PROMINENCE CORRELATES IN SWEDISH PROSODY

Gunnar Fant and Anita Kruckenberg
Royal Institute of Technology, KTH, Sweden

ABSTRACT

This is a contribution to a widened insight in acoustical correlates of prominence, largely based on studies of prose reading. A novelty is the establishment of a continuously scaled perceptually determined prominence parameter as a supplement to discrete phonological categories of stress and accentuation. The acoustical analysis is supplemented by continuous records of sub- and supraglottal pressures. Two intensity measures and a new parameter representing spectral tilt are introduced. Results from linear regression analysis are discussed with possible references to production constraints.

1. INTRODUCTION

The purpose of our study is to widen the knowledge base of prosody and in specific the prosody of Standard Swedish. A follow up of earlier work [3-7] is the introduction of a continuous scaling of perceived syllable and word prominence as a supplement to discrete phonological categories. On the experimental level we have introduced an extended acoustical analysis by continuous records of sub-and supraglottal pressure and syllable prominence in synchrony with records of the speechwave oscillogram, spectrogram, F0 and intensity.

On a theoretical level we are facing the demands of a more integrated view, not only by reference to an extended set of parameters, but also by looking into more general problems concerning the interaction of prosodics and segmentals within a multilevel contextual frame. A well recognized topic is the interaction of prominence and grouping, stress and intonation [1-2, 9-10]. FO correlates of stress have to employ different parameters for the two word accents in Swedish, accent 1 and accent 2. We are now in a position to compare 5 subjects´ reading of a one minute long passage from a novel and study effects of imposing focus and identifying default focus in a sentence with reference to a complete set of acoustic parameters.

A novelty is the introduction of two intensity measures, with and without high frequency preemphasis. The difference is adopted as a measure of spectral tilt which we have found to be useful.

Our major tool for studies of relations between acoustic parameters and prominence is regression analysis supported by attempts to explain covariation in terms of underlying constraints of the production mechanism. How and to what extent is intensity determined by subglottal pressure and F0?

A specific problem of old origin [12-15] is the role of subglottal pressure as an intensity and prominence determining factor. Is it confined to high prominence levels only? Is there a relation between subglottal temporal dynamics and prominence?

2. EXPERIMENTAL TECHNIQUES

The technique for prominence rating was described in [3]. A listener crew was engaged in the assessment of each syllable or word in a read text presented over a loudspeaker in repeated chunks of the order of a sentence. The direct estimate technique involves the setting of a pencil mark on a vertical line scaled from 0 to 35 for each syllable or word. This is a continuous interval scale, which we label Rs. As the only guideline, subjects were told that typical values for unstressed syllables would be Rs=10 and for stressed syllables Rs=20.

The speech material originated from a session [5] in which simultaneous measures of true subglottal and supraglottal pessures had been recorded. The speaker, SH, was a medical doctor specializing in voice research. He has a good voice and a standard Swedish pronunciation. Fifteen staff members and students graded the entire corpus of 213 syllables within the nine-sentence paragraph of our standard text. Each sentence was played twenty times in succession over a loudspeaker system. The whole test took about an hour to complete. It incorporated a larger corpus of sentences than that of [3] and should be more representative. The standard deviation among the 15 subjects in our listening crew was of the order of 3 Rs-units only, which implies an uncertainty of the means, $0.7\sigma/(N^{0.5})$, of the order of 0.4 units.

Our earlier tests from the pilot study in [3] showed that word prominence assessments closely follow those of the dominating syllable in the word, i.e. the syllable carrying maximum stress in isolated lexical pronunciation. Word prominence has accordingly been quantified indirectly as that of the dominant syllable of the word. These data are shown in Table 1.

Content words	Rs	N	*Function words*	Rs	N
Numerals	22.8	1	Pronouns	12.5	22
Nouns	19.8	31	Prepositions	11.1	18
Adjectives	18.2	5	Auxiliary verbs	10.7	8
Verbs	17.1	22	Others	9.4	18
Adverbs	17.0	6			
Weighted mean	18.6	65	*Weighted mean*	11.0	64

Table 1. Syllable prominence Rs versus word class. N = number of words.

An example of our routine for assembly of analysis data is given in Figure 1 which shows syllable prominence Rs, oscillogram, spectrogram, sub- and supraglottal pressure, FO and two intensity traces. One is the sound pressure level with flat weighting, SPL and the other, SPLH, is the same but for our standard pre-emphasis

$$G(f) = 10\log10\{(1+f^2/200^2)/(1+f^2/5000^2)\} \quad \text{dB} \qquad (1)$$

which has a gain of 3 dB at 200 Hz, 14 dB at 1000 Hz and 25 dB at 5000 Hz.

SPLH is more sensitive to variations in the region of the second and the third formant, F2 and F3, than is SPL and should accordingly provide a better match to the concept of sonority. Moreover, the difference SPLH-SPL is related to the contribution of formants above F1. We will adopt SPLH-SPL as a measure of spectral tilt, but with an understanding that it is influenced not only by the source but also by the filter function, i.e. by Fa as well as by the particular formant pattern.

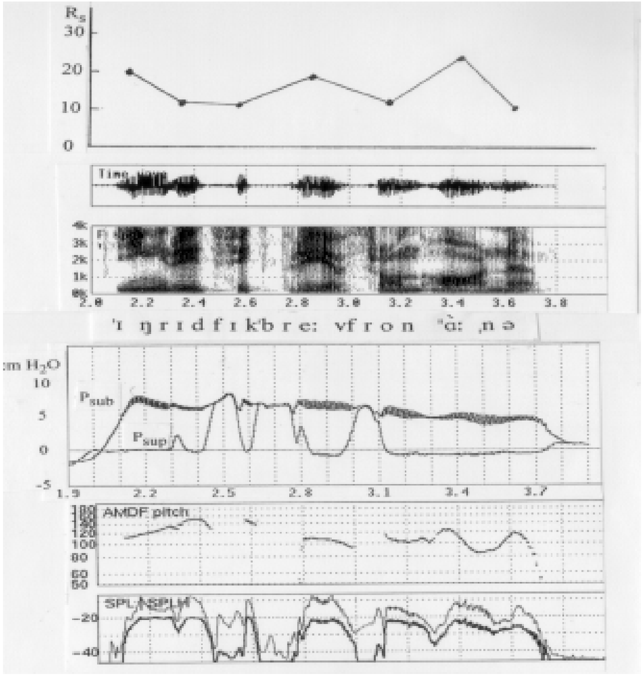

Figure 1. Ingrid fick brev från Arne. “Ingrid received a letter from Arne”,

3. RESULTS

Here follows brief summaries of how prominence is mirrored in the separate acoustic analysis dimensions. One observation of major systematic importance is the contextual variability of correlates which implies the need of context specific specifications and generalizations to relational features, “ceteris paribus” [11]. Our major tool of correlation is linear regression analysis of the growth of prominence Rs with increments of the acoustic parameters.

3.1.Subglottal pressure

Apart from initial rise and decay phases of the order of 150 ms the average contour of the subglottal pressure Psub within a breathgroup of our data, see Figure 1, is a decline from about 6 to 4 cm H_2O. The associated F0 declination is of the order of 4 semitones. Accordingly, correlating Rs versus Psub of each of the 213 syllables provides a very low score. On the other hand, confining the analysis to a restricted context,.e.g. of vowels [a] in phrase initial position as in Figure 2, there is a clear correlation but it can be argued that it is partially maintained by a clustering of the data into two groups, low and high stress.

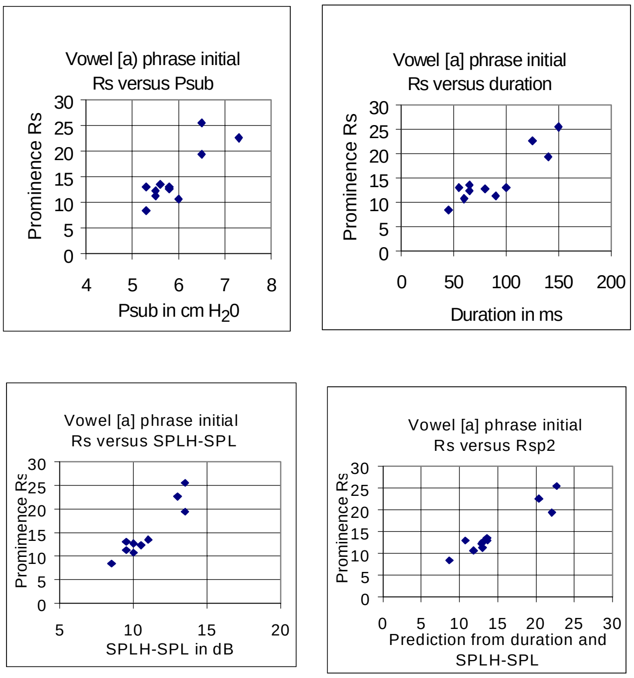

Figure 2. Vowel [a] phrase initial. Rs versus Psub, duration and (SPLH-SPL). Lower right: Rs predicted from duration and (SPLH-SPL).

A more detailed analysis shows a remarkable temporal pattern of Psub starting to build up well in advance of the stressed syllable continuing up to its left boundary followed by a decaying contour [5]. The location of the turning point coincides with the P-center of rhythmical analysis [14].

Furthermore, the rate of decay within the stressed syllable is positively correlated to its prominence, Rs, with a correlation coefficient of r=0.5. There is accordingly some evidence that the subglottal pressure promotes not only focal stress but has also a role in moderate degrees of stress. On the other hand we have evidence that focal stress can be activated without a raised Psub especially in breathgroup final positions.

3.2. Duration

Duration appears to be the most rubust prominence correlate. Figure 3 shows syllable duration as a function of the number of phonemes per syllable under two conditions, a stressed versus unstressed tagging of all syllables in the text, discarding those affected by final lengthening. The two subjects showed similar results. In terms of relative prominence the unstressed averaged Rs=11.8 and the stressed Rs= 19. The corresponding difference in average duration was 100 ms within both 2, 3 and 4-phoneme syllables.

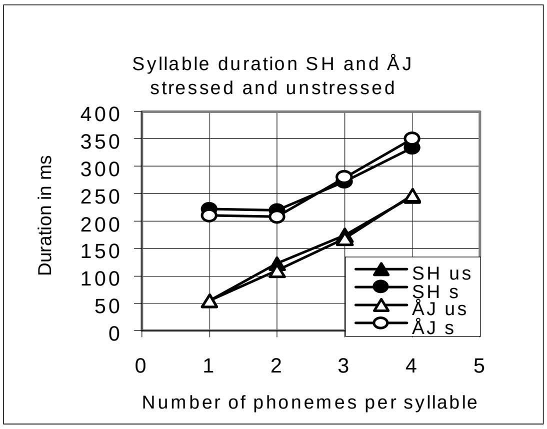

Figure 3. Stressed and unstressed syllable duration.

By means of linear interpolation or extrapolation we may now translate syllable duration to a first order predicted Rs or inversely, predict duration from assumed Rs values. Accordingly, a step in Rs from 10 to 20 would correspond to an increase of syllable duration by 125 ms. A similar correlational analysis involving all vowels in the text showed that short vowels increased by about 80 ms in duration from Rs=10 to 20, for details see [8].

3.3. Intensity and spectral tilt

Regression analysis of prominence Rs versus duration, P-sub and intensity parameters for the selected [a] vowels, Figure 2, showed rather similar patterns and thus a positive pronounced co-variation. The best correlation was found for (SPLH-SPL), (r=0.93), closely followed by SPLH, (r=0,91). For SPL we noted a correlation of r=0.79. The duration data scored r=0.89 and Psub r=0.84. A prediction of Rs from the joint data of SPLH-SPL and duration gave a score of r=0.95. As expected, the gain from combining two strong and covarying predictors is rather small.

An increase of Rs by 10 units, in the domain of accented syllables from Rs=15 to 25, was found to be associated with 6 dB increase in SPL, 9 dB in SPLH and thus 3 dB in the spectral tilt measure SPLH-SPL. The latter is quite sensitive to the particular vowel quality, i. e. the formant pattern, and increases with the degree of articulatory openess. With increasing stress the gain in the SPLH-SPL parameter is only in part related to a less steeply falling voice source slope. For the [a] vowel, the contribution from the vocal tract transfer function is substantial. The combination of source and filter enhances formant amplitudes and thus the perceived sonority.

3.4 Fundamental frequency

A binary classification of syllables as stressed versus unstressed has been useful in many applications as in Figure 3. A more detailed classification is needed for dealing with accentuation, i.e. when prominence may include a significant element of F0 modulation. A minimum of four prominence levels are generally considered in Swedish prosody [2,10]. We have added one more. These are tabulated below with an approximate mapping onto the Rs parameter.

A. Binary system			
	unstressed		Rs<15, average Rs=11
	stressed		Rs>15, average Rs=19
B. Differentiated system			
	Unstressed		Rs = 5-12
	Stressed unaccented		Rs =12-15
	Stressed accented		
		non-focal	Rs =15-20
		focal	Rs =20-25
		higher levels	Rs >25

Table 2. The prominence parameter Rs and prosodic categories.

The F0-patterns of the Swedish word accents as structured by Bruce [1] have a basic component, an F0 fall which in accent 1 is from the H of a the syllable preceding the accented syllable to a minimum L*in the stressed syllable. In accent 2 the drop is from an H* at the left boundary of the stressed vowel to a low level L in the voiced part of the stressed syllable. The accent 2 drop thus occurs later in time than the accent 1 drop. This is a minimum specification of nonfocal accentuation. Focal accentuation adds a rise from L* to a high point Ha in the accent 1 stressed syllable. In accent 2 the rise is confined to the next syllable or later.

In real speech we have to consider prominence levels intermediate between focal and non-focal accentuations. The greater the size of the F0 rise in accent 1 or accent 2 the greater is the perceived prominence. In order to establish quantified relations between F0 measures and prominence, similar to what was suggested in [16],we have adopted a routine of confining the L* of accent 1 to the left boundary of the stressed syllable and introducing a point Ha at its right boundary. In accent 2 the high point following L is labeled Hg. The F0 rises, (Ha-L*) and (Hg-L) are significantly correlated to Rs especially in the range of Rs>17.5. However, at lower Rs values these measures tend to be zero or even negative depending on the superimposed phrase intonation. This is typical of text reading in passages out of focus. In the range of Rs between 15 and 20 the size of the accent 2 (H*-L) fall adds significantly to the prominence but it does not increase much in the range of Rs>20 where the (Hg-L) takes over. The accent 1 (H-L*) initial fall shows a weak negative correlation with Rs and is not a useful prominence correlate. The H, when present, and also L* are influenced more by the overall intonation contour than by the degree of prominence. There are also contextual rules related to the position of an accentuation within a phrase or sentence which we are studying [8].

4. FOCAL ACCENTUATION

One part of our study has been devoted to individual variations. Three males, including subject SH and two females, read the same text referred to above. One major observation is the relevance of the semitone scale of F0 measures which efficiently minimized the female-male differences.

One object was the realization of focal accentuation within the simple declarative sentence. “Ingrid fick brev från Arne”. A typical neutral realization showing the detailed F0-contour was that of Figure 1. We have now compared a neutral reading of this sentence with readings when relatively strong focus was placed on one of the 5 words. The differences, i.e. the changes induced by focal accentuation, define the ordinates in Figure 4. These values were determined at each of the 7 successive syllables, *Ingrid-fick-brev-från-A-rne.* when "Ingrid", "brev" and "Arne" respectively was in focus. The individual spread of data is rather small as indicated by contours of the mean plus and minus one standard deviation.

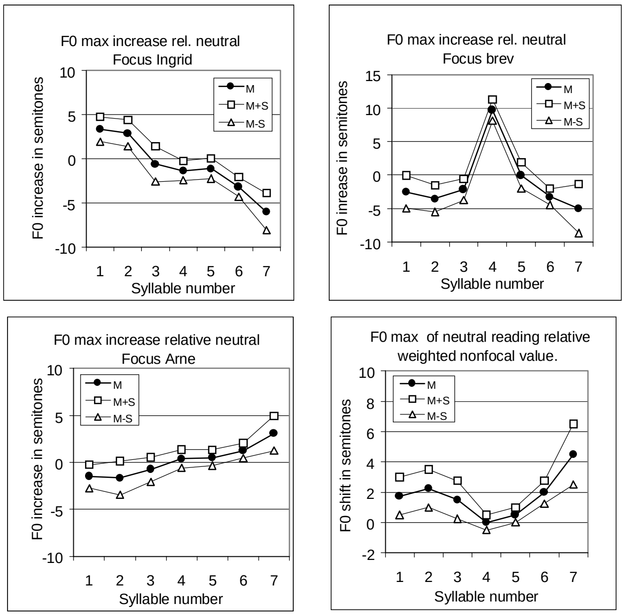

Figure 4. Upper graphs and lower left: F0 shifts induced by focal accentuation. Lower right: Default F0 shifts relative to a neutral baseline.

The words "Ingrid" and "brev" have accent 1 and "Arne" accent 2. The focal F0 peak areas are carried by the Ha in accent 1 and by Hg in accent 2 .They extend over an entire word and represent pivots of the intonation grid [9,10]. The magnitude was of the order of 3-5 semitones. De-accentuation after a focus is considered to be more apparent after than before a focus [1] which applies to "Ingrid" in focus compared to the last word "Arne". However, focus on the intermediate word "brev" shows a rather symmetrical pattern.

A special study was devoted to quantifying the extent and magnitude of "default" focus in the neutral reading. For each syllable we have accordingly calculated the mean of a group of five conditions, that of the neutral version and the values the syllable obtains when focus is placed on the other four words of the utterance. We shall refer to this as the weighted neutral reference as an alternative to the neutral reading alone.

The neutral reading minus the weighted neutral reference is shown in the lower right graph of Figure 4. We noted a difference of the order of 2 semitones in both syllables of the first word "Ingrid", reflecting new information, 2 semitones in the first syllable of the sentence final word "Arne" and 4 semitones in its second syllable carrying the focal accent 2 Hg peak. The sum of default and extra focal increase was of the order of 6-10 semitones. Similar studies with duration, SPL, SPLH and SPLH-SPL as parameters have been carried out [8]. Duration contributed much more to the default prominence of the first word than the last word. Focal accentuation on syllable 5, the preposition "från", caused a much larger duration increase than in any of the content words when in focus. The SPL and (SPLH-SPL) measures showed trends similar to those of F0.

5. ADDITIONAL CORRELATES

A basic component of prominence is articulatory induced segmental distinctiveness versus reduction, hyper- versus hypo forms, [13], which in part influences the spectral tilt parameter (SPLH-SPL). The degree of segmental contrast in vowel-consonant boundaries, e.g. the depth of an [r] dip, increases with prominence at the onset of a stressed word preceded by a sonorant. A brief intensity minimum induced by glottalization is often seen at vocalic word boundaries. A pause preceding a word is occasionally inserted as a prompter.

A cue of strong focal accentuation of a word is the appearance of a fall after the F0 peak which has the combined function of a juncture and signaling extra prominence. Grouping and accentuation are mutually dependent.

REFERENCES

[1] Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

[2] Bruce, G. and B. Granström. 1993. Prosodic modelling in Swedish speech synthesis. *Speech Communication,* 13, 63-74.

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

[4] Fant, G. and A. Kruckenberg. 1994. Notes on stress and word accent in Swedish. *Proceedings of the International Symposium on Prosody, Yokohama.* Also published in *STL-QPSR* 2-3/1994, 125-144.

[5] Fant G., Kruckenberg A., Hertegård S. and Liljencrants J. 1997. Accentuation and subglottal pressure in Swedish, *ESCA workshop on Intonation.* Athens, Greece, Sept 18-20, 1997.

[6] Fant, G., Hertegård, S.,Kruckenberg, A. and Liljencrants, J. 1997. Covariation of subglottal pressure, F0 and glottal parameters. *Eurospeech 97*, 453 – 456.

[7] Fant, G.and A. Kruckenberg. 1998. Prominence and accentuation. Acoustical correlates. In P. Branderud and H. Traunmüller (eds), *Proc of Fonetik* 98 (Swedish Phonetics Conference). Stockholm University, 142-145.

[8] Fant, G. and A Kruckenberg. Acoustic-phonetic analysis of prominencce in Swedish. To be published in Antonis Botinis (editor) *Intonation*, Cambridge University Press.

[9] Gårding, E. 1981. Contrastive prosody: a model and its application. *Studia Lingvistica* 35, 146-166.

[10] Gårding, E. 1989. Intonation in Swedish. *Working papers* 35. Lund University, Department of Linguistics, 63-88.

[11] Jakobson, R., Fant, G. and Halle, M. 1967. *Preliminaries to speech analysis. The distinctive features and their correlates*. The MIT Press, Seventh printing 1967.

[12] Ladefoged, P. 1967. *Three Areas of Experimental Phonetics.* Oxford University Press.

[13] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In W.J. Hardcastle and A. Marchal (eds.), *Speech Production and Modelling*. Kluwer Academic Publishers, 403-439.

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

[15] Ohala J. 1990. Respiratory activity in speech. In W.J. Hardcastle and A. Marchal (eds*), Speech Production and Speech Modelling.* Kluwer Academic Publishers, 23-53

[16] Strangert, E. and M. Heldner. 1995. The labelling of prominence in Swedish by phonetically trained transcribers. *ICPhS 95*, vol 4, 204-207.