

A STOCHASTIC MECHANICAL MODEL TO GENERATE JITTER IN THE PRODUCTION OF VOICED SOUNDS

E. Cataldo¹ and C. Soize²

¹Universidade Federal Fluminense
Applied Mathematics Department – Graduate Program in Telecommunications Engineering
Rua Mário Santos Braga, S/N – Centro – Niterói – RJ – 24020-140 – Brazil
e-mail: ecataldo@im.uff.br

² Université Paris-Est
Laboratoire Modélisation et Simulation Multi Echelle – MSME FRE 3160 CNRS – 5, Bd Descartes,
77454 – Marne-la-Vallée – France
e-mail: christian.soize@univ-paris-est.fr

Keywords: Stochastic modeling, voice production, mechanical models, uncertainties modelling.

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 jitter has been the subject for researchers due to its important applications such as identification of pathological voices (nodulus in the vocal folds, paralysis of the vocal folds, or even, the vocal aging, among others). Large values for jitter variation can indicate a pathological characteristic of the voice. The objective of this paper is to construct a stochastic model of jitter using a mechanical model of the vocal folds. This model assumes complete right-left symmetry of the vocal folds and considers motion of the vocal folds only in the horizontal direction. The corresponding stiffness of each vocal fold is considered as a stochastic process and its modelling is proposed. The probability density functions of the fundamental frequency related to the voices produced are constructed and compared for different levels of jitter. Some samples of synthesized voices in these cases are obtained.*

1 INTRODUCTION

The systems of voice production are important sensorial structures which permit human beings communicate, share information, exchange ideas, feelings, emotions, intentions, etc. The inefficiency of these structures or its absent 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 in evaluating the vocal structures remains in their importance for the human voice production.

One can say, in a figurative sense, that speech does not start in the lungs, but in the brain. After the message is created in the mind, it is necessary to make a plan to physically produce the corresponding sound to transmit the information. 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, before continuous, 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. It can be used to measure voice quality, to indicate the presence of pathologies related to the voice, and even to recognize speakers [1, 2, 12].

The objective of this paper is to construct a stochastic model of jitter based on the deterministic model created by Flanagan and Landgraf [4] to produce voice, with some modifications brought from the Ishizaka and Flanagan [5] model and other introduced by the authors. The idea is to consider the stiffness of the one-mass model as a stochastic process, to model it and to solve the corresponding stochastic differential equations. The synthesis of voice signals will be obtained considering different levels of jitter.

2 Deterministic model used

The deterministic model used as start is the 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 constitute a symmetric system composed by two vocal folds. The vocal tract is represented by a standard configuration of concatenated tubes [3, 11].

Figure 1 illustrates a sketch of the model.

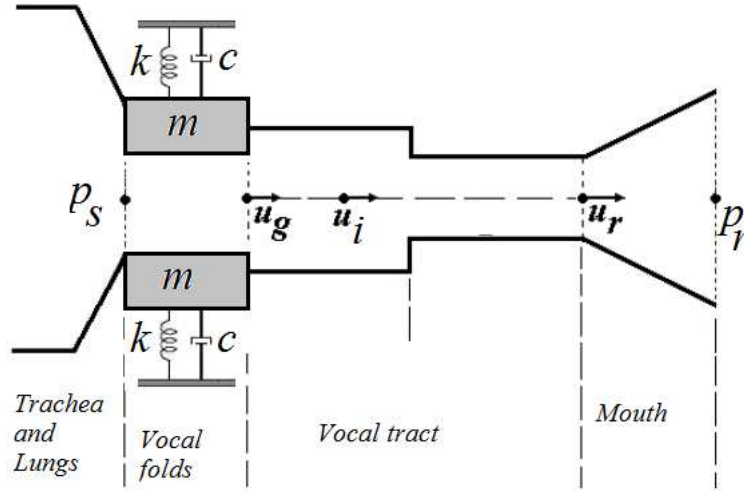


Figure 1: Sketch of the Flanagan and Landgraf model (1968).

The system of differential equations to be solved can be divided in three parts: (i) A coupling equation between the glottal flow and the vocal tract (Eq. (1)); (ii) A system of differential equations related to the sound acoustic propagation through the vocal tract vocal (Eq. (2)); (iii) A equation related to the dynamics of the vocal folds (Eq. (3)).

The complete model considered here has some modifications in relation to the original Flanagan and Landgraf model. Some of them are in the Ishizaka and Flanagan model, and other according to Cataldo et al. [6, 8].

The coupling equation is given by Eq. (1):

$$(R_v + R_k)U_g + L_g \frac{dU_g}{dt} + L_1 \frac{dU_g}{dt} + \frac{1}{C_1} \int_0^t (U_g - U_1) d\hat{t} - P_s(t) = 0, \quad (1)$$

where $R_v = 12\mu d\ell^2 Ag^{-3}$, $R_k = 0,44\rho |U_g| Ag^{-2}$ and $L_g = \rho d Ag^{-1}$, μ is the air cinematics viscosity, d is the vocal fold thickness, ℓ is the length of each vocal fold, A_g is the glottal area, ρ is the air density and U_g is the acoustic volume velocity through the glottis.

The equations related to the wave acoustic propagation through the vocal tract are given by Eq. (2):

$$\left\{ \begin{array}{l} (L_1 + L_2) \frac{dU_1}{dt} + (R_1 + R_2)U_1 + \frac{1}{C_2} \int_0^t (U_1 - U_2)dt' + \frac{1}{C_1} \int_0^t (U_1 - U_g)dt' = 0 \\ (L_2 + L_3) \frac{dU_2}{dt} + (R_2 + R_3)U_2 + \frac{1}{C_3} \int_0^t (U_2 - U_3)dt' + \frac{1}{C_2} \int_0^t (U_2 - U_1)dt' = 0 \\ \vdots \\ (L_{n-1} + L_n) \frac{dU_{n-1}}{dt} + (R_{n-1} + R_n)U_{n-1} + \frac{1}{C_n} \int_0^t (U_{n-1} - U_n)dt' \\ \quad + \frac{1}{C_{n-1}} \int_0^t (U_{n-1} - U_{n-2})dt' = 0 \\ (L_n + L_R) \frac{dU_n}{dt} + R_n U_n - L_R \frac{dU_R}{dt} + \frac{1}{C_n} \int_0^t (U_n - U_{n-1})dt' = 0 \\ L_R \frac{d(U_R - U_n)}{dt} + R_R U_R = 0 \end{array} \right. \quad (2)$$

where n is the number of tubes considered and the configuration of the vocal tract is the same used in [11].

These equations consider the vocal tract representation as a transmission line of n cylindrical tubes, with area sections given by A_1, A_2, \dots, A_n , and tube lengths given by $\ell_1, \ell_2, \dots, \ell_n$. The corresponding *inductances* are given by $L_n = \frac{\rho \ell_n}{2A_n}$ and the *capacitances* given by $C_n = (\ell_n A_n / \rho c^2)$, with c as the sound velocity. To take into account the lost of the vocal tract, *resistances* in series are used and calculated by $R_n = (S_n / A_n^2) \sqrt{\rho \mu \omega / 2}$, where S_n is the length of the n -th circumference and ω is the oscillation frequency of the vocal folds. The line transmission ends with a radiation load such that the inductance is given by $L_R = (8\rho / 3\pi) \sqrt{\pi A_n}$ and the resistance given $R_R = (128\rho c / 9\pi^2 A_n)$, where A_n is the last section cylindrical area (corresponding to the mouth).

The vocal folds dynamics is given by Eq. (3):

$$M \frac{dx^2}{dt^2} + C \frac{dx}{dt} + Kx = F(t), \quad (3)$$

where $x(t)$ is the displacement of the mass, $F(t) = \frac{1}{2}(P_1(t) + P_2(t))\ell d$ is the force upon the mass, with $P_1(t) = P_s(t) - 1, 37P_B(t)$, P_s is the subglottal pressure, $P_2 = -0.50P_B(t)$ and $P_B(t) = \frac{1}{2}\rho |U_g(t)|^2 A_g(t)^{-2}$, where U_g is the glottal flow velocity and A_g is the glottal area given by $A_g(t) = A_{g0} + \ell x(t)$, with ℓ the length of each vocal fold.

At the moment of the collision of the vocal folds, when the displacement of the mass reaches the critic value $x_0 = -\frac{A_{g0}}{\ell}$, the glottis closes and $A_g(t)$ and $U_g(t)$ remains null until $x(t) > x_0$.

At this moment, the damping is given by $C + C^*$ and the force is given by $F(t) = \frac{1}{2}P_s(t)\ell d$, with $C^* = 2\alpha\sqrt{MK}$. The values of the parameters used can be found in [7, 8, 10].

In previous works [8, 9, 10], the tension parameter of the vocal folds was considered as a random variable and its probability density function was constructed, identified by solving an

inverse stochastic problem and updated using Bayesian statistics. Here, K is considered as a stochastic process, named $K(t)$. The corresponding differential stochastic equation is obtained by substituting K in Eq. (3) by $K(t)$. The modelling used for $K(t)$ will be explained in details in the following of this paper.

3 Jitter modelling

Jitter has called the attention of researchers who believe it can be measured using the voice signal and can be used for discriminating healthy speakers from those with voice pathological characteristics, and other characteristics related, as aging voice, for example. From the second middle of eighties years, the research has become less clinical and more methodological. The methods used to measure jitter, and other characteristics of the voice signal, changed considerably.

One of the first works to quantify jitter was proposed by Lieberman [3] and it was characterized by a factor which considered all perturbations greater than 0.5 ms. Other preliminary studies were based on 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 studies agree with the fact that typical values of jitter are between 0.1% and 1% of the fundamental period, for the so-called normal voices; that is, without characteristics of pathologies. It is important to say that, in general, jitter decreases as the fundamental frequency increases. The majority of the authors conclude that it is possible to discriminate healthy voices from pathological voices using jitter characteristics and even to recognize speakers [13, 19, 20, 21, 22].

In general, the authors who work with models of jitter do not consider mathematical models for the voice production and just a few consider stochastic models [14, 15, 16].

3.1 $K(t)$ modelling

Let $K(t)$ be a stochastic process corresponding to the stiffness in the model used here. So, the Eq. (3) can be rewritten as Eq. (4):

$$M \frac{dx^2}{dt^2} + C \frac{dx}{dt} + K(t)x = F(t). \quad (4)$$

The following properties of the stochastic process $\{K(t), t \in \mathbb{R}\}$, indexed by \mathbb{R} , with values in \mathbb{R}^+ , are introduced in order to obtain a solution for Eq. (3):

(i) For all t , $0 < K_0 \leq K(t)$, where K_0 is a positive constant.

(ii) $\{K(t), t \in \mathbb{R}\}$ is a stationary stochastic process.

(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 $K_c(t)$ is such that $K(t) = K_c(t) + \underline{K}$. The autocorrelation function of $K_c(t)$ is written as $R_{K_c}(\tau) = \int_{-\infty}^{+\infty} e^{i\omega\tau} S_{K_c}(\omega) d\omega$ in which $S_{K_c}(\omega)$ is the power spectral density function.

(iv) The variance σ_K^2 of the random variable $K(t)$, for all fixed t , is such that $\sigma_K^2 = R_{K_c}(0)$.

For all t fixed, $K(t)$ is modeled as follows:

$$K(t) = K_0 + (\underline{K} - K_0)f(\underline{Z} + Z(t)), \quad (5)$$

with $0 < K_0 < \underline{K}$ and the function $f : \mathbb{R} \rightarrow \mathbb{R}^+$ is a continuous function and should be chosen such that

$$E\{f(\underline{Z} + Z(t))\} = 1 \text{ and } E\{f(\underline{Z} + Z(t))^2\} < +\infty.$$

Choosing $f(\tilde{z}) = \tilde{z}^2$, $K(t)$ is rewritten as $K(t) = K_0 + (\underline{K} - K_0)(\underline{Z} + Z(t))^2$, where $\{Z(t), t \in \mathbb{R}\}$ is a second-order stochastic process, indexed by \mathbb{R} , with real values, centered, mean square continuous, spectral density $S_Z(\omega)$ and physically realizable; that is, $S_z(\omega)$ satisfies the following condition:

$$\int_{-\infty}^{+\infty} \frac{\ell n S_z(\omega)}{1 + \omega^2} d\omega > -\infty. \quad (6)$$

Stochastic process Z is constructed according to a finite dimension Markov realization [18] that is written as:

$$\begin{cases} Z(t) = UY(t) \\ dY(t) = -BY(t)dt + AdW(t), \end{cases} \quad (7)$$

in which W is the \mathbb{R} -valued normalized Wiener process indexed by $[0, +\infty[$, where A is a real constant, B a positive real constant and U a real constant. Taking the particular case $U = 1$, the previous Itô stochastic differential equation can be written as an equality of generalized stochastic processes [17, 18]:

$$\frac{dZ(t)}{dt} = -BZ(t) + A \frac{dW(t)}{dt}, \quad (8)$$

with $Z(0) = Z_0$ a \mathbb{R} -valued second-order random variable. Note that W can defined a generalized stochastic process whose generalized derivative $\frac{dW}{dt}$ is a \mathbb{R} -valued generalized stochastic process indexed by \mathbb{R}^+ , whose extension on \mathbb{R} is stochastically equivalent in the wide sense to the normalized Gaussian white noise $N = \frac{dW(t)}{dt}$, which is not a second-order stochastic process, but for which the power spectral density function is written, for all real ω , $S_N(\omega) = \frac{1}{2\pi}$. The frequency response function H of the linear filter $h*$ associated with Eq. (8) is written as

$$H(\omega) = \frac{A}{i\omega + B}.$$

The power spectral density function of the Gaussian stationary stochastic second-order stochastic process $Z = h * N$ is such that

$$S_Z(\omega) = |H(\omega)|^2 S_N(\omega),$$

which is thus written as

$$S_Z(\omega) = \frac{1}{2\pi} \frac{A^2}{\omega^2 + B^2}.$$

Therefore,

$$\begin{aligned} E\{(\underline{Z} + Z(t))^2\} = 1 &\Leftrightarrow \underline{Z}^2 + E\{Z(t)\}^2 = 1 \Leftrightarrow \underline{Z}^2 + \int_{-\infty}^{+\infty} S_Z(\omega) d\omega = 1 \\ &\Leftrightarrow \underline{Z}^2 + \int_{-\infty}^{+\infty} S_Z(\omega) d\omega = 1 \Leftrightarrow \underline{Z}^2 + \int_{-\infty}^{+\infty} \frac{A^2}{2\pi(\omega^2 + B^2)} d\omega = 1 \\ &\Leftrightarrow \underline{Z}^2 + \frac{A^2}{2B} = 1 \Leftrightarrow \underline{Z}^2 = 1 - \frac{A^2}{2B} \geq 0. \end{aligned}$$

So, $1 - \frac{A^2}{2B} > 0 \Leftrightarrow A < \sqrt{2B}$.

4 Simulations

The objective of this section is to simulate voice signals considering the stochastic model proposed and, consequently, with jitter.

The subglottal pressure is not considered constant during all the time of simulation, but a function p_s , according to the results obtained in [6], given by:

$$p_s(t) = \begin{cases} 8000 \sin(5\pi t) & , 0 \leq t < 0.1 \\ 8000 & , 0.1 \leq t \leq 1.9 \\ 8000 \sin\left(\frac{5\pi t}{9}\right) & , 1.9 \leq t < 2. \end{cases}$$

and the corresponding plot is given in Fig. 2. The parameters used corresponding to the

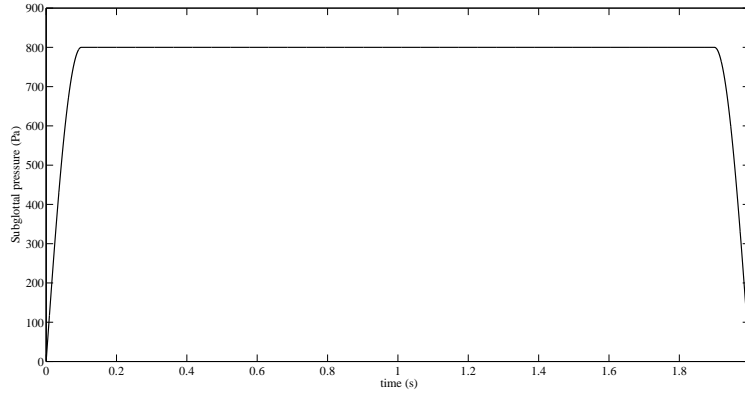


Figure 2: Subglottal pressure.

deterministic model have the following values: $A_{g0} = 0.05 \text{ cm}^2$, $\rho = 1.2 \times 10^{-3} \text{ g/cm}^3$, $c = 346.3 \times 10^2 \text{ cm/s}$, $\mu = 1.86 \times 10^{-5} \text{ g/(cm}^2\text{s)}$, $M = 0.24 \text{ g}$, $L = 1.4 \text{ cm}$, $d = 0.3 \text{ cm}$, $K_0 = 40.000 \text{ g/s}^2$. For the damping coefficient, it was considered $C = 0$ and $\alpha = 1$, i.e, only during the collision the damping was considered, as in [4]. An analysis has been performed in order to determine the number of necessary realizations. The constants A and B in Eq. 8 are taken as $A = 40$ and $B = 1,000,000$. The number of realizations is $N = 2 \times 88,200 = 176,400$, with a time step $\Delta t = 1/f_s = 1/88200 \text{ s}$. The time of simulation for each realization is then given by $N \times \Delta t = 2 \text{ s}$. The expected values $E\{K(t)\}$ and $E\{K(t)^2\}$ are independent of t , and are estimated by

$$E\{K(t)\} = \lim_{t \rightarrow +\infty} \overline{K}(t) \quad , \quad \overline{K}(t) = \frac{1}{t} \int_0^t K(t') dt' , \quad (9)$$

$$E\{K(t)^2\} = \lim_{t \rightarrow +\infty} \overline{K^2}(t) \quad , \quad \overline{K^2}(t) = \frac{1}{t} \int_0^t K(t')^2 dt' , \quad (10)$$

which allows for verifying when the ergodicity property is reached.

Only the first 1,000 points are considered for plotting the graphs shown in Fig. 3.

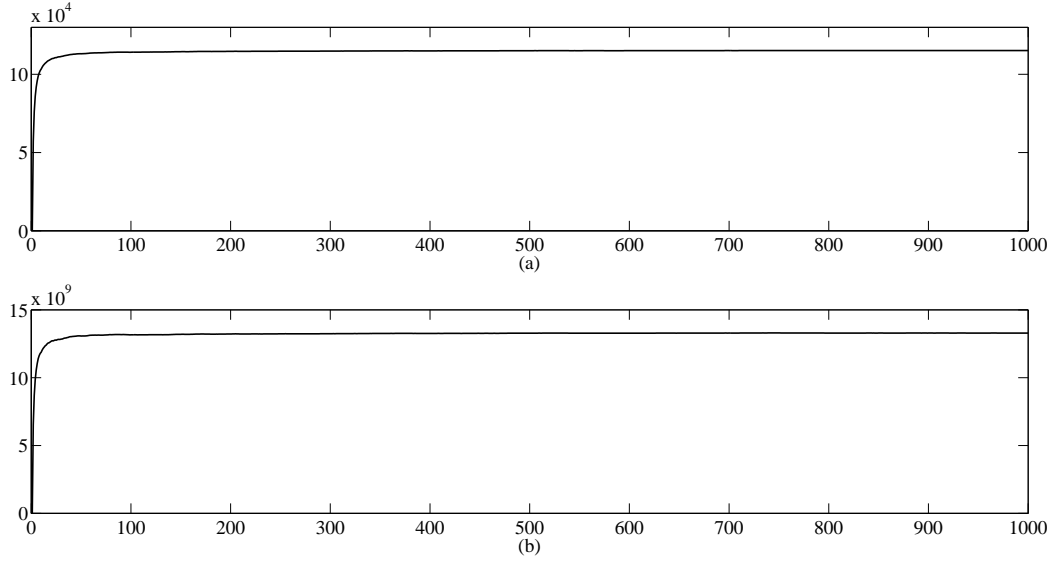


Figure 3: (a) Graph $t \mapsto \overline{K}(t)$. (b) Graph $t \mapsto \overline{K^2}(t)$.

Since stochastic process is stationary, the Cumulative Distribution Function (CDF) of the random variable $K(t)$ or any fixed t , is independent of t is denoted by $F_K(k)$, and such that

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

$$1_{[-\infty, k]}(k') = \begin{cases} 1, & \text{if } k' \leq k \\ 0, & \text{if } k' > k. \end{cases}$$

Due to the ergodic property of stochastic process K ,

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

Figure 4 shows the graphs of $t \mapsto \overline{F_K}(k; t)$ for different values of k , considering only the first 1000 time steps of the time simulation.

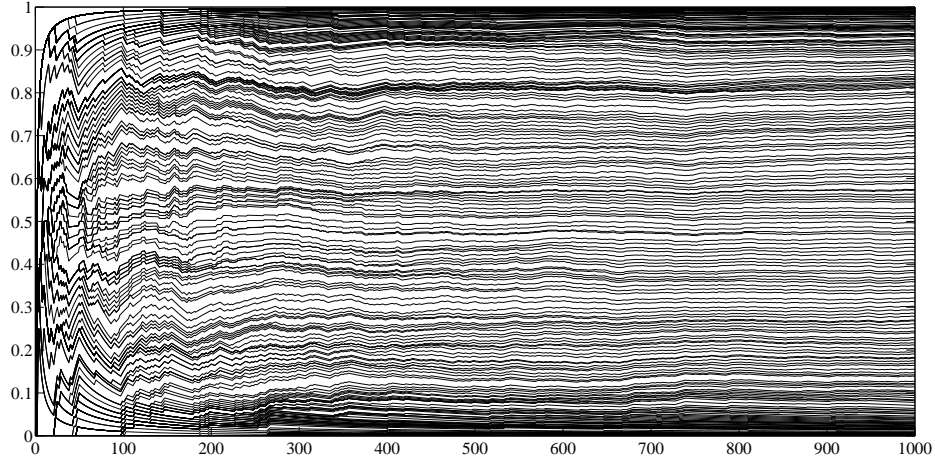


Figure 4: For several values of k , graph $t \mapsto \overline{F_K}(k; t)$. The lower graph corresponds to the minimum value of k , and the upper graph to its maximum value.

Two samples of voice signals are simulated, one considering only the deterministic model, without jitter ($A = 0$) and the others with different levels of jitter ($A = 40$ and $A = 160$). Samples of the glottal signals (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.

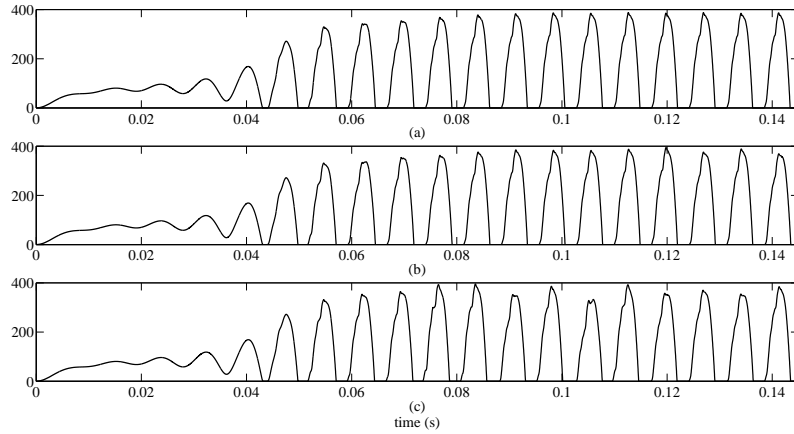


Figure 5: Glottal signal without jitter(a) and with jitter (b) $A = 40$ and (c) $A = 160$, synthesized by the described model.

The 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.

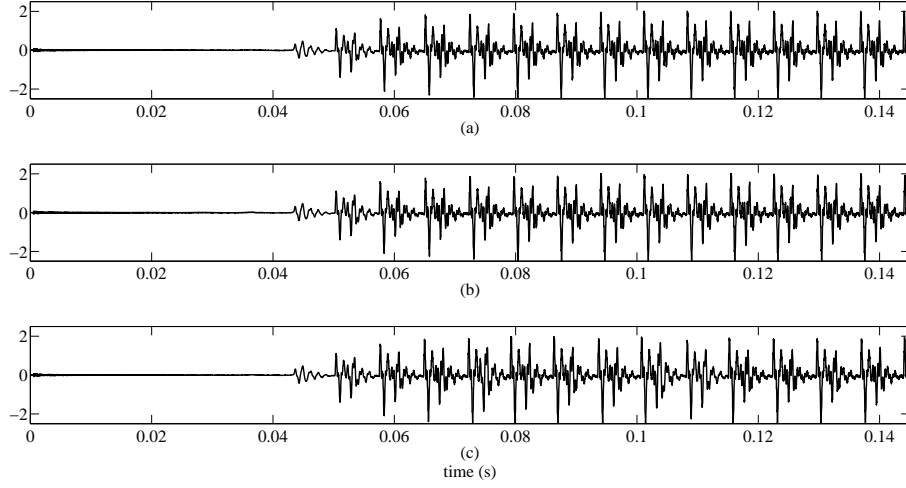


Figure 6: Output pressure signal without jitter(a) and with jitter (b) $A = 40$ and (c) $A = 160$, synthesized by the described model.

It is difficult to observe the variation of the fundamental frequency in these plots. A good way to observe the 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 , or two different levels of jitter, and the pdf's are constructed (Fig. 7).

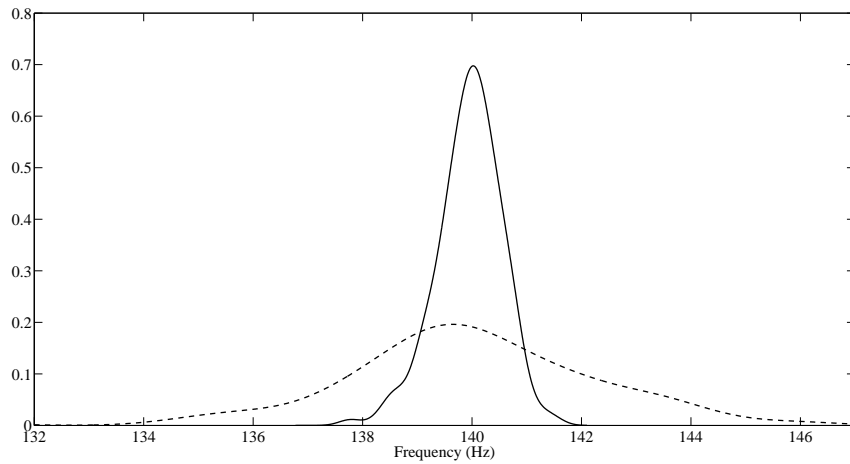


Figure 7: Probability density functions of the fundamental frequency considering two different levels of Jitter: $A = 40$ (continuous line) and $A = 160$ (dashed line).

Some 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>, where *detn* corresponds to the case without jitter, N_1 corresponds to $A = 40$ and N_2 corresponds to $A = 160$.

4.1 Jitter measurements

The jitter features are also calculated using the Praat voice analysis software [23]. Praat reports different types of measures for jitter, which are listed below.

4.1.1 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 Eq. (11):

$$\text{Jit}_{\text{abs}} = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}| \quad (11)$$

where T_i are the lengths of each glottal cycle and N the number of periods considered.

4.1.2 Local

It is the average absolute difference between consecutive periods, divided by the average period. It is given by Eq. (12):

$$\text{Jit}_{\text{loc}} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (12)$$

In general, 1.040 is considered as a threshold for pathology.

4.1.3 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 pathology.

4.1.4 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.

4.1.5 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.

4.1.6 Summary of the jitter measurements

Using the signals synthesized, with two 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. 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

A methodology to construct a stochastic model for creating jitter in a mechanical model for producing voice is proposed. Such a model considers the stiffness related to the vocal folds as a stochastic process and the corresponding voice signals are simulated and the probability density function of the fundamental frequency constructed for different values of the parameters associated to the stochastic model. The comparison between the probability density functions shows that the fundamental frequency has variations in relation to a mean value, showing that jitter was generated. The voice signals were also synthesized and it can be perceived the different sounds related to a normal voice, without jitter, and to two different levels of jitter, one of them very similar to a normal voice, with a low percentage of variation of the fundamental frequency and other with a much greater variation, characterizing a hoarse voice, which can indicate a characteristic of pathology.

6 Acknowledgments

This work was supported by CNPq, CAPES and FAPERJ.

The author Edson Cataldo received financial support from CAPES to attend the congress.

REFERENCES

- [1] I. Titze, *Principles of voice production*. Prentice Hall, Englewood Cliffs, NJ, 1994.
- [2] M. Farrus, J. Hernando, Using jitter and shimmer in speaker verification. *Signal Processing, IET*, **3(4)**, 247–257, 2008.
- [3] P. Lieberman, Some acoustic measures of the fundamental periodicity of normal and pathologic larynges. *J. Acoust. Soc. Am.*, **35**, 344–353, 1963.

- [4] J. Flanagan, L. Landgraf, Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics*, **AU-16** (1), 1968.
- [5] K. Ishizaka, J. Flanagan, Synthesis of voiced sounds from a two-mass model of the vocal folds. *Bell Syst. Tech. J.*, **51**, 1233–1268.
- [6] E. Cataldo, J. Lucero, F. Leta, L. Nicolato, Synthesis of voiced sounds using low-dimensional models of the vocal cords time-varying subglottal pressure. *Mechanics Research Communications*, **33**(2), 250–260, 2005.
- [7] E. Cataldo, R. Sampaio, J. Lucero, C. Soize, Modeling random uncertainties in voice production using a parametric approach. *Mechanics Research Communication*, **35**, 429–490, 2008.
- [8] E. Cataldo, C. Soize, R. Sampaio, C. Descelier, Probabilistic modeling of a nonlinear dynamical system used for producing voice. *Computational Mechanics*, **43**, 265–275, 2009.
- [9] E. Cataldo, C. Soize, R. Sampaio, Uncertainty quantification of voice signal production mechanical model and experimental updating. *Mechanical Systems and Signal Processing*, **40**, 718–726, 2013.
- [10] J. Mauprivez, E. Cataldo, R. Sampaio, 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, 2009.
- [11] G. Fant, *The acoustic theory of speech production*. Mouton, The Hague, 1960.
- [12] R. Baken, F. Orlikof, *Clinical measurement of speech and voice*, second edition. USA, singular, 2000.
- [13] L. Mendonza, M. Vellasco, E. Cataldo, M. B. Silva, A. A. Apolinario, Classification of Vocal Aging Using Parameters Extracted From the Glottal Signal. *J Voice*, **21**(2), 157–68, 2014.
- [14] J. Schoengten and R. De Guchteneere. Predictable and random components of jitter. *Speech Communication*, **21**, 255–272, 1997.
- [15] J. Schoengten. Stochastic models of Jitter. *J. Acoust. Soc. of Am.*, **109**, 1631–1650, 2001.
- [16] J. Schoengten, S. Fraj, J. C. Lucero. Testing the reliability of Grade, Roughness and Breathiness scores by means of synthetic speech stimuli. *Logopedics Phoniatrics Vocology*, **1**, 1–9, 2013.
- [17] P. Krée and C. Soize, *Mathematics of Random Phenomena*. Reidel, Dordrecht, 1986.
- [18] C. Soize, *The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions*. World Scientific, Singapore, 1994.
- [19] N. Henrich, C. d’Alessandro, B. Doval, M. Castellengo, 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.

- [20] J. Londono, J. I. G. Llorente, N. S. Lechien, V. O. Ruiz, G. C. Dominguez, 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.
- [21] J. Kreiman and B. R. Gerrat. Perception of aperiodicity in pathological voice. *Acoustical Society of America*, **117**, 2201–2211, 2005.
- [22] P. H. Dejonckerea, A. Giordano, J. Schoentgen, S. Fraj., L. Bocchid and C. Manfredid. 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.
- [23] Praat software website: <http://www.fon.hum.uva.nl/praat/>