ACUSTICA · acta acustica
Vol. 84 (1998) 1135 – 1150

A Symmetrical Two-Mass Vocal-Fold Model Coupled to Vocal Tract and Trachea, with Application to Prosthesis Design

N. J. C. Lous*, G. C. J. Hofmans, R. N. J. Veldhuis*, A. Hirschberg
Eindhoven University of Technology, P.O. box 513, 5600 MB Eindhoven, The Netherlands

Summary
We propose a two-mass model for the vocal folds. The aerodynamic force resulting from the air flow through the glottis is distributed over both masses, as opposed to some earlier models in which the force is allocated to the upstream mass only. This allows the choice of a symmetrical vocal-fold structure, with two identical mass-spring systems. The number of mechanical parameters is thus reduced. Their choice is inspired upon analysis of the eigenmodes of the actual vocal folds. The new aerodynamic force distribution allows acoustic feedback from vocal tract and trachea to be considered; sub- and supra-glottal systems are modelled as transmission lines. The flow model includes a simple flow-separation description.
Parameter values are based on physiological measurements and calculations reported in literature. However, as in any vocal-fold model it is inevitable that parameters also partly make up for simplifications in the model.
The model predictions are shown to compare well to in-vivo experimental data from literature concerning acoustic feedback. The two-mass model is applied to study the effect of size and position of a vocal-fold prosthesis on its performance.
We indicate how the model should be modified to provide order-of-magnitude estimates of the effect of flow inertia and fluid viscosity. The described improvement of the flow model is not of predictive value within the framework of the simplified mechanical model we propose.

PACS no. 43.70.Bk

1. Introduction

A detailed ad-initio numerical simulation of human speech production would involve a very elaborate physiological model, including neural control of lungs, trachea and vocal tract. It is therefore impossible to achieve. Even when we focus on the vocal folds, and use simplified acoustic models for lungs and vocal tract, it is extremely difficult to simulate the speech production process numerically. The equations of motion of the fluid (air) and of the tissues show dramatic non-linearities. In the fluid, we have to predict unsteady flow separation from a smoothly curved surface. The vocal-fold collision involves a complex non-linear mechanical behaviour. For instance, the collision generates elastic waves propagating in the vocal folds, and in the tissue surrounding them, as demonstrated by the vibration of ones throat during phonation. This vibration costs energy, which is dissipated from the vocal-fold oscillations, thus causing the oscillations to be damped. For a proper treatment of the radiation of elastic waves generated by the collision, a detailed physiological description of the trachea is needed. Altogether, a complex analysis, and huge computing power would be involved.

There is obviously room for simplified models, some of which are still quite detailed. Alipour and Titze [1] have presented a complex two-dimensional model that takes the layered structure of the vocal folds into account, and in which the flow is quite accurately described. Such a model is at the upper edge of our current numerical power. At the other extreme we find the two-mass model of Ishizaka and Flanagan [2]. In spite of its simplicity and shortcomings, it has caught some essential features of human speech production. Models with such a level of simplification can be used for real-time speech synthesis, which makes them very attractive. As pointed out by Berry *et al.* [3], a systematic study of two-mass models provides information on pathological behaviour of the vocal folds. Lucero [4] used the original Ishizaka and Flanagan two-mass model [2], and a 'body-cover' modification of this model by Story and Titze [5], to show that by suitable parameter variation, a two-mass model can even simulate falsetto-type vocal-fold oscillations. Herzel [6] applied a two-mass model to investigate into the influence of vocal-fold asymmetry. The above illustrates that two-mass models are useful research tools.

The major draw-back of this kind of vocal-fold models is pointed out by Story and Titze [5]: there is no simple relationship between the parameters in the model and the physiology of the vocal folds. Although many parameter values are based on physiological measurements, they have to be tuned to compensate for over-simplifications in the model.

Simplified models of the vocal folds are not only useful for speech synthesis or medical diagnostics. We applied a two-mass model to explore the acoustic response of a vocal-fold prosthesis. Such a prosthesis could replace the vocal folds when these have been removed as a consequence of cancer. The prosthesis would probably be much smaller than the vocal folds, and it may not be possible to place it exactly at the original vocal-fold position. We consider both the influence of size reduction and implantation location.

The geometry of our model is inspired by similar work by Liljencrants [7]. We consider a piece-wise straight glottis wall, supported by two point masses. The model aerodynam-

Received 13 May 1997,
accepted 20 March 1998.

* Institute for Perception Research (IPO)

ACUSTICA · acta acustica
Vol. 84 (1998)

ics are based on work by Pelorson *et al.* [8], Bogaert [9] and Veldhuis *et al.* [10]. The aerodynamic description is combined with a flow-separation model due to Liljencrants [7], tuned to our geometry. The aerodynamic force is assumed to act on the three mass-less plates representing the flow channel wall. This allows direct calculation of the force acting on each of the point masses in the model by considering equilibrium of force and moments.

The distribution of the aerodynamic force over both the point masses differs from the approach of Pelorson *et al.* [8] and Veldhuis *et al.* [10]. In the application of their model, they assume the complete force to act on the masses modelling the pharynx side of the vocal folds. No force is assumed to act on the masses modelling the larynx side of the vocal folds, and the latter masses are considered to be smaller than those modelling the pharynx side. We assume the masses at the larynx and pharinx side of the vocal-fold model to be indentical, and associate all masses in the model with identical mass-spring systems. This symmetry allows a 'fair' force allocation to each of the masses, and a reduction of the number of parameters.

The mechanical parameters can easily be determined from the eigenmode frequencies of the actual vocal folds as calculated by Berry and Titze [11]. However, for the model to perform well, we have chosen eigenmode frequencies that more closely correspond to the eigenmode frequencies of the two-mass model by Ishizaka and Flanagan [2]. Again, parameter values are not only dictated by physiology, but also have to make up for over-simplifications in the model.

To include acoustic effects, the two-mass model is coupled to a transmission-line description of the vocal tract. Vocal-tract shapes corresponding to the vowels used in the model are taken from Fant [12]. The trachea and lungs are modelled in a similar way, following a description by Ishizaka *et al.* [13]. The lung model is designed to realise the proper cut-off frequency of about 500 Hz.

A great surprise of our model is, that it produces typical glottal pulses with the assumed symmetrical vocal fold structure, featuring identical mass-spring systems! This may seem obvious from a physiological point of view, but it is new to the world of two-mass models, in which an asymmetrical vocal-fold structure has been assumed since the original paper by Ishizaka and Flanagen [2]. The model predictions will be shown to improve when approximated inertive and viscous terms are included in the flow description.

As mentioned before, we apply our two-mass model to prosthesis design. As a precise relationship between the model and an actual structure for the prosthesis is still to be found, the results obtained are not more than a first approximation. Even if they are not directly relevant for the design of an actual prosthesis, they provide some more insight into the effect of coupling vocal-fold oscillations to the acoustics of sub- and supra-glottal systems.

We also hope that our non-conventional two-mass model, presented in this paper, will provide inspiration for new vocal-fold descriptions.

In Section 2 we present a basic model, in which we assume a quasi-stationary incompressible frictionless flow within the glottis. We systematically discuss the choice of parameters, and their relationship to literature or reality, in Section 3. In Section 4, comparison with in-vivo data obtained by Cranen [14] and Rothenberg [15] shows that the model provides a reasonable prediction of acoustic feedback on the vocal-fold oscillations from both vocal tract and sub-glottal system. Application of the model to explore the effect of size and position of a prosthesis on its performance is discussed in Section 5. Finally, in Section 6 we discuss a number of aerodynamic aspects in more detail. We reconsider the flow-separation model, and partly drop the assumption that the flow is quasi-stationary and frictionless to determine order-of-magnitude estimates for the influence of inertia and viscosity on the flow.

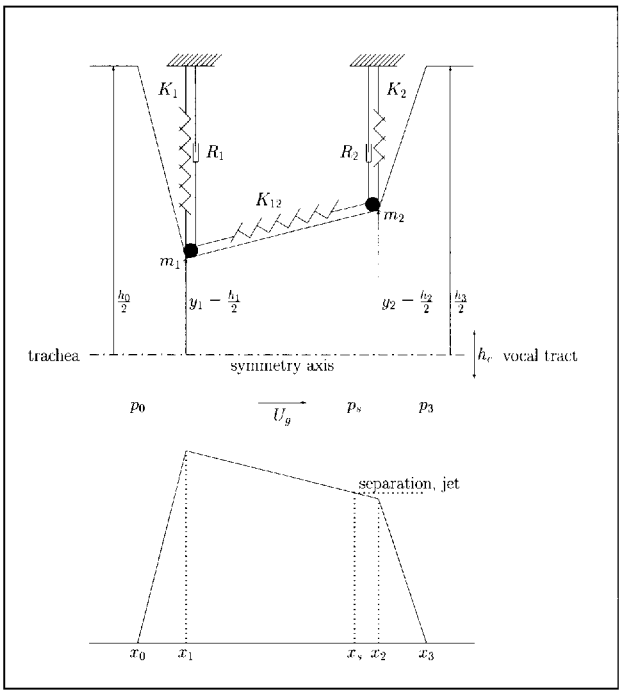

Figure 1. Yet another two-mass model for the vocal folds.

2. Yet another two-mass model for the vocal folds

In this section we describe the two-mass model applied in our study. Using a two-mass model corresponds to the assumption that there is a capture of elastic waves in the vocal-fold structure, and that these waves can be reasonably well described by a lumped model. One could also consider a single-mass model driving a surface wave; this may be a subject for further research.

Our two-mass model is based on existing models by Ishizaka and Flanagan [2], Pelorson *et al.* [8], and Veldhuis *et al.* [10]. We do not yet justify the assumptions used to obtain the model, nor do we discuss any parameter values. These topics, and comparison to the two-mass models by Ishizaka and Flanagan [2] and Pelorson *et al.* [8] will be addressed in Section 3.

2.1. Vocal-fold geometry

We assume a two-dimensional structure of the vocal folds as shown in Figure 1; the third dimension is taken into account by assuming the vocal folds to have a length L_g. The air flows from left to right, from trachea to vocal tract. We assume symmetry with respect to the flow channel axis, so that we only need to consider one vocal fold in detail. The shape of each of the vocal folds is described by three rigid mass-less plates. These three plates join the points $i-1$ and i with co-ordinates (x_{i-1}, y_{i-1}) and (x_i, y_i) for $i = 1, 2, 3$; the abscissa are fixed. Between these points, the flow channel height is a linear function of x,

$$h(x,t) = \frac{h_i(t) - h_{i-1}(t)}{x_i - x_{i-1}}(x - x_{i-1}) + h_{i-1}(t), \qquad (1)$$

in which h_0 and h_3 are constant. For $i = 1, 2$, $h_i(t)$ equals twice the displacement $y_i(t)$ of point i with respect to its rest position on the symmetry axis. If $y_i(t) \leq 0$, we set $h_i(t) = 0$.

2.2. Aerodynamics in the glottis

2.2.1. Pressure and flow

At the entrance x_0 of the glottis, a pressure discontinuity $p_{\mathrm{phon}}(t)$ is imposed, modelling the lung pressure that drives phonation. The total pressure at the glottis side of x_0 equals $p_0(t) = p_{\mathrm{phon}}(t) + p_{0,\mathrm{ac}}(t)$, where $p_{0,\mathrm{ac}}(t)$ is the acoustic pressure at the trachea side of x_0 resulting from interaction of the vocal folds with the sub-glottal system.

We assume a one-dimensional quasi-stationary frictionless and incompressible flow from the trachea to a point $x_s(t)$ where the flow separates from the wall to form a free jet. The pressure is described by the corresponding approximation of Bernoulli's equation,

$$p(x,t) + \frac{\rho_0}{2}\left(\frac{U_g(t)}{h(x,t)L_g}\right)^2 = p_0(t) + \frac{\rho_0}{2}\left(\frac{U_g(t)}{h_0 L_g}\right)^2, \qquad (2)$$

where ρ_0 is the mean air density, and $U_g(t)$ denotes the volume flux through the glottis. As we assumed the flow between trachea and $x_s(t)$ to be one-dimensional, quasi-stationary and incompressible, continuity of flux implies that $U_g(x,t)$ is independent of x within the glottis.

We propose to use the model of Liljencrants [7] to determine $x_s(t)$. If $h_2(t) > s h_1(t)$, we assume $x_s(t) = x_2$. Otherwise, the separation point $x_s(t)$ is determined from the condition $h_s(t)/h_1(t) = s$, in which we use that $h(x,t)$ is a linear function of x between x_1 and x_2. The constant parameter s is referred to as the separation constant.

We assume turbulence in the free jet to dissipate all the kinetic energy with a negligible recovery of pressure. This implies that for $x > x_s(t)$ the pressure in the glottis is equal to the pressure $p_3(t)$ at the entrance of the vocal tract. This free jet model introduces the dissipation necessary to obtain a control of the flow by the movement of the vocal folds.

If the glottis is closed, which is the case if either $h_1(t) \leq 0$ or $h_2(t) \leq 0$, the volume flow $U_g(t)$ is zero. Within the closed part of the glottis, where the tissue of both vocal folds actually touches, we use $p(x,t) = 0$. The pressure equals $p_0(t)$ at the upstream side of the closure, and $p_3(t) = p_{3,\mathrm{ac}}(t)$ at the downstream side of the closure. The acoustic pressure $p_{3,\mathrm{ac}}(t)$ results from acoustic effects in the vocal tract.

2.2.2. Aerodynamic force

The aerodynamic forces on the plates are induced by the pressure $p(x,t)$ along the flow channel. Since the air flow is assumed to be frictionless, $p(x,t)$ completely determines these forces.

As m_1 and m_2 support the plates, an aerodynamic force $\mathbf{F}_{hi}$ is transmitted to each mass m_i, $i = 1, 2$. By considering a force and torque balance on each of the mass-less plates, the y-component of the aerodynamic force on point i, $i = 1, 2$, is found to be

$$F_{hi}(t) = \int_{x_{i-1}}^{x_i} \left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) p(x,t)\mathrm{d}x + \int_{x_i}^{x_{i+1}} \left(\frac{x_{i+1} - x}{x_{i+1} - x_i}\right) p(x,t)\mathrm{d}x. \qquad (3)$$

This integral can be solved analytically for the pressure distribution $p(x,t)$ resulting from the Bernoulli equation (2). We refer to Appendix A1 for details.

If either $h_1(t) \leq 0$ or $h_2(t) \leq 0$, the glottis is closed. Assuming $h_1(t) > 0$ and $h_2(t) \leq 0$, there is a point x_{trach} between x_1 and x_2 where $h(x) = 0$, as $h(x)$ is a linear function of x. Similarly, there is a point x_{tract} between x_2 and x_3 where $h(x) = 0$ because h_3 is constant and positive. The pressure $p(x,t)$ then is $p_0(t)$ between x_0 and x_{trach}, and $p_3(t)$ between x_{tract} and x_3. Equation (3) yields the corresponding aerodynamic forces on points 1 and 2. Notice that between x_{trach} and x_{tract}, where the vocal folds are in mechanical contact, we have set $p(x,t) = 0$. This contact area does therefore not contribute to the aerodynamic force $F_{hi}(t)$. The vocal fold collision is accounted for in the mechanical model, which we describe in the sequel.

If $h_1(t) \leq 0$, x_{trach} and x_{tract} can be determined in similar way, and so can the aerodynamic forces.

2.3. Vocal-fold mechanics

The mechanical behaviour of the vocal-fold tissue is described in terms of lumped inertia, elasticity, and damping. The vocal-fold inertia is modelled by two point masses (or, actually, line masses of length L_g) in each fold. These point masses m_1 and m_2 are assumed to move in the y-direction only, perpendicular to the flow channel axis. The origin $y = 0$ is taken on the axis, corresponding to the assumed symmetry.

To mimic elasticity and damping, each mass m_i is attached to an (imaginary) rigid wall by means of a spring/dashpot combination which exerts a force $(-K_i y_i - R_i \dot{y}_i)$ on mass m_i for $i = 1, 2$. The spring behaviour is described by K_i,

ACUSTICA · acta acustica
Vol. 84 (1998)

and $R_i = 2\xi_i\sqrt{K_i m_i}$ controls the dashpot damping. The value of R_i is assumed to include the damping caused by radiation of elastic waves generated by collision of the vocal folds. An additional spring between m_i and m_j induces a coupling force between the masses. The rest length of the coupling spring is defined such that the force on m_i is well approximated by $K_{12}(y_i - y_j)$. Notice that all forces are functions of the mass positions y_i and y_j.

The coefficients K_i and R_i are not constant, as we assume the vocal-fold collision to induce a non-linear behaviour. We assume a stepwise increase of the spring constants K_i from an initial value K_{io} to $K_{ic} = 4K_{io}$ when $h(x_i) < h_c$, where h_c denotes a constant critical height. The value of the damping coefficient is then increased from $\xi_{io} < 1$ to $\xi_{ic} \approx 1$, such that R_{io} increases to R_{ic} and describes a critical damping behaviour of mass/spring system i.

The mechanical model can be summarised by the equations of motion

$$m_i\ddot{y}_i + R_i\dot{y}_i + K_i y_i + K_{ij}(y_i - y_j) = F_{hi}. \quad (4)$$

If required, this mechanical model can easily be altered to allow mass movements in the x-direction as well. Elasticity and damping in the x-direction can be taken into account by introducing two extra spring/dasphot combinations with coefficients $\hat{K}_i$ and $\hat{R}_i$, each attached to an (imaginary) rigid wall. These spring/dashpot combinations exert a force $(-\hat{K}_i x_i - \hat{R}_i\dot{x}_i)$ on m_i. A coupling spring K'_{12} between the masses induces an extra force on m_i, acting parallel to the x-axis, that is well approximated by $K'_{12}(x_i - x_j)$ (note: in general $K'_{12} \neq K_{12}$). This yields an equation of motion similar to (4), in which the rest positions of masses m_i have to be included.

2.4. Acoustics in vocal tract, trachea, and lungs

The vocal tract is represented as a transmission line of N cylindrical, hard-walled sections of fixed length L_i each, and with cross-sectional areas A_i, $i = 1, \ldots, N$. In each section i, we assume one-dimensional acoustic pressure wave propagation $p_{\mathrm{ac},i}(x,t) = p_i^+(x - c_0 t) + p_i^-(x + c_0 t)$, ignoring convective effects. The speed of sound in air c_0 is assumed to be constant, neglecting temperature gradients. The particle velocity is given by $u_i(x,t) = (p_i^+(x - c_0 t) - p_i^-(x + c_0 t))/\rho_0 c_0$.

At the interface between sections i and $i+1$, the jump in cross-sectional area from A_i to A_{i+1} is taken into account by introducing reflection and transmission relations following from continuity of flow ($A_i u_i = A_{i+1} u_{i+1}$) and pressure ($p_i^+ + p_i^- = p_{i+1}^+ + p_{i+1}^-$).

At the lips, the pressure waves are partly radiated and partly reflected, depending on the lip impedance. In the frequency domain, this boundary condition can be written as

$$\frac{A_N Z_N}{\rho_0 c_0} = \frac{p_N^+(\omega) + p_N^-(\omega)}{p_N^+(\omega) - p_N^-(\omega)}, \quad (5)$$

where A_N is the cross-sectional area of the vocal tract at the lips, and Z_N denotes the corresponding impedance. Z_N can be approximated as

$$Z_N = \frac{\rho_0 c_0}{A_N}\left[\mathrm{j}\frac{ka}{2} + \frac{(ka)^2}{2}\right] \approx \frac{1}{2}\frac{\rho_0 c_0}{A_N}\frac{\mathrm{j}ka}{1 + \mathrm{j}ka}; \quad (6)$$

$k = \omega/c_0 = 2\pi f/c_0$, and $a = \sqrt{A_N/\pi}$ denotes the effective radius of the mouth opening. This corresponds to a typical low frequency radiation behaviour of a flanged pipe in free space. These frequency-domain results are implemented in the time domain by means of a time-invariant filter (see Appendix A2).

Trachea and lungs are similarly modelled as a transmission line. A straight tube of constant cross-sectional area A_{trach} and length L_{trach} is chosen to describe the trachea. The lungs are modelled as an exponential horn, extending from right to left in the convention of Figure 1. Assuming $x = 0$ at the leftmost horn end, we can express its cross-sectional area as $A_{\mathrm{horn}}(x) = A_{\mathrm{trach}}\exp(-\alpha(x - L_{\mathrm{horn}}))$. The cross-sectional areas A_i of the sub-glottal transmission line are chosen as a stepwise approximation of the trachea and lung configuration.

For an infinite horn, the plane-wave approximation predicts all sound waves to be transmitted for frequencies exceeding the horn cut-off frequency $f_c = 4\pi\alpha/c_0$, while sound reflection does not occur for such frequencies. An appropriate choice for f_c, from which the value of α follows, is discussed in Section 3.4.

Sound radiation at the end of the horn is taken into account in the same way as at the lips. In our model, there is only little reflection above f_c because $A_{\mathrm{trach}}\exp(\alpha L_{\mathrm{horn}})$ is relatively large.

2.5. Coupling the vocal-fold model to vocal-tract, trachea, and lung acoustics

The acoustic model is connected to the incompressible quasi-stationary frictionless flow description within the glottis by assuming continuity of flow and pressure at x_0 and x_3. Denoting by $Z_1 = \rho_0 c_0/A_1$ the characteristic impedance of the first vocal-tract section, continuity of flow implies $Z_1 U_g(t) = p_1^+(t) - p_1^-(t)$. Continuity of pressure at x_3 then gives

$$\begin{aligned} p_3(t) &= p_{3,\mathrm{ac}}(t) = p_1^+(t) + p_1^-(t) \\ &= Z_1 U_g(t) + 2p_1^-(t). \end{aligned} \quad (7)$$

When the glottis is closed, the same equations can be used with $U_g(t) = 0$, corresponding to total reflection of the sound waves at the glottis. At x_0, a similar expression results for $p_{0,\mathrm{ac}}(t)$.

2.6. Implementation

The two-mass model for the vocal folds, and the acoustic description of vocal tract, trachea, and lungs, are implemented in Matlab. Details concerning this implementation, based on a programme by Bogaert [9] and Veldhuis *et al.* [10], can be found in Appendix A2.

3. Parameter choice and discussion

3.1. Geometrical parameters

The vocal folds are assumed to be symmetrical with respect to the flow channel axis. This is a simplification, as perfect symmetry is an unlikely biological feature. Berry *et al.* [3] discuss the necessity of an asymmetrical model in order to include chaotic effects. However, symmetry is a reasonable assumption if the vocal-fold model is used in prosthesis design, the symmetry of which can be designed.

The choice of geometrical parameters is based on measurements. The length L_g of the oscillating, membranous portion of the vocal folds is 14 mm as in Ishizaka and Flanagan's model [2], and in correspondence with measurements by Hirano *et al.* [16]. However, the depth $x_2 - x_1$ of the membranous part of the vocal folds is chosen to be 2 mm, instead of the 2.5 mm measured by Bear [17] or the 3 mm that Hirano *et al.* report [16]. The inlet depth $x_1 - x_0$, and the outlet depth $x_3 - x_2$ are both set to 0.2 mm, a choice that particularly under-estimates the inlet depth.

The model predictions degrade rapidly if some or all of the geometrical parameters are increased to their measured values. This may indicate that our model over-estimates the aerodynamic forces, that are directly proportional to the flow channel wall area they act on. Under-estimating the vocal-fold size compensates for this model artefact. This is particularly true for the inlet depth $x_1 - x_0$, which is a very critical parameter, far more than the outlet depth $x_3 - x_2$. The aerodynamic force acting on inlet and outlet is directly proportional to the inlet and outlet area, and therefore to their depth. Furthermore, although the aerodynamic force at the pharynx side of the vocal tract is influenced by acoustic feedback from the sub-glottal system, it is dominated by the imposed phonation pressure of 800 Pa. Thus, the force acting on the inlet has a non-zero DC-component. At the vocal-tract side, no pressure is imposed, and there is no DC-component in the aerodynamic force. This aerodynamic asymmetry explains the difference in sensitivity of the model to changes in inlet- and outlet depths, and clarifies the model's fierce reaction to increasing $x_1 - x_0$ to more than about 2 mm.

The aerodynamic forces can also be (relatively) decreased by increasing the total vocal fold mass, as we have chosen to do (see Section 3.3). None of these parameter tunings correspond to measurements. They make up for over-simplifications in the model. Among the set of parameters discussed above, it does not really matter which ones we use to compensate for the over-estimation of the aerodynamic force. Within reason, the only relevant parameter seems to be the ratio of F_{hi} and m_i, since this ratio determines the mass acceleration and therefore the vocal-fold movement. As an example, the glottal-pulse shapes do not change essentially if we set $x_2 - x_1 = 3$ mm as measured, while decreasing the vocal-fold mass to 0.1 g and L_g to 10 mm. This is the parameter set used by Story and Titze [5].

The height h_0 at the vocal-fold entrance is found from the trachea area A_{trach} as $h_0 = A_{\text{trach}}/L_g$. Our choice of A_{trach} (see Section 3.4) yields $h_0 \approx 1.8$ cm. The glottis height h_3 at x_3 is determined from the vocal-tract shape approximations by Fant [12]. For the /a/, Fant reports the first vocal-tract section to have an area $A_1 = 2.6\,\text{cm}^2$, from which $h_3 = A_1/L_g \approx 1.9$ cm. Similarly, $h_3 \approx 2.3$ cm if the vocal tract corresponds to the /i/.

3.2. Aerodynamic parameters

We have chosen a pressure discontinuity at x_0 to model the lung pressure that drives phonation. This makes it possible to use nearly identical transmission lines to describe acoustics in the sub- and supra-glottal systems. In particular, modelling sound radiation from the lungs would be more complex than it is in our description if the driving pressure were modelled at the end of the lung horn. The imposed pressure discontinuity at x_0 is $p_{\text{phon}}(t) = 800$ Pa, as in Story and Titze's body-cover model [5] and close to the value used by Ishizaka and Flanagan [2].

Experimental work by Pelorson *et al.* [8] has demonstrated that there are no inlet losses as assumed by Van den Berg *et al.* [18]. Ishizaka and Flanagan [2] include these inlet losses by assuming flow separation at a sharp inlet of the vocal folds. However, such edges are introduced for convenience, mainly to ease calculations. Within the framework of a quasi-one-dimensional flow model, one should not a priori assume flow separation at these 'virtual' edges, but rather ignore the local details of the flow. The absence of inlet losses in our model partly explains why for the same $p_{\text{phon}}(t)$ we find an aerodynamic force that is larger than in Ishizaka and Flanagan's model [2], and why we need to over-estimate the total vocal-fold mass compared to Ishizaka and Flanagan [2].

Flow separation in the diverging part of the glottis, however, is essential to describe how the vocal-fold oscillations control the volume flow. Teager and Teager [19] and Hirschberg *et al.* [20] have pointed out that ignoring flow separation results into a paradoxical situation in which the movement of the vocal folds has almost no influence on the flow through the glottis (d'Alembert paradox). When flow separation is assumed to occur at a fixed point, e.g. at x_2, our two-mass model grossly over-estimates the vocal-fold velocity upon closure. This is in line with findings of Pelorson *et al.* [8].

In Ishizaka and Flanagan's model [2], flow separation occurs at edges that are fixed in space. Pelorson *et al.* [8] explored the use of a smooth glottal geometry, combined with a simple flow-separation description derived from a boundary layer theory. The use of such a flow model appeared to dramatically influence the performance of the two-mass model.

We apply an even simpler flow-separation model by Liljencrants [7], in which he proposes a separation constant $s = 1.2$. Both measurements and theoretical calculations show that the exact value to be chosen for s depends on the assumed vocal-fold geometry. The separation constant is therefore yet another parameter of which the value (partly) has to make up for simplifications in other aspects of the model. A critical discussion of the separation criterion is given in Section 6.

ACUSTICA · acta acustica
Vol. 84 (1998)

3.3. Mechanical parameters

Pelorson *et al.* [8] extended their two-mass model with a more elaborate flow description. However, in their mechanical model the aerodynamic force is assumed to act on the first (and larger) mass only, at the upstream-side of the vocal folds. The second mass, which Pelorson *et al.* [8] and Veldhuis *et al.* [10] assume to be much smaller as proposed by Ishizaka and Flanagan [2], is pulled along by the first. It behaves like a child which is held by the hand by a running parent: the child is always out of phase. The movement of the second mass thus realises the converging glottis shape in the opening phase, and the diverging shape while closing. These alternating glottis shapes result into a net energy transfer from the flow to the vocal-fold oscillation.

We propose a distribution of the aerodynamic force over both masses. The choice of mechanical parameters is heavily influenced by this force allocation. To obtain reasonable glottal-flux pulses, we have to reduce the ratio m_1/m_2 quite drastically compared to the models of Ishizaka and Flanagan [2], Pelorson *et al.* [8], and Veldhuis *et al.* [10]. This may seem unfortunate, but it also indicates that the allocation of forces to the point masses is extremely influential – more so than the inclusion of inertive or viscous terms in the flow description. The flow models of Ishizaka and Flanagan [2], Pelorson *et al.* [8], and Veldhuis *et al.* [10] allow approximations of such inertive (and viscous) terms to be included, resulting in improvements of the glottal-pulse shapes they predict. Starting from a similar base model, we find that changing the aerodynamic force allocation has a far larger effect, requiring actual parameter changes for the model to remain operative. We therefore do include a more detailed description of the aerodynamic force distribution, but for the moment neglect inertive and viscous terms in the flow model.

The total vocal-fold mass is slightly increased from the 0.15 g proposed by Ishizaka and Flanagan [2] to 0.2 g. This increase compensates for the absence of inlet losses in our model as we already discussed in Section 3.2. The new force distribution makes it reasonable to set $m_1 = m_2 = 0.1$ g, and correspondingly $K_1 = K_2$. This reduces the number of mechanical parameters, and allows a simple approach to a functional choice of their values. The spring constants $K_1 = K_2 = K$ are determined from analysis of the modes of vibration of the mechanical system. We want to fit the frequencies of these modes to the frequencies of the main modes of vibration of the actual vocal folds. Berry and Titze [11] show that two modes are dominant. One of them is the mode at angular eigenfrequency ω_1 in which masses m_1 and m_2 move in phase, i.e. $y_1(t) = y_2(t)$. The second dominant mode, at angular eigenfrequency ω_2, is the one in which they move in counter-phase. A simple derivation by Morse and Ingard [21] shows that the corresponding modes of the frictionless mechanical system (without dashpots) occur at $\omega_1^2 = K/m$, and $\omega_2^2 = (K + 2K_{12})/m$. This yields

$$K = m\omega_1^2 \qquad \text{and}$$
$$K_{12} = \frac{1}{2}m(\omega_2^2 - \omega_1^2) \approx \sigma K, \tag{8}$$

where $\sigma = (\omega_2 - \omega_1)/\omega_1$. The simplicity of these relationships between spring constants and angular eigenfrequencies is a consequence of the symmetry in the model. For a continuum model, Berry and Titze [11] report results that correspond to $\omega_1/(2\pi) \approx 150$ Hz, and $0.01 \leq \sigma \leq 0.06$. Again, the relationship between Story and Titze's continuum model [5] and the lumped two-mass model is not obvious. We found the best results for $\omega_1/(2\pi) = 100$ Hz, so that the first mechanical eigenmode corresponds to the fundamental oscillation frequency of the vocal folds. We follow Ishizaka and Flanagan [2] in the choice $\sigma \approx 0.6$. Usings Story and Titze's results for ω_1 and ω_2 we did not obtain oscillations of the two-mass model.

Similar to Ishizaka and Flanagan [2], and Story and Titze [5], we describe the dashpot damping with the coefficients $R_i = 2\xi_i\sqrt{K_i m_i}$. The value of the damping coefficient is $\xi_{io} = 0.1$ in the open phase, as proposed by Ishizaka and Flanagan [2]. Story and Titze [5] choose a damping constant of $\xi_{io} = 0.4$, which strongly influences the model performance. In the closed phase, we have chosen $\xi_{ic} = 1.1$ as in Ishizaka and Flanagan's model [2]. This is slightly more than Story and Titze's $\xi_{ic} = 0.8$ [5]. It corresponds to a critical damping upon vocal-fold closure.

In the collision model, the critical height h_c is usually negative. Pelorson *et al.* [8] discuss the assumption of mechanical contact just before the actual vocal-fold contact, which allows some aspects of the vocal-fold oscillations to be predicted with a two-mass model. The assumption that $h_c > 0$ can be justified as a correction for three-dimensional effects in a two-dimensional vocal-fold representation. Another example of such a correction is the leaky-glottis model of Cranen [14], which certainly improves the quality of a number of model predictions. As in the models by Pelorson *et al.* [8] and Veldhuis *et al.* [10], we find the best results for $h_c > 0$, which is not in agreement with Ishizaka and Flanagan [2]. We choose $h_c = 0.02$ mm, as proposed by Pelorson *et al.* [8]. The value of h_c is critical: small variations lead to large changes in the glottal flow predicted by the model. The collision model is completed by increasing the spring constant $K = K_1 = K_2$ by a factor 4 if the glottis height becomes smaller than h_c. We follow Ishizaka and Flanagan [2] in the choice of this factor.

We would like to stress that, as observed by Pelorson *et al.* [8], the two-mass model results are very sensitive to the choice of a collision model. The collision description we present is basic, and the parameter choice is once more dictated by the need to compensate for model simplifications rather than by physiological measurements. The factor 4 by which the spring constant K is increased to model vocal-fold collision seems quite arbitrary, but realises a satisfactory behaviour at closure. The choice of a positive h_c may possibly be avoided by considering the three-dimensional vocal-fold structure in more detail. A systematic study of the vocal-fold collision by means of finite-element simulations might provide a less caricatural collision model.

3.4. Acoustic parameters

The mean air density ρ_0 equals 1.2 kg/m^3, and the speed of sound in air is set to $c_0 = 350$ m/s.

The stepwise approximations of the vocal-tract shapes are due to Fant [12]. Non-linear losses in the vocal tract, at the teeth and at the lips are ignored, although Davies *et al.* [22] show that they are significant for certain vowels. The effect of temperature gradients within the vocal tract is also neglected.

The exponential-horn model for the lungs realises a cut-off frequency of 500 Hz. The trachea has a length of $L_{\mathrm{trach}} = 20$ cm, and an area $A_{\mathrm{trach}} = 2.5\,\mathrm{cm}^2$. The horn is 10 cm long. Its stepwise approximation corresponds to measurements by Van den Berg, Rohrer and Weibel, as Ishizaka *et al.* [13] report. Again, non-linear losses and temperature gradients are neglected for the sake of simplicity.

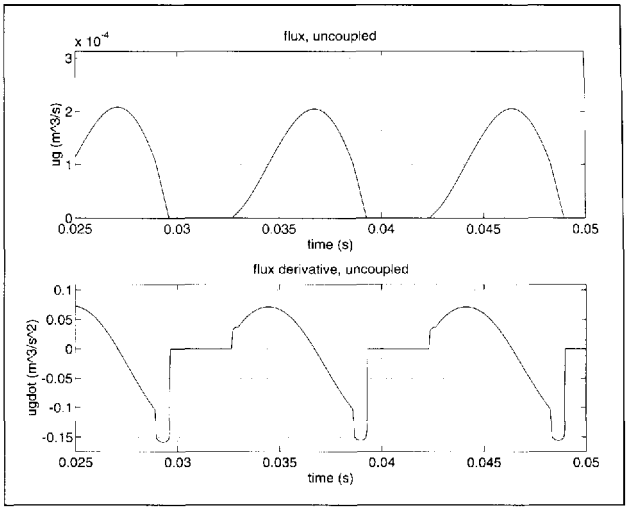

Figure 2. Glottal flux and flux derivative at 100 Hz, without coupling to vocal tract or trachea. The two-mass model yields reasonable results, although the flux derivative displays 'clicks'.

4. Acoustic feedback: observations and predictions

4.1. The influence of coupling the vocal-fold model to vocal tract, trachea, and lungs

Figure 2 displays the glottal flux and its time derivative predicted by our vocal-fold model, without coupling to vocal tract or trachea. The phonation frequency is 100 Hz.

In the uncoupled model, the vocal-fold opening is very sudden. This is a consequence of the fact that we ignore inertive and viscous terms in the flow description. These terms smoothen the glottal-flux pulses particularly when the glottis height is small (see Section 6.2), i.e. at opening and at closure of the vocal folds. The activation of the separation criterion causes another discontinuity in the derivative of U_g, to be called a 'click', just before the glottis actually closes. In general the separation criterion is inactive since $h_2/h_1 < s$, so the flow separates at x_2. Only during a short period just before closure the separation criterion becomes active, and the separation point x_s starts to move from x_2 towards x_1. The pressure difference between x_1 and x_s, used in Bernoulli's equation (2) to calculate the flux, thus suddenly diminishes more rapidly, inducing a rapid decrease of U_g. This sudden decrease of flux is visible as a 'click'. Notice that a smaller flux corresponds to smaller aerodynamic forces, and therefore eventually to a slowing-down of the mass movements. These mechanical effects are secondary, however, as a consequence of the inertia of the point masses. This also explains why the discrete changes in the spring constant K, inherent to our collision model, do not re-appear as clicks.

Finally, we observe a dramatic click at closure of the vocal folds. It is due to the absence of viscosity in the flow model, and to the fact that we ignore the contribution to the flux induced by glottal wall displacements. We refer to Section 6 for a more extensive discussion.

Figure 3 shows glottal-flux pulses and their derivative after extending the vocal-fold model with a vocal-tract description. The vocal-tract shapes correspond to the vowels /a/ and /i/ according to Fant [12]. We have chosen these vowels because of the availability of both measurements by means of inverse

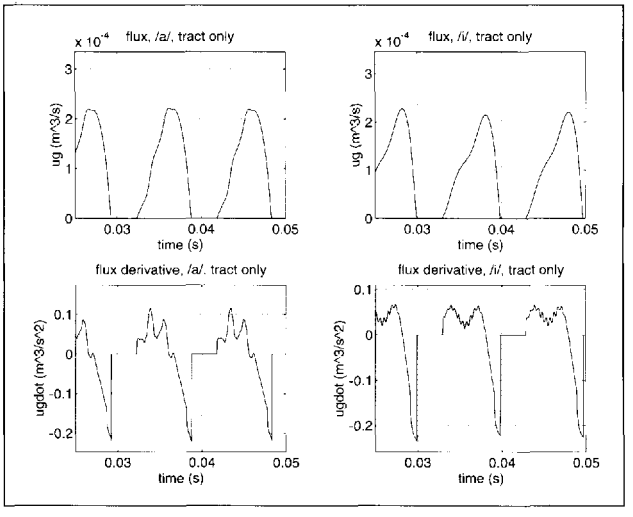

Figure 3. Glottal flux and flux derivative for /a/ and /i/ at 100 Hz, with coupling to vocal tract only. Acoustic feedback from the vocal tract causes formant ripples on the flux pulse and its derivative. The frequency of the larger ripples corresponds to the first formant frequency F_1; the smaller ones result from acoustic feedback via higher formants. The flux pulses become more asymmetrical. Constricted vowels (e.g. /i/) are more affected by acoustic feedback than unconstricted vowels (e.g. /a/).

filtering, and direct in-vivo measurements by Cranen [14]. The phonation frequency is 100 Hz.

Acoustic feedback from the vocal tract causes 'formant ripples' on the glottal-flux pulses and their derivative. The larger ripples occur at a frequency that corresponds to the first formant frequency F_1. The smaller ones, mainly visible on the pulse derivative of the /i/, result from acoustic feedback via higher formant frequencies. In particular at these higher frequencies, the assumed quasi-stationarity in the flow model is likely to cause an over-estimation of the acoustic feedback. Acoustic feedback also accentuates the asymmetry of the glottal-pulse shape, as observed by Rothenberg [15].

ACUSTICA · acta acustica
Vol. 84 (1998)

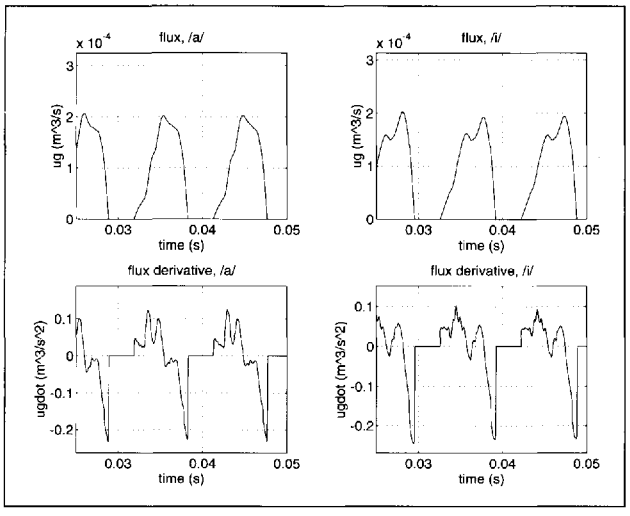

Figure 4. Glottal flux and flux derivative for /a/ and /i/ at 100 Hz, with coupling to vocal tract and trachea. Coupling to the trachea causes additional formant ripples on the flux pulse and its derivative, and increases the effect of acoustic feedback. Constricted vowels (e.g. /i/) are more affected by acoustic feedback than unconstricted vowels (e.g. /a/).

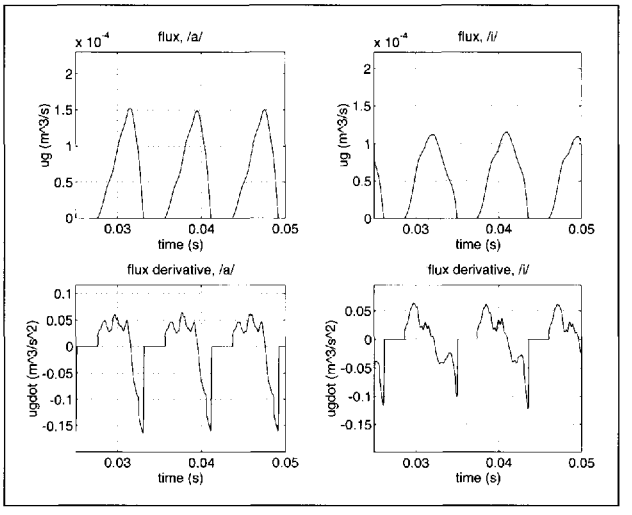

Figure 5. Glottal flux and flux derivative for /a/ and /i/ at 120 Hz, with coupling to vocal tract and trachea. The acoustic feedback is less pronounced if the phonation frequency F_0 does not correspond to any resonance frequency of vocal tract of trachea.

Figure 4 displays predictions obtained after extending our model to include both vocal tract and sub-glottal system. We used the same vocal-tract shapes as before, and a phonation frequency of 100 Hz.

Formant ripples and glottal-pulse shape asymmetry become more pronounced after coupling vocal folds and trachea, particularly for the /i/. The vocal-tract shape corresponding to the /i/ has a smaller lip opening than the /a/, and correspondingly a lower first formant F_1. As a consequence, the first (and dominant) formant of an /i/ radiates its energy less efficiently than the first formant of an /a/ (see equation (6)). More energy is therefore reflected back into the vocal tract when pronouncing an /i/. This explains the larger sensitivity of the /i/ to acoustical feedback, which our model especially predicts at higher phonation frequencies. In reality, this effect may be less pronounced than the two-mass model predicts, as the constriction at the lips for the /i/ may induce significant convective dissipation by flow separation as discussed by Davies *et al.* [22].

The acoustics in the vocal tract affect the pressure p_3 at x_3. This directly influences U_g via Bernoulli's equation. As p_3 appears in the expressions for F_{h1} and F_{h2} as well, acoustic feedback also causes changes in the mass movements. If we set $p_0 = p_{\text{phon}}$ and $p_3 = 0$ in the mechanical model only, and leave the aerodynamic model unchanged, the vocal-tract acoustics do not affect the mass movements directly. This allows an estimation of the influence of acoustic feedback via Bernoulli's equation, used to calculate the glottal flux U_g. Similarly, if we set $p_0 = p_{\text{phon}}$ and $p_3 = 0$ in Bernoulli's equation, not altering the mechanical model, the consequences of acoustic feedback for the movement of m_1 and m_2 can be estimated. This procedure reveals that the formant ripples are mainly due to feedback via Bernoulli's equation. Feedback via the aerodynamic forces accounts for the accentuation of the glottal-flux pulse asymmetry after coupling.

Furthermore, the amount of acoustic feedback is related to the ratio of phonation frequency F_0 and first formant frequency F_1. For the /a/, $F_1 \approx 500$ Hz (see Figure 9, upper-left graph). At $F_0 = 100$ Hz, the fifth harmonic in the spectrum of the glottal flux therefore practically coincides with the first formant frequency of the /a/: $5F_0 = 500 \approx F_1$. This harmonic coincidence causes a large interaction between vocal folds and vocal tract. At a phonation frequency of 120 Hz, the first formant frequency F_1 of the /a/ lies between the fourth and fifth harmonics in the glottal flux spectrum, where the fourth harmonic is nearest to F_1 at 480 Hz. There is no (or certainly less) harmonic coincidence in this case. Comparision of the leftmost graphs in Figures 4 and 5 shows that for the /a/, acoustic feedback is indeed less pronounced at a phonation frequency of 120 Hz. As a comparision, consider the /i/, with a first formant frequency $F_1 \approx 175$ Hz (see Figure 9, upper-right graph). The harmonic of the glottal flux spectrum nearest to F_1 is the second (at 200 Hz) for a phonation frequency of $F_0 = 100$ Hz, and the first (at 120 Hz) for a phonation frequency of $F_0 = 120$ Hz. This discrepancy between F_1 and the nearest glottal flux spectrum harmonic is 25 Hz for $F_0 = 100$ Hz, and 55 Hz for $F_0 = 120$ Hz. Although harmonic coincidence does not actually occur for either of these phonation frequencies, the situation where $F_0 = 100$ Hz is closest to coincidence. Comparision of the rightmost graphs in Figures 4 and 5 confirms that for the /i/, acoustic feedback is most pronounced at a phonation frequency of 100 Hz.

A particularly strong acoustic feedback is to be expected at phonation frequency $F_0 = F_1$, as discussed by Hertegard and Gauffin [23], Lucero [4], and Schutte and Miller [24]. Their observation involves a falsetto voice, which is not predicted by our model in the current configuration; comparison to model results is therefore not possible.

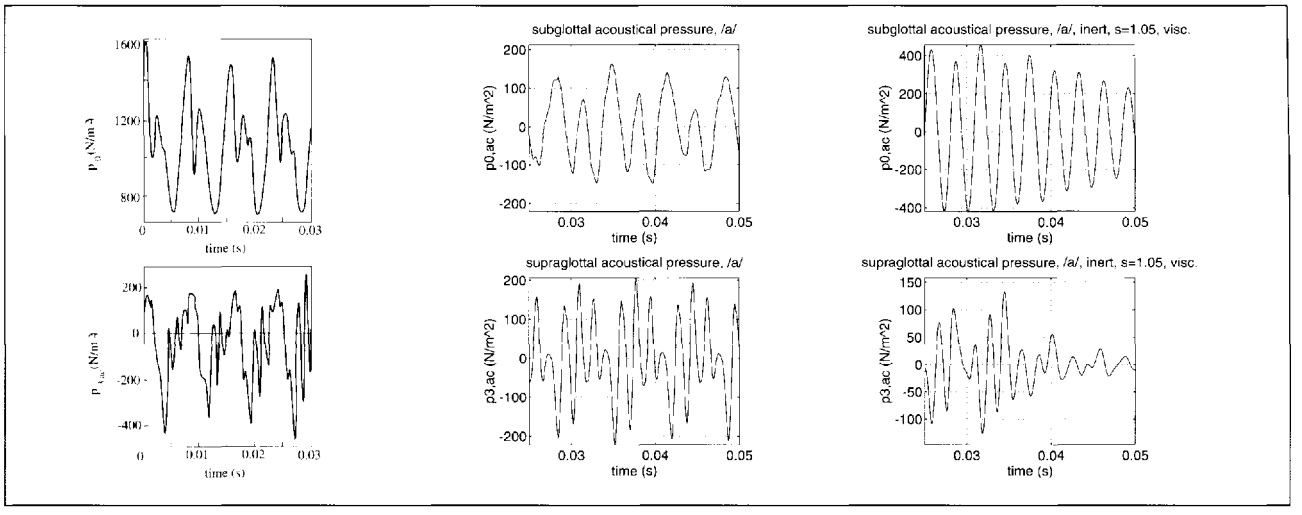

Figure 6. Sub- and supra-glottal pressure for /a/ at 145 Hz, with coupling to vocal tract and trachea. The measurements by Cranen (left) are well predicted by the model (middle). The discrepancy between measured and predicted sub-glottal acoustical pressure amplitude is of the order of magnitude of differences observed between speakers. The model predictions deteriorate upon including approximated inertive and viscous terms in the flow model, and reducing the separation constant to $s = 1.05$ (right).

The two-mass model performs reasonably well in the frequency domain, too. The sound spectra resulting at the lips in the completely coupled model compare well to spectra of /a/ and /i/ presented in literature. However, the details of these frequency domain results may not correspond to measurements. The spectra include the effect of simplifications in the model, like the 'clicks' in the glottal pulse derivative shown in Figure 2. In particular, we have already mentioned the crucial role of the collision in the actual sound production by the vocal folds. The collision description in our two-mass model is rather elementary; the same may be expected of its contribution to realistic vowel sound production.

4.2. Comparison to sub- and supra-glottal pressure measurements

Cranen [14] reports in-vivo measurements of sub- and supra-glottal pressures. His experimental results [14, p.43, speaker BC] are shown in the graphs on the left in Figure 6. The graphs in the middle display the pressures p_0 and p_3 predicted by our model. The vocal-tract shape corresponds to an /a/, and the phonation frequency is 145 Hz as in [14].

Both p_0 and p_3 are well predicted by our model, although the amplitude of the sub-glottal acoustic pressure p_0 is under-estimated by a factor 4. However, the reference speaker BC phonates at a relatively high frequency of 145 Hz. When we compare the predicted sub-glottal acoustic pressure amplitude to measured values at about 100 Hz, realised by another speaker [14, p.43, speaker LB], the under-estimation of p_0 practically vanishes. This indicates that the sub-glottal pressure variations are strongly speaker-dependent, and that they may vary significantly with frequency.

The model results also compare well to measurements by Koike [25].

4.3. Comparison to glottal-flux pulse measurements in air and in helium

Rothenberg [15] describes experiments comparing the effect of acoustic feedback in phonations made in air, and in air with a large proportion of helium. The leftmost graph of Figure 7 shows his results. The middle graph displays the (normalised) glottal pulse resulting from the model with $\rho = \rho_{He} \approx (4/29)\rho_0$, and $c = c_{He} \approx c_0\sqrt{29/4}$, compared to a similar pulse in air. We use the vocal-tract shape corresponding to an /a/, and a phonation frequency of 110 Hz as reported by Rothenberg in [15].

The two-mass model clearly reproduces the glottal-flux pulse smoothing in helium observed by Rothenberg [15]. As the speed of sound is much larger in helium than it is in air, the first formant frequency of the /a/ shifts from about 500 Hz in air to over 1000 Hz in helium. If we couple the two-mass vocal-fold model to the vocal-tract transmission-line model for an /a/, and use the acoustic pressure at the lips thus predicted to determine sound spectra in air and in helium, the formant frequency rise is clearly reproduced. In helium, F_1 is far more distant from the fundamental vocal-fold oscillation frequency F_0 than in air. This drastically decreases the influence of acoustic feedback, and explains the glottal-flux pulse smoothing.

Rothenberg [15] reports an increased glottal-flux pulse symmetry in helium, and a slower vocal-fold closure than in air. The larger pulse symmetry in helium is caused by a reduction of the influence of inertia in the vocal tract. Similarly, Rothenberg [15] mainly attributes the difference in closure behaviour in air and helium to inertive effects. However, inertia has not been included in our flow model so that these effects are not well predicted. In Section 6.2, we propose crude approximations for both inertive and viscous terms in the flow description, and we discuss their influence on the model predictions.

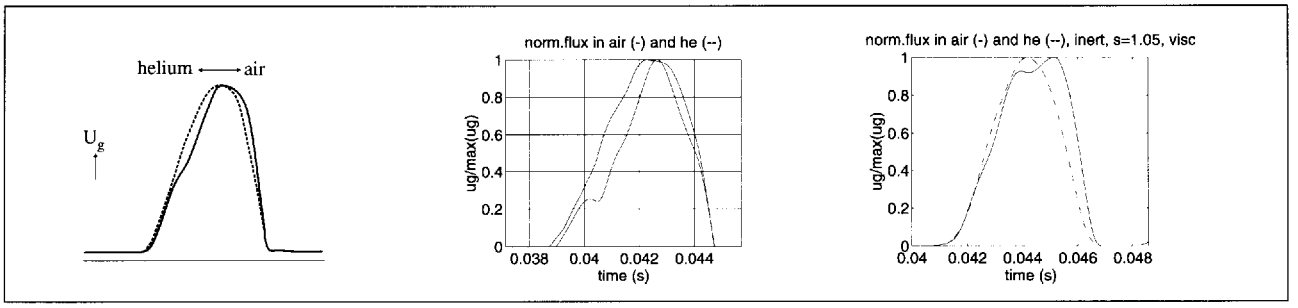

Figure 7. Glottal flux for /a/ at 110 Hz, in air and in helium, according to Rothenberg (left), and according to the model (middle). In helium, the flux pulse becomes smoother and more symmetrical. These results are globally well predicted by the two-mass model. The model does not predict the more sudden closure in helium, unless approximated inertive and viscous terms are included in the flow model with $s = 1.05$ (right).

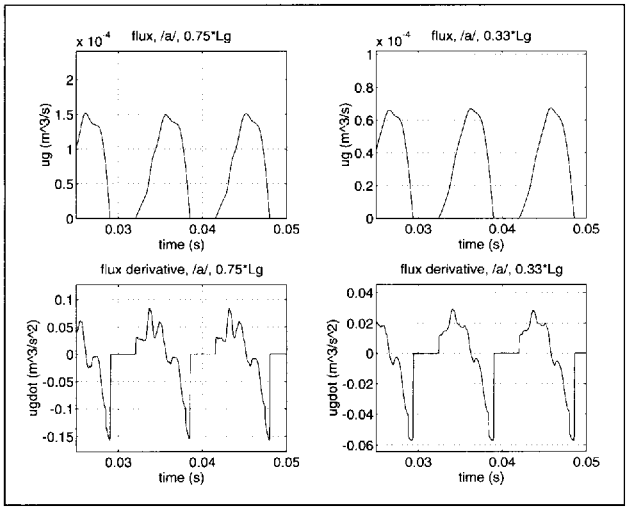

Figure 8. Glottal flux and flux derivative for /a/ at 100 Hz, with coupling to vocal tract and trachea. Vocal-fold length L_g and mechanical model scaled by a factor β, for $\beta = 0.75$ and $\beta = 0.33$. A vocal-fold prosthesis scaled down in this way is less affected by acoustic feedback, but also yields a smaller glottal flux, and a reduced sound pressure at the lips.

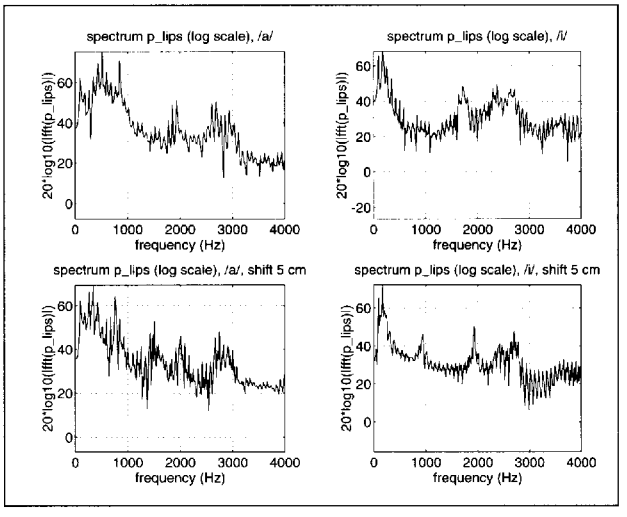

Figure 9. Spectrum of the sound pressure radiated at the lips for /a/ and /i/ at 100 Hz, with coupling to vocal tract and trachea (upper graphs). If the vocal-fold model is shifted 5 cm down into trachea, both /a/ and /i/ start to sound like an /oe/ (lower graphs). Implanting a prosthesis elsewhere than at the original vocal-fold location will complicate articulation.

5. Application of the model to vocal-fold prosthesis design

5.1. Influence of prosthesis size reduction

A vocal-fold prosthesis will, for surgical reasons, have to be about three times smaller than the actual vocal folds. We applied the model to investigate the effects of scaling down the (prosthetic) glottis.

In Figure 8 we display results for the vowel /a/ at a phonation frequency of 100 Hz. The vocal-fold length is reduced to βL_g, and its mass is decreased by the same factor. To keep the relevant vibration modes $\omega_1 = K/m$, and $\omega_2 = (K + K_{12})/m$ fixed, we also have to scale down the spring constants K and K_{12} by a factor β. This scaling procedure does not alter the fundamental vocal-fold oscillation frequency F_0, as we have chosen $F_0 = \omega_1/(2\pi) = 100$ Hz.

Upon scaling the vocal-fold length L_g to βL_g, the glottal flux (maximum) is reduced by a factor β. Reduction of the glottal flux implies a reduction of acoustical pressure fluctuations in the vocal tract and the trachea. This implies a reduction of a acoustical feedback on the glottal-flux pulse. Scaling down L_g indeed results in a less pronounced acoustic feedback, and in smoother glottal-flux pulses and glottal-flux pulse derivatives. Decreasing the vocal-fold length also affects the spectrum of the sound radiated at the lips. Apart from a general reduction of the radiated power, the second formant at frequency F_2 becomes relatively less pronounced. Notice that for $\beta = 0.75$, the scaled vocal-fold length is nearly equal to the 1 cm proposed by Story and Titze (5). A scaling factor $\beta = 0.33$ reduces the vocal-fold length to 0.5 cm, which would be a reasonable size for a prosthesis from a surgical point of view. However, for $\beta = 0.33$ the power of the radiated sound decreases significantly. The bearer of a prosthesis of this size might not be able to speak up.

When we scale down the prosthesis geometry by a factor β in all three dimensions, masses and spring constants have to be reduced by a factor β^3 to keep F_0, ω_1, and ω_2 unchanged. Three-dimensional scaling with factors $\beta = 0.75$ and $\beta =$

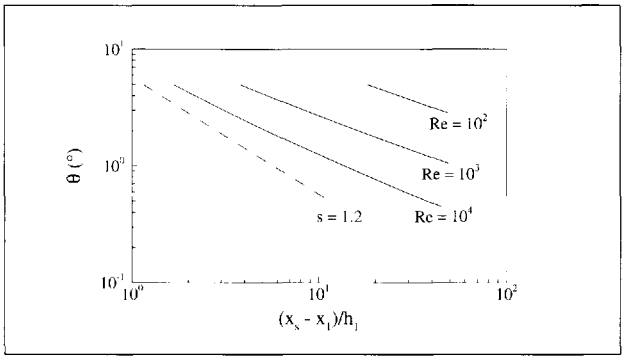

Figure 10. The angle of divergence θ versus the (normalised) separation point $(x_s - x_1)/h_1$ for three values of the Reynolds number. The results correspond to theoretical calculations concerning a stationary, diverging geometry (diffusor), as in the vocal-fold closing phase. Two-mass model predictions correspond to $Re < 1000$. The line marked $s = 1.2$ respresents the separation criterion used in our model. The separation criterion does not correspond to a physical description of flow separation.

0.33 appears to have no effect on the predicted glottal-flux pulses and their derivatives for the /a/ at 100 Hz, apart from an increase of the mass movement amplitudes. However, it may be very difficult to find a material that satisfies both the conditions of strength and large compliance corresponding to a scaling factor $\beta = 0.33$ in particular. Furthermore, such a compliant prosthesis is not a barrier for food, and it is likely to be most sensitive to dirt.

5.2. Influence of prosthesis implantation location

After surgery, it may be necessary to implant a vocal-fold prosthesis slightly below the original location of the vocal folds. The spectral effect of a large displacement of 5 cm on the glottal-flux pulse and its derivative is shown in Figure 9. The phonation frequency is 100 Hz.

The difference between the radiated sound spectra of /a/ and /i/ decreases after shifting the vocal-fold prosthesis 5 cm downward in the trachea. Both vowels become more similar to the /oe/, the vocal-tract shape of which roughly corresponds to a straight pipe of constant diameter. This indicates that the more the prosthesis position deviates from the original vocal-fold location, the more its bearer will have to alter his articulation in order to pronounce understandable vowels.

6. Critical discussion of the flow model

6.1. Physical relevance of the separation criterion

To evaluate the choice of the separation constant s, it is useful to study in more detail the actual physical separation behaviour. We consider a stationary, linear, diverging channel, to be thought of as the glottis geometry predicted by our two-mass model at one instant during closure of the vocal folds. In this stationary channel, we assume a quasi-steady flow, so that the physical separation point is only a function of the geometry and the Reynolds number. The Reynolds number Re is a non-dimensional flow parameter, that gives an indication of the importance of viscosity in a flow. In our case, it is defined as the two-dimensional flux U_g/L_g times the ratio of density ρ_0, and dynamic viscosity η, i.e. $Re = U_g\rho_0/L_g\eta$. For air at room temperature we have $\eta \approx 1.8 \cdot 10^{-5}$ Ns/m^2. The stationary geometry is defined by the initial height h_1 at the beginning of the diverging part of the glottis, and the angle θ between the diverging flow channel walls. This angle of divergence is given by $\theta = \arctan[(y_2 - y_1)/(x_2 - x_1)]$ (see Figure 1). We use the method of Polhausen as described by Pelorson *et al.* [8] to predict flow separation in this geometry. The velocity profile in the viscous boundary layer has been approximated by a third-order polynomial. Flow separation occurs when the direction of the flow near the wall becomes opposite to the direction of the main flow.

Figure 10 displays the relationship between the angle of divergence and the (normalised) separation point $(x_s - x_1)/h_1$ on a logarithmic scale. Three values of the Reynolds number have been considered, corresponding to the typical range of Re relevant for speech. In our vocal-fold geometry, and for the flux our model predicts, we have $Re \leq 1000$. We also plotted the angle of divergence versus the (normalised) separation point for the separation condition $s = 1.2$ proposed by Liljencrants [7].

Figure 10 clearly shows that the choice $s = 1.2$ does not represent an accurate physical description of flow separation. The separation constant appears to be yet another parameter that has to compensate for over-simplifications in our model. The particular choice $s = 1.2$ partly makes up for the neglected inertive term in the flow model. If inertia is included in the flow description, the value of s should be decreased to 1.05 or 1.1 to avoid over-compensation. This would imply an even lower value of θ for a given (normalised) separation point $(x_s - x_1)/h_1$. For values of s smaller than $s = 1.2$, the discrepancy between the predicted flow separation according to Liljencrants [7] and the theoretical results from the Polhausen method become more pronounced.

The larger we choose the separation constant to be, the more sudden is the closure of the vocal folds. Since the main contribution to the aerodynamic force was seen to result from the pressure in the glottis between x_0 and x_s, the separation point largely influences the force acting on the vocal folds. The smaller $x_s - x_1$ is, the smaller the aerodynamic forces are. The flow-separation model causes the separation point to start moving from x_2 toward x_1 as soon as $h_2/h_1 = s$ in the vocal-fold closing phase. The aerodynamic forces then decrease, and the glottis closure slows down. If s is relatively large, this slowing down takes place relatively late, so that the vocal-fold closure as a whole speeds up. Fixing the separation point at the exit of the glottis (corresponding to a very large value of s) yields an unrealistic and extremely sudden closing behaviour, as already shown by Pelorson *et al.* [8]. Small values of s correspond to a more relaxed vocal-fold closure.

With the parameter choice discussed in Section 3, the separation criterion appears to induce a movement of x_s only just before the glottis closes. As its physical relevance has been shown to be questionable, the separation criterion in our model is in fact not more than a way to soften the 'clicks' in the glottal flux and aerodynamic force upon closure.

ACUSTICA · acta acustica
Vol. 84 (1998)

6.2. Flow inertia and viscosity

6.2.1. Approximation of inertive and viscous terms

It is not easy to extend the flow model to include an inertive term. The present model assumes the glottal flux to be independent of x within the glottis, which corresponds to the equation of continuity $\partial U_g(x,t)/\partial x = 0$. Modelling inertia would require this mass conservation equation to be changed to $\partial U_g(x,t)/\partial x = -L_g \partial h(x,t)/\partial t$ to take flow channel wall movements into account. Furthermore, an extra inertive term $\rho \partial\phi/\partial t$ would have to be added in the right-hand side of Bernoulli's equation (2). Here, $\phi = \int_{x_0}^{x} v(x,t)\mathrm{d}x$, in which $v(x,t)$ denotes the x-component of the particle linear velocity. This would no longer allow explicit calculation of the aerodynamic forces, as presented in Appendix A1.

To avoid this complication of the model, and still roughly estimate the order of magnitude of the influence of inertia, we focus on the effect of the unsteady term in Bernoulli's equation. We neglect the change in mass-conservation equation, and assume the glottal flux to be independent of x. The inertive term in Bernoulli's equation can then be approximated as

$$\begin{aligned}\Delta p_{inert} &= \frac{\partial}{\partial t}\left(\rho \int_{x_0}^{x_2} v(x,t)dx\right) \\ &= \frac{\rho_0}{L_g}\frac{\partial}{\partial t}\left(U_g(t)\int_{x_0}^{x_2}\frac{1}{h(x,t)}\mathrm{d}x\right) \\ &\approx \frac{\rho_0}{L_g}\frac{\mathrm{d}U_g}{\mathrm{d}t}\frac{x_2 - x_1}{\min(h_1,h_2)}.\end{aligned} \tag{9}$$

We used incompressibility, and ignored the wall displacement term in the derivative as proposed by Ishizaka and Flanagan [2]. To roughly estimate the effect of inertia on the glottal flow, we add the expression for $\Delta p_{\mathrm{inert}}$ to the left-hand side of the discretised Bernoulli equation (A15) for calculation of the flux U_g. The inertive flow correction is also taken into account when determining the aerodynamic forces. In our simple two-mass model, we slightly over-estimate the value of s to compensate for the neglected inertive term. After inclusion of such a term in the flow model, s should be decreased.

Apart from inertive effects, we neglected viscosity in the description of the glottal flow. Again, it is hard to model viscosity correctly. A rough estimate of its influence on the glottal flow is found by extending Bernoulli's equation with a viscous pressure loss term

$$\begin{aligned}\Delta p_{\mathrm{visc}} &= \frac{12\eta U_g}{L_g}\int_{x_0}^{x_2}\frac{1}{h(x,t)^3}\mathrm{d}x \\ &\approx \frac{12\eta U_g}{L_g}\frac{x_2 - x_1}{\min(h_1,h_2)^3}.\end{aligned} \tag{10}$$

The term Δp_{visc} corresponds to the pressure loss due to viscosity in a channel of small height h in which the flow has a fully developed Poiseuille velocity profile. We add the Poiseuille-term (10) to the left-hand side of the discretised Bernoulli equation (A15) to determine the glottal flux and the aerodynamic forces.

In the expressions for $\Delta p_{\mathrm{inert}}$ and Δp_{visc}, the diverging or converging shape of the glottis is approximated by a channel of constant height $\min(h_1, h_2)$. The flow channel portion between x_0 and x_1 is ignored when approximating the integrals in the inertive and viscous expressions (9) and (10). A mathematically more precise approach, in which we used

$$\Delta p_{\mathrm{inert}} = \frac{\rho_0}{L_g}\frac{\mathrm{d}U_g}{\mathrm{d}t}\int_{x_0}^{x}\frac{1}{h(x,t)}\mathrm{d}x,$$

and

$$\Delta p_{visc} = \frac{12\eta U_g}{L_g}\int_{x_0}^{x}\frac{1}{h(x,t)^3}\mathrm{d}x, \tag{11}$$

did not change the model predictions significantly.

Numerical results obtained after inclusion of an inertive term in the flow model are displayed in the leftmost graphs of Figure 11. The middle graphs display the model predictions if inertia is taken into account, but with a reduced separation constant $s = 1.05$ instead of $s = 1.2$. The glottal-flux pulses and their derivative, resulting after modelling viscosity as well, are shown in the rightmost graphs. Model predictions have been made for an /a/ at 100 Hz in a model with acoustic *feedback from vocal tract and sub-glottal system.*

Equations (9) and (10) show that the (approximated) inertive and viscous terms particularly influence the flow when the glottis height is small, i.e. at opening and at closure of the vocal folds. Both inertia and viscosity tend to slow down the opening of the vocal folds, as they reduce the mean flow velocity in the x-direction. The glottal-flux pulses are smoothed. Inertia of the air causes the vocal-fold closure to be more sudden, while viscous effects smoothen the resulting 'peak' in the glottal flow derivative. These observations are clearly reflected by the model predictions. The inertive effects become less pronounced after decreasing the separation constant s.

The rightmost graphs of Figure 6 show that the model predictions for an /a/ at 145 Hz are strongly affected when inertive and viscous term approximations are included in the flow model. The separation constant is set to $s = 1.05$. The sub- and supra-glottal pressure amplitudes change significantly. The effect of inertia and viscosity seems to be over-estimated at this relatively high phonation frequency, and causes the vocal-fold oscillations to be damped. In comparison to the measurement results by Cranen [14], it appears to be counter-productive to use our 'more detailed' flow model.

After taking inertia and viscosity into account, the model does predict the slower vocal-fold closure in helium observed by Rothenberg [15]. This is displayed in the rightmost graphs of Figure 7 for an /a/ at 110 Hz. The separation constant is $s = 1.05$ instead of $s = 1.2$ to avoid over-estimation of inertive effects.

6.2.2. Physical relevance of inertive and viscous terms

We wish to emphasise that the inertive and viscous terms used in our model are not more than rough approximations, so that conclusions are to be considered with care. The results are

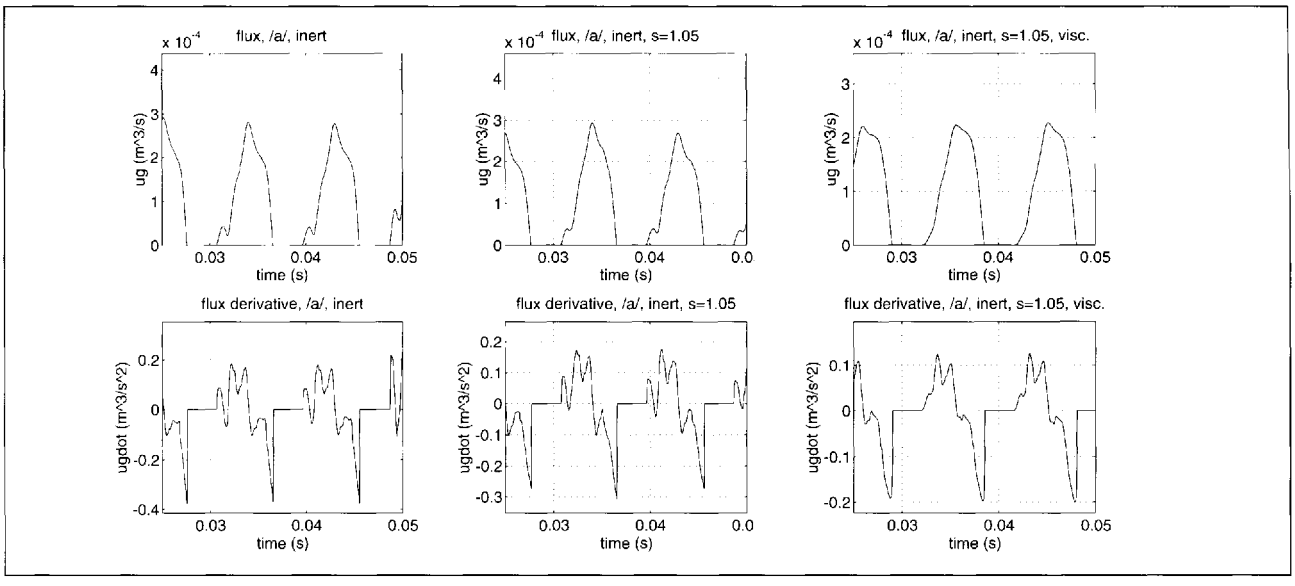

Figure 11. Glottal flux and flux derivative for /a/ and /i/ at 100 Hz, with coupling to vocal tract and trachea. The flow model includes an inertive term (left), an inertive term with separation constant $s = 1.05$ instead of $s = 1.2$ to avoid over-estimation of inertive effects (middle), and inertive and viscous terms with $s = 1.05$ (right). Inertia slows down the vocal-fold opening, whereas the glottis closes more rapidly. Decreasing the separation constant is equivalent to a reduction of inertive effects. Viscosity smoothens the flux pulses and their derivatives at opening and at closure.

to be interpreted as crude order-of-magnitude estimations. Any detailed description of inertia is only consistent if the whole vocal-fold model – including geometry, mechanics and acoustics – is of similar precision. Inertive or viscous term approximations used in combination with the basic geometrical, mechanical, and acoustical vocal-fold model as presented above are no more than elegant ways of fitting model results to experimental or measurement data (see Figure 7). We have seen that adding approximated terms to model inertia and viscosity may also have negative effects on the model predictions (see Figure 6).

As an example of the difficulties occurring in an 'exact' description of inertive losses, we consider the equation of mass conservation in the form $\partial U_g(x,t)/\partial x = -L_g \partial h(x,t)/\partial t$. A first problem is that we cannot combine this equation with a quasi-stationary flow-separation description. Second, the non-stationary mass conservation equation introduces a volume injection due to flow channel wall movements when applied to our mechanical model. As the vocal folds have a small compliance, we expect such a volume injection at the glottis to be compensated by wall movements nearby, i.e. we expect the integral of $\partial h(x,t)/\partial t$ over a fold to be close to zero. This cannot be properly predicted without a detailed mechanical model including the elastic behaviour of the entire throat. However, ignoring the wall displacement in the glottis is certainly a poor approximation at closure of the vocal folds. We expect most of the aerodynamic and mechanical details relevant for sound production to occur at that moment. It is therefore unlikely that accurate vowel sounds could be produced on the basis of the simplified mechanical description used in our model.

If the two-mass model is to be used for actual vowel synthesis, it may be practical to separate the prediction of vocal-fold oscillations from the prediction of the sound production. A two-mass model as simple as the one we presented may be used to give reasonable predictions of the global mass-movements, and of the corresponding glottal-flux pulses. Once the global mass-movement is predicted by means of this model, an additional module may be added in which geometry, aerodynamics and mechanics are modelled in more detail at closure of the vocal folds.

7. Conclusions

We presented a simple two-mass model for the vocal folds, based on earlier models by Ishizaka and Flanagan [2], Pelorson *et al.* [8], and Veldhuis *et al.* [10]. In contrast to Pelorson *et al.* [8], we do not assume the complete aerodynamic force to act on one mass only. We propose a distribution of the aerodynamic force over both masses. This distribution appears to have a large influence on the global vocal-fold movement, far more than adding possible approximations of inertive and viscous terms to the flow description. A consequence of the new force distribution appears to be, that symmetry can be assumed in both geometrical and mechanical models. In our model, the vocal-fold inlet and outlet depths are equal, and the two mass-spring systems are identical.

The choice of model parameters has been shown to be very difficult. Parameter values can partially be based on physiological measurement, or on theoretical calculations, but they also have to make up for deficiencies or simplifications in the model. While a smoother geometry or a three-dimensional mechanical model would ease the choice of certain parameters, they would at the same time introduce new unknowns.

ACUSTICA · acta acustica
Vol. 84 (1998)

Nevertheless, our two-mass model yields reasonable predictions of glottal-flux pulses and sub- and supra-glottal pressures in the presence of acoustical feedback from trachea and vocal tract. This allows application of the model to vocal-fold prosthesis design – in which case the geometrical symmetry may be seen as an advantage. We explored the effect of size and position of such a prosthesis. Decreasing the prosthesis length only, without changing its mechanical properties, reduces the radiated power at the lips. The acoustical feedback on the glottal-flux pulses becomes less pronounced. Decreasing the prosthesis size in all three dimensions does not essentially affect the predicted glottal-flux pulses, or the radiated power. However, such scaling requires the prosthesis material to be extremely compliant. A very compliant prosthesis does not form a barrier for food transport toward the lungs, and is likely to be very sensitive to dirt.

Extending the flow description with crudely approximated inertive and viscous terms does influence the model results. However, a more complex flow description will not really improve the model prediction unless geometry and mechanical model are described with the same level of complexity. Furthermore, a detailed flow model would call for a mechanical model that takes into account the three-dimensional vocal-fold movement observed under certain circumstances. An improved flow model applied to a simplified geometry may even have a negative influence on the overall model performance, because it does not provide room for compensation of errors induced by the basic mechanical model.

Acknowledgement

The research leading to this paper has been financed in part by grants from the Dutch Technology Foundation (STW), project numbers GGN55.3712 and ETN33.2952.

Appendix

A1. Explicit expressions for $F_{h,1}$ and $F_{h,2}$

We consider the aerodynamic force acting on the flow channel wall between x_{i-1} and x_i, for $i = 1, 2, 3$. As time dependencies are not relevant for the following calculations, they will be omitted. The pressure $p(x)$ leads to reaction forces perpendicular to the wall. As the masses are allowed to move in the y-direction only, we are interested in the corresponding component of the aerodynamic force. This can be found as

$$\int_{x_{i-1}}^{x_i} p(x)\mathrm{d}x \tag{A1}$$

Notice that we integrate along the x-axis ($\mathrm{d}x$), and not along the flow channel wall, to obtain the y-component of the aerodynamic force only. As the wall is assumed to be rigid and mass-less, this force leads to reaction forces at points $i = 0, 1, 2, 3$ parallel to the y-axis, marked with an index hi in Figure A1. The superscript l indicates that a reaction force results from the pressure distribution in the flow channel to the left of point i, between x_{i-1} and x_i. If a reaction force follows from $p(x)$ in the glottis to the right of point i, between x_i and x_{i+1}, it is marked with a superscript r.

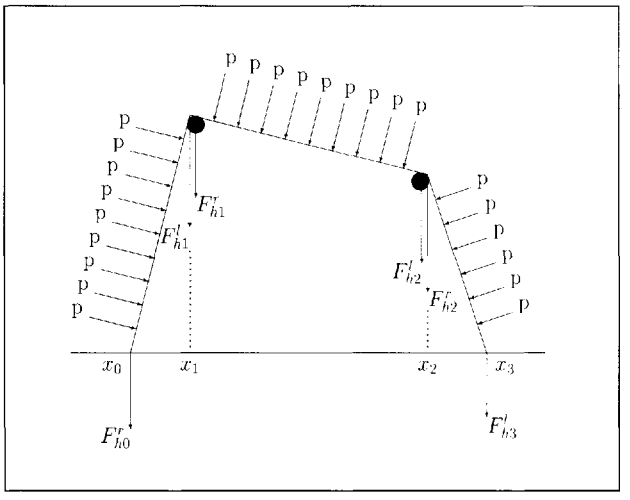

Figure A1. Aerodynamic reaction forces resulting from a pressure distribution $p(x)$ in the glottis. The force components F^r_{h1} at x_1, and F^l_{h2} at x_2, result from the pressure distribution on the mass-less plate connecting masses m_1 and m_2. Similar force components can be determined resulting from the aerodynamic pressure on the plates modelling the glottis inlet (F^r_{h0} and F^l_{h1}) and outlet (F^r_{h2} and F^l_{h3}).

Force balance with respect to the wall section between points $i-1$ and i, and torque balance at $i-1$ yield

$$F^r_{h,i-1} + F^l_{hi} = \int_{x_{i-1}}^{x_i} p(x)\mathrm{d}x \tag{A2}$$

$$(x_i - x_{i-1})F^l_{hi} = \int_{x_{i-1}}^{x_i} (x - x_{i-1})p(x)\mathrm{d}x, \tag{A3}$$

from which

$$F^r_{h,i-1} = \int_{x_{i-1}}^{x_i} \frac{x_i - x}{x_i - x_{i-1}} p(x)\mathrm{d}x,$$

and

$$F^l_{hi} = \int_{x_{i-1}}^{x_i} \frac{x - x_{i-1}}{x_i - x_{i-1}} p(x)\mathrm{d}x, \tag{A4}$$

for $i = 1, 2, 3$. As the glottis inlet at x_0, and the outlet at x_3, are fixed, the aerodynamic force components F^r_{h0} and F^l_{h3} are not relevant. For $i = 1, 2$, the total aerodynamic force acting on x_i equals $F_{hi} = F^l_{hi} + F^r_{hi}$, as shown in Figure A1. This corresponds to equation (3).

From Bernoulli's equation (2),

$$p(x) = p_0 + \frac{\rho_0}{2}\left(\frac{U_g}{L_g}\right)^2 \left(\frac{1}{h_0^2} - \frac{1}{h(x)^2}\right)$$
$$=: P_0 - \frac{P_1}{h(x)^2} \tag{A5}$$

between x_0 and x_s, while $p(x) = p_3$ for $x > x_s$. Hence

$$\int_{x_0}^{x_1} p(x)\mathrm{d}x = P_0(x_1 - x_0) - P_1 \int_{x_0}^{x_1} \frac{1}{h(x)^2}\mathrm{d}x \tag{A6}$$

$$\int_{x_1}^{x_2} p(x)\mathrm{d}x = P_0(x_s - x_1) - P_1 \int_{x_1}^{x_s} \frac{1}{h(x)^2}\mathrm{d}x + p_3(x_2 - x_s) \quad \text{(A7)}$$

$$\int_{x_0}^{x_1} xp(x)\mathrm{d}x = \frac{P_0}{2}(x_1^2 - x_0^2) - P_1 \int_{x_0}^{x_1} \frac{x}{h(x)^2}\mathrm{d}x \quad \text{(A8)}$$

$$\int_{x_1}^{x_2} xp(x)\mathrm{d}x = \frac{P_0}{2}(x_s^2 - x_1^2) - P_1 \int_{x_1}^{x_s} \frac{x}{h(x)^2}\mathrm{d}x + \frac{p_3}{2}(x_2^2 - x_s^2). \quad \text{(A9)}$$

With these expressions, F_{h1} and F_{h2} can be determined; the only missing links left are the explicit formulas for $\int_{x_{i-1}}^{x_i} 1/h(x)^2\mathrm{d}x$ and $\int_{x_{i-1}}^{x_i} x/h(x)^2\mathrm{d}x$, $i = 1, 2$.

We rewrite equation (1) to $h(x) = H_1 x + H_2$, with $H_1 = (h_i - h_{i-1})/(x_i - x_{i-1})$, and $H_2 = -H_1 x_{i-1} + h_{i-1} = (x_i h_{i-1} - x_{i-1} h_i)/(x_i - x_{i-1})$. The following results can then be derived:

$$\int_{x_{i-1}}^{x_i} \frac{1}{h(x)}\mathrm{d}x = \frac{1}{H_1}\int_{x_{i-1}}^{x_i} \frac{1}{h(x)}\mathrm{d}h(x) = \frac{1}{H_1}\ln(h(x))\Big]_{x=x_{i-1}}^{x_i} = \frac{1}{H_1}\ln\left(\frac{h_i}{h_{i-1}}\right), \quad \text{(A10)}$$

$$\int_{x_{i-1}}^{x_i} \frac{1}{h(x)^2}\mathrm{d}x = \frac{1}{H_1}\int_{x_{i-1}}^{x_i} \frac{1}{h(x)^2}\mathrm{d}h(x) = \frac{-1}{H_1}\frac{1}{h(x)}\Big]_{x=x_{i-1}}^{x_i} = \frac{-1}{H_1}\left(\frac{1}{h_i} - \frac{1}{h_{i-1}}\right). \quad \text{(A11)}$$

Furthermore, as $x/h(x)^2 = (h(x) - H_2)/(H_1 h(x)^2)$,

$$\int_{x_{i-1}}^{x_i} \frac{x}{h(x)^2}\mathrm{d}x = \frac{1}{H_1}\int_{x_{i-1}}^{x_i} \frac{1}{h(x)}\mathrm{d}x - \frac{H_2}{H_1}\int_{x_{i-1}}^{x_i} \frac{1}{h(x)^2}\mathrm{d}x \quad \text{(A12)}$$

$$= \frac{1}{H_1^2}\left(\ln\left(\frac{h_i}{h_{i-1}}\right) + H_2\left(\frac{1}{h_i} - \frac{1}{h_{i-1}}\right)\right). \quad \text{(A13)}$$

A2. Numerical scheme

The mechanical model is summarised in equations (4). They can be regarded as the differential equations of a linear time-invariant system with F_{h1} and F_{h2} as inputs and the mass positions and velocities as outputs. Setting $\mathbf{y}(t) = [y_1(t), \dot{y}_1(t), y_2(t), \dot{y}_2(t)]^T$ and $\mathbf{F}(t) = [F_{h1}(t), F_{h2}(t)]^T$, equation (4) can be rewritten to $\dot{\mathbf{y}}(t) = A\mathbf{y}(t) + B\mathbf{F}(t)$ for appropriate matrices A and B. The general form of a discrete-time state-space description is

$$\dot{\mathbf{y}}[n] = E\mathbf{y}[n-1] + F\mathbf{F}[n], \quad \text{(A14)}$$

in which n denotes to the n-th iteration step. The matrices E and F have the dimensions of A and B, respectively. The objective is to design the discrete-time system such that if $\mathbf{F}[n] = \mathbf{F}(nT_s)$, then $\mathbf{y}[n] = \mathbf{y}(nT_s)$, with T_s the sampling period. This cannot be achieved in general, but the approximation is close if T_s is small compared to the smallest time constant in the system. We then may choose $E = \exp(AT_s)$, and $F = BT_s$.

Given the mass positions $y_1[n]$ and $y_2[n]$, the separation point position $x_s[n]$ can be determined. We then apply a discretised version of Bernoulli's equation (2) between x_0 and $x_s[n]$,

$$p_{3,\mathrm{ac}}[n] + \frac{\rho_0}{2}\left(\frac{U_g[n]}{h_s[n]L_g}\right)^2 = p_{0,\mathrm{phon}}[n] + p_{0,\mathrm{ac}}[n] + \frac{\rho_0}{2}\left(\frac{U_g[n]}{h_0 L_g}\right)^2. \quad \text{(A15)}$$

After substitution of the coupling equation (7) for $p_{3,\mathrm{ac}}[n]$, and similarly for $p_{0,\mathrm{ac}}[n]$, a quadratic equation results from which $U_g[n]$ can be solved if the incoming sound wave $p_1^-[n]$ from the vocal tract, and the corresponding sound pressure in the trachea, are known. Subsequently, $p_1^+[n]$ is found from $p_1^+[n] = Z_1 U_g[n] + p_1^-[n]$, and similarly for the outgoing sound wave in the trachea.

The aerodynamic forces F_{h1} and F_{h2} are determined from equation (3) as described in Appendix A1.

To discretise the acoustic model, the reflection- and transmission relations are rewritten to

$$\begin{pmatrix} p_i^+[n] \\ p_i^-[n] \end{pmatrix} = M \begin{pmatrix} p_{i-1}^+[n] \\ p_{i-1}^-[n] \end{pmatrix}. \quad \text{(A16)}$$

The components of M are functions of the coefficients $r_i = (A_i - A_{i-1})/(A_i + A_{i-1})$ and $t_i = 1 - R_i$. Given $p_1^+[n]$ and $p_1^-[n]$, this allows the discretised pressure distribution along the vocal tract (and, similarly, in trachea and lungs) to be determined.

The lip impedance, and the corresponding reflection and transmission behaviour of sound waves at the lips, is approximated by a discrete filter. Equation (6) for the lip impedance is of the form $Z_N(s) = (b_1 s + b_0)/(a_1 s + a_0)$ with $s = \mathrm{j}\omega$. We use the bi-linear transformation $z = (2 + sT_s)/(2 - sT_s)$ to design a suitable discrete time-invariant filter. This transformation has the advantage over traditional discretisation that it yields a stable filter of the form $Z_N(z) = (\beta_1 + \beta_0 z^{-1})/(1 + \alpha_1 z^{-1})$ if the time-continuous expression for $Z_N(s)$ is stable (see e.g. [26]). The discrete filter $Z_N(z)$ can be realised with a second order delay line. Similar filters can be designed for the frequency-dependent reflection- and transmission coefficients at the end of the exponential horn modelling the lungs.

Bogaert [9] and Veldhuis *et al.* [10] have reported the discrete-time state-space approach to the discretisation of

ACUSTICA · acta acustica
Vol. 84 (1998)

the mechanical vocal-fold model to reduce computing time drastically with respect to traditional numerical integration of the equations of motion.

In each iteration of numerical scheme, implemented in Matlab, the numerical calculations are performed in the order in which they are described. The sampling rate F_s is chosen to be 32 kHz.

References

[1] F. Alipour, I. R. Titze: Combined simulation of two-dimensional airflow and vocal fold vibration. – In: Vocal Fold Physiology: Controlling complexity and chaos. P. J. Davis, N. H. Fletcher (eds.). Singular Pub. Group, San Diego, 1996, 17–29.

[2] K. Ishizaka, J. L. Flanagan: Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell Syst. Tech. J. **51** (1972) 1233–1267.

[3] D. A. Berry, H. Herzel, I. R. Titze, K. Krischer: Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions. J. Acoust. Soc. Am. **95** (1994) 3595–3604.

[4] C. L. Lucero: Chest and falsetto-like oscillations in a two-mass model of the vocal folds. J. Acoust. Soc. Am. **100** (1996) 3355–3359.

[5] B. H. Story, I. R. Titze: Voice simulation with a body-cover model of the vocal folds. J. Acoust. Soc. Am. **97** (1995) 1249–1260.

[6] H. Herzel: Possible mechanisms of vocal instabilities. – In: Vocal Fold Physiology: Controlling complexity and chaos. P. J. Davis, N. H. Fletcher (eds.). Singular Pub. Group, San Diego, 1996, 63–75.

[7] J. Liljencrants. private communication, 1993.

[8] X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands, Y. Auregan: Theoretical and experimental study of quasisteady flow separation within the glottis during phonation. J. Acoust. Soc. Am. **96** (1994) 3416–3431.

[9] I. J. M. Bogaert: Speech production by means of a hydrodynamic model and a discrete-time description. IPO-report 1000, Institute for Perception Research, Einhoven, The Netherlands, 1994.

[10] R. N. J. Veldhuis, I. J. M. Bogaert, N. J. C. Lous: Two-mass models for speech synthesis. Proceedings of the 4th European Conference on Speech Communication Technology, Madrid, Spain, 1995. 1853–1856.

[11] D. A. Berry, I. R. Titze: Normal modes in a continuum model of vocal fold tissues. J. Acoust. Soc. Am. **100** (1996) 3345–3354.

[12] G. Fant: Acoustic theory of speech production. Mouton & Co, 's-Gravenhage, The Netherlands, 1960, 115.

[13] K. Ishizaka, M. Matsudaira, T. Kaneko: Input acoustic-impedance measurement of the subglottal system. J. Acoust. Soc. Am. **60** (1976) 190–197.

[14] B. Cranen: The acoustic impedance of the glottis; measurements and modelling. Dissertation. University of Nijmegen, The Netherlands, 1987.

[15] M. Rothenberg: Acoustic interaction between the glottal source and vocal tract. – In: Proceedings of the Kurume Vocal Fold Physiology Conference. K. N. Stevens, M. Hirano (eds.). University of Tokyo Press, 1981, 305–328.

[16] M. Hirano, S. Kurita, T. Nakashima: The structure of the vocal folds. – In: Proceedings of the Kurume Vocal Fold Physiology Conference. K. N. Stevens, M. Hirano (eds.). University of Tokyo Press, 1981, 33–44.

[17] T. Bear: Investigation of the phonatory mechanism. ASHA Reports **11** (1981) 38–46.

[18] J. W. van den Berg, J. T. Zantema, P. Doornenbal, Jr.: On the air resistance and the Bernoulli effect of the human larynx. J. Acoust. Soc. Am. **29** (1957) 626–631.

[19] H. M. Teager, S. M. Teager: Evidence for non-linear mechanisms in the vocal tract. – In: Speech Production and Speech Modeling. W. J. Hardcastle, A. Marchal (eds.). Kluwer, Dordrecht, The Netherlands, 1990.

[20] A. Hirschberg, X. Pelorson, G. C. J. Hofmans, R. R. van Hassel, A. P. J. Wijnands: Starting transient of the flow through an in-vitro model of the vocal folds. – In: Vocal Fold Physiology: Controlling complexity and chaos. P. J. Davis, N. H. Fletcher (eds.). Singular Pub. Group, San Diego, 1996, 31–46.

[21] P. M. Morse, K. U. Ingard: Theoretical acoustics. McGraw-Hill, New-York, 1968.

[22] P. O. A. L. Davies, R. S. McGowan, Shadle: Practical flow duct acoustics applied to the vocal tract. – In: Vocal Fold Physiology: Frontiers in basic science. I. R. Titze (ed.). Singular Pub., San Diego, 1993, 93–142.

[23] S. Hertegard, J. Gauffin: Voice source-vocal tract interaction during high-pitched female singing. – In: Proceedings of the Stockholm Music Acoustic Conference SMAC93. A. Friberg, J. Iwarsson, E. Jansson, J. Sundberg (eds.). Roy. Swedish Aca. of Music 79, 1994, 177–182.

[24] H. K. Schutte, D. G. Miller: The effect of F_0/F_1 coincidence in soprano high notes on pressure at the glottis. J. of Phonetics **14** (1986) 385–392.

[25] Y. Koike: Sub- and supraglottal pressure variation during phonation. – In: Proceedings of the Kurume Vocal Fold Physiology Conference. K. N. Stevens, M. Hirano (eds.). University of Tokyo Press, 1981, 181–192.

[26] A. Papoulis: Signal analysis. McGraw-Hill, London, 1984, 170-173.