



**HAL**  
open science

## Jitter generation in voice signals produced by a two-mass stochastic mechanical model

Edson Cataldo, Christian Soize

► **To cite this version:**

Edson Cataldo, Christian Soize. Jitter generation in voice signals produced by a two-mass stochastic mechanical model. *Biomedical Signal Processing and Control*, 2016, 27, pp.87-95. 10.1016/j.bspc.2016.02.003 . hal-01279812

**HAL Id: hal-01279812**

**<https://hal.science/hal-01279812>**

Submitted on 27 Feb 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Jitter generation in voice signals produced by a two-mass stochastic mechanical model

E. Cataldo<sup>a</sup>, C. Soize<sup>b</sup>

<sup>a</sup>*Universidade Federal Fluminense, Applied Mathematics Department and Graduate program in Telecommunications Engineering, Rua Mário Santos Braga, S/N, Centro, Niterói, RJ, CEP: 24020-140, Brazil*

<sup>b</sup>*Université Paris-Est, Laboratoire Modélisation et Simulation Multi Echelle, MSME UMR 8208 CNRS, 5 Bd Descartes, 77454 Marne-La-Vallée, France*

---

## Abstract

Jitter is a phenomenon caused by the perturbation in the length of the glottal cycles due to the quasi-periodic oscillation of the vocal folds in the production of the voice. It can be modeled as a random phenomenon described by the deviations of the glottal cycle length in relation to a mean value. Its study has been developed due to important applications such as aid in identification of voices with pathological characteristics, when its values are large, because a normal voice has naturally a low level of jitter. The aim of this paper is to construct a stochastic model of jitter using a two-mass mechanical model of the vocal folds, assuming complete right-left symmetry of the vocal folds and considering the motion of the vocal folds only in the horizontal direction. The stiffnesses taken into account in the model are considered as stochastic processes and their modeling are proposed. Glottal signals and voice signals are generated with jitter and the probability density function of the fundamental frequency is constructed for several values of the hyperparameters that control the level of jitter.

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

---

## 1. Introduction

The voice production process, in particular in voiced speech production, where vowels are included, is due to the oscillation of the vocal folds, which modulates the airflow coming from the lungs, and the air pulses then generated (called the glottal signal) will be filtered and amplified by the vocal tract and, further, radiated by the mouth.

However, the glottal signal is not exactly periodic. The (small) fluctuation in each glottal cycle length is called jitter and its study is justified because

---

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

it can be used for measuring the voice quality and indicating the presence of pathologies related to the voice or even helping the speech recognition (Titze, 1994; Orlikof, 2000; Farrus and Hernando, 2008; Baken and Manfredi et al., 2012).

Voice pathology can cause increased noise components in the voice signal such as: fundamental frequency and amplitude irregularities and variations with different patterns, sub-harmonic frequency components, turbulent noise, voice breaks and tremors (Koike, 1973; Hirano, 1981; Muahamad and Mulhen, 2014; Wilcox, 2015).

To help pre-diagnosis of pathologies related to the voice, in general, it is necessary to extract not only jitter from the voice signal, but also other measures, like *shimmer* and *HNR* (*harmonic-noise ration*). A large level of jitter can indicate a possibility of a pathology, but to help the identification of the type of pathology, in general, it is not enough. Some authors have discussed this point and proposed strategies to use jitter and other measures to classify pathologies related to the voice (Hirose, 1998; Silva et al., 2009; Vieira, 2001; Zhang, 2008; Mendoza et al., 2014; Mendoza et al., 2015).

Some motivations for developing models of jitter include the discussion about the mechanisms that may cause the movements of the vocal folds to be non periodic. In addition, models of jitter may help to improve naturalness or mimic hoarse voices when synthesizers are used, for example.

Some researchers have developed models to produce voice and even voice with pathological characteristics (Flanagan and Landgraf, 1968; Ishizaka and Flanagan, 1972; Titze, 1973; Story and Titze, 1995). Erath et al. (2013) made a good review of models used to produce voice, including mechanical models and pathological phonation. Shinji Deguchi and Juki Kawahara (2011) present a continuum-based numerical model of phonation to simulate human phonation with vocal nodules. Deguchi and Kawahara (2011) have simulate human phonation with vocal nodules. Fraile et al. (2015) simulate vocal tremor using a high-dimensional discrete vocal fold model. However, all of these models are deterministic.

The objective of this paper is to construct a stochastic model of the vocal folds in order to generate jitter. The underlying deterministic model used is the one introduced by Ishizaka and Flanagan model (1972). Previous works have discussed stochastic mechanical models to produce voice (Cataldo et al., 2009, 2013) considering some parameters modeled as random variables, prior probability distributions have been constructed and then updated.

The approach used here is different and consists in modeling the stiffnesses as stochastic processes. A system of nonlinear stochastic differential equation is then solved and the voiced signals are synthesized for different levels of jitter.

## 2. Deterministic model used

The underlying deterministic model used is the nonlinear two-mass model proposed by Ishizaka and Flanagan 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. The vocal tract is represented by a standard configuration of concatenated tubes (Fant, 1960; Liberman, 1963).

The complete model considered here presents some modifications in relation to the original Ishizaka and Flanagan model (Cataldo et al. , 2005, 2008, 2009). The system of differential equations to be solved, and which is detailed in Appendix A, can be divided in three parts (see Fig. 1 that illustrates a sketch of the model):

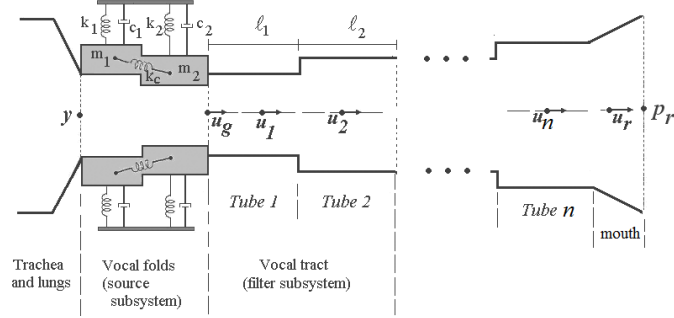


Figure 1: Sketch of the Ishizaka and Flanagan model (1972).

- A nonlinear integro-differential equation for the glottal flow that is coupled with the vocal tract, called the *coupling equation* (see Eqs. (A.1) to (A.9)). It 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.
- A system of linear integro-differential equations related to the sound acoustic propagation through the vocal tract and called the *sound acoustic propagation equation* (see Eq. (A.10)). It is constituted of  $n + 1$  scalar linear integro-differential equations for which the  $n + 1$  unknown time-dependent functions are the real-valued functions  $t \mapsto u_1(t), \dots, u_n(t), u_r(t)$ . The function  $t \mapsto p_r(t)$  is the sound acoustic pressure through the mouth.
- A system of nonlinear differential equations related to the dynamics of the vocal folds and called the *vocal folds dynamic equation* (see Eq. (A.11)). It is a system of nonlinear differential equations whose unknown time-dependent functions are the real-valued functions  $t \mapsto x_1(t)$  that is the displacement of the mass 1 and  $x_2(t)$  is the displacement of the mass 2, corresponding together to the vocal folds. For all  $t$ ,  $x_1(t)$  and  $x_2(t)$  are generated by the *vocal folds dynamic equations*. The solutions  $x_1$  and  $x_2$  of such equations are constructed for all  $t$  and are used as follows:

- The collision of the vocal folds starts at a time  $t_1$  when  $x_1(t)$  reaches a given critical value  $x_{01}$  (defined in Appendix A). This process continues, and at a time  $t_2$ ,  $x_2(t)$  reaches a given critical value  $x_{02}$ . At this time, the two parts of the model corresponding to the vocal folds are in collision. Then, at a third time  $t_3$ , the upper part of the vocal folds leaves the contact, but the mass lower part continues in contact (mass  $m_2$ ). Finally, at a time  $t_4$ , the lower part leaves the collision.
- While the glottis is closed (during time  $t_1$  up to  $t_4$ ),  $u_g$  and  $du_g/dt$  remain zero,  $u_g(\tau) = du_g(\tau)/d\tau = 0$ , the stiffnesses and the dampings of the models are modified in the vocal folds dynamic equation, but the propagation of sound goes on in the vocal tract .

### 3. Jitter modeling

The jitter phenomenon has been studied for many researchers in order to be used as a measure for discriminating voices with pathological characteristics or even aging voice characteristics. Since the second middle of the eighties, the research about jitter has become less clinical and more methodological. The methods used to measure jitter, and other characteristics of the voice signal, changed considerably, and one of the first works for quantifying the jitter was proposed by Lieberman (1963). 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. In general, the 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). Some works discussing jitter and its application (Henrich et al., 2005; Kreiman and Gerratt, 2005; Londono et al., 2010; Dejonckerea et al., 2012; Mendoza et al., 2014). 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. Only a few authors consider stochastic models (Schoengten and De Guchteneere, 1997; Schoengten, 2001; Cataldo et al., 2009; Shoengten, 2013; Cataldo et al., 2013).

#### 3.1. System of stochastic equations

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 stiffnesses  $k_1$ ,  $k_2$ , and  $k_c$  are constant, it is coherent to introduce a stochastic modeling of these stiffnesses in replacing them by stochastic quantities  $K_1(t)$ ,  $K_2(t)$ , and  $K_c(t)$ . We thus introduce the stationary stochastic processes  $\{K_1(t), t \in \mathbb{R}\}$ ,  $\{K_2(t), t \in \mathbb{R}\}$ , and  $\{K_c(t), t \in \mathbb{R}\}$ , with values in  $\mathbb{R}^+$ , which models the stiffnesses  $k_1$ ,  $k_2$  and  $k_c$  of the deterministic model. Consequently, the deterministic equations (that are detailed in Appendix A) for the deterministic time functions  $u_g(t)$ ,  $u_1(t), \dots, u_n(t)$ ,  $u_R(t)$ ,  $x_1(t)$ , and  $x_2(t)$ , become stochastic equations for the stochastic processes denoted by  $U_g(t)$ ,

$U_1(t), \dots, U_n(t), U_R(t), X_1(t),$  and  $X_2(t)$ . The corresponding stochastic equations are detailed in Appendix B.

### 3.2. Construction of a stochastic model for $[K(t)]$

For all fixed  $t$ , the real-valued random variables  $K_1(t), K_2(t),$  and  $K_c(t)$  must be positive as explained above. Nevertheless, this property is not sufficient for ensuring the coherence of the construction of the stochastic model of the stochastic stiffnesses. It is necessary to write that, for all fixed  $t$ , the random matrix  $[K(t)]$  defined by

$$\begin{bmatrix} K_{11}(t) & K_{12}(t) \\ K_{21}(t) & K_{22}(t) \end{bmatrix} \quad (1)$$

in which

$$K_{11}(t) = K_1(t) + K_c(t), \quad (2)$$

$$K_{22}(t) = K_2(t) + K_c(t), \quad (3)$$

$$K_{12}(t) = K_{21}(t) = -K_c(t), \quad (4)$$

is a random matrix with values in the set  $\mathbb{M}_2^+(\mathbb{R})$  of all the positive-definite real  $(2 \times 2)$  matrices. Consequently, the construction of the stochastic model of stiffnesses consists in constructing a stochastic process  $\{[K(t)], t \in \mathbb{R}\}$  with values in  $\mathbb{M}_2^+(\mathbb{R})$ , which can then be not Gaussian (due to the positivity property). It can then be deduced the stochastic model of the dependent non-Gaussian stochastic processes  $\{K_1(t), t \in \mathbb{R}\}, \{K_2(t), t \in \mathbb{R}\},$  and  $\{K_c(t), t \in \mathbb{R}\},$  which are such that, for all fixed  $t$ ,

$$K_c(t) = -K_{12}(t), \quad (5)$$

$$K_1(t) = K_{11}(t) - K_c(t), \quad (6)$$

$$K_2(t) = K_{22}(t) - K_c(t). \quad (7)$$

In the framework of the jitter effects, matrix-valued stochastic process  $\{[K(t)], t \in \mathbb{R}\}$  is chosen as a stationary stochastic process that must be of second-order for physical reason. In order to ensure the positivity property, for all  $t$  fixed in  $\mathbb{R}$ , random matrix  $[K(t)]$  is written as

$$[K(t)] = [\underline{L}_K]^T [G(t)] [\underline{L}_K], \quad (8)$$

in which  $[G(t)]$  is a random matrix with values in  $\mathbb{M}_2^+(\mathbb{R})$  such that its mean values  $E\{[G(t)]\} = [I_2]$  is the identity matrix. In Eq. (8),  $[\underline{L}_K]$  is an upper triangular real  $(2 \times 2)$  matrix such that the mean value  $[\underline{K}] = E\{[K(t)]\} \in \mathbb{M}_2^+(\mathbb{R})$  of the random matrix  $[K(t)]$  is written as

$$[\underline{K}] = [\underline{L}_K]^T [\underline{L}_K], \quad (9)$$

which is the Cholesky decomposition of the positive-definite real matrix  $[K]$ . As explained in (Soize, 2012), for all  $t$  in  $\mathbb{R}$ , the random matrix  $[G(t)]$  must be written as

$$[G(t)] = \frac{1}{1+\epsilon}([G_0(t)] + \epsilon[I_2]), \quad (10)$$

in which  $\epsilon > 0$  is a given regularization parameter and where  $\{[G_0(t)], t \in \mathbb{R}\}$  is a second-order stationary stochastic process with values in  $\mathbb{M}_2^+(\mathbb{R})$  such that

$$E\{[G_0(t)]\} = [I_2]. \quad (11)$$

The stochastic model is constructed as explained in (Soize, 2005, 2006). For all fixed  $t$ , random matrix  $[G_0(t)]$  is written as

$$[G_0(t)] = [L_0(t)]^T [L_0(t)], \quad (12)$$

in which  $[L_0(t)]$  is a random upper triangular matrix whose diagonal entries are positive-valued random variable,

$$[L_0(t)] = \begin{bmatrix} L_0(t)_{11} & L_0(t)_{12} \\ 0 & L_0(t)_{22} \end{bmatrix}. \quad (13)$$

For  $j = 1, 2$ , the positive-valued random variable  $L_0(t)_{jj}$  is written as

$$L_0(t)_{jj} = \sigma \sqrt{2h(\alpha_j, U_{jj}(t))}, \quad (14)$$

in which the mapping  $u \mapsto h(\alpha, u)$  from  $\mathbb{R}$  into  $]0, +\infty[$  is defined in (Soize, 2006) and where

$$L_0(t)_{12} = \sigma U_{12}(t). \quad (15)$$

In Eqs. (14) and (15), the different quantities are defined hereinafter.

- $\sigma = \delta/\sqrt{3}$  in which  $\delta$  is a hyperparameter of the stochastic model, which allows for controlling the level of the statistical fluctuations, and which must satisfy  $0 \leq \delta < \sqrt{3/7}$  (note that  $\delta = 0$  will correspond to the deterministic case, there are no statistical fluctuations).
- For  $j = 1, 2$ , the real-valued parameter  $\alpha_j$  is such that  $\alpha_j = (1 - j + \delta/\sqrt{3})/2$ .
- $\{U_{11}(t), t \in \mathbb{R}\}$ ,  $\{U_{12}(t), t \in \mathbb{R}\}$ , and  $\{U_{22}(t), t \in \mathbb{R}\}$  are three independent copies of a Gaussian stationary second-order centered normalized stochastic process  $\{U(t), t \in \mathbb{R}\}$  such that, for all fixed  $t$ ,  $E\{U(t)\} = 0$  and  $E\{U^2(t)\} = 1$ .
- The Gaussian stationary centered stochastic process  $U$  is thus completely defined by its autocorrelation function  $R_U(\tau) = E\{U(t+\tau)U(t)\}$  which is chosen as  $R_U(\tau) = \exp\{-b|\tau|\}$ , in which  $b$  is the second positive-valued

hyperparameter of the stochastic model (therefore, for all  $b > 0$ , we have effectively  $E\{U^2(t)\} = 1$ ). Therefore, the corresponding power spectral density function,  $S_U(\omega)$ , of stochastic process  $U$  is such that

$$R_U(\tau) = \int_{\mathbb{R}} e^{i\omega\tau} S_U(\omega) d\omega \quad , \quad S_U(\omega) = \frac{1}{2\pi} \frac{2b}{\omega^2 + b^2} ,$$

The correlation time,  $\tau_c$ , is defined by

$$\tau_c = \int_0^{+\infty} \exp\{-b|\tau|\} d\tau = \frac{1}{b} .$$

Consequently, hyperparameter  $b$  is related to the correlation time.

- The copies of Gaussian stationary stochastic process  $U$  will be generated in introducing a linear Itô stochastic differential equation.

Taking into account the definition of stochastic process  $U$ , a realization (sample path or trajectory) of stochastic process  $U$  for  $t \geq 0$  can be obtained by solving the following linear Itô stochastic differential equation,

$$dU(t) = -bU(t) dt + \sqrt{2b} dW(t) \quad , \quad t > 0 , \quad (16)$$

with the random initial condition

$$U(0) = U_0 . \quad (17)$$

In Eq. (16),  $\{W(t), t \geq 0\}$  is the normalized Wiener stochastic process (Krée, 1986; Soize, 1994). The random initial condition  $U_0$  is a normalized Gaussian real-valued random variable ( $E\{U_0\} = 0$  and  $E\{U_0^2\} = 1$ ), which is independent of Wiener process  $\{W(t), t \geq 0\}$ . Note that the construction of the three copies  $U_{11}$ ,  $U_{12}$ , and  $U_{22}$  of stochastic process  $U$  is obtained by solving Eq. (16) with Eq. (17) by using, for  $\{W, U_0\}$ , three independent copies  $\{W, U_0\}_{11}$ ,  $\{W, U_0\}_{12}$ , and  $\{W, U_0\}_{22}$  of the random variable  $U$  and of the Wiener stochastic process  $W$ .

#### 4. Simulations and production of voice signals with jitter

The objective of this section is to perform the numerical simulation of the system of nonlinear stochastic integro-differential equations in order to produce voice signals with jitter. The subglottal pressure  $p_s(t)$  is a given deterministic function of time (according to the results obtained in (Cataldo et al., 2005)).

The subglottal pressure  $p_s(t)$  is deterministic and, without jitter, the glottal signal  $u_g(t)$  is a periodic signal denoted by  $u_g^p(t)$ . Then, if  $u_g^p(t)$  was considered as a stochastic process, denoted by  $U_g(t)$ , then  $U_g(t)$  that would be a periodic function, would not be a stationary stochastic process on the real line, and consequently, would not be ergodic. However, considering the presence of jitter



in the voice signals generated, the stochastic process  $U_g(t)$  could be written as  $U_g(t) = u_g^p(t) + U_g^s(t)$ , in which  $u_g^p(t)$  would be a periodic function and where the perturbation,  $U_g^s(t)$ , would be assumed to be a stationary and ergodic stochastic process. It is not easy to prove that the perturbation  $U_g^s(t)$  is stationary and ergodic. However, it is reasonable to introduce such a hypothesis that allows for giving a sense to the construction of the fundamental frequency as a random variable (by construction, the fundamental frequency is considered as a random variable that does not depend on time). In such a context, an important point is the construction of the estimation of the probability density function of this random fundamental frequency by using a time averaging related to a trajectory of the stochastic response. In addition, the convergence analysis of this probability density function with respect to the duration of the time averaging must be carried out. Such an analysis has been performed and the convergence has been obtained. Such properties can be viewed in looking at the results, which will be shown further in this section.

It is important to say that there are three hyperparameters,  $\epsilon > 0$ ,  $b > 0$ , and  $\delta > 0$ , that have been introduced in Section 3.2 devoted to the construction of the stochastic modeling of the stiffnesses. Hyperparameter  $\epsilon$  is a regularization parameter that can be chosen small and thus is fixed to the value  $10^{-6}$ . Hyperparameter  $b$  allows for controlling the correlation time  $\tau_c$ , and thus, for controlling the bandwidth of power spectral density function  $S_U$ . Therefore the color of the stiffnesses stationary signals, which impacts the coloring voice is controlled by this hyperparameter  $b$ . Hyperparameter  $\delta$  controls the level of the statistical fluctuations that are injected in the stochastic model through the stochastic process  $[K(t)]$ . For the stochastic model proposed,  $\delta$  must be such that  $0 < \delta < \sqrt{37}$  (Soize, 2005).

For the numerical simulations carried out with the deterministic model and with the stochastic model, the following nominal values of the parameters are:  $A_{g0} = 0.05 \times 10^{-2} m^2$ ,  $\rho = 0.12 kg/m^3$ ,  $c_a = 346.3 m/s$ ,  $\mu = 1.86 \times 10^{-4} kg/(m^2 s)$ ,  $m_1 = 0.125 \times 10^{-3} kg$ ,  $m_2 = 0.025 \times 10^{-3} kg$ ,  $\ell = 1.4 \times 10^{-2} m$ ,  $d_1 = 0.25 \times 10^{-2} m$ ,  $d_2 = 0.05 \times 10^{-2} m$ .

Concerning the numerical simulations carried out with the stochastic model, the mean values  $\underline{K}_1$ ,  $\underline{K}_2$ , and  $\underline{K}_c$ , which are the entries of the matrix  $[\underline{K}]$  are  $\underline{K}_1 = 371 N/m$ ,  $\underline{K}_2 = 64 N/m$ , and  $\underline{K}_c = 105 N/m$ . The stochastic solver used is the Monte Carlo method for which only one realization is computed and for which the ergodic property is used for the estimates of the statistics of interest. The realization is sampled with a sampling frequency  $f_s = 88,200 Hz$ , which yields a time step  $\Delta t = 1/(2f_s) s$ . The time of simulation for the realization is then given by  $N \times \Delta t = 0.5 s$ . This number is enough in order to ensure the convergence of the solution to a stationary and ergodic stochastic process.

At first, we take  $b = 200$  and three values of  $\delta$  that are  $\delta = 0.001$  (very small dispersion equivalent to the case without jitter),  $\delta = 0.05$  and  $\delta = 0.1$  for two cases with a significant jitter effect. Then, the three realizations of the glottal signal (stochastic process  $U_g$ ) are shown in Fig. 2 corresponding to these three values of  $(b, \delta)$ . In Fig. 2, it can be observed the variation of the amplitude,

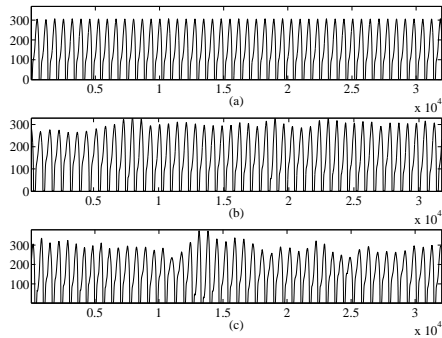


Figure 2: For  $b = 200$ , realization of the glottal signal ( $U_g$ ) for  $\delta = 0.001$  (equivalent to the case without jitter, (a)), and for  $\delta = 0.05$  (b) and  $0.1$  (c) (with jitter).

called shimmer, associated to the jitter. Let us consider the duration between to successive times, the first one corresponding to the instant the glottis opens and the second one the instant for it closes completely. This duration, denoted by  $T_{\text{fund}}$  is a random variable, and its inverse is defined as the fundamental frequency that is the random variable  $F_{\text{fund}} = 1/T_{\text{fund}}$ . For each value of  $(b, \delta)$ , the probability density function of  $F_{\text{fund}}$  is estimated using the realization of the stochastic process  $U_g$  that is numerically simulated, and is shown in Fig. 3 for each value of  $(b, \delta)$ .

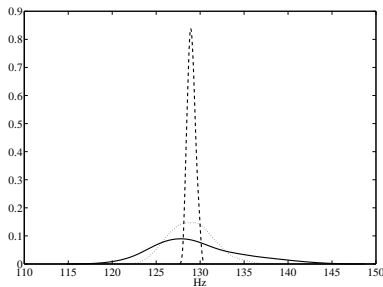


Figure 3: For  $(b, \delta) = (200, 0.001)$  (dashed line),  $(200, 0.05)$  (dotted line),  $(200, 0.1)$  (full line), probability density function of random fundamental frequency  $F_{\text{fund}}$ , estimated using each realization displayed in Fig. 2

In order to analyze the effects of the variation of hyperparameter  $b$  (that controls the time correlation) on the glottal signal,  $\delta$  is fixed to the value 0.1 for  $b = \{50, 100, 150\}$ . Fig. 4 displays the realizations of stochastic process  $U_g$  for these three values of  $(b, \delta)$ .

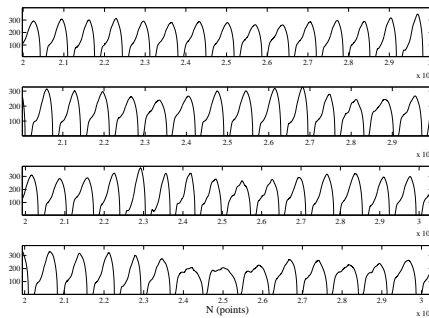


Figure 4: For  $\delta = 0.1$ , realization of the glottal signal ( $U_g$ ) with jitter for  $b = 50$  (a), 100 (b), and 150 (c).

For each value  $(b, \delta)$  of the hyperparameters, the probability density function of  $F_{\text{fund}}$ , which is estimated using the realization of the stochastic glottal signal ( $U_g$ ) shown in Fig. 4, is displayed in Fig. 5.

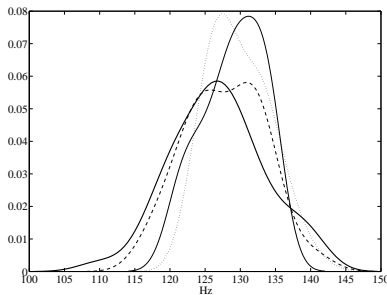


Figure 5: For  $(b, \delta) = (50, 0.1)$  (dashed line),  $(100, 0.1)$  (dotted line),  $(150, 0.1)$  (full line), probability density function of random fundamental frequency  $F_{\text{fund}}$ , estimated using each realization displayed in Fig. 4.

In Figs. 3 and 5, it can be seen that the probability density function of the random fundamental frequency,  $F_{\text{fund}}$ , is very sensitive to the values of hyperparameters  $b$  (controlling the color of the signal) and  $\delta$  (controlling the level of the statistical fluctuations). This means that the stochastic model proposed has the capability to generate a large family of voice signals with a jitter effect. Figs. 2 and 4 show that the realizations of the glottal signal is sensitive to the value of

$\delta$ , but would seem to be less or not sensitive to the value of  $b$ . In fact, the voice signals are effectively sensitive to the value of  $b$  as it can be listened in:

<https://www.dropbox.com/s/4vkwrxjenxlix4ol/IF72SoundsJitter.zip?dl=0>

The utterance of vowel /a/, simulated with the stochastic model with jitter, for the different values of  $b$  and  $\delta$ , can be listened using the internet link given before. The cases considered for the sounds are the following ones:

- $b = 50$  and  $\delta = 0.1$ : *Male-b50delta01*
- $b = 100$  and  $\delta = 0.1$ : *Male-b100delta01*
- $b = 150$  and  $\delta = 0.1$ : *Male-b150delta01*
- $b = 200$  and  $\delta = 0.001$ : *Male-b200delta0001*
- $b = 200$  and  $\delta = 0.1$ : *Male-b200delta01*
- $b = 200$  and  $\delta = 0.05$ : *Male-b200delta005*
- Deterministic case: *Male-deterministic.wav*

#### 4.1. Jitter measurements

The jitter features are also calculated using the Praat voice analysis software: <http://www.fon.hum.uva.nl/praat/> that reports different types of measures for jitter listed below. The terminology used below corresponds to the one used by this software.

Let  $T_1, \dots, T_N$  be  $N$  realizations of the random variable  $T_{\text{fund}}$ , which correspond to successive values extracted from one realization of the stochastic glottal signal  $U_g$  (the total duration of the simulated realization of the glottal signal is thus  $T_1 + T_2 + \dots + T_N$ ). Consequently, the statistical estimation of the mean value of the random variable  $T_{\text{fund}}$  can be written as

$$\underline{T}_{\text{fund}} = \frac{1}{N} \sum_{i=1}^N T_i. \quad (18)$$

Below, five quantities (introduced by the literature) are considered: Absolute jitter, local jitter, RAP, PPQ5, and DDP. In general, the value 1.040% for the local jitter is considered as a threshold for the occurrence of a pathology, the value 0.680% for RAP is considered as a threshold for the occurrence of a pathology, and the value 0.840% for PPQ5 is considered as a threshold for pathology (as this number was based on jitter measurements influenced by noise, the correct threshold is probably lower). For instance the absolute jitter and the local jitter are defined as follows:

$$\text{Absolute jitter: } \text{Jit}_{\text{abs}} = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}|.$$

$$\text{Local jitter: } \text{Jit}_{\text{loc}} = \text{Jit}_{\text{abs}} / \underline{T}_{\text{fund}}.$$

*Summary of the jitter measures.* Using the signals synthesized, with different levels of jitter, the Praat software was used for measuring jitter, considering approximately the same number of periods and during the time the subglottal pressure is constant. The results obtained are given in Tab. (1). The measures

b	$\delta$	Abs	Loc	RAP	PPQ5	DDP
200	0.001	$19.145 \times 10^{-6} s$	0.247 %	0.108 %	0.187 %	0.323 %
50	0.1	$213.169 \times 10^{-6} s$	<b>2.716</b> %	<b>1.513</b> %	<b>2.132</b> %	4.538 %
100	0.1	$255.873 \times 10^{-6} s$	<b>3.384</b> %	<b>1.950</b> %	<b>2.184</b> %	5.849 %
150	0.1	$214.653 \times 10^{-6} s$	<b>3.008</b> %	<b>1.732</b> %	<b>2.165</b> %	5.197 %
200	0.1	$267.724 \times 10^{-6} s$	<b>3.867</b> %	<b>2.157</b> %	<b>3.116</b> %	6.472 %

Table 1: Jitter measurements calculated by using Praat software. The numbers in bold represent those indices that have overpast the pathology thresholds.

must not be considered as having precise values because they depend upon the length of the signal window considered. However, it can easily be seen that, for  $b = 200$  and  $\delta = 0.001$ , the voice signal produced with the stochastic model corresponds to a normal voice (without characteristics of pathologies), while for all the other cases, the voice signal produced corresponds to voices with characteristics of pathologies. The mean value of the random fundamental frequency  $F_{\text{fund}}$  of the voice signal considered is about  $130 Hz$ . Then, when  $b$  is near to this value ( $b = 150$ ), it can be seen that the values of jitter are less than when  $b$  is 100 or 200. The correlation time (equal to  $1/b$ ) is near to the mean value  $T_{\text{fund}}$  of the random fundamental period.

## 5. Conclusions

An approach has been proposed for constructing a stochastic model for creating jitter in a two-mass mechanical model used for producing voiced sounds. The stiffnesses of the model have been considered as stochastic processes, a stochastic stiffness matrix was constructed and its modeling proposed. The construction of the corresponding stochastic matrix permitted to generate jitter and its level can be controlled through different values of three hyperparameters. The probability density function of the so-called random fundamental frequency has been constructed and compared for different values of the hyperparameters. The voice signals have been synthesized with different levels of jitter, which could be easily perceived when they are listened. The model proposed gives excellent results for producing a jitter effect in the voice signals.

## 6. Acknowledgments

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

## Appendix A. Deterministic Model

*Coupling equation.* The coupling nonlinear equation relates  $u_g$  and  $u_1$ , for which coefficients depend on  $x_1(t)$  and  $x_2(t)$ , is written as

$$\begin{aligned} & (R_{k1}(x_1(t)) + R_{k2}(x_2(t))) |u_g(t)| u_g(t) \\ & + (R_{v1}(x_1(t)) + R_{v2}(x_2(t))) u_g(t) \\ & + (L_{g1}(x_1(t)) + L_{g2}(x_2(t))) \frac{du_g}{dt} + L_1 \frac{du_g}{dt} \\ & + R_1 u_g(t) + \frac{1}{c_1} \int_0^t \{u_g(\tau) - u_1(\tau)\} d\tau - p_s(t) = 0, \end{aligned} \quad (\text{A.1})$$

where the coefficients  $R_{k1}(x_1(t))$ ,  $R_{k2}(x_2(t))$ ,  $\mathcal{R}_{k1}(x_1(t), u_g(t))$ ,  $\mathcal{R}_{k2}(x_2(t), u_g(t))$ ,  $L_{g1}(x_1(t))$  and  $L_{g2}(x_2(t))$  are defined by

$$R_{v1}(x_1(t)) = 12 \mu d_1 \ell^2 [A_{g1}(x_1(t))]^{-3}, \quad (\text{A.2})$$

$$R_{v2}(x_2(t)) = 12 \mu d_2 \ell^2 [A_{g2}(x_2(t))]^{-3}, \quad (\text{A.3})$$

$$\mathcal{R}_{k1}(x_1(t), u_g(t)) = 0.19 \rho |u_g(t)| [A_{g1}(x_1(t))]^{-2}, \quad (\text{A.4})$$

$$\mathcal{R}_{k2}(x_2(t), u_g(t)) = \left[ \frac{0.5 - \frac{A_{g2}(x_2(t))}{A_1} \left(1 - \frac{A_{g2}(x_2(t))}{A_1}\right)}{A_{g2}(x_2(t))^2} \right], \quad (\text{A.5})$$

$$L_{g1}(x_1(t)) = \rho d_1 [A_{g1}(x_1(t))]^{-1}, \quad (\text{A.6})$$

$$L_{g2}(x_2(t)) = \rho d_2 [A_{g2}(x_2(t))]^{-1}, \quad (\text{A.7})$$

in which  $A_{g1}(x_1(t))$  and  $A_{g2}(x_2(t))$  are the glottal areas such that

$$A_{g1}(x_1(t)) = A_{g01} + \ell x_1(t), \quad A_{g2}(x_2(t)) = A_{g02} + \ell x_2(t), \quad (\text{A.8})$$

with  $\ell$  the length of each vocal fold, and  $A_{g01}$  and  $A_{g02}$  are such that the critical value  $x_{01}$  and  $x_{02}$  are written as

$$x_{01} = -A_{g01}/\ell, \quad x_{02} = -A_{g02}/\ell. \quad (\text{A.9})$$

In Eqs. (A.2) to (A.7),  $\mu$  is the air kinematic viscosity,  $d_1$  and  $d_2$  are the vocal fold thicknesses, and  $\rho$  is the air density. In Eq. (A.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  $t_1$ , Eq. (A.1) becomes

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

up to time  $t_4$ .

*Sound acoustic propagation equation.* We consider the configuration of the vocal tract proposed by (Fant, 1960). The vocal tract is represented as a transmission line of  $n$  cylindrical tubes, for which the section areas are  $A_1, \dots, A_n$  (the last area  $A_n$  corresponds to the mouth) and where the tube lengths are  $\ell_1, \dots, \ell_n$ .

For  $i = 1, \dots, n$ , 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. There are two eigenfrequencies related to the masses, which characterize the dynamics of the vocal folds. The model is constructed in choosing  $\omega = (k_1 / m_1)^{1/2}$  as the eigenfrequency controlling the dynamics. 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. (A.1) by time-dependent function  $u_1$ , are written as

$$\left\{ \begin{array}{l} (L_1 + L_2) \frac{du_1(t)}{dt} + (r_1 + r_2) u_1(t) + \frac{1}{c_2} \int_0^t \{u_1(\tau) - u_2(\tau)\} d\tau + \\ \frac{1}{c_1} \int_0^t \{u_1(\tau) - u_g(\tau)\} d\tau = 0, \\ \\ (L_i + L_{i+1}) \frac{du_i(t)}{dt} + (r_i + r_{i+1}) u_i(t) + \\ \frac{1}{c_{i+1}} \int_0^t \{u_i(\tau) - u_{i+1}(\tau)\} d\tau + \\ \frac{1}{c_i} \int_0^t \{u_i(\tau) - u_{i-1}(\tau)\} d\tau = 0 \quad , \quad i = 2, \dots, n-1, \\ \\ (L_n + L_R) \frac{du_n(t)}{dt} + r_n u_n(t) - L_R \frac{du_R(t)}{dt} + \\ \frac{1}{c_n} \int_0^t \{u_n(\tau) - u_{n-1}(\tau)\} d\tau = 0, \\ \\ L_R \frac{d(u_R(t) - u_n(t))}{dt} + r_R u_R(t) = 0. \end{array} \right. \quad (\text{A.10})$$

*Vocal folds dynamic equation.* The system of nonlinear differential equations in  $x_1$  and  $x_2$  for the vocal folds dynamics, which is coupled with the vocal-tract (through  $u_g(t)$ ) are written as

$$\left\{ \begin{array}{l} m_1 \frac{d^2 x_1(t)}{dt^2} + r_1 \frac{dx_1(t)}{dt} + s_1(x_1) + k_c(x_1 - x_2) = F_1(x_1(t), u_g(t)), \\ m_2 \frac{d^2 x_2(t)}{dt^2} + r_2 \frac{dx_2(t)}{dt} + s_2(x_2) + k_c(x_2 - x_1) = F_2(x_2(t), u_g(t)). \end{array} \right. \quad (\text{A.11})$$

When the glottis is open, that is,  $x_1(t) > x_{01}$  and  $x_2(t) > x_{02}$ , then

$$F_1(x_1(t), u_g(t)) = P_1(x_1(t), u_g(t)) \ell d_1,$$

with

$$P_1(x_1(t), u_g(t)) = p_s(t) - 1.37 \frac{\rho}{2} \left( \frac{u_g(t)}{A_{g1}(x_1(t))} \right)^2 - \frac{1}{2} \left( R_{v1}(x_1(t)) u_g(t) + L_{g1}(x_1(t)) \frac{du_g}{dt} \right)$$

and

$$F_2(x_2(t), u_g(t)) = P_2 \ell d_2,$$

with

$$P_2(x_1(t), x_2(t), u_g(t)) = P_1(x_1(t), u_g(t)) - \frac{1}{2} \left\{ (R_{v1}(x_1(t)) + R_{v2}(x_2(t))) u_g(t) + (L_{g1}(x_1(t)) + L_{g2}(x_2(t))) \frac{du_g}{dt} \right\} - \frac{\rho}{2} u_g(t)^2 \left( \frac{1}{A_{g2}(x_2(t))^2} - \frac{1}{A_{g1}^2(x_1(t))} \right).$$

We also have

$$s_i(x_i) = k_i(x_i + \eta_{k_i} x_i^3) \quad , \quad i = 1, 2.$$

It can then be seen that the vocal folds dynamic equation depends on the two masses  $m_1$  and  $m_2$  and on three stiffness parameters that are  $k_1$ ,  $k_2$ , and  $k_c$ .

When the glottis is closed,  $x_i \leq -\frac{A_{g0i}}{2\ell}$  and

$$s_i(x_i) = k_i(x_i + \eta_{k_i} x_i^3) + h_1 \left\{ \left( x_i + \frac{A_{g0i}}{2\ell} \right) + \eta_{k_i} \left( x_i + \frac{A_{g0i}}{2\ell} \right)^3 \right\} \quad , \quad i = 1, 2.$$

The values for  $F_1$  and  $F_2$  depend on the configuration of the glottis when it is closed ( $u_g(t) = 0$  and  $\frac{du_g(t)}{dt} = 0$ ):

- If  $x_1(t) \leq x_{01}$  and  $x_2(t) > x_{02}$ ,

$$F_1 = P_s \ell d_1 \text{ and } F_2 = 0. \tag{A.12}$$

- If  $x_1(t) > x_{01}$  and  $x_2(t) \leq x_{02}$ ,

$$F_1 = P_s \ell d_1 \text{ and } F_2 = P_s \ell d_2. \tag{A.13}$$

- If  $x_1(t) \leq x_{01}$  and  $x_2(t) \leq x_{02}$ ,

$$F_1 = P_s \ell d_1 \text{ and } F_2 = 0. \tag{A.14}$$



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

- In Eqs. (A.1) to (A.14), the values of the parameters can be found in (Cataldo et al., 2008 and 2009; Mauprivez et al., 2012).
- Eqs. (A.1) to (A.14) constitute a set of nonlinear coupled equations.
- In Eq. (A.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 (see Cataldo et al., 2008 and 2009, and see also a discussion in Sciamarella et al. 2004; Lucero, 2005).
- 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 defined by Eqs. (A.1) to (A.14) can effectively produce the phonation using only the few parameters introduced in the model. It is important to note that there is a range for the values of the 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 non periodic manner; it means that the voice signal generated can be chaotic, for example. The values of 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. (A.1) to (A.14).
- Preliminary studies have shown that small variations of the 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 parameters, in particular for the masses  $m_1$  and  $m_2$ , and for the stiffnesses  $k_1$ ,  $k_2$ , and  $k_c$  of the vocal folds.
- In previous works (Cataldo et al., 2009; Mauprivez et al., 2012, Cataldo et al., 2013), the tension parameter of the vocal folds (which is a parameter describing a relation between mass and stiffness of the vocal folds, and which can be found, for instance, in Ishizaka and Flanagan (1972)) 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 (Cataldo et al., 2013). In this paper, we consider a stochastic model for  $k_1$ ,  $k_2$  and  $k_c$  that are modeled by stochastic processes  $K_1(t)$ ,  $K_2(t)$  and  $K_c(t)$ , in order to generate a jitter in the voice production for the phonation.

The corresponding system of nonlinear stochastic differential equations is obtained by substituting stiffnesses  $k_1$ ,  $k_2$  and  $k_c$  in the deterministic model by the stochastic processes  $K_1(t)$ ,  $K_2(t)$  and  $K_c(t)$ , respectively. Consequently,  $u_g, u_1, \dots, u_n, u_R$  and also  $x_1, x_2$  and  $x_c$  become stochastic processes  $U_g, U_1, \dots, U_n, U_R$  and  $X_1$  and  $X_2$  (the stochastic equations are detailed in Appendix B).

- We are interested in constructing the asymptotic stationary solution for the stochastic equations, which corresponds to a "quasi-periodic" steady state solution for the deterministic equation. The stochastic model allows for producing the normal voices and also some pathological voices.

### Appendix B. Stochastic vocal folds dynamic equations

*Stochastic vocal folds dynamic equation.* The vocal folds dynamic equation (defined by Eq. (A.11)) in  $x_1(t)$  and  $x_2(t)$  depending on  $u_g(t)$  becomes a system of nonlinear stochastic differential equations for the stochastic processes  $X_1(t)$  and  $X_2(t)$  coupled with the stochastic process  $U_g(t)$ . The system of stochastic differential equations can be written as

$$[M]\ddot{X}(t) + [C]\dot{X}(t) + [K(t)]X(t) = -S(X(t)) + F(X(t), U_g(t)), \quad (\text{B.1})$$

in which

$$[M] = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \quad [C] = \begin{bmatrix} r_1 & 0 \\ 0 & r_2 \end{bmatrix},$$

$$[K(t)] = \begin{bmatrix} K_1(t) + K_c(t) & -K_c(t) \\ -K_c(t) & K_2(t) + K_c(t) \end{bmatrix}.$$

The vector valued-function  $F(X(t), U_g(t))$  is written as

$$F(X(t), U_g(t)) = (F_1(X_1(t), U_g(t)), F_2(X_2(t), U_g(t))),$$

for which  $F_1(X_1(t), U_g(t))$  and  $F_2(X_2(t), U_g(t))$  are obtained in replacing, in the equations that define  $F_1$  and  $F_2$  introduced in Appendix A, the deterministic quantities  $x_1(t)$ ,  $x_2(t)$ , and  $u_g(t)$  by the stochastic quantities  $X_1(t)$ ,  $X_2(t)$ , and  $U_g(t)$ . It should be noted that the random quantities  $F(X(t), U_g(t))$  and  $S(X(t))$  depend on the collision of the vocal folds.

When there is no collision of the vocal folds, the expression of  $S(X(t))$  is given by

$$S(X(t)) = \begin{bmatrix} K_1(t) \eta_{k_1} X_1(t)^3 \\ K_2(t) \eta_{k_2} X_2(t)^3 \end{bmatrix}.$$

When there is collision of the vocal folds, the expression of  $S(X(t))$  is given by

$$S(X(t)) = \begin{bmatrix} K_1(t) \left( \eta_{k_1} X_1(t)^3 + \frac{h_1}{k_1} \left\{ \left( X_1(t) + \frac{Ag_{01}}{2\ell} \right) + \eta_{\epsilon_1} \left( X_1(t) + \frac{Ag_{01}}{2\ell} \right)^3 \right\} \right) \\ K_2(t) \left( \eta_{k_2} X_2(t)^3 + \frac{h_2}{k_2} \left\{ \left( X_2(t) + \frac{Ag_{02}}{2\ell} \right) + \eta_{\epsilon_2} \left( X_2(t) + \frac{Ag_{02}}{2\ell} \right)^3 \right\} \right) \end{bmatrix}.$$

*Stochastic coupling equation.* The deterministic coupling nonlinear equation (defined by Eq. (A.1) with Eqs. (A.2) to (A.9)) in  $u_g(t)$  and coupled with  $u_1(t)$ , for which the coefficients depend on  $x_1(t)$  and  $x_2(t)$ , become a stochastic coupling equation for the stochastic process  $U_g(t)$  and coupled with the stochastic process  $U_1(t)$ , for which the coefficients depend on stochastic processes  $X_1(t)$  and  $X_2(t)$ , and is written as

$$\begin{aligned} & (R_{k1}(X_1(t)) + R_{k2}(X_2(t))) |U_g(t)| U_g(t) + (R_{v1}(X_1(t)) + R_{v2}(X_2(t))) U_g(t) \\ & + (L_{g1}(X_1(t)) + L_{g2}(X_2(t))) \frac{dU_g}{dt} + L_1 \frac{dU_g}{dt} + R_1 U_g(t) \\ & + \frac{1}{c_1} \int_0^t \{U_g(\tau) - U_1(\tau)\} d\tau - p_s(t) = 0, \quad (\text{B.2}) \end{aligned}$$

where  $R_{v1}(X_1(t))$ ,  $R_{v2}(X_2(t))$ ,  $L_{g1}(X_1(t))$ , and  $L_{g2}(X_2(t))$  are defined by Eqs. (A.2) to (A.8) in which  $x_1(t)$ ,  $x_2(t)$ , and  $u_g(t)$  are replaced by the stochastic quantities  $X_1(t)$ ,  $X_2(t)$ , and  $U_g(t)$ . When the glottis is closed, the *coupling equation* becomes

$$\frac{1}{c_1} \int_0^t \{U_g(\tau) - U_1(\tau)\} d\tau - p_s(t) = 0. \quad (\text{B.3})$$

*Stochastic sound acoustic propagation equation.* The deterministic resistances  $r_1, \dots, r_n$  of the vocal tract (which are independent of  $t$ ) become stochastic processes  $R_1(t), \dots, R_n(t)$  for the stochastic model, and are such that, for all  $t$  and for all  $i = 1, \dots, n$ ,

$$R_i(t) = (S_i/A_i^2) \sqrt{\rho \mu \Omega_1(t)/2} \quad , \quad \Omega_1(t) = (K_1(t)/m)^{1/2}. \quad (\text{B.4})$$

Consequently, the deterministic sound acoustic propagation equation (defined by Eq. (A.10)) in  $u_1(t), \dots, u_n(t), u_R(t)$  becomes a stochastic sound acoustic propagation equation with stochastic coefficients for the stochastic processes

$U_1(t), \dots, U_n(t), U_R(t)$ , which is written as

$$\left\{ \begin{array}{l} (L_1 + L_2) \frac{dU_1(t)}{dt} + (R_1(t) + R_2(t)) U_1(t) + \frac{1}{c_2} \int_0^t \{U_1(\tau) - U_2(\tau)\} d\tau + \\ \quad \frac{1}{c_1} \int_0^t \{U_1(\tau) - U_g(\tau)\} d\tau = 0, \\ \\ (L_i + L_{i+1}) \frac{dU_i(t)}{dt} + (R_i(t) + R_{i+1}(t)) U_i(t) + \\ \quad \frac{1}{c_{i+1}} \int_0^t \{U_i(\tau) - U_{i+1}(\tau)\} d\tau + \\ \quad \frac{1}{c_i} \int_0^t \{U_i(\tau) - U_{i-1}(\tau)\} d\tau = 0 \quad , \quad i = 2, \dots, n-1, \\ \\ (L_n + L_R) \frac{dU_n(t)}{dt} + R_n(t) U_n(t) - L_R \frac{dU_R(t)}{dt} + \\ \quad \frac{1}{c_n} \int_0^t \{U_n(\tau) - U_{n-1}(\tau)\} d\tau = 0, \\ \\ L_R \frac{d(U_R(t) - U_n(t))}{dt} + r_R U_R(t) = 0. \end{array} \right. \quad (\text{B.5})$$

## REFERENCES

- Baken, R., Orlikof, F., 2000. Clinical measurement of speech and voice, second edition. USA, Singular.
- Cataldo, E., Lucero, J., Leta, F., Nicolato, L., 2005. Synthesis of voiced sounds using low-dimensional models of the vocal cords time-varying subglottal pressure. *Mechanics Research Communication* 33(2), 250–260.
- Cataldo, E., Sampaio R., Lucero, J., Soize, C., 2008. Modeling random uncertainties in voice production using a parametric approach. *Mechanics Research Communication* 35(7), 429–490.
- Cataldo, E., Soize, C., Sampaio, R., Desceliers, C., 2009. Probabilistic modeling of a nonlinear dynamical system used for producing voice. *Computational Mechanics* 43(2), 265–275.
- Cataldo, E., Soize, C., Sampaio, R., 2013. Uncertainty quantification of voice signal production mechanical model and experimental updating. *Mechanical Systems and Signal Processing* 40(2), 718–726.
- Deguchi, S., Kawahara, Y., 2011. Simulation of human phonation with vocal nodules. *American Journal of Computational Mathematics*, 1, 189–201.
- Dejonckerea, P. H., Giordano, A., Schoentgen, J., Fraj, S., Bocchid, L., Manfredid, C., 2012. To what degree of voice perturbation are jitter mea-

- surements valid? A novel approach with synthesized vowels and visuo-perceptual pattern recognition. *Biomedical Signal Processing and Control* 7(1), 37–42.
- Erath, B. D., Zañartu, M., Stewart, K. C., Plesniak, M. W., Sommer, D. E., Peterson, S. D., 2013. A review of lumped-element models of voiced speech. *Speech Communications*, 55, 667–690.
- Fant, G., 1960. *The acoustic theory of speech production*. Mouton, The Hague.
- Farrus, M., Hernando, J., 2008. Using jitter and shimmer in speaker verification. *Signal Processing IET* 3(4), 247–257.
- Flanagan, J., Landgraf, L., 1968. Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics AU-16*, 57–64.
- Fraile, R., Godino-Llorente, J. I., Kob, M., 2015. Simulation of tremulous voices using biomechanical model. *Eurasip Journal on Audio, Speech, and Music Processing*, 1, 1–12.
- Henrich, N., d’Alessandro, C., Doval, B., Castellengo, M., 2005. 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.
- Hirano, M., 1981. *Clinical Examination of Voice*. Springer Verlag, Vienna.
- Hirose, H., 1998. *Clinical Aspects of Voice Disorders*. Interuna Publishers, Tokyo.
- 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.
- Krée, P., Soize, C., 1986. *Mathematics of Random Phenomena*. Reidel, Dordrecht.
- Koike, Y., 1973. Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Stud. Phonol.* VII, 17–23.
- Kreiman, J., Gerratt, B. R., 2005. Perception of aperiodicity in pathological voice. *Journal of the Acoustical Society of America* 117(4), 2201–2211.
- Lieberman, P., 1963. Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *Journal of the Acoustical Society of America* 35(3), 344–353.
- Li, L., Saigusa, H., Nakazawa, Y., Nakamura, T., Komachi, T., Yamaguchi, S., Liu A., Sugisaki, Y., Shinya, E., Shen, H., 2010. A pathological study of bamboo nodule of the vocal fold. *Journal of Voice*, 24 (6), 738–741.
- Londono, J., Llorente, J. I. G., Lechien, N. S. , Ruiz, V. O. , Dominguez, G. C., 2010. An improved method for voice pathology detection by means of a HMM-based feature space transformation. *Journal of Pattern Recognition* 43 (9), 3100–3112.
- Lucero, J. C., 2005. Bifurcations and limit cycles in a model for a vocal fold oscillator. *Communication in Mathematical Sciences* 3(4), 517–529.

- Manfredi, C., Giordano, A., Schoentgen, J., Fraj, S., Bocchi, L., Dejonckere, P. H., 2012. Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools. *Biomedical Signal Processing and Control* 7 (4), 409–416.
- Mauprivez, J., Cataldo, E., Sampaio, R., 2012. 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.
- Mendonza, L., Vellasco, M., Cataldo, E., Silva, M. B., Apolinario, A. A., 2014. Classification of vocal aging using parameters extracted from the glottal signal. *Journal of Voice*, 21(2), 157–68.
- Mendoza L., Kohler, M., Vellasco, M., Cataldo, E., 2015. Analysis and Classification of Voice Pathologies using glottal signal parameters. *Journal of Voice*, 1, 1–10.
- Muhammad, G., Melhem, M., 2014. Pathological voice detection and binary classification using MPEG-7 audio features. *Biomedical Signal Processing and Control* 11, 1–9.
- Pinto, N. B., Titze, I. R., 1990. Unification of perturbation measures in speech signals. *The Journal of the Acoustical Society of America*, 87, 1278–1289.
- Praat software website: <http://www.fon.hum.uva.nl/praat/>.
- Schoentgen, J., De Guchteneere, R., 1997. Predictable and random components of jitter. *Speech Communication* 21, 255–272.
- Schoentgen, J., 2001. Stochastic models of Jitter. *Journal of the Acoustical Society of America* 109, 1631–1650.
- Schoentgen, J., Fraj, S., Lucero, J. C., 2015. Testing the reliability of grade, roughness and breathiness scores by means of synthetic speech stimuli. *Logopedics Phoniatrics Vocology* 40(1) 5–13.
- Silva, D. G., Oliveira, L. C., Andrea, M., 2009. 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.
- Soize, C., 1994. *The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions*. World Scientific, Singapore.
- Soize, C., 2005. Random matrix theory for modeling uncertainties in computational mechanics. *Computer methods in Applied Mechanics and Engineering* 194(12-16), 1333–1366.
- Soize, C., 2006. Non-Gaussian positive-definite matrix-valued random fields for elliptic stochastic partial differential operators. *Computer methods in Applied Mechanics and Engineering*, 195(1-3), 26–64.
- Soize, C., 2012. *Stochastic Models of Uncertainties in Computational Mechanics*. American Society of Civil Engineers (ASCE), Reston.

- Story, B. H., Titze, I. R., 1995. Voice simulation with a body-cover model of the vocal folds. *The Journal of the Acoustical Society of America*, 97(2), 1249–1260.
- Titze, I. R., 1973. The human vocal cords: a mathematical model. Part I. *Phonetica*, 28, 129–170.
- Titze, I., 1994. *Principles of voice production*. Prentice Hall, Englewood Cliffs, NJ.
- Vieira, M. N., McInnes, F. R., Jack M. A., 2001. On the influence of laryngeal pathologies on acoustic and electroglottographic jitter measures. *The Journal of the Acoustical Society of America*, 111, 1045–1055.
- Wilcox, K. A., Horii, Y., 2015. Age and Changes in vocal jitter. *Journal of Gerontology*, 35(2),194–198.
- Zhang Y., Jiang, J. J., 2008. Acoustic analyses of sustained and running voices from patients with laryngeal pathologies. *Journal of Voice*, 22(1), 1–9.