Some results from a preliminary study of British English speech rhythm

by

David R. Hill, Ian H. Witten*, and W. Jassem†

University of Calgary Computer Science Department
Research Report 78/26/5
(Slightly revised 2007)

DEPARTMENT OF
COMPUTER SCIENCE

[This page is intentionally blank]

MAN-MACHINE SYSTEMS LABORATORY REPORT

Department of ComputerScience
The University of Calgary
CALGARY, Alberta, Canada, T2N 1N4

Some results from a preliminary study of British English speech rhythm

by

David R. Hill, Ian H. Witten*, and W. Jassem†

University of Calgary Computer Science Department
Research Report 78/26/5
(Slightly revised 2007)

The original version of this paper was presented at the 94th Meeting of the Acoustical Society of America as the first paper in Session V. Speech Communication V: Linguistics and Prosodic Features (Peter McNeilage, Chairman) on Thursday December 15 1977.

* Dept. of Electrical Engineering Science, The University of Essex, COLCHESTER, Essex, England CO4 3S4
(Current address: Dept. of Computer Science, University of Waikato, Te Whare Wānanga o Waikato Hamilton, New Zealand)

† Instytut Podstawowych Problemow Techniki PAN, Pracownia Fonetyki Akustycznej, Noslowskiego 10, 61-704, POZNAN, Polland

[This page is intentionally blank]

CONTENTS

[This page is intentionally blank]

SUMMARY

The paper reports a statistical study that was carried out in order: (a) to increase our understanding concerning the nature and underlying determinants of rhythm in spoken British English; and (b) to serve as a basis for generating good rhythm for computer speech output.

Two sets of utterances were selected for the study, because they were available as published audio tapes associated with a book (Halliday 1970), and based on the fact they were carefully, but naturally spoken to illustrate British English intonation for teaching English as a foreign language. It also seemed advantageous that we had nothing whatever to do with their generation, nor were the tapes specifically aimed at teaching or illustrating rhythm.

Segmental analyses were performed, checked by a second phonetician, and the resulting data on segment durations, together with information concerning higher level structure, were prepared for computer analysis. It was found that 3 main factors account for about 70% of the rhythmic structure of spoken British English, and can be used to produce quite good models of speech rhythm that are suitable for practical application in computer speech output. The contributions to rhythmic structure are: (a) about 45% due to the identity of the constituent sounds; (b) about 15% arising from differences in segment duration between the information points and other speech (marked versus unmarked segments) and (c) about 9% representing a tendency towards equal spacing in time of the stressed syllables (an isochrony effect that was most significant for marked speech). The work was supported by the Natural Sciences & Engineering Research Council of Canada under under Grant Number A5261.

[This page is intentionally blank]

Introduction

A knowledge of the rhythmic structure of spoken English is of interest for a variety of reasons. Three of these are: (a) to advance our knowledge of language production processes in general; (b) to provide an accurate basis for teaching non-native speakers an acceptable way to speak the language; and (c) to allow the rhythm of spoken English to be defined for purposes of computer speech output.

It is generally agreed that the perceived rhythm of spoken English depends primarily on the timing of successive accented syllables. English is said to be a stress timed language. Accented syllables normally occur in information (content) words such as verbs and nouns, rather than in form words like *and, the, for* and *to*. Form words are associated with the structure of a message as opposed to the information to be conveyed.

Theories of rhythm for spoken English, on both sides of the Atlantic, have suggested that rhythmic units delineated by successive accented syllables have a tendency to be of more equal duration (more isochronous) than might be expected from the total number of syllables or elementary speech segments comprising each rhythmic unit. This is in contrast to a language like French in which it is suggested that successive syllables tend to be of relatively equal length. French is said to be a syllable timed language. Both theories are strongly disputed because objective measurements on speech reveal that rhythmic units vary in duration by a factor of as much as 6 or 7 to 1 (for example, the present results, whilst Wenk & Wioland (1982), cited by Williams & Hiller (1994) found that doubling the number of syllables in French did not result in doubling the time for the speaking time for those syllables. It has also been suggested that the perceived effect is mainly subjective (Lehiste 1973).

It is assumed that a rhythmic unit in English stretches from just before one accented syllable to just before the next, so that all syllables are taken into account in assessing the duration of rhythmic units. This is the form implicit in the theories of Jones (1918), Abercrombie (1967) and Halliday (1970) in Britain, or Pike (1945) and Ladefoged (1975) in the U.S.A. Halliday calls these units *feet*, which he likens to bars in music, and these form the rhythmic basis for speech in his book (Halliday 1970). The book is intended to help non-native speakers of British English produce a reasonable approximation to the native speaker's intonation pattern. This attempt depends intimately on some reasonable model of natural rhythm, but rhythm *per se* is not what is being taught in the book.

In Jassem's theory (Jassem 1952), which also traces its origins to Daniel Jones, not all syllables that are spoken are counted in assessing the duration of rhythmic units (termed *rhythm units* in his theory). Some syllables at the end of Halliday's feet that belong, syntactically, to what follows (so-called *proclitic* syllables) are omitted when totalling up the duration of the rhythm units, and it only the rhythm units that are considered to have a tendency towards isochrony is Jassem's theory. The discounted syllables are termed *anacruses*, again by analogy with their musical equivalents. Though anacruses are part of the rhythm, in the sense that they occur, they are not part of the isochronous rhythmic structure. Thus, if programming a computer to speak English by using a speech synthesiser and suitable driving rules, the results of the two theories would be expected to be rather different, even though both theories assume a tendency towards isochrony of rhythmic units based on stressed syllables and, in principle, mark the beginnings of the rhythmic units in the same place (ignoring Halliday's so-called silent beats, which seem to us to be a purely subjective effect, in the sense that objective measurements do not reveal silence). As a result of our attempts to produce spoken output for computers, the question arose as to whether such theories might serve as a basis for specifying rhythm, and, if so, whether Jassem's theory or Halliday's theory was better in some sense.

These questions involved some further questions. For example: what might a "tendency towards isochrony" mean, in quantative terms; and is either theory superior as a predictor of actual duration measurements taken from spoken English. There was also the question, given such measurements, as to whether some alternative pattern of rhythmic structure might be discerned, or at least whether additional factors playing a part in rhythmic structure might be formalised and quantified. We also hoped to check out other results in the linguistics literature concerning factors affecting durations.

However, the emphasis in this paper is on the analysis and application of the rhythm data obtained from spectrographic analysis in what must be considered a preliminary study. A detailed comparison of the two theories in the light of the data obtained has already appeared in another paper (Jassem, Hill, Witten 1984).

Discernable factors in speech rhythm

For this preliminary study, two samples of spoken British English were analysed, both being taken from the illustrative tape recordings supplied with Halliday's book (1970). The book and its associated tapes was considered suitable for our study because: (a) the tapes are published; (b) the tapes are primarily concerned with illustrations of intonation, and seem less likely to involve manipulated rhythmic structure; (c) the speakers have accents falling in the general range of interest; and (d) we were not involved in the production of the tapes in any way. We chose Study Units 30 and 39.

Study Unit 30 comprised 40 assorted utterances, statements and questions, by a single speaker, whilst Study Unit 39 comprised 49 utterances that, taken together, made up a discussion between two speakers of similar speech habits, about a poem. The durations of the individual speech sounds for both study units were determined by careful, doubly-checked hand analysis of sound spectrograms. Checking was carried out by a second phonetically trained and experienced reader. About 10% of the material was analysed completely independently. Agreement was surprisingly good, in that, after a few obvious mistakes had been corrected, the discrepanices fell within ±2.5 msecs. This agreement is in accord with error estimates in the field. A total of 217 seconds of speech (close to 2500 phones) were analysed. It should be noted that traditional phonetic criteria were used in the segmental analysis—a procedure that cannot, in the ultimate analysis, be entirely self-consistent because there are differing criteria for placing segment boundaries, depending on the particular local environment of any given segment.

The data thus collected were processed by computer, rechecked for numerical and notational consistency, and then subjected to detailed statistical analysis. The durations of either Halliday-type or Jassem-type rhythmic units ("feet" and "rhythm units" respectively) were readily calculated from the durations of the individual segmental elements.

There were three main levels of analysis: the segment (or individual speech sound) level; the syllables level; and the rhythmic unit level. At the segment level, we were interested to know what factors contributed to setting the durations of individual speech sounds, and in what proportion. Thus the basis of this analysis was *contribution to variance of mean segment duration*. At the syllable level, we wished to know what factors contributed to variance in mean syllable duration, and in what proportion. Finally, at the level of rhythmic units, we wished to know what factors contributed to the variance of mean duration of rhythmic units (both Halliday's feet and Jassem's rhythm units), and in what proportion. From these processed data we hoped to deduce other things.

From the point of view of synthesising speech using a computer, if the duration of each speech sound can be determined according to known and specifiable factors, then the rhythm of the utterance will follow. The segment level of analysis was, therefore, of prime importance. An important part of the two higher levels of analysis was to see how well the durations of syllables and rhythmic units were modelled on the basis of data gathered at the segment level, taking account of increasing numbers of factors. The figures for the two study units were surprisingly similar in many respects and so only the figures associated with Study Unit 30 are quoted below, where main results are involved, although details of both analyses appear in the appendix. Study Unit 39 involved two speakers and was less formal speech, both characteristics involving greater variance (which was the main difference, and the reason for focussing on Study Unit 30).

The chief determinant of speech sound duration, as judged by contribution to variance in mean segment duration, was (as might be expected) the kind of phoneme (speech posture) the speech sound represented; that is its *phoneme type*. Phoneme type (47 types total) accounted for 45% of the variance in mean segment duration. It should be noted that, had the phoneme type accounted for 100% of the variance in mean segment duration, we should probably have needed to look for no other factors. However, it should also be noted that the various factors that were examined were not always independent. For example, syllables were divided into four types -- *strong, weak, proclitic* and *enclitic*. Strong and weak syllables were both accented, and were distinguished by their composition in terms of speech sounds. Proclitic syllables were unaccented, and distinguished as described above, whilst enclitic syllables were unaccented syllables that were not proclitic. Thus the effect of syllable type on mean segment duration was not independent of the effect of phoneme type on mean segment duration, since phoneme type partially distinguished syllable type. For this reason, although the type of syllable into which a segment fell was found to accountfor 14% of the variance

in mean segment duration, it was not found necessary to take this into account as an additional, independent factor in our initial modeling attempts. A second factor that was reasonably independent was the kind of rhythmic unit into which a segment fell. The units could be of four major types: they could fall at the end of an utterance (*final*); they could contain the word that conveyed the main point of the phrase and therefore carry additional accent (*tonic*); they could be subject to both the conditions just specified (*tonic-final*); or they could be *unmarked*. We found that, for either feet (Halliday) or rhythm units (Jassem), this factor accounted for about 15% of the variance in mean segment duration. For most purposes, it was sufficient to consider only the two classes *marked* (the first three types) and *unmarked*.

Another factor of importance in determining speech sound duration, as judged by its independent contribution to variance in mean segment duration, was the *size* of a rhythmic unit (i.e. the number of segments in it). As the size of a rhythmic unit increased, for both kinds of unit, the mean segment duration decreased, so that the duration of rhythmic units did not increase in direct proportion to their increase in size. Although other mechanisms whereby rhythmic units tend to be of equal duration can be postulated, this particular effect clearly causes the units to be more nearly of equal duration than they would have been without it. The effect accounted for 9% of the variance in mean segment duration and is an objectively measurable tedency towards isochrony. The intervals between successive stresses is more equal than might be expected. No corresponding effect of syllable size could be discerned so that we believe there is no evidence, in our data, of any syllable timed feature of spoken British English. It is true that syllables consisting of one segment gave consistently longer mean segment durations, regardless of type,but this was almost certainly due to the fact that such segments were necessarily syllabic nucleii (i.e. mostly vowels or diphthongs). It was shown that being a syllabic nucleus was itself a measurable factor relating to mean segment suration, accounting for 5% of the variance, but it was not independent of phoneme type, and it was not found necessary to take it into account in our modelling of rhythmic structure.

Other sources of variance in mean segment duration were tested, including the well-reported effects of being in an initial or final consonant cluster: the effect of voiced versus voiceless consonant termination on the length of the syllable nucleus—also compared to open nucleii; and the effect of being in different utterances. Despite reasonable care to exploit combinations of factors that might enhance such effects, they did not show up as being significant; indeed the differences were not infrequently the "wrong" way.

We noticed that some marked rhythmic units seemed to be marked only by pitch change, so that the associated marked segments would have increased the variance of the population disproportionately. As we had no theoretical framework to predict such occurrences, we were not able to take this effect into account. We also observed the opposite—rhythmic units that were marked only by duration increase. Though these would not have affected our analysis for rhythm, they are a problem in modelling intonation. Another problem is that, when in doubt, we had to take the published transcription of the tapes we analysed. It is clear that, whether the tapes were produced from the script, or the script was generated from the tapes, errors and conflicts of opinion would and did occur. Finally, there was undoubtedly some error in our own measurements—despite our care to keep this to a minimum. Such factors undoubtedly all contribute to unexplained variance.

Appendix 1 provides a reasonably detailed digest of the results of our statistical analysis for both study units, on which the preceding discussion is based. It also presents the results of our rhythmic structure modelling according to different criteria, based on that analysis. The segment duration statistics and regression data, used in our modelling experiments, and in our subsequent speech synthesis, are included. It is to the modelling experiments that we now turn.

Modelling syllables and rhythmic unit durations

The three factors that were used in our initial attempts to model rhythmic structure at the syllable and rhythmic unit level were: (1) the kind of segment; (2) whether the segment occurred in a marked or an unmarked rhythmic unit; and (3) the number of segments in the rhythmic unit.. By "modelling rhythmic unit structure", we mean the process of assigning segment durations, applying any regression correction for the isochrony effect, and then comparing the results with the original data.

Because of their assumed nature, anacruses were modelled solely on the basis of their constituent seg-

ments, with no marked/unmarked distinction and no regression correction. Modelling was carried out using both feet and rhythm units as the guiding framework to see if either provided a better "fit", or other advantages.

Two possible classifications of phoneme identity were used. The first divided phonemes into 11 crude classes (long vowel, short vowel, voiced plosive, nasal, etc.), whilst the second used the original 47 types implicit in the analysis. Combining these classifications with the marked/unmarked distinction led to one table containing 22 mean segment durations (on which the so-called *phoneme-class* model was based), and another table containing 94 mean segment durations (on which was based the so-called *phoneme-type* model). By looking up appropriate entries in these tables, given a syllable comprising particular segments, a duration could be predicted for that syllable according to either a phoneme-class model or to a phoneme-type model. This modelling exercise was carried out. A similar exercise was also carried out for both kinds of rhythmic unit except that, for these, a further correction—namely a diminution of segment duration in proportion to rhythmic unit size—was also applied. The reduction did not, of course, produce anywhere near complete compensation for duration, just the 9% or so reducation we found in the analysis. In all cases the durations predicted by the models were paired for comparison with the measured durations. Finally, the syllable durations were also predicted according to a syllable-type model that used eight numbers (four types: *strong, weak, enclitic* and *proclitic*; by two categories *marked* and *unmarked*) to predict syllable durations which were also pair-wise compared with the measured durations.

We found that this syllable-type model accounted for only 47% of the overall variance in mean syllable duration. This exceeds the percentage accounted for using a basis of only syllable type, in either marked or unmarked categories alone (39% and 34% respectively), but not by the full 16% of the total variance that is apparently accounted for by the marked versus unmarked distinction. This is a typical effect of non-independence of factors.

Interestingly, we found little difference between the syllable durations predicted (modelled) and observed, using the phoneme-level models, regardless of whether we used the phoneme-class or the phoneme-type model in predicting these syllable durations. The former accounted for about 74% of the variance, and the latter for about 78%. Thus both low-level (i.e. segment-based) models of syllable duration are considerably better at predicting syllable duration than the syllable-type model, the crude phoneme-class-based model being only slightly worse than the more detailed phoneme-type-based model.

At the rhythmic unit level of analysis, there were obvious similarities between the results for (Abercrombie/Halliday) foot-oriented duration predictions and predictions based on (Jassem) rhythm units. For unmarked units, the phoneme-class model accounted for about 70% of the variation in either kind of rhythmic unit duration, and the type model for about 76% of the variation. Correcting the durations for the isochrony effect, using a simple linear regression, only bettered these figures by about 1%. For marked feet, the picture was significantly different. The phoneme-class model accounted for only about 40% of the variance in rhythmic unit duration, and the type model for about 48%. However, adding the isochrony factor improved these figures by about 15%. It is also true that the rhythm unit theory (Jassem) seemed a rather more accurate predictive base for marked units than the foot-based theory (Abercrombie/Halliday), the figures above being averaged over the two theories.

However, the difference in isochrony effect between anacruses (negligible) and narrow rhythm units (strong) showed up as highly significant (Jassem, Hill, Witten 1984), and is masked when the foot-based framework is used. This is certainly a reason for preferring to use the Jassem theory in modelling British English rhythm.

Conclusions

Conclusions from this comparitively limited initial study must necessarily be tentative and hedged with qualifications about the effects of small sample sizes and large numbers of statistical inferences, drawn from a fairly small body of data. It is partly considerations of this nature that have inhibited, temporarily, attempts at further refinement of our modelling of rhythmic structure, based on our present data. However, all the effects reported, unless otherwise stated, were significant at the 1% level. It should be remembered that the spectrographic analysis, followed by checking through several stages of integrity and statistical analysis, even

though assisted by computer, represents a prodigious amount of labour, and we should welcome parallel studies by other research groups.

We conclude that, in assigning duration to the basic elements (segments) of speech, for purposes of speech synthesis by computer, the primary factor (within the limitations of this initial study) is the type of phoneme represented by the segment (45% of the variance). A further 15% of the variance is, we found, accounted for by the type of rhythmic unit into which the segment falls (marked or unmarked). A further 9% of the variance is accounted for by a requirement to shorten mean segment duration somewhat, in proportion to the size of (number of segments in) a rhythmic unit, but is predominantly associated with marked rhythmic units, in which it accounts for 15% of the variance in unit duration.

Other factors play a minor role, but are more obviously non-independent, so that we have accounted for at most 70% or so of the contributions to choosing mean segment duration (and hence rhythmic structure) in spoken British English. Using only the three main factors mentioned above, we found that it was possible to model (predict) the durations of rhythmic units according to either of the major theories of British English speech rhythm, and acount for about 76% of the variance in unmarked rhythmic units, and about 63% of the variance in marked rhythmic units. It was much more important to take into account the tendency towards isochrony in marked rhythmic units than in unmarked rhythmic units. We found no evidence of any syllable-timed component of rhythm in spoken British English.

For computer speech synthesis, therefore, a table containing two sets of mean phoneme-type segment durations, one for segments falling in marked rhythmic units and one for those falling in unmarked rhythmic units, coupled with simple linear regression correction of duration for marked feet, may be expected to produce reasonable approximations to the rhythm of spoken British English. However, at least 25% of the variance in rhythmic unit duration, and hence of the determinants of the rhythmic structure remain unaccounted for. This may be due partly to natural variation, partly experimental errors of various kinds, and partly due to missing portions of the theoretical framework for analysis, such as O'Hala's comb model of speech production pre-planning, or the effect we observed in which some rhythmic units were clearly marked, but apparently only by pitch movement.

Acknowledgements

The authors wish to acknowledge, with gratitude, the supportof the National Research Council of Canada for this work under grant number A5261.

Accompanying material

At the original presentation of this paper to the 94th.meeting of the Acoustical Society of America, tape recordings of nine speech utterances were played to demonstrate the practical results of using the reported model for synthetic speech. The utterances comprised the first nine utterances from Halliday's Study Unit 30, and represented: the original; speech synthesised by segmental rules, with copied rhythm and intonation; speech synthesised by segmental rules plus rules embodying a simple interpretation of isochrony for rhythm, and a simple form of Halliday's recommendations for intonation; and, finally, speech synthesised by segmental rules,with the intonation copied, but with the rhythmic structure generated according to the model specified in this paper and its appendix. Copies of this recording, with the order arranged to facilitate comparitive listening, may be obtained by writing to the first author, and may soon be available on the internet.

References

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

HALLIDAY MAK (1970) *Spoken English: Intonation*. Oxford University Press: Oxford

HILL DR, JASSEM W & WITTEN IH (1978) A statistical approach to the problem of isochrony in spoken British English. *Research Report Number 78/27/6*, Dept. of Computer Science, U. of Calgary (Paper presented to the 1977 meeting of the International Society of Phonetic Sciences, Miami Beach, Florida,December 1977, by invitation)

HILL DR, MANZARA L & SCHOCK C (1995) Real-time articulatory speech-synthesis-by-rules. *Proc. AVIOS 95 14th Int Voice Technologies Applications Conference*, San Jose, pp 27-44

JASSEM W (1952) Stress in modern English. *Bulletin de la Société Linguistique Polonaise* **XII**, 189-194

JASSEM W, HILL DR & WITTEN IH (1984) Isochrony in English speech: its statistical validity and linguistic relevance. in *Intonation, Accent and Rhythm: Studies in Discourse Phonology* (eds. D. Gibbon & H. Richter), New York: Walter de Gruyter, 203-225

JONES D (1918) *An Outline of English Phonetics.* (pp 237-242, 9th. edition), Heffer: Cambridge, U.K.
LADEFOGED P (1975) *A Course in Phonetics*. (pp 102-103), Harcourt, Brace, Jovanovich: New York
LEHISTE I (1973) Rhythmic units and syntactic units in production and perception. *J. Acoustical Soc. Amer.* 54 (5) 1228-1234
PIKE KL (1945) *Intonation in American English.* U. Michigan Press: Ann Arbor (reprinted 1960)
WILLIAMS B & HILLER SM (1994) The question of randomness in English foot timing: a control experiment. J *Phonetics* **22**, 423-439
WENK BJ & WIOLAND F (1982) Is French really syllable-timed? *J Phonetics* **10**, 193-216.

APPENDIX 1: RESULTS OF ANALYSIS OF UTTERANCES SU3O AND SU39

The following tables and comments represent a summary digest of some of the main statistics arising from and used in this study. It will be noticed that, at the various levels of analysis, out of all the units at that level (segments, syllables, etc.), only a certain percentage of those available were used. This reflects a number of factors, including the fact that some units were broken by hesitations, and some segments (e.g. initial and final stops, and some glide-vowel combinations) could not be assigned durations consistent with the rest of the analysis. Thus the closure period of an initial stop was indeterminate, and they were not analysed separately. For such segments, the higher level units containing them could not be used either. Such problems are inherent in traditional phonetic analysis.

The appendix falls into four main divisions:
(a) Segment level analysis;
(b) Syllable level analysis and modelling;
(c) Foot and rhythm unit analysis and modelling; and
(d) Perspective overview.

In each of the first three sections, the emphasis is on the attempt to apportion independent sources of variance for the units concerned. In (b) and (c) the results of modelling experiments at those levels are also summarised.

The authors are aware of the limitations inherent in modelling the rhythm of utterances, based on statistics derived from those same utterances, but argue that the modelling is a fair test of the extent to which relevant, independent factors have been isolated. It also demonstrates clearly that, even when thus constrained, some factors are still apparently undetermined.

More importantly, synthetic speech generated using the preferred model, based on statistics from Study Unit 30, has a very acceptable rhythm that is preferred by listeners in formal listening trials, including numerous independent judgements based on comparisons with alternative methods. One immediate goal is to generate synthetic utterances, using the model and the original data in paired comparison tests to determine whether listeners are able to distinguish between the two kinds of rhythm.

The rhythm model based on phoneme types and Abercrombie/Halliday feet was used as the basis for computing rhythm in the Trillium Sound Research Inc. "TextToSpeech" kit developed for the NeXT computer and NeXTSTEP operating system. This software is now available at no cost under a General Public Licence from the Free Software Foundation website at http://savannah.gnu.org/projects/gnuspeech.

As noted in the summary, and the body of the paper, independent sources for approximately 70% of the variation in segment duration were found—45% due to segment identity, 15% to whether the segment was in a marked or an unmarked rhythmic unit, and 9% due to what was seen as an isochrony effect.

It should be noted that, since the effect of vowel types and rhythmic unit types were accounted for indepenently in seeking sources of variance, the usual objection that a tendency towards isochrony is somhow an artifact of the relation between tense vowels, syllable types, etc. was circumvented. The tendency towards isochrony found in rhythmic units—accounting for around 9% of the variation in segment duration—was independent of the segment and syllable types.

Williams and Hiller (1994—see references), tackling the same question, but using a rather different approach, came to a very similar conclusion. That there is a linguistically mediated tendency towards isochrony based on stressed/accented syllables in the resulting "feet" delineated by these beats"".

The results we obtained are tabulated in the following pages.

Notes relevant to all tables

0% is highly significant; 100% is totally insignificant. Figure which are significant at the 5% level or better are underlined. Significances are computed by analyses of variance or by 2-tailed t-tests, as appropriate. In the latter case a pooled estimate of variance is used.

Numbers given in parenthesis after means show the number of elements which contributed towards the mean.

Throughout the data listings, the convention "msec2" is used to represent "$msec^2$".

GENERAL

	SU30	**SU39**
length of speech sample (excluding hesitations and pauses)	95 secs	122 secs
number of utterances	40	49
number of tonegroups	69	80
number of feet	230	264
number of feet used in analysis	156 (68%)	161 (61%)
number of rhythm units	230	264
number of rhythm units used in analysis	162 (70%)	166 (63%)
number of anacruses	117	128
number of anacruses used in analysis	77 (66%)	60 (47%)
number of syllables	529	636
number of syllables used in analysis	449 (85%)	461 (71%)
number of segments	1348	1623
number of segments used in analysis	1146 (85%)	1372 (85%)
mean segment duration	71 msec	75 msec
mean segment rate (segments/see)	14.1	13.3
mean syllable duration	166 msec	168 msec
mean syllable rate (syllables/see)	6.0	6.0
mean rhythm unit duration	311 msec	356 msec
mean rhythm unit rate (rhythm units/see)	3.2	2.8
mean anacrusis duration	138 msec	181 msec
mean anacrusis rate (anacruses/sec)	7.2	5.5
mean foot duration	384 msec	423 msec
mean foot rate (feet/sec)	2.6	2.4
variance in segment duration	1380 msec2	1830 msec2
variance in segment duration—rhythm units (no proclitics)	1558 msec2	2011 msec2
variance in syllable duration	8220 msec2	8040 msec2
variance in rhythm unit duration	12630 msec2	18310 msec2
variance in anacrusis duration	4740 msec2	9400 msec2
variance in foot duration	15210 msec2	25050 msec2
standard deviation of segment durations	37 msec	43 msec
normalized to percentage of mean	52%	57%
standard deviation of syllable durations	91 msec	90 msec
normalized to percentage of mean	55%	54%
standard deviation of rhythm unit durations	112 msec	135 msec
normalized to percentage of mean	36%	38%
standard deviation of anacrusis durations	69 msec	97 msec
normalised to percentage of mean	50%	54%
standard deviation of foot durations	123 msec	158 msec
normalised to percentage of mean	33%	37%

SEGMENT DURATIONS	**SU30**	**SU39**
Segment durations: utterance effects variance in segment durations	1380 msec2	1830 msec2
between utterance variance	59 msec2	160 msec2
expressed as a percentage of total variance	4%	9%
significance of between utterance variation	12%	**0.0%**

	Foot effects		**Rhythm unit effects**	
	SU30	**SU39**	**SU30**	**SU39**
Segment durations: rhythmic unit effects				
mean segment durations (in msecs)				
(a) tonic and utterance-final rhythmic units	100 (131)	99 (171)	100 (131)	99 (169)
(b) non-tonic but utterance final rhythmic units	112 (25)	102 (43)	112 (25)	102 (43)
(c) tonic but non-final rhythmic units	80 (244)	80 (343)	88 (176)	86 (218)
(d) non-tonic, non-final rhythmic units	61 (747)	67 (815)	66 (506)	73 (539)
variance accounted for by this breakdown	209 msec2	138 msec2	220 msec2	119 msec2
expressed as a percentage of total variance	15 %	8 %	14 %	6 %
significance of between group variation	**0.0 %**	**0.0 %**	**0.0 %**	**0.0 %**
significance of difference between (a) and (b)	24 %	77%	24 %	79 %

NOTE: Tests on Rhythm Units were repetitions of those done on feet, after removing all proclitic syllables. Utterance final RUs produce identical results to Foot Units since such units cannot contain proclitic syllables. Both sets of figures are given above for completeness though they are quite similar.

Consider rhythmic units which are:
1. not the first of the utterance;
2. not tonic;
3. not final

and divide them into pre-tonic units and post-tonic units.

	Foot effects		**Rhythm unit effects**	
Segment durations				
between these all these units	79 %	**0.0 %*** **	98 %	15 %
between pre-tonic units	99 %	**4 %*** **	100 %	24 %
between post-tonic units	24 %	8 %	59 %	31 %
between all pre-tonic units & all post-tonic units	61 %	**0.0 %*** **	96 %	7 %
mean segment duration (in msec)				
within pre-tonic units	63 (332)	63 (414)	67	69
within post-tonic units	64 (274	72 (389)	67	75
mean segment duration (in msec) in rhythmic units:				
having 1 segment	170 (2)	168 (3)	172 (3)	168 (3)
having 2 segments	116 (19)	111 (23)	105 (32)	119 (43)
having 3 segments	87 (84)	100 (64)	92 (123)	96 (96)
having 4 segments	81 (81)	90 (96)	81 (137)	95 (123)
having 5 segments	80 (115)	76 (140)	79 (120)	74 (150)
having 6 segments	73 (181)	72 (193)	70 (181)	76 (178)
having 7 segments	70 (175)	74 (140)	68 (90)	79 (117)
having 8 segments	60 (155)	71 (132)	63 (44)	75 (101)

*** These figures suggest a significant slowing following the tonic in SU39 tone groups, as opposed to SU30. It may reflect the use of a wider range of rhythmic resources in conversation, but does not show up for Rhythm Units so is, presumably, largely confined to proclitic syllables. (See als "Syllable effects" in the tables).

having 9 segments	63 (158)	73 (153)	63 (79)	75 (92)
having 10 segments	62 (77)	68 (197)	66 (30)	66 (68)
having 11 segments	63 (53)	71 (84)	-	76 (22)
having 12 segments	57 (38)	68 (28)	-	54 (7)
having 13 segments	-	72 (46)	-	-
having 14 segments	-	73 (51)	-	-
having 15 segments	68 (11)	-	-	-
having 16 segments	-	57 (22)	-	-
variance accounted for by these groups	120 msec2	104 msec2	147 msec2	170 msec2
expressed as a percentage of total variance	9%	6%	9%	8%
significance of between group variance	**0.0%**	**0.0 %**	**0.0%**	**0.0%**

SEGMENT DURATIONS: SYLLABLE EFFECTS

	SU30	**SU39**
mean segment duration (in msec) in:		
(a) strong syllables	90 (356)	94 (434)
(b) weak syllables	70 (188)	77 (222)
(c) enclitic syllables	65 (293)	69 (342)
(d) proclitic syllables	54 (310)	57 (374)
variance accounted for by this breakdown	199 msec2	200 msec2
expressed as a percentage of total variance	14 %	11 %
significance of between-group variation	**0.0 %**	**0.0 %**
significance of difference between (a) and (b)	12 %	**1.2 %**
mean segment durations broken down by size of syllable:		
strong syllables		
with 1 segment	154 (6)	160 (6)
with 2 segments	104 (63)	113 (61)
with 3 segments	89 (161)	91 (212)
with 4 segments	78 (101)	88 (142)
with 5 segments	96 (25)	88 (14)
weak syllables		
with 2 segments	61 (24)	80 (34)
with 3 segments	70 (140)	79 (167)
with 4 segments	77 (25)	60 (22)
enclitic syllables		
with 1 segment	93 (1)	78 (6)
with 2 segments	71 (97)	71 (104)
with 3 segments	62 (178)	69 (190)
with 4 segments	61 (18)	67 (40)
with 5 segments	-	50 (4)
proclitic syllables		
with 1 segment	84 (10)	67 (21)
with 2 segments	51 (155)	55 (157)
with 3 segments	55 (137)	59 (183)
with 4 segments	58 (8)	57 (11)

mean segment durations (in msecs) for:		
(a) consonants in syllable-initial position	65 (362)	69 (412)
(b)syllable nucleii	80 (481)	83 (547)
(c) consonants in syllable-final position	62 (272)	70 (352)
variance accounted for by this breakdown	68 msec2	38 msec2
expressed as a percentage of the total variance	5%	2%
significance of between group variation	**0.0%**	**0.0 %**
significance of difference between (a) and (c)	24%	80%
mean segment durations (in msec) for:		
(a) single syllable-initial consonants	65 (316)	70 (350)
(b) consonants in syllable-initial clusters	62 (46)	64 (63)
(c) single syllable-final consonants	61 (217)	72 (272)
(d) consonants in syllable-final clusters	66 (55)	64 (80)
significance of difference betweem (a) and (b)	50%	19%
significance of difference between (c) and (d)	32%	11%
mean segment durations (in msec), considering only foot-initial syllables, for:		
(a) single syllable-initial consonants	88 (48)	91 (67)
(b) consonants in syllable-initial clusters	71 (18)	76 (20)
(c) single syllable-final consonants	81 (39)	88 (56)
(d) consonants in syllable-final clusters	80 (11)	81 (11)
significance of difference betweem (a) and (b)	7%	14%
significance of difference between (c) and (d)	89%	64%
mean segment durations (in msec) in:		
(a) open syllable nucleii	78 (172)	82 (173)
(b) nucleii with unvoiced termination	82 (105)	85 (120)
(c) nucleii with voiced termination	82 (204)	82 (254)
significance of between-group variation	59%	80%
mean segment durations (in msec), for segments in foot-initial syllables, in:		
(a) open syllable nucleii	147 (20)	173 (18)
(b) nucleii with unvoiced termination	119 (18)	113 (27)
(c) nucleii with voiced termination	118 (37)	115 (47)
significance of between-group variation	9%	**0.2%**
mean segment durations (in msec), for long vowels in:		
(a) open syllable nucleii	107 (27)	116 (35)
(b) nucleii with unvoiced termination	101 (14)	113 (14)
(c) nucleii with voiced termination	124 (33)	107 (33)
significance of between-group variation	37%	81%
mean segment durations (in msec), for long vowels		
in foot-initial syllables, tonic feet only, in:		
(a) open syllable nucleii	153 (10)	163 (9)
(b) nucleii with unvoiced termination	148 (3)	106 (5)
(c) nucleii with voiced termination	156 (9)	160 (7)
significance of between-group variation	97%	25%

SEGMENT DURATIONS: SEGMENT LEVEL EFFECTS

	SU30	**SU39**
mean segment durations (in msec) by broad phonetic category		
(a) short vowels	63 (312)	64 (360)
(b) long vowels	114 (75)	111 (82)
(c) diphthongs	124 (73)	139 (78)
(d) glides	55 (114)	57 (140)
(e) nasals	58 (145)	67 (166)
(f) unvoiced plosives	69 (141)	80 (174)
(g) voiced plosives	57 (71)	57 (91)
(h) unvoiced fricatives	86 (95)	98 (96)
(i) voiced fricatives	50 (86)	57 (139)
(j) affricates	108 (15)	116 (26)
(k) aspirate (H)	58 (18)	67 (20)
variance accounted for by this breakdown	443 msec2	511 msec2
expressed as a percentage of the total variance	32%	28%

NOTE: most of these classes are not statistically homogeneous—at the 5% level of significance. The only exceptions to this are diphthongs and glides. For these classes, the within-class significance of the differences between segment durations are:

diphthongs	67%	10%
glides	60%	31%

For part of the work, we modelled segment durations on the basis of a segment class model which used the above classes, plus a distinction between: (a) segments in marked feet (tonic, final or both); and (b) unmarked feet (the remainder). The durations are as follows:

PHONEME CLASS MODEL

(22 durations categories per study unit dataset)

	SU30		**SU39**	
	unmarked	**marked**	**unmarked**	**marked**
short vowel	55.4	89.4	59.3	78.1
long vowel	88.7	160.5	96.0	138.1
diphthong	109.8	151.8	121.0	155.9
glide	48.2	67.8	49.2	66.6
nasal	50.3	74.7	61.8	78.5
unvoiced plosive	63.7	84.3	71.4	95.5
voiced plosive	51.1	71.3	53.3	62.8
unvoiced fricative	74.3	109.7	86.4	118.8
voiced fricative	47.7	68.6	51.6	74.6
affricate	107.1	106.9	109.3	134.1
aspirate (H)	49.7	66.5	61.4	76.1

variance accounted for by segment class breakdown	668 msec2	633 msec2
expressed as a percentage of the total variance	48%	35%

A more detailed breakdown was based on individual segment categories—approximating to phoneme categories, with some allophones (aspirated [p^h, t^h, k^h, b^h, d^h]) classified separately. This gave a total of 47 categories. The actual breakdown is omitted to save space, as the variance accounted for was comparable to the class model above

	SU30	**SU39**
variance accounted for by 47 type segment breakdown	624 msec2	723 msec2
expressed as a percentage of the total variance	45%	37%

The 47 type breakdown led to a more detailed model called the segment-type model, the essence of which is a dataset having 94 durations—one for each segment type (47) for both marked and unmarked feet (47 x 2 = 94). These datasets for both SU30 and SU 39 appear in the next table.

SEGMENT TYPE MODEL
(94 duration categories per study unit data-set)

(The notation PX below represents that realisation of the /p/ phonem exhibiting an extended aspirated and/or fricated region following release. Similar notation for the other stops have the same meaning.)

		SU30		**SU39**	
Segment ID	IPA symbol	**unmarked**	**marked**	**unmarked**	**marked**
AA	æ	84.1	122.1	81.4	111.4
AH	a	65.4	118.5	72.3	43.3
A	ʌ	77.0	132.4	70.9	87.7
E	ɛ	61.0	77.7	75.8	86.1
I	ɪ	53.3	76.3	60.9	73.1
O	ɔ	70.5	103.9	78.0	102.5
UH	ə	46.2	74.1	45.0	57.2
U	ʊ	51.0	-	43.3	92.5
AR	ɑ	106.8	181.7	135.0	183.3
AW	ɔː	114.1	194.1	89.5	139.5
EE	i	82.4	141.8	83.8	97.9
ER	əː	132.2	167.7	99.8	140.7
UU	u	63.7	125.0	103.0	143.0
AH-I	æɪ	117.7	186.7	117.0	151.9
AH-UU	æʊ	112.4	148.2	109.2	135.7
E-I	ɛɪ	99.0	132.1	128.2	108.5
O-I	oɪ	-	135.0	90.7	213.1
UH-UU	ʌʊ	104.8	168.0	136.6	168.3
R	r	40.3	70.7	41.8	53.6

W	w	47.4	68.6	42.3	74.3
L	l	51.6	60.3	55.8	70.5
Y	j	54.5	84.4	53.5	71.7
M	m	49.7	94.3	57.2	84.1
N	n	50.6	65.9	59.5	77.1
NG	ŋ	49.7	68.0	78.7	66.7
P	p	76.0	81.6	54.3	91.3
T	t	49.7	70.3	53.7	63.9
K	k	62.7	83.5	63.2	55.6
PX	p^h	89.4	116.2	100.1	147.8
TX	t^h	89.9	107.0	94.9	115.5
KX	k^h	87.2	114.6	83.0	112.9
B	b	61.0	70.7	64.5	59.8
D	d	45.6	73.2	43.3	59.4
G	g	53.3	68.2	53.5	53.8
BX	b^h	73.9	-	-	-
DX	d^h	-	65.4	73.2	98.8
F	f	70.1	97.9	86.0	76.8
TH	θ	54.6	116.1	74.4	78.8
S	s	78.1	111.5	86.1	132.1
SH	ʃ	63.8	124.2	115.3	108.7
V	v	48.4	68.9	51.2	47.5
DH	ð	41.3	41.1	43.4	57.8
Z	z	56.5	84.8	60.8	94.8
ZH	ʒ	-	-	58.0	-
CH	tʃ	118.5	118.8	129.6	130.3
J	dʒ	93.4	100.0	75.1	164.6
H	h	49.7	66.5	61.4	76.1

Variance accounted for by type breakdown:	825 msec2	857 msec2
As a percentage of total variance	60%	47%

SUMMARY OF VARIANCE ACCOUNTED FOR BY THE VARIOUS SEGMENT LEVEL MODELS

	SU30	**SU39**
The marked/unmarked distinction accounts for	224 msec2	151 msec2
Expressed as a percentage of the total variance	16%	8%
The class model (11 categories)	443 msec2	511 msec2
Expressed as a percentage of the total variance	32%	28%
The class model plus the marked/unmarked distinction (22 categories)	668 msec2	633 msec2
Expressed as a percentage of the total variance	48%	35%
The type model (47 categories)	624 msec2	723 msec2
Expressed as a percentage of the total variance	45%	37%
The type model plus the marked/unmarked distinction (94 categories)	825 msec2	857 msec2
Expressed as a percentage of the total variance	60%	47%

SYLLABLE DURATIONS AND MODELLING

SU30 measured mean syllable durations (msec)	**strong**	**weak**	**enclitic**	**proclitic**
(a) in tonic final feet	315 (17)	226 (6)	227 (17)	-
(b) in non-tonic, final feet	460 (4)	135 (1)	-	-
(c) in tonic, non-final feet	276 (24)	215 (17)	182 (24)	123 (32)
(d) in unmarked feet	207 (72)	152 (43)	127 (76)	106 (116)
significance of difference between (a), (b) and (c) (marked); analysis of variance:	**0.5%**	61%	9%	-

SU39 measured mean syllable durations (msec)	**strong**	**weak**	**enclitic**	**proclitic**
(a) in tonic final feet	298 (9)	187 (10)	95 (5)	-
(b) in non-tonic, final feet	225 (2)	135 (1)	-	-
(c) in tonic, non-final feet	269 (36)	218 (16)	169 (39)	124 (38)
(d) in unmarked feet	239 (64)	171 (42)	134 (75)	120 (124)
significance of difference between (a), (b) and (c) (marked); analysis of variance:	53%	38%	**1.3%**	-

There are some significant differences between syllable durationsin the different types of marked feet, but there is no discernable pattern and no obvious explanatory theory. Lumping all the data from marked feet together (combining (a), (b) and (c)), we obtain two simpler tables.

SU30 measured mean syllable durations (msec)	**strong**	**weak**	**enclitic**	**proclitic**
(a) in marked feet	307 (45)	214 (24)	201 (41)	123 (32)
(b) in unmarked feet	207 (72)	152 (43)	127 (76)	106 (116)
significance of difference (2-tailed t-test):	**0.0%**	**0.1%**	**0.0%**	5.4%
SU39 measured mean syllable durations (msec)				
(a) in marked feet	273 (47)	203 (27)	161 (44)	124 (38)
(b) in unmarked feet	239 (68)	171 (42)	134 (75)	120 (124)
significance of difference (2-tailed t-test):	6.4%	**4.8%**	**1.7%**	73%

These figures formed the basis of the syllable-type model for syllable durations.

The measured syllable durations compare with those computed from the phoneme-type and phoneme-class model as follows:

SU39: comparison of measured versus computed syllable durations in unmarked feet (msec)	**strong**	**weak**	**enclitic**	**proclitic**
a. measured	239 (64)	171 (42)	134 (75)	120 (124)
b. phoneme-type model	228 (64)	170 (42)	133 (75)	127 (124)
c. phoneme-class model	223 (64)	156 (42)	133 (75)	132 (124)
significance of difference between (a) and (b) (2-tailed paired t-test)	17%	83%	73%	17%
significance of difference between (a) and (c) (2-tailed paired t-test)	5.9%	**2.9%**	82%	**1.7%**

In fact, four tables like this were calculated, for:

* utterance SU30: marked feet
* utterance SU30: unmarked feet
* utterance SU39: marked feet
* utterance SU39: unmarked feet

The one shown is representative of the agreement obtained between the measured syllable durations and those calculated according to the models.

The models account for the following amounts of the variance in syllable duration:

	SU30	**SU39**
unmarked feet:		
variance in syllable durations	4610 msec2	7310 msec2
variance due to syllable types	1550 msec2	2140 msec2
expressed as a percentage	34%	29%
variance accounted for by phoneme-class model	3310 msec2	4180 msec2
expressed as a percentage	72%	57%
variance accounted for by phoneme-type model	3600 msec2	4550 msec2
expressed as a percentage	78%	62%
marked feet:		
variance in syllable durations	11920 msec2	8610 msec2
variance due to syllable types	4650 msec2	3400 msec2
expressed as a percentage	39%	39%
variance accounted for by phoneme-class model	8010 msec2	5980 msec2
expressed as a percentage	67%	69%
variance accounted for by phoneme-type model	8410 msec2	6660 msec2
expressed as a percentage	71%	77%

(continued on next page)

overall:

variance in syllable durations	8220 msec2	8040 msec2
variance due to marked/unmarked distinction	1290 msec2	840 msec2
expressed as a percentage	16%	10%
variance accounted for by syllable type model	3830 msec2	2860 msec2
expressed as a percentage	47%	36%

We conclude:

* both phoneme models are good models of syllable duration
* they are much better than modelling syllable durations on the basis of syllable type
* the phoneme-class model is nearly as good as the phoneme-duration model (at least as far as modelling syllable duration is concerned).

FOOT DURATIONS AND MODELLING

SU30: duration of unmarked feet	**number of segments in foot**										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
measured duration	170	193	210	262	283	363	396	382	456	537	492
sample size	2	1	12	7	12	19	15	15	9	3	5
phoneme-type model	111	150	193	265	287	360	383	400	442	484	504
discrepancy	59	43	17	-3	-3	4	12	-19	14	53	-12
phoneme-class model	110	161	191	264	284	350	382	409	441	497	499
discrepancy (D)	61	32	19	-2	-1	14	14	-27	15	40	-7

SU39: duration of unmarked feet	**number of segments in foot**												
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**14**
measured duration	116	150	304	284	307	346	379	482	563	613	684	673	706
sample size	1	4	8	8	16	16	8	3	10	6	1	2	1
phoneme-type model	117	153	232	263	303	347	401	485	531	615	654	742	824
discrepancy (D)	-1	-3	72	21	4	-1	-22	-3	32	-2	30	-69	-118
phoneme-class model	121	160	225	252	298	351	396	472	532	608	616	715	837
discrepancy (D)	-5	-10	79	32	9	-5	-17	10	31	5	68	-42	-131

Similar tables were produced for marked feet. The generally decreasing trend of the discrepancies can be seen from the attached graphs, which show them for both for marked and unmarked feet for both models, SU30. The equivalent rhythm unit graphs also appear.

Both the phoneme-class model and the phoneme-type model of foot duration were combined with a straight-line regression (See below), which corrected for the decreasing discrepancy with increasing foot size by weighting the phoneme model durations proportionally to the number of segments in the foot, in a manner which decreased the resulting variance as much as is possible with this simple linear correction. The variances accounted for were as follows:

Portioning out variance	**SU30**	**SU39**
unmarked feet:		
variance in foot durations	13160 msec2	26170 msec2
variance accounted for by phoneme-class model	9320 msec2	18600 msec2
expressed as a percentage	71%	71%

variance accounted for by phoneme-class model corrected by a linear regression with foot size	9360 msec2	18760 msec2
expressed as a percentage	71%	72%
variance accounted for by phoneme-type model	10050 msec2	19530 msec2
expressed as a percentage	76%	75%
variance accounted for by phoneme-type model corrected by a linear regression with foot size	10070 msec2	19730 msec2
expressed as a percentage	77%	75%
marked feet:		
variance in foot durations	13360 msec2	20970 msec2
variance accounted for by phoneme-class model	4610 msec2	7531 msec2
expressed as a percentage	35%	36%
variance accounted for by phoneme-class model corrected by a linear regression with foot size	7210 msec2	11760 msec2
expressed as a percentage	54%	56%
variance accounted for by phoneme-type model	5490 msec2	11580 msec2
expressed as a percentage	41%	55%
variance accounted for by phoneme-type model corrected by a linear regression with foot size	7570 msec2	13470 msec2
expressed as a percentage	57%	64%

RHYTHM UNIT DURATIONS AND MODELLING

SU30: duration of unmarked rhythm units	**number of segments in rhythm unit**

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
measured duration	170	200	221	256	267	317	383	428	473
sample size	2	4	32	16	16	20	6	4	2
phoneme-type model	111	163	199	258	267	311	394	440	439
discrepancy (D)	59	37	22	-2	0	5	-11	-12	34
phoneme-class model	110	167	191	250	262	307	385	441	419
discrepancy (D)	61	33	30	6	6	9	-2	-13	54

SU39: duration of unmarked rhythm units	**number of segments in rhythm unit**

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
measured duration	116	215	232	286	310	349	400	457	484
sample size	1	10	15	14	21	12	8	5	1
phoneme-type model	117	154	207	282	300	359	398	509	587
discrepancy (D)	-1	61	25	4	10	-10	2	-52	-103
phoneme-class model	121	153	196	270	287	356	399	470	574
discrepancy (D)	-5	62	36	16	23	-7	1	-13	-90

Similar tables were produced for marked rhythm units.

Both the phoneme-class model and the phoneme-type model of rhythm unit duration were combined with a straight-line regression (See below), which corrected for the decreasing discrepancy with increasing rhythm unit size by weighting the phoneme model durations proportionally to the number of segments in the unit, in a manner which decreased the resulting variance as much as is possible with this simple linear correction.

The variances accounted for were as follows:

Portioning out variance	**SU30**	**SU39**
unmarked rhythm units:		
variance in rhythm unit durations	8646 msec2	11850 msec2
variance accounted for by phoneme-class model	6089 msec2	6451 msec2
expressed as a percentage	70%	54%
variance accounted for by phoneme-class model cor-		
rected by a linear regression with rhythm unit size	6170 msec2	6884 msec2
expressed as a percentage	71%	58%
variance accounted for by phoneme-type model	6475 msec2	7054 msec2
expressed as a percentage	75%	60%
variance accounted for by phoneme-type model		
corrected by a linear regression with rhythm unit size	6561 msec2	7619 msec2
expressed as a percentage	76%	64%
marked rhythm units:		
variance in rhythm unit durations	11720 msec2	19170 msec2
variance accounted for by phoneme-class model	5658 msec2	7228 msec2
expressed as a percentage	48%	38%
variance accounted for by phoneme-class model		
corrected by a linear regression with rhythm unit size	7151 msec2	10740 msec2
expressed as a percentage	61%	56%
variance accounted for by phoneme-type model	6506 msec2	10770 msec2
expressed as a percentage	56%	56%
variance accounted for by phoneme-type model		
corrected by a linear regression with rhythm unit size	7851 msec2	12340 msec2
expressed as a percentage	67%	64%

PERSPECTIVE

How much of the variance in segment, syllable, and foot durations has been tentatively accounted for?

UTTERANCE SU30	**segment**	**syllable**	**foot**
total variance	1380 msec2	8220 msec2	15210 msec2
phoneme-class model:			
residual variance	712 msec2	2120 msec2	5560 msec2
variance accounted for	668 msec2	6100 msec2	9650 msec2
expressed as percentage	48%	74%	63%
hypothesized error (15 msec)	225 msec2	225 msec2	225 msec2
variance accounted for (with error)	893 msec2	6325 msec2	9875 msec2
expressed as percentage	65%	77%	65%
phoneme-type model:			
residual variance	555 msec2	1790 msec2	4770 msec2
variance accounted for	825 msec2	6430 msec2	10440 msec2

expressed as percentage	60%	78%	69%
hypothesized error (12 msec)	144 msec2	144 msec2	144 msec2
variance accounted for (with error)	969 msec2	6574 msec2	10584 msec2
expressed as percentage	70%	80%	70%
phoneme-type model with regression:			
residual variance	-	-	4030 msec2
variance accounted for	-	-	11180 msec2
expressed as percentage	-	-	74%
hypothesized error (12 msec)	-	-	144 msec2
variance accounted for (with error)	-	-	11324 msec2
expressed as percentage	-	-	74%

UTTERANCE SU39	**segment**	**syllable**	**foot**
total variance	1830 msec2	8040 msec2	25050 msec2
phoneme-class model:			
residual variance	1197 msec2	2950 msec2	10310 msec2
variance accounted for	633 msec2	7040 msec2	14740 msec2
expressed as percentage	35%	70%	59%
hypothesized error (22 msec)	484 msec2	484 msec2	484 msec2
variance accounted for (with error)	1117 msec2	7524 msec2	15224 msec2
expressed as percentage	61%	75%	61%
phoneme-type model:			
residual variance	973 msec2	2480 msec2	7900 msec2
variance accounted for	857 msec2	7510 msec2	17150 msec2
expressed as percentage	47%	75%	68%
hypothesized error (20 msec)	400 msec2	400 msec2	400 msec2
variance accounted for (with error)	1257 msec2	7910 msec2	17550 msec2
expressed as percentage	69%	79%	70%
phoneme-type model with regression:			
residual variance	-	-	6900 msec
variance accounted for	-	-	18150 msec2
expressed as a percentage	-	-	72%
hypothesized error (20 msec)	-	-	400 msec
variance accounted for (with error)	-	-	18550 msec2
expressed as a percentage	-	-	74%

It can be seen that each model accounts for less of the segment duration variance than it does of the foot-duration variance. However, this can be explained by postulating a small error in measuring the duration of each of the segments. Under the (reasonable) assumption that this error is a random variable independent of segment durations, its variance can simply be added to the variance accounted for, in each case. The error figures given above are calculated to make the percentage duration accounted for in the case of segments and feet coincide, and varies from 12 to 22 msecs. However, the percentage duration of syllables accounted for is much greater in both study units than the percentage segment and foot durations accounted for. This is an anomaly that we cannot yet explain, though it could simply reflect a relatively better rule for syllable division than for segment or rhythmic unit division.

In postulating errors of various sizes we do not intend to imply that we have any basis other than pure conjecture for such errors.

REGRESSIONS OF RHYTHMIC UNIT DURATION v SIZE

For purposes of speech synthesis it is useful to provide easily used quantitative data on the regression of rhythmic unit duration against rhythmic unit size. Since there are two possible models (phoneme-class or phoneme type), two categories of rhythmic unit (marked or unmarked), two types of rhythmic unit (feet or rhythm units) and two study groups (SU30 and SU39) there are a total of 2x2x2x2 = 16 regression equations involved. These are all given below. Each equation predicts the mean discrepancy (D) between the rhythmic unit durations computed by a particular model and the corresponding measured durations for those units, given size of unit (RUS). On the next pages, some of this information is also presented graphically.

Regression equations giving the Discrepancy (D) in msec for differing Rhythmic Unit Size (RUS) in number of segments. The graphs that follow on the next pages illustrate the phoneme-type model results graphically.

Study unit, class or type-model marked/unmarked	**Equation**	**Significance of regression**
SU30, class model, feet:		
unmarked	D = 24.0 - 2.79RUS	30%
marked	D = 130.7 - 21.62RUS	**0.0%**
SU30, type model, feet:		
unmarked	D = 18.5 - 2.08RUS	39%
marked	D = 117.7 - 19.36RUS	**0.0%**
SU30, class model, rhythm units:		
unmarked	D = 38.5 - 5.05RUS	7%
marked	D = 99.8 - 21.25RUS	**0.0%**
SU30, type model, rhythm units:		
unmarked	D = 33.2 - 5.20RUS	**4.5%**
marked	D = 92.0 - 20.17RUS	**0.0%**
SU39, class model, feet:		
unmarked	D = 41.6 - 4.80RUS	19%
marked	D = 145.2 - 25.28RUS	**0.0%**
SU39, type model, feet:		
unmarked	D = 41.1 - 5.38RUS	12%
marked	D = 102.5 - 16.92RUS	**0.0%**
SU39, class model, rhythm units:		
unmarked	D = 72.6 - 11.66RUS	**0.8%**
marked	D = 128.2 - 25.49RUS	**0.0%**
SU39, type model, rhythm units:		
unmarked	D = 71.0 - 13.2RUS	**0.1%**
marked	D = 88.5 - 17.12RUS	**0.0%**

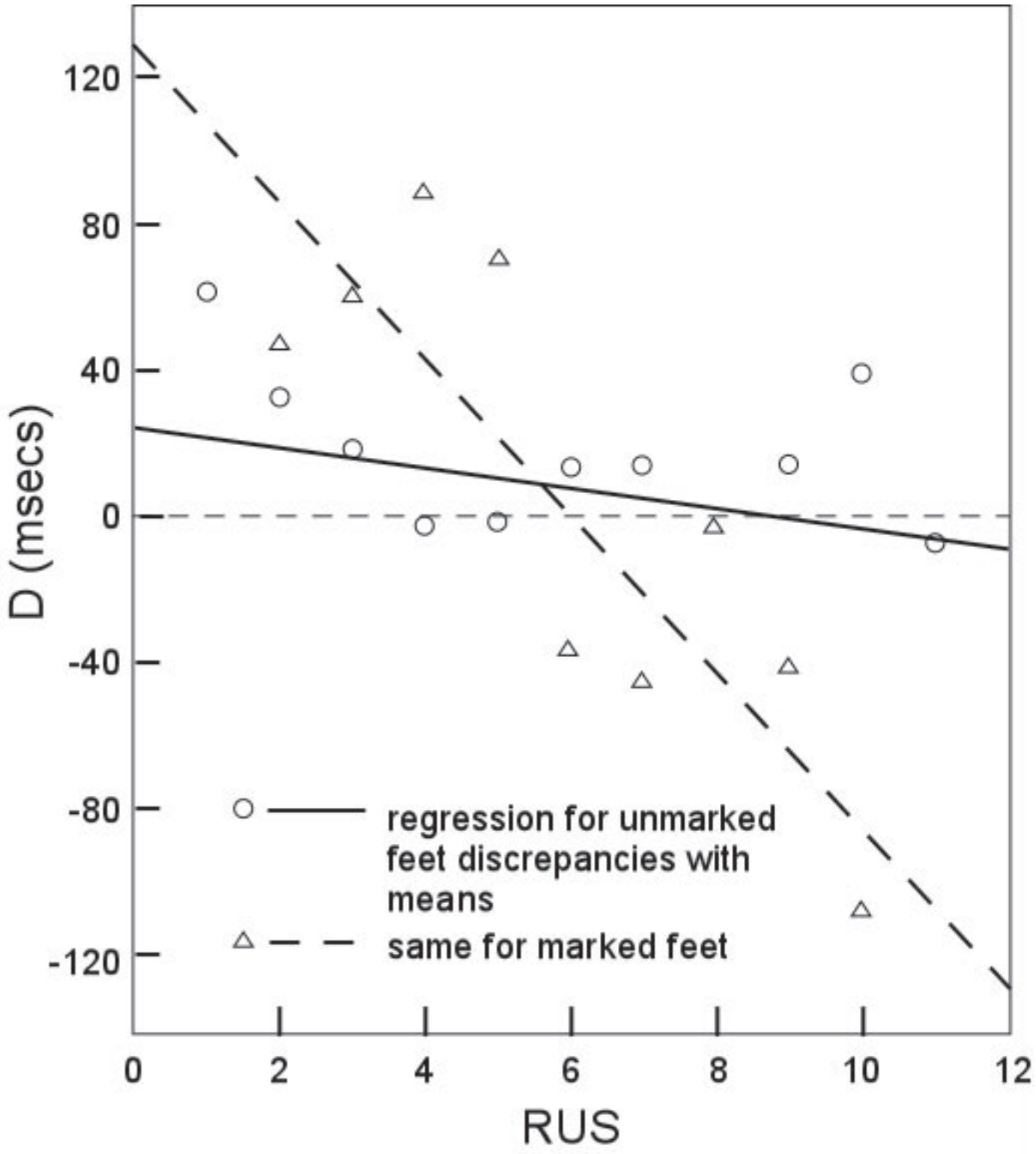

Figure A1: SU30 regression lines for correcting class model feet

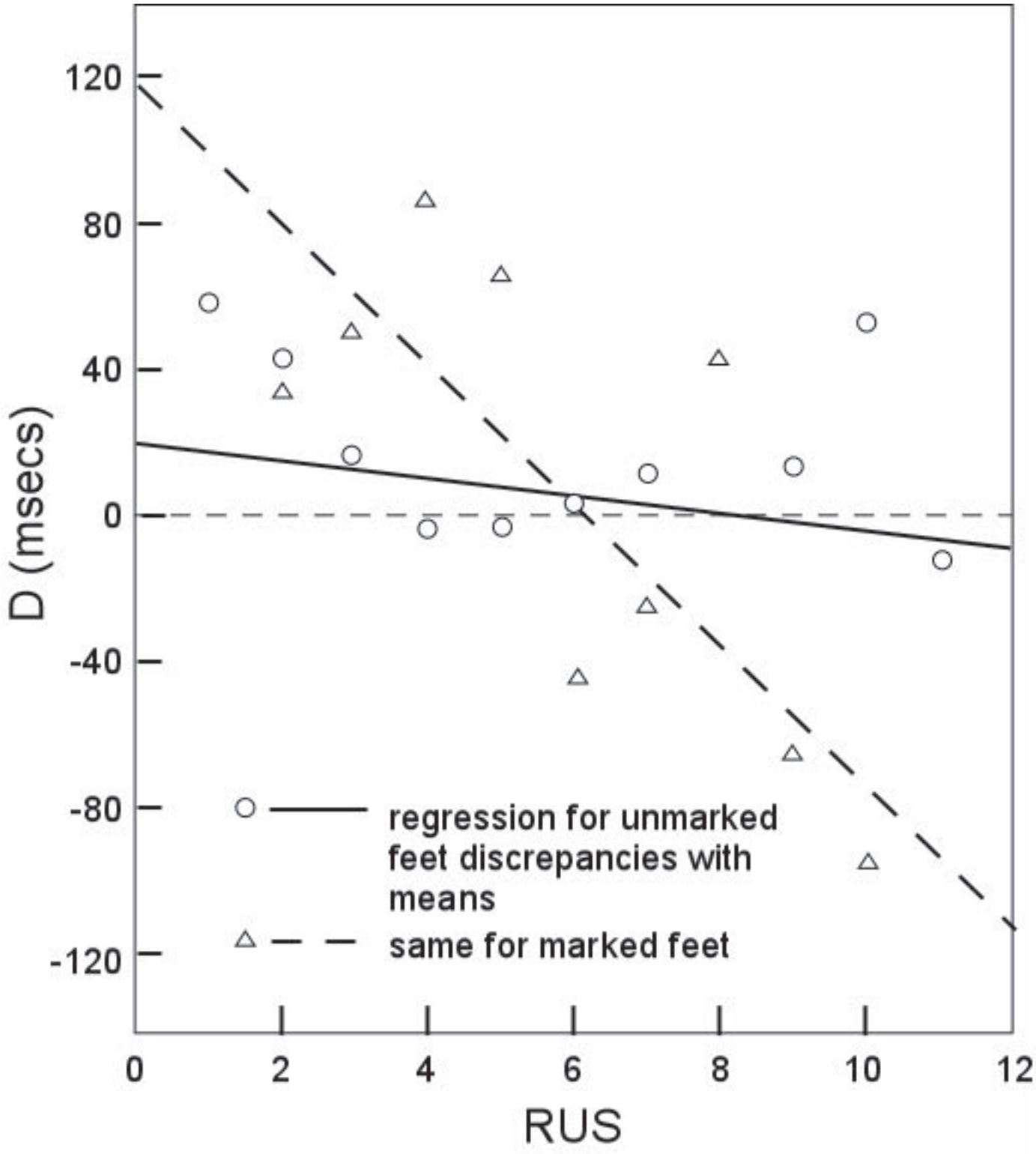

Figure A2: SU30 regression lines for correcting type model feet

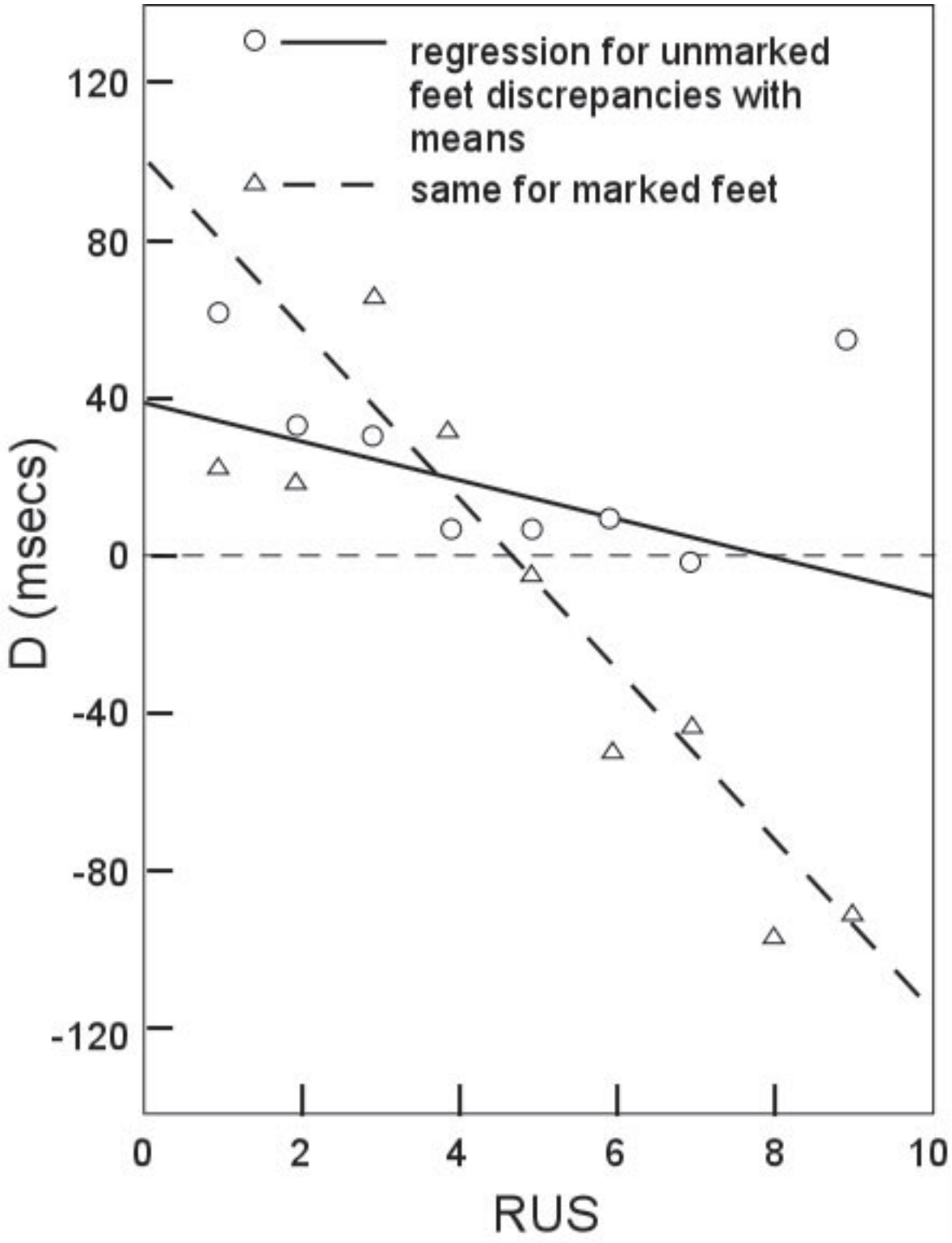

Figure A3: SU30 regression lines for correcting class model rhythm units

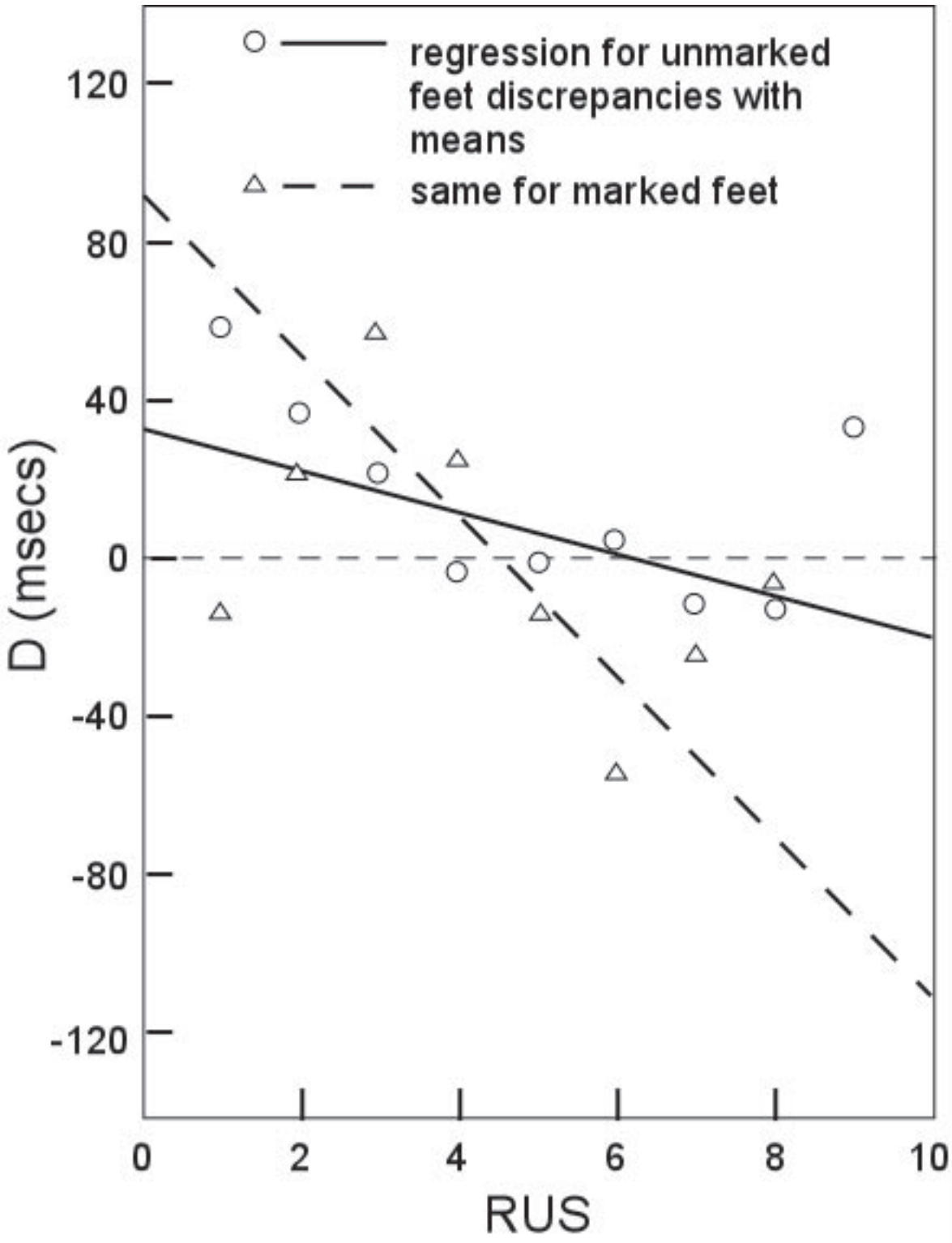

Figure A4: SU30 regression lines for correcting type model rhythm units