1-1-2012

The Vietnamese Vowel System

Giang Huong Emerich
University of Pennsylvania, gianghnguyen.upenn@gmail.com

The Vietnamese Vowel System

Abstract
In this dissertation, I provide a new analysis of the Vietnamese vowel system as a system with fourteen monophthongs and nineteen diphthongs based on phonetic and phonological data. I propose that these Vietnamese contour vowels - /ie/, /ɯɤ/ and /uo/ - should be grouped with these eleven monophthongs /i e ɛ a ɐ ʌ ɤ ɯ u o ɔ/ based on their similarities in phonetic and phonological behaviors. The phonetic characteristics of these vowels are studied acoustically using normalized and scaled acoustic values of 13,925 tokens, spoken by female Hanoian speakers from my speech corpus, "The Vietnamese Speech Corpus". Phonetic analysis shows that the eleven monophthongs and three contour vowels are similar in terms of formant frequency targets, formant dynamic trajectories, and duration. Phonologically, monophthongs and contour vowel can be rhymed with each other in poems, and the two elements within each contour vowel should be analyzed as two halves of one root node in the syllable structure. In chapters 1 and 2, I give the current analysis of the Vietnamese sound system, review different approaches to the acoustic features of vowels, and the phonemic status of diphthongs. In chapter 3, I give a detailed description of the Vietnamese Speech Corpus. In chapter 4, I show the difference in formant targets between monophthongs and glides, as well as the importance of duration in distinguishing vowels in Vietnamese. I also give evidence for the differences in duration between the diphthongs and the monophthongs-and-contour-vowels group. In chapter 5, I analyze the natural class of monophthongs and contour vowels in terms of feature geometry and give evidence from Vietnamese phonological processes to support the analysis of contour vowels as being in the same natural class as monophthongs. I also re-analyze Vietnamese triphthongs as diphthongs in this chapter. Finally, in chapter 6, I summarize the similarities and differences across the monophthongs, contour vowels, and diphthongs, and suggest possible future studies to test this hypothesis of the Vietnamese vowel system.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Linguistics

First Advisor
Jiahong Yuan

Keywords
diphthong, dynamic trajectory, Hanoi dialect, triphthong, Vietnamese, vowel

Subject Categories
Anthropological Linguistics and Sociolinguistics | Asian Studies | Linguistics

THE VIETNAMESE VOWEL SYSTEM

Giang Huong Emerich

A DISSERTATION

in

Linguistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2012

Supervisor of Dissertation	Graduate Group Chairperson
______________________	______________________
Jiahong Yuan	Robin Clark
Assistant Professor of Linguistics	Associate Professor of Linguistics
	Acting Graduate Chair of Linguistics

Dissertation Committee

Jiahong Yuan	Assistant Professor of Linguistics
Eugene Buckley	Associate Professor of Linguistics
William Labov	Fassit Professor of Linguistics

Acknowledgements

I would like to thank my advisor, Jiahong Yuan, for giving me his utmost support and advice throughout the process of researching and writing this dissertation as well as during my years here at Penn. He has been very patient, inspiring and helpful in more ways that I can describe. I would like to thank my committee members, Eugene Buckley and William Labov, whose advice and input on my dissertation topics and writing I truly appreciate.

I would like to express my special gratitude to my lab mates, Joshua Tauberer and Catherine Lai, with whom I have shared so much for the last seven years. In addition to being my great friends and providing input and advice on my research, Joshua Tauberer has also written the Praat-Py and Python scripts for me to process my speech data in its initial state, and Catherine Lai has helped me review my writings numerous times. And for that, I would like to thank them.

Numerous people have played very important roles in my life here at Penn and I would like to send my warmest appreciation to them. Among those are Laurel MacKenzie, Aviad Eilam, Toni Cook, Keelan Evanini and Josef Fruehwald, who have shared with me responsibilities in organizing events, the fun in sociolinguistic research trips, and their brilliant advice. Another important person is Amy Forsyth, our Administrative Coordinator, who has had to put up with so many of my crazy ideas and schedules, and somehow she still managed to be so helpful. Tony Kroch was one of my very first professors here at Penn, who is not only brilliant, but was also very supportive and caring when I was going through some stressful parts of life. There is Gillian Sankoff, whose classes and smiles I absolutely love, and who has set a great example in balancing one's intellectual life with commitments to one's family, friends and community.

Outside of Penn, I would like to send my special thanks to Marc Brunelle for spending part of his vacation time to give me input on my research. I would like to send my thanks to my Bachelor Thesis advisor, Patricia Donegan, who guided me throughout my studies at the University of Hawaii at Manoa, and who recommended that I attend Penn.

I would like to thank all my friends, especially Martina Nguyen, Hanh Bui, Anh Tran, and Minkyoung Song, for they have been there for me every step of the way, through all the all-night research, through all the failings and achievements, through all the stress and the excitement.

Last but not least, I would like to dedicate this dissertation with my special thanks and gratitude to my parents, Nguyễn Ngọc Khôi and Trần Liên Hương, who have made me who I am today, to my little sister, Nguyễn Ngân Giang, who always makes me smile and proud, and to my husband, Chad Calvin Emerich, whose love I treasure.

ABSTRACT

THE VIETNAMESE VOWEL SYSTEM

Giang Huong Emerich

Jiahong Yuan

In this dissertation, I provide a new analysis of the Vietnamese vowel system as a system with fourteen monophthongs and nineteen diphthongs based on phonetic and phonological data. I propose that these Vietnamese contour vowels – /ie/, /ɯɤ/ and /uo/ – should be grouped with these eleven monophthongs /i e ɛ a ɐ ʌ ɤ ɯ u o ɔ/ based on their similarities in phonetic and phonological behaviors. The phonetic characteristics of these vowels are studied acoustically using normalized and scaled acoustic values of 13,925 tokens, spoken by female Hanoian speakers from my speech corpus, "The Vietnamese Speech Corpus". Phonetic analysis shows that the eleven monophthongs and three contour vowels are similar in terms of formant frequency targets, formant dynamic trajectories, and duration. Phonologically, monophthongs and contour vowel can be rhymed with each other in poems, and the two elements within each contour vowel should be analyzed as two halves of one root node in the syllable structure. In chapters

1 and 2, I give the current analysis of the Vietnamese sound system, review different approaches to the acoustic features of vowels, and the phonemic status of diphthongs. In chapter 3, I give a detailed description of the Vietnamese Speech Corpus. In chapter 4, I show the difference in formant targets between monophthongs and glides, as well as the importance of duration in distinguishing vowels in Vietnamese. I also give evidence for the differences in duration between the diphthongs and the monophthongs-and-contour-vowels group. In chapter 5, I analyze the natural class of monophthongs and contour vowels in terms of feature geometry and give evidence from Vietnamese phonological processes to support the analysis of contour vowels as being in the same natural class as monophthongs. I also re-analyze Vietnamese triphthongs as diphthongs in this chapter. Finally, in chapter 6, I summarize the similarities and differences across the monophthongs, contour vowels, and diphthongs, and suggest possible future studies to test this hypothesis of the Vietnamese vowel system.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Vietnamese has been described as a language with a 9-vowel system (Nguyễn Bạt Tụy 1949, 1959; Haudricourt 1952, Đoàn Thiện Thuật 1977, Kirby 2011), a 10-vowel system (Smalley 1957, Le-Van-Ly 1960, Crothers 1978), or with an 11-vowel system (Thompson 1965, Nguyễn Đình Hòa 1966, Han 1968). The 9-vowel system analysis lists these vowels /i e ɛ a ɔ o u ɤ ɯ/ as phonemes and the vowels /ɐ ʌ/ as allophones of phonemes /a ɤ/. However, this analysis, as well as the 10-vowel system analysis, does not sufficiently explain the minimal pairs of the vowel pair /a ɐ/ which are followed by a nasal or by a glide, as shown in the table 1.1 (Thompson 1965).

No.	Orthography	IPA	Gloss	Orthography	IPA	Gloss
1	*tan*	/tan ˧/	melt	*tăn*	/tɐn ˧//	bubble
2	*cán*	/kan ˧˥/	roll on	*cắn*	/kɐn ˧˥/	bite
3	*tai*	/taj ˧/	ear	*tay*	/tɐj ˧//	hand
4	*sai*	/saj ˧/	wrong	*say*	/sɐj ˧//	drunk

Table 1.1: The minimal pairs of the vowels /a ɐ/

All these previous analyses, regardless of the number of vowel phonemes they agree upon, describe Vietnamese diphthongs as vowel-plus-glide sequences, and

Vietnamese triphthongs as glide-vowel-glide sequences, in which the glides could be /j/, /w/, and /ɰ/. However, these three diphthongs – /ie/, /ɯɤ/ and /uo/ – are different from the rest of the diphthongs, and they have allophones /iə/, /ɯə/, and /uə/ in syllable-final positions, and /ie/, /ɯɤ/ and /uo/ in non-final positions. Nguyễn Đình Hòa (1997) described these three diphthongs as "double vowels" (Nguyễn Đình Hòa 1997, p. 22). When these three diphthongs /ie/, /ɯɤ/ and /uo/ precede the /w/ and /j/ glides, they are actually sequences of vowel-vowel-glide instead of the usual glide-vowel-glide sequences of all other triphthongs.

No.	Orthography	IPA	Gloss
1	*tia*	/tiə ˧/	ray
2	*đưa*	/ɗɯə ˧/	give
3	*đua*	/ɗuə ˧/	race
4	*tiên*	/tien ˧/	angel
5	*chương*	/cɯɤŋ ˧/	chapter
6	*tuôn*	/tuon ˧/	flow
7	*điêu*	/ɗiew ˧/	lie
8	*đuôi*	/ɗuoj ˧/	tail

Table 1.2: Examples of diphthongs and triphthongs

My hypothesis is that the three diphthongs /ie/, /ɯɤ/ and /uo/ are actually not diphthongs and they should be grouped with the monophthongs to make Vietnamese a 14-vowel system with all the diphthongs and triphthongs being analyzed as having only two elements. This hypothesis is proven by the acoustic features of these monophthongs and diphthongs as well as their phonological behaviors. In chapter 2, I will review the different approaches to vowel description using vowel formant values, vowel temporal and dynamic trajectory, as well as the difference between vowels and glides. These studies show that beside formants, duration is an important feature and using a combination of features is effective in vowel classification. The studies also discuss the phonemic status of diphthongs. I also give the description of the Vietnamese sound system including consonants, vowels and tones. In chapter 3, I give a detailed description of the Vietnamese Speech Corpus and its structure, as well as the speech collection and recording techniques employed in building this corpus. I also provide the annotation rules, acoustic measurement techniques as well as normalization formulas used in this research. Even though I only use the speech of six female Hanoian speakers for this study, as the annotation for the entire corpus is not yet complete by this point of time, the detailed description of the entire corpus would give a better understanding of the data set. Chapter 4 is the phonetic study of the Vietnamese vowels, providing evidence for the importance of duration in distinguishing vowels in Vietnamese and presenting the results on formant, duration for the monophthongs, contour vowels, and diphthongs. These values and the perception test results support the classification of monophthongs and contour vowels as one natural class, and all other Vietnamese diphthongs as another natural class. In chapter 5, I analyze the natural class of monophthongs and contour vowels in terms of feature geometry and give evidence in Vietnamese phonological processes to support the analysis of contour vowels in the same natural class as monophthongs. I also give a re-analysis of Vietnamese triphthongs as diphthongs in this chapter. Finally, in chapter 6, I summarize the similarities and differences of all the monophthongs, contour vowels, and diphthongs, and discuss possible further studies of the Vietnamese sound system.

I will mark the tones in Vietnamese using IPA tone level symbols, e.g. / ˧/ for level tone, in combination with the glottal stop, placed right next to the IPA transcription of the word within the bracket, e.g. *tan* /tan˧/ "melt". The details of the IPA transcription of the Vietnamese tones are listed in table 2.2 in chapter 2. For the ease of discussion of the Vietnamese monophthongs, contour vowels and diphthongs, I name them informally using capitalized corresponding Vietnamese words, which are listed in the tables 1.3, 1.4 and 1.5.

No.	Monophthong	Name	IPA	Gloss
1	a	*GÀ*	/ɣa ˧˩/	chicken
2	ɐ	*RẮN*	/rɐn ˧˥/	snake
3	ɤ	*MƠ*	/mɤ ˧/	apricot
4	ʌ	*MẬN*	/mʌn ˧ʔ˩/	plum
5	i	*CHIM*	/cim ˧/	bird
6	e	*SÊN*	/sen ˧/	snail
7	ɛ	*TÉP*	/tɛp ˧˥/	shrimp
8	ɯ	*ƯNG*	/ɯŋ ˧/	hawk
9	u	*CÚT*	/kut ˧˥/	quail
10	o	*CÔNG*	/koŋ ˧/	peacock
11	ɔ	*CHÓ*	/cɔ ˧˥/	dog

Table 1.3: Names for the Vietnamese monophthongs

No.	Contour vowels	Name	IPA	Gloss
1	ie	KIẾN	/kien ˧˥/	ant
2	ɯɤ	LƯƠN	/lɯɤn ˧/	eel
3	uo	CHUỘT	/cuot ˧ʔ˩/	mouse

Table 1.4: Names for the Vietnamese contour vowels

No.	Diphthong	Name	IPA	Gloss
1	/aj/	NAI	/naj ˧/	deer
2	/ɐj/	TAY	/tɐj ˧/	hand
3	/ʌj/	CÂY	/kʌj ˧/	tree
4	/ɔj/	VOI	/vɔj ˧/	elephant
5	/oj/	MÔI	/moj ˧/	lips
6	/ɤj/	GIỜI	/zɤj ˧˩/	centipede
7	/uj/	MŨI	/muj ˧ʔ˥/	nose
8	/ɯj/	NGỬI	/ŋɯj ˧˩˧/	smell (v.)
9	/aw/	CÁO	/kaw ˧˥/	fox
10	/ɐw/	CAU	/kɐw ˧/	betel fruit
11	/ʌw/	GẤU	/ɣʌw ˧˥/	bear
12	/ɛw/	HEO	/hɛw ˧/	pig
13	/ew/	SẾU	/sew ˧˥/	crane
14	/iw/	NÍU	/niw ˧˥/	pull (v.)
15	/ɯw/	CỪU	/kɯw ˧˩/	sheep

Table 1.5: Names for the Vietnamese diphthongs

Chapter 2

Previous studies and the Vietnamese sound system

2.1. Articulatory qualities of vowels and consonants

Approximants and vowels are divided based on the positions of their articulators and the voice source. Vowels are produced with the articulators being apart and the airstream passage relatively unobstructed (Ladefoged 2001). Approximants are consonants, which are characterized by an obstructed vocal tract, but they are produced with the vocal tract being narrow enough to create turbulent airflow but not completely obstructed. Some approximants, for example /w/ and /j/, are considered glides because they are approximants with vowel features, being more resonant than consonants and produced with similar articulatory positions. These glides, however, are also similar to consonants because their vocal tracts are more constricted than that of /u/ and /i/; for example, /j/ has a tongue body position that is between a half-close to close vowel. In addition, glides behave like consonants because they can be devoiced after other voiceless consonants, they can fill a syllable onset position when it is required, and they have less energy than the vowels /u/ and /i/. And while the vowels /i ɯ u y/ are the nuclei of the syllable, their glide counterparts /j ɰ w ɥ/ cannot fill the role of the nucleus.

As for articulation, vowels are voiced and sonorant and they can be described by the height of the tongue body, the frontness of the tongue, and the rounding of the lips. Due to the limited space in the vocal tract and movements of our oral articulators, the vowel space, which is drawn out in figure 2.1, is defined by the most extreme positions of the tongue. The highest position of the tongue body is associated with the highest vowel. The lowest position of the tongue body is associated with the lowest vowel. The most front and the most back positions of the tongue are associated with the most front and the most back vowels, respectively. These two pairs of features – high/low and front/back features – are the most basic parameters to distinguish vowels across languages, and these features have multi values instead of binary values to show the various levels of the tongue (Lindau 1978).

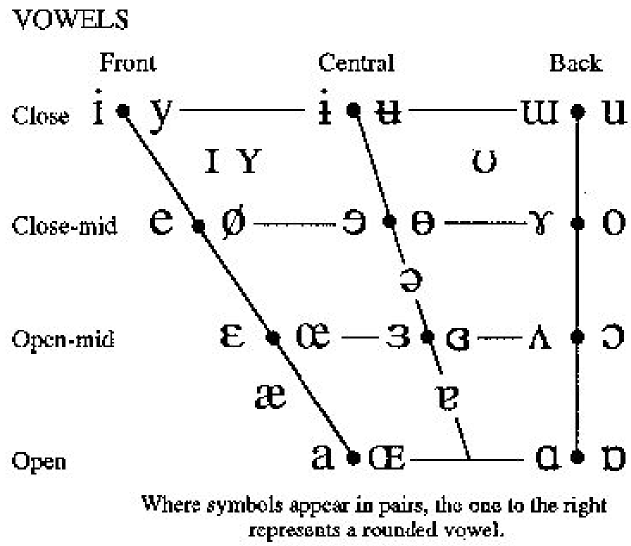


Figure 2.1: IPA vowel chart

2.2. Formant values of vowels and glides

2.2.1. Targets of monophthongs

Acoustically, vowels can be described and distinguished in terms of the frequency position of their resonances (Potter & Peterson 1948). These resonances, which are called formants, are measured in hertz and can be used to map the vowels onto the vowels space, with the x-axis representing the second formant (F2) and the y-axis the first formant (F1).

Based on the acoustic measurement of formants, Peterson and Barney (1952) investigate the relation between the vowel phoneme intended by a speaker and the vowel phoneme identified by a listener. In this study, tokens of ten monosyllabic English words were recorded in random order from 76 speakers and they were presented to 70 listeners in an identification task. The production word list and the identification word list were compared with one another, and the tokens' formants were measured acoustically. The results show that the ease of identification varies greatly across vowels. For example, 143 tokens out of the total of 152 tokens of [i] were unanimously classified by listeners, while only 9 tokens out of 152 tokens of [ɑ] were identified unanimously. In the acoustic measurement of these tokens, their first, second and third formants are measured at the most stable part of the monophthongs – the steady state. Using only the static values of F1 and F2 is sufficient to separate the English monophthongs in the vowel perception space even though several adjacent vowels overlap one another, e.g. [ɝ] overlaps with both [ɛ] and [ʊ], which may be distinguished using the third formant values (Peterson and Barney 1952).

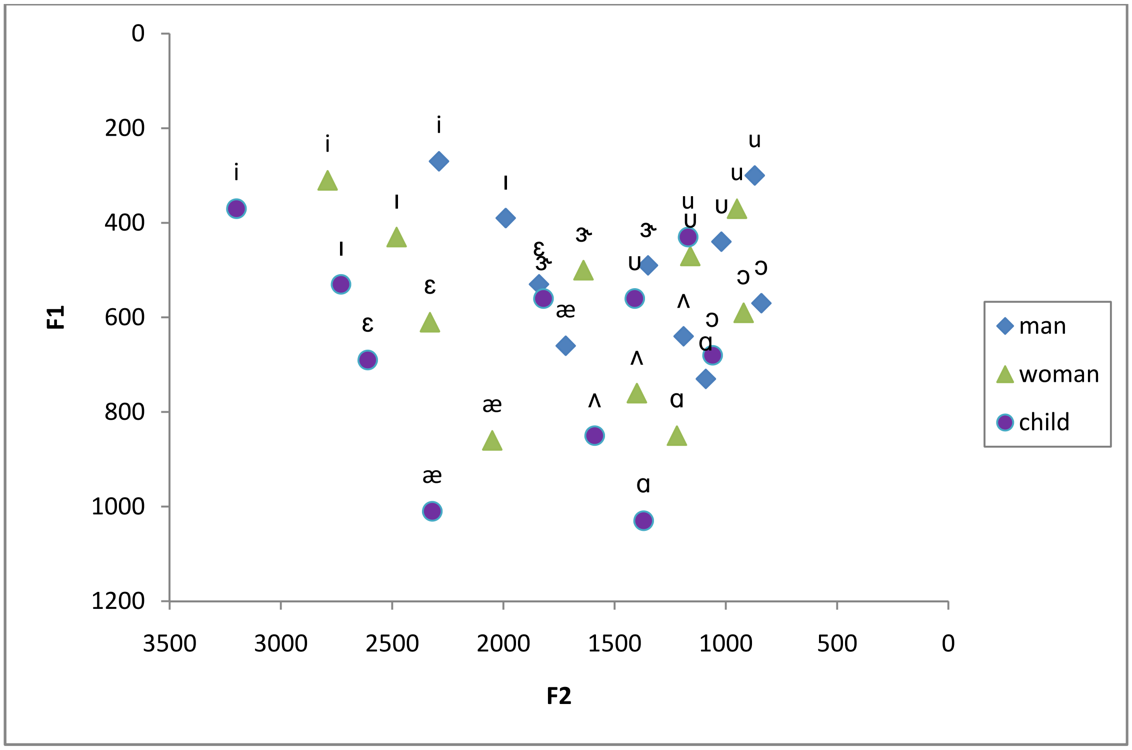


Figure 2.2: Average formants for woman, man and child speakers
Note: These data points are adapted from Peterson and Barney 1952 (p. 183)

2.2.2. Targets of glides

Glides do not reach the same formant values as the pure vowels do. The differences between vowels and glides can be described in formant values, as glides like /w/ and /j/ tend to have a mid-to-high first formant because their higher tongue position creates more constriction in the vocal tract.

Gay (1968, p. 1570) notices from previous studies that diphthong targets are not necessarily phonetically compatible with the vowels used to describe them; for example, the onset of /aj/ can vary from [a] to [æ] and the offset from [ɛ] to [i]. He argues that the vowel-inherent spectral change plays a major role in the perception of both the phonetic and phonemic diphthongs of English and the distinction between diphthongs is governed by the rate of change of the formant transition rather than the onset or offset target positions. In his study (Gay 1968), he records and analyzes both duration and formant-frequency measurements of five English diphthongs and glide nuclei with ten tokens of each diphthong and glide nucleus. All of these tokens are CVC monosyllables and are recorded in a sentence context at three speaking rates: slow, moderate, and fast.

The results from Gay's (1968) experiment show that for all diphthongs, onset frequency positions are stable for all three duration conditions. The offsets, however, show a consistent pattern in F1-F2 across the three duration conditions. The first formant offsets of /ɔj/, /aj/, and /ej/ are higher and their second formant offsets are lower as phoneme duration decreases. Both first and second formant offsets of /aw/ and /ow/ are higher when duration decreases. This pattern is consistent for all five speakers, recorded in this experiment. And it shows that the glide course remains stable and is unique for each diphthong. If duration is reduced, the diphthong gesture is terminated before reaching the offset target, which explains the variation in offset targets among tokens. Figure 2.3 shows the formant slope for "boy" in three speaking rates and their offset positions. Thus Gay (1968) argues that the formant slope, as an inherent feature of diphthongs, and the onset position are important to discriminate diphthongs.

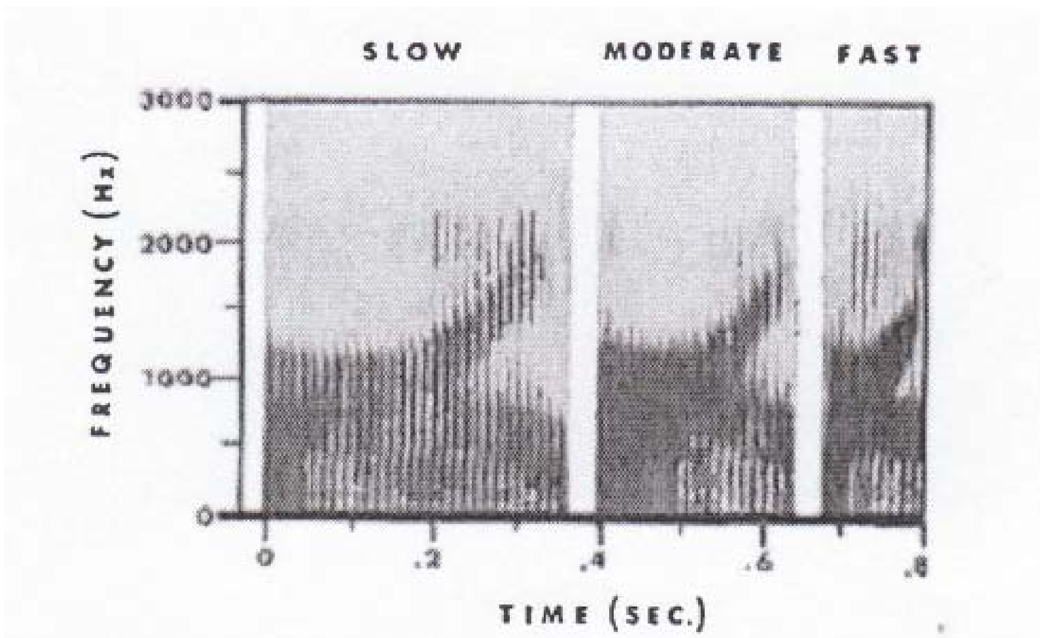


Figure 2.3: Spectrogram of the word "boy" for three conditions of speaking rate.
F2 offset is 1750 Hz (slow), 1600 Hz (moderate), 1340 Hz (fast).
F2 rates of change are 7.0, 7.2, 7.2 Hz/ms, respectively (Gay 1968, p. 1571)

The position of the vowels within the syllable also affects their formant values, potentially making them closer to a consonant than pure vowels. The variants of /i/ and /u/ in Spanish diphthongs exhibit similar frequency patterns to that of consonants in the absolute initial position, to semi-consonants after a consonant, and to glides in a final position (Borzone de Manrique 1976). In absolute initial position, the first formant is of lower frequency due to the vocal tract constriction but the palatalization of /i/ causes an increase in the frequency of the second and third formants. However, the noise at the higher frequency, which represents the turbulence in the vocal tract due to constriction, is not as apparent as that of the pure consonant /j/. In the case of /i,u/ being glides in syllables like Consonant + /ie/ or Consonant + /ei/ (Borzone De Manrique 1976), the frequency aims of /i,u/ are not reached, which is different from the pure vowels /i,u/. Collier et. al. (1982), along with Kakita et. al. (1976) via their X-ray results, have observed that the tongue root is less advanced for the glide /j/ compared to the vowel /i/.

2.3. Duration and its role in distinguishing vowels and glides

2.3.1. Duration characteristics of vowels and diphthongs

Besides the formant frequency differences, duration plays an important role in the identification of vowels and glides. In previous studies where diphthongs were mapped out with two points of values of F1 and F2, the three English diphthongs /aw/, /aj/, and /ɔj/ are distinguishable, as shown in figure 2.4.

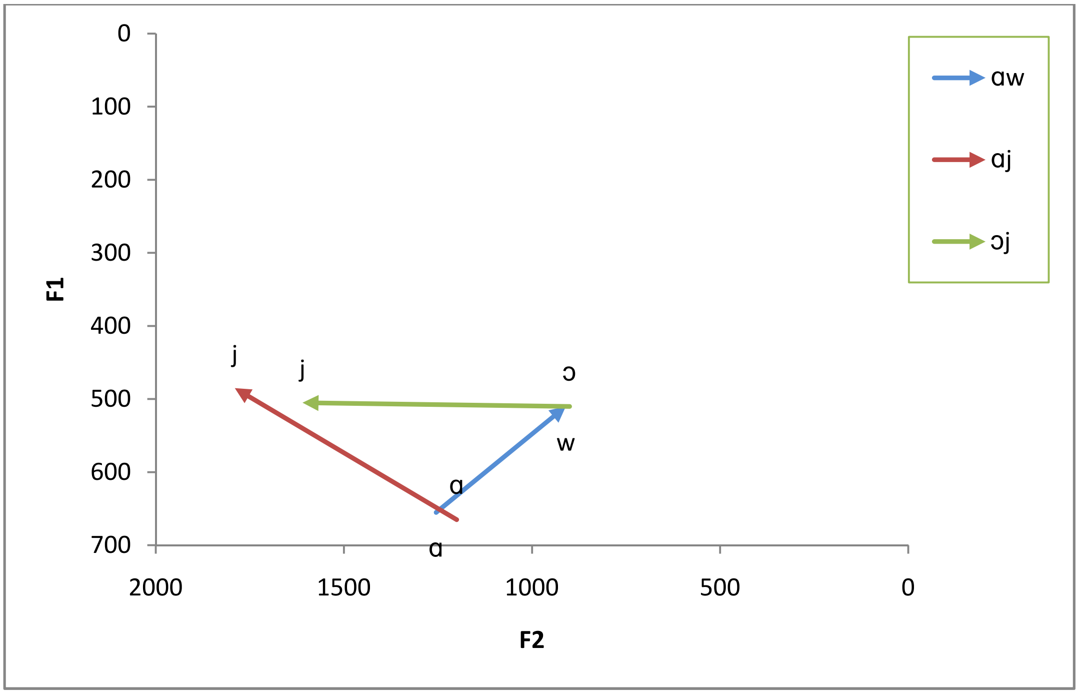


Figure 2.4: Average formants for English diphthongs /ɑw/, /ɑj/, and /ɔj/
Note: These data points are adapted from Lehiste and Peterson 1961 (p. 269)

However, as mentioned in Peterson and Barney (1952), some vowels are easier to discriminate – i.e. /i/ and /u/ – because they are at the limit edge of the vowel space. Meanwhile, the more central vowels are harder to discriminate. To study the nature of these vowels, instead of measuring the formant values at the targets of the nuclei, with targets defined here as the time interval within the syllable nuclei where the formants are parallel to the time axis and with a minimum length of 20 ms, Lehiste and Peterson (1961) suggest a different method of measuring diphthongs using both formant values and duration. For each token in their study, they measured the formant values at the beginning of the onglide and the onglide duration, the formant values at the steady state and the steady state duration, and then the formant values at the end of the offglide and the offglide duration.

The results in figures 2.5 and 2.6 show that the duration of each part of the nucleus adds much information about the nucleus itself. Syllable nuclei are divided into short and long. The short vowels /ɪ ɛ ə ʊ/ are monophthongs with a relatively long offglide and a correspondingly shorter target. The long vowels /i æ ɑ ɔ u/, on the other hand, have a relatively longer target and a short offglide. Temporal values can contribute toward the distinction between long and short monophthongs, or in other words, the distinction between tense and lax vowels.

The term "tense" from this point on is used loosely to describe the "longer" vowels – i.e. vowels /i æ ɑ ɔ u/ in English and at least /a ɤ/ in Vietnamese. These "tense" vowels could either have a slightly longer duration or an actual formant quality difference in comparison to the "lax" vowels. However, as there is no other neutral term to sum up the difference in both duration and formants between these two groups of vowels, I will use these two terms "tense" and "lax" loosely to distinguish between the longer vowels /a ɤ/ and the shorter vowels /ɐ ʌ/ in Vietnamese.

	Onglide	Target (shaded)	Offglide
$o^{ʊ}$	15	25	60
ɔ	16	59	25
u	17	63	20
ɑ	18	60	22
ʊ	19	40	41
æ	20	53	27
ɝ	22	31	47
ɪ	23	32	45
ɛ	23	34	43
ə	24	41	35
i	29	47	24
$e^{ɪ}$	50	30	20

Figure 2.5: Duration in percentage for onglides, steady states and offglides of English monophthongs (Lehiste and Peterson 1961, p. 275)

	Onglide	Target 1	Glide	Target 2	Offglide
ɑj	12	26	31	17	14
ɔj	13	25	30	17	15
ɑw	16	26	27	16	15

Figure 2.6: Duration in percentage for English diphthongs (Lehiste and Peterson 1961)

The importance of duration might also lie with the length of the steady state instead of the length of the entire phone. Studies show that glides have a short steady state of about 30ms while their vowel counterparts can have long steady states. Thus, in diphthongs that contain a vowel and glide, there is a difference in duration of the steady states of the vowel component and the glide component. An acoustic study of Dutch genuine diphthongs, pseudo diphthongs, and long tense vowels that are diphthongized shows that the vowel component has a long steady state while the glide component does not necessarily have a steady state and the same formant values are not reached in closed syllable situations (Collier 1982). Thus the steady states of vowels and glides have different values in duration.

Vowel duration, as well as other time-varying information, is necessary for the accurate identification of vowels in ongoing speech. One of the hypotheses that are suggested by Strange (1989) is that the temporal trajectory of vowels is the primary cue for perception. This hypothesis is supported by Lehiste and Peterson (1961), who show systematic differences in the temporal structure of formant trajectories, especially between tense and lax vowels. It is also supported by the research in Huang (1985), who shows the perceptual salience of such differences in the temporal structure of the formant trajectories. Huang (1985) demonstrates that changing from a tense trajectory to a lax trajectory yields a large shift in the perception of tense-lax vowel pairs and a small shift for lax-lax vowel pairs that differ in tongue height. Also, total syllable duration is an important determinant of category boundaries because tokens with short durations tend to be perceived more as lax. Last but not least, the temporal trajectory hypothesis is supported by Di Benedetto (1989), who shows that among mid-to-high front vowels, the process of vowel identification is a function of the temporal location of the F1 maxima as well as their spectral location.

2.3.2. Duration differences in different environments

In identifying vowels, duration is considered a phonological feature. In Nooteboom's (1972) study of fifteen Dutch vowels, the vowels are divided into four long-versus-short pairs, four separate vowels that do not participate in duration distinction, four genuine short vowels, and three genuine diphthongs. Results from this study show that duration actually divides these fifteen vowels into two groups: a short vowel group and a long vowel group which includes the diphthongs. Therefore, this study supports the hypothesis that intrinsic duration is independent of vowel quality, stress and syllable positions, and that all three Dutch diphthongs are treated as single phonological units in terms of duration.

Duration values in vowels and diphthongs can be affected phonetically by the surrounding environments in some languages, such as English, and a general belief is that there is a correlation between the vowel height and its duration. However, the normalized duration values of English vowels across dialects, which neutralize the effect from the surroundings, show that English vowels have different duration values in different dialects. Vowel duration is a phonologically intrinsic value, thus when vowel height changes, vowel duration does not follow the change (Tauberer and Evanini 2009).

Duration of different sound classes can also vary according to the influence of the environment. In a previous study by Salza (1987), Italian glides /i,u/ are classified as semi-consonants when they precede other vowels, i.e. in the /ia/ and /ua/ sequences, and as glides when they follow other vowels, i.e. in the /ai/ and /au/ sequences. And even though the experiment results are not conclusive for unnatural tokens and stressed tokens, the study shows that the glide in the unstressed /ai/ is shorter in comparison to that of the vowel /i/ in a bisyllabic vowel clusters.

Hualde (2002) compares hiatus tokens and diphthong tokens in Spanish with the hypothesis that the duration of pure vowels is longer than that of glides; thus, if a sequence of vowels is identified as hiatus, their durations would be longer. His research focuses on hiatus and diphthong tokens of /ia/ and is based on speaker intuitions to classify the tokens, but the results show that it is not possible to simply rely on durational measurements alone to distinguish between hiatus and diphthongs.

The total duration of diphthongs is reported to be affected by the syllable structure in the same way as the total duration of monophthongs. Roengpitya (2007) studies the duration of three Thai diphthongs /ia/, /ɯa/ and /ua/ in four syllable structures "piia", "piaʔ", "piat", and "pian". The results show that the total duration of Thai diphthongs is shorter in closed syllables and longer in open syllables; however, the duration of the first element of the diphthong, which is always longer than the second element, stays about the same in all syllable structures. The change in duration appears only in the second vocalic element /a/.

2.4. Dynamic trajectory as a way to classify vowels

Nearey and Assmann (1986), who study the importance of inherent spectral change in perception for ten Canadian English monophthongs in isolated syllables, state that a statistical analysis of F1 and F2 measurements for both diphthongs and monophthongs shows significant formant change. Figure 2.7, taken from Nearey and Assmann's (1986) graph, shows that the "nominal monophthongs" /i/, /e/, and /æ/, as well as the "phonetic diphthongs" /ej/ and /ow/, have significant formant movement in either F1, F2, or both.

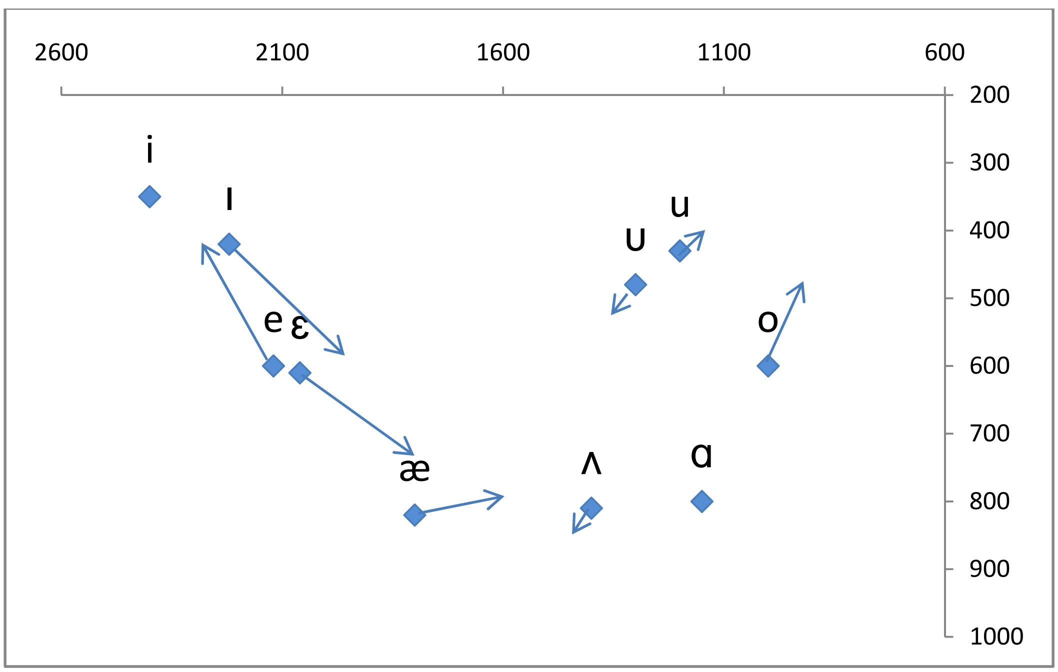


Figure 2.7: Direction for Canadian English monophthongs
(from Nearey and Assmann 1986, p. 1300)

First, out of a series of 40 naturally spoken isolated vowel tokens, Nearey and Assmann (1986) extract a 30-ms *nucleus* segment (A) at the 24% mark of the tokens and a 30-ms *offglide* segment (B) at the 64% mark of the tokens. These extractions are then multiplied by a 30-ms Hamming window and combined into four listening conditions: *natural* order (A-B) with a 10 ms window of silence between them, *repeated* order (A-A) with a 10 ms window of silence between them, *reverse* order (B-A) with a 10 ms window of silence between them, and *baseline* (full token) with a 10 ms window of silence between them. The perception results show that both the *repeated* order and *reverse* order have a significantly higher error rate in comparison with that of *baseline* and *natural* order. Thus, Nearey and Assmann (1986) conclude that vowel-inherent dynamic information is important in the identification of isolated vowels in Canadian English.

In order to establish the importance of the dynamic information of vowels, Strange (1989) did three perception experiments with CVC syllables in a carrier sentence to test the weight of three sources of nuclei structure: (1) the target information available in the vocalic nuclei, (2) the intrinsic duration information specified by the syllable length, and (3) the dynamic spectral information defined over the syllable onsets and offsets. The first experiment yielded a low perception error rate for silent-center syllables (syllables that only have an onset and an offset). The silent-center syllables were identified correctly about 85% of the time and the majority of the vowels were perceived with greater than 90% accuracy. The error rate increased for silent-center syllables whose intrinsic duration differences were neutralized. In the second experiment, perception was significantly better when only initial transitions or final transitions were presented alone. In the third experiment, while the silent-center syllables and excised syllable nuclei with appropriate intrinsic durations were identified with the same error rate as the unmodified controls, the silent-center syllables with neutralized duration were identified with a slightly higher error rate, and truncated vocalic nuclei yielded the highest increase in perception error rate. These results suggest that when duration differences are neutralized, dynamic spectral information is more important than the target information available in the truncated vocalic nuclei. In addition, the static target information is not sufficient for accurate identification of vowels in coarticulated speech.

2.5. Diphthongs as monophonemic and multi-phonemic

In the phonology of vowel systems, there are some debates on the status of diphthongs as phonemes. Hockett (1955) classifies long vowels and diphthongs as nuclei clusters and excludes them from the basic vowel system. However, at least in some languages, diphthongs and long vowels must be treated in the same way as monophthongs (Lass 1984, p. 138). In Icelandic, for example, diphthongs and monophthongs behave in the same way with respect to length: their length is controlled by the same consonantism – shortened before long consonants or clusters, and lengthened before single consonants. Therefore, /i/ is long in [i:s] *ís* 'ice' (nom sg) and short in [iss] *íss* (gen sg), and /ai/ is long in [lai:s] *læs* 'literate' (non-neuter) and short in [laist] *læst* (neuter).

In historical changes across languages, monophthongs have become diphthongs, and diphthongs have reduced to monophthongs (Lipski 1979). In Portuguese, for example, a historical change in the diphthong /aw/ leads to a process of diphthongizing monophthongs /o/ and /e/. First, Portuguese /aw/ retained its final glide and raised to /ow/ in stressed position, after which – through a complex process of morphological analogy affecting some words – /ow/ became /oj/, and thus *causam* –

coisa. At the same time, Portuguese /o/ diphthongized to /aw/ while Portuguese oral and nasal /e/ diphthongized to /aj/.

In the history of Old English and Middle English, the long high vowels /i:/ and /u:/ had been unconditionally diphthongized into /aj/ and /aw/ respectively, which created a pull chain in what is now known as the Great Vowel Shift.

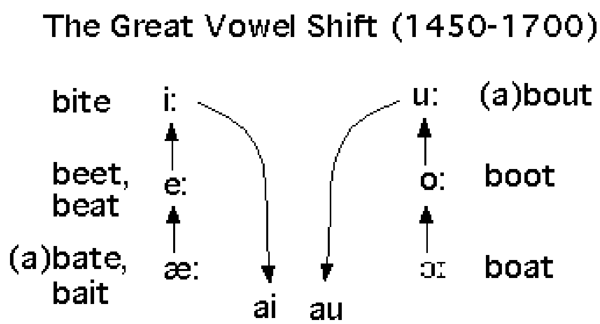


Figure 2.8: The Great Vowel Shift in English

Even though Italian has only seven vowels, there are 167 phonemically different vowel combinations including 106 bi-, 45 tri- and 16 tetra-phonematic sequences (Salza 1987). Among these sequences, the diphthongization – gliding of /i/ and /u/ contiguous to another vowel – results in tautosyllabic diphthongs rather than heterosyllabic vowel clusters. Traditionally, diphthongs have two phones with two targets but only one syllable nucleus. However, if the sequence is biphonematic, the glides /i,u/ are not parts of the nucleus but act as non-syllabic vowels with a lesser degree of sonority. Thus, if the vowel sequences are classified as diphthongs, they are monophonemic and should behave phonetically and phonologically as such.

Taking a different approach, Meyer (2005) studies the poem texts of the early 19th century to find the syllabic classification of Tahitian vowel sequences. He finds that the /ai/, /ei/, /oi/, /ao/, /au/, /ou/, /ae/, and /oe/ vowel sequences are always classified as single syllables, while the /ea/, /eo/, /eu/, /ou/, /iu/, /oa/, /ua/, and /ue/ vowel sequences are always classified as two syllables. This result shows that syllabification can be used as one of the methods to distinguish diphthongs from bisyllabic vowel sequences.

Dutch diphthongs have been grouped into "genuine" and "pseudo" types based on data from Dutch and some other languages (Collier *et al* 1982). Genuine and pseudo diphthongs differ in the degree of openness and advancement in their onsets and offsets. Genuine diphthongs start with a relatively open vocal tract and end with a relatively closed one, while pseudo diphthongs start with an open, half-open, or closed vocal tract before ending with a glide.

These two groups also differ in the harmony of tongue and lip position between the beginning and ending configurations. The vocal tract of genuine diphthongs have the same level of tongue advancement and lip position at the beginning and the end of the diphthongs, whereas the tongue advancement and lip position are changed in the case of pseudo diphthongs. As a result, the genuine diphthongs are overall described as single gestures whereas the pseudo diphthongs are described as two discrete concatenated gestures (Collier *et al* 1982, p. 321).

While studies of the Dutch vowel system divide the Dutch diphthongs into two groups – a group of genuine diphthongs and a group of pseudo diphthongs – based on their type of movements (Collier 1982), diphthongs in other languages are classified based on their glide positions. "Falling" diphthongs start with a more prominent vowel and end with a glide, e.g. English /aj/. "Rising" diphthongs start with a glide and end with a more prominent vowel, e.g. Chinese /ja/. The terms "falling" and "rising" here describe the change in sonority within the diphthongs. Thus, a "falling" diphthong starts at a more sonorant segment, the vowel, and ends with a less sonorant segment, the glide. A "rising" diphthong starts with a less sonorant segment, the glide, and ends with a more sonorant segment, the vowel. "Centralizing" diphthongs start with a more peripheral vowel and end with a more central one, i.e. Irish /iə/. There are also "closing" diphthongs, in which the second element is more close than the first, and "opening" diphthongs, in which the second element is more open than the first. Closing diphthongs tend to be falling, and opening diphthongs – rising (Ren 1986 and Wikipedia).

To describe the characteristics of diphthongs, they are written using a vowel and a glide or using two vowels. Some diphthongs are written using a vowel and a glide, which reflects the idea that diphthongs have, in various orders, one steady state (the vowel) and a transition state (the glide); for example, the "pseudo" diphthongs in Dutch /aj oj uj iw ew/. Some, however, are written with two vowels, as in the case of Dutch "genuine" diphthongs /ei ʌy ɑu/, to show that they contain two steady states (Collier *et al* 1982, Ren 1986, Chitoran 2002). The identification of components of triphthongs also follows this same principle to identify steady states and glides.

2.6. Description of the Vietnamese sound system

Vietnamese is an isolating language with more free morphemes than bound morphemes (Nguyễn Đình Hòa 1997), and most Vietnamese words are single-syllable words. Borrowed words, though they may have several syllables in the source language, are adapted to the monosyllabic format of Vietnamese words. The syllable structure of

Vietnamese is C_1VC_2, in which C_1 can be any of the 20 consonants in the Vietnamese system, and C_2 can be one of these eight consonants /p t k m n ŋ w j/. The labial and alveolar voiced stops /ɓ ɗ/ are described as implosives in Thompson (1965) but as plosives in other studies (Pham 2009, Han 1968). The grapheme *g* is a velar fricative /ɣ/ since for most speakers it is a fricative, and it is considered as such for this study. However, many speakers of the Hanoian dialect have this phone as a voiced plosive stop /g/ instead of the fricative /ɣ/. In foreign loan words, the consonant /p/ and /r/ also appear in syllable-initial position for some speakers (Kirby 2011). The orthographic consonants *ch and tr* are described as an unaspirated unaffricated palatal stop /c/ in Thompson (1965) but as an affricate /tç/ in Kirby (2011). For this study, I will use Thompson's (1965) description of this consonant.

	Bilabial	Labio-dental	Dental	Alveolar	Palatal	Velar	Glottal
Stop	/ɓ/[1] *b*		/t/ *t* /tʰ/ *th*	/ɗ/ *đ*	/c/ *ch, tr*	/k/ *c, k, q*	/ʔ/ *ø*
Nasal	/m/ *m*		/n/ *n*		/ɲ/ *nh*	/ŋ/ *ng, ngh*	
Fricative		/f/ *ph* /v/ *v*		/s/ *x, s* /z/ *d, gi, r*		/x/ *kh* /ɣ/ *g, gh*	/h/ *h*
Approximant	/w/ *o, u*		/l/ *l*				

Table 2.1: Consonants in the Hanoi dialect

[1] /ɓ/ only appear in onsets and its allophone /p/ – in coda for native words

In terms of tones, Thompson (1965) describes the Hanoi dialect to have six tones, although only two of these tones – the Rising tone and the Drop tone – are realized in syllables with unreleased oral stops (Kirby 2011).

No.	Name	Ortho-graphy	Pitch description	Other phonetic qualities	Examples
1	**Level**	a ˧	Mid or high-mid trailing pitch, nearly level when syllable is not final in pause group; falls to low range in final syllables	Laxness	*ta* /ta ˧/ 'I, me'
2	**Rising**	á ˧˥	High rising pitch, might be heard as high level in rapid speech	Tenseness	*tá* /ta ˧˥/ 'dozen'
3	**Falling**	à ˧˩	Low trailing pitch	Often breathy and lax	*tà* /ta ˧˩/ 'evil'
4	**Curve**	ả ˧˩˧	Mid-low dropping pitch, not too abrupt, with a rise at the end	Tenseness	*tả* /ta ˧˩˧/ 'describe'
5	**Broken**	ã ˧ʔ˥	High rising pitch	Glottalization	*tã* /ta ˧ʔ˥/ 'diaper'
6	**Drop**	ạ ˧ʔ˩	Low dropping pitch with an abrupt falling	Syllable ends in stops or is cut off abruptly by a glottal stop; tenseness	*tạ* /ta ˧ʔ˩/ 'weight''

Table 2.2: Vietnamese tones

Nguyễn Đình Hòa (1997), as well as Thompson (1965), describes Vietnamese as a language with eleven monophthongs /a ɐ ʌ ɛ e i ɔ o ɤ u ɯ/. Nine out of these eleven vowels appear in all contexts except for /ɐ/ and /ʌ/, which only appear in closed syllables. These vowels /ɐ ʌ/ are also described as short vowels. The unrounded vowels are /i/, /e/, /ɛ/, /ɤ/, and /ɯ/, and the rounded vowels are /u/, /o/, and /ɔ/ (Nguyễn Đình Hòa 1997).

	Front	Central	Back
High	i		u ɯ
Mid	e	ʌ ɤ	o
Low	ɛ	ɐ a	ɔ

Table 2.3: Vietnamese vowels (based on Nguyễn Đình Hòa 1997)

Not all researchers agree that Vietnamese has eleven monophthongs. In his typological study of vowel systems, Crothers (1978, p. 115) classifies Vietnamese as a [10:2] language with a total of ten monophthongs /i e ɛ a ɔ o u ʌ ɤ ɯ/, /ɯ ɤ/ being the interior vowels . The phone [ɐ] is not included as a separate phoneme and interpreted as a shorter /a/ (Smalley 1957, Le-Van-Ly 1960, Crothers 1978).

In another groups of studies (Nguyễn Bạt Tụy 1949, 1959; Haudricourt 1952, Đoàn Thiện Thuật 1977, Kirby 2011), Vietnamese is described as a language with nine monophthongs /i e ɛ a ɔ o u ɤ ɯ/ with two allophones /ɐ ʌ/. The differences in tension between the allophone /ɐ/ and the vowel /a/ are explained as being the effect of the following consonants onto the actual vowels (Nguyễn Bạt-Tụy 1959). If vowel /a/ is followed by a voiced consonant, it would surface as the longer allophone /a/. If vowel /a/ is followed by a voiceless consonant, it then would be the shorter allophone /ɐ/. In this analysis by Nguyễn Bạt-Tụy (1959), these Vietnamese word pairs *cáp* - *cắp*, *cát* - *cắt*, *các* - *cắc* are actually not minimal pairs as they are transcribed with the same vowel /a/ but with different consonants: /kab ˧˥/ - /kap ˧˥/, /kad ˧˥/ - /kat ˧˥/, /kag ˧˥/ - /kak ˧˥/. However, this analysis does not explain the minimal pairs ending in nasals and glides in table 1.1 and the minimal pairs for vowels /ɤ/ and /ʌ/ in table 2.4.

No.	Orthography	IPA	Gloss	Orthography	IPA	Gloss
1	*trơn*	/cɤn ˧/	polished	*chân*	/cʌn ˧//	foot
2	*đơm*	/ɗɤm ˧/	measure	*đâm*	/ɗʌm ˧/	stab
3	*tơi*	/tɤj ˧/	loosely	*tây*	/tʌj ˧//	west
4	*sơi*	/sɤj ˧/	eat	*xây*	/sʌj ˧//	build

Table 2.4: The minimal pairs of the vowels /ɤ ʌ/

To a non-native speaker, the vowels within these pairs /a ɐ/ and /ɤ ʌ/ have similar formant quality but different in length, and other previous studies (Thompson 1965, Nguyễn Đình Hòa 1966, and Han 1968) treat /a/, /ɐ/, /ɤ/, and /ʌ/ as different phonemes. Han (1968) shows that feature of duration operates as the primary cue for distinguishing the vowels within each pair. The duration ratio of /a/ versus /ɐ/ and /ɤ/ versus /ʌ/ is approximately 2:1 in various phonetic environments (Han 1968). The vowels in /a ɐ/ and /ɤ ʌ/ overlap with each other in the vowel space (Han 1968), and the small formant differences between them might be influenced by their duration differences.

Of the eleven vowels in Vietnamese, six – /i/, /e/, /ɯ/, /ɤ/, /u/, and /o/ – are described by Thompson (1965) as diphthongs in open syllables, as shown in table 2.5. They are nonetheless identified phonologically as monophthongs rather than diphthongs in this study.

Orthography	Phone	Closed syllables	Open syllables	Structure
i/y	/i/	[i]	[ji]	Glide + Vowel
ê	/e/	[e]	[ej]	Vowel + Glide
ư	/ɯ/	[ɯ]	[ɰɯ]	Glide + Vowel
ơ	/ɤ/	[ɤ]	[ɤɰ]	Vowel + Glide
u	/u/	[u]	[wu]	Glide + Vowel
ô	/o/	[o]	[ow]	Vowel + Glide

Table 2.5: Monophthongs that diphthongize in open syllables (Thompson 1965)

As for diphthongs in Vietnamese, Thompson (1978) describes them as "two-element nuclei" – sequences of a vowel followed by a glide. He also identified several diphthongs, i.e. /iu/, /ie/, /ɯɤ/ and /uo/, as vowel clusters instead of diphthongs because the glides in these are as prominent as the preceding vowels. In this study, I will divide Vietnamese diphthongs into two groups, partially based on Thompson (1965): the diphthong group and the contour vowel group. Diphthong is defined as a sequence in which a vowel is followed or preceded by a glide, while a contour vowel is defined as a sequence of two vowels. The diphthong group has 25 diphthongs in total, and this group does not include the six long vowels provided in table 2.5. The diphthongs are then divided into three subgroups: falling diphthongs in Vj form, falling diphthongs in Vw form, and rising diphthongs in wV form (table 2.6).

Vj diphthongs		Vw diphthongs		wV diphthongs	
ai	/aj/	*ao*	/aw/	*oa*	/wa/
ay	/ɐj/	*au*	/ɐw/	*oă*	/wɐ/
ây	/ʌj/	*âu*	/ʌw/	*uâ*	/wʌ/
oi	/ɔj/	*eo*	/ɛw/	*oe*	/wɛ/
ôi	/oj/	*êu*	/ew/	*uê*	/we/
ơi	/ɤj/	---------------		*uơ/uâ*	/wɤ/
ui	/uj/	*iu*	/iw/	*uy*	/wi/
ưi	/ɯj/	*ưu*	/ɯw/	----------------	

Table 2.6: Diphthongs in Vietnamese (based on Thompson 1965)

The contour vowel group has three members and are all from the upper vocalic group (Thompson 1965). The first contour vowel is /ie/, which is written as *iê, yê* or *ia*. The second is /ɯɤ/, written as *ươ* or *ưa*, and the third is /uo/, written *uô* or *ua*. In closed syllables, the second segments in the contour vowels are lower in the vowel space, in comparison to the first segments, but they have the same frontness and backness quality. When the contour vowels appear in open syllables, their second segments are centralized and have similar quality to a central vowel /ʌ/ (Nguyễn Đình Hòa 1966, Emeneau 1951). These contour vowels are grouped into a different group, separated from the diphthongs in table 2.6, in Han's (1966) study as well.

Orthography		IPA
(1)	*iê*	/ie/
	yê	
(2)	*uô*	/uo/
(3)	*ươ*	/ɯɤ/

Table 2.7: Vietnamese contour vowels

Thompson (1965) classifies as triphthongs all nuclei that have a sequence of three elements. These can be a diphthong plus a glide or a contour vowel plus a glide.

Vowel-Vowel-Glide			Glide-Vowel-Vowel			Glide-Vowel-Glide		
(1)	*iêu* or *yêu*	/iew/	(2)	*uyê*	/wie/	(3)	*uya*	/wiə/
	uôi	/uoj/					*uyu*	/wiw/
	ươi	/ɯɤj/					*oai*	/waj/
	ươu	/ɯɤw/					oây	/wʌj/
							oay	/wɐj/

Table 2.8: Vietnamese triphthongs (based on Thompson 1965)

Chapter 3

The Vietnamese Speech Corpus

3.1. Introduction

The Vietnamese Speech Corpus is a collection of recordings I made in Vietnam during the summer of 2008 to build a research speech corpus for future Vietnamese linguistic studies. The corpus contains a total of 29 hours of speech: 18 hours of spontaneous speech in interview format, 5.5 hours of read speech in news reading format, and 6 hours of read speech in embedded word list format. There are 26 speakers recorded in the Vietnamese Speech Corpus: 13 Hanoi dialect speakers and 13 Saigon dialect speakers. The entire corpus is recorded one-on-one in a sound-proof studio using the Sony PCM-D50 Recorder at the sampling rate of 44.1 kHz.

3.2. Recording techniques

All of the audio recordings were recorded digitally at the sampling rate of 44.1 kHz using a Sony PCM-D50 Recorder with a high quality stationary stereo microphone. All speakers were individually recorded in sound studios, one sound studio in Hanoi and one in Hochiminh City. For each speaker, all the recordings were done completely in one session, which ranged from one and a half hour to two hours with short breaks every 30 minutes. In these recording sessions, I was the only interviewer and only one speaker was present inside the studio at a time. Both studios are built specifically for music recordings and radio recording, thus completely sound-proof and these rooms have a dimension of about 8 feet by 8 feet. During the recording, the speaker sat about two feet across from me, the interviewer, with the microphone attached to a microphone holder and placed 10 inches from the speaker's mouth.

The 26 speakers in this corpus were chosen through a friend and family network. Most speakers are schoolmates of each other, and they are all between the age of 18 and 24 with lower middleclass to upper middleclass socio-economic status. They were all born and raised in the cities of the dialects and at least one of their parents is a native speaker of that dialect. For the Hanoi dialect, two of the speakers are within the author's social circle, two others are in the social circles of the author's acquaintances, and the other nine speakers are in the social circles of other Hanoian speakers (figure 3.1). For the Saigon dialect, four speakers belong to the social circles of the author's acquaintances, and the other nine speakers are in the social circles of other Saigon speakers (figure 3.2).

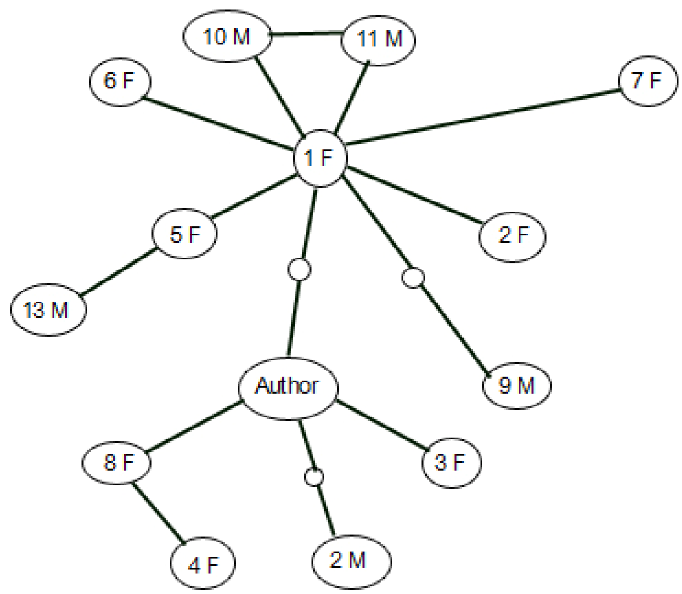


Figure 3.1: The network of Hanoi dialect speakers

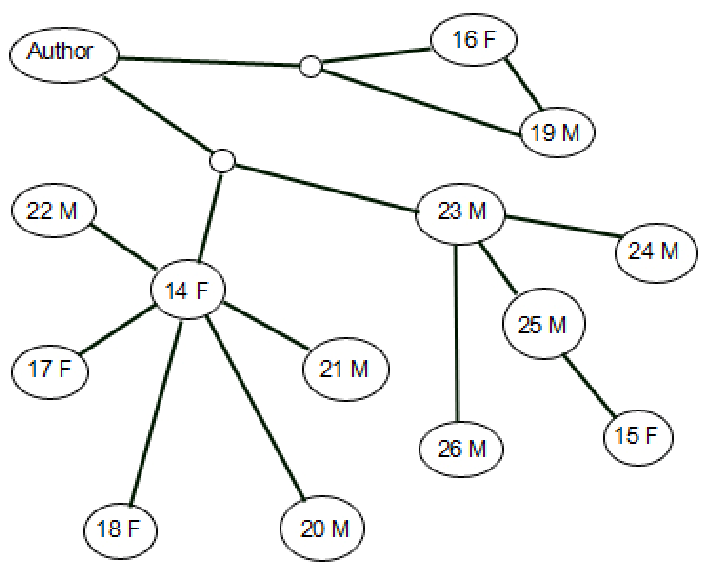


Figure 3.2: The network of Saigon dialect speakers

The first portion of each recording session is the spontaneous speech portion, recorded in the format of the sociolinguistic interview (Labov 1984). For all speakers, this portion lasted about 60 minutes with questions drawn from an adapted set of question modules, similarly to Q-GEN-II (Labov 1984), including topics on speakers' demography, love relationships, fears, dreams, family, neighborhood relationships and changes, hobbies, and Vietnamese linguistic topics such as loanword and reduplication. The purpose of these sociolinguistic interviews is to obtain natural speech from speakers by eliciting narratives of their lives and encouraging the speakers to lead the conversation. Thus even though the recordings were set in a studio setting, the topics made the speakers comfortable and relaxed. While the interview portion is about 60 minutes for most speakers, I only used the first 15 minutes of interview speech from six female speakers of the Hanoi dialect for this dissertation.

The next read speech portion contains speakers reading three pieces of news with a total of 2679 words, extracted from a current issue of the popular newspaper *Tuổi Trẻ* ("Youth"). The first news piece is about the rising price of rice grain in the market and it has 578 words, containing many number words. The second news piece is about the current music show business in Vietnam and it has 780 words total, containing some loanwords. The third news piece is about popular vacation destinations for the summer and it has 1321 words total, containing many proper names of local attractions. This news reading portion was recorded with a short break between the news pieces. For all speakers except speaker 2, all three news pieces are read and recorded. Only the first news piece was recorded for speaker 2 because she started to feel uneasy about her weak reading skills and her speech reflected this insecurity after the first news piece. For this dissertation, I only use the recordings of the first news piece of all six female speakers in Hanoi dialect.

The word list contains 160 minimal pairs eliciting all Vietnamese monophthongs, diphthongs and triphthongs in two environments: CV and CVC. Each word is placed in a carrier sentence *Tôi đọc từ X thêm ba lần nữa. X, X, X* ("I say the word X three more times. X, X, X"); thus, there are four tokens of each word. The sentences are randomized and then grouped together into four groups so that the speakers could have short breaks between them. During the recording session, the speakers are asked to read the number of the sentence and then the sentence with a read speech rate. The words embedded in these framed sentences are organized in the word list table 3.1, which is transcribed for the Hanoi dialect. Each token in the word list is listed in orthography, IPA transcription, and with an English gloss. The shaded boxes show the non-existent environments for those vowels and diphthongs.

Table 3.1: The word list in the framed sentence portion of the recordings

No.	Phone	CV	IPA	Gloss	CVC	IPA	Gloss
1	a	*ta*	/ta ˧/	I, me	*tan*	/tan ˧/	melt
		toa	/twa ˧/	train car	*toan*	/twan ˧/	intent
		tá	/ta ˧˥/	dozen	*tát*	/tat ˧˥/	slap
		xóa	/swa˧˥/	erase	*toát*	/twat ˧˥/	sweat out
					tám	/tam ˧˥/	eight
2	ɐ				*tăm*	/tɐm ˧/	bubble
					thoăn	/tʰwɐn ˧/	fast
					xoăn	/swɐn ˧/	curly
					tắt	/tɐt ˧˥/	turn off
					cắt	/kɐt ˧˥/	cut
					xoắn	/swɐn˧˥/	curly
3	ʌ				*vân*	/vʌn ˧/	pattern
					tân	/tʌn ˧/	new
					tuân	/twʌn ˧/	follow
					tuất	/twʌt ˧˥/	year, age
					bất	/ɓʌt ˧˥/	not
					quất	/kwʌt ˧˥/	whip
					cất	/kʌt ˧˥/	put away
					tất	/tʌt ˧˥/	socks
					gấc	/ɣʌk ˧˥/	a fruit
4	ɛ	*toe*	/twɛ ˧/	fan out	*khoen*	/xwɛn ˧/	ring
		xe	/sɛ ˧/	car	*xen*	/sɛn ˧/	interrupt
		té	/tɛ ˧˥/	fall down	*tét*	/tɛt ˧˥/	split out
		toé	/twɛ ˧˥/	splash	*toét*	/twɛt ˧˥/	tear out

5	e	*tê* *quê* *thuê* *tế* *tuế*	/te ˧/ /kwe ˧/ /tʰwe ˧/ /te ˧˥/ /twe ˧˥/	numb origin hire pray old age	*tên* *tết*	/ten ˧/ /tet ˧˥/	name new year
6	i	*ti* *quy* *tuy* *tí* *túy*	/ti ˧/ /kwi ˧/ /twi ˧/ /ti ˧˥/ /twi ˧˥/	chest turtle however small drunk	*tinh* *tin* *tít*	/tiɲ ˧/ /tin ˧/ /tit ˧˥/	sharp believe further
7	ɔ	*to* *có*	/tɔ ˧/ /kɔ ˧˥/	large have	*con* *tót*	/kɔn ˧/ /tɔt ˧˥/	kid jump on
8	o	*tô* *tố*	/to ˧/ /to ˧˥/	big bowl sue	*tôn* *tốn* *tốt*	/ton ˧/ /ton ˧˥/ /tot ˧˥/	aluminum cost good
9	ɤ	*quơ* *khuơ* *tơ* *tớ* *quớ*	/kwɤ ˧/ /xwɤ ˧/ /tɤ ˧/ /tɤ ˧˥/ /kwɤ˧˥/	grab wave silk I, me mistaken	*tơn* *vớt*	/tɤn ˧/ /vɤt ˧˥/	hurry scoop
10	u	*tu* *tú*	/tu ˧/ /tu ˧˥/	abstinence beautiful	*phun* *cút*	/fun ˧/ /kut ˧˥/	spray go away
11	ɯ	*tư* *tứ*	/tɯ ˧/ /tɯ ˧˥/	private four	*tưng* *vứt*	/tɯŋ ˧/ /vɯt ˧˥/	bounce throw

12	ie	*khuya* *kia* *tia* *tía*	/xwie˧/ /kie ˧/ /tie ˧/ /tie ˧˥/	late night over there ray purple	*khuyên* *tiên* *tuyên* *tuyết* *tiết* *kiết* *tiêu* *kiêu* *tiếu*	/xwien ˧/ /tien ˧/ /twien ˧/ /twiet ˧˥/ /tiet ˧˥/ /kiet ˧˥/ /tiew ˧/ /kiew ˧/ /tiew ˧˥/	advise angel announce snow secrete constipate digest proud funny
13	ɯɤ	*cưa* *tưa* *tứa*	/kɯɤ ˧/ /tɯɤ ˧/ /tɯɤ˧˥/	saw fray seep out	*tươm* *thướt* *hươu* *cưới* *khướu* *bươu* *tươi* *tước* *tưới*	/tɯɤm ˧/ /tʰɯɤt ˧˥/ /hɯɤw˧/ /kɯɤj ˧˥/ /xɯɤw˧˥/ /ɓɯɤw ˧/ /tɯɤj ˧/ /tɯɤk ˧˥/ /tɯɤj ˧˥/	neat soft sway deer marry a bird swelled fresh peel water (v.)
14	uo	*tua* *cua* *túa*	/tuo ˧/ /kuo ˧/ /tuo ˧˥/	rewind crab spill out	*khuôn* *tuôn* *xuôi* *suối* *tuốt* *cuối*	/xuon ˧/ /tuon ˧/ /suoj ˧/ /suoj ˧˥/ /tuot ˧˥/ /kuoj ˧˥/	mold spill out follow stream peel final
15	iw	*tiu* *thiu* *khuỷu* *tíu*	/tiw ˧/ /tʰiw ˧/ /xwiw˧˩˧/ /tiw ˧˥/	sad soured elbow chirpy			

16	ew	*thêu* *kêu* *tếu*	/tʰew ˧/ /kew ˧/ /tew ˧˥/	embroidery call funny			
17	ɛw	*keo* *teo* *xéo*	/kɛw ˧/ /tɛw ˧/ /sɛw˧˥/	glue shrink tilted			
18	aw	*cao* *tao* *táo* *cáo*	/kaw ˧/ /taw ˧/ /taw ˧˥/ /kaw˧˥/	tall I, me apple fox			
19	ɐw	*sau* *cau* *cáu*	/sɐw ˧/ /kɐw ˧/ /kɐw˧˥/	after betel upset			
20	ɯw	*sưu* *khưu* *cưu* *cứu*	/sɯw˧/ /xɯw˧/ /kɯw˧/ /kɯw˧˥/	collect a name protect rescue			
21	ʌw	*câu* *tâu* *tấu*	/kʌw ˧/ /tʌw ˧/ /tʌw ˧˥/	fishing report tell on			
22	aj	*tai* *thoai* *khoai* *cai* *tái* *khoái*	/taj ˧/ /tʰwaj ˧/ /xwaj ˧/ /kaj ˧/ /taj ˧˥/ /xwaj˧˥/	ear slight slope potato cut off rare like			

23	ɐj	*xoay* *tay* *hoay* *cay* *xoáy* *táy*	/swɐj ˧/ /tɐj ˧/ /hwɐj ˧/ /kɐj ˧/ /xwɐj˧˥/ /tɐj ˧˥/	turn hand turning spicy spiral fidget			
24	ɯj	*gửi* *chửi* *cửi*	/ɣɯj˧˩˧/ /cɯj˧˩˧/ /kɯj˧˩˧/	send swear loom			
25	ɤj	*cơi* *tơi* *tới*	/kɤj ˧/ /tɤj ˧/ /tɤj ˧˥/	stir separated arrive			
26	ʌj	*tây* *cây* *tấy*	/tʌj ˧/ /kʌj ˧/ /tʌj ˧˥/	western tree swollen			
27	ɔj	*coi* *toi* *cói*	/kɔj ˧/ /tɔj ˧/ /kɔj ˧˥/	watch die a grass			
28	oj	*tôi* *côi* *tối*	/toj ˧/ /koj ˧/ /toj ˧˥/	I, me alone dark			
29	uj	*vui* *tui* *cúi* *túi*	/vuj ˧/ /tuj ˧/ /kuj ˧˥/ /tuj ˧˥/	happy I, me bend over bag			

3.3. The structure of the corpus

The master recordings for each speaker are split into the speech of the interviewer, which is the author, and the speech of the speakers based on the transcription. The speech of the interviewer is removed from the corpus, and the recordings of the speaker are divided into smaller .wav files of about one minute in length for further analysis. Each of these .wav files is paired with a .TextGrid file of the same name. The file names contain the information on speakers, speech style and the order of recording.

File names for .wav and its companion .TextGrid file:

XXY_ZZZZ.wav and XXY_ZZZZ.TextGrid

in which:

XX is the speaker's number, ranging from 01 to 26

Y is the speech style:

- 3 for spontaneous speech,
- 6 for news reading
- 7 for framed sentences

ZZZZ is the order of recording, starting at 0001 for each speaker.

Example 1: 013_0005.wav and 013_005.TextGrid are the 5th files for speaker 1 and they belong to the Spontaneous Speech portion of the recording.

Example 2: 176_0082.wav and 176_0082.TextGrid are the 82nd files for speaker 17 and they belong to the News Reading portion of the recording.

Figure 3.3: Structure of file names in the Vietnamese Speech Corpus

Table 3.2 below has the corpus structures with dialectal and speaker information, gender, speech style and file names for all six female speakers of the Hanoi dialect, which is the data I used for this dissertation.

Dialect	Speaker	Gender	Speech Style	Total file	Names
Hanoi	1	female	Spontaneous 3 News 6 Sentences 7	11 15 17	013_0001 to 013_0011 016_0012 to 016_0026 017_0027 to 017_0043
Hanoi	2	female	Spontaneous 3 News 6 Sentences 7	63 3 19	023_0001 to 023_0063 026_0064 to 026_0066 027_0067 to 027_0085
Hanoi	3	female	Spontaneous 3 News 6 Sentences 7	150 10 17	033_0001 to 033_00150 036_0151_036_0160 037_0161 to 037_0177
Hanoi	4	female	Spontaneous 3 News 6 Sentences 7	53 9 15	043_0001 to 043_0053 046_0054 to 046_0062 047_0063 to 047_0078
Hanoi	5*	female	Spontaneous 3 News 6 Sentences 7	186 9 16	053_0001 to 053_0187* 053_0197 to 053_0204* 053_0208 and 053_0224 056_0188 to 056_0196 057_0205 and 057_0206 057_0209 to 057_0222
Hanoi	6	female	Spontaneous 3 News 6 Sentences 7	71 8 16	063_0001 to 063_0071 066_0072 to 066_0079 067_0080 to 067_0095

Table 3.2: Data set of six female Hanoi dialect speakers

Note: * Files 177-179, 181, 183, 185, 198, 200, 202, 207 and 223 are excluded

3.4. Annotation

3.4.1. Transcription system and rules

The speech data from all the recordings, both Hanoi and Saigon dialects, has been first transcribed at the word level by a native speaker of Vietnamese, who is a speaker of the Saigon dialect and has had contact with Hanoian speakers as well. The transcriptions are then checked by the author, who is a speaker of both dialects. In these transcriptions, the Vietnamese orthography is replaced with a transcription which is similar to the Vietnamese orthography without the attached diacritics. The diacritics are replaced with the symbols available in the English keyboard, and tones are marked by numbers (as shown in figure 3.4). The basic rules for transcriptions are similar to those of the SWITCHBOARD transcriptions (Switchboard: A User's Manual), including rules about capitalization, writing out numbers, and punctuation. The rules for special words are listed below.

1. Foreign words are marked in < > with origins, e.g. <English show> means the English word "show".
2. Hyphenation is used for foreign names such as “Dda-ra-hoa”
3. Acronyms are preceded by the @ symbol, e.g. @NASA
4. Abbreviations with a s sequence of individual letters should be written in all caps, with each individual letter preceded by a ~ tilde symbol, e.g. ~K ~T ~X
5. In partially-pronounced words, a single dash - is used to indicate point at which word was broken off, e.g. lo- long.
6. Hard-to-understand sections are marked with (( )) with the guessed transcriptions inside, if possible, e.g. ((tao1 co2 bie^2t)) or just (( )).
7. Proper nouns are marked with a caret ^, e.g. ^Ha3 ^No^i5.
8. Filled pauses are marked with %, e.g. %uh, %oh.
9. Speaker noises are marked inside { }, e.g. {laugh}.
10. Background noise are marked in < >, e.g. <noise>.
11. Interviewer speech has time stamps and marked as <giang>.

Figure 3.4: Special transcription rules

Ortho-graphy	Transcription	Ortho-graphy	Transcription	Ortho-graphy	Transcription
b *báo*	b bao2	*r* *ráng*	r ra2ng	*ô* *hồng*	o^ ho^3ng
c *cá*	c ca2	*s* *săn*	s sa(1n	*ơ* *chợ*	o+ cho+6
d *dông*	d do^1ng	*t* *tinh*	t ti1nh	*u* *xung*	u xu1ng
đ *đận*	dd dda^6n	*v* *vũ*	v vu5	*ư* *trưng*	u+ tru+1ng
g *gụ*	g gu6	*x* *xuân*	x xua^1n	*y* *hy*	y hy1
h *hát*	h ha2t	*a* *anh*	a a1nh	level *a*	1 a1
k *khôi*	k kho^i1	*ă* *trăng*	a(tra(1ng	rising *á*	2 a2
l *lũ*	l lu5	*â* *chân*	a^ cha^1n	falling *à*	3 a3
m *miền*	m mie^3n	*e* *hẹn*	e he6n	curve *ả*	4 a4
n *núi*	n nui2	*ê* *hên*	e^ he^1n	broken *ã*	5 a5
p *phở*	p pho+4	*i* *xin*	i xi1n	drop *ạ*	6 a6
q *quê*	q que^1	*o* *cóc*	o co2c		

Table 3.3: Transcriptions in the Vietnamese Speech Corpus

Explanations	Actual transcription script
time stamp	109.66
"%" marks pause filler	The^2 la3 bo^2 me6 a1nh y2 cu5ng dde^2n nha3- %ah me6 a1nh y2 cu5ng dde^2n nha3 ro^i3.
"-" marks partially-pronounced word	Me6 c-me6 a1nh y2 muo^2n mua1 nha3 rie^1ng cho1 e1m vo+i2 a1nh a^y2 o+4 rie^1ng chu+2 kho^1ng muo^2n ai1 la3m da^u1.
"{ }" marks speaker's noise	Kho^1ng o+4 chu1ng vo+i2 ai1 he^2t {laugh}.
time stamp	144.79
	E1m cu5ng muo^2n la3 o+4 ne^u2 m-ma3 e1m cu5ng muo^2n la3 o+4 chu1ng vo+i2 me6 cho^3ng y2.
	Ti2nh me6 cho^3ng e1m no2ng la(2m vo+i2 ca4 cu6c ti2nh la(2m.
time stamp	187.27
marked for interviewer's speech	<giang>
time stamp	190.06
	Me6 e1m thi3 ga^3n la(2m, cai2 gi3 cu5ng ta^1m su+6 he^2t.
	Me6 e1m cai2 gi3 cu5ng ta^1m su+6 he^2t.
time stamp	200.70
marked for interviewer's speech	<giang>

Figure 3.5: A part of the transcription with explanations

The time stamps in the transcript are used to cut the large .wav files from the recordings into smaller .wav files for the corpus. All .wav files are cut into .wav files of about one minute in length. Scripts are cut at the same time stamps and .TextGrid files are created with two tiers: a Word tier and a Phone tier. The Word tier has boundaries placed between each syllable, and has silence markers such as SP and SIL as well. The syllables in the Word tier have the exact format as in the scripts, thus have the information about the onset, nucleus, coda and tones. The Phone tier is created from the Word tier with boundaries between all the phones, consonants and vowels alike without the tones. Then the .TextGrid files are forced-aligned with the .wav files using the Penn Phonetics Lab Forced Aligner (P2FA) (Yuan and Liberman 2008). After processing the .wav files and aligning the .TextGrid files, the author checks and hand-corrects all the boundaries in both tiers for all speakers.

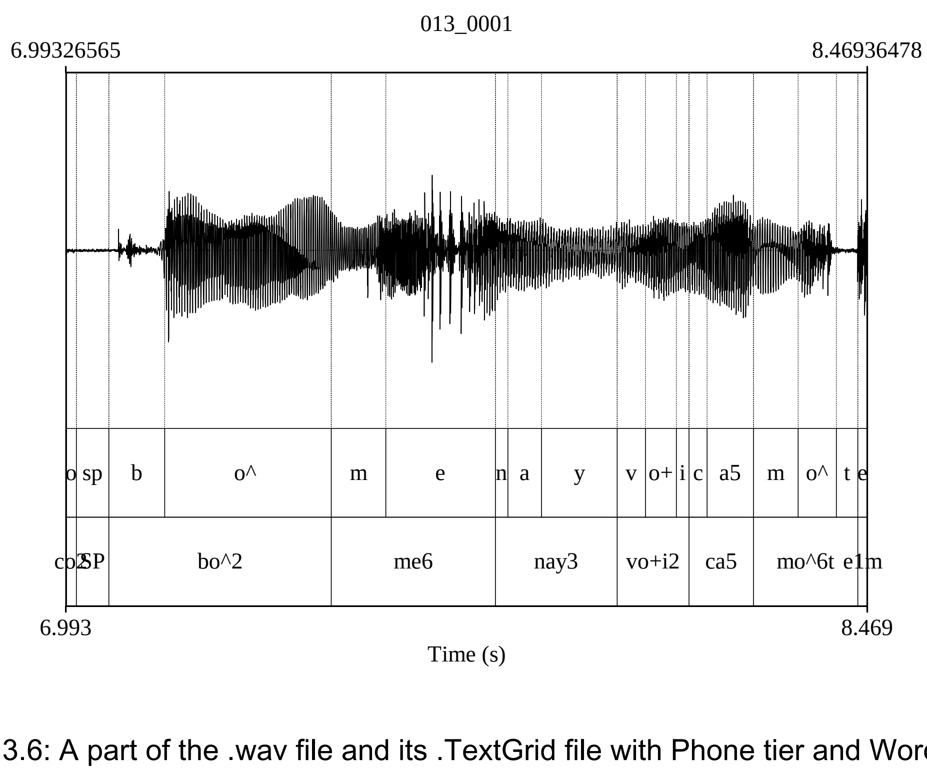


Figure 3.6: A part of the .wav file and its .TextGrid file with Phone tier and Word tier

3.4.2. Rules for boundary alignment

As this study focuses on the vowel system, the rules for aligning the boundaries of the vowels and diphthongs are very important. For all the boundaries, the most important standard is my native speaker's perception. I move the boundaries to their correct places based on the expected spectrograms and waveform of each sound. Boundaries between consonants and vowels are based on voice bar, formants, intensity and the existence of noise. Each monophthong is one phone. All the diphthongs and contour vowels are divided into two phones, and triphthongs – three phones. Boundaries between vowels are based on the change in formants, and the transitional glides between vowels are divided equally into phone on each side unless otherwise noted from the actual sounds. My final decisions are based on whether I, as a native speaker, can hear just that sound, or I can hear the neighboring sounds as well. The ideal target is to be able to hear only one segment between each pair of boundaries.

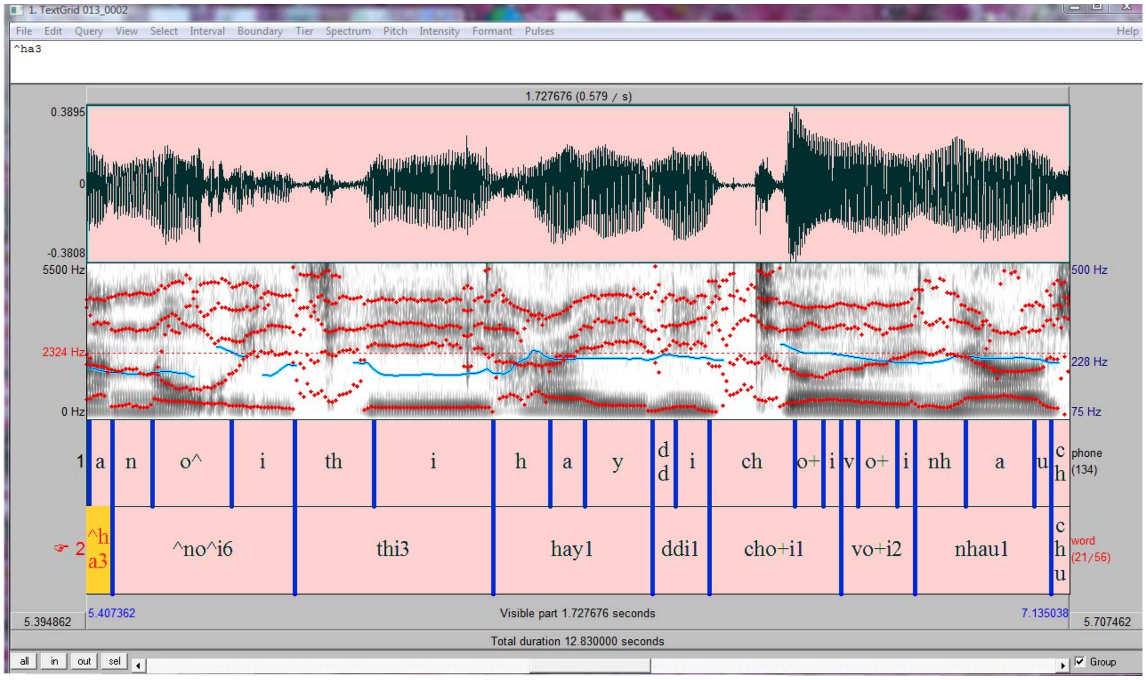


Figure 3.7: A screen shot of the .wav and .TextGrid file with correct boundaries

3.4.3. Acoustic measurements and normalizations

The total number of tokens for monophthongs, diphthongs and triphthongs extracted from this portion of the corpus is 25,233 tokens. Nearly half of these tokens, 11,310 tokens, are excluded from the study, leaving a total of 13,923 tokens of monophthongs and diphthongs for all three speech styles, as shown in table 3.4.

	Interviews	News	Framed sentences	Total
Monophthongs	4,744	3,153	1,137	9,033
VV clusters	529	396	438	1,360
Vw & Vj diphthongs	1,584	787	1,158	3,529
Total	6,856	4,336	2,733	13,925

Table 3.4: The total number of usable tokens from the speech data

The first reason for the exclusion of some of the 11,310 tokens is speaker noise, e.g. {laugh}. The words in the framed sentences other than the target word list are also excluded, i.e. all extra words in this carrier sentence *Tôi đọc từ X thêm ba lần nữa. X, X, X* ("I say the word X three more times. X, X, X"). In addition, some particle word classes are also excluded because of their interaction with the sentence intonation. This group of particles includes interjections, initial particles, final particles, and clause particles (Thompson 1965, Hạ 2010). They generally have a different intonation, duration, and intensity due to their functions, and they appear in much high frequency than other function and content words. Thus their exclusion would prevent the data results being distorted. These particles are listed with examples in table 3.5.

No.	Particles	Word – IPA	Examples and Gloss
1	Final	*ý* /i ˧˥/	*...với cả người ta cũng không thích ý.* ... and also people don't like that, yeah!
2	Final	*nhá* /ɲa ˧˥/	*... không được chê nhá.* ... can't talk bad about it, yeah!
3	Final	*nha* /ɲa ˧/	*... nên mệt nha.* ... thus tired, yeah!
4	Final	*á* /a ˧˥/	*... đặt là gửi thôi á.* ... ordered and we sent out, yeah!
5	Final	*à* /a ˧˩/	*... chứ thế à?* ... is that so?
6	Final	*ạ* /a ˧ʔ˩/	*Hai anh em ạ.* And older brother and a younger sibling.
7	Final	*ừ* /ɯ ˧˩/	*Ừ.* Yes.
8	Final	*ấy* /ʌj ˧˥/	*Khách người ta đến muộn ấy...* Customer who comes late, yeah...
9	Final	*đấy* /dʌj ˧˥/	*Em mới về làm nội thất ô tô đấy chứ.* I then work with car interior, yeah.
10	Initial	*thế* /tʰe ˧˥/	*... thế bắt đầu mới hiểu về ô tô.* ... then started to understand about car.
11	Final	*chứ* /cɯ ˧˥/	*Em mới về làm nội thất ô tô đấy chứ.* I then work with car interior, yeah.
12	Clause	*và* /va ˧˩/	*... cần lắp cái gì và không cần lắp cái gì.* ... need to install some and not to install some.
13	Clause	*thì* /tʰi ˧˩/	... vào mùa cưới thì em mới làm thôi. ... only in the wedding season I work.
14	Clause	*là* /la ˧˩/	*Anh trai bảo em là thôi.* My brother told me "no".
15	Clause	*mà* /ma ˧˩/	*... gần như là Trung Quốc hết mà chị.* ... mostly China, yeah, sister.
16	Clause	*nhưng* /ɲɯŋ ˧/	*... nhưng em với anh ý ...* ... but I and him, ...
17	Interjections	*vâng* /vʌŋ ˧/	*Vâng ạ.* Yes
18	Initial	*dạ* /za ˧ʔ˩/	*Dạ, ông ngoại em cũng cho ...* Yes, my grandpa also gives ....

Table 3.5: List of excluded particle (examples extracted from the speech of speaker 2)

To be able to learn about the differences in characteristics of vowels and glides, I measure formants at five points in each phone: the 10% point, 30% point, 50% point, 70% point and 90% point. The phone boundaries are used to measure the phone duration and word duration. A Praat script, written by Joshua Tauberer, is used to automatically extract duration values and to call up the Praat commands "Get formants" for measure formants.

Formant frequencies of all vowel tokens are important to this study and since there are six speakers in this data set – all female speakers of the same age range and of the same dialect – I normalize all formant frequency values to minimize the intra-speaker variations. The method I employed is the Lobanov normalization method using standard deviation (Lobanov 1971, Thomas and Kendall 2007).

Lobanov normalization: $F_{n[V]}^{N} = (F_{n[V]} - MEAN_n)/S_n$

Where $F_{n[V]}^{N}$ is the normalized value for $F_{n[V]}$ (i.e., for formant *n* of vowel *V*).

$MEAN_n$ is the mean value for formant *n* for the speaker in question

S_n is the standard deviation for the speaker's formant *n*.

Figure 3.8: Lobanov normalization

The mean value and the standard deviation for individual speaker's formants are derived from all the formant values measured at the 10%, 30%, 50%, 70% and 90% points of each vowel.

Speaker	Value	F1	F2	F3
Speaker 1	mean	601	1762	2900
	stdev	170	435	318
Speaker 2	mean	599	1767	2931
	stdev	196	398	299
Speaker 3	mean	622	1746	2894
	stdev	207	449	319
Speaker 4	mean	581	1675	2828
	stdev	183	469	282
Speaker 5	mean	557	1748	2726
	stdev	163	416	209
Speaker 6	mean	608	1766	2741
	stdev	184	391	271

Table 3.6: The means and standard deviations for all speakers and all formants

The normalized data is then scaled back into the Hertz values for ease of analysis. The scaling is based on scaling formulas introduce by Thomas and Kendall (2007), which use the Minimum and Maximum values of all speakers' normalized formant values.

Scaling formula:

$$F'_{i[V]} = F_{iMIN} + F_{iRANGE} *(F^N_{i[v]} - F^N_{iMIN}) / (F^N_{iMAX} - F^N_{iMIN})$$

Where F_{iMIN} is the minimum value for formant *i* of all vowels before normalization

F_{iRANGE} is the difference between the maximum and minimum values of formant *i* of all vowels and all speaker before normalization.

$F^N_{i[V]}$ is a normalized value for formant *i* of vowel V.

F^N_{iMIN} is the minimum normalized formant values for formant *i*.

F^N_{iMAX} is the maximum normalized formant values for formant *i*.

Figure 3.9: Scaling formula for formant values

Based on the minimum and maximum values in the speech data of this six Hanoian speakers, the actual scaling formulas for scaling the first formant, second formant and third formant are as follows, which is used to scale all tokens of monophthongs and diphthongs.

Scaling formula for speakers 1 to 6:

$$F'_{1[V]} = 140+1735*(F^N_{1[v]} -(- 2.824755))/(12.845244 - (-2.824755))$$

$$F'_{2[V]} = 450+2740*(F^N_{2[v]} -(- 3.278188))/(4.117761 - (-3.278188))$$

$$F'_{3[V]} = 1390+2860*(F^N_{3[v]} -(- 5.281911))/(10.089378 - (-5.281911))$$

where $F^N_{1[V]}$ is a normalized value for first formant of vowel V.

$F^N_{2[V]}$ is a normalized value for first formant of vowel V.

$F^N_{3[V]}$ *is* a normalized value for first formant of vowel V.

Figure 3.10: Scaling formula for speakers 1 to 6

Chapter 4

A phonetic study of the Vietnamese vowel system

4.1. Formant values of monophthongs, contour vowels and diphthongs

Previous studies, as mentioned in Gay (1968), show that the targets of diphthongs are not the same as those of monophthongs, e.g. the onset of /aj/ can vary from [a] to [æ] and the offset from [ɛ] to [i]. My hypothesis is that the components of the Vietnamese contour vowels have targets, measured at the 50% point for each component, that are similar to the targets of monophthongs.

4.1.1. Formant values and mean targets

4.1.1.1. Data and methods

To study the formant structures of eleven Vietnamese monophthongs, I use 9,034 tokens of monophthongs, which are tokens of 741 unique words, from the recordings of six female Hanoian speakers. These 9,034 tokens include 4,744 tokens in spontaneous speech, 3,153 tokens in news reading format, and 1,137 tokens in framed sentence format. Minimal pairs for the eleven monophthongs are listed in the table 4.1. The two short vowels /ɐ ʌ/ do not exist in open syllables, thus there are no tokens of CV syllables of these two vowels in the table 4.1. All of the tokens of monophthongs in the data are measured for formant frequencies at five points: 10%, 30%, 50%, 70%, and 90% of the vowel duration. The formant values are all normalized using the Lobanov method and then scaled using Thomas and Kendall (2007) method as described in chapter 3.

Monophthong	Minimal pair 1			Minimal pair 2		
	Word	IPA	Gloss	Word	IPA	Gloss
a	*ba*	/ɓa ˧/	three	*can*	/kan ˧/	stop
ɐ				*căn*	/kɐn ˧/	root
ɤ	*bờ*	/ɓɤ ˧˨/	shore	*tơn*	/tɤn ˧/	hurry
ʌ				*cân*	/kʌn ˧/	weight
i	*bí*	/ɓi ˧˥/	stuck	*tin*	/tin ˧/	believe
e	*bê*	/ɓe ˧/	calf	*bên*	/ɓen ˧/	side
ɛ	*bé*	/ɓɛ ˧˥/	small	*đen*	/ɗɛn ˧/	black
ɯ	*cứ*	/kɯ ˧˥/	continue	*tưng*	/tɯŋ ˧/	bouncy
ɔ	*bỏ*	/ɓɔ ˧˨˧/	abandon	*con*	/kɔn ˧/	kid
o	*bố*	/ɓo ˧˥/	father	*côn*	/kon ˧/	stick
u	*đủ*	/ɗu ˧˨˧/	enough	*cún*	/kun ˧˥/	puppy

Table 4.1: Minimal pairs of the eleven monophthongs
extracted from the spontaneous speech data

The first, and most traditional, method for describing the differences between monophthongs is to map vowels onto the vowel space using a single measure of first and second formant frequency (Peterson and Barney 1952). The third formant can be used as an additional measure to distinguish the vowels from one another. As the 50%

point is considered the most stable part of a vowel, free from the influence of the surrounding environment, the formant frequency values used for mapping the vowels in the traditional way are the measured values at the 50% point. The first and second formant mean values of these monophthongs are also taken as targets for later analysis of the behavior of glides.

As for the contour vowels and diphthongs, they are all annotated as containing two components in the data set, as described in chapter 2. For the contours, both components are considered vowel components, and for the diphthongs, the first component is a contour component and the second component is a glide component, based on my previous hypothesis. All of these components are also measured for first, second and third formants at the 10%, 30%, 50%, 70% and 90% points, and the measurements are normalized and scaled according to the methods described in chapter 2. To compare these components with the monophthongs, I use their first and second formant values at the 50% point to map their distribution onto the vowel space.

4.1.1.2. Results

The mean values of the first and second formants measured at the 50% points, as listed in table 4.2, show that nine of the eleven monophthongs, i.e. /i e ɛ ɯ ɤ ɔ o u ʌ/, are spaced apart from each other while the two vowels /a ɐ/ overlap.

No.	Monophthong	# of tokens	First formant	Second formant
1	a	1867	600	1693
2	ɐ	456	598	1618
3	ɤ	491	458	1455
4	ʌ	553	512	1650
5	i	1136	355	2193
6	e	673	409	2079
7	ɛ	659	486	2058
8	ɯ	641	381	1676
9	ɔ	1104	543	1372
10	o	814	442	1137
11	u	640	386	1166

Table 4.2: Mean formant values of the eleven Vietnamese monophthongs

The formant measurements at the 50% point of the vowel components in the contour vowels /ie/, /ɯɤ/ and /uo/ are mapped onto the vowel space with their mean values and the mean values of the corresponding monophthongs in figure 4.1. The mean values of the vowel components are very close to those of the monophthongs, especially in the case of /i/ and /ɤ/, for which the values are nearly identical. The distance between the means of the vowel components /ɯ/ and /u/ and the means of their corresponding monophthongs are also small, with a distance of 4 Hz in the first formant and 120 Hz in the second formant for /u/, and 14 Hz in the first formant and 120 Hz in the second formant for /ɯ/.

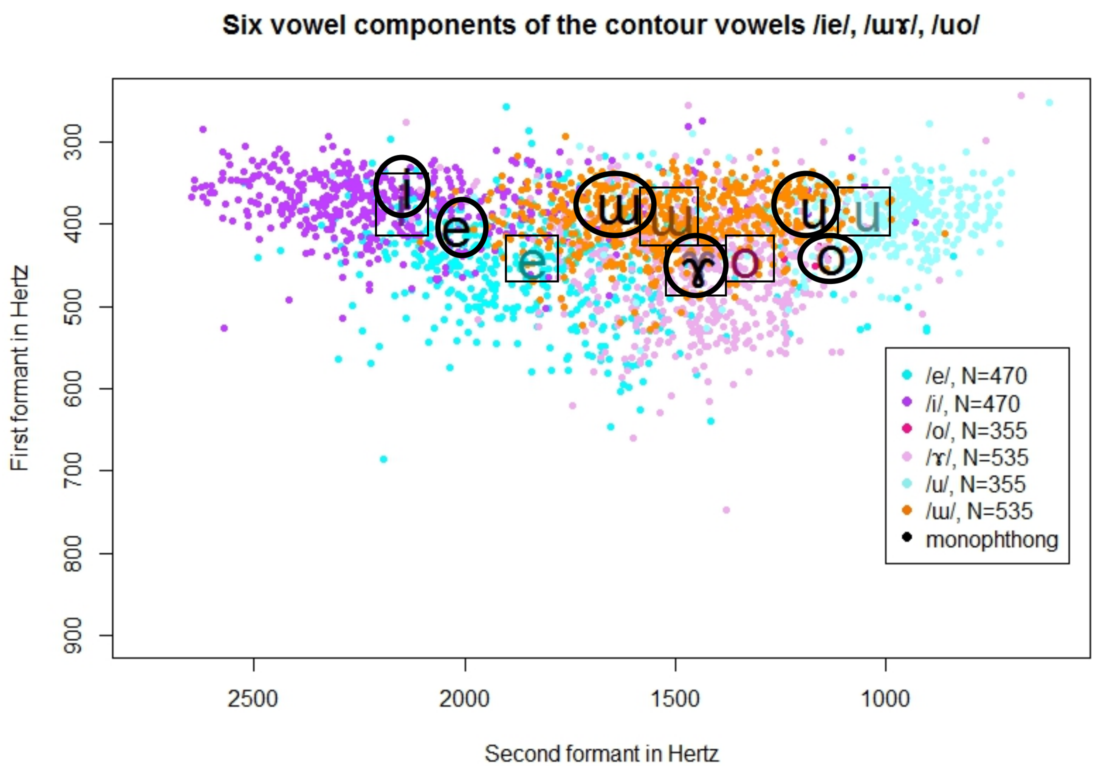


Figure 4.1: Formants of the vowel components in the contour vowels (in squares) and their corresponding monophthongs (in circles)

In both /j/-glide and /w/-glide diphthongs, the same pattern is true for the vowel components, having their mean formants values closer to those of their corresponding monophthongs.

The means of these vowel and glide components in different environments, i.e. components of contour vowels in open syllables, components of contour vowels in closed syllables, components of diphthongs, monophthongs in open syllables, and monophthongs in closed syllables, are mapped with their confidence intervals, which is the standard 95% range of confidence interval calculated in R, in figure 4.2. The means of the vowel components in each environment are mapped as dots, asterisks and squares in this graph while the letters mark the total mean values of each monophthong. The mean values for the glides /j/ and /w/ – the second components of the diphthongs – are marked with the upside-down triangles with their confidence intervals. The large ovals enclosing the means for each vowels are drawn in for ease of identification only and it does not represent any statistic significance. Figure 4.2 shows that for most environments, the mean formant values of all tokens are located around the mean formant values of the monophthongs. The means of the first vowel components of the contour vowels, i.e. the filled circles and the asterisks for /i/, /ɯ/, and /u/, are also closer to the means of other environments. The means of the second vowel components of the contour vowels in open syllables, i.e. the filled circles and the asterisks for /e/, /ɤ/ and /o/, are centralized. This is consistent with the process of dorsalization of the contour vowels in final syllable position, in which /e/, /ɤ/ and /o/ become more or less like /ʌ/. I will discuss the details of the dorsalization process further in chapter 5.

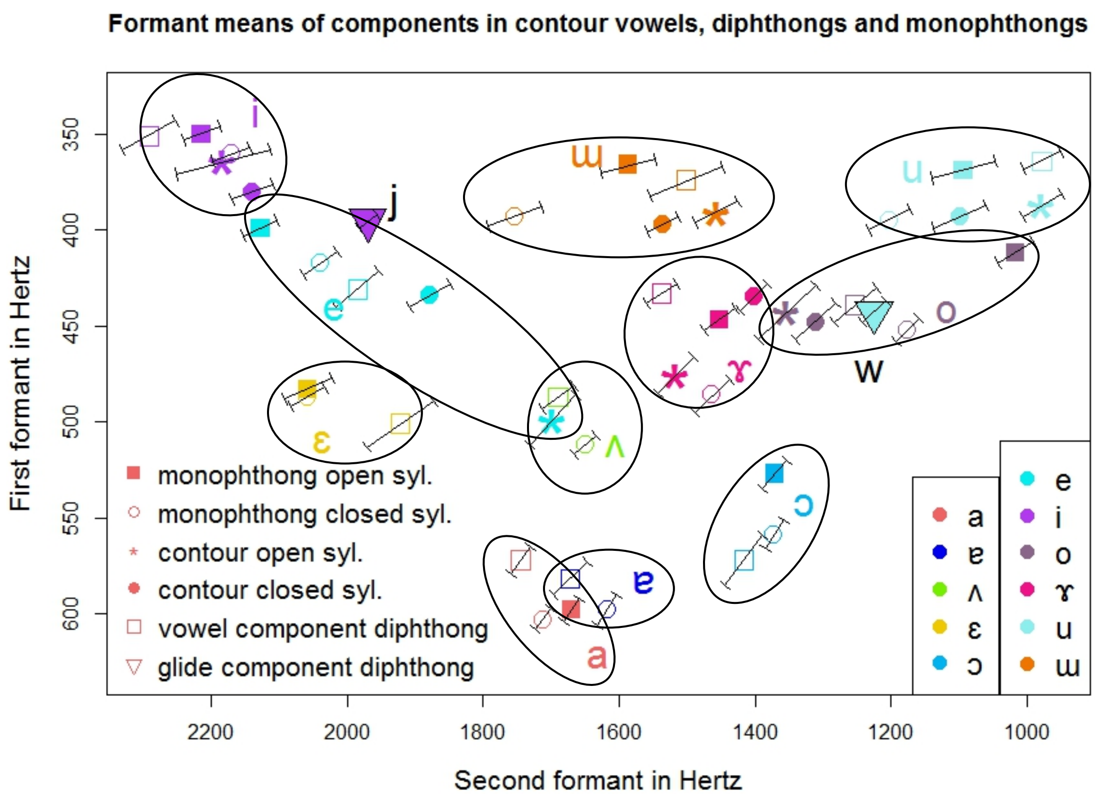


Figure 4.2: Mean formant values in different environment

The glides /j/ and /w/ have their means plotted separately from the mean values of /i/ and /u/. When I chart all of the tokens of monophthongal /i/ and /u/ against all the tokens of glide /j/ and /w/ from the diphthongs, as in figure 4.3, all the tokens of monophthongal /i/ and /u/, which are in red and blue with most them enclosed in the ovals for ease of identification, are distributed toward the two upper limits of the vowel space. The tokens of the gliding components in the diphthongs, however, are spread out across the vowel space, showing that the targets are not always reached as diphthong offglides.

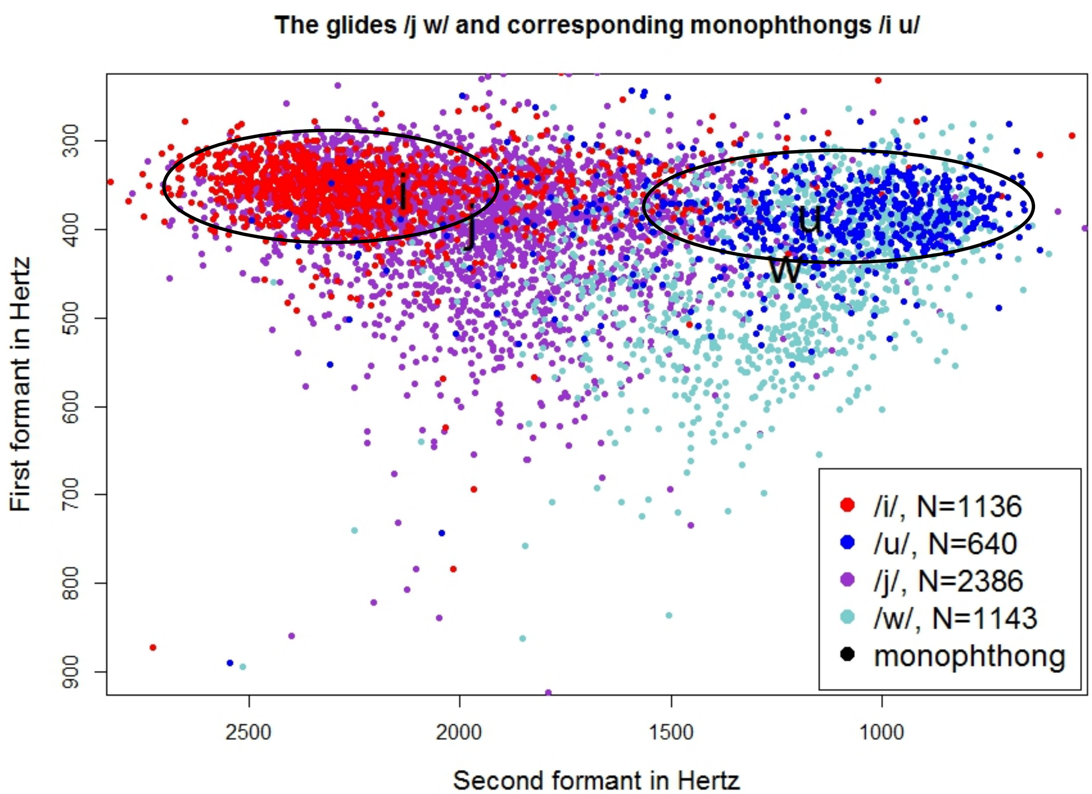


Figure 4.3: The glides /j w/ and their corresponding monophthongs /i u/

4.1.1.3. Discussion

The results show that the two components of the contour vowels, being measured at the 50% points for each component, and the vowel components of the diphthongs always reach their formant targets in both first and second formant frequencies. For the second vowel components of the contour vowels in open syllables, the targets are actually the central vowel [ʌ] – due to the process of dorsalization, which these tokens have reached. As for the vowel component of diphthongs, their mean values of the first and second formants are close to those of the monophthongs. But the second component – the glide component [j] or [w], however, does not always reach the same first and second formant frequency values as the monophthongs /i/ and /u/. The formant values for the glides vary greatly, thus the glide [j] ranges between an [ɛ] or an [ʌ] up to an [i], and the glide [w] ranges between an [ɔ] or an [ɤ] to an [u]. These results are consistent with the findings in previous studies on the formant differences between vowels and glides (Gay 1968, Collier 1982). The results show the difference between contour vowels and diphthongs, and they support the grouping of contour vowels and monophthongs into one natural class.

4.1.2. Formants and duration of the tense-lax vowel pairs

Among the eleven monophthongs, there are two pairs of vowels, i.e. /a ɐ/ and /ɤ ʌ/, that have been described as being different in duration (Thompson 1965, Han 1968). In some studies, the phones in these pairs have been described as single phonemes with /ɐ/ and /ʌ/ being the allophones of /a/ and /ɤ/, respectively (Smalley 1957, Le-Van-Ly 1960, Crothers 1978, Nguyễn Đình Hòa 1997), but this description does not fully explain the existence of all the minimal pairs. In other studies, they are described as different phonemes with similar quality but different duration (Thompson 1965, Han 1968). These two vowel pairs will be called the "tense-and-lax" pairs of vowels for the reasons I presented in chapter 2. In this section, I will be looking at the difference and similarity within the tense-lax vowel pairs in terms of formant values.

4.1.2.1. Data and methods

There are 1,867 tokens of /a/, 456 tokens of /ɐ/, 491 tokens of /ɤ/, and 553 tokens of /ʌ/ in the data set. These tense and lax vowels were measured at the 50% point for first, second and third formants, and the values were normalized using the methods described in chapter 3. The duration of the vowels were also measured and means were calculated.

4.1.2.2. Results

When the tense and lax monophthongs are mapped onto the vowel space using first and second formant values measured at the 50% point, the tokens of the pair /ɤ ʌ/ show different distribution in the vowel space. Their first and second formant means are 458Hz and 1443Hz for monophthong /ɤ/ and 512Hz and 1617Hz for monophthong /ʌ/.

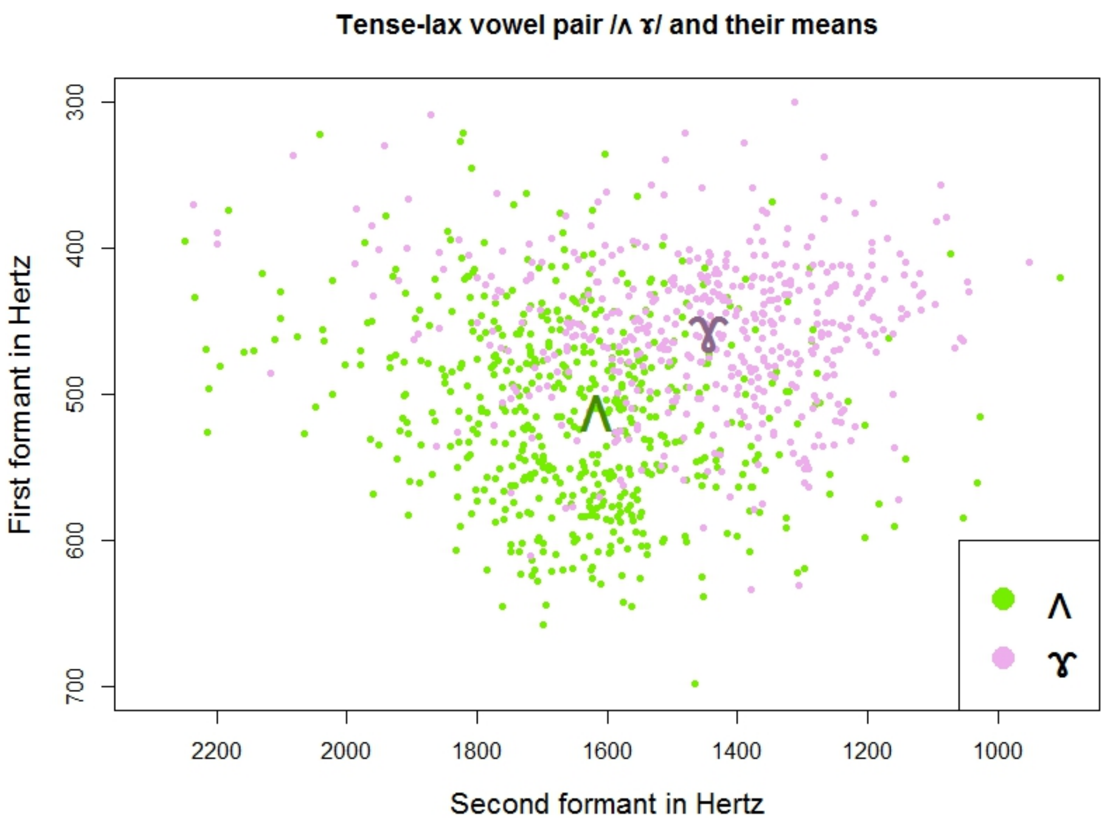


Figure 4.4: The vowel pair /ɤ ʌ/ and their means

The mean duration values of vowels /ɤ/ and /ʌ/ are different from one another by about 10ms in Spontaneous and News speech, and by 50ms in Word List speech, as shown in figure 4.5.

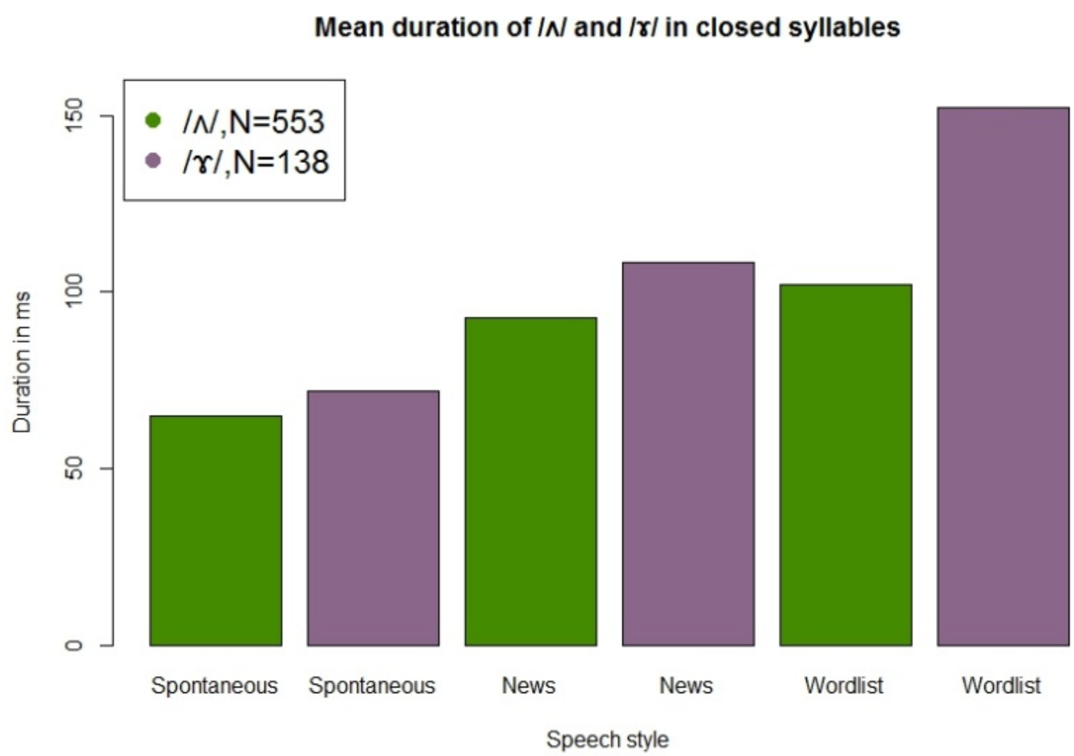


Figure 4.5: Mean duration of / ɤ / and / ʌ / in closed syllables

On the other hand, the formant values for the other tense-lax pair, i.e. /a ɐ/, overlap (figure 4.6), and their mean values are nearly the same, with a first formant mean of 600Hz and 598Hz, and a second formant mean of 1630Hz and 1578Hz for /a/ and /ɐ/ respectively.

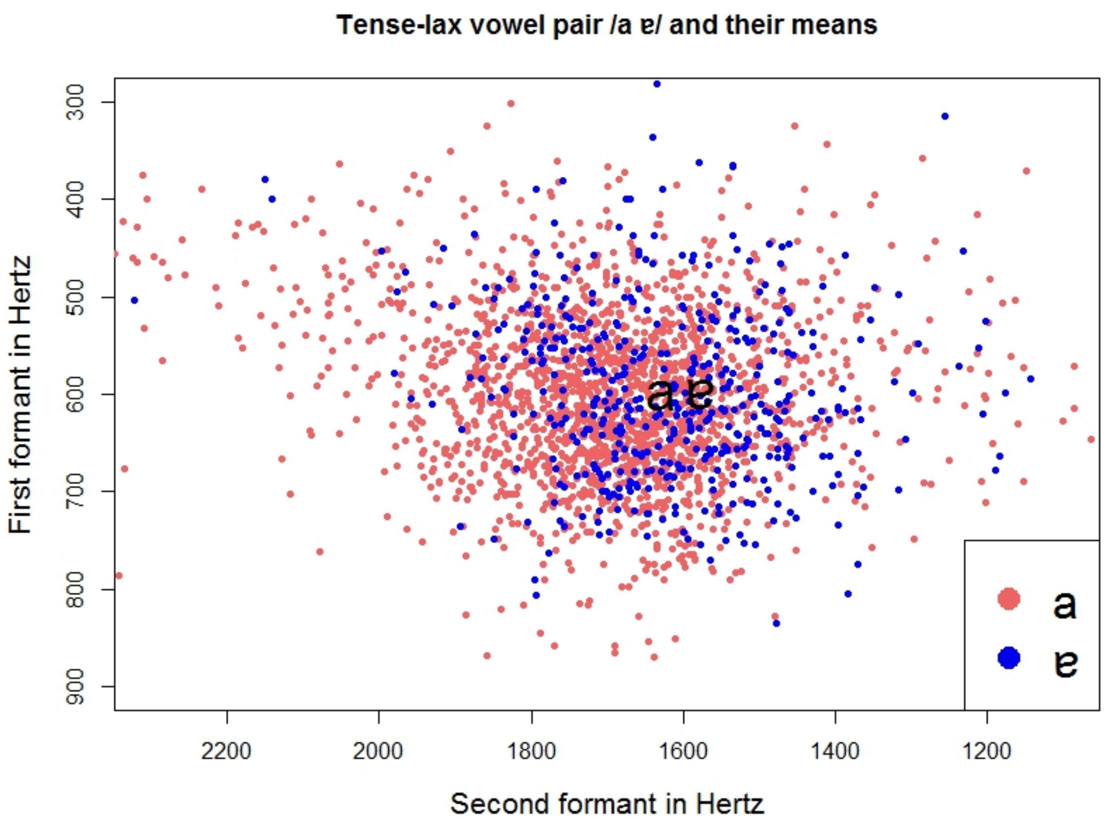


Figure 4.6: The vowel pair /a ɐ/ and their formant means

Figure 4.7 below shows that the means of all three formants of vowel /a/ is extremely similar to those of the vowel /ɐ/. The third formant of vowels is said to be important as a perception cue for surrounding consonants, r-coloring, and lip-rounding characteristics (Peterson and Barney 1952, Harris *et al* 1958). Much as the first and second formants of /a/ and /ɐ/ are not enough to separate them in the vowel space, the third formant also does not show much difference between the two vowels (figure 4.7). The third formant means are 2227Hz and 2245Hz for /a/ and /ɐ/ respectively.

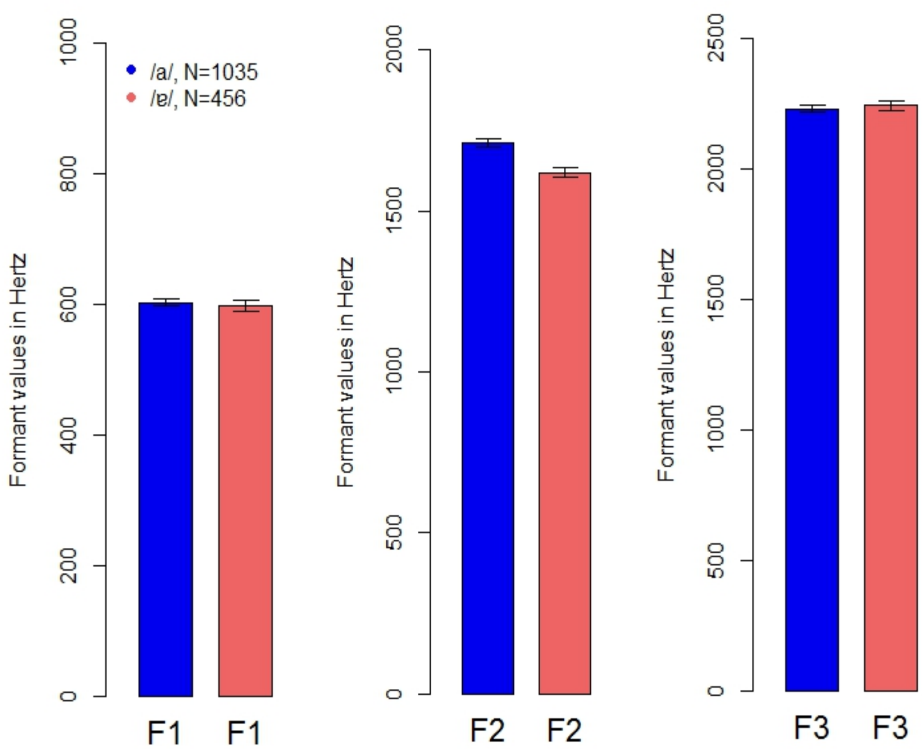


Figure 4.7: Mean formant values for vowel pair /a ɐ/

Note: The error bar is calculated using confidence interval of 95%

The distributions of the third formant frequencies are also overlapped completely with each other, as shown in figure 4.8.

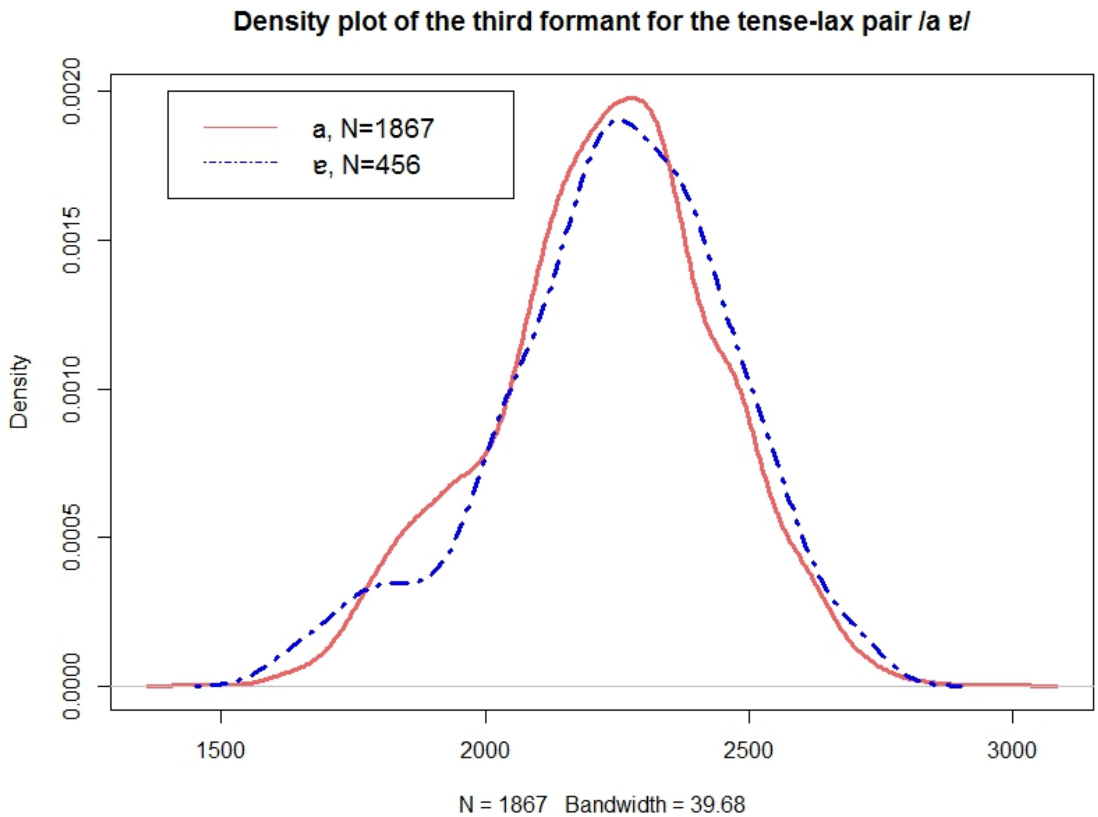


Figure 4.8: Distribution for the third formant of all tokens of /a ɐ/ in closed-syllables

As for the tense-lax pair of monophthongs /a/ and /ɐ/, the mean duration of /ɐ/ is shorter by 20ms to 30ms in spontaneous and news reading speech and by an average of 70ms in framed sentences. The difference in duration between /a/ and /ɐ/ is evident in all styles of speech, as demonstrated in figure 4.9, even though the amount of difference varies.

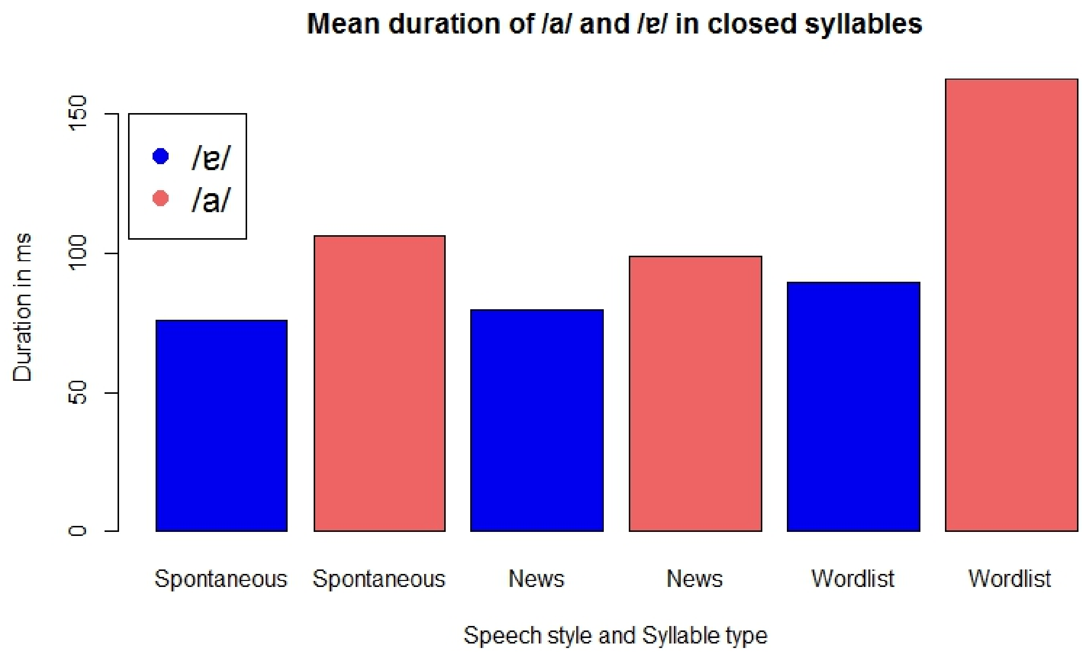


Figure 4.9: Mean duration of all token of /a/ and /ɐ/ in closed syllables

	Spontaneous speech		News reading		Word list	
	/ɐ/	/a/	/ɐ/	/a/	/ɐ/	/a/
# of tokens	228	485	160	497	68	53
Duration in ms	76	106	79	99	90	162

Table 4.3: All mean duration values of /a/ and /ɐ/ in closed syllables

The density plots for the duration of /a/ and /ɐ/ in figure 4.10 show that even though there is a large overlapping area in the density plot, the duration of /ɐ/ is more concentrated between 50ms and 100ms, while the duration of /a/ extends from 50ms to 150ms or even more. The actual mean value of all tokens of /a/ in closed syllables is 106ms and mean value of all tokens of /ɐ/ in closed syllables is 79ms with a p-value < 2.2e-16 based on t-test.

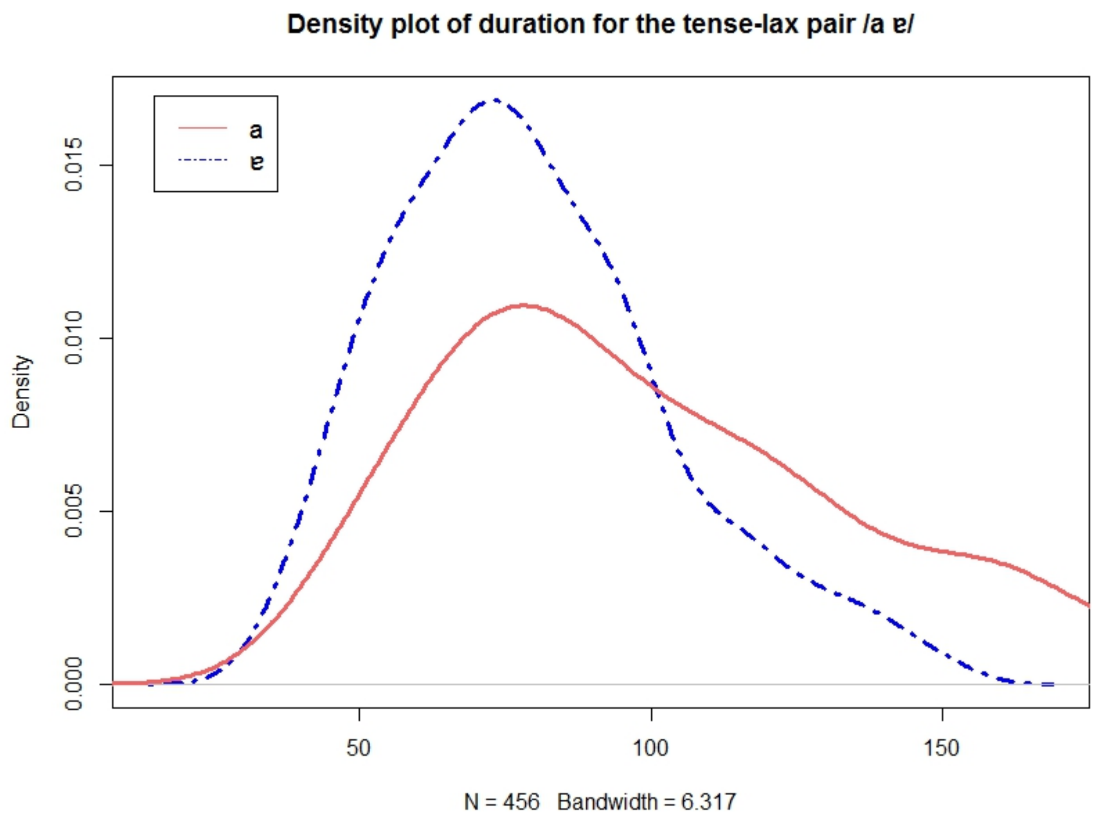


Figure 4.10: Density plot of all duration values of /a/ and /ɐ/ in closed syllables

There are eleven minimal pairs for the tense-lax vowel pair /a ɐ/, which were extracted from the data set with different speech styles from different speakers. While some of these minimal pairs appear in the speech of more than one speaker, only one set of tokens was picked for each of the minimal pairs to demonstrate the difference in duration. These minimal pairs show the same pattern of vowel /a/ being longer than vowel /ɐ/ but the difference in duration is not always the same. The minimal pairs in figure 4.11 are sorted by increasing difference between the token of /ɐ/ and the token of /a/ in each pair.

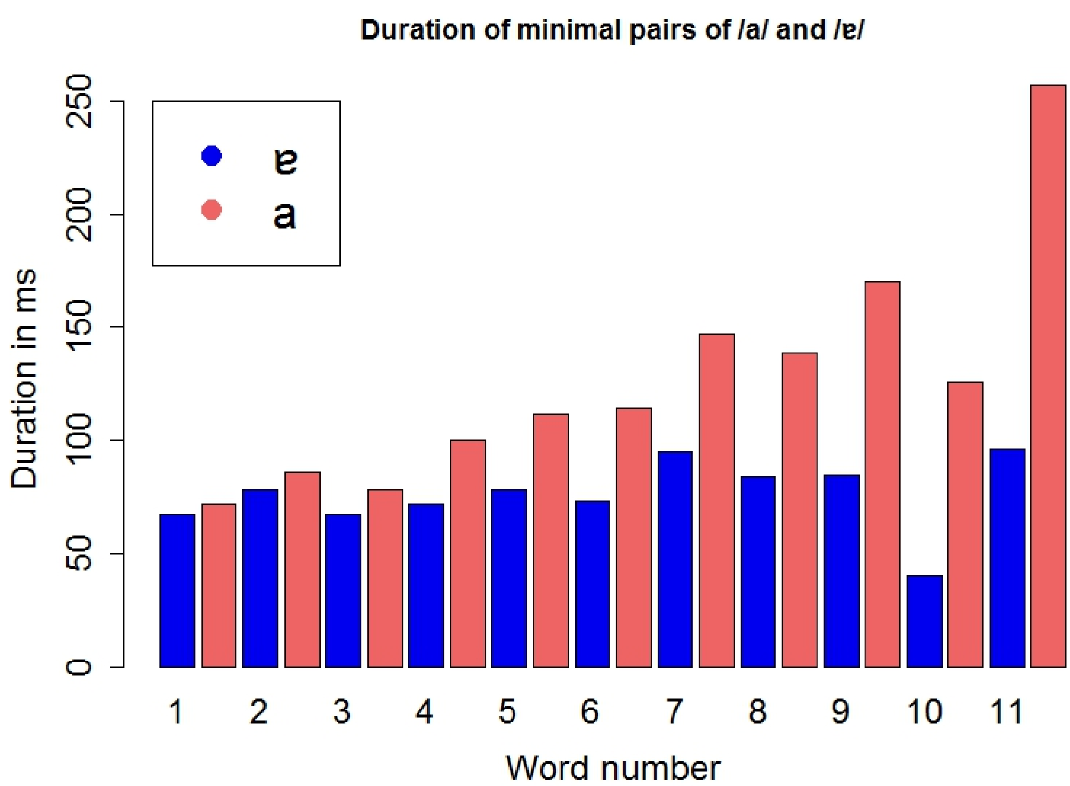


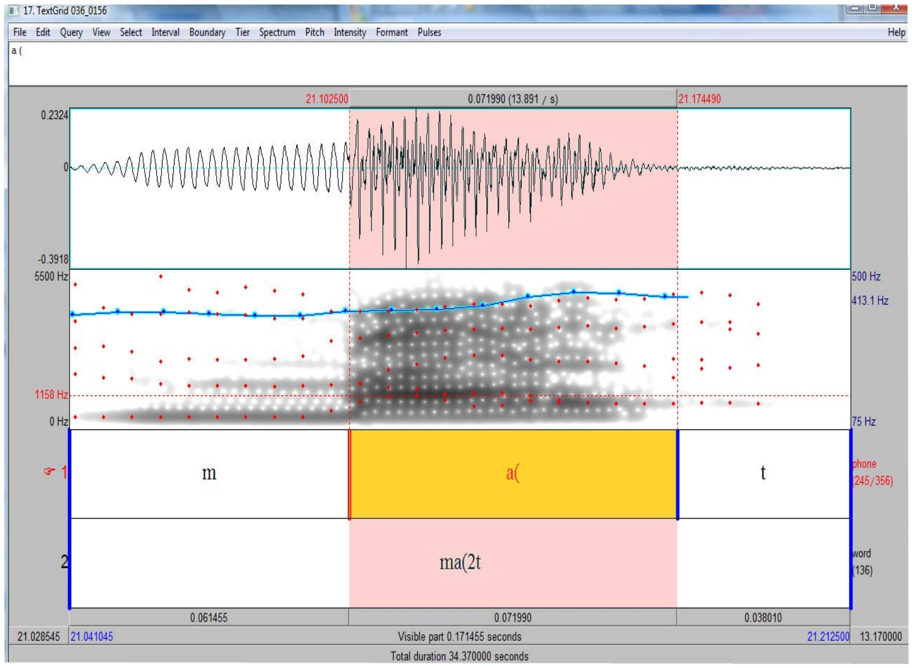
Figure 4.11: Duration of minimal pairs of /a/ and /ɐ/

Note: The words are listed by numbers in table 4.4 below

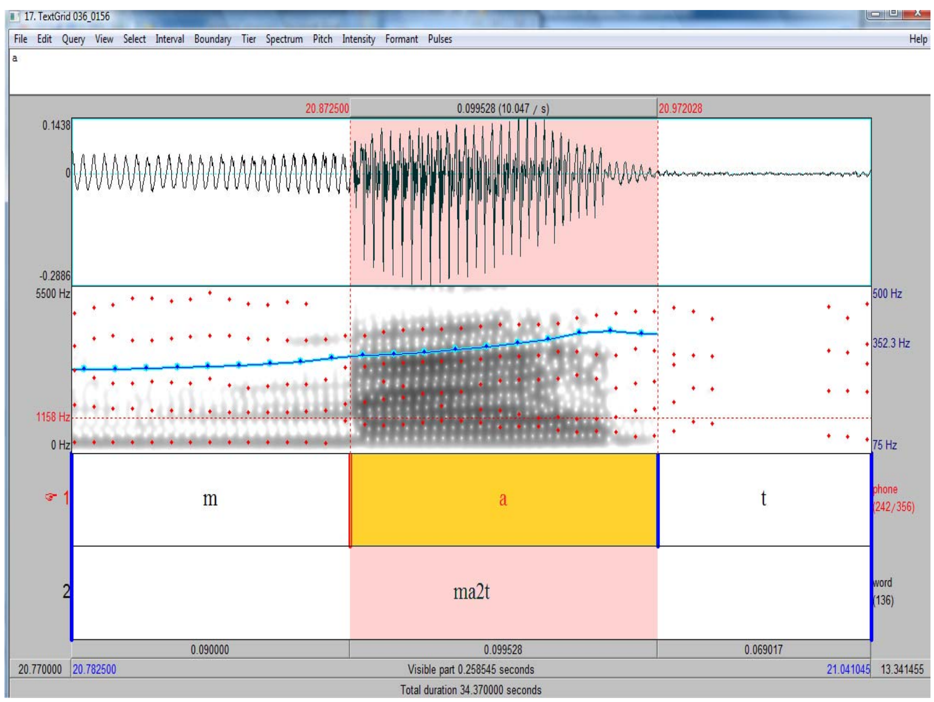
No.	Speech style	/a/ word	IPA	Gloss	Speech style	/ɐ/ word	IPA	Gloss
1	news	*đạc*	/ɗak˧ʔ˩/	stuff	news	*đặc*	/ɗɐk ˧ʔ˩/	condense
2	interview	*đang*	/ɗaŋ ˧/	continue	interview	*đăng*	/ɗɐŋ ˧/	post
3	news	*cát*	/kat ˧˥/	sand	interview	*cắt*	/kɐt ˧˥/	cut
4	news	*mát*	/mat ˧˥/	cool	news	*mắt*	/mɐt ˧˥/	eye
5	news	*nang*	/naŋ ˧/	pocket	interview	*năng*	/nɐŋ ˧/	active
6	news	*lạng*	/laŋ ˧ʔ˩/	100mg	news	*lặng*	/lɐŋ ˧ʔ˩/	quiet
7	interview	*bác*	/ɓak ˧˥/	uncle	news	*bắc*	/ɓɐk ˧˥/	north
8	word list	*tát*	/tat ˧˥/	slap	word list	*tắt*	/tɐt ˧˥/	turn off
9	interview	*an*	/an ˧/	calm	news	*ăn*	/ɐn ˧/	eat
10	interview	*nam*	/nam ˧/	south	interview	*năm*	/nɐm ˧/	year
11	interview	*đạt*	/ɗat ˧ʔ˩/	achieve	news	*đặt*	/ɗɐt ˧ʔ˩/	book

Table 4.4: List of minimal pairs in figure 4.11

Figures 4.12 and 4.13 show two minimal pairs of /ɐ/ and /a/. In the first pair (figure 4.12), the two monophthongs /ɐ/ and /a/ have a duration of 72ms and 100ms while the total word duration of *mắt* and *mát* are 171ms and 259ms respectively. In the second example (figure 4.13), the word pair *năm* and *nam* have similar durations of 200ms and 210ms, but the phone durations of /ɐ/ and /a/ are 40ms and 126ms. In both examples, the formants of monophthongs /ɐ/ and /a/ are quite similar to one another regardless of their duration.

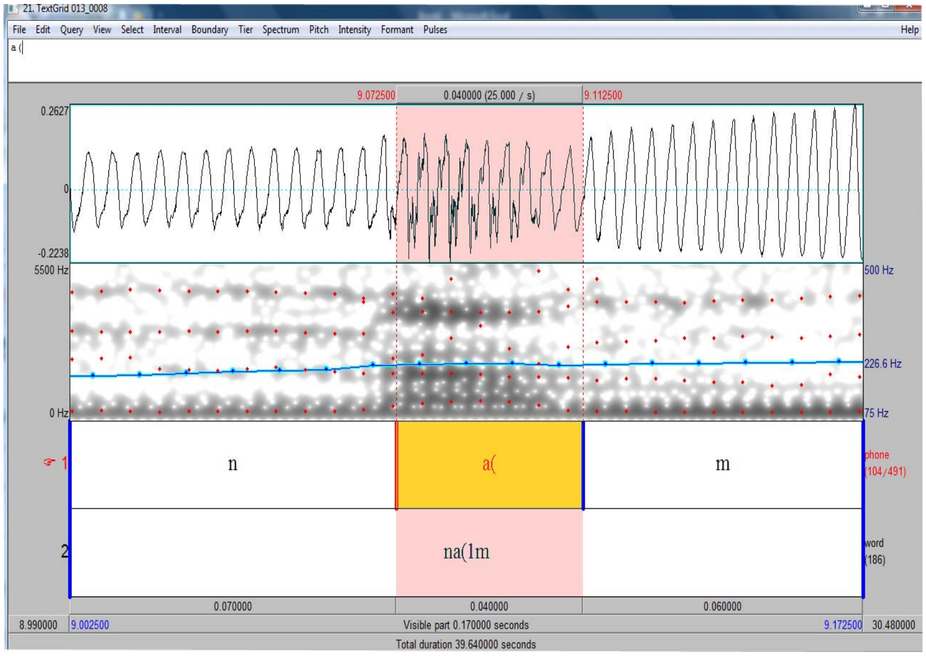


word *mắt* /mɐt ˧˥/ "eye" word 171ms, phone 72ms

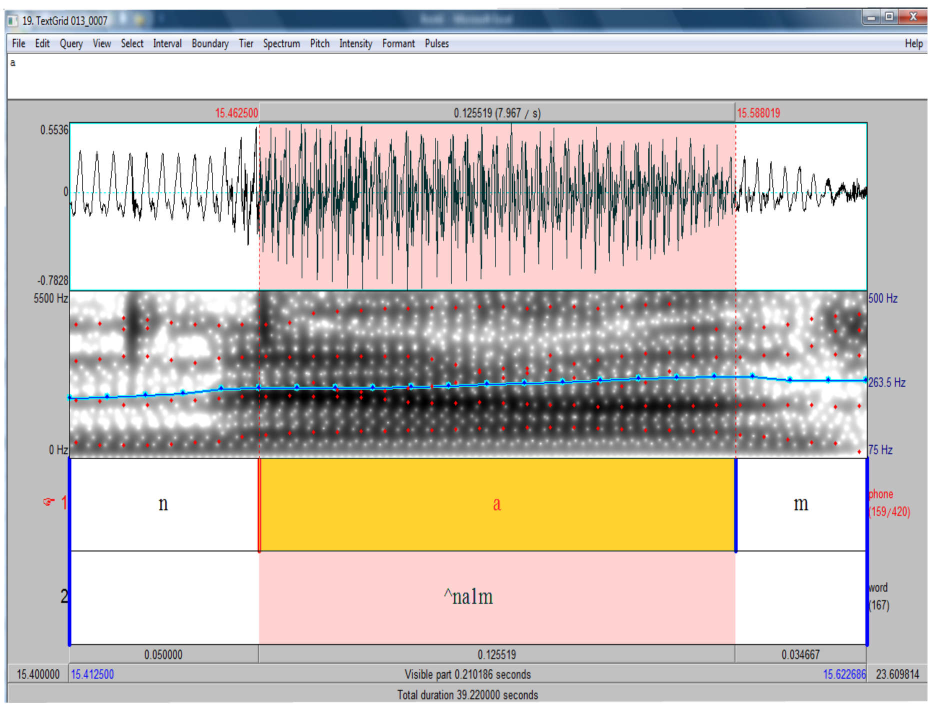


word *mát* /mat ˧˥/ "cool" word 259ms, phone 100ms

Figure 4.12: A minimal pair of /a/ and /ɐ/ and their duration (extracted from the news reading portion of speaker 3)

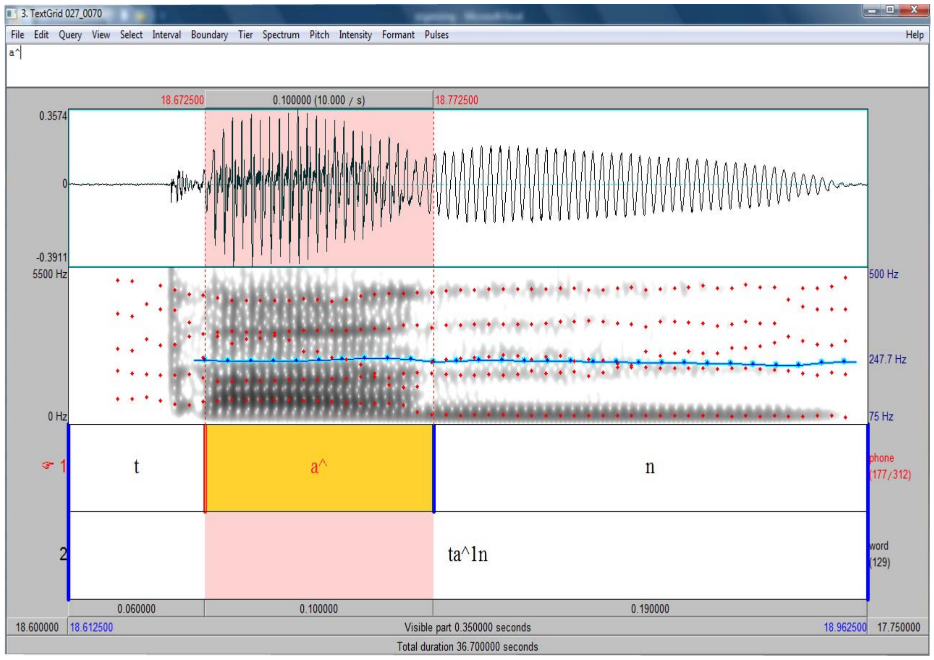


word *năm* /nɐm ˧/ "year" word 200ms, phone 40ms

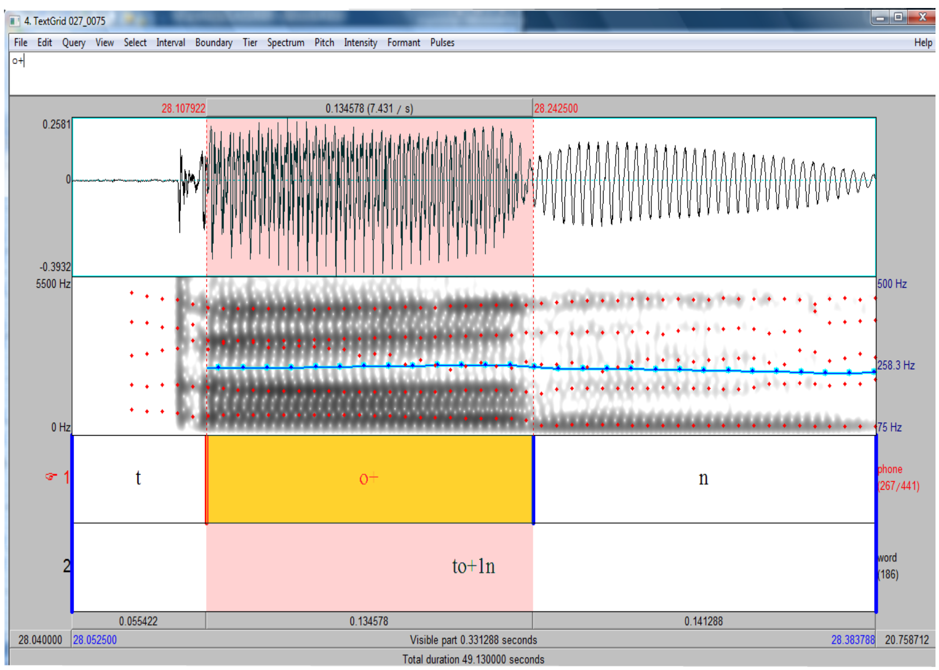


word *nam* /nam ˧/ "south" word 210ms, phone 126ms

Figure 4.13: Another minimal pair of /a/ and /ɐ/ and their duration (extracted from the spontaneous speech portion of speaker 1)



word *tân* /tʌn/ "new" word 350ms, phone 100ms



word *tơn* /tɤn / "hurry" word 331ms, phone 135ms

Figure 4.14: A minimal pair of /ɤ/ and /ʌ/ and their duration (extracted from the framed sentence portion of speaker 2)

4.1.2.3. Discussion

Among the eleven monophthongs of Vietnamese, nine vowels, i.e. /i e ɛ ɤ ɯ ʌ ɔ o u/, are distinguishable by formant values measured at the steady state in the vowel space of Hanoi dialect speakers. The other two vowels /a/ and /ɐ/, however, are not distinguishable by first, second or third formants. These two vowels have minimal pairs and thus cannot be grouped into a single phoneme, as some previous studies have argued. They are also described in previous studies as two phonemes with similar formant frequency values but differing in duration (Thompson 1965, Han 1966). While the vowel pair /a ɐ/ is not the only pair of Vietnamese vowels that exhibits a difference in duration, as the vowel pair /ɤ ʌ/ does as well, the latter pair shows only a mean difference of 10ms between the vowels rather than the 20ms to 30ms difference exhibited by /a ɐ/. In the next section of this chapter, I will demonstrate that duration is important in distinguishing the tense monophthongs /a/ and /ɤ/ from the lax monophthongs /ɐ/ and /ʌ/, especially for the tense-lax pair /a ɐ/.

4.1.3. Perception

There are two pairs of tense-lax monophthongs in Vietnamese, i.e. /a ɐ/ and /ɤ ʌ/, and they are all said to have audible differences in duration (Thompson 1965, Han 1966). In previous parts of this study, I have shown that the tense-lax pair /ɤ ʌ/ can be separated in the vowel space but the tense-lax pair /a ɐ/ cannot because they have similar formant values. I also showed that duration is a key difference between the monophthongs of the tense-lax pair /a ɐ/. To actually test the role of duration in the perception, I did a perception test for these four monophthongs.

4.1.3.1. Data and methods

The perception test includes an Identification test and a Discrimination test for the tense-lax pairs of monophthongs /a ɐ/ and /ɤ ʌ/. A minimal pair for the tense-lax pair /a ɐ/ and a minimal pair for the tense-lax pair /ɤ ʌ/ contain natural tokens of these four monophthongs, taken from previous framed-sentence style recordings. The frequencies of these four words are low based on a corpus of 1,055,617 Vietnamese words (Pham *et al* 2008), and their frequencies in percentage and total number of tokens in that corpus are listed in table 4.5. The percentage values are rounded up to the nearest ten thousandth. There are three other minimal pairs that are presented to the testing subjects in separate parts of this perception test, but their results are excluded from this study because these are not natural tokens from previous recordings.

No.	Word	IPA	Gloss	Number of tokens	Percentage
1	*tát*	/tat ˦˥/	slap	17	0.00161%
2	*tắt*	/tɐt ˦˥/	turn off	77	0.00729%
3	*tơn*	/tɤn ˦/	hurry	1	0.00001%
4	*tân*	/tʌn ˦/	new	294	0.02785%

Table 4.5: The minimal pairs in perception test

The subjects for the perception test are chosen among the Vietnamese students studying in Philadelphia, who were born and raised in Vietnam and came to Philadelphia for higher education. These subjects are of the same age group as the recording speakers and there are a total of eight Hanoi dialect test subjects and six Saigon dialect test subjects, half of each are female, and half male. The perception tests were all administered in my office and each subject listened to the test file on a computer with a personal headphone. The test file is one long .wav file, which has the author's voice reading the direction and the line number and the testing tokens.

The Identification task, as shown in the sampled script below, has three parts dedicated to the first pair of tense-lax /a ɐ/, and the other three – the second pair of tense-lax /ɤ ʌ/. All these parts contain unmodified tokens extracted from previous recordings. Each part has a total of 36 testing lines. In this task, each line of the answer sheet contains two written words. The test subjects would hear one spoken token for each line, and they are asked to identify the corresponding word on the answer sheet. The Discrimination task also has two parts and each part is dedicated to one pair of the minimal pairs with a total of 40 testing lines. In this task, each line in the answer sheet has two choices: *cùng một từ* ["same word"] and *hai từ khác nhau* ["two different words"]. The subjects would hear two spoken tokens for each line, and they are asked to decide if those are two tokens of the same word or of different word.

Sample script and answer sheet of part 4: {SP} stands for {silence}

... một {SP} *tắt* {SP} *hai* {SP} *tắt* {SP} *ba* {SP} *tát ...*

...[mot ˧ʔ˩] {SP} [tɐt ˧˥] {SP} [haj ˧] {SP} [tɐt ˧˥] {SP} [ɓa ˧] {SP} [tat ˧˥]...

'... one {SP} *tắt* {SP} two {SP} *tắt* {SP} three {SP} *tát*...'

Part 4: Please circle what you hear

1	tát	tắt	19	tát	tắt
2	tát	tắt	20	tát	tắt
3	tát	tắt	21	tát	tắt
4	tát	tắt	22	tát	tắt
5	tát	tắt	23	tát	tắt

Sample script and answer sheet of part 5: {SP} stands for {silence}

... một {SP} *tát tắt* {SP} *hai* {SP} *tắt tắt* {SP} *ba* {SP} *tắt tắt* ....

..[mot ˧ʔ˩] {SP} [tat ˧˥] [tɐt ˧˥] {SP} [haj ˧] {SP} [tɐt ˧˥] [tɐt ˧˥] {SP} [ɓa ˧] {SP} [tɐt ˧˥] [tɐt ˧˥]..

'... one {SP} *tát tắt* {SP} two {SP} *tắt tắt* {SP} three {SP} *tắt tắt*...'

Part 5: Please circle your answer

1	cùng một từ	hai từ khác nhau	21	cùng một từ	hai từ khác nhau
2	cùng một từ	hai từ khác nhau	22	cùng một từ	hai từ khác nhau
3	cùng một từ	hai từ khác nhau	23	cùng một từ	hai từ khác nhau
4	cùng một từ	hai từ khác nhau	24	cùng một từ	hai từ khác nhau
5	cùng một từ	hai từ khác nhau	25	cùng một từ	hai từ khác nhau

Note: *cùng một từ* means 'same word'; *hai từ khác nhau* means 'two different words'

Figure 4.15: Sample scripts and answer sheets for the perception test

4.1.3.2. Results

For Hanoi dialect test subjects, the tense vowel /a/ is perceived as the lax vowel /ɐ/ 30% of the time in the Identification test. The tense vowel /a/ is perceived as the lax vowel /ɐ/ 23% of the time by the Saigon dialect test subjects. Meanwhile, the lax vowel /ɐ/ is rarely mistaken for the tense vowel /a/, as Hanoi dialect test subjects and Saigon dialect test subjects only made this mistake 5% and 11% of the time, respectively. The subjects' one-direction confusion between the tokens of the tense-lax pair /a ɐ/, shown in the results of the Identification test, is confirmed by their error rate of 20% in the Discrimination test for this pair of vowels.

The results for the other tense-lax pair /ɤ ʌ/ follow a similar pattern. The tense vowel /ɤ/ is mistaken for the lax vowel /ʌ/ 27% and 39% of the time for Hanoi dialect test subjects and Saigon dialect test subjects, respectively. The misperception of the lax vowel /ʌ/ as the tense vowel /ɤ/ was also low: 6% among Hanoi dialect test subjects and 8% among Saigon dialect test subjects. This error pattern is confirmed by the Discrimination test result, in which error rate for the Hanoi dialect test subject is 20% and Saigon subjects – 5% higher than that of the Hanoi subjects.

The higher error rate in the Identification test among Saigon dialect test subjects might suggest that the vowels in these pairs are indistinguishable for Saigon dialect test subjects, thus their error rate is closer to that of a second language identification task than the error rate among Hanoi dialect test subjects, especially for the vowel pair /ɤ ʌ/. However, the direction of the misidentification is one-way, and the results from the Discrimination test for both pairs /a ɐ/ and /ɤ ʌ/ are confirming the subjects' confusion.

	Vowel pair /a ɐ/	Vowel pair /ɤ ʌ/
Hanoi subjects	20%	20%
Saigon subjects	20%	25%

Table 4.6: Percentage of errors in the Discrimination test

4.1.3.3. Discussion

The Discrimination test results for both pairs of vowels show that subjects can distinguish the vowels in each pair but without high accuracy rate. As the vowels in each pair have similar formant frequency values, the primary perception cue for the test subjects to rely on is duration, an attribute that I am using the features "tense" and "lax" to capture. The vowels with the longer duration in the pairs, i.e. the tense /a/ and /ɤ/, are mistaken for their shorter lax counterparts more than 25% of the time, while the lax counterparts are usually not mistaken for the longer tense vowels. As the tokens are presented to the subjects in isolation, and these tokens are all natural tokens, the explanation for the misperception of the longer tense monophthongs as the shorter lax monophthongs is that tense monophthongs have a duration threshold and they are distinguishable only if this threshold is achieved. And as these monosyllable tokens were presented to the subjects in isolation, this threshold is unclear to the subject due to lack of reference points and due to the effect of being in isolated monosyllables, thus subjects lose the key feature to identify the tense vowels. This perception test further strengthens the role of duration in monophthongs and it suggests that duration is an intrinsic feature of Vietnamese tense monophthongs in addition to their formant values. This duration is an intrinsic feature, which means it is an essential and natural property of the vowels, regardless of the surrounding influence.

4.2. Dynamic trajectories of monophthongs, contour vowels and diphthongs

4.2.1. Data and methods

Diphthongs are considered to have two components in their basic structure: a vowel and a glide. There are a total of fifteen vowel-glide diphthongs in Vietnamese, eight of which are vowel-/j/-glide diphthongs, and the other seven are vowel-/w/-glide diphthongs. All of these vowel-glide diphthongs only appear in open syllables.

No.	Diphthong	Example	IPA	Gloss	# of tokens
1	aj	*tai*	/taj ˧/	ear	596
2	ʌj	*tây*	/tʌj ˧/	west	307
3	ɐj	*tay*	/tɐj ˧/	hand	247
4	oj	*tôi*	/toj ˧/	me, I	364
5	ɤj	*tơi*	/tɤj ˧/	loose	391
6	ɔj	*toi*	/tɔj ˧/	die	238
7	ɯj	*gửi*	/ɣɯj ˧˩˧/	send	72
8	uj	*chui*	/cuj ˧/	go under	171
9	ʌw	*tâu*	/tʌw ˧/	report	171
10	aw	*tao*	/taw ˧/	me, I	391
11	ɐw	*đau*	/ɗɐw ˧/	hurt	189
12	ew	*têu*	/tew ˧/	ridicule	105
13	ɛw	*teo*	/tɛw ˧/	shrink	97
14	iw	*chịu*	/ciw ˧˩˧/	agree	71
15	ɯw	*hưu*	/hɯw ˧/	retire	119

Table 4.7: Fifteen Vietnamese vowel-glide diphthongs
(total 3,549 tokens of diphthongs)

There are three contour vowels in Vietnamese (Thompson 1965): /ie/, /ɯɤ/ and /uo/. They can appear in both open and closed syllables. In this data set, there are a total of 529 tokens of contour vowels in interview style, 396 tokens of vowels clusters in news reading style, and 438 tokens of contour vowels in framed sentence style.

No.	Diphthong	Example	IPA	Gloss	# of tokens
1	ie	*tiên* *tia*	/tien ˧/ /tie ˧/	angel ray	470
2	ɯɤ	*tươm* *tưa*	/tɯɤm ˧/ /tɯɤ ˧/	dressed fray	535
3	uo	*tuôn* *tua*	/tuon ˧/ /tuo ˧/	flow tuft	355

Table 4.8: Three Vietnamese contour vowels
(total 1,360 tokens of contour vowels)

Besides the 9,034 tokens of monophthongs, there are a total of 3,549 tokens of vowel-glide diphthongs in my data set with examples for all fifteen diphthongs. These diphthongs are annotated as two vowels next to each other, for example, diphthong /aj/ is marked as /a/ and /i/ in the phone tier. The choice of boundary location between these sounds is explained in chapter 2. For each sound in the diphthongs, formant frequencies are measured at five points: 10%, 30%, 50%, 70%, and 90% of the duration. Thus, there are ten measuring points for the entire diphthong. All the measurements are normalized and scaled, as explained in chapter 2 as well.

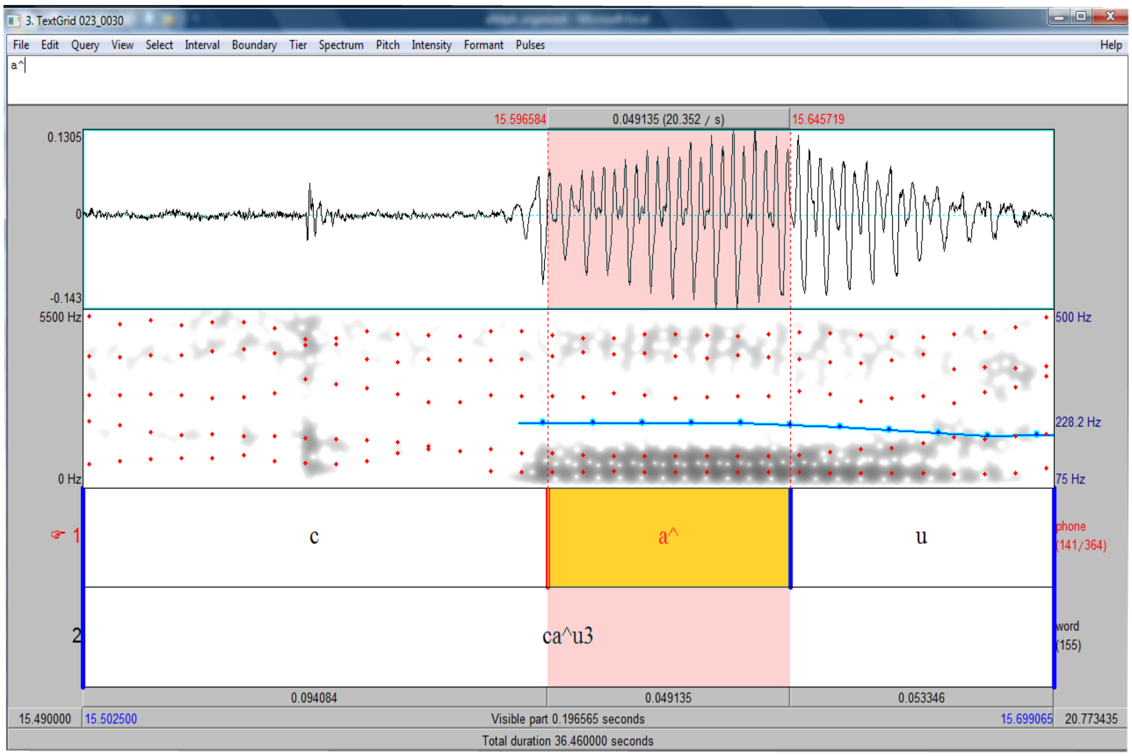


Figure 4.16: An example of /ʌw/ in *cầu* /kʌw ˧˩/ "bridge" of speaker 2

An example of /ʌw/	Formant	10%	30%	50%	70%	90%
Segment /ʌ/	1st formant	374	378	375	371	370
	2nd formant	954	907	893	893	907
Segment /w/	1st formant	363	356	356	337	372
	2nd formant	962	1097	1254	1404	1492

Table 4.9: Formant values at five points for each part of
the diphthong [ʌw] in the word *cầu* /kʌw ˧˩/ "bridge" of speaker 2
Note: These values are normalized and scaled

The contour vowels are also annotated as two vowels next to each other, for example, contour vowel /ie/ is marked as /i/ and /e/ in the phone tier. The choice of boundary location between these sounds is also explained in chapter 2. For each sound in the clusters, formant frequencies are measured at five points: 10%, 30%, 50%, 70%, and 90%. Thus there are ten measuring points for the entire cluster. All the measurements are normalized and scaled, as explained in chapter 2 as well.

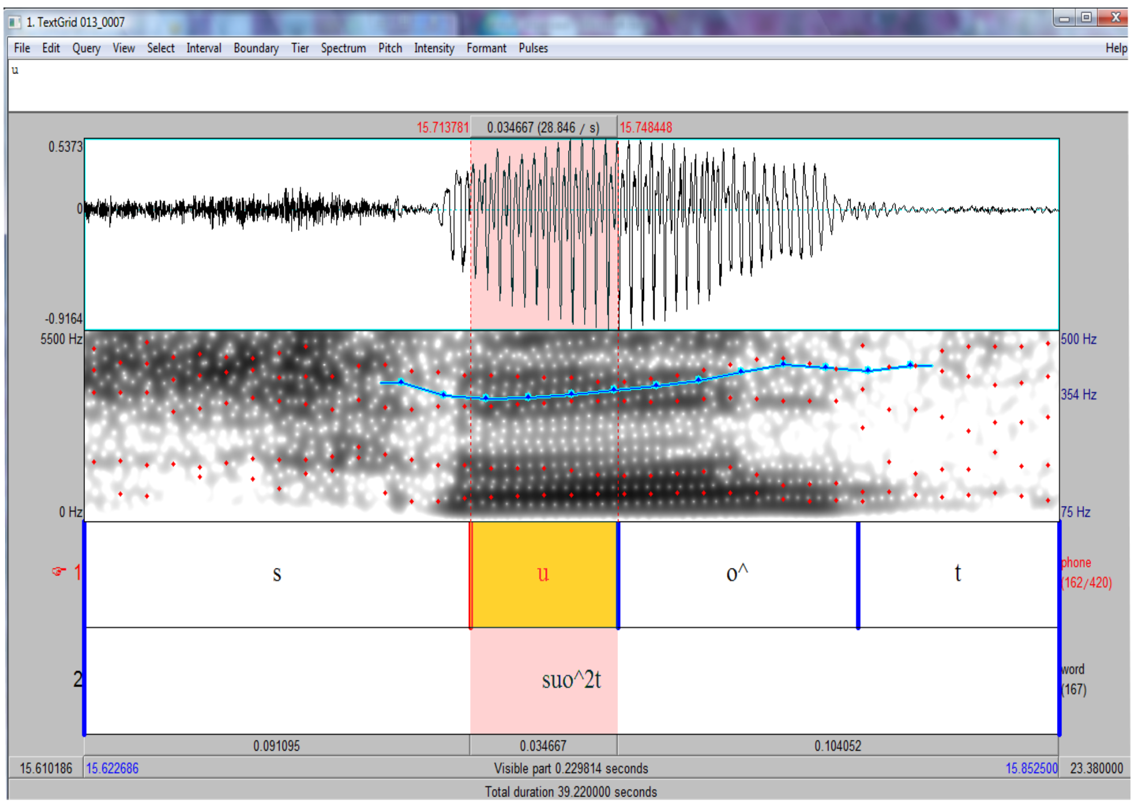


Figure 4.17: An example of /uo/ in *suốt* /suot ˧˥/ "all the way" of speaker 1

An example of /uo/	Formant	10%	30%	50%	70%	90%
Part /u/	1st formant	494	509	527	544	548
	2nd formant	1465	1432	1425	1374	1276
Part /o/	1st formant	1256	1343	1420	1031	1067
	2nd formant	962	1097	1254	1404	1492

Table 4.10: Formant values at five points for each part of
the contour vowel /uo/ in the word *suốt* /suot ˧˥/ "all the way" of speaker 1
Note: These values are normalized and scaled

In order to learn the differences in dynamic trajectories for the monophthongs, diphthongs and contour vowels, I measure the travel distance in the vowel space of each diphthong, which is calculated by the distance in the vowel space between two points in the dynamic trajectory. Displacement of the vowel is defined as the square root of the sum of the squared difference in first and second formant as presented in figure 4.18. For this study, however, I am not doing any discrimination experiments to test the role the dynamic trajectory plays in vowel perception.

$D_{BA} = \sqrt{(F1B - F1A)^2 + (F2B - F2A)^2}$

in which, D_{BA} is the Displacement between point B and point A in the trajectory,
$F1B$ is the first formant frequency of point B,
$F1A$ is the first formant frequency of point A,
$F2B$ is the second formant frequency of point B,
$F2A$ is the second formant frequency of point A,

Figure 4.18: Formula for Displacement

4.2.2. Results

The dynamic trajectories of the eleven monophthongs are drawn using the mean formant values for five points of measurements in figure 4.19. The dynamic trajectories of all these eleven monophthongs, including the first and the last points, are located within the usual vowel space areas of these monophthongs.

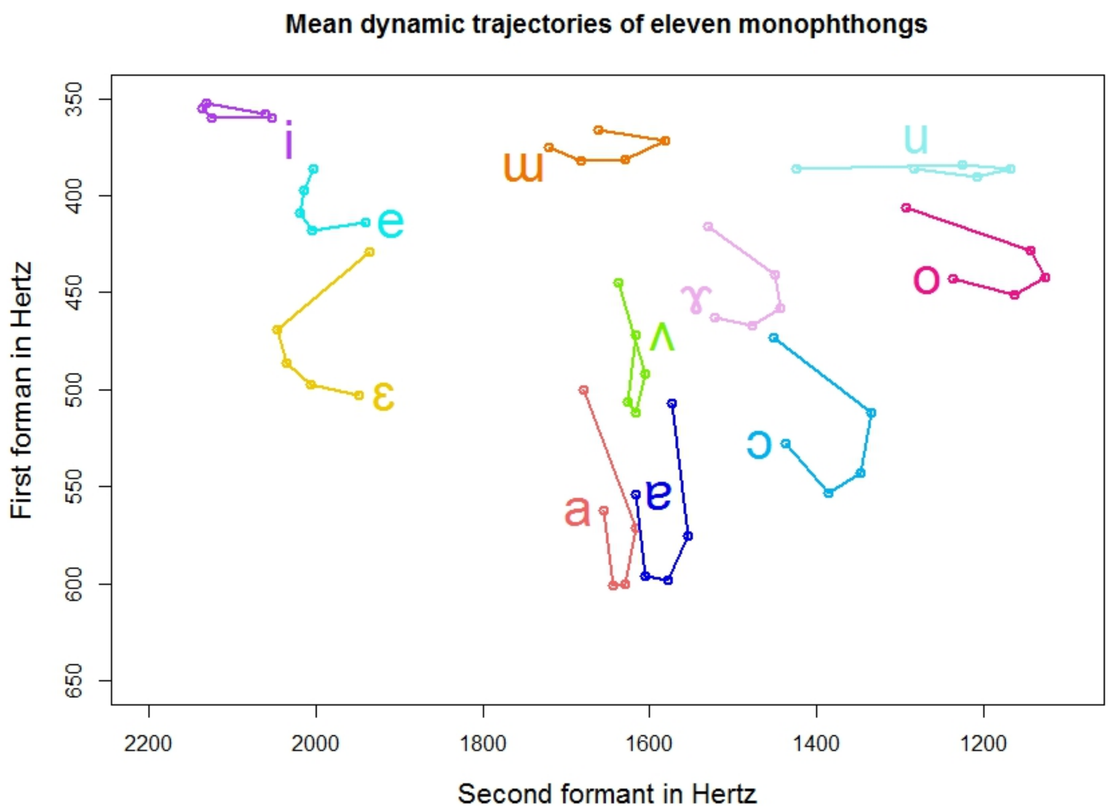


Figure 4.19: Dynamic trajectory of formants for eleven monophthongs

Note: The monophthongs mark the beginning of the trajectories.

Meanwhile, the dynamic trajectories of the vowel-glide diphthongs, drawn from the mean values of the first formant frequency and second formant frequency at all ten points, show changes over the course of the entire diphthongs. Figure 4.20 shows the dynamic trajectories of eight vowel-/j/-glide diphthongs, which move from the vowel location toward the location in the vowel /i/.

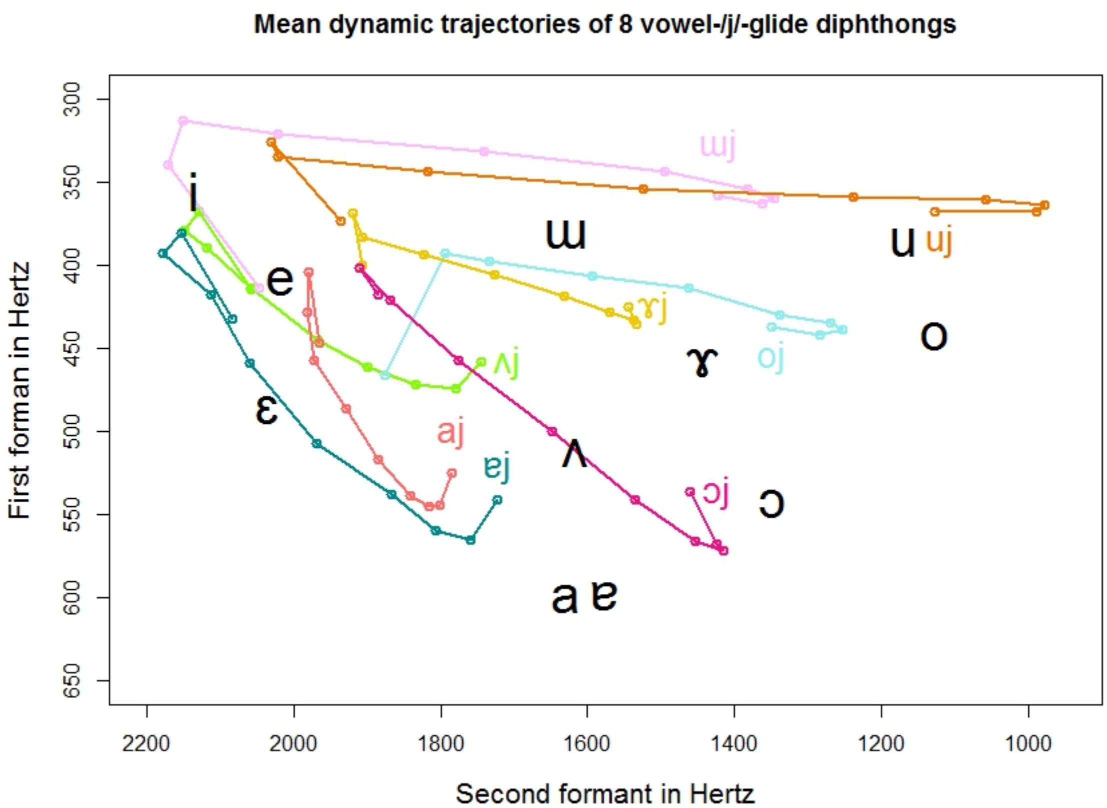


Figure 4.20: Eight Vietnamese vowel-/j/-glide diphthongs and their trajectories

Note: The black larger letters mark the mean values of the monophthongs

The dynamic trajectories of the vowel-/w/-glide diphthongs also show dramatic changes in their courses. These backing diphthongs have their trajectories moving away from the vowel position toward the vowel /u/ position in the vowel space.

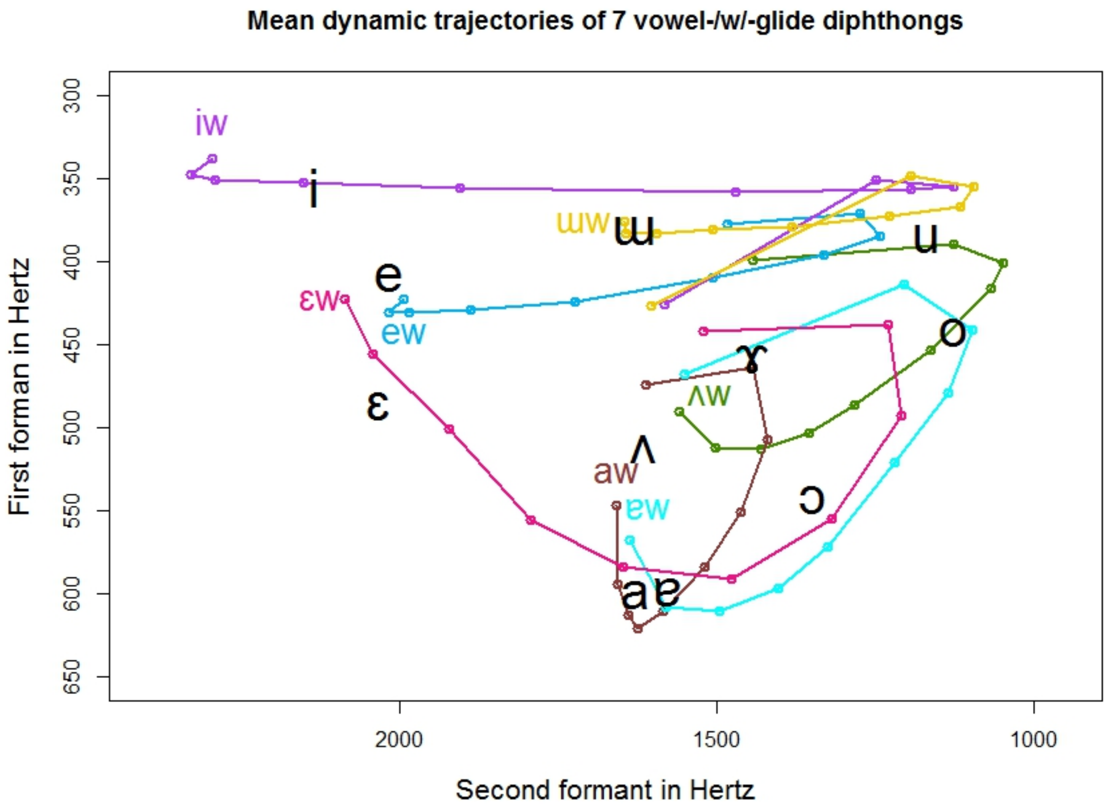


Figure 4.21: Seven Vietnamese vowel-/w/-glide diphthongs and their trajectories

Note: The black letters mark the mean values of the monophthongs

The Vowel-Glide diphthongs have different values in their Summed Displacement, which is the sum of Displacement between the adjacent points, and their Ends Displacement, which is the Displacement between the 10% point of the vowel part and the 90% point of the glide part. For example, the word *cầu* (/kʌw ˧˩/ "bridge") of speaker 2, as mentioned earlier, has a Summed Displacement of 673 , calculated by the sum of all individual Displacements after 20% of duration, and a Ends Displacement of 538, calculated by the difference in formant values between the 10% point of /ʌ/ and the 90% point of /w/. The calculated values of Summed Displacement and Ends Displacement are rounded to the nearest whole number.

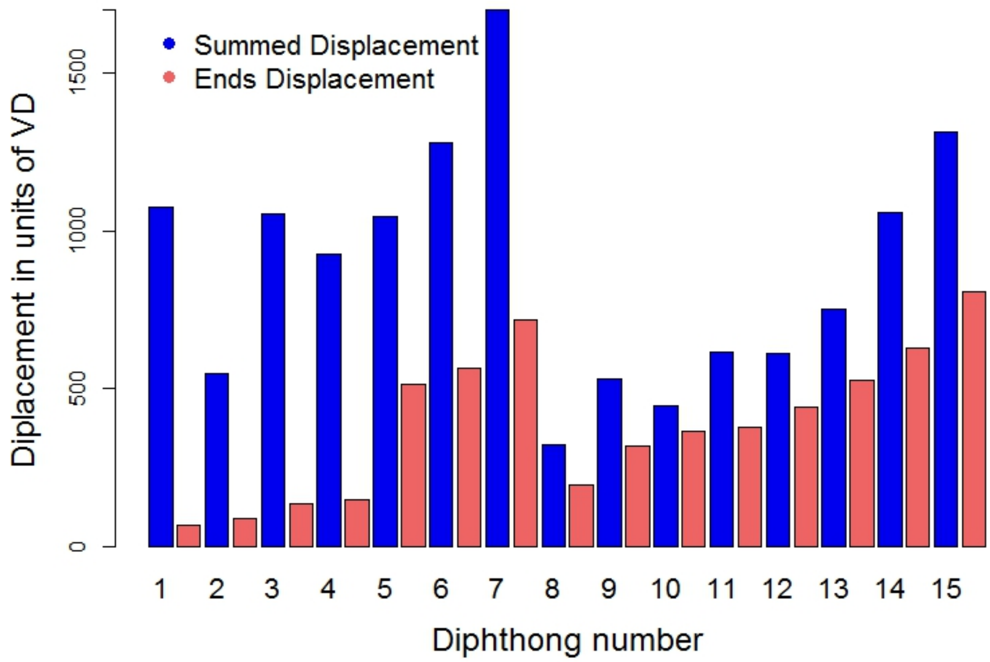


Figure 4.22: Displacements of Vietnamese vowel-glide diphthongs

Note: The numbers can be matched to their diphthongs in table 4.11

No.	Diph-thong	Ends Displace-ment	Summed Displace-ment	No.	Diph-thong	Ends Displace-ment	Summed Displace-ment
1	ɯw	1075	66	8	aj	321	196
2	aw	548	86	9	ʌj	532	316
3	ɐw	1056	133	10	ɤj	446	364
4	ʌw	928	147	11	ɐj	614	376
5	ew	1048	512	12	ɔj	613	440
6	ɛw	1280	565	13	oj	750	527
7	iw	1701	718	14	ɯj	1060	627
				15	uj	1313	808

Table 4.11: Ends Displacement and Summed Displacement of vowel-glide diphthongs (based on their mean values of the first and second formant frequencies)

The contour vowels share among themselves the same pattern of movement: moving from a high vowel to a lower vowel with the same frontness and backness. Their starting points and ending points are nearly identical to the actual monophthongs, unlike the target and starting points of diphthongs, as shown in figures 4.20 and 4.21.

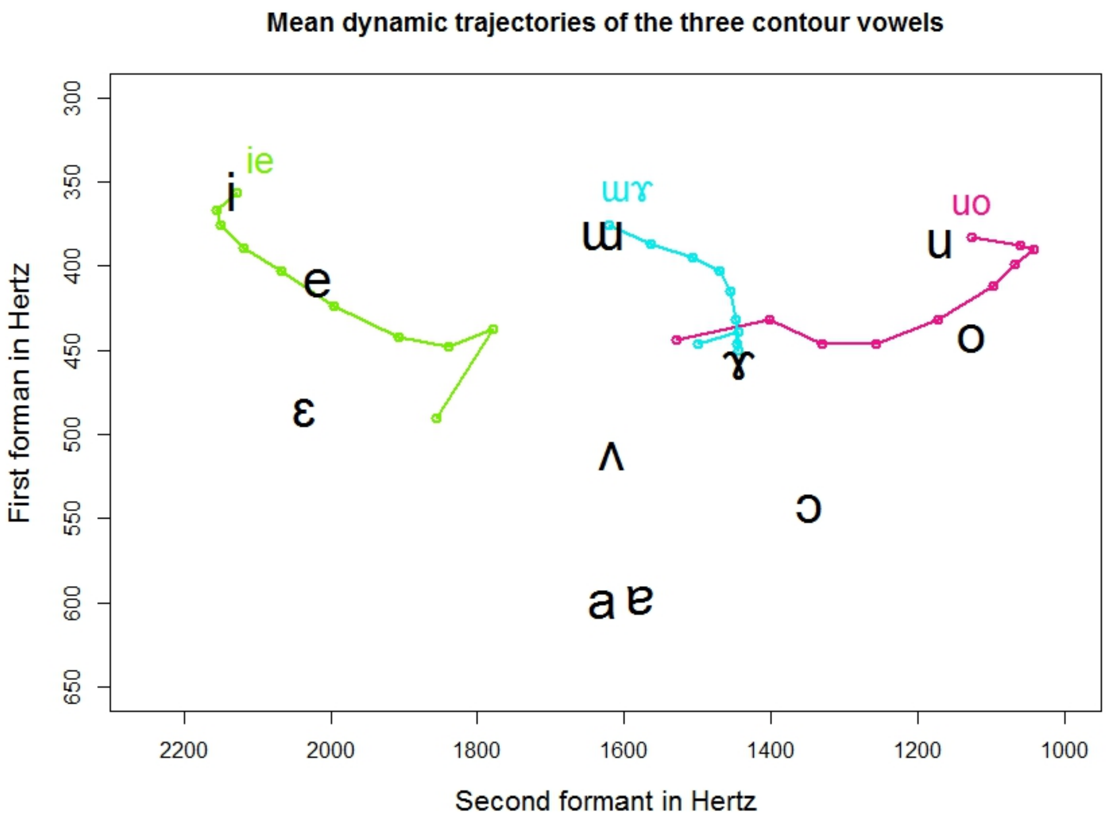


Figure 4.23: Mean dynamic trajectories of contour vowels /ie/, /ɯɤ/ and /uo/

All three contour vowels also share the same pattern in their Summed Displacement and Ends Displacement with the Ends Displacement values being about a half or two-thirds of the Summed Displacement values. This pattern, however, is not distinctively different from that of the vowel-/j/-glide diphthongs, thus it is given for reference purposes only.

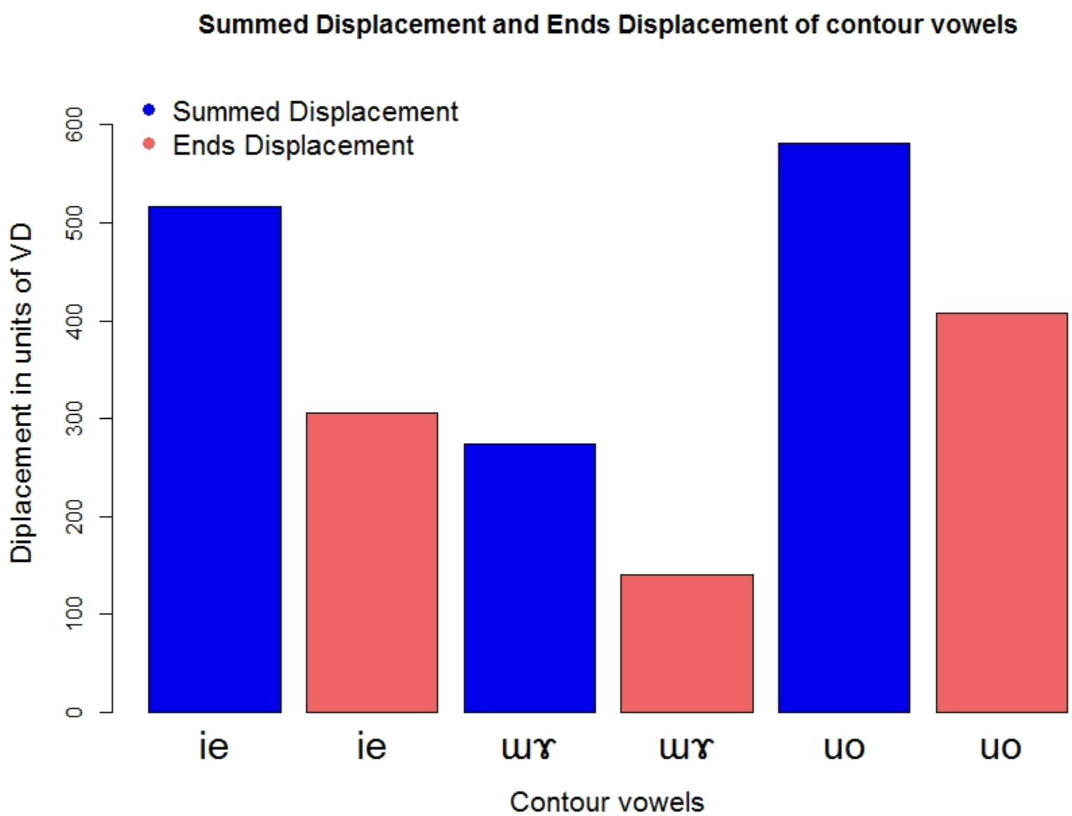


Figure 4.24: Summed Displacement and Ends Displacement of contour vowels

4.2.3. Discussion

Dynamic trajectory is important even for the monophthongs, which are traditionally measured for formant values at the 50% points, because trajectory shows how each vowel travels within the vowel space around that 50% point. For all monophthongs, their path is navigated within 50Hz for the first formant and 100Hz for the second formant. The trajectories also show that all monophthongs have steady states, during which the movements within the vowel space are small and the difference in first formant is only around 25Hz and that in second formant is around 50Hz in the second formant. These steady states last about 40% of the monophthong duration, showing only small changes in the trajectories between the 30% point, 50% point and the 70% point in figure 4.19. The trajectories of the tense-lax vowel pair /a ɐ/ also show similarity in their locations and movements. These results are consistent with the findings using only formant values at the 50% point and further strengthen the argument for the importance of duration difference in this vowel pair.

Among the vowel-glide diphthongs, the dynamic trajectory pattern is the large displacement in the first and second formants between the first and second components of these gliding diphthongs. For the /j/-glide diphthongs, the dynamic trajectories all start from the non-front locations of the vowel space moving toward the vowel [i] position. Diphthongs that would have a smaller displacement between the first and the vowel [i] position, i.e. diphthongs /ej/ and /ɛj/, do not exist in the Vietnamese diphthong inventory. Similarly, the /w/-glide diphthongs all start from the non-back locations moving toward the [u] target, and diphthongs such as /ow/ and /ɔw/ do not exist in the Vietnamese diphthong inventory. The non-existence of vowel-/j/-diphthongs and vowel-/w/-diphthongs with a shorter displacement might be proof that there is a minimum displacement threshold required for diphthongs.

The minimum distance between the mean values at 10% point of the vowel part and the 90% point of the glide part is 86 for vowel-/w/-glide diphthongs and 196 for vowel-/j/-glide diphthongs. The vowel-/w/-glide diphthongs have a larger Ends Displacement while having a smaller Summed displacement in comparison to the vowel-/j/-glide diphthongs. These differences among the vowel-glide diphthongs confirm the division of vowel-glide diphthongs into two groups – the [w]-glide group and the [j]-glide group, based on their intrinsic dynamic trajectory as a natural and unique characteristic of each diphthong group.

The dynamic trajectories of the contour vowels are very different from that of the two groups of diphthongs. For the diphthongs, the dynamic trajectories start close to the means of corresponding vowels of the vowel components, and end in the direction of the vowels /i u/ but rarely meet the mean values of the actual monophthongs /i u/. The dynamic trajectories of contour vowels, however, are more target-oriented instead of direction-oriented, having the beginning points and ending points close to the mean values of the corresponding monophthongs.

Another unique characteristic of the dynamic trajectories of the contour vowels is that the displacement of first and second formants is spread out nearly equally throughout the entire contour vowels. For the diphthongs, the displacement of formants is in smaller steady steps for the first 50% of the entire diphthong duration, which is the vowel component. The displacements of formants in the first 50% part of the diphthongs are similar to that of the monophthongs, small and steady with each 20% covering about 25Hz in first formant and 50Hz in second formants for Vj diphthongs, and 50Hz and

75Hz respectively for Vw diphthongs. The glide components of the diphthongs have much larger and irregular displacement values, being nearly twice as much as those of the vowel components. The contour vowels, on the other hand, have smaller steady steps for nearly 100% of the total duration of all three contour vowels. These smaller steady steps in the dynamic trajectories of contour vowels, as well as monophthongs and diphthongs, not only show the direction and displacement of the first and second formants, but represent the steady states in the nucleus as well. The existence of steady states through-out the contour vowels is evidence for the similarity between both components of the contour vowels and the corresponding monophthongs.

4.3. The changing duration in contour vowels and diphthongs

4.3.1. Duration of monophthongs

4.3.1.1. Data and methods

Of the total 9,034 tokens of monophthongs in my speech data from the Hanoi dialect, 5,513 tokens are of closed syllables and 3,521 tokens are of open syllables. In closed syllables, the possible codas for closed syllables include only voiceless stops /p t c k/ and nasal stops /m n ɲ ŋ/. These codas are also the only one that can appear after contour vowels in closed syllables. Nine of the monophthongs, all the vowel clusters and all the diphthongs can appear in open syllables, and the next sound segments could either be a silence speech or the onsets of the next syllables – vowels, voiced and voiceless stops, voiced and voiceless fricatives, voiced nasals, and a voiced lateral approximant. To study the duration changes in different syllable structures and environments, I measured the duration of monophthongs in closed syllables before voiceless stops and voiced nasals, and the duration of the monophthongs in open syllables before different onset environments.

4.3.1.2. Results

For all eleven monophthongs of Vietnamese, the phone duration is shortened by an average of 45ms in closed syllables for spontaneous speech and news reading speech, and by an average of 150ms in closed syllables in framed sentences.

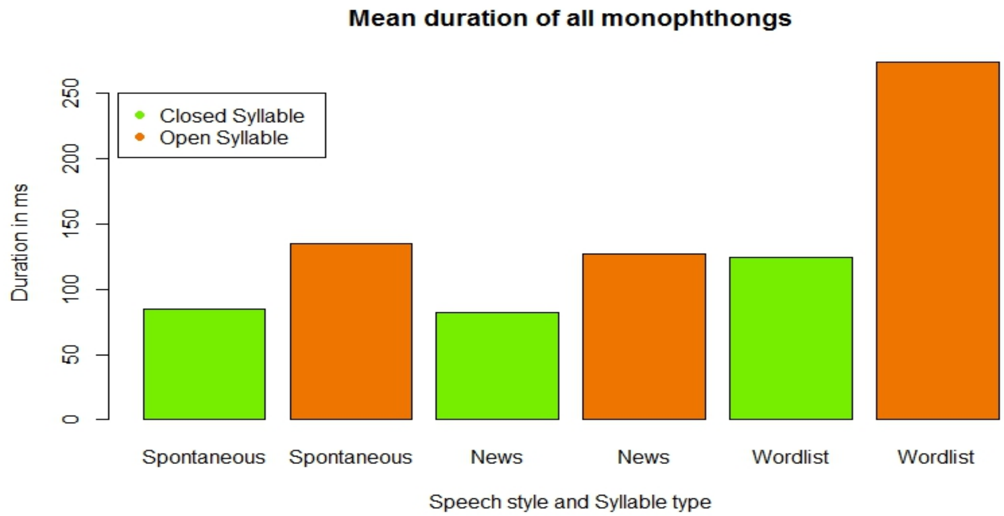
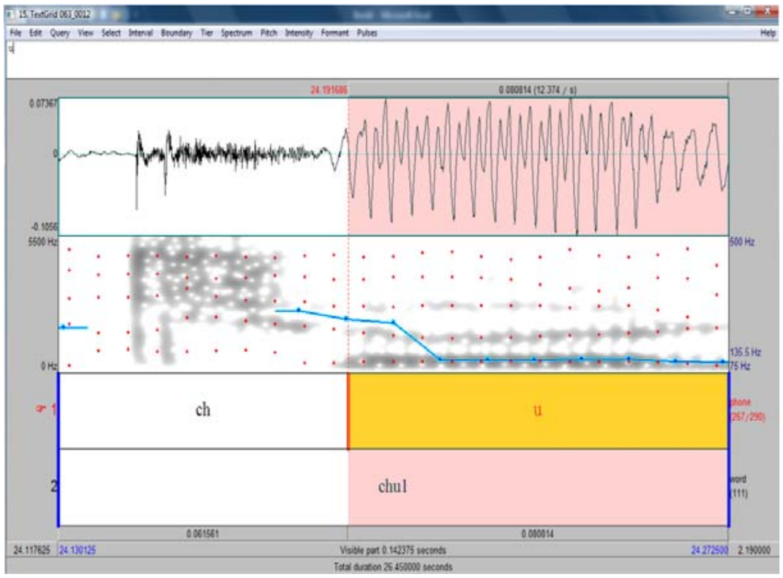


Figure 4.25: Mean values of duration for eleven monophthongs

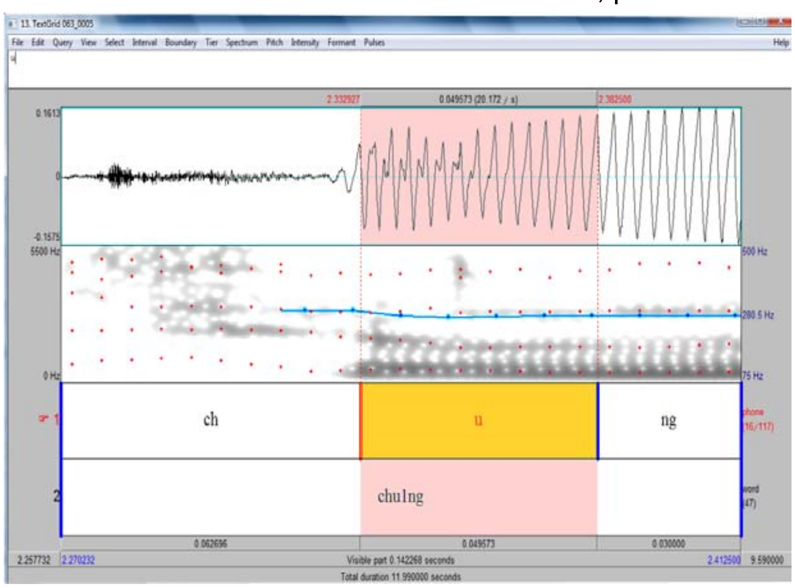
	Spontaneous speech		News reading		Word list	
	Closed syllables	Open Syllables	Closed syllables	Open Syllables	Closed syllables	Open Syllables
# of tokens	2885	1860	1924	1231	699	435
Duration in ms	85	135	82	127	125	274

Table 4.12: Mean values of duration for eleven monophthongs

The examples, extracted from the spontaneous speech of speaker 1 and speaker 6, demonstrate the difference in duration between a closed syllable and an open syllable. In the first example in figure 4.26, the word pair *chu* and *chung* have the same word duration of 142ms, but the phone /u/, which has nearly identical formant, has a duration of 80ms and 52ms respectively. In the second example, i.e. figure 4.27, the word pair *bố* and *bốn* have the total word duration of 410ms and 416ms, and the duration of phone /o/ in those words are 307ms and 170ms respectively. These two words have a more dramatic difference in duration while their formant values are nearly identical.

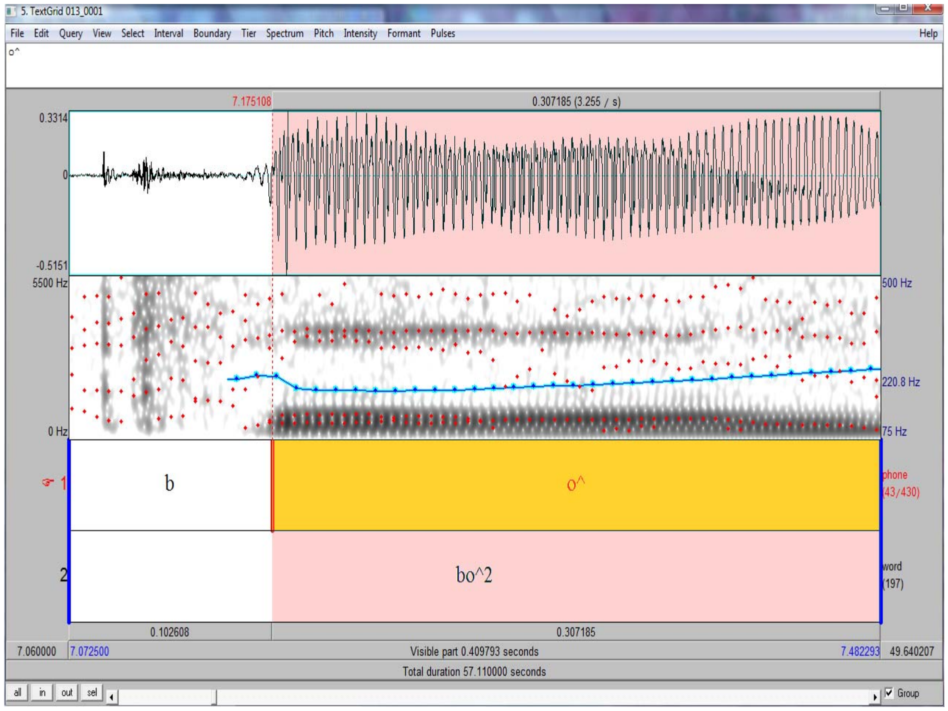


word *chu* /cu ˧/ "careful" word 142ms, phone 81ms

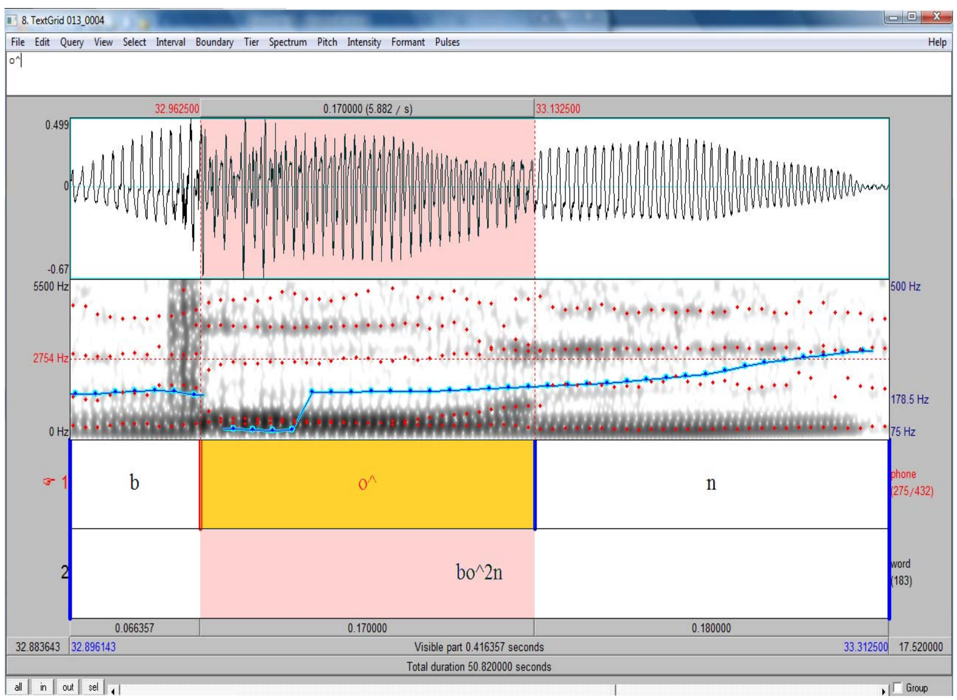


word *chung* /cuŋ ˧/ "general" word 142ms, phone 50ms

Figure 4.26: An example of open and closed syllables and their duration (extracted from the spontaneous speech portion of speaker 6)



word *bố* /ɓo ˧˥/ "father" word 410ms, phone 307ms



word *bốn* /ɓon ˧˥/ "four" word 416ms, phone 170ms

Figure 4.27: Another example of open and closed syllables and their duration (extracted from the spontaneous speech portion of speaker 1)

All the nine monophthongs that appear in both closed and open syllables have the same pattern of being shorter in closed syllables and being longer in open syllables, as shown in figure 4.28. The duration of two vowels /ɐ/ and /ɤ/ is among the shortest among all vowels in closed syllables.

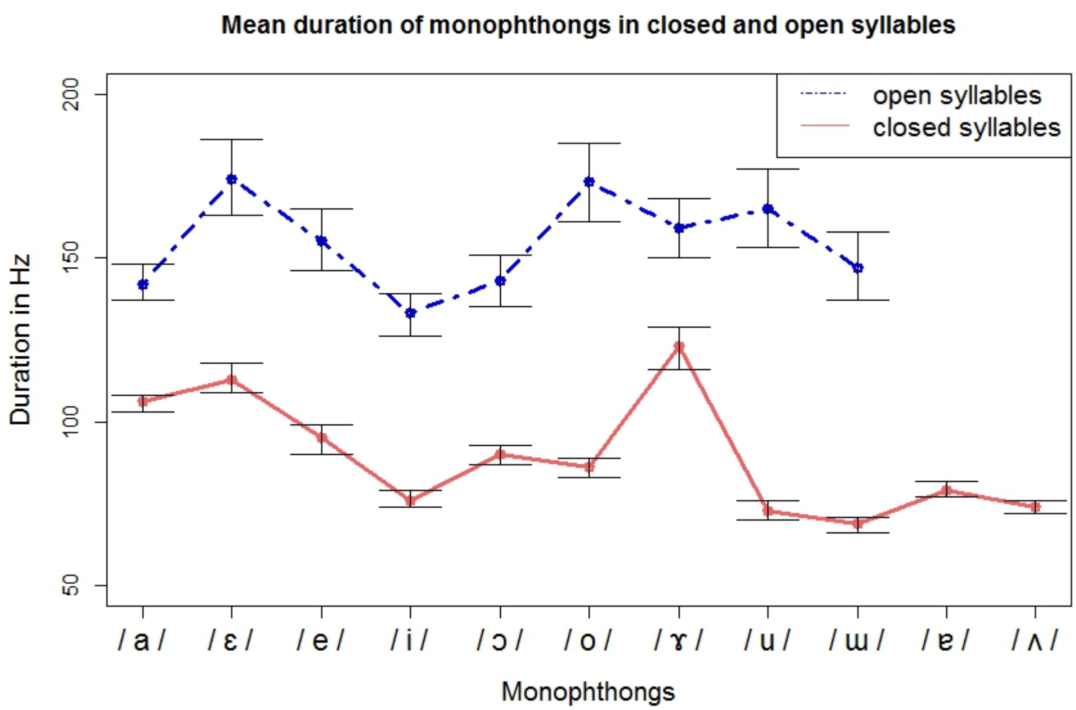


Figure 4.28: Mean duration of monophthongs in closed and open syllables

Monophthongs	Closed syllables	Open syllables
/ɐ/	456	N/A
/ʌ/	553	N/A
/a/	1035	832
/ɛ/	430	229
/e/	379	294
/i/	568	568
/ɔ/	565	539
/o/	617	197
/ɤ/	138	353
/u/	425	215
/ɯ/	342	299

Table 4.13: Number of tokens for monophthongs in closed and open syllables

Among the open syllable tokens, however, the monophthongs which precede silence speech have a much longer duration than the monophthongs preceding a voiced or voiceless segments in the next onset, as shown in figure 4.29.

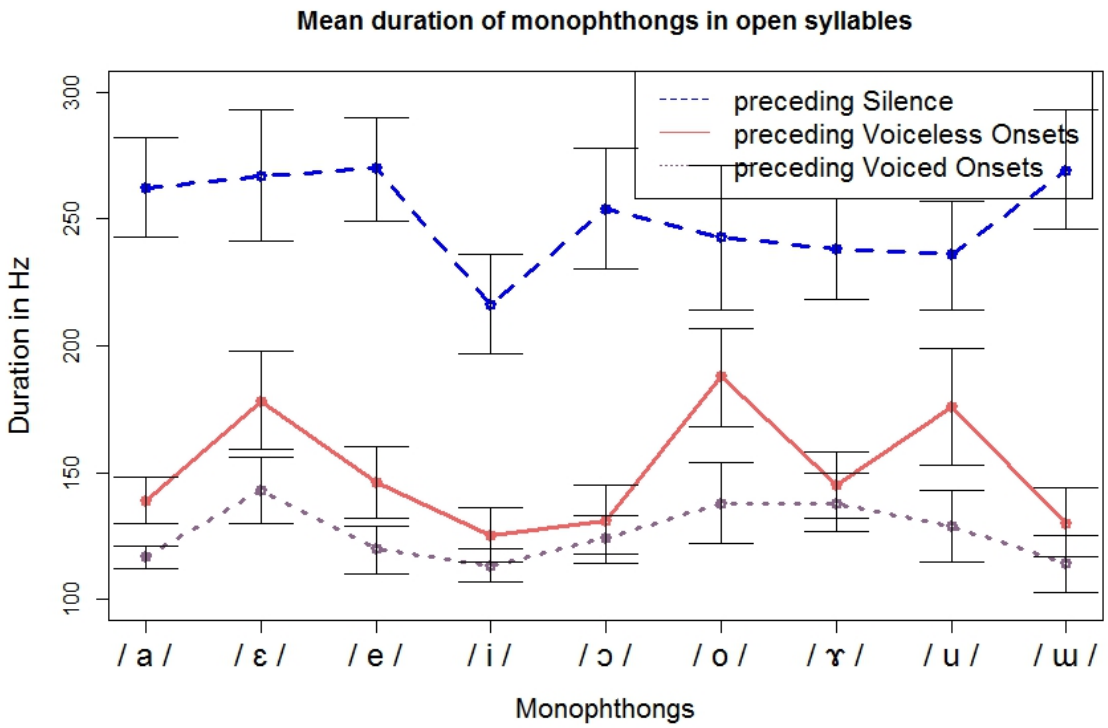


Figure 4.29: Mean duration of monophthongs in open syllables

Preceding	/a/	/ɛ/	/e/	/i/	/ɔ/	/o/	/ɤ/	/u/	/ɯ/
Silence speech	103	36	48	87	69	32	65	41	52
Voiceless	281	78	127	178	181	71	114	70	116
Voiced	448	115	119	303	289	94	174	104	131

Table 4.14: Number of tokens in open syllables in different environments

4.3.1.3 Discussion

Monophthongs have durations shortened by 50ms on average in closed syllables, as opposed to open syllables. This is a pattern that all monophthongs follow. In addition, monophthongs in open syllables are longer by an average of 100ms when they precede silence speech, and they have shorter duration preceding voiced and voiceless segments. The observation is that contour vowels can appear in both closed and open syllables just like most monophthongs. They have a similar mean duration compared to that of the monophthongs. Therefore, the duration of the entire contour vowels should behave in the same way as that of the monophthongs.

4.3.2. Duration of diphthongs and contour vowels

4.3.2.1. Data and methods

For each of the 3,549 tokens of fifteen Vietnamese vowel-glide diphthongs and 1,360 tokens of three Vietnamese contour vowels in my data set, the two components are measured separately for duration. The total duration of a diphthong or a contour vowel is the sum of duration of its components. The closed-syllable tokens of monophthongs and contour vowels are compared to one another in terms of duration. The open-syllable tokens of monophthongs and contour vowels are grouped with the tokens of diphthongs for duration comparison purposes as diphthongs only appear in open syllables. The environment of open-syllable tokens is decided by the following onset segments or following silence speech. If the onset of the next word is a voiced segment, i.e. vowel, nasal, or voiced stop and fricative, the phone is marked as preceding voiced segments. If the onset segment of the next word is a voiceless stop or voiceless fricative, the phone is marked as preceding voiceless segments. If there is long silence speech after the token, i.e. the token is at the end of the phrase, or the token is followed by a pause, it is considered as preceding silence speech.

I first studied the distribution of duration in monophthongs, diphthongs and contour vowels. Then, I compared the total phone duration of contour vowels in closed and open syllables to that of the monophthongs. I also studied the total duration of diphthongs in open syllables before silence speech versus before other segments.

4.3.2.2. Results

The tokens of contour vowels show a similar distribution of density in comparison to monophthongs, as shown in the density plot of monophthongs, contour vowels and diphthongs in figure 4.30. The density curves of the monophthongs and the contour vowels have one peak each with the mean duration of 112ms and 139ms, respectively. The diphthongs have a two-peak density curve with a mean duration of is 188ms.

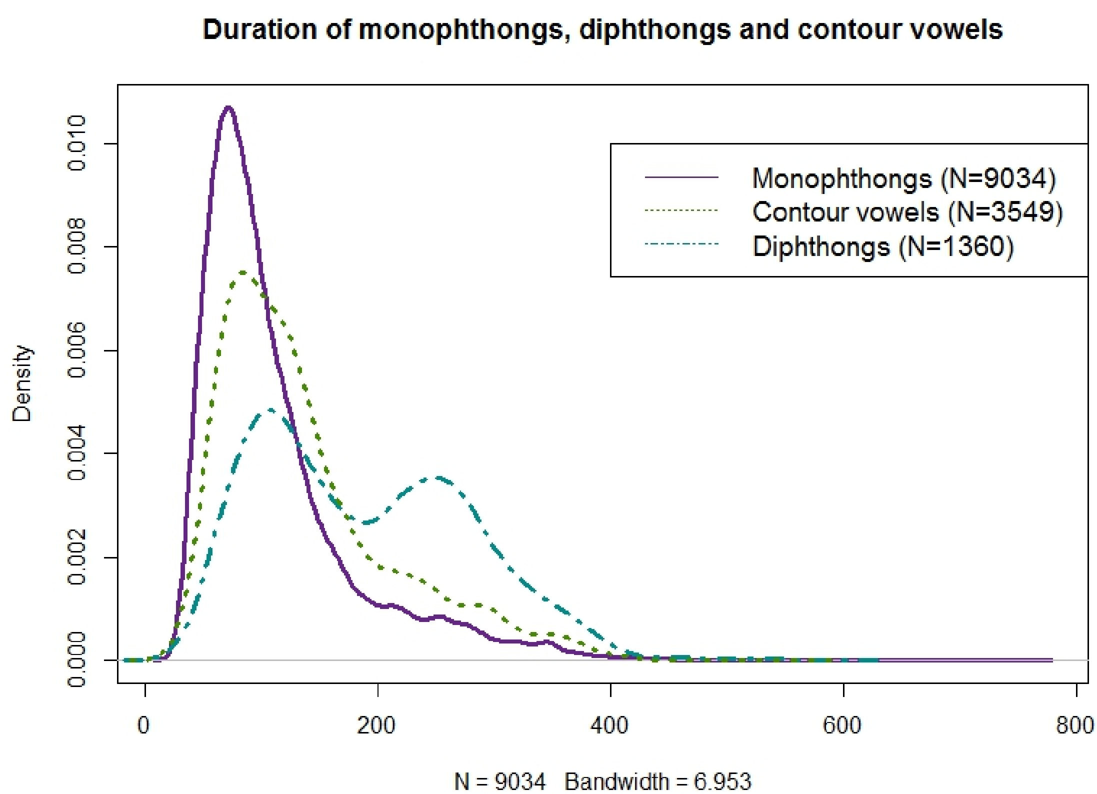


Figure 4.30: Density plot of duration of monophthongs, diphthongs and contour vowels

The mean duration values of the contour vowels in open and closed syllables show the same pattern as that of the monophthongs, with the total duration being longer in open syllables and shorted in closed syllables, as shown in figure 4.31. The difference in the mean duration of the open-syllable tokens and the closed-syllable token varies from 55Hz for /uo/ to 100Hz for /ie/.

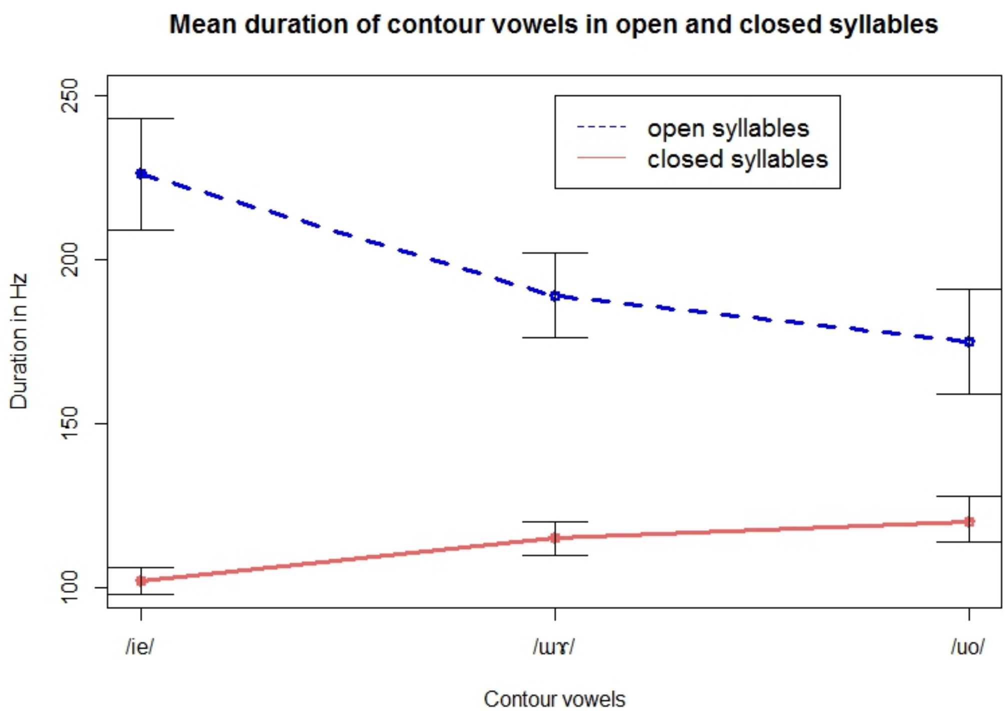


Figure 4.31: Mean duration of contour vowels in open and closed syllables

	Open syllables	Closed syllables
/ie/	104	366
/ɯɤ/	194	341
/uo/	167	188

Table 4.15: Total number of tokens of contour vowels

In open syllables, the contour vowels /ie/ and /ɯɤ/ have the longest duration before silence speech and shortest duration preceding voiced onsets, as shown in figure 4.32. The contour vowel /uo/, however, has its longest duration preceding voiceless onsets.

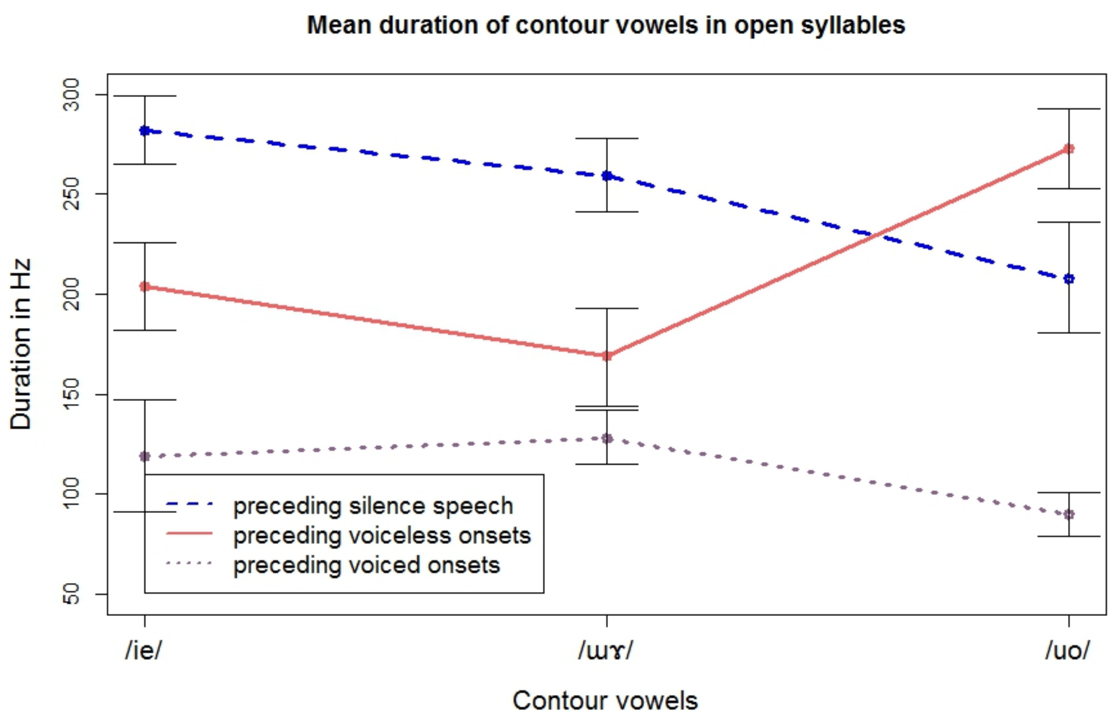


Figure 4.32: Contour vowels in open syllables

Precede	Silence speech	Voiceless segments	Voiced segments
/ie/	47	41	16
/ɯɤ/	73	57	64
/uo/	51	41	75

Table 4.16: Total number of tokens of contour vowels in open syllables

The diphthongs, which only appear in open syllables, have similar behavior to the contour vowels in regard to their duration being longest before silence speech, longer before voiceless onsets and shortest before voiced onsets (figure 4.33 and 4.34). However, the duration values here are the total duration values of both components of the diphthongs and of the contour vowels. This similar behavior is true for both vowel-/w/-glide diphthongs and the vowel-/j/-glide diphthongs, except for the diphthong /ɯj/ (figure 4.34).

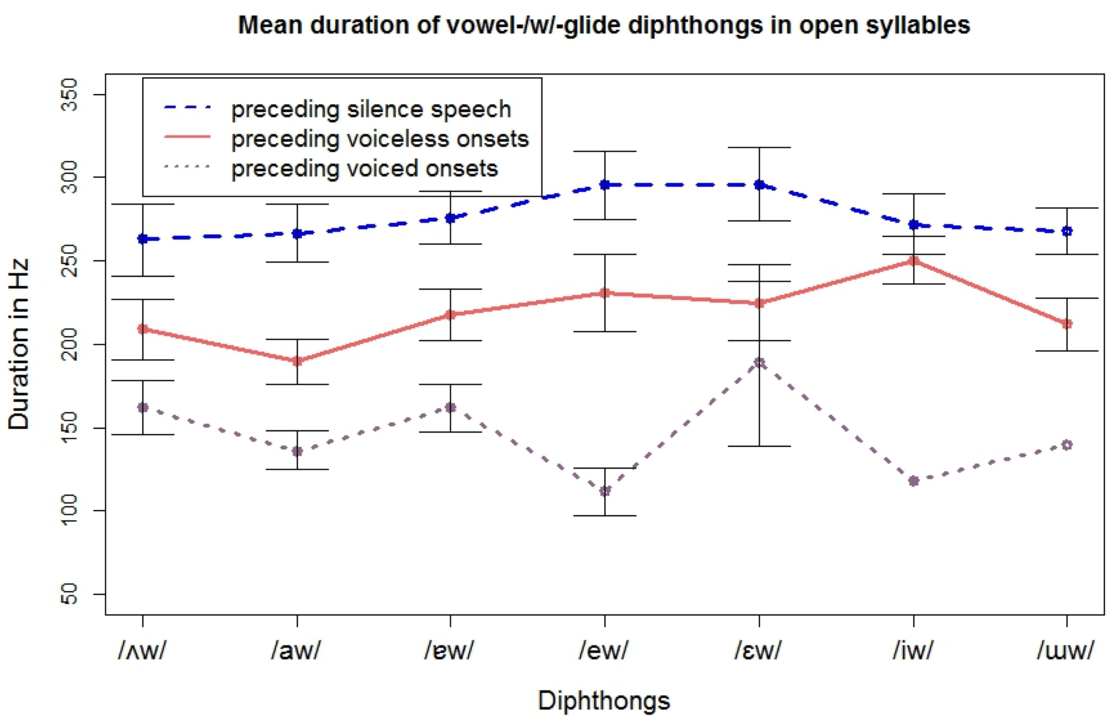


Figure 4.33: Mean duration of vowel-/w/-glide diphthongs in open syllables

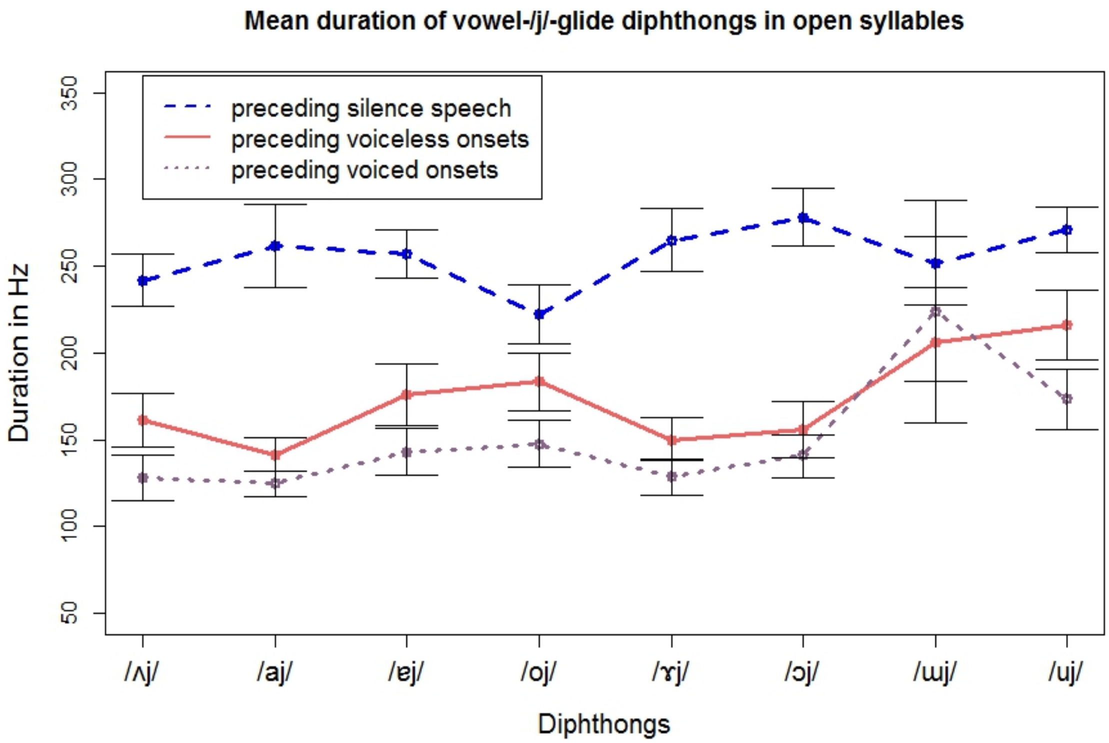


Figure 4.34: Mean duration of vowel-/j/-glide diphthongs in open syllables

Precede	Silence speech	Voiceless segments	Voiced segments
/ʌw/	66	54	51
/aw/	106	128	157
/ɐw/	62	57	70
/ew/	33	53	19
/ɛw/	32	49	16
/iw/	40	30	1
/ɯw/	54	62	3
/ʌj/	100	92	115
/aj/	78	230	288
/ɐj/	77	88	82
/oj/	118	113	133
/ɤj/	78	145	168
/ɔj/	57	101	80
/ɯj/	46	20	6
/uj/	70	45	56

Table 4.17: Total number of tokens of diphthongs in open syllables

Note: the shaded boxes show diphthongs with very few tokens

4.3.2.3 Discussion

In this part of the study, only the total duration of contour vowels and diphthongs is evaluated. This value of total duration is the sum of duration of the two components in the contour vowels and diphthongs. In closed syllables, the total duration of all three contour vowels behaves in the same way as that of the monophthongs, being longer in open syllables. This result supports the claim of contour vowels being more alike to monophthongs and the two components within the contour vowels are more or less parts of one unit instead of being two separate independent parts. Contour vowels and diphthongs can not be compared in a closed-syllable environment because none of the Vietnamese diphthongs appear in this environment.

In open syllables, the total duration of Vietnamese vowel-glide diphthongs and contour vowels are influenced by their following segments. The contour vowels /ie/ and /ɯɤ/ are longest before silence speech and all diphthongs are also at their longest duration before silence speech. Both vowel-/w/-glide and vowel-/j/-glide diphthongs behave similarly before voiced and voiceless segments with the total duration of each diphthong being longer before voiceless segments, except for the /ɯj/ diphthong. The difference in duration values of diphthongs before voiceless segments and voiced segments varies among diphthongs and this difference in values seems to be larger in vowel-/w/-glide diphthongs. The overall results show that the whole duration of the diphthongs and contour vowels is influenced by the environments in the same pattern. However, I suspect that the individual components within the diphthongs and contour vowels do not follow the same pattern.

4.3.3. Changes in duration of the components in diphthongs and contour vowels

4.3.3.1. Data and methods

For the 3,549 tokens of diphthongs and 1360 tokens of contour vowels, the duration of each component is measured separately. As diphthongs only appear in open syllables, only the components of contour vowels in open syllables are taken into account for this comparison. I will show that both components in the contour vowels are more likely to be influenced by the environments, i.e. both are more likely to have longer duration in open syllables preceding silence speech, while the glide component of diphthongs are less likely to be influenced by the environments.

4.3.3.2. Results

The vowel component within diphthongs shows a difference in length in different environments. In the case of vowel-/w/-glide diphthongs preceding silence speech, the mean durations of the vowel components in these vowel-/w/-glide diphthongs are longer than those of the corresponding monophthongs. Figure 4.35 shows the bar plot of duration values for each vowel component with the mean duration of corresponding monophthongs represented by a solid line zigzagging across all the bar plots.

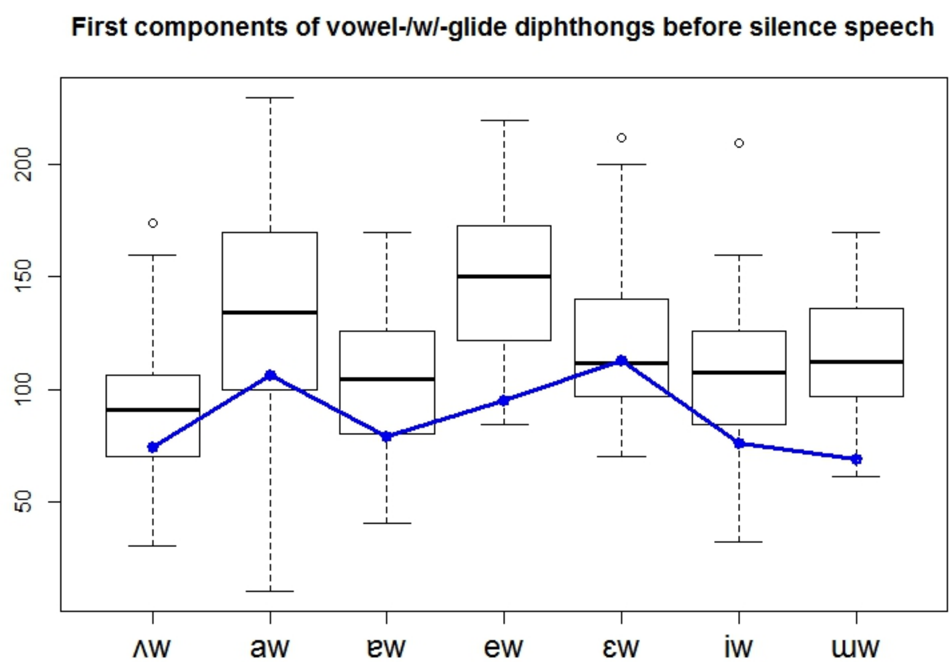


Figure 4.35: Duration of the first components of vowel-/w/-glide diphthongs before silence speech

When these diphthongs precede a voiceless onset, however, the mean durations of the vowel components and of their corresponding monophthongs are nearly the same, as shown in figure 4.36. When the vowel-/w/-glide diphthongs precede a voiced onset, the mean durations of the vowel components are actually shorter than those of the corresponding monophthongs, as shown in figure 4.37.

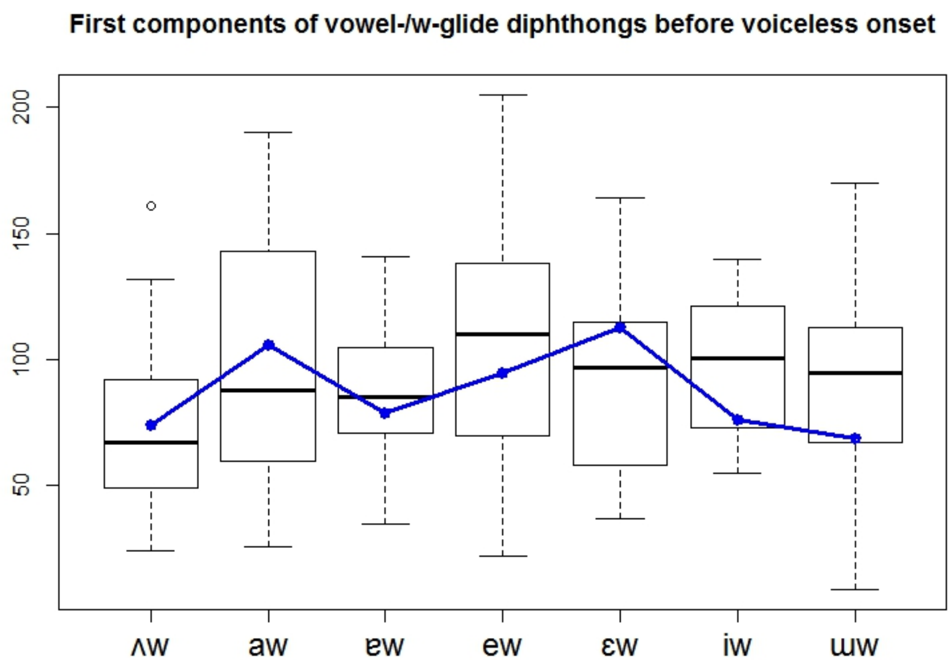


Figure 4.36: Duration of the first components of vowel-/w/-glide diphthongs before voiceless onsets

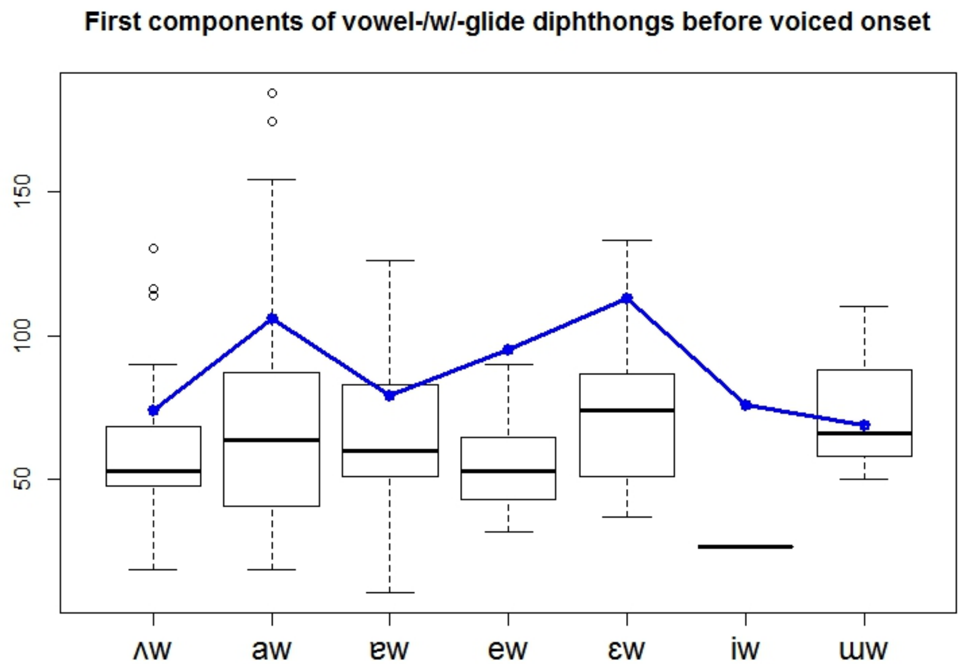


Figure 4.37: Duration of the first components of vowel-/w/-glide diphthongs before voiced onsets

The glide component, i.e. the second component, of the vowel-/w/-glide diphthongs, however, do not follow this pattern of duration changes due to the environment influence. The bar plots in figure 4.38, 4.39, and 4.40 show the duration values of the second component of the vowel-/w/-glide diphthongs being compared to the mean duration value of the corresponding vowel /u/, shown in figure 4.38, 4.39, and 4.40 as the solid line across all the bar plots. The mean duration of the second component is longer than that of vowel /u/ before silence speech and voiceless onsets.

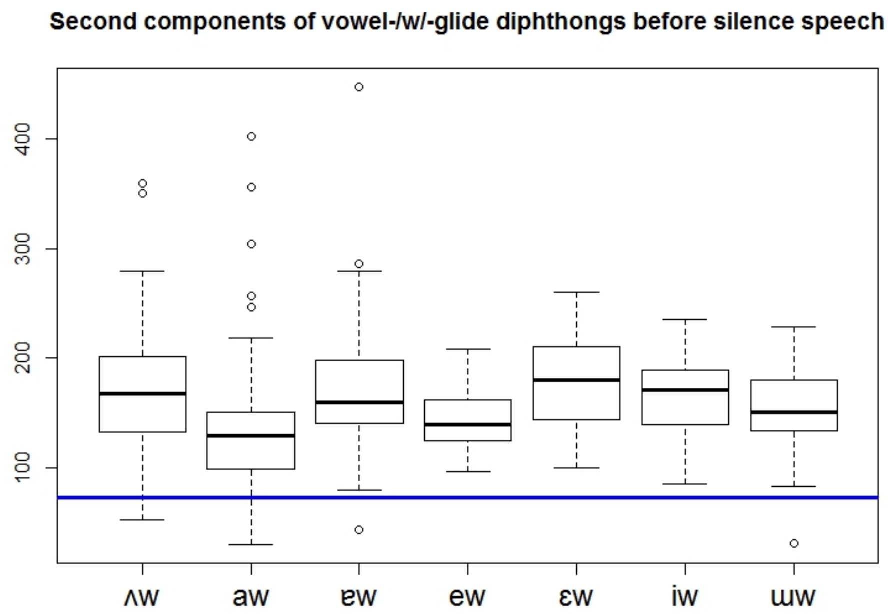


Figure 4.38: Duration of the second components of vowel-/w/-glide diphthongs before silence speech

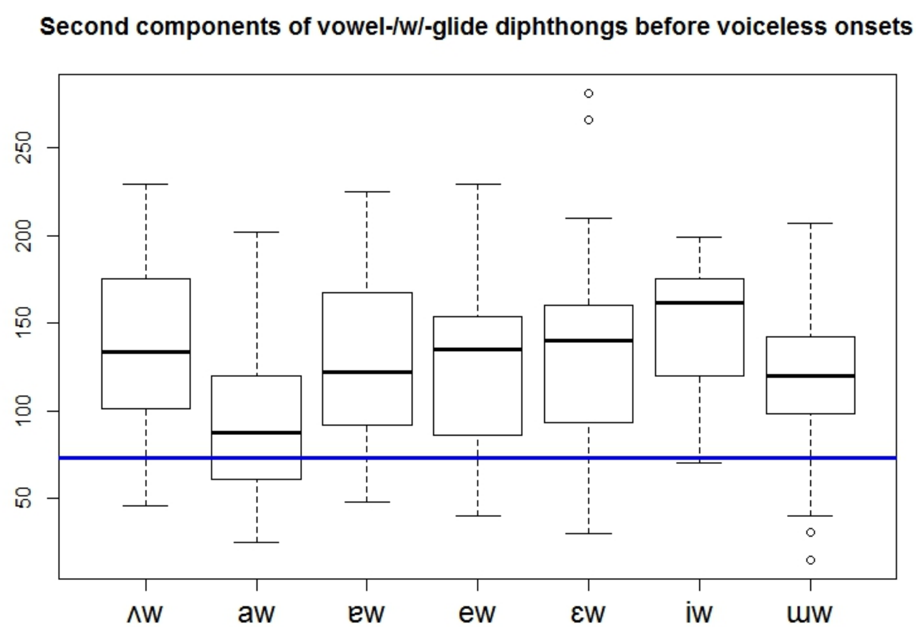


Figure 4.39: Duration of the second components of vowel-/w/-glide diphthongs before voiceless onsets

Before voiced onsets, the second components of the vowel-/w/-glide diphthongs have similar duration values in comparison to the mean duration value of the corresponding vowel /u/, as shown in figure 4.40.

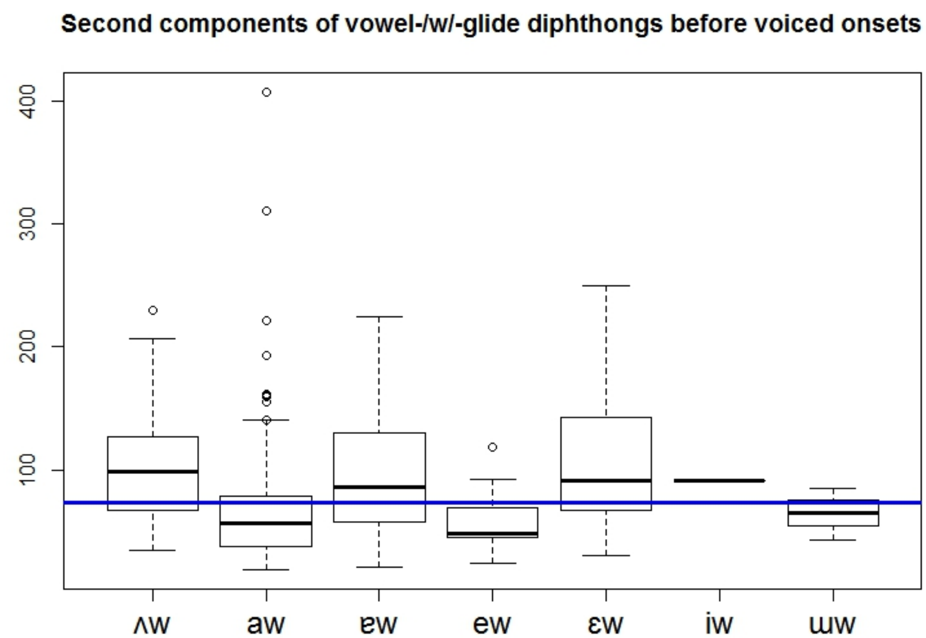


Figure 4.40: Duration of the second components of vowel-/w/-glide diphthongs before voiced onsets

This same pattern of duration change in the vowel components and the duration stability in the glide components is exhibited in the vowel-/j/-glide diphthongs as well, as shown in figures 4.41, 4.42, 4.43, 4.44, 4.45, and 4.46 below. The duration of the vowel component of the vowel-/j/-glide diphthongs in comparison to the mean duration value of the corresponding vowel /i/, is longer before silence speech, similar before voiceless onsets, and shorter before voiced onsets.

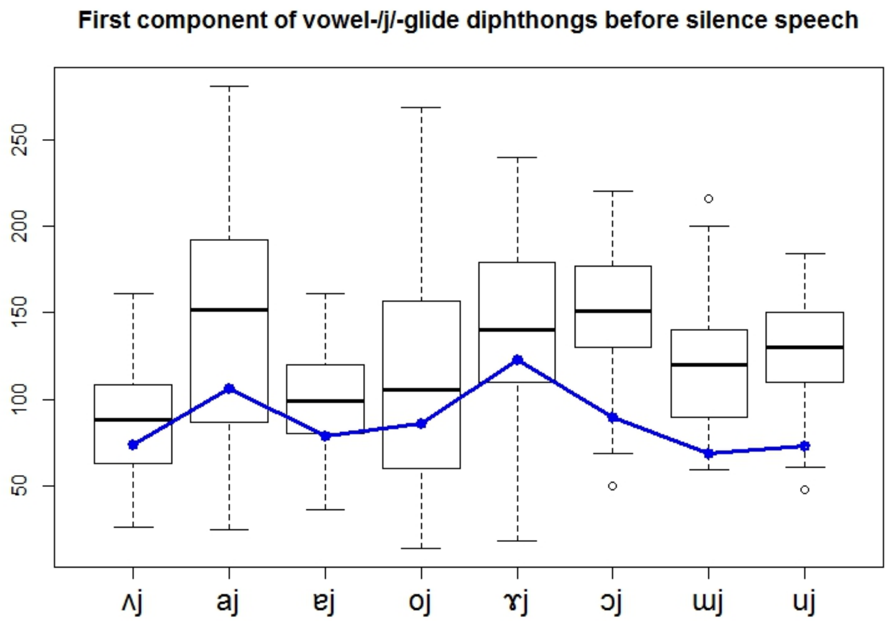


Figure 4.41: Duration of the first components of vowel-/j/-glide diphthongs before silence speech

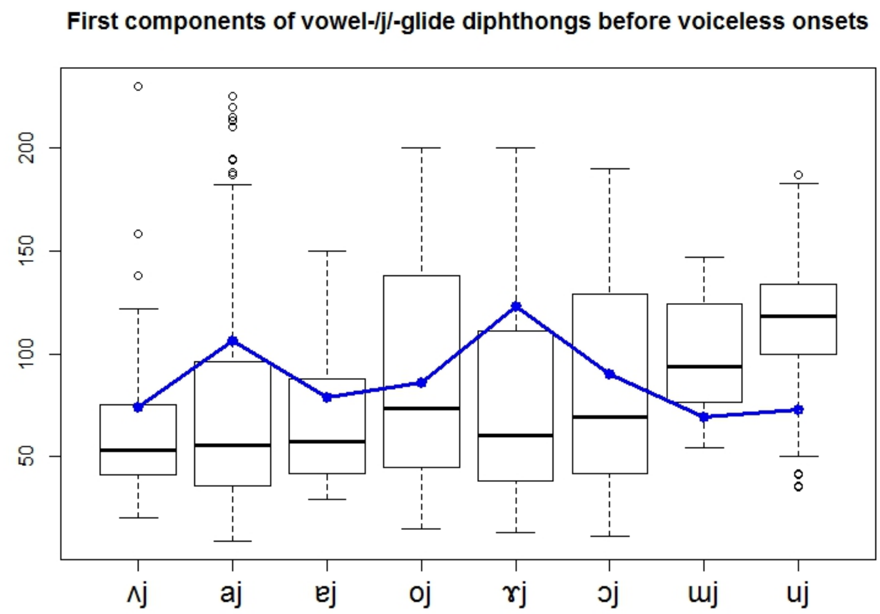


Figure 4.42: Duration of the first components of vowel-/j/-glide diphthong before voiceless onsets

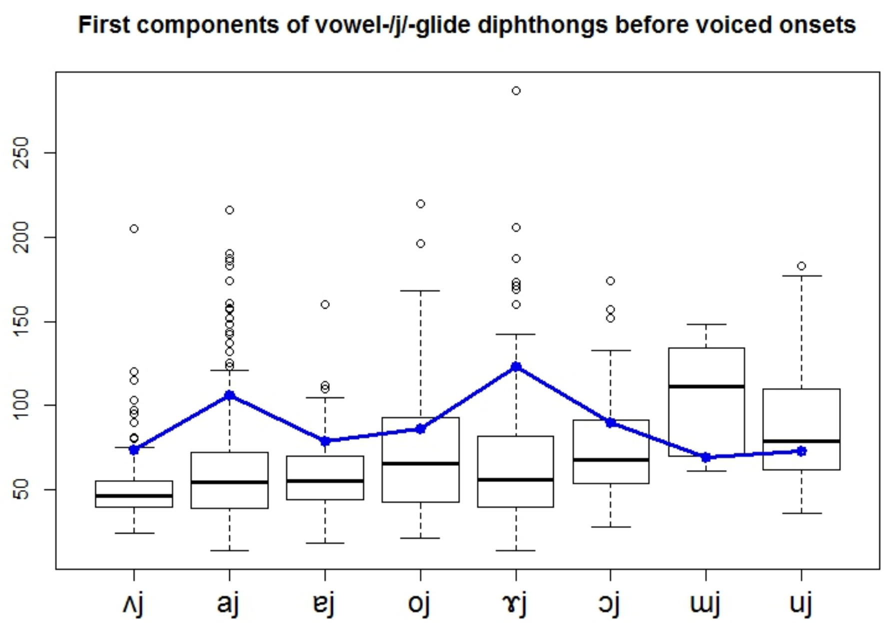


Figure 4.43: Duration of the first components of vowel-/j/-glide diphthong before voiced onsets

The second components of these vowel-/j/-glide diphthongs also have a similar duration pattern as the second component of the vowel-/w/-glide diphthongs before silence speech and before voiced speech, being longer before silence speech while having similar duration before voiced onsets, as shown in figure 4.44 and 4.46. When the diphthongs are followed by voiceless onsets, however, the second components of the vowel-/j/-glide diphthongs have a similar duration to that of the corresponding vowel /i/, as shown in figure 4.45.

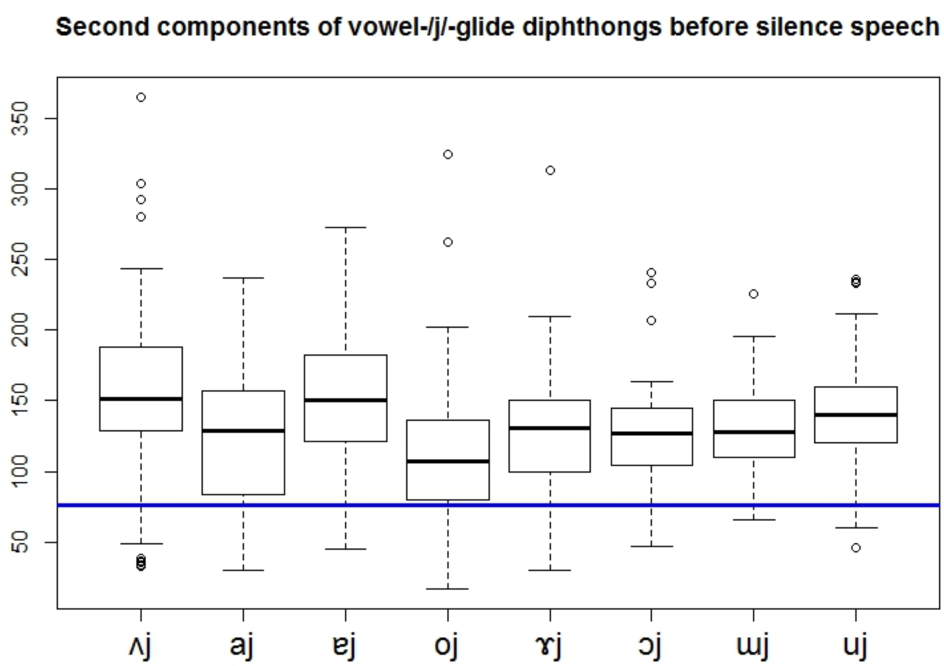


Figure 4.44: Duration of the second components of vowel-/j/-glide diphthongs before silence speech

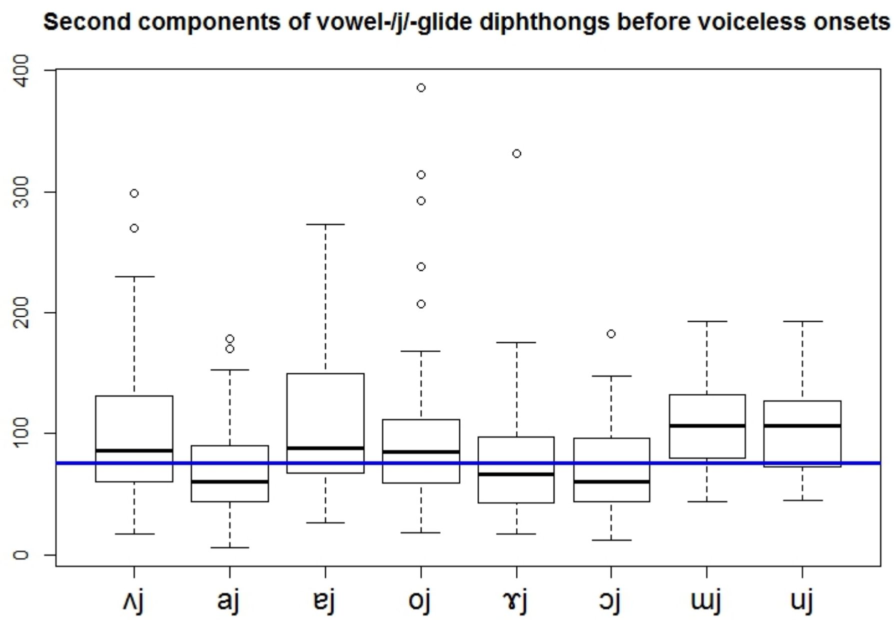


Figure 4.45: Duration of the second components of vowel-/j/-glide diphthongs before voiceless onsets

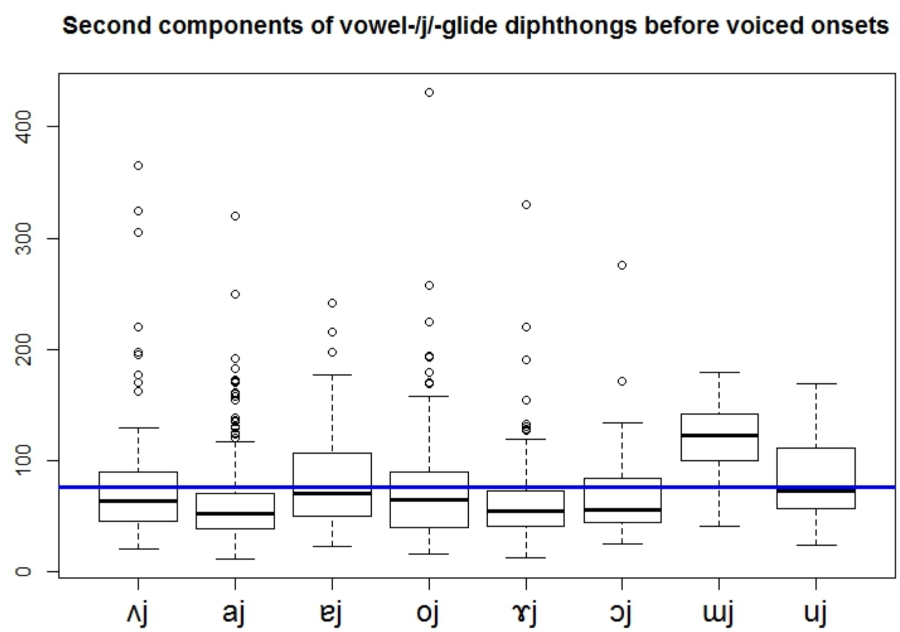


Figure 4.46: Duration of the second components of vowel-/j/-glide diphthongs before voiced onsets

Contour vowels, however, behave differently from the diphthongs. Both of the components of the contour vowels behave like the first vowel components of the diphthongs, being longer before silence speech, at average length before voiceless onsets and being shorter before voiced onsets. Figures 4.47, 4.48, and 4.49 show a similar pattern in duration of the first components and the second components in contour vowels, which is the pattern of the vowel components of the diphthongs, not that of the glide components of the diphthongs.

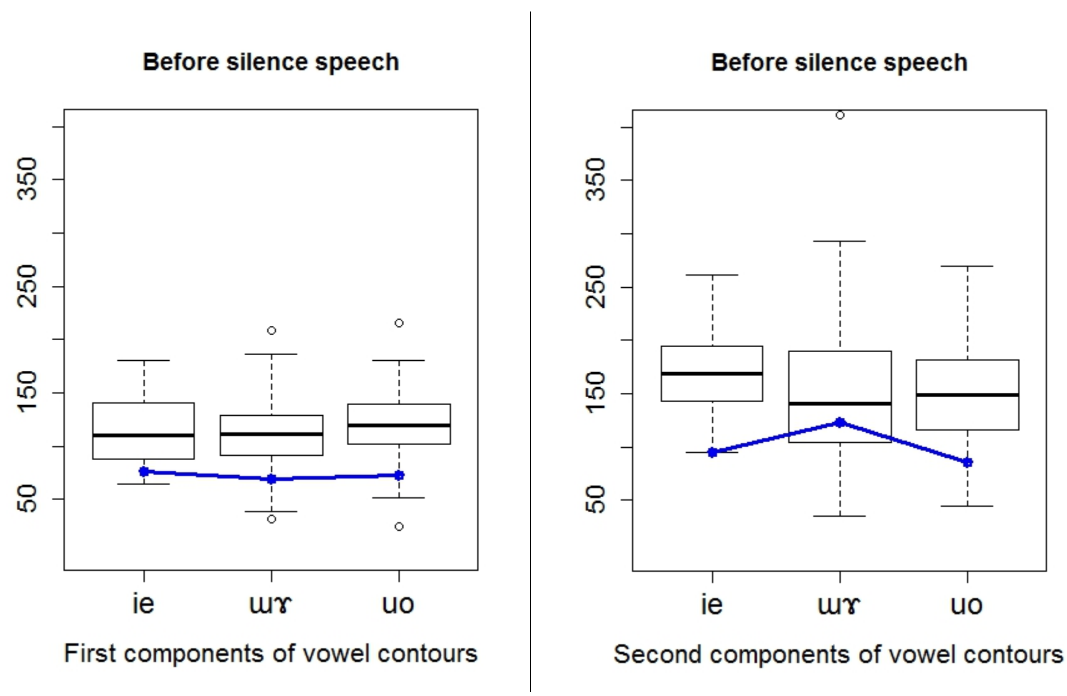


Figure 4.47: Duration of the first and second components of contour vowels before silence speech

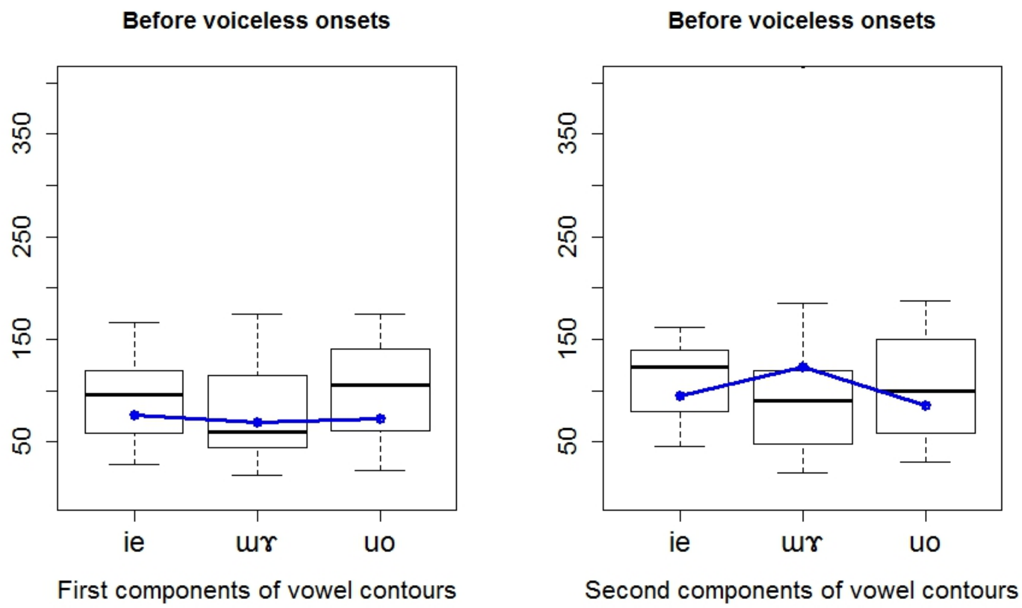


Figure 4.48: Duration of the first and second components of contour vowels before voiceless onsets

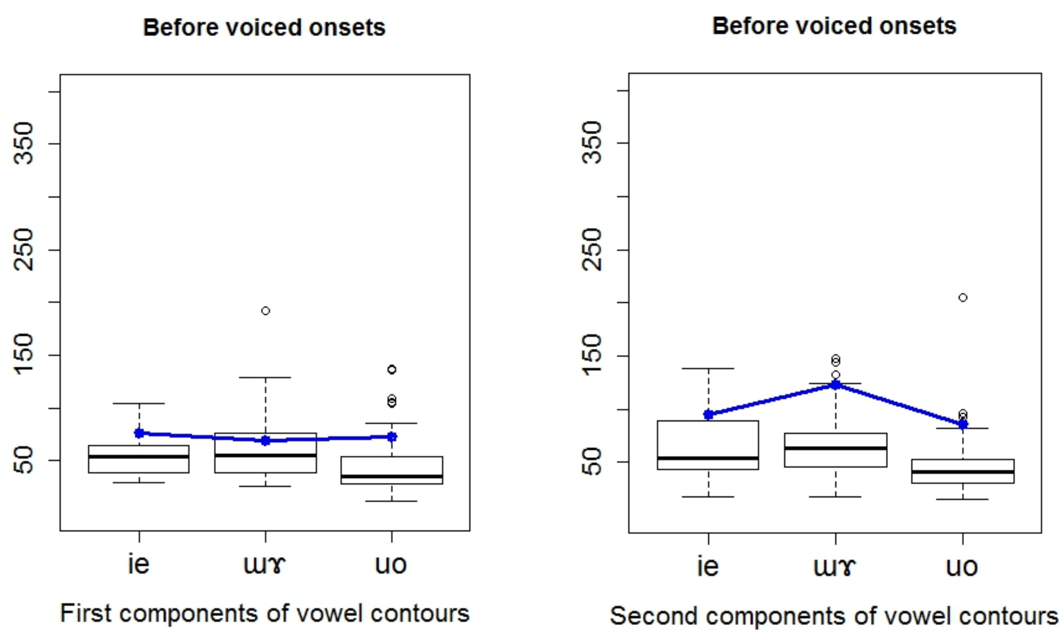


Figure 4.49: Duration of the first and second components of contour vowels before voiced onsets

4.3.3.3. Discussion

While the whole duration of contour vowels and diphthongs changes in the same direction under the influence of the following onsets, the duration of the individual components do not. Both components in the contour vowels follow the same patterns as the vowel components in the diphthongs and the monophthongs. The glide components of the diphthongs do not exhibit the same pattern of duration change before voiceless and voiced onsets and the changes in the whole duration of diphthong are the results of the changes in the first component of the diphthongs.

The behaviors in duration of contour vowels prove that the contour vowels /ie/, /ɯɤ/ and /uo/ do not contain glide segments and their duration behaves similarly to that of the Vietnamese monophthongs. Thus, together with the similarities in formant targets, Vietnamese contour vowels should be classified in the same group with Vietnamese monophthongs in terms of phonetic behavior, and Vietnamese diphthongs constitute a different natural class.

Chapter 5

Phonological arguments for the status of contour vowels

5.1. Description of Vietnamese monophthongs using feature geometry

5.1.1. Vietnamese vowel system

Thompson (1965) described the syllable structure for Vietnamese as (C1)V(C2) with a tone, in which C1 can be an optional single consonant or a consonant plus a /w/ glide, and C2 is an optional consonant. The nucleus V can be a monophthong, a diphthong or a contour vowel. This approach includes the glide /w/ in *quê* /kwe/ with the consonant /k/ and describes that the combination of consonant and /w/ is possible for all Vietnamese consonants except for /n/ and labial stops (Thompson 1965, p. 46). The common rhyming practice in Vietnamese poems, as shown in the rhyming examples 1 and 2 below (figure 5.1 and 5.2), also shows that the glide /w/ in the /w/-vowel sequences are included in the onset of the syllable, not in the nucleus, as in the examples in table 5.1.

Figure 5.1: Rhyming example 1

A folklore poem *Người nông dân và con trâu* (The farmer and the water buffalo)
(Nguyên 2001, p. 172)

Trâu ơi, ta bảo Trâu ***này***$_{1}$,	Dear Buffalo, let me tell you
Trâu ra ngoài ruộng, Trâu ***cày***$_{1}$ *với* ***ta***$_{2}$.	You go to the field, you work with me
Cấy cày vốn nghiệp nông ***gia***$_{2}$	Working in the field is the farmer's job
Ta đây Trâu đấy, ai ***mà***$_{2}$ *quản* ***công***$_{3}$.	If you and me are there, we are our managers
Bao giờ cây lúa còn ***bông***$_{3}$,	As long as the rice is blooming,
Thời còn ngọn cỏ ngoài ***đồng***$_{3}$ *Trâu ăn.*	There will be grass in the field for you

Rhyme No.	Word 1	IPA	Word 2	IPA	Word 3	IPA
1	*này*	/nɐj ˧˩/	*cày*	/kɐj ˧˩/		
2	*ta*	/ta ˧/	*gia*	/za ˧/	*mà*	/ma ˧˩/
3	*công*	/koŋ ˧/	*bông*	/ɓoŋ ˧/	*đồng*	/ɗoŋ ˧˩/

Figure 5.2: Rhyming example 2

A well-known verse (line 1307 and 1307) from
Truyện Kiều (The story of a girl named Kieu)
(Nguyễn 1997, p. 173)

Dưới trăng quyên đã gọi ***hè***	In the moonlight, the birds call out for summer
Đầu tường lửa lựu lập ***lòe*** *đâm bông*	By the wall, the red buds are glowing.

Rhyme: ***hè*** /hɛ ˧˩/ is rhymed with ***lòe*** /lwɛ ˧˩/

No.	Original word	IPA	Matching Rhyme	IPA	Reference
1	*nhà*	/ɲa ˧˩/	*quà*	/kwa ˧˩/	p. 47
2	*tuông*	/tuoŋ ˧/	*buôn*	/ɓuon ˧/	p. 39
3	*vàng*	/vaŋ ˧˩/	*hoàng*	/hwaŋ ˧˩/	p. 74, p. 170
4	*ai*	/aj ˧/	*ngoài*	/ŋwaj ˧˩/	p. 76
5	*già*	/za ˧˩/	*qua*	/kwa ˧/	p. 84
6	*què*	/kwɛ ˧˩/	*chè*	/cɛ ˧˩/	p. 89
7	*tre*	/cɛ ˧/	*què*	/kwɛ ˧˩/	p. 103
8	*khoai*	/xwaj ˧/	*hai*	/haj ˧/	p. 108
9	*thuyền*	/tʰwien ˧˩/	*tiền*	/tien ˧˩/	p. 119
10	*chài*	/caj ˧˩/	*ngoài*	/ŋwaj ˧˩/	p. 162
11	*tre*	/cɛ ˧/	*chòe*	/cwɛ ˧˩/	p. 210
12	*nhện*	/ɲen ˧ʔ˩/	*quện*	/kwen ˧ʔ˩/	p. 224

Table 5.1: Examples of rhyming of Cw and C from Nguyên (2001)

The analysis of the medial /w/ as part of the syllable onsets would mean the Vietnamese vowel system has eleven monophthongs, seven vowel-/w/ diphthongs, eight vowel-/j/ diphthongs, and three contour vowels, as listed in table 5.2. The medial /w/ in the Group 2 triphthongs in table 5.3 are reanalyzed as a medial /w/, which is part of the onset, and a nucleus containing either a vowel-/w/-glide diphthong, a vowel-/j/-glide diphthong or a contour vowel. The only "true" triphthongs at this point are in the Group 1 Triphthong in table 5.3, which will be proven to be diphthongs later in this chapter.

Monophthongs (open & closed syllables)	Vw diphthong (open syllables only)	Vj diphthongs (open syllables only)	VV clusters (open & closed syllables)
i	iw	---	ie
e	ew	---	---
ɛ	ɛw	---	---
a	aw	aj	---
ɐC	ɐw	ɐj	---
ɯ	ɯw	ɯj	ɯɤ
ɤ	---	ɤj	---
ʌC	ʌw	ʌj	---
ɔ	---	ɔj	---
o	---	oj	---
u	---	uj	uo

Table 5.2: Vietnamese monophthongs and diphthongs

Note: vowels followed by "C" can only appear in closed syllables

Group 1 Triphthongs	Structure	Group 2 Triphthongs	Structure
iew	VV diphthong + /w/	wiw	/w/ + Vw diphthong
ɯɤw	VV diphthong + /w/	wew	/w/ + Vw diphthong
uoj	VV diphthong + /j/	wɛw	/w/ + Vw diphthong
ɯɤj	VV diphthong + /j/	waw	/w/ + Vw diphthong
		wɐw	/w/ + Vw diphthong
		wʌw	/w/ + Vw diphthong
		waj	/w/ + Vj diphthong
		wɐj	/w/ + Vj diphthong
		wʌj	/w/ + Vj diphthong
		wie[1]	/w/ + VV diphthong

Table 5.3: Triphthongs with the possible environments (Thompson 1965)

Note: [1] Triphthong /wie/ appears in both open and closed syllables

5.1.2. Feature geometry for Vietnamese monophthongs

The feature geometry theory (Jakobson 1941, Chomsky & Halle 1968, and Clements & Keyser 1983, Halle and Clements 1983) has a group of binary features which can be used to divide sounds into natural classes. These binary features are arranged in a hierarchy to explain the dependent co-existence of features and phonological processes such as assimilation. The features to distinguish and classify English vowels are [vocalic], [high], [back], [low], [round] and [tense] (Jakobson 1941, Chomsky & Halle 1968, and Clements & Keyser 1983, Halle and Clements 1983).These six features group the ten American English monophthongs into their natural classes.

Vietnamese monophthongs can be divided into three levels of height by using the features of [high] and [low], separating /i ɯ u/ and /a ɐ/ from the six [-high] [-low] monophthongs /e ɛ ʌ ɤ o ɔ/. The feature [back] separates three [-back] vowel /i e ɛ/ from the rest of the monophthongs. The three features [high], [low], and [back], however, can not separate the vowels /a/ and /ɤ/ from the vowels /ɐ/ and /ʌ/, which are different in their phonotactics constraints as /a ɤ/ can appear in open syllables and /ɐ ʌ/ can not, and in their formants and duration values, as described in chapter 4.

These four vowels can be distinguished using the [tense] feature since [+tense] can be displayed phonetically with longer duration. The [+tense] vowels /a ɤ/ have mean duration values of 142ms and 159ms in open syllables, respectively and 106ms and 123ms in closed syllables, respectively. Meanwhile, the [-tense] vowels /ɐ ʌ/ can only appear in closed syllables and their mean duration values in these closed syllables are 79ms and 74ms, respectively. The feature [round] is necessary for distinguishing vowel /u/ from its unrounded counterpart /ɯ/. Half of the [+back] monophthongs have [+round] and it plays an important role in Vietnamese phonology, which I will explain in part 2 of this chapter.

Vowel	[high]	[low]	[tense]	[back]	[round]
i	+	-	+	-	-
e	-	-	+	-	-
ɛ	-	-	-	-	-
a	-	+	+	+	-
ɐ	-	+	-	+	-
ʌ	-	-	-	+	-
ɤ	-	-	+	+	-
ɯ	+	-	-	+	-
u	+	-	+	+	+
o	-	-	+	+	+
ɔ	-	-	-	+	+

Table 5.4: Distinctive features of Vietnamese monophthongs

Chapter 4 shows that the contour vowels /ie/, /ɯɤ/, and /uo/ behave phonetically like monophthongs: they all reach the first and second formant targets, and they have the same pattern of duration change for the entire phone and for each component in different environments. These contour vowels thus should be included into the natural class of the eleven monophthongs, but the contour vowels are separated from the monophthongs by an additional feature, which defines the natural class of contour vowels.

The difference between the contour vowels and the monophthongs is the change in height within the contour vowels, moving from a [+high] position to a [-high] position within the same front-back area. This internal contour of height in the contour vowels can be analyzed as two separate values of [high] attached to the place node. This analysis of contour vowels as contour segments is based on the analysis of contour segments and their branching root nodes by Sagey (1986, p. 93). Sagey's analysis of affricates shows that each affricate is "a sequence of articulations on a single timing unit" (Sagey 1986, p. 81), and that both of the [-cont] and [+cont] values of the affricate would be attached to the root node of the sequence. Similarly, monophthongs as a non-contour segment would have either [+high] or [-high], while contour vowels /ie/, /ɯɤ/ and /uo/ would have two values of [high] attached to the [VPlace] node, marked as having [±high] value in the feature geometry. The [VPlace] node is a separate node containing all the vowel features. This [VPlace] lies underneath the [Place] node for both vowels and consonants (Steriade 1987, Clements 1991, Odden 1991, Ni Chiosáin and Padgett 1993). This internal contour is also explained by the dynamic trajectories of the contour vowels, as seen in figure 4.23 in chapter 4, starting around high vowel positions heading toward mid-high positions while maintaining the same [back] value.

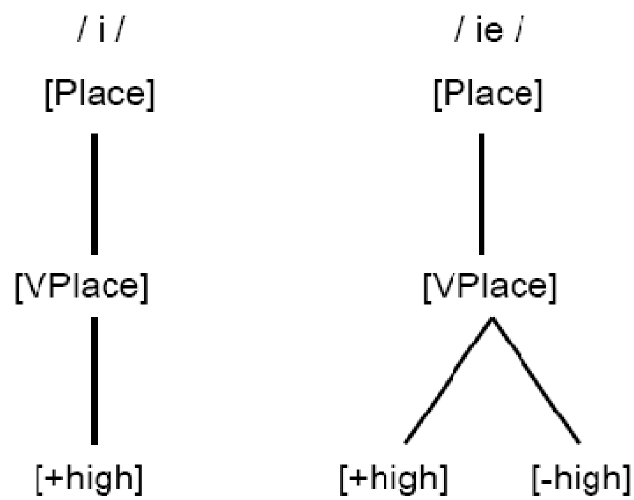


Figure 5.3: Examples of [high] values of monophthong /i/ and contour vowel /ie/

As for the feature [back], I first divide the vowel space vertically into two areas: a [-back] area, corresponding to the front positions of the tongue, and the [+back] area, corresponding to the central and back position of the tongue body. The eleven monophthongs and the three contour vowels thus divided into a [-back] group with /i e ɛ a/ and /ie/, a [+back] group of /ɐ ʌ ɤ ɯ u o ɔ/ and /ɯɤ, uo/. The [-back] group can be divided into a [+back, -round] group of /ɐ ʌ ɤ ɯ/ and /ɯɤ/, and a [+back, +round] group of /u o ɔ/ and /uo/.

Vowel	[high]	[low]	[tense]	[back]	[round]
/ie/	±	-	+	-	-
/i/	+	-	+	-	-
/e/	-	-	+	-	-
/ɛ/	-	-	-	-	-
/a/	-	+	+	+	-
/ɐ/	-	+	-	+	-
/ʌ/	-	-	-	+	-
/ɤ/	-	-	+	+	-
/ɯ/	+	-	-	+	-
/ɯɤ/	±	-	-	+	-
/uo/	±	-	+	+	+
/u/	+	-	+	+	+
/o/	-	-	+	+	+
/ɔ/	-	-	-	+	+

Table 5.5: Distinctive features of Vietnamese monophthongs and contour vowels

5.1.3. Hierarchical syllable structure in Vietnamese

Vietnamese syllables can be mapped into a hierarchy structure, in which each syllable has one to two moras. This hierarchical structure corresponds to the syllabic structure $(C_1)V(C_2)$. The onset consonant can either be a single consonant or a onset cluster, which has one consonant plus a glide /w/. The consonant in the onset cluster can be any of the Vietnamese consonants, for example, /kw/, /hw/, /ŋw/ and /lw/.

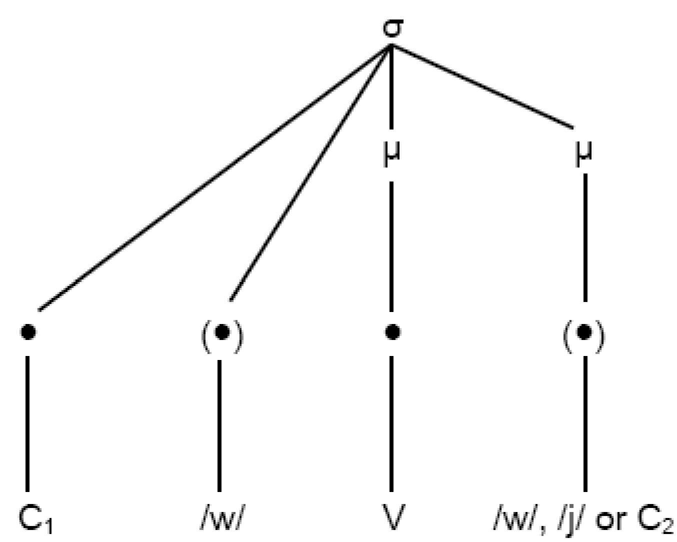


Figure 5.4: Basic hierarchical structure for Vietnamese syllables

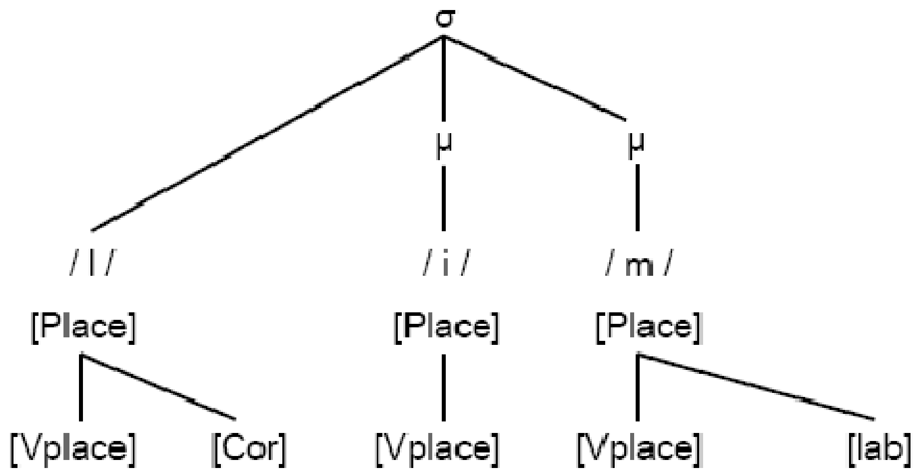


Figure 5.5: Example of *lim* /lim˧/ "a type of wood"

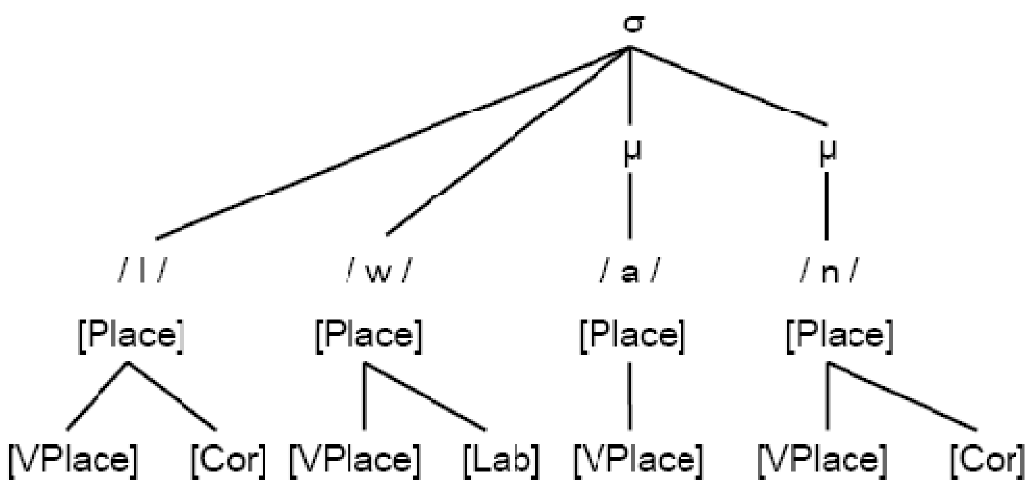


Figure 5.6: Example of *loan* /lwan˧/ "a type of bird"

When V is a monophthong, the second timing unit is optional, which can be occupied by a consonant, hence C_2. Since there are two monophthongs, i.e. /ɐ ʌ/, that can not appear in open syllables and they have the shortest duration among all eleven Vietnamese monophthongs, there might be a constraint on minimal mora weight required for each Vietnamese syllable and the two short vowels can not fulfill the minimal mora weight requirement by themselves. However, I am not going to investigate this possible minimal mora weight requirement in this study.

When the syllable contains a diphthong, the vowel component would occupy the first mora of the syllable. The glide component would occupy the second mora, thus the syllable can not have another consonantal coda. This is why all Vietnamese diphthongs only appear in open syllables.

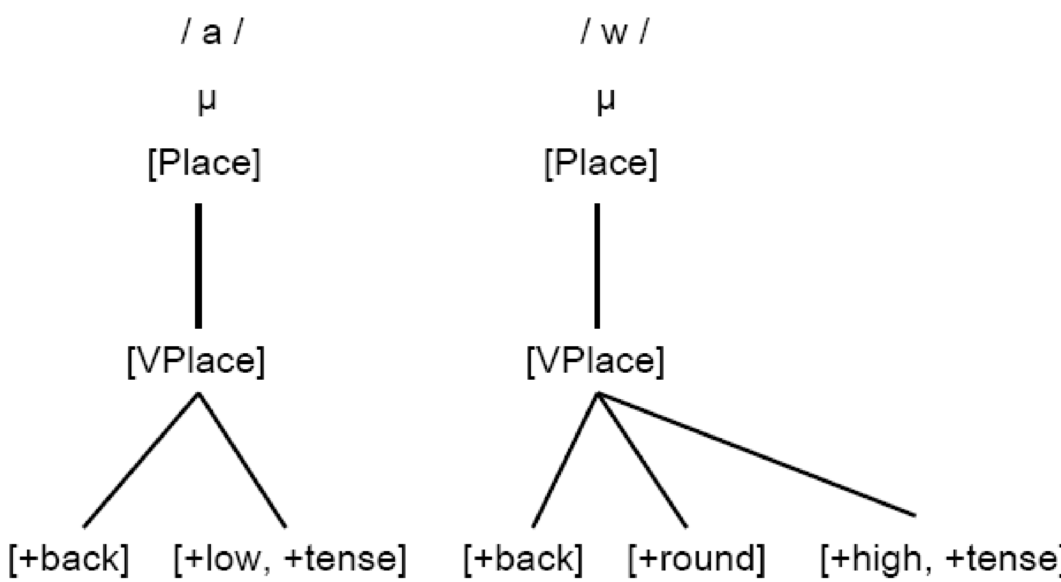


Figure 5.7: An example of /aw/

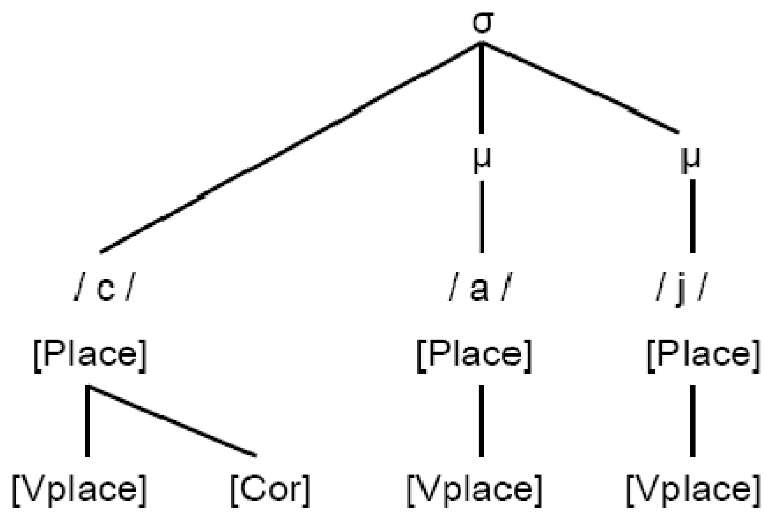


Figure 5.8: Example of *chai* /caj˧/ "bottle"

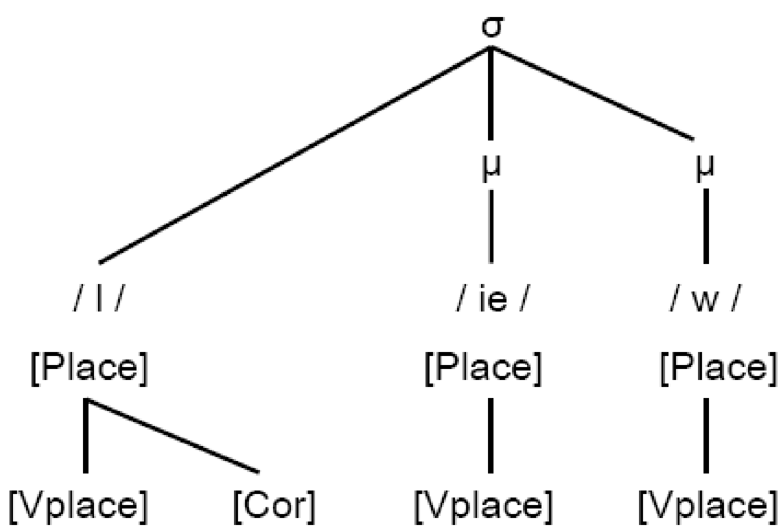


Figure 5.9: Example of *liêu* /liew˧/ "a proper name"

5.2. Evidence for the feature geometry in the language

5.2.1. Labio-velarization

Labio-velarization in Vietnamese is a phonological process in which the velar codas are labio-velarized after round vowels. As consonants and vowels all share the same set of VPlace features (Ni Chiosáin & Padgett 1993), the VPlace feature, which has the [+round] value of the vowels, is shared by the [Dor] consonant, making the consonant a complex consonant with both [+round], which is translated as [Lab], and [Dor] values.

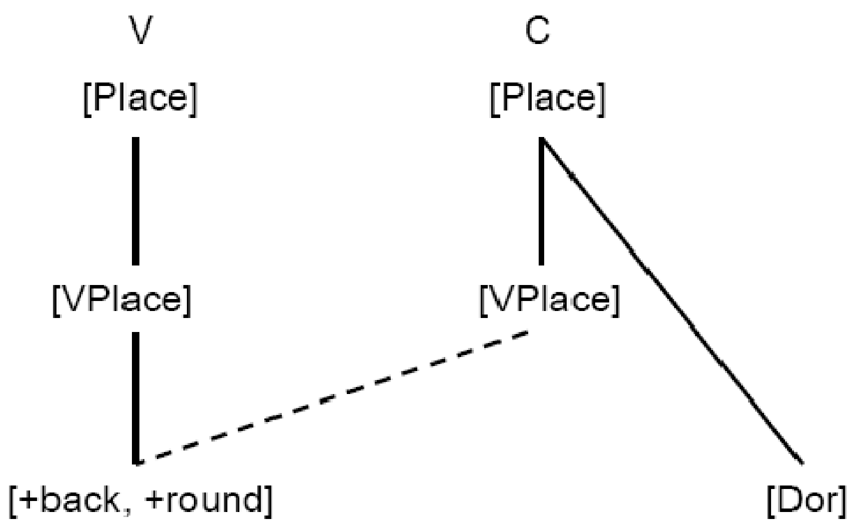


Figure 5.10: Labio-velarization rule

	Original form	→	Labio-velarized form
1.	*xong* /sɔŋ ˧/	→	[sɔŋm ˧]
2.	*trong* /cɔŋ ˧/	→	[cɔŋm ˧]
3.	*múc* /muk ˧˥/	→	[mukp ˧˥]
4.	*công* /koŋ ˧/	→	[koŋm ˧]

Table 5.6: Examples of the labio-velarization rule

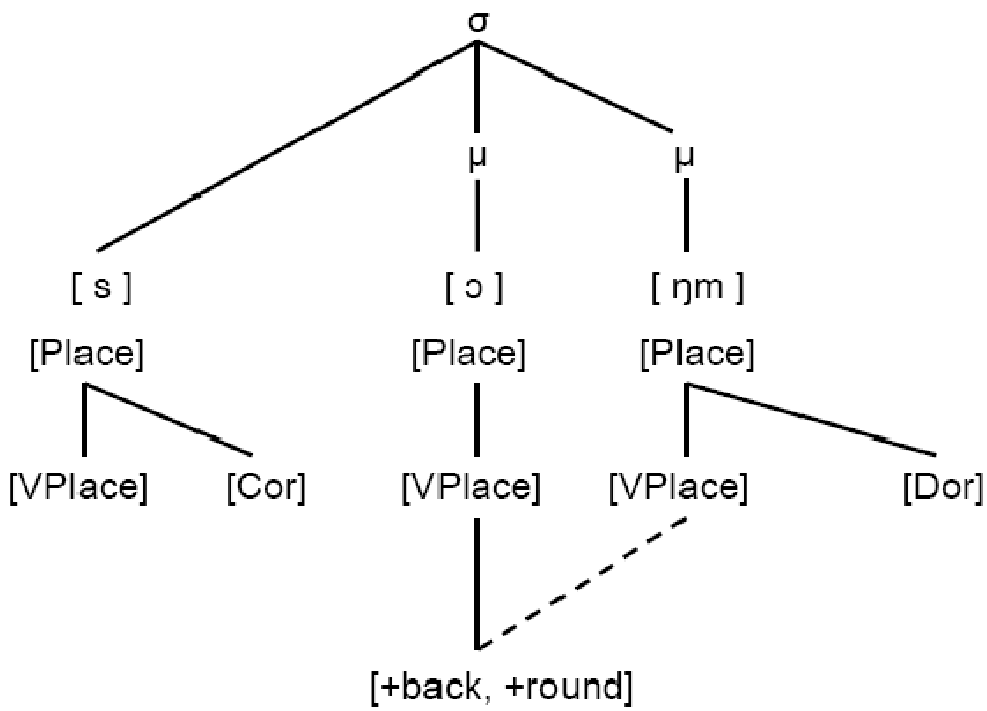


Figure 5.11: Example of *xong* /sɔŋ ˧/, labio-velarized as [sɔŋm ˧]

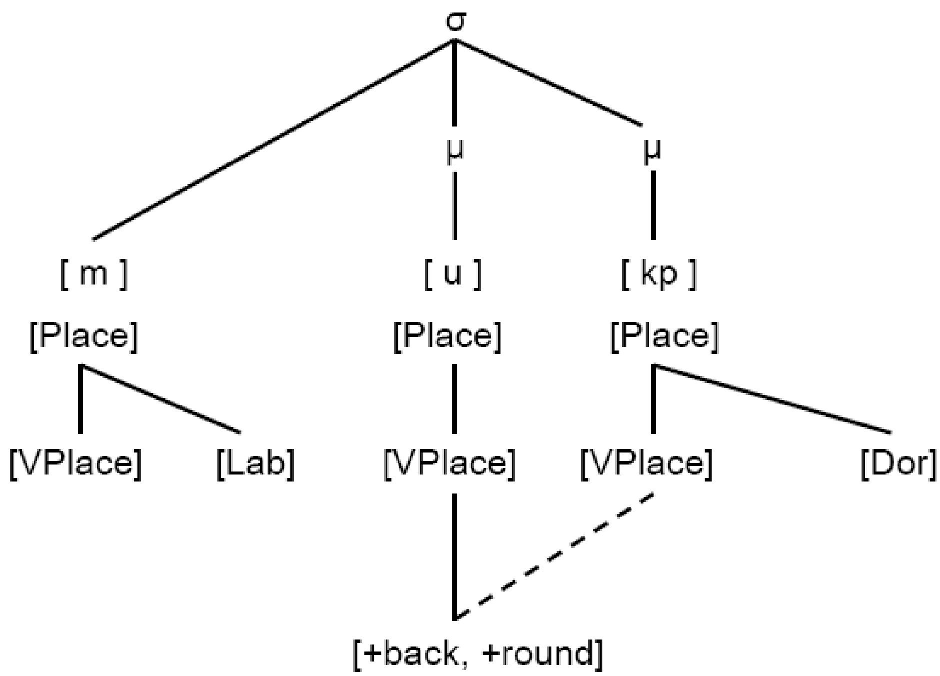


Figure 5.12: Example *"múc"* /muk ˧˥/, labio-velarized as [mukp ˧˥]

If Vietnamese contour vowels and Vietnamese monophthongs belong to a natural class, a phonological rule that applies to monophthongs should affect the contour vowels as well unless the rule is blocked by some other processes. In the case of the labio-velarization rule, in which the [+round] feature is spread onto the [Dor] consonant via sharing of the VPlace features between the vowel and consonant, it is actually blocked in the case of the contour vowel /uo/.

The contour vowel /uo/ is marked as a contour vowel with a [±high] value for height. If the [+round] value at the VPlace node is spread onto the [Dor] consonant, as happening under the labio-velarization rule, the syllable would either have one mora with a [±high] value and one mora with [Lab, Dor] complex value, or one complex segment with [Lab, Dor] value for place and [±high] value for height. This level of complexity is marked in languages in general, thus the labio-velarization rule is blocked after the contour vowel /uo/ in Vietnamese.

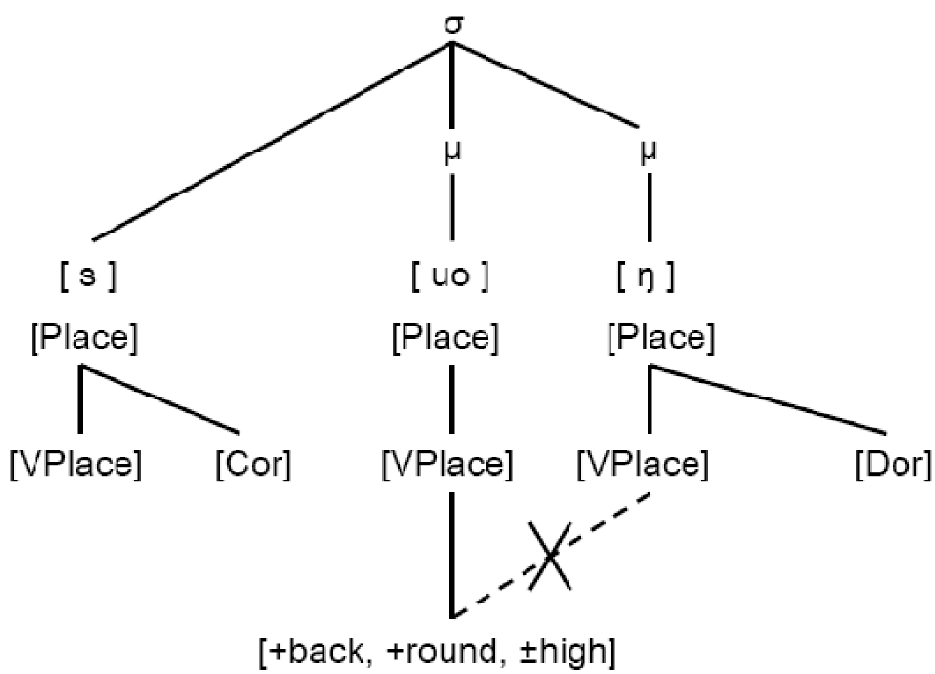


Figure 5.13: Example *suông* /suoŋ˧/ "flow smoothly", not affected by the Labio-velarization rule

5.2.2. Reduplication words in Vietnamese and their patterns

Reduplication in Vietnamese is a historically productive process with forms ranging from full reduplication to partial reduplication (Thompson 1965). There are "about 5000 reduplicated words in Vietnamese, as published in the Dictionary of Vietnamese Reduplication (Hoàng *et al* 1998), which, compared with a regular dictionary of 50,000 entries, is a significant amount" (Alves 1999, p. 223). The reduplicated words are mostly learned by the modern native speakers and were recorded from folklores and early Vietnamese literature, but there are a small number of these reduplications collected from the modern literature after the year 1945, which might have been created by those authors. These sources are listed in the index of the Dictionary of Vietnamese Reduplicated Words (Hoàng *et al* 1998, p. 413).

Full reduplication is a process, in which a full syllable including tone is duplicated to express a different amount or degree, a continuous or repetitive action, or different intensity levels. Otherwise, any part of the syllable can be duplicated to reduce or intensify the meaning but this partial reduplication process is no longer productive. Partial reduplication can be the duplication of just the onset, of just the rime, of the onset and the coda, or of everything but tones. The treatment of syllable components in the Vietnamese reduplication process is consistent with the syllabic hierarchical structure presented in figure 5.5 of this chapter.

Full reduplication, as demonstrated with examples in table 5.7, is a productive process in Vietnamese, and in this process, the entire syllable is copied to create a new word with an increased level of intensity. When nouns are fully duplicated, they mean "many of that object". When a verbal is duplicated, the new action is a repetitive and continuous action. When an adjective is duplicated, the new quality is a more intense one.

No.	Base	Base meaning	Reduplicated word	New meaning	Duplication
1	*nhăn*	wrinkle	*nhăn nhăn*	slightly wrinkled	whole word
2	*im*	quiet	*im im*	all quiet	whole word
3	*hôi*	stink	*hôi hôi*	a little stinky	whole word
4	*quýnh*	nervous	*quýnh quýnh*	clumsy	whole word
5	*khuều*	grab	*khuều khuều*	keep reaching	whole word

Table 5.7: Examples of full reduplication

The closest process to full reduplication is the reduplication of all the segments except for tones, demonstrated with examples in table 5.8, and this process is semi-productive with a similar semantic effect to that of the full reduplication.

No.	Base	Base meaning	Reduplicated word	New meaning	Duplication
1	*nhăn*	wrinkle	*nhăn nhẳn*	stomachache	all but tone
2	*im*	quiet	*im ỉm*	really quiet	all but tone
3	*hống*	empty	*hông hống*	uncovered	all but tone
4	*quàng*	fast	*quáng quàng*	hurry	all but tone
5	*khuều*	grab	*khuều khuễu*	clumsily tall	all but tone

Table 5.8: Examples of full reduplication except for tone

Partial reduplication is a process that changes the meaning of the word by duplicating certain segments in the syllable to create a new two-syllable word. The more different the two syllables are from one another, the more different the new meaning is. In the partial reduplication of the syllable onset, the onsets could be consonant-only or consonant-and-/w/. The duplicated segments are generally the identical duplications of the original segments.

No.	Base	Base meaning	Reduplicated word	New meaning	Duplication
1	*nhăn*	wrinkle	*nhăn nhó*	grumpy	# [ɲ]
2	*nèo*	pursuade	*nì nèo*	beg	# [n]
3	*hôi*	stink	*hôi hám*	stinky	# [h]
4	*quýnh*	nervous	*quýnh quáng*	rushing	# [kw]
5	*khoèo*	grab, curve	*khoằng khoèo*	unorganized	# [xw]

Table 5.9: Examples of reduplication of the onsets

Other reduplication processes duplicate a full segment or parts of the segments of the syllables to create new meanings. One can just duplicate the onset, the coda and the tone of a syllable, and replace the prominent vowel target. In the case of the example 1 in table 5.10 below, the rising tone, the onset consonant /t/ and the coda consonant /m/ are duplicated while the vowel /ɛ/ is replaced by the vowel /ɔ/.

No.	Base	Base meaning	Reduplicated word	New meaning	Duplication
1	*tém*	tight	*tóm tém*	closed lips	[t_m ˧˥]
2	*hóc*	obstructive	*hóc hách* [1]	object to sth.	[h_c ˧˥]
3	*ách*	pressured	*óc ách* [1]	full of water	[#_k ˧˥]
4	*hốc*	hole	*hốc hác*	bony (face)	[h_k ˧˥]
5	*khuều*	grab	*khuều khoào*	weak grabbing	[xw_w ˧˩]

Table 5.10: Examples of reduplication by vowel replacement

Note: [1] The interchanging /c/ and /k/ are results of a different phonological rule, which I will not discuss in this study.

The last group of duplication processes includes the reduplication of the syllable rimes, i.e. the examples in table 5.11, and the reduplication of the onset and vowels but not coda, i.e. the examples in table 5.12. Neither of these processes duplicate the tones. The tones have a couple of patterns which they follow, such as replacing the original tones with one of the tones in the same tone register. The examples in table 5.11 show that the reduplication process groups the glide /w/ with the onset consonant. Therefore, in the example 2 of table 5.11, the rime /aŋ/ in *mang* /maŋ ˧/ is matched by the reduplication process with the base rime /aŋ/ of *hoang* /hwaŋ ˧/.

No.	Base	Base meaning	Reduplicated word	New meaning	Duplication
1	*tịch*	quiet	*tịch mịch*	quiet, calm	[_ic ˧ʔ˩]
2	*hoang*	wild	*hoang mang*	unsure	[_aŋ ˧]
3	*tốc*	speedy	*hộc tốc*	rushing, hurry	[_ok]
4	*quýnh*	nervous	*luýnh quýnh*	clumsy, unsure	[_iɲ ˧˥]
5	*khuều*	grab	*lều khuều*	too tall	[_ew ˧˩]

Table 5.11: Examples of reduplication of the rimes

No.	Base	Base meaning	Reduplicated word	New meaning	Duplication
1	*oạch*	fall	*oành oạch*	keep falling	[wa_]
2	*dặc*	far	*dằng dặc*	very far	[zɐ_]
3	*hốc*	hole	*hông hốc*	uncovered	[ho_]
4	*quýnh*	nervous	*quýnh quýt*	really hurry	[kwi_ ˧˥]
5	*bộp*	loud noise	*bồm bộp*	noisy	[ɓo_]

Table 5.12: Examples of reduplication by coda replacement

The reduplication processes in Vietnamese also treat the contour vowels /ie/, /ɯɤ/ and /uo/ in the same way they treat monophthongs. First of all, there are reduplications in which the contour vowels are copied as a whole in the same way as the monophthongs, i.e. examples in table 5.13. Second, there are duplicated tokens, as shown in table 5.14, in which the contour vowels and monophthongs are paired and given the same weight in the structures. These behaviors show that the reduplication processes give monophthongs and contour vowels the same treatment, and for reduplication purposes, contour vowels and monophthongs belong to the same class.

No.	Base	Base meaning	Reduplicated word	New meaning	Duplication
1	*hiếc*	tilt	*hiếc hiếc*	a little tilted	full redup.
2	*ướt*	wet	*ướt ướt*	a little wet	full redup.
3	*hiếng*	lopsided	*hiêng hiếng*	a little tilted	all but tone
4	*ướm*	try out	*ươm ướm*	test opinion	all but tone
5	*hiếc*	tilt	*hiêng hiếc*	tilted	coda replacement
6	*ướt*	wet	*ươn ướt*	a little damp	coda replacement

Table 5.13: Examples of different reduplication processes with the contour vowels

No.	Base	Base meaning	Reduplicated word	New meaning	Duplication
1	*hiển*	clear	*hiển hách*	lustrious	# [h]
2	*thiết*	interested	*thiết tha*	insistent	# [t^h]
3	*tươm*	neat	*tinh tươm*	neat, organized	[t_n/m ˧]
4	*ướt*	wet	*ướt át*	very wet	[#_t ˧˥]
5	*chiều*	please	*chiều chuộng*	spoil	# [c_]
6	*gớm*	gross	*gớm guốc*	ugly, terrible	# [ɣ_]

Table 5.14: Examples of reduplication with interchangeable monophthongs and contour vowels

5.2.3. Simplifying the contour vowels

Another evidence for a natural class which includes all the Vietnamese monophthongs /i e ɛ a ɐ ʌ ɤ ɯ u o ɔ/ and contour vowels /ie ɯɤ uo/ is the existence of vowel reduction processes in various Vietnamese dialects. While tokens of the vowel-glide diphthongs do not show variation across dialects, the reduction of the contour vowels is quite common in various Vietnamese dialects.

Assuming that the current Vietnamese orthography system, "Chữ Quốc Ngữ", was an accurate reflection of the standard Vietnamese dialect in the northern part of Vietnam in the 19th century, the differences between current pronunciations and the writing system actually reflect various completed sound changes from the 19th century contemporary sounds to the current sound system of the northern dialect.

First, the Dissimilation process, i.e. the examples in table 5.16 below, leads to the change of the contour vowel /ɯɤ/ into the [-back] contour vowel /ie/ in Hanoi dialect. The contour vowel /ɯɤ/ of the Hanoi dialect remains intact at syllable-final positions and before most sound segments in closed syllables, i.e. the examples in table 5.15. However, it changes to /ie/ before a labial-velar glide /w/, e.g. the word *hươu* /hɯɤw ˧/ "deer" is actually /hiew ˧/. Similarly, the monophthong /ɯ/ in the diphthong /ɯw/ is also changed into /i/ in Hanoi dialect, e.g. the word *hưu* /hɯw ˧/ "retirement" is actually /hiw˧/. These changes in Hanoi dialect reflect a Dissimilation process of the [+back] feature before a similar [+back] segment without affecting any other features, e.g. the [±high] feature of /ɯɤ/ stays intact as it changes to /ie/.

No.	Orthography	Gloss	IPA	Hanoian pronunciation
1	*chứa*	store	/cɯɤ ˧˥/	[cɯə ˧˥]
2	*tương*	sauce	/tɯɤŋ ˧/	[tɯɤŋ ˧]
3	*trượt*	slide	/cɯɤt ˧ʔ˩/	[cɯɤt ˧ʔ˩]
4	*chườm*	ice (verb)	/cɯɤm ˧˩/	[cɯɤm ˧˩]

Table 5.15: Examples of contour vowel /ɯɤ/ in syllable-final positions and before non-/w/ codas

No.	Orthography	Gloss	IPA	Hanoian pronunciation
1	*hươu*	deer	/hɯɤw ˧/	[hiew ˧]
2	*rượu*	wine	/rɯɤw ˧ʔ˩/	[riew ˧ʔ˩]
3	*hưu*	retirement	/hɯw ˧/	[hiw ˧]
4	*lưu*	save	/lɯw ˧/	[liw ˧]

Table 5.16: Examples of contour vowel /ɯɤ/ and /ɯ/ before a /w/ coda

This Dissimilation process is active within the domain of VPlace node only and [Dor] consonants such as velar voiceless stop /k/ and velar nasal /ŋ/ do not trigger this process. Similarly to the argument in Sagey (1986, p. 141) for the case of Simplification in Nootka, the Dissimilation process in Vietnamese is actually the delinking of the [+back] value from the VPlace node of the first mora, which is triggered by the existence of the [+back] value in the VPlace node of the second mora. This Dissimilation process supports the hierarchical grouping of /ɯɤ/ and /w/ in /ɯɤw/ as well as /ɯ/ and /w/ in /ɯw/ under the two separate moras, as shown in the figure 5.14 and 5.15 below.

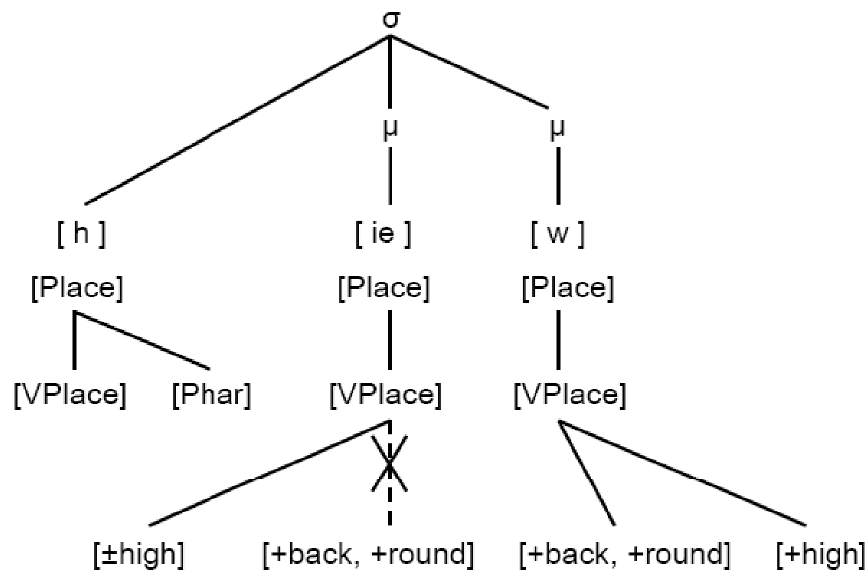


Figure 5.14: Example 1 *hươu* /hɯɤw ˧/ ("deer") as [hiew ˧]

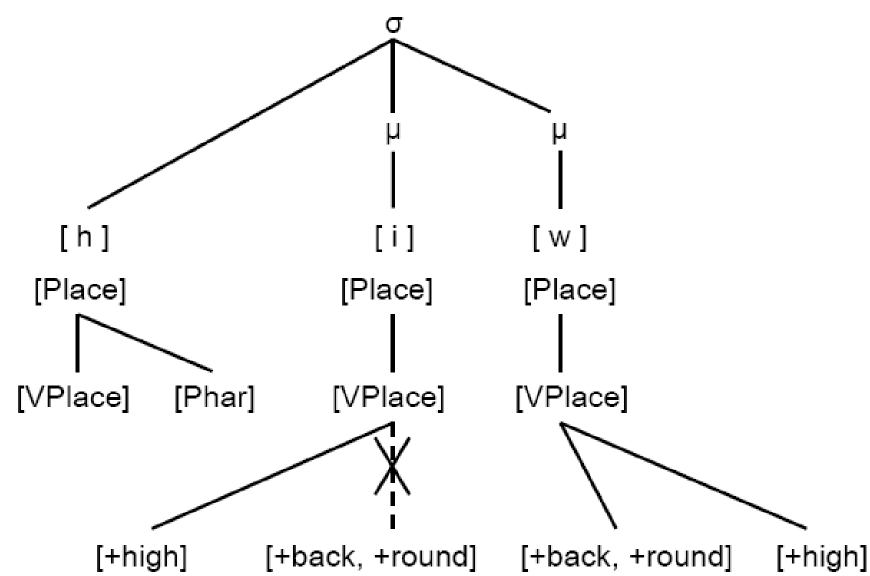


Figure 5.15: Example 2 *hưu* /hɯw ˧/ "retirement" as [hiw ˧]

Beside the Dissimilation process, there is a Dorsalization process that changes the contour vowels /ie/, /ɯɤ/, and /uo/ into /iʌ/, /ɯʌ/ and /uʌ/, respectively, in syllable-final positions ended in glottal stops, or if they precede velar consonants /k/ and /ŋ/ in all dialects. And since the [Cor] consonants in Saigon dialect are first velarized in syllable-final positions, the Dorsalization process applies to all contour vowels with [Cor] or [Dor] coda consonants in Saigon dialect, as shown in the first five examples in table 5.17, 5.18, and 5.19. This process, however, is not analyzed further in this dissertation.

No.	Orthography	17th century	Hanoi	Saigon	gloss
1	tia[1]	/tie ˧/	[tiʌ ˧]	[tiʌ ˧]	ray
2	tiết	/tiet ˧˥/	[tiet ˧˥]	[tiʌk ˧˥]	secrete
3	tiến	/tien ˧˥/	[tien ˧˥]	[tiʌŋ ˧˥]	progress
4	tiếc	/tiek ˧˥/	[tiʌk ˧˥]	[tiʌk ˧˥]	regret
5	tiếng	/tien ˧˥/	[tiʌŋ ˧˥]	[tiʌŋ ˧˥]	sound
6	tiếp	/tiep ˧˥/	[tiep ˧˥]	[tip ˧˥]	continue
7	tiêm	/tiem ˧/	[tiem ˧]	[tim ˧]	inject
8	tiêu	/tiew ˧/	[tiew ˧]	[tiw ˧]	pepper

Table 5.17: The contour vowel /ie/ in different environments

Note: [1] In open-syllables, /ie/ is written in orthography as *ia*

No.	Orthography	17th century	Hanoi	Saigon	Gloss
1	*cứa* [1]	/kɯɤ ˧˥/	[kɯʌ ˧˥]	[kɯʌ ˧˥]	slice
2	*lướt*	/lɯɤt ˧˥/	[lɯɤt ˧˥]	[lɯʌk ˧˥]	glide
3	*mướn*	/mɯɤn ˧˥/	[mɯɤn ˧˥]	[mɯʌŋ ˧˥]	hire
4	*cước*	/kɯɤk ˧˥/	[kɯɤk ˧˥]	[kɯʌk ˧˥]	custom
5	*cường*	/kɯɤŋ ˧˩/	[kɯɤŋ ˧˩]	[kɯʌŋ ˧˩]	strong
6	*cướp*	/kɯɤp ˧˥/	[kɯɤp ˧˥]	[kɯp ˧˥]	rob
7	*cườm*	/kɯɤm ˧˩/	[kɯɤm ˧˩]	[kɯm ˧˩]	bead
8	*hươu*	/hɯɤw ˧/	[hiew ˧]	[hɯw ˧]	deer
9	*cưới*	/kɯɤj ˧˥/	[kɯɤj ˧˥]	[kɯj ˧˥]	marry

Table 5.18: The contour vowel [ɯɤ] in different environments

Note: [1] In open-syllables, /ɯɤ/ is written in orthography as *ưa*

No.	Orthography	17th century	Hanoi	Saigon	gloss
1	*tua* [1]	/tuo ˧/	[tuʌ ˧]	[tuʌ ˧]	rewind
2	*tuốt*	/tuot ˧˥/	[tuot ˧˥]	[tuʌt ˧˥]	peel
3	*khuôn*	/xuon ˧/	[xuon ˧]	[xuʌn ˧]	mold
4	*cuốc*	/kuok ˧˥/	[kuʌk ˧˥]	[kuʌk ˧˥]	hoe
5	*xuống*	/suoŋ ˧˥/	[suʌŋ ˧˥]	[suʌŋ ˧˥]	go down
6	*cuỗm*	/kuom ˧ʔ˥/	[kuom ˧ʔ˥]	[kum ˧ʔ˥]	steal
7	*đuôi*	/ɗuoj ˧/	[ɗuoj ˧]	[ɗuj ˧]	tail

Table 5.19: The contour vowel [uo] in different environments

Note: [1] In open-syllables, /uo/ is written in orthography as *ua*

The phonological processes that argue for the status of the [±high] feature in contour vowels is the Simplification process and the Assimilation process, both of which affect the [±high] value of the contour vowels. These processes are active only in the Saigon dialect with the Simplification process triggered by the [Lab] feature under the

[Place] node of the consonants, and the Assimilation process triggered by the [+high] value under the [VPlace] node of the glide segments.

In Saigon dialect, when contour vowels are followed by [Lab] consonants, which is either /p/ or /m/ in Vietnamese, the Simplification process keeps the [+high] value, which is the default height feature for vowels, and discard the [-high] part of the [±high] value of the contour vowels, as shown in figure 5.16.

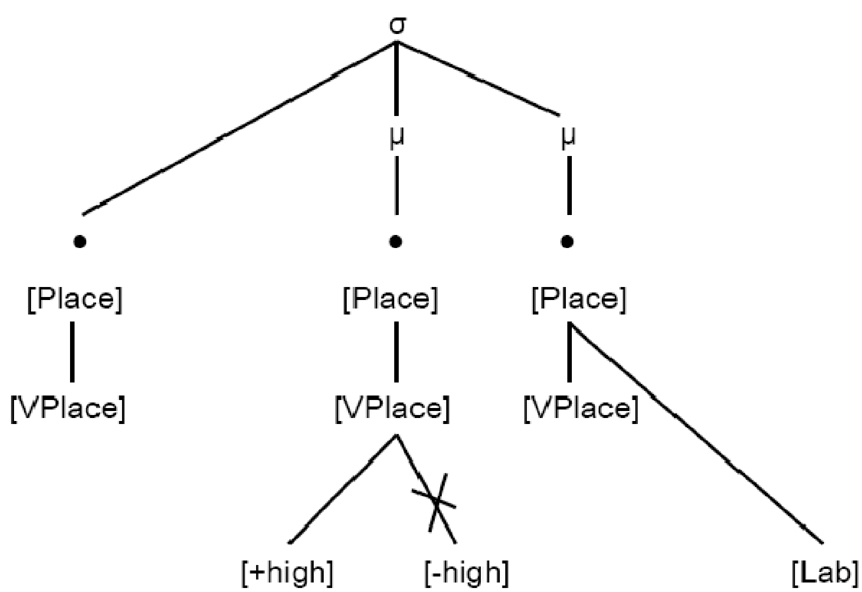


Figure 5.16: Process of Simplification

If preceding the glide /u/ and /i/, however, the contour vowels in Saigon dialect also go through an Assimilation process, which delinks the entire [±high] contour value, and associates the vowel with the [+high] value of the glides. The examples of these are in the shaded part of tables 5.17, 5.18 and 5.19.

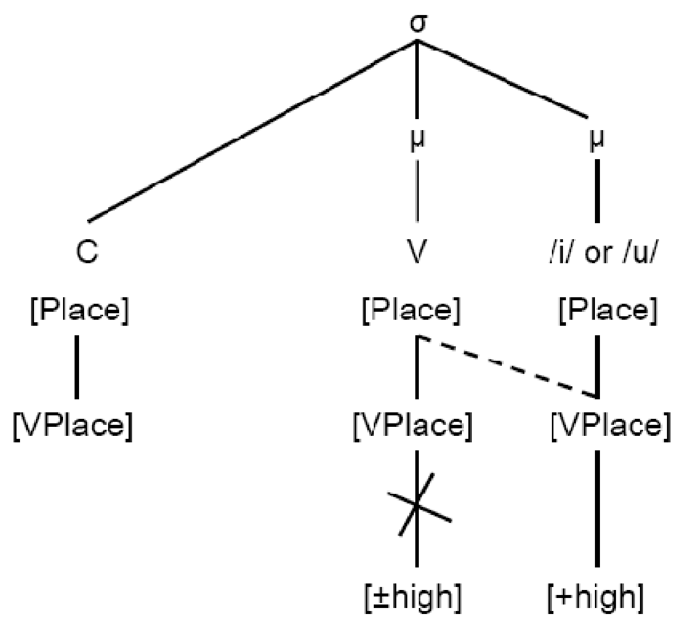


Figure 5.17: Process of Assimilation

These two phonological processes, Simplification and Assimilation, show that the contour vowels are treated by the Vietnamese phonological system as one unit with a [±high] feature under the [VPlace] node and this contour feature can be split into a [+high] value and a [-high] value, which explain the dialectal variations in contour vowels.

5.2.4. Analysis of contour vowels as actual clusters of monophthongs

One might argue that instead of the [±high] feature, the contour vowels should be analyzed as actual clusters of monophthongs, which occupy two mora nodes in the syllables. The coda and the onset are linked directly to the syllable node and they do not count toward the mora number of the syllables. The syllable structure for Vietnamese would then be $(C_1)(w)V_1(V_2)(C_2)$ and it could be one to two moras. The monophthongs would occupy V_1, and the contour vowels would occupy V_1 and V_2. The vowel-glide sequence would have to occupy V_1 and C_2 to cover the cases of diphthongs and triphthongs.

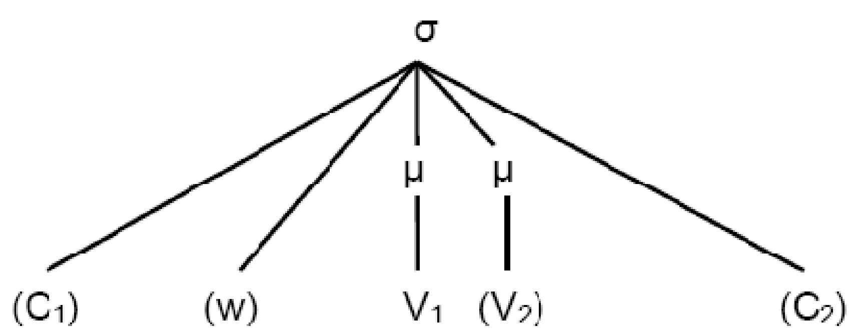


Figure 5.18: Alternative syllable structure for Vietnamese

This analysis implies that the glides in diphthongs and triphthongs have the same moraic weight as the coda consonant. In addition, this analysis does not explain the non-triggered labio-velarization process in the case of /uo/, as the [VPlace] nodes of both V_1 and V_2 have the [+round] value and that value should be able to be shared with the [Dor] consonant C_2. One would need to stipulate a reason why C_2 is not labialized if both V_1

and V_2 has [Lab] value, but why it is labialized when only V_1 alone is occupied and have [Lab] value. This alternative analysis also has to imply that mora weight is not important in Vietnamese to explain the interchangeability quality of monophthongs and contour vowels in reduplication in table 5.14. Last, this analysis needs to either assign mora weight to [Lab] consonants and the /i u/ glides or give a reason to forbid a two-mora nucleus before [Lab] consonants and the /i u/ glides as a dialectal feature in the Saigon dialect to explain the reduction of contour vowels into monophthongs. Thus this alternative analysis is inferior to the analysis of including the contour vowels and monophthongs of Vietnamese into one natural class.

5.2.5. Discussion

Contour vowels [ie], [ɯɤ] and [uo] traditionally have been analyzed as diphthongs in Vietnamese, which means they would occupy two single-mora root nodes in the syllables, but these contour vowels should be analyzed as monophthongs, occupying only one single-mora root node with a [±high] value underneath the [VPlace] node, based on their behaviors in several phonological processes in Vietnamese. Under this analysis, the Vietnamese word *nguyên* /ŋwien ˧/ "whole" would have similar moraic and syllabic structure as the English word *Gwen,* as shown in figure 5.19 and 5.20 below.

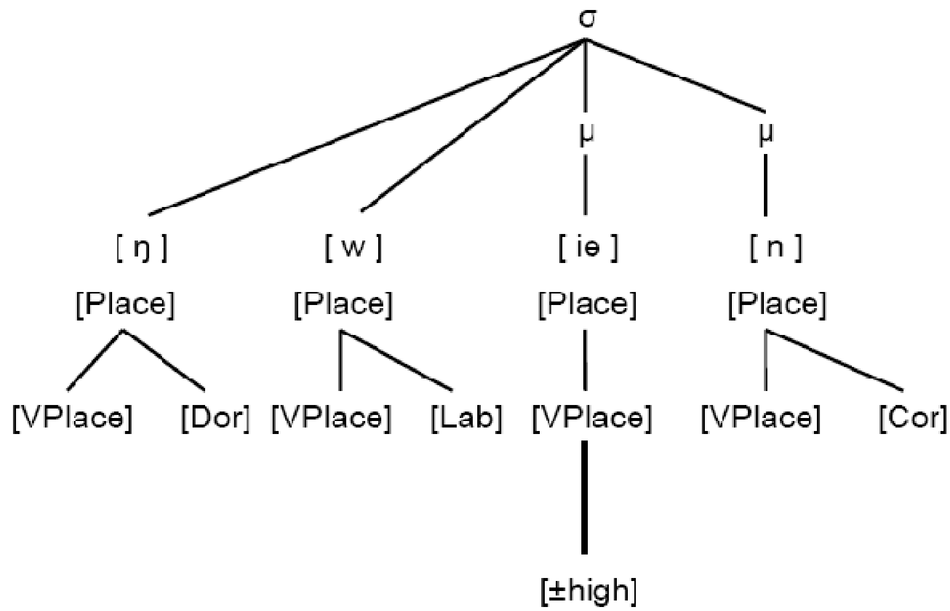


Figure 5.19: Syllabic structure of the Vietnamese word *nguyên* /ŋwien ˧/ "whole"

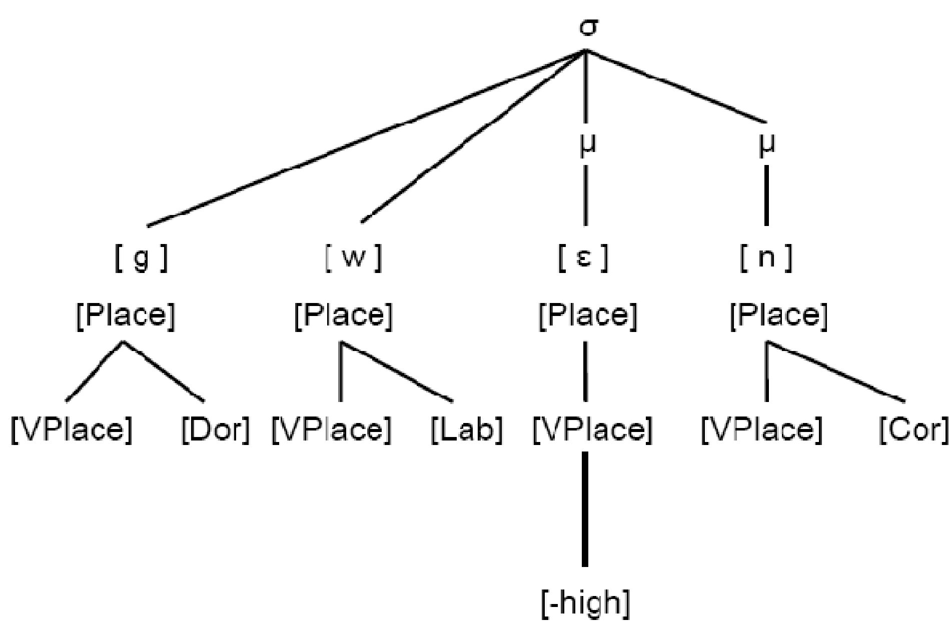


Figure 5.20: Syllabic structure of the English word *Gwen* /gwɛn/

First, these contour vowels can be defined as a separate natural class within the monophthong group based on the [±high] feature. This [±high] feature blocks the Labio-velarization process, which affects all [+back, +round] monophthongs except for the contour vowels. Second, the patterns in reduplication shows that contour vowels /ie/, /ɯɤ/ and /uo/ are duplicated in the same way monophthongs are. And last but not least, the dialectal reduction of contour vowels in Hanoi and Saigon dialect further support the equal role of contour vowels among the monophthongs, making the reduction process either the Simplification process to keep [+high] value only if preceding a [Lab] segment, or the Assimilation process to replace [±high] value with a [+high] value of the following high /i/.

With the separation the /w/-vowel sequences into an medial /w/ and a nucleus vowel, the structures in the Group 2 triphthongs, except for /wie/, are no longer triphthongs but vowel-glide diphthongs preceded by complex onsets containing /w/ (as shown in table 5.20). The triphthong /wie/ in the Group 2 Triphthongs is actually a monophthong with a complex consonant-/w/ onset. As for the Group 1 triphthongs, they are actually diphthongs with two moras, one attached to the contour vowels, i.e. /ie/, /ɯɤ/

and /uo/, one attached to the glides (table 5.21). The new diphthongs /iew/, /ɯɤw/, /uoj/ and /ɯɤj/ are added into the original fifteen diphthongs making the number of diphthongs in Vietnamese nineteen.

No.	Triph-thong	Previous analysis	Example	IPA	Gloss	New analysis
1	wiw	/w/ + Vw diphthong	*khuỷu*	/xwiw˧˩˧/	curve	/xw/ + /i/ + /w/
2	wew	/w/ + Vw diphthong	*khuều*	/xwew˧˩/	grab	/xw/ + /e/ + /w/
3	wɛw	/w/ + Vw diphthong	*khoèo*	/xwɛw˧˩/	hooked	/xw/ + /ɛ/ + /w/
4	waw	/w/ + Vw diphthong	*quạu*	/kwaw˧ʔ˩/	upset	/kw/ + /a/ + /w/
5	wɐw	/w/ + Vw diphthong	*quào*	/kwɐw˧˩/	scratch	/kw/ + /ɐ/ + /w/
6	wʌw	/w/ + Vw diphthong	*khuấu*	/xwʌw˧˥/	skim	/xw/ + /ʌ/ + /w/
7	waj	/w/ + Vj diphthong	*khoai*	/xwaj˧/	potato	/xw/ + /a/ + /j/
8	wɐj	/w/ + Vj diphthong	*khoáy*	/xwɐj˧˥/	cowlick	/xw/ + /ɐ/ + /j/
9	wʌj	/w/ + Vj diphthong	*khuấy*	/xwʌj˧˥/	stir	/xw/ + /ʌ/ + /j/

Table 5.20: Reanalysis of Group 2 triphthongs as diphthongs

Note: /wie/ is excluded from this table to be included table 5.21

No	Triph-thong	Previous analysis	Example	IPA	Gloss	New analysis
1	iew	VV diphthong + /w/	*xiêu*	/siew˧/	slanted	/x/ + /ie/ + /w/
2	ɯɤw	VV diphthong + /w/	*hươu*	/hɯɤw˧/	deer	/h/ + /ɯɤ/ + /w/
3	uoj	VV diphthong + /j/	*xuôi*	/suoj ˧/	flow	/s/ + /uo/ + /j/
4	ɯɤj	VV diphthong + /j/	*tươi*	/tɯɤj ˧/	fresh	/t/ + /ɯɤ/ + /j/

Table 5.21: Reanalysis of Group 1 triphthongs as diphthongs

Chapter 6

Conclusion remarks

While previous studies (Thompson 1965, Nguyễn Đình Hòa 1966, Han 1968) analyze Vietnamese as a language with eleven monophthongs, the phonetic and phonological evidences in this study support my hypothesis of re-analyzing Vietnamese as a language with fourteen monophthongs, nineteen rising diphthongs, and no triphthong. Three of the monophthongs, i.e. /ie/, /ɯɤ/ and /uo/, have the feature of [±high], setting themselves apart from the other eleven monophthongs /i e ɛ a ɐ ʌ ɤ ɯ u o ɔ/.

First, the phonetic study results of the six female Hanoi dialect speakers, collected as parts of the Vietnamese Speech Corpus, show that the contour vowels /ie/, /ɯɤ/ and /uo/ have two vowel components within themselves. Both of these vowel components reach the same formant targets, i.e. having similar mean formant values and dynamic trajectory structures, as those of the corresponding monophthongs. In addition, the individual duration of these vowel components is influenced by the surrounding environments in the same pattern as that of the monophthongs and that of the vowel components of diphthongs. There is no resemblance in phonetic properties between the components of the contour vowels and the glide components in diphthongs, i.e. stable duration without influence from the environment as well as glide-like formant targets, which act more like a guideline than a target. As my study shows that the behavior in formant quality and duration of contour vowels is more similar to that of the monophthongs instead of the diphthongs, the Vietnamese contour vowels /ie/, /ɯɤ/, and /uo/ should be placed in the same natural class with the Vietnamese monophthongs. In addition, the re-analysis of contour vowels as monophthongs also leads to the re-analysis of Vietnamese triphthongs as diphthongs, thus leaving no triphthong in the Vietnamese vowel system.

In the phonological analysis, the natural class of Vietnamese monophthongs and contour vowels has one additional [±high] feature designed to define the contour vowels themselves as a natural sub-class. The [±high] feature belongs to the [VPlace] node, which follow the [VPlace] theory in previous studies (Steriade 1987, Clements 1991, Odden 1991, Ni Chiosáin and Padgett 1993). This distinctive feature [±high] is actually a sequence of [+high] and [-high], similar to the analysis of affricates by Sagey (1986).

This [±high] feature explains the behavior of contour vowels in Vietnamese phonological processes such as Labio-velarization, Assimilation and Simplification.

Vietnamese contour vowels are treated as single phonological units with the same moraic weight as that of the monophthongs. Both contour vowels and monophthongs occupy the one single mora node in the syllable structure while the coda consonants and the glide components in the diphthongs occupy the optional second mora node. The reduplication process and the rhyming patterns in Vietnamese treat the three contour vowels and the monophthongs as interchangeable units. They can be rhymed with one another. They can go through the same duplication processes and they can be replaced interchangeably by one another in reduplication. Phonologically, the contour vowels should be grouped with the monophthongs into one natural class.

In conclusion, this corpus-based study shows that the Vietnamese vowel system is a system with fourteen monophthongs /i e ɛ a ɐ ɤ ʌ ɯ ɔ o u ie ɯɤ uo/ and nineteen diphthongs, comprising of ten vowel-/j/-glide diphthongs /aj ɐj ʌj ɔj oj ɤj ɯj uj ɯɤj uoj/ and nine vowel-/w/-glide diphthongs /aw ɐw ʌw ɛw ew iw ɯw iew ɯɤw/. More perception tests, especially discrimination tests, should be done to learn if the shapes or the directions of the dynamic trajectories can be the primary cues in the identification of Vietnamese monophthongs, contour vowels and diphthongs. In addition, a phonetic analysis of both male and female speakers of different dialects of Vietnamese would be important to learn whether these similarities and differences within the natural class of monophthongs and contour vowels are truly intrinsic characteristics or dialectal features instead. Extending this analysis to the diphthongs in other languages that are similar to the Vietnamese contour vowels would also help in understanding this natural sub-class within the monophthong class. These future studies would help in understanding of characteristics of these fourteen Vietnamese monophthongs /i e ɛ a ɐ ɤ ʌ ɯ ɔ o u ie ɯɤ uo/.

Bibliography

Andersen, Henning. 1972. Diphthongization. *Language* 48:1-50.

Andruski, J. E. and T. M. Nearey. 1992. On the sufficiency of compound target specification of isolated vowels in /bVb/ syllables. *The Journal of the Acoustical Society of America* 91:390-410.

Benguerel, A. P. and T. U. McFadden. 1989. The effect of coarticulation on the role of transitions in vowel perception. *Phonetica* 46:80-96.

Borzone de Manrique, Ana Maria. 1976. Acoustic study of /i,u/ in the Spanish diphthong. *Language and Speech* 19:121-128.

Borzone de Manrique, Ana Maria. 1979. Acoustic study of the Spanish diphthongs. *Phonetica* 36:194-206.

Canh, N-P. 1974. A contribution to the phonological interpretation of the diphthongs in Modern Vietnamese. *Phonetics Pragensia* 4:133-142.

Chioran, Ioana. 2002. A perception-production study of Romanian diphthongs and glide-vowel sequences. *Journal of the International Phonetic Association* 32:203-222.

Clements, G.N. 1991. Place of articulation in consonants and vowels: a Unified Theory. L'Architecture et la Géometrie des Représentations Phonologiques, eds. B. Laks and A. Rialland. Paris: Editions du C.N.R.S. [In French translation; preliminary English version in *Working Papers of the Cornell Phonetics Laboratory* 5:37-76.]

Collier, René, Fredericka Bell-Berti, and Lawrence J. Raphael. 1982. Some acoustic and physiological observations on diphthongs. *Language and Speech* 25:305-323.

Crothers, John. 1978. Typology and universals of vowel systems. *Universals of Human Language, Volume 2: Phonology*, ed. J. H. Greenberg, 95-152. Stanford, California: Stanford University Press.

Donegan, Patricia J. 1978. On the natural phonology of vowels. *Working Papers in Linguistics* 23. Ohio State University.

Đoàn Thiện Thuật. 1977. *Ngữ âm tiếng Việt* [*Vietnamese phonetics*]. Hà Nội: Nhà Xuất Bản Đại Học Quốc Gia.

Emeneau, Murray B. 1951. Studies in Vietnamese Annamese grammar. *University of California Publications in Linguistics* 8.

Fowler, C. 1980. Coarticulation and theories of extrinsic timing control. *Journal of Phonetics* 80:113-133.

Fox, R. 1983. Perceptual structure of monophthongs and diphthongs in English. *Language and Speech* 26:21-49.

Fox, R. 1989. Dynamic information in the identification of discrimination of vowels. *Phonetica* 46:97-116.

Gottfried, M., J. D. Miller and D. J. Meyer. 1993. Three approaches to the classification of American English diphthongs. *The Journal of Phonetics* 21:205-229.

Han, M. S. 1968. Complex syllable nuclei in Vietnamese. *Studies in the Phonology of Asian Languages* 6. University of Southern California.

Haudricourt, André G. 1952. Les Voyelles brèves du vietnamien. *Bulletin de la Société de Linguistique de Paris* 48(1):90-93.

Hillenbrand J., L. A. Getty, M. J. Clark, and K. Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* 97:3099-3111.

Hillenbrand, J. and R. T. Gayvert. 1993. Vowel classification based on fundamental frequency and formant frequencies. *The Journal of Speech and Hearing Research* 36:647-700.

Hạ. Kiều Phương. 2010. Prosody of Vietnamese from an interactional perspective: ờ, ừ and vâng in backchannels and requests for information. *Journal of the Southeast Asian Linguistics Society* 3(1):56-76.

Hoàng, Văn Hành, Hà Quang Năng, Nguyễn Văn Khang, Phạm Hùng Việt, and Nguyễn Công Đức. 1998. *Từ điển Từ láy tiếng Việt* [*Dictionary of Vietnamese Reduplicated Words*]. Nhà xuất bản Khoa học Xã hội.

Hockett, Charles F. 1955. A manual of phonology. *International Journal of American Linguistics* 21. Baltimore: Waverly Press and Indiana University.

Hollbrook, A and C. Fairbanks. 1962. Diphthong formants and their movements. *Journal of Speech and Hearing Research* 5:38-58.

Hualde, José Ignacio and Mónica Prieto. 2002. On the diphthong/hiatus contrast in Spanish: some experimental results. *Linguistics* 40:217-34.

Jenkins, J. J., W. Strange, and S. Miranda. 1994. Vowel identification in mixed-speaker silent-center syllables. *The Journal of the Acoustical Society of America* 95:1030-1043.

Jenkins, J., W. Strange, and T. Edman. 1983. Identification of vowels in "vowelless" syllables. *Perception and Psychophysics* 34:441-450.

Jones, Charles. 1989. *A History of English Phonology*. New York: Longman Inc.

Kirby, James P. 2011. Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association* 41(3):381-392.

Kehoe, Margaret, Geraldine Hilaire-Debove, Katherine Demuth and Conxita Lleó. 2008. The structure of branching onsets and rising diphthongs: Evidence from the acquisition of French and Spanish. *Language Acquisition* 15:5-57.

Ladefoged, Peter. 2001. *Vowels and Consonants: An Introduction to the Sounds of Languages*. Malden, Massachusetts: Blackwell Publishers Inc.

Lass, Roger. 1984. *Phonology: An Introduction to Basic Concepts*. Cambridge: Cambridge University Press.

Lehiste, Ilse and Gordon E. Peterson. 1961. Transitions, glides and diphthongs. *Journal of the Acoustical Society of America* 38:268-77.

Le-Van-Ly. 1960. *Le Parler Vietnamien* (second edition). Saigon: Bo Quoc Gia Giao Duc.

Lindblom, B. 1963. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35:1773-1781.

Lipski, John M. 1979. On vowel-diphthong transitions. *Current Issues in the Phonetics Sciences*, 551-561.

Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different listeners. *Journal of the Acoustical Society of America* 49:606-608.

Nearey, T. M. 1989. Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America* 85:2088-2113.

Nearey, T. M. and P. F. Assmann. 1986. Modeling the role of inherent spectral change in vowel identification. *The Journal of the Acoustical Society of America* 80:1297-1308.

Nguyên, Triều. 2001. *Ca Dao Ngụ Ngôn Người Việt: Tuyển chọn và Bình giải* [*Vietnamese folklore poem and sayings – A collection and analysis*]. Nhà xuất bản Thuận Hóa.

Nguyễn, Bạt Tụy. 1949. *Chữ và Vần Việt Nam Khoa Học* [*Scientific study of Vietnamese letters and syllables*]. Sài Gòn: Ngôn Ngữ.

Nguyễn, Bạt-Tụy. 1959. *Ngôn Ngữ học Việt Nam* [*Vietnamese linguistics*]. Sài Gòn: Ngôn Ngữ.

Nguyễn, Du. 1997. *Truyện Kiều* [*A story of a girl named Kieu*], ed. Nguyễn Quảng Tuân. Nhà xuất bản Khoa học Xã hội.

Nguyễn, Đình-Hòa. 1966. *Speak Vietnamese*. Rutland & Tokyo: Charles E. Turtle Co. Publishers.

Nguyễn, Đình-Hòa. 1997. *Tiếng Việt Không Son Phấn* [*Vietnamese without veneer*]. Amsterdam: John Benjamins.

Ni Chiosáin, Maire and Jaye Padgett. 1993. Inherent VPlace. *Report no. LRC-93-09*. Linguistics Research Center, Department of Linguistics, UC Santa Cruz.

Nooteboom, S. G. and I. H. Slis. 1972. The phonetic feature of vowel length in Dutch. *Language and Speech* 15:301-16.

Odden, David. 1991. Vowel Geometry. *Phonology* 8:261-90.

Parker, E. M. and R. L. Diehl. 1984. Identifying vowels in CVC syllables: effects of inserting silence and noise. *Perception and Psychophysics* 36:369-380.

Peeters, Wilhelmus Johannes Maria. 1991. *Diphthong Dynamics: A Cross-linguistic Perceptual Analysis of Temporal Patterns in Dutch, English and German*. Utrecht, the Netherlands.

Peterson, G. and H. Barney. 1952. Control methods used in a study of vowels. *The Journal of the Acoustical Society of America* 24:175-184.

Peterson, Gordon E. and Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32:693-703.

Pham, Andrea Hòa. 2003. *Vietnamese Tone: A Reanalysis*. New York: Routledge.

Pham, Andrea Hòa. 2006. Vietnamese Rhyme. *Southwest Journal of Linguistics* 25:107-142.

Pham, Andrea Hòa. 2009. The identity of non-identified sounds: Glottal stop, prevocalic /w/ and triphthongs in Vietnamese. *Toronto Working Papers in Linguistics TWPL* 34.

Pham, G., Kohnert, K., & Carney, E. 2008. Corpora of Vietnamese Texts: Lexical Effects of Intended Audience and Publication Place. *Behavior Research Methods* 40:154-163.

Pike, K. L. 1947. On the phonemic status of English diphthongs. *Language* 23:151-159.

Potter, Ralph K. and Gordon E. Peterson. 1948. The representation of vowels and their movements. *The Journal of the Acoustical Society of America* 20:528-535.

Ren, HongMo. 1986. *Working Papers in Phonetics: On the Acoustic Structure of Diphthongal Syllables*. University of California, Los Angeles.

Roengpitya, Rungpat. 2007. A new look on diphthongs in Thai. *SEALSXIII: Papers form the 13th Meeting of the Southeast Asian Linguistics Society* (2003), eds. Shoichi Iwasaki, Andrew Simpson, Karen Adams, and Paul Sidwell, 231-237. Canberra, Pacific Linguistics.

Sagey, Elizabeth Caroline. 1993. *The Representation of Features and Relations in Non-linear phonology*, Ph.D. dissertation, MIT, Cambridge, Ma. (Revised version published 1991 by Garland Publications, New York.)

Salza, Pier Luigi. 1988. Durations of Italian diphthongs and vowel clusters. *Language and Speech* 31:97-113.

Smalley, William A. and Nguyen Van-Van. 1957. *Vietnamese for Missionaries: A Course in the Spoken and Written Language of Central Viet Nam, I & II*. Saigon.

Steriade, Donca. 1987. Locality Conditions and Feature Geometry. *Proceedings of NELS* 17, Graduate Linguistics Student Association, University of Massachusetts, Amherst. 595-617.

Stokes, Stephanie F. and I. Man Wong. 2002. Vowel and diphthong development in Cantonese-speaking children. *Clinical Linguistics & Phonetics* 16:597-617.

Strange, W. 1989a. Dynamic specification of coarticulated vowels spoken in sentence context. *The Journal of the Acoustical Society of America* 85:2135-53.

Strange, W. 1989b. Evolving theories of vowel perception. *The Journal of the Acoustical Society of America* 85:2081-7.

Strange, W., J. J. Jenkins, and T. L. Johnson. 1983. Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America* 74:695-705.

Strange, W., R. R. Verbrugge, D. P. Shankweiler, and T. R. Edman. 1976. Consonant environment specifies vowels identity. *The Journal of the Acoustical Society of America* 60:213-224.

Swets, Francine and Marc van Oostendorp. 2003. The structure of the Dutch /au/. *Linguistics in the Netherlands*, 167-76. John Benjamins Publishing Company.

Syrdal, Ann K. and H.S. Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79:1086-100.

Tauberer, Joshua and Keelan Evanini. 2009. Intrinsic vowel duration and the post-vocalic voicing effect: Some evidence from dialects of North American English. *Proceedings of Interspeech*.

Thomas, Erik R. and Kendall, Tyler. 2007. *NORM: The Vowel Normalization and Plotting Suite*.

Thompson, Laurence C. 1965. *A Vietnamese Reference Grammar*. Honolulu: University of Hawai'i Press.

Verbrugge, R. R. and B. Rakerd. 1986. Evidence of talker-independent information for vowels. *Language and Speech* 29:39-55.

Watson, Catherine I. and Jonathan Harrington. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America* 106:458-468.

Zahorian, S. and A. Jagharghi. 1993. Spectral-shape features versus formants as acoustic correlates for vowels. *The Journal of the Acoustical Society of America* 94:1966-1982.