Voice signals produced with jitter through a stochastic one-mass mechanical model

Edson Cataldo, Christian Soize

Voice signals produced with jitter through a stochastic one-mass mechanical model

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

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

Abstract

The quasi-periodic oscillation of the vocal folds causes perturbations in the length of the glottal cycles which are known as jitter. The observation of the glottal cycles variations suggests that jitter is a random phenomenon described by random deviations of the glottal cycle lengths in relation to a corresponding mean value and, in general, its values are expressed as a percentage of the duration of the glottal pulse. The objective of the paper is the construction of a stochastic model for jitter using an one-mass mechanical model of the vocal folds, which assumes complete right-left symmetry of the vocal folds, and which considers motions of the vocal folds only in the horizontal direction. Concerning the study design, the jitter has been the subject for researchers due to its important applications such as the identification of pathological voices (nodules in the vocal folds, paralysis of the vocal folds, or even, the vocal aging, among others). Large values for jitter variations can indicate a pathological characteristic of the voice. Concerning the model, the corresponding stiffness of each vocal fold is considered as a stochastic process and its modeling is proposed. The results that are presented concern the probability density function of the fundamental frequency related to the voice signals produced, which are constructed and are compared for different levels of jitter. Some samples of synthesized voices in these cases are obtained. As conclusions, it is showed that jitter could be obtained using the model proposed. The Praat software was also used in order to verify the measures of jitter in the synthesized voice signals.

Keywords: Stochastic modeling, voice production, mechanical models, jitter.

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

1. Introduction

The systems of voice production are important sensorial structures, which permit to the human beings communicate, share information, exchange ideas, feelings, emotions, intentions, etc. The inefficiency of these structures or its absence can make even the social life more difficult. In addition, there are people who depend upon their voices to work, such as broadcasters, singers and other. So, the interest of evaluating the vocal structures remains important for the human voice production. Roughly speaking, an air stream coming from the lungs passes through the trachea, vocal and nasal structures, and reaches the mouth. In particular, in voiced speech production, where vowels are included, the production of the voice signal is due to the oscillation of the vocal folds, which modifies the airflow into pulses of air (the so-called glottal signal) which will be further filtered and amplified by the vocal tract and, finally, radiated by the mouth. However, the oscillations of the vocal folds are not exactly periodic and the pulses of air, which compose the glottal signal, have not exactly the same time duration. The small random fluctuation in each glottal cycle length is called jitter and its study is particularly important in different areas related to the voice generation. One of the first works for quantifying the jitter was proposed by Lieberman [1] who has characterized it by introducing a factor representing all perturbations greater than 0.5 ms. Other preliminary works were based on the calculations of a typical value related to the differences between the lengths of the cycles and their mean values or, more rarely, from the instantaneous frequencies and their mean values. Basically, these works agree with the fact that typical values of the jitter are between 0.1% and 1% of the fundamental period, for the so-called normal voices; that is, without presence of pathologies. The jitter value can be seen as a measure of the irregularity of a quasi-periodic signal and it can be a good indicator of the presence of pathologies such as vocal fold nodules or a vocal fold polyp [2, 3, 4, 5]. It is important to say that, in general, jitter decreases as the fundamental frequency increases. The majority of the authors concludes that it is possible to discriminate healthy voices from pathological voices using jitter characteristics and even to recognize speakers [6, 7, 8, 9, 10]. Jitter can be used for measuring the voice quality, for indicating the presence of pathologies related to the voice, and even for helping the speech recognition [11, 12, 13]. Some authors have even used jitter to discuss the relation between age and changes in vocal jitter [14, 15]. In general, to investigate the presence of pathologies related to the voice it is necessary to extract not only jitter from the voice signal, but also other measures, like *shimmer* and *HNR (harmonic-noise ration)* [16, 17]. There are some important mechanical models discussed in the literature to produce voice and even to simulate some pathologies or irregularities related to voice production [18, 19, 20, 21]. Erath et al. [22] made a good review of lumped-element models of voiced speech, discussing since the anatomy and physiology of the vocal folds up to applications of lumped-element vocal fold models in speech research, including the discussing of mechanical models and pathological phonation. Shinji Deguchi and Juki Kawahara [23] present a continuum-based numerical model

of phonation to simulate human phonation with vocal nodules. Fraile et al. [24] simulate vocal tremor using a high-dimensional discrete vocal fold model. However, all of these models are deterministic. There are also some discussions about generation of chaotic voice signals using mechanical models as the one proposed by Wong et al. [2], who develops a mass-spring model which is a hybrid of the two-mass and the longitudinal string models, proposed by Ishizaka and Flanagan [19] and Titze [20], respectively. The model is used to simulate the motion of normal and asymmetric vocal folds. With variation of tissue mass and stiffness, subharmonic and chaotic vibrations in the displacement of the vocal folds are obtained. It is concluded that similar vibratory characteristics also appeared in pathological speech data analyzed using time domain jitter and shimmer measures and a harmonics-to-noise ratio metric. This model is also deterministic and some authors, as Lieberman [1] and Schoengten [25], suggest that jitter designates feeble random cycle-to-cycle perturbations of the glottal cycle lengths. In general, the authors who work with models of jitter (or the variations of the fundamental frequency) do not introduce mathematical models for the voice production and only a few authors consider stochastic models [26, 27, 28, 29]. Some motivations for developing models of jitter include the discussion about the mechanisms that may cause the movements of the vocal folds to be aperiodic. The causes of glottal aperiodicities are multiple and in this paper we discuss one of these causes. Models of jitter may also help to improve naturalness or mimic hoarse voices and also a motivation is to confirm the mathematical form of markers that would characterize perturbed cycle lengths statistically rather than heuristically [30]. The objective of this paper is to construct a stochastic model of jitter based on the use of the voice production deterministic model introduced by Flanagan and Landgraf [18], including the modifications brought from the Ishizaka and Flanagan model and those introduced by the authors. Previous works have discussed stochastic mechanical models to produce voice [26, 27] considering some model parameters as uncertain and modeled by random variables for which prior probability distributions have been constructed and then updated. However, the approach used here is different and consists of modeling the stiffness as a stochastic process. Once the stochastic modeling is done, a nonlinear stochastic differential equation has to be solved. The synthesis of voice signals is then obtained in taking into account different levels of jitter.

2. Deterministic model used

The deterministic model used as start is the nonlinear one-mass model proposed by Flanagan and Landgraf to generate voice. The complete model is composed by two subsystems: the subsystem of the vocal folds (*source*) and the subsystem of the vocal tract (*filter*). The two subsystems are coupled by the glottal flow. During the phonation, the filter is excited by the sequence of pulses of the glottal signal. Each vocal fold is represented by a mass-stiffness-damper system and a symmetric system composed by two vocal folds is constituted. The vocal tract is represented by a standard configuration of concatenated tubes

[1, 31]. The complete model considered here presents some modifications in relation to the original Flanagan and Landgraf model. Some of them have been introduced by Ishizaka and Flanagan, and others by Cataldo et al. [26, 32]. The system of differential equations to be solved can be divided in three parts (see Fig. 1 that illustrates a sketch of the model):

- A nonlinear integro-differential equation for the glottal flow that is coupled with the vocal tract, called the *coupling equation* (Eq. (1)).
- A system of linear integro-differential equations related to the sound acoustic propagation through the vocal tract and called the *sound acoustic propagation equation* (Eq. (7)).
- A nonlinear differential equation related to the dynamics of the vocal folds and called the *vocal folds dynamic equation* (Eq. (8)).

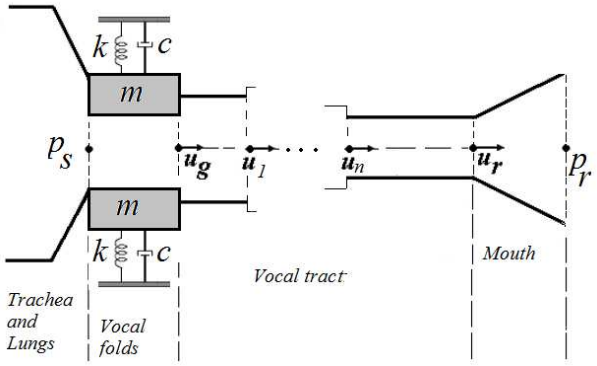

Figure 1: Sketch of the Flanagan and Landgraf model (1968).

Before describing these three equations, we define what are the unknowns of these equations.

- The *coupling equation* is a scalar nonlinear integro-differential equation whose unknown time-dependent function is the real-valued function $t \mapsto u_g(t)$ that models the acoustic volume velocity through the glottis. This equation depends on the real-valued function $t \mapsto u_1(t)$ that is related to the first tube of the sound acoustic propagation into the vocal tract (see hereinafter).
- The *sound acoustic propagation equation* is constituted of $n+1$ scalar linear integro-differential equations for which the $n+1$ unknown time-dependent functions are the real-valued functions $t \mapsto u_1(t), \ldots, u_n(t), u_R(t)$.
- The *vocal folds dynamic equation* is a scalar nonlinear differential equation whose unknown time-dependent function is the real-valued function $t \mapsto$

$x(t)$ that is the displacement of the mass of the vocal folds. For all t, $x(t)$ is generated by the *vocal folds dynamic equation*. The solution x of such an equation is constructed for all t and is used as follows:

- The collision of the vocal folds occurs at a time t when $x(t)$ reaches a given critical value x_0 (defined after) and, at this time t, the glottis closes. The glottis remains closed so that the values $\{x(\tau), t \leq \tau \leq t'\}$ of x (that are generated by the *vocal folds dynamic equation* are such that $x(\tau) \leq x_0$, and until the time t' that is such that $x(\tau') > x_0$ for $\tau' > t'$; x_0 is a value related to the initial space between the vocal folds, before phonation starting (Eq. (6)).
- For $\{t \leq \tau \leq t'\}$, when the glottis is closed, u_g and du_g/dt remain zero, $u_g(\tau) = du_g(\tau)/d\tau = 0$, the damping is modified in the vocal folds dynamic equation, but the propagation of sound goes on in the vocal tract .

Coupling equation. This coupling nonlinear equation in u_g and u_1, for which coefficients depend on $x(t)$, is written as

$$\begin{aligned}\{R_v(x(t)) + \mathcal{R}_k(x(t), u_g(t))\}\, u_g(t) + \{L_g(x(t)) + L_1\}\frac{du_g(t)}{dt} + \\ \frac{1}{c_1}\int_0^t \{u_g(\tau) - u_1(\tau)\}\, d\tau - p_s(t) = 0\,,\end{aligned} \tag{1}$$

where the coefficients $R_v(x(t))$, $\mathcal{R}_k(x(t), u_g(t))$, and $L_g(x(t))$ are defined by

$$R_v(x(t)) = 12\,\mu\,\mathrm{d}\,\ell^2\,[A_g(x(t))]^{-3}\,, \tag{2}$$

$$\mathcal{R}_k(x(t), u_g(t)) = 0.44\,\rho\,|u_g(t)|\,[A_g(x(t))]^{-2}\,, \tag{3}$$

$$L_g(x(t)) = \rho\,\mathrm{d}\,[A_g(x(t))]^{-1}\,, \tag{4}$$

in which $A_g(x(t))$ is the glottal area that depends on $x(t)$ and that is written as

$$A_g(x(t)) = A_{g0} + \ell\,x(t)\,, \tag{5}$$

with ℓ the length of each vocal fold, and where A_{g0} is such that the critical value x_0 is written as

$$x_0 = -A_{g0}/\ell\,. \tag{6}$$

In Eqs. (2) to (4), μ is the air kinematic viscosity, d is the vocal fold thickness, and ρ is the air density. In Eq. (1), $p_s(t)$ is the subglottal pressure that is given and the coefficients c_1 and L_1 are defined hereinafter. When the glottis is closed at a time t, Eq. (1) becomes

$$\frac{1}{c_1}\int_0^t \{u_g(\tau) - u_1(\tau)\}\, d\tau - p_s(t) = 0\,.$$

Sound acoustic propagation equation. We consider the configuration of the vocal tract proposed by [31]. The vocal tract is represented as a transmission line of

n cylindrical tubes, for which the section areas are $A_1, \ldots, A_n$ (the last area A_n corresponds to the mouth) and where the tube lengths are $\ell_1, \ldots, \ell_n$. For $i = 1, \ldots, n$, the corresponding *inductances* are given by $L_i = \rho\, \ell_i/(2A_i)$ and the *capacitances* by $c_i = \ell_i A_i/(\rho c_a^2)$ in which c_a is the sound velocity in air. To take into account the lost of the vocal tract, *resistances* are introduced in series and are such that $r_i = (S_i/A_i^2)\sqrt{\rho\, \mu\, \omega/2}$ where S_i is the length of the i-th circumference and $\omega = (k/m)^{1/2}$ is the eigenfrequency frequency of the undamped vocal folds. The line transmission ends with a radiation load for which the inductance is written as $L_R = (8\,\rho/(3\pi))\sqrt{\pi A_n}$ and the resistance as $r_R = 128\,\rho\, c_a/(9\pi^2 A_n)$. Consequently, the linear integro-differential equations related to the wave acoustic propagation through the vocal tract, which is coupled to Eq. (1) by time-dependent function u_1, are written as

$$
\left\{
\begin{array}{l}
(L_1 + L_2)\dfrac{du_1(t)}{dt} + (r_1 + r_2)\, u_1(t) + \dfrac{1}{c_2}\displaystyle\int_0^t \{u_1(\tau) - u_2(\tau)\}\, d\tau + \\
\qquad\qquad \dfrac{1}{c_1}\displaystyle\int_0^t \{u_1(\tau) - u_g(\tau)\}\, d\tau = 0\,, \\
\\
(L_i + L_{i+1})\dfrac{du_i(t)}{dt} + (r_i + r_{i+1})\, u_i(t) + \\
\qquad\qquad \dfrac{1}{c_{i+1}}\displaystyle\int_0^t \{u_i(\tau) - u_{i+1}(\tau)\}\, d\tau + \\
\qquad\qquad \dfrac{1}{c_i}\displaystyle\int_0^t \{u_i(\tau) - u_{i-1}(\tau)\}\, d\tau = 0 \quad , \quad i = 2, \ldots, n-1\,, \\
\\
(L_n + L_R)\dfrac{du_n(t)}{dt} + r_n\, u_n(t) - L_R\dfrac{du_R(t)}{dt} + \\
\qquad\qquad \dfrac{1}{c_n}\displaystyle\int_0^t \{u_n(\tau) - u_{n-1}(\tau)\}\, d\tau = 0\,, \\
\\
L_R\dfrac{d(u_R(t) - u_n(t))}{dt} + r_R\, u_R(t) = 0\,.
\end{array}
\right. \tag{7}
$$

Vocal folds dynamic equation. The nonlinear differential equation in x for the vocal folds dynamics, which is coupled with the vocal-tract (through $u_g(t)$) is written as

$$
m\,\frac{d^2x(t)}{dt^2} + \{c + c^*(x(t))\}\,\frac{dx(t)}{dt} + k\,x(t) + a_1\,p_B(x(t), u_g(t)) = a_2\,p_s(t)\,, \tag{8}
$$

in which $a_1 = 1.87\,\frac{\ell\,\mathrm{d}}{2}$ and $a_2 = \frac{\ell\,\mathrm{d}}{2}$, where $x(t)$ is the displacement of the mass m of one vocal fold, k is its stiffness, and c is its damping coefficient when the glottis is opened (when the glottis is closed, there is an additional damping). The coefficient $c^*(x(t))$ and the nonlinear function $p_B(x(t), u_g(t))$ are defined as follows:

- If $x(t) \geq x_0$ (the glottis is closed), then

$$c^*(x(t)) = 2\,\alpha\,\sqrt{m\,k} \quad , \quad p_B(x(t), u_g(t)) = 0\,, \tag{9}$$

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

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

$$c^*(x(t)) = 0 \quad , \quad p_B(x(t), u_g(t)) = \frac{(1/2)\,\rho\,|u_g(t)|^2}{(A_{g0} + \ell\,x(t))^2}\,. \tag{10}$$

Remarks on the deterministic system and the stochastic modeling of the jitter.

- In Eqs. (1) to (10), the values of the parameters can be found in [26, 33, 34].
- Eqs. (1) to (10) constitute a set of nonlinear coupled equations.
- In Eq. (1), $du_g(t)/dt$ does not exist at a time t for which the glottis is closing or is opening. Such a non existence is taken into account by the numerical scheme of time integration during the computation [26, 34].
- The analysis of the existence and uniqueness of a solution and the possible bifurcations are very difficult to analyze from a mathematical point of view. However, these equations have been numerically studied by several authors and a knowledge on the type of the solutions that can be obtained is available (in particular, see hereinafter).
- The voice production model constructed defined by Eqs. (1) to (10) can effectively produce the phonation using only the few control parameters introduced in the model. It is important to note that there is a range for the values of the control parameters, which allows for obtaining a regular phonation. For example, if the subglottal pressure is too low, the phonation will not be possible. On the other side, if it is too high, the vocal folds can oscillate in a nonperiodic manner . The parameters considered in this paper, called the typical glottal condition, make sure that glottal-flow signal reaches a periodic steady-state for the deterministic model defined by Eqs. (1) to (10).
- Preliminary studies have shown that small variations of the control parameters can be associated with a physiological action that allows for producing the sounds that are targeted. Some types of pathologies can be simulated for certain values of the control parameters, in particular for the mass m and the stiffness k of the vocal folds.
- In previous works [26, 27, 33], the tension parameter of the vocal folds (which is a parameter describing a relation between mass m and stiffness k of the vocal folds and which can be found, for instance, in [19], was considered as a random variable and its probability density function was constructed and identified by solving an inverse stochastic problem. Further,

it was updated using the Bayesian method [27]. In this paper, we consider a stochastic model for k that is modeled by a stochastic process $K(t)$ in order to generate a jitter in the voice production for the phonation. The corresponding nonlinear stochastic differential equation is obtained by substituting in Eq. (8), stiffness k by $K(t)$. Consequently, $u_g, u_1, \ldots, u_n, u_R$ and x become stochastic processes $U_g, U_1, \ldots, U_n, U_R, X$, which have to verify Eqs. (1) to (10) and then, these nonlinear stochastic integro-differential differential equations have to be solved. We are interested in constructing the asymptotic stationary solution which corresponds to a "quasi-periodic" steady state, showing that the normal voices and also some pathological voices can be characterized with such a stochastic model of the voice production.

- The construction of the stochastic process $K(t)$ is explained in hereinafter.

3. Generating Jitter

3.1. Stochastic modelling of jitter

Let $\{K(t), t \in \mathbb{R}\}$ be a stochastic process indexed by the real line $\mathbb{R}$, with values in $\mathbb{R}^+$, which models stiffness k in Eq. (8). As the objective of the stochastic model is to enrich the deterministic model of the voice production, and since we are interested in constructing a *stochastic perturbation* (the jitter effect) of the periodic solution that is produced when the stiffness is a constant k, it is coherent to introduce a stationary stochastic process for $\{K(t), t \in \mathbb{R}\}$ (because, if a constant k can be viewed as a particular stationary stochastic process, while a constant k cannot be viewed as a nonstationary stochastic process). Consequently, the deterministic equations defined by Eqs. (1), (7), and (8) for deterministic time functions $u_g, u_1, \ldots, u_n, u_R$ and x become stochastic equations for stochastic processes denoted by $U_g, U_1, \ldots, U_n, U_R, X$, which are written as follows.

Stochastic vocal folds dynamic equation. The vocal folds dynamic equation (defined by Eqs. (8) to (10)) in $x(t)$ depending on $u_g(t)$ becomes a nonlinear stochastic differential equation for the stochastic process $X(t)$ coupled with the stochastic process U_g, such that

$$m\,\frac{d^2X(t)}{dt^2}+\{c+c^*(X(t))\}\,\frac{dX(t)}{dt}+K(t)\,X(t)+a_1\,p_B(X(t),U_g(t))=a_2\,p_s(t)\,, \tag{11}$$

in which $c^*(X(t))$ and $p_B(X(t),U_g(t))$ are such that

- If $X(t) \geq x_0 \quad a.s.$ (the glottis is closed), then

$$c^*(X(t))=2\,\alpha\,\sqrt{m\,K(t)} \quad , \quad p_B(X(t),U_g(t))=0 \quad a.s\,. \tag{12}$$

- If $X(t) < x_0 \quad a.s.$ (the glottis is opened), then

$$c^*(X(t)) = 0 \quad a.s. \quad , \quad p_B(X(t), U_g(t)) = \frac{(1/2)\,\rho\,|U_g(t)|^2}{(A_{g0} + \ell\,X(t))^2}\,. \tag{13}$$

Stochastic coupling equation. The coupling nonlinear equation (defined by Eq. (1) with Eqs. (2) to (6)) in u_g and coupled with u_1, for which the coefficients depend on $x(t)$, become a stochastic coupling equation for the stochastic process U_g and coupled with the stochastic process U_1, for which the coefficients depend on stochastic process X, and is written as

$$\begin{aligned}\{R_v(X(t)) + \mathcal{R}_k(X(t), U_g(t))\}\,U_g(t) + \{L_g(X(t)) + L_1\}\frac{dU_g(t)}{dt} + \\ \frac{1}{c_1}\int_0^t \{U_g(\tau) - U_1(\tau)\}\,d\tau - p_s(t) = 0\,,\end{aligned} \tag{14}$$

where the stochastic coefficients $R_v(X(t))$, $\mathcal{R}_k(X(t), U_g(t))$, and $L_g(X(t))$ are defined by Eqs. (2) to (5). When the glottis is closed at a time t, Eq. (14) becomes

$$\frac{1}{c_1}\int_0^t \{U_g(\tau) - U_1(\tau)\}\,d\tau - p_s(t) = 0\,. \tag{15}$$

Stochastic sound acoustic propagation equation. The resistances $r_1, \ldots, r_n$ of the vocal tract become stochastic processes $R_1, \ldots, R_n$ such that, for all t and for all $i = 1, \ldots, n$,

$$R_i(t) = (S_i/A_i^2)\sqrt{\rho\,\mu\,\Omega(t)/2} \quad , \quad \Omega(t) = (K(t)/m)^{1/2}\,, \tag{16}$$

and, consequently, the sound acoustic propagation equation (defined by Eq. (7)) in $u_1, \ldots, u_n, u_R$ becomes a stochastic sound acoustic propagation equation with stochastic coefficients for the stochastic processes $U_1, \ldots, U_n, U_R$, which is writ-

ten as

$$\left\{\begin{array}{l} (L_1+L_2)\dfrac{dU_1(t)}{dt}+(R_1(t)+R_2(t))\,U_1(t)+\dfrac{1}{c_2}\displaystyle\int_0^t\{U_1(\tau)-U_2(\tau)\}\,d\tau+ \\ \qquad\qquad\qquad \dfrac{1}{c_1}\displaystyle\int_0^t\{U_1(\tau)-U_g(\tau)\}\,d\tau=0\,, \\ \\ (L_i+L_{i+1})\dfrac{dU_i(t)}{dt}+(R_i(t)+R_{i+1}(t))\,U_i(t)+ \\ \qquad\qquad\qquad \dfrac{1}{c_{i+1}}\displaystyle\int_0^t\{U_i(\tau)-U_{i+1}(\tau)\}\,d\tau+ \\ \qquad\qquad\qquad \dfrac{1}{c_i}\displaystyle\int_0^t\{U_i(\tau)-U_{i-1}(\tau)\}\,d\tau=0 \quad , \quad i=2,\ldots,n-1\,, \\ \\ (L_n+L_R)\dfrac{dU_n(t)}{dt}+R_n(t)\,U_n(t)-L_R\dfrac{dU_R(t)}{dt}+ \\ \qquad\qquad\qquad \dfrac{1}{c_n}\displaystyle\int_0^t\{U_n(\tau)-U_{n-1}(\tau)\}\,d\tau=0\,, \\ \\ L_R\dfrac{d(U_R(t)-U_n(t))}{dt}+r_R\,U_R(t)=0\,. \end{array}\right. \tag{17}$$

3.1.1. Construction of a stochastic model for $K(t)$

The following properties of the stochastic process $\{K(t), t\in\mathbb{R}\}$ are introduced in order to obtain a suitable solution for stochastic equations: Eqs. (1) to (7) and Eqs. (11) to (13):

(i) For all t, $0<k_0\le K(t)$ *a.s.* , where k_0 is a positive constant.

(ii) $\{K(t), t\in\mathbb{R}\}$ is a stationary stochastic process (for the reason given before).

(iii) $\{K(t), t\in\mathbb{R}\}$ is a second-order stochastic process, mean-square continuous, with mean value $\underline{k}=E\{K(t)\}>k_0>0$. The centered stochastic process K_c is such that $K(t)=K_c(t)+\underline{k}$. The autocorrelation function of stochastic process K_c is written, for all real τ, as $R_{K_c}(\tau)=\int_{-\infty}^{+\infty}e^{i\omega\tau}S_{K_c}(\omega)\,d\omega$ in which the positive-valued function $S_{K_c}(\omega)$ is the power spectral density function and, for all fixed t, the variance of the random variable $K(t)$ is such that $\sigma_K^2=R_{K_c}(0)=\int_{-\infty}^{+\infty}S_{K_c}(\omega)\,d\omega$.

(iv) For all fixed t in $\mathbb{R}$, the random variable $K(t)$ is written as

$$K(t)=k_0+(\underline{k}-k_0)(\underline{z}+Z(t))^2\,. \tag{18}$$

The stochastic process Z and the real constant $\underline{z}$ must be constructed in order that, for all t in $\mathbb{R}$, $E\{(\underline{z}+Z(t))^2\}=1$ and $E\{(\underline{z}+Z(t))^4\}<+\infty$. The stochastic process $\{Z(t), t\in\mathbb{R}\}$ is constructed as a second-order Gaussian stochastic

process, indexed by $\mathbb{R}$, with values in $\mathbb{R}$, which is centered, mean-square continuous, stationary and ergodic, physically realizable, whose power spectral density function $S_Z(\omega)$ is written as

$$S_Z(\omega) = \frac{1}{2\pi}\frac{a^2}{\omega^2 + b^2} \quad , \quad a > 0 \quad , \quad b > 0\,, \tag{19}$$

in which a and b must satisfy the constraint equation $E\left\{(\underline{z} + Z(t))^2\right\} = 1$ that can be written as

$$\underline{z}^2 + \int_{-\infty}^{+\infty} \frac{a^2}{2\pi(\omega^2 + b^2)} d\omega = 1 \quad \Longrightarrow \quad \underline{z}^2 = 1 - \frac{a^2}{2b}\,, \tag{20}$$

which yields the following constraint inequality for a and b,

$$b > 0 \quad , \quad 0 < a < \sqrt{2b}\,. \tag{21}$$

Consequently, Gaussian stochastic process Z can be viewed as the linear filtering $Z = h * N_\infty$ of the centered Gaussian white noise N_∞ (generalized stochastic process) whose power spectral spectral density function is $S_{N_\infty}(\omega) = 1/(2\pi)$, by the causal and stable linear filter whose frequency response function $\widehat{h}(\omega) = \int_0^{+\infty} e^{-i\omega t} h(t)\, dt = a/(i\omega + b)$ (because $S_Z(\omega) = |\widehat{h}(\omega)|^2\, S_{N_\infty}(\omega)$). Introducing the linear Itô stochastic differential equation,

$$dY(t) = -b\, Y(t)\, dt + a\, dW(t) \quad t > 0\,, \tag{22}$$

with the initial condition $Y(0) = 0$ *a.s.*, in which W is the real-valued normalized Wiener process indexed by $[0, +\infty[$, it can be proved [35, 36] that Eq. (22) has a unique solution $\{Y(t), t \geq 0\}$ such that, for $t_0 \to +\infty$, the stochastic process $\{Y(t), t \geq t_0\}$ is stochastically equivalent to the stationary stochastic process Z (note that Y is not stationary on $\mathbb{R}^+$ for the positive shift, but is asymptotically stationary). In practice, this means that, if t_0 is chosen sufficiently large, Y and Z are the *same* Gaussian stationary and ergodic second-order centered stochastic process for which the power spectral density function is given by Eq. (19). Consequently, Eq. (22) can be used for generating trajectories of stochastic process Z.

3.2. Jitter measurements

There are different types of measures for jitter listed below.

(i) *Absolute.* It is the cycle-to-cycle variation of the fundamental frequency, *i.e*, the average absolute difference between consecutive periods, in seconds, expressed as

$$\mathrm{Jit}_{\mathrm{abs}} = \frac{1}{N-1}\sum_{i-1}^{N-1} \mid T_i - T_{i+1} \mid\,, \tag{23}$$

in which T_i are the lengths of each glottal cycle and N is the number of periods considered.

(ii) *Local*. It is the average absolute difference between consecutive periods, divided by the average period, and given by

$$\mathrm{Jit}_{\mathrm{loc}} = \frac{\dfrac{1}{N-1}\displaystyle\sum_{i-1}^{N-1} \mid T_i - T_{i+1} \mid}{\dfrac{1}{N}\displaystyle\sum_{i=1}^{N} T_i} \,. \tag{24}$$

In general, the value 1.040% is considered as a threshold for the occurrence of a pathology.

(iii) *RAP*. It is the relative average perturbation, the average absolute difference between a period and the average of it and its two neighbors, divided by the average period. In general, 0.680% is considered as a threshold for the occurrence of a pathology.

(iv) *PPQ5*. It is the five-point period perturbation quotient, computed as the average absolute difference between a period and the average of it and its four closest neighbors, divided by the average period. In general, 0.840% is considered as a threshold for pathology; as this number was based on jitter measurements influenced by noise, the correct threshold is probably lower.

(v) *DDP*. It is the five-point period perturbation quotient, computed as the average absolute difference between a period and the average of it and its four closest neighbors, divided by the average period.

4. Simulations

The objective of this section is to simulate voice signals considering the stochastic model proposed and, consequently, with jitter. The subglottal pressure $p_s(t)$ (given in Pa) is a function of time defined (according to the results obtained in [32]) by

$$p_s(t) = \begin{cases} 800 \, \sin\left(5\pi t\right) \ , \ 0 \leq t < 0.1 \\ 800 \ , \ 0.1 \leq t \leq 1.9 \\ 800 \, \sin\left(\dfrac{5\pi t}{9}\right) \ , \ 1.9 < t \leq 2 \,. \end{cases} \tag{25}$$

The graph of function p_s is displayed in Fig. 2. The values of the parameters for the deterministic model are the following: $A_{g0} = 0.05 \times 10^{-2}\, m^2$, $\rho = 0.12\, kg/m^3$, $c_a = 346.3\, m/s$, $\mu = 1.86 \times 10^{-4}\, kg/(m^2 s)$, $m = 0.24 \times 10^{-2}\, kg$,

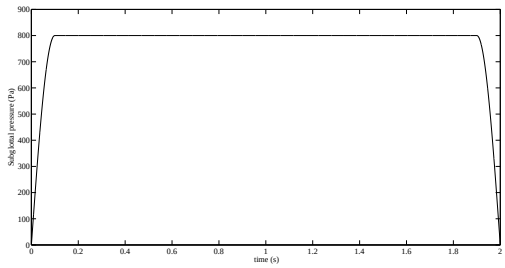

Figure 2: Graph of the subglottal pressure $t \mapsto p_s(t)$.

$\ell = 1.4 \times 10^{-2} m$, d $= 0.3 \times 10^{-2}\, m$, $k_0 = 40\, N/m$, $\underline{k} = 115\, N/m$. For the damping coefficient, it was considered $c = 0$ and $\alpha = 1$, i.e, only during the collision the damping was considered, as in [**?**]. The constants a and b (verifying Eq. (21)) are taken as $a = 40$ and $b = 1,000,000$. The stochastic solver is the Monte Carlo method. The number of realizations is $N = 2 \times 88,200 = 176,400$, with a time step $\Delta t = 1/fs = 1/88200\, s$. The time of simulation for each realization is then given by $N \times \Delta t = 2\, s$. This number is enough in order to warrant the convergence of the solution to a stationary and ergodic stochastic process. Below, we consider only the asymptotic stationary and ergodic solution.

For such an asymptotic solution, the expected values $E\{K(t)\}$ and $E\{K(t)^2\}$ are thus independent of t and can be estimated by

$$E\{K(t)\} = \lim_{t\to+\infty} \overline{K}(t) \quad , \quad \overline{K}(t) = \frac{1}{t}\int_0^t K(t')dt' \,, \tag{26}$$

$$E\{K(t)^2\} = \lim_{t\to+\infty} \overline{K^2}(t) \quad , \quad \overline{K^2}(t) = \frac{1}{t}\int_0^t K(t')^2 dt' \,, \tag{27}$$

which allows for verifying when the ergodicity property is reached (due to the use of stochastic process Y instead of stochastic process Z). Only the first $1,000$ points are considered for plotting the graphs shown in Fig. 3. Since stochastic

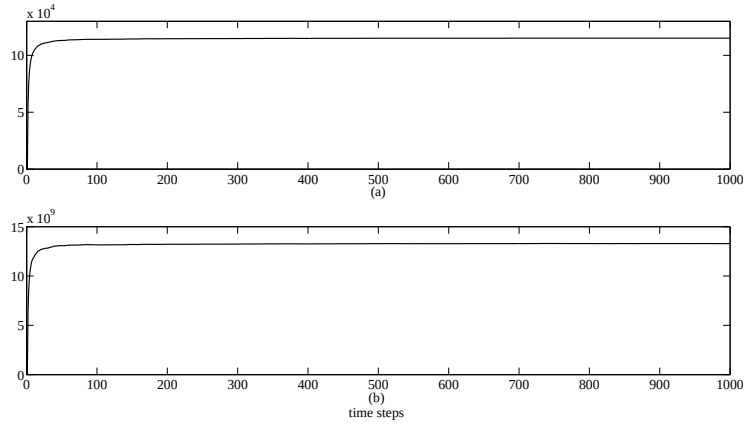

Figure 3: (a) Graph of the mean value $\overline{K}$, in relation to the number of realizations. (b) Graph of the second-moment of K; that is, K^2, in relation to the number of realizations.

process K is stationary, for any fixed t, the cumulative distribution function $F_K(k)$ of the random variable $K(t)$ is independent of t and is such that

$$F_K(k) = \text{Proba}\{K(t) \leq k\} = E\{1_{[-\infty,k]}(K(t))\} \,,$$

$$1_{]-\infty,k]}(k') = \left\{ \begin{array}{l} 1 \ , \text{ if } k' \leq k \\ 0 \ , \text{ if } k' > k. \end{array} \right.$$

Due to the ergodic property of stochastic process K,

$$F_K(k) = \lim_{t\to+\infty} \overline{F_K}(k;t) \quad , \quad \overline{F_K}(k;t) = \frac{1}{t}\int_0^t 1_{]-\infty,k]}(K(t'))dt' \,.$$

Fig. 4 shows the graphs of $t \mapsto \overline{F_K}(k;t)$ for different values of k, considering only the first $1,000$ time steps of the time simulation. In order to illustrate how

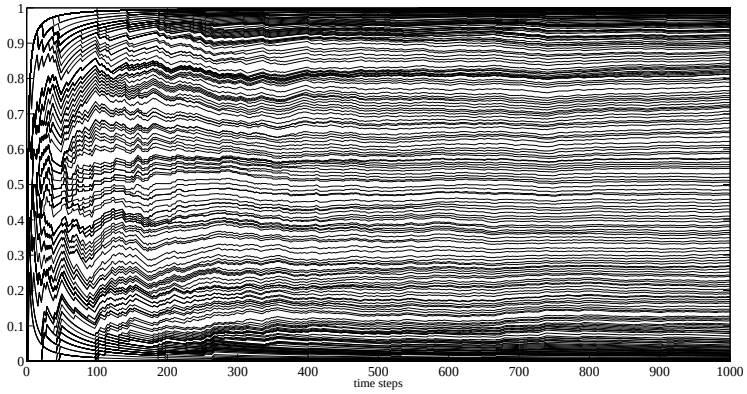

Figure 4: For several values of k, graph of the cumulative distribution function F_K, in relation to the number of realizations. The lower graph corresponds to the minimum value of k, and the upper graph to its maximum value.

the variation of the fundamental frequency (jitter) is achieved, three samples of voice signals are simulated: one considering only the deterministic model, without jitter ($a = 0$), and the two others with different levels of jitter ($a = 40$ and $a = 160$).

Samples of the glottal signal (U_g) simulated are shown in Fig. 5, considering the case in which an /a/ vowel is produced. It can be observed the variation of the amplitude, called shimmer, associated to the jitter. It can be noted that

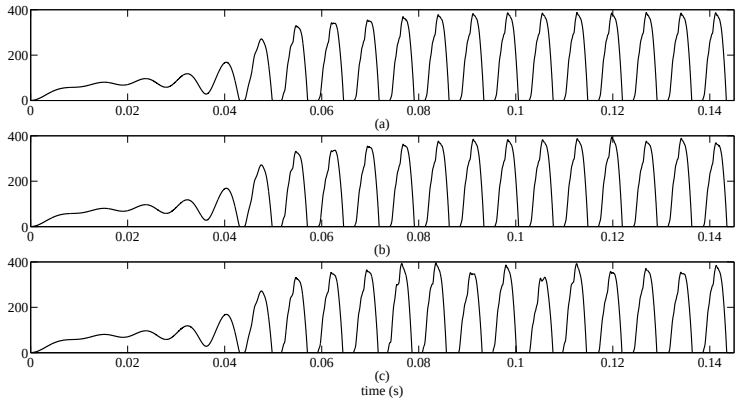

Figure 5: Glottal signal without jitter (a) and with jitter (b) $a = 40$ and (c) $a = 160$, amplitude variation synthesized by the described model.

the variation of the time interval for the glottal pulses varies and this implies variation in its amplitudes. In Fig. 5 there are not variations in the time interval and consequently in the amplitudes; that is, there are neither *jitter* nor *shimmer*.

The stochastic output pressure is calculated by $P_R(t) = \frac{d}{dt}U_R(t)$ and the plots corresponding to the glottal flows (Fig. 5) are given in Fig. 6. Maybe it

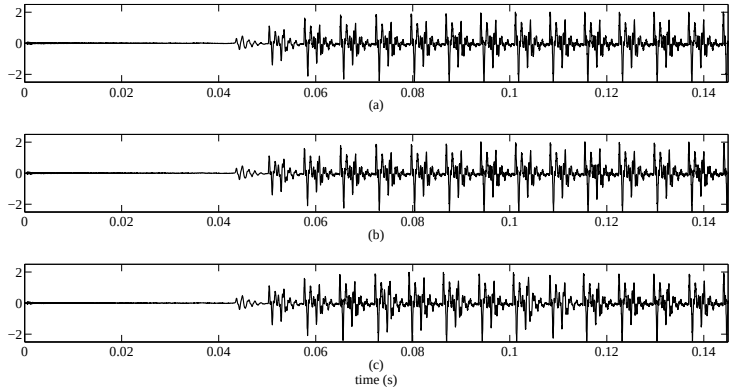

Figure 6: Output pressure signal without jitter (a) and with jitter (b) $a = 40$ and (c) $a = 160$, synthesized by the described model.

is not too easy to observe the variation of the fundamental frequency in these plots. A good way to observe such a variation of the fundamental frequency is to construct the probability density function (pdf) associated to it.

Then, two voice signals are simulated corresponding to different values of a (two different levels of the jitter), and the pdf's are constructed (Fig. 7). Some

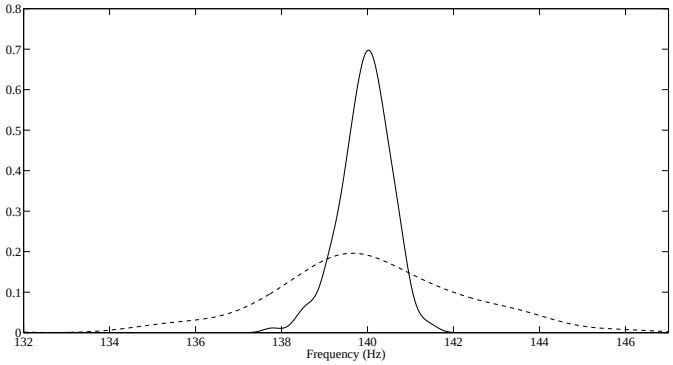

Figure 7: Probability density functions of the fundamental frequency considering two different levels of Jitter: $a = 40$ (continuous line) and $a = 160$ (dashed line).

results obtained with the vowels synthesis, in the deterministic case, and with two different levels of jitter ($a = 0$, $a = 40$ and $a = 160$) can be found and heard in https://www.dropbox.com/s/mwaq3u6ad96po7x/male140Hz.zip?dl=0, for which *detn* corresponds to the case without jitter, N_1 corresponds to $a = 40$ and N_2 corresponds to $a = 160$.

Using the signals synthesized, with two different levels of jitter, the Praat software [37] was used for measuring jitter, considering approximately the same number of periods and during the time the subglottal pressure is constant. Then, Tab. (1) is constructed:

Vowel	level	Abs	Loc	RAP	PPQ5	DDP
A	N_1	43.2e-6	0.6 %	0.361 %	0.367 %	1.084 %
	N_2	182.21e-6	2.553 %	1.522 %	1.504 %	4.566 %
E	N_1	32.122e-6	0.448 %	0.254 %	0.272 %	0.761 %
	N_2	133.35e-6	1.575 %	0.808 %	0.852 %	2.453 %
I	N_1	36.821e-6	0.504 %	0.313 %	0.285 %	0.939 %
	N_2	116.614e-6	1.599 %	0.948 %	1.028 %	2.845 %
O	N_1	32.200e-6	0.483 %	0.271 %	0.318%	0.813 %
	N_2	149.048e-6	2.197 %	1.225 %	1.388 %	3.674 %
U	N_1	41.421e-6	0.577 %	0.360 %	0.313 %	1.080 %
	N_2	190.009e-6	2.650 %	1.659 %	1.414 %	4.977 %

Table 1: Jitter measurements calculated by using Praat software.

5. Conclusions

An approach has been proposed for constructing a stochastic model for creating jitter in a mechanical model that allows for producing voice. Such a model considers the stiffness related to the vocal folds as a stochastic process and the corresponding voice signals have been simulated. The probability density function of the fundamental frequency constructed for different values of the parameters associated to the stochastic model can then be estimated. The comparison between the probability density functions shows that the fundamental frequency has variations in relation to a mean value, showing that jitter has effectively been generated. The voice signals have also been synthesized and it can be perceived the different sounds related to a normal voice, without jitter, and with jitter for two different levels of jitter. One of them is very similar to a normal voice with a low percentage of variation of the fundamental frequency while the other one has a much greater variation that characterizes a hoarse voice that can indicate the occurrence of a pathology.

6. Acknowledgments

This work was supported by CAPES (grant Grant BEX 2623/15-3), CNPq and FAPERJ (APQ1).

7. References

[1] Lieberman P. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. J. Acoust. Soc. Am. 35, 344–353, 1963.

[2] Wong, D., Ito M. R., Cox N. B., Titze, I. R. 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, 1991.

[3] Hirose, H. Clinical Aspects of Voice Disorders. Interuna Publishers, Tokyo, 1998.

[4] Silva, D. G., Oliveira, L. C., Andrea, M. Jitter estimation algorithms for detection of pathological voices. Eurasip Journal on Advances in Signal Processing - Special issue on analysis and signal processing of oesophageal and pathological voices, Article No. 9, 2009.

[5] LiL, Saigusa, H., Hakazawa, Y., Nakamura, T., Komachi, T., Yamaguchi, S., Liu A., Sugisaki, Y., Shinya, E., Shen, H. A pathological sutdy of bamboo nodule of the vocal fold. Journal of Voice, 24 (6), 738–741, 2010.

[6] Henrich N., d'Alessandro C., Doval B., Castellengo M. Open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. Journal of the Acoustical Society of America 117 (3), 1417–1430, 2005.

[7] Kreiman J., Gerrat B. R. Perception of aperiodicity in pathological voice. Acoustical Society of America 117, 2201–2211, 2005.

[8] Londono J., Llorente J. I. G., Lechien N. S. , Ruiz V. O. , Dominguez G. C. An improved method for voice pathology detection by means of a HMM-based feature space transformation. Journal of Pattern Recognition 43 (9), 3100–3112, 2010.

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

[10] Mendoza L.. Kohler, M.; Vellasco, M.; Cataldo, E. Analysis and Classification of Voice Pathologies using glottal signal parameters. Journal of Voice, 1, 1–10, 2015.

[11] Titze I. Principles of voice production. Prentice Hall, Englewood Cliffs, NJ, 1994.

[12] Baken R., Orlikof F. Clinical measurement of speech and voice, second edition. USA, singular, 2000.

[13] Farrus M., Hernando, J. Using jitter and shimmer in speaker verification. Signal Processing, IET, 3(4), 247–257. 2008.

[14] Mendonza L., Vellasco M., Cataldo E., Silva M. B., Apolinario A. A. Classification of Vocal Aging Using Parameters Extracted From the Glottal Signal. Journal of Voice, 21(2), 157–68, 2014.

[15] Wilcox, K. A., MS, Horii, Y., Age and Changes in vocal jitter. Journal of Gerontology, 35 (2),194–198, 2015.

[16] Vieira, M. N., McInnes, F. R., Jack M. A. On the influence of laryngeal pathologies on acoustic and electroglottographic jitter measures. The Journal of the Acoustical Society of America, 111, 1045–1055, 2001.

[17] Zhang Y., Jiang, J. J., Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. Journal of Voice, 22 (1), 1–9, 2008.

[18] Flanagan J., Landgraf L. Self-oscillating source for vocal-tract synthesizers. IEEE Transactions on Audio and Electroacoustics, AU-16 (1), 1968.

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

[20] Titze, I. R., The human vocal cords: a mathematical model. Part I. Phonetica, 28, 129-170.

[21] Story, B. H., Titze, I. R. Voice simulation with a body-cover model of the vocal folds. The Journal of the Acoustical Society of America, 97 (2), 1249–1260, 1995.

[22] Erath, B. D., Zañartu, M., Stewart, K. C., Plesniak, M. W., Sommer, D. E., Peterson, S. D. A review of lumped-element models of voiced speech. Speech Communications, 55, 667–690, 2013.

[23] Deguchi, S., Kawahara, Y., Simulation of human phonation with vocal nodules. American Journal of Computational Mathematics, 1, 189–201, 2011.

[24] Fraile, R., Godino-Llorente, J. I., Kob, M., Simulation of tremulous voices using biomechanical model. Eurasip Journal on Audio, Speech, and Music Processing, 1, 1–12, 2015.

[25] Schoengten J., Stochastic models of Jitter. The Journal of the Acoustical Society of America, 109, 1631–1650, 2001.

[26] Cataldo E., Soize C., Sampaio R., Desceliers C. Probabilistic modeling of a nonlinear dynamical system used for producing voice. Computational Mechanics, 43, 265–275, 2009.

[27] Cataldo E., Soize C., Sampaio R. Uncertainty quantification of voice signal production mechanical model and experimental updating. Mechanical Systems and Signal Processing, 40, 718–726, 2013.

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

[29] Schoengten J., Fraj S., Lucero J. C. Testing the reliability of Grade, Roughness and Breathiness scores by means of synthetic speech stimuli. Logopedics Phoniatrics Vocology, 1, 1–9, 2013.

[30] Pinto, N. B., and Titze, I. R. Unification of perturbation measures in speech signals. The Journal of the Acoustical Society of America, 87, 1278–1289, 1990.

[31] Fant G. The acoustic theory of speech production. Mouton, The Hague, 1963.

[32] Cataldo E., Lucero J., Leta F., Nicolato L. Synthesis of voiced sounds using low-dimensional models of the vocal cords time-varying subglottal pressure. Mechanics Research Communication, 33 (2), 250–260, 2005.

[33] Mauprivez J., Cataldo E., Sampaio R. Artificial neural networks applied to the estimation of random variables associated to a two-mass model for the vocal folds. Inverse problems in Science & Engineering, 20, 209–225, 2012.

[34] Cataldo E., Sampaio R., Lucero J., Soize C. Modeling random uncertainties in voice production using a parametric approach. Mechanics Research Communication, 35, 429–490, 2008.

[35] Krée P., Soize C. Mathematics of Random Phenomena. Reidel, Dordrecht, 1996.

[36] Soize C. The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions. World Scientific, Singapore, 1994.

[37] Praat software website: http://www.fon.hum.uva.nl/praat/.