentally-informed descriptions of the voice quality in contemporary Northern Vietnamese are basically like those described by Thompson thirty years ago. Only one major point of disagreement arose, namely, *ngang* tone today in the capital is not often lax but usually only modal in voice quality. *Huyền* is lax and for some speakers even lax to the point of breathiness with somewhat lowered subglottal air pressure. Our data indicate *hỏi* and *ngã* are both tense, of course, but their tension is not alike and is not distributed across the syllable in the same way; our findings basically support the description in Thompson's classic account.

We believe the above study has shown, that the Vietnamese language without any doubt uses voice quality in its tonal system. Some speakers may use it more than others but it figures in the phonological system of all the speakers we examined. Some have suggested that Vietnamese is a language that relies for prosodic contrasts only on pitch differences, registers and contours, but we believe that view is overlooking a very important aspect of the sound system. It remains, of course, to be seen if in time Vietnamese will develop toward a state in which voice quality becomes only an enhancement to pitch, as it is perhaps in some forms of Chinese, and that pitch alone carries the bulk of the information burden. But as for today, voice quality remains a significant part of the language. Moreover, one of the surprises to us was the presence of suggestive harsh voice in the *hỏi* tone of some speakers. Harsh voice and creaky voice (from original -s/-h/creaky voice) are in some ways kindred phenomena. If more examples can be found, then that discovery would indeed be important evidence in support of the developmental history as outlined in Haudricourt (1954).

REFERENCES

Andrev, N.D. and M.V. Gordina. 1957. *Sistema tonov vjetnamskogo jazyka. Vestnik* Leningradskogo gosudarstvennogo universitela: Serija Istorii, Jazyka i Literatury 8.132-48.

Diffloth, Gerard. 1992. "Vietnamese as a Mon-Khmer language." In Ratliff and Schiller, eds. *Papers from the First Annual Meeting of the Southeast Asian Linguistics Society*. Tempe, AZ: Arizona State University, Program for Southeast Asian Studies.

Đoàn Thiện Thuật. 1977. *Ngữ âm tiếng Việt*. Hà Nội: Nhà Xuất Bản Đại Học và Trung Học Chuyên Nghiệp.

Earle, M.A. 1975. *An Acoustic Phonetic Study of Northern Vietnamese Tones*. Santa Barbara: Speech Communications Research Laboratory, Inc.

Edmondson, Jerold A. and Li Shaoni. 1994. "Voice quality and voice quality change in the Bai language of Yunnan Province." *Linguistics of the Tibeto-Burman Area* 17.2.49-68.

Emeneau, Murray B. 1951. *Studies in Vietnamese (Annamese) Grammar*. University of California Publication Series VIII.

Fant, Gunnar and B. Sonesson. 1962. "Indirect studies of glottal cycles by synchronous inverse filtering and photoelectric glottography." *Speech Transmission Laboratory Quarterly Progress and Status Report*-4/1962.1-3.

Fant, Gunnar. 1960. *The Acoustic theory of speech production*. The Hague: Mouton.

Fant, Gunnar. 1961. "A new anti-resonance circuit for inverse filtering." *Speech Transmission Laboratory Quarterly Progress and Status Report*-4/1961.1-6.

Ferlus, Michel. 1979. "Lexique thavung-français." *Cahiers de linguistique Asie orientale* 5.71-94.

Ferlus, Michel. 1994. "Quelques particularitiés du cuôi cham, une langue Viet-Muong du Nghe An (Vietnam)." *Neuviemes journées de linguistiques de l'Asie Orientale CRLAO (CNRN-EHESS)*. 4, p.

Ferlus, Michel. 1995. "Particularités du dialecte vietnamien de Cao Lao Ha (Quang Binh, Vietnam)." *Dixièmes journées de linguistiques de l'Asie Orientale CRLAO (CNRN-EHESS)*. 6pp.

Gregerson, Kenneth J. 1969. *A Study of Middle Vietnamese Phonology*. Saigon: Extrait du Bulletin de la Société des Etudes Indochinoises, Nouvelle Série, Tome XLIV, No. 2.

Han, Mieko S. 1969. *Studies in the Phonology of Asian Languages VIII: Vietnamese tones*. Los Angeles: University of Southern California.

Haudricourt, André-Georges. 1954. "De l'origine des tons en vietnamien." *Journal Asiatique* 142.1.

Henderson, Eugenie. 1952. "The main features of Cambodian pronunciation." *Bulletin of the School of Oriental and African Studies* 14:149-74.

Hoàng Tuệ and Hoàng Minh. 1975. *Remarques sur la structure phonologique du vietnamien*. Étude Vietnamienne 40, Hanoi.

Laver, John. 1980. *The Phonetic Description of Voice Quality*. Cambridge University Press.

Lê Văn Lý. 1948. *Le parler vietnamien*. Paris: Imp. éd. Hương Anh.

Lindqvist, J. 1965. "Studies of the voice source by means of inverse filtering." *Speech Transmission Laboratory Quarterly Progress and Status Report-2/* 1965.8-13.

Maspéro, Henri. 1912. "Études sur la phonétique historique de langue annamite." Les initales. *BEFEO* 12.1-127.

Nguyễn Hàm Dương. 1962. *Sistema tonov i spectry glasnyx vjetnamskogo jazyka*. Kandidatskua dissertacija. Moskva.

Nguyễn Văn Lợi. 1993. *Tiếng Rục*. Hà Nội: Nhà Xuất Bản Khoa Học Xã Hội (Ngôn Ngữ Học).

Phạm Văn Hải. 1973. *A Study of Vietnamese Tones*. Carbondale Ill: Southern Illinois University.

Rhodes, Alexandre de. 1651. *Dictionarium annamiticum, latinum et lusitanum*. Rome.

Sagart, Laurent. 1988. *On the Departing Tone, Glottalised Tones in China and South East Asia*. Prosodic analysis and Asian linguistics to honour R. K. Sprigg. Canberra: Pacific Linguistics C104.

Thompson, Laurence. 1965. *A Vietnamese Grammar*. Seattle: University of Washington Press

Trần Hương Mai, Aurélie. 1969. "Stress, Tones and Intonation in South Vietnamese." Ph.D. dissertation, Australian National University.

Vũ Thanh Phương. 1982. "Phonetic properties of Vietnamese tones across dialects." In D. Bradley, ed. Papers in South-east Asian linguistics, No.8 Tonation. Canberra: Pacific Linguistics A62.

Received: 15 January 1997

Program in Linguistics
University of Texas at Arlington
Arlington, TX 76019
USA

Linguistics Institute
National Centre for
Social and Human Sciences
Hanoi, Vietnam