 DOI 10.2478/v10015-010-0001-x

IS METRICAL FOOT A PHONETIC OBJECT?

TOMASZ CISZEWSKI
angtc@univ.gda.pl
University of Gdańsk

Abstract
The assumption behind this pilot study is that metrical feet are not 'groups of syllables' or 'interstress intervals' but rather 'groups of vowels' extracted from the phonetic material contained between two stresses. We analysed the duration, pitch, intensity and acoustic energy of all vowels in isolated pronunciations of 72 initially stressed items (mono-, di- and trisyllables). The results reveal that pre-fortis clipping of the stressed vowel and final lengthening are interrelated, which suggests that stressed and unstressed final vowels are able to 'negotiate' their durations. Such 'communication' between the stressed vowels and the final unstressed ones is possible only if a mediating constituent (the foot) is postulated. Most importantly, we found no significant differences ($p < .05$) between the total acoustic energy and the total vowel duration in words having a different number of syllables, which supports the assumption of foot-level isochrony in English. It was also observed that the significant increase in vowel duration in stressed CVC monosyllables co-occurs with a significantly greater pitch slope, which we interpret to be a tonally driven implementation of minimal foot binarity requirement.

Key words: metrical foot, isochrony, duration, pitch, intensity, acoustic energy

1. Introduction

Although the concept of the metrical foot has been widely used in phonological literature, little attention has been devoted to formulating a properly constrained formal definition that could be empirically testable at the same time. Most of the definitions available so far (Abercrombie 1967, Hayes 1995: 40, Giegerich 1992: 181) refer to the traditional notion of the syllable (which, however, is in itself highly disputable, both phonologically and phonetically) and the idea of an interstress interval. It seems likely that the persistent definitional problem has its sources in the original poetic usage of the term, whereby the foot is used as a cover term for all sorts of interstress intervals within a rhythmical piece of poetry. Rhythmicity in poetry, however, is arrived at rather artificially and results from a conscious artistic manipulation of lexical and syntactic structure, subordinated to the intended semantic result. As such, it is quite different from a (potential) rhythmicity of naturally produced language. Thus, it should not be taken for granted that the poetic and the linguistic stress foot are identical in terms of size, internal structure and the acoustic characteristics of their components.

However, most phonological approaches seem to tacitly rely on the assumption that foot heads (i.e. stressed elements, be it whole syllables, syllable rhymes or nuclei, depending on the framework) share the same, formal and/or acoustic properties and, what seems even less empirically grounded, that the universal foot template is binary (Hayes 1995: 71). It was the latter assumption in particular that has led to a lot of *ad hoc* theorising, i.e. extrametricality rules in Metrical Phonology (Hayes 1995, among others), the ‘superfoot’ (Selkirk 1980, Harris 1994, among others), whose aim was to cater for ternary stress patterns through binary footing. The dogmatic insistence on binarity in Optimality Theory and Government Phonology, on the other hand, results in disregard for the data that do not support universal foot binarity.

As observed by de Lacy (2007), the strong theoretical predictions are usually based on empirically poor and impressionistic data. Unfortunately, this also refers to English. Despite abundant literature on the formal aspects of the English foot structure, its acoustic properties remain largely unexplored. In this pilot study we analyse the acoustic properties of both stressed and unstressed vowels as well as the relations that hold between them. In particular, our aim is to establish a set of relations between different acoustic properties of consecutive vowels which may indicate that the foot is a real phonetic object.

2. Pre-theoretical assumptions and hypotheses

Given the fairly complex consonantal phonotactics of English and cross-linguistic insensitivity of stress to syllable onsets (see Gordon 2005 for an extensive discussion), combined with a variable cross-linguistic sensitivity of stress to the coda consonant (its presence and/or type), we decided to rely entirely on the only undisputable building blocks of putative metrical feet, i.e. vowels/nuclei. Therefore, all test items are composed of ...CV... sequences, where all the consonants belong to the onset.

(1)	a. 1-syll: (CV)C	1-syll: (CV:)C
	b. 2-syll: (CV)(CV)	2-syll: (CV:)(CV)
	c. 3-syll: (CV)(CV)(CV)	3-syll: (CV:)(CV)(CV)

The theoretical assumption behind such selection of test items is that metrical feet are not interstress intervals (where all phonetic material between two consecutive stresses contributes to the overall foot duration) but rather groups of vowels ‘extracted’ from an interstress interval. In terms of phonological representations metrical feet are thus assumed to be constructed on a separate level of nuclear projection.[1]

```
(2)     Foot  =  { V     V     V }
                   |     |     |
                 C V  C  V  C  V
```

While we do not in principle rule out the possibility of post-lexical pedification, for the purposes of the present study only morphologically simple items were selected. This follows from an assumption that non-derived forms provide a direct insight into the ‘canonical’ or ‘templatic’ structure of metrical feet, while in suffixed forms and on the

[1] Cf. Gussmann (2002: 215).

phrase-level pedifications may be influenced by morphology and syntax. Thus, true metrical regularities, if existent, must be first sought in the lexicon itself. A consequence of this assumption, which requires a separate analysis, is that in morphologically complex forms and post-lexically the same regularities should be observed.

The final consonants in monosyllables (1a) are not analysed as codas and therefore do not contribute to the overall duration of the foot. Their 'non-coda' status is supported by the fact that there are no phonotactic restrictions on the quantity of the preceding vowel in CV(:)C monosyllables, while long vowels are generally absent before a true coda consonant, e.g. *bean* vs. **beanding*. Thus, the 'closed syllable shortness' regularity is observed only in non-final closed syllables and final CVCC ones but not in final CV:C and in monosyllabic CV:C words (Harris and Gussmann 1998).

In this study the following hypotheses were tested:

(3) a. H_0= Nuclei do not form feet, hence:

- no special relations hold between/among the phonetic properties of vowels within the postulated constituent.
- stressed elements (nuclei in this study; alternatively rhymes) are similar in terms of phonetic properties.
- stress-independent phonetic processes which affect vowels, e.g. pre-fortis clipping (PFC) and final lengthening (FL) are not mutually related.
- the total duration and acoustic energies of vowels in 1-, 2- and 3-syllable items must differ significantly.

b. H_1= Nuclei form feet, hence:

- there must be some systematic relations between vowels.
- these relations must be constant.

3. Data and measurements

Since this is a pilot study, only one informant was recorded (male, aged 51, a speaker of Standard Southern English). The wav. recordings were made in a quiet room with a SFX-Pro microphone placed about 30 cm from the informant's face. The test items were presented in the form of a PowerPoint presentation in a randomised order at a steady tempo (approx. 15 s per slide) to avoid the 'list reading' effect which could potentially influence the pitch of stressed vowels in particular.

The total number of test items was 72, of which 22 were CVC monosyllables, 22 were CVCV disyllables and 23 CVCVCV initially stressed trisyllables and 5 quadrisyllables with antepenultimate stress. (The full list of test items is provided in the Appendix.) Most items were morphologically simple and the stressed vowel was always placed in the same consonantal context, i.e. preceded by a voiced stop and followed by both a voiced and voiceless stop, e.g. *bit/bid/beat/bead/biddy/bitty/beady/Beatty,* etc. When the desired context was not represented in the lexicon, a similar item was chosen in such a way that the difference in the consonantal context was reduced to the minimum, e.g. *obesity* (the stressed vowel followed by a fricative rather than a plosive). Items with a sonorant in the post-stress position, e.g. *ban*, were not used in order to avoid segmentation ambiguities. It is noteworthy that some combinations required for the study were either non-existent or heavily underrespresented in the lexicon, e.g. a stressed [ʊ] followed by two syllables or a long vowel followed by two syllables, respectively. Also, initially stressed quadrisyllabic forms, except *-ory*/*-ary* formations, e.g. *cemetery*, are extremely rare. The *-ory*/*-ary* items were not included in the study since their pre-final vowel is generally elided in British English (and strengthened by secondary stress in General American). However, these lexical gaps or lexical 'repair' strategies (e.g. elision vs. secondary stress in *-ory*/*-ary* forms) also provide interesting information about the metrical structure of the English lexicon and indirectly support the preliminary conclusions of the present study.

Although the total number of test items is rather small, the number of all individual measurements made with PRAAT (duration, pitch, intensity) (Boersma and Weenink 2005) and Cricket software (acoustic energy) (available at *http://www.linguistics.ecsb.edu/faculty/gordon/projects.html*) was 1,122 (duration: 150; pitch max/mean: 324; pitch slope: 108; intensity max/mean: 324; intensity slope: 108; acoustic energy: 108).

For each vowel the duration was measured with PRAAT using waveforms and spectrograms from the point where the target vowel formant structure was reached up to the release of the following stop. An alternative method, i.e. from the moment of release of the preceding stop to the beginning of the closure phase for the following stop, was rejected. Although the durational measurements obtained by both methods were expected to be comparable, it was observed that the pitch and intensity for a particular vowel differ remarkably depending on the method. Since the pitch and intensity values in the post-target phase were invariably higher than in the pre-target phase, we assumed that the former ones may be salient for stress perception even though they are no longer accompanied by vowel periodicity.

Duration was measured in milliseconds for each individual vowel within an item, i.e. for both stressed (V_1) and unstressed vowels (V_2 and V_3), where applicable. A one-way Anova with an alpha of .05 was used to analyse the significance of the differences in mean vowel durations. First, we checked the significance of pre-fortis shortening effects on stressed vowels in words having the same number of syllables. The items containing phonemically short vowels and those having long vowels were tested separately. Secondly, the PFC effects were compared for words with different number of syllables in order to establish whether the degree of stressed vowel shortening is related to the number of syllables an item contains. Finally, mean duration of word-final vowels in di- and trisyllables was analysed. Since final lengthening is positionally conditioned, the assumption was that the differences in mean durations of word-final vowels should be insignificant and entirely independent of the PFC effects, the phonemic length of the initial stressed vowel or the number of syllables in an item. Thus, we first analysed the significance of the differences in mean durations of word-final vowels in words having the same number of syllables and then between the groups of 2- and 3-syllable items.

Additionally, the total vowel duration was calculated for each item: $T_{DUR}=V_1+(V_2 (+V_3))$. Since the intrinsic vowel durations differ considerably (Peterson and Lehiste 1960) due to the minimal execution time of an articulatory movement (Klatt 1976), the total vowel durations were compared for 1-, 2- and 3-syllable words with the same stressed vowel, e.g. *bid/biddy/bigamy*.

As far as pitch (Hz) and intensity (dB) are concerned, the maximal and mean values were measured for all vowels within the same durational selections on the spectrogram. The significance of the differences in maximal/mean pitch and intensity of stressed vowels in monosyllabic, disyllabic and trisyllabic items were tested with one-way Anova (alpha of .05). Additionally, we analysed the significance of maximal and mean intensity of unstressed vowels (V_2 and V_3) in disyllabic and trisyllabic items with a view to finding possible correlations with the intensity of the stressed vowel V_1.

Pitch and intensity slopes were also calculated for the stressed and unstressed vowels in the following way:

(4) $Pitch_{SLOPE}=P_{MAX} – P_{MIN}$ (V-finally)
$Intensity_{SLOPE}=Int_{MAX} – Int_{MIN}$ (V-finally)

Pitch and intensity slopes were thus expressed ‘statically’ (in Hz and dB, respectively) rather than ‘dynamically’ (in Hz/s and dB/s). Given the significant differences in the durations of stressed vowels in particular, the dynamic measurement was likely to produce different results for two different vowels even if their maximal and minimal pitch values were the same. For this reason, we have chosen the ‘static’ measurement which keeps pitch and intensity slopes independent from vowel duration. A one-way Anova with an alpha .05 was used to test the significance of mean pitch and intensity slope differences within the stressed vowels in 1-, 2- and 3-syllable words and their possible correlations with PFC effects and phonemic vowel length. Pitch and intensity slopes were also analysed for all unstressed vowels in 2- and 3-syllable words. Then, we checked the significance of pitch and intensity slope differences between the stressed vowels in the following combinations: (i) 1-syll. vs. 2-syll. words, (ii) 1-syll. vs. 3-syll. words and (iii) 2-syll. vs. 3-syll. words.

Finally, the acoustic energy (in decibel milliseconds, i.e. the sum total of decibel values over the entire selected window) of each vowel was measured with Cricket and the total acoustic energy was calculated for each item by summing up the energies of its component vowels. The differences in mean total acoustic energies in monosyllables, disyllables and trisyllables were tested for their significance (one-way Anova alpha .05).

4. Results and discussion

4.1 Duration of stressed vowels

The increased duration of stressed syllables (vowels) has been generally accepted to be one of the main phonetic correlates of stress (Laver 1995). The inversely proportionate relation between the duration of the stressed syllable and the number of syllables that follow has also been observed (e.g. Kim and Cole 2005). Since in this study, however, our aim was to investigate the phonetic characteristics of the entire foot, we decided to analyse the duration of stressed vowels and the interdependence between their duration and the total number of syllables within a word in relation to a phonetic regularity (PFC) which also affects the vowel duration on the one hand but is assumed to be contextually independent of stress. The assumption was that regardless of stress-dependent durational differences between the stressed vowels in shorter vs. longer items and stressed vs. unstressed vowels, the PFC effects (VC_{VOICED}>$VC_{VOICELESS}$ for alpha .05), which are related to the voicing of the consonant following the stressed vowel, should be constant. If this is not the case, i.e. the PFC effects turn out to be insignificant for some group of items, the conditioning factor must be singled out which is responsible for PFC suspension. If, as we hypothesised, the PFC effects in V_1 correlate negatively with the number syllables within an item, a higher-level constituent must be postulated 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. The PFC effects were significant in monosyllables and disyllables regardless of the phonemic length of the stressed vowel. The results are presented in (5) and (6) below.

(5) Monosyllables

a. short vowels:

followed by voiced C (*bid*):
DUR_{MEAN}: 218.6 ms $p = 0.04$
followed by voiceless C (*bit*):
DUR_{MEAN}: 154.3 ms

b. long vowels:

followed by voiced C (*bead*):
DUR_{MEAN}: 342.5 ms $p = 0.00008$
followed by voiceless C (*beat*):
DUR_{MEAN}: 175.7 ms

(6) Disyllables

a. Short Vowels:

followed by voiced C (*biddy*):
DUR_{MEAN}: 106.8 ms $p = 0.01$
followed by voiceless C (*bitty*):
DUR_{MEAN}: 74.04 ms

b. Long Vowels:

followed by voiced C (*beady*):
DUR_{MEAN}: 158 ms $p = 0.047$
followed by voiceless C (*Beatty)*:
DUR_{MEAN}: 112.7 ms

However, in trisyllabic items the effect of PFC turned out to be insignificant.

(7) a. Short Vowels:

followed by a voiced C (*bigamy*):
DUR_{MEAN}: 84.6 ms $p = 0.063$
followed by a voiceless C (*rickety*):
DUR_{MEAN}: 73.1 ms

b. Long Vowels:
followed by a voiced C (*naivety*):
DUR_{MEAN}: 125.4 ms $p = 0.47$
followed by a voiceless C (*obesity*):
DUR_{MEAN}: 103.8 ms

Since PFC significantly affects stressed vowels in mono- and disyllables but not in trisyllables, its application must be related to the number of syllables that follow the stressed one within the same morphologically simple word.

Then, we analysed the degree of V_1 shortening as a function of the number of syllables that follow. The mean durations of stressed vowels were tested for their significance in 1-, 2- and 3-syllable words, e.g. *bit/bitty/rickety*, for four groups of items separately, i.e. (i) phonemically short V_1 followed by a voiceless consonant (8a), (ii) phonemically short V_1 followed by a voiced consonant (8b), (iii) phonemically long V_1 followed by a voiceless consonant (8c), and (iv) phonemically long V_1 followed by a voiced consonant (8d). The idea behind such grouping of examples was to eliminate the possible differences in duration that may be due to the phonemic length of V_1 and PFC effects. The following results were obtained.

(8) a. Short V_1 followed by a voiceless C

	A	B	C	D
V_1	DUR_{MEAN}	DUR_{MEAN}	DUR_{MEAN}	DUR_{MEAN}
1-syll. items (*bit*)	154.3 ms	154.3 ms	154.3 ms	-----
2-syll. items (*bitty*)	74.1 ms	74.1 ms	-----	74.1 ms
3-syll. items (*rickety*)	70.4 ms	-----	70.4 ms	70.4 ms
$p < 0.05$	$p = 3E-07$	$p=0.0002$	$p = 4E-06$	$p = 0.55$

b. Short V_1 followed by a voiced C

	A	B	C	D
V_1	DUR_{MEAN}	DUR_{MEAN}	DUR_{MEAN}	DUR_{MEAN}
1-syll. items (*bid*)	211.6 ms	211.6 ms	211.6 ms	-----
2-syll. items (*biddy*)	102.7 ms	102.7 ms	-----	102.7 ms
3-syll. items (*bigamy*)	82.9 ms	-----	82.9 ms	82.9 ms
$p < 0.05$	$p = 5E-07$	$p = 0.002$	$p = 2E-06$	$p = 0.043$

c. Long V_1 followed by a voiceless C:

	A	B	C	D
V_1	DUR_{MEAN}	DUR_{MEAN}	DUR_{MEAN}	DUR_{MEAN}
1-syll. items (*beat*)	175.7 ms	175.7 ms	175.7 ms	-----
2-syll. items (*Beatty*)	112.7 ms	112.7 ms	-----	112.7 ms
3-syll. items (*obesity*)	103.8 ms	-----	103.8 ms	103.8 ms
$p < 0.05$	$p = 0.015$	$p = 0.008$	$p = 0.03$	$p = 0.69$

d. Long V_1 followed by a voiced C:

	A	B	C	D
V_1	DUR_{MEAN}	DUR_{MEAN}	DUR_{MEAN}	DUR_{MEAN}
1-syll. items (*bead*)	342.5 ms	342.5 ms	342.5 ms	-----
2-syll. items (*beady*)	158.0 ms	158.0 ms	-----	158.0 ms
3-syll. items (*naivety*)	125.4 ms	-----	125.4 ms	125.4 ms
$p < 0.05$	$p = 3E\text{-}06$	$p = 5E\text{-}05$	$p = 3E\text{-}05$	$p = 0.23$

The differences in mean V_1 durations in 1-, 2- and 3-syllable words were highly significant for both long and short vowels regardless of the PFC context (8a vs. 8b and 8c vs. 8d, column A). However, it is the increased mean V_1 duration in monosyllables that is responsible for this significance. When the three groups of items were tested separately in three possible combinations, i.e. 1syll. vs. 2-syll. words, 1-syll. vs. 3 syll. words and 2-syll. vs. 3-syll. words, it turned out that mean V_1 durations remain significantly different only when monosyllabic items are compared with di- and trisyllabic ones (columns B and C). The differences in mean durations of V_1 between 2-syllable words and 3-syllable words (column D) were generally insignificant. The only context in which these differences were significant ($p = 0.043$) was the one in which a short V_1 was followed by a voiced consonant (8b) (no PFC effects). For the other three groups of items, the *p*-values of $p = 0.55$ (8a), $p = 0.69$ (8c) and $p = 0.23$ (8d) indicate that the duration of the stressed vowel is not inversely proportionate to the number of syllables within an item. Thus, it is not directly related to the overall length of an item. In effect, the results suggest that stressed vowels are lengthened in monosyllables rather than shortened in polysyllabic forms. As further discussion will show, this conjecture is independently supported by the analysis of pitch slope within the stressed syllables.

4.2 Duration of final vowels

Interestingly, statistically significant differences were found in the mean durations of word-final unstressed syllables. Since all test items end in a vowel, we expected that in trisyllabic words final unstressed vowels would be significantly longer than non-final unstressed vowels due to word-final lengthening. This was fully confirmed by the results (V_2 and V_3 mean durations in trisyllables were 59 ms and 74 ms, respectively; p = 1,92143E-08). Due to the fact that final lengthening (FL) is positionally conditioned, there should be no significant differences in the duration of word-final vowels in 2- and 3-syllable items. However, the duration of final vowels was insignificantly different only in groups of words having the same number of syllables.

(9) a. Disyllables; Short V_1
V Final DUR_{MEAN}: 88.5 ms (*biddy)* $p = 0.79$
V Final DUR_{MEAN}: 86.5 ms (*bitty)*
b. Trisyllables: Short V_1
V Final DUR_{MEAN}: 75.6 ms (*bigamy*) $p = 0.88$
V Final DUR_{MEAN}: 73.8 ms (*rickety*)

c. Disyllables: Long V_1
V Final DUR_{MEAN}: 77.6 ms (*beady)* $p = 0.12$
V Final DUR_{MEAN}: 91.7 ms (*Beatty)*
d. Trisyllables; Long V_1
V Final DUR_{MEAN}: 55.7 ms (*naivety*) $p = 0.26$
V Final DUR_{MEAN}: 66.8 ms (*obesity*)

The differences in mean durations of final vowels were significantly different for words of two and three syllables, regardless of the phonemic length of the stressed vowel and its PFC context.

(10) a. disyllables: V_1 (Long Vowel):
followed by voiced C *bead*y: $p = 0.015$
V Final DUR_{MEAN}: 77.7 ms
trisyllables: V_1 (Long Vowel):
followed by voiced C *naivety*:
V Final DUR_{MEAN}: 63.1 ms

b. disyllables: V_1 (Long Vowel):
followed by voiceless C *Beatty*: $p = 0.035$
V Final DUR_{MEAN}: 91.7 ms
trisyllables: V_1 (Long Vowel):
followed by voiceless C *obesity*:
V Final DUR_{MEAN}: 66.8 ms

c. disyllables: V_1 (Short Vowel):
followed by voiced C *biddy*: $p = 0.008$
V Final DUR_{MEAN}: 88.5 ms
trisyllables: V_1 (Short Vowel):
followed by voiced C *bigamy*:
V Final DUR_{MEAN}: 75.6 ms

d. disyllables: V_1 (Short Vowel):
followed by voiceless C *bitty*: $p = 0.001$
V Final DUR_{MEAN}: 99.9 ms
trisyllables: V_1 (Short Vowel):
followed by voiceless C *rickety*:
V Final DUR_{MEAN}: 84.8 ms

Thus, the degree of FL is dependent upon the syllabic ‘distance’ between the stressed vowel and the word-final one. There is a negative correlation between the actual duration of word-final vowels and the number of syllables within an item in three groups of examples: *beady* vs. *naivety* (r=-0.889), *Beatty*-*obesity* (r=-0.878), *bitty*-*rickety* (r=-0.489). In one group, i.e. short V_1 followed by a voiced consonant *biddy*-*bigamy* (r=0.03) negative correlation was not confirmed, though. Strangely enough, it is the same context in which the differences in mean V_1 durations between disyllables and trisyllables were, somehow exceptionally, significant (8b, column D). We interpret these two facts as being interrelated, i.e. since the duration of stressed vowels in *biddy-bigamy* items happens to be significantly different, the durations of final vowels in the *bigamy* type of items, despite being significantly shorter than final vowels in the *biddy* type of items, do not need to negatively correlate with the number of syllables within an item. As the analysis of the total vowel durations will show, this is due to the equalisation of the T_{DUR} within all three groups of words.

The discussion so far indicates that PFC and FL are interrelated. Within the groups of 2-syll. items and 3-syll. items with phonemically long V_1 the final nucleus is longer if PFC affects the stressed Nucleus. Since long vowels are affected by PFC to a greater extent, there is a need for length compensation in the final vowels. Thus, the PFC effects are compensated for by the duration of the final vowel. This indicates that the stressed vowel and the final vowel are able to 'negotiate' their durations. Such 'communication' between the first (stressed) vowel and the final one is possible only if a mediating constituent (foot) is postulated.

Both the duration of the stressed vowel and the duration of the final vowel depend on total number of nuclei ('syllables'). Thus, as the number of nuclei following the stressed one increases, the DUR_{MEAN} of *all* Nuclei decreases. Crucially, the durational characteristics of stressed vowels is variable. Since they significantly differ in duration, stress generalisations are unattainable in relation to this phonetic property of vowels alone.

4.3 Total vowel duration

In order to test the hypothesis, according to which the significant differences in the durations of stressed and final unstressed vowels are ancillary to the equalisation of total vowel duration (T_{DUR}) in 1-, 2- and 3-syllable words, we analysed the significance of the differences in T_{DUR} for items with a phonemically identical stressed vowel, e.g. *bid*/*biddy*/*bigamy*. If we roughly assume that the duration of a stressed vowel is 100%, and the following unstressed one(s) 50%, than the following T_{DUR} are expected:

Monosyllable:	= 100%
Disyllable:	≈ 150%
Trisyllable:	≈ 200%

Thus, the resulting *p* values for the differences in T_{DUR} in monosyllables vs. disyllables vs. trisyllables should be much below .05. However, for items sharing the same stressed vowel the differences in T_{DUR} were always non-significant regardless of the PFC context. (The vowels /ʊ/ and /ɜː/ are not included since they are underrepresented in trisyllabic items.)

(11) a. Short vowels:

	PFC	no PFC	
/ɪ/	176.8 ms	185.5 ms	$p = 0.75$
/e/	165.1 ms	189.0 ms	$p = 0.13$

/æ/	202.7 ms	250.3 ms	$p = 0.13$
/ʌ/	178.6 ms	205.7 ms	$p = 0.16$
/ɒ/	189.2 ms	197.4 ms	$p = 0.5$

b. Long vowels:

	PFC	no PFC	
/iː/	177.9 ms	252.7 ms	$p = 0.12$
/ɑː/	228.4 ms	315.0 ms	$p = 0.14$
/ɔː/	211.1 ms	267.5 ms	$p = 0.22$
/uː/	165.1 ms	246.7 ms	$p = 0.21$

Moreover, the differences in mean T_{DUR} in monosyllables, disyllables and trisyllables with a phonemically short V_1 (196.9 ms when followed by a voiced consonant and 183.5 ms when followed by a voiceless one) were also insignificant ($p = 0.07$), regardless of the actual vowel quality. For items with a phonemically long V_1 the differences in mean T_{DUR} (270.8 ms when V_1 was followed by a voiced consonant and 193 ms when it was followed by a voiceless consonant) were significant, though ($p = 0.002$) when the vowel quality was disregarded. However, the differences in mean T_{DUR} in disyllables and trisyllables (231.6 ms when followed by a voiced consonant and 202.5 ms when followed by a voiceless one) were insignificant ($p = 0.14$). This indicates that in isolated pronunciations monosyllables containing a phonemically short vowel (unlike those with phonemically long Vs) lengthen their vowels significantly.

The analysis of total vowel durations points at a strong tendency to equalise the T_{DUR} in words having a different number of syllables. The hypothesis concerning the foot isochrony in English needs to be reconsidered, i.e. rather than being a durational property of interstress intervals, isochrony must be sought in the total duration of stress bearing parts of the syllable only (nuclei and possibly codas, although the latter option was not analysed in the present study).

As far as the phonetic properties of the foot are concerned, the duration of stressed vowels alone turns out to be an unreliable predictor of stress. Stressed vowels were shown to shorten as the number of the syllables increases. However, no further shortening was observed in 3-syllable items. This indicates that the maximum number of syllables within the English foot is three. Further reductions of V_1 duration may be impossible due to articulatory constraints (minimal execution time for an articulatory gesture). They may also be perceptually costly. Insufficient duration of stressed vowels may also make the realisation of other stress related phonetic properties (like pitch and intensity) impossible. If further reductions, on the other hand, did not occur in feet longer than 3 nuclei, the 'equalisation' of T_{DUR} in 1-, 2- and 3-syll. items would remain inexplicable.

The constraint on the maximal foot size is also independently confirmed by the fact that English lexicon disfavours morphologically simple, initially stressed forms of more than three syllables. When they do occur, e.g. *cemetery*, *territory*, *Mandarin*, the foot trisyllabicity is restored either through vowel elision (RP English) or vowel strengthening and secondary stress (General American).

In conclusion, as far as duration is concerned, the most important results of the present pilot study are the equalisation of T_{DUR} observed in items of different number of syllables and the interdependence of PFC and FL.

4.4 Pitch

In this study, for isolated pronunciations of test items the differences in pitch (both maximal and mean) of stressed vowels turn out to be insignificant. The differences in P_{MAX} were significantly different (p = 0.046) only when V_1 was followed by a voiced consonant:

Monosyllables	P_{MAX}: 142.83 Hz
Disyllables	P_{MAX}: 83.49 Hz
Trisyllables	P_{MAX}: 102.92 Hz

However, no systematic relation was found between P_{MAX} of V_1 and the number of syllables:

P_{MAX} (1-syll) > P_{MAX} (2-syll) and P_{MAX} (3-syll)

but

P_{MAX} (2-syll) < P_{MAX} (3-syll)

Similarly, P_{MEAN} was found significantly different (p = 0.013) only for short stressed vowels (and only when the following consonant was voiceless):

Monosyllables	P_{MEAN}: 60.95 Hz
Disyllables	P_{MEAN}: 107.28 Hz
Trisyllables	P_{MEAN}: 100.82 Hz

Again, there is no systematic relation between P_{MAX} of V_1 and the number of syllables:

P_{MEAN} (1-syll) < P_{MEAN} (2-syll) and P_{MEAN} (3-syll)

but

P_{MEAN} (2-syll) > P_{MEAN} (3-syll)

All other differences in P_{MAX} and P_{MEAN} of V_1 were statistically insignificant (p > .05). Moreover, standard deviation was always very high within each group of items, making generalisations concerning P_{MAX} and P_{MEAN} of V_1 impossible. Very often P_{MAX} of the same vowel in similar items differed considerably, e.g. by approx. 324 Hz (/ʊ/ in *good* and *foot*) and by 377 Hz (final /i/ in *body* and *buddy*).

Thus, P_{MAX} and P_{MEAN} turn out to be phonetically and phonologically irrelevant properties of stressed vowels in words produced in isolation. This result is not as surprising as it may seem, though. If P_{MAX}/P_{MEAN} of both stressed and unstressed nuclei were rigidly tied up (even as a relational property), pitch would not be available for other functions it must perform (e.g. focal/emphatic, structural (Q/S), intonational or musical).

Moreover, pitch may also be speaker-/sex-/style-dependent. Thus, pitch must be set free lexically since its main stress-related functions are post-lexical and non-phonological.

4.5 Pitch slope

Pitch slope (P_{SLOPE}=P_{MAX} - P_{MIN} [V-finally]) was first analysed for groups of items with the same number of syllables. The results show that in syllabically homogeneous words the differences in mean pitch slope in V_1 are generally insignificant and independent of the voicing of the following consonant (*bit*/*bid* $p = 0.14$; *biddy*/*bitty* $p = 0.17$; *bigamy*/*rickety* $p = 0.18$; *bead*/*beat* $p = 0.61$; *beady*/*Beatty* $p = 0.23$; *naivety*/*obesity* $p = 0.76$) and the phonemic length of the stressed vowel (*bit*/*beat* $p = 0.4$; *bid*/*bead* $p = 0.45$; *bitty*/*Beatty* $p = 0.65$; *biddy*/*beady* $p = 0.21$; *rickety*/*obesity* $p = 0.27$).[2] However, the differences in mean P_{SLOPE} between the stressed vowels in 1-syll. items, on the one hand, and 2- and 3-syll. items on the other, were always highly significant. For all stressed vowels in both PFC contexts the following results were obtained.

(12) Mean P_{SLOPE} V_1 1-syll: 66.53 Hz $p = 0.00000000007$
Mean P_{SLOPE} V_1 2-syll: 11.05 Hz
Mean P_{SLOPE} V_1 3-syll: 10.33 Hz

However, there is no significant difference in pitch slope of V_1 between disyllables and trisyllables. Thus, P_{SLOPE} in V_1 is significantly greater in 1-syll items.

(13) Mean P_{SLOPE} V_1 2-syll: 11.05 Hz $p = 0.81$
Mean P_{SLOPE} V_1 3-syll: 10.33 Hz

The significance of pitch slope in V_1 was also tested separately for the following groups of items: (i) short vowels (with and without the PFC variable), (ii) long vowels (with and without the PFC variable). In all cases the stressed vowels in monosyllables had greater pitch slope than the corresponding vowels in di- and trisyllables.

(14) Mean P_{SLOPE} V_1
a. short vowel followed by a voiceless C
1-syll: 50.32 Hz $p = 0.00004$
2-syll: 11.63 Hz
3-syll: 9.45 Hz

2-syll: 11.63 Hz $p = 0.59$
3-syll: 9.45 Hz

[2] Pitch slope was significantly different only in trisyllables (*bigamy*/*naivety* p=0.002) in which V_1 was followed by a voiced consonant (*bigamy*/*naivety* p=0.002). Admittedly, this result remains problematic.

b. short vowel followed by a voiced C

1-syll:	96.52 Hz	$p = 0.0005$
2-syll:	6.42 Hz	
3-syll:	12.27 Hz	
2-syll:	6.42 Hz	$p = 0.54$
3-syll:	12.27 Hz	

c. long vowel followed by a voiceless C

1-syll:	47.24 Hz	$p = 0.03$
2-syll:	15.07 Hz	
3-syll:	15.74 Hz	
2-syll:	15.07 Hz	$p = 0.91$
3-syll:	15.74 Hz	

d. long vowel followed by a voiced C

1-syll:	69.29 Hz	$p = 0.027$
2-syll:	12.76 Hz	
3-syll:	18.34 Hz	
2-syll:	12.76 Hz	$p = 0.25$
3-syll:	18.34 Hz	

Pitch slope differences in short and long stressed vowels in an identical PFC context followed the same pattern, i.e. the vowels in monosyllables had significantly greater pitch slope than the stressed vowels in 2- and 3-syllable words.

(15) Mean P_{SLOPE} V_1

a. short vowels:

1-syll:	73.42 Hz	$p = 3.1E\text{-}07$
2-syll:	9.02 Hz	
3-syll:	7.60 Hz	
2-syll:	9.02 Hz	$p = 0.55$
3-syll:	7.60 Hz	

b. long vowels

1-syll:	88.91 Hz	$p = 0.037$
2-syll:	13.48 Hz	
3-syll:	17.21 Hz	
2-syll:	13.48 Hz	$p = 0.6$
3-syll:	17.21 Hz	

The significance of P_{SLOPE} within the stressed vowels in 1-syll. items as opposed to 2-/3-syll. items and the significantly greater vowel duration in monosyllables implies that these two properties are related. We interpret this fact as being related to foot minimality requirement. The articulatory execution of P_{SLOPE} within the vowel in 1-syll. items requires more time. Hence, the significantly greater V_{DUR} in 1-syll. items pronounced in isolation (=pre-pausal context). P_{SLOPE}, then, is a phonetic implementation of the phonological requirement of foot 'binarity', i.e. feet must contain (at least) two distinct tones (T). (cf. Gordon's (2000) idea of 'tonal crowding' and the cross-linguistic avoidance of final stress). By inference, what has traditionally been referred to as 'Trisyllabic Laxing' in fact seems to be a tonally conditioned 'Monosyllabic Lengthening'. In 2- and 3-syll. items (=feet) the distinct tones are distributed over 2 or 3 consecutive Nuclei, hence no need for the stressed vowel to be longer.

4.6 Intensity

The relations between the maximal and mean intensity values obtained for V_1, V_2 and V_3 generally mirror the durational relations between the corresponding vowels. For instance, the reduced V_1 duration in the PFC context (*bid* vs. *bit*) corresponds to lower V_1 intensity ($p = 0.028$). In 2- and 3-syll. items, on the other hand, Int_{MAX} of V_1 is independent of the PFC context, e.g. $p = 0.35$ in 2-syll. items (*biddy* vs. *bitty*) and $p = 0.65$ in 3-syll. items (*bigamy* vs. *rickety*). Unlike duration, however, $Int_{MAX/MEAN}$ in disyllables and trisyllables decreases $V_1 > V_2 > V_3$[3] and $Int_{MAX/MEAN}$ of V_1 in 1-syll. items on the one hand and 2- and 3-syll. items on the other was not significantly different (in absolute values it was almost identical), e.g.

Int_{MAX} V_1 1-syll:	78.15 dB	$p = 0.27$
Int_{MAX} V_1 2-syll:	78.45 dB	
Int_{MAX} V_1 3-syll:	79.59 dB	

Since the integration of intensity and duration represents the acoustic energy of a sound, the identical intensity values at the beginning of the putative foot suggest that each foot starts with the same amount of acoustic energy. The differences in $Int_{MAX/MEAN}$ of V_2 and V_3 between 2-syll. and 3-syll. feet depend on the distribution of the (roughly identical) amount of acoustic energy among the foot recessive nuclei (=Int_{SLOPE}). Indirectly, the identical Int_{MAX} of V_1 in all items suggests that the amount of air available foot-initially is similar (at least for words pronounced in isolation). What makes 1-/2- and 3-syll. items (=feet) different is the distribution of the energy over the component Nuclei. The total acoustic energy within each foot must be, therefore, constant.

[3] Note that V_3 was invariably longer than V_2 in trisyllabic items due to the FL effects.

4.7 Intensity slope

Similarly to pitch slope, the differences in mean Int_{SLOPE} within the stressed vowels in words having the same number of syllables were generally insignificant and, again, independent of the PFC context (*biddy*/*bitty* $p = 0.06$; *bigamy*/*rickety* $p = 0.96$; *bead*/*beat* $p = 0.36$; *naivety*/*obesity* $p = 0.83$) and the phonemic length of the stressed vowel (*bit*/*beat* $p = 0.19$; *bid*/*bead* $p = 0.49$; *bitty*/*Beatty* $p = 0.4$; *biddy*/*beady* $p = 0.31$; *bigamy*/*naivety* $p = 0.84$; *rickety*/*obesity* $p = 0.96$). Significant differences in Int_{SLOPE} which are related to the voicing of the following consonant were observed only in two groups of items: *bit*/*bid* ($p = 0.035$) and *beady*/*Beatty* ($p = 0.015$).

Since the relation between the Int_{SLOPE} on the one hand and the phonemic length of the stressed vowel and the voicing of the following consonant on the other turned out to be insignificant, we analysed the significance of Int_{SLOPE} differences in V_1 in relation to another variable, namely the number of syllables within an item. This provided compelling arguments for the interdependence between the length of an item and degree of intensity slope in V_1:

Int_{SLOPE} V_1 1-syll: 16.26 dB $p = 4.1E\text{-}23$
Int_{SLOPE} V_1 2-syll: 8.02 dB
Int_{SLOPE} V_1 3-syll: 4.01 dB

Unlike pitch slope, however, Int_{SLOPE} was also significantly different within V_1 in di- and trisyllables:

Int_{SLOPE} V_1 2-syll: 8.02 dB $p = 6.9E\text{-}06$
Int_{SLOPE} V_1 3-syll: 4.01 dB

Another important fact is that the decrease in Int_{SLOPE} within the stressed vowels is strikingly regular, i.e. 16.26 dB in monosyllables > 8.02 dB in disyllables > 4.01 dB in trisyllables, i.e. in disyllables Int_{SLOPE} of the stressed vowels is reduced by 50% and in trisyllables by 75%. Given the facts that (i) the stressed vowels in all items have nearly identical Int_{MAX} and (ii) stressed vowels in monosyllables are significantly longer than those in di- and trisyllables, this regularity must be related to the total acoustic energy (intensity over time) of the putative foot. Therefore, the most dramatic intensity tilt is observed in monosyllables, since all acoustic energy must be 'consumed' over one (albeit lengthened) vowel. In words of two and three syllables the same amount of energy must be distributed more economically over the available syllables. Hence, we observe a steady (≈50%) decrease of V_1 intensity as the number of the syllables increases as well as significant differences in Int_{SLOPE} of final unstressed vowels in di- and trisyllables.

Int_{SLOPE} V_{FINAL} 2-syll: 3.97 dB $p = 1.6E\text{-}06$
Int_{SLOPE} V_{FINAL} 3-syll: 1.44 dB

Thus, the durational differences between the stressed vowels in words of different number of syllables prove ancillary to the phonetic realisation of other acoustic features (intensity and pitch) of stressed vowels.

4.8 Total acoustic energy

The total acoustic energies in decibel milliseconds were calculated for all 72 test items by summing the energies of all component vowels within a word. Then, we analysed the significance of the differences in mean total acoustic energies in mono-, di- and trisyllables disregarding the PFC context and the phonemic length of the stressed vowels. Given that the differences in mean and maximal intensities of stressed vowels and the differences in total vowel durations in mono-, di and trisyllabic words are non-significant, we expected that the differences total acoustic vowel energies should also be non-significant. The results obtained fully support this prediction.

(16)	a. Total E_{AC} 1-syll (V_1):	218187	$p = 0.97$
	Total E_{AC} 2-syll (V_1+V_2):	218943	
	b. Total E_{AC} 1-syll (V_1):	218187	$p = 0.6$
	Total E_{AC} 3-syll (V_1+V_2+V_2):	227134	
	c. Total E_{AC} 2-syll (V_1+V_2):	218943	$p = 0.49$
	Total E_{AC} 3-syll (V_1+V_2+V_2):	227134	
	d. Total E_{AC} 1-syll (V_1):	218187	$p = 0.81$
	Total E_{AC} 2-syll (V_1+V_2):	218943	
	Total E_{AC} 3-syll (V_1+V_2+V_2):	227134	

The ‘equalisation’ of acoustic energy within the foot (similarly to the ‘equalisation’ of total foot duration in items having a different number of syllables) suggests that the foot may be defined as an equalised portion of acoustic energy of its component vowels (or rhymes) rather than an interstress interval. This approach eliminates the need to resort to parameters like *%V*, *ΔC* and *ΔV* (cf. Ramus et al. 1999), naturally explains why stress is cross-linguistically insensitive to onset structure and why PFC and FL effects within a foot are interdependent.

5. Conclusion

Since there are no significant differences in mean total E_{AC}, mean T_{DUR} and $Int_{MAX/MEAN}$ of stressed vowels and given the fact that – apart from intensity – the stressed nuclei (V_1) do not share any other phonetic properties, H_0 must be rejected. This means that nuclei (possibly rhymes, though this hypothesis was not tested here) are grouped together into feet in such a way that the following relations hold:

$$\text{Foot} = \left\{ \begin{array}{ll} \text{Total } E_{AC} = V_1+(V_2\ (+V_3)) & \\ T_{DUR} = V_1+(V_2\ (+V_3)) & \\ INT_{MAX/MEAN}\ V_1 & \text{1-syll=2-syll=3-syll} \\ Int_{SLOPE}\ V_1 & \text{1-syll>2-syll>3-syll} \\ P_{SLOPE}\ V_1 & \text{1-syll>2-syll/3-syll } \textit{but} \text{ 2-syll} \approx \text{3-syll} \end{array} \right\}$$

The relational requirements on foot structure explicitly rule out only ‘unary’ feet (due to their insufficient duration (=the lack of space for the execution of pitch/intensity slope). Thus, in isolated (=pre-pausal) pronunciations short monosyllables must lengthen their vowels, which makes the realisation of pitch slope possible.

Given the conditions above, the ‘optimal foot’ appears to be ‘binary’, since binarity provides optimal conditions for the realisation two distinct tones. However, the binarity requirement is neither syllabic nor moraic but tonal. The durational differences between the stressed vowels in words having a different number of vowels (‘syllables’) are thus epiphenomenal to pitch contrasts required within the foot. The stressed vowels in monosyllables, however, require a greater pitch slope than those in di- and trisyllables. This, we believe, is related to the lack of a consonantal ‘recovery’ period (cf. Delgutte 1982) between the two tones realised over one vowel. Since the stressed vowels in di- and trisyllables are followed by an onset consonant, a less robust contrast in pitch between V_1 and $V_2/V_3$is sufficient.

This does not mean, however, that a trochee is the only possible English foot type, as some formal phonological approaches maintain. The fact that both mean total E_{AC} and T_{DUR} in trisyllabic items are not significantly different from those in mono- and disyllables suggests that ternary (dactylic) feet are equally viable structures. Moreover, the articulatorily motivated suspension of PFC effects on the stressed vowels in trisyllabic words together the lexical scarcity of morphologically simple words stressed on the pre-antepenult indicate that ternary feet are in fact maximal in English. Our results do not support the word-final syllable ‘extrametricality’ assumption (Liberman and Prince 1977) either, since the acoustic characteristics of V_3 (e.g. through the reduction of FL effects) in trisyllabic words is indeed related to the acoustic characteristics of the two preceding vowels. Finally, the equalisation of total E_{AC} and T_{DUR} of vowels ‘extracted’ from the syllabic interstress interval seems to support the assumption of foot-level isochrony in English.

References

Abercrombie, D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Boersma, P. and D. Weenink. 2005. Praat: doing phonetics by computer (version 5.0.32). http://www.praat.org/ (13 August 2008).

de Lacy, P. 2007. Quality of data in metrical stress theory. *Cambridge Extra Magazine.* Issue 2.

Delgutte, B. 1982. Some correlates of phonetic distinctions at the level of the auditory nerve. In Carlson, R and B. Granström (eds.), *The representation of speech in the peripheral auditory system*: 131–150. Amsterdam: Elsevier.

Giegerich, H. 1992. *English Phonology: An Introduction*. Cambridge: Cambridge University Press.

Gordon, M. 2000. The tonal basis of final weight criteria. *Chicago Linguistics Society (CLS)* 36 *(Main Session)*: 141–156.

Gordon, M. 2005. A perceptually-driven account of onset-sensitive stress. *Natural Language and Linguistic Theory* 23: 595–653.

Gussmann, E. 2002. *Phonology: Analysis and Theory.* Cambridge: Cambridge University Press.

Harris, J. 1994. *English Sound Structure*. Oxford: Blackwell.

Harris, J. and E. Gussmann. 1998. Final Codas: why the west was wrong. In Eugeniusz Cyran (ed.), *Structure and Interpretation. Studies in Phonology*: 139–162. Lublin: Folium.

Hayes, B. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Kim, H. and J.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.

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

Laver, J. 1995. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Liberman, M. and A. Prince. 1977. On Stress and Linguistic Rhythm. *Linguistic Inquiry* 8.2: 249–336.

Peterson, G. E. and I. Lehiste. 1960. Duration of Syllable Nuclei in English. *Journal of the Acoustical Society of America.* 32: 693–703.

Ramus, F., M. Nespor and J. Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 72: 1–28.

Selkirk, E. 1980. The role of prosodic categories in English word-stress. *Linguistic Inquiry* 11: 563–605.

Appendix:

List of test items

a. <u>Monosyllables</u>: (22)

Short vowels:	**Long vowels:**
bid-bit	bead-beat
bed-bet	bard-Bart
bad-bat	board-bought
bud-butt	booed-boot
bod-dot	bird-Bert
bood-foot	

b. <u>Disyllables</u> (22)

Short vowels:	**Long vowels:**
biddy-bitty	beady-Beatty
beddy-Betty	bardy-barty
baddie-batty	bawdy-corty
buddy-butty	moody-booty
body-dottie	birdy-bertie
goodie-footie	

c. <u>Trisyllables</u> (28)

Short vowels:	**Long vowels:**
bigamy-rickety	naivety-obesity
liberty-pickardy*	cardamon-Arcady
legacy-Cecily	Cordoba-sportily
editor-metaphor	nudity-mutiny
abbacy-academy	
Agatha-Vatican	
thuggery-tuppeny	
buggery-uppity	
monogamy-hippopotamus	
prodigy-locative	

*nonce word