Uncertainties in mechanical models of larynx and vocal tract for voice production

E. Cataldo, Christian Soize, Christophe Desceliers, R. Sampaio

Proceedings of the XII International Symposium on Dynamic Problems of Mechanics (DINAME 2007),
P.S.Varoto and M.A.Trindade (editors), ABCM, Ilhabela, SP, Brazil, February 26 - March 2, 2007

Uncertainties in mechanical models of larynx and vocal tract for voice production

Edson Cataldo[1], **Christian Soize**[2], **Christophe Desceliers**[2], **Rubens Sampaio**[3]

[1] Departamento de Matemática Aplicada - Programa de Pós-graduação em Eng. de Telecomunicações - PGMEC, Universidade Federal Fluminense, Rua Mário Santos Braga, S/N, Centro, Niterói, RJ, Brasil - CEP: 24020-140

[2] Laboratoire de Mécanique, Université de Marne-La-Vallée, 5 Bd Descartes, 77454, Marne-la-Vallée Cedex 2, France

[3] Departamento de Engenharia Mecânica, Pontifícia Universidade Católica, Rua Marquês de São Vicente, 225, Gávea, Rio de Janeiro, RJ, Brasil - CEP: 22453-900

Abstract: We have reviewed a number of studies of voiced sound articulatory synthesis, but all of them have been conducted using deterministic models. However, real voices are random processes. This paper is concerned with the discussion of uncertainties and random variables involved in the process of voice production, based on the maximum entropy principle, and using the mathematical-mechanical model proposed by Ishizaka and Flanagan. We discuss an implicit numeric method for synthesizing voice, which allows us to solve the associated inverse dynamics problem; that is, to identify the parameters of the model, given a specific voice signal. We can, then, analyze the sensitivity and random characteristics of the frequencies involved in the process, due to uncertainties in the parameters.

Keywords: Voice production, Uncertainties, Identification, Modeling, Signal Processing.

NOMENCLATURE

a_{g_0} = neutral glottal area, area, m2
a = section or glottal area, area, m2
$\tilde{a}$ = coefficient in the Newmark method, dimensionless
A = random variable related to the area, area, m2
c = damping coefficient, damping, Ns/m
$\tilde{c}_n$ = expression related to the damping, m3/N
$\underline{c}_n$ = Fourier coefficients
$\mathbf{c}$ = vector related to the Fourier coefficients
$[C]$ = damping matrix, damping, Ns/m
d = stickness, length, m
f_0 = fundamental frequency, frequency, Hz
$\underline{f}$ = force acting in the mass, force, N
$\widehat{h}$ = transfer function of the vocal tract
$\mathbf{h}$ = force vector function, force, N
k = linear stiffness, stiffness, N/m
k_c = linear stiffness, stiffness, N/m
$[K]$ = stiffness matrix, stiffness, N/m
ℓ = length of the tube or the vocal fold, length, m
$\tilde{\ell}$ = expression related to the length, kg/m4
L = random variable related to the length, length, m
m = mass of the vocal cord, mass, kg
m_X = mean value of X
$[M]$ = mass matrix, mass, kg
N = number of points of the period
p_m = pressure acting in the mass, pressure, Pa
P_c = probability used for the confidence region, dimensionless
p = radiated pressure, pressure, N/m2
p_X = probability distribution of X
q = parameter to be identified, dimensionless
Q = random variable related to q, dimensionless
qu = quantile
r = acoustic resistance in the mouth, Pa/m3/s
$\widehat{s}_r$ = radiation impedance, Pa/m3/s
s = nonlinear stiffness function, stiffness, N/m
t = damping function, damping, Ns/m
T = period, time, s
$\widehat{u}$ = volume velocity, in the frequency domain, density, m3/Hz
u = air volume velocity across the tubes, velocity, m3/s
$\dot{u}$ = derivative of u, m3/s2
v_c = sound velocity, velocity, m/s
x^+ = upper envelope related to confidence region
x^- = lower envelope related to confidence region
x = displacement of the mass, distance, m
$\mathbf{w}$ = vector containing x_1, x_2, u_n
y = radius of the tubes, length, m
$\mathbf{z}_i$ = vector used in the Newmark method

Greek Symbols
α = index used in the algorithm
$\tilde{\alpha}$ = constant used in the Newmark method
β = index used in the algorithm
Γ = Gamma function
δ = dispersion coefficient used in the probability distribution
$\tilde{\delta}$ = constant used in the Newmark method, dimensionless
Δt =sampling time, time, s
η_h = nonlinear coefficient of the springs during collision, dimensionless
η_k = nonlinear coefficient of the springs, dimensionless
ρ = air density, density, kg/m3
μ = shear viscosity coefficient, viscosity, Pas
υ = function used in the coupling equation
$\tilde{\delta}$ = coefficient used in the Newmark method
ν = number of the Fourier coefficients
σ_X = standard deviation of X
θ = realization
ξ = damping ratio, dimensionless

Subscripts
g = relative to glottal
n = relative to the number of the tubes
r = relative to the radiated sound
1 = relative to the mass 1 in the model
2 = relative to the mass 2 in the model

Superscripts
exp = relative to the given signal
mod = relative to the model
reg = relative to the regenerated signal

INTRODUCTION

The voice production process has been studied by many researchers, and for many different reasons such as to obtain synthesis of voiced sounds (Ishizaka and Flanagan, 1972; Koizumi et al., 1987; Cataldo et al., 2006), to simulate pathological vocal-fold vibrations (Ishizaka and Isshiki, 1976; Zhang et al., 2005), to discuss nonlinearities related to the process (Steinecke and Herzel, 1995; Herzel et al., 1995; Lucero, 1999).

We are interested here in *phonation*, which is one of the larynx functions. The laryngeal complex is located between the pharynx and the trachea, and consists of a number of cartilages and muscles. In the larynx, we can find the vocal folds, small muscular cushions that adduct (come together) to close the laryngeal airway, or abduct (separate) to open this airway. The opening between the vocal folds is called the *glottis*, and the term *glottal* has come to be used as a general term for the laryngeal function. Sound results from the vibration of vocal folds, which alternately snap together and apart, colliding with one another in a basically (quasi-)periodic fashion.

The laryngeal function is highly similar within major groups of sounds produced. For example, vocal fold vibration differs little across vowels, which gain their distinctiveness by the shaping of the articulatory system above the larynx, i.e., the portion that goes from the glottis up to the mouth, called *vocal tract*. For this reason, the phonetic description of speech is based largely on supraglottal articulatory features.

The vocal folds, together with glottal airflow, constitute a highly nonlinear self-oscillating system. According to the accepted myoelastic theory of voice production proposed by Van den Berg (1968) and Titze (1980), the vocal folds are set into vibration by the combined effect of the subglottal pressure, the viscoelastic properties of the folds, and the Bernoulli effect. We cannot forget the coupling between the vocal folds dynamics and the vocal tract acoustics. The effective length, mass, and tension of the vocal folds are determined by muscle action; and in this way, the fundamental frequency (*pitch*) and the waveform of the glottal pulses can be controlled. The vocal tract acts as a filter which transforms the primary signals into meaningful voiced speech.

The two-mass model of the vocal folds (Ishizaka and Flanagan, 1972) has been widely used and the capability of this well-known model to reproduce the oscillation in detail has been successfully demonstrated. In addition to this model of the vocal folds, it is necessary to have a model for the vocal tract; and as usual, we use an acoustic tube for it. This complete model (vocal folds + vocal tract) will be subsequently referred as the IF72 model in this paper.

The system just discussed has been used for producing and studying voiced sounds in a deterministic way. However, the human voice production system is not deterministic. One of our purposes here is to take into account uncertainties, by using a probabilistic approach related to the parameters present in the IF72 model; or more generally, to the voice production process.

We divided our work, basically, into three parts:

(1) Modeling: we discuss an implicit numerical method to synthesize voiced sounds using the IF72 model.

(2) Inverse dynamics problem (deterministic): the fundamental frequency is controlled by the parameters of the vocal folds and the distinctiveness between the vowel sounds is related to the parameters of the vocal tract. Then, we discuss a way to solve the inverse problem; that is, given a recorded voice signal, how to identify the parameters of the IF72 model that best approach it.

(3) Randomization of the direct problem: the parameters of the vocal folds and the vocal tract are modeled as random variables by taking into account that the voice production is a stochastic process. An experimental database is constructed with experimental realizations of the stochastic process and, for each realization from that database, the random parameters of the model are identified by solving an optimization problem.

MODELING

Figure 1 shows a sketch of the IF72 model. Each vocal fold is represented by two (nonlinear) mass-damper-spring systems, coupled through a (linear) spring (k_c) and the vocal tract is represented by a standard two-tube configuration for vowel /a/ (Ishizaka and Flanagan, 1972; Titze, 1994).

The IF72 model assumes motion in the direction perpendicular to airflow only, which is in turn assumed to be quasi-steady and described by Bernoulli's energy equation. Other assumptions may be found in the cited references.

The complete model (vocal folds and vocal tract) is called a source-filter model (Fant, 1960; Ishizaka and Flanagan, 1972) and it is a well known representation of speech production acoustics, layered on the theory of linear systems. A fundamental assumption is that the vocal tract can be decomposed into dynamic and mutually independent parts (in our case, two) that interact linearly.

We divide the source-filter system in two parts: the subsystem of the vocal folds (source) and the subsystem of the vocal tract (filter). They are coupled by the glottal volume velocity.

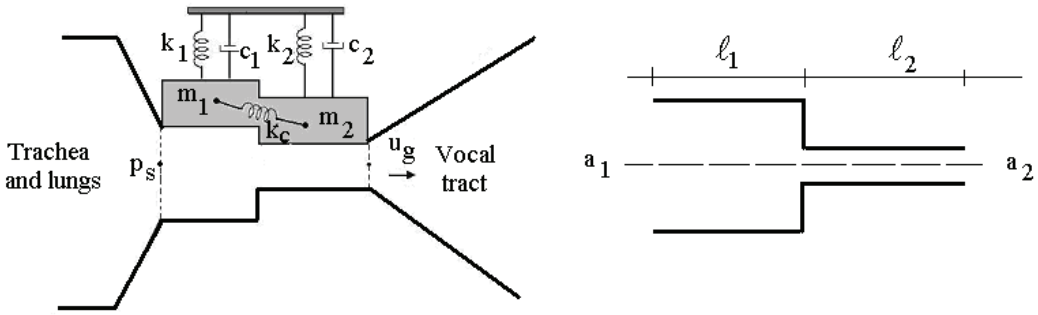

Figure 1 – Two-mass model of the vocal folds; vocal tract model for vowel /a/.

During phonation, the filter is excited by a sequence of airflow pulses with a frequency, say f_0, which we call the fundamental frequency of the voice signal. The full interaction between the glottal flow and vocal tract is included by solving the differential equations related to the airflow in the vocal tract, and the equations that specify the mechanical vibration mechanism of the vocal folds.

In summary, the output pressure (the voice produced) can be written, in the frequency domain, as

$$\widehat{p}_r(\omega) = \widehat{u}_g(\omega)\widehat{h}(\omega)\widehat{s}_r(\omega) \tag{1}$$

where $\widehat{p}_r(\omega)$ refers to the radiated sound pressure, $\widehat{u}_g(\omega)$ refers to the air volume velocity, $\widehat{h}(\omega)$ represents the transfer function of the vocal tract and $\widehat{s}_r(\omega)$ denotes radiation characteristic, in the mouth. Putting this equation into words, we could say that the radiated sound pressure waveform of speech is the product of the laryngeal spectrum, the vocal tract transfer function, and the radiation characteristic.

The subglottal pressure p_s is the input of the subsystem source, which output is the air volume velocity u_g. Consequently, u_g is the input of the subsystem filter, which output is the radiated pressure, denoted by p_r. If we consider the complete system, p_s is the input and the output is the radiated pressure p_r.

The complete IF72 model, can be described, in a simplified form, by Eq. 2 and Eq. 3

$$\phi_1(\mathbf{w})|u_g|u_g + \phi_2(\mathbf{w})u_g + \phi_3(\mathbf{w})\dot{u}_g + \frac{1}{\tilde{c}_1}\int_0^t (u_g(\tau) - w_3(\tau))d\tau - p_s = 0 \tag{2}$$

$$[M]\ddot{\mathbf{w}} + [C]\dot{\mathbf{w}} + [K]\mathbf{w} + \mathbf{h}(\mathbf{w}, \dot{\mathbf{w}}, u_g, \dot{u}_g) = 0 \tag{3}$$

where

$$\mathbf{w}(t) = \begin{pmatrix} w_1(t) \\ w_2(t) \\ w_3(t) \\ w_4(t) \\ w_5(t) \end{pmatrix} = \begin{pmatrix} x_1(t) \\ x_2(t) \\ u_1(t) \\ u_2(t) \\ u_r(t) \end{pmatrix} \tag{4}$$

The functions $t \mapsto x_1(t)$ and $t \mapsto x_2(t)$ are the displacements of the masses, $t \mapsto u_1(t)$ and $t \mapsto u_2(t)$ describe the air volume velocity through the (two) tubes that model the vocal tract and $t \mapsto u_r(t)$ is the air volume velocity through the mouth. The function, that we call radiated pressure, $t \mapsto p_r(t)$, is defined by Eq. 5.

$$p_r(t) = u_r(t)r_r \tag{5}$$

where $r_r = \frac{128\rho v_c}{9\pi^3 y_2^2}$, ρ is the air density, v_c is the sound velocity and y_2 is the radius of the second tube.

And also,

$$[M] = \begin{bmatrix} m_1 & 0 & 0 & 0 & 0 \\ 0 & m_2 & 0 & 0 & 0 \\ 0 & 0 & \tilde{\ell}_1+\tilde{\ell}_2 & 0 & 0 \\ 0 & 0 & 0 & \tilde{\ell}_2+\tilde{\ell}_r & -\tilde{\ell}_r \\ 0 & 0 & 0 & -\tilde{\ell}_r & \tilde{\ell}_r \end{bmatrix}, \quad [C] = \begin{bmatrix} r_1 & 0 & 0 & 0 & 0 \\ 0 & r_2 & 0 & 0 & 0 \\ 0 & 0 & r_1+r_2 & 0 & 0 \\ 0 & 0 & 0 & r_2 & 0 \\ 0 & 0 & 0 & 0 & r_r \end{bmatrix}, \tag{6}$$

$$[K] = \begin{bmatrix} k_1+k_c & -k_c & 0 & 0 & 0 \\ -k_c & k_2+k_c & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\tilde{c}_1}+\frac{1}{\tilde{c}_2} & -\frac{1}{\tilde{c}_2} & 0 \\ 0 & 0 & -\frac{1}{\tilde{c}_2} & \frac{1}{\tilde{c}_2} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \mathbf{h}(\mathbf{w},\dot{\mathbf{w}},u_g,\dot{u}_g) = \begin{bmatrix} s_1(w_1)+t_1(w_1)\dot{w}_1 - f_1(w_1,u_g,\dot{u}_g) \\ s_2(w_2)+t_2(w_2)\dot{w}_2 - f_2(w_1,w_2,u_g,\dot{u}_g) \\ -\frac{1}{\tilde{c}_1}u_g \\ 0 \\ 0 \end{bmatrix}, \tag{7}$$

where

$\tilde{\ell}_n = \frac{\rho\ell_n}{2\pi y_n^2}$, $\tilde{\ell}_r = \frac{8\rho}{3\pi^2 y_n}$, $r_n = \frac{2}{y_n}\sqrt{\rho\mu\frac{\omega}{2}}$, $\omega = \sqrt{\frac{k_1}{m_1}}$, $a_n = \pi y_n^2$, $\tilde{c}_n = \frac{\ell_n \pi y_n^2}{\rho v_c^2}$, ℓ_n is the length of the n th tube, y_n is the radius of the n th tube, and μ is the shear viscosity coefficient.

The functions $\mathbf{w} \mapsto \phi_1(\mathbf{w})$, $\mathbf{w} \mapsto \phi_2(\mathbf{w})$, $\mathbf{w} \mapsto \phi_3(\mathbf{w})$ and $(\mathbf{w},\dot{\mathbf{w}},u_g,\dot{u}_g) \mapsto \mathbf{h}(\mathbf{w},\dot{\mathbf{w}},u_g,\dot{u}_g)$; and also the functions $w_1 \mapsto s_1(w_1)$, $w_2 \mapsto s_2(w_2)$, $t_1 \mapsto s_1(t_1)$, $t_2 \mapsto s_2(t_2)$, $(w_1,u_g,\dot{u}_g) \mapsto f_1(w_1,u_g,\dot{u}_g)$ and $(w_1,w_2,u_g,\dot{u}_g) \mapsto f_2(w_1,w_2,u_g,\dot{u}_g)$ are described in the appendix.

In order to solve the system (Eq. 2 and Eq. 3); that is, find u_g and $\mathbf{w}$, given p_s, a *centered finite difference* scheme is used for Eq. 2 and an unconditionnally stable Newmark scheme is used for Eq. 3. All of the values used here are the same used by Ishizaka and Flanagan (1972).

Let Δt be the sampling time and $\mathbf{w}_i = \mathbf{w}(i\Delta t)$, $\dot{\mathbf{w}}_i = \dot{\mathbf{w}}(i\Delta t)$, $\ddot{\mathbf{w}}_i = \ddot{\mathbf{w}}(i\Delta t)$, $u_{g_i} = u_g(i\Delta t)$ and $\dot{u}_{g_i} = \dot{u}_g(i\Delta t)$.

Then, for all $i \geq 1$, we can write

$$\phi_1(\mathbf{w}_i)|u_{g_i}|u_{gi} + \phi_2(\mathbf{w}_i)u_{g_i} + \phi_3(\mathbf{w}_i)\frac{1}{\Delta t}(u_{g_i} - u_{g_{i-1}}) + \frac{1}{\tilde{c}_1}\Delta t\sum_{k=0}^{i-1}(u_{g_k} - w_{1_k}) - p_s = 0 \tag{8}$$

and

$$[A]\mathbf{w}_i + \mathbf{h}\left(\mathbf{w}_i, \frac{\mathbf{w}_i - \mathbf{w}_{i-1}}{\Delta t}, u_{g_i}, \frac{u_{g_i} - u_{g_{i-1}}}{\Delta t}\right) = \mathbf{z}_i \tag{9}$$

where

$$\begin{aligned} [A] &= [K] + \tilde{a}_0[M] + \tilde{a}_1[C] \\ \mathbf{z}_i &= [M](\tilde{a}_0\mathbf{w}_{i-1} + \tilde{a}_2\dot{\mathbf{w}}_{i-1} + \tilde{a}_3\ddot{\mathbf{w}}_{i-1}) + [C](\tilde{a}_1\mathbf{w}_{i-1} + a_4\dot{\mathbf{w}}_{i-1} + \tilde{a}_5\ddot{\mathbf{w}}_{i-1}) \end{aligned} \tag{10}$$

and

$$\begin{cases} \ddot{\mathbf{w}}_i = \tilde{a}_0(\mathbf{w}_i - \mathbf{w}_{i-1}) - \tilde{a}_2\dot{\mathbf{w}}_{i-1} - \tilde{a}_3\ddot{\mathbf{w}}_{i-1} \\ \dot{\mathbf{w}}_i = \dot{\mathbf{w}}_{i-1} + \tilde{a}_6\ddot{\mathbf{w}}_{i-1} + \tilde{a}_7\ddot{\mathbf{w}}_i \\ \tilde{a}_0 = \frac{1}{\tilde{\alpha}\Delta t^2}, \tilde{a}_1 = \frac{\tilde{\delta}}{\tilde{\alpha}\Delta t}, \tilde{a}_2 = \frac{1}{\tilde{\alpha}\Delta t} \\ \tilde{a}_4 = \frac{\tilde{\delta}}{\tilde{\alpha}} - 1, \tilde{a}_5 = \frac{\tilde{\delta}}{2}\left(\frac{\tilde{\delta}}{\tilde{\alpha}} - 2\right), \tilde{a}_6 = \Delta t(1-\tilde{\delta}), \tilde{a}_7 = \tilde{\delta}\Delta t \end{cases} \tag{11}$$

with $u_{g_0} = 0$, $\mathbf{w_0} = 0$, $\dot{\mathbf{w}}_\mathbf{0} = 0$, $\ddot{\mathbf{w}}_\mathbf{0} = 0$, $\tilde{\delta} = 0.5$ and $\tilde{\alpha} = 0.25$.

Algorithm used

The method used to construct the solution of Eq. 8 and Eq. 9 consists in finding u_{g_i} as the limit of the sequence $\{u_{g_i}^\alpha\}$, $\alpha \geq 0$, when α tends to infinity such that, for all $\alpha \geq 1$, $i \geq 1$,

$$\phi_1(\mathbf{w}_i^{\alpha-1})|u_{g_i}^\alpha|u_{g_i}^\alpha + \phi_2(\mathbf{w}_i^{\alpha-1})u_{g_i}^\alpha + \phi_3(\mathbf{w}_i^{\alpha-1})\frac{1}{\Delta t}(u_{g_i}^\alpha - u_{g_{i-1}}) + \frac{1}{\tilde{c}_1}\Delta t\sum_{k=0}^{i-1}(u_{g_k} - w_{1_k}) - p_s = 0\,, \tag{12}$$

with $\mathbf{w}_i^0 = \mathbf{w}_{i-1}$, $u_{g_i}^0 = u_{g_{i-1}}$.

In Eq. 12, $\mathbf{w}_i^{\alpha-1}$ is the limit of the sequence $\{\mathbf{w}_i^{\alpha-1,\beta}\}$, $\beta \geq 0$, when β tends to infinity; such that for all $\beta \geq 1$, $\alpha > 1$, $i \geq 1$,

$$[A]\mathbf{w}_i^{\alpha-1,\beta} = \mathbf{z}_i - \mathbf{h}\left(\mathbf{w}_i^{\alpha-1,\beta-1}, \frac{\mathbf{w}_i^{\alpha-1,\beta-1} - \mathbf{w}_{i-1}}{\Delta t}, u_{g_i}^{\alpha-1}, \frac{u_{g_i}^{\alpha-1} - u_{g_{i-1}}}{\Delta t}\right) \tag{13}$$

with $\mathbf{w}_i^{\alpha,0} = \mathbf{w}_i^{\alpha-1}$.

In other words, our strategy consists in considering two loops (outer α and inner β). For each iteration i, we find u_g and we use this value for finding $\mathbf{w}$, but with this value of $\mathbf{w}$ found, we must return to correct u_g and we repeat this process up to the tolerances are reached.

As an example, we simulated a vowel /a/, considering two tubes for the vocal tract and the same data used by Ishizaka and Flanagan (1972): $d_1 = 0.25\,cm$, $d_2 = 0.05\,cm$, $a_{g_0} = 0.05\,cm^2$, $p_s = 8000\,Pa$, $m_1 = 0.125\,g$, $m_2 = 0.025\,g$, $k_1 = 80.000\,dyn/cm$, $k_2 = 8.000\,dyn/cm$, $k_c = 25.000\,dyn/cm$, $\xi_1 = 0.1$, $\xi_2 = 0.6$, $\eta_{k_1} = \eta_{k_2} = 100$, $\eta_{h_1} = \eta_{h_2} = 500$. The graphs obtained are showed in the Fig. 2. These results are coherent with those presented by Ishizaka and Flanagan (1972).

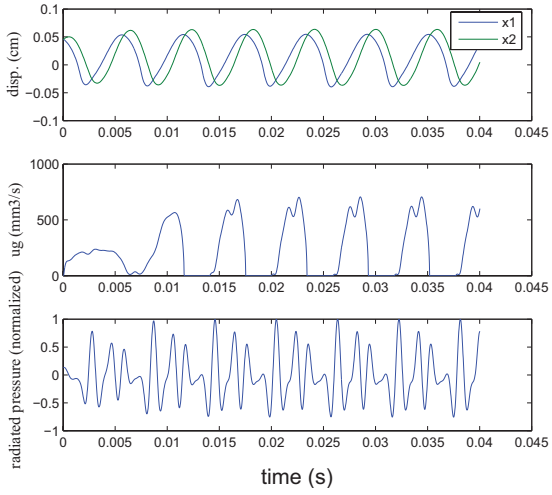

Figure 2 – Displacements of the two masses (x_1 and x_2); glottal volume velocity (u_g); radiated pressure (p_r). Production of a vowel /a/.

INVERSE DYNAMICS PROBLEM (DETERMINISTIC)

Most speech simulation models are based on the assumption of one-dimensional wave propagation, as the IF72 model. This means that the tubular vocal tract shape can be approximated as a finite number of cylindrical elements that are *stacked* consecutively from the larynx to the mouth. A particular vocal tract shape can be imposed on a model by specifying the cross-sectional area of each cylindrical element as a function of the distance from the glottis. For modelling purposes, any vocal tract shape can be defined by its unique *area function*. Hence, a necessary component for the simulation of natural sounding speech is an inventory of vocal tract area functions that correspond to the vowels (and consonants) used to produce human speech. The success of speech simulators has been limited, in part, by the lack of a body of morphological information about the vocal tract shape on which to base these area functions and many efforts have been made to find the description of the vocal tract configurations for vowels relative to their acoustic output (Fant, 1960; Adachi and Yamada, 1999; Titze and Story, 1996; Takemoto and Honda, 2006). Here, we will discuss a method for finding the parameters of the vocal tract, from a given voice signal.

We also introduced a factor q, as done by Ishizaka and Flanagan, called *tension parameter*, that controls the fundamental frequency f_0 of the vocal folds oscillation, because vocal fold abduction and tension should be the main factor used by speakers to control phonation (Koenig, 2000; McGowan et al., 1995). We write $k_1 = q^2\widehat{k}_1$, $k_2 = q^2\widehat{k}_2$, $k_c = q^2\widehat{k}_c$ and we control the parameter q, where $\widehat{k}_1$, $\widehat{k}_2$ and $\widehat{k}_c$ are values fixed (we used the same values that Ishizaka and Flanagan (1972)).

One of the objectives of this paper is to identify the values of q (*tension parameter*), ℓ_1, ℓ_2, a_1 and a_2 (geometrical

parameters of the vocal tract) corresponding to the recorded signal, by solving an optimization problem in terms of q, ℓ_1, ℓ_2, a_1 and a_2 in a mean-square sense. This identifies the appropriated parameters to reconstruct a recorded signal.

This optimization problem consists in: (1) finding the value of q that minimizes the function $q \mapsto |f_0^{exp} - f_0^{mod}(q)|^2$, where f_0^{exp} is the fundamental frequency of the recorded signal and $f_0(q)$ is the fundamental frequency of the signal generated by the model, for a given q; (2) finding the values of ℓ_1, ℓ_2, a_1, a_2 that minimizes the function $(\ell_1, \ell_2, a_1, a_2) \mapsto ||\mathbf{c}^{mod}(\ell_1, \ell_2, a_1, a_2) - \mathbf{c}^{exp}||^2$.

The components of the vector $\mathbf{c}^{mod}$ are the Fourier coefficients obtained from an signal generated by the model $t \mapsto p_r^{mod}(t)$, where $p_r^{mod}(t) = u_r(t)\, r_r$ and the components of the vector $\mathbf{c}^{exp}$ are the coefficients of a *pseudo* Fourier expansion of the quasi-periodic recorded signal $t \mapsto p_r^{exp}(t)$; $p_r^{exp}(t)$ is the radiated pressure obtained from a real voice at time t.

We write $\mathbf{c}^{exp} = (\underline{c}_{-\nu}^{exp}, \ldots, \underline{c}_{\nu}^{exp})$ and $\mathbf{c}^{mod} = (c_{-\nu}^{mod}, \ldots, c_{\nu}^{mod})$, where

$$\underline{c}_n^{exp} = \frac{1}{N - N^{exp}} \Delta t \sum_{m=0}^{N-N^{exp}} c_n^{exp,m} \ , \ -\nu \le n \le \nu \ , \tag{14}$$

$$c_n^{exp,m} = \frac{1}{T^{exp}} \Delta t \sum_{k=m}^{m+N^{exp}-1} p_r^{exp}(k\Delta t) exp(-jn2\pi f_0^{exp} k\Delta t) \ , \ T^{exp} = \frac{1}{f_0^{exp}} \ , \ N^{exp} = \frac{T^{exp}}{\Delta t} \ , \ N = \frac{T}{\Delta t} \ , \tag{15}$$

$$c_n^{mod} = \frac{1}{T^{mod}} \Delta t \sum_{k=0}^{N^{mod}-1} p_r^{mod}(k\Delta t) exp(-jn2\pi f_0^{mod} k\Delta t) \ , \ -\nu \le n \le \nu \ , \ T^{mod} = \frac{1}{f_0^{mod}} \ , \ N^{mod} = \frac{T^{mod}}{\Delta t} \ , \ N = \frac{T^{mod}}{\Delta t} \ . \tag{16}$$

We will consider a signal, collected from a brazilian person, speaking a (sustained) vowel /a/. The data were read into MATLAB software, which was used for all further processing.

Figure. 3 shows the signal recorded, in the time domain and in the frequency domain.

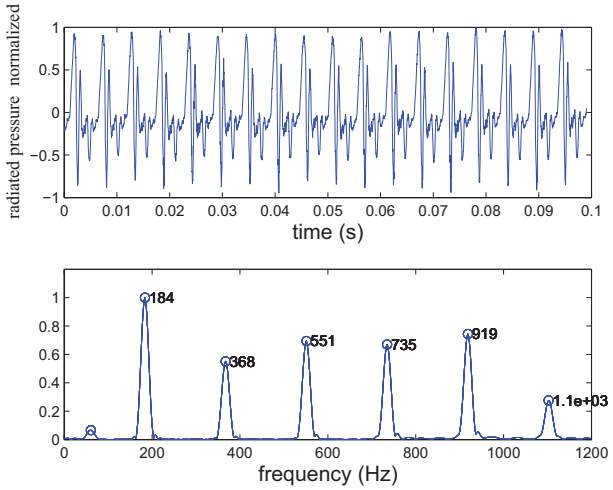

Figure 3 – Signal recorded. (a) signal in the time domain; (b) signal in the frequency domain.

Figure 4 shows the graph of the recorded signal, $t \mapsto p_r^{exp}(t)$, and the graph of the signal $t \mapsto p_r^{reg}(t)$, given by

$$p_r^{reg}(t) = \sum_{n=-\nu}^{\nu} \underline{c}_n^{exp} exp(jn2\pi f_0^{exp} t)$$

using $\nu = 30$ *pseudo* Fourier coefficients.

We can observe that the regenerated signal is well reproduced.

RANDOMIZATION OF THE DIRECT PROBLEM

Up to now, we have discussed the deterministic problem, direct and inverse. However, our principal goal is to analyse the system when uncertainties are present.

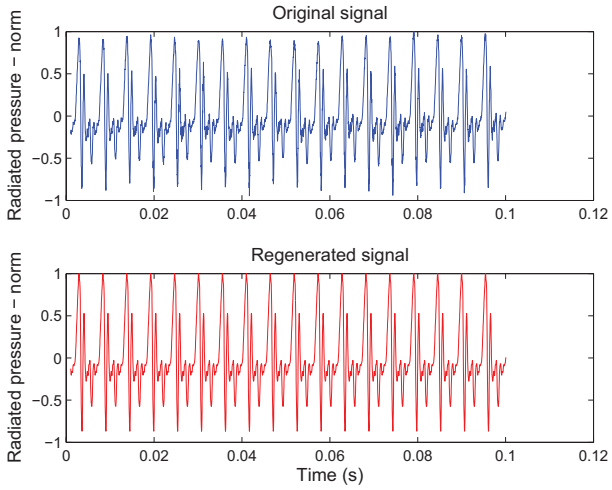

Figure 4 – Signal recorded: section of the original signal; section of the regenerated signal.

We collected three signals, frow two different people (P_1 and P_2), each one spoke the sustained vowel /a/. Then, for each signal, we solved the inverse problem discussed in the previous section. We found the values showed in the Tab. 1.

Table 1 – Identified parameters for P_1 and P_2.

	Parameters	q	a_1 (cm^2)	a_2 (cm^2)	ℓ_1 (cm)	ℓ_2 (cm)
P_1	Signal 1	0.6300	1.5975	5.9258	9.9780	9.4872
	Signal 2	0.6000	1.8052	5.8557	8.6928	8.9612
	Signal 3	0.6150	1.7948	5.9964	8.2503	9.4466
P_2	Signal 1	1.25	1.66	6.1871	8.6303	9.2965
	Signal 2	1.25	1.6576	6.0046	9.1714	9.1696
	Signal 3	1.27	1.7900	5.1090	9.2100	9.1600

As we can see, the values for the parameter q and for the parameters of the vocal are different, for each signal produced, as we expected.

We consider, then, the random variables Q, A_1, A_2, L_1 and L_2 related to the variables q, a_1, a_2, l_1 and l_2 and we are going to study the results obtained from the system when these quantities vary.

Construction of the probabilistic model for the random parameters

To take into account the uncertainties, a probabilist model is constructed (Soize, 2000), which consists in modeling the parameters q, a_1, a_2, ℓ_1 and ℓ_2 as random variables Q, A_1, A_2, L_1 and L_2, respectively.

The probabilistic model of each random variable (Q, A_1, A_2, L_1 and L_2) is constructed taking into account the following available informations:

(1) $E\{Q\} = q, E\{A_1\} = a_1, E\{A_2\} = a_2, E\{L_1\} = \ell_1, E\{L_2\} = \ell_2$

(2) $E\{Q^{-2}\} < +\infty, E\{A_1^{-2}\} < +\infty, , E\{A_2^{-2}\} < +\infty, E\{L_1^{-2}\} < +\infty, E\{L_2^{-2}\} < +\infty$

where $E\{.\}$ denotes the mathematical expectation operator. The maximum entropy principle (Shannon, 1948) yields the following probability distribution

$$p_X(x) = 1_{]0,+\infty[}(x) \frac{1}{m_X} \left(\frac{1}{\delta^2}\right)^{\frac{1}{\delta^2}} \frac{1}{\Gamma(1/\delta^2)} \left(\frac{x}{m_X}\right)^{\frac{1}{\delta^2}-1} exp\left(-\frac{x}{\delta^2 m_X}\right) \tag{17}$$

where X represents Q, L_1, L_2, A_1, A_2; $\delta = \frac{\sigma_X}{m_X}$ is a dispersion coefficient such that $0 \leq \delta \leq 1/\sqrt{2}$; σ_X is the standard deviation of X; m_X is the mean value of X and $\alpha \mapsto \Gamma(\alpha)$ is the Gamma function defined by $\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt$.

Confidence region for the random frequencies

We construct the confidence region associated with a probability level P_c (we considered $P_c = 0.95$) for each one of the frequencies (we considered the three first significant frequencies in the spectrum of the signals). The confidence region is constructed using the quantiles (Serfling, 1980). Let $F_X(\xi)(x)$ be the distribution function of random variable $X(\xi)$. For $0 < p < 1$, the pth quantile (denoted by $qu(p)$) of F_X is defined as

$$qu(p) = inf\{x : F_X(\xi)(x) \geq p\} \, . \tag{18}$$

Then, the upper envelope $x^+(\xi)$ and the lower envelope $x^-(\xi)$ of the confidence region are defined by

$$x^+(\xi) = qu((1+P_c)/2) \text{ and } x^-(\xi) = qu((1+P_c)/2) \tag{19}$$

Let $x_1(\xi) = X(\xi;\theta_1), \ldots, x_n(\xi) = X(\xi;\theta_n)$ be the n independent realizations of random variable $X(\xi)$.
Let $\tilde{x}_1(\xi) < \ldots < \tilde{x}_n(\xi)$ be the order statistics associated with $x_1(\xi), \ldots, x_n(\xi)$. Therefore, one has the following estimation:

$$x^+(\xi) \simeq \tilde{x}_{j^+}(\xi) \, , \; j^+ = fix(n(1+P_c)/2) \text{ and } x^-(\xi) \simeq \tilde{x}_{j^-}(\xi) \, , \; j^- = fix(n(1-P_c)/2) \, , \tag{20}$$

in which $fix(z)$ is the integer part of the real number z.

First, the geometrical parameters of the vocal tract are modeled as random variables A_1, A_2, L_1 and L_2, while the *tension parameter* is fixed at $q = 1.25$. Figure 5(a) shows the confidence region, obtained in this case, for the first three frequencies considered. It can be seen that these frequencies are not sensitive to the geometrical parameters.

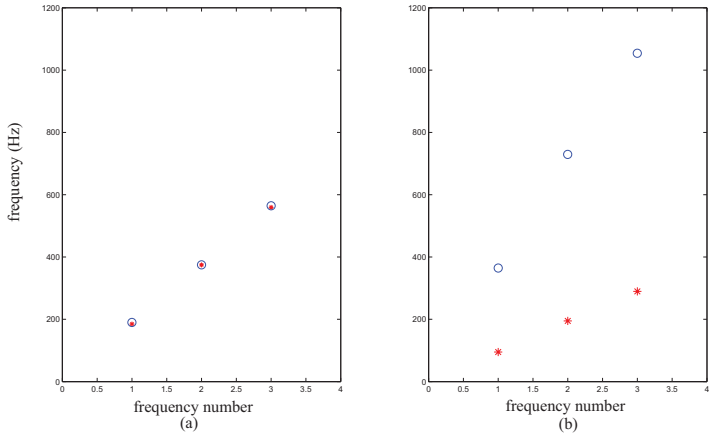

Figure 5 – Confidence region for the first three frequencies: $* \rightarrow x^-$ and $o \rightarrow x^+$. (a) Geometrical parameters are random variables and $q = 1.25$; (b) *Tension parameter* is random variable and $a_1 = 1.7\,cm^2$, $a_2 = 5.9\,cm^2$, $\ell_1 = 9.0\,cm$, $\ell_2 = 9.0\,cm$.

Secondly, the *tension parameter* is considered as the random variable Q and the geometrical parameters of the vocal tract are values fixed at $a_1 = 1.7\,cm^2$, $a_2 = 5.9\,cm^2$, $\ell_1 = 9.0\,cm$ and $\ell_2 = 9.0\,cm$. Figure 5 (b) shows the confidence region obtained for the three first frequencies and it can be seen, in this case, that these frequencies are very sensitive to the *tension parameter*.

The values considered for the realizations above were: $E\{Q\} = 1.25$, $E\{A_1\} = 1.7\,cm^2$, $E\{A_2\} = 5.9\,cm^2$, $E\{L_1\} = 9.0\,cm$, $E\{L_2\} = 9.0\,cm$ and $\frac{\delta}{E\{Q\}} = 0.4$, $\frac{\delta}{E\{A_1\}} = 0.4$, $\frac{\delta}{E\{A_2\}} = 0.2$, $\frac{\delta}{E\{L_1\}} = 0.1$, $\frac{\delta}{E\{L_2\}} = 0.1$.

These two graphs justify the method used for solving the inverse problem, for which two independent optimization problems were set, one for the *tension parameter*, related to the fundamental frequency, and other for the geometrical parameters of the vocal tract, related to the *pseudo* Fourier coefficients of the signal.

Figure 6, we show the results for the probability distribution for Q, whose analytical expression is given by Eq. 17 (with $X = Q$), and we show also the probability distribution for the fundamental frequency, which were obtained from the model, considering Q as a random variable.

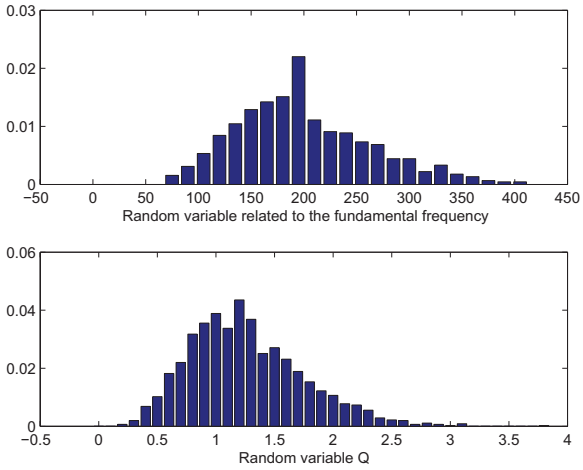

Figure 6 – Probability distribution for the fundamental frequency (top); Probability distribution for Q (bottom).

CONCLUSIONS

We used an implicit numerical method for solving the system of differential equations related to the voice production problem, modeled by Ishizaka and Flanagan (1972). The results for the direct problem are consistent with those previously presented in the literature. This particular method was chosen because not only it is more efficient in terms of computational cost, but it also allows the solution of the associated inverse dynamics problems; that is, the identification of parameters used in the model, given a recorded signal. The inverse problem was divided in two parts, one to identify the parameter that controls the fundamental frequency, and another to identify the geometrical parameters of the vocal tract. The results obtained here support the strategy used and also confirm that the voice production system is stochastic. In addition, we constructed a probabilistic model to analyze the uncertainties in the parameters involved in the process, considering only the available information and the maximum entropy principle. We also investigated the sensitivity of some frequencies involved in the process to the parameter uncertainty. Finally, we constructed the probability distribution for the fundamental frequency, considering the *tension parameter* as a random variable, using the maximum entropy principle. These results indicate our next step, that will be pursued in a future work: to solve the stochastic inverse problem; that is, to identify the statistics of the parameters involved in the process of voice production.

APPENDIX

$$\phi_1(\mathbf{w}) = \left(\frac{0.19\rho}{a_{g_0}+2\ell_g w_1} + 2\ell_g w_1\right) + \frac{\rho}{(a_{g_0}+2\ell_g w_2)^2}\left[0.5 - \frac{a_{g_0}+2\ell_g w_2}{a_1}\left(1 - \frac{a_{g_0}+2\ell_g w_2}{a_1}\right)\right]$$

$$\phi_2(\mathbf{w}) = \left(12\mu\ell_g \frac{d_1}{(a_{g_0}+2\ell_g w_1)^3} + 12\ell_g^2 \frac{d_2}{(a_{g_0}+2\ell_g w_2)^3} + r_1\right) \;,\; \phi_3(\mathbf{w}) = \left(\frac{\rho d_1}{a_{g_0}+2\ell_g w_1} + \frac{\rho d_2}{a_{g_0}+2\ell_g w_2} + \tilde{\ell}_1\right)$$

$$s_1(w_1) = \begin{cases} k_1\eta_{k_1}w_1^3 \;,\; w_1 > -\frac{a_{g_0}}{2\ell_g} \\ k_1\eta_{k_1}w_1^3 + 3k_1\left\{\left(w_1 + \frac{a_{g_0}}{2\ell_g}\right) + \eta_{h_1}\left(w_1 + \frac{a_{g_0}}{2\ell_g}\right)^3\right\}, w_1 \le -\frac{a_{g_0}}{2\ell_g} \end{cases}$$

$$s_2(w_2) = \begin{cases} k_2\eta_{k_2}w_2^3 \;,\; w_2 > -\frac{a_{g_0}}{2\ell_g} \\ k_2\eta_{k_2}w_2^3 + 3k_2\left\{\left(w_2 + \frac{a_{g_0}}{2\ell_g}\right) + \eta_{h_2}\left(w_2 + \frac{a_{g_0}}{2\ell_g}\right)^3\right\}, w_2 \le -\frac{a_{g_0}}{2\ell_g} \end{cases}$$

$$t_1(w_1) = \begin{cases} 0\,, w_1 > -\frac{a_{g_0}}{2\ell_g} \\ 2\xi\sqrt{m_1 k_1}\,, w_1 \le -\frac{a_{g_0}}{2\ell_g} \end{cases} \quad , \quad t_2(w_2) = \begin{cases} 0\,, w_2 > -\frac{a_{g_0}}{2\ell_g} \\ 2\xi\sqrt{m_2 k_2}\,, w_2 \le -\frac{a_{g_0}}{2\ell_g} \end{cases}$$

$$f_1(w_1, u_g, \dot{u}_g) = \begin{cases} \ell_g d_1 p_{m_1}(w_1, u_g, \dot{u}_g)\,, w_1 > -\frac{a_{g_0}}{2\ell_g} \\ 0\,, otherwise \end{cases}$$

$$p_{m_1}(w_1, u_g, \dot{u}_g) = p_s - 1.37\frac{\rho}{2}\left(\frac{u_g}{a_{g_0}+2\ell_g w_1}\right)^2 - \frac{1}{2}\left(12\mu\ell_g \frac{d_1}{(a_{g_0}+2\ell_g w_1)^3} + \frac{\rho d_1}{a_{g_0}+2\ell_g w_1}\right)\dot{u}_g$$

$$f_2(w_1,w_2,u_g,\dot{u}_g) = \begin{cases} \ell_g d_2 p_{m_2}(w_1,w_2,u_g,\dot{u}_g)\,, w_1 > -\frac{a_{g0}}{2\ell_g} \text{ and } w_2 > -\frac{a_{g0}}{2\ell_g} \\ \ell_g d_2 p_s\,, w_1 > -\frac{a_{g0}}{2\ell_g} \text{ and } w_2 \leq -\frac{a_{g0}}{2\ell_g} \\ 0\,,\ otherwise \end{cases}$$

$$p_{m_2}(w_1,w_2,u_g,\dot{u}_g) = p_{m_1} - *$$
$$* = \frac{1}{2}\left\{(12\mu\ell_g \frac{d_1}{(a_{g0}+2\ell_g w_1)^3} + 12\ell_g^2 \frac{d_2}{(a_{g0}+2\ell_g w_2)^3})u_g + (\frac{\rho d_1}{a_{g0}+2\ell_g w_1} + \frac{\rho d_2}{a_{g0}+2\ell_g w_2})\dot{u}_g\right\} - \frac{\rho}{2}u_g^2\left(\frac{1}{(a_{g0}+2\ell_g w_2)^2} - \frac{1}{(a_{g0}+2\ell_g w_1)^2}\right)$$

ACKNOWLEDGMENTS

This work was founded by the International Cooperation Project CAPES-COFECUB N.476/04, by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and by FAPERJ (Fundação Carlos Chagas Filho de Amparo Pesquisa do Estado do Rio de Janeiro).

REFERENCES

Adachi, S.,Yamada, M.,1999, "An acoustical study of sound production in biphonic singing, Xmij", J. Acoust. Soc. Am., Vol. 105, pp. 2920-2932.

Cataldo, E., Leta, F. R., Lucero, J., Nicolato, L., 2006, "Synthesis of voiced sounds using low-dimensional models of the vocal cords and time-varying subglottal pressure", Mechanics Research Communications, Vol. 33, pp. 250-260.

Fant, G., 1960, "The acoustic theory of speech production", Mouton, The Hague.

Goldstein, 1980, "An articulatory model for the vocal tracts of growing children", Doctoral dissertation, Massachussetts Institute of Technology, Cambridge, MA.

Herzel, H., Berry, D., Titze, I., Steinecke, I., 1995, "Nonlinear dynamics of the voice: Signal analysis and biomechanical modeling", Chaos, Vol. 5 (1).

Ishizaka, K., Flanagan, J.L., 1972,"Synthesis of voiced sounds from a two-mass model of the vocal cords", Bell Syst. Tech. J., Vol. 51, pp. 1233-1268.

Ishizaka, K., Isshiki, N., 1976, "Computer simulation of pathological vocal-cord vibration", J. Acoust. Soc. Am., Vol. 60, No. 5, pp. 1193-1198.

McGowan, R. S., Koenig, L.L. and Löfqvist, A., 1995, "Vocal tract aerodynamics in /aCa/ utterances: Simulations", Speech Commun., Vol. 16, pp. 67-88.

Koenig, L.L., 2000, "Laryngeal factors in voiceless consonant production in men, women, and 5-year-olds", J. Speech Lang. Hear. Res., Vol. 43, pp. 1211-1228.

Koizumi, T., Taniguchi, S., Hiromitsu, S., 1987, "Two-mass models of the vocal cords for natural sounding voice synthesis", J. Acoust. Soc. Am., Vol. 82, pp.1179-1192.

Lucero, J. C., 1999, "A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset", J. Acoust. Soc. Am., Vol. 105, pp. 423-431.

Shannon, C. E., 1948, "A mathematical theory of communication", Bell System Tech. J., Vol. 27, pp. 379-423 and pp. 623-659.

Soize, C., 2000, "A nonparametric model of random uncertainties for reduced matrix models in structural dynamics", Probabilistic Engineering Mechnics, Vol. 15, No. 3, pp. 277-294 pp. 623-652.

Steinecke, I., Herzel, H., 1995, "Bifurcation in an asymmetric vocal-fold model", J. Acoust. Soc. Am., Vol. 97, pp. 1874-1884.

Takemoto, H., Honda, K., 2006, "Measurement of temporal changes in vocal tract area function from 3D cine-MRI data", J. Acoust. Soc. Am., Vol. 119, pp. 1037-1049.

Titze, I.R., 1980, "Comments on the myoelastic-aerodynamic theory of phonation", J. Acoust. Soc. of Am., Vol. 23, pp.495-510.

Titze, I. R.,1994, "Principles of Voice Production", Ed. Prentice-Hall, NJ, Englewood Cliffs, NJ, 530 p.

Titze, I.R., Story, B. H., Hoffman, E. A., 1996, "Vocal tract area functions from magnetic resonance imaging", J. Acoust. Soc. of Am., Vol. 100, pp.537-554.

Van den Berg, J., 1968, "Myoelastic-aerodynamic theory of voice production", J. Speech Hear., Rs. 1, pp. 227-244.

Zhang, Y., Jiang, J., Rahn III, D. A., 2005, "Studying vocal fold vibrations in Parkinson's disease with a nonlinear model", Chaos, Vol. 15, No. 033903, pp. 1-10.

RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.