ON THE QUANTAL NATURE OF SPEECH TIMING

Gunnar Fant and Anita Kruckenberg

Dept. of Speech, Music and Hearing, KTH,
Box 70014, Stockholm 10044.
email:Gunnar@speech.kth.se

ABSTRACT

This is a review of regularities we have observed in the analysis of text reading, mostly Swedish, directed to the timing of vowels and consonants, syllables, interstress intervals and pauses. We have found tendencies of quantal aspects of temporal structure, superimposed on more gradual variations, which add quasi-rhythmical elements to speech. A local average of interstress intervals of the order of 0.5 sec appears to function as a reference quantum for the planning of pause durations. A recent study, confirming our previous findings of multiple peaks with about 0.5 sec spacing in histograms of pause durations, provides support to this model. It is well established that pause durations tend to increase with increasing syntactic level of boundaries. However, these variations tend to be quantally scaled even within a specific boundary category, e.g. between sentences or between paragraphs. Relatively short pauses, as between phrases or clauses, show durations in complementary relation to terminal lengthening.

There are indications of approximately 1, 1/2, 1/4, 1/8 ratios of average durations of interstress intervals, stressed syllables, unstressed syllables and phoneme segments which adds to the observed regularities.

The timing of syllables and phonetic segments with due regard to relative distinctiveness and reading speed will be discussed and also tempovariations within a sentence.

1. SEGMENTS AND SYLLABLES

The main source of data to be discussed here derives from our own studies [1-7]. The text was a passage of about 8 minutes´ duration from a Swedish novel read by our reference subject, a Swedish language expert. A databank search system organized within a linguistic frame was developed for the processing. Our analysis has been concerned with individual vowels and consonants, syllables, interstress intervals and pauses. In addition we have data from 15 other subjects reading a limited part of the text. Durations were measured by hand from broad band spectrograms.

The concept of quantally structured durational data is not new. Gårding [8], in a study of contrastive prosody, proposed a timing model for read Swedish in which the duration of an unstressed CV-syllable is the unit. Syllables with either a long vowel or a long consonant, i.e. stressed syllables were given two such units and phrase final lengthening one extra unit.

Our accumulated experience from speech analysis, including a recent unpublished databank survey, allows a more extensive modelling. We find a clear tendency of factor 2 relations between major categories. Interstress intervals, measured from the onset of the vowel in a stressed syllable to the onset of a vowel in the next stressed syllable, excluding those spanning a pause or a syntactic boundary, averaged 540 ms. The average duration of primary stressed syllables as well as those of secondary stress in compound words was 270 ms. Unstressed syllables averaged 132 ms. Mean phoneme duration was 70 ms. Unstressed vowels averaged 59 ms and unstressed consonants 51 ms. There are thus approximately 1, 1/2, 1/4, 1/8 relations in the timing of interstress intervals, stressed syllables, unstressed syllables and phonemes

The data above refer to contexts excluding prepause locations. Within this regular frame there exists a continuity of variations of segment durations and positional variants but one still finds regularity traits. Thus consonants after short stressed vowels are of about twice the length of unstressed consonants which holds for voiced as well as for unvoiced consonants, according to [1] a ratio of 87/44 for voiced and 135/67 for unvoiced consonants.

A basic distinction in Swedish phonology is that of "quantity". A stressed syllable contains either a long or a short vowel. A stressed short vowel is followed by a long consonant or vice versa. The relation of the duration of a long stressed vowel to a short stressed vowel is not 2 to 1 but of the order of 1.6 to 1. Lexically stressed vowels in function words generally lose their stress in connected speech. As a mean trend over all contexts and tempos and several data corpora we derived in [1] a relation between long and short stressed vowels of

$$V_{long} = 1.9V_{short} - 45 \text{ ms} \qquad (1)$$

The durational distinction is lost when V_{short} approaches 50 ms. A fully stressed VC: is about 10 % shorter than a V:C and of the order of 210 ms.

The average number of phonemes per syllable is close to 2.9 for stressed and 2.2 for unstressed syllables, but text specific variations occur. In our standard prose passage we noted 3.0 phonemes per stressed syllable. Alternatively with a non-conventional definition of syllables, constrained by root morphemic criteria such that the word "legat" would be segmented as [leg-at] opposed to the conventional [le-gat], the average number of phonemes per stressed syllable in the corpus

morphemic definition would be that the duration of the consonant following the stressed vowel is lengthened. In our statistics, retaining the conventional definition of syllables, we have accordingly introduced a special category for initial consonants of unstressed syllables that are preceded by a stressed vowel in an open syllable. The mean duration of such syllables is 192 ms and the average number of phonemes is 2.55, i.e. substantially greater than for the main category of unstressed syllables.

2. PAUSES. RHYTHMICAL CONTINUITY

Lea [9] introduced the concept of rhythmical continuity of stress intervals (feet) spanning a pause, stating that mean values of such intervals equaled an integer of the average duration of interstress intervals that are not interrupted by a syntactic boundary. From readings of the Rainbow Passage he noted quantal steps of 0.5 seconds. Pauses before clauses averaged 0.5 seconds and before a new sentence 1 second, which implies pause spanning feet of 1 and 1.5 seconds.

This model was further developed by us [1]. We noted a complementary relation between the duration of pauses of the first quantal order within sentences and prepause lengthening.

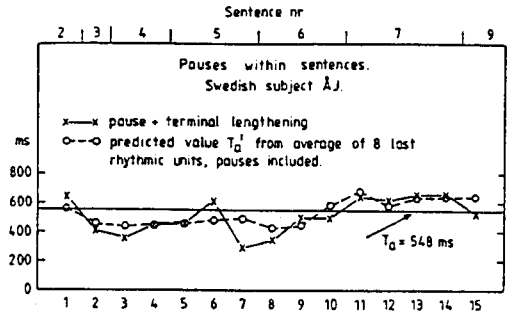

Figure 1. Inter-sentence pause durations plus final lengthening compared to local, 8 feet average interstress intervals.

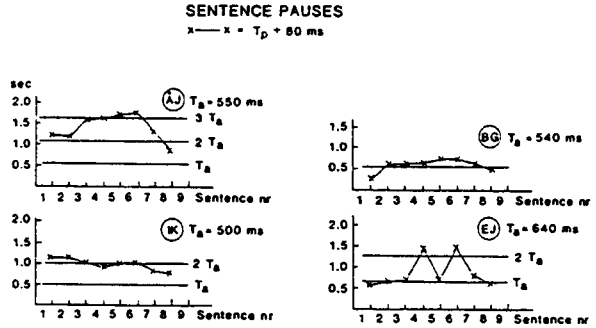

Figure 2. Examples of four Swedish subject's inter-sentence pauses showing quantal tendencies.

As shown in Fig.1 from [4] the sum of these two components tend to match the average free-foot duration derived from a short time memory span of about 8 free feet, or 4 seconds. We have observed quantal distributions of pause durations within one and the same boundary category, for sentence as well as for paragraph endings. These findings are speaker specific, some producing more rhythmically coherent patterns than others, and some favouring a larger number of quanta than others. An example from [4] is shown in Fig. 2.

Here the duration of the pause plus a standard value of prepause lenghtening equals an integer of the subjects average interstress interval. We have also exemplified such trends for English as well as for French, [4]

Examples of multimode pause durations are illustrated in Fig 3. pertaining to our reference speaker and in Fig 4 to some recently collected data for a female subject introducing a Linguaphone course. Distances between peaks are of the order of 0.5 seconds. Observe how this trend is apparent in terms of pause durations of 2 and 3 quanta and at paragraph boundaries 3, 4 and 5 quanta.

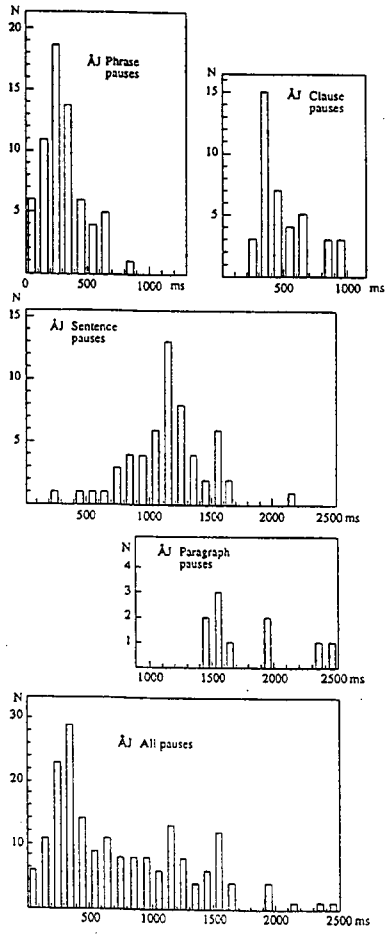

Figure 3. Phrase, clause, sentence, paragraph and all pauses in an 8 minute long text reading, subj ÅJ.

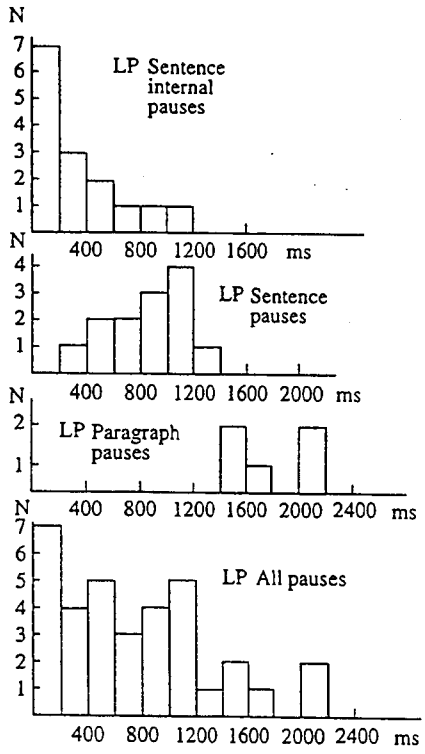

Figure 4. Sentence-internal, sentence, paragraph and all pauses in a 3 minute introductory reading for a Linguaphone course.

Strangert [10] found pause durations systematically increasing with syntactic level comparing phrase, clause, sentence and paragraph boundaries. These are comparable to our data but the possibility of quantal effects are hidden in the averaging process. An analysis of histograms supplied by Heldner and Strangert [11] also shows some trends of bimodal distributions but they are less apparent than in our reference data.

An additional support for the quantal nature of pausing derives from data on breathing, [13] (see [1] page 36). In this study sentence pauses gave a histogram peak at 500 ms without breathing and 1000 ms with breathing. Clause boundaries peaked at 500 ms with breathing and at 300 ms without breathing.

3. INTERSTRESS INTERVALS. FINAL LENGTHENING

It is by now well established that the duration of an interstress interval (foot) T_n increases with the number of phonemes, n, or with the number of syllables, m, in the foot.

For a passage read by our reference subject, excluding boundary spanning feet, we noted

$$T_n = 158 + 53n \qquad (2)$$

where n is the number of phonemes in the foot. Alternatively, in terms of the number of syllables, m

$$T_m = 190 + 120m \qquad (3)$$

The average foot length was T_n =548 ms corresponding to n=7.5 phonemes or m=3 syllables

A sequence of stress beats is thus quasi-rhythmical only [14], and the standard deviation is of the order of 40 %. However, the prosodic importance is considerable for stresstimed languages such as Swedish.

From our databank study of 8 minutes of prose reading we recently found average values of final lengthening of unstressed syllables ranging from 95 ms for pauseless juncture, 95 ms for sentence internal boundaries, 70 ms for sentence boundaries and 35 ms for paragraph boundaries. Stressed final syllable lengthening was of the order of twice those of the unstressed values above. These data are comparable to those in [1] where we reported a mean value of final lengthening of T_f=110 ms for within-sentence pauses of the order of 300-500 ms and more specifically

$$T_f = 190\text{-}0.2\ T_p \qquad (4)$$
(r=0.4)

Clause or sentence initial shortening was found to be of less magnitude and of the order of 40 ms for stressed and 15 ms for unstressed syllables.

Ideally, our model of rhythmical continuity across a pause implies that the sum of pause duration and final lengthening (initial shortening or lengthening included) of segments within a pause spanning foot equals an integer of the average foot length. The remaining part of the foot containing segments before and after the pause (excluding preboundary lengthening and postboundary shortening effects) has a duration which is determined by the number of phonemes according to Eq. 2 and which averages a free foot quantum.

However, as confirmed in [12] the complementary relation of final lengthening and pause duration appears to hold for relatively short pauses only. The decrease of final lengthening before longer pauses should also be considered. More extensive data are needed for a refinement of these durational models.

4. DISTINCTIVENESS AND TEMPO.

The durational contrast between stressed and unstressed syllables is a major correlate to distinctiveness [2,5,7]. A part of the contrast derives from the larger average number of phonemes per syllable in stressed than in unstressed syllables. However, in Swedish, the major part of the order of 100 ms, lies in the lengthening of vowels and consonants in stressed positions.

For our reference subject we noted an increase of the unstressed syllable duration D_u with the number of phonemes n in the syllable by a linear regression [5,7]

$D_u = 9 + 51n$ (5)

Correspondingly for stressed syllables

$D_s = 62 + 72n$ (6)

In a more distinct reading mode, stressed syllables increased relatively more than unstressed syllables which remained rather stable. The relative constancy of unstressed syllable durations also holds true of individual variations.

The stressed/unstressed contrast is speaker and language dependent, smaller in French than in Swedish [7] and is largely carried by stressed syllables. The statistics for unstressed syllables were rather similar comparing both speakers and languages. The same trend is also maintained comparing lower tempo and normal tempo speech. However, there is a reversal in fast speech where the unstressed syllables are relatively more reduced than stressed syllables [3].

4.1. Tempo variations

The local tempo in terms of average segment durations within a sentence or a phrase is considerably influenced by the density of content words and thus of potential stresses within the text. In addition there exist deviations from normally predicted durations that reflect reductions and expansions around and within focal regions. Such deviations tend to cancel within a sentence, [3,5] and reflect a finite pulmonary and articulatory energy at disposal [15]. In addition, alternating slowing down and speeding up of the tempo within a paragraph adds to the naturalness of reading.

5. CONCLUDING REMARKS

We have demonstrated two regularity aspects of speech timing.
(1) Quantal steps of the order of 500 ms in pause durations related to the average duration of interstress intervals.
(2) Average durations of stressed syllables, unstressed syllables and phoneme segments, of the order of 250 ms, 125 ms and 62.5 ms. which suggests 1/2, 1/4 and 1/8 ratios of the basic 500 ms quantum.
(3) The trend of rhythmical continuity across pauses is speaker specific in manifestation.
(4) The choice of quantal level is influenced by syntactic criteria and individual habits.

Much more work could be devoted to problems of statistical significance and influence of tempo and reading style and specific language dependencies.

6. ACKNOWLEDGEMENTS

This work has been financed by grants from the Bank of Sweden Tercentenary Foundation, the Swedish Council for research in the Humanities and Social Sciences and the Carl Trygger Foundation.

7. REFERENCES

1. Fant, G. and Kruckenberg, A. "Preliminaries to the study of Swedish prose reading and reading style", *STL-QPSR* 2/1989, 1-83.
2. Fant, G., Kruckenberg, A. & Nord, L. "Prosodic and segmental speaker variations", *Speech Communication* 10, 521-531, 1991.
3. Fant, G., Kruckenberg, A. and Nord, L. "Some observations on tempo and speaking style in Swedish text reading". *ESCA Workshop on "The phonetics and phonology of speaking styles", Barcelona*, 1991.
4. Fant, G., Kruckenberg, A. and Nord, L. "Stress patterns and rhythm in the reading of prose and poetry with analogies to music performance". In: *Music, Language, Speech, and Brain*, Wenner-Gren International Symposium (J. Sundberg, L. Nord, R. Carlson, eds.), Series Vol. 59, pp. 380-407, 1991.
5. Fant, G., Kruckenberg, A. and Nord, L. "Prediction of syllable duration, speech rate and tempo", *Proc. ICSLP 92, Banff*, Vol 1, 667-670, 1992.
6. Kruckenberg, A. and Fant, G. "Iambic versus trochaic patterns in poetry reading", *Nordic Prosody VI*, Stockholm,. 123-135**,** 1993.
7. Kruckenberg, A. and Fant, G. "Notes on syllable duration in French and Swedish", *Proc. XIIIth ICPhS* , 158-161, 1995.
8. Gårding, E., "Contrastive prosody: a model and its application." AILA Congr. 181. *Studia ling*. 35 146-166, 1981.
9. Lea, W.A. *Trends in Speech Recognition*, Prentice Hall, Inc.,1980.
10. Strangert, "Pausing in texts read aloud"*Proc.XIIth ICPhS* Vol. 4, 238-241, 1991.
11. Heldner, M. & Strangert, E. Personal communication of data. 1996.
12. Horne, M, Strangert, E and Heldner, M. "Prosodic boundary strength in Swedish: final lengthening and silent interval duration.", *Proc. XIIIth ICPhS*, Vol. 1, 170-173. 1995.
13. Base, A "Pauser i tal", *EE thesis work, KTH, dept. of speech, music and hearing***,** 1983.
14. Lehiste, I. "Isochrony reconsidered", *J. Phonetics* 5, 253-263, 1977.
15. Öhman, S., "Word and sentence intonation: a quantitative model*," STL-QPSR* 2-3/1967, 20-54.