

A preliminary quantitative study on the characteristics of Vietnamese vowels and English vowels

Nguyen Bach*, Srihari Reddy**

* Department of Computer Science, nguyen@cs.jhu.edu

** Department of Electrical & Computer Engineering, srihari@jhu.edu

Johns Hopkins University

1. Introduction

Vietnamese, a language in South-East Asia, has nearly 80 million speakers in Vietnam and around 3 million speakers overseas. There are 29 letters in the writing system of the Vietnamese language.

a	ă	â	b	c	d	đ	e
ê	g	h	i	k	l	m	n
o	ô	ơ	p	q	r	s	t
u	ư	v	x	y			



Vietnamese is a monosyllabic and tonal language which has 11 vowels, 19 consonants, and 6 tones. Vietnamese vowels are *i, u, u, e, ɤ, o, ɛ, ɔ, ɐ, a, ɑ*. The Vietnamese vocalic system was divided into upper and lower vocalics (Thompson 1987). The upper vocalics include six vowels, /*i u u e ɤ o*/. They are formed relatively high in the mouth and characterized by a three-way position (front, back unrounded, and back rounded). Lower vocalics include five vowels, /*ɛ ɔ ɐ a ɑ*/. They are formed relatively low and characterized by a two-way position distinction (front, back). Figure 1 shows the Vietnamese vowel quadrilateral.

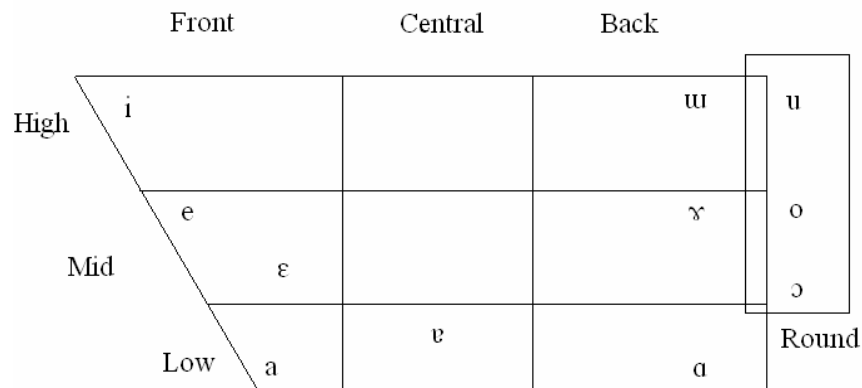


Figure 1: The Vietnamese vowel quadrilateral

These are (numbers indicate the indices to be used throughout this report): Level (1), sometimes also referred to as ‘mid-level’, rising (2), broken (3), falling (4), curve (5), and drop (6) tones, see also appendix 1.

Vowels are not always evenly distributed throughout a vowel chart, for example English vowel chart. The current study aims at providing a preliminary quantitative description of formant values for F1 and F2 for each vowel and plot the vowel chart of Vietnamese. In addition, the project also verify two hypotheses which are 1) the distance between front vowels is the same as the distance between back vowels, and 2) the distance between high vowels is the same as the distance between low vowels.

2. Methods

In this section we would like to provide information about the subject, data, recording procedures, and measurement criteria. In order to examine the characteristics of each vowel a set of 11 utterances was recorded by a 24-year-old native male speaker of Hanoi dialect, the standard dialect of Vietnam. The speaker can speak English fluently but not well-trained in phonetics. The utterances were recorded three times as mono sounds in the frequency of 11025 Hz. The word list as follows:

No	Vowel	Meaning in Vietnamese	Meaning in English	Transcription
1	/i/	tí	tiny	[ti]
2	/e/	tê	to sacrifice	[te]
3	/u/	tứ	four	[tu]
4	/u/	tú	bachelor	[tu]
5	/ɤ/	tớ	I	[tɤ]
6	/o/	tỏ	to denounce	[to]
7	/ɛ/	té	to fall down	[te]
8	/ɔ/	tó	no meaning ¹	[tɔ]
9	/a/	tá	dozen	[ta]
10	/ɐ/	ắt	surely	[ɐt]
11	/ɑ/	ta	we	[ta]

The major concerns of project are the vowels therefore the word list is chosen so that consonants and tones have less affection on vowels. To test this hypothesis, the words are minimal pairs only and should have the same tone so that all other influences on voice-onset-time are controlled as much as possible. However, it is very hard to select the list in Vietnamese. Vowel /ɐ/ is the only one that does not begin with the consonant /t/, while vowel /ɑ/ begins with /t/ and without tone 2. Others begin with the same consonant /t/ and tone 2.

Each vowel is represented in two parameters, the first and second formant. To identify vowels form the acoustics, F1 and F2 are measured near the center of the vowel by using Praat. F1 and F2 are measured in Hz domain. JPlotFormants program uses F1 and F2 values to plot the vowel chart of Vietnamese. Note that JPlotFormants does not use an

¹ “tỏ” has no meaning when it stands separately but it’s a real sound in the word “quả tỏ” – catch in the act.

IPA font. We are to use the following set of symbols within JPlotFormants: /i/ ii; /e/ e; /u/ w; /u/ u; /ɣ/ v; /o/ o; /ɛ/ eh; /ɔ/ ao; /a/ a; /ɐ/ ac; /ɑ/ aa. Figure 2 illustrates the technique in Praat.

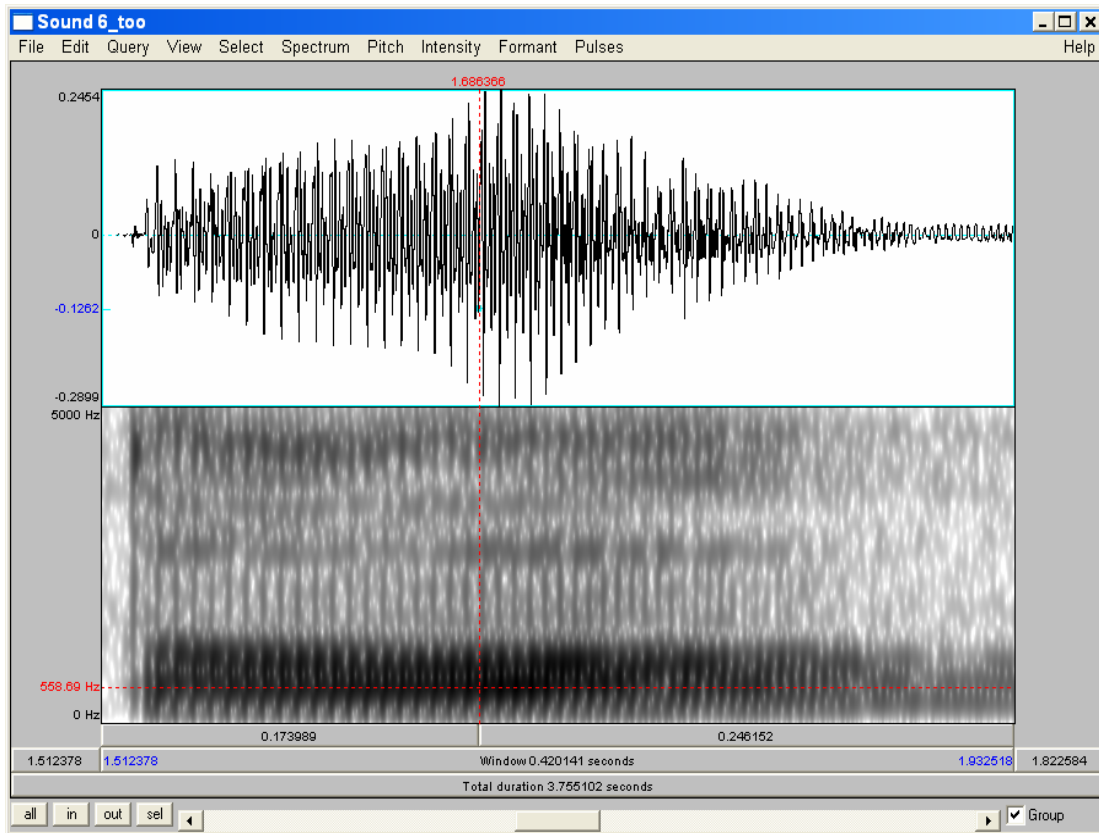


Figure 2: Measure F1 and F2 using Praat

To measure the distance between two vowels, the absolute value of the difference two tokens of adjacent vowels. We also need to compute mean and standard deviation for further analysis. For example, we calculate the distance between /i/ and /e/ for Vietnamese as:

	F1 for /i/	F1 for /e/	F1 distance	F2 for /i/	F2 for /e/	F2 distance
token_1	342	427	85	1001	1206	205
token_2	340	426	86	1051	1200	149
token_3	341	425	84	1026	1203	177
Mean	341	426	85	1026	1203	177
STDEV			1			28

Finally, the data set is rather small so statistical differences are based on two-tailed t tests and the alpha-level for p-value is 0.05.

3. Analysis & results

We show the formant values for F1 and F2 below.

tí [1 ti.wav]	F1 in Hz	F2 in Hz	tó [6 too.wav]	F1 in Hz	F2 in Hz
token_1	431	2138	token_1	554	1008
token_2	431	2121	token_2	558	1025
token_3	448	2121	token_3	554	990
Mean	436.66667	2126.666667	Mean	555.3333333	1007.666667
Std. Dev	9.8149546	9.814954576	Std. Dev	1.88561808	17.5023808
té [2 tee.wav]	F1 in Hz	F2 in Hz	té [7 te.wav]	F1 in Hz	F2 in Hz
token_1	552	1926	token_1	800	2067
token_2	569	1960	token_2	797	2016
token_3	552	1943	token_3	797	2016
Mean	557.66667	1943	Mean	798	2033
Std. Dev	9.8149546	17	Std. Dev	1.73205081	29.4448637
tú [3 tuw.wav]	F1 in Hz	F2 in Hz	tó [8 to.wav]	F1 in Hz	F2 in Hz
token_1	452	1294	token_1	778	1230
token_2	452	1295	token_2	778	1230
token_3	435	1277	token_3	760	1213
Mean	446.33333	1288.666667	Mean	772	1224.33333
Std. Dev	9.8149546	10.11599394	Std. Dev	10.3923048	9.81495458
tú [4 tu.wav]	F1 in Hz	F2 in Hz	tá [9 ta.wav]	F1 in Hz	F2 in Hz
token_1	416	922	token_1	794	1344
token_2	400	924	token_2	794	1358
token_3	416	924	token_3	811	1360
Mean	410.66667	923.3333333	Mean	799.666667	1354
Std. Dev	9.2376043	1.154700538	Std. Dev	9.81495458	8.71779789
tó [5 tow.wav]	F1 in Hz	F2 in Hz	át [10 at.wav]	F1 in Hz	F2 in Hz
token_1	643	1342	token_1	775	1433
token_2	646	1324	token_2	775	1417
token_3	645	1324	token_3	792	1451
Mean	644.66667	1330	Mean	780.666667	1433.66667
Std. Dev	1.5275252	10.39230485	Std. Dev	9.81495458	17.0098011
ta [11 ta.wav]	F1 in Hz	F2 in Hz			
token_1	830	1546			
token_2	829	1563			
token_3	812	1564			
Mean	823.66667	1557.66667			
Std. Dev	8.25967	10.11599			

To test the hypothesis 1, “the distance between the front vowels is the same as the distance between back vowels in Vietnamese”, the distance of front vowels and back vowels is computed in F1 domain. The next table reports tokens, means, and standard deviation.

Front Vowels

	F1 for /i/ in Hz	F1 for /e/ in Hz	Distance in Hz b/w /i/ and /e/
token_1	431	552	121
token_2	431	569	138
token_3	448	552	104

	F1 for /e/ in Hz	F1 for /ɛ/ in Hz	Distance in Hz b/w /e/ and /ɛ/
token_1	552	800	248
token_2	569	797	228
token_3	552	797	245

	F1 for /ɛ/ in Hz	F1 for /a/ in Hz	Distance in Hz b/w /ɛ/ and /a/
token_1	800	794	6
token_2	797	794	3
token_3	797	811	14

Mean distance between the front vowels: 123
Std. Deviation of distance between front vowels: 101.3002961

p-value for an alpha level of 0.05: 0.9454

Back Vowels

	F1 for /u/ in Hz	F1 for /ʊ/ in Hz	Distance in Hz b/w /u/ and /ʊ/
token_1	452	643	191
token_2	452	646	194
token_3	435	645	210

	F1 for /ɔ/ in Hz	F1 for /ɒ/ in Hz	Distance in Hz b/w /ɔ/ and /ɒ/
token_1	643	778	135
token_2	646	778	132
token_3	645	760	115

	F1 for /ɒ/ in Hz	F1 for /a/ in Hz	Distance in Hz b/w /ɒ/ and /a/
token_1	778	830	52
token_2	778	829	51
token_3	760	812	52

Mean distance between the back vowels: 125.7778
Std. Deviation of distance between back vowels: 63.95267

Thus the probability that the difference between the distances between the two groups is due to chance is 0.9454. This is greater than the alpha-level. The two distances are not statistically different for our alpha-level. We can conclude that the hypothesis is true. The distance between the front vowels and the back vowels is same.

To test the hypothesis 2, “the distance between the high vowels is the same as the distance between low vowels in Vietnamese”, the distance of front vowels and back vowels is computed in F2 domain. The next table reports tokens, means, and standard deviation.

High Vowels

	F2 for /i/	F2 for /u/	Distance in Hz b/w /i/ and /u/
token_1	2138	1294	844
token_2	2121	1295	826
token_3	2121	1277	844
Mean distance			838
Std. Deviation of distance			10.3923

Low Vowels

	F2 for /a/	F2 for /ɑ/	Distance in Hz b/w /a/ and /ɑ/
token_1	1344	1546	202
token_2	1358	1563	205
token_3	1360	1564	204
Mean distance			203.6667
Std. Deviation of distance			1.527525
p-value for an alpha level of 0.05:			5.01E-08

Thus the probability that the difference between the distances between the two groups is due to chance is 5.01E-0.8. This is significantly smaller than the alpha-level. Thus the two distances are statistically different for our alpha-level. We derive the conclusion that the hypothesis is not true. The distance between the high vowels and the low vowels is not same.

By using the formant pairs, we come up with a possible vowel space for the Vietnamese language in Figure 3. All vowels fall into the possible vowel space with the F1 in range of 200 and 1000, while F2 in between 500 and 2500. The vowel chart shows that the distance between the front high unrounded vowel /i/ and the front low unrounded vowel /a/ is around 400 in Hz, while the distance of /i/ and /u/ is around 1200.

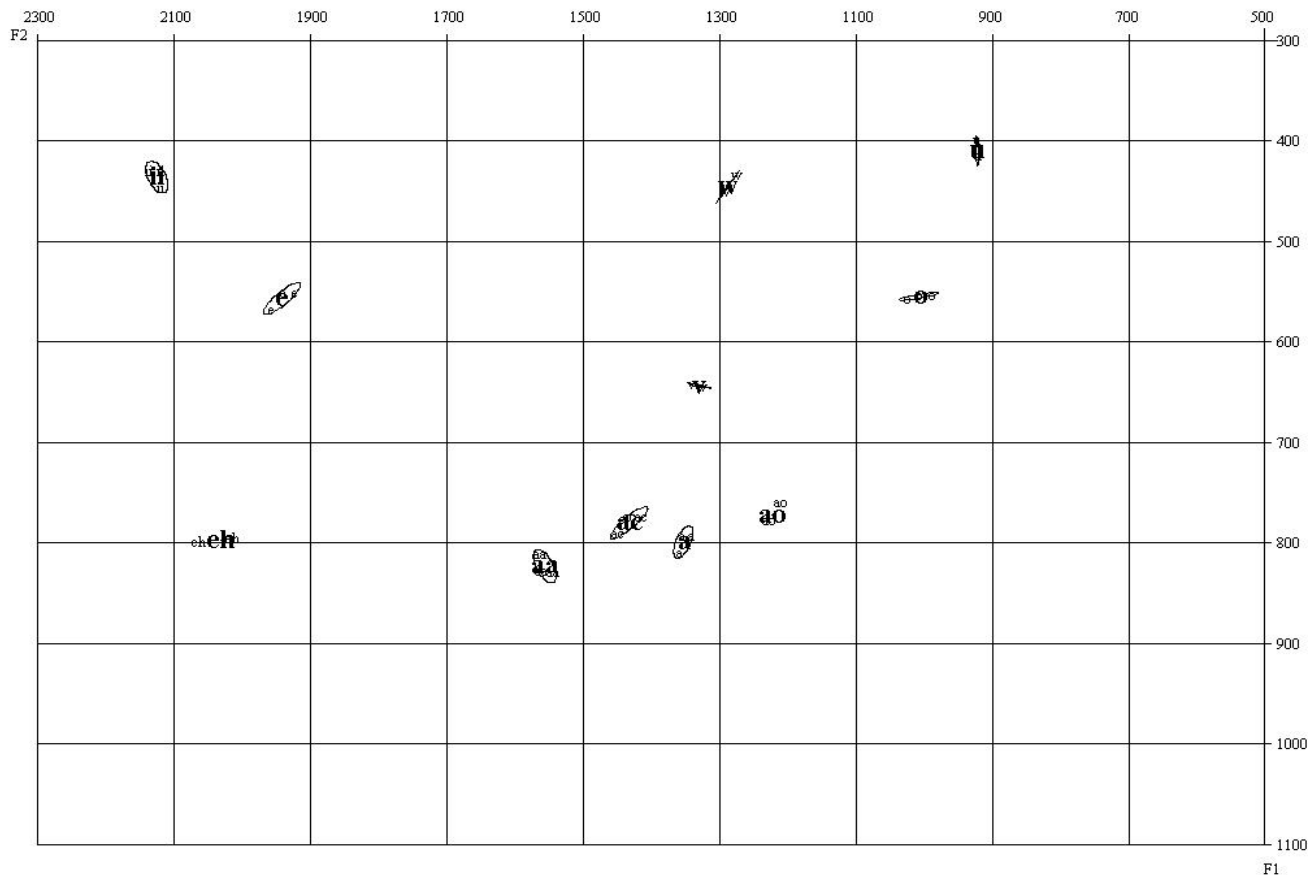


Figure 3: Vietnamese Vowel Space

4. Reference

- [1] Thompson, Laurence. 1987. *A Vietnamese Reference Grammar*. Hawaii: University of Hawaii.
- [2] H. Mixdorff, N. Bach, et al., *Quantitative Analysis and Synthesis of Syllabic Tones in Vietnamese*, Proceeding of The 8th European Conference on Speech Communication and Technology 2003 in Switzerland, Sep 2003, pp 177 - 180.
- [3] <http://www.saigonnet.vn/english/edu/learning-vietnamese/>
- [4] P. Ladefoged, *Vowels and Consonants*, Blackwell Publishing, 2001.
- [5] <http://www.de-han.org/vietnam/chuliau/lunsoat/sound/>
- [6] <http://www.praat.org>
- [7] <http://www.linguistics.ucla.edu/people/grads/billerey/PlotFrog.htm>

Appendix 1 (except from [2])

Vietnamese is known as a monosyllabic tone language having six different lexical tones. These are (numbers indicate the indices to be used throughout this article): Level (1), sometimes also referred to as ‘mid-level’, rising (2), broken (3), falling (4), curve (5), and

drop (6) tones. Tones 2-6 are marked by diacritics in the Vietnamese script which uses the Latin alphabet. The widely cited description by Thompson [1] gives the following account which is also summarized in the below table:

No	Vietnamese Name	English name	F0 contour	Diacritic	Additional features
1	Ngang	level	Trailing/falling	none	Laxness
2	Sắc	rising	Rising	Á	Tenseness
3	Ngã	broken	Rising	Ã	Glottalization
4	Hỏi	falling	Falling	À	Tenseness
5	Huyền	curve	Falling	À	Laxness, breathiness
6	Nặng	drop	Dropping	Ạ	Glottalization/tenseness

Tone 1 is modal and its contour is nearly level in non-final syllables not accompanied by heavy stress, although even in these cases it probably trails downward slightly. Although tone 1 is phonetically slightly falling, it is phonemically regarded as a level tone similar to Mandarin tone 1, but with relatively lower pitch. Tone 2 is high and rising (perhaps nearly level in rapid speech) and tense, and similar to tone 2 in Mandarin Chinese. Tone 3 is also high and rising, the *F0* contour being similar to that of tone 2, but it is accompanied by the rasping voice quality occasioned by tense glottal stricture. In careful speech such syllables are sometimes interrupted completely by a glottal stop (or a rapid series of glottal stops). Its trajectory therefore sometimes shows a characteristic break in the voicing at about half of the total duration of the syllable.

Tone 4 is tense; it starts somewhat higher than tone 5 and drops rather abruptly. In final syllables, and especially in citation forms, this is followed by a sweeping rise at the end, and for this reason it is often called the ‘dipping’ tone. However, nonfinal syllables seem only to have a brief level portion at the end, and this is exceedingly elusive in rapid speech. Although tone 4 is usually described as a low falling and then rising tone, not all Vietnamese speakers have the rising part. When tone 4 consists of a falling and a rising contour, it is similar to Beijing Mandarin tone 3.

Tone 5 is also lax, starts quite low and trails downward toward the bottom of the voice range. It is often accompanied by a kind of breathy voicing, reminiscent of a sigh. Tone 6 is also tense; it starts somewhat lower than tone 4. With syllables ending in a stop [p t c k] it drops only a little more sharply than tone 5, but it is never accompanied by the breathy quality of that tone. Other syllables have the same rasping voice quality as tone 3, drop very sharply and are almost immediately cut off by a strong glottal stop. Tone 6 is much shorter than other tones with a tendency to go lower.