Original Paper

Phonetica 2019;76:1–30
DOI: 10.1159/000485103

Received: March 3, 2016
Accepted after revision: October 27, 2017
Published online: May 28, 2018

Effects of Word Position on the Acoustic Realization of Vietnamese Final Consonants

Thi Thuy Hien Tran[a] Nathalie Vallée[a] Lionel Granjon[b]

[a]GIPSA-lab, Speech and Cognition Department, Université Grenoble Alpes, CNRS 5216, Grenoble, and [b]Laboratoire Psychologie de la Perception, Université Paris Descartes, UMR 8242, Paris, France

Abstract

A variety of studies have shown differences between phonetic features of consonants according to their prosodic and/or syllable (onset vs. coda) positions. However, differences are not always found, and interactions between the various factors involved are complex and not well understood. Our study compares acoustical characteristics of coda consonants in Vietnamese taking into account their position within words. Traditionally described as monosyllabic, Vietnamese is partially polysyllabic at the lexical level. In this language, tautosyllabic consonant sequences are prohibited, and adjacent consonants are only found at syllable boundaries either within polysyllabic words (CVC.CVC) or across monosyllabic words (CVC#CVC). This study is designed to examine whether or not syllable boundary types (interword vs. intraword) have an effect on the acoustic realization of codas. The results show significant acoustic differences in consonant realizations according to syllable boundary type, suggesting different coarticulation patterns between nuclei and codas. In addition, as Vietnamese voiceless stops are generally unreleased in coda position, with no burst to carry consonantal information, our results show that a vowel's second half contains acoustic cues which are available to aid in the discrimination of place of articulation of the vowel's following consonant.

Vietnamese Phonological Background

Vietnamese, a Mon-Khmer language in the Austroasiatic family, is traditionally described as an "isolating" language in which all words are invariable and syllables generally have an independent meaning in isolation (Đoàn, 1999, pp. 65; see also Trương, 1970). Most Vietnamese syllables are lexical or grammatical morphemes, and a very large proportion of its words are monosyllabic and monomorphemic, which led many authors to classify Vietnamese as a monosyllabic language (Nguyễn, 1989, p. 27). The few polysyllabic monomorphemic words are either place names (e.g. *Sài Gòn*) or loans (e.g. the French borrowing *va li* "suitcase"). This one-to-one mapping of syllables (*âm tiết*) and morphemes (*hình vị*) produces the concept of *syllabeme* (*tiết*

KARGER

© 2018 S. Karger AG, Basel

E-Mail karger@karger.com
www.karger.com/pho

Thi Thuy Hien Tran
GIPSA-lab, Département Parole et Cognition
Université Grenoble Alpes, 1180, avenue Centrale, BP 25
FR–38040 Grenoble Cedex 9 (France)
E-Mail thi-thuy-hien.tran@gipsa-lab.fr

vị or *tiếng*) in Vietnamese linguistics (Ngô, 1984; Cao, 1985), the minimal meaningful unit of this language. As in the following example[1], syllables are monosyllabic words and provide an important part of the Vietnamese basic words (Lê, 2011), e.g. *ngôi nhà đẹp* [ŋoj$^{\mathrm{A1}}$ ɲa$^{\mathrm{A2}}$ dɛp$^{\mathrm{D2}}$] "the house is beautiful" (classifier-house-beautiful).

Vietnamese is therefore typically characterized as phonologically monosyllabic, as the basic structure of the lexicon is monosyllabic (Michaud, 2009), and the segmental phonology operates exclusively within the domain of the syllable. However, not all Vietnamese words are monosyllables. There is a substantial number of polysyllabic lexical words with from 2 to 4 syllables, such as in example 1:

(1)	*học tập*	[hɔk$^{\mathrm{D2}}$ tɤ̆p$^{\mathrm{D2}}$]	"to study"
	hợp tác xã	[hɤp$^{\mathrm{D2}}$ tak$^{\mathrm{D1}}$ sa$^{\mathrm{C2}}$]	"cooperative"
	khập khà khập khiễng	[xɤ̆p$^{\mathrm{D2}}$ xa$^{\mathrm{A2}}$ xɤ̆p$^{\mathrm{D2}}$ xieŋ$^{\mathrm{C2}}$]	"limping"

From a study on segmental co-occurrences in Vietnamese lexical units, Trần and Vallée (2009) found in a lexicon of 5,000 in current use that disyllabic words account for half of the lemmas (49.67%) and monosyllabic ones for 49.65%. Trisyllabic words are very limited (0.68%), while 4-syllable lemmas do not appear among the Vietnamese common words. As a result, Vietnamese is a monosyllabic language at the phonological level, but a polysyllabic language at the lexical level.

The Vietnamese Lexicon: Polysyllabic Word Formation Strategies

A brief overview of different formation processes of Vietnamese polysyllabic words provides a good illustration of the "monosyllabic status but partially polysyllabic at the lexical level" for this language. In Vietnamese, polysyllabic words are basically formed either by composition or by reduplication.

Traditional descriptions classify the Vietnamese compound words into 2 semantically differentiated main categories, namely coordinate compounds versus subordinate compounds (Thompson, 1965; Mai et al., 1997; Nguyễn, 1998; Diệp, 2004; see Schiering et al., 2010, for a summary).

Semantically coordinate compound words are basically formed by at least 2 juxtaposed monosyllabic elements. The compound may consist of component elements which, if taken separately, each have a different and specific meaning. The combination of these elements manifests a common sense of the word. For example, *xăng dầu* "fuel" consists of 2 elements, *xăng* "gasoline" and *dầu* "oil." The coordinate compounds may also be formed by 2 elements having similar or neighboring meanings. For example, all words in example 2 are composed of synonymous elements.

(2)	*tìm kiếm*	[tim$^{\mathrm{A2}}$ kiem$^{\mathrm{B1}}$]	"to search"
	đợi chờ	[dɤj$^{\mathrm{B2}}$ tɕɤ$^{\mathrm{A2}}$]	"to wait"
	bởi vì	[bɤj$^{\mathrm{C1}}$ vi$^{\mathrm{A2}}$]	"because"
	thay đổi	[tʰɛj$^{\mathrm{A1}}$ doj$^{\mathrm{C1}}$]	"to change"
	núi non	[nuj$^{\mathrm{B1}}$ nɔn$^{\mathrm{A1}}$]	"mountains"

[1] In all examples, tones are indicated in superscript according to Michaud (2004), with conventional alphanumerical labels reflecting their diachronic origin; the transcription corresponds to surface forms as pronounced in standard Hanoi Vietnamese: *ngang* high-level (A1)*, huyền* low-falling (A2)*, sắc* low-rising (B1)*, nặng* mid-falling-glottalized (B2)*, hỏi* high-falling-rising (C1)*, ngã* mid-rising-glottalized (C2); with 2 additional tones in stop-final rhymes: *sắc* high-rising (D1) and *nặng* low-falling (D2).

Semantically subordinate compounds consist of an element having a generic sense and a second one playing a subordinate role to the first element. Thus, in the following example 3, the first monosyllabic component *xe* is the generic term for vehicle, while *máy, đạp, ngựa, điện* are subordinate elements that express the discrimination and through which different types of vehicle can be distinguished:

(3)	*xe máy*	[sɛ^{A1} măj^{B1}]	“motorcycle”
	xe đạp	[sɛ^{A1} dap^{D2}]	“bicycle”
	xe ngựa	[sɛ^{A1} ŋɯɤ^{C2}]	“horse-cart”
	xe điện	[sɛ^{A1} dien^{C2}]	“tramway”

Another usual way to create polysyllabic words is by reduplicative derivation, i.e. by repeating a whole or only parts of a word (Thompson, 1965; Mai et al., 1997; see Noyer, 1998, p. 85, for a summary). Reduplication is a process that expresses a number of meanings, such as distributive, iterative, attenuative, intensive, and emphatic. A reduplicated word consists of 2–4 syllables that have a phonetic link between them at the suprasegmental level, either tone or the subsyllabic constituents of onset and/or rhyme, as illustrated in example 4 (duplicated part in bold). However, 98% of the Vietnamese reduplicative words are disyllabic (Lê et al., 2009).

(4)	*chuồn chuồn*	[**tɕuon**$^{\mathbf{A2}}$ **tɕuon**$^{\mathbf{A2}}$]	“dragonfly”
	khe khẽ	[**xɛ**^{A1} **xɛ**^{C2}]	“softly”
	lang thang	[l**aŋ**$^{\mathbf{A1}}$ t^{h}**aŋ**$^{\mathbf{A1}}$]	“to wander”
	khít khìn khin	[**x**it^{D1} **x**in^{A2} **x**in^{A1}]	“close-fitting”
	lủng là lủng lẳng	[**l**uŋ^{C1} **l**a^{A2} **l**uŋ^{C1} **l**ăŋ^{C1}]	“hanging down loosely”

In addition, there are also complex polysyllabic words in Vietnamese, phonetically transcribed from foreign languages. Because of long contact with Chinese and French, the Vietnamese vocabulary includes a large number of Sino-Vietnamese words (as illustrated in example 5) and has also more recently been enriched with loanwords from French (as e.g. in example 6).

(5)	*đại lý*	[daj^{B2} li^{B1}]	“agency”
	du lịch	[zu^{A1} lik^{D2}]	“tourism”
	bất động	[bɤ̆t^{D1} doŋ^{B2}]	“immobile”
(6)	*ban công*	[ban^{A1} koŋ^{A1}]	“balcony” (from French, *balcon*)
	bê tông	[be^{A1} toŋ^{A1}]	“concrete” (from French, *béton*)
	ô tô	[ʔo^{A1} to^{A1}]	“car” (from French, *auto*)

Vietnamese: A Tonal Language with CVC Dominant Syllable Structure

Vietnamese is typologically an isolating CVC language (almost 70% of the whole Vietnamese syllables are of this type) while CV syllables are more limited (24%) (Trần and Vallée, 2009). The syllable structure is C(w)V(C) with each syllable bearing a lexical tone and the (w) and final (C) optional. This pattern implies that the glottal stop which appears only in syllable-initial position is phonemic. The nucleus is either a long or short vowel or a diphthong. A strong phonotactic restriction process occurs in Vietnamese syllable-final position, since only 6 obstruents of the language’s 22 consonants are allowed in the coda, specifically either a voiceless stop or a nasal: /p/ /t/ /k/ /m/ /n/ /ŋ/. It should also be noted that a glide, /w/ or /j/, is permitted in coda and that the voiceless bilabial plosive /p/ is never found in syllable-initial position in native Vietnamese words. Another particularity of Vietnamese phonetics and phonology is that voiceless stops are released in syllable onset position with a fast and audible burst, but they typically occur as an unreleased (burstless) allophone when in syllable-final position (Cao, 1985; Đoàn, 1999; Trần and Vallée, 2009; Trần, 2011; Kirby, 2011).

Vietnamese is a tonal language in which the meaning of each syllable varies according to its tone. The rhyme is the tone-bearing unit, and the syllable-initial consonant does not carry tone information (Trần et al., 2005). The Vietnamese tones are characterized both by pitch modulation and by voice quality (see Michaud, 2004, and Brunelle et al., 2010, for more details), but there is considerable tone variation between Vietnamese dialects. Generally, the northern varieties have 6 tones while those in other regions (southern and central) have 5, merging 2 of the tones into 1, while still other north-central varieties have a merger of 3 tones resulting in a 4-tone system. Moreover, while northern and central varieties have glottalized tones (Nguyễn and Edmondson, 1997; Phạm, 2003; Michaud, 2004), southern dialects make no use of voice quality in production in tonal contrasts (Brunelle, 2009). However, southern listeners can still access these cues in perception when listening to northern speakers (Kirby, 2010).

For the purpose of this paper, we will focus only on the Hanoi dialect which has a phonologically 6-tone system: *ngang* high-level (A1)*, huyền* low-falling (A2)*, sắc* low-rising (B1)*, nặng* mid-falling-glottalized (B2)*, hỏi* high-falling-rising (C1)*, ngã* mid-rising-glottalized (C2). Stop-final rhymes are associated with 2 tonal variants a high-rising *sắc* (D1) and a low-falling *nặng* (D2), both produced with a modal voice, making the final stops /p/ /t/ /k/ produced without glottalization (Michaud, 2004).

Research Questions

The 3 final stops of Vietnamese are traditionally described as unreleased and are therefore lacking the distinguishing information that would be carried in a release. Nevertheless, native subjects are able to identify them at very high rates (81% on average) (Trần and Vallée, 2010). This raises interesting questions about what acoustic cues underlie this ability and allow for that discrimination.

Furthermore, many experimental studies in articulatory phonetics have shown differences in the production of consonants, according to their position within the syllable (Lindblom, 1983; Keating, 1983; Browman and Goldstein, 1988, 1995; Byrd, 1995; Krakow, 1999; Kingston, 2008) and within prosodic domains (e.g., utterance) (Malécot, 1968; Fougeron and Keating, 1997). Previous works suggest differences in coarticulation strength between tautosyllabic consonant clusters[2] (#CC or CC#) and heterosyllabic consonant sequences (C#C) (Davidson, 2003, 2007). Differences have also been found in the acoustic signal between syllable-initial and syllable-final consonants, whether preconsonantal ($\underline{\text{C}}$.C) or word-final ($\underline{\text{C}}$#C) consonants (Redford and Diehl, 1999; Wright, 2004). Such findings raise the question of whether heterosyllabic but word-internal sequences (C.C) differ from sequences which span a word

[2] It should be noted that we call a series of consonants a *cluster* if they occur in the same syllable, and a *sequence* if in two consecutive syllables, according to the terminology proposed by Pulgram (1965).

boundary (C#C). In other words, are there differences in the acoustic properties of Vietnamese finals when they are word-internal (C̲.C) versus spanning a word boundary (C̲#C)?

In a pilot study on 2-consonant sequences produced by an adult native speaker of Vietnamese, Trần and Vallée (2009) found acoustic differences in the preconsonantal consonant according to whether the consonant sequences cross a word boundary or a word-internal syllable boundary. If the findings of the 2009 study are replicated with multiple speakers, it would reinforce the finding of differences at syllable boundaries in word-internal and word-final positions in Vietnamese. One of the main goals of this paper is therefore to extend the pilot acoustical investigation with a greater number of speakers and to try to uncover evidence about the word position effects by examining the acoustic properties of Vietnamese coda consonants.

Two main questions are therefore addressed:

1 What acoustic cues are available to potentially aid in the discrimination of Vietnamese final consonants since they are not released?

2 Are word-internal codas different from word-final codas?

Method

We conducted an acoustic study of final stops and final nasals. The acoustic parameters of the 6 Vietnamese final obstruents /p t k m n ŋ/ were measured for 10 native speakers of Hanoi Vietnamese in a production experiment using a speech-reading task, and by comparing the productions of the consonants according to their position in the word and in the syllable.

Corpus

We first selected a lexicon of the 5,000 most commonly used Vietnamese lemmas (i.e., the uninflected base form of content words). The source lexicon from which ours was selected was originally collected by Lê et al. (2003) and Lê (2006) in the framework of an automatic speech recognition project. The vocabulary was automatically created from textual data extracted from Web resources (*VnExpress News*, the most read Vietnamese online newspaper) on which various forms of data treatment, such as filtering, were performed (see Lê, 2006, p. 48). Next, we used the text2wfreq tool from the CMU-Cambridge Statistical Language Modeling tool kit in order to extract 5,000 words from the list of vocabulary on the criterion of their numbers of occurrences. We verified that the lexicon contained the 100 words "as indispensable as possible" from the Swadesh list for Vietnamese of Lê (2004). We also compared the list of vocabulary to the most common 1,000 words of the Assimil method "Learning Vietnamese as Foreign Language" (Đỗ and Lê, 1994). The resulting 5,000-word Vietnamese lexicon was then integrated into the G-ULSID database (Grenoble and UCLA Lexical and Syllabic Inventory Database) (Maddieson and Precoda, 1992; Rousset, 2004; Vallée et al., 2009). Each lexical item corresponds to a phonologically transcribed and syllabified word form. The subsyllabic onset and rhyme units of each syllable are indicated, with the nucleus and coda at the rhyme level.

From the Vietnamese lexicon, 62 lexical monosyllables and disyllables were selected for the present study (4 CV, 23 CVC and 35 CVC.CVC). Each of these lexical items contains 1 of the 6 target consonants in the context of the low vowel /a/ and in the rising tonal context (*sắc* tones, B1 and D1).

The reason for choosing this tonal context is twofold. First, stop-final syllables permit only 2 tones (*sắc*-D1 or *nặng*-D2). An examination of the 5,000-word lexicon has shown that words with final stops matching our criteria (a /CaC/ initial syllable in a polysyllabic item) are 3 times more frequent under tone D1 than under tone D2, 62 versus 22 words, respectively. Furthermore in Northern Vietnamese, tones D1 and D2 are not produced with glottalization (Michaud, 2004). Second, syllables with final sonorants only occur with the *sắc*-B1 and *nặng*-B2 tones, of which B1 is produced in modal voice, whereas B2 is glottalized. In the B2 tonal context, acoustically speaking, the final nasals almost entirely disappear as segments under the influence of glottal interrupt (Michaud, 2004; Michaud et al., 2006). In order to facilitate the audio signal segmentation, then acoustic measurements, *sắc* tones (B1 and D1) were therefore selected for the corpus. The vowel /a/ was chosen to provide a better contrast between the vocalic segment and the adjacent stop in onset as well as in coda position.

In the limited lexical inventory remaining after our initial syllable selection, the tonal context of either *sắc* rising tones (B1, D1) or *nặng* falling tones (B2, D2) was selected for the second syllable of a polysyllabic word, leading to a corpus 18 of the former and 17 of the latter. It should be noted that the tonal context of the second syllable has minimal influence on acoustic measurements (and specifically F$_0$) of the first syllable, which is the target of our acoustical study, because in a polysyllabic word, the tonal coarticulation is generally progressive rather than regressive, and if a regressive tone influence is found, it is much weaker than its progressive counterpart (Brunelle, 2009).

In the resulting lexical corpus, the 6 target consonants /p t k m n ŋ/ are located in various within-word positions: (i) in syllable-initial onset position C- of monosyllables in an open /**C**a/ or closed /**C**aC/ syllabic structure (e.g., /ta/, /ka/, /ma/, /tat/, /kat/, /nat/); (ii) in word-final coda position -C$_{\#}$ in monosyllables /Ca**C**/ (e.g., /kap/, /mat/, /kak/, /tam/, /tan/, /kaŋ/) or in word-internal coda position -C$_{\sigma}$ in the case of disyllabic words /Ca**C**.CVC/ (e.g., /fap.li/, /fat.zak/, /sak.daŋ/, /xam.fa/, /ban.ket/, /saŋ.tak/).

The target polysyllabic words were selected so that the paired consonants to be analyzed disagreed in voicing: all voiceless coda consonants preceded a voiced onset consonant, and conversely, all voiced codas preceded a voiceless onset. For the same reason, 2 carrier sentences were designed for the monosyllables, to ensure that the phonetic context of the target phonemes was comparable regarding voicing, thus:

If the target word ends with a vowel or a sonorant (**-C$_{\#}$** = /m n ŋ/), the following word in the carrier sentence began with the voiceless alveolar [s]: "*Bạn sẽ gặp từ __ xuất hiện trong bài khóa*" [ban^{B2} sɛ^{C2} ɣăp^{D2} tɯ^{A2} __ swɤ̆t^{D1} hien^{B2} ʨɔŋ^{A1} baj^{A2} xwa^{B1}] (you will find the words appear in the text).

If the target word ends with a voiceless stop (**-C$_{\#}$** = /p t k/), the next word in the carrier sentence began with the lateral voiced consonant [l]: "*Bạn sẽ gặp từ __ liên tiếp xuất hiện trong bài khóa*" [ban^{B2} sɛ^{C2} ɣăp^{D2} tɯ^{A2} __ lien^{A1} tiep^{D1} swɤ̆t^{D1} hien^{B2} ʨɔŋ^{A1} baj^{A2} xwa^{B1}] (you will find the words successively appear in the text).

As a reminder, /p/ does not exist in onset position in the native Vietnamese vocabulary, and the loanwords where it is now established do not meet the criteria of the corpus (vowel /a/, *sắc* tones). Therefore, there are no p-initial words as part of the materials. In the end, 62 monosyllabic and disyllabic words matching our criteria were selected from the 5,000-word lexical database (see the Appendix 1 for a complete list of the stimuli). The 35 disyllabic items selected are either semantically compound (32) or reduplicative (3) words. The recording corpus consisted of 4 randomized repetitions of the selected words in their carrier sentences, which resulted in (62 words × 4 repetitions × 10 speakers) = 2,480 tokens for analysis.

Speakers

The 5 male and 5 female participants were between 23 and 28 years of age and were native speakers of the Hanoi dialect of northern Vietnamese. The speakers were asked to read each sentence in one setting at a normal and fluent speech rate, paying attention to not introduce a phrase boundary or a pause after the target word. The sentences were written in 18-point font size and were listed and presented on paper. Breaks were inserted during the reading task for participant comfort. The recording was performed in the GIPSA-lab's sound-proof room (Grenoble, France), using a Marantz PMD 670 digital audio recorder at a 44.1-kHz sampling rate and an AKG C1000S cardioid microphone.

Data Processing

The 2,480 recorded sentences (3 h of read speech) were manually segmented and annotated using Praat speech-processing software. Acoustic parameters were then semiautomatically analyzed using Praat and Matlab. The measurements made were:

- Target consonant duration: initial consonants C- in both **C**/a/ and **C**/a/C$_{\#}$, final consonants in both C/a/**C**$_{\#}$ and C/a/**C**$_{\sigma}$.CVC
- Vowel /a/ duration
- Voice onset time (VOT) duration
- Stop closure duration

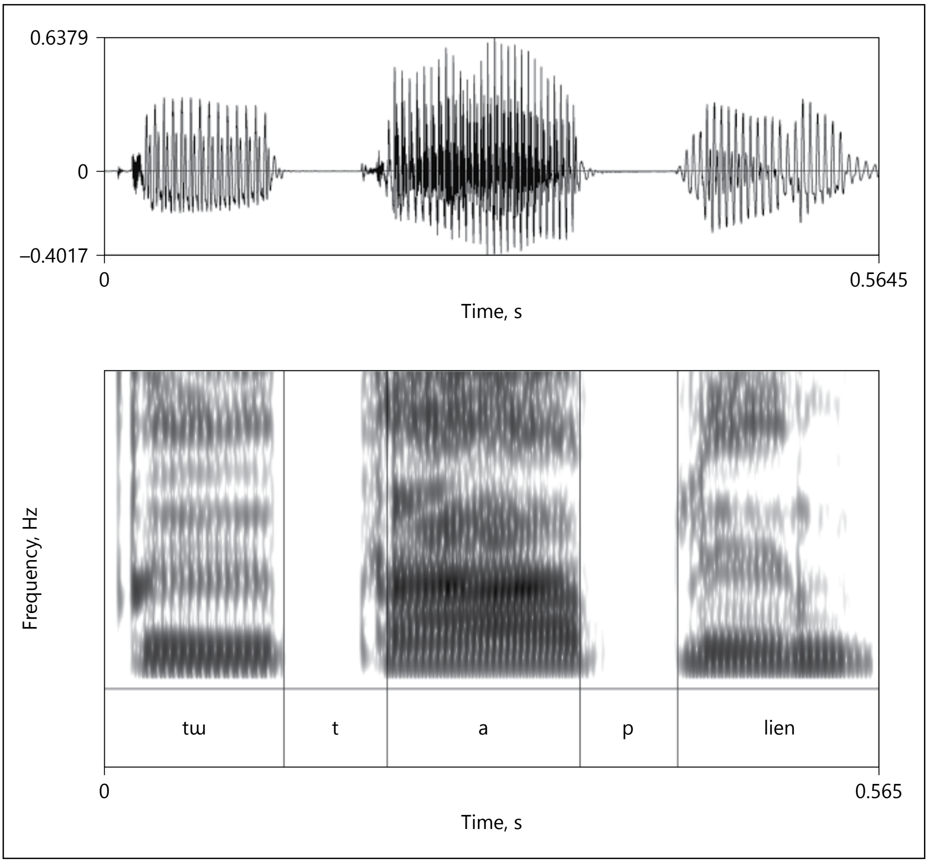


Fig. 1. Segmentation example of C-, V, and -$C_{\#}$ durations of the word [tap].

- Burst amplitude and duration (when a burst was visible on the spectrogram)
- Time evolution of acoustic parameters: intensity, fundamental frequencies (F_0), and the first 3 formants (F_1, F_2, F_3) were measured from time points of 40 and 50% of the vowel length (i.e., the vocalic part less affected by consonantal environment), to time points of 60, 70, 80, and 90% of the vowel length (i.e. the transition part between /a/ and the adjacent final consonant)

These VC transition parameters were added because of the generally unreleased stops found in coda position in Vietnamese. Previous studies have shown that final consonants can be identified by listeners through variations of the rate and direction of formant transitions of the preceding vowel (Sharf and Hemeyer, 1972; Dorman et al., 1977; Cao, 1985, p. 83; Serniclaes, 1987). Thus, in our study, variations (Δ) of F_1, F_2, F_3, also F_0, and intensity (ΔI) were obtained by calculating the difference values between time points T_2 (from 50 to 90% of the vowel length) and time points T_1 (40% of the vowel duration corresponding to the steady-state portion of the vowel).

For stop consonants, C- duration was measured on the spectrogram by taking the time interval between the last periodic pulse of the immediately preceding vowel /ɯ/ aligned with the apparent end of the formant structure, and the start of the glottal vibration for the following /a/ aligned with the sharp beginning of its formant structure. The total stop duration therefore includes the closure phase, the burst, and the VOT durations (Fig. 1). Final stop duration in -$C_{\#}$ and in -C_{σ} was calculated as the difference between the time marked as the voicing termination of the vowel /a/ and the time marked as the start of the glottal vibration of the following voiced consonant.

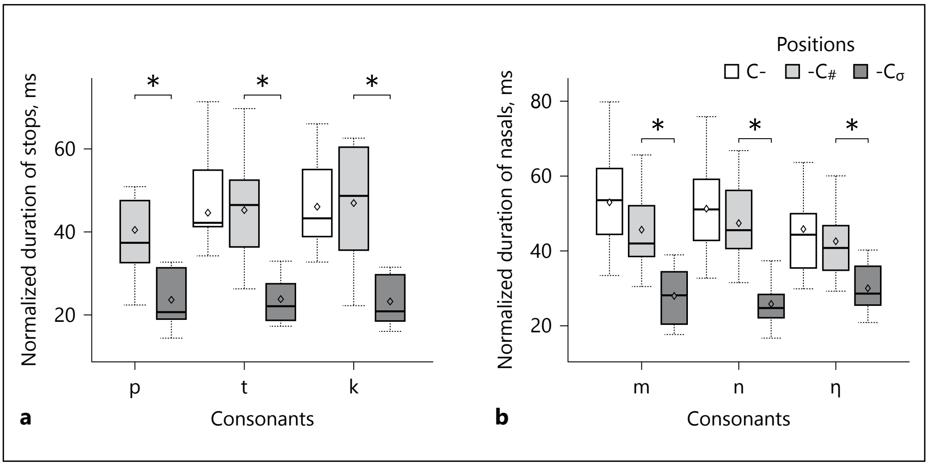


Fig. 2. Boxplot with symbol (◊) representing the means of normalized duration (ms) of stops (**a**) and nasals (**b**) according to their position within words. * $p < 0.05$.

VOT was measured from wide-band spectrograms according to the method of Lisker and Abramson (1964). VOT is the time interval between the onset of the energy burst corresponding to the release of stop consonant closure and the beginning of the first periodic component of regular glottal vibration corresponding to the voicing onset of the following vowel or of the following voiced fricative (in the case of final consonants -$C_{\#}$ and -C_{σ}). Burst duration and intensity were measured from wide-band spectrograms when release noise (burst energy) was visible on the spectrogram.

All of the duration measurements were normalized to the speech rate (syllables per minute) before statistical treatment. The speech rate of the target word is calculated by taking the number of syllables of the word and dividing by its raw duration. The normalized consonant durations are calculated by taking the raw duration and dividing by the corresponding local speech rate. The same procedure is applied for all other duration measurements.

Statistical Analysis

Acoustic measurements were analyzed with a series of repeated-measures analyses of variance (ANOVAs) using SPSS© (Statistical Package for the Social Sciences) for effects of syllable position, or as interaction between position and place of articulation (labial, coronal, or velar). Investigated effects were:

- Effect of syllable position (C-, -$C_{\#}$, or -C_{σ}) on acoustic variables of target stops (total duration, closure duration, burst duration, burst intensity, VOT). For nasal consonants, a separate test was conducted on the total duration of the consonant
- Effect of boundary type (-$C_{\#}$ or -C_{σ}) on VC transitions (changes in intensity, in fundamental and formant frequencies from 40 to 90% of the preconsonantal vowel duration)

A statistical significance level of $p = 0.05$ was used for all tests.

Results

Duration

Figure 2 shows the mean normalized values of stop and nasal durations (in milliseconds) according to their position within the syllable. Stops and nasals are of comparable length in monosyllable onsets (C-) and codas (-$C_{\#}$), while both are

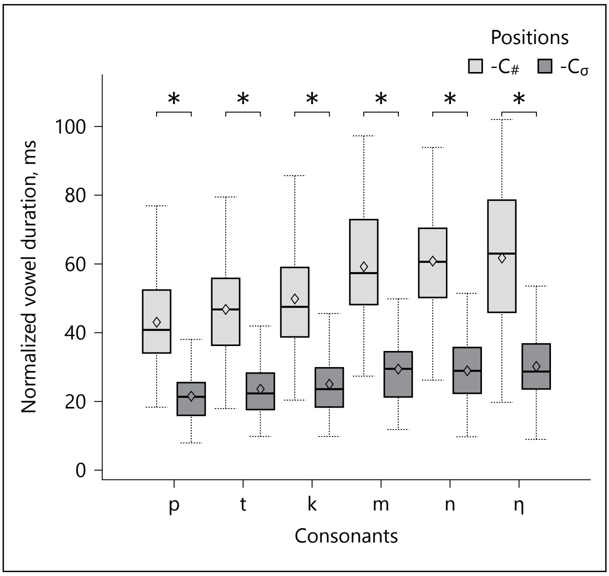


Fig. 3. Boxplot with symbol (◊) representing the means of normalized duration (ms) of vowel /a/ when followed by stops and nasals, according to the boundary types (interword -$C_{\#}$ or intraword -C_{σ}). * $p < 0.05$.

shorter in disyllabic word-internal codas (-C_{σ}) (the detailed results are given in the Appendices).

The within-subject analyses reveal significant effects overall for positions on the duration for stops (/t/ and /k/) [F(2, 18) = 34.636; $p < 0.001$] and for nasals [F(2, 18) = 36.287; $p < 0.001$]. However, for C- and -$C_{\#}$ positions in monosyllables (CV, CVC) specifically, no significant differences are found for /t/ and /k/ (recalling that Vietnamese prohibits syllable-initial /p/) [F(1, 9) = 0.068; $p > 0.7$], or for nasals [F(1, 9) = 2.235; $p > 0.1$]. Focusing more narrowly on C- and -$C_{\#}$ of the CVC structures, there is again no difference in the durations observed for stops [F(1, 9) = 0.045; $p > 0.8$] or for nasals [F(1, 9) = 1.979; $p > 0.1$]. It is also observed that /a/ is realized longer in CVC when preceded by a stop and followed by a nasal (e.g., /tam/) than in the opposite pattern (e.g., /mat/) [F(1, 9) = 19.188; $p < 0.002$].

Comparison of consonant durations in C- and in {-$C_{\#}$, -C_{σ}} shows differences between these 2 sets of position (i.e., onset vs. coda), attested within subjects for stops [F(1, 9) = 17.831; $p < 0.003$] and for nasals [F(1, 9) = 28.465; $p < 0.001$].

No significant effects of place of articulation on duration are found for stops [F(1, 9) = 0.011; $p > 0.9$]. There is a significant difference in duration for nasal obstruents but only between the labial /m/ and velar /ŋ/ consonants [F(1, 9) = 10.317; $p < 0.02$].

As found previously in the single-speaker study of Trần and Vallée (2009), these within-subject results show differences in the production of both stop and nasal consonants as a function of coda type (-$C_{\#}$ or -C_{σ}), namely:

- Consonant durations are significantly different for stops [F(1, 9) = 78.618; $p < 0.001$] and for nasals [F(1, 9) = 40.863; $p < 0.001$], thus showing an effect of the boundary type on the consonant duration and suggesting that speakers might use different degrees of coarticulation between word-internal consonant sequences compared to those across word boundaries. No significant effects of place of articulation on the consonant duration are observed here, neither for stops [F(2, 18) = 0.938; $p > 0.4$] nor for nasals [F(2, 18) = 0.122; $p > 0.8$]

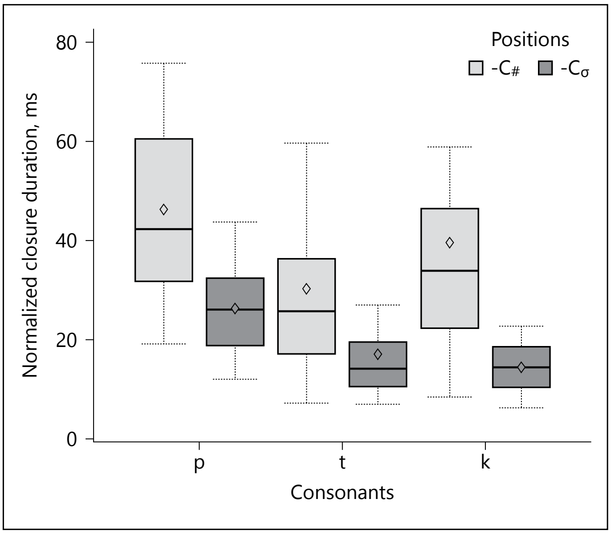


Fig. 4. Boxplot with symbol (◊) representing the means of normalized closure duration (ms) of stops according to within-word positions.

- The vowel /a/ is significantly longer before -$C_{\#}$ than before -C_{σ} (Fig. 3) for both stops [F(1, 9) = 104.923; $p < 0.001$] and nasals [F(1, 9) = 169.927; $p < 0.001$]. Vowel duration also varies significantly according to the place of articulation of the coda consonant: [a^k] duration > [a^t] duration ($p < 0.03$), and [a^k] duration > [a^p] duration ($p < 0.009$); however [a^t] duration ~ [a^p] duration ($p > 0.05$). This place effect is not observed for final nasals [F(2, 18) = 0.921; $p > 0.4$]

This result indicates that vowel duration contains information on the place of articulation of the following coda stop in syllables with tone D1, as the vowel is longer when followed by the velar stop /k/ than when followed by the coronal /t/ or bilabial /p/.

- For closure duration of coda consonants, word-final stops have proportionally longer closing phases than within-word coda stops [F(1, 133) = 6.723; $p < 0.015$] (Fig. 4). No significant effect is observed according to the place of articulation [F(2, 133) = 1.485; $p > 0.2$]

Burst

Vietnamese stop consonants are always accompanied by bursts in word-initial positions, but they are traditionally described as unreleased in the coda (Cao, 1985; Đoàn, 1999). However, analysis of our multispeaker data shows the presence of bursts after word-final coda stops (22.55%) and to a lesser extent before word-internal coda stops (4.72%). Unreleased stops are thus more frequent at word boundaries than at word-internal syllable boundaries.

Burst durations are significantly shorter in coda positions (-$C_{\#}$ and -C_{σ}) than in onset position C-, for /k/ [F(1, 309) = 103.568; $p < 0.001$] and for /t/ [F(1, 322) = 12.969; $p < 0.001$]. Mean normalized burst duration is 4.78 ms for /k/ in C- but 2.26 ms in -$C_{\#}$ and 1.52 ms in -C_{σ}, while for /t/, it is 2.51 ms in C-, different from -$C_{\#}$ with 2.04 ms and -C_{σ} with 1.68 ms. As for the coda positions, the burst duration in -$C_{\#}$ is not significantly different from that in -C_{σ}, neither for /p/ [F(1, 60) = 4.176; $p > 0.1$] nor for /k/ [F(1, 29) = 2.224; $p > 0.1$] nor for /t/ [F(1, 44) = 1.816; $p > 0.1$].

Burst intensities of stop consonants are significantly different depending on within-word initial versus final positions C- vs. -$C_{\#}$ and -C_{σ} for /k/ [F(1, 19) = 41.325; $p < 0.001$] and for /t/ [F(1, 19) = 17.590; $p < 0.001$]. Mean energy bursts are higher in C- (65.43 dB for [k] and 61.68 dB for [t]) than in -$C_{\#}$ (56.95 dB for [k]; 56.38 dB for [t]) or in -C_{σ} (58 dB for [k]; 57.26 dB for [t]). No significant differences in burst intensity are found for word-internal versus word-final coda stops [F(1, 31) = 0.07; $p > 0.8$].

To generalize, coda stops are not typically released in Vietnamese, but when released, their bursts are significantly shorter and less intense than stops in onset position. However, no significant differences in duration or intensity are found between coda stops in word-internal and word-final positions.

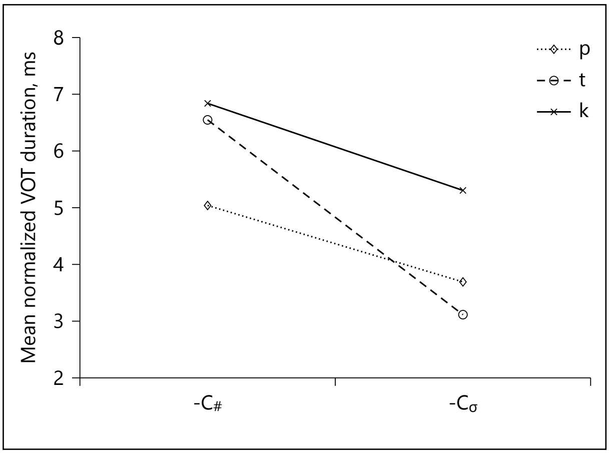


Fig. 5. Mean normalized voice onset time (VOT) duration (ms) of stops as a function of within-word positions.

Voice Onset Time

Mean normalized VOT values of bilabial and alveolar stops vary significantly with within-word positions. The VOT is proportionally longer in -$C_{\#}$ than in -C_{σ} for /p/ [F(1, 60) = 8.617; $p < 0.005$] and for /t/ [F(1, 40) = 4.638; $p < 0.04$] (Fig. 5). VOT differences depending on word positions was not found for /k/ [F(1, 29) = 0.514; $p > 0.4$].

As expected, VOTs also depend on stop places of articulation, with the longest mean normalized VOT for the voiceless velar stop, regardless of its position in the syllable or in the word (8.88 ms for /k/ vs. 4.94 ms for /t/ in C- position; 5.31 ms for /k/ vs. 3.17 ms for /t/ in word-internal position -C_{σ}). However, VOT differences between /k/ and /t/ are negligible in word-final position (6.48 and 6.03 ms, respectively), while word-final VOTs of /p/ are distinctly shorter (5.04 ms).

Vowel-Consonant Transition

In our data, we note a large number of coda stops in which the consonant burst is absent, as shown in the following spectrogram *gác bút* [ɣak^{D1}. but^{D1}] "to stop writing" *or* "to abandon the writing profession" (Fig. 6). The spectrogram shows no release burst for /k/ realized as [k̚]. As the silent closure phase is not followed by a noise burst, final consonants are necessarily identified from the transitional phase with the preceding vowel segment, as noted by Cao (1985, p. 83). Since Lisker (1957), Liberman et al. (1954), Delattre et al. (1955), or Delattre (1963), many studies have also demonstrated that formant transitions are cues for consonant perception and identification, and that they prevail over burst when a conflict occurs between burst and formant transitions (Walley and Carrell, 1983). Because of the high proportion of unreleased final consonants in our data, the acoustic parameters of vowel-consonant transitions (the first 3 formants, fundamental frequency, and intensity) were measured from 40 to 90% of the vowel duration in order to examine whether acoustic information of final stops depends on boundary types (inter-word vs. intraword).

Formant measurements in this acoustic signal portion are very tricky because of the influence of anticipatory coarticulation on formants. After testing several methods for transition analysis, LPC (linear predictive coding) was retained for this study, because it has been used successfully in many studies on formant detection and analysis (Markel and Gray, 1976; Spanias, 1994; Nearey et al., 2002; Adank et al., 2004), and it has proven to be the most reliable method for the analysis of highly variable vowels, like those produced by children (Ménard, 2002; Ménard et al., 2010).

LPC analyses were performed to extract values of the first 3 formants in the following way. The sound was resampled at 11,000 Hz and the signal preamplified from 50 Hz. The LPC parameters were adjusted for each analyzed vowel. The first parameter adjusted was the number of poles, ranging from 12 to 16. Then, for each of the specified time points (40, 50, 60, 70, 80, and 90% of the vowel duration), 5 separate analyses were computed to avoid spurious formant detection. Each of these 5 analyses

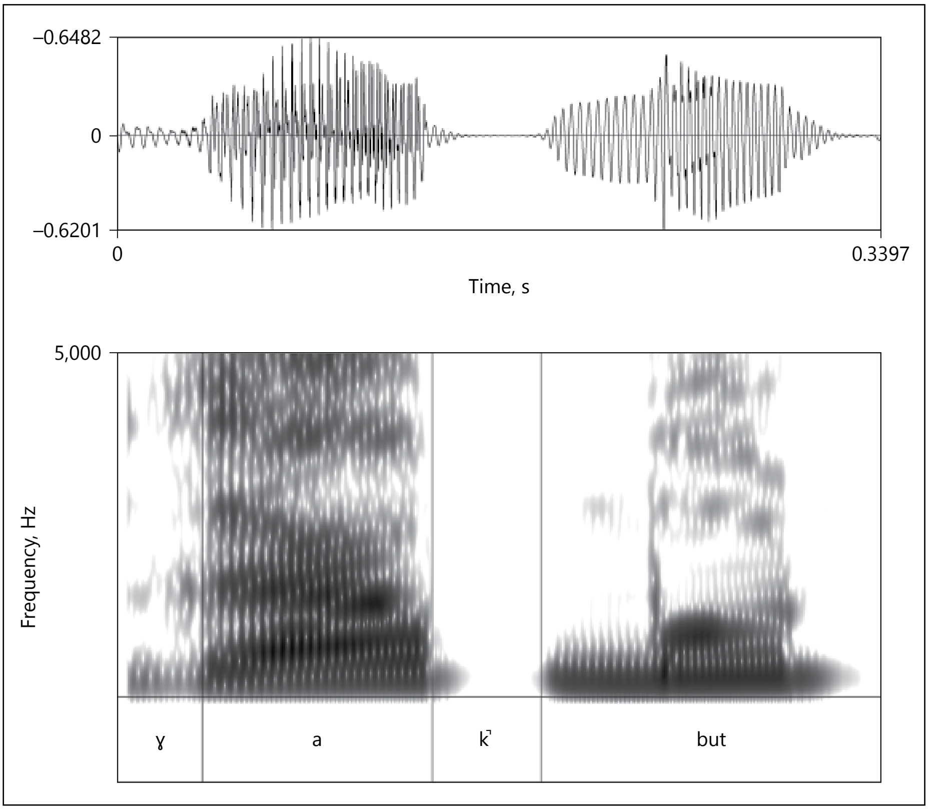


Fig. 6. Audio signal and spectrographic representation of the Vietnamese disyllabic word *gác bút* [ɣak^{D1}. but^{D1}].

used a time window of 23.3 ms, centered on the specified time point with an offset of respectively –2, –1, 0, +1, and +2 ms. Each parameter pair (number of poles, time offset) gave a spectral response curve, resulting in a set from which we selected the closest curve to the calculated average one. The reference values (average and maximum) for the vowel were provided to the formant detection algorithm. To ensure measurement accuracy by the LPC algorithm, automatically detected formant values were verified on a broad-band spectrogram. When automatic detection did not correspond to the spectrographic representation, the number of poles was readjusted, and the LPC procedure was applied again.

F_0 values were extracted by using the automatic detection algorithm RAPT (robust algorithm for pitch tracking). The raw intensity data were collected using Praat, and the smoothing was performed with Matlab (Talkin, 1995).

The data were analyzed using a repeated-measures ANOVA with the factors *place of articulation* (bilabial, coronal, velar) of final consonant -$C_{\#}$ vs. -C_{σ} and *the syllable tone* (neutral tone A1, rising tones B1-D1, falling tones B2-D2) of the target word and/or syllable.

Formant Transitions

Mean values of the first 3 formants measured at the 6 points located from 40 to 90% of the vowel duration by steps of 10% were analyzed in relation to the final-stop consonant *place of articulation* and to the *tone* of the next syllable. Recall that monosyllabic target words and first syllable of disyllabic target words all had the same rising tonal context (*sắc* tones, B1 and D1).

Table 1. Significance levels for the means of F_1, F_2, and F_3 measured from 40 to 90% of vowel duration, as a function of the factor *consonant place*

	df	40%		50%		60%		70%		80%		90%	
		F	*p*	F	*p*	F	*p*	F	*p*	F	*p*	F	*p*
F_1	2; 18	9.645	*0.001*	4.677	*0.023*	1.577	0.234	0.742	0.490	0.269	0.767	0.920	0.410
F_2	2; 18	2.294	0.130	1.266	0.306	2.206	0.139	9.906	*0.001*	9.132	*0.002*	12.644	*0.000*
F_3	2; 18	2.786	0.880	5.565	*0.013*	2.208	0.139	4.585	*0.025*	1.956	0.17	1.876	0.181

Significant *p* values are italicized.

Table 2. Significance levels for the means of F_1, F_2, and F_3 measured from 40 to 90% of vowel duration, as a function of the factor *tone*

	df	40%		50%		60%		70%		80%		90%	
		F	*p*	F	*p*	F	p	F	*p*	F	*p*	F	*p*
F_1	2; 18	0.147	0.865	0.648	0.535	2.478	0.112	9.808	*0.001*	21.035	*0.000*	6.270	*0.009*
F_2	2; 18	73.306	*0.000*	83.199	*0.000*	27.6996	*0.000*	40.590	*0.000*	32.745	*0.000*	16.715	*0.000*
F_3	2; 18	0.049	0.953	0.101	0.904	0.048	0.954	0.561	0.580	0.578	0.571	0.156	0.857

Significant *p* values are italicized.

Table 1 shows F_1, F_2, and F_3 differences at various points of the vowel duration, according to the place of articulation of stops in coda, and regardless of the boundary types which immediately follows the final consonant (interword or intraword).

Differences in F_1 values according to the final-consonant place of articulation are significant at the vowel center (both 40 and 50%). A more detailed analysis shows that a bilabial place of articulation reinforces the differences (significantly higher than differences for other consonant places). The significant effect observed at 50% for F_3 is caused by the coronal place.

For the computed mean values of F_2, significant effects are found at 70% ($p < 0.001$), 80% ($p < 0.002$), and 90% ($p < 0.001$). F_3 values are significantly different at 70% of the vowel duration ($p < 0.03$), while the differences in F_1 are never found to be significant in this portion of VC transition ($p > 0.05$). These results clearly suggest that the final acoustic part of the vowel, and more specifically F_2 transition, provides clues of postvocalic consonant place of articulation.

Regarding the factor *tone*, we note that a neutral level tone A1 on the syllable following a monosyllabic word influences the values of both F_1 from 70% and F_2 from 40% of the vowel duration (Table 2). In disyllabic words, no significant differences are found in F_1, F_2, and F_3 in the first syllable due to the tone (rising *sắc* or falling *nặng*) on the second syllable ($p > 0.05$).

Mean values and standard deviations of F_1, F_2, and F_3 changes from 50 to 90% of the vowel duration are given in the Appendices. The slopes were obtained by calculating the difference between the formant values at 40% and those at each interval (from 50 to 90% by steps of 10%). For example, ΔF_1 at 90% = $F_{1(90\%)} - F_{1(40\%)}$; ΔF_2 at 90% = $F_{2(90\%)} - F_{2(40\%)}$; ΔF_3 at 90% = $F_{3(90\%)} - F_{3(40\%)}$. ΔF_0 and ΔI were obtained in the same way. The calculations were carried out according to the coda consonants /p/, /t/, /k/, and to the tone of the following adjacent syllable.

Change over time in the 3 formants in the final portion of the vowel is presented in Figures 7–9, respectively. In these figures we labeled the curves by coda consonant (p, t, or k) and the following tone (level, rising, or falling). Codas in monosyllabic words are followed by A1 level tone (in their

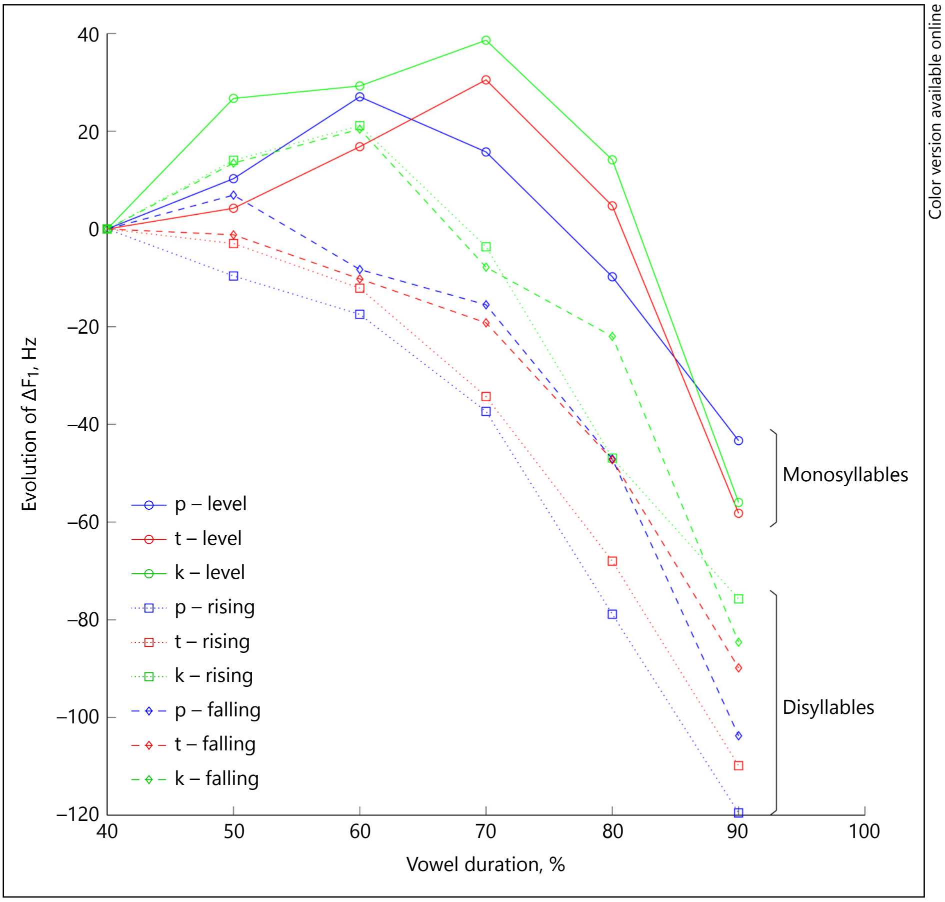


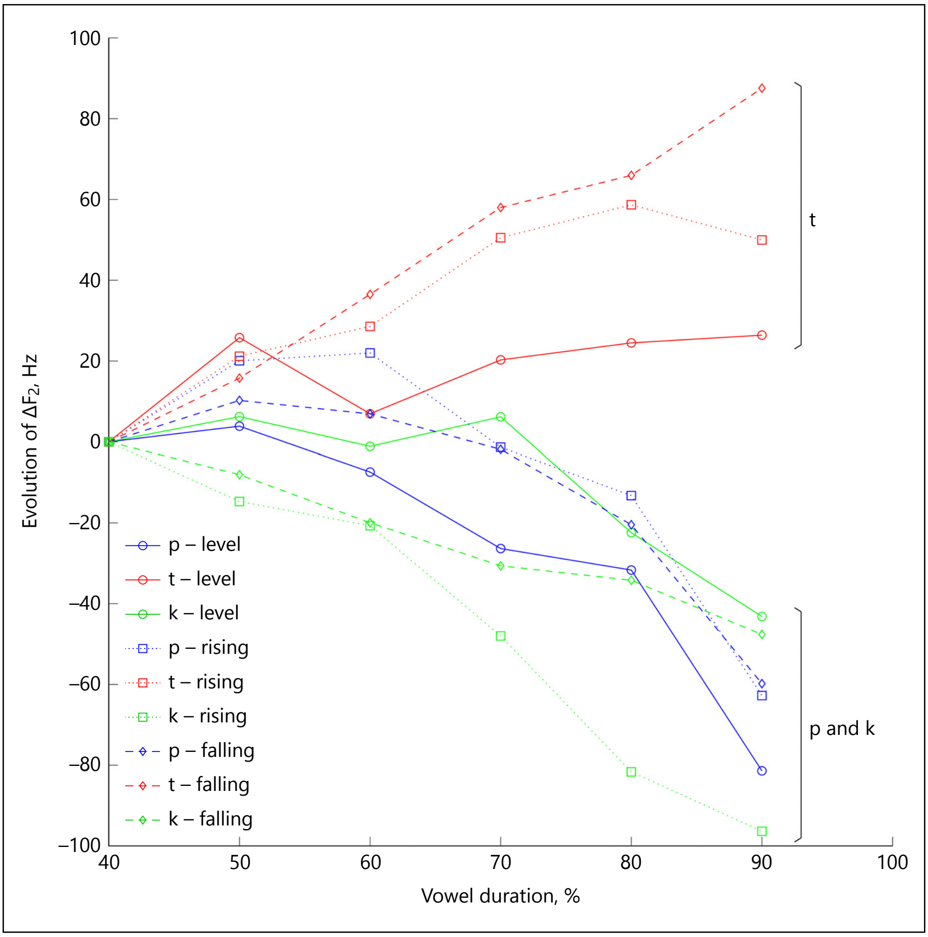
Fig. 7. Time evolution of ΔF_1 estimated from 50 to 90% of the vowel duration according to the coda place and to the tone of the following syllable.

frame sentences), and first syllable coda consonants in disyllabic words are followed in the second syllable by B1 or D1 rising tones or by B2 or D2 falling tones.

From 50% of the vowel duration, ΔF_2 is already significantly different, depending on the place of articulation of the coda consonant, and remains significantly different until the end of the vowel. Figure 8 shows that from 70%, the slopes of F_2 are positive in the context of /t/ realizations, while those of /p/ and /k/ are all negatives. In this VC transition portion, ΔF_3 has no significant difference due to the coda's place of articulation ($p > 0.05$), probably because of the variability of the observed F_3 values (Fig. 9). ΔF_1 is significantly different at 60%, which may be due to the randomness of the sample.

A gender effect is observed, with an interaction between gender and coda consonant place of articulation, showing significant effects on both ΔF_2 and ΔF_3 at 80% ($p < 0.001$ and $p < 0.04$, respectively), and also at 90% of the vowel duration ($p < 0.001$ and $p < 0.008$, respectively). This means that ΔF_2 and ΔF_3 distinguish coda place of articulation differently between males and females. An effect of the following tone is observable only for ΔF_1 from 60 to 90% of the vowel duration ($p < 0.05$) (Table 3; Fig. 7). The significant difference is driven by monosyllabic words (p – level, t – level, k – level).

Regarding disyllabic words, whatever the tone of the following syllable (rising or falling), no significant effect on ΔF_1 is detected ($p > 0.6$), and also no significant interaction between *tone* and *gender* on ΔF_1 is found.



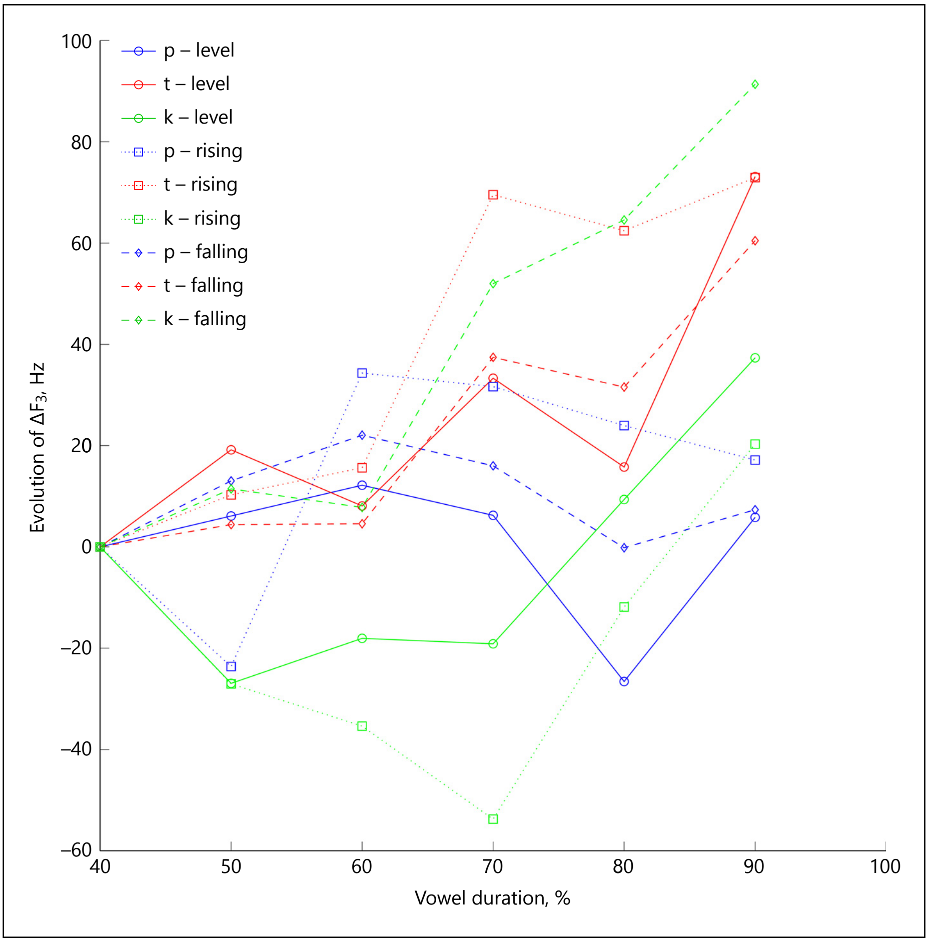
Color version available online

Fig. 8. Time evolution of ΔF_2 estimated from 50 to 90% of the vowel duration according to the coda place and to the tone of the following syllable.

In order to observe effects of nasal consonant place of articulation on formant transitions, we measured formants at 50 and 90% of the vowel duration only. Given their specific acoustic characteristics, especially regarding the presence of nasal formants on the spectrogram (Maeda, 1982; Rossato, 2000; Vaissière, 2007, 2008), we decided to perform ANOVA tests between VC transitions (oral stops) and VN transitions (nasal obstruents) with C and N having the same place of articulation. The goal was to verify the absence of nasal formants in the measured values of the first 3 formants of the preceding vowel.

Comparisons between ΔF_1, ΔF_2, and ΔF_3 values of the 3 possibilities /ap/ versus /am/, /at/ versus /an/, and /ak/ versus /aŋ/ show significant differences between [k] and [ŋ], and [p] and [m] (Table 4). This result indicates the possibility that nasal formant measurements are included in the data. Therefore, for nasal consonants, the approach used for stops was not retained here, to avoid extraction of unreliable and potentially misleading information regarding acoustic cues to coda place of articulation in the preceding /a/ vowel.

Additional analysis based on the calculation of 12 MFCC coefficients (mel-frequency cepstrum coefficients) and their derivatives over a sliding window confirms the existence of spectral



Color version available online

Fig. 9. Time evolution of ΔF_3 estimated from 50 to 90% of the vowel duration according to the coda place and to the tone of the following syllable.

characteristics of C in the VC transition portion at 90% of the vowel duration. A principal component analysis was performed based on these data (Fig. 10).

The principal component analysis shows that at 90% of its duration, the vowel contains information on the place of articulation of the postvocalic consonant: the dispersion areas are separate for velar consonants (/k/ and /ŋ/), and the coronal areas (/t/ and /n/) are located between those of velars and bilabials. The information in the MFCC that separates bilabials and velars is not visible in the examination of ΔF_2 (Fig. 8).

Intensity Transition

As was done for formant transitions, ΔI values were obtained by measuring intensity at different points through the vowel duration (from 50 to 90%) and then by comparing with measurements taken at 40%. A repeated-measures ANOVA was performed to observe effects of the 2 factors (place of articulation and tonal context) on ΔI.

Figure 11 shows mean values of ΔI from 50 to 90% of vowel duration calculated according to both the stop place of articulation and the tonal context. As a reminder, the *ngang* tone (high level A1)

Table 3. Significance levels for ΔF_1, ΔF_2, and ΔF_3 from 50 to 90% of vowel duration, as a function of factors *place* and *tone*

Factors	Δ	df	50%		60%		70%		80%		90%	
			F	*p*	F	*p*	F	*p*	F	*p*	F	*p*
Place	ΔF_1	2; 16	3.430	0.058	4.214	*0.034*	1.607	0.231	2.102	0.155	0.818	0.459
	ΔF_2	2; 16	7.114	*0.006*	14.831	*0.000*	23.642	*0.000*	84.929	*0.000*	66.156	*0.000*
	ΔF_3	2; 16	1.943	0.176	1.012	0.386	2.130	0.151	0.755	0.486	1.376	0.281
Tone	ΔF_1	2; 16	3.058	0.075	9.264	*0.002*	21.570	*0.000*	9.655	*0.002*	3.516	0.054
	ΔF_2	2; 16	0.893	0.429	0.639	0.541	0.613	0.554	1.356	0.286	2.402	0.122
	ΔF_3	2; 16	1.060	0.370	0.339	0.718	2.214	0.142	1.412	0.272	0.414	0.668

Significant *p* values are italicized.

Table 4. Comparison of ΔF_1, ΔF_2, and ΔF_3 between oral and nasal consonants

Oral vs. nasal	df	ΔF_1		ΔF_2		ΔF_3	
		F	*p*	F	*p*	F	*p*
[p] vs. [m]	1; 8	0.591	0.464	6.254	*0.037*	0.628	0 .451
[t] vs. [n]	1; 8	0.178	0.684	4.495	0.067	2.386	0.161
[k] vs. [ŋ]	1; 8	5.406	*0.049*	5.393	*0.049*	0.266	0.620

Significant *p* values are italicized.

indicates that the coda stop is word final ($-C_{\#}$), while the sắc tones (rising B1 or D1) and the nặng tones (falling B2 or D2) indicate that the coda stop is word internal ($-C_{\sigma}$). Mean values and standard deviations of ΔI are listed in the Appendices.

At 90% of the phonetic realization of the vowel, differences in ΔI occur depending on the tone of the following syllable [F(2, 18) = 7.196; $p < 0.005$] (Table 5). Intensity measured in VC transition of monosyllabic words (followed by a neutral tone A1) is significantly lower than intensity of within-word VC transition followed by a rising sắc tone B1 or D1 ($p < 0.02$) or by a falling nặng tone B2 or D2 ($p < 0.03$). No relevant differences between intensity of within-word VC transition according to tonal context (rising or falling) are found ($p > 0.3$). No effect of tone is observed from 50 to 80%.

At 90% of the total vowel duration, the decrease in intensity is significantly different according to the place of articulation of the following consonant [F(2, 18) = 29.344; $p < 0.001$]. The decrease is more pronounced with velar consonants than with bilabials ($p < 0.001$) or coronals ($p < 0.001$) (Fig. 11). Differences between bilabials and coronals are not significant ($p > 0.5$). A significant effect of the intensity decrease is already observed at 80, 70, and 60% as a function of the place of articulation [F(2, 18) = 28.95; $p < 0.001$]. Over the changing VC transitions, the greatest negative intensity slopes are obtained with velar consonants as opposed to bilabials ($p < 0.001$) and coronals ($p < 0.001$). Unlike with plosives, no significant effect of place of articulation on intensity is observed with nasals.

F_0 Transition

The same approach used for formant and intensity transitions was retained for ΔF_0. Mean values and standard deviations of F_0 slopes from 50 to 90% of the total vowel duration are presented in the

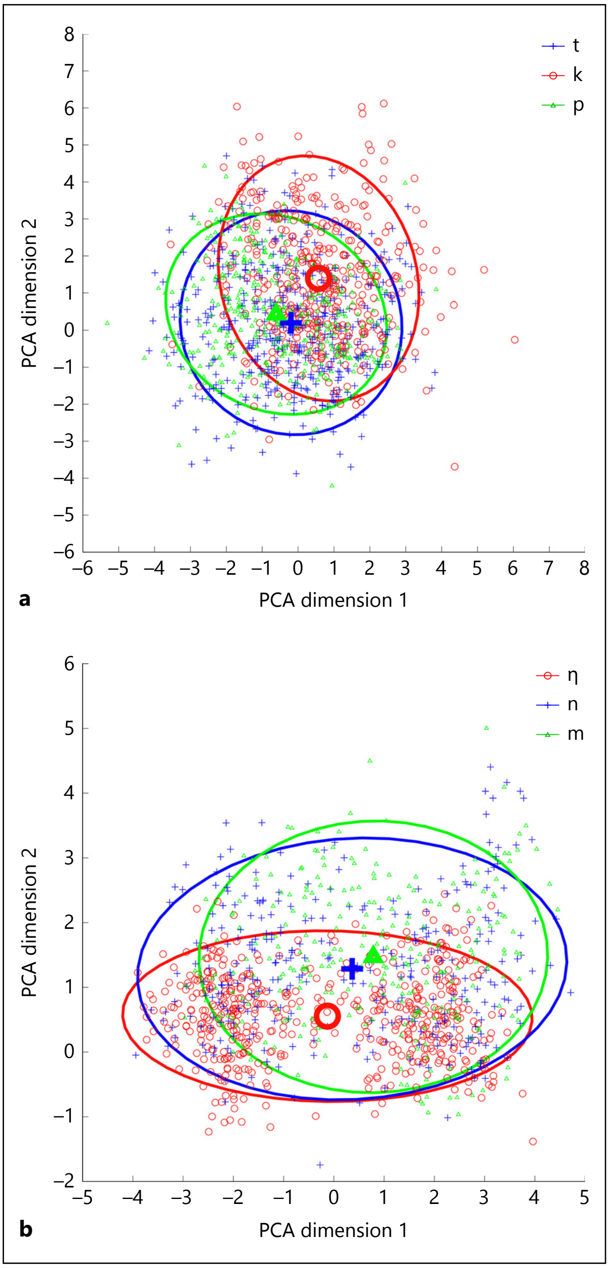


Fig. 10. Principal component analysis (PCA) of mel-frequency cepstrum coefficients of the transition portion at 90% of the vowel duration in the context of stops /p t k/ (**a**) and of nasals /ŋ n m/ (**b**).

Appendices. ΔF_0 was compared in terms of the place of articulation of the coda consonant (/p t k/) and the tone of the following syllable. Time evolutions of F_0 are presented in Figure 12.

As a reminder, the observed VC transitions were all in syllables with the rising tone *sắc* D1, which exhibited a rising pattern toward the syllable final part in all contexts. Significant differences in ΔF_0 values were observed from 70 to 90% of the total vowel duration according to the consonant place of articulation ($p < 0.007$, $p < 0.003$, $p < 0.001$, respectively) (Table 6). The rise of F_0 measured at 70,

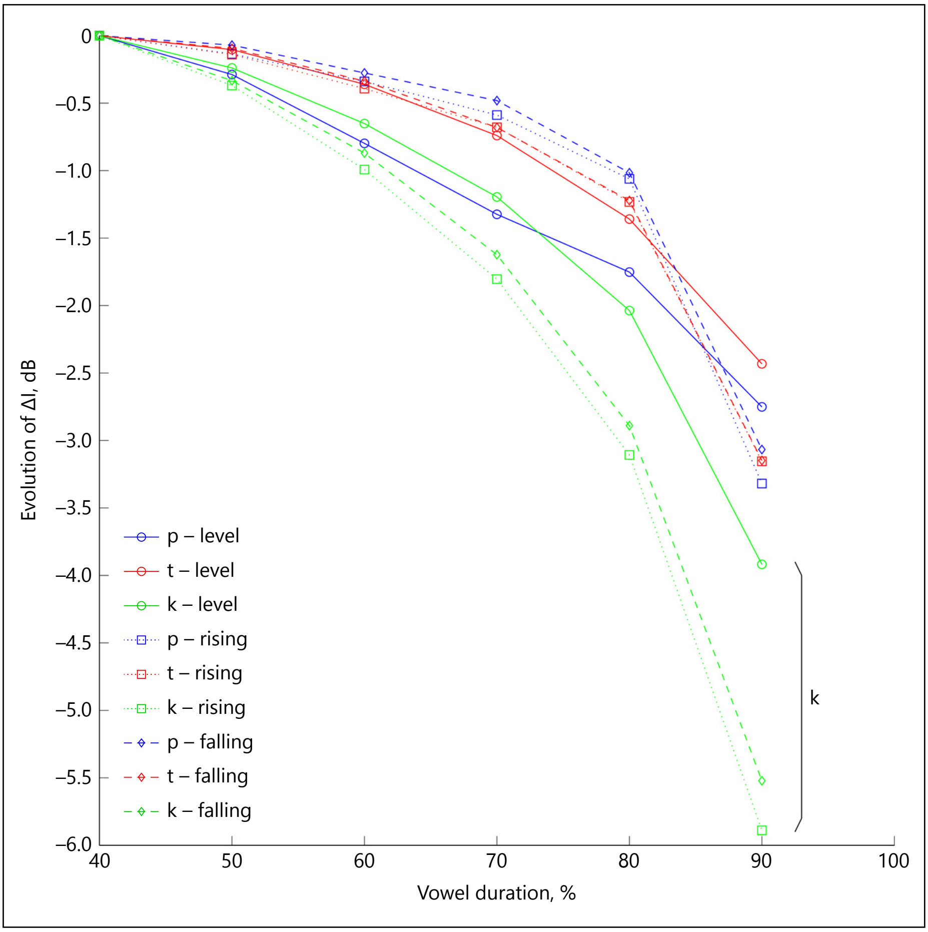


Fig. 11. Temporal evolution of intensity (mean ΔI) estimated from 50 to 90% of vowel duration, according to the coda place and to the tone of the following syllable.

Table 5. Significance levels of ΔI from 50 to 90% of vowel duration, as a function of factors *place* and *tone*

ΔI	df	50%		60%		70%		80%		90%	
		F	*p*	F	*p*	F	*p*	F	*p*	F	*p*
Place	2; 18	6.085	*0.010*	10.854	*0.001*	16.139	*0.000*	28.946	*0.000*	29.344	*0.000*
Tone	2; 18	0.237	0.792	0.272	0.765	0.242	0.788	0.059	0.943	7.196	*0.005*

Significant *p* values are italicized.

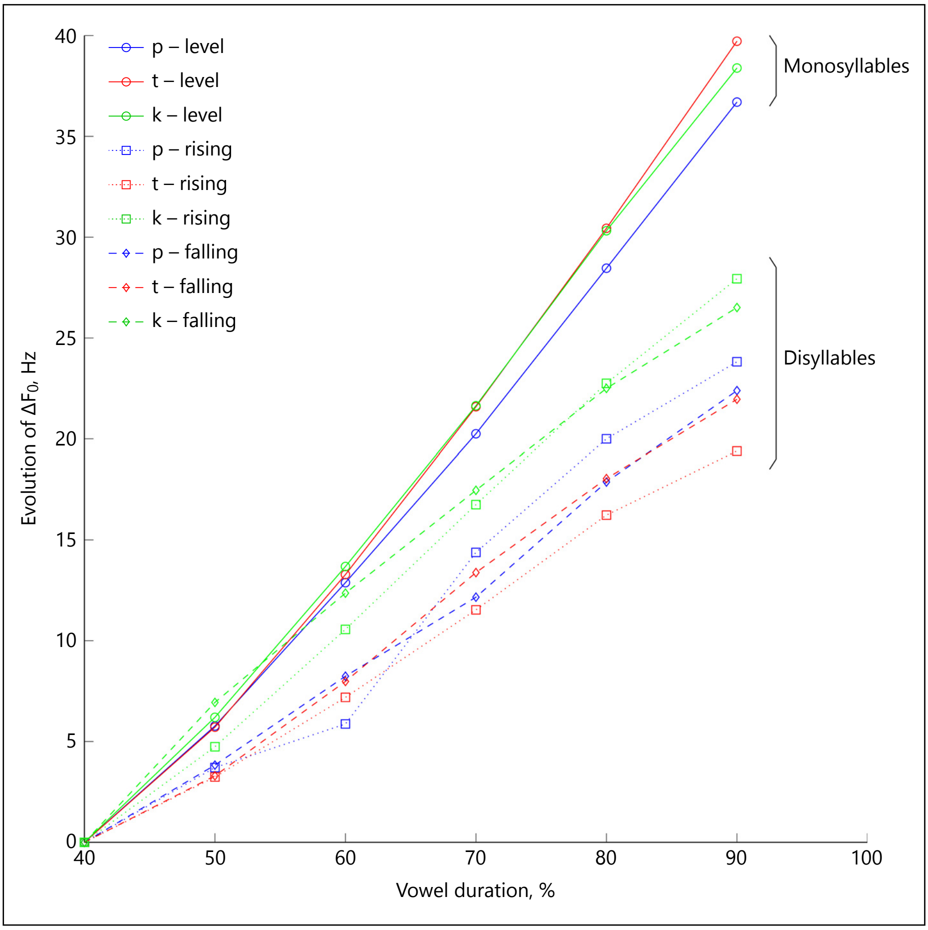


Fig. 12. Temporal evolution of ΔF_0 estimated from 50 to 90% of vowel duration in VC transitions, according to the coda place and to the tone of the following syllable.

Table 6. Significance levels of ΔF_0 from 50 to 90% of vowel duration, as a function of factors *place* and *tone*

ΔF_0	df	50%		60%		70%		80%		90%	
		F	*p*	F	*p*	F	*p*	F	*p*	F	*p*
Place	2; 18	3.884	0.061	1.023	0.379	12.054	*0.000*	9.348	*0.002*	6.784	*0.006*
Tone	2; 18	2.534	0.107	6.854	*0.006*	43.406	*0.000*	49.414	*0.000*	64.258	*0.000*

Significant p values are italicized.

80, and 90% of the vowel duration in V + velar sequences is generally greater than the rise of F_0 in V + bilabial ($p < 0.04$) or in V + coronal sequences ($p < 0.009$).

An effect of the following syllable's tone on ΔF_0 is observed from 60 to 90% of the vowel duration. Figure 12 shows that the F_0 rise rate is significantly greater from 60% of the vowel portion when it is within a monosyllabic word followed by an *ngang* high level tone than within the initial syllable of a disyllabic word followed by rising *sắc* tones or falling *nặng* tones. No significant ΔF_0 differences are found between the 2 tone types of disyllabic-initial syllables, high (*sắc*) and low (*nặng*) ($p > 0.9$).

Discussion

Unreleased Final Consonants and Acoustic Cues for Place of Articulation

Vietnamese stop consonants are traditionally described as unreleased in coda position (Cao, 1985; Đoàn, 1999). However, our results show that syllable-final stops are not always unreleased, and indeed 27% of them are produced with a burst, but with shorter duration and lower intensity than those of syllable-initial stops. Probably due to coarticulation, Vietnamese stops are found without a burst more frequently before an intrasyllabic word boundary (-C_σ) than before an interword boundary (-$C_\#$), and never in a word-initial position. The results of the ANOVA performed on acoustic measurements of final consonants show no significant variation in burst duration or burst intensity according to coda word-final versus word-internal positions ($C_\#$ *vs*. C_σ). Other linguists as Hoàng and Hoàng (1975) and Cao (1985) have highlighted the difference between 2 categories of released initial versus unreleased final stops with minimal pairs of bimorphemic and bisyllabic constructions, such as the words *phát hành* [fat^{D1}. hĕŋ^{A2}] "to publish" and *phá thành* [fa^{B1}. tʰĕŋ^{A2}] "to demolish the town walls". Resyllabification is not allowed across syllable or word boundaries in Vietnamese (Cao, 1985), thus final stops, often unreleased, never move from a coda position to the next syllable onset position.

Trần and Vallée (2010) have shown, in a perceptual identification test of Vietnamese final consonants by native listeners, that the presence of a burst improves consonant identification scores, even though the number of released final stops /p t k/ (with presence of burst) in the test is much smaller (25, 45, and 22%, respectively) relative to the number of unreleased ones. However, for the latter, the identification scores were above the threshold of chance (i.e., >50%). This result indicates that unreleased stops include other acoustic cues than just the burst, with information on their place of articulation.

Analysis of F_0 contours, formant trajectories, and intensity curves observed in the VC transition indeed confirms that the second half of a vowel carries acoustic information on the place of articulation of the following consonant (which is not released in most cases). At 70% of the vowel duration, mean values of F_2 and F_3 change significantly depending on the syllable-final consonant's place of articulation. At 80% and 90%, differences in F_2 are significant, but there are no notable differences in F_3. ΔF_2 values measured from the time point corresponding to the vowel center are only significantly different according to the following consonant's place of articulation. However, we do not observe the rising F_2 transition for the velar consonant /k/ as proposed by Delattre et al. (1955) and Delattre (1963) from synthetic speech. F_2 transition is falling for both /p/ and /k/. It would make sense if there were forms like *túc* [tu$\widehat{k^{p}}$], *tốc* [to$\widehat{k^{p}}$], *tóc* [tɔ$\widehat{k^{p}}$], with a labialized velar final in our corpus. But as the vowel was always /a/, it is somewhat mysterious. An extension of the study with a wider corpus, including all vowel contexts, would be interesting to understand this result.

Spectral information obtained with MFCC at 90% of the vowel duration shows, on the other hand, that the acoustic dispersion areas for the Vietnamese velar stop are more separate from those of labial and coronal ones (back vs. front stop).

In addition, significant differences in ΔI or ΔF_0 are found depending on the final stop's place of articulation. Velar stops show trends both in intensity decrease and F_0 rise in a more important manner than bilabials and coronals. These findings are consistent with Cao's proposal (1985) that identification of place of articulation of unreleased final stops is possible through recoverable information located in the transition portion of the preceding vowel nucleus. Our study suggests that information for consonant place identification can be found in dynamic acoustic cues from F_0, F_2, and intensity slopes. To confirm this finding, a perception experiment with native speakers should be designed to determine the degree of pertinence or the relative importance of these dynamic acoustic parameters in unreleased stop identification.

As it is well known, VOT varies with consonant place of articulation (Cho and Ladefoged, 1999). On average, in our study, VOT values of /k/ are always higher than those of /p/ or /t/, irrespective of within-syllable or within-word position. For any given place of articulation, there are differences from one language to another (Cho and Ladefoged, 1999). Comparing our results with those of Serniclaes (1987) on the French voiceless stops /p t k/ shows that mean VOT in Vietnamese (/k/ being the longest in onset position, approx. 23.94 ms) is about 30% shorter than in French (approx. 35 ms). The variability of positive VOT in Vietnamese (from 12 to 24 ms) is much less than in French, where the range extends from 10 to 70 ms (Serniclaes, 1987, p. 125). The much higher variability of French VOTs is due to the stops' places of articulation and the places of the following adjacent vowels: VOT is lengthened with both retraction of the place ([p] > [t] > [k]) and vowel aperture ([a] > [i], [u]) (Serniclaes, 1987, p. 127). Other works show the same effect of vowel context on VOT length in stop production, mostly in French and English (Fischer-Jørgensen, 1972; Yeni-Komshian et al., 1977; Terrance and Bernard, 1994). In Vietnamese, we also think that VOT variability depends not only on the consonant articulation places, but also on various vowel contexts. Having said that, at this time, we need further work to clearly address this issue.

Final Consonants and Effect of the Syllable Boundary Type on Their Production

Our study examines acoustic features of Vietnamese stops and nasals according to their position within words and, more precisely, differences between the first segment of 2 adjacent consonants depending on whether they are within-word or across-word consonant sequences. Our multispeaker experiment confirms results of the pilot study of Trần and Vallée (2009) on speech production of a single adult native speaker and shows clear differences in the acoustic realization of Vietnamese coda consonants in word-internal versus word-final positions.

Vietnamese coda stops and nasals are significantly longer word-finally in monosyllables (CV$\mathbf{C}_{\#}$) than word-internally in the initial syllable of a disyllabic word (CV$\mathbf{C}_{\sigma}$.CVC). The fact that all duration measurements are normalized to the speech rate clearly shows that observed differences in duration (consonant, closure, and vowel) between -$\mathbf{C}_{\#}$ and -$\mathbf{C}_{\sigma}$ are not a result of disyllables being produced at an overall faster rate than monosyllables and thus are not related to speech rate. This result suggests that observed duration differences are probably the result of different degrees of coarticulation between intraword versus interword consonants: due to a stronger blending (more overlap) between successive consonants across syllable boundaries within polysyllables, temporal coproduction is greater in consonant sequences -**C**$_σ$.C- and less if the consonants are heterosyllabic -**C**$_{\#}$C- across word boundaries. These results are consistent with those of many previous studies which have found that phonetic features of phonemes in various languages were influenced by within-syllable and/or within-word positions (Lindblom, 1983; Keating, 1983; Browman, Goldstein, 1995; Byrd, 1996; Fougeron and Keating, 1997; Keating et al., 1999; Redford, 1999; for a review, see Krakow, 1999).

A new finding of our study is that the effect of syllable boundary type is also present in the nucleus, more precisely in the second half of the preconsonantal vowel. Indeed, a number of relevant results regarding duration parameters (final consonant duration, closure duration, vowel duration, VOT duration) show acoustic effects of syllable boundary type (inter- or intraword) on syllable rhyme. Stop consonants in -C$_{\#}$ have a longer VOT duration and a longer closing phase than in -C$_σ$. The vowel nucleus is significantly longer when the following consonant is before a word boundary than before an intraword boundary. This latter result holds for stops as well as for nasals.

Characteristics of the acoustic transition from the vowel nucleus into the final consonant also highlight effects of syllable boundary type. The F$_1$ trajectory from 60 to 90% of vowel duration is significantly different between -C$_{\#}$ and -C$_σ$. The intensity decrease at 90% of vowel duration is less severe before -C$_σ$ than before -C$_{\#}$. The continuous F$_0$ trajectory of each target syllable differs significantly from 60% of the vowel duration between monosyllabic words and disyllabic-initial syllables.

These results clearly suggest that VC transitions contain acoustic cues at the end of a 1-syllable word or a word-internal nonfinal syllable. Our data support the findings of Cutler and Norris (1988), Cutler and Butterfield (1991) that in clear speech, lengthening of preboundary syllables was used as one of the strategies by the speakers to mark word boundaries. Furthermore, their prosodic analyses of the syllable following a critical word boundary showed that more lengthening as well as greater increases in intensity and in F$_0$ were also applied in clear speech to weak syllables compared to strong ones in order to mark word boundary. These effects were, however, very small in comparison to the durational effects they observed for syllables preceding the boundary and for pauses at the boundary.

No significant ΔF$_0$ difference is observed in our data between disyllabic-initial syllables followed either by rising tones (B1 or D1) or by falling tones (B2 or D2). This result confirms previous studies on tonal effects in Vietnamese polysyllabic compound words (Han and Kim, 1974; Brunelle, 2003) that Vietnamese has a perseverative rather than an anticipatory tonal coarticulation pattern. This means that F$_0$ of the first syllable is not affected by the following syllable tone. An additional explanation for this result may come, in our view, from a phonotactic rule which constrains stop consonant sequences internally to the disyllabic words selected for the study: cross-syllable sequences within words contain a voiceless consonant followed by a voiced one. Such a pattern of coarticulation (voiceless-voiced combination) involves a short period of time (during the production of the voiceless stop) in which the vocal folds may have time to adjust their vibration rate to the production of the following syllable’s tone.

Conclusion and Perspectives

As in some other Asian languages such as Cantonese, Thai, and Korean, syllable-final stops in Vietnamese are usually unreleased. Evidence of acoustic cues carrying information on coda consonant place of articulation is found in our study (F_0 contours, formant trajectories, and intensity curves observed in VC transition). This finding is consistent with previous studies according to which "the lack of release bursts does not impair intelligibility at least for native speakers of those languages" because "the closing gestures of the stops include a component that compensates somewhat for the absence of release" (Abramson and Tingsabadh, 1999; see also Wright, 2004).

We plan to complete this work with a perceptual study in order to examine whether native listeners are able to use acoustic feature information to identify the correct nature of a syllable. In other words, do they know, from solely auditory input corresponding to isolated identical syllables, whether these make up a monosyllable or the first syllable of a disyllabic word? Trần (2011) suggested in a preliminary perceptual study that Vietnamese subjects were able to distinguish by ear whether target syllables in isolation (i.e., extracted from their surrounding context) belonged originally to a disyllabic or monosyllabic word. Trần's study should be extended by using the same syllabic sequences produced both in isolation and in a word-medial context in order to demonstrate that listeners are able to use acoustic properties of the rhyme to make decisions about a syllable's position in the word and that their performance is not influenced by phonological neighborhood density. Such an experiment would provide evidence on native speakers' capacity to distinguish between different types of syllabic boundaries (intraword vs. interword) from acoustic cues of the rhyme.

Our results show an effect of syllable boundary type on the production of Vietnamese final consonants. A word-final syllable boundary was found to be acoustically different from a word-internal syllable boundary within a compound word. Indeed, it has been shown that the realization of a number of acoustic parameters (final consonant duration, closure duration, VOT duration, vowel duration, F_0, F_1, and intensity) varies as a function of syllable position (word-internal vs. word-final). These differences in phonetic characteristics of coda consonants between monosyllables and disyllables suggest that Vietnamese disyllables have their own prosodic status rather than being simple juxtapositions. On this point, our findings are not consistent with Noyer (1998), who found "no evidence for separating 'word'-sized units from 'morpheme'-sized units in Vietnamese" (p. 92), and with Schiering et al. (2010), who added that "a monosyllabic word in Vietnamese is prosodically indistinct from other syllables and polysyllabic words behave prosodically like other polysyllabic strings, i.e. phrases" (p. 661). Since compounds and phrases often have the same syntactic structure, the acoustic differences we found between monosyllables and disyllables may help listeners to know where a word ends. Consequently, it will be worthwhile to set up another experimental protocol and to carry out a perceptual study for a better understanding of word types in Vietnamese.

A few gender differences are found in the production test with an interaction of final-consonant place of articulation. Interspeaker variation should be better investigated in order to check whether subgroups of homogeneous speakers exist.

As a reminder, our corpus consists of 35 disyllables (compound and reduplicated words, including opaque Sino-Vietnamese compounds and polysyllabic borrowings).

No phonetic differences are found in VC$_\sigma$ transitions between disyllables used in this study, whatever the tone of the second syllable. Although we did not set up an experimental protocol to highlight these grammatical aspects, but these results suggest that different types of disyllables may be not prosodically distinct from each other. Further works focusing on different types of disyllables will be required to clearly address the issue.

Interesting findings from the present study in read speech also encourage us to extend our multispeaker investigations in Vietnamese spontaneous speech. The first preliminary results confirm, as in the read speech, the existence of word position effects on final consonant realizations (Trần, 2015). In addition, a number of articulatory studies have investigated consonant overlap patterns in English, Russian, Korean, Taiwanese, Cantonese, and Georgian (Browman and Goldstein, 1990; Zsiga, 1994; Byrd, 1996; Zsiga, 2000; Cho, 2001; Chitoran et al., 2002; Wright, 2004; Yanagawa, 2006; Gao et al., 2011; Kochetov, 2013). They found that gestural overlap between consonant sequences can differ considerably across languages. For example, consonants across syllable boundaries in Taiwanese showed less overlap than in English (Yanagawa, 2006). Also, the greater overlap between heterosyllabic coda and onset consonants in Taiwanese may be related to the unreleased nature of the coda consonant in this language (Gao et al., 2011]. In perspective, it would be interesting to investigate articulatory overlap in Vietnamese consonant sequences with kinematic data and to discuss the relationship between degrees of overlap and syllable boundary types.

Appendices

Appendix 1

List of 62 items selected from the Vietnamese lexicon for the acoustic study of stops and nasals. Target consonants are highlighted in bold.

No.	Stimuli (orthographic shape)	Transcription (IPA, revised to 2015)	Meaning
1	tá	**t** a	dozen
2	tác	**t** a **k**	the deer’s cry
3	tát	**t** a **t**	slap
4	tám	**t** a **m**	eight
5	tán	**t** a **n**	to crush, to grind, to court, to chat
6	táp	**t** a **p**	to snap, to lap
7	táng	**t** a **ŋ**	throw a punch (familial)
8	cá	**k** a	fish
9	cát	**k** a **t**	sand
10	các	**k** a **k**	every, all (adv.)
11	cán	**k** a **n**	handle
12	cáp	**k** a **p**	cable
13	cám	**k** a **m**	bran
14	cáng	**k** a **ŋ**	stretcher
15	má	**m** a	mother, cheek
16	mát	**m** a **t**	fresh

17	mác	**m** a **k**	scimitar
18	mán	**m** a **n**	people of San Chay (in the Northeast of Vietnam)
19	máng	**m** a **ŋ**	gutter
20	ná	**n** a	crossbow
21	nát	**n** a **t**	crushed
22	nám	**n** a **m**	burnt
23	ngáng	**ŋ** a **ŋ**	to bar
24	ngáp	**ŋ** a **p**	to yawn
25	ngác	**ŋ** a **k**	disoriented
26	ngát	**ŋ** a **t**	very [sweet] (perfume)
27	ngán	**ŋ** a **n**	to be depressed, be tired of
28	pháp lý	f a **p** . l i	legal
29	áp bức	ʔ a **p** . b ɯ k	to oppress
30	bát ngát	b a **t** . ŋ a t	vast, immense, limitless
31	khát nước	x a **t** . n ɯɤ k	be thirsty
32	phát giác	f a **t** . z a k	to reveal, to discover, to find out
33	hát ví	h a **t** . v i	tune of popular song
34	lác đác	l a **k** . d a k	scattered
35	gác bút	ɣ a **k** . b u t	abandon the job of writer
36	xác đáng	s a **k** . d a ŋ	pertinent
37	đám cưới	d a **m** . k ɯɤ j	marriage
38	khám phá	x a **m** . f a	to discover
39	bán kết	b a **n** . k e t	semifinal
40	gián tiếp	z a **n** . t ie p	indirect
41	đáng kính	d a **ŋ** . k i ŋ	venerable
42	đáng tiếc	d a **ŋ** . t ie k	regrettable
43	sáng kiến	s a **ŋ** . k ie n	initiative
44	sáng tác	s a **ŋ** . t a k	to compose
45	tháng tám	tʰ a **ŋ** . t a m	August
46	áp dụng	ʔ a **p** . z u ŋ	to apply
47	áp lực	ʔ a **p** . l ɯ k	pressure
48	pháp luật	f a **p** . l w ɤ̆ t	law
49	pháp lệnh	f a **p** . l e ŋ	order
50	khát vọng	x a **t** . v ɔ ŋ	aspiration
51	phát động	f a **t** . d o ŋ	to mobilize
52	mát dạ	m a **t** . z a	satisfied
53	ác liệt	ʔ a **k** . l ie t	very fierce, very violent
54	tác dụng	t a **k** . z u ŋ	utility, usefulness
55	xám xịt	s a **m** . s i t	leaden (color)
56	bám trụ	b a **m** . tɕ u	hold on to
57	thán phục	tʰ a **n** . f u k	to admire
58	cán sự	k a **n** . s ɯ	junior staff member
59	sáng tạo	s a **ŋ** . t a w	to invent
60	kháng cự	x a **ŋ** . k ɯ	to resist
61	tháng chạp	tʰ a **ŋ** . tɕ a p	December
62	đáng sợ	d a **ŋ** . s ɤ	frightening

Appendix 2

Means and standard deviations (in parentheses) of normalized duration (ms) of each Vietnamese consonant according to their within-word position: word-initial (C-), word-final (-C$_{\#}$) and coda of the first syllable of a disyllabic word (-C$_{\sigma}$). /p/ is illicit in syllable-initial position.

Consonants	C-	-C$_{\#}$	-C$_{\sigma}$
p		40.5 (15)	23.6 (7)
t	44.6 (12)	45.2 (13)	23.8 (6)
k	46.1 (14)	46.9 (14)	23.2 (6)
m	52.9 (14)	45.7 (12)	28.0 (8)
n	51.3 (13)	47.4 (11)	25.8 (6)
ŋ	45.9 (13)	42.6 (10)	30.0 (6)

Appendix 3

Mean values and standard deviations (SD) of ΔF_1, ΔF_2, and ΔF_3 (Hz) as a function of 2 factors: coda consonant /p/, /t/, /k/ and tone of the syllable following the target consonant (A1, D1, D2).

Percent	Δ		p_A1	t_A1	k_A1	p_D1	t_D1	k_D1	p_D2	t_D2	k_D2
50%	F_1	Mean	10.30	4.24	25.02	–5.07	–3.10	14.06	8.89	–1.19	13.49
		SD	18.07	11.01	32.22	38.70	8.42	22.07	9.24	35.16	29.32
	F_2	Mean	3.87	25.44	6.23	16.50	21.18	–14.71	10.34	13.70	–4.81
		SD	30.63	29.25	18.31	14.35	21.69	34.87	17.44	15.12	15.74
	F_3	Mean	7.26	19.76	–32.93	–28.33	10.27	–27.05	13.03	4.40	11.88
		SD	108.14	62.82	89.24	30.77	54.59	38.84	16.58	32.92	41.32
60%	F_1	Mean	27.06	16.85	26.52	–16.1	–15.5	21.15	–8.29	–10.22	20.44
		SD	39.34	36.54	47.96	36.50	22.83	24.28	26.79	34.69	44.13
	F_2	Mean	–7.43	3.32	–4.18	20.7	28.56	–20.7	6.35	34.45	–24.09
		SD	38.33	46.52	37.42	21.26	16.05	26.13	23.78	18.93	39.85
	F_3	Mean	8	3.31	–22.8	22.27	15.62	–35.4	22.1	4.57	8.23
		SD	114.74	77.92	103.49	78.22	116.79	74.13	23.78	101.59	52.32
70%	F_1	Mean	15.77	30.53	36.92	–37.3	–34.3	–6.5	–15.5	–19.21	–7.83
		SD	51.71	29.72	59.04	38.30	39.09	42.53	37.74	40.89	45.89
	F_2	Mean	–31.22	18.41	6.18	–4.15	50.53	–48	–2.41	55.94	–30.63
		SD	81.47	37.47	22.38	24.16	23.33	29.48	25.00	25.22	38.84
	F_3	Mean	6.23	15.78	–30.3	19.59	69.55	–53.8	16	37.44	51.99
		SD	122.14	100.08	41.49	81.46	78.59	71.52	42.28	94.66	80.74
80%	F_1	Mean	–9.78	4.74	9.42	–79.8	–67.9	–46.3	–47.1	–47.26	–22
		SD	55.14	53.62	42.16	30.86	54.11	36.00	71.87	61.39	52.43
	F_2	Mean	–35.85	18.4	–22.4	–16.1	58.69	–81.7	–21	63.91	–34.21
		SD	37.29	38.74	43.13	23.69	18.34	65.40	36.23	40.31	59.68
	F_3	Mean	–35.75	6.62	3.38	11.9	63.01	–11.9	–0.17	27.73	64.56
		SD	133.57	130.33	79.29	85.48	152.71	100.87	60.18	129.88	95.55
90%	F_1	Mean	–46.32	–58.16	–57.7	–120	–112.9	–76.4	–104	–97.57	–85.48
		SD	55.12	125.73	60.59	44.80	59.70	70.34	73.30	36.43	61.84
	F_2	Mean	–82.35	17.41	–43.2	–67.5	49.97	–97.2	–62.9	85.49	–49.9
		SD	72.01	52.20	73.65	40.05	30.85	81.01	57.91	64.74	65.56
	F_3	Mean	3.5	54.61	23.19	5.5	72.93	20.3	5.38	46.05	88.95
		SD	111.67	124.82	77.34	92.46	169.09	125.35	65.44	144.31	114.73

Acknowledgments

The authors warmly wish to express their gratitude to Marc Brunelle and to anonymous reviewers for their thorough and valuable comments on our manuscript. Special thanks are due to Thomas Sawallis for his careful proofreading of the manuscript and for his meticulous comments on this work. We would like to thank Julien Meyer and Christophe Savariaux for reading earlier versions of this paper and for providing helpful comments. Many thanks go to Alexis Michaud for stimulating discussions about Vietnamese final consonants and audiences at the Workshop on Vietnamese Prosody in Cologne 2015. We are also thankful to Nguyễn Chu Đức for helping out with various Matlab and Latex scripts. Last but not least, many thanks go to the Vietnamese students in Grenoble who kindly participated in different experiments.

This work was supported by l'Agence Universitaire de la Francophonie (PC-411/2460) and by the French National Research Agency, grant No. APPSy-0449560.

Disclosure Statement

All authors have no conflict of interest to report.

References

Abramson AS, Tingsabadh K (1999): Thai final stops: cross-language perception. Phonetica 56:111–122.

Adank P, Van Hout R, Smits R (2004): An acoustic description of the vowels of Northern and Southern Standard Dutch. J Acoust Soc Am 116:1729–1738.

Browman CP, Goldstein L (1988): Some notes on syllable structure in articulatory phonology. Phonetica 45:140–155.

Browman CP, Goldstein L (1990): Gestural specification using dynamically defined articulatory structures. J Phon 18:299–320.

Browman CP, Goldstein L (1995): Gestural syllable position effects in American English; in Bell-Berti F, Raphael LJ (eds): Producing Speech: Contemporary Issues: For Katherine Safford Harris. Woodbury, AIP Press, pp 19–33.

Brunelle M (2003): Tone coarticulation in Northern Vietnamese. Proc 15th Int Congr Phonet Sci, Barcelona, pp 2673–2676.

Brunelle M (2009): Northern and Southern Vietnamese tone coarticulation: a comparative case study. J Southeast Asian Ling 1:49–62.

Brunelle M, Nguyễn DD, Nguyễn KH (2010): A laryngographic and laryngoscopic study of northern Vietnamese tones. Phonetica 67:147–169.

Byrd D (1995): C-centers revisited. Phonetica 52:285–306.

Byrd D (1996): Influences on articulatory timing in consonant sequences. J Phon 24:209–244.

Cao XH (1985): Phonologie et linéarité. Réflexions critiques sur les postulats de la phonologie contemporaine. Paris, Société d'Etudes Linguistiques et Anthropologiques de France, vol 18.

Chitoran I, Goldstein L, Byrd D (2002): Gestural overlap and recoverability: articulatory evidence from Georgian. Lab Phonol 7:419–447.

Cho T (2001): Effects of morpheme boundaries on intergestural timing: evidence from Korean. Phonetica 58:129–162.

Cho T, Ladefoged P (1999): Variation and universals in VOT: evidence from 18 languages. J Phon 27:207–229.

Cutler A, Butterfield S (1991): Word boundary cues in clear speech: a supplementary report. Speech Commun 9:485–495.

Cutler A, Norris D (1988): The role of strong syllables in segmentation for lexical access. J Exp Psychol Hum Percept Perform 14:113–121.

Davidson L (2003): The Atoms of Phonological Representation: Gestures, Coordination and Perceptual Features in Consonant Cluster Phonotactics; unpubl doctoral dissertation, Johns Hopkins University.

Davidson L (2007): Coarticulation in contrastive Russian stop sequences. Proc 16th Int Congr Phonet Sci, Saarbrücken, pp 417–420.

Delattre P (1963): Le jeu des transitions de formants et la perception des consonnes. Proc 4th Int Congr Phonet Sci, La Haye, Mouton, pp 407–418.

Delattre P, Liberman AM, Cooper FS (1955): Acoustic loci and transitional cues for consonants. J Acoust Soc Am 27:769–774.

Diệp QB (2004): Ngữ pháp tiếng Việt (Vietnamese grammar). Hanoi, Nhà xuất bản Giáo dục (Education Press).

Dorman MF, Studdert-Kennedy M, Raphael LJ (1977): Stop-consonant recognition: release bursts and formant transitions as functionally equivalent, context-dependent cues. Percept Psychophys 22:109–122.

Đoàn TT (1999): Ngữ âm tiếng Việt (Vietnamese Phonetics). Hanoi, Hanoi National University Publishing.

Đỗ TD, Lê TT (1994): Le vietnamien sans peine (Méthode d'apprentissage de la langue vietnamienne). Chennevières-sur-Marne, Méthode Assimil.

Fischer-Jorgensen E (1972): p t k et b d g français en position intervocalique accentuée; in Valdman P (ed): Papers in Linguistics and Phonetics to the Memory of Pierre Delattre. Den Haag, Mouton, pp 143–200.

Fougeron C, Keating PA (1997): Articulatory strengthening at edges of prosodic domains. J Acoust Soc Am 101:3728–3740.

Gao M, Mooshammer C, Hagedorn C, Nam H, Tiede M, Chang YC, Hsieh FF, Goldstein L (2011): Intra- and inter-syllabic coordination: an articulatory study of Taiwanese and English. Proc 17th Int Congr Phonet Sci, Hong Kong, pp 723–726.

Han MS, Kim K (1974): Phonetic variation of Vietnamese tones in disyllabic utterances. J Phon 2:223–232.

Hoàng T, Hoàng M (1975): Remarques sur la structure phonologique du vietnamien. Hanoi, Essais Linguistiques, Etudes Vietnamiennes, vol 40.

Keating PA (1983): Comments on the jaw and syllable structure. J Phon 11:401–406.

Keating P, Wright R, Zhang J (1999): Word-level asymmetries in consonant articulation. UCLA Work Pap Phon 97:157–173.

Kingston J (2008): Lenition. Proc 3rd Conf Lab Approaches Spanish Phonol. Somerville, Cascadilla Press, pp 1–31.

Kirby J (2010): Dialect experience in Vietnamese tone perception. J Acoust Soc Am 127:3749–3757.

Kirby J (2011): Vietnamese (Hanoi Vietnamese). J Int Phonet Assoc 41:381–392.

Kochetov A (2013): Production, Perception, and Phonotactic Patterns: A Case of Contrastive Palatalization. Abingdon, Routledge.

Krakow RA (1999): Physiological organization of syllables: a review. J Phon 27:23–54.

Lê HP, Nguyễn TMH, Roussanaly A (2009): Finite-State Description of Vietnamese Reduplication. Proc 7th Workshop Asian Lang Resour. Singapore, Suntec, pp 63–69.

Lê TX (2011): Eléments archaïques et éléments nouveaux dans le lexique vietnamien contemporain. 4th Congr Asia Pacific Netw, Paris. http://www.reseauasie.com/userfiles/file/I06_xuyen_lexique_vietnamien.pdf (accessed April 17, 2014).

Lê VB (2006): Reconnaissance automatique de la parole pour des langues peu dotées; thèse de doctorat en informatique, Université Joseph-Fourier, Grenoble.

Lê VB, Bigi B, Besacier L, Castelli E (2003): Using the web for fast language model construction in minority languages. Proc 8th Eur Conf Speech Commun Technol, Geneva, pp 3117–3120.

Lê VT (2004): Vị trí tiếng Nùng Dín trong quan hệ với các phương ngữ Nùng và Tày ở Việt Nam (Nung Din Language Position in Relation to the Nung and Tay Dialect in Vietnam); PhD dissertation, Hanoi Institute of Linguistics. http://ngonngu.net/index.php?p=264 (Swadesh list of words in Vietnamese) (accessed December 10, 2010).

Liberman AM, Delattre PC, Cooper FS, Gerstman LJ (1954): The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychol Monogr Gen Appl 68:1–13.

Lindblom B (1983): Economy of speech gestures; in MacNeilage PF (ed): The Production of Speech. New York, Springer, pp 217–245.

Lisker L (1957): Minimal cues for separating /w, r, l, y/ in intervocalic position. Word 13:256–267.

Lisker L, Abramson AS (1964): A cross-language study of voicing in initial stops: acoustical measurements. Word 20:384–422.

Maddieson I, Precoda K (1992): Syllable structure and phonetic models. Phonology 9:45–60.

Maeda S (1982): The role of the sinus cavities in the production of nasal vowels. Acoustics, Speech, and Signal Processing. IEEE ICASSP '82, Paris, vol 7, pp 911–914.

Mai NC, Vũ ĐN, Hoàng TP (1997): Cơ sở ngôn ngữ học và tiếng Việt (Base Linguistics and Vietnamese Language). Hanoi, Nhà xuất bản Đại hoc và Giáo dục chuyên nghiệp (University and Professional Education Press).

Malécot A (1968): The force of articulation of American stops and fricatives as a function of position. Phonetica 18:95–102.

Markel JD, Gray AH Jr (1976): Linear Prediction of Speech. Communication and Cybernetics. Berlin, Springer, vol 12.

Ménard L (2002): Production et perception des voyelles au cours de la croissance du conduit vocal: variabilité, invariance et normalization; thèse de doctorat en sciences du langage, Université Stendhal, Grenoble.

Ménard L, Chrétien J, Lachapelle R, Marleau I (2010): Corrélats acoustiques de la perception des voyelles produites par des locuteurs sourds. Spectrum 2:19–31.

Michaud A (2004): Final consonants and glottalization: new perspectives from Hanoi Vietnamese. Phonetica 61:119–146.

Michaud A (2009): Monosyllabicization: patterns of evolution in Asian languages; in Stolz T, Nau N, Stroh C (eds): Monosyllables: From Phonology to Typology. Bremen, University of Bremen, pp 115–130.

Michaud A, Vũ NT, Amelot A, Roubeau B (2006): Nasal release, nasal finals and tonal contrasts in Hanoi Vietnamese: an aerodynamic experiment. Mon-Khmer Studies 36:121–137.

Nearey TM, Assmann PF, Hillenbrand JM (2002): Evaluation of a strategy for automatic formant tracking. J Acoust Soc Am 112:2323–2323.

Ngô TN (1984): The Syllabeme and Patterns of Word Formation in Vietnamese; PhD dissertation, New York University.

Nguyễn PP (1989): Le vietnamien, un cas de romanisation inachevée. Cahiers d'études vietnamiennes. Paris, Université de Paris-7, vol 10, pp 25–32.

Nguyễn TG (1998): Từ vựng học tiếng Việt (Vietnamese Lexicology). Hanoi, Nhà xuất bản Giáo dục (Education Press).

Nguyễn VL, Edmondson J (1997): Tones and voice quality in modern northern Vietnamese: instrumental case studies. Mon-Khmer Studies 28:1–18.

Noyer R (1998): Vietnamese "morphology" and the definition of word. Univ Pa Work Pap Linguist 5:65–89.

Phạm AH (2003): Vietnamese Tones: A New Analysis. New York, Routledge.

Pulgram E (1965): Consonant cluster, consonant sequence, and the syllable. Phonetica 13:76–81.

Redford M (1999): An Articulatory Basis for the Syllable; PhD dissertation, University of Texas, Austin.

Redford M, Diehl R (1999): The relative perceptual distinctiveness of initial and final consonants in CVC syllables. J Acoust Soc Am 106:1555–1565.

Rossato S (2000): Du son au geste, inversion de la parole: le cas des voyelles nasales; thèse de doctorat en signal image parole télécoms, Institut National Polytechnique de Grenoble.

Rousset I (2004): Structures syllabiques et lexicales des langues du monde: données, typologies, tendances universelles et contraintes substantielles; thèse de doctorat en sciences du langage, Université Stendhal, Grenoble.

Schiering R, Bickel B, Hildebrandt KA (2010): The prosodic word is not universal, but emergent. J Linguist 46:657–709.

Serniclaes W (1987): Étude expérimentale de la perception du trait de voisement des occlusives du français; thèse de doctorat en sciences psychologiques et pédagogiques, Institut de Phonétique, Université Libre de Bruxelles.

Sharf DJ, Hemeyer T (1972): Identification of place of consonant articulation from vowel formant transitions. J Acoust Soc Am 51:652–658.

Spanias AS (1994): Speech coding: a tutorial review. Proc IEEE 82:1541–1582.

Talkin D (1995): A robust algorithm for pitch tracking (RAPT); in Kleijn WB, Paliwal KK (eds): Speech Coding and Synthesis. New York, Elsevier Science, vol 495, p 518.

Terrance MN, Bernard LR (1994): Effects of place of articulation and vowel context on VOT production and perception for French and English stops. J Int Phon Assoc 24:1–18.

Thompson LC (1965): A Vietnamese Grammar. Seattle, University of Washington Press.

Trần ĐĐ, Castelli E, Serignat JF, Lê XH, Trịnh VL (2005): Influence of F_0 on Vietnamese syllable perception. Interspeech, Lisbon, pp 1697–1700.

Trần TTH (2011): Processus d'acquisition des clusters et autres séquences de consonnes en langue seconde: de l'analyse acoustico-perceptive des séquences consonantiques du vietnamien à l'analyse de la perception et production des clusters du français par des apprenants vietnamiens du FLE; thèse de doctorat en sciences du langage, Université Stendhal, Grenoble.

Trần TTH (2015): A word prosodic structure study in Vietnamese read and spontaneous speech. Workshop on Vietnamese Prosody, Cologne.

Trần TTH, Vallée N (2009): An acoustic study of interword consonant sequences in Vietnamese. J Southeast Asian Linguist 1:231–249.

Trần TTH, Vallée N (2010): Corrélats acoustico-perceptifs des consonnes non relâchées du vietnamien. Actes des 28e Journées d'Etudes sur la Parole, Université de Mons, pp 377–380.

Trương VC (1970): Structure de la langue vietnamienne. Paris, Librairie Orientaliste Paul Geuthner.

Vaissière J (2007): Identifying the feature [+nasal] in syllable-initial and syllable-final nasal consonants. Conférence invitée d'ACI-Prosodie: Where the Features Come from, Paris.

Vaissière J (2008): Aspects aérodynamiques et acoustiques de la nasalité. Cours invité de l'École Thématique CNRS: Dynamique de la nasalité, Porquerolles.

Vallée N, Rossato S, Rousset I (2009): Favoured syllabic patterns in the world's languages and sensorimotor constraints; in Pellegrino F, Marsico E, Chitoran I, Coupé C (eds): Approaches to Phonological Complexity. Berlin, Mouton de Gruyter, pp 111–139.

Walley AC, Carrell TD (1983): Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. J Acoust Soc Am 73:1011–1022.

Wright R (2004): A review of perceptual cues and cue robustness; in Hayes B, Kirchner R, Steriade D (eds): Phonetically Based Phonology. Cambridge, Cambridge University Press, pp 34–57.

Yanagawa M (2006): Articulatory Timing in First and Second Language: A Cross-Linguistic Study; dissertation, Yale University.

Yeni-Komshian GH, Caramazza A, Preston MS (1977): A study of voicing in Lebanese Arabic. J Phon 5:35–48.

Zsiga EC (1994): Acoustic evidence for gestural overlap in consonant sequences. J Phon 22:121–140.

Zsiga EC (2000): Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. J Phon 28:69–102.