

Depth-kymography of vocal fold vibrations: part II. Simulations and direct comparisons with 3D profile measurements

Frits F M de Mul, Nibu A George, Qingjun Qiu, Gerhard Rakhorst and Harm K Schutte

Department of Biomedical Engineering BMSA, Faculty of Medicine, University Medical Center Groningen UMCG, University of Groningen, PO Box 196, 9700 AD Groningen, The Netherlands

E-mail: ffm@demul.net

Received 29 July 2008, in final form 14 March 2009

Published 3 June 2009

Online at stacks.iop.org/PMB/54/3955

Abstract

We report novel direct quantitative comparisons between 3D profiling measurements and simulations of human vocal fold vibrations. Until now, in human vocal folds research, only imaging in a horizontal plane was possible. However, for the investigation of several diseases, depth information is needed, especially when the two folds act differently, e.g. in the case of tumour growth. Recently, with our novel depth-kymographic laryngoscope, we obtained calibrated data about the horizontal and vertical positions of the visible surface of the vibrating vocal folds. In order to find relations with physical parameters such as elasticity and damping constants, we numerically simulated the horizontal and vertical positions and movements of the human vocal folds while vibrating and investigated the effect of varying several parameters on the characteristics of the phonation: the masses and their dimensions, the respective forces and pressures, and the details of the vocal tract compartments. Direct one-to-one comparison with measured 3D positions presents—for the first time—a direct means of validation of these calculations. This may start a new field in vocal folds research.

 This article features online multimedia enhancements

(Some figures in this article are in colour only in the electronic version)

1. Introduction

We are interested in the horizontal and vertical positions and movements of the human vocal folds. Recently, we developed for this purpose a novel fast 3D profiling laryngoscope ('depth-kymograph') (George *et al* 2008). Since the introduction of videokymography (VKG) (Svec and Schutte 1996, Schutte *et al* 1998, Svec *et al* 2007, 2009), in which one single line crossing

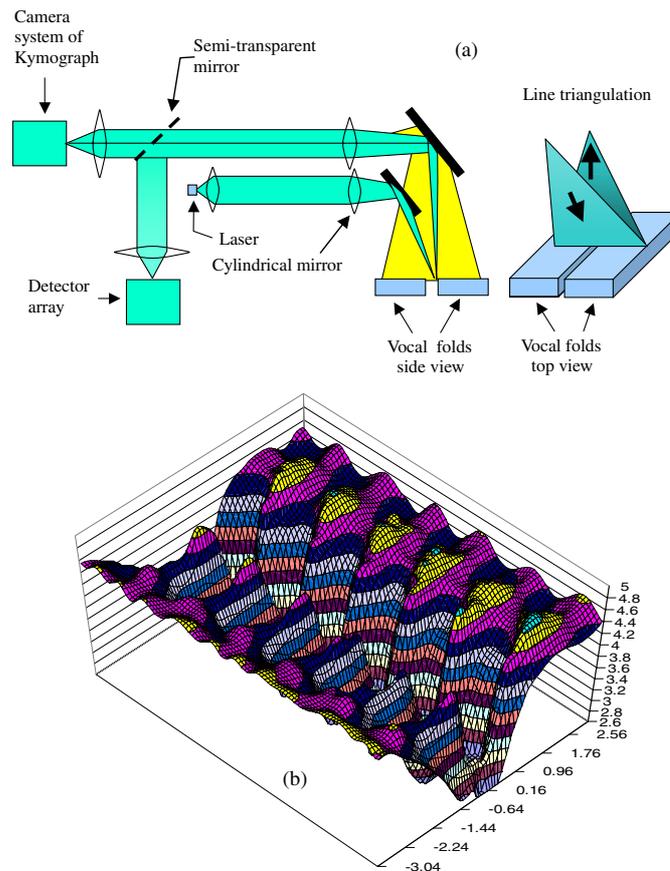


Figure 1. (a). Depth-kymography. A laser light sheet is projected onto the vocal folds, and the light reflected at a small angle (triangulation angle) is analysed using a separate detector array (in general). For clarity, in this sketch the optics from the mirrors on, and the vocal folds, have been rotated 90° in the horizontal plane, with respect to the real situation in the mouth. In the present experimental situation, for the detector array, a high-speed Kymo camera is used, and a semi-transparent mirror is absent. The three spatial dimensions are (1) along the laser line, (2) the depth and (3) by shifting the laser line perpendicular to the glottal midline. (b) Depth-kymography. Time recording of a registration. Horizontal axis: distance to the glottal midline (mm); vertical axis: profile height (mm). Third axis: time, interval = 0.25 ms (4 kHz sampling rate).

the two vocal folds is imaged over time using a fast camera system, enabling to study the horizontal movements of the vocal folds as a function of time, the third dimension (vertical) appears to be physiologically interesting as well. The novel depth-kymograph makes use of the same single line imaging as normal (2D) VKG. See figure 1. In section 2.1, we will give a more detailed description.

With three spatial dimensions we mean (1) the imaging line across the two vocal folds, (2) the depth profile of the surface along this line and (3) shifting (presently manually) this line over the vocal folds, perpendicular to the glottal midline, in the anterior–posterior direction.

A distinctable vertical displacement of the vocal fold, mainly of its mucosa, is clinically related to loosely connected mucosa to the underlying vocal fold body. This contributes typically to a better and more defined closed quotient of the glottal cycle. Establishing the extent (height) and pattern of the profile incorporates a clue of significance about the health

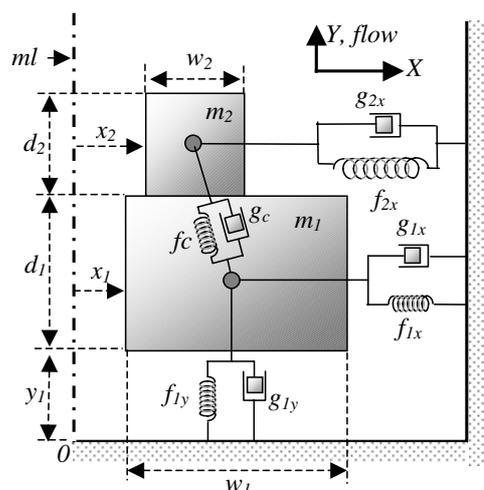


Figure 2. Outline of the vocal fold two-mass model. Only one vocal fold is shown. ml is the midline of the glottis and provides the 0-point for the x -coordinates. Mass m_1 is connected to the wall (at the side and below) with the spring forces f_{1x} and f_{1y} , and the damping forces g_{1x} and g_{1y} . Mass m_2 is connected to m_1 by coupling the forces f_c and g_c which both have x - and y -components. Mass m_2 is also connected to the sidewall by the forces f_{2x} and g_{2x} . d_1 and d_2 are the thicknesses of the masses, and w_1 and w_2 are their widths. The third dimension (length of the glottis) is l_g (not drawn).

and possibilities of the tissue. Polyps, abscesses, carcinoma, asymmetries and other voice disorders, encountered in singing and speech, will be assessed much more clearly. It becomes gradually better known how important the glottis closure by the mucosal wave is. The present tool of depth-videokymography gives us an excellent tool for examining this clinical detail of voice production.

The first results of this 3D method make it worthwhile to find a method that would enable us to relate the 3D findings to physiologically interesting data, like material constants, such as stiffness and viscosity data of the vocal fold masses. For this purpose, we investigated the potential of numerical 3D simulations of the dynamical positions and movements of the vocal folds during self-sustained phonation. Since these simulations are based on physical reasoning, constants and expressions, in principle, relations of the measurement details with physical constants can be made.

Simulations of the dynamics of the human vocal folds have been the subject of a series of papers, starting from Flanagan (1968), Flanagan and Cherry (1969) and Dudgeon (1970). This model makes use of a single vibrating mass for each vocal fold, subject to spring forces connecting it to the wall and pressure forces due to air pressures in the vocal tract. This mass was allowed to move in horizontal directions only. The shape of the vocal tract (lengths and cross sections of compartments) could be altered, and in doing so, sound could be obtained.

Since the one-mass model turned out to be not sufficient to produce phonation with adequate physiological accuracy, multi-mass models were introduced (Ishizaka 1966, Ishizaka and Flanagan 1972). Here the masses are connected to the surrounding walls and to each other by spring-and-damper force combinations. This model is schematically shown in figure 2. A thorough account of this approach was given by Ishizaka and Flanagan (1972), albeit for horizontal freedom only (no vertical forces or movements). These authors provided the framework for all the following modelling: they gave expressions for the pressures in the various parts of the vocal duct and for the spring-and-damper forces. For the vocal tract they

introduced a simple electronic analogue circuit model, consisting of a transmission line of T-circuits representing cylindrical containers with rigid walls, each containing air resistance and inertia components in both 'horizontal' branches of the T and an air compliance in the 'vertical' branch. The mouth was modelled according to the radiation load equal to that of a circular piston in an infinite baffle, as an air resistance and inertia in parallel. Two equations of motion were used for the horizontal (X)-coordinate of the two masses, respectively. The forces were exerted by the connecting springs and dampers, and by the pressure force exerted on the surface areas of the masses, exposed to the air flow. With this model, they studied self-sustained oscillation and phonation for various vowels and they were able to extract characteristic values for the parameters used in the model. A multiple-mass model was introduced by Titze, in which a set of up to 16 masses were handled, followed by a continuum-mass model (Titze 1973, 1974, 1984, Titze and Strong 1975).

These simulations revealed that the two-mass model might be too simple to be capable of simulating the laryngeal movements to fine physiological details, but that it can be used in a satisfactory way as far as the production of phonation by speech synthesis is concerned. This was the underlying justification of the work of Koizumi and co-workers (Koizumi *et al* 1987, 1989, Nudelman and Hoyt 1989). They added vertical freedom to the Ishizaka–Flanagan model, in the form of a varying vertical (Y)-coordinate and of vertical components to the spring-and-damper force combinations, but they discarded the direct connection of the upper mass to the sidewall. Further, they allowed the masses to vary in depth (and in mass, but not in width), under the constraint that the total mass should remain constant. For the expressions and numerical values of the spring-and-damper forces, the pressures and pressure forces, they relied on the work of Ishizaka and Flanagan (1972). Numerical details for the vocal tract components (lengths and areas of components) were obtained from work published by Titze and co-workers (Story *et al* 1996, 1998, Story and Titze 1998). The effect of yielding vocal tract walls could be included as well, by using an adapted circuitry for the electrical analogue of the vocal tract T-circuits (Ishizaka *et al* 1975).

Based on these previous models, we constructed a two-mass model, allowing vertical freedom, with the possibility of mass transfer between the two masses during the vibrations. In doing this, a description of a moving mucosal wave on top of the vocal folds could be included.

2. Method

2.1. Experiments

Using the depth-kymograph (George *et al* 2008), we recorded depth-kymographic data of various persons. The laryngoscope was inserted in the mouth, and the laser light was projected as a line crossing both vocal folds, perpendicular to the glottal midline. The line was positioned using the illumination of the conventional 'horizontal' kymograph system. All subjects gave informed consent for these measurements.

In figure 1(a), an overview of the instrument is sketched. The laser projection optics consists of a semiconductor laser, emitting at 635 nm, with 1.8 mW mm^{-2} power density, together with an aspheric lens system to focus the beam onto the folds and a cylindrical lens system for stretching the beam to a line across the vocal folds. These optics are mounted in an extra optical tube, along the normal endoscope tube. The triangulation angle is about 8° , which appears to be the experimental maximum. The detection distance is 60–70 mm from the endoscope tip, and the laser line is 18–20 mm long and about 0.4 mm wide. Detection takes place using a high-speed colour camera (Wolf, HResEndocam 5562; 256×256 pixels), at a

sampling rate of 4 kHz, via the normal endoscope channel. For simultaneous 2D kymographic and stroboscopic imaging, we used a 300 W white light source, also coupled to the endoscope. However, the 2D image can also be obtained from the depth-kymographic image, by adding all intensity in the ‘column’ of pixels perpendicular to the ‘row’ of pixels with the laser beam image. A typical width of the envelope of the image of one single point of the laser beam line is 20 pixels.

Typically several seconds of phonation were recorded, and the resulting data were processed using a dedicated algorithm and plotted as cross sections in a plane perpendicular to the glottal midline, as a function of time (figure 1(b)). See George *et al* (2008) for more details.

We have recorded the vocal fold vibrations of a number of subjects, all with normal vocal folds. For now, for the comparison with simulations, we use only the measured results obtained with one single male person, about 65 years of age, and with no laryngeal pathology leading to vocal folds disorders (after usual laryngoscopic examination by an experienced laryngologist). These recordings were chosen, since this person showed a left–right asymmetry in vertical motions and profile, which was not seen with stroboscopy or ‘normal’ 2D kymography, looking ‘from above’ only.

2.2. Simulations

We investigated an extended model including the vertical dimension. We also included the option to use both the widths and the depths of both masses as dynamic variables, so as to allow mass transfer between the two masses. In addition, following Ishizaka and Flanagan (1972), we added a spring-and-damper connection of the upper mass to the respective sidewall, since we observed from our 3D measurements that this seems to be the case in reality as well. We named this extended model the ‘combined model’. We designed the programming in such a way that it would be easy to return to the two underlying more limited models. In figure 2, a sketch of the model is shown. A full account of the model is given in the appendix.

Basically, for the iterations of the positions, eight variables are relevant: the horizontal (x) and vertical (y) positions of the centre of mass of both masses and their thicknesses d and widths w . With these eight variables, it is hard to obtain stable iteration solutions. However, since the total mass has to be conserved and the upper mass always rests on top of the lower mass, two of the variables can be skipped, leading to a slight increase in the stability only. Therefore, we decided to reduce the number of variables to four, by investigating three situations (for each mass):

- S.1. variables x , y , w and d , with the constraint that at the wall side the mass is fixed;
- S.2. variables x , y and d , but constant w and
- S.3. constants d and w , thus implying a constant mass.

Apart from the extra freedom in the y -direction, the third situation corresponds to that of Ishizaka and Flanagan (1972).

For typical input values of the parameters, we generally relied upon Ishizaka and Flanagan (1972) and Koizumi *et al* (1987). Table 1 lists parameters used as inputs in the simulations, and table 2 lists the variables resulting from the simulations. In addition, during the simulations, all actual values for all coordinates, masses, pressures and flows, as functions of time, are recorded in the form of plots or tables.

In the simulations, the time step for the calculations is chosen to be 4 μ s or less. Output time intervals (e.g. for plotting and frequency analysis) were of 50 μ s.

The frequency analysis is done by fast Fourier transformations of the time signals of all coordinates, pressures and flows, as the real Fourier input function, with zeros for the imaginary

Table 1. Typical input parameters of the simulations (see also Ishizaka and Flanagan (1972) and Koizumi *et al* (1987)).

Parameter	Value	Parameter	Value
rhoA: air density (g cm ⁻³)	0.001 02	mass 2: m20: mass in rest (g)	0.04
lambda: heat conduction coefficient (W (mK) ⁻¹)	0.026	-----: d20: depth in rest (cm)	0.09
Cp: specific heat constant pressure (J kg K ⁻¹)	1005	-----: x20: x-position in rest (cm)	0.02
eta: adiabatic constant (-)	1.4	-----: k2x: spring constant x (N m ⁻¹)	15
c: sound velocity (m s ⁻¹)	333	-----: eta2x: nonlinear spring constant x (cm ⁻²)	100
mu: air viscosity (Pa s) or (N s m ⁻²)	1.8 × 10 ⁻⁵	-----: h2x: collision spring	180
rhoT: tissue density (g cm ⁻³)	1.04	-----: etah2x: nonlinear collision constant x (cm ⁻²)	500
lg: glottis length (cm)	1.4	-----: zeta2x: damping ratio x (-)	0.1
mass 1: m10: mass in rest (g)	0.09	coupling: kcx: spring constant x (N m ⁻¹)	15
-----: d10: depth in rest (cm)	0.2	-----: kcy: spring constant y (N m ⁻¹)	15
-----: x10: x-position in rest (cm)	0.02	-----: etacx: nonlinear spring constant x (cm ⁻²)	100
-----: y10: y-position in rest (cm)	0.15	-----: etacy: nonlinear spring constant y (cm ⁻²)	100
-----: k1x: spring constant x (N m ⁻¹)	60	-----: zetacx: damping ratio x (-)	0.7
-----: k1y: spring constant y (N m ⁻¹)	100	-----: zetacy: damping ratio y (-)	0.7
-----: eta1x: nonlinear spring constant x (cm ⁻²)	500	lungs: Plungs: pressure (Pa)	1000
-----: eta1y: nonlinear spring constant y (cm ⁻²)	100	-----: Rlungs: resistance (kg (m ⁴ s) ⁻¹)	8 × 10 ⁵
-----: h1x: collision spring constant x (N m ⁻¹)	180	trachea: length (cm)	5.0
-----: etah1x: nonlinear collision constant x (cm ⁻²)	500	-----: area (cm ²)	3.0
-----: zeta1x: damping ratio x (-)	0.1	contraction: length (cm)	1.0
-----: zeta1y: damping ratio y (-)	0.2	-----: area at entrance (cm ²)	0.5
		-----: area at mid point (cm ²)	0.3
		-----: area at exit (cm ²)	0.1
		mouth: area (cm ²)	2.0

Table 2. Output variables of the simulations.

Variable (o)	Symbol	Variable	Symbol
Model code ^a		DC flow (l s ⁻¹)	DCF
Lower mass: (g)	⟨m1⟩ ± dm1	Mouth pressure: PresM (Pa)	⟨MP⟩ ± dMP
Horizontal position (mm)	⟨x1⟩ ± dx1	Mouth flow: FlowM (l s ⁻¹)	⟨MF⟩ ± dMF
Vertical position	⟨y1⟩ ± dy1	Frequencies: first peak (Hz) (FM1)	1st Freq
Slope	dy1/dx1	Ratio Second/first, . . . , fifth/first peak intensity	Rat.2/1 . . .
Thickness	⟨d1⟩ ± dd1	Glottis: closed quotient (%)	Clos%
Width	⟨w1⟩ ± dw1	Phase delay upper vs. lower mass (degree)	Ph.2/1
Upper mass (g)	⟨m2⟩ ± dm2	Averaged duration of open period (ms)	AD; time open
Horizontal position (mm)	⟨x2⟩ ± dx2	Max. open width (mm)	MW
Vertical position	⟨y2⟩ ± dy2	Max. open area (mm ²)	MA
Slope	dy2/dx2		
Thickness	⟨d2⟩ ± dd2		
Width	⟨w2⟩ ± dw2		

Notation: for variable x: ⟨x⟩ = time average; dx = amplitude of vibration during one cycle.

^a Model code: 'ABC':

A = 1, 2, 3: according to Ishizaka and Flanagan, Koizumi and (present) combined model, respectively.

B = 1: lower mass: both width *w* and thickness *d* variable; vocal folds fixed at wall side (situation S1); 2: id.: width constant, thickness variable (S2); 3: id.: width and thickness constant (S3).

C = 1, 2, 3: for upper mass, as with B (for lower mass).

part. The frequency spectra are then obtained as the modulus of the real and imaginary parts of the Fourier output function. The maximum frequency f_{\max} is normally chosen at 1 or 5 kHz, implying time steps of $1/(2f_{\max}) = 500$ and $200 \mu\text{s}$. The frequency step is f_{\max}/N_{freq} , with N_{freq} the number of Fourier points. From the time signals, sound can be produced using

the internal computer synthesizer, with a sampling rate of $11\,025\text{ s}^{-1}$. As long as the elapsed simulation time is shorter than the inserted sound period, the time signals are repeated.

Most of the simulations were performed without the option of yielding walls, since we found that mostly the effects of inclusion of yielding walls were small to negligible and since inclusion of yielding wall effects increased the calculation times considerably.

2.3. Comparison of simulations and measurements

The computer program produces images of the surface of the vocal folds along the laser line, as a function of time. Afterwards, these subsequent images were combined to a movie, and overlaid in the movie of the experimental data (available at stacks.iop.org/PMB/54/3955).

To obtain an optimum correspondence, first the frequency of the simulations was corrected so as to match that of the measurements. This was done by adapting the masses and thicknesses of the two masses. Among other options are the lengths of the vocal folds or the values of the springs and dampers. In this case, for this subject, the correction factor was 0.80. It can be seen in the movie that the frequencies of the measurements and the simulations still differ slightly (240.2 versus 239.6 Hz). We did this deliberately, so as to show a gradual growth of the respective phase difference.

Since in general the duration of the measured data will not correspond to an exact integer multiple of the cycle period, in the overlay comparison procedure we skipped the last incomplete cycle, in order to obtain a self-repeating in-phase sequence, needed for the comparison with the simulated data.

2.3.1. Correspondence criterion. After matching the frequencies, the phases were matched by shifting the phase of the measured data, so as to obtain a minimum value for the correspondence criterion CC, which we simply defined as

$$CC = \sqrt{\left[\frac{1}{N} \sum_{n=1}^N D(\text{simulated} - \text{measured})_n^2 \right]}, \quad (1)$$

where $D(\cdot\cdot)_n$ stands for the actual smallest distance between the upper edge point of the upper mass (modelled with rectangular cross section), nearest to the glottal midline, and the measured depth-kymographic curve. The averaging summation is performed over all available time points ($n = 1, \dots, N$) in all completed cycles of the measured data.

3. Results of simulations

As an example, plots of the simulations obtained with the parameters listed in table 1 are shown in table 3 and figure 3. In this simulation, use has been made of the vocal tract area function, listed by Story and Titze (Story and Titze 1998, Story *et al* 1996, 1998) as corresponding to the vowel /o/. This function consists of 44 compartments (T-circuits), which makes the pressures-and-flows array (equations (A.15) and (A.19)) to have dimensions 91×91 , or 46×46 , when yielding wall effects (Ishizaka *et al* 1975) are taken into account or not, respectively. In the present work, we did not include the yielding wall option, since we found that this option has a relatively small effect only at the expense of considerably increased calculation time.

We started the simulation from zero pressures and flows by linearly increasing the lung pressure from zero to the parameter value as shown in table 1, within 5 ms.

Figure 3 contains time and frequency plots, for a time window between 380 and 400 ms after the start. The available plots are plots of the six coordinate variables, the pressures at

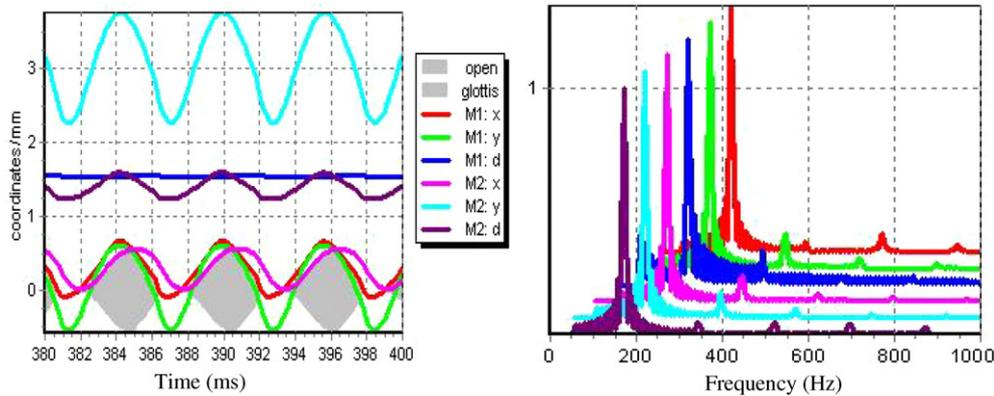


Figure 3. Coordinates as time and frequency functions. Model code: 312 (combined model; situations: S2 and S1 for upper and lower mass (M2 and M1), respectively). Calculation time step = 0.004 ms. Output time step (plotted points) = 0.05 ms (also used in Fourier transforms). Fourier spectra (maximum frequency = 5000 Hz with 1024 points; time step = 0.100 ms). The open glottis is indicated with shading. For clarity, the frequency spectra are peak-normalized and shifted over multiples of 1/20 of the full horizontal and vertical scales.

Table 3. Output data of the simulation, shown in figure 3.

Variable	Lower mass		Upper mass	
	Initial	Final	Initial	Final
Mass (g)	0.09	0.067 ± 0.008	0.04	0.063 ± 0.008
Horizontal position (mm) ^a	0.2	0.29 ± 0.38	0.2	0.32 ± 0.28
Vertical position	0	0.12 ± 0.58	2.9	3.08 ± 0.74
Thickness	2.0	1.55 ± 0.01	0.9	1.41 ± 0.18
Width	3.09	2.99 ± 0.33	3.05	3.05 ± 0
Slope dy/dx ^b		1.53		2.68
DC flow (l s ⁻¹)				0.20784
Mouth pressure (MP) (Pa)				-0.040 ± 61.7
Mouth flow (MF) (l s ⁻¹)				0.194 ± 0.565
First frequency peak FM1 (Hz)				171.0
Peak ratios 2/1; 3/1; 4/1; 5/1				1.39; 1.13; 0.55; 1.03
Closure % (closed quotient)				17.3
Phase delay upper versus lower mass (degree)				45.1
Open glottis: (duration) (ms)				4.84
Id. max. area (mm ²)				16.1

^a From the edge of the glottis to the glottal midline.

^b Vertical versus horizontal movement.

different positions along the air tract and the flows calculated for the lungs, the glottis and the mouth, but we show here the time and frequency plots of the coordinates only.

From the panels it is seen that a vibration of the glottis has been developed, with a phase delay between the two masses.

From the time development of the pressure or flow through the mouth, a sound signal can be derived.

For the simulations, a set of all nine combinations were made out of the three situations (S1, S2, S3) as indicated above, for the two masses, in which the parameter values shown in table 1 were taken as the ‘original’ situation. The ‘combined model’, as described in the appendix, was compared with the Ishizaka–Flanagan and the Koizumi model in several relevant situations. The following simulations were performed.

- (1) Using the parameter values shown in table 1, several vocal tract area functions were inserted. See table 4 for an overview. The area functions obtained by Story and Titze (Story and Titze 1998, Story *et al* 1996, 1998) are of special interest, since the shape of the vocal tract compartments differs considerably. It is expected that the intensity ratios of the higher harmonics will differ accordingly.
- (2) For all combinations, all parameters from the list shown in table 1 (from the glottal length on) were varied around their original values by multiplying with ten different factors: 1.01, 0.1, 0.2, 0.5, 0.8, 1.2, 1.5, 2.0, 5.0 and 10.0, and compared with the ‘original’ combination (factor 1.0). The factor 1.01 was used especially to calculate relative slopes of the changes in the variables upon variation of the parameters. Not all simulations provided sustained vibrations, especially those with the factors 0.1 and 10. These can be excluded afterwards by allowing only simulations with horizontal vibration amplitudes for both masses larger than a specified value (e.g. 0.01 mm). This leads to $9 \times 34 \times 9$ simulations. In a 1.2 GHz computer, each simulation of 400 ms lasted about 2 min, in total 230 h. Therefore, we included an option in our program to perform series of simulations automatically.
- (3) Since the parameter y_0 of lower mass 1 has the ‘original’ value 0, it was treated separately using an initial value of 0.05, which then was handled as in (1).
- (4) Because with the procedure in (1), changing the mass or thickness implies a change in the width (to keep the thickness or the mass constant respectively), we also performed simulations in which the input values of both mass and thickness of the two masses were altered together, both with factors 1.1 and 0.9. In these simulations, we only made changes for the parameters mass, thickness and x -position at rest of both masses.
- (5) The mouth is modelled using a flow resistance and inertia in parallel, see the appendix, equation (A.12). This implies that for dc flows, the mouth would offer zero impedance to the flow. This effect was investigated by adding a small resistance in series with the inertia in the mouth.

We record the time development of all coordinates and masses, all pressures and flows and the variables as mentioned in table 2, as functions of time and frequency.

The averaged values of the output variables and their amplitudes are written in an ASCII file, and can be plotted in the form of ‘pivot plots’, containing the output variables V as functions of a chosen input parameter P , in a dimensionless, normalized way (i.e. divided by the values obtained with the ‘original’ situation), thus around the pivot value (1,1). There is also the option of producing ‘slope plots’, which contains the normalized slopes of the variables around the parameter values of the ‘original’ situation, in the form $(dV/V)/(dP/P)$.

4. Discussion

Due to the large amount of available calculated data, we limit ourselves to a selection.

The model code is defined as ‘ABC’, with ‘A’ as the model; ‘A’ = 1, 2, 3: Ishizaka–Flanagan, Koizumi and combined, respectively; ‘B’ and ‘C’ = 1, 2, 3 according to the three situations S1, S2 and S3 mentioned above, for the lower and the upper mass (masses 1 and 2), respectively.

The calculation time step, the frequency time step and the total simulation time per simulation were $4 \mu\text{s}$, $200 \mu\text{s}$ and 400 ms , respectively. The frequency spectra are refreshed each 20 ms .

Table 3 contains data of the output variables corresponding to the present simulations, for model 312. In figure 3, the coordinates are plotted as a function of time (left panel) and frequency (right panel). The grey shaded area in the time panel indicates the opening periods of the glottis. The shape of the glottis edge function is not harmonic, since it is composed of the vibrations of both masses. It is seen from the table that closure is obtained for both masses, but shifted in time. Here, the total closure percentage ('closed quotient') is 17%. There is a phase difference of about 45° between the two masses, and the closure of the glottis is obtained by the successive closure of the two constituting masses (which can be seen from the x -coordinates becoming negative). Although the coordinates have only one relevant frequency peak, the pressures and flows have various higher harmonics, in general up into the kHz region. This is probably related to the nonlinear behaviour of the vocal tract compartments. In figure 8, we will address this point again.

Also, the vibration frequency spectra of the pressures and characteristic flows are available. These spectra show a series of peaks, around 1 kHz and above, which can be assumed to correspond to the formants of the sound. It is remarkable that the coordinates show a single frequency peak only, while the flows and pressures have much more significant higher harmonics. For the mouth pressure, the higher harmonics frequently have higher intensities than the fundamental frequency (we will give results in figure 8).

In figure 4, a comparison of the various models is given. From top to bottom the panels show (a) the masses, (b) the coordinates, (c) other variables such as the mouth pressure and flow, and glottis variables, and (d) the ratios of the second, third, fourth and fifth frequency peaks with that of the first peak. This first peak corresponds to the (nearly) sole peak in the frequency spectra of the coordinates. The values shown in figure (b) are relative to those of model 333 (reference values shown in the legend). Models 323 and 332 are equivalent to model 333 and are therefore not listed.

In the plots, the average values—where relevant—are shown as bars, with error bars denoting the amplitudes during the vibration cycles. The 'original' parameter situation (table 1) is listed as 'model' 0. In several models, a dynamic redistribution of the total mass takes place as compared with the original situation (as in model 111; see also table 3), up to almost a reverse in masses in models 311 and 321. This effect was observed in almost all simulations we performed. After the initial period of re-division, the masses become practically constant, with small amplitudes, but the upper mass shows a considerable exchange in thickness versus width.

In all models, the average mouth pressure is relatively small compared with the amplitude. Sometimes it becomes even slightly negative, especially in the models with situation S1 for the lower mass (in this submodel, mass 1 is fixed to the sidewall). This might be caused by round-off errors in the simulations, due to the relatively large difference between the average value and the amplitude.

The fact that the average value is so much smaller than the amplitude can be related to the findings that professional singers might produce phonation with a large amplitude, together with using a small quantity of air, thus being capable of sustaining the tone for a relatively long time (Schutte 1986, Schutte and Miller 1986a, 1986b).

All simulations with the combined model produce similar average and amplitude values for the different coordinates. The amplitudes are about 25% of the average values, indicating a considerable vibration. This is also reflected in the values for the ratios of the vertical and

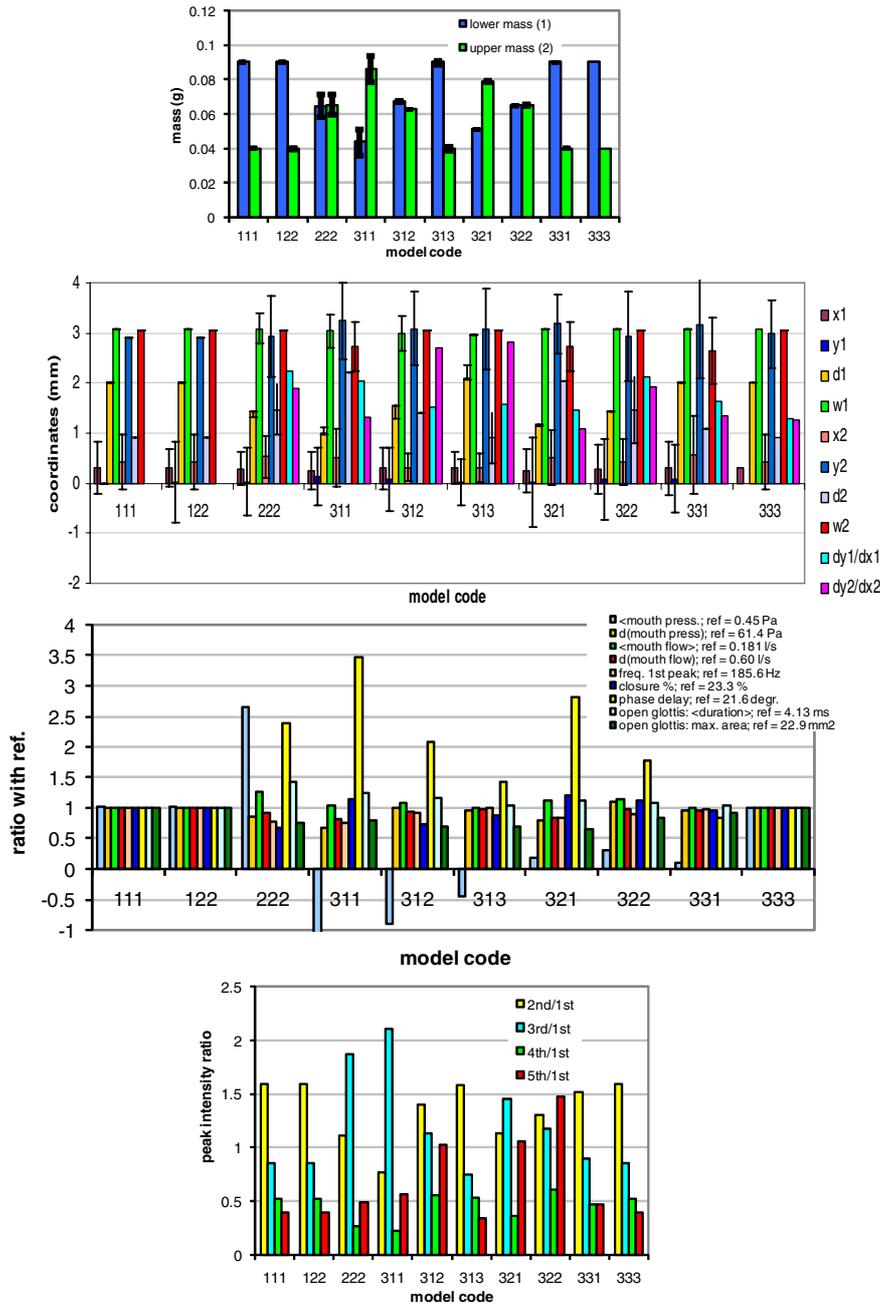


Figure 4. The effects of different models. Top to bottom panels: masses, coordinates, other variables, frequency peak ratios. Model code: ‘ABC’, A = model type (here: 1 = ‘Ishizaka–Flanagan’, 2 = ‘Koizumi’, 3 = ‘Combined Model’); B = 1, 2, 3 refers to situations S1 (variable width), S2 (fixed width) and S3 (fixed width and thickness), respectively, for the lower mass; C: as B, for the upper mass. See table 3. For models A and B, only the relevant models are shown. As can be expected, models 323, 332 and 333 gave similar results. Bar values are averaged values during vibrations; the error bars indicate the amplitudes during vibration. The ratios of the frequency peaks are calculated for PM (the pressure in the mouth). The values in the lower panel are relative to the reference value (model 333).

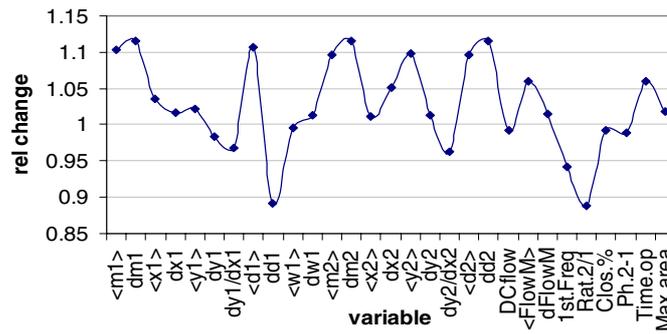


Figure 5. Relative change of variables with respect to the values listed in table 3, upon changing the rest masses and the rest thicknesses of both masses by 10%, relative to the values listed in table 1.

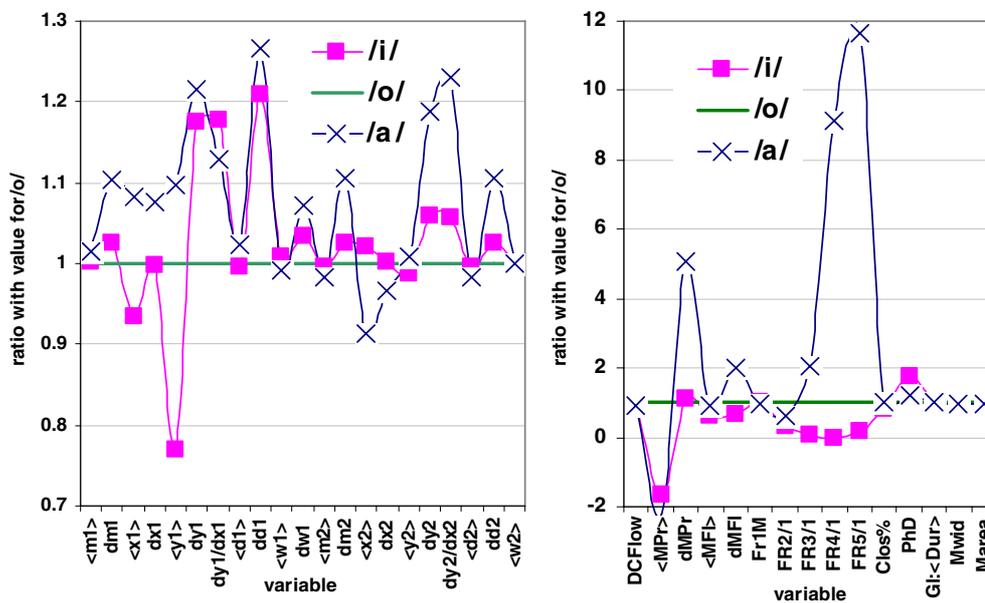


Figure 6. Effects of different vocal tract area functions (see Titze and Strong 1975 and Titze 1984). Model 312. The variable values are relative to the values of /o/. Note the differences in the mouth pressure and the frequency peak intensity ratios.

the horizontal amplitudes, which mostly are larger than 1. It implies that for certain phonation types, the vertical vibration is larger than the horizontal vibration.

The ratios of the frequency peak intensities (here for vocal tract area function /o/) indicate that the higher harmonics have considerable significance, as could already be seen in figure 3. We will discuss this further with figures 5 and 8.

In the following figures, model 312 is used, if not otherwise mentioned.

Mass and thickness of both masses are input parameters in the models. This implies that when changing the rest mass of only one of the two masses, the width of that mass will change as well, since the density will remain constant. Figure 5 shows the effect of changing

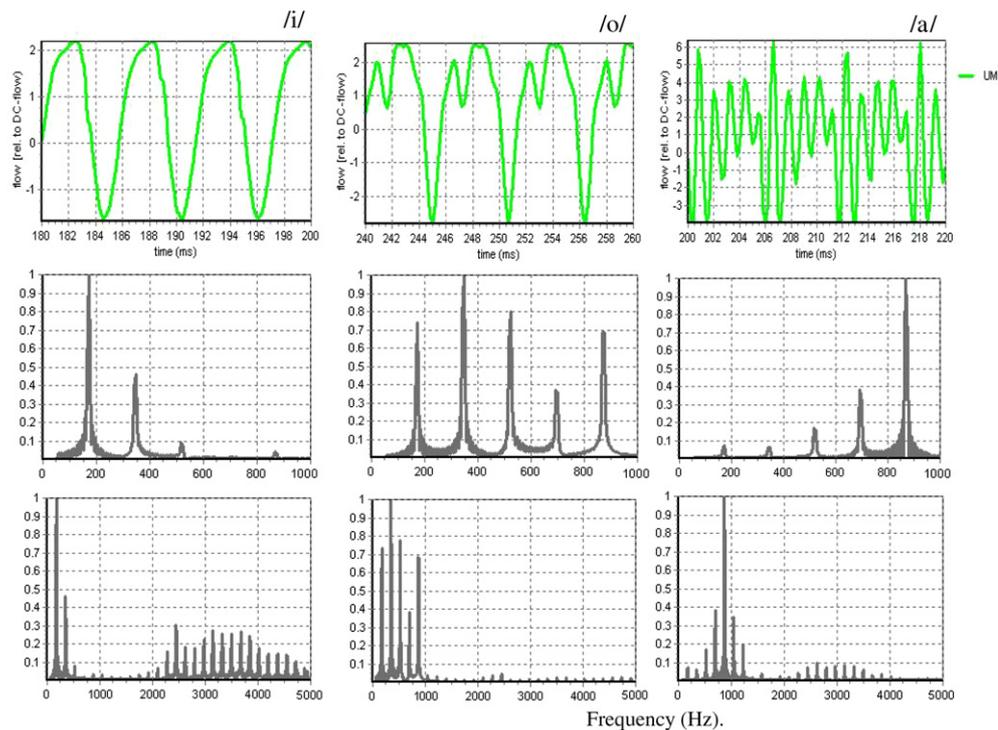


Figure 7. Time and frequency spectra of the pressure in the mouth of three vowels: /i/, /o/ and /a/, according to Story and Titze (Titze and Strong 1975, Titze 1984, Koizumi *et al* 1987). Model 312. Note the differences in the higher harmonics, at about 1 kHz and in the region of 3–5 kHz.

both the masses and thicknesses together (with +10%), keeping the width constant. The relative changes can roughly be distinguished in two categories: about 0% change and about 10% change. The latter correspond to changes in the averaged masses and thicknesses and their amplitudes, as could be expected. Also the frequency ratios show larger changes. The changes are smaller for the x -, y - and w -coordinates of mass 1, and the x -coordinate of mass 2. Apparently, these variables are not very sensitive to this type of parameter changes. The changes are also substantial for the average flow in the mouth, the frequency of the first peak and the second to first peak ratio. That can be expected since the frequency is directly related to the square root of the masses.

Figure 6 plots the effects of changing the vocal tract area function for different vowels: /i/, /o/ and /a/, as defined by Story and Titze (1998), Story *et al* (1996, 1998), relative to the results of one of those functions (/o/). The /i/- and /a/- vowels have larger amplitudes in the thicknesses of both masses and larger slopes (ratios of vertical and horizontal amplitudes). Significant differences also occur for the average mouth pressure and the peak intensity ratios. It appears that functions /i/, /o/ and /a/ have increasing higher harmonics intensities in the 0–1 kHz region, especially the fifth harmonic of function /a/. Other variables are not particularly influenced by the choice of the vocal tract area function. This might indicate that in those situations, the dynamics of the glottis part and the reaction of the vocal tract are not very tightly coupled, meaning that the vibrations of the vocal folds can probably be considered as a relatively independent driving motor for the air vibrations in the vocal tract.

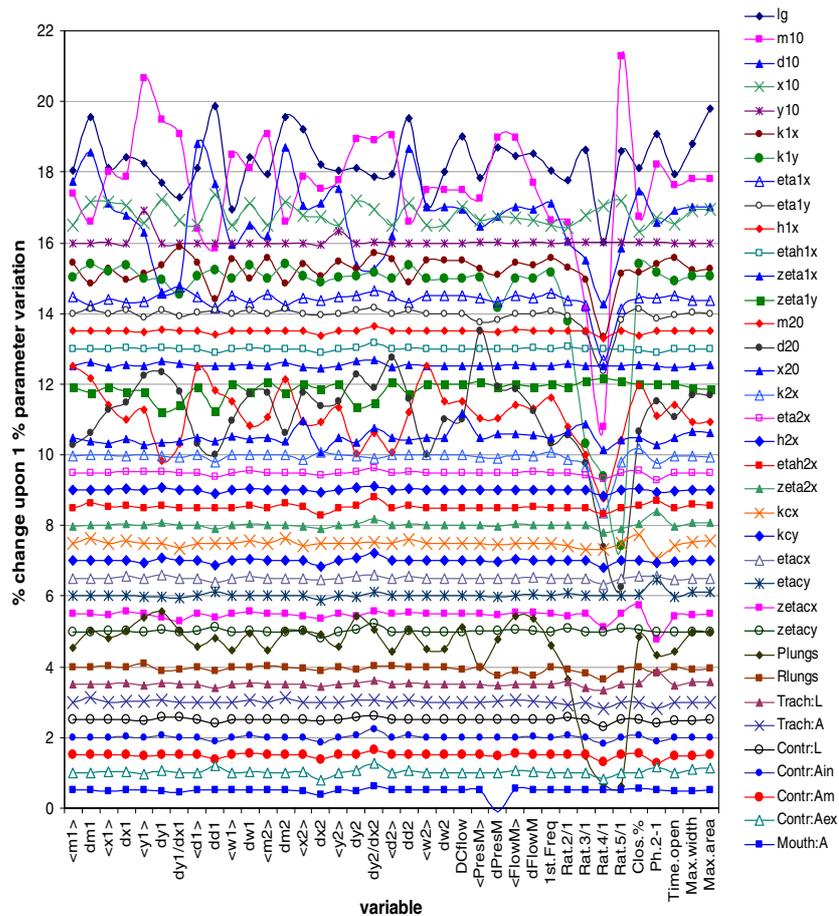


Figure 8. Influence of variations of parameters (see legend) on the variables (at the horizontal axis). Model 312. Cumulative offset: +0.5.

In figure 7, waveforms and frequency spectra for those three vowels are shown in more detail. In the low-frequency region, below 1 kHz, vowel /i/ closely follows the fundamental frequency of the glottis, while vowel /o/ has comparable intensities in the first few higher harmonics, and vowel /a/ has a high intensity in the fourth harmonics around 1 kHz, at the cost of the ground frequency intensity, probably corresponding to brilliance. The vowels /a/ and especially /i/ have higher harmonics in the intermediate region of 3–5 kHz.

In figure 8, the influence of variations of the parameters upon the variables is schematically plotted, as the percentage of change upon 1% change in the parameter value. In the plot, the parameters have cumulative offsets of +0.5. Most results are (much) smaller than unity, except for changes in the frequency peak intensities. The parameters which also have large significant effects are the length of the glottis (l_g), the rest values of the masses, thicknesses and x -coordinates of the two masses, and the driving lung pressure. The most sensitive variables are the x - and y -coordinates of the two masses and the peak intensity ratios. These effects need further study.

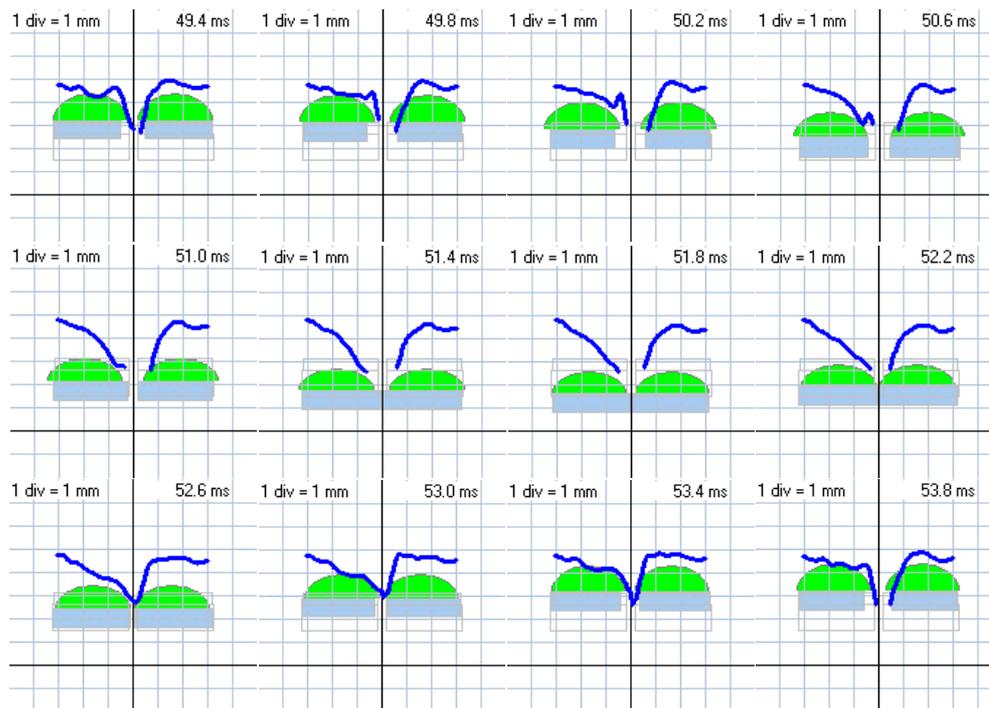


Figure 9. Snapshots from the movie (available from stacks.iop.org/PMB/54/3955), made of successive vertical cross sections of the simulations, during one vibration cycle. Curve: overlay of the corresponding depth-kymographic measurement. One division = 1 mm. Filled shapes: actual positions; open rectangles: positions at the start of simulations. Correspondence criterion value (see section 2.3): $CC = 0.45$ mm. Note the exchange of mass between the two masses during the cycle.

4.1. Comparison of simulations and measurements

Finally, in figure 9, we present overlays of depth measurements and simulations, in a plane perpendicular to the glottal midline. We used the measured results obtained with one single male person, about 65 years of age, and with no vocal folds disorders, for the comparison with simulations. These recordings were chosen, since this person showed a left–right asymmetry in vertical motions and profile, which could not be seen with stroboscopy or ‘normal’ 2D kymography, since these techniques produce top view images only.

Figure 9 contains 12 events out of one of the vibration cycles. The measurements are shown as a line, overlaying the two masses of the simulations. We have drawn the cross section of the upper mass in the form of a half-ellipse, so as to resemble the real shape of that upper mass (as assumed in the case of mucosal wave presence) more closely than the rectangle, used in the model. The half-ellipse has a long axis equal to the width of the rectangle and the short axis so as to keep the mass equal to that of the rectangle. In the accompanying movie (available at stacks.iop.org/PMB/54/3955), this comparison is illustrated. The exchange of mass between the two masses during the cycle is marked. The mass content of the upper mass is higher when its vertical position is higher up.

In section 2.3, the correspondence criterion for estimating the correspondence between the measured data and the simulated data was introduced. This criterion determines the closest distance between the upper surface of the vocal folds in measurements and simulations, averaged over all available completed vibration cycles. For the situation, corresponding to figure 9, the CC value was 0.45 mm. To obtain this value, first the parameters in the simulation were adjusted so as to produce results with the proper frequency, corresponding to that of the measurements (240.2 Hz), and secondly the phase was adjusted to that of the measurements so as to obtain the minimum criterion value.

From the overlays shown in figure 9 and from the movie, it is seen that—at least for the underlying measurements—the model needs some correction: the mass seems to be firmly attached to its sidewall. This may indicate that for the upper mass, model situation S1 is most appropriate. In addition, to incorporate mucosal wave presence as well, the shape of the upper mass (now rectangular or half-elliptic) has to be reconsidered.

5. Conclusions

We summarize the results of this study as follows.

- (1) With the ‘combined model’ as described, we were able to produce simulations that resemble the horizontal and vertical vibrations of the vocal folds. It was also possible to extract data about phonation frequencies and about pressures and flows throughout the vocal air duct. We compared the results with previously published models and found proper correspondence.
- (2) With this model we investigated the effects of varying the physical parameters, such as the driving lung pressure, the dimensions of the air duct compartments and the glottis, on the variables of the vibrations, such as the masses and their dimensions, the coupling elastic springs and dampers, the frequency spectra (peak height positions and ratios), and found different results for different parameter/variable combinations.
- (3) We also investigated how the adoption of different vocal tract area functions (corresponding to different vowels) affected the frequency spectra for the sound production.
- (4) The developed computer program enables us to produce sound, by feeding the resulting time recording of the signal produced for the pressure or the flow in the mouth, to a software synthesizer, coupled to the computer’s (or external) speakers. Depending on the quality of the audio system, the sound reasonably corresponds to expectations.
- (5) We also produced overlays of pictures of vertical cross sections of the two-mass model, while vibrating, with pictures taken from the depth-kymography measurements, and combined those in a accompanying movie. After slight adjustment of the measured and simulated frequencies, and of the phase delay, the overlapping of both images is very good. This overlay method enables us to compare measurements and simulations in a quantitative way. Hopefully, the method might trigger new series of research of vocal fold vibrations.
- (6) As a means to quantify the correspondence between measurements and simulations, we introduced a correspondence criterion, being the shortest distance between the upper point of the upper mass at the midline side and the measured curve of the surface of the vocal folds, and averaged over as many vibration cycles as available.

Acknowledgment

This work was supported by The Netherlands Technology Foundation STW (Stichting Technische Wetenschappen) project 06633, Applied Science Division of NWO (Natuurwetenschappelijk Onderzoek).

Appendix. Vocal folds dynamics model

We partly have used a model of Koizumi *et al* (Koizumi *et al* 1987, 1989, Nudelman and Hoyt 1989), which in turn was based on the model of Ishizaka and Flanagan (1972). The model consists of two masses on top of each other and connected by a spring-and-damper combination, on each side of the vocal folds. Different from these previous models, both masses are attached to the wall of the glottis by other spring-and-damper combinations. All spring-and-damper combinations have a horizontal and vertical component. In figure 1, the situation is sketched.

In principle, we consider all x - and y -coordinates of the two masses as variables in the model. Only the length of the glottis l_g , is taken constant (but can be used as a parameter), and both the widths w and depths d can be taken as variables in the simulations. This differs from the previous models.

To distinguish our model from those previous models, we will name it the ‘combined model’.

The rest positions of all coordinates (x , y , d , w) are denoted with the subscript 0, and the deviations from that rest position are indicated with a prime, e.g.

$$x(t) = x_0 + x'(t). \quad (\text{A.1})$$

The two masses are subject to three types of forces: spring forces (f), damping forces (g) and forces exerted by air pressures (F_p). The spring forces are determined by the actual position of the masses with respect to each other and (for the lower mass) to the walls of the glottis. The damping forces are proportional to the actual velocity of the masses. The pressure forces are determined by the actual areas of the masses exposed to the air pressures in the x - and y -directions (which may vary according to the position and thickness of the masses), the impedances for the air flow in the various compartments of the air duct and the air flow itself. These impedances are flow resistances (R), inertias (L) and compliances (C). This means that in the calculations, in each time step we have a double problem to solve.

- Calculation of the pressures and flows; this can be achieved by using an electrical analogue with circuits for the pressures, flows and impedances.
- Solution of the equations of motion for the two masses; for this goal the actual values for pressures and flows are needed.

The cross-sectional areas of the two parts of the glottis, when open, are given by

$$A_{gi}(t) = A_{gi0} + A'_{gi}(t), \quad \text{with} \quad A_{gi}(t) = 2l_g x_i(t), \quad \text{etc} \quad \text{and} \quad i = 1, 2. \quad (\text{A.2})$$

We have conservation of the total mass (ρ_i is the tissue density in kg m^{-3}):

$$m_1 + m_2 = \text{constant}, \quad \text{with} \quad m_i = \rho_i l_g d_i w_i, \quad i = 1, 2, \quad (\text{A.3})$$

or

$$d_1 w_1 + d_2 w_2 = \text{constant}, \quad \text{with all variables } d_1, d_2, w_1 \text{ and } w_2 > 0. \quad (\text{A.4})$$

The equations of motion are given by

$$\begin{aligned}
 EM_1 &\equiv m_1 \ddot{x}'_{c1} + \dot{m}_1 \dot{x}'_{c1} + g_{1x} + f_{1x} + g_{cx} + f_{cx} - F_{p,1x} = 0 \\
 EM_2 &\equiv m_1 \ddot{y}'_{c1} + \dot{m}_1 \dot{y}'_{c1} + g_{1y} + f_{1y} - g_{cy} - f_{cy} - F_{p,1y} = 0 \\
 EM_3 &\equiv m_2 \ddot{x}'_{c2} + \dot{m}_2 \dot{x}'_{c2} + g_{2x} + f_{2x} - g_{cx} - f_{cx} - F_{p,2x} = 0 \\
 EM_4 &\equiv m_2 \ddot{y}'_{c2} + \dot{m}_2 \dot{y}'_{c2} + g_{cy} + f_{cy} - F_{p,2y} = 0, \quad y_2 = y_1 + d_1 + d_2,
 \end{aligned} \tag{A.5}$$

with x_{c1} , x_{c2} , y_{c1} and y_{c2} the centre-of-mass coordinates, and f , g and F_p representing forces exerted by springs, dampers and air pressures, respectively. (In their papers (Koizumi *et al* 1987, 1989), Koizumi *et al* did not use the centre-of-mass coordinates.)

The centre-of-mass coordinates and the spring and damping forces are given by

$$\begin{aligned}
 x_{c1} &= x_1 + w_1/2, & y_{c1} &= y_1 + d_1/2, & x_{c2} &= x_2 + w_2/2, & y_{c2} &= y_2 - d_2/2, \\
 f_{1x} &= k_{1x} x'_1 (1 + \eta_{1x} x_1'^2), & f_{1y} &= k_{1y} y'_1 (1 + \eta_{1y} y_1'^2), \\
 f_{2x} &= k_{2x} x'_2 (1 + \eta_{2x} x_2'^2), & f_{2y} &= 0, \\
 f_{cx} &= k_{cx} x_c (1 + \eta_{cx} x_c^2), & x_c &= x_{c1} - x_{c2}, \\
 f_{cy} &= k_{cy} y_c (1 + \eta_{cy} y_c^2), & y_c &= \frac{1}{2}(d_1 + d_2), \\
 g_{1x} &= r_{1x} \dot{x}'_{c1}, & g_{1y} &= r_{1y} \dot{y}'_{c1}, & g_{2x} &= r_{2x} \dot{x}'_{c2}, & g_{2y} &= 0, & g_{cx} &= r_{cx} \dot{x}_c, \\
 g_{cy} &= r_{cy} \dot{y}_c, \\
 r_{ij} &= 2(q_{ij} + \zeta_{ij}) \sqrt{k_{ij} m_i} \quad (i = 1, 2, c; j = x, y).
 \end{aligned} \tag{A.6}$$

where k_{ij} and η_{ij} are linear and nonlinear spring constants and ζ_{ij} are damping constants. The value of q_{ij} is given by 1 if the masses are in the ‘closure’ condition, and 0 otherwise (‘open’). For the coupling forces, closure is not applicable. In the closure condition, we have extra repulsing spring forces, given by

$$f_{h,1x} = h_{1x} x_1 (1 + \eta_{h,1x} x_1^2), \quad f_{h,2x} = h_{2x} x_2 (1 + \eta_{h,2x} x_2^2), \tag{A.7}$$

to be added to f_{1x} and f_{2x} , respectively, and where now h_{ix} and $\eta_{h,ix}$ ($i = 1, 2$) are the ‘extra’ spring constants in the ‘closure’ situation. In the Ishizaka–Flanagan model, no y -coordinates or y -forces are present. In the Koizumi model, the forces f_{2x} and g_{2x} are absent, and the forces do not work on the centre-of-masses.

The forces exerted by pressures are determined by the distribution of pressures over the whole region (Ishizaka and Flanagan 1972). We denote by P_s the pressure at the beginning of the contraction to the glottis, P_{11} and P_{12} at the entrance and exit of the lower mass in the glottis, respectively, P_{21} and P_{22} the same for the upper mass and P_1 at the expansion to the vocal tract (Ishizaka and Flanagan 1972). The forces are given in tables 4 and 5.

The pressures can be expressed using flow resistances and inertias:

$$\begin{aligned}
 R_g &= R_{sg} + R_{m1} + R_{g,\text{mid}} + R_{m2} + R_{\text{exp}}, \\
 L_g &= L_{sg} + L_{m1} + L_{m2},
 \end{aligned} \tag{A.8}$$

Table 4. Forces from pressures in the Ishizaka–Flanagan and Koizumi models.

		Situation of masses	F_{1x}	F_{1y}	F_{2x}	F_{2y}
A	$x_1 > x_2 > 0$		P_{m1}	[K]: $P_{11}(w_{10}-x_1')l_g$	P_{m2}	[K]: $(P_{12}-P_{22}) \cdot (x_1'-x_2') \cdot l_g$
	$x_1 = x_2 > 0$					
	$x_2 > x_1 > 0$					
B	$x_1 > 0;$ $x_2 < 0$		$P_s d_1 l_g$	= A	$P_s d_2 l_g$	[K]: $+P_s (x_1'-x_2') \cdot l_g$ [I]: 0
C	$x_1 < 0;$ $x_2 > 0$		= B	[K]: $P_{11}w_{10}l_g$ [I]: 0	0	0
D	x_1 and $x_2 < 0$		= B	= C	0	0

In the column ‘Situation of masses’ the position of the two masses is sketched with respect to the midline of the glottis. I = Ishizaka–Flanagan only; K = Koizumi only. $P_{m1} = 1/2(P_{11}+P_{12})d_1l_g$; $P_{m2} = 1/2(P_{21} + P_{22})d_2l_g$.

Table 5. Forces from pressures in the combined model.

		Situation of masses	F_{1x}	F_{1y}	F_{2x}	F_{2y}
A	$x_1 > x_2 > 0$		P_{m1}	$P_{11} \cdot w_1 l_g$	P_{m2}	$P_{21}(x_1-x_2)l_g - P_{22} w_2 l_g$
	$x_1 = x_2 > 0$			= A		
	$x_2 > x_1 > 0$			$P_{11} \cdot w_1 l_g - P_{12}(x_2-x_1) l_g$		
B	$x_1 > 0;$ $x_2 < 0$		$P_s d_1 l_g$	$P_{11} \cdot w_1 l_g$	$P_s d_2 l_g$	$-P_{22} w_2 l_g + P_{21} x_1 l_g$
C	$x_1 < 0;$ $x_2 > 0$		= B	$P_{11}(w_{10}+x_{10}) l_g - P_{12} x_2 l_g$	0	$-P_{22} w_2 l_g$
D	x_1 and $x_2 < 0$		= B	$P_{11}(w_{10}+x_{10}) l_g$	0	$-P_{22} w_2 l_g$

In case of closure, w_i is maximum (then $w_i = w_{i0} + x_{i0}$, provided that w_i is taken as variable). For P_{m1} and P_{m2} , see table 4.

with (Ishizaka and Flanagan 1972)

$$\begin{aligned}
 R_{sg} &= 1.37 \frac{\rho_a}{2} \frac{|U_g|}{A_{g1}^2}; & L_{sg} &= \int_0^{l_c} \frac{\rho_a dy}{A_c(y)} \\
 R_{m,i} &= \frac{12\mu \cdot l_g^2 d_i}{A_{gi}^3}; & L_{mi} &= \frac{\rho_a d_i}{A_{gi}}; & i &= 1, 2 \\
 R_{mid} &= \frac{1}{2} \rho_a |U_g| \cdot [A_{g2}^{-2} - A_{g1}^{-2}] \\
 R_{exp} &= -\frac{1}{2} \rho_a \frac{|U_g|}{A_{g2}^2} \cdot 2 \frac{A_{g2}}{A_t} \left(1 - \frac{A_{g2}}{A_t} \right),
 \end{aligned} \tag{A.9}$$

with ρ_a the air density (kg m^{-3}) and μ the air viscosity (Pa s). For A_t , the cross-sectional area of the first component of the vocal tract (first T-circuit) can be taken. It is seen that some resistances are flow dependent, and therefore, the circuitry is not linear, as in electrical circuits.

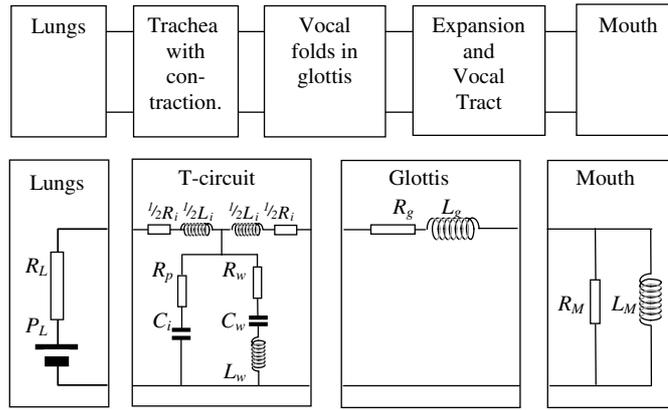


Figure 10. The pressures–flows system, modelled analogous to an electronic circuit. The T-circuit applies to the trachea and the vocal tract compartments. The components are explained in the text. After Ishizaka and Flanagan (1972), Titze (1973, 1974, 1984), Titze and Strong (1975) and Koizumi *et al* (1987).

This leads to the following scheme for calculating the pressures in the glottis:

$$\begin{aligned}
 P_s - P_{11} &= R_{sg} U_g + L_{sg} V_g \\
 P_{11} - P_{12} &= R_{m,1} U_g + L_{m,1} V_g \\
 P_{12} - P_{21} &= R_{\text{mid}} U_g \\
 P_{21} - P_{22} &= R_{m,2} U_g + L_{m,2} V_g \\
 P_{22} - P_t &= R_{\text{exp}} U_g
 \end{aligned} \tag{A.10}$$

where U_g is the glottis air flow ($\text{m}^3 \text{s}^{-1}$), V stands for dU/dt and the subscripts c , g and t refer to the contraction part, the glottis and the expansion part, respectively. A_c is the cross-sectional area (dependent on the y -coordinate) of the contraction part. A_c and l_c refer to the cross-sectional area and length of the contraction part before the glottis, respectively. A_t is the area of the expansion opening after leaving the glottis.

In order to avoid singularities at the closure point, we inserted a smoothing factor CF, given by

$$\begin{aligned}
 \text{CF} &= 1 && \text{if } x_1 > s \cdot x_{10}; && \text{with } 0 < s < 1, \\
 &= \frac{1}{2} \left(1 + \frac{x_1}{2s \cdot x_{10}} \right) && \text{if } |x_1| \leq s \cdot x_{10} \\
 &= 0 && \text{if } x_1 < -s \cdot x_{10}
 \end{aligned} \tag{A.11}$$

and analogously for x_2 . Normally we choose $s = 0.01$.

We incorporated two options for the widths of the masses:

- variable width, according to the x -variable, thus setting $w_i(t) + x_i(t) = w_0 + x_0$,
- constant width, thus with $w_i(t) = w_0$.

Both options have different consequences for the distribution of the total mass over the two masses. The first option implies that the wall side of the mass is fixed.

For the calculation of the pressures and the flows, we adopted a circuit model, as described in figure 10. It is an electrical analogue, consisting of a series of blocks, each containing

resistances (R), inertias (L) and compliances (C), based on a similar model of Ishizaka and Flanagan (1972). The trachea part contains one T-circuit compartment and the vocal tract part contains several T-circuit compartments, each with its own characteristics, depending on the sound to be phonated.

The details of the various components of the T-circuits are (Ishizaka and Flanagan 1972)

$$R_i = \frac{S_i l_i}{A_i^2} \sqrt{\frac{1}{2} \rho_a \mu \omega}; \quad L_i = \frac{\rho_a l_i}{A_i}; \quad C_i = \frac{A_i l_i}{\rho_a c^2}, \quad (\text{A.12})$$

with c the sound velocity (m s^{-1}), and l_i , S_i and A_i the length, circumference and cross-sectional area, respectively, of the vocal tract segment, imagined as a cylinder compartment. The variables R , L and C have units ($\text{kg (m}^4 \text{ s)}^{-1}$), (kg m^{-4}) and ($\text{m}^4 \text{ s}^2 \text{ kg}^{-1}$) respectively. The initial frequency ω is given by the eigenfrequency of the set of masses:

$$\omega = \sqrt{\frac{k_{1x} + k_{cx} + k_{2x}}{m_1 + m_2}}. \quad (\text{A.13})$$

The values for the lengths and areas of the compartments of the vocal tract depend on the phonation and can be found in Story and Titze (1998) and Story *et al* (1996, 1998). We included an extra resistance R_p for calculation reasons. The right vertical part of the T in the circuit allows us to include effects of yielding walls (R_w , L_w , C_w). Details are given by Ishizaka *et al* (1975).

The mouth is modelled with a resistance R_M and an inertia L_M in parallel (Flanagan 1968, Flanagan and Cherry 1969, Ishizaka 1966, Dudgeon 1970, Ishizaka and Flanagan 1972), as the load of a circular piston in an infinite baffle with the cross-sectional area A_M (which may be equal to that of the final tract compartment):

$$R_M = \frac{128 \rho_a c}{9 \pi^2 A_M}; \quad L_M = \frac{8 \rho_a}{3 \pi \sqrt{\pi A_M}}. \quad (\text{A.14})$$

The complete electrical analogue is shown in figure 10 for N compartments in the vocal tract. We have a set of $2(N+1)$ linear circuit equations to solve. The circuit equations are

$$\begin{aligned} U_{0,0}: & \quad -P_L + \left(R_L + \frac{1}{2}R_{i,0}\right)U_{0,0} + \frac{1}{2}L_{i,0}V_{0,0} + \frac{1}{C_{i,0}} \int [U_{0,0} - U_{0,1}] dt \\ & \quad + R_{p,0}(U_{0,0} - U_{0,1}) = 0 \\ U_{n,0}: & \quad -R_{w,n-1}(U_{n-1,1} - U_{n,0}) - \frac{1}{C_{w,n-1}} \int [U_{n-1,1} - U_{n,0}] dt - L_{w,n-1}(V_{n-1,0} - V_{n,0}) \\ & \quad + \frac{1}{2}(R_{i,n-1} + 2Z_{Rg} + R_{i,n})U_{n,0} + \frac{1}{2}(L_{i,n-1} + 2Z_{Lg} + L_{i,n})V_{n,0} \\ & \quad + R_{p,n}(U_{n,0} - U_{n,1}) + \frac{1}{C_{i,n}} \int [U_{n,0} - U_{n,1}] dt = 0 \quad \text{for } n = 1, \dots, N \\ U_{n,1}: & \quad -\frac{1}{C_{i,n}} \int [U_{n,0} - U_{n,1}] dt - R_{p,n}(U_{n,0} - U_{n,1}) + L_{w,n}(V_{n,1} - V_{n+1,0}) \\ & \quad + \frac{1}{C_{w,n}} \int [U_{n,1} - U_{n+1,0}] dt + R_{w,n}(U_{n,1} - U_{n+1,0}) = 0 \\ & \quad \text{for } n = 0, 1, \dots, N \\ U_{N+1,0}: & \quad -R_{w,N}(U_{N,1} - U_{N+1,0}) - \frac{1}{C_{w,N}} \int [U_{N,1} - U_{N+1,0}] dt - L_{w,N}(V_{N,0} - V_{N+1,0}) \\ & \quad + \left(\frac{1}{2}R_{i,N} + R_M\right)U_{N+1,0} - R_M U_{N+1,1} + \frac{1}{2}L_{i,N}V_{N+1,0} = 0 \\ U_{N+1,1}: & \quad -R_M(U_{N+1,0} - U_{N+1,1}) + L_M V_{N+1,1} = 0 \end{aligned} \quad (\text{A.15})$$

where circuits $N + 1,0$ and $N + 1,1$ denote the pre-mouth and mouth regions, respectively, and where $z_{Rg} = R_g$ and $z_{Lg} = L_g$ for $n = 1$ (glottis), and $z_{Rg} = 0$ and $z_{Lg} = 0$ otherwise. Since usually the vocal tract area functions, as published by Titze *et al* (Story *et al* 1996, 1998, Story and Titze 1998), may have about 45 compartments, the complete array may contain about 92 linear equations with 92 unknown variables to solve, each time step over and over.

In the simulations, the first and second time derivatives \dot{X} and \ddot{X} of a variable X are calculated as

$$\begin{aligned}\dot{X}(t) &= \frac{X(t) - X(t - \Delta t)}{\Delta t} \\ \ddot{X}(t) &= \frac{X(t) - 2X(t - \Delta t) + X(t - 2\Delta t)}{(\Delta t)^2},\end{aligned}\tag{A.16}$$

where Δt is the time step. The time integral is evaluated as

$$\int_0^t X(t) \cdot dt = \left[X(t) + \sum_{k=1}^K X(t - i \cdot \Delta t) \right] \cdot \Delta t, \quad \text{with} \quad K = \frac{t}{\Delta t}.\tag{A.17}$$

In the calculations of pressures and flows, the values obtained at previous times are added to the right-hand side quantities of the equations. This set of equations can be solved analytically using band-matrix algorithms.

The four equations of motion have to be solved in an iterative way, using x_1 , x_2 , y_1 and d_1 as the variables. Then y_2 and d_2 are known using the constraints in equations (A.3) and (A.4). The iterative procedure is like the Newton–Raphson algorithm, in which the variables subsequently are varied around their previous value, until

$$\sum_{k=1}^4 (EM_k)^2 = \text{minimum}\tag{A.18}$$

is reached. In doing so, the respective step sizes are gradually reduced and the iteration process is repeated until all step sizes are smaller than the original values times a small number (order 10^{-6}).

In case the effects of yielding walls are negligible or not present, we can simplify equation (A.15) to

$$\begin{aligned}U_0 : \quad & -P_L + \left(R_L + \frac{1}{2} R_{i,0} \right) U_0 + \frac{1}{2} L_{i,0} V_0 + R_{p,0} (U_0 - U_1) + \frac{1}{C_{i,0}} \int [U_0 - U_1] dt = 0 \\ U_n : \quad & R_{p,n-1} (U_n - U_{n-1}) + \frac{1}{C_{i,n-1}} \int [U_{n,0} - U_{n-1}] dt + \frac{1}{2} (R_{i,n-1} + 2Z_{Rg} + R_{i,n}) U_n \\ & + \frac{1}{2} (L_{i,n-1} + 2Z_{Lg} + L_{i,n}) V_n + R_{p,n} (U_n - U_{n+1}) + \frac{1}{C_{i,n}} \int [U_{n,0} - U_{n+1}] dt = 0 \\ \text{for} \quad & n = 1, \dots, N \\ U_{N+1} : \quad & R_{p,N} (U_{N+1} - U_N) + \frac{1}{C_{i,N}} \int [U_{N+1} - U_N] dt + \frac{1}{2} L_{i,N} V_{N+1} \\ & + \frac{1}{2} R_{i,N} U_{N+1} + R_M (U_{N+1} - U_{N+2}) = 0 \\ U_{N+2} : \quad & R_M (U_{N+1} - U_{N+2}) + L_M V_{N+2} = 0\end{aligned}\tag{A.19}$$

where circuits $N + 1$ and $N + 2$ denote the pre-mouth and mouth regions, respectively. The set of equations reduces to $N + 3$ equations.

In order to prevent instabilities in the iteration process, some additional constraints are built in.

- All four variables are limited to a certain maximum value.
- All step sizes have a maximum and a minimum limit.
- In case the width of the mass is taken as variable, this width has a minimum value of half the value in the rest position.

The flow in the steady-state situation (dc flow) can be calculated by assuming the impedances of all inertias being zero and those of all compliances being infinite. Then all local flows will be equal. Since some resistances are flow dependent, a quadratic equation with the flow as the unknown variable results.

References

- Dudgeon D E 1970 Two-mass model of the vocal cords *J. Acoust. Soc. Am.* **48** 118A
- Flanagan J L 1968 Use of an interactive laboratory computer to study an acoustic oscillator model of the vocal cords *IEEE Trans. Audio Electroacoust.* **17** 2–6
- Flanagan J L and Cherry L 1969 Excitation of vocal tract synthesizers *J. Acoust. Soc. Am.* **45** 764–9
- George N A, de Mul F F M, Qiu Q, Rakhorst G and Schutte H K 2008 Depth-kymography: high-speed calibrated 3D imaging of human vocal fold vibration dynamics *Phys. Med. Biol.* **53** 2667–75
- Ishizaka K 1966 On models of the larynx *J. Acoust. Soc. Japan* **22** 293–4
- Ishizaka K and Flanagan J L 1972 Synthesis of voiced sounds from a two-mass model of the vocal cords *Bell Syst. Tech. J.* **51** 1233–67
- Ishizaka K, French J and Flanagan J L 1975 Direct determination of vocal tract wall impedance *IEEE Trans. Acoust. Speech Signal Process.* **23** 370–3
- Koizumi T, Taniguchi S and Hiromitsu S 1987 Two-mass models of the vocal cords for natural sounding voice synthesis *J. Acoust. Soc. Am.* **82** 1179–92
- Koizumi T, Taniguchi S and Hiromitsu S 1989 Reply to comments on ‘Two-mass models of the vocal cords for natural sounding voice synthesis’ *J. Acoust. Soc. Am.* **85** 2221–3
- Nudelman H B and Hoyt B D 1989 Comments on ‘Two-mass models of the vocal cords for natural sounding voice synthesis’ *J. Acoust. Soc. Am.* **85** 2220–1
- Schutte H K 1986 Aerodynamics in phonation *Acta Otorhinolaryngologica Belg.* **40** 344–57
- Schutte H K and Miller D G 1986a Transglottal pressures in professional singing *Acta Otorhinolaryngologica Belg.* **40** 395–404
- Schutte H K and Miller D G 1986b The effect of F0/F1 coincidence in soprano high notes on pressure at the glottis *J. Phonetics* **14** 385–92
- Schutte H K, Svec J G and Sram F 1998 First results of clinical application of videokymography *Laryngoscope* **108** 1206–10
- Story B and Titze I 1998 Parameterization of vocal tract area functions by empirical orthogonal modes *J. Phonetics* **26** 233–60
- Story B, Titze I and Hoffman E A 1996 Vocal tract area functions from magnetic resonance imaging *J. Acoust. Soc. Am.* **100** 537–53
- Story B, Titze I and Hoffman E A 1998 Vocal tract area functions for an adult female speaker based on volumetric imaging *J. Acoust. Soc. Am.* **104** 471–87
- Svec J G and Schutte H K 1996 Videokymography: high-speed line scanning of vocal fold vibration *J. Voice* **10** 201–5
- Svec J G, Sram F and Schutte H K 2007 Videokymography in voice disorders: what to look for? *Ann. Otol. Rhinol. Laryngol.* **116** 172–80
- Svec J G, Sram F and Schutte H K 2009 Videokymography *The Larynx* 3rd edn, ed M P Fried and A Ferlito (San Diego, CA: Plural Publishing)
- Titze I 1973 The human vocal cords: a mathematical model, part I *Phonetica* **28** 129–70
- Titze I 1974 The human vocal cords: a mathematical model, part II *Phonetica* **29** 1–21
- Titze I 1984 Parameterization of the glottal area, glottal flow and vocal fold contact area *J. Acoust. Soc. Am.* **75** 570–80
- Titze I and Strong W 1975 Normal model in vocal cord tissue *J. Acoust. Soc. Am.* **57** 736–44