This is the author's final proof.
The article appears in Borowsky, Kawahara, Shinya & Sugahara (Eds.). (2012).
Prosody Matters: Essays in Honor of Elisabeth Selkirk (pp. 208-234).
Sheffield, UK: Equinox.

Erratum:
Tables 7.3-7.4, 'F0 mean' at the top row should be 'Duration mean'.

7 Phonetic evidence for prosodic word prominence in American English

Mariko Sugahara[a]

7.1 Introduction

In the theory of Prosodic Phonology (Selkirk, 1978, 1980, 1986; Booij, 1983; Beckman and Pierrehumbert, 1986; Nespor and Vogel, 1986; and others), there are layered phonological constituents that are related to but not necessarily isomorphic with the morpho-syntactic constituent representation. The dominant view is that the prosodic hierarchy in English consists of at least five levels of constituents as shown in (1).

(1) Prosodic Hierarchy
Intonational Phrase (IPh)
Phonological Phrase (PPh)[1]
Phonological Word (PW)
Foot (Ft)
Syllable (σ)

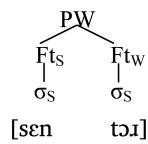
One of the important characteristics of these phonological constituents is that each level of constituents has a head, which is the most prominent constituent of one level below. The head is normally associated with a particular phonological element in English, for instance: the head of σ is associated with a vowel, the head of Ft is associated with a full vowel syllable (Selkirk, 1980; Beckman and Edwards, 1990; Ewen and Van Der Hulst, 2001), the head of

[a] Mariko Sugahara: Department of English, Doshisha University, Imadegawa-Karasuma, Kamigyo-ku, Kyoto, Japan. Email: msugahar@mail.doshisha.ac.jp

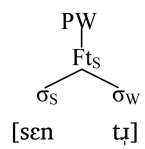
IPh is associated with a nuclear pitch accent (Beckman and Edwards, 1990, and others). Among the elements of the prosodic hierarchy in English, the head of PW, i.e., the main stress Ft, lacks a phonological associate. It is the goal of this paper to investigate whether there is any phonetic correlate of the head of PW.

Consider the following pair of words that demonstrate the relation between prosodic heads and their phonological correlates: *center* [ˈsɛn. ṯɻ] vs *centaur* [ˈsɛn. ˌtɔɹ].[2] They are both disyllabic and trochaic but differ in the quality of their second syllable nuclei. The second syllable nucleus of *centaur* is a full vowel with secondary stress [ɔ], while that of *center* is an unstressed syllabic sonorant [ɻ̩]. Selkirk (1980) and Hayes (1980) propose that the difference between the two words is captured as the difference in their Ft-level constituent structure and prominence representations. The former consists of two single-syllable feet, and therefore both syllables are the heads of separate feet as shown in (2), and the latter consists of a single trochaic foot dominating two syllables where the initial syllable is the head of the foot as shown in (3). In (2), the initial single-syllable foot is the head of PW while the second one is not. The subscripts 's' and 'w' here stand for the head and a non-head of a constituent of one level above respectively. Therefore, 'σ_S' stands for the head of Ft and 'Ft_S' for the head of PW.

(2)



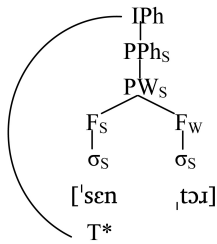
(3)



When a trochaic word like *centaur* that consists of two feet as in (2) is pronounced in isolation, the initial full vowel syllable (primary stress), e.g., *cen-* in *centaur*, is usually acoustically and perceptually more salient than the second full vowel syllable (secondary stress), e.g., *-taur* in *centaur*, partly because it is where a nuclear pitch accent appears (Gussenhoven, 2004: 21).

That the initial foot dominating *cen* in (2) is the head of PW does not necessarily guarantee that the syllable will automatically bear a pitch accent: pitch accents are aligned with the designated terminal element (DTE) of IPh, as shown in (4).[3] PW's in a post-nuclear pitch accent position that are not the head of IPh have no pitch accent on their primary stress syllables (Pierrehumbert, 1980, among others).

(4) T* stands for a pitch accent



Segmental and suprasegmental elements, like full vowels and pitch accents, are not the only correlates of prosodic representation. Other acoustic parameters such as duration also correlate with prosodic structure representation, e.g., pre-boundary lengthening (Beckman and Edwards, 1990; Wightman *et al.*, 1992; among many others), the lengthening of syllables bearing the phrasal prominence (Turk and Sawush, 1997; Turk and White, 1999; Cho and Keating, 2009). Also the formant frequencies of vowels are reported to vary to the extent they do not cross phoneme boundaries as the prominence levels of the vowels change (Erickson, 2002, 2003).

This paper examines whether the head-Foot of PW has any phonetic correlates to distinguish it from a non-head-Foot even in an unaccented environment where the presence or absence of a pitch accent does not distinguish the two. More specific questions are (a) whether the designated terminal element (DTE) of the head-Foot of PW, i.e., a vowel with primary stress, is associated with longer durations than the DTE of a non-head-Foot, i.e., a vowel with secondary stress (Sections 7.3 and 7.4, Experiments I and II), and (b) whether the head-Foot of PW as a whole undergoes lengthening (Section 7.4, Experiment II).

7.2 Previous studies: Primary stress and its acoustic strengthening

The phonetic correlates of English primary stress syllables in an unaccented post-nuclear focus environment have been studied by Campbell and Beckman (1997) and de Jong, (2004), Sluijter *et al.* (1995), Sluijter and van Heuven (1996), Huss (1978), and Okobi (2006). Among other things most of these studies report that primary and secondary stress vowels are acoustically distinguished by their duration and other acoustic properties such as spectral tilt.[4] Campbell and Beckman (1997), however, did not find any reliable durational and spectral tilt differences to distinguish the two levels of stress. One reason for the disagreement in these findings might be that the materials used in these studies were not well-controlled. For example, it is not clear whether primary and secondary stressed vowels were compared or whether primary and unstressed vowels were compared in Huss's (1978) and Okobi's (2006) study.[5] In addition the morphological structure and syllable count were not controlled in the studies of de Jong (2004) and Campbell and Beckman (1997). Sluijter *et al*. (1995) and Sluijter and van Heuven (1996) elicited primary and secondary stress vowels by presenting written texts with bold face letters for the primary stress syllables to their speakers. Such visual aid might have cued speakers to exaggerate acoustic differences between the syllables with and without the bold face letters.

Therefore it is worth revisiting the question whether acoustic differences between primary and secondary stressed syllables in the absence of a pitch accent adopting proper stimuli. Furthermore, none of these previous studies investigate what the domain of PW-prominence strengthening is: is it only the primary stressed syllable or the entire head-Ft of PW?

7.3 Experiment I

This experiment used two-syllable noun-verb pairs, where nouns and verbs are distinguished by the location of primary stress and secondary stress. None of the syllables in those word pairs were pronounced with reduced vowels. We compared (a) F0 values and (b) durations of full vowels with primary stress and those of full vowels with secondary stress in both a nuclear pitch accent environment (accented context) and a post-nuclear environment where no pitch accent was present (unaccented context).

The comparison of F0 values guaranteed that in the post-focus part of an utterance, not only secondary stress syllables but also primary stress syllables were unaccented. As for the duration analyses, primary stress vowels of 150 ms to 160 ms turned out to be longer than secondary stress vowels even in the unaccented environment.

7.3.1 Methods

7.3.1.1 Subjects

Eight paid speakers of American English participated in the experiment: five female and three male students at the University of Massachusetts at Amherst, aged between 18 and 30 years, without any known hearing or speaking disorders.

7.3.1.2 Materials

Three noun-verb disyllabic minimal pairs were recorded, which are shown in (5). Nouns are trochaic, i.e., primary stress followed by secondary stress, while verbs are iambic, i.e., secondary stress followed by primary stress.

(5)

a. *DIgest* vs. *diGEST*[6]
b. *MISprint* vs. *misPRINT*
c. *TRANSplant vs. transPLANT*

The target words were embedded in two different contexts: one was a neutral context where they were interpreted as presentational focus (new information), and the other was a post-focus context where they were interpreted as old information in a post-nuclear pitch accent position. We manipulated the former context so that the target words would always carry a pitch accent on a primary stress syllable, and the latter context so that the target words would be pronounced in a lower flat pitch range without any pitch accent. The contexts are provided in Appendix 7.1.

7.3.1.3 Recording

Recordings took place in a sound proof studio at the University of Massachusetts, Amherst. Scripts were presented to the speakers on a computer monitor, which the speakers read aloud. The speakers controlled the speed of text presentation by pressing the space key when they finished reading one text and move to another. The texts were ordered randomly and at least one

filler text was inserted between the texts containing the target words. Each text was recorded only once in one recording session. Speakers participated in two recording sessions conducted on different days. Therefore, 48 noun-verb pairs in total were recorded (3 pairs × 8 speakers × 2 repetitions). Their speech was directly recorded onto a hard disk (44.1 kHz, 16 bit), using an AKG C420 Cardioid Headset Condenser Microphone.

7.3.1.4 Segmentation and measurement

The vowel periods of the target words were demarcated by the beginning of voicing and the end of F2, and are henceforth called V1 (the initial syllable vowel) and V2 (the final syllable vowel). The durations and the mean F0 values were measured in Praat (Boersma and Weenick, 2009).

Some tokens of V2 in the pairs of *misprint* and *transplant* were nasalized without being followed by nasal constriction intervals, but others had a vowel region followed by separate nasal intervals. Since Repeated Measures Analyses (RMANOVA) is adopted for the comparisons of primary and secondary stress, data from nouns and those from their verbal counterparts obtained from the same speaker in the same recording session were paired with each other. The V2 data from word-pairs in which its members did not agree in the presence or absence of the following independent nasal intervals were all excluded from the analyses (eight pairs from the neutral context and two pairs from the post-focus context), just in case the presence or absence of the nasal region affects the durational outcome of V2. As a result, 40 pairs and 46 pairs were available for the analyses of V2 in the neutral and the post-focus context respectively.

In addition, the extraction of F0 values from some vowels was not possible because of their creakiness. Here, too, we only used data from cases where F0 values were available for both a noun and its verb counterpart in the same word-pair produced by the same speaker in the same recording session. The number of pairs available for the F0 analyses of V1 was 37 in the neutral context and 30 in the post-focus context, and that of V2 was 33 in the neutral context and 31 in the post-focus context.

7.3.2 Results

7.3.2.1 F0 Analyses

In order to confirm our assumption that target words in the neutral context bear an H* nuclear pitch accent (= phrasal prominence) and those in the post-

focus context are unaccented, we first considered F0. If we are on the right track, the F0 of primary stress vowels should be higher than that of secondary stress vowels in the neutral context due to the H* accent. In the post-focus context, however, the F0 of primary and that of secondary stress vowels should be almost equal.

RMANOVA separately compared data for male and female speakers because their intrinsic pitch ranges are substantially different. The results are summarized in Table 7.1 for V1 and Table 7.2 for V2. Since 16 comparisons were carried out, i.e., 2 vowels $\times$ 2 contexts $\times$ 2 analyses (by-speaker and by-word analyses) $\times$ 2 genders, the alpha value was adjusted to $0.05/16 = 0.00313$ (roughly 0.003).

Table 7.1 shows that the mean F0 of V1 in the primary stress (noun) condition was 27 Hz (14%) higher than that of V1 in the secondary stress (verb) condition for female speakers, which was statistically significant at the adjusted alpha level ($\alpha = 0.003$) in both by-word and by-speaker analyses. For male speakers, however, by-word measures were only 6 Hz (5%) higher than by-speaker, which turned out to be marginally significant at $\alpha = 0.05$.

In the post-focus context, primary stress V1 was unexpectedly associated with slightly lower F0 mean than secondary stress V1 for both male speakers (−4 Hz) and female speakers (−5 Hz). This result, however, does not contradict our prediction that F0 of primary stress V1 should be higher than that of secondary stress V1 in the neutral context but not in the post-focus context.

In Table 7.2, we also see that primary stress V2 was higher than secondary stress V2 in the neutral context for both male and speakers (19 Hz, 18%) and female speakers (32 Hz, 19%), and the difference was significant for both the by-word and the by-speaker analyses. In the post-focus context, however, there was no significant difference between primary stress V2 and secondary stress V2.

Table 7.1: Mean F0 values and the results of RMANOVA (by-word, by-speaker): primary stress V1 vs. secondary stress V1.

			F0 mean (StD)	Num of pairs	by-word F (df)	by-speaker F (df)
Male (3 speakers)	Neutral	Primary	134 Hz (28.7)	18	3.54* (1,15)	3.71* (1,15)
		Secondary	128 Hz (39.8)			
	Post-Focus	Primary	111 Hz (21.8)	17	−4.5 (1,14)	−5.8 (1,14)
		Secondary	116 Hz (19.8)			
Female (4 speakers)	Neutral	Primary	218 Hz (28.2)	19	14.7*** (1,16)	14.5*** (1,14)
		Secondary	191 Hz (26.1)			
	Post-Focus	Primary	157 Hz (12.6)	13	−4.9 (1,10)	−5.5 (1,9)
		Secondary	162 Hz (14)			

***p<0.003, **p<0.01, *p<0.05

Table 7.2: Mean F0 values and the results of RMANOVA (by-word, by-speaker): primary stress V2 vs. secondary stress V2.

			F0 mean (StD)	Num of pairs	by-word F (df)	by-speaker F (df)
Male (3 speakers)	Neutral	Primary	123 Hz (16.1)	11	22.9*** (1,8)	21.2*** (1,8)
		Secondary	104 Hz (22.9)			
	Post-Focus	Primary	105 Hz (25.6)	15	0.014 (1,12)	0.006 (1,12)
		Secondary	105 Hz (27.4)			
Female (5 speakers)	Neutral	Primary	197 Hz (38.9)	22	18.7*** (1,19)	14.6*** (1,17)
		Secondary	165 (22.4)			
	Post-Focus	Primary	165 Hz (25.9)	16	–0.26 (1,14)	–0.52 (1,11)
		Secondary	170 Hz (32.7)			

***p<0.003, **p<0.01, *p<0.05

For all of the comparisons above, there was neither a significant interaction between stress types and words nor an interaction between stress types and speakers.

The results confirm that neither secondary nor primary stressed vowels are accented in the post-focus context, i.e., neither carries phrasal prominence. In the following sections, the post-focus context is called the 'unaccented' environment while the neutral context the 'accented' environment.

7.3.2.2 Durational analyses

RMANOVA compared the mean durations of primary stress and those of secondary stress vowels in the accented and the unaccented environment. Since six comparisons were carried out for V1 and V2 respectively, i.e., 3 word pairs $\times$ 2 accent environments, the alpha value was adjusted to 0.05/6 = 0.0083 (roughly 0.008).

V1 durations

As shown in Table 7.3, the significant lengthening effect of primary stress on V1 was observed only in the *digest* pair regardless of the accent environments: the primary stress V1 of *digest* was more than 20 percent longer than its secondary stress V1 counterpart in both the accented and the unaccented environment.

Table 7.3: Mean durations and the results of RMANOVA (by-speaker): P (primary stress V1) vs. S (secondary stress V1).

			F0 mean (StD)	Num of pairs	by-speaker F (df)
Accented	[aɪ] in digest	P	163 ms (19.3)	16	105.4** (1,8)
		S	127 ms (13.4)		
	[ɪ] in misprint	P	71 ms (8.8)	16	1.84 (1,8)
		S	68 ms (15.8)		
	[æ] in transplant	P	119 ms (22.6)	16	1.77 (1,8)
		S	113 ms (19.9)		
Unaccented	[aɪ] in digest	P	159 ms (20.1)	16	45.6** (1,8)
		S	131 ms (13.8)		
	[ɪ] in misprint	P	70 ms (7.2)	16	1.78 (1,8)
		S	66 ms (13.9)		
	[æ] in transplant	P	104 ms (18.9)	16	–3.48 (1,8)
		S	115ms (22.9)		

**$p < 0.008$, *$p < 0.05$

There was also no significant accentual lengthening effect: the accented primary V1 was not significantly longer than its unaccented counterpart.

V2 durations

The mean duration of primary V2 of the *digest* pair [ɛ] was about 14 percent longer than that of secondary V2, and it was statistically significant at the adjusted alpha value ($\alpha = 0.008$) in both the accented and the unaccented environment.

For [æ] in *transpl<u>a</u>nt*, the mean duration of primary V2 was about 17 percent longer than that of secondary V2 in the unaccented environment, which was statistically significant at $\alpha = 0.008$. In the accented environment, the primary stress V2 of *transpl<u>a</u>nt* was 8 percent longer than secondary stress

V2, which was marginally significant at $\alpha = 0.05$. The V2 durations of the pair of *misprint*, however, showed no primary stress lengthening effect regardless of the accentual environments. Table 7.4 shows these results. We also find that there was no accentual lengthening effect: accented primary stress was not significantly longer than unaccented primary stress.

Table 7.4: Mean durations and the results of RMANOVA (by-speaker): P (primary stress V2) vs. S (secondary stress V2).

			F0 mean (StD)	Num of pairs	by-speaker F (df)
Accented	[ɛ] in digest	P	150 ms (24.6)	16	12.3** (1,8)
		S	132 ms (21.1)		
	[ɪ] in misprint	P	99 ms (22.7)	13	0.33 (1,6)
		S	96 ms (24.3)		
	[æ] in transplant	P	155 ms (30.2)	11	7.45* (1,4)
		S	143 ms (27.4)		
Unaccented	[ɛ] in digest	P	148 ms (20.3)	16	27** (1,8)
		S	130 ms (21.6)		
	[ɪ] in misprint	P	98 ms (23.5)	16	0.65 (1,8)
		S	94 ms (20.4)		
	[æ] in transplant	P	151 ms (20)	14	11** (1,6)
		S	129ms (22.3)		

**p <0 .008, *p< 0.05

7.3.2.3 Discussion

Our major finding here is that the lengthening of primary stressed syllables is not ubiquitous. For V1, only the *digest* pair showed a significant difference between the two types of stress in both accentual environments. Similarly for V2, only the *digest* and *transplant* pairs showed significant differences in both

accentual environments. What these cases have in common is that the mean durations of their primary stressed vowels were 150 ms to 160 ms regardless of the accent conditions. In contrast, the mean durations of primary stress vowels that did not undergo lengthening were less than 120 ms (see in Tables 7.3 and 7.4). This suggests that longer syllables may be more likely to show stress-related durational adjustments.

Another finding is that there was no accentual lengthening effect, contra Turk and Sawusch's (1997), Turk and White's (1999) and Cho and Keating's (2009) results. They showed that the mean vowel duration was longer for the accented primary stress than for the unaccented primary stress. It may be due to the difference in the types of accent used in their studies and in the present study. In Turk and Sawusch's (1997), Turk and White's (1999) and Cho and Keating's (2009) studies, accented words were interpreted as contrastive narrow focus (a.k.a. correction focus), whereas in the present study, accented words were interpreted as presentational focus. Selkirk (2002) reports that contrastive narrow focus and presentational focus are associated with different shapes of pitch accents: the former with a bitonal accent L + H* while the latter with a monotonal accent H*. The bitonality of the accented words with narrow focus could have added extra length to the accented words in Turk and Sawusch (1997), Turk and White (1999) and Cho and Keating (2009), which might have resulted in the significant durational difference between accented and unaccented vowels with primary stress.

7.4 Experiment II

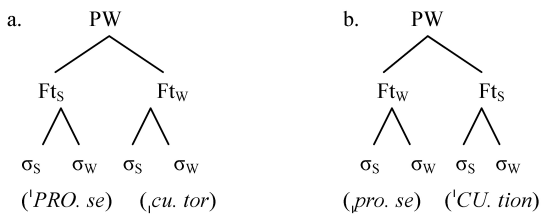
The results of Experiment I do not provide clear answers to the question of what the domain of PW-prominence lengthening is, because in these cases the head-Foot of each PW consisted only of the syllable with primary stress. Experiment II addresses this issue.

In Experiment I the members of each noun-verb pair have the same segmental sequence, so speakers had to rely on the syntactic and semantic information in carrier sentences to detect the grammatical category of each word. In Experiment II we manipulated the stimuli so that the members of each word pair did not share entirely the same segmental content. Here the target words consist of four-syllables, and the two members of each word pair partially share the same segmental sequence up to their penultimate syllables, e.g., *prosecutor* vs. *prosecution*. The position of the primary stress and that of

the secondary stress is also different: the former has primary stress on its initial syllable and secondary stress on its penultimate syllable, while the latter has it the other way round. Because the final syllables help indicate the stress pattern differences, it was not necessary to manipulate the syntactic and the semantic content of the carrier sentences to indicate the location of primary and secondary stress. Therefore, the target words used in this experiment were embedded in syntactically simple carrier sentences as in Appendix 7.2 (e.g., *I said _____, you know*).

Despite the difference in the location of primary and secondary stress, each pair of words has two trochaic feet. The presence of two feet in these word-pairs makes it possible to examine whether the domain of lengthening related to PW-prominence is the primary stressed syllable only or the entire head-Ft of PW. Consider the minimal pair: ˈ*PROse*ˌ*cutor* and ˌ*prose*ˈ*CUtion* in more detail: the initial foot (ˈPRO.se) of ˈ*PROse*ˌ*cutor* is the head of PW as shown in (6) while the initial foot in ˌ*prose*ˈ*CUtion* is not. (The head-Ft of PW is marked with a subscript 's'.)

(6)



If the entire head-Ft of PW lengthens then both the initial stressed syllable and the following unstressed syllable in (6a) should be longer than the parallel constituents in (6b).

7.4.1 Methods

7.4.1.1 Speakers

Five paid native speakers of American English participated in the experiment (two female and three male). The participants were undergraduate exchange students at Doshisha University, aged between 20 and 25 years.[7] They had no known hearing or speaking disorders.

7.4.1.2 Materials

The list of word-pairs used in this experiment is shown in (7).

(7)

(a) ˈ*DOmiˌnating* vs. ˌ*domiˈNAtion*
(b) ˈ*FAsciˌnating* vs. ˌ*fasciˈNAtion*
(c) ˈ*NAviˌgator* vs. ˌ*naviˈGAtion*
(d) ˈ*PROseˌcutor* vs. ˌ*proseˈCUtion*
(e) ˈ*TERmiˌnating* vs. ˌ*termiˈNAtion*

Each of the words in (7) was embedded in two contexts, as in Experiment I i.e. neutral accented context and post-focus unaccented context. See Appendix 7.2 for details. A pitch accent appears at the primary stress location in the accented context while no accent is present in the unaccented context.

7.4.1.3 Recording

Recordings took place in a quiet room at Doshisha University. Speakers' speech was directly recorded to a hard disk (44.1 KHz, 16 bit) using a RØDE NT2-A microphone. Speakers read each aloud four times in two sessions: two repetitions in each recording session on different days. Since there are five speakers, five word pairs, two contexts and four repetitions, the total number of utterances recorded was 400 (200 pairs).

7.4.1.4 Segmentation and measurement

We labeled the vowel period of the first stressed syllable (V1) and that of the second unstressed syllable (V2) following the same labeling procedure as for Experiment I. For the V1 interval of the pair of *terminating* and *termination*, the rhotic part that follows the vowel was also included in the V1 interval, because it was impossible to segment the vocalic and rhotic intervals; instead the whole interval was labeled as /ɝ/. Duration and mean F0 values were automatically extracted by Praat (Boersma and Weenick, 2009).

7.4.2 Results

7.4.2.1 The analyses of V1

F0 analyses

In order to make sure that vowels with primary stress in the neutral context were accented while those in the post-focus context were not, the mean F0

values of each V1 period was considered first. If V1 with primary stress is associated with a higher mean F0 value than that with secondary stress in the neutral context, then we can conclude that V1 with primary stress is accented. If V1 with primary stress and V1 with secondary stress are associated with almost equal mean F0 values in the post-focus context, then we can conclude that V1 with primary stress is unaccented. Since four comparisons (RMANOVA) were carried out (2 genders × 2 contexts), the alpha value was adjusted to 0.05/4 = 0.0125 (roughly 0.01). Table 7.5 shows the results.

The mean F0 values of V1 with primary stress in the neutral context were significantly higher than V1 with secondary stress vowels in the same context. In the post-focus context, however, the mean F0 values were almost the same for both primary stress and secondary stress. This result supports the hypothesis that V1 with primary stress in the neutral context is accented while that in the post-focus context is not. There was no significant interaction between stress patterns and word pairs/speakers.

Table 7.5: Mean F0 values and the results of RMANOVA (by-word, by-speaker): primary stress V1 vs. secondary stress V1.

			F0 mean (StD)	Num of pairs	$F_{\text{by-word}}$ (df)	$F_{\text{by-speaker}}$ (df)
Male (2 speakers)	Neutral	P	124 Hz (18.5)	40	24.6** (1,35)	61.03** (1,37)
		S	116 Hz (25.3)			
	Post-Focus	P	97 Hz (18.5)	40	0.053 (1,35)	0.048 (1,37)
		S	97 Hz (19.1)			
Female (3 speakers)	Neutral	P	236 Hz (22.3)	60	93.6** (1,55)	234.3** (1,57)
		S	207 Hz (13.4)			
	Post-Focus	P	181 Hz (15.9)	60	0.084 (1,55)	0.083 (1,57)
		S	181 Hz (17.1)			

**$p<0.01$, *$p<0.05$

Durational analyses

RMANOVA (by-word and by-subject) tested whether the mean duration of V1 with primary stress was longer than that of V1 with secondary stress both in an accented environment and in an unaccented environment. Since separate comparisons were carried out for the accented and the unaccented environment, the alpha value was adjusted to 0.05/2 = 0.025. Table 7.6 summarizes the results: V1 with primary stress is significantly longer than V1 with secondary stress regardless of the accentual environments in both the by-word and the by-speaker analyses.

Table 7.6: Mean durations and the results of RMANOVA (by-word, by-speaker): P (primary stress V1) vs. S (secondary stress V1).

			Duration mean (StD)	Num of pairs	by-word F (df)	by-speaker F (df)
5 speakers	Accented (Neutral)	P	109ms (23.6)	100	125** (1, 95)	111** (1, 95)
		S	96ms (20)			
	Unaccented (Pos-Focus)	P	103ms (22.1)	100	14.9** (1, 95)	14.5** (1, 95)
		S	98ms (22.8)			

**p<0.01, *p<0.05

There was, however, an interaction between stress patterns and word-pairs according to the by-word analyses: in the accented environment $F(4, 95) = 4.73$, $p = 0.002$; in the unaccented environment $F(4, 95) = 2.55$, $p = 0.02$. Given this, additional RMANOVA (by-speaker) tests were carried out separately for each word pair to examine the durational difference between primary and secondary stress in each word-pair. Since ten comparisons were carried out, i.e., 5 word pairs × 2 accent environments, the alpha value for this statistical test was adjusted to 0.05/10 = 0.005. Results are summarized in Table 7.7.

Primary stress was consistently longer than secondary stress for all word-pairs in the accented environment: the difference was significant at $\alpha = 0.005$ for four pairs. One pair: *terminating* vs. *termination*, was marginally significant at $\alpha = 0.05$. In the unaccented environment, however, this pair: *termi-*

nating vs *termination,* was the only one that showed a significant difference at $\alpha = 0.005$. Two pairs, i.e., *fascinating* vs. *fascination* and *dominating* vs. *domination,* were marginally significant at $\alpha = 0.01$ and $\alpha = 0.05$ respectively, whereas the pair of *prosecuting* vs. *prosecution* showed no significant difference between the two stress conditions. That is, although there was a general tendency for primary stress to be longer than secondary stress regardless of the accentual environments, the tendency did not hold for all word pairs in the unaccented environment.

The mean durations of the primary stress vowels in this experiment were much shorter than 150 ms. Nonetheless, they were longer than secondary stress vowels. In contrast, primary stress vowels that underwent primary stress lengthening in Experiment I were either approximately 150 ms or more than 150 ms. We will come back to this point again in Section 7.4.3.

Table 7.7: Mean durations and the results of RMANOVA (by-speaker): P (primary stress V1) vs. S (secondary stress V1).

Word Pairs	Stress	Accented			Unaccented		
		Duration (Std)	Num of pairs	$F_{by\text{-}speaker}$ (Df)	Duration (Std)	Num of pairs	$F_{by\text{-}speaker}$ (Df)
[ɑ] in *domi-*	P	124 ms (17.6)	20	38.5*** (1,15)	114 ms (11.3)	20	6.29* (1,15)
	S	110 ms (15)			110 ms (12.2)		
[ɑ] in *prose-*	P	102 ms (16.5)	20	13.5*** (1,15)	97 ms (14.3)	20	−1.43 (1,15)
	S	94 ms (12.3)			100 ms (13.9)		
[æ] in *fasci-*	P	118 ms (12.2)	20	25.6*** (1,15)	117 ms (16.9)	20	6.77** (1,15)
	S	105 ms (14)			112 ms (18.1)		
[æ] in *navi-*	P	124 ms (14)	20	56.1*** (1,15)	114 ms (14.2)	20	3.94* (1,15)
	S	103 ms (15.8)			107 ms (13.3)		
[ɝ] in *termi-*	P	75 ms (11.9)	20	4.96* (1,15)	71 ms (12.5)	20	8.71*** (1,15)
	S	68 ms (13.3)			64 ms (14.3)		

***$p < 0.005$, **$p<0.01$, *$p< 0.05$

An additional finding is that there was accentual lengthening of primary stress vowels: accented primary stress vowels (109 ms) was about 6 percent longer than the unaccented primary stress vowels (103 ms). This accentual lengthening effect was statistically significant according to ANOVA: $F(1, 198) = 3.38$, $p = 0.03$.

7.4.2.2 The analysis of unstressed V2

The mean durations of the unstressed V2 were also compared between a post-primary stress position and a post-secondary stress position. As discussed later, there were cases where V2 lacked its independent voicing period, which were treated as zero ms in the analyses here. We carried out RMANOVA (both by-word and by-subject). The alpha value was adjusted to .05/2=.025. Post-primary stress V2 was significantly longer than post-secondary stress V2 in both the accented and the unaccented environment as shown in Table 7.8. Furthermore, there was no interaction between the word-pair factor and the speaker factor, which means that the durational relationship between post-primary stress V2 and post-secondary stress V2 is consistent across all word pairs and all speakers.

Table 7.8: Mean durations and the results of RMANOVA (by-speaker): P (post-primary stress $V2_{Unstressed}$) vs. S (post-secondary stress $V2_{Unstressed}$).

			Duration mean (StD)	Num of pairs	by-word	by-speaker
					F (df)	*F* (df)
5 speakers	Accented (Neutral)	P	40 ms (13.9)	100	6.75** (1, 95)	7.09** (1, 95)
		S	37 ms (14.6)			
	Unaccented (Post-focus)	P	35 ms (17.3)	100	9.59** (1, 95)	9.67** (1, 95)
		S	31 ms (17.9)			

**$p<0.01$, *$p<0.05$

An additional finding is that the unstressed V2 of *fasci-* and that of *prose-* often lacked voicing periods, and the phrasal prominence (pitch accent) and the PW-prominence (primary stress) additively affected the frequency of the V2 devoicing of *prose-*. V2 in the foot lacking PW-prominence (V2 following V1 with secondary stress) was more likely to be devoiced than that in the foot

associated with PW-prominence (V2 following V1 with primary stress) though it was not statistically significant for accented nor for unaccented cases according to Fisher's exact test. V2 in PW lacking the phrasal prominence (V2 following unaccented V1) was more frequently devoiced than that of PW associated with the phrasal prominence (V2 following accented V1), which was marginally significant according Fisher's exact test (+PW-prom: $p<0.1$; -PW-prom: $p<0.1$). That is, V2 in the foot with both the PW-prominence and the phrasal prominence (V2 following V1 with accented primary stress) was the least likely to be devoiced while that in the foot with neither of the prominence (V2 following V1 with secondary stress in an unaccented environment) was the most likely to be devoiced for the *prose-* words. Such a tendency was not obtained for the V2 devoicing of *fasci-*, however. Figure 7.1 shows the frequency of devoicing in percentages, and how it varies depending on the presence or absence of PW-prominence (±PW Prom) and phrasal prominence (±IPh Prom).

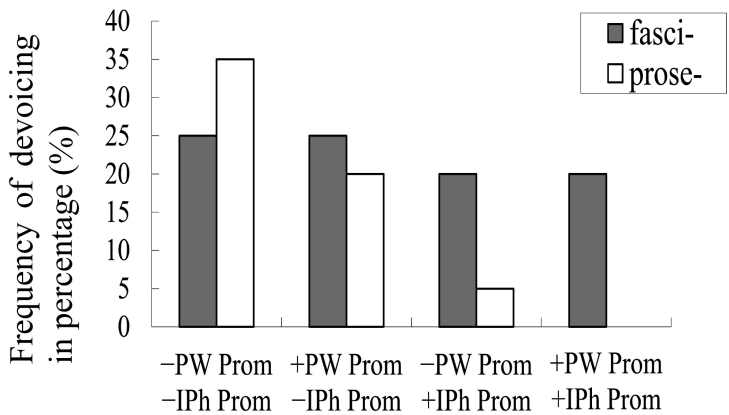


Figure 7.1: Percent V2 devoicing

7.4.3 Discussion

We observe that both the primary stress V1 and the unstressed V2 that follows the primary stress V1 undergo lengthening. The domain of lengthening triggered by the PW-prominence extends to the entire foot.

In this experiment, primary stress vowels of less than 150 ms underwent PW-prominence lengthening. This result differs from that found in Experiment I, where only those of 150 ms to 160 ms underwent PW-prominence in Experiment I. The discrepancies between the two experiments may be due to the difference in the rhythmic patterns of the target words used in these two experiments. Experiment I used disyllabic words in which two stressed syllables were next to each other while in this experiment, words with two feet were used in which an unstressed reduced syllable intervened between the two stressed syllables. Although the primary stress V1 by itself was shorter than 150 ms in this experiment, the foot as a whole undergoes lengthening. The fact that the relatively short primary stress V1 lengthened in this experiment may be a side effect of the lengthening applied to the entire foot that bears the PW-prominence.

Additionally, we found an effect of phrasal prominence (accent) on vowel duration: the accented primary stress V1 was longer than the unaccented primary stress V1. We also see in Tables 7.6 and 7.7 that the unaccented primary stress V1 was longer than not only the secondary stressed V1 in the unaccented environment but also that in the accented environment. This result indicates that both PW prominence and phrasal (IPh) prominence cumulatively contribute to the lengthening of vowels: the DTE of IPh (= accented primary stress vowels) > the DTE of PW (= unaccented primary stress vowels) > the DTE of Ft (= secondary stress vowels in both accented and unaccented environments). This outcome contradicts Cho and Keating's (2009) observation: they do not positively show that primary stress vowels in an unaccented environment were longer than secondary stress vowels. The discrepancy between Cho and Keating's results and ours may be due to the difference in the types of focus/accents involved or the kinds of target words used in the two studies. In Cho and Keating's experiment, target words were interpreted as 'contrastive narrow focus' while those in this experiment were not. They also used nonce words such as *ˈNEbaˌben* vs. *ˌnebaˈBEN*, while we used existing words. It could be that when producing nonce words, speakers can roughly distinguish the acoustic properties of the accented primary stress syllables and the rest of the syllables, but cannot make more subtle distinctions, for example, between the unaccented primary stress and the unaccented secondary stress. When producing real words, however, they can adjust more

subtle durational differences with different levels of prosodic prominence. A possible hypothesis is that familiar and more frequently used words are stored in the mental lexicon together with more fine-grained prosodic and acoustic information. In contrast, no such information is stored for nonce words, and speakers cannot therefore perform as well in distinguishing the different levels of prominence when producing them.

Another finding is that the PW-prominence and the phrasal prominence additively contributed to the frequency of post-stress unstressed V2 devoicing of *prose-* (see Fig. 7.1): the devoicing tended to be the least frequent when the post-stress V2 was dominated by the head-Ft of PW and the head-PW of IPh at the same time, and the most frequent when it was dominated by neither of them though it was not necessarily statistically significant. This outcome may also indicate that the accentual (phrasal-prominence) factor and the PW-prominence factor each contribute to the acoustic property adjustment of vowels independently.

7.5 Conclusions

The present study investigated the effects of PW-prominence on the duration of vowels and the domain of the lengthening associated with PW-prominence.

The results accord well with previous studies such as Huss (1978), Sluijter and van Heuven (1996), de Jong (2004) and Okobi (2006): primary stress vowels were longer than secondary stress vowels not only in an accented context but also in an unaccented context. In the comparison of disyllabic noun-verb pairs where the location of primary stress and that of secondary stress alternated such as *ˈDIˌgest* vs. *ˌdiˈGEST* (Experiment I), only vowels that were longer than 150 ms showed a durational difference between the primary stress and the secondary stress in both the accented and the unaccented contexts. In the comparison of four-syllable words, where the location of primary stress and secondary stress alternated between the initial and the penultimate syllable (Experiment II), e.g., *ˈDOmiˌnating* vs. *ˌdomiˈNAtion*, the initial primary stress vowels were consistently longer than their secondary stress counterparts in both the accented and the unaccented contexts despite the fact that all of the target vowels were less than 150 ms. Apart from the subtle discrepancy between Experiment I and Experiment II in duration of vowels,

results obtained in these two experiments are convincing enough to support the claim that PW-prominence, i.e., the prosodic prominence that is responsible for the assignment of the primary stress, contributes to the durational adjustments of vowels.

Finally, both primary stress vowels and the following unstressed vowels undergo lengthening (Experiment II). From this result, we can conclude that the domain of PW-prominence lengthening is the entire head-Foot of PW. Furthermore, in Experiment II, the PW-prominence and the IPh-prominence additively affect both duration of primary stress vowels and the frequency of devoicing of post-primary unstressed vowels. These results suggest that the PW-prominence is an independent prosodic factor contributing to the acoustic property of an entire foot.

Acknowledgments

I am grateful to Donna Erickson, Shigeto Kawahara and anonymous reviewers for their helpful comments and suggestions to the earlier version of this paper, and to Toni Borowsky and Shigeto Kawahara for editing this paper. All mistakes are mine, however. I would like to also thank Elisabeth Selkirk, Stefanie Shattuck-Hufnagel, and Anthony Okobi for having insightful discussions at an earlier stage of this study, during the summer of 2006. This study is supported by MEXT (Grant No.18720133) to the author.

Notes

1 Beckman and Pierrehumbert (1986) further distinguish the two levels of PPh in English: the Intermediate Phrase and the Accentual Phrase. The Intermediate Phrase in English, according to Beckman and Pierrehumbert, is a domain which is demarcated by a phrase-edge marking tone (a.k.a. phrase accent) and in which the downstep of H* pitch accents takes place. They define the Accentual Phrase as a domain in which one and only one pitch accent is realized, which has been already attested in Japanese. In English, however, there are neither boundary tones nor edge-marking tones to support the level of Accentual Phrase. Beckman and Pierrehumbert's Intermediate Phrase and Accentual Phrase correspond to Selkirk's Major Phrase and Minor Phrase respectively (Selkirk, 1986; Selkirk and Tateishi, 1991).

2 The example pair of *center* and *centaur* was taken from Sugahara and Turk (2009).

3 A designated terminal element of a constituent Ci is a terminal element that is exclusively dominated by strong nodes in Ci (Liberman and Prince, 1977; Selkirk, 1984). Therefore, DTE of IPh is the head (the nucleus vowel) of the head-σ that is dominated by the head-Ft of the head-PW.

4 Spectral tilt (H1-A3) is a measure of H1 (the amplitude of first harmonic) relative to A3 (the amplitude of the third formant (F3)).

5 Huss (1978), for example, used pairs of words such as *DEcrease* (noun) vs. *deCREASE* (verb). However, his noun-verb pairs do not necessarily constitute minimal pairs in terms of primary stress and secondary stress, because the initial syllable of Huss' iambic verbs may be pronounced with a reduced unstressed vowel as in [dəˈkris]. Huss, then compared the strong initial syllable in *DEcrease* (noun) and the weak and possibly unstressed syllable in *deCREASE*. In Okobi's (2006) study, he compared the acoustic properties of the initial primary stress syllables and those of the second syllables in *DIdi* [di.di], *DOdo* [doʊ.doʊ] and *DAda* [dɑ.dɑ], for example. In English, however, unstressed vowels may contrast between [i] and [oʊ] at a word-final position as discussed in Kahn (1976), Flemming (2009: 91) and Flemming and Johnson (2007: 91–93), and it is not clear whether the weaker second syllables in *DIdi* and *DOdo* really carried secondary stress.

6 According to Upton *et al*. (2003) and Wells (2007), the initial syllable of the verb form *diGEST* may be pronounced as reduced. However, in our experiment, none of our speakers produced the word with an initial reduced syllable: they always pronounced it as [daɪ. ʤɛst].

7 The students are from California, the Midwest or the Northeast.

References

Boersma, Paul and Weenink, David. (2009) Praat: doing phonetics by computer (Version 5.1.15) [Computer program]. Retrieved August 30, 2009, from http://www.praat.org/.

Booij, Geert. (1983). Principles and parameters in prosodic phonology. *Linguistics* 21: 249–280.

Beckman, Mary E. and Edwards, Jan. (1990) Lengthening and shortening and the nature of prosodic constituency. In J. Kingston and M. E. Beckman (eds) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* 152–214. Cambridge: Cambridge University Press.

Beckman, Mary E. and Pierrehumbert, Janet. (1986) Intonational structure in Japanese and English. *Phonology Yearbook* 3: 255–309.

Campbell, Nick and Beckman, Mary E. (1997) Stress, prominence, and spectral tilt. *Proceedings of ESCA Workshop on Intonation*, 18–21.

Cho, Taehong and Keating, Patricia. (2009) Effects of initial position versus prominence in English. *Journal of Phonetics* 37 (4): 466–485.

De Jong, Kenneth. (2004) Stress, lexical focus, and segmental focus in English: Patterns of variation in vowel duration. *Journal of Phonetics* 32 (4): 493–516.

Erickson, Donna. (2002) Articulation of extreme formant patterns for emphasized vowels. *Phonetica* 59 (2–3): 134–149.

Erickson, Donna. (2003) Some effects of prosody on articulation in American English. In T. Honma, M. Okazaki, T. Tabata and S. Tanaka (eds) *A New Century of Phonology and Phonological Theory: A Festschrift for Prof. Haraguchi* 473–491. Tokyo: Kaitakusha.

Ewen, Colin J. and Van Der Hulst, Harry. (2001) *The Phonological Structure of Words*. Cambridge: Cambridge University Press.

Flemming, Edward. (2009) The phonetics of schwa vowels. In D. Minkova (ed.) *Phonological Weakness in English* 78–95. Hampshire: Palgrave MacMillan.

Flemming, Edward and Johnson, Stephanie. (2007) Rosa's roses: reduced vowels in American English. *Journal of the International Phonetic Association* 37 (1): 83–96.

Gussenhoven, Carlos. (2004) *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hayes, Bruce. (1980) *A Metrical Theory of Stress Rules*. PhD dissertation, MIT. Published 1985, New York: Garland.

Hayes, Bruce. (1995) *Metrical Stress Theory: Principles and Case Studies*. Chicago, IL: University of Chicago Press.

Huss, Volker. (1978) English word stress in the post-nuclear position. *Phonetica* 35: 86–105.

Kahn, Daniel. (1976) *Syllable-based Generalizations in English Phonology*. PhD dissertation, MIT. Published 1980, New York: Garland.

Liberman, Mark Y. and Prince, Alan. (1977) On stress and linguistic rhythm. *Linguistic Inquiry* 8 (2): 249–336.

Nespor, Marina and Vogel, Irene. (1986) *Prosodic Phonology*. Dordrecht: Foris.

Okobi, Anthony. (2006) *Acoustic Correlates of Word Stress in American English*. PhD dissertation, MIT.

Pierrehumbert, Janet. (1980) *The Phonology and Phonetics of English Intonation*. PhD dissertation, Cambridge, MA: MIT.

Selkirk, Elisabeth O. (1978) On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed.), *Nordic Prosody II* 1–31. Trondheim: TAPIR.

Selkirk, Elisabeth O. (1980) The role of prosodic categories in English word stress. *Linguistic Inquiry* 11 (3): 563–605.

Selkirk, Elisabeth O. (1984) *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.

Selkirk, Elisabeth O. (1986) On derived domains in sentence phonology. *Phonology Yearbook* 3: 371–405.

Selkirk, Elisabeth O. (1995) Sentence prosody: intonation, stress and phrasing. In J. Goldsmith (ed.) *The Handbook of Phonological Theory* 550-569. Oxford: Blackwell.

Selkirk, Elisabeth O. (2002) Contrastive *FOCUS* vs. presentational *focus*: prosodic evidence from right node raising in English. *Proceedings of Speech Prosidy 2002*, 643–646.

Selkirk, Elisabeth O. and Tateishi Koichi. (1991) Syntax and downstep in Japanese. In C. Georgopoulos and R. Ishihara (eds), *Interdisciplinary Approaches to Language* 519–543. Dordrecht: Kluwer Academic Publishing.

Sluijter, Agaath M. C., Shattuck-Hufnagel, Stefanie, Stevens, Kenneth N. and van Heuven, Vincent J. (1995) Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English. *Proceedings of the 13th ICPhS* vol 2, 630–633.

Sluijter, Agaath M. C. and van Heuven, Vincent J. (1996) Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of ICSLP* vol 2, 630–633.

Sugahara, Mariko and Turk, Alice. (2009) Durational correlates of English sublexical constituent structure. *Phonology* 26: 477–524.

Turk, Alice and Sawusch, James R. (1997) The domain of accentual lengthening in American English. *Journal of Phonetics* 25 (1): 25–41.

Turk, Alice and White, Laurence. (1999) Structural influences on accentual lengthening in English. *Journal of Phonetics* 27 (2): 171–206.

Upton, Clive, Kretzschmar, William A. Jr and Konopka, Rafal. (2003) *Oxford Dictionary of Pronunciation for Current English.* Oxford: Oxford University Press.

Wells, J. C. (2007) *Longman Pronunciation Dictionary*. Harlow: Pearson Education Limited.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P. J. (1992) Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91 (3): 1707–1717.

Appendix 7.1. Test materials used in Experiment I

Target words are italicized here for clarity. They were not italicized in the scripts presented to speakers.

1.1. Neutral (Accented) Context

Trochaic Noun Forms

(a) What is Amy doing? Amy is reading the news and *digest*, I think.
(b) What does the paper look like? The paper suffers from typos and *misprints*, I think.
(c) Tell me about your heart. My heart suffers from a cut and *transplant*, I think.

Iambic Verb Forms

(a) What is good about cookies? Cookies are easy to eat and *digest*, I think.
(b) Tell me about these new fonts. The fonts are easy to type and *misprint*, I think.
(c) Please tell me about hearts. Hearts are easy to move and *transplant*, I think.

1.2. Post-Focus (Unaccented) Context

Trochaic Noun Forms

(a) Mary is reading the news and digest. But Amy ISN'T reading the news and *digest*, I think.
(b) Their paper suffers from typos and misprint. But our paper DOESN'T suffer from typos and *misprint*, I think.
(c) His heart suffers from a cut and transplant. But your heart DOESN'T suffer from a cut and *transplant*, I think.

Iambic Verb Forms

(a) Bread is easy to eat and digest. But cookies AREN'T easy to eat and *digest*, I think.
(b) These fonts are easy to type and misprint. But those fonts AREN'T easy to type and *misprint*, I think.
(c) Livers are easy to move and transplant. But hearts AREN'T easy to move and transplant, I think.

Appendix 7.2. Test materials used in Experiment II

Target words are italicized here for clarity. They were not italicised in the scripts presented to speakers.

2.1. Neutral (Accented) Context:

Initial Primary stress

a. I said '*prosecutor*', you know.
b. I said '*navigator*', you know.
c. I said '*fascinating*', you know.
d. I said '*dominating*', you know.
e. I said '*terminating*', you know.

Initial Secondary stress

(a) I said '*prosecution*', you know.
(b) I said '*navigation*', you know.
(c) I said '*fascination*' you know.
(d) I said '*domination*', you know.
(e) I said '*termination*', you know.

2.2. Post-Focus (Unaccented) Context

Initial Primary stress

(a) I didn't say 'prosecutor'. HE said '*prosecutor*', you know.
(b) I didn't say 'navigator'. HE said '*navigator*', you know.
(c) I didn't say 'fascinating'. HE said '*fascinating*', you know.
(d) I didn't say 'dominating'. HE said '*dominating*', you know.
(e) I didn't say 'terminating'. HE said '*terminating*', you know.

Initial Secondary stress

(a) I didn't say 'prosecution'. HE said '*prosecution*', you know.
(b) I didn't say 'navigation'. HE said '*navigation*', you know.
(c) I didn't say 'fascination'. HE said '*fascination*', you know.
(d) I didn't say 'domination'. HE said '*domination*', you know.
(e) I didn't say 'termination'. HE said '*termination*', you know.