Asociación Argentina de Mecánica Computacional

Mecánica Computacional Vol XXX, págs. 3331-3338 (artículo completo)
Oscar Möller, Javier W. Signorelli, Mario A. Storti (Eds.)
Rosario, Argentina, 1-4 Noviembre 2011

ESTIMATION OF A PARAMETER OF A NON-LINEAR STOCHASTIC MODEL FOR THE VOCAL FOLDS THROUGH A MODIFIED MCMC ALGORITHM

Julien Mauprivez[a], Edson Cataldo[b] and Rubens Sampaio[a]

[a] *Mechanical Engineering Department – PUC-Rio, Rua Marquês de São Vicente, 225, Gávea, RJ, CEP: 22453-900, Brazil, e-mail: mobmaup@gmail.com, http://www.mec.puc-rio.br/*

[b] *Applied Mathematics Department – Universidade Federal Fluminense, Graduate Program in Telecommunications Engineering, Rua M'ario Santos Braga, S/N, Centro, Niteroi, CEP: 24020-140, RJ, Brazil, http://www.uff.br/gma/*

Keywords: Inverse model, vocal folds model, stochastic mechanics.

Abstract. Low order non-linear mechanical models for vocal folds, in the phonation process, have been shown to be useful in the case of normal and disordered voice studies. Despite their relative simplicity, they are able to simulate the main features of the vocal fold dynamics. A good example is the so-called two-mass Lous model, which uses few input parameters and has shown excellent results in understanding phonation phenomena. However, to model a real voice, it is required to infer a set of parameters of the model. Recently, some authors pointed out the advantage of using probabilistic approaches to characterize vocal fold dynamics. In this paper, a numerical stochastic model for voice production is used to simulate several vowel utterances. Then, the vocal fold tension probability density function is considered unknown and estimated from vowel utterances, using a Monte Carlo Markov Chain. Results show a good match between the estimated and actual probability densities.

1 INTRODUCTION

The physical process responsible for voice production involves various phenomena such as turbulence, vibration of biological structures, and aero-acoustical couplings, which can be modeled and simulated in details, but at a high computational cost (Cook et al. (2009)). A good compromise between computational cost and accuracy is obtained using low-order models, such as Lous et al. (1998) model, and referred to as Lous model in the present paper. A discussion about the accuracy of the model can be found in Ruty et al. (2007).

In Sec. 2, Lous model for vowel utterance production is described together with its stochastic reformulation inspired by a recent paper by Cataldo et al. (2009). As no experimental data are available for random realizations of vowel utterances, the numerical model described in Sec. 2 is used to build a consistent set of data. This numerical model consists of generating independent realizations of the vocal fold tension, and simulating the corresponding set of voice utterances, which is a stochastic process.

Section 3 is devoted to the description of a modified Metropolis-Hastings Monte Carlo Markov Chain (M-H MCMC). The algorithm herein described aims to infer the supposedly unknown probability density function (p.d.f.) of the tension factor, based on the observable p.d.f. of the voice fundamental frequency.

Finally, an application, is discussed together with qualitative results in Sec. 4.

2 DETERMINISTIC AND ASSOCIATED STOCHASTIC MODEL OF THE PHONATORY SYSTEM FOR VOWEL PRODUCTION

2.1 Deterministic model

The two-mass model used here is the one created by Lous et al. (1998), and was constructed after Ishizaka and Flanagan (1972) model, considering an improved description of the airflow.

The complete model is composed of two coupled subsystems: one subsystem modeling the vocal folds, which is called *source*, and one subsystem modeling the vocal tract, which is called *filter*.

The *source* subsystem is composed by two mass-spring-damper oscillators coupled by a linear spring, as represented in Fig. 1.

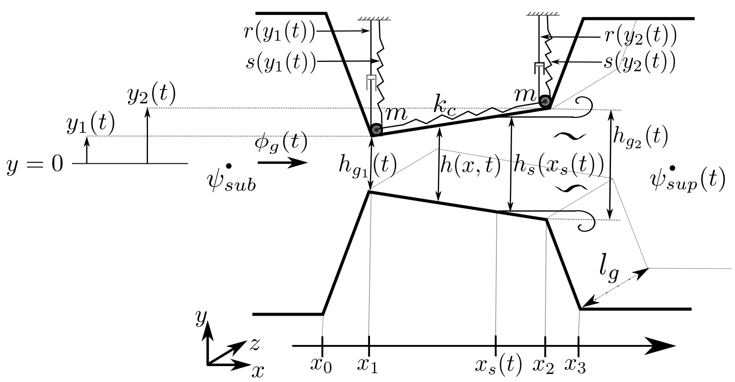

Figure 1: Schematic representation of Lous model.

The geometry of the space between the two plates representing the vocal folds is described by three quantities: the glottal height at point x at the instant t ($h(x,t)$), the glottal depth (l_g) towards z-direction, and the glottal length (given by the distance $x_3 - x_0$). The glottal flow is denoted by $\phi_g(t)$ and the dynamics of the vocal folds are given by Eq. 1:

$$\begin{cases} m\frac{d^2y_1(t)}{dt^2} + r(y_1(t))\frac{dy_1(t)}{dt} + s(y_1(t))y_1(t) \\ \quad +k_c(y_1(t) - y_2(t)) = f_1(p_{sub}, p_{sup}(t), h_{g_1}(t), h_{g_2}(t)) \\ \\ m\frac{d^2y_2(t)}{dt^2} + r(y_2(t))\frac{dy_2(t)}{dt} + s(y_2(t))y_2(t) \\ \quad +k_c(y_2(t) - y_1(t)) = f_2(p_{sub}, p_{sup}(t), h_{g_1}(t), h_{g_2}(t)) \end{cases} \tag{1}$$

where $h_{g_1}(t)$ and $h_{g_2}(t)$ are the glottal heights, $y_1(t)$ and $y_2(t)$ are the corresponding position for masses 1 and 2 relative to their rest position, $r(y_{1,2}(t))$ and $s(y_{1,2}(t))$ are, respectively, the damping and stiffness functions, which will be described later. k_c is the stiffness constant of the linear spring which couples the two mass-spring-damper systems. f_1 and f_2 are the forces applied to the vocal folds due to the pressure field in the glottis and the acoustic pressure at the vocal tract inlet, as defined later. The sub-glottal pressure is given by the constant ψ_{sub} and the supra-glottal pressure is given by the function $\psi_{sup}(t)$. The displacement of each vocal fold is considered to be perpendicular to the direction of the airflow.

The elasticity and damping functions ($s(y_{1,2}(t))$ and $r(y_{1,2}(t))$) are piecewise linear functions of the position of the vocal folds and they take into account the vocal-folds collision.

The elasticity function is given by Eq. 2:

$$s(y_i(t)) = \begin{cases} ky_i(t) & , \quad h_{g_i}(t) > h_{lim}, \\ (k+3k)y_i(t) & , \quad h_{g_i}(t) \leqslant h_{lim}. \end{cases} \quad i = 1, 2. \tag{2}$$

where k is a constant. As proposed in Pelorson et al. (1994), the contact between the vocal folds occurs before the eventual full glottis closure, considering contact for $h_{g_{1,2}}(t) \leqslant h_{lim}$, where h_{lim} is a positive constant.

The damping function is given by Eq. 3:

$$r(y_i(t)) = \begin{cases} 2\xi\sqrt{mk} & , \quad h_{g_i}(t) > h_{lim}, \\ 2(\xi+1)\sqrt{mk} & , \quad h_{g_i}(t) \leqslant h_{lim}. \end{cases} \quad i = 1, 2. \tag{3}$$

where ξ is the damping factor of the oscillators and is constant.

The fundamental frequency of the vocal folds is very sensitive to the variation of the vocal fold tension factor q (Ishizaka and Flanagan (1972),Cataldo et al. (2009)), related to the stiffness and mass associated to the mass-spring-damper systems modeling the vocal folds (respectively k and m). The values of mass and stiffness to be used are defined by $m = \frac{\widehat{m}}{q}$, and $k = q\widehat{k}$. Then, an increase in q causes a diminution of the mass participating of the vibration m and an increase of the stiffness k. This tension factor were introduced in Ishizaka and Flanagan (1972) to simulated the diminution of the active mass during vocal fold vibration as the stiffness increases.

The airflow through the glottis is assumed to be quasi-steady, incompressible, and unidimensional (along the x axis Pelorson et al. (1994)). As suggested in Lous et al. (1998), the viscosity and fluid inertia are approximated by adding an inertive term and a Poiseuille term to Bernoulli equation. The pressure distribution along the glottis, denoted by $\psi(x,t)$, can be described by the modified Bernoulli's energy equation and it is given by Eq.4:

$$\psi(x,t) = \begin{cases} \psi_{sub} - \frac{\rho}{2}\left(\frac{\phi_g(t)}{l_g(h(x,t)-h_{sub})}\right)^2 \\ \quad -12\mu l_g^2 \phi_g(t) \int_{x_0}^{x} \frac{1}{l_g^3 h^3(x,t)} dx & , x < x_s \\ \quad -\rho \frac{d\phi_g(t)}{dt} \int_{x_0}^{x} \frac{1}{l_g h(x,t)} dx \\ \psi_{sup}(t) & , x \geqslant x_s \end{cases} \tag{4}$$

where $\phi_g(t)$ is the volumic flow inside the glottis, x_s the position of detachment of a free jet from the vocal folds, h_{sub} the height at x_0, ρ the density of air and μ the air dynamic viscosity. The position of the free jet detachment is defined by Eq.5.

$$h_s(t) = \min(\alpha h_{g_1}(t), h_{g_2}(t)). \tag{5}$$

The value of α is set to 1.1 Lous et al. (1998), when viscous and inertive terms are considered.

The force applied on the oscillators at $y_1(t)$ and $y_2(t)$, due to the pressure field, considering only the component normal to the plates, is given by Eq.6:

$$f_i(t) = \int_{x_{i-1}}^{x_i} \left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) \psi(x,t)dx + \int_{x_i}^{x_{i+1}} \left(\frac{x_{i+1} - x}{x_{i+1} - x_i}\right) \psi(x,t)dx, \tag{6}$$

where $i = 1, 2$.

The *filter* subsystem, coupled to the *source* subsystem, is described as an acoustic tube with variable cross section, having for input the supraglottal pressure $\psi_{sup}(t)$ and for output $\psi_r(t)$. For the sake of simplicity, the vocal tract is represented by a concatenation of cylindrical tubes and at the end of the last acoustic tube a radiation load equivalent to that of a disc in an infinite plane is imposed (Fant (1960)).

For the whole system (*source* + *filter*), the input is the subglottal pressure ψ_{sub}, which is constant, and the output is the function $\psi_r(t)$, the pressure at the lips. Details about Lous model can be found in Lous et al. (1998).

2.2 Stochastic model

The present section briefly presents the methodology used to generate the testing data sets. Measurement of the mechanical parameters of the vocal folds for various independent vowel utterance were not available, neither in-vitro nor in-vivo, to us. Then, for practical reasons, all of the data presented in is this work are simulated numerically.

Recently, Cataldo et al. (2009) discussed the uncertainties of some parameters in Ishizaka and Flanagan's model and probability density functions were constructed for them, using the Maximum Entropy Principle. Herein, the same idea is used, but applied to the Lous model, for parameter q.

The random variable Q is then associated with the parameter q, and, the corresponding stochastic model is constructed by substituting in the deterministic Lous model.

The p.d.f $p_Q(q)$ of the random variable Q has to verify the constraints given in Cataldo et al. (2009).
Applying the Maximum Entropy Principle yields to the gamma probability density function given by Eq. 7:

$$p_Q(q) = \mathbf{1}_{]0,+\infty[}(q)\frac{1}{\underline{Q}}\left(\frac{1}{\delta_Q^2}\right)^{\frac{1}{\delta_Q^2}} \times$$
$$\times\frac{1}{\Gamma\left(1/\delta_Q^2\right)}\left(\frac{q}{\underline{Q}}\right)^{\frac{1}{\delta_Q^2}-1} exp\left(-\frac{q}{\delta_Q^2\underline{Q}}\right) \quad (7)$$

where the positive parameter $\delta_Q = \sigma_Q/\underline{Q}$ is the relative deviation of the random variable Q such that $\delta_Q < 1/\sqrt{2}$ and where σ_Q is the standard deviation of Q.

The Gamma function Γ is defined by $\Gamma(\alpha) = \displaystyle\int_0^{+\infty} t^{\alpha-1}e^{-t}dt$.

To generate the realizations of the output radiated pressure, the Monte Carlo Method is used. First, independent realizations $q(\theta_i)$ of the random variable Q are generated using the probability density function defined by Eq. 7. For each realization $Q(\theta)$ of the random variable Q, a realization of the output acoustic pressure, $\Psi_r(t,\theta)$, is calculated using the Lous model equations.

3 INFERENCE OF THE TENSION'S FACTOR P.D.F.

The stochastic process $\Psi_r(t)$ is represented in this work as a non-linear mapping of a random variable Q, by Lous model. Due to this non-linear mapping, the analytical resolution of the inverse problem, i.e. the inference of the p.d.f of Q given $\Psi_r(t)$, is intractable. Nevertheless, using numerical approximations, an estimate of this p.d.f can be obtained.

To infer the p.d.f of Q from $\Psi_r(t)$, the voice signal is first parametrized. For each realization $\Psi_r(t,\theta)$ of $\Psi_r(t)$, the corresponding realization of the fundamental frequency $F_0(\theta)$ is calculated. Then, the equations shown in sections 2.1 and 2.2 are used to map from Q to F_0 through a nonlinear function g(.),

$$F_0 = g(Q). \quad (8)$$

As the analytical inverse mapping $Q = g^{-1}(F_0)$ is not available, a Metropolis-Hastings Markov Chain Monte Carlo (M-H MCMC) algorithm (Chib and Greenberg (1995)) is implemented to infer Q from F_0.

It is supposed that $p_{F_0}(f_0^t)$, the p.d.f. of the target fundamental frequency is known. Given the experimental distribution for F_0, $p_{F_0}(f_0^t)$, and the mapping $F_0 = g(Q)$, through Lous model, the Metropolis-Hastings algorithm, for a fixed iteration number N, can be implemented as follows:

1. Choose $Q(\theta_0)$, and find by simulation $F_0(\theta_0) = g(Q(\theta_0))$,
2. set $i = 1$,
3. at step i, generate a candidate $Q(\theta_c)$ from the internal Markov kernel $\rho(.|Q(\theta_{i-1}))$, and find by simulation $F_0(\theta_c) = g(Q(\theta_c))$,

4. compute $\alpha = \frac{p(g(Q(\theta_c)))\rho(Q(\theta_{i-1})|Q(\theta_c))}{p(g(Q(\theta_{i-1})))\rho(Q(\theta_c)|Q(\theta_{i-1}))} = \frac{p(F_0(\theta_c))\rho(Q(\theta_{i-1})|Q(\theta_c))}{p(F_0(\theta_{i-1}))\rho(Q(\theta_c)|Q(\theta_{i-1}))}$,

5. set $Q(\theta_i) = Q(\theta_c)$ with probability $argmin(1, \alpha)$, else set $Q(\theta_i) = Q(\theta_{i-1})$ with probability $1 - \alpha$,

6. set $i = i + 1$,

7. if $i \leq N$ return to 3.

The internal Markov kernel $\rho(.|Q(\theta_{i-1}))$ causes the support of the posterior distribution to be progressively explored. It is worth noting that $\rho(.|Q(\theta_{i-1}))$ should be chosen so that the candidate $Q(\theta_c)$ can effectively explore the whole support of the posterior distribution.
Due to steps 4 and 5, values of $Q(\theta_i)$ that maps to more likely $F_0(\theta_i)$ are chosen with higher probability than those mapping to less likely values.

By choosing a symmetric probability function for the transition kernel, the acceptance rate α can be simplified to Eq.9.

$$\alpha = \frac{p(g(Q(\theta_c)))}{p(g(Q(\theta_{i-1})))} = \frac{p(F_0(\theta_c))}{p(F_0(\theta_{i-1}))}. \tag{9}$$

In this work, the acceptance rate have a uniform p.d.f. of mean $Q(\theta_{i-1})$ and support 2σ.

$$\rho(Q(\theta_c)|Q(\theta_{i-1})) = U(Q(\theta_{i-1}) - \sigma, Q(\theta_{i-1}) + \sigma), \tag{10}$$

4 APPLICATION AND DISCUSSION

A target set Q is generated from the p.d.f given by Eq.7 and the statistics presented in Tab.1.

Table 1: Summary of the statistics of the target set.

Parameter	Mean	Relative deviation	Number of realizations
Q	0.94	2.5%	500

From these realizations, and using the forward Lous model, the associated target p.d.f. $p_{F_0}(f_0^t)$ for F_0 is simulated.

During the simulation, the set $p_{F_0}(f_0^t)$ is related to the observable data. The p.d.f. $p_Q(q^t)$, used to generate the target p.d.f. $p_{F_0}(f_0^t)$, is only used for comparison with the estimated distribution, i.e. to validate the algorithm.

The choice of the transition kernel is of great importance, in order to work with a reasonable computational cost (Gilks et al. (1996)). Its probability density should be chosen so that the candidate $Q(\theta_c)$ can effectively explore the whole support of the posterior distribution.
Another important parameter is its support 2σ. If it is too big, the support of the posterior distribution is quickly explored, at cost of the rejection of many candidates. On the other hand, using a small σ, most of the candidates $Q(\theta_c)$ will be accepted, nevertheless, few will sample the regions of low probability.

In the following application the starting point of the chain is $\mathbf{Q}(\theta_0) = .\mathbf{84}$. This value was set by running the deterministic model for different tension factors $Q(\theta_0)$ until the simulation of a fundamental frequency $F_0(\theta_0)$ of reasonable probability, given $p_{F_0}(f_0^t)$. The support of the kernel density is $\mathbf{2}\sigma = \mathbf{0.30}$.

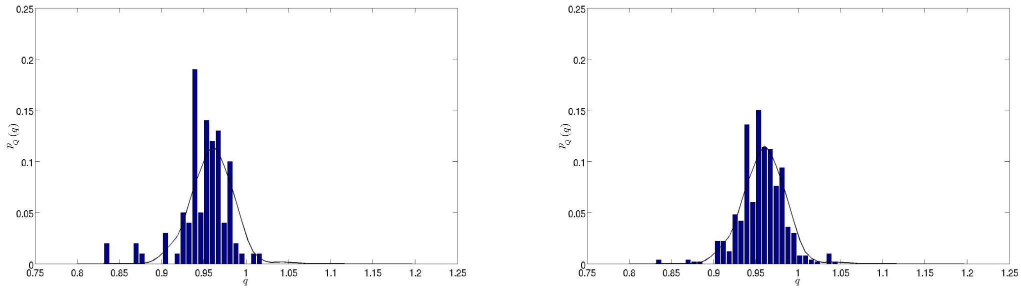

Figure 2: Posterior (bars) and target (line) $p_Q(q)$ for: $n = 100$ (left), and $n = 500$ (right)

Figure 2 shows how the sampling is concentrated in the high probability region during the first 500 iterations. Very few realizations are sampled from the tails of the distribution.

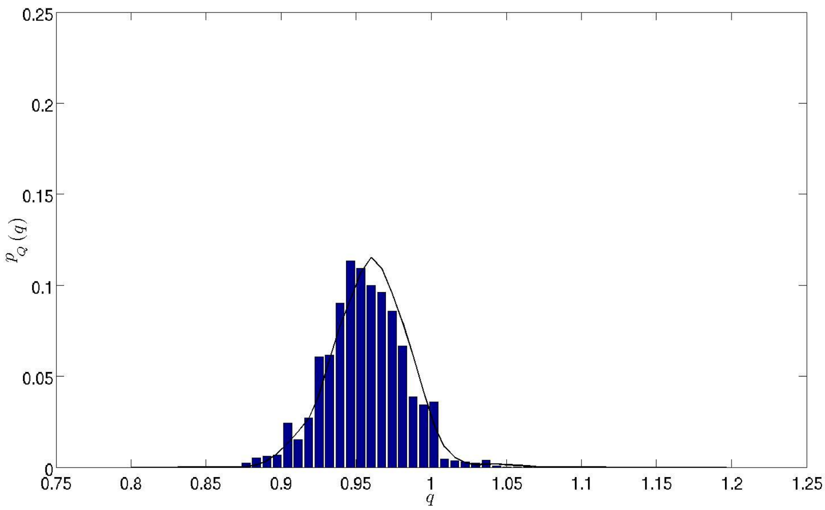

Figure 3: Posterior (bars) and target (line) histograms for $p_Q(q)$ for $n = 4000$

After 4000 steps, the right tail remains almost unexplored, suggesting σ is too small (0.15). Nevertheless, as shown on Fig.3, a good match is obtained between the target and simulated distributions.

5 CONCLUSION

The inverse mapping of a stochastic non-linear model have been implemented using Metropolis-Hastings Monte Carlo Markov Chain algorithm. Very satisfying results were obtained for the estimation of the vocal fold tension probability density function when compared to the actual

one.
Nevertheless, the probability of the values lying in the tails of the p.d.f. are not well inferred, suggesting that the arbitrarily chosen number of iterations or properties of the Markov Kernel are not optimal.
As a next step, to be less dependent of arbitrarily chosen variables, an implementation of a more recent algorithm such as Sequential Monte Carlo algorithm is being made.

ACKNOWLEDGEMENTS

The authors acknowledge FAPERJ (Fundação de Amparo à Pesquisa no Rio de Janeiro, CAPES (CAPES/COFECUB project N. 672/10) and CNPq (Brazilian Agency: Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the financial support they gave to this research.

REFERENCES

Cataldo E., Soize C., Sampaio R., and Desceliers C. Probabilistic modeling of a nonlinear dynamical system used for producing voice. *Computational Mechanics*, 43(2):265–275, 2009.

Chib S. and Greenberg E. Understanding the metropolis-hastings algorithm. *American Statistician*, pages 327–335, 1995.

Cook D., Nauman E., and Mongeau L. Ranking vocal fold model parameters by their influence on modal frequencies. *The Journal of the Acoustical Society of America*, 126(4):2002, 2009.

Fant G. *Acoustic Theory of Speech Production with Calculations Based on X-ray Studies of Russian Articulations*. Mouton, The Hague, 1960.

Gilks W., Richardson S., and Spiegelhalter D. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1996.

Ishizaka K. and Flanagan J. Synthesis of Voiced Sounds from a Two-Mass Model of the Vocal Cords. *Bell System Tech. J. SI*, pages 1233–1268, 1972.

Lous N., Hofmans G., Veldhuis R., and Hirschberg A. A Symmetrical Two-Mass Vocal-Fold Model Coupled to Vocal Tract and Trachea, with Application to Prosthesis Design. *Acta Acustica united with Acustica*, 84(6):1135–1150, 1998.

Pelorson X., Hirschberg A., van Hassel R., Wijnands A., and Auregan Y. Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model. *The Journal of the Acoustical Society of America*, 96:3416, 1994.

Ruty N., Pelorson X., Van Hirtum A., Lopez-Arteaga I., and Hirschberg A. An in vitro setup to test the relevance and the accuracy of low-order vocal folds models. *The Journal of the Acoustical Society of America*, 121:479, 2007.