A Rhythmic-Prosodic Model of Poetic Speech

Jörg Bröggelwirth
Institut für Kommunikationsforschung und Phonetik
Universität Bonn, Germany
jbr@ikp.uni-bonn.de

Abstract
In this paper a new approach towards the analysis of speech rhythm is presented. In the speech rhythm literature it was often discussed that rhythmic phenomena are more transparent in the metrical structure of orally produced poetry. However, up to now only a few phoneticians have worked on this special speaking style. For analyzing the rhythmic and prosodic patterns of this kind of speech, a corpus of read German poetry, including four different meters, was recorded. This study gives a first sight on durational and intonational effects in the data. A final prosodic modeling and its perceptual evaluation is currently under development.

1. Introduction
Pike [1] and Abercrombie [2] classified languages by the regularity of the occurrence of syllables or stresses in the time domain. Their isochrony hypothesis claimed that any language was either syllable-timed (equal duration of consecutive syllables) or stress-timed (equal duration of inter-stress-intervals). Even though Kohler ([3], [4]) showed a tendency towards isochrony in German, the strong form of the hypothesis was rejected [5]. Thus, new approaches of describing speech rhythm were needed. As for language typology the dichotomy of syllable-timed vs. stress-timed is still maintained. However, the assignment of acoustic correlates to either of these categories changed. In [6] the two rhythm classes were described by the percentage of vowels (%V) and the standard deviation of consonant cluster durations (ΔC) in the overall utterance. Cummins [7] criticizes this formalism for not capturing perceived rhythm directly. According to this an adequate account for the description of speech rhythm is yet to be developed.

As segmental and suprasegmental features both influence the rhythm of speech [8] a different perspective of speech rhythm is presented in [9] by analyzing the beats of rhythmic units instead of measuring the durational patterns of speech segments only. The results of a speech cycling task expose the so-called harmonic timing effect, i.e. the preferred placement of accents at points of 1/3, 1/2 and 2/3 of the time of a superordinated cycle or phrase, respectively. Additionally, Morton and colleagues [10] show that the physical onset of a speech signal is not the same as its psychological onset, i.e. the P-center. Their study yields that the P-center lies somewhere in the consonant-vowel transition.

The aforementioned studies point out that on the one hand there are different views of speech rhythm and on the other hand speech rhythm is difficult to measure. One way of dealing with these difficulties is to investigate special speaking styles which are supposed to be more rhythmic than others. One candidate of such a style is poetry. The poetic speaking style is said to have optimal prosodic oppositions with a clearer realization of most of the effects outlined above ([3], [4], [12], [13]).

Up to now there are only rare studies about the acoustic-phonetic characteristics of poetic speech rhythm. Kruckenberg and Fant [14] analyzed poetry in Swedish. They recorded a set of poems comprising iambic and trochaic meters spoken in prose as well as poetic style. The analysis of the data indicates a certain poetic attitude which speakers show while reading poetry. This attitude is especially characterized by a lower tempo, higher pitch, higher intensity, a smaller modulation depth of local F0 variations and a lack of a verse final drop of pitch. Furthermore, the durations of metrical feet have a smaller standard deviation and decrease linearly with the position within the actual verse. Additional to this, the contrast between the pitch of the stressed and the unstressed syllable of an iamb is higher than that of a trochee. Moreover, the poetic speaking style shows a stronger tendency towards isochrony than prose does. This is underlined by the rhythmic continuity across verse lines meaning that pause spanning feet have durations which are whole-numbered multiples of the duration of the average foot late in the verse. At the same time the duration of these pauses decreases with an increase of the number of phones within the interval.

Additionally, an interesting observation regarding read poetry is made by Kien and Kemp [15]. Their study points out that the duration of biological action units, i.e. the basic units of any motoric behavior, with about 1-4 s roughly corresponds to the mean duration of poetic verses. Therefore the verse is an important semantic unit with a strong inner coherence.

2. A corpus of read German poetry
As described above, poetry reading can be seen as a phonetic subject useful for the analysis of rhythmic-prosodic phenomena. So far there does not exist a large corpus of that kind for German. For this reason, a corpus of read German poetry was recorded. This database had to meet certain criteria.

2.1. Text selection
For the purpose of an analysis of the acoustic-phonetic characteristics of the metrical structure of poetry in German, the corpus to be recorded had to be large enough to yield valid statistics and small enough to be manageable. At the same time the data should cover a range of different criteria. The first important criterion was to include the most important meters of German poetry, i.e. iamb, trochee, dactyl and song.

For iamb and trochee three poems or extracts of poems, respectively, were selected, whereas for the other two meters two poems were considered. In this respect it was not easy to assign the mentioned meters to certain poems. Many poems, especially the ones supposed to be dactylic, consist of a mixture of different meters. Anyway, with a little support
from Dr. von Tiedemann from the German literature department Bonn it was possible to find quite modern poems which fitted very well.

2.2. Recording

For the recording of the corpus we consulted 12 speakers, 6 professional and 6 nonprofessional ones. Each of the two groups comprises 3 male and 3 female speakers. Educated actors were taken as professional speakers. The nonprofessional speakers were students of a variety of different subjects. Because the corpus was supposed to provide information about the difference between poetic and prosaic reading, both speaking styles were recorded in a manner following Kruckenberg and Fant [14]. Each text was first manipulated in such a way that it looked like prose with changes regarding for example orthography and punctuation. Then the readers were instructed to read all of these poems in a prose mode followed by the reading of the poetic originals. In the latter case the speakers were told to realize the meter assigned to each poem. Because the nonprofessional speakers had a tendency to realize the meter in a staccato like style, they were encouraged to go on like that if only they kept their style consistent.

Every reader was recorded during one session of about 2 hours. The participants were paid for this. Moreover, a stereo signal of a microphone as well as a laryngograph was recorded in the soundproof chamber of the Institute of Communication Research and Phonetics of the University of Bonn. Both channels were digitized with 16 kHz and 16 bits. Some passages of the recording material had to be repeated so that the resulting data had to be cut.

2.3. Annotation

For a profound analysis of the recorded corpus of German poetry, a linguistic-phonetic annotation of the data was necessary. Among these features are, for instance, word and phrase boundaries, metrical feet (inter-stress intervals), word stress, perceptual syllable prominence, global F0 movement at boundaries, acoustic labeling of intonation as well as (automatically generated) GToBI labels [16]. A couple of additional variables relevant for a statistical evaluation of the corpus’ prosody was annotated. Syllable duration, for example, is particularly dependent on information about a syllable’s position in its corresponding foot or phrase and its constituent number of phones ([17], [18]).

3. Analysis

The current study has two main aims. First, it is attempted to gather basic knowledge about the prosodic characteristics of the poetic speaking style in German. Second, it is aimed at finding out what kind of acoustic-prosodic information is necessary for a listener to identify this special style. These two goals are connected with each other by modeling and thus predicting the syllable-based prosody, i.e. duration and intonation disregarding intensity, of the poetic speaking style. This is work in progress. Anyway, some first results focusing on the iambic and trochaic meters are presented in the following sections.

3.1. Duration

As for the durational evaluation of poetic speaking style some interesting observations were made. The rhythmic continuity described by Kruckenberg and Fant [14] for Swedish was not found in our German data. This became clear by an investigation of the durations of pause spanning feet which did not show the expected regularity described in [14].

Figures 1 and 2 show a comparison between the distributions of phrase and verse durations of the poetic versions of the iambic and trochaic data. In contrast to the phrase duration distribution the steep fall at the lower end of the verse duration distribution with a minimum of about 2 s (excluding short outliers) is very typical for the poetic verse unit. Furthermore, the poetic variants of our iambic and trochaic verses have a durational median of 3.3 s suggesting that they might be regarded as articulatory action units as described in [15].
An investigation of foot durations showed medians of 518 ms for the iambs in poetic mode, 423 ms for the iambs in prosaic mode, 477 ms for the trochees in poetic mode and 385 ms for the trochees in prosaic mode. These values reveal that iambic feet have longer durations than trochaic feet and poetic feet have longer durations than prosaic ones. Thus foot duration specifies the actual meter as well as the speaking style, i.e. poetry or prose. Furthermore, the relation between the duration of the unaccented and the accented syllable of a foot is characteristic for its corresponding meter. The medians of this unaccented/accented quotient are as such: 60.8% for the poetic iambs, 61.0% for the prosaic iambs, 67.7% for the poetic trochees and 77.2% for the prosaic trochees. The difference between the medians of the two poetic variants is statistically significant (Mann-Whitney test; p < 0.001).

Moreover, a stronger tendency of the poetic feet as a whole towards isochrony could not be found. Figures 3 and 4 depict two scatterplots showing the dependency between the number of phones within a foot and its corresponding foot duration of the iambs and trochees in poetic mode. The two plots compare the values of the nonprofessional with the professional speakers. As can be seen the nonprofessional speakers tend to produce feet with a small number of phones with higher durations than the professional speakers do. This fact combined with the slightly flatter regression line of the nonprofessional speakers’ data suggests that there is a stronger tendency towards isochrony for nonprofessional speakers articulating iambs and trochees in a poetic mode.

3.2. F0

For a pitch analysis the mean F0 for each syllable in the corpus was computed. Because the F0 values are dependent on a speaker’s individuality (especially gender), the F0 data were corrected for the latter factor by means of a z-normalization. The z-values of the iambs and trochees in poetic mode are higher than their prosaic counterparts. At the same time trochees have higher values than iambs. The F0 difference caused by the higher pitch of the poetic data as compared to the prosaic data is statistically significant (T-test; p < 0.001). As for F0 analysis there is still work to be done.

4. Prosodic models of poetic speech

A major aim of this study is to model the rhythmic-prosodic characteristics of poetic speech. Based on the aforementioned statistics syllable durations of read poetry for the meters iamb and trochee are to be modeled. Dactyl and song are deferred at the moment.

4.1. Duration models

The duration of poetic syllables is modeled in two ways. The first way is a rule-based one using the concept of isochrony. The second way is a statistical one using regression models. Both models are supposed to be used as syllable duration predictors for a prosody manipulation of speech signals. In the near future these two models will be evaluated by listening experiments with the manipulated stimuli.

4.1.1. Isochronous model

Based on the relations of the duration of the unaccented and accented syllables of a metrical foot combined with the meter specific medians of iambs and trochees read in the poetic mode, syllable duration can be roughly modeled by an isochronous approach.

\begin{equation}
\text{x} + 0.608 \times \text{x} = 518
\end{equation}
In the same way the durational median of the poetic trochee is used to compute its accented and unaccented syllable durations. This can be verified in formula (2).

\[ x + 0.677 \times x = 477 \]  

(2)

4.1.2. Regression model

An expectedly more natural alternative to the isochronous model is a regression model of poetic syllable duration. After a factor analysis and a stepwise linear regression the predictor space could be reduced to 3 predictors for each meter. The iamb’s syllable durations can be predicted by accentuation of the syllable, position of the syllable in the phrase and number of phones within the syllable. The trochee’s syllable durations, on the other hand, can be predicted by position of the syllable in the foot, position of the syllable in the phrase and number of phones in the syllable.

The quality of the regressions is meter and speaker dependent. The correlation coefficients for the iamb’s lie between 0.69 for the worst and 0.76 for the best speaker approximation. The range of the trochees is 0.53 to 0.75. These results are comparable to the prediction quality known from speech synthesis systems. For the forthcoming duration manipulation of speech signals the model of the speaker with the best approximation will be used.

5. Discussion

This study presents a new approach of analyzing the rhythmic-prosodic characteristics of speech by focusing on the special speaking style of read poetry in German. A large corpus of that kind was recorded and segmentally as well as prosodically annotated. The first analyses show meter specific differences of duration and intonation. These differences are regarded as rhythmic differences.

6. Conclusions

It was often stated by phoneticians that poetry has a stronger tendency towards isochrony. But although nonprofessional speakers are closer to isochrony than professional speakers, this assumption has to be rejected. Nevertheless the concept of isochrony will be used for an upcoming experimental prosody manipulation of speech signals. Two models for a poetic duration prediction, an isochronous and a regression method, respectively, are proposed. In the future it is planned to add an intonation model of poetic speech and to evaluate the quality of the duration and intonation models by listening experiments. Furthermore, it will be tested which one of the two modeled prosodic cues, intonation and duration, is the more important for the identification of the poetic speaking style. Is it possible to identify a certain meter relying on prosody alone? Which prosodic feature characterizes a meter the best? Duration and intonation will be tested independently from each other.

7. Acknowledgements

The research reported here was made possible by a research grant of the German Research Council DFG, PN 84026. We would like to thank Dr. von Tiedemann for his support.

8. References