TIME-DOMAIN MODEL OF THE SINGING VOICE

Malte Kob, Nils Alhäuser, Ulrich Reiter

Institute of Technical Acoustics
Technical University Aachen
Templergraben 55, D-52056 Aachen, Germany
kob@akustik.rwth-aachen.de

ABSTRACT

A combined physical model for the human vocal folds and vocal tract is presented. The vocal fold model is based on a symmetrical 16 mass model by Titze. Each vocal fold is modeled with 8 masses that represent the mucosal membrane coupled by non-linear springs to another 8 masses for the vocalis muscle together with the ligament. Iteratively, the value of the glottal flow is calculated and taken as input for calculation of the aerodynamic forces. Together with the spring forces and damping forces they yield the new positions of the masses that are then used for the calculation of a new glottal flow value. The vocal tract model consists of a number of uniform cylinders of fixed length. At each discontinuity incident, reflected and transmitted waves are calculated including damping. Assuming a linear system, the pressure signal generated by the vocal fold model is either convoluted with the Green's function calculated by the vocal tract model or calculated interactively assuming variable reflection coefficients for the glottis and the vocal tract during phonation. The algorithms aim at real-time performance and are implemented in MATLAB.

1. INTRODUCTION

In the last years, the quality of the synthesized human voice could be significantly improved due to the availability of fast computers that are capable to solve iteratively complex systems of equations that describe the physics of flow generation and sound propagation in the human vocal folds and vocal tract. The main advantage of the direct description of the physical laws of the voice generation is the direct access to all important parameters during the calculation. This is not possible when a simple oscillator-filter model approach is used. As a drawback, the calculations become very time-consuming the more accurate the model describes reality. Therefore, approximations must be assumed which have to be chosen carefully. This is why at present, for the aerodynamic generation of noise sound no simple models are available.

The models presented here will not include sound generated by turbulences as it would be necessary for an adequate generation of speech. At this stage of accuracy, we aim at simulation that is based on measured quantities wherever they can be derived from in vivo or in vitro investigations.

2. MODELS

In the following, the physical models for the vocal folds and the vocal tract are briefly described.

2.1. Vocal fold model

The model used for the glottis is based on the 16 mass model by Titze described in [1] and [2] but includes some results of recent research. Each vocal fold is modeled with 8 masses m that represent the mucosal membrane coupled to another 8 masses M for the vocalis muscle together with the ligament. Figure 1 illustrates the model.

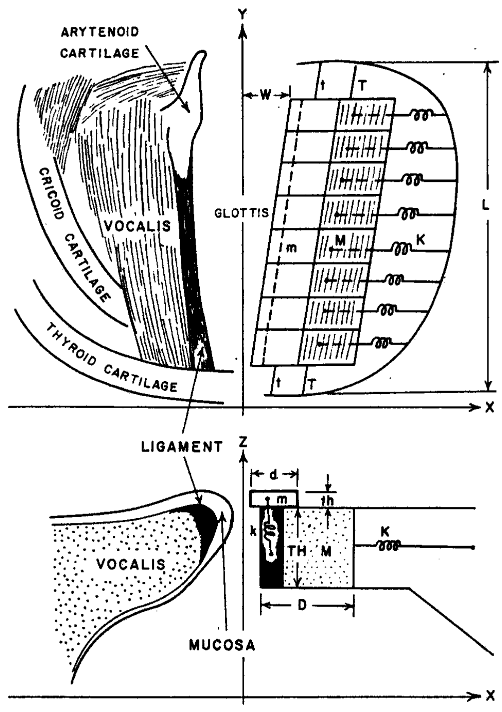

Figure 1. *Model of the vocal folds after Titze [3].*

Each of the masses is coupled to its neighbored masses or the cartilage by springs allowing for vertical and horizontal transversal movement. Although this relatively high number of degrees of freedom is not necessary for an oscillating fold model (see e.g. [4]), the adjustment of geometrical and physical parameters for the generation of different voice registers can be related to measured values more easily. As an example, the effect of higher modes of the vocal folds that has been reported for high pitches [3] can be modeled. For each time step, the value of the glottal flow is calculated and taken as input for calculation of the aerodynamic forces that result from Bernoulli's equation if jet generation is assumed at the point of minimum area between the masses. This point depends on the actual position of the masses during the phonation cycle as illustrated in Figure 2.

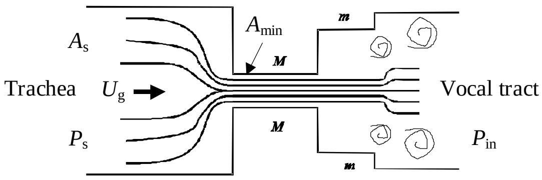

Figure 2. *Jet generation in the glottis*.

For the transglottal pressure, this assumption yields the following expression:

$$P_{tg} = P_s - P_{in} = \frac{1}{2}\rho U_g^2 \left(\frac{1-k_e}{A_{min}^2} - \frac{1}{A_s^2} \right), \qquad (1)$$

with the subglottal pressure P_s and area A_s, the supraglottal pressure P_{in}, the density of air ρ, the glottal flow U_g, the pressure recovery coefficient k_e [5] and the minimum area between the vocal folds A_{min}.

Together with the non-linear spring forces (including additional springs during contact) and non-linear damping forces, the aerodynamic forces yield the new positions of the masses that are then used for the calculation of a new glottal flow value.

Most physical parameters that are relevant for the generation of the singing voice can be changed during phonation. This makes it possible to simulate a change of register or varying active stress of the vocalis muscle. As example, in Figure 3 the change of the calculated vocal fold pressure is shown as the voice is switched from modal to head register.

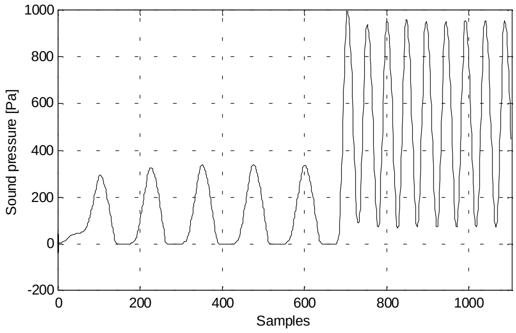

Figure 3. *Vocal fold signal change from modal to head voice*.

For this change of register, the following parameters have been modified:

Parameter	Modal register	Head register
Distance of the folds	0.025 mm	0.175 mm
Active Stress factor	1	0.1
Mass factor	1	0.1
Mucosa tension factor	1	2
Vocalis tension factor	1	0.1

Table I. *Parameter modification for change from modal to head register*.

These parameter changes are taken from literature [6],[7]. Of course, the change of parameters due to the complex interaction of singing and hearing with its feedback on the neural level is not included.

2.2. Vocal tract model

Most modern physical models for the vocal tract are based on measured geometrical data from MRI (nuclear magnetic resonance images) and EBCT (electron beam computed tomography). The data used in this vocal tract model are the calculated equivalent area functions, i.e. the radius of a circle representing the actual vocal tract area vs. the distance of the segment from the glottis. The area functions are taken from [8].

2.2.1. Cylindrical model

The vocal tract is described by a number of evenly spaced cylinder segments of approx. 4 mm length. Figure 4 describes the superposition of the reflected and transmitted waves at the discontinuities k.

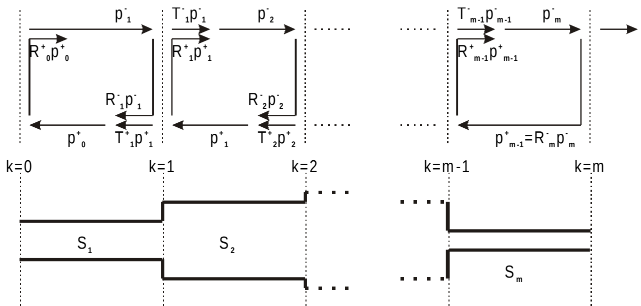

Figure 4. *Reflections and transmissions at the discontinuities of the cylinder segments*.

At every interface k, the reflected and transmitted wave pressures are calculated by multiplication of the incident wave pressure with the reflection coefficients R or transmission coefficients T and a damping constant that has been derived from measurements on a vocal tract model. R and T are calculated from the area values on both sides of each discontinuity. The contributing waves are superimposed for each time step. Because of the small and equal spacing of the segments the sample rate for this model is fixed to

$$F_s = \frac{c_0}{\Delta x} = \frac{350\ m/s}{3.968\ mm} = 88200 \text{ Hz} . \qquad (2)$$

At the glottis and at the mouth the calculation of the coefficients is a bit more difficult. Due to the time-varying area of the glottis, the reflection coefficient can be calculated for each time step or set to an average value. The radiation impedance at the mouth is modeled by the Green's function of a sphere with equivalent mouth area that yields the following expression for the reflection coefficient at the mouth opening:

$$R_m^- = -\frac{c_0}{2r}\exp\left(-\frac{c_0}{2r}t\right)\varepsilon(t) , \qquad (3)$$

with speed of sound c_0, mouth radius r and Heaviside function $\varepsilon(t)$. The pressure superposition at the last interface finally gives the sound pressure at the mouth.
For verification of the calculated vocal tract transfer functions for different vowels, a "washer" kit was built consisting of a set of 120 discs with bore radii ranging from 3 mm to 29 mm. For representation of the vocal tract area functions, between 40 and 46 discs were assembled, depending on the vowel. Figure 5 shows the comparison of calculated and measured transfer functions for the vowel |i|.

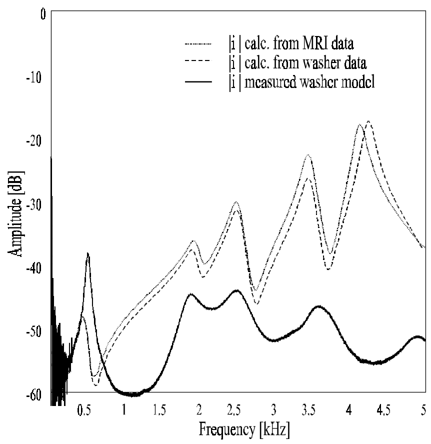

Figure 5. *Comparison of transfer functions.*

The upper two curves show the calculated transfer functions based on the exact area function values from the MRI data (dotted line) and the approximated areas derived from bores with integer mm radius (dashed line). The latter area values were consistent with those within the vocal tract washer model of which the transfer function (solid line) has been measured using a two channel FFT measurement setup. It is obvious that the discretisation of the area functions has little influence on the transfer function whereas the measured transfer function values differ significantly from the calculated ones, less in frequency but more in amplitude. The effect of the too strong formants could be due to the frequency-independent damping that should increase with frequency. Most formant frequencies match within 10% the values measured by Story et al. [8].

2.2.2. *Conical model*

Preliminary investigations [8] indicate that the number of cylinders can be reduced to 8 conical ones without a significant change of the area functions. Application of a fast recursive algorithm based on work by Barjau et al. [9] allows for arbitrary sampling rates and interactive variation of geometric and acoustic parameters. The application of this algorithm for the vocal tract and nasal tract is under development and will increase the calculation time significantly.

3. IMPLEMENTATION IN MATLAB

Although an implementation in DSP code or C^{++} might have reduced the calculation time, MATLAB was chosen as the programming language because of the easy structure and graphical interface facilities. The calculation times for the simulations are at present still far from real time (approx. 500 for the vocal fold model on a Pentium II processor with 450 MHz). Nevertheless, the parameter variations can be studied in acceptable time.

3.1. Convolution

The classical oscillator-filter model assumes a linear system for the fold-tract combination (see Figure 6).

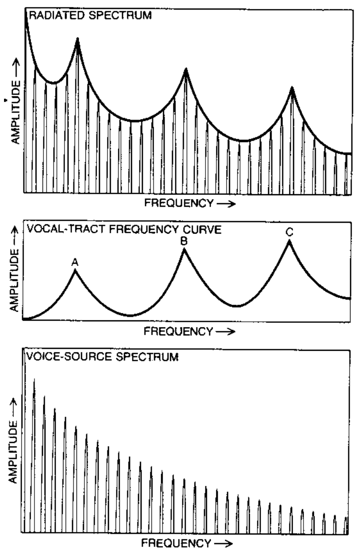

Figure 6. *Schematic representation of the synthesized voice spectrum (top) as combination of oscillator (bottom) and filter (middle), taken from [10].*

Therefore, no variation of the impedances on both sides of the glottis are included during phonation. In the program, this case is performed by separate calculation of the fold signal, assuming a constant trachea and vocal tract impedance, and a separate calculation of the transfer function of the vocal tract, assuming a constant, average cross section of the area between the vocal folds. After calculation of each, the combined signal results as the convolution of both.

3.2. Interactive Iteration

The second calculation way available is the interactive calculation of both models. At each timestep pressure values P_{in}^{+} and P_{s}^{-} are generated in the glottis from P_{phon}, using reflected pressure from trachea P_{s}^{+} and vocal tract P_{in}^{-} (which was p_{0}^{+} in Figure 4).
Figure 7 illustrates the wave propagation in and from the glottis.

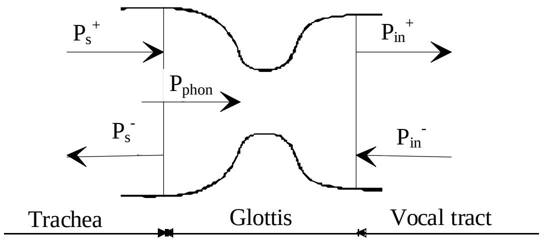

Figure 7. *Wave propagation in and from the glottis.*

These values can be monitored during calculation. The output signal at the other end of the waveguide ladder of the vocal tract is the sound pressure radiated from the mouth opening.
The pressure signal derived from the interactive calculation method sounds more lively than the convoluted signal. Also, the vowels sound much more realistic. Sound examples will be given at the conference.

3.3. Graphic user interface

The choice of the most important parameters for the calculation of the models can be accessed from the graphical user interface that was programmed with MATLAB. Figure 8 shows a screenshot of the working window.

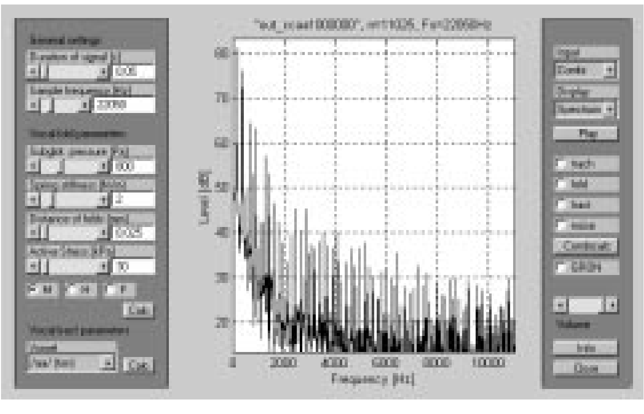

Figure 8. *Picture of the MATLAB graphic user interface.*

The left section shows various inputs for the adjustment of general settings like sample rate and signal duration and for each model the most important variables. Among those are physical parameters for the fold signal generation, the choice of the voice register and the vowel choice menu. On the right, the display section allows for choosing the file to be displayed and the representation of the signal (time, frequency, spectrogram). Also, the interaction mode for the combined calculation can be chosen.

4. CONCLUSIONS

The described MATLAB program models the generation of the singing voice with direct access to many parameters relevant for phonation. The time domain approach makes it possible to change physical parameters and to investigate various pressure values during the calculation. The vocal tract model has still to be improved with respect to frequency-dependent damping and a frequency-independent sample rate. Up to now, a model for the trachea and the nasal tract is still missing. These models are subject to current research.
Future research will investigate the quantitative contribution of the noise generated by turbulences that occur when the flow has to pass through narrow passages as are the glottis and the mouth opening. An interesting feature of the waveguide approach is the possibility to “inject” the noise signal at any position within the propagation line.
As symmetry of both folds is assumed, some effects like biphonation or bifurcation cannot be modeled. The extension of the present program code into a non-symmetrical description of the vocal folds will give insight into the behavior when higher modes of the folds are excited.

5. REFERENCES

[1] Titze, I. R. “The Human Vocal Cords: A Mathematical Model, Part I”, Phonetica Vol. 28, 129-170, 1973.

[2] Titze, I.R. “The Human Vocal Cords: A Mathematical Model, Part II” Phonetica Vol. 29, 1-21, 1974.

[3] Titze, I.R., and Strong, W.J. “Normal modes in vocal cord tissues” J. Acoust. Soc. Amer., Vol. 57 (3), 736-744, 1975.

[4] Lous, N.J.C., Hofmans, G.C.J., Veldhuis, R.N.J., and Hirschberg, A. “A Symmetrical Two-Mass Vocal-Fold Model Coupled to Vocal Tract and Trachea, with Application to Prosthesis Design” ACUSTICA – acta acustica, 84, 1135-1150, 1998.

[5] Ishizaka, K., and Flanagan, J.L. “Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords” Bell Syst. Tech. J. 51 (6), 1233-1268, 1972.

[6] Biesalski, P., and Frank, F. “Phoniatrie – Pädaudiologie” Vol. 1, “Phoniatrie”, 2nd Ed., Stuttgart; New York, Thieme, 1994.

[7] Wirth, G. “Stimmstörungen: Lehrbuch für Ärzte, Logopäden, Sprachheilpädagogen und Sprecherzieher” 4. Ed., Köln, Dt. Ärzte-Verl., 1995.

[8] Story, B.H., Titze, I.R., Hoffman, E.A. ”Vocal tract area functions from magnetic resonance imaging” J. Acoust. Soc. Amer., Vol. 100 (1), 537-554, 1996.

[9] Barjau, A., Keefe, D.H., Cardona, S. “Time-domain simulation of acoustical waveguides with arbitrarily spaced discontinuities” J. Acoust. Soc. Amer. Vol. 105 (3), 1951-1964, 1999.

[10] Sundberg, J. “The Acoustics of the Singing Voice” in “The physics of music”, Scientific American, W. H. Freeman and Company, San Francisco, 16-23, 1978.