Quantal theory of speech timing

Fant, G. and Kruckenberg, A.

journal: TMH-QPSR
volume: 37
number: 2
year: 1996
pages: 041-044

Quantal theory of speech timing

Gunnar Fant and Anita Kruckenberg

Abstract

This is a summary view of temporal patters we have observed in the analysis of Swedish text reading. Special attention is devoted to trends of quantal structures in the timing of vowels and consonants, syllables, interstress intervals and pauses. It is well known that pause durations increase with increasing syntactic order of boundaries. A recent study supports our previous findings of multiple peaks in the histograms of pause durations and our theory of neural coordination of pause durations and prepause final lengthening. These add to an integer of a basic time-constant of about 0.5 sec which reflects a local average of interstress duration and preserves a quasi-rhythmical continuity of interstress intervals spanning a pause.

Regularities in the timing of syllables and phonetic segments with due regard to relative distinctiveness and reading speed will be discussed and, on a higher level, tempovariations within a sentence.

Segments and syllables

The main source of data to be discussed here derives from studies by Fant & Kruckenberg, (1989) and Fant, Kruckenberg & Nord (1991A,B). The text was a passage from a Swedish novel of about 10 minutes duration read by our reference subject, a Swedish language expert. A databank search system organized within a linguisticframe 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 (1981), in a study of contrastive prosody, proposed a model for read Swedish in which the duration of an unstressed CV syllable is the unit. Two such units are allotted syllables with either a long vowel or a long consonant, i.e. stressed syllables. Phrase final lengthening is allotted one extra unit.

Our modelling supports these simple rules but is more extensive and shows a clear tendency of octave 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 is thus approximately a 1:2:4:8 relation in the timing of phonemes, unstressed syllables, stressed syllables and interstress intervals.

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 about twice the length of unstressed consonants which holds for voiced as well as unvoiced consonants, in Fant & Kruckenberg (1989, page 81), 44 and 87 ms for voiced and 67 and 135 for unvoiced consonants.

A basic distinction in Swedish phonology is that of "vowel quantity". A stressed syllable contains either a long or a short vowel. Their relation 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 have found a relation between long and short stressed vowels as follows

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

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 passage we noted 3.0 phonemes per stressed syllable. With a non-conventional definition of syllables to be constrained by root morphemic criteria, e.g. "leg-at" versus the conventional "le-gat", the average number of phonemes per stressed syllable will increase to 3.3. One argument in favour of the morphemic definition is that the duration of the consonant following the stressed vowel is prolonged. In our statistics, retaining the conventional definition of syllables, we have accordingly noted 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.

Pauses and rhythmical continuity

Lea (1980) introduced the concept of rhythmical continuity of stress intervals (feet) spanning a pause, stating that mean values of such intervals equalled 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 Fant & Kruckenberg (1989) who noted a complementary relation between prepause lengthening and pause duration The sum of these two components tended to match the free-foot duration or an integer of this quantum. Furthermore, there is evidence that the quantal reference derives from a short time memory span of about 8 free feet, or 4 seconds.

We also observed bimodal distributions of pause durations within one and the same boundary category, for sentence as well as 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, Fant & Kruckenberg (1991B). We have also exemplified such trends for English as well as for French.

Examples of multimode pause durations are illustrated in Fig 1. pertaining to our reference speaker and in Fig 2 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.

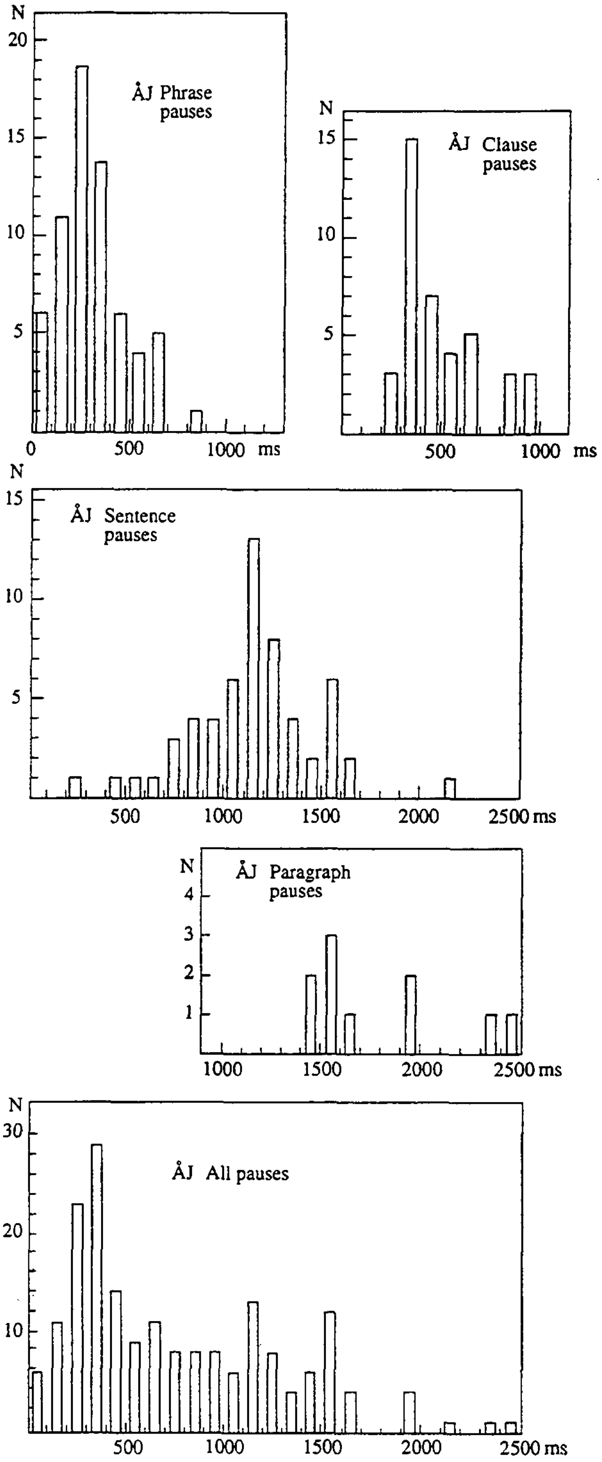

Fig. 1. Phrase, clause, sentence, paragraph and all pauses in 7 minutes text reading, subj ÅJ.

Strangert (1991) 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. A closer analysis of histograms supplied by Heldner and Strangert (1996) shows some trends of bimodal.

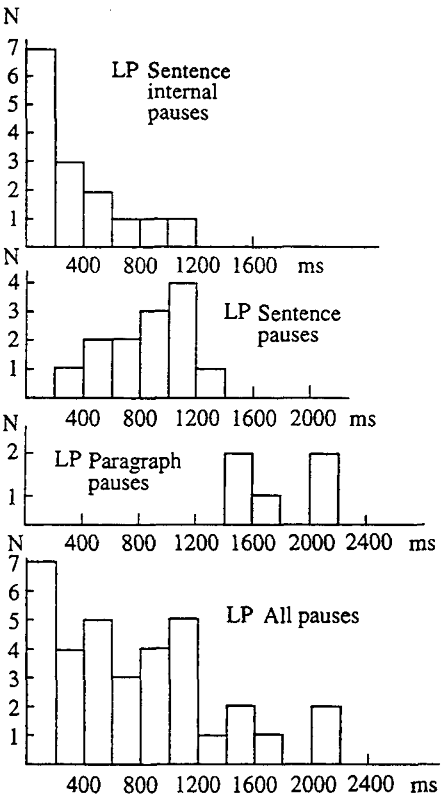

Fig. 2. Sentence internal, sentence, paragraph and all pauses in a 3 minute introductory reading for a Linguaphone course.

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, Base (1983), see Fant & Kruckenberg (1989, page 36). She reported on sentence pauses with histograms peaks at 500 ms without breathing and 1000 ms with breathing. Clause boundaries peaked at 500 ms with breathing and at 300 ms without breathing.

Interstress intervals and final lengthening

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

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

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

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

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

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

A sequence of stress beats is thus quasi-rhythmical only and the standard deviation is of the order of 30 %. However, the prosodic importance is considerable for stress-timed languages such as Swedish.

From our databank study of 7 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 of Fant & Kruckenberg (1989) who reported a mean value of final lengthening of Tf=110 ms for within-sentence pauses of the order of 300-500 ms and more specifically Tf= 190-0.2 Tp, (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 duration 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 averages a free foot quantum.

However, as confirmed by Horne et al (1995) 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.

Distinctiveness and tempo

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

For our reference subject Kruckenberg and Fant (1995) noted an increase of unstressed syllable durations with the number of phonemes by a linear regression

$$Ds= 9 +51n \qquad (4)$$

and for stressed syllables

$$Ds = 62 +72n \qquad (5)$$

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, Kruckenberg & Fant (1995) 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, Fant et al (1991A).

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, Fant & Kruckenberg (1989), Fant, Kruckenberg & Nord (1982) and reflect a finite pulmonary and articulatory energy at disposal, Öhman (1967). In addition, an alternating slowing down and speeding up of the tempo within a paragraph adds to the naturalness of reading.

Concluding remarks

We have demonstrated two regularity aspects of speech timing, (1) quantal steps of about 500 ms in pause durations related to the average duration of interstress intervals, (2) factor 2 relations of the durations of stressed syllables, unstressed syllables and phoneme segments, of the order of 250 ms, 125 ms and 62.5 ms.

The trend of rhythmical continuity across pauses is speaker specific in manifestation.

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

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.

References

Fant G & Kruckenberg A (1989). Preliminaries to the study of Swedish prose reading and reading style, *STL-QPSR* 2/1989, 1-83.

Fant G, Kruckenberg A & Nord L (1991A). Some observations on tempo and speaking style in Swedish text reading. *ESCA Workshop on The phonetics and phonology of speaking styles, Barcelona*, 1991.

Fant G, Kruckenberg A & Nord L (1991B). Stress patterns and rhythm in the reading of prose and poetry with analogies to music performance. In: Sundberg J, Nord L & Carlson R, eds., *Music, Language, Speech, and Brain*, Wenner-Gren International Symposium, Series 59: 380-407.

Fant G, Kruckenberg A & Nord L (1992B). Prediction of syllable duration, speech rate and tempo, *Proc. ICSLP 92, Banff,* 1: 667-670.

Gårding E (1981). Contrastive prosody: a model and its application. AILA Congr. 181. *Studia ling.* 35: 146-166.

Heldner M & Strangert E (1996). Personal communication of data.

Horne M, Strangert E & Heldner M (1995). Prosodic boundary strength in Swedish: final lengthening and silent interval duration., *Proc. XIIIth ICPhS,* 1: 170-173.

Kruckenberg A & Fant G (1995). Notes on syllable duration in French and Swedish, *Proc. XIIIth ICPhS, 158-161.*

Lea WA (1980). *Trends in Speech Recognition*, Prentice Hall, Inc.

Strangert E (1991). Pausing in texts read aloud *Proc. XIIth ICPhS* 4: 238-241.

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