Intonation, Accent and Rhythm

Studies in Discourse Phonology

Edited by
Dafydd Gibbon and Helmut Richter

Sonderdruck

Walter de Gruyter · Berlin · New York
1984

W. JASSEM, D. R. HILL, I. H. WITTEN

Isochrony in English Speech: its Statistical Validity and Linguistic Relevance

1. Introduction

In her review of recent work on rhythm in English speech, Lehiste (1977) shows that there has been disappointingly little agreement among specialists as to the validity of the isochrony principle, ranging from fairly strong textbook affirmation, such as the oft-cited description by Pike (1945: 34) or a more up-to-date formulation by Ladefoged (1975: 102-103) through doubt and scepticism (e.g. O'Connor, 1965; Bolinger, 1965; Uldall, 1971) to downright rejection (e. g. Lea, 1974; Shen and Paterson, 1962). Her own experiments lead Lehiste (1973; 1975; 1977) to the conclusion that "... there were some aspects of the data that spoke for the presence of isochrony, and other aspects that spoke against it" (1977: 256). According to Lehiste, isochrony is most evident at the perceptual level, but seeing that the sensation of rhythmicality must reflect some properties of the signal, she also made measurements of the sound wave and produced evidence in favour of a certain measure of isochrony, which she related to syntax. The relation is an inverse one in the sense that if the (tendency for) isochrony is destroyed by "an increase in the interstress interval", then this is a sign of the presence of a syntactic boundary. One might perhaps re-interpret Lehiste's conclusions in a direct sense as assigning to isochrony a function of internal syntactic cohesion. One of the major merits of the extensive treatise on the rhythm of spoken English by Adams (1979) is a most exhaustive historical survey of previous work in the area, with laudable emphasis on the widely overlooked ideas of the early phoneticians. It transpires that much of the present-day thinking and argumentation on the rhythmicality of spoken English was anticipated in the preceding centuries, beginning with John Hart and Joshua Steele, the latter having strongly influenced Abercrombie's (1964; 1973) theory, which will figure prominently further on in this paper. Also, in direct relation to one particular problem considered here, John Hart (Adams, 1973: 22-23) related 'accent' (equivalent to the early 20th century 'stress') to 'melody and rhythmus' rather than loudness and he was aware of rhythmical (accentual) units of speech such as *of an apple*, *from the Citie*, "rediscovered" 30 years ago by one of the present writers (Jassem, 1949; 1952).

It is also noteworthy that a difference of opinion as to whether pauses in speech should be counted in the rhythmical structure of English can be traced to the late 18th and early 19th century, J. Steele supporting the pause and J. Odell opposing it (Adams, 1979: 29-30).

2. The Theory

The general problem is so well-known that it will scarcely be necessary to expand on it. Hardly any present-day textbook of English phonetics (or phonology) fails to mention rhythmicality as reflected in the (approximate) isochrony 'interstress intervals'. Some of the major, more specific, questions, are:

1. What are the relations between rhythmicality (and isochrony) in the acoustic signal and rhythmicality as a percept?

2. What are the appropriate measurement procedures to be applied to the speech wave which would test the hypothesis of rhythmicality?

3. What is the experimental design of appropriate perceptual tests?

4. How are the results of measurements and perceptual tests to be evaluated?

5. If the 'isochrony effect' is present, does it primarily affect the length of the entire syllables, or the duration of the constituent segments (phones)?

6. If isochrony is primarily reflected in the length of entire syllables, does it affect all the syllables of the rhythm unit ('bar', 'foot', or whatever) or only the accented ('stressed') syllables?

7. If isochrony is primarily reflected in the duration of the phonetic segments (phones), does it affect all the segments, or some of them only, and how?

8. Is rhythmicality equally effective in all styles of speech, or is it a dependent variable, being, for instance, more evident in verse reading, less so in prose reading, slow and deliberate speech, and perhaps least evident in fast, casual discourse?

9. Is rhythmicality an effect of accent, or is accent the effect of rhythmicality?

9a. If accent is the cause rather than the effect of rhythm, then which phonetic attributes are relevant for accentuation?[1]

10. If there is an isochrony effect, then (a) how can it be quantified, (b) how can the quantitative statement be used to describe the strength of the effect, and (c) how strong has the effect to be in order to be treated as phonologically relevant?

---

[1] 9a is only included here because the question has often been begged. But the answer is in fact a crucial premise in the theory of isochrony and rhythm. This has most fully been realized by Adams (1979), see especially her introduction and Chapter 6.

11. What are the relations, if any, between the rhythm units and syntax or morphology?

The investigation reported on below proposes a method which may give partial answers to points 2 and 4. It also assumes a hypothesis related to point 7 and tests it statistically, leading to a possible answer to 10 a and b. Reference will also be made to point 11.

3. Two Specific Theories of English Speech Rhythm

Two specific theories of the rhythm of spoken English have been proposed. The one, which will here be referred to as (A), was first put forward by Abercrombie (1964, 1973), and the other, referred to as (B), by Jassem (1952), (slightly modified in Jassem, 1980 and 1981).[2]

Apart from the fairly obvious postulate that both theories submit, viz. one of a tendency towards equality of interstress intervals, they have one important premise in common. They do not start off with any higher-order *syntactic* units of which the rhythm units would be constituents. In fact, the 'beats', or 'bars', or 'feet', etc. of either theory, though possibly correlated with syntactic entities, are *independent* of them.[3]

Abercrombie's theory was further developed by Halliday (1970) and Witten (1977). The two theories may be summarized as follows:

(A) Abercrombie

1. The rhythm unit, called FOOT, always begins with a stressed syllable, consequently any unstressed syllable follows a stressed one within the same Foot. All unstressed syllables may therefore be described as postaccentual (or postictic).

2. If any utterance begins with an unstressed syllable, a silent stress is posited, this being an abstraction manifested as zero sound, i.e. not materialized objectively, though real psychologically (subjectively).

3. A disyllabic foot is triple-timed and may be represented by one of the following structures: ∪ — (short-long), ∩ ∩ (medium-medium), or

[2] Jassem 1980 and 1981 are recent editions of earlier works first published in 1954 and 1962 respectively. The modifications of the original 1952 version of (B) contained in these works were proposed long before the results of the present investigation were available, though they are largely borne out by it.

[3] In this, as in other aspects, the position of both authors are drastically different from those assumed by the proponents of any variety of Generative Phonology. But it may be interesting to note that Jassem's 'rhythm units' seem to coincide with Chomsky and Halle's 'phonological words'. The locution (utterance, tone-group, sentence, etc.) *The book was in an unlikely place* is analysed into three 'phonological words' (Chomsky and Halle 1968: 367-368): *the-book was-in-an-unlikely place*. Exactly the same division is obtained by applying the principles expounded in Jassem 1952. Abercrombie's interpretation would be entirely different: *the* | *book-was-in-an-un-* | *likely* | *place*.

— ∪ (long-short). The original version of the theory did not discuss feet of more than two syllables, and this part of the theory was supplied by Witten (1977). This mora-based structure of rhythm was not insisted on by Halliday (1970).

4. The internal rhythmic structure of a Foot is inherently related to its segmental structure, e.g. (C)V^1CV(C)[4] and (C)V^2(C)V(C) both produce ∩ ∩, etc. Abercrombie adds, however, that "the phonematic structure of the syllable may ... at times be quite irrelevant" (1964: 217).

5. There are certain relations between rhythm and syntax. For instance, the quantities depend on the presence of a word boundary" (ibid, p. 219). There is a rhythmic difference between a one-syllable word followed by an unstressed syllable of a word which is not directly related syntactically, e. g., *(take) Grey to (London)* as opposed to *Greater (London)*—and a word followed by an enclitic, e. g. *take it, tell him*. It is pointed out that enclitic treatment of monosyllables is "not entirely clear" (p. 221), some other cases of "rhythmic linking" being *piece of, may there (be)*, etc.[5]

6. Stress is assumed (see, e. g., Abercrombie, 1967: 35; Ladefoged, 1975: 222) to be increased effort (or energy). The notion of stress is primary in relation to the notion of rhythm.

(B) Jassem

1. English speech consists of two kinds of rhythm units: (a) Narrow Rhythm Units (NRU) and (b) Anacruses (ANA). For a given tempo, the length of a narrow rhythm unit depends on the number of syllables. This length is a constant for a monosyllabic rhythm unit and a given tempo, and may be denoted by Y As the number of syllables in a narrow rhythm unit increases, the length of the narrow rhythm unit (NRU) also increases, but *not proportionately*. A two-syllable NRU is longer than a monosyllabic one, but it is *distinctly less* than 2Y. A three-syllable NRU is also longer than a two-syllable one, but its length is *significantly less than* 3Y The length of longer rhythm units is determined analogously.

2. Individual syllables within a multisyllable NRU tend to be of equal length, i.e., the complete length of a polysyllabic NRU tends to be somewhat equally divided among the constituent syllables.[6]

---

[4] V^1 - short vowel, V^2 = long vowel or diphthong, C = consonant.

[5] As the phonematic structure of the Foot is only sometimes relevant to its internal rhythmical structure, and the relations between syntax and rhythm are not always clear, it is not possible, within the framework of theory (A), to deduce the rhythm of an utterance from its phonemic transcription. Nor does it seem to be possible to make simple additions to a transcription so as to indicate the internal rhythmical structure of the Foot.

[6] It follows from (I) and (2) that the relative lengths of NRUs and their constituent syllables may be graphically represented like this:

3. 'Stress', which is now termed ACCENT (Jassem and Gibbon 1980), is the *effect* of the temporal organization of utterances. A complete monosyllabic utterance is accented by definition. It is the tendency described above under (2) that is the basis of accent: *The only or the first syllable of a Narrow Rhythm Unit is accented.*

4. The duration of phones is necessarily affected by variations in the length of syllables. Thus /i/ is longer in *read* than in *reading* and this, in turn, is slightly longer than /i/ in *reading it*, at least in fairly slow speech. It is not yet known how and to what extent rhythm affects the individual phonemes or phoneme types.

5. Besides the NRUs discussed above, an utterance may include a syllable, or a sequence of syllables, which is characterized by being as short as possible, i.e. as short as is compatible with sufficiently distinct articulation of the constituent phones. Such a syllable, or syllables, constitute the ANACRUSIS (ANA). The length of an ANA, consequently, tends to be proportionate to the number of the constituent syllables, or perhaps more directly, the number of constituent phones. The ANA, if present in an utterance, always *precedes* an NRU and belongs to that NRU. An NRU, together with a preceding ANA (if any) forms a TOTAL RHYTHM UNIT (TRU).

6. The rhythm of English speech is a *phonetic* phenomenon and is determined on purely phonetic principles with no recourse to any other level of analysis-synthesis, such as grammar or semantics. But there are interre-

---

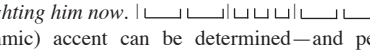
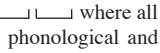
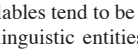
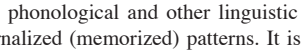
1 syllable

2 syllables

3 syllables 

4 syllables

etc.

[7] It is assumed that accent is perceived in longer utterances due to the durational variation of syllables, e. g., in *David's fighting him now*.  But one might justifiably ask how the (rhythmic) accent can be determined—and perceived—in an utterance like *Dinner's ready*  where all the syllables tend to be of equal length. The answer is that accent, like all phonetic, phonological and other linguistic entities, is recognized, in the speech signal, by reference to internalized (memorized) patterns. It is assumed that patterns like those shown in fn. 6 above are remembered, together with a quasi-absolute 'beat' length, typical for the given speech tempo. —— Mental traces for various quasi-absolute beat lengths are necessary elements of a conductor's musical memory. —— Thus, *dinner's ready*  two NRUs of two syllables each, rather than four one-syllable NRUs because these would be almost twice as long, as in *Jack bought four dogs*  . It should also be noted that speech perception is 'heterarchical', with parallel signal processing mechanisms active at several levels, and with feedback. These complex processes concur in the resolution of possible ambiguities.

[8] Cf. Ladefoged's examples: *speed, speedy, speedily* (1975, p. l03).

lations between the rhythmical structure and syntax. These interrelations are described in detail, though perhaps still not fully, in Jassem, 1952. Suffice it to say here that words belonging to a TRU usually form syntactic entities, and phraseology often disrupts a simple rhythm-syntax relation. In *the minute hand of my watch, "of"* is in ANA (is proclictic): *(the minute hand) of my watch*, but in *a kind of fruit*, "*of*" belongs to one rhythm unit together with *kind*: *a kind of fruit*.[9]

7. The description of rhythm as explained under 1 to 5 above may very simply be incorporated into a phonemic transcription of General British English (RP), as proposed by Jassem (1949), by observing that (1) the accented syllable is preceded by an accent mark (tonal, e.g. [ˈ], [ˌ], [ˆ] etc.) or atonal-rhythmic [.], or general ['], (2) TRUs are separated by spaces, (3) a NRU, by implication, extends between an accent mark and the following space, and (4) the length of the syllables of the NRU, and the length of the ANA are determined by rules 1, 2 and S above. Thus, Abercrombie's (1964; 216) | *This is the* | *house that* | *John built* | is, according to (B), *This is the house that John built* and is transcribed /ˌðɪsɪz ðəˈhaʊs ðət ʤɒnˈbɪlt/.[10]

A very essential difference between the two conceptions of rhythm in spoken English is the treatment of the unaccented syllables, which—according to (A)—always follow the accented syllable. According to (B), they are either preaccentual (preictic) or postaccentual (postictic), the two categories being subject to fundamentally different rhythmic patterning.

If isochrony makes any sense at all, the length of the accented syllables must be related to the number of unaccented syllables within the same construct (foot, rhythm unit, or whatever). According to (A), in

John's pleased

John was pleased

John would be pleased

John would have been pleased

John would have been extremely pleased

the syllable /ʤɒn/ should get gradually shorter to accomodate the unaccented (unstressed) syllables. According to (B), the length of /ʤɒn/ in these utterances would not be significantly affected by the presence or absence of any following unaccented syllable(s) in our example because the TRU ends with /n/.[11]

---

[9] Cf. such unconventional spellings as *kinder (kind of), sorta (sort of), cupper (cup of)*, etc.

[10] Further examples of this type of transcription may be found in Jassem 1949, 1952, 1980 and 1981, and also in O'Connor 1967.

[11] According to yet another theory of English speech rhythm proposed by Allen (1968), one unstressed syllable following a stressed one does not affect the length of the interstress interval, but a further increase of the number of unstressed syllables results in an increase of the total length of that interval. The isochrony effect is maintained by a "negative correla-

4. Syllable or phone?

The smallest unit of speech rhythm is usually assumed to be the syllable and this is certainly appropriate for poetry. Describing English rhythm, Jones (1975: 238-244) uses musical notation with note lengths indicated in the same manner as in usual musical scores. But the length of the note does not here refer to the duration of the syllable but rather to the time from one vowel to the next. All the same, Jones does speak of syllable length in this connection (cf., e.g., 1976: 238 ff.). In Jassem's formulation of English rhythm (1949; 1952) it was the duration of the syllable that was assumed to be regulated by the isochrony principle. Abercrombie (1964) distinctly refers to syllable length. But within this theory, the length of the syllable is, partly at least, determined by the component phones (cf. above, Sec. 3). In 1965, O'Connor wrote "... this contention [the isochrony principle, W. J.] has never been satisfactorily tested and an investigation of the durational aspect of speech might throw light on this, and on the general pattern of rhythm in English" (O'Connor, 1965: 11). The same author, three years later (O'Connor, 1968), pointed out that if the syllable was an elementary unit of rhythm, then, in a fixed frame, it should maintain its length irrespective of the number of constituent phones, so that the duration of the component segments should be . . in more or less inverse proportion to their number." (O'Connor, 1968: 1). Although the experiment performed by O'Connor was limited in scope, the results are of considerable significance. It was found that, in a fixed rhythmical frame, the length of the syllable increased quite consistently with the number of the constituent phones. However, "The duration is in any case not directly proportional to the number of segments; there is therefore a compressive tendency which might correspond to isochrony mentioned in phonetic literature" (ibid, p. 3).

In order to measure the duration of any entity in the speech signal, it is obviously necessary to know where, along the time axis, the entity begins and ends. In a relatively simple case like the one investigated by O'Connor[12] there is no problem about the beginning and the end of the syllable in question. But there are cases in English where there is "no point of syl-

tion between the length of a given unstressed syllable and the number of syllables in the interstress interval" (p. 52); ". . . as we add more unstressed syllables, the interval gets longer, but the longer it gets, the more it resists any further increase in length" (p. 53). However: "Pre-clitics are generally shorter than post-clitics and they may undergo different kinds of changes because of these intrinsic differences" (ibid.). Lehiste (1972) also noted the fact that an unstressed syllable may "shorten" a preceding stressed syllable, or very nearly fail to do so - according to whether it belongs to the same word or not. The loss of the 'shortening effect' is particularly noticeable between subject and predicate.

[12] Syllables consisting of between 3 and 9 phones were embedded in a fixed frame "Take_____ Park", e. g., 'Take /ses/ Park', etc.

lable division" (Hockett, 1955: 52) because two peaks are joined by an interlude. A syllabification like /smol-ə/ (smaller) is morphophonemic, not phonetic or phonemic. China and finer are perfect rhymes in Standard British and there is no phonological ground for treating them differently at the level of phone-(phoneme)-to-syllable synthesis. In fact /n/ is an interlude in both words. Such cases make it very difficult, if at all possible, to measure the length of English syllables in connected speech. But the boundaries between segments (and phones) can, at least in principle, always be located in spectrograms. It is therefore preferable to examine rhythm units, such as Feet or NRUs, or whatever, as sequences of segments rather than sequences of syllables. Following O'Connor's train of thought, it seems appropriate to investigate the relations between the length of a rhythm unit and that of the constituent phones.

5. The Experiment

It will have been gathered from the preceding sections that much more research is needed, both at the speech wave and the perception levels, before a completely reliable description of English speech rhythm and the underlying principle of isochrony can be formulated. The main aim of the present experiment is (a) to test a hypothesis on the reality of isochrony in the speech signal on the basis of some reasonably representative material, and (b) to see whether there is any statistically justifiable reason for preferring one of the two specific theories presented in Sec. 3 over the other. Interrelations between rhythm and syntax in the light of the two theories will also come under discussion.

Considering the complexity of the entire problem of rhythm and isochrony in English, outlined above in Sec. 2, the present study can only hope to be one step in the direction of a complete solution.

Study Units Nos. 30 and 39 of M. A. K. Halliday: *A Course in Spoken English: Intonation* (1970) have been selected because the recorded materials are readily available commercially, so that—if necessary—measurements or other experiments may be made on the same material by other interested specialists. The pronunciation of the texts seemed to the present authors to be a good compromise between careful and deliberate, and casual and natural. Since a reliable automatic method of segmentation has not yet been devised, it was necessary to make the measurements visually from conventional spectrograms. The duration of individual phonetic segments was measured with an accuracy of about 5 ms and all the measurements were double-checked by at least two of the authors. The incidence of 'stress' is marked in the printed texts, but was checked by the authors in careful, though informal, listening tests and was confirmed, except for one or two cases.

6. Phones and Rhythm Units: Basic Data

After extensive preliminary statistical testing, the phones were divided into 18 classes, as follows:

F - flaps and initial voiceless lenis stops,
D - the weak-friction lenis fricatives / ð, v /,
G - the non-syllabic vocoids / w, j, r /,
E - the checked non-open vowels / e, ɪ, ə, ʊ, ʌ /,
B - non-initial lenis stops,
N - non-syllabic nasals,
H - the aspirate and the initial voiceless fortis aspirated stops,
K - fortis unaspirated stops,
Z -. the heavy-friction lenis fricatives / z, ʒ /,
SC- syllabic contoids,
KH - aspirated unaccented fortis stops,
S - fortis fricatives,
AFV - the lenis affricates / ʤ, dr /,
O - the close unchecked and the open checked vowels / iː, uː, æ (ː), ɒ /,
KHA -accented fortis stops, AF- the fords affricates / ʧ, tr /,
A - the mid and open unchecked monophthongs / ɜː, ɑː, ɔː / and the diphthongs
FTH - aspirated final fortis stops.

It will be noted that the classes include types of *phones* rather than types of phonemes. Preliminary statistical testing revealed that there are systematic durational differences between allophones of one phoneme, whilst phones belonging to different phonemes may be of the same duration.

Initially, a simplifying assumption was made that the mean durations of the phone types may be calculated irrespective of the two kinds of rhythm units posited under theory (B).

Table 1 presents the mean durations, with the variances, standard deviations and coefficients of variability.

The differences between the means need not necessarily be all statistically significant. An analysis of variance showed that at least some of the means were different at $\alpha = .01$, as shown in Table 2.

The null hypothesis on equality of all means being rejected at $\alpha = .01$, a means-clustering analysis as proposed by Gabriel (1964) was performed. This analysis groups together those of the means that do not differ significantly. It is based on the principle of minimum within-group variance with maximum between-group variance. This test led (after the 5th step) to the following grouping:

(F) (D) (G E B N) (H K Z SC) (KH S AFV O) (KHA AF A FTH)

The differences between the means within groups have not been shown to be

Table 1: Phone class duration

rank	class	mean ms	variance ms²	st. dev. ms	coeff. of var. %
1	F	16.6	32.5	5.7	34.3
2	D	47.6	303.2	17.4	36.6
3	G	56.7	500.7	22.4	39.5
4	E	59.2	643.1	25.4	42.9
5	B	60.2	486.5	22.1	36.7
6	N	61.8	809.4	28.4	46.0
7	H	64.9	1165.6	34.1	52.5
8	K	65.3	523.9	22.9	35.1
9	Z	70.3	1144.5	33.8	48.1
10	SC	76.2	685.3	26.2	34.3
11	KH	85.5	538.8	23.2	27.2
12	S	87.9	1157.9	34.0	38.7
13	AFV	93.9	832.9	28.9	30.7
14	O	96.3	1550.6	39.4	40.9
15	KHA	117.2	1074.8	32.8	28.0
16	AF	126.9	835.5	28.9	22.8
17	A	132.1	2971.7	54.5	41.3
18	FTH	137.5	709.1	26.2	19.4

Table 2: One-way analysis of variance for phone class duration

variance	SS	df	MS	F calculated	$F_{.05}$	$F_{.01}$
between	1636860	17	96286.1	98.985	1.62	1.96
error	2413350	2481	972.1			
total	4050210	2498	972.7			

significant at $\alpha = .05$, whilst any difference between any two means not belonging to one group is significant at $\alpha = .05$.

The phone duration data were subsequently broken down according to the type of rhythm unit, as shown in Table 3.

The small differences between the FOOT means in Table 3 and the corresponding means in Table I are due to slightly different data included in the calculations (e. g., the phones in Feet with 'silent stress' were disregarded in the calculations for Foot means).

The following observations can be made on inspection of the figures in Table 3:

Table 3: Mean durations of phone classes in rhythm units (ms). Size of class in parentheses.

phone class	FOOT		NRU		ANA	
F	(17)	16.8	(9)	18.5	(10)	14.5
D	(135)	48.4	(73)	49.8	(74)	45.2
G	(193)	56.2	(166)	57.9	(37)	43.2
E	(513)	59.8	(303)	66.0	(244)	50.1
B	(131)	60.4	(108)	60.5	(26)	59.6
N	(259)	62.2	(208)	62.9	(65)	60.5
H	(37)	65.8	(33)	69.3	(5)	35.6
K	(179)	66.1	(144)	68.5	(43)	55.4
Z	(69)	73.2	(57)	77.2	(20)	55.4
SC	(44)	75.9	(38)	77.6	(6)	65.0
KH	(49)	86.3	(37)	88.5	(12)	77.4
S	(165)	90.4	(144)	93.4	(27)	65.5
AFV	(18)	93.8	(16)	95.3	(1)	80.6
O	(141)	97.3	(123)	100.0	(25)	78.8
KHA	(41)	118.3	(42)	118.0	–	–
AF	(30)	127.7	(30)	127.7	–	–
A	(231)	134.2	(199)	139.9	(47)	100.2
FTH	(6)	137.5	(5)	136.5	–	–

Table 4: Relative durations of the phone classes (NRU = 1.0)

phone class	FOOT	ANA
F	.907	.785
D	.972	.907
G	.970	.755
E	.907	.759
B	.999	.985
N	.989	.962
H	.948	.514
K	.965	.809
Z	.949	.718
SC	.978	.838
KH	.975	.874
S	.967	.701
AFV	.985	.846
O	.973	.788
KHA	1.000	–
AF	1.000	–
A	.959	.717
FTH	1.010	–

(1) With but a few exceptions, the ranking of the means is the same within each of the three sets of data (and the same as that in Table 2).

(2) Most of the shifts of ranking occur in the ANA column, in which the means are less reliable because of smaller sizes of the classes.

(3) With very few exceptions, whenever the three means are available, the ANA mean is the smallest and the NRU mean the largest. Table 4 shows the relative means assuming that for each phone class the duration in NRU is equal to unity.

Table 5 includes mean values that are critical for the problem at issue here.

Table 5: Statistical characteristics of rhythm units

type of rhythm unit	grand mean of phone duration (ms)	mean rhythm unit duration (ms)	mean number of phones (mean length)	mean rate (phones per second)
FOOT	75.32	427	5.7	13.3
NRU	80.90	339	4.2	12.4
ANA	56.57	162	2.9	17.7

The data are not sufficient for a hypothesis to be put forward as to whether the distinction NRU *vs*. ANA affects the different phone classes in the same way, i.e., as to possible interaction between NRU/ANA and the phone classes. There is no indication, for instance, that vowels might be subject to a more drastic 'shortening' in ANA. But *the average duration of NRU is more than twice the average duration of ANA*. The number of phones in ANA is, on an average, 0.69 the number of phones in NRU whilst the average duration of ANA is 0.48 that of NRU. Actually *the average phone duration in ANA is 0.7 of the average phone duration in NRU*.

This difference is significant at $\alpha = .001$ and is tantamount to the phoneme rate being 1.43 faster in ANA than in NRU.

7. Rhythm Models

7.1. A model with two variables

Let us denote the duration of any rhythm unit by d, its 'size', expressed as the number of constituent phonemes, by n, and the duration of each of the constituent phonemes, by p. We ignore here the differences due to membership in different phone classes. By definition, $d = np$. Under an assumption of functional relationship between d and n, we may consider two extreme cases:

(A) No isochrony: $p = \text{const.} = \frac{d}{n}$ and

(B) Strict isochrony: $d = \text{const.} = np$

In both cases we are assuming that d is a function of n. The two cases are illustrated in Fig. 1

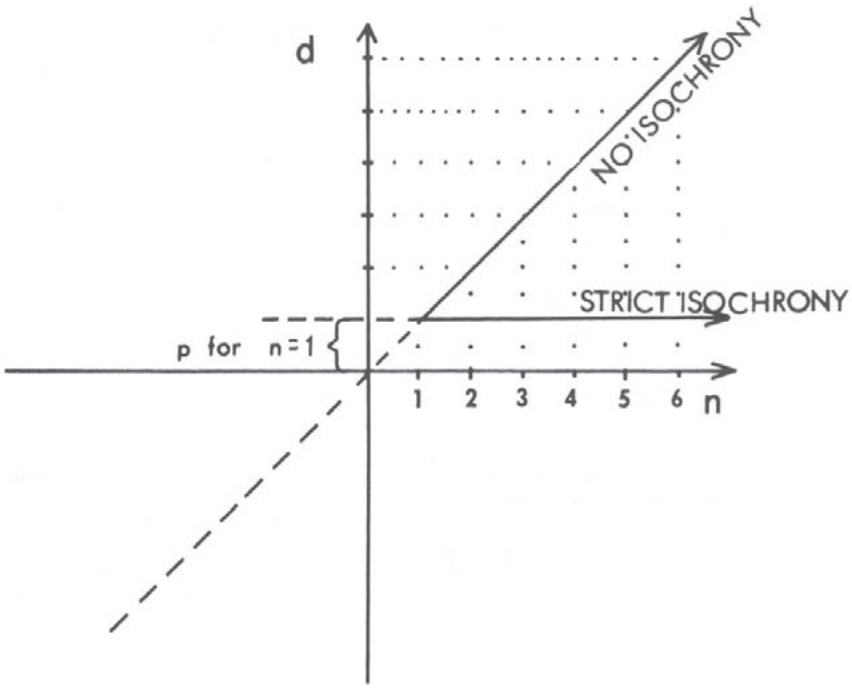


Figure 1: A functional relationship between d – the duration of a rhythm unit and n – the number of phones in the unit.

The functions in Fig. 1 have physical sense only for n = {real positive integer}. The upper limit of it is not known.

In reality, there is no functional relationship between d and n, at least because (a) there are systematic differences in the duration of phones, and rhythm units of a given size may consist of different phone classes, thus differing in duration, and (b) there is a certain measure of random variation even if the rhythm units are of the same size and have the same structure in terms of classes to which the constituent phones belong. There are no doubt other sources of variation, e. g., the position in the utterance (utterance-final rhythm units may tend to be longer), consequently, a realistic model of isochrony is a regression model in which $d = a + bn$, i.e., one in which the duration of the rhythm unit is estimated from its size on the basis of a best- fitting analytical relationship between the two variables, as shown in Fig. 2.

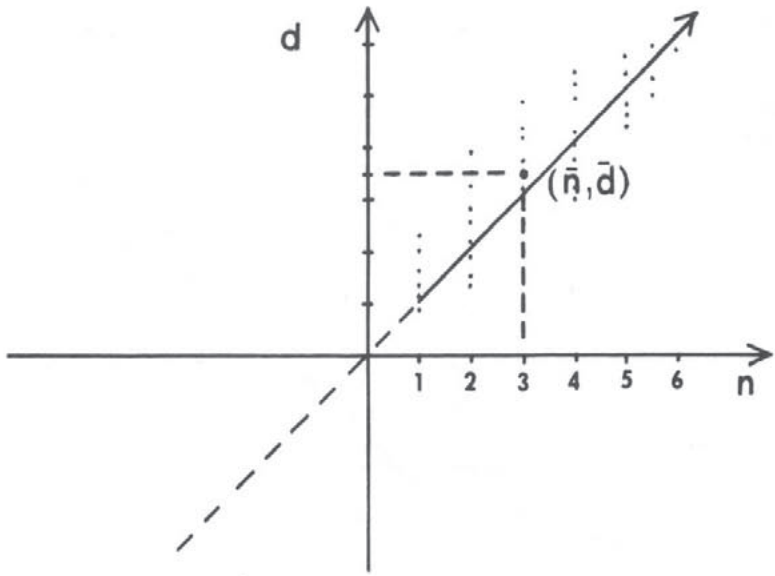


Figure 2: A regression model of the relationship between d – the duration of a rhythm unit and n – the number of phones in the unit.

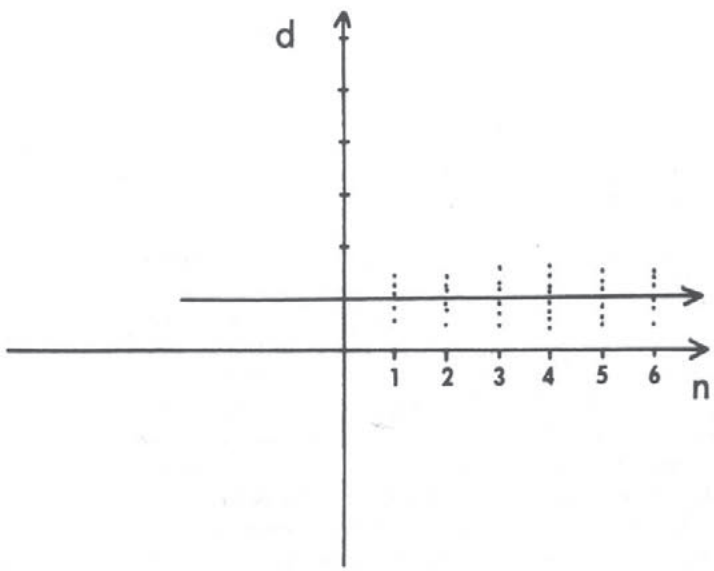


Figure 3: Regression coefficient b = 0. Values of d cluster closely about a 'strict isochrony' line.

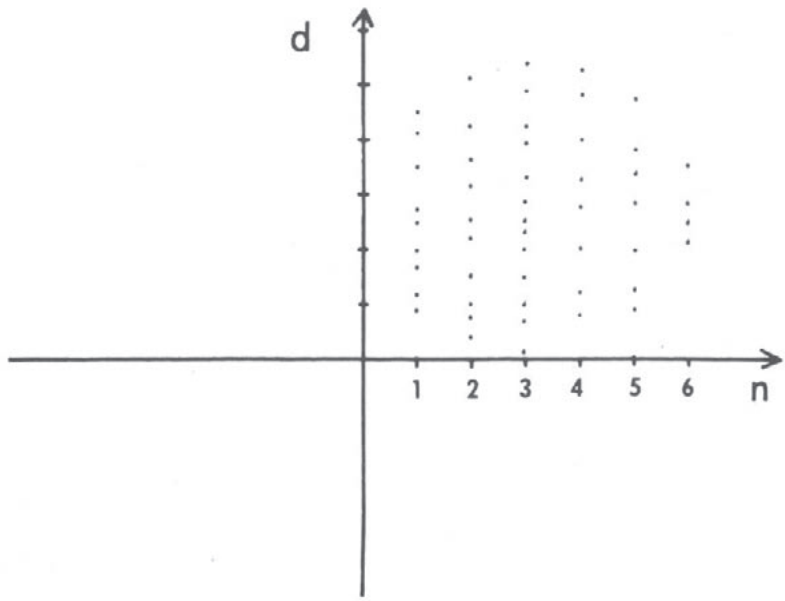


Figure 4: Regression cofficient b = 0. Values of *d* completely random.

Note that if the regression coefficient $b = 0$, then isochrony is indefinite, as shown in Figs. 3 and 4.

We shall consider two models with two variables. In one of these, *d* will be estimated from the sum of the mean durations of the constituent phone classes. This sum will be denoted by $\overline{d}$.

7.2. A model with three variables

Another possibility that has to be investigated is that *d*, the duration of the rhythm unit, depends both on $\overline{d}$, the mean cumulative duration of the constituent phone classes, and on their number *n*. The value of $\overline{d}$ may be expected to correlate highly with *n*, yet the interrelations between $\overline{d}$ and *n* may be such that a better estimate of *d* is obtained if both $\overline{d}$ and *n* are assumed to have an effect on *d*. In a general from, this multiple association is expressed by a regression equation $d = a + b\overline{d} + cn$.

8. Analysis of Regression

8.7. Two variables: d *and* n

8.1.7. Linear regression

In a bivariate regression model, *d* is first made to depend on *n* only. For no isochronism, the regression line can be made to pass through the origin and form a 45° angle with the abscissa by performing a linear transformation of both axes,

with $(d - \bar{d}) / p$ on the ordinate and $(n - \bar{n})$ on the abscissa. The regression equation then takes the form

$$\frac{d - \bar{d}}{\bar{p}} = a + b(n - \bar{n}) \qquad (1)$$

Ideally, under this transformation, a should equal zero, but in reality $a \approx 0$ because $\bar{p}$ is not calculated from exactly the same raw data as $\bar{d}$. As mentioned before, some rhythm units had to be left out because their duration was not measurable. Also, in a few cases, the total duration of the rhythm unit could be measured, though segmentation was doubtful so the individual values of p were not measured. It will be seen that the resultant discrepancy is entirely negligible.

Table 6

	$\overline{n - \bar{n}}$	$\overline{\frac{d - \bar{d}}{\bar{p}}}$	$\sigma(n-\bar{n})$	$\sigma\frac{d-\bar{d}}{\bar{p}}$	r	$r^2 \cdot 100\%$	a	b	arctg b
FOOT	0.00	0.00	2.21	1.98	0.72	51.7	−0.0026	0.644	32.8°
NRU	0.00	0.00	1.63	1.47	0.62	39.0	−0.0018	0.561	29.3°
ANA	0.00	0.00	1.53	1.68	0.83	69.6	−0.0036	0.916	42.5°

Table 6 gives the results of an analysis of regression with the variables d and n, as expressed by eq. (1).

The magnitude that is most directly related to isochrony is either the regression coefficient b or the corresponding arctg b. The following conclusions can be drawn from Table 6:

(1) The regression coefficient for ANA is quite close to unity, and the corresponding arctg b, i. e., the angle of the regression line with the abscissa, is close to 45°, consequently there is very little isochrony in ANA.

(2) The NRU has a coefficient of regression which is nearly half (exactly .613) the coefficient of regression for ANA, and the angle of the regression line for NRU is 0.69 the angle for ANA[13]. There is distinct tendency towards isochrony in NRU, though it is not very close to strict isochrony.

(3) The regression coefficient for FOOT is intermediate, but closer to that of NRU.

(4) The coefficient of determination $r^2 \cdot 100\%$ indicates that though there is distinctly more variance unaccounted for in the regression of NRU than in the regression for ANA, both may be considered as satisfactory in the sense of their predictive power. The fact that the coefficient of determination is high for ANA and lower for NRU makes it plausible that there is a factor which is active in the latter but absent in the former. Prob-

[13] Cf. above, Sec. 6, on the relative mean durations of NRU and ANA.

ably this factor is the distinction between final and nonfinal position. The ANA can, by definition, only stand in a nonfinal position. The coefficient of determination for the FOOT is intermediate between the other two and indicates that theory (A) is not, in a statistical sense, unacceptable. But it is shown to *obliterate a distinction which is statistically very highly significant* (viz. that between ANAs and NRUs). The isochrony effect is shown in Fig. 5.

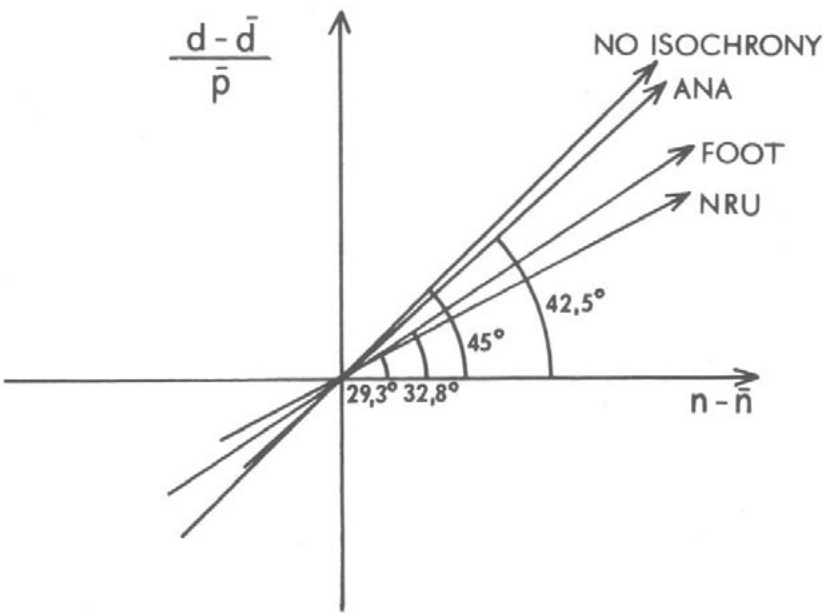


Figure 5: Linear regression of *d* on *n* in Foot, NRU and Ana.

8.1.2. Quadratic regression

Table 7 gives the results of a quadratic regression of *d* on *n* and n^2. After normalization of the variables, the regression equation here is of the form

$$\frac{d - \bar{d}}{\bar{p}} = a + b(n - \bar{n}) + c(n - \bar{n})^2 \qquad (2)$$

Table 7

	$\overline{(n-\bar{n})^2}$	$\sigma(n-\bar{n})^2$	$r[(n-\bar{n})^2,d]$	$r[(n-\bar{n}),d]$	$r[(n-\bar{n}),(n-n)^2]$	a	b	c	$r^2 \cdot 100\%$
FOOT	4.90	7.14	0.29	0.72	0.27	−0.141	0.619	0.0282	52.7
NRU	2.67	4.52	0.37	0.62	0.47	−0.078	0.523	0.0287	39.6
ANA	2.34	3.81	0.55	0.83	0.56	0.117	0.844	0.0515	70.3

It can be seen from Table 7 that the coefficients of determination for all three types of rhythm unit are marginally better than in the linear model, which is more directly interpretable. Therefore, there is very little, if anything, to be gained from a quadratic regression model.

8.2. Two variables: d *and* $\bar{d}$.

It is also possible to estimate the duration of a rhythm unit from the cumulative average duration of the constituent phones. In other words, we now take the mean duration of each phone in the rhythm unit, as shown in Table 3 in the appropriate column, add the figures obtaining $\bar{d}$ and estimate d from this. If there is no isochrony, then, on an average, $d = \bar{d}$. In the case of strict isochrony, d is constant, so there must be a 'coefficient of compression'. We have found it convenient to express the relation between d and $\bar{d}$ by

$$\frac{d - \bar{d}}{\bar{p}} = a + b \frac{\bar{d}}{\bar{p}} \qquad (3)$$

because, again, the regression coefficient and its corresponding angle are easily interpretable.

Table 8

	$\frac{\bar{d}}{\bar{p}}$	$\sigma\frac{d}{\bar{p}}$	r	$r^2 \cdot 100\%$	a	b	arctg b
FOOT	5.74	2.09	0.80	63.8	−4.35	0.758	37.1°
NRU	4.25	1.48	0.73	52.8	−3.07	0.724	35.9°
ANA	2.92	1.58	0.90	80.4	−2.78	0.951	43.6°

The results of the analysis of regression are contained in Table 8. As the means for d are here normalized by $\bar{p}$ i. e., $\bar{p}_{Foot}$, $\bar{p}_{\mathrm{NRU}}$ and $\bar{p}_{\mathrm{ANA}}$ respectively, they represent in fact the mean number of phones in the different rhythm units. As in Table 6, the correlation coefficient is highest for ANA and smallest for NRU, but each is distinctly higher than its counterpart in Table 6. The values of the coefficients of determination are therefore also each better. Thus, by taking the means for the various phone classes which go to make d we have accounted for part of the variance still unaccounted for in the previous bivariate models. Fig. 6 shows the isochrony effect.

8.3. Regression with three variables

Even though the simple regression models are quite satisfactory as judged by the high coefficients of determination, it is tempting to see whether some further improvement might not be achieved by predicting d from both $\bar{d}$ and n.

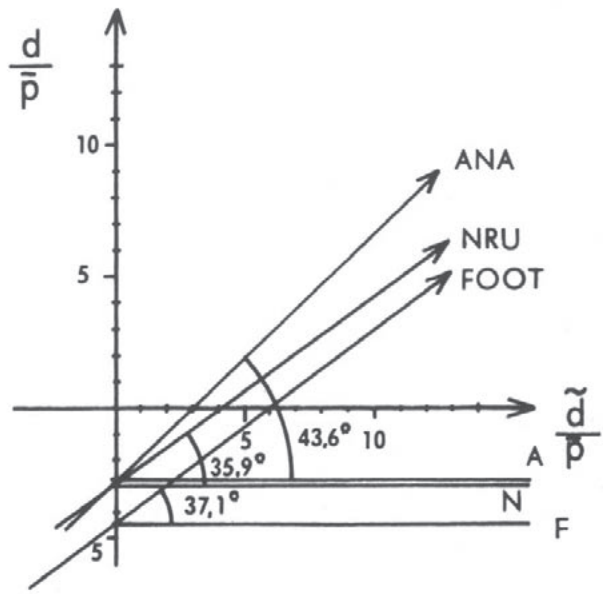


Figure 6: Linear regression of d on $\bar{d}$ in Foot, NRU and Ana according to Table 8.

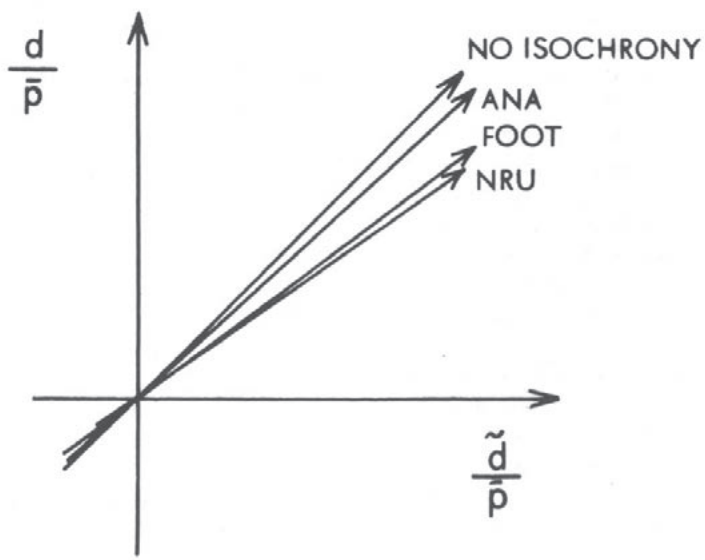


Figure 7: Linear regression of d on $\bar{d}$. Regression lines brought to a common origin.

Table 9

	$\bar{n}$	$\frac{\bar{d}}{\bar{p}}$	$\frac{\bar{d}}{\bar{p}}$	σ (n)	$\sigma(\frac{\bar{d}}{\bar{p}})$	$r(n,\frac{\bar{d}}{\bar{p}})$	$r(n,\frac{d}{\bar{p}})$	$r(\frac{d}{\bar{p}},\frac{\bar{d}}{\bar{p}})$	a	b	c	$r^2 \cdot 100\ \%$
FOOT	5.76	5.76	5.74	2.21	2.09	0.93	0.72	0.80	1.33	−0.187	0.942	64.4
NRU	4.25	4.25	4.25	1.63	1.48	0.91	0.62	0.73	1.11	−0.221	0.950	53.8
ANA	2.94	2.92	2.92	1.53	1.58	0.96	0.83	0.90	0.17	−0.378	1.302	81.3

The following linear regression was tested:

$$\frac{d - \bar{d}}{\bar{p}} = a + b\,\frac{\bar{d}}{\bar{p}} + cn, \qquad (4)$$

and Table 9 gives the results of the regression analysis; cf. also Fig. 7.

Ideally, the three means: $\bar{n}$, $\bar{d}/\bar{p}$ and $\bar{d}/\bar{p}$ should be equal. Again, the slight differences are due to small differences in the raw data. The correlation coefficients for (n, $\bar{d}/\bar{p}$) are naturally very high. Both for (n, $\bar{d}/\bar{p}$) and (d/p, $\bar{d}/p$) the correlation coefficients are highest for ANA and lowest for NRU, but all values are high. As can be seen from the values of the coefficients of determination, the three-variate model accounts for more variance than any of the _bivariate models, though it is only marginally better than that for d and $\bar{d}$.

9. Some Linguistic Considerations

It was explained in Jassem (1949; 1952) that if the model of rhythm there proposed is accepted, then rhythm can very simply be indicated in the phonemic transcription of a running English text by observing the rules quoted above in Section 2. The beginning of Unit 30 in Halliday (1970) transcribed according to the rules reads as follows:

['ɪzðæt 'ʧiz tɷ'it ɔ'ɪzɪt tə'pɷdɪn ə'maɷs 'træp ‖ ɪfaɪd'nɜɷn 'ɜlɪər ə'baɷt ðə'wedɪŋ aɪdəv'sent ə'keɪbl]

The beginning of Unit 39 reads:

[aɪ'laɪk ðæt'pɜɪm baɪðætɪ'levnjɜr ɜɷld ɔs'treɪlɪən 'bɔɪ 'dɪdjɷ 'sɪɪt]

Is that /'ɪzðæt/, *cheese* /'ʧiz/, earlier /'ɜlɪər/ are examples of NRUs that are co-extensive with TRUs because they include no anacrusis. *To eat* /tɷ'it/, *or is it* /ɔ'ɪzɪt/, *to put in* /tə'pɷdɪn/, *if I'd known* /ɪfaɪd'nɜɷn/, are examples of TRUs that begin with an anacrusis.

No attempt has ever been made to indicate rhythm as described by Abercrombie (1967) and Witten (1977) in a running transcription of an English text.

As mentioned earlier, both models are independent of syntax, but both admit interrelations between syntax and rhythm. Abercrombie's model, as applied by Halliday, sometimes results in very peculiar tone groups such as // *if I'd known earlier about the wedding I'd have* // *sent a cable* //. The

first tone group includes the subordinate clause plus the subject and part of the predicate of the main clause, the remainder of which forms the second tone group. Many peculiar tone group boundaries may be found in Halliday 1970. Here are some more examples: // *he was grey and he was woolly and his* // *pride was inordinate, he danced on a sandbank in the* // *middIe of Australia and he* // *went to the Big God Ngong* // (p. 121). // *on the Isle of Man you can* // *still ride in a horse-drawn tram* // (p. 117). Such tone group boundaries, strange from the syntactical point of view, are due to the assumption that unstressed (unaccented) syllables always belong to the same rhythm unit as the preceding stressed (accented) syllables and from the assumption that silence (pause) is a marker of a tone-group boundary[14]. Model (B) of English rhythm has a simpler relation to syntax and does not result in such disconcerting discrepancies between the phonological and the syntactical structure of running speech.

10. Summary and Conclusions

A statistical method has been applied to express isochrony in quantitative terms, and an attempt has been made to find isochrony in the acoustic speech signal. It was assumed that if isochrony was at all detectable in the speech wave, it should affect the duration of the phones which constitute the rhythm units.

Tape recordings of continuous, naturally spoken General British English ('RP'), consisting of a total of almost 2500 successive phones served as experimental material. The duration of the phones was measured spectrographically and the phones were grouped into classes according to their mean duration.

Two theories of English rhythm were tested: Abercrombie's—called (A)—which postulates one type of quasi-isochronous rhythm unit, the FOOT, and Jassem's—called (B)—which posits two types, viz. ANACRUSIS with no isochrony, and NARROW RHYTHM UNIT which tends towards isochrony. Four regression models were applied making the duration of a rhythm unit depend (a) linearly on the number of phones in the unit, (b) curvilinearly on the number of phones in the unit, (c) on the sum of the mean durations of the phones in the unit and (d) on both the sum of the mean durations of the phones in the unit and their number. The results of the regression analysis show that in all models the tendency towards isochrony is minimal in ANACRUSIS and quite distinct, if not very strong, in the NARROW RHYTHM UNIT. Isochrony is also present in the FEET, but the FOOT averages out and obliterates the distinction between

[14] On a distributional view of the tone-group, see Jassem 1978.

ANACRUSIS and NARROW RHYTHM UNIT which is shown to be statistically very highly significant.

In keeping with theory (B), rhythm and isochrony can be very simply indicated in running transcription of English text, which does not appear to be possible within theory (A). This is of particular importance for computer-controlled speech synthesis by rule. Using a very simple algorithm based on rules supplied by theory (B), the temporal organization of speech may be generated from a transcription indicating the incidence of accent and boundaries between TOTAL RHYTHM UNITS plus a table of mean phone durations.

Theory (B) relates the syntactic component of spoken text to its phonological component much more simply than does theory (A).

Acknowledgements

The authors wish to express their appreciation of a grant from the National Research Council of Canada to the Department of Computer Science, University of Calgary, Alberta which enabled WJ to work there during an extended visit to Canada, and to thank the British Council for supporting a shorter working visit by WJ to the University of Essex, Colchester and one by IHW to Poznan. The co-operation of dr M. Krzyśko and Mr. P. Stolarski of the Computer Centre of Mickiewicz University, Poznań, in the computing labour is also gratefully appreciated.

References

Abercrombie, D. (1964). Syllable quantity and enclitics in English. In D. Abercrombie & al. (eds.), *In Honour of Daniel Jones*. Longmans: London. 216-222.

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

Abercrombie, D. (1973). A phonetician's view of verse structure. In *Phonetics in Linguistics*. Longman: London. 6-13.

Adams, C. (1979). *English Speech Rhythm and the Foreign Learner*. Mouton: The Hague.

Allan, C. D. (1968). On testing for certain stress-timing effects. *UCLA Working Papers in Phonetics* 10: 47-59.

Bolinger, D. L. (1965). Pitch accent and sentence rhythm. In D. L. Bolinger, *Forms of English*. Hokuou PubI. Co.: Tokyo. 139-180.

Gabriel, K. R. (1964). A procedure for testing the homogeneity of all sets of means in analysis of variance. *Biometrics* 20 (3): 459-477.

Halliday, M. A. K. (1970). *A Course of Spoken English: Intonation*. Oxford University Press: Oxford.

Hockett, C. F. (1955). *A Manual of Phonology. Indiana University Publications in Anthropology and Linguistics,* Memoir 11.

Jassem, W. (1949). Indication of rhythm in the transcription of Educated Southern English. *Le Maitre phonétique* III/92: 22-24.

Jassem, W. (1952). Stress in Modern English. *Bulletin de la Societé Linguistique Polonaise* XII: 189-194.
Jassem, W. (1978). On the distributional analysis of pitch phenomena. *Language and Speech* 21:362-372.
Jassem, W. (1980). *Fonetyka języka angielskiego (English Phonetics)* PWN: Warszawa, 7th ed.
Jassem, W. (1981). *Podręcznik wymowy angielskiej (A Handbook of English Pronunciation)* PWN: Warszawa, 7th ed.
Jassem, W. & Gibbon, D. (1980). Re-defining English accent and stress. *Journal of the International Phonetic Association* 10: 2-16.
Jones, D. (1976). *Outline of English Phonetics*. Heffer: Cambridge. 9th ed. repr.
Ladefoged, P. (1975). *A Course in Phonetics*. Harcourt, Brace, Jovanovich: New York.
Lea, W. A. (1974). *Prosodic aids to speech recognition: IV A general strategy for phonologically guided speech understanding*. Univac Rep. PX 10791.
Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society ofAmerica* 54: 1228-1234.
Lehiste, I. (1975). The role of temporal factors in the establishment of linguistic units and boundaries. In W. U. Dressler & al. (eds.), *Phonologica* 1972. 115-122.
Lehiste, 1. (1977). Isochrony reconsidered. *Journal of Phonetics* 5: 253-263.
O'Connor, J. D. (1965). The perception of time intervals. *Progress Report, Sept. 1965*. Phonetics Lab. UCL. 11-13.
O'Connor, J. D. (1967). *Better English Pronunciation*. Cambridge University Press: Cambridge.
O'Connor, J. D. (1968). The duration of the foot in relation to the number of component sound- segments. *Progress Report, June 1968*. Phonetics Lab., UCL. 1-6.
Pike, K. L. (1945). *The Intonation ofAmeri can English*. University of Michigan Press: Ann Arbor.
Shen, Y. and Peterson G. G. (1962). Isochronism in English. *University of Buffalo Studies in Linguistics, Occasional Papers* 9: 1-36.
Uldal, E. (1971). Isochronous stresses in R. P. In L. L. Hammerich et al. (eds.), *Form and Substance*. Akademisk Forlag: Odense. 205-210.
Witten, I. H. (1977). Flexible scheme for assigning timing and pitch to synthetic speech. *Language and Speech* 20: 240-260.