

Some problems in voice source analysis

Gunnar Fant

*Department of Speech Communication and Music Acoustics, Royal Institute of Technology, KTH, Box 70014,
S-10044 Stockholm, Sweden*

Received 7 May 1993

Abstract. This is an overview of some recent studies of voice source acoustics and glottal flow analysis and modelling performed at the KTH. Time and frequency domain aspects of the production process are discussed with a view of relating glottal flow parameters from inverse filtering and vocal tract transfer functions to formant amplitudes and bandwidths. Alternative methods of determining the time constant $T_a = 1/(2\pi F_a)$ in the return phase of glottal flow derivative after the instant of excitation, and thus of spectral tilt, are discussed. Selective inverse filtering, removing all but one formant, is potentially useful for this purpose. The influence of uncertainties in quantifying the vocal tract transfer function is exemplified by a calculation of the effects of introducing a finite baffle effect of the human head adding a high-frequency emphasis above the standard +6 dB/octave. Particular attention has been paid to temporal variations within an utterance as derived from continuous inverse filtering. Aspects of breathy voicing and female–male differences in voice production are discussed. It is demonstrated that the temporal profile of the excitation amplitude, $E_e(t)$, within an utterance derived from a male speaker can be approximated by the envelope of the negative part of the speech wave.

Zusammenfassung. Hier wird eine Bilanz von einigen, kürzlich im KTH über die Akustik der Stimmquelle sowie über die Analyse und das Rechenmodell des entsprechenden Stimmritzendurchflusses durchgeführten Untersuchungen beschrieben. Es werden die Zeit und die Frequenz des Erzeugungsverfahrens untersucht, um die durch Gegenfilterung erhaltenen Parameter des Stimmritzendurchflusses und die Funktionen der Übertragung des Stimmkanals mit der Amplitude und der Bandbreite in Verbindung zu bringen. Es werden neue Methoden zur Bestimmung der Zeitkonstante $T_a = 1/(2\pi F_a)$ in der Rückleitungsphase der Ableitung des Stimmritzenflusses nach dem Zeitpunkt der Erregung und somit des globalen Gefälles des Stimmritzenspektrums vorgeschlagen. Eine selektive Gegenfilterung, die nur ein Gebilde hält, scheint diesbezüglich eine versprechende Methode zu sein. Der Einfluss der Ungenauigkeiten in der Quantisierung der Transferfunktionen des Stimmkanals wird anhand eines Beispiels erläutert; es wird durch Berechnung gezeigt, dass die Eingabe einer Endwirkung der Kopfdämpfung eine Hochfrequenzverstärkung zusätzlich zu der Normalverstärkung von 6 dB/Oktave ergibt. Besonders beachtet wurden die zeitlichen Variationen innerhalb einer Aussprache, wie z.B. die Ableitungen einer kontinuierlichen Gegenfilterung. Es werden die Stimmgeräusche und die Unterschiede Mann–Frau in der Stimmerzeugung untersucht. Es wird anhand einer von einem männlichen Sprecher erzeugten Aussprache gezeigt, dass das zeitliche Profil der Erregungsamplitude $E_e(t)$ durch den negativen Teil der Stimmkurve des Sprachsignals genähert werden kann.

Résumé. On présente ici un bilan de certaines études récentes menées au KTH sur l'acoustique de la source vocale et sur l'analyse et la modélisation du flux glottique associé. On examine les aspects temporels et fréquentiels du processus de production dans le but de relier les paramètres du flux glottique obtenus par filtrage inverse et les fonctions de transfert du conduit vocal à l'amplitude et à la largeur de bande des formants. On propose de nouvelles méthodes de détermination de la constante de temps $T_a = 1/(2\pi F_a)$ dans la phase de retour de la dérivée du flux glottique après l'instant d'excitation, et donc de la pente globale du spectre glottique. Un filtrage inverse sélectif, ne conservant qu'un seul formant, semble être à cet égard une méthode prometteuse. L'influence des imprécisions dans la quantification des fonctions de transfert du conduit vocal est illustrée sur un exemple: on montre par calcul que l'introduction d'un effet fini d'atténuation de la tête ajoute une amplification supplémentaire haute-fréquence à l'amplification standard de 6 dB/octave. Une attention particulière a été apportée aux variations temporelles au sein d'un énoncé, telles que dérivées d'un filtrage inverse continu. On examine les questions de voisement bruité et de différences homme–femme dans la production de la voix. On démontre que, sur un énoncé produit par un locuteur masculin, le profil temporel de l'amplitude d'excitation $E_e(t)$ peut être approché par l'enveloppe de la partie négative du signal de parole.

Keywords. Voice production theory; inverse filtering; glottal flow; voice source dynamics; source spectrum.

1. Introduction

One of the most engaging and complex problems in speech research is how to analyze and model the human voice source. There is a need for a more profound general knowledge of voice source characteristics and production mechanisms, not the least for speech synthesis purposes. This is one of the many fields in which Hiroya Fujisaki has made significant contributions (Fujisaki and Ljungqvist, 1986, 1987).

A major tool for studies of the voice source is inverse filtering of the speech wave, which is a process of cancelling the formant structure thereby regenerating a replica of the glottal volume velocity time function or its derivative. For descriptive purposes we need suitable models to approximate these time functions with a few parameters.

At the KTH, we have a long tradition of activities within this field, see (Fant and Lin, 1991). Our first experiments on inverse filtering were reported by Fant (1959b). Early landmarks included the demonstration of double excitations in a hoarse voice (Briess and Fant, 1962), and photoelectric glottography combined with inverse filtering (Fant and Sonesson, 1962). Lindqvist-Gauffin (1964, 1965, 1970) studied glottal flow waveforms within a phonatory frame of varying F_0 and SPL which was followed up by more extensive studies of Sundberg and Gauffin (1979). Experiments on continuous inverse filtering and theory of voice source modelling have been reported in a series of articles by Fant (1979a, 1979b, 1980, 1982a, 1982b, 1986). A report of Mártony (1965) on voice source spectra has a renewed interest in the light of recent reports on frequency domain analysis techniques (Fant and Lin, 1988, 1991).

Other important contributions to inverse filtering techniques and voice source studies are those of Rothenberg (1973, 1983). More complex aspects of acoustic-aerodynamic interaction were first treated by Ananthapadmanabha and Fant (1982), and followed up by Fant et al. (1985b), Fant (1986), Fant and Lin (1987, 1988) and Lin (1990).

A three-parameter model of vocal flow allowing a variation of open quotient and pulse sym-

metry was introduced by Fant (1979a, 1979b, 1980). It was applied to studies of source filter characteristics of connected speech with discussions of temporal variations including glottal abduction in voiced aspiration.

A major advance was the LF-model (Fant et al., 1985a), which is now widely used. The novelty was the introduction of a gradual return phase after the flow discontinuity at closure. This extension proved to be of greater perceptual importance than any of the traditional glottal flow pulse shape parameters. A variant of the LF-model is that of Ananthapadmanabha (1984) who suggested a parabolic instead of an exponential return phase. A review of voice source models appears in (Fujisaki and Ljungqvist, 1986).

Applied descriptive work with the LF-model has been reported in (Gobl, 1988; Karlsson, 1990, 1991; Gobl and Karlsson, 1989). More general problems of voice source modeling with special reference to breathy phonation have been treated in (Gobl and Ní Chasaide, 1988; Fant and Lin, 1988). Voice source rules for speech synthesis have been discussed in (Carlson et al., 1989, 1991).

The purpose of the present article is to summarize basic theory and results from our recent work on voice source analysis and modelling. One problem is how to define source and filter characteristics and to describe their interactions. This is generally a matter of choosing appropriate approximations for the particular purpose of analysis, e.g., inverse filtering. How do we optimally combine time-domain and frequency-domain analyses for our purposes? What is the relation between what goes on inside the vocal tract to what comes out as a speech wave? In particular, how are amplitude levels of formants and of the voice fundamental related to vocal transfer functions and voice source characteristics? How do these relations explain some of the male-female differences? Can we infer the time-domain properties of glottal flow from spectral data of the speech wave alone?

A few principal and technical matters of inverse filtering will be dealt with. One is the determination of the time constant T_a of the return phase at glottal closure, which has a profound influence on the spectral tilt of the source

but is hard to measure and is sensitive to many factors related to the overall frequency response in speech production as well as to the particular inverse filtering set-up. In this connection, the possible role of the baffle effect of the speaker's head and body will be discussed. Novel time-domain methods for determining the return phase time constant T_a will be discussed in relation to conventional methods.

We are badly in need of more data on voice source dynamics, e.g., concerning the initiation and termination of voicing and the temporal variations in boundary regions between voiced and unvoiced segments. An inhibiting factor for advance is the lack of satisfactory methods for continuous and automatic extraction of voice source parameters. One attempt to solve this problem is that of Strik and Boves (1991). Meanwhile, as will be discussed, hand-edited continuous inverse filtering can provide important insights with suggestions for routine analysis.

2. Source filter decomposition and interaction

There are three stages of sophistication in voice production theory and in voice source modelling. The simplest one is to employ a parameterized model of glottal flow input to the supraglottal vocal tract, thereby by definition excluding the influence of the subglottal system. A more complete functional model includes finite specifications of the subglottal system and/or the time-varying glottal impedance in which the main control variables are the time-varying glottal area function and the lung pressure. The most sophisticated approach is to apply a self-oscillating model of the vocal folds with appropriate mechanical-aerodynamic coupling and control parameters related to vocal fold tension and degree of abduction. This is a useful component of an articulatory synthesizer, but the model lacks explanatory power unless an analysis is performed of the mechanical, aerodynamic and acoustic consequences including glottal area function and flow (Flanagan et al., 1975; Bickley, 1991; Liljencrants, 1991).

There are several sources of interaction to keep in mind. One is that of the influence of the

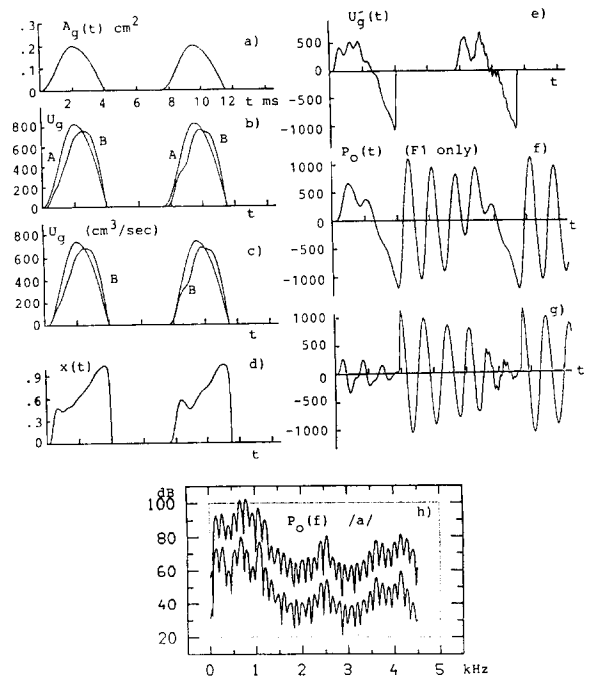


Fig. 1. Acoustic interaction in voice production, vowel [a]. Left column: (a) glottal area function, (b) glottal flow, A without and B with glottal plus supraglottal load, (c) the same as (b) but without the glottal load, (d) glottal air particle velocity profile. Right column: (e) glottal flow derivative for the first two pulses, (f) F_1 only excited, (g) F_1 -oscillation isolated. Below: (h) Spectrum of three successive voice periods, top interactive, bottom non-interactive.

supraglottal load, which is enhanced by supraglottal constrictions, on the vibratory pattern of the vocal folds and, thereby, on the shape of the glottal area function (Bickley and Stevens, 1986; Stevens, 1987). Another source of interaction is the nonlinear transform from glottal area function to glottal flow and the influence of formant oscillations superimposed on the transglottal pressure at the instance of excitation.

Because of these interactions, the true glottal flow as revealed by inverse filtering is more complex than the underlying glottal area function. The flow is skewed to the right and the positive going part of the flow derivative may display multiple peaks (Fant et al., 1985b, Fant and Lin, 1987). This is illustrated in Figure 1 from (Fant and Lin, 1988), which shows successive stages in a time-domain analysis of voice production, glottal area, particle velocity, glottal flow and flow derivative and the excitation of a single formant.

These model experiments predict typical patterns of human speech, see Figure 2, which shows the succession of glottal flow derivative periods in the final syllable [jø:] of the Swedish word "adjö", [aj'ø:]. The dual peak pattern in the first part of the [ø] is apparent, while the final part of the vowel displays a smoother shape as in standard models. The double peak can be explained as the influence of an F_1 oscillatory component evoked in the previous voice cycle which adds to the supraglottal pressure in the glottal opening phase, producing a minimum in the transglottal pressure and, thus, in the positive part of the glottal flow derivative. The distance between the peaks is approximately half a period of F_1 . The spectral consequence is a zero at about $2F_1$, which is confirmed by both modelling and inverse filtering experiments. The appearance of the double peak or other ripple components depends on the particular phase of the superimposed F_1 oscillation and the particular voice mode. The residual F_1 can also affect the flow derivative at closure and thus the excitation amplitude. The perceptual importance of these kinds of perturbations is probably not very great (Nord et al., 1986), but they add to the very complex random fluctuations

from one voice period to the next, which constitutes an element of naturalness of human speech.

An extreme case of superposition and nonlinear interaction is when F_1 and F_0 coincide in frequency which has the consequence of minimizing the transglottal pressure in the main part of the glottal open period while maintaining a high value of flow derivative at the closing edge. Air consumption is minimized while formant excitation is retained (Fant, 1986; Schutte and Miller, 1986). This is apparently a mechanism adding to the importance of the F_1F_0 mutual gain factor in soprano singing discussed by Sundberg (1973), but it also has implications for studies of voice production in female speech that has not been fully exploited (Fant and Lin, 1988).

Inverse filtering does not necessarily portray the true glottal flow. This ideal is attained under the condition of a tuning appropriate for complete glottal closure only. In routine practice, the inverse filter is adjusted for maximum formant cancellation in the presumed maximally closed glottal phase which may have a finite opening allowing some degree of subglottal coupling. This setting will cause errors in the reconstruction of the true glottal flow. On the other hand, if the

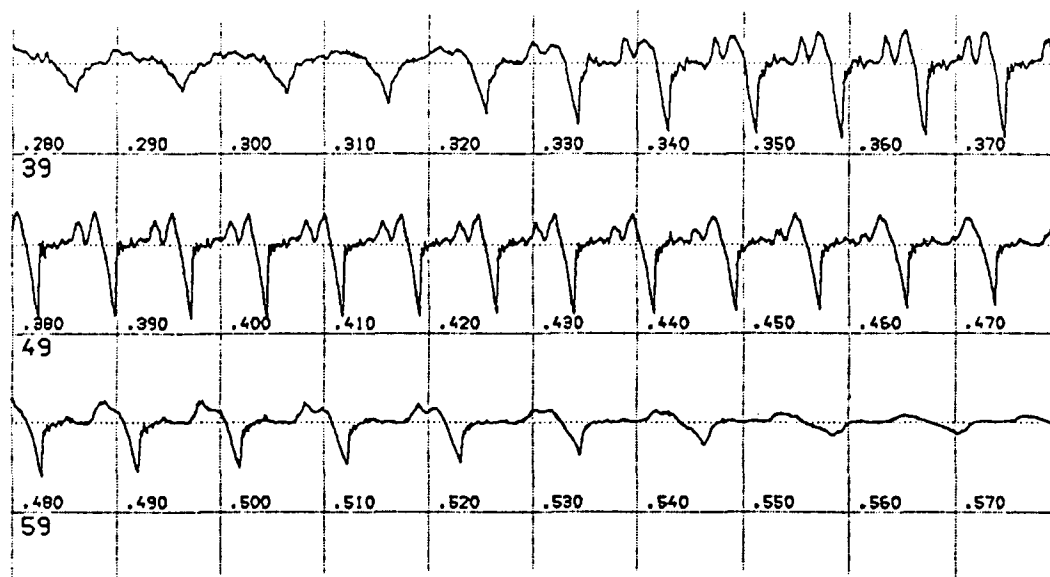


Fig. 2. Continuous inverse filtering of the syllable [jø:] from the word "adjö", [aj'ø:]. The first part of the top row is the consonant [j].

inverse filter is tuned according to a correct assumption about the supraglottal transfer function, there will appear residual formant oscillations originating from incomplete cancellation of true VT modes (Lin, 1990). In other words, the true glottal flow always contains components related to uncompensated VT modes that depend on the time-varying overall system function within a glottal period. This is partially a nonlinear process. The setting for maximal cancellation in the maximally closed phase is generally preferred since it provides a more pragmatic base for formant synthesis.

The occurrence of pole-zero pairs in the VT transfer functions of nasals, nasalized voiced sounds and in laterals is well established from acoustic theory (Fant, 1960) and has been taken into account in inverse filtering, see, e.g., (Fujisaki and Ljungqvist, 1987; Karlsson, 1991). Subglottal coupling may also manifest as extra formants that complicate the spectral pattern (Fant et al., 1972; Fant and Lin, 1988; Klatt and Klatt, 1990). This is frequently the case for unvoiced aspiration after a stop release but may also penetrate into the following vowel (Fant and Lin, 1988).

In addition, we have to consider distributed zeros entering via the source spectrum, e.g., due to the finite duration properties of glottal pulses. A related problem has to do with the time-variable decay constant and frequency of formant oscillations in the open part of the glottal cycle. The tendency towards truncation of formant oscillations at the beginning of the open phase results in the well-known $\sin x/x$ pattern of distributed zeros (Fant and Ananthapadmanabha, 1982). These are rightly properties of the transfer function but will appear as components of the source spectrum after inverse filtering. In addition, we have all the voice source interaction phenomena to consider.

The general conclusion is that there are many sources of irregularities of voice source spectra. However, the situation is less complex if we apply windows of analysis substantially shorter than a voice fundamental period permitting a separation of the glottal maximally open phase from the glottal maximally closed phase, allowing a separation of source residue waveforms from formant oscillations. This very short-time analysis is con-

ceptually close to a pure time-domain analysis but has important ties to general frequency-domain descriptions.

In summary, the theory of voice production is complicated by several interaction phenomena. The perceptual importance of some of these may be marginal (Nord et al., 1986), while others have an important role in determining dynamic properties of continuous speech. In the following section, the emphasis will be on a basic and more approximate source-filter theory that can be described by relatively straight-forward analytical tools.

3. The LF-model

The LF-model (Fant et al., 1985a), see Figure 3, provides a practical idealization of the time

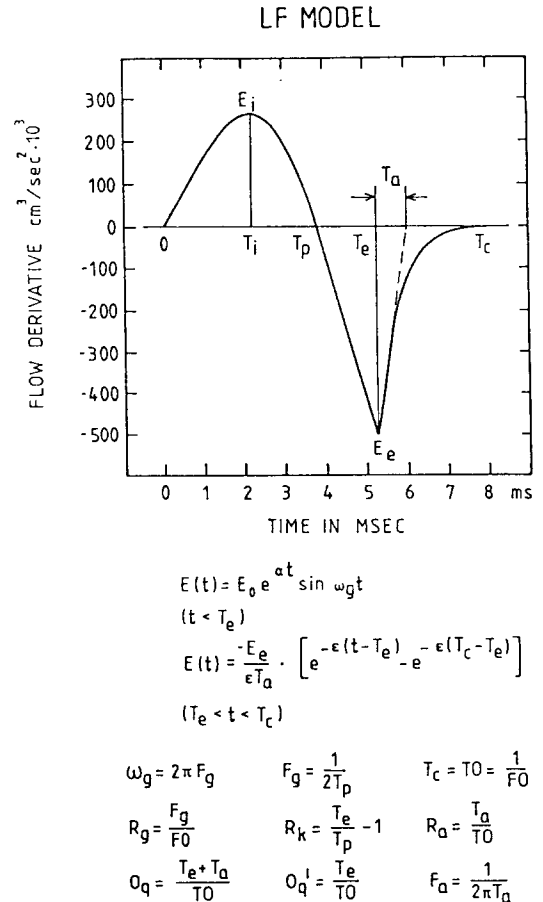


Fig. 3. The LF-model.

derivative of the glottal volume velocity flow with reference to critical events in the underlying glottal flow waveshape.

The main body of the flow derivative is modeled as a sinusoid of exponentially expanding amplitude, starting at $t = 0$, reaching a maximum value E_i at $t = T_i$, passing through zero at $t = T_p$, continuing to negative values and abruptly returning to zero after having reached a value of $-E_e$ at a time T_e . E_e is the most important parameter since it sets the levels of formant amplitudes.

The time event T_p corresponds to the location of the peak value U_o of the glottal flow and T_i to the location of the inflection point of the rising branch of the flow. The main part of the flow pulse is terminated at the time T_e . This is the instance of main excitation within the glottal period. The following return path towards zero amplitude is modelled as an exponential which is merely a corner effect in the flow profile but a more apparent feature in the flow derivative. The projection of the derivative of the return phase at time T_e on the zero line is labelled T_a . The greater T_a , the more pronounced is the spectrum role off in excess of the ideal -6 dB/octave of the flow derivative or of the -12 dB/octave flow reference. The effect of the return branch is

equivalent to that of a low-pass filter of the first order with a cut-off frequency

$$F_a = 1/(2\pi T_a). \quad (1)$$

The parameters $R_g = T_0/2T_p = F_g/F_0$ and $R_k = (T_e - T_p)/T_p$, see Figure 3, define the main pulse shape. Increasing R_g implies a shortening of the rise time of the flow pulse in relation to the total duration T_0 of the glottal cycle, and decreasing R_k implies a shortening of the decay time of glottal flow versus the rise time. If the decrease of R_k occurs at constant maximum flow, U_o , the increased steepness of the falling branch will result in an increase of the excitation parameter E_e and, thus, of the overall spectrum level.

The open quotient increases with decreasing R_g and increasing R_k . It can be defined either as

$$O_q = (T_e + T_a)/T_0 = (1 + R_k)/2R_g + T_a, \quad (2)$$

which conforms with conventional concepts or simply as

$$O_q' = T_e/T_a = (1 + R_k)/2R_g. \quad (3)$$

The pulse shape parameters R_g and R_k have a local influence on the low frequency part of the spectrum. Decreasing R_k at constant E_e will

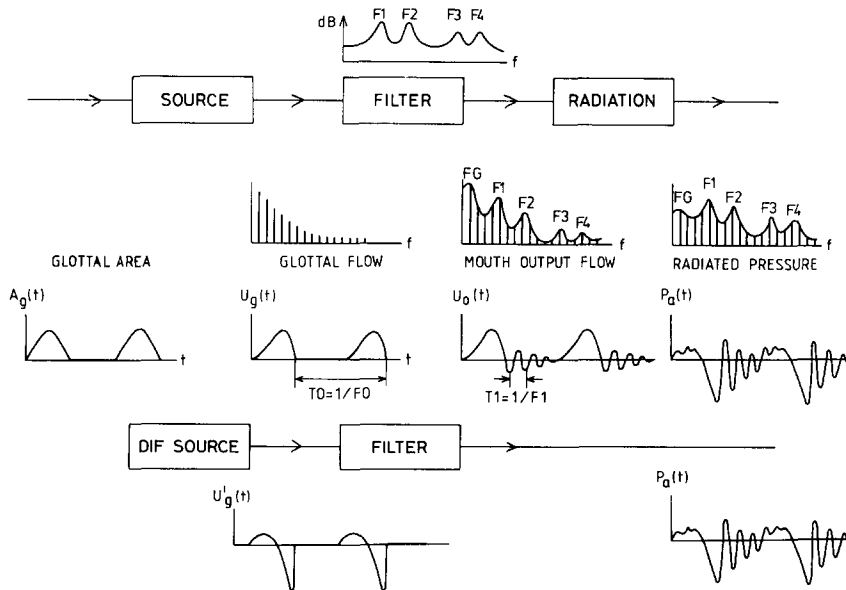


Fig. 4. General frequency-domain and time-domain illustrations of voice production.

reduce the amplitude of the voice fundamental thus enhancing the second and possibly the third harmonic as in a pressed male voice. A similar effect is attained with increasing R_g . In general, R_k and R_g are of less perceptual importance than the return phase parameter T_a or its equivalent F_a .

4. From vocal tract to sound. Basic theory

The understanding of voice production is enhanced by combining frequency- and time-domain models of source and vocal tract transfer functions to determine the main parameters of the radiated speech wave. A conceptual presentation is provided in Figure 4 as a general frame for the following analysis.

Consider a transform representation of the radiated sound as

$$P(s) = G(s)H(s)R(s), \quad (4)$$

where $G(s)$ is the source, $H(s)$ is the vocal tract transfer function and $R(s)$ is the radiation transfer.

The absolute value of the radiation transfer is

$$R(\omega) = K_T(\omega)(\rho\omega/4\pi a), \quad (5)$$

where ρ is the density of air and a is the distance from the speakers mouth. $K_T(\omega)$ represents a radiation transfer in excess of the basic +6 dB/octave and is usually disregarded.

The amplitude of the voice fundamental A_0 of a voiced sound has a close relation to the peak amplitude U_o of the glottal volume velocity pulse train. A derivation of an analytical relation starts by noting that the amplitude of the fundamental in a Fourier series expansion of glottal flow is approximately $U_o/2$. When transferred to the radiated sound by (4) with $K_T(\omega) = 1$ we find

$$A_o = U_o k \pi F_0 H(F_0)(\rho/4\pi a), \quad (6)$$

in which the constant k is close to 1 for commonly encountered glottal waveforms and open quotients of the order of $O_q = 0.6$ – 0.8 . For $O_q = 0.3$, typical of a low F_0 pressed male voice, $k = 0.7$. It should be observed that A_0 is proportional to the voice fundamental frequency F_0 and obtains

a boost from the vocal tract transfer function, $H(F_0)$, which can be ignored if F_1 is well above F_0 .

Basic relations between formant amplitudes and source characteristics can be derived by considering the excitation to be an abrupt positive-going step function of amplitude E_e in the time derivative of the glottal volume velocity flow. This is a feature inherent in all source models and accounts for an overall -12 dB/octave slope of the flow spectrum, i.e., -6 dB/octave in the flow derivative. The initial amplitude A_i of the evoked damped oscillation, assuming a one-formant vocal tract transfer function with F_n well above F_0 , is derived from an inverse transform

$$A_i = E_e(\rho/4\pi a). \quad (7)$$

This is a base formula that needs to be expanded to take into account at least three major factors. One factor, $K_n(f)$, corrects for the presence of more than one vocal tract mode. The effect of a finite return time, T_a , of the glottal pulse at closure is represented by the factor $K_a(f)$. The third factor, $K_T(f)$, is the already mentioned baffle effect, see (5).

$$A_{in} = k_n(f)k_a(f)K_T(f)(\rho/4\pi a). \quad (8)$$

One important feature is that the initial amplitudes, A_i , are independent of the bandwidth. If a formant is not too close to F_0 , its envelope peak in a harmonic spectrum has an amplitude of

$$A_{sn} = A_{in}F_0/(\pi B_n), \quad (9)$$

where B_n is the formant bandwidth.

The spectrum envelope peak measure, A_{sn} , is a practical reference for formant amplitude measurements. An alternative is to perform an rms summation of individual harmonics within the formant. Both methods become uncertain at high F_0 values when the spectrum is less well defined. The specific mode of superposition from previous periods is important (Fant et al., 1963). An approximate expression from Fant (1959a) relating the rms value, A_e , to the spectrum peak amplitude, A_s , is

$$A_e/A_s = [(1 - e^{-2y})y/2]^{1/2}(1 - e^{-y})^{-1}, \quad (10)$$

where $y = \pi B_n / F_0$. This ratio approaches 1 for small y -values. It is 1.04 for $y = 1$, 1.15 for $y = 2$ and 1.44 for $y = 4$.

In a more general analysis we have to consider a system function with an arbitrary number of resonant modes. The Laplace transform of an allpole VT transfer function with an added factor $1/s$ for the combined effect of source and radiation is

$$P_1(s) = \frac{1}{s} \prod_1^{\infty} \omega_{on}^2 / \left[(s + a_n)^2 + \omega_n^2 \right]^{-1}. \quad (11)$$

The inverse transform is the corresponding time function

$$p(t) = \sum_1^{\infty} A_{in} e^{-\pi B_n t} \cos(\omega_n t + \varphi_n). \quad (12)$$

With a finite number of r formants and correction K_{rr} for poles higher than r , we may write

$$A_{in} = \prod_{\substack{m=r \\ m \neq n}}^{m=r} \left[1 - F_n^2 / F_m^2 \right]^{-1} K_{rr}, \quad (13)$$

$$K_{rr} = \prod_{m=r+1}^{m=\infty} \left[1 - F_n^2 / F_m^2 \right]^{-1}, \quad (14)$$

where K_{rr} is the correction for poles higher than $n = r$. For a four-formant case (Fant, 1959a, 1960)

$$20 \log_{10} K_{r4} = 0.54 x_1^2 + 0.00143 x_1^4, \quad (15)$$

$$x_1 = f / F_1 = F_n / F_1 = 4l / c. \quad (16)$$

A special case is the single tube ideal neutral vowel with formants at $F_n = (2n - 1)F_1$ and $F_1 = c / 4l = 500$ Hz for a total length of $l = 17.5$ cm. Its transfer function derived from transmission line theory with the $1/s$ factor for source plus radiation added is

$$P(s) = (1/s) [\cosh(sl/c + \alpha l)]^{-1}. \quad (17)$$

The corresponding time function is identical in form to that of (12). The initial amplitudes are

$$A_{in} = (4/\pi) (-1)^n (F_1 / F_n), \quad (18)$$

and the bandwidths are

$$B_n = \alpha c / \pi. \quad (19)$$

The initial amplitudes A_{in} are apparently inversely proportional to their frequencies F_n . In

the transform equations (13) and (14), this inverse proportionality is preserved by the factors relating formant amplitudes to formant frequencies. Another consequence of the identity of equations (13) and (18) is that the amplitude of the first formant, A_{i1} , when acting together with all other formants, is a factor $4/\pi$ larger than if it were excited alone. The $4/\pi$ factor is identical to the higher pole correction (14) at F_1 for formants above F_1 . Seemingly, this difference would upset the temporal continuity of source function and transient response at the instance of excitation. However, the continuity is restored by the alternating sign property of all superimposed formant oscillations, the sum of which in the neutral vowel constitutes an exponentially damped square wave of the base frequency F_1 and $F_2 = 3F_1$, $F_3 = 5f_1$, etc.

The $1/s$ approximation of the source + radiation transform is not sufficient for cases where F_1 comes close to F_0 or rather to $F_g = 1/2T_p$. In this case, there is mutual reinforcement. The amplitude of the first formant gains a support from the relative prominent low-frequency part of the source spectrum which can be said to possess an " R_g " formant, see the more detailed derivation in (Fant, 1979a, 1980). This situation is of special importance in female speech.

5. Estimation of the return time constant

In the case of a single formant transfer function, there is continuity at the point of excitation, T_e , between the excitation amplitude, $-E_e$, and the initial value, $-A_i$, of the evoked minus cosine formant oscillation. This is the situation encountered in inverse filtering of speech, when all formants but one have been cancelled. What happens with a finite return time constant T_a of the glottal pulse? An analysis of the transient response reveals that the onset of the oscillation is delayed versus T_e and that the initial amplitude, A_i , is reduced. The reduction is the same as the insertion loss at a frequency, F_n , of an equivalent low-pass filter of cut-off frequency $F_a = 1/(2\pi T_a)$.

$$K_a = A_i / E_e = \left[1 + F_n^2 / F_a^2 \right]^{-1/2}. \quad (20)$$

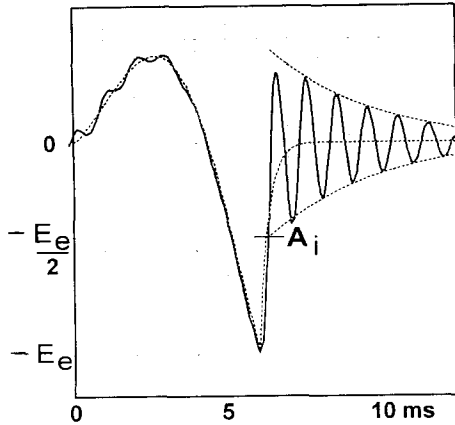


Fig. 5. The time-domain equivalence of a single formant of frequency $F_n = 1000$ Hz and bandwidth $B_n = 100$ Hz excited by an LF-pulse of $F_a = F_n/2 = 500$ Hz. In accordance with (20) $A_i/E_e = 0.45$.

Equation (20) suggests an alternative method for determining F_a as

$$F_a = F_n [E_e^2/A_i^2 - 1]^{-1/2}. \quad (21)$$

In practice, this operation can be performed after cancelling all formants but one, e.g., F_1 or F_2 or F_3 . One can also cancel all formants and insert a probe formant of suitable frequency and bandwidth and apply (21). Here follows a list of corresponding values:

F_n/F_a	A_i/E_e
0.5	0.90
0.7	0.82
1.0	0.71
2.0	0.45
5.0	0.20
10.0	0.10

The case of $F_n/F_a = 2$ and $A_i/E_e = 0.45$ is illustrated in Figure 5. ($F_a = 500$ Hz, $F_n = 1000$ Hz and $B_n = 100$ Hz.)

A more practical method is to add a real zero as a compensation filter inverse to the T_a filter (20) and set its cut-off frequency for best cancellation which implies a step function return without the probe formant, and a continuity of $A_i = E_e$ when the probe formant is applied. The accuracy is of the order of 20%. One can also attempt to vary the first order T_a compensation filter for a best match over a certain frequency domain. All

these methods are convenient and can be used as alternatives to the standard routine of varying the T_a of a superimposed LF-model for a best match.

It should also be kept in mind that any deviation from a perfectly flat audio response will affect the measurement of T_a . One well-known source of uncertainty lies in the particular choice of higher pole correction.

Another component that can affect the interpretation of inverse filter curves is the possible baffle effect of the speaker's head and body, i.e., the factor $K_T(\omega)$ of (5). The theoretical model that was adopted (Fant, 1960) is that of the mouth radiating as a circular piston on the surface of a sphere of radius 9 cm. This would account for a high frequency emphasis in excess of the basic +6 dB per octave of radiation which is 0.4 dB at 314 Hz, 2 dB at 625 Hz, 4 dB at 1250 Hz, 5.4 dB at 2500 Hz and 6 dB at 5000 Hz. A good approximation (Fant, 1980) is a first order one-pole one-zero filter.

$$H_b(s) = 2(s + 2\pi 600)/(s + 2\pi 1200). \quad (22)$$

Figure 6 exemplifies the influence of adding this system function to a glottal flow derivative waveform with $F_a = 1200$ Hz return phase. In agreement with analytical calculations, T_a is halved and, thus, F_a is doubled. Furthermore, the return curve overshoots the zero line by 13%.

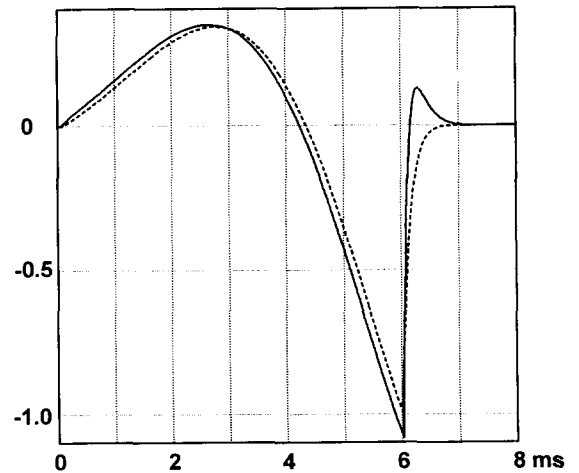


Fig. 6. Solid line shows overshoot of an LF-pulse if a hypothetical preemphasis representing the baffle effect of the head is added.

The latter effect is, as expected, the same as is found when overcompensating T_a in the just mentioned experimental set-up for T_a measurements.

However, the relevance of the baffle effect may be questioned. The effects we have predicted above seem rather extreme. Flanagan (1972) did not find any noteworthy extra preemphasis in his manikin experiment. Further experiments are needed to estimate $K_T(f)$ in real speech.

6. Breathy voicing

Selective inverse filtering retaining the F_1 -oscillation is exemplified in Figure 7 where the upper curves pertain to an isolated vowel [æ] produced with a high F_a of the order of 3000 Hz.

The lower curves pertain to a short stressed vowel [a] preceding an aspirated stop [k]. Here, the features of increasing T_a , i.e., decreasing F_a , and the associated A_i reduction and the increase of damping toward the end of the vowel are apparent. The underlying cause is the progressing vocal fold abduction which accounts for a large leakage component of air flow superimposed on the glottal pulse train. Formant amplitudes are, thus, progressively decreased both by the increasing duration of the return phase and by the increasing damping.

These two factors are further demonstrated in Figure 8 which shows 13 successive voice fundamental periods of mean $F_0 = 118$ Hz in the transition from a voiced [h] to a vowel [æ]. In the

F_1 -selective inverse filtering, there is almost no trace of an F_1 -oscillation in the voiced [h], whilst it is fully developed in the vowel [æ]. A quantitative analysis of F_a and B_1 within the sequence was undertaken. Measures of $T_a = 1/F_a$ were determined from the LF-model definition when large and, otherwise, from the A_i/E_e ratio (21).

F_a varied from 90 Hz in the [h] to 1200 Hz at the end of the [æ]. At the same time B_1 varied from a large value greater than 500 Hz in the [h] to a minimum value of 52 Hz at the end of the [æ]. A major part of the transition is located within a time span of three pitch periods in which F_a increases from 250 Hz to 1000 Hz, and B_1 decreases from 300 Hz to 130 Hz. There are traces of random noise ripple in the [h] part of the oscillogram.

Within the voiced [h], the wave shape of the selective inverse filtered function, as well as that of the extracted glottal flow derivative and the speech wave, closely resembles a full-wave rectified sinusoid which is well suited for an analytical derivation of its LF-parameters. The result is $R_g = 0.9$, $R_k = 0.39$, $T_a = 0.2T_0$ and, thus, $F_a = 1/(2\pi T_a) = F_0/(0.4\pi)$. With $F_0 = 118$ Hz inserted, we note $F_a = 94$ Hz which is close to the $F_a = 90$ Hz measure in Figure 8. There is a general agreement between the LF-data of the full-wave rectified sinewave above and data from voiced-voiceless boundary regions reported in (Gobl, 1988; Karlsson, 1990).

The variation of B_1 values reflects the increasing load of the glottal resistance at large values of glottal opening and, thus, of flow. The contribution of glottal damping to B_1 in a leaky phonation can be estimated from

$$\Delta B_1 = 50(F_1/500)^2 U_g(t)/55, \quad (23)$$

where $U_g(t)$ is average flow in cm^3/sec of the leakage component. This expression is an extension of equation (20) in (Fant and Lin, 1988), see also (Fant, 1979a) for a more detailed formula.

As was illustrated in Figures 7 and 8, the selective inverse filtering provides an overall dynamic view of parameters closely related to the source spectrum tilt and formant amplitudes and bandwidths. A more direct approach to studies of the effects of glottal abduction or adduction ges-

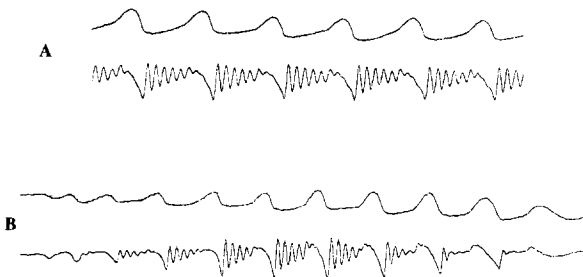


Fig. 7. Glottal flow pulses and F_1 selective inverse filtering. Male voice, low F_0 . The top curves, A, are from a vowel [æ] and the bottom curves, B, from a short stressed vowel [a] preceding an aspirated stop [k].

tures is to display the temporal course of the amplitudes of four formants and of the voice fundamental, as has been demonstrated in (Gobl and Ní Chasaide, 1988).

7. Continuous tracking of source amplitude

The most important of the voice source parameters is the source amplitude E_e , i.e., the amplitude of the negative going peak in the glottal volume velocity time derivative. The temporal variation of $E_e(t)$ can be constructed from LF-parameter determinations in successive voice periods. A smarter and more accurate method which preserves continuity is to perform a continuous inverse filtering from predetermined formant frequency tracking, as was used in connection with Figures 2, 8 and 9.

For studies of temporal variations of E_e , it is convenient to display a continuous inverse filter function in synchrony with a spectrogram, as shown in Figure 9. One remarkable property of the $E_e(t)$ function is its robustness with respect to inaccuracies in formant frequency matching and to the influence of interfering pole-zero pairs. A test with constant setting of the inverse filter to a neutral vowel pattern (B) gave much the same E_e contour as the complete inverse filtering (A). Moreover, an even more radical finding ap-

peared. The negative side of the speech wave oscillogram (C) also provided a very good approximation, as is also confirmed by Figure 10.

In view of the production theory outlined in the previous sections, this conformity can be anticipated. The glottal flow derivative waveshape reappears as a component in the speech wave and in absence of superposition from one period to the next, there should exist a continuity between the E_e spike and the onset of the sum of evoked formant oscillations.

A practical difficulty is to find out which side of the oscillogram is the negative part. This is a matter of experience. The negative part usually has larger peak amplitudes than the positive part, at least in nasal consonants and nasalized vowels, where the formant pattern is reduced by zeros and, thus, in part already inverse filtered. However, in a test of a female voice, we did not find the same conformity between the oscillogram and the proper E_e contour. A related technique is to approximate the glottal flow peak amplitude by simply integrating the speech wave (Fant, 1979a), see also Figure 4. Further tests are needed.

Another example of $E_e(t)$ compared to the oscillogram is shown in Figure 10. This is an assembly of data from (Fant, 1987) and from (Gobl, 1988) who studied LF-parameter variations within a word uttered in focal and in pre-focal positions within a sentence. The measured E_e

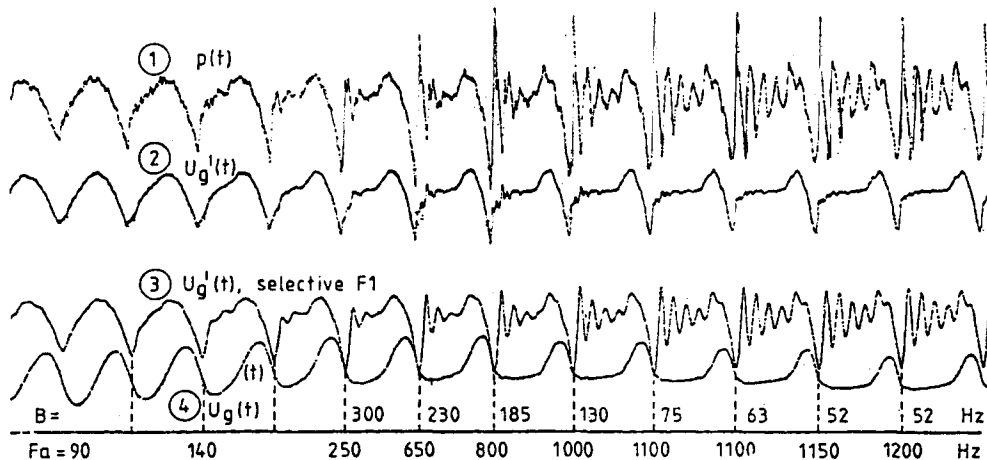


Fig. 8. (1) Oscillogram, (2) inverse filter output, (3) selective inverse filtering and (4) integrated inverse filter output for a voiced [h] and a following vowel [æ] with calculated values of B_1 and F_0 indicated. From (Fant and Lin, 1988).

contours, indicated with solid lines, conform well with the oscillograms.

This figure illustrates the general rule of increase in vowel-consonant contrast with emphasis (hyper-articulation) versus the reduced contrast in de-emphasis (hypo-articulation), as discussed in (Fant, 1987). Articulatory emphasis ensures more effective supraglottal consonantal constrictions which have the effect of reducing transglottal pressure and, thus, the amplitude and the shape of glottal pulses. In addition, the overall muscular tension associated with emphasis causes

an active glottal abduction for the [h] which reduces the voicing component or eliminates it and promotes noise generation.

Figure 9 provides additional insights in voice source dynamics. A reduction of E_c indirectly serves as a measure of VT-constriction. It is of the order of 50% for sonorant consonants like [l]. Narrow vowels like [i:] and [u:] in Swedish, see the middle and the right part of the figure, display decreasing E_c with increasing stress because of increased articulatory narrowing and, thus, interference with flow dynamics. The closure ges-

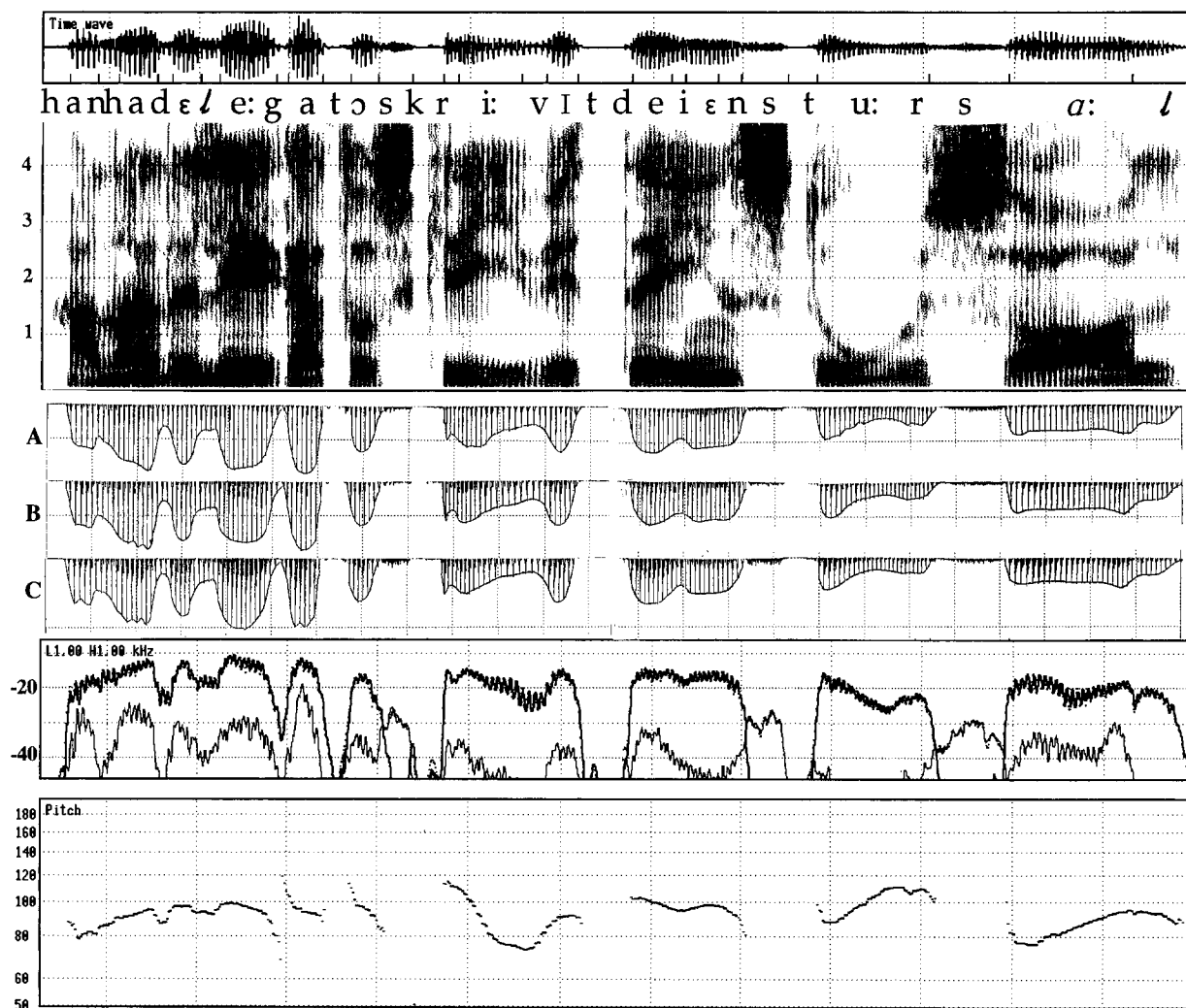


Fig. 9. Similarity of the negative side of (A) continuously inverse filtered speech, (B) the same with neutral vowel constant setting and (C) unfiltered speech wave, within the frame of a spectrogram with associated F_0 and intensity (low-pass and high-pass 1 kHz). Swedish sentence, "Han hade legat och skrivit det i en stor sal?"

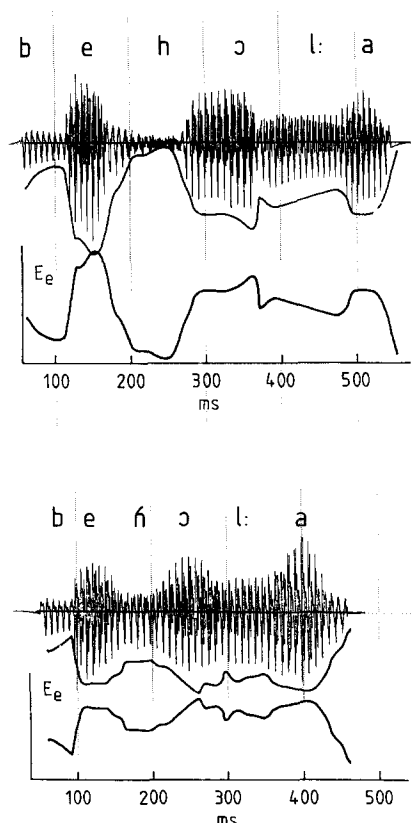


Fig. 10. Oscillogram and source excitation amplitude $E_e(t)$ of a word in sentence focal and below in sentence pre-focal position.

ture when enhanced by stress is diphthongal towards a consonantal target. More open vowels display an increase in E_e with stress. However, in connected speech, the contrast in intensity and in E_e comparing vowels of stressed and unstressed syllables averages a few dB only with higher values for emphatic stress. Increasing voice intensity is also associated with a relative emphasis of the high-frequency part of the source spectrum. More data on voice source parameter variation and interactions with the segmental structure are needed.

8. Male-female differences

What can be said about male-female differences in voice functions? From earlier studies

(Fant and Lin, 1988) we know that female glottal flow pulses are not simply scaled down versions of male voice pulses. If this had been the case, the average E_e values would have been the same. In addition, an F_0 - F_1 proximity is more often encountered in a female vowel which adds to the relative prominence of the voice fundamental. The glottal flow, U_0 , as well as the amplitude, E_e , of the derivative at closure and thus initial amplitudes, A_i , of formant oscillations are all about 3–7 dB higher in male than in female voices. Accordingly, U_0/E_e tends to be the same and of the order of 0.7–1.1 ms. An additional factor favouring higher formant levels in male speech than in female speech is lower bandwidths. On the other hand, observed differences in formant spectral levels are a few dB only to the advantage of male speech which is less than implied by E_e values. This can be explained by the higher F_0 of female voices which implies a larger number of excitations per unit time and, thus, of delivered energy. In terms of spectral amplitude levels, the female spectrum gains a factor of 5 dB in F_0 which compensates for the lower E_e values, see (9) and the more detailed discussion in (Fant and Lin, 1988).

Breathy voicing is more common in female speech than in male speech, and the overall ratio between voiceless and voiced parts of speech tends to be greater in female than in male speech. One instance is the more frequently occurring pre-occlusion aspiration in female speech. An additional feature that can be found in a female breathy voice is the occurrence of zeros and extra formants due to subglottal coupling (Fant and Lin, 1988; Klatt and Klatt, 1990; Fant et al., 1991). In a female voice, the steeper fall of the voice source spectrum at lower and medium frequencies is followed by a leveling off at higher frequencies which is generally ascribed to noise components. This might not be the whole truth. Modelling experiments taking into account a small but finite posterior glottal chink in the female glottis, which tends to stay open during phonation, would predict such a break in the source spectrum (Lin, 1990).

The relative prominence of the voice fundamental in a female voice is related to a greater open quotient, an F_1 closer to F_0 and a lower F_a .

Soprano singers tend to choose articulations that will ensure an F_1 close to F_0 , thus reinforcing the amplitude of the voice fundamental (Sundberg, 1973). An additional advantage of the F_1 – F_0 proximity is a reduction of air consumption, as already discussed (Schutte and Miller, 1986; Fant, 1986).

These are just a few experiences from applied theory and modelling. There remains much to be learned about male–female–child differences in voice functions, not the least for categorizing the large individual range of variabilities.

9. General discussion

Available data on voice source dynamics in connected speech are rather meager. Apart from the examples given in the previous sections and the limited material that has been analyzed in (Fant, 1979b, 1980; Fant and Lin, 1988; Gobl, 1988; Karlsson, 1990, 1991; Gobl and Karlsson, 1989; Gobl and Ní Chasaide, 1988; Strik and Boves, 1991), not much has been done. Inverse filtering and parameter extraction is a time-consuming undertaking. Fully automatic and reliable methods are highly needed. One difficulty in inverse filtering is to perform a pole-zero decomposition of vocal tract transfer functions without interfering with zeros that belong to the source-function (Karlsson, 1991).

We thus need to learn more about tracking temporal variations of vocal tract system functions, including subglottal coupling, and we should learn more about nonlinear source–signal interaction and noise generation as covarying factors in voice production. The acoustical correlates of breathiness are, thus, not found in the source function only, but to a substantial part also in the overall transfer function, e.g., as a radical increase of B_1 (equation (23)), some formant frequency shifts, mainly in F_1 and the appearance of subglottal formants which by now are well established facts.

The same argument holds for speech naturalness which may be enhanced by the choice of a more realistic voice pulse shape and by taking interaction phenomena into account. Of greater importance are the covarying variations of both

source and transfer functions at segmental and phrasal boundaries. It is the properties of the complete sound rather than the properties of the source alone that have to be considered. Advanced schemes of articulatory modelling and synthesis, e.g., (Lin, 1990; Lin and Fant, 1992), will pave the way for a deeper understanding.

Speech intensity has a complex relation to source and to VT-characteristics. Moreover, there is not a simple relation between intensity and stress. Increased stress results in hyper-articulation and a greater contrast between open and closed articulations. A high front vowel like [i:], articulated with greater tension, will become more close and thereby lose intensity, both because of a less efficient source and a more consonantal-like formant pattern. In Swedish, the tense [i:] will approach [j]. On the other hand, a de-emphasis results in hypo-articulation and, thereby, in incomplete consonantal closures which will reduce vowel-consonant contrasts, see (Fant, 1987; Gobl, 1988; Fant et al., 1991).

A possible simplification in collecting data on LF-model parameters would be to concentrate on E_e , T_a (or F_a) and the open quotient O_q and let R_k and R_g become dependent parameters. Instead of direct period-by-period measurements of T_a one could employ running high-pass and low-pass measures of the voice source from which approximate T_a measures could be derived.

Acknowledgments

These studies have been undertaken with partial support from The Bank of Sweden Tercentenary Foundation, The Carl Trygger Foundation and ESPRIT grant SPEECHMAPS (via NUTEK). Valuable help in preparing illustrations was provided by Johan Liljencrants.

References

T.V. Ananthapadmanabha (1984), "Acoustic analysis of voice source dynamics", *STL-QPSR*¹, Nos. 2–3/1984, pp. 1–24.

¹ *STL-QPSR* is the Speech Transmission Laboratory Quarterly Progress and Status Report published by the Department of Speech Communication and Music Acoustics, KTH, Stockholm.

- T.V. Ananthapadmanabha and G. Fant (1982), "Calculation of true glottal flow and its components", *Speech Communication*, Vol. 1, Nos. 3–4, pp. 167–184.
- C.A. Bickley (1991), "Vocal-fold vibration in a computer model of the larynx", in *Vocal Fold Physiology*, ed. by J. Gauffin and B. Hammarberg (Singular Publishing Group, San Diego, CA), pp. 37–46.
- C.A. Bickley and K.N. Stevens (1986), "Effects of a vocal tract constriction on the glottal source: Experimental and modelling studies", *J. Phonetics*, Vol. 14, pp. 373–382.
- B. Briess and G. Fant (1962), "Studies of voice pathology by means of inverse filtering", *STL-QPSR*, No. 1/1962, p. 6.
- R. Carlson, G. Fant, C. Gobl, B. Granström, I. Karlsson and Q. Lin (1989), "Voice source rules for text-to-speech synthesis", *Internat. Conf. Acoust. Speech Signal Process.*, Vol. 1, pp. 223–226.
- R. Carlson, B. Granström and I. Karlsson (1991), "Experiments with voice modelling in speech synthesis", *Speech Communication*, Vol. 10, Nos. 5–6, pp. 481–489.
- G. Fant (1959a), "Acoustic analysis and synthesis of speech with applications to Swedish", *Ericsson Technics*, No. 1.
- G. Fant (1959b), "The acoustics of speech", *Proc. 3rd Internat. Congress Acoustics, Stuttgart*, pp. 187–201.
- G. Fant (1960), *Acoustic Theory of Speech Production* (Mouton, the Hague).
- G. Fant (1979a), "Glottal source and excitation analysis", *STL-QPSR*, No. 1/1979, pp. 85–107.
- G. Fant (1979b), "Voice source analysis – A progress report", *STL-QPSR*, Nos. 3–4/1979, pp. 31–54.
- G. Fant (1980), "Voice source dynamics", *STL-QPSR*, Nos. 2–3/1980, pp. 17–37.
- G. Fant (1982a), "Preliminaries to the analysis of the human voice source", *STL-QPSR*, No. 4/1982, pp. 1–27.
- G. Fant (1982b), "The voice source, acoustic modeling", *STL-QPSR*, No. 4/1982, pp. 28–48.
- G. Fant (1986), "Glottal flow: Models and interaction", *J. Phonetics*, Vol. 14, Nos. 3/4, pp. 393–399.
- G. Fant (1987), "Interactive phenomena in speech production", *Proc. XIth Internat. Congress Phonetic Sciences (ICPhS), Tallinn, USSR*, Vol. 3, pp. 376–381.
- G. Fant and T.V. Ananthapadmanabha (1982), "Truncation and superposition", *STL-QPSR*, Nos. 2–3/1982, pp. 1–17.
- G. Fant and Q. Lin (1987), "Glottal source – Vocal tract acoustic interaction", *STL-QPSR*, No. 1/1987, pp. 133–127.
- G. Fant and Q. Lin (1988), "Frequency domain interpretation and derivation of glottal flow parameters", *STL-QPSR*, Nos. 2–3/1988, pp. 1–21.
- G. Fant and Q. Lin (1991), "Comments on glottal flow modelling and analysis", in *Vocal Fold Physiology. Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, ed. by J. Gauffin and B. Hammarberg (Singular Publishing Group, San Diego, CA), pp. 47–56.
- G. Fant and B. Sonesson (1966), "Indirect studies of glottal cycles by synchronous inverse filtering and photoelectric glottography", *STL-QPSR*, No. 4/1966, pp. 1–3.
- G. Fant, K. Pintoft, J. Liljencrants, B. Lindblom and J. Mártony (1963), "Formant-amplitude measurements", *J. Acoust. Soc. Amer.*, Vol. 35, No. 11, pp. 1753–1761.
- G. Fant, K. Ishizaka, J. Lindqvist-Gauffin and J. Sundberg (1972), "Subglottal formants", *STL-QPSR*, No. 1/1972, pp. 1–12.
- G. Fant, J. Liljencrants and Q. Lin (1985a), "A four-parameter model of glottal flow", *STL-QPSR*, No. 4/1985, pp. 1–13.
- G. Fant, Q. Lin and C. Gobl (1985b), "Notes on glottal flow interaction", *STL-QPSR*, Nos. 2–3/1985, pp. 21–45.
- G. Fant, A. Kruckenberg and L. Nord (1991), "Prosodic and segmental speaker variations", *Speech Communication*, Vol. 10, Nos. 5–6, pp. 521–531.
- J.L. Flanagan (1972), *Speech Analysis Synthesis and Perception* (Springer, New York) (1st Edition 1965).
- J.L. Flanagan, K. Ishizaka and K.L. Shipley (1975), "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", *Bell System Tech. J.*, Vol. 54, pp. 485–505.
- H. Fujisaki and M. Ljungqvist (1986), "Proposal and evaluation of models for the glottal source waveform", *Internat. Conf. Acoust. Speech Signal Process.*, Vol. 3, pp. 1605–1608.
- H. Fujisaki and M. Ljungqvist (1987), "Estimation of voice source and vocal tract parameters based on ARMA analysis and a model for the glottal source waveform", *Internat. Conf. Acoust. Speech Signal Process.*, Vol. 2, pp. 637–640.
- C. Gobl (1988), "Voice source dynamics in connected speech", *STL-QPSR*, No. 1/1988, pp. 123–159.
- C. Gobl and I. Karlsson (1989), "Male and female voice source dynamics", in *Vocal Fold Physiology*, ed. by J. Gauffin and B. Hammarberg (Singular Publishing Group, San Diego, CA), pp. 121–128.
- C. Gobl and A. Ni Chasaide (1988), "The effects of adjacent voiced/voiceless consonants on the vowel voice source: A cross language study", *STL-QPSR*, Nos. 2–3/1988, pp. 23–59.
- I. Karlsson (1990), "Voice source dynamics for female speakers", *Proc. 1990 Internat. Conf. Spoken Language Processing, Kobe*, pp. 69–72.
- I. Karlsson (1991), "Dynamic voice quality variations in natural female speech", *Proc. Internat. Conf. Phonetic Science, Aix-en-Provence*, Vol. 4, pp. 10–13.
- D.H. Klatt and L.C. Klatt (1990), "Analysis, synthesis and perception of voice quality variations among female and male talkers", *J. Acoust. Soc. Amer.*, Vol. 87, pp. 820–857.
- J. Liljencrants (1991), "A translating and rotating mass model of the vocal tract", *STL-QPSR*, No. 1/1991, pp. 1–18.
- Q. Lin (1990), *Speech production theory and articulatory speech synthesis*, DSc. Thesis, Department of Speech Communication and Music Acoustics, KTH, Stockholm.
- Q. Lin and G. Fant (1992), "An articulatory speech synthesizer based on a frequency domain simulation of the vocal tract", *Internat. Conf. Acoust. Speech Signal Process.*, Vol. 2, pp. 57–60.
- J. Lindqvist-Gauffin (1964), "Inverse filtering – Instrumentation and techniques", *STL-QPSR*, No. 4/1964, pp. 1–4.
- J. Lindqvist-Gauffin (1965), "Studies of the voice source by means of inverse filtering", *STL-QPSR*, No. 2/1965, pp. 8–13.
- J. Lindqvist-Gauffin (1970), "The voice source studied by means of inverse filtering", *STL-QPSR*, No. 2/1970, pp. 3–9.

- J. Mártony (1965), "Studies of the voice source", *STL-QPSR*, No. 1/1965, pp. 4–9.
- L. Nord, T.V. Ananthapadmanabha and G. Fant (1986), "Perceptual tests using an interactive source filter model and considerations for synthesis strategies", *J. Phonetics*, Vol. 14, pp. 401–404.
- M. Rothenberg (1973), "A new inverse filtering technique for deriving the glottal airflow waveform during voicing", *J. Acoust. Soc. Amer.*, Vol. 53, pp. 1632–1645.
- M. Rothenberg (1983), "Source-tract interaction in breathy voice", in *Vocal Fold Physiology*, ed. by I. Titze and R. Scherer (The Denver Center for the Performing Arts, Denver, CO), pp. 465–481.
- H.K. Schutte and D.G. Miller (1986), "The effect of F_0/F_1 coincidence in soprano high notes on pressure at the glottis", *J. Phonetics*, Vol. 14, pp. 385–392.
- K.N. Stevens (1987), "Interaction between acoustic sources and vocal tract configuration for consonants", *Proc. XIth Internat. Conf. Phonetic Science, Tallinn*, pp. 385–389.
- H. Strik and L. Boves (1991), "On the relation between source parameters and prosodic features in connected speech", *Speech Communication*, Vol. 11, pp. 167–174.
- J. Sundberg (1973), "Observations on a professional soprano singer", *STL-QPSR*, No. 1/1973, pp. 14–24.
- J. Sundberg and J. Gauffin (1979), "Wave form and spectrum of the glottal voice source", in *Frontiers of Speech Communication Research*, ed. by B. Lindblom and S. Öhman (Academic Press, London), pp. 301–322.