Acoustic and perceptual cues for compound-phrasal contrasts in Vietnamese

Anh-Thu T. Nguyễn[a] and John C. L. Ingram[b]
School of English, Media Studies, and Art History, University of Queensland, St Lucia, QLD 4072, Australia

(Received 4 January 2006; revised 31 March 2007; accepted 11 May 2007)

This paper reports two series of experiments that examined the phonetic correlates of lexical stress in Vietnamese compounds in comparison to their phrasal constructions. In the first series of experiments, acoustic and perceptual characteristics of Vietnamese compound words and their phrasal counterparts were investigated on five likely acoustic correlates of stress or prominence (f0 range and contour, duration, intensity and spectral slope, vowel reduction), elicited under two distinct speaking conditions: a "normal speaking" condition and a "maximum contrast" condition which encouraged speakers to employ prosodic strategies for disambiguation. The results suggested that Vietnamese lacks phonetic resources for distinguishing compounds from phrases lexically and that native speakers may employ a phrase-level prosodic disambiguation strategy (juncture marking), when required to do so. However, in a second series of experiments, minimal pairs of bisyllabic coordinative compounds with reversible syllable positions were examined for acoustic evidence of asymmetrical prominence relations. Clear evidence of asymmetric prominences in coordinative compounds was found, supporting independent results obtained from an analysis of reduplicative compounds and tone sandhi in Vietnamese [Nguyễn and Ingram, 2006]. A reconciliation of these apparently conflicting findings on word stress in Vietnamese is presented and discussed. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2747169]

PACS number(s): 43.70.Fq, 43.71.Es, 43.70.Bk, 43.70.Kv [PEI] Pages: 1746–1757

I. INTRODUCTION

Compounding is a highly productive process in Vietnamese and in many languages. Compounding may be characterized as a device for creating words (new lexical items) from phrases (compositional syntactic constructions). Phonologically, this may involve a prosodic contrast in which the prosodic characteristics of the phrasal construction are altered to conform to the prosodic template of the word. This is evidenced in many languages which mark the distinction between compound words and their phrasal counterparts prosodically, in ways that are language specific and possibly related to their prosodic type (see Nguyễn *et al.*, in press). In English, where words tend to take initial stress (Cutler and Carter, 1987), this is achieved by assigning left-edge prominence to compounds, in contrast to the right-edged prominence of a phrase. Along with their characteristic left prominence, most English compounds also take on the prosodic characteristics of a word, deaccenting and temporally compressing the second element so as to conform to rhythmic constraints of stress-timing (e.g., *blackberry* versus *black berry*; Nguyễn *et al.*, in press). In tone languages such as Mandarin Chinese (Shen, 1993) and Hmong (Ratliff, 1992), compounds are prosodically distinguished from phrases on the basis of lexical tone sandhi. In Vietnamese, tone sandhi is restricted to reduplications (Thompson, 1965; Ngô, 1984; among others); therefore, in the absence of a general rule of tone sandhi in Vietnamese compounds, it is questioned whether there is an independent prosodic mechanism for distinguishing Vietnamese compounds from their phrasal counterparts, or whether listeners are required to rely exclusively on context to disambiguate phrasal from compound constructions.

The aim of our research was to examine whether there is any prosodically distinctive marking of compound words in Vietnamese (in relation to their corresponding phrases). Two series of experiments were conducted to test (a) whether there is any prominence related contrastive pattern between compounds and phrases in Vietnamese such as one observes in compound phrasal pairs such as *black-bird* versus *black bird* in stress based languages like English, and (b) whether compounds in Vietnamese as disyllabic word forms show any evidence of prosodic asymmetry, such as deaccenting one of the syllables of a disyllabic word.

In the first series, acoustic evidence from native speaker production and perception was sought as to whether there is acoustical and perceptual evidence for contrastive stress patterns in Vietnamese (strong–weak in noun phrases versus weak–strong in compounds). Native speaker production and perceptual discrimination of compounds and phrases was tested under two speaking conditions: (1) a condition of "maximal contrast," where subjects were asked to produce minimal prosodic pairs of the target stimuli in a manner that would "bring out the differences in meaning" between a compound and its corresponding phrase; and (2) in a "picture naming task," where there was no particular focus upon the possible differentiating phonetic features of the compound-phrasal contrast. The maximal contrast condition was in-

[a]Electronic mail: thunguyen@uq.edu.au
[b]Electronic mail: j.ingram@uq.edu.au

tended to test whether the language provides a way of prosodically marking distinctions in meaning between syntactically generated phrases and their corresponding lexical compounds. The picture naming task asks the same question but under conditions intended to more faithfully reflect circumstances of normal speech production.

In a second series of experiments, bisyllabic coordinating compounds which have a free variation form with reversed constituent syllable positions (e.g., *bàn*[table] *ghế*[chair] versus *ghế*[chair] *bàn*[table]: furniture) were used to investigate prosodic asymmetry (left or right prominence) at the level of the disyllabic word.

The patterns of syllable comparisons consist of a paradigmatic contrast between homophonous syllables of a compound and a phrase (e.g., *hoa hồng*: noun phrase: a pink flower and *hoa hồng*: compound: a rose) and a crossover contrast between homophonous syllables in a coordinative compound (*bàn ghế*) and its free variation form (*ghế bàn*). Thus, in the first series of experiments, a set of paradigmatic comparisons could be made to test for acoustic correlates of a possible stress contrast between compound and phrasal minimal prosodic pairs and to assess the perceptual impact of such a contrast, under two conditions of elicitation. In a second experiment, syntagmatic comparisons could be made between the first and second position within a class of coordinative compounds that happen to be freely reversible in phonological form. In both cases, it was possible to control for effects of speaker, speech rate, tonal and segmental composition in making the comparisons, but obviously, in the second case it was not possible to test for a compound-phrasal perceptual contrast, as both forms were allomorphs of the same compound word.

A. Compounding in Vietnamese

Vietnamese noun phrases and compounds are, of course, contrastive in terms of meanings (compositional versus noncompositional) and morphosyntactic structures (phrases versus words). There are different types of compounds in Vietnamese. Thompson (1965) identified six different types, namely, idiom compounds, generalizing coordinative compounds, specializing compounds, reinforcing compounds, attributive compounds, and pseudocompounds. The present study restricts itself to the investigation of prosodic features of two compound types: (1) specializing adjunctive compounds whose meanings are "fundamentally related to the meanings of the corresponding phrases but are generally highly specialized in reference" (e.g., *hoa hồng*: noun phrase: *a pink flower* versus compound: *a rose*) (Thompson, 1965, p. 129), and (2) generalizing coordinative compounds made up of two coordinative bases/heads which denote a general class (e.g., *bàn ghế* versus *ghế bàn*: compound: *furniture*); In terms of syntactic structure, Vietnamese compounds and phrases have a reversed word order from English (e.g., Vietnamese: head noun +adjective modifier: *hoa*[flower] *hồng*[pink] versus English: Adjective+Noun: *black berry*). Phonologically, it is claimed by some linguists that the pattern of prominence in Vietnamese compounds, particularly specializing compounds, is the reverse of the English pattern; that is *weak-strong* for compounds "with weak stress on their first base/syllable" (Thompson, 1965; Trần, 1969; Ngô, 1984) and *strong-weak* for noun phrases (Thompson, 1965, p. 121). By contrast, there are contradictory views about the stress patterns of coordinating compounds. According to Thompson (1965), this compound type "frequently occurs with weak stress on their first syllable" (p. 128), while Ngô (1984) and Cao (2003), relying on the "morphosyntactic weight" of the syllable, claimed that, "both constituents received strong stress" (Ngô, 1984, p. 275). Nevertheless, whether the contrastive pattern of prominence of Vietnamese compounds and noun phrases is acoustically supported and whether Vietnamese compounds show evidence of prosodic asymmetry (e.g., deaccenting one of the syllables of a disyllabic word) needs to be investigated.

B. Stress in Vietnamese

Lexical stress has been found to be instantiated by a variety of phonetic cues, i.e., duration, pitch (fundamental frequency—F0), intensity, and vowel quality (Beckman, 1986; among others). It is widely agreed that stress is phonetically realized in a language-specific way. Stress in a tone language such as Mandarin Chinese was shown to affect duration and the surface output of lexical tones in Mandarin: F0 range is expanded in stressed (i.e., focused) syllables and narrower or compressed in unstressed or postfocal contexts (Xu, 1999). In Vietnamese, a tone language, no system of lexical word stress has been found; nevertheless, it is widely accepted that there is stress in the sense of accentual prominence at the phrasal level (Thompson, 1965; Nguyễn, 1970). Some studies have attempted to investigate the phonetic correlates of Vietnamese stress, but mostly in the sense of phrase level accentual prominence and generally based on auditory impressions. Đỗ (1986) has shown that duration and intensity are important parameters for describing stress in Vietnamese. Chaudhary (1983) remarks that intensity is one stable acoustic correlate of Vietnamese stress. Some other authors, such as Hoàng and Hoàng (1975), or Gsell (1980) consider that full tonal realization of accented syllables is one of the positive marks of accent. Jones and Huỳnh (1960) remark that normally stresses in a Vietnamese utterance are conditioned by junctures.

A review of the literature shows that previous studies of stress in Vietnamese (Cao, 2003; Ngô, 1984; among others) have been largely based on a notion of "morphosyntactic" weight of syllables (e.g., "Assign[stress] to the core of a morphosyntactic category" Ngô, 1984, p. 101) and "stress" in the sense of phrase level accentual prominence (Thompson, 1965; Nguyễn, 1970; Chaudhary, 1983; Hoàng and Hoàng, 1975; Jones and Huỳnh, 1960), but have not addressed the question of stress at the word level, from the perspective of acoustic phonetics and native speakers' perception, which is the aim of this study.

II. EXPERIMENT 1. COMPOUND AND PHRASAL STRESS PRODUCTION

In this section, we report the first series of experiments involving minimal prosodic pairs of compound words and their corresponding phrasal constructions elicited under different conditions of "maximal contrast" and "picture naming." We report an analysis of the acoustic characteristics that might be expected to differentiate compounds and phrases if speakers are producing them in a manner that distinguishes the two forms. This is followed (in Sec. III) by a perceptual experiment that assessed the discriminability of the compound-phrasal pairs by native Vietnamese listeners, from tokens obtained under the two conditions of elicitation.

Two competing hypotheses are evaluated in these experiments;

(1) Hypothesis 1: This hypothesis is to test whether there is acoustical evidence of contrastive stress patterns (strong-weak in noun phrases versus weak-strong in compounds). This hypothesis predicts that the vowel of the first element (V1) in a noun phrase will have longer duration, stronger intensity, larger F0 range, possibly higher F0, and fuller vowel quality than the same vowel in a compound and vice versa, the vowel in the second element (V2) of the compound will have longer duration, stronger intensity, larger F0 range, possibly higher F0, and fuller vowel quality than the same vowel of the phrase.
(2) Hypothesis 2: This hypothesis predicts that in case no strong evidence of duration, intensity, and pitch as cues to the contrastive stress patterns is found, a junctural pause between two constituents of the phrase may serve (under special conditions of maximal contrast) as the acoustic and perceptual cue to the distinction between compounds and noun phrases in Vietnamese. If hypothesis 2 turns out to be correct, it raises the question of whether a compound-phrasal prosodic contrast is part of the phonology of Vietnamese or whether Vietnamese speakers are making use of a prosodic device for syntactic phrase disambiguation (juncture) which is available universally. "Juncture" as a phonological category is probably a complex phonetic construct, cued by preboundary lengthening, pausing, and possibly also pitch contour modification (though this may well be much less prominent in a tone language). In this study, the pausing between the two constituent syllables of the test items and the prepausal lengthening of the first elements are examined in order to decide whether or not there is a juncture.

A. Linguistic materials

Fifteen pairs of two-syllable compounds and their corresponding phrases were constructed from three types of compounds, formed on the basis of their grammatical structures: noun-adjective (NA type), noun-verb (NV type), and noun-noun (NN type), illustrated in Table I.

There is an important distinction between the phrase types that correspond to three types of compound used in the experiment. The NA and NV compounds derive from syntactic constituents that form a clause and are dominated by an S node in standard syntactic representation. The NN compounds, on the other hand, derive from noun phrases, dominated by the lower-ranked syntactic constituent NP. It is well known from prosodic phonology that juncture breaks are much more likely to occur between major clausal constituents than within a phrasal constituent, such as an NP (Selkirk, 1986; Nespor and Vogel, 1986).

TABLE I. Type of compound, source phrase, and syntactic rank.

Type	Example	Compound	Phrase	Rank
N-A	hoa hồng	*Rose*	*Flower [is] pink*	Clause
N-V	bò cày	*Ox-plough*	*Ox [is] ploughing*	Clause
N-N	chân vịt	*Propellar*	*Duck's foot*	NP

B. Subjects and test conditions

Two groups of native speakers ($n_1=30$, $n_2=15$), aged 18–22, balanced for selection by gender and by the three major dialect regions of Vietnam (northern: Hanoi, central: Hue, and southern: Saigon) were separately tested for production of compound and phrasal constructions under two elicitation conditions.

Under the *maximal contrast* condition, subjects were asked to read minimal sentence pairs in a natural way, such that listeners could distinguish between the meaning of a compound and its corresponding phrase. Subjects in this experiment were 30 first-year university students in Vietnam (speakers of Hanoi [$n=10$], Hue [$n=10$], and Saigon dialects [$n=10$); half males and half females in each dialect group, age ranged 18–22). The test items were embedded in sentences having the same grammatical structure and word order. All test items occurred in utterance nonfinal position. Examples of these sentences are:

Compound: ***Hoa hồng*** thì đẹp (A *rose* is beautiful)

Phrase: ***Hoa hồng*** thì đẹp (A *pink flower* is beautiful)

For the *picture naming* task, a different group of 15 overseas Vietnamese students (speakers of Hanoi [$n=5$], Hue [$n=5$] and Saigon [$n=5$]; nine females and six males, age ranged 22–45) at the University of Queensland were asked to describe a picture, using a constant carrier sentence to ensure that the target phrase or compound word appeared in nonfinal position in an utterance with the same number of words, tone coarticulation, and speaking rate effects. Two examples:

Có(there is/are)+numeral/classifier
+compound/phrase+ ở đây (here)

Có một bông **hoa hồng** ở đây (There is a rose here)
Có hai con **cá mập** ở đây (There are two sharks here)

Recordings were made in a quiet room using sound recording and editing computer software (PRAAT, Boersma and Weenink, 2007), at 20 kHz sampling rate and 16 bit precision.

C. Measurements

The test items were segmented via the Emu Speech Tools, (Cassidy, 1999). First, the Emu Labeller was used to

TABLE II. Percentages of juncture pauses in noun phrases elicited under minimal pair sentence and picture naming tasks by phrase types and dialects.

	Minimal pair sentence			Picture naming		
	Hanoi	Hue	Saigon	Hanoi	Hue	Saigon
NA	84	80	82	0	0	0
NN	16	14	15	0	0	0
NV	66	60	66	0	0	0

mark the edges of the target syllables and vowels, relying primarily on the spectrographic display in the Labeller. The segmentation criteria were generally based on the major discontinuities of the energy distribution over frequency and time visible on the spectrograms. Then, the Emu-R statistical software was used to extract vowel duration (ms), vowel formant (Hz), intensity (db), spectral tilt, and fundamental frequency (Hz). The following acoustic parameters were measured:

(1) Duration of first and second syllables (onset[$O1O2$], vowel[$V1V2$], coda[$C1C2$], and whole syllable[$S1S2$]); duration of the pause (if longer than 100 ms) between the syllable constituents of the phrases/compounds.
(2) First formant and second formant of both first and second syllables ($S1F1$, $S1F2$, $S2F1$, and $S2F2$). Formant at vowel midpoint was taken for monophthongs while formant at ten equidistant points of the formant trajectory was taken for diphthongs/triphthongs.
(3) Tone range (F0range=$F0$max-F0min): $S1F0range$ and $S2F0range$.
(4) Mean of vowel intensity (db) at four equidistant points ($V1$ *intensity mean and* $V2$ *intensity mean*).
(5) Vowel spectral tilt (H1-A3): third formant is compared with the first harmonic using Stevens and Hanson's model (1995) ($V1$ *spectral tilt and* $V2$ *spectral tilt*). An acoustic correlate of greater vocal effort (and thus stress) is expected to be a decrease of a negative spectral tilt (more gradual fall of the spectrum).

D. Statistical analysis of stress contrasts

Rather than regarding the maximal contrast (minimal pair) and picture naming production tasks as levels of a single treatment factor in a unified factorial design, the two conditions were analyzed in separate experiments. Different subjects were required so that the effects of exposure to one elicitation condition did not contaminate or bias responses elicited under the other condition. There were unequal numbers of subjects under the minimal pair [$n=30$] and picture naming tasks [$n=15$]. Also, the two experiments were undertaken separately, with the picture naming experiment conceived as a replication and extension of the original experiment which was conducted under conditions of maximal contrast elicitation.

A mixed (fixed and random) effects analysis of variance (ANOVA) model, using the restricted maximum likelihood method to estimate variance components was applied in all of the foregoing analyses. The fixed effects included prosodic type (compound versus phrase), compound type (NA, NN, NV), dialects (Hanoi, Hue, and Saigon), and their two- and three-way interactions. The random effects were speakers and words. A Tukey post-hoc test was then conducted to determine the significant differences among levels of the main fixed factors and their interaction effects.

In addition, a ratio measure of the dependent variable in a phrase over its corresponding compound (e.g., for the duration parameter: duration $S1_{phrase}$/ duration $S1_{compound}$) was calculated for presentation in graphs. Thus, values greater than 1.0 reflect a higher magnitude for the phrasal token on a given acoustic parameter, and conversely, a value significantly less than 1.0, a greater magnitude for the compound token than its corresponding phrase. In this way, any effects upon the acoustic parameter of interest caused by intrinsic segmental composition, tone type, or speaker were controlled for in testing for the effects of prosodic type.

Spectrographic analysis was used to identify and measure any juncture pauses (defined as breaks longer than 100 ms) that tended to occur between constituents of a phrase under the maximal contrast elicitation condition.

E. Results on juncture pauses

Juncture pauses were only observed in the maximal contrast elicitation condition and appeared mainly in two types of noun phrase constructions: noun-adjective and noun-verb, not in noun-noun phrases. Table II shows the percentage of juncture pauses used in phrases by phrase types. Juncture pauses occurred in approximately 80% of the noun-adjective phrase constructions, about 65% of the noun-verb phrase constructions, but in only 15% of the noun-noun phrase constructions. Pausing was *never* used (0%) in compounds and in the picture naming task, regardless of prosodic type (compound nor phrase).

These results strongly suggest that the speakers were employing a juncture (phrase boundary) marking strategy to differentiate phrases and compounds in the maximal contrast (minimal pair) condition, but were not doing so under the picture naming elicitation condition. We sought to confirm this and to determine whether any *other* mechanism might be operative from a series of ANOVAs of the paradigmatic comparisons on the five prosodic parameters described earlier.

F. Statistical results of prosodic parameters

The effects of interest to this study are prosodic type (compound versus phrase), and its interaction with the fac-

TABLE III. Summary of the patterns of acoustic difference between CP and PH. Symbol < >: sig. at $p<0.01$, Ns.: insignificant.

Acoustic cues	MAX-S1	MAX-S2	PIC-S1	PIC-S2
Duration	PH>CP		Ns.	Ns.
F0 range	PH>CP	CP>PH: inconclusive	Ns.	Ns.
Formant F1	PH>CP		Ns.	Ns.
Intensity	PH<CP	PH<CP	Ns.	Ns.
Spectral tilt	Ns.	Ns.	Ns.	Ns.

tors compound type (NA, NN, NV) and dialect (Hanoi, Hue, Saigon). The main effects of compound type and dialect, while of some intrinsic interest, are not central to the hypotheses tested in this paper if significant. Particularly, the significant main effect of prosodic type was shown to be consistent across speakers of the three dialects in a preliminary analysis; therefore, dialectal variation is discussed only when relevant. The pattern of significant effects of the ANOVA results (Table III) closely paralleled the distribution of junctural pauses (Table II), providing *prima facie* support for the hypothesis that *all* the prosodic effects which differentiate compounds from phrases are linked to the distribution of junctural pauses (phrase boundary effects).

There were no significant effects observed under the picture naming elicitation (PIC), whereas robust main effects of prosodic type and interactions between prosodic type and compound type were found under the maximal contrast elicitation for the dependent variables of syllable duration, F0 range, and vowel first formant. Intensity yielded a main effect for the phrasal-compound contrast under minimal pair elicitation (the maximal contrast condition, MAX) but no interaction effect with compound type. We turn now to a detailed consideration of these effects in terms of each acoustic parameter type, which is reported in two main sections: (1) results of the PIC experiment and (2) the results of the MAX experiment.

1. Results of the picture naming experiment

There were no significant main or interaction effects for any factors of the picture-naming task. This is illustrated in the PH/CP ratio measurement in Fig. 1 that there was no significant difference in any acoustic parameters between compounds and phrases across three compound types indicated by the PH/CP ratio clustering around a value of 1 (unity).

2. Results of the maximal contrast experiment

The statistical results and ratio measurements (Fig. 2) showed that there were highly significant main effects of prosodic type for syllable duration, F0 range, intensity, and vowel first formant (F1) for the first syllable in the minimal pair sentence task, while there were no or very marginally significant effects for the second syllable.

a. S1 duration. The ANOVA results showed that the first syllable (S1) had a significant main effect of prosodic type ($F(1,860)=971.16$, $p<0.0001$) and its interaction with compound type ($F(2,860)=19.21$, $p<0.0001$). Tukey post-hoc comparisons and a plot of mean phrase/compound (PH/CP) ratios by compound types (Fig. 2) showed that the first syllables of all three word/phrase types (NA, NN, NV) of the phrase was significantly longer than that of the corresponding compound (mean PH/CP ratio >1). Nevertheless, Fig. 2 also shows that the magnitude of lengthening was much greater for NA and NV than NN constructions, indicated by greater PH/CP mean ratios. Further examination of the duration of onset, vowel, and coda of the first syllables showed that all segments (onset, vowel, and coda) of the first syllable of the phrase were significantly longer than their counterparts in the compound ($p<0.01$).

b. S2 duration. There was a significant main effect of prosodic type ($F(1,859)=8.61$, $p<0.01$) and two-way interactions not only with compound type ($F(2,859)=20.93$, $p<0.0001$) but also with dialect ($F(2,859)=9.68$, $p<0.0001$). Post hoc tests showed that significant effects for the second syllables were only found for the NV construction (CP >PH: mean PH/CP ratio <1, Fig. 2). Examination of the syllable constituents showed that while the onset of the second syllable of the phrase was longer than its counterpart in the compound across the three compound types, a significant effect (compound>phrase, $p<0.01$) was found for only the vowel and coda of the NV compound type, this is argued to be an artifact of the juncture that separated the compound from its following preposition phrase (see Sec. V E). Post-hoc pairwise examination of the interaction of prosodic type and dialect showed a dialectal effect that the second syllable of the NA and NN noun phrases of only the Hanoi speakers was significantly longer than its compound counterpart (PH >CP: mean PH/CP ratio >1).

To sum up, the duration results showed that all segments (onset, vowel, and coda) of the first syllable of the phrase under maximal contrast elicitation were significantly longer than their counterparts in the compound across the three dialects and three compound types (though NN had smaller lengthening magnitude than NV and NA). By contrast, only the vowel and coda of the second syllable of the NV compound was longer than its counterpart in phrases. There was also a dialectal effect that in the northern (Hanoi) dialect, the second syllable of the NN and NA noun phrase was longer than its compound counterpart, suggesting that Hanoi speakers seem to place emphasis on both syllables of the noun phrase. Further discussion on duration results is presented in Sec. V E.

c. S1 F0 range. The first syllable (S1) had a significant main effect for the factor prosodic type ($F(1,858)=24.94$, $p<0.0001$), its two-way interaction (prosodic type × compound types: $F(2,858)=20.11$, $p<0.0001$), and three-way interaction (prosodic type × compound types × dialects: $F(4,858)=3.8$, $p<0.01$). Post hoc pairwise comparisons sup-

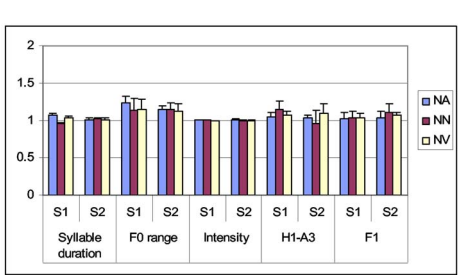


FIG. 1. (Color online) Mean of PH/CP ratios of the acoustic parameters by syllables (S1 and S2) and compound types (NA, NN, NV) of the picture naming experiment.

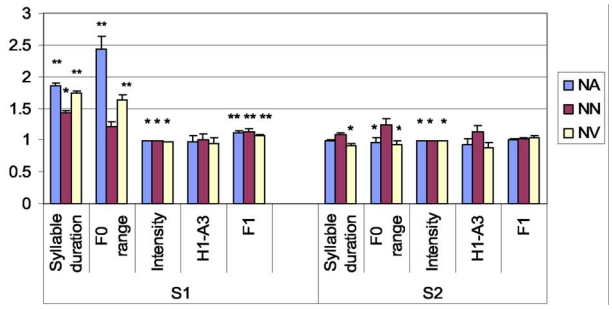


FIG. 2. (Color online) Mean of PH/CP ratios of the acoustic parameters by syllables (S1 and S2) and compound types (NA, NN, NV) of the maximal contrast experiment. (**) significant at $p<0.0001$, (*) $p<0.01$.

ported by a graphical display of the mean of PH/CP ratio by compound types (Fig. 2) showed significant compound-phrase differences for the NA and NV compound types only, with no significant effect for the NN type. The first syllable of the NA and NV noun phrase had a significantly greater F0 range than its counterpart in compounds (PH/CP ratio >1) across the three dialects. A detailed analysis of the common tone types (falling, dropping, level, rising) across three compound types showed that the tonal range expansion was found for the falling and rising tones only (PH$>$CP, $p<0.01$), while the level and the dropping tone did not show significant effects, possibly due to articulatory constraints on their F0 expansion (level is a flat tone; drop tends to end with creakiness).

d. S2 F0 range. The ANOVA results showed a significant main effect of prosodic type ($F(1,859)=22.78$, $p<0.0001$), a two-way interaction (prosodic type$\times$dialect: $F(2,859)=4.92$, $p<0.01$) and a weak three-way interaction (prosodic type$\times$compound types$\times$dialects: $F(4,859)=2.39$, $p<0.05$). Post hoc results showed that the second syllable of the NA and NV compound had significantly greater F0 range than that of the phrase (PH/CP ratio <1, Fig. 2). However, detailed analysis by the common tone types (curve, falling, rising) showed significant effects for the curve tone and the falling tone of the NA type (CP$>$PH, $p<0.01$) while no significant difference was found for the same falling tone of the NV type. The significant effect of the falling tone of the NA compound types can be explained as an artifact effect of tonal coarticulation in the compound pattern, due to the high level and rising tone of the preceding syllables (*áo*, *bánh*, *cá*, *hoa*); this is well-supported by a statistically significant difference in the starting F0 of this falling tone between (compound and phrase, $F(1,90)=3.4$, $p<0.03$). The significant effect of the curve tone of the NV compound types was found to be contributed by the Southern dialect only (Southern: $F(1,60)=5.6$, $p<0.03$; Central: $F(1,30)=1.5$, $p=0.2$ ns., Northern: $F(1,30)=0.9$, $p=0.3$ ns). Examination of the F0 contour showed that the Southern speakers produced the curve tone in the compound with a higher rising tail after the creakiness than in the phrase, probably due to more articulatory space given by a following juncture break after the compound pattern of this NV compound type, which will be discussed in detail in Sec. V E on the duration effect.

An examination of the three-way interaction effect (prosodic type$\times$compound type$\times$dialects) on the second syllable showed that the Southern (Saigon) speakers had significantly greater F0 range expansion than speakers of the other two dialects. This is consistent with the findings of previous cross-dialect tonal studies that the Southern dialect has greater F0 range than the Northern and Central dialects (Seitz, 1986).

In summary, the F0 range results showed a robust effect (PH$>$CP) for the first syllable across the three dialects, while the effect for the second syllable appears to be only an artifact of tonal coarticulation.

e. Intensity and spectral tilt. The ANOVA results on mean intensity of four equidistant points along the vowel showed a significant main effect of prosodic type for both syllables (S1: $F(1,860)=19.64$, $p<0.0001$; S2: $F(1,859)=7.54$, $p<0.01$). Mean PH/CP ratios by compound types (Fig. 2) showed that mean intensity of both syllables (S1 and S2) of compound patterns were significantly greater than their counterparts in the phrase (mean PH/CP ratio <1). Further examination of the four equidistant points on the intensity contour showed that there were significant differences (CP$>$PH) in the last two points of the first syllable and in the first three points of the second syllable (statistical significance at $p<0.01$).

Spectral tilt: The spectral *slope* measurements (H1-A3) showed no significant effects.

f. Vowel formants. The ANOVA analysis of F1 and F2 formant values of both syllables (S1 and S2) showed only a significant main effect of prosodic type (PH$>$CP, Fig. 2) for the first syllable ($F(1,860)=22.51$, $p<0.0001$). As confirmed in the monophthong vowel plots (Fig. 3), the raising of the F1 of the vowel of the phrases in comparison with that of the compound was found only in the first syllable under the minimal pair sentence task [MAX-S1: Fig. 3(a)] across three dialects, while vowels of both prosodic types (CP and PH) in the second syllable [MAX-S2, Fig. 3(b)] tend to assemble in the same region of the vowel space.

An examination of the formant trajectories of the diphthongs/triphthongs showed similar patterns; there was a significant difference in formant trajectory between phrase and compound for the first vowel of the minimal pair sentence task only (MAX-S1) in comparison to the formant trajectories of diphthongs under the three other positions (MAX-S2, PIC-S1, and PIC-S2). The diphthong/triphthongs /aw/(in *áo*) of the first syllable of the phrase had significantly lower F2 in the offglide than its counterpart in compound, the diphthong /wa/ (in *hoa*) had a significantly higher F2, and the triphthong /ɯɤi/(in *ngu'ò'i*) had a lower F2 in the onglide /ɯ/ portion and higher F2 in the offglide.

G. Discussion

1. Duration

The duration results showed that all segments (onset, vowel, and coda) of the first syllable of a phrase were significantly longer than their counterparts in a compound across the three dialects and three compound types (though NN had a smaller lengthening effect than NV and NA)—but only under the maximal contrast elicitation condition. The lengthening of the first syllable of the phrases seemed to bear a close relationship with the juncture pause between the constituents of that phrase (Table II) elicited under the minimal pair sentence task: The NN word/phrase type is different from the other two phrase types NA and NV. There were very few junctures between two syllables of the NN type and

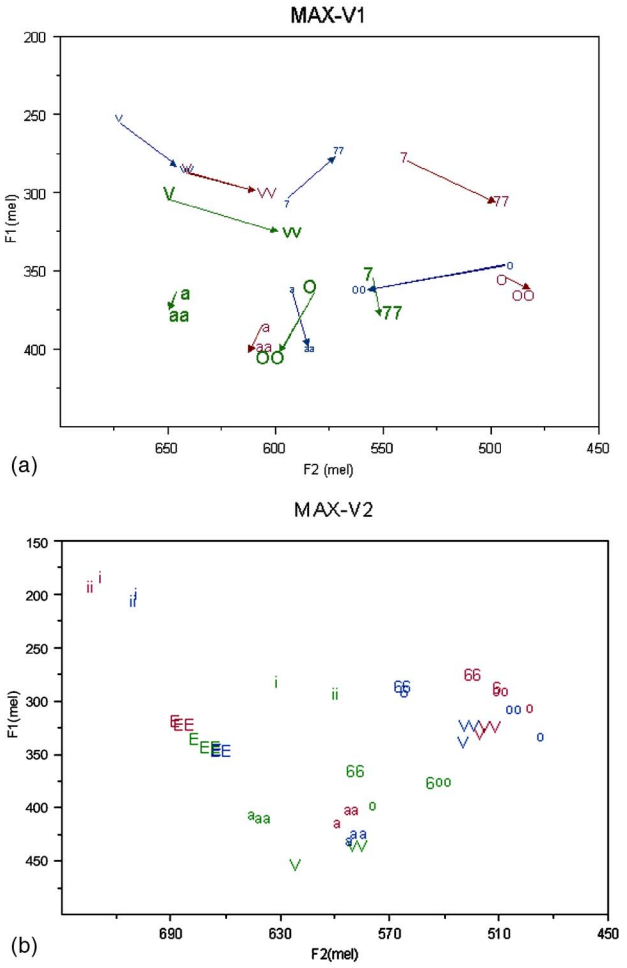
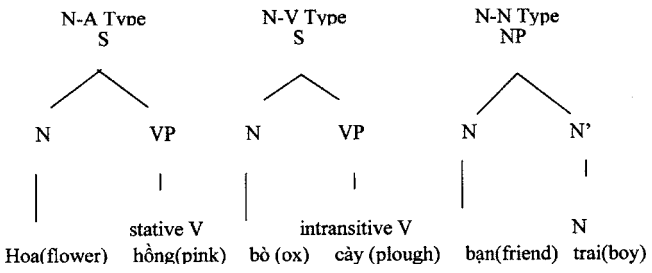


FIG. 3. (Color online) (a), (b) Formant plots of monothong by dialects and prosodic type. Vowels are represented in SAMPA symbols. Single character: CP, double character: PH. Legends for dialects: Largest font: Southern [Saigon], smaller font: Northern [Hanoi]; smallest font: Central[Hue]. Vowel symbols: V: /ʌ/, 7: /ɤ/, a:/a/, 0:/ɔ/ (fig. a), o:/o/ (fig. b), E:/ɛ/, i:/i/.

the lengthening effect on V1 of the NN type was less than that of the NA and NV types (Fig. 2). In fact, as mentioned previously and discussed below, juncture is not allowed between elements of the NN type. The few junctures that were found in this type appear to be the result of some speakers' carry-over effect from the NA and NV types. The absence of a juncture between the constituents of the items in the NN group and the lengthening to a smaller extent of the first constituent in comparison to that of the NA and NV types can be explained on the basis of prosodic boundary strength in the theory of prosodic phonology (Nespor and Vogel, 1986 or Selkirk, 1986). The following diagram shows the syntactic trees of the three phrase types.



The propensity to prosodically mark a constituent boundary by a juncture pause or by preboundary lengthening is related to the rank of the syntactic constituent so marked. The boundary between two phrases in a clause, outranks any boundary between two constituents of a noun phrase. Vietnamese requires an intervening copular verb "*là*" (e.g., *bạn trai*: bạn *là* nam: friend *is* boy/male) or a verb phrase (nhà đá: nhà được *làm bằng* đá: house *is made of* stone) to form a predicative phrase from a sequence of two nouns.

Another effect of duration is the significant lengthening of the second constituent of compounds compared with their counterparts in noun phrases, but this only applied to compounds of the NV type. Supported by evidence that only the vowel and the coda of the compound was significantly lengthened in comparison to their counterparts in the phrase, it can be argued that this lengthening could be due to an artifact of a juncture pause that most speakers made between the compounds and its following prepositional phrase, which occurred in all five sentences of this phrase type (NV) but not in the two other phrase types (NA and NN).

There was also a dialectal effect on duration. In the northern (Hanoi) dialect, the second syllable of NN and NA noun phrases was significantly longer than their compound counterparts, suggesting that Hanoi speakers may put emphasis on both syllables of the noun phrases to distinguish it from compounds.

2. F0 range

The first syllable of the NA and NV noun phrase had significantly greater F0 range than its counterpart in compounds across three dialects. By contrast, the significant effect for the second syllable is only an artifact of tonal coarticulation. Combined with other acoustic correlates (duration, vowel quality, and intensity), the greater F0 range only on the first syllable of the phrase in comparison to compounds suggests a fuller tone realization thanks to prepausal syllable lengthening. This is supported by the fact that for the NN word/phrase types, there was almost no juncture and thus less syllable lengthening, no significant difference was found for F0 range between phrase and compound.

3. Intensity and spectral tilt

Both syllables of the phrase in the minimal pair sentence task had lower intensity (db) in comparison to their counterparts in the compound. This suggests an effect of speaking rate deceleration and gesture diminution at a phrase boundary. Again, the effect is confined to the maximal contrast elicitation and appears to be a by-product of a juncture pause distinguishing the phrase from the compound.

4. Vowel formants

The raising of the F1 of the vowel of a phrase in comparison to that of a compound was confined to the first syllable of the maximal contrast elicitation across the three dialects. This is arguably an effect of jaw lowering in response to syllable lengthening.

Taken together, as shown in Table III, the differences between phrases and compounds seem to be confined mainly to the first syllable of the phrase in the minimal pair sentence task, which is accompanied by a junctural pause. The results suggest a boundary-induced lengthening effect which is supported by not only duration but also intensity and vowel formant results. However, the evidence of a more fully realized tone and a lengthening of all syllable segments (including onset) also seems to be indicative of an additional accentual effect which may have been "conditioned by juncture," as observed previously by Jones and Huỳnh (1960) that "normally the stresses in a Vietnamese utterance are conditioned by junctures." Speakers might have intended to put more emphasis on the grammatical head noun of the noun phrase (which is also parsed as the subject of the adjective clause or the communicative theme) to distinguish it from the compound word. This is consistent with the dialectal effect that Northern (Hanoi) speakers lengthened both words of the phrase to distinguish it from the compound.

In summary, the acoustic analysis shows that Hypothesis 1 is not supported: there was no significant effect on any of the acoustic parameters between compounds and noun phrases elicited under the picture naming task, which more closely represents spontaneous natural speech. Even under conditions of maximal contrast in the minimal pair sentence task, there was no consistent significant asymmetrical prominence in terms of spectral tilt, intensity, or duration and F0 range to distinguish between contrastive elements of compounds and noun phrases. Only a phrase-level junctural effect, prepausal lengthening, and expanded F0 range on the first syllable was found between phrasal constituents, supporting hypothesis 2. No evidence was found of an independent word prosody to distinguish compounds from phrases, such as operates in a stress language like English. Given that the only acoustic features in native speakers' production that appear to distinguish compounds from phrases in Vietnamese emerged under conditions of maximal contrast and apparently involved a "juncture marking strategy," it remained to be investigated whether a parallel situation pertained in the domain of perception of compound–phrase contrasts. The answer to this question was the purpose of the following perception experiment.

III. PERCEPTION EXPERIMENT

Having argued on the basis of acoustic analysis of production data that Vietnamese speakers rely on juncture to mark phonological contrasts between compounds and phrasal constructions under conditions of maximal contrast, it remains to show that Vietnamese listeners may only be able to distinguish between phrasal and compound constructions when junctural cues are provided by the speaker. There is evidence (from the production experiment) that the junctural cues are not possible in NN phrasal constructions. The prediction is, therefore, that even under conditions of maximal perceptual contrast (minimal prosodic pair discrimination) Vietnamese listeners will fail to distinguish compounds from NN phrases.

A. Subjects

Eighteen subjects (six Hanoi, six Hue, and six Saigon; ten females) with no known auditory deficiencies, participated in the perception experiment. Only three of the listeners were also speakers for the production experiments. Since Hanoi dialect is a prescribed national standard in instruction and national broadcasting, listeners from other dialects, particularly the well-educated ones, generally have no difficulty understanding Hanoi dialects, but the reverse does not hold. Listeners of this study are all tertiary students at the University of Queensland and have been in Australia from 4 months to 1.5 years.

B. Stimuli

Stimuli of four Hanoi female speakers from the production experiment (two speakers from the picture-naming task and two from the minimal pair sentence task) were used for the perception study. The criteria used for the selection of stimuli were that all four speakers spoke standard Hanoi dialects and had an easy-to-hear voice, which was judged to be clear, loud enough to hear, and easy to understand for speakers of other dialects by the experimenter. The two speakers from the minimal pair sentence task consistently produced a juncture break between syllables of the NA and NV noun phrases but no juncture between the NN noun phrases and compound stimuli. The two speakers from the picture-naming tasks consistently produced no pause, gap, break, or whatever between syllables of phrases and compounds.

There were two types of stimuli: (1) compounds/phrases in isolation and (2) compounds/phrases embedded in contextual sentences. The test items in isolation were segmented from the contextual sentences via speech analysis software (PRAAT). There were in total 240 tokens: 15 test items $\times 2$ prosodic types (CP vs PH) $\times 4$ speakers (2 in picture naming task versus 2 in minimal pair sentence task) $\times 2$ stimulus types (compounds/phrases in isolation versus compounds/phrases in contextual sentence). The tokens were put in random order each with two immediate repetitions in block of ten with a gap of about 6 s between each item. There were in total 24 blocks (12 blocks of target items in isolation and 12 blocks of contextual sentences containing the target items).

C. Procedures

The listening identification test consisted of two parts: in part one, listeners heard the target items presented in isolation; in part two they heard a contextual sentence containing the target item. Part one was performed before part two to avoid a carry-over effect of contextual sentence perception. Two different meanings of the target item were given in the answer sheet: one as a compound and one as a noun phrase. The subjects' task was to choose the meaning which they judged was expressed by the speaker by circling the letter corresponding to their response in the answer sheet.

The experiment was delivered online so that subjects could either perform the test via their personal computer at home or at the researcher's office.

D. Results

A five-way fixed effect ANOVA (2 prosodic types [CP vs PH] ×2 experiments [MAX vs PIC] ×2 task types [word/phrase vs sentence] ×3 compound types [NA, NN, NV] ×3 dialects [HN, HU, SG]) was conducted on the percentage number of correct responses (the number of listeners over the total listeners in a dialect group who chose the correct response for a testing item). It is worth noting that a preliminary analysis showed that there was no significant difference between two speakers in the same experiment (MAX and PIC), and thus speakers as a factor was excluded from this analysis. Significant main effects were obtained for all factors except dialects (prosodic types: $F(1,1)=258$, $p<0.0001$, experiments: $F(1,1)=144$, $p<0.0001$; task types: $F(1,1)=17$, $p<0.0001$, compound types: $F(1,2)=10$, $p<0.0001$, dialects: $F(1,2)=0.3$ ns). Five significant two-way interaction effects were found including, prosodic types ×experiments: $F(1,1)=137$, $p<0.0001$, prosodic types ×compound types: $F(1,2)=9.35$, $p<0.0001$, prosodic types×dialects: $F(1,2)=11.93$, $p<0.0001$, prosodic types ×tasks: $F(1,1)=11.24$, $p<0.001$ and experiments ×compound types: $F(1,2)=18.15$, $p<0.0001$. The significant three-way interactions consisted of prosodic types ×experiments×compound types: $F(1,2)=40.32$, $p<0.0001$, prosodic types×compound types×tasks: $F(1,2)=4.69$, $p<0.01$, and a four-way interaction prosodic types ×experiments×compound types×tasks: $F(1,2)=11.27$, $p<0.0001$. Examination of the five main effects showed that compounds had significantly more correct response than phrases (mean scores: 73.5% vs 47.3%, $p<0.01$). The test items elicited under the maximal contrast experiment (MAX) also had more accurate responses than those elicited under the picture naming task (PIC) (mean: 70 vs 50.6, $p<0.01$). Test items embedded in a contextual sentence (S) yielded more correct responses than those presented in isolation (W) (mean: 63.7 vs 57.0, $p<0.01$). The NA and NV compound types with a juncture break between syllables of the phrase had more correct responses than the NN type (NA: 64.9, NV: 60.3, NN: 56, $p<0.01$). In contrast, there was no significant difference in correct perception scores among the three dialectal groups of listeners (HN: 59, HU: 60, SG: 61).

Examination of the prosodic types×dialects interaction showed that Northern (Hanoi) listeners had more correct compound responses but fewer correct phrase responses than listeners in the other two dialects (mean percentage scores compound versus phrase, respectively: Hanoi: 78 vs 40, Hue: 73 vs 51; Saigon: 70 vs 50, $p<0.01$), indicating that they misclassified phrases more for compounds even though they were listening to speakers of their own dialects.

Examination of the four-way interaction effect prosodic types×experiments×compound types×tasks (Fig. 4) showed three things which were consistent with the main effects. (1) The compound stimulus items from the PIC experiment had significantly more correct responses than the phrase, indicating a response bias toward the compound form and that the phrase in this experiment tended to be misclassified as compound due to the absence of a juncture break. (2) This is further supported by evidence that the phrase of the NN word/phrase type from the MAX experiment, which also lacks a juncture break between its two syllables, was more misheard as compound, indicating that listeners expected a juncture pause between components of phrases as a distinction from compounds. In addition, performance for stimuli with an absence of a juncture between constituents of the phrase was at chance level (clustering around 50%), indicating listeners' confusion between the compounds and phrases. In contrast, in patterns that have a juncture between syllables of phrases (NA and NV under MAX experiment), identification was better, though not optimal (approximately 70%) and had equal rate of success between the compound and phrase patterns. (3) Generally, there was a very marginal difference between items embedded in a contextual sentence and items presented in isolation, indicated by similar patterns of variation across three compound types for both the MAX and the PIC experiments. This further strengthened the finding that a juncture break between the two components of the phrase is essential for the distinction between phrase and compound regardless of whether they were presented in context or isolation. In other words, to listeners, the presence of a junctural break is more important than the contextual sentence in assisting the discrimination between meanings. This also lends further support to the acoustic results of the production experiment in indicating that there is neither acoustic nor perceptually salient evidence for a prosodically distinctive prominence contrast between the phrase and compound in Vietnamese, except for a juncture break between two components of the phrase but only under maximal contrast conditions. The difference between the NN and the other two compound/phrase types NA and NV under condition of maximal contrast also indicated speakers' parsing strategy in trying to convey the distinction between phrase and compound by relying on syntactic structure. The adjectives are also stative verbs in Vietnamese and thus the adjective in NA and the verb in NV constructions can be parsed as verb phrases of the adjective clauses while the noun cannot.

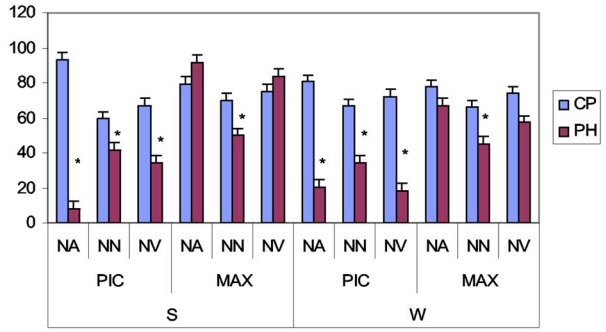


FIG. 4. (Color online) Mean of responses by four way interactions: prosodic types×compound types×experiments×task types. (*) sig. $p<0.01$.

IV. CONCLUSION: COMPOUNDS AND PHRASES IN PRODUCTION AND PERCEPTION

The results of the production experiments indicate that Vietnamese speakers only produce consistent acoustic cues to distinguish compounds from phrases when required to do so by speaking under conditions of maximal contrast and do

so by employing a juncture marking strategy that is probably universally available. Consistent with this strategy, the first components of the phrasal constructions were lengthened in comparison with their corresponding compounds accompanied by an expanded tone range, intensity dropping, and jaw lowering, which suggests a prepausal articulatory effect. No conclusive evidence of enhanced fundamental frequency, duration, or intensity prominence was found on the second components of compounds. Compound words as a whole were not temporally compressed in comparison to their phrasal counterparts as in English, a stressed language.

The results of the perception experiments indicate that listeners relied only on the juncture between the two components of noun phrases as a cue to distinguish between noun phrases and compounds. The results that listeners failed to distinguish noun phrases from compounds in the NN types and in stimuli elicited under picture-naming task where there is no juncture between the two constituents and performed better in patterns that have a juncture between syllables of noun phrases (NA and NV under MAX experiment) regardless of whether they were presented in context or isolation strengthens the case that there is neither significant acoustic nor perceptual evidence of contrastive stress patterns (strong-weak versus weak-strong) between noun phrases and compounds, contrary to previous claims in the literature (Thompson, 1965; Trần, 1969; among others). There was no significant dialectal variation in the production or perception of compound — phrasal contrasts. Hanoi speakers differed from speakers of the other two dialects in lengthening both syllables of the phrase in comparison to compound but were not more successful than listeners of Hue and Saigon in perceptually distinguishing compounds and phrases produced by speakers of their own dialect.

The acoustic and perceptual evidence appears to converge and is consistent with native speaker intuitions that phrases and their corresponding compounds are usually pronounced identically in Vietnamese, even in careful speech. However, it must be conceded that the *paradigmatic comparisons* upon which the acoustic evidence of the foregoing experiments rely is less than ideal to investigate this question. Prosodic features, particularly lexical stress patterns, probably rest heavily on *syntagmatic* comparisons between syllable sized units in the speech stream. Such comparisons were not possible given the phonetic structure of the compound phrasal constructions examined in this experiment. But Vietnamese provides us with another subclass of compound expressions—coordinative compounds with a free variation of reversible syllable constituents, which provide segmental feature control that enables a sensitive evaluation of asymmetries of prosodic prominence between adjacent syllables of disyllabic words.

V. EXPERIMENT 3: COORDINATIVE COMPOUNDS

In this experiment, essentially the same acoustic parameters as experiment 1 were analyzed, including syllable duration (onset, nucleus, and coda), F0 range, F0 contour, vowel intensity, spectral tilt, and vowel formant structure, to determine which syllable of the two (the first or second in a crossover contrast (e.g., in *bàn ghế* vs. *ghế bàn*) is more prominent.

A. Hypothesis

The basic hypothesis for testing was that coordinative compounds in Vietnamese are right headed i.e., show signs of acoustic prominence on the second syllable. But it could be argued that any acoustic prominence observed on the second syllable of a disyllabic compound may simply be a (word) boundary effect (such as final lengthening), rather than an accentual prominence effect. To meet this objection, the fine-structure of the temporal organization of the target syllable (including onset, vowel, and coda) will be examined. Accentual prominence characterized by vocal and articulatory gesture enhancement on a syllable might be expected to express itself as much in the onset as the coda, whereas word boundary lengthening may be expected to differentially lengthen the coda (Edwards and Beckman, 1988). Therefore, apart from other accompanying acoustic correlates (e.g., a larger tone range, greater intensity, and/or a full/unreduced vowel), the second syllable (S2) is considered to be more acoustically/accentually prominent than its corresponding syllable in the first position (S1) if it has a longer duration across all segment constituents including onset, vowel, and coda. By contrast, if only the syllable coda gets lengthened, the effect may be simply a preboundary lengthening.

B. Method

Ten minimal pairs of coordinative compounds were used in this experiment to elicit speech from the same subjects and using the same procedure as in the picture naming task; i.e., the same 15 subjects were asked to describe a picture, using the target word in a constant carrier sentence.

C. Analysis

Signal processing and measurements of fundamental frequency, duration, intensity, spectral tilt, and vowel formant were identical to those reported in experiment 1 (Sec. II C).

The statistical analysis involved a mixed model ANOVA, with two fixed factors: Syllable position (S1 and S2) and dialects (Hanoi, Hue, and Saigon), and two random factors: syllables (20 syllables) and speakers (15 speakers). The acoustic parameters (listed in Sec. II C) were the dependent variables. In addition, a ratio measure of S2/S1 (e.g., for the duration parameter: S2 duration/S1 duration) was also calculated and presented in graphs.

D. Results

The ANOVA results showed a significant main effect of syllable position for syllable duration and its constituent segments (syllable: $F(1,1551)=175.48$, $p<0.0001$, onset: $F(1,1524)=6.08$, $p<0.02$, vowel: $F(1,1551)=281.98$, $p<0.0001$, coda: $F(1,1153)=8.56$, $p<0.01$), tone duration: $F(1,1551)=156.59$, $p<0.0001$, F0 range: $F(1,1551)=19.39$, $p<0.0001$, mean intensity: $F(1,1551)=7.03$, $p<0.01$ and vowel first formant: $F(1,1551)=7.8$, $p<0.01$.

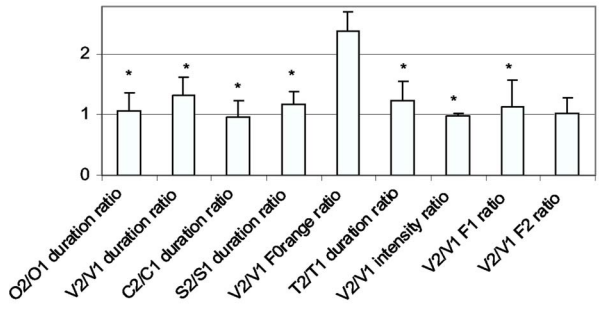


FIG. 5. (Color online) Mean S2/S1 acoustic values of coordinative compounds. (*) sig. $p<0.01$. Legend: O2/O1: onset, V2/V1: vowel, C2/C1: coda, T2/T1: tone.

No significant effect was found for the factor of dialect or its interaction with syllable positions. There was no significant effect for spectral tilt values. Mean ratio in Fig. 5 showed that the constituent syllable of the coordinative compound has longer syllable duration (longer onset and vowel but shorter coda), longer tones, larger tone range, and higher vowel first formant (except a lower intensity value) when it is in the second position than in the first position.

1. Duration

The second syllable was shown (Fig. 5) to have significantly longer duration than the first syllable. Examination of the lengthening mechanism of the constituent segments of syllables that have an onset, vowel, and coda (e.g., bạn, chồng, quần, sách, vịt, xuồng, Fig. 6) showed that the syllable in the second position had significantly longer onset (O duration), vowel (V duration) and overall syllable duration (S duration) but a marginally shorter coda (C duration) than the same syllable in the first position. This evidence, particularly the negative effect of final lengthening on the coda, is counterindicative of a preboundary lengthening effect but supportive of the hypothesis of a right-headed accentual effect, which is also supported by other acoustic cues: F0 range and shape and vowel formant measurements (indicative of F1—F2 space expansion).

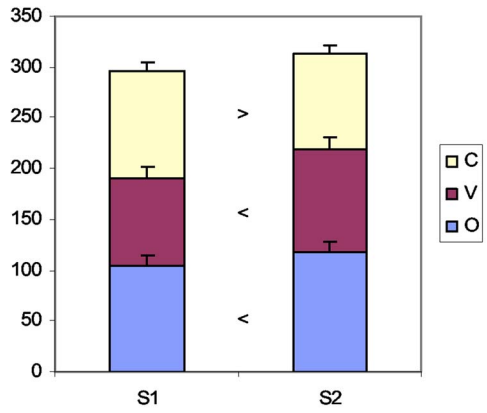


FIG. 6. (Color online) Duration of segments (onset, vowel, coda) of syllables in coordinative compounds. ⟨or⟩: significant patterns at $p<0.01$.

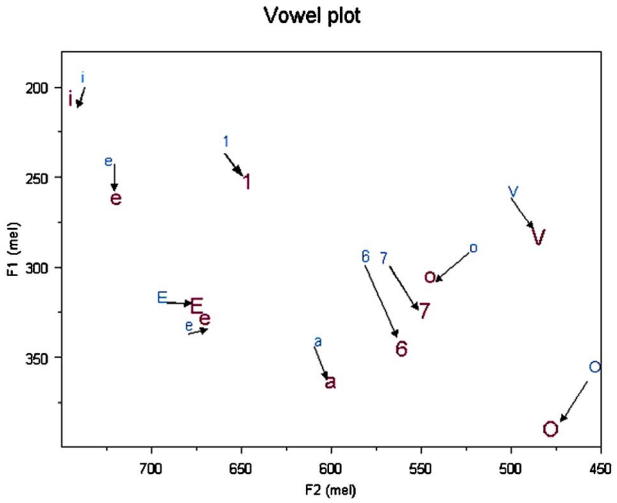


FIG. 7. (Color online) Vowel plots of coordinative compounds. Symbols: V: /ʌ/, 7: /ɤ/, a:/a/, O:/ɔ/, o:/o/ E:/ɛ/, e:/e, i:/i/, l:/ɨ/, 6:/ɐ/. Smaller font: first syllable, larger font: second syllable. The vowel /ɨ/ produced by Southern (Saigon) speakers.

2. F0 contour and F0 range

The examination of F0 contour showed that a syllable in the second position has a longer and fuller tonal shape in addition to a larger tonal range than one in the first position. For example, the second rising (sac) tone rose higher. The second falling (huyen) tone fell more sharply and lower. The second curve(hoi) and drop(nang) tone fell more sharply and lower (across three dialects) and then rose higher than its first position counterpart in the Southern dialect, consistent with findings from previous studies about the effect of stress on the surface output of lexical tones (Xu, 1999; Gsell, 1980; Hoàng and Hoàng, 1975).

3. Vowel formants

Figure 7 shows that the vowel in the second syllable has higher F1 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.

4. Intensity and spectral tilt

No significant effect was found for spectral tilt. The second syllable had a marginally lower intensity value (dB) than the first syllable, suggesting a possible diminuation of vocal effort word or phrase finally at word boundaries.

E. Discussion

The results of this experiment showed that the second syllables of coordinative compounds are more acoustically prominent than first syllables, evidenced by a longer duration, larger tone range, fuller tonal shape, and gesture enhancement, suggesting a right-headed stress pattern. This supports Thompson's observation (1965) and discredits the prediction of Cao (2003) and Ngô (1984) of an "equivocal"

stress pattern for coordinative compounds on the ground of "morphosyntactic weight." Moreover, the results of this experiment are consistent with our recent findings on reduplicative compounds which were shown to have a right-headed prominence pattern on both phonetic and phonological grounds (Nguyễn and Ingram, in press). Taken together, these results strongly suggest a word-level stress pattern in Vietnamese.

VI. CONCLUSION

There is superficial tension (but no underlying inconsistency) between the principal findings reported in this paper. From the first series of experiments, it is clear that Vietnamese speakers only produce, and listeners only consistently hear phonetic contrasts between compounds and phrases under conditions of maximal contrast, in which speakers mark a phrasal construction with a juncture between the left and right elements of the phrase, to distinguish it from the compound word—something that they would not normally do unless specific demands of communication require it. Vietnamese apparently lacks the equivalent of a word level "compound rule" such as English and many other languages possess, which assigns a distinctive stress pattern to compound words via a mechanism such as lexical deaccentuation.

On the other hand, from the second experiment, involving syntagmatic comparisons in reversible adjunctive compounds (and other special compound constructions (Nguyễn and Ingram, in press) there was clear acoustic evidence that Vietnamese disyllabic word forms are not symmetrical in terms of accentual prominence, but "right-headed" or biased in weight toward the second element. These findings address a long-standing controversy over the existence of word stress in Vietnamese and more recently, the question of the "prosodic word" in a highly syllable oriented tone language such as Vietnamese. It is possible that word stress levels exist only as a phonetic tendency in Vietnamese, upon which no distinctive phonological contrasts depend. This is a matter of active inquiry, with theoretical implications for a typology of prosodic systems and of practical interest for phonetic and phonological transfer effects in second language learning.

ACKNOWLEDGMENTS

The postdoctoral research fellowship granted to T. A. T. N. by the University of Queensland is gratefully acknowledged. We would like to thank our subjects for their voluntary participation. We are grateful to the anonymous reviewers of the previous manuscript for their constructive criticisms and the anonymous reviewers of the current journal for their valuable comments and suggestions. Thanks to Dr. Bert Remijsen for sharing his praat scripts and N. H. N. for his statistical advice.

Beckman, M. E. (**1986**). *Stress and Non-Stress Accent* (Foris, Dordrecht).
Boersma, Paul & Weenink, David (2007). Praat: doing phonetics by computer (Version 4.6.09) [Computer program]. Retrieved June 24, 2007, from http://www.praat.org/
Cao, X. H. (**2003**). *Tiếng Việt: Mấy Vấn đề Ngữ âm ngữ Pháp Ngữ nghĩa*, Vietnamese: issues of phonology, syntax, and semantics (Giáo dục Press, Danang).
Cassidy, S. (**1999**). "Compiling multi-tiered speech databases into the relational model: Experiments with the Emu system," Proceedings of Eurospeech '99, Budapest, September 1999.
Chaudhary, C. C. (**1983**). "Word stress in Vietnamese: A preliminary investigation," Indian Linguist. **44**, 1–10.
Cutler, A., and Carter, D. M. (**1987**). "The predominance of strong initial syllables in the English vocabulary," Comput. Speech Lang. **2**, 133–142.
Đỗ, 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.
Edwards, J., and Beckman, M. E. (**1988**). "Articulatory timing and the prosodic interpretation of syllable duration," Phonetica **45**, 156–174.
Erickson, D. (**2002**). "Articulation of extreme formant patterns for emphasized vowels," Phonetica **59**, 134–149.
Gsell, R. (**1980**). "Remarques sur la structure de l' espace tonal en Vietnamien du sud (Parler de Saigon)," (Remarks on the structure of tonal space in Southern Vietnamese (spoken in Saigon)), Cahiers d'etudes Vietnamiennes, 4, Universite'de Paris 7.
Hoàng, T., and Hoàng, M. (**1975**). "Remarques sur la structure phonologique du Vietnamien. Essais Linguistiques (Remarks on the phonological structure of Vietnamese)," Etudes Vietnamiennes (Vietnamese studies), Vol. **40**, Hanoi.
Jones, R. B., and Huýnh, S. T. (**1960**). *Introduction to Spoken Vietnamese* (American Council of Learned Societies, Washington, DC).
Nespor, M., and Vogel, I. (**1986**). *Prosodic Phonology* (Foris, Dordrecht).
Ngô, T. N. (**1984**). "The syllabeme and pattern of word formation in Vietnamese," Ph.D dissertation, New York University, New York.
Nguyễn, D. L. (**1970**). "A contrastive phonological analysis of English and Vietnamese," *Pacific Linguistics Series* (Australian National University, Canberra), Vol. **8**.
Nguyễn, T. A. T., and Ingram, J. (**2006**). "Reduplication and word stress in Vietnamese," Proceedings of the 11th Australian International Conference on Speech Science & Technology, University of Auckland, New Zealand, 6–8 December.
Nguyễn, T. A. T., Ingram, J., and Pensalfini, R. (**2008**). "Prosodic transfer in Vietnamese acquisition of Vietnamese contrastive stress patterns," J. Phonetics **36**.
Ratliff, M. (**1992**). "Meaningful tone: A study of tonal morphology in compounds, form classes, and expressive phrases in White Hmong," in *Monograph Series on Southeast Asia* (Northern Illinois University Center for Southeast Asian Studies, DeKalb, IL), Vol. **27**.
Seitz, P. (**1986**). "Relationships between tones and segments in Vietnamese," Ph.D. thesis, University of Pensylvania (UMI).
Selkirk, E. (**1986**). "On derived domains in sentence phonology," Phonology Yearbook **3**, 371–405.
Shen, X. S. (**1993**). "On the temporal dimension of Mandarin," Acta-Llinguistica-Hafniensia **26**, 143–159.
Stevens, K. N., and Hanson, H. M. (**1995**). "Classification of glottal vibration from acoustic measurements," in *Vocal Fold Physiology: Voice Quality Control*, edited by O. Fujimura and H. Hirano (Singular, San Diego).
Thompson, L. (**1965**). *A Vietnamese Reference Grammar* (University of Hawaii Press, Honolulu).
Trần, H. M. (**1969**). "Stress, tone and intonation in South Vietnamese," Ph.D thesis. Australian National University, Malden.
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.
Xu, Y. (**1999**). "Effects of tone and focus on the formation and alignment of F0 contours," J. Phonetics **27**, 55–105.