UNITS OF TEMPORAL ORGANIZATION. STRESS GROUPS VERSUS SYLLABLES AND WORDS.

Gunnar Fant

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

Phone 46 8 790 7872, Fax 46 8 790 7854

1. INTRODUCTION

This is a contribution to the discussion of a keynote paper, "Some observations on the temporal organisation and rhythm of speech", by Sieb Nooteboom for the XIIth International Congress of Phonetic Sciences, 1991, in Aix-en-Provence. Over the years, Nooteboom has made important contributions to these aspects of speech prosody. Several of his main issues I find uncontroversal. We may all agree that we must have full insight into the particular overall contextual frame that may influence observed data. We need reliable, quantitative models accounting for speech sound durations and ways of testing these models.

Nooteboom is somewhat sceptical about statistical studies on corpuses of connected speech, which may obscure real regularities, and he advocates for well controlled laboratory experiments for testing specific ideas. Indeed - but this latter statement could be turned around to claim that there is a real need for insightful interpretations within a linguistic and pragmatic frame of data from large corpuses of connected speech, and that tendencies observed in "lab speech" might not be equally valid for normal text reading. An optimal combination is needed. There is also a need to relate words in connected speech quantitatively to single word utterances. I have a feeling that short "lab speech" sentences occupy an intermediate position which needs to be better understood and modelled in relation to the two extremes.

I shall have reason to expand on Nooteboom's main issue about stress groups versus words. A large part of Nooteboom's paper is devoted to the defence of the word as a basic unit and to express scepticism about the stress group. The controversy outlined by Nooteboom goes beyond the intentions of Fant and Kruckenberg, [4], but it provides him with an incitement to review recent work, some from his own department. This is interesting material per se but merely adds to the established notion of the word as a basic unit, which we do not deny. The controversy appears somewhat superficial. Nooteboom leaves it to us and others to provide evidence in support of the stress group. This will be one of the objects of my review. Nooteboom's discussion of rhythmical properties is limited to within-word structures. I shall provide a broader basis for the discussion of speech rhythm in relation to stress groups and pauses and to temporal units larger than the stress group.

2. THE STRESS GROUP AS A UNIT OF TEMPORAL ORGANIZATION

The stress group or foot is a domain of speech which incorporates one main stress. The boundaries may be defined so that the stress group comprizes a number of complete syllables. This is the case of the metrical foot. However, the most common definition of the domain of a stress group is the interval between two successive stressed vowels. The latter convention is usually adopted for the study of stressed timed languages as Swedish and English in accordance with rhythmical consideration of stressed vowel onsets approximating the so called P-centers, the locations of perceived beats. These are found to be displaced ahead of the vowel onset if pre-

ceded by a cluster or an unvoiced consonant, [14, 16]. In the analysis of French prosody the stress group is generally considered to end with the last phoneme of an accented syllable. Martin [15] refers to such stress groups as "prosodic words", which become minimal units in an intonational analysis. We have followed a similar principle in the analysis of French prose reading, and we have found specific patterns of durational increase associated not only with prepause stresses but also with minor stresses inside a clause or a phrase [7, 8]. In order to attain a closer conformity with syntactic units Jassem et al [10] suggest a wider definition of the stress group not restricted to ending or starting with a stressed syllable. The stress group has also played a role in phonological systems, e.g. Selkirk [17].

The stress group is accordingly an accepted unit in phonetics and it remains to evaluate its merits and limitations. Here follows a brief summary:

(1) The basic function of the stress group is to serve as a frame for studying quasi-rhythmical aspects of speech as a sequence and alternation of stressed and unstressed syllables [12].

(2) In connected speech the duration of a stress group not spanning a pause or a region of phrase juncture lengthening increases with the number of phonemes or syllables contained in an approximately linear fashion

$$T_n = a + bn \quad (1)$$

where b is the average increment per added unit, syllable or phoneme, and a is an offset value which represents the average stress induced lengthening. We may accordingly identify the average duration of an unstressed syllable as b, whilst a + b represents the average duration of a stressed syllable. Apparently the ratio (a + b)/ b is a measure of stressed/unstressed contrast that can be used as a correlate of an individual or general speaking style [4, 5].

(3) A more detailed analysis reveals a weak tendency of isochrony in stress timed languages, e.g. stressed syllable compensatory shortening when the number of following unstressed syllables is increased. This issue has been taken up in several papers to this congress. In our experience the effect is small, in our Swedish data base about 15 ms per added unstressed syllable. Campbell [2] reports somewhat larger values. It is my impression that it is more pronounced in isolated polysyllabic words or short lab sentences than in connected speech. Also it is associated more with the first and second added unstressed syllables than with additional syllables, see e.g. Strangert [18].

(4) A closer approximation to isochrony appears in read poetry and may be studied in terms of stress groups as a supplement to the formal syllable based metrical foot [11].

(5) The stress group is a convenient unit for discussing tempo, i.e. speech rate. The average duration of a phoneme within a stress group is T_n/n, where T_n is the duration and n is the number of phonemes contained. This observed value may be compared to a predicted value of b + a/n from free foot statistics, Eq.1. One aspect of speech rhythm is the alternation of tempo within and between phrases. Some of these variations are predictable from the text in terms of the density of stressed syllables of content words, which shows systematic variations. A text neutral phrase rhythm of decelerations and accelerations may thus be computed as a reference to which adds the speaker's subjective interpretations [8].

(6) An apparent tie exists between mean stress group durations and pause durations. We have observed a tendency of pauses plus associated prepause lengthening to approximate an average interstress interval [4]. This is typically the case of sentence internal pauses of the order of 300 - 500 ms. Pauses between sentences are longer. We have found multi-modal distributions of pause durations, with some additional correction for prepause lengthening, to approximate two or three or four quanta of the order of an average interstress interval. This is found not only for Swedish but also for English and French. A rhythmical coherence of pauses and tempo with obvious analogy to music performance is typical of a relaxed reading of good speakers, and it is also typical of the reading of metrically structured verse. We have evidence that the average interstress interval within a short time memory span of about 4 seconds preceding a pause or something like the last eight free feet

synchronizes an internal beat generating clock which sets a preferred pause duration. This is exemplified by Fig.1 for Swedish sentence internal pauses.

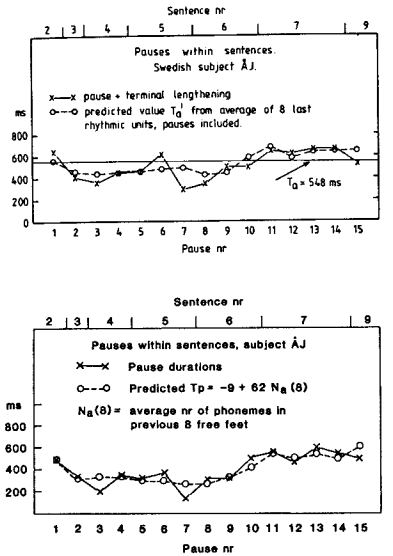

Fig.1. Sentence internal pauses predicted from the last eight free feet, above in terms of mean foot duration, below in terms of average number of phonemes per foot.

Here we have gone one step further and tested the hypothesis that the number of phonemes within a stress group would have the same predictive power as durations. There is a close correlation between number of phonemes per stress group and pause duration. A prediction of pause durations between complete sentences in French is shown in Fig.2.
It is apparent that the eight free feet local reference provides a better prediction than the average foot duration of the whole text. However, it must be stressed that these rhythmical traits are speaker dependent and become upset in conscious efforts to change the overall speech rate or speaking mode. Also, there remains to clarify the underlying perceptual and speech motor mechanisms.

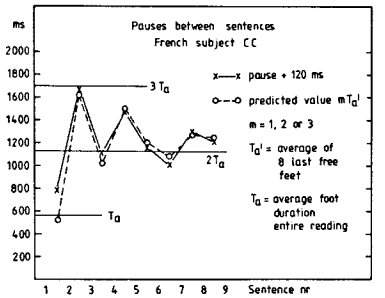

Fig.2. Prediction of pause lengths between sentences in French. The local eight free feet reference provides a better prediction than the long time average foot duration.

3. FINAL REMARKS

Nooteboom's main argument is that much knowledge of speech structure is tied to the word as an organizational unit, whilst this is not the case for stress groups. In my discussion I have stressed the role of the stress group as a unit for structuring rhythmical phenomena and for quantifying stressed/unstressed contrasts. The statement of Fant and Kruckenberg [4] that, in connected speech, the stress group overrides the word is valid in the sense that the stress group as a unit is orthogonal to the word, i.e. it occupies an independent tier of temporal structure above the word level imposing additional constraints. One example is the finding of Bruce [1] that the alternation of weaker and stronger unstressed syllables in Swedish constitutes a rhythmical pattern with a greater consistency within stress groups than within words.
In retrospect, our provocative statement has a wider significance than the role of the stress group. The underlying notion is that of the large differencies frequently encountered between words in isolation and words in context affecting overall duration as well as relative segment durations and patterns. As Nooteboom points out, these depend on a complex of conditioning factors and interactions within and outside a linguistic frame that need to be better understood. The differencies are at times

drastic. The average duration of prepositions is only 20% of the isolated citation form value. Data from "lab speech" experiments are not always representative of the reading of continuous texts. Thus, the recursive segment duration models of Lindblom [13], based on systematically enlarged word and sentence structures, have been quite influential, but they do not seem to have sufficient predictive power for connected text reading. On the other hand, Carlson et al [3] report a significance of stress group alignments in preserving synthesis quality.

At last, a few words about the limitations of stress group statistics. For studies of stressed/unstressed contrasts the regression constants a and b provide a gross measure only. For a more detailed analysis we have to go inside the stress group and perform separate studies of stressed and unstressed syllables and their segmental components [7, 8]. In science we need to use each unit at its best advantage. The stress group is not without interest. It is an established unit.

REFERENCES.

[1] Bruce, G. (1984), "Rhythmic alternation in Swedish", in *Nordic Prosody III*, 31-41, University of Umeå, Almqvist & Wiksell Int.

[2] Campbell, W.N. (1988), "Foot-level shortening in the spoken English corpus", in *Proceedings of FASE 88*, Edinburgh, 489-494.

[3] Carlson, R., Granström, B. and Klatt, D. (1979), "Some notes on the perception of temporal patterns", *Frontiers of Speech Communication Research* (B. Lindblom & S. Öhman, editors), 233-244, Academic Press.

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

[5] Fant, G., Kruckenberg, A. and Nord, L. (1990), "Acoustic correlates of rhythmical structures in text reading", *Nordic Prosody V*, 70-86, University of Turku, Painosalama Oy.

[6] Fant, G., Kruckenberg, A. and Nord, L. (1989), "Rhythmical structures in text reading. A language contrasting study", *Eurospeech 89*, Vol. 1, 498-501.

[7] Fant, G., Kruckenberg, A. and Nord, L. (forthcoming), "Durational correlates of stress in Swedish, French and English", Proceedings of the Second Seminar on Speech Production, Leeds, May 1990. To be published in *Journal of Phonetics*.

[8] Fant, G. Kruckenberg, A. and Nord, L. (1991), "Temporal organization and rhythm in Swedish", *ICPhS* 1991.

[9] Fant, G., Kruckenberg, A. and Nord, L. (1991), "Language specific patterns of prosodic and segmental structures in Swedish, French and English", *ICPhS* 1991.

[10] Jassem, W., Hill, D.R. and Witten I.H. (1984), "Isochrony in English Speech: its Statistical Validity and Linguistic Relevance", (D. Gibbon & H Richter, eds.) *Intonation, Accent and Rhythm. Studies in Discourse Phonology*. Walter de Gruyter, Berlin, New York.

[11] Kruckenberg, A., Fant, G. and Nord, L. (1991), "Rhythmical structures in poetry reading", *ICPhS* 1991.

[12] Lehiste, I. (1977), "Isochrony reconsidered", *Journal of Phonetics*, 5, 253-263.

[13] Lindblom, B. (1975), "Some temporal regularities in spoken Swedish," (G. Fant & M. Tatham, eds.) *Auditory Analysis and Perception of Speech*, 387-396, Academic Press, London.

[14] Marcus, S.M. (1981), "Acoustic determinants of perceptual center (P-center) location", *Perception & Psychophysics* 30, 3, 247-256.

[15] Martin, Ph. (1982), "Phonetic realisations of prosodic contours in French", *Speech Communication* 1, 283-294.

[16] Rapp, K. (1971), "A study of syllable timing," *STL-QPSR* 4/1971, 14-19.

[17] Selkirk, E.O. (1984), *Phonology and Syntax. The Relation between Sound and Structure*, The MIT Press, Cambridge, MA.

[18] Strangert, E. (1985), *Swedish speech rhythm in a cross-language perspective*, Almqvist & Wiksell International, Stockholm.