Vietnamese, the official language of Vietnam, is spoken natively by over seventy-five million people in Vietnam and greater Southeast Asia as well as by some two million overseas, predominantly in France, Australia, and the United States. The genetic affiliation of Vietnamese has been at times the subject of considerable debate (Diffloth 1992). Scholars such as Tabard (1838) maintained a relation to Chinese, while Maspero (1912), despite noting similarities to Mon-Khmer, argued for an affiliation with Tai. However, at least since the work of Haudricourt (1953), most scholars now agree that Vietnamese and related Vietic languages belong to the Mon-Khmer branch of the Austroasiatic family.

It is important to make a distinction between ‘literary Vietnamese’, a prescriptive construct in which several orthographic distinctions are maintained in production, and the colloquial or standard speech of a given dialect region. This illustration describes the modern Hanoi dialect of Northern Vietnamese; segmental and tonal inventories, as well as lexicon, vary considerably between Vietnamese dialects, including those spoken in areas adjacent to Hanoi.


1 The Vietic branch is sometimes referred to as Việt-Mường, although this latter term is also used to refer exclusively to a sub-branch of Vietic containing Vietnamese and Mường. See Diffloth (1992) and Hayes (1992) for further discussion.

The recordings accompanying this illustration are of a 32-year-old male native of Hanoi.

Consonants

Initials

<table>
<thead>
<tr>
<th></th>
<th>Labial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plosive</td>
<td>ɓ</td>
<td>t  tʰ</td>
<td>d</td>
<td>ɗ c</td>
<td>k</td>
<td>ʔ</td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>f v</td>
<td>s z</td>
<td>x γ h</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td>w</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral approximant</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ɓaŋ ɓa ‘three’  daŋ ɗa ‘banyan tree’  kaŋ ɗa ‘mug’
maŋ ma ‘ghost’  naŋ ɗa ‘custard apple’  ngaŋ ɗa ‘Russia’
tɑŋ ta ‘we, our’  thɑŋ thɑ ‘to forgive’  lɑŋ là (existential copula)
faŋ pha ‘to brew’  vaŋ və ‘and’  wanŋ oan ‘unjustly’
saŋ xa ‘far’  zaŋ ɗa ‘skin’  haŋ hà ‘river’
tcaŋ cha ‘father’  naŋ nhã ‘house’  hɔŋ ɗa ‘flower’
xɑŋ khɑ ‘rather’  ɣaŋ gã ‘chicken’  ʔaŋ â (question particle)

The voiced plosives are canonically, but not consistently, realized as implosives. Initial /t tʰ/ are apico-dental [t tʰ], lamino-alveolar [t tʰ], or contiguous apico-dental lamino-alveolar (‘denti-alveolar’, Harris 2006), while /d n l/ are apico-alveolar.

Some previous treatments such as that of Thompson (1965) recognize an unaspirated, unaffricated palatal stop /c/. However, in the speech of many younger Vietnamese native speakers from Hanoi, such as that of the present consultant, this segment is consistently realized as an affricate [tʃ], a well-attested areal feature (Harris 2006). The tongue body contacts the alveolar or post-alveolar region during the production of both the palatal nasal [n] and the palatal affricate [tʃ] in initial position (Henderson 1965).

While some varieties of Vietnamese maintain a distinction in the phonetic realizations of orthographic ⟨tr-⟩ and ⟨ch-⟩, these onsets are completely merged in modern Hanoi Vietnamese. The highly salient (and socially stigmatized) merger of /l/ and /n/ > /ŋ/, characteristic of the speech of many lower- and working-class Vietnamese in the Red River Delta, is sometimes consciously manipulated to humorous and/or pejorative effect in colloquial Hanoi speech, as in e.g. /nŋwŋl/ nâu ‘brown’ + /ŋŋɔŋmŋl/ nɔŋ ‘hot’ = ‘hot coffee with milk’ > [lŋwŋ lojɔŋl].

In syllable-initial position /p j f/ occur in a small number of foreign (mainly French) loans, e.g. [pɑŋ] < panne ‘breakdown’, [ɣaŋ ɗaŋ] < garage, [biŋ ɗaŋ] < billiard. For many speakers, however, /p/ is realized as [b/ɓ] and /f/ as [z].
Finals

Hanoi Vietnamese licenses eight segments in coda position: three unreleased voiceless obstruents /p t k/ (\([p^t^k]\))\(^2\), three nasals /m n n\(\breve{n}\)/, and two approximants /j w/\(^2\). In final position /t n/ are canonically alveolar, though it is not clear if they are chiefly laminal or apical. While the EGG study of Michaud (2004) found no evidence of glottalization accompanying unreleased final stops /p t k/, the laryngoscopic study of Edmondson et al. (2010) suggests that glottal reinforcement (in the sense of Esling, Fraser & Harris 2005) may not always be absent in this context.

Velar fronting

Although the phonetic realization of the stops /ŋ k/ following /i e ei/ have sometimes been described as palatal [n c], they are actually pre-velar [ŋ] and [k], with no point of alveolar contact (Henderson 1965). The conditioning vowels tend to be shortened and centralized, and may be produced with a noticeable palatal offglide.

| Khô | ‘Vietnamese’ |
| Keŋ̄ | ‘channel’ |
| Krŋ̄ | ‘broth’ |
| Sík̄ | ‘chain’ |
| Sek̄ | ‘slanting’ |
| Srk̄ | ‘book’ |

There do exist a few instances of true velars following /el/, e.g. [sE:nŋ] xêng ‘shovel’.

Labial-velar finals

Following back rounded vowels /u o ø/, the velar stops /k ŋ/ are produced as doubly articulated labial-velars [kp ŋn]. This articulation is sometimes accompanied by a visible puffing of the cheeks as air becomes trapped in the oral cavity.

| UmĦ̄ | ‘tumor’ |
| OmĦ̄ | ‘grandfather’ |
| Oŋŋ̄ | ‘bee’ |
| Ukŋ̄ | ‘Australia’ |
| Okŋ̄ | ‘snail’ |
| Ėkŋ̄ | ‘mind, brain’ |

Note the differences between the doubly articulated labial-velars and plain final bilabials:

| Sük̄ | ‘to scoop’ |
| Sup̄ | ‘soup’ |
| Hōŋ̄ | ‘hip’ |
| Hōm̄ | ‘day’ |
| Hōkŋ̄ | ‘to study’ |
| Hōp̄ | ‘to meet’ |
| Sōŋŋ̄ | ‘wave’ |
| Ŝōm̄ | ‘hamlet’ |

\(^2\) Whether these segments are transcribed as final approximants /j w/ or as semivowels /h ŋ/ is largely a matter of analytic perspective. From a phonological standpoint, these segments may be regarded as approximants (consonants) on the grounds that they may not be followed by another consonant. However, these segments are articulated somewhat differently from the initial approximants, with a lesser degree of closure.
As with velar fronting, there are rare exceptions to the realization of final velars as labial-velar after back rounded vowels: compare e.g. [boŋmɨ̝] bong ‘to come loose’ with [boŋmɨ̝] boong ‘deck (of ship)’ (< French pont; Nguyễn Bạt Tụy 1949; Haudricourt 1952; Sampson 1969).

Vowels

Hanoi Vietnamese distinguishes nine vowel qualities /i e æ a ɯ y o ø/ and three falling diphthongs /ɨə ɯə ɯə/. Length is normally distinctive only in closed syllables and then only for the vowels /ɑ ɒ/ and /ɔ/; although there do exist a small number of lexical pairs which provide evidence for a length distinction between the vowels /ə ɛ/ such as [seqɨ̝] xếng ‘shovel’ and [zɐŋ] xanh ‘green’ or [zʊŋ] xoong ‘saucepan’ and [zʊŋ] xong ‘to finish’. Although these differences are phonetically robust, the fact that such pairs are also distinguished by differences in the articulation of the coda segment has led to some debate on the proper phonological treatment of the vowel system (Nguyễn Bạt Tụy 1949, 1959; Đoàn Thiện Thuật 1977).

While acoustic analysis of the accompanying audio files reveals small but consistent spectral differences between long and short /ɤ/, it has not been established that these differences are perceptually or psychoacoustically salient; therefore, they are transcribed here as instances of the same vowel quality, i.e. [ɤ ɨ]. /tu/ is frequently realized as mid-centralized [ɨ], leading some authors to transcribe it as [i] (Han 1966; Brunelle 2003). /uw ɯəw/ neutralize to [iɯ] in colloquial Hanoi speech, although speakers who control a formal register may still be able to produce a difference based on the spelling, as can be heard by comparing the accompanying recordings of [ziwɨ̝] diu ‘to soften’ and [zɯɔwɨ̝] râu ‘liquor’.

| Monophthongs |  |  |  |  |  |
|----------------|----------------|----------------|----------------|----------------|
| ɨ | thi | ‘test’ | ɨ | tũ | ‘word’ | ɨ | tù | ‘prison’ | ɨ | tum | (placename) |
| ɨ | tim | ‘heart’ | ɨ | tĩn | ‘news’ | ɨ | tũm | (placename) | ɨ | tũ | ‘short’ |
| ɨ | dip | ‘occasion’ | ɨ | mɨt | ‘jackfruit’ | ɨ | mɨt | ‘jam’ | ɨ | mũt | ‘gun’ |
| ɨ | xinɨ̝ | ‘pretty’ | ɨ | sunɨ̝ | ‘to swell’ | ɨ | sɨng | ‘to swell’ | ɨ | sɨng | ‘to scoop’ |
| ɨ | thɨc | ‘to like’ | ɨ | sukɨ̝ | ‘energy’ | ɨ | ngũ | ‘to smell’ | ɨ | mũ | ‘smell, taste’ |
| ɨ | diu | ‘to soften’ | ɨ | kɨwũ | ‘to rescue’ | ɨ | cɨu | ‘to rescue’ | ɨ | sɨ | ‘to rescue’ |
Hanoi Vietnamese distinguishes eight tones: a six-tone paradigm in open or sonorant-final syllables and a two-tone paradigm in syllables ending in an unreleased oral stop. For convenience, the traditional Vietnamese names of the tones are provided here along with an alphanumeric code indicative of the tones’ historical origins (Michaud 2004).
Figure 2  F0 tracks of tones for a male speaker of Hanoi Vietnamese. Panel (a) shows the six tones found in open sonorant-final syllables; panel (b) compares the pitch of rising and low glottalized tones in open or sonorant-final syllables (black lines) with their checked counterparts (gray lines).

Like many languages of mainland Southeast Asia, pitch is not the only or even primary cue to tone in Hanoi Vietnamese. Instead, tones are realized by a complex of pitch and voice quality features (Nguyễn Văn Lợi & Edmondson 1998; Phạm 2001, 2003), which serve as crucial perceptual cues for native speakers (Brunelle 2009b). In particular, glottalization plays an important role in the production and perception of the broken (C2) and glottalized (B2) tones. The falling tones (A2, C1) have been described by some researchers as accompanied by a breathy voice quality (Thompson 1965; Phạm 2001, 2003); the low falling tone (C1) has also been described as accompanied by light final laryngealization (Nguyễn Văn Lợi & Edmondson 1998; Michaud 2004; Kirby 2010). However, in a laryngoscopic and laryngographic study of Northern Vietnamese tones, Brunelle et al. (2010) found that tone production did not systematically involve visually detectable constrictions other than glottal constriction.

Although duration has not been shown to be a salient perceptual cue to Vietnamese tone, syllables bearing tones B2 [a] and C1 [ŋ] are often shorter than syllables bearing other tones due to the effects of final glottalization. Tone C1 [ŋ], sometimes pronounced with a falling–rising contour in conservative or careful speech, is colloquially realized as a low fall.

Coda–tone restriction
Syllables with obstruent codas are subject to a tonal co-occurrence restriction. Citation tones D1 and D2 [₁] occur only on syllables ending in a voiceless oral stop (‘checked syllables’), and these are the only tones which occur on these syllables. While the D tones may be
analyzed as allophones of the B tones, they are phonetically quite distinct. Tone B1 \[\ddagger\] differs from D1 \[\ddagger\]\] both in pitch onset as well as trajectory (see Figure 2b), and while tone B2 \[\ddagger\]\] is characterized by strong final glottalization, tone D2 \[\ddagger\]\] is produced with modal voice quality (Michaud 2004).

**Voice quality**

Differences in the realization of voice quality in Hanoi Vietnamese tones can be seen in Figure 3, which shows waveforms and pitch tracks for the production of six tones on the carrier syllable /ba/ uttered by a female native speaker. Irregular glottal pulses are clearly visible in the waveforms of the B2 and C2 tokens, although in different locations: tone B2 \[\ddagger\]\] is characterized by initial periodicity followed by strong glottalization, whereas tone C2 \[\ddagger\]\] is interrupted by a strong medial glottal constriction. Some aperiodicity is also visible in the second half of the low falling token C1 \[\ddagger\].

Despite its central role in the production and perception of Vietnamese tone, voice quality is not transcribed in the present illustration for three reasons. First, as emphasized by Phạm (2001, 2003), voice quality is an intrinsic property of the TONE, not of the vocalic nucleus, and at present the IPA transcription provides no way to reflect this important difference. Second, the existing system of IPA diacritics cannot impart the relevant details of the temporal alignment between voice quality and pitch. As illustrated in Figure 3, glottal constriction can literally interrupt the realization of the vocalic nucleus during production of the broken tone (C2), while nonmodal voicing is restricted to the final portion of the glottalized tone (B2); conversely, syllabic rimes bearing a low falling tone (C1) may be laryngealized or breathy throughout. This is consistent with the findings of Nguyễn Văn...
Lôi & Edmondson (1998), who used airflow data to show changes of different degrees in the glottal stricture settings over the course of production of these three tones. Finally, as has been noted in all previous phonetic accounts, there exists considerable variation in the realization and magnitude of voice quality features between speakers and even within the speech of a single speaker, further complicating the issue of how such phonetically complex tones are best represented.

**Tonal coarticulation**

Although Vietnamese tones are not subject to phonological tone sandhi (i.e. the realization of a tone is not affected by the surrounding tonal environment), tonal realization in connected speech is subject to phonetic coarticulation effects. Although tonal height coarticulation is bidirectional, progressive tonal coarticulation is much stronger than anticipatory coarticulation in Hanoi Vietnamese; dissimilatory coarticulation is completely absent (Han & Kim 1974; Brunelle 2003, 2009a).

**Conventions**

**Syllable structure**

A Vietnamese syllable consists of three obligatory elements: an onset, a tone and a vowel. The syllable may optionally contain an obstruent, nasal, or approximant coda. The initial consonant may be accompanied by a secondary labial articulation, as in [hwa Ŝa] hoa ‘flower’. This articulation never follows labial onsets except in loanwords, e.g. [tiœ Ŝa] tiền boa ‘tip’ (< French pourboire). The approximant /j/ never follows the front vowels /i i ı e e/, while /w/ never follows rounded vowels /u u ơ o/.

**Transcription of recorded passage**

Gió bấc và mặt trời cãi nhau xem ai mạnh hơn, trong lúc đó một du khách mặc một áo khoác ấm đi qua. Họ giao kèo với nhau rằng ai là người đầu tiên mà có thể bắt người du khách kia cởi áo thì sẽ được coi là mạnh hơn. Sau đó gió bấc...

Acknowledgements

The author would like to thank the Multimedia, Information, Communications and Applications Center (MICA) at the Hanoi University of Technology for graciously allowing access to their recording facilities; to Marc Brunelle, Jerry Edmondson, Hạ Kiều Phương, and an anonymous reviewer for many detailed comments and suggestions, which have substantially improved this manuscript; and to Mạc Đăng Khoa and Trần Đỗ Đạt for their assistance in translating ‘The North Wind and the Sun’. Any errors, omissions, or oversights are the sole responsibility of the author.

References


Andreev, Nikolai D. & Mirra V. Gordina. 1957. Sistema tonov v’etnamskogo iazyka (po eksperimental’nym dannym) [The system of tones in the Vietnamese language (according to experimental data)]. Vestnik Leningradskogo gosudarstvennogo universiteta 8, 132–148.


Cao Xuân Hao. 1978. Số phận cạc vần có nguyên âm hẹp qua các phương ngữ lớn của Việt Nam [The fate of syllables containing close vowels in the major dialects of Vietnam]. In Thông báo ngữ âm học. TP Hồ Chí Minh: Viện Khoa học Xã hội.


Ferlus, Michel. 1991. Le dialecte vietnamien de Vinh. Presented at the 24th International Conference on Sino-Tibetan Languages and Linguistics, Ramkhamhaeng University, Bangkok.


Han, Mieko S. 1966. Vietnamese vowels (Studies in the Phonology of Asian Languages 4). Los Angeles, CA: University of Southern California, Acoustic Phonetics Research Laboratory.


