Biomechanical modeling of laryngeal dynamics using inverse filtering of speech signals

Alan P. Pinheiro*, Carlos D. Maciel, José C. Pereira, Marcelo B. Joaquim, Lianet Selpúlveda-Torres
Department of Electrical Engineering
University of São Paulo
São Carlos, Brazil
*alanpp@sc.usp.br

Abstract— Voice production occurs within the larynx where the vocal folds produce the glottal flow signal through vibrational movements that generate the primary voice signal. In this context, this paper introduces a method based on inverse filtering of speech signals and a biomechanical model of vocal folds to simulate the production of this primary signal. This method allows the formulation of a deterministic model of the vocal folds that reproduces their dynamic behavior and can be used to evaluate the functional state of the subject's voice using only acoustic speech signals. The results indicate that the proposed method can simulate the main components of the glottal flow signal.

Keywords-glottal flow; vocal folds; inverse filtering.

I. INTRODUCTION

The voice production process comprehends a set of biological systems that involves the lungs, larynx, pharynx, trachea, and nasal and oral cavities. This process starts in the larynx, where the vocal folds produce the excitation signal of vocal tract through vibratory motions that modulate the air (from the trachea) generating the glottal flow. This signal will define the fundamental frequency of voice and many of its basic features. Due to the intense activity in the larynx, many voice pathologies originate in this structure [1] and may result in a change in the behavior of its dynamics, producing a perturbation on the voice.

In this context, a detailed study of voice must include an evaluation of the properties of the glottal flow signal. For this purpose, some methods have been developed to evaluate the larynx activity and/or the waveform of the glottal flow generated, among which videolaryngoscopy [2], electroglottography [3] and inverse filtering of speech signals [4] can be emphasized.

Videolaryngoscopy is considered the gold standard method [5], as it is capable of directly evaluating the functional state of vocal folds using visual recordings of their movements. However, it uses an endoscopy linked to a video camera, which is inserted in the subject's mouth (or nasal cavity) interfering with the natural voice production during the exam. Moreover, its cost may be impracticable in many applications. Electroglottography uses a pair of noninvasive electrodes fixed on the neck skin, next to the larynx, sensitive to the vibrational activity present in this region. Its drawback is associated with the fact that the tissues where the electrodes are fixed attenuate some signal components generating a distortion in the waveform of the glottal flow recorded by the device. Lastly, the inverse filtering method uses acoustical speech signals recorded by a microphone to reconstruct the waveform of the glottal flow through an algorithm which can compensate the effects of vocal tract on the signal analyzed.

However, a single analysis of the glottal flow signal would not be enough to describe the complex mechanisms of voice production in the larynx. For this reason, many authors [6-7] have developed biomechanical models of vocal folds to reproduce their dynamics through their vibrational behavior. In this sense, the use of inverse filtering as a method to estimate the glottal flow waveform is a useful way to produce information regarding a determined subject. This information can be used for a biomechanical model to reconstruct the vocal folds dynamics of this subject and simulate their glottal flow signal. This approach needs few hardware resources and can generate valuable information concerning the evaluation of subject voice in clinical practice.

The aim of this paper is to describe a technique that allows the evaluation of the vocal folds dynamics of a subject using a biomechanical model. The glottal flow produced by the subject is extracted from his/her speech signal using an inverse filtering method. This extracted signal is used for an optimization procedure able to change the model parameters until it could reproduce, with a good precision, the waveform of the extracted glottal flow. As a contribution, this method allows the simulation of the vocal folds behavior of a specific subject, evaluation of the functional state of voice and its dynamics using only acoustic speech signals.

II. METHODOLOGY

The first step in this research to simulate the behavior of vocal folds was the extraction of the glottal flow waveform using a speech signal recorded by a microphone. An inverse filtering method is applied to the signal of four subjects (with healthy voices) that vocalize the sustained vowel "a". Next, an optimization

procedure combining genetic algorithm and simplex method is used to find the best values for the parameters of the biomechanical model employed, such that the model can accurately reproduce the glottal flow waveform extracted from the subject voice signal. The methods employed in this research are described in the following sections.

A. *Estimation of the glottal flow waveform*

The extraction of the glottal flow waveform is performed by the Iterative Adaptive Inverse Filtering (IAIF) [8]. Basically, this method uses the Linear Predictive Coding (LPC) to estimate the vocal tract transfer function and compensate the vocal tract effects on the analyzed signal using an inverse filtering process. The IAIF adopts a voice production model based on three interdependent processes: production of glottal excitation, (2) vocal tract equalization and (3) lip radiation effect.

Following the IAIF method, to estimate the glottal waveform it is necessary to evaluate the contributions of the vocal tract and lips radiation in the signal. Considering that the lips radiation effects can be modeled accurately enough with a fixed differentiator, the problem concentrates on the estimation of the vocal tract transfer function. The glottal waveform is then estimated by an inverse filtering which will cancel the vocal tract effects and posteriorly the lips radiation in the speech signal. Figure 1 illustrates the steps involved in the IAIF method.

In the first step of the IAIF method, the speech signal is high-pass filtered with a second-order Butterworth filter with cut-off frequency of 30 Hz. If the low frequency components of the speech are not removed, the resulting glottal wave starts to fluctuate. In the second step a preliminary estimate of excitation effects on the signal is computed by LPC analysis. The glottal contribution is eliminated from the signal through an inverse filtering process (Step 3).

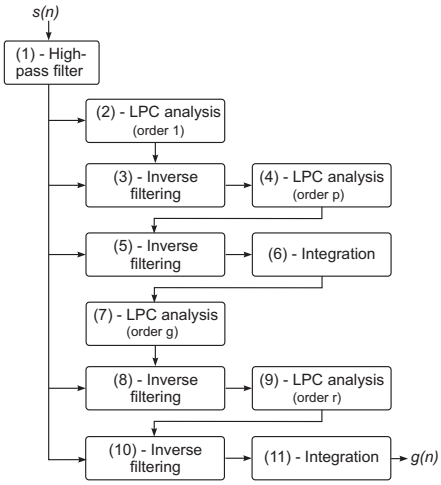

Figure 1. Block diagram of the IAIF method used to estimate the glottal flow waveform. The input s(n) is a speech signal recorded by a microphone and the output g(n) is a time series corresponding to a glottal flow which produced the signal s(n).

In Step 4 an LPC filter is again applied resulting in a vocal tract model used in Step 5 to cancel the effects of this vocal tract on the analyzed speech signal. At the end, the lips radiation is also canceled by the numerical integration in Step 6. The result of this first iteration is a preliminary estimate of the glottal flow wave in the Step 6 output. To increase the reliability of this estimate, the IAIF uses a second iteration that comprehends Steps 7 to 11 in the diagram of Figure 1.

At the beginning of the second iteration (Step 7), a new estimate of the glottal excitation is performed using an LPC analysis. Unlike the first iteration, the input of this step is the preliminary estimate of the glottal flow evaluated by former steps, and not by the speech signal, as in Step 2. This approach allows the glottal contribution to be evaluated more accurately when compared with the start of the first iteration. In Step 8 this glottal contribution is suppressed by inverse filtering and a new LPC analysis (present in Step 9) is applied to the signal whose glottal contribution had been suppressed, constituting a final model for the vocal tract. The glottal flow waveform is finally obtained by the inverse filtering in Steps 10 (which had suppressed the vocal tract contributions from the original speech signal) and 11, canceling the lips radiation.

During the analysis, signals with about 25 ms were used with a rectangular window. The LPC order described in Figure 1 diagram was p = r = 8 and g = 6. In agreement with Alku *et al* [8], these values are enough for this type of analysis.

B. *Biomechanical modeling of vocal folds*

The vocal folds consist of a set of tissue layers able to vibrate due to the aerodynamic interaction that this system has with the airflow from the trachea. Based on its myoelastic and aerodynamic properties, Ishizaka and Flanagan [6] developed a biomechanical model of the vocal folds which simulates their main vibrational motions and reproduce their dynamics. This model, termed as two-mass model, assumes one vocal fold to be represented by a pair of two coupled oscillators using two masses, three springs and two dampers, as in Figure 2.

The two masses of the model were used to reproduce the mucosal wave phase difference while the other lumped elements (springs and dampers) represent

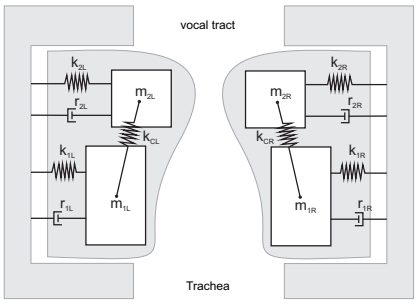

Figure 2. Biomechanical model of the vocal folds elaborated by Ishizaka and Flanagan [6].

a stiff layer and its viscoelastic properties. In this model the masses are vibrated by aerodynamic forces caused by the interaction between the subglottal pressure and

vocal fold tissues and can be described by Bernoulli laws. Steinecke and Herzel [7] proposed an adaptation for this classical model in which some assumptions were adopted to simplify the model and decrease the computational cost, keeping the most important features of vocal fold dynamics and its physiological base. The model equations are described in (1).

$$\begin{aligned} m_{1\alpha}\ddot{x}_{1\alpha} &= -r_{1\alpha}\dot{x}_{1\alpha} - k_{1\alpha}x_{1\alpha} - k_{c\alpha}(x_{1\alpha} - x_{2\alpha}) + F^{I}_{1\alpha}(x_{1\alpha}) \\ &\quad + F^{B}(P_s, L, d, x_{1\alpha}, x_{2\alpha}) \\ m_{2\alpha}\ddot{x}_{2\alpha} &= -r_{2\alpha}\dot{x}_{1\alpha} - k_{2\alpha}x_{2\alpha} - k_{c\alpha}(x_{2\alpha} - x_{1\alpha}) + F^{I}_{2\alpha}(x_{2\alpha}) \end{aligned} \quad (1)$$

The $x_{i\alpha}$ variable corresponds to oscillation amplitudes of the masses. Index i represents the lower (i=1) and upper (i=2) portion of vocal folds while index α represents the left (α=L) and right (α=R) portion. The main parameters of the model are denoted by $m_{i\alpha}$ (mass), $k_{i\alpha}$ (stiffness coefficient), $k_{c\alpha}$ (coupling spring constant), $r_{i\alpha}$ (damping coefficients) and P_s (subglottal pressure) and can be represented by a parameter vector p such as $p := [m_{i\alpha}, k_{i\alpha}, k_{c\alpha}, r_{i\alpha}, P_s]$. The impact forces due to vocal fold collisions and Bernoulli forces were represented as F^{I}_{α} and F^{B}, respectively, and are described in details in [7]. The standard parameters of this model, defined by the authors, are: $m_{1\alpha}$=0.125, $m_{2\alpha}$=0.025, $k_{1\alpha}$=0.08, $k_{2\alpha}$=0.008, $k_{c\alpha}$=0.025, $r_{1\alpha}$=$r_{2\alpha}$=0.02 and P_s=0.008. All units are given in centimeters, grams, milliseconds and their corresponding combinations. In this research, the equations showed in (1) were solved using the standard fourth-order Runge-Kutta method. All simulations were processed using Matlab (Mathworks, USA).

C. Optimization of biomechanical model parameters

By using the parameters of the two-mass model represented above by p, it is possible to reproduce the glottal signal. As each subject has their own voice characteristics - which reflects their vocal fold dynamics - it is important to find the value of p that best reproduces the glottal flow extracted by IAIF method.

For this purpose, an optimization procedure combining genetic algorithms and the simplex method was used. This method compares the signal extracted by IAIF method with the one simulated by the two-mass model. As the model equations are nonsmooth and might have a non-convex search space, in the first step a genetic algorithm is applied to search for a rough approximation in order to avoid inappropriate local minima. In the second step, this approximate solution serves as the starting point for a simplex method refining the approximate solution. This approach will help to reduce the optimization time and avoid the problem of convergence to local minima.

The genetic algorithm used in this research utilized the roulette wheel selection rule and elitism techniques [9] to accelerate the convergence. A population of five hundred individuals was adopted and the algorithm was programmed to process only 50 generations to avoid excessive processing time. The search space comprised the [0.0 to 0.4] interval for the first seven mechanical parameters of vector p while the eighth parameter should be contained in the [0.008 to 0.040] interval. The objective function used for both methods was defined as

$$\Psi(p) = \sum_{k=1}^{m} (x_E[k] - x_S[k])^2. \quad (2)$$

Were, m indicates the sample number of the speech signal. The value of Ψ represents the approximation error between the signal simulated by the model (x_S) and the glottal flow signal extracted from the speech signal (x_E). The minimum value of Ψ gives the model parameters most able to reproduce the vibrational movements of the folds.

III. RESULTS

Figure 3 illustrates some results involving the glottal signal extracted from the speech signal of a subject and the simulation performed by the biomechanical model.

The described method could simulate the signal produced by glottal vibrations of a specific subject with relative precision. The difference between the extracted and simulated signals (Figure 3b) might be attributed to the model behavior, which cannot simulate certain patterns, and/or the optimization procedure, which does not produce an optimum solution. Stochastic components in the extracted signal may have increased

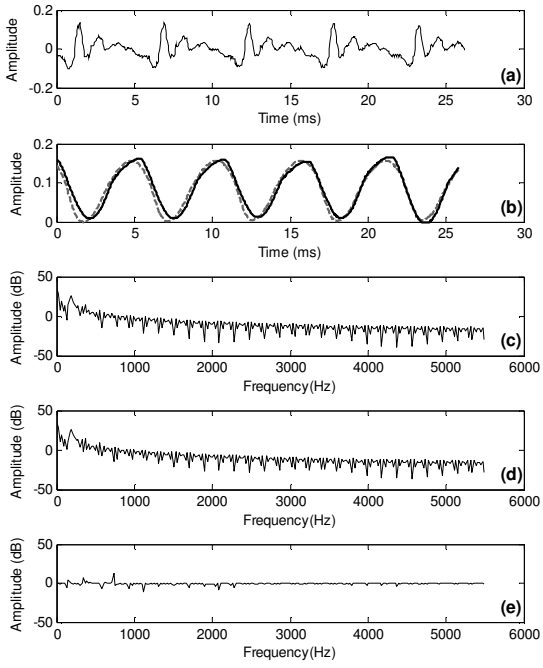

Figure 3. (a) Windowed speech signal of vowel /a/ pronounced by subject 1. (b) Glottal flow signal extracted by inverse filtering method (dotted line) and its corresponding simulated signal (continuous line) generated by the biomechanical model. (c) Fourier spectrum of the extracted signal and (d) simulated signal. (e) Spectrum error between the simulated and extracted signals.

the error value. The graphics of Figures 3c and 3d show that the simulated signal can reproduce the most

important components of the extracted signal. Table 1 shows the mechanical parameters found by the optimization procedure for the four subjects analyzed.

TABLE I. PARAMETERS OF THE BIOMECHANICAL MODEL CALCULATED FOR THE FOUR SUBJECTS

Parameters	Subjects			
	1	*2*	*3*	*4*
m_1, m_2	0.1900, 0.0819	0.1668, 0.0700	0.1894, 0.1065	0.1392, 0.0569
k_1, k_2	0.1234, 0.0491	0.0922, 0.0474	0.1666, 0.0046	0.1511, 0.0202
r_1, r_2	0.0611, 0.0223	0.1960, 0.0087	0.0099, 0.0282	0.0097, 0.0245
k_c, P_s	0.1244, 0.0237	0.0824, 0.0232	0.1994, 0.0319	0.1918, 0.0222
Error[a]	0.29	0.49	0.32	0.58

a. Error between the extracted and simulated signals calculated by equation (2).

IV. DISCUSSION

This paper has described a method which allows simulating the vocal folds dynamics of a subject using his/her speech signal and a standard biomechanical model. Emphasis was given to the extracted and simulated glottal flow signals used to describe the dynamical behavior of the larynx during the phonation.

As an example, the glottal flow waveform of a healthy subject (illustrated in Figure 1) was estimated by the IAIF method and later used to optimize the parameters of the biomechanical model. The optimization procedure showed a good performance and the curves produced by the model could reflect the most important vibration patterns present in the signal extracted from this subject. The error signal between the spectra confirms this result. Similar results were achieved in the analysis of the other subjects.

The mechanical parameters exhibited in Table 1 were significantly different from those defined as standard to the model. This is a strong evidence that (1) the proposed method can explore a large search space, and (2) the model parameters are sensitive to the vibration pattern showed for each subject. Furthermore, the simulated signals do not take into account merely the amplitude, fundamental frequency or phase, but also the peculiar features of the extracted waveform.

Albeit the biomechanical model has been successfully used to elucidate the vocal folds dynamics in scientific researches [2, 3, 10], the physiologic relevance of these parameters, found here to four subjects, is limited. Nevertheless, these parameters had an association with some physiologic properties of fold tissues related to elasticity, Young module, viscosity, tension, etc [11].

The importance of the proposed method is related to the clinical evaluation of voice and vocal folds [11], pathologies assessment [10], voice synthesis [12] and evaluation of voice dynamics [2, 3]. In this sense, some authors [13] have demonstrated that dynamic analysis methods can be used as very useful tools to quantify voice properties and have equal or superior performance than classic methods, such as jitter and shimmer. The limitations of the this method here proposed are related to the optimization procedure, which does not ensure a global optimum, and the IAIF drawbacks, which are very sensitive to noise in the signal input [8].

The development of a deterministic model of vocal folds has brought new perspectives to the use of emergent techniques in dynamic systems that can be applied to voice research. In particular, the nonlinear normal mode [14] theory has been used to study the structural nonlinearities in dynamical systems, detection and identification of complex behaviors in systems like those used here. Future works involving these methods will contribute to the characterization of the dynamic behavior of voice production in the larynx.

ACKNOWLEDGMENT

The authors would like to acknowledge FAPESP (grant 09/516.982) and CNPq (grant 143453/2008-4) for the financial support given to this project.

REFERENCES

[1] G. Fant, “Some problems in voice source analysis”, Speech Commun, vol. 13, pp. 7-22, 1993.

[2] P. Mergell, H. Herzel and I. R. Titze, “Irregular vocal-fold vibration—High-speed observation and modeling”, J. Acoust. Soc. Am., vol. 108, pp. 2996-3002, 2000.

[3] X. Qin, S. Wang and M. Wan, “Improving Reliability and Accuracy of Vibration Parameters of Vocal Folds Based on High-Speed Video and Electroglottography”, IEEE Trans Biomed Eng., vol. 56, pp. 1744-1754, 2009.

[4] E. Moore and J. Torres, “A performance assessment of objective measures for evaluating the quality of glottal waveform estimates”, Speech Commun, vol. 50, pp. 56-66, 2008.

[5] Y. Yan, E. Damrose and D. Bless, “Functional Analysis of Voice Using Simultaneous High-Speed Imaging and Acoustic Recordings”, J. of Voice, in press.

[6] K. Ishizaka and J.L. Flanagan, “Synthesis of voiced sounds from a two-mass model of the vocal cords”, Bell Syst. Tech. J., vol. 51, pp. 1233–1268, 1972.

[7] I. Steinecke and H. Herzel, “Bifurcations in an asymmetric vocal fold model”, J. Acoust. Soc. Amer., vol. 97, pp. 1874–1884, 1995.

[8] P. Alku, E. Vilkman and A.-M. Laukkanen, “Estimation of amplitude features of the glottal flow by inverse filtering speech pressure signals”, Speech Commun, vol. 24, pp. 123-132, 1998.

[9] D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, New York: Addison-Wesley, 1989.

[10] M. Döllinger et. al., “Vibration Parameter Extraction From Endoscopic Image Series of the Vocal Folds”, IEEE Trans Biomed Eng., vol. 42, pp. 773-781, 2002.

[11] C. Tao, Y. Zhang and J. Jiang, “Extracting Physiologically Relevant Parameters of Vocal Folds From High-Speed Video Image Series”, IEEE Trans Biomed Eng., vol. 54, pp. 794-801, 2007.

[12] B. H. Story, “An overview of the physiology, physics and modeling of the sound source for vowels”, Acoust. Sci. & Tech., vol. 23, pp. 195-206, 2002.

[13] J. J. Jiang, Y. Zhang and C. McGilligan, “Chaos in Voice, From Modeling to Measurement”, J. of Voice, vol. 20, pp. 2-17, 2006.

[14] A. F. Vakakis, “Non-linear normal modes and their applications in vibration theory: an overview”, Mech. Syst. Signal Process., vol. 11, pp. 3-22, 1997.