Dept. for Speech, Music and Hearing

Quarterly Progress and Status Report

Analysis by synthesis of glottal airflow in a physical model

Liljencrants, J.

journal: TMH-QPSR
volume: 37
number: 2
year: 1996
pages: 139-142

KTH Computer Science and Communication

Analysis by synthesis of glottal airflow in a physical model

Johan Liljencrants

A mechanical glottal model of the vocal flods with two basic degrees of freedom is self oscillating and is driven by a slowly varying lung pressure. A substantial number of static parameters describe its metrics and mechanical properties. The model is executed in an environment of tracheal and vocal tract loads. Contrary to the shape descriptive LF model this physical model purports to mimic details in the glottal oscillation process and the aim in the analysis by synthesis process is to find direct physical, anatomical, and articulatory causes to features observed in inverse filtered waveforms from natural utterances.

The physical glottal model

The glottal model used here is a variation of the classical two-mass model of Ishizaka and Flanagan (1972). It was developed with the particular aim of controlling boundary movements in a simulation of glottal aerodynamics by Liljencrants (1989, 1991), and purports to be anatomically more realistic. A distinctive characteristic is that the two resonators are here organized as a translational system and a superimposed rotational system, as shown in Fig. 1. 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.

The more novel feature of the model is the second resonance mechanism which is taken to be a rotational oscillation around its center of gravity of the same mass element as in the translational system.

The modelling can be subdivided into a number of basic steps, relating to different domains, and which are interconnected by successive transformations. The first domain is the mechanical which describes the movements of simple mass/spring systems under influence of their driving forces. 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. Additionally, any number of the controlling parameters may be modified from outside in small increments at each sample interval.

The mechanical system

The basic mechanical variables computed during simulation are the instantaneous deflections x, y, and rotation φ.

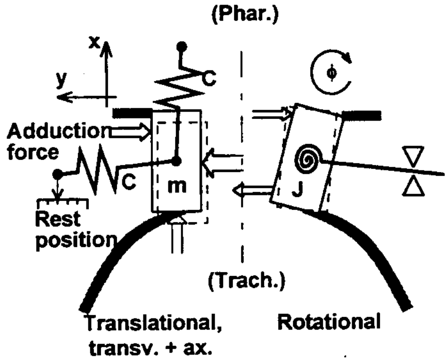

Fig. 1. The basic mechanical elements of the glottal model. Both sides have translational and rotational components, here separated for clarity.

This is done with the elementary equation of motion for the force f acting on a mass M suspended on a stiffness k and a viscous resistance R, which among several alternatives can be discretized in time as

$$f_0 = k\,y_1 + R\frac{y_0 - y_2}{2T} + M\frac{y_0 - 2y_1 + y_2}{T^2} \qquad (1)$$

where subscript 0 denotes the current values. y_1 and y_2 are earlier deflections, 1 and 2 sample interval steps T ago. Solving for y_0 gives a recursion formula to compute it from the present input f_0 and the two past deflections.

A special and important complication in the mechanical system is that the stiffness is nonlinear. The representation used for this is in the same spirit as used by Ishizaka and Flanagan (1972), and is put in terms of a critical deflection y_c, such that at the deflection y the stiffness is $k\{y\} = k_0\,(1 + (y/y_c)^2)$. The nonlinearity of the stiffness is due to the deformation of the vocal folds under influence of the instantaneous deflection y. The critical value y_c also is given two distinctly different values. One is of the order of 1 mm and applies when glottis is open, the other is smaller and applies to the glottis closed case.

The geometrical system

From the instantaneous mechanical spatial deflections the active glottal surfaces are visualised as two ribbons with sinusoidal edges, moving like skipping ropes with a superimposed twisting, as shown in Fig. 2.

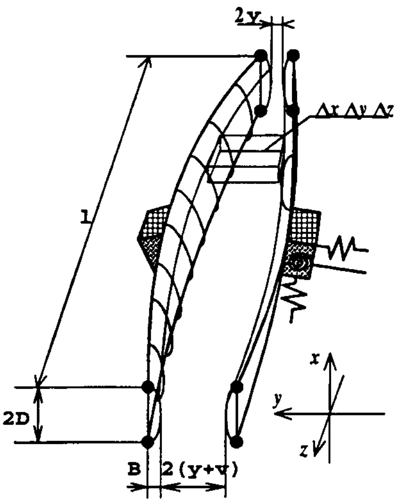

Fig. 2. Geometric development of glottis as the space between two bulging and sinusoidally edged ribbons.

The space between these ribbons defines the volume to be considered for the aerodynamics. In the present version of the model, the ribbons are also given a static bulge along the middle. The shape of the bulge is elliptical, with a separate parameter to control the amount of bulging. The bulge may make the vocal cords tend to roll rather than clap shut, and this can be used to control features in the differentiated flow curve, often to make it more rounded. This bulge is similar, but not identical to one used by Titze (1989) for modelling the vocal cord contact area.

Apart from its direct influence on air flow, the bulge is also used to find a contact area as fraction of the total fold area, a simulated version of an electroglottogram. This measure is additionally used for gradual interpolation of the mechanical damping factors between those for the open glottis and for the closed. The bulge also helps define a hypothetical crossover volume. This volume is in a physical reality of course transformed into a deformation. In the model it is used to compute a displacement flow. This may be added to the main glottal flow under control of a weighting factor. Furthermore the geometical stage allows for simple 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 weighting factors. However, these normally give rather insignificant contributions to the finally interesting flow derivative, as was already stated by Flanagan (1977).

The aerodynamical system

The aerodynamic component of the model is based on general knowledge of the pressure profile from classical work like that compiled by Ishizaka and Flanagan (1972), supplemented with experiences from detailed treatment with the Navier-Stokes equations, Liljencrants (1991b).

The equation that describes the pressure P is for simplicity taken to be one-dimensional, as a function of the glottal flow U and the axial x dimension. We assume that the pressure variations are small enough that we can regard the medium to be incompressible, and thus the flow to be the same for all x. Starting from a subglottal pressure P_s just before glottis (where the particle speed U/A is negligible) we can express the pressure as

$$P\{x\} = P_s - B\{x\} \cdot U^2 - R\{x\} \cdot U - \mathrm{d}(L\{x\} \cdot U)/\mathrm{d}t \qquad (2)$$

The U^2 term represents the Bernoulli law where the 'Bernoulliance' B can be written as

$$B\{x\} = 0.5 \cdot \rho/(A\{x\})^2, \qquad (3)$$

where ρ is the density of air and A is the glottal area as a function of x. When the glottis is open this term dominates and singly settles to a basic shape of the pressure profile $P\{x\}$ according to the area spatial variation. In (2) this is supplemented with a U term to account for viscous pressure drop in a resistance *R* and a final one for drop in the inertance *L*. It could be noted that the final term by virtue of partial derivation should be viewed as a true inertive term $L \cdot dU/dt$ and an additional resistive one $U \cdot dL/dt$. The resistive and inertive terms dominate when the glottal area is very small (a glottal diameter typically <0.1 mm) and/or when the change rate is large. This happens at the glottal opening and closure instants.

A special complication arises from the fact that the Bernoulli relation is valid only for 'channel flow', that is, when the area is filled up with the flow. At a point slightly downstream the point of minimum area A_m the flow will however separate from the channel walls and form a jet of area A_s, fig. 3.

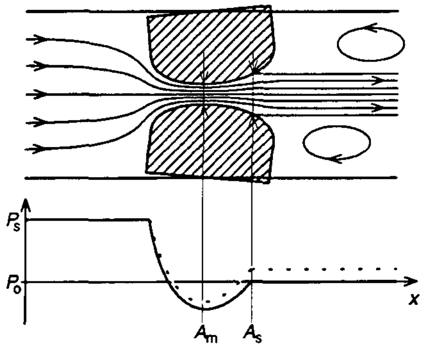

Fig. 3. Top: stylized view of the airflow in the glottal passage with flow separation and jet formation at the area As. Bottom: corresponding pressure profile built up from the Bernoulli relation (dotted) and corrected for resistive and inductive pressure drops.

By virtue of (2) and (3) the pressure has its minimum near the constriction of area A_m and then rises again with the area squared until flow separates. In classical literature like the empirical work of van den Berg et al (1957) this rise is quantified with a normalized 'recovery coefficient', typically with a value in the range 0.5 .. 0.7, and which in our notation would equal $(A_s^2-A_m^2)/A_m^2$. In the present work we have adopted this in the form to assume A_s to be a constant factor greater than A_m, however with the limitation that it is not allowed to exceed the area at the exit of the glottal slit. This scheme compares favourably to aerodynamic simulations by Liljencrants (1991b) and to the 'S-G equation' which approximates empirical measurements on similar glottal shapes by Scherer and Guo (1990). As a competing scheme Pelorson et al (1994) have developed an approximate method of predicting the separation point, based on boundary layer theory. In our present experience that method appears to

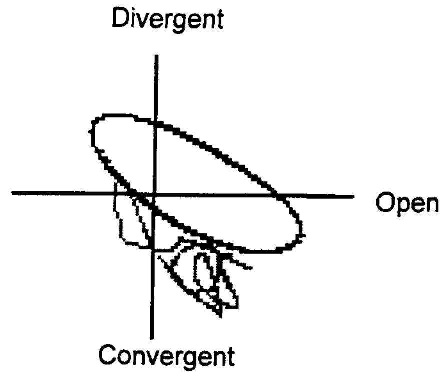

Fig. 4. Traces of divergence angle (up) vs opening displacement (right) for one fold, oscillating in normal chest register. The solid graph represents the motion and the light shaded the scaled excitation driving force.

overestimate the pressure recovery, but further studies of this will be pursued.

The mechanical model acts as second-order lowpass filters (with nonlinear properties) 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. Eq. (1) suggests that the output displacement *y* may be directly compared to the input *f/k*. The inputs are shown in fig. 5 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 detail is due partly to the nonlinear behaviour 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. Fig. 4 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.

Results

Liljencrants (1994) reports in more detail a study of the present model, mimicking a set of utterances from two singers, where pitch, loudness, and pressedness were systematically varied. It was then possible to isolate a few useful coefficients that make it possible to predict what values of parameters like subglottal pressure, vocal fold tension, and adduction are required to produce specified values of pitch, loudness and pressedness.

Another study by Karlsson (1995) focussed more on details in the inverse filtered waveforms from a speaker, again varying his articulation in similar dimensions. This study illustrated the importance of such more esoteric parameters like the bulge of the vocal folds and the frequency of the tracheal first resonance.

Conclusions

The present physical model of the glottis has shown great flexibility and is able to mimic a wide range of glottal oscillation types, including falsetto and 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.

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 require rather accurate parameter tuning, and moreover it may be important to follow specific combined parameter changes rather than changes in single parameters. A sometimes frustrating consequence of the non-linearity is that a specific oscillation mode may change considerably as one controlling parameter is changed a minimal amount, and if the parameter is restored the new oscillation mode will persist.

Acknowledgements

This research has been supported by ESPRIT BR no. 6975 Speech Maps, the Swedish Technical Research Council (TFR), and the Royal Institute of Technology in Stockholm (KTH).

References

van den Berg JW, Zantema JT and Doornenbal P (1957). On the air resistance and the Bernoulli effect of the human larynx. *JASA* 29: 626-631.

Ishizaka K, Flanagan JL (1972). Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Techn J* 51/6: 1233-1268.

Ishizaka K, Flanagan JL (1977). Acoustic properties of longitudinal displacement in vocal cord vibration. *Bell System Techn J* 56:6, Jul/Aug, 889-917.

Karlsson I, Liljencrants J (1994). Wrestling the two-mass model to conform with real glottal waveforms. *Proc Intl Congr of Spoken Language Processing,* Sept 18-22, Yokohama, 151-154.

Karlsson I & Liljencrants J (1996). Diverse voice qualities, models and data. *Ibid.*

Liljencrants J (1989). Numerical simulation of glottal flow. In: Gauffin J & Hammarberg B, eds. *Vocal Fold Physiology.* San Diego: Singular Publishing Group, Inc, 99-104.

Liljencrants J (1991a). A translating and rotating mass model of the vocal folds. *STL-QPSR* 1/1991: 1-18.

Liljencrants J (1991b). Numerical simulations of glottal flow. *Proc Eurospeech 91*, Genova, 255-258.

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, eds. *Vocal fold physiology, voice quality control.* San Diego: Singular Publishing Group, Inc.

Pelorson X, Hirschberg A, van Hassel RR, Wijnands APJ & Auregan Y (1994). Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *JASA* 96/6: 3416-3431.

Scherer RC & Guo C-G (1990). Laryngeal modeling: translaryngeal pressure for a model with many glottal shapes. *Proc Intl Conf Spoken Language Processing,* Kobe, 1: 57-60.

Titze IR (1977). On the mechanics of vocal fold vibration. *JASA* 60/6: 1366-1380.

Titze IR (1989). Four-parameter model of the glottis and vocal fold contact area. *Speech Communication* 8/3: 191-201.