A translating and rotating mass model of the vocal folds

Liljencrants, J.

journal: STL-QPSR
volume: 32
number: 1
year: 1991
pages: 001-018

http://www.speech.kth.se/qpsr
A TRANSLATING AND ROTATING MASS MODEL OF THE VOCAL FOLDS

Johan Liljencrants

Abstract

The motion of a vocal fold is modelled with two mechanical resonators. One is for the translational movement of the body of the fold and is driven by the spatial average pressure in the glottal constriction. The other is rotational, it models the covering mucosa and is driven by the axial pressure gradient.

Unlike the classical Ishizaka-Flanagan two-mass model, these resonators are mechanically uncoupled. Since the excitations of the two resonators are interrelated they still oscillate at the same frequency, but for oscillation, the rotational resonance must be higher than the translational. Also a pair of moderately detuned folds generate a well defined pitch rather than a composite diplophonic signal.

Apart from fixed anatomical configuration parameters and lung pressure, the model is controlled with two parameters only, for cord tension and adduction. Regimes of oscillation are shown for relevant pairs of parameters.

The aerodynamic drive power is predominantly supplied by transglottal pressure in the opening phase while the Bernoulli force at closure gives a rather small contribution.

INTRODUCTION

The first self-oscillating model of the vocal folds was reported by Flanagan & Landgraf (1968), and by Flanagan (1969). In this, a fold is represented by a mechanical resonator built from a mass suspended on a non-linear stiffness. One surface of the mass element bounds the glottal passage, and the element moves laterally to operate as a shutter valve controlling the area open to glottal airflow. The resonator is driven by the pressure exerted on this surface as consequence of the air stream fed into the passage by the lung pressure. To overcome deficiencies in oscillation at certain vocal tract loads and to improve anatomical likeness, this one-mass model was later developed into the now classical two-mass model of Ishizaka & Flanagan (1972) (the IF model). In this, a fold is represented by two resonators in tandem downstream the glottal passage, sharing an elastic mutual coupling element between them.

This layout suggests to regard the folds as transmission lines where waves propagate. Such waves do exist in the mucosal cover of the human folds, they are known to be important to the oscillation mechanism and have inspired much work in their own right. However, the waves carry only a minor part of the oscillatory energy in the folds and a corroborating observation is that the two-mass model works best when the second mass is much smaller than the first.

The present model was developed with the particular aim to control boundary movements in a simulation of glottal aerodynamics (Liljencrants, 1989; 1991), but it also purports to be anatomically slightly more realistic than the IF model of which it is a variation. The proposition is that the two resonators are organized as a translational system and a superimposed rotational system.

The primary model element is thus a mechanical resonator comprising a compliance C=1/k, a mass M, and a damping resistance R. An essential articulatory pitch control comes via the compliance, predominantly defined by the vocal fold tension. The translation may be lateral like in the primitive one-mass model. We can also include one more degree of freedom to allow for longitudinal movement. The folds would then follow essentially elliptic paths, lateral and axial relative to the airflow. Such extension does not require more control parameters in the model, nor does it complicate its implementation significantly.
The novel proposed feature is the second resonance mechanism. This is taken to be a rotational oscillation of the same mass $M$ around its center of gravity. Physiologically there may be some little rotation in the body of the vocal fold around its axis, but the explicit purpose of the rotor is here rather to model the wave motion in the fold cover in a maximally simple way. We can visualize the model mass as a translating and rotating rectangle, Fig. 1, while the physiological counterpart perhaps rather is a translating mass with a surface layer sloshing axially back and forth to modulate convergence or divergence of the glottal air passage.

The model has no mechanical signal connection between the translational and rotational movements, the most significant difference of philosophy from the IF model. Instead 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. But these two driving signals, the force and the torque, are both derived from the transglottal pressure profile and are highly interdependent. The resonators are thus indirectly coupled by the aerodynamics.

**CONTROL PARAMETERS**

Table I lists the parameters used to specify the model and the control of it, typical values used for them, and key relations to alternative descriptors.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>0.2 g</td>
<td>Mass</td>
</tr>
<tr>
<td>$i$</td>
<td>1.5 mm</td>
<td>Gyration radius, Mom of inertia $J = M \cdot i^2$</td>
</tr>
<tr>
<td>$s$</td>
<td>1.0</td>
<td>Lateral tuning skew, Left/Right res. freq. ratio</td>
</tr>
<tr>
<td>$r$</td>
<td>1.4</td>
<td>Rot./Transl. resonance frequency ratio</td>
</tr>
<tr>
<td>$k$</td>
<td>1.0 mm</td>
<td>Critical transl. for open fold stiffness</td>
</tr>
<tr>
<td>$K$</td>
<td>0.5 mm</td>
<td>Critical transl for closed fold stiffness</td>
</tr>
<tr>
<td>$d$</td>
<td>0.1</td>
<td>Open damping factor $I/Q$</td>
</tr>
<tr>
<td>$e$</td>
<td>0.5</td>
<td>Closed damping factor</td>
</tr>
<tr>
<td>$f$</td>
<td>140 Hz</td>
<td>Base frequency</td>
</tr>
<tr>
<td>$y$</td>
<td>0.05 mm</td>
<td>Rest transl gap, one-sided</td>
</tr>
<tr>
<td>$v$</td>
<td>0 mm</td>
<td>Rest rot gap, one-sided</td>
</tr>
<tr>
<td>$P$</td>
<td>1000 Pa</td>
<td>Lung pressure</td>
</tr>
<tr>
<td>$a$</td>
<td>4 cm$^2$</td>
<td>VT area, sets load impedance level</td>
</tr>
<tr>
<td>$h$</td>
<td>500 Hz</td>
<td>$F_i$ formant frequency</td>
</tr>
<tr>
<td>$q$</td>
<td>0.1</td>
<td>$B_i/F_i$ formant damping factor</td>
</tr>
<tr>
<td>$A$</td>
<td>1 m$^2$</td>
<td>Tracheal area, sets source impedance level</td>
</tr>
<tr>
<td>$H$</td>
<td>500 Hz</td>
<td>Tracheal formant frequency</td>
</tr>
<tr>
<td>$Q$</td>
<td>0.1</td>
<td>Tracheal formant damping factor</td>
</tr>
</tbody>
</table>

*Table I. Control parameters and standard values.*
To a first group we may refer parameters relating to the fold anatomy. Here the mass $M$ follows from the fold dimensions while the gyration radius $i$ moderates the rotational moment of inertia $J$ in relation to $M$, the magnitude balance between translation and rotation. This radius is a rather abstract parameter that, however, in principle can be related to the detailed geometry. This was not attempted, instead an arbitrary value was selected such that the magnitude of the rotation came out reasonably large. The particular value also proved to be very uncritical within the range of 1-5 mm. In conjunction the ratio $r$ of rotational to translational resonance frequencies defines the torsional stiffness assigned to the rotation. To examine the influence of left-right asymmetry, a skew factor $s$ is also included to define a resonance frequency ratio between the two sides. This group concludes with critical deflections to describe the stiffness non-linearity for the cases of open and closed folds, and two corresponding damping factors. These latter four are difficult to assess by physiological measurement, and it would be of interest to know how critical their values are to the model operation.

A second parameter group relates to glottal articulation. Here the base frequency $f$ is by arbitrary choice set to be the geometric mean frequency to all four resonators (left and right, transl. and rot.). Together with $M$ and the tuning factors this relates to the stiffness imposed by fold tension, possibly biased by the elastic properties of the folds. The rest gap $y$ reflects adduction gestures. The lung pressure $P$ could also be included in this group rather than in the next.

The 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. Higher formants could also have been included here for completeness, but were omitted since these generally have little influence on the vocal fold function. Additional similar parameters are specified for trachea to introduce a source impedance for the lung pressure.

THE AERODYNAMIC DRIVING FORCES

To exercise the fold model we must assume some driving input from the aerodynamics. At this stage we are more interested in the qualitative behaviour of the model than in details of the driving forces, such as would come from detailed aerodynamic simulation. Therefore a much simplified formulation was used, based on general knowledge of the pressure profile. The primary simplification is that for the open channel we neglect viscosity, on the other hand, for a very narrow or closed channel we neglect the kinetics.

**Pressure profile**

The conventional formulation for the total kinetic transglottal pressure drop stems from van den Berg, Zantema, & Doornenbal (1957). For the constriction having area $A_x$ it is of the form

$$P_y = (c_e + l - c_e) \frac{0.5}{p} \rho (U/A_y)^2$$

employing the corrective "entry coefficient" $c_e$, "recovery coefficient" $c_r$, the density $\rho$, and the externally measurable flow $U$.

To obtain a pressure profile relatable to this we directly use the Bernoulli law of kinetic pressure, starting from a subglottal pressure $P_s$

$$P(y) = P_s - 0.5 \rho v(y)^2$$

where the velocity $v$ now by contrast is the maximum velocity over a cross section of the channel at the axial coordinate $y$. Further we assume the flow $U$ and passage depth $L$ (equal the length of the fold) to be constant, and the cross diameter $d(y)$ to vary with $y$. With a side glance at (1) we put the velocity as

$$v(y) = U(1+c_e)^{1/2}/Ld(y)$$

such as to give the proper pressure drop from input to the narrowest point. The entry coefficient accounts for non-uniformity of velocity in the cross-section.
At a certain point \( y_j \) downstream the narrowest passage the flow separates from the wall and forms a jet of effective diameter \( d_j \), and from there and further downstream the pressure is zero. From this it is now simple to eliminate \( U \) to find the pressure profile

\[
P[y] = P_g \left( 1 - \left( d_j / d_j(y) \right)^2 \right), \quad y < y_j. \tag{4}\]

The assumption made now is that this \( d_j \) equals either 1.1 times the diameter at the trailing edge, or 1.3 times the minimum diameter, whichever is the smaller. The latter will apply in case of a diverging passage. This formulation is now in harmony with (1). The specific magnitude assumption on \( d_j \) with a diverging passage corresponds to a recovery coefficient \( c_\text{r} \) to be maximally \( 1.32^2 - 1 = 0.69 \), going down to 0.21 with a converging passage. The recovery mediates that pressure at the narrowest point may be negative.

When the passage is closed the pressure profile is assumed to have a linear fall from \( P_g \) to 0 over the axial length of the closure. The same could be used when the passage is so narrow and straight that viscosity comes to importance, should one care to distinguish this phase despite it gives rather little contribution to the glottal drive because of its short duration. Since such refinements were omitted here the driving forces as functions of time show minor discontinuities rather than smooth transitions when the model folds open and close.

Several stylized pressure profiles made to this scheme are shown in Fig. 2.

**Glottal flow**

From \( d_j \) and \( P_g \) (which initially equals the lung pressure \( P \)), a corresponding flow \( U \) can be found with (3) and (4). The former involves the entry coefficient which was taken to be 0.3. The particular value chosen has no influence on the shape of the profile and thus no direct implications for the present study of the model kinetics, but it does come in as being a scale factor on flow when the acoustic load is accounted for.

The model was also tried with a slightly modified constriction area, different from \( L_d \). This assumed a deflected shape to be triangular lengthwise the cords which will then introduce a more gradual transition between open and closed glottis, a feature known to be particularly relevant to the magnitude of high frequency components in the generated acoustic signal. Again, this will affect the flow directly, but the mechanical movements only marginally.

**Driving forces**

From the pressure profile the driving force for the translational resonator is found by integrating the pressure over the inner surface of the fold. Similarly the driving torque for the rotation comes from integrating the pressure times the axial distance from the fold center.

With an early even more simplified formula for the pressure profile (neglecting pressure recovery at convergent and uniform shape) it was necessary to extend this integration somewhat upstream the glottal passage (10\% of the glottal axial length was used) to ensure a non-zero driving force with a typical start geometry of parallel fold surfaces. This extended integration is justifiable also from the anatomical configuration and was therefore kept later.

Notably the force and torque become the same on both sides of the passage, but the resulting deflections will come out asymmetric, should the two folds differ in tuning.

**RESONATOR IMPLEMENTATIONS**

The driving force \( f \) and the resulting fold deflection \( x \) of the translational resonator are connected by the differential equation

\[
f = I / C \cdot \frac{dx}{dt} + M \frac{d^2x}{dt^2} \tag{5}\]

which among several alternatives can be discretized in time as
where subscript 0 denotes the current values. $x_1$ and $x_2$ are earlier deflections, 1 and 2 time steps $T$ ago. Solving for $x_0$ gives a recursion formula to compute it from the present input $f_0$ and the two past deflections

$$
x_0 = \frac{2CT^2 f_0 - (2T^2 - 4MC) x_1 - (2MC - RCT) x_2}{2MC + RCT}.
$$

Fig 1. The basic components in the translational and rotational movements.

Fig 2. Typical simplified pressure profiles from sequential times within one glottal cycle.

Fig 3. Display from the test program. Left is from top a record of glottal diameter, stylized vocal folds, and pressure profile. Right is a magnified trace of the trailing edge movement and a histogram of pitch.
This is a direct formulation like the one used by Ishizaka-Flanagan. It is obvious how to introduce nonlinearity by modifying the coefficients dynamically, according to prescriptions on how the elements $R$ and $C$ may depend on $x$.

An alternative way to proceed is to use the complex frequency $s$ domain formulation of (5)

$$f = \left( \frac{1}{C} + R \frac{s}{s + m s^2} \right) x \quad (8)$$

and, in terms of resonance angular frequency $\omega = (MC)^{-1/2}$ and damping factor $d = R/\omega M$, write it as a mechanical complex compliance (admittance divided by $s$)

$$f = C \frac{\omega^2}{s^2 + d \omega s + \omega^2} \quad (9)$$

The dimensionless final factor for frequency dependence is a second order low-pass function. It can be $z$-transformed with the conventional textbook "impulse invariant method" to finally result in the recursive computation formula

$$x_0 = (1 + a_1 + a_2)Cf_0 - a_1x_1 - a_2x_2 \quad (10)$$

where a "gain factor" has been applied to the input $Cf_0$ to set the zero frequency scaling correct, and with the coefficients

$$a_1 = -2 e^{d \omega T} \cos(\omega T) \quad (11)$$

$$a_2 = e^{d \omega T}. \quad (12)$$

The expressions (7) and (10) are largely equivalent and both of them were operated successfully, the selection between them is much a matter of taste. It should, however, be remembered that a historical rationale behind the $z$-transform method is that the direct form (7) suffers from inferior accuracy unless the sampling frequency $1/T$ is much higher than the resonance frequency.

The $z$-transform variant (10) was used for a major part of this study because it allowed for a relatively low sampling frequency to speed computations, typically 5 kHz was used.

Also to improve speed attempts were made to switch between three pre-computed sets of $a$ coefficients, depending on the actual deflection $x$. These sets corresponded to the cases of closed glottis, open glottis with small deflection, and open glottis with deflection larger than a given threshold. This implies a fairly coarse stepwise approximation to the nonlinear stiffness, a piecewise linear force vs. deflection characteristic. Compared to the use of a more elaborate gradual stiffness characteristic, with recomputation of coefficients at every sample interval, it did not appear to influence the general behaviour of the model significantly in quality.

An important exception was, however, the jitter and shimmer characteristics. With a piecewise linear stiffness these measures came out larger, more irregular, and also undesirably much dependent on sampling frequency. For that reason all measurements shown here apply only to continuously varied stiffness. The formula used is in the same spirit as used by IF, but put in terms of a critical deflection $x_c$, such that at the deflection $x$ the stiffness is

$$k(x) = k_0 (1 + (x/x_c)^2) \quad (13)$$

$$k_0 = 1/C_0 = \omega^2 M = (2\pi f)^2 M. \quad (14)$$

When changing the $a$ coefficients care must of course also be taken to scale the input $Cf$ properly.

**Acoustic loads**

To accommodate for source and load impedances of the trachea and the vocal tract their corresponding pressure drops were computed from past and present values of $U$ using the same direct or $z$ transform formulas as those outlined for the mechanical resonators. The only difference is that these now represent acoustic impedances (instead of mechanical compliances), with levels
corresponding to the areas, and with resonances modelling the tracheal and vocal formant frequencies and bandwidths.

From the pressure drops in the acoustic loads it is elementary to bias and scale the transglottal pressure profile for the next time sample such that the total pressure drop equals the lung pressure. Assuming zero mouth pressure the vocal tract load will bias the supraglottal pressure to vary on both sides of zero which eventually will influence the glottal drive force.

![Figure 4](image-url)

*Fig 4.* Time waveforms in the model operated with standard parameter values as in Tab I, but driven by a varied pressure as shown in the top trace.
DESCRIPTION OF THE FOLD OSCILLATION

Fig. 3 represents a computer screen plot used in the study, including an instantaneous pressure profile and a record of the passage diameter. The ellipsoid patterns at upper right are enlarged tracings of the trailing edges of the two folds which were detuned by a factor of 0.9. Here also axial translation was enabled, using the same dynamic elements as for the lateral movement, but with a separate driving force also derived from the pressure profile. Such axial movement did not influence the model performance, likely because no provisions were made to let it feed back to the pressure and flow. The two trajectories have different size because of the detuning. They also show slight irregular deviations from their mean path.

Fig. 4 shows a typical recording of data from the model as functions of time. The only varying input parameter is the driving lung pressure shaped by two cosinusoidal half-periods into a smooth pulse. All other parameters have the standard values given in Table I.

In the initial phase of a period when the folds are pressed apart the axially decreasing pressure also induces a rotation such as to make the passage convergent. The mass acquires kinetic energy which, however, gradually is transformed into potential energy in the stiffness, and eventually the deflection is so large that the resonators swing back. Reaching this point it is essential that the pressure profile has changed such that the outward force is reduced, the passage must go toward a diverging shape. Otherwise the force will counteract the oscillation which will then die out, and the passage will stay open and convergent.

Here it is critical that the rotational resonance frequency is higher than the translational, such that the rotation in principle tends to anticipate the deflection. The movements of the fold leading and trailing edges are seen in the second from top set of traces of Fig 4 together with their difference that represents the rotation.

The following graphs in this figure show the driving force and torque, then the same multiplied to the velocities of their respective elements. This becomes the instantaneous power supplied by the air to entrain oscillation. For sustained oscillation the average power supplied must of course be positive even if it instantaneously may be negative. An interesting feature in the drive power is that a major part of it is supplied during the opening phase while the folds are pressed apart by the lung pressure. In the closing phase where oscillation is supported by the "Bernoulli underpressure" the power contribution is relatively small by comparison.

PARAMETER VALUES AND PERFORMANCE

To evaluate the behaviour of the model it was installed in a test program which could scan selected pairs of parameters. For each set of values the model was operated from a rest position until it had completed a number (20) of cycles or a maximum time (1024 samples), and then measurements were done on the model output. When the scan is complete these measurements can be summarized by iso contours in the plane of the two parameters, such as in Fig. 5 and following. The measurements include amplitude, pitch, jitter, and shimmer.

Amplitude

The amplitudes of the cycles (except the first one) were averaged and displayed as a dB measure. The reference level was arbitrarily taken at the deflection that would result if the glottal surfaces were statically pushed apart by the driving lung pressure. This amplitude measure suggests the model’s willingness to oscillate, when positive it indicates reliable oscillation and higher values reflect increasing efficiency in the drive-resonator mechanisms. Negative values mostly correspond to waveforms with decaying amplitude, often combined with multi-frequency interferences. This amplitude measure is easy to assess and gave more reliable and less noisy results than initial attempts to determine amplitude growth rate in the first few cycles.

Pitch

Pitch periods were individually located by simple detection of negative peaks. This gave raw results in form of integer N multiples of the sampling interval T. To improve resolution consecutive periods were then pairwise compared for an optimal time alignment. That was done with a heuristic
algorithm, invented by accident and presented in the appendix.

The average pitch from the sequence of measurement samples is shown in the isograms in terms of percentage above the base frequency parameter $f$.

**Jitter and shimmer**

Perturbations in frequency and amplitude are an important feature to make speech sound natural. Such perturbations are bound to develop in a non-linear multi-resonator system like the present and it would be interesting to know if any specific parameters are of special significance to the "chaoticity".

Perturbations were measured as follows from the same pairs of periods as used for the pitch measurement. First determine a sequence of relative interperiod changes

$$\Delta_j = 2(y_{j+1} - y_j)/(y_{j+1} + y_j), \quad j = 2 \ldots 19$$

where $y_j$ for the jitter represents the pitch (arbitrarily frequency or period time) found from the $j$th period pair. For the shimmer $y_j$ is instead the amplitude of the $j$th period.

This formula differs importantly from common practice in that the sign of the perturbation is kept. Also the change is normalized to the mean of the two compared samples rather than to only one of them as is sometimes done.

To find a conclusive jitter or shimmer measure first the mean value of these $\Delta_j$ is subtracted and then the RMS value of the remainders is taken. This refinement is essential to counteract artificially high values that will otherwise result should there be a superposed monotonous frequency or amplitude change.

It is important that the underlying measurements of frequency and amplitude are uncompromised by noise from quantization in time and level - the smallest jitter or shimmer that can be discerned is of the same order of magnitude as the measurement error for an individual $y_j$. For the frequency measurement the algorithm in the appendix was useful. For amplitude it was not very accurate to use the peak sample alone, neither to use the sum of all points within a pulse (the pulse area). It turned out superior to use extrapolation by fitting a parabola to the three samples closest to the peak.

With such refinements the jitter and shimmer measures were tested on ideal sinusoidal waveforms, asynchronous to the sampling rate. The measurement noise background then reached the 0.1% level when sampling frequency was as low as typically 50 times the signal frequency (pitch).

**RESULTS AND DISCUSSION**

Each of Fig. 5 and following show results from 21*21 parameter value pairs, uniformly distributed over the parameter plane. The figures also tabulate all parameter values used. In the table initial characters X and Y denote which parameters are used for the axes in the diagrams. Glottal minimum diameter/area waveforms are shown at lower left and pertain to the middle $x$ coordinate in the isograms

**Tuning**

Fig. 5 shows the effect of the tuning parameters. Horizontally is the ratio $s$ of the left to right fold resonance frequencies where unity would be a normal value with the two folds equal. Vertically is the ratio $r$ of rotational to translational resonance frequencies. The highest several amplitude contours clearly identify a normal operating area. Essential conclusions are that

a. The rotational resonance must be higher than the translational. An $r$ factor between 1.1 and 2 is usable, the optimum is close to the lower limit.

b. Lateral equality is not critical, and tolerance increases with a higher rotational resonance. The more obvious consequence of lateral detuning is that the mid-line of the glottal passage oscillates, but this is probably acoustically insignificant.
c. Within the efficient range jitter and shimmer are very low and not much influenced by the tuning parameters.

Fig 5. Effects of lateral detuning (horizontal) and rotational to translational frequency ratio (vertical).
Fig 6. Effects of lung pressure (horizontal) and fold tuning/tension (vertical).
Fig 7. Effects of lung pressure (horizontal) and abduction (vertical).
Fig 8. Effects of damping factor in the open glottis resonators (horizontal) and in the closed (vertical).
These interesting findings suggest there is a rather wide tuning tolerance for the folds in a "normal" voice, and they contradict popular belief that differently tuned folds should promote irregular or diplophonic voice.

Such "pathological" operation does not occur in the model until the lateral detuning $s$ is so large that oscillation efficiency suffers - then the low frequency fold yields too much so that the other gets impeded acceleration, and their recoils come grossly out of phase.

With extreme detuning efficiency improves slightly again. That is when one of the folds is made so stiff that the other makes most of the work alone and with little interference.

Fig. 5 lower left shows that within the optimal area the pitch is close to 1.2 times the base frequency. This is indeed a reason the tuning factors were defined the particular way shown in Table I; it would seem plausible the resulting frequency is proportional to some mean of the component resonance frequencies.

**Lung pressure vs fold resonance**

Fig. 6 shows horizontally the driving lung pressure and vertically the mean fold resonance frequency. The iso contours for frequency approach the square root shape that delineates a constant ratio of pressure to stiffness. The amplitude curves show very little variation. This may be misleading, one must be aware that the amplitude norm (static deflection under the actual pressure) varies considerably along both axes in the diagram. With low tuning and stiffness the folds are driven far out in the nonlinear region which increases the pitch much over the small signal resonance frequency. Concurrently jitter and shimmer increase markedly.

**Lung pressure vs abduction**

Fig. 7 horizontally shows pressure as before, while the vertical axis shows abduction in terms of the (single sided) rest gap between the folds. Abduction appears to be a major controller for amplitude, and at lower pressures there is need for a rather precise control. A wide gap causes oscillation to start more gradually. Informal listening to this also conveyed the impression of indistinct breathy onset of phonation. A negative rest gap will hold the folds closed at low pressures while high pressures yield a sharp, pressed voice. The pressure at which oscillation starts also strongly depends on the fold stiffness which was held constant for this figure.

The abduction has little influence on pitch and shimmer. When the folds are pressed together the pitch, however, rises somewhat and jitter increases.

**Damping factors**

In Fig. 8 the abscissa is the damping factor of the open folds and the ordinate for the closed folds. The damping should be higher when the folds are closed, so the lower right corner of the figure is outside reasonable range.

The amplitude diagram indicates that the open folds’ must have a damping factor $d=1/Q$ lower than about 0.4 for sustained oscillation, and that the value of damping in the closed phase is not critical. The lower left corner where damping is very low is also not realistic. Here the waveforms have initial transients that cover several cycles, this also shows up as high values in the diagram of jitter and shimmer.

Oscillation remains even with rather high damping. This signifies that the aerodynamic drive is very efficient, only a moderate part of the kinetic energy in the folds need be saved for coming cycles. The folds have a short memory and quickly adapt to changes in pressure and tension.

**Vocal tract interaction**

The abscissa in Fig. 9 shows variation of the magnitude of the vocal tract load in terms of a nominal tract cross section area. The ordinate is a (first) formant frequency and the formant $Q$ is held constant. Naturally interaction is the strongest when pitch is close to the formant frequency which will effect wide variation in transglottal pressure synchronized with the fold motion. It is interesting that the pitch is lowered by this interaction while amplitude is not so much affected.
Fig. 9. Effects of vocal tract area (horizontal) and first formant frequency (vertical).
CONCLUSIONS

A vocal fold model was developed with the distinctive feature that its two basic degrees of freedom are translational and rotational, and that these movements take place in mechanically uncoupled resonators. The resonators are instead indirectly coupled via aerodynamic forces. The rotational movement models the mucosal vocal fold cover and serves to shape the air passage, thus essentially controlling the transglottal pressure profile while expending a comparatively small mechanical power. This pressure then acts to inject a larger amount of power into the translational movement of the folds to maintain their vibration so that they can open and shut the airflow into the vocal tract. This power is mainly received from subglottal pressure when the folds open while the power generated by the Bernoulli underpressure at closure is small by comparison.

Reliable oscillation requires the rotational resonance to have a higher frequency than the translational. Other parameters such as damping factors and relative detuning of two folds show a wide tolerance range for an efficient regime of oscillation. Within this range the model does generate chaotic perturbations, but at a low level, typically up to the order of 1% in amplitude and frequency.

The model can flexibly produce several variations in a normal voice, including such extremes as those associated with fold infections. The model failed to produce creaky and diplophonic voice which suggests that such call for additional degrees of freedom.

ACKNOWLEDGEMENT

This work was made possible by support from the Royal Institute of Technology, Stockholm (KTH), and the Research Council of the Swedish Board for Technical Development (STUF).

REFERENCES


APPENDIX

ALGORITHM TO FIND A PERIOD TIME, INTERPOLATED BETWEEN THE SAMPLING INSTANTS

Let $x_i$ be the samples of a time sequence. Suppose that we by some pitch detection scheme have found that it is nearly periodic with the period $NT$, where $N$ is an integer and $T$ is the sampling interval. We want to find an interpolated refinement of the period time. To that end we compute a timewise misalignment between two sequential periods of duration $NT$.

For each pair of points, $N$ samples apart, where the waveforms $x_i$ and $x_{N+i}$ are similar, determine a mean slope $d_i$ (of dimension $dx/di$) and an error $e_i$ (of dimension $dx$)

$$d_i = 0.5(x_{i+1} - x_i + x_{N+i+1} - x_{N+i}) \quad (a1)$$
$$e_i = x_i - x_{N+i}, \quad i = 0 \ldots N-1 \quad (a2)$$

and compute the expression of sums over $i$:

$$\Delta N = \sum d_i e_i / \sum (d_i)^2. \quad (a3)$$

That is, we sum the error over one period using the slope as a weight. This indicates whether the coarse estimate $N$ is too small or too large. The result is normalized with the power of the weight to render a fractional correction $\Delta N$ (of dimension $di$). We add this correction to $N$ to get a final pitch estimate from

$$T_0 = (N + \Delta N) T = 1/F_0 \quad (a4)$$

If the coarse estimate $N$ was correct then $\Delta N$ should be less than 0.5 in magnitude, but the method also extrapolates outside that range.

The strength of this method lies in that it integrates all data of the two periods. As consequence it is robust to noise and to where in the cycle the time reference is placed. However, much high frequency noise tends to make the denominator in (a3) too large so that the magnitude of $\Delta N$ is underestimated, thus it may be beneficial to smooth $x_i$ somewhat beforehand.

Fig. a1 shows test results with $N = 10, 20, \text{ and } 50$, using samples of an ideal sinusoid as the signal. The abscissa shows how much the true non-integer number of sample intervals in one period deviates from $N$. This period time was swept in small increments. The ordinate shows the remaining error after interpolation, also in units of sample intervals. It is obvious how higher sampling frequency (larger $N$) also gives larger tolerance to the initial coarse estimate $N$.

This algorithm is not suited to detect $N$ by exploration because it gives no guidance to whether the original $N$ pertains to a single or two pitch periods, or to a formant period. Fig. a2 shows the plateau found for different $N$ when the true period was 24.667 samples. The test signals are a pure fundamental, the same plus second harmonic, and plus second and third harmonics, all components with the same amplitude.
Fig. a1. The ordinate shows the remaining error after interpolation in units of sample intervals, as function of true period time minus coarse estimate $N$. The signal is a sinusoid and $N = 10, 20, \text{ and } 50$.

Fig. a2. Plateau of computed pitch periods as function of original estimate $N$. The true period was 24.667 samples in the sinusoidal test signal, also with one and with two harmonics.