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

Calculation of true glottal flow and its components

Ananthapadmanabha, T. V. and Fant, G.

journal: STL-QPSR
volume: 23
number: 1
year: 1982
pages: 001-030

KTH Computer Science and Communication

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

I. SPEECH PRODUCTION

A. CALCULATION OF TRUE GLOTTAL FLOW AND ITS COMPONENTS*

T.V. Ananthapadmanabha and G. Fant

Abstract

An iterative numerical algorithm is presented for computing the true glottal flow given the glottal area function and lung pressure. The effects of glottal inertance and sub- and supraglottal impedance are discussed. It is shown that the effect of glottal inertance is small and that it is adequate to consider subglottal and supraglottal systems by one-formant loads only. An equivalent circuit for the glottis considering the nonlinear and time-varying pressure-flow relation is derived. In addition to the dynamic glottal resistance, there exists a hypothetical inductance in the equivalent circuit which is mainly responsible for the skewing of the source pulses under load. An analytical equation for the main pulse shape of the true glottal flow is derived as the source residue. The role of ripple components in the true glottal flow is discussed.

1. Introduction

The volume velocity of air flow through the glottis, the glottal pulses, is the main source of voiced sounds. Glottal air flow as well as vocal folds vibratory patterns may be affected by changes in subglottal pressure and supraglottal articulations. In the classical approach for the synthesis of voiced sounds, viz., the source-filter modeling philosophy (Fant 1960), the interaction between the volume velocity source and the physiological parameters is assumed to be linearly separable. This approach is empirical but the implementation of synthesis is simple. On the other hand, an integrated vocal cord, vocal tract model (Flanagan 1968, 1972) preserves inherent constraints between the air flow through the glottis and sub- and supraglottal states. This approach is sophisticated but the practical implementation in synthesis becomes highly complex. From general experience in speech analysis we expect the source-filter interaction to be mainly confined to the first formant. A pragmatic model for synthesis may lie in-between these two extreme modeling philosophies, the simple formant analog and the complete interactive VT-model. Recent studies by Fant (1979, 1980, 1981) are aimed towards the development of a pragmatic model for the voice source. The present study is also in the same direction.

The success of an interactive source-filter model depends on its ability to represent the complex three-dimensional movements of the vocal folds by means of analytical equations. The early work of van den Berg (1957) provided us with insight in the glottal impedance and aerodynamic forces within the glottis during

---

* Parts of this paper were presented at the 102nd meeting of the Acoustical Society of America, Dec. 1-4, 1981, Miami Beach, FL, USA by G. Fant

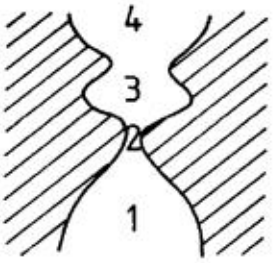
stationary flow. However, recent studies by Ishizaka and Matsudiara (1972), Gauffin et al (1981), Scherer (1981) have questioned the validity of van den Berg's classical formulae. We believe that the knowledge of aerodynamic forces within the glottis is not sufficiently well developed and that the vibratory pattern of vocal folds too complex and too detailed to be directly modeled in a practical synthesis scheme. For these reasons we base our calculations of volume velocity on the underlying glottal area function. Given the lung pressure and glottal area function, the problem is to compute the true glottal flow, and to study the effect of (a) glottal impedance, (b) subglottal system, and (c) vocal tract load on the glottal flow. In particular, we are developing analytical expressions to decompose the true glottal flow into its main pulse form and overlayed ripple components.

2. Glottal impedance

The glottis is a narrow constriction formed between two vibrating vocal folds, Fig. I-A-1. It may be noted that the lower and upper lips of the folds do not move in phase, the longitudinal component producing alternatively converging and diverging shapes. The excess air pressure in the lungs causes air to flow through the trachea and the glottis. This pulsating air flow not only causes acoustic excitation of vocal cavity resonances but is also responsible for sustaining vocal fold vibration. van den Berg proposed an empirical formula for predicting the steady (dc) flow based on extensive experiments on static larynx models with uniform glottis. His formula for the pressure drop ΔP is:

$$\Delta P = (k\varsigma U^2)/(2A^2) + (12\mu DlU)/A^3 \qquad \text{Eq. (1)}$$

where ς is the density of air, μ is the coefficient of viscosity, U is the volume velocity flow, P is the excess subglottal pressure, and A is the glottal area, D is the glottal depth, and l is the length of the glottis (see Fig. I-A-1). The constant $k=(k_1-k_2)$ is determined empirically. The form of the above equation is guided by aerodynamic theories. The first term in Eq. (1) is called the kinetic resistance drop and the second term is called the viscous drop. The constants k_1 and k_2 are the so-called entry drop and exit recovery coefficients. According to van den Berg, $k_1 = 1.375$, and $k_2 = 0.5$. Thus with Eq. (1), the volume velocity of air flow, U, can be calculated for given glottal dimensions and pressure. The above formula has been used in most of the studies on glottal flow reported so far. But, recent investigations, Ishizaka and Matsudiara, 1972, Scherer 1981, suggest a revision of the value for k_2, which according to them is much smaller, of the order of 0.01-0.05.



1. SUBGLOTTAL SYSTEM: TRACHEA
2. GLOTTIS
3. VENTRICLES & PHARYNX

Fig. I-A-1a

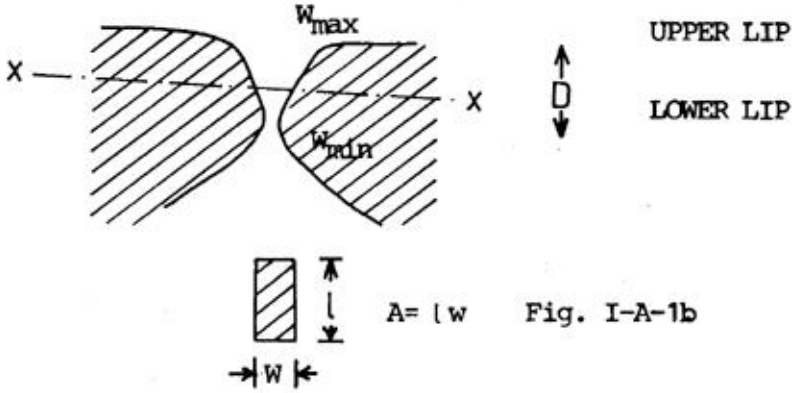


Fig. I-A-1b

Fig. I-A-1. The glottis.

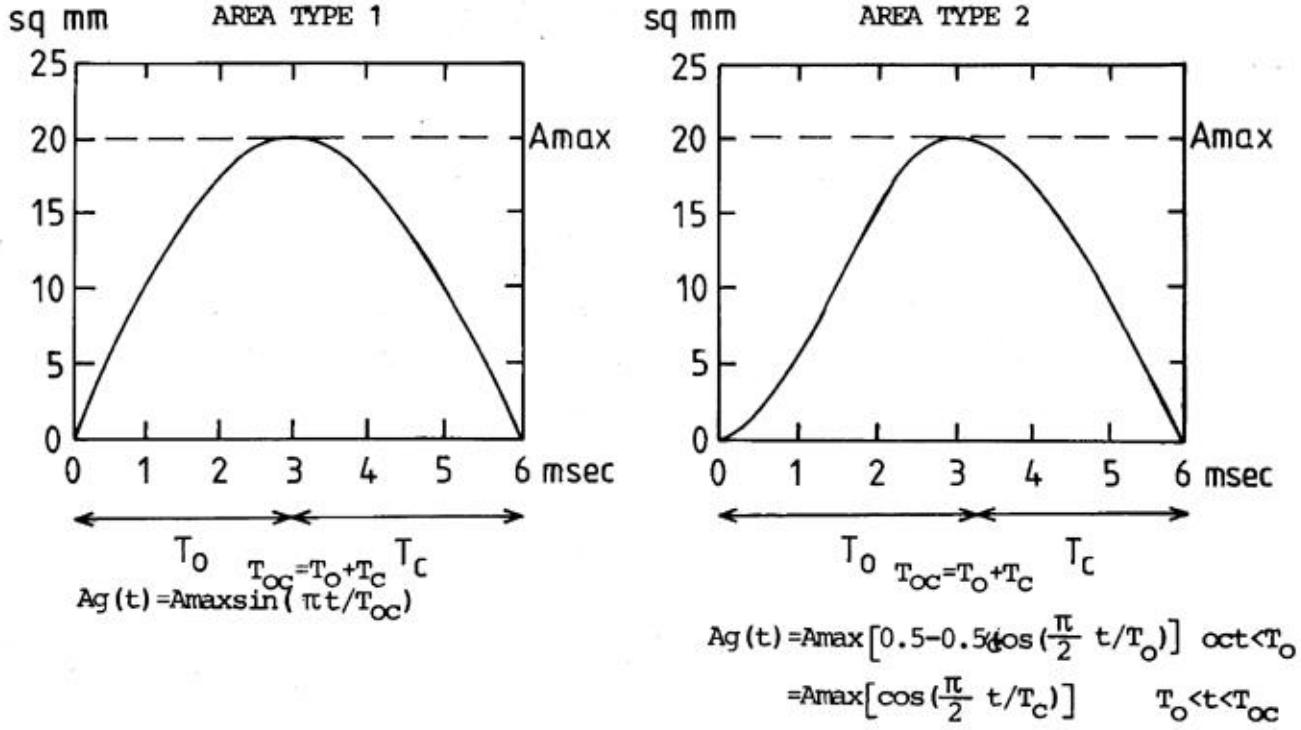


Fig. I-A-2. Typical glottal area functions.

The above discussion is restricted to the glottal flow through a uniform glottis. Experiments on nonuniform glottal shapes strongly indicate the dependence of glottal resistance on the shape of the glottis, Gauffin et al 1981. In our view, one major difference between the uniform glottis and the nonuniform glottis is that the viscous drop in the latter is much smaller. This can be visualized since the viscous drop term depends inversely on the width of the glottis, see Eq. (1). Based on the preliminary results of Gauffin et al (1981), we can to a first order approximation estimate the glottal resistance of a nonuniform glottis to be the same as that of the glottal resistance of a uniform glottis of width corresponding to the minimum width of the nonuniform glottis, but with the viscous drop term omitted. We propose a constant value for the kinetic resistance term k as 1.1. Thus for a static flow, we have the following equation:

$$\Delta P = (1.1\varsigma/2A^2)U^2 \qquad \text{Eq. (2)}$$

where A corresponds to the minimum area of the glottis. Photoglottography (Childers et al 1980, Kitzing 1977) gives us data directly on the minimum area of the glottis. The glottal area function in general depends on phonetic context, phonatory mode, voice intensity, pitch etc. These relationships are as yet not very well known. An extensive experimental work is needed. We have adopted reasonable estimates of glottal area functions in the chest register as a starting point for flow calculations. Two typical area functions are shown in Fig. I-A-2.

2.1 Glottal inertance

Eqs. (1) and (2) represent the glottal resistance for static flow. But, during phonation glottal area changes continually. Thus, under dynamic conditions the glottal impedance now incorporates an additional term corresponding to the glottal inertance. The role of the subglottal system and of the vocal tract load under dynamic conditions will be discussed subsequently.

Assuming a constant glottal depth D the glottal inertance is

$$L_g(t) = \varsigma D/A_g(t) \qquad \text{Eq. (3)}$$

Since L_g changes in time with $A_g(t)$, the pressure drop across L_g is

$$\Delta P_L = L_g dU_g/dt + U_g dL_g/dt \qquad \text{Eq. (4)}$$

The complete equation for the flow without load is given by

$$K_k U^2 + R_v U + L_g dU/dt + U dL_g/dt = P_{tg} \qquad \text{Eq. (5)}$$

where we have included R_v, the viscous resistance = $12\mu dl/A_g^3$, and $K_k = R_k/U_g = 1.1\varsigma/2A_g^2$ from Eq. (1). Glottal flow is calculated for area functions of type 2 and transglottal pressure P_{tg} = 8 cm water pressure. Different combinations of glottal impedance terms are used in the calculation. The results are shown in Fig. I-A-3. The effect of the viscous term is to render the flow to be more gradual both at onset and closure. In the spectral domain this implies a low-pass filtering effect above 2-5 kHz. The effect of glottal inertance is rather small and the net effect of the two terms is even smaller. When R_v is zero, the net effect of glottal inertance is exactly zero. This can be explained as follows:

$$\Delta P_L = d(L_g U)/dt = d(\varsigma D \cdot v)/dt \qquad \text{Eq. (6)}$$

where v is the air particle velocity = U/A. In the presence of only the kinetic term, the particle velocity is given by

$$v = (U_g/A_g) = \sqrt{(2P/k\varsigma)} \qquad \text{Eq. (7)}$$

which is a constant. Thus, ΔP_L which depends on the rate of change of velocity is zero. In the presence of a subglottal system and vocal tract load, however, the velocity is not a constant. The acceleration is typically 10^6cm/sec^2. Thus, ΔP_L is of the order of 0.3 cm water, which is small compared to transglottal pressure variations from induced pressures below and above the glottis. In all further discussions we represent the glottal impedance by the kinetic term only, Eq. (2).

3. Dynamic volume velocity flow

Under phonation the glottal area is continually changing and, hence, we have to compute the dynamic glottal flow. This cannot be made by considering the glottal area function as a sequence of quasistatic configurations, since, each quasistatic configuration assumes steady flow to have been reached. However, the transient pressure variations in subglottal and supraglottal cavities cause significant changes in volume velocity air flow. The equivalent circuit for calculating the dynamic glottal flow is shown in Fig. I-A-4.

3.1 Load representation

Both the subglottal cavity and the vocal cavity are distributed parameter systems. But, in this study we propose a lumped parameter representation for the

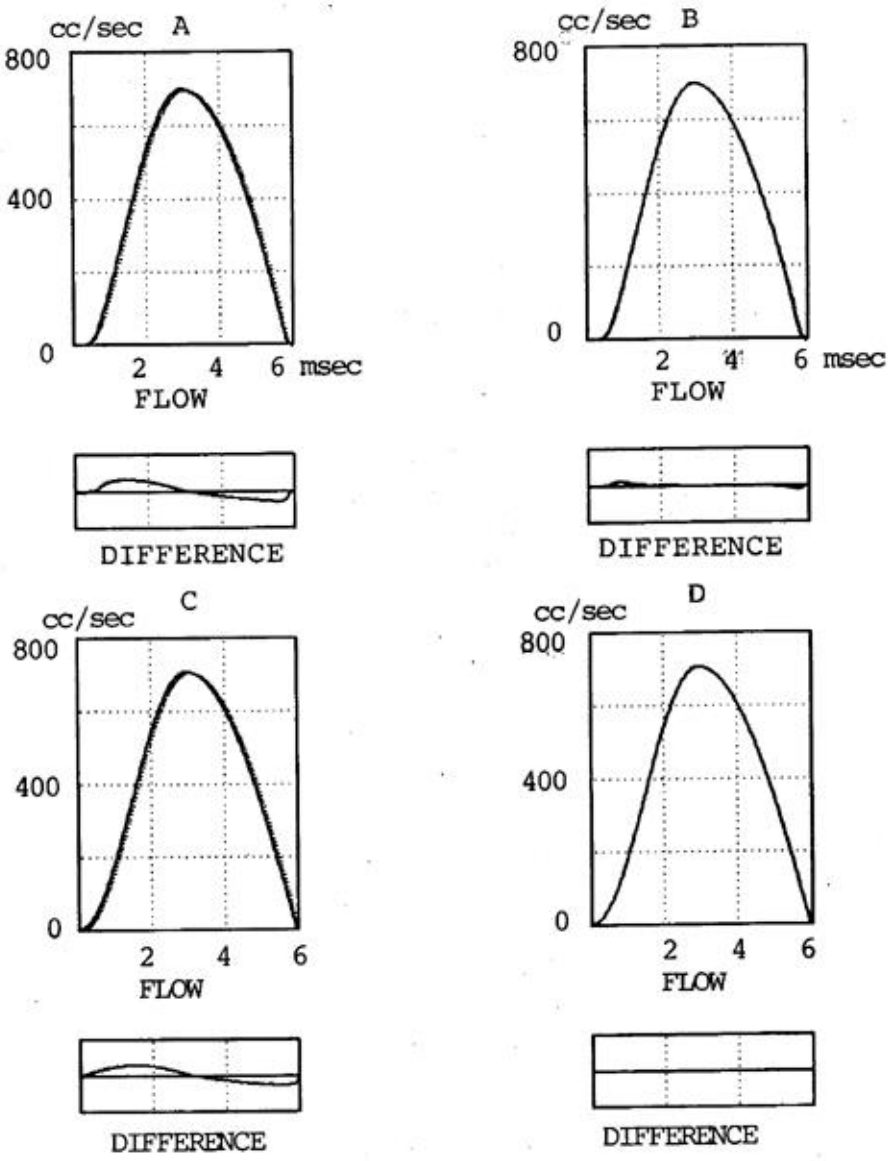


Fig. I-A-3. The effect of glottal impedance elements on the flow.

A. FULL: K_k and R_v DOTTED: K_k, R_v, LgdU/dt

B. FULL: K_k and R_v DOTTED: K_k, R_v, d(LgU)/dt

C. FULL: Only K_k DOTTED: K_k and LgdU/dt

D. FULL: Only K_k DOTTED: K_k and d(LgU)/dt

loads, a cascade of parallel 'RLC' tuned circuits, see also Guerin et al (1976). For the subglottal system we use the experimental data of the input impedance given by Ishizaka et al (1976). Data of formant frequencies, bandwidths, and inductance elements are given in Table I-A-I. The reason for specifying inductance instead of resistance, as done conventionally, will become apparent in Section 4.4.

TABLE I-A-I. Input impedance of subglottal system load

Resonant frequency Hz	615	1355	2110
Bandwidth Hz	246	155	140
Inductance mh	3.8	0.72	0.27

Based on the Wakita and Fant (1978) model for the vocal tract, the input impedance is calculated (U. Laine, personal communication) for vocal tract area functions of Russian vowels (Fant 1960). The lumped parameter equivalent circuit values are given in Table I-A-II up to the third formant.

TABLE I-A-II. Input impedance of vocal tract load

Vowel	Resonant frequency Hz	Bandwidth Hz	Inductance mh
/a/	659	32	6.14
	1060	31	1.80
	2418	56	0.12
/o/	524	35	5.73
	847	24	1.46
	2357	28	0.06
/u/	277	69	7.97
	611	23	1.39
	2374	20	0.03
/i/	269	63	7.31
	2257	22	0.38
	2876	26	0.07
/e/	444	31	2.77
	1940	49	0.41
	2716	192	0.43

3.2 Numerical algorithm

The volume velocity of air flow is related to the transglottal pressure through the nonlinear equation (2). The transglottal pressure will be the difference between the assumed constant lung pressure and the pressure drops across the load elements. Thus, from Fig. I-A-4C, applying the Loop Equation:

$$\left[(k\varsigma)/2A^2(t)\right] U_g^2(t) = P_{tg} = P_L - \sum_{i=1}^{m} V_{si} - \sum_{i=1}^{n} V_i \qquad \text{Eq. (8a)}$$

where V_{si} is the pressure drop across the i-th formant subglottal load, and V_i is the pressure drop across the i-th formant vocal tract load, given by

$$C_{si}dV_{si}/dt + V_{si}/R_{si} + (1/L_{si})\int V_{si}dt = U_g(t) \qquad \text{Eq. (8b)}$$

$$C_i\, dV_i/dt + V_i/R_i + (1/L_i)\int V_i\, dt = U_g(t) \qquad \text{Eq. (8c)}$$

The above set of equations have to be solved simultaneously. This can conveniently be done by a numerical iterative algorithm. Initially we assume $U_g(t)$ to be the no load current, $U_{sc}(t)$

$$\left[(k\varsigma)/2A^2(t)\right] U_{sc}^2(t) = P_{tg} \qquad \text{Eq. (9)}$$

The initial estimate $U_g^0(t) = U_{sc}(t)$ is substituted in Eqs. (8b) and (8c) to estimate $\left\{V_{si}^0(t)\right\}$ and $\left\{V_i^0(t)\right\}$. These values are substituted back into Eq. (8a) and a new estimate $U_g^1(t)$ is obtained. This iteration proceeds untill the solution converges to a desired accuracy. In our calculations, the iteration terminates when $\left|U_g^i(t) - U_g^{i-1}(t)\right| \leq 0.001 U_g^i(t)$. Interestingly, we have found that for almost all cases studied, the solution converges within a maximum of four iterations and the average number of iterations is about two. The subglottal pressure, P_s, and the oral pressure, P_v, are also available as by-products in the computation for glottal flow.

3.3 Study on vowels

A. Effect of subglottal system: The no load flow (full curve) is compared with the flow computed by including only the first subglottal formant load (dotted curve) in Fig. I-A-5A. The absolute maximum value in the difference signal is about 13% of the average glottal flow. The increased steepness of closure in the presence of subglottal load may be noted. The glottal flow computed by including successively the second and third formant loads of the subglottal system is shown in Figs. I-A-5B and I-A-5C. The difference signal is negligible. This shows that it is sufficient to use only the first formant load of the subglottal system to compute the flow.

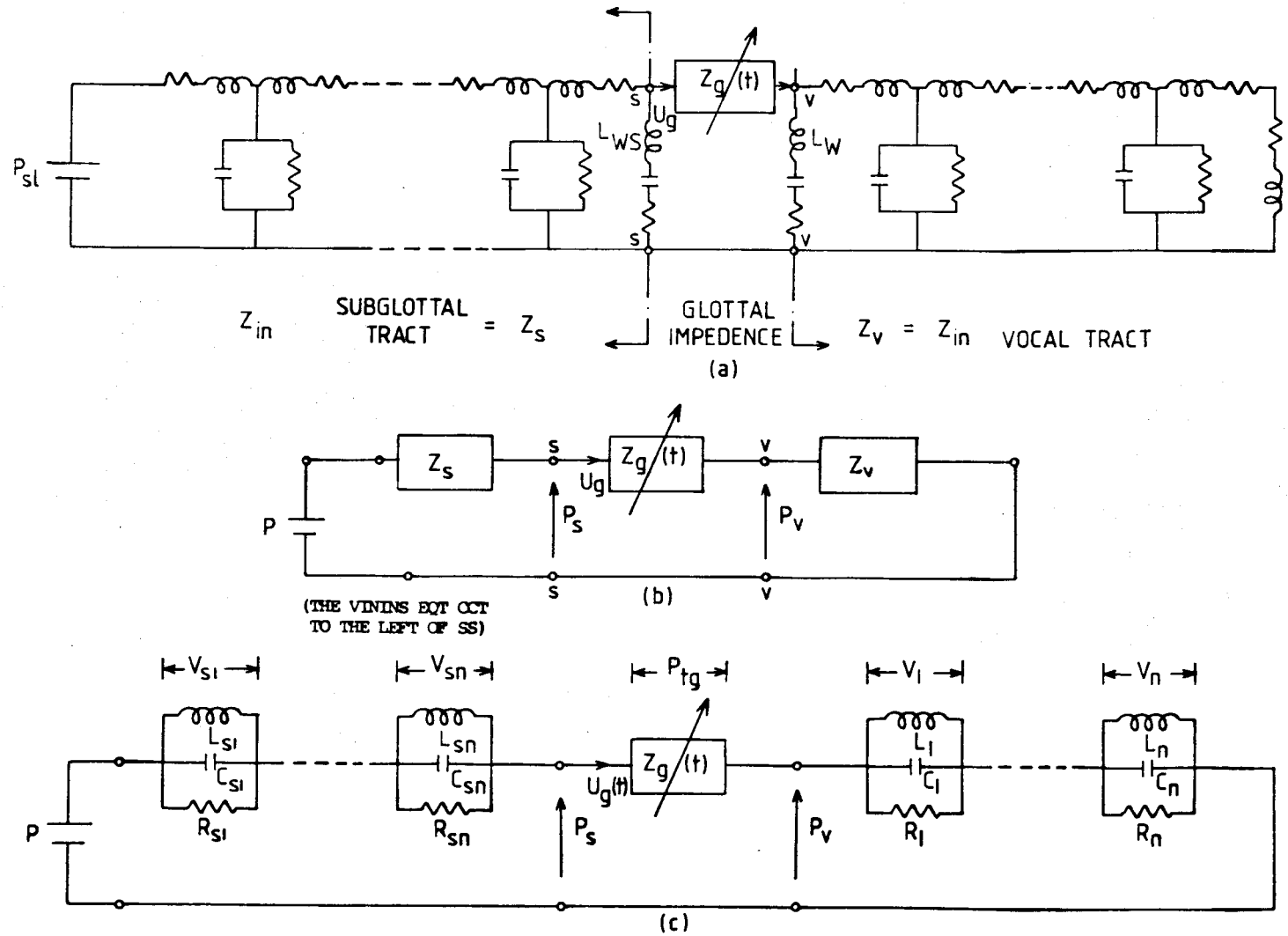


Fig. I-A-4. Equivalent circuit over open phase.

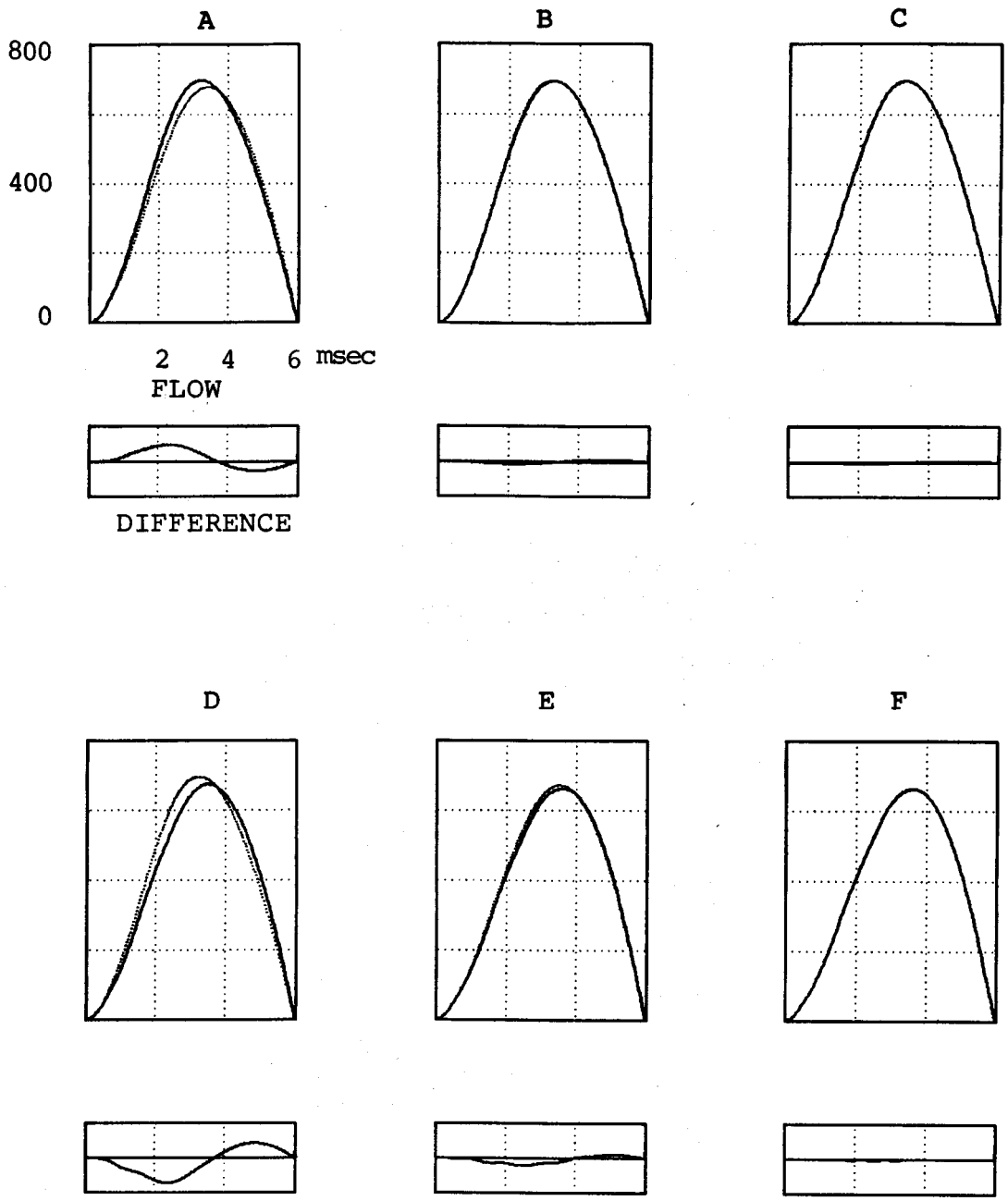


Fig. I-A-5. The effect of subglottal system and vocal tract load.

A: F_{1s}, no load

B: F_{1s}, $F_{1s}+F_{2s}$

C: $F_{1s}+F_{2s}$, $F_{1s}+F_{2s}+F_{3s}$

D: $F_{1v}+F_{1s}$

E: $F_{1v}+F_{2v}+F_{3v}$, $F_{1v}+F_{1s}$

F: $(F_{1v}+F_{2v}+F_{3v}+F_{1s})$, $(F_{1v}+F_{2v}+F_{1s})$

B. Effect of higher formants of vocal tract: The glottal flow is computed with subglottal system only. This flow (dotted curve) is compared with the glottal flow (full curve) calculated by including the first formant load of the vocal tract of the vowel /a/ (Fig. I-A-5D). The difference signal shows the effect of the first formant load of the vocal tract. It may be noted that the effect of the sugblottal system is comparable to the effect of the first formant load of the vocal tract. Now the glottal flow is computed by including the second and third formant loads, Figs. I-A-5E and I-A-5F. From the difference signal it may be noted that the effect of the higher formant loads is rather small.

C. Glottal flow for vowel sounds: The volume velocity air flow through the glottis for five vowel sounds (see Table I-A-II) is calculated by including three formant loads of the subglottal system and the vocal tract load. The results are shown in Fig. I-A-6. The glottal pulses for vowels /u/ and /i/ are distinctly different. The relative skewing of the pulses compares well with the recent results given by Matuasek and Batalov (1980) and Rothenberg (1980). In our calculations we have used the same glottal area function and the same open phase duration for all the vowels. In the actual practice, differences in pitch have to be considered. But, these calculations illustrate the relative nature of the source pulses for different vowels. The slope at closure is seen to depend on the total input inductance. Thus, for example, vowels /i/ and /u/ have a somewhat steeper closure compared to other vowels. Since the inductance associated with higher formant loads is small, the effect of higher formant loads on the volume velocity flow is negligible.

D. Effect of radiation, the cavity wall vibration: Radiation load, cavity wall vibration load, will influence the calculation of the input impedance of the vocal tract cavity. The input impedance of the vocal tract may be calculated with and without these effects. The results presented here are representative of the cases with these losses. The effect of radiation on the input inductance of the first formant load is very small, of the order of 5%. The effect of the cavity wall vibration is to reduce the input inductance, especially of the first formant. This effect is significant for some of the vowels, for example /u/. It is recommended that the input impedance be calculated with these effects. The role of the input inductance in shaping the volume velocity flow is discussed in Section 4.4.

E. Effect of superposition: In all the calculations presented so far we have assumed the load to be 'at rest'. In practice, since the excitation is periodic, there is a finite energy at the instant of glottal opening. To study the effect of

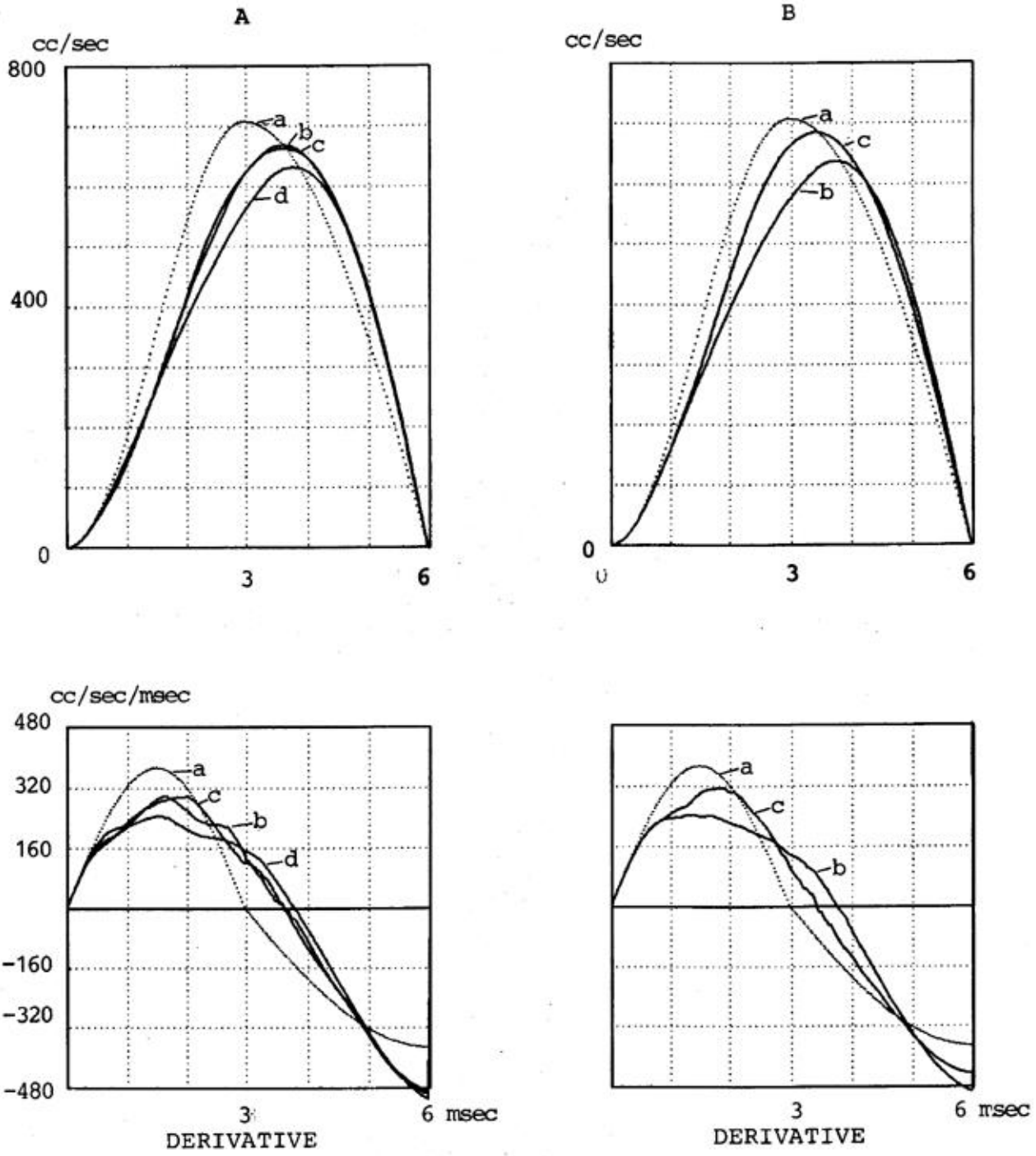


Fig. I-A-6. Source pulse and its derivative for vowel sounds.
A: a) No load, b) vowel /a/, c) vowel /o/, d) vowel /u/
B: a) No load, b) vowel /i/, c) vowel /ɛ/

the initial conditions, a single formant load is considered. The glottal flow is computed for the first period with zero initial conditions. Over the closed phase, the energy charged during the open phase discharges exponentially. By controlling the instant of the next glottal onset the effect of the relative phase of the formant oscillation on the flow may be studied. In Fig. I-A-7 the computed glottal flow for the second cycle is drawn over the glottal flow for the first cycle. When the first formant frequency is low, the initial conditions have a significant influence on the shape of the glottal pulse. Similar observations were reported by Flanagan and Landgraf (1968) and Flanagan et al (1975).

F. Spectra of glottal pulses: We next examine the effect of the vocal tract load on the frequency domain parameters of the source pulses. We have chosen the vowel /i/ for illustration since for this case the source pulse under load deviates markedly from the no load pulse. The source pulse and its log spectrum (up to 1600 Hz) for no load, with only subglottal load, and with sub- and supraglottal loads, is shown in Fig. I-A-8. The effect of subglottal load on the spectrum is small. The effect of the total load is to increase the asymptotic spectral level by about 3 dB and to reduce the spectral notch of the first zero. It may be noted that the spectral level of the second peak (second harmonic) attains a higher level with load. This is the frequency domain effect of the time domain skewing of the pulses. The effect of the load is seen more dramatically in the time domain than in the frequency domain.

4. Analytical approach

4.1 Pseudo-equivalent circuit for the glottis

We shall now develop an equivalent circuit for the glottis which has a time-varying nonlinear impedance. We refer to the equivalent circuit as a pseudoequivalent circuit since the circuit elements of the equivalent circuit will be considered to be linear and time invariant. This enables us to apply the Laplace transform methods to derive analytical equations for the glottal flow.

Consider a one-formant load as shown in Fig. I-A-9A. The Norton's equivalent circuit can be written as shown in Fig. I-A-9B, where $U_{sc}(t)$ is obtained by short-circuiting the load. We have from Eq. (2)

$$U_{sc}(t) = A_g(t)\sqrt{2P/k\varsigma} \qquad \text{Eq. (10)}$$

Let the Norton's source conductance be $g_{sc}(t)$. According to the linear theory,

$g_{sc}(t)$ would be $U_{sc}(t)/P$. But since the glottal flow pressure relation is non-linear, we have to derive an expression for $g_{sc}(t)$. Writing the nodal equation at the load terminal,

$$CdV/dt + V/R + (1/L)\int Vdt = U_g(t)$$
$$= A_g(t)\sqrt{2P_{tg}/k\varsigma} \qquad \text{Eq. (11)}$$

where $P_{tg} = P-V$ is the transglottal pressure. Assuming $V < P$, we get

$$CdV/dt + V/R + (1/L)\int Vdt + Vg_0(t)/2 = U_{sc}(t) \qquad \text{Eq. (12)}$$

where

$$g_0(t) = U_{sc}(t)/P$$
$$= A_g(t)\sqrt{\frac{2}{k\varsigma P}} \qquad \text{Eq. (13)}$$

The above equation is the nodal equation referring to Norton's equivalent circuit. Thus $Vg_0(t)/2$ can be interpreted as the current through $g_{sc}(t)$. Hence, we identify $g_{sc}=g_0/2$. This corresponds to the well known result that the dynamic resistance of the glottis is twice the static resistance.

In the above derivation we have considered only the nonlinearity of the glottal conductance. But $g_0(t)$ is also time-varying. Differentiating Eq. (12) wrt't' we get

$$d^2V/dt^2 + \left[1/R + g_0(t)/2\right](1/C)dV/dt + (1/LC)\left[1 + .5Lg_0'(t)\right]V = (1/C)\,U_{sc}'(t) \qquad \text{Eq. (14)}$$

The above equation is a linear differential equation with time-varying coefficients. The standard Laplace transform methods cannot directly be applied. Interpreting the above equation as a load with an undamped resonant frequency

$$\omega_1^2 = \omega_0^2\left[1 + 0.5Lg_0'(\tau)\right] = \left[1/LC\right]\left[1 + 0.5Lg_0'(\tau)\right] \qquad \text{Eq. (15)}$$

and a damping coefficient

$$2\alpha_1 = \left[1 + 0.5Rg_0(\tau)\right]\,(1/RC) = 2\alpha_0\left[1 + 0.5Rg_0(\tau)\right] \qquad \text{Eq. (16)}$$

we can write a pseudo-Laplace transform as

$$V(s) = \omega_0^2 LsU_{sc}(s)/\left[s^2 + 2\alpha_1 s + \omega_1^2\right] \qquad \text{Eq. (17)}$$

We replace the time variable 't' by 'τ' to signify that we assume the equivalent glottal impedance to be stationary for writing the pseudo Laplace transform.

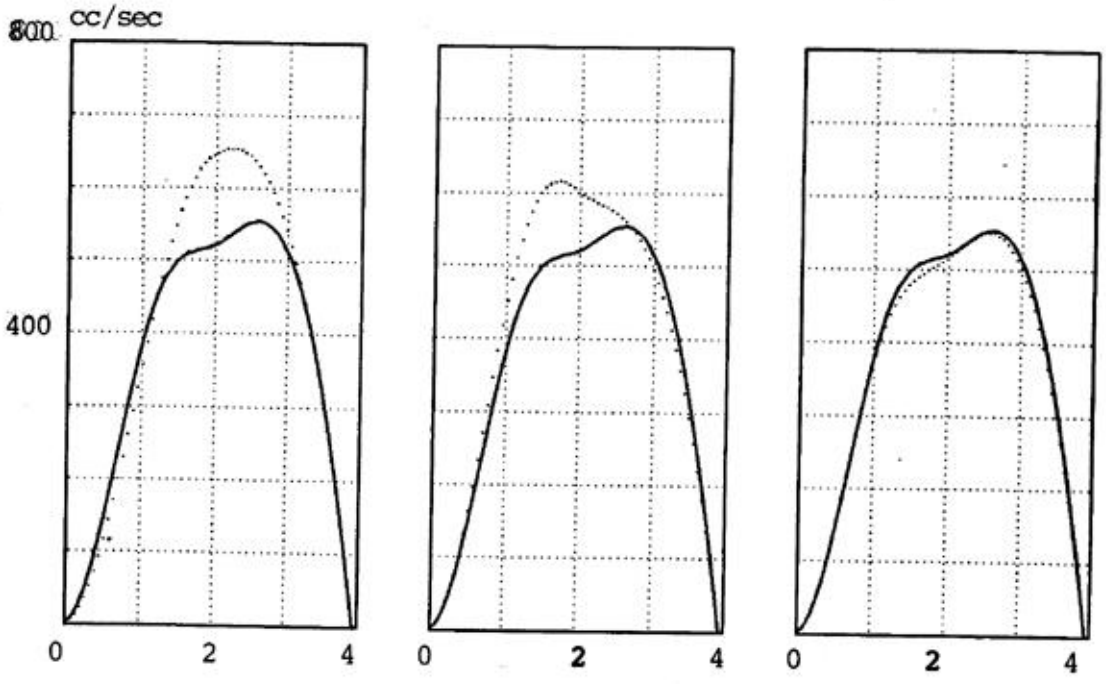


Fig. I-A-7. The effect of the superposition component on the glottal flow. Open phase = 4 msec. Load: F_1 of /i/.
A: closed phase = 2.5 msec, B: closed phase = 4 msec,
C: closed phase = 5.5 msec.

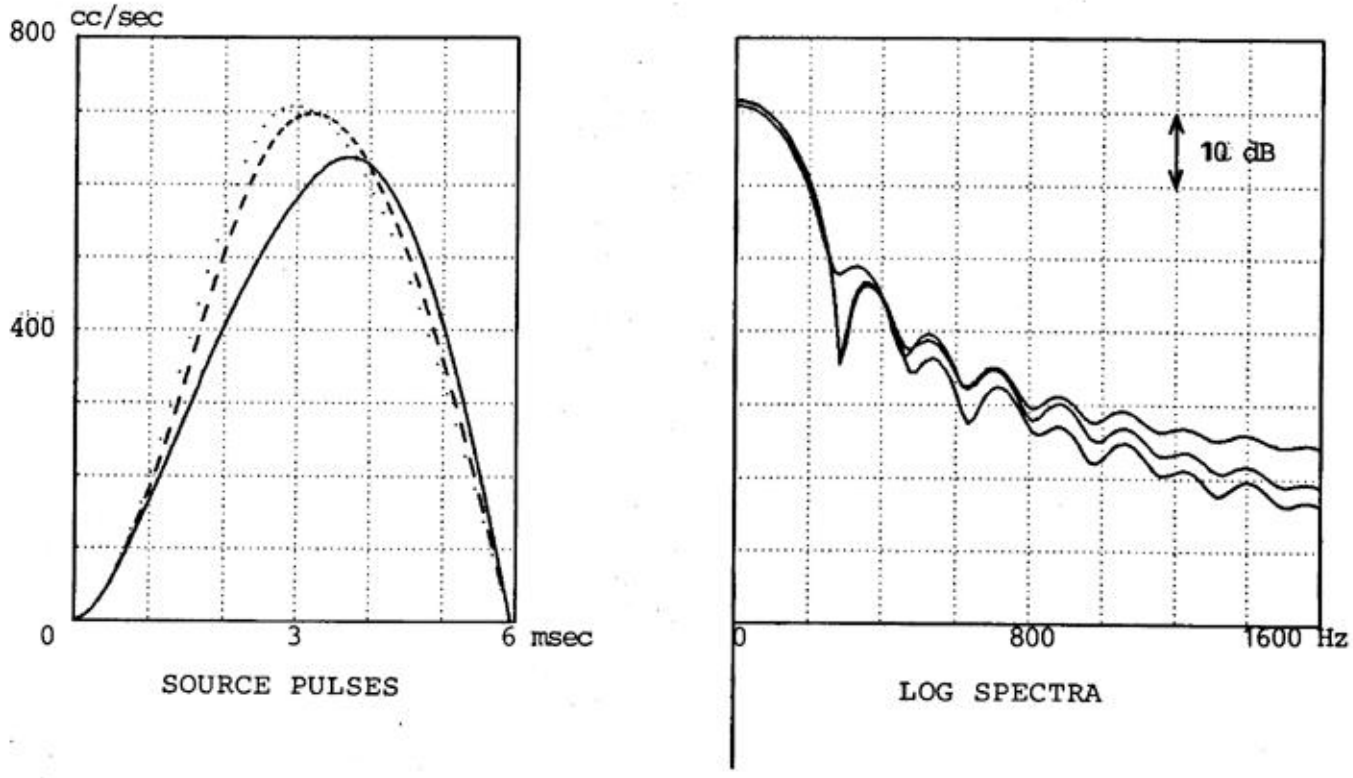
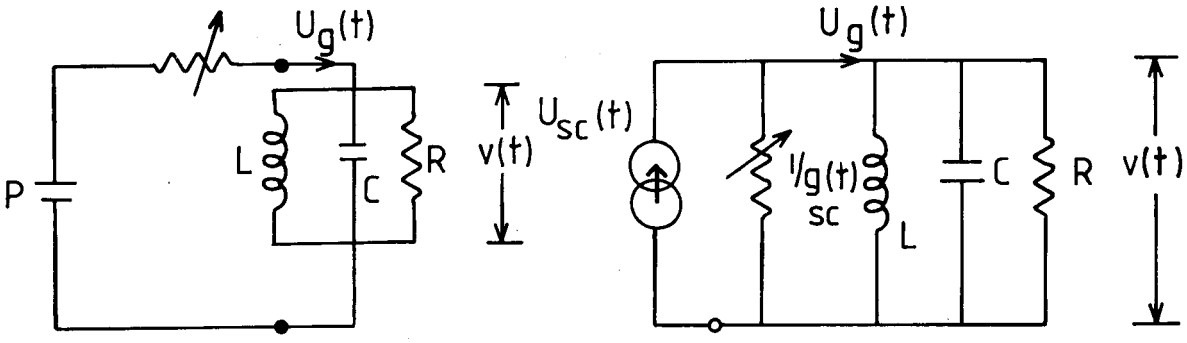


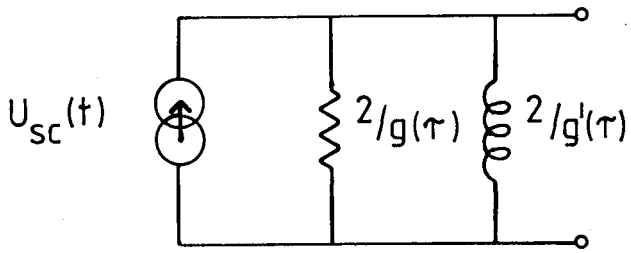
Fig. I-A-8. The source pulse and its spectrum.



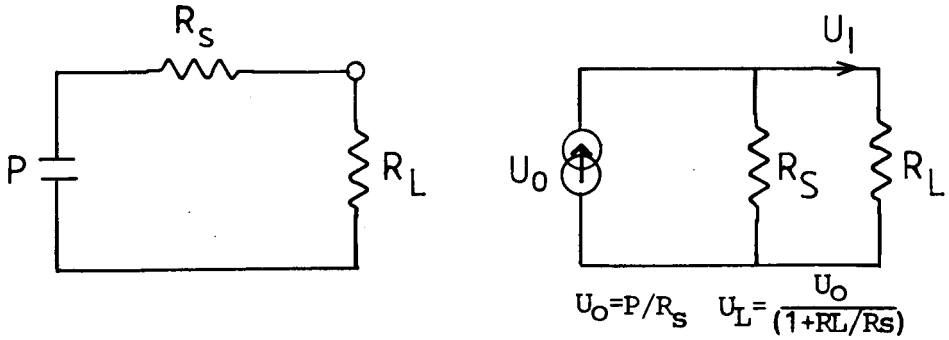
A. One formant load

$U_{sc}(t)$: short circuit flow $U_g(t)$: true flow
$g_{sc}(t)$: Norton's source conductance

B. NORTON'S EQT CCT



C. PSEUDO EQUIVALENT CIRCUIT



D. A SIMPLE DC CIRCUIT

Fig. I-A-9. Equivalent circuit for the glottis.

Thus, according to the nonlinear theory, we find the source conductance to be $0.5g_0(\tau)$, which causes a bandwidth modulation and due to the time-varying nature of $g_0(t)$, we find an equivalent hypothetical inductance of value $2/g_0'(\tau)$ which causes a frequency modulation of the resonant frequency of the load. The pseudoequivalent circuit is shown in Fig. I-A-9c.

4.2 Components of true glottal flow

The true glottal flow entering the load is related to the short circuit flow in an interesting manner (Fant 1981). Equating the Laplace transform of V(t) obtained from Eq. (11) with Eq. (17), we get

$$U_g(s) = U_{sc}(s)\left[\frac{s^2 + 2\alpha_0 s + \omega_0^2}{s^2 + 2\alpha_1 s + \omega_1^2}\right] \qquad \text{Eq. (18)}$$

In the time domain, $U_g(t)$ has three components:

(a) Source residue component - This arises due to the poles of $U_{sc}(s)$.

(b) Ripple component - This arises due to the time-varying complex conjugate pole pair. In a time-domain analysis this component can also be interpreted as the source for the transient response of the F_1 load, where the transients are excited at the points of discontinuity or epochs in $U_{sc}(t)$.

(c) Superposition component - Eq. (18) is derived assuming zero initial conditions of the load. The finite initial conditions of the load causes yet another component in $U_g(t)$. In a time-domain analysis this component of the current is responsible to simulate the transient response of the load during the open phase.

We shall discuss the time domain and frequency domain properties of these components in greater detail in Sections 4.4 and 4.5.

4.3 Candidates for voice source

As discussed by Fant (1981) there are two candidates for the voice source, viz., the short circuit source and the true glottal flow. The use of the short circuit source $U_{sc}(t)$ has the advantage that it can be determined independent of the articulation. Thus $U_{sc}(t)$ can be modeled directly from a knowledge of glottal area function and lung pressure. In this view, the subglottal system is also considered as a load in series with the vocal tract load. But, the use of $U_{sc}(t)$ as a source complicates the specification of the transfer function of the system: separate transfer functions have to be specified for open and closed phases of the glottal cycle. Alternatively, using true glottal flow for the voice source requires the specification of only the closed phase transfer function.

This also has the advantage that the voice source can be studied experimentally using inverse filtering. But, now the true glottal flow depends on the articulation. Also true glottal flow has ripple components which makes the modeling of the voice source difficult.

The short circuit flow and the true glottal flow for the F_1 load of the vowel /i/ is shown in Fig. I-A-10. The true glottal flow includes the effect of initial conditions on the load. It may be seen that the source-filter interaction has a significant effect on the short circuit flow. The log spectra of the flow pulses are also shown in the figure. From Eq. (18) it follows that the log spectral difference should correspond to the response of a pole-zero pair. The log spectral difference has almost an all pass trend. The spectral dip and the peak in the neighborhood of F_1 has a dynamic range of about 20 dB. Also, there is a 3 dB asymptotic increase in the spectral level due to the increased steepness of the closure of the true glottal flow. The above illustration shows the importance of considering F_1 - load interaction with the source. However, for the purpose of modeling the true glottal flow, we have to study the components of the flow in greater detail.

4.4 Source residue

Of the three components of the true glottal flow, the source residue component is the most dominant and it describes the main pulse shape of the true glottal flow. We will now derive an analytical expression for the residue component considering a one-formant load. Let $U_{sc}(t)$ be a finite duration signal with continuously differentiable derivatives in the open interval of the pulse. Then $U_{sc}(s)$ has poles only at s=0. The source residue of $U_g(s)$ at s=0 can be written as (see Appendix-A for the derivation)

$$U_r(t) = U_{sc}(t)\left[\frac{1}{1+.5Lg'(t)}\right] - U'_{sc}(t)\left[\frac{0.5L\{g(t)-g''(t)/R\}}{\{1+.5Lg'(t)\}^2}\right] + \ldots \quad \text{Eq. (19)}$$

It may be noted from Eq. (19) that $U_r(t)$ is determined mainly by the inductance element of the load. This suggests a convenient method for approximately finding $U_r(t)$, viz., to compute the flow assuming only inductive elements of the load. Thus by solving the following equation we get an estimate of $U_r(t)$:

$$\left[k\varsigma/2A^2(t)\right] U_r^2(t) + LdU_r(t)/dt = P \qquad \text{Eq. (20)}$$

The above derivation is the theoretical basis for Rothenberg's method (Rothenberg

1981). However, Rothenberg models the glottal conductance instead of the glottal area function and uses a linear equation. Also the value of inductance to be used in Eq. (20) according to Rothenberg does not correspond to the total input inductance.

The residue component of the flow computed based on Eq. (19) for a half sine area function (area type 1) is compared with the true glottal flow in Fig. I-A-11A. Also the residue components computed based on Eqs. (19) and (20) are compared in Fig. I-A-11B. In spite of the pseudo Laplace transform method used in deriving Eq. (19), the residue component gives a good match to the true glottal flow.

Based on Eq. (20), the computed residue component is compared with the true flow for the area function of type 2 in Fig. I-A-12A. We note that the residue component describes the smooth outline of the pulse accurately. The log spectral difference has a zero level at high frequencies since the residue component gives the correct slope of closure. The spectral dip and peak in the neighborhood of F_1 has now a lower dynamic range. Fig. I-A-12B shows a comparison of the true glottal flow and the glottal flow with ripple component but not the superposition component.

The glottal flow minus the residue component correspond to the sum of the ripple component and the superposition component. Conventionally, these ripple components are supposed to be narrow band F_1 ripple. The log spectrum shows the ripple component to be a very broad, bandpass type signal. Further diesuccion on ripple components is given in Section 4.5.

From Eq. (19), retaining only the first term on RHS gives

$$U_r(t) = U_{sc}(t) / \left[1 + .5Lg'(t)\right] \qquad \text{Eq. (21)}$$

The above equation is very powerful in predicting many aspects of volume velocity flow: (i) Since $g'(t)$ is positive over the opening phase and negative over the closing phase, $U_r(t)$ rises more gradually but falls more steeply than $U_{sc}(t)$. Also the steepness of the closure depends on the load inductance. (ii) Consider a simple dc circuit as shown in Fig. I-A-9D. The load current U_1 is related to no load current U_0 as

$$U_1 = U_0 / \left[1 + R_1/R_s\right] \qquad \text{Eq. (22)}$$

where R_1 and R_s are, respectively, the load and the source resistances. When $R_1 << R_s$, the current regulation is good and $U_1 \simeq U_0$. By an analogue of Eqs. (21) and (22), we can say that the source-filter interaction is strongest when the load inductance is comparable to the hypothetical inductance of the glottis, $2/g_0'(t)$.

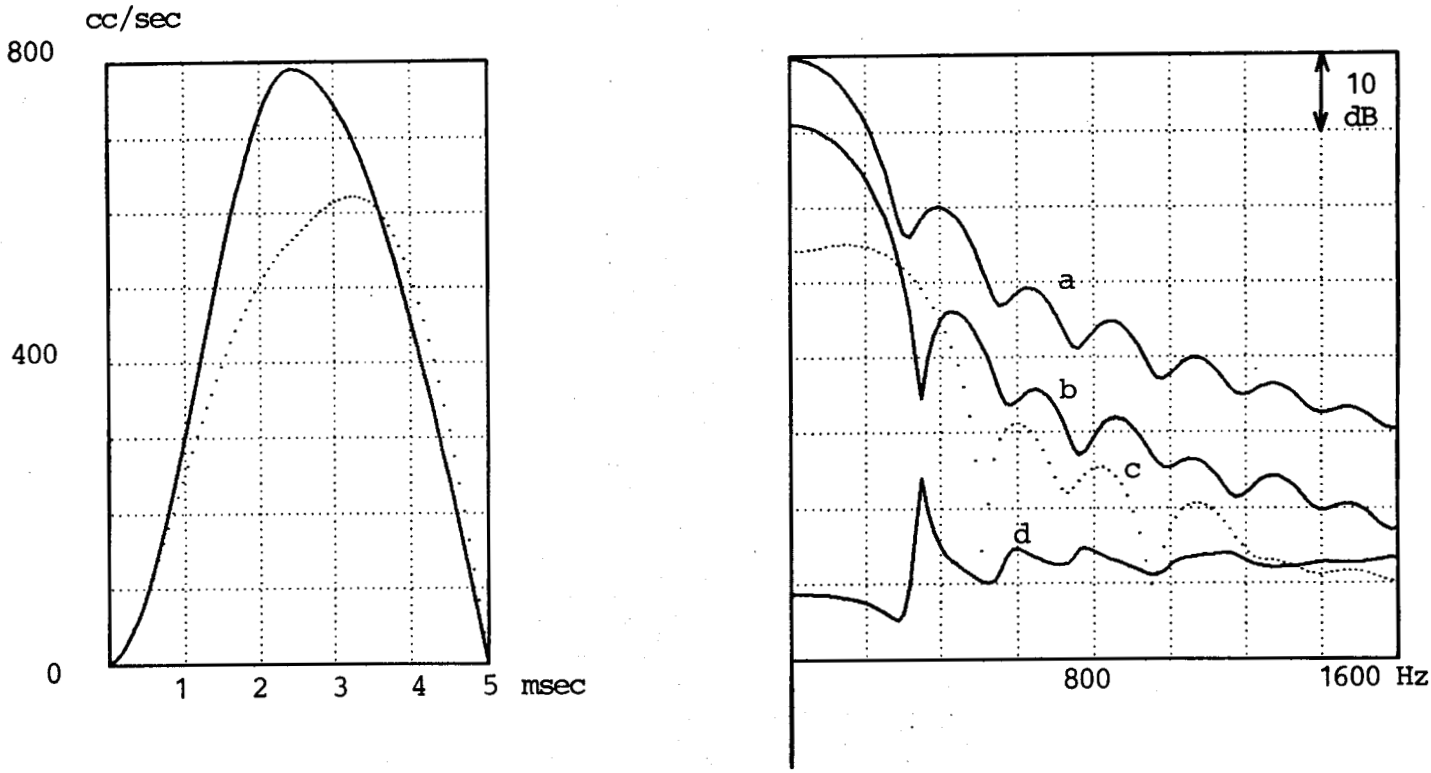


Fig. I-A-10. Source-filter interaction.
A: FULL: No load flow $U_{sc}(t)$; DOT: Flow with load $U_g(t)$
B: a) log spectrum of $U_g(t)$, b) log spectrum of $U_{sc}(t)$
c) log spectrum of difference component $U_g(t)-U_{sc}(t)$
d) (a)-(b)

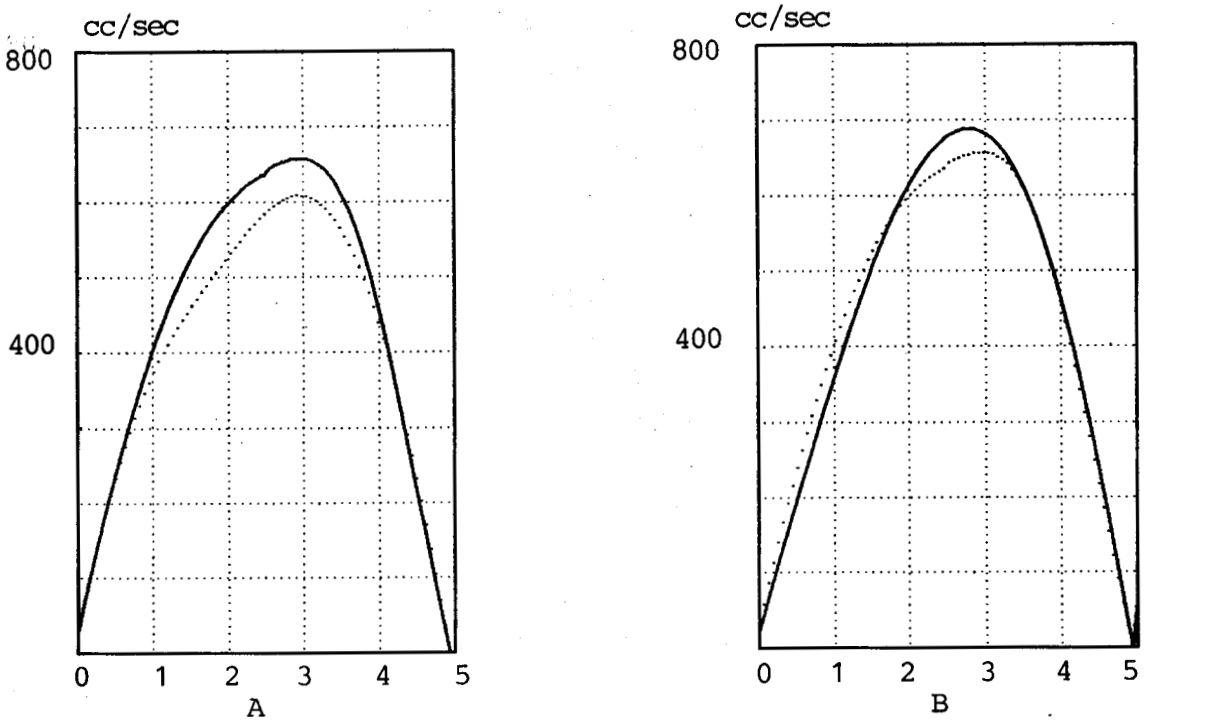


Fig. I-A-11. Residue component of glottal flow.
A: FULL: Residue DOT: True flow
B: FULL: Flow for only inductive load DOT: Residue

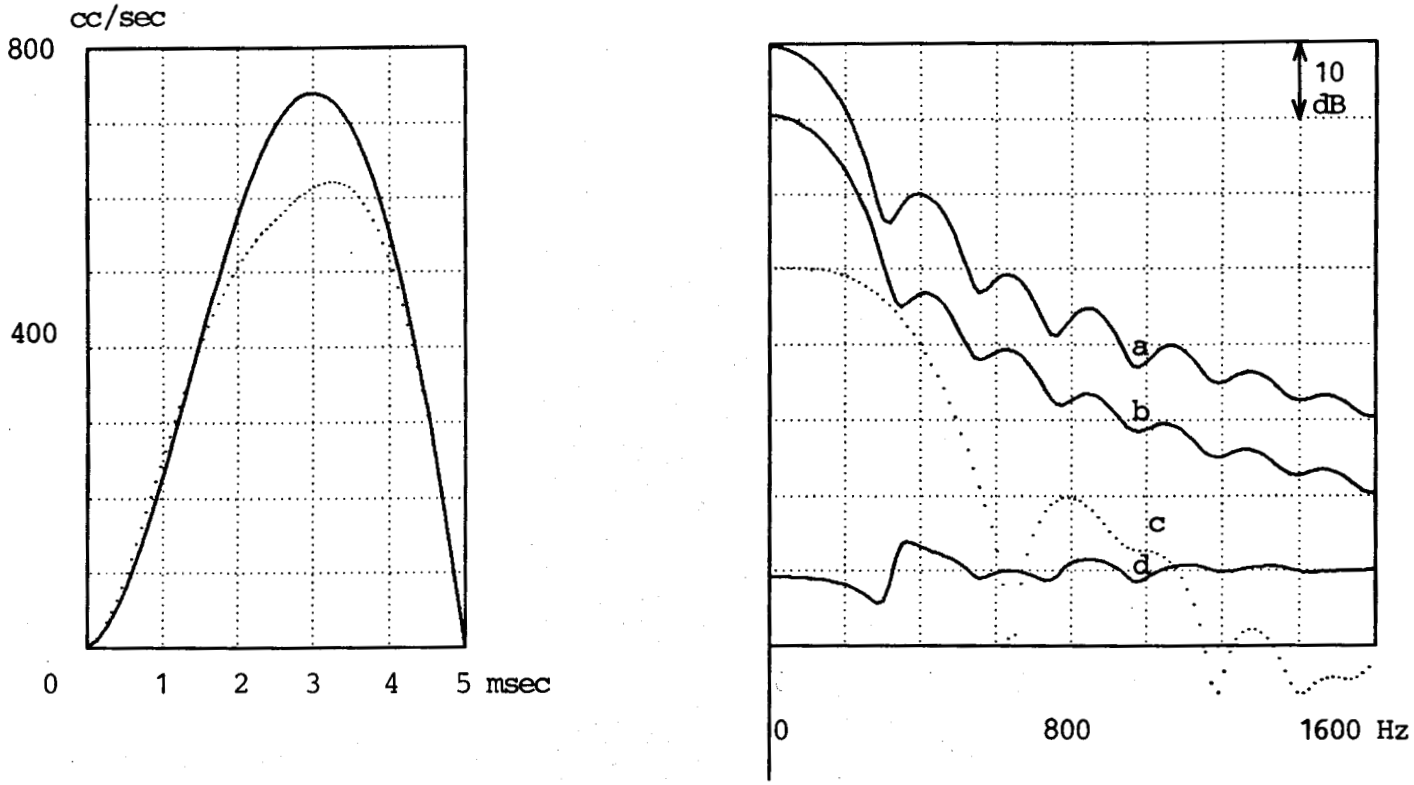


Fig. I-A-12A.

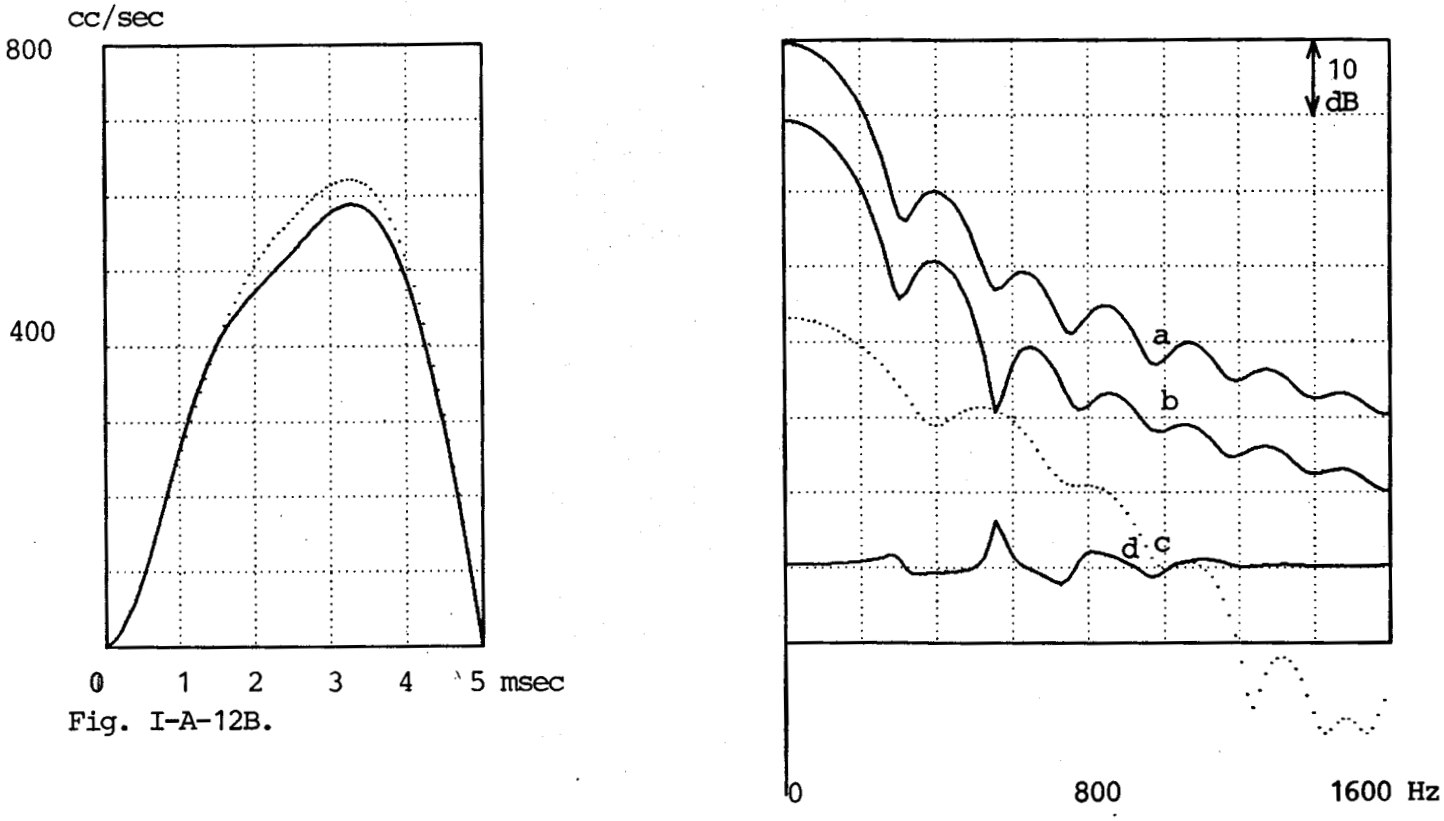


Fig. I-A-12B.

Fig. I-A-12. Components of the glottal flow.
A: True flow (dot) compared with residue
B: True flow (dot) compared with flow without superposition component.

For example, for the low pitch case, with the peak glottal area as 20 mm^2, $g_0'(t)$ maximum is about 40. Thus, $.5Lg'(t)$ is about 0.1 for 5 mh load inductance. (iii) Let $g_0(t)=G_0\sin(\omega_g t)$, $U_{sc}(t)=U_0\sin(\omega_g t)$, then $U_r(t)\simeq U_0\sin(\omega_g t)-\left[G_0U_0L\omega_g/4\right]\sin(2\omega_g)$. This shows that the main effect of load in the frequency domain is to control the amplitude of a $2\omega_g$ component. This is the frequency domain interpretation of time domain skewing of the source pulse. (iv) Dividing both sides of Eq. (21) by $A_g(t)$ we have a relation between the particle velocity under load and the no load constant particle velocity.

4.5 Role of ripple component and superposition component

The voltage response V(t) has both a residue component and transients, as can be seen from Eq. (17). Since the transient response is also the force-free response, there is no ripple current in the input short circuit current $U_{sc}(t)$. But, the ripple component is present in true glottal flow $U_g(t)$. If we disconnect the source-filter coupling, this ripple component of the true flow is necessary to represent the effect of a time varying bandwidth and resonant frequency modulation due to the source coupling over the open phase. (See Appendix-B for mathematical reasonings.) The same argument holds for the superposition component. In the absence of the superposition component, the excitation due to the previous cycle would undergo an exponential damping over the open phase. To simulate the coupling effect on the initial conditions, the superposition component is necessary.

The impulse response of an F_1 load circuit is deduced using periodic pulses of (a) the true glottal flow, and (b) the residue component, Fig. I-A-13. When the residue component is used for excitation, the response can be seen to possess only an exponential damping. However, the superposition component in the true glottal flow correctly simulates the source coupling effect.

4.6 A pragmatic model for the voice source

The discussion presented in the previous section suggests an interesting alternative for defining a voice source. We could use the source residue component of the true glottal flow omitting the ripple component and the superposition component. Then, the F_1 resonant circuit should be designed to have the required bandwidth modulation. Referring to the equivalent circuit of the glottis, Fig. I-A-9, with a one-formant load only, the maximum instantaneous glottal bandwidth component is given by $0.5\omega_1^2LG_0$, where G_0 is the peak glottal conductance, which is of the order of 0.1 mho. For high F_1 vowels, such as /a/, this maximum instantaneous bandwidth could be as large as 500 Hz, thus causing almost a complete truncation of the F_1 response, Fig. I-A-13A.

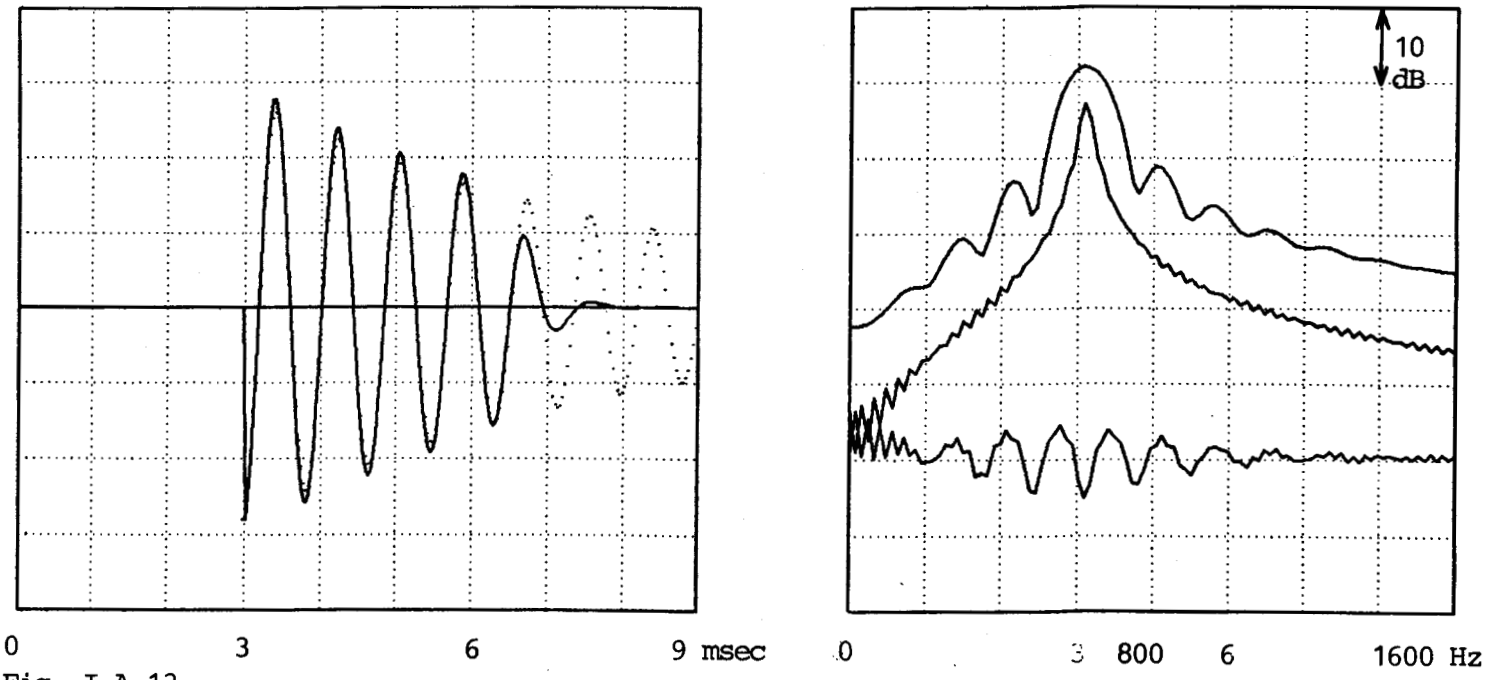
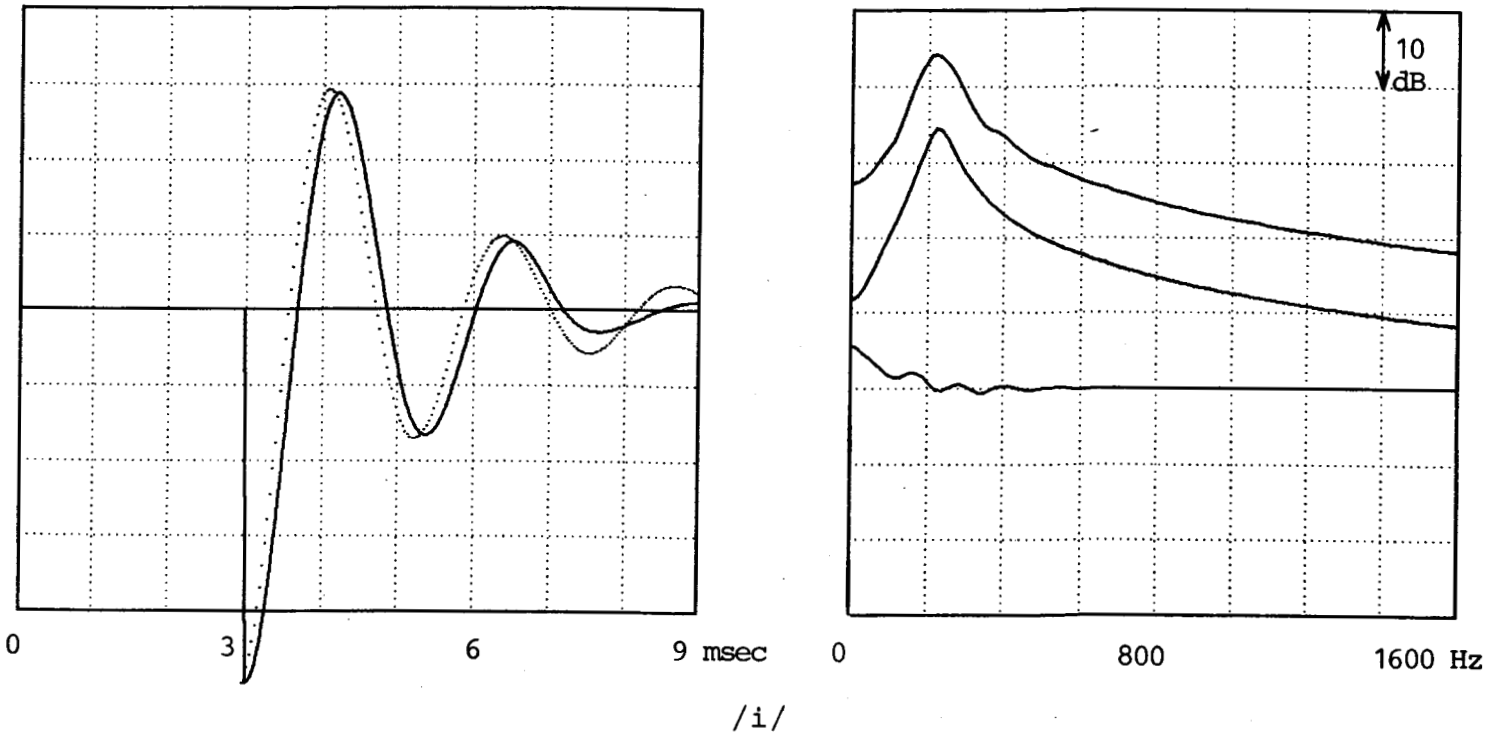


Fig. I-A-13.

/a/



/i/

Fig. I-A-13. Impulse response of F_1 load coupled to glottis only during open phase.

An interesting problem concerns the perceptual importance of this "truncation effect". Does the time course of the formant oscillation envelope need to be preserved in detail or does it suffice to increase the bandwidth of an equivalent conventional decay to compensate for the excess damping in the open phase? Perceptual matching experiments on truncated periodic F_1 responses with untruncated responses (Fant and Liljencrants 1979) suggest that subjects tend to match for an equal loudness level of F1 which implies equal mean value of the envelope. Computation of equivalent bandwidths for different vowels under conditions of varying maximum glottal damping are needed. Also, see Ananthapadmanabha et al (1982).

Ripple components in the voice source can thus be accounted for by making appropriate temporal changes in F1 frequency and bandwidth. But, the ripple components may add in different phases with the residue component thus causing either an increase or a decrease in the slope at closure. This would be exemplified when F_0 and F_1 have simple ratios (Fant et al. 1963). Such changes would affect the strength of excitation of higher formants. This will be referred to as a nonlinear 'superposition'. (See forthcoming article by G. Fant and T.V. Ananthapadmanabha.)

The effect of coupling to the subglottal system on the source residue can be considered by increasing the input inductance. Since the subglottal system can be represented as an inductance in series with the glottal impedance, the effect of the subglottal system on damping is to reduce the effect of truncation for high F_1 vowels, and its effect is negligible for low F_1 vowels. Further studies on perceptual importance of fine temporal structure in damping, the truncation effect and the superposition effect are in progress.

Despite the truncation and the superposition effects, we can say that the use of the source residue component as voice source provides an accurate pragmatic description of the true glottal flow. Since the source residue component is a smooth function, it can be described parametrically.

5. Conclusion

Based on a system theoretic approach, we have given numerical and analytical methods for computing the glottal flow which retains the major aspects of the source-filter interaction. The computed volume velocity of the air flow compares well with the practical data of oscillographic waveforms obtained from speech analysis and inverse filtering studies (Fant 1981). So far, not much attention has been given to the study of ripple components in the glottal flow. With the approach presented here, it has been possible to separate out the components in true glottal

flow and thus study the temporal and spectral properties of the ripple components. Throughout this work, a typical glottal area function was assumed for the flow calculations. The next step in the study of the voice source dynamics is to obtain the covariation of the glottal area function with respect to the subglottic pressure, pitch, abduction etc. This will require a well coordinated experimental investigation by means of photoglottography, inverse filtering, and subglottal pressure measurements.

References

ANANTHAPADMANABHA, T.V., NORD, L. and FANT, G. (1982): "Perceptual discriminability of nonexponential/exponential damping of the first formant of vowel sounds", to be publ. in Proc. of the Symp. on The Representation of Speech in the Peripheral Auditory System, May 17-19, 1982; Elsevier/North-Holland.

van den BERG, Jw. (1957): "On the air response and the Bernouilli effect of the human larynx", J.Acoust.Soc.Am. 29.

CHILDERS, D.G., MOTT, J.S. and MOORE, G.P. (1980): "Automatic parameterization of vocal cord motion from ultra high speed films", Proc. ICASSP, Vol. 1, pp. 65-68.

FANT, G. (1960): Acoustic Theory of Speech Production, Mouton, Hague. 2nd edition 1970.

FANT, G., FINTOFT, K., LILJENCRANTS, J., LINDBLOM, B. and MÁRTONY, J.(1963): "Formant-amplitude measurements", J.Acoust.Soc.Am. 35.

FANT, G. (1979a): "Glottal source and excitation analysis", STL-QPSR 1/1979.

FANT, G. (1979b): "Vocal source analysis - a progress report", STL-QPSR 3-4/1979.

FANT, G. (1980): "Voice source dynamics", STL-QPSR 2-3/1980.

FANT, G. (1981): "The source filter concept in voice production", STL-QPSR 1/1981.

FANT, G. and ANANTHAPADMANABHA, T.V. (forthcoming): "Truncation and superposition", to be publ. in STL-QPSR.

FLANAGAN, J.L. and LANDGRAF, L. (1968): "Self-oscillating source for vocal tract synthesizers", IEEE Trans. Audio and Electroacoust., AU-16, pp. 57-64.

FLANAGAN, J.L., ISHIZAKA, K. and SHIPLEY, K.L. (1975): "Synthesis of speech from a dynamic model of the vocal cords and vocal tract", Bell System Techn. J., 54.

GAUFFIN, J., BINH, N., ANANTHAPADMANABHA, T.V. and FANT, G. (1981): "Glottal geometry and volume velocity waveform", to be publ. in Proc. of the Research Conf. on Voice Physiology, Madison, WI, USA.

GUERIN, B., MRYATI, M. and CARRE, R. (1976): "A voice source taking account of coupling with the supraglottal cavities", Conf.Rec. IEEE, ASSP, pp. 47-50.

ISHIZAKA, K. and MATSUDIARA, M. (1972): "Fluid mechanical considerations of vocal cord vibration", SCRL Monograph No. 8.

ISHIZAKA, K., MATSUDIARA, M. and KANEKO, T. (1976): "Input acoustic impedance measurement of the subglottal system", J.Acoust.Soc.A. 60, pp. 910-917.

KITZING, P. (1977): "Methode zur kombinierten photo- und elektroglottographischen Registrierung von Stimmlippenschwingungen", Fol.Phon. 29, pp. 249-260.

MATAUSEK, M.R. and BATALOV, V.S. (1980): "A new approach to the determination of the glottal waveform", Trans.IEEE ASSP 28, pp. 616-622.

ROTHENBERG, M. (1980): "Acoustic interaction between the glottal source and the vocal tract", Proc.Conf. Vocal Fold Physiology, Kurume, Japan.

SCHERER, R.W. (1981): "Laryngeal fluid mechanics: Steady flow considerations using static models", Ph.D. Thesis, University of Iowa.

WAKITA, H. and FANT, G. (1978): "Toward a better vocal tract model", STL-QPSR 1/1978.

APPENDIX-A

Let $U_{sc}(t)$ be a finite duration signal with continuously differentiable derivatives in the interval $(0, T_0)$. Then we can write a series expansion for $U_{sc}(s)$ as

$$U_{sc}(s) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0) - f^{(n)}(T_0)e^{-sT_0}}{s^n} \tag{A.1}$$

where $f^{(n)}(t)$ is the n-th derivative of $U_{sc}(t)$. Consider a typical term of $U_{sc}(s)$ in Eq. (A.1) as

$$U_{sc}(s,k) = k!U_R/s^{k+1} \tag{A.2}$$

which has a time domain correspondence $U_{sc}(t,k)=U_k t^k$. From Eq. (18), the k-th term of $U_g(s)$ becomes

$$U_g(s,k) = \frac{k!U_R}{s^{k+1}} T(s) \tag{A.3}$$

where

$$T(s) = \frac{s^2 + 2\alpha_0 s + \omega_0^2}{s^2 + 2\alpha_1 s + \omega_1^2} \tag{A.4}$$

The source residue component of $U_g(s,k)$ is given by

$$U_r(s,k) = k!U_k\left[\frac{A_1}{s^{k+1}} + \frac{A_2}{s^k} + \frac{A_3}{s^{k-1}} + \ldots\right] \tag{A.5}$$

where

$$A_1 = \alpha t \underset{s\to 0}{} T(s), \quad A_2 = \alpha t \underset{s\to 0}{} \frac{dT(s)}{ds}$$

$$A_j = \alpha t \underset{s\to 0}{} \frac{1}{(j-1)!} \frac{d^{j-1}T(s)}{ds^{j-1}} . \tag{A.6}$$

It may be noted that constants A_1, A_2, ... are independent of $U_{sc}(t)$. The Laplace inverse of (A.5) gives the time domain k-th term as

$$U_r(t,k) = A_1U_k t^k + A_2U_k kt^{k-1} + A_3U_k(k)(k-1)t^{k-2} + \ldots$$

$$= A_1U_{sc}(t,k) + A_2U'_{sc}(t,k) + A_3U''_{sc}(t,k) + A_3U''_{sc}(t,k) + \ldots \tag{A.7}$$

Replacing the k-th term by the series expansion (i.e., summing both sides with respect to k), and since $A_1, A_2, A_3, \ldots$ are independent of $U_{sc}(t)$, we get

$$U_r(t) = A_1 U_{sc}(t) + A_2 U'_{sc}(t) + A_3 U''_{sc}(t) + \ldots \qquad (A.8)$$

Solving for $A_1, A_2, \ldots$ from (A.6), we get

$$U_r(t) = \frac{1}{\left[1+0.5Lg'(\tau)\right]} \left[U_{sc}(t) - \frac{0.5L\ g(\tau) - Lg''(\tau)/R}{\left[1 + 0.5Lg'(\tau)\right]} U'_{sc}(t) \ldots \right]$$

Replacing ' ' by 't' in the final step we get the desired result:

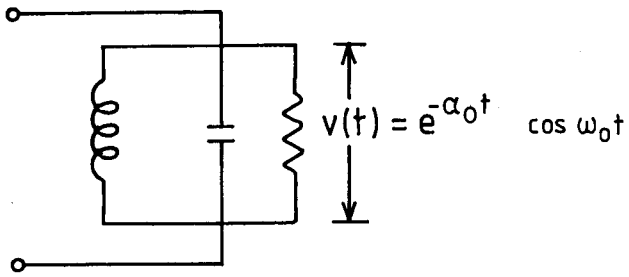
$$U_r(t) = \frac{1}{\left[1 + 0.5Lg'(t)\right]} \left[U_{sc}(t) - \frac{0.5L\ g(t) - Lg''(t)/R}{\left[1 + 0.5Lg'(t)\right]} U'_{sc}(t) \ldots \right] \qquad (A.9)$$

APPENDIX-B

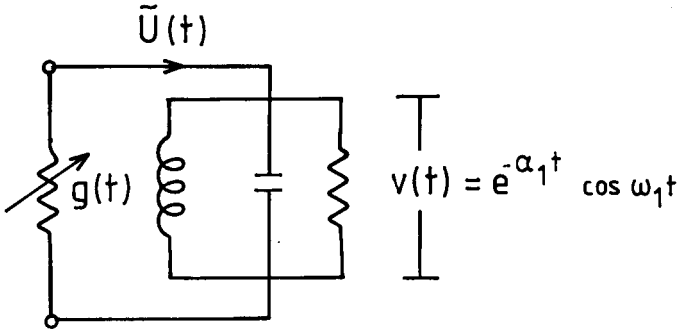
The main excitation at closure causes an exponential damping over the closed phase with a response of the form $e^{-\alpha_0 t}\cos(\omega_0 t)$, where ω_0 and α_0 correspond to the closed phase resonant frequency and damping constant, see Fig. I-A-B.1. This transient response does not require any input current. Over the open phase, due to source coupling, the transient response changes to the form $e^{-\alpha_1 t}\cos(\omega_1 t)$, Fig. I-A-B.2, where only the glottal resistance is shown. When true glottal flow is used as the excitation, there is no coupling to the source resistance, Fig. I-A-B.3. Then a component of the current $\tilde{U}(t)$ is required to simulate the open phase damping changes. Comparing Figs. I-A-B.2 and I-A-B.3, it can be inferred that $\tilde{U}(t)$ corresponds to the (negative) of the current through $g(t)$:

$$\tilde{U}(t) = -g(t)e^{-\alpha_1 t}\cos(\omega_1 t).$$

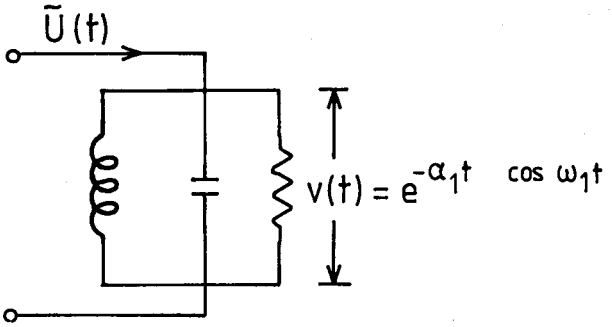
It may be noted that $\tilde{U}(t)$ has the usual F_1 ripple but modulated by $g(t)$. Thus, the envelope of $\tilde{U}(t)$ has the form of $g(t)$, Fig. I-A-B.4.



CLOSED PHASE
TRANSFER FUNCTION
Fig. I-A-B.1



OPEN PHASE
TRANSFER FUNCTION
Fig. I-A-B.2



SIMULATED OPEN PHASE
Fig. I-A-B.3

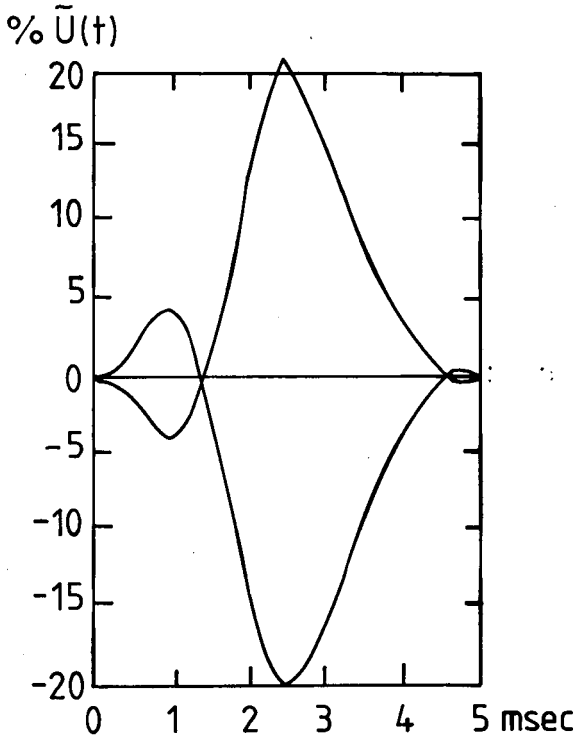


Fig. I-A-B.4