The question of randomness in English foot timing: a control experiment

Briony Williams and Steven M. Hiller (Centre for Speech Technology Research, 80 South Bridge, Edinburgh EH1 1HN, UK)

Suggested running title: Randomness in English foot timing

Abstract

Isochrony has been considered only in terms of stressed syllables. However, it may also be a random property of unstressed syllables, and a control experiment was deemed necessary. A hand-transcribed database of 98 sentences, each produced by three speakers, formed the input to an algorithm calculating durations of feet, number of syllables per foot, and mean syllable duration within each foot. In each output dataset, feet were based on one of the following criteria: stressed, tense, unreduced, random, or arbitrary syllables (the latter based on ordinal numbers of syllables within the utterance). Calculations were made of the correlations between foot duration and number of syllables per foot, and between mean syllable duration and number of syllables per foot. The 'foot compression effect' was shown to be nonrandom, and due to linguistic rather than arbitrary factors. A detailed examination of actual syllable durations was then carried out. The main determinants of syllable duration were the number of constituent segments, and syllable status in terms of target/non-target. A small but significant syllable shortening effect was also found, dependent on the number of syllables in the foot, which was linguistic rather than random.

1 Introduction

1.1 *The concept of isochrony*

Much has been written on the concept of isochrony in English: the alleged tendency for stressed syllables to occur at equal intervals regardless of the number of intervening unstressed syllables (Lehiste 1977 reviews the study of isochrony in English). Measurements of speech rhythm in English have disproved the claim that stressed syllables occur at precisely equal intervals (references are given in Dauer 1983), and the more usual claim is the weaker one that stressed syllables merely tend to occur at intervals more regular than can be accounted for by the number of intervening unstressed syllables (eg. Halliday 1967). Isochrony is then said to be psychologically real for the listener, who perceives more regularity than is actually present in the acoustic signal (Lehiste 1977, Buxton 1983).

In English, it is claimed that stressed syllables tend to recur at regular intervals, and so this is an example of a `stress-timed' language. In French, on the other hand, the claim is that syllables tend to recur at equal intervals whether stressed or unstressed, and so this is an example of a `syllable-timed' language. The distinction between stress-timed and syllable-timed languages was first made by Pike, who comments that "the tendency toward uniform spacing of stresses in material which has uneven numbers of syllables within its rhythm groups can be achieved only by destroying any possibility of even time spacing of syllables" (Pike 1946: 34). This means that the two timing modes should be incompatible. The distinction was later upheld by Abercrombie, who writes: "As far as is known, every language in the world is spoken with one kind of rhythm or with the other" (Abercrombie 1967: 97). Research since this dictum has tended to concentrate on whether particular languages should be assigned to the stressed-timed or syllable-timed category.

The classification of French as syllable-timed has been disputed by Wenk & Wioland (1982), who find that a sequence of twelve syllables has less than twice the duration of a sequence of six syllables. This is contrary to the predictions of syllable-timing, which would forecast a duration of exactly twice as long. They suggest an alternative analysis of French, based on `rhythmic groups' rather than individual syllables, stating that `the greater the number of rhythmic groups in an utterance, the greater the amount of time the utterance will be allotted' (op. cit.: 194-5). They prefer to categorise English and French in terms of `leader-timed' versus `trailer-timed'. In a leader-timed language such as English, the stressed syllable is the first one in the rhythmic unit; while in a trailer-timed language such as French, the stressed syllable is the final one in the rhythmic unit. This categorisation is upheld by Fletcher (1991), whose experiments lead her to the observation that "French is as accent trailer-timed as English is stress-timed".

Roach (1982) measures syllable durations in six languages (three stress-timed, three syllable-timed) and compares the standard deviations of these durations. He finds that the standard deviations for French (75.5 ms) and English (86.0 ms) syllable durations do not differ markedly, and observes that this seems to be counter-evidence to the claim that French, as a syllable-timed language, would have less variation in syllable duration than would English (a stress-timed language). He then measures interstress intervals in the same six languages and compares the standard deviations of these interstress intervals. He finds that the stress-timed languages in fact show a greater amount of variance in the deviations (of observed from expected durations) than do the syllable-timed languages. This is contrary to the predictions of stress-timing, which would forecast the opposite result. A similar finding occurs in Dauer (1983), who observes that "stresses occur no more regularly in English than they do in any other language with clearly definable stress".

The overall impression gained from these studies is that the distinction between stress-timed and syllable-timed languages is not satisfactory. Roach (1982) suggests that "there is no language which is totally syllable-timed or totally stress-timed -- all languages display both sorts of timing; languages will, however, differ in which type of timing predominates" (op. cit.: 78). This seeemingly sensible compromise runs the risk of defeating the whole point of the concepts `stress-timed' and `syllable-timed' (as observed above, they were originally regarded as mutually exclusive). In the light of this risk, it is appropriate to ask just what it is that is being measured, and whether it is the correct thing to be measuring. This paper is an attempt to make a start on this question by asking whether what is being measured differs significantly from chance -- i.e. whether it is worth measuring at all.

1.2 *Need for a control experiment*

In the work outlined above, and in similar studies, the working assumption has been that the unit of isochrony in stress-timed languages is the interstress interval (or `foot'), which has usually been taken to begin with (or, in French, end with) the stressed syllable. However, there have been no investigations into the question of whether the same tendency towards isochrony could be obtained if one were to delimit the foot according to randomly-chosen syllables, or according to another type of syllable chosen on a different basis from stressedness. In short, there has been no control experiment to test the assumption that the relevant patterning factor is stress.

This is what the present work is designed to investigate. The experiment confines itself to the question of whether the timing effects said to be due to stress are in fact significantly different from chance. If it were the case that a tendency towards isochrony of feet could be shown when `feet' are held to begin at, for example, every third syllable, or at randomly-chosen syllables, then even though isochrony would indeed have been demonstrated, it could not be said to have any linguistic significance. It has been assumed in previous studies that any isochrony found must necessarily have linguistic significance: this remains to be proven.

Given the above considerations, the following statements can be said to encapsulate the three separate working assumptions made in isochrony studies of English to date.

1) Isochrony exists in English (even if only as a tendency).
2) Isochrony is linguistically significant.
3) Isochrony is based on stressed syllables.

These are the separate assumptions that this experiment aims to test. It is emphasised that the object of study is not the fundamental mechanism underlying any isochrony found, but rather the simple question of whether any isochrony found is significantly greater than chance and linguistically significant. The language to be studied is English, which is traditionally taken to show a tendency towards isochrony based on stressed syllables. The term `foot' is here preferred to `interstress interval' as the latter term implies that the patterning factor is the stressed syllable, while the former does not explicitly carry any such implication.

1.3 *What counts as evidence for isochrony?*

Before the assumptions can be tested, it is necessary to define the kind of result that will be taken as evidence for isochrony. It is unrealistic to expect `metronomic' isochrony in English, as feet are by no means all of the same duration. Roach (1982) found that, out of six languages, English exhibited the greatest amount of variation in foot duration.

A useful measure will be the 'compression effect' on foot duration. If there were no such effect, then (for example) feet of four syllables would be twice as long as feet of two syllables, and feet of six syllables would be twice as long as feet of three syllables and three times as long as feet of two syllables. This is what would traditionally be expected in a syllable-timed language. However, if it can be shown that feet of four syllables are less than twice as long as those of two syllables, or feet of six syllables are less than three times as long as those of two syllables, then there is evidence of a compression effect on foot durations. Nakatani et al. (1981) found no compression effect for English reiterant speech (repeated `ma' syllables mimicking the prosody of the intended natural speech). The work reported here looks for a compression effect for natural (i.e. not reiterant) English speech.

A second measure of isochrony is syllable duration. If isochrony exists, then syllables from, for example, a five-syllable foot will tend to be shorter than comparable syllables from a two-syllable foot. In the present study, to begin with, an approximation to syllable duration was obtained by dividing the duration of each foot by the number of syllables in the foot. This gave mean syllable duration per foot, rather than true syllable durations. In their study of reiterant English speech, using actual syllable durations, Nakatani et al. (1981: 101) found a very small foot-conditioned syllable shortening effect of this kind, which was independent of word-final effects. The work reported here looks for a syllable shortening effect in non-reiterant English speech. Crystal & House (1990), working with two scripts of read speech (16 and 17 sentences per script), found no shortening of stressed or unstressed syllables with increasing number of syllables per foot, but did find some shortening of unstressed vowels with increasing number of segments per syllable. On the other hand, Campbell (1988), working with a dramatically-read literary passage, found shortening of syllable durations with increasing number of syllables in the foot.

Given the above two criteria for a tendency towards isochrony, the necessary evidence for confirming assumptions 1, 2 and 3 is as shown below under 4, 5 and 6 respectively. The experiment reported here aims to assess whether such evidence can be found in English.

4) One or more types of syllable can be found that function as a patterning principle for feet that show a tendency towards isochrony that is significantly greater than chance.
5) These types are based solely on linguistic criteria.
6) Stressed syllables lead to feet showing the strongest tendency towards isochrony of all these types.

1.4 *Hypotheses tested*

Given the types of evidence required under 4, 5 and 6 above, the task is to manufacture feet based on different patterning principles, but taken from the same speech data. These will then be tested for any tendency towards isochrony (as defined in 1.3 above). Different types of patterning principle are to be tested, as follows.

Random feet, with number of syllables per foot determined by a random number generator, provide a much-needed control condition. If other types of foot show no more tendency towards isochrony than this type of foot, then it cannot be concluded that isochrony in English is significantly greater than chance. Therefore this type of foot is needed in the search for the kind of evidence outlined in 4 above.

Arbitrary feet, with number of syllables per foot determined by a fixed but arbitrary algorithm (e.g. every third syllable starts a new foot) provide a means of looking for the type of evidence under 5 above. If linguistically-based feet (i.e. the four types outlined below) show no more tendency towards isochrony than this type of foot, then it cannot be concluded that the tendency towards isochrony is linguistically significant.

Tense feet, with a new foot starting at every syllable containing a tense vowel (i.e. diphthong or long monophthong) are a means of testing whether isochrony is dependent on vowel type (tense versus lax, where lax vowels are all short monophthongs). This type of foot provides a means of testing for the type of evidence outlined in 6 above, since if feet based on stressed syllables show no more tendency towards isochrony than feet based on tense vowels, then it cannot be concluded that isochrony is based on stressed syllables.

Full feet, with a new foot starting at every full (i.e. unreduced) vowel, are similar to tense feet in providing a means of testing for the type of evidence under 6 above. Reduced vowels are schwa and the centralised versions of short /i/ and /u/; full vowels are all other vowels (syllabic consonants are included under 'reduced vowels').

Lexically stressed feet, with a new foot beginning at every syllable that bears the primary lexical stress of an orthographic word, is one form of testing for stress-based isochrony. A lexically stressed syllable is merely a candidate for sentence-level stress, and may or may not receive stress in any given utterance: not every full vowel is also lexically stressed. As the word `stress' has been used to mean many different things, it is necessary to test more than one definition of it.

Accented feet, with a new foot beginning at every syllable that bears a pitch accent (i.e. where there is an easily-perceived peak in the pitch contour), is the other form of stress to be tested. Pitch accented syllables may be determined auditorily, in contrast to lexically stressed syllables, which are merely the phonological potential locations for accent.

1.5 *Redundancies in syllable types*

There are some redundancies in syllable types which may affect the interpretation of the results. These are as follows.

7) Tense vowels are also full, i.e. unreduced (but not always vice-versa).

8) Accented syllables are also lexically stressed (but not always vice-versa).

9) The vowels of lexically stressed syllables are also full, at least for content words (not always vice-versa, but, in this data, nearly always).

2 Data processing

The types of foot outlined above were calculated over a hand-transcribed speech database of 98 phonemically dense sentences recorded by three adult male speakers of RP English (JL, GW and PJ). The sentences each exemplified one class of phonemes: for example, "I'm naming one man among many" was one of those exemplifying nasal phonemes, while "Dad would buy a big dog" was one of those exemplifying voiced stops. Each speaker produced 774, 763 and 739 syllables respectively, the total therefore being 2276 syllables.

2.1 *Hand transcriptions*

The segmentation and labelling of the data had been carried out by trained phoneticians at the Centre for Speech Technology Research, Edinburgh. The labelling was assigned at the phonemic level, with no allophonic information. Schwa vowels were labelled as such where they occurred, and consonants deleted by the speaker were not subsequently added by the transcriber, but no distinction was made between the various positional allophones of consonants. Syllable-initial consonants were included in the relevant syllable, as in the work carried out by Crystal & House (1990).

Linguistic features in the form of diacritics were added to the phoneme symbols by the first author. These features were: prepausal (usually signalling end of utterance); tense (for diphthongs and long monophthongs); lexically stressed (for syllables bearing the primary lexical stress of an orthographic word, which could include 'the' or 'a'); and accented (for the syllables heard as manifesting a pitch accent). The initial segment of each syllable was likewise determined by hand.

2.2 *Formation of the data matrix*

The transcribed database then formed the input to an algorithm which calculated the duration of feet over the entire database. In each case, feet were taken to begin at the onset of the first consonant belonging to the relevant syllable, assuming a maximal syllable onset principle while respecting the phonotactics of the syllable. Six versions of the algorithm were used, corresponding to the six different foot types, as follows.

Random feet were constructed by choosing syllables at random to form the first syllable of such feet. This was done by using a random number generator to yield integers between one and five. Each integer was then used as the number of syllables contained in the current foot. Five was chosen as a practical upper limit, since previous work suggests that few stress-based feet in English exceed this length.

Arbitrary feet were constructed by beginning a new foot every *n* syllables from the current syllable, beginning at the *m*th syllable of the utterance, where *n* and *m* varied systematically between 1 and 5, as exemplified below.

10) Starting at the first syllable of the utterance, making a new foot every 2 syllables.
11) Starting at the second syllable of the utterance, making 4-syllable feet.
12) Starting at the fifth syllable of the utterance, making 5-syllable feet.

These permutations of *n* and *m* were chosen to cover all possibilities within the set limits of 1 to 5 syllables per foot. No redundant permutation was used: for example, the algorithm did not begin at the fourth syllable of the utterance, making 3-syllable feet, as the feet thus formed would already have been covered under the permutation starting at the first syllable of the utterance and making three-syllable feet.

Tense feet were constructed by beginning a new foot at every syllable containing a tense vowel (ie. a long monophthong or a diphthong). A few utterances did not contain any tense vowels and so were not represented in this sub-case of the experiment.

Full feet were constructed by beginning a new foot at every syllable containing a full vowel. A full vowel was taken to be any syllable nucleus except schwa, syllabic consonant, morpheme-final unstressed /ii/ or /ou/ (eg. in *'city"*, *'willow"*), the /i/ of suffix '*-ing"*, or the vowels of some stressless function words such as *'we"*, *'you"*. Many vowels were 'full' in this sense, and so shorter feet predominated in this sub-case of the experiment.

Lexically stressed feet were constructed by beginning a new foot at every syllable that formed the primary lexical stress of an orthographic word. This syllable formed the potential for stress location, and may not have been prominent in any given utterance. Even the lexically-stressed syllables of function words were included (eg. *'the"*, *'we"*), in an attempt to obtain data that was less dependent on phonological criteria, in order to assess the significance of any increase in isochrony that might be displayed by more phonologically-defined feet.

Accented feet were constructed by beginning a new foot at every syllable that displayed an easily perceivable pitch peak, defined auditorily. Given the nature of the data (carefully-read short sentences) accented syllables were not difficult to locate.

For each foot type, a data matrix was formed, with one row per foot constructed, and four columns. The four columns gave values for the following four variables:

13) Duration of foot in msec.
14) Number of syllables in the foot.
15) Mean syllable duration in the foot (i.e. 13 above divided by 14 above).
16) Prepausal (1) or non-prepausal (0) foot.

In the subsequent statistical analyses, prepausal (utterance-final) feet were discarded, and only non-prepausal feet were analysed. This was in order to avoid any bias in the results introduced by utterance-final lengthening of the foot.

2.3 Observed foot durations

Table I shows the mean foot durations and mean number of syllables per foot observed, by speaker and foot type, for five foot types. The durations are generally shorter than those observed by Fant *et al* (1991), who found a mean duration of 565 msec. for English. This is probably because the mean number of syllables per foot is smaller in this data, as seen in Table I. In another paper, Fant *et al*. (1989) report a mean for English of 2.8 syllables per foot, using a read connected passage rather than the isolated sentences currently under discussion.

Fig. 1 shows the mean foot durations observed for speaker JL (whose duration values were intermediate between the other two speakers). The four 'linguistic' feet, plus the 'random' feet are shown. The values in each case are shown with an arrow giving one standard deviation above and below the mean. The final small graph compares the values for all five foot types directly. It can be seen that there is no levelling-off in the curve in the case of the 'random' feet, and thus, it would seem, no 'foot compression effect'. In order to examine the size of any foot compression effect in the four 'linguistic' feet, however, it is necessary to undertake a detailed statistical analysis, as described below.

3 Statistical analyses

The data matrix constructed for each foot type, for each of three speakers (JL, GW and PJ) then formed the input to a statistical analysis. The independent variable in each case was the number of syllables in the foot. The dependent variable was either foot duration (section 3.1) or averaged syllable duration (section 3.2), according to the two tests for isochrony described in section 1.3 above.

3.1 *Foot duration*

3.1.1 *Linear regression*

Using the same approach as Fant *et al* (1991), a linear regression of foot duration on number of syllables per foot was carried out for each speaker, for each of the six foot types. Foot duration (*y*) was a function of the y-axis intercept (*a*) and a coefficient *b* multiplying the number of syllables per foot (*x*), in the following equation:

(17) $y = a + bx$

3.1.2 *Results*

For each regression, the overall slope was positive, and the F value and r-squared value were high, indicating a highly significant model that explains a great deal of the variance in the dependent variable (foot duration). Results are shown in Table II by foot type. For each speaker and each foot type, the *a* and *b* values of the equation in (17) are given, together with the overall r-squared value for the regression model, and also the significance (i.e. not significant, significant to $p < 0.05$ (single asterisk), or significant to $p < 0.02$ (double asterisk), using a two-tailed test in each case). For the cateory of arbitrary feet, all feet were included in the model, rather than concentrating only on feet beginning at, for example, the third syllable of the utterance.

3.1.3 Discussion

It can be seen from Table II that feet increase with the number of syllables per foot, for all foot types. The random and arbitrary feet show generally higher r-squared values, indicating that this factor (number of syllables) accounts for a higher proportion of the variance than in the other foot types. In other words, in the 'linguistic' foot types there may be some other factor also at work in the determination of foot duration.

3.2 *Averaged syllable duration*

3.2.1 *Linear regression*

A second regression was then carried out, this time regressing averaged syllable duration on number of syllables in the foot. It should be borne in mind that these are not true syllable durations, being a mean duration of syllables within a given foot. However, these mean durations will suffice to build up a rank ordering, though not a quantification, of syllable durational behaviour across the different foot types.

The crucial measure for establishing isochrony in the case of syllable duration is the presence of an overall negative slope for the regression curve (ie. the '*b*' term in (17) above), such that syllable duration is inversely proportional to the number of syllables in the foot. This would indicate that average syllable duration became shorter as the number of syllables per foot increased.

A least-squares regression of mean syllable duration on number of syllables per foot was carried out in the same way as before, using only non-prepausal feet. Also as before, all types of arbitrary foot were merged into one large dataset for arbitrary feet in general.

3.2.2 *Results*

The results are shown in Table III in the same format as Table II, where 'ns' indicates 'not significant', '*' indicates $p < 0.05$, and '**' indicates $p < 0.02$ (two-tailed tests)..

It can be seen from Table III that shortening of averaged syllable duration appears to occur in syllables from accent-defined feet, accounting for a significant minority of the variance in syllable duration. It also occurs, though to a lesser extent, in syllables from feet defined by full syllables. The results are less clear in the case of syllables from 'tense' or 'lexically-stressed' feet. In the case of syllables from random or arbitrary feet, however, there appear to be few cases of a significant syllable shortening effect, and in any case the magnitude of any such effect is negligible and so can be discounted.

3.2.3 Discussion

There appears to be a syllable shortening effect greater than chance in the case of the four linguistic feet, since these show a higher negative *b* coefficient, and a larger r-squared than the 'random' feet. Since the four linguistic feet show a greater syllable shortening effect than the arbitrary feet, it can be concluded that the cause of syllable shortening is linguistic rather than arbitrary. In other words, both statements (4) and (5) above are borne out, with support for statement (6) if 'stressed' is interpreted as 'accented'.

However, it should be borne in mind that these results are obtained on the basis of syllable durations that were averaged across each foot. Therefore, it could be that the syllable shortening effect is merely an artefact of the fact that linguistically-defined foot-initial syllables will consistently be longer than non-foot-initial syllables (whether defined in terms of 'full', 'lexical', 'accented' or 'tense'). This is because these initial syllables tend to have a larger number of consonants, a higher proportion of phonologically long vowels, a lower proportion of (or zero) especially short vowels (eg. schwa), and also stress-induced lengthening of segments. Thus, the fewer the number of syllables included in the foot-wide averaging, the larger the mean duration will appear to be. As Fletcher (1991) observes, "[t]he effects of including a longer accented syllable in the calculation of mean syllable duration become less and less pronounced as units become longer". Therefore, it is necessary to examine possible foot-based syllable shortening on the basis of actual observed syllable durations. This was carried out in the next experiment.

3.3 Actual syllable duration

3.3.1 Number of segments per syllable

Actual syllable durations, together with the corresponding number of syllables in the foot, were obtained for the three speakers (in this case, foot type was not relevant, as all types of syllables were being considered). As before, only non-prepausal feet were considered. A linear regression was carried out, where the dependent variable was syllable duration and the independent variable was number of segments per syllable. For speaker JL, the constant '*a*' term (the intercept on the *y* axis) was -13 (18 for GW, 6 for PJ), while the '*b*' coefficient was 100 (90 for GW, 84 for PJ), all significant at $p < 0.02$. For JL, the overall r-squared value was 0.58 (0.51 for GW, 0.50 for PJ), which is fairly high. A one-way analysis of variance was also carried out, where the dependent variable was syllable duration, and the independent variable was the number of segments per syllable. The resulting F values were very large, and were significant to $p < 0.02$ for all three speakers (F=297, 3 and 669 d.f. for JL, F=224, 3 and 668 d.f. for GW, F=174, 4 and 672 d.f. for PJ).

Clearly the number of segments in a syllable is an important determinant of the syllable's duration, but is not the only factor. The patterning factor (full vowel identity, or accent, etc.) might supply another determinant. Thus another test was performed to check this hypothesis.

3.3.2 Target versus non-target factor

Syllables had been tagged according to whether or not they were target syllables, where a 'target' syllable was, as appropriate, 'full', 'lexically-stressed', 'accented', 'tense', or 'randomly-chosen'. A one-way analysis of variance was performed for each speaker and target type, where the dependent variable was syllable duration, and the independent variable was target category (target/non-target). The F values were significant for all speakers and target types at $p < 0.02$, indicating that there were significant differences in variance between the two groups.

In order to focus on the origin of these differences, t-tests were performed on the durations of target and non-target syllables, by speaker and target type. The results are shown in Table IV, which gives the level of significance of the t value, the mean duration of target syllables in msec., the mean duration of non-target syllables in msec., and the difference between these two values in msec. It can be seen from Table IV that target syllables were significantly longer in duration than non-target syllables for all four of the 'linguistic' target types. The difference was largest in the case of the 'full' syllables. For random targets, there was either no significant difference, or a less significant and very small difference (for speaker JL).

The conclusion is that there is no difference in mean syllable durations in the randomly-chosen case, while there is a significant difference (to varying degrees) in the four linguistic cases, of which the best cases are generally 'full' and 'accented'. An obvious cause of these differences is the number of segments per syllable, the mean values of which are shown in Table V by speaker and syllable type (target or non-target), together with the difference between the means in msec. It is clear from Table V that the 'full' case shows the greatest difference between target and non-target syllables as regards number of segments per syllable, closely followed by the 'accented' case. The 'random' case, on the other hand, shows a negligible difference. This pattern corresponds to that seen in Table IV for mean syllable durations, which are thereby explained as being mainly due to the number of segments per syllable.

Not only number of segments, but also the type of those segments, could be a determinant of syllable duration. This is indicated by Table VI, which shows the mean durations and number of observations for syllables containing two segments (in order to control for number of segments per syllable). It is clear that there is no consistent difference in duration between target and non-target two-segment syllables in the case where syllables are chosen at random. In the four linguistic cases, however, target syllables are consistently longer in duration than non-target syllables. This is probably due to reasons of vowel quality: that is, tense vowels are longer in duration than non-tense vowels, and non-full vowels (schwa, etc) are shorter in duration than full vowels. There is also a frequency effect: that is, accented syllables are likely to contain a higher proportion of tense (and therefore longer) vowels than are unaccented syllables, and lexically-stressed syllables are likely to contain a lower proportion of reduced (and therefore shorter) vowels than are non-lexically stressed syllables. Thus, both the number of segments per syllable, and the target/non-target status of the syllable, are determinants of syllable duration. In order to test, for example, for an accent-induced lengthening effect on syllables, it would be necessary to include information about segment identity in order to control for factors such as vowel quality. There was insufficient data for this to be carried out in the present investigation.

3.3.3 Number of syllables per foot, for two-segment syllables

To control for number of segments per syllable, syllables of two segments were chosen for further testing. This sub-group (two-segment syllables) was the most numerous of all sub-groups by segment number, and was chosen because other sub-groups did not contain sufficient observations to enable conclusions to be drawn. It was not possible to divide the sub-group of two-segment syllables between target and non-target syllables, as the groups thus formed were not large enough for conclusions to be drawn. It was also not possible to differentiate between cases of CV and VC syllable structure for these syllables. However, certain results were obtained despite these limitations, as seen below.

A linear regression was carried out for two-segment syllables, by speaker and syllable type, where the dependent variable was syllable duration and the independent variable was the number of syllables in the given syllable's foot. The results are shown in Table VII, which gives the intercept ('*a*' value), the coefficient ('*b*' value), the r-squared value and overall significance of the regression.

Table VII shows that there is a small but significant tendency for the duration of two-segment syllables to decrease with increasing number of syllables in the foot (ie. a negative '*b*' coefficient), for all linguistic foot types and especially for the 'full' and 'lexical' feet. The 'random' feet, on the other hand, showed no such tendency at all. It is true that the r-squared values are small (the best case, for lexical feet for speaker PJ, explains no more than 5% of the variance in syllable duration). However, the effect is significant to $p < 0.02$ in most cases, and is clearly not present at all in 'random' feet (the control case). Thus a small tendency to true isochrony in production has been demonstrated, at least in the case of two-segment syllables. While this tendency is greater than chance, it is unclear which patterning principle ('full vowel' or 'lexically stressed syllable') is the one controlling it. Thus, although statements (4) and (5) above have been borne out in the statistical analyses, it is not possible, given this data, to make a definitive judgement of the truth of statement (6) above.

3.3.4 Number of syllables per foot, for target syllables

To control for the factor target/non-target syllable, syllables were divided between these two groups, and a linear regression of syllable duration on number of syllables in the relevant foot was carried out. The results for non-target syllables (non-full, non-lexically-stressed, non-accented, non-tense, non-randomly-chosen, as appropriate) were not significant and showed no consistent pattern across speakers. The results for target syllables are shown in Table VIII. A further regression was carried out on target syllables from feet greater than one syllable in length. The results are shown in Table IX in the same format as before.

Table VIII, for all target syllables, shows a small syllable shortening effect in the case of 'full' and 'accented' feet. That is, for these foot types, it would appear that the duration of syllables decreases with increasing number of syllables per foot. However, the results from Table IX, for feet with two or more syllables, show a different pattern. Here, the small syllable shortening effect is seen in the case of 'lexically-stressed' and 'accented' feet, for two speakers only. This change in pattern suggests that a large part of the appearent syllable shortening effect in the case of 'full' feet derives from the lengthening of a given syllable when it forms a monosyllabic foot, as compared to cases where it forms the initial syllable of a polysyllabic foot. Thus the syllable shortening effect seen for 'lexically-stressed' and 'accented' feet in Table IX can be said to indicate more reliably a small amount of isochrony in the case of these foot types.

4 Discussion

4.1 Summary

From the data given in Fig. 1, and in Tables II and III, it can be concluded that there is no foot compression effect in random feet, where 'foot compression' refers to the tendency for the growth in foot duration to level off with increasing number of syllables per foot. Equivalently, there is no tendency for averaged syllable duration to decrease with increasing foot size in the case of random feet (Table III). It also appears, from Tables II and III, that the same statements are true of arbitrary feet, and so the foot compression effect found in the other four foot types may be said to be non-random and also linguistic rather than arbitrary.

This is far from proving isochrony in 'linguistic' feet, however. This is because of the difference in mean duration between foot-initial and non-foot-initial syllables in the case of the four linguistic foot types, as seen in Table IV. The difference (always in favour of the foot-initial syllable) means that there is a spurious shortening of averaged syllable durations as the number of syllables in the feet increases. Thus investigations must be made on the basis of actual syllable durations.

A major determinant of actual syllable durations is the number of segments per syllable (section 3.3.1), as well as the target/non-target status of the syllable (section 3.3.2) in the case of syllables chosen on a linguistic basis. If the number of segments in the syllable is held constant at two (section 3.3.3) then a small but significant foot-based syllable compression effect can be demonstrated in production, ie. true isochrony (Table VII). Similarly, if only target syllables from polysyllabic feet are considered (section 3.3.4) then a foot-based syllable compression effect is likewise found (Table IX).

4.2 Implications

Whether this size of effect is above the threshold of perception is an important question. The largest value of syllable compression (for lexical feet for speaker PJ), gives a decrease of 39 msec. for each added syllable in the foot, while the corresponding value for speaker GW is 29 msec. (Table IX). The mean duration of unstressed syllables in lexical feet for speaker PJ is 137 msec., and that for speaker GW is 173 (Table IV). The difference limen for the perception of changes in duration, given a reference sound of 110 msec., is 21.56 msec., while the difference limen given a reference sound of 175 msec. is 32.90 msec. (Lehiste 1970). For speaker PJ, the syllable compression effect is therefore above the just-noticeable difference in duration for speech sounds of this order, while for speaker GW it is marginally below the threshold of perception. Thus the perceptual validity of this syllable compression effect is open to doubt, at least in the case of speaker GW.

What is not open to doubt, however, is that the size of this isochronous effect on syllable durations is dwarfed by the factors of number of segments per syllable, and syllable identity (as accented or not accented, etc.). The perceptual salience of syllable shortening, examined above, is probably outweighed by linguistic factors such as the tendency for accented syllables to have a greater number of consonants than unaccented syllables, or the tendency for lexically stressed syllables to have more intrinsically long vowels ('tense' vowels) than non-lexically stressed syllables. Dauer (1983) concludes that it is factors such as these that lead to the perceptual impression of rhythmic regularity in languages such as English, noting that "[t]he hypothesis by Nakatani *et al* (1981, p.103) that 'isochrony should show up best in reiterant speech because there is no perturbation of the underlying rhythm by segmental variation' seems to be backwards. It is precisely the language structure with all its language specific segmental variation that is responsible for perceived differences in language rhythm".

The work reported in the present paper has shown that 'foot compression' in English (the levelling off of growth in foot duration with increasing number of syllables) is neither random nor arbitrary, but is linked to linguistic factors such as accentedness of the foot-initial syllable. The same is true of the small foot-dependent syllable shortening effect (using actual syllable durations) found in 'linguistic' feet but not in 'random' feet. The conclusion may not be particularly unexpected, but it is reassuring that this control experiment has confirmed the approach taken in previous studies of stress in English: ie. that the patterning factor for rhythmic units is non-random and is related to a linguistic variable such as accent. The precise nature and cause of rhythm, and the way in which to measure a language's degree of rhythmicality, still await further research. However, it can now be stated that it is indeed worth attempting to measure rhythmicality, as this has been shown not to be random.

References

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

Buxton, H. (1983) Temporal Predictability in the Perception of English Speech. In A. Cutler & D.R. Ladd (eds.), *Prosody: Models and Measurements*. Berlin: Springer-Verlag.

Campbell, W.N. (1988) Foot-level shortening in the Spoken English Corpus. *Proceedings of the 7th FASE Symposium*, September 1988, Edinburgh, UK.

Crystal, T.H. & House, A.S. (1990) Articulation rate and the duration of syllables and stress groups in connected speech. *Journal of the Acoustical Society of America*, **88** (1): 101-112.

Dauer, R.M. (1983) Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, **11**: 51-62.

Fant, G., Kruckenberg, A. & Nord, L. (1989) Rhythmical structures in text reading: A language contrasting study. *Proceedings of the European Conference on Speech Communication and Technology* (vol. 1):498-501, September 1989, Paris, France.

Fant, G., Kruckenberg, A. & Nord, L. (1991) Durational correlates of stress in Swedish, French and English. *Journal of Phonetics*, **19**: 351-365.

Fletcher, J. (1991) Rhythm and final lengthening in French. *Journal of Phonetics*, **19**: 193-212.

Halliday, M.A.K. (1967) *Intonation and Grammar in British English*. The Hague: Mouton.

Lehiste, I. (1970) *Suprasegmentals*. Cambridge, Mass. and London, England: MIT Press.

Lehiste, I. (1977) Isochrony reconsidered. *Journal of Phonetics*, **5**: 253-263.

Nakatani, L.H., O'Connor, K.D. & Aston, C.H. (1981) Prosodic Aspects of American English Speech Rhythm. *Phonetica*, 38: 84-106.

Pike, K.L. (1946) *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Roach, P. (1982) On the distinction between `stress-timed' and `syllable-timed' languages. In: D. Crystal (ed.), *Linguistic Controversies*. London: Edward Arnold.

Wenk, B.J. & Wioland, F. (1982) Is French really syllable-timed? *Journal of Phonetics*, **10**: 193-216.

Acknowledgements

Thanks are due to Mary Beckman for her incisive and illuminating comments, and also to Janet Fletcher, for her interest and for helpful discussions. The paper has benefitted greatly from the comments of four sets of anonymous reviewers.

Table I: Observed mean foot durations in msec. and mean number of syllables per foot.

Speaker:	JL		GW		PJ	
Foot type:	Duration	Syll. no.	Duration	Syll. no.	Duration	Syll. no.
Full	372	1.77	396	1.79	343	1.77
Lexical	271	1.30	283	1.31	250	1.32
Accented	607	2.84	654	2.94	633	3.24
Tense	472	2.17	494	2.18	439	2.24
Random	536	2.57	561	2.65	516	2.74

Table **II**: Regression of foot duration on no. of syllables per foot (* = $p < 0.05$, ** = $p < 0.02$)

Speaker:	JL				GW				PJ			
Foot type	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig
Full	133	135	0.56	**	141	142	0.62	**	128	122	0.56	**
Lexical	79	147	0.36	**	77	157	0.39	**	89	123	0.32	**
Accented	176	151	0.69	**	192	157	0.76	**	199	134	0.70	**
Tense	72	176	0.55	**	79	190	0.87	**	92	155	0.85	**
Random	24	199	0.76	**	-15	218	0.83	**	-7	191	0.79	**
Arbitrary	11	204	0.81	**	9	213	0.84	**	16	186	0.82	**

Table III: Regression of averaged syllable duration on no. of syllables per foot (ns = not significant, * = p < 0.05, ** = p < 0.02)

Speaker:	JL				GW				PJ			
Foot type	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig
Full	302	-42	0.19	**	317	-43	0.24	**	282	-41	0.22	**
Lexical	256	-31	0.03	**	263	-31	0.03	**	245	-36	0.05	**
Accented	304	-26	0.26	**	319	-27	0.37	**	276	-21	0.30	**
Tense	256	-14	0.03	*	270	-13	0.08	**	242	-13	0.11	**
Random	223	-5	0.01	ns	192	6	0.02	ns	177	3	0.01	ns
Arbitrary	214	-2	0.00	*	222	-2	0.00	ns	201	-3	0.00	**

Table IV: Results of t-tests on actual syllable durations of target/non-target syllables, by speaker and target type (ns = not significant, * = p < 0.05, ** = p < 0.01, tar = mean duration of target syllable in msec., non-tar = mean duration of non-target syllable in msec., diff = difference between these two values in msec.).

Speaker:	JL				GW				PJ			
Target type	sig	tar	non-tar	diff	sig	tar	non-tar	diff	sig	tar	non-tar	diff
Full	**	257	143	114	**	269	152	117	**	239	130	109
Lexical	**	223	156	67	**	229	173	56	**	207	140	67
Accented	**	271	174	97	**	280	183	97	**	250	164	86
Tense	**	266	178	88	**	276	184	92	**	247	162	85
Random	*	217	200	17	ns	213	219	-6	ns	191	192	-1

Table V: Mean number of segments per syllable by speaker and syllable type, together with difference between the means (tar = target syllables, non-tar = non-target syllables, diff = difference).

Speaker:	JL			GW			PJ		
Target type	tar	non-tar	diff	tar	non-tar	diff	tar	non-tar	diff
Full	2.40	1.92	0.48	2.41	1.95	0.46	2.41	1.94	0.47
Lexical	2.26	2.00	0.26	2.25	2.05	0.20	2.25	2.04	0.21
Accented	2.49	2.05	0.44	2.46	2.07	0.39	2.46	2.08	0.38
Tense	2.33	2.14	0.19	2.30	2.15	0.15	2.28	2.16	0.12
Random	2.24	2.16	0.08	2.17	2.23	-0.06	2.20	2.20	0.00

Table VI: Mean syllable durations in msec. for syllables of two segments by speaker and syllable type (tar = target, non-tar = non-target, n = no. of observations).

Speaker:	JL				GW				PJ			
Syll. type	tar	n	non-tar	n	tar	n	non-tar	n	tar	n	non-tar	n
Full	222	(212)	148	(173)	236	(205)	156	(167)	214	(205)	136	(169)
Lexical	201	(280)	155	(102)	210	(282)	170	(90)	193	(280)	137	(94)
Accented	223	(134)	170	(248)	239	(131)	179	(241)	212	(123)	162	(251)
Tense	238	(145)	159	(237)	255	(147)	164	(225)	230	(149)	145	(225)
Random	196	(149)	184	(236)	197	(143)	203	(229)	179	(139)	178	(235)

Table VII: Results of linear regression of syllables per foot on actual syllable durations for two-segment syllables ('a' = intercept, 'b' = coefficient of x, ns = not significant, * = p < 0.05, ** = p < 0.02, two-tailed tests).

Speaker:	JL				GW				PJ			
Foot type	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig
Full	210	-10	0.03	**	220	-9	0.02	**	205	-12	0.03	**
Lexical	210	-12	0.02	**	226	-15	0.02	**	213	-20	0.05	**
Accented	197	-3	0.01	ns	212	-15	0.02	ns	194	-4	0.02	**
Tense	198	-3	0.02	*	209	-3	0.01	*	197	-5	0.05	**
Random	197	-3	0.00	ns	191	3	0.00	ns	178	0	0.00	ns

Table VIII: Results of linear regression of syllables per foot on actual syllable durations for (all) target syllables ('a' = intercept, 'b' = coefficient of x, ns = not significant, * = p < 0.05, ** = p < 0.02, two-tailed tests).

Speaker:	JL				GW				PJ			
Foot type	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig
Full	280	-13	0.01	*	300	-17	0.03	**	260	-11	0.01	*
Lexical	233	-7	0.00	ns	246	-12	0.01	ns	221	-10	0.00	ns
Accented	296	-9	0.02	(0.10)	337	-20	0.09	**	283	-11	0.04	**
Tense	264	1	0.00	ns	271	2	0.00	ns	252	-2	0.00	ns
Random	209	3	0.00	ns	195	7	0.01	ns	178	5	0.01	ns

Table IX: Results of linear regression of syllables per foot on actual syllable durations for target syllables from feet of more than one syllable in length ('a' = intercept, 'b' = coefficient of x, ns = not significant, * = p < 0.05, ** = p < 0.02, two-tailed tests).

Speaker:	JL				GW				PJ			
Foot type	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig	*a*	*b*	r^2	sig
Full	253	-2	0.00	ns	282	-10	0.01	ns	232	-1	0.00	ns
Lexical	239	-10	0.00	ns	284	-29	0.03	*	284	-39	0.04	**
Accented	270	-2	0.00	ns	325	-17	0.05	**	277	-9	0.02	*
Tense	260	2	0.00	ns	274	1	0.00	ns	261	-4	0.01	ns
Random	198	6	0.00	ns	220	0	0.00	ns	189	2	0.00	ns