Vowel Reduction in Conversational Speech in Vietnamese

Anh-Thư T. Nguyễn
Mountain Creek. QLD, Australia
Email: thunguyen1972@gmail.com

Abstract

This paper reported a study which investigated vowel reduction in conversational speech in Vietnamese. Vowel duration and spectral quality are examined with respect to (1) the interaction of durational and spectral values, (2) the effect of accent, (3) the position of vowels in words and (4) word category (function vs. content words). The analyses were conducted on vowels produced in interview conversations by 6 speakers. The results showed that (1) vowel duration and spectral quality in Vietnamese are highly related; (2) vowels in accented condition have longer duration and are more peripheral than those in unaccented counterparts; nevertheless, this effect is confounded with pre-boundary effect because of stress being a syntactic boundary in Vietnamese phrases (Cao 1978, 1998); (3) Vowels of the first syllables of disyllabic words are more reduced compared to those in second syllables and (4) vowels are more reduced in monosyllabic function words than in content words. Overall, the study shows that vowel reduction is more pervasive in spontaneous Vietnamese than previously documented.

Keywords

Vowel reduction, Vietnamese, spontaneous speech, accent, function words, acoustic

1. Introduction

1.1. Vowel Reduction

In spontaneous speech, words are often reduced compared to their pronunciations in canonical forms. This involved vowel reduction, which is considered to be "a phenomenon of undershoot related to unattained targets, generally involving shortened vowel duration leading to a reduced vowel space. This phenomenon affects vowel quality so that the values

observed under certain conditions are often far removed from prototypical ones." (Meunier & Espesser 2011: 2).

Vowel reduction has been observed under many conditions and prosodic contexts. Lindblom (1963) found that vowels are reduced in unstressed position, at lower prosodic boundaries and in increased speech tempo. This is realised in such a way that unstressed vowels are shorter than their stressed counterparts and decreased vowel duration is known to correlate with formant undershoot. General durational variation in connected speech has been found to affect vowel spectral (Gendrot & Adda-Decker 2005), i.e. the longer the vowels, the larger the acoustic space they will occupy, being thus more and more distinct from one another. Deaccentuation also leads to vowel reduction. It has been found that focusing and accentuation in English are accompanied by hyperarticulation of the vowel space: that is, the vowels in accented words are phonetically more peripheral (Harrington, Fletcher & Beckman 2000). Peripheral vowels are those located at the edge of the vowel envelope, and non-peripheral nuclei are those located on the inside. Vowels are generally more reduced in function words as compared to lexical words (van Bergem 1993). Because in most of these conditions vowel reduction is accompanied by a shortening of the vowel, vowel reduction is often seen as a consequence of the durational shortening (Lindblom 1963, Flemming 2004). This is called "target undershoot" (Lindblom 1963). This term implies that due to temporal constraints the articulators do not reach the vowel-specific target resulting in formant undershoot. Target undershoot has been found for most kinds of vowel reduction phenomena and causes a general shrinkage of the vowel space. In a recent study, Meunier and Espesser (2011) found that vowel reduction in conversational French speech depends on several factors related to lexical properties (word category: function words versus content words) and to prosodic properties (stress and final lengthening). This paper reported a study which investigated vowel reduction in conversational speech in Vietnamese. Vowel duration and spectral quality are examined with respect to (1) the interaction of durational and spectral values, (2) the effect of accent, (3) the position of vowels in words and (4) word category (function vs. content words). The motivation behind the study of vowel reduction is to support the investigation of prosodic asymmetry in Vietnamese

1.2. Background

The term 'stress' is used to refer to the perceptual salience at certain places in strings of syllables, but it has several different referents: (a) relative syllable salience in an utterance; this is syllable-, not word-oriented; (b) stress in a word; this is part of the lexical phonology; (c) stressing of words in utterances for various aspects of propositional and expressive meaning, often called 'accent(uation)' (Kohler 2008). Stress and accentuation depends

crucially on the speaker's ability to make certain syllables more noticeable than others. A syllable which "stands out" in this way is a prominent syllable. There are many ways in which a syllable can be made prominent. In English, prominence is associated with greater length, greater loudness, pitch prominence and with "full" vowels and diphthongs. Despite the complexity of this set of interrelated factors it seems that perceptually the listener simply hears syllables as more prominent or less prominent (Roach 2002).

Vietnamese is a contour tone language that has no system of culminative word stress (Nguyễn, Ingram and Pensalfini 2008; Nguyễn and Ingram 2007a); nevertheless, it is widely accepted that there is stress in the sense of accentual prominence at the phrasal level (Thompson 1965; Nguyễn Đăng Liêm 1970). Duration, intensity, and full tonal realisation of accented syllables have been observed to be important parameters for describing stress in Vietnamese (Đỗ 1986; Chaudhary 1983; Hoàng & Hoàng 1975; Gsell 1980). Jones and Huỳnh (1960) stated that "normally the stresses in a Vietnamese utterance are conditioned by the junctures," and regarded the fundamental stress pattern of Vietnamese as consisting of the alternating occurrence of a weak and strong stress, with the last word of the phrase receiving a strong stress. Cao (1978, 1998) provides convincing evidence for stress being a syntactic boundary in Vietnamese phrases, in classifying compounds, and a quality of content words as opposed to functional words, with longer duration, stronger amplitude and more completed tonal contour. In recent studies on carefully phonetically controlled and specialized sets of Vietnamese disyllabic compounds and reduplications , Nguyen and Ingram (2007a, b) have found that the right hand elements of a disyllabic compound words were more prosodically prominent by a number of relevant phonetic measures: greater tonal f0 range, higher intensity, greater duration of the second syllable, and formant measurements indicative of more centralized vowel nuclei (vowel reduction) on the first syllable. In this paper, the terms stress, accent and prominence are used to refer to phenomena specific to Vietnamese prosody as specified above.

The status of vowel reduction in Vietnamese has not received much attention in empirical research. Only the segmental reduction of unstressed syllables has been found in Nguyen and Ingram (2007a, b) and Pham (2008). Nguyen and Ingram (2007a) examined vowel formants of disyllabic compounds vs. phrases in picture naming and minimal pair sentence tasks (e.g., *hoa hồng*: noun phrase: a pink flower and *hoa hồng*: compound: a rose). They found that the raising of the F1 (the first formant) of the vowel of the phrases in comparison with that of the compound was evidenced only in the first syllable under the minimal pair sentence task (i.e. under accented condition). That is, vowels under accented condition tend to be more peripheral than those under normal unaccented condition. In addition, in the same study, they also found that in reversible coordinative compounds (e.g., in *bàn ghế* : tables and chairs vs.

ghế bàn: chairs and tables), the vowel in the second syllables have longer duration and are more peripheral than its first position counterparts, suggesting an articulatory gesture enhancement with larger mouth opening and jaw lowering which have been shown by previous researchers (Erickson 2002; Van Summers 1987) to be indicative of stress or prosodic strengthening. Nguyen and Ingram (2007b) investigated vowel formants of disyllabic reduplications (e.g., *sáng sáng*: bright bright). They found that the vowel of the reduplicated syllable (the first syllable) tended to be more centralised than that of the base syllable (the second syllable). Phạm (2008) found that Vietnamese function words are typically cliticized (phonetically reduced) and appended to prosodic words. This study seeks to investigate vowel duration and spectral quality with respect to (1) the interaction of durational and spectral values, (2) the effect of accent, (3) the position of vowels in words and (4) word category (function vs. content words).

Since Vietnamese has no system of culminative word stress and the status of accent as well as the existence and realization of higher prosodic constituents (phonological phrase, intonational phrase) has never been systematically investigated in Vietnamese (only a basic description of Vietnamese is found in Thompson 1965), a study was conducted to investigate the acoustic correlates of prosodic prominence and boundary, as they are perceived by naïve listeners, in spontaneous speech of Vietnamese interview conversations (Nguyen, submitted). The detailed methodology of the perception test and the summary of its results are presented in the section 2.1 below.

2. Method

2.1. The speech corpus

A total of 30 sound files were created from excerpts of interviews of six speakers (3 females and 3 males) of Saigon dialect. They were students at University of Queensland and came from Ho Chi Minh city. They had been in Australia from 1-3 years. These excerpts consisted of a single-speaker passages of spontaneous conversational interview speech. In the interviews, the speakers answered the experimenter's open-ended questions about their hobbies, their experiences and their lifestyles. The selected samples include five excerpts from each of the six speakers. The duration of these excerpts is from 15 to 30 seconds. Excerpts were selected according to the following two criteria. (1) speech excerpts were selected from segments of the interviews in which there are no technical recording problems, and (2) speech excerpts were selected to minimize the occurrence of disfluencies, though it was not possible to avoid all the disfluent regions for the selection of speech excerpts.

Orthographic transcriptions of the excerpts were prepared by the experimenter (a native speaker of Southern Vietnamese). In these transcriptions which were prepared for use

in the listening test, punctuation and line breaks were removed except as necessary to fit on the page, aiming to avoid providing any hints as to the structure. Disfluencies such as repeated or partial words were included in the transcripts but filled pauses ("à, ư…") were not indicated.

Forty naive listeners without training in phonetics or prosody were recruited at a university in the South of Vietnam. Most were tertiary students and in the age range from 18 to 30 years of age. The experiment was administered in a sound-attenuated room. Each listener was presented with a packet containing instructions and the printed transcripts of the practice and test extracts. They marked their responses on these printed answer sheets.

Listeners were randomly assigned to one of two groups. 20 listeners were asked to mark a vertical line between words at locations where they perceived a boundary between different chunks of the utterance. The remaining 20 listeners were instructed to underline words that they heard as "prominent". A prominent word was defined for the listeners as a word that is "highlighted for the listener, and stands out from other non-prominent words", while a chunk is defined as a grouping of words "that helps the listener interpret the utterance", and that chunking is "especially important when the speaker produces long stretches of continuous speech". All listeners heard the extracts twice with brief pauses between them. The extracts in each group were presented in a random order.

A modified form of Cohen's Kappa was used to assess agreement among listeners. This method takes into account the amount of agreement that can be expected by chance. Kappa values can vary between 0 and 1. The particular form of Kappa used here is based on Brennan and Prediger (1981), which is suitable for tasks with multiple raters in which the raters are not constrained as to how many items they assigned to each category ("free marginal"). Calculations were made using the Online Kappa Calculator (Randolph 2008). Kappa values were determined for each extract, pooling across all the listeners.

The results show that Kappa values for boundary marking for each excerpt ranged from 0.79 to 0.97 with a mean of 0.89. Kappa values for prominent marking for each excerpt ranged from 0.77 to 0.84 with a mean of 0.88. Randolph (2008) suggests that for this form of kappa, 0.7 or above is "adequate", so the values obtained in this study are well above this proposed cut-off. The result also showed that there was a strong correlation between boundaries scores and prominent scores ($r= 0.71$, $p<0.001$). In addition, 74% of words that were marked as prominence were also marked as preceding a boundary, indicating that listeners tend to perceive prominence on words where they perceive a boundary following, consistent with the observations of Cao (1978 1998) and Jones and Huỳnh (1960). The examination of the effect of accent on vowel reduction in this paper is based on these results: these listener-identified prominences. Those words marked as prominence by two-thirds or

more of listeners (14 or more of the 20: equal to or above 70%) were considered to have "consensus" agreement and chosen as accented words for this study.

2.2. Vietnamese vowels in the corpus

The Vietnamese vowel system contains 9 long vowels, 2 short vowels and 3 diphthongs. The long vowels are /i, ɯ, u, e, ɤ, o, ɛ, a, ɔ/; short vowels are /ă, ɤ̆/; diphthongs are /ie, ɯɤ, uo/ (Đoàn 1977). The spectral analysis of diphthongs is often more complicated, and I therefore chose to analyse only the eleven monophthongs. The following table shows the frequency of each vowel in the corpus as well as their IPA and SAMPA symbols. Due to the difficulty of representing IPA symbols for some Vietnamese vowels (particularly ă and ɤ̆) in the vowel plots, the phonetic notation employed in the following analysis is derived from the Speech Assessment Methods Phonetic Alphabet (SAMPA) – a computer readable phonetic alphabet (Wells, 1997). Duration and formants (first: F1, second: F2 and third formant: F3) of all monophthong vowels in the corpus was taken manually via Praat (Boersma & Weenink 2009).Vowel formants (Hz) was taken at vowel midpoint.

Conventional Vietnamese orthography	IPA vowel	SAMPA	Frequency	Percent
ă	ă	6	46	4.4
ơ	ɤ	7	53	5
a	a	A	284	26.9
ê	e	e	41	3.9
e	ɛ	E	67	6.4
i	i	I	163	15.5
ư	ɯ	M	60	5.7
ô	o	o	76	7.2
o	ɔ	O	153	14.5
u	u	U	68	6.4
â	ɤ̆	V	44	4.2
Total			1055	100

Table 1. The frequency of vowels in the corpus as well as their IPA and SAMPA symbols

2.3 Analysis

In order to account for the effect of speakers' differences and the intrinsic segmental and

tonal effects, a restricted maximum likelihood (REML) applied to mixed model methodology was performed on each of the acoustic parameters. The dependant variables were duration (ms), F1 (Hz), F2 (Hz) and F3 (Hz) values. The fixed effects included a condition (e.g., prominence (prominent vs. non-prominent), vowel positions (first vs. second in a bisyllabic word) or word categories (function words vs. content words)) and vowels (11 vowels) and their two-way interactions. The random effect was speakers (6 speakers), syllable structures (CVC, CV, V, and VC; C for a consonant and V for a vowel), tones (Rising, Level, Falling, Dropping and Curve/ Broken: in Southern Vietnamese, Curve and Broken are merged) and items. A Tukey post-hoc test was then conducted to determine the significant differences among the levels of the main effects and their interaction effects. The use of REML overcomes the potentially serious deficiency of the ANOVA-based methods which assumed that data are sampled from a random population and normally distributed. REML also avoids bias arising from maximum likelihood estimators in which all fixed effects are known without errors, consequently tend to downwardly bias estimates of variance components. Moreover, REML can handle unbalanced data. The data analysis was carried out using the SAS program. It is noted that in a preliminary analysis with syllable structure as one of the fixed effects, the results show only a marginal significant main effect for vowel duration while there was no significant effect for vowel spectrum, therefore, syllable structure was included in this model as a random factor.

3. Results

3.1. The interaction of durational and spectral values

In order to examine the interaction of durational and spectral values, the vowels were split into three different duration-dependent data sets corresponding to short, intermediate and long vowels in the corpus: set 1: 30ms to 79ms, set 2: 80ms to 119 ms, set 3:120 ms to 400 ms. Table 2 shows female and male spectral values (F1, F2 and F3 in Hz) according to the three duration sets in the corpus.

As shown in table 2 and figure 1, spectral reduction (i.e., more centralized spectral /formant values) was observed in relation to duration variation. In set 1 (shortest vowels) the vowel space is clearly less peripheral than in set 3 (longest vowels), with set 2 being intermediate. The extent of the reduction depends on vowel identity and spectral parameters (F1, F2 or F3). F1 values of the vowels /a/, / ă/, / ɤ̆ /, /ɤ/, /ɛ/, /e/, and / ɔ/ were decreased to a great extent particularly by the female speakers, indicating a less open jaw and/or higher (less low) tongue position. F2 values show a decrease for vowels /i/, /e/ and /ɛ/, an increase for /ɯ/, /u/ and /o/, indicating that front and back vowels show more central values. A decrease in F3 values is observed for the vowel /i/ as duration decreases, while an increase in F3 is shown for

vowels /ɯ/, /ɔ/, /ɤ/ and /u/. F3 variations tend to show that, for short vowels, lip position may be less extreme (less rounded for rounded vowels and less spread for unrounded vowels).

Formant values - Female speakers												
	Vowel durations	a	ɛ	ă	i	ɯ	ɔ	ɤ	u	ɤ̆	e	o
F1	30-79ms	825	581	876	509	498	772	692	500	656	591	632
	80-119 ms	867	677	1001	452	505	851	747	436	824	656	637
	120-400 ms.	953	674	1029	419	451	921	761	427	887	670	651
F2	30-79ms	1869	2390	1773	2171	2417	1551	1882	1141	1783	2271	1199
	80-119 ms	1859	2220	1778	2436	1953	1521	1738	1235	1788	2032	1233
	120-400 ms.	1936	2427	1750	2644	1689	1364	1536	1046	1902	2384	1003
F3	30-79ms	3052	3228	3011	3097	3272	3199	3049	3050	2958	3276	3234
	80-119 ms	2929	2919	2676	3243	3096	3105	3178	3029	3217	2951	3165
	120-400 ms.	2954	3375	2638	3373	2997	2970	3083	2945	2645	3155	3442
Formant values - Male speakers												
	Vowel durations	a	ɛ	ă	i	ɯ	ɔ	ɤ	u	ɤ̆	e	o
F1	30-79ms	617	527	680	429	397	596	571	422	539	463	523
	80-119 ms	620	530	702	392	390	668	620	458	724	534	546
	120-400 ms.	661	598	827	367	405	663	600	414	430	559	506
F2	30-79ms	1501	1804	1532	1749	1979	1233	1436	1127	1435	1663	995
	80-119 ms	1578	1869	1490	1938	1946	1184	1454	1237	1538	1618	940
	120-400 ms.	1667	1822	1675	2068	1752	1140	1371	846	1179	1782	893
F3	30-79ms	2460	2411	2428	2642	2727	2531	2793	2661	2591	2520	2614
	80-119 ms	2542	2521	2280	2647	2631	2541	2693	2446	2769	2238	2772
	120-400 ms.	2625	2583	3546	2678	2447	2657	2624	2812	2706	2450	2488

Table 2. Female and male spectral values (F1, F2 and F3 in Hz) according to the three duration sets in the corpus. Set 1: 30ms to 79ms, set 2: 80ms to 119 ms, set 3:120 ms to 400 ms.

These results are consistent with findings in previous studies in other languages (Lindblom 1963; Gendrot & Adda-Decker 2005; Meunier & Espesser 2011) and confirm the relationship between decrease in vowel durations and hypoarticulation.

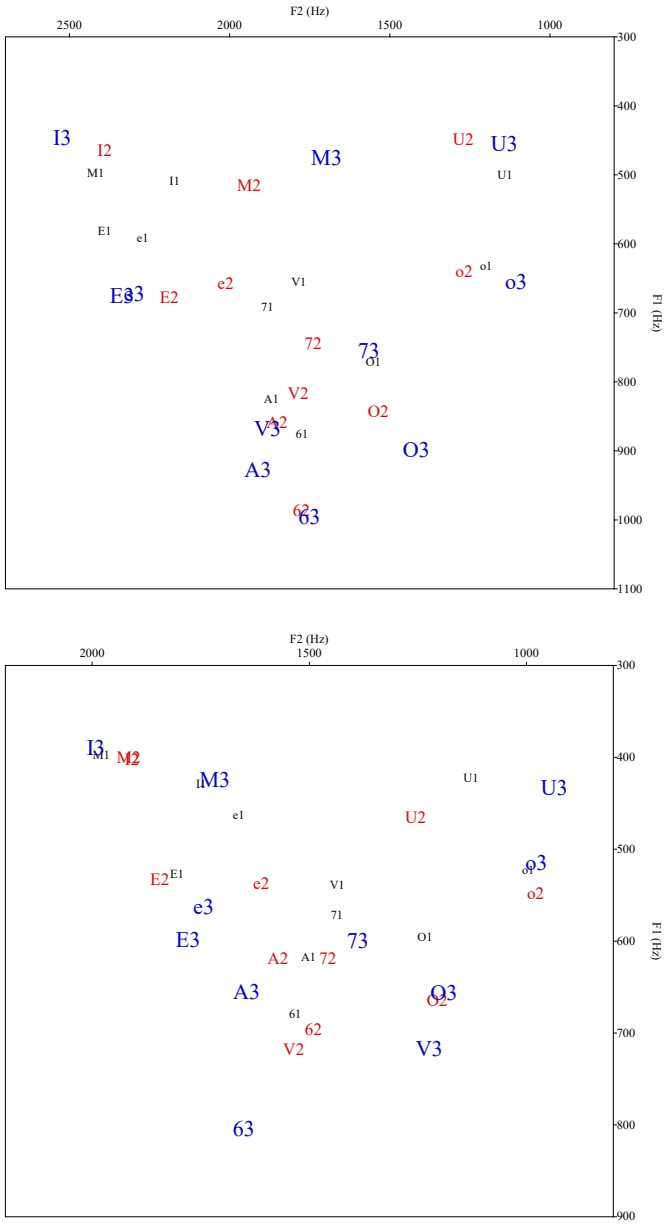

Figure 1.Vowel plots for female (1a:top fig.) and male (1b: bottom fig.) speakers (F1 and F2 values in Hz) according to duration sets (set 1: 30-79ms: number 1 black and smallest font, set 2: 80-119 ms: number 2 red and bigger font, set 3:120-400 ms: number 3 blue and biggest font). Vowels are represented in SAMPA: 6:/ă/, 7: /ɤ/, A:/a/, e:/e/, E:/ɛ/, I: /i/, M:/ ɯ/, o: /o/, O: /ɔ/, U:/u/, and V: /ɤ̆/.

3.2 Vowel spectral and the effect of accent

In order to examine the effect of accent on vowel reduction, mixed effects ANOVA models were conducted on vowel duration and F1, F2 and F3 values. The fixed effects were prominence (prominent vs. non- prominent conditions) and vowels (11 vowels). The random effect was speakers, syllable structures, tones and items. The accented words/syllables were determined based on the listener-identified prominences of another experiment reported elsewhere (Nguyen, submitted). Those words marked as prominence by two-thirds or more of listeners (14 or more of the 20: equal to or above 70%) were considered to have "consensus" agreement and were chosen as accented words for this study.

The results of the ANOVA on vowel duration show significant effect for the main factors prominence: $F(1, 1028) = 79.48$, $p<.0001$ and vowels: $F(10, 1028) = 4.13$, $p<.0001$ and the interaction effect of prominence x vowels : $F(10, 1028) = 5.35$, $p<.0001$. The examination of the interaction effect by the post-hoc Tukey analysis showed that the vowels /a/, /e/, /ɯ/, /ɔ/, /u/, and /o/ were significantly longer in prominence condition than those in non prominence conditions while there was no significant effect for other vowels.

The results of the ANOVA on F1 values show a robust significant effect for the main factors prominence: $F (1, 1028) = 21.45$, $p<.0001$ and vowels: $F(10, 1028) = 56.51$, $p<.0001$ while there was no significant effect for the interaction of prominence and vowels: $F(10, 1028) = 1.43$, $p=0.16$ ns. This raising of F1 values in prominence condition across almost all vowels is confirmed in the vowel plots below. This is consistent with Nguyen and Ingram (2007a, b) and suggests an articulatory gesture enhancement with larger mouth opening and jaw lowering which have been shown by previous researchers (Erickson 2002; Van Summers 1987) to be indicative of stress or prosodic strengthening

The results of the ANOVA on F2 values show a significant effect for the main factors prominence: $F (1, 1028) = 8.64$, $p<.01$ and vowels: $F(10, 1028) = 40.61$, $p<.0001$ and a significant effect for the interaction of prominence and vowels: $F(10, 1028) = 3.63$, $p<.0001$. The examination of the interaction effect by Tukey post-hoc analysis showed that the F2 values of the vowels /ɯ/ and /i/ were significantly lower in prominence condition than those in non prominence conditions while there was no significant effect for other vowels. This is further confirmed in figure 2.

The results of the ANOVA on F3 values show a non-significant effect for the main factor prominence: $F (1, 1028) = 1.2$, $p = 0.27$ ns., a significant effect for vowels: $F(10, 1028) = 3.89$, $p<.0001$ and a non-significant effect for the interaction of prominence and vowels: $F(10, 1028) = 1.39$, $p=0.18$ ns.

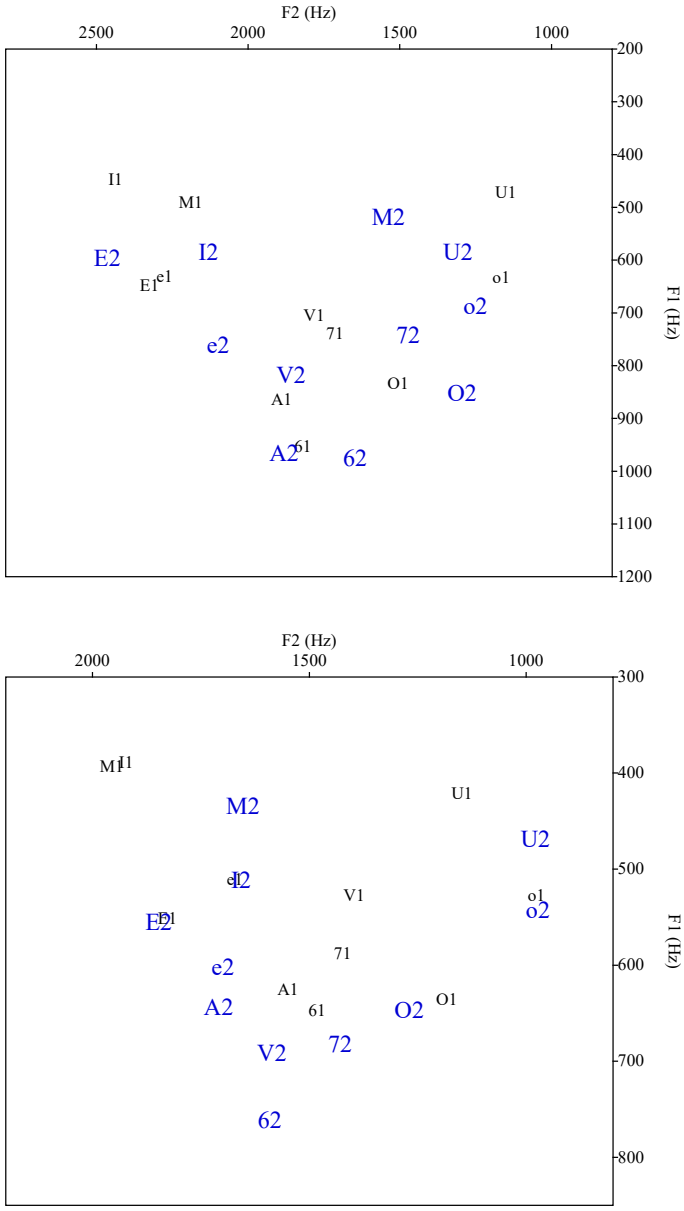

Figure 2. (Colour online) Vowel plots for female (2a: top fig.) and male (2b: bottom fig.) speakers (F1 and F2 values in Hz) according to prominence conditions (non-prominence: number 1 black and smaller font, prominence: number 2 blue and bigger font). Vowels are represented in SAMPA: 6:/ă/, 7: /ɤ/, A:/a/, e:/e/, E:/ɛ/, I: /i/, M:/ ɯ/, o: /o/, O: /ɔ/, U:/u/, and V: /ɤ̆/.

Since more than 70% of the prominent words are pre-boundary as mentioned in the summary of results of another study (Nguyen submitted) in section 2.1 above, the effect of accent on vowel reduction on this section can be said to be confounded with the effect of pre-boundary lengthening. Nevertheless, as this is an intrinsic characteristic of Vietnamese prosody that stress being a syntactic boundary in Vietnamese phrases (Cao 1978, 1998), it is very difficult to tease out the two effects given the nature of the spontaneous conversation data corpus used in this study.

3.3. The effect of positions of vowels in bisyllabic words

Since there is a tendency for Vietnamese words to have two syllables (disyllabic), (it is found that among 41850 entries in the Vietnamese dictionary, 65% were disyllabic, only 20% were monosyllabic and 15% were loanwords and idioms, Nguyen and Ingram 2013), the examination of the effect of positions focuses only on bisyllabic words. Since the status of "word" in Vietnamese is controversial, the determination of monosyllabic words and disyllabic words in the corpus of this study is based on the lexical entries in a Vietnamese dictionary (Nguyễn, Hồ and Nguyễn 2005). A sample list of bisyllabic and monosyllabic words is presented in the appendix. There was in total 308 disyllabic words with monophthong vowels in the corpus.

In order to examine the effect of positions of vowels within a bisyllabic word on vowel reduction, mixed effects ANOVA models were conducted on vowel duration and F1, F2 and F3 values. The fixed effects were vowel positions (first and second positions) and vowels (11 vowels). The random effect was speakers, syllable structures, tones and items.

The results of the ANOVA on vowel duration show a robust significant effect for the main factors positions: $F(1, 308) = 24.43$, $p<.0001$ and vowels: $F(10, 308) = 2.58$, $p<.01$ but the interaction effect of positions x vowels was not significant: $F(10, 308) = 2.52$, $p=0.2$ ns. Mean duration of the vowels was 78ms for the first vowels and 119 ms for the second vowels. This result indicates that the vowels of the first syllables were significantly shorter than those of the second syllables.

The results of the ANOVA on F1 values show a significant effect for the main factors positions: $F (1, 308) = 9.05$, $p<.0001$ and vowels: $F(10, 308) = 40.7$, $p<.0001$ and a weak significant effect for the interaction of position and vowels: $F(10, 308) = 2.08$, $p<.05$. The examination of the interaction effect showed that all vowels, except the vowels /ɤ/, had higher F1 on the second syllables. This raising of F1 values in the second syllables across almost all vowels is consistent with Nguyen and Ingram (2007a, b) and suggests that while the second syllables have articulatory gesture enhancement with larger mouth opening and jaw lowering to be indicative of prosodic strengthening, the first syllables tend to be greatly

reduced as shown in figure 3 below. Particularly, figure 3 shows that the first syllables tend to be shorter, have lower intensity and less full plus shorter tonal shapes compared to the second syllables of the bisyllabic words.

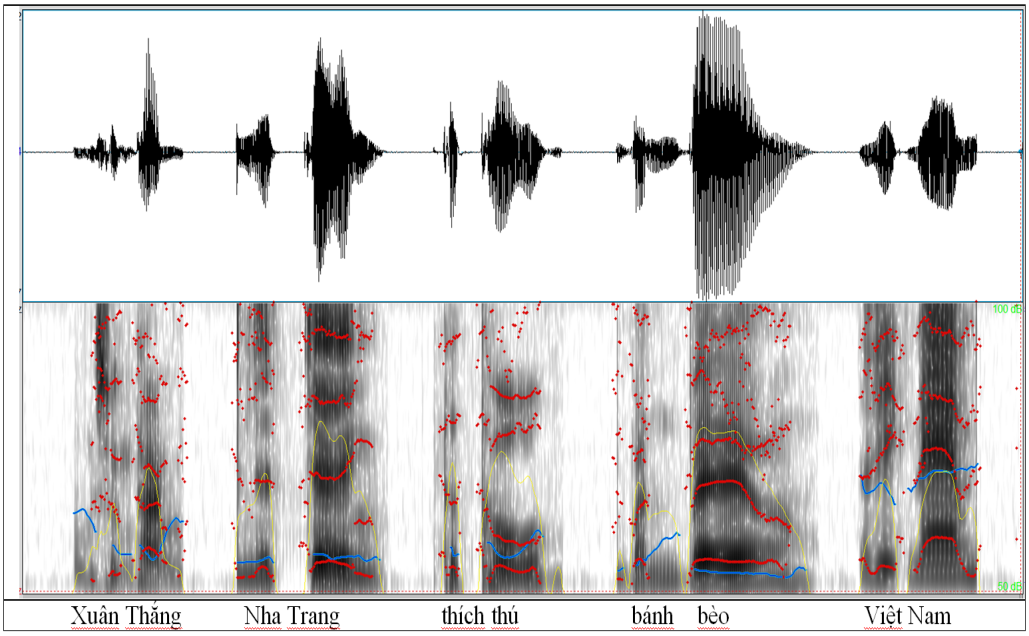

Figure 3. (Color online) Examples of first syllable reduction in disyllabic words. Xuân Thắng: a proper name, Nha Trang: a proper name of place, thích thú: enjoyable, bánh bèo: a kind of cake, Việt Nam: a country proper name.

The results of the ANOVA on F2 values show a non-significant effect for the main factor positions: F (1, 308) = 0.01, p=0.8 ns., a significance for vowels: F(10, 308) = 47.71, p<.0001 and a non-significant effect for the interaction of positions and vowels: F(10, 308) = 0.92, p=0.6 ns.

Similarly, the results of the ANOVA on F3 values show a non-significant effect for the main factor positions: F (1, 308) = 0.13, p = 0.7 ns., a significant effect for vowels: F(10, 308 = 2.98, p<.001 and a non-significant effect for the interaction of positions and vowels: F(10, 308) = 1.13, p=0.5 ns.

3.4. The effect of word category (function vs. content words)

In order to examine the effect of word category on vowel reduction, monosyllabic content words and function words were selected from the excerpts. Content words consisted of nouns

(e.g., nhà: house, quán: shop, phim: movie), verbs (e.g., đi: go, đến: arrive, học: learn), adjectives (e.g., nhỏ: small, đẹp: beautiful, lạnh: cold) and adverbs (e.g., rất: very). Function words included prepositions (e.g., của: of, ở: at), conjunctions (e.g., và : and, nhưng: but, hoặc: or), pronouns (tôi: I, họ: they), copular verbs (e.g., thì, là: to be) and auxiliary verbs (e.g., bị, được: passive form verbs, sẽ: will, đã: past form verb, đang: progressive verb form). There were in total 722 monosyllabic words (350 content words and 372 function words)

Mixed effects ANOVA models were conducted on vowel duration and F1, F2 and F3 values. The fixed effects were word category (function vs. content words) and vowels (11 vowels). The random effect was speakers, syllable structures, tones and items.

The results of the ANOVA on vowel duration show a robust significant effect for the main factors word category: $F(1, 722) = 43.25$, $p<.0001$ and vowels: $F(10, 722) = 3.94$, $p<.0001$ but the interaction effect of word category x vowels was not significant: $F(10, 722) = 0.6$, $p= 0.4$ ns. Mean duration of the vowels was 86ms for the function words and 113 ms for the content words. This result indicates that the vowels of the functions words were significantly shorter than those of the content words.

The results of the ANOVA on F1 values show a significant effect for the main factors word category: $F (1, 722) = 7.16$, $p<.0001$ and vowels: $F(10, 722) = 94.82$, $p<.0001$ and a weak significant effect for the interaction of word category and vowels: $F(10, 722) = 2.09$, $p<.05$. The examination of the interaction effect showed that all vowels, except the vowels /ɛ/, /i/, /ɯ/, and /ɤ/, had lower F1 on the function words. This lowering of F1 values in the function words is consistent with the results of weak prosodic conditions reported in sections 3.2 (unaccented vs. accented) and 3.3 (first vs. second syllables) above and Nguyen and Ingram (2007a, b).

The results of the ANOVA on F2 values show a non-significant effect for the main factors word category: $F (1, 722) = 2.12$, $p=0.18$ ns., a significant effect for vowels: $F(10, 722) = 77.17$, $p<.0001$ and a significant effect for the interaction of word category and vowels: $F(10, 722) = 7.27$, $p<.0001$. The examination of the interaction effect showed that the vowels / ă/, /ɯ/, and /ɔ/ had significantly higher F2 in the function words than in the content words, while /a/ and /u/ had lower F2 in the function words.

The results of the ANOVA on F3 values show a non-significant effect for the main factor word category: $F (1, 722) = 0.07$, $p = 0.7$ ns., a significant effect for vowels: $F(10, 722 = 3.2$, $p<.0001$ and a non-significant effect for the interaction of word category and vowels: $F(10, 722) = 2.2$, $p=0.3$ ns.

In summary, the vowels in the function words tend to have shorter duration and lower F1 values than their content word counterparts, indicating a lesser degree of mouth opening and jaw lowering, an opposite condition to the prosodic strengthening in content words. This

suggests that functions words tend to be reduced, supporting Pham (2008) and findings in other languages (van Bergem 1993; Meunier & Espesser 2011)

4. Discussion and conclusion

First, the results of this study show that vowel duration and spectral quality in Vietnamese are highly related. This is consistent with findings in previous studies in other languages (Lindblom 1963; Gendrot & Adda-Decker 2005; Meunier & Espesser 2011) and confirm the relationship between decrease in vowel durations and hypoarticulation.

Second, the study provides evidence of the effect of accent/prominence on vowel duration and quality. Vowels in accented condition are significantly longer than those in unaccented condition. Particularly, the raising of F1 values in prominence condition across almost all vowels is consistent with Nguyen and Ingram (2007a) and suggests an articulatory gesture enhancement with larger mouth opening and jaw lowering which have been shown by previous researchers (Erickson 2002; Van Summers 1987) to be indicative of stress or prosodic strengthening. This mirrors the findings of the effect of accentuation on vowel reduction in other languages (Swedish: Lindblom 1963; Dutch: van Bergem 1993; Australian English: Harrington, Fletcher & Beckman 2000). However, since more than 70% of the prominent words are pre-boundary as mentioned in the summary of results of another study (Nguyen, submitted) in section 2.1 above, the effect of accent on vowel reduction on this section can be said to be confounded with the effect of pre-boundary lengthening. Nevertheless, as this is an intrinsic characteristic of Vietnamese prosody that stress being a syntactic boundary in Vietnamese phrases (Cao 1978, 1998; Jones and Huỳnh 1960), it is very difficult to tease out the two effects given the nature of the spontaneous conversation data corpus used in this study. Therefore, the effect of accent on vowel reduction in Vietnamese need to be further investigated in future studies.

Third, the study also shows a within-word position effect on vowel duration and spectral quality. Vowels in the first syllables of bisyllabic words were found to have significantly shorter duration and lower F1 values, suggesting that while the second syllables have articulatory gesture enhancement with larger mouth opening and jaw lowering to be indicative of prosodic strengthening, the first syllables tend to be greatly reduced. This is in agreement with Nguyen and Ingram (2007a, b)'s results and consistent with findings in other languages (Adda-Decker 2008; Meunier & Espesser 2011). This result also confirms the prosodic asymmetry of bisyllabic words in Vietnamese (Nguyen and Ingram 2007a, b; Nguyen 2010; Nguyen 2014).

Fourth, the result of this study shows that the production of Vietnamese vowels depends on word category. The vowels in the monosyllabic function words tend to have shorter

duration and lower F1 values than their content word counterparts, indicating a less mouth opening and jaw lowering, an opposite condition to the prosodic strengthening in content words. This suggests that functions words tend to be reduced in Vietnamese, supporting Pham (2008) and findings in other languages (van Bergem 1993; Adda-Decker et al 2008; Meunier & Espesser 2011)

These are some limitations of the analyses on spontaneous speech. That is, some data are unbalanced. Studies on non-controlled speech (or conversational, spontaneous speech) are obviously constrained by the real distribution of words in the speech corpus. Nevertheless, the unbalanced data in this study has been accounted for by the restricted maximum likelihood (REML) applied to mixed model methodology.

A few studies have examined phonetic vowel reduction in conversational speech (Koopmans-Van Beinum 1980 for Dutch; Deterding 1997 for English; Harmegnies & Poch-Olivé 1992 for Spanish), demonstrating shorter vowel durations and 'shrinkage' of the vowel space in conversational speech compared to read speech. Vowels in conversational speech were found to occupy a more central position in the vowel space than vowels in read speech with unstressed vowels being closer to the centroid in both speaking styles (Koopmans-Van Beinum, 1980; Deterding, 1997; Harmegnies & Poch-Olivé, 1992). Vowels in read speech were also found to be less peripheral when compared to formant values given for Greek vowels spoken in isolated sentences, indicating that these vowels were reduced too but to a lesser degree than vowels in conversational speech (Lengeris 2012). In Vietnamese, Nguyen and Ingram (2007a,b) found that in short utterances that are read aloud under well-controlled conditions ('laboratory speech'), vowels of the first syllables of the dissyllabic compounds and reduplications were more reduced (shorter and more centralized) compared to those of the second syllables. Nevertheless, no study on vowel reduction has been done in Vietnamese read speech, and thus this needs to be further investigated in further research.

In summary, this study shows a significant effect of prominence (plus boundary), syllable position and word category on vowel reduction in Vietnamese. Overall, the study provides a precise quantification of vowel reduction and shows that vowel reduction is more pervasive in spontaneous Vietnamese than previously documented.

Acknowledgements

I would like to thank the subjects for their voluntary participation in the experiments and the anonymous reviewers for their constructive comments.

Appendix: A sample list of monosyllabic and disyllabic words

Monosyllabic words	Disyllabic words
Và: and	Bình Dương: proper name
Nhưng: but	Thành phố: city
Cũng: also	Thích thú: enjoyable
Rất: very	Sinh học: biology
đi: go	Sách vở: books and notebooks
Đến: arrive	Thanh bình: peaceful
Tôi: I	Không gian: space
Tìm: find	Học trò: students
Sẽ: will	Môn học: subjects of study
Nhanh: fast	Độc lập: independent

5. References

Adda-Decker, M., Gendrot, C. and Nguyen, N., 2008, Contributions du traitement automatique de la paroleà l'étude des voyelles orales du français. *Traitement Automatique des Langues, 49* (3).

Boersma, P.& Weenink, D. ,2007, Praat: doing phonetics by computer (Version 4.6.09). Computer program. Retrieved June 24, 2007, from http://www.praat.org/ California, Los Angeles.

Brennan, R.& Prediger, D. 1981. Coefficient Kappa: Some uses, misuses and alternatives. *Educational and Psychological Measurement 41*, 687-699.

Cao, X. H., 1978, Trọng âm và các quan hệ ngữ pháp trong tiếng Việt (Stress and syntactic relations in Vietnamese). *Thông báo Ngữ âm học (Bulletin of Phonetics)* (Institute of Social Science, Hồ Chí Minh city).

Cao, X. H. 1998. *Tiếng Việt- Mấy vấn đề ngữ âm, ngữ pháp, ngữ nghĩa* (Some issues in phonetics, syntax and semantics in Vietnamese) (Giáo dục, Hà Nội .

Chaudhary, C. C., 1983, Word stress in Vietnamese: A preliminary investigation. *Indian Linguistics 44*, 1-10.

Deterding, David., 1997, The Formants of Monophthong Vowels in Standard Southern British English Pronunciation. *Journal of the International Phonetic Association 27,* 47-55.

Đoàn, T. T., 1977, *Ngữ Âm Tiếng Việt* (Vietnamese phonology) (Đại học và trung học chuyên nghiệp, Hà Nội).

Đô˜, T. D.,1986, Elements pour une e’tude comparative de l’ intonation en Francais et en

Vietnamien: L'accent de mots en Vietnamien (Elements for a comparative studies of intonation in French and in Vietnamese: the accent of words in Vietnamese). *Memoire de DEA*, Universite' de Paris 3 ILPGA, Paris.

Erickson, D., 2002, Articulation of extreme formant patterns for emphasized vowels. *Phonetica 59,* 134–149.

Flemming, E., 1995, *Auditory representations in phonology*. Ph.D. dissertation. (University of California, Los Angeles).

Gendrot, C. & Adda-Decker, M. , 2005, Impact of duration on F1/F2 formant values of oral vowels: An automatic analysis of large broadcast news corpora in French and German. *Interspeech:* 2453–2456.

Gsell, R., 1980, Remarques sur la structure de l' espace tonal en Vietnamien du sud (Parler de Saigon). *Cahiers d'etudes Vietnamiennes, 4,* 1-26.

Harmegnies, Bernard, and Poch-Olivé Dolors., 1992, A study of style-induced vowel variability: laboratory versus conversational speech in Spanish. *Speech Communication 11*(4-5), 429-437.

Harrington, J., Fletcher, J.& Beckman, M., 2000, Manner and place conflicts in the articulation of accent in Australian English. In Broe. Pierrehumbert. *Papers in Laboratory Phonology V: Acquisition and the lexicon. pp. 40–51* (Cambridge University Press, Cambridge).

Hoàng, T. & Hoàng, M., 1975, Remarques sur la structure phonologique du Vietnamien. Essais Linguistiques (Remarks on the phonological structure of Vietnamese). *Etudes Vietnamiennes (Vietnamese studies) 40*.

Kohler, K.J., 2008, The Perception of Prominence Patterns. *Phonetica 65*, 257–269.

Koopmans-Van Beinum, Florien J.,1980, Vowel Contrast Reduction. An Acoustic and Perceptual Study of Dutch Vowels in Various Speech Conditions. PhD. diss., University of Amsterdam.

Lengeris, A., 2012, Phonetic vowel reduction in Standard Modern Greek. In Z. Gavriilidou, A. Efthymiou, E. Thomadaki & P. Kambakis-Vougiouklis (eds), Selected papers of the 10th International Conference on Greek Linguistics, Komotini/Greece: Democritus University of Thrace, 401-407.

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

Meunier, C.& Espesser, R., 2011, Vowel reduction in conversational speech in French: The role of lexical factors. *Journal of Phonetics 39* (3), 271-278.

Nguyễn , D. L.,1970, A contrastive phonological analysis of English and Vietnamese. *Pacific*

Linguistics Series 8. (Australian National University, Canberra).

Nguyễn, K. T., Hồ, H. T.& Nguyễn, Đ. D. , 2005, *Tự Điển Tiếng Việt* (Vietnamese Dictionary) Culture Publishing House, Saigon.

Nguyễn, T.A.T. & Ingram J.C. ,2007a, Acoustic and perceptual cues for compound - phrasal contrasts in Vietnamese. *The Journal of the Acoustical Society of America 112 (3),*1746-1757.

Nguyễn, T.A.T. & Ingram J.C. , 2007b, Stress and tone Sandhi in Vietnamese reduplications. *Mon-Khmer studies Journal 37,* 15-40.

Nguyễn, T.A.T., Ingram, J. C. & Pensalfini, R.,2008, Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics 36:* 158-190.

Nguyễn, T.A. T. , 2010, Rhythmic pattern and corrective focus in Vietnamese polysyllabic words. *The Mon-Khmer Studies Journal 39,* 1-28.

Nguyễn , T.A.T.& Ingram J.C. , 2013, Perception of prominence patterns in Vietnamese dissyllabic words. *The Mon-Khmer Studies Journal 42,* 89-101.

Nguyễn, T.A. T., 2014, Acoustic correlates of rhythmic structure of Vietnamese narrative speech. *The Mon-Khmer Studies Journal 43*.

Nguyễn, T.A. T., Acoustic correlates of listener-identified boundaries and prominences in spontaneous Vietnamese speech. (submitted).

Pham, A. H., 2008, Is there a prosodic word in Vietnamese? *Toronto Working Papers in Linguistics 29,*1-23.

Randolph, J., 2008, Online Kappa Calculator. http://justus.randolph.name/kappa (date last view April 30th, 2013) .

Roach, P. 2002, A Little Encyclopaedia of Phonetics: http://www.personal.rdg.ac.uk /~llsroach/encyc.pdf (originally Roach, P. 1992. Introducing Phonetics) (Penguin, London 2002).

Thompson, L. 1965. *A Vietnamese Reference Grammar* (University of Hawaii Press, Honolulu).

van Bergem, D. R.,1993, Acoustic vowel reduction as a function of sentence accent, word stress and word class. *Speech Communication 12,* 1–23.

van Summers, W., 1987, Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *J. Acoust. Soc.Am. 82,* 847–863.

Wells, J. C., 1997, SAMPA computer readable phonetic alphabet. In Gibbon. Moore. Winski. Handbook of standards and resources for spoken language systems (Part IV, section B) (Mouton de Gruyter, Berlin and New York).