HELSINKI UNIVERSITY OF TECHNOLOGY
Faculty of Information and Natural Sciences

Atte Aalto

A Low-order Glottis Model with Nonturbulent Flow and Mechanically Coupled Acoustic Load

Master's Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology.

Espoo, March 6, 2009

Supervisor:	Professor Timo Eirola
Instructor:	Dr. Jarmo Malinen

TEKNILLINEN KORKEAKOULU	DIPLOMITYÖN TIIVISTELMÄ

Informaatio- ja luonnontieteiden tiedekunta

Tekijä:	Atte Aalto
Koulutusohjelma:	Teknillisen fysiikan ja matematiikan koulutusohjelma
Pääaine:	Matematiikka
Sivuaine:	Systeemitekniikka
Työn nimi:	Akustiseen kuormaan mekaanisesti kytketty matalan kertaluvun glottismalli pyörteettömällä virtauksella
Title in English:	A low-order glottis model with nonturbulent flow and mechanically coupled acoustic load
Professuuri:	Mat-1 Matematiikka
Työn valvoja:	Prof. Timo Eirola
Työn ohjaaja:	TkT Jarmo Malinen

Työssä muodostetaan matalan kertaluvun massajousimalli ihmisen äänihuulista. Äänihuulien värähtelyn aiheuttaa Bernoullin lain mukainen virtaus, jossa virtausnopeuden kasvu kapeassa raossa aiheuttaa paineen alenemisen. Tämä johtaa glottiksen (äänihuulien välinen rako) sulkeutumiseen. Suljetun glottiksen läpi ei luonnollisesti ole virtausta, jolloin kudosvoimat palauttavat äänihuulet jälleen irti toisistaan.

Virtausnopeus glottiksen läpi saadaan erillisestä yksiulotteisesta kokoonpuristumattomasta virtausmallista. Malli ottaa huomioon ääntöväylän ilmamassan inertian sekä viskoosin painehäviön ääntöväylässä ja glottiksessa. Painehäviön muutos äänihuulien liikkuessa säätelee virtauksen nopeutta ja glottiksen sulkeutuessa pysäyttää virtauksen kokonaan.

Tämän mallin ulostulo syötetään ääntöväylämalliin, joka on siirtolinjatyyppinen Websterin yhtälömalli kaarevalle putkelle. Tämän yhtälön ratkaisusta luetaan ilmanpainetta äänihuulien yläpuolella. Tämä paine syötetään takaisin äänihuulimalliin, jossa se näkyy terminä liikeyhtälöiden kuormavoimassa. Yksi työn tavoitteista on tutkia tämän takaisinkytkennän vaikutusta glottiksen toimintaan ja todentaa muiden tutkijoiden tuloksia.

Websterin yhtälössä käytetty ääntöväylädata on peräisin muiden tutkijoiden suorittamista anatomisista magneettiresonanssikuvauksista. Datasta on määritetty Websterin yhtälössä tarvittavat ääntöväylän halkileikkauksen pinta-alat sekä ääntöväylän kaarevuus. Laskennassa ääntöväylä on diskretoitu elementtimenetelmällä käyttäen Websterin yhtälön luonnollista energianormia.

Sivumäärä: 46 + 11	Avainsanat: puheentuotto, glottismalli, Bernoulli-virtaus, virtauksen indusoima värähtely, Websterin yhtälö
Hyväksytty:	Kirjasto:

HELSINKI UNIVERSITY OF TECHNOLOGY　　ABSTRACT OF MASTER'S THESIS

Faculty of Information and Natural Sciences

Author:	Atte Aalto
Degree Programme:	Engineering Physics and Mathematics
Major subject:	Mathematics
Minor subject:	Systems technology
Title:	A low-order glottis model with nonturbulent flow and mechanically coupled acoustic load
Title in Finnish:	Akustiseen kuormaan mekaanisesti kytketty matalan kertaluvun glottismalli pyörteettömällä virtauksella
Chair:	Mat-1 Mathematics
Supervisor:	Prof. Timo Eirola
Instructor:	Dr. Jarmo Malinen

A low order mass-spring model of human vocal cords is constructed. The vibration of the cords is caused by a Bernoulli-type flow. This means that increasing flow velocity in a narrowing causes a drop in the pressure finally leading to closure of the glottis (the orifice between the vocal cords). When the glottis is closed, there is no flow and the forces of the tissues push the cords apart again.

The flow velocity through the glottis is obtained from a separate one-dimensional incompressible flow model. The inertia of the air contained in the vocal tract is taken into account as well as viscous pressure losses in the vocal tract and the glottis. Changes in the pressure loss due to vocal cords' movement regulate the glottal flow velocity and, when the glottis closes, stop the flow altogether.

The output of this model is fed to the vocal tract model, which is a transmission line model given by the Webster's equation for a curved tube. The solution of the Webster's equation gives us the air pressure distribution in the vocal tract. The air pressure at the glottis end is fed back to the glottis model, where it appears as a term in the load force for the equations of motion. One of the purposes of this work is to study the effect of this feedback on the glottal behaviour and verify results obtained by other researchers.

The vocal tract data used in the Webster's equation is obtained by magnetic resonance imaging (MRI) performed by other researchers on human subjects. The vocal tract cross sectional area and the tract curvature are determined from such MRI data. For computations, the vocal tract was discretized using the Finite Element Method based on the natural energy norm of the Webster's equation.

Number of pages: 46 + 11	Keywords: speech production, glottis model, Bernoulli flow, flow induced vibrations, Webster's equation
Approved:	Library code:

Acknowledgements

First of all, I want to thank my instructor Jarmo Malinen for guiding me through this task. His guidelines for the work have been instructive, yet flexible enough allowing me to experiment and carry out own ideas. I also thank Pertti Palo for helping me with many things along the way. He has also provided me with valuable comments and advices concerning both the content of this work and research work in general. Of course I want to thank my supervisor, Professor Timo Eirola.

I would also like to thank Professor Paavo Alku from the TKK Acoustics Laboratory for taking some time to read and comment this work. He also provided us with some data, which would had been rather laborous for us to collect ourselves.

Last but not least I would like to thank my fellow workers at the Institute of Mathematics for pleasant company.

Otaniemi, March 6, 2009

Atte Aalto

Contents

Chapter 1

Introduction

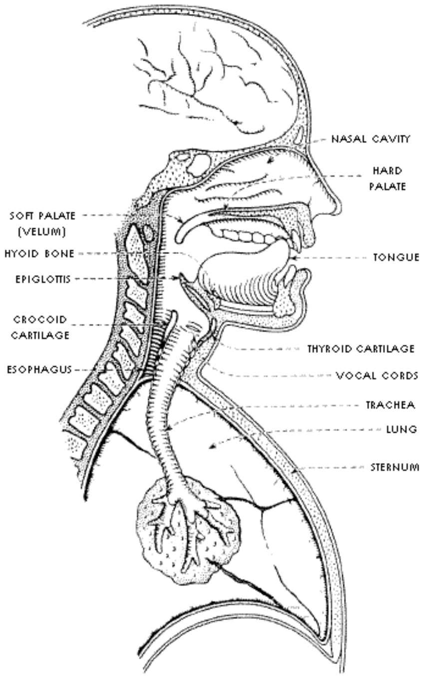
1.1 Human voice production

Figure 1.1 shows a schematic diagram of the human vocal mechanism. In a simplified model of human voice production, lungs can be considered as a large air reservoir in constant pressure. This pressure is caused by the breathing muscles contracting the lungs. The air escapes from the lungs through a channel consisting of two parts, the trachea and the vocal tract (VT). These parts are separated by a slit-like narrowing, formed by vocal cords. The orifice between the cords is called the glottis. At the other end, the vocal tract is terminated by the lips. The voice also has a secondary transmission channel, namely the nasal tract diverging from the VT at velum and ending at the nostrils. The velum opening regulates the influence of the nasal coupling.

In the basic configuration the voice is generated by the flow induced vibrations of the vocal cords which act like a valve, periodically opening and closing the glottis, and thereby generate short flow pulses. This oscillation occurs because the cords have no (stable) equilibrium states for a flow exceeding a certain value, known as the *phonation threshold*. When the glottis is closed, there is a transglottal pressure difference, that will eventually force the glottis open. After the glottis opens, the flow accelerates and — due to Bernoulli effect — the local pressure at the glottis drops. The pressure drop sucks the vocal cords together again. The glottal flow pulses excite the acoustics of the air column in the vocal tract. The acoustic voice signal is filtered by the vocal tract and the sound signal is eventually transmitted to the exterior space through the mouth and/or the nostrils.

1.2 Speech sounds

The geometry of the VT varies during phonation due to the movement of the *articulators*, of which the most important ones are the lips, jaw, tongue and velum. Let us briefly introduce the production mechanisms of typical speech



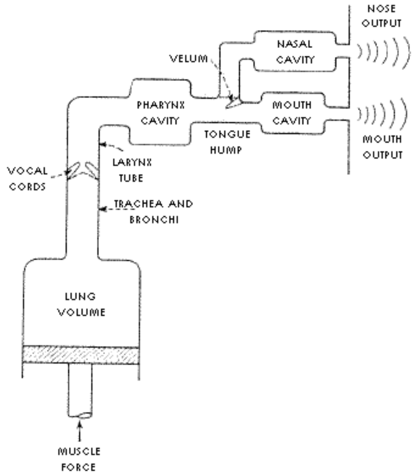


Figure 1.1: Schematic diagrams of the human voice mechanism and functional components of the vocal tract by Flanagan (1972)

sounds in the General American (GA) dialect (see Flanagan (1972)).

In *vowel* production, the VT is more or less open at every point, and the sound is transmitted principally through mouth and, to a lesser degree, through nostrils. The vowels can be classified by two properties of the configuration, the position of the tongue hump (front, central and back) and degree of constriction of the VT at mouth. Altering the VT geometry has an effect on the acoustic eigenfrequencies of the air column in the VT. In phonetics these frequencies are known as the *formant frequencies*.

In the English language the production of some consonants resembles vowel production. For example the production of the *glides* [w, j] pronounced as in words "we" and "you" respectively, is very close to the production of [u] and [i] (pronounced as in words "boot" and "eve"). Also the production of the *semivowels* [r, l] pronounced as in "read" and "let" resembles that of vowels. The only difference is that the tongue is up creating a constriction at the mouth.

Also the *nasals* [m, n, ŋ] resemble vowels to some extent. They are produced by closing the vocal tract — either by lips in [m], the tip of the tongue against the hard palate in [n] or the back of the tongue against the soft palate [ŋ] — and holding the nasal tract widely open. The sound is then transmitted only through nostrils.

One class of the physically more complicated consonants are the *fricatives* [v, f, ð, Θ, z, s, ʒ, ʃ, h] pronounced as in words ”vote”, ”for”, ”then”, ”thin”, ”zoo”, ”see”, ”azure”, ”she” and ”he” respectively. The fricatives are produced by constricting the VT at certain point so that turbulent flow is formed at the constriction. For example [v] and [f] are produced by constricting the mouth opening by the teeth and the lower lip. The difference between these two is that voicing (that is, vocal cord oscillation with closure) occurs during the production of [v] which is not the case during the production of [f]. This way fricatives can be further classified into voiced [v, ð, z, ʒ] and their voiceless pairs [f, Θ, s, ʃ]. The so called *glottal fricative* [h] has no voiced counterpart.

Another class are the *stop consonants* [p, t, k, b, d, g], pronounced as in words ”pay”, ”to”, ”key”, ”be”, ”day” and ”go” respectively. They are produced by initially closing the VT at certain point and letting the lungs build up a pressure behind the closure. This pressure is then abruptly released by opening the closure. For example, when pronouncing [d] or [t], the VT is initially closed by pressing the tongue against the palate. Like fricatives, also the stop consonants can be subcategorized into voiced [b, d, g] and voiceless [p, t, k], depending on whether voicing occurs during the pressure buildup.

Of speech sounds not included in GA speech, let us present few examples whose production differs from any GA sound. One example is the Finnish [r], which is produced by letting the tip of the tongue vibrate against the hard palate. Another one is the French (or guttural) [r] which is produced by letting the velum vibrate against the back of the tongue.

1.3 Modelling human phonation

The demand for phonation models has increased constantly during the last fifty years. Applications of such models can be found in telephony and speech synthesizing technologies as well as some medical sciences such as surgery (see, *i.e.*, Sváček and Horáček (2006)). Perhaps the best known class of models consist of a low-order mass-spring model of glottis, coupled to some kind of static acoustic load representing the vocal tract (see, *e.g.*, Ishizaka and Flanagan (1972)). The model constructed in this thesis also falls under this category. These models are suitable for modelling the production of vowel (and vowel-like) speech sounds. Physically more complicated speech sounds, such as stop consonants and fricatives are outside these models' range.

One approach for studying human phonation are inverse filtering techniques (see, *i.e.*, Alku (1992) and Alku *et al.* (2006)) which constitute a demand for a prior model of the glottis signal. Such signal models are presented in *e.g.* Fant (1979) and Fant *et al.* (1986).

One of the earliest widely known physical glottis models is presented by Ishizaka and Flanagan (1972). Their glottis model is symmetric and it consists of two masses per cord. The aerodynamic force acting on glottis takes into account the Bernoulli effect and a viscous pressure drop according to the Hagen-Poiseuille equation. Their VT-model consists of four cylindrical tube-elements.

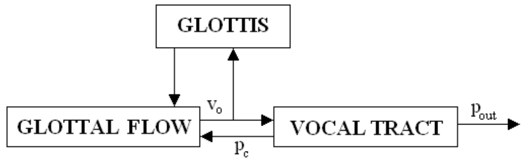


Figure 1.2: Block diagram of a model with a feedback configuration as in the model by Ishizaka and Flanagan (1972)

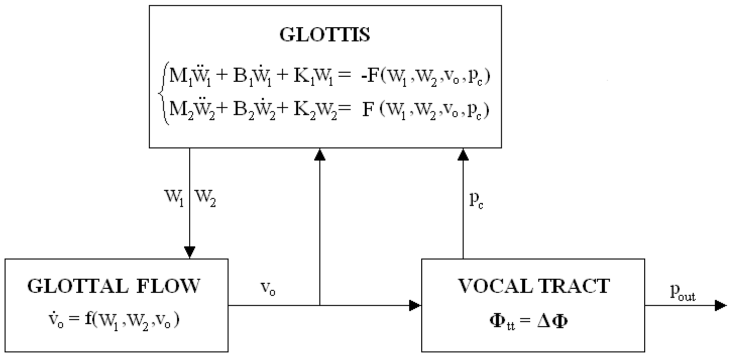


Figure 1.3: The block diagram of the model presented in this paper

The VT pressure at the glottis end is taken into account when evaluating the glottal flow. This kind of feedback configuration is illustrated in Fig. 1.2.

A more recent model of phonation is presented by Titze (2008). There the effect of the VT feedback to both glottal flow and vocal fold mechanics is studied first separately and then with a computational model.

1.4 Outline of this work

Fig. 1.3 shows the block diagram of the model constructed in the present work. Our design philosophy is to keep the model simple enough to be mathematically tractable. We want all the blocks to be physically realistic on a subsystem level. However, considering the whole system, there are some model simplifications on the subsystem level that would exclude each other.

First, in Chapter 2, a mass-spring model of the glottis is developed. The geometry of the model as well as the equivalent aerodynamic forces are highly simplified. The model has two degrees of freedom per cord and no symmetry assumption is made. This means that both vocal cords are allowed to vibrate independently. Thus, modelling of the effect of nonsymmetric parameters is possible. For the closed glottis, a nonlinear spring force is applied. This force is a simple version of the Hertz model of impact force as in another glottis model by Horáček *et al.* (2005). This paper and its predecessor (Horáček and Švec (2002)) have proven valuable references concerning also many other glottis model details.

This work also presents a model of the glottal flow. This model takes into account viscous pressure losses in the glottis and the vocal tract. It also takes into account the inertance of the VT. However, in the derivation of the flow equation, it is assumed that the air is incompressible. That is, the mass and volume flow through every cross-section in the VT is constant at a given moment.

At the end of Chapter 2, the behaviour of the glottis model is investigated by numerical simulation with different parameter configurations including a simulation with asymmetric glottis parameters. In these simulations there is no feedback from the VT directly to the mass-spring model. However, the flow model implicitly contains an inertive counter pressure from the vocal tract, which is always present.

In Chapter 3 the vocal tract model is presented. The model is a Webster's horn equation model which approximates the solution of the 3-D wave equation averaged over the VT cross-sections (for an early treatment of the Webster's equation, see Chiba and Kajiyama (1941)). Here we use a more general variant of the Webster's equation, derived by Lukkari and Malinen (2008b). The curvature of the tube is taken into account as a correction factor for the speed of sound. However, energy dissipation at the tube walls is not taken into account here. A solver based on the Finite Element Method is written for the VT model. At the end of Chapter 3, the lowest formant frequencies and corresponding pressure/velocity potential distributions are computed from an eigenvalue equation. The formants are compared to those given by a 3-D wave equation model by Hannukainen *et al.* (2007). These two models are constructed by using exactly the same magnetic resonance imaging (MRI) data for the VT, making this comparison reasonable.

In Chapter 4, the glottis and VT models are coupled together. For comparison, a simulation without the VT feedback is run. Then the effect of the feedback is investigated first for the actual VT geometry and then by using a straight tube as the resonator. The length of the tube is varied for tuning the formant frequencies.

It is particularly interesting to see what happens when the lowest formant frequency crosses the glottal fundamental frequency or its lowest multiples. This has been studied also by Titze (2008) and Hatzikirou *et al.* (2006) with a model similar to the one in Ishizaka and Flanagan (1972).

Chapter 2

The glottis model

In this chapter, we shall introduce the two blocks on the left in the block diagram (Fig. 1.3). First, we shall construct the mass-spring model of glottis in Section 2.1. Then, a 1-D model of the (incompressible) glottal flow with viscous pressure loss is constructed in Section 2.2. The coupling from the flow to the glottis model through the load force F is developed in Section 2.3.

The geometry of the vocal folds is as simple as possible. There is little point in refining the model geometry, when many of the material parameters are more or less neglected, and the aerodynamics in the flow model are based on somewhat harsh laminarity and incompressibility assumptions. The same applies also for the omission of trigonometric functions in the formulas of the load force F.

Simulations will be performed for the glottis model before it is connected to the vocal tract.

2.1 The mechanics of the glottis model

We consider a physical system shown in Fig. 2.1. The system consists of two wedge-shaped vibrating bodies having two degrees of freedom each. The system is practically two-dimensional, meaning that all cross-sections in the glottis are rectangular. The width of the vocal cords and the channel between them (to the direction perpendicular to the paper) is denoted by h.

This system can be replaced by an equivalent system that consists of altogether six masses, three each side. These three masses are attached to a rod of length L, so that there is one mass in both ends and one at the midpoint. This rod is connected to the wall of the channel with two sets of springs and dampers. The dampers are located at the endpoints of the rod whereas the springs are located at points whose distance from the midpoint is l. The reason for the placement of the springs is that the tuning properties are better than if the springs would be at the endpoints of the rod as well. This will be discussed in Section 2.4.2 in more detail. In addition, in the equivalent system

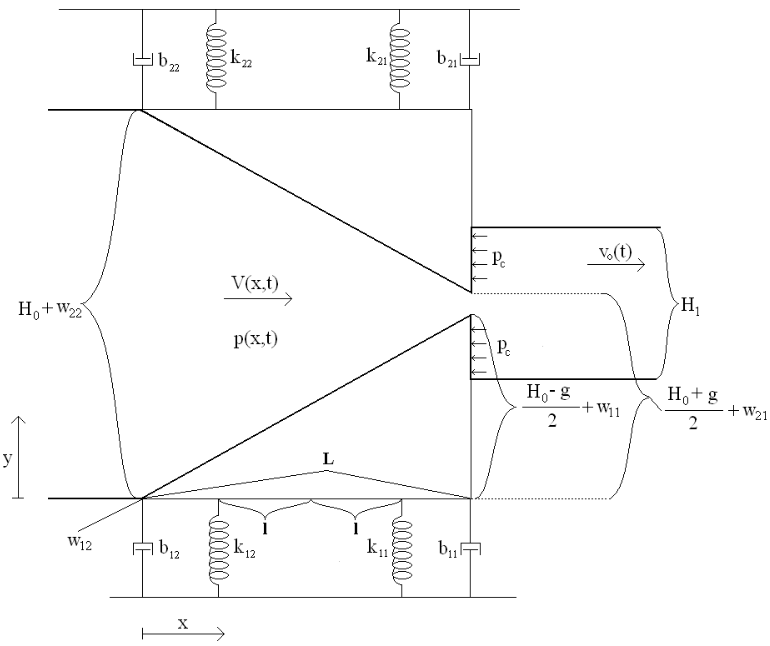


Figure 2.1: The geometry of our model

there are load forces F_1 and F_2 that depend only on the glottal openings at the narrow end of the glottis (point $x = L$) and in the wide end ($x = 0$), namely $\Delta W_1 := g - w_{11} + w_{21}$ and $\Delta W_2 := H_0 - w_{12} + w_{22}$. Here g is the glottal gap, when the displacements are zero. This gap is a control parameter in the model. When the glottis is open, F_1 and F_2 correspond to the force and moment caused by the dynamic pressure $p(x,t)$. When the glottis is closed, there is no air flow. Instead of the air pressure there is a contact force between the vocal cords pushing the cords apart.

The equations of motion for the cords are

$$\begin{cases} M_1\ddot{W}_1(t) + B_1\dot{W}_1(t) + K_1W_1(t) = -F \\ M_2\ddot{W}_2(t) + B_2\dot{W}_2(t) + K_2W_2(t) = F, \qquad t \in \mathbb{R} \end{cases} \tag{2.1}$$

where $W_j = (w_{j1} \quad w_{j2})^T$ are the displacements of the endpoints of the j^{th} cord ($j = 1,2$) and $F = (F_1 \quad F_2)^T$ is the external load force. M_j is the mass matrix, B_j is the damping matrix and K_j is the stiffness matrix.

The equilibrium position of the masses is taken to be $w_{ji} = 0, \ i,j = 1,2$ which occurs when there is no flow, and constant pressure p_{sub} at all sides of the vocal cords. Then $F \equiv 0$ and since the system is at rest, that is $\dot{W}_j = \ddot{W}_j = 0$, by Eq. (2.1) we have $W_j = 0$.

Next we shall calculate the entries of the mass and stiffness matrices by means of Lagrangian mechanics. First, we need to express the kinetic energy T_j

and potential energy V_j as functions of variables w_{ji} and their time derivatives $\dot{w}_{ji}$. For the j^{th} cord we get

$$T_j = \frac{1}{2} m_{j1} \dot{w}_{j1}^2 + \frac{1}{2} m_{j2} \dot{w}_{j2}^2 + \frac{1}{2} m_{j3} \left(\frac{\dot{w}_{j1} + \dot{w}_{j2}}{2} \right)^2 \tag{2.2}$$

and

$$V_j = \frac{1}{2} k_{j1} \left(a w_{j1} + b w_{j2} \right)^2 + \frac{1}{2} k_{j2} \left(b w_{j1} + a w_{j2} \right)^2 , \tag{2.3}$$

where $a = \frac{L/2+l}{L}$ and $b = \frac{L/2-l}{L}$.

The Lagrangian function is defined as $\mathscr{L}_j = T_j - V_j$ and it satisfies the Lagrange equations

$$\frac{d}{dt} \left(\frac{\partial \mathscr{L}_j}{\partial \dot{w}_{ji}} \right) - \frac{\partial \mathscr{L}_j}{\partial w_{ji}} = 0, \qquad i, j = 1, 2. \tag{2.4}$$

By substituting (2.2) and (2.3) into (2.4) we get the unloaded and undamped equations of motion

$$\begin{cases} m_{j1} \ddot{w}_{j1} + m_{j3} \frac{\ddot{w}_{j1} + \ddot{w}_{j2}}{4} + \left(a^2 k_{j1} + b^2 k_{j2} \right) w_{j1} + ab(k_{j1} + k_{j2}) w_{j2} = 0, \\ m_{j2} \ddot{w}_{j2} + m_{j3} \frac{\ddot{w}_{j1} + \ddot{w}_{j2}}{4} + \left(b^2 k_{j1} + a^2 k_{j2} \right) w_{j2} + ab(k_{j1} + k_{j2}) w_{j1} = 0. \end{cases}$$

Thus, the mass and stiffness matrices are

$$\begin{aligned} M_j &= \begin{bmatrix} m_{j1} + \frac{m_{j3}}{4} & \frac{m_{j3}}{4} \\ \frac{m_{j3}}{4} & m_{j2} + \frac{m_{j3}}{4} \end{bmatrix}, \\ K_j &= \begin{bmatrix} a^2 k_{j1} + b^2 k_{j2} & ab(k_{j1} + k_{j2}) \\ ab(k_{j1} + k_{j2}) & b^2 k_{j1} + a^2 k_{j2} \end{bmatrix}. \end{aligned} \tag{2.5}$$

Since the dampers are located at the endpoints of the cords, the damping matrices are diagonal

$$B_j = \begin{bmatrix} b_{j1} & 0 \\ 0 & b_{j2} \end{bmatrix}.$$

Numerical values of the physical constants are determined in Section 2.4.2. The damping coefficients b_{ji} remain tuning parameters.

2.2 Glottal flow

We denote the subglottal pressure with p_{sub} and take the pressure in the exterior space to be zero. We assume, that the pressure changes along the glottis and vocal tract for three reasons. Firstly, there is a Bernoulli flow through the mouth with velocity v_m. Secondly, there is a viscous pressure loss in the glottis and the VT. Thirdly, the air in the VT is in accelerating (decelerating) motion when the glottis is opening (closing) causing an inertive pressure. Mathematically

$$p_{sub} = \frac{1}{2}\rho v_m^2 + p_{loss}(\Delta W_1, v_o) + p_a, \tag{2.6}$$

where p_a is the pressure accelerating/decelerating the air in the VT, and p_{loss} is the pressure loss in the glottis and the VT. The supraglottal flow velocity is denoted by v_o, which is the quantity we are interested in. This pressure loss depends on the flow velocity and the state of the glottis through opening ΔW_1. It is here assumed that the air is incompressible in the VT too.

Recall that the pressure loss in a tube with circular cross-section is given by the Hagen-Poiseuille equation

$$\frac{dp}{dx} = -\frac{8\mu Q}{Ar^2} \tag{2.7}$$

where μ is the dynamic viscosity of the gas (unit $Pa \cdot s$), Q is the gas flux (m^3/s), A is the tube cross-sectional area, and r is the radius of the channel. The derivation of the Hagen-Poiseuille equation can be found in Fetter and Walecka (1980), pages 445-448. The Hagen-Poiseuille equation is derived for a laminar flow in a channel with circular cross-section (in which case $A = \pi r^2$) but it can be used also for other profile shapes. In that case the radius r must be replaced with the *hydraulic radius*, defined as

$$r_h = \frac{2A}{C}, \tag{2.8}$$

where A is the area and C is the circumference of the cross-section of the channel. For a tube with circular cross-section the hydraulic radius coincides with the radius of the cross-section.

The pressure loss in the VT is computed by integrating (2.7) over the VT. The VT geometry is presented in Section 3.2.5. Here we need the hydraulic radius r_h which is shown in Fig. 3.4, and the area function shown in Fig. 3.2. Thus the pressure loss in the VT is

$$p_{loss,VT} = v_o \frac{8\mu A_o}{\pi} \int_0^{L_{VT}} \frac{ds}{A(s) r_h(s)^2} =: v_o C_{VT}.$$

Between two parallel planes within distance H from each other, the Hagen-Poiseuille law is

$$\frac{dp}{dx} = -\frac{12\mu Q}{hH^3}. \tag{2.9}$$

One way to compute the pressure loss in the glottis would be to set H to be the height of the channel in the glottis, that is $H = H(x,t)$ and integrate this expression over the glottis. However, this pressure loss was experimentally found to be rather mild. Therefore, motivated by Eq. (2.9), the pressure loss in the glottis was taken to be of the form

$$p_{loss,g} = \frac{C_g}{\Delta W_1^3} v_o, \tag{2.10}$$

where the loss coefficient $C_g = \frac{12\mu A_o L_2}{h}$. This corresponds to a pressure loss in a rectangular tube with height ΔW_1, width h and length L_2, where $\Delta W_1 << h$. Here L_2 remains a tuning parameter to be determined experimentally.

Now the whole pressure loss can be written as

$$p_{loss}(\Delta W_1, v_o) = \left(\frac{C_g}{\Delta W_1^3} + C_{VT} \right) v_o. \tag{2.11}$$

Next we shall deal with the acceleration of the air in the VT. The power accelerating/decelerating the air in the VT is $p_a Q = p_a A_o v_o$. This power is equal to the change rate of the total kinetic energy of the air column, that is

$$\begin{aligned} p_a(t) A_o v_o(t) &= \frac{d}{dt} \int_{VT} \frac{1}{2} \rho v(\mathbf{r}, t)^2 d\mathbf{r} \\ &= \int_{VT} \rho v(\mathbf{r}, t) \dot{v}(\mathbf{r}, t) d\mathbf{r} \\ &= v_o(t) \dot{v}_o(t) \rho \int_{VT} \frac{A_o^2}{A(\mathbf{r})^2} d\mathbf{r} \qquad \left| \frac{d\mathbf{r}}{A(\mathbf{r})} = ds \right. \\ &= v_o(t) \dot{v}_o(t) \rho A_o^2 \int_0^{L_{VT}} \frac{ds}{A(s)}, \end{aligned}$$

where $A(\mathbf{r}) = A(s)$ is the area of the cross-section that contains $\mathbf{r}$ and whose distance from the glottis is s (measured along the VT centerline). Here we used $v(\mathbf{r}, t) = \frac{A_o}{A(\mathbf{r})} v_o(t)$ (and the same for $\dot{v}_o$) which follows from the incompressibility. By denoting the tube *inertance* by $C_{iner} := \rho \int_0^{L_{VT}} \frac{ds}{A(s)}$ we get $p_a(t) = C_{iner} A_o \cdot \dot{v}_o(t)$. Now Eq. (2.6) yields

$$\dot{v}_o(t) = \frac{1}{C_{iner} A_o} \left(p_{sub} - \frac{1}{2} \rho \left(\frac{A_o}{A_m} \right)^2 v_o(t)^2 - \left(\frac{C_g}{\Delta W_1^3} + C_{VT} \right) v_o(t) \right) \tag{2.12}$$

where the flow velocity at the mouth v_m is replaced with $\frac{A_o}{A_m} v_o$, and $A_m = A(L_{VT})$ is the area of the mouth opening. The constants C_{iner} and C_{VT} are determined by numerical integration from data presented in Figs. 3.4 and 3.2. The subglottal pressure p_{sub} remains a control parameter which is directly related to the average glottal volume flow.

2.3 The load force F

2.3.1 Aerodynamic force for the open glottis

We shall assume that the flow is one dimensional. That is, both the flow velocity $V = V(x, t)$ and the pressure $p = p(x, t)$, where x denotes the distance from the wide end of the glottis.

We shall use the static version of the law of conservation of mass for incompressible flow

$$H(x, t) V(x, t) = H_1 v_o, \tag{2.13}$$

where H_1 is the supraglottal channel height, which is set so that the channel area coincides $A(0)$ in the VT model developed in Chapter 3, that is $H_1 = A(0)/h$, where h is the channel width. In the glottis, the height of the channel is

$$H(x,t) = \Delta W_2(t) + \frac{x}{L}(\Delta W_1(t) - \Delta W_2(t)), \qquad x \in [0, L]. \tag{2.14}$$

The pressure and velocity distributions are connected through the continuity equation

$$\frac{\partial p(x,t)}{\partial x} + \rho V(x,t)\frac{\partial V(x,t)}{\partial x} + \frac{\partial V(x,t)}{\partial t} = 0, \qquad x \in [0, L].$$

However, the time derivative part is neglected here, and so we get the familiar Bernoulli law

$$p(x,t) + \frac{1}{2}\rho V(x,t)^2 = p_{sub} \tag{2.15}$$

where p_{sub} is the subglottal pressure.

Now we solve $V(x,t)$ from (2.13), and $p(x,t)$ from (2.15) and finally by using (2.14) we get

$$p(x,t) - p_{sub} = -\frac{1}{2}\rho v_o^2 \frac{H_1^2}{\left(\Delta W_2 + \frac{x}{L}(\Delta W_1 - \Delta W_2)\right)^2} \tag{2.16}$$

Thus we have connected the velocity distribution to the relative positions of the cords and the pressure distribution to the velocity distribution. The aerodynamic load force for the open glottis

$$F_A = \begin{pmatrix} F_{A,1} \\ F_{A,2} \end{pmatrix}, \qquad \Delta W_1 > 0$$

can now be determined by two integrals:

$$F_{A,1} + F_{A,2} = h \int_0^L (p(x,t) - p_{sub})\, dx \tag{2.17}$$

and

$$L \cdot F_{A,1} = h \int_0^L x(p(x,t) - p_{sub})\, dx - p_c \cdot h \frac{H_1}{2} \frac{H_0 - H_1}{2}, \tag{2.18}$$

where p_c is the supraglottal perturbation pressure from the vocal tract. The area of influence of pressure p_c is $hH_1/2$ (assuming the glottal gap to be negligible) and the moment arm of the corresponding force is $(H_0 - H_1)/2$. Here p_{sub} is subtracted from the pressure $p(x,t)$ because of our assumption that $w_{ij} = 0 \;\; \forall\, i,j = 1,2$ is the equilibrium position under subglottal pressure p_{sub}, and therefore forces F_1 and F_2 must vanish when $p(x,t) \equiv p_{sub}$ and $p_c = 0$.

Finally, using (2.16), the evaluation of integrals (2.17) and (2.18) yields

$$F_{A,1} + F_{A,2} = -\frac{\rho v_o^2 h L}{2} \cdot \frac{H_1^2}{\Delta W_1 \Delta W_2} \tag{2.19}$$

and

$$
\begin{aligned}
F_{A,1} \quad &= \quad \frac{\rho v_o^2 hL}{2}\left(-\frac{H_1^2}{\Delta W_1(\Delta W_2-\Delta W_1)}+\frac{H_1^2}{(\Delta W_1-\Delta W_2)^2}\ln\left(\frac{\Delta W_2}{\Delta W_1}\right)\right)- \\
&\quad -\frac{H_1(H_0-H_1)}{4L}h\cdot p_c. \qquad (2.20)
\end{aligned}
$$

Then by subtracting (2.20) from (2.19) we get

$$
\begin{aligned}
F_{A,2} = \quad &= \quad \frac{\rho v_o^2 hL}{2}\left(\frac{H_1^2}{\Delta W_2(\Delta W_2-\Delta W_1)}-\frac{H_1^2}{(\Delta W_1-\Delta W_2)^2}\ln\left(\frac{\Delta W_2}{\Delta W_1}\right)\right)+ \\
&\quad +\frac{H_1(H_0-H_1)}{4L}h\cdot p_c. \qquad (2.21)
\end{aligned}
$$

Note that if the supraglottal perturbation $p_c = 0$, we get (2.20) from (2.21) by interchanging $\Delta W_1 \longleftrightarrow \Delta W_2$. This symmetry could be expected because the flow direction has no effect on the aerodynamic forces in our simple flow model.

2.3.2 Contact force for the closed glottis

When the glottis is closed, the aerodynamic force is zero. Instead, there is an impact force due to collision of the vocal cords. Horáček *et al.* (2005) model this force by using a slightly simplified version of the Hertz model of impact forces (see Landau and Lifshitz (1970), pages 30-35). This impact force is of the form

$$f_H = k_H|\Delta W_1|^{3/2}, \quad \text{when } \Delta W_1 < 0.$$

In the Hertz model, the coefficient k_H depends on the material of the colliding objects and also on their shape, more precisely the radius of curvature at the contact point. Therefore, the coefficient cannot be defined by the Hertz model in our geometry. Despite this, using a nonlinear spring as impact force is physically justifiable, and we shall apply one.

Of course, the effect of the counter pressure p_c does not vanish when the glottis is closed. Together with the impact force the load force for the closed glottis becomes (see Eq. (2.18) and the discussion following it)

$$F_H = \begin{bmatrix} k_H|\Delta W_1|^{3/2} - \frac{H_0-H_1}{2L}\frac{H_1}{2}h\cdot p_c \\ \frac{H_0-H_1}{2L}\frac{H_1}{2}h\cdot p_c \end{bmatrix} \quad \text{when } \Delta W_1 < 0.$$

2.4 Numerical solution

2.4.1 Method

We have written MATLAB code for the numerical solution of the equations of motion (2.1) and the flow equation (2.12). This code can be found in Appendix A. The code uses the classical fourth order Runge-Kutta (RK) method for the equations of motion and implicit Euler method for the flow equation. The load function F in the equations is discontinuous, and this causes problems. In addition, the aerodynamic forces (2.20)-(2.21) are singular at the point of discontinuity. We can get rid of the singularity by replacing the point of discontinuity slightly above zero. This means stopping the flow when the glottal gap is under a certain threshold ϵ. Of course the viscosity in the glottis stops the flow anyway, and ϵ is chosen to be so small that the flow already is rather low. The meaning of this trick is merely that now we can use constant time step length for almost every step. Because of this we spare one matrix inversion on every step in the FEM solver for the VT model. This makes the numerical solution faster. The numerical solution is not sensitive to the choice of ϵ.

Thus the load function for the equations of motion is piecewise defined

$$F\big(\Delta W_1(t), \Delta W_2(t)\big) = \begin{cases} F_A\big(\Delta W_1(t), \Delta W_2(t)\big), & \text{when } \Delta W_1(t) > \epsilon \\ 0, & \text{when } \Delta W_1(t) \in [0, \epsilon] \\ F_H\big(\Delta W_1(t)\big), & \text{when } \Delta W_1(t) < 0. \end{cases}$$

Note that F has only one discontinuity at $\Delta W_1(t) = \epsilon$.

So we got rid of the singularity but the discontinuity still causes a problem in numerical solution. This is dealt with the following procedure. If at certain moment the glottis is open, meaning $\Delta W_{1,k} > \epsilon$, we use only values of F_A in the next RK-step, even on the ”wrong” side of the discontinuity if needed. Here we must be careful with the choice of the timestep length. It has to be chosen small enough, so that the change of ΔW_1 in one step does not exceed ϵ.

If the glottis closes at the next timestep, meaning $\Delta W_{1,k+1} < \epsilon$, we'll interpolate the point where the threshold ϵ is crossed. For this we use the second degree interpolating polynomial for which values $\Delta W_{1,k-1}$, $\Delta W_{1,k}$ and $\Delta W_{1,k+1}$ are needed.

We shall demonstrate this interpolation with an example. We assume that the threshold $\epsilon = 0.2$ and that by using F_A as load function we have solution points $\Delta W_{1,3} = 0.4$, $\Delta W_{1,4} = 0.3$ and $\Delta W_{1,5} = 0.16667$ (see Fig. 2.2). The threshold was crossed at the step $4 \to 5$. Now we shall interpolate by fitting a second degree polynomial to the solution points and solving the point where the threshold is crossed. In this example the point is $t = 4,772h$ where h is the length of the timestep. After this we fit interpolating polynomials for every variable and evaluate their values at $t = 4,772h$ and set these values for the new solution point. On the next step F_H is used as the load function because now the glottis is closed.

In this interpolation, the order of the error is $O(h^3)$ since we use the second degree interpolation polynomials. In one RK-step the order of the error is $O(h^4)$. However, the number of the steps where this interpolation is performed, does not depend on h, but only on the length of the simulation time interval. This means that the overall order of the error is $O(h^3)$.

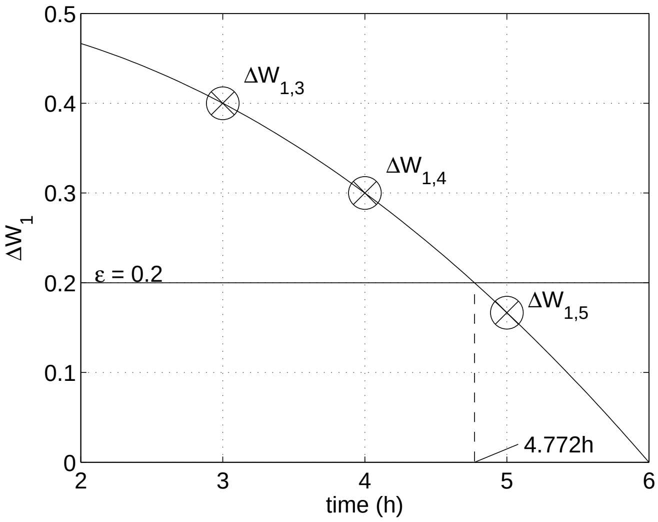


Figure 2.2: Interpolation example

2.4.2 Physical constants

The geometry of our model is as simple as possible. Therefore we shall determine the cords' total mass, static moment, and the moment of inertia by using a somewhat more realistic geometry than the one used for determining the aerodynamic forces. This geometry is the one used by Horáček and Švec (2002). They approximated the shape of the vocal fold by a parabolic function

$$a(x) = -159.861(x - 5.812 \cdot 10^{-3})^2 + 5.4 \cdot 10^3 \; [m] \qquad x \in [0, L].$$

The total mass, static moment and moment of inertia with respect to point $x = 0$ can now be evaluated by integrals

$$m = h\rho_h \int_0^L a(x)\,dx,$$

$$T = h\rho_h \int_0^L xa(x)\,dx,$$

$$I = h\rho_h \int_0^L x^2 a(x)\,dx,$$

where h is the width of the channel and ρ_h is the density of the vocal cords. Now the entries of the mass matrix (2.5) can be determined through conditions

$$\begin{aligned} m_{j1} + m_{j2} + m_{j3} &= m, \\ \tfrac{L}{2} m_{j3} + L m_{j1} &= T, \\ \left(\tfrac{L}{2}\right)^2 m_{j3} + L^2 m_{j1} &= I, \qquad j = 1, 2. \end{aligned} \tag{2.22}$$

As in Horáček and Švec (2002) the parameters in preceding equations were taken as follows: $L = 6.8\ mm$, $h = 18\ mm$ and $\rho_h = 1020\ kg/m^3$. With these parameters, by solving Eqs. (2.22) we get $m_{j1} = 1.686 \cdot 10^{-4}\ kg$, $m_{j2} = 0.595 \cdot 10^{-4}\ kg$ and $m_{j3} = 2.531 \cdot 10^{-4}\ kg$.

The height of the channel, which is also the glottal gap at $x = 0$ when the displacements $W_1 = W_2 = 0$, was taken as $H_0 = 11.2\ mm$. The glottal gap at the narrowest point $x = L$ was $g = 0.4\ mm$, when the displacements were zero. The air density was $\rho = 1.2\ kg/m^3$ and the dynamic viscosity $\mu = 18.7 \cdot 10^{-6} Pas$. The stiffness coefficient for the contact force was $k_H = 730\ N/m^{3/2}$. The subglottal pressure was $p_{sub} = 800\ Pa$ above the ambient pressure. The length L_2 in the expression of the glottal pressure loss coefficient was $0.8\ mm$ (see explanation related to Eq. (2.10)).

The Laplace-transformation of the undamped ($B = 0$) system yields

$$s^2 M\hat{W}(s) + K\hat{W}(s) = \hat{F}(s),$$

where M and K are as in (2.5). The transfer function from F to W is

$$G(s) = (s^2 M + K)^{-1}.$$

The natural (angular) frequencies of the system are the imaginary parts of the poles of the transfer function. Thus, they are obtained as the roots of the polynomial

$$r(s) = \det(s^2 M + K).$$

However, we want to solve an inverse problem. We want to fit the stiffness coefficients k_1 and k_2 so that they correspond to desired natural frequencies f_1 and f_2. Thus we want to solve equations

$$\begin{cases} r(2\pi i f_1) = 0 \\ r(2\pi i f_2) = 0 \end{cases} \tag{2.23}$$

with respect to stiffness coefficients k_1 and k_2. Here the problem was that a real solution did not always exist if the natural frequencies were close to each other. This problem is solved by adjusting the parameter l, which is the distance between the midpoint $x = L/2$ and the springs. When $l = 0.35L$ the stiffness coefficients were real in all simulated cases.

Bounds for the damping parameters b_{ji} were experimentally found so that the damped system was stable but not overdamped, meaning that the oscillation did not stop once it had started. We used values $b_{ji} = 0.1\ Nm/s$ for $i, j = 1, 2$, when there was no feedback from the vocal tract.

2.4.3 Results

First we set $f_1 = 100\ Hz$ and $f_2 = 105\ Hz$. Solving equations (2.23) with these frequencies gives $k_{11} = k_{21} = 124\ N/m$ and $k_{12} = k_{22} = 69\ N/m$. Fig. 2.3 shows the eigenmodes of the cords vibrating *in vacuo* and their corresponding eigenfrequencies with these parameters.

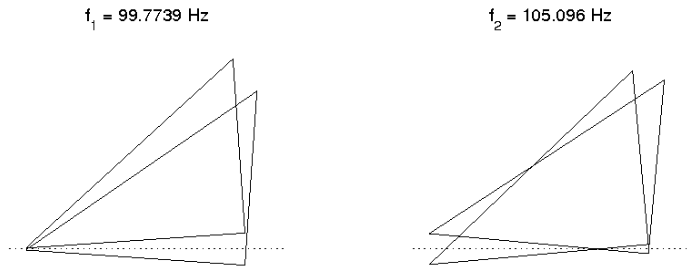


Figure 2.3: Eigenmodes and corresponding eigenfrequencies for the cords vibrating *in vacuo*

The timestep in all simulations was $0.02\ ms$. First time domain simulation was performed with all-symmetric parameters and initial conditions. The results of this simulation are shown in Fig. 2.4. The upper picture shows the positions of the cords in the narrow end of the glottis ($x = L$). The picture in the middle shows the oscillation of the lower cord at the wide end of the glottis ($x = 0$). The lowest picture shows the glottal area,

$$A_g = \begin{cases} h\Delta W_1(t), & \text{when } \Delta W_1(t) > 0, \\ 0, & \text{when } \Delta W_1(t) \leq 0. \end{cases}$$

The behaviour of the model is regular. The frequency of the oscillations is $F_0 = 118\ Hz$ and the open quotient (OQ) is 0.63, meaning that the glottis is open 63 % of the time. The average glottal volume flow is $\frac{1}{T}\int_0^T A_o v_{out}(t)dt = 0.30\ l/s$, where T is one period duration. Fig. 2.5 shows the glottal area function and the flow through the glottis during one open phase.

We also carried out a simulation with non-symmetric masses. We set the mass m_{21} 20 % greater than m_{11}. Other parameters were as in the first simulation. The positions of the cords in the narrow end of the glottis are shown in the upper picture in Fig. 2.6. The asymmetry causes a phase difference between the cords' oscillation, and reduces the oscillation frequency to 114 Hz. The OQ is again 0.63 and the average glottal volume flow is 0.31 l/s. The phase difference is illustrated in the lower pictures which show the phase diagrams $(w_{11}(t), w_{12}(t))$ and $(w_{12}(t), w_{22}(t))$. However, besides the frequency, the only thing that can be "heard" from the glottal behaviour is the glottal area function, and it is not remarkably influenced by the asymmetry.

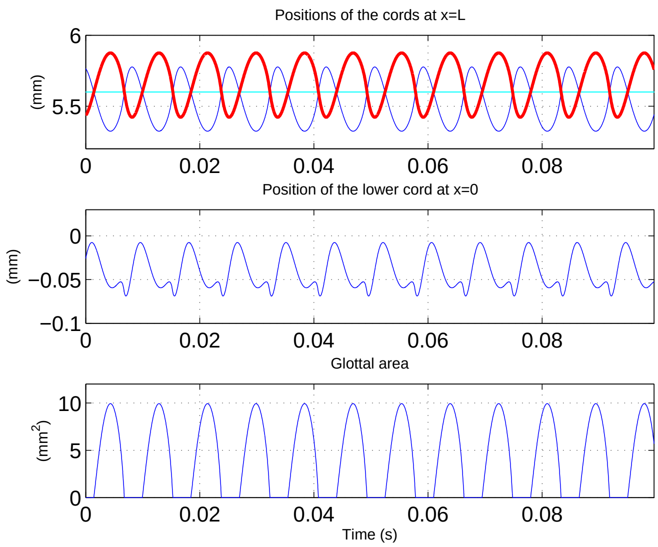


Figure 2.4: Results of the symmetric simulation; $f_1 = 100\ Hz,\ f_2 = 105\ Hz$

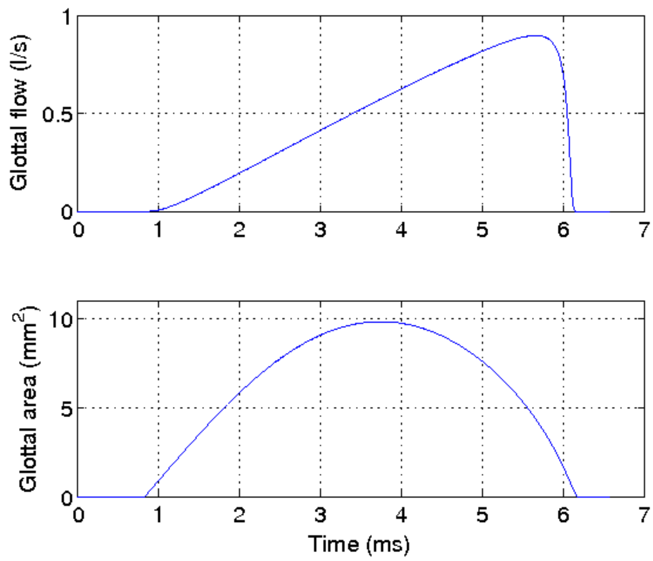


Figure 2.5: The output velocity and the glottal area function during one pulse. The simulation parameters are as in the first simulation (Fig. 2.4).

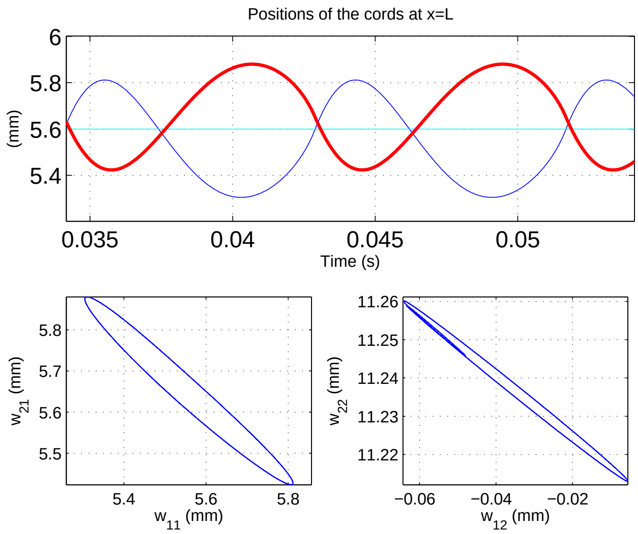


Figure 2.6: Results of the asymmetric simulation. The mass m_{21} is 20 % greater than m_{11}. The cord nr. 2 corresponds to the thicker line in the upper picture.

2.4.4 Parameter identification of the F-model

Fant (1979) used a three-parameter model (often referred to as the F-model) to describe the glottal flow pulse. This model was later improved by Fant *et al.* (1986) (known as the LF-model). They removed the abrupt flow termination in the F-model and added an exponential decay to the end of the flow derivative. The parameters for these pulses are determined by inverse filtering of measurements of the volume velocity at the lips (see, *i.e.*, Alku *et al.* (2006) and the references therein).

In our model the end of the pulse is smooth (that is, continuously differentiable). Despite this, the pulse has more resemblance to the F-model. Therefore we shall compare our velocity pulse to the three-parameter F-model pulse fitted into our pulse. These pulses are presented in Fig. 2.7.

The glottal volume velocity pulse in Fant (1979) consists of two pieces, a rising and a falling branch:

$$U(t) = \begin{cases} \frac{1}{2}U_0(1-\cos(\omega t)), & \text{when } t \in (T_1, T_{max}), \\ U_0\big(K\cos\big(\omega(t-T_{max})\big) - K + 1\big), & \text{when } t \in [T_{max}, T_2). \end{cases}$$

The three parameters are the peak value U_0, the pulse rise frequency $\omega = \frac{\pi}{T_{max}-T_1}$, where T_1 is the time, when glottis opens ($T_1 = 0$ in the picture) and T_{max} is the peak time. The third parameter is the steepness factor for the falling branch $K = \big(1-\cos\big(\omega(T_2-T_{max})\big)\big)^{-1}$, where T_2 is the time, when glottis closes again.

In Fig. 2.7 there is the output volume velocity pulse (the same pulse as in Fig. 2.5) and the Fant-model pulse that is fitted to our pulse as described above. The parameters used for fitting are also shown in the picture.

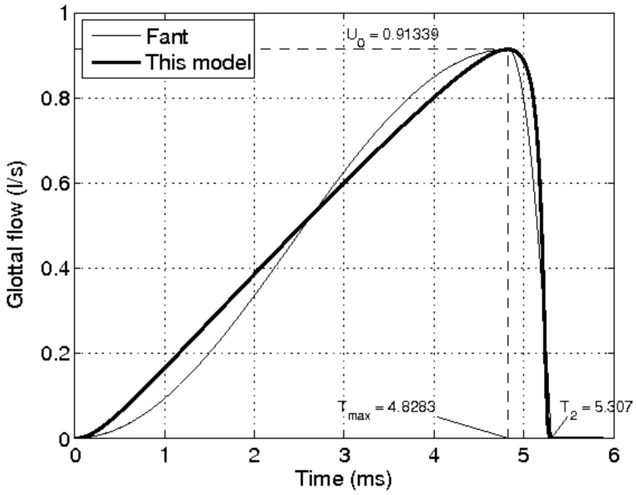


Figure 2.7: The output volume velocity pulse given by our model and a fitted Fant-model pulse.

Chapter 3

The vocal tract model

Our purpose is to connect our glottis model to an acoustic load which is modelled by the Webster's equation.

Consider first the solution of the wave equation for the velocity potential. Since we are handling a tube-like domain, we know that the wave motion propagates mainly in the direction of the tube. This motivates us to study only the solution's average over each cross-section of the tube. Our goal is to write an equation approximating the behaviour of this averaged solution, that would be simpler than the 3-D wave equation. This equation is known as the Webster's horn equation.

A complete derivation of this equation can be found in Lukkari and Malinen (2008b). They also take into account the curvature of the tube, which causes a correction factor for the speed of sound. The derivation of the Webster's equation with curvature will be outlined here.

Before connecting the glottis model and the VT model together, the formant frequencies and corresponding pressure distributions will be computed in this chapter. These results can be compared with a 3-D wave equation model by Hannukainen *et al.* (2007). This comparison is reasonable because the models are constructed by using the same data for the VT-geometry.

3.1 The Webster's equation

3.1.1 Preliminaries

We are looking for an approximate solution to the wave equation

$$\begin{cases} \Phi_{tt} = c^2 \Delta \Phi, & \text{in } \Omega, \\ \Phi_t + \theta c \frac{\partial \Phi}{\partial \nu} = 0, & \text{on } \Gamma_1, \\ \frac{\partial \Phi}{\partial \nu} = 0, & \text{on } \Gamma_2, \\ \frac{\partial \Phi}{\partial \nu} = u, & \text{on } \Gamma_3, \end{cases} \tag{3.1}$$

where Ω is the interior of the vocal tract, Γ_1 is the mouth opening, Γ_2 denotes the walls of the vocal tract, and Γ_3 is a control surface above the glottis. The function Φ is a velocity potential, that is, a function satisfying $-\nabla\Phi = \mathbf{v}$. The coefficient θ in the mouth boundary condition is the *normalized acoustic resistance* — a dimensionless coefficient regulating the radiation resistance at lips.

We shall begin with a path $\gamma : [0, L_{VT}] \to \mathbb{R}^3$, which is parameterized by its arch length, L_{VT} being the length of the vocal tract. This is the *centerline* of our curved tube. We define the *curvature* of the path at point $\gamma(s)$ by $\kappa(s) := ||\gamma''(s)||$.

An orthonormal coordinate system is fixed to every point of γ. The three unit vectors are defined by

$$\mathbf{t}(s) := \gamma'(s), \quad \mathbf{n}(s) := \frac{\mathbf{t}'(s)}{\kappa(s)} \quad \text{and} \quad \mathbf{b}(s) := \mathbf{t}(s) \times \mathbf{n}(s).$$

The vector $\mathbf{t}(s)$ is called the tangent vector, $\mathbf{n}(s)$ is the normal vector and $\mathbf{b}(s)$ is the binormal vector. This orthonormal coordinate system is called the Frenet frame and it is a right hand coordinate system for $\mathbb{R}^3$ at all points of the curve, where $\kappa(s) > 0$. In the derivation of the Webster's equation it is assumed that $\kappa(s) > 0 \ \forall \ s \in [0, L_{VT}]$.

Next we shall form the tube around the centerline γ. To every point $\gamma(s)$ we attach a $\gamma(s)$-centered disc with radius $R(s)$, which lies on the plane whose normal vector is $\mathbf{t}(s)$. This disc is denoted by $\Gamma(s)$ and it is parameterized with polar coordinates by using vectors $\mathbf{n}(s)$ and $\mathbf{b}(s)$ as the basis vectors for the plane. Thus the tube representing the vocal tract can be written in parameterized form

$$\Omega = \left\{ \gamma(s) + r\cos\theta\mathbf{n}(s) + r\sin\theta\mathbf{b}(s) \;\middle|\; s \in [0, L_{VT}],\ r \in [0, R(s)),\ \theta \in [0, 2\pi) \right\}.$$

The parameters (s, r, θ) can be used as coordinates in the tube and henceforth they are called the *tube coordinates*. We make a standing assumption

$$\eta(s) := R(s)\kappa(s) < 1 \quad \forall s \in [0, L_{VT}]$$

which says that the tube does not fold onto itself guaranteeing that the coordinate transformation $(s, r, \theta) \mapsto (x, y, z)$ is bijective. The number $\eta(s)$ is called the *curvature ratio*.

3.1.2 The derivation of the Webster's equation

As mentioned before, a complete derivation will not be presented here, and the readers looking for one are referred to Lukkari and Malinen (2008b). First, it is assumed that Φ is the solution of (3.1). Then the averaged solution is defined as

$$\overline{\Phi}(s, t) := \frac{1}{A(s)} \int_{\Gamma(s)} \Phi dA, \tag{3.2}$$

where $A(s) = \pi R(s)^2$.

The next steps in the derivation of the Webster's equation in Lukkari and Malinen (2008b) are rather lengthy and would require much more preliminary work, so unfortunately some of the definitions presented here are not very well-motivated. After writing the wave equation in integral form and using the divergence theorem, Neumann boundary conditions on the walls of the VT, a function $L(\cdot,\cdot)$ is defined by

$$L(s_0, s_1) := \int_{\Gamma(s_1)} \frac{\partial \Phi}{\partial s} dA - \int_{\Gamma(s_0)} \frac{\partial \Phi}{\partial s} dA - \int_{s_0}^{s_1} \left(\int_{\Gamma(s)} \frac{1}{c^2 \Xi^2} \frac{\partial^2 \Phi}{\partial t^2} dA \right) ds, \quad (3.3)$$

where $\Xi(s, r, \theta) := (1 - r\kappa(s)\cos\theta)^{-1}$ is the *curvature factor*. To gain some motivation for this definition, let us note that the first two terms here can be interpreted as the most significant term of $\frac{\Delta \Phi}{\Xi}$ integrated over the piece of the tube between $\Gamma(s_0)$ and $\Gamma(s_1)$.

To obtain the desired equation it is necessary to study the limit $\lim_{s' \to s} \frac{L(s,s')}{s'-s}$. In Lukkari and Malinen (2008b) it is shown that (under certain smoothness assumptions) we have for the limit

$$\lim_{s' \to s} L(s, s') = \int_{\Gamma(s)} \frac{1}{\Xi} \nabla \left(\frac{1}{\Xi} \right) \cdot \nabla \Phi dA. \quad (3.4)$$

The right hand side of (3.4) is the residual of $\frac{\Delta \Phi}{\Xi}$ that was not included in the definition of L. It is assumed to be small and it is included in the error term. Next, this limit of L is calculated starting from the definition (3.3).

The first two terms in (3.3) are dealt with by showing that for the averaged solution (3.2) it holds that

$$\begin{aligned} A(s)\frac{\partial \overline{\Phi}}{\partial s} &= -A'(s)\overline{\Phi} + \frac{\partial}{\partial s}\left(\int_{\Gamma(s)} \Phi dA \right) \\ &= -A'(s)\overline{\Phi} + \int_{\Gamma(s)} \frac{\partial \Phi}{\partial s} dA + \frac{A'(s)}{2\pi} \int_0^{2\pi} \Phi(s, R(s), \theta) d\theta. \end{aligned}$$

It now directly follows that

$$\int_{\Gamma(s)} \frac{\partial \Phi}{\partial s} dA = A(s)\frac{\partial \bar{\Phi}}{\partial s} + A'(s)\left(\bar{\Phi}(s) - \frac{1}{2\pi} \int_0^{2\pi} \Phi(s, R(s), \theta) d\theta \right). \quad (3.5)$$

Note that the expression inside the parenthesis is a difference of two means of Φ, on $\Gamma(s)$ and $\partial\Gamma(s)$, respectively.

For the last term in (3.3) the limit of the desired form is easy to see and thus, by using (3.5) we get for the limit

$$\begin{aligned}
\lim_{s'\to s} \frac{L(s,s')}{s'-s} &= \frac{\partial}{\partial s}\left(\int_{\Gamma(s)} \frac{\partial \Phi}{\partial s} dA\right) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left(\int_{\Gamma(s)} \frac{\Phi}{\Xi^2} dA\right) \\
&= \frac{\partial}{\partial s}\left(A(s)\frac{\partial \bar{\Phi}}{\partial s}\right) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left(\int_{\Gamma(s)} \frac{\Phi}{\Xi^2} dA\right) \\
&+ \frac{\partial}{\partial s}\left(A'(s)\left(\bar{\Phi} - \frac{1}{2\pi}\int_0^{2\pi} \Phi(s, R(s), \theta) d\theta\right)\right).
\end{aligned} \tag{3.6}$$

Here the first term looks good, and the last term is included in the error. However, in the middle term we have Ξ^{-2} multiplying Φ inside the integral, and since it depends on r and θ, it cannot be brought out from the integral without due punishment. Therefore, we shall define the *sound speed correction factor* as the average of Ξ^{-2}:

$$\frac{1}{\Sigma(s)^2} := \frac{1}{A(s)}\int_{\Gamma(s)} \frac{dA}{\Xi^2} = 1 + \frac{1}{4}\eta(s)^2,$$

where the latter equivalence is obtained by a straightforward calculation from the definition of Ξ.

In the sense of least squares, the average $\Sigma(s)^{-2}$ is the best constant estimate for function $\Xi(s,r,\theta)^{-2}$ over $\Gamma(s)$. We define the error function

$$E(s,r,\theta) := \frac{1}{\Xi(s,r,\theta)^2} - \frac{1}{\Sigma(s)^2} \tag{3.7}$$

allowing us to write the middle term in (3.6) in the form

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\left(\int_{\Gamma(s)} \frac{\Phi}{\Xi^2} dA\right) = \frac{A(s)}{c^2\Sigma(s)^2}\frac{\partial^2 \overline{\Phi}}{\partial t^2} + \int_{\Gamma(s)} \frac{E}{c^2}\frac{\partial^2 \Phi}{\partial t^2} dA. \tag{3.8}$$

By using (3.4), (3.6) and (3.8) we get

$$\frac{1}{c^2\Sigma(s)^2}\frac{\partial^2 \overline{\Phi}}{\partial t^2} - \frac{1}{A(s)}\frac{\partial}{\partial s}\left(A(s)\frac{\partial \overline{\Phi}}{\partial s}\right) = F(s,t) + G(s,t), \tag{3.9}$$

where F and G contain the error terms gathered from (3.4), (3.6) and (3.8):

$$\begin{aligned}
F(s,t) &:= \frac{1}{A(s)}\frac{\partial}{\partial s}\left(A'(s)\left(\bar{\Phi} - \frac{1}{2\pi}\int_0^{2\pi} \Phi(s, R(s), \theta) d\theta\right)\right), \\
G(s,t) &:= \frac{1}{A(s)}\int_{\Gamma(s)} \left(E\Delta\Phi - \frac{1}{\Xi}\nabla\left(\frac{1}{\Xi}\right)\cdot \nabla\Phi\right) dA.
\end{aligned}$$

Now $F(s,t)$ contains a difference of two averages of the solution of the wave equation. This difference is small, if the tube area is small. In $G(s,t)$ the term $\Delta\Phi$ is limited and the error function $E(s,r,\theta)$ in (Eq. (3.7)) is a difference of a function and its average over the disc. This difference small, if the curvature factor is close to one ($E \equiv 0$ for an uncurved tube). The second term in $G(s,t)$ is small if the curvature factor $\eta(s)$ and the components of $\nabla\Phi$ that are perpendicular to the tube centerline are small.

Now, for the solution of the wave equation, Eq. (3.9) holds. The Webster's horn equation with curvature is

$$\frac{1}{c^2\Sigma(s)^2}\frac{\partial^2\psi}{\partial t^2} - \frac{1}{A(s)}\frac{\partial}{\partial s}\left(A(s)\frac{\partial\psi}{\partial s}\right) = 0. \tag{3.10}$$

3.2 Numerical solution

We solve numerically the Webster's equation (3.10) with boundary conditions corresponding to the wave equation (3.1), that is

$$\begin{cases} \frac{\partial\psi}{\partial s}(0,t) = -v_o(t) \\ \psi_t(L_{VT},t) + \theta c\frac{\partial\psi(L_{VT},t)}{\partial s} = 0, \end{cases} \tag{3.11}$$

where L_{VT} denotes the length of the vocal tract. Here $v_o(t)$ is the glottal flow derived in Section 2.2. Note that the channel area after glottis, denoted by A_o in Section 2.2, is equal to $A(0)$ allowing v_o to be used directly as the VT input. The latter boundary condition models boundary dissipation in the form of flow resistance $p = \theta\rho cv$.

3.2.1 Weak formulation of the Webster's equation

Let us first write a weak formulation of the Webster's equation. First, we shall write the Webster's equation in first order form by introducing an auxiliary function $\pi(s,t) = \rho\dot{\psi}(s,t)$. Then we define $\mathbf{W} := \frac{1}{A(s)}\frac{\partial}{\partial s}\left(A(s)\frac{\partial}{\partial s}\right)$, and we get

$$\frac{d}{dt}\begin{bmatrix}\psi\\ \pi\end{bmatrix} = \begin{bmatrix}0 & \rho^{-1}\\ \rho c(s)^2\mathbf{W} & 0\end{bmatrix}\begin{bmatrix}\psi\\ \pi\end{bmatrix}.$$

Henceforth let $\mathbf{L} := \begin{bmatrix}0 & \rho^{-1}\\ \rho c(s)^2\mathbf{W} & 0\end{bmatrix} : \mathcal{Z} \to \mathcal{X}$, where

$$\mathcal{Z} := \left(H^1(0,L_{VT}) \cap H^2(0,L_{VT})\right) \times H^1(0,L_{VT});$$

$$\mathcal{X} := H^1(0,L_{VT}) \times L^2(0,L_{VT}).$$

We equip the Hilbert space $\mathcal{X}$ with the inner product

$$\left\langle \begin{bmatrix}y_1\\ y_2\end{bmatrix}, \begin{bmatrix}x_1\\ x_2\end{bmatrix}\right\rangle_{\mathcal{X}} := \frac{1}{2}\left(\rho\int_0^{L_{VT}} y_1'(s)x_1'(s)A(s)ds + \frac{1}{\rho c^2}\int_0^{L_{VT}} y_2(s)x_2(s)\frac{A(s)}{\Sigma(s)^2}ds\right).$$

The norm induced by this inner product is the physical *energy norm*.

The endpoint control and observation operators are defined by

$$\mathbf{G}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} := \begin{bmatrix} -z_1'(0) \\ z_2(L_{VT}) + \theta\rho c z_1'(L_{VT}) \end{bmatrix} \quad \text{and} \quad \mathbf{H}\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} := z_2(0),$$

where $(z_1 \ z_2)^T \in \mathcal{Z}$. Now the vocal tract model can be written as a linear boundary control system

$$\begin{cases} \dot{z}(t) &= \mathbf{L}z(t) \\ \mathbf{G}z(t) &= \begin{bmatrix} v_o(t) \\ 0 \end{bmatrix} \\ \mathbf{H}z(t) &= p_c(t) \\ z(0) &= z_0 \end{cases} \tag{3.12}$$

Here the first, second and fourth equation define the solution $z(t)$ and the output is given by the third equation. Malinen and Staffans (2006) and Malinen and Staffans (2007) treat the solvability of such boundary control systems and in Lukkari and Malinen (2008a) it is shown that (3.12) satisfies the conditions required for conservativity and solvability. The reason why the control operator $\mathbf{G}$ is defined in this manner is that now the mouth boundary term is included in the control term. Thus, the system can be shown to be conservative also with boundary conditions (3.11) with a small modification of the argument in Malinen and Staffans (2007).

In order to obtain the weak formulation, we take a test function $\begin{bmatrix} v(s) \\ 0 \end{bmatrix} \in \mathcal{X}$ and take the inner product of the top row of (3.12) and this test function:

$$\left\langle \begin{bmatrix} \dot{\psi}(s,t) \\ \dot{\pi}(s,t) \end{bmatrix}, \begin{bmatrix} v(s) \\ 0 \end{bmatrix} \right\rangle_{\mathcal{X}} = \left\langle \mathbf{L}\begin{bmatrix} \psi(s,t) \\ \pi(s,t) \end{bmatrix}, \begin{bmatrix} v(s) \\ 0 \end{bmatrix} \right\rangle_{\mathcal{X}}. \tag{3.13}$$

For the left hand side of this we get

$$\left\langle \begin{bmatrix} \dot{\psi}(s,t) \\ \dot{\pi}(s,t) \end{bmatrix}, \begin{bmatrix} v(s) \\ 0 \end{bmatrix} \right\rangle_{\mathcal{X}} = \frac{\rho}{2}\int_0^{L_{VT}} \frac{\partial^2 \psi(s,t)}{\partial s \partial t}\frac{\partial v(s)}{\partial s} A(s) ds$$

and the right hand side

$$\begin{aligned} \left\langle \mathbf{L}\begin{bmatrix} \psi(s,t) \\ \pi(s,t) \end{bmatrix}, \begin{bmatrix} v(s) \\ 0 \end{bmatrix} \right\rangle_{\mathcal{X}} &= \left\langle \begin{bmatrix} 0 & \rho^{-1} \\ \rho c(s)^2 \mathbf{W} & 0 \end{bmatrix} \begin{bmatrix} \psi(s,t) \\ \pi(s,t) \end{bmatrix}, \begin{bmatrix} v(s) \\ 0 \end{bmatrix} \right\rangle_{\mathcal{X}} \\ &= \frac{\rho}{2}\int_0^{L_{VT}} \rho^{-1} \frac{\partial \pi(s,t)}{\partial s}\frac{\partial v(s)}{\partial s} A(s) ds. \end{aligned}$$

We do the same thing for another test function $\begin{bmatrix} 0 \\ v(s) \end{bmatrix} \in \mathcal{X}$ to obtain

$$\left\langle \begin{bmatrix} \dot{\psi}(s,t) \\ \dot{\pi}(s,t) \end{bmatrix}, \begin{bmatrix} 0 \\ v(s) \end{bmatrix} \right\rangle_{\mathcal{X}} = \left\langle \mathbf{L}\begin{bmatrix} \psi(s,t) \\ \pi(s,t) \end{bmatrix}, \begin{bmatrix} 0 \\ v(s) \end{bmatrix} \right\rangle_{\mathcal{X}}. \tag{3.14}$$

Now the left hand side is

$$\left\langle \begin{bmatrix} \dot{\psi}(s,t) \\ \dot{\pi}(s,t) \end{bmatrix}, \begin{bmatrix} v(s) \\ 0 \end{bmatrix} \right\rangle_{\mathcal{X}} = \frac{1}{2\rho c^2} \int_0^{L_{VT}} \frac{\partial \pi(s,t)}{\partial t} v(s) \frac{A(s)}{\Sigma(s)^2} ds$$

and the right hand side

$$\begin{aligned} \left\langle \mathbf{L} \begin{bmatrix} \psi \\ \pi \end{bmatrix}, \begin{bmatrix} 0 \\ v(s) \end{bmatrix} \right\rangle_{\mathcal{X}} &= \left\langle \begin{bmatrix} 0 & \rho^{-1} \\ \rho c(s)^2 \mathbf{W} & 0 \end{bmatrix} \begin{bmatrix} \psi \\ \pi \end{bmatrix}, \begin{bmatrix} 0 \\ v(s) \end{bmatrix} \right\rangle_{\mathcal{X}} \\ &= \frac{1}{2\rho c^2} \int_0^{L_{VT}} \rho c^2 \Sigma(s)^2 \mathbf{W} \psi(s,t) v(s) \frac{A(s)}{\Sigma(s)^2} ds \\ &= \frac{1}{2} \int_0^{L_{VT}} \frac{\partial}{\partial s} \left(A(s) \frac{\partial \psi(s,t)}{\partial s} \right) v(s) ds. \end{aligned}$$

Partial integration yields

$$\left\langle \mathbf{L} \begin{bmatrix} \psi \\ \pi \end{bmatrix}, \begin{bmatrix} 0 \\ v(s) \end{bmatrix} \right\rangle_{\mathcal{X}} = \frac{1}{2} \Big|_0^{L_{VT}} A(s) \frac{\partial \psi(s,t)}{\partial s} v(s) - \frac{1}{2} \int_0^{L_{VT}} A(s) \frac{\partial \psi(s,t)}{\partial s} \frac{\partial v(s)}{\partial s} ds \tag{3.15}$$

3.2.2 Spatial discretization

The basis functions of the element space are formed next. First, the vocal tract is divided into N slices of equal length $\Delta s := L_{VT}/N$. Then, we shall define piecewise linear functions $v_j(s),\ j = 1, ..., N+1$ by

$$v_j(s) := \begin{cases} \frac{s-(j-2)\Delta s}{\Delta s}, & s \in [(j-2)\Delta s, (j-1)\Delta s], \\ -\frac{s-j\Delta s}{\Delta s}, & s \in [(j-1)\Delta s, j\Delta s], \\ 0, & s \notin [(j-2)\Delta s, j\Delta s]. \end{cases}$$

For $j = 1$, the definition on the top row, and for $j = N+1$ the definition on the middle row do not apply. The functions v_j are called hat functions because of their form. The function v_j reaches value 1 at point $s = (j-1)\Delta s$. This means, that for the first basis function $v_1(0) = 1$ and for the last basis function $v_{N+1}(L_{VT}) = 1$.

Thus, we are looking for an approximate solution of (3.12) of the form

$$\begin{bmatrix} \psi(s,t) \\ \pi(s,t) \end{bmatrix} = \sum_{i=1}^{N+1} \left(\xi_i(t) \begin{bmatrix} v_i(s) \\ 0 \end{bmatrix} + \mu_i(t) \begin{bmatrix} 0 \\ v_i(s) \end{bmatrix} \right), \tag{3.16}$$

such that the residual is orthogonal to all of the basis functions. If we now insert this into Eqs. (3.13) and (3.14) and instead of some $v(s)$ we take the inner product with all the basis functions, we get $2(N+1)$ equations which can be written in matrix form

$$\begin{cases} \rho \mathbf{K} \dot{\xi}(t) &= \mathbf{K} \mu(t), \\ \mathbf{M} \dot{\mu}(t) &= -\mathbf{K} \xi(t) - \mathbf{R} \mu + \mathbf{b}(t) \end{cases} \tag{3.17}$$

corresponding to (3.12). Here

$$\begin{aligned}
\mathbf{M}_{ij} &= \frac{1}{2\rho c^2}\int_0^{L_{VT}} v_i(s)v_j(s)\frac{A(s)}{\Sigma(s)^2}ds, \\
\mathbf{K}_{ij} &= \frac{1}{2}\int_0^{L_{VT}} v_i'(s)v_j'(s)A(s)ds, \\
\mathbf{R}_{ij} &= \begin{cases} \frac{A(L_{VT})}{2\theta\rho c}, & \text{when } i=j=N+1; \\ 0, & \text{otherwise,} \end{cases} \\
\mathbf{b}_j(t) &= \begin{cases} \frac{A(0)}{2}v_o(t), & \text{when } j=1; \\ 0, & \text{when } j\neq 1. \end{cases}
\end{aligned} \tag{3.18}$$

The damping matrix $\mathbf{R}$ and the load vector $\mathbf{b}$ are gathered from the substitution term in (3.15) by using the boundary conditions (3.11) for the Webster's equation. Since the stiffness matrix $\mathbf{K}$ is invertible, it can be eliminated from the first equation of (3.17).

3.2.3 Temporal discretization

Next task is the time discretization. We replace $\xi(t)$ and $\mu(t)$ with approximative solutions $\xi^n \approx \xi(t_n)$ and $\mu^n \approx \mu(t_n)$, for which the Crank-Nicholson method (see Malinen and Havu (2007)) can be written as

$$\begin{cases} \rho\frac{\xi^n-\xi^{n-1}}{\Delta t} &= \frac{\mu^n+\mu^{n-1}}{2}, \\ \mathbf{M}\frac{\mu^n-\mu^{n-1}}{\Delta t} &= -\mathbf{K}\frac{\xi^n+\xi^{n-1}}{2} - \mathbf{R}\frac{\mu^n+\mu^{n-1}}{2} + \mathbf{b}(t_n), \end{cases}$$

To avoid inverting an ill-conditioned matrix, this system of equations is written as a double recursion instead of a $2(N+1)$-sized matrix equation. Thus, we eliminate μ^n from the lower equation obtaining the *update equations*

$$\begin{cases} \left(\frac{\Delta t}{2}\mathbf{K} + \frac{2\rho}{\Delta t}\mathbf{M} + \rho\mathbf{R}\right)\xi^n &= \left(-\frac{\Delta t}{2}\mathbf{K} + \frac{2\rho}{\Delta t}\mathbf{M} + \rho\mathbf{R}\right)\xi^{n-1} + 2\mathbf{M}\mu^{n-1} + \Delta t\mathbf{b}(t_n) \\ \rho\xi^n - \frac{\Delta t}{2}\mu^n &= \rho\xi^{n-1} + \frac{\Delta t}{2}\mu^{n-1} \end{cases} \tag{3.19}$$

so ξ^n is first solved from the first equation by matrix inversion, and it is then inserted to the second equation from which μ^n is solved.

The time steps are the same that are used in solving the equations of motion and the ODE for v_o, so we readily have the value $v_o(t_n)$, which is needed in the evaluation of $\mathbf{b}(t_n)$. Since the time step is constant except on the steps when the glottis closes or opens, the inverse of the matrix on the left hand side of the first equation in (3.19) is pre-computed in order to make the simulation faster. When time step is not constant, the matrix equation must be solved separately.

3.2.4 Resonance model

Before performing any time domain simulations, we shall compute the formant frequencies from the Webster's equation. This will be done first for an uncurved tube and then for a curved one.

The resonances of the Webster's equation can be solved by finding the discrete frequencies λ and their corresponding eigenfunctions (pressure distributions) $[\psi_\lambda(s), \pi_\lambda(s)]^T$ satisfying

$$\left\{ \begin{array}{rcl} \mathbf{L}\begin{bmatrix}\psi_\lambda \\ \pi_\lambda\end{bmatrix} & = & \lambda \begin{bmatrix}\psi_\lambda \\ \pi_\lambda\end{bmatrix}; \\ \mathbf{G}\begin{bmatrix}\psi_\lambda \\ \pi_\lambda\end{bmatrix} & = & 0. \end{array} \right. \tag{3.20}$$

The time harmonic extension $\begin{bmatrix}\psi_\lambda(s,t) \\ \pi_\lambda(s,t)\end{bmatrix} = e^{\lambda t}\begin{bmatrix}\psi_\lambda(s) \\ \pi_\lambda(s)\end{bmatrix}$ of the eigenfunction clearly satisfies Eq. (3.12). Thus, the imaginary part of λ is an (angular) resonance frequency.

Again, by writing the weak formulation for Eq. (3.20), setting the control to zero and applying spatial discretization, we obtain a generalized matrix eigenvalue problem

$$\mathbf{K}\mu_\lambda = \lambda^2 \rho \mathbf{M}\mu_\lambda. \tag{3.21}$$

In order to be able to compare these frequencies with those given by the 3-D wave equation (computed in Hannukainen *et al.* (2007)), we have used the Dirichlet boundary condition at the mouth here. This explains the absence of the damping matrix $\mathbf{R}$. Also $\mathbf{K}$ and $\mathbf{M}$ are $N \times N$ matrices instead of $(N+1)\times(N+1)$ as in Eq. (3.17). If the number of elements N is high enough, the eigenvalues of the discretized system are good approximations of the eigenvalues of the original system, especially in the case of the smallest eigenvalues. In our simulations we have used $N = 100$.

3.2.5 Data

We shall use the MRI data provided by Olov Engwall from KTH, Stockholm. The raw data was collected from a native male Swedish speaker pronouncing a prolonged vowel [ø:] in supine position. Engwall and Badin (1999) describe the MR imaging procedure and also present the corresponding formant measurement data.

The same data was also used in a 3-dimensional wave equation model by Hannukainen *et al.* (2007). For this reason, we can compare the 1-dimensional Webster's equation to the actual 3-dimensional wave equation — at least in frequency domain.

The MRI data consists of 29 cross-sectional slices of the vocal tract. However, the slices were not perpendicular to the centerline of the tract, so the slices could not be used as such. First, we determined the centerline of the vocal tract by connecting the centers of mass of each slice. Then, a tangent vector of the path was numerically evaluated at all 29 points, and the slices were projected on the plane perpendicular to this tangent vector. The processed data is shown in Fig. 3.1. The areas of each of the slices were then calculated as well as the circumferences which are needed for evaluating the hydraulic radius as described in Section 2.2. The cross-sectional area is shown in Fig. 3.2 and the hydraulic

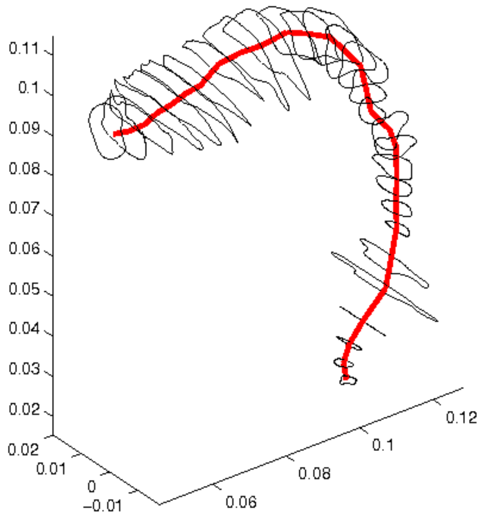


Figure 3.1: The processed MRI data used for constructing the VT-model and the centerline of the tract. The units are in meters. The mouth is at the top left corner and the glottis at the lower right corner.

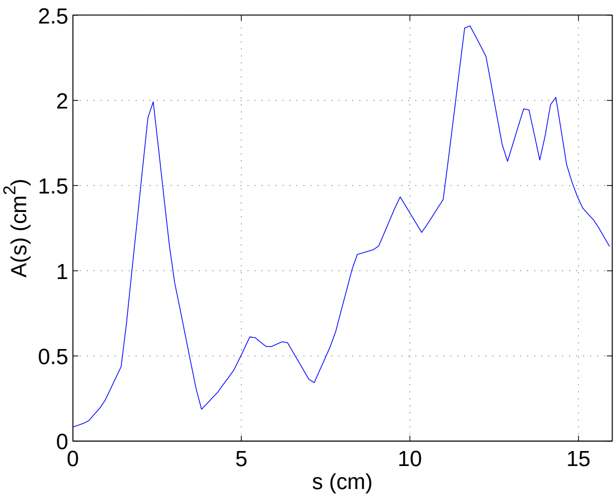


Figure 3.2: The cross-sectional area of the VT, perpendicular to the VT centerline. The s-axis is parameterized as the distance from the glottis measured along the centerline.

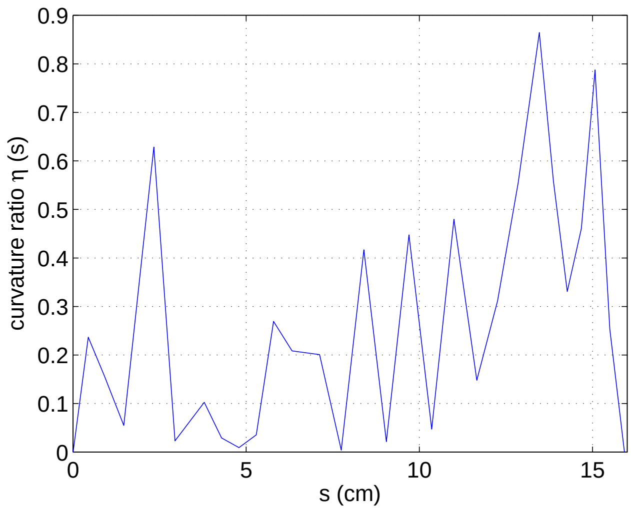


Figure 3.3: The curvature ratio of the VT-tube.

radius in Fig. 3.4. In these figures, the s-axis is parameterized by the arch length of the centerline.

Finally, the sound speed correction factor was computed by numerically evaluating the curvature $\kappa(s)$ of the centerline and approximating the tube radius by $R(s) = \sqrt{A(s)/\pi}$. Fig. 3.3 shows the curvature ratio $\eta(s)$ along the VT. Let us note that the curvature ratio is always distinctly less than one, as assumed in the derivation of the Webster's equation. Even though the curvature ratio varies a lot along the VT, the sound speed correction factor $\Sigma(s)^{-2} = 1 + \frac{1}{4}\eta(s)^2$ varies between 1 and 1.132.

The value for the normalized acoustic resistance θ (see Eq. (3.1) and the expression of $\mathbf{R}_{ij}$ in Eq. (3.18)) was experimentally chosen to be 0.06. There are many approaches in the literature for the VT termination and most of these produce a frequency dependent (and complex) impedance.

One approach is to use the impedance for a piston-like source set in a sphere. This kind of model yields an analytical expression in form of an infinite series. For this reason, a more widely used model is obtained by letting the ratio of the radii of the piston and the sphere approach zero corresponding to a piston set into an infinite wall. Then the acoustic resistance factor θ with low frequencies is approximately $\frac{\omega^2 r_m^2}{2c^2}$ where ω is the angular frequency of the acoustic radiation and r_m is the radius of the piston. Both of these approaches are treated in *e.g.* Morse and Ingard (1968) (Chapter 7). Our choice for θ corresponds to source frequency of around 2200 Hz.

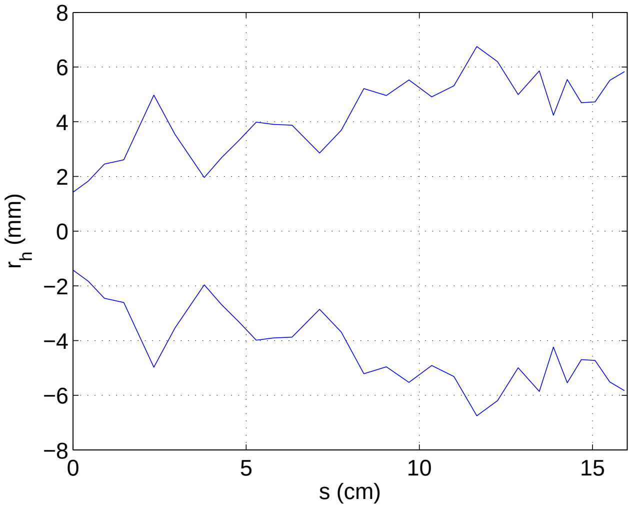


Figure 3.4: The hydraulic radius of the vocal tract by equation (2.8) and the data presented in Section 3.2.5.

3.2.6 Results

The lowest formant frequencies F1,...,F4 for an uncurved ($\Sigma(s) \equiv 1$) and a curved tube are presented in Table 3.1. For comparison, there are also the corresponding frequencies from a 3-D wave equation model by Hannukainen *et al.* (2007) and the formants measured by Engwall and Badin (1999) from the same test subject. To make the comparison reasonable, we have used the Dirichlet boundary condition at mouth as in Hannukainen *et al.* (2007).

Our principal purpose is to compare the Webster's equation to the 3-D wave equation. These formants are very close to each other. However, for some reason, the uncurved tube seems to be even better than the curved tube. Some of the reasons for the discrepancy between the computed and measured formants is discussed in Hannukainen *et al.* (2007).

Table 3.1: Formants for [ø:] in kHz, from our Webster's equation in an uncurved and a curved tube, from the 3-D wave equation by Hannukainen *et al.* (2007) and formants measured by Engwall and Badin (1999).

	F1	F2	F3	F4
Webster, uncurved	0.66	1.35	2.68	3.76
Webster, curved	0.64	1.32	2.64	3.71
HLMP07	0.68	1.35	2.71	3.79
EB99	0.50	1.06	2.48	3.24

Fig. 3.5 shows the pressure distributions (element approximations of $\pi_\lambda = \sum_{k=1}^{N+1} \mu_{\lambda,k} v_k(s)$, see Eq. (3.21)) related to formant frequencies F1,...,F4. These are computed for the curved tube, but here the difference between the curved and uncurved tube was insignificant. When comparing this with the corresponding figure (Fig. 2) in Hannukainen *et al.* (2007), it is very difficult to see any difference. This could be expected because the pressure varies mainly in the direction of the VT. However, they report a weak cross-mode resonance in the oral cavity related to F4. This kind of phenomenon is, of course, not accounted for by a 1-D model such as the Webster's equation.

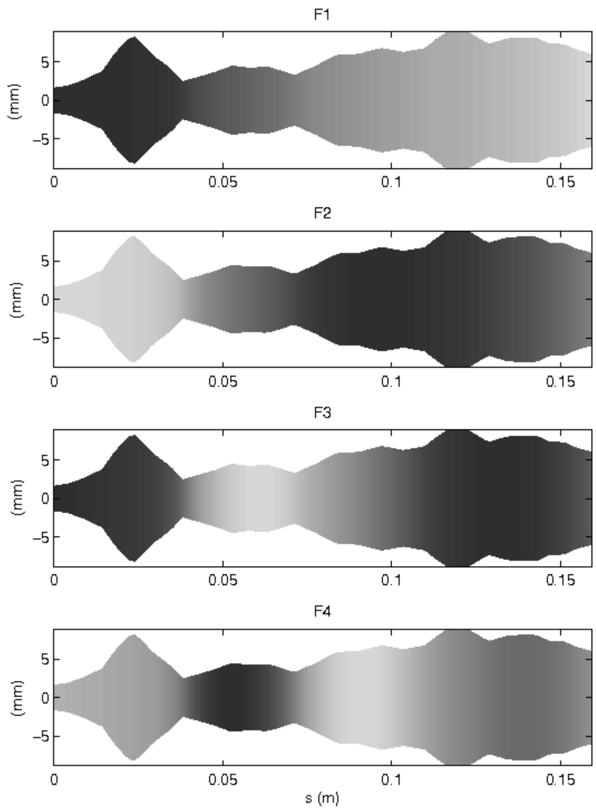


Figure 3.5: Pressure distributions corresponding to formants F1,...,F4.

Chapter 4

Full model simulations

In this chapter we shall present the results of time domain simulations of the full model. In Section 4.1 the model is simulated as a feedforward model so that the VT model is simply excited with the glottis pulse and the pressure at the lips is observed. In Section 4.2.1 we shall investigate the effect of the mechanical feedback from the VT to the glottis introduced in Section 2.3.1. Finally, in Section 4.2.2 the glottis model is coupled to a tube with constant area function. The length of this tube is varied for tuning the lowest formant frequency.

4.1 Simulations without feedback

First simulation was performed with the same parameters as the first glottis model simulation (Fig. 2.4). That is, symmetric glottis parameters and the fundamental frequencies of the vocal fold vibrating modes were 100 Hz and 105 Hz. The result is shown in Fig. 4.1. The top picture shows the volume flows through the glottis and the mouth. Note that the acoustic vibration does not proceed through the open glottis but the flow there is fully determined by the glottal flow model. The pressure at mouth opening is shown in the second picture and the spectrum of this signal in the third picture in Fig. 4.1. The lowest picture shows the spectrum of the glottal flow. The spectra contain peaks at frequencies mF_0, where m is an integer and F_0 is the vocal fold oscillation frequency (118 Hz). The VT formant frequencies cannot be seen as such, but in the speech signal spectrum the peaks that are close to formant frequencies are clearly amplified. For example the first formant frequency $F_1 = 640$ Hz is between the peaks at $5F_0$ and $6F_0$. Between every multiple of F_0 there are five subharmonics with intervals of 16.8 Hz in both spectra.

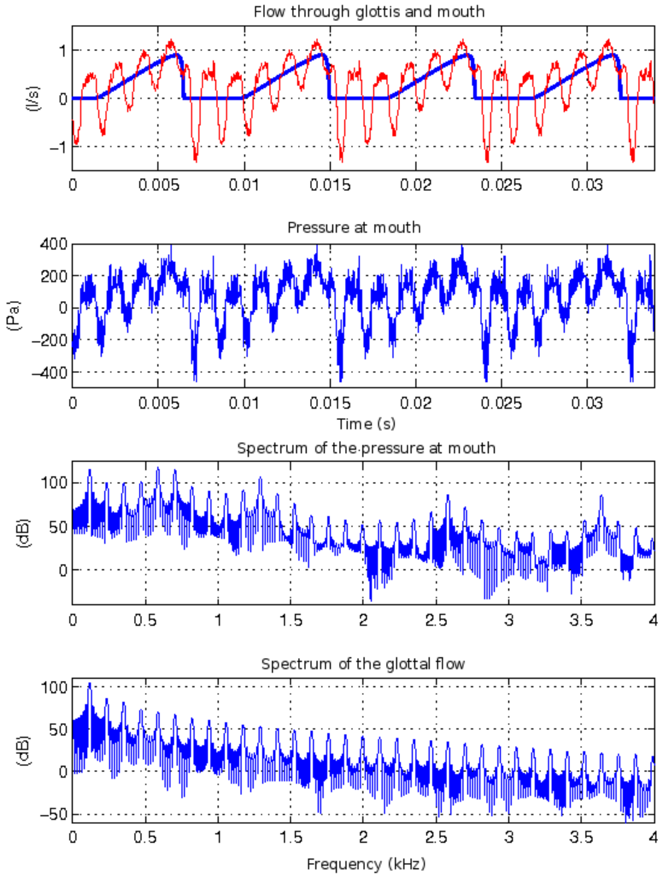


Figure 4.1: The volume flows through the glottis (VT-model input) and through the mouth and the pressure at mouth in a feedforward simulation. Below there are the spectra of the pressure at mouth and the glottal flow.

4.1.1 Inverse filtering the obtained signal

In order to validate our model, the pressure at mouth was inverse filtered by *iterative adaptive inverse filtering* (IAIF) method developed in Alku (1992). For this we used a MATLAB-based toolkit, TKK Aparat (see Airas (2008)). This method estimates the VT transfer function in an iterative manner using all-pole modelling. This transfer function is then used together with a lip radiation model for inverse filtering.

in the Aparat the maximum number of formant frequencies to be modelled by the vocal tract filter can be chosen by the user as well as the value of the first order lip radiation model. Fig. 4.2 shows the glottal flow given by our model and the inverse filtered signal. In the transfer function estimation we set the maximum number of the formants to be fitted to 11 and the lip radiation coefficient to 0.97. Table 4.1 shows the values of the estimated VT formants, that were below one half of the sampling frequency (here 19 kHz), and those computed from the Webster's equation with boundary conditions (3.11). Note that the formants in Table 3.1 were computed using Dirichlet boundary condition at mouth, which explains the small discrepancy between these two. With greater values of the acoustic resistance coefficient θ this discrepancy obviously grows. Three lowest formants are estimated rather well, whereas the rest are systematically smaller.

The glottal flow estimated by inverse filtering seems to have problems in capturing the rapid ending of the pulse. The reason for this is that rapid changes in the signal correspond to higher frequencies in the spectrum. Since there seems to be a systematic error in the estimated transfer function related to the higher formants, it can be expected that these changes cause error in the inverse filtering procedure.

Table 4.1: Formants for [øː] in kHz given by our Webster's equation and formants estimated by the IAIF method

	F1	F2	F3	F4	F5	F6	F7	F8	F9
Webster	0.65	1.31	2.65	3.71	5.15	6.81	7.23	8.30	9.23
Estimated	0.66	1.32	2.61	3.65	5.06	6.47	6.84	7.73	8.62

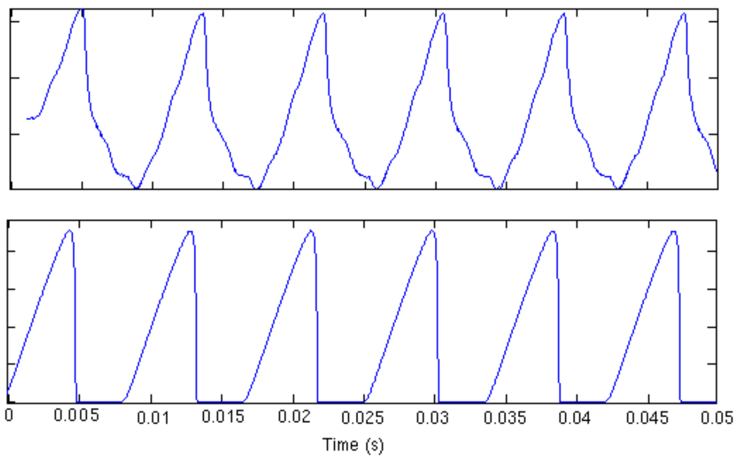


Figure 4.2: The glottal flow obtained by inverse filtering and the flow given by our model

4.2 Simulations with feedback

4.2.1 Realistic VT-geometry

The mechanically coupled counter pressure from the vocal tract always seemed to have a damping effect on the vocal cords. For this reason we had to diminish the glottal damping terms b_{ij} from 0.1 Nm/s in the feedforward system to 0.065 Nm/s in the feedback system to sustain continued oscillation. As before, this value was found experimentally. Other parameters were first kept the same as earlier.

Here the effect of the feedback is rather mild, even so, that the different situations could not be identified only by observing the pressure at mouth. A small ripple can be seen in the glottal area function, but the glottal flow pulse is not very sensitive to this ripple. The spectrum of the glottal flow is not influenced by the feedback. A slight additional skewing of the pulse can be observed. If we calculate the ratio of the pulse acceleration time to the whole open phase duration, that is $\frac{T_{max}-T_1}{T_2-T_1}$ (see Section 2.4.4) it is 90.6 % for the system without feedback and 91.5 % for the system with feedback. The change in the glottis dynamics is illuminated in Fig. 4.3, which shows the phase diagrams of the glottal oscillation in two cases. The upper left picture

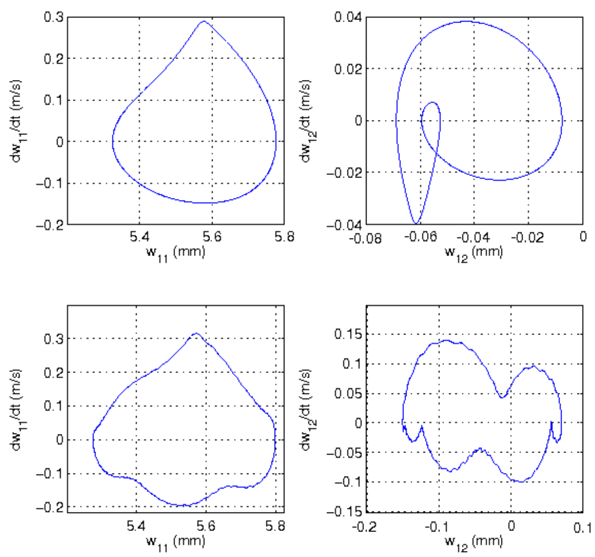


Figure 4.3: The phase diagrams of the glottal oscillation from a simulation without feedback (top) and a simulation with feedback (bottom). Pictures on the left show the behaviour of the cords in the narrow end of the glottis and pictures on the right show the behaviour of the cords in the wide end.

shows the curve $(w_{11}(t), \dot{w}_{11}(t))$ and the upper right picture shows the curve $(w_{12}(t), \dot{w}_{12}(t))$ in the simulation without feedback. In the lower pictures there are the same curves in the simulation with feedback. In both cases the oscillation is perfectly periodic, meaning that the cycles in the phase diagrams are stable. The vibration pattern of w_{12} changes significantly when the feedback is present, but this could be expected since the aerodynamic force is much weaker in the wide end of the glottis thus making the feedback more influential.

4.2.2 Straight tube as the resonator

More interesting is what happens when the vocal fold vibration frequency F_0 is closer to the lowest formant frequency F_1 or when $2F_0 \approx F_1$. Here this effect is studied by using an uncurved tube shown in Fig. 4.4 as the resonator. The area of the tube at $s = 0$ is chosen so that it coincides with $A(0)$ of the realistic geometry used earlier. The area after the expansion is the same as the area of mouth. Also the boundary conditions in both ends of the tube were the same as in the earlier simulations with the realistic VT geometry (Eq. (3.11)).

Two sets of simulations were performed. In the first one, the glottis model parameters were the same as earlier. The tube length was varied between $0.20\ m$ ... $0.71\ m$ thus spanning the frequency range $123.2\ Hz$ ... $438.8\ Hz$ covering three multiples of the source frequency F_0. These tube lengths are rather unrealistic considering human VT, but the sole purpose of this experiment is to study the feedback effect when F_0 and F_1 are close to each other. By varying only the tube length we can exclude any internal changes in the glottis so that all changes in the glottal vibration pattern are caused by the coupling. In reality, $F_0 - F_1$ crossovers can occur, but obviously with higher source frequencies F_0 (see Titze *et al.* (2008)).

The spectrogram with different values of F_1 is shown in Fig. 4.6. This is a slightly nontypical spectrogram, because the x-axis variable is not time, but the formant frequency F_1. All simulations are independent with default initial conditions. The simulations have been long enough and the beginning of each simulation has been excluded from the data, so that there is no effect

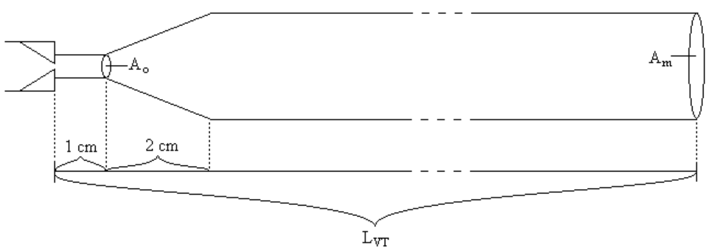


Figure 4.4: The geometry for testing the feedback effect for different resonator formant frequencies F_1. The length of the tube was varied for tuning F_1.

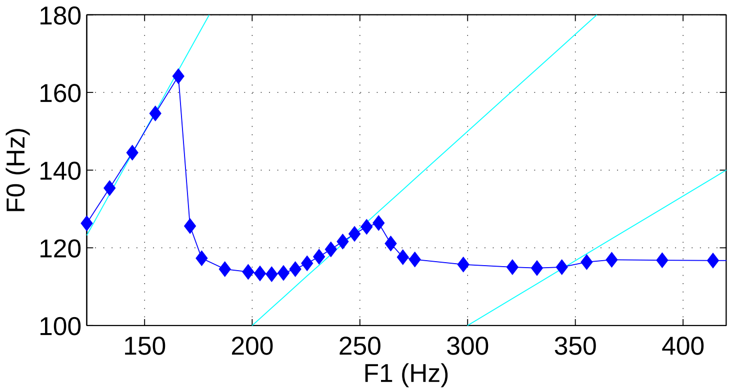


Figure 4.5: The source frequency F_0 as a function of the lowest VT formant frequency F_1 in the first set of simulations with the straight tube. The auxiliary lines are $F_0 = F_1$, $2F_0 = F_1$ and $3F_0 = F_1$.

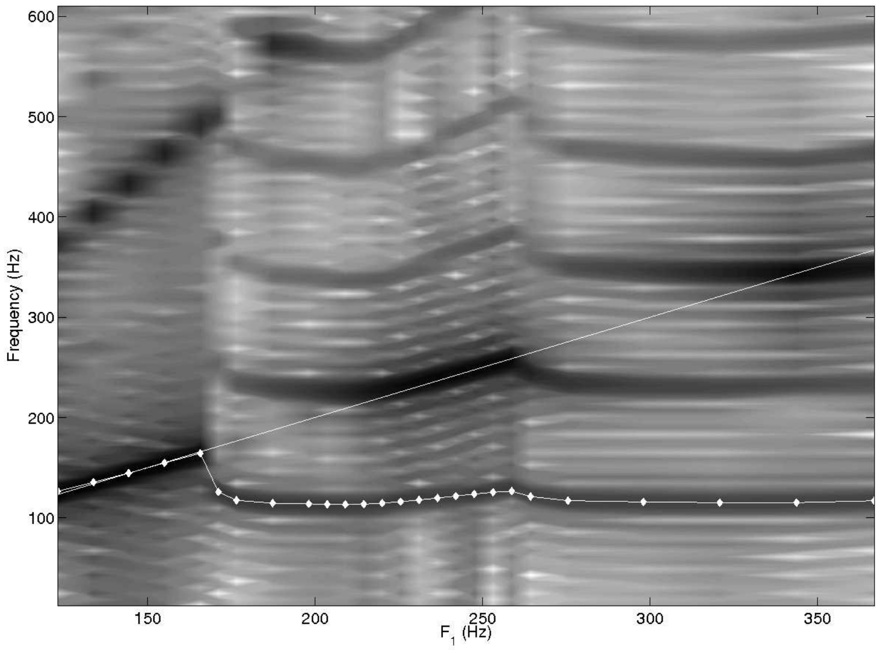


Figure 4.6: The spectrogram of the pressure signal when F_1 is varied. The line shows F_1 in the spectra and the diamonds show the source frequency F_0 in each simulation.

from initial transitions. One simulation corresponds to one tube length. The figure also shows the source frequencies F_0 in every simulation and an auxiliary line showing F_1 in the spectrogram. The source frequencies are also plotted in a more illustrative scale in Fig. 4.5 with lines $F_0 = F_1$, $2F_0 = F_1$ and $3F_0 = F_1$. This picture clearly shows what happens to F_0 when F_1 crosses some of its multiples. When $F_1 \approx F_0$, the source frequency locks in to the formant frequency, until it gets too far from the natural source frequency. A similar but weaker phenomenon can be seen when $2F_0 \approx F_1$ (and also when $3F_0 \approx F_1$).

In the second set of simulations the glottis model was tuned so that its natural frequency was higher (233 Hz). This was achieved by increasing the stiffness coefficients to $k_{11} = k_{21} = 682\ N/m$ and $k_{12} = k_{22} = 379.5\ N/m$. Also the glottal gap was narrowed to $g = 0.2\ mm$ and the subglottal pressure was increased to 1800 Pa. Now the tube length was varied between 0.16 m ... 0.50 m so that the corresponding F_1 frequency range was 175 $Hz, ..., 560\ Hz$ covering two multiples of F_0. The source frequency's dependence on F_1 is shown in Fig. 4.7. Now F_0 remains locked in to F_1 much longer and F_0 climbs as high as 400 Hz. After the frequency drop, F_0 settles on a level about 20 Hz higher than before the "climb". The effect of $2F_0 - F_1$-crossover is milder now than in the first set. The bump in F_0 in this crossover is here only about 8.5 Hz compared to 13 Hz in the first set.

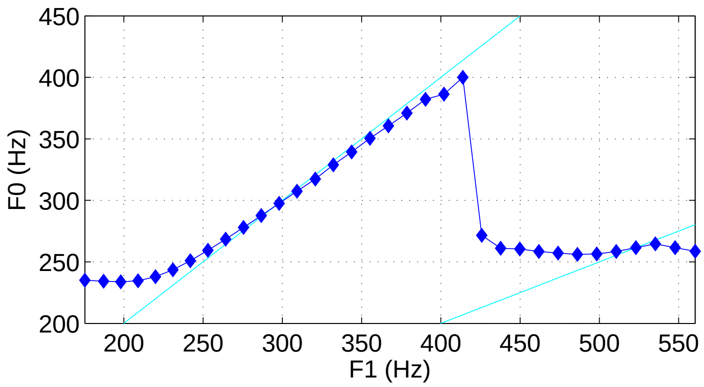


Figure 4.7: The source frequency F_0 as a function of the lowest VT formant frequency F_1 in the second set of simulations with the straight tube. The auxiliary lines are $F_0 = F_1$ and $2F_0 = F_1$.

4.3 Comparison to other works

Titze (2008) has created a nonlinear source-filter coupling theory and Titze *et al.* (2008) created three vocal exercises for human test subjects for studying this coupling in practise. They reported that when the interaction between the source and filter is mild, that is, when the dominant source frequency lies well below the lowest VT formant frequency, the effect of the coupling can be seen in the glottal flow pulse skewing and pulse ripple. Our results are well in line with this observation (see Section 4.2.1 and Fig. 4.3). When the source frequency F_0 and the lowest formant frequency F_1 are closer to each other (or even when $2F_0 \approx F_1$), the feedback can cause a sudden jump in the source frequency. Our model reveals a *synchronization* phenomenon. This means, that when F_1 is close to F_0 or some of its multiples, the oscillation frequency of the mass-spring system changes so that the two systems are synchronized. This synchronization could contribute to some of the phenomena reported by Titze *et al.* (2008) (Figs. 5C and 10D).

They also reported two other kinds of bifurcations besides frequency jumps, namely *subharmonic regimes* (spectral peaks at frequencies $\frac{k}{2}F_0,\ k = 1, 3, 5, \dots$) and *chaotic oscillation*. Our model reveals five subharmonics between every multiple of F_0, and they are stronger near $2F_0 - F_1$ crossover but not remarkably (see Fig. 4.6). Chaotic oscillation never occurred in our simulations, even when the subglottal pressure p_{sub} was increased up to 3300 Pa, or when the glottis model parameters were set unsymmetric ($m_{21} = 1.2 \cdot m_{11}$).

Hatzikirou *et al.* (2006) have also created a similar two mass model of glottis and simulated it with a tube of varying length as the acoustic load. They also report *frequency pulling* by F_1. In addition, the subharmonics occur much clearer in their simulations as they do here.

Chapter 5

Discussion

Chapter 2

The primary target of this work was to construct a low order nonsymmetric mass-spring model with a 1-D flow model. This task was carried out in Chapter 2. The used flow model takes into account viscous pressure losses in the glottis and VT. The vocal tract inertance is also included in the flow equation, Eq. (2.12). However, the flow pulse (Fig. 2.7) seems to be slightly too much skewed towards the end of the open phase. Reasons for this lie in our harsh assumptions that the flow is laminar and incompressible. Because of the laminarity assumption, the pressure loss in the glottis and vocal tract given by our model is likely to be smaller than in reality. This is because turbulent flow and excluded phenomena on the tissue surface (*e.g.* mucosal vibrations) might cause energy dissipation to heat.

The incompressibility assumption has an effect on the inertia of the air column in the VT. (coefficient C_{iner} in Eq. (2.12)). Because the flow is, in fact, compressible, there is hidden spring reaction which would temporally divide the change in momentum in a different way. For this reason the inertia coefficient in the model may appear too large. Also the pressure loss in the VT effectively grows, if the incompressibility assumption is omitted.

Instead of constructing a dynamical compressibility model, the inertance C_{iner} and the pressure loss coefficients C_g and C_{VT} could be fitted in an optimal way so that the pulse would match as well as possible the glottal pulses obtained by inverse filtering. This procedure is illustrated in Fig. 5.1. It shows an LF-model pulse which was obtained by first inverse filtering with the IAIF method a natural [a] vowel, produced by a male speaker using pressed phonation. The LF-model parameters were obtained using the Aparat toolkit. In creating pressed speech, subjects typically increase adduction of their vocal folds, hence resulting in a glottal flow with long closed phase and a short closing phase.

The parameters for the modelled pulse are obtained by creating a pulse using Eq. (2.12) and approximating $\Delta W_1 = A\sin(\omega t)$, where $t \in [0, \pi/\omega]$. The squared error between this pulse and the LF-pulse was minimized by adjusting the parameters C_{iner}, C_g and C_{VT} in the flow equation. This was done by using MATLAB's built-in command *fminsearch*. However, after optimal parameter estimation, the pressure loss due to the flow through mouth and the pressure loss in the VT turned out to be negligible. After omitting these terms we are left with three parameters but only two terms in the flow equation. This means that same pulse is obtained with infinitely many parameter combinations. A reasonable combination minimizing the squared error is $p_{sub} = 550\ Pa$, $C_{iner} = 2.35 \cdot 10^3\ kg/m^4$ and $C_g = 8.24 \cdot 10^{-9}\ Ns$. The values computed earlier are $C_{iner} = 3.30 \cdot 10^3\ kg/m^4$ and $C_g = 8.22 \cdot 10^{-11}\ Ns$.

The pulses in Fig. 5.1 are very close to each other. This suggests that the LF-pulse can be faithfully constructed with a crude physical model.

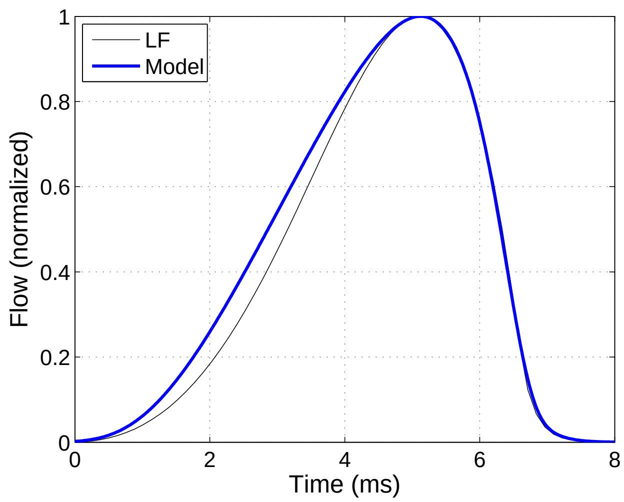


Figure 5.1: The modelled pulse and an LF-pulse obtained by inverse filtering [a] vowel produced by male speaker using pressed phonation

Chapter 3

The VT model was presented in Chapter 3. First, such a variant of the Webster's equation was presented, that includes a contribution due to the tube curvature. This variant is derived in a manuscript Lukkari and Malinen (2008b). Then the state space was discretized by a FE method, using the physical energy norm of the state space. Crank-Nicholson discretization was applied in the time variable.

A secondary purpose of this work was to compare the spectral properties of the Webster's equation (with and without curvature) with the 3-D wave equation. Hannukainen *et al.* (2007) computed the formant frequencies by using the 3-D wave equation. The data for the Webster's equation, that is, the VT cross-sectional area function and the curvature of the VT centerline, were obtained from the same MRI data that was used also by Hannukainen *et al.* (2007).

It was noticed that the formant frequencies given by both variants of the Webster's equation were quite close to the formants given by the wave equation. The four lowest formants given by the uncurved Webster's equation were on average 1.2 % lower, and the formants from the curved equation were on average 3.2 % lower than the formants given by the wave equation (see Table 3.1). This difference could be explained by the choice of the VT centerline as the center of mass of each cross-section. This way the tube might become effectively longer, than in the 3-D equation, because in the 3-D geometry a wave propagating in the VT can "take a shortcut" in the curves of the tube. In the Webster's equations case, this is of course impossible. Furthermore, since the sound speed correction factor in the Webster's equation with curvature is always less than or equal to one, the curvature factor in the equation makes the tube effectively even longer, which lowers the formant frequencies even more. This questions the usability of the Webster's equation with curvature as such, at least in acoustic applications. When the curvature ratio is small, the Webster's equation with curvature becomes more accurate but the effect of the curvature is negligible. When the curvature ratio is greater, the curved equation fails to describe the curvature effect correctly.

One possible way to fix the situation is to scale the total length of the tube. By dimension analysis, the formant frequencies would then be scaled similarly. So instead of studying the absolute values of the formant frequencies, we should compare the relative frequencies F_n/F_1. However, these were very close to each other for both curved and uncurved case so we cannot make any conclusions based on these computations. In addition, we cannot use the higher formant frequencies because the formants from the wave equation are distorted (upwards) by the crossmode resonances. Also the geometry used for constructing the data is certainly not exactly such as it is assumed in the derivation of the Webster's equation. That is, the tube cross-sections are not circular.

Another shortcoming of our model is the lack of dissipative terms in the vocal tract. The physical interpretation of the Neumann boundary condition at the walls of the VT is that the material of the tube walls is absolutely inflexible. In reality, the walls of the VT are elastic and the vibration of the air is transmitted to the tissue causing dissipation at walls. Viscous losses are not included in the Webster's equation either. Thus the only dissipation in the model is the flow resistance at lips: $p_{res} = \theta \rho c v_m$. The normalized acoustic resistance θ is here more or less arbitrarily chosen, but refining the model here by physical considerations would be rather useless as long as other dissipation is excluded.

Chapter 4

The results of the full model simulations are shown in Chapter 4. The model output seems all right and the spectra of both mouth pressure signal and the glottal flow are believable. The results were also well in line with earlier findings: when the source frequency F_0 is well below the lowest VT formant frequency F_1, the VT feedback effect is rather weak. Only when the frequencies were close to each other, the feedback caused bifurcations in the source vibration. However, the only type of bifurcation revealed by our model was the source frequency lock-in to the formant frequency F_1 when F_1 approached F_0 or when $F_1 \approx 2F_0$.

Experimental studies by Titze *et al.* (2008) revealed also subharmonic and, with some test subjects, even chaotic regimes at these frequency crossovers. In their experiments the frequencies F_0 and F_1 changed dynamically, so that it is impossible to say whether these phenomena were steady or only transitional. Our simulations for studying the feedback effect in Section 4.2.2 were separate for different values of F_1 so that no transitional bifurcations can be detected. In reality, there are of course other phenomena besides the VT acoustics that can have an effect on the glottal behaviour.

Hatzikirou *et al.* (2006) also performed similar simulations as we in Section 4.2.2, but with such a feedback configuration that the VT feedback had a direct effect on the glottal flow. Their model revealed subharmonics in the spectrum of the position of one mass in their mass spring model of the glottis.

Bibliography

Airas, M. (**2008**). “TKK Aparat: An environment for voice inverse filtering and parameterization,” Locopedics Phoniatrics Vocology **33**, 49–64.

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

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

Chiba, T. and Kajiyama, M. (**1941**). The Vowel, its Nature and Structure, Tokyo-Kaiseikan Publishing Company Ltd.

Engwall, O. and Badin, P. (**1999**). “Collecting and analysing two- and three-dimensional mri data for swedish,” Tech. rep., QPRS: Dept. for Speech, Music and Hearing, Stockholm.

Fant, G. (**1979**). “Glottal source and excitation analysis,” Tech. Rep. 1, QPRS: Dept. for Speech, Music and Hearing, Stockholm.

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

Fetter, A. and Walecka, J. (**1980**). Theoretical Mechanics of Particles and Continua, McGraw–Hill.

Flanagan, J. L. (**1972**). Speech Analysis Synthesis and Perception, Springer-Verlag.

Hannukainen, A., Lukkari, T., Malinen, J., and Palo, P. (**2007**). “Vowel formants from the wave equation,” Journal of Acoustical Society of America Express Letters **122**.

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

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

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

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

Landau, L. D. and Lifshitz, E. M. (**1970**). Theory of Elasticity, Pergamon press.

Lukkari, T. and Malinen, J. (**2008**a). "Webster's equation as a conservative linear system," Manuscript, 22 pp .

Lukkari, T. and Malinen, J. (**2008**b). "Webster's model with curvature and dissipation," Manuscript, 16 pp .

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

Malinen, J. and Staffans, O. (**2006**). "Conservative boundary control systems," Journal of Differential Equations **231**, 290–312.

Malinen, J. and Staffans, O. (**2007**). "Impedance passive and conservative boundary control systems," Complex Analysis and Operator Theory **1**, 279–300.

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

Sváček, P. and Horáček, J. (**2006**). "Numerical modelling of effect of tonsillectomy on production of Czech vowels," Acta Acustica united with Acustica **92**, 681–688.

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

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

Appendix A

The MATLAB-code

The structure of the code is such that there are two initialization files, init.m and VTdata.m. The init.m-script has to be run before every simulation. The file VTdata.m only needs to be run once unless changes are made. Unfortunately the original VT data can not be provided.

The file solver.m does the simulation. It calls functions ff.m, which is the time derivative of the state vector (from the equations of motion of the glottis), NewV.m, which computes the glottal flow and interpol.m which performs the interpolation as described in Section 2.4.1.

A.1 File init.m

```
%In this file the physical parameters of the (glottis) model are
%initialized

global d step rho width L mu;

%===SIMULATION PARAMETERS===

NumIts=5000;            %Number of iterations
step=0.00002;           %Time step length
N=100;                  %Number of discretization points in the VT
fb=1;                   %Feedback on (1) or off (0)
continue=0;             %Continue previous simulation? 1/0

%===PHYSICAL PARAMETERS===

rho=1.2;                %Air density
c=343;                  %Speed of sound
mu=18.7e-6;             %Dynamic viscosity of air
H0=11.2*10^-3;          %Height of the subglottal channel
```

```
width=18*10^-3;         %Width of the subglottal channel
L=6.8*10^-3;            %Length (or thickness) of the glottis
Kh=730;                 %Retrieving stiffness in vocal cord contact

aa=0.85;      %The x-coordinates of the springs are aa*L and bee*L
bee=0.15;

theta=.06;              %Mouth resistance coefficient in VT boundary condition

% When glottis is narrower than this, it is closed and flow is set to
% zero (epsilon in the report)
d=2.5e-5;

%---Parameters for cord #1---
%Masses
m11=1.686e-4;
m12=0.595e-4;
m13=2.531e-4;

%Stiffness coefficients
k11=124;
k12=69;

%Damping coefficients
b11=.065;
b12=.065;

%The equilibrium state when there is no flow
Y_110=5.4*10^-3;
Y_120=0;


%---Parameters for cord #2---
%Masses
m21=m11;
m22=m12;
m23=m13;

%Stiffness coefficients
k21=k11;
k22=k12;

%Damping coefficients
b21=b11;
b22=b12;
```

```
%The equilibrium state when there is no flow
Y_210=5.8*10^-3;
Y_220=H0;

%===INITIAL CONDITIONS===

%---Initial values for the glottis---
if continue > .5
    xend=x(1:8,end);       %Storing the final state of previous
end                        %simulation
x=zeros(9,NumIts+1);       %The solution points are stored here

if continue > .5
    %Continuing previous simulation
    x(1:8,1)=xend;
else
    %Glottis initially closed (determined by simulating with constant flow)
    x(1,1)=0.00574388213041;
    x(2,1)=0.07929750365587;
    x(3,1)=-0.00013732214913;
    x(4,1)=-0.03349122869515;
    x(5,1)=0.00545611786959;
    x(6,1)=-0.07929750365587;
    x(7,1)=0.01133732214913;
    x(8,1)=0.03349122869516;
end

%---Initial value for the glottal flow---

if continue > .5
    Vend=Vout(end);
else
    Vend=0;
end
Vout=zeros(NumIts+1,1);
Vout(1)=Vend;

%---The initial state for the VT---

if continue < .5
    xi=zeros(N+1,1);
    eta=zeros(N+1,1);
end

%---Forming the mass, stiffness and damping matrices---
M1=[m11+m13/4,m13/4;m13/4,m12+m13/4];
```

```
M2=[m21+m23/4,m23/4;m23/4,m22+m23/4];
M1_inv=inv(M1);
M2_inv=inv(M2);
K1=[aa^2*k11+bee^2*k12, aa*(1-aa)*k11+bee*(1-bee)*k12;
   aa*(1-aa)*k11+bee*(1-bee)*k12, (1-aa)^2*k11+(1-bee)^2*k12];
K2=[aa^2*k21+bee^2*k22, aa*(1-aa)*k21+bee*(1-bee)*k22;
   aa*(1-aa)*k21+bee*(1-bee)*k22, (1-aa)^2*k21+(1-bee)^2*k22];
B1=diag([b11 b12]);
B2=diag([b21 b22]);
kerr=L*width*rho*H0^2;  %Auxiliary coefficient
open=y210-y110>d;       %Test whether the glottis is initially closed
```

A.2 File VTdata.m

```
%This file processes the VT data and determines the centerline of the VT,
%the curvature of it and the cross-sectional area function

global Ao Am vakio1 vakio2

%Importing the data and removing the false slices from the mouth. The data
%consists of three matrices containing the X-, Y- and Z-coordinates of the
%VT boundary points. One row contains the information of one slice.

neutral_tract;      %This imports the data
X3D(29,:)=X3D(34,:);
Y3D(29,:)=Y3D(34,:);
Z3D(29,:)=Z3D(34,:);
X3D=X3D(1:29,:);
Y3D=Y3D(1:29,:);
Z3D=Z3D(1:29,:);

%Change of units: cm -> m
X3D=.01*X3D;
Y3D=.01*Y3D;
Z3D=.01*Z3D;

nsl=size(X3D,1);    %Number of SLices
pps=size(X3D,2);    %Points Per Slice

%Initially the VT centerline is determined as the center of mass of the
%boundary points. The Y-coordinate (corresponding to right-left direction)
%is left zero. Other coordinates are X (forward-backward) and Z (up-down)
cpath=zeros(3,nsl);
cpath(1,:)=mean(X3D');
```

```
cpath(3,:)=mean(Z3D');

%In matrix VT the first row contains the parameter s (centerline arch length
%parameter), second row contains the hydraulic radius, third is the cross
%sectional area and fourth is the sound speed correction factor 1/Sigma(s)^2.
VT=zeros(4,nsl);
VT(1,2:end)=cumsum(sum((cpath(:,2:end)-cpath(:,1:end-1)).^2).^.5);

%Then the slices are projected on planes whose normals are tangents of the
%centerline at each slice. In addition, the area and the hydraulic radius
%of each slice are determined. Also the centerline is corrected to match
%the center of mass of the slice.

%First the tangent vectors of the centerline are determined and stored
normals=[cpath(:,2)-cpath(:,1),cpath(:,3:end)-cpath(:,1:end-2),
         cpath(:,end)-cpath(:,end-1)];

for k=1:nsl
    %k:th normal vector is rotated 90 degrees and normalized to unit length
    abu=[normals(3,k);0;-normals(1,k)];
    abu=abu/norm(abu);

    %The previous projection is stored here. Initially it is the last point to
    %be projected (that is pps:th point)
    old_proj=abu'*([X3D(k,pps);Y3D(k,pps);Z3D(k,pps)]-cpath(:,k))*abu+
             [0;Y3D(k,pps)-cpath(2,k);0];

    %This vector is the correction to the centerline at k:th slice
    correction=zeros(3,1);
    for j=1:pps
        %proj is the datapoint projected on the plane with respect to
        %origin at the center of the slice
        proj=abu'*([X3D(k,j);Y3D(k,j);Z3D(k,j)]-cpath(:,k))*abu+[0;Y3D(k,j)-cpath(2,k);0];

        %change of the origin to the original one
        X3D(k,j)=cpath(1,k)+proj(1);
        Z3D(k,j)=cpath(3,k)+proj(3);

        %This "area" is the area of a triangle formed by the center of the slice
        %and points "proj" and "old_proj".
        %This area can be negative if the slice is not convex.
        area=sign(normals(:,k)'*cross(proj,old_proj))*abs(acos(proj'*old_proj/
             norm(proj)/norm(old_proj))/8*(norm(proj)+norm(old_proj))^2);
        VT(3,k)=VT(3,k)+area;

        %The average of "proj" and "old_proj" is weighted with the area of
```

```
        %the triangle and it is added to the "correction"-vector
        correction=correction+area/2*(proj+old_proj);
        %First the circumferences are stored here
        VT(2,k)=VT(2,k)+norm(old_proj-proj);
        old_proj=proj;
    end
    %The "correction" is divided with the whole slice area because of the
    %weighting. The centerline is then corrected.
    correction=2/3/VT(3,k)*correction;
    cpath(1,k)=cpath(1,k)+correction(1);
    cpath(3,k)=cpath(3,k)+correction(3);
end
clear('abu','proj','old_proj')
VT(2,:)=2*VT(3,:)./VT(2,:);     %The hydraulic radius is r_h=2A/C.

%The fourth row of "VT" contains the correction factor for the speed of
%sound. If they are replaced with ones, the tube is assumed uncurved.
kaps=zeros(nsl,1);
for k=1:size(VT,2)-2
    hyp=norm(cpath(:,k+2)-cpath(:,k));
    l1=cpath(:,k+1)-cpath(:,k);
    l2=cpath(:,k+2)-cpath(:,k+1);
    kappa=2*(1-(l1'*l2)^2/norm(l1)^2/norm(l2)^2)^.5/hyp;
    kaps(k+1)=kappa;
end
VT(4,:)=1+VT(4,:)./VT(3,:);
VT(4,:)=1+.25*VT(3,:)/pi.*kaps'.^2;
clear('hyp','l1','l2','normals','kappa','kaps')

%The s-axis is discretized for the FEM-solver
Lvt=VT(1,end);  %Length of the VT (=arch length of the centerline)
ds=Lvt/N;       %Discretization interval

%The data is modified to correspond to this discretization, that is the
%data is interpolated in the points of discretization. This data is stored
%to the matrix "VT2".
VT2=zeros(4,N+1);
VT2(1,:)=ds*(0:N);
VT2(2:4,1)=VT(2:4,1);
VT2(2:4,end)=VT(2:4,end);
for k=2:N
    ind=find(VT(1,:)<=(k-1)*ds,1,'last');
    VT2(2:4,k)=((k-1)*ds-VT(1,ind))/(VT(1,ind+1)-VT(1,ind))*VT(2:4,ind+1)+
               (VT(1,ind+1)-(k-1)*ds)/(VT(1,ind+1)-VT(1,ind))*VT(2:4,ind);
end
clear('ind');
```

```
Ao=VT2(3,1);    %VT area after glottis
Am=VT2(3,end);  %Area of mouth

%Computing the constant C_{iner}
integrandi=1./VT2(3,:);
integraali=(.5*integrandi(1)+.5*integrandi(end)+sum(integrandi(2:end-1)))*ds;
vakio1=rho*Ao*integraali;

%Computing the constant C_{VT}
integrandi=1./VT2(2,:).^4;
integraali=(.5*integrandi(1)+.5*integrandi(end)+sum(integrandi(2:end-1)))*ds;
vakio2=integraali;
clear('integrandi','integraali');

%Finally, the mass matrix M, stiffness matrix K and dissipative matrix R
%corresponding to the boundary condition at mouth

R=sparse(zeros(N+1,N+1));
R(N+1,N+1)=Am/2/(rho*c*theta);

M=sparse(zeros(N+1,N+1));
M(1,1)=1/4*VT2(3,1)*VT2(4,1)+1/12*VT2(3,2)*VT2(4,2);
for k=2:N
    M(k,k)=1/12*VT2(3,k-1)*VT2(4,k-1)+1/2*VT2(3,k)*VT2(4,k)+1/12*VT2(3,k+1)*VT2(4,k+1);
    M(k,k-1)=1/12*VT2(3,k-1)*VT2(4,k-1)+1/12*VT2(3,k)*VT2(4,k);
end
M(N+1,N+1)=1/12*VT2(3,N)*VT2(4,N)+1/4*VT2(3,N+1)*VT2(4,N+1);
M(N+1,N)=1/12*VT2(3,N)*VT2(4,N)+1/12*VT2(3,N+1)*VT2(4,N+1);
M=M+M'-diag(diag(M));
M=ds/2/rho/c^2*M;
M_inv=M^-1;

K=sparse(zeros(N+1,N+1));
K(1,1)=VT2(3,1)/2+VT2(3,2)/2;
for k=2:N
    K(k,k)=VT2(3,k-1)/2+VT2(3,k)+VT2(3,k+1)/2;
    K(k,k-1)=-VT2(3,k-1)/2-VT2(3,k)/2;
end
K(N+1,N+1)=VT2(3,N)/2+VT2(3,N+1)/2;
K(N+1,N)=-VT2(3,N)/2-VT2(3,N+1)/2;
K=K+K'-diag(diag(K));
K=1/2/ds*K;

%The matrix in the update equations are precomputed and -inverted here for
%the time step "step" in order to make computation faster.
```

```
xici=full((step/2*K+2/step*rho*M+rho*R)^-1);
xic=(-step/2*K+2/step*rho*M+rho*R);
```

A.3 File solver.m

```
%This file solves the equations of motion of the glottis, the
%glottal flow ODE and the Webster's equation one step at a time.

op=zeros(NumIts+1,1);   %1/0 glottis open or closed
op(1)=open;
dt=step;
Pm=zeros(NumIts+1,1);   %Vector for pressure at mouth
Vm=Pm;                  %Vector for flow velocity at mouth
Pc=Pm;                  %Counter pressure (feedback)
bhat=zeros(N+1,1);      %FEM-solver load vector

for n=1:NumIts

    %EQUATIONS OF MOTION

    Vf=Vout(n)*Ao/H0/width;   %Subglottal flow velocity
    X=x(1:8,n);               %Current state

    %RK4 steps
    s1=ff(X,fb*.0817*Pc(n),Vf,open,M1_inv,K1,B1,M2_inv,K2,B2,
          Y_110,Y_120,Y_210,Y_220,width,kerr,L,Kh,d);
    s2=ff(X+step/2*s1,fb*.0817*Pc(n),Vf,open,M1_inv,K1,B1,M2_inv,K2,B2,
          Y_110,Y_120,Y_210,Y_220,width,kerr,L,Kh,d);
    s3=ff(X+step/2*s2,fb*.0817*Pc(n),Vf,open,M1_inv,K1,B1,M2_inv,K2,B2,
          Y_110,Y_120,Y_210,Y_220,width,kerr,L,Kh,d);
    s4=ff(X+step*s3,fb*.0817*Pc(n),Vf,open,M1_inv,K1,B1,M2_inv,K2,B2,
          Y_110,Y_120,Y_210,Y_220,width,kerr,L,Kh,d);
    Xnew=X+step/6*(s1+2*s2+2*s3+s4);
    dt=step;

    %Testing whether the glottis closes/opens at current step. If so,
    %then interpolate as described in the report
    if abs((Xnew(5)-Xnew(1)>d)-open)>.5
        open=1-open;
        [Xnew,dt]=interpol(x(:,n-1),[X;x(9,n)],[Xnew;x(9,n)+step]);
    end
    op(n+1)=Xnew(5)-Xnew(1)>0;
    x(1:8,n+1)=Xnew;
    x(9,n+1)=x(9,n)+dt;
```

```
    %GLOTTAL FLOW

    if op(n+1) > .5
         Vout(n+1)=NewV(Xnew(5)-Xnew(1),Xnew(7)-Xnew(3),Vout(n),dt);
    end
    Vout(n+1)=(Vout(n+1)>0)*Vout(n+1);  %No negative flow

    %WEBSTER'S EQUATION

    %Load vector for the FEM-solver
    bhat(1)=dt*.25*Ao*(Vout(n)+Vout(n+1));

    %Crank-nicholson time discretization. The matrices in the update
    %equations are precomputed in VTdata.m for steps with time step "step".
    if dt<step
        xi_old=xi;
        xi=(dt/2*K+2/dt*rho*M+rho*R)\((-dt/2*K+2/dt*rho*M+rho*R)*xi_old+2*M*eta+bhat);
        eta=2/dt*rho*(xi-xi_old)-eta;
    else
        xi_old=xi;
        xi=xici*(xic*xi_old+2*M*eta+bhat);
        eta=2/step*rho*(xi-xi_old)-eta;
    end

    Pm(n+1)=eta(N);                   %Pressure at mouth
    Vm(n+1)=-(xi(N)-xi(N-1))/ds;      %Flow velocity at mouth
    Pc(n+1)=eta(1);                   %Counter pressure
 end

[OQ,flux]=suhde(op,Vout,x(9,:));     %Calculating the open quotient and
OQ                                    %glottal net flux and printing them
flux
```

A.4 File ff.m

```
function f=ff(x,Pc,Vf,open,M1_inv,K1,B1,M2_inv,K2,B2,Y_110,Y_120,
              Y_210,Y_220,width,kerr,L,Kh,d)

%This is the function f of the equation x'(t)=f(x(t)), which are the
%equations of motion of the glottis

gap1=x(5)-x(1);
gap2=x(7)-x(3);
```

```
if open > .5

    if gap1<0
        error('neglog')
    end

    %Load force for open glottis
    Fsum=-kerr*Vf^2/2/gap1/gap2;
    F1=kerr*Vf^2/2*(-1/gap1/(gap2-gap1)+1/(gap1-gap2)^2*log(gap2/gap1));
    F2=Fsum-F1;
    F=[F1-Y_110^2/2/L*width*Pc;F2+Y_110^2/2/L*width*Pc];
else
    %Load force for closed glottis
    F=[(gap1<0)*Kh*(-gap1)^1.5-Y_110^2/2/L*width*Pc;Y_110^2/2/L*width*Pc];
end

dW1=M1_inv*(-B1*[x(2);x(4)]-K1*[x(1)-Y_110;x(3)-Y_120]-F);
dW2=M2_inv*(-B2*[x(6);x(8)]-K2*[x(5)-Y_210;x(7)-Y_220]+F);

f=[x(2);dW1(1);x(4);dW1(2);x(6);dW2(1);x(8);dW2(2)];
```

A.5 File NewV.m

```
function new_v=NewV(gap1,gap2,v,dt)

%This function calculates the glottal flow by using a semi-implicit
%Euler-method

global vakio1 vakio2 rho Ao Am width L mu

%Driving pressure (p_{sub} in the flow-ODE)
Plung=800;

%No flow if the glottis is closed
if gap1<=0
    new_v=0;
    return
end

CC=12*mu*Ao*0.8e-3/width/gap1^3+8*mu*Ao/pi*vakio2;

%Semi-implicit Euler method
new_v=1/(1+CC*dt/vakio1)*(v+dt/vakio1*(Plung-.5*rho*(Ao/Am)^2*v^2));
```

A.6 File interpol.m

```
function [new,dt]=interpol(Xold,X,Xnew)

global d step;

gap0=Xold(5)-Xold(1);
gap1=X(5)-X(1);
gap2=Xnew(5)-Xnew(1);

%Interpolate the point of closure (stored in "root")
p=polyfit([Xold(9),X(9),Xnew(9)],[gap0-d,gap1-d,gap2-d],2);
r=roots(p);
root=r(1);
if and(r(2)>X(9),r(2)<Xnew(9))
    root=r(2);
end

%"new" is the interpolated solution
new=zeros(8,1);

p=polyfit([Xold(9),X(9),Xnew(9)],[Xold(1),X(1),Xnew(1)],2);
new(1)=polyval(p,root);

p=polyfit([Xold(9),X(9),Xnew(9)],[Xold(2),X(2),Xnew(2)],2);
new(2)=polyval(p,root);

p=polyfit([Xold(9),X(9),Xnew(9)],[Xold(3),X(3),Xnew(3)],2);
new(3)=polyval(p,root);

p=polyfit([Xold(9),X(9),Xnew(9)],[Xold(4),X(4),Xnew(4)],2);
new(4)=polyval(p,root);

p=polyfit([Xold(9),X(9),Xnew(9)],[Xold(5),X(5),Xnew(5)],2);
new(5)=polyval(p,root);

p=polyfit([Xold(9),X(9),Xnew(9)],[Xold(6),X(6),Xnew(6)],2);
new(6)=polyval(p,root);

p=polyfit([Xold(9),X(9),Xnew(9)],[Xold(7),X(7),Xnew(7)],2);
new(7)=polyval(p,root);

p=polyfit([Xold(9),X(9),Xnew(9)],[Xold(8),X(8),Xnew(8)],2);
new(8)=polyval(p,root);

dt=root-X(9);
```