Co-variation of acoustic parameters in prosody

Gunnar Fant & Anita Kruckenberg

Speech, Music and Hearing, KTH, CSC computer Science and Communication

Abstract

We have studied two aspects of co-variation. One is inherent in the speech production mechanism, in particular the voice source and its dependency of subglottal pressure and F0. These relations explain how speech intensity co-varies with F0 in connected speech, which has led us to define a mid-frequency F0r in a speaker's available intonation range. The upper part, F0>Fr, conveys sentence and focal accentuation. Here the sound pressure level SPL saturates with increasing prominence. The role of subglottal pressure contours in shaping prosodic boundaries, e.g. at terminal junctures, has been given a renewed attention. We have observed an asynchronic timing of F0 and subglottal pressure. F0 peaks tend to lag subglottal maxima. Phoneme and syllable duration, F0 peak levels and also SPL are highly correlated with our prominence parameter RS. Individual variations exist, especially at very high prominence levels.

1. Introduction

Our studies of speech prosody were initiated in 1986 and have gradually resulted in a complete model for Swedish with potential applications in other languages, the FK system (Fant & Kruckenberg, 2004; Fant 2004). It has been derived from a corpus of prose reading of a novel employing three male and two female subjects and is supplemented by a phonetically structured database of durations from a larger corpus.

A recurrent topic has been parameter co-variation (Fant, Kruckenberg & Liljencrants, 2000). The purpose of the present article is to summarize our findings and clarify some points of general interest, including the role of subglottal pressure in shaping prosodic boundary regions.

A unique resource is our continuously scaled word and syllable prominence parameter RS derived from perceptual tests. It was introduced by Fant & Kruckenberg (1989) and involved a single speaker. Our present RS data is the average from our five subjects in prose reading determined by the authors.

We are concerned with two major aspects of co-variation, what is inherent in the production process and what is a part of the speech code in the frame of the RS parameter. To what extent can observed patterns of co-variation in connected speech be related to production?

2. Parameter co-variation

2.1 Subglottal Pressure

A laryngologist, SH, one of the five subjects in our prose reading, served as a subject for measurement of sub- and supraglottal pressures during the recording of a one-minute long paragraph from our main corpus. A number of contrasting sentences were also included. Subglottal pressure was measured through a tracheal puncturing probe, and supraglottal pressure through a probe inserted through the nasal pathways (Fant, Hertegård, Kruckenberg & Liljencrants, 1997; Fant, 2004). The entire material from 20 sentences is documented in a major publication (Fant, Kruckenberg, Liljencrants & Hertegård, 2000).

2.2 The F0 mid-frequency in voice production

2.22 Glissando studies

The term glissando implies sustained phonation at a gliding pitch. Several studies have indicated the role of an F0 mid-frequency, here labelled F0r, at which the co-variation of F0, intensity and also subglottal pressure, Psub, change. We have indications from glissando sustained phonation as well as from inverse filtering and models of connected speech.

Our reference subject SH provided us with a set of glissando phonation of the vowel [ae]. Measures of the voice source amplitude Ee and Psub as a function of F0 at varying overall voice effort were derived. A stylised summary of the results appears in Figure 1.

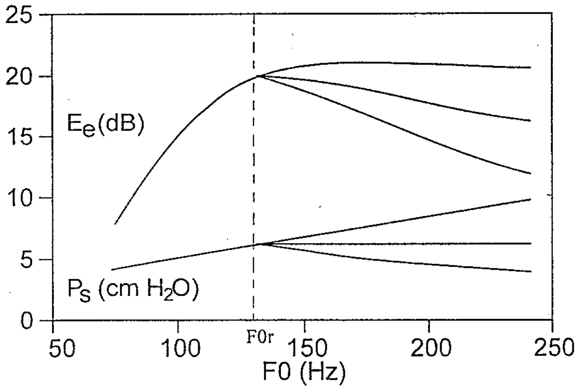

Figure 1. Overall trends of the co-variation of subglottal pressure Ps and voice source amplitude Ee in a set of glissando phonations. Mean values from varying voice efforts. Subject SH.

There is a distinct break at a mid-frequency F0r at about 130 Hz, where Psub and Ee saturate or decay depending on how the voice effort is maintained at the end of the phonation. A continuing rise of Psub above F0r is found in singing (Sundberg, Andersson & Hultqvist, 1999). Voice effort induced variations in Psub in the region below F0r occur but have been omitted for the sake of clarity.

2.22. Connected speech

The lower part of Figure 2 pertains to inverse filtering from connected speech of a third subject.

Here the Uo maximum is found at 90 Hz. Ee increases by about 12 dB per octave of F0 and reaches a maximum at about 110 Hz.

We accordingly have support from both glissando phonation and connected speech to define a critical F0r in the middle of a speaker's available intonation range. A representative model is that intonation covers a range of two octaves in F0, with F0r situated in the centre at about 110-130 Hz for a male voice and 220-260 Hz for a female voice.

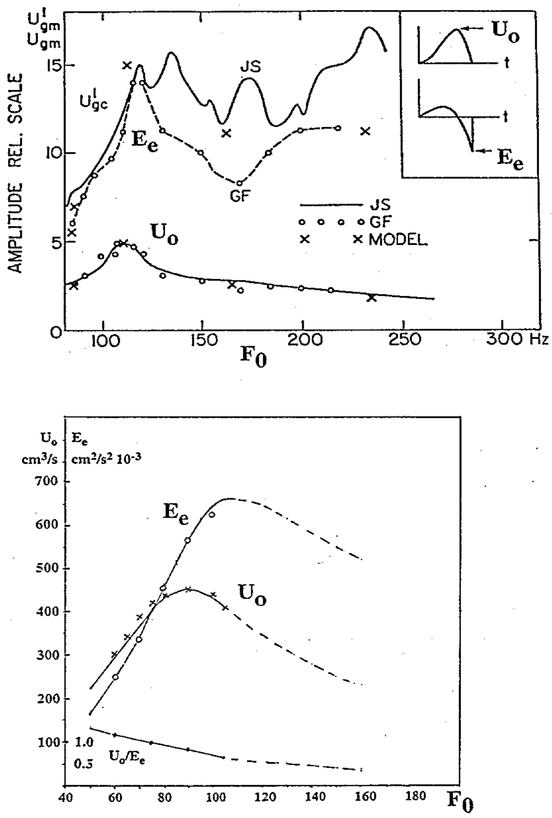

Figure 2. Inverse filtered data from glissando phonation above, and from connected speech, below.

These data derive from the early findings of Fant (1982). The top part of Figure 2 pertains to inverse filtering of glissando phonation of the vowel [ae] with JS and GF as subjects. The amplitude of glottal flow Uo as a function of F0 was quite similar for the two subjects, with maxima at 115 Hz. Also the voice source scale factor Ee ,derived from the negative going peak of the flow derivative, showed similar trends. There was a distinct rise up to a limiting frequency of 120 Hz followed by a saturation.

The concept of a lower and a higher part of a speaker's intonation range explains findings from studies of co-variation of F0 and SPL in our prose reading material.

2.3 F0 and intensity

We have detailed results from studies of the co-variation of sound pressure level SPL and F0 in our main corpus of prose reading, see Figure 3.

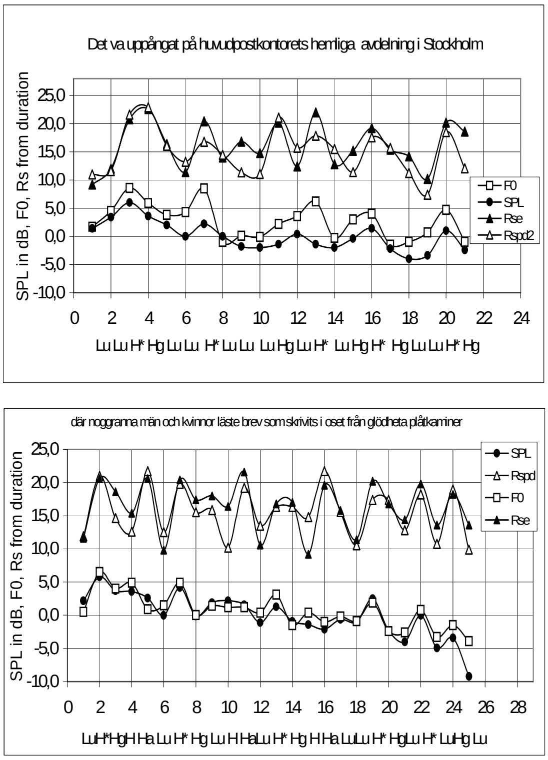

Figure 3. A main clause above, and a subordinate clause below. The lower curves portray the co-variation of F0 and SPL. The top curves show syllable prominence predicted from duration, Rspd, and assessed from listening, RS.

Here syllables carrying sentence or focal accent occupy F0 values in a region above F0r. This may be seen in the upper part of Figure 3 pertaining to a main clause. SPL follows F0 except in prominent syllables where SPL reaches limiting levels. In the following relative clause, the lower part of the figure, all F0 peaks are below F0r and SPL shows an increase by about 1 dB per semitone of F0.

In Figure 3 has been added a comparison of predicted RS from the five subjects' mean of observed syllable durations and our perceptually based RS data. There is a fair agreement.

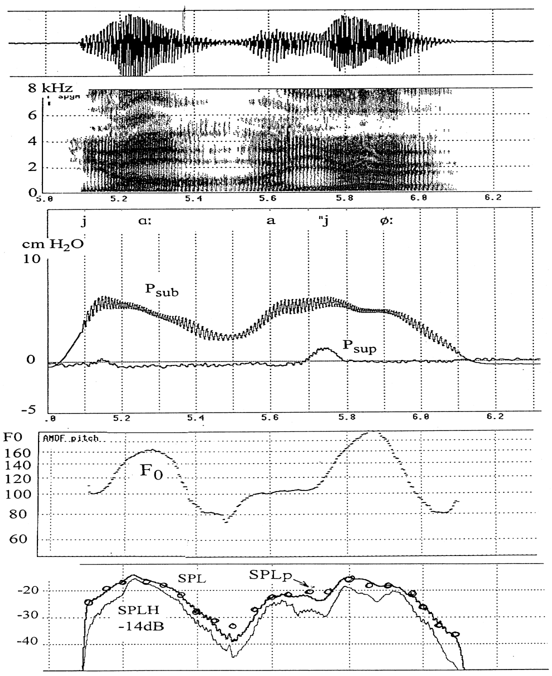

Figure 4. A two word sentence "Ja ajö" (Yes, goodbye) illustrating the location of F0 peaks in regions of falling Psub, and of SPL predicted from F0 and Psub, small open circles, which are close to measured data, solid line.

4. Subglottal pressure as a prosodic boundary marker

A rather recent finding is the active role of a local Psub fall in the termination of a prosodic group and its synchrony with an F0 peak. An example is shown in Figure 4.

The subglottal pressure, Psub, is raised in advance of a syllable boundary in anticipation of an F0 peak located in a region of falling Psub. This is typically the case of long Swedish vowels terminating a prosodic group. In phonemically short vowels the effect is less apparent.

Figure 4 also demonstrates the predictability of SPL from F0 and Psub.

A more difficult task is to predict Psub from F0 and SPL, which is limited by articulatory interference, but can bring out some essentials (Fant & Kruckenberg, 2005).

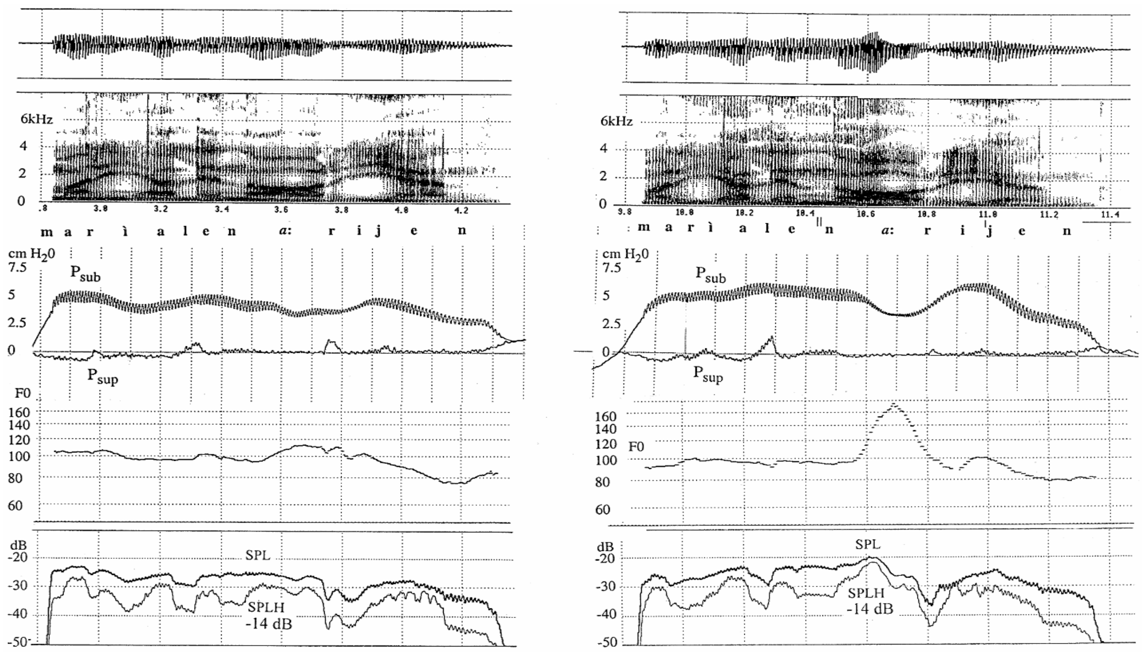

Figure 5. Maria Lenar igen. (mariale’n :rijen). *Left neutral, right high prominence.*

In Figure 5 we have access to the full set of synchronized time functions of the Wspect display (Liljencrants, 2006). In the right hand part we may observe how the intensity parameters SPL and SPLH, intensity with high frequency pre-emphasis, show a distinct drop synchronized with the drop in subglottal pressure Psub, while the F0 peak is delayed towards the Psub minimum. Observe the relation to the bottom curve of Figure 1.

A detailed study of other juncture correlates appeared in Fant & Kruckenberg (1989).

5. Concluding remarks

The significance of an intonation mid-frequency F0r has been demonstrated in sustained speech as well as in connected speech. Focal and sentence accent occupy the upper half of the intonation range, F0>F0r.

A falling subglottal pressure contour has an important role as a terminal juncture boundary marker.

References

Fant G (1982). Preliminaries to analysis of the human voice source. KTH STL-QPSR 4/1982, 1-27.

Fant G (2004). Speech Acoustics and Phonetics. Selected Writings. Kluwer Academic Publishers, Springer.

Fant G, Hertegård S, Kruckenberg A & Liljencrants J (1997). Covariation of subglottal pressure, F0 and glottal parameters. Eurospeech 97, 453-456.

Fant G & Kruckenberg A (1989). Preliminaries to the study of Swedish prose reading and reading style. KTH STL-QPSR, 2/1989, 1-83.

Fant G & Kruckenberg A (2004). The FK prosody model. Analysis and synthesis. Nordic Prosody IX Lund, Sweden.

Fant G & Kruckenberg A (2005). Covariation of subglottal pressure, F0 and intensity. 9th European Conference on Speech Technology and Processing, Interspeech Lisboa, Paper ID #2411.

Fant G, Kruckenberg A, Liljencrants J & Hertegård S (2000). Acoustic phonetic studies of prominence in Swedish. KTH TMH-QPSR 2/3 2000, 1-52.

Fant G, Kruckenberg A & Liljencrants J (2000). Acoustic-phonetic analysis of prosody in Swedish. In A Botinis (ed). Intonation, Analysis, Modelling and Technology. Kluwer Academic Publishers, 55-86.

Liljencrants J (2006). Wspec. http://www.fonema.se/wspect/wspect.html <johan@fonema.se>

Sundberg J, Andersson M & Hultqvist C (1999). Effects of subglottal pressure variations on professional baritone singers' voice sources. J. Acoust. Soc. Am. 105/3, 1965-1971.

More references are to be found in http://www.speech.kth.se/~gunnar

KTH QPSR reports can be down-loaded from http://www.speech.kth.se/publications/.