CONTRIBUTIONS TO TEMPORAL ANALYSIS OF READ SWEDISH

Gunnar Fant & Anita Kruckenberg
Dept. of Speech Communication and Music Acoustics
Royal Institute of Technology, Stockholm

Introduction

This is a summary of some of the findings from the last few years' work on a project centered around the data bank now developing within our group at the Royal Institute of Technology. Earlier reports dealing with both segmental and prosodic aspects have appeared in Fant, Nord, & Kruckenberg (1986; 1987). A more comprehensive report on these studies will be given in a forthcoming issue of the STL-QPSR; see also Fant & Kruckenberg (1988). The prosody project to be reported here aims at studies of: (1) General relations between objective measures - durations and pitch excursions - and subjective estimates of syllabic stress. (2) Inter-stress intervals and their interaction with syntactic boundaries and pauses. (3) Relations between objective boundary region measures and subjective estimates of degree of juncture-boundary marking. (4) Distribution and individual variations of pauses.

Text material and processing

For the analysis performed in the present study, we have selected a paragraph of nine sentences from a novel by Kerstin Ekman, in all 133 words of about 50 seconds' reading time. This is but a small part of the data bank text material. A major part of the analysis was devoted to the reading of a reference subject, ÅJ, who has a clear and engaging reading style without mannerisms. We also analyzed one of the nine sentences read by 15 other subjects, including five females.

Computer-generated spectrograms with synchronous FO, oscillogram, and intensity trace were made of this limited speech material. A segmentation into successive speech sounds corresponding to a broad phonetic transcript was undertaken. The occasional difficulties and ambiguities involved have been discussed previously (Fant, Nord, & Kruckenberg, 1986).

Objective and subjective measures of syllabic stress

Swedish is a stress-timed language with sequences of unstressed syllables alternating with stress syllables. A stressed syllable carries a nucleus of a long vowel followed by one or two short consonants, or no consonant, or the vowel is short and followed by one consonant, or a consonant cluster. A stressed syllable also carries one of two contrasting tones, accent 1 or accent 2.

Duration appears to be the main correlate of stress in Swedish, at least it is more readily quantifiable than associated FO measures. The zone of durational increase with increasing stress is the entire syllable but a larger part is confined to the vowel and the following consonant. This VC nucleus will serve as our major objective measure but we have also been studying durational patterns of entire syllables and vowel-to-vowel units. For each of these different objects, we have constructed normalized measures, syllable duration indexes, to account for variations with the number of phonemes within the unit.

We define the syllable duration index by

$$S_i = 1 + (T-T_u)/(T_s - T_u),$$

where T is the measured duration of a syllable or a vowel-to-vowel unit or a VC unit and T_u is the typical duration of unstressed units and T_s of stressed units determined separately for the specific number of phonemes in a unit. The duration index S_i is thus attained by interpolation/extrapolation with respect to mean values of 1 for unstressed and 2 for stressed syllable units. Most of our studies are based on VC and syllable final or single V units. For our reference speaker ÅJ, we noted T_u=53 ms for V, T_s=185 for V:, and T_u=109 ms for unstressed VC. For stressed VC, we made use of separate references of V:C=236 ms and VC:=212 ms.

A similar process was carried out for S_i calculated on the basis of syllables and V-V units. These two latter alternatives gave rather similar results. In other words, with proper normalization, the durational correlates of stress are very much the same independent of the unit of observation.

SYLLABIC STRESS. OBJECTIVE AND SUBJECTIVE.

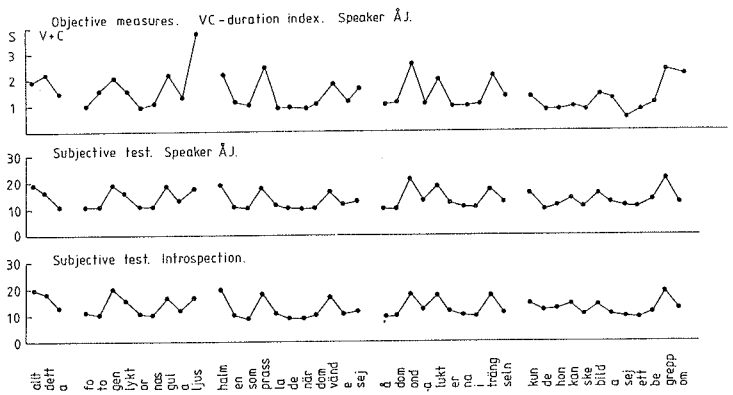

Fig. 1. Syllabic stress. Objective and subjective measures.

In Fig. 1, the VC duration index has been compared to subjective evaluations. Fifteen subjects were asked to make a direct estimate of perceived prominence of each of the syllables in sentence 7. They first made an introspective evaluation from silent reading of the text and then listened to subject ÅJ reading the same passage. The consistency was quite good, standard deviations for a single estimate were of the order of three units within the given frame of 10 for unstressed and 20 for typical stressed syllables. Apart from minor deviations in unstressed syllables, the overall profiles display an apparent similarity. Deviations between objective and subjective scalings of ÅJ's reading may in part be explained by the occasional influence of more extreme inherent durations, such as for /s/ and /a/. The sentence contains four phrases. Those that end with a stressed syllable receive a final lengthening which apparently is ignored in the subjective estimate.

A conspicuous trait is the great similarity between the introspective performance and the listening to subject ÅJ. This could imply that subjects might rely more on their inner "top down" expectancy than on what they hear. However, control experiments involving listening to

several subjects' readings of one and the same sentence showed that listeners estimates did follow the individual variations of produced stress patterns, as evidenced from objective measures and expert listening. One conclusion is that subject ÅJ's interpretation of the text is quite similar to the average of the subjects in the listening test and thus not extreme.

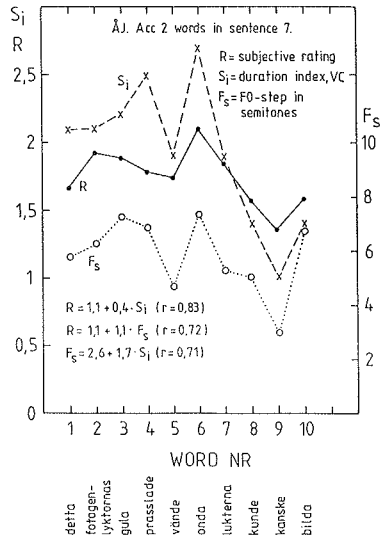

Fig. 2.

In Fig. 2 we have included local F0 excursions as a second objective measure. For this purpose, we have selected the ten stressed syllables of sentence 7 carrying accent 2. The F0 fall within the voiced part of a syllable has been plotted in semitones together with S_i values corrected for inherent durations and the subjective response. the overall correlation is apparent and of the order of 0.7 to 0.8. Larger durations of the voiced part allow larger total F0 excursions, as expected. We have also been able to demonstrate similar covariations with the relative F0 rise in accent 1 syllables. Male and female subjects produce similar F0 contours when measured on a semitone scale.

Connective F0 traits

A few comments will be made on the F0-contour of the reference sentence, Fig. 3. It contains five intonational phrases supporting the syntactic structure. Following Eva Gårding (1984), we have sketched an intonation grid. In the accent domains the grid has a width of the order of half an octave. The overall declination within a phrase is also of this order, i.e., six semitones. The final rise at the end of the sentence could be described as indicating a focal domain. Standard descriptions of Swedish intonation, e.g., Bruce (1977), treat the secondary stress of grave accent words as a discrete unit. The secondary hump is thus considered to be present only in compounds or under the influence of a sentence accent. We often find a weak secondary F0-peak appearing in the second syllable after the main stress, e.g., "vände sig", "kunde hon", "bilda sig", i.e., a pattern of alternation. We feel that in connected non-laboratory speech there is place for a more continuous aspect of stress, the measurable correlates including a gradual appearance of a secondary F0 accent 2 peak. It is an open question whether the high F0 in the second syllable of the adjective "gula" belongs to the accent 2 domain of the word or whether it is a high F0 reference point for the following accent 1 domain of the word "ljus". Perhaps we have an additive effect here.

Other aspects of stress correlates that need to be quantified are voice source characteristics and reduction phenomena including consonant-vowel contrast.

The speech code

A special attention has been laid on the study of rhythmical qualities in text reading. We have fresh evidence supporting a suggestion of Lea (1980) that inter-stress intervals, defined by distances from a stressed vowel to the next stressed vowel of the sequence, act as synchronizing impulses for an internal clock which guides the time we take in making pauses between phrases and sentences.

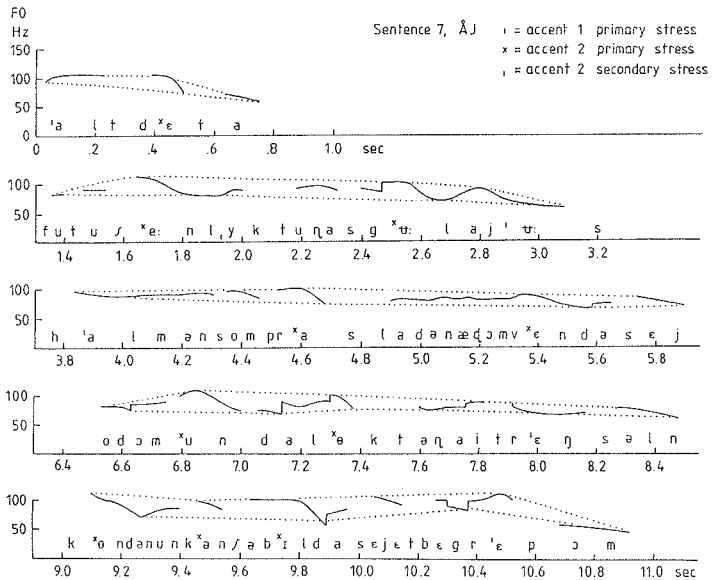

Fig. 3.

We have found that the average duration of the inter-stress intervals not spanning a syntactic boundary sets the basic temporal module of the internal clock. In rhythmical reading, the duration of an inter-stress interval spanning a phrase boundary with a pause tends towards the sum of an inter-stress interval predicted from the number of phonemes contained plus one modular unit of the inner clock, i.e., an average inter-stress interval is added. At sentence boundaries with longer pauses, there is a tendency of one or two additional clock units being added. In other words, noting that we always have a terminal lengthening before and to some extent after pauses, we may restate this finding as an expectancy that the sum of the physical pause and terminal lengthening equals an integer multiple of the rhythmical time constant. This is illustrated in Fig. 4.

Durations of inter-stress intervals are thus largely imposed by language and do not appear to be adjusted o retain isochrony while the rhythmical demand becomes apparent in the planning of pauses. Even without such perfect synchrony, e.g., when a phrase boundary is realized without a pause, the final lengthening alone is capable of signalling the appearance of the boundary.

The constants of the regression line relating inter-stress intervals to number of phonemes appear to be a key to the analysis of components of individual reading and talking styles. These constants should also be of interest in contrastive language studies.

The subjective markedness of phrase boundaries as determined by listening tests was found to correlate well with the duration of the boundary spanning inter-stress interval and with F0-dip measures and appearance of voice creak. In spite of large inter-subject variations of pause durations, the group means showed a clear tendency towards what was observed as a rhythmical norm for a single speaker.

Relative pause time, i.e., the ratio of pause time to reading time, was found to be a consistent speaker-dependent characteristic. The general tendency in speech that the variance of a larger unit is greater than the sum of the variances of its parts as in, e.g., V:C and C:V units apparently also holds for pause durations and terminal lengthening as parts of an inter-stress interval. There is also a weak tendency in this direction relating the total reading time to its parts, effective speech time, and pause time. The larger unit is more stable than the parts.

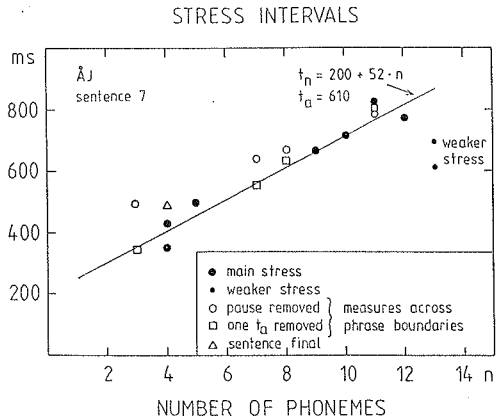

Fig. 4.

References

Bruce, G. (1988): Swedish Word Accents in Sentence Perspective, CWK Gleerup, Lund.

Fant, G. & Kruckenberg, A. (1988): "Temporal structures in Swedish text reading", STL-QPSR 2-3/1988.

Fant, G., Nord, L., & Kruckenberg (1986): "Individual variations in text reading. A data-bank pilot study", STL-QPSR 4/1986, pp. 1-17.

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

Gårding, E. (1984): "Comparing intonation", Working Papers No. 27, Linguistics Dept., University of Lund, pp. 75-99.

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