Covariation of subglottal pressure, F0 and intensity

Gunnar Fant and Anita Kruckenborg
Department of Speech Music and Hearing, KTH, Stockholm
gunnar@speech.kth.se

Abstract
This is a report summarising results from studies of true subglottal pressure, supraglottal pressure and speech wave data. We have derived co-variation patterns, which allow a prediction of intensity from subglottal pressure and F0, and conversely a prediction of a subglottal pressure contour from F0 and intensity.

Of special interest is the significance of a mid-point in a speakers’ available F0 range at which the relations change. In the lower F0 range subglottal pressure and intensity rise with F0, whilst in the upper part they tend to saturate. In connected speech we find a build-up of subglottal pressure well in advance of a stressed word, and a decay of subglottal pressure in the final part of a phrase starting already in the falling branch of an F0 peak.

1. Documentary material
Our study is concerned with the co-variation of subglottal pressure, F0 and intensity within a prosodic frame. A substantial part of our material includes a unique set of synchronized recordings of true subglottal pressure obtained from a sond inserted externally into the larynx, together with supraglottal pressure and the speech wave. A male laryngologist served as a subject [1-3].

A most important tool has been the speech processing and display system Wspect, developed by Johan Liljencrans. Besides oscillogram and spectrogram it generates F0 on a log scale, with a standard of one semitone per 2 mm. Two intensity traces are provided, the regular SPL and one with high frequency pre-emphasis, SPLH [1-2, 5-7] defined by a gain

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

(1)

In relation to SPL it emphasizes the second and higher formants. Accordingly, SPLH-SPL serves as a measure of spectral tilt within the constraints of a given formant pattern and it increases with stress [5-7].

When available, records of sub- and supraglottal pressures, Psub and Psup, have been included as well as synchronized traces of perceived syllable prominence RS, derived from listening tests [7]. Our complete set of 19 illustrations, covering one minute of prose reading appear in [1]. They constitute a documentary material of general interest.

The upper part of Figure 1 is an example of a neutral declarative sentence. At the onset of a breath group, voicing starts at a Psub of 3-5 cm H2O. The duration of the rise to an initial value of the order of 6-8 cm H2O is 120-180 ms.
The initial voice onset requires vocal fold adduction, and the normal voice offset is usually caused by abduction. As required, \( \text{Psub} \) equals \( \text{Psup} \) at vocal tract closure, which usually are regions of pressure build-up. At less complete degrees of articulatory constrictions the driving pressure becomes \( \text{Pdr} = \text{Psub} - \text{Psup} \). The aéra-dynamical consequences of articulatory interaction in specific phonetic frames have been studied in [1] and [2].

An example of high focal stress is illustrated in the lower part of Figure 1: “å de drog bakom dom…”., “there was a draught behind the”. The F0 peak occurs in a region of falling \( \text{Psub} \) and elevated SPL and SPLH.

## 2. Parameter covariation

### 2.1. The F0 midfrequency, \( F_{0\text{ref}} \)

Of special interest is the extent to which the intensity contour follows the F0 contour and how the co-variation is mediated by subglottal pressure. We have found evidence for the relevance of a mid-frequency \( F_{0\text{ref}} \) in a speaker’s available F0 range, below which the SPL intensity follows the F0 contour. In the upper frequency range, \( F_{0} > F_{0\text{ref}} \), the SPL measures tend to saturate with increasing F0, [8].

These relations are governed by the subglottal pressure which rises with F0 up to \( F_{0\text{ref}} \) and then tends to level off. This is less so in the speech of singers who usually maintain a rising \( \text{Psub} \) and SPL above their \( F_{0\text{ref}} \).

An indication of a breaking point in F0 was found already in inversed filtered data of an early publication [4]. The upper part of Figure 2 pertains to glissando phonations, i.e. sustained phonation with gliding pitch of a vowel [ae:]. The glottal flow parameter \( \text{Uo} \) and the corresponding voice source excitation amplitude \( \text{Ee} \) are plotted as a function of F0 for two subjects, JS and GF.

![Figure 2. LF-parameters \( \text{Ee} \) and \( \text{Uo} \) as a function of F0. Above, subjects JS and GF in glissando phonation. Below, subject ÅJ, data sampled from prose reading.](image)

The \( \text{Ee} \) parameter is a direct determinant of SPL as governed by the basic proportionality between \( \text{Ee} \) and formant amplitudes [5-8].

The two subjects have quite similar patterns of Uo as a function of F0, but they differ in Ee. Both show the same breaking points of 110 Hz for Uo and 120 Hz for Ee, but subject JS maintains higher Ee levels above the breaking point.

The lower graph of Figure 2 shows Uo and Ee data as a function of F0 obtained from inverse filtering of passages from a prose text read by our reference subject in earlier studies, ÅJ. His Uo curve levels off at 90 Hz and Ee at 100 Hz. Evidently, the breaking points for Ee could be labelled \( F_{0\text{ref}} \).

Figure 3. \( \text{Ee} \) and subglottal pressure \( \text{Psub} \) as a function of F0 in four glissando phonations, male subject SH.

### 2.2. Covariation of F0, \( \text{Psub} \) and SPL.

We now have access to more recent findings from the session where sub-and supraglottal pressures were recorded. A high speed fiberscope photography of our subject indicates a maximum of glottal area in the mid-frequency range, here around 130 Hz.

Our major source for deriving analytical rules connecting F0, \( \text{Psub} \) and SPL is the data in Figure 3, [1], [2] and [8]. It pertains to glissando phonations of the vowel [ae] at four different levels of voice effort. The \( \text{Psub} \) parameter is in cm H2O. The \( \text{Ee} \) parameter, here plotted in dB, was derived from inverse filtering. There is a rising contour of \( \text{Psub} \) and \( \text{Ee} \) up to the mid-frequency \( F_{0\text{ref}} \). At higher frequencies \( \text{Psub} \) levels off but is subject to increase with voice effort, whilst \( \text{Ee} \) shows clear tendencies to increase or decrease with F0 depending on the voice effort.

From a statistical analysis of the co-variation of \( \text{Ee} \) and \( \text{Psub} \) with F0 in the range F0<\( F_{0\text{ref}} \) we have derived the following relations: \( \text{Ee} \sim F_{0}^{1.35} \) at constant \( \text{Psub} \); \( \text{Ee} \sim \text{Psub}^{1.1} \) at constant F0 and \( \text{Psub} \sim F_{0}^{0.7} \) at constant Ee. Here \( \text{Ee} \) is not in dB but a scalar value and F0 is in Hertz.

Accordingly, with co-varying \( \text{Psub} \):

\[
\text{Ee} \sim F_{0}^{1.35} \cdot \text{Psub}^{1.1} \sim F_{0}^{2.1}
\]

(2)

which implies that \( \text{Ee} \) is increased by 12.5 dB/oct in F0 when the co-varying \( \text{Psub} \) is taken into account. This is roughly one dB per semitone increase of F0. On the other hand, in terms of \( \text{Psub} \) and a normal co-varying F0 we find...
A general model of \( E_e \) in dB, as a function of \( P_{\text{sub}} \) and \( F_0 \), valid for the entire \( F_0 \) range has been derived

\[
E_e = K + 20 \log_{10} \left( P_{\text{tr}} 1.1 x_n 1.35 \left[ (1-x_n) 2 + x_n 2/Q^2 \right]^{-0.5} \right) \tag{4}
\]

where \( x_n = F_0/F_0^{\text{ref}} \) and \( Q \) is of the order of 1.25. A higher \( Q \) will enhance the turning point. \( P_{\text{tr}} \) is the transglottal pressure drop, \( P_{\text{sub}} - P_{\text{sup}} \), which substitutes \( P_{\text{sub}} \) under conditions of a finite supraglottal pressure drop, \( P_{\text{sup}} \).

Extending this formula to the sound pressure level \( \text{SPL} \) we arrive at

\[
\text{SPL} = K + 20 \log_{10} \left( P_{\text{tr}} 1.6 x_n 1.85 \left[ (1-x_n) 2 + x_n 2/Q^2 \right]^{-0.5} \right) \tag{5}
\]

Figure 4. Illustrating the predictability of SPL in a two-word phrase from \( P_{\text{sub}} \) and \( F_0 \), and the prediction of \( P_{\text{sub}} = P_{\text{tr}} \) from \( F_0 \) and SPL. Predicted values are marked by circles.

Here we have included an additional exponent of 0.5 in the \( F_0 \) dependency at increasing pulse rates. The theoretical motivation is that a doubling of the number of voice periods in a given time would account for a 3 dB increase. However other factors may come into play.

Also, results from our prediction experiments suggest a better fit to experimental data with the use of a higher \( P_{\text{tr}} \) exponent, i.e. \( P_{\text{tr}}^{1.6} \). This value is identical to what has been reported in both [9] and [10].

An interesting property of Equation 5 is that it allows for solving \( P_{\text{tr}} \) given \( F_0 \) and SPL.

\[
P_{\text{tr}} = \left\{ 10^{\left( \frac{\text{SPL} - K}{20} \right)} x_n 1.85 \left[ (1-x_n) 2 + x_n 2/Q^2 \right]^{0.5} \right\}^{1.6} \tag{6}
\]

These equations have been successfully tested in our material. Figure 4 shows the two word Swedish sentence “Ja adjö” [j\(:\) a’j \(\check{o}\):]. In the lower part the predicted SPL is included as small circles around the measured SPL. There is a close match except at the word boundary. Observe that the SPL measure saturates in the \( F_0 \) peak region of the final vowel in a falling branch of \( P_{\text{sub}} \). As indicated, there was also a fairly close match between predicted and measured \( P_{\text{sub}} \). The \( F_0^{\text{ref}} \) of the subject was 130 Hz.

A more drastic example is shown in Figure 5. It illustrates a Swedish sentence: “å Maria Lenár igen”, “and Maria Lenár again” [o \( m \) a \( r \) i \( a \) l e ‘n \( \check{a} \) ri \( j \) e \( n \)] with high prominence of the second syllable of “Lenár”, spoken by a female subject. Her estimated \( F_0^{\text{ref}} \) was 220 Hz.

The main \( F_0 \) peak occurs in the middle of the long stressed vowel [\( a \)], which happens to coincide with a deep minimum in SPL. Could it be inferred from \( P_{\text{sub}} \)? In the lack of actual data we have resorted to a prediction from Eq. 6. The result is included in the lower part of Figure 5 together with the \( F_0 \) contour. Indeed, it shows a pronounced minimum of \( P_{\text{sub}} \) in the \( F_0 \) peak region, but other details are more uncertain.

However, we have access to the same sentence uttered by our reference male subject with a \( P_{\text{sub}} \) trace included. Figure 6 provides a view of sampled data of the recorded and predicted \( P_{\text{tr}} = P_{\text{sub}} \). A pattern similar to that in Figure 5 appears with a pronounced \( P_{\text{sub}} \) minimum in the vowel [\( a \)], but less radical.
The prediction preserves essentials, but there are deviations in some details. These could be attributed to failures in semi-closed segments of low SPL such as in the voiced consonant [j]. Also, the normal relation between the source parameter Ee and SPL is upset when the voice fundamental carries the main part of SPL.

Anyhow, our prediction rules have enabled us to explain an instance of a rather extreme pattern of a local dip in SPL within an F0 peak area. It appears to function as a part of a prosodic boundary.

The variability of temporal synchrony between F0 and Psub is a source of difficulty in deriving regression equations. Locally, in the sequence of an unstressed and a moderately stressed syllable, Psub usually stays rather constant along a gradually declining slope, and the same is true of the glottal source scale factor Ee. In this context there is not time for a change in lung pressure or glottal adjustments to take place. In the range F0>F0ref and prominence RS=10-20, the average increase of Psub from unstressed to stressed conditions was merely 1 cm H2O.

Here we have noted an increase in Ee by 3 dB, in SPL 4 dB, in SPLH 7 dB and in F0 peaks 4 semitones [7]. This accounts for SPL increasing 1 dB per semitone in F0. In the upper frequency range, F0>F0ref and RS>22, SPL and also SPLH tend to saturate, while F0 continues to increase. There is an increase of Psub with voice effort, but timed earlier than a major F0 peak.

A major conclusion is that an elevated Psub is required for higher degrees of prominence only, and its timing is adjusted to fit a breath group as well as prosodic groupings.

### 4. Acknowledgements

Johan Liljencrants and Kjell Gustafson have provided valuable support in the data processing.

### 5. References


**Figure 6. Measured and predicted Ptr. Male subject SH [omarialenə rijen].**

But how can the pronounced dip in Psub of the two subjects be explained phonetically? It is evidently a manifestation of the termination of a focally emphasized syllable at a prosodic boundary, here before a single word adverbial phrase (“igen” = “again”). It includes a combination of final lengthening and subglottal pressure drop. In our experience this pattern may occur in long highly emphasized vowels only, and functions as a prosodic boundary signal.

Accordingly, in the same sentence spoken with the stress accent placed on the first syllable of the test word “Lenar”, i.e. [lénar], the Psub contour did not show a minimum.

A quite general observation from our collected data, valid for both short and long vowels, is that Psub is raised well in advance of a stressed word to be available at the onset of the stressed syllable where it starts to decay, in part due to air consumption.

### 3. Summary and discussion

A main part of our study has been concerned with the co-variation of true subglottal pressure, F0 and intensity. Our data support the basic function of Psub as an energy supply, which SPL, Psub and duration cease to increase with F0. This is contrary to singing where Psub continues to rise in the whole F0 range [10].

A novel contribution to speech production theory is the derivation of rules for predicting, within a constant, the SPL contour from F0 and Psub, Eq. 5, and conversely for predicting Psub from SPL and F0, Eq. 6. These rules are subject to some limitations due to phonetic context. Complete articulatory closure implies a high Psub equal to Psup. At finite degrees of opening as in [j] and other voiced continuants the SPL is governed more by the voice fundamental than by the source amplitude parameter Ee, which impedes the prediction.

A pertinent question is to what extent the co-variation rules we have derived are specific to our subject and to speaking style. This remains to be found.