Prominence correlates. A study of Swedish

Gunnar, Fant1,2, Anita Kruckenber1, Johan Liljencrants1 and Antonis Botinis2

1Department of Speech. Music and Hearing, KTH, Sweden
2Department of Languages, University of Skövde, Sweden

Abstract

This is a summary of studies of word and syllable prominence in Swedish performed during several years. A unique feature is the correlation of observed acoustic data with a continuously scaled parameter of perceived prominence. Besides the established parameters of duration, F0, intensity, and spectral tilt we have also data on true subglottal pressure. Studies of co-variation within the set of acoustic parameters reveal some interesting relations, some of which can be related to the production mechanism. The major part of the material derives from prose reading, but we have also data from contrasting “lab type” sentences. Some systematic differences appear. Our findings have applications in the development of text-to-speech rules.

1. Introduction

1.1 The prominence parameter

Our technique for prominence rating has been described in [1, 2, 3, 4]. A crew of 15 listeners was engaged in the assessment of each syllable or word from a recording, presented in repeated chunks of the order of a sentence. A scale, Rs, from 0 to 30 was established. It was found that lexically unstressed syllables averaged Rs=11 and stressed syllables Rs=19. Word prominence assessments closely followed those of syllables carrying maximum stress in isolated lexical pronunciation. Nouns and numerals received an average score of 20, adjectives 18, verbs and adverbs 17. All function words received scores lower than 13. These values were obtained for a single speaker, SH, whose recording included true subglottal pressure sensed through a tracheal puncturing probe and mouth pressure through a nasal probe [5].

In connection with our more recent studies of intonation [6, 7] we used the same text, a paragraph of one minute’s length, corpus 1, but now read by five speakers, three males and two females. Rs determinations were limited to accented words and were performed by two expert judges only. We noted on the average about one Rs unit higher values of content words than in the earlier single subject reading. F0 predictions derived from this material have been tested on a second paragraph of the same length, corpus 2, which have been reported in [7].

It is the purpose of the present article to review our earlier work and to report on more recent studies on focal accentuation and parameter interrelation.

2. Studies of prose reading

2.1 Duration

Over the years a considerable amount of work has been devoted to duration in Swedish prose reading [3, 8]. A major finding is the systematic growth of syllable duration with the degree of prominence and the number of phonemes in the syllable. For non-terminal locations we found a difference close to 100 ms comparing average values of stressed and unstressed syllables of the same number of phonemes, e.g. 2, 3, 4 or 5. This is demonstrated in Figure 1.

![Figure 1. Syllable duration.](image-url)
2.1 Intensity parameters

Intensity is usually measured as sound pressure level, SPL, in dB. We have introduced an additional parameter, SPLH, which differs from the SPL by the introduction of our standard pre-emphasis.

\[
G(f) = 10 \log_{10} \left( \frac{1+f^2/200^2}{1+f^2/5000^2} \right) \text{ dB} \tag{1}
\]

It 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 formants, F2 and F3, than is SPL, and could accordingly match the concept of sonority. Moreover, the difference, SPLH-SPL, is a measure of the relative weight of formants in a region above F0 and F1. SPLH-SPL is in part related to the source and in part to the filter function, i.e. the formant pattern. At constant articulation, variations in the SPLH-SPL measure accordingly brings out variations in the high frequency contents of the voice source, which in turn is related to the concept of spectral tilt [11].

A source parameter closely associated with SPL is the Ee parameter, the negative amplitude of the differentiated glottal flow at the closing point of the glottal cycle [12]. All formant amplitudes are proportional to Ee but are also determined by the pattern of formant frequencies. In the lower part of a speaker’s available F0 range, and taking into account the covarying increase of subglottal pressure, the Ee parameter increases in proportion to F0² [12, 13].

The second power relation implies that one semitone increase in F0 is associated with 1 dB increase in Ee and thus in SPL, which is the tendency observed in Figure 2.

2.2 Rs correlates

Results from linear regression analysis relating Rs to acoustical parameters have been reported in [3, 4, 13]. They pertain to the prose reading of subject SH. In order to minimise the influence of context the sampling was limited to [a] vowels in an early position of a sentence. Two examples are shown in Figure 3. One is a graph of Rs versus SPLH-SPL which attained a correlation coefficient of R²=0,87. The other is a prediction of Rs from the joint data of duration (DUR) and (SPLH-SPL) where we noted R²=0,90.

These happened to be the best predictors. Next in order came SPLH with R²=0,82, DUR with R²=0,80, SPL with R²=0,76, Ee with R²=0,60 and F0 with R²=0,45. The low F0 score does not reflect a true importance since the sampling represented raw data outside a proper intonation modelling frame.

The clustering of the data points in two regions centered at Rs values of 23 and 12, reflects alternations between stressed and unstressed syllables originating from content words and function words in this particular context. They are somewhat higher than our mean values for the entire subject SH corpus, Rs=19 respectively Rs=11.

We may summarise average trends as follows. An increase of prominence from Rs=11 to Rs=19 is associated with 4 dB in SPL, 6 dB in SPLH, 2 dB in SPLH-SPL, 1 cm H2O in Psub and 60 ms in DUR. These values are in part based on the [a] data above, but are fairly representative for other vowels, though possibly somewhat lower in less open vowels. It should be kept in mind that the vowel specific, inherent, values of SPLH-SPL are highly influenced by the formant pattern. They vary between 4-15 dB, the lowest for [u:] the highest for none-close front vowels. The step in syllable duration in this Rs interval is close to 100 ms.
3. Accentuation and focus

The linear increase of acoustic parameters with Rs is valid up to a limit of the order of Rs =22-25. At higher Rs levels we encounter non-linear effects, usually saturation, which have to be taken into account. This is especially apparent for duration which does not increase much above Rs=25 whereas F0 and intensity parameters gain dominance. A specific feature is that the H*+L fall of F0 in the major syllable of an accent 2 word tends to saturate above Rs=22 whilst the F0 peak in the secondary syllable carrying sentence accent gains a dominance [6, 7].

With these limitations in mind we may refer to the following set of parameter increase from Rs=15, the lowest limit of accentuation to Rs=25, a prominent focal accentuation. In this range F0 increases by 4-8 semitones. The co-varying increase in SPL is of the order of 6 dB, in SPLH 9 dB, and in SPLH-SPL 3 dB. These values were largely obtained from systematic shifts of the location of focus within a short sentences [3, 4].

Our most recent study [4] provides illustrations of spectrograms in synchrony with F0, subglottal pressure, F0, SPL and SPLH of sentences contrasting in accent type and in neutral versus focal prominence of a test word. These were "Maria Lénar igen, Maria Lenár igen, Maria Lenár igen, Maria lenár igen, Maria lenár igen. Sampled data from these appear in Figure 4.

The temporal pattern is almost the same in the neutral and in the focal version. In the two accent 1 versions, lenár and lenár, the focal reinforcement raises the level of the F0 peak in the accented syllable by 8 and 7 semitones respectively.

Figure 3. Above, Rs versus SPLH-SPL, below Rs versus a joint prediction from (SPLH-SPL) and DUR

Figure 4. Focal and neutral versions of a test word [lenar] with three different accent patterns. The data points pertain to the two vowels and a preceding vowel [a]
engaged. A general impression from a preliminary analysis, see Figure 5, is the primary role of F0 as a focal determinant and the less influence of duration. Also, subjects differed much with respect the relative weight given the various parameters. Thus one of the subjects read the whole sentence louder when in focus but retained prominent SPL contrasts. Another subject made little use of SPL but maintained prominent F0 contours including a pre-focal reduction of F0 thus enhancing a transitional contrast.

![Duration in ms](image1)

![Duration in ms](image2)

5. References


