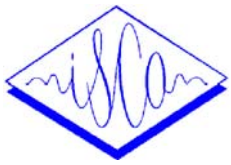
ISCA Archive
http://www.isca-speech.org/archive



ITRW on Nonlinear Speech
Processing (NOLISP 05)
Barcelona, Spain
April 19-22, 2005

# Estimation of vocal cord biomechanical parameters by non-linear inverse filtering of voice

Pedro Gómez, Rafael Martínez, Francisco Díaz, Carlos Lázaro, Agustín Álvarez, Victoria Rodellar, Víctor Nieto

Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n
28660 Boadilla del Monte, Madrid, Spain
pedro@pino.datsi.fi.upm.es

**Abstract.** Voice pathologies are a growing social concern, due to the role that voice and speech play in the performance of certain professions, and in the general population quality of life. In these last years emphasis has been placed in the early detection of these pathologies, for which computer tools performing voice processing are used to evaluate certain time and spectrum domain parameters to infer the presence of pathology. Going one step ahead the present work is aimed to estimate the values of the biomechanical parameters of the vocal fold system, as mass, stiffness and losses by the inversion of the vocal fold structure, which could help in classifying the specific patient's pathology. The model structure of the vocal cord will be presented, and a method to estimate the biomechanical parameters of the cord body structure will be described. Results for normal and pathological cases will be presented and discussed.

## 1 Introduction

Voice pathologies are a field of important research area in voice and speech processing regarding the quality of life of the population, especially for those persons who make voice a means of their professional activity, as speakers, singers, actors, lawyers, broadcasters, priests, teachers, call center workers, etc [8][11][17]. The success in treating voice pathologies depend on their early detection, and as such simple yet powerful inspection procedures are desirable. Among those procedures patient's voice inspection is a simple, low cost and fast method to obtain an estimation of the presence of pathology, which can be used as a screening routine to decide in using specialized inspection methods –as videoendoscopy- which can be more precise in pathology classification, but at the same time are less comfortable, more expensive and complicate. The key technique to use voice as a source for screening information, is to obtain a set of biomechanical parameters associated to the structure of the phonation source, which is the vocal fold system. Up to date techniques use time- and frequency-domain estimation of what are called *perturbation parameters*, which measure the deviation of the specific patient's voice from certain normal standards [4][15][10][5]. These techniques have revealed efficient themselves in the detection

of pathology, but supply little information on the nature of the pathology. Trying to go one step ahead our research group initiated a study to determine the possibility of estimating the values of the biomechanical parameters of the vocal fold system, as mass, stiffness and losses from the glottal source obtained from the residual after removing the vocal tract transfer function. This procedure is well documented in the literature (see for example [2]), and produces a trace which may be shown to be directly related with the glottal aperture (the average light measured between vocal cords during the phonatory cycle) [6]. The use of *k-mass* vocal fold models [13][1] help in determining that there are two main components in the movement of the vocal cord, contributed by the precise structure of the cord: the movement of the bulk muscular tissue of the cord body (see **Fig. 1.a**) and a traveling wave known as the *mucosal wave* [16][12], which is contributed by the epithelial tissue of the cord cover (see **Fig. 1.b**).

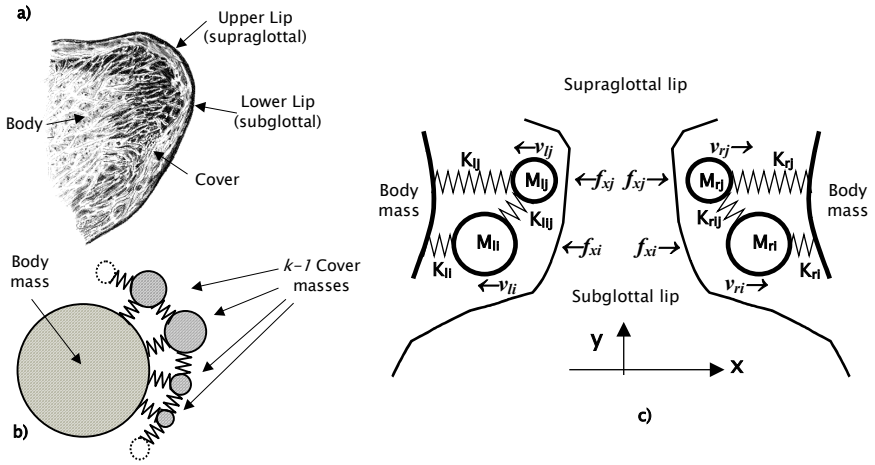



**Fig. 1.a**) Cross-section of the left vocal cord showing the body and cover structures (taken from [14]). **b**) *k-mass* model of the body and cover. **c**) *3-mass* model used to establish the dynamics of the body-cover system

In previous research [6][7] it has been shown that both contributions can be separated from the glottal aperture to produce two traces, known as the *average glottal aperture (AGA)* and the *mucosal wave correlate (MWC)*. Their relative energy ratio may be used as a clue for the presence of certain pathologies which induce the diminishing or complete disappearance of the *mucosal wave* [9]. In the present study the emphasis will be placed in using the *unfolded glottal aperture (UGA)* to measure the main biomechanical parameters involved in the dynamics of the cord body. For such the model structure of the vocal cord will be presented, and a method to estimate the biomechanical parameters of the cord body structure will be described. This method is based on the hypothesis stating that the fingerprint of the cord body dynamics is responsible for the *power spectral density (psd)* of the *UGA*, thus allowing the identification of the biomechanical parameters of the cord body from the theoretical dynamical transfer function between forces and speeds in the cord body model. For such a first estimation of the biomechanical parameters is obtained, which will be later refined by an adaptation process.

## 2 Estimating the cord movement

The vocal cords are two folds which can be found in the phonatory system located in the larynx supported by a complex structure of cartilage and muscles. These folds can be brought to a close contact to stop the flow of air through the respiratory system, and under convenient lung pressure can produce a vibration which is the basics of the phonation. A good explanation of the phonatory function can be found in [15]. A cross section of a vocal cord can be seen in **Fig. 1.a**, showing its tissular structure, which will be composed by the body and the cover as mentioned before. In **Fig. 1.b** an equivalent *k-mass* model is presented, where the main structure (the body) has been represented by a large lump which will be referred to as $M_{lb}$ for the left cord and as $M_{rb}$ for the right one. The cover will be represented by a set of *k-1* lumped masses $M_{li}$ and $M_{ri}$, $1 \le i \le k-1$ linked by springs among themselves and to the body mass. Each spring will be represented by an elasticity parameter given by $K_{lij}$ (left cord) and $K_{rij}$ (right cord) where $i$ and $j$ refer to the masses linked ($i,j=0$ will point to the body mass). It will be assumed that a loss factor $R_{l,rij}$ will be also associated to each spring to have viscous and other losses into account. A representation of the vocal fold dynamical relations may be seen in **Fig. 1.c** including a body mass and two cover masses. This is the simplest model which can grant a proper study of the *mucosal wave* phenomenon, and has been widely studied in the literature [1][12]. The estimation of the cord movement is carried out in several steps as explained in **Fig. 2**. The vocal tract is removed from the voice trace by inverting the voice producing model ([3], pg. 193) to generate the glottal source $u_g(n)$.

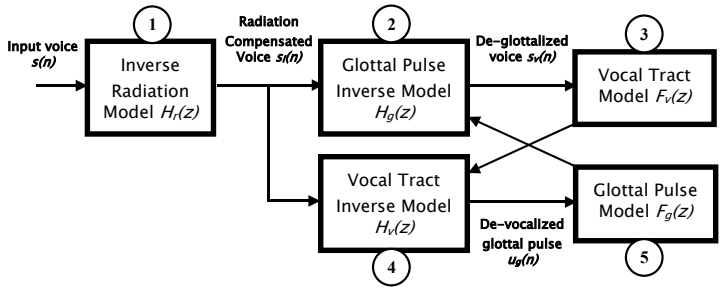



**Fig. 2.** Estimation of the *glottal pulse* $\boldsymbol{u_g(n)}$ by coupled model estimation and inversion

Step 1 consists in removing the radiation effects from voice $s(n)$ (see **Fig. 3.a**) by filtering with $H_r(z)$. Step 2 consists in removing the glottal pulse generating model $H_g(z)$ from the radiation compensated voice $s_l(n)$. In the first iteration $H_g(z)$ need not be a very precise estimation, as it will be refined by successive iterations. In step 3 the vocal tract model $F_v(z)$ is estimated from the de-glottalized voice $s_v(n)$. Step 4 will consist in removing the vocal tract model by filtering with the inverse function $H_v(z)$ to obtain a better estimation of the glottal source $u_g(n)$. Step 5 produces a more precise model of the glottal source $F_g(z)$, which could be used to refine $H_g(z)$. The procedure will repeat steps 2-5 to a desired end. The whole process is described in more detail in previous work [6]. The *glottal aperture* $u_g(n)$ as shown in **Fig. 3.c** is composed by the body mass movement and by the *mucosal wave oscillation* produced by

the cover masses. In the present study emphasis will be placed in adjusting the spectral behavior of the *UGA* to estimate the biomechanical parameters of the cord body. The mathematical process to separate the body component is described in [6] and [7]. The main hypothesis is that the envelope of the *power spectral density of the unfolded glottal aperture* is determined by the admittance of the cord body dynamics, as explained later. The *unfolded glottal aperture (UGA)* will be defined as:

$$u_{gu}(n) = (-1)^k u_g(n \in w_k) \quad (\mathbf{1})$$

where $u_g(n)$ is the *glottal aperture*, and $w_k$ is the *k-th period window* on $u_{gu}(n)$.

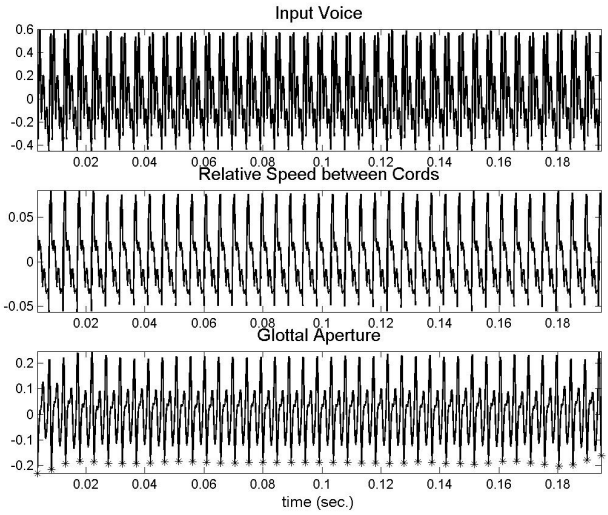



**Fig. 3.** a) Input voice *s(n)*. b) Glottal speed. c) Glottal aperture $u_g(n)$, (*) unfolding points (**1**)

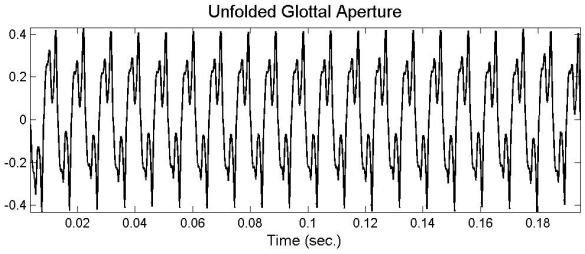



**Fig. 4.** Unfolded glottal aperture $u_{gu}(n)$

The unfolded glottal aperture in **Fig. 4** could be seen as the excursion that one of the cords would describe if vibrating freely (no opposite cord).

## 3 Estimation of the body biomechanical parameters

The estimation of the body biomechanical parameters is related to the inversion of the integro-differential equations for the cord system, as given for the left cord by:

$$f_{xl} = v_{lb} R_{lb} + M_{lb} \frac{dv_{xl}}{dt} + K_{lb} \int_{-\infty}^{t} v_{lb} dt \tag{2}$$

where the biomechanical parameters are the *lumped masses* $M_{lb}$ (left cord) and $M_{rb}$ (right cord), the *elastic parameters* $K_{lb}$ and $K_{rb}$ and the losses $R_{lb}$ and $R_{rb}$. The behavior of this model will be studied in the frequency domain, associating the force $f_{xl}$ on the body with the velocity $v_{lb}$ in the frequency domain. The study will be carried out for the left cord, the right cord showing a similar behavior, which will not be given here for the sake of brevity. The relationship between velocity and force in the frequency domain could be given in the terms of the electromechanical equivalent admittance (cord body admittance). Our interest will be to use the frequency behavior of this admittance in the process of cord-body biomechanical parameter estimation. It will be assumed that the *UGA psd* is determined exclusively by the square modulus of the body admittance $Y_{bl}(s)$ as given in (**3**). To explore this assumption the long-time power spectral density of a frame of *0.2 sec*. of $u_{gu}(n)$ will be matched against the square modulus of the admittance

$$T_{mb}(\omega) = \left|Y_{bl}\right|^2 = \left|\frac{V_{lb}(\omega)}{F_{xl}(\omega)}\right|^2 = \left[\left(\omega M_{lb} - \omega^{-1} K_{lb}\right)^2 + R_{lb}^2\right]^{-1} \tag{3}$$

showing a maximum value at the position of the pitch which is given by

$$T_1 = T_{mb}(\omega = \omega_r) = \frac{1}{R_{lb}^2}; \quad \omega_r = \sqrt{\frac{K_{lb}}{M_{lb}}} \tag{4}$$

The value of the third harmonic on the same power spectral density will be given by

$$T_3 = T_{mb}(\omega = 3\omega_r) = \frac{1}{\left(\frac{8}{3}\right)^2 \omega_r^2 M_{lb}^2 + R_{lb}^2} \tag{5}$$

where (**4**) has been used. From this expression the following estimate for the body mass will be obtained

$$M_{lb} = \frac{3}{8\omega_r}\left[\frac{T_1 - T_3}{T_1 T_3}\right]^{\frac{1}{2}} = \frac{3}{8\omega_r}\sqrt{r_{13}}; \quad r_{13} = \frac{T_1 - T_3}{T_1 T_3} \tag{6}$$

On its turn the value of $K_{lb}$ could be derived from (**4**). When plotting the corresponding function $T$ from (**3**) as given in **Fig. 5.1.c** what is found is that the matching between curves is not complete.

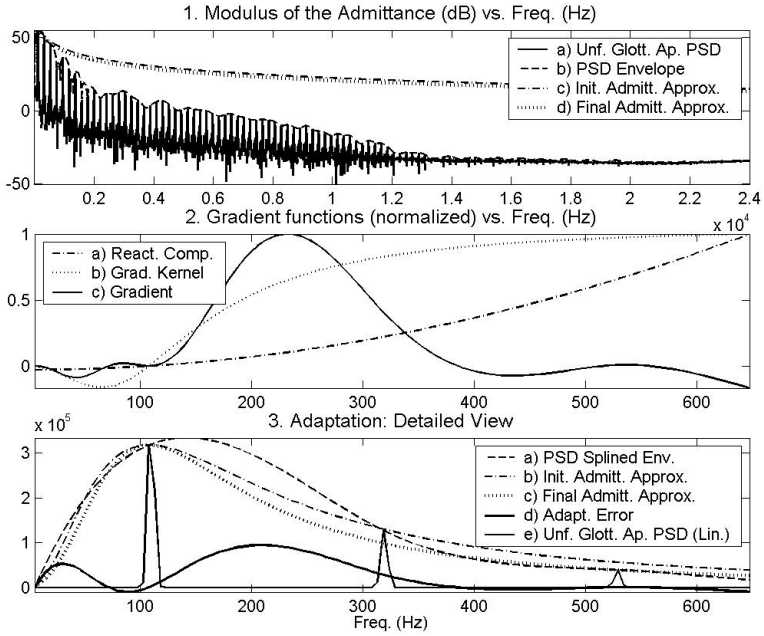



**Fig. 5.1.a**) Power spectral density of the unfolded glottal aperture. **1.b**) Interpolation of its envelope using splines. **1.c**) Approximation of the envelope by the function (**3**) using the estimate (**6**). **1.d**) Approximation of the envelope using the adaptive estimate (**9**). **2.a**) Reactive component of the Admittance. **2.b**) Kernel of the Admittance used in the projection given in (**8**) to estimate the adaptation gradient. **2.c**) Gradient function vs. frequency. **3.a**) Detailed view of the envelope interpolated by splines. **3.b**) Initial approximation by (**3**) using the mass estimate in (**6**). **3.c**) Final estimation using the adaptive approximation given in (**9**). **3.d**) Adaptation error. **3.f**) First harmonic lines of the unfolded glottal aperture (linear)

To refine the estimation an adaptive process is set forth, consisting in minimizing a cost function $L$ defined on the mean square error between the splined approach of the envelope of the *UGA psd* $T_{sb}$ (see **Fig. 5.3.a**) versus the square modulus of the cord body admittance $T_{mb}$

$$L = \int_0^{\omega_s} \varepsilon_T^2 d\omega; \quad \varepsilon_T = T_{sb} - T_{mb} \tag{7}$$

where $\omega_s$ is the Nyquist frequency. To achieve this approximation, the component of the gradient of $L$ with respect to $M$ will be used, which can be derived as

$$\frac{\partial L}{\partial M} = 4\int_0^{\omega_s} \varepsilon_T \, T_{mb}(\omega) Q_Y(\omega) d\omega; \quad Q_Y(\omega) = (\omega^2 M_{lb} - K_{lb}) \tag{8}$$

with which the process of adaptive estimation may be carried out as

$$M_{lb}^{k} = M_{lb}^{k-1} - \mu \frac{\partial L}{\partial M_{lb}} \tag{9}$$

$\mu$ being the classical adaptation step. Using this adaptation process a new estimation for the body mass may be obtained (see **Fig. 5.3.c**), which produces a relative refinement on the estimation of the mass relative to the one based solely in (**6**). To see this more clearly compare the functions in linear plot, as the splined approach (**Fig. 5.3.a**) against the third formant-based estimate (**Fig. 5.3.b**) and the adaptive estimate (**Fig. 5.3.c**). The adaptation error for this last case is plotted in **Fig. 5.3.d** in closer view. The normalized values of the gradient kernels are shown in **Fig. 5.2**. $Q_Y$ is given in **Fig. 5.2.a**, $T_{mb}Q_Y$ is represented in **Fig. 5.2.b**, and the gradient $\partial L/\partial M$ is given in **Fig. 5.2.c**. To compare the refinement obtained in the estimation of the body mass using the adaptive process **Table 1** illustrate the values obtained for the biomechanical parameters of the cord body accordingly with the two estimation methods used (direct and adaptative) for the case under study (voice from a 2-mass model synthetic trace).

**Table 1.** Comparison between the biomechanical parameters obtained from the third harmonic and the adaptive estimation methods

| **Estimation method** | **Body Mass ($M_{lb}$)** | **Losses ($R_{lb}$)** | **Elasticity ($K_{lb}$)** |
|---|---|---|---|
| Direct (3rd harmonic) | 2,1500e-004 | 5,5331e-005 | 138.500 |
| Adaptive | 2,6710e-004 | 5,5331e-005 | 171,900 |

It may be seen that in this case the refinement obtained with the adaptive method is on the order of a 24%. The fact that the mass of the cord body seems to be clearly related to the ratio between the values of the *UGA psd* for the first and third harmonics gives substantial support to the use of this parameter as an important distortion measure as certain studies on pathological voice suggest [9].

## 4 Results and discussion

At this point what seems most crucial is to evaluate the accuracy in the exposed method. To obtain direct *in vivo* estimations of the biomechanical parameters and voice records from normophonic and pathological cases to establish the accuracy of the method seems to be rather difficult. Another more practical approach is to use a *k-mass* model of the vocal folds to produce voice traces assigning *a priori* known values for the biomechanical parameters, and use the estimation methods proposed in the present study to infer the values of the parameters, comparing the estimates obtained against the values introduced in the model. For such 16 voice traces where synthesized using a *2-mass* model of the vocal folds. The value of the subglottal mass ($M_{l1}$=$M_{r1}$) was fixed to *0.2 g.* The supraglottal mass was varied from *0.005 to 0.05 g.* (see **Fig. 6.1.b** and **c**). On its turn the elasticities linking both masses to the reference wall ($K_{l1}$=$K_{r1}$) were set to *110 g.sec*$^{-2}$ whereas the interelasticity linking subglottal and supraglottal masses ($K_{l12}$=$K_{r12}$) was varied from *5* to *255 g.sec*$^{-2}$ in alternating

steps as shown in **Fig. 6.3.b** and **c**. The value for the theoretical pitch generated by the model values was fixed to *120 Hz* for all cases. The value of the losses was fixed to $4.10^{-2}$ $g.sec^{-1}$ for the whole set of traces. A model of the acoustic tube (vocal tract) with *64 sections* for the vowel */a/* was chained to the vocal fold model to generate vowel-like voice. Traces lasting *0.5 sec.* were generated at a sampling frequency of *48.000 Hz*. These traces were subject to the set of procedures described in **section 2** to obtain the unfolded glottal aperture, and used in determining their power spectral density and the body biomechanical parameters as described in **section 3**. The resulting estimations are displayed in **Fig. 6** and listed in **Table 2**.

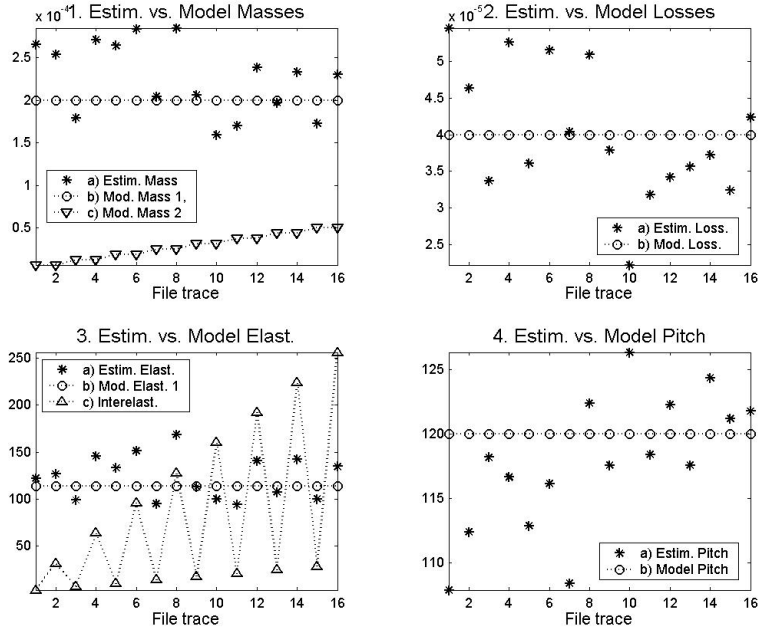



**Fig. 6.1.a)** Estimated values for cord body masses. **1.b**) Model values for subglottal masses. **1.c**) Model values for supraglottal masses. **2.a**) Estimated values for cord body losses. **2.b**) Model values for subglottal and supraglottal losses. **3.a**) Estimated values for cord body elasticity. **3.b**) Model values for subglottal and supraglottal elasticity. **3.c**) Model values for interelasticity. **4.a**) Estimated values for the pitch. **4.b**) Model values for pitch

What may be appreciated from **Fig. 6** is that the estimation of the body mass is centered around the value fixed in the model for the subglottal masses, the estimates showing slight apparent contamination by crosstalk from the supraglottal masses. This is also the case of body elasticities, where a small influence from the interelasticity seems to slightly contaminate the estimates for the body elasticity. Interelasticity crosstalk seems to exert also some influence in the estimation of the losses. The estimation of the pitch as obtained from the power spectral density of the unfolded glottal aperture is also reasonable. The dispersion of the parameters as seen from **Table 2** seems to be in the order of a 25 %.

**Table 2.** Values of the estimated body parameters for the set of synthetic voice traces plotted in **Fig. 6**

| File No. | Mb | Rb | Kb | fp |
|---|---|---|---|---|
| 1 | 2.660e-004 | 5.456e-005 | 122.255 | 107.900 |
| 2 | 2.539e-004 | 4.638e-005 | 126.641 | 112.400 |
| 3 | 1.793e-004 | 3.369e-005 | 98.919 | 118.201 |
| 4 | 2.714e-004 | 5.262e-005 | 145.814 | 116.665 |
| 5 | 2.643e-004 | 3.607e-005 | 132.950 | 112.873 |
| 6 | 2.838e-004 | 5.156e-005 | 151.128 | 116.149 |
| 7 | 2.046e-004 | 4.041e-005 | 94.956 | 108.414 |
| 8 | 2.850e-004 | 5.090e-005 | 168.515 | 122.384 |
| 9 | 2.064e-004 | 3.790e-005 | 112.653 | 117.584 |
| 10 | 1.591e-004 | 2.221e-005 | 100.238 | 126.342 |
| 11 | 1.700e-004 | 3.184e-005 | 94.090 | 118.407 |
| 12 | 2.385e-004 | 3.424e-005 | 140.773 | 122.277 |
| 13 | 1.971e-004 | 3.562e-005 | 107.553 | 117.576 |
| 14 | 2.334e-004 | 3.723e-005 | 142.443 | 124.347 |
| 15 | 1.726e-004 | 3.239e-005 | 100.078 | 121.194 |
| 16 | 2.300e-004 | 4.238e-005 | 134.663 | 121.771 |
| **Means:** | 2.260e-004 | 4.000e-005 | 123.354 | 117.780 |
| **Std. Dev.:** | 4.264e-005 | 9.060e-006 | 22.972 | 5.344 |

The referencing of traces have been carried out comparing the mass and elasticity average estimates against the values used in the models. The relative gains for mass and elasticity coefficients have been found to be $G_{ma}=0.0056$, $G_{ka}=0.0040$, which are in good agreement. The absolute referencing for the determination of the losses is very much related to the energy of the trace as obtained from its autocorrelation function, and is still under study. Practical estimations have yielded the value of $G_{ra}=32.5312$ for this set of experiments, but the question is not closed yet. Another important question which is also left for future research is the issue of mass unbalance, as it is of most interest to infer mass differences between cords related to several critical pathologies. This study is being conducted defining the common and differential modes of cord vibration, and from these a contribution associated to each cord body could be established. The same may be said for cord elasticities. Results on this line are being prepared for further publication.

## 5 Conclusions

Through the present paper the possibility of obtaining indirect estimates of the vocal cord biomechanical parameters from the voice trace has been shown. This could open new possibilities for the non-invasive distant exploration of the larynx of patients both for pathology detection and classification by analysis of the voice trace. The method is still subject to revision to reduce the influence of second-order biomechanical parameters regarding accuracy Its possible extension to unbalanced parameter estimation is also under study.

## Acknowledgments

This research is being carried out under grants TIC2002-02273 and TIC2003-08756 from the *Programa Nacional de las Tecnologías de la Información y las Comunicaciones (Spain)*.

## References

[1] Berry, D. A.: Mechanisms of modal and nonmodal phonation. Journal of Phonetics, Vol. 29, (2001), pp. 431-450.
[2] Childers, D. G.: Speech Processing and Synthesis Toolboxes. John Wiley & Sons, New York (2000).
[3] Deller, J. R., Hansen, J. H. L., Proakis, J. G.: Discrete-Time Processing of Speech Signals. John Wiley & Sons, New York, 2000.
[4] Dibazar, A. A., Narayanan, S.: A System for Automatic Detection of Pathological Speech. Proc. of the 36th Asilomar Conf. on Signals, Systems and Computers (2002).
[5] Godino, J. I., Gómez, P.: Automatic Detection of Voice Impairments by means of Short Term Cepstral Parameters and Neural Network based Detectors. IEEE Trans. on Biomed. Eng. Vol. 51, No. 2 (2004), pp. 380-384.
[6] Gómez, P., Díaz, F., Martínez, R., Godino, J. I., Álvarez, A., Rodríguez, F. , Rodellar, V.: Precise Reconstruction of the Mucosal Wave for Voice Pathology Detection and Classification. Proceedings of the EUSIPCO'04 (2004) 297-300.
[7] Gómez, P., Godino, J. I., Díaz, F., Álvarez, A., Martínez, R., Rodellar, V.: Biomechanical Parameter Fingerprint in the Mucosal Wave Power Spectral Density, Proc. of the ICSLP'04 (2004) 842-845.
[8] Growing Problem for Call Centre Workers Suffering from Repetitive Voice Injury. http://www.healthandsafetytips.co.uk/News.htm
[9] Kuo, J., Holmberg, E. B., Hillman, R. E.: Discriminating Speakers with Vocal Nodules Using Aerodynamic and Acoustic Features. Proc. of the ICASSP'99, Vol. 1, 15-19 March (1999), pp. 77-80.
[10] Ritchings, T.; McGillion, M.; Moore, C.: Pathological voice quality assessment using artificial neural networks. Medical Engineering and Physics, vol. 24 (2002), pp. 561-564.
[11] Simberg, S.: Prevalence of Vocal Symptoms and Voice Disorders among Teacher Students and Teachers and a Model of Early Intervention, Ph.D. Thesis, Department of Speech Sciences, University of Helsinki, http://ethesis.helsinki.fi/.
[12] Story, B. H.: An overview of the physiology, physics and modeling of the sound source for vowels. Acoustic Sci. & Tech., Vol. 23, 4 (2002), pp. 195-206.
[13] Story, B. H., and Titze, I. R.: Voice simulation with a bodycover model of the vocal folds. J. Acoust. Soc. Am., Vol. 97 (1995), pp. 1249–1260.
[14] The Voice Center of Eastern Virginia Medical School. http://www.voice-center.com/larynx_ca.html.
[15] Titze, I. R.: Principles of Voice Production. Prentice-Hall, Englewood Cliffs (1994).
[16] Titze, I. R.: The physics of small-amplitude oscillation of the vocal folds. Journal of the Acoustical Society of America, Vol. 83, No. 4 (1988) 1536-1552.
[17] Titze, I.R., Lemke, J., & Montequin, D.: Populations in the U.S. workforce who rely on voice as a primary tool of trade: A preliminary report. Journal of Voice 11, (1997) 254–259.