A four-parameter model of glottal flow

Fant, G. and Liljencrants, J. and Lin, Q.

journal: STL-QPSR
volume: 26
number: 4
year: 1985
pages: 001-013

http://www.speech.kth.se/qpsr
A. A FOUR-PARAMETER MODEL OF GLOTTAL FLOW

Gunnar Fant, Johan Liljencrants, and Qi-guang Lin

Abstract
A glottal flow model with four independent parameters is described. It is referred to as the LF-model. Three of these pertain to the frequency, amplitude, and the exponential growth constant of a sinusoid. The fourth parameter is the time constant of an exponential recovery, i.e., return phase, from the point of maximum closing discontinuity towards maximum closure. The four parameters are interrelated by a condition of net flow gain within a fundamental period which is usually set to zero. The finite return phase with a time constant $t_a$ is partly equivalent to a first order low-pass filtering with cutoff frequency $F_a = (2\pi t_a)^{-1}$.

The LF-model is optimal for non-interactive flow parameterization in the sense that it ensures an overall fit to commonly encountered wave shapes with a minimum number of parameters and is flexible in its ability to match extreme phonations. Apart from analytically complicated parameter interdependencies, it should lend itself to simple digital implementations.

The L-model
The four-parameter model, here referred to as the LF-model, has developed in two stages. The first stage was a three-parameter model of flow derivative, introduced by Liljencrants.

$$\frac{dU_g(t)}{dt} = E(t) = E_0 e^{at} \sin \omega t$$

It will be referred to as the L-model. It has the advantage of continuity whilst the early Fant (1979) model is composed of two parts, a rising branch:

\[ U_g(t) = \frac{1}{2} U_0 (1 - \cos \omega_g t) \]

with derivative

\[ E_1(t) = \frac{\omega_g U_0}{2} \sin \omega_g t \]

and a descending branch at \( t_p < t < t_p + \frac{\arccos \frac{K+1}{K}}{\omega_g} \)

\[ U_g(t) = U_0 [K \cos \omega_g (t-t_p) - K] \]

\[ E_2(t) = -\omega_g U_0 \sin \omega_g (t-t_p) \]

This F-model has a discontinuity at the flow peak which adds a secondary weak excitation, see Fant (1979). One potential advantage of the L-model, Eq. (1), is that it can be implemented with a standard second-order digital filter with positive exponent (negative damping). The generated time function is interrupted at a time \( t_e \) when the flow

\[ U(t) = \frac{[E_0 e^{\alpha t} (\alpha \sin \omega_g t - \omega_g \cos \omega_g t) + \omega_g]}{2^{\alpha + \omega_g}} \]

has reached zero.

The three algebraic parameters of the L-model, Eq. (1), \( E_0, \alpha, \) and \( \omega_g \) map on to the three basic flow derivative parameters

\[ t_p = \frac{1}{2 \omega_g}, \quad \omega_g = 2 \pi F_g \]

\( t_e \) from Eq. (4)

\[ E_e = -E_0 e^{\alpha t} \sin \omega_g t_e \]

The F- and L-models share the parameter \( t_p \) or \( F_g = 1/2 t_p \). As shown in Fig. 1, the L-model displays a more gradual rise than the F-model given the constraint of equal \( t_e \) and \( E_e/E_0 \). This asymmetry increases with increasing \( E_e/E_0 \). The spectral differences comparing the L- and F-models are not great. The L-model exhibits a lower degree of spectral ripple which is an advantage.

It is convenient to introduce the dimensionless parameters:

\[ R_d = \frac{2 \alpha}{\omega_g} = \frac{-R}{F_g} \]

\[ R_k = \frac{t_e-t_p}{t_p} = \frac{t_n}{t_p} \]
Flow derivative

Fig. 1b. Frequency-domain comparison of the F- and I-models.

Fig. 1a. Time-domain comparison of the F- and I-models.
Here, $B$ is the "negative bandwidth" of the L-model exponent. $R_d$ and $R_k$ and $E_e/E_i$ are mutually dependent shape parameters. The three-parameter L-model may thus be conceived of having one shape parameter, one temporal scale factor, and one amplitude scale factor.

The shape factor of the F-model (Fant, 1979) is

$$K = \frac{1}{8} \left( \frac{E_e}{E_i} \right)^2 + \frac{1}{2}$$

(7)

Referring to a basic shape parameter, $E_e/E_i$, we may thus compare the F- and L-models as follows:

<table>
<thead>
<tr>
<th>$E_e/E_i$</th>
<th>$K$</th>
<th>$R_d = \frac{2\alpha}{\omega_g}$</th>
<th>$R_k = \frac{t_e - t_p}{t_p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.625</td>
<td>0.12</td>
<td>0.73</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.43</td>
<td>0.54</td>
</tr>
<tr>
<td>3</td>
<td>1.625</td>
<td>0.84</td>
<td>0.42</td>
</tr>
<tr>
<td>4</td>
<td>2.5</td>
<td>1.35</td>
<td>0.33</td>
</tr>
</tbody>
</table>

The location $t_i$ of the flow derivative negative maximum $E_i$ is

$$t_i = \frac{t_p}{2} \left[ 1 + \arctan \left( \frac{R_d}{2} \right) \right]$$

(8)

which defines an asymmetry factor

$$R_1 = \frac{t_i - t_p/2}{t_p/2} = \frac{2}{\pi} \arctan \left( \frac{R_d}{2} \right)$$

(9)

At sufficiently small $K$-values in the F-model, $K < 1$, there is a negative smooth turning point prior to the closure step. In the L-model a corresponding negative smooth turning point occurs at

$$t = \frac{3t_p}{2} (1 + \frac{1}{3\pi} \arctan \left( \frac{R_d}{2} \right))$$

(10)

which requires $R_k > 0.54$ or $R_d < 0.43$. This property allows both the F- and the L-model in a limit to approach a sine wave, $K = 0.5$ and $R_d = 0$, which is of some relevance to the termination of voicing.
The obvious shortcoming of a model with abrupt flow termination is that it does not allow for an incomplete closure or for a residual phase of progressing closure after the major discontinuity. Either and generally both of these conditions account for a residual phase of decreasing flow after the discontinuity. In voiced h-sounds, i.e., in breathy phonations, the discontinuity may occur in the middle of the descending branch at the flow followed by a less steep descent and a final trailing off corner effect (see Fant, 1980, 1982). In less breathy phonations, the discontinuity is found further down towards the base line of the flow.

Accumulated experimental evidence the last years has shown that this corner effect is more a rule than an exception. In his five-parameter model, Ananthapadmanabha (1984) has accordingly included a terminal return phase which is modeled as a parabolic function. We have chosen to adopt an exponential function

$$E_2(t) = \frac{E_e}{e^{t/a}} \left[ e^{-\varepsilon(t-t_e)} - e^{-\varepsilon(t_c-t_e)} \right]$$  \hspace{1cm} (11)

$$t_e < t < t_c$$

As shown in Fig. 2, the time $t_a$ is the projection of the derivative of $E_2(t)$ at $t=t_e$ on the time axis. For a small $t_a$, $\varepsilon = 1/t_a$, and otherwise

$$\varepsilon t_a = \left[ 1 - e^{-\varepsilon(t_c-t_e)} \right]$$  \hspace{1cm} (12)

In practice it is convenient to set $t_c = T_0$, i.e., the complete fundamental period.

The sole remaining independent parameter of the return phase is, thus, its experimentally determined effective duration $t_a$. The area under the return branch, i.e., the residual flow at $t_e$ is written:

$$U_e = \frac{E \cdot t}{2} . K_a$$  \hspace{1cm} (13)

When $t_a$ is small, $K_a = 2$, otherwise an integration of Eq. (11) may be approximated by the following procedure. Calculate the parameter
The LF-model of differentiated glottal flow. The four wave-shape parameters \( t_p, t_e, t_a, \) and \( E_e \) uniquely determine the pulse \( (t_c = T_0 = 1/F_0) \).
\[ R_a = \frac{t_a}{t_c - t_e} \]  

(14)

If \( R_a < 0.5 \)

\[ K_a = 2 - 2.34 R_a^2 + 1.34 R_a^4 \]

If \( R_a > 0.5 \)

\[ K_a = 2.16 - 1.32 R_a + 0.64(R_a - 0.5)^2 \]  

(15)

If \( R_a < 0.1 \)

\[ K_a = 2.0 \]

We now know the residual flow \( U_e \). By imposing the final requirement of area balance, i.e., zero net gain of flow during a fundamental period

\[ \int_0^T E(t) = 0 \]

we may, accordingly, start out from the four waveform parameters, \( t_p, t_e, t_a \), and \( E_e \) and solve for the \( \alpha \) or \( R_\alpha = 2\alpha/\omega_e \) of Eqs. (4) and (5) to provide a flow matching that of Eq. (13).

Prior to this operation it is convenient to normalize flow functions by a division by \( E_e \) and \( t_p \). The scale factor \( E_0 \) for synthesis can now be determined from Eq. (5).

The amplitude \( E_1 \) of the positive maximum is found from

\[ E_1 = E_e \left( \frac{R_3}{2} \left( -\frac{\pi}{2} + \arctg \frac{R_3}{2} - \frac{\pi R_\alpha}{2} \right) \cdot \sin \left( \frac{\pi}{2} + \arctg \frac{R_3}{2} \right) \right) \]

(16)

The computer program for implementation is organized as follows.

Perform inverse filtering. Shift, if necessary the base line to achieve area balance, i.e., zero net change of flow. If the zero line is judged reliable, note the residual net flow for further use. Mark with a
cursor the starting time $t_0 = 0$ and the subsequent locations of $t_p$, $t_e$, $t_e + t_a = t_r$ and the complete period length $T_0$. Measure $E_e$.

Go through the procedure above, Eqs. (4)(5), and (13) with normalized flow functions $U_e/(E_e \cdot t_p)$. Apply an iterative solution starting from zero return flow to find $R_d$ and then $E_0$ and $E_1$. Synthesize the model curve and compare it with the inverse filtered curve. If $E_e/E_1$ does not fit, try to shift the starting point by adding a fixed quantity to $t_p$ and $t_e$. Update the parameters and check the result.

The results so far have shown this procedure to provide not less reliable results than the five-parameter model of Ananthapadmanabha (1984) and will be adopted as our standard. It also has the advantage of peak continuity and that it better matches highly aspirated waveforms and allows a sinusoid as an extreme limit.

Properties of the LF-model

An introduction of dynamic leakage, i.e., of a finite return phase, has to be compensated by a decrease of $\alpha$, i.e., of $R_d = 2\alpha/\omega_s$ and, thus, also of the peak ratio $A_e = E_e/E_1$ in order to maintain the area balance, i.e., the zero net gain of flow within a fundamental period. These trading relations are apparent from the nomograms of Figs. 3 and 4, which can be used for quantifying parameter interdependencies.

The effect on the flow derivative and flow pulse shape on the second derivative flow spectrum ($+12$ dB/oct versus flow spectrum) of a finite $t_a$ is illustrated in Fig. 5. Here $R_k = (t_e - t_c)/t_p = 0.5$. Increasing $t_a$ causes an increased high frequency deemphasis, as noted by Ananthapadmanabha (1984). Due to the exponential wave shape of the return phase, it may be modeled as an additional first-order low-pass function of cutoff frequency

$$F_a = \frac{1}{2\pi t_a}$$

(17)

As evidenced from Fig. 5, this is a good approximation for assessing the main spectral consequence of dynamic leakage. Even a very small departure from abrupt termination causes a significant spectrum roll-off in addition to the standard -12 dB/oct glottal flow spectrum. Thus, $t_a = 0.15$ ms accounts for a 3-dB fall at $F_a = 1060$ Hz and 12 dB at 4000 Hz. In general, the loss is

$$\Delta L = 10 \log(1 + f^2/F_a^2) \text{ dB}$$

A variation of $A_e = E_e/E_1$ at constant $E_e$ and $t_a = 0.1$ ms and $R_k = 0.5$ is illustrated in Fig. 6. Observe the significance of the negative spike $E_0$ to set the spectrum level at frequencies above $G_y$. The low-pass filter effect is approximately independent of $A_e$. The spectral ripple increases with decreasing $A_e$ and with increasing $t_a$. Increasing $A_e$ at constant excitation $E_9$ is related to a more "pressed" voice which
Fig. 3. LF-model nomograms relating the exponent to the relative excitation timing.

Fig. 4. LF-model nomograms relating the negative to positive peak ratio to the relative excitation timing.
Fig. 5. Waveshape and spectral correlates of increasing the return time constant $t_a$. 

\[ \Delta L = -10 \log_{10} \left[ 1 + (2\pi t_a f)^2 \right] \]
Fig. 6. Increasing the peak factor $A_0$ at constant excitation amplitude $X_0$ and a constant return time constant $t_0 = 0.1$ ms.

Fig. 7. LP-modeling of breathy voicing.
attains a "thinner" quality due to the relative lower low-frequency level.

Fig. 7 shows a sequence of two complete periods illustrating typical wave shapes and spectra of breathy phonation as in voiced [h], $t_a = 1.3$ ms, and an intermediate wave shape, $t_a = 0.2$ ms in addition to abrupt closure, $t_a = 0$. The constants have been chosen to provide similarity with the inverse filter data of Fig. 7 in Fant (1980).

Results from our present laboratory set-up for inverse filtering and glottal source analysis developed by Ananthapadmanabha (1984) are shown in Fig. 8. The object is a male [a] which has assimilated some breathiness from a following aspirated stop. Here there is a combination of an extreme large peak ratio $A_g = 4$ and a significant leakage, $T_a = 0.43$ ms, similar to that of Fig. 7. Under the oscillogram follows a five-parameter match with the A-model and then the four-parameter LF-model. In this case the LF-model provides a better overall fit. In less extreme cases the two models perform equally well. The initial LPC-spectrum analysis prior to inverse filtering does not fit the spectrum very well. This is in part due to the low sampling frequency in this special case.

The LF-model is mainly intended for a noninteractive modeling of the voice source but it could, like the Fant model, also be applied to the description of glottal area functions. A discussion of interactive phenomena is found in Ananthapadmanabha & Fant (1982), Fant & Ananthapadmanabha (1982), and Fant (1982). A perceptual evaluation of interaction has been reported by Nord, Ananthapadmanabha, & Fant (1984).

References


Fig. 8. Waveform matching with A-model and LF-model.