

EXPERIMENTS WITH ANALYSIS BY SYNTHESIS OF GLOTTAL AIRFLOW

Johan Liljencrants

KTH, dept. of Speech, Music and Hearing

Stockholm, Sweden

ABSTRACT

The glottal model is based on a mechanical system with two basic degrees of freedom and is a variation on the classical two-mass model of Ishizaka and Flanagan (1972). A distinctive characteristic is that the two resonators are here organized as a translational system and a superimposed rotational system. Both resonator systems use one and the same mass element. The resonators are separately driven by the aerodynamics, the translational by the space average pressure in the glottal passage, and the rotational by the pressure gradient in the flow direction. The resonators are thus indirectly coupled by the aerodynamics. Experiences in static and dynamic approximations to natural voices using an analysis by synthesis strategy are summarized.

1. THE PHYSICAL GLOTTAL MODEL

The modeling is performed in four basic steps, relating to different domains. The first of these is the mechanical domain which describes the mechanical movements of simple mass/spring systems under influence of their driving forces, fig. 1. The second is a geometrical domain which represents a specific shape of the glottal slit, and this shape is derived from the instantaneous displacements of the mechanical system. The third is the aerodynamic step where pressures and flows in the glottal slit are derived from the detailed geometry. Finally an acoustic fourth step delivers boundary conditions in form of subglottal and supraglottal pressures, as influenced by the acoustic loads of the trachea and the vocal tract. When the pressure in the glottal slit has been derived it defines the forces and torques that act on the mechanical system. This closes the loop and thus the simulation of one time sample involves the sequential application of all four steps.

1.1 Control parameters

The rather considerable number of parameters used to specify the model are enumerated in Table 1. Their symbols are given here with boldface characters and pertain to the particular codes used with the modeling computer program.

Among the parameters a first group relates to the fold anatomy such as the length measures \mathbf{l} and \mathbf{D} , and the mass \mathbf{m} . The complete model also includes other parameters, not yet extensively used, for instance \mathbf{s} which accounts for left-right asymmetry, and \mathbf{s} for the mass ratio between the front and back halves of a cord. An imbalance here ($\mathbf{s} \neq 1$) will introduce a second harmonic movement when we look on the vocal fold as a

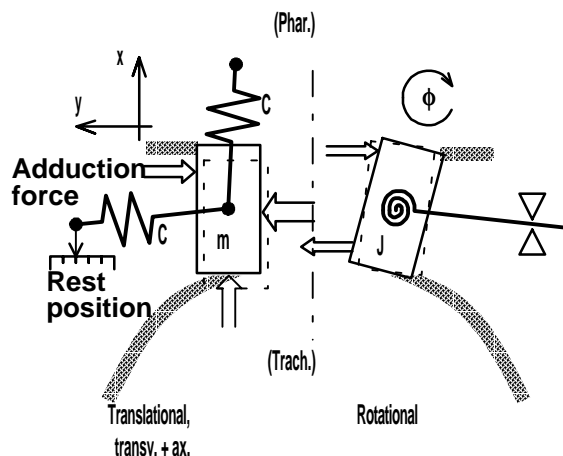


Fig. 1. The basic mechanical elements of the glottal model. Both sides have translational and rotational components, for clarity only one kind per side is shown here.

string and implies that the model is extended to contain two interconnected resonator systems for each fold.

\mathbf{B} is the amplitude of an elliptical bulge shape of the fold surface that also enters to define a hypothetical crossover volume when the folds collide. This volume is in the physical reality transformed into a deformation. In the model it is used to compute a displacement flow which may be added to the main glottal flow under control of the weighting factor \mathbf{j} . Furthermore the geometrical stage allows for computation of two small corrective flows, namely the 'lateral' pumping flow induced by the lateral motion of the free glottal surfaces, and the 'axial' pumping flow induced by the x displacement. These are optionally included using the weighting factors \mathbf{J} and \mathbf{L} .

A second parameter group relates to glottal articulation. Here the tension measure \mathbf{f} pertains to the fold modeled as an ideal string of length \mathbf{l} and total mass \mathbf{m} which would have its fundamental resonance at a frequency $f_0 = (\mathbf{f}/\mathbf{m}\mathbf{l})^{0.5}/2$. In a basic mass-stiffness resonator with the same resonance frequency this corresponds to a stiffness $k = (2\pi f_0)^2 \mathbf{m} = \pi^2 \mathbf{f}/\mathbf{l}$. In conjunction with this a ratio \mathbf{r} of rotational to translational resonance frequencies defines a torsional stiffness assigned to the rotation. The rest gaps \mathbf{y} and \mathbf{v} reflect abduction gestures. \mathbf{y} is a 'parallel' component of the fold rest placement found at the frontal cord end, and \mathbf{v} is an 'angular' component giving the additional distance at the other cord end.

symb	typ	dim.	meaning
g	100	μs	sample interval
Glottal anatomy			
l	11	mm	vocal fold length
D	1	mm	vocal fold thickness/2
m	70	mg	vocal fold mass
i	0.8	mm	gyration radius.
I	0	mm	center of gravity offset
B	0.3	mm	contact surface bulge
C	0	mm	chink width
k	1	mm	crit. y_c open fold stiffness
K	0.5	mm	crit. y_c closed fold stiffness
n	0.5	rad	critical rotational excursion
d	0.05	-	open fold damping $1/Q$
e	0.5	-	closed fold damping factor
E	1	-	factor of rotational damping
s	1	-	longitudinal tuning skew
S	1	-	lateral tuning skew
L	0.5	-	factor for axial pumping
J	0.5	-	factor for lateral pumping
j	0.5	-	fac. f. surf. wave pumping
Glottal articulation			
p	600	Pa	lung pressure
f	0.05	N	fold longitudinal tension
F	1	-	abduction factor
r	1.7	-	rot. to trans. res. freq. ratio
y	0	mm	rest gap/2
v	0	mm	rest angular gap/2
V	0	rad	rest divergence angle/2
Vocal tract articulation			
A	2	cm^2	area of trachea
H	600	Hz	1st tracheal res. frequency
Q	0.1	-	1st tracheal res. damping
a	3	cm^2	area of vocal tract
h	500	Hz	1st vocal res. frequency F1
G	1500	Hz	2nd vocal res. frequency F2
g	0.1	-	vocal res. damping factor
R	0	Ns/m^5	vocal DC resistance

Table 1. Parameters of the glottal model.

The lung pressure p should also be included in this parameter group. Liljencrants (1994) introduced an abduction variable, additional to y and v , which is interpreted as a (muscular) force operating to compress or separate the folds. This force is normalized with respect to the lung pressure and the nominal cord contact area, such that the abduction force is $f_{ab}=(F-1) \cdot p \cdot 2D \cdot l$. The value $F=0$ represents that the cords are pressed together with the same force as the lung pressure would develop on the cord area. This is with some margin enough to keep the glottis constantly closed, causing a total flow cutoff. $F=1$ is a neutral value giving zero adduction force, while higher values of F correspond to active abduction.

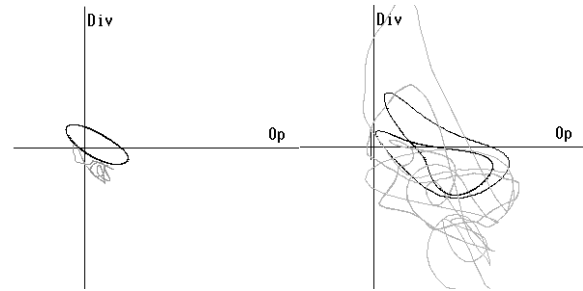


Fig. 2. Traces of divergence angle vs. opening displacement (solid) for one oscillating fold and the scaled excitation driving the fold (light shade). Left is normal chest register oscillation, right is a form of creaky oscillation.

A third parameter group describes oral articulation in terms of first formant frequency and bandwidth, complemented with an oral area to define the load impedance level. Additional similar parameters are specified for trachea to introduce a source impedance for the lung pressure.

2. OPERATING THE MODEL.

The non-linear properties of the aerodynamical as well as the mechanical system are very important, especially with voice qualities that deviate from 'normal'. Such voices are often difficult to synthesize (as well as to produce by a human speaker) since they may require rather accurate parameter tuning, and moreover it may be important to follow specific combined parameter changes rather than changes in single parameters.

The mechanical model acts as nonlinear second-order lowpass filters on the force and torque inputs which are derived from the aerodynamics. Using the proper scaling these inputs can be shown at commensurate scales in the same diagram - the output excursion y may be directly compared to the input force f divided by the stiffness k . The inputs are shown in fig. 2 as the irregular traces in lighter shade. They are generally much more complicated in detail structure than are the resulting lowpass smoothed movements of the folds. This is due partly to the highly nonlinear behavior of the aerodynamics, partly to the pressure variations of higher frequency that are inflicted by the acoustic systems of the trachea and the vocal tract. The left example of fig. 2 shows a number of minor loops at the lower right part of the input trace which are due to the acoustic resonance in the trachea.

The right part of fig. 2 represents an example of creaky voice which can be fabricated making delicate adjustments of several parameters, specifically a high closed fold damping e and an abduction effected by an increased rest position y . Here the ellipse has been much deformed and indeed split up into a double loop. The corresponding glottal flow exhibits a typical pattern of alternate higher and lower pulses. The specific mechanism in operation is that when the folds close a major part of the kinetic energy is dissipated such that oscillation will have to restart from

almost quiescent. The next cycle will then have such low amplitude that the folds barely close, if at all.

2.1. Static matching of glottal waveforms

The model has been used interactively with manual control of its parameters. Attempts were made to match in detail the shapes of human glottal flow waves for various voice qualities, obtained by inverse filtering. Data obtained for a number of samples with different glottal articulation types are given by Karlsson and Liljencrants (1994), and by Karlsson (1995).

As general comment it must be stated that because of the large number of controlling parameters there is a severe demand on the operator to use a conservative strategy in the analysis by synthesis process. Since a specific feature in the glottal wave may often be modified or controlled by more than one parameter there is an obvious risk the operator is lead into unjustified conclusions. It appears to be a good approach to match several samples of speech of vastly differing quality, but from one and the same speaker. A primary strategy would then be to hold as many parameters as possible at identical values for all speech samples, especially, of course, those for anatomic measures. The secondary step is to identify a small number of parameters, and perturb those in small steps toward the various individual voice qualities.

Initially most of the anatomical parameters describing the glottal dimensions have to be decided on. If no measurements on the speaker exist one can infer reasonable values for the length l and the driving pressure p from gender, age, and speaking style. These parameters are well behaved in the sense that they in general do not drastically influence the oscillation behavior but rather operate as scale factors. Assuming a fold thickness $2D$ the mass m can be given some reasonable value, perhaps guided by a nominal cord volume $(2D)^2 l$. Now generally a fold tension f can be found, such that the correct pitch is approximated. An empirical matrix formula by Liljencrants (1994) may be helpful to go from desired pitch, level and pressedness specifications into approximately corresponding p , f , and F parameters.

In attempting to control the vibration mode the tuning factor r of the rotational resonance is essential, and here the display of fig. 2 is helpful. In general a lowering of r toward 1 will make the ellipse more eccentric (translation and rotation tend to come in phase) and slope less (rotation decreases and aerodynamic drive is less efficient). Increasing r toward 2 often makes the ellipse more upright and eventually may cause falsetto-like vibration.

Since the average pressure is constantly higher at the subglottal side there is a tendency the cords experience a mean torque toward convergence of glottis, not always offset by the Bernoulli effect. This is obvious in the graph of fig. 2 left where the input trace is mostly below the horizontal axis. In typical 'chest register' oscillation the ellipse crosses this axis, i. e. the shape alternates between convergent and divergent. The divergence rest angle parameter v may be used to provide an angular offset between input and output. This is a critical adjustment, usually

very small, if too large the cords are easily trapped in a constantly divergent or convergent position.

The abduction parameter F is useful to adjust the magnitude of the flow at given pressure p and fold tension f . Another consequence of F is to modify the open quotient of the glottal cycle. Conceivably a similar action could be expected from a manipulation of the rest positions y and v which define the spacing between the cord supports. The latter parameters should however rather be used to control the steepness of the differentiated flow at vocal closure, the point of main acoustic excitation, and where most of the high frequency components are generated. With $y=v=0$ there is a distinct transition into zero glottal area at closure, while at greater values there will always be a certain leak at one or both ends of the cords. An additional parallel passage to model a glottal 'chink' is supplied by the parameter C . This has been suggested by Båvegård and Fant (1994) and by Cranen and Schroeter (1995) to have some additional influence at high frequencies, but this is small and has not yet been fully evaluated in the present model.

The damping factors d and e are in general not critical unless oscillation is only marginally maintained. It should be noted that the damping e for the closed case can take rather high values - this then means a major part of the oscillating energy is dissipated in the closed interval, and that each cycle to a large extent becomes independent of the previous ones. However, since e is further modulated by the relative crossover area (the 'EGG' signal) its influence is on the average less than the parameter value might suggest. For further flexibility an extra multiplier E for the rotational damping was added, but has not proven to be of particular interest.

The acoustic loads from the vocal tract and the trachea cause an 'interaction ripple' on the flow and some fine structure, specifically in the 1 kHz range, and can be shaped by manipulating the tracheal parameters A , H , and Q . This appears to be a convenient way to estimate the tracheal dimensions on which data is rather hard to find otherwise. One consequence of the tracheal load is a characteristic dip in the flow spectrum, usually in the 800 Hz range, and it may be noted that this dip generally is located at a higher frequency than the tracheal formant.

The non-linear properties of the aerodynamical as well as the mechanical system are very important, especially with voice qualities that deviate from 'normal'. Such voices are often difficult to synthesize (as well as to produce in a stationary way by a human speaker) since they require rather accurate parameter tuning, and moreover it may be important to follow specific combined parameter changes rather than changes in single parameters. A simple example is when the vibration is weakly excited such that the stationary amplitude is small enough to stay within an essentially linear stiffness range - then even minute parameter changes may cause the oscillation to die out slowly. A more frustrating one is when the nonlinear effects are crucial, specifically when the folds close and bounce. Oscillation may then stop quickly when a parameter value is changed, and it will

not start again if the same parameter is restored. In such cases you may have to restart by doing some 'gesture', for instance in form of a temporary increase in lung pressure or a temporary opening of glottis with the \mathbf{y} parameter.

2.2. Dynamic simulations

Having established suitable anatomical parameters for a particular speaker the next step attempted was to mimic short dynamic utterances, varying only a few of the controlling parameters. The obvious ones to use are the driving pressure \mathbf{p} , mainly relating to sound level (or more precisely: to speaking effort), the fold tension \mathbf{f} essentially proportional to pitch squared, and rest position \mathbf{y} (or \mathbf{v}) plus the abduction \mathbf{F} for on-off control. Though there was little data on what the speakers actually did there was no difficulty in intuitively estimating suitable initial values from a spectrogram including pitch and level tracks. Fair synthetic approximations to the natural speech could be obtained after a few iterations.

The dynamic runs show a difference as compared to the static fine tunings in that it is easier to keep the model oscillating. The 'gestures' in e. g. \mathbf{p} , \mathbf{y} and \mathbf{f} give rise to substantial excitations such that the system is seldom completely at rest.

At this stage the dynamic simulations give an opportunity to evaluate the relative perceptual importance of several parameters which are not used for the primary control. So far this has only been done with informal listening. Changes coming from perturbation of the glottal metrics are mostly easy to predict, like the pitch lowering and retarded voice onset from an increase in \mathbf{m} to simulate a glottal oedema.

Among present observations few may be mentioned. The damping factors \mathbf{d} for open and \mathbf{e} for closed glottis can be to some extent be traded against each other, but this does affect the voice quality and the voice onset. The displacement flow from the fold collisions (controlled by the \mathbf{j} parameter) has a marked influence at higher frequencies. In the flow waveform this manifests as a 'hump' trailing after the main pulse as elucidated by Hertegård (1994). Contrarily the low frequency components from axial pumping (\mathbf{L} parameter) gives a negligible contribution to the spectrum as was already stated by Ishizaka and Flanagan (1977). Detuning the two folds (using \mathbf{s}) up to about 20% gives little or no audible result even though the mechanical oscillation becomes asymmetric, the center of the glottal passage oscillates.

3. CONCLUSIONS

The present physical model of the glottis has shown great flexibility and is able to mimic a vast range of glottal oscillation types, including falsetto and instances of creaky voice. However, because of the large number of controlling parameters that are set manually, there is a severe demand on the operator to use a conservative strategy in the analysis by synthesis process. Since a specific feature in the glottal wave may often be modified or controlled by more than one parameter there is an obvious risk the operator is lead into unjustified conclusions. It appears to be a good approach to match several samples of speech of vastly

differing quality, but from one and the same speaker. A primary strategy would then be to make as many parameters as possible converge into identical values for all speech samples, especially those for anatomic measures. The secondary step is then to identify a small number of parameters, and perturb those in small steps toward the various individual voice qualities.

4. ACKNOWLEDGMENTS

This research has been supported by ESPRIT BR # 6975 Speech Maps, the Swedish Technical Research Council (TFR), and the Royal Institute of Technology in Stockholm (KTH).

5. REFERENCES

1. Båvegård, M., Fant, G. (1995): Interactive voice source modelling. Proc ICPhS 95, Stockholm, Vol 2, 634-637.
2. Cranen B., Schroeter J. (1995): Physiologically motivated modeling of the voice source in articulatory analysis/synthesis. Submitted to Speech Communication.
3. Hertegård, S. (1994): Vocal fold vibrations as studied with flow inverse filtering. Studies in logopedics and phoniatrics No 5, Huddinge University Hospital, Stockholm.
4. Ishizaka, K., Flanagan, J. L. (1972): Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell System Techn. J. 51 (6), 1233-1268.
5. Ishizaka, K., Flanagan, J. L. (1977): Acoustic properties of longitudinal displacement in vocal cord vibration. Bell System Techn. J. 56:6, Jul/Aug, 889-917.
6. Karlsson, I., Liljencrants, J. (1994): Wrestling the two-mass model to conform with real glottal waveforms. Proc. Int. Congr. of Spoken Language Processing, Sept. 18-22, Yokohama, 151-154.
7. Karlsson, I. (1995): Extreme voice quality, models and data. Speech Maps WP1.3, deliverable 27.
8. Liljencrants, J. (1991a): A translating and rotating mass model of the vocal folds. STL-QPSR 1, 1-18.
9. Liljencrants J (1991b): Numerical simulations of glottal flow. Proc. Eurospeech 91, Genova, 255-258.
10. Liljencrants, J. (1994): Control of voice quality in a glottal model. Proc. 8th Vocal Fold Physiology Conference, April 6-8 1994, Kurume, Japan. In Fujimura, O., Hirano, M.: *Vocal fold physiology, voice quality control*. San Diego: Singular Publishing Group, Inc.