# Reproducing laryngeal mechanisms with a two-mass model

Denisse Sciamarella, Christophe d'Alessandro

LIMSI-CNRS, BP 133, F91403, Orsay France

denisse@limsi.fr, cda@limsi.fr

# Abstract

Evidence is produced for the correspondence between the oscillation regimes of an up-to-date two-mass model and laryngeal mechanisms. Features presented by experimental electroglottographic signals during transition between laryngeal mechanisms are shown to be reproduced by the model.

# 1. Introduction

One of the main challenges in voice-production research has for long been the construction of a deterministic physical model which could describe the different mechanisms characterising vocal-fold motion as well as their acoustic correlate.

The essential features in the behavior of the glottal source have been reasonably described by a series of simplified vocalfold models which are apt for real time speech synthesis and which have followed and improved the pioneering 1972 Ishizaka and Flanagan's model [1]. In this kind of lumped models, self-sustained vocal-fold oscillations are attributed 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*.

A systematic acoustic study of the traditional Ishizaka and Flanagan's two-mass model has shown that vocal-fold dynamics could present distinct oscillation regimes which could be associated to what is known as laryngeal mechanisms of voice production [2]. A laryngeal mechanism is defined as a phonation mode associated to a particular glottal configuration, which is characterized by the effective mass, largeness and length of the vocal cords taking part in vibration, as well as by muscular tension. The question of laryngeal mechanism reproduction with simple vocal-fold models is of great importance in vocalfold modeling research, since it constitutes a well-known acoustic phenomenon in direct connection with vocal-fold motion.

This article revisits the problem of laryngeal mechanism reproduction, in the context of a recently proposed symmetrical two-mass model which includes an up-to-date aerodynamic description of the flow through the glottis. A brief description of the production model will be given in section 2. The method used to compute vocal-fold contact area from this production model will also be precised, since this will allow confrontation of the model predictions with certain features observed in experimental electroglottograms during transition between laryngeal mechanisms. The procedure used to identify different oscillation regimes is outlined in section 3. Results concerning oscillations regimes will be presented in section 4. Section 5 will concentrate on transitions between regimes. Conclusions will be presented in section 6.

# 2. The vocal-fold two-mass model

For vocal-fold motion simulation, we will be using the model implemented by Niels Lous *et al* in 1998 [3]. This two-mass model 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 largeness  $L_g$ . As usual, symmetry is assumed with respect to the flow channel axis. The flow channel height h(x, t) is a piecewise linear function (see figure 1) determined by  $h_{1,0}, h_{2,1}, h_{3,2}$ :



Figure 1: Geometrical structure of the symmetrical two-mass model.

$$h_{i,i-1}(x,t) = \frac{h_i(t) - h_{i-1}(t)}{x_i - x_{i-1}}(x - x_{i-1}) + h_{i-1}(t) \quad (1)$$

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

The main flow through the glottis is approximated by a quasi-stationary, inviscid, locally incompressible, and quasiparallel flow from the trachea up to a point  $x_s$  where the flow separates from the wall to form a free jet. The turbulence in the free jet introduces the necessary dissipation to obtain the glottal flow modulation caused by the movement of the vocal folds. Further details concerning the computation of  $x_s$  from a simple geometrical flow-separation model can be found in [3]. In figure 1,  $h_c$  is the critical height at which mechanical contact is assumed to take place. Actual contact between vocal cords takes place when  $h_{2,1}(x,t) \leq 0$ . Thus, the instantaneous contact area during vocal-fold motion simulation can be computed as follows:

$$a(t) = L_g . x_c(t) \tag{2}$$

where  $x_c(t)$  is the distance along which  $h_{2,1}(x, t) \leq 0$ . Contact area computation is an important result among the model predictions since experimental electroglottograms are conjectured to be related to the contact area between the folds [4].

Vocal-fold dynamics depends on inertia, elasticity and damping. 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 relevant quantities for glottal flow production in the case of a symmetrical glottal structure (no asymmetry is assumed between upstream and downstream masses) amount to seven control parameters: the effective largeness  $L_g$ , length d and mass m of the vocal cords, the linear springs constant k attached to the masses, the spring constant  $k_c$  coupling both masses, the subglottal pressure  $P_s$ , and the mass damping coefficient  $\zeta$ . Control parameters  $\zeta$  and k undergo a stepwise increase on vocal-fold collision which is assumed to be invariant. Default values of control parameters at a typical glottal condition are:  $d \approx 0.2 \ cm$ ,  $m \approx 0.1 \ g$ ,  $k \approx 40 \ N/m$ ,  $k_c \approx 25 \ N/m$ ,  $L_g \approx 1.4 \ cm$  and  $\zeta \approx 0.1$ ,  $P_s = 8 \ cmH_2O$ .

Throughout this work, simulations are performed with a viscous flow correction that slows down opening and closure of the folds so that spurious discontinuities (or "clicks") in the glottal-flow derivative are eliminated. The glottal-flow pulse  $U_g(t)$  and its time derivative generated by the above-described model at the typical glottal condition and without acoustic coupling with the supraglottal system are shown in figure 2.

Acoustic coupling of the glottal model to vocal-tract and the consequent feedback on the glottal-flow signal (which produces formant ripples in  $U_g(t)$  and its time derivative) will not be considered in this paper. Systematic acoustic studies of twomass models with and without vocal-tract acoustic coupling have shown that vocal-fold motion simulation in the absence of vocal-tract constitutes an ideal scenario to detect distinct oscillation regimes of the dynamical system of voice production [2] [5]. Besides, waveforms of glottal area and fundamental frequency are almost independent of the vocal-tract shape.



Figure 2: (a) Glottal volume velocity in  $cm^3/s$  for the uncoupled Niels Lous model with viscous flow correction. (b) Glottal flow derivative in  $m^3/s^2$  corresponding to (a).

#### 2.1. Laryngeal mechanisms in terms of control parameters

Laryngeal mechanisms are usually defined in terms of glottal configuration and muscular tension. In the presence of a production 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 eventually  $k_c$ .

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 to 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-cords become tense, slim and short. This vibration mode differs from  $M_I$  in aspects regarding glottal configuration, muscular tension and glottal closure. The reason for the reduction in the largeness of the folds that participates in vibration is an accentuated compression between the arytenoids. On the other hand, vibration in  $M_{II}$  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 small values of m, d and  $L_g$ , while k and  $k_c$  are considerably higher.

In view of this description, 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 physical model is sound enough. Whether the glottal-flow signals generated with the model effectively correspond to phonation in a certain mechanism is a question that will be discussed in the following sections.

# 2.2. Laryngeal mechanisms in terms of acoustic parameters

Laryngeal mechanisms can be charaterized in terms of acoustic parameters. As fundamental frequency  $F_0$  is increased, one can notice a voice break corresponding to the change between  $M_I$ and  $M_{II}$ . Generally,  $M_I$  corresponds to lower values of  $F_0$ , a low open quotient, and a stronger intensity.  $M_{II}$  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.

### 3. Identifying oscillation regimes

Identifying oscillation regimes of a vocal-fold model during the variation of one or many control parameters is quite simple. Within the range of allowed oscillations, the glottal-flow signal  $U_g(t)$  and its derivative  $U'_g(t)$  are positive-definite periodic functions, which are continuous and derivable except maybe at the opening and closure instants. Each cycle presents a shape qualitatively close to that of waveforms shown in figure 2. Close to the boundary of an allowed oscillation region, either periodicity or glottal pulse shape is altered. We say then that glottal-flow is irregular. Based on these simple criteria, a C + + algorithm has been written to detect oscillation regimes.

The program has two functioning modes, both comprising vocal-fold motion simulation. The first mode iteratively sweeps a selected range of a number of control parameters and computes within each iteration  $U_g(t)$ ,  $U'_g(t)$  and a(t) during a voicing time interval  $\Delta t$  which greatly exceeds the build-up time required for the oscillation to settle to a steady state. The glottal volume velocity is inspected backwards to establish the fundamental frequency  $F_0$  of the signal. If there are long-term irregularities, the control parameters responsible for them are stored in a separate file. Next, a sample of the glottal-flow period is isolated and the number and nature of the local extrema in  $U'_{a}(t)$ are examined in order to verify if the pulse is regular. If it is, the code checks for glottal leakage: if  $U_q(t) \neq 0$  at the sample borders, control parameters are supposed to induce glottal leakage and are grouped in a different file from those inducing complete complete closure. The fundamental frequency for the

glottal-flow signal is always saved as a function of control parameters. Further details concerning the algorithm procedure may be found in [2] and [5].

The second mode allows to vary a given control parameter in one single simulation. An output sound file allows to listen to the variation, and a data file permits visual analysis of the evolution of glottal flow and contact area, the latter being particularly relevant during the transition between mechanisms.

# 4. Laryngeal mechanisms and oscillation regimes

In this section we examine whether the symmetrical two-mass model described in section 1 is capable of reproducing different laryngeal mechanisms.

Our 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 3 shows parameter space for some of these physical parameters, in which we encounter two distinct regions within which regular vocal-fold oscillations take place. In these examples, the dark regions correspond to signals with glottal leakage, while the clearer regions correspond to signals with complete glottal closure. Notice that within a single region in parameter space, the variation of fundamental frequency is smooth.



Figure 3: 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. Dark points correspond to signals with incomplete glottal closure, i.e. glottal leakage.

Regimes with glottal leakage sistematically 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 regimes with glottal leakage and  $M_{II}$ .

Distinct oscillation regions may also appear for oscillations without glottal leakage. An example is shown in figure 4 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 4, an identification of this oscillation regime with  $M_0$  is not possible since the correspondent glottalflow signals do not present a sufficiently short open phase. Attempts of a simultaneous lowering of k and  $L_g$  as d is increased (with respect to the typical glottal condition) have not allowed us to find oscillation regimes ressembling  $M_0$ , which is in theory described by a physiological action of this kind.



Figure 4: Parameter space and variation of  $F_0$  for m and  $P_s$ . The black points correspond to the lowest values of  $F_0$ .

# 5. Transition between regimes

### 5.1. The nature of the transition

The transition from one regime to another is generally marked by a jump in fundamental frequency. Consider figure 3 and notice that moving from the clear to the dark regions involves a jump in  $F_0$ . However, remark 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 frequency. In the upper right corner of (c), for instance, or in the lower left corner of (a), it is possible to pass from the clear to the dark 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, i.e. to a sudden qualitative change in the behavior of the system during a smooth variation of control parameters.

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 [6]. Concerning this open question, 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 increase  $F_0$ . If this is true, the degree of training of a speaker in the control of his vocal apparatus may result in different physiological solutions to produce a desired effect (such as an increase in  $F_0$ ).

### 5.2. Transitions and electroglottographic signals

The use of electroglottographic signals (EGG) to measure  $F_0$ and the open quotient, independently of the effects of the supraglottal system, is an interesting alternative to other traditional but indirect methods such as inverse filtering. This techinque consists in passing a high frequency electric signal between two electrodes positioned at two different locations on the neck. Tissues in the neck act as conductors whereas airspace narrows the conducting path. The EGG gives thus an indication of the sealing of the glottis.

In terms of vocal-fold models, the electroglottogram has been conjectured to be related to the contact area between the cords. The traditional two-mass model does not allow calculation of contact area because the projected area in this model is always rectangular and there is no gradation in opening or closing [7]. Instead, the vocal-fold geometry depicted in figure 1, admits a gradual variation of contact area in time, which is given by a(t) in equation 2.

Henrich [8] reports the existence of peak doubling in experimental DEGG signals (a'(t)) during the transition between the first and second laryngeal mechanisms. Our numerical simulations show that this phenomenon can be effectively reproduced when a transition between oscillation regimes is occuring. As an example, figure 5 shows a cycle of a(t) and its derivative a'(t) before and during the transition between the clear and dark regions depicted in figure 3(c). Just as observed in experimental DEGG signals, a'(t) presents an opening peak doubling, which constitutes another element in favour of interpretation of oscillation regimes in terms of laryngeal mechanisms.



Figure 5: EGG and DEGG signals generated by vocal-fold motion simulation with the Niels Lous model (a) before the transition and (b) during the transition between the clear and dark regions of figure 3(c) at  $P_s = 450 \ cmH_2O$ .

# 6. Conclusions

Laryngeal mechanisms constitute a fundamental aspect of vocal production for several reasons. In the first place, they are essential to account for the wide range of phonation frequencies covered by the human vocal apparatus. Secondly, their characterisation in terms of glottal configuration provides an indication of the vibration modes that a suitable vocal production model should consider, and of mode activation in terms of control parameters. Finally, laryngeal mechanisms admit a precise acoustic description which is ideal for validation of the model through numerical simulation of glottal flow.

This work revisits the problem of laryngeal mechanism reproduction with a voice production two-mass model. The model proposed by Niels Lous *et al* [3] in 1998 constitutes a new testbench for the acoustic response of a simple deterministic production model based on an aerodynamic description of the flow through the glottis comprising flow separation in its divering part.

An algorithm that allows detection of oscillation regimes via glottal-flow simulation and inspection has been proposed. Distinct oscillation regimes are detected, and evidence is produced for their identification with the first and second laryngeal mechanisms, which are the most common mechanisms used in human phonation. Nevertheless, identification of lowfrequency oscillation regimes with  $M_0$  or vocal fry has not been possible, at least as far as a symmetrical glottal structure is assumed.

Transitions between oscillation regimes are shown to share features experimentally observed for transitions between laryngeal mechanisms. The peak doubling present in experimental electroglottographic signals during such transitions has been reproduced using contact area functions generated with the production model. Such a result constitutes further evidence for the identification of laryngeal mechanisms with oscillation regimes. In this context, the evolution of control parameters during transitions provides a physiological interpretation of the transition between mechanisms. Simulations show that the change from a regime to another may be attributed either to a sudden change in the activity of the muscles (as initially supposed for transitions between different laryngeal mechanisms) or to underlying bifurcations of the dynamical system, inducing a sudden change in the behavior of the source in the course of a smooth variation of control parameters (as recently proposed). Which of the two options effectively occurs will depend on the choice and magnitude of the parameters used by the speaker to control the transition.

# 7. Acknowledgements

The authours would like to thank Nathalie Henrich, Coriandre Vilain and Avraham Hirschberg for useful discussions.

# 8. References

- K. Ishizaka and J. L. Flanagan, Synthesis of Voiced Sounds form a two-mass model of the vocal cords. Bell. Syst. Tech. J., 51, pp. 1233-1268 (1972).
- [2] D. Sciamarella and C. d'Alessandro: "A study of the Two-Mass Model in terms of Acoustic parameters", Proc. IC-SLP pp.2313-2316 (2002).
- [3] 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 1135-1150 (1998)
- [4] I. R. Titze: "Interpretation of the Electroglottographic Signal" Journal of Voice (4) 1, pp.1-9 (1990)
- [5] D. Sciamarella and C. d'Alessandro: "An acoustic analysis of a symmetrical two-mass model" (to be published)
- [6] Herzel, Bifurcations and chaos in voice signals. App. Mech. Rev. 46 (7) (1993).
- [7] D. G. Childers et al: "A model for vocal fold vibratory motion, contact area, and the electroglottogram" J. Acoust. Soc. Am. 80 (5) 1309-1320 (1986)
- [8] N. Henrich, Thèse de Doctorat de l'Université Paris 6. Etude de la source glottique en voix parlée et chantée (2001)