Native and Vietnamese Production of Compound and Phrasal Stress Patterns.

Thu Nguyen and John Ingram

School of English, Media studies, and Art history
University of Queensland, QLD 4072, Australia

Abstract
This study examines prosodic transfer effects in the production of three contrastive English stress patterns at the level of word and phrase prosody by Vietnamese learners of English. Both languages employ distinctive patterns of pitch (F0) and intensity prominence to signal contrasts between otherwise homophonous compound and phrasal constructions, though reversed in headedness (English: black-bird vs black bird; Vietnamese: flower-pink [rose], flower pink [pink flower]). However, English but not Vietnamese requires compounds to conform to temporal constraints of word prosody. Both languages also signal contrasts between broad and narrow (contrastive) focus (e.g. ‘A black bird...not a white one’). Comparisons between advanced and beginner Vietnamese learners productions of these three stress patterns reveal good accommodation to L2 pitch and intensity targets but imperfect adaptation to timing constraints that distinguish word from phrasal constructions in English.

1. Introduction
This paper examines the acoustic features that native and non-native speakers (Vietnamese learners of English) use to produce the three contrastive stress patterns: compound, broad and narrow focus phrases, as shown by the following minimally contrastive triplet:

a. This is a kind of fruit. It is a blackberry. (compound word)
b. This berry is black. It is a black BERRy. (broad focus noun phrase)
c. This berry isn’t green. It is a BLACK berry. (narrow focus noun phrase)

The triplet illustrated above can be said to exemplify three types of prominence. The compound blackberry has a primary lexical stress on black, which is the accent-bearing syllable in a three-syllable word. The broad focus noun phrase black berry has phrasal stress (or default accent assignment) on BER-, the first syllable of berry. In the narrow focus noun phrase black berry the syllable BLACK receives an emphatic or contrastive stress. Both the compound word (a.) and the narrow focus phrase (c.) have accentual prominence on black the left hand element of the construction, distinguishing them from the broad focus phrase (b.), which is often deemed to be “right headed”. The compound blackberry as a single three-syllable word should be temporally or rhythmically distinct from either of the two-word phrasal constructions. Specifically, the constituent syllables in blackberry may be expected to be shorter than their counterparts in the phrasal construction: a well-attested effect in English and other stress-timed languages (Lehiste, 1972; Fowler, 1977).

Vietnamese also permits the construction of contrastive prosodic triplets that distinguish broad from narrow focus in a noun phrase and contrastive patterns of headedness or prominence for compound and phrasal constructions, though the manner of phonetic realization is different. The Vietnamese tonal system severely restricts F0 being used as marker of ‘stress’, which is a controversial feature in Vietnamese phonology. Also, while there is evidence that Vietnamese may use juncture (pausing) and pre-pausal lengthening to signal contrasting head/dependency relations in compounds and phrases (Ingram and Nguyen, 2002), there is no evidence that Vietnamese compounds are subject to temporal or rhythmic constraints of word prosody, because Vietnamese entirely lacks a system of culminating word stress. The main aim of this and related investigations (Ingram, Nguyen & Pensalfini in review and Nguyen & Ingram in review)) was to gain a better understanding of how typological prosodic differences (Vietnamese: a tonal language, syllable timed; English: a stress-accent language, foot timed) impact upon the acquisition of L2 prosody in speech production.

2. Production Experiment

2.1. Methods
2.1.1. Linguistic Materials
Three sets of minimally contrastive triplets of compound (word = C), broad focus (phrase = B), and narrow focus (phrase = N) were constructed, using three syllabic templates: monosyllabic first element plus disyllabic second element (e.g.: black berry); disyllabic first element plus monosyllabic second element (e.g.: butter fish); disyllabic first and second element (e.g.: English teacher). There were four tokens for each syllable type, yielding 12 sets of triplets or 36 items for each speaker. Each item was made up of a short context sentence, followed by fixed carrier sentence (This/it/he is a ....... ) which ensured that target contrasts appeared as the final elements in each sentence and were in approximately the same position in the sentence intonation contour. (See item example 1a-c above).

2.1.2. Subjects
Three groups of subjects were used for this experiment. Ten Vietnamese beginning learners of English were first year English-major university students in Hanoi. Ten postgraduate students at the University of Queensland who were teachers of English and had been in Australia from 8 months to 10 years served as advanced Vietnamese speakers of English, and a control group consisted of four native Australian speakers of English. There were half male and half female in each group.

2.1.3 Procedures
Before the recording, subjects were provided sufficient time for familiarization and practice. They then read the text aloud three times in their normal speaking manner. Only the third repetition was recorded and used for analysis. The recording was made in a quiet room using a sound recording and editing computer software (Speech Station) at 20 kHz sampling rate and 16 bit precision. It is
noted that a five-minute review of the three stress patterns and explanation of the meaning of the triplets were given to the beginner group only so as to make sure they produce the correct stress patterns.

2.2. Measurements
2.2.1. Duration
Studies of the effects of stress and accent on duration in English have shown that not only the rhymes but also the initial consonants are lengthened relative to their counterparts in unstressed syllables (Umeda, 1977; Ingrisano & Weismer, 1979; among others). Therefore, in this experiment, the duration of the whole syllable, including the onset and the rhyme was measured.

2.2.2. Fundamental Frequency and Intensity
The peak point of F0 and intensity was taken for stressed syllables of the test items. The point of measurement was taken at the center of vowels.

2.3. Analysis
The acoustic analysis concerns fundamental frequency (F0), duration and intensity of the constituents of test items. Nine acoustic parameters were investigated: mid-vowel pitch value of the first and second stressed syllable (e.g., English teacher: V1F0, V2F0), mid-vowel intensity value of the first and second stressed syllable(V1 and V2 intensity), pitch change(V1F0-V2F0), intensity change(V1 intensity-V2 intensity), duration of the first and second stressed syllables (e.g., English teacher: S1 and S2), and duration of the whole compound words or noun phrases (blackberry, English teacher).

Two-way ANOVAs (3 Stress patterns x 3 Proficiency levels) with pair comparisons between stress patterns (e.g., Broad vs. Narrow, Broad vs. Compound, and Narrow vs. Compound) by Tukey method were conducted on each parameter (1-9 above) in turn as the dependent variable. If there is a significant interaction effect between proficiency levels and stress patterns, further ANOVAs with pair comparison between stress patterns were carried out separately for each speaker group. Note that prosodic transfer effects will gain expression as significant interaction effects between Stress Patterns and Proficiency levels. On the other hand, simple main effects of Stress pattern with no interaction with Proficiency level indicate a common production strategy across the three subject groups. Main effects of Proficiency levels alone indicate group differences in F0, intensity, or speech rate. Such differences were observed, but are not regarded critical for the present analysis.

2.3.1. Pitch and Intensity
Significant main effects of Stress pattern (all at p<.001), without significant interactions with Proficiency level (highest interaction effect p=.04), were found for all 6 of the F0 and Intensity measures. These main effects of Stress pattern on F0, Intensity, and F0 change and Intensity change are shown in figure 1a (all figures are at the final page of this paper). Pair-wise comparisons between the three stress patterns are shown in Table 1.

Table 1: The significant level (p<0.01, flagged) of pair comparison among the three stress patterns.

<table>
<thead>
<tr>
<th></th>
<th>V1F0</th>
<th>V2F0</th>
<th>V1int.</th>
<th>V2int.</th>
<th>F0 change</th>
<th>Int.change</th>
</tr>
</thead>
<tbody>
<tr>
<td>N – C</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N – B</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>C – B</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

Figure 1a and 1b indicate (and paired comparisons in Table 1 confirm) that narrow focus items were distinguished from broad focus items and compounds by higher peak F0 and intensity on the initial stressed syllable. By contrast, the broad focus items were distinguished from the narrow focus and compounds by higher F0 and intensity on the second stressed syllable.

Pair comparison of pitch change from the first stressed syllable to the second stressed syllable shows significant difference in degree of pitch change among the three patterns: pitch change was greatest in narrow focus pattern, less in compound pattern and least in broad focus pattern. Intensity change was greater in the narrow and compound patterns than the broad pattern (Fig. 1c).

2.3.2. Duration and Pause
2.3.2.1. Duration
The ANOVAs on S1 duration, S2 duration and word/phrase duration showed a significant main effect of Stress patterns(p<0.01). Proficiency levels(p<0.01) and significant interaction Stress patterns x Proficiency levels (S1: p<0.001, S2p<0.02, word/phrase duration: p<0.0001). In order to examine the interaction effect, a series of ANOVAs with planned pair comparison between stress patterns were carried out on each dependent variable separately for each speaker group. The results are reported in Table 2 below and Figure 2.

Table 2: The significant levels (p<0.01, flagged) of pair comparison among the three stress patterns.

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>word/phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td>NATIVE</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Narrow – Compound</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Broad – Compound</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>ADVANCED</td>
<td>Narrow – Compound</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Broad – Compound</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>BEGINNERS</td>
<td>Narrow – Compound</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Narrow – Compound</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Broad – Compound</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Narrow – Broad</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

As shown in Table 2 and Figure 2, there were significant differences in duration among the stress patterns for only the native and advanced speaker group, while beginning Vietnamese speakers of English showed no difference in duration among the three stress patterns. The duration of first stressed syllables in compounds spoken by native speakers was significantly shorter than the same syllables in the two other phrasal constructions. By contrast, advanced speakers produced longer first stressed syllables in the narrow focus pattern than the same syllables in the broad and compound patterns. This lengthening effect can be a by-product of emphatic accentuation of the first syllable of the narrow pattern. Both native and advanced speakers showed significant lengthening of the second stressed syllable of the broad pattern in comparison to the other two patterns, which stem from the accentuation of the phrasal stress. It is shown in figure 3c that compounds produced by both the native and advanced speakers were significantly shorter than their phrasal counterparts. This indicates that advanced speakers made some compression of the compound words even though at a lesser extent than native English speakers.

2.3.2.2. Pauses
It is noted from the analysis of the speech data that while native and advanced speakers made no pause, beginners often paused between the two constituents of the phrases. The number of items with pauses over the total testing items is 0/144 for native
speakers, 0/360 for advanced speakers and 59/360 (16%) for beginners (of which 17 broad focus phrases, 7 compounds and 35 narrow focus phrases). It is shown from the figure 3 that beginning Vietnamese speakers of English tended to pause between components of a phrase than a compound.

2.3.3. Discussion
The foregoing statistical analysis of the F0 and intensity traces suggested that the narrow focus stress pattern is distinguished from broad focus items and compounds by higher peak F0 and intensity on the initial stressed syllable. By contrast, the broad focus items were distinguished from the narrow focus and compounds by higher F0 and intensity on the second stressed syllable. However, when the inverse patterns of prominence on V1 and V2 are jointly considered, e.g.: by expressing F0 as a change of voice pitch over the first and second nuclear constituents of the phrase or compound, it is apparent that a ternary level pitch contrast (N>C>B) may be supported over the conventionally recognized binary contrast (N-C [left-headed] > B [right-headed]); see Fig. 1c. Changes of intensity over the 3 phrase types (Fig. 1c.) parallel changes of voice pitch but have less discriminating power; a finding consistent with classical experiments on the phonetics of English word stress (Fry, 1955).

The results on duration show evidence of a word shortening effect in compounds in comparison with the two phrasal constructions. This is consistent with findings from previous studies (Faure, Hirst, and Chafcouloff, 1980; Farnetani and Cosi, 1988); the syllables in compounds were compressed to conform to the temporal template of a word by the word shortening effect. Nevertheless, this is evidenced in only the native and advanced speakers’ items while beginning Vietnamese speakers of English made no compression of compound words at all.

Generally the pitch and intensity patterns of the two groups of non-native speakers were very similar to those of native speakers, while only advanced speakers could discriminate compounds from phrases by means of a duration contrast. This indicates that the beginning Vietnamese speakers of English could encode the pitch and intensity contrast among the three stress patterns but they have not yet acquired the duration contrast. Their reliance on pitch and intensity and their pause pattern indicate a transfer of properties from their first language.

The results of this study show that the production cues that speakers used are highly consistent with the perception cues that they exploited in the discrimination of the three stress patterns reported in the other paper (Nguyen & Ingram, in review). The perception experiment indicated that native speakers and nonnative speakers differ in their sensitivity to acoustic cues. Native speakers made use of both pitch and duration cues in their perception and production of stress patterns; they relied on contrastive pitch prominence for narrow and broad identification and on duration for compound perception. In contrast, Vietnamese listeners and speakers tended to be sensitive to pitch and intensity (especially on the first stressed syllables of the items) for stress pattern perception and production. Pitch prominence on the first stressed syllable was an indication of the narrow pattern, lower pitch and/or intensity cues on the first stressed syllables indicated the compound pattern, and stronger pitch and intensity on second stressed syllables were cues for the broad pattern. Duration was not used either in perception or production by the non-native beginning speakers. This is evidenced from the lack of compression of compounds in comparison to phrasal construction in their production and their failure to recognize compounds in the perception test. Nevertheless, the advanced speakers’ compression of compounds in their production of the three contrasts indicates that the duration as a cue is learnable.

2.3.4 Discriminant Analysis
In order to assess the efficiency with which F0 and duration characteristics of the speaker productions provide a basis for identifying the three target prosodic types, a Discriminant analysis was undertaken. Two orthogonal dimensions were used to classify the stimuli: a) the pitch change measure (V1f0-V2f0), and b) the normalized duration of word/phrase tokens (expressed as a t-score or modified z-score). The scatter plots are in Figure 4.

The scatter plot of native speaker’s items showed that all but only six of the experimental stimuli could be correctly classified into their Broad, Narrow and Compound groupings on the basis of the F0 change measure and the normalized ‘word’ duration scores (fig. 4a). Many items spoken by the advanced speakers were misclassified by the discrimination function. However, the three stress patterns could be separately–grouped (fig. 4b). The scatter plot 4c showed that items spoken by beginning speakers were unclassifiable into three separate groups. This suggests that even though beginners could make some pitch and intensity contrast, they fail to discriminate the three stress patterns in their production without using the duration cue.

3. Conclusion
The results of this study confirm that native speakers and nonnative speakers differ in their use of acoustic cues which are optimally suited to their respective first language phonologies for identifying the three English stress patterns. This study also shows a consistent relationship between production and perception strategies. Native speakers of English employed a combination of pitch and duration cues in both production and perception of the three stress patterns. Vietnamese listeners showed tonal transfer effects in their production and perception of the English stress patterns in their reliance mainly on pitch and intensity and failure to use duration cue. Nevertheless, the advanced speakers’ ability to compress the compound words to some extent indicates that duration as a cue is indeed learnable.

References
Ingram, J; Nguyen, T. A. T., & Pensalfini, R. An acoustic analysis of compound and phrasal stress patterns in Australian English (being reviewed for the Journal of Phonetics)
Nguyen, T. A. T, and Ingram, J. Native &Vietnamese perception of the three stress patterns (being reviewed for the Journal of Phonetics).


**Figure 1** (above): Comparison of mean V1F0, V2F0, V1 intensity, V2 intensity, F0 change and intensity change among three stress patterns across three speaker groups. Fig 1 a: V1F0 and V2F0, Fig. 1b: V1 and V2 intensity, and Fig. 1c: F0 and intensity change.

**Figure 2**: Comparison of mean S1 duration, S2 duration, and word/phrase duration among three stress patterns and across three speaker groups. Fig 2a: S1 duration, Fig. 2b: S2 duration, and Fig. 2c: word/phrase duration.

**Figure 3**: The pause patterns of Beginners

**Figure 4** (below): scatter plots: Circle: Broad, Square: Narrow, Triangle: Compound