SECONDARY STRESS VOWELS IN AMERICAN ENGLISH: THE TARGET UNDERSHOOT OF F1 AND F2 FORMANT VALUES

Mariko Sugahara

Doshisha University
msugahar@mail.doshisha.ac.jp

ABSTRACT

The purpose of this paper is to investigate whether a distinction between primary stress and secondary stress is correlated with F1 and F2 formant frequency differences, even in an environment where the two levels of stress were not distinguished by the presence or absence of pitch accent, i.e., in a post-focus environment. Our investigation of the F1 and F2 values of the first syllable of a *DIgest*(n)-*diGEST*(v) pair and those of an *IMport*(n)-*imPORT*(v) pair revealed that the vowels of unaccented secondary stress syllables underwent phonetic reduction, i.e., the F1 and F2 values of such vowels did not necessarily achieve their target values as those of unaccented primary stress syllables did. Such an observation supports the idea that primary and secondary stress bear different levels of prosodic prominence even in the absence of a phrasal pitch accent.

Keywords: primary stress, secondary stress, target undershoot, F1, F2, noun-verb pairs, post-focus prosody, phonetic vowel reduction

1. INTRODUCTION

One possible way to capture the distinction between primary and secondary stress in English is that the former is the location for a nuclear pitch accent, i.e., phrasal prominence, when words are pronounced in isolation ([2]). There is, however, another view that the two kinds of stress bear different levels of prosodic prominence independent of the presence or absence of a nuclear pitch accent ([6]). According to that view primary stress occupies the strong position of a foot that is the head of a given prosodic word. On the other hand secondary stress occupies the strong position of a less prominent non-head foot. That is to say, the former bears the prominence of a prosodic word level, while the latter bears the prominence of a lower foot level only. There has been increasing evidence for the two kinds of stress bearing such different levels of prosodic prominence independent of the presence or absence of a nuclear pitch accent. For example it has been reported that the two types of stress are durationally distinguished even in an unaccented environment ([3], [5]). They found that in such an environment, the primary stressed syllables were longer than the secondary stressed syllables were. Furthermore, these two types of stress are not only different in their duration but also in their spectral tilt, according to [7] and [5]. They showed that the amplitude of F3 (or F2) relative to that of the first harmonic was greater in unaccented primary stressed syllables than in unaccented secondary stressed syllables. [1] The investigations of such duration and amplitude-related spectral-tilt parameters, however, do not reveal whether or not there is any vowel quality difference between the two levels of stress. Vowel quality is acoustically correlated with first and second formants (F1 and F2), and the main goal of this paper is to investigate how the values of these formants differ between the two levels of stress in a post-focus unaccented region.

Given the fact that secondary stress syllables are shorter in their duration than primary stress syllables even in a post-focus unaccented environment and following the model of target undershoot by [4], we expect formant-target-undershoot, i.e., phonetic vowel reduction, to be more likely to take place at post-focus stressed vowels than at post-focus primary stressed vowels. We examine such a prediction by observing the F1 and F2 transitions of vowels in two-syllable noun-verb pairs, such as in *DIgest-diGEST,* where the location of primary stress and that of secondary stress alternate.

2. THE EXPERIMENT

We carried out a production experiment, and its procedure and results are shown below.

2.1. Experimental procedure

Two noun-verb pairs were used: *DIgest*(n)–*diGEST*(v) and *IMport*(n)–*imPORT*(v). Nouns in these pairs have a trochaic pattern, that is the first syllable bears primary stress while the second syllable bears secondary stress. The other way round is true of the verbs: the first syllable bears secondary stress while the second syllable bears primary stress.

Each word was embedded in texts where they appeared in a post-focus region and were interpreted as already given, as shown in Appendix, so that they would not bear any phrasal accent (i.e. pitch accent).

Three native speakers of American English, D, B and L participated in the experiment. They were all college students at the time of the recording. They read aloud the texts that contained the target nouns and verbs, which were presented at a computer screen, in a random order. Filler texts were also inserted between these texts. Speaker D and Speaker L read each text 8 times and Speaker B 6 times.

The recording took place in a quiet room, using a SONY DAT recorder. The recorded speech was redigitized at the sampling rate of 44.1kHz using Adobe Audition on a PC.

2.2. Measurements

We used Praat and measured the initial, mid and final points of the first formant (F1) and the second formant (F2) of the first vowels of the target words. A vowel interval is defined as the period between the onset and the ending of vowel voicing.

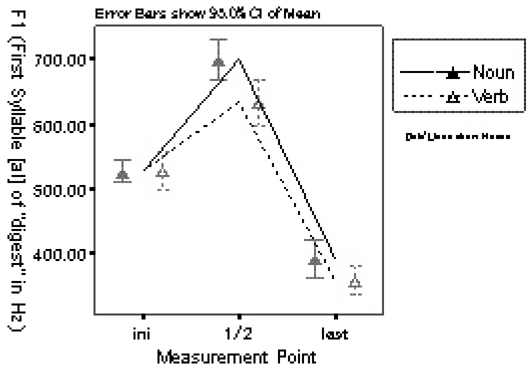
2.3. Results and analyses

Since Speaker L always put an audible H* pitch accent on primary stress syllables even in a post-focus context, her data are not reported in this result section.

2.3.1. The F1 and F2 of DIgest (n) - diGEST (v)

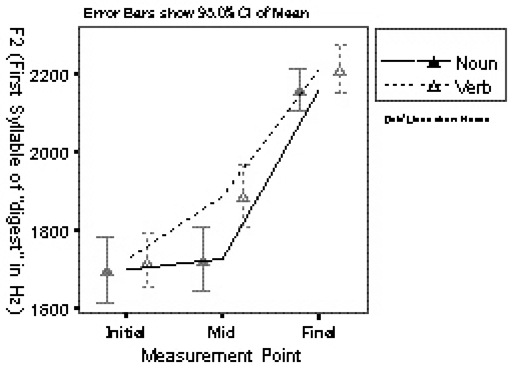
The F1 and F2 values of the initial syllable vowel of the *DIgest*(n)-*diGEST*(v) pair is reported first. Since the initial syllable diphthong [aɪ] requires a shift from a low vowel to a high vowel, we expect F1 to start from a high frequency and end in a low frequency. However, such F1 transition is predicted to undergo target undershoot when the vowel bears secondary stress as in *diGEST*(v). That is, its F1 might move to the lower frequency of the second part vowel of the diphthong, i.e., [ɪ], without fully achieving the target high frequency of the first part vowel [a]. This prediction is on the right track. As shown in Figure 1, the mean mid point frequency of F1 is 633 Hz when the vowel bears secondary stress as in *diGEST*(v), while the mean is 699 Hz when it bears primary stress as in *DIgest*(n). The former is 66Hz lower than the latter. The 95% confidence intervals of these two mean values, i.e., the error bars of Figure 1, do not overlap, and therefore we consider the mean difference to be statistically significant. What this means is that the F1 frequency of the secondary stress vowel starts shifting to a lower frequency already at the mid measurement point, without reaching the higher target frequency of the first part of the diphthong [a].

Figure1: F1 transition from the initial measurement point to the final measurement point of the first vowel [aɪ] in the *DIgest*(n)-*diGEST*(v) pair. The solid line represents the noun and the dotted line the verb. Error bars show the 95% intervals of each mean value.



The diphthong [aɪ] also requires the transition of F2 from a lower to a higher frequency, because of the movement from the more central vowel [a] to the more front vowel [ɪ]. Therefore, we expect the F2 to shift to a higher frequency of [ɪ] without fully achieving the lower target value of [a] when the vowel bears secondary stress. This prediction is also correct as shown in Figure 2. The mean mid point frequency of the initial secondary stress vowel in *diGEST*(v) is 1,888 Hz, about 160 Hz higher than that of the primary stress vowel of the noun form which shows 1,727 Hz. The 95% interval error bars of these two mean values do not overlap, either.

Figure2: F2 transition from the initial measurement point to the final measurement point of the first vowel [aɪ] in the <u>*DI*</u>*gest*(n)-<u>*di*</u>*GEST*(v) pair. The solid line represents the noun and the dotted line the verb. Error bars show the 95% intervals of each mean value.



2.3.2. *The F1 and F2 of IMport (n) – imPORT (v)*

Next consider the F1 and F2 values of the initial syllable vowel of the *IMport*(n)-*imPORT*(v) pair.

The F1 value of the initial vowel [ɪ] is inherently lower than other vowels because of its higher tongue position. The immediately following coda labial nasal [m], however, has even a lower F1 value. Therefore, if target undershoot takes place in the secondary stress syllable of the verb form, we expect the F1 frequency of the secondary stress vowel to be dragged to a further lower frequency range by the following labial nasal. As predicted, the F1 frequency mean at the mid point of the secondary stressed vowel turned out to be about 65 Hz lower than that of the primary stress counterpart. Figure 3 shows that the former is 449 Hz while the latter is 515 Hz. At the same time, their 95% confidence intervals do not significantly overlap.

The F2 transition of the same vowel was also examined. The high vowel [ɪ] intrinsically has a relatively high F2 target frequency, while the immediately following labial nasal has a lower F2 frequency. If the secondary stress vowel undergoes the target undershoot of F2, it is predicted to be in a lower frequency range than the F2 formant of the primary stress counterpart. Figure 4, however, shows that the prediction was not fully born out. Although the outcome was still in the direction of what we expected, i.e., the F2 initial point starts slightly lower in the secondary stress syllable, the confidence intervals of these two kinds of stressed vowels significantly overlap. Therefore, the target undershoot of F2 formant values is not observed here. It is probably due to tongues easily moving back and forth even in a short period of time, and as a result target undershoot may be less likely to take place in F2 formant transitions than in F1 formant transitions.

Figure3: F1 transition from the initial measurement point to the final measurement point of the first vowel [ɪ] in the <u>*IM*</u>*port* (n)-<u>*im*</u>*PORT*(v) pair. The solid line represents the noun and the dotted line the verb. Error bars show the 95% intervals of each mean value.

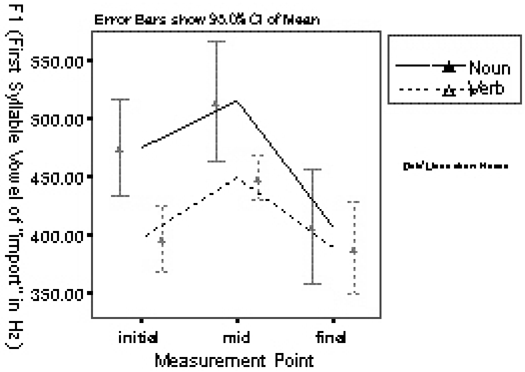
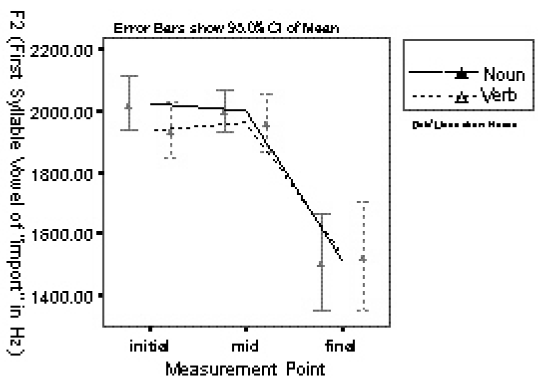


Figure4: F2 transition from the initial measurement point to the final measurement point of the first vowel [ɪ] in the <u>*IM*</u>*port* (n)-<u>*im*</u>*PORT*(v) pair. The solid line represents the noun and the dotted line the verb. Error bars show the 95% intervals of each mean value.



3. CONCLUSION

We investigated whether the distinction between primary stress and secondary stress was correlated with F1 and F2 formant frequency differences, even in an environment where the two levels of stress were not distinguished by the presence or absence of pitch accent. Our investigation of the F1 and F2 values of the first syllable of a *DIgest*(n)-*diGEST*(v) pair and those of an *IMport*(n)-*imPORT*(v) pair revealed that the vowels of unaccented secondary stress syllables were likely to undergo phonetic reduction, i.e., the F1 and F2 values of such vowels did not necessarily achieve their target values as those of unaccented primary stress syllables did. This observation supports the idea that primary and secondary stress bear different levels of prosodic prominence even in the absence of phrasal prominence.

It is still necessary to collect more production data from more speakers and word pairs with other vowels. Also we will examine whether the formant reduction of secondary stressed vowels is audible to listeners and whether they use such a phonetic cue to distinguish primary and secondary stressed syllables in their perception.

4. REFERENCES

[1] Campbel, N., Beckman, M. 1997. Stress, Prominence, and Spectral Tilt. *Proceedings of an ESCA Workshop*, 67-70.
[2] Gussenhoven, C. 2005. *The Phonology of Tone and Intonation.* Cambridge: Cambridge University Press.
[3] Huss, V. 1977. English word stress in post-nuclear position. *Phonetica* 35, 86-105.
[4] Lindblom, B. Spectrographic study of vowel reduction. *JASA* 35, 1773-1781.
[5] Okobi, T. 2006. *Acoustic Correlates of Word Stress in American English*. PhD dissertation, MIT.
[6] Prince, A. & Smolensky, P. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar* (Tech. Rep. Np.2). Rutgers University & University of Colorado.
[7] Sluijter, A. M. C., Shattuck-Hufnagel, S., Stevens, K. N., van Heuven, V. J. 1995. Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English. *Proc. ICPhS* Stockholm 2, 630-633.

APPENDIX

Texts presented to speakers are listed in (a) and (b).

(1) *DIgest* (n) in a post-nuclear position
- (a) Which digests surprised us? The film's <u>digest</u> surprised us.
- (b) Which digest surprised us? The reader's <u>digest</u> surprised us.

(2) *diGEST* (v) in a post-nuclear position
- (c) Who digests steaks very well? The boys <u>digest</u> steaks very well.
- (d) Who digests steaks very well? The eaters <u>digest</u> steaks very well.

(3) *IMport* (n) in a post-nuclear position
- (e) Whose imports go well? Joan's <u>imports</u> go well.
- (f) Whose imports go well? Brenda's <u>imports</u> go well.

(4) *imPORT* (v) in a post-nuclear position
- (g) Who imports goods from France? Joans <u>imports</u> goods from France.
- (h) Who imports goods from France? Higgins <u>imports</u> goods from France.

ACKNOWLEDGEMENTS

This work is supported by the JSPS-Monbu Kagakusho Grant-in-Aid for Young Scientists (B) under Grant No. 18720133 to Mariko Sugahara, Doshisha University.

[1] [1], however, did not find such a spectral tilt difference between the two levels of stress in the absence of pitch accent. It may be due to different types of word pairs used by [1] and by [5] and [4].