3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

RHYTHM ANALYSIS OF SPEECH AND MUSIC SIGNALS

Andrej Miksic and Bogomir Horvat

University of Maribor, Faculty of Technical Sciences,
Maribor, Slovenia

ABSTRACT

The paper describes some of the rhythmical features of speech and music signals. Some parallels in duration analysis between music performance and spoken text are drawn. We focused on the rhythmical properties that could be easily applied to speech synthesis or to automatic segmentation of continuous, natural speech, as well as to segmentation of sampled music. In the analysis of rhythmical patterns of continuous speech, we discuss timing and stress at the syllable, the word and the sentence levels. We examine how stress, word endings and sentence endings effect the absolute syllable duration and the span of the deviations due to different speakers.

Keywords: *rhythm, rhythmical pattern, prosody, intonation, segmentation, speech database*

1. INTRODUCTION

Now that the computational power and memory limits are no more the restrictive elements in speech analysis equipment, there is a renewed interest in handling longer sequences of speech. Determining the pitch contours and the timing rules, that apply over groups of words are essential and very complex problems in speech analysis. Mechanised analysis of natural-language sentences in extremely difficult, because of the richness and complexity of natural-language grammars.

Readers actually behave more of less rhythmically. All languages display a tendency for stresses to occur at constant time intervals. This is helpful to the listeners, as they can direct their attention ahead and are thus able to pick up important notes.

Since rhythmical demands are much stricter in music than in speech, it is easier to make an acoustic analysis of a song and afterwards test similar features on speech.

2. RHYTHM IN MUSIC

Every harmonised motion like marching, dancing, breathing, or heart beating can be called rhythmical. Rhythm in music relates to a regular ordering of stronger and weaker elements. There is no specific component of the acoustic signal that can specify the rhythm. However, it is known that some characteristics of the physical wave such as duration, frequency, intensity, or silence greatly effect our perception of rhythm.

We do not have a rhythm centre in the brain, nor do we have specialised time perceptors. The experience of rhythm involves regularity and grouping of more strongly accented elements together with the weaker ones.

In music rhythm is usually easy to follow. During musical performance it is possible to count beats inside the measure (bar notation only contributes to a better survey of rhythmical flow). In this manner a music sequence, including pauses, can be divided into equal time units. It is well known of course, that certain deviations to the strict beat timing can occur (final lengthening at the end of the music phrase, prepause lengthening).

3. SPEECH RHYTHMS

Some sounds can be distinguished from others, for example, as louder, longer, or higher in pitch, and occur in time relationships to the sounds that are softer, shorter, or lower in pitch. Those time relations between the alternating sounds make up a stress pattern, which contributes to the perception of speech rhythm.

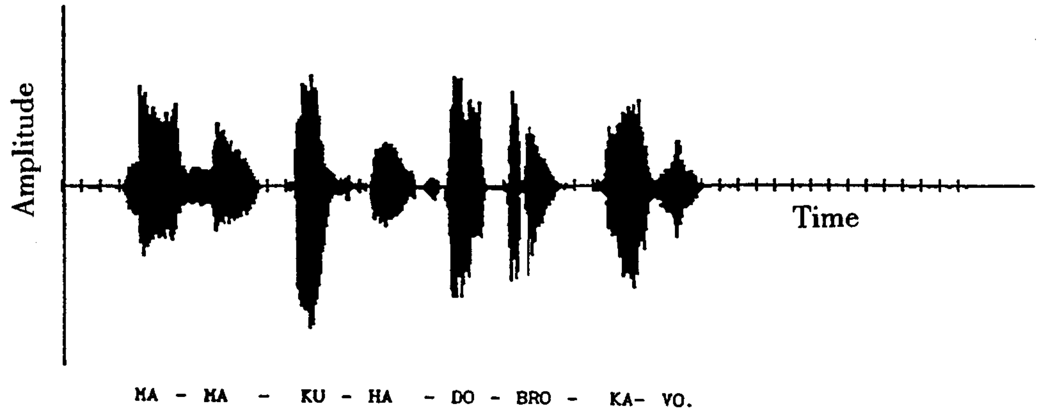

Figure 1. Alternating strong and weak syllables in a spoken Slovene sentence.

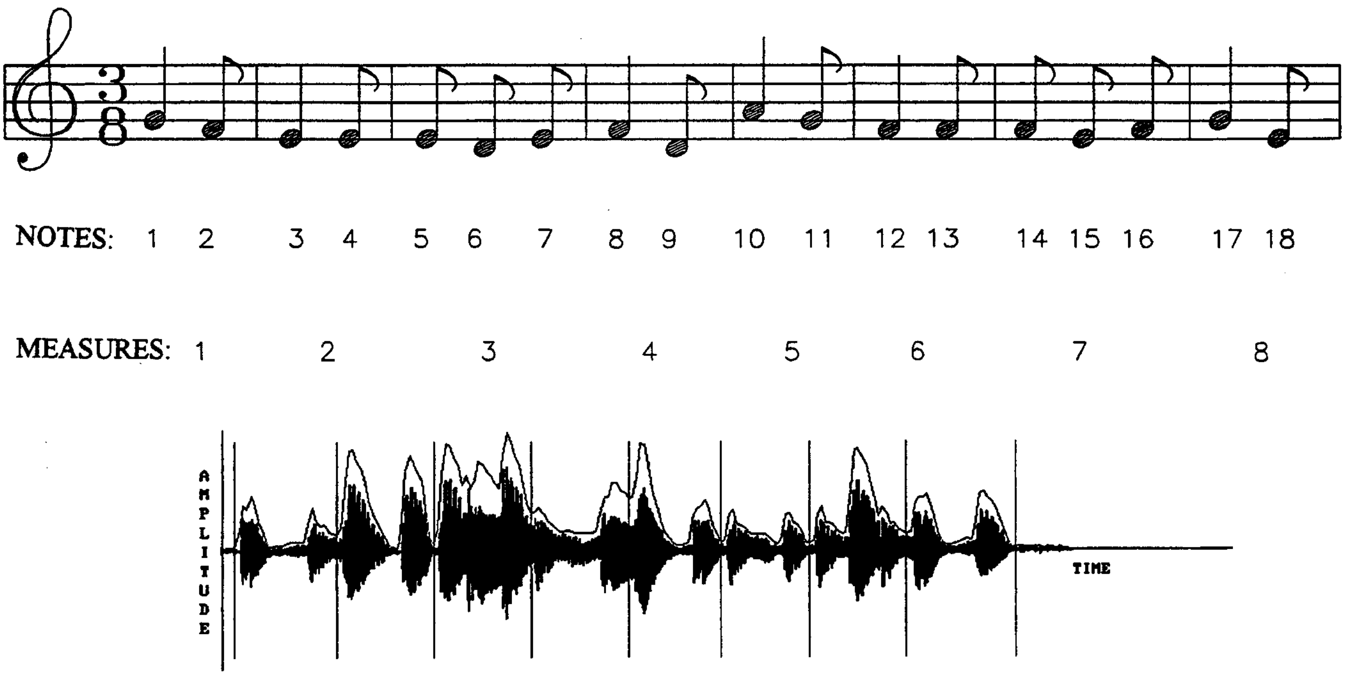

Figure 2a and 2b. Music notation of the song and waveform of the sampled performance

Like meters in music, speech rhythms refer to the perception of a regular ordering of stronger and weaker elements [2]. Stressed and unstressed syllables often alternate. Figure 1 shows the waveform of a Slovene sentence "Mama kuha dobro kavo" (Mom cooks a good coffee.). We can see how syllables with higher and lower energy regularly follow each other. This is the pattern typical for child speech. Adults rather use prolonging of the vowels and pitch shifts to emphasize important segments. In general, there is no automatic correspondence between the perceived stress and any acoustic measure.

There are limitations in producing speech rhythms. Speech involves the complex coordination of articulatory components (lips, tongue, glottis, jaw) that can not move independently.

Speech rhythms emerge from diverse rhythmic levels; there is a rhythm of syllables, words, phrases, the rhythm that reflects emotions, the rhythm of a regional dialect, and so on. In continuous, natural speech, there is an apparent tendency to rhythmic continuity even across pauses [3].

4. ILLUSTRATIONS OF RHYTHM ANALYSIS

As we mentioned before, our primary concern in rhythm analysis was to study timing rules and stress patterns, without concentrating on preference rules in music or on intonations in speech.

Figure 2a shows a music notation of the first few measures of a Slovene folk song (3 beats per measure). Folk songs tend to have clear rhythms. Music performances of these songs even reflect the (memorised) perception. Figure 2b shows the time waveform of the sampled piano performance of the same song. Vertical lines coincide with the bars in the music notation. The energy envelope of the signal is plotted above the signal.

Duration analysis of the piece leads us to the following conclusions.

1. Rhythmical pattern, which is indicated by the bar lines, in most cases starts with a stressed beat. (In stressed time languages, like English and Slovene, rhythmical groups tend to start with a stressed vowel.) At every fourth measure (starting with the third), which is set up of three quarter notes, we noticed the lengthening of the first note. This might be due to performers' subconscious following of the rhythmical patterning of the previous measures.

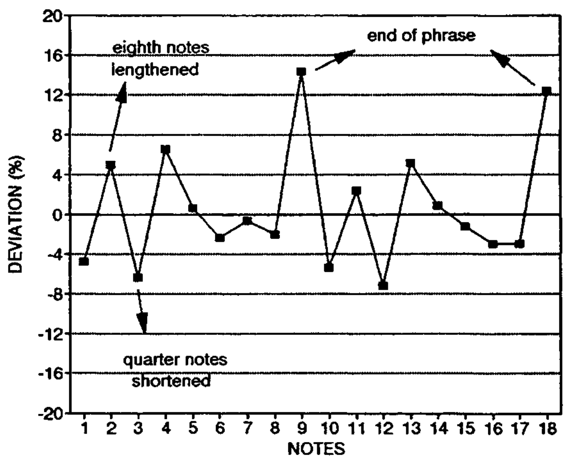

Figure 3. Duration deviations of each note from the nominal values.

2. The song from the figure 2 is composed of quarter notes (2 beats) and eight notes (1 beat). Figure 3 illustrates the duration deviations of each note from mechanical regularity. Durations of notes were measured from energy onset of the note to the energy onset of the following note. We see a typical shortening of quarter notes (from 2.0 beats to 1.90 beats) and prolonging of eighth notes (from 1.0 beats to 1.08 beats). Deviations from the nominal ratio (1.76:1.0 instead of 2.0:1.0) show how a performer highlights a melodic character of the song.

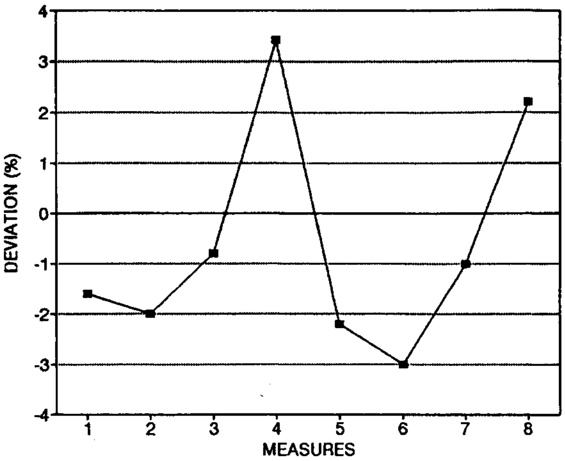

Figure 4. Duration deviations on the level of measures.

3. The sequence could be broken into two phrases of 4 measures. Figure 4 shows duration deviations on the level of whole measures. The middle measures of each phrase were played quicker and the final measures (measures 4 and 8) significantly slower. It is clear (figure 4) how the timing deviations at the phrase level dominate over the timing changes at the note level.

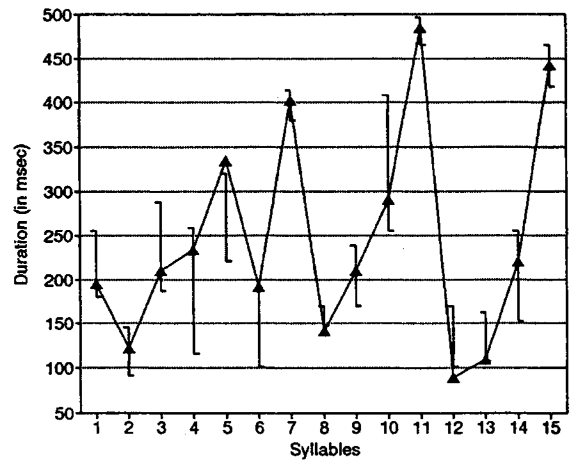

Figure 5. An example of timing analysis of continuos speech

In speech, duration plays a more important role than in music. At the word level, duration indicates stress. Stressed words are extended. At the sentence level, duration indicates sentence structure. The final syllable is often extended.

Figure 5 shows the timing analysis of the continuous speech sequence "Radio Slovenija. Prvi program. Porocila." (Radio Slovenia. First program. The news.). The comparison of the repeated pronouncing of the text (10 speakers as subjects) and the female radio announcer was carried out. The sequence of 15 syllables was set up. Vertical lines show the span of the deviations that are due to different speakers. Continuous line shows a time pattern of the voice of the radio announcer.

The stressed syllables are longer, but the prolongings of the last syllables in each word (syllables 3, 7, 9, 11 and 15) are overlaid on the stressed syllables (syllables 1, 5, 8, 11 and 14). Final lengthening that sets off each sentence (syllables 7, 11 and 15) is clearly noticed. A very interesting effect seen here is the lowest span of the deviations at those syllables (figure 5).

5. CONCLUSION

Some features derived from the rhythm analysis can be used in speech synthesis. However, in order to be able to generalize conclusions, the extensive analysis of speech data base (Slovene language speech database is still not available) should be done. For such database analysis present knowledge of speech prosody could be very helpful in creating the algorithm for the mechanical analysis of natural language.

REFERENCES

[1]: Bengsson, I., Gabrielsson, A.: *Analysis and synthesis of musical rhythm.* Studies of Music Performance - Publications by the Royal Swedish Academy of Music, No.39, 1983

[2]: Handel S.: *Listening.* MIT Press, Cambridge, 1989

[3]: Fant, G., Kruckenberg, A., Nord L.: *Stress patterns, pauses and timing in prose reading.* STL-QPSR 1/1989, pp.7-11, 1989

[4]: Prek S.: *Teorija glasbe.* Drzavna zalozba Slovenije, Ljubljana, 1989

[5]: Zwicker E., Fastl H.: *Psychoacoustics: Facts and models.* Springer Verlag, Berlin, 1990.

[6]: Allen G.D.: *Speech rhythm: its relation to performance universals and articulatory timing.* J Phonetics 3, p 75, 1975