Speech rhythms – modelling the groove

Dafydd Gibbon

Universität Bielefeld (2012-05-19)

1 Rhythm – listening closely

Thesis: There is no speech rhythm. There are only speech rhythms.

1.1 Background

The rhythms of speech are an intensely debated topic, and have attracted increasing attention since the early studies by Pike (1945) and Jassem (1949, 1952). Detailed overviews of previous approaches have been provided at different times (Gibbon & Richter 1984; Gibbon 2006; Gibbon, Hirst & Campbell 2012). Solutions to the problem of characterising rhythms, 'modelling the groove', are many, yet none is definitive. Consequently, more intensive 'listening' is called for. The objective of this study is not to apply *a priori* models or simple 'rhythm metrics', but to take a fresh view of the nature of speech rhythms from intuitive, phonetic and formal points of view. On this basis, approaches to capturing the rhythms themselves will be outlined, rather than particular properties such as isochrony, regularity, 'smoothness', with the aim of developing a more comprehensive and integrative view of rhythms than has been available so far.

The present study has a methodological focus, rather than reporting on specific descriptive issues. Binary, ternary and other rhythm models are considered, first from a pre-theoretical point of view, and in some detail, with reference to the ternary rhythm model of Jassem and the binary rhythm model of Abercrombie, and a generic *pre-peak peak post-peak* basic alternation template for rhythm patterns is proposed. On the basis of the discussion of rhythm models, the physical properties of speech rhythms are investigated, and a generic model of syllables and feet is proposed, permitting the integration of notions such as *syllable-timed* and *foot-timed* rhythms into a single framework: the *Alternating Syllable Model* (*ASM*) and the *Syllable Extension Model* (*SEM*). Some popular quantitative models are queried in respect of their validity as rhythm metrics, and the outloook for the future development of new 'genuine rhythm models' based on recently proposed oscillator systems is sketched.

1.2 Rhythm schema, rhythm interpretation and rhythm performance

The problem is, though, that rhythms are elusive, in speech as in music. In music, pedantic iterations are not rhythm: whatever the style, there is a 'feeling', 'swing' or 'groove', with subtle *accelerandi* and *rallentandi*, with *anacruses*, 'grace notes' and syncopations which conspire against an exact definition. It is useful to make a three-way distinction which is important for disentangling different perspectives on the rhythms of speech and music.

First, there are the conventional types of *metre* in rhetoric and in traditional poetry, such as the iambic pentameter, and the times and metric patterns in music and dance, such as the waltz and the samba. Second, there are specific *interpretations* of these metres and structures in context, such as syncopations in music which modify the underlying beat, or like adaptations to grammatical patterns in speech (the 'THIRteen MEN rule', as opposed to 'there were thirTEEN'). Third, there are the perceivable, and measureable, spontaneously varied performed rhythms of the performance of speech and music, which I will refer to as the *groove*.

Figure 1 shows the three distinctions in music. The first and second lines in each example correspond. The basic rhythmic and melodic template of Gershwin's *I got Rhythm* (Figure 1 top) does not reflect the syncopations noted in the lead sheet (Figure 1 centre) of a jazz musician's interpretation and even less the melodic and rhythmical swing of Ella Fitzgerald's performance (Figure 1 bottom). The categorial properties of standard musical notation mirror the categorial properties of more abstract linguistic notations, particularly phonological and prosodic notations, while the pitch trace is a more appropriate expression of individual melismata and glissandi. The following discussion will deal with similar issues in speech.

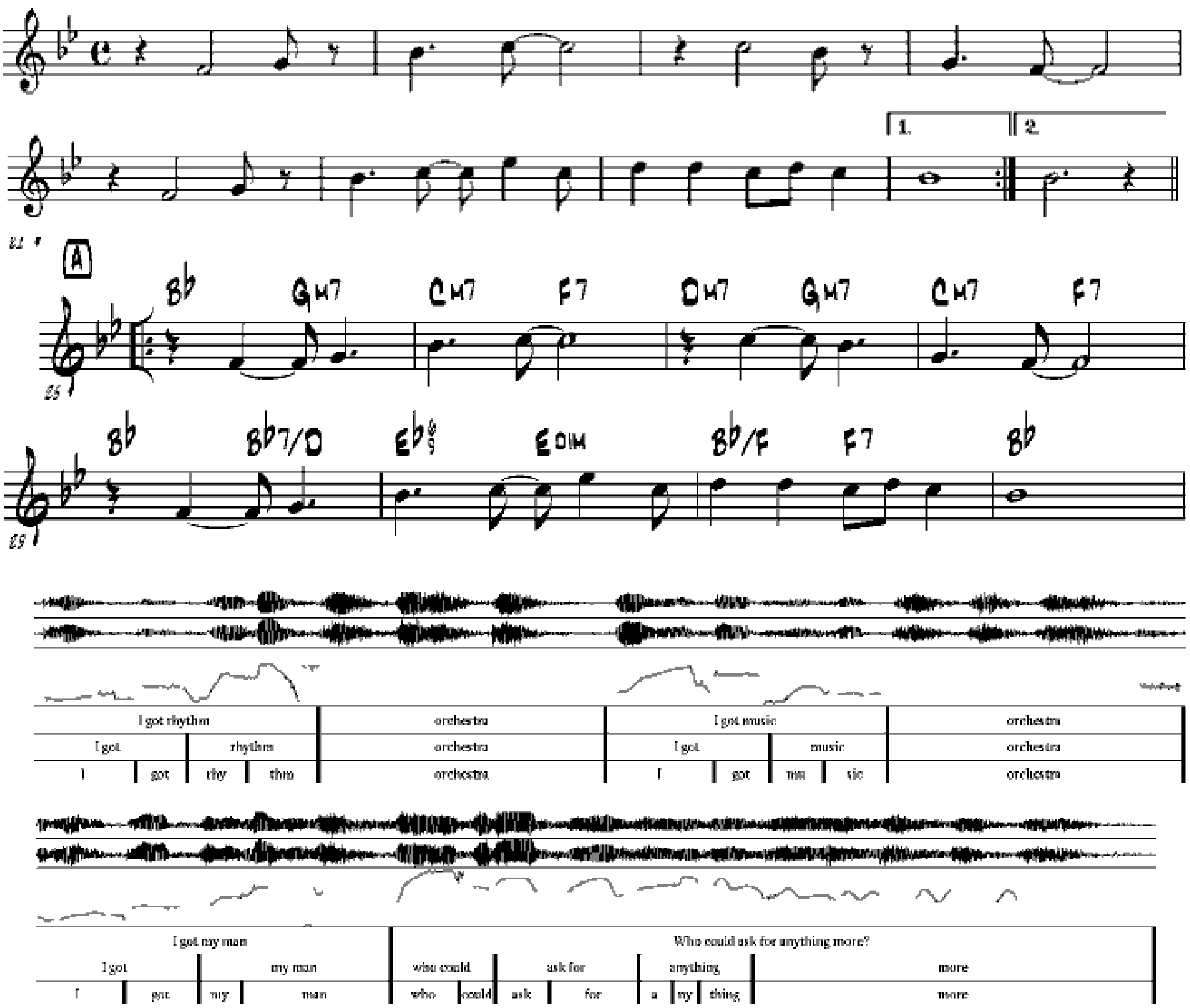

Figure 1: "I got rhythm", Gershwin (1930): schema (top), syncopation (middle, Guy Bergeron, 2010), performance (bottom, Ella Fitzgerald, 1996).

Speech rhythms, like rhythms in music, are hard to capture. But rhythms evidently have a physical and physiological reality, and not illusions or cognitive constructs alone. If this were so, we would not be able to identify 'wrong rhythms' - but we can do this. Rhythms have a physical component, even though this component is notoriously hard to pin down, in distinction to general principles of structured timing which may or may not be rhythmic.

We have to start somewhere, though. And starting points have been very different, so that rhythm studies have varied greatly over the years in the methods used, and consqeuently phoneticians have come to very different conclusions. Earlier studies were based on perceptual impressions of rhythmically relevant prominence relations, such as 'primary stress' and 'secondary stress', by pedagogical phoneticians, but also by academic phoneticians (Pike 1945; Jassem 1949, 1952). Later phonological 'rhythm' studies combined these with data structures such as trees and histograms ('grids', i.e. visualisations of numerical vectors) to express relative prominence relations (Chomsky & Halle 1968; Liberman & Prince 1977) and to derive the relative prominence relations systematically from word and sentence structure. Later still, phoneticians introduced quantitative 'rhythm metrics' in order to quantify the physical basis of perceived rhythms by means of measurements of the acoustic signal. A classification of approaches to such rhythm models and metrics was proposed by Gibbon (2006), and some of these will be referred to in the present study.

The central insight in the present context is, first, that by concentrating on isochrony, the evenness of event durations, the essential character of rhythms, that 'dum-de-de-dum-de' factor which distinguishes rhythms from other sequences, has been lost, and second, that models must be developed which take 'genuine rhythms' into account.

The structure of the study is as follows. In Section 2, 'Grooves and models', the intuitive empirical starting points are outlined in terms of the *Jassem Rhythm Model* (*JRM*) and the *Abercrombie Rhythm Model* (*ARM*). In Section 3, 'The syllable is the mother of the groove', an integrative model for relating *syllable-timed* rhythms and *foot-timed* rhythms is developed, with a model of syllables as rhythmically alternating units, the *Alternating Syllable Model* (*ASM*), and a model of feet as extensions of syllable *Onsets* and *Codas*, the *Syllable Extension Model* (*ASM*). In Section 4, 'Groovy phonetics', the phonetic properties of rhythms are discussed using a specific example as an illustration and source of experimental hypotheses, relating the timing patterns of phone sequences, syllable sequences, and sequences of ternary (Jassem type) and binary (Abercrombie type) foot sequences. In Section 5, 'Measuring the groove', the validity of a number of popular 'rhythm metrics' as rhythm models is questioned, and in Section 6, 'Rocking the groove', the outlook for developing 'genuine rhythm models' on the basis of recent oscillator rhythm systems is briefly discussed, before a summary and outlook is presented in Section 7, 'The future of the groove'.

2 Grooves and models

2.1 Definition and *explicandum*

This intuitive definition is taken to be consensual:

> Rhythms are temporally regular iterations of events which embody alternating strong and weak values of an observable parameter.

The concept of 'observable parameter' is neutral between rhythm as an epiphenomenon in human perception and cognition on the one hand, and physical measurements on the other. The alternating values of the observable parameter are commonly referred to as *strong-weak*, *light-dark*, *loud-soft*, *stressed-unstressed*, *conspicuous-inconspicuous*, *prominent-nonprominent, consonant-vowel*, *hand raised vs. hand lowered*, and in many other ways. The parameter may be a *single* feature type or a *complex* combination of many, it may be *hierarchical* in structure, it may appear in any *modality*, perhaps even the gustatory and olfactory modalities. An essential distinction must be made between *physical rhythms* and *semiotic rhythms*.

Physical rhythms may be *natural rhythms* (waves, regular limb movements during locomotion) or *artefactual rhythms* such as the rhythms of motors, the ticking of clocks, or the visual 'rhythms' (of fences or moiré silk.

The basic function of semiotic rhythms is to mark cohesion in event sequences, whether in speech or in other modes of behaviour. Semiotic rhythms may be *aesthetic rhythms* (as in metre, the rhythms of music and dancing, the patterns of abstract art) or *communicative rhythms* (the patterns of speech and gesture).

Physical rhythms (like other physical events) may be interpreted, for example in religioius or poetic contexts, as semiotic, and any of these rhythm types may be 'reconstructed' on the basis of cognitive expectations and superimposed on physiological sensations and percepts.

Starting with the intuitive definition of rhythms as temporally regular alternations, the form of a rhythm (whether physical or semiotic) may be analysed in terms of several key properties, being (a) a *time series* of (b) *rhythm events*, with (c) each event containing (at least) a pair of *different* *observable values of a parameter* over (d) *intervals of time of relatively fixed perceived duration*. Finally, (e) '*it takes (at least) two to make a rhythm*': one alternation of parameter values is not yet a rhythm.

The key properties of rhythm form are visualised as a basic *Generic Binary Rhythm Model* (*GBRM*) in Figure 2, which shows rhythms as a prominent *peak* events followed by less prominent *post-peak* events. The *GBRM* visualises an intuitive understanding of rhythms according to the initial definition, as a regular iteration of alternating strong and weak values of a parameter. The definition of *event* as a pair of an *interval* and a *property* is taken from event logic.

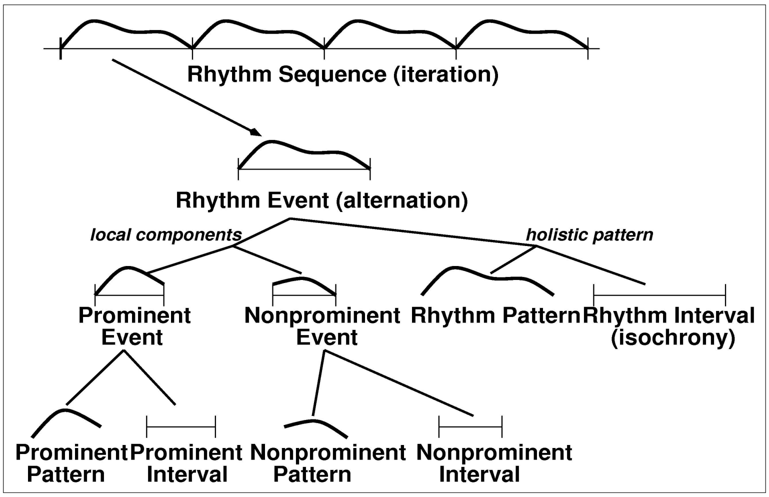

Figure 2: Visualisation of a Generic Binary Rhythm Model (GBRM).

However, speech rhythms are much more complex than the intuition-based *Generic Binary Rhythm Model* (*GBRM*) permits, as plausible pronunciations of examples (1) – (4) demonstrate (the comments are in the terminology of poetic metre, with no implication that the lines are poetic).

(1) This | fine | bear | swam | fast | near | Jane's | boat. (Singlets, syllable-timed, *dum dum* ...)

(2) And then | a car | arrived. (Iambs, *de-dum* ...)

(3) This is | Johnny's | sofa. (Trochees, *dum-de* ...)

(4) Jonathan | Appleby | carried it | awkwardly. (Dactyls, *dum-de-de* ...)

(5) It's a shame that he fell | in the pond. (Anapaests, *de-de-dum* ...)

(6) A lady | has found it | and Tony | has claimed it. (Amphibrachs, *de-dum-de* ...)

In fast speech, the numbers of unstressed syllables surrounding a stressed syllable may be greater than the one or two illustrated here.

These speech rhythm variants suggest that a model of basic speech rhythm units should not have a purely binary structure, i.e. *peak post-peak*, but a ternary structure, i.e. *pre-peak*, *peak*, *post-peak*, as a more comprehensive *Generic Rhythm Model*. A descriptively adequate model of speech rhythms in general needs to go beyond simple binary foot structures, and to be sufficiently flexible to capture not only trochaic rhythms but the other types, too.

2.2 The *Jassem Rhythm Model* and the *Abercrombie Rhythm Model*

Two classic speech rhythm types have been proposed: *syllable-timed* rhythms and *foot-timed* (or *stress-timed*) rhythms. The basic units of each type, syllable and foot, are claimed to be 'isochronous': the syllable and the foot, respectively, are said to be spoken in relatively constant temporal intervals. There are two prominent phonetic models for *foot-timing* : the *Jassem Rhythm*

Model and the *Abercrombie Rhythm Model* (Jassem 1949, 1952; Abercrombie, 1967). Phonetic properties of the models have been discussed in detail by Hirst & Bouzon (e.g. 2005).

The *ARM* is binary, and its *Rhythm Unit* (*RU*) contains two constituents, the *Ictus* (stressed syllable) followed by the optional *Remiss* (sequence of unstressed syllables up to but not including the next *Ictus*. The patterns captured are just the trochee and dactyl types. Where a sequence starts with unstressed syllables, a 'silent *Ictus*' is postulated.

The *JRM* is hierarchical with two hierarchical divisions: the *Total Rhythm Unit* (*TRU*) is a ternary sequence of *Anacrusis* (*ANA*), *Narrow Rhythm Unit* (*NRU*) and *Rhythmical Juncture* (*RJ*). The *Anacrusis* is a rapidly pronounced sequence of unstressed syllables between the last *RJ* and the next *NRU*, does not have the property of isochrony, and is largely determined by the coincidence of a grammatical break with the preceding *RJ*. The definition of Jassem's *NRU* is the same as the definition of Abercrombie's *RU*: a stressed syllable followed by a sequence of unstressed syllables. Jassem does not use the terms *Ictus* and *Remiss* for the constituents of the *NRU*, but they will be used here for convenience of reference. The difference between the *JRM* and the *ARM* is that the syllables in the *JRM* sequence of *ANA* and *NRU* constitute a ternary structure of *ANA*, *Ictus* and *Remiss*, while in the *ARM* the syllables constitute a simpler binary structure of *Ictus* and *Remiss*. In each case, only the *Ictus* is obligatory.

Given the rhythm patterns documented in (1) – (6), it appears *prima facie* that the *ARM* is inadequate to describe most of the patterns, while the *JRM* captures all of them, and also the longer stretches of unstressed syllables found in fast speech. The properties of the two models are summarised and discussed in detail by elsewhere (Gibbon, Hirst & Campbell 2012).

Without prejudicing the issue of whether these two types are systemically and phonetically valid or not, it is straightforward to interpret the *Generic Rhythm Model* in terms of the *JRM* and *ARM* rhythm models (Table 1), with a syllabification of the sequence *it was terrifying to see*.

Table 1: Comparison of Generic Rhythm Model (GRM), Jassem Rhythm Model (JRM) and Abercrombie Rhythm Model (ARM).

	–	**it**	**was**	**TER**	**ri**	**fy**	**ing**	**to**	**see**
GRM:	–	*pre-peak*		*peak*	*post-peak*			*pre-peak*	*peak*
Jassem (JRM):	–	ANA		*Ictus*	*Remiss* + RJ			ANA	*Ictus*
Abercrombie (ARM):	*Ictus*	*Remiss*		*Ictus*	*Remiss*				*Ictus*
Grammar:	–	*it was terrifying*						*to see*	

The direct comparison in Table 1 shows clear structural and functional differences between the ternary Jassem model and the binary Abercrombie model. First, tbe binary *Ictus-Remiss* model forces Abercrombie to postulate a silent *Ictus* in cases where a sequence begins with unstressed syllables. Jassem's model obviates such multiplication of entities *praeter necessitate* by introducing the *ANA* category, which is independently empirically motivated by having its own timing pattern which differs from timing in the *NRU* (Jassem, Hill & Witten 1984; Hill 2012), and thus is not introduced *praeter necessitate*. With the ternary *ANA Ictus Remiss* pattern, the Jassem model fulfils the requirements for a *GRM pre-peak peak post-peak* pattern.

Second, the Jassem model differs from the Abercrombie model by introducing an explicit *RJ* boundary category, which is motivated phonetically as (a) ending an *NRU* (for example by 'final lengthening' at the end of a rhythm sequence, and (b) separating an *NRU* from a following *ANA*, and thus is also not introduced *praeter necessitate*.

Third, Jassem's examples show that the *JRM* differs in systemic motivation from the *ARM* by implicitly recognising grammatical properties of rhythm patterns (cf. also Pike 1945:33ff.), in two ways: (a) in the tendency for *ANA* to mark proclitic or premodifying grammatical items in English right-headed constructions, and (b) in the tendency for the *RJ* to correspond with grammatical boundaries.

Correspondence of grammatical and prosodic boundaries is neither a necessary nor a sufficient condition for a rhythm model, since grammatical boundaries may occur within rhythm units in fast speech, and rhythm units may occur within larger grammatical units in deliberative discussion styles (cf. contributions to Dechert & Raupach 1980). A priori there is no necessity for grammatical and prosodic units to coincide at the grammatical and rhythmical junctures. Nevertheless, the relation between grammatical and prosodic boundaries is a frequently noticed (and actually rather obvious) tendency, and inclusion of this well-motivated systemic correspondence property, together with phonetic motivation for the ternary pattern, and without introducing silent entities, weighs in favour of the Jassem model.

3 The syllable is the mother of the groove

3.1 The Case of the Missing *Ictus* and the Missing *Remiss*

There is a puzzling feature of *foot-timed* rhythms in English and prosodically similar languages, which has not been adequately discussed in the extensive literature so far. If rhythm is *alternation* between stressed and unstressed syllables, then how are cases such as (1) accounted for, in which only stressed syllables occur? The Abercrombian model postulates that in sequences which begin with unstressed syllables, there is a 'silent *Ictus*', and sequences which end with an *Ictus* have a 'silent *Remiss*', yielding yet another entity *praeter necessitate*. But in the Jassem model, too, there is no account of what is rhythmical in the pattern of the missing *Remiss*: to state that *ANA* and *Remiss* are optional is a descriptive statement with no systemic or phonetic explanatory value. The 'puzzle of the absent *Remiss*' which is shared by both the *JRM* and the *ARM* models is representative of the entire literature.

3.2 The Alternating Syllable Model: syllables as feet

More generally, what are the units which alternate in rhythms of languages (and corpus events in general) which are *syllable-timed*? The answer to this question is the key to resolving the puzzle of the missing *Remiss* in both the Abercrombie and the Jassem models, with its apparent lack of alternation and therefore lack of rhythm.

The solution to this problem is surprisingly simple and, once stated, surprisingly obvious, but leads through a detailed discussion of syllable phonotactics and morphonotactics.

First, in cases such as (1), and in syllable timing in general, the *pre-peak*, *peak post-peak* alternation is between *vowels* and *consonants*. The CVC alternation is evidenced most clearly in languages like English which have complex syllables, such as *splints* /splɪnts/, with complex consonant clusters in both *Onset* (*O*) and *Coda* (*C*) of the syllable, which alternate with the sonorant *Nucleus* (*N*) in the pattern $(O\ N\ C)^+$ (the superscript '+' indicates a sequence consisting of at least one occurrence of the pattern in parentheses). Interactions between the *Nucleus* and the *Coda* within the *Kernel* (the term 'rhyme' is a misleading metaphor in view of its specific and different meaning in poetry) are discussed below.

English has many closely interrelated syllable templates; the *splints* /splɪnts/ type is just one. The sequence /s/+/p/+/l/ɪ/+/n/+/t/+/s/ can be categorised as *s+O+L+V+S+O+s*, i.e. /s/ followed by obstruent, liquid, vowel, sonorant, obstruent, and terminated by /s/, with each subsequence subject to complex co-occurrence constraints (Gibbon 2001).

The category sequence *s O L V S O s* is parsed on distributional grounds as *s O L* (*Onset*), *V S* (*Nucleus*), *O s* (*Coda*). Distributionally, the *Nucleus* is /ɪn/ with the pattern *V S*, a short vowel followed by a sonorant consonant. Likewise, the *Coda* is distributionally relatively simple, with the template *Os* for obstruent followed by the morphologically determined /s/. The parse /spl/ + /ɪn/ + /ts/ is done differently in traditional syllable analyses, but is justified at the lexical level by distributional constraints: the *V S* pattern of /ɪn/ is substitutable by a long vowel pattern *V V*, a vowel plus approximant *V A*, or a short vowel *V*, retaining the same *Onset* and *Coda* structures.

Syllable models which use an *a priori* phonetic categorisation of sonorants as consonants rather than a systemic distributional motivation assign the *S* component to the *Coda*, which conflicts with the distributional facts about English syllable structure. However, using a criterion of perceptual prominence, which is relevant for rhythm, the distributionally motivated parse /spl/ + /ɪn/ + /ts/ has to give way to /spl/ + /ɪ/ + /nts/: /ɪ/ is a prominence peak and /n/ has intermediate prominence between /ɪ/ and /t/ on sonority scales.

There is no 'right' and 'wrong' between distribution based analysis and sonority based analysis. There is no single syllable structure: in systemic terms, there are simply two perspectives on syllable structure, justified by independent empirical criteria of distribution and sonority, respectively. The lexical distributional pattern /spl/ + /ɪn/ + /nts/ is thus reanalysed as /spl/ + /ɪ/ + /nts/ in terms of sonority.

The sonority alternation criterion provides valuable further information: the sonority criterion which explains – rather than just describing – the preferences of languages of the world for CV patterns over V syllable patterns (Jakobson 1941 *et multi*): the CV pattern guarantees a rhythmic alternation of vowels and consonants.

In cases where both *Onset* and *Coda* are missing, as in sequences of V-only syllables, this is not the end of all alternation: other low sonority phonetic means intervene to preserve the alternation, for example glottal stops, approximants, or amplitude and pitch modulations. In English, the second mora of long vowels, or the glide of diphthongs serves as the weak element. Where vowels are adjoined with *liaison*, a transition between sonority levels occurs. In interjections such as *A-a-a-a!* [aʔaʔaʔa], a glottal stop intervenes and preserves alternation.

3.3 The *Syllable Extension Model*

The basis for distinguishing *foot-timed* rhythms and *syllable-timed* rhythms has now been explained. But how is the co-occurrence of *syllable-timed* cases such as example (1) in English with various different kinds of *foot-timed* rhythms accounted for?

The solution to relating *syllable-timed* and *foot-timed* events is, as with the justification of the *JRM*, partly systemic: the *Onset* and the *Coda* of the basic syllable event is extended by extrametrical 'weak syllables' with very different phonotactic and often morphotactic status from the 'strong' stressed syllable. Some weak syllables are morphologically determined (affixes), some are historically opaque and purely phonological (e.g. in latinate words such as *complaint* or *solid*), others are short grammatical words such as determiners, pronouns, auxiliary verbs, prepositions, conjunctions, which have clitic-like behaviour in informal and fast speech styles.

The word *splints* is already an example of morphological extension: the sequence /ts/ has a morphonotactic juncture {t+s}. Words like *solidly* /ˈsɔlɪdlɪ/ have post-*Coda* syllables (the first /l/ is ambisyllabic, i.e. both *Coda* and *Onset*), which have distributionally highly constrained 'weak syllables', in one case morphologically opaque, /lɪd/, in the other morphologically transparent, /lɪ/, a derivational morpheme. The morphophonotactic structure is {ˈsɔlɪd+lɪ}. Other words like *complaint* /kəmˈpleint/ have a pre-*Onset* syllable such as /kəm/ whose original Latin morphological status is opaque in English. In cases such as *unlikely* /ənˈlaɪklɪ/ {ən+ˈlaɪk+lɪ} the pre-*Onset* extension is morphologically transparent as a derivational morpheme. Typically, the weak *Syllable Extensions* are either unstressed or acquire alternating secondary stresses.

3.4 Syllables as feet, feet as *Syllable Extensions*

Having characterised syllables as alternating units with the *ASM* analysis, and as *Onset extensions* and *Coda extensions* in the *SEM* analysis, the remaining step to linking *syllable-timed* and *foot-timed* rhythms is a very small one: the basic rhythm event type is the *ASM*. In the *SEM*, the *Onset* and the pre-*Onset* extension join as the *Anacrusis*, the pre-peak component of the *GSM* pattern, and the *Coda* and the post-*Coda* extension join as the *Remiss*. Consequently, where there is no *Remiss* in the traditional sense as a sequence of unstressed syllables, the *Coda* steps in as the minimal *Remiss*.

The principles of *syllable-timed* and *foot-timed* therefore no longer need to be seen as mutually exclusive categories: they are the ends of a continuum which, at the 'foot' end may be extended arbitrarily by segment and syllable reduction in informal and fast speech. The problem of showing the relationship between *syllable-timing* and *foot-timing*, both when they co-occur in a given language and typologically when they occur in corpora of different languages, is therefore solved in a coherent and explanatory way by combining the *Alternating Syllable Model* (*ASM*) with the *Syllable extension Model* (*SEM*): differences arise from differing phonotactic and timing constraints on *Onset Extension* and *Coda Extension* sequences.

It is tempting to search for further similarities between syllables and their components and feet and their components: syllables and *TRU*s share a similar hierarchical structure; the *Onset* has properties of distributional independence from the *Kernel* and the *Anacrusis* has properties of temporal independence from the *NRU*; the *peak* property is shared by *Nucleus* and *Ictus* (which for stressed vowels are in any case identical by definition); the *Nucleus* and *Coda* are distributionally interdependent and *Ictus* and *Remiss* are temporally interdependent; both the *Coda* and the *Remiss* are variable in respect of timing, lenitions, assimilations and reductions.

On the basis of the preceding argumentation, the conclusion is drawn that the syllable is at the core of all rhythmical timing patterns. But rhythms are not necessarily based only on durational relationships at one level: there are other contributory factors to *peak* prominence such as timing properties of constituents, as well as pitch patterning. An obvious relation to look for is the patterning of durations between hierarchically close constituents of a rhythmic pattern. Accordingly, discussion of phonetic timing properties of syllables and their constituents is the next step.

4 Groovy phonetics

4.1 Timing patterns: durations and duration differences

In the preceding sections, discussion has been purely on phonotactic and morphonotactic lines. A detailed quantitative empirical analysis is not feasible in the present context, but the main lines of *ASM-SEM* based research can be outlined straightforwardly, starting by pointing out structural similarities between ternary syllable structure and the ternary *JRM* as the basis for rhythmic patterning (Figure 3).

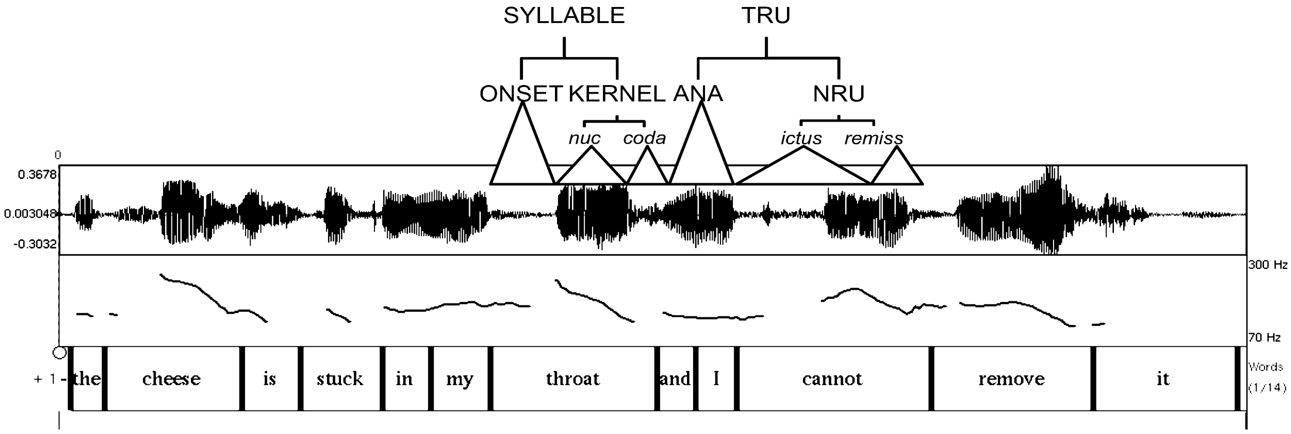

Figure 3: Structural similarity between syllable and rhythm unit patterns.

Next, the timing patterns of syllables and syllable components are discussed, followed by discussion of the phonetic consequences of the *JRM* and *ARM* timing patterns. The methods start with an annotated speech recording (Figure 4).

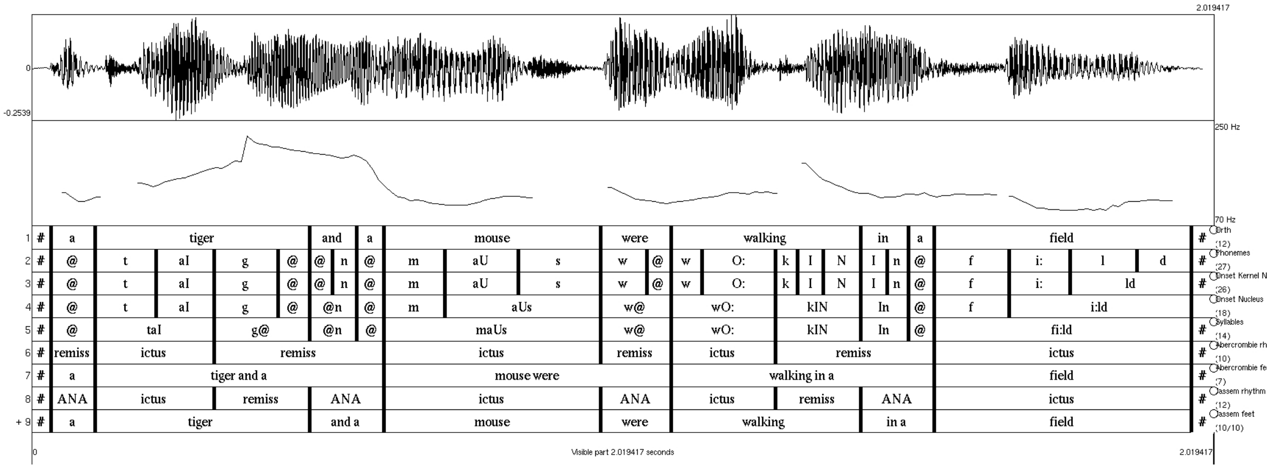
2.019417
0
-0.2539
250 Hz
70 Hz
1 # a | tiger | and | a | mouse | were | walking | in | a | field #
2 # @ | t | aI | g | @ | @ | n | @ | m | aU | s | w | @ | w | O: | k | I | N | I | n | @ | f | i: | l | d #
3 # @ | t | aI | g | @ | @ | n | @ | m | aU | s | w | @ | w | O: | k | I | N | I | n | @ | f | i: | ld #
4 # @ | t | aI | g | @ | @n | @ | m | aUs | w@ | wO: | kIN | In | @ | f | i:ld #
5 # @ | taI | g@ | @n | @ | maUs | w@ | wO: | kIN | In | @ | fi:ld #
6 # remiss | ictus | remiss | ictus | remiss | ictus | remiss | ictus #
7 # a | tiger and a | mouse were | walking in a | field #
8 # ANA | ictus | remiss | ANA | ictus | ANA | ictus | remiss | ANA | ictus #
+ 9 # a | tiger | and a | mouse | were | walking | in a | field #
Orth (12)
Phonemes (27)
Onset Kernel Nucleus (26)
Onset Nucleus (18)
Syllables (14)
Abercrombie rhythm (10)
Abercrombie feet (7)
Jassem rhythm (12)
Jassem feet (10/10)
0
Visible part 2.019417 seconds
2.019417

Figure 4: Recording of "a tiger and a mouse were walking in a field", manually annotated on 9 tiers (syllable structure, the Jassem Rhythm Model and the Abercrombie Rhythm Model).

The procedure involves manual annotation of the data, and automatic processing of the manually annotated data:

1. Manual annotation of recorded speech on separate tiers, by syllable and syllable constituent labels, and by the categories of the *JRM* and the *ARM*.
2. Extraction of the durations of intervals on each tier.
3. For the *n* intervals on each tier, calculation of the *n*-1 differences between duration of each interval *k* and the following interval *k*+1, starting with the first interval.
4. Calculation of the absolute difference between neighbouring durations (i.e. conversion of negative difference values into positive difference values).
5. Calculation of linear regression over duration patterns.
6. Display of the resulting patterns.

This procedure, with normalised durations, was designed for quantitative studies. In the present context, no normalisation of durations or duration values is performed because the displays are intended for 'eyeballing' visual interpretation rather than for quantitative analysis.

The rationale behind these visualisations of timing patterns is that popular rhythm metrics' rely on these kinds of empirical data as a starting point for measurement (cf. Section 6).

4.2 Syllable timing patterns

Using the method outlined above, the duration relations in phoneme sequences and in syllable sequences in the extract from a read-aloud spoken narrative shown in Figure 4 were calculated. The ultimate constituents of syllables are segmental phones (corresponding to phonemes); the durations and duration differences in the phone pattern are shown in Figure 5.[1]

[1] Interpretation of lines in Figure 5, Figure 6 and Figure 7: dark line: durations; dark grey line extending below zero: differences between neighbouring durations; light grey line above zero, meeting the difference line on positive peaks: absolute – positive – difference between durations; straight line: linear regression over durations.

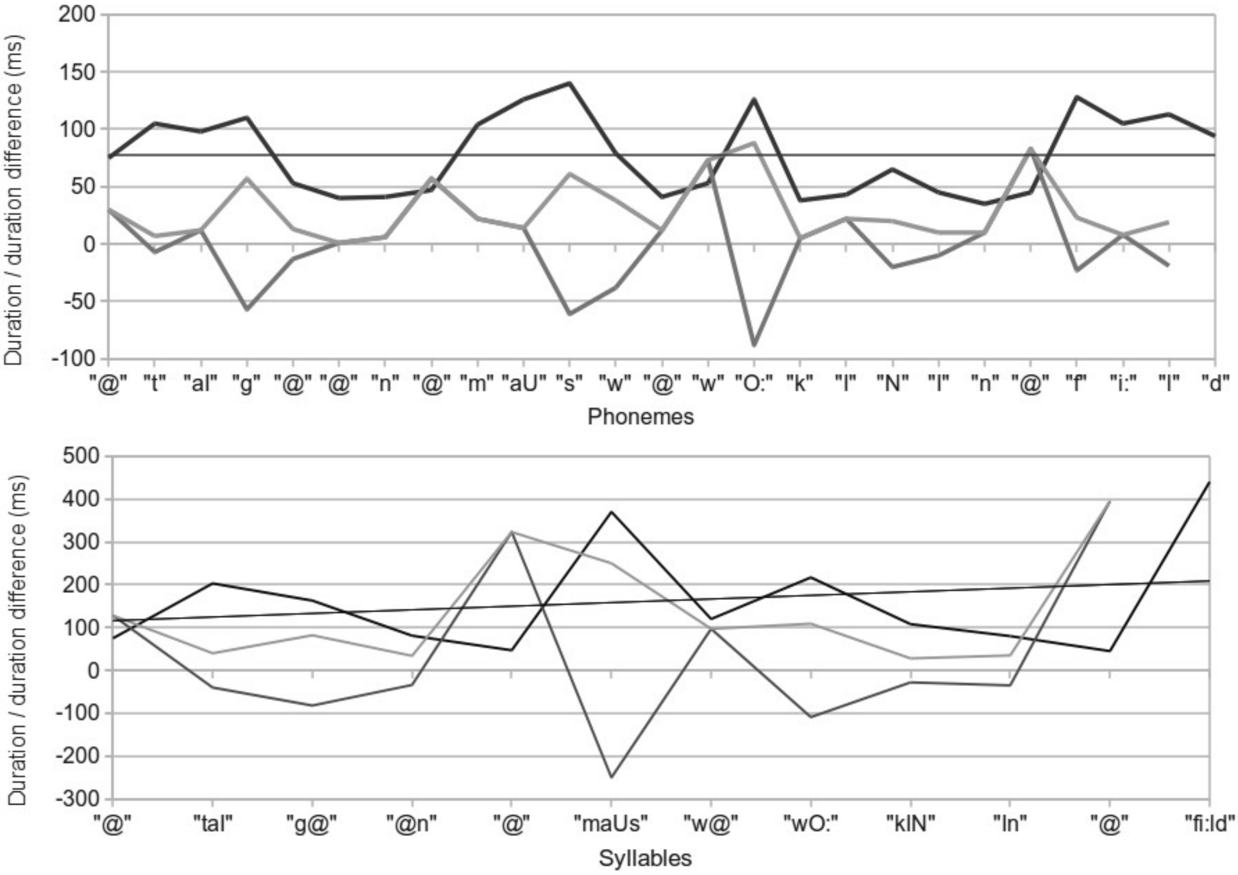

Figure 5: Phoneme and syllable sequence durations and duration differences.

Inspection of the phone duration sequence shows, as expected, that the durations peak on the stressed long vowels, i.e. the diphthongs /aɪ/ and /aʊ/, and the long vowels /ɔː/ and /iː/. The *Onset* consonants /t/, /m/ and /f/ are approximately the same length as the following vocalic segment (the /f/ is even slightly longer); the *Sonorant* /l/ in /fiːld/ is also longer than the /iː/.

The difference function is an 'edge detection' function, and effectively indicates boundaries or transitions between longer and shorter units; the plain difference function shows the *weak-strong* or *strong-weak* directionality of the transition, and the absolute difference function simply shows boundaries of whatever direction. *Prima facie*, the absolute difference function shows boundaries which occur in relatively even distances in terms of segment counts (not necessarily in terms of time; in this case the regularity may be an accident of the phonotactic structure of the utterance.

Another obvious regularity is found in the durations of the stressed vowels, which are between about 110 and 120 milliseconds. The flat regression line shows an unexpected property: that the duration distributions of the phones tend to remain constant throughout the utterance.

The syllable timing patterns, on the other hand, show a different picture: there is no obvious regularity of in the syllable boundary distribution, unlike the phone distribution. Further, Syllable lengths increase over the length of the utterance, both overall (as shown by the slightly rising regression line), but also the stressed syllables /taɪ/, /maʊs/ increase in length, a pattern which is repeated in the second half, also overall slightly higher, with the syllables /wɔːk/ and /fiːld/, marking a hierarchical timing structure (cf. Campbell 1992; Gibbon 2006).

This increase in the duration of syllables in general and of the stressed syllables in particular, in the first part 'a tiger and a mouse', in the second part 'were walking in a field' reflects a grammatical Subject-Predicate structure ('a tiger and a mouse', 'were walking in a field'), a systemic correspondence already noted in the cases of the Pike (1945) and Jassem (1949, 1952) approaches.

Before moving on to an analysis of the foot timing patterns, a higher level patterning can thus already be discerned. However, there may be a very simple explanation for these increases in length: final syllable lengthening, shorter in the initial group, longer in the final group.[2]

4.3 Jassem timing patterns

The same data collection, analysis and duration display method was applied for the *JRM*, looking first at the ternary sequences *ANA*-*Ictus*-*Remiss* (top) and then at the two types of rhythmical unit *ANA* and *NRU* (Figure 6).

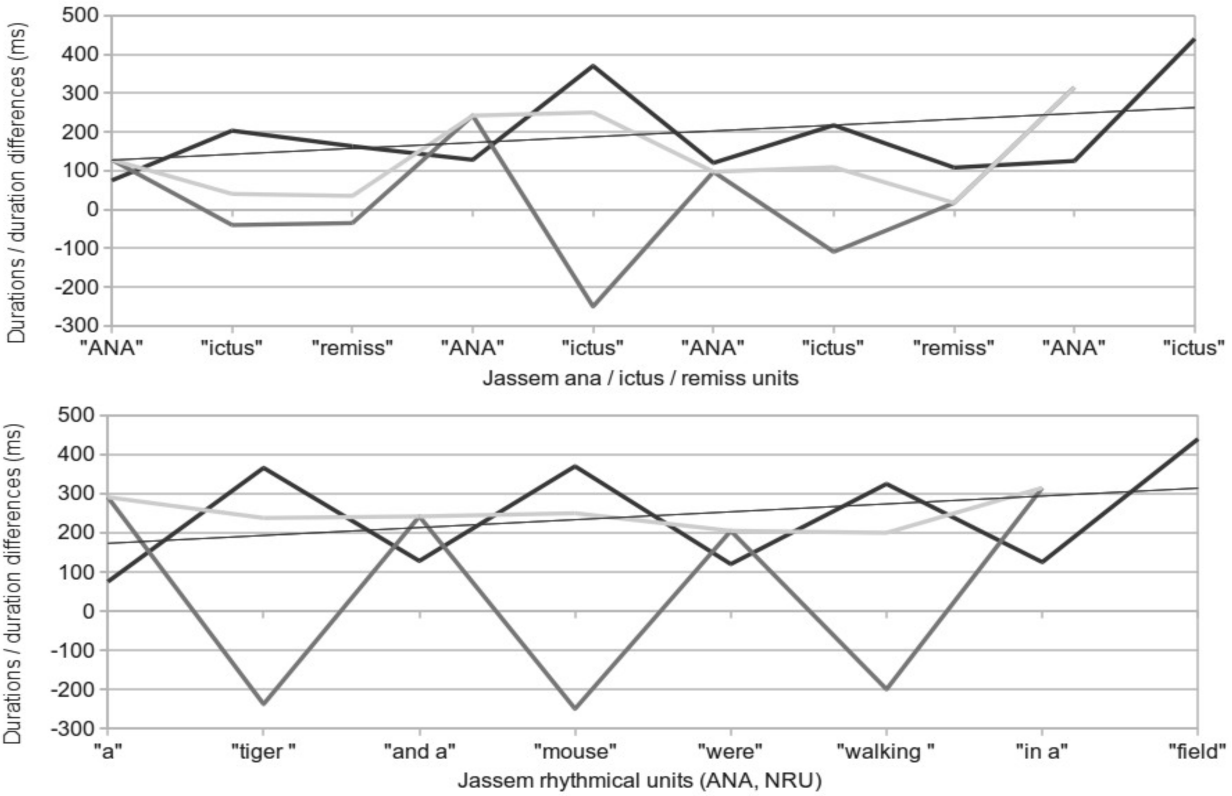

Figure 6: Foot and foot constituent durations and duration differences – Jassem Rhythm Model.

The regression lines in both the displays of the Jassem model confirm the impression given by the syllable pattern (though not the phone pattern), that the units increase in length over the course of this utterance in each case. In the top display of Figure 6, the lengths of the *Ictus* instances reflect the increase in length of the stressed syllables which has already been, unsurprisingly, since the *Ictus* is a stressed syllable by definition. There is no obvious pattern over the entire ternary sequence (top). However, the bottom display, which just relates the two rhythm unit types *ANA* and *NRU* shows a very striking distribution indeed: the durations show an unexpected conspicuous alternating ‘zig-zag’ pattern, with *NRU* lengths of about 360 milliseconds and *ANA* lengths of just over 100 milliseconds (including the initial *ANA* ‘a’), with an overall fairly even *TRU* length of about 460 milliseconds.

Obviously, if both *ANA* and *NRU* sequences separately tend to isochrony, then the *TRU* also tends to isochrony and there is at first glance no clear advantage to regarding the *ANA* as a separate category in this example: *duration*(*ANA*+*NRU*) = *duration*(*NRU*+*ANA*) = *duration*(*TRU*).

[2] A strict *caveat* is in order: the example was chosen with the aim of illustrating a method method, and therefore represents only an informal demonstration of hypotheses leading to possible pgeneralisations. Further, the read-aloud data may not be generalisable to other, more spontaneous speech registers. For the generalisations themselves, quantitative analysis is required, and this is the topic of separate studies (Gibbon 2006).

What does come out very clearly, though, is that – unlike the length of the *Anacrusis* – the length of the *NRU* is not a function of the number of syllables: the *NRU* retains its length whether there is a *Remiss* (as in 'tiger', 'walking') or not ('mouse', 'field'); in the latter cases, the *ASM-SEM* model of 'syllables as feet' comes to bear. In consequence, what also comes out is that there is a clear difference between the *Remiss* sequence of unstressed syllables, whose length is interdependent with the length of the *Ictus*, and the *ANA* sequence of unstressed syllables, whose length is independent of that of the *Ictus*. It is an interesting new possibility that Anacrusis and *NRU* may be equi-durational.

These are useful hypotheses for quantitative studies, though of course they may be refuted. Again, the read-aloud data are used with illustrative, not quantitive intention.

4.4 Abercrombie timing patterns

The *ARM* was investigated (Figure 7) using the same criteria as were used for the *JRM*.

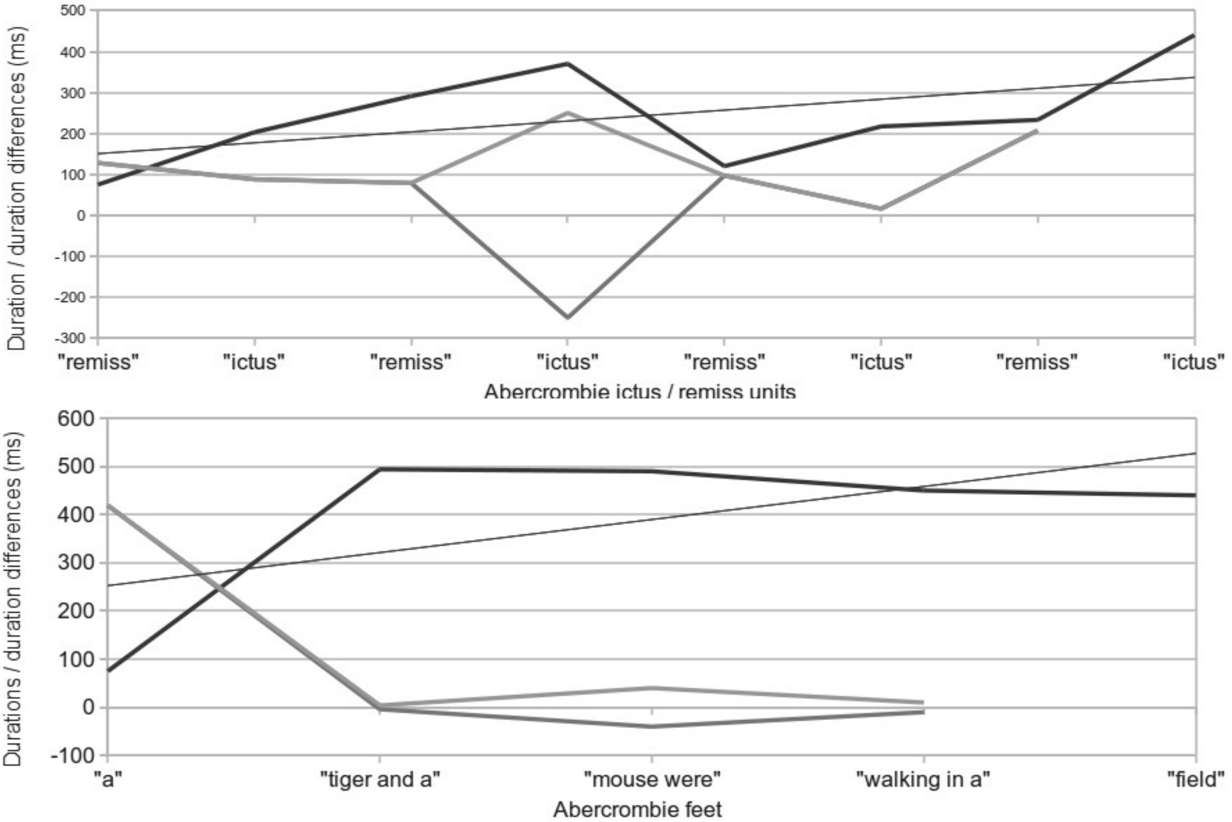

Figure 7: Foot and foot constituent durations and duration differences – Abercrombie Rhythm model.

As with the syllable and the *JRU* analyses, the regression lines show a slight increase in durations during the utterance. However, the *ARM* bundles Jassem's *ANA* together with the *Remiss* and this does not result in any obvious kind of evenness or alternation. The bottom display, on the other hand, repeats the result of the *JRM* investigation: when *ANA* is bundled together with *NRU* (= *Ictus* + *Remiss*), as expected, very even patterns result. The 'odd-man-out' is the initial unstressed syllable 'a', which does not fit comfortably into the *ARM*, while it is easily handled by the *JRM*.

4.5 Conspectus of syllable and foot models

A number of interesting observations have emerged as a source of possible generalisations and hypotheses for quantitative modelling:

1. The distribution of phone segment durations tends to remain constant over an utterance.

2. The distribution of durations of syllables and foot based units tends to increase over an utterance (which may be accidental and due to the chance distribution of phone segments within syllables).
3. *Ictus* durations tend to increase hierarchically over grammatical units such as Subject sequences, Predicate sequences and Subject-Predicate sequences.
4. *Anacrusis* sequences tend to be equal in length, i.e. isochronous (a hypothesis which is contrary to Jassem's idea that they are not isochronous.
5. *Narrow Rhythm Units* tend to be equal in length, i.e. isochronous (which is Jassem's hypothesis).
6. *Anacrusis* unstressed syllable sequences tend to be as long as the entire *Narrow Rhythm Unit* (and *a fortiori* longer than *Remiss* sequences of unstressed syllables).
7. The *JRM* produces a range of very clear hypotheses, while the *ARM* only produces one hypothesis, which is a subset of the *Jassem Rhythm Model* hypotheses, namely that Abercrombie's *RU* and Jassem's *TRU* both tend to be isochronous in this particular example, but that Abercrombie's model fails to integrate the initial unstressed syllable into the pattern.

5 Measuring the groove

Starting with a clear understanding of the nature of rhythm, and in particular of the relationship between the *syllable-timed* and *foot-timed* rhythm styles, a number of popular 'rhythm metrics' of recent years can be briefly investigated. These (and many other) metrics and their empirical significance have been discussed in detail elsewhere (Gibbon 2006; Gut 2012), so discussion in the present context can be restricted to just a few representative metrics and their formal validity for 'modelling the groove'.

5.1 Mean Foot Length (MFL) and Percentage Foot Deviation (PFD) metric

The metric proposed by Roach (1982) is a little difficult to extract from the textual description. A plausible interpretation is that the *MFL-PFD* metric defines the *Mean Foot Length* in the obvious way by averaging foot lengths, and then derives the *Percentage Foot Deviation* as a simplified analogy for standard deviation. While for standard deviation the square root of the squared difference between each value and the mean is calculated, in the Roach formula the sum of the absolute differences between each length and the mean length is divided by the overal length of the sequence and converted to a percentage:

$$MFL = \frac{\sum_{i=1}^{n} |foot_i|}{n} \qquad PFD = 100 \times \frac{\sum_{i=1}^{n} |MFL - len(foot_i)|}{n \times MFL}$$

The expression $|foot_i|$ here means the length of $foot_i$, not its absolute value. The total differences from the mean length are expressed as a fraction of the overall length. Evidently, if all feet are equal in length, the differences are zero, the sum of differences is zero, and the fraction of the overall length is zero. Therefore if *PFD* approaches *zero*, the feet are more perfectly isochronous, and the more the *PFD* differs from zero, more irregular the timing is. That is the theory. But it does not work as advertised, for several reasons:

1. The measure is not normalised for speech rate: if the speech rate varies over the utterance, this can lead to an artificially high *PFD*, even though the local duration differences are rather small.
2. More seriously: the more the *PFD* differs from zero, the less we know about what it is actually measuring, because it is a global measure, averaging over the lengths of all feet (or whatever unit is used) within the entire unit. Different – even random – orderings of the

same duration values, can be scattered over the utterance in arbitrary orders, still yielding the same *PFD*.

Evidently, the *PFD* is measuring something like ‘smoothness’ of durations averaged over a whole utterance. It says nothing about the actual alternating rhythmic structure. Consequently, for very low values, while the *PFD* can be a useful indicator of ‘smoothness’ of sequences of a particular type of unit such as the foot, but higher values are uninterpretable, meaning simply ‘roughness’. As a measure of relative isochrony, i.e. of the relative ‘smoothness’ of syllable and foot timing the, *PFD* metric has successfully and consistently discriminated between corpora of different languages, but it has neither measured rhythms nor explained them.

5.2 Rhythmic Irregularity Measure (RIM) metric

The *Rhythmic Irregularity Measure*, *RIM*, (Scot et al. 1986) calculates the sum of the logarithm of the ratios between all durations of non-identical intervals in the utterance:

$$RI = \sum_{i \neq j} \left| \log \frac{I_i}{I_j} \right|$$

Although the ratio looks like a very different kind of measure, it is also a global measure which can also only differentiate along a scale between between ‘smooth’ and ‘rough’. As it stands, it is also dependent on the length of the utterance: the log of the ratios is summed for each pair of intervals. Crucially, like the *PFD*, the ordering of the intervals does not matter: a random ordering of the same values yields the same index. Consequently, the *RIM*, like the *PFD*, measures relative isochrony, i.e. relative ‘temporal smoothness’ as opposed to ‘temporal roughness’ or inequality, and not rhythm. Nevertheless, in this capacity, the *RIM* has consistently succeeded in discriminating between corpora of languages with different temporal patterning, but it has neither measured nor explained rhythms.

5.3 The normalised Pairwise Variability Index metric

One of the most popular ‘rhythm metrics’ (Low, Grabe & Nolan 2000) is the *normalised Pairwise Variability Index* (*nPVI*), which averages the normalised durations differences between neighbouring intervals and multiplying by 100. Normalisation is carried out by dividing the difference between durations by their mean duration:

$$nPVI = 100 \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m-1)$$

The *nPVI* ranges from 0, for totally equal durations, towards an asymptote of 200 for ever ‘noisier’ sets of unequal durations. The limit of 200 arises from normalisation by average (here: division by 2). If normalisation were by the sum of durations, then the asymptote would be 100, a percentage. The *nPVI* has also been used successfully to classify corpora of different languages relatively consistently, but cf. Gut (2012) for discussion of inconsistencies. Like the other metrics, the *nPVI* has its problems as a model of rhythm, though in principle the model looks as though it would work for binary rhythms, where neighbours alternate. But this turns out to be a vain hope:

1. It was noted in cases (1) to (6) that speech rhythm in English is not binary.
2. Taking the absolute values of differences destroys the *strong-weak* versus *weak-strong* directionality of duration change which characterises rhythms: there is no alternation any more.
3. There is another formal problem with taking the absolute difference: many different kinds of duration value can produce the same index. It is easy to check that a sequence such as <1, 2, 1, 2>, a regular, rhythmical alternation, produces a *nPVI* value of 66.66'. It is also easy to

check that monotonically increasing or decreasing geometrical series such as <1, 2, 4, 8> or <8, 4, 2, 1> yield the same *nPVI* value of 66.66'.

4. Also any combination of such series, such as <1, 2, 1, 2, 4, 8, 4, 2, 1, 2, 1> etc. yields the same value of 66.66'. Similarly, the series <1, 3, 1, 3, 1> yields a *nPVI* of 100, and so do corresponding geometrical series such as <1, 3, 9, 27>.

Normalisation for speed rate evidently has a serious down side. The *nPVI* is thus also a measure of 'smoothness' and 'roughness', either of alternation or of geometrical progression, though it implicitly embodies a constraint against random orderings, unlike the other metrics, and it normalises for speech tempo changes.

Like the other metrics, the low values, which indicate evenness, may be related to rhythm, but for the high values it is not clear what is being measured, apart from a degree of unevenness in the duration set. Again, like the other metrics, the *nPVI* has been used successfully in discriminating different timing patterns in corpora of different languages, though it has been neither measuring rhythms nor explaining them.

5.4 *ΔC, %C; ΔV, %V* segmental sequence ratio metrics

A number of descriptive statistical measures were introduced by Ramus and associates (2002; 1999), and were successfully used to discriminate between corpora of different languages. However, a close look shows that the measures essentially reflect the phonotactic structure of the language:

1. The standard deviation of consonantal interval durations (ΔC) relates directly to the complexity of these clusters: in *CV* languages (often associated with 'syllable timing'), ΔC may be predicted to be low, in *CCCVCCC* languages like English (often associated with 'foot timing' or 'stress timing'), ΔC may be predicted to be very high.
2. The standard deviation of vocalic interval durations also relates directly to syllable phonotactics: if a language has a vowel length contrast (English, German), then ΔV may be predicted to be higher than if a language does not have a vowel length contrast (Polish).
3. The percentage of consonantal intervals (%C) in relation to vocalic intervals, or the converse (%V) is a function of the complexity of the phonotactics of the language.

Like the other metrics, the Ramus metrics are 'smoothness' metrics, but in addition, being easily relatable to the phonotactics of languages, they can potentially reflect the genuinely rhythmical property of alternation, since we know – *a priori*, as it were – that consonants and vowels alternate with each other: *syllable-timed* corpora will evidently tend to cluster towards the low ends of the *ΔC* and *ΔV* scales for reasons which are independent of any phonetic measurements, and to have a lower *%C* and higher *%V* ratios. And this is indeed what the studies find by phonetic measurements. Like the other metrics, the segment sequence ratio metrics have also been successfully used to discriminate between corpora of different languages. But, like the other metrics, the segment sequence ratio metrics also do not measure rhythm.

5.5 What do we do with the 'smoothness metrics'?

The *ΔC, ΔV,* %C and %V ratio metrics reflect the syllable patterns in speech corpora and hence, in principle, they can be used to detect the presence or absence of different kinds of alternating pattern or rhythm, though in practice they are not used in this way: if all the consonants are put together in random order, and all the vowels are put together in random order, the metrics still yield the same values.

The situation with the other metrics which are concerned with 'smoothness' versus 'roughness' measures is a little different. They are general metrics, and can in principle be applied to any kind of flow, whether units of speech or distances between cars on the road. They can provide a measure for whether units at a given level are more or less isochronous, but they cannot tell what non-isochrony, i.e. temporal inequality, means, and are therefore ignorant about rhythmic alternations of lack of rhythmic alternations. In order to obtain information about rhythmic alternations, the temporal properties of the constituents of the units measured must be considered (Asu & Nolan 2006; Nolan & Asu 2009).

So if, for example, only with one size of interval, such as the foot, is investigated, not much can be said about the internal structure of the interval. But if they are used to compare the 'smoothness' of interval sequences at different levels, such as syllables as well as feet, then the results can at least be used to determine whether the data are located on a scale between *syllable-timed* and *foot-timed* rhythms: if the *syllable* sequence has the lower *PFD*, *RIM* or *nPVI*, then the utterance is more *syllable-timed*, and if the *foot* has the lower *PFD*, *RIM* or *nPVI*, then the utterance is more *foot-timed*. By triangulating different locations in the architecture of prosody in this way, from phones through syllables and feet to larger units, general statements about the 'syllableness' or the 'footness' of timing can be proposed.

The *ASM-SEM* model of syllable extension developed in the present study predicts that the position of a speech corpus on a scale between *syllable-timed* events and *foot-timed* events is a function of the number *Syllable Extensions* and their duration.

Nevertheless, the metrics still do not explain, in the sense in which, for example, the Jassem model together with the *ASM-SEM* explain, the mechanisms by which *syllable-timed* or *foot-timed* rhythms operate.

6 Rocking the groove

The remaining open issue is, then: if the descriptive 'smoothness metrics' cannot capture rhythm in any transparent fashion, what can? What would actually count as a model of an alternating, 'rocking' rhythmic sequence, as opposed to a model of temporal regularity and irregularity?

The answer lies in recognising the underlying rhythm mechanism as an *oscillator*. A number of oscillator models of rhythm have been proposed (O'Dell & Nieminen 1999; Cummins 2001; Barbosa 2002; Barbosa & da Silva 2012; cf. also Wachsmuth 2002). Barbosa's *Coupled Oscillator Model* will be singled out for brief mention (Figure 8).

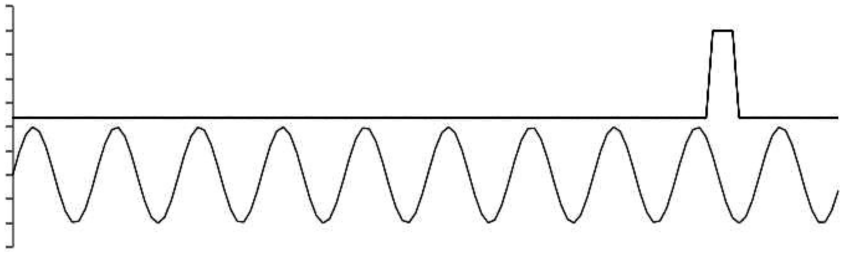

Figure 8: Two-level oscillator model (after Barbosa 2002).

In Barbosa's model, two basic rhythms, the phrase rhythm and the syllable rhythm are postulated (for Brazilian Portuguese – different constructions must be provided for English), and the syllable rhythm is influenced, 'entrained', by the higher level phrase rhythm. It is this interaction of rhythms at different levels which accounts for some of the complexity of the 'groove' of naturally performed speech rhythm. A similarly structured two-level model has been postulated by Fujisaki (1988) for the pitch patterning of accent and intonation.

The oscillator models still have a long way to go, as they are too simple to account for many of the patterns already discussed in this study, but at least they model 'genuine rhythm', which have a well-defined formal, mathematical basis.

The converse issue of oscillator models for analysing rather than generating rhythm is dealt with by Tilsen & Johnson (2008), who propose a 'Rhythm Comb' for identifying the low frequencies in speech which represent rhythms (Figure 9).

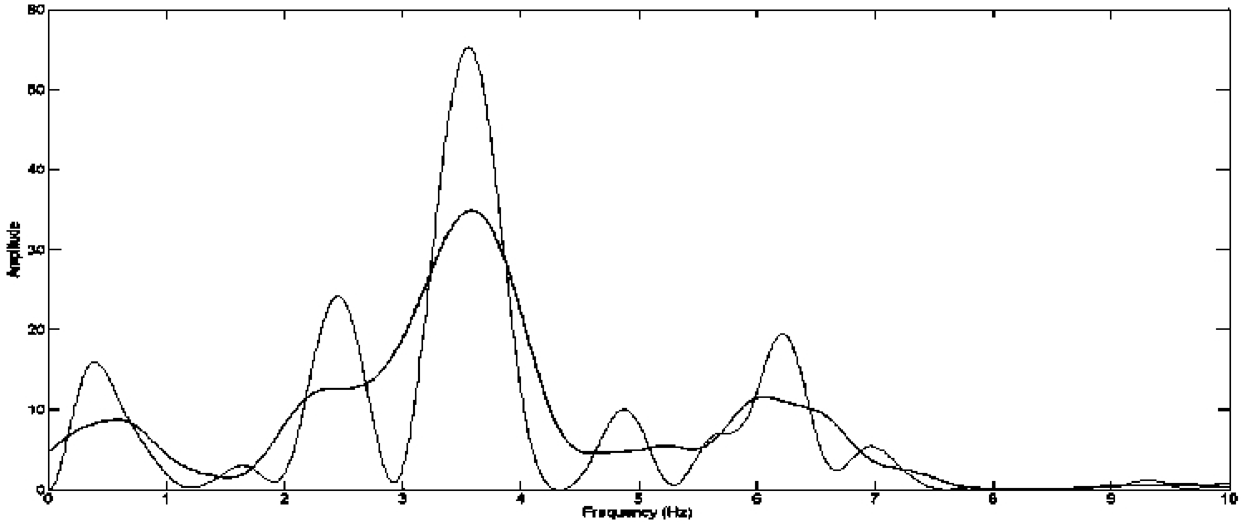

Fig. 4. Raw (blue) and smoothed (red) spectrum. L = 31 points. N = 2048.

Figure 9: Tilsen & Johnson Rhythm Comb Model (2008).

The *Rhythm Comb Model* is essentially a spectrum analysis of long term properties of the speech signal in order to determine not the frequencies of the vocal spectrum, but the frequencies involved in rhythm variation. The *Rhythm Comb Model* of Tilsen & Johnson and Barbosa's *Coupled Oscillator Model* are mathematically closely related.

But the issue of how to relate these oscillator models to explanatory models of prosodic structure is still open. A first step along the road to integrating different approaches to modelling and measuring into an explanatory model has been developed by Wagner (2000), using *Finite State Machine* (*FSM*) models. A related approach in the form of an extension of the *Jassem Rhythm Model* to permit the required iterations of syllables and feet, also using *FSM* models (Figure 10) is discussed in detail elsewhere (Gibbon 2006, 2012).

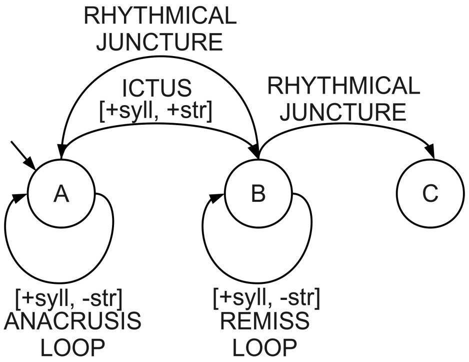

Figure 10: The Rhythm Oscillator Model as an extension of the Jassem Rhythm Model.

An empirical basis for *FSM* models of rhythmic oscillation may be expected in statistically enhanced *FSM*s, which are widely used for many purposes, for instance in the form of *Hidden Markov Models* (*HMM*s) in speech technology (and many other fields). The standard *HMM*s used in speech technology require modification for use in studying duration, however, since in general they factor duration out in favour of generalising over phone intervals.

7 The future of the groove

In this study, methodologies used in the study of rhythm have been queried. It turns out that many of the metrics and models used are not really concerned with 'modelling the groove', that is, the alternating beats of rhythm, but with the temporal evenness or 'smoothness' of sequences, focussing on the criterion of isochrony. Modelling conventions have been explored in some detail, with the aim of developing systemically and phonetically explanatory methods for integrating perspectives which have often been seen as irreconcilable. Starting with a simple model of alternating binary rhythm, the Jassem and Abercrombie models of foot or stress based English rhythm were investigated.

Moving beyond the limitations of these models, an integrative model of syllable and foot based rhythms was proposed, the *Alternating Syllable Model* coupled with the *Syllable Extension Model*, in which the syllable functions as the minimal foot, with a *pre-peak peak post-peak* alternating pattern. The *foot-timed* patterns are derived from *syllable-timed* patterning by *Onset Extension* (*Anacrusis*), stressed *Nucleus* (*Ictus*) and *Coda Extension* (*Remiss*), in the *pre-peak peak post-peak* positions of the *Generic Rhythm Model*, respectively. With the *ASM-SEM* approach, a coherent and comprehensive explanation of how variation between *syllable-timed* and *foot-timed* rhythms take place both in one and the same language, and among languages, is available for the first time.

The field is still wide open. Results of rhythm analysis do not only vary by the methods used and by the modelling conventions followed. Speech rhythm is highly complex and has semiotic functions of cohesion creation, in addition to its formal features of isochrony and alternation. Variation of these forms and semiotic functions by language and dialect, by social formality of style and by functional register of use, have contributed to many of the inconsistencies found in rhythm analyses to date, and are as yet largely unexplored avenues. There is much still to do in 'modelling the groove'.

References

[1] Abercrombie, D. 1967. Elements of General Phonetics. Edinburgh: Edinburgh University Press.

[2] Asu, E. L. & F. Nolan. 2006. Estonian and English rhythm: a two-dimensional quantification based on syllables and feet. In *Proceedings of Speech Prosody 2006, Dresden, Germany*.

[3] Barbosa, P. A. 2002. Explaining Cross-Linguistic Rhythmic Variability via a Coupled Oscillator Model of Rhythm Production. In: Proceedings of Speech Prosody 2002, Aix-en-Provence, 163–166.

[4] Barbosa, P. & W. da Silva. 2012. A New Methodology for Comparing Speech Rhythm Structure between Utterances: Beyond Typological Approaches . In H. Caseli, A. Villavicencio, A. Teixeira, F. Perdigao, eds. *Proceedings of Computational Processing of the Portuguese Language: 10th International Conference, PROPOR 2012*, Coimbra, Portugal, April 17-20, 2012. Berlin: Springer.

[5] Campbell, N. 1992: *Multi-level timing in speech*. Ph.D. thesis, University of Sussex.

[6] Cummins, F. (2002): Speech Rhythm and Rhythmic Taxonomy. In *Proceedings of Speech Prosody 2002, Aix-en-Provence*. 121–126.

[7] Dechert, H. W. & M. Raupach, eds. 1980. *Temporal variables in speech. Studies in Honour of Frieda Goldmann-Eisler*. The Hague: Mouton.

[8] Fujisaki, H. 1988. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In O. Fujimura, ed., *Vocal physiology: voice production, mechanisms and functions*, New York: Raven. 347-355.

[9] Gibbon, D. 2001. Preferences as defaults in computational phonology. In K. Dziubalska-Kołaczyk, ed., *Constraints and Preferences*. Trends in Linguistics, Studies and Monographs 134. Berlin: Mouton de Gruyter. 143-199.

[1] Gibbon, D. 2006. Time Types and Time Trees: Prosodic Mining and Alignment of Temporally Annotated Data. In: S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter & J. Schließer, eds., *Methods in Empirical Prosody Research*. Berlin: Walter de Gruyter. 281-209.

[2] Gut, U. 2012. Rhythm in L2 speech. In D. Gibbon, D. Hirst & N. Campbell, eds, *Rhythm, Melody and Harmony in Speech. Studies in Honour of Wiktor Jassem*. Poznań: Polskie towarzystwo Fonetychne (Polish Phonetics Association).

[3] Hirst, D. & C. Bouzon. 2005. The effect of stress and boundaries on segmental duration in a corpus of authentic speech (British English. In *Proceedings of Interspeech 2005, Lisbon*. 29-32.

[4] Jakobson, Roman. 1940. *Kindersprache, Aphasie und allgemeine Lautgesetze.* Uppsala: Almqvist & Wiksells.

[5] Jassem, W. 1949. indikeiʃn əv spiːtʃ riðm in ðə traːnskripʃn əv edjukeitid sʌðən ingliʃ (Indication of speech rhythm in the transcription of educated Southern English). *Le Maître Phonétique*, III/92, 22-24. [Republished in Journal of the International Phonetic Association with the original IPA version and a new orthographic version by D. Hirst].

[6] Jassem, W. 1952. *Intonation of Conversational English (Educated Southern British). Prace Wrocławskiego Towarzystwa Naukowego (Travaux de la Société des Sciences et des Lettres de Wrocław).* Seria A. Nr. 45. Wrocław: Nakładem Wrocławskiego Towarzystwa Naukowego.

[7] Jassem, W., D. R. Hill & I. H. Witten. 1984. Isochrony in English Speech: its Statistical Validity and Linguistic Relevance. In D. Gibbon & H. Richter, eds. *Intonation, Accent and Rhythm: Studies in Discourse Phonology*. Berlin: Mouton de Gruyter. 203-225.

[8] Low, E. L., E. Grabe & F. Nolan. 2000. Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech* 43, 4, 377–401.

[9] Nolan, F. & E. L. Asu. 2009. The Pairwise Variability Index and coexisting rhythms in language. *Phonetica* 66, 64-77.

[10] O'Dell, M. L. & T. Nieminen (1999): Coupled oscillator model of speech rhythm. In *Proceedings of the International Congress of Phonetic Sciences*, San Francisco 1999.

[11] Ramus, F. (2002): Acoustic correlates of linguistic rhythm: Perspectives. In *Proceedings of Speech Prosody 2002, Aix–en–Provence*. 115–120.

[12] Ramus, F., M. Nespor & J. Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 3, 265–292.

[13] Roach, P. 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. In Crystal, D., ed., *Linguistic Controversies: Essays in Linguistic Theory and Practice*. London: Edward Arnold. 73–79.

[14] Scott, D. R., S. D. Isard & B. de Boysson-Bardies. 1986. On the measurement of rhythmic irregularity: a reply to Benguerel. *Journal of Phonetics* 14, 327–330.

[15] Tilsen, S. & Johnson, K. 2008. Low-frequency Fourier analysis of speech rhythm. *Journal of the Acoustical Society of America*, 124, 2: 34-39.

[16] Wachsmuth, I. 2002. Communicative rhythm in gesture and speech. In McKevitt, P., C. Mulvihill & S. O'Nuallain eds. *Language, Vision and Music*. Amsterdam: John Benjamin. 117–132.

[17] Wagner, P. 2001. Rhythmic alternations in German read speech. In *Proceedings of Prosody 2000, Poznan*. 237–245.