On the Acoustic Sensitivity of a Symmetrical Two-Mass Model of the Vocal Folds to the Variation of Control Parameters

Denisse Sciamarella, Christophe d'Alessandro
LIMSI-CNRS, BP 133, F-91403, Orsay, France

Summary

The acoustic properties of a recently proposed two-mass model for vocal-fold oscillations are analysed in terms of a set of acoustic parameters borrowed from phenomenological glottal-flow signal models. The analysed vocal-fold model includes a novel description of flow separation within the glottal channel at a point whose position may vary in time when the channel adopts a divergent configuration. It also assumes a vertically symmetrical glottal structure, a hypothesis that does not hinder reproduction of glottal-flow signals and that reduces the number of control parameters of the dynamical system governing vocal-fold oscillations. Measuring the sensitivity of acoustic parameters to the variation of the model control parameters is essential to describe the actions that the modelled glottis employs to produce voiced sounds of different characteristics. In order to classify these actions, we applied an algorithmic procedure in which the implementation of the vocal-fold model is followed by a numerical measurement of the acoustic parameters describing the generated glottal-flow signal. We use this algorithm to generate a large database with the variation of acoustic parameters in terms of the model control parameters. We present results concerning fundamental frequency, intensity and pulse shape control in terms of subglottal pressure, muscular tension, and the effective mass of the folds participating in vocal-fold vibration. We also produce evidence for the identification of vocal-fold oscillation regimes with the first and second laryngeal mechanisms, which are the most common phonation modes used in voiced-sound production. In terms of the model, the distinction between these mechanisms is closely related to the detection of glottal leakage, i.e. to an incomplete glottal closure during vocal-fold vibration. The algorithm is set to detect glottal leakage when transglottal air flow does not reach zero during the quasi-closed phase. It is also designed to simulate electroglottographic signals with the vocal-fold model. Numerical results are compared with experimental electroglottograms. In particular, a strong correspondence is found between the features of experimental and numerical electroglottograms during the transition between different laryngeal mechanisms.

PACS no. 43.64.-q, 47.85.-g, 43.60.-c, 05.45.-a, 43.70.-h

1. Introduction

One of the main challenges in voice production research has for long been the construction of a deterministic vocal-fold model which could describe, in particular, the mechanisms responsible for different voice qualities. Presently, a qualitative distinction between pressed, modal, breathy, whispery, tense, lax, creaky or flow voice is often made in terms of the acoustic parameters describing one cycle of the glottal flow derivative [1, 2, 3]. Quantitative aspects, such as frequency or intensity, are also readable from this kind of glottal-flow phenomenological model. However, these acoustic parameters do not account for the subtle features linked to the behavior of the source: they just provide us with an empirical description of the signal at the exit of the glottis. On the other hand, modelling and numerical

simulation of the speech production process is a difficult task which implies coping with the complex nonlinearities of a fluid-structure interaction problem where the driving parameters are subject to neural control.

Since 1972, a series of simplified vocal-fold models which are apt for real-time speech synthesis have followed and improved the pioneering Ishizaka and Flanagan's two-mass model [4]. In this kind of lumped models, self-sustained vocal-fold oscillations are mainly due to a varying glottal geometry that creates different intraglottal pressure distributions during the opening and closing phases of the vocal-fold oscillation cycle. The non-uniform deformation of vocal-fold tissue is assured by a mechanical model having at least two degrees of freedom. For this reason, the most simple lumped vocal-fold models are known as *two-mass models*.

It has often been remarked that the main weakness of this approach lies in the absence of a simple relationship between the parameters in the model and the physiology

of the vocal folds [5]. Most of the parameters in the model are initially chosen according to physiological measurements [6], but afterwards they have to be tuned to compensate for over-simplifications of the model. These tunings are performed by trial and error, so that the signals predicted by the model share the features presented by experimental glottal-flow waveforms. But the task is not simple, mainly because the parameters characterizing the signal are greatly outnumbered by the control parameters of the model, and because the intricate correlation between acoustic and control parameters has not been unveiled.

Research is therefore needed not only to build a bridge between physiology and physics but also between physics and the acoustic phenomenological models describing glottal-flow waveforms. Devoting efforts to the second issue is certainly necessary in order to bring together the phenomena of voice production and perception, and eventually to decide whether a production model with a few control parameters related to acoustic parameters is realizable [7]. The existence of such a production model would constitute a first step towards the eventual long-term construction of a certainly more ambitious voice production model capable of relating neural activities to glottal driving parameters (as has been recently done for the syrinx in the case of birds [8]).

In this context, studying the acoustic response of vocal-fold two-mass models is essential to unveil the actions that the modelled source employs to produce different acoustic effects. A systematic study of acoustic and control parameter correlations has been performed in the case of the traditional Ishizaka and Flanagan's (IF) two-mass model [9]. This preliminary study has shown that the smooth variation of control parameters can be associated with a physiological action producing a specific acoustic effect which can be compared to those reported in the literature [1].

The aim of this paper is to perform an acoustic characterization of a two-mass model with an up-to-date aerodynamic description of glottal flow which takes into account the formation of a free jet downstream of a moving separation point in the closing phase of the glottal cycle [10, 11, 12, 13]. The choice of a model with a symmetrical glottal structure as introduced in [14] will be adopted, mainly because it allows a reduction in the number of control parameters which narrows the gap with the low number of acoustic parameters used to describe glottal-flow signals in phenomenological models. The fact that this assumption does not hinder reproduction of glottal pulses is a remarkable property of this kind of approach. Symmetrical two-mass models thus constitute a new testbench for correlation analysis between acoustic and control parameters, as well as a promising scenario for vocal-fold modelling in terms of acoustic parameters. A model of such characteristics was implemented by Niels Lous *et al* [14] in 1998.

The article is organised as follows. The theoretical background concerning the invoked models is given in section 2. This section provides a self-contained description of the Niels Lous model, a quick reference to glottal-flow sig-
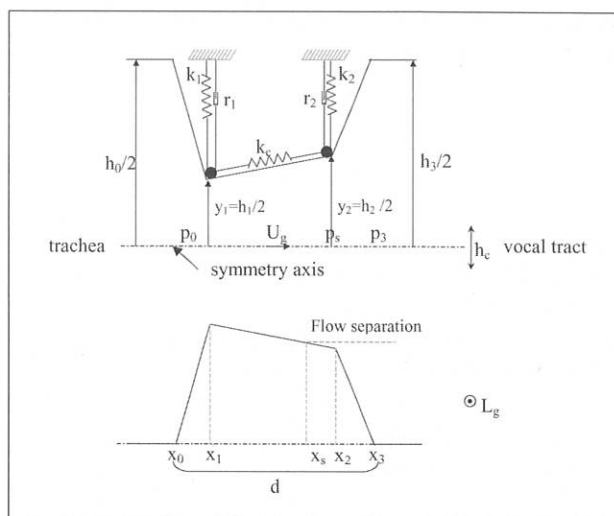

Figure 1. Sketch of the glottal channel geometry in the Niels Lous two-mass model.

nal models in order to introduce the so-called acoustic parameters and a subsection devoted to what we will refer to as control parameters of the model. Section 3 is devoted to the description of the algorithmic procedure designed to generate the data that will be subsequently analysed. The acoustic analysis is developed in section 4. We present results concerning the effects on glottal-flow signals of flow separation and of the acoustic feedback of the vocal tract. The subsection presenting the sensitivity of acoustic parameters to the variation of control parameters has been outlined to show, in terms of the data, how the model controls fundamental frequency, intensity and pulse shape. We also report the observation of oscillation regimes when acoustic measurements are plotted in control parameter space, and provide an interpretation of oscillation regimes in terms of laryngeal mechanisms. Finally, we show that the reported behavior of experimental electroglottographic signals during a transition between mechanisms may be encountered in numerical electroglottographic signals when the mechanical system traverses an underlying bifurcation. General conclusions are drawn in section 5.

2. Background models

2.1. The vocal-fold model

Any lumped vocal-fold model is composed of a description of the vocal-fold geometry, the aerodynamics of the flow through the glottis, the vocal-fold mechanics and the coupling to vocal-tract, trachea and lung acoustics.

The two-mass model proposed by Niels Lous *et al* [14] assumes that the vocal-fold geometry is described by a couple of three mass-less plates as shown in Figure 1. The model considers a two-dimensional structure with the third dimension taken into account by assuming vocal folds have a length L_g (compare to [15]). As usual, symmetry is assumed with respect to the flow channel axis. The flow

channel height $h(x,t)$ is a piecewise linear function of x (see Figure 1) determined by $h_{1,0}, h_{2,1}, h_{3,2}$:

$$h_{q,q-1}(x,t) = \frac{h_q(t) - h_{q-1}(t)}{x_q - x_{q-1}}(x - x_{q-1}) + h_{q-1}(t), \quad (1)$$

where $q = 1, 2, 3$ and h_0 and h_3 are constant.

Vocal-fold mechanical behavior during the production of voiced sounds depends on lumped inertia m_i, elasticity k_i, viscous loss ζ_i and damping $r_i = 2\zeta_i\sqrt{k_i m_i}$. The position of each of the two-point masses ($y_i, i = 1, 2$) is animated with a motion which is perpendicular to the flow channel axis. The coupling between the masses is assured by an additional spring k_c. Unlike in [4], non-linearities in the springs characteristics are absent in this model: the non-linear behavior of the system is assured by vocal-fold collision. Glottal closure is associated with a stepwise increase in spring stiffness k_i and viscous loss ζ_i that will represent the stickiness of the soft, moist contacting surfaces as they form together, just as in the traditional IF model [4]. The equations of motion for each of the masses of this vocal-fold model read:

$$m_i \frac{d^2 y_i}{dt} + r_i \frac{dy_i}{dt} + k_i y_i + k_c(y_j - y_i) = f_i(P_s, L_g, d, \rho_0, \mu_0), \quad (2)$$

where $i, j = 1, 2$ ($j \neq i$) and f_i is the y-component of the aerodynamic force acting on point i. The force depends on subglottal pressure P_s, vocal-fold dimensions (L_g, d), air density $\rho_0 = 1.2 \, kg/m^3$ and air viscosity $\mu_0 = 1.8610^{-5} \, kg/m\,s$.

The aerodynamics of the flow within the glottis plays a fundamental role in a voice production model. An analysis based on the evaluation of dimensionless numbers [16] shows that the main flow through the glottis can be approximated by a quasi-stationary, inviscid, locally incompressible and quasi-parallel flow from the trachea up to a point x_s where the flow separates from the wall to form a free jet. The pressure before x_s can hence be calculated from Bernoulli's equation:

$$p(x,t) + \frac{\rho_0}{2}\left(\frac{U_g(t)}{h(x,t)L_g}\right)^2 = p_0(t) + \frac{\rho_0}{2}\left(\frac{U_g(t)}{h_0 L_g}\right)^2, \quad (3)$$

with $U_g(t)$ the volume flux through the glottis. These approximations do not hold for the boundary layer that separates the main flow from the walls, in which viscosity is relevant and the flow is no longer quasi-parallel. Although very thin, the boundary layer is important since it explains the phenomenon of flow separation.

Experimental work by Pelorson *et al* [10] shows that the occurrence of flow separation within the glottal channel, combined with no pressure recovery for the flow past the glottis, is not a second order effect. In fact, at high Reynolds number, the volume flux control by the movement of the vocal folds is due to the formation of the free jet downstream of the glottis as a result of flow separation

in the diverging part of the glottis. As the jet width is small compared with the diameter of the pharynx, most of the kinetic energy will be dissipated before the flow reattaches. Flow separation is shown to occur not at a fixed position but at a location which depends on the flow characteristics as well as on glottal geometry.

For simplicity, the boundary-layer theory necessary to explain and predict this behavior is substituted in the model with a geometrical separation criterion that will determine the position x_s of the separation point during the closing phase. This criterion has been recently proposed by Liljencrants (see [14, 16]). It is based on the hypothesis that flow separation is mainly sensitive to the channel geometry so that when $h_2(t) > sh_1(t) > 0$, $x_s(t)$ may be determined from the condition $h_s(t)/h_1(t) = s$, where s is referred to as the separation constant. Otherwise, i.e. when the separation criterion is inactive, the flow separates at x_2 ($x_s = x_2$) for an open glottis. When the glottis is closed x_s is assumed to be zero.

Regarding the aerodynamic force driving vocal-fold oscillations, Pelorson *et al* [10] assume that there are no forces acting on the masses next to the larynx side of the vocal folds. The traditional IF two-mass model does not make this assumption but considers the latter masses to be smaller than those modelling the pharynx side. Niels Lous *et al* [14] have shown that neither of these asymmetries are necessary to produce reasonable glottal waveforms. This simplification is new to the world of vocal-fold lumped models, and has coined the notion of a *symmetrical* two-mass model.

It is clear that the aerodynamical portrait of transglottal flow breaks down near vocal-fold collision: the apertures involved are too small to justify a quasi-stationary, high-Reynolds-number approximation. In such a case, a viscous flow model should be considered. However, a numerical resolution of the full equations holding near glottal closure is computationally too expensive for real-time speech synthesis. This point is quite delicate since it is particularly near glottal closure that high frequency energy is produced, to which the ear is very sensitive. Vocal-fold collision is accounted for in the rough manner described within the mechanical model. As observed in [14], a systematic study of vocal-fold collision by means of finite-element simulation could be useful to improve glottal-flow modelling.

The representation of the vocal tract in this symmetrical vocal-fold model does not differ from the one used in the traditional IF two-mass model: the glottis is coupled to a transmission line of cylindrical, hard-walled sections of fixed length. In each section, one-dimensional acoustic pressure wave propagation is assumed. In this model, trachea and lungs are similarly modelled as a transmission line. The trachea is described as a straight tube of constant cross-sectional area and length, and lungs are modelled as an exponential horn. Coupling with the incompressible quasi-stationary frictionless flow description within the glottis is obtained by assuming continuity of flow and pressure.
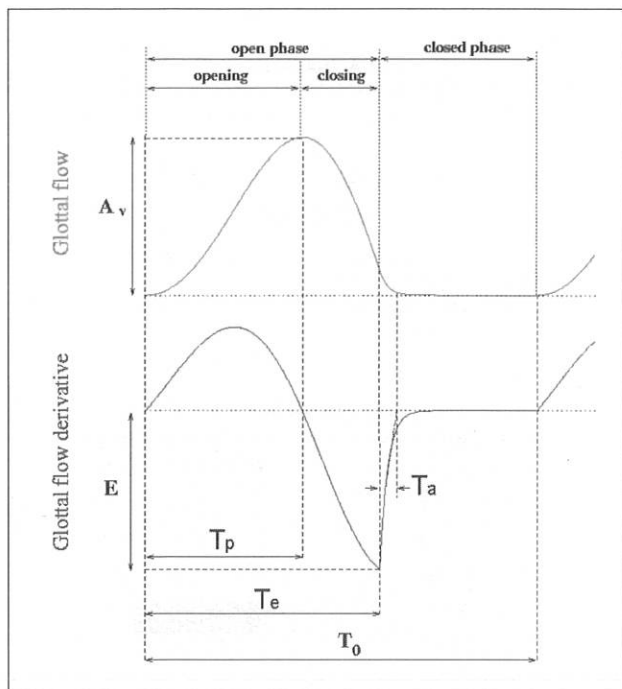
Figure 2. Definition of parameters describing the glottal-flow pulse (above) and its derivative (below). The fundamental period, T_0, is a global parameter, which controls the speech melody; T_e is the duration of the open phase; T_p is the duration of the opening phase; T_a the effective duration of the return phase.

2.2. Glottal-flow signal models

Glottal-flow signal models, which provide a description of glottal-flow waveforms in terms of the definition of a few acoustic parameters, have proved to be particularly useful for vocal intensity and timbre description. A wide variety of signal models is available in the literature [17, 18, 19, 20, 21], differing in the number and choice of acoustic parameters. Doval and D'Alessandro [2] have shown, however, that these models may all be described in terms of a unique set of acoustic parameters, closely linked to the physiological aspect of the vocal-folds vibratory motion. The glottal flow signal is assumed to be a periodic positive-definite function, continuous and derivable except maybe at the opening and closure instants.

In order to define a suitable set of acoustic parameters, let T_0 be the fundamental period of the signal and $F_0 = 1/T_0$ the fundamental frequency. Consider the glottal pulse shape depicted in Figure 2.

In order to describe the glottal-flow pulse and its derivative in time we introduce the following parameters:

- the *open quotient* $O_q = T_e/T_0$, where T_e is the duration of the open phase,
- the *speed quotient* $S_q = T_p/(T_e - T_p)$ (which conveys the degree of asymmetry of the pulse), T_p being the duration of the opening phase and
- the effective *duration of the return phase* T_a (which measures the abruptness of the glottal closure).

Description of the pulse height requires an additional parameter: the amplitude of voicing A_v (the distance be-

tween the minimum and maximum value of the glottal volume velocity) or alternatively,

- the *speed of closure E* which corresponds to the glottal volume velocity at the moment of closure, whose main perceptual correlate is intensity.

2.3. Control parameters

Consider equations (2) and (3): our dynamical variables are y_1, y_2 and U_g; f_1, f_2 and h are prescribed functions, and the remaining quantities are the model parameters. As mentioned in (2.1), we follow [14] in the assumption that the glottis has a symmetrical structure, i.e. $m_i = m$, $k_i = k$, $r_i = r$. The stepwise variation of elasticity and damping on collision is also symmetrical: when $h(x_i) < 0$, k is increased to $c_k k$ and ζ to $\zeta + c_\zeta$

Typical values for these parameters are: $d \approx 0.2\,\text{cm}$, $m \approx 0.1\,\text{g}$, $k \approx 40\,\text{N/m}$, $k_c \approx 25\,\text{N/m}$, $\zeta \approx 0.1$, $L_g \approx 1.4\,\text{cm}$ and $P_s \approx 8\,\text{cm}\,\text{H}_2\text{O}$ ($h_0 = h_3 = 1.78\,\text{cm}$, $h_c = 0$, $c_k = 4$, $c_\zeta = 1.5$). This set of values will be hereafter referred to as the *typical glottal condition*, and the waveforms obtained for this set of values will be called *typical glottal waveforms*. The values assigned to the collision constants c_k and c_ζ are chosen so that a satisfactory behavior at closure is attained. Vocal-fold length can take values between $1.3\,\text{cm} < L_g < 1.7\,\text{cm}$ for women and $1.7\,\text{cm} < L_g < 2.4\,\text{cm}$ for men. L_g can be stretched in 3 or 4 mm during phonation [22]. Subglottal pressure P_s may vary from $8\,\text{cm}\,\text{H}_2\text{O}$ in normal conversation (60 dB SPL) to $360\,\text{cm}\,\text{H}_2\text{O}$ (120 dB SPL) for a tenor singing at full volume [23].

Throughout this article, we will assume that some of these parameters (namely, $h_0, h_3, h_c, c_k, c_\zeta$) are fixed. This does not mean that the model is not acoustically sensitive to the variation of these parameters. It is a decision we make in order to restrict our control parameters to those which can be directly interpreted in terms of a physiological action. It is worth remarking that m, d and L_g make part of the active control parameters since a speaker can vary the vocal-fold mass, length and thickness participating in vocal-fold vibration.

The additional symmetry imposed by the assumption of a symmetrical glottal structure entails an interesting reduction in the number of mechanical control parameters. Let us recall that the traditional two-mass model needs at least *twenty-one* parameters to reproduce characterisitc glottal-flow signals, while the phenomenological description of the glottal-flow signal itself can be attained with as few as *five* acoustic parameters, including fundamental frequency. The control parameters in the symmetrical model amount to *seven* quantities, namely $d, m, k, k_c, \zeta, L_g, P_s$, thus reducing the gap between acoustic and physical parameters for voiced sound reproduction.

It is worth noting that nothing in this formalism forbids an eventual distinction between upper and lower masses. The model admits an asymmetrical vocal-fold structure as well, but as we will show throughout our acoustic analysis, the assumption of a symmetrical vocal-fold structure does

not hinder reproduction of the wide variety of acoustic properties observed in experimental glottal-flow signals.

3. Algorithmic procedures

Data generation for an acoustic analysis of the above-described vocal-fold model is carried out by an algorithmic procedure comprising a numerical simulation of vocal-fold motion according to equations (2) and (3). Such simulations compute the dynamical variables $U_g(t)$, $y_1(t)$, $y_2(t)$ by means of an iterative process in time. For the implementation of vocal-fold motion simulation with the Niels Lous model we follow [16].

In order to study the response of the model to the variation of control parameters, three additional tasks have to be performed: prescribing the way in which control parameters will be varied, extracting dynamical variables which can be compared with experimental data, and measuring acoustic parameters from glottal-flow signals.

Let p be one of the control parameters of the model. It can be varied in two different ways: either
(a) we set p to vary in time within the vocal-fold motion simulation, so that $p = p(t)$ as $U_g(t), y_1(t), y_2(t)$ are calculated, or
(b) we set p to adopt a number of values within a given range and we compute $U_g^p(t), y_1^p(t), y_2^p(t)$ for each p.

We will use (a) to compare real-time control parameter variation with experimental data, in particular with experimental electroglottographic signals, and (b) for a numerical measurement of acoustic parameters. Further details on the algorithms performing these tasks is given below.

3.1. Numerical simulation of electroglottographic signals

In order to compute glottal-flow evolution throughout the real-time variation of one of the control parameters of the model over a chosen range, an algorithm is implemented (see the flow diagram in figure 3). The initialisation box requires input for:
- the algorithm parameters (voicing time t_{fin}, sampling rate),
- the control parameters of the model,
- the inclusion or discarding of acoustic coupling to the vocal-tract in the simulation.

The control parameter, p, and its range of variation, (p_{ini}, p_{fin}), can be selected. The increment Δp is computed in order to attain p_{fin} at t_{fin}. Notice that if Δp is sufficiently small, the variation of p does not produce transients and the simulation corresponds to a smoothly varying glottal-flow signal which actually resembles the result of a physiological gradual action.

The shaded box in Figure 3, corresponding to vocal-fold motion simulation with the Niels Lous two-mass model, contains the iterative process in time that allows calculation of $y_1(t), y_2(t)$ and $U_g(t)$ as in [16]. This iterative process is slightly modified to compute dU_g/dt, $x_s(t)$ and $a(t)$, where $a(t)$ denotes the contact area between
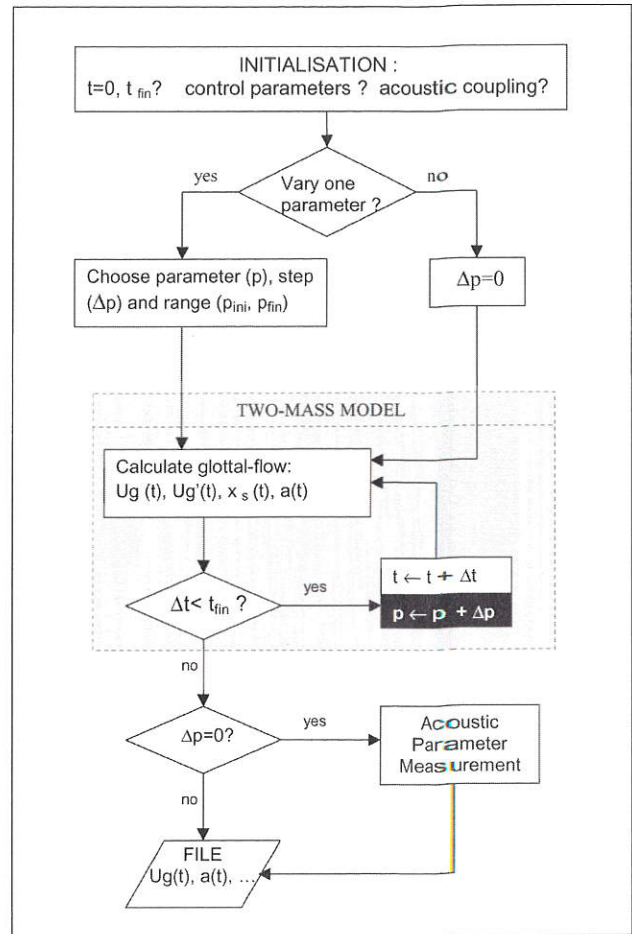

Figure 3. Flow diagram of the algorithm simulating real-time variation of one of the control parameters of the model.

the folds. Notice that the traditional two-mass model does not allow calculation of contact area because the projected area in IF is always rectangular and there is no gradation in opening or closing [24]. Instead, the vocal-fold geometry depicted in Figure 1, admits a gradual variation of contact area in time, which is given by:

$$a(t) = L_g \cdot x_c(t), \qquad (4)$$

where $x_c(t)$ is the distance along which $h_{2,1}(x,t) \leq 0$. Computing $a(t)$ is important since the contact area between the folds has been conjectured to correspond to electroglottographic measurements [24]. The electroglottographic technique consists in passing a high frequency electric signal (2–5 MHz typically) between two electrodes positioned at two different locations on the neck. Tissues in the neck act as conductors whereas airspace narrows the conducting path. When airgaps are reduced, the overall conductance between the electrodes increases. Glottal closing (opening) is consequently associated with an increase (decrease) in the electroglottographic signal. The electroglottographic signal (EGG) gives thus an indication of the sealing of the glottis, and constitutes a direct measurement of vocal-fold vibration. The numerical simulation of electroglottographic signals is obtained by run-

ning the algorithm and plotting $a(t)$. If $\Delta p \neq 0$, the underlying variation of a control parameter provides an EGG simulation in the course of a hypothetical physiological action.

The data output file contains $U_g(t)$, dU_g/dt, $h_1(t)$, $h_2(t)$, $a(t)$ and $x_s(t)$. The glottal-flow volume derivative can be used to generate synthetic sound files for perception analysis. In fact, dU_g/dt is a good approximation to the radiated sound pressure [4, 9]. The sound output file allows the listener to perceive the effect of the variation of a control parameter and hence of the associated physiological action, regardless of whether such an action is effectively possible for a human speaker without inducing variations of the rest of the physical parameters which have been kept constant during the simulation.

Notice that if Δp has been set to zero, control parameters are all kept constant, and therefore an additional action can be performed: acoustic parameter measurement. The procedure used to measure acoustic parameters from steady glottal-flow time series is discussed in the next paragraph.

3.2. Numerical measurement of acoustic parameters

The flow diagram corresponding to the algorithm used to compute acoustic parameters as a function of control parameters is shown in figure 4. The initialisation box will prompt the user to set the voicing time t_{fin}, the sampling rate and the control parameters that will be varied (p_q with $1 \leq q \leq 3$, i.e three at most) with their respective ranges of variation and increment steps. Simultaneous variation of more than one control parameter is important to seize the intercorrelations between them. Variation of a single control parameter is also necessary to understand the acoustic correlate of its variation. While the selected control parameters p_q are varied, the remaining control parameters are set to their default values, which are those of the typical glottal condition. The algorithm will iterate over the allowed values of p_q. For each set of values given to p_q, the algorithm performs four actions, namely
- simulating vocal-fold motion with the Niels Lous model (i.e. generating a vector type variable containing $U_g(t)$ and $dU_g/dt \ \forall t < t_{fin}$),
- computing acoustic parameteres for the resulting glottal-flow signals (using both $U_g(t)$ and dU_g/dt),
- storing p_q followed by the acoustic parameters in a file and
- incrementing p_q.
At the end of the $q-$multiple loop, the output file contains $q+5$ columns with the values of $p_q, F_0, E, O_q, S_q, T_a$ obtained within each iteration.

It is worth remarking that t_{fin} must be adjusted to a value which greatly exceeds the build-up time required for the oscillations to settle to a steady state ($t_{fin} > 0.1$ s). Notice however that for certain values of p_q, steady-state oscillations may not settle at all. The limits of the model to produce oscillations should *a priori* correspond to the limits of the phonation apparatus, which is uncapable of producing voiced sounds beyond certain physiological possi-
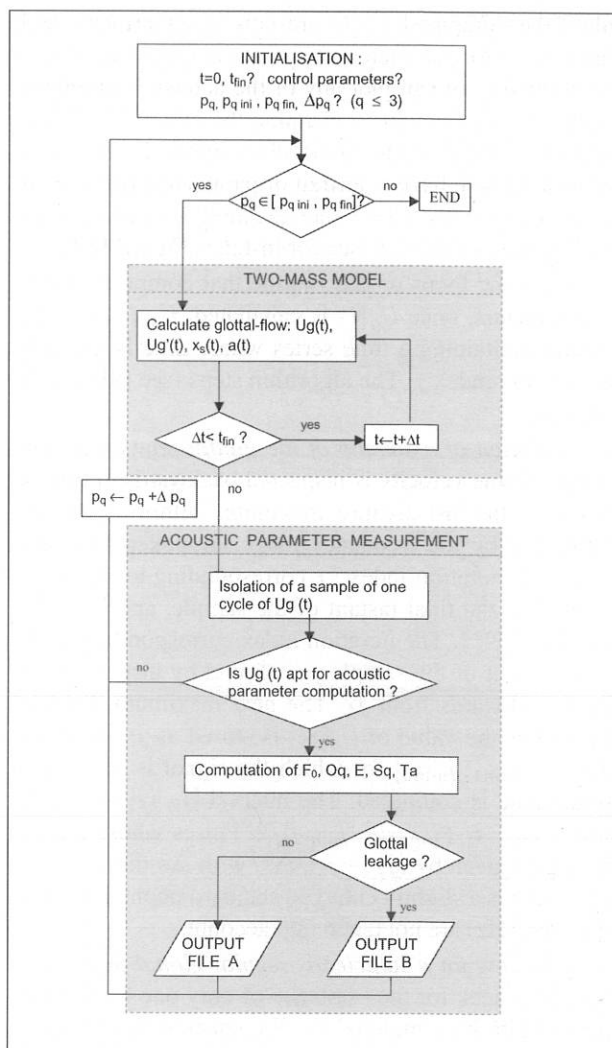

Figure 4. Flow diagram for the algorithm of numerical measurement of acoustic parameters.

bilities. The reader must bear in mind that these physiological constraints do not only correspond to, for instance, a maximum value of subglottal pressure that the lungs can attain. It may also happen that the lungs are capable of producing high values of subglottal pressure for which the vocal-fold mechanical system is unable to oscillate, unless the rigidness of the folds is high enough, for instance. In this example, the vocal folds will not reach steady-state oscillations for a high P_s and a low k_c, even if the lungs can effectively attain such a value of P_s. In such cases, the algorithm computes $U_g(t)$, but the glottal-flow signal does not present the expected periodic shape necessary for acoustic parameter computation (Figure 2). The algorithm will then skip this phase and directly increment the varied parameters without storing results in the output file.

To illustrate the algorithm procedure, let us consider an example. Let us choose to vary two control parameters: $k \in [10 \, \text{N/m}, 110 \, \text{N/m}]$ in steps of 5 N/m and $m \in [0.01 \, \text{g}, 0.14 \, \text{g}]$ in steps of 0.01 g. The program will iterate over the values of k and m and store in the output file the values of $m, k, F_0, E, O_q, S_q, T_a$ corresponding to each iteration,

unless the computed $U_g(t)$ presents irregularities which inhibit acoustic parameter computation. Once the process is completed, we can plot any of the acoustic parameters versus $\{m, k\}$ in order to examine the effect of the variation of m and k on the glottal-flow signal. If we plot m versus k we will have a portrait of parameter space, i.e. of the values of m and k for which the model predicts regular steady-state oscillations (see for instance Figure 15d).

Let us now focus on the routine that computes acoustic parameters, once $U_g(t)$ is calculated. $U_g[j]$ is in fact a vector containing a time series where time is given by the iteration index j. The algorithm steps (see [9]) are the following:

1) *Isolation of a sample of the glottal-flow cycle*: The glottal volume velocity is inspected backwards in time to search for the last greatest maximum within an interval established by the frequency range in spoken and sung voice. The iteration index j_f corresponding to this event is stored as the final instant of the sample, and $U_g[j_f]$ is stored as U_g^{max}. The iteration index corresponding to the initial instant of the sample j_i is found by inspecting the signal backwards from j_f. The next maximum that best approaches the value of $U_g[j_f]$ is stored as j_i. Next, the interval $[j_{min1}, j_{min2}]$ for which the signal is at its minimum value is computed. The interval $[j_i, j_f]$ is reset to start at $j_{min} = (j_{min1} + j_{min2})/2$. Pulses whose temporal length (given by $(j_f - j_i)/\Delta s$, with Δs the sampling rate) exceeds a slightly enlarged standard phonation range ($[30, 1500]$ Hz) are not taken into account.

2) *Checking for a sufficiently regular glottal-flow waveform*: We check for the existence of only one local maximum within the sample of U_g. We check if this property is fulfilled during the cycles preceding the chosen sample of U_g (the oscillations build-up phase is excluded from this verification). In this way, we make sure the glottal-flow signal has reached a periodic steady-state. Similarly, we count the local extrema within the sample of dU_g/dt. In the absence of vocal-tract coupling, dU_g/dt should exhibit one local maximum and one local minimum, as in Figure 2. Other conditions, such as $|Ug[j_i] - Ug[j_f]| \leq U_g^{max}$, or $U_g[j_{min}] \leq U_g^{max}/2$, contribute to confirm that U_g has the suitable shape for acoustic parameter computation. If any of these conditions is not satisfied, irregularities for the corresponding control parameters are reported to the screen, and the next steps (acoustic parameter computation, glottal leakage detection and storing results in the output file) are skipped. Notice that we have not conditioned dU_g/dt to be derivable. In fact, the activation of the separation criterion is expected to produce additional discontinuities, which a priori do not prevent acoustic parameter computation.

3) *Calculating acoustic parameters for the given sample*: We inspect dU_g/dt within $[j_i, j_f]$. We compute T_p by substracting the iteration index (j_1) corresponding to the first non zero value of dU_g/dt and the iteration index (j_2) associated with the maximum of U_g. A_v is directly $U_g[j_2]$. We compute T_e from $(j_3 - j_1)$ where (j_3) corresponds to the minimum value of dU_g/dt. E is directly $dU_g/dt[j_3]$.

Finally, T_a is computed by substracting the iteration index j for which $U_g'[j] > E/4$ and j_3. The acoustic parameters are calculated in terms of these values following the definitions presented in the previous paragraph.

4) *Checking for glottal leakage*: If $U_g[j_{min}] \neq 0$ (incomplete closure of the glottis) the control parameter values for when this happens are stored in a separate file.

Notice that the measurement of T_e is performed in terms of the glottogram derivative. Hence, when there is glottal leakage (i.e. the transglottal air flow does not reach zero during the quasi-closed phase), T_e no longer stands for the duration of the open phase but simply for the time needed to attain the maximum rate of decrease in flow. Therefore, the reader should keep in mind that, throughout this work, glottal leakage is not represented by a unit value of O_q but by a separately measured non-zero minimum value of the glottal flow.

4. Results

4.1. The typical glottal condition

Let us first consider the symmetrical two-mass model, without coupling to the vocal tract, and with the control parameters taking the values of the typical glottal condition listed in section 2.3.

The model predictions are reproduced in Figure 5a and b for a phonation frequency of about 100 Hz. The discontinuities at the vocal-fold opening and closure instants are mainly due to the absence of viscosity in the flow model (notice that glottal-flow signal models do not assume that dU_g/dt should be derivable at the opening and closure instants). The additional discontinuity in the derivative of $U_g(t)$ before closure is due to the activation of the separation criterion. Figure 6 shows the instantaneous values taken by x_s during the cycle shown in Figure 5a and b. When $h_2(t) > sh_1(t) > 0$ ($s = 1.2$) the separation point x_s moves from x_2 towards x_1 and hence, the pressure difference between x_1 and x_s used in equation 3 to calculate the flux decreases more rapidly, inducing a rapid decrease of U_g which is clearly visible in the glottal-flow derivative. Even if this kind of discontinuity is not prescribed in glottal-flow signal models, acoustic parameters are still meaningful in terms of the zeros and extrema of dU_g/dt within a period (see Figure 2), as anticipated in the algorithm for numerical measurement of acoustic parameters presented in the previous section.

Viscosity tends to slow down the opening and closing of the folds. Following [14] in the estimation of the pressure loss due to viscosity, the model predicts the smooth glottal-flow shown in figure 5(c) and (d). Notice that inclusion of the viscous term removes the discontinuity corresponding to the activation of the separation criterion as well. In fact, we have found that the viscous-flow correction will demand, for instance, higher subglottal pressures for the criterion to become active. In order not to favour an unrealistic (too sudden) closing behavior, a viscosity term corresponding to an approximation of a fully devel-
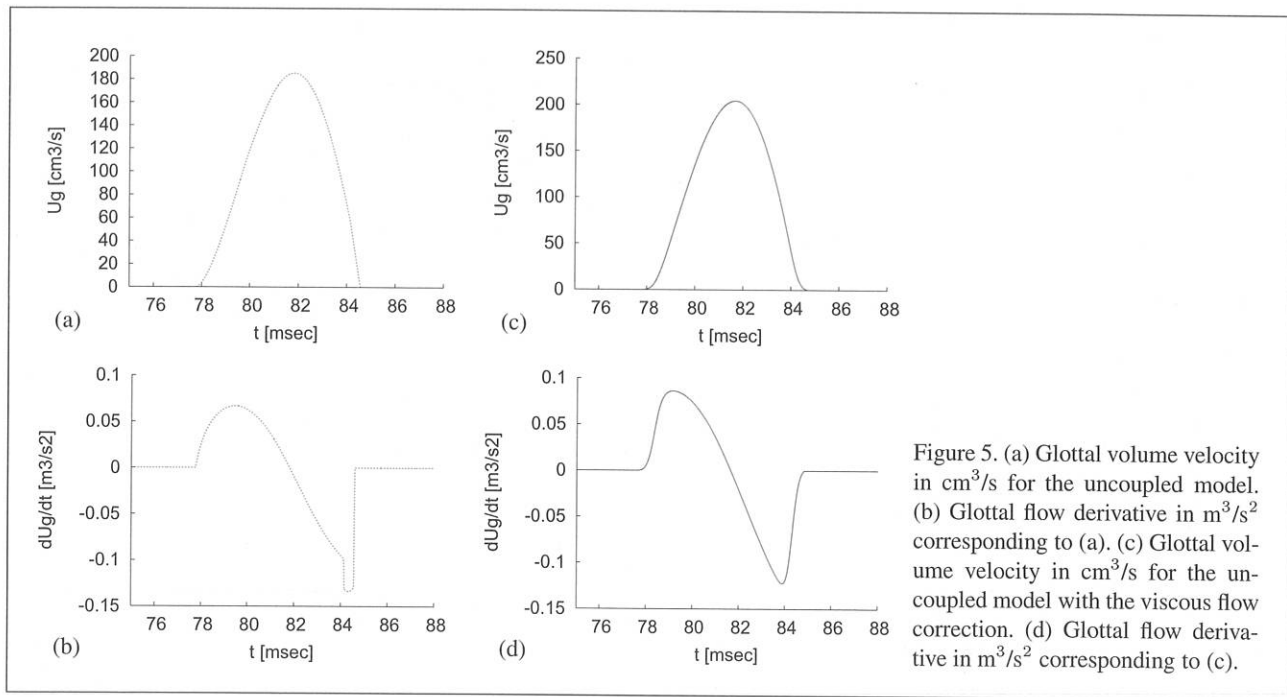
Figure 5. (a) Glottal volume velocity in cm^3/s for the uncoupled model. (b) Glottal flow derivative in m^3/s^2 corresponding to (a). (c) Glottal volume velocity in cm^3/s for the uncoupled model with the viscous flow correction. (d) Glottal flow derivative in m^3/s^2 corresponding to (c).
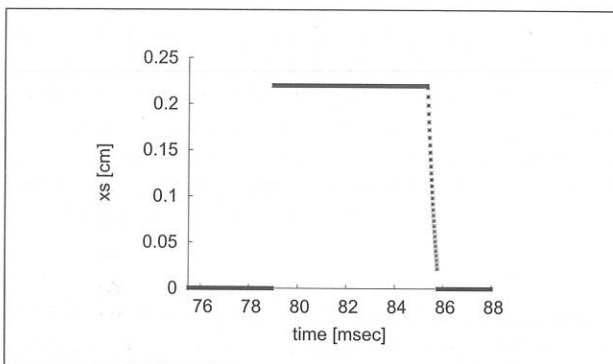

Figure 6. Position x_s of the separation point corresponding to Figure 5a and b.

oped Poiseuille velocity profile is hereafter included in our simulations.

$$\Delta p_{visc} \approx \frac{12\mu U_g}{L_g} \frac{x_2 - x_1}{\min{(h_1, h_2)}^3}. \qquad (5)$$

Examples of the effect of the vocal tract on the glottal flow waveform are given in Figure 7. Compare the glottogram generated by the uncoupled model to the one corresponding to the glottis coupled to the vocal tract for vowel /a/. The values of the control parameters are set in both cases according to the typical glottal condition (see section 2.3). Notice that even if $y_1(t)$, $y_2(t)$ and F_0 remain almost invariant when the vocal-tract shape is altered, the acoustic interaction between the vocal tract configuration and the glottal volume flow accentuates the asymmetry of the glottal-pulse shape and introduces formant ripples in the glottal flow waveform.

These results (concerning the sensitivity of the glottal-flow waveform to the vocal-tract shape in this model) are essentially similar to those obtained with previous two-mass models. This is not surprising: the representation of the vocal tract in the symmetrical two-mass model does not essentially differ from [4]. In order to concentrate on the new elements of this model, namely, the symmetry assumption and the geometry-dependent position of the separation point, we will hereafter disregard the acoustic load of the vocal tract and constrain our analysis to the acoustic effects originated by the parameters controlling glottal configuration. Certainly, the acoustic parameters measured in this work will not strictly correspond to a "true" glottal airflow, but their variation in terms of control parameters will not be masked by formant ripples and will be consequently more neatly evaluated [25, 26]. For recent discussions on the importance of acoustic feedback into fold oscillations from the vocal tract, see [9, 27, 28].

4.2. Acoustic parameter sensitivity to control parameters

The acoustic characterization of this symmetrical vocal-fold model poses a number of questions among which the first is whether it is able to reproduce the whole range of values for acoustic parameters as measured in experimental glottal-flow signals. Our analysis shows that there is a positive answer to this question and that acoustic parameters may attain values with the Niels Lous model that cannot be attained with the asymmetrical IF model [9].

The variation of m, k and P_s suffice to reproduce the standard phonation frequencies ($F_0 = [30, 1500]$ Hz). The open quotient can also be made to vary from $[0.3, 1]$ if we assume here that the value 1 represents glottal leakage. Likewise, $S_q \in [0.8, 9.0]$, $E \in [0, 160]$ m^3/s^2 and $R_a = T_a/T_0 \in [0.02, 0.18]$.

The sensitivity of acoustic parameters to the variation of physical control parameters is a good indicator of the ac-
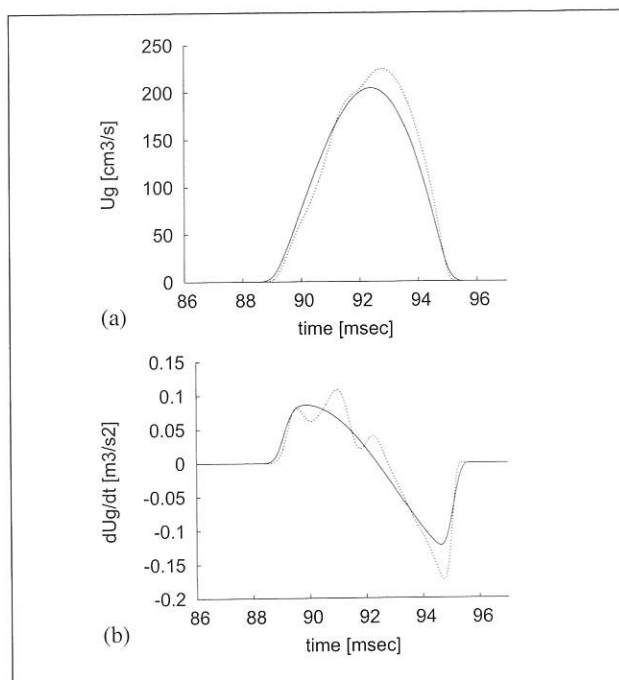
Figure 7. (a) Glottal volume velocity in cm^3/s in the absence of acoustic coupling with the vocal tract (full line), and with vocal tract as in vowel /a/ (dotted line). (b) Glottal flow derivative in m^3/s^2 corresponding to (a).
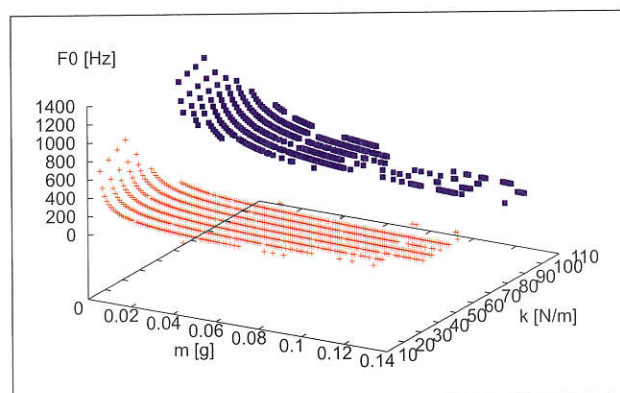

Figure 8. Variation of fundamental frequency as vibrating mass (m) and vocal-fold tension (k) are varied. The region in red represents phonation with complete glottal closure while the region in blue corresponds to phonation with glottal leakage.

Figure 9. Variation of fundamental frequency with subglottal pressure: (a) for the symmetrical model for $P_s > 10$ cm H_2O, (b) for the range of subglottal pressure in which both models (IF and Niels Lous) oscillate. The points in the upper left corner correspond to the symmetrical model with glottal leakage, and the points below correspond to the symmetrical model without glottal leakage. The points in the center correspond to the IF model. Values of control parameters other than subglottal pressure have been chosen to follow in both cases the typical glottal condition.

tions that the modelled glottis employs to produce voiced sounds of different characteristics. We will therefore outline the general tendencies observed in the variation of acoustic parameters as control parameters are varied.

4.2.1. Fundamental frequency control

Titze [29] has observed that increasing fundamental frequency is mainly the effect of four possible actions: a contraction of the vocalis (increase of the vocal-fold tension, i.e. of their spring constant in a two-mass model), a decrease in the vibrating mass, an increase in the subglottal pressure and a decrease in the vibrating length.
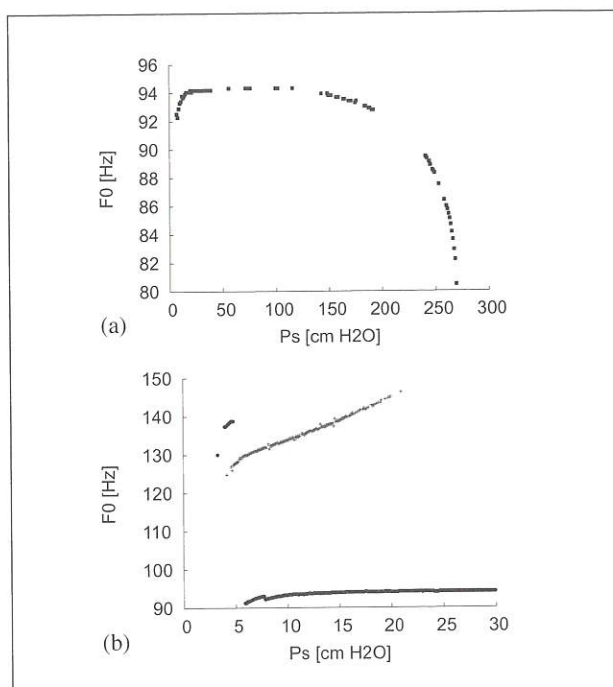
Our acoustic analysis shows that a symmetrical two-mass model attains the highest values of F_0 by decreasing m and increasing k: this is specially efficient if both actions take place simultaneously, as shown in Figure 8.

Increasing P_s also induces an increase in the fundamental frequency when $P_s < 40$ cm H_2O. For 40 cm $H_2O < P_s < 150$ cm H_2O, subglottal pressure does not induce substantial changes in frequency. Finally, for $P_s > 150$ cm H_2O, the effect is the opposite: increasing subglottal pressure induces a decrease in F_0 (see Figure 9a). It is interesting to compare these results to those predicted by the traditional two-mass model. The evolution of F_0 with subglottal pressure for the IF model is shown in Figure 9b. The points in the upper left corner correspond to the symmetrical model with glottal leakage, the points in the center correspond to the IF model and the points below correspond to the symmetrical model without glottal leakage. First of all, it is worth noting that the IF model does not oscillate for $P_s > 20$ cm H_2O: it only oscillates for low values of subglottal pressure, inducing an increase in F_0. The symmetrical model predicts a much more complex behavior: there is glottal leakage when the subglottal pressure is very low and this produces higher frequencies than those obtained when there is complete glottal closure.

As Titze observes [29], a decrease in the vibrating thickness d entails a slight increase in F_0 according to our sim-
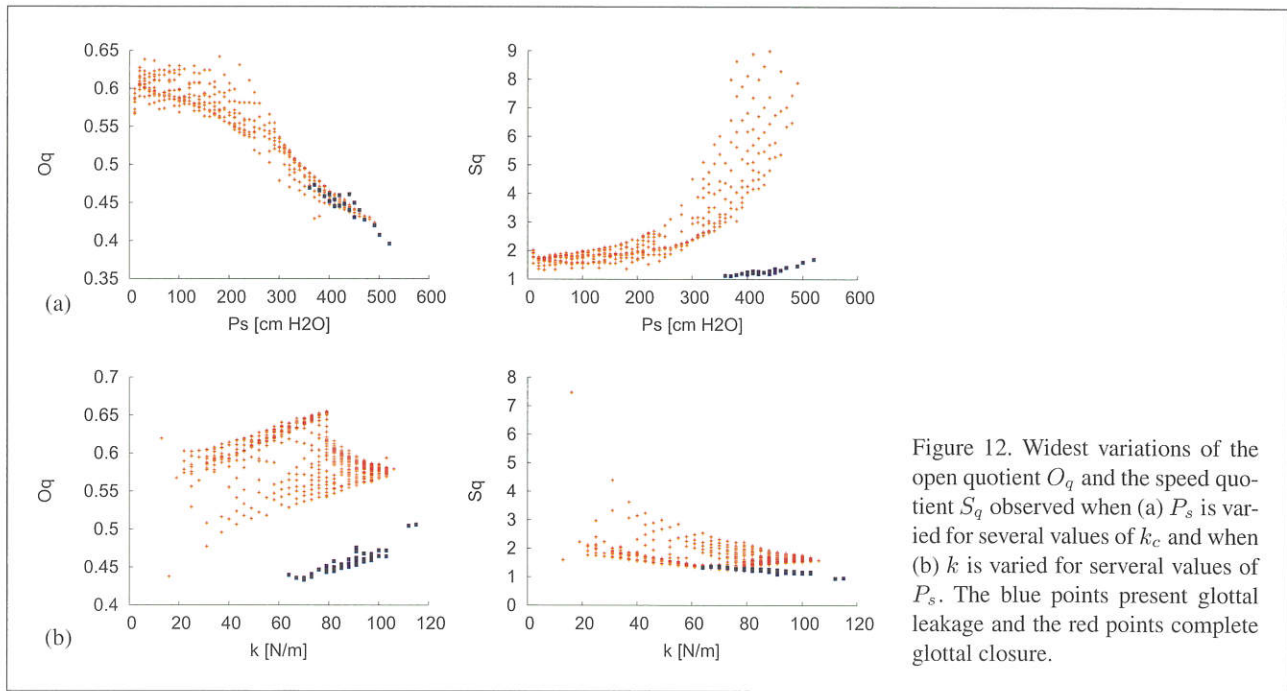
Figure 12. Widest variations of the open quotient O_q and the speed quotient S_q observed when (a) P_s is varied for several values of k_c and when (b) k is varied for serveral values of P_s. The blue points present glottal leakage and the red points complete glottal closure.
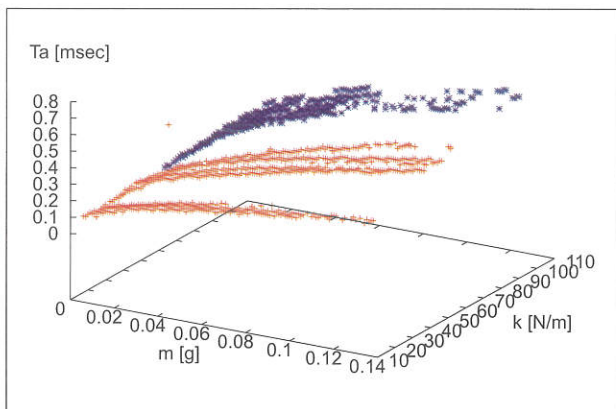

Figure 13. Variation of T_a with m and k. The blue points indicate glottal leakage. The red points indicate oscillations with complete glottal closure.
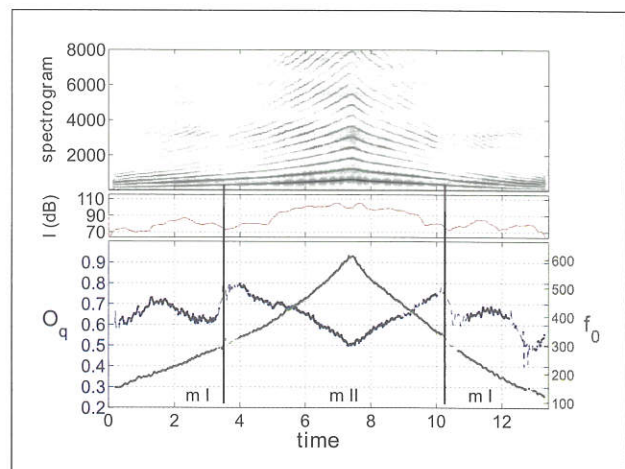

Figure 14. Spectrogram, variation of intensity, and variation of fundamental frequency and open quotient for a glissando sung by a tenor, as reported by N. Henrich in [34].

4.3. Oscillation regimes and laryngeal mechanisms

4.3.1. Laryngeal mechanisms

Laryngeal mechanisms denote different phonation modes with well-defined acoustic characteristics. The question of laryngeal mechanism reproduction with low-dimensional vocal-fold models is of great importance in vocal-fold modelling research, as it constitutes a well-known acoustic phenomenon in direct connection with vocal-fold motion [33].

Laryngeal mechanisms are usually defined in terms of glottal configuration and muscular tension. In a vocal-fold model, glottal configuration is easily quantified by some of the control parameters mentioned above, namely m, d and L_g, while muscular tension is represented by k and k_c.

For instance, the glottal configuration adopted in what is called mechanism 0 (m_0) or *vocal fry* corresponds to

k and L_g small and d high. The vibration in this mechanism presents a very short open phase (i.e. glottal-flow is non-zero during a small fraction of the oscillation period). Glottal configuration adopted in mechanism I (m_I), corresponding to the so-called *modal voice* or *chest register*, is such that the vibrating tissue is long, large and dense. In terms of control parameters, m_I is associated with high values of m, d and L_g. During phonation in mechanism II (m_{II}), corresponding to the so-called *falsetto voice* or *head register*, vocal-folds become tense, slim and short. This laryngeal mode differs from m_I in aspects regarding glottal configuration, muscular tension and glottal closure. The reduction in the length of the folds that participates in vibration is caused by an accentuated compression between the arytenoids. On the other hand, vibration in m_{II}

ulations, but this effect is much less important than the effects mentioned above. The effect of the remaining parameters is the following: an increase in ζ induces a slight decrease in F_0, while an increase in k_c or L_g induces a slight increase in F_0.

4.2.2. Intensity control

Gauffin and Sundberg [30] have found that the SPL of a sustained vowel shows a strong relationship with the negative peak amplitude of the differentiated glottogram, which we have called speed of closure E.

For a male speaker, Fant *et al* [31] found that E was proportional to $P_s^{1.1}$, which is very close to the linear relation observed in [32]. Numerical computation of E for the symmetrical model as subglottal pressure is varied, yields the relation shown in Figure 10.

The model induces a relation between E and P_s which is reasonably approximated by Fant's relation. The detail obtained in our numerical results may be attributed to the strict invariance of the other physical parameters in our simulation. In fact, if we consider the effect of varying subglottal pressure with an underlying variation of another parameter (e.g. k_c in Figure 11), $E(P_s)$ presents a dispersion which resembles measurements presented by [32] and which makes the detailed behavior observed in Figure 10 no longer visible. Figure 11 also shows that beyond $300\,\text{cm}\,H_2O$, glottal leakage allows to maintain an increase in E following Fant's relation.

Considering the variation of E with the seven control parameters, we have found that the highest values of E are attained by increasing P_s and k_c: once more, this is specially efficient if both actions take place simultaneously, as shown in Figure 11. The effect of other parameters is less important. Increasing d or L_g tends to favor an increase in intensity while a big vibrating mass m would produce the opposite effect. The influence of ζ or k on intensity is quite weak.

4.2.3. Control of the glottal pulse shape

For the typical glottal condition, phonation at $100\,\text{Hz}$ presents $O_q \approx 0.5$, $S_q \approx 2$ and $T_a \approx 0.5\,\text{ms}$. *Breathiness* is easily indicated by the existence of glottal leakage, which is usually accompanied by an increase of T_a and a decrease of S_q.

The widest ranges of variation for O_q and S_q are generated when P_s, k and k_c are varied. An increase in P_s or k_c entails a reduction of O_q and an increase in S_q, while the effect of k is quite the opposite. This is shown in Figure 12.

When P_s, k, k_c keep values close to the typical glottal condition, O_q and S_q are bounded to smaller ranges, namely, $O_q \in [0.45, 0.65]$ (recall that glottal leakage is calculated separately), $S_q \in [1, 3]$. An inverse proportionality between O_q and S_q is generally present. In other words, when either k or L_g are increased, O_q increases and S_q decreases and when either k_c or P_s are increased, O_q decreases and S_q increases. A simultaneous increase (or decrease) of O_q with S_q in phonation would imply - in the context of this model- a simultaneous and balanced variation of parameters inducing opposite effects.
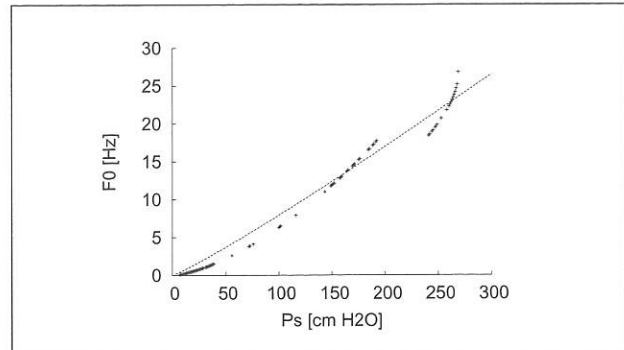

Figure 10. Variation of E as subglottal pressure (P_s) is varied from numerical measurements in the symmetrical model (pluses). The dotted line corresponds to the values of E predicted by Fant's relation [31].
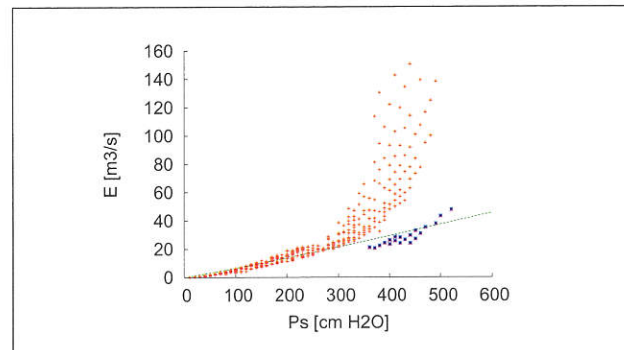

Figure 11. Variation of E as subglottal pressure (P_s) is varied for several values of k_c. There is complete glottal closure for the points in red and glottal leakage for the points in blue. The green line corresponds to the values of E predicted by Fant's relation.

Our numerical measurements show that glottal leakage is invariably associated with low values of S_q and high values of T_a in comparison with the values of these acoustic parameters when there is complete glottal closure. This regularity is in accordance with the above description of breathy voice. Physiological actions related to breathiness will be further discussed in the following section.

Abrupt glottal closure ($T_a \approx 0$) is typically present when parameters in set $C = \{\zeta, m, k, P_s\}$ have low values (with respect to the typical glottal condition). See Figure 13 for an example. This is also bound to happen for large values of d or L_g. Values of T_a are certainly dependent on F_0: the highest values of T_a (which may reach $4\,\text{ms}$) are attainable when the fundamental frequency is low enough.

It has been observed that S_q is generally correlated with T_a. In fact, this holds during the variation of any of the control parameters with the exception of the vibrating mass m (which entails an increase in T_a while S_q remains almost constant), as well as for the coupling spring constant k_c.
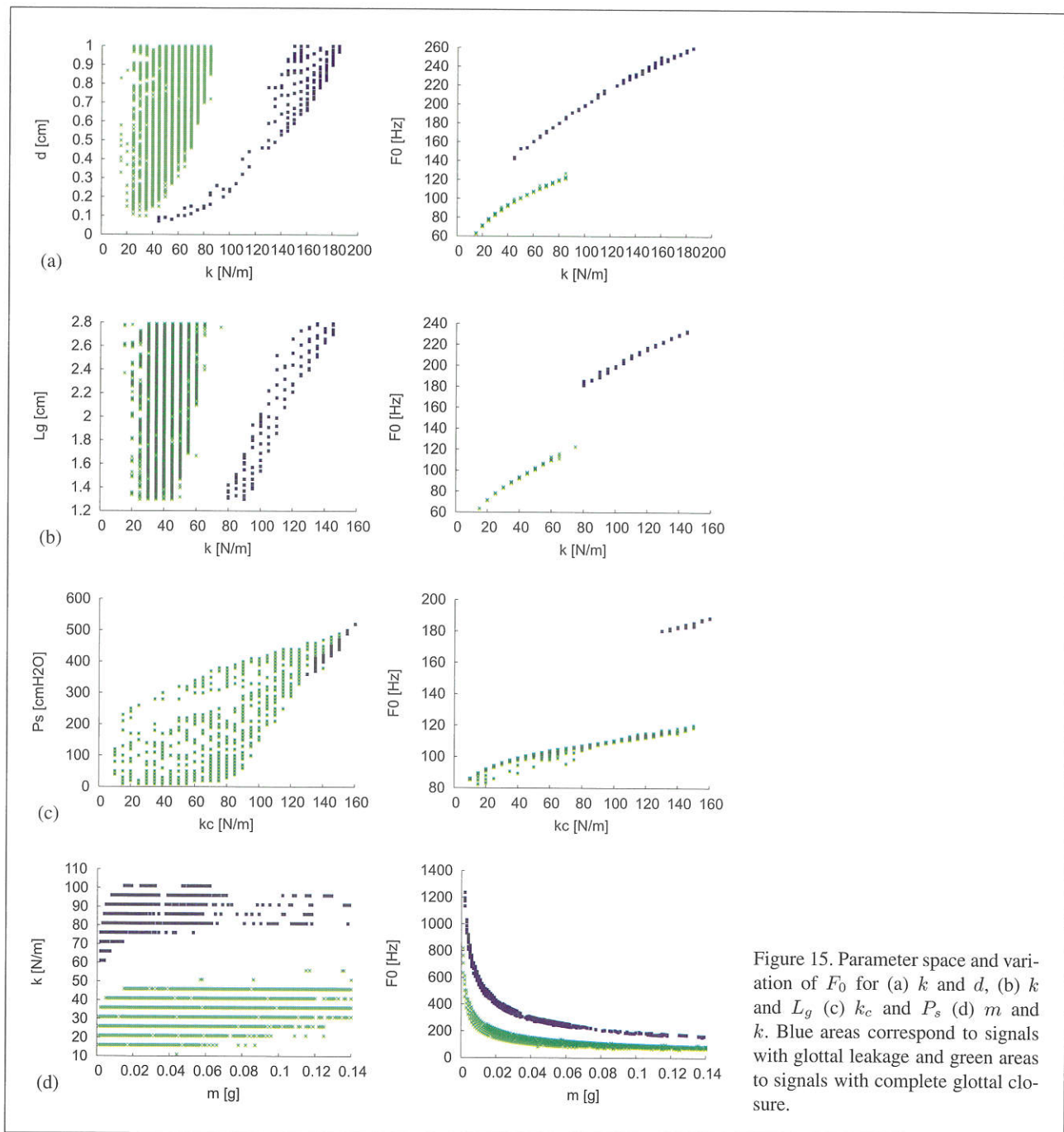
Figure 15. Parameter space and variation of F_0 for (a) k and d, (b) k and L_g (c) k_c and P_s (d) m and k. Blue areas correspond to signals with glottal leakage and green areas to signals with complete glottal closure.

usually implies a certain degree of glottal leakage: the transglottal airflow does not reach zero during the quasi-closed phase as a consequence of an incomplete glottal closure. In terms of the model, m_{II} means low values of m, d and L_g, while k and k_c are considerably higher.

Laryngeal mechanisms can also be identified in terms of acoustic parameters [1]. As fundamental frequency F_0 is increased, one can notice a voice break corresponding to the change between m_I and m_{II} (see Figure 14). Generally, m_I corresponds to lower values of F_0, a low O_q, and a stronger intensity. Instead, m_{II} corresponds to higher values of F_0, a high open quotient and a weaker intensity. Vocal fry (or m_0) may be activated when the vocal apparatus is forced to produce frequencies lower than 30 Hz.

4.3.2. Oscillation regimes

The preceding section suggests that simulations with different values of m, d, L_g, k and k_c should in principle be able to reproduce different laryngeal mechanisms, provided the vocal-fold model is sound enough. Whether glottal-flow signals generated with a symmetrical model effectively correspond to phonation in a certain mechanism is a question that we will attempt to answer from the results of our numerical simulations.

Numerical experiments show that as m, k, d, L_g, P_s, or k_c are varied in pairs, distinct oscillation regimes are clearly visible. Figure 15 shows parameter space for some of these control parameters, in which we encounter two distinct regions within which regular vocal-fold oscilla-

tions take place. In these examples, the blue square points correspond to signals with glottal leakage, while the green crosses correspond to signals with complete glottal closure. Notice that within a single region in parameter space, the variation of fundamental frequency is smooth.

Regimes with glottal leakage systematically present higher values of F_0, a lower intensity and a higher open quotient. Besides, they are activated as k or k_c increase and reaching them implies less muscular effort if d or L_g are small. In order to attain the highest frequencies, it is necessary to lower m. All these features suggest a correspondence between m_{II} and the oscillation regimes of the symmetrical two-mass model which present glottal leakage.

Distinct oscillation regions may also appear for oscillations without glottal leakage. An example is shown in Figure 16 where m and P_s are simultaneously varied. The transition from one region to another implies a jump in F_0. However low F_0 is in the right region of Figure 16, an identification of this oscillation regime with m_0 is not possible since the correspondent glottal-flow signals do not present a sufficiently short open phase. A simultaneous lowering of k and L_g as d is increased (with respect to the typical glottal condition) has been simulated in search of an oscillation regime which could be identified with m_0, since this laryngeal mechanism is described by a physiological action of this kind. However, these numerical experiments have not allowed us to find oscillation regimes resembling m_0.

4.3.3. Transition between regimes

• The nature of the transition:

The transition from one regime to another is generally marked by a jump in fundamental frequency. Consider Figure 15 and notice that moving from the green to the blue regions involves a jump in F_0. However, note that moving from one regime to another in parameter space does not necessarily imply a sudden change in control parameters to produce the jump in F_0. In the upper right corner of (c), for instance, or in the lower left corner of (a), it is possible to pass from the green to the blue region with a smooth variation in (k_c, P_s) or in (k, d) and this smooth variation will anyway induce a jump in fundamental frequency. These situations correspond to a bifurcation of the dynamical system governing vocal-fold oscillations, in the sense that a sudden qualitative change in the behavior of the system takes place during a smooth variation of control parameters [35].

This distinction is important since laryngeal mechanisms have been first attributed to a sudden modification of the activity of the muscles, whereas recently it has been suggested that transitions may be due to bifurcations in the dynamical system [35]. Our calculations show that, *a priori*, both possibilities may hold. According to our results, it is the choice and value of the control parameters which are varied during the transition that will determine whether a discontinuous physiological action is necessary to induce a jump in F_0. If this is true, the degree of train-
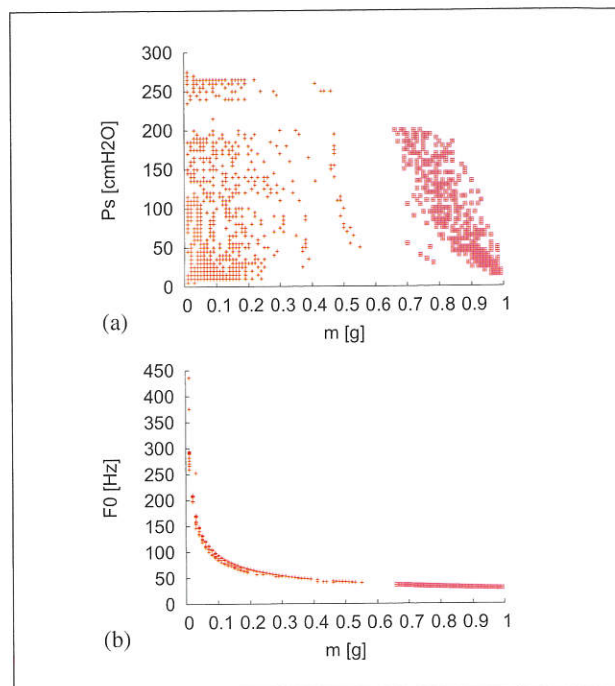

Figure 16. Parameter space and variation of F_0 for m and P_s. The points corresponding to the signals attaining the lowest values of F_0 are colored in pink.
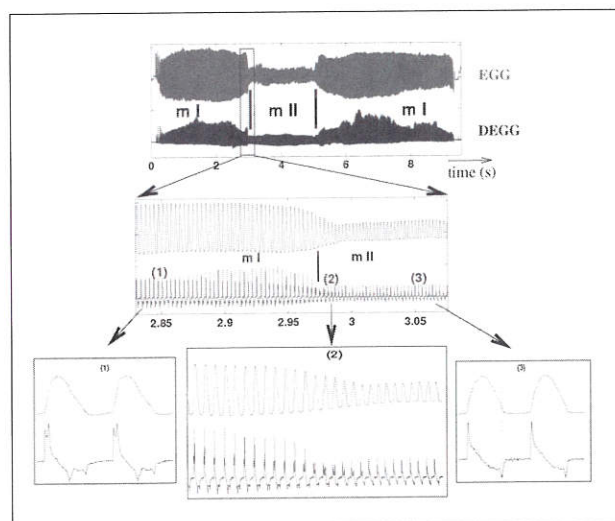

Figure 17. EGG and DEGG signals exhibiting peak doubling during a transition between laryngeal mechanisms m_I and m_{II}, observed in a glissando sung by a baritone, as reported by N. Henrich in [34]. The top panel presents the shape of both signals over the whole glissando. The middle and bottom pannels zoom on the transition.

ing of a speaker in the control of his vocal apparatus may result in different physiological solutions to produce a desired effect (such as increasing F_0 in a glissando).

• Transitions and electroglottographic signals:

Henrich [1] reports the existence of peak doubling in experimental DEGG signals $(\mathrm{d}a(t)/\mathrm{d}t)$, particularly next to or during the transition between the first and second laryngeal mechanisms [34]. Figure 17 shows that *right before*
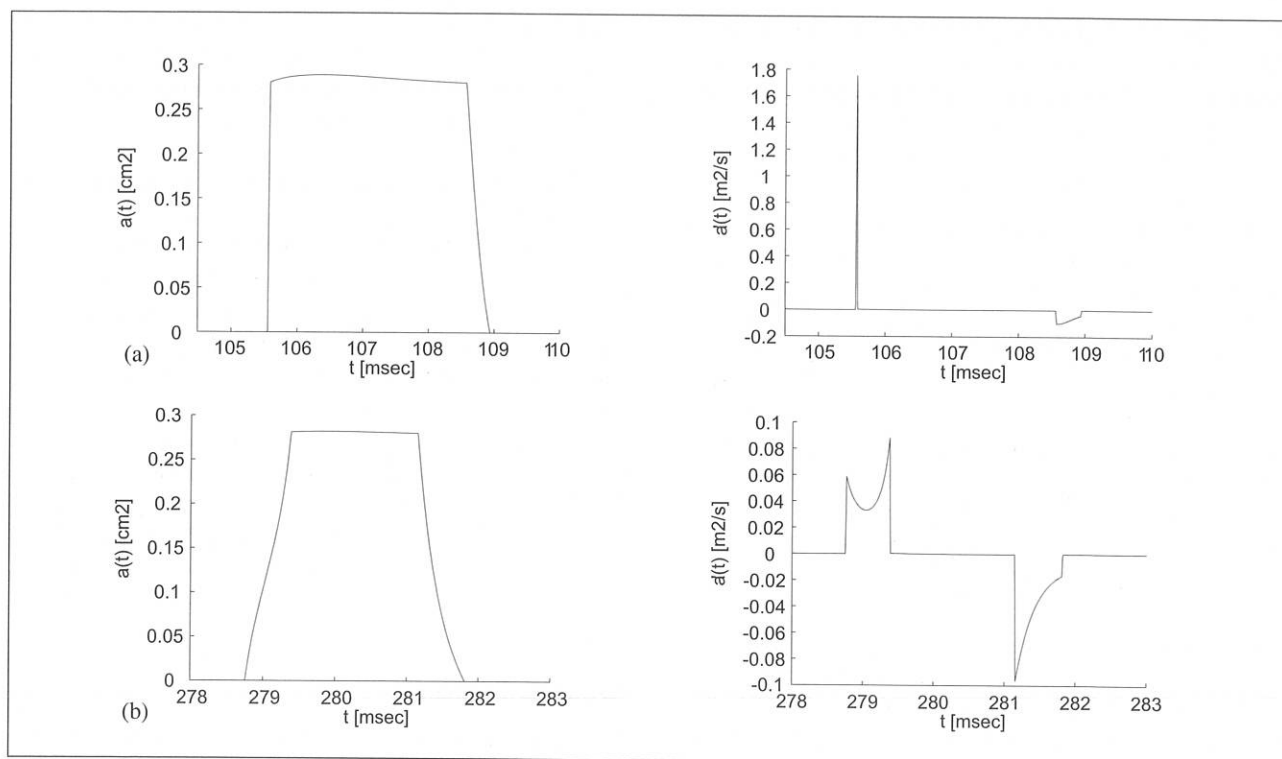
Figure 18. EGG and DEGG signals generated by vocal-fold motion simulation with the symmetrical model (a) before the transition and (b) during the transition between the green and blue regions of Figure 15c at $P_s = 450 \, \text{cm} \, H_2O$.

the transition (panel 1) both the opening and the closure peaks are doubled. During the transition (panel 2), some periods present double closure peaks and single opening peaks. After the transition (panel 3), both closure and opening peaks are single. Opening peaks are generally less clearly marked, while closure peaks are either extremely precise and unique, or they are neatly doubled. This phenomenon has been considered in a couple of experimental studies [36] and [37]. It has first been conjectured to be linked to (a) a slightly dephased contact along the length of the folds. If this is so, this kind of effect should be reproduced by a vocal-fold model in which a structure is assigned to the folds along L_g, as in Titze's model [15]. A second hypothesis has attributed double peaks to (b) a rapid contact along the $x-$direction followed by a contact along L_g.

Even if our simple and essentially $2D$ two-mass model does not allow either for (a) or (b), our numerical simulations show that double closure peaks can be clearly reproduced when a transition between oscillation regimes is occuring. As an example, Figure 18 shows a cycle of $a(t)$ and its derivative $da(t)/dt$, *well before* (a) and during (b) the transition between the green and blue regions in Figure 15(c). Just as observed in Figure 17, $da(t)/dt$ presents double closure peaks during the transition. The fact that the model reproduces double closure peaks during a transition between regimes constitutes another element in favour of the interpretation of oscillation regimes in terms of laryngeal mechanisms. These results suggest that peak-doubling at closure may occur due to a time-lag closure in the $x-$direction exclusively, provided that an

underlying variation of certain control parameters is producing a qualitative change in the behavior of the mechanical system.

5. Conclusions

Symmetrical two-mass models of vocal-fold oscillations constitute a new testbench in the quest for a physical phonation model capable of linking physiological actions to voice acoustics. It has been shown that the assumption of a symmetrical glottal structure does not hinder generation of glottal pulses covering the full parameter space, while a reduction in the number of control parameters is gained. We have examined the acoustic properties of the symmetrical two-mass model proposed by Niels Lous *et al* in [14], in which flow separation takes place at a variable position depending on the glottal geometry. For the characterization of glottal-flow waveforms, we have resorted to a set of acoustic parameters borrowed from phenomenological glottal-flow signal models [2], which is particularly useful for vocal intensity and timbre description.

An algorithm is developed in order to compute the acoustic characteristics of the model by generating the glottal airflow signal for different settings of the control parameters of the model. The algorithm allows examination of the glottal volume velocity, the position of the masses, the contact area between the folds and the position of the separation point as a function of time. It also simulates real-time control parameter variations for perception analysis and calculates the contact area function between the folds which can be compared with results ob-

tained from electroglottographic signals. From salient timing events of the glottal waveform, a number of source parameters are estimated for each glottal pulse. This approach allows for the mapping between the control parameters of the two-mass model and typical parameters used for characterising the voice source signal.

With this tool, we have determined the conditions under which the phenomenological description provided by the signal model can be applied to two-mass-model generated signals. Simulations without acoustic coupling to the vocal tract show that the activation of the separation criterion proposed by Liljencrants produces a discontinuity in the derivative of glottal volume velocity. This discontinuity is not prescribed in glottal-flow signal models but does not prevent acoustic parameter computation. The inclusion of a viscous-flow correction is shown to demand higher subglottal pressures for the separation criterion to become active (apart from predicting a smooth opening an closing of the vocal folds).

Simulations with acoustic coupling to the vocal tract show the degree in which the acoustic feedback of the vocal tract affects the glottogram shape, producing formant ripples in the glottal-flux derivative and accentuating the asymmetry of the glottal-pulse shape, just as observed for previous vocal-fold models. The effects of the vocal tract are left out from the correlation analysis between acoustic and control parameters, in order to concentrate on the acoustic effects of the variation of the source control parameters originated by the new elements introduced in [14].

The symmetrical vocal-fold model is shown to reproduce the whole range of values for acoustic parameters observed in experimental glottal-flow signals. These ranges are even wider than those attained with the traditional asymmetrical two-mass model. In fact, the symmetrical model admits oscillations in regions of parameter space that the asymmetrical two-mass model cannot reach (e.g. regions where $P_s > 20\,\mathrm{cm\,H_2O}$).

The sensitivity of acoustic parameters is an indicator of the actions that the modelled glottis employs to produce voiced sounds of different characteristics. Our study shows that the control of fundamental frequency is mainly obtained with a simultaneous increase in elasticity and a decrease in the vibrating mass of the folds. Intensity is particularly sensitive to subglottal pressure and vocal-fold rigidness. The open quotient is mainly controlled by a combined action of subglottal pressure and vocal-fold elasticity. In turn, variations in the abruptness of the glottal closure are produced by a simultaneous adjustement of the mechanical properties of the folds, including damping, as well as of subglottal pressure. Breathiness is determined by the vibrating thickness and length of the folds, as well as by their elasticity and rigidness.

Finally, our simulations show that the model produces distinct 'oscillation regimes' and that these can be identified with different phonatory modes (laryngeal mechanisms). Evidence is produced for the identification of some of these regimes with the first and second laryngeal mechanisms, which are the most common mechanisms used in human phonation. On the other hand, identification of low-frequency oscillation regimes with mechanism 0 (vocal fry) has not been possible, at least for a symmetrical glottal structure.

Transitions between oscillation regimes are shown to share features experimentally observed for transitions between laryngeal mechanisms. The double closure peaks reported in [1] for experimental electroglottographic signals during such transitions, has been reproduced using the contact area functions generated with the symmetrical production model. Such a result constitutes further evidence for the identification of laryngeal mechanisms with oscillation regimes. According to the symmetrical two-mass model, the nature of the transition between regimes may be of two types: either there is a sudden change in the activity of the muscles or there is an underlying bifurcation of the dynamical system. Which of both possibilities takes place will depend on the region of parameter space visited during the transition.

Acknowledgement

References

[1] N. Henrich: Etude de la source glottique en voix parlée et chantée. Thèse de Doctorat de l'Université Paris 6, 2001.

[2] B. Doval, C. d'Alessandro: Spectral correlates of glottal waveform models: an analytic study. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Munich, Germany, 1997, 446–452.

[3] C. Gobl, A. Ní Chasaide: Acoustic characteristics of voice quality. Speech Communication 11 (1992) 481–490.

[4] K. Ishizaka, J. L. Flanagan: Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell. Syst. Tech. J. 51 (1972) 1233–1268.

[5] B. H. Story, I. R. Titze: Voice simulation with a body-cover model of the vocal folds. J. Acoust. Soc. Am. 97 (1995) 1249–1260.

[6] J. W. Van den Berg, J. T. Zantema, P. Doornenbal: On the air resistance and the bernoulli effect of the human larynx. J. Acoust. Soc. Am. 29 (1957) 626–631.

[7] D. Sciamarella, G. B. Mindlin: Topological structure of flows from human speech data. Phys. Rev. Letters 82 (1999) 1450.

[8] R. Laje, G. B. Mindlin: Diversity within a birdsong. Phys. Rev. Lett. 89 (2002) 28, 288102–1/4.

[9] D. Sciamarella, C. d'Alessandro: A study of the two-mass model in terms of acoustic parameters. International Conference on Spoken Language Processing (ICSLP), 2002, 2313–2316.

[10] X. Pelorson, A. Hirschberg, R. R. van Hassel, A. P. J. Wijnands, Y. Auregan: Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model. J. Acoust. Soc. Am. 1994 (96) 3416–3431.

[11] I. J. M. Bogaert: Speech prodcution by means of hydrodynamic model and a discrete-time description. IPO-Report 1000, Institute for Perception Research, Eindhoven, The Netherlands, 1994.

[12] R. N. J. Veldhuis, I. J. M. Bogaert, N. J. C. Lous: Two mass models for speech synthesis. Proceedings of the 4th European Conference on Speech Communication Technology, Madrid, Spain, 1995, 1854–1856.

[13] A. Hirschberg, J. Kergomard, G. Weinreich: Mechanics of musical instruments. – In: CISM Courses and Lectures No. 355. Spinger-Verlag, 1995.

[14] N. J. C. Lous, G. C. Hofmans, R. N. J. Veldhuis, A. Hirschberg: A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design. Acta Acustica 84 (1998) 1135–1150.

[15] I. R. Titze, J. W. Strong: Normal modes in vocal cord tissues. J. Acoust. Soc. Amer. 57 (1975) 736–744.

[16] C. Vilain: Contribution à la syntheèse de la parole par modèle physique. Thèse de Doctorat de l'Institut National Polytechnique de Grenoble, 2002.

[17] A. E. Rosenberg: Effect of glottal pulse shape on the quality of natural vowels. J. Acous. Soc. Am. 49 (1971) 583–590.

[18] G. Fant, J. Liljencrants, Q. Lin: A four parameter model of glottal flow. STL-QSPR 4 (1985) 1–13.

[19] D. Klatt, L. Klatt: Analysis, synthesis and perception of voice quality variations among female and male talkers. J. Acous. Soc. Am. 87 (1990) 820–857.

[20] P. H. Milenkovic: Voice source model for continuous control of pitch period. J. Acous. Soc. Am. 93 (1993) 1087–1096.

[21] D. G. Childers, T. H. Hu: Speech synthesis by glottal excited linear prediction. J. Acous. Soc. Am. 96 (1994) 2026–2036.

[22] D. G. Childers: Speech processing and synthesis toolboxes. John Wiley and Sons, New York, 2000.

[23] R. Husson: Physiologie de la phonation. Masson, Paris, 1962.

[24] D. G. Childers, D. M. Hicks, G. P. Moore, Y. A. Alsaka: A model for vocal fold vibratory motion, contact area, and the electroglottogram. J. Acoust. Soc. Am. 80 (1986) 1309–1320.

[25] G. Fant: Glottal source and excitation analysis. STL-QPSR, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 1979, 1, 85–107.

[26] G. Fant: The source filter concept in voice production. STL-QPSR, Speech, Music and Hearing, Royal Institute of Technology, Stockholm, 1981, 1, 21–37.

[27] A. Van Hirtum, I. Lopez, A. Hirschberg, X. Pelorson: On the relationship between input parameters in the two-mass vocal-fold model with acoustical coupling ans signal parameters in the glottal flow. Proc. Voice Quality: functions, analysis and synthesis (VOQUAL03) August 2003, Geneva, Swiss, 2003, 47–50.

[28] R. Laje, T. Gardner, G. B. Mindlin: The effect of feedback in the dynamics of the vocal folds. Phys. Rev. E 64 (2001) 056201.

[29] I. R. Titze: Principles of voice production. Prentice-Hall Inc., Englewood Cliffs, New York, 1994.

[30] J. Gauffin, J. Sundberg: Spectral correlates of glottal voice source waveform characteristics. Journal of Speech and Hearing Research 32 (1989) 556–565.

[31] G. Fant, A. Kruckenberg: Voice source properties of the speech code. TMH-QPSR 4/1996, 1996, 45–46.

[32] J. Sundberg, M. Andersson, C. Hulqvist: Effects of subglottal pressure variation on professional baritone singers' voice sources. J. Acoust. Soc. Am. 105 (1999) 1965–1971.

[33] D. Sciamarella, C. d'Alessandro: Reproducing laryngeal mechanisms with a two-mass model. European Conference on Speech Communication and Technology - Eurospeech, 2003.

[34] N. Henrich, C. d'Alessandro, M. Castelengo, B. Doval: Open quotient in speech and singing. Notes et documents LIMSI 2003-05, 2003, 1–19.

[35] H. Herzel: Bifurcation and chaos in voice signals. Appl. Mech. Rev. 46 (1993) 399–413.

[36] M. P. Karnell: Synchronized videostroboscopy and electroglottography. J. Voice 3 (1989) 68–75.

[37] M. H. Hess, M. Ludwigs: Strobophotoglottographic transillumination as a method for the analysis of vocal fold vibration patterns. J. Voice 14 (2000) 255–271.