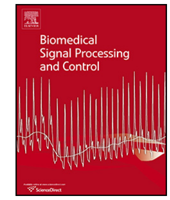
Biomedical Signal Processing and Control 27 (2016) 87–95

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



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

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

[a] Universidade Federal Fluminense, Applied Mathematics Department and Graduate program in Telecommunications Engineering, Rua Mário 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

ARTICLE INFO

Article history:
Received 21 October 2015
Received in revised form
29 December 2015
Accepted 5 February 2016
Available online 27 February 2016

Keywords:
Stochastic modeling
Voice production
Mechanical models
Jitter

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.

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 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 [1–4].

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 [5–8].

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 [9–14].

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 [15–18]. Erath et al. [19] made a good review of models used to produce voice, including mechanical models and pathological phonation. Deguchi and Kawahara [20] present a continuum-based numerical model of phonation to simulate human phonation with vocal nodules. Deguchi and Kawahara [20] have simulate human phonation with vocal nodules. Fraile et al. [21] 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 [16]. Previous works have discussed stochastic

* Corresponding author. Tel.: +55 2196318344.
E-mail addresses: ecataldo@im.uff.br (E. Cataldo), christian.soize@univ-paris-est.fr (C. Soize).

http://dx.doi.org/10.1016/j.bspc.2016.02.003
1746-8094/

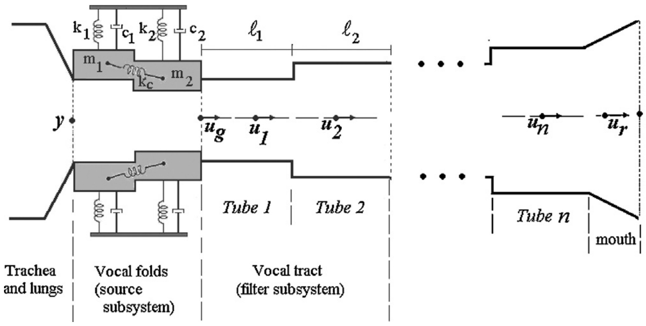


Fig. 1. Sketch of the Ishizaka and Flanagan model [16].

mechanical models to produce voice [22,23] 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 [24,25].

The complete model considered here presents some modifications in relation to the original Ishizaka and Flanagan model [26,27,22]. 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):

- A nonlinear integro-differential equation for the glottal flow that is coupled with the vocal tract, called the *coupling equation* (see Eqs. (A.1)–(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 $t \mapsto u_1(t), \ldots, u_n(t), u_R(t)$. 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 [25]. 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 the presence of pathologies). Some works discussing jitter and its application [28–31,13]. 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 [32,33,22,23].

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 $u_g(t), u_1(t), \ldots, u_n(t), u_R(t), x_1(t)$, and $x_2(t)$, become stochastic equations for the stochastic processes denoted by $U_g(t), U_1(t), \ldots, U_n(t), U_R(t), X_1(t)$, 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

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

in which

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

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

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