Voice Pathology Detection by Vocal Cord Biomechanical Parameter Estimation

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

Facultad de Informática, Universidad Politécnica de Madrid,
Campus de Montegancedo, s/n28660 Boadilla del Monte, Madrid, Spain
pedro@pino.datsi.fi.upm.es

Abstract. Voice pathologies have become a social concern, as voice and speech play an important role in certain professions, and in the general population quality of life. In these last years emphasis has been placed in early pathology detection, for which classical perturbation measurements (*jitter, shimmer, HNR*, etc.) have been used. 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 non only in pathology detection, but in classifying the specific patient's pathology as well. 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. From these, deviations from normophonic cases, and unbalance between cords may be extracted to serve as pathology correlates. The relevance of deviations and unbalance in Pathology Detection is shown through Principal Component Analysis. Results for normal and pathological cases will be presented and discussed.

1 Introduction

Voice pathology detection is a field of important research area in voice and speech processing as it may affect the quality of life of the population, especially in people who use voice extensively in their professional activity, as speakers, singers, actors, lawyers, broadcasters, priests, teachers, call center workers, etc [13][16][22]. 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 if other specialized inspection methods –as videoendoscopy- are to be used, as these being more precise in pathology classification, are at the same time less comfortable, more expensive and complicate, and their use should be obliviated if a simple inspection could help in screening patients before being subject to full inspection procedures. The estimation of biomechanical parameters associated to the structure of the phonation organs would suppose an important improvement in the use of voice for pathology screening. Up-to-date techniques use time- and frequency-domain estimation of *perturbation parameters*, which measure the deviation of the specific patient's

M. Faundez-Zanuy et al. (Eds.): NOLISP 2005, LNAI 3817, pp. 242 – 256, 2005.

voice from certain normal standards [6][20][15][7]. 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 a study has been initiated to estimate the values of the biomechanical parameters of the vocal fold system (mass, stiffness and losses) from the glottal source obtained from voice after removing the vocal tract transfer function. This procedure is well documented in the literature (see for example [2] [5] [1]), and produces a trace which can be shown to be directly related with the glottal source (the average aperture measured between vocal cords during the phonatory cycle) [8]. The use of *k-mass* vocal fold models [18][3] help in determining that there are two main components in the movement of the vocal cord, contributed by the structure of the cord: the movement of the bulk muscular tissue of the cord body (see **Figure 1.a**) and a traveling wave known as the *mucosal wave* [21][17], which is contributed by the epithelial tissue of the cord cover (see **Figure 1.b**).

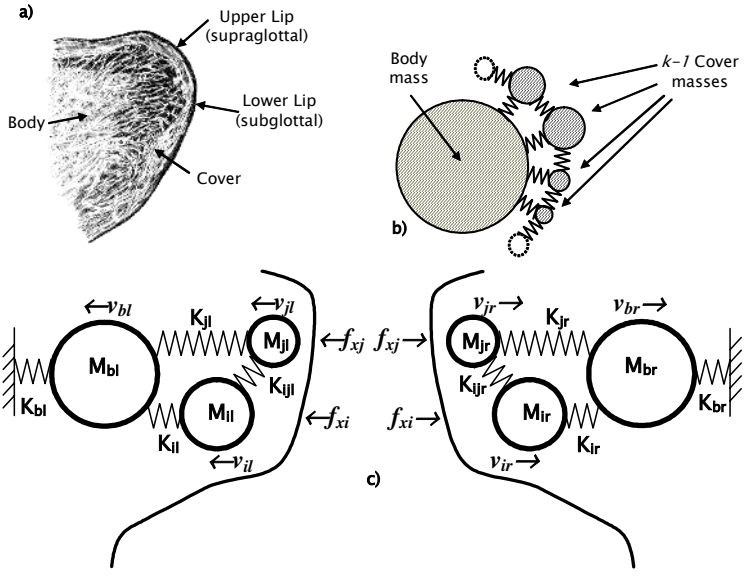

Fig. 1. a) Cross-section of the left vocal cord showing the body and cover structures (taken from [19]). **b**) *k-mass* model of the body and cover. **c**) *3-mass* model used to establish the dynamics of the body-cover system.

In previous research [8][9] it has been shown that both contributions present in the glottal source can be separated to produce two traces, known as the *avrage acoustic wave* or *average glottal source (AGS)* 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 reduction or complete disappearance of the *mucosal wave* [14]. In the present study the emphasis will be placed in using the *average glottal source* 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 hypothesizing that the fingerprint of the cord body dynamics

is responsible for the *power spectral density (psd)* of the *AGS*, 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. In this way a first estimation of the biomechanical parameters is obtained, which can be later refined adaptively.

2 Estimating Cord Dynamics

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 [20]. A cross section of a vocal cord can be seen in 0.a, showing its tissular structure, which is composed by the body and the cover as mentioned before. In **Figure 1.b** an equivalent *k-mass* model is presented, where the main structure (the body) has been represented by a large lump which referred to as M_{lb} (left cord) and M_{rb} (right cord). The cover is represented by a set of *k-1* lumped masses M_{li} and M_{ri}, $1 \leq i \leq k-1$, linked by springs among themselves and to the body mass. Each spring is represented by a stiffness 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 0.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 ([3][17][18]). The estimation of the cord movement is based on the pioneering work by Alku ([2][23]), which has been modified for an iterative implementation in several steps as shown in **Figure 2**.

Step 1 consists in removing the radiation effects from voice *s(n)* (see **Figure3.a**) by filtering with $H_r(z)$. Step 2 consists in removing the glottal pulse generating model $F_g(z)$ by by its inverse $H_g(z)$ from the radiation compensated voice $s_l(n)$. In the first

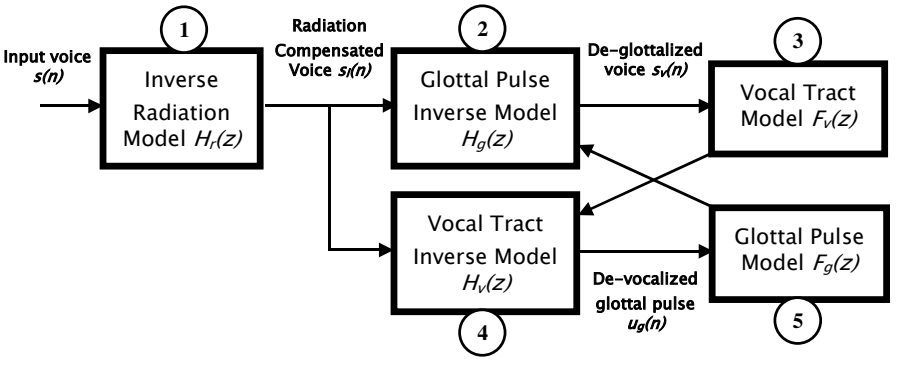

Fig. 2. Estimation of the *glottal pulse* $\boldsymbol{u_g(n)}$ by coupled model estimation and inversion

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 deglottalized voice $s_v(n)$. Step 4 will consist in removing the vocal tract model by filtering $s_v(n)$ with the vocal tract 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 [8]. The *glottal source* $u_g(n)$ as shown in **Figure 3.c** is composed by the body mass movement (cord body dynamics) and by the *mucosal wave oscillation* produced by the cover masses (cord cover dynamics). The *mucosal wave correlate* will be defined as:

$$y_m(n) = u_g(n) - \overline{u}_g(n); \quad n \in W_k \tag{1}$$

where $y_m(n)$ is the *mucosal wave correlate*, and W_k is the *k-th period window* on $u_g(n)$.

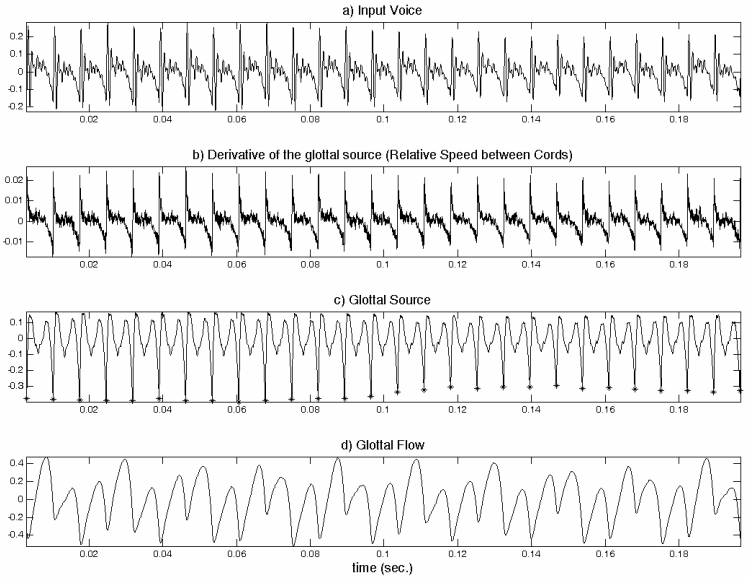

Fig. 3. a) Input voice s(n). b) Glottal source derivative. c) Glottal source $u_g(n)$ unfolding points (*). d) Unleveled glottal pulse.

The effects of vocal tract coupling have been neglected.The traces shown in **Figure 4.a**, **b** and **c** are respectively the ground leveled version of the *glottal source* $u_g(n)$, the *average glottal wave* $\overline{u}_g(n)$, and the leveled *glottal pulse*. As the *average glottal source* may be associated with the simplest cord body dynamics, the difference between the *glottal source* and the *average glottal source* may be considered as contributed by the cord cover dynamics, and can be seen as the *mucosal wave correlate*, as shown in **Figure 4.d**. In the present study emphasis will be placed in adjusting the spectral behavior of the *mucosal wave correlate* to estimate the biomechanical

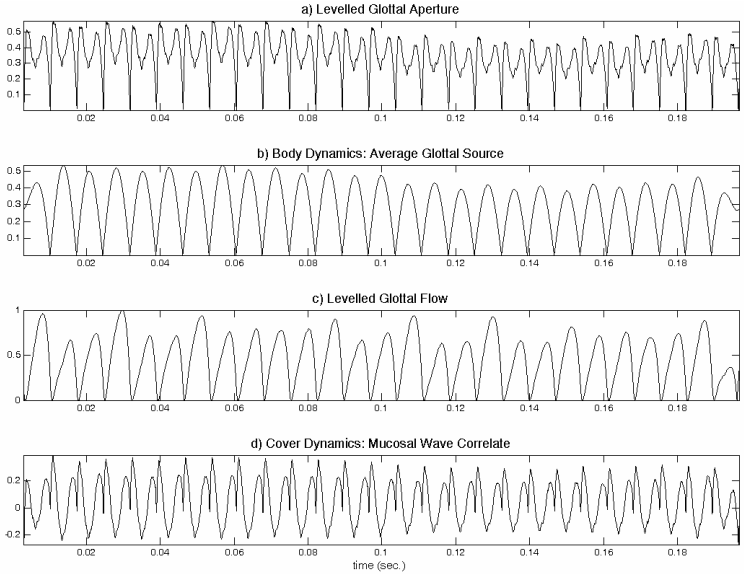

Fig. 4. a) Leveled glottal source, b) average glottal source, c) Leveled glottal pulse, d) cord body dynamics: mucosal wave correlate

parameters of the cord body as described in [8] and [9]. The main hypothesis is that the envelope of the *power spectral density of the mucosal wave correlate* is determined by the admittance of the cord body dynamics, as explained later.

The estimation of $\overline{u}_g(n)$ is carried out for each pitch period (cycle) as the subtraction of a half-wave sinusoidal arch with the same semi-period as the source, using an adaptive method to evaluate the amplitude of the arch, based on the minimization of the energy of the error between the glottal source $u_g(n)$ and the sinusoidal average $\overline{u}_g(n)$ as

$$L = \sum_{n \in W_k} y_m^2(n) = \sum_{n \in W_k} \left(u_g(n) - \overline{u}_g(n)\right)^2 \tag{2}$$

with

$$\overline{u}_g(n) = u_{ok}\, sin(\omega_k n \tau); \quad n \in W_k \tag{3}$$

ω_k being the angular frequency associated to the *k-th* cycle semi-period and τ being the sampling period. The optimization of the amplitude of each sinusoidal arch will be derived minimizing the cost function L in terms of u_{0k} as

$$\frac{\partial L}{\partial u_{0k}} = 0 \quad \Rightarrow \quad u_{0k} = \frac{\sum_{n \in W_k} y_{gk}(n)\, sin(\omega_k n \tau)}{\sum_{n \in W_k} sin^2(\omega_k n \tau)}. \tag{4}$$

3 Estimation of the Body Biomechanical Parameters

Detecting the cord body mass, stiffness and damping is based on the inversion of the integro-differential equation of the one-mass cord model, which for the left vocal cord would be

$$f_{xl} = v_{lb}R_{lb} + M_{lb}\frac{dv_{xl}}{dt} + K_{lb}\int_{-\infty}^{t} v_{xl}dt \tag{5}$$

where the biomechanical parameters involved are the lumped masses M_{lb}, the stiffness K_{lb} and the losses R_{lb}. The equivalent model is shown in **Figure 5**. The estimation of the body biomechanical parameters is related to the inversion of this model, associating the force f_{xl} on the body with the velocity of the cord centre of masses v_{xl} in the frequency domain.

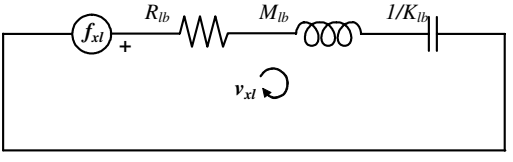

Fig. 5. Electromechanical equivalent of a cord body

The relationship between velocity and force in the frequency domain is expressed as the cord body admittance. The working hypothesis for the process of biomechanical parameter estimation will be based on the assumption that the envelope of the power spectral distribution of the *mucosal wave correlate* (cover dynamic component) is directly related with the square modulus of the input admittance to the electromechanical equivalent $Y_{bl}(s)$ given as

$$T_b(\omega) = |Y_{bl}|^2 = \left|\frac{V_{xl}(\omega)}{F_{xl}(\omega)}\right|^2 = \left[\left(\omega M_{lb} - \omega^{-1}K_{lb}\right)^2 + R_{lb}^2\right]^{-1} \tag{6}$$

The robust estimation of the model parameters is based in the determination of two points on the power spectral density of the cover dynamic component [10] $\{T_{b1}, \omega_1\}$ and $\{T_{b2}, \omega_2\}$, from which the lumped body mass (BM) may be estimated as

$$M_{lb} = \frac{\omega_2}{\omega_2^2 - \omega_1^2}\sqrt{\frac{T_{b1} - T_{b2}}{T_{b1}T_{b2}}} \tag{7}$$

On its turn the elastic parameter (body stiffness: BS) K_{lb} may be estimated from the precise determination of the position of the resonant peak, this being $\{T_r, \omega_r\}$

$$K_{lb} = M_{lb}\omega_r^2 \tag{8}$$

whereas the of body losses (BL) may be estimated (but for a scale factor G_b) as

$$R_{lb} = \frac{G_b}{\sqrt{T_r}} \tag{9}$$

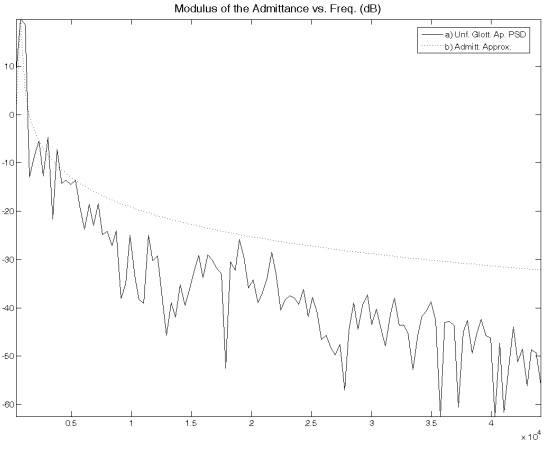

Fig. 6. Parametric fitting of the *mucosal wave power spectral density* for a cycle of the sample trace (full line) against the admittance approximation (dot line)

The estimations obtained from a phonation cycle of a normophonic voice trace have been used to reconstruct the approximated square modulus of the admittance, which is presented in **Figure 6** against the power spectral density of the cover dynamics component for comparison.

Table 1 illustrates the values obtained for the biomechanical parameters of the cord body accordingly with two estimation algorithms (direct and adaptative) from a 2-mass model synthetic voice trace.

Table 1. Comparison between the biomechanical parameters obtained from (6-9) by direct and adaptive estimations

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 the divergence between both methods 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 *mucosal wave correlate power spectral density* for the first and third harmonics if $\omega_1 = \omega_r$ and $\omega_2 = 3\omega_r$ gives substantial support to the use of this parameter as an important distortion measure as certain studies on pathological voice suggest [14][4].

4 Results for Synthetic Voice

At this point what seems most crucial is to evaluate the accuracy in the exposed method. Obtaining 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 **Figure 7.1.b** and **c**).

On its turn the springs linking both masses to the reference wall (K_{l1}=K_{r1}) were set to *110* $g.sec^{-2}$ whereas the stiffness linking subglottal and supraglottal masses (K_{l12}=K_{r12}) was varied from *5* to *255* $g.sec^{-2}$ in alternating steps as shown in Figure **7.3.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 were treated as described in **section 2** to obtain the *mucosal wave correlate*, and used in determining its power spectral density and the body biomechanical parameters as

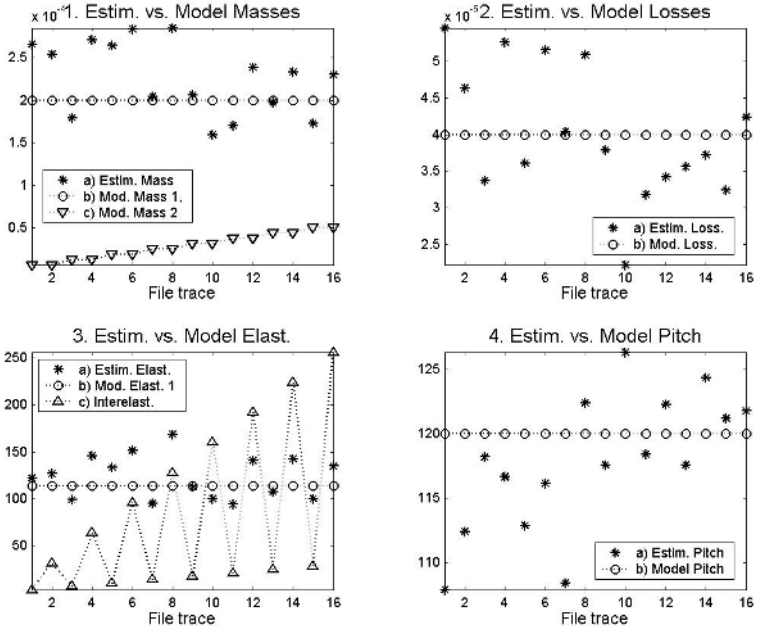

Fig. 7. 1.a) Estimated values for cord body masses. 1.b) Model values for subglottal masses. 1.c) Model values for supraglottal masses. 2.a) Estimated values for cord body losses. 2.b) Model values for subglottal and supraglottal losses. 3.a) Estimated values for cord body elasticity. 3.b) Model values for subglottal and supraglottal elasticity. 3.c) Model values for interelasticity. 4.a) Estimated values for the pitch. 4.b) Model values for pitch.

Table 2. Values of the estimated body parameters for the set of synthetic voice traces plotted in Figure 7

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

described in **section 3**. The resulting estimations are displayed in **Table 2** and listed in **Figure 7**. It may be appreciated from **Figure 7** 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 stiffness, where a small influence from the interelasticity seems to slightly contaminate the estimates. 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 source is also reasonable. The dispersion of the parameters as seen from **Table 2** seems to be in the order of a 25 %.

The referencing of traces has 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 $G_{ra}= 32.53$ for this set of experiments, but the question is not closed yet. Another important question 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 stiffness. A slight unbalance between waveform cycles may be observed in **Figure 4.a)** and **c)**. As estimations of mass, stiffness and losses will be available by cycles, the unbalance of these parameters (**BMU** – Body Mass

Unbalance, **BLU** – Body Losses Unbalance and **BSU** – Body Stiffness Unbalance) may be defined as

$$\begin{aligned} \mu_{uk} &= \left(\hat{M}_{bk} - \hat{M}_{bk-1}\right) / \left(\hat{M}_{bk} + \hat{M}_{bk-1}\right) \\ \rho_{uk} &= \left(\hat{R}_{bk} - \hat{R}_{bk-1}\right) / \left(\hat{R}_{bk} + \hat{R}_{bk-1}\right) \\ \gamma_{uk} &= \left(\hat{K}_{bk} - \hat{K}_{bk-1}\right) / \left(\hat{K}_{bk} + \hat{K}_{bk-1}\right) \end{aligned} \tag{10}$$

where $1 \leq k \leq K$ is the *cycle window* index and $\hat{M}_{bk}, \hat{R}_{bk}, and\ \hat{K}_{bk}$ are the *k-th cycle* estimates of mass, losses and stiffness on a given voice sample (intra-speaker). Other parameters of interest are the deviations of the average values of mass, losses and compliance for the *j-th* sample $\overline{M}_{bj}, \overline{R}_{bj}, and\ \hat{K}_{bj}$ relative to average estimates from a normophonic set of speakers (inter-speaker) as

$$\begin{aligned} \mu_{dj} &= \left(\overline{M}_{bj} - \overline{M}_{bs}\right) / \overline{M}_{bs} \\ \rho_{dj} &= \left(\overline{R}_{bj} - \overline{R}_{bs}\right) / \overline{R}_{bs} \\ \gamma_{dj} &= \left(\overline{K}_{bj} - \overline{K}_{bs}\right) / \overline{C}_{bs} \end{aligned} \tag{11}$$

these parameters are known as **BMD** (Body Mass Deviation), **BLD** (Body Losses Deviation) and **BSD** (Body Stiffness Deviation).

5 Results from Natural Voice

A variant of Principal Component Analysis (PCA) known as *multivariate measurements analysis* (see [12], pp. 429-30) Hierarchical Clustering and have been used with the distortion parameters given in **Table 3** [11].

PCA is conceived as the optimal solution to find the minimum order of a linear combination of random variables $\boldsymbol{x}_j$ showing the same variance as the original set, where the components of $\boldsymbol{x}_j$ correspond to different observations (samples) of a given input parameter (*j-th* parameter) for a set of *20* normophonic and *20* pathologic samples (*4* samples with polyps, *6* samples with bilateral nodules, *5* samples with Reinke's Edema, and *5* samples with reflux inflammation) as listed in **Table 4**.

Table 3. List of parameters estimated from voice

Coeff.	Description
x_1	pitch
x_2	jitter
$x_{3\text{-}5}$	shimmer-related
$x_{6\text{-}7}$	glottal closure-related
$x_{8\text{-}10}$	HNR-related
$x_{11\text{-}14}$	mucosal wave psd in energy bins
$x_{15\text{-}23}$	mucosal wave psd singular point values
$x_{24\text{-}32}$	mucosal wave psd singular point positions
$x_{33\text{-}34}$	mucosal wave psd singularity profiles
$x_{35\text{-}37}$	biomechanical parameter deviations (11)
$x_{38\text{-}40}$	biomechanical parameter unbalance (10)

Table 4. Values of $x_{35\text{-}39}$ for the samples studied. Sample conditions are: *N – Normophonic; BP – Bilateral Polyp; LVCP – Left Vocal Cord Polyp; BRE – Bilateral Reinke's Edema; BN – Bilateral Noduli; LR – Larynx Reflux; RE – Reinke's Edema; RVCP – Right Vocal Cord Polyp.*

Trace	Condit.	BMD	BLD	BSD	BMU	BLU
001	N	-0.632	-0.136	-0.540	0.027	0.039
003	N	-0.154	-0.145	-0.137	0.079	0.056
005	N	-0.039	-0.299	-0.213	0.078	0.044
007	N	-0.492	-0.461	-0.573	0.036	0.046
00A	N	-0.542	-0.207	-0.567	0.065	0.064
00B	N?	1.320	0.642	1.250	0.149	0.191
00E	N	-0.054	0.012	-0.128	0.159	0.098
010	N	-0.408	0.164	-0.491	0.115	0.103
018	N	-0.031	-0.205	-0.167	0.078	0.076
01C	N	-0.557	-0.315	-0.581	0.058	0.052
024	N?	0.631	1.330	1.200	0.120	0.124
029	N	0.101	-0.111	0.416	0.057	0.048
02C	N	-0.329	-0.253	-0.079	0.035	0.040
02D	N	-0.227	-0.193	0.022	0.116	0.053
032	N	-0.507	-0.019	-0.367	0.038	0.071
035	N	0.424	-0.302	-0.021	0.099	0.065
043	N	0.219	0.156	0.466	0.059	0.030
047	N	-0.497	1.070	-0.180	0.076	0.052
049	N	-0.157	0.160	0.029	0.113	0.079
04A	N	-0.005	1.770	0.073	0.098	0.075
065	BP	0.240	7.490	3.220	0.835	0.712
069	LVCP	0.560	3.490	2.460	0.408	0.318
06A	BRE	0.142	2.860	1.760	0.300	0.331
06B	BN	0.427	3.860	2.150	0.339	0.326
06D	BN	0.573	3.540	2.160	0.338	0.339
071	BRE	0.417	3.210	1.870	0.306	0.348
077	LR	2.000	3.170	3.660	0.460	0.320
079	RE	0.658	2.860	2.170	0.396	0.333
07E	BN	0.843	2.990	2.340	0.328	0.303
07F	LR	0.420	2.850	1.950	0.332	0.309
083	LR	0.253	2.880	1.900	0.391	0.333
092	BRE	0.216	2.750	1.720	0.469	0.353
098	RE	0.187	2.830	1.720	0.360	0.339
09E	BN	1.400	11.700	5.510	0.637	0.518
09F	LR	0.062	2.920	1.660	0.309	0.334
0A0	RVCP	0.156	3.020	1.720	0.333	0.338
0A9	LVCP	0.012	3.600	1.660	0.293	0.311
0AA	LR	-0.091	2.970	1.600	0.268	0.315
0B4	BN	0.154	4.280	1.870	0.305	0.338
0CA	BN	-0.057	3.040	1.630	0.310	0.361

These samples were processed to extract the set of *40* parameters listed in **Table 3**, of which two subsets were defined for classification: $S_1=\{x_{2\text{-}39}\}$, including most of the parameters available, and $S_2=\{x_2, x_3, x_8, x_{35\text{-}39}\}$ including *jitter*, *shimmer*, *HNR,* deviations (**BMD**, **BLD** and **BSD**), and unbalances (**BMU** and **BLU**). The results of the

clustering process are shown in **Figure 8** as biplots against the two first principal components from PCA analysis. It may be seen that the clustering process assigned most of normophonic samples to one cluster (with the exception of *00B* and *024*) both for S_1 as well as for S_2. The results using S_2 are given in **Table 5**.

Table 5. Clustering results for S_2

Cluster	Samples
c_{21} (o)	001, 003, 005, 007, 00A, 00E, 010, 018, 01C, 029, 02C, 02D, 032, 035, 043, 047, 049, 04A
c_{22} (◊)	00B, 024, 065, 069, 06A, 06B, 06D, 071, 077, 079, 07E, 07F, 083, 092, 098, 09E, 09F, 0A0, 0A9, 0AA, 0B4, 0CA

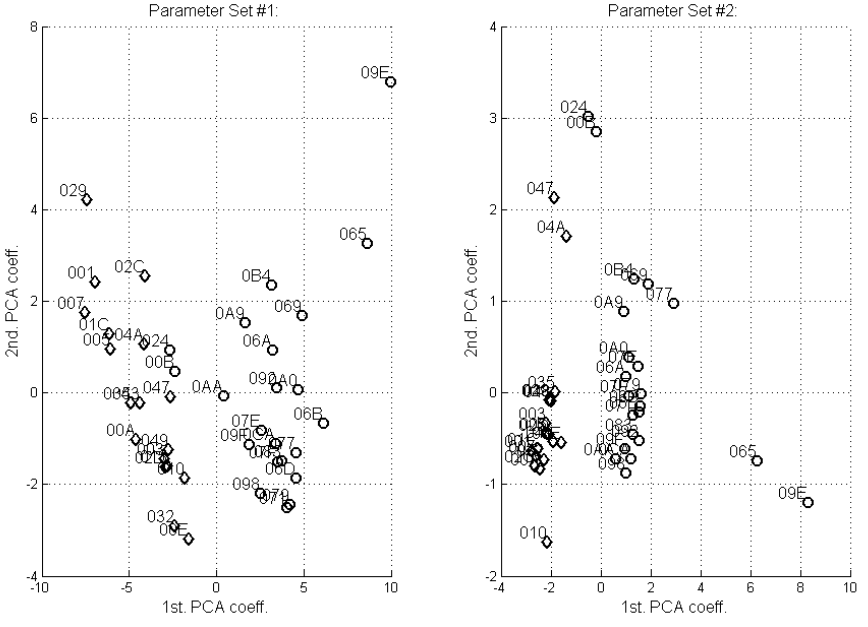

Fig. 8. Left) Clusters for S_1. Right) Clusters for S_2.

To further clarify the analysis a 3D plot of the results vs the three most relevant input parameters in S_2 as established by PCA is presented in **Figure 9**. The most relevant parameter according to this combination seems to be **BSD** (x_{37}). The larger x_{37}, the stiffer the cord and the less normophonic the production. The second most relevant parameter seems to be **jitter** (x_2). The third most relevant parameter is **BLD** (x_{36}) associated to the profile of the spectral profile peak (Q factor).

The behaviour of cases *00B* and *024*, classified as pathological by PCA analysis deserves a brief comment. These appear in **Figure 9** (encircled) not quite far from normal cases 001-04A, but showing a stiffness that doubles those of normophonic

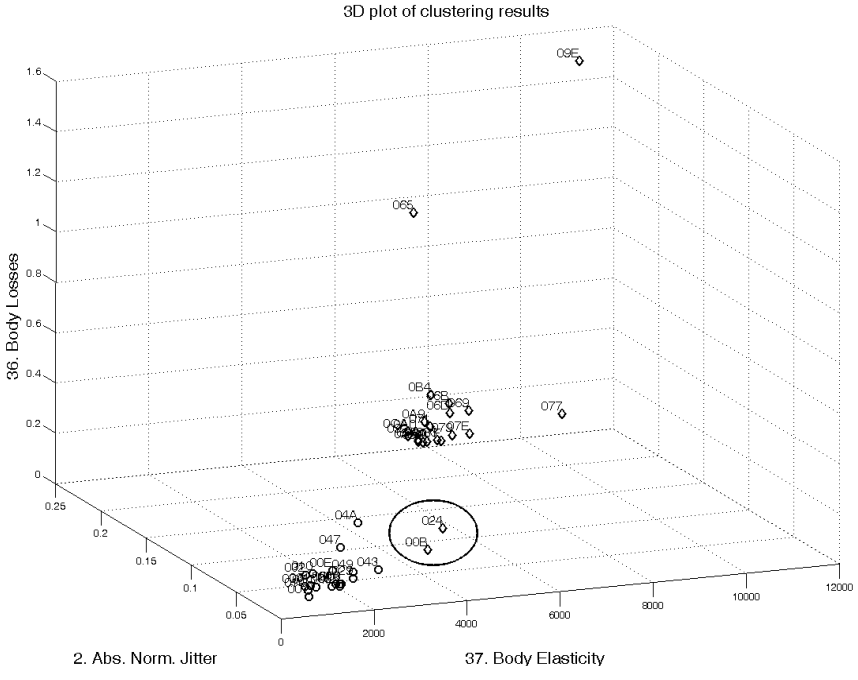

Fig. 9. 3D Clustering Plot showing the separation in the manifold defined by the parameter subset {x37, x2 and x36} – ordered by relevance

samples. Apparently this detail was determinant in their classification as not normophonic by PCA. This fact was confirmed by their values for the **BSD** in **Table 4**, being 1.25 and 1.2 respectively, or 225% and 220%.

6 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 patients both for pathology detection and classification by analysis of the voice trace. The method is still subject to revision to take into account the influence of second-order biomechanical parameters. Its possible extension to unbalanced parameter estimation is also under study. The methodology presented detects biomechanical unbalance from voice records for pathology detection by common pattern recognition techniques. Normophonic samples show small unbalance indices, as opposed to pathologic ones. There is not a specific pattern of unbalance related to a given pathology (although more cases need to be studied). Biomechanical parameter unbalance is a correlate to pathology quantity rather than quality. Although mild pathologies may appear as normophonic from subjective analysis the use of the proposed methods may spot them and help in keeping trace of their evolution in time. Adequately combining classical

distortion parameters with deviation parameters renders fairly good results in pathology detection. These conclusions need to be confirmed by more experiments.

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] Akande, O. O. and Murphy, P. J., "Estimation of the vocal tract transfer function with application to glottal wave analysis, Speech Communication, Vol. 46, No. 1, May 2005, pp. 1-13.

[2] Alku, P., "An Automatic Method to Estimate the Time-Based Parameters of the Glottal Pulseform", ICASSP'92, pp. II/29-32.

[3] Berry, D. A.: Mechanisms of modal and nonmodal phonation. Journal of Phonetics, Vol. 29, (2001), pp. 431-450.

[4] Cannito, M. P., Buder, E. H., Chorna, L. B., "Spectral Amplitude Measures of Adductor Spasmodic Dysphonic Speech", Journal of Voice, Vol. 19, No. 3 (2005), pp. 391-410

[5] Childers, D. G.: Speech Processing and Synthesis Toolboxes. John Wiley & Sons, New York (2000).

[6] 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).

[7] 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.

[8] 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.

[9] 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.

[10] Gómez, P., Godino, J. I., Álvarez, A., Martínez, R., Nieto, V., Rodellar, V., "Evidence of Glottal Source Spectral Features found in Vocal Fold Dynamics", Proceedings of the 2004 Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP'05 (2005), pp. 441-444.

[11] Gómez, P., Díaz, F., Álvarez, A., Martínez, R., Rodellar, V., Fernández, R., Nieto, A., Fernández, F. J.: PCA of Perturbation Parameters in Voice Pathology Detection. Proc. of EUROSPEECH'05. (2005), pp. 645-648.

[12] Johnson, R. A., Wichern, D. W.: Applied Multivariate Statistical Analysis. Prentice-Hall, Upper Saddle River, NJ, 2002.

[13] Growing Problem for Call Centre Workers Suffering from Repetitive Voice Injury. http://www.healthandsafetytips.co.uk/News.htm

[14] 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.

[15] Ritchings, T.; McGillion, M.; Moore, C.: Pathological voice quality assessment using artificial neural networks. Medical Engineering and Physics, vol. 24 (2002), pp. 561-564.
[16] 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/.
[17] 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.
[18] 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.
[19] The Voice Center of Eastern Virginia Medical School. http://www.voice-center.com/larynx_ca.html.
[20] Titze, I. R.: Principles of Voice Production. Prentice-Hall, Englewood Cliffs (1994).
[21] 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.
[22] 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.
[23] Varho, S., Alku, P., "A new predictive method for all-pole modelling of speech spectra with a compressed set of parameters", *Proc. ICASSP99*, 1999, pp. III/126-129.