Modal locking between vocal fold oscillations and vocal tract acoustics

Tiina Murtola[1,2], Atte Aalto[2,3], Jarmo Malinen[2], Daniel Aalto[4,5], Martti Vainio[4]

[1]Dept. of Signal Processing and Acoustics, Aalto University, Finland

[2]Dept. of Mathematics and Systems Analysis, Aalto University, Finland

[3]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg

[4]Institute of Behavioural Sciences (SigMe group), University of Helsinki, Finland

[5]Communication Sciences and Disorders, University of Alberta, Canada.

Summary

During voiced speech, vocal folds interact with the vocal tract acoustics. The resulting glottal source–resonator coupling has been observed using mathematical and physical models as well as in *in vivo* phonation. We propose a computational time-domain model of the full speech apparatus that contains a feedback mechanism from the vocal tract acoustics to the vocal fold oscillations. It is based on numerical solution of ordinary and partial differential equations defined on vocal tract geometries that have been obtained by magnetic resonance imaging. The model is used to simulate rising and falling pitch glides of [ɑ, i] in the fundamental frequency (f_o) interval $[145\,\mathrm{Hz}, 315\,\mathrm{Hz}]$. The interval contains the first vocal tract resonance f_{R1} and the first formant F_1 of [i] as well as the fractions of the first resonance $f_{R1}/5$, $f_{R1}/4$, and $f_{R1}/3$ of [ɑ]. The glide simulations reveal a locking pattern in the f_o trajectory approximately at f_{R1} of [i]. The resonance fractions of [ɑ] produce perturbations in the pressure signal at the lips but no locking.

1 Introduction

The classical source–filter theory of vowel production assumes that the source (i.e., the vocal fold vibration) operates independently of the filter (i.e., the vocal tract, henceforth VT) whose resonances modulate the resulting sound [1, 2]. Even though this approach captures a wide range of phenomena in speech production, some observations remain unexplained by the source–filter model lacking feedback. The purpose of this article is to address some of these observations using computational modelling.

In this work, simulations where the fundamental frequency (f_o) rises and falls over the range $[145\,\mathrm{Hz}, 315\,\mathrm{Hz}]$ are considered for vowels [ɑ] and [i]. Similar glides recorded from eleven female test subjects are treated in the companion article [3]. Such glides are particularly interesting when the f_o range intersects an isolated acoustic resonance of the supra- or subglottal cavity. Since the lowest formant F_1 usually lies high above f_o in adult male phonation, this situation is more typical in females and children when they are producing vowels with low F_1 such as [i]. As reported in Section 5, simulations reveal (in addition to other observations) a characteristic locking behaviour of f_o at the VT acoustic resonance[1] $f_{R1} \approx F_1$.

This article has two equally important objectives. Firstly, we pursue better understanding of the time-domain dynamics of glottal pulse perturbations near f_{R1} of [i]. An acoustic and flow-mechanical model of the speech apparatus is a well-suited tool for this purpose. Secondly, we introduce and validate a computational model that meets these requirements. The proposed model has been originally designed to be a glottal source for a high-resolution 3D computational acoustics model of the VT which is being developed for medical purposes. There is also an emerging application for such models as a development platform of speech signal processing algorithms [5, 6, 7]. Since perturbations of f_o near F_1 are a widely researched, yet quite multifaceted phenomenon, as discussed next, it is a good candidate for model validation experiments.

The simulations carried out in this article indicate special kinds of perturbations in vocal folds vibrations near a VT resonance. The mere existence of such perturbations is not surprising considering the wide range of existing literature. Since the seminal work of [8], a wide range of glottal source perturbation patterns related to acoustic loading has been investigated. Experiments were carried out in [9] on excised larynges mounted on a resonator to determine how glottal amplitude ratio changes with the subglottal resonator length. Physical models were used in [10] with a subglottal resonator to study phonation onsets and offsets, and in [11] with sub- and supraglottal resonators to study phonation onsets. The latter also considered

[1]The notation of [4] is used to differentiate resonances and formants though, of course, we expect $f_{Rj} \approx F_j$ for $j = 1, 2, \ldots$.

the dynamics of frequency jumps as the natural frequency of their physical model was varied over time. Similarly, a physical model of phonation with a tubular, variable length supraglottal resonator was studied in [12, 13], and it was used to validate a flow-acoustic model somewhat resembling the one proposed in this article.

The source–filter interaction problem was approached in [14] using both reasoning based on sub- and supraglottal impedances and a non-computational flow model as well as computational model comprising a multi-mass vocal fold model and wave-reflection models of the subglottal and supraglottal systems. A two-mass model of vocal folds, coupled with a variable-length resonator tube, was used in [15], and pitch glides were simulated using a four-mass model to analyse the interactions between vocal register transitions and VT resonances in [16].

These works reveal a consistent picture of the existence of perturbations caused by resonant loads, and this phenomenon has also been detected experimentally in [17] using speech recordings, in [18] using simultaneous recordings of laryngeal endoscopy, acoustics, aerodynamics, electroglottography, and acceleration sensors, and in [19] using simultaneous speech, electroglottography and accelerometer recordings combined with separate resonance estimation measurements.

Although the existence of these perturbations has been well reported, speech modelling studies have given only limited attention to the time-domain dynamics of fundamental frequency glides where such perturbations would be expected to occur. Of the above mentioned studies, upward glides were simulated in [11] by varying the natural frequency of their physical model over time. Their small amplitude oscillation model exhibited a frequency jump when crossing the resonance of their downstream tube when the acoustic coupling was sufficiently strong. Downward glides were simulated in [14] followed by upward glides by varying the parameters of a multi-mass vocal fold model. Frequency jumps, subharmonics and amplitude changes were observed in the regions where load reactances were changing rapidly. Changes in the rate of change of the fundamental frequency in these regions can also be seen in their Figures 10-14. In [16] upward glides were simulated followed by downward glides by adjusting the tension parameter (i.e., decreasing masses and increasing stiffness parameters by the same factor) in their four-mass vocal fold model. They observed frequency jumps associated with register changes, which in turn were shown to occur at different frequencies depending on the VT load.

Some of the most popular approaches to modelling phonation are based on the Kelly–Lochbaum VT [20] or various transmission line analogues [21, 22, 23]. Contrary to these approaches, the proposed model consists of (ordinary and partial) differential equations, conservation laws, and coupling equations. In this modelling paradigm, the temporal and spatial discretisation is conceptually and practically separated from the actual mathematical model of speech. The computational model is simply a numerical solver for the model equations, written in MATLAB environment. The modular design makes it easy to decouple model components for assessing their significance to simulated behaviour.[2] Since the generalised Webster's equation for the VT acoustics assumes intersectional area functions as its geometric data, VT configurations from magnetic resonance imaging (MRI) can be used without transcription to non-geometric model parameters. Further advantages of speech modelling with Webster's equation have been explained in [25].

The proposed model is of low order: it aims at qualitatively realistic functionality, tunability by a low number of parameters, and tractability of model components, equations, and their relation to biophysics. Similar functionality in higher precision can be obtained using computational fluid dynamics with elastic tissue boundaries. Such approaches aim to model the speech apparatus as undivided whole [26], but the computational cost is much higher compared to our model or the models proposed in, e.g., [25] and [27]. Numerical efficiency is a key issue because some parameter values or their feasible ranges (in particular, for hard-to-get physiological parameters) can only be determined by trial and error, leading to a high number of required simulations as discussed in [30, Chapter 4]. The proposed model is hence suitable for investigating speech phenomena where realistic model output is only produced with a narrow range of control parameter values.

2 Phonation Model

2.1 Vocal Fold Mechanics

Voiced speech sounds originate from self-sustained quasi-periodic oscillations of the vocal folds where the closure of the aperture between the vocal folds, i.e. the glottis, cuts off the airflow from lungs in a process called phonation. A single period of the glottal flow produced by phonation is known as a glottal pulse.

The main mechanism controlling the f_o of voiced speech is contraction of the cricothyroid muscles which leads to stretching the vocal folds and hence increased stress. Secondary mechanisms of f_o control include the vertical movement of larynx and changes in the subglottal pressure through the control of respiratory muscles.

[2] Some economy of modelled features is desirable to prevent "overfitting" while explaining experimental facts. Good modelling practices in mathematical acoustics have been discussed in [24, Chapter 8].

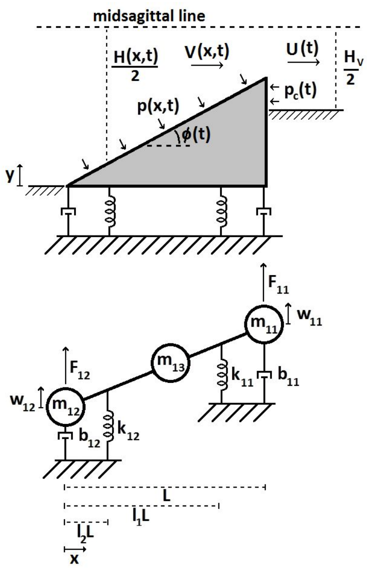

Figure 1: Top: The geometry of the glottis model with the trachea to the left and the vocal tract to the right. Bottom: Lumped-element representation of the lower vocal fold ($j = 1$) with two degrees of freedom.

2.1.1 Equations of motion

The anatomic vocal fold configuration is idealised as a low-order mass-spring system with aerodynamic surfaces as shown in Figure 1. For previous uses and more detailes on this model, see [28, 29, 30] and [31]. Such lumped-element models have been used frequently (see, e.g., [13, 32, 33, 34, 35, 36] and the reviews [37, 38]) since the introduction of the classic two-mass model [8].

The radically simplified glottis geometry in Figure 1 (top) corresponds to the coronal section through the center of the vocal folds. Both f_o and the phonation type can be changed by adjusting parameter values [30, Section 4]. However, register shifts are not within the scope of this model.

The vocal fold model consists of two wedge-shaped moving elements whose distributed mass is reduced to three mass points which, for the j^{th} fold, $j = 1, 2$, are located so that m_{j1} is at $x = L$, m_{j2} at $x = 0$, and m_{j3} at $x = L/2$. Here L denotes the thickness of the vocal fold structures. The masses are calculated so that the reduced system retains the mass, and static and inertial moments of a parabolic vocal fold shape (for details, see [31, p. 14]). Each vocal fold has two degrees of freedom: m_{j1} and m_{j2} can move in the y-direction. Although this causes some distortion to the shape of the wedges, the displacements in the x-direction are small enough that the effect is negligible. The elastic support of the vocal ligament is approximated by two springs at points $x = l_1 L$ and $x = l_2 L$, and losses caused by internal resistance of the tissues to movement and deformation is represented by two dampers at points $x = 0$ and $x = L$.

The equations of motion for the vocal folds are

$$\begin{cases} M_1\ddot{W}_1(t) + B_1\dot{W}_1(t) + K_1W_1(t) = F_1(t), \\ M_2\ddot{W}_2(t) + B_2\dot{W}_2(t) + K_2W_2(t) = F_2(t), \qquad t \in \mathbb{R}, \end{cases} \tag{1}$$

where $W_j(t) = \left[w_{j1}(t)\ w_{j2}(t)\right]^T$ are the displacements of m_{j1} and m_{j2} in the y-direction as shown in Figure 1 (bottom). The load force pair $F_j(t) = \left[F_{j1}(t)\ F_{j2}(t)\right]^T$ comprises acoustic pressure forces as well as aerodynamic pressure forces when the glottis is open (equation (9)) and collision forces when the glottis is closed (equation (5)). The mass, damping, and stiffness matrices M_j, B_j, and K_j, respectively, in (1) are

$$M_j = \begin{bmatrix} m_{j1} + \frac{m_{j3}}{4} & \frac{m_{j3}}{4} \\ \frac{m_{j3}}{4} & m_{j2} + \frac{m_{j3}}{4} \end{bmatrix}, \quad B_j = \begin{bmatrix} b_{j1} & 0 \\ 0 & b_{j2} \end{bmatrix},$$

$$\text{and} \quad K_j = \sum_{i=1}^{2} k_{ji} \begin{bmatrix} l_i^2 & l_i(1-l_i) \\ l_i(1-l_i) & (1-l_i)^2 \end{bmatrix}. \tag{2}$$

The entries of these matrices have been computed using Lagrangian mechanics. The damping matrices B_j are diagonal since the dampers are located at the endpoints of the vocal folds. The model supports asymmetric vocal fold vibrations but for this work, symmetry of left and right vocal folds is imposed by using parameters $M = M_j$, $K = K_j$, and $B = B_j$, $j = 1, 2$, and by setting $F(t) = F_2(t) = -F_1(t)$. As a further simplification, tissue damping is assumed to be uniform everywhere, i.e., $b_i = \beta$ for $i = 1, 2$. The parameters in (2) as well as the load force components in (1) are illustrated in Figure 1.

The gap between the vocal folds is denoted by $H(x, t)$, and in the model geometry (Figure 1 (top))

$$H(x,t) = H_0(t) + \frac{x}{L}(H_L(t) - H_0(t)), \quad x \in [0, L], \tag{3}$$

where inferior glottal gap $H_0(t) = H(0, t)$ and superior glottal gap $H_L(t) = H(L, t)$ are related to (1) through

$$\begin{bmatrix} H_L(t) \\ H_0(t) \end{bmatrix} = W_2(t) - W_1(t) + \begin{bmatrix} g_L \\ g_0 \end{bmatrix}. \tag{4}$$

The rest gap parameters g_0 and g_L correspond to the points $x = 0$ and $x = L$, respectively.

2.1.2 Vocal fold collision

When the glottis is closed (i.e., $H_L(t) < 0$), there is no airflow between the vocal folds and hence no force arising from it affecting the vocal folds. There are, however, nonlinear spring forces with parameter k_H, accounting for the contact force of the vocal folds. They are accompanied by the acoustic counter pressure from the VT and subglottal tract (SGT), denoted by $p_c = p_c(t)$ in (15). Thus, the force pair for equation (1) during glottal closed phase is given by

$$F = F_H = \begin{bmatrix} k_H |H_L|^{3/2} - A_{pc} p_c \\ A_{pc} p_c \end{bmatrix}, \quad \text{for} \quad H_L < 0, \tag{5}$$

where the area $A_{pc} = A_{pc}(t)$ is the nominal area on which p_c acts corrected with relative moment arms (see equation (16)).

This approach is related to the Hertz impact model that has been used similarly in [32] and [39]. When the glottis is open (i.e., $H_L(t) > 0$), the spring force in (5) is not enabled. Then the load terms in equation (1) are given by $F(t) = F_A(t)$ as introduced in equation (9) in terms of the aerodynamic forces from the glottal flow.

2.2 Glottal Flow Aerodynamics

The main component of the airflow within the speech apparatus, to which the acoustic component acts as a perturbation, is assumed to be incompressible and one-dimensional, and to satisfy mass conservation and Newton's second law. The flow is also assumed to be lossless everywhere except at the glottal opening. This main glottal flow (volume velocity) component is described by

$$\dot{U}(t) = \frac{1}{I_L} \left(p_s(t) - R_g(t) U(t) \right), \tag{6}$$

where $p_s(t)$ is the driving stagnation pressure at the lungs whose time variation is assumed to be slow, I_L regulates the inertia of the load air column, and $R_g(t)$ represents non-recoverable losses in the glottis.

Equation (6) is related to Newton's second law for the air column in motion, and it can be derived (following [31, Section 2.2]) from the pressure balance $p_s = p_g + p_a$, where the pressure change from the lungs to the outside space is the sum of the glottal pressure loss p_g and the accelerating pressure p_a of the fluid column in the airways. To obtain an expression for p_a, the power of accelerating an (incompressible) fluid column is considered. This power is equal to the derivative of the kinetic energy of the fluid column, yielding $p_a(t)U(t) = \rho U(t)\dot{U}(t) \int \frac{d\vec{r}}{A(\vec{r})^2}$, where the integration is extended over the VT and SGT volumes. Here, $A(\vec{r})$ denotes the area of the fluid column cross-section that contains the position vector $\vec{r}$, and incompressibility $A(\vec{r})v(\vec{r}, t) = U(t)$ was used. By denoting the nominal value of inertance $I_L = \rho \int \frac{d\vec{r}}{A(\vec{r})^2}$, these equations yield $p_a = I_L \dot{U}(t)$. In the context of the airways, the nominal inertance can be split into VT and SGT contributions $I_V = \rho \int_0^{L_{VT}} \frac{ds}{A(s)}$ and $I_S = \rho \int_0^{L_S} \frac{ds}{A_S(s)}$, respectively, so that $I_L = I_V + I_S$; see Sections 2.3 and 2.4.

Unfortunately, the integration over the volume of airways (even if the SGT geometry was available) does not necessarily yield the correct total inertance. The flow outside of mouth as well as the masses of the lungs, diaphragm, etc., are coupled to the flow. For the same reason, the inertial effect for VT and SGT, observed in the low frequency limit of the acoustic equations (10) and (14), does not give a sufficient account of the total intertance since not all of it is due to acoustics. Thus, the inertance parameter I_L must, in general, be used as a tuning parameter. The high frequency feedback from the VT acoustics to the glottal flow, a particularly notable effect in phonations where the glottis does not fully close, is not included in (6).

The glottal pressure loss consists of two components following [40]

$$p_g = R_g(t)U(t) = \frac{12\mu L_g U(t)}{h H_L(t)^3} + \frac{k_g \rho U(t)^2}{2h^2 H_L(t)^2}. \tag{7}$$

The first term represents the viscous pressure loss, and it is motivated by the Hagen–Poiseuille law in a narrow aperture. It approximates the pressure loss in the glottis using a rectangular tube of width h, height H_L, and length L_g. The parameter μ is the kinematic viscosity of air. The second term takes into account the pressure losses not attributable to viscosity in the same sense as the first. The coefficient k_g represents the difference between pressure drop at the glottal inlet and recovery at the outlet. This coefficient depends not only on the glottal geometry but also on the glottal opening, driving pressure, and flow through the glottis [41]. Equations (6)–(7) bear resemblance to the description of airflow in [12, 13] where the pressure drop, loss, and recovery effects, however, are accounted for by flow separation in a diverging channel.

The pressure $p(x, t)$ in the glottis is given in terms of $U = U(t)$ by making use of the Bernoulli theorem $p(x, t) + \frac{1}{2}\rho V(x, t)^2 = p_s$ for the Venturi effect, where $V(x, t)$ is the velocity within the glottis, and the mass conservation law $hH(x, t)V(x, t) = U(t)$. Since each vocal fold has two degrees of freedom, $p(x, t)$ and the VT/SGT counter pressure p_c can be reduced to an aerodynamic force pair $F_A = \begin{bmatrix} F_{A,1} & F_{A,2} \end{bmatrix}^T$ where $F_{A,1}$ acts at $x = L$ and $F_{A,2}$ at $x = 0$ in Figure 1 (bottom). This reduction can be carried out by using the total force and moment balance equations

$$\begin{aligned} F_{A,1} + F_{A,2} &= h \int_0^L \left(p(x, t) - p_r \right) dx \text{ and} \\ L F_{A,1} &= \frac{h}{\cos^2 \phi} \int_0^L x \left(p(x, t) - p_r \right) dx - L A_{pc} p_c, \end{aligned} \tag{8}$$

where $\phi = \phi(t)$ is the angle of the inclined vocal fold surface as shown in Figure 1 (top), A_{pc} accounts for the moment arms and areas on which p_c acts (see equation (16)), and p_r is the reference pressure corresponding to the equilibrium position $w_{ij} = 0$ for $i, j = 1, 2$. Since the displacements w_{ij} are in the y-direction only, the aerodynamic forces have been assumed to act in this direction as well. 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.

The force calculations are done using the pressure difference $p(x,t)-p_r$ so that $F_{A,1}$ and $F_{A,2}$ vanish when $p(x,t)=p_r$ and $p_c=0$. The reference pressure is associated with the hydrostatic pressure reference level in vibrating tissues, and it is expected to satisfy $p_r \le p_s$. If $p_r = p_s$ is used, the aerodynamic force always tries to close the glottis. For small flow velocities $V(x,t)$, using $p_r < p_s$ results in the driving pressure p_s pushing the vocal folds open more strongly than the aerodynamic force pulls them close. There is no obvious way to determine the true magnitude of p_r as it is an outcome of dynamic pressure equalisation processes related to p_s and the additional partial pressure due to haemodynamics in tissues. For this work, it is assumed that $p_r = 0.5p_s^0$, where $p_s^0 = p_s(0)$, and the equilibrium gap parameter $g_L > 0$ so that starting simulations with a closed glottis is not necessary.

Evaluation of the integrals in (8) yields, for $H_L > 0$,

$$\begin{aligned} F_{A,1} = {} & \frac{hL}{2\cos^2\phi}\Bigg(-\frac{\rho U^2}{h^2 H_L (H_0 - H_L)} \\ & + \frac{\rho U^2}{h^2 (H_L - H_0)^2}\ln\left(\frac{H_0}{H_L}\right) + (p_s - p_r)\Bigg) \\ & - A_{pc} p_c, \quad \text{and} \\ F_{A,2} = {} & \frac{hL}{2\cos^2\phi}\Bigg(\frac{\rho U^2\left(H_0 \sin^2\phi + H_L\cos^2\phi\right)}{h^2 H_L H_0 (H_0 - H_L)} \\ & - \frac{\rho U^2}{h^2 (H_L - H_0)^2}\ln\left(\frac{H_0}{H_L}\right) + \cos(2\phi)\,(p_s - p_r)\Bigg) \\ & + A_{pc} p_c. \end{aligned} \tag{9}$$

During the glottal closed phase (i.e., when $H_L(t) < 0$), the aerodynamic force (9) is not enabled, and the vocal fold load force is instead given by equation (5).

2.3 Vocal Tract Acoustics

A generalised version of Webster's horn model resonator is used as acoustic loads to represent both the VT and the SGT. It is given by

$$\frac{A(s)}{c^2\Sigma(s)^2}\frac{\partial^2\psi}{\partial t^2} + 2\pi\alpha W(s)\frac{\partial\psi}{\partial t} - \frac{\partial}{\partial s}\left(A(s)\frac{\partial\psi}{\partial s}\right) = 0, \tag{10}$$

where c denotes the speed of sound, the parameter $\alpha \ge 0$ regulates the energy dissipation through air/tissue interface, and the solution $\psi = \psi(s,t)$ is the velocity potential of the acoustic field; i.e., $v = -\frac{\partial\psi}{\partial s}$. Then the sound pressure is given by $p = \rho\frac{\partial\psi}{\partial t}$, where ρ denotes the density of air. The generalised Webster's model for acoustic waveguides has been derived from the wave equation in a tubular domain in [42], its solvability and energy notions have been treated in [43], and the approximation properties in [44].

The generalised Webster's equation (10) is applicable if the VT is approximated as a curved tube of varying cross-sectional area and length L_{VT}. The three-dimensional centreline $\gamma(s)$ of the tube is parametrised using distance $s \in [0, L_{VT}]$ from the superior end of the glottis. At every s, the cross-sectional area of the tube perpendicular to the centreline is given by the area function $A(s)$, and the (hydrodynamic) radius of the tube, denoted by $R(s)$, is defined by $A(s) = \pi R(s)^2$. The curvature of the tube is $\kappa(s) = \|\gamma''(s)\|$, and the curvature ratio $\eta(s) = R(s)\kappa(s) < 1$.

The final parameters appearing in (10) are the stretching factor $W(s)$ and the sound speed correction factor $\Sigma(s)$ for curvature, defined by

$$\begin{aligned} & W(s) = R(s)\sqrt{R'(s)^2 + (\eta(s)-1)^2}, \;\text{ and} \\ & \Sigma(s) = \left(1 + \tfrac{1}{4}\eta(s)^2\right)^{-1/2}. \end{aligned} \tag{11}$$

2.3.1 Boundary conditions

The VT resonator is coupled to the glottal flow given by equation (6) with

$$\frac{\partial\psi}{\partial s}(0,t) = -\frac{U_{AC}(t)}{A(0)}, \tag{12}$$

where the DC component has been removed from the glottal flow, i.e., $U_{AC}(t) = U(t) - \frac{1}{T}\int_{t-T}^{t} U(\tau)\,d\tau$ with $T = 2/f_o$. The effect of this removal is negligible when phonation has become stable, but it is more pronounced at the beginning of simulations when a stable waveform has not yet developed. Equations (10)–(12) characterise a variant of the source–filter model in the sense that the acoustics of the VT is only excited at the glottis.

At the lips, the reactive acoustic response of the exterior space is modelled by the differential equation

$$\begin{aligned} & -R_m L_m \frac{\partial\psi}{\partial s}\left(L_{VT}, t\right) \\ & = \frac{\rho}{A(L_{VT})}\left(R_m \psi\left(L_{VT}, t\right) + L_m \frac{\partial\psi}{\partial s}\left(L_{VT}, t\right)\right), \end{aligned} \tag{13}$$

which corresponds to the impedance $Z(\xi) = \frac{\xi R_m L_m}{R_m + \xi L_m}$ of the same form as the "first-order high pass model" for termination of an acoustic horn in [45, Section 4.1]. The circuit topology of this model is the parallel coupling of a resistor and an inductor.

2.4 Subglottal acoustics

Anatomically, the SGT consists of the airways below the larynx: trachea, bronchi, bronchioles, alveolar ducts, alveolar sacs, and alveoli. This system has been modelled either as a tree-like structure [27] or, more simply, as an acoustic horn whose area increases towards the lungs [34, 46]. We take the latter approach and denote the cross-sectional area and the horn radius by $A_S(s)$ and $R_S(s)$ (see equation (17)), respectively, where $s \in [0, L_S]$ and L_S is the nominal length of the SGT.

Since the subglottal horn is assumed to be straight, we have $\eta = 0$, $\Sigma = 1$ and $W_s(s) = R_S(s)\sqrt{R_S'(s)^2 + 1}$. Then equations (10)–(12) translate to

$$\begin{cases} \frac{A_S(s)}{c^2}\frac{\partial^2 \widetilde{\psi}}{\partial t^2} + 2\pi\alpha W_s(s)\frac{\partial \widetilde{\psi}}{\partial t} - \frac{\partial}{\partial s}\left(A_S(s)\frac{\partial \widetilde{\psi}}{\partial s}\right) = 0, \\ \frac{\partial \widetilde{\psi}}{\partial t}(L_S, t) + \theta_s c \frac{\partial \widetilde{\psi}}{\partial s}(L_S, t) = 0, \\ \frac{\partial \widetilde{\psi}}{\partial s}(0, t) = \frac{U_{AC}(t)}{A_S(0)}, \end{cases} \tag{14}$$

where the solution $\widetilde{\psi}$ is the velocity potential for the SGT acoustics. Instead of using the reactive boundary dynamics (13), the termination loss at lungs is characterised by normalised acoustic resistance $\theta_s \geq 0$ in equation (14). SGT acoustics is a important factor in phonation in general but its contribution to changes occurring during glide simulations is negligible as long as f_o is far from the subglottal resonances.

2.5 Acoustic counter pressure

The feedback coupling from VT/SGT acoustics back to vocal fold surfaces is realised as the product of the acoustic counter pressure $p_c = p_c(t)$ and the moment corrected area $A_{pc} = A_{pc}(t)$ as already shown in equations (5) and (9) above.

The counter pressure is the resultant of VT and SGT pressure components, and it is given in terms of velocity potentials from equations (10) and (14) by

$$p_c(t) = Q_{pc}\rho\left(\psi_t(0, t) - \widetilde{\psi}_t(0, t)\right), \tag{15}$$

where tuning parameter $Q_{pc} \in [0, 1]$ enables scaling the magnitude of the feedback. The parameter Q_{pc} is necessary because the wedge geometry tends to overestimate the area of the vocal fold surface on which p_c can do work, and further, it is difficult to directly estimate the proportions of the underlying flow and the superimposed acoustics. In simulations, overestimation of the acoustic feedback forces leads to permanently non-stationary, even chaotic vibrations of the vocal folds, which are outside the scope of this work.

The area A_{pc} is best understood in reference to the moment balance in equation (8), although it appears in the same way in both equations (5) and (9). For each vocal fold, p_c acts on the area $\frac{1}{2}(H_V - H_L)h$ and produces a moment arm of $\frac{1}{4}(2H_0 - H_V - H_L)$ around points $(x, y) = (0, 0)$ and $(x, y) = (0, H_0)$ for the lower and upper folds, respectively. Hence

$$A_{pc} = \frac{h}{8L}(H_V - H_L)(2H_0 - H_V - H_L). \tag{16}$$

Equations (15) and (16) assume that both the VT and SGT pressure components act in the x-direction only (i.e., horizontally in Figure 1 (top)). This assumption minimises the tendency of the wedge geometry to overestimate the effect of the SGT compared to the effect of the VT.

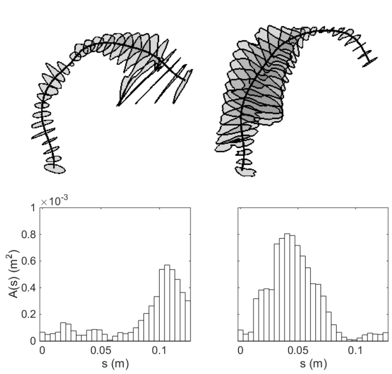

Figure 2: Top: The VT intersections extracted during phonation of [ɑ] and [i]. Bottom: The resulting area functions for equation (10) as a function of distance from the glottis.

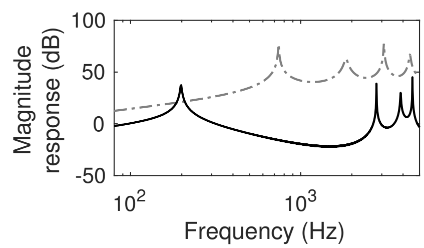

Figure 3: The magnitude responses of the VT acoustic loads obtained by simulating output for an impulse input for [ɑ] (dashed gray) and [i] (solid black). The response of [ɑ] has been raised by 50 dB for clarity.

3 Parameters

3.1 Vocal tract

Table 1: VT parameter values.

Parameter	[ɑ]	[i]
Inertance, I_V	$2540\,\frac{\mathrm{kg}}{\mathrm{m}^4}$	$2820\,\frac{\mathrm{kg}}{\mathrm{m}^4}$
Length, L_{VT}	132 mm	136 mm
1st resonance, f_{R1}	742 Hz	198 Hz
2nd resonance, f_{R2}	1846 Hz	2791 Hz
Area at mouth	299 mm^2	66 mm^2
R_m	$1.98 \cdot 10^6\,\frac{\mathrm{kg}}{\mathrm{s\ m}^4}$	$8.96 \cdot 10^4\,\frac{\mathrm{kg}}{\mathrm{s\ m}^4}$
L_m	$33.2\,\frac{\mathrm{kg}}{\mathrm{m}^4}$	$70.6\,\frac{\mathrm{kg}}{\mathrm{m}^4}$
$\mathrm{Re}(Z(400\pi i))$	879	$4.44 \cdot 10^4$
$\mathrm{Im}(Z(400\pi i))$	$4.17 \cdot 10^4$	$4.48 \cdot 10^4$

Solving Webster's equation requires that the VT is represented with an area function and a centreline, from which curvature information can be computed. Two different VT geometries corresponding to vowels from a healthy 26 years old female are used: A prolonged [ɑ] produced at $f_o = 168$ Hz and similarly produced [i] at $f_o = 210$ Hz. These geometries have been obtained by MRI using the experimental setting described in [47]; see also [48, 49, 50] for earlier approaches. The extraction of the computational geometry from raw MRI data has been carried out by the custom software described in [51, 52]. The VT geometries and their area functions are shown in Figure 2, their simulated frequency responses in Figure 3, and and the VT geometry dependent parameter values are given in Table 1.

The reactive acoustic loading (13) at the lips requires values for R_m and L_m. The values in Table 1 were obtained by interpolation at 200 Hz from the piston model given in [53, Chapter 7, Eq. (7.4.31)] and tuning of R_m to remove excessive fluctuations in simulated waveforms. The low order rational model $Z(\xi) = \frac{\xi R_m L_m}{R_m + \xi L_m}$ approximates the irrational piston model impedance very well for frequencies within 100 Hz . . . 2 kHz, and the frequency responses in Figure 3 are reasonable as well.

3.2 Subglottal tract

Full SGT geometry cannot be constructed from the MRI data that is used for the VT. Instead, an exponential horn is used as the SGT area function for equation (14)

$$A_S(s) = A_S(0)e^{\epsilon s}, \text{ where } \epsilon = \frac{1}{L_S} \ln\left(\frac{A_S(L_S)}{A_S(0)}\right) \quad (17)$$

following [46]. The values for $A_S(0) = 2\,\text{cm}^2$ and $A_S(L_S) = 10\,\text{cm}^2$ are taken from [46, Figure 1]. The horn length L_S is selected so that the lowest subglottal resonance is $f'_{R1} = 500\,\text{Hz}$ which results in the second lowest resonance at $f'_{R2} = 1.0\,\text{kHz}$. This is a reasonable value for f_{R1} based on [9, Table 1]; see also [39, 54, 55] and [27, Figure 1]. The SGT lung termination resistance in equation (14) is given the value $\theta_s = 1$ which corresponds to an absorbing boundary condition. The air column in this SGT model has a inertia parameter value $I_S = 1040\,\text{kg/m}^4$.

Table 2: Physical and physiological constants.

Parameter	Value
speed of sound in air, c	$343\,\frac{m}{s}$
density of air, ρ	$1.2\,\frac{kg}{m^3}$
kinematic viscosity of air, μ	$18.27\,\frac{\mu Ns}{m^2}$
VT/SGT loss coeff., α	$76\frac{\mu s}{m}$
glottal gap at rest at $x = 0$, g_0	10.9 mm
glottal gap at rest at $x = L$, g_L	0.4 mm
control gap above glottis, H_V	2 mm
vocal fold length [56], h	10 mm
vocal fold thickness [32], L	6.8 mm
1^{st} vocal fold spring location, l_1	0.85
2^{st} vocal fold spring location, l_2	0.15
contact spring constant [32], k_H	$730\,\frac{N}{m^{3/2}}$
viscous thickness, L_g	1.5 mm
SGT length, L_S	350 mm
resistance at lungs, θ_s	1
entrance/exit coeff., k_g	0.6
initial driving pressure, p_s^0	650 Pa

3.3 Static parameter values

Table 2 lists the numerical values of physiological and physical constants used in all simulations. Note that the vocal fold springs are, for this study, placed symmetrically about the midpoint of the vocal folds.

The masses in M are calculated by combining the vocal fold shape function used in [32] with female vocal fold length reported in [56], yielding a total vibrating mass $m_1 + m_2 + m_3 = 0.27\,\text{g}$. A first estimate for the spring coefficients in K is calculated by assuming that the first eigenfrequency of the vocal folds matches the starting frequency for the simulations. The spring coefficients are then adjusted until simulations produce $f_o \approx 145$ Hz, giving the initial K^0 for equations (18)–(19) with total spring coefficients $k_1 + k_2 = 248\,\text{N/m}$. For details of these calculations, see [31] and [30].

The vocal fold damping parameter β plays an important but problematic role in vocal fold models. If there is too much damping, sustained oscillations do not occur. Conversely, too low damping causes instability in simulated vocal fold oscillations. The magnitude of physically realistic damping in vibrating tissues is not available, and, due to its simplifications, the present model could fail to produce quasi-stationary phonation even if realistic experimental damping values were used. For this article, $\beta = 0.009\,\text{kg/s}$ is used as it produces slowly changing glottal pulse amplitudes when simulations are carried out with constants parameters as well as in feedback free glides. This damping is small enough that the resonances of the mass-spring-damper system (1) are defined approximately by M and K alone.

In this work, the nominal values of I_V and I_S, given in Table 1 and Section 3.2, are used without tuning.

4 Computational Aspects

4.1 Production of pitch glides

The f_o-glides are simulated by controlling two parameter values dynamically. First, the matrix K is scaled while keeping the matrix M constant as the relative magnitudes of M and K essentially determine the resonance frequencies of vocal fold model (1). This approach is based on the assumption that the vibrating mass and the length of the vocal folds are not significantly changed when the speaker's pitch increases; a reasonable simplification as far as the frequency range is small and register changes are excluded.

The driving pressure p_s is the second parameter used to control the glide. The dependence of f_o on p_s has been observed in simulations [8, 57], physical experiments using upscaled replicas [12], as well as in humans [58] and excised canine larynges [59]. The

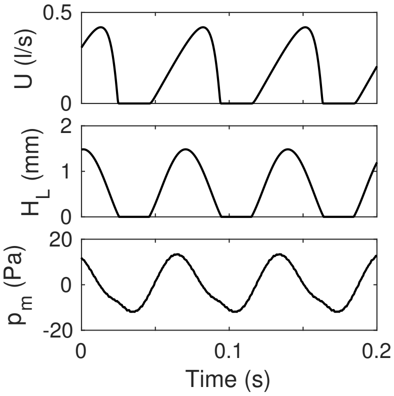

Figure 4: Simulated pulse shapes for [i] with feedback ($Q_{pc} = 0.1$) before the glide begins: glottal flow U, glottal gap H_L, and sound pressure at the lips p_m.

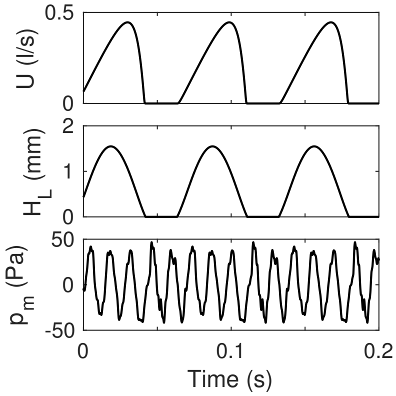

Figure 5: Simulated pulse shapes for [ɑ] with feedback ($Q_{pc} = 0.1$) before glide: glottal flow U, glottal gap H_L, and sound pressure at the lips p_m.

impact of p_s on f_o is, however, secondary in these glides (the f_o trajectories with and without p_s control differ by at most 10%). Instead, p_s is scaled in order to maintain phonation and to prevent large changes in phonation type as the stiffness of the vocal folds changes. It was found by trial and error, that equal scaling of p_s and K best maintained the glottal open quotient OQ (proportion of glottal cycle during which the glottis is open, see [60, Figure 4]), the closing quotient ClQ (proportion of the glottal cycle during which the flow is decreasing), and the maximum of H_L approximately steady over the upward glide when acoustic feedback was disabled.

The parameters are scaled exponentially with time

$$K(t) = 2.2^{2t/T} K^0, \quad p_s(t) = 2.2^{t/T} p_s^0 \tag{18}$$

for rising glides, and

$$K(t) = 2.2^{2-2t/T} K^0, \quad p_s(t) = 2.2^{1-t/T} p_s^0 \tag{19}$$

for falling glides. The duration of the glide is $T = 3\,\mathrm{s}$, and t is the time from the beginning of the glide. Note that the temporal scale of the glides is long compared to glottal cycles, and hence the control parameters K and p_s can be regarded as static from the point of view of the vocal fold dynamics. Other starting conditions (particularly, vocal fold displacements and velocities, and pressure and velocity distributions in the resonators) are taken from stabilised simulations. These parameters produce glides with f_o approximately in the range $[145\,\mathrm{Hz}, 315\,\mathrm{Hz}]$, although the exact range depends on the VT geometry and feedback level.

4.2 Numerical realisation

The model equations are solved numerically using MATLAB software and custom-made code. The vocal fold equations of motion (1) are solved by the fourth order Runge–Kutta time discretisation scheme. The flow equation (6) is solved by the backward Euler method. The VT and SGT are discretised by FEM using piecewise linear elements ($N = 29$ for VT and $N = 10$ for SGT) and the physical energy norm of Webster's equation. Energy preserving Crank–Nicolson time discretisation (i.e., Tustin's method [61]) is used for the resonators. The time step is generally 10 μs which is small enough to keep the frequency warping in Tustin's method under one semitone for frequencies under 13 kHz. Reduced time step, however, is used near glottal closure. This is due to the discontinuity in the aerodynamic force (9) at the closure which requires numerical treatment by interpolation and time step reduction as explained in [31, Section 2.4.1].

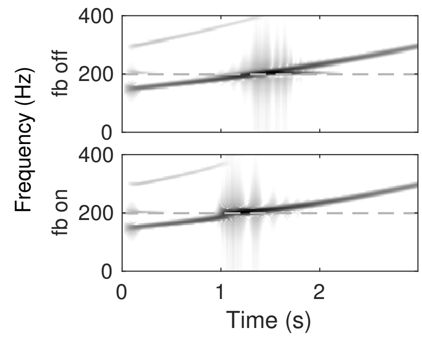

Figure 6: Spectrogram of pressure at lips during glide for [i]. Top: without feedback ($Q_{pc} = 0$). Bottom: with feedback ($Q_{pc} = 0.1$). Dashed gray line is f_{R1}.

Solving the equations of motion of the vocal folds is the computationally most expensive part of the model, taking approximately 55% of the running time in simulations of steady phonation with constant parameter values. In comparison, solving the Webster's equations with precomputed mass, stiffness, and damping matrices takes approximately 10% of the simulation time, and the flow equation solver less than 2%. Simulation of 1 s takes approximately 20 s on a standard professional desktop computer.

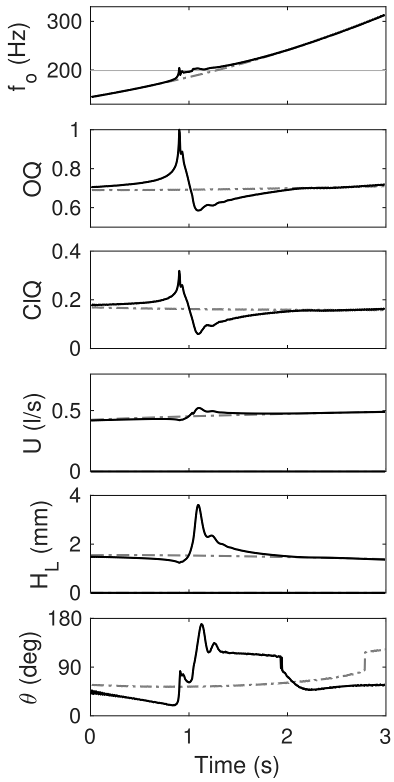

Figure 7: Glide for [i] with feedback ($Q_{pc} = 0.1$) (solid black) and without feedback ($Q_{pc} = 0$) (dashed gray). Shown are fundamental frequency f_o (horizontal gray line is f_{R1}), open quotient OQ, closing quotient ClQ, envelopes of glottal flow U and gap H_L, and phase difference θ between m_{j1} and m_{j2}.

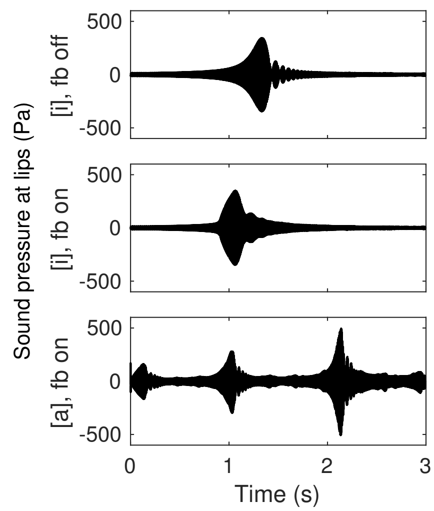

Figure 8: Sound pressures at the lips during upward glides. Top: [i] without feedback ($Q_{pc} = 0$). Middle: [i] with feedback ($Q_{pc} = 0.1$). Bottom: [ɑ] with feedback ($Q_{pc} = 0.1$).

5 Simulation Results

The glottal flow U and gap H_L (or more generally the glottal area hH_L) pulses produced by the model (Figures 4–5) appear realistic when compared to the experimental data presented in [54, Figures 4–7], the signals produced by different numerical models (see [8, Figures 14a–14c], [27, Figures 10–11], [39, Figures 8 and 10], [62, Figure 6], [63, Figure 5]), and the glottal pulse waveforms obtained by inverse filtering in, e.g., [64, Figures 10–13], [60, Figures 3 and 6], and [65, Figures 5.3, 5.4, and 5.17]. Quantitative comparison of the model to the LF model can be found in [66]. The skewing of U relative to H_L – an effect that has been observed in natural speech, e.g., with the help of inverse filtering in [67, 68] – is mainly produced by the inertial term in (6).

The results of upward glide simulations for [i] are shown in Figures 6–7. Figure 6 displays spectrograms of the sound pressure signal at the lips with and without feedback. For Figure 7, the f_o trajectory, OQ, and ClQ have been extracted from U pulse by pulse. Envelopes of U, and H_L are also displayed, and the phase difference θ between m_{j1} and m_{j2} has been estimated based on how much peaks in H_0 are delayed compared to H_L.

The simulations indicate a consistent locking pattern in f_o trajectories at $f_{R1}[i]$ which vanishes if the VT feedback is decoupled by setting $Q_{pc} = 0$. This locking pattern for rising glides can be seen in Figure 6 as a discontinuity in the f_o contour near f_{R1} followed by an interval where f_o appears to be approximately constant. More details are visible in the f_o trajectory in Figure 7: a rapid rise in f_o (hereafter referred to as a jump), a locking to a plateau at approximately f_{R1}, and a smooth release. The height of the jump, degree of overshoot and oscillations about the plateau level, as well as the duration of the locking event depend on parameter choices (see, e.g., Figure 11). In the glide displayed in Figure 7, the f_o trajectories deviate by over 1% in the range 178–215 Hz as measured from feedback free trajectory, and the overshoot at the frequency jump reaches 205 Hz. The flattest part of the locking, which follows the overshoot, occurs at 195-197 Hz.

The frequency jump in the simulations is preceded by a decrease in vocal fold oscillation and glottal flow amplitudes (Figure 7), and a decrease in the phase difference between upper and lower vocal fold masses. This is accompanied by increased breathiness in the phonation, as characterised by increasing OQ and ClQ values, which reduces the effect of the feedback from the acoustics to the vocal folds. The locking plateau coincides with a nearly constant rate of decreasing OQ and ClQ, and increasing amplitude of, in particular, H_L. At the same time, there are large but smooth changes in θ. After the release of f_o the glottal pulse characteristics return gradually to the feedback free trajectories, except for θ. The sudden

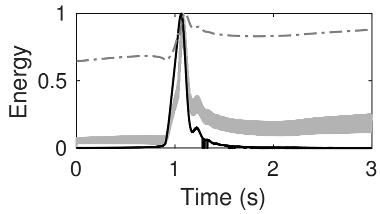

Figure 9: Normalised envelope of energy in VT acoustics (solid black) and in the glottal flow U (dashed gray), and energy in vocal fold vibrations (solid gray) in upward glide for [i] with $Q_{pc} = 0.1$.

changes in θ seen at 1.9 s with feedback and at 2.8 s without feedback are caused by the method of estimating θ. Near these instants H_0 pulses have shallow double peaks, and the sudden change occurs when the dominant peak shifts from one to the other. Note, however, that changes in pulse shapes are smooth near these instants. Further, H_0 and H_L have well defined single peaks at and near the locking event, so changes in θ there are not caused by this same phenomenon.

This locking behaviour of f_o or the related waveform changes are not observed for glides of [ɑ] where f_{R1}[ɑ] is not inside the simulated frequency range $[145\,\mathrm{Hz}, 315\,\mathrm{Hz}]$. The differences in the f_o trajectories and glottal pulse characteristics between feedback ($Q_{pc} = 0.1$) and feedback free ($Q_{pc} = 0$) configurations are negligible for [ɑ].

The VT resonance $f_{R1}[i]$ and the resonance fractions f_{R1}[ɑ]$/5 = 148\,\mathrm{Hz}$, f_{R1}[ɑ]$/4 = 186\,\mathrm{Hz}$ and f_{R1}[ɑ]$/3 = 247\,\mathrm{Hz}$ are within the frequency range, and the corresponding events are visible in the sound pressure signal at the lips (Figure 8). Note that despite this visibility, corresponding events can be seen in the glottis only for the event in the middle panel, i.e. $f_{R1}[i]$ with feedback. For [ɑ], the pressure signals with and without feedback are nearly identical (only glide with feedback is shown in Figure 8). For [i], the largest difference in the pressures is the timing of the resonance event.

When feedback is disabled, energy cannot be transferred from the resonating vocal tract to the oscillating vocal folds or to the glottal flow. Figure 9 shows how energy, normalised to one, in each of the subsystems develops when feedback is on. As the resonance nears, p_c does work on the vocal folds increasing the energy in the vocal fold oscillations which in turn feeds energy into U. Since p_c has an increasingly strong periodic component at $f_{R1}[i]$, all three subsystems get locked to this frequency. Unlocking occurs when the first vocal fold eiqenfrequency has been raised sufficiently for the energy in the oscillations to win out over the frequency of p_c.

Rising and falling glides show different perturbation patterns as shown in Figure 10. The x-axis in this figure is the relative vocal fold stiffness, which for rising glides is $2.2^{t/T}$ and for falling glides $2.2^{1-t/T}$

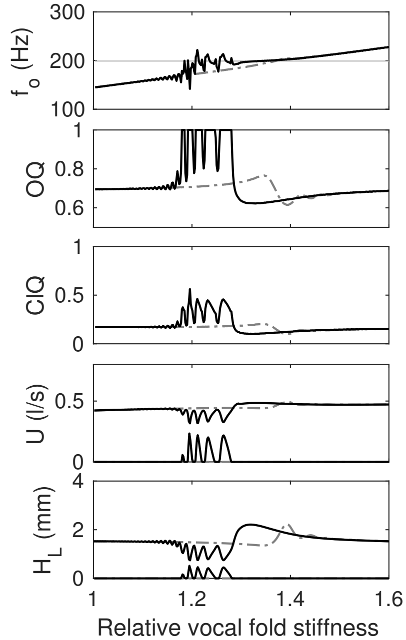

Figure 10: Upward (dashed gray) and downward (solid black) glides for vowel [i] with $Q_{pc} = 0.04$. Shown are fundamental frequency f_o (f_{R1} indicated by horizontal gray line), open quotient OQ, closing quotient ClQ, and the envelopes of glottal flow U and gap H_L. On the x-axis, relative vocal fold stiffness refers to the coefficient of the K^0 matrix in equations (18) and (19).

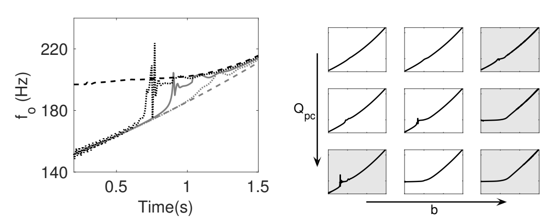

Figure 11: Left: f_o trajectories for [i] with different values of Q_{pc}: gray dashed 0.0, gray dotted 0.05, gray solid 0.1, black dotted 0.15, and black dashed 0.2. Right: f_o trajectories for [i] qualitatively as Q_{pc} and β increase in the direction of the arrow. Light gray background indicates that small parameter changes can lead to loss of quasi-stable glides.

as given in equations (18) and (19). For given model parameter values, falling glides exhibit more fluctuations in glottal pulse parameters at the locking event and the perturbation lasts longer. The fluctuations in f_o in the falling glides during the locking and at frequencies below this are qualitatively similar to what occurs at extreme values of Q_{pc} and β for rising glides.

The feedback parameter Q_{pc} plays, unsurprisingly, a key role in the f_o jump and locking in glides for [i] as shown in Figure 11 (left). With no acoustic feedback to the vocal folds, there are no perturbations in f_o, whereas with a high Q_{pc}, starting a glide with f_o below f_{R1} is not possible without decreasing K^0. If a starting f_o below f_{R1} is obtainable, a high Q_{pc} value results in a large overshoot at the jump, and fluctuations in f_o both before the jump and at the beginning of the plateau.

Besides Q_{pc}, the locking pattern is also sensitive to other model parameters, in particular the vocal fold damping β. In fact, β and Q_{pc} affect the locking behaviour in complementary ways, as qualitatively shown in Figure 11 (right). The full frequency range $[145\,\text{Hz}, 315\,\text{Hz}]$ for f_o can be obtained with modal locking if $Q_{pc} \in [0.05, 0.12]$ and $\beta \in [0.005, 0.015]$. Beyond these ranges, an increase in one parameter needs to be compensated for with a decrease in the other. Otherwise, the locking pattern disappears or the simulated f_o range is reduced to above f_{R1}[i].

The stability of glide simulations (understood as slowly changing amplitude envelope of glottal flow U) becomes a serious issue at high values of one or both of the parameters Q_{pc} and β. The driving pressure p_s in glide simulations is dynamically controlled as given in equations (18)–(19). If p_s were instead kept constant, we would observe an increasing OQ and decreasing amplitudes of glottal flow and vocal fold oscillations throughout the glide but the qualitative behaviour of modal locking events, including the behaviour of phonation type parameters around these events, would remain very similar.

6 Discussion

We have reported observations on the locking of f_o at a resonance of the VT in simulated pitch glides. The locking behaviour shows a consistent time-dependent behaviour that is similar for rising and falling glides. The f_o jump at the beginning of the locking in rising glides and end of the locking in falling glides occurs together with and increased breathiness of phonation as characterised by open quotient OQ and closing quotient ClQ. During the locking plateau, these parameters indicated an approximately steady decrease in breathiness.

The locking takes place only at frequencies determined by supraglottal resonances. Use of p_s as a secondary control parameter for the glides ensure that the main cause for changes in OQ and ClQ is the acoustic loading. By modifying the strength of the acoustic feedback (i.e., the parameter Q_{pc} in equation (15)) and vocal fold tissue losses (i.e., the parameter β), the locking tendency at f_{R1}[i] may be modulated from non-existent (where both Q_{pc} and β have low values) to extreme locking at f_{R1}[i] without release (where Q_{pc} and/or β have large values); see Figure 11. Small changes to the model (as discussed below) leave the locking behaviour at f_{R1}[i] unchanged, even though the model parameter values required for the desired glottal waveform change (cf. [28, 29]). We conclude that the simulation results on vowel glides reported in Section 5 reflect the model behaviour in a consistent and robust manner.

To what extent do the simulation results validate the proposed model? The model produces perturbations of the glottal pulses at VT resonances and, additionally, sound pressure perturbations at some of the VT resonance fractions. Of the former, a wide existing literature was reviewed in Section 1. Observations on perturbations in speech at formant fractions have not been reported, to our knowledge, in experimental literature. There is a particular temporal pattern of locking in simulated perturbations at f_{R1}[i] as shown in Figures 6 and 7 (topmost panel). A similar pattern can be seen in the speech spectrograms given in [17, Figure 5], [16, Figure 4], as well as in the vowel glide samples in the data set of [3]. The pitch trajectory and speech spectrogram in [19, Figure 4] also show locking but no release. A similar locking behaviour can also be interpreted to lie behind the experimental results shown in [12, Figures 10b and 13b], and it also tends to emerge in model simulations even if the acoustic feedback is realised in different manner; see, e.g., [14, Figures 13 and 14] and [69, Figure 6].

6.1 Acoustics

The effect of physically realistic values of parameter α in model simulations is negligible; see [25, Section 5] and [30, Section 3.3.2]. These losses move the VT resonance positions computed from equations (10) slightly. On the other hand, the VT resonances are quite sensitive to the parameters of the parallel RL model in equation (13), similar to the simplified model proposed in [45, Eq. (28)]. In its most general form, the model in [45, Eq. (39)] is an integro-differential delay equation with nine parameters and a single delay lag. Unfortunately, it cannot be introduced to Webster's model as a boundary condition: this is the salient feature of equation (13) that simplifies the implementation of the FEM solver.

It is expected that the otherwise small subglottal effect in simulations will get more pronounced when $f_o \to f'_{R1}$, and similarly VT impact for [ɑ] will increase when $f_o \to f_{R1}$[ɑ]. These resonance frequencies, as well as the fractions $f_{R1}[\text{i}]/n$, $n = 2, 3, ...$, are not included in the glides because the two glide controls appear to be insufficient to maintain phonation through such a large frequency range. Such glides would likely require dynamic control of vocal fold length and mass as well. The similarity of the VT and SGT resonators is visible near the resonances fractions in the presented glides, however: The first subglottal resonance fraction $f'_{R1}/2$ shows up in the counter pressure (15) in the same way as $f_{R1}[\text{ɑ}]/n$.

The SGT acoustics model proposed in [27] is likely to produce the correct resonance distribution and frequency-dependent energy dissipation rate at the lung end without tuning. The horn model requires tuning of the horn geometry in order to get the lowest subglottal resonance realistic $f'_{R1} = 500\,\mathrm{Hz}$. Doing so freezes all the higher subglottal resonances at fixed positions, e.g., $f'_{R2} = 1.0\,\mathrm{kHz}$. The branching subglottal models given in [27, Figure 8] have the second subglottal resonance between 1.3 kHz and 1.5 kHz. It was observed in [70] that the soft tissues introduce an additional nonacoustic resonance to the subglottal system that is lower than the first subglottal formant f'_{R1} attributed to air column dynamics. There is no obvious way how a horn model could be used to accommodate such a resonance at ca. 350 Hz due to the yielding wall dynamics.

6.2 Vocal folds and glottal flow

The vocal fold geometry shown in Figure 1 (top) leads to a simple expression for the aerodynamic force in equation (9). The further simplification of keeping the direction of $p(x,t)$ constant (i.e., considering changes in ϕ negligible) is possible without affecting the qualitative behaviour of the model. The difference between the driving pressure p_s and the reference pressure p_r can be included in the force balance when the glottis is closed (equation (5)) although the wedge-shaped vocal folds, their point-like collision, and the assumption of incompressible glottal flow lead to overestimation of the effect. This addition causes an increase in the open quotient throughout simulations, but if the model parameters are adjusted to achieve a phonation similar to Figures 4–5 before the glides, the locking behaviour remains qualitatively unchanged.

Replacing the sharp peaks by flat tops in Figure 1 results in phonation that has typically lower open quotient (OQ) compared to the original wedge-like geometry. This change makes it easier to adjust the parametrisation of the model to obtain some phonation targets. In particular, the value of the glottal loss parameter k_g can then be based on experimental values (e.g., [41]) since the model geometry becomes more similar to the experimental model geometry (M5).

The importance of entrance and exit effects represented by k_g can be seen, for example, by comparing simulated volume velocities and glottal areas with the experimental curves in [40, Figure 3], obtained from a physical model of the glottis. In model simulations, leaving out this transglottal pressure loss term changes the glottal pulse waveform significantly if other model parameters are kept the same, as shown in [30, Figure 3.7]. About half of the total pressure loss in simulations is due to entrance and exit effects at the peak of opening of the glottis; see [30, Figure 3.6]. However, the behaviour of the simulated f_o trajectories over f_{R1}[i] does not change if $k_g = 0$. Then, however, the vowel glide must be produced by different model parameter values.

The glottal flow has been studied extensively since 1950's. Compared to the flow model used here, physiologically more faithful glottal flow solvers have been proposed in, e.g., [35, 46, 62, 71, 72] and [73]. As pointed out in [72], more sophisticated flow models are challenging to couple to acoustic resonators since the interface between the flow-mechanical (in particular, the turbulent) and the acoustic components is no longer clearly defined.

Direct feedback from VT acoustics to the glottal flow can be added to the model although it has been left outside the scope of this work. In implementing this feedback mode, particular care must be taken to remove the additional acoustic contribution in the inertial effect, which is already accounted for by (6). The impact of this feedback mechanism is expected to be notable around the f_o jump, when the glottal closure is short or non-existent.

Turbulence in supraglottal space is a spatially distributed acoustic source, and it does not provide a spatially localised acoustic signal for the resonator in equation (12). Much of the turbulence noise energy lies above 4 kHz where Webster's model equation (10) is not an accurate description [74, 75]. The unmodelled supraglottal jet may even exert an additional aerodynamic force to vocal folds that would not be part of the acoustic counter pressure p_c.

7 Conclusions

We have presented a model for vowel production, based on (partial) differential equations, that consists of submodels for glottal flow, vocal folds oscillations, and acoustic responses of the VT and SGT cavities. The model was used for simulations of rising and falling vowel glides of [ɑ, i] in frequencies that span one octave [145 Hz, 315 Hz]. This interval contains the lowest VT resonance f_{R1} of [i] but not that of [ɑ]. Perturbation events in simulated vowel glides were observed at VT acoustic resonances, or at some of their fractions but nowhere else.

The fundamental frequency f_o of the simulated vowel was observed to lock to f_{R1}[i] but similar locking was not seen at any of the resonance fractions of [ɑ]. The locking events were accompanied by changes in the phonation: increased breathiness below and partially at the locking frequency and steady change in breathiness during most of the lock. If these changes can also be detected in glides produced by human speakers, e.g., by using electroglottography, they may provide an indirect means of identifying locking events when coincidence of f_o and f_{R1} makes it challenging to track them both.

The locking event takes place only when the acoustic feedback from VT to vocal folds is present, and then it has a characteristic time-dependent behaviour. A large number of simulation experiments were carried out with different parameter settings of the model to verify the robustness and consistency of all observations. The similarity between simulated pitch pattern and experimental results in literature was achieved by using feedback from acoustics to vocal fold tissues, indicating that this feedback mode can be strong enough to affect speech outcomes.

The simulation model does not include the neural control actions on the vocal fold structures or dynamic modifications of the VT geometry. There is also a significant control action affecting the driving stagnation pressure and it has been used as a control variable in equations (18)–(19) for glide productions. In humans, neural control actions are part of feedback loops, of which some are auditive, and some others operate directly through tissue innervation and the central nervous system. So little is known about these feedback mechanisms that their explicit mathematical modelling seems infeasible. Instead, the model parameters for simulations are tuned so that the simulated glottal pulse waveform corresponds to experimental speech data. Despite these simplifications the model appears to be sufficiently detailed to replicate the observations found in literature.

Acknowledgements

This study was funded by the Academy of Finland (projects no. 284671 and 312490), the Finnish graduate school in engineering mechanics, Finnish Academy project Lastu 135005, 128204, and 125940; European Union grant Simple4All (grant no. 287678), Aalto Starting Grant 915587, and Åbo Akademi Institute of Mathematics. The authors would like to thank the four anonymous reviewers in 2013 and 2016 for comments leading to many improvements of the model.

References

[1] T. Chiba, M. Kajiyama: The vowel, its nature and structure. Phonetic Society of Japan, Tokyo, 1941.

[2] G. Fant: Acoustic theory of speech production. Mouton, The Hague, 1960.

[3] D. Aalto, J. Malinen, M. Vainio, Modal locking between vocal fold and vocal tract oscillations: Experiments and statistical analysis. ArXiv e-prints, arXiv:1211.4788, 2016.

[4] I. R. Titze, R. J. Baken, K. W. Bozeman, S. Granqvist, N. Henrich, C. T. Herbst, D. M. Howard, E. J. Hunter, D. Kaelin, R. D. Kent, J. Kreiman, M. Kob, A. Löfqvist, S. McCoy, D. G. Miller, H. Noé, R. C. Scherer, J. R. Smith, B. H. Story, J. G. Švec, S. Ternström, J. Wolfe: Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. The Journal of the Acoustical Society of America **137** (2015) 3005–3007.

[5] P. Alku, J. Horáček, M. Airas, F. Griffond-Boitier, A.-M. Laukkanen: Performance of glottal inverse filtering as tested by aeroelastic modelling of phonation and FE modelling of vocal tract. Acta Acustica united with Acustica **92** (2006) 717–724.

[6] P. Alku, J. Pohjalainen, M. Vainio, A.-M. Laukkanen, B. H. Story: Formant frequency estimation of high-pitched vowels using weighted linear prediction. The Journal of the Acoustical Society of America **134** (2013) 1295–1313.

[7] J. Guðnason, D. D. Mehta, T. F. Quatieri: Evaluation of speech inverse filtering techniques using a physiologically based synthesizer. Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal (ICASSP), April 2015, 4245–4249.

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

[9] S. F. Austin, I. R. Titze: The effect of subglottal resonance upon vocal fold vibration. Journal of Voice **11** (1997) 391–402.

[10] Z. Zhang, J. Neubauer, D. A. Berry: The influence of subglottal acoustics on laboratory models of phonation. The Journal of the Acoustical Society of America **120** (2006) 1558–1569.

[11] J. C. Lucero, K. G. Lourenço, N. Hermant, A. Van Hirtum, X. Pelorson: Effect of source-tract acoustical coupling on the oscillation onset of the vocal folds. The Journal of the Acoustical Society of America **132** (2012) 403–411.

[12] N. Ruty, X. Pelorson, A. Van Hirtum, I. Lopez-Arteaga, A. Hirschberg: An in-vitro setup to test the relevance and the accuracy of low-order vocal folds models. The Journal of the Acoustical Society of America **121** (2007) 479–490.

[13] N. Ruty, X. Pelorson, A. Van Hirtum: Influence of acoustic waveguides lengths on self-sustained oscillations: Theoretical prediction and experimental validation. The Journal of the Acoustical Society of America **123** (2008) 3121–3121.

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

[15] H. Hatzikirou, W. T. Fitch, H. Herzel: Voice instabilities due to source-tract interactions. Acta Acustica united with Acustica **92** (2006) 468–475.

[16] I. T. Tokuda, M. Zemke, M. Kob, H. Herzel: Biomechanical modeling of register transitions and the role of vocal tract resonators. The Journal of the Acoustical Society of America **127** (2010) 1528–1536.

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

[18] M. Zañartu, D. D. Mehta, J. C. Ho, G. R. Wodicka, R. E. Hillman: Observation and analysis of in vivo vocal fold tissue instabilities produced by nonlinear source-filter coupling: A case study. The Journal of the Acoustical Society of America **129** (2011) 326–339.

[19] L. Wade, N. Hanna, J. Smith, J. Wolfe: The role of vocal tract and subglottal resonances in producing vocal instabilities. The Journal of the Acoustical Society of America **141** (2017) 1546–1559.

[20] K. L. Kelly, C. C. Lochbaum: Speech synthesis. Proceedings of the Fourth International Congress on Acoustics, 1962, Paper G42, 1–4.

[21] H. K. Dunn: The calculation of vowel resonances, and an electrical vocal tract. The Journal of the Acoustical Society of America **22** (1950) 740–753.

[22] S. El-Masri, X. Pelorson, P. Saguet, P. Badin: Development of the transmission line matrix method in acoustics. Applications to higher modes in the vocal tract and other complex ducts. International Journal of Numerical Modelling: Electronic Networks, Devices and Fields **11** (1998) 133–151.

[23] J. Mullen, D. Howard, D. Murphy: Waveguide physical modeling of vocal tract acoustics: Flexible formant bandwith control from increased model dimensionality. IEEE Transactions on Audio, Speech, and Language Processing **14** (2006) 964–971.

[24] S. Rienstra, A. Hirschberg: An introduction to acoustics. Eindhoven University of Technology, 2013.

[25] K. van den Doel U. M. Ascher: Real-time numerical solution of Webster's equation on a nonuniform grid. IEEE Transactions on Audio, Speech, and Language Processing **16** (2008) 1163–1172.

[26] J. Horáček, V. Uruba, V. Radolf, J. Veselỳ, V. Bula: Airflow visualization in a model of human glottis near the self-oscillating vocal folds model. Applied and Computational Mechanics **5** (2011) 21–28.

[27] J. C. Ho, M. Zañartu, G. R. Wodicka: An anatomically based, time-domain acoustic model of the subglottal system for speech production. The Journal of the Acoustical Society of America **129** (2011) 1531–1547.

[28] T. Murtola, J. Malinen: Waveform patterns in pitch glides near a vocal tract resonance. Proceedings of INTERSPEECH 2017, Stockholm, 2017, 3487-3491

[29] A. Aalto, T. Murtola, J. Malinen, D. Aalto, M. Vainio: Modal locking between vocal fold and vocal tract oscillations: Simulations in time domain. ArXiv e-prints, arXiv:1506.01395, 2017.

[30] T. Murtola: Modelling vowel production. Licentiate thesis, Aalto University School of Science, Espoo, Finland, 2014.

[31] A. Aalto: A low-order glottis model with nonturbulent flow and mechanically coupled acoustic load. Master's thesis, Helsinki University of Technology, Espoo, Finland, 2009.

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

[33] J. Liljencrants: A translating and rotating mass model of the vocal folds. STL-QPSR **32** (1991) 1–18.

[34] N. J. C. Lous, G. C. J. Hofmans, R. N. J. Veldhuis, A. Hirschberg: A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design. Acta Acustica united with Acustica **84** (1998) 1135–1150.

[35] 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. Application to a modified two-mass model. The Journal of the Acoustical Society of America **96** (1994) 3416–3431.

[36] B. H. Story, I. R. Titze: Voice simulation with a body-cover model of the vocal folds. The Journal of the Acoustical Society of America **97** (1995) 1249–1260.

[37] B. D. Erath, M. Zañartu, K. C. Stewart, M. W. Plesniak, D. E. Sommer, S. D. Peterson: A review of lumped-element models of voiced speech. Speech Communication **55** (2013) 667–690.

[38] P. Birkholz: A survey of self-oscillating lumped-element models of the vocal folds. In: B. J. Kröger, P. Birkholz, eds., Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung, 2011, 47–58.

[39] M. Zañartu, L. Mongeau, G. R. Wodicka: Influence of acoustic loading on an effective single mass model of the vocal folds. The Journal of the Acoustical Society of America **121** (2007) 1119–1129.

[40] J. van den Berg, J. T. Zantema, P. Doornenbal: On the air resistance and the Bernoulli effect of the human larynx. Journal of the Acoustical Society of America **29** (1957) 626–631.

[41] L. P. Fulcher, R. C. Scherer, T. Powell: Pressure distributions in a static physical model of the uniform glottis: Entrance and exit coefficients. The Journal of the Acoustical Society of America **129** (2011) 1548–1553.

[42] T. Lukkari, J. Malinen: Webster's equation with curvature and dissipation. ArXiv e-prints, arXiv:1204.4075, 2013.

[43] A. Aalto, T. Lukkari, J. Malinen: Acoustic wave guides as infinite-dimensional dynamical systems. ESAIM: Control, Optimisation and Calculus of Variations **21** (2015) 324–347.

[44] T. Lukkari, J. Malinen: A posteriori error estimates for Webster's equation in wave propagation. Journal of Mathematical Analysis and Applications **427** (2015) 941–961.

[45] T. Hélie, X. Rodet: Radiation of a pulsating portion of a sphere: application to horn radiation. Acta Acustica united with Acustica **89** (2003) 565–577.

[46] P. Birkholz, D. Jackel, B. Kröger: Simulation of losses due to turbulence in the time-varying vocal system. IEEE Transactions on Audio, Speech, and Language Processing **15** (2007) 1218–1226.

[47] D. Aalto, O. Aaltonen, R.-P. Happonen, P. Jääsaari, A. Kivelä, J. Kuortti, J.-M. Luukinen, J. Malinen, T. Murtola, R. Parkkola, J. Saunavaara, T. Soukka, M. Vainio: Large scale data acquisition of simultaneous MRI and speech. Applied Acoustics **83** (2014) 64–75.

[48] B. Story, I. Titze, E. Hoffman: Vocal tract area functions from magnetic resonance imaging. The Journal of the Acoustical Society of America **100** (1996) 537–554.

[49] B. H. Story, I. R. Titze, E. A. Hoffman: Vocal tract area functions for an adult female speaker based on volumetric imaging. The Journal of the Acoustical Society of America **104** (1998) 471–487.

[50] B. H. Story, I. R. Titze: Parameterization of vocal tract area functions by empirical orthogonal modes. Journal of Phonetics **26** (1998) 223–260.

[51] A. Kivelä: Acoustics of the Vocal Tract: MR image segmentation for modelling. Master's thesis, Aalto University School of Science, Espoo, Finland, 2015.

[52] A. Ojalammi, J. Malinen: Automated segmentation of upper airways from MRI: Vocal tract geometry extraction. Proceedings of BIOIMAGING 2017, 2017, 77–84

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

[54] B. Cranen, L. Boves: Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production. The Journal of the Acoustical Society of America **77** (1985) 1543–1551.

[55] B. Cranen, L. Boves: On subglottal formant analysis. The Journal of the Acoustical Society of America **81** (1987) 734–746.

[56] I. R. Titze: Physiologic and acoustic differences between male and female voices. The Journal of the Acoustical Society of America **85** (1989) 1699–1707.

[57] D. Scimarella, C. d'Alessandro: On the acoustic sensitivity of a symmetric two-mass model of the vocal folds to the variation of control parameters. Acta Acustica united with Acustica **90** (2004) 746–761.

[58] P. Lieberman, R. Knudson, J. Mead: Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation. The Journal of the Acoustical Society of America **45** (1969) 1537–1543.

[59] I. R. Titze: On the relation between subglottal pressure and fundamental frequency in phonation. The Journal of the Acoustical Society of America **85** (1989) 901–906.

[60] P. Alku: Glottal inverse filtering analysis of human voice production - a review of estimation and parameterization methods of the glottal excitation and their applications. Sadhana **36** (2011) 623–650.

[61] V. Havu, J. Malinen: The Cayley transform as a time discretization scheme. Numerical Functional Analysis and Optimization **28** (2007) 825–851.

[62] I. R. Titze: Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model. The Journal of the Acoustical Society of America **111** (2002) 367–376.

[63] I. R. Titze: Parameterization of the glottal area, glottal flow, and vocal fold contact area. The Journal of the Acoustical Society of America **75** (1984) 570–580.

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

[65] H. Pulakka: Analysis of human voice production using inverse filtering, high-speed imaging, and electroglottography. Master's thesis, Helsinki University of Technology, Espoo, Finland, 2005.

[66] A. Aalto, P. Alku, J. Malinen: A LF-pulse from a simple glottal flow model. Proceedings of the 6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA2009), Florence, 2009, 199–202.

[67] M. Berouti, D. Childers, A. Paige: Glottal area versus glottal volume-velocity. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP '77 (1977) **2** 33–36.

[68] S. Granqvist, S. Hertegård, H. Larsson, J. Sundberg: Simultaneous analysis of vocal fold vibration and transglottal airflow: exploring a new experimental setup. Journal of Voice **17** (2003) 319–330.

[69] N. Ruty, X. Pelorson, A. van Hirtum: Influence of acoustic waveguides lengths on self-sustained oscillations: Theoretical prediction and experimental validation. Proceedings of Acoustics '08, Paris, June 29-July 4, 2008, 1243–1247.

[70] S. M. Lulich, H. Arsikere: Tracheo-bronchial soft tissue and cartilage resonances in the subglottal acoustic input impedance. Journal of the Acoustical Society of America **137** (2015) 3436–3446.

[71] B. D. Erath, S. D. Peterson, M. Zañartu, G. R. Wodicka, M. W. Plesniak: A theoretical model of the pressure field arising from asymmetric intraglottal flows applied to a two-mass model of the vocal folds. The Journal of the Acoustical Society of America **130** (2011) 389–403.

[72] P. Punčochářová-Pořízková, K. Kozel, J. Horáček, J. Fürst: Numerical simulation of unsteady compressible low Mach number flow in a channel. Engineering Mechanics **17** (2010) 83–97.

[73] P. Šidlof, J. Horáček, V. Řidký: Parallel CFD simulation of flow in a 3D model of vibrating human vocal folds. Computers & Fluids **80** (2013) 290–300.

[74] T. Vampola, J. Horáček, A.-M. Laukkanen, J. G. Švec: Human vocal tract resonances and the corresponding mode shapes investigated by three-dimensional finite-element modelling based on CT measurement. Logopedics Phoniatrics Vocology **40** (2013) 1–10.

[75] T. Vampola, A.-M. Laukkanen, J. Horáček, J. G. Švec: Finite element modelling of vocal tract changes after voice therapy. Applied and Computational Mechanics **5** (2011) 77–88.