

Modeling random uncertainties in voice production using a parametric approach

E. Cataldo, R. Sampaio, J. Lucero, Christian Soize

► **To cite this version:**

E. Cataldo, R. Sampaio, J. Lucero, Christian Soize. Modeling random uncertainties in voice production using a parametric approach. *Mechanics Research Communications*, Elsevier, 2008, 35 (7), pp.454-459. 10.1016/j.mechrescom.2008.05.006 . hal-00684827

HAL Id: hal-00684827

<https://hal-upec-upem.archives-ouvertes.fr/hal-00684827>

Submitted on 3 Apr 2012

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.



Modeling random uncertainties in voice production using a parametric approach

Edson Cataldo ^{a,*}, Rubens Sampaio ^b, Jorge Lucero ^c, Christian Soize ^d

^a Departamento de Matemática Aplicada, Programa de Pós-graduação em Eng. de Telecomunicações, Universidade Federal Fluminense, Rua Mario Santos Braga, S/N, Centro, Cep: 24, 020-140 Niterói, Brazil

^b Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro, Brazil

^c Departamento de Matemática, Universidade de Brasília, Brazil

^d Université Paris-Est, Laboratoire Modélisation et Simulation Multi Echelle, MSME FRE 3160 CNRS, 5, Bd Descartes, 77454 Marne-la-Vallée, France

ARTICLE INFO

Article history:

Received 5 October 2007

Received in revised form 9 March 2008

Available online 10 June 2008

Keywords:

Voice production

Uncertainties

Parametric probabilistic approach

Biomechanical models

ABSTRACT

The biomechanical system of voice production has been usually modeled as deterministic. However, when uncertainties of the parameters are considered, voice production must be regarded as a stochastic process. We follow a parametric approach for stochastic modeling, which requires the adoption of random variables to represent the uncertain parameters, and argue that such an approach improves predictability of the model. For each random variable, a probability density function is constructed using the maximum entropy principle. From the output of the model, a probability density function of the voice fundamental frequency is constructed, for configurations corresponding to male and female larynges, and compared to experimental data.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

In past years several models of voice production (phonation) have been proposed to explain its underlying mechanisms and to synthesize voice by computer simulation (Cataldo et al., 2006). However, most models are deterministic, whereas phonation must be regarded as a stochastic process. This stochasticity and its applications have been discussed by many researchers. Titze (1994) stated that perturbation analysis should be conducted to determine the degree to which the fundamental frequency (F_0) deviates from a prescribed model. Further, he showed that the distribution of F_0 is nearly gaussian, suggesting that perturbations are primarily random. Pinto and Titze (1990) discussed distributional characteristics of perturbations, and argued that different types of vocal perturbations may have different distributions. Some discussions have been made to characterize probability distributions of vocal jitter, considering voice signals from normal subjects (e.g., Schoengen, 2001). Iwata and von Leden (1970) concluded that data from normal adult males yielded normal synchronous curves. Later, Stone et al. (1976) contested those findings claiming that some results indicated non gaussian distributions.

In this work, the voice production system is considered to be non-deterministic. A version of the popular two-mass model of the vocal folds (Ishizaka and Flanagan, 1972) is adopted. This model has been widely used (Zhang et al., 2005; Adachi and Yu, 2005) and the capability of this well-known model to reproduce the vocal folds vibrations has been successfully demonstrated. Some parameters of this model will be characterized as uncertain by using the maximum entropy principle (Kapur and Kesavan, 1992). This is a powerful technique which allows us to construct probability density functions based on scarce available information.

* Corresponding author. Tel.: +55 21 2629 5483; fax: +55 21 2629 5515.

E-mail address: ecataldo@im.uff.br (E. Cataldo).

¹ Rua Mário Santos Braga, S/N, Centro, Niterói, RJ, Brazil.

2. Mean model

For the so-called mean model we adopt a version of the two-mass model of the vocal folds proposed by Ishizaka and Flanagan (1972), shown in Fig. 1.

Each vocal fold is represented by two-mass systems. The two vocal folds are assumed identical and they move symmetrically with respect to the glottal midline, in the horizontal direction.

The dynamics of the system can be written, in a simplified form, by Eq. (1) and (2):

$$\psi_1(\mathbf{w})\dot{u}_g + \psi_2(\mathbf{w})|u_g|u_g + \psi_3(\mathbf{w})u_g + \frac{1}{\tilde{c}_1} \int_0^t (u_g(\tau) - u_1(\tau))d\tau - y = 0 \tag{1}$$

$$[M]\ddot{\mathbf{w}} + [C]\dot{\mathbf{w}} + [K]\mathbf{w} + \mathbf{h}(\mathbf{w}, \dot{\mathbf{w}}, u_g, \dot{u}_g) = 0 \tag{2}$$

where $\mathbf{w}(t) = (x_1(t), x_2(t), u_1(t), u_2(t), u_r(t))^t$, x_1 and x_2 are the displacements of the masses, u_1 and u_2 describe the air volume flow through the (two) tubes that model the vocal tract, u_r is the air volume flow through the mouth, y is the subglottal pressure is denoted by y , u_g is the glottal signal, and p_r is the radiated pressure at the output and it is given by $p_r(t) = u_r(t)r_r$, where $r_r = 128\rho v_c/9\pi^3 y_2^2$, ρ is the air density, v_c is the sound velocity, and y_2 is the radius of the second tube. And also $\tilde{c}_1 = \ell_1 \pi r_1^2 / \rho v_c^2$, where ℓ_1 is the length of the first tube, r_1 is the radius of the first tube, and μ is the shear viscosity coefficient. Parameters $\psi_1, \psi_2, \psi_3, \mathbf{h}$, as well as matrices $[M], [C], [K]$ are obtained from the model proposed by Ishizaka and Flanagan (1972), after algebraic manipulations.

The expressions of ψ_1, ψ_2, ψ_3 are given here because they contain the parameter a_{g0} in their equations:

$$\phi_1(\mathbf{w}) = \left(\frac{0.19\rho}{(a_{g0} + 2\ell_g x_1)} + 2\ell_g x_1 \right) + \frac{\rho}{(a_{g0} + 2\ell_g x_2)^2} \left[0.5 - \frac{a_{g0} + 2\ell_g x_2}{a_1} \left(1 - \frac{a_{g0} + 2\ell_g x_2}{a_1} \right) \right]$$

$$\phi_2(\mathbf{w}) = \left(12\mu\ell_g \frac{d_1}{(a_{g0} + 2\ell_g x_1)^3} + 12\ell_g^2 \frac{d_2}{(a_{g0} + 2\ell_g x_2)^3} + \tilde{r}_1 \right), \quad \phi_3(\mathbf{w}) = \left(\frac{\rho d_1}{a_{g0} + 2\ell_g x_1} + \frac{\rho d_2}{a_{g0} + 2\ell_g x_2} + \tilde{\ell}_1 \right)$$

where $\tilde{\ell}_1 = \frac{\rho \ell_1}{2\pi y_1^2}$, $\tilde{r}_1 = \frac{8\rho}{3\pi^2 y_n}$, $r_1 = \frac{2}{y_1} \sqrt{\rho\mu\omega}$, $\omega = \sqrt{\frac{k_1}{m_1}}$, $a_1 = \pi y_1^2$, $\tilde{c}_1 = \frac{\ell_1 \pi y_1^2}{\rho v_c^2}$, ℓ_1 is the length of the first tube, y_1 is the radius of the first tube, and μ is the shear viscosity coefficient. And also, ℓ_g is the length of the vocal fold and d_n is the vocal fold width, with $n = 1, 2$.

We can note that Eq. (2) describes the vibration problem in each of the two subsystems (vocal folds and vocal tract) and Eq. (1) is the equation that couples the two subsystems.

The tension parameter is used in order to control the fundamental frequency of the vocal folds. Then, the parameters m_1, k_1, m_2, k_2, k_c are written as $m_1 = \hat{m}_1/q, k_1 = q\hat{k}_1, m_2 = \hat{m}_2/q, k_2 = q\hat{k}_2, k_c = q\hat{k}_c$, in which $\hat{m}_1, \hat{k}_1, \hat{m}_2, \hat{k}_2, \hat{k}_c$ are fixed values.

And, the parameters m_1, k_1, m_2, k_2, k_c are written as $m_1 = \hat{m}_1/q, k_1 = q\hat{k}_1, m_2 = \hat{m}_2/q, k_2 = q\hat{k}_2, k_c = q\hat{k}_c$, in which $\hat{m}_1, \hat{k}_1, \hat{m}_2, \hat{k}_2, \hat{k}_c$ are included in the expressions of the matrices M, C and K .

3. Stochastic model

Three parameters, which are the main responsible for changes of the fundamental frequency, are considered as uncertain: tension parameter (q), subglottal pressure (y), and neutral glottal area (a_{g0}). The corresponding stochastic equations are obtained from Eqs. (1) and (2), by substituting the parameters by the corresponding random variables Q, Y , and A_{g0} , respectively.

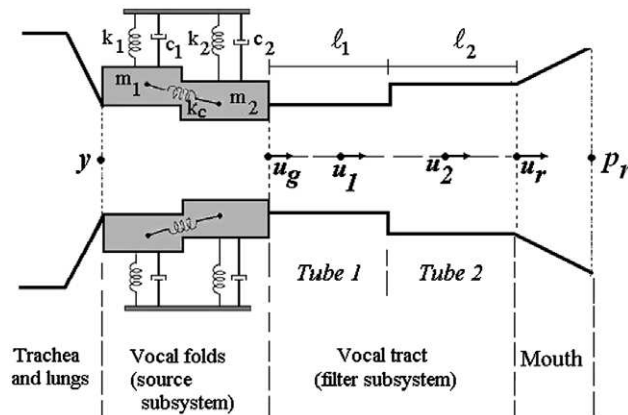


Fig. 1. Model used for describing the voice production process.

Instead of considering that perturbations occur cycle-to-cycle for a fixed set of parameters, an output signal is obtained for each realization of the three parameters and variations of fundamental frequency of each signal are analyzed. This is possible if voice production is assumed to be a stationary and ergodic process (see, e.g., Schoengten, 2001). To construct the density probability functions of the random variables Q , Y and A_{g0} we follow the Maximum Entropy Principle (Kapur and Kesavan, 1992; Jaynes, 1957a,b; Soize, 2001), which consists in maximizing the entropy subjected to constraints defined by the usable information.

The maximum entropy principle is invoked whenever we have only partial information about a system which can be expressed in the form of some averages. Uncertainty is the characteristic of such systems. We can note that Ishizaka and Flanagan (1972) do not give much information about the values of the parameters, principally about their dispersions.

Letting X be a continuum random variable, the Entropy $S(p_X)$ (Shannon (1948)) of its probability density function p_X is defined by Eq. (3).

$$S(p_X) = - \int_{\mathbb{R}} p_X(x) \ln(p_X(x)) dx. \tag{3}$$

The probability density function p_X to be constructed is the one with the largest uncertainty, measured by the entropy, among the sets of all of the probability density functions that verify the constraints defined by the usable information.

The tension parameter is modeled by the random variable Q . The usable information are: (i) its support is $]0, +\infty[$, (ii) its mean value is $E\{Q\} = \underline{Y}$, (iii) $E\{1/Q^2\} < +\infty$. Information (iii) is due to $M_1 = \hat{m}_1/Q$ is to be a second-order random variable. Then, it is necessary that $E\{M_1^2\} < +\infty$ yielding $E\{1/Q^2\} < +\infty$. Its probability density function is given by

$$p_Q(q) = \mathbf{1}_{]0,+\infty[}(q) \frac{1}{\underline{Q}} \left(\frac{1}{\delta_Q^2}\right)^{\frac{1}{\delta_Q}} \frac{1}{\Gamma(1/\delta_Q^2)} \left(\frac{q}{\underline{Q}}\right)^{\frac{1}{\delta_Q}-1} \exp\left(-\frac{q}{\delta_Q^2 \underline{Q}}\right), \tag{4}$$

where the positive parameter $\delta_Q = \sigma_Q/\underline{Q}$ is the dispersion coefficient, satisfying $\delta_Q < 1/\sqrt{2}$, and σ_Q is the standard deviation of Q .

The subglottal pressure is modeled by the random variable Y . The usable information are: (i) its support is $]0, +\infty[$, (ii) its mean value is $E\{Y\} = \underline{Y}$, (iii) The second-order moment of its inverse is finite $E\{1/Y^2\} < +\infty$. Information (iii) is used because 0, and values near to it, should be repulsive values for Y , since there is a minimum pressure necessary to produce phonation. The probability density is

$$p_Y(y) = \mathbf{1}_{]0,+\infty[}(y) \frac{1}{\underline{Y}} \left(\frac{1}{\delta_Y^2}\right)^{\frac{1}{\delta_Y}} \frac{1}{\Gamma(1/\delta_Y^2)} \left(\frac{y}{\underline{Y}}\right)^{\frac{1}{\delta_Y}-1} \exp\left(-\frac{y}{\delta_Y^2 \underline{Y}}\right), \tag{5}$$

where the positive parameter $\delta_Y = \sigma_Y/\underline{Y}$ is the dispersion coefficient, satisfying $\delta_Y < 1/\sqrt{2}$, and σ_Y is the standard deviation of Y .

The usable information for constructing the probability density function of A_{g0} are: (i) its support is $]0, +\infty[$, (ii) its mean value is $E\{A_{g0}\} = \underline{A}_{g0}$, and (iii) it is a second-order random variable; it means $E\{A_{g0}^2\} < +\infty$. The probability density function of A_{g0} is

$$p_{A_{g0}}(a_{g0}) = \mathbf{1}_{]0,+\infty[} e^{-\lambda_0 - \lambda_1 a_{g0} - \lambda_2 a_{g0}^2} \tag{6}$$

where λ_0, λ_1 and λ_2 are the values that minimize the function Δ , given by Eq. (7)

$$\Delta = \lambda_0 + \lambda_1 \underline{A}_{g0} + \lambda_2 c \tag{7}$$

with $E\{X^2\} = c, c < +\infty$. A coefficient of dispersion $\delta_{A_{g0}} = \frac{\sigma_{A_{g0}}}{\underline{A}_{g0}}$ is created, where $\sigma_{A_{g0}}$ is the standard deviation of A_{g0} . It can be proved that $c = \underline{A}_{g0}^2 (1 + \delta_{A_{g0}}^2)$.

4. Numerical examples

A stochastic solver based on a Monte Carlo numerical simulation was used to compute the radiated pressure. For each independent realization $A_{g0}(\theta), Y(\theta)$, and $Q(\theta)$, a realization of the random fundamental frequency $F_0(\theta)$ is calculated in the same way as described for the mean model. A convergence analysis with respect to n was carried out studying the convergence of the estimated second-order moment of F_0 . This convergence analysis was performed for different values of $\delta_Y, \delta_Q, \delta_{A_{g0}}$, for $n \geq 500$, convergence was always reached. Then, $n = 500$ was taken for all further estimations. The values used for simulations to reproduce the signals of voice produced by men and women are based on those ones discussed by Lucero and Koenig (2005). The density probability functions for 500 realizations of the random variables Q is shown in Fig. 2 shows, Y and A_{g0} , and coefficients of dispersion $\delta_{A_{g0}} = \delta_Q = \delta_Y = 0.05$.

The results are in good agreement with previous experimental measures, (e.g., Titze, 1994). Here, the comparison is simply qualitative to show that the shapes of the probability density functions are in agree with those ones showed, experimentally, by Titze. And also, we could see that the fundamental frequencies obtained for women are greater than those ones obtained for men.

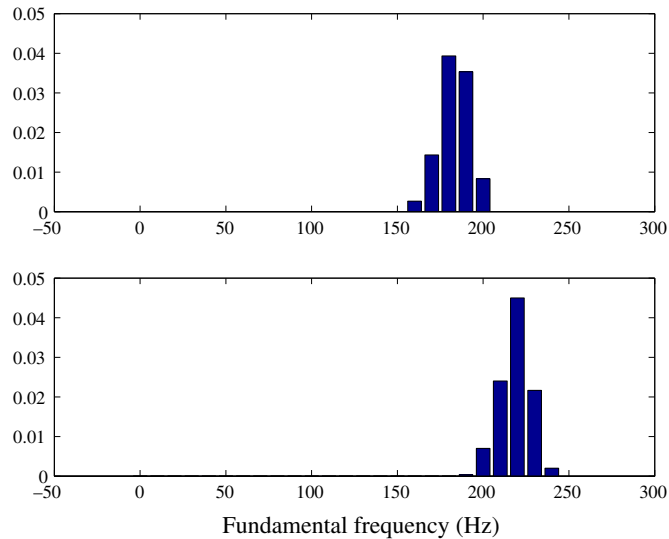


Fig. 2. Probability density functions of the fundamental frequency for men (top) and women (bottom).

In the following, we will try to validate the development presented here, with solving an inverse problem and comparing the results obtained with experimental data. Voice signals produced by one person have been analyzed and their statistics have been compared with simulations. Two examples based on this idea are showed.

In the first example, seven hundred samples of voice signals of a sustained vowel/a/ were recorded from one and each sample had 0.3s of duration. From the samples, the fundamental frequency was estimated and a probability density function was constructed and it is shown in Fig. 3 (top), in which the mean value of the fundamental frequency is $\hat{m}_{F_0}^{\text{exp}} = 113$ Hz and the coefficient of dispersion $\hat{\delta}_{F_0}^{\text{exp}} = \hat{m}_{F_0}^{\text{exp}} / \hat{\sigma}_{F_0}^{\text{exp}} = 0.06$. Next, the inverse problem was solved, using the measured data as targets. A simple trial-and-error method yielded $\underline{A}_{g0} = 5 \times 10^{-2} \text{m}^2$, $\underline{Y} = 6600 \text{Pa}$, $\underline{Q} = 0.63$, $\delta_Y = 0.13$ and $\delta_Q = 0.05$. The mean value of the fundamental frequency obtained by numerical simulation was $\hat{m}_{F_0} = 112.8 \text{Hz}$, the dispersion coefficient was $\hat{\delta}_{F_0} = \hat{m}_{F_0} / \hat{\sigma}_{F_0} = 0.05$ and the probability density function obtained is shown in Fig. 3, bottom. As shown, the theoretical results are in good agreement with the experimental data, which indicates the possibility of characterizing a set of voice signals by solving an inverse stochastic problem.

A second example is then performed. A voice signal corresponding to a sustained vowel/a/ has been recorded from one person and 675 frames were obtained from this signal, each one with 0.01s of length and the corresponding fundamental frequency was calculated for each frame. From the samples, the fundamental frequency was estimated and a probability density function was constructed and it is shown in Fig. 4 (top). In this case, the mean value of the fundamental frequency is $\hat{m}_{F_0}^{\text{exp}} = 120.9525 \text{Hz}$ and the coefficient of dispersion $\hat{\delta}_{F_0}^{\text{exp}} = \hat{m}_{F_0}^{\text{exp}} / \hat{\sigma}_{F_0}^{\text{exp}} = 0.0171$. Next, the inverse problem was solved, using the measured data as targets. A simple trial-and-error method yielded $\underline{A}_{g0} = 5 \times 10^{-2} \text{m}^2$, $\underline{Y} = 750 \text{Pa}$, $\underline{Q} = 0.66$, $\delta_{A_{g0}} =$

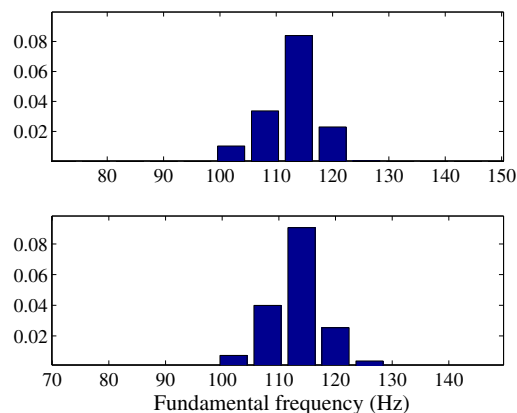


Fig. 3. Experimental probability density function of the fundamental frequency related to experimental voice signals produced by one person (top) compared with the probability density function estimated with the stochastic model (bottom) – first example.

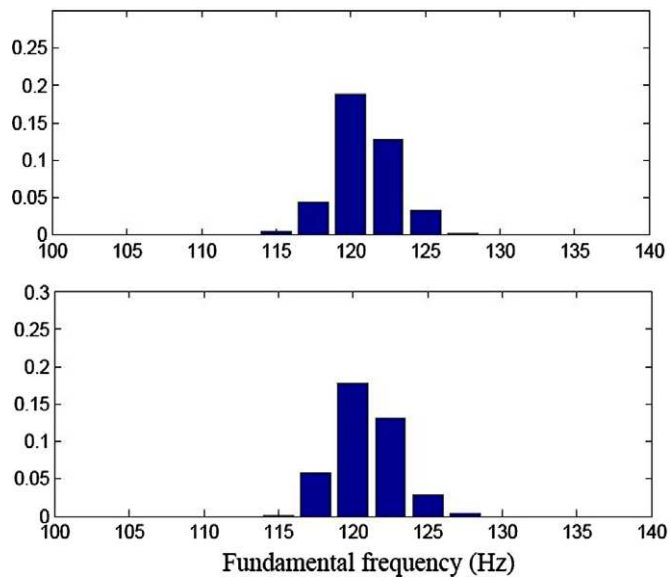


Fig. 4. Experimental probability density function of the fundamental frequency related to experimental voice signals produced by one person (top) compared with the probability density function estimated with the stochastic model (bottom) – second example.

0.03, $\delta_V = 0.01$, $\delta_Q = 0.01$. The mean value of the fundamental frequency obtained by numerical simulation was $\hat{m}_{F_0} = 120.7694\text{Hz}$ and the dispersion coefficient was $\hat{\delta}_{F_0} = \hat{m}_{F_0}/\hat{\sigma}_{F_0} = 0.0173$. The probability density function obtained is shown in Fig. 4, bottom. Again, one can note that the theoretical results are in good agreement with the experimental data.

5. Conclusions

A parametric probabilistic approach, based on the Maximum Entropy Principle, was proposed to take into account uncertainties in a biomechanical voice production model. This technique is especially powerful when available experimental data sets are not sufficiently large, as in the case of phonation.

It may be questioned if a gaussian distribution could have been used to the random variables associated to the uncertain parameters. Some disadvantages of such distributions are: (1) gaussian random variables vary from minus infinity to plus infinity, whereas the biomechanical parameters of the model (pressure, vocal fold tension, and glottal area) can not be negative; (2) let f be the force on the spring, X its displacement and K its stiffness, so that $X = f/K$. Assume the force is deterministic, and that both the stiffness and the displacement are random. It is natural to consider the expected value of X finite; that is, $E\{X^2\} < +\infty \Rightarrow E\{f^2/K^2\} < +\infty \Rightarrow E\{1/K^2\} < +\infty$, which is not true for a gaussian random variable.

Experimental data were compared with those ones obtained numerically and the results showed that this model is useful to understand a little more about the phonation phenomenon.

Acknowledgement

This work was supported by the Brazilian Agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), by the International Cooperation Project Capes-Cofecub, No. 476/04, and by Faperj.

References

- Adachi, S., Yu, J., 2005. Two-dimensional model of vocal fold vibration for sound synthesis of voice and soprano singing. *J. Acoust. Soc. Am.* 117, 3213–3224.
- 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. *Mech. Res. Commun.* 33, 250–260.
- Ishizaka, K., Flanagan, J.L., 1972. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell Syst. Tech. J.* 51, 1233–1268.
- Iwata, S., von Leden, H., 1970. Pitch perturbations in normal and pathologic voices. *Folia Phoniatr.* 22, 413–424.
- Jaynes, E., 1957a. Information theory and statistical mechanics. *Phys. Rev.* 106 (4), 620–630.
- Jaynes, E., 1957b. Information theory and statistical mechanics. *Phys. Rev.* 108, 171–190.
- Kapur, J.N., Kesavan, H.K., 1992. *Entropy Optimization Principles with Applications*. Academic Press, San Diego.
- Lucero, J.C., Koenig, L.L., 2005. Simulations of temporal patterns of oral airflow in men and women using a two-mass model of the vocal folds under dynamic control. *J. Acoust. Soc. Am.* 117, 1–11.
- Pinto, N., Titze, I., 1990. Unification of perturbation measures in speech signals. *J. Acoust. Soc. Am.* 87 (3), 1278–1289.
- Schoengten, J., 2001. Stochastic models of jitter. *J. Acoust. Soc. Am.* 109, 1631–1650.
- Shannon, C.E., 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423. 623–659.

- Soize, C., 2001. Maximum entropy approach for modeling random uncertainties in transient elastodynamics. *J. Acoust. Soc. Am.* 109, 1979–1996.
- Titze, I.R., 1994. *Principles of Voice Production*. Prentice-Hall, NJ, Englewood Cliffs, NJ.
- Stone, R.E., Deller, J.R., Anderson, D.J., 1976. Distributional characteristics of jitter in sustained phonation. In: Presented at a conference of American Speech and Hearing Association, Texas.
- Zhang, Y., Jiang, J., Rahn III, D.A., 2005. Studying vocal fold vibrations in Parkinson's disease with a nonlinear model. *Chaos* 159, 1–10. N. 033903.