USING BIOMECHANICAL PARAMETER ESTIMATES IN VOICE PATHOLOGY DETECTION

P. Gómez, C. Lázaro, R. Fernández, A. Nieto, J. I. Godino, R. Martínez, F. Díaz, A. Álvarez, K. Murphy, V. Nieto, V. Rodellar, F. J. Fernández-Camacho

GIAPSI, Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo, s/n, 28660, Boadilla del Monte, Madrid, Spain

It has been shown in previous work that biomechanical parameters related to the cord body dynamics can be indirectly estimated from the power spectral density of the mucosal wave correlate [4]. In the present study the use of these measurements to estimate the presence of parameter unbalance will be shown. The role of these parameters together with the classical distortion ones in relation to pathology detection and classification will be explored. Results using normophonic as well as pathologic voice will be presented and discussed.

I. INTRODUCTION

Classically, Voice Processing focused onto detecting pathological voice by means of distortion parameter estimation directly from the voice trace [7][2], albeit the detection process being masked by the vocal tract and other supra-structures of the vocal apparatus. More advanced methods remove the influence of the vocal tract, to obtain an indirect estimation of the glottal source [1]. The first and second derivatives of the glottal source are correlates of the glottal aperture and the relative speed between cord centers of mass [3][5]. The glottal aperture correlate can be seen as being composed of two main parts: a slow-varying average movement, which is referred to as "the average acoustic waveform" [8], and a fast-varying waveform, resulting from the mucosal wave traveling on the body-cover structure [10][11]. The dynamics of the body would be reflected in the average glottal aperture, whereas the dynamics of the cover would be retained by the mucosal wave correlate (see Figure 1).

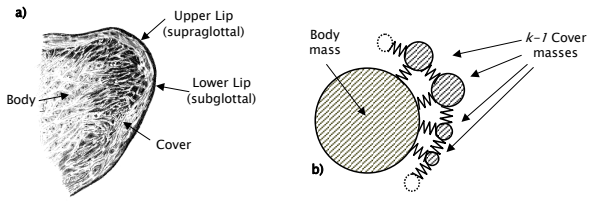

Figure 1. a) Cross-section of the left vocal cord showing the body and cover structures (taken from [9]). b) *k-mass* model of the body and cover.

It may be expected that the power spectral density (psd) of the average acoustic wave would be determined by the dynamics of the cord center of masses, whereas the power spectral density of the mucosal wave correlate would be mostly influenced by the cover dynamics. Moving one step ahead, separating both signals would become an important target for estimating vocal fold biomechanics. In a previous work [4] it was shown that estimates of the cord mass and stiffness could be obtained from the power spectral density of the average acoustic waveform. Through this paper the methodology for parameter unbalance estimation will be presented. Experiments using pathologic and normophonic samples will also be given.

II. CORD BODY BIOMECHANICAL ESTIMATES

Glottal source reconstruction by inverse filtering, as used in the present study, is due to Alku [1]. Relevant details on its recursive implementation by paired lattices are to be found in [5]. By removing the vocal tract influence a given voice trace can be processed to render the relative speed between cords, the glottal aperture and the glottal source as shown in Figure 2.

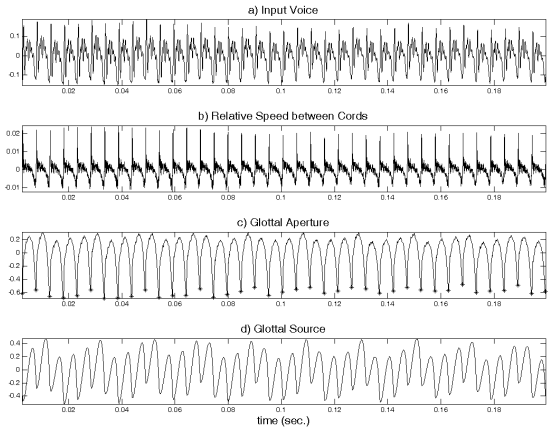

Figure 2. Glottal source estimation. a) input voice (sample *00B*), b) second and c) first derivatives of the glottal source, d) glottal source (unlevelled).

Detecting the cord body mass, stiffness and damping is based on the inversion of the integro-differential equation of the one-mass cord model

$$f_{xl} = v_{lb} R_{lb} + M_{lb} \frac{dv_{xl}}{dt} + \frac{1}{C_{lb}} \int_{-\infty}^{t} v_{lb} dt \qquad (1)$$

Models and analysis of vocal emissions for biomedical applications. 4th international workshop. October 29-31, 2005 – Firenze, Italy.
Edited by C. Manfredi. ISBN 88-8453-320-1 (online)

where the biomechanical parameters involved are the lumped masses M_{lb}, the elastic parameters C_{lb} and the losses R_{lb}. The equivalent model is shown in Figure 4.

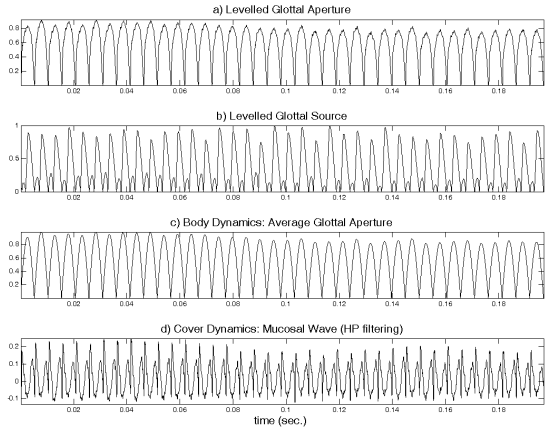

Figure 3. Estimation of the mucosal wave correlate: a) Levelled first derivative of the glottal source, b) Levelled glottal source, c) average acoustic waveform, d) mucosal wave correlate

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_{lb} in the frequency domain.

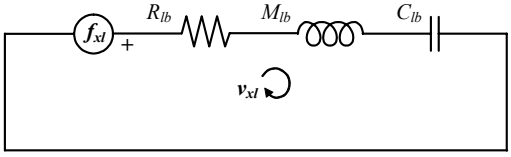

Figure 4. Electromechanical equivalent of a cord body

The relationship between velocity and force in the frequency domain is expressed as the cord body admittance. It will be assumed that the power spectral density of the levelled glottal aperture (1st derivative of the glottal source) is related to the square modulus of the body admittance $Y_{bl}(s)$ as

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

which shows a maximum value at

$$T_1 = T_{mb}(\omega = \omega_r) = \frac{G_b}{R_{lb}^2} \quad (3)$$

$$\omega_r = \sqrt{\frac{1}{M_{lb}C_{lb}}} \quad (4)$$

where G_b is a factor of scale between the average acoustic waveform power spectral density and the square modulus of the cord body admittance. The value of the third harmonic will be given by

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

From this expression the following estimate for the body mass may be obtained

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

$$r_{13} = \frac{T_1 - T_3}{T_1 T_3} \quad (7)$$

The value of C_{lb} could be derived from (4). The curve fitting after estimating the biomechanical parameters for a real trace (sample *00B*) is shown in Figure 5.

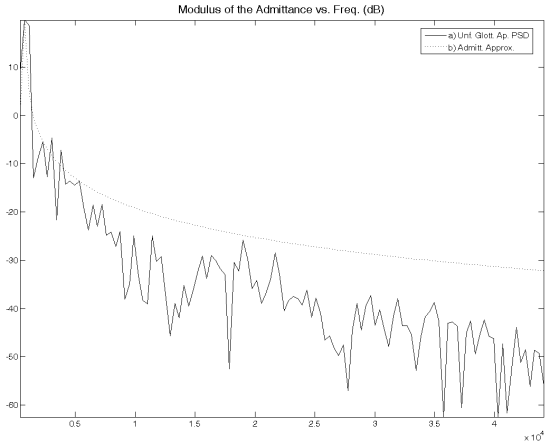

Figure 5. Parametric fitting of a specific average acoustic waveform for sample *00B* (full line) against the admittance approximation (dot line)

III. PARAMETER UNBALANCE

A slight unbalance between waveform cycles may be observed in Figure 3.a) and c). Even cycles appear to be larger than odd ones. As estimations of mass, stiffness and damping will be available on a cycle frame basis, the unbalance of these parameters (**BMU** – Body Mass Unbalance, **BLU** – Body Losses Unbalance and **BSU** – Body Stiffness Unbalance) may be defined as

$$m_{uk} = \left(\hat{M}_{bk} - \hat{M}_{bk-1}\right)/\left(\hat{M}_{bk} + \hat{M}_{bk-1}\right)$$
$$r_{uk} = \left(\hat{R}_{bk} - \hat{R}_{bk-1}\right)/\left(\hat{R}_{bk} + \hat{R}_{bk-1}\right) \quad (8)$$
$$c_{uk} = \left(\hat{C}_{bk} - \hat{C}_{bk-1}\right)/\left(\hat{C}_{bk} + \hat{C}_{bk-1}\right)$$

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

$$m_{dj} = \left(\overline{M}_{bj} - \overline{M}_{bs}\right)/\overline{M}_{bs}$$
$$r_{dj} = \left(\overline{R}_{bj} - \overline{R}_{bs}\right)/\overline{R}_{bs} \qquad (9)$$
$$c_{dj} = \left(\overline{C}_{bj} - \overline{C}_{bs}\right)/\overline{C}_{bs}$$

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

IV. Results and Discussion

The key tool in the classification into pathologic and normophonic samples used in this research is Principal Component Analysis (PCA), 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). A variant of Principal Component Analysis known as *multivariate measurements analysis* (see [6], pp. 429-30) has been used with the distortion parameters given in Table 1.

Table 1. List of parameters produced 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 (8)
$x_{38\text{-}40}$	biomechanical parameter unbalance (9)

This methodology has been applied to *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 2. 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

These samples were processed to extract the set of *40* parameters listed in Table 1, 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 6 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 3.

Table 2. Values of $x_{35\text{-}39}$ for the samples studied

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

Table 3. 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

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 7. 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).

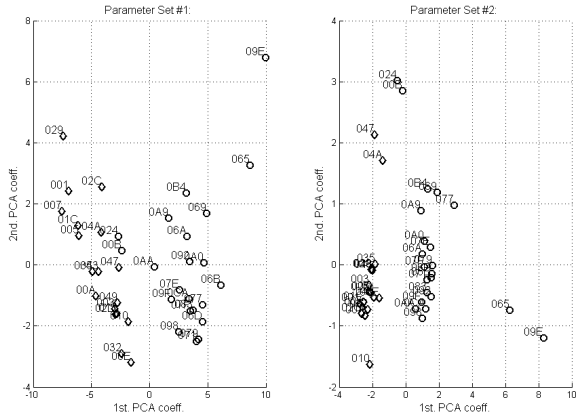

Figure 6. Left) Clusters for S_1. Right) Clusters for S_2.

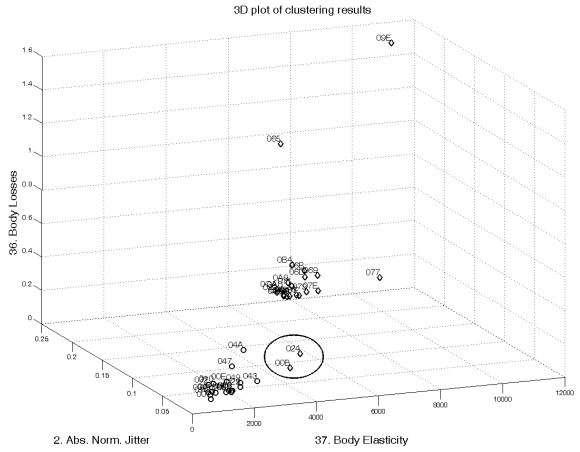

Figure 7. 3D Clustering Plot showing the separation in the manifold defined by the parameter subset {x37, x2 and x36} – ordered by relevance

The behaviour of cases *00B* and *024*, classified as pathological by PCA analysis deserves a brief comment. These appear in Figure 7 (encircled) not quite far from normal cases 001-04A, but showing a stiffness that doubles those of normophonic 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 2, being 1.25 and 1.2 respectively, or 225% and 220%.

V. CONCLUSIONS

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. Mild pathologies may appear as normophonic from subjective analysis. Adequately combining classical distortion parameters with deviation parameters renders fairly good results in pathology detection. These conclusions need to be confirmed by more experiments.

VI. ACKNOWLEDGMENTS

This research carried out under grant Nos. TIC2002-2273, TIC2003-08756 and TIC2003-08956-C02-00, from Programa de las Tecnologías de la Información y las Comunicaciones, Ministry of Education and Science, Spain.

VII. REFERENCES

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

[2] 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.

[3] 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, pp. 842-845.

[4] Gómez, P., Martínez, R., Díaz, F., Lázaro, C., Álvarez, A., Rodellar, V., Nieto, V., "Estimation of vocal cord biomechanical parameters by non-linear inverse filtering of voice", *Proc. of the 3rd Int. Conf. on Non-Linear Speech Processing NOLISP'05*, Barcelona, Spain, April 19-22 2005, pp. 174-183.

[5] Gómez, P., Godino, J. I., Álvarez, A., Martínez, R., Nieto, V., Rodellar, V., "Evidence of Glottal Source Spectral Features found in Vocal Fold Dynamics", *Proc. of the ICASSP'05*, 2005, pp. V.441-444.

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

[7] Kuo, J., Holmberg, E. B., Hillman, R. E., "Discriminating Speakers with Vocal Nodules Using Aerodynamic and Acoustic Features", *Proc. of the ICASSP'99*, 1999, pp. I.77-80.

[8] Titze, I., "Summary Statement", *Workshop on Acoustic Voice analysis, National Center for Voice and Speech*, 1994.

[9] The Voice Center of Eastern Virginia Med. School: http://www.voice-center.com/larynx_ca.html.

[10] 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.

[11] Titze, I. R., "The physics of small amplitude oscillation of the vocal folds", *J. Acoust. Soc. Am.*, Vol. 83, 1988, pp. 1436-1552.