# THE VOICE SOURCE IN PROSODY

*Gunnar Fant and Anita Kruckenberg*
*Dept of Speech communication and Music Acoustics, KTH, Stockholm, Sweden*

## ABSTRACT

The role of voice source parameters as acoustic correlates to prosodic categories have been studied experimentally and within the frame of production theory. Source parameters contribute more to shaping phrase group intensity profiles and to emphatic stress than to non-focal stressed/unstressed contrasts. The covariation of $F_0$ and source amplitude and source-tract interaction effects are important components of a generative theory.

## INTRODUCTION

This is a progress report from an ongoing project on Swedish prosody [1]. One purpose is to discuss the role of the intensity parameter which in early Swedish phonology [2] was supposed to be crucial for the so called "Expiratory accent", while experience from the last decades of speech analysis and synthesis points at segmental durations and $F_0$-patterns as the primary acoustic correlates of stress in Swedish and English. A recent attempt to reconsider the relative salience of the intensity domain [3] has focused on spectral tilt as a more important parameter than overall intensity.

We shall also expand on the relation between $F_0$ and source parameters in the realisation of accentuation and attempt physiological and acoustic explanations. Source tract interaction phenomena [4] promote a production oriented view of articulatory effort as a component of prominence.

The intensity of a voiced sound has a complex relation to source and filter functions. The source may be extracted by inverse filtering, i.e. a process of removing the filter function from the sound. The glottal flow, thus recovered, is specified by an amplitude factor $E_e$ and a few shape parameters. $E_e$ is defined as the slope of the glottal flow at the closing discontinuity which usually coincides with the amplitude of the peak of the glottal flow derivative [5]. The most important shape parameter within the LF model of glottal flow [5-6] is the frequency $F_a$ at which the source spectrum attains an additional -6dB/oct slope. An increase of emphasis is usually accompanied by an increase in both $E_e$ and $F_a$. Formant amplitudes are proportional to $E_e$ while an increase in $F_a$ provides a relative gain at higher frequencies.

## GLOBAL CONTOURS

The voice source has an important role in establishing groups and boundaries, e.g. shaping the onset, rises and declination within a phrase. From studies of prose reading we have found a typical phrase intensity contour with an initial rise lasting about 100 ms usually followed by a declination of about 4 dB per second and a more rapid decay in the last 400 ms ending with a final abduction gesture in prepause voiced segments or with a creaky voice termination at a voiced juncture. The main part of the contour may be level or show a rise-fall indicating a prominent focal domain in partial conformity with the $F_0$-contour. The declination of intensity is in part the automatic consequence of the $F_0$ dependency of $E_e$, in part a decline of about 3 dB per octave fall in $F_0$ at constant $E_e$ and to a part the consequence of a decreasing lung pressure within the phrase. A final abduction gesture in the last 50-100 ms is associated with the main attributes of breathy voicing, e.g. increasing decline of the spectral slope (decreased $F_a$) and increased F1 bandwidth. Phrase junctures with continued voicing are often produced with creaky voicing accompanying the local F0 minimum.

Another aspect of the global contour is an alternation of rises and falls of overall voice intensity and tempo within a paragraph which adds an element of engagement.

## STRESS AND ACCENTUATION

Our studies of prose reading have confirmed the relative subordninate role of intensity as a stress correlate [1]. Within a corpus of about 200 syllables



the average difference in intensity, defined as sound pressure level in a lowpass (LP) 1000 Hz band comparing stressed and unstressed vowels was 2.5 dB and about 1 dB higher values with highpass (HP) 1000 Hz measures. $E_e$ measures were closely proportional to LP measures. No significant differences between phonemically short and long vowels were found.. Maximally open stressed vowels showed 1-2 dB higher LP and $E_e$ values than maximally close vowels and also a greater stressed/unstressed contrast than more close vowels. The maximally constricted phase of long stressed [u:] [ʉ:] [i:] [y:] are not included in these data. They showed an additional weakening of the order of 3-6 dB.

Source amplitude and intensity play a greater role in emphatically stressed words. A separate study of specially constructed "lab sentences" contrasting in word accent types and in a variation of the place of focal accentuation showed a more substantial range of variation.

These are illustrated in Figures 1-2. As discussed in [1] the source amplitude $E_e$ follows $F_0$ up to a critical frequency above which $E_e$ tends to decrease. For the male subject, Figure 1, it was found at $F_0$=130 Hz. A closer analysis reveals a 1.7 power proportionality of $E_e$ with respect to $F_0$ in the low frequency ascending branch. The break is especially apparent in the female voice, Figure 2 where the focal intonation peak overshoots the critical $F_0$=215 Hz causing a local intensity minimum. The asymmetry of the surrounding maxima suggests the presence of a larger subglottal pressure in the rising than in the falling part of the $F_0$ contour. The minimum is not always present. It was found in two out of 4 subjects and was occasionally missing for the subject AK of Figure 2. It can be less apparent in the intensity than in the $E_e$ or in the speech waveform display, see Figure 1, which to some extent might be explained by the general relation of intensity being proportional to $F_0$ at constant $F_0$.

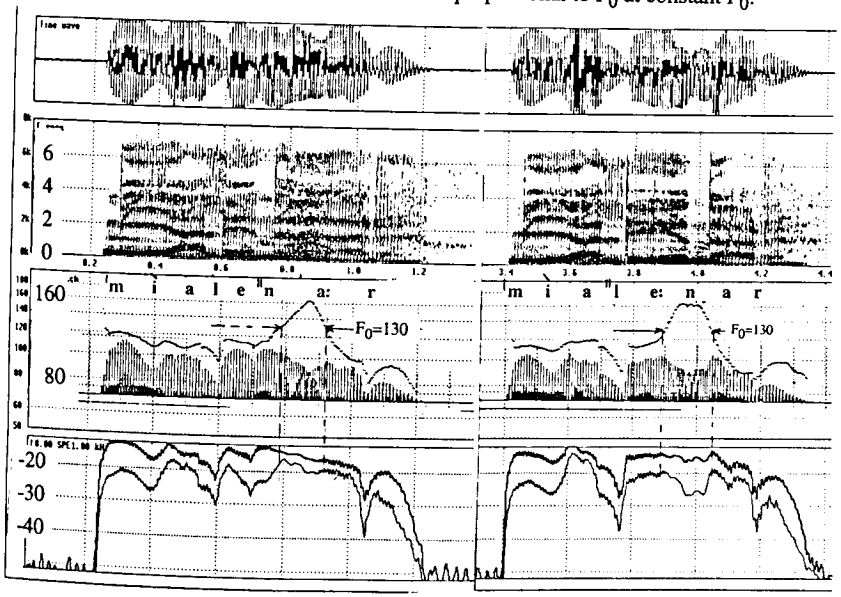


*Figure 1. Male subject, contrasting stress location. $E_e$ is superimposed on the $F_0$ display. The lower intensity curve is produced with high frequency preemphasis.*

A study of the first and second syllables of the contrasting words "Lenár" [lenɑ:r] and "Lénar" [le:nar] both in focus showed large intensity effects for subject AK. The overall unweighted step in SPL from the first to the second vowel was -6 dB for "Lénar" and + 3dB for "Lenár".

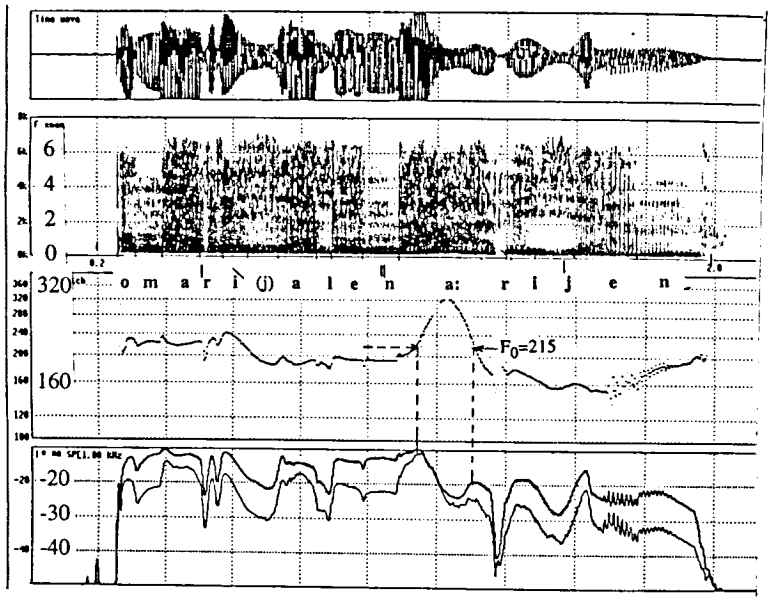


*Figure 2. Female subject. Lexical stress on second syllable of "Lenár". Processing and display as in Figure 1.*

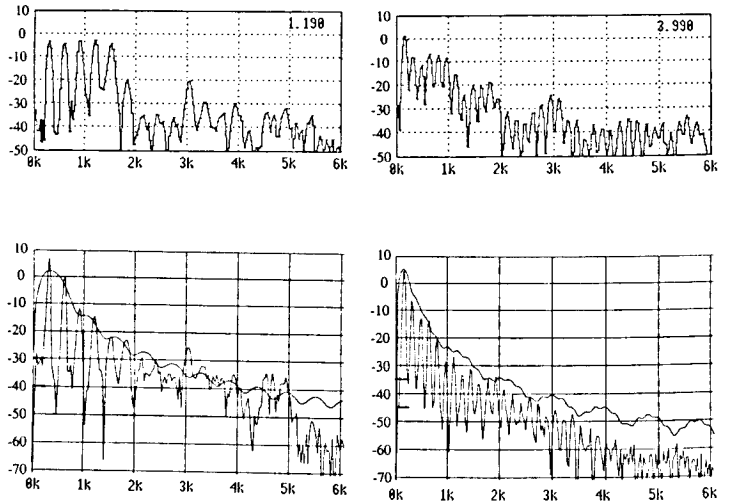


*Figure 3. Spectra of stressed long [ɑ:] upper left from "Lenár" in Figure 2 and of unstressed [a] upper right from the contrasting "Lénar" together with their corresponding source spectra below. The envelopes in the source spectra pertain to a LF model match.*

The corresponding high frequency preemphasized measures (+22 dB increase from 200 Hz to 5000 Hz) were -6 dB and +9dB respectively which reveals a significant high frequency gain in the stressed [ɑ:].



This is further illustrated in Figure 3 by the corresponding amplitude-frequency spectra of the two vowels and their source spectra derived from inverse filtering. They differ by a factor 2 in the $F_a$ parameter which was 400 Hz for the stressed vowel and 200 Hz for the unstressed vowel. Subjects GF, on the other hand, produced much less intensity contrast, and even a reverse trend with the two contrasting words out of focus.

The intensity minimum in the long vowel [i:] of "Maria" in Figure 2 and "Mia". in Figure 1 is a typical instance of the articulatory gesture towards a [j] element in the middle of the vowel imposing a source filter interaction. The local dip in source amplitude $E_e$ is substantially increased under influence of focal stress. Similarly, the $E_e$ and the intensity minimum of the [*l*] in "Lénar" is enhanced and prolonged versus the [*l*] in "Lenár", see Figure 2, or versus the same word out of focus which enhances the consonant-vowel contrast at the onset of the stress gesture, i.e. at the P-center [7]. This is an additional stress correlate to consider [1].

## DISCUSSION

Stressed syllables do not differ significantly from unstressed syllables in terms of source amplitude and high frequency contents unless produced with a marked emphasis which implies an increased subglottal pressure. This finding is coherent with the general conclusions in [8]. Continuously scaled measures of perceived prominence of syllables in prose reading have been correlated to duration and local $F_0$ measures [1] which both provided correlation coefficients of the order of r=0.9 to be compared to r=0.5 for recent studies of intensity and $E_e$.

The essential element of a focal versus nonfocal accentuation is the $F_0$ contour with or without significant increases in duration or in source properties. The individual variability is large and need to be related to the possible presence of a local subglottal pressure pulse and its synchrony with the glottal $F_0$ gesture and the stressed vowel onset.

Increase of stress also affects the consonant vowel articulatory and acoustic contrast not only in terms of formant patterns but also in terms of glottal source efficiency, i.e. a reduction of source amplitude and high frequency contents with increasing articulatory narrowing. This interaction may cause a narrow vowel to lose intensity at increased stress levels. Another instance of reversed stress intensity relation is when $F_0$ in focal accentuations overshoots a high critical value above which source amplitude and intensity is reduced unless backed up by an increasing subglottal pressure.

## ACKNOWLEDGEMENT

This research has been supported by a grant from The Bank of Sweden Tercentenary Foundation.

## REFERENCES

[1] Fant, G. and Kruckenberg, A. (1994), "Notes on stress and word accent in Swedish", *Proc. Int. Symp. on Prosody*, Sept. 18 1994, Yokohama. Also published in *STL-QPSR* 2-3 1994, pp.125-144.

[2] Elert, C.C (1968), *"Allmän och svensk fonetik"*, Almquist & Wiksell.

[3] Sluijter, A.M.C. and van Heuver, V.J. (1994), "Spectral balance as an acoustic correlate of linguistic stress". Manuscript submitted to *J.A.S.A.*

[4] Bickley, C.C. and Stevens, K.N. (1986), "Effects of a vocal tract constriction on the glottal source: Experimental and modelling studies", *Journal of Phonetics* 14, pp. 373-382.

[5] Fant, G., Liljencrants, J. & Lin, Q. (1985), "A four-parameter model of glottal flow", STL-QPSR 4/1985, pp. 1-13.

[6] Fant, G., Kruckenberg, A., Liljencrants J. & Båvegård, M. (1994), "Voice source parameters in continuous speech. Transformation of LF-parameters", *ICSLP-94*, Yokohama.

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

[8] Stevens, K.N (1994), "Prosodic influences on glottal waveform: Preliminary data", *Int. Symp. on Prosody*, Sept. 18 1994, Yokohama, pp. 53-63.