Stressed Vowel Duration and Phonemic Length Contrast

Tomasz Ciszewski
University of Gdańsk

Abstract
It has been generally accepted that greater vowel/syllable duration is a reliable correlate of stress and that absolute durational differences between vowels underlie phonemic length contrasts. In this paper we shall demonstrate that duration is not an independent stress correlate, but rather it is derivative of another stress correlate, namely pitch. Phonemic contrast, on the other hand, is qualitative rather than quantitative.

These findings are based on the results of an experiment in which four speakers of SBrE read 162 mono-, di- and trisyllabic target items (made of CV sequences) both in isolation and in carrier phrases. In the stressed syllables all Southern British English vowels and diphthongs were represented and each vowel was placed in 3 consonantal contexts: (a) followed by a voiced obstruent, (b) voiceless obstruent and (c) a sonorant. Then, all vowels (both stressed and unstressed) were extracted from target items and measured with PRAAT.

The results indicate that stressed vowels *may* be longer than unstressed ones. Their durational superiority, however, is not stress-related, but follows mainly from vowel-intrinsic durational characteristics and, to some extent, from the prosodic context (i.e. the number of following unstressed vowels) in which it is placed. In CV_1CV_2 disyllables, when V_1 is phonemically short, the following word-final unstressed vowel is almost always longer. It is only when V_1 is a phonemically long vowel that V_2 may be shorter. As far as diphthongal V_1 is concerned, the durational $V_1{\sim}V_2$ relation is variable. Interestingly, the $V_1{\sim}V_3$ relation in trisyllables follows the same durational pattern. In both types of items the rare cases when a phonemically short V_1 is indeed longer than the word-final vowel involve a stressed vowel which is open, e.g. [æ,ɒ], and whose minimal execution time is longer due to a more extensive jaw movement. These observations imply that both in acoustic and perceptual terms the realisation of word stress is not based on the durational superiority of stressed vowels over unstressed ones. When it is, it is only an epiphenomenon of intrinsic duration of the stressed vowel and extra shortness of non-final unstressed vowel.

As far as phonemic length contrast is concerned, we observe a high degree of durational overlap between phonemically long and short vowels in monosyllabic CVC words (which is enforced by a greater pitch excursion), whereas in polysyllables the differences seem to be perceptually non-salient (>40 ms, cf. Lehiste 1970). This suggests that the differences in vowel duration are not significant enough to underlie phonological length contrasts.

1. Introduction

Vowel duration has been given an enormous amount of research attention, both phonetic and phonological. It has also been generally accepted that duration is one of the major phonetic correlates of stress (cf. Fry 1955, 1958). In this paper we will concentrate on how phonemic length contrasts are curtailed by the operation of pre-fortis clipping (PFC) and the prosodic context (i.e. the number of the following unstressed syllables, or foot structure) in which the stressed vowel is placed. We will argue that PFC and the size of the foot obliterate quantitative vowel contrasts.

2. Experiment design

Four male speakers of Southern British English took part in a controlled experiment. Each subject read 162 target items (54 monosyllables, 54 disyllables and 54 trisyllables). All items were presented in two contexts: in isolation and phrase-finally (*Say the word...*). Target items were selected according to the following criteria: (i) all monosyllables were of the CVC type, (ii) all di- and trisyllables terminated in [i] (incidentally schwa), (iii) in the stressed vowel position all RP vowels an diphthongs were represented, (iv) the post-stress consonants were of three types: voiceless obstruents, voiced obstruents and sonorants (each vowel and diphthong was placed in all three consonantal contexts), (v) where possible, the initial C was a voiced obstruent. Only vowels were measured in the present study. The total number of observations amounts to 652 (162 vowels x 2 contexts x 4 subjects). The significance of the durational differences between stressed vowels in isolated vs. phrase-final context was tested for all vowels in all three groups of target items (mono-, di- and trisyllables) separately. We hypothesised that both isolated and phrase-final pronunciations are in fact identical by virtue of being followed by silence. Thus, if the phrase-final lengthening effects occur (for individual vowels or globally for all vowels within an item in terms of their total duration), they should be observed in both contexts. One-way Anova (with an alpha of .05) confirms that there is no significant effect of the context on both stressed and unstressed vowel duration (p>.05). Thus, the two sets of data were combined which increased the sensitivity of further statistical tests (n=104 for an individual subject in each group of items, i.e. 1-, 2- and 3-syllables).

Vowel duration was measured with PRAAT (Boersma and Weenink 2005) using waveforms and spectrograms. For vowels followed by consonants, vowel onset was identified as the point where the target vowel full formant structure was reached and the end of the vowel corresponded to the beginning of the closure phase. The termination of word-final vowels was assumed to coincide with the end of periodic wave accompanied by dispersion of F2/F3.

3. Vowels duration: a problematic stress correlate

Earlier studies have shown that there exist three acoustic correlates of stress, i.e. *f*0, duration and intensity. According to Fry (1955, 1958), Bolinger (1958) and Morton and

Jassem (1965) the correlates differ in their contribution to stress perception: $f0$ provides the strongest cue, increased duration has a slightly lesser perceptual value and intensity is the weakest correlate. As argued by Lieberman (1960), however, vowel duration is the weakest correlate. A different point of view is presented by Cutler, Dahan and Donselaar (1997: 154) who argue that there is "peculiar redundancy of stress cues in English" and it is also segmental structure that provides robust information about stress.

In essence, the null hypothesis tested in the present study assumes that there exists a fixed $V_{STRESSED} > V_{UNSTRESSED}$ relation that holds for all three phonetic correlates of stress, duration being one of them. Thus, V_1 in polysyllabic items should invariably be longer than the following unstressed vowels (V_2 in 2- and 3-syllable words and V_3 in 3-syllable words). The durational superiority of the stressed vowel over the unstressed ones within a lexical item, however, is not as obvious as it may seem. Admittedly, in trisyllabic words V_1 was found to be generally longer than the following unstressed vowel (V_2). The mean differences between the two vowels for each subject were as follows: S1=61 ms; S2=69 ms; S3=52 ms and S4=77 ms. However, not in all cases was the difference between V_1 and V_2 positive. V_1 did happen to be shorter than V_2 (S1=5.5%; S2=4.6%; S3=14.5% and S4=0.6% of items in the sample). Although such instances were relatively infrequent in each sample, the very fact that they did occur raises doubts about the validity of $V_{STRESSED} > V_{UNSTRESSED}$ relation. We do not think, however, that this provides sufficient arguments for rejecting it. It has to be mentioned that V_2 was longer than V_1 only very specific contexts: (i) when V_1 was followed by a coda consonant (e.g. *density, dignity*) and/or the consonant following V_2 was a stop (e.g. *Kennedy, Canada*). The former context accounts for the extra shortness of V_1 and the latter one for the lengthening of V_2 due to a slightly longer closure phase before the following stop. Furthermore, since the coda consonant is generally assumed to contribute to the phonological weight of the syllable rhyme, its duration should also be taken into account. If added to the pre-coda vowel, the total duration of the CV rhyme would have certainly eliminated all instances in which V_1 alone was shorter than V_2 in trisyllables.

Much stronger doubts concerning the durational domination of the stressed vowels over the unstressed ones appear when V_1 duration in di- and trisyllables is compared with that of word-final unstressed vowels (e.g. *biddy, bigamy*). In disyllables, when V_1 is phonemically short, the following word-final unstressed syllable is almost always longer. It is only when V_1 is a phonemically long vowel that V_2 is shorter. As far as diphthongal V_1 is concerned, the durational $V_1 \sim V_2$ relation is variable. Interestingly, the $V_1 \sim V_3$ relation in trisyllables follows the same durational pattern. In both types of items the rare cases when a phonemically short V_1 is indeed longer than the word-final vowel involve a stressed vowel which is open, e.g. [æ,ɒ] and whose minimal execution time (Klatt 1986) is longer due to a more extensive jaw movement. These observations imply that both in acoustic and perceptual terms the realisation of word stress is not based on the durational superiority of stressed vowels over unstressed ones. When it is, it is only an epiphenomenon of intrinsic duration of the stressed vowel and extra shortness of non-final unstressed vowels, as illustrated in the graphs (1) and (2). Hence, to a large extent it is accidental.
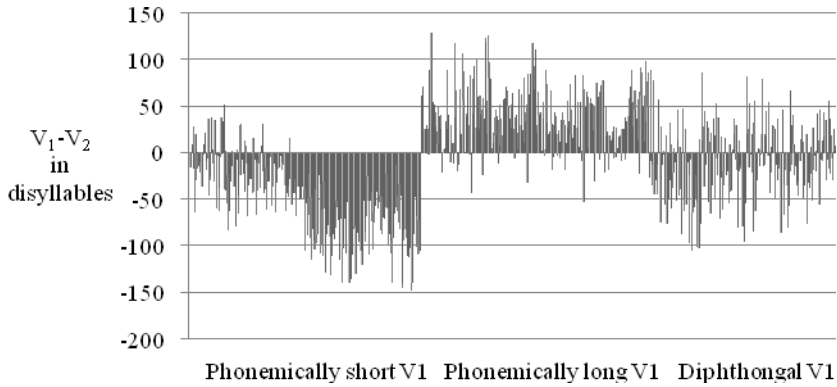
Graph 1: V₁-V₂ difference in duration (ms) in 2-syllable items (all subjects)
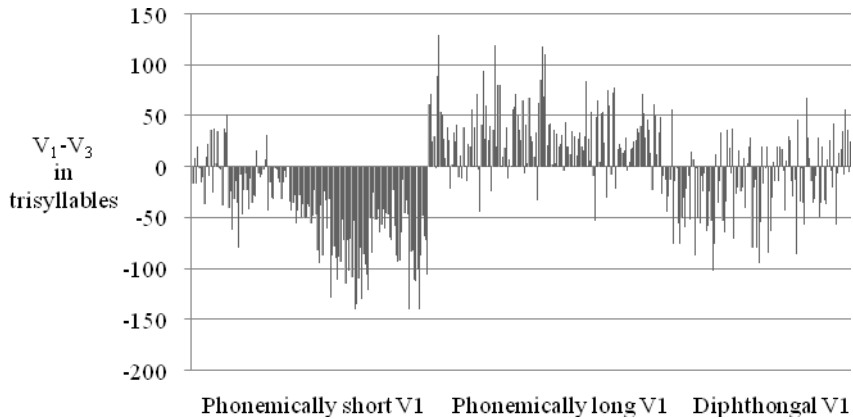
Graph 2: V₁-V₃ difference in duration (ms) in 3-syllable items (all subjects)

In consideration of the above, we have to reject the idea that stressed vowels are longer than the unstressed ones within the same item. In terms of duration the $V_{STRESSED} > V_{UNSTRESSED}$ relation is neither stable nor does it seem to be stress-related.

4. Pre-fortis clipping effects, or how phonemic contrast gets neutralised

In principle the PFC effects should be observed in all vowels which are followed by a voiceless obstruent regardless of the vowel position and the prosodic context. Thus, it should affect stressed and unstressed vowels alike as it is a stress-independent

phenomenon. As observed by Kim and Cole (2005), there exists an inversely proportionate relation between the duration of the stressed syllable and the number of syllables that follow. However, this regularity is also contextually independent of PFC. While on average the duration of the stressed vowels is expected to decrease in longer items, the compression effect may not suspend the operation of PFC. Thus, the mean difference in milliseconds between the duration of stressed vowels followed by a voiceless obstruent and those followed by a voiced one is expected to diminish without threatening the significance of the difference itself.

Thus, according to the null hypothesis, regardless of the durational differences between the stressed vowels in shorter vs. longer items and stressed vs. unstressed vowels, the PFC effects, which are merely related to the voicing of the following consonant, should be constant. If this claim is falsified, i.e. the PFC effects turn out to be insignificant for some group of items or some prosodic context, the conditioning factor must be singled out which is responsible for the PFC suspension. An alternative hypothesis, in our view, must assume that it is caused by the intervocalic durational relations within polysyllabic items. The existence of such interdependences entails a postulation of a higher-level constituent which controls the interactions between the total number of syllables and the degree of stressed vowel shortening before a fortis consonant. We assume that this constituent is the metrical foot.

First, let us consider the durational differences relating to the PFC in the group of monosyllables ending in voiced vs. voiceless obstruent. Rather unsurprisingly, the one-way Anova test (alpha .05) confirms that PFC has a highly significant effect (S1 $p=1.28E-14$; S2 $p=2.59E-08$; S3 $p=.0007$; S4 $p=1.6E-15$) on vowel duration for all subjects regardless of the phonemic length of the vowel.

In disyllables the pre-voiced/pre-voiceless durational difference between stressed vowels (V_1) remains statistically significant, although it has to emphasised that the *p*-values are generally higher and the mean differences are smaller (S1 $p=.0002$; S2 $p=.004$; S3=.0007; S4 $p=.02$).

As far as trisyllabic items are concerned, however, for all subjects the PFC effects on V_1 duration turn out to be non-significant (S1 $p=.02$; S2 $p=.07$; S3 $p=.56$; S4 $p=.23$). Moreover, the mean differences in duration between $V_1+C_{VOICELESS}$ and V_1+C_{VOICED} are further reduced, both generally and for an individual subject. Noteworthy is also the fact that while the mean difference in the duration of pre-voiced vs. pre-voiceless vowels in monosyllables (53.6 ms~113.7 ms) may be safely assumed to be perceptually salient, this is not so obvious in the case of di- and trisyllables, where the difference range is 27~33.9 ms and 20.9~26.6 ms, respectively.

In conclusion, PFC affects stressed vowels to a different degree depending on the number of syllables that follow. Thus, the probability of its occurrence is inversely proportionate to the overall vowel duration of the word.

Let us now pay attention to another surprising fact, namely that phonemically identical vowels followed by a voiceless obstruent are not necessarily shorter than that those followed by a voiced one. The percentage of cases when the vowel in $VC_{VOICELESS}$ is longer than VC_{VOICED} is presented in Table 1 below. For each speaker, the left-hand column shows the number of items where the pre-voiced vowel was actually longer than the phonemically identical pre-voiceless one and the right-hand one the percentage of such occurrences in the data sample (n=36).

	S1		S2		S3		S4	
Monosyllables	0	0%	1	3%	10	28%	0	0%
Disyllables	1	3%	7	19%	13	36%	8	22%
Trisyllables	12	33%	11	30%	12	33%	14	39%

Table 1: Number of instances in which a stressed vowel is longer than a phonemically identical vowel despite the PFC context.

This seems to undermine the very relation between the duration of a vowel and the voicing of the following consonant. This observation does not necessarily falsify PFC. As argued by Kingston and Diehl (1994), PFC is a feature which enhances phonemic contrast and as such it facilitates speech perception. As Gussenhoven (2007: 146) puts it "the implementation of pre-fortis clipping [...] is a concession to the hearer by way of compensation for the frequent devoicing of the voiced obstruent." Thus, this compensation is more likely to occur when the phonemic distinctiveness is threatened. Its degree observed in experimental conditions will then depend upon the organisation of the input. Since in our experiment the order of target items was randomised (i.e. items like *bit* and *bid* were never placed consecutively), there was no (or very little) necessity of contrast enhancement.

Since on the other hand, PFC is aerodynamically conditioned 'because the transglottal pressure difference creating the airflow driving vocal fold vibration is hard to maintain in the face of the impedence by the oral constriction of obstruents' (Gussenhoven: ibid.), its effect on vowel duration is likely to be observed even if distinctiveness is not threatened (e.g. in a randomised experimental input). This does not mean, however, that it *must* occur as the aerodynamic conditioning may be successfully counterbalanced by the prosodic one (which may also be aerodynamic in nature). Pre-fortis clipping, then, is both an articulatorily motivated and speaker-controllable parameter which may be latent (i.e. producing statistically and perceptually insignificant differences in vowel duration) when the vowel contrast is safe.[1] In terms of speech processing, considering the fact that the perceptual information load is directly proportionate to the number of the syllables within an item (cf. the cohort theory by Marlsen-Wilson and Tyler (1980)), in monosyllables the number of instances in which a vowel followed by $C_{VOICELESS}$ is longer than the phonemically identical vowel followed by C_{VOICED} is the lowest.

To sum up, PFC has been shown to have the greatest effect on vowel duration in monosyllabic items. The degree of durational difference between pre-voiced and pre-voiceless vowels in the stressed position is inversely proportionate to the overall length of an item, i.e. the effect is lesser on the stressed vowels in disyllables than on those in monosyllables and it becomes insignificant in trisyllabic items. Pre-fortis clipping appears to be both an articulatorily motivated and speaker-controllable process which may be latent (i.e. producing statistically and perceptually insignificant differences in vowel duration) when the vowel contrast is safe.

---

[1] A typical context for its activation is the presentation of length contrast in minimal pairs (*beat~bead*), e.g. in the process of phonetic instruction.

5. Durational overlap between phonemically long and short vowels

We observed that (i) mean stressed vowel durations systematically decrease as the number of following unstressed syllables increases and (ii) the differences between stressed vowel durations in mono- and disyllables are significantly greater (67-97 ms) than those between di- and trisyllables (15-43 ms).
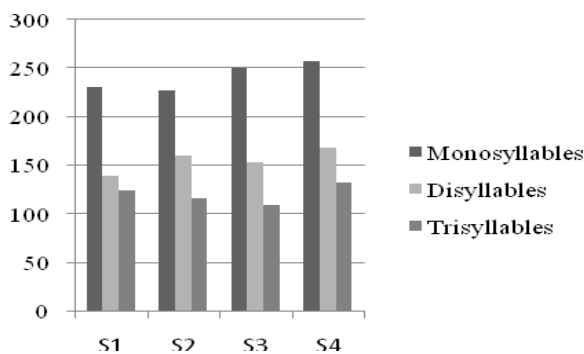

Graph 3: Mean stressed vowel durations (ms) in mono-, di- and trisyllabic items

Theoretically, one would expect that the systematic decrease in V_1 duration in 2- and 3-syllable words should result in a simultaneous obliteration of phonemic length distinctions and, consequently, pose a potential threat to their perception. However, the danger of eliminating phonemic length distinction in polysyllabic items is not as serious as it may seem. Recall that the inter-speaker variation ranges from 18ms to 43.9 ms, which does fit neatly in the non-distinguishable window (10~40 ms) established by Lehiste (1970: 13). The durational deficiency of V_1 in polysyllabic items may also be successfully compensated for by a more robust segmental context. Note that, paradoxically, due to the fact that as the number of the syllables grows, the number of potential vowel-consonant permutations increases rapidly, which reduces the chances of generating, for instance, a trisyllabic minimal pair (whose semantic contrast relies entirely upon V_1 quantity) virtually to zero. Thus, the substantially reduced V_1 recognition time in di- and trisyllables can hardly impede the process of the whole word recognition. Language economy should, therefore, allow to loosen the length contrast requirements where intelligibility is not threatened, i.e. in polysyllabic forms, and strengthen it if the recognition of an item is largely dependent on the recognition of the vowel, i.e. in monosyllables. So much of the theory. What emerges from our data, however, is a completely opposite regularity. It is in the monosyllabic items where the stressed long and short vowels display durational convergence rather than in di- and trisyllables. This conclusion was arrived at by mapping the mean durations of phonemically long and short vowels onto the corresponding standard deviation values. Thus, we have calculated the span of a durational window for the two classes of stressed vowels in 1-, 2- and 3-syllable words by adding the standard deviation for each group to its mean duration on the one hand and subtracting the standard deviation from the

corresponding mean duration on the other. The resulting windows for phonemically short and long vowel durations in each group of items were then compared for each subject with a view to extracting the degree of overlap, which was calculated in the following way: $(V_{MEAN\ DUR.} + V_{STD\ DEV.}) - (V:_{MEAN\ DUR.} - V:_{STD\ DEV.})$. We assumed that there is an inversely proportionate relation between the degree of the durational overlap and the robustness of the phonemic length contrast in a particular group of items.

It turns out that for all subjects the durational overlap was observed only in monosyllabic items (S1=60.2 ms; S2=8.2 ms; S3=48 ms and S4=29.8 ms) and not in di- and trisyllables. This is graphically illustrated in (x) below. Mean duration values are represented by ♦.
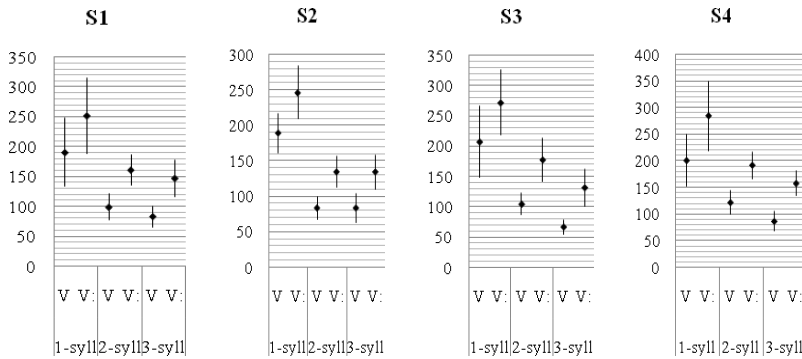
Graph 4: Long/short durational overlap in 1-, 2- and 3-syllable items

Thus, despite the (misleading) fact that the differences in mean durations between long vs. short vowels remain constant for all three groups of target items (cf. the distances between ♦s in each V/V: pair), the durational overlap between long and short vowels in monosyllables indicates that the phonemic contrast is, at least to some extent, suspended in this particular context. Bearing in mind the doubtful perceptual value of the long-short V_1 contrast in polysyllables and a fair amount of durational long-short overlap in monosyllables, we have to conclude that in general the phonemic contrast, at least in the dialect of English investigated in this study, is qualitative rather than quantitative in nature. What follows is that the perception of phonemic length contrast and the production of phonemically conditioned differences in vowel durations may be two different phenomena. While the distinctions do have their articulatory manifestation, their perception (due to the fact that they are below just noticeable difference) are based on quality rather than quantity.

6. Conclusions

The present findings may be summarised as follows. Duration alone is not an independent stress correlate. It is rather a derivative of other correlates (pitch in particular). Stressed vowels may be longer than unstressed ones. Their durational

superiority, however, is not stress-related but follows mainly from vowel-intrinsic durational characteristics. The operation of PFC obliterates the durational contrasts. Phonemic contrast is qualitative rather than quantitative. In monosyllables there is a high degree of durational overlap between phonemically long and short vowels (which is enforced by a greater pitch excursion), whereas in polysyllables the differences do exist but are perceptually non-salient.

References

Bolinger, Dwight. 1958. A theory of pitch accent in English. *Word* 14: 109-149.

Cutler, Anne, Dahan Delphine and van Donsellar, Wilma. 1997. Prosody in the comprehension of spoken language: a literature review. *Language and Speech* 40: 141-202.

Fry, Denis B. 1955. Duration and intensity as acoustic correlates of linguistic stress. *JASA* 27: 765-768.

Fry, Denis B. 1958. Experiments in the perception of stress. *Language and Speech* 1: 126-152.

Gussenhoven, Carlos. 2007. A vowel height split explained. Compensatory listening and Speaker Control. In J. Cole and J. I. Hualde (eds.) *Laboratory Phonology* 9*:* Mouton de Gruyter: 145-172.

Kim, Heejin and Jennifer Cole. 2005. The stress foot as a unit of planned timing: evidence from shortening in the prosodic phrase. *Proceedings of Interspeech* 2005, Lisbon, Portugal: 2365-2368.

Kingston, John and Randy Diehl. 1994. Phonetic knowledge. *Language* 70(3): 419-454.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59: 1208-1221.

Lehiste, Ilse. 1970. *Suprasegmentals.* Cambridge, MA: The MIT press.

Lieberman, Philip. 1960. Some acoustic correlates of word stress in American English. *JASA* 32: 451-454.

Marslen-Wilson. William D. and Loraine K. Tyler. 1980. The temporal structure of spoken language understanding *Cognition* 8: 1-71.

Morton, John and Wiktor Jassem 1965. Acoustic correlates of stress. *Language and Speech* 8*:* 159-181.