COVARIATION OF SUBGLOTTAL PRESSURE, F_0 AND GLOTTAL PARAMETERS

Gunnar Fant, Stellan Hertegård, Anita Kruckenberg and Johan Liljencrants*
Dept. of Speech, Music and Hearing, KTH, Stockholm, S-10044, E-mail gunnar@speech.kth.se
*Huddinge University Hospital Stockholm

Abstract

Sub- and supraglottal pressures, P_{sub} and P_{sup}, have been recorded during glissando phonation of sustained vowels, isolated vowels, and continuous speech including a one minute long reading of a novel. Studies of the covarition of P_{sub}, F_0, voice excitation amplitude E_e and overall sound pressure level reveal systematic relations some of which can be expressed in closed form by regression equations.

Systematic differences in P_{sub} with respect to position in a breathgroup, vowel and consonant category and the degree of stress have been observed. The domain of subglottal increase with stress is of the order of one or a few words rather than a single syllable.

The global contour of P_{sub} within a breathgroup and the local finestructures of P_{sub} and transglottal pressure, $P_{tr}=P_{sub}-P_{sup}$, associated with specific articulatory events are described and discussed.

INTRODUCTION

Our study is intended as a contribution to a more integrated view of speech production. One part is directed to the covariation of voice source parameters, sound pressure level, F_0, and sub- and supraglottal pressures in continuous speech. According to Fant, [4] and Fant et al. [7] these relations have to be modelled differently in the lower and in the upper part of a speaker's F_0-range indicating systematic differences in the phonatory mechanism.

Another object of our study is the temporal contour of subglottal pressure within a phrase or a breathgroup, its onset and offset characteristics, the overall declination and local variations. From the earlier work of Ladefoged [10,11], Atkinson [1], Ohala [12,13] we know that the domain of raised subglottal pressure in connection with contrastive stress is larger than a single syllable and can span over more than a single word.

A specific object of our study will be to assess the magnitude of raised subglottal pressure in relation to the degree of emphasis. An apparent rise can be expected in contrastive and focal accent but what about stress of a lower degree?

Segment specific values of subglottal pressure exist. To what extent are they passively induced and to what extent do they reflect specific pulmonary commands? A fine stucture of more rapid variations of subglottal pressure reflects passive, articulatory induced, interaction effects [12,13].

In order to promote the interpretation of our accumulated data on sub- and supraglottal pressure contours we are now in the process of applying equivalent network modelling of the respiratory system.

DATA COLLECTION

One of the authors, SH, served as the subject for the recording of sub- and supraglottal pressures during the reading of isolated vowels, specially composed test sentences and a paragraph from a novel. In addition, the subject performed glissando phonations of sustained vowels in which F_0 varied continuously in a range from 65 to 250 Hz. Subglottal pressure, P_{sub}, was measured through a tracheal puncturing probe and supraglottal pressure, P_{sup}, by means of a sond inserted through the nasal pathways. In order to optimise the size of superimposed voice ripple the P_{sub}- and P_{sup} time functions were lowpass filtered at 75 Hz.

In order to create a tie between observed acoustic and aerodynamic data and prosodic marking we have performed listening test providing continuously scaled assessments of the degree of prominence of each of the 213 syllables of the prose reading.

RESULTS

Glissando phonations and isolated vowels.

Basic parameter relations.

The purpose of these experiments was to study the covariation of F_0,, P_{sup}, E_e and the overall sound pressure level (SPL). E_e is the amplitude of the negative peak of the differentiated glottal flow and serves as a scale factor of formant excitation.

Fig. 1 shows E_e and P_{sub} as a function of F_0 from four glissando phonations differing in vocal effort and temporal contour. These data confirm our earlier observations of the existence of a critical voice fundamental frequency F_{0r} in the middle of a speakers available F_0-range where the F_0-dependency of E_e changes, in Fig.1 to be found in the region of 130 Hz. If P_{sub} at increasing F_0 decreases, or is level and low, there is a maximum of E_e at F_{0r}. Otherwise F_{0r} appears as a breaking point where E_e saturates or shows a less steep increase with F_0.

The presence of the F_{0r} was first observed in the glissando experiments of Fant [2] and have been confirmed to exist also in sampled data of E_e versus F_0 in connected speech [5-7]. For female voices we found F_{0r} values around 220 Hz.

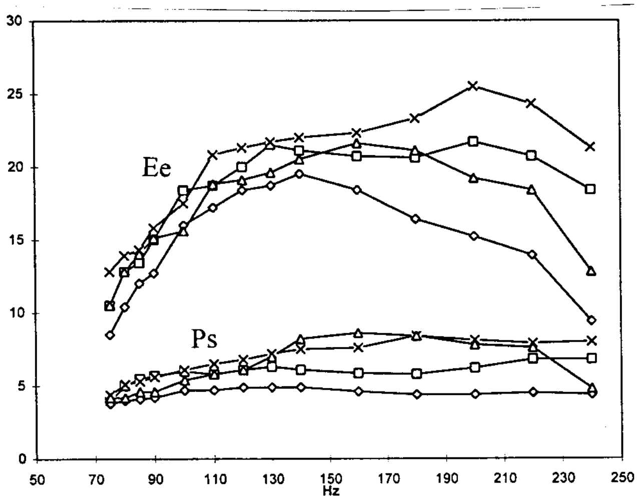

Fig. 1. Glottal excitation amplitude (E_e) and supraglottal pressure (P_s) as a function of F_0 during four glissando phonations.

From the data of Fig. 1 we have sampled E_e at constant F_0 and varying P_{sub} as well as at constant P_{sub} and varying F_0. A correlational analysis showed that in the range of $F_0<F_{0r}$, E_e was proportional to $P_{sub}^{1.1}$ and to $F_0^{1.35}$ and that P_{sub} increased in proportion to $F_0^{0.7}$. Accordingly, eliminating P_s, the result is that E_e varies in proportion to $F_0^{2.1}$ which is of the same or slightly higher order than found in continuous speech, [5].

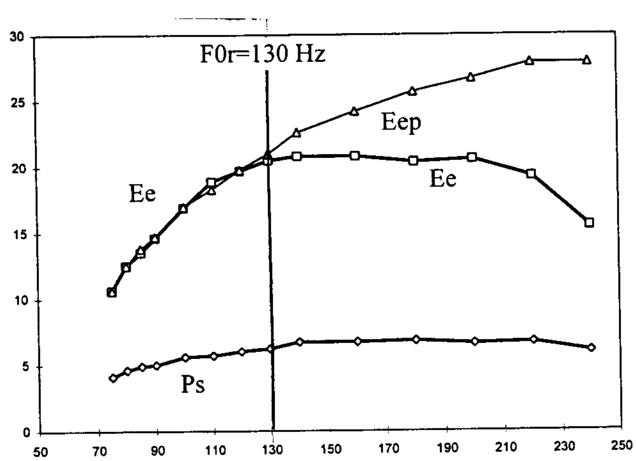

Fig. 2. Mean values of E_e and P_s from Fig. 1 as a function of F_0 together with the predicted E_{ep} derived from the region of $F_0<F_{0r}=130$ Hz.

The average data from our glissando phonations are shown in Fig. 2 together with E_{ep} predicted from the range $F_0<F_{0r}$. The breaking point F_{0r} = 130 Hz is now more clearly seen. At $F_0>F_{0r}$ E_e stays rather constant with increasing F_0 while the prediction from the $F_0<F_{0r}$ region provides an overestimate.

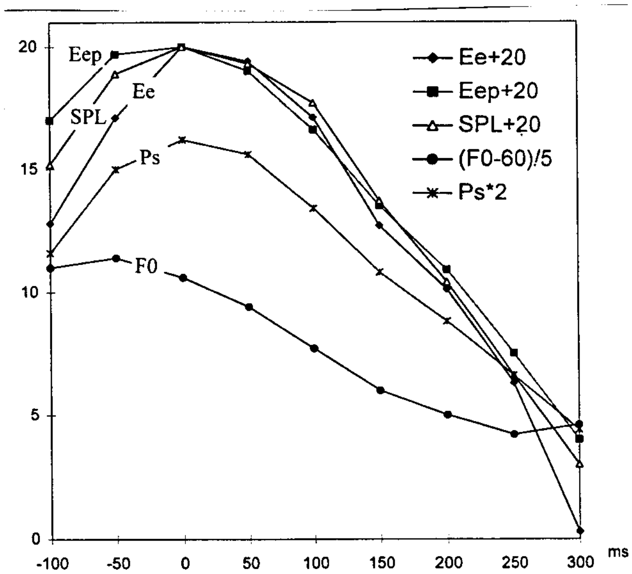

Fig. 3. Mean traces of F_0, P_s, E_e, E_{ep} and SPL within eight single vowel utterances.

The validity of the $P_{sub}^{1.1}$ $F_0^{1.35}$ proportionality of E_e has been tested in the single vowel utterances, see Fig. 3 which pertains to the mean of eight vowels. Their duration averaged 0.5 seconds. The P_{sub} as well as the E_e and SPL reached a maximum after 100 ms. The intonation was essentially falling from F_0=115 to 75 Hz. The predicted E_{ep} and the SPL closely follow E_e in the major part of the falling contour. In the onset interval of the first 100 ms E_{ep} exceeds E_e. The same is true at the termination, because of a less tense and more abducted phonation increasing the spectral tilt of the voice source.

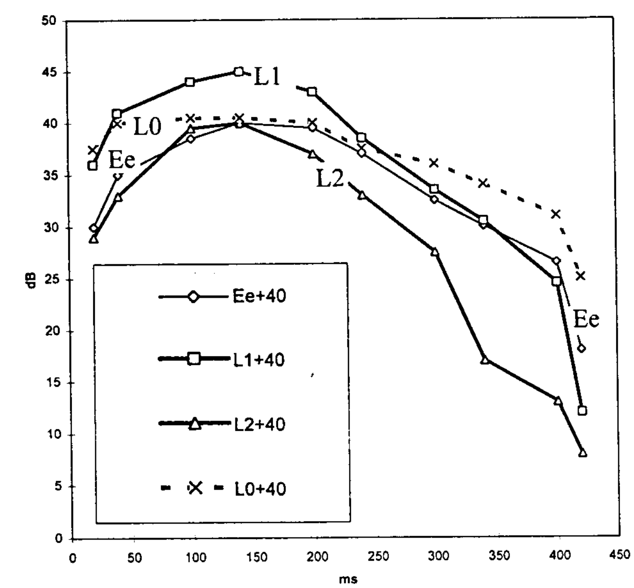

Fig. 4 (Lower right) Temporal variation of first and second formant amplitudes, L1 and L2, the amplitude L0 of the voice fundamental, and E_e within a single vowel [æ:].

This is further illustrated in Fig. 4. The relatively soft voice quality at the onset and especially at the end of the vowel [æ:] is apparent from the relative dominance of the amplitude L0 of the voice fundamental versus the amplitudes L1 and L2 of formants 1 and 2. This is a normal attribute of a breathgroup, [4,6], see also Strikes and Boves [16]. The $P_{sub}^{1.1}F_0^{1.35}$ proportionality of E_e at $F_0<F_{0r}$. has been verified in parts of our sentence material. At $F_0>F_{0r}$, as in focal accentuation, E_e reaches a limiting value or decreases with increasing F_0 [4,6]. The relations observed in sustained phonation, Fig. 1, thus appear to have a general significance.

It remains to develop a more complete model introducing the glottal area and glottal articulation as additional components. Recent high speed fiberscope photography of the vocal cords, Hertegård [9], indicate that the glottal cross-sectional area increases up to $F_0=F_{0r}$ and decreases above F_{0r} as could be expected from glottal volume velocity profile data [2].

The subglottal pressure profile

Of interest to our study is the global contour of P_{sub}, its onset, declination and termination and to infer effects of pulmonary activity as separate from those induced by glottal and supraglottal articulations.

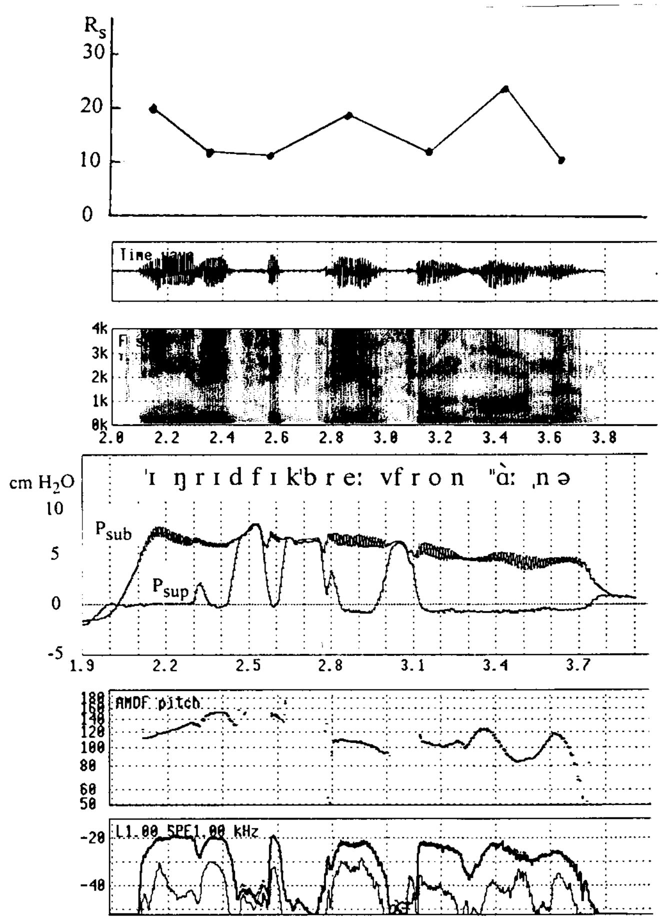

Fig. 5. A complete declarative sentence: "Ingrid fick brev från Arne" (Ingrid received a letter from Arne).

In Fig. 5 and 6 we have illustrated spoken sentences with conventional display of waveform, spectrogram, F_0 and intensity (sound pressure level SPL and high frequency weighted SPLH) to which we have added traces of P_{sub} and P_{sup} and also the perceived degree of syllable prominence, R_s.

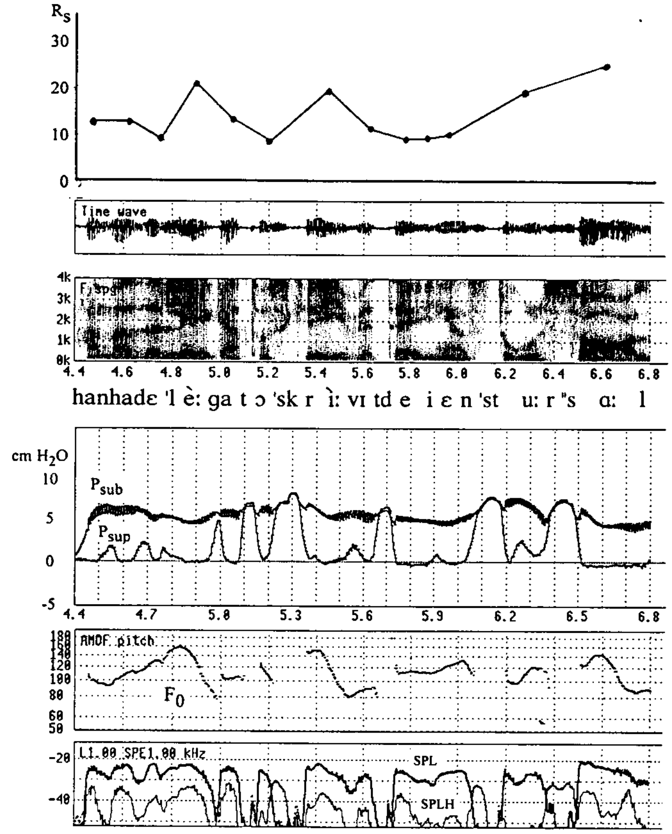

Fig. 6. Clause: "Han hade legat å skrivit de i en stor sal," (He had been lying writing it in a big hall,)

General observations from these and other utterances is that voicing starts at a transglottal pressure, $P_{tr}=P_{sub}-P_{sup}$ of about 3.5 cm H_2O but can continue down to values of the order of 1 cm H_2O in abducted voice terminations. During inhalation between breathgroups the transglottal pressure attains negative values of the order of 1-2 cm H_2O.

A typical P_{sub} contour shows a rise in the first 80-160 ms reaching a peak value averaging 6.8 cm H_2O. The onset time is of the same order of magnitude as can be seen in the data of Ladefoged [10,11], Atkinson [1], Strik and Boves [16]. At the end of a breathgroup there is a corresponding offset interval of the same duration as the onset and enhanced by glottal abduction. The declination of P_{sub} after the onset and up to the offset interval averaged 1.8 cm H_2O with a clear tendency to be independent of the duration of the breathgroup. The average duration of a breathgroup was 1.8 sec. In collecting average data of P_{sub} we therefore made corrections for the particular position within a breathgroup.

Within a breathgroup stressed vowels averaged P_{sub}=5.7 cm H_2O and unstressed vowels 5.2 cm H_2O. A similar influence of stress was found in the P_{sub} of unvoiced consonants which averaged 6.2 cm H_2O in stressed and 5.4 cm H_2O in unstressed contexts. Because of the declination an unstressed vowel in initial position was often found to have a higher P_{sub} than a stressed vowel at the end of a breathgroup. This was expected from our experience concerning E_e data in connected speech [3].

The supraglottal pressure P_{sup} was zero or very small for vowels but attained finite values in voiced consonants. In highly constricted unvoiced consonant segments P_{sup} equaled P_{sub}. As an average for [b, d and g] we noted P_{sub}=5.8 and P_{sup}=3.9 with the highest value of P_{sup}=5.0 for [b]. For the continuants [r, l and v] we noted P_{sub}=5.2 and P_{sup}=1.7 cm H_2O. In the maximally constricted interval of a long highly stressed [u:] we noted P_{sub}=6.1 and P_{sup}=2.1 cm H_2O.

A typical feature associated with an abduction is a local drop of P_{sub} as in pre-occlusion and post-release intervals of aspirated stops and more generally at a voiced/unvoiced boundary, see the dip before the final [s] in Fig. 6. A drop in P_{sub} was also found in samples of voiced [h] for which we noted P_{sub}=3.1 cm H_2O. In high vowel contexts of [h] we noted P_{sup} of the order of 1-2 cm H_2O indicating a pronounced vocal tract narrowing.

In Fig. 5 illustrating a neutral declarative sentence the P_{sub} declination contour is evenly falling. A different stress distribution is exemplified in Fig. 6. Here, the subglottal pressure is raised in the compound "legat å skrivit" (lying and writing) and at the end, "stor sal" (big hall) which attains focal accent. These are gross pulmonary gestures overriding the finestructure of individual syllables, some of them unstressed as documented in the R_s curve.

In general we can expect an increase of P_{sub} which sets in at the P-center of an inter-stress interval. The size varies from a few cm H_2O in marked emphasis down to zero values in low but finite accentuation levels. The decay of P_{sub} within the inter-stress interval is in part associated with an escape of air at constant pulmonary force at an average rate of 1 cm H_2O per second, but there is also evidence of a relaxation of the pulmonary support within an interstress interval resulting in larger drops of P_{sub}. Such drops, with time constants of the order of 150 ms, have been found in focal accentuations involving long vowel nuclei at the end of the stressed syllable and also at the termination of a phrase [7].

The fine structure within a subglottal contour varies more rapidly. It is related to variations in glottal and supraglottal flow resistance and also to the effects of vocal tract wall expansion following sudden supraglottal closure or in voicing maintained during closure. This is the cause of the higher P_{sub} in unvoiced than in voiced sounds.

We have verified the findings of Atkinson [1] that high vowels such as [i] and [u] have not only higher intrinsic F_0 than open vowels such as [æ] and [a] but also slightly higher P_{sub} (of the order of 0.5 cm H_2O in our data). This might be explained by the lower air consumption at the higher F_0 of high vowels.

These are tentative conclusions backed up by an ongoing modelling of the respiratory system inspired by the work of K.N. Stevens[15] with reference to the modelling of Rothenberg (14) and in agreement with those of Ohala [12,13]. An analysis-by-synthesis of our recorded data should enable a closer separation of linguistically meaningful control of emphasis and grouping versus passive glottal and supraglottal articulatory induced variations.

Acknowledgments

This work has been supported by grants from The Bank of Sweden Tercentenary Foundation and the Swedish Research Council for Engineering Sciences.

References

[1] Atkinson J. (1978). *Aspects of intonation in speech: Implications from an experimental study of fundamental frequency.* Ph.D. diss., University of Connecticut

[2] Fant G. (1982). "Preliminaries to the analysis of the human voice source." *STL-QPSR* 4/1982, pp. 1-27.

[3] Fant G. (1993). "Some problems in voice source analysis", *Speech Communication* Vol. 13 (1993) pp. 7-22.

[4] Fant G. (1997). "The voice source in connected speech" To be published in *Speech Communication.*

[5] Fant G. and Kruckenberg A. (1994). "Notes on stress and word accent in Swedish". *STL-QPSR* 2-3 1994, pp. 125-144.

[6] Fant G. and Kruckenberg A (1996) "Voice source properties of the speech code". *TMH-QPSR* 4/1996, pp. 45-56.

[7] Fant G., Hertegård S. and Kruckenberg A. (1996). "Focal accent and subglottal pressure". *TMH-QPSR* 2-1996, pp.29-32.

[8] 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.

[9] Hertegård S. (1997), Personal communication.

[10] Ladefoged P. (1967). *"Three Areas of Experimental Phonetics".* Oxford University Press.

[11] Ladefoged P. (1961). "Physiological studies of speech." *STL-QPSR* 3/1961, pp. 16-21.

[12] Ohala J. (1974). "A mathematical model of speech aerodynamics." Speech Communication Seminar, Stockholm, *1974, Almqvist and Wiksell*, Uppsala.

[13] 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.

[14] Rothenberg M. (1968), *The Breath Stream Dynamics of Simple-Released-Plosive Production*, Bibliotheca Phonetica No. 6, Base: S. Karger.

[15] Stevens K.N. (1997) Manuscript of a book, "Acoustic Phonetics.

[16] Strik H. and Boves L. (1992). "Control of fundamental frequency, intensity and voice quality in speech." *Journal of Phonetics*, 20, pp. 15-25.