Speaker's Gender Detection from Glottal Biometry

Pedro Gómez-Vilda, Roberto Fernández-Baillo, Agustín Álvarez-Marquina, Luis Miguel Mazaira-Fernández, Rafael Martínez-Olalla, Victoria Rodellar-Biarge

Grupo de Informática Aplicada al Tratamiento de Señal e Imagen, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo, s/n, 28660 Madrid

e-mail: pedro@pino.datsi.fi.upm.es

Abstract. Through the present work a biometric signature of a speaker's voice is proposed for the detection of the speaker's gender. The estimation method relies on the extraction of the glottal flow derivative from voice after removing the vocal tract transfer function by inverse filtering. This spectral density is related to the vocal fold cover biomechanics, and it is well known that certain speaker's features as gender, age or pathologic condition are present in it. For such a database of 100 pathology-free speakers equally balanced in gender and age is used as an experimental framework to draft the results exposed in the work. As the estimated biometric parameters show a certain degree of cross-correlation Principal Component Analysis (PCA) is used to reduce parameter dimension. The principal components are used in unsupervised *k-means* clustering of speakers (unsupervised gender detection). The outcome grouping shows an almost complete separation of speakers by gender in terms of the most relevant parameters derived from a statistical dispersion study. Possible applications of the study can be found in forensic acoustics as well as in speaker identification and verification tasks.

Keywords: Voice Biometry, Speaker's Identification, Speaker Biometrical Characterization, Forensic Acoustics, Glottal Source

1. Introduction

The present work is oriented to voice characterization to determine a biometric signature of voice based on the parameterization of the glottal biomechanics for voice characterization (gender, age and speaker's pathological voice condition being the primary targets, among others). A comprehensive review of the characterization of voice may be found in [1]. Traditionally the characterization of the speaker has been oriented to gender and age as the main goals. Good studies have been published in this sense during the last two decades [2][3][4][5][6]. These works show the way to establish a more structured study regarding voice characterization. On one side they point out to the use of time or frequency domain parameters as the basis of the study. On the other side, they deal with Vocal Tract or Glottal Source biometry. In the present approach the Glottal Source has been selected as the object of the research. A generalized signature is proposed on a full description of the Glottal Source spectral envelope, concentrating on the singularities appearing on this pattern (peaks and troughs). This generalization is based on the biomechanical foundations of the Glottal Source spectral envelope [7], whose singularities may be shown to be strongly conditioned by the biomechanical relations among parameters in well-known *k-mass* models [8]. Principal Component Analysis [9] is proposed to produce more compact data sets which can improve detection and classification results. This first approach to a more general study is oriented to the detection of specific speaker's characteristics as gender on the glottal biometrical signature of voice. Classically most works dealing with the biometry of voice have considered the voice signal as a whole, not establishing a clear separation among the roles played by the different organs implied in voice production (vocal tract vs vocal folds) [10]. After the early work of Brookes and Chan [11] it has been only in the last years when an interest has appeared in studying the characteristics of the Glottal Source separately from the vocal tract for speaker recognition [12][13]. Nevertheless, it seems intuitive that in treating voice biometry following a deconstructive way, important improvements could be obtained. This means that glottal parameters have to be treated separately accordingly to their statistical inter-speaker and intra-speaker characteristic distributions. Considering the classical source-filter voice generation model [14] composed by an excitation (Glottal Source) and a modulating structure (Vocal Tract), it may be expected that the excitation will depend on the biometric low-level characteristics of the speaker (glottal system physiology) being weakly influenced by the message (text), but strongly conditioned by the production process (physiological and emotional conditions, prosody, tonal height, production gesture, pathology, etc.). An analytical description of voice biometry is proposed in Figure 1.

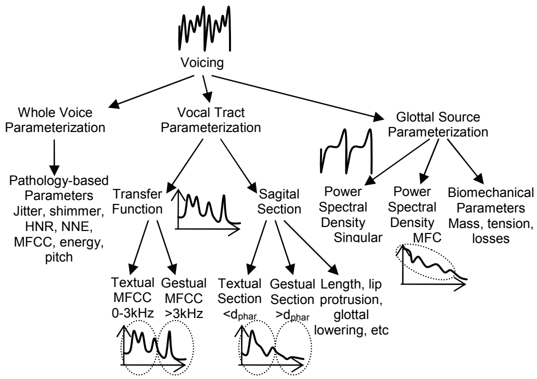

Figure 1. Analytic description of voice biometry in terms of vocal characteristics (mainly message dependent) or glottal characteristics (mainly biometric).

The parameterization of voice may be carried out using estimates of one of the following main categories (from left to right in the picture):

- The Whole Voice Power Spectral Density (WVPSD), estimated by FFT or LPC. The short-time power spectrum is coded as Mel-Frequency Cepstral Coefficients (MFCC) [15].
- The Vocal Tract Transfer Function Modulus (VTTFM). The WVPSD reflects the influence of the Glottal Source spectral envelope as a *1/f* spectral tilt, which distorts the Vocal Tract Transfer Function. A separation between Vocal Tract and Glottal Source could render better results in the decoding of message (Speech Recognition) as well as in the characterization of the source (Speaker Recognition).
- The Glottal Source Power Spectral Density (GSPSD). The Glottal Source can be parameterized in the time or in the frequency domain. Time domain methods are based in the well-known Liljiencrants-Fant model [16]. Frequency domain methods are preferred as they tend to be more robust facing noise or pathology [5].

Within the VTTFM a clear distinction could be made between frequency regions below 3000 Hz, which are more influenced by the message, and above 3000 Hz which are more influenced by the speaker's gesture and personality. The parameterization of the Vocal Tract can be given as well in terms of its associated area functions (Sagital Section). In this case it is also possible to establish two segments: the oral part and the glottal part. The former is more influenced by articulation, the later is more related to the speaker's characteristics. Concerning the parameterization of the Glottal Source the time domain methods are oriented to the estimation of OC, SC, ClQ, RQ and NAQ (Open, Speed, Closing, Return and Normalized Amplitude Quotients). The frequency domain (GSPSD) is oriented to the estimation of H_1-H_2 which is known to be related to the CQ (Close Quotient), as well as the Maximum Flow Declination Rate (MFDR) and the Spectral Slope. Other methods are based on MFCC or LPCC parameterization of the power spectral density of the glottal signals (glottal flow derivative, glottal source derivative, etc.) similarly to VTTFM. Another line is related with the parameterization of the Glottal Source frequency envelope and the extraction of the biomechanical parameters of a k-mass glottal model by inversion as in [21].

2. Estimation of the glottal source

The methodology proposed in this work is based in a frequency domain parameterization of the glottal source power spectral density, with the following distinctive characteristics:

- It is carried out either on the Glottal Source or on the Mucosal Wave Correlate (MWC), derived from the Glottal Source by removing the Acoustic Average Wave (AAW) [17].
- It estimates the singularities of the Mucosal Wave Correlate Power Spectral Density (MWCPDS) as sets of peaks and notches relative to F_0. Therefore it can be considered as a generalization of the parameters used in [5].

Its biometrical character is granted by its inter- and intra-speaker statistical variability mainly conditioned by the personal characteristics of the speaker (gender, age, tension, glottal gesture, etc). The methodology used for the estimation of the Glottal Source is based on the elimination of the vocal tract by inverse filtering by well-known methods [18], and in the separation of the Glottal Source into the two referred components (AAW and MWC). An example of the glottal signal estimation results from inverse filtering may be seen in Figure 2 from quasi-stationary utterances of the vowel /a/ by typical male and female speakers.

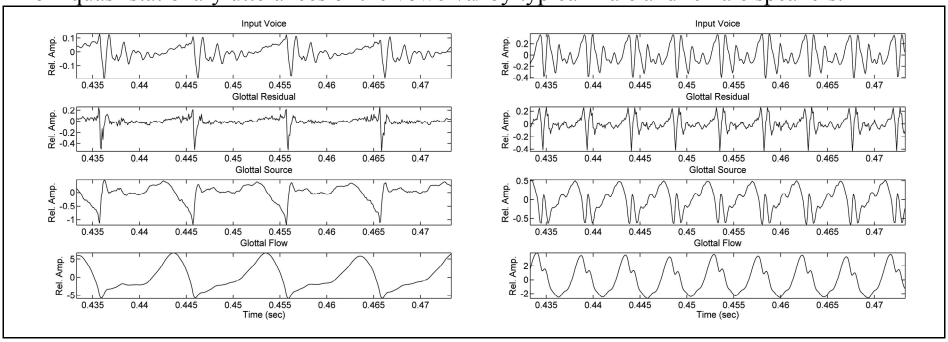

Figure 2. Examples of reconstructed glottal signals from vowel /a/ for prototype male and female speakers (*#185* –left-, and *#158* -right). From top to bottom: input voice, glottal residual, source and flow (four left templates: male prototype; four right templates: female prototype). Horizontal axes are given in sec for a sampling frequency of 11,050 Hz.

The plot in Figure 3 reproduces in detail the time evolution of a cycle of the Glottal Source (full line) where the four phonation phases may be observed separated by vertical dot lines from left to right: return, closure, open and closing phases. Two other variables are plotted as well: the Average Acoustic Wave (dash-dot) and the Mucosal Wave Correlate (dash). The dash-dot plot corresponds to the ideal Glottal Source if no Vocal and Pharyngeal Tracts were present under non inertial load conditions assuming that each Vocal Fold could be represented by a single body mass (1-mass model). This would be equivalent to two ideal Vocal Folds with a single mass behaviour attached to the walls of the tract by single elastic springs. The vibration would describe perfect semi-sinusoidal arches accordingly to the relation between mass and spring constants. This signal coincides with the Acoustic Average Wave (AAW) and has been evaluated by optimally fitting a semi-sinusoid arch to the Liljencrants-Fant pattern [7].

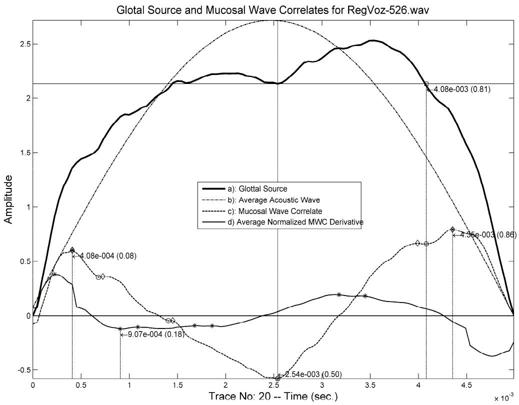

Figure 3. Splitting the Glottal Source (a) into the Average Acoustic Wave (b) and the Mucosal Wave Correlate (c). The minimum in the derivative of the MWC (d) marks the end of the return phase. The vertical middle dot line divides the phonation cycle into the close phase (left) and the open phase (right). The vertical side dot lines mark the return and closing points.

The dash plot corresponds to the difference between the AAW and the L-F Glottal Source plots, and is thus referred as the Mucosal Wave Correlate. This signal shows interesting properties, such as the ability of pointing out the start of the open phase, which takes place at its minimum (middle vertical dot line). This property may be used in detecting the open and close intervals of the phonation cycle.

3. Glottal-Source based Biometric Signature

Through the present approach a methodology to derive biometrical parameters of the Glottal Source in the frequency domain is proposed. The biometrical parameters are estimated on the power spectral density of either the Glottal Source or the Mucosal Wave Correlate. The signature obtained from the Mucosal Wave Correlate is more specifically related to the biomechanics of the vocal fold cover, while that from the Glottal Source includes the biomechanics of both the body and the cover of the vocal fold. The estimates based on this last approach are more suitable for biometric applications, the estimates from the Mucosal Wave Correlate being more suitable for studies in vocal fold pathology. In both cases the parameter estimation methodology to be applied is the same. The power spectral densities shown in Figure 4 correspond to the Glottal Source from prototype male and female voices. A common behaviour may be observed in both cases regarding the envelopes of the power spectral densities: a fast raise from low frequencies to a maximum and a decay towards lower frequencies with a general trend of *12 dB/oct*. In between a series of valleys or local minima may be appreciated surrounded by peaks.

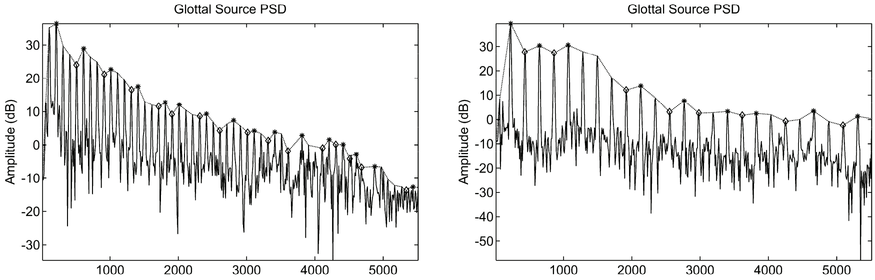

Figure 4. a) Power spectral density of the glottal source from vowel /a/ for prototype male and female speakers (*#185* –left-, and *#158* -right) showing the singularities superimposed: *-maxima; ◊-minima. Relative amplitude is given in dB. Horizontal axes are given in Hz.

In Figure 5 the envelope of the glottal source power spectral density of the male prototype has been extracted showing a first maximum T_{M1} centered at a frequency f_{M1} followed by a descent to a minimum T_{m1} in f_{m1} and to a new maximum T_{M2} at a frequency f_{M2}. This type of notch may appear several more times as the general trend of the power spectral density is decaying. The presence of two maxima enclosing a minimum is explained by the resonances and anti-resonances in the system of masses and springs on the vocal fold body and cover structures [8].

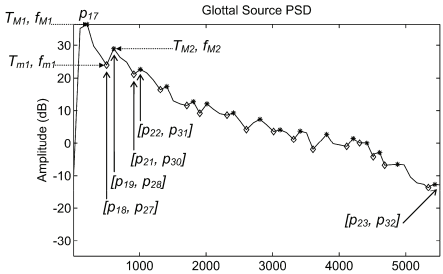

Figure 5. Power spectral density envelope of the glottal source for speaker #185 showing the first notch profile $\{T_{M1}, f_{M1}\}$, $\{T_{m1}, f_{m1}\}$ and $\{T_{M2}, f_{M2}\}$, and the meaning of 10 of the singularity parameters used in the study $\{p_{17}, p_{18}, p_{19}, p_{21}, p_{22}, p_{27}, p_{28}, p_{30}, p_{31}$ and $p_{32}\}$. Relative amplitude is given in dB. Horizontal axis is given in Hz.

Therefore a glottal signature of voice may be established detecting each notch by estimating the amplitude and position of its singularity points and its slenderness factor as described in [7]. In a practical case the biometrical signature is estimated from the singularities of the power spectral density of either the MWC or the Glottal Source as follows

- The Glottal Source is windowed in 512-sample frames and the power spectral density of each window is estimated by FFT in dB as in Figure 4.
- The envelopes of the power spectral densities are estimated for each frame.
- The maxima (*) and minima (◊) on the respective envelopes are detected and their amplitudes and frequencies collected as two lists of ordered pairs: $\{T_{Mk}, f_{Mk}\}$ and $\{T_{mk}, f_{mk}\}$, with k the ordering index.
- The first (and usually the largest of all maxima) (T_{M1}, f_{M1}) is used as a normalization reference both in amplitude and in frequency.
- The normalized singularity points and the approximate envelope of the power spectral densities for the MWC are assigned to the parameters for the study accordingly to Table 1.

Table 1. MWC singularity parameters used in the study.

Parameter No. and Description	Parameter No. and Description
p_{17} - Amplitude of the first maximum in dB T_{M1}	p_{26} - Absolute pos. of first maximum f_{M1}
p_{18} - Normalized ampl. of first minimum in dB τ_{m1}	p_{27} - Norm. pos. of first minimum φ_{m1}
p_{19} - Norm. ampl. of second maximum in dB τ_{M2}	p_{28} - Norm. pos. of second maximum φ_{M2}
p_{21} - Norm. ampl. of second minimum in dB τ_{m2}	p_{30} - Norm. pos. of second minimum φ_{m2}
p_{22} - Norm. ampl. of third maximum in dB τ_{M3}	p_{31} - Norm. position of third maximum φ_{M3}
p_{23} - Norm. ampl. of spec. prof. at max. freq. in dB τ_{fm}	p_{32} - Norm. position of end value φ_{fm}
p_{24} - Norm. position of initial value in freq. φ_i	p_{33} - Slenderness of first notch σ_{m1}
p_{25} - Norm. pos. of first min. before the first max. φ_{m0}	p_{34} - Slenderness of second notch σ_{m2}

Another possible parameterization strategy would be based on clipping voicing frames in segments aligned with the pitch cycle. In this way a different estimation would be produced for each cycle-like segment. In the present study voice frame durations of *0.2* sec. long are used producing different numbers of pitch cycles for male and female voice (typically ranging from 20-40). The number of pitch cycles used is designated as *M*. Assuming reasonable stationary conditions along the frame duration (considering that a stable vowel is produced) estimations of the parameter means and standard deviations could be used in classification as

$$x_{ij} = \frac{1}{M}\sum_{m=1}^{M} p_{ijm} \tag{1}$$

$$\sigma_{ij} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(p_{ijm} - x_{ij}\right)^2} \tag{2}$$

where *i* is the parameter index and *m* is the cycle index. In this way the estimations are more robust to intra-speaker variability as will be shown in the sequel.

4. Materials and methods

A corpus of 100 normal speakers equally distributed by gender was randomly recruited from a wider database recorded during the life of project MAPACI [19]. 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 electroglottography, video-endoscopiy and GRBAS evaluation [20]. The recordings consisted in utterances of the vowel */a/* of about *3 sec* per record. A *0.2 sec* frame from the record centre was used in the estimations. The spectral profile parameters $\{p_{17-34}\}$ As each parameter was estimated on a phonation cycle basis, for a prototype male voice (with pitch around *100 Hz*) an average of $M=20$ values was obtained, which for female voice (with a typical pitch of *200 Hz*) should be around $M=40$. In this way $J=46$ observation parameters x_{ij} were obtained as the average of each observation parameter p_{im} over $1 \le m \le M$ phonation cycles following (1) with $1 \le j \le J$ for each speaker $1 \le i \le I$ in the set of $I=100$ speakers. The estimations of observation parameter *j* for all the speakers $1 \le i \le I$ in the set are stacked as a column vector

$$\boldsymbol{x}_j = \left[x_{1j}, x_{2j} \dots x_{ij}, \dots x_{Ij}\right]^T \tag{3}$$

Similarly the estimations for the whole set of parameters are piled as a matrix of observations

$$\boldsymbol{X} = \left[\boldsymbol{x}_1, \dots \boldsymbol{x}_j, \dots \boldsymbol{x}_J\right] \tag{4}$$

Principal Component Analysis was applied to this dataset as described in [9] and [22] to re-evaluate the set of observation parameters as

$$\boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{e}_j; \quad 1 \le j \le J \qquad (5)$$

where the vectors $\boldsymbol{y}_j$ contain the new parameters (principal components) for each speaker in the list $1 \le i \le I$ their variance diminishing with component order. PCA was applied as follows:

a) Pre-selection of a database $\boldsymbol{X}_{17\text{-}34}$ from the original parameter set $\boldsymbol{S}_0 = \{x_{1\text{-}46}\}$ for the whole set of speakers. The resulting subset of parameters $\boldsymbol{S}_1 = \{x_{17\text{-}19}, x_{21\text{-}28}, x_{30\text{-}34}\}$ included the normalized estimates of the power spectral density singularities, as given in Table 1.
b) Z-score the database $\boldsymbol{X}_{17\text{-}34}$ by subtracting means and normalizing to standard deviations.
c) Split the database $\boldsymbol{X}(\boldsymbol{S}_1)$ in two clusters by *k-means* blindly (unsupervisedly).
d) Apply PCA on $\boldsymbol{X}(\boldsymbol{S}_1)$ to transform it to a new manifold for *16* principal components producing a matrix $\boldsymbol{Y}_{1\text{-}16}$ ordered by component relevance.
e) Select the three first components for 3-D presentation purposes.

5. Results and discussion

The results in $\boldsymbol{Y}_{1\text{-}16}$ have been plotted in terms of the three first principal components, as described in a)-e) as given in Figure 6.

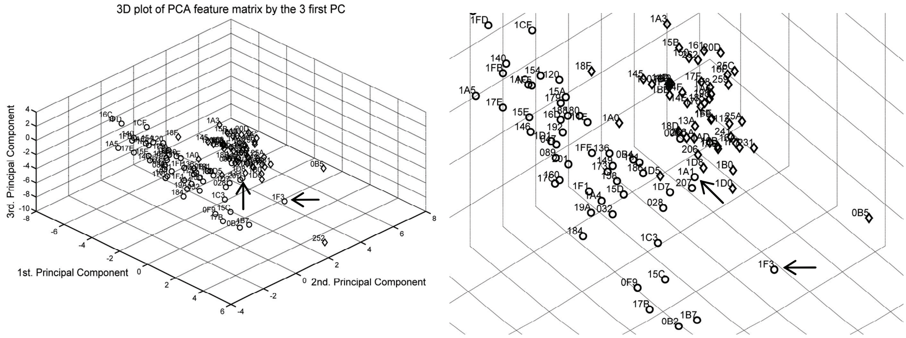

Figure 6. Classification results in the PCA manifold in terms of the first 3 principal components. Left: The set of samples is clustered into two main groups. Samples labelled (◊) are from male subjects, whereas those labelled (o) are from female subjects with two exceptions for #1A1 and #1F3, pinpointed by arrows. Right: Close-up view of the same plot.

These results show that the unsupervised clustering succeeded in accurately separating speakers by gender with the exception of the two male subjects grouped within the female cluster (*#1A1* and *#1F3*). The female cluster shows a broader branch-like inter-speaker variability than the male cluster, which is less spread-out. This may imply that different branches may be found within the mainly-female cluster and would deserve a further investigation. As a consequence it may be said that it will be easier to establish classifications within female than in male groups.

The main question to be answered at this point is which parameters will be more sensitive to gender, as there is a clear dependence of sample distributions on gender. To gain a better view on intra- and inter-speaker variability the following steps were covered:

- The average values and standard deviations of each speaker were evaluated for each parameter in their respective templates in terms of *M* accordingly with (1) and (2), thus serving as estimates of intra-speaker variability.
- The statistical dispersion of the parameter templates for the set of male and female subjects was presented as box plots, thus serving as estimates of inter-speaker variability.

Results are presented in Figure 7 (contrasted to the ones from the male and female prototypes).

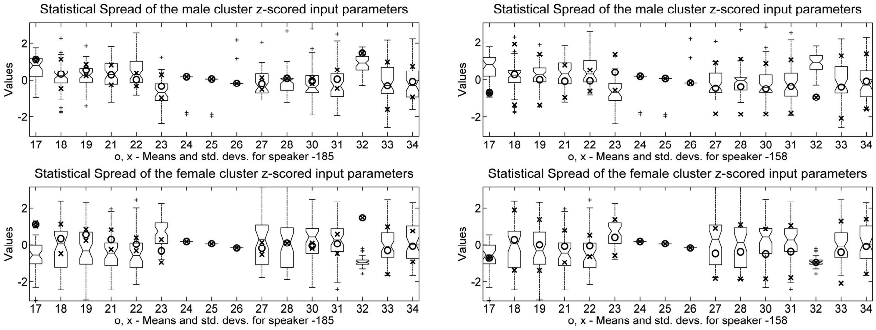

Figure 7. Statistical dispersion of the profile parameters used in the study from the male (left) and female (right) groups compared against their respective prototype male (#185) and female (#158) templates. The intra-speaker variability is expressed by the marks (o: mean) and (×: standard deviation). The inter-speaker variability is represented as notch-box plots.

First of all it must be mentioned that certain parameters (as x_{24}, x_{25} and x_{26}) show almost no variability, therefore they are not of specific interest for our study. The impression derived from Figure 6 about the wider spread presented by female distributions when comparing male and female clusters is clearly confirmed. This dispersion is especially relevant regarding parameters x_{18}, x_{19}, x_{27} and x_{28}, which are related to estimates of the first minimum and second maximum. Besides, certain parameter distributions from male and female groups do not overlap, or do so slightly, as is the case of x_{17}, x_{23} and x_{32}. This means that if used in differential clustering experiments they will render the best results, as will be shown in the sequel. Another interesting result is derived from the comparison between the prototype intra-speaker male template (*#185*) against the spread of the male and female groups. It may be seen that the prototype male template fits within the male parameter spread (showing slight deviations for x_{19}, x_{23}, x_{28} and x_{32}). The same template when compared against the female distribution is in clear disagreement with respect to parameters x_{17}, x_{23} and x_{32}. A similar comparison may be carried out on the female prototype (*#158*) against male inter-speaker variability, showing strong disagreements again with respect to parameters x_{17}, x_{23} and x_{32}. On the contrary only parameters x_{30} and x_{31} show slight deviations between the prototype female and the female inter-speaker dispersion. A further confirmation of these observations is obtained using Fisher's Discriminant Ratio

$$fdr_j = \frac{\left(\mu_{mj} - \mu_{fj}\right)^2}{\sigma_{mj}^2 + \sigma_{fj}^2}; \quad 1 \le j \le J \tag{6}$$

where (μ_{mj}, σ_{mj}) and (μ_{fj}, σ_{fj}) are the means and standard deviations of the male and female distributions for parameter j. The results in Table 2 confirm the observations in the sense that x_{17}, x_{23} and x_{32} are the most relevant parameters in gender detection.

Table 2. Relevance of singularity parameters from FDR

Parameter index and name	Relevance	Parameter index and name	Relevance
32. MW PSD End Val. Pos. rel.	0.2083	23. MW PSD End Val. rel.	0.1430
17. MW PSD 1st Max. ABS.	0.1365	33. MW PSD 1st Min NSF	0.0213
18. MW PSD 1st Min. rel.	0.0146	31. MW PSD 4th Max. Pos. rel.	0.0092
19. MW PSD 2nd Max. rel.	0.0078	30. MW PSD 2nd Min. Pos. rel.	0.0040
34. MW PSD 2nd Min NSF	0.0029	28. MW PSD 2nd Max. Pos. rel.	0.0008
27. MW PSD 1st Min. Pos. rel.	0.0006	21. MW PSD 2nd Min. rel.	0.0005
22. MW PSD 4th Max. rel.	0.0003	24. MW PSD Origin Pos. rel.	0.0000
25. MW PSD In. Min. Pos. rel.	0.0000	26. MW PSD 1st Max. Pos. ABS.	0.0000

These results may be used in improving clustering strategies as shown in Figure 8, where clear differential groupings may be obtained when a non-overlapping parameter as x_{32} is used in a typical differentiating experiment.

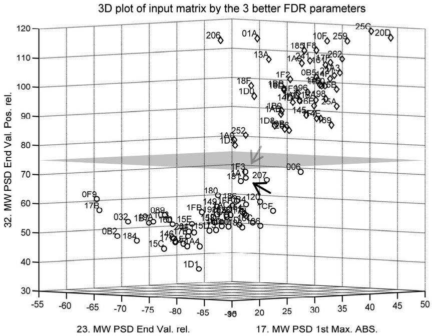

Figure 8. Differential clustering in terms of the most relevant parameters according to FDR: x_{17}, x_{23} and x_{32} where clear gender separation is produced by the plane at x_{32}=*75*.

The fact that gender splitting is feasible by hyperplane separation (x_{32}=75) suggests that more sophisticated techniques as Support Vector Machines may be used in complex multiple-feature separation (as gender, age or phonation modality).

6. Conclusions

From what has been shown the following important conclusions may be derived:

- The general decay trend of the glottal signal is coded in parameters p_{17}, p_{23} and p_{32}.
- Parameters p_{17}, p_{23}, p_{32} and p_{33} are the most sensitive ones to gender.
- Genders show different parameter dispersions, being broader in female than in male voice.
- In general intra-speaker parameter dispersion is lower than inter-speaker dispersion.
- PCA helps interpreting results by dimensionality reduction.

As a general conclusion it may be said that a structured classification of the biometry of voice is a real need, as specific and clearly differentiated biometric information is present in the glottal components of voice, independently from features observed in vocal tract features. Therefore splitting voice into vocal and glottal components is a reasonable technique when articulation and biometry are two different objectives, as for example, in forensic applications of voice. As speaker identification and characterization algorithms strongly rely on joint probability densities of the parameters used in the experiments the production of glottal and vocal parameter descriptions statistically independent may be the clue to more accurate speaker recognition methods. This is especially important as far as the False Acceptance rates in security applications are critical to determine the suitability of these techniques in a given scenario. In this respect a combination of vocal and glottal feature descriptors independently and in fusion experiments may help in establishing efficient strategies for the improvement of detection rates. The implementation of this methodology may rely in pitch-synchronous or pitch-independent strategies, both having been tested with similar results. This makes it suitable for its application in real scenarios in forensic and security frameworks. The methodology presented may be also generalized to the study of speaker features as age, voice profile, emotional features and others alike.

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] R. M. Nickel, Automatic Speech Character Identification, IEEE Circuits and Systems Magazine 6 (4) (2006) 8-29.
[2] P. J. Price, Male and female voice source characteristics: Inverse filtering results, Speech Comm. 8 (1989) 261-277.
[3] D. G. Childers et al., Automatic recognition of gender by voice, Proc. of the ICASSP 1 (1988) 603-606.
[4] A. R. Sulter and H. P. Wit, Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age, J. Acoust. Soc. Am. 100 (1996) 3360-3373.
[5] P. Alku, Parameterisation Methods of the Glottal Flow Estimated by Inverse Filtering, Proc. of VOQUAL'03, Geneva, August 27-29 (2003) 81-87.
[6] S. P. Whiteside, Sex-specific fundamental and formant frequency patterns in a cross-sectional study, J. Acoust. Soc. Am. 110 (1) (2001) 464–478.
[7] P. Gómez et al., Evaluation of voice pathology based on the estimation of vocal fold biomechanical parameters, J. Voice 21 (4) (2007) 450-476.
[8] D. A. Berry, Mechanisms of modal and non-modal phonation, J. Phonetics 29 (2001) 431-450.
[9] Johnson, R. A., Wichern, D. W., Applied Multivariate Statistical Analysis, Prentice-Hall, Upper Saddle River, NJ, 2002.
[10] F. Bimbot et al., A tutorial on text-independent speaker verification, Eurasip Journal on Applied Signal Processing (4) (2004) 430–451.
[11] D. M. Brookes and D. S. F. Chan, Speaker characteristics from a glottal airflow model using robust inverse filtering, Proc. Inst. of Acoustics 16 (5) (1994) 501–508.
[12] M. D. Plumpe et al., Modeling of the Glottal Flow Derivative Waveform with Application to Speaker Identification, IEEE Trans. on Speech and Audio Proc. 7 (5) (1999) 569-586.
[13] N. Zheng et al., Integration of Complementary Acoustic Features for Speaker Recognition, IEEE Signal Proc. Letters 14 (3) (2007) 181-184.
[14] G. Fant, Theory of Speech Production, Mouton, The Hague, Netherlands (1960).
[15] J. R. Deller, J. G. Proakis J. H. L. Hansen, Discrete-Time Processing of Speech Signals, Macmillan, NY (1993).
[16] G. Fant et al., A four-parameter model of glottal flow, STL-QSPR 4 (1985) 1-13. Reprinted in: Speech Acoustics and Phonetics: Selected Writings, G. Fant, Kluwer Academic Publishers, Dordrecht (2004) 95-108.
[17] I. R. Titze, Summary Statement, Workshop on Acoustic Voice Analysis, National Center for Voice and Speech (1994).
[18] P. Alku, An Automatic Method to Estimate the Time-Based Parameters of the Glottal Pulseform, Proc. of the ICASSP'92 (1992) II/29-32.
[19] Project MAPACI: http://www.mapaci.com.
[20] M. Hirano et al., Acoustic analysis of pathological voice. Some results of clinical application, Acta Otolaryngologica 105 (5-6) (1988) 432-438.
[21] P. Gómez et al., Biometrical Speaker Description from Vocal Cord Parameterization, Proc. of ICASSP'06, Toulouse, France, (2006) 1036-1039.
[22] P. Gómez et al., PCA of Perturbation Parameters in Voice Pathology Detection, Proc. of INTERSPEECH'05 (2005) 645-648.