Stochastic mechanical model of vocal folds for producing jitter and for identifying pathologies through real voices

Edson Cataldo, Christian Soize

Stochastic mechanical model of vocal folds for producing jitter and for identifying pathologies through real voices

E. Cataldo[a], C. Soize[b]

[a] *Universidade Federal Fluminense, Graduate program in Electrical and Telecommunications Engineering, Rua Mário Santos Braga, S/N, Centro, Niterói, RJ, CEP: 24020-140, Brazil*
[b] *Université Paris-Est, Laboratoire Modélisation et Simulation Multi Echelle, MSME UMR 8208 CNRS, 5 Bd Descartes, 77454 Marne-La-Vallée, France*

Abstract

Jitter, in voice production applications, is a random phenomenon characterized by the deviation of the glottal cycle length with respect to a mean value. Its study can help in identifying pathologies related to the vocal folds according to the values obtained through the different ways to measure it. This paper aims to propose a stochastic model, considering three control parameters, to generate jitter based on a deterministic one-mass model for the dynamics of the vocal folds and to identify parameters from the stochastic model taking into account real voice signals experimentally obtained. To solve the corresponding stochastic inverse problem, the cost function used is based on the distance between probability density functions of the random variables associated with the fundamental frequencies obtained by the experimental voices and the simulated ones, and also on the distance between features extracted from the voice signals, simulated and experimental, to calculate jitter. The results obtained show that the model proposed is valid and some samples of voices are synthesized considering the identified parameters for normal and pathological cases. The strategy adopted is also a novelty and mainly because a solution was obtained. In addition to the use of three parameters to construct the model of jitter, is the discussion of a parameter related to the bandwidth of the power spectral density function of the stochastic process to measure the quality of the signal generated. A study about the influence of all the main parameters is also performed. The identification of the parameters of the model considering pathological cases is maybe of all novelties introduced by the paper the most interesting.

Keywords: Voice production, jitter, stochastic mechanical models, experimental identification, statistical inverse problem.

Email addresses: ecataldo@im.uff.br (E. Cataldo), christian.soize@univ-paris-est.fr (C. Soize)

1. Introduction

The production of a voiced sound starts when the airflow coming from the lungs is modified into the glottal signal, a quasi-periodic signal after passing through the glottis, where the vocal folds are located. The main examples of voiced sounds are the vowels and this paper is based on their production.

The acoustic pressure signal, after passing by the vocal folds, is filtered and amplified by the vocal tract and then radiated by the mouth originating the voice signal. As the vocal folds displacements are not exactly symmetric the time intervals corresponding to the air pulses of the glottal signal have random fluctuations, called jitter.

There are different ways to measure jitter and its study is important to identify irregularities on the phonation. The values of jitter considered to a normal voice is between 0.1% and, at the maximum, 1% in relation to the mean of the time glottal intervals. Other acoustic measures can also be used, as Shimmer and HNR (Ratio Harmonic-Noise), to help in identifying pathologies on the vocal folds, vocal aging or even to help in problems of speaker recognition or stress situations related to the voice. However, the main feature that should be considered is jitter (Wong, 1991; Jiang et al., 2009; Dejonckerea et al., 2012; Mongia and Sharma, 2014; Silva et al., 2016) and this paper is focused in its generation.

Some models of jitter have been proposed but, in general, they do not consider mechanical models, they are created directly on the voice signals, considering some perturbations as, for example, a controlled noise (Schoengten et al. 1997, 2013).

Some mechanical models of jitter have been proposed by the same authors of this paper (Cataldo et al., 2012; Cataldo and Soize, 2016, 2017) and, now, a new mechanical stochastic model is then proposed but considering three control parameters, which gives more possibilities to generate jitter, including a way to change the quality of the voice generated. A new parameter is introduced to discuss this quality, related to the bandwidth of the power spectral density function and, mainly, an inverse stochastic problem is solved to identify parameters and, consequently, to validate the model proposed. With these new possibilities, specific pathologies of the vocal folds can be created and identified, such as paralysis of the vocal folds.

The stochastic model proposed here has the origin based on the deterministic model created by Flanagan and Landgraf (1968), known as the first model used to generate voice using a nonlinear one-mass mechanical model. More complete deterministic models were created (Ishizaka and Flanagan, 1972; Avanzini, 2008; Zhang and Jiang, 2008; Pickup and Thomson, 2009; Cveticanin, L., 2012; Erath and al., 2013; Pinheiro and Kerschen, 2013) even considering pathological cases in the vocal folds (Gunter, 2004) or stress situation (Luzan et al., 2015) but the idea here is to show that it is possible to generate jitter and voice signal with quality from the primary model considering the stiffness as a stochastic process and, mainly, validate the model proposed identifying parameters solving an statistical inverse problem taking into account experimental normal voices and

also with pathological characteristics.

2. Primary deterministic model

Figure. 1 illustrates a sketch of the model.

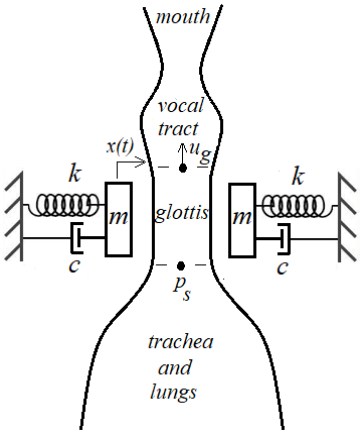

Figure 1: Sketch based on the Flanagan and Landgraf (1968) model.

Each vocal fold is represented by a nonlinear mass-stiffness-damper system and the complete model is composed by the subsystem of the vocal folds (*source*) coupled by the glottal flow to the subsystem of the vocal tract (*filter*). To generate jitter the stiffness will be considered as a stochastic process for which a model is proposed.

3. Stochastic modeling of jitter

The stiffness k is modeled by a stochastic process $\{K(t), t \in \mathbb{R}\}$ with values in $\mathbb{R}^+$. Consequently, the dynamical position of each vocal fold will be given by a stochastic process, named $X(t)$, coupled with the stochastic process associated with the glottal flow (volume flow velocity), noted $U_g(t)$. The stochastic dynamics of the vocal folds is described by Eq. 1:

$$m\,\frac{d^2X(t)}{dt^2}+\{c+c^*(X(t))\}\,\frac{dX(t)}{dt}+K(t)\,X(t)+a_1\,p_B(X(t),U_g(t))=a_2\,p_s(t)\,, \tag{1}$$

where $a_1 = 1.87\,\frac{\ell\,\mathrm{d}}{2}$ and $a_2 = \frac{\ell\,\mathrm{d}}{2}$, with ℓ the length of each vocal fold and d the vocal fold thickness. The stochastic process $X(t)$ is the displacement of the mass m of one vocal fold, $K(t)$ is its stiffness and c is its damping coefficient when the glottis is opened; when the glottis is closed, there is an additional

damping given by $c^*(X(t))$ described in the following, where the Bernoulli pressure $p_B(X(t), U_g(t))$ is also described.

- If $X(t) \geq x_0$ (the glottis is closed and x_0 is a minimum value corresponding to normal vocal folds when they are in relaxed position), then

$$c^*(X(t)) = 2\,\alpha\,\sqrt{m\,K(t)} \quad , \quad p_B(X(t), U_g(t)) = 0\,, \tag{2}$$

  in which $\alpha > 0$ is a given damping rate.

- If $X(t) < x_0$ (the glottis is opened), then

$$c^*(X(t)) = 0 \quad , \quad p_B(X(t), U_g(t)) = \frac{(1/2)\,\rho\,|U_g(t)|^2}{(A_{g0} + \ell\,X(t))^2}\,. \tag{3}$$

  where ρ is the air density and A_{g0} (the so-called neutral glottal area) is such that the critical value x_0 is written as $x_0 = -A_{g0}/\ell$.

The stochastic process $U_g(t)$ is the acoustic volume velocity through the glottal orifice (the glottal flow). The air pressure that comes from the lungs and forces the vocal folds is called the subglottal pressure and is denoted by $p_s(t)$. The constant parameters have been discussed in the original paper about the corresponding deterministic model (Flanagan and Landgraf, 1968). Some information about values can also be found in (Cataldo and Soize, 2017).

In this paper, the stochastic process $K = \{K(t), t \in \mathbb{R}\}$, indexed by $\mathbb{R}$, is constructed according to the properties defined as follows.

(i) For all t, $0 < k_0 \leq K(t)$ where k_0 is a positive constant independent of t.

(ii) As the idea is to construct the jitter effect, modeled as a stochastic perturbation of the corresponding periodic movement of the vocal folds produced when k is a constant, stochastic process $K(t)$ is assumed to be a stationary stochastic process that cannot be Gaussian (because it is a positive-valued stochastic process).

(iii) Denoting by E the mathematical expectation, $\{K(t), t \in \mathbb{R}\}$ is thus a non-Gaussian stationary stochastic process such that $E\{K(t)^2\} < +\infty$ for all t (second-order stochastic process), for which its mean function (that is independent of t) is written as $E\{K(t)\} = \underline{k} > k_0 > 0$, and which is assumed to be mean-square continuous in order to guaranty the existence of a power spectral measure.

A representation of non-Gaussian stochastic process $K(t)$ can be constructed using Information Theory as explained in (Soize, 2017). Following such a construction, we introduce a Gaussian second-order real-valued stochastic process, $Y = \{Y(t), t \in \mathbb{R}\}$, centered, mean-square continuous, stationary and ergodic,

physically realizable. A representation of stochastic process K can then be written as

$$K(t) = k_0 + (\underline{k} - k_0)(\underline{y} + Y(t))^2 \quad , \quad \forall\, t \in \mathbb{R}\,, \tag{4}$$

in which $\underline{y}$ is a parameter (that will be defined later) such that

$$E\{(\underline{y} + Y(t))^2\} = 1 \quad , \quad E\{(\underline{y} + Y(t))^4\} < +\infty\,. \tag{5}$$

The conditions defined by Eq. (5) effectively yields, for all t, $E\{K(t)\} = \underline{k}$ and $E\{K(t)^2\} < +\infty$. Let ω be the angular frequency in rad/s and f be the circular frequency in Hz such that $\omega = 2\pi f$. The Gaussian stochastic process Y is constructed as the linear filtering, $Y = h * N_\infty$, of the centered Gaussian white noise N_∞ (generalized stochastic process) whose power spectral density function is written, for all real ω, as

$$S_N(\omega) = \frac{1}{2\pi}\,, \tag{6}$$

and where $h = \mathcal{F}^{-1}\{H\}$ is the inverse Fourier transform of the complex-valued frequency response function $\omega \mapsto H(\omega)$ that we defined, for all real ω, by

$$H(\omega) = \frac{a}{-\omega^2 + 2i\omega\xi b + b^2}\,, \tag{7}$$

in which, a, b, and ξ are three positive parameters that will be defined later.

Consequently, the power spectral density function $S_Y(\omega)$ of Gaussian stationary stochastic process Y is written, for all real ω, as

$$S_Y(\omega) = \frac{1}{2\pi}\frac{a^2}{(b^2 - \omega^2)^2 + 4\xi^2 b^2 \omega^2} \quad , \quad a > 0 \quad , \quad b > 0 \quad , \quad \xi > 0\,. \tag{8}$$

From Eq. (8), it can be deduced that the mean-square derivative $\{\dot{Y}(t), t \in \mathbb{R}\}$ of stochastic process $\{Y(t), t \in \mathbb{R}\}$ is a second-order stochastic process because $\int_{\mathbb{R}} \omega^2\, S_Y(\omega)\, d\omega < +\infty$. Let $\{\mathbb{Z}(t) = (Y(t), \dot{Y}(t)), t \geq 0\}$ be the stochastic process with values in $\mathbb{R}^2$ solution of the following Itô stochastic differential equation,

$$d\mathbb{Z} = [\alpha]\, \mathbb{Z}\, dt + \beta\, dW(t) \quad , \quad t > 0\,, \tag{9}$$

with the initial condition $\mathbb{Z}(0) = (0, 0)$, in which $\{W(t), t \geq 0\}$ is the real-valued normalized Wiener stochastic process indexed by $[0, +\infty[$, where $[\alpha]$ is the (2×2) real matrix and β is the real vector such that

$$[\alpha] = \begin{bmatrix} 0 & 1 \\ -b^2 & -2\xi b \end{bmatrix} \quad , \quad \beta = \begin{bmatrix} 0 \\ a \end{bmatrix}. \tag{10}$$

It can be proved (see for instance, Krée and Soize, 1986) that Eq. (9) has a unique solution $\{\mathbb{Z}(t), t \geq 0\}$ such that, for $t_0 \to +\infty$, the stochastic process

$\{\mathbb{Z}(t), t \geq t_0\}$ is asymptotically stationary and tends to the stationary Gaussian stochastic process $\{(Y(t), \dot{Y}(t)), t \in \mathbb{R}\}$ in which $Y = h * N_\infty$. The first condition defined by Eq. (5) yields,

$$\underline{y}^2 + \int_{-\infty}^{+\infty} S_Y(\omega) d\omega = 1 \quad \Longrightarrow \quad \underline{y}^2 = 1 - \frac{a^2}{4\xi b^3} \,. \tag{11}$$

Consequently, the parameters must satisfied the following conditions,

$$0 < a^2 < 4\xi b^3 \quad , \quad b > 0 \quad , \quad \xi > 0. \tag{12}$$

In order to control the bandwidth of stationary stochastic process Y, we introduce the parameter $\epsilon > 0$ (Krée and Soize, 1986) that is defined by

$$\epsilon = \sqrt{1 - \frac{m_2^2}{m_0 m_4}} \quad , \quad m_{2p} = \int_{\mathbb{R}} \omega^{2p} \, S_Y(\omega) \, d\omega \quad , \quad p = 0, 1, 2 \,. \tag{13}$$

Parameter ϵ is estimated using the simulated signals and is discussed in the next section. It is important to say that the bandwidth is related to the quality of the synthesized sounds (Rabiner and Schafer, 2011) and this is one of the main reasons to introduce it in this paper, discussing the relation between the presence of jitter and the quality of the voice.

4. Simulation

4.1. General ideas

The objective of this section is to generate voice signals with jitter using the stochastic model proposed and to analyze the sensitivity of the stochastic model with respect to parameters a, b, and ξ. As the main idea is to generate jitter, a way to measure it will also be discussed. There are different ways to analyze jitter effects (Mongia, 2014). At first, it is important to define the random variable associated with the duration of the glottal cycle, which is defined as the duration between two successive times, the first one corresponding to the instant the vocal folds (glottis) opens and the second one the instant when it closes completely. The corresponding random variable will be denoted by T_{fund}. To calculate $T_f und$ from $X(t)$, it was used an algorithm based on an implementation of the RAPT pitch tracker (Talkin, 1995). For each glottal cycle k, a duration denoted by $T_{\text{fund}}(\theta_k)$ can be associated. Considering that the set $\{T_{\text{fund}}(\theta_k), k = 1, \ldots, N\}$ constitutes N realizations of random variable T_{fund} (corresponding to all the glottal cycles of the voice signal), jitter can be measured by the following equations.
(i) The absolute jitter, denoted by *JitterAbs*, is defined by

$$\textit{JitterAbs} = \frac{1}{N-1} \sum_{k=1}^{N-1} |T_{\text{fund}}(\theta_{k+1}) - T_{\text{fund}}(\theta_k)| \,. \tag{14}$$

(ii) The relative jitter, denoted by *JitterRel*, is defined by

$$\textit{JitterRel} = \frac{\dfrac{1}{N-1}\displaystyle\sum_{k=1}^{N-1} |T_{\text{fund}}(\theta_k) - T_{\text{fund}}(\theta_{k+1})|}{\dfrac{1}{N-1}\displaystyle\sum_{k=1}^{N-1} T_{\text{fund}}(\theta_k)}. \tag{15}$$

(iii) The relative average perturbation, denoted by *JitterRAP* , is defined as the average absolute difference between a period and the average of it and its two neighbors, divided by the average period.
(iv) The five-point period perturbation quotient, denoted by *JitterPPQ5*, is defined as the average absolute difference between a period and the average of it and its four closest neighbors, divided by the average period.
(v) Another important way for verifying the jitter generation is to use the probability density function associated with the random variable $F_{\text{fund}} = 1/T_{\text{fund}}$, which will be called the fundamental frequency.

4.2. Sensitivity analysis with respect to the parameters of S_Y

The variations of parameters a, b, and ξ are taken into account and for each triplet (a, b, ξ), the value of the relative jitter given by Eq. (15) is calculated. Before showing the results obtained, and understanding what happens when the ξ parameter varies, power spectral density function S_Y is calculated considering different values for the parameters but emphasizing the variation of ξ. For performing such a sensitivity analysis of S_Y with respect to its parameters, we introduce the normalization factor, S_Y^{ref}, defined as $\max_f S_Y(2\pi f; a, b, \xi) = a^2/(8\pi b^4 \xi^2(1-\xi^2))$ in which a, b, and ξ are fixed to 10, $2\pi\times180$, and 0.1, respectively.

For $a = 10$ and $b = 2\pi f_p$ with $f_p = 180\,Hz$, Fig. 2 shows the graph of the dimensionless and normalized power spectral density function $f \mapsto S_Y(2\pi f; a, b, \xi)/S_Y^{\text{ref}}$ as a function of $\xi \in \{0.1, 0.15, 0.3, 0.7\}$."The frequency f_p is chosen such that frequencies values of the voice will be around it, that is, its value is near of the fundamental frequency.´´ The maximum of each curve hap-

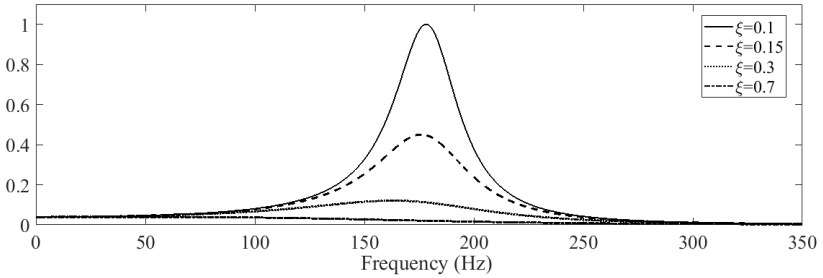

Figure 2: For $a = 10$ and $b = 2\pi f_p$ with $f_p = 180\,Hz$, graph of function $f \mapsto S_Y(2\pi f; a, b, \xi)/S_Y^{\text{ref}}$ as a function of $\xi \in \{0.1, 0.15, 0.3, 0.7\}$.

pens when the frequency is equal to $2\pi f_p\sqrt{1-2\xi^2}$. It is important to note that, as the value of ξ increases, the bandwidth of S_Y becomes larger. For $a = 10$ and $\xi = 0.1$, Figure 3 shows the graphf of the dimensionless and normalized power spectral density function $f \mapsto S_Y(2\pi f; a, b, \xi)/S_Y^{\text{ref}}$ as a function of $b = 2\pi\, c_b\, f_p$ in which $f_p = 180\,Hz$ and where $c_b \in \{1, 1.5, 2\}$. Again, it is important to note that the value of the

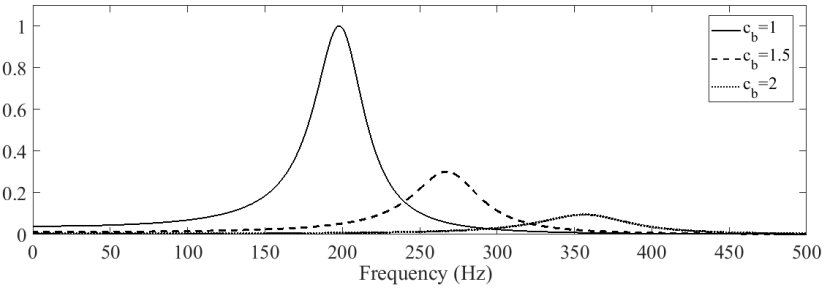

Figure 3: For $a = 10$ and $\xi = 0.1$, graph of function $f \mapsto S_Y(2\pi f; a, b, \xi)/S_Y^{\text{ref}}$ as a function of $b = 2\pi\, c_b\, f_p$ in which $f_p = 180\,Hz$ and where $c_b \in \{1, 1.5, 2\}$.

frequency corresponding to the maximum of each curve changes with the values of b and as b increases, the curve shifts to the right. Therefore, it is possible to change the frequency of the voice modifying this parameter although the fundamental frequency is fixed.

4.3. Cases simulated

During the simulations, the values of parameters a, b, and ξ, as well as the mean of the fundamental frequency, will vary. All the other parameters will be fixed and their values are $p_s(t) = 800\,Pa$, $m = 0.24 \times 10^{-2}\,kg$, $c = 346.3\,m/s$, $k_0 = 40\,N/m$, $\underline{k} = 115\,N/m$, $a_1 = 1.87\,\ell\,d/2$ and $a_2 = \ell\,d/2$, with $\ell = 1.4 \times 10^{-2} m$ and d $= 0.3 \times 10^{-2}\,m$. The other parameters that are necessary to produce the sounds, including values related to the vocal tract, are given in (Cataldo and Soize, 2017). In particular, the parameter A_{g0} and the air density ρ are chosen such that $A_{g0} = 0.04 \times 10^{-2}\,m^2$ and $\rho = 0.12\,kg/m^3$. The objective of this section is to perform a sensitivity analysis of the parameters in order to better understand how to proceed to solve the inverse problem to identify parameters of the model corresponding to experimental voice which will be discussed further in the paper. Another important objective of this section is to show that with these three parameters a, b and ξ, there are different possibilities to generate jitter, but also to control the distribution of the fundamental frequency. Although, jitter is a variation of the glottal cycle and consequently this variation is related to the random variable $F_{Fund} = 1/T_{Fund}$, the shape of the curve corresponding to the probability density function of F_{Fund} is not so easily controlled with the variation of jitter. It means that, to identify parameters of the model, it is important to minimize the distance between measures of jitter (from simulated and experimental signals) but also distance between probability density functions of F_{Fund} (simulated and experimental). So, this section will give the feeling of how to vary the parameters in order to better minimize those distances. Thirteen cases were simulated. For each case simulated, the corresponding voice signal will be synthesized and available to be heard following the link:

https://www.dropbox.com/sh/eo49b4usr1n4iz4/AADU8gI-JGWAeWqmxwE5u7nwa?dl=0 .

All the values of the parameters considered and also the value calculated for ε for all the simulations considered are summarized in Tab. 1.

Case	a	c_b	ξ	Relative jitter	ϵ
I	10	1	0.01	0.16%	0.43
II	200	1	0.01	0.32%	0.43
III	600	1	0.01	0.77%	0.43
IV	1200	1	0.01	3.48%	0.43
V	10	1	0.2	0.09%	0.9
VI	600	1	0.2	0.48%	0.9
VII	1800	1	0.2	1.43%	0.9
VIII	3000	1	0.2	2.51%	0.9
IX	10	1	0.5	0.09%	0.96
X	1000	1	0.5	0.80%	0.96
XI	3000	1	0.5	2.23%	0.96
XII	3000	1.5	0.5	0.90%	0.93
XIII	3000	2	0.5	0.64%	0.92

Table 1: Value of bandwidth parameter ε of stationary stochastic process Y for all the simulation cases.

As the idea is to discuss the sensitivity of the parameters with the specific objective of solving the inverse stochastic problem, some of these cases will be selected and the graphs showed and discussed. Graphs of the probability density function will be constructed for some cases. To construct the probability density function of the random variable $F_{Fund} = 1/T_{Fund}$, related to the variation of the fundamental frequency, it is necessary to calculate the glottal time interval, for each glottal cycle. It means the evaluation of the realizations of the random variable T_{Fund}. the case X is used to illustrate a piece of the voice signal generated and the corresponding graph of the normalized output pressure is shown, with five glottal time intervals highlighted, in Fig. 4. The correspond-

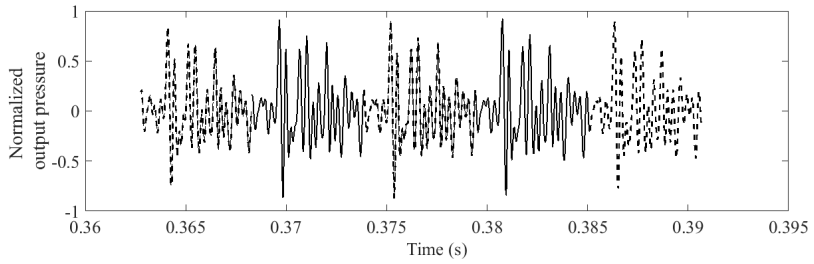

Figure 4: Piece of a voice signal: normalized output pressure.

ing values for the glottal time intervals evaluated in this case are: $0.005543\,s$, $0.005574\,s$, $0.005612\,s$, $0.005687\,s$ and $0.005515\,s$. And the corresponding fre-

quencies: $180.40\,Hz$, $179.37\,Hz$, $178.18\,Hz$, $175.83\,Hz$ and $181.33\,Hz$. These are five realizations of the random variables T_{Fund} and F_{Fund}, respectively. From Tab. 1, cases VIII, X, XI and XII will be chosen to be discussed and compared. At first, comparing case X with case XI, it can be observed a large variation of the parameter a, and the other parameters are fixed. Consequently, the level of jitter was much increased, passing from a case without characteristic of pathology (jitter less than 1%) to a case with pathological characteristics. Comparing case XI with case XII, the parameter c_b increases and it is interesting to note that the level of jitter decreases, maintaining all the other parameters fixed. And, finally, the reason for using the case VIII is that the parameter ξ is decreased, in relation to the other cases (X, XI and XII), with all the other parameters fixed. Then, the bandwidth of the power spectral density of the stochastic process is decreased, but the level of jitter is increased. So, the distribution of the fundamental frequency has to be analyzed when one wants to compare two voice signals, and not only the level of jitter can be compared. Although jitter is, in some way, directly related to the frequency variation, the distribution of the frequencies give some kind of information which cannot be perceived through verifying only the level of jitter. In summary, the probability density functions of the random variable $F_{Fund} = 1/T_{Fund}$ related to the cases VIII, X, XI and XII are constructed and shown together in Fig. 5. The impor-

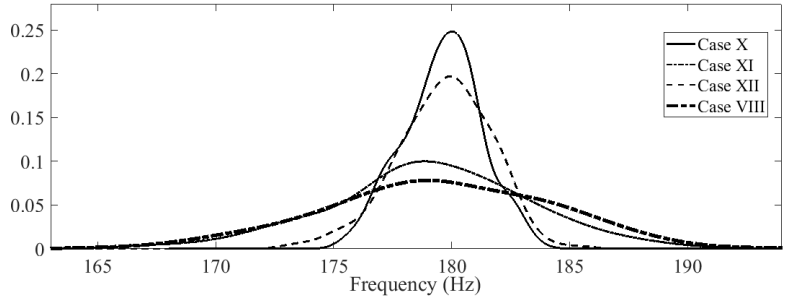

Figure 5: Probability density functions corresponding to cases VIII, X, XI, and XII(from the highest to the lowest).

tance of these cases is to compare the shape of the probability density functions even with different values of jitter. An important general observation is that, with different combination of the parameters a, b, and ξ, the same values of jitter can be obtained. However, the quality of the voice generated is not the same, because it also depends on the bandwidth of stationary stochastic process Y, which is defined by Eq. 13. This is one of the main reasons for calculating the bandwidth parameter ϵ and also to put it inside the Tab. 1. It is important to hear the synthesized sounds to better understand what it means.

5. Statistical inverse problem

In order to validate the model proposed, parameters a, b, and ξ are identified using experimental voice signals. This identification is carried out by introducing a cost function that is constructed writing that the probability density function associated with the simulated voice is close to the probability density function of the experimental voice and also, the jitter obtained for the simulated voice is close to the jitter of the experimental voice. The four measures of jitter are used. The cost function, denoted by $J_{\text{cost}}(a,b,\xi)$, is then defined by

$$\begin{aligned} J_{\text{cost}}(a,b,\xi) = & \frac{1}{2}\mathit{Dist}_{\text{dens}}(a,b,\xi) + \frac{1}{4}\mathit{JitterRel}_{\text{dist}}(a,b,\xi) \\ & + \frac{1}{4}\mathit{JitterAbs}_{\text{dist}}(a,b,\xi) + \frac{1}{4}\mathit{JitterRAP}_{\text{dist}}(a,b,\xi) \\ & + \frac{1}{4}\mathit{JitterPPQ5}_{\text{dist}}(a,b,\xi)\,, \end{aligned} \tag{16}$$

in which, each quantity appearing in the right-hand side member is defined hereinafter.
(i) Let $f \mapsto f_S(f;a,b,\xi)$ be the probability density function on $[0\,,+\infty[$ of random variable $F_{\text{fund}}(a,b,\xi)$ associated with the simulated voice and $f \mapsto f_R(f)$ be the probability density function on $[0\,,+\infty[$ of the random variable associated with the experimental voice. The distance between these two probability density functions is written as

$$\mathit{Dist}_{\text{dens}}(a,b,\xi) = \frac{1}{2}\int_0^{+\infty} \mid f_S(f;a,b) - f_R(f) \mid \, df\,. \tag{17}$$

The probability density functions are estimated by using the Gaussian kernel estimation method from the nonparametric statistics (Bowman and Azzalini, 1997). For each value of (a,b,ξ), probability density function $f_S(.;a,b,\xi)$ of $F_{\text{fund}}(a,b,\xi)$ is estimated using the realization of the stochastic process corresponding to the glottal flow computed with the stochastic model and probability density function f_R of F_{fund} is estimated using the realization of the experimental glottal signal obtained through a filtering inverse algorithm (PSIAIF) (Pavo,1992) of the experimental voice.

(ii) For each given value of vector (a,b,ξ), N realizations $\{\theta_k, k=1,\ldots,N\}$ of the voice signal are computed, which allows for computing the jitter quantities defined in Section 4.1. Let $\mathit{Jitter}_{\text{sim}}$ represent one of these four jitter quantities: $\mathit{JitterRel}_{\text{sim}}$, $\mathit{JitterAbs}_{\text{sim}}$, $\mathit{JitterRAP}_{\text{sim}}$, or $\mathit{JitterPPQ5}_{\text{sim}}$. Let $\mathit{Jitter}_{\text{exp}}$ be the jitter calculated with the experimental signal. Then, a distance between $\mathit{Jitter}_{\text{sim}}$ and $\mathit{Jitter}_{\text{exp}}$ can be defined by

$$\mathit{Jitter}_{\text{dist}} = \frac{|\mathit{Jitter}_{\text{sim}} - \mathit{Jitter}_{\text{exp}}|}{\mathit{Jitter}_{\text{exp}}}\,. \tag{18}$$

The optimal values a^{opt}, b^{opt} , and ξ^{opt} are then computed by solving the following optimization problem,

$$(a^{\text{opt}}, b^{\text{opt}}, \xi^{\text{opt}}) = \arg \min_{(a,b,\xi)\in\mathcal{C}} J_{\text{cost}}(a, b, \xi)\,, \tag{19}$$

in which the admissible set $\mathcal{C}$ is defined, using Eq. (12), by

$$\mathcal{C} = \{(a, b, \xi) \in \mathbb{R}^3 \quad \text{such that} \quad 0 < a^2 < 4\xi b^3 \quad , \quad b > 0 \quad , \quad \xi > 0\}\,. \tag{20}$$

The values of the fixed parameters considered for the corresponding deterministic model are the same as considered for the simulations. The first case to be taken into account is a voice signal from a woman producing an /e/ vowel. The parameters corresponding to the mean value $\underline{k}$ of K is considered in a way that the mean of the random variable associated with the fundamental frequency simulated is very near of the one for the real voices. Then, the optimal values a^{opt}, b^{opt} and ξ^{opt} of parameters a, b and ξ are identified by solving the optimization problem defined by Eq. (19).

5.1. Algorithm used

- Step 1: From the experimental voice signal obtained with the vowel produced all the values corresponding to the random variable $F_{\text{fund}} = 1/T_{\text{fund}}$ are obtained using the algorithm (Talkin, 1995) and the probability density function $f \mapsto F_R(f)$ associated is estimated. The mean value of random variable F_{fund} is calculated and is used in the other steps. From this signal, the four measures of jitter are obtained: $\textit{JitterRel}_{\text{exp}}$, $\textit{JitterAbs}_{\text{exp}}$, $\textit{JitterRAP}_{\text{exp}}$, and $\textit{JitterPPQ5}_{\text{exp}}$.

- Step 2: Using the model proposed, one signal is simulated in a way that the mean of random variable F_{fund} of this signal was near from the mean value calculated in step 1. It is not difficult to generate this signal because there are parameters in the model directly related to the fundamental frequency as, for example, f_p. However, some essays are necessary in order to obtain a mean value near the one wished. At the same time, values for a, b, and ξ have been calculated so that the estimated probability density function of random variable F_{fund} for the simulated voice signal is near from the probability density function estimated in step 1. This step 2 takes some time because it is a step of essays. Values are obtained and they will serve as start for the grid variation of the values of the fundamental frequency and also of the parameters a, b, and ξ, consequently four loops are constructed.

- Step 3: For each value of the fundamental frequency and of the triplet (a, b, ξ), the Monte Carlo Method is used for the estimation of the probability density functions and the computation of cost function $J_{cost}(a, b, \xi)$.

- Step 4: The minimum value of the cost function estimated in Step 3 is the objective that has to be reached.

5.2. Identifying parameters

5.2.1. The first experimental voice signal considered is a female production of a vowel /e/.

After solving the inverse stochastic problem, the optimal values obtained were: $a^{opt} = 200$, $b^{opt} = (1/2)\pi f_p$, $f_p = 200\,Hz$ and $\xi^{opt} = 0.9$. Table 2 shows the values of jitter calculated for the experimental voice and for the simulated voice, after solving the inverse problem. The value obtained for the bandwidth parameter is $\epsilon = 0.98$. As we have already discussed, not only the measures of

Jitter	Experimental	Simulated
JitterRel	0.48%	0.52%
JitterAbs	$2.37e-05\,s$	$2.54e-05\,s$
JitterRAP	0.26%	0.30%
JitterPPQ5	0.29%	0.32%

Table 2: Jitter values for a female production of a vowel /a/, without pathological characteristics

.

jitter have to be taken into account but also the distance between the probability density functions $f_S(.\,; a^{\text{opt}}, b^{\text{opt}}, \xi^{\text{opt}})$ and f_R (simulated and experimental) shown in Fig. 6.

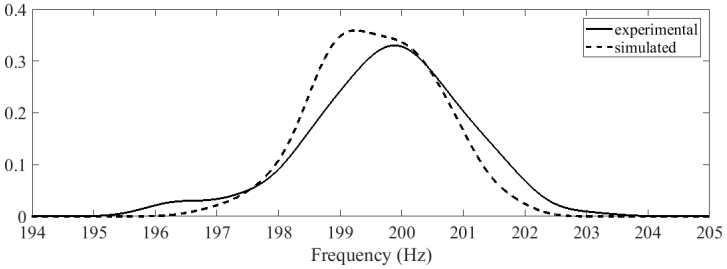

Figure 6: pdf of the random fundamental frequency corresponding to a real voice (solid line) and corresponding to the simulated for the optimal values of the parameters (dotted line) in the case without pathological characteristics.

It is important to say that if the distance between the pdfs was not taken into consideration inside the cost function, the values obtained to the jitter measures would be: *JitterRel* $= 0.50\%$, *JitterAbs* $= 2.4404e-05\,s$, *JitterRAP* $= 0.30\%$ and *JitterPPQ5*$5 = 0.31\%$ and in this case the distance between the pdfs would be a little bit greater. As a way to verify what happens when a sound is synthesized considering these optimal values of the parameters, a voice signal has been simulated with the optimal values of the parameters. The experimental signal ($exper_1.wav$) and the corresponding optimal simulated one ($simulated_1.wav$), in the same link presented before for all simulations.

5.2.2. The second case considered is a voice signal from a woman with paralysis of the vocal folds.

After solving the inverse stochastic problem, the optimal values obtained were: $a^{opt} = 1050$, $b^{opt} = (1.5\pi f_p$, $f_p = 226\,Hz$ and $\xi^{opt} = 0.4$. Table 3 shows the values of jitter calculated for the real voice and for the simulated voice, after solving the inverse problem. In this case, the value obtained for the bandwidth parameter is $\epsilon = 0.96$.

Jitter	Experimental	Simulated
JitterRel	3.24%	3.40%
JitterAbs	$1.44e-04\,s$	$1.42e-04\,s$
JitterRAP	2.03%	2.00%
JitterPPQ5	2.05%	2.45%

Table 3: Jitter values for a female production of a vowel /a/, with a pathology

.

Figure 7 shows probability density functions $f_S(.\,; a^{\text{opt}}, b^{\text{opt}}, \xi^{\text{opt}})$ and f_R (simulated and experimental). It is important to note the difference between the values obtained for jitter, but mainly showing that it is possible to solve the inverse stochastic problem even considering a pathological case.

In this case, the pathological one, the results are better for the distance between the values of jitter for the experimental and simulated signals than those obtained for the normal voice.

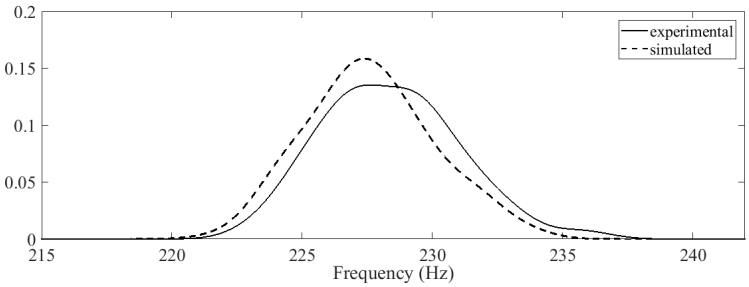

Figure 7: pdf of the random fundamental frequency corresponding to a real voice (solid line) and corresponding to the simulated for the optimal values of the parameters (dotted line) in the case without pathological characteristics.

In the normal case or in the pathological case, the value of ϵ is high. It shows that it is not directly related to the pathology, but to the quality of the synthesized sound.

6. Conclusions

A stochastic model has been proposed using three control parameters for generating jitter considering a mechanical model for producing voiced sounds. Some pathological cases have been generated and the model has been validated considering an inverse stochastic problem to identify the parameters. With three control parameters more possibilities of different sound are obtained, including different levels of jitter and, mainly, it is possible to control the quality of the synthesized voice. The inverse stochastic problem that is solved to identify parameters of the model uses different measures of jitter and also the distance between probability density functions, showing that with more measured features the voices synthesized are more similar of the corresponding experimental voices. A pathological case caused by an unilateral paralysis of the vocal folds has been considered and, even in this case, the parameters of the model has been identified. The bandwidth parameter has been used as a measure of quality of the synthesized voice and it has also been considered when the inverse problem has been solved.

7. Acknowledgments

This work was supported by CNPq.

REFERENCES

Avanzini, F., 2008. Simulation of vocal fold oscillation with a pseudo-one-mass physical model. Speech Communication, 50 (2), 95–108.

Bowman, A. W., Azzalini, A., 1997. Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations. Oxford University Press.

Cataldo E., Soize C., Sampaio R., 2012. Using Bayesian method for updating the probability density function related to the tension parameter in a voice production model. Journal of Biomechanics, 45 (1), S481. http://dx.doi.org/10.1016/S0021-9290(12)70482-7.

Cataldo, E., Soize, C., 2016. Jitter generation in voice signals produced by a two-mass stochastic mechanical model. Biomedical Signal Processing and Control (Print), 27, 87–95.

Cataldo, E., Soize, C., 2017. Voice signals produced with jitter through a stochastic one-mass mechanical model. Journal of Voice, 31 (1), 111e9–111e18.

Cveticanin, L., 2012. Review on mathematical and mechanical models of the vocal cord. http://dx.doi.org/10.1155/2012/928591.

Dejonckerea, P. H., Giordano, A., Schoentgen, J., Fraj, S., Bocchid, L., Manfredid, C., 2012. To what degree of voice perturbation are jitter measurements valid? A novel approach with synthesized vowels and visuoperceptual pattern recognition. Biomedical Signal Processing and Control, 7(1), 37–42.

Erath, B. D., Zañartu, M., Stewart, K. C., Plesniak, M. W., Peterson, S. D., 2013. A review of lumped-element models of voiced speech. Speech Communication, 55 (5), 667–690.

Flanagan, J., Landgraf, L., 1968. Self-oscillating source for vocal-tract synthesizers. IEEE Transactions on Audio and Electroacoustics, AU-16, 57–64.

Gunter, H. E., 2004. Modeling mechanical stresses as a factor in the etiology of benign vocal fold lesions. Journal of Biomechanics, 37 (7), 1119–1124.

Luzan, C. F., Mihaescu, J. C. M., Khosla, S.M., Gutmark, E., 2015. Computational study of false vocal folds effects on unsteady airflows through static models of the human larynx. Journal of Biomechanics, 48 (7), 1248–1257.

Ishizaka, K., Flanagan, J., 1972. Synthesis of voiced sounds from a two-mass model of the vocal folds. Bell Syst. Tech. J., 51, 1233–1268.

Jiang, J. J., Zhang, Y., MacCallum, J., Sprecher, A., Zhou, L., 2009. Objective acoustic analysis of pathological voices from patients with vocal nodules and polyps. Folia Phoniatrica et Logopaedica, 61, 342–349.

Krée P., Soize C., 1986. Mathematics of Random Phenomena. Reidel, Dordrecht.

Mongia, P. K., Sharma, R.K., 2014. Estimation and statistical analysis of human voice parameters to investigate the influence of psychological stress and to determine the vocal tract transfer function of an individual. Journal of Computer Networks and Communications, 1–17.

Pradeep, M., Tech M., 2012. Improving Sound Quality by Bandwidth. International Journal of Scientific & Engineering Research, 3(9),1–9.

Talkin, D., 1995. A robust algorithm for pitch tracking (rapt). Speech coding and synthesis, 495–518.

Pavo, A., 1992. Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering. Speech Communication, V. 11 (2–3), 109–118.

Pickup, B. A., Thomson, S. L., 2009. Influence of asymmetric stiffness on the structural and aerodynamic response of synthetic vocal fold models. Journal of Biomechanics, 42 (14), 2219-2225.

Pinheiro, A. P., Kerschen, G., 2013. Vibrational dynamics of vocal folds using nonlinear normal modes. Medical Engineering & Physics, 35(8), 1079–1088.

Rabiner, L., Schafer, R., 2011. Theory and Application of Digital Signal Processing. Pearson Education.

Schoengten, J., De Guchteneere, R., 1997. Predictable and random components of jitter. Speech Communication 21, 255–272.

Silva, M., Vellasco, M. M. B., Cataldo, E., 2016. Evolving spiking neural networks for recognition of aged voices. Journal of voice, 31, 25–33.

Soize, C., 2017. Uncertainty Quantification - An accelerated Course with Advanced Applications in Computational Engineering. Interdisciplinary Applied Mathematics Series, Springer, New York.

Titze, I. R., 1994. Principles of voice production. Prentice Hall, Englewood Cliffs, NJ.

Titze, I. R., 1994. Mechanical stress in phonation. Journal of Voice, 8, 99–105.

Wong, D., Ito, M. R., Cox, N. B., Titze, I. R., 1991. Observation of perturbations in a lumped-element model of the vocal folds with application to some pathological cases. The Journal of the Acoustical Society of America 89(1), 383–394.

Zhang, Y., Jiang, J. J., 2008. Nonlinear dynamic mechanism of vocal tremor from voice analysis and model simulations. Journal of Sound and Vibration, 316 (1–5), pages 248–262.