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

Glottal source - vocal tract acoustic interaction

Fant, G. and Lin, Q.

KTH Computer Science and Communication

journal: STL-QPSR
volume: 28
number: 1
year: 1987
pages: 013-027

B. GLOTTAL SOURCE - VOCAL TRACT ACOUSTIC INTERACTION*

Gunnar Fant and Qiguang Lin

Abstract

Recent developments within our group of voice source - vocal tract acoustic interaction are reviewed. Special emphasis is layed on non-linear superposition phenomena, i.e., how the excitation within a period is dependent on the past history of vocal tract oscillations and their residual components within the transglottal pressure. A study of breathy phonation shows that constant leakage affects the voice source slope less than does the dynamic leakage in terms of a residual closing phase. A simulation of a female voice source is attempted.

Introduction

Acoustic interaction we define as the dependency of the glottal volume velocity flow $U_g(t)$ on supraglottal articulations under the constraints of a prescribed glottal area function $A_g(t)$. Mechanical interaction, on the other hand, is the dependency of glottal vibratory patterns and thus of $A_g(t)$ on the overall state and aerodynamics. The complete interaction thus has a mechanical and an acoustic component.

The source filter model we employ for simulations of interaction as well as for inverse filtering states that a voiced speech sound is the convolution of the true glottal flow and the supraglottal transfer function. It is thus the combination of a complex and dependent source and a linear, theoretically well defined transfer function. A practical complication in inverse filtering is the difficulty of attaining an accurate formant tracking and to estimate "closed glottal conditions" even if they do not exist as in breathy voicing.

The objectives of our simulations presented here are to contribute to an insight in the factors that determine acoustic interaction. The glottal flow $U_g(t)$ and its derivative $U'_g(t)$ and the spectrum of flow derivative $U'_g(f)$ have been calculated with a variation of vocal tract network parameters and glottal area functions. A main finding is the great dependency of interaction on the past history of the pressure-velocity state within the vocal tract. Basic theory and more detail data have been presented in earlier publications. This is to be considered an informal progress report of results from simulations supplementing earlier work; see the list of references.

---

*Paper EE24, 113th Meeting of the Acoustical Society of America, May, 1987.

We have not yet exploited the full potentialities of the theoretical modeling. Much remains to be done in systematic studies of the influence of the many components involved and of different voice types including female and childen's voices. Also we need more experience from confronting the model with data from inverse filtering of real speech and evaluating the perceptual importance of various aspects of interaction.

Theory

The basic theory, see Ananthapadmanabha & Fant (1982) and more recently, Fant, Liljencrants, & Lin (1985a); Fant, Lin, & Gobl (1985b); Lin (1987), focuses on the instantaneous transglottal pressure.

$$\Delta P = P_\ell - (P_{in} + P_{sg}) = \frac{k\rho\, v_0^2 |x| x}{2} + \frac{12\mu d \ell^2}{A_g^2} v_0 x + \rho d v_0 x' \qquad \text{Eq. (1)}$$

The three terms at the right represent pressure drops associated with glottal kinetics, viscosity and inductance. We have introduced the normalized particle velocity x.

$$\left.\begin{array}{l} U_g(t) = A_g(t) \cdot v_0 \cdot x(t) \\ x = v_g(t)/v_0 \quad v_0 = (2P\ell/k\rho)^{\frac{1}{2}} \quad \rho d = A_g(t) \cdot L_g(t) \end{array}\right\} \qquad \text{Eq. (2)}$$

P_ℓ is lung pressure, d denotes glottal depth and ℓ glottal length. The glottal shape factor k is set to 1.

The convolution $P_i(t)=U_g(t)*Z_iU_g(t)*Z_id(t)/(t)$ is handled by treating $Z_i(t)$ as the inverse Lapalce transform of the partial fraction expansion of Z(s) up to five formants and a realization with standard second order digital filters. Similarly, three formants are included for the subglottal impedance. Eq. (1) is discretized and solved as a quadratic equation without any iterations which is an improvement compared to the iterative procedure employed by Ananthapdamanabha & Fant (1982), see also Fant et al. (1985a; 1985b). In order to allow negative going flow in the glottal impedance, which is not unrealistic, we have substituted $|x|x$ for x^2.

The introduction of the relative particle velocity x allows a greater computational accuracy, simplifies the expression for the pressure drop across the glottal inductance and is a direct compound measure of interaction. If we temporarily neglect sub- and supraglottal loads, and the glottal inductance LPV_g and the viscosity term R_g, we end up with a constant x=1 and a constant reference particle velocity

$$v_0 = 3709(P\ell/8)^{0.5} \qquad \text{Eq. (3)}$$

where $P\ell$ is the lung pressure in cm H_2O.

The flow pulse

$$U_g(t) = A_g(t) \cdot v_0 \cdot x(t) = A_g(t) \cdot v_0 \qquad \text{Eq. (4)}$$

is now fully proportional to the glottal area function. This is the so-called "short circuit" flow $U_{sc}(t)$ when the complete load is introduced. The function x(t) varies significantly within the glottal open period and accounts for:

(1) Pulse-skewing, generally to the right due to total vocal tract and glottal inductance. The flow is delayed with respect to the glottal area. This effect has been closely studied by Rothenberg (1981).

(2) Overlayed ripple caused by short-time transglottal pressure variations from formant oscillations evoked at the pulse onset and in previous glottal periods. As a rule, the contribution from excitations in a previous pulse overrides components originating from the particular flow pulse under observation. The quadratic kinetic term accounts for a non-linear conversion from instantaneous transglottal pressure to flow. We thus have a nonlinear superposition, the past history of vocal tract oscillations adding to the details of a glottal source pulse.

(3) Short-term variations of formant bandwidth and frequency within the time span of a glottal pulse. These effects may be hard to separate from multiple excitations within a flow pulse adding to discontinuities of instantaneous frequencies and time-domain envelopes. A heavy damping is referred to as truncation. Ripple and truncation effects have been extensively studied by Ananthapadmanabha & Fant (1982); Fant & Ananthapadmanabha (1982); Fant & al. (1985a; 1985b).

In the present study we find significant ripple and truncation effects even in higher formants, F_2, F_3 and F_4. All these interaction effects add to a random fine structure of shape amplitude and excitation time locations within a glottal pulse train. In the frequency domain, they are reflected by a few extra humps and valleys in the source flow spectrum. These features, as well as the mechanical component of interaction, seem to contribute to the naturalness of speech.

One finding of relevance to the characteristics of children's and female voices is that source spectrum slope is less influenced by the presence of a leakage than by concommittant changes in glottal area function shape, specifically the dynamic leakage, i.e., the residual closing phase.

References

Ananthapadmanabha, T.V. & Fant, G. (1982): "Calculation of true glottal flow and its components", Speech Communication 1, pp. 167-184.

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

Fant, G. (1986): "Glottal flow: Models and interaction", pp. 393-399 in (B. Fritzell & G. Fant, eds.), Voice Acoustics and Dysphonia, Theme Issue, J. of Phonetics 14, No. 3/4.

Fant, G. & Ananthapadmanabha, T.V. (1982): "Truncation and superposition", STL-QPSR 2-3/1982, pp. 1-17.

Fant, G., Liljencrants, J., & Lin, Q. (1985a): "A four-parameter model of glottal flow", STL-QPSR 4/1985, pp. 1-13.

Fant, G., Lin, Q., & Gobl, C. (1985b): "Notes on glottal flow interaction", STL-QPSR 2-3/1985, pp. 21-45.

Lin, Q. (1987): "Nonlinear interaction in human voice production", paper to be presented at the Int. Conf. of Chinese Information Processing, Beijing, P.R. of China.

Rothenberg, M. (1981): "An interactive model for the voice source", STL-QPSR 4/1985, pp. 1-17.

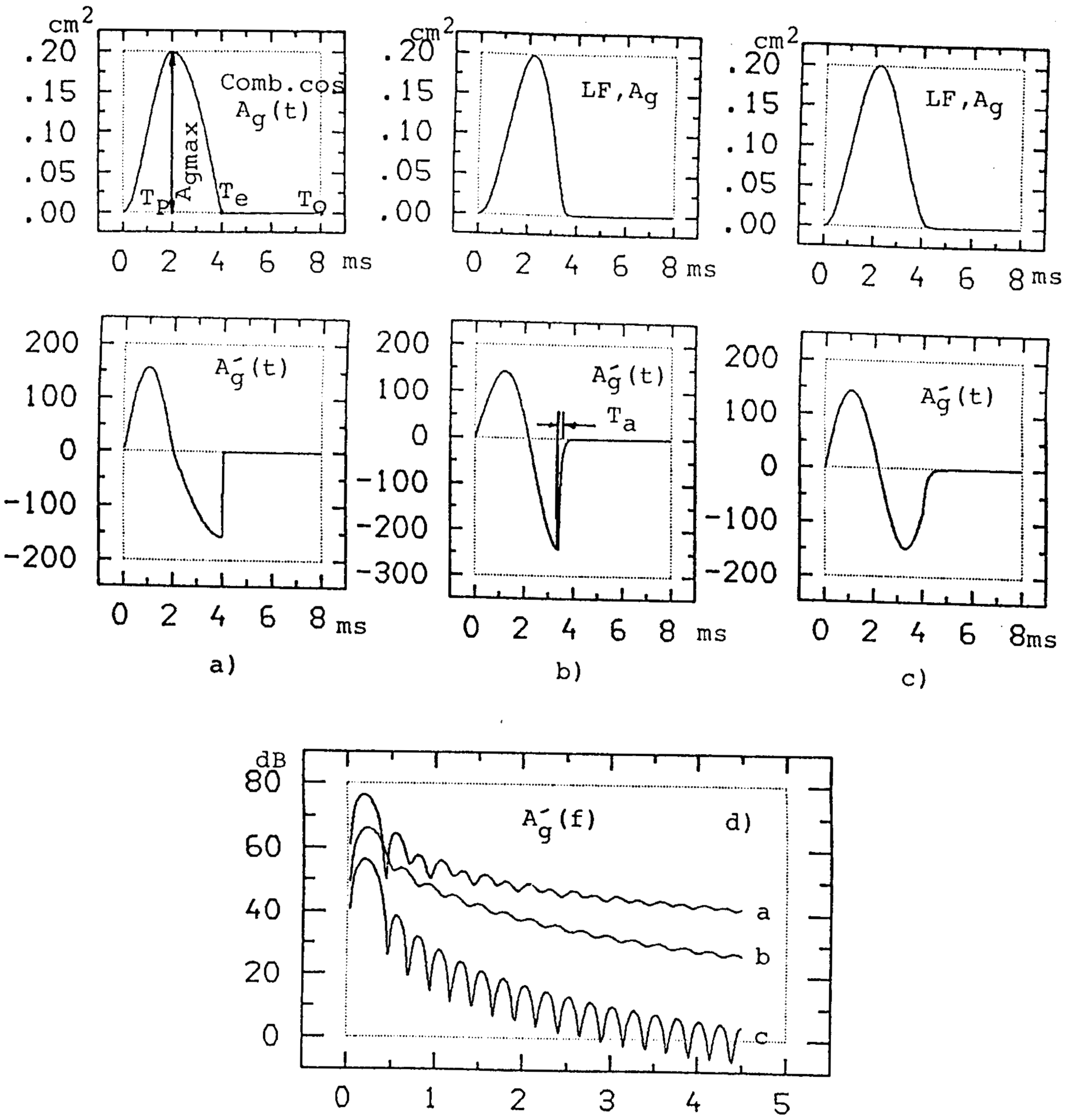


Fig. 1. Examples of the glottal area function models employed in this study.

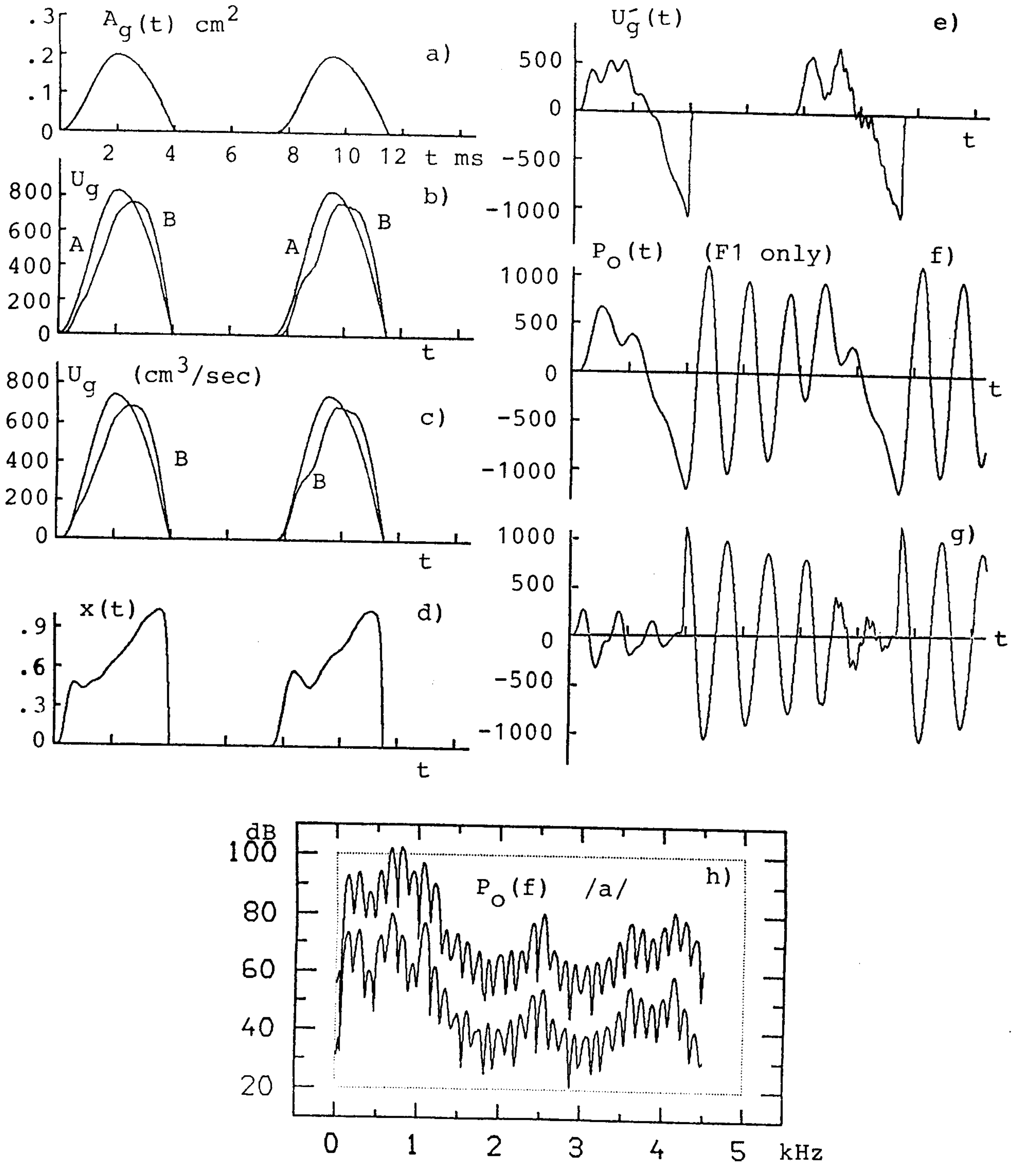
a) A raised cosine waveform and its derivative form;

b) An LF model and its derivative form whose main characteristic is the presence of the gradual return phase after the major closure which we used for simulating breathy voice and female and children's voices.

c) Another example of the LF model is shown.

d) The corresponding spectra, $A_g'(f)$.

The open quotient is defined as the ratio of T_e/T_o, and the symmetry factor as $T_p/(T_e-T_p)$.



A_g(t) cm^2
a)
.3
.2
.1
0
2 4 6 8 10 12 t ms
U_g
b)
A
B
800
600
400
200
0
t
U_g (cm^3/sec)
c)
B
t
x(t)
d)
.9
.6
.3
0
t
U'_g(t)
e)
500
0
-500
-1000
t
P_o(t) (F1 only)
f)
1000
500
0
-500
-1000
t
g)
1000
500
0
-500
-1000
t
dB
100
80
60
40
20
P_o(f) /a/
h)
0 1 2 3 4 5 kHz

Fig. 2.

Fig. 2. a) Tracing of glottal area function (a combined cosine wave). Here T_p=2 ms, T_e=4 ms, T_o=8 ms, A_{gmax}=0.2cm^2.

b) Glottal source flow. A comparison is made between a short circuit (glottal inductance and glottal viscosity are set to zero) and a compound vowel, /a/, loading. Curve A pertains to the short circuit case, while Curve B corresponds to the interactive model. It can be seen that Curves A and B differ in SKEWNESS: Curve A is skewed to the right side and RIPPLE: overlayed ripple components are seen in Curve A. These effects are the main consequences of interaction.

c) The same as b) except for the presence of a finite glottal inductance and glottal viscosity when computing the glottal flow for the short circuit case. Curve B is identical to Curve B in b). It is shown by comparing b) and c) that under the short circuit circumstance, the finite glottal inductance and viscosity will make the glottal flow onset more gradual, i.e., a corner-rounding effect.

d) $x=V_g/V_o$, the normalized particle velocity, see text for details; x is an indicator of interaction. In a conventional linear source model, it is constant, x=1 when the glottis is open and x=0 when the glottis is closed. The product $x(t)A_g(t)V_o$ is the computed true glottal flow.

e) Differentiated glottal flow, $U_g'(t)$. Here the ripple is more apparent than in $U_g(t)$. The negative peak v amplitude becomes is a scale factor of formant excitation.

f) The lip pressure output (defined as the derivative of the volume flow at the lips) after selected inverse-filtering (without cancellation of the F1 oscillation).

g) The same as f) except that the source component is eliminated here by a double-differentiation. The F1 oscillation decays faster as the glottis is open due to the coupling to the glottal internal impedance.

h) FFT spectrum of the sound pressure output, vowel /a/. The Hamming window covers a time span of three fundamental periods. A close comparison is made between the interactive (top curve) and the linear models (bottom). It is seen that the peaks of F1, F2, F4 and F5 are broadened in the top curve and that the valleys between F0 and F1 and between F1 and F2 have been "filled in" in the interaction case. See Fig. 3 for further discussions on spectral details. The interaction has also increased the formant amplitudes, viz., the fundamental and the second harmonic amplitude which derive from the skewing effect.

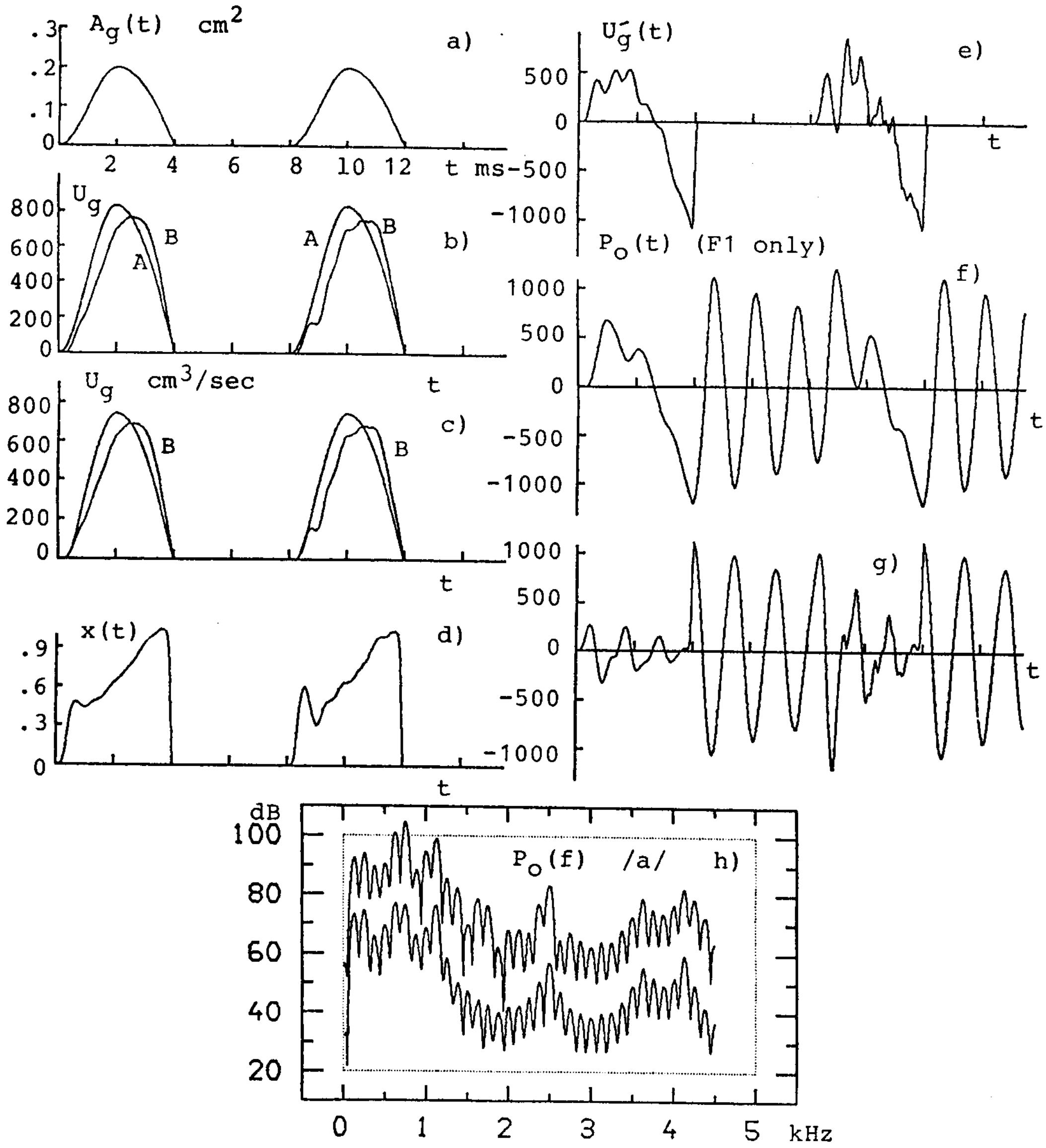


Fig. 3. The same simulation condition as in Fig. 2 except for T_o=8 ms instead of 7.5 ms as in Fig. 2.
It is observed that a pseudo-peak between F2 and F3 appears in the sound pressure spectrum of the interaction model (the top curve in the bottom panel). Such a peak is not visible when T_o=7.5 ms. This means that the nonlinear superposition effect is very critical. This is especially true when some formants of a sound have a small bandwidth.

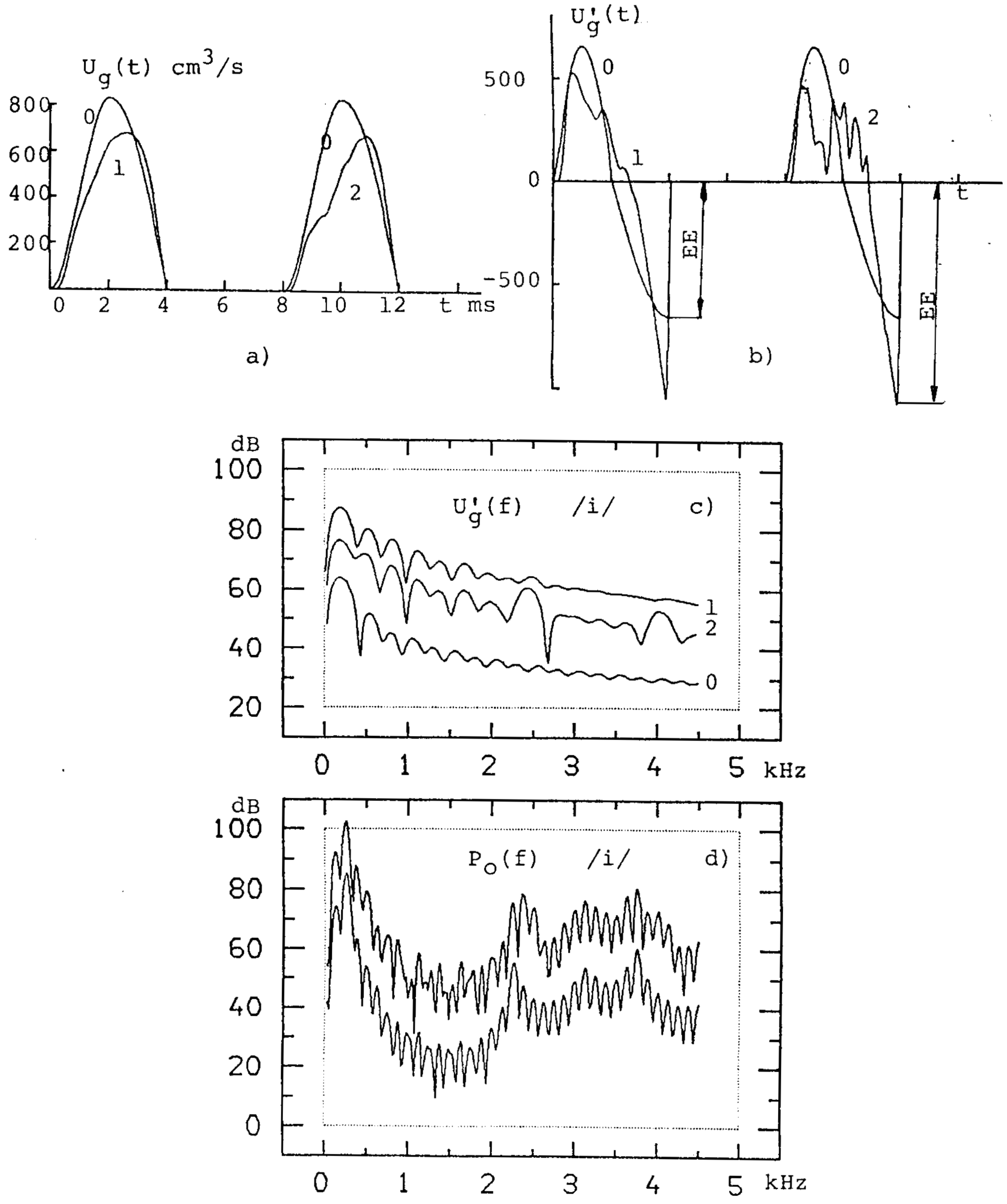
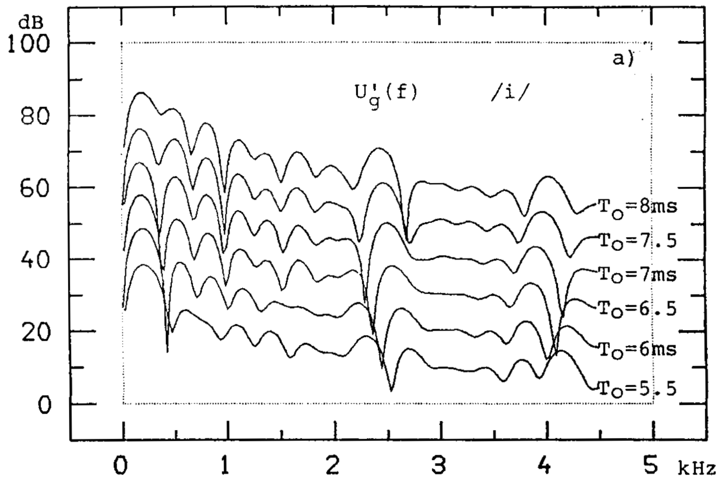


Fig. 4. a) The vowel /i/. A comparison between the glottal flow in the linear model, (curves labelled 0) and the interactive model, (curves labelled 1 and 2, the initial and second pulse, respectively).

b) Corresponding flow derivatives. The excitation level defined as the maximum negative peak at closure, EE, is about 4 dB higher in the interactive model than its counterpart in the linear source model.

c) The corresponding FFT spectrum of the curves in b). A couple of extra humps are clearly seen in the source spectrum due to the interaction effect.

d) FFT spectrum of the sound pressure within a frame of three complete periods. The interaction (top curve) causes a broadening of F1 and F2 and a small positive shift in the F2 frequency.



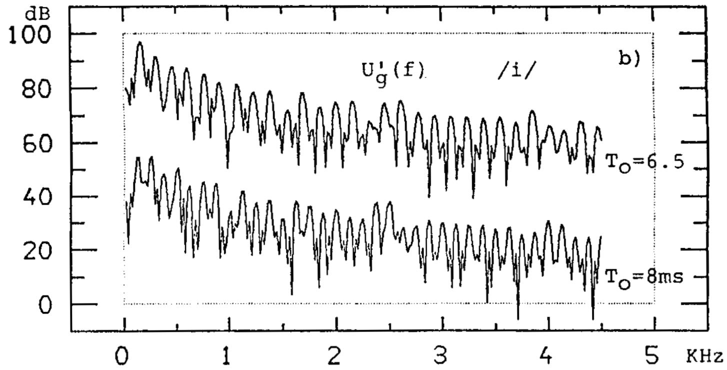


Fig. 5. a) The vowel /i/. Nonlinear superposition as a function of a varying duration of the closed phase. It should be noted that through T_O=7 ms, the notch changes its location relative to the extra peak in the vicinity of 2.2 kHz.

b) $U_g'(f)$, the Hamming window covering three voice cycles displays the harmonic structure. Top curve: T_O=6.5 ms; Bottom curve: T_O=8 ms.

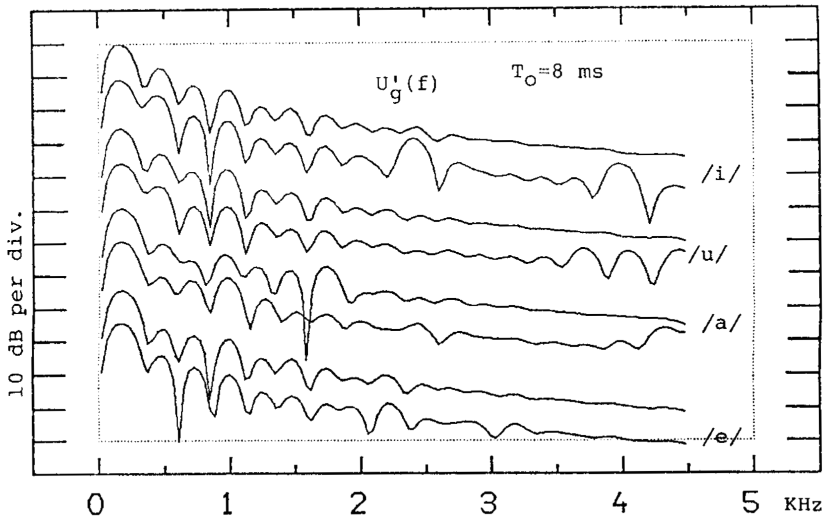


Fig. 6. The first and second pulse source spectra of different vowels: /i/, /u/, /a/, and /e/. The curve shift between the vowels is 15 dB. The first and second pulses are separated 10 dB. A small inherent difference in excitation level exists. It increases within the series /e/, /u/, /i/ and finally to /a/ with a total span of 1.5 dB. It is observed that the source spectra are inter-periodically different for each vowel.

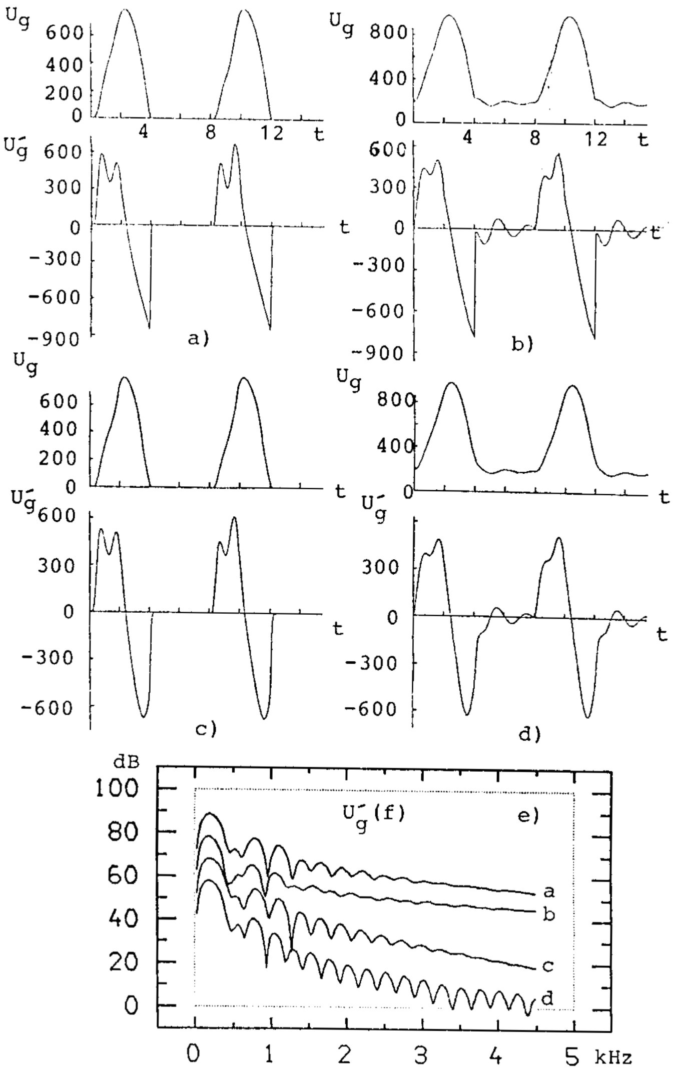


Fig. 7A. Leakage simulation. The acoustic load is modelled by as single formant. F/B=500/50, L=5 mH.

a) No leakage. The glottal area function is a raised cosine; T_o=8 ms, T_e=4 ms, T_p=4 ms, A_{gmax}=0.2 m^2.

b) A constant leakage (the leaky area is 25% of the glottal peak area) by adding the leaky area to the glottal function given in 1). The load is the same as in 1). Note the reduction of ripple components during the glottal open phase adn the appearence of formant oscillation in the "closed" phase.

c) A dynamic leakage, the glottal area function is simulated by an LF model. T_e=4 ms, T_p=2.2 ms, T_a=0.15 ms and EE=80. The maximum glottal area is approximately the same as in 1).

d) The same as 3) except that an addition of a constant leakage area is introduced to simulate both constant and dynamic leakage. Observe the remaining F1 ripples in the closed phase.

e) The corresponding source flow derivative spectra.

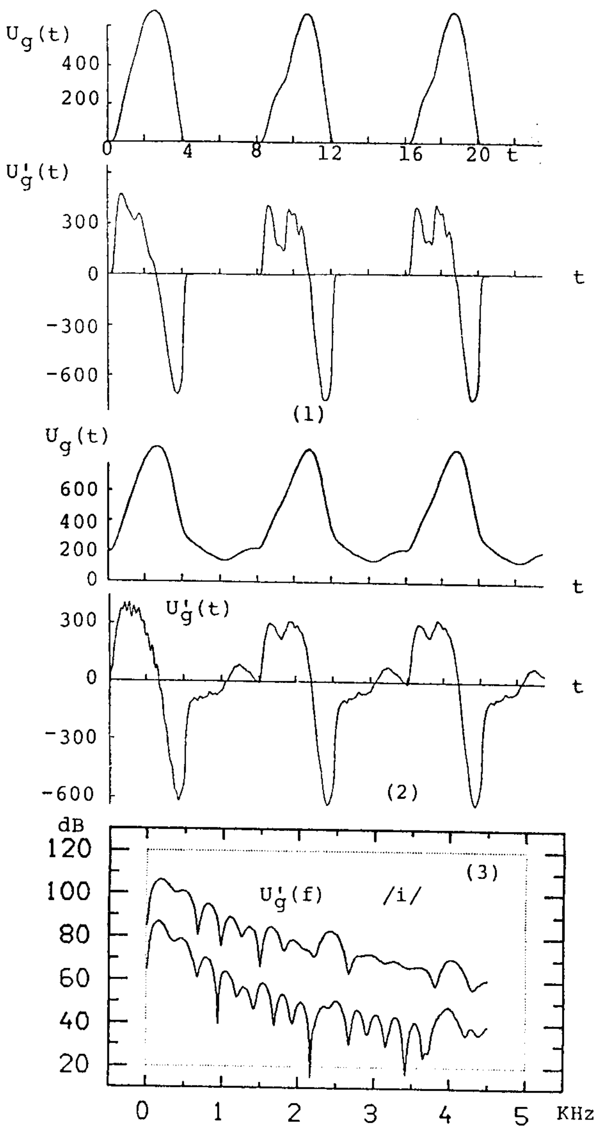


Fig. 7B. Leakage simulation. The acoustic load is a compound vowel /i/.

1) Dynamic leakage only, A_g the same as in Fig. 7A:3).
2) Both dynamic and constant leak. Glottal area conditions are the same as in Fig. 7A:4).
3) The corresponding source spectrum of 1) and 2). For the second pulse only.

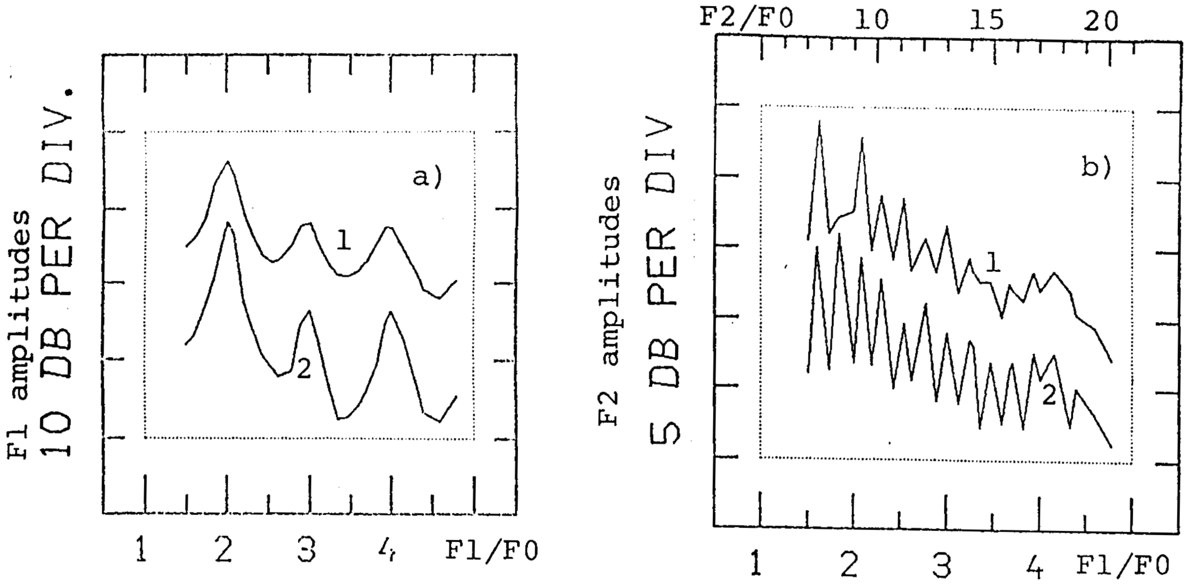


Fig. 8. a) Comparison of F1 amplitudes between 1: interactive and 2: linear models as a function of F1/F0.
b) Comparison of F2 amplitudes between 1: interactive and 2: linear models as a function of F1/F0 and F2/F0.

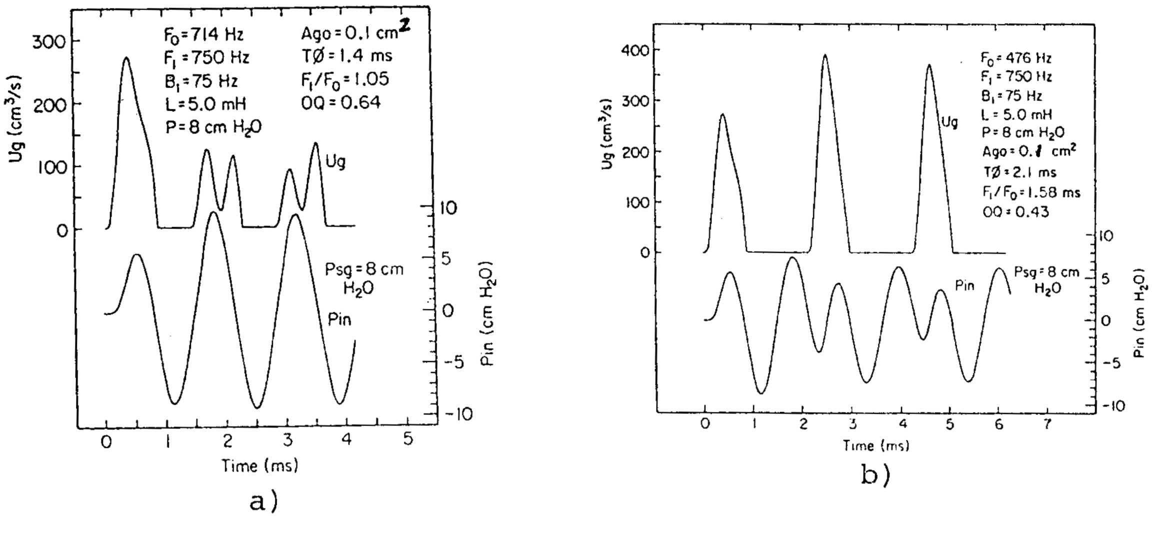


Fig. 9. Glottal flow and supraglottal pressure assuming a one-formant loading.
a) F1 close to F0, F1/F0=1.05 simulating a soprano. This is an optimum condition for larger output and low air consumption.
b) Glottal flow and supraglottal pressure at a lower F0=F1/1.58. Observe the larger air consumption (from Fant & al., 1985a).

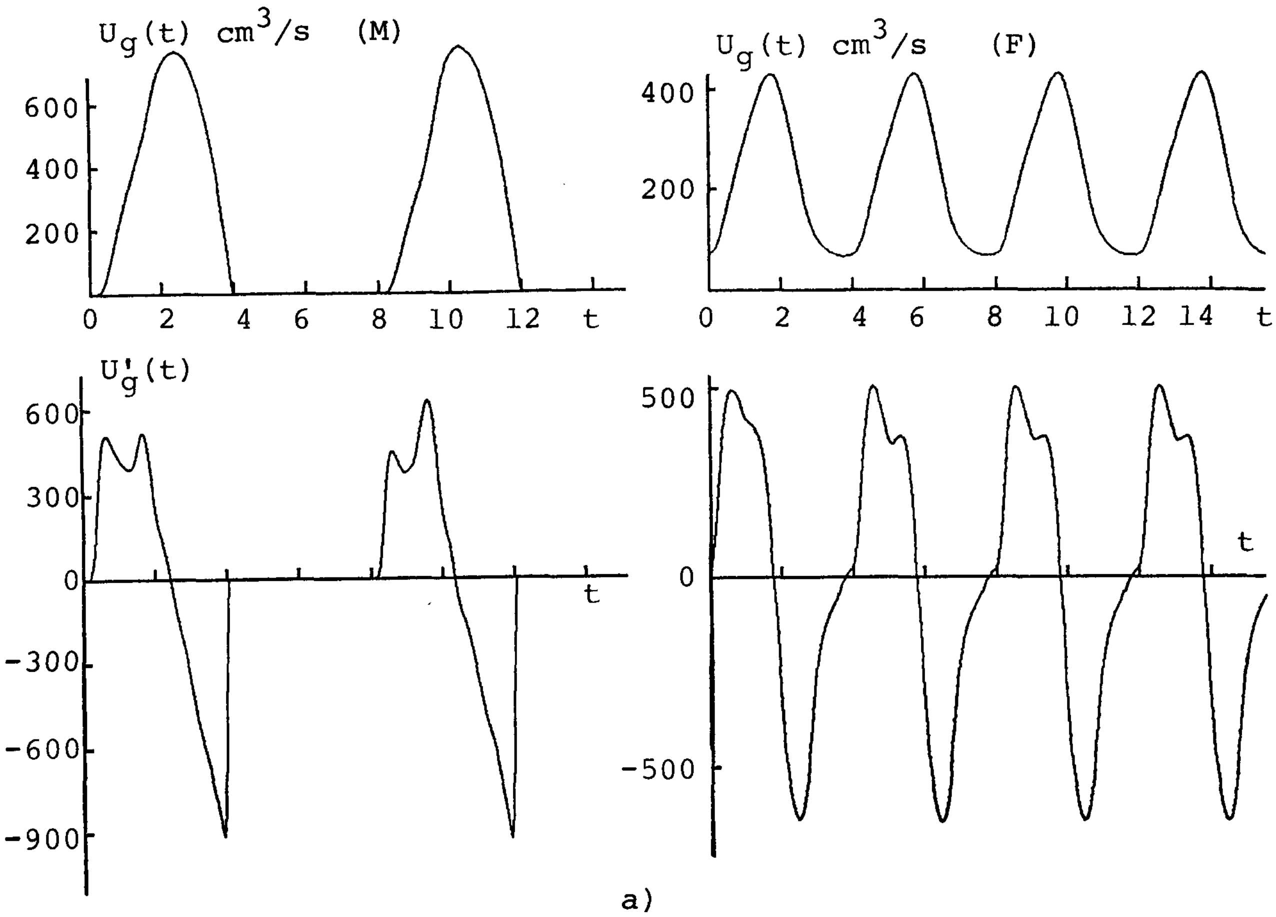


Fig. 10. An attempt to simulate male and female phonation of a neutral vowel. Male: A_g is modelled by a raised cosine function, T_o=8 ms, open quotient is 50%, T_p/T_e=50%. A_{gmax}=0.2 cm^2. Female: A_g is simulated by an LF-model, T_e=2.8 ms, T_p=1.5 ms, T_a=0.2 ms, T_o=4ms. A_{gmax}=0.1 cm^2. This could represent a somewhat breathy phonation.
Formant frequencies and bandwidths of the female sample are adjusted to representative values.
a) Glottal flow and flow derivatives.
b) Glottal flow derivative spectra, $U_g'(f)$, computed for the second pulse only.
c) The same as b) except that $U_g'(f)$ covers the first three periods which enhances the harmonic fine structure.
d) Sound pressure spectra. The length of the Hamming window is one and the same, 24 ms, for both sexes.

COMMENTS: The difference in source spectrum slope is primarily due to the "dynamic leakage", i.e., longer return phase of the female simulation whilst the added constant leakage has a smaller spectral effect, compare Fig. 7. Although female voices differ much, the prolonged return phase has been verified in human speech samples. The constant leakage mainly adds formant bandwidth increase, accounts for residual F1 ripple in the closed phase and adds somewhat to the spectral slope if combined with dynamic leakage.

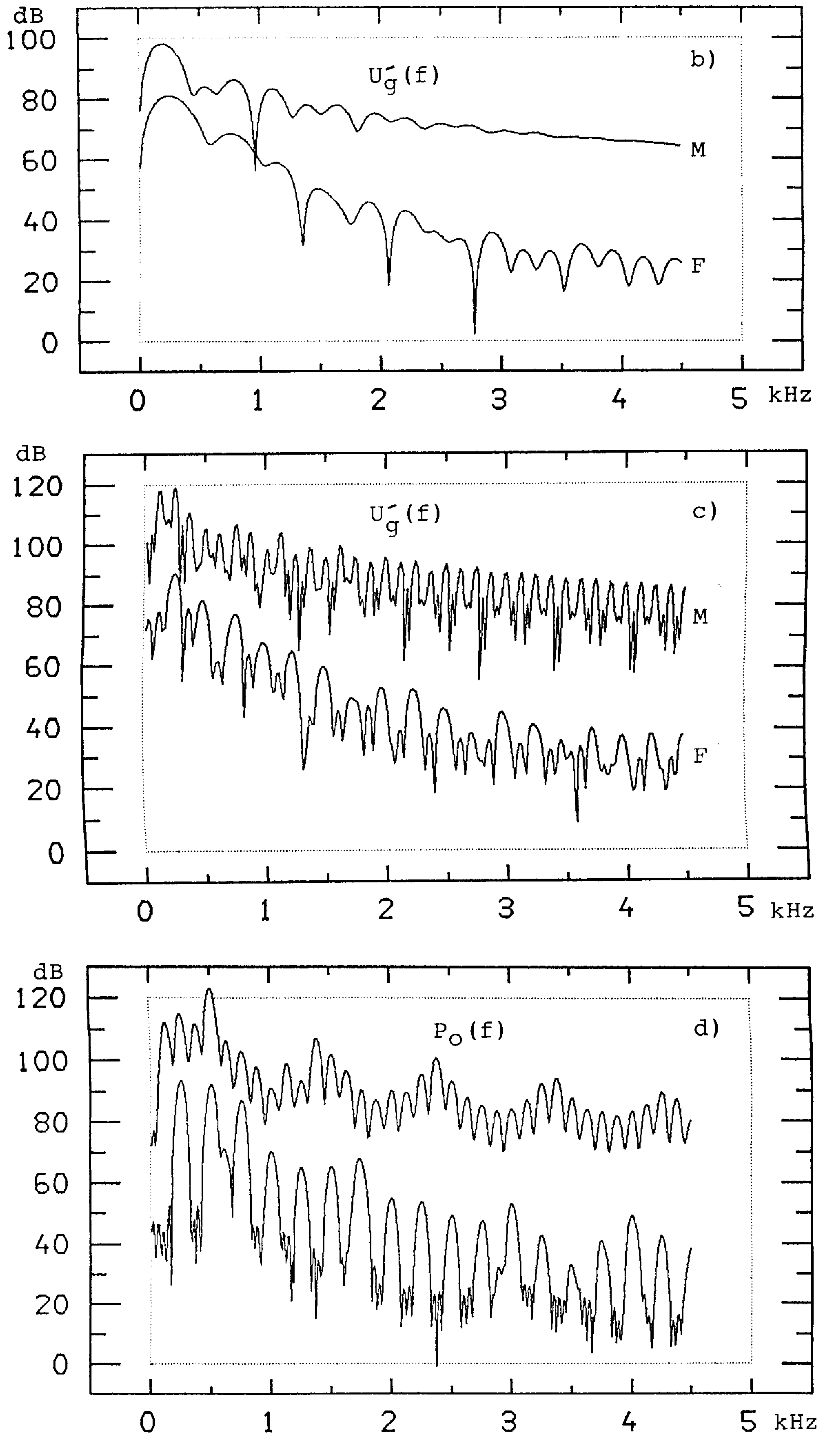

dB
100
80
60
40
20
0
$U_g'(f)$
b)
M
F
0
1
2
3
4
5
kHz
dB
120
100
80
60
40
20
0
$U_g'(f)$
c)
M
F
0
1
2
3
4
5
kHz
dB
120
100
80
60
40
20
0
$P_o(f)$
d)
0
1
2
3
4
5
kHz

Fig. 10 b) - d)