A LF-PULSE FROM A SIMPLE GLOTTAL FLOW MODEL

A. Aalto[1], P. Alku[2] and J. Malinen[1]
[1]Dept. of Mathematics and Systems Analysis, Helsinki Univ. Tech., Espoo, Finland
[2]Dept. of Signal Processing and Acoustics, Helsinki Univ. Tech., Espoo, Finland

***Abstract*: We discuss a novel low-order mass-spring model of human vocal folds with incompressible 1D flow. Our model consists of three subsystems: a flow model, a nonsymmetric mass-spring model for the vocal folds, and a resonator representing the vocal tract (VT).**
***Keywords*: Glottis model, Bernoulli flow, flow induced vibrations**

I. Introduction

This study addresses a mechanical glottis simulator to act as the source for the wave equation model of vowel production, see [4]. We require a validated glottal flow/pressure signal simulator that is computationally less demanding than Navier–Stokes and/or elasticity equations. We also use this model for studying the vocal tract feedback effect in [1] where we observe some phenomena reported in [10] and [11]. In this paper we validate our glottis model against the LF-signal model proposed in [3]. This is carried out by fitting the simulated pulse to LF-pulses measured from three different types of phonation.

Our model consists of three subsystems: a 1D flow model; a nonsymmetric, low-order mass-spring model for the vocal folds; and a resonator representing the vocal tract (VT), based on the Webster's equation. These subsystems are modelled using physically sound mathematical approximations. Firstly, we use the Bernoulli law and the mass conservation law for a static flow. Secondly, the flow through the glottis is assumed to be incompressible. Obviously, the flow is not truly static because of the moving vocal folds, and the Webster's equation is based on the very assumption that air is compressible. In addition to these inconsistencies between subsystems, also the glottis geometry is extremely simplified as, e.g., in the seminal work [7].

II. Mathematical model

A. Flow

We assume an incompressible 1D air flow through the glottal opening whose velocity v_o satisfies

$$\dot{v}_o(t) = \frac{1}{C_{iner} h H_1}\left(p_{sub} - \frac{C_g}{\Delta W_1(t)^3} v_o(t)\right), \quad (1)$$

motivated by the Hagen–Poiseuille law; here p_{sub} is the subglottal pressure (subtracted by the ambient air pressure), and h is the width of the rectangular channel. The parameter C_{iner} regulates the flow inertia and C_g regulates the pressure loss in the glottis. The glottal opening is given by $\Delta W_1 = g + w_{21} - w_{11}$ at the narrow end (i.e., towards the supraglottal cavity). The opening at the wide end (towards the trachea) is $\Delta W_2 = H_0 + w_{22} - w_{12}$; see Fig. 1 for these and other used symbols.

In the glottis, the flow velocity $V(x,t)$ is assumed to satisfy the static mass conservation law for incompressible flow

$$H(x,t)V(x,t) = H_1 v_o(t)$$

where $H(x,t)$ is the height of the channel in the glottis. In our simple geometry it is

$$H(x,t) = \Delta W_2(t) + \frac{x}{L}(\Delta W_1(t) - \Delta W_2(t)), \;\; x \in [0, L].$$

Now the pressure $p(x,t)$ in the glottis is given by the two equations above and the (static) Bernoulli law

$$p(x,t) + \frac{1}{2}\rho V(x,t)^2 = p_{sub}.$$

Since both vocal folds have two degrees of freedom, this pressure can be reduced to a force pair $(F_{A,1}, F_{A,2})^T$ where $F_{A,1}$ effects at the narrow end of the glottis ($x = L$) and $F_{A,2}$ at the wide end ($x = 0$). This reduction is done by using the total force and moment balance equations

$$F_{A,1} + F_{A,2} = h \int_0^L (p(x,t) - p_{sub})\, dx,$$

$$L \cdot F_{A,1} = h \int_0^L x(p(x,t) - p_{sub})\, dx - p_c \cdot h \frac{H_1}{2} \frac{H_0 - H_1}{2}.$$

The moment is evaluated with respect to point $(x, y) = (0, 0)$ for the lower fold and $(x, y) = (0, H_0)$ for the upper fold.

Evaluation of these integrals yields

$$\begin{cases} F_{A,1} &= \frac{1}{2}\rho v_o^2 hL\left(-\frac{H_1^2}{\Delta W_1(\Delta W_2-\Delta W_1)}\ \cdots\right. \\ &\quad \left.+\frac{H_1^2}{(\Delta W_1-\Delta W_2)^2}\ln\left(\frac{\Delta W_2}{\Delta W_1}\right)\right) - \frac{H_1(H_0-H_1/2)}{4L}hp_c, \\ F_{A,2} &= \frac{1}{2}\rho v_o^2 hL\left(\frac{H_1^2}{\Delta W_2(\Delta W_2-\Delta W_1)}\ \cdots\right. \\ &\quad \left.-\frac{H_1^2}{(\Delta W_1-\Delta W_2)^2}\ln\left(\frac{\Delta W_2}{\Delta W_1}\right)\right) + \frac{H_1(H_0-H_1/2)}{4L}hp_c. \end{cases} \quad (2)$$

B. Vocal folds

The vocal fold model consists of two wedge-shaped vibrating elements that have two degrees of freedom each (see Fig. 1). The distributed mass of these elements can be reduced into three mass points which are located so that m_{j1} is at $x = L$, m_{j2} at $x = 0$, and m_{j3} at $x = L/2$. The elastic support of the vocal folds is approximated by two springs at points $x = aL$ and $x = bL$. Thus the equations of motion for the vocal folds are

$$\begin{cases} M_1\ddot{W}_1(t) + B_1\dot{W}_1(t) + K_1W_1(t) = -F(t), \\ M_2\ddot{W}_2(t) + B_2\dot{W}_2(t) + K_2W_2(t) = F(t) \end{cases} \quad (3)$$

where $W_j = (w_{j1}, w_{j2})^T$ are the displacements of the right and left endpoints of the j^{th} fold, $j = 1, 2$. Here M_j, B_j, and K_j are the mass, damping, and stiffness matrices, respectively

$$\begin{aligned} M_j &= P\begin{bmatrix} m_{j1}+\frac{m_{j3}}{4} & \frac{m_{j3}}{4} \\ \frac{m_{j3}}{4} & m_{j2}+\frac{m_{j3}}{4} \end{bmatrix}, \\ B_j &= \begin{bmatrix} b_{j1} & 0 \\ 0 & b_{j2} \end{bmatrix}, \\ K_j &= \frac{1}{P}\begin{bmatrix} a^2k_{j1}+b^2k_{j2} & ab(k_{j1}+k_{j2}) \\ ab(k_{j1}+k_{j2}) & b^2k_{j1}+a^2k_{j2} \end{bmatrix}. \end{aligned} \quad (4)$$

The entries of these matrices are computed by means of Lagrangian mechanics. The damping matrices B_j are diagonal since the dampers are located at the endpoints of the vocal folds. The springs are located symmetrically around the midpoint $x = L/2$, so that $a = (L/2+l)/L$ and $b = (L/2-l)/L$. The parameter P is used for tuning the oscillation frequency.

During the glottal open phase, the load terms of (3) are given by $F = (F_{A,1}, F_{A,2})^T$ as given in Eq. (2). During the glottal closed phase ($\Delta W_1 < 0$), there are no aerodynamic forces except the counter pressure from the VT. Instead, there is a nonlinear spring force given by the Hertz impact model for the collision of the vocal folds (see [5]):

$$F_H = \begin{bmatrix} k_H|\Delta W_1|^{3/2} - \frac{H_0-H_1/2}{2L}\frac{H_1}{2}h\cdot p_c \\ \frac{H_0-H_1/2}{2L}\frac{H_1}{2}h\cdot p_c \end{bmatrix}.$$

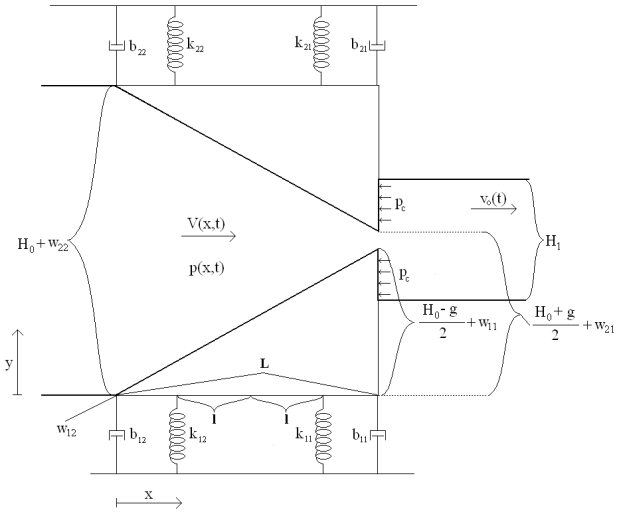

Figure 1: The geometry of the glottis model and the symbols used

C. Vocal tract

We use Webster's horn model resonator as an acoustic load. The Webster's horn equation is

$$\Psi_{tt}(s,t) - \frac{c^2}{A(s)}\frac{\partial}{\partial s}\left(A(s)\frac{\partial\Psi(s,t)}{\partial s}\right) = 0$$

where c is the sound velocity and $\Psi(s,t)$ is a velocity potential. Note that $p = \rho\Psi_t$ in the VT. The parameter $s \in [0, L_{VT}]$ is the distance from the narrow end of the glottis measured along the VT centreline and L_{VT} is the length of the VT. The area function $A(\cdot)$ is the cross-sectional area of the VT, perpendicular to the VT centreline. It is taken from the geometry of [Section 4, 4] corresponding [ø:].

The resonator is controlled by the flow velocity v_o from Eq. (1) through the boundary condition at the glottis end

$$\Psi_s(0,t) = -v_o(t).$$

The boundary condition at lips is a frequency-independent acoustic resistance of the form

$$\Psi_t(L_{VT},t) + \theta c\Psi_s(L_{VT},t) = 0$$

where θ is the normalised acoustic resistance (see [Chapter 7, 8]) regarded as a tuning parameter. This boundary condition represents flow resistance $p = \theta\rho cv$ where v is the flow velocity through the mouth.

The resonator exerts a counter pressure $p_c(t) = \rho\Psi_t(0,t)$ to the vocal folds equations (3) through Eqs. (2), thereby forming a mechanical feedback loop between the vocal folds and the vocal tract.

III. Numerical simulation

The equations of motion (3) are solved with the fourth order Runge–Kutta (RK) method, and the flow equation (1) is solved with the implicit Euler method. The VT is discretised by the Finite Element Method using piecewise linear elements and the physical energy norm of the Webster's equation. The Crank–Nicolson method is applied for temporal discretisation. Especially the FEM-solver performs faster with a constant time step, because the state update equations are the same, and we spare one matrix inversion on every step. Thus we keep time steps ΔT constant except when the glottis either closes or opens.

The load force in Eq. (3) is discontinuous (in fact, singular) when $\Delta W_1 \to 0^+$. The singularity is removed by replacing the aerodynamic force by zero when $\Delta W_1 < \epsilon$. Since v_o is small when ΔW_1 is small, the solution is not sensitive to the choice of $\epsilon > 0$. The discontinuity is dealt with by locating the time of closure/opening by interpolation and restarting computation from there, see [pp. 12-14, 1]. For this we use the second degree interpolating polynomial so that numerical error is of order $\mathcal{O}(\Delta T^3)$. Because the number of the exceptional steps (at times of opening or closure) does not depend on ΔT, the total error is also of order $\mathcal{O}(\Delta T^3)$. The total error of the RK-method is of order $\mathcal{O}(\Delta T^4)$ but considering its smaller computational burden, this interpolation method is an appropriate way to treat the discontinuous load in Eq. (3).

IV. Parameter estimation and model validation

The model parameters are listed in Table 2. When estimating model parameters, we use three LF-pulses, obtained by fitting the LF-waveforms to glottal flow derivatives, estimated from natural speech using an automatic inverse filtering method [2]. Speech data consisted of vowels [aː] produced in breathy, normal, and pressed phonation by a male speaker.

The damping matrix in Eq. (3) is adjusted so that the vibration is stable and sustained oscillation occurs. The rest of the parameters in Eqs. (2) and (3) are from literature (see Table 2). Note that only the relative values of parameters p_{sub}, C_{iner}, and C_g in Eq. (1) matter, and increasing the subglottal pressure p_{sub} only increases the height of the pulse.

If the changes in $\Delta W_1(t)$ are neglected, we have three parameters characterizing the flow pulse: the pulse length (parameter P), the height (parameter p_{sub}), and the inclination (parameter C_{iner}). The inertance C_{iner} and the parameter P in Eqs. (1) and (4)

Table 1: The inertance C_{iner} in Eq. (1) for different LF-pulses. The Mean error 1 is the error from the beginning of the pulse to the peak value and Mean error 2 from the peak value to the closure.

	Breathy	Normal	Pressed
C_{iner} $(\frac{kg}{m^4 s})$	524.8	530.9	630.8
Mean error 1 (%)	2.93	1.45	2.09
Mean error 2 (%)	4.14	3.07	2.14

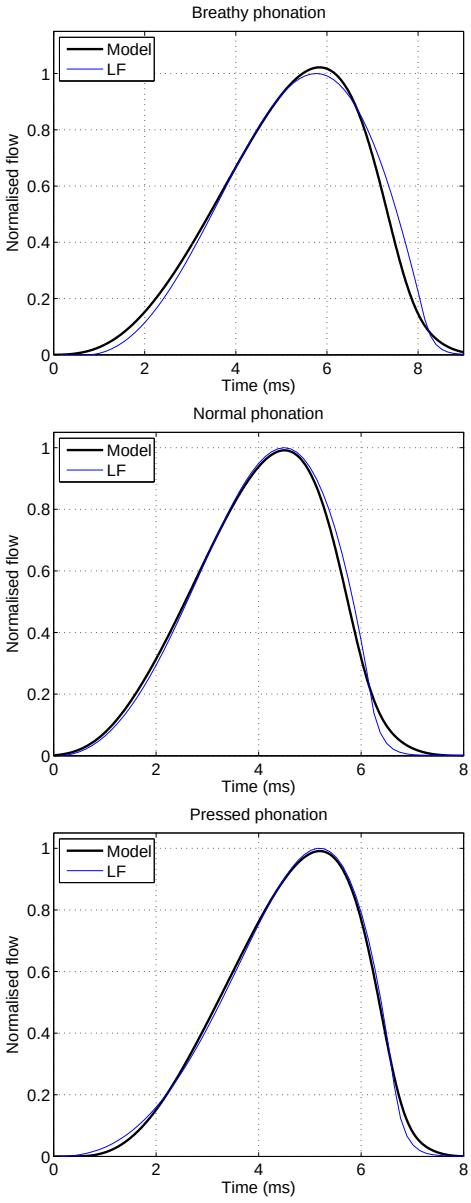

Figure 2: Three LF-pulses corresponding to breathy, normal, and pressed phonation and the corresponding simulated pulses after parameter tuning

are tuned so that the pulse shape matches optimally a measured LF-pulse. Since the measured pulse height has no scale, the parameters p_{sub} and C_g are tuned so that their magnitudes are realistic and the total flow (area) of the pulse matches that of the LF-pulse.

This parameter tuning is performed so that the instants of maximal flow in the measured LF-pulse and the simulated pulse coincide. The length of the pulse (that is, parameter P) and the parameter C_{iner} are then varied in order to minimise the squared error of the two pulses. The measured and the simulated pulses after parameter tuning are shown in Fig. 2. The optimal values of C_{iner} are shown in Table 1.

V. Conclusions

We presented a flow mechanical glottis model to be used as a real-time source for a VT simulator. The model is validated against three LF-pulses that have been estimated by parameterising glottal flow derivatives, obtained by inverse filtering natural utterances. We conclude that LF-pulses corresponding to different types of phonation can be faithfully constructed.

It is interesting to observe how the inertance term C_{iner} in Eq. (1) increases monotonically when the phonation type changes from breathy to normal, and then to pressed. When producing pressed voices, speakers use a vibration mode which is characterised by a small abduction quotient (see [pp. 260, 9]) which, in turn, results in a flow pulse with a shorter relative length of the closing phase. The present study indicates that this phenomenon is reflected as the increase of the parameter value C_{iner}.

It is worth noting that here we have only optimised the flow parameters of Eq. (1) according to measured data. Depending on the type of phonation, a speaker controls not only the subglottal pressure p_{sub} but also the mechanical properties of the vocal folds, corresponding to Eq. (3). For example, the glottis barely closes during breathy phonation, whereas the glottal vibration patterns are similar in all of our simulations.

Table 2: The model parameters

	Source of the value
h, H_0, L, l	From [6]
H_1	Through condition $hH_1 = A(0)$
p_{sub}, C_g	Only the relative magnitudes of p_{sub}, C_{iner} and C_g are relevant.
g	Set so that glottal area matches typical observations
m_{ij}, k_{ij}	From mechanical properties of the model presented in [6]
b_{ij}	Tuned so that sustained, stable oscillation occurs
k_H	From [5]
$A(\cdot)$, L_{VT}	From [4]
θ	Tuning parameter regulating energy dissipation in the VT
P, C_{iner}	Optimised (see Sec. IV)

References

[1] Aalto, A. (**2009**). A low-order glottis model with nonturbulent flow and mechanically coupled acoustic load, Master's thesis, TKK, Helsinki. Available at `http://math.tkk.fi/research/sysnum/`.

[2] Alku, P. (**1992**). "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," Speech Communication **11**, 109–118.

[3] Fant, G., Liljencrants, J., and Lin, Q. (**1986**). "A four-parameter model of glottal flow," Tech. rep., QPRS: Dept. for Speech, Music and Hearing, Stockholm.

[4] Hannukainen, A., Lukkari, T., Malinen, J., and Palo, P. (**2007**). "Vowel formants from the wave equation," Journal of the Acoustical Society of America Express Letters **122**, EL1–EL7.

[5] Horáček, J., Šidlof, P., and Švec, J. (**2005**). "Numerical simulation of self-oscillations of human vocal folds with Hertz model of impact forces," Journal of Fluids and Structures **20**, 853–869.

[6] Horáček, J. and Švec, J. (**2002**). "Aeroelastic model of vocal-fold-shaped vibrating element for studying the phonation threshold," Journal of Fluids and Structures **16**, 931–955.

[7] Ishizaka, K. and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell System Technical Journal **51**, 1233–1268.

[8] Morse, P. and Ingard, K. (**1968**). Theoretical Acoustics, McGraw–Hill.

[9] Titze, I. (**1994**). Principles of voice production, Prentice Hall.

[10] Titze, I. (**2008**). "Nonlinear source-filter coupling in phonation: Theory," Journal of the Acoustical Society of America **123**, 2733–2749.

[11] Titze, I., Riede, T., and Popolo, P. (**2008**). "Nonlinear source-filter coupling in phonation: Vocal exercises," Journal of the Acoustical Society of America **123**, 1902–1915.