SOME TIMING STUDIES OF PROSE, POETRY AND MUSIC

Lennart Nord, Anita Krucken berg and Gunnar Fant

Department of Speech Communication and Music Acoustics, KTH,
Box 700 14, S-100 44 Stockholm, Sweden

ABSTRACT

Prosodic relations in prose, poetry and music are discussed with an emphasis on durational properties. In order to gain a deeper understanding of speech prosody, we are presently engaged in a comparison of the timing relations in such activities as the reading of poetry and music performance, where there usually is a strong and obvious rhythmic patterning of the produced sound sequences. Also there are interesting parallels to be drawn by comparing the formal notations of prose, poetry and music. Generally, there are no simple relations between abstract notations and performance, and moreover, notations have varied with tradition and particular needs. However, it is a challenge to tie descriptive systems closer to common human constraints in production and perception.

INTRODUCTION

We have recently extended our interest in speech prosody and rhythmical structures in reading to cover analogous phenomena in poetry, and also with an ambition to relate our findings to established knowledge of music structure and music performance, thus adding to the perspective of a common denominator - a common code.

We benefit from our close contacts with the music acoustics group at KTH [1]. In a recent article, our colleagues in the speech and music research, [2] compared their modelling of speech and music performance. We also rely on earlier work on Swedish poetry reading by Goude and Malmström [3] and Goude et al [4], who have investigated the perception of rhythm in poetry, and furthermore, the rhythm research at Uppsala University by Bengtsson and Gabrielsson [5].

We have a general interest in the acoustic correlates of rhythmical performance in prose, poetry and music. Our main research objective is to model the prosodic structure of fluent speech, including individual variations of reading styles. The degree of rhythmicality varies considerably in speech and is generally perceived at an intuitive level. A natural variation of stress pattern and speaking rhythm is essential for good speech synthesis quality.

In this paper, an outline for research will be made, as well as some comparisons between prose, poetry and music. We are concentrating on the durational properties, postponing a treatment of other prosodic signals, notably pitch moments.

In timing studies, with the focus on objective and subjective measures of rhythmical structures, a number of questions can be formulated, such as: what are the rhythmic entities in speech and music; what is the timing accuracy in the perception and production of a rhythmic pulse in read texts; what is the correspondence between rhythm and meter in read poetry, the relevance of the concept of rising and falling metres (e.g. iambic versus trochaic feet), etc.

PROSE

Style

Reading and speaking style has not been studied in phonetics for a very long time. In the 18th century, books were available with instructions on how to read aloud, with careful advice on how to make a pause and silently count to one for a comma in the text, to two for a semi-colon, three for a colon, four for a period etc., in order to improve the reading by a correct timing. See [6] for a review of this area.

Speech production models

In the seventies, rhythmical speech was studied at a detailed level by phoneticians with the aim to better understand the coordination of articulatory movements in the modelling of speech production, [7-10]. In a typical experimental paradigm, subjects were told to synchronize their speech to a pulse train or to press a button at the perceived onset of a syllable. Those tests gave an insight about the planning and coordination of syllable production. The location of so called "p-centers" (p for pulse) for CV syllables, differed in distance to the vowel onset depending on the complexity of the preceding consonantal part [11].

Fluent speech analysis

There is a renewed interest in studies of text reading, prosody and reading styles. More powerful speech analysis facilities have become available and data bases are created at many places, one of the aims being to handle long stretches of speech, [12]. In an earlier stage phonetics laboratories were mostly concerned with "lab speech", i.e. isolated words and short sentences, which suited old spectrographs, but said little about the prosody of natural, fluent speech.

Table 1

<table>
<thead>
<tr>
<th>Level</th>
<th>Communication Form</th>
<th>Abstract Notation</th>
<th>Signal Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Speech</td>
<td>Text (prose)</td>
<td>Speech</td>
</tr>
<tr>
<td>II</td>
<td>Poem (language theory)</td>
<td>Text (verse)</td>
<td>Speech</td>
</tr>
<tr>
<td>III</td>
<td>Music (music theory)</td>
<td>Musical notation</td>
<td>Music performance</td>
</tr>
</tbody>
</table>

In order to gain a deeper understanding of speech prosody, we are presently engaged in a comparison of the timing relations in such activities as the reading of poetry and music performance, where there usually is a strong and obvious rhythmic patterning of the produced sound sequences. Also there are interesting parallels to be drawn by comparing the formal notations of prose, poetry and music. Generally, there are no simple relations between abstract notations and performance, and moreover, notations have varied with tradition and particular needs. However, it is a challenge to tie descriptive systems closer to common human constraints in production and perception.

We benefit from our close contacts with the music acoustics group at KTH [1]. In a recent article, our colleagues in the speech and music research, [2] compared their modelling of speech and music performance. We also rely on earlier work on Swedish poetry reading by Goude and Malmström [3] and Goude et al [4], who have investigated the perception of rhythm in poetry, and furthermore, the rhythm research at Uppsala University by Bengtsson and Gabrielsson [5].

We have a general interest in the acoustic correlates of rhythmical performance in prose, poetry and music. Our main research objective is to model the prosodic structure of fluent speech, including individual variations of reading styles. The degree of rhythmicality varies considerably in speech and is generally perceived at an intuitive level. A natural variation of stress pattern and speaking rhythm is essential for good speech synthesis quality.

In this paper, an outline for research will be made, as well as some comparisons between prose, poetry and music. We are concentrating on the durational properties, postponing a treatment of other prosodic signals, notably pitch moments.

The following three levels form the basis in a comparison between the three forms of communication:

<table>
<thead>
<tr>
<th>Level</th>
<th>Communication Form</th>
<th>Abstract Notation</th>
<th>Signal Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Speech</td>
<td>Text (prose)</td>
<td>Speech</td>
</tr>
<tr>
<td>II</td>
<td>Poem (language theory)</td>
<td>Text (verse)</td>
<td>Speech</td>
</tr>
<tr>
<td>III</td>
<td>Music (music theory)</td>
<td>Musical notation</td>
<td>Music performance</td>
</tr>
</tbody>
</table>

In order to gain a deeper understanding of speech prosody, we are presently engaged in a comparison of the timing relations in such activities as the reading of poetry and music performance, where there usually is a strong and obvious rhythmic patterning of the produced sound sequences. Also there are interesting parallels to be drawn by comparing the formal notations of prose, poetry and music. Generally, there are no simple relations between abstract notations and performance, and moreover, notations have varied with tradition and particular needs. However, it is a challenge to tie descriptive systems closer to common human constraints in production and perception.

We benefit from our close contacts with the music acoustics group at KTH [1]. In a recent article, our colleagues in the speech and music research, [2] compared their modelling of speech and music performance. We also rely on earlier work on Swedish poetry reading by Goude and Malmström [3] and Goude et al [4], who have investigated the perception of rhythm in poetry, and furthermore, the rhythm research at Uppsala University by Bengtsson and Gabrielsson [5].

We have a general interest in the acoustic correlates of rhythmical performance in prose, poetry and music. Our main research objective is to model the prosodic structure of fluent speech, including individual variations of reading styles. The degree of rhythmicality varies considerably in speech and is generally perceived at an intuitive level. A natural variation of stress pattern and speaking rhythm is essential for good speech synthesis quality.

In this paper, an outline for research will be made, as well as some comparisons between prose, poetry and music. We are concentrating on the durational properties, postponing a treatment of other prosodic signals, notably pitch moments.

The following three levels form the basis in a comparison between the three forms of communication:

<table>
<thead>
<tr>
<th>Level</th>
<th>Communication Form</th>
<th>Abstract Notation</th>
<th>Signal Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Speech</td>
<td>Text (prose)</td>
<td>Speech</td>
</tr>
<tr>
<td>II</td>
<td>Poem (language theory)</td>
<td>Text (verse)</td>
<td>Speech</td>
</tr>
<tr>
<td>III</td>
<td>Music (music theory)</td>
<td>Musical notation</td>
<td>Music performance</td>
</tr>
</tbody>
</table>
Rhythmic grouping in ordinary prose reading may not seem to be very important for our perception. However, readers behave more or less rhythmically in their speech performances. Note, though, that this does not imply that it is a simple task to quantify the degree of rhythmicality. As listeners we reject a too rigid pulse train, hence the moderate success of the strict rule-books mentioned earlier.

A change of tempo and reading style affects vowel and consonant durations differently. Obviously, there is a need to pursue studies of such relations. For a recent contribution in this field, see [13].

Rhythmic entities and prose reading

Our perception of rhythm in speech and prose reading is related to the presence of marked events, which imposes a pattern of rhythmical groups. In so called "stressed timed" languages like English and Swedish, Pike [14], it is the alternation between syllables of lower and higher degree of stress which sets the rhythmical pattern. A rhythm group, or in other words an interstress interval or stress foot, is for analysis purposes usually defined as the interval from the onset of a stressed vowel to the onset of the next stressed vowel. Such a foot may span a pause or an otherwise marked syntactic boundary. If not, it is referred to as a "free foot". The main stress correlates are syllable durations and F0 contours but vowel reduction and vocal source intensity and waveshape is also of some importance.

In a language like French, by Pike referred to as "syllable timed", word stress is less apparent and rhythmical groups are generally considered to end instead of starting with a stressed syllable. Wenk and Winland [15] do not find it motivated to call French syllable timed. Instead, they refer to French as "trailier timed", in distinction to English which is "leader timed".

Isochrony does not exist in an absolute sense [16-20], but there is a clear trend in all languages that the foot duration increases with the number of syllables or phonemes contained. However, there is a physiological and perceptual reality of the mean stress interval in text reading, which sets the time base for control of the joint duration of pauses and prepause lengthening. In rhythmical reading the sum of pause and final lengthening appears to follow the local rhythm, as set by the mean free foot length within a memory span of the order of a sentence. There is a continuity of rhythm across the pause just as in music performance, where the final note before a pause may be lengthened as long as there is rhythmical continuity across the pause.

POETRY

Theoretic description.

Two different abstractions guide the reading of a poem. One is the linguistic frame, i.e. the natural rhythmical pattern of spoken language. The other is the metrical grid implied by a scanned reading of the poem. These two patterns may be in harmony or in conflict, which creates a tension that might add to the poetic value. In recitation, more or less attendance may be given to the metrical norm.

Not much work has been devoted to acoustic-phonetic analysis of read poetry. For a recent critical study, see Loots [21]. A specific problem, treated by Newton [22], is the difference between iambic and trochaic meter. We have confirmed his statement that the weak syllable of a trochaic foot is relatively longer than the weak syllable of a iambic foot. In our pilot experiment [23] on scanned reading, we found an 1.2:1 ratio of the strong and weak syllable in a trochaic foot, whereas in the iambic foot the weak to strong ratio was 1:2.2. This can probably be explained by the effect of final lengthening, prolonging the unstressed final syllable in a trochee and a stressed final syllable in a iamb.

MUSIC

Theory

To the layman it is difficult to make a division between the contents of a piece of music and its form. Mainly because the semantics of music can not be expressed in other terms than musical! However, tree structures of the kind used in linguistic analysis have been applied in music theory [24] and the concepts of phrases of different dignity have fitted into this type of formal description.

It is striking that in English (and Swedish) there are no independent terms for the form and the substance of music. As depicted in Table 1, the "language-speech" dichotomy thus corresponds to "music-music performance".

Bar notation

Are there any analogies between music and poetry in terms of the concepts of bar notation and foot analysis? Some proponents of modern metrical theory, [25], prefer to regard an initial weak syllable as a foot in poetry. There is a strong analogy between bars in music notation and stress locations in the analysis of prose.

Foot pattern in music is usually chosen to have an initial strong beat, followed by weak beats. This resembles the strategy chosen in prose, the "phonetic foot". The interpretation of note sequences and the notation of bars gives a choice for the music theorist. Basically, two ways are open: either one can ignore any formal metric scheme, i.e. the placement of the bars and completely rely on the melodic line. Or, one can regard the formal placement of notes within the bar pattern as a help to decode the tree structure, that is, the perceptual metrical scheme. In a sense, the bar format with its underlying rhythmical interpretation (in 4/4, for instance, the first and third beats are usually accented) has the same role as the metrical scheme in poetry, by imposing an accent pattern that can be more or less contrasting to the rhythm that is deduced from the melodic line, or in poetry from the prosodic patterns underlying the word string. Whether to include the bar format in an analysis or not is difficult to decide. Another complication is also, that the composer might more or less explicitly make use of the bars, by placing the accented notes in the melody line on, or close to the 1 or 3 in a 4/4 bar, and if not doing so (e.g. by syncopation), clearly indicate it in the musical flow. Modern music (and also renaissance music!) has a different attitude in this respect. Any rhythmic interpretation of the position of notes in a bar might instead create difficulties, and accordingly, the best strategy would be to neglect the bar lines altogether.

RHYTHMICAL CONSTRAINTS

How strict are the demands on rhythmical patterns in prose, in poetry and in music? The rhythmical demands in music are high for a number of reasons. In a collective musical performance, e.g., a common rhythm must be easy to follow. And, even if the rhythm changes with time, it should at least not be unpredictable. This means that we can talk about this phenomenon at two levels. At the micro level it is interesting to know whether a shortening of one segment leads to a prolongation of the following segment, and as quickly as possible conform with the word string. Whether to include the bar format in an analysis or not is difficult to decide. Another complication is also, that the composer might more or less explicitly make use of the bars, by placing the accented notes in the melody line on, or close to the 1 or 3 in a 4/4 bar, and if not doing so (e.g. by syncopation), clearly indicate it in the musical flow. Modern music (and also renaissance music!) has a different attitude in this respect. Any rhythmic interpretation of the position of notes in a bar might instead create difficulties, and accordingly, the best strategy would be to neglect the bar lines altogether.

EUROSPEECH '89, Paris, France, September 1989 2691
RHYTHMIC PULSE ACROSS PAUSES

The problems of continuity across lines are seldom treated in the analysis of read poetry. In music, the interval between musical phrases is coherent with the rhythm, and in the abstract notation it is possible to count the number of rhythmic pulses across pauses. In practice, larger deviations may occur at those places, due to phrasing. (As already mentioned, in rhythmical reading of prose we have found an apparent tendency of rhythmic continuity across pauses [20].) We intend to look closer at the rhythmic structures over a complete read poem, with the intention of also incorporating the pauses in the analysis.

ILLUSTRATIONS

Examples of notation (II) and signal representation (III) of prose, poetry and music, using an excerpt from a Swedish poem (song) by B. Sjöberg [27] will illustrate the contents of Table I.

A poem, rewritten as a prose passage with a phonetic transcription:

Den första gång jag såg dig, det var en sommardag på förmiddan, då solen lyste klar. Och ängens alla blommor av många hundra slag, de stod bugande i par vid par.

The poem with metrical analysis (iambic feet):

\[ \begin{align*}
\text{Det första gång jag såg dig, det var en sommardag på förmiddan, då solen lyste klar,} \\
& \text{och ängens alla blommor av många hundra slag, de stod bugande i par vid par.}
\end{align*} \]

Music notation of the song (one note for each syllable):

Duration analysis of spoken texts and played music.

Duration analyses of spoken text and music performance, as described above, was made in a pilot experiment with one of the authors as a subject (LN). (Examples of simulated versions of text and music will be presented at the conference.)

In the first text version the aim was to adhere to the scanned rhythm indicated in the musical version of the poem. The text was sung in a scanned manner and a tapping was afterwards synchronized to what was perceived as the rhythmic pattern. The duration between taps, corresponding to the beat, starting with a strong beat, is marked with \( * \) in Fig 1. As can be seen the pattern is very stable, approximately 500 ms intervals, and the speech has accordingly been forced into a rhythmic pattern.

Contrary to this pattern is the foot pattern of the prose-like reading (vowel onset to vowel onset of next stressed syllable), as indicated by "*", where the duration varies much more as a consequence of linguistic context (stress pattern and syntax), allowing for this type of variation.

The song was also played on a keyboard instrument according to the musical notation with a natural, not too exaggerated phrasing. The note-to-note durations are plotted in Fig 2, not as absolute duration values, but as difference values in relation to the nominal lengths of the notes. A nominal value of 350 ms for one eighth note (\( \text{\textup{Mr}} \)) was somewhat arbitrary chosen, with respect to which shortened and prolonged notes appear as negative and positive values. Notes within a bar are connected with a thin line and the notes appearing on beat 1 and 3 in a bar are marked with circles (o) and stars (*), respectively. A typical effect, seen here, is the final lengthening of notes appearing at the end of a phrase (notes number 13, 23, 36 and 46).

Acknowledgement: Our thanks to Rolf Carlson and Anders Friberg for the simulation of speech and music performances, respectively. This work has been financed by the Swedish Council for Research in the Humanities and Social Sciences.
REFERENCES


[17] E den Os, "Rhythm and tempo of Dutch and Italian", Amsterdam: 1988


EUROSPEECH '89, Paris, France, September 1989