

Dept. for Speech, Music and Hearing
**Quarterly Progress and
Status Report**

**The source filter concept in
voice production**

Fant, G.

journal: STL-QPSR
volume: 22
number: 1
year: 1981
pages: 021-037



**KTH Computer Science
and Communication**

<http://www.speech.kth.se/qpsr>

B. THE SOURCE FILTER CONCEPT IN VOICE PRODUCTION*

G. Fant

ABSTRACT - This is a review of recent work on source-filter composition of voiced sounds, describing various alternatives of defining a source function with implications for a more profound understanding of the acoustic processes involved and their modelling in equivalent networks. The theoretical study is backed up by experiments with inverse filtering and other types of glottography. Special attention is devoted to time-varying bandwidths within a glottal open period and the implications for defining overall effective bandwidths. The presence in the speech wave of voice source components as well as the presence of formant oscillation in the glottal flow are discussed. Expressions for supra- and sub-glottal pressure fluctuations are also derived. An earlier proposed parametric model of glottal flow and formant excitation is reviewed and discussed with respect to voice intensity and some dynamic factor in connected speech.

1. INTRODUCTION

The basic philosophy of acoustic theory of speech production is that speech involves a source function and a vocal-tract filtering process, i.e., a raw material and a sound shaping. Since vocal-tract filtering conveys a larger part of the information in speech less attention has been paid to source functions. In practical applications, such as speech synthesis and vocoders and in general descriptive theory, source functions are mostly considered as simple on/off switching of a quasiperiodic excitation for voiced sounds alternating with discreet transients and noise sources for unvoiced sounds with the recognition that mixed excitation of voicing modulating noise sources also occurs.

In recent years the demand for greater naturalness in speech synthesis has developed a renewed interest in studies of the voice source and source filter interactions. This demand which is paralleled by a general descriptive need in acoustic phonetics, e.g., with respect to speech prosody and individual voice qualities and speaking habits, is revealed when we attempt to synthesize female speech rule. Voice pathology and theory of singing are related topics, where the need for improved voice source models are apparent, Gauffin and Sundberg (1980) and Sundberg and Gauffin (1979).

The following presentation is limited to considerations of the aerodynamic and acoustic stages of voice production dealing with basic definitions of source and filter functions, their interaction and common dependency on articulatory patterns with a view of establishing improved quantitative models for the production process. This involved both considerations of how to decompose a stationary

* This paper was an invited lecture presented at the IV F.A.S.E. Symposium on Acoustics and Speech, Venice, April 21-24, 1981

voiced speech segment into source and filter parameters, synthesis strategies, and the study of dynamic variations in connected speech. Readers are referred to Titze (1980) for a review of the vocal cord vibrational mechanism.

2. ELEMENTARY THEORY, TIME AND FREQUENCY DOMAIN REPRESENTATIONS

The acoustics of voice production involves two intimately connected aspects, aerodynamics and sound, which in a sense are analogous to direct current and alternating current. The aerodynamics describing the mean pressures and flow of air in the respiratory pathways relates to the power supply, whereas the sound derives from disturbances in flow and pressure distributions. The concept of source could be applied to the underlying exhalation pressures and flow. An efficiency measure can be defined as the ratio of radiated power to the exhalation power which becomes extremely small and of not much interest except in a comparative sense.

However, from a signal analysis point of view, we generally define the source as the pulsating flow passing through the time-varying glottis opening between the vocal folds. The output of the filtering process is the sound pressure wave in front of the speaker, which is related to the intermediate stage of air flow passing the lips of the speaker by a simple differentiation, i.e., a 6-dB per octave spectral boost.

The corresponding time-domain representations are less well known. Technically they may be derived from inverse filtering of the sound pressure wave or of the mouth output flow to cancel the filtering effect thus regenerating the source properties. The very first stage of this transform from the speech wave to the mouth output, which is a simple integration, provides a zero order approximation to inverse filtering. The integration enhances the time domain by the appearance of a usually clear pulse component resembling that of the glottal flow and in the frequency domain there is a clear dominance of a low-frequency area below that of the first formant.

The time-domain representation can be brought back to a level preceding that of the glottal flow, i.e., the time-varying glottal opening. Due to inertia effects the glottal flow generally builds up more slowly than implied by the glottal area function but terminates more steeply to meet the requirement of zero flow after closure.

The successive transformations from source to sound may be taken in a different order, i.e., (source $\times$ radiation) $\times$ filter, which defines the differentiated glottal flow as the source and the vocal-tract volume velocity transfer as the filter. With this view one can identify in the waveshape of the radiated sound certain characteristics which derive from the waveshape of the differentiated source. The source residue is apparent as a negative starting amplitude of the first formant oscillation, which is continuous with the preceding negative-going excursion reflecting the derivative of the glottal flow during the closing phase. The source residue has a spectral correspondence in the spectrum level in the region of the fundamental or first few harmonics, which may be enhanced at low intensity voicing to the same order of magnitude as the first formant. In this respect we could speak of a "glottal formant", but I would prefer to refer it as the baseband source peak.

Another topic which I will expand on later in this paper is the temporal envelope of the damped oscillations representing vocal-tract resonances. The main excitation appears at the instant of glottal closure after which there is a more or less uniform decay until the beginning of next glottal opening phase when the rate of decay may increase rapidly. At low fundamental frequencies the first formant oscillation may be effectively terminated before the beginning of the next excitation. This is the situation illustrated by Fig. I-B-1 in which for graphical simplicity wave components of F2 and higher formants are omitted.

3. INVERSE FILTERING

Inverse filtering techniques has been used over 25 years to extract source

characteristics from a speech wave. The more wellknown studies are those of Miller (1959), Lindqvist (1965), Rothenberg (1973), Holmes (1976), Sundberg and Gauffin (1979). For my own studies in the past two years, Fant (1979; 1980) I built an inverse filtering analog equipment operating from an FM-taperecorder and scaled down 16 times in play-back speed, which allow for high resolution graphical recording on an Oscillomink. A part of a sentence to be analyzed was processed repeatedly with different settings of formant frequencies measured from oscillograms and spectrograms. One such example is shown in Fig. I-B-2 which supports the general presentation already given schematically in Fig. I-B-1. The first trace is the speech wave of a vowel [æ], the second is the differentiated inverse filtered version, and the third function is the corresponding undifferentiated source function. The fourth trace shows the speech wave after removing all formants but F1.

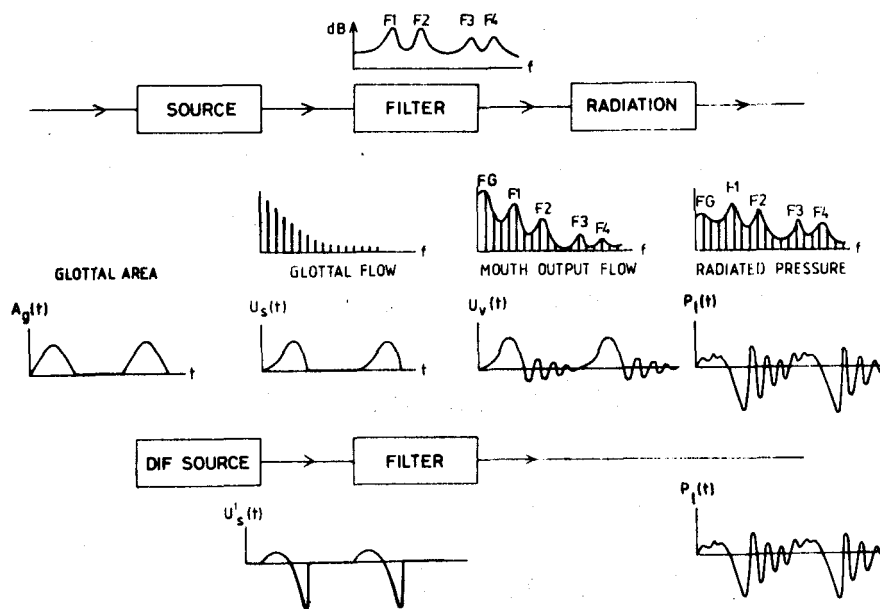


Fig. I-B-1. Frequency and time-domain general concepts of source-filter decomposition of voiced sounds.

æ (här)

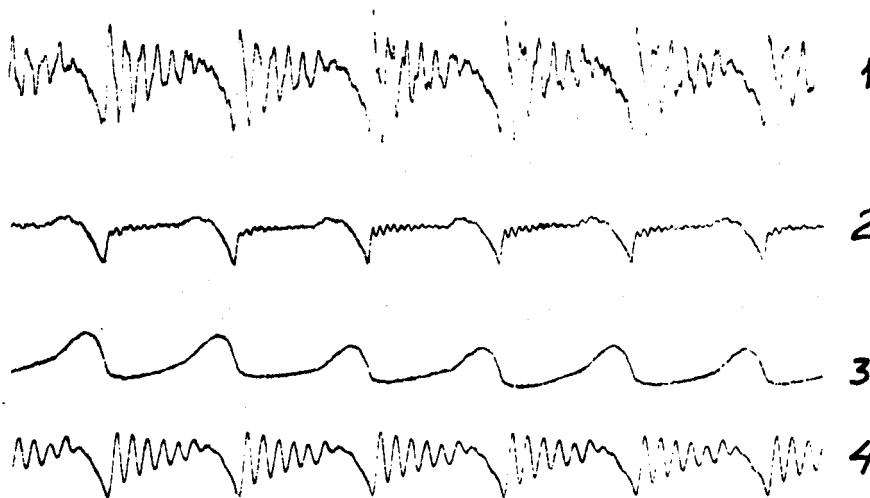


Fig. I-B-2. Oscillogram of vowel [æ], male voice, $F_0=75$ Hz. From top pressure wave, inverse filtering with differentiation added, the same without differentiation and differentiated inverse filtered function without F1 cancellation.

This is a picture typical of male speech at low voice fundamental frequencies. Observe the continuity between the negative going glottal flow differential and the starting amplitude of the first formant oscillation. A high quality condenser microphone was used in this recording.

Another technique illustrated in Fig. I-B-3 is based on the use of a Rothenberg type face mask, which preserves a dynamic measure of mouth output flow.*

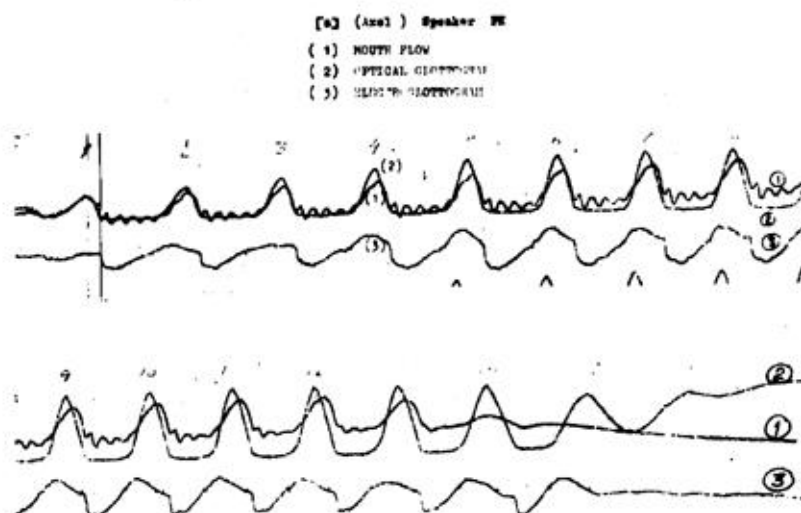


Fig. I-B-3. Stressed Swedish vowel [a] preceding k-occlusion as in the word "Axel" showing (1) mouth flow, (2) optical glottogram, (3) electroglottogram.

The recording was made simultaneously with an electroglottogram and with an optical glottogram produced by transglottal illuminations and a light pick-up in the pharynx, Kitzing (1979). In this processing inverse filtering was omitted in order to bring out the F1-ripple in the glottal closed phase of the mouth flow to conform with the schematic presentation in Fig. I-B-1. The synchrony between the optical, acoustical, and electrical glottograms at the points of glottal closure has been enhanced by a correction for the vocal tract and mask transit time. In the electrical glottogram it is the rapid resistance fall which signals the instant of vocal cords reaching contact.

The shape of the area function peaks derived from the optical glottograms is symmetrical and something between a half sinewave and a triangle, as generally seen from calculations with the Flanagan-Ishizaka two-mass model. There is also to be seen in Fig. I-B-3 a latency of the peaks of glottal flow with respect to corresponding area peaks, which we shall discuss later.

The phonetic frame of Fig. I-B-3 is that of a stressed short Swedish vowel [a] before an unvoiced stop. There is a gradual increase of the glottis area towards the end of the vowel with a terminating incomplete closure, which can be considered as pre-occlusion aspiration. It compares well with Fig. 10 in Flanagan et al (1975) experiments with the two-mass model of the vocal cords.

* Fig. I-B-3 derives from a pilot study in Stockholm 1980 in cooperation with Martin Rothenberg, Anders Löfqvist, Peter Kitzing, Jan Gauffin, and Johan Sundberg. The optical and electrical glottograms were produced with Peter Kitzing's (1979) instrumentation.

4. CIRCUIT THEORY OF VOICE PRODUCTION

Source filter theory of voice production is the theory of its approximations. Time-varying and non-linear functions introduce difficulties which obscure an insight in essentials unless we take resort to simplified and approximate transformations. The physically most correct would be to regard the lung pressure P_g and its associated muscular back-up as the source and derive the output from a set of differential equations representing the state of flow and pressure in the combined subglottal, glottal, and supraglottal systems in extremely short successive intervals of time. This is actually the principle of the Ishizaka-Flanagan (1972) model which has become most influential in speech research. However, we need an "alternating current" source to define the temporal or spectral raw material to be shaped by the filtering process. One approach which I have used since my early work is to derive an equivalent current $U_s = P_g(Z_{sg} + Z_g)^{-1}$ which defines the source as the flow through the glottis impedance with the vocal tract input terminals short-circuited, see Fig. I-B-4. This source function is then applied to the input terminals of the vocal tract in parallel with the source impedance ($Z_{sg} + Z_g$)

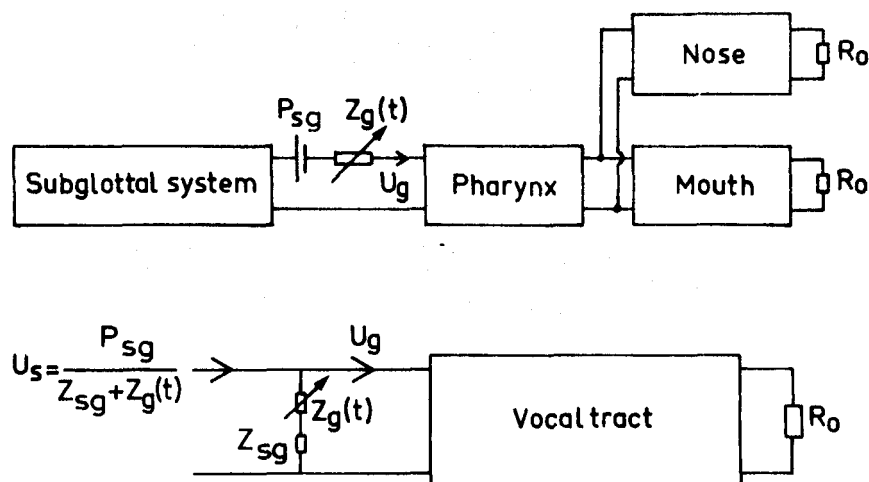


Fig. I-B-4. Generalized voice production network.

The true expression for the transglottal pressure drop time function is

$$P_{tg} = P_{sg} - P_i = (R_k + R_v) U_g + \frac{d}{dt} (L_g U_g) \quad \text{Eq. (1)}$$

where all the quantities are time variable, see Flanagan (1972). The "kinetic" resistance R_k dominates and is related to the flow by the approximate expression

$$R_k = \left(\frac{\rho P_{tg}}{2} \right)^{\frac{1}{2}} \cdot \frac{1}{A_g(t)} = \frac{\rho v_g}{2 A_g(t)} \quad \text{Eq. (2)}$$

in which the particle velocity

$$v_g = (2P_{tg}/\rho)^{1/2} = U_g/A_g \quad \text{Eq. (3)}$$

is constant over the entire extent of the glottal opening. This is in practice not so and an appreciable acceleration is to be expected from the overall inertia effecting the flow in the entire vocal tract. However, this simplification provides a first approximation

$$U_g(t) = v_g(t)A_g(t) \approx (2P_{tg}/\rho)^{1/2} \cdot A_g(t) \quad \text{Eq. (4)}$$

whereby glottal flow is proportional to glottal area.

In a quasitransform expression referring to Fig. I-B-4 we may relate the actual glottal flow to the glottal, subglottal, and supraglottal impedances and the lung pressure P_g

$$U_g(s,t) = \frac{P_g}{Z_{sg}(s) + sL_g(t) + R_g(t) + Z_v(s)} \quad \text{Eq. (5)}$$

where $Z_v(s)$ is the vocal tract input impedance. From this expression we may make the general conclusion that the glottal flow would have zeros whenever either the vocal tract input impedance or the subglottal impedance has poles, which would occur at the resonance frequencies of the uncoupled supra- and subglottal tracts. However, the glottal flow also contains resonance components from the entire supra- plus subglottal systems, which tend to cancel the effects of the zeros in the output sound. An exception is unvoiced aspirated sounds, where the frequency of the subglottal formants may obscure the interpretation of the spectrum, Fant et al (1972).

The two alternatives of defining a current source, the glottal volume flow, and the short-circuit flow both have their drawbacks. The true glottal flow U_g becomes an object of incremental calculations in the time domain of a complexity not less than that of solving for the output. Moreover, U_g contains superimposed formant oscillations which implies a linear dependency on the particular articulation. The true advantage of U_g as a source candidate is that it enables us to specify the initial conditions for the transient response of the supraglottal tract after closure, which determines a major part of the formant energy.

The linear source-filter dependency is avoided if we define the source by the underlying short circuit current U_s but we lose the control of the source dependency of aerodynamic pressure drops in supraglottal constriction as well as the specific glottal flow shaping induced by the overall inertia effects. The best we can do then if we want to define a source by a reasonable transfer equation is to choose the transform parameters for the short-circuit source, such as to anticipate certain overall aspects of source filter interaction which can be made as a part of a synthesis program. Essential of the source filter decomposition may be studied in a simplified circuit, Fig. I-B-5, with constant subglottal pressure and the supraglottal tract modelled by a single parallel resonance circuit representing the first formant F_1 . The solution can then easily be extended to a multiple pole transmission line representation of the vocal tract.

The main excitation at the closing edge of the glottal flow pulse U_g or of the corresponding short current flow U_s may be modelled by a ramp source

$$S(s) = \frac{-dU_s(t)}{dt} \cdot \frac{1}{s^2} = -U'_s/s^2 \quad \text{Eq. (6)}$$

where we may alternatively choose U_g instead of U_s with consequences to be derived in the following analysis.

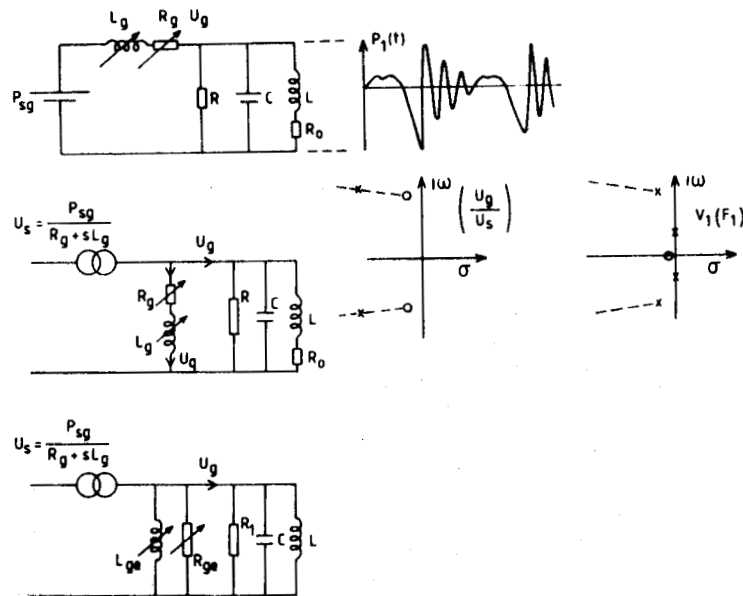


Fig. I-B-5. Simplified networks representing a single resonator generation of the first formant.

The filter function relating output current $U_R(s)$ through the radiation resistance R_0 to the source $S(s)$ may in both instances be put in the general form

$$H_1(s) = \frac{\omega_0^2}{(s + \alpha_1)^2 + \omega_1^2} \quad \text{Eq. (7)}$$

where $\omega_0^2 = 1/L_e C$, $1/L_e = 1/L + 1/L_{ge}$

$$L_{ge} = L_g(1 + R_g^2/\omega^2 L_g^2); \quad R_{ge} = R_g(1 + \omega^2 L_g^2/R_g^2)$$

$$-\alpha_1 = \pi B_1 = 1/2\pi R_e C, \quad 1/R_e = 1/R + 1/R_{ge} + \omega^2 L^2/R_0$$

with the understanding that the source elements R_{ge} and L_{ge} are infinite during an ideal closure phase and over the entire voice period, if the $U_g(s)$ source is chosen.

The radiation transfer from the current through the lips to the sound pressure at a cm from the speaker is proportional to frequency, Fant (1960)

$$H_r(s) = \frac{Q \cdot S}{4\pi \cdot a} \quad \text{Eq. (8)}$$

The combined transform of source vocal transfer and radiation is thus

$$H_v(s) = S(s) \cdot H_1(s) \cdot H_r(s) = \frac{-U_s'}{4\pi \cdot a} \cdot \frac{\omega_0^2}{s[(s + \alpha_1)^2 + \omega_1^2]} \quad \text{Eq. (9)}$$

and the corresponding time function becomes

$$p_v(t,a) = -U'_s \cdot \frac{g}{4\pi a} \cdot e^{-\pi B_1 t} \cos(2\pi F_1 t + \varphi_1) \quad \text{Eq. (10)}$$

The next step is to define the entire extent of the source pulse by a smooth time function, for instance with a rising branch $U = \frac{1}{2}U_0 [1 - \cos \omega_g(t-T_1)]$ and a falling branch $U = U_0 [K \cos \omega_g(t-T_2) - K + 1]$,

where $\omega_g = 2\pi F_g$ and $F_g = \frac{1}{2(T_2-T_1)}$, see Fig. I-B-6 from Fant (1979).

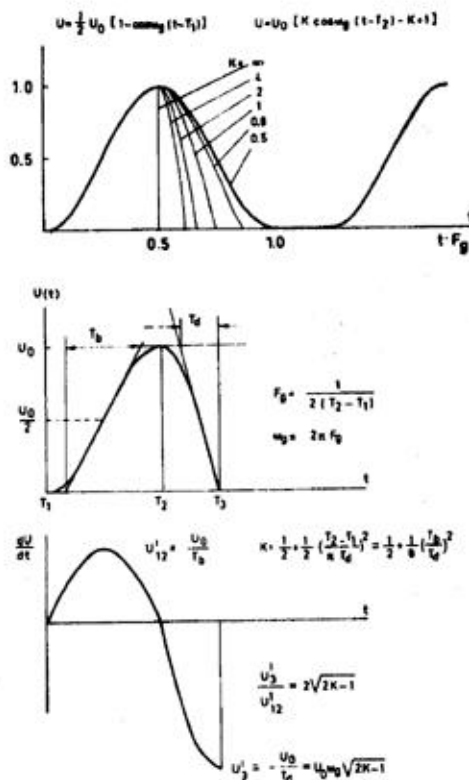


Fig. I-B-6. The source model, Fant (1979).

This source pulse has discontinuities in the derivatives at the instant of opening T_1 , peak T_2 , and closure T_3 . The transient response at T_2 and T_3 are damped oscillations with lower initial amplitudes than the response to the T_1 ramp already derived above and they decay 6-dB per octave faster with frequency. The inverse Laplace transform also provides expressions for the residue in the speech wave of the differentiated source pulses. Providing F_1 is large compared to F_g they are preserved in almost the original source shape but for a small phase shift. The analysis further proves the continuity between the negative peak of the source residue in the speech wave and the starting amplitude of the oscillations at the instant of opening.

We are now in the position to describe the essentials of the history of events within a single voice period. The T_1 and T_2 oscillations, as well as pos-

sible residual T_3 oscillations from previous periods, are subjected to a slight time variable increase in frequency during glottal opening and a time variable damping due to the glottal differential resistance which we may approximate with twice the $R_k(t)$ previously derived. Thus $L_g = \rho d / A_g(t)$ and $R_g = 2R_k = \rho v_g / A_g(t)$ with the possible generalization $v_g(t) = U_g(t) / A_g(t)$. We thus have an instantaneous bandwidth which is the linear sum of the

$$B_1(t) = 1/2 \pi R C + R_0 / 2 \pi \cdot L + 1/2 \pi R_{ge}(t) C \quad \text{Eq. (11)}$$

contributions from vocal tract internal losses, radiation losses, and glottal losses. The latter term may reach high values of the order of F_1 in open sounds as [a] which is a situation of almost impedance match comparing the characteristic impedance of the pharyngeal tract with the glottis resistance. In this situation the rate of oscillation decay is accordingly very high, as illustrated by Fig. I-B-7, the formant oscillation is effectively truncated in the glottal open part and the waveshape becomes dominated by the source residue. In a leaky voice there is a permanent increase of losses during the maximally closed phase. The greater the glottal opening, the smaller is R_g and the larger is the damping effect.

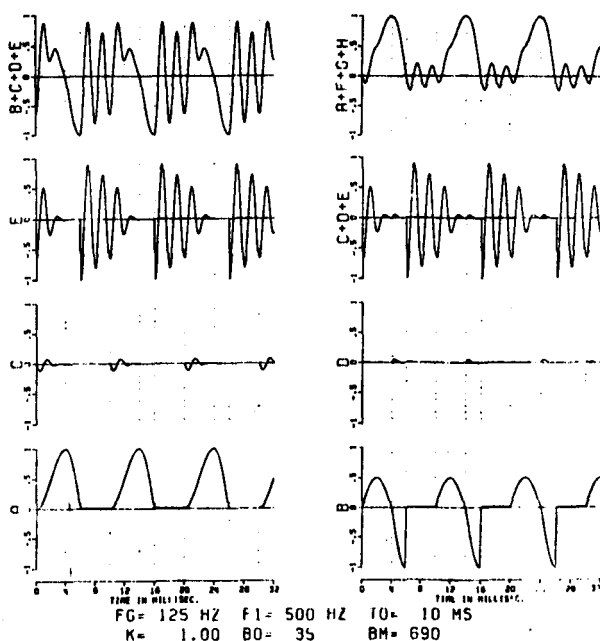


Fig. I-B-7. Time-domain calculation of the response of a one-formant approximation of the vocal tract. Left column: pressure wave, F_1 oscillation excited at closure, at opening, the source flow. Right column: integrated speech wave, sum of F_1 oscillation, F_1 oscillation at the glottal peak, and finally, the differentiated source function.

What is the relation of the glottal flow to the preceding short-circuit source U_s ? From the equivalent network we derive

$$\frac{U_g(s)}{U_s(s)} = \frac{[(s + \bar{\alpha}_1)^2 + \bar{\omega}_1^2]}{[(s + \alpha_1)^2 + \omega_1^2]} \quad \text{Eq. (12)}$$

which contains the time variable resonance frequency of a bandwidth already discussed and a zero representing the maximum of the input impedance, or, in other words, the first formant without regard to the glottal load. This zero-pole pair may be regarded as a spectral notch filter with a pole frequency

$$F_1 = \frac{1}{2\pi} \left[\left(\frac{1}{L} + \frac{1}{L_{ge}} \right) \frac{1}{C} \right]^{\frac{1}{2}} \text{ somewhat larger than the zero frequency } \bar{F}_1 = \frac{1}{2\pi} \left(\frac{1}{L_C} \right)^{\frac{1}{2}} \text{ and}$$

with a pole bandwidth which exceeds the zero bandwidth by the time variable glottal bandwidth $1/2\pi R_{ge}C$. The depth of the spectral notch ignoring the difference between $\bar{F}_1$ and F_1 is $R_g/(R_g+R_1)$. However, in the expression for the sound output the zero disappears and there merely remains a formant peak which has an instantaneous peak level of $R_g(t)/R_g(t)+R_1$. This reduction, which may be quite appreciable, more than 20 dB, will, however, not show up in a spectrum averaged over a voice fundamental period or longer. On the contrary, if the VT internal losses are relatively high, the additional R_g damping accounts for only a slight decrease in the overall signal level.

How do we define an effective formant bandwidth in this case? A study of Fant and Liljencrants (1979) supports the view, see Fig. I-B-8, that if a truncated damped oscillation has the same starting amplitude as an ideal exponentially damped oscillation, it is perceived maximally similar if the bandwidth of the ideal exponentials satisfies a condition of the same integrated mean value of signal amplitude, see Fig. I-B-8.

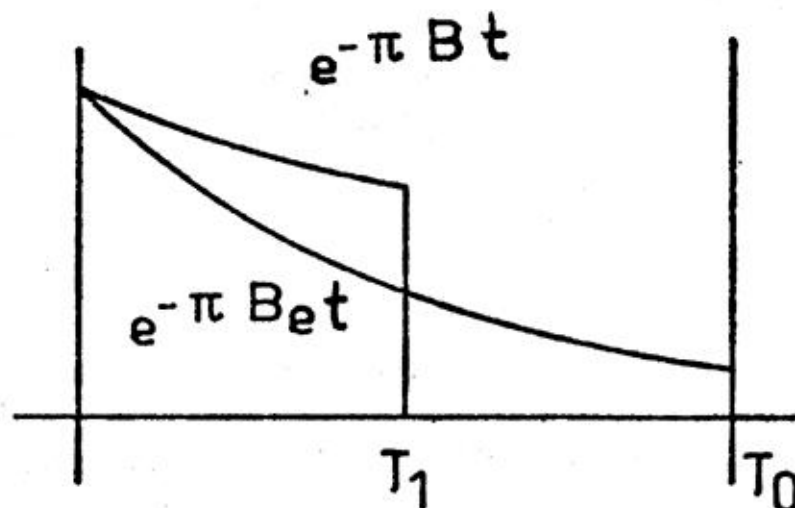


Fig. I-B-8. Bandwidth of truncated exponential equals that of a continuing reference.

$$\frac{1 - e^{-\pi B_1 T_0}}{\pi B_1 T_0} = \frac{1 - e^{-\pi B_1 t}}{\pi B_1 \cdot T_0} \quad \text{Eq. (13)}$$

Incidentally, this is also the criterion of the same effective peak spectrum level which may explain the perceptual results. The following tabulation exemplifies the equivalent bandwidth increase due to truncation. The bandwidth during glottal closure is $B_1 = 1/2R_1C = 35$ Hz and $F_0 = 100$ Hz and 200 Hz.

Table	t/T_0	$F_0 = 100$ Hz, ΔB	$F_0 = 200$ Hz, ΔB
I-B-I	0.2	140	
	0.3	87	
	0.4	63	132
	0.5	40	87
	0.6	29	60
	0.8	10	23
	1.0	0	0

We may conclude that the bandwidth increase due to truncation is rather small when the opening quotient of (glottal pulse duration) / (voice fundamental period) is small and when the voice fundamental frequency is low. The F_0 proportionality is apparent. Also the effect of a truncation is larger when the bandwidth during glottal closure is low. (An alternative 3 dB down spectral criterion does not in itself have any perceptual relevance and provides bandwidths larger than those derived from our signal mean value criterion.)

What can be said about the waveshape of the true glottal flow $U_g(t)$ compared to the assumed $U_s(t)$ source? One expected feature is an F1 oscillation superimposed on $U_g(t)$. In order to find a sensitive measure we can calculate $U_s(t) - U_g(t) = U_q(t)$, which is the current through the $R_g L_g$ vocal tract input shunt. If we neglect L_g , we could determine $U_g(t)$ as the vocal tract input pressure $P_i(t)$ divided by R_g . A solution for $P_i(t)$ by means of an inverse Laplace transform within the one-formant model of Fig. I-B-5, provides a solution differing from that of the sound pressure wave in front of the speaker by a constant factor only. Thus the F1 oscillatory component is

$$P_i(t) = -L U'_s e^{-\alpha_1 t} \cos(\omega_1 t + \varphi_1) \quad \text{Eq. (14)}$$

The amplitude of the negative pressure spike at glottal closure point T_3 is simply the evoked pressure

$L \frac{dU_s(t)}{dt}$ across the inductance L due to the rate of change of the current $U_s(t)$.

With the particular choice of pulse shape I have selected the flow derivative at closure is

$$U'_s(t=T_3) = -U_0 \cdot \omega_g \sqrt{2k-1} \quad \text{Eq. (15)}$$

where U_0 is the peak flow P_{sg}/R_0 and R_0 is the minimum static flow resistance which is half the differential resistance. The maximal value of instantaneous glottal bandwidth is thus

$$B_{go} = \frac{1}{2\pi \cdot 2R_0 C} \quad \text{and a corresponding } Q_{go} = \frac{F_1}{B_{go}}$$

$$\text{Thus } P_{io} = -U'_s \cdot L = P_{sg} \cdot \frac{\omega_g}{\omega_1} \cdot \frac{2}{Q_{go}} \cdot \sqrt{2k-1} \quad \text{Eq. (16)}$$

$$\text{and } \frac{U_q(t)}{U_0} = \frac{P_i(t)}{P_{sg}} \cdot \frac{R_{go}}{R_g(t)} \quad \text{Eq. (17)}$$

$$\text{or } U_g(t) = P_i(t)/R_g(t) \quad \text{Eq. (18)}$$

The ratio of the vocal tract input pressure to subglottal pressure $P_i(t)/P_{sg}$ can now be estimated. Assuming $F_1=600$ Hz, $F_g=150$ Hz, $B_{go}=300$ Hz, and $K=\frac{1}{2}$, we get a ratio of 25% which is of the right order of magnitude compared to experimental data from the Flanagan et al (1978) model and the direct measurements on a human speaker by Kitzing and Löfqvist (1975).

The damped F1 oscillation evoked at T_3 does not penetrate the glottal flow until next glottal open interval, when it has lost an appreciable part of its amplitude and is subjected to additional higher rate of damping and competes with secondary excitations at T_1 and T_2 .

Now, assuming $F_0=100$ Hz and a 5 ms duration of the glottal pulse with a peak location 3 ms after opening and a mean glottal bandwidth of 200 Hz in the rising part of the pulse, there would remain an F1 oscillation in U_g of the order of about 40 dB below the peak flow. This is rather small in view of the relative large glottal bandwidth. However, the gain in U_g current due to the low R_g is counteracted by the faster damping within the glottal pulse.

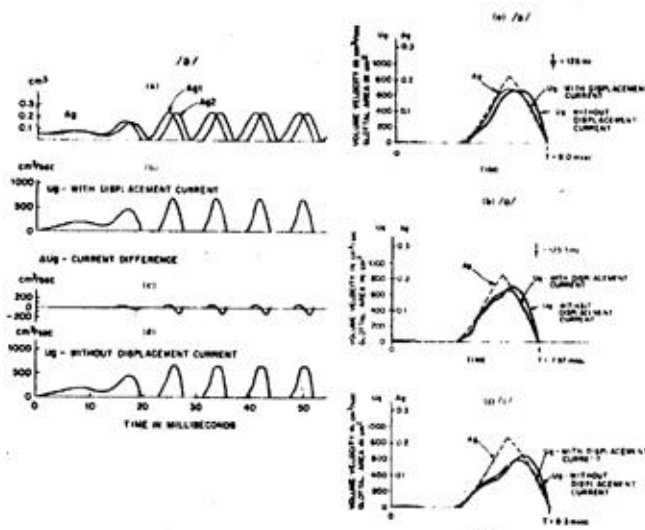


Fig. I-B-9 Flanagan and Ishizaka (1978) illustrations of glottal flow latency and the effect of displacement current.

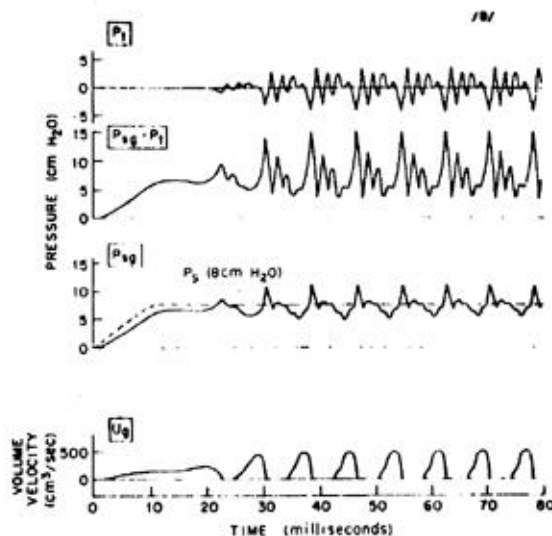


Fig. I-B-10 Supraglottal, transglottal, and subglottal pressure fluctuations (Flanagan and Ishizaka 1978)

A simpler derivation independent of the particular glottal flow model is the following:

$$\frac{U_g(t)}{U_o} = \frac{-U'_s \cdot L}{U_o R_g(t)} = \frac{-L}{T_d \cdot R_g(t)} \quad \text{Eq. (19)}$$

in which the derivate U'_s is defined as U_o/T_d and T_d is the pulse declination time, Fant (1979). We now need absolute measures of total vocal tract inductances. A narrow constriction has an inductance $g\ell/A$, where ℓ is the length and A the area. For a transmission line input instead of the single resonator implied in Fig. I-B-5 the expression for the input voltage response in Eq. (14) shall be modified substituting L for $(4\ell\Delta L)/\pi^2$, where ℓ is the total length of the tube and ΔL the inductance per unit length.

Furthermore, there enters an amplitude factor $(2n-1)^{-2}$ for $F_n = F_1, F_2, F_3$ etc. It should be observed that these damped oscillations start in the same phase and not in alternating phases as in the output. This is because of the alternation of poles and zeros in the input impedance. Thus, it would not pay to make a spectral analysis of the subglottal pressure ripple with LPC-techniques.

The total lumped vocal tract inductance would accordingly be $1.2g$ for a tube of 17.5 cm length and 6 cm² cross-section area and larger for any sound with local constrictions. From the tabulations in Fant (1960) we may quote ℓ/A ratios ranging from 2 for the vowel [ε] to 6 for [a] and 8 for [i], which implies that the initial amplitude of the starting negative spike in the input pressure varies with the vowel in proportion to ℓ/A . Furthermore, if we take into consideration the entire subglottal tract as a tube of 20 cm length and 2.5 cm² cross-sectional area it would have an ℓ/A ratio of 3.2. At the instant of glottal closure the subglottal pressure has reached a positive peak of $U'_s L_{sg}$, which is of the same order of magnitude as for the supraglottal negative spike. The particular ratios of sub- and supraglottal spikes implied by the particular ℓ/A ratios fit well with the Kitzing and Löfqvist data.

The superposition from one period to the next is more apparent for female than for male voices and, accordingly, we could expect a greater F1 component in the glottal flow. It remains to be seen. From inverse filtering experiments, Lindqvist (1965), and from the Flanagan model data we know that the glottal flow shows greater irregularity for [i] than for other vowels. This fact conforms with the greater lumped inductance for [i] and the less glottal damping which allows the residual F1 ripple to sustain within the glottal open period. However, the glottal flow waveshapes described by Flanagan and Ishizaka appear to contain more F1 ripple than we have seen from inverse filter experiments.

Another effect associated with the total lumped inductance in the vocal tract and its relation to glottal resistance is the latency of the glottal flow peak versus the glottal area function peak. The total flow inertia also includes that of the glottis which in case of maximal opening of 0.16 cm^2 and a depth of 3.2 mm of the glottal passage has an inductance of 2 g . With a representative glottal minimum resistance of 0.8 g c , the time constant of $(L_{\text{tot}})/R_{\text{go}}$ becomes of the order of 1 ms for the [i] and about 0.5 ms for a more open sound, which is of the right order of magnitude compared to data from the Flanagan-Ishizaka experiment, Figs. I-B-9, 10 and the inverse filter experiment of Fig. I-B-3. This flow latency due to the inductance was already observed by Flanagan and Landgraf (1968) and has been especially studied by Rothenberg (1980). One consequence of this inertia effect is the acceleration of glottal particle velocity and the increase of the steepness of the closing slope of the glottal pulse compared to the glottal area function and also compared to the assumed U_s source. One special problem is that of superposition. It was observed by Fant et al (1963) that in a series of phonations of sustained [æ] vowels with a large span of F_0 values F_1 rightly displayed maxima, when F_1/F_0 was a whole integer 1, 2, 3, 4 etc. The second formant amplitude, however, did not vary in amplitude synchronously with the F_2/F_0 ratio but seemed rather synchronized with the F_1/F_0 ratio. This could imply that superimposed F1 ripple in the glottal flow had increased the steepness of the pulse closure by an extent which was effective over the entire spectrum. These effects need to be studied in more detail.

Another secondary effect was described by Flanagan and Ishizaka (1978). Just before glottal closure the displacement current in towards the glottis, due to the upper parts of the glottis expanding during the contraction of the bottom part, counteracts the steepness of the descending slope and thus reduces the spectral level of the source somewhat. Holmes (1976) claims the presence of significant excitation during the interval of glottal closure. This also remains to be verified.

5. VOICE SOURCE PARAMETERS AND THEIR DYNAMICS

Most of this paper has been concerned with the problems of defining source functions of vowels and to describe source-filter interactions. For the practical use of a source model it should have parameters that have sufficient specificational power and can be readily quantified from acoustic analysis. These aspects and the dynamic variations in connected speech have been the topics of several articles in our quarterly reports, Fant (1979, 1980). I shall here only make some general comments.

There are many different combinations of three parameters that can specify a pulse shape. In retrospect I now feel that the K and F_g parameters are less important than (1) the direct measure of the closing steepness U'_s , which is the scale factor of all formant amplitudes and (2) the volume of air contained in a glottal pulse, which sets the amplitude level of the baseband peak.

Accordingly, within the constraints $U'_s = U_0 \omega_g (2K-1)^{\frac{1}{2}}$ of the source model, Fig. I-B-6, formant amplitude levels increase either by an increase in the asymmetry factor K or by an increase of F_g , i.e., by an overall shortening of the glottal pulse, which both result in increased U'_s . Spectrum shifts at constant source

peak F_0 and varying K and a selected normalizing $F_0=125$ Hz is shown in Fig. I-B-11 referring to initial amplitudes of formant oscillations and in Fig. I-B-12 referring to the +12 dB per octave pre-emphasized source spectrum (including special directivity and baffle effects, see Fant (1960, 1979b)).

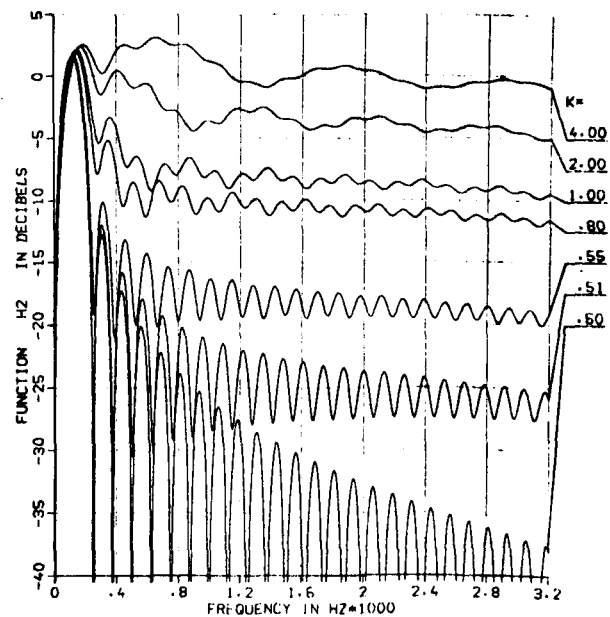
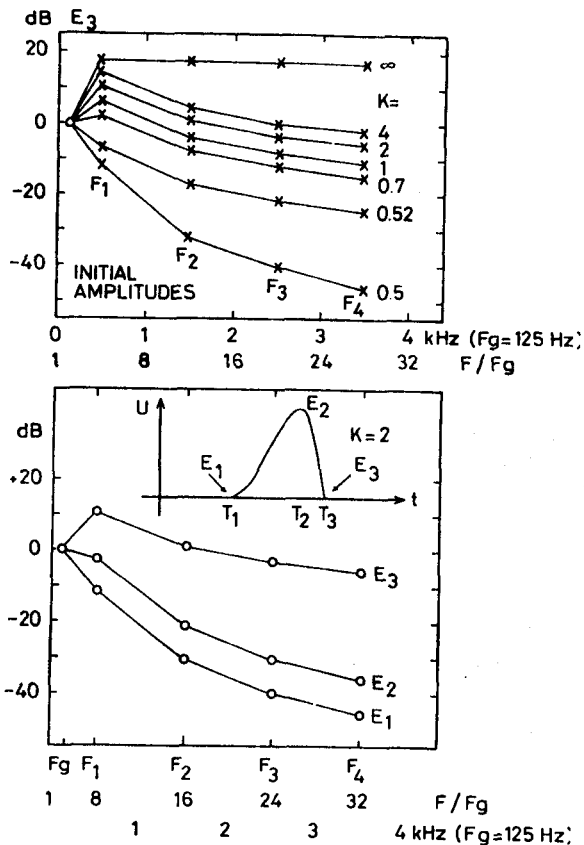


Fig. I-B-11 Initial amplitudes of each of F_1, F_2 and F_3 oscillations evoked at the glottal opening T_1 , peak T_2 , and closure.

Fig. I-B-12 The voice source spectrum variation with normalized frequency scale and varying steepness factor K .

At a low voice intensity level, the F_0 peak may override the F_1 peak in amplitude, see Fig. I-B-13, and this happens dynamically at the end of voicing in an open breathgroup final position. A reduction of F_1 initial amplitude and an increase of bandwidth cooccur as a result of abduction in aspirated voiced sound segments. Similar phenomena occur in nasalized boundary regions, Fant (1979, 1980). A certain extension of the source model may be necessary for representing the dynamics of incomplete closure within a voice period.

There appear to be two modes available to produce an intensity increase. One is a rise in the overall scale factor of glottal flow pulses which is a main consequence of increased subglottal pressure. The other is an adduction of the vocal cords physiologically induced by a medial compression, which may increase the steepness of the closing branch of glottal pulses while maintaining or even reducing the amount of air contained in a single pulse. This is typically the pattern found in syllabic stress, whilst an overall emphasis of a word or a part of a sentence appears to be accomplished by the first mentioned or by some combination of the two modes, compare the discussion of Rothenberg et al (1974) and

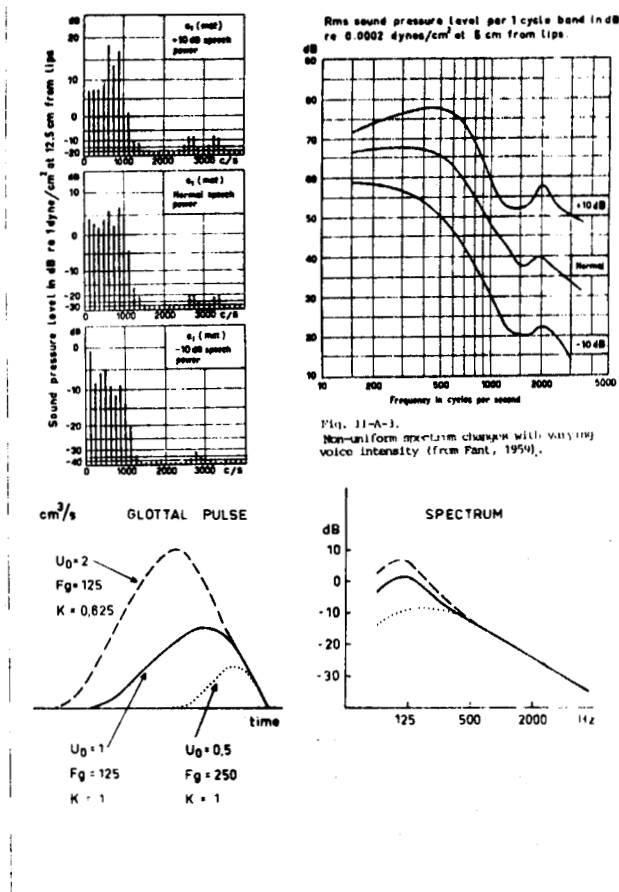


Fig. I-B-13. Non-linear spectrum shift at increasing voice effort. Below, three pulses with the same flow derivative at closure but varying absolute value of pulse areas.

Lindqvist (1965). However, the pilot study of Fant (1980) shows that a syllabic stress contrast need not be associated with a voice source change, the durational patterns being of main importance.

ACKNOWLEDGMENTS

Parts of the material discussed in this paper has appeared in recent years' articles in our quarterly reports. We have had a stimulating period of work the last year in Stockholm with Martin Rothenberg from Syracuse, USA as our guest. Some of the source-filter interactions described in this paper were initially discussed with him and my colleagues Johan Sundberg and Jan Gauffin.

REFERENCES

- Ananthapadmanabha, T.V. and Yegnanarayana, B. (1979): "Epoch extraction from linear prediction residual for identification of closed glottis interval", IEEE Trans., Vol. ASSP-27, No. 4, August.
- Carré, R. and Guérin, B. (1980): "Glottal impedance of a two mass model of the vocal cords", paper at the 99th ASA-meeting, Atlanta, USA.
- Fant, G. (1959): "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics No. 1.
- Fant, G. (1960): Acoustic Theory of Speech Production, Mouton, Haag (2nd edition 1970).

- Fant, G. (1979a): "Temporal fine structure of formant damping and excitation", paper at the ASA meeting in Cambridge, MA, USA, June.
- Fant, G. (1979b): "Glottal source and excitation analysis", STL-QPSR 1/79.
- Fant, G. (1979c): "Vocal source analysis - a progress report", STL-QPSR 3-4/79.
- Fant, G. (1980): "Voice source dynamics", STL-QPSR 2-3/80.
- Fant, G. and Liljencrants, J. (1979): "Perception of vowels with truncated intraperiod decay envelopes", STL-QPSR 1/79.
- Fant, G. et al (1963): "Formant amplitude measurements", J.Acoust. Soc.Am. 35, pp. 1753-1761.
- Fant, G. (1972): "Vocal tract wall effects, losses and resonance bandwidths", STL-QPSR 2-3/72.
- Flanagan, J.L. (1972): Speech Analysis Synthesis and Perception, Springer Verlag, Berlin and Heidelberg.
- Flanagan, J.L., Ishizaka, K., and Shipley, K.L.: "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", The Bell System Techn.J. 54, March 1975.
- Gauffin, J. and Sundberg, J. (1980): "Data on the glottal voice source behavior in vowel production", STL-QPSR 2-3/80.
- Fant, G. (1980): "The relations between area functions and the acoustic signal", Phonetica 37.
- Guérin, B., Mrayati, M., and Carré, R.: "A voice source taking account of coupling with the supraglottal cavities", paper 1975.
- Holmes, J.N. (1976): "Formant excitation before and after glottal closure", IEEE Conf. on ASSP, Philadelphia, PA, USA.
- Ishizaka, K. and Flanagan, J.L. (1972): "Synthesis of voiced sounds from a two-mass model of the vocal cords", Bell System Tech.J. 51.
- Kitzing, P.: "Methode zur kombinierten photo- und elektroglossographischen Registrierung von Stimmlippenschwingungen", Fol.Phon. 29, 1977.
- Kitzing, P. and Löfqvist, A.; "Subglottal and oral air pressures during phonation - preliminary investigation using a miniature transducer system", Med. and Biol. Eng. Sept. 1975.
- Lindqvist, J. (1965): "Studies of the voice source by means of inverse filtering", STL-QPSR 2/65.
- Miller, R.L. (1959): "Nature of the vocal cord wave", J.Acoust. Soc. Am. 31.
- Mrayati, M. and Guérin, B. (1976): "Etude de l'impédance d'entrée du conduit vocal-Couplage source-conduit vocal", Acustica 35.
- Flanagan, J.L. and Landgraf, L.L. (1968): "Self-oscillating source for vocal-tract synthesizers", Proc.IEEE-AFCRL, Boston, Mass. 1967.
- Rothenberg, M. (1973): "A new inverse-filtering technique for deriving the glottal air flow waveform during voicing", J.Acoust.Soc. Am. 53.
- Rothenberg, M. (1980): "Acoustic interaction between the glottal source and the vocal tract", Proc.Conf.Vocal Fold Physiol.Kurume.

Rothenberg, M., et al: "A three-parameter voice source for speech synthesis", Proc. Speech Communication (ed. Fant) Vol. 2, Almqvist & Wiksell, Stockholm.

Sundberg, J. and Gauffin, J. (1978): "Waveforms and spectrum of the glottal voice source", STL-QPSR 2-3/78.

Titze, I. (1980): "Comments on the myoelastic - aerodynamic theory of phonation", J.Speech and Hearing Res. 23.