GLOTTAL-SOURCE SPECTRAL BIOMETRY FOR VOICE CHARACTERIZATION

P. Gómez, R. Fernández, A. Álvarez, L. M. Mazaira, V. Rodellar, R. Martínez, C. Muñoz

Grupo de Informática Aplicada al Tratamiento de Señal e Imagen, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28660 Madrid
phone: + (34) 913367384, fax: + (34) 913366601, e-mail: pedro@pino.datsi.fi.upm.es
web: www.mapaci.com

ABSTRACT

The biometric signature derived from the estimation of the power spectral density singularities of a speaker's glottal source is described in the present work. This consists in the collection of peak-trough profiles found in the spectral density, as related to the biomechanics of the vocal folds. Samples of parameter estimations from a set of 100 normophonic (pathology-free) speakers are produced. Mapping the set of speaker's samples to a manifold defined by Principal Component Analysis and clustering them by k-means in terms of the most relevant principal components shows the separation of speakers by gender. This means that the proposed signature conveys relevant speaker's meta-information, which may be useful in security and forensic applications for which contextual side information is considered relevant.

1. INTRODUCTION

Speech Processing Technologies have developed powerful tools to extract information present in voice and speech under different appearances. This is what Nickel in a recent overview [12] defines as "textual contents" and "contextual side information". This last concept includes meta-information present on the speaker's voice, as gender, age, emotional state, voice conditions (healthy or pathological), language and dialectal issues, function of prosody and intonation in the message, etc. Classically the methodology used to characterize these concepts is based in the estimation of certain parameters which are assumed to contain a description of general voice characteristics. These parameters are used in the training of classification engines, amenable of separating speakers in clusters accordingly to this contextual side information. This is so in speaker's identification and verification tasks (SIV) [14] or in voice pathology detection (VPD) [5], where MFCC's are used as classical parameter templates, and Gaussian Mixture Models (GMM's) or Support Vector Machines (SVM's) are the usual classification engines [15]. The efficiency of both parameterization and classification approaches is well proven, although Detection Error Tradeoff rates have reached a certain lower limit which is difficult to overcome. To find alternative techniques to reduce DET's, a possible approach is to analyze the textual and contextual nature of voice, separating both profiles, using the most suitable classification engines for both flows of information, and fusing the results as is customary in SIV tasks [14]. The analytical technique presented here splits unvoicing and voicing frames. The spectral information present in these last ones would be separated into vocal tract spectral contents (mainly of textual character: related with the articulatory features of the message) and glottal source spectral contents (mainly contextual: gesture, tension, stress, pathology, intonation, etc.). The use of the glottal source in speaker's identification tasks was proposed by Plumpe [13] to improve DET rates over classical methods using the temporal features of the glottal source. That technique presented several inconveniences derived from the relative separation of real glottal source temporal patterns (open and close quotients, etc.) from those of the ideal Liljencrants-Fant paradigm [4]. The aim of the present study is to devise a biometric voice signature based on the spectral characteristics of the glottal source [9], instead on its temporal counterparts. The power spectral distribution of the glottal source is estimated and the relevant spectral parameters on its power spectral density are used as biometric descriptors. This distribution is separated into two components: the one contributed by the vocal fold body, and the one due to the vocal fold cover. Through the present study the vocal fold cover biomechanical signature will be used to derive the speaker's biometric characterization. The Glottal Source Biometric Signature will be defined in section 2. Section 3 will be devoted to present the detection of gender using this signature among a database of 100 speakers equally balanced. Section 4 will present and discuss the detection results and Section 5 will summarize the main conclusions.

2. GLOTTAL SOURCE BIOMETRIC SIGNATURE

The proposed methodology for the estimation of the Glottal Signature is based on the derivation of the glottal source, once the vocal tract influence is removed from the voice trace by inverse filtering by means of an iterative implementation of Alku's method [1], [2], [8] and in the separation of the glottal source in its body and cover components under a phonation cycle basis as described in detail in [6]. A typical power spectral density profile of the cover component is shown in Figure 1. The following features may be appreciated: a rapid rise from low frequencies to a first amplitude maximum given by T_{M1} centered at a frequency f_{M1}, followed by a minimum T_{m1} at f_{m1} and a new maximum T_{M2} at f_{M2}. Several of these "V" profiles may be found, the spectral tendency showing a decay of *1/f* (dot line). These "V" grooves or troughs are present in all normal speakers examined. They

are almost unnoticeable in cases with over-tense voice, generally associated to the presence of certain pathological conditions of the vocal fold.

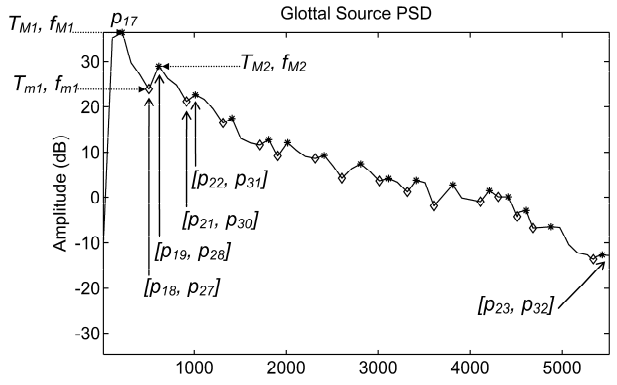

Figure 1. Power spectral density plot of the mucosal wave correlate for speaker 14E showing the vertices of the "V" profile $\{T_{M1}, f_{M1}\}$, $\{T_{m1}, f_{m1}\}$ and $\{T_{M2}, f_{M2}\}$, and the meaning of the 10 most relevant singularity parameters used in the study. Relative amplitude is given in dB. Horizontal axis given in Hz.

Where the parameter pairs [p_{18}, p_{27}], [p_{19}, p_{28}], [p_{21}, p_{30}], [p_{22}, p_{32}] and [p_{23}, p_{32}] encode the relative amplitude in dB and frequency position of each singularity as defined in 1)a)i)(4). The numbering of the parameters has to see with their relative position in a wider parameter list including distortion and biomechanical parameters as well as described in [7]. The ultimate reasons explaining this spectral behaviour are to be found in vocal fold biomechanics [8]. Therefore a "glottal signature" may rely on each "V" profile as shown in [9], measuring the amplitude and position of each minimum and the two maxima on its sides. Another related parameter is trough slenderness, which is the ratio between the width and depth of the trough. The glottal signature may be estimated by the relative amplitude differences for each singularity (maxima and minima) with respect to the first maximum found. For such, the normalized power spectral density of the glottal source or the mucosal wave correlate (after subtracting the glottal average wave) is used to obtain the positions of envelope singularities as follows:

- Pitch-synchronous frames of the mucosal wave correlate power spectral density in dB are FFT-estimated
- Each frame spectral envelope is obtained
- The envelope maxima (*) and minima (◊) in amplitude and frequency are estimated as ordered pairs with order index k: $\{T_{Mk}, f_{Mk}\}$ and $\{T_{mk}, f_{mk}\}$.
- The largest of all maxima $\{T_{Mm}, f_{Mm}\}$ is used as a normalization reference both in amplitude and in frequency for maxima and minima as given by:

$$\left.\begin{aligned}\tau_{Mk} &= T_{Mk} - T_{Mm}\\ \tau_{mk} &= T_{mk} - T_{Mm}\end{aligned}\right\};\quad 1 \le k \le K \qquad (1)$$

$$\left.\begin{aligned}\varphi_{Mk} &= \frac{f_{Mk}}{f_{Mm}}\\ \varphi_{mk} &= \frac{f_{mk}}{f_{Mm}}\end{aligned}\right\};\quad 1 \le k \le K \qquad (2)$$

- The slenderness factor of each minimum is defined as follows:

$$\sigma_{mk} = \frac{f_{Mm}\left(2T_{mk} - T_{Mk+1} - T_{Mk}\right)}{2\left(f_{Mk+1} - f_{Mk}\right)};\quad 1 \le k \le K \qquad (3)$$

The definition of the glottal signature is based on these predefined parameters as:

$$\left.\begin{aligned}&p_{17} = T_{M1}; p_{18} = \tau_{m1}; p_{19} = \tau_{M2}; p_{21} = \tau_{m2};\\ &p_{22} = \tau_{M3}; p_{27} = \varphi_{m1}; p_{28} = \varphi_{M2}; p_{30} = \varphi_{m2};\\ &p_{31} = \varphi_{M3}; p_{32} = \tau_{Nf}; p_{33} = \sigma_{m1}; p_{34} = \sigma_{m2};\end{aligned}\right\} \qquad (4)$$

where the difference in amplitudes between the largest local maximum and the end frequency limit has been also used to define p_{23} and p_{32}. Figure 2 shows the intra-speaker variability of the first nine ordered pairs $\{T_{Mk}, f_{Mk}\}$ and $\{T_{mk}, f_{mk}\}$ for typical male and female subjects. Note the normalization properties of this parameterization, both in amplitude and frequency.

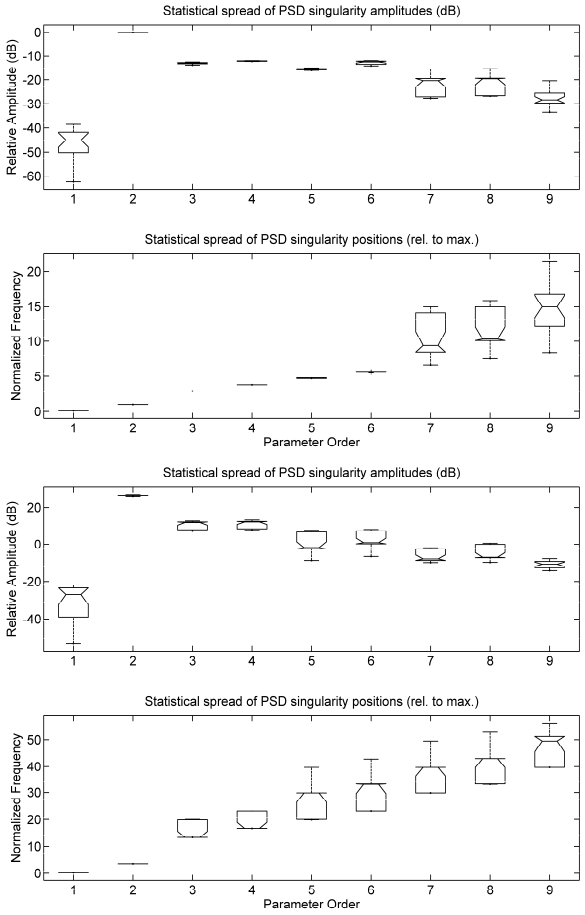

Figure 2. Intra-speaker variability of male (label #185: upper two templates) and female (label #158: lower two templates) for the first nine spectral singularities. Frequency positions are normalized to the first peak f_{M1}.

It may be seen that singularities in this specific male case appear less dispersed than in the female case. Nevertheless this need not be so in all cases, as one can find also pretty stable female signatures. Peaks and troughs show similar average positions in the female than in the male case as well. These differences are mainly due to voice quality features in both specific subjects. This particular behaviour of the glottal signature between gender scan not be generalized to all cases, as it depends on subject-specific factors, voice organic

or functional pathology among them. What is generalizable for both genders is that female singularities appear at higher values in frequency (normalized).

3. MATERIALS AND METHODS

A set of 100 speakers equally distributed by gender was randomly recruited from a wider database. Speaker ages ranged from *19* to *39*, with an average of *26.77* years and a standard deviation of *5.75* years. The normal phonation condition of speakers was determined by electroglottographic, videoendoscopic and GRBAS [10] evaluations. The recordings consisted in three utterances of the vowel */a/* produced in different sessions of about *3 sec* per record. A *0.2 sec* frame from the record centre was used in the estimations. The spectral profile parameters $\{p_{18\text{-}34}\}$ as well as the body and cover biomechanical parameters $\{p_{35\text{-}46}\}$ were obtained for each speaker [9]. Parameters $\{p_{1\text{-}14}\}$, corresponding to distortion measurements (*pitch*, *jitter*, *shimmer*, *NHR*, etc.) [5] were also estimated, although not used in the experiments presented. As each parameter was estimated on a phonation cycle basis (pitch-synchronous), for a typical male voice (with pitch around *100 Hz*) an average of $M=20$ estimations was obtained, which should be around $M=40$ for female voice (with a typical pitch of *200 Hz*). In this way up to $J=46$ observation parameters x_{ij} with $1 \le j \le J$ for each speaker $1 \le i \le I$ in the set of $I=100$ speakers were evaluated as the average of each observation parameter p_{im} over $1 \le m \le M$ phonation cycles:

$$x_{ij} = \frac{1}{M} \sum_{m=1}^{M} p_{im} \qquad (5)$$

These values are stored in a column vector:

$$\boldsymbol{x}_j = \left[x_{1j}, x_{2j} \ldots x_{ij}, \ldots x_{Ij} \right]^T \qquad (6)$$

containing the estimations of parameter j for all the speakers $1 \le i \le I$ in the set. The estimations for the whole set of speakers will be organized as a matrix of observations:

$$\boldsymbol{X} = \left[\boldsymbol{x}_1, \ldots \boldsymbol{x}_j, \ldots \boldsymbol{x}_J \right] \qquad (7)$$

Principal Component Analysis will be applied to this dataset as described in [11]. The set of eigenvalues and eigenvectors $\{\lambda i, ei\}$ of the covariance matrix $\boldsymbol{C}$ of $\boldsymbol{X}$ will be evaluated. The set of parameters is re-calculated in terms of principal components as:

$$\boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{e}_j; \quad 1 \le j \le J \qquad (8)$$

Column vectors $\boldsymbol{y}_j$ contain the new parameters (principal components) for each speaker in the list $1 \le i \le I$, ordered by their variance diminishing with the component order, according to their the respective eigenvalues $\{\lambda_i\}$, provided that $\lambda_i \ge \lambda_{i+1}$. This means that after a certain point, let's suppose it be $j=p \ll J$, the residual variance contained in the remaining components can be considered negligible, which allows truncating the component set to the first p column vectors, thus reducing the size of the data set substantially. Another important application of PCA may come through what would be called "reverse annotation". As the first components are ordered by relevance and these are linear combinations of the original parameter set, the structure of the first (most relevant) component for speaker i:

$$y_{i1} = x_{i1}e_{11} + x_{i2}e_{21} + \cdots + x_{ij}e_{j1} + \cdots + x_{iJ}e_{J1} \qquad (9)$$

is a linear combination of the original parameter row vectors weighted by the components of the first (column) eigenvector $\boldsymbol{e}_1$. Therefore, the relative contribution of each original parameter to this first component will depend on the relative values of the elements of this eigenvector, the ones with largest absolute value contributing more than those with lowest absolute value. This same consideration would be extensible to the reduced set of the first p components, having in mind that their relevance is also graded by their respective latency (absolute value of their eigenvalues). Therefore, a sorting function to grade the relative contribution of the original parameters in the new data set in cases where the first eigenvalue is substantially larger than the rest ($p=1$) is expressed by the absolute value of the first column eigenvector elements. The methodology implemented in the present study used the original parameters recorded for the set of speakers as defined in (1)-(4) on which PCA was applied as described in [7]. The following steps were covered:

1. Pre-selection of a database $\boldsymbol{X}_{17\text{-}34}$ comprising the original parameter set $\boldsymbol{S}_0 = \{x_{17\text{-}34}\}$ from the set of selected speakers.
2. From this subset, x_{20} and x_{29} were not used as they eventually correlated with x_{19} and x_{28} for all the speakers.
3. The resulting subset of parameters $\boldsymbol{S}_1 = \{x_{17}, x_{18}, x_{19}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{30}, x_{31}, x_{32}, x_{33}, x_{34}\}$ included the normalized version of the singularities on the power spectral density of the mucosal wave correlate, and the slenderness factors, as given in (3).
4. Separate the resulting database $\boldsymbol{X}(\boldsymbol{S}_1)$ set into two clusters by k-means blindly (with no "a priori" information).
5. Apply PCA on $\boldsymbol{X}(\boldsymbol{S}_1)$ to transform it on a new manifold for $p=16$ principal components producing a matrix $\boldsymbol{Y}_{1\text{-}16}$ ordered by the relevance of its principal components.
6. Select the three first components for the 3-D presentation of results in the principal component subspace.
7. Evaluate the relevance of the original parameters by reverse annotation, as in Table 1. Select the original parameters with highest relevance corresponding to $\boldsymbol{S}_2 = \{x_{28}, x_{30}, x_{19}, x_{31}, x_{18}, x_{27}, x_{22}, x_{21}\}$.
8. Select the 3 most relevant original parameters from $\boldsymbol{S}_2$ corresponding to $S_3 = \{x_{30}, x_{28}, x_{19}\}$. This subset may be used for 3-D presentation of results in the original parameter subspace.

Table 1. Relevance of singularity parameters from PCA (1st eigenvector)

Parameter order and name	Latency
28. MW PSD 2nd Max. Pos. rel.	*0.3446*
30. MW PSD 2nd Min. Pos. rel.	*0.3408*
19. MW PSD 2nd Max. rel.	*0.3359*
31. MW PSD 4th Max. Pos. rel.	*0.3344*
18. MW PSD 1st Min. rel.	*0.3341*
27. MW PSD 1st Min. Pos. rel.	*0.3340*
22. MW PSD 4th Max. rel.	*0.3222*
21. MW PSD 2nd Min. rel.	*0.3093*
33. MW PSD 1st Min NSF	*0.1563*

34. MW PSD 2nd Min NSF	*0.1408*
26. MW PSD 1st Max. Pos. ABS.	*0.1377*
24. MW PSD Origin Pos. rel.	*0.1377*
25. MW PSD In. Min. Pos. rel.	*0.1377*
17. MW PSD 1st Max. ABS.	*0.1139*
23. MW PSD End Val. rel.	*0.0565*
32. MW PSD End Val. Pos. rel.	*0.0202*

The meaning of the most relevant singularity parameters may be clearly deduced from Figure 1, where they are associated to the main features of the first two "V" profiles.

4. RESULTS AND DISCUSSION

The results after PCA have been plotted in terms of the three main principal components as shown in Figure 3. The most relevant original parameters in $\boldsymbol{S}_3$ correspond to the relative position in frequency of the 2nd minimum (x_{30}) and the second maximum (x_{28}) and the relative amplitude of the second maximum (x_{19}), all of them related to the second "V" profile.

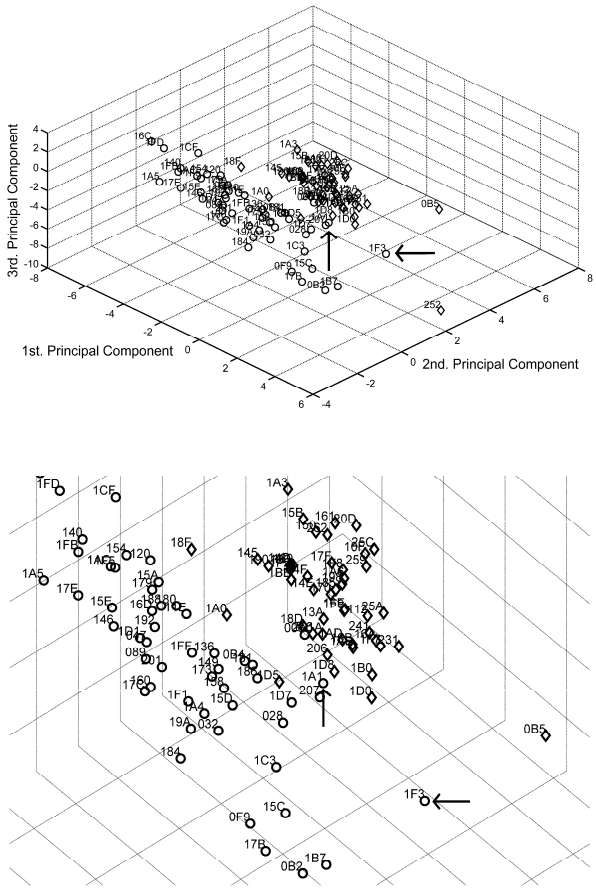

Figure 3. Top: Classification results in terms of the main 3 principal components. The samples are clustered into two groups linked by a tiny isthmus. Most of the samples labeled as (o) are in the left hand side group, whereas those with (◊) are in the right hand side one split by the 2nd principal component. Bottom: Isthmus enlarged for better view. Samples pinpointed by arrows are two misclassified cases. Labels give speaker identities.

Table 2 gives speaker-to-cluster assignments. The first observation is that samples were "blindly" separated according to speaker's gender, as most of male samples (48) were assigned to cluster (◊), where all female samples (50) were assigned to cluster (o). The rest of male samples (2) were assigned to cluster (o). Therefore there were 2/100 False Acceptances and 2/100 False Rejections. These results show clearly that the influence of gender is hidden in the set of observation parameters with stronger influence in clustering, i. e., those reflected as the most relevant ones by PCA analysis. It may be shown [8] that the parameters measuring the amplitude of the peaks relative to the first maximum {x_{19} and x_{22}} are related to the two most important massive structures on the cord cover accordingly to 2-mass vocal fold biomechanics [3]. The relative depth of the troughs in the "V" profiles {x_{18} and x_{21}} is related with the elastic parameters of the springs linking the masses. Tense vocal cords would present shallower troughs and would be more often seen in female than in male voice. On the other hand, male voice would be expected to show higher peaks than female voice. Although a deep study of the statistical distribution of these parameters is still pending, one may expect that the positional coefficients {x_{28}, x_{30}, x_{31}}, would be more compressed (tighter packed) in male than in female voice (corresponding to normalized frequency positions 4, 5 and 6 as given in the two bottom templates of Figure 2). This would be in good agreement with other studies treating the influence of gender in voice spectral parameters [16].

Table 2. Clustering of speakers

	Male speaker labels	**Female speaker labels**
Cluster (o)	*01A, 0B5, 10F, 112, 13A, 145, 14D, 14E, 14F, 150, 15B, 161, 169, 16A, 16B, 16E, 170, 17F, 185, 18D, 18F, 190, 196, 198, 1A0, 1A3, 1A6, 1AB, 1AD, 1B0, 1BB, 1D0, 1D5, 1D8, 1F2, 1F7, 1F8, 1F9, 206, 20D, 231, 241, 252, 256, 259, 25A, 25C, 262*	
	N=48	**N=0**
Cluster (◊)	*1A1, 1F3*	*006, 028, 032, 047, 089, 0B2, 0B4, 0F9, 11E, 120, 136, 140, 146,149, 154, 158, 15A, 15C, 15D, 15E, 160, 16C, 16D, 173, 179, 17B, 17C, 17E, 180, 181, 184, 186, 188, 192, 19A, 1A4, 1A5, 1AC, 1B7, 1C3, 1CF, 1D1, 1D7, 1F1, 1F5, 1FB, 1FD, 1FF, 201, 207*
	N= 2	**N= 50**

These results may explain why the gender condition is so neatly expressed in a completely blind clustering experiment as the one shown. It could be argued that gender appears to be related to pitch, and therefore it would be present anyway on the results, but this analysis would be too simplistic, as pitch influence has already been removed by the normalization process in expressions (1) and (2), therefore gender information must be encoded in the relative singularity val-

ues (mainly in troughs). The mis-clustered cases (speakers 1A1 and 1F3) are located near in the "isthmus" close to the boundary limits between both clusters where classification may be prone to ambiguity.

5. CONCLUSIONS

The power spectral density of the mucosal wave correlate contains relevant information directly related with vocal cord biomechanics. The singularities present in the spectral profile appear as "V" troughs, and once quantized in amplitude and frequency relative to the first maximum, may serve as biometric descriptors of the speaker's glottal function. A clear correlation is present among certain parameters estimated on the power spectral density of the mucosal wave correlate and certain speaker's features. In this sense, the speaker's gender is influencing strongly the power spectral distribution of glottal signals. It could be expected that other aspects of the speaker's meta-features, as age, tenseness, stress, nasalization, production gesture (chest, head, falsetto, etc.), and even pathology-related ones (creaky, fry, roughness, breathiness, etc), would appear reflected in the distribution of the specific glottal spectral profile defined, this being a matter for further study. These results are of most interest for forensic applications in cases where scarce records are available from a speaker, making it almost impossible to assign a specific identity to the records. Under such assumption meta-feature information could help in building robot phonetic pictures ("vocal passport") which could be later used in intelligent database searches, substantially reducing the span of search. The normalization of glottal biometry and its combination with vocal tract features may be applied in speaker identification and verification using fusion methods [13] to improve DET rates. The fields of security applications, as well as speech therapy, singing and others related, could also benefit from this kind of representations.

ACKNOWLEDGMENTS

This work is being funded by grants TIC2003-08756 and TEC2006-12887-C02-01/02 from Plan Nacional de I+D+i, Ministry of Education and Science, 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, August 27-29, 2003, pp. 81-87.

[2] 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.

[3] Berry, D. A., "Mechanisms of modal and non-modal phonation", *J. Phonetics*, Vol. 29, 2001, pp. 431-450.

[4] Fant G., Liljentcrants J., Lin Q., "A four-parameter model of glottal flow", *STL-QSPR*, Vol. 4, 1985, pp 1-13. Reprinted in *Speech Acoustics and Phonetics: Selected Writings*, G. Fant, Kluwer Academic Publishers, Dordrecht 2004, pp. 95-108.

[5] Godino, J. I., Gomez, P., "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors", *IEEE Trans Biomed. Eng*. Vol. 51, 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", *Proc. of the EUSIPCO'04, 2004*, pp. 297-300.

[7] 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 INTERSPEECH'05*, 2005, pp. 645-648.

[8] 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.

[9] Gómez, P., Rodellar, V., Álvarez, A., Lázaro, J. C., Murphy, K., Díaz, F., Fernández, R., "Biometrical Speaker Description from Vocal Cord Parameterization", *Proc. of ICASSP'06*, Toulouse, France, 2006, pp. 1036-1039.

[10] Hirano, M., Hibi, S., Yoshida, T., Hirade, Y., Kasuya, H., and Kikuchi, Y., "Acoustic analysis of pathological voice. Some results of clinical application," *Acta Otolaryngologica*, vol. 105, no. 5-6, pp. 432-438, 1988.

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

[12] Nickel, R. M., "Automatic Speech Character Identification", *IEEE Circuits and Systems Magazine*, Vol. 6, No. 4, 2006, pp. 8-29.

[13] Plumpe, M. D., Quatieri, T. F., Reynolds, D. A., "Modeling of the Glottal Flow Derivative Waveform with Application to Speaker Identification", *IEEE Trans. on Speech and Audio Proc.*, Vol. 7, No. 5, 1999, pp. 569-586.

[14] Reynolds, D. A., Campbell, W., Gleason, T., Quillen, C., Sturim, D., Torres-Carrasquillo, P., Adami, A., "The 2004 MIT Lincoln Laboratory Speaker Recognition System", *Proc. of the ICASSP'05*, 2005, pp. I-177-180.

[15] Van der Heiden, F., Duin, R. P. W., De Ridder, D., Tax, D. M. J., *Classification, Parameter Estimation and State Estimation*, John Wiley, Chichester, England, 2004.

[16] Whiteside, S. P., "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," *J. Acoust. Soc. Am.*, Vol. 110, No. 1, pp. 464–478, 2001.