Journal of Biomechanics xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomechanics

journal homepage: www.elsevier.com/locate/jbiomech
www.JBiomech.com

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

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

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

ARTICLE INFO

Article history:
Accepted 14 April 2018
Available online xxxx

Keywords:
Voice production
Jitter
Stochastic mechanical models
Experimental identification
Statistical inverse problem

ABSTRACT

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

1. Introduction

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

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

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

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

Some mechanical models of jitter have been proposed by the same authors of this paper (Cataldo et al., 2012; Cataldo and Soize, 2016, 2017) and, now, a new mechanical stochastic model is then proposed but considering three control parameters, which gives more possibilities to generate jitter, including a way to

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

https://doi.org/10.1016/j.jbiomech.2018.04.031
0021-9290/© 2018 Elsevier Ltd. All rights reserved.

change the quality of the voice generated. A new parameter is introduced to discuss this quality, related to the bandwidth of the power spectral density function and, mainly, an inverse stochastic problem is solved to identify parameters and, consequently, to validate the model proposed. With these new possibilities, specific pathologies of the vocal folds can be created and identified, such as paralysis of the vocal folds.

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

2. Primary deterministic model

Fig. 1 illustrates a sketch of the model.

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

3. Stochastic modeling of jitter

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

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

where $a_1 = 1.87\,\frac{\ell d}{2}$ and $a_2 = \frac{\ell d}{2}$, with ℓ the length of each vocal fold and d the vocal fold thickness. The stochastic process $X(t)$ is the

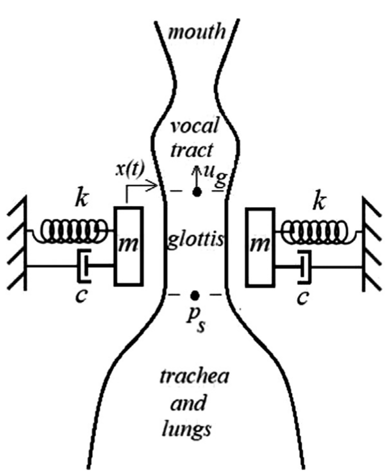

Fig. 1. Sketch based on the Flanagan and Landgraf (1968) model.

displacement of the mass m of one vocal fold, $K(t)$ is its stiffness and c is its damping coefficient when the glottis is opened; when the glottis is closed, there is an additional damping given by $c^*(X(t))$ described in the following, where the Bernoulli pressure $p_B(X(t), U_g(t))$ is also described.

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

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

  in which $\alpha > 0$ is a given damping rate.
- If $X(t) < x_0$ (the glottis is opened), then

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

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

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

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

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

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

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

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

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

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

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

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

Please cite this article in press as: Cataldo, E., Soize, C. Stochastic mechanical model of vocal folds for producing jitter and for identifying pathologies through real voices. J. Biomech. (2018), https://doi.org/10.1016/j.jbiomech.2018.04.031