**Original Paper**

Phonetica 2009;66:78–94
DOI: 10.1159/000208932

Received: September 15, 2008
Accepted: January 20, 2009

# Do Rhythm Measures Reflect Perceived Rhythm?

William Barry[a] Bistra Andreeva[a] Jacques Koreman[b]

[a]Saarland University, Saarbrücken, Germany; [b]Norwegian University of Science and Technology, Trondheim, Norway

**Abstract**

In a production study, Bulgarian, English and German verses with regular poetic metrical metres of different types and elicited prose utterances with varied accentual patterns are produced in textual and iterative (*dada*) form and measured at syllable level according to the pairwise variability index (PVI) principle. Systematic differences in PVI values show that the measure is sensitive to metrical differences. But variations for utterances with the same metrical structure and comparable measures for accentually different utterances show the measure to be insensitive to the temporal distribution of accents. A perceptual experiment with Bulgarian, English and German subjects confirms the hypothesis that the perceived strength of rhythmicity in a line of verse is determined not only by its temporal structure, but also by other acoustic properties, most clearly by $F_0$ change within the metrical foot.

Copyright © 2009 S. Karger AG, Basel

## Introduction

Rhythm typology has its roots in auditory observation [Lloyd James, 1940] popularly explained as perceived regularities in the occurrence of feet, syllables and (a later addition) morae. The terminology, *stress-timing, syllable-timing* and *mora-timing*, has, however, become increasingly divorced from the auditory impressions, taking on the status of abstract typological features for the languages of the world. The lack of any physical evidence of isochrony corresponding to the perceived regularities has led more recently to a reversal in the focus of the search, to differences in the degree of (the ubiquitous) durational variability. Syllable variability measures have successfully separated languages statistically [Ramus et al., 1999; Grabe and Low, 2002]. But the wide scatter of measures behind the averaged points in the 'rhythm space' (whether for language, text or speaker) casts doubt on a perceptual rhythmic basis to the measures or at least on their status as a characterization of a language rhythm type [Barry et al., 2003; Russo and Barry, 2008]. In fact, since the abandonment of the (perceptually understandable) isochrony search, there has been no statement of what differences between

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2009 S. Karger AG, Basel
0031–8388/09/0662–0078
$26.00/0
Accessible online at:
www.karger.com/pho

William J. Barry
Myrtle Cottage, Bearwood Road
Sindlesham-Wokingham
Berkshire RG41 5BB (UK)
E-Mail wbarry@coli.uni-saarland.de

Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

languages in the variation measures mean in perceptual terms. On the other hand, the everyday experience that different utterances (or prosodic phrases within utterances) can sound rhythmically different, i.e., that each individual, prosodically coherent unit has its own (information-) structurally determined rhythm, would lead one to expect a wide scatter of values. This would not necessarily undermine the idea of the measure as a reflex of the realized surface rhythm, though it again raises the difficult question of what is actually meant by the 'rhythm of a language'. Perceptually speaking, the 'rhythm' prototype of a particular language would be built up from the statistical variation among the myriad rhythmically different utterances. The fact remains, however, that very little research effort has been directed towards the link between the measures and the perceivable realized rhythm.

Limited perceptual studies serving the concept of rhythm typology show some ability to *categorize* delexicalized utterances as belonging to languages of different rhythm types [Ramus et al. 2003], but a marked asymmetrical error pattern argues against any general typological validity. Utterances with sequences of low-complexity syllables from a language with a high variability in syllable complexity (e.g. English) may be categorized as belonging to a language with a low variability in syllable complexity (e.g. Spanish, Italian, Japanese), but not vice versa. This result supports the above-mentioned suggestion that rhythm type, in terms of language rhythm, is a perceptual prototype. It would appear, however, that no work has explicitly addressed the relationship between the concrete rhythmic measure and the rhythmic impression of the utterance that produced it. This article attempts to do so. Thus the issue addressed here is not a rhythm-typology one (though the inclusion of three languages retains a strong typological implication). It addresses the relationship between the rhythmic impression, which can be assumed to arise with any utterance, whatever the language, and the physical structure of that utterance. This it does in two ways: firstly, to avoid problems of perceptual judgement of postulated rhythm categories, we compare utterances that are defined, in terms of poetic metrical structure, as rhythmically the same or rhythmically different. We also attempt to separate the syllabic-structural element by measuring both the original textual versions and iterative *dada* versions of the same utterances. To gain insight into the manifestation (in terms of rhythm measures) of metrical rhythmic structure across languages, we consider comparable patterns in Bulgarian, English and German. The measure we employ is a syllable-based pairwise variability index (PVI-syllable). In this first part, we exploit defined (poetic) rhythmic types and perform a production analysis.

Secondly, taking up observations from the production results which indicate that (i) perceived rhythm is the product of other properties than just duration, (ii) the same metrical rhythm can be realized with differing relative contributions of duration and melodic change, and (iii) languages differ in the degree to which they modify duration for accentual purposes (see 'Results' section below), we addressed the question of whether the *perceptual weighting* of the parameters contributing to the strong-weak alternation of trochaic rhythm differs across languages. There are clear implications for rhythm typology if this were the case. To this end, we examined the physical basis of degrees of perceived rhythmicity within one metrical pattern. Together with duration and fundamental frequency ($F_0$), the prime candidates implicated in the production study, intensity and vowel quality were manipulated in order to ascertain the relative contribution of the four parameters to the impression of rhythmicity. Again we compare across languages, taking the same stimuli for Bulgarian, English and German listeners.



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

## Production Study

### *Material and Speakers*

Two speakers per language (1 F, 1 M) read or recited verses with an iambic/trochaic, dactylic/anapest or a more complex metre. These were children's rhymes or specially constructed verses. The verses are given in the 'Appendix'. In addition, short sentences, presented visually, were produced in answer to questions which required a shift of focus (broad-focus or with narrow focus on an early but not initial or a late but not final content word) so that the same segmental material was produced with different rhythmic patterns. This material was considered to bridge the gap between rhythmically defined verse and normal prose. In order to have rhythmically defined material *without* the variation in structural complexity, the speakers also produced all verses and utterances in an iterative *dada* form, repeating meaningful sections of the verses or the sentence immediately after reading or reciting the textual version.

### *Hypotheses*

In the material with poetic metrical structure, the weak-strong or strong-weak alternations (and therefore alternating longer and shorter syllables) in the iambic/trochaic lines are expected to have a higher PVI than dactylic/anapest or more complex metrical forms, where two or more unstressed (shorter) syllables occur in sequence, reducing the sum of differences between consecutive syllables. Figure 1 illustrates this expectation.

In the non-poetic material with focus shift, one hypothesis is that a broad-focus utterance will have a higher PVI than either narrow-focus utterance because the greater number of phrasally accented words in a broad-focus utterance (with longer lexically stressed syllables, cf. Turk and Sawusch [1997], de Jong [2004] for English) should increase the number of large differences between accented and unaccented syllables.

An alternative hypothesis could postulate lower variability for *early* focus (due to de-accenting after the nuclear accent) than for *late* focus, which may have only slight de-accenting on prenuclear elements. This, combined with a stronger nuclear contrast, could lead to a PVI which is equal to, or even exceeds the broad-focus index.

### *Results*

Results are presented separately for each language. The results for the English children's verses broadly support the above hypotheses (fig. 2). The dactylic and more complex metrical patterns ['Humpty Dumpty' (hd) and 'Christmas is Coming' (xmas)] have lower syllabic variability indices than the (predominantly) iambic/trochaic verses ['Mary had a Little Lamb' (mary) and 'Mary Mary Quite Contrary' (mc)].

The differences are greater in the *dada* versions, but there is unexpected disagreement between speakers 1 and 2 in the values for 'mary' (compare mary1 and mary2), which is not found in the textual versions. Auditory analysis revealed that speaker 2 produced the *dada* version with clear iambic/trochaic rhythm but (atypically for an English speaker) retained a full vowel in the unstressed syllable rather than reducing it to schwa. As a result, the durational contrast between strong and weak was less.

The average variability indices for the Bulgarian poems contradict the hypotheses. The variability indices for the dactylic/anapest poems lie between those of the two iambic/trochaic poems in the *dada* versions, while in the textual versions, the values for the dactylic/anapest poems are higher (with one exception) than for the iambic/



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

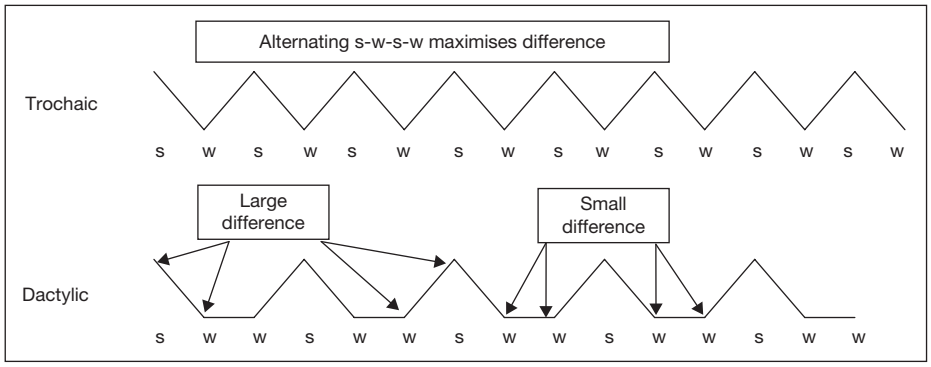


**Fig. 1.** Schematic illustration of relationship between PVI value and rhythmic pattern. s = Strong; w = weak.

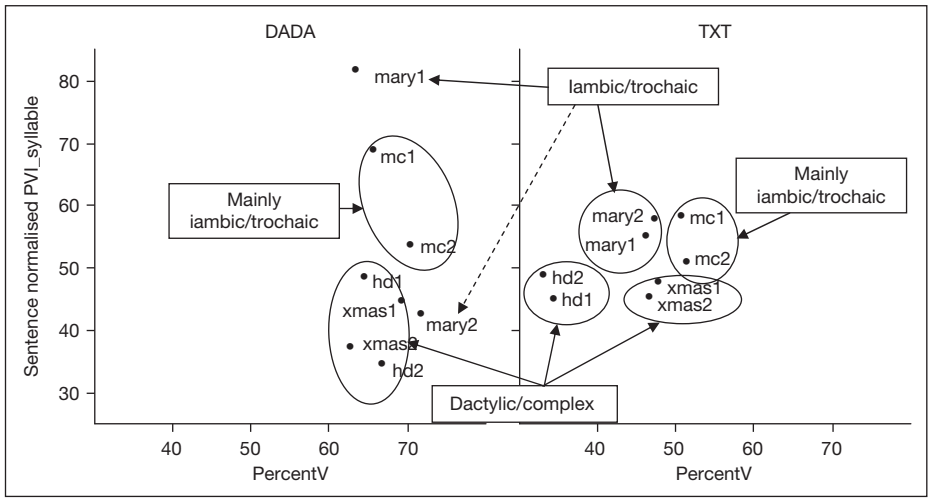


**Fig. 2.** Normalized syllabic PVI values (plotted against percent vowel) for English children's verses (text and *dada* versions, see 'Appendix' for abbreviations).

trochaic ones. Again, an auditory analysis revealed divergence in the surface realization from the strict metrical pattern assumed for the verses. For example, nominally iambic/trochaic lines were produced with a de-accenting of the central feet, resulting in extended sequences of non-accented (and therefore minimally varying) syllables. This is illustrated in figure 3, showing syllabic PVIs for 12 lines (line numbers are identified in the figure) of the *dada* version of the third Bulgarian poem (Ов'чарски 'куче'та са 'двете,...., see 'Appendix'), which is nominally trochaic with an upbeat on the first syllable. Those lines which are realized with a strictly alternating rhythm have higher variability indices than the majority, which have one de-accented foot. Line 6, in which two consecutive feet are de-accented by both speakers, resulting in a sequence of 5 unstressed syllables, has the lowest value. An important reason for this systematic deviation from the nominal metrical pattern lies in the fact that trochees represent a



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

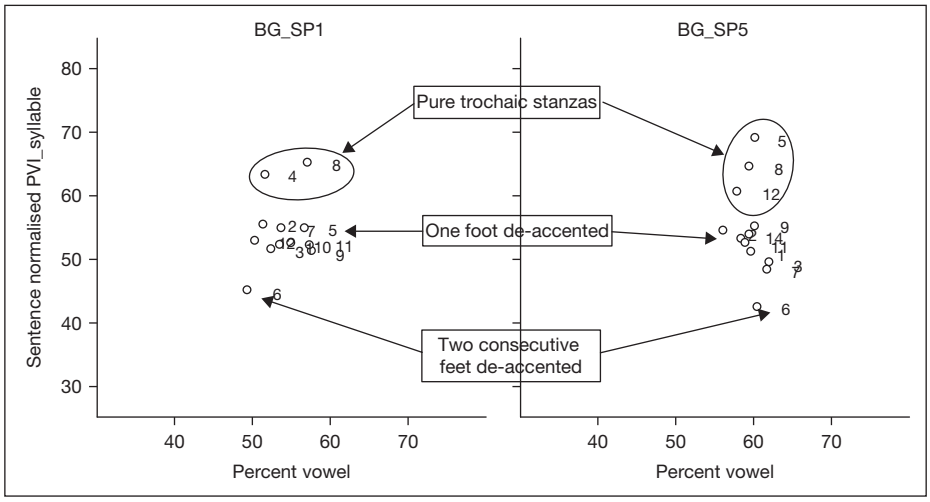


**Fig. 3.** Normalized syllabic PVI values (plotted against percent vowel) for the 12 individual lines of the nominally trochaic Bulgarian verse (text version, speakers shown separately).

rhythmic pattern which is not ideally suited to Bulgarian morpho-syntax (e.g., the definite and indefinite article are post-appended, requiring monosyllabic nouns – which are relatively rare – to produce a trochaic NP pattern).

In the German data there is agreement between the textual and the *dada* versions. However, contrary to the hypothesis, the highest variability is found for the dactylic/anapest and the lowest for the ostensibly iambic/trochaic poem. Auditory examination of the iambic poem 'Es war einmal ein Mann' reveals that the lines were phrased in couplets with a *less strongly* differentiated weak-strong pattern in the odd lines, the central beat being audibly weakened. According to the hypothesis, this should lead to a lower PVI measure for the odd lines. However, figure 4 indicates a *higher* value in most couplets. An explanation for this paradox may lie the fact that the clearer prominences on the strong syllables of the even lines are carried more by tonal accents than by duration, whereas the central beat in the odd lines has a 'force accent' or 'accent of pressure' (German 'Druckakzent', i.e. an accent without pitch movement or pitch contrast to the surrounding syllables), which is less auditorily prominent despite (generally) greater durational differentiation. The PVI is, of course, purely duration-based and cannot reflect the stronger beat that stems from the stronger tonal modulation.

Overall, the PVI values are relatively low with considerable overlap between the odd and even lines. This is attributable to the lack of phonological vowel reduction in the unstressed syllables in German and parallels the production pattern by the second English speaker's *dada* version of 'Mary' reported above.

*Results for Non-Poetic 'Rhythm'*

Figure 5 gives average syllable durations for the text version of the same sentence produced under different focus conditions (averages over 6 repetitions each by 2



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

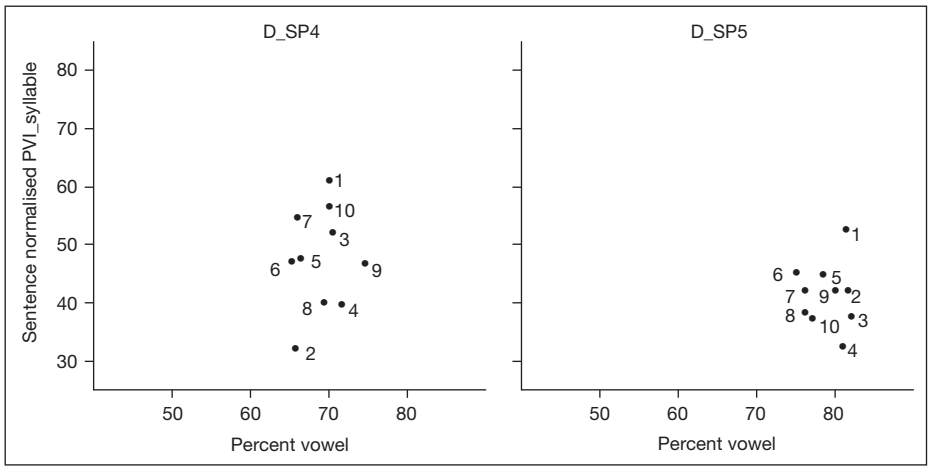


**Fig. 4.** Normalized syllabic PVI values (plotted against percent vowel) for individual lines (numbered 1–10) of the iambic German children's poem 'Es war einmal ein Mann' (text version), in which the lines were produced in couplets, the first line with a weaker and the second line with a perceptually stronger iambic rhythm.

speakers of each focus condition). The arrows on the duration plots for the contrastive conditions indicate the change in duration for the particular syllable in comparison to the broad-focus condition.

The patterns of change over the focus conditions and the degree of PVI change vary considerably over the three languages. In the English sentence *The cop found the burglar there.* there is very little change of PVI value with change of focus. In the change from broad to early focus there is a strongly *increased* durational contrast between the first *the* and *cop* and a slight increase between *found* and the second *the*. This is counteracted by a decrease of contrast in the second half (lengthened *-glar* and shortened *there*). The change from broad to late focus shows almost no change in the first half (decreased first *the* and increased *found*) and, in the second half, a larger increase in the duration of *bur-* than *-glar* in the focused word *burglar*. This results in a slight increase in the PVI value from 78.4 to 80.0.

In the Bulgarian sentence (*Igrax na dama bez kaka ti* 'I played at draughts without my elder sister'), both narrow-focus versions show considerable increase in PVI value over the broad-focus condition. The change from broad to early focus brings a decrease in the duration of *-grax* and the preposition *na*, and an increase in the stressed syllable of the focused word *dama*. This increases the contrasts between *na* and *da-* and between *da-* and *-ma* in the first part of the sentence. The only change of any magnitude in the second half is the reduction in the final syllable, lowering the *ka-ti* contrast. For the late focus condition, there is greater contrast in both halves: in the second half through a slightly shorter *bez* and a much longer first syllable in the focused word *kaka*, and in the first half through a reduction of *na* and a slight deaccentuation of *da-*.

In the German sentence (*Der Mann fuhr den Wagen vor* 'The man drove the car up'), the early-focus condition shows a large increase in PVI value over the broad-focus condition while the increase in the late-focus version is very slight (comparable



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

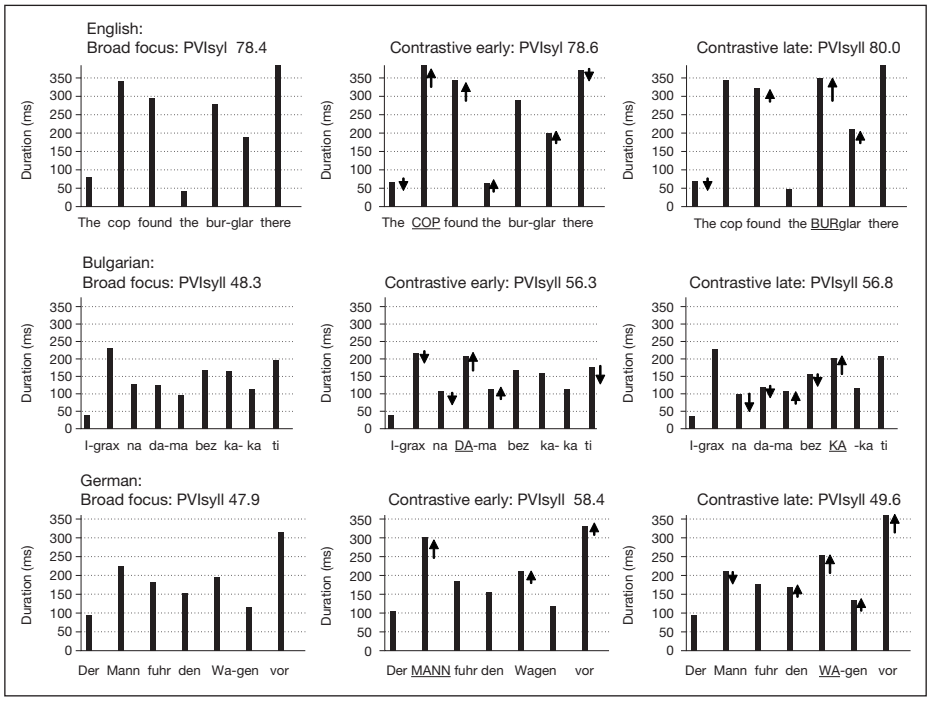


**Fig. 5.** Average syllable durations for the text version of the same sentence produced under different focus conditions in English, Bulgarian and German (6 repetitions each by 2 speakers per language of each focus version). Arrows indicate durational change from broad focus; length of arrow shows relative amount.

to the English sentence). The change from broad to early focus is almost exclusively carried by an increase in the duration of the focused word *Mann* (though the force-accented *Wa-* also increases slightly over the nuclear *Wa*' in the broad-accent version). The late-focus condition shows a big increase in duration of the first syllable of *Wagen*, which results in greater contrast in the second half, though not as much as there would be if *den* and *-gen* had stayed the same. In any case, the increase is counteracted by a slight de-accenting of *Mann*, which reduces the contrast between the syllables *der Mann* and *fuhr* and, with it, the PVI value.

### *Discussion*

The PVI clearly captures more than just differences in syllabic structure. Systematic shifts in PVI-syllable values between textual versions of poems with differing metrical rhythms remain when the structural differences have been removed and replaced by an iterative *dada* structure. The same general statement is valid for changes in the 'rhythm' of non-poetic material as a result of focus shift (fig. 5).

As a global variability measure across a given stretch of text (here it was a line of a poem or a sentence in the non-poetic material, corresponding in both cases to an intonation unit), the measure reflects the average change in duration from syllable to



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

syllable, without sensitivity to the location of the large and small changes. Therefore a change in metre from e.g., iambic to anapest will normally result in a decreased PVI-syllable value, as our schematic representation shows, but so will the reduction of an accent *anywhere* in the line or sentence. The nature of the measure as a *global* reflection of syllable-to-syllable duration change means that massive changes in non-poetic 'rhythm', with de-accentuation of one part of a sentence combined with stronger accentuation of another part, can result in exactly the same PVI measure as the same sentence without either modification.

Differences are apparent between the three languages examined here even when the structural differences have been removed. However, at least some of the differences are traceable to structural origins: (1) The lower PVI value for *dada*-iterative iambic lines in the German rhymes may be attributed to the lack of (phonological) reduction (to schwa) of the /a(:)/ vowel in unstressed syllables. (2) The lower PVI value for nominally iambic lines in Bulgarian may be attributed to the frequent occurrence of multisyllabic words (partly through cliticization) without secondary accents, making an iambic or trochaic metre difficult to maintain in the surface realization. Finally, non-durational (i.e. tonal) properties also contribute to the rhythmic beats of an utterance, as indicated for instance in (i) the inverse relation of the PVI values to the perceived prominence differences found in the 'Mann' poem in German; (ii) the shifts of focus in the English non-poetic material, which was auditorily very clear but had almost no reflex in the PVI measure. It is important to bear in mind that, just as intonation should not be seen as tonal structure alone, so rhythm should not be seen exclusively as the product of durational relations.

Our revised critique of the PVI measures must therefore be that they do not *only* reflect statistical differences between languages in terms of their syllable structures, a representation which is open to variation with the choice of textual materials, the speakers selected and the style and tempo with which they produce the texts. Part of the textual and style-based variation *is* clearly 'rhythm'-dependent, be it poetic metrical rhythm or non-verse phrasal accentuation patterning. However, here as in other aspects, differences between PVI indices are non-specific, in that different accentual changes can result in the same PVI shift. Furthermore, there is no evidence yet that this rhythmic/accentual component of the PVI would contribute to an informative rhythm-typological separation of languages.

## Perception Experiment

To test more systematically for parameter interactions of the kind observed in the production data, a perceptual experiment was carried out. The aim was to examine the relative perceptual weight of the four parameters that are generally considered to correlate with prominence and may therefore be expected to contribute to perceived rhythmicity: $F_0$, duration, vowel quality and intensity.

### *Task and Material*

Pairs of 8-syllable trochaic nonsense (*dada*) lines which differed in the values of the prominence-lending parameters were presented for judgement. Subjects heard the two lines with a 0.5-second



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

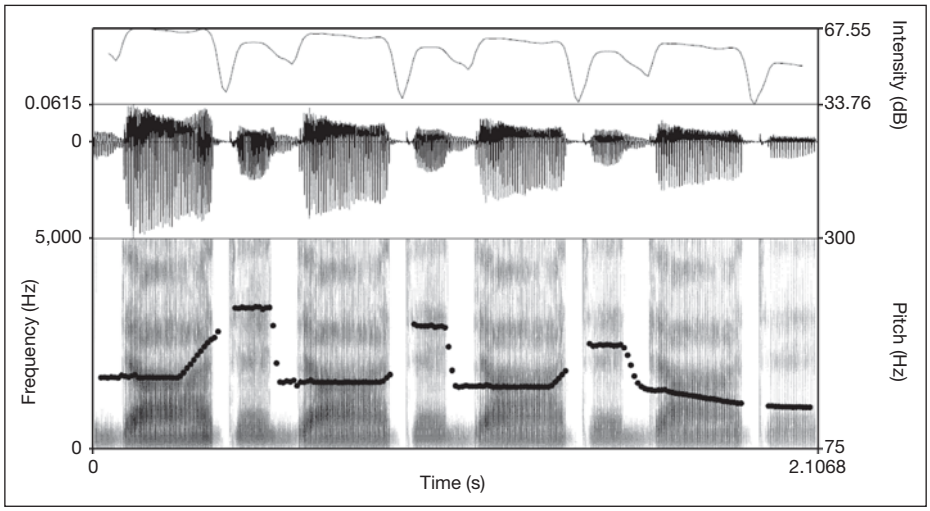


**Fig. 6.** Spectrogram of fully structured 8-syllable stimulus showing the generated $F_0$ contour, and the intensity time course (above).

separation and adjusted a slide on a monitor to indicate the degree to which one or the other stanza was 'more strongly rhythmical'. An upward shift from the neutral midway slide position indicated greater rhythmicity of the first line, a downward shift indicated the stronger rhythm of the second line. Apart from a mid-point marker, there was no overt scale (e.g. ± 1–10). Subjects chose their own range, and the positions on the hidden scale of ±100 were normalized for each subject by calculating z scores.

The pairs were controlled to differ in one, two, three or all four prominence parameters. They were constructed from a /ˈdaːdə/ foot extracted from an English *dada* rendition of the Max and Moritz stanza 'Of two youths named Max and Moritz' (cf. www.fln.vcu.edu/mm/mm-vor_dual.html for a dual-language version of the poem) and concatenated to form the required 8-syllable sequence. English was selected because it had the greatest long-short syllable-duration contrast among the three languages examined, together with a clear vowel-category change to schwa on the unstressed syllable. The vowel-quality shift was also found in the Bulgarian material, but the quality of the English speaker's voice was better suited to resynthesis for repeated listening. The 8-syllable constructed line was given a linear 25-Hz declination from 150 to 125 Hz over 7 syllables and a further 5-Hz fall-off over the final syllable which served as a baseline for all stimuli. A linear intensity declination of 2 dB per syllable was imposed on syllables 1–7 with a 4-dB fall-off on the 8th syllable which also served as the intensity baseline for all stimuli.

A 'trochaic' and a 'neutral' condition was defined for each parameter:

(i) *Duration.* The original durations of 352 and 164 ms for the stressed and unstressed syllables, respectively, were maintained for the 'trochaic' condition (dur+). An equal duration of (352 + 164)/2 = 258 ms was imposed on both syllables for 'neutral' (dur–).

(ii) *$F_0$.* The intonation of a German line was schematized and imposed on the $F_0$ declination baseline for the trochaic condition ($F_0$+). This comprised a L* on the stressed syllable rising to a H on the unstressed syllable for foot 1–3 followed by downstepped !H* on the 4th stressed syllable with a low phrase boundary on the final (unstressed) syllable, falling to a low boundary tone (see the final fall-off in the declination pattern described above). Figure 6 shows the spectrogram with the $F_0$ contour



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

superimposed. The neutral contour ($F_0$–) was identical with the declination baseline (i.e. the 150- to 125-Hz fall-off over 7 syllables and a further 5-Hz fall-off over the final syllable described above).

(iii) *Intensity.* The degree of intensity change between stressed and unstressed syllables was modeled on the Bulgarian data, where 3- to 4-dB change was typical, much higher than in German or English. To enhance the possibility of intensity contributing to rhythmicity a change of 5 dB was defined for the trochaic condition (int+) and superimposed on the intensity declination baseline. The resulting intensity contour is shown in figure 6. The neutral condition (int–) followed the intensity declination baseline pattern of 2 dB per syllable and 4 dB fall-off finally.

(iv) *Vowel Quality.* The original /ɑː/ – /ə/ qualities were retained from the trochaic condition (vq+). For the neutral condition (vq–) the /ɑː/ was replaced by /ə/. The choice of two /ə/ syllables rather than two /ɑː/ syllables was dictated by naturalness; the concatenation of a durationally reduced full /ɑː/ quality was perceived as unnaturally hyperarticulated.

The stepwise modification of the full trochaic lines results in 16 decreasingly 'trochaically defined' stanzas (table 1).

Stimulus 1 and 16 were paired with each of the other stimuli, resulting in 30 stimulus pairs, 15 with the fully defined trochaic line in the first position (1 + n pairs) and 15 with the completely neutral line in the first position (16 + n stimuli). The stimulus combinations and the parameter modifications are shown in table 2. These 30 stimulus pairs were presented four times in different randomizations (120 pairs) preceded by five practice pairs for which the judgements were not registered. Subjects were 18 Bulgarian (Sofia), 16 English (South East) and 20 German (South West) native speakers.

The analysis focused on possible differences between the three languages and/or between subjects within a language in the effect of a given parameter change on the perceived strength of rhythmicity.

### *Results*

To compare judgements across subjects and languages, the raw difference values were transformed to z-values. A two-way univariate analysis of variance with language and parameter as independent variables showed a significant effect of language (d.f. 2, $F = 18.6$, $p < 0.001$) and parameter (d.f. = 29, $F = 71.67$, $p < 0.001$) with a significant interaction of language and parameter (d.f. = 58, $F = 6.14$, $p < 0.001$). The Scheffé post hoc test showed that each of the languages differed significantly from the others. A parallel two-way analysis of variance for subject and parameter (a three-way analysis language × subject × parameter is not possible because the subjects are included in the languages) also showed a highly significant effect for subjects (d.f. = 53, $F = 35.985$, $p < 0.001$) but examination of the post hoc grouping of subjects showed no language-dependent tendency. In other words, the main language effect did not manifest itself in any recognizable grouping of the subjects in terms of the strength of their reaction to particular parameter manipulations.

Since there was a significant subject × parameter interaction (d.f. = 1,537, F 4.36, $p < 0.001$), we tested the degree of agreement between the subjects in the strength of the parameter influence on perceived rhythmicity. Spearman rho correlations were calculated for possible differences in reactions to the two directions of parameter manipulation (compare table 2: 1 + n: neutralization and 16 + n: enhancement). Taking all three languages together, the rank order of the stimulus pairs in their effect on perceived strength of rhythmicity was highly correlated for the 1 + n and the 16 + n stimuli ($n = 15$; $r = 0.829$, $p < 0.001$); i.e., averaging the judgements over all the subjects, the scores were similar irrespective of whether the differences between the stimuli within the stimulus pair were the result of parameter neutralization or enhancement. Table 3



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

**Table 1.** Parameter-value combinations employed in the stimulus definition

| Stimulus | $F_0$ | Duration | Intensity | Vowel quality |
|---|---|---|---|---|
| 1 | + | + | + | + |
| 2 | + | – | + | + |
| 3 | + | + | – | + |
| 4 | + | + | + | – |
| 5 | + | – | – | + |
| 6 | + | + | – | – |
| 7 | + | – | + | – |
| 8 | + | – | – | – |
| 9 | – | + | + | + |
| 10 | – | – | + | + |
| 11 | – | + | – | + |
| 12 | – | + | + | – |
| 13 | – | – | – | + |
| 14 | – | – | + | – |
| 15 | – | + | – | – |
| 16 | – | – | – | – |

**Table 2.** Stimulus combinations employed in the rhythmicity judgement task and associated parameter manipulations

| Stimulus combination | Parameter change | Stimulus combination | Parameter change |
|---|---|---|---|
| 1-2 | – duration difference | 16-15 | + duration difference |
| 1-3 | – intensity difference | 16-14 | + intensity difference |
| 1-4 | – vowel-quality difference | 16-13 | + vowel-quality difference |
| 1-5 | – duration and intensity | 16-12 | + duration and intensity |
| 1-6 | – duration and quality | 16-11 | + duration and quality |
| 1-7 | – intensity and quality | 16-10 | + intensity and quality |
| 1-8 | – duration, intensity and quality | 16-9 | + duration, intensity and quality |
| 1-9 | – $F_0$ difference | 16-8 | + $F_0$ difference |
| 1-10 | – $F_0$ and duration | 16-7 | + $F_0$ and duration |
| 1-11 | – $F_0$ and intensity | 16-6 | + $F_0$ and intensity |
| 1-12 | – $F_0$ and quality | 16-5 | + $F_0$ and quality |
| 1-13 | – $F_0$, duration and intensity | 16-4 | + $F_0$, duration and intensity |
| 1- 14 | – $F_0$, intensity and quality | 16-3 | + $F_0$, intensity and quality |
| 1-15 | – $F_0$, duration and quality | 16-2 | + $F_0$, duration and quality |
| 1-16 | – $F_0$, duration, intensity and quality | 16-1 | + $F_0$, duration, intensity and quality |

– and + indicate that the specified parameter contrasts were neutralized (–) and enhanced (+), respectively.



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

shows the relative position of the stimuli in the ranked order from most to the least important.

Averaging over the subjects of each language individually, however, it was found that Bulgarian differed considerably from the other two languages: $r = 0.211$, $p = 0.451$ for Bulgarian, while $r = 0.829$, $p < 0.001$ for German and $r = 0.825$, $p < 0.001$ for English; i.e., for the Bulgarian subjects, the impression of rhythmic strength was influenced by whether the fully structured or the fully neutralized line was heard first. Correlations calculated between languages for the two presentation directions show that the Bulgarian subjects differ from German and English only for the 16 + n stimuli, i.e. when the fully neutralized line was heard first. Tables 4a, b show that there was high agreement between the languages for the 1 + n stimulus pairs.

Against this background of basically systematic rhythmicity judgements, we addressed the main question behind the perceptual task, namely the relative importance of the four modified parameters.

Taking the order for the stimulus pairs shown in table 3 and averaging the summed rank for the 8 occurrences of each parameter as a 'rhythm index', we find the values as shown in table 5. To clarify the meaning of the numerical values: since each parameter occurs 8 times in different combinations, the numerical value of highest possible average rank is the sum of ranks 1–8 divided by 8 = 4.5; the lowest possible average rank is the sum of ranks 9–16 divided by 8 = 11.5. Table 5 indicates that duration is the most important factor, followed by $F_0$, intensity and vowel quality.

The inter-language correlation coefficients indicated, however, that the weighting of the parameters is not constant from one language to another. The 'rhythm indices' calculated from the stimulus rankings are given in table 6. The most striking differences between the languages are seen in the Bulgarian results. In contrast to the clear ranking: 'duration higher than $F_0$' found for German and English, the Bulgarian subjects rank $F_0$ equally highly, mainly due to a lower overall rating for duration compared to the other two languages. Vowel quality and to some extent intensity are also rated higher (lower numerical value) by Bulgarian subjects than by German and English subjects, resulting in a smaller weighting range across the four parameters. One-way ANOVAS for each parameter (using the difference scores for stimuli sharing the same manipulated parameter as the dependent factor, irrespective of which other parameters were also manipulated) showed significant differences between the languages (d.f. = 2, duration: $F = 6.356$, $p = 0.002$; $F_0$: $F = 6.633$, $p = 0.001$; intensity: $F = 5.229$, $p = 0.005$: vowel quality: $F = 8.103$, $p < 0.001$). Post hoc tests showed in each case a separation of Bulgarian and English, with German taking an intermediate position.

### *Discussion*

The main finding from the perception experiment is that perceived rhythm is certainly not just a product of durational structuring, though the strong position that duration enjoys in the rhythm-index hierarchy confirms its importance. Change in $F_0$ also contributes strongly to the impression of rhythm, supporting the observation made on the German production data, where stronger $F_0$ differences with weaker durational differences (between stressed and unstressed syllables of the trochaic feet) sounded more rhythmical than the more monotone line with stronger durational differences.



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

**Table 3.** Overall order of degree-of-rhythmicity difference for the 1 + n and 16 + n stimulus pairs

| Order of importance | 1 + n stimuli | 16 + n stimuli |
|---|---|---|
| 1 | +4 fdiq | –3 fdi |
| 2 | +3 fdq | –1 d |
| 3 | +3 fdi | –2 di |
| 4 | +2 di | –2 fd |
| 5 | +2 fd | –3 fdq |
| 6 | +1 d | –4 fdiq |
| 7 | +3 diq | –2 dq |
| 8 | +2 fi | –2 fi |
| 9 | +1 f | –3 diq |
| 10 | +2 dq | –1 f |
| 11 | +1 i | –2 fq |
| 12 | +2 fq | –2 iq |
| 13 | +3 fiq | –3 fiq |
| 14 | +1 q | –1 q |
| 15 | +2 iq | –1 i |

f = $F_0$; d = duration; i = intensity; q =quality.

**Table 4.** Inter-language correlations for ranked strength of stimulus effect on rhythmicity judgements: 1 + n stimuli (a) and 16 + n stimuli (b)

| | | 1-first_BG | 1-first_D | 1-first_GB |
|---|---|---|---|---|
| a. 1-first | | | | |
| 1-first_BG | | 1.000 | 0.711(**) | 0.721(**) |
| | sig. | | 0.003 | 0.002 |
| 1-first_D | | 0.711(**) | 1.000 | 0.871(**) |
| | sig. | 0.003 | | 0.000 |
| 1-first_GB | | 0.721(**) | 0.871(**) | 1.000 |
| | sig. | 0.002 | 0.000 | |
| b. 16-first | | | | |
| 16-first_BG | | 1.000 | 0.118 | 0.196 |
| | sig. | | 0.676 | 0.483 |
| 16-first_D | | 0.118 | 1.000 | 0.486 |
| | sig. | 0.676 | | 0.066 |

It is tempting to ascribe the different perceptual weighting of the parameters to language-inherent properties. It has recently been argued [Dimitrova, 1998] that Bulgarian is a more syllable-timed language than previously assumed. Despite the well-known vowel reduction in unstressed syllables and the existence of quite complex syllable structures in the syllabary, consideration of Dauer's [1987] catalogue of linguistic characteristics places it further from the stress-timed pole than either German or English. The lack of a short-long vowel opposition in Bulgarian is an important



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

**Table 5.** 'Rhythm indices' for the four parameters modified in the stimulus pairs averaged over all subjects

| | 1 + n stimuli | 16 + n stimuli |
|---|---|---|
| Duration | 4.75 | 4.63 |
| $F_0$ | 6.63 | 7.25 |
| Intensity | 7.75 | 8.38 |
| Vowel quality | 9.25 | 9.63 |

feature of this placement. In Barry et al. [2003], rhythm measures also showed different styles of Bulgarian grouping closer to the 'syllable-timed' area of the rhythm space than either German or English. Thus, the lower weighting of duration (higher numerical value) found in this perceptual study might be seen as confirmation of this claim. The overall weighting was, however, influenced strongly by 4 Bulgarian subjects who systematically judged the stimuli with durationally equal syllables to be *more* rhythmic than the durationally trochaic ones (and vice versa). It therefore remains a moot point whether their rhythm-judging strategy and the fact that *no* subjects from the German or English groups judged the stimuli in that manner may be seen as supporting evidence for the tendency towards more 'syllable timing' in Bulgarian. The alternative is to see these Bulgarian subjects as just applying their own (idiosyncratic) criteria. One thing is clear, despite significant language differences: between-subject variation is strong, as the lack of grouping of subjects according to language showed, indicating the non-categoriality of rhythm and the possibility of adopting various strategies.

Differences between the languages in the weighting of vowel *quality* as a factor in rhythm perception also argue against a linguistic grounding. The higher weighting from the Bulgarian subjects might be seen as a corollary of the phonologized vowel reduction found in unstressed syllables. The contrast to German would seem to support this standpoint. However, it stands in direct contrast to the very low weighting by the English subjects despite a similar process of vowel reduction in English.

## Conclusions

The results of this study cast some light on the complex phenomenon of rhythm in speech, but it also opens up further questions. Returning to the impressionistic roots of speech rhythm and avoiding the typological abstractions that grew out of them, we could confirm that within recent quantitative approaches different metrical rhythms *can* be reflected in different rhythm measures. But it is certainly not the case that different measures are necessarily an indication of a perceptually different metrical rhythm. Particularly the freedom to vary the prominence of beats within a given metre (which may possibly be described as varying the 'strength' of the same rhythmic pattern) can result in widely diverging values. Conversely, changing the position of strong beats while maintaining the relative strength can result in the same measure but a different perceived rhythmic pattern. The latter case is typical of non-poetic speech, where different information-structural demands can result in very different accent placement and associated rhythm. In cross-language (and thus potentially typological) terms, there was some evidence that different morpho-syntactic regularities lead to differences in



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

**Table 6.** Parameter ‘rhythm indices’ for Bulgarian, German and English subjects

| | Bulgarian | | German | | English | |
|---|---|---|---|---|---|---|
| | 1 + n stimuli | 16 + n stimuli | 1 + n stimuli | 16 + n stimuli | 1 + n stimuli | 16 + n stimuli |
| Duration | 5.88 | 6.75 | 5.88 | 5.13 | 4.50 | 4.50 |
| $F_0$ | 6.00 | 6.38 | 6.25 | 7.25 | 6.75 | 7.00 |
| Intensity | 8.00 | 7.88 | 8.13 | 10.00 | 8.00 | 8.63 |
| Vowel quality | 8.25 | 6.13 | 9.75 | 10.38 | 9.00 | 9.25 |

what is a typical number of un- or de-accented syllables between phrasally prominent syllables, which can affect both the typical rhythm measure and the typical rhythmic impression.

Within the production study, nearly all of the variation in ‘rhythmic strength’ was reflected in the durationally based rhythm measure, but there was evidence from some of the variation that duration and other properties – $F_0$ at least – were complementing one another. This was taken up in the perception experiment, which confirmed that $F_0$ changes within the foot function as a very strong secondary, or in the case of Bulgarian subjects almost equal factor in determining rhythmic strength. Intensity and vowel quality appear to be almost negligible factors for (most) German subjects (the average rhythm index is >9 and >10, respectively). For the English subjects they are slightly higher ranked, but they are most important for (at least some) Bulgarian subjects.

The present data do not suggest that the perceptual patterning has a linguistic basis, but further investigation must ascertain whether the broad inter-subject variability observed is a product of the method or in the nature of the speech-rhythm phenomenon itself.

## Appendix

*Bulgarian Verses*

1. Шамˈпиони по ˈбягане ˈмного боˈгати, ˈистински ˈкучета ˈаристоˈкрати.
Не ˈискат цеˈлувки, не ˈискат преˈгръдки, а ˈдивеч да ˈгонят, заˈщото са ˈхрътки.
Виж ˈзайчето ˈбяга, сърˈцето му ˈтупка, след ˈмиг ще се ˈскрие във ˈсвоята ˈдупка.
За ˈтях то е ˈплячка и ˈжива иˈграчка, доˈбре че до ˈвкъщи е ˈсамо на ˈкрачка.
2. С ˈято ˈпатки ˈцяло ˈлято ˈси жиˈвях боˈгато в ˈблато.
ˈПълнех ˈгушка, ˈно със ˈпушка, ˈвзе ме ˈмлад лоˈвец на ˈмушка.
С ˈловно ˈкуче ˈнова ˈмода, ˈМюнстерˈлендер по поˈрода,
ˈСтреля ˈцялаˈта неˈделя, ˈгръмна ˈстараˈта ми ˈлеля.
3. Овˈчарски ˈкучеˈта са ˈдвете, | Те ˈпазят от злиˈни овˈцете.
Оˈглежда ˈхълмоˈвете ˈголи | Безˈстрашноˈто шотˈландско ˈколи.
То ˈима ˈдълга ˈтопла ˈгрива, | През ˈзимаˈта да ˈне насˈтива.
До ˈнего ˈвярнаˈта съˈпруга | Си ‘има ˈднес заˈдача ˈдруга –
Тя ˈпази ˈагънˈцата ˈвсяко | С ˈплаква ˈсе, че ˈиска ˈмляко.
Поˈрасналˈли са ˈвече ˈвсички, | Ще ˈтрябва ˈда паˈсат треˈвички.
4. Пролетˈта изпълˈзя сред цветˈя, | Долеˈтя радостˈта сред деˈца,
Изпълˈзя таз мечˈта от дуˈша | И заˈпя песенˈта за мечˈта.
Любовˈта за деˈца тя поˈся, | Радостˈта от дуˈша тя заˈпя,
Таз мечˈта с пролетˈта долеˈтя, | Таз мечˈта за деˈца в песенˈта.



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

*English Verses*
1. (‘mary’) ˈMary ˈhad a ˈlittle ˈlamb, its ˈfleece was ˈwhite as ˈsnow,
And ˈeveryˈwhere that ˈMary ˈwent the ˈlamb was ˈsure to ˈgo.
It ˈfollowed ˈher to ˈschool one ˈday, which ˈwas aˈgainst the ˈrule.
And ˈall the ˈchildren ˈlaughed and ˈsang to ˈsee a ˈlamb at ˈschool.
2. (‘mc’) ˈMary ˈMary ˈquite conˈtrary ˈhow does your ˈgarden ˈgrow?
With ˈsilver ˈbells and ˈcockle ˈshells and ˈlittle maids ˈall in a ˈrow.
3. (‘hd’) ˈHumpty ˈDumpty ˈsat on a ˈwall | ˈHumpty ˈDumpty ˈhad a great ˈfall.
ˈAll the king’s ˈhorses and ˈall the king’s ˈmen | ˈcouldn’t put ˈHumpty toˈgether aˈgain.
4. (‘xmas’) ˈChristmas is ˈcoming and the ˈgoose is getting ˈfat.
ˈPlease put a ˈpenny in the ˈold man’s ˈhat.
If you ˈhaven’t got a ˈpenny a ˈhaˈpˈny will ˈdo,
If you ˈhaven’t got a ˈha’p’ny a ˈfarthing will ˈdo.
If you ˈhaven’t got a ˈfarthing then ˈGod bless ˈyou!

*German Verses*
1. Es ˈwar einˈmal ein ˈMann, | Der ˈhatte ˈeinen ˈSchwamm.
Der ˈSchwamm war ˈihm zu ˈnass, | Da ˈging er ˈauf die ˈGass.
Die ˈGass war ˈihm zu ˈkalt, | Da ˈging er ˈin den ˈWald.
Der ˈWald war ˈihm zu ˈgrün, | Da ˈging er ˈnach Berˈlin.
Berˈlin war ˈihm zu ˈvoll, | Da ˈging er ˈnach Tiˈrol.
2. Auf dem ˈBaum sitzt ein ˈMann | Mit dem ˈSpecht in der ˈHand
Und erˈzählt, was er ˈsieht. | Wenn der ˈSpecht von ihm ˈfliegt,
Kommt der ˈSpecht zu dem ˈMann | Setzt er ˈsich auf die ˈHand
Von dem ˈMann auf dem ˈBaum | Der den ˈSpecht dort beˈstaunt.
3. ˈHeute ˈbin ich ˈwild und ˈböse,
ˈBin ein ˈWolf im ˈgrauen ˈFell,
ˈBin ein ˈDrache, ˈbin ein ˈLöwe,
ˈUnd ich ˈbeiße ˈund ich ˈbell.
ˈIch zerˈtrete ˈzwanzig ˈSchnecken,
ˈUnd ich ˈmache ˈganz viel ˈKrach,
ˈSchneide ˈLöcher ˈin die ˈDecken,
ˈMache ˈmeine ˈSchwester ˈwach.
ˈHeute ˈbin ich ˈwild und ˈböse,
ˈUnd ich ˈgehe ˈnicht ins ˈBett,
ˈKnalle ˈTüren ˈmit Geˈtöse,
ˈBin ganz ˈkratzig ˈbin nicht ˈnett.

## Acknowledgements

Our grateful thanks to Sibylle Kötzer for producing the stimuli for the perception experiment. The work was supported by the German Research Council, grant No. BA 737/10–1.

## References

Barry, W.J.; Andreeva, B.; Russo, M.; Dimitrova, S.; Kostadinova, T.: Do rhythm measures tell us anything about language type? Proc. 15th Int. Congr. Phonet. Sci., Barcelona 2003, pp. 2693–2696.
Dauer, R.M.: Phonetic and phonological components of language rhythm. Proc. 11th Int. Congr. Phonet. Sci., Tallinn 1987, vol. 5, pp. 447–450.
de Jong, K.: Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. J. Phonet. *32:* 493–516 (2004).
Dimitrova, S.: Bulgarian speech rhythm: stress-timed or syllable-timed? J. int. phonet. Ass. *27*: 27–33 (1998).
Grabe, E.; Low, E.L.: Durational variability and the rhythm class hypothesis; in Gussenhoven, Warner, Papers in Laboratory Phonology, vol. VII, pp. 515–546 (2002).



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM

Lloyd James, A.: Speech signals in telephony (Pitman & Sons, London 1940).
Ramus, F.; Dupoux, E.; Mehler, J.: The psychological reality of rhythm classes: perceptual studies. Proc. 15th int. Congr. Phonet. Sci., Barcelona 2003, pp. 337–342.
Ramus, F.; Nespor, M.; Mehler, J.: Correlates of linguistic rhythm in the speech signal. Cognition *73*: 265–292 (1999).
Russo, M.; Barry, W.: Isochrony reconsidered: objectifying relations between rhythm measures and speech tempo. Proc. Speech Prosody, Campinas 2008, pp. 419–422.
Turk, A.E.; Sawusch, J.R.: The domain of accentual lengthening in American English. J. Phonet. *25:* 25–41 (1997).



Downloaded by:
Saarländische Universitäts und Landesbibliothek
134.96.104.233 - 6/5/2014 1:52:25 PM