Voice Signals Produced With Jitter Through a Stochastic One-mass Mechanical Model

***Edson Cataldo and †Christian Soize,** **Niteroi, RJ, Brazil, and †Marne-La-Vallée, France*

Summary: Background. The quasiperiodic 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.

Objective. The objective of this paper is the construction of a stochastic model for jitter using a one-mass mechanical model of the vocal folds, which assumes complete right-left symmetry of the vocal folds, and which considers motions of the vocal folds only in the horizontal direction.

Study design. The jitter has been the subject for researchers due to its important applications such as the identification of pathological voices (nodules in the vocal folds, paralysis of the vocal folds, or even, the vocal aging, among others). Large values for jitter variations can indicate a pathological characteristic of the voice.

Method. The corresponding stiffness of each vocal fold is considered as a stochastic process, and its modeling is proposed.

Results. 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.

Conclusions. It is showed that jitter could be obtained using the model proposed. The Praat software was also used to verify the measures of jitter in the synthesized voice signals.

Key Words: Stochastic modeling–Voice production–Mechanical models–Jitter–Mathematical models.

INTRODUCTION

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

Accepted for publication January 4, 2016.

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

Address correspondence and reprint requests to Edson Cataldo, Universidade Federal Fluminense, Applied Mathematics Departament and Graduate Program in Telecommunications Engineering, Rua Mário Santos Braga, S/N, Centro, Niteroi, RJ, CEP: 24020-140, Brazil. E-mail: ecataldo@im.uff.br

Journal of Voice, Vol. 31, No. 1, pp. 111.e9–111.e18
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.01.001

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

DETERMINISTIC MODEL USED

The deterministic model used as start is the nonlinear one-mass model proposed by Flanagan and Landgraf to generate voice. The complete model is composed by two subsystems: the subsystem of the vocal folds (*source*) and the subsystem of the vocal tract (*filter*). The two subsystems are coupled by the glottal flow. During the phonation, the filter is excited by the sequence of pulses of the glottal signal. Each vocal fold is represented by a mass-stiffness-damper system, and a symmetric system composed by two vocal folds is constituted. The vocal tract is represented by a standard configuration of concatenated tubes.[1,31] The complete model considered here presents some modifications in relation to the original Flanagan and Landgraf model. Some of them have been introduced by Ishizaka and Flanagan,

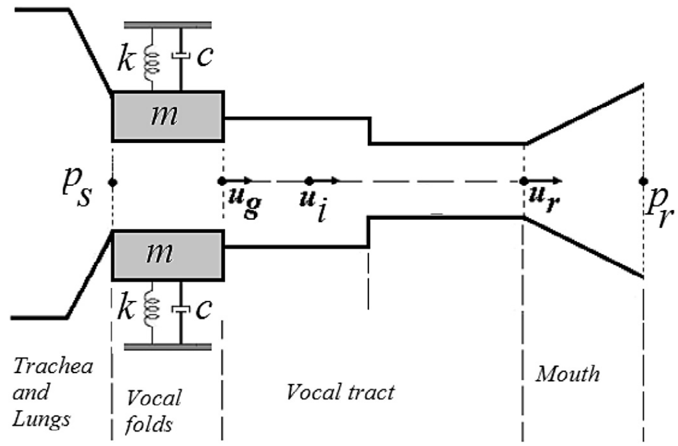

FIGURE 1. Sketch of the Flanagan and Landgraf model (1968).

and others by Cataldo et al.[26,32] The system of differential equations to be solved can be divided in three parts (see Figure 1 that illustrates a sketch of the model):

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

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

- The *coupling equation* is a scalar nonlinear integro-differential equation whose unknown time-dependent function is the real-valued function $t \mapsto u_g(t)$ that models the acoustic volume velocity through the glottis. This equation depends on the real-valued function $t \mapsto u_1(t)$ that is related to the first tube of the sound acoustic propagation into the vocal tract (see hereinafter).
- The *sound acoustic propagation equation* is constituted of $n+1$ scalar linear integro-differential equations for which the $n+1$ unknown time-dependent functions are the real-valued functions $t \mapsto u_1(t), \ldots, u_n(t), u_R(t)$.
- The *vocal folds dynamic equation* is a scalar nonlinear differential equation whose unknown time-dependent function is the real-valued function $t \mapsto x(t)$ that is the displacement of the mass of the vocal folds. For all t, $x(t)$ is generated by the *vocal folds dynamic equation*. The solution x of such an equation is constructed for all t and is used as follows:
  - The collision of the vocal folds occurs at a time t when $x(t)$ reaches a given critical value x_0 (defined after) and, at this time t, the glottis closes. The glottis remains closed so that the values $\{x(\tau), t \le \tau \le t'\}$ of x (that are