GLOTTAL BIOMETRIC FEATURES: ARE PATHOLOGICAL VOICE STUDIES APPLIABLE TO VOICE BIOMETRY?

P. Gómez-Vilda, A. Álvarez-Marquina, L. M. Mazaira-Fernández, R. Fernández-Baíllo, V. Rodellar-Biarge, V. Nieto-Lluis.
[1] Grupo de Informática Aplicada al Procesado de Señal e Imagen
Facultad de Informática, Universidad Politécnica de Madrid
Campus de Montegancedo, s/n, 28660 Boadilla del Monte, Madrid, Spain
e-mail: pedro@pino.datsi.fi.upm.es

Abstract: **The purpose of the present paper is to introduce a methodology successfully used already in voice pathology detection for its possible adaptation to biometric speaker characterization as well. For such, the behavior of the same GMM classifiers used in the detection of pathology will be exploited. The work will show specific cases derived from running speech typically used in NIST contests against a Universal Background Model built from the population of normophonic subjects in specific vs general evaluation paradigms. Results are contrasted against a set of impostors derived from the same population of normophonic subjects. The relevance of the parameters used in the study will also be discussed.**

Keywords: **Speaker Characterization, Glottal Source, GRBAS, Voice Pathology Grading, Gaussian Mixture Models**

1. INTRODUCTION

The purpose of the present paper is to explore to which extent results in voice pathology detection and grading studies can be extended to give an accurate description of the speaker's voice biometry. In past studies our group has proposed new sets of parameters derived from the glottal component of voice which have been shown to be highly resolving in the detection and grading of pathology [4][9]. These may be grouped into three different classes:

- Glottal Source Spectral Profile Features (GSSPF), which are produced pinpointing the singularities of the Glottal Source Power Spectral Density (GSPSD), specifically the first two "V-grooves" resulting from anti-resonances in the vocal fold biomechanical behavior [3].

- Vocal Fold Biomechanical Parameter Descriptors, which result from the inversion of the electro-mechanical equivalents of the vocal folds when the power spectral profiles are fit to the transfer functions associated to the biomechanical parameters of the vocal folds [3][4].

- Glottal Phonation Cycle Features, which result from the parameterization of the time-domain behavior of the reconstructed Glottal Source. Open, Close and Return Quotients are among the most widely used ones [14], although others based on the vocal gap are introduced as well [4].

The importance of some of these parameters in voice pathology detection is more than evident. Some of the parameters showing better correlation to voice pathology are highly sensitive and mark the presence of pathology with high accuracy. The intention of the present work is to explore if these parameters or others alike may be applied as well to determine personality features or biometric markers of a speaker with a similar degree of accuracy.

The paper is divided in the following sections: an overview of the methodology used in voice pathology detection is given in section 2; section 3 is devoted to the formulation of this study for voice biometry; section 4 is intended to describe the materials and methods for voice biometrical differentiation in intra- and inter-speaker experiments; results are discussed in section 5, and finally conclusions drawn from the present study are presented.

2. VOICE PATHOLOGY DETECTION

Voice Pathology may be detected using different strategies, classically mel-cepstrum parameterization and GMM (Gaussian Mixture Models) classification [8]. Nevertheless the use of mel-cepstral coefficients on the whole voice signal, although efficient, lacks semantics, i.e., it is really difficult to infer which factors convey to successful detection, and from this point it seems really difficult to infer which are the clues to successful classification of pathologies, this being a major aim in the field far from being completed. A different approach is that one of biometric and biomechanical parameter

extraction based on the glottal excitation, which produces parameter sets directly related with spectral singularities or vocal fold parameters as dynamic masses or tensions. This approach has been used in the recent past yielding interesting results [9][10]. The combination of specific parameter cocktails may yield quite accurate results not only in voice pathology detection, but in estimating the degree of pathology as well, mimicking the objective estimation of GRBAS [6]. The methodology relies in the accurate determination of a set of individuals which may be considered "healthy" or "pathology-free" from examinations including electroglottogram and endoscopy of the vocal folds. This set of "normophonic" speakers is the key to the correct evaluation of pathology. Normophonic speakers need to be recruited for both genders, as morphologic differentiations between male and female are meaningful [15], and a normophonic male subject may appear as dysphonic if contrasted against a female database. From the inversion of the Liljencrants-Fant source-filter model the glottal source (excitation) is reconstructed [1]. Advanced parameterization techniques are used for the estimation of observation vectors, where each speaker *i* is represented by a parameter vector:

$$\mathbf{x}_i = [x_{i1}, x_{i2}, ...x_{iJ}] \quad (1)$$

composed of J values x_{ij} produced from a 200 msec. segment of voice corresponding to a sustained utterance of /a/ accordingly with the description given in [5].

Table 1. Description of the parameter set

Param.	Description
x_1	*pitch*
x_2	*jitter*
$x_{3\text{-}5}$	3 variants of *shimmer*
$x_{6\text{-}7}$	Glottal closure parameters
$x_{8\text{-}10}$	Harmonic-Noise and H_2-H_1 Ratios
$x_{11\text{-}14}$	4 first cepstral coefficients of the mucosal wave correlate power spectral density
$x_{15\text{-}23}$	Singularities of mucosal wave correlate power spectral density (amplitude)
$x_{24\text{-}32}$	Singularities of mucosal wave correlate power spectral density (frecuency)
$x_{33\text{-}34}$	Slenderness of the two first "V troughs"
$x_{35\text{-}37}$	Biomechanical parameters of vocal fold body (masses, losses, tensions)
$x_{38\text{-}40}$	Intra-speaker period-synchronous variations of body biomechanics
$x_{41\text{-}43}$	Biomechanical parameters of vocal fold cover (masses, losses, tensions)
$x_{44\text{-}46}$	Intra-speaker period-synchronous variations of cover biomechanics

The observations derived from a given speaker are not used as such, but transformed according to Principal Component Analysis procedures [5] (PCA projection). The reasons are two-fold: on one side the reduction of correlation among the observations improves the data inversion process and results in more stable GMM's; on the other side the dimensions of the vectors can be reduced, thus implying less computational expenses. Once the normophonic male (*m*) and female (*f*) sets are completed the model observation matrices are produced:

$$\begin{aligned} \mathbf{X}_{Mm} &= [\mathbf{x}_{1m}, ... \mathbf{x}_{im}, ... \mathbf{x}_{Im}]^T \\ \mathbf{X}_{Mf} &= [\mathbf{x}_{1f}, ... \mathbf{x}_{if}, ... \mathbf{x}_{If}]^T \end{aligned} \quad (2)$$

Similarly the control observation matrices $\mathbf{X}_{Cm}$ and $\mathbf{X}_{Cf}$ are produced using observations from the dysphonic male and female sets. The PCA projection is based on the joint model-control covariance matrix [12]:

$$\begin{aligned} \mathbf{X}_P &= [\mathbf{X}_{Mm,f}^T, \mathbf{X}_{Cm,f}^T]^T \\ \mathbf{C}_P &= \mathbf{X}_P \mathbf{X}_P^T \end{aligned} \quad (3)$$

The matrix ($\mathbf{E}_P$) of eigenvalues of $\mathbf{C}_P$ is used to project the original observations matrices on the new principal component matrices:

$$\begin{aligned} \mathbf{Y}_m &= \mathbf{X}_m \mathbf{E}_P \\ \mathbf{Y}_f &= \mathbf{X}_d \mathbf{E}_p \end{aligned} \quad (4)$$

Once the enrolment of enough normophonic individuals of both genders is available, a GMM for each gender set is produced (Γ_m for the male set and Γ_f for the female one). For such the mean vectors $\boldsymbol{\psi}_{Mm}$ and $\boldsymbol{\psi}_{Mf}$ as well as the corresponding covariance matrices $\mathbf{C}_{Mm}$ and $\mathbf{C}_{Mf}$ are estimated. The GMM is defined by a set of Gaussian multivariate functions of the kind:

$$p(\mathbf{y}_{ti} / \Gamma_{Mm,f}) = \frac{1}{(2\pi)^{Q_m/2} \left|\mathbf{C}_{Mm,f}\right|^{1/2}} e^{-1/2(\mathbf{y}_{ti}-\boldsymbol{\psi}_{Mm,f})^T \mathbf{C}_{Mm,f}^{-1}(\mathbf{y}_{ti}-\boldsymbol{\psi}_{Mm,f})} \quad (5)$$

$\mathbf{y}_{ti}$, $\boldsymbol{\psi}_n$, and $\mathbf{C}_n$ being respectively the data vector under test of subject *i*, the centroids of the parameter Gaussians GMM's and the Covariance Matrices of each observation set, *p* being the conditional probability of an observation vector being a member of the specific set represented by the specific Gaussian. As a generalization, if the normophonic GMM is composed by a certain number of Gaussians the joint probability will be expressed as:

$$p_T(\mathbf{y}_{ti} / \Gamma_{Nm,f}) = \sum_k w_k p_k(\mathbf{y}_{ti} / \Gamma_{km,f}) \quad (6)$$

where w_k are the weights of the linear combination generating the overall probability. In the present case mono-Gaussian Models show to be accurate enough. Finally the issue of voice pathology detection may be stated in terms of a score usually given as a Log-Likelihood Ratio (LLR) of the odds:

$$\begin{aligned} &\Lambda_p(\mathbf{y}_{tmi}) = \\ &log\{p(\mathbf{y}_{tmi} / \Gamma_{nm})\} - log\{p(\mathbf{y}_{tmi} / \Gamma_{\bar{n}m})\} \end{aligned} \quad (7)$$

This score is based on distance metrics as shown in Figure 1, and it may be used for detecting the pathological condition of the subject using classical ROC-DET (Receiver Operator Characteristics or Detection Error Trade-Off) plots. Depending if the LLR is over or under a given threshold θ ($\Lambda_p(\boldsymbol{y}_{tmi}/\Gamma_{nm})>\theta$ or $\Lambda_p(\boldsymbol{y}_{tfi}/\Gamma_{nf})<\theta$) the voice of the subject under test is considered normal or dysphonic.

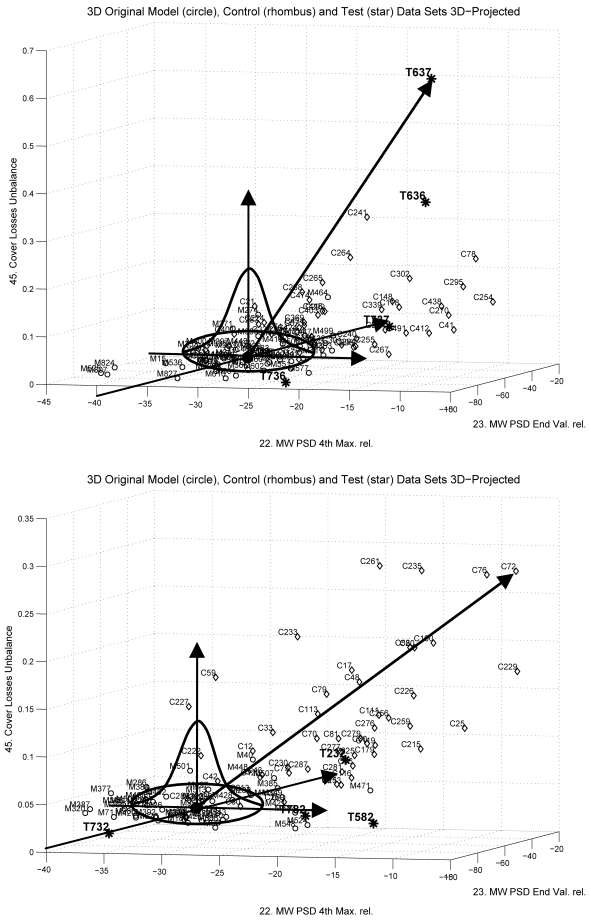

Figure 1. Top: Male cluster set with joint normophonic and pathologic distributions. The distance to the normal distribution may serve as a measure of the pathology grade. Bottom: Idem for the female cluster set.

The figures give an idealized idea on how each respective GMM quantifies the membership probability of each subject relative to its respective model set (Universal Background Model) plotted on the three parameters with largest FDR's. The normalized distance of each subject to the respective model centroid is used as a voice quality evaluation factor (grade) for each individual (g_i) [9]. This distance is marked by arrows for each set farthest cases.

3. Application to Voice Biometry

The main problem in applying the above conclusions to voice biometrical studies is the intra-speaker variability. In other words: to which extent the parameters obtained for a given speaker under a given phonation modality are similar to the speaker's other phonation modalities and distinct at the same time to the parameters obtained from other speakers' phonations? To answer this crucial question one has to take into account the sources of intra- and inter-speaker variability. For intra-speaker studies these may be the main sources of variability:

- The modality of the phonation, this being normal (modal), over-pressed or under-pressed. The modal phonation is considered to be associated with the relaxed (emotion-less) speaker, whilst the over-pressed corresponds to emotional excitation (anger, exultation, wrath...), and the under-pressed has to see with anguish, fatigue, depression, etc. Thus modality is very much related to the speaker's emotional state.

- Vocalization. Usually the decomposition of the voice under the source-filter model into the glottal source and vocal tract transfer function is highly dependent on this last pattern. Therefore the results will be different for open than for close vowels, and for voicing consonants. This characteristic has to see with articulation or acoustic-phonetic issues.

- Prosody. The stress and emphasis of the phonation in running speech is of most importance. Raising or lowering the pitch reduces or adds duration to the glottal phonation cycle, and consequently to the resulting parameterization. This situation is similar to the study of voice in singing, as prosody may be considered as the "music" of running speech. The raising or lowering of pitch in speech can produce quite different results in the parameter description of the glottal source in interrogative, declarative or imperative sentences.

With all this information in mind examples will be given from voice samples corresponding to different articulation and prosody cases, and consequences will be drawn regarding their use in speaker recognition studies. The relevance of the speaker's emotional state will be left for further elaboration.

The study will be conducted in terms of the well-known Prosecutor's vs Defender's approach as a classical Log-Likelihood Ratio (LLR) estimation by the specificity-typicality two-stage paradigm [11]:

$$p(I_u / I_a) = \frac{p(I_a / I_u)\,p(I_u)}{p(I_a)} \qquad (8)$$

where I is in general de information available from a specific speaker (I_u from the questioned or unasserted speaker, I_a from the asserted or suspect). The above probabilistic model will be formalized by the classical LLR evaluating the Prosecutor's Hypothesis (H_p) against the Defender's Hypothesis (H_d):

$$\Lambda_{u/a} = \log\left\{\frac{f(E/H_p, I)}{f(E/H_d, I)}\right\} = \\ = \log\left\{\frac{p(I_u/I_a)}{p(I_u)}\right\} = \log\left\{\frac{p(I_a/I_u)}{p(I_a)}\right\} = \\ = \log\left\{p(\mathbf{x}_i^u/\Gamma_A)\right\} - \log\left\{p(\mathbf{x}_i^u/\Gamma_B)\right\} \quad (9)$$

E, H_p and H_d being respectively the Evidence, the Prosecutor and the Defender Hypotheses. The general speaker information, composed by the set of observations (parameter medians of the set of parameters in Table 1) from the asserted or suspect (a) and the unasserted or questioned (u) observations are defined as:

$$\mathbf{x}_i^a = \left[x_{1i}^a, x_{2i}^a, \ldots x_{mi}^a\right]^T \\ \mathbf{x}_i^u = \left[x_{1i}^u, x_{2i}^u, \ldots x_{mi}^u\right]^T \quad (10)$$

The Universal Background Gaussian Model (UBGM) Γ_B will be composed by the covariance matrix $\mathbf{C}_B$, and mean vector ψ_B for the reference population data set. The Asserted Gaussian Model Γ_A is to be built in a similar way from all the data available from the suspect, resulting in $\mathbf{C}_A$, and ψ_A. The evaluation of the membership of a given questioned frame $\mathbf{y}_i^u$ with respect to the UBMG or the AGM will be estimated in terms of the conditioned probability:

$$p(\mathbf{y}_i^u / \Gamma_{A,B}) = \\ \frac{1}{(2\pi)^{Q_{A,B}/2}\left|\mathbf{C}_{A,B}\right|^{1/2}} e^{-1/2\left(\mathbf{y}_i^u - \boldsymbol{\psi}_{A,B}\right)\mathbf{C}_M^{-1}\left(\mathbf{y}_i^u - \boldsymbol{\psi}_{A,B}\right)^T} \quad (11)$$

Once the relative membership probabilities are produced, the LLR of the Prosecutor's vs the Defender's Hypothesis given in (9) will be estimated. Results for a practical study case will illustrate this technique in the next section.

4. Results and Discussion

For the purposes of the present study a set of 30 normophonic male speakers from [7] will be used in the experiments. Part of this subset, specifically 20 speakers will serve as the Universal Background Model Set, and 10 speakers more will be used as imposters for T-norm contrast. The questioned and suspect frames will be obtained from a 300-sec. recording of running speech (test 4, channel a) from the last NIST SRE10 HARS1 competition [13] selecting 12 frames where the utterance /ah/ or /uh/ have been produced, either in long vowels or in fillings, these frames being given in Table 2.

Table 2. Frames under study

Frame start	Frame end	Frame #
9.2	9.4	4009
28.7	28.9	4028
43.95	44.20	4043
201.55	201.75	4201
213.85	214.00	4213
232.30	323.55	4232
243.55	243.85	4243
248.80	249.35	4248
267.00	267.35	4267
276.00	276.20	4276
289.95	290.25	4289
291.30	291.60	4291

The whole set of 20+10+12 frames taken at 8kHz are parameterized and PCA projected. The corresponding Model, Control and Test sets are described in Table 3.

Table 3. Model, Control and Test Sets used in the experiments

Set	Frames
Model	15 271 274 314 333 334 335 347 353 361 362 363 366 368 372 383 397 399 400 406
Control	4009 4028 4043 4201 4213 4232 4243 4248 4289 4291
Test	408 416 417 419 422 427 429 432 443 464 4267 4276

It may be seen that the PCA projection will be carried out on the Model and Control Sets, the first constructed exclusively from 20 frames of different normophonic speakers. The Control Set is integrated by 10 frames from the same speaker. The Test set includes 10 frames from different normophonic speakers and 2 more frames from the questioned speaker.

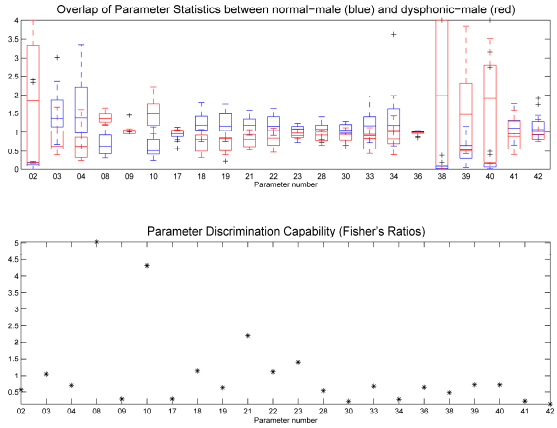

Figure 2. Top: Parameter Distribution comparisons of the Model Set (blue) and Control Set (red). Bottom: Values of Fisher's Discriminant Ratios for the same parameters.

In this way the objective is twofold: on one side to determine if the samples taken at different time instants from the same speaker present some similarity among themselves, on the other side to determine if they can be differentiated from a Universal Background Model in

terms of possible dysphonia reflected in certain parameters. As a side objective, the parameters reflecting the dysphonic condition are to be determined. This last objective is achieved using Fisher's Discriminant Ratios (FDR), as given by:

$$fdr_j = \frac{(\mu_{Mj} - \mu_{Cj})^2}{\sigma_{Mj}^2 + \sigma_{Cj}^2}; \quad 1 \le j \le J \qquad (12)$$

the resulting estimations are given in Figure 2. The set of parameters being investigated have been selected among the most resolving ones, although not all of them are sensitive enough. The three most resolving are the HNR (x_8), the f_{H1}-f_{H2} (x_{10}) and the 2nd minimum in the GSPSD (Glottal Source Power Spectral Density: x_{21}). It is interesting to see that shimmers ($x_{3\text{-}4}$) are more resolving than jitter (x_2), and that the spectral singularities of the GSPSD ($x_{18\text{-}23}$) show also important discriminating capabilities in this case. Thus a good balance between classical distortion parameters and biometrical ones is expected to enhance discrimination results. The Model (o), Control (◊) and Test (*) Sets given by matrices $\mathbf{X}_M$, $\mathbf{X}_C$ and $\mathbf{X}_T$ in Table 3 are shown in Figure 3.

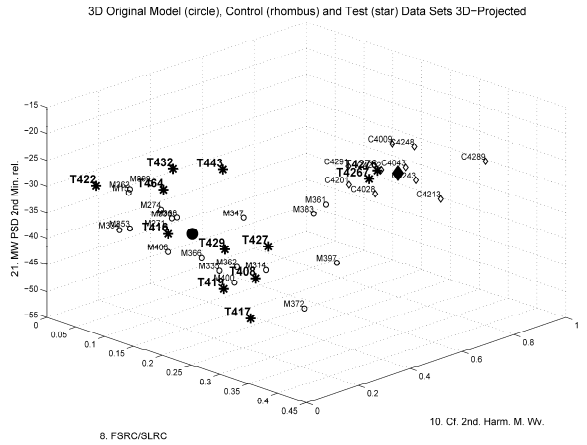

Figure 3. 3D Projection of the Data Set used in the experiments: Model (o), Control (◊) and Test (*) Sets. Centroids of the Model and Test sets are given by a filled circle and rhombus.

The projection is given in terms of the three most resolving parameters accordingly to the values of FDR from (12). It may be seen that the Model Set frames (labeled as Mxxx and o) are located altogether around a (more or less) well defined cluster (except for M361 and M383). It may be seen also that the Test Set frames (labeled as Txxx and *) corresponding to normophonic subjects are grouped themselves in the neighborhood of the Model Set. Clearly the Control Set frames (labeled Cxxxx and ◊) are grouped apart mixed with the two Test frames taken from the questioned speaker (T4267 and T4276). This points out to the questioned frames as being produced also by the suspect (evidence would favor H_p in detriment of H_d). This is more clearly expressed by the data given in Table 4.

Table 4. Results from the detection process. Rec#: Number of the frame record. Λ_p: Log Likelihood Ratio referred to Dysphonia. G: Grade of Dysphonia. sDo: Square of Norm. Distance to the Model Set Centroid. sD◊: Id. to the Control Set. $p(y/\Gamma_B)$: Probability of membership to the Model Set. $p(y/\Gamma_A)$: Id. to the Control Set. $\Lambda_{u/a}$: Likelihood Ratio referred to the Prosecutor's Hypothesis vs the Defense Hypothesis.

Rec#	Λ_p	G	sDo	sD◊	$p(y/\Gamma_B)$	$p(y/\Gamma_A)$	$\Lambda_{u/a}$
408	-11,17	0	2,26	40,69	4,33E-07	1,93E-13	-14,62
416	-10,96	0	3,44	33,15	2,40E-07	8,36E-12	-10,26
417	-10,62	0	21,68	416,24	2,63E-11	5,45E-95	-192,69
419	-12,10	0	8,33	211,35	2,08E-08	1,69E-50	-96,92
422	-11,66	0	18,98	63,07	1,01E-10	2,66E-18	-17,45
427	-11,94	0	2,72	50,10	3,43E-07	1,75E-15	-19,09
429	-11,06	0	6,46	167,51	5,30E-08	5,58E-41	-75,94
432	-10,32	0	7,58	102,44	3,03E-08	7,55E-27	-42,84
443	-10,94	0	12,22	37,38	2,98E-09	1,01E-12	-7,99
464	-9,72	0	5,61	51,86	8,13E-08	7,26E-16	-18,53
4267	-11,79	0	23,52	15,03	1,04E-11	7,20E-08	**8,84**
4276	-13,22	0	35,00	26,71	3,36E-14	2,09E-10	**8,74**

As seen in the table, for each frame in the Test Set (first column to the left) the pathologic LLR in (7) is given. It may be seen in the second column that accordingly to the data available the two suspect/questioned frames can not be considered pathological, this fact being reinforced by the objective grade G (third column), which results null, this corresponding with no dysphonia. The fourth column gives the squared Mahalanobis distance from each sample to the Model Centroid. The two more distant frames are the ones extracted from the questioned speech segment. The fifth column gives the same figure for each sample relative to the Test Centroid estimated from frames extracted from the suspect speech segment (which happen to be the same than the questioned one in the present experiment). The two closer frames to the Test Centroid are now the ones extracted from the questioned speech segment (T4267 and T4276). The next two columns give the relative membership probabilities of each Test frame to both the Control and Model Sets. The membership probability of the upper ten frames relative to the UBGM is clearly larger than their respective membership probability relative to the AGM. This results in negative LLR's favoring the H_d. On the contrary, the last two frames show membership probabilites larger for the AGM than for the UBGM. The respective LLR's are positive and similar, favoring the H_p in their case.

5. CONCLUSIONS

Interesting consequences may be derived from the present study. First of all it seems that parameters classically derived for the study of voice pathology as *shimmer*, *HNR* or H_2-H_1 can be used for the biometrical characterization of the speaker as well. This conclusion is very important, as these parameters have clear semantics as far as the characterization of a speaker is concerned. A second conclusion is that the Glottal Source Power Spectral Distribution (GSPSD), and especially its singularities (peaks and troughs) are relevant for the biometrical characterization of the speaker. It is known that the Glottal Source may be altered by articulation as well as by modality, vocalization or prosody, as explained in section 3. The frames selected from the running speech segment were not especially conditioned by any factor except by vowel coloring (in fact most of them correspond to the kind of fillers /uh/'s and /ah/'s, which are spontaneously produced by Native Speakers of English). The modality is different in most of them, as well as the prosody (some present questioning or surprise marks). Nevertheless, the system identified clearly all of them as being different from the Model Set, selected from sustained vowels, and similar among themselves. This fact may indicate that the parameters selected are robust to modal information and sensitive to biometrical differences. Of course the work is still far from being completed. Massive tests on model and test running speech segments as the ones proposed in the last NIST SRE HARS2 contest need to be processed. For such automatic vowel selection and framing is to be put into work and the discussed methodology applied on blind tests to measure its capability in speaker characterization tasks.

ACKNOWLEDGMENTS

This work has been funded by grants TIC2003-08756, TEC2006-12887-C02-01/02 and TEC2009-14123-C04-03 from Plan Nacional de I+D+i, Ministry of Science and Technology, by grant CCG06-UPM/TIC-0028 from CAM/UPM, and by project HESPERIA (http.//www.proyecto-hesperia.org) from the Programme CENIT, Centro para el Desarrollo Tecnológico Industrial, Ministry of Industry, Spain.

REFERENCES

[1] Alku, P., "Parameterisation Methods of the Glottal Flow Estimated by Inverse Filtering", Proc. of VOQUAL'03, Geneva, 2003, pp. 81-87.

[2] Baken, R. J. and Orlikoff, R., *Clinical measurement of speech and voice*, 2 ed., Singular Pub. Group, 2000.

[3] Gómez. P. et al., "Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters", *J. Voice*, Vol. 21, No. 4, 2007, pp. 450-476.

[4] Gómez. P. et al., "Glottal Source Biometrical Signature for Voice Pathology Detection", *Speech Communication*, Vol. 51, 2009, pp. 759-781.

[5] Gómez. P. et al., "PCA of Perturbation Parameters in Voice Pathology Detection", Proc. of INTERSPEECH'05, 2005, pp. 645-648.

[6] M. Hirano et al., "Acoustic analysis of pathological voice. Some results of clinical application", Acta Otolaryngologica, Vol. 105 (5-6), 1988, pp. 432-438.

[7] Project MAPACI: http://www.mapaci.com.

[8] Fraile, R., Sáenz, N., Godino, J. I., Osma, V., Fredouille, C., "Automatic Detection of Laryngeal Pathologies in Records of Sustained Vowels by Means of Mel-Frequency Cepstral Coefficient Parameters and Differentiation of Patients by Sex", Folia Phoniatrica et Logopaedica, Vol. 61, 2009, pp. 146-152.

[9] Gómez, P., "Voice Pathology Grading by Gaussian Mixture Models: Study Cases", Proc. of MAVEBA09, Firenze, Italy, December 2009, pp. 45-48.

[10] Godino, J. I., Osma, V., Sáenz, N., Gómez, P., Blanco, M., Cruz, F., "The Effectiveness of the Glottal to Noise Excitation Ratio for the Screening of Voice Disorders", J. of Voice, Vol. 24, No.1, 2010, pp. 47-56.

[11] González, J., et al., "Emulating DNA: Rigorous Quantification of Evidential Weight in Transparent and Testable Forensic Speaker Recognition, IEEE Trans. on ASLP, Vol. 15, No. 7, 2007, pp. 2104-2115.

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

[13] http://www.itl.nist.gov/iad/mig//tests/sre/2010/index.html, NIST SRE10 Evaluation Workshop, 24-25 June 2010, Brno, Czech Republic.

[14] Orr, R. , Cranen, B., de Jong, F. I. C. R. S., "An investigation of the parameters derived from the inverse filtering of flow and microphone signals", Proc. VOQUAL'03, 2003, pp. 35–40.

[15] Gómez. P. et al., "Detecting Pathology in the Glottal Spectral Signature of Female Voice", Proc. of MAVEBA07, Firenze, Italy, December 2007, pp. 183-186.