Dept. for Speech, Music and Hearing

Quarterly Progress and Status Report

Stress patterns, pauses, and timing in prose reading

Fant, G. and Kruckenberg, A. and Nord, L.

journal: STL-QPSR
volume: 30
number: 1
year: 1989
pages: 007-012

STRESS PATTERNS, PAUSES, AND TIMING IN PROSE READING

Gunnar Fant, Anita Kruckenberg and Lennart Nord
Dept. of Speech Communication and Music Acoustics,KTH

Abstract

A decomposition of a read text into a succession of interstress intervals reveals basic structures of stress timing and stress distribution within a sentence and is the bases for analysis of quasi-rhythmical phenomena, speech tempo, and reading style. In a case study, we describe how stress foot statistics may be adopted for the analysis of various reading modes such as normal, low voice effort, high voice effort, and distinct reading.

INTRODUCTION

The long term objective of our project is to learn about what constitutes good reading style and to develop a perspective of the demands of a text and individual variations. We also intend to extend experience from prose reading to the study of poetry reading and relate our findings to the theory of musical composition and performance.

In the course of our work we have found it necessary to devote a major part of our time to methodological problems combined with small scale pilot studies. As described by Fant, Nord, & Kruckenberg (1987); Fant, Kruckenberg, & Nord (1988); Fant & Kruckenberg (1986a, 1986b, 1988a, 1988b), we have developed quantitative continuous scales of syllabic stress, with reference to durations and FO patterns and also to subjectively scaled measures. A substantial amount of work has been devoted to studies of stress groups, i.e., interstress intervals or "stress feet". We have shown that such intervals, defined from the onset of a stressed vowel to the onset of the next stressed vowel, see Fig. 1, provide a convenient basis for analysis of rhythmical qualities and also of stressed-unstressed contrasts which is one of many parameters that enter descriptions of reading style.

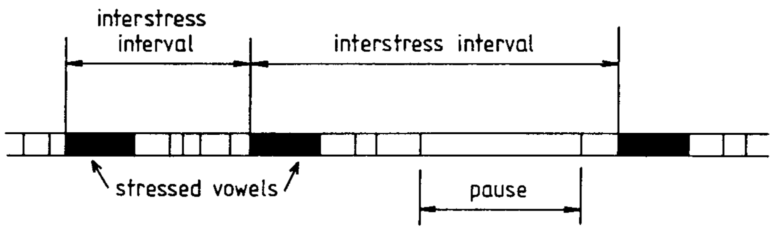

Fig. 1. Interstress intervals are bounded by the onsets of stressed vowels.

STRESS GROUP ANALYSIS

Interstress intervals are of the order of 500 ms, ranging from 0.25 sec to 1 sec with a standard deviation of about 35%. Thus, there is no absolute isochrony. However, according to Lehiste (1975), there still exists a trend of perceptual isochrony.

The duration of a stress foot increases systematically with the number of phonemes or syllables contained. From a linear regression analysis of free feet, i.e., those that do not span a pause or an otherwise marked boundary, see Fig. 2, we may derive the foot constants a and b, of which b is the increment of duration per additional phoneme and a is the sum of all stress induced segment lengthenings within the foot. Typical values are a=100 ms and b=59 ms per phoneme with an average foot duration of T_a=535 ms corresponding to 7.4 phonemes or 2.9 syllables.

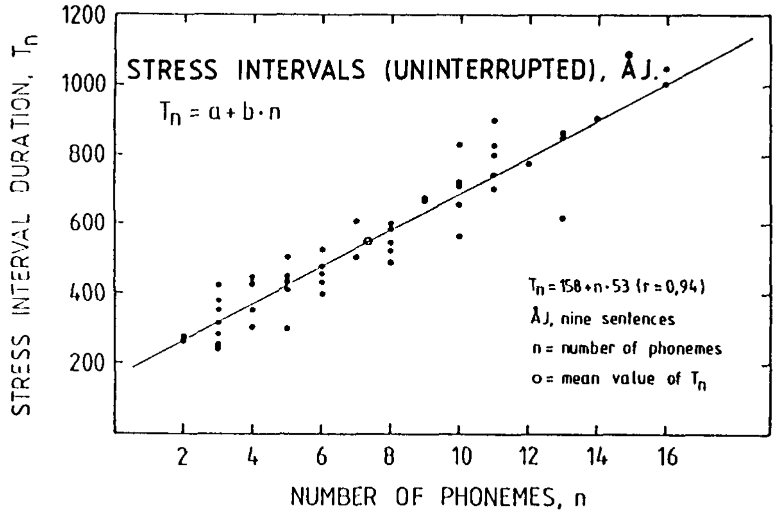

Fig. 2. *Duration of uninterrupted interstress intervals, "free feet", as a function of the number of phonemes contained.*

For our reference subject, we found a=158 ms and b=53 ms per phoneme and an average interstress interval of T_a=550 ms. Within a group of 15 subjects, we found a variation of a from 50 ms to 200 ms whilst b had a more restricted range of variation, from 50 to 65 ms per phoneme, which points at a general stability of the duration of unstressed phonemes and overall speech rate. The higher the a and the smaller the b, the greater is the contrast between stressed and unstressed syllables, which also conditions a somewhat lower relative spread of foot durations. An increase of a is associated with a greater overall increase of consonant durations in the stressed syllable than of the duration of the stressed vowel.

Differences in inherent phoneme durations tend to average out within a stress foot. Besides the number of phonemes contained, the main source of variation of the duration of a stress foot is the particular degree of stress. We have quantified stress by a normalized syllable duration index which has a mean value of 2 for stressed syllables and 1 for unstressed syllables. The standard deviation of the duration index of syllables carrying the stress within the feet was 0.4. The essential variance of stress feet durations was the same as that of stressed syllables. Within a sentence, positive departures from the mean were compensated by negative deviations, see Fig. 3, and to a far greater extent than could be predicted from standard deviations of prediction errors. This tendency conforms with a notion of stress planning within the constraints of a total energy expenditure assigned to a sentence, Öhman (1976), and can also be interpreted as a tendency to minimize deviations from an overall synchrony plan.

RHYTHM ACROSS PAUSES

An apparent aspect of stress timing is the realisation of boundaries containing a pause. In rhythmical reading a stress foot spanning a clause or a phrase boundary has a duration close to what could be predicted from the number of phonemes plus an additional average free foot unit T_a. Under these ideal circumstances the sum of pause duration and "terminal lengthening" equals T_a. The terminal lengthening is the difference between observed and predicted sound duration assuming a normalized stress level. The average free foot spacing may thus be conceived of as synchronizing an internalized pulse interval generator which ensures a continuation of speech rhythm across pauses.

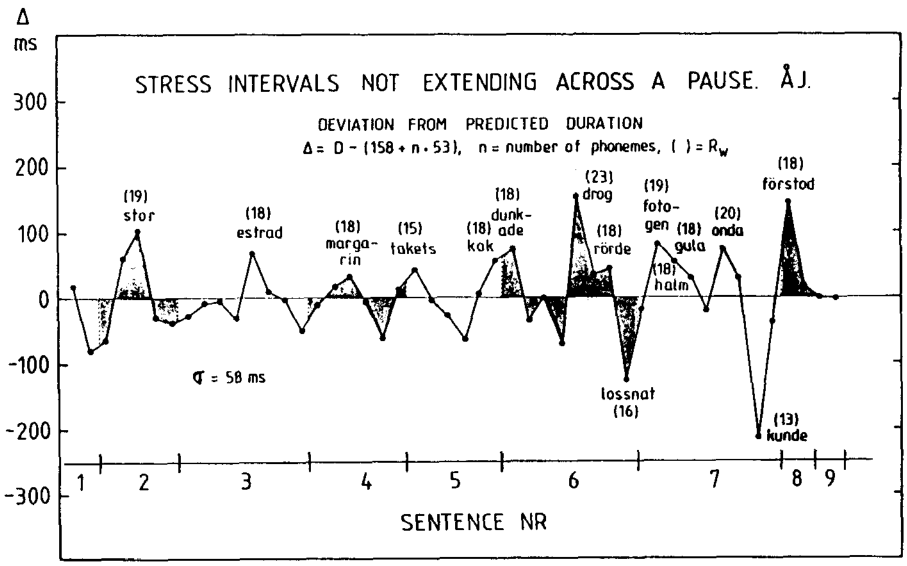

Fig. 3. *Deviation of free feet durations from values predicted from the number of phonemes only.*

If we adopt the paragraph average data of the constants a, b, and T_a, the departures from ideal boundary realization become of the order of 50-100 ms but are found to systematically vary from sentence to sentence, see Fig. 4. The errors are found to be positively correlated to the average interstress interval of the previous sentence. This tendency was also found for synchrony errors in transcending pauses between sentences. Such pauses, typically 1 second or 1.5 seconds plus associated terminal lengthening, have apparently absorbed one or two additional pulse interval quanta.

These observations suggest that the rhythm generator to some extent follows the past history of reading within a memory span of about 5-10 feet. To what extent it looks ahead is not clear.

SPEECH RATE AND STRESS RATE

There are two important aspects of tempo or speech rate to consider. A basic measure of speech rate is the number of syllables or phonemes per second. We shall here be concerned with phonemes per second or rather the inverse, the average duration of phonemes within a part of the read text. This measure, of the order of 80 ms, will be dependent on the relative density of stressed syllables which may vary considerably from one sentence to the next. With a fewer number of phonemes per foot, there will be a predominance of the stressed part and thus a greater than average phoneme duration.

Another measure is that of average phoneme duration within a free foot, which because of the absence of terminal effects is about 5% smaller than the overall phoneme duration, see Fig. 4.

The other concept is that of the stress rate, i.e., the inverse of the average duration of a nonspanning foot. The stress rate which is text dependent has an inverse relation to the speech rate, as we have seen. However, the foot constants a and b need not change with varying stress rate and thus constitute a more stable indicator of the reader's performance which may be tied down to a long time average duration of phonemes per free foot.

ANALOGY TO MUSIC

As outlined by our theory a local average of stress rate synchronizes an internal pulse rate which is the reference for rhythmical realizations of pause spanning feet, securing a total duration of pause and terminal lengthening to equal T_a or $2T_a$ or $3T_a$. The analogy to music is obvious. A rhythmical continuity across a musical pause involves similar constraints on pause duration and final lengthening. The interval between two bars, the "measure" often contains two beats. In a common march tempo the basic beat is 500 ms, which is of the same order as we have found in rhythmical reading. However, in rhythmical reading synchrony is more apparent in pauses than in the speech part.

A CASE STUDY

We shall now follow up this general presentation with some results from a case study. We have analyzed five versions of our reference subject ÅJ's reading of sentence 7 in our standard text: "Allt detta, fotogenlyktornas gula ljus, halmen som prasslade när de vände sig och de onda lukterna i trängseln, kunde hon kanske bilda sig ett begrepp om." It contains four noun phrases, inviting to a rhythmical disposition with four major pauses followed by a predicate phrase read at a considerably higher speech rate. Our subject read the entire paragraph from which this sentence was collected in four different modes: normal, low voice effort, distinct reading, high voice effort. The normal reading was repeated once and then came out somewhat "forced", less rhythmical in the pausing.

For the low voice effort, we noted pauses of even durations, 200 ms longer than anticipated, which is a remarkable deviation from our pause spanning rule. The pauses were homogeneous but no longer rhythmically related to the tempo. As can be seen in Fig. 5 the normal, the forced and the low voice effort versions were read at about the same speech rate of 75 ms average free foot duration (excluding the predicate final phrase). The a and b constants were about the same for normal and low voice effort, of the order of b=53 ms per phoneme and a=185 ms, whereas the high voice effort and distinct reading constituted another group averaging b=59 ms per phoneme and a=260 ms at an average free foot duration of 94 ms. The increased a/b ratio reflects an increased stressed/unstressed contrast. The pausing in these two readings were uneven, but averaged the order of magnitude expected from the rhythmicality criteria. The "forced" version of reading has an interesting relation to normal reading. The speech rate is the same which requires that $b+a/n_a$ is the same. Here n_a = 7.8 is the average number of phonemes per foot within the ensemble of foot data (which excluded the predicate phrase). Compared to the normal reading the lower a/b ratio of the "forced" version indicates a lower stressed/unstressed contrast.

Another observation is that in the high voice and the distinct mode and also in the forced mode the ratio of speech rates in the final and main part of the sentence was increased, of the order of 1.5 compared to 1.3 in the normal and low voice effort reading. Such tempo changes reflecting alternating semantic emphasis and economy of overall energy expenditure are typical features of an engaged reading.

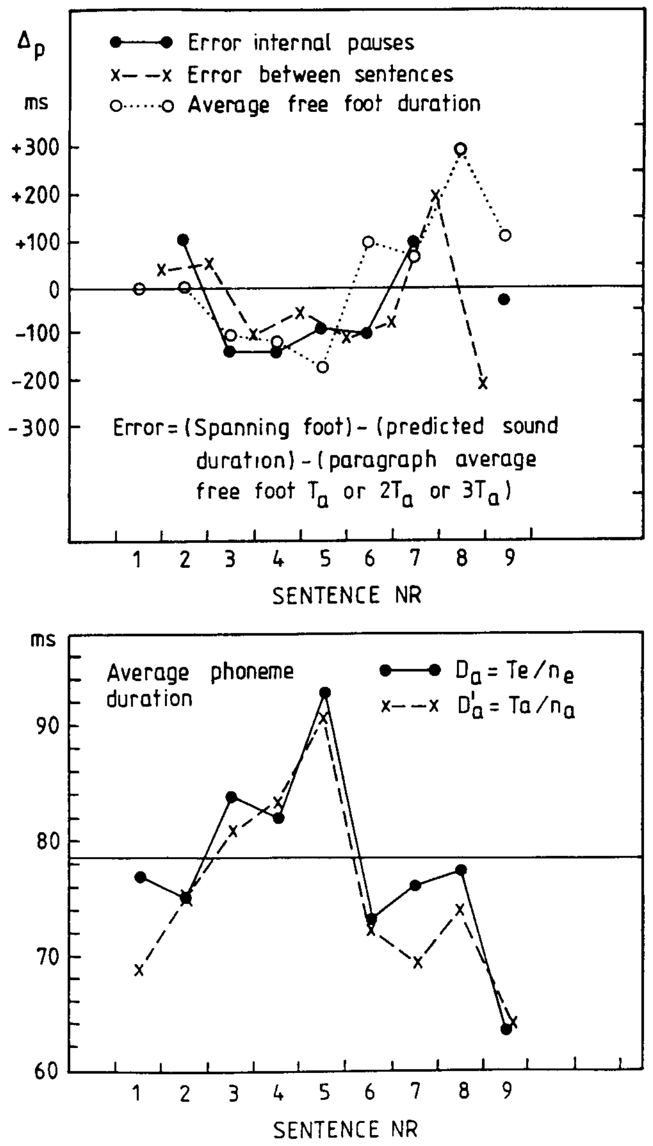

Fig. 4. *Above: covariation of pause spanning errors and sentence average free foot durations. Below: variation of sentence average values of phoneme durations, Da, and free foot average phoneme duration D'a.*

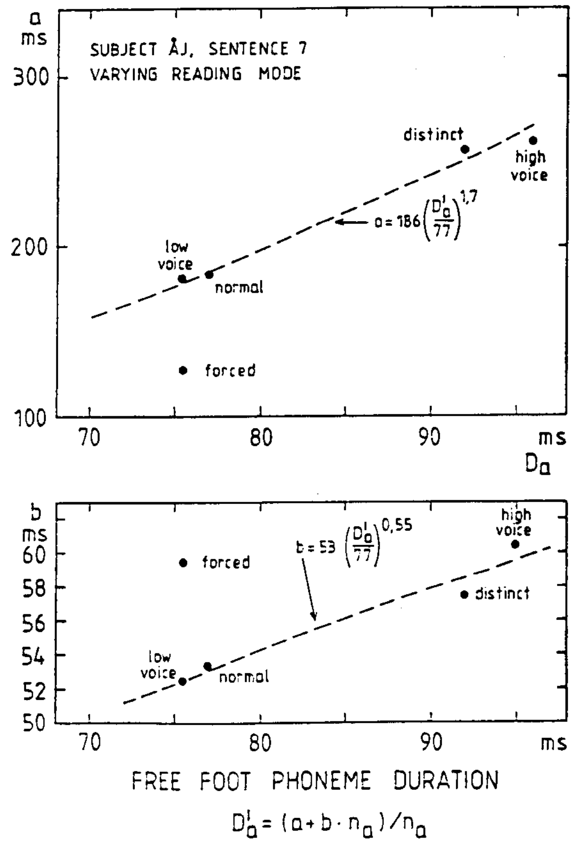

Fig. 5. *The a and b free foot parameters as a function of a corresponding average free foot phoneme duration. Subject ÅJ, five reading modes.*

CONCLUSIONS

We may conclude that perfect synchrony in pause realisations is found in relaxed rhythmical reading only. In rhythmical reading stress timing preserves synchrony across pauses in terms of one or more extra rhythm units added to the boundary. These units are the time intervals of an internalized clock which follows a relative short time average of interstress intervals already executed and perhaps also those that are just about to be executed.

It is important to keep apart the notions of speech rate and stress rate, the latter being conditioned by the text and exerting an influence of the criteria for rhythmicality in pause planning. Pause and final lengthening are similarly structured in music and in speech. The basic rhythmical unit is the same. It remains to gain experience about these phenomena in poetry reading. A more complete account of these and related studies will be given in a forthcoming issue of the STL-QPSR.

Acknowledgments

This work has been financed by the Bank of Sweden Tercentenary foundation and the Swedish Board for Technical Development.

References

Fant, G. & Kruckenberg, A. (1986a): "Projekt god svenska - databas", Dept. of Speech Communication and Music Acoustics, KTH, Stockholm, Internal report.

Fant, G. & Kruckenberg, A. (1988a): "Contributions to temporal analysis of read Swedish," pp. 37-41 in *Working Papers 34* (Dept. of Linguistics, University of Lund), Papers from the Second Phonetics Conference, Lund, May.

Fant, G., Kruckenberg, A. (1986b): "Individual variations in text reading. A data-bank pilot study", *STL-QPSR 4/1986*, pp. 1-17; also pp. 104-114 in (O. Engstrand, ed.) *Papers from the Swedish Phonetics Conference, Uppsala, Oct. 1986*, Reports from Uppsala University, Dept of Linguistics, No. 17, 1987.

Fant, G. & Kruckenberg, A. (1988b): "Some durational correlates of Swedish prosody," pp. 465-502 in *Proc. SPEECH '88, Book 2* (7th FASE-Symposium), Institute of Acoustics, Edinburgh.

Fant, G., Kruckenberg, A., & Nord, L. (1988): "Data-bank analysis of speech prosody", paper presented at the 2nd Symp on Advanced Man-Machine Interface Through Spoken Language, Hawaii, November.

Fant, G., Nord, L., & Kruckenberg, A. (1987): "Segmental and prosodic variabilities in connected speech. An applied data-bank study," pp. 102-105 in *Proc. from XIth Int.Congr. of Phonetic Sciences, (ICPhS), Tallinn, USSR, Vol.6*, Estonian Academy of Sciences, Tallinn.

Lea, W.A. (1980): *Trends in Speech Recognition*, Prentice Hall Int.

Lehiste, I. (1975): "Isochrony reconsidered," J. Phonetics 5, pp. 253-263, 1975.

Strangert, E. (1985): *Swedish Speech Rhythm in a Cross-Language Perspective*, Almqvist & Wiksell International, Stockholm

Öhman, S. (1967): "Word and sentence intonation: A quantitative model," *STL-QPSR 2-3/1967*, pp. 20-54.