

Identifying a stochastic process related to the stiffness in a voice production mechanical model

Edson Cataldo, Christian Soize

► **To cite this version:**

Edson Cataldo, Christian Soize. Identifying a stochastic process related to the stiffness in a voice production mechanical model. 10th International Conference on Structural Dynamics, EURODDYN 2017, Sep 2017, Rome, Italy. Elsevier, 199, pp.912-917, 2017, Procedia Engineering. <hal-01585707>

HAL Id: hal-01585707

<https://hal-upec-upem.archives-ouvertes.fr/hal-01585707>

Submitted on 11 Sep 2017

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.

Identifying a stochastic process related to the stiffness in a voice production mechanical model

E. Cataldo^a, C. Soize^b

^a*Universidade Federal Fluminense, Applied Mathematics Department and Graduate program in Telecommunications Engineering, Rua Mrio 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 that 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 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. In this paper the construction a stochastic model for jitter using an one-mass mechanical model of the vocal folds is proposed. The corresponding stiffness of each vocal fold is considered as a stochastic process and its modeling is propose d. The probability density function of the fundamental frequency related to the voice signals produced are constructed and compared for different levels of jitter. Some samples of synthesized voices in these cases are obtained. Then, using experimental voice signals, the stochastic process modelled will be identified. The idea is to generate voice signals which the level of jitter which can be associated to the experimental voice used.

Keywords: Stochastic modeling, Identification of stochastic processes, voice production, mechanical models, jitter.

1. Introduction

In voiced speech production, the production of the voice signal is due to the oscillation of the vocal folds, which modifies the airflow coming from the lungs 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. The oscillations of the vocal folds are not exactly periodic and, consequently, the pulses of air which compose the glottal signal have not the same time duration. The small random fluctuation in each glottal cycle length is called jitter and it is a way to characterize voice signals even those with pathological characteristics. Typical values of 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

*Edson Cataldo. Tel.: +5521996318344.

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

E-mail address: ecataldo@im.uff.br

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 [1, 2]. 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 [7–9, 11, 12]. 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, which can be modeled as a random phenomenon. The objective of this paper is to use the stochastic model of jitter proposed by Cataldo and Soize [9], which is based on the use of the voice production deterministic model introduced by Flanagan and Landgraf [10], including the modifications carried over from the Ishizaka and Flanagan model [5] and those introduced by the authors. This model proposes a stochastic model for the stiffness that is a stochastic process. In this paper, the objective is to identify some parameters of this stochastic model solving a statistical inverse problem.

2. Deterministic model used

The underlying deterministic model used 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 is composed of two vocal folds. The vocal tract is represented by a standard configuration of concatenated tubes. 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. [7, 13]. The system of differential equations to be solved can be divided in three parts (see Fig. 1 that illustrates a sketch of the model)

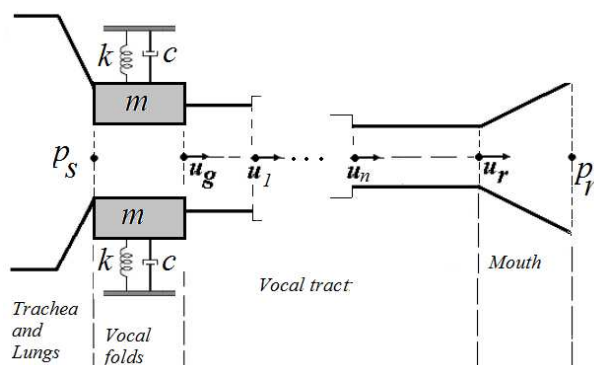


Figure 1: Sketch of the Flanagan and Landgraf model (1968).

All the equations related to this model can be seen in details in [9].

3. Stochastic modeling of jitter

All the procedure to be followed to generate jitter is the same that the one used in [9]. Summarizing, $\{K(t), t \in \mathbb{R}\}$ is a stochastic process indexed by the real line \mathbb{R} , with values in \mathbb{R}^+ , which models stiffness k . The vocal folds dynamic equation in $x(t)$, the position of the vocal focal, depending on $u_g(t)$, the glottal flow, becomes a nonlinear stochastic differential equation for the stochastic process $X(t)$ coupled with the stochastic process U_g , such that

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

where m is the corresponding mass of the vocal folds, c and c^* are values related to the damping of the vocal folds, a_1 and a_2 are constants, p_B is a function related to the air pressure and $p_s(t)$ is the subglottal pressure. The stochastic

process $K(t)$ is written as

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

where k_0 is a constant. 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

$$S_Z(\omega) = \frac{1}{2\pi} \frac{a^2}{\omega^2 + b^2}, \quad a > 0, \quad b > 0, \quad (3)$$

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

$$\underline{z}^2 + \int_{-\infty}^{+\infty} \frac{a^2}{2\pi(\omega^2 + b^2)} d\omega = 1 \quad \implies \quad \underline{z}^2 = 1 - \frac{a^2}{2b}, \quad (4)$$

which yields the following constraint inequality for a and b ,

$$b > 0, \quad 0 < a < \sqrt{2b}. \quad (5)$$

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 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,

$$dY(t) = -b Y(t) dt + a dW(t) \quad t > 0, \quad (6)$$

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 [17] that Eq. (6) has a unique solution $\{Y(t), t \geq 0\}$ such that, for $t_0 \rightarrow +\infty$, 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. (3). Consequently, Eq. (6) can be used for generating trajectories of stochastic process Z . Some results obtained with the vowels synthesis, in the deterministic case, and with two different levels of jitter ($a = 40$ and $a = 160$) can be found and heard in <https://www.dropbox.com/s/mwq3u6ad96po7x/male140Hz.zip?dl=0>, for which *detm* corresponds to the case without jitter, N_1 corresponds to $a = 40$ and N_2 corresponds to $a = 160$. The idea is then to solve the corresponding stochastic inverse problem to identify the parameters a and b associated to real voices obtained experimentally.

4. The corresponding inverse problem

Let us consider the duration between two 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_{fund} , is a random variable, and its inverse is defined as the fundamental frequency that is the random variable $F_{fund} = 1/T_{fund}$. For each value of (a, b) , the probability density function of F_{fund} is estimated using the realization of the stochastic process corresponding to the output pressure. The objective is then to identify parameters a and b such that the probability density functions of the random variable F_{fund} associated to a simulated voice, f_S , is near to the probability density function associated to the real voice, f_R . The distance between the two probability density functions is calculated by

$$d(f_S, f_R) = \int_I |f_S(t) - f_R(t)| dt. \quad (7)$$

Details about simulation of the stochastic model can be seen in [9]. The values of the parameters for the deterministic model are the following:

$A_{g0} = 0.05 \times 10^{-2} m^2$, $\rho = 0.12 kg/m^3$, $c_a = 346.3 m/s$, $\mu = 1.86 \times 10^{-4} kg/(m^2 s)$, $m = 0.24 \times 10^{-2} kg$, $\ell = 1.4 \times 10^{-2} m$, $d = 0.3 \times 10^{-2} m$, $k_0 = 40 N/m$. For the damping coefficient, it was considered $c = 0$ and $\alpha = 1$, i.e, only during the collision the damping was considered. The stochastic solver is the Monte Carlo method.

Two experimental signals were taken into account. One of them is called $exper_1$ and the other one is $exper_2$. One is a voice signal from a man and the another from a woman. At first, the parameter corresponding to the mean value of K was considered in a way that the mean of the random variable associated with the fundamental frequency simulated was very near to the mean value associated to the real voice. The parameters a and b were used to fit the probability density function.

4.1. Identifying parameters for $exper_1$

The first voice signal to be considered is the production of an vowel /a/ and the output pressure is shown in Fig. 2. After solving the inverse stochastic problem, the values obtained were $a = 3.2834$ and $b = -1000$. The

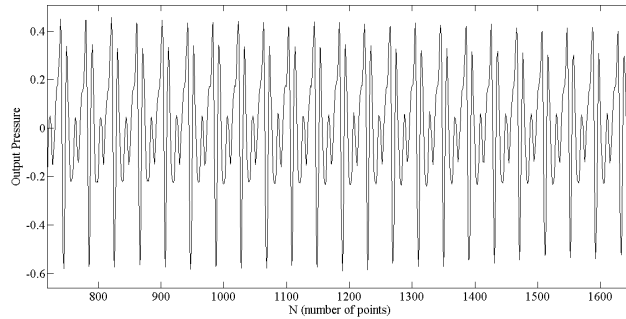


Figure 2: A window of the output pressure of the voice signal produced by the experimental voice.

mean value of the fundamental frequency of the real voice was $272.05 Hz$ and the standard deviation was $1.52 Hz$. For the simulated voice, the mean value of the fundamental frequency was $272.17 Hz$ and the standard deviation was $1.51 Hz$. Figure 3 shows the two pdfs (simulated and real). The value for $d(f_S, f_R)$ is 0.098. The parameters

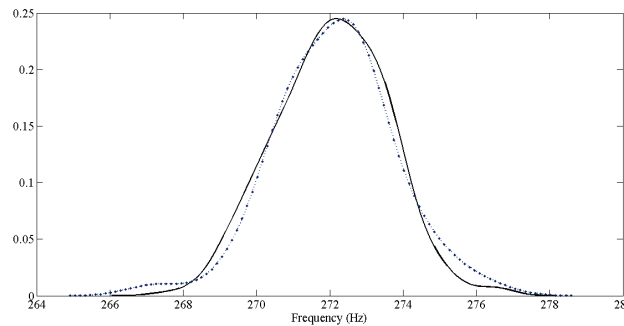


Figure 3: pdf of the random fundamental frequency corresponding to a real voice (solid line) and the corresponding simulated (dotted line).

a and b have been identified taking into account the two probability density functions related to the fundamental frequency. As a way to verify what happens when a sound is synthesized considering these parameters, a voice signal was simulated. The original signal and the corresponding simulated one can be heard following the link <https://www.dropbox.com/sh/3kuhooxuypgtmuh/AAAWiYNkdIIN3xUxajaV6HQPpa?dl=0>. It is interesting to say that only the similarity between fundamental frequencies was taking into account. Maybe other quantities should

be taken into account to become the sounds more similar. IT can be said that this signal is a good signal. It can also be observed that the amplitude of this signal does not suffer high variations. The next signal to be considered is not as good as this one.

4.2. Identifying parameters for *exper₂*

The second voice signal to be considered is the production of an vowel /a/ and the output pressure is shown in Fig. 4. It can be noted that the variation of maximum amplitudes is larger than the signal *exper1*. After solving

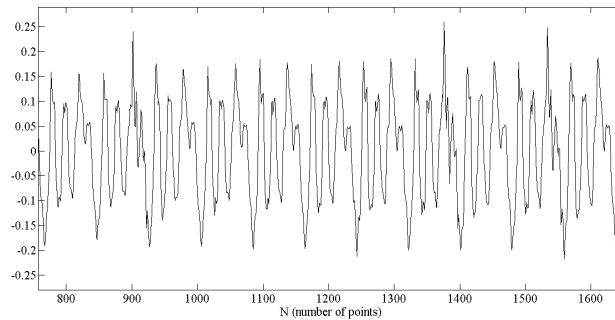


Figure 4: A window of the output pressure of the voice signal produced by the experimental voice.

the inverse stochastic problem, the values obtained are $a = 3.2834$ and $b = -1000$. Figure 5 shows the two pdfs (simulated and real). It can also be observed that the pdf constructed from the experimental signal is not so symmetric

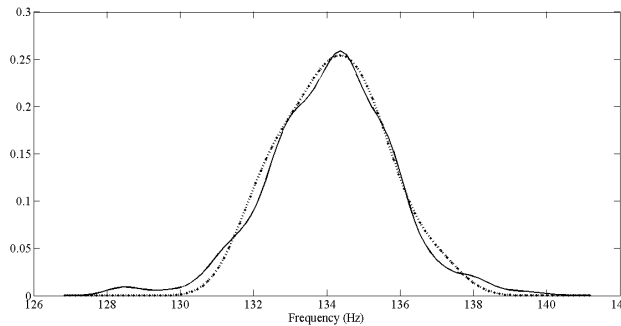


Figure 5: pdf of the random fundamental frequency corresponding to a real voice (solid line) and the corresponding simulated (dotted line).

as the one constructed with the signal *exper1*. The original signal and the corresponding simulated one can also be heard following the link <https://www.dropbox.com/sh/3kuhooxuypgtmuh/AAAWiYNkdIIN3xUxajaV6HQPpa?dl=0>. In this case, it can be noted that the simulated sound is more different from the original than the first case considered, the *exper1*.

5. Conclusions

An approach has been proposed for constructing a stochastic model for creating jitter in a mechanical model that allows for producing voice and identifying two parameters of this model. A stochastic inverse problem was solved and the criterion used was the distance between the probability density functions related to the random variable associated

to the fundamental frequency. Two parameters were then identified and the pdf constructed from the simulated voices were so near of the pdf constructed from real voices. The corresponding sounds obtained were also available. The next idea is use other parameters, and not only the fundamental frequency, to solve the stochastic problem and verify if the sounds are more similar.

Acknowledgements

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

References

- [1] 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.
- [2] Hirose, H. *Clinical Aspects of Voice Disorders*. Interuna Publishers, Tokyo, 1998.
- [3] 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.
- [4] Flanagan J., Landgraf L. Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics*, AU-16 (1), 1968.
- [5] 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.
- [6] Titze, I. R., The human vocal cords: a mathematical model. Part I. *Phonetica*, 28, 129–170.
- [7] 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.
- [8] 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.
- [9] Cataldo E., Soize C. *Voice Signals Produced With Jitter Through a Stochastic One-mass Mechanical Model*. Volume 31, Issue 1, Pages 111.e9111.e18, 2017.
- [10] Flanagan J., Landgraf L. Self-oscillating source for vocal-tract synthesizers. *IEEE Transactions on Audio and Electroacoustics*, AU-16 (1), 1968.
- [11] Schoengten J., De Guchteneere R. Predictable and random components of jitter. *Speech Communication*, 21, 255–272, 1997.
- [12] Schoengten J., Fraj S., Lucero J. C. Testing the reliability of Grade, Roughness and Breathiness scores by means of synthetic speech stimuli. *Logopedics Phoniatics Vocology*, 1, 1–9, 2013.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Krée P., Soize C. *Mathematics of Random Phenomena*. Reidel, Dordrecht, 1996.
- [17] Soize C. *The Fokker-Planck Equation for Stochastic Dynamical Systems and its Explicit Steady State Solutions*. World Scientific, Singapore, 1994.