Durational cues to stress and phrasing are preserved post-focally in English

Danfeng Wu[1]

[1]Massachusetts Institute of Technology
dfwu@mit.edu

Abstract

This paper studies two questions in English prosody through an investigation of prosody in post-focal contexts: i) whether phrasal stress has a phonetic basis, and should be distinguished from pitch accent; and ii) whether an intermediate phrase must contain pitch accent. The post-focal contexts are good test grounds for these questions because they are claimed to undergo 'deaccentuation', i.e. they lack pitch accents. This paper shows with preliminary results from a production study that there is durational evidence that indicates that phrasal stress can exist in the absence of pitch accent, and therefore phrasal stress can be realized acoustically, and should be distinguished from pitch accent. Furthermore, intermediate phrase boundaries are preserved post-focally, implying that intermediate phrases do not have to contain pitch accent.

Index Terms: pitch accent, stress, phrasing, duration, post-focal deaccenting, English

1. Introduction

Words following narrow focus in English are known to be deaccented, a phenomenon known as 'post-focal deaccenting' [1], [2]. For example, when the sentence 'Freddie has been thinking over the problem' is uttered in a neutral context with broad focus (e.g. as an answer to the question 'What has happened?'), *Freddie*, *over*, and *problem* tend to carry pitch accents, and the nuclear accent falls on *problem*. When focus falls on the subject *Freddie* (e.g. by making the sentence an answer to the question 'Who has been thinking over the problem?'), *Freddie* has nuclear accent, and there is no pitch accent on any word following Freddie. Typically, this means that F0 falls sharply after *Freddie*, and remains low and level through the rest of the sentence.

A simple analysis of this phenomenon posits that focus must be marked by a nuclear accent, and nuclear accent is the last pitch accent in a phrase, therefore there can be no pitch accent following focus in a phrase. Furthermore, Beckman and Pierrehumbert [3] and Beckman [4] have argued that every intermediate phrase (iP) must contain at least one pitch accent. Then the fact that there is no pitch accent after focus implies that there is no iP phrasing post-focally. In other words, post-focal material should be in the same iP as the focus. This paper presents preliminary experimental results that suggest the contrary, i.e. post-focal phrasing is preserved, as phrase-final lengthening effects are preserved post-focally. These results suggest that the requirement that every iP must contain a pitch accent is a violable constraint rather than an inviolable rule.

In addition, there has been a lot of debate about whether pitch accent is the only way to realize prominence. For example, Bolinger [5] claimed that pitch is the primary cue to stress, and there should not be phrasal stress distinctions without pitch accent. In contrast, Halliday [6], Vanderslice and Ladefoged

[7], and others distinguish pitch accent from stress, and claim that there can be lexically stressed words that do not have a pitch accent in an utterance. Subsequent works have supported the latter view by showing that stress can be identified in the absence of intonational pitch features. For example, Sluijter and van Heuven [8] have shown that in Dutch, spectral balance (the relative energy in different parts of the spectrum) is a reliable acoustic correlate of word-level stress in the absence of pitch accent. However, these works are mostly concerned with word-level stress, or a language different from English, and the question remains of whether phrasal stress can be distinguished from pitch accent in English.

Post-focal contexts in English are a good test ground for this question due to deaccenting. Assuming that pitch accents are removed post-focally, if we can find cues to phrasal stress, then we can say that phrasal stress may exist independent of pitch accent, and that the two must be distinguished. This paper presents preliminary experimental results indicating that durational cues to phrasal stress are preserved post-focally, which suggest that phrasal stress exists in the absence of pitch accent, and post-focal deaccenting should be understood as removal of pitch accent, but not phrasal stress.

2. Syntactic structures with different prosodic patterns

The experiment looks for evidence of phrasing and phrasal stress post-focally by putting syntactic structures that are usually distinguished by accentuation and phrasing patterns in a post-focal context. Then we can see if they continue to be distinguished prosodically.

First, we identified a suitable syntactic distinction. Then we confirmed the prosodic difference associated with this syntactic distinction with suitable durational measures in a neutral context, i.e. under broad focus. Finally, we examined the same syntactic structures in the post-focal context using the same durational measures.

In a neutral context, when a verb is followed by a word that is ambiguous between a particle and a preposition, different syntactic structures result in different prosodies [9], [11]. Take the string *thinking over* as an example:

> *Verb + Particle (V + Part)*   (1)
> Freddie has been thinking over the problem.

> *Verb + Preposition (V + Prep)*   (2)
> Freddie has been thinking over the last hour.

Following are our observations about their prosodies in the region from the verb to the particle / preposition, which are consistent with Price *et al.*'s [9] observations:

- The particle has a pitch accent, whereas the preposition does not.

- The verb tends to carry a pitch accent in V+Prep, but not in V+Part.
- The prosodic boundary between the verb and the particle is smaller than the boundary immediately following the particle, whereas the boundary between the verb and the preposition is greater than the boundary immediately following the preposition. Specifically, there tends to be a phrase accent (L-) immediately after the particle but not after the verb in V+Part, and immediately after the verb but not after the preposition in V+Prep.

We annotate these observations below with ToBI transcription terms, where H* indicates a high pitch accent, L- indicates a low phrase accent, and ')' indicates the stronger boundary in the region from V to P:

Verb + Particle (V + Part) (3)
Freddie has been thinking over the problem.

…	thin	king	o	ver	…
		H*		L-)	

Verb + Preposition (V + Prep) (4)
Freddie has been thinking over the last hour.

…	thin	king	o	ver	…
	H*		L-)		

Since the rime of an accented syllable is longer than the rime of an unaccented syllable [12], the stressed rime of the verb (also the first rime of the verb, therefore labeled *V1*) is longer in V+Prep than in V+Part, and the stressed rime of the following word (also its first rime, labeled *P1*) is longer in V+Part than in V+Prep. All the verbs, prepositions, and particles used in this experiment are disyllabic and trochaic.

In addition, the final rime of a word is lengthened before a boundary, and the bigger this break, the longer the rime [13]. We call the final rime of the verb (also its second rime) *V2*, and the final rime of the following word (also its second rime) *P2*. While V and P may both be followed by a prosodic boundary, the relative strength of the break differs, so the degree of pre-boundary lengthening should differ, too. P2 should be lengthened more than V2 in V+Part, and P2 should be lengthened less than V2 in V+Prep. In other words, the ratio of V2 to P2 is greater in V+Prep than in V+Part. We will compare the ratio of V2 to P2 across syntactic structures instead of V2 and P2 independently because it has been suggested that the strength of a phrase boundary is not marked absolutely, but relative to other phrase boundaries [14], [15].

The research question is whether these prosodic distinctions under broad focus are maintained in a post-focal context. If elements that are lengthened due to accentuation in broad focus are still lengthened post-focally, we expect post-focal V1 to be longer in V+Prep than in V+Part, and post-focal P1 to be longer in V+Part than in V+Prep. This result would suggest that phrasal stress is maintained post-focally.

If phrasing is preserved post-focally, we expect the ratio of V2 to P2 to be greater in V+Prep than in V+Part.

If these prosodic distinctions are preserved post-focally, we can further ask whether such post-focal effects are similar in size compared to those under broad focus. If so, we do not expect a difference between differences in the duration of relevant rimes. On the other hand, if the difference in the duration of relevant rimes is significantly smaller post-focally

than under broad focus, then the prosodic differences in terms of prominence and phrasing between the two syntactic structures are reduced post-focally.

3. Methods

3.1. Materials

The materials consisted of 68 target sentences (2 syntactic structures x 2 focus structures x 17 sets). The two syntactic conditions were V+Part and V+Prep, and the two focus conditions were broad focus and narrow focus. Focus was elicited with a leading wh-question. For example, the following two questions were respectively used to elicit broad focus and narrow focus in the answer:

Question for broad focus: What has happened? (5)
Question for narrow focus: Who has been thinking over the problem?
Answer: Freddie has been thinking over the problem.

When the intended syntactic structure was uncommon or the question was odd out of the blue, a sentence was added before the question to introduce the context and make the conversation natural. The speaker was to say the context (if there is one), the question, and the answer in the given order, to elicit the intended syntactic structure and focus in the target sentence (the answer). There were 59 filler items, some of which contained a context, a question and an answer, and others lacked a context.

Note that we did not intend to study whether V+Part and V+Prep *always* create the prosodies described in (3) and (4), but to ask when they do, whether these prosodic differences are preserved post-focally. Therefore, we only considered sets of items where the speakers' pitch accent placement in broad focus was as described in (3) and (4). We identified the pitch accent placement using ToBI, by listening and inspecting pitch tracks. Some sets failed to elicit such prosody consistently for most speakers, which we thought had to do with the naturalness of the particular context and conversation. We discarded those sets for all the speakers, and ended up with 7 sets in the analysis. The V+P combinations in these 7 sets are: *going over, stopping over, looking over, starting over, thinking over, blowing over*, and *checking over*. P is *over* for all of them because there are very few trochaic words that have the ambiguity we were looking for, i.e. that can be combined with a trochaic verb and create the syntactic ambiguity of V+Part and V+Prep. The other only word that has this property is *under* in *going under*, which was unfortunately eliminated due to its inability to elicit the prosodic patterns described in (3) and (4) consistently.

3.2. Participants

We conducted a production study with five native speakers of North American English. The participants were all graduate students. They were remunerated a small sum for their time, and granted their written consent to being tested. At the end of the experiment, we asked each participant to guess the purpose of the experiment, and no one was able to make a correct guess. We suspect the large number of filler items and the diverse form of items helped conceal the purpose of the experiment to the subjects.

3.3. Data collection

Recording took place in a sound-attenuated booth in the Linguistics Department of MIT. Participants were seated in

front of a computer, which displayed one context-question-answer trio at a time. The stimuli plus fillers were presented in random order, so minimal pairs were not placed next to each other. Participants were given instructions about the task at the beginning of the experiment, and time to read quietly before reading out loud each trio. They were encouraged to act out the dialogues naturally rather than reading mechanically.

3.4. Data analysis

We labeled V1, V2, P1, and P2 manually in Praat, and measured their durations. We relied on cues to segmental boundaries to label the rimes. For example, the criterion for the boundary between a stop and a following vowel (e.g. the beginning of V2, [ɪŋ] in *thinking over*) is the onset of voicing after the release of [k]. We fitted 3 linear mixed effects models via Satterthwaite's degrees of freedom method, with the durations of V1 and P1, and the log of the ratio of the duration of V2 to the duration of P2 (*log(V2/P2)*) as the dependent variable in each model, and syntax (V+Part vs. V+Prep) and focus (broad vs. narrow) as fixed effects. We chose log(V2/P2) instead of V2/P2 because V2/P2 was not normally distributed. The models included random intercepts and slopes by speaker and item where those effects made a significant contribution to model fit [16].

4. Results

Before conducting the analysis of durations, we confirmed that subjects did deaccent post-focally, as was expected.

Contrast Figure 1 and Figure 2, which show the F0 tracings and waveforms from the same speaker. In Figure 2, F0 falls sharply after the focused subject, and declines steadily through the rest of the sentence (the slight perturbations are idiosyncratic properties of the consonants), an indication of deaccentuation, whereas F0 rises and falls throughout the sentence in Figure 1. We examined ten samples of waveforms of narrow-focus sentences, and found the same pattern as Figure 2. We therefore confirmed that subjects did deaccent post-focal material.
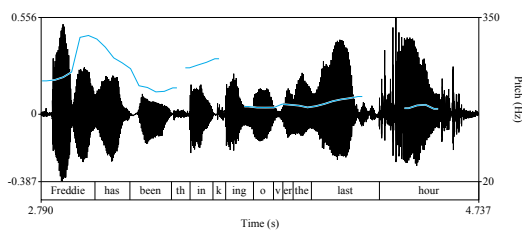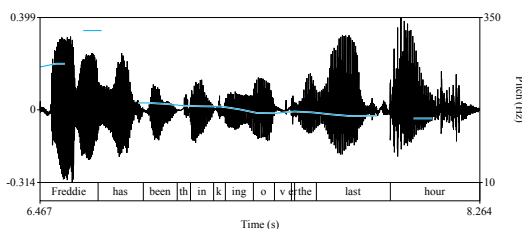

Figure 1: *Waveform of a sentence in broad focus*

Figure 2: *Waveform of a sentence with subject focus*

The durational analysis found that as expected, under broad focus, V1 is on average longer by 21.0 ms in V+Prep than in V+Part ($p<0.05$), as is shown by Figure 3. This mean difference, as well as all the other mean differences described in this section, is the coefficient of the fixed variables in the

model. This means that the verb rimes that carry pitch accent are lengthened, confirming results by Price *et al.* [9].

Post-focally, V1 is on average longer by 20.6 ms in V+Prep than in V+Part ($p<0.05$). This means that post-focal verb rimes that would otherwise be lengthened due to pitch accent under broad focus are still lengthened, indicating that post-focal material can still be prominent, and that prominence can be realized by durational cues.

V1 is not significantly longer in broad-focus sentences than in narrow-focus sentences. We did not find a significant interaction between syntax and focus, indicating that the difference between the two syntactic structures does not differ significantly across focus conditions.
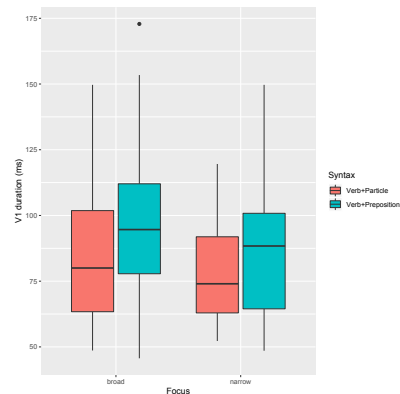

Figure 3: *Duration of the verb's stressed rime in different syntactic structures in broad focus and narrow focus.*

Figure 4 shows that as expected, under broad focus, P1 is on average longer by 24.2 ms in V+Part than in V+Prep ($p<0.01$). This again confirms Price *et al.*'s [9] results, indicating that the stressed rime of the particle (which carries pitch accent) is longer than that of the preposition (which is not pitch accented).

Post-focally, P1 in V+Part is on average longer by 16.6 ms than P1 in V+Prep ($p<0.05$), indicating that the stressed rime of the particle is longer than the stressed rime of the preposition. Therefore, not only are particles more prominent than prepositions post-focally, but this relevant prominence can be realized by longer duration.

P1 is on average 12.8 ms longer in broad-focus sentences than in narrow-focus sentences ($p<0.01$). We found no significant interaction between syntax and focus, again indicating that the difference between the two syntactic structures does not differ significantly across focus conditions.
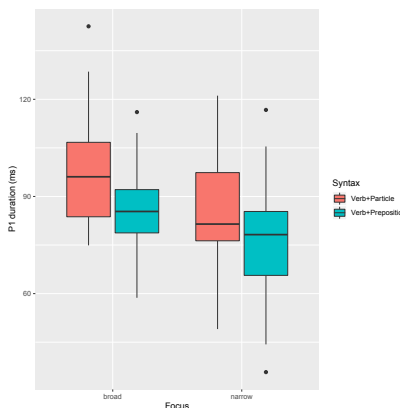

Figure 4: *Duration of P's stressed rime in different syntactic structures in broad focus and narrow focus.*

As Figure 5 shows, under broad focus, log(V2/P2) is significantly greater in V+Prep than in V+Part (p<0.05). This means that in V+Prep, the greater boundary immediately follows V, and in V+Part, the greater boundary immediately follows Part. This result is expected and consistent with Price *et al.* [9] and Norcliffe and Jaeger [11], since pre-boundary rimes are lengthened to different degrees depending on the strength of the boundary.

Post-focally, log(V2/P2) is significantly greater in V+Prep than in V+Part (p<0.05). This suggests that post-focally, the relative degree of phrase boundaries is preserved [11], as is indicated by the different degrees of lengthening effects on the pre-boundary rimes.

We found no significant effect of focus or interaction between syntax and focus. The non-significant interaction indicates that the relative strength of post-focal phrase boundaries is not significantly different from the relative degree of phrase boundaries in broad-focus contexts.
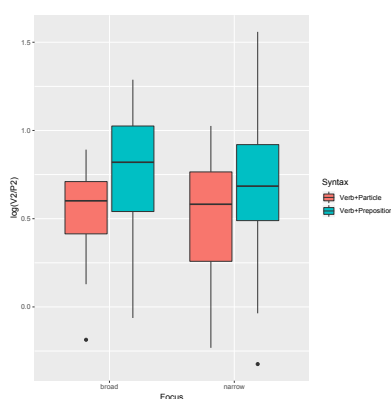

Figure 5: *Log of ratio of verb's second rime duration to P's second rime duration in broad focus and narrow focus.*

5. Discussion

Rimes that carry pitch accent under broad focus are still lengthened post-focally, and the post-focal lengthening effects do not differ significantly from those under broad focus. Since post-focal contexts undergo deaccentuation, the fact that rimes that would otherwise be pitch accented under broad focus are still lengthened suggests that phrasal stress must be present post-focally, and the lengthening effects we observed are due to phrasal stress. This calls for distinguishing between pitch accent and phrasal stress, and implies that phrasal stress can be realized acoustically by duration.

Our observation is consistent with existing theories of stress and pitch accent assignment, such as Pierrehumbert [17]. Pierrehumbert follows Liberman and Prince [18] in positing the metrical grid, which determines the relative strength of the stress on each syllable. Then pitch accents are aligned with syllables on the highest levels of the metrical grid. In V+Part, for instance, the pitch accent falls on the particle and not the verb, an indication that the particle is higher on the metrical grid than the verb. Since the domain that includes the particle and the verb is phrasal, the particle must carry phrasal stress. Likewise, for V+Prep, where the verb has pitch accent and phrasal stress. Post-focal contexts undergo deaccentuation, but the phrasal stress on the particle and the verb in V+Prep remains.

Post-focally, the relative degree of phrase boundaries is preserved, as is indicated by different degrees of lengthening

effects on the pre-boundary rimes. Assume that under broad focus, the stronger phrase boundary between V and P (i.e. the boundary following the particle, and the boundary following the verb in V+Prep, which tend to coincide with a phrase accent) is an iP, and the weaker boundary (i.e. the boundary following the verb in V+Part, and the boundary following the preposition, which do not have a phrase accent) is smaller than an iP (likely a word boundary or clitic-group boundary [9]). Then the fact that the relative boundary strength is preserved post-focally suggests that the stronger phrase boundary is still iP, and the weaker boundary is still smaller than an iP.

Since post-focal contexts undergo deaccentuation, this result suggests that an iP can exist without containing any pitch accent, contra Beckman and Pierrehumbert [1] and Beckman [4]. However, this result is perhaps less surprising under the view that prosodic boundaries tend to align with syntactic boundaries. Since the syntactic structure has not changed in these post-focal contexts, the prosodic boundaries should not alter, if the alignment of prosody to syntax is a higher priority than satisfying the requirement that every iP contains at least one pitch accent. Our results then suggest that the latter requirement may be considered as a violable constraint rather than an inviolable rule. In other words, the requirement that every iP contains at least one pitch accent should be relaxed.

The main effect of focus points to different directions in the three statistical analysis models. While P1 is significantly longer in broad-focus contexts than in narrow-focus contexts, V1 and log(V2/P2) are not significantly different between the two focus conditions. This result is puzzling because while it is possible that word duration is shortened in post-focal contexts, it is unlikely that only some parts (e.g. P1) are shortened, while others (e.g. V1) are not. For this reason, we do not have an interpretation for this effect.

Another question worth studying is the preservation of phrase accents post-focally. There has been a view in the literature that every iP must end with a phrase accent [1]. Our impression from this study was that post-focal contexts lack tones altogether, including phrase accents and pitch accents, as can be seen from Figure 2. If this is true, and intermediate phrasing is nevertheless preserved, this suggests that post-focal iP can lack any phrase accent.

6. Conclusion and future work

This paper reports a production study of English prosody in post-focal contexts. Our observation that post-focal contexts lack pitch accents but still have evidence of phrasal stress supports the view that the two should be distinguished in theory. We also presented results that suggest that iP boundaries are preserved post-focally, and that an iP may exist without containing any pitch accent. In the next stage of the project, we will collect data from more speakers. We will also look for evidence of phrase accents post-focally in order to answer the question of whether every iP must have a phrase accent.

7. Acknowledgements

8. References

[1] K. de Jong, "Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration." *Journal of Phonetics*, vol. 32(4), pp. 493–516, 2004.

[2] Y. Xu and C. X. Xu, "Phonetic realization of focus in English declarative intonation." *Journal of Phonetics*, vol. 33(2), pp. 159–197, 2005.

[3] M. E. Beckman and J. B. Pierrehumbert, "Intonational structure in English and Japanese," *Phonology Yearbook*, vol. 3, pp. 255–309, 1986.

[4] M. E. Beckman, "The parsing of prosody", *Language and Cognitive Processes*, pp. 17–67, 1996.

[5] D. Bolinger, "A theory of pitch accent in English," *Word*, vol. 14, pp. 109–149, 1958.

[6] M. A. K. Halliday, *Intonation and Grammar in British English*, The Hague: Mouton, 1967.

[7] R. Vanderslice and P. Ladefoged, "Binary suprasegmental features and transformational word-accentuation rules," *Language*, vol. 48, pp. 819–838, 1972.

[8] A. M. C. Sluijter and V. J. van Heuven, "Spectral balance as an acoustic correlate of linguistic stress," The Journal of the Acoustical Society of America, vol. 100, pp. 2471–2485, 1996.

[9] P. J. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong, "The use of prosody in syntactic disambiguation," *Journal of the Acoustical Society of America*, vol. 9, pp. 2956–2970, 1991.

[10] M. Wagner, "Prosody as diagonalization of syntax. Evidence from complex predicates," *Proceeding of NELS 34*, 2003.

[11] E. Norcliffe and T. F. Jaeger, "Accent-free prosodic phrases? Accents and phrasing in the post-nuclear domain," *Proceedings of Interspeech*, 2005.

[12] S. Dimitrova and A. Turk, "Patterns of accentual lengthening in English four-syllable words," *Journal of Phonetics*, vol. 40, 403–418, 2012.

[13] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf and P. J. Price, "Segmental durations in the vicinity of prosodic boundaries," *Journal of the American Acoustic Association*, vol. 91, pp. 1707–1717, 1992.

[14] K. Carlson, C. Jr. Clifton, and L. Frazier, "Prosodic boundaries in adjunct attachment," *Journal of Memory and Language*, vol. 45, pp. 58–81, 2001.

[15] M. Wagner, "Prosody as diagonalization of syntax. Evidence from complex predicates," *Proceeding of NELS*, vol. 34, pp. 587–602, 2004.

[16] H. Matushek, R. Kliegl, S. Vasishth, H. Baayen and D. Bates, "Balancing type 1 error and power in linear mixed models," *Journal of Memory and Language*, vol. 94, pp. 305–315, 2017.

[17] J. B. Pierrehumbert, "The phonetics and phonology of English intonation," Ph.D. dissertation, MIT, 1980.

[18] M. Y. Liberman and A. Prince, "On stress and linguistic rhythm," *Linguistic Inquiry*, vol. 8, pp. 249–336, 1977.