

# Quantitative Intervocalic Relations: Evidence for Foot Structure

Tomasz Ciszewski

**Abstract** The paper discusses intriguing differences between the acoustic characteristics of stressed vowels placed in different prosodic contexts. We have analysed the recordings of speakers of SBrE. The experimental design was as follows. Two male native speakers of Southern British English read 162 mono-, di- and trisyllabic words made of CV sequences. The target items were presented in two contexts: (i) isolated and (ii) phrase-final. The stressed syllables contained all RP vowels and diphthongs, followed by a voiced obstruent, a voiceless obstruent and a sonorant. Vowels were then extracted from target words and their duration, pitch and intensity were measured with PRAAT. Pitch and intensity measurements included three parameters: mean/max values and the intravocalic slope. The total number of observations was 9072. The significance of various types of differences was tested with one-way ANOVA and correlation tests. The results suggest that there exist significant discrepancies between the properties of stressed vowels which depend on the distance between the stressed vowel and the end of the word. These differences on the one hand follow from general laws of speech aerodynamics and on the other serve as important cues in word recognition process.

## 1 Introduction

Qualitative interdependencies between non-adjacent vowels within lexical items have been observed in a number of languages. These include vowel harmonic processes (van der Hulst et al. 1995), tone assimilation and spreading (Yip 1995) or formant frequencies co-articulation effects (Öhman Sven 1966).

---

T. Ciszewski (✉)  
University of Gdańsk, Gdańsk, Poland  
e-mail: [angtc@univ.gda.pl](mailto:angtc@univ.gda.pl)

In this paper we concentrate on quantitative relations between vowels within initially stressed, morphologically simple words and the ways in which vowel duration interacts with other correlates of stress, namely pitch and intensity. The observed variability of stressed vowel duration indicates that duration is not a reliable correlate of stress in English, since stressed vowels are not necessarily longer than unstressed vowels within the same lexical item. Their durational superiority is often overridden by stress-independent quantitative processes, like pre-fortis clipping (PFC) and final lengthening (FL). Since PFC is contextually conditioned, its effect should be significant irrespective of the prosodic context in which the stressed vowel is placed. Earlier studies (Ciszewski 2010b) have shown, however, that in initially stressed trisyllables produced in isolation stressed vowels in pre-voiced and pre-voiceless contexts do not differ significantly in their durations. As far as the latter regularity (FL) is concerned, its positional conditioning suggests that the durations of word-final vowels should be independent of the number of the preceding syllables in an item and the distance between the stressed vowel and the final unstressed one. The results do not confirm this prediction; word-final vowels in trisyllables prove to be systematically shorter than those in disyllabic words.

The article is organised as follows. In Sect. 2 the experiment design, including the subjects' profile, the stimulus and the measurement criteria, is presented. In the Results and Discussion part we first analyse the differences between the duration of word-final vowels in di- and trisyllables (Sect. 3.1). In the next section durational correlations between stressed and unstressed vowels are discussed which point at the existence of a superordinate durational template whose function is to equalise the total vowel duration within items having a different number of syllables. In Sect. 3.3 the differences in total vowel duration are analysed. Finally, Sect. 3.4 addresses the question of mutual relations between pitch and duration in stressed and unstressed vowels.

## 2 Experiment Design

Two 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 and 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.

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 Results and Discussion

#### 3.1 *Final Lengthening in di- and Trisyllables*

Final lengthening is a regularity whereby the presence of a major syntactic boundary lengthens the word immediately preceding the boundary (for an extensive discussion see: Fletcher 2010). Since in our experiment all items were placed in a pre-boundary position, no significant durational differences should be observed, both as far as whole words and their component vowels are concerned. All polysyllabic target items end in an open syllable, which –except for a few isolated cases– contains the short vowel [i]. In isolated and phrase-final position this vowel is known to undergo a lengthening process known as ‘happy tensing’ (Fabricius 2002).

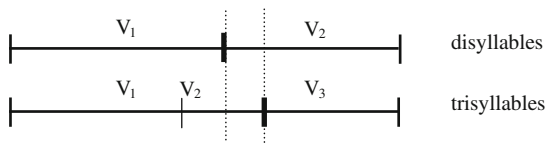
The significance of the difference in duration of word-final vowels in di- and trisyllables was tested for both subjects with one-way Anova (alpha 0.05;  $n = 108$ ). The results point to a highly significant effect of the number of syllables in an item on the duration of the word-final vowel (S1  $p = 2.5E - 40$ ; S2  $p = 2.5E - 44$ ). Thus, FL proves to be sensitive to the overall duration of the word, or more precisely to the distance to the opposite edge of the word. Therefore, the positional motivation of the processes is overridden by some sort of non-local conditioning which controls the degree to which the process affects its target vowels. The differences in final vowel duration in di- and trisyllabic words are illustrated in the graph below (Fig. 1).

As far as the degree of FL is concerned, the only conceivable explanation for the variation observed between 2- and 3-syllable words seems to be the total duration of preceding (non-final) vowel(s). It has been observed that the accumulated duration of  $V_1$  and  $V_2$  in trisyllables is systematically greater than the duration  $V_1$  in disyllables. Simultaneously, final vowels in trisyllables are shorter than final vowels in disyllables. While the former regularity is obvious (given the fact that mean  $V_2$  duration in trisyllables is greater than the mean difference between  $V_1$  duration in di- and trisyllables), the latter one concerning word-final vowels remains entirely accidental unless it is interlinked with the former. Thus, in order to check whether the duration of word-final vowels is related to the duration of non-final vowels, one needs to compare the differences between (i)  $V_1$  duration in disyllables and the accumulated ( $V_1 + V_2$ ) duration in trisyllables on the one hand with (ii) the differences between final vowel durations in the two groups of words on the other. Our assumption is that the two differences should be

**Fig. 1** Duration of final vowels in 2- and 3-syllable items



comparable. This is schematically illustrated below. (The interval corresponding to the differences in question is enclosed between the dashed lines.).

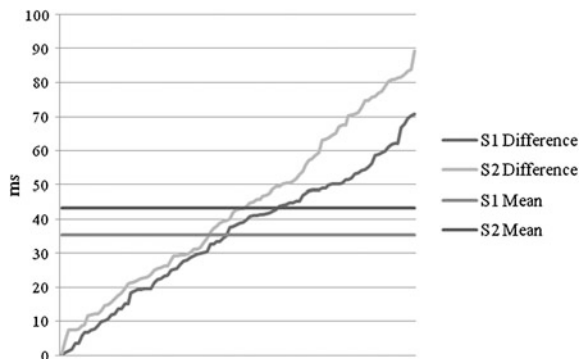


Ideally, the two differences should neutralise each other, i.e. the difference between them should be close to zero. One has to remember, though, that (i) the duration of stressed vowels varies significantly due to the differences in their phonemic length (especially in disyllables) and the influence of PFC and (ii) the total vowel duration in di- and trisyllabic words also displays a certain amount of variation. Thus, it is unlikely that the differences between the durations of non-final vowels and the differences between the durations of final ones in the two groups of items will be identical. However, mean differences between the durations of non-final and final vowels prove negligible.

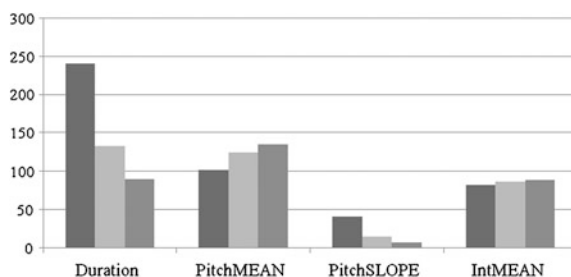
Mean differences: non-final vs. final V (ms)	S1	S2
A. $(V_1 + V_2)^{\text{TRISYLLABLES}} - V_1^{\text{DISYLLABLES}}$	20.9	22.0
B. $V_2^{\text{DISYLLABLES}} - V_3^{\text{TRISYLLABLES}}$	35.3	43.3
Mean A—Mean B	14.4	21.3

Apart from the difference between the mean values (A-B above), we have also calculated the differences between the durations of non-final vowel(s) in di- and trisyllabic items with a phonemically identical  $V_1$  placed in the same PFC context, e.g. *biddy* ~ *bigamy*, and compared them with the difference in final vowel duration for each pair of items. Graph 2 below illustrates the results. (Subject-individual differences between  $(V_1 + V_2)^{\text{TRISYLLABLES}} - V_1^{\text{DISYLLABLES}}$  and  $V_2^{\text{DISYLLABLES}} - V_3^{\text{TRISYLLABLES}}$  have been arranged in ascending order and are labelled

**Fig. 2** Differences between the duration of non-final vowels versus the differences between final vowels in di- and trisyllables



**Fig. 3** Differences in stressed vowel duration, mean pitch, pitch slope and mean intensity in monosyllables (*bid*), disyllables (*bidy*) and trisyllables (*bigamy*) (S2)



as ‘Difference’ for convenience; mean values of these differences (S1 = 35.3 ms and S2 = 43.3 ms) are represented by horizontal lines.)

Surprisingly, despite considerable inter- and intra-speaker variation, the durational differences between  $V_1$  in disyllables and ( $V_1 + V_2$ ) in trisyllables on the one hand, and those between the final vowels in di- and trisyllables on the other are nearly identical for both subjects, which shows that the relation between them is almost perfectly proportionate, i.e. regardless of speaker-individual differences in the absolute duration of final and non-final vowels, the differences between di- and trisyllabic words are constant and amount roughly to 30 ~ 40 ms.<sup>1</sup> This is a much more realistic result than the proposed ‘zero difference’ (note that the difference is a derivative of a number of variables that affect  $V_1$  duration (phonemic length, PFC effects),  $V_2$  duration in trisyllables (weak PFC effect) and the natural intra-speaker variation in final vowel duration. The remaining non-reducible 30 ~ 40 ms difference may then be interpreted as being mechanistically conditioned by insurmountable articulatory requirements on  $V_1$  duration which are imposed by the following consonant and, to a some extent by its intrinsic phonemic length.<sup>2</sup> This, in turn, suggests that not only is the degree of FL conditioned

<sup>1</sup> Which corresponds to approximately 10 % of total vowel duration. This result is supported by a number of analyses (for an overview of studies on just-noticeable durational differences see: Waniek-Klimczak 2005).

by total vowel duration, but that it is also primarily related to the duration of the stressed vowel. Hence, FL and PFC must also be interdependent.

In conclusion, the differences in degree of FL in 2- and 3-syllable words are indeed coupled with the differences in the accumulated duration of the preceding vowel(s). The 'equilibrium' is not perfect, though. It is disturbed by a complex network of durational interrelations between phonemic length of  $V_1$  and the PFC context (which has an effect on both  $V_1$  in disyllables and  $V_2$  in trisyllables) and partly by the natural variation in the duration of final vowels.

### 3.2 *Durational Correlations Between Stressed and Unstressed Vowels*

Bearing in mind the complex combination of factors that influence  $V_1$  duration, the lengthening and tensing of  $V_2$  and the global pre-pausal lengthening of the whole word (which to some extent also affects  $V_1$ ), the possibility of a systematic relation between  $V_1$  and  $V_2$  in disyllables duration appears unlikely. The analysis of correlation, however, provides arguments that cast doubts on the durational independence of the two vowels.

$V_1 \sim V_2$ correlation	S1	S2
Correlation coefficient	-0.14	-0.1
<i>t</i> test (n = 108)	-3.7	-2.74

Moreover, mean  $V_2$  durations differ depending on the phonemic length of  $V_1$ , i.e. when  $V_1$  is a phonemically short vowel, mean  $V_2$  duration is always slightly greater than mean  $V_2$  duration in items with a phonemically long/diphthongal  $V_1$ .

	S1	S2
Mean $V_2$ duration (short $V_1$ )	143	204
Mean $V_2$ duration (long $V_1$ )	141	199

One may argue that despite their statistical significance, the negative  $V_1 \sim V_2$  correlations are rather weak and unconvincing. In our view this objection is unfounded and the result points at more than a chance regularity. It has to be remembered that while  $V_1$  duration is naturally diversified due to phonemic length differences, PFC effects and intrinsic duration,  $V_2$  is phonemically identical for all

<sup>2</sup> Note that the differences between  $V_1$  mean duration in di- and trisyllables correspond roughly to the 30 ~ 40 ms difference (S1 = 42 ms; S2 = 34 ms).

items as is its context (final open syllable). In such circumstances no correlation should be observed. It has to be admitted, though, that the differences in  $V_2$  duration are unlikely to carry any perceptual load (cf. Lehiste 1970).

In trisyllabic items the conceivable set of temporal intervocalic relations is much larger than in disyllables and it includes the following possibilities:

- $V_1 \sim V_2$
- $V_1 \sim V_3$
- $V_2 \sim V_3$
- $(V_1 + V_2) \sim V_3$
- $(V_1 + V_3) \sim V_2$
- $(V_2 + V_3) \sim V_1$

It is only  $V_1 \sim V_2$  interdependence that results in a statistically significant (negative) correlation ( $S1 = -0.11$ ;  $t$  test =  $-2.98$  and  $S2 = -0.25$ ;  $t$  test =  $-7.20$ ). The other durational relations are either statistically insignificant (for one or both subjects), statistically significant but of opposite value (positive vs. negative) or are a mixture of the two possibilities. The statistically significant and negative  $V_1 \sim V_2$  correlation in trisyllables seems to be analogous to that observed between corresponding vowels in disyllables. The crucial difference, however, is that in disyllables  $V_2$  was the word-final vowel, whereas in trisyllables it is the medial one. This may suggest that durational correlations between vowels are local, i.e. they involve vowels only in consecutive syllables. An alternative interpretation of the apparently non-systematic distribution of statistically significant intervocalic correlations in trisyllables is that individual speakers employ different networks of intervocalic durational correlations.

A fundamental problem, however, is why in the first place systematic and unsystematic (speaker-individual) significant correlations are observed. As far as qualitative interrelations between vowels in consecutive syllables (or within the entire word) are concerned, their explanation may be of articulatory nature, i.e. formant frequencies at the beginning of the following vowel are in a way 'inherited' from the formant frequencies observed in the final phase of a preceding vowel (Öhman Sven 1966) and vowel harmonies are related to a particular articulatory setting (nasality, openness/closeness). Generally, intervocalic qualitative relations may be considered 'spreading' or 'co-articulation' phenomena emanating from the stressed vowel. Intervocalic qualitative co-articulations, therefore, involve promoting a particular feature of the stressed vowel onto the unstressed ones within a domain. In terms of quantity, the only remote analogy we can think of is a simultaneous lengthening/shortening of all vowels within a domain connected with faster/slower tempo of delivery or phrase-final lengthening of the whole item. The durational correlations between the stressed and the unstressed vowels should then of necessity be positive. The significant  $V_1 \sim V_2$  correlations observed in our data, however, are negative.

Given our experimental conditions, i.e. steady tempo of stimulus presentation, the fact that the increase in  $V_1$  duration entails  $V_2$  shortening suggests that there exists some pre-programmed durational pattern which controls the duration of both vowels

in disyllabic items. In other words, the duration of one vowel is checked against the duration of the other. This constitutes a serious argument against the ‘no foot’ hypothesis (e.g. Selkirk 1984) since it points to a superordinate temporal unit. Simultaneously, it accounts for why  $V_1$  does not have to be longer than  $V_2$  in disyllables or  $V_3$  in trisyllables. The curtailed duration of  $V_1$ , e.g. when it is phonemically short and additionally affected by PFC, is thus ‘compensated for’ by the increased duration of  $V_2$ . If so, the durations of  $V_1$  and  $V_2$  in disyllabic items are both fine-tuned to fit a durational template whose overall duration, as we will argue in the next section, oscillates around 300 ms (similar results are reported by Kohno 1992).

In the following section additional arguments will be provided which support the assumption that intervocalic durational correlations are indeed superimposed by a higher-order durational template which not only enforces the equalisation of total vowel duration within items of the same number of syllables, but also levels off the durational differences between items having a different number of syllables.

### 3.3 Total Vowel Duration

As far as the degree of variation in total vowel duration within the groups of items with the same number of syllables is concerned, we observe that it is remarkably greater in the group of monosyllables than in polysyllables and in di- and trisyllabic words it is nearly identical. This also holds true for the standard deviation values.

	S1		S2	
	CoV	Std dev	CoV	Std dev.
Monosyllables	19.2	43.4	28.1	72.0
Disyllables	13.2	39.7	11.7	43.1
Trisyllables	13.9	38.3	12.7	43.0

Had there been no tendency to equalise the total vowel duration, an opposite regularity should be observed, i.e. the overall degree of variation in polysyllables, i.e. the summation of all individual vowel variations, should be higher than in monosyllables. Thus, the increase in the number of syllables should result in the increase of variation in total vowel duration, i.e. the more variables, the greater the variation. This, as we see, is not the case.

Another argument which directly supports the equalisation hypothesis is provided by the analysis of variation coefficients for individual vowels and their comparison with those obtained for total vowel duration in a particular group of items. In principle, the mean variation coefficient for the component vowels and for the total duration should be identical. The coefficients of variation (%) for individual vowels and total vowel duration are presented below.



(continued)

Disyllables	V <sub>1</sub>	V <sub>2</sub>		Mean		Total Vowel Duration
Disyllables	V <sub>1</sub>	V <sub>2</sub>		Mean		Total Vowel Duration
S1	12.9	12.1		18.0	>	13.2
S2	24.3	9.1		16.7	>	11.7
Trisyllables	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	Mean		Total Vowel Duration
S1	28.9	22.7	10.9	20.8	>	13.9
S2	30.2	23.3	10.7	21.4	>	12.7

We observe instead that mean variation coefficients for individual vowels in di- and trisyllables are invariably greater than those obtained for total vowel duration and that for both subjects the difference between the mean variation coefficient for  $V_1/V_2/(V_3)$  and total V duration CoV is slightly greater in tri- than in disyllables. Again, the increase in the number of variables (i.e. CoV of particular vowels) that may influence the total vowel duration is counterbalanced by the decrease in the degree of overall variability within each sample. The only motivation for this, somehow paradoxical, regularity seems to be the superimposed pressure on individual vowels to adjust their durations<sup>3</sup> in such a way that their accumulated duration fits a certain durational template. These facts, in our view, do point at a strong tendency towards the equalisation of total vowel duration within each group of items, which, as suggested above, directly refutes the ‘no-foot’ hypothesis. If, however, as we assume, the tendency has neural and aerodynamic foundations, then it should also, if not primarily, manifest itself in the equalisation of total vowel durations in items of different number of syllables.

Admittedly, when analysed in purely statistical terms, the differences in total vowel duration between 1-, 2- and 3-syllable words are significant ( $p < 0.05$ ). However, the differences in mean total duration between di- and trisyllables are remarkably smaller than those between mono- and polysyllables. Interestingly, for both subjects the accumulated duration of vowels in disyllables is greater than that in trisyllables. This stands at variance with an intuition that the increase in the number of syllables within an item must entail the increase of its total vowel duration and indirectly supports the equalisation hypothesis.

Mean total vowel duration (ms)	S1	S2
Monosyllables	227.3	256.8
Disyllables	301.8	369.4
Trisyllables	274.2	338.7

<sup>3</sup> Understandably, the degree of the adjustment is restricted by a number of factors (vowel-intrinsic duration, phonemic length, consonantal context, minimal execution time, etc.).

In the light of previous research on perception of durational differences (e.g. Lehiste 1970), it is evident that the differences of the range observed for di- and trisyllabic items ( $S1 = 27.6$  ms;  $S2 = 30.7$  ms) are well below the level of perceptual significance.<sup>4</sup> For this reason, the actual statistical significance of these differences is of secondary importance. It is rather the relations between particular significances, mapped onto corresponding mean differences, that support the ‘equalisation’ hypothesis. Since inter-speaker variation in total vowel duration, however, does seem to be perceptually salient ( $>60$  ms), isochrony is fundamentally a perceptual phenomenon (cf. Lehiste 1977), which nonetheless has its acoustic foundations. In other words, there exist non-reducible speaker-independent differences in total vowel duration (both within each group of items and between the two groups) which are mainly ‘inherited’ from the differences in stressed vowel duration (phonemic length, PFC effect). These differences, however, are counterbalanced by the variable duration of word-final vowels. In effect, the remaining discrepancy is neutralized perceptually. Thus, given the natural – often stylistically conditioned – variations in tempo in connected speech, the impression of rhythmicality is temporally local, i.e. it is confined to smaller parts of an utterance, e.g. tone units, as argued in Cauldwell (2002), or individual polysyllabic words, as in our experiment) and is then ‘reset’ before the following one begins. In effect, in a longer utterance there may be a few ‘isochronies’, corresponding to different parts of the same utterance. This interpretation is, on the one hand congruent with Cauldwell’s postulate of ‘functional irrhythmicality’ and, on the other, it explains why, despite stylistically conditioned and speaker-individual differences in tempo, the impression of isochrony is almost unanimously reported by listeners and consequently refuted by researchers (e.g. Roach 1982).<sup>5</sup>

### ***3.4 Pitch-Duration Relations in Stressed and Unstressed Vowels***

Apart from intervocalic quantitative relations we also analysed intravocalic pitch-duration interdependence. Although it may, at first sight, seem digressive, this interdependence is crucial for the interpretation of both the intervocalic durational correlations and the divergence of total vowel duration in di- and trisyllabic items.

---

<sup>4</sup> On the other hand, mean vowel duration in monosyllables is far greater than the minimal execution time (cf. Klatt 1976). This suggests that vowels in isolated and phrase final monosyllables undergo an extra lengthening, whose aim arguably is to accommodate a complex contour tone (Ciszewski, in prep.).

<sup>5</sup> Our study also shows that vowel-only approach to rhythm provides much less diversified results than the approach based on interstress interval. In our opinion, the latter has been wrong in assuming that both vowels and consonants contribute to the production/perception of rhythm (for more arguments see: Ciszewski 2010a).

Previous studies have shown that vowel duration (in acoustic and perceptual terms) is negatively correlated with mean  $f_0$ , i.e. low-tone vowels are longer than high-tone ones (Gandour 1977) and that dynamic tones require greater vowel duration (Gordon 2001; Zhang 2001; Yu 2002). A number of possible interpretations for this phenomenon are discussed in Ohala (1973). These include: (i) dynamogenetic theory by Taylor (1933), (ii) air pressure increase behind vowel constriction (Mohr 1971), (iii) vocal tract and vocal cords acoustic coupling (Atkinson and James 1972) and (iv) mechanical tongue-larynx interaction leading to vertical tensing of vocal folds (Ladefoged 1964). However interesting these studies may be, they all concentrate on duration-pitch interrelations in unreduced (hence stressed) vowels. Our results fully confirm earlier findings, but only as far as  $V_1$  is concerned. Unstressed vowels seem to be subject to different laws of speech mechanics and aerodynamics.

In all three groups of target items (mono-, di- and trisyllables) a statistically significant negative correlation has been confirmed:

$V_1$ pitch-duration correlations	S1		S2	
	r	<i>t</i> test	r	<i>t</i> test
Monosyllables	-0.57	-5.97	-0.48	-20.46
Disyllables	-0.17	-4.60	-0.22	-7.59
Trisyllables	-0.49	-17.29	-0.18	-6.15

Since a systematic decrease in stressed vowel duration was independently observed as the number of the following unstressed syllables increases which is accompanied by a simultaneous increase in  $V_1$  mean pitch, the negative pitch-duration correlation must, by inference, be much stronger if calculated for all groups of items collectively. This is indeed the case (S1:  $r = -0.67$ ,  $t$  test = -96; S2:  $r = -0.65$ ,  $t$  test = -89). Whichever theory is assumed to explain this regularity, its cross-linguistic validity does imply a mechanistic/aerodynamic motivation (which assumption the theories discussed by Ohala (1973) seem to share).

Any purely mechanistic or aerodynamic explanation, however, should cater simultaneously for stressed and unstressed vowels. What transpires from our data is that pitch and duration of unstressed vowels (both word-medial and word-final) are also coupled but the correlation is positive, rather than negative, as it was observed for stressed vowels.

$V_1$ pitch-duration correlations	S1		S2	
	r	<i>t</i> test	r	<i>t</i> test
Disyllables $V_2$ (final)	0.17	4.55	0.09	2.92
Trisyllables $V_2$	0.16	4.29	0.08	2.52
Trisyllables $V_3$ (final)	0.28	8.13	0.43	16.87

Thus, either the explanation is not mechanistic at all (which is unlikely) or some important factor(s?) has/have not been taken into consideration in earlier research.

### 3.5 Interpretation of Results

As we emphasised at the beginning of this section, the comparison of pitch-duration correlations in stressed and unstressed vowels is not at all digressive and it is directly related to the systematic intervocalic quantitative (negative)  $V_1 \sim V_2$  correlation and the equalisation of total vowel duration. Only when these types of observations are ‘mapped’ onto each other, can the –somewhat paradoxical– difference in pitch-duration correlation between stressed and unstressed vowels be fully understood.

Our explanation is as follows. The onset of the stressed vowel coincides with a sudden increase of subglottal pressure (cf. Ladefoged 1967: 46) produced by appropriate muscular constrictions, which in turn results in the increase of short time average volume velocity. Due to Bernoulli’s effect the relation between the velocity of air passing through the glottis and time is inversely proportionate. Thus, the longer the time between the outburst of acoustic energy (which is a function of volume velocity and subglottal pressure) and the complete occlusion caused by the following consonant, the greater the decrease in velocity, and consequently, the average pitch of the stressed vowel (= negative pitch-duration correlation in stressed syllables). This mechanism is also responsible for  $V_1$  pitch increase in di- and trisyllables (note that  $V_1$  duration is inversely proportionate to the number of the following unstressed syllables) and explains why  $V_1$  pitch slope is significantly greater in monosyllables than in polysyllables (the greater distance between the initial energy outburst and consonant occlusion does not allow to maintain stable air velocity; the steady decrease in velocity results in the simultaneous decrease of  $f_0$ ) (Fig. 3).

Unstressed vowels, on the other hand, are not accompanied by the increase of subglottal pressure. They ‘inherit’ their energy from the preceding vowel ( $V_1 > V_2 > V_3$ ).<sup>6</sup> Since subglottal pressure does not increase, air velocity is relatively stable. Depending on the amount of energy which remains after the articulation of the stressed vowel (note that it may vary due to the differences in phonemic length and/or the operation of PFC), the following unstressed vowels may also vary in duration. Thus, if  $V_1$  is shorter, it has higher pitch; this in turn elevates  $V_2$  ( $V_3$ ) pitch proportionately. For the same reason,  $V_2$  is proportionately longer (= negative durational  $V_1 \sim V_2$  correlation). This is reflected in the positive pitch-duration correlation in unstressed syllables.

---

<sup>6</sup> Similarly to the terminal frequencies of formants in VCV utterances which have been shown to depend on the entire vowel context (Öhman 1966).

Although both the quantitative intervocalic relations and intravocalic pitch-duration relations may be interpreted mechanically and aerodynamically, they may also perform important communicative functions. In particular we hypothesise that it is the acoustic characteristics of  $V_1$  that facilitates word processing and recognition. For instance, relatively low pitch, accompanied by substantial pitch slope and greater duration signals the proximity of a word boundary. In contrast, higher but level pitch together with reduced duration and greater intensity signal a more remote word boundary.

	Duration	Pitch (mean)	Pitch slope	Perceptual message
$V_1$ <i>bid</i>	241 ms	102 Hz	41 Hz	THE END
$V_1$ <i>biddy</i>	133 ms	125 Hz	15 Hz	1 SYLL. TO THE END
$V_1$ <i>bigamy</i>	91 ms	135 Hz	7 Hz	2 SYLL. TO THE END

Thus, the relations and regularities which first of all are mechanistically/aerodynamically conditioned have an independent perceptual value.

## 4 Conclusions

The results obtain in our experiment suggest that intervocalic relations in morphologically simple, initially stressed lexical items are not only qualitative, but also quantitative. It has been found that the durations of stressed and post-stress vowels are bound by statistically significant negative correlation. This indicates that the vowels in question 'negotiate' their durations, i.e. the increase in the duration of  $V_1$  entails the decrease in  $V_2$  duration. The fact that the quantitative relation is negative may be adequately explained only if we assume that it is controlled by a superimposed durational template. Thus, the durational adjustments aim at equalising the total vowel duration not only within items having the same number of syllables (but differing in  $V_1$  duration) but also in di- and tri-syllabic words. The equalisation is additionally supported by statistically significant differences in final vowel duration, which suggests that the degree of final lengthening is also controlled by the same durational template. We hypothesise that this template corresponds to stress foot. The vowel-only approach, which was assumed in this study, provides strong arguments in favour of the isochrony hypothesis which has been consequently refuted for interstress intervals which comprise consonants as well. The unit of pre-programmed timing which emerges from our data corresponds to approximately 300 ms.<sup>7</sup>

<sup>7</sup> Admittedly, this figure may be dependent on speech rate and an analysis of spontaneous speech is likely to reveal a much greater durational variability.

Apart from the quantitative intervocalic relations we also observed intriguing intravocalic correlations between duration and pitch. The analysis confirms earlier findings which point at a cross-linguistically valid negative correlation between the two correlates of stress. This correlation, however, holds only for stressed vowels. Unstressed vowels display an opposite pitch-duration correlation.

Both the quantitative intervocalic correlations and the variable pitch-duration correlations in stressed and unstressed vowels are, in our view, mechanistically and aerodynamically conditioned. Their overall acoustic effect, i.e. a combination of  $V_1$  duration, equalised total vowel duration,  $V_1$  pitch increase in longer items and the remarkably greater pitch and intensity slope in monosyllables, however, are perceptually informative. They aid word processing and lexical access (as suggested by McAllister 1991, van Donselaar, Koster and Cutler 2005, among others) and serve as prosodic boundary distance markers.

Further research on perception of these phenomena is required to verify our findings.

## References

- Atkinson, James E. 1972. Intrinsic F0 in vowels: Physiological correlates. Paper presented at Fall ASA meeting, Miami Beach
- Boersma, Paul and David Weenink. 2005. Praat: doing phonetics by computer (version 5.0.32). <http://www.praat.org/> (13 August 2008)
- Cauldwell, Richard. 2002. The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythm. *Applied Language Studies*, Apples 2/1: 1–24
- Ciszewski, Tomasz. 2010a. Theoretical modelling of word-level isochrony in English. In *Issues in Accents of English 2: Variability and Norm*: 261–275. Cambridge Scholars Publishing
- Ciszewski, Tomasz. 2010b. Is metrical foot a phonetic object? In *Research in Language*, Waniek-Klimczak, E. & L. Shockey (eds.), Łódź University Press: 115–135
- Ciszewski, Tomasz (in prep.) *The anatomy of the English metrical foot: acoustics, perception and structure*
- Fabricius, Anne. 2002. Weak vowels in modern RP: An acoustic study of happyY-tensing and KIT/schwa shift. *Language Variation and Change* 14:211–237. Cambridge University Press
- Fletcher, Janet. 2010. The prosody of Speech: Timing and Rhythm. In Hardcastle, W.J., Laver, F. and F. E. Gibbon (eds.), *The Handbook of Phonetic Sciences*. (Second edition). Blackwell Publishing Ltd
- Gandour, Jackson. 1977. On the interaction between tone and vowel length: Evidence from Thai dialects. *Phonetica* 3:54–65
- Gordon, Matthew 2001. A typology of contour tone restriction. *Studies in Language* 25(3):423–462
- 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
- Kohno, Morio. 1992. Two mechanisms of processing sound sentences. In Tohkura, Y., Vatikiotis-Bateson, E. and Y. Sagisaka (eds.), *Speech Perception, Production and Linguistic Structure*: 287–29. Amsterdam, Washington, Oxford: IOS Press
- Ladefoged, Peter. 1964. Some possibilities in speech synthesis. *Language and Speech* 7:205–14
- Ladefoged, Peter. 1967. *Three areas of experimental phonetics*. Oxford University Press
- Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: The MIT Press
- Lehiste, Ilse. 1977. Isochrony reconsidered. *Journal of Phonetics* 5: 193–212

- McAllister, Jan. 1991. The Processing of Lexically Stressed Syllables in Read and Spontaneous Speech. *Language and Speech January 1*, 34:1–26
- Mohr, Buurckhard. 1971. Intrinsic variations in the speech signal. *Phonetica* 23: 65–93.
- Ohala, John. 1973. Explanations for the intrinsic pitch of vowels. Unpublished manuscript
- Öhman Sven E. G. 1966. Coarticulation in VCV utterances: Spectrographic measures. *Journal of the Acoustical Society of America*, 39: 151-168.
- Roach, Peter. 1982. On the Distinction between “Stressed Timed” and “Syllable Timed” Languages. In Crystal, D. (ed.): *Linguistic Controversies. Essays in Linguistic Theory and Practice*. In Honour of F. R. Palmer: 73–79.
- Selkirk, Elisabeth. 1984. *Phonology and Syntax. The relation between sound and structure*. Cambridge, MA: MIT Press
- Taylor, Herbert. C. 1933. The fundamental pitch of English vowels. *Journal of Experimental Psychology* 16: 565–582.
- van der Hulst, Harry and Jeroen van de Weijer. 1995. Vowel Harmony. In Goldsmith, J.A. (ed.): *Handbook of Phonological Theory*, Cambridge, MA: Blackwell Publishers
- van Donselaar, Wilma, Koster, Mariëtte & Cutler, Anne. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology*, 58A(2): 251–273
- Waniek-Klimczak, Ewa. 2005. *Temporal parameters in second language speech*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego
- Yip, Moira. 1995. Tone in East Asian Languages. In Goldsmith, J.A. (ed.): *Handbook of Phonological Theory*, Cambridge, MA: Blackwell Publishers
- Zhang, Jiajie. 2001. *The effects of duration and sonority on contour tone distribution-typological survey and formal analysis*. PhD dissertation. Department of Linguistics. University of California, Los Angeles