Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics

Patricia Henríquez, Jesús B. Alonso, Miguel A. Ferrer, Carlos M. Travieso, Juan I. Godino-Llorente, and Fernando Díaz-de-María

***Abstract*—In this paper, we propose to quantify the quality of the recorded voice through objective nonlinear measures. Quantification of speech signal quality has been traditionally carried out with linear techniques since the classical model of voice production is a linear approximation. Nevertheless, nonlinear behaviors in the voice production process have been shown. This paper studies the usefulness of six nonlinear chaotic measures based on nonlinear dynamics theory in the discrimination between two levels of voice quality: healthy and pathological. The studied measures are first- and second-order Rényi entropies, the correlation entropy and the correlation dimension. These measures were obtained from the speech signal in the phase-space domain. The values of the first minimum of mutual information function and Shannon entropy were also studied. Two databases were used to assess the usefulness of the measures: a multiquality database composed of four levels of voice quality (healthy voice and three levels of pathological voice); and a commercial database (MEEI Voice Disorders) composed of two levels of voice quality (healthy and pathological voices). A classifier based on standard neural networks was implemented in order to evaluate the measures proposed. Global success rates of 82.47% (multiquality database) and 99.69% (commercial database) were obtained.**

***Index Terms*—Chaos, disordered speech, entropy, nonlinearity.**

I. Introduction

THE main methods used by the medical community to evaluate the speech production system and diagnose pathologies are either direct ones which require direct inspection of vocal folds (using laryngoscopical techniques such as fiberscope) and cause discomfort to the patient, or subjective ones in which voice quality is evaluated by a doctor's direct audition (GRBAS and RBH methods [45], [46]). These techniques require trained expert doctors. The use of voice quality measures obtained from recorded voice allows us to quantify the voice quality and to document the patient evolution using objective measures. They are noninvasive, quick and automatic techniques and can be a help to traditional techniques used in medicine.

The use of these techniques combined with classification methods provides the development of expert aided systems for the detection of speech system pathologies. These systems can be applied as portable tools in preventive medicine, especially for professional singers or presenters who have more risks of suffering from voice disorders. They are also useful tools for postoperative monitoring. Moreover, their use in telemedicine environments is possible as a remote and automatic screening method. Finally, they can be used as a medical–legal documentation tool to express in a quantitative manner the success of a surgical intervention.

In the last decades, some studies have provided objective measures of voice quality. Measures are obtained of the voice signal in time, spectral and cepstral domains. The most important measures used in existing literature are: fundamental frequency [1], [2], whose determination is important because several measures depend on its correct estimation, pitch perturbation (jitter) [3], [4], amplitude perturbation (shimmer) [3], [4], harmonic to noise ratio [5], low to high energy ratio [6], normalized noise energy [7], glottal to noise excitation ratio (GNE) [8], dynamic time warping and Itakura–Saito distorsion measure [47]. Using a combination of these sets of measures, laryngeal pathologies detection systems using recorded voice signal have been developed obtaining different success rates in the classification between healthy and pathological voices: 93.5% [9], 85.8% [10], 76.67% [11], 96.1% [12]. The comparison of mentioned rates is difficult because each system has been evaluated with different databases, since a reference database does not exist. Moreover, as reported in [43], the evaluation of the results is far from being robust.

Nevertheless, most measures considered in these works do not take into account nonlinearity in the speech system despite the fact that some studies show the underlying process of speech generation exhibiting nonlinear components [13]–[17]. As a result, recent works consider this new approach in order to reveal discriminative measures between healthy and pathological voices. Examples are measures based on high order statistics (HOS) [18], [19] and AM–FM modelling of voice signal [20]–[23].

Chaos theory, an area of nonlinear dynamics systems theory, applied to nonlinear time series has recently been adopted as a new nonlinear approach to speech signal processing. The application of nonlinear chaotic techniques in speech signal

Manuscript received August 19, 2008; revised January 26, 2009. Current version published July 06, 2009. This work was supported in part by the Spanish Government under Grant TIC2003-08956-C02-02. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mark Hasegawa-Johnson.

P. Henríquez, J. B. Alonso, M. A. Ferrer, and C. M. Travieso are with the Technological Centre for Innovation in Communications (CeTIC), Department of Signal and Communication, University of Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain (e-mail: phenriquez@cetic.eu; mferrer@dsc.ulpgc.es; ctravieso@dsc.ulpgc.es).

J. I. Godino-LLorente is with the Department of Circuits and Systems Engineering, University Politécnica de Madrid, 28040 Madrid, Spain (e-mail: igodino@ics.upm.es).

F. Díaz de María is with the Department of Signal Theory and Communications, University of Carlos III de Madrid, 28911 Leganés Madrid, Spain (e-mail: fdiaz@tsc.uc3m.es).

Digital Object Identifier 10.1109/TASL.2009.2016734

processing so far are based on chaotic modelling or extraction of chaotic characteristics (Lyapunov exponents, correlation dimension, etc.). The main chaotic characteristics studied are the Lyapunov exponents [24]–[26] and dimensions of attractor, especially the correlation dimension. The correlation dimension has been shown to be capable of distinguishing healthy voices from pathological ones [24], [27]–[29] and even distinguishing between different types of pathologies such as ataxic dysarthia and hyperkinetic extrapyramidal dysarthia [29]. A high-quality vowel synthesizer based on chaotic techniques has also been developed [30]. The entropy has been applied to detect complex dynamics in disordered speech [31] in a preliminary study.

Our aim is to study the usefulness of six nonlinear chaotic measures in the automatic discrimination of two levels of voice quality (healthy and pathological speakers) measured from speech recordings. In order to assess the usefulness of the measures an automatic classification system is used. As far as we know, this is the first time that the six measures proposed in this work have been used together in order to discriminate between healthy and pathological speakers for screening purposes. Furthermore, two databases have been used to assess the usefulness of the measures and to compare results. As the multiquality database is labelled with different levels of voice quality a deeper analysis is carried out with this database. Four levels of voice qualities have been considered, from healthy to severe pathological voice. The more pathological a voice is the less quality it presents. The results obtained show that these characteristics provide high classification rates.

This paper is divided into the following sections: methods, data, experiments, results, and conclusion. The methods section is devoted to a brief theoretical description of nonlinear dynamics system. The data section is related to the two databases used in the experiments. In the experiments section, the experimental procedure is explained. In the results section, the results are shown and discussed.

II. Methods

A. Nonlinear Dynamics System

Equations of several complex systems are usually unknown. As a result they can only be analyzed from the information within a time series as an output of the system. If some indicators show that time series have nonlinear behavior, nonlinear techniques such as chaos theory can be applied to extract nonlinear characteristics of the system.

Deterministic dynamical systems describe the time evolution of a system in some phase space $\Gamma \in \Re^m$ (m-dimensional vectorial space), where a state is specified by a vector $\overline{x} \in \Re^m$. The evolution in time t can be expressed by ordinary differential equations

$$\frac{d}{dt}\overline{x}(t) = f\left(t, \overline{x}(t)\right), \quad t \in \Re \tag{1}$$

or in discrete time $t = n\Delta t$ by maps

$$\overline{x}_{n+1} = F(\overline{x}_n), \quad n \in \mathrm{Z}. \tag{2}$$

A sequence of points ($\overline{x}(t)$ or $\overline{x}_n$) that solve the equations of the system are called trajectories. The initial conditions are $\overline{x}(0)$ or $\overline{x}_0$, respectively. The solution depends on f or F and on initial conditions. The region of the phase space in which all trajectories originated in a range of initial conditions converge after a transition time is called attractor. It represents the long-term behavior of the system [31], [33].

The dynamical system underlying the speech production process is very complex and its equations are not known. Nevertheless, Takens' embedding theorem [34] establishes that it is possible to reconstruct a phase space diffeomorphically equivalent to the original one from the time series of a system. The delays method is used to reconstruct the state-space vector ($\overline{s}_n$) formed by time-delayed samples of the observation (the speech signal)

$$\overline{s}_n = [s[n], s[n-T], \ldots, s\left[n-(m-1)T\right]] \tag{3}$$

where $s[n]$ is the speech signal, m is the embedding dimension of phase space reconstructed and T is the time delay. The speech signal is embedded in the reconstructed phase space. Its long-term evolution in the reconstructed phase space is called attractor. When $m > 2D+1$ (D is the attractor dimension) the reconstructed phase space is diffeomorphically equivalent to the original one.

Takens' theorem is strictly an existence theorem and does not suggest how to find the embedding dimension (m) and time delay (T). Nevertheless, T can be estimated by the first minimum of mutual information function (FMMI) [35] or by the first zero of the autocorrelation function (FZA). The false neighbors method [36] and false strands method [37] can be used to estimate m.

B. Value of First Minimum of Mutual Information Function

The mutual information function measures the mutual dependency between two variables. When these two variables are a discrete signal $s[n]$ and its delayed version $s[n+\tau]$ (being τ the delay), the mutual information function measures the quantity of information we already possess about the value of $s[n+\tau]$ if we know $s[n]$. A histogram for the probability distribution of the data is created, being p_i the probability of finding a time series value inside the ith bin of the histogram and $p_{ij}(\tau)$ the joint probability that $s[n]$ is in bin i and $s[n+\tau]$ in bin j. The mutual information estimator reads as [32]

$$I(\tau) = \sum_{i,j} p_{ij}(\tau) \ln\left[\frac{p_{ij}(\tau)}{p_i p_j(\tau)}\right]. \tag{4}$$

The FMMI function marks the delay where mutual information adds maximal information to the knowledge we have from $s[n]$.

C. Correlation Dimension (Taken's Estimator)

The correlation dimension gives an idea of the complexity of the dynamics and the attractor. More complex systems have a higher correlation dimension. In random processes, the correlation dimension is not bounded, while in deterministic systems there tends to be a finite value and it can be a non integer number (fractal dimension). The correlation dimension is given as [32]

$$D_2 = \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{\ln C(\varepsilon, N)}{\ln \varepsilon} \tag{5}$$

with $C(\varepsilon, N)$ being the correlation sum of a set of points $\overline{s}_n (n = 1 \ldots N)$ of the speech signal attractor in the reconstructed phase space

$$C(\varepsilon, N) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \Theta\left(\varepsilon - \|\overline{s}_i - \overline{s}_j\|\right) \quad (6)$$

where $\Theta(s) = 0$ if $s \leq 0$ and $\Theta(s) = 1$ if $s > 0$, which counts the number of points inside the sphere with radius ε around $\overline{s}_i$. $C(\varepsilon, N)$ is the average fraction of points within a distance of ε from any other point. Equation (6) converges very slowly as ε tends to zero. To circumvent this problem the local slope can be estimated as

$$D_2 = \frac{d \ln C(\varepsilon, N)}{d \ln \varepsilon} \cong \lim_{\Delta \ln \varepsilon \to 0} \frac{\Delta \ln C(\varepsilon, N)}{\Delta \ln \varepsilon}. \quad (7)$$

D_2 is estimated by calculating the local slope of the curve $\ln(C(\varepsilon))$ against $\ln(\varepsilon)$ when the curve has a plateau. When N is significantly large, D_2 converges with the increase of m. Since m is not known *a priori*, the convergence of D_2 is checked varying the value of m.

The application of a maximum-likelihood estimator to obtain optimal values for D_2, Takens–Theiler estimator [42] has been suggested. It can be obtained as follows

$$D_{TT}(\varepsilon) = \frac{C(\varepsilon)}{\int_0^{\varepsilon} \frac{C(\varepsilon')}{\varepsilon'} d\varepsilon'}. \quad (8)$$

D. Entropies

Entropy describes the quantity of disorder or complexity of a system. Shannon entropy estimator and Rényi entropy estimators were studied.

1) Shannon Entropy: Let us consider a system in which its output falls into the unit interval and divide it into P bins. Denote by p_i the probability that one of the outputs falls into ith bin. Entropy of a system reads as

$$H = -\sum_i p_i \ln p_i. \quad (9)$$

When H is maximum, the amount of additional information needed to specify the result of a measurement is at a maximum. If $H = 0$, then no additional information is needed. The more chaotic and nonlinear a signal is the higher is its entropy because its values fall in several different bins.

2) Rényi Entropies: Rényi entropies, based on transition probabilities, quantify the loss of information in time. Let us consider observables where the partition elements are intervals I_j of size ε. Let us introduce joint probabilities $p_{i_1, i_2, \ldots, i_m}$ that an arbitrary time n the observable falls into the interval I_{i_1} and in time $n + 1$ it falls into interval I_{i_2} and so on. Then, block entropies of block size m reads as [32]

$$H_q(m, \varepsilon) = \frac{1}{1-q} \ln \sum_{i_1, i_2, \ldots, i_m} P^q_{i_1, i_2, \ldots, i_m}. \quad (10)$$

Order q Rényi entropy is defined as

$$h_q = \sup_I \lim_{m \to \infty} \frac{1}{m} H_q(m, \varepsilon) \quad (11)$$

Rényi entropies quantify the loss of information in time in a dynamic system. In a nonchaotic system, initially nearby points in phase space will be nearby in another region of the phase space at any later point in time; therefore, Rényi entropies will tend to zero. In a chaotic system, the property of sensitivity to the initial conditions implies the divergence of nearby trajectories. It will be more difficult to predict subsequent states. More information is necessary to specify a state of the system with precision adequate for prediction. Nearby points in phase space will evolve to far points, so Rényi entropies are greater than zero. In random systems all phase space regions are possible in the long-term so Rényi entropies are infinite.

We consider the following measurements: h_2 $(q = 2)$ called correlation entropy, H_1 called first-order Rényi block entropy and H_2 called second-order Rényi block entropy for $m = 2$.

III. Databases

A. Multiquality Database

The multiquality database was recorded at the General Hospital "Doctor Negrín" in Gran Canaria (Spain) [18]. Specialist doctors diagnosed the healthy and pathological voices according to the degree of hoarseness (G) of the GRBAS scale [45]. The different levels are graded form 0 to 3. 0, 1, 2, and 3 correspond to healthy voice, light pathological (LP) voice, moderate pathological (MP) voice, and severe pathological (SP) voice, respectively.

The recordings of abnormal quality of voice were obtained from speakers with disordered speech, considering a great range of speech system pathologies (hypofunction, hyperfunction, vocal fold paralysis, vocal folds nodule, sessile polyp, pedunculated polyp, Reinke's edema, adult papiloma, carcinoma, ulcer, chronic laryngitis, etc.).

The database consists of 142 speakers, 85 healthy speakers, and 57 pathological speakers. Each sample of the database comprises the five Spanish vowels (/a/, /e/, /i/, /o/, /u/ in the International Phonetic Alphabet) pronounced in a sustained way for approximately two seconds for each vowel separated by silences. Sustained vowels were used because the voice production system uses most part of its mechanisms (e.g., glottal flux of constant air, vibration of the vocal folds in a continuous way, etc.) in the phonation of this kind of sound. This way, many types of anomalies of these mechanisms can be detected. Besides, sustained phonations are independent of the language. Additional information about sex, type, and level of pathology and labels that indicates the beginning of each vowel is stored in each sample.

Speaker voice was recorded with a conventional sound card (SoundBlaster) and a basic microphone (VIVANCO MF 15/13166 with a linear range up to 10 kHz). Speakers were recorded in a hospital room with realistic acoustic conditions, taking care that the signals presented neither inadequate level of intensity nor saturation. The speaker was situated about 25

cm from the microphone and the voice signals were digitized with a sampling rate of 22 050 Hz and 16 bits per sample.

B. MEEI Voice Disorders database

The Massachusetts Eye and Ear Infirmary (MEEI) Voice Disorders Database distributed by Kay Elemetrics [40] was also used. To the present date, it is the only database that is commercially available. It contains recordings of sustained phonation of vowel /a/, 53 healthy and 657 pathological files at a sampling rate of 25 kHz. We have considered only a subset of the database, 53 healthy, and 173 pathological voices. The selection was accomplished, identical to the subset considered by Parsa and Jamieson [41], to assure that all the files have a diagnosis, and gender and age characteristics are uniformily distributed between normal and pathological files. The duration of these vowel samples was 3 s for healthy voices and 1 s for pathological voices. All the files were down-sampled to 22 050 Hz.

IV. Experiments

In this paper, we present a study of nonlinear properties of the speech signal versus voice quality. The characteristics studied are: the Takens–Theiler estimator of correlation dimension (CD), first- and second-order Rényi entropies (RE1, RE2) correlation entropy (CE), the Shannon entropy (SE) and the value of the FMMI function. The motivation of this study is to assess the usefulness of these measures to discriminate between healthy and pathological voices. Two databases were used in order to validate and compare results: a multiquality database [18] and the MEEI database [40].

The process used in the experiment is basically the same for both databases. The next subsection is focused on the experiment with the multiquality database. The differences with the MEEI database are pointed in the second subsection.

A. Experiment With the Multiquality Database

The process used in the experiment is divided in three stages: signal preprocessing, extraction of the measurements and classification. In the signal preprocessing stage, the samples of the database (the five Spanish vowels) are normalized between -1 and 1 and the mean is removed. Then, a selection of the stable part of the phonation for each vowel is carried out. The 20% of the vowel length is eliminated, the 10% of the beginning and the 10% of the end of the vowel because they show transitory behavior. Then, the central second of each vowel is used for the experiment.

The extraction of the measurements is computed for each vowel independently because the value of measurements depends on the vowel [39]. Consequently, five sets of measurements for each sample (a set for each vowel) are computed. Each vowel is segmented into 10 equally spaced asynchronous frames (nonoverlapped) using rectangular windows. The length of each frame is 30 ms. Measurements are extracted from these frames. In the case of the Shannon entropy, the entire voice is used. Finally, the values of the measurements per frame are averaged. This way, it is obtained a value per vowel and measurement. In order to extract the measurements, the TISEAN software [38] has been used.

The delay (T) was chosen as a tradeoff between the FMMI function and the FZA function. FMMI and FZA were computed for each sample of the database. Then, the average values of FMMI and FZA were obtained. Finally, the mean value of these values was computed. The value of the delay is 8 samples (0.36 ms because the sample frequency used was 22 050 Hz). The embedding dimension (m) was varied between 1 and 10.

In the classification stage, each set of measurements per vowel is the input of a classifier, so five classifiers are used, one for each vowel. Each classifier is based on a standard neural network. They evaluate the measurements in a quantitative way and discriminate between healthy and pathological vowels. A sample of the database is diagnosed as pathological if the number of pathological vowels detected is equal to or more than three.

For each classifier, multilayer feedforward neural networks with one hidden layer are used. Supervised learning is carried out using backpropagation train algorithm. The input layer is made up of either as many inputs as characteristics (when all characteristics are evaluated combined) or is made up of one input (when only one characteristic is evaluated). The output layer has one node. The activation functions on the hidden nodes are tansigmoids (hyperbolic tangents) and the activation function of the output node is linear. The connection weights and biases are initialized according to the Nguyen–Widrow initialization algorithm [44]. The training process is stopped when a relative error of 0.005 is reached.

The database is split into a training subset and a testing subset with 70% and 30% of each type of voice, respectively. The data in the training set are z-score normalized. The test set is normalized by subtracting the training set mean and dividing by the training set standard deviation for each characteristic. The test set is normalized according to the normalization values used for the training set. The characteristics are evaluated individually and combined. The experiments were repeated 20 times, each time using different training and test sets randomly chosen.

The equal error rate (EER), the point for which the false positives rate (healthy files classified as pathological files) equals the false negative rate (pathological files classified as healthy files), is obtained varying the threshold in the output of each classifier and computing the false positive rate and false negative rate (each characteristic individually and combined).

B. Experiment With the MEEI Database

In the case of the MEEI Voice Disorders Database [40], as the vowel /a/ samples appear to include only the stable part of the phonation, the selection of the stable part was skipped in the preprocessing stage. In the classification stage, as this database is comprised of samples of vowel /a/, only one classifier was used and the EER was obtained only for vowel /a/. The delay computed in the multiquality database was applied to the MEEI database.

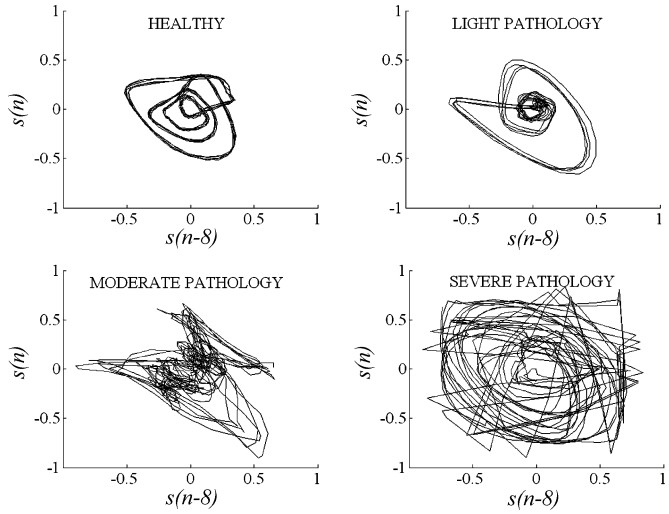

Fig. 1. Attractors of different kinds of 30-ms frame voice from vowel /a/: healthy voice (upper left), light pathological voice (upper right), moderate pathological voice (bottom left), and severe pathological voice (bottom right). Embedding dimension $m = 2$ and samples delay $T = 8$ (0.36 ms).

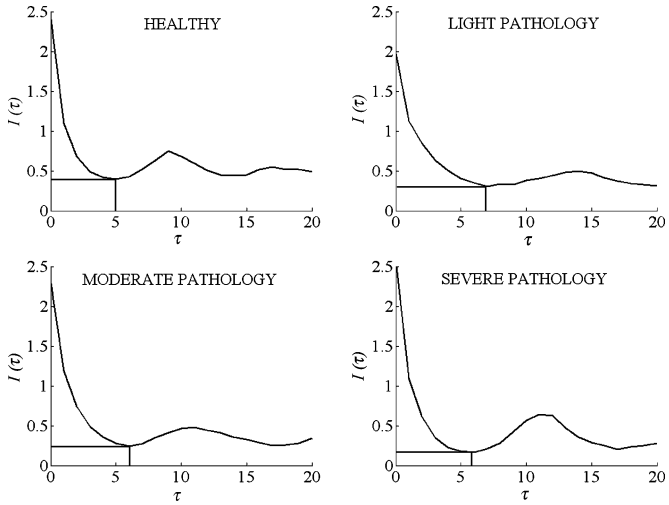

Fig. 2. Information function for each kind of 30-ms frame voice from vowel /a/: healthy voice (upper left), light pathological voice (upper right), moderate pathological voice (bottom left), and severe pathological voice (bottom right). The estimated value of the first minimum of mutual information function is marked.

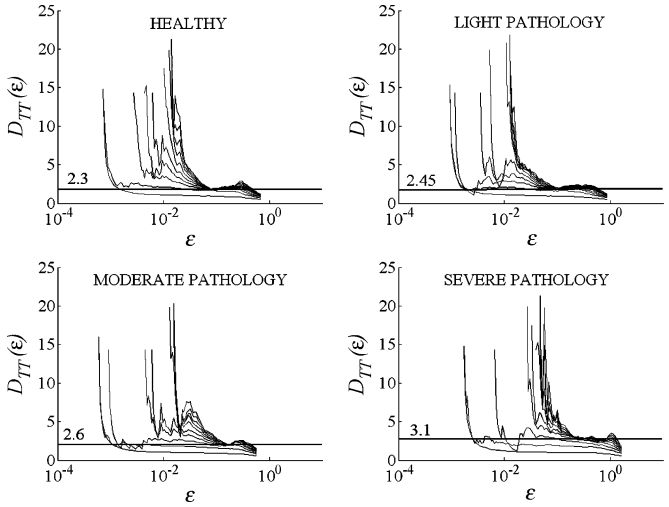

Fig. 3. Takens–Theiler estimator for each kind of 30-ms frame voice from vowel /a/: healthy voice (upper left), light pathological voice (upper right), moderate pathological voice (bottom left), and severe pathological voice (bottom right). The lower curve corresponds to embedding dimension $m = 1$ and the upper curve to $m = 10$. The straight line is the value of the Takens–Theiler estimator. The value of the estimator is written above the straight line.

V. Results

A. Study of the Discrimination of the Measurements

The graphical results for some characteristics (from multiquality database) are illustrated for a healthy frame of vowel /a/ and for different levels of pathological frames of vowel /a/ in order to observe the differences between them (Figs. 1–3) and to explain the way the measurements were computed. Besides, the data distribution of each measurement is shown in box plots (Fig. 4). Fig. 5 also shows the data distribution of each measurement for the MEEI database. This way, a comparison between different kinds of voice can be accomplished. In the next paragraphs, each illustration is discussed.

In Figs. 4 and 5, the boxes have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data. Boxes whose notches do not overlap indicate that the medians of the two groups differ at the 5% significance level.

Fig. 1 illustrates the attractors from a healthy and different levels of pathological 30-ms frames of voices of vowel /a/ (from the multiquality database [18]) with $m = 2$ and $T = 0.36$ ms. The attractor corresponding to the healthy voice is more regular than the attractor corresponding to the pathological voices. The more pathological the voice is, the more irregular is the corresponding attractor.

The curve of the mutual information function for each kind of voice is illustrated in Fig. 2. The value of the FMMI function, marked in the figure, is estimated. The Shannon entropy is estimated as the value of the mutual information function at zero time for the entire vowel. The minimum value of the mutual information between a signal and its delayed version is higher in healthy voices. This means that in the time of maximum difference (i.e., when the FMMI occurs) of a signal with its delayed version, this difference is lower in healthy voices than in pathological voices. Fig. 4 proves this fact. This figure shows the data distribution for healthy (H), pathological (P) and the different levels of pathologic voice (LP, MP, and SP) for each measurement and for /a/ vowel of the multiquality database. LP, MP, and SP voices are subgroups of P voices. According to Fig. 4, clear differences are observed among the medians of H and P voices and even between LP, MP, and SP voices. It indicates that the more pathological the voice is, the more irregular.

The CD curves (obtained with Takens–Theiler estimator) are depicted in Fig. 3. Each curve represents the values $m = 1, \ldots, 10$. The x-axis (ε) represents the scale in which the correlation dimension is computed (the size of the sphere within which the neighbors of a point in the phase space are counted for). When a plateau is found (the scaling range where the value of the Takens–Theiler estimator is independent of

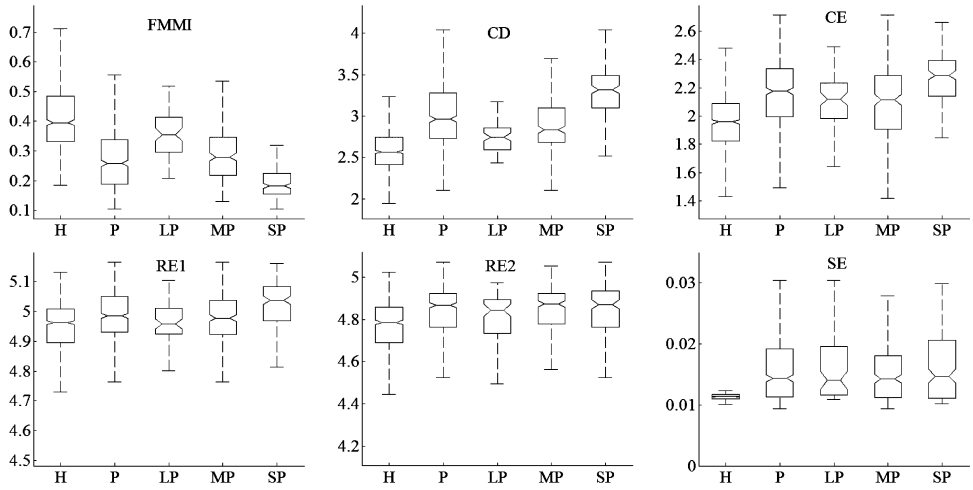

Fig. 4. Data distribution of each kind of voice (H: healthy voice, P: pathological voice, LP: light pathological voice, MP: moderate pathological voice, SP: severe pathological voice) for each measurement extracted from the /a/ vowel of the multiquality database (FMMI: first minimum of the mutual information function. CD: correlation dimension. CE: correlation entropy. RE1: first-order Rényi block entropy. RE2: second-order Rényi block entropy. SE: Shannon entropy).

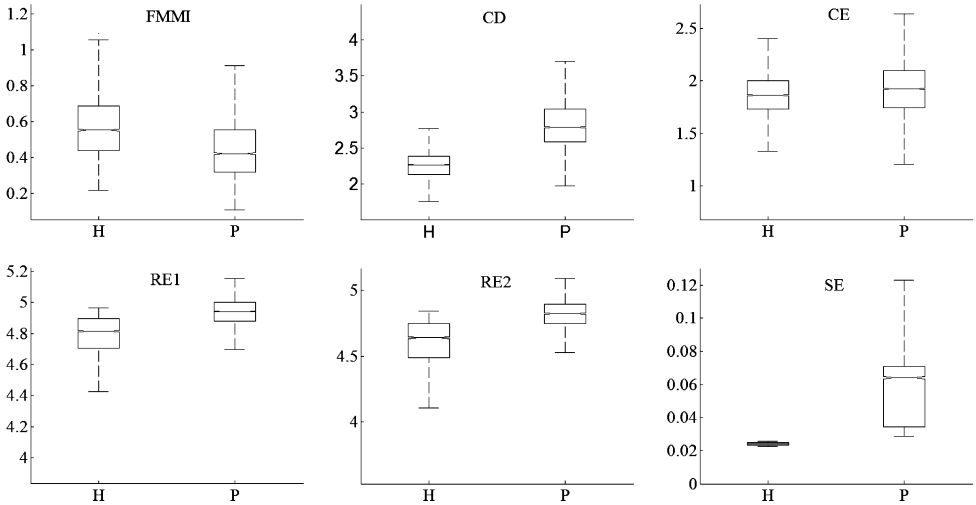

Fig. 5. Data distribution of each kind of voice (H: healthy voice, P: pathological voice) for each measurement extracted from the MEEI database (FMMI: first minimum of the mutual information function. CD: correlation dimension. CE: correlation entropy. RE1: first-order Rényi block entropy. RE2: second-order Rényi block entropy. SE: Shannon entropy).

m and ε), the value of the straight line that fits this plateau is a good CD estimator. In Fig. 3, a straight line marks the value of the correlation dimension estimated. A more complex system has a higher CD, up to infinite value for stochastic signals. In the case of frames of healthy voices the CD has a lower value than in the case of frames of pathological voices. This is an indicator of a more complex geometrical structure in a pathological voice. According to Fig. 4, the differences between the medians of H, P, LP, MP, and SP are evident. As a conclusion, CD is discriminative between H and P voices and between different kinds of pathological voices.

The same procedure as the correlation dimension estimator is used to estimate the value of the correlation entropy. Fig. 4 shows clear differences between the medians of H and P voices. The medians are higher in frames of pathological voices, as in the case of correlation dimension. This is an indicator that a P voice presents more loss of information in time than a H voice. Besides, the value of the CE is higher in SP than in MP and LP voices. The median of the MP is slightly higher than the median of LP voices. The CE is also discriminative between the different kinds of quality voices.

In Fig. 4, the RE1, RE2, and SE also show higher values in P voices than in H voices. This is an indicator of a more complex geometrical structure in pathological voices. These measurements are discriminative between H and P voices. However, they are less discriminative between the different kinds of pathological voices and even between LP and H voices (in the case of the RE1 measurement).

Fig. 5 shows the data distribution for healthy (H) and pathological (P) voices of the MEEI database. Fig. 5 shows similar results to the multiquality database in the discrimination between H and P voices. However, the medians are more separate in CD and SE than in the same measurements of the multiquality database.

The main conclusion after observing Figs. 4 and 5 is that the medians of H and P voices differ at the 5% significance level in all the cases. This leads to the application of a classification

TABLE I
EQUAL ERROR RATES FOR MULTIQUALITY DATABASE: CHARACTERISTICS INDIVIDUALLY AND COMBINED

	Equal Error Rate (%)						
Vowel	FMMI *	CD*	CE*	RE1*	RE2*	SE*	Combined
A /a/	29.23	29.17	39.84	42.69	40.00	43.69	20
E /e/	43.86	45	35.28	37.78	42.5	40.38	29.44
I /i/	49.04	50.28	50.56	48.61	44.17	39.78	33.62
O /o/	35.28	28.85	37.22	41.11	41.54	34.72	31.9
U /u/	38.18	47.5	41.54	46.11	48.69	36.75	28.08

*FMMI: First minimum of the mutual information function. CD: Correlation dimension. CE: Correlation entropy. RE1: first-order Rényi block entropy. RE2: second-order Rényi block entropy. SE: Shannon entropy.

TABLE II
SUCCESS RATES IN A CONFUSION MATRIX FOR MULTIQUALITY DATABASE

Detector's decision (%)	Actual diagnosis	
	Pathological	Normal
Pathological	81.67 ($\sigma = 7.68$)	16.73 ($\sigma = 5.19$)
Normal	18.33 ($\sigma = 7.68$)	83.27 ($\sigma = 5.19$)

method to obtain a quantitative value of the discrimination between H and P voices.

B. Classification Results

It has been shown that the studied measurements have different values for different kinds of voices. Once this fact is evident, we use a classification system in order to evaluate the discriminative usefulness of the characteristics against two voice qualities (healthy and pathological voices). In the next lines the results obtained using the classification system are discussed.

The results for the multiquality database [18] are shown in Table I. The EER is shown for each vowel and each characteristic individually and combined. These results were obtained after evaluating the characteristics with different numbers of neurons in the hidden layer of the neural network. Finally, the best results were obtained with 60 neurons in the hidden layer. According to Table I, combination of all characteristics yields the lowest EER for each vowel.

Once a threshold for each vowel has been chosen, the performance of the system is computed. Then a voice is diagnosed as pathological if the number of pathological vowels is equal to or more than three. Table II shows the confusion matrix of the system with the mean and standard deviation values obtained averaging the results for each individual experiment. The averaged global success of the system is 82.47% with a standard deviation (σ) of 3.1.

As the multiquality database is labeled with four levels of quality, a deeper study of the classification rates is carried out. As the level of pathology of each sample is known, a separation of the percentage of each of the three pathological levels classified as pathological voices and wrongly classified as normal can be made. Table III shows the confusion matrix with the scores of different levels of pathologies added: percentage of LP speaker classified as pathological speaker, percentage of MP speaker classified as pathological speaker and percentage of SP speaker classified as pathological speaker. The classification rate decreases in LP speakers since they are more likely to be confused with healthy speakers.

TABLE III
SUCCESS RATES IN A CONFUSION MATRIX FOR MULTIQUALITY DATABASE: LEVELS OF VOICE PATHOLOGIES ADDED

Detector's decision (%)	Actual diagnosis			
	LP*	MP *	SP*	Normal
Pathological	63.33(σ=30)	77.22(σ=9.85)	97.50(σ = 6.10)	16.73(σ=5.19)
Normal	36.67(σ=30)	22.78(σ=9.85)	2.50(σ = 6.10)	83.27(σ=5.19)

*LP: light pathological voice, MP: moderate pathological voice, SP: severe pathological voice.

TABLE IV
EQUAL ERROR RATES FOR MEEI DATABASE: CHARACTERISTICS INDIVIDUALLY AND COMBINED

	Equal Error Rate (%)						
Vowel	FMMI *	CD*	CE*	RE1*	RE2*	SE*	Combined
A /a/	38.81	8.66	41.96	44.42	41.25	3.12	0.31

*FMMI: First minimum of the mutual information function. CD: Correlation dimension. CE: Correlation entropy. RE1: first-order Rényi block entropy. RE2: second-order Rényi block entropy. SE: Shannon entropy.

TABLE V
SUCCESS RATES IN A CONFUSION MATRIX FOR MEEI DATABASE

Detector's decision (%)	Actual diagnosis	
	Pathological	Normal
Pathological	99.69 ($\sigma = 0.12$)	0.31 ($\sigma = 0.12$)
Normal	0.31 ($\sigma = 0.12$)	99.69 ($\sigma = 0.12$)

In the case of the MEEI database the same procedure was followed. In this case, the best results were obtained using ten neurons in the hidden layer. Table IV shows the EER for each characteristic individually and combined. Shannon entropy and correlation dimension are the characteristics that show the better EER. Consequently, they show the better success rate. According to Table V, in which the confusion matrix for MEEI database is shown for the EER point, the success rate for all characteristics combined is 99.69% with a standard deviation (σ) of 0.2.

C. Comparison With Works in Literature

The global results obtained in this paper are compared with some results obtained in the literature in Table VI. This table lists some studies on automatic detection and classification of voice pathologies using different databases (the number of normal and pathologic samples is detailed), measurements and classification techniques (the results of the present paper are added).

The classification rates obtained in these works are similar to the present one for the multiquality database. For example, a set of classical measurements (jitter, shimmer, etc.) are studied with a classification based on neural network technique in Linder [49] and a 80% classification rate was obtained. Shimmer, periodicity, spectral and chaos measurements, neural network technique, and an extended version of our multiquality database were used in Alonso [24] with a global performance of 92.76%. In our work, using only six chaos measurements the success rate is fairly good with the multiquality database (82.87%).

In the case of the results obtained with the MEEI database, the comparison is only fair in the case of Parsa (98.7%) [41], [48] and Saénz–Lechón (89.6%) [43] works because the same

TABLE VI
COMPARISON WITH SOME RESEARCH WORKS ON VOICE PATHOLOGY DETECTION

First author	Database (normal+pathologic)	Measurements	Classifier	Best results (%)
Linder [49]	8 + 112	Jitter, Shimmer, standard deviation of fundamental frequency, glottal-to-noise excitation ratio	Neural Network	80
Wallen [10]	9 + 20	Perturbation, cepstral, LPC	Multi-layer perceptron	85.8
Boyanov [9]	50 + 150	Perturbation, noise, energies	K-nearest neighbours, linear discriminant analysis, self-organized maps	93.5
Alonso [24]	100 + 68	Noise, Periodicity, Spectral, shimmer, chaos	Neural Network	92.76
Current paper	85 + 57	Chaos	Neural Network	82.47
Parsa [41][48]	MEEI 53+173	Noise	Linear discriminant analysis	98.7
Saénz-Lechón [43]	MEEI 53+173	Mel-Frequency Cepstral Coefficients	Multi-layer perceptron	89.6
Hadjitodorov [12]	MEEI 53+638	Perturbation, noise	K-nearest neighbors	96.1
Current paper	MEEI 53+173	Chaos	Neural Network	99.69

data subset was used and the methodology is quite similar to the one used in this paper. The global classification rate for the MEEI database in this work (99.69%) is the highest found in the literature (comparing only the works with the same data subset and similar methodology).

VI. CONCLUSION

The usefulness of six nonlinear chaotic characteristics: first- and second-order Rényi entropies, the correlation entropy, the correlation dimension, the value of the first minimum of mutual information function and Shannon entropy, has been evaluated in order to distinguish between two voice qualities (healthy and pathological voices). Two databases were used to evaluate the characteristics, a multiquality database [18] and a commercial one (MEEI Voice Disorders [40]) in order to obtain comparative results between them. The multiquality database comprises samples labeled with different kinds of voice quality according to the hoarseness (G) of the GRBAS scale. The MEEI database only has samples labeled as healthy and pathological voices.

A previous statistical study was carried to check the discrimination of the characteristics for both databases. In the multiquality database, the statistical study showed remarkable differences between healthy and pathological voices and even between the three different levels of pathologies for each of the characteristics studied. Generally, the quantitative evaluation of the measurements was correlated with the medical assessment. The correlation dimension and the value of the first minimum of mutual information function were the characteristics that better discriminated among the different voice qualities of the multiquality database. The MEEI database also showed significative differences between the medians of the two classes of voice (healthy and pathological voices). As a conclusion, the medians of the healthy and pathological voices differ at the 5% significance level in both databases and in all the characteristics.

The characteristics were evaluated with neural networks to discriminate between healthy and pathological voices. Successful results were obtained for both databases. The global success rate obtained with the multiquality database [18] was 82.47% $(\sigma = 3.1)$ and with the MEEI database [40] 99.69% $(\sigma = 0.2)$. This demonstrates that the six proposed characteristics are useful to discriminate between healthy and pathological speakers.

The difference between the two classification rates of both databases is due to the nonexistence of LP speakers in the MEEI database. LP speakers are more likely to be classified as normal speakers because the difference between healthy and LP or MP speakers is lower than between a healthy and an MP or an SP speaker. In the results, if the LP and MP speakers are removed, the success rate of the multiquality database is similar to the MEEI database (see SP speaker in Table III and compare with the MEEI results in Table V).

The measurements studied in this research can be used to document the patient evolution. They can also be used in help systems for pathology diagnosis in the speech production system. As a new step, we propose the combination of the nonlinear characteristics evaluated here and classical characteristics used previously in order to evaluate if the combination results in better classification rates.

REFERENCES

[1] B. Boyanov, S. Hadjotodorov, B. Teston, and D. Doskov, "Robust hybrid pitch detector for pathological voice analysis," in *Proc. Larynx 97*, 1997, pp. 55–58.

[2] B. Boyanov, T. Ivanov, and G. Cholet, "Robust hybrid pitch detector," *Electron. Lett.*, vol. 29, no. 22, pp. 1924–1926, 1993.

[3] H. Kasuya, Y. Endo, and S. Saliu, "Novel acoustic measurements of jitter and shimmer characteristics from pathologic voice," in *Proc. Eurospeech'93*, Berlin, Germany, 1993, pp. 1973–1976.

[4] C. Ludlow, C. Bassich, N. Connor, D. Coulter, and Y. Lee, "The validity of using phonatory jitter and shimmer to detect laryngeal pathology," in *Laryngeal Function in Phonation and Respiration*. Boston, MA: Brown, 1987, pp. 492–508.

[5] E. Yumoto and W. Gould, "Harmonics to noise ratio as an index of the degree of hoarseness," *J. Acoust. Soc. Amer.*, vol. 71, no. 6, pp. 1544–1550, 1982.
[6] M. Yunik and B. Boyanov, "Method for evaluation of the noise-to-harmonic- component ratios in pathological and normal voices," *Acustica*, vol. 70, pp. 89–91, 1990.
[7] H. Kasuya, S. Ogawa, K. Mashie, and S. Ebihara, "Normalized noise energy as an acoustic measure to evaluate pathologic voice," *J. Acoust. Soc. Amer.*, vol. 80, no. 5, pp. 1329–1334, 1986.
[8] M. Frohlich, D. Michaelis, and H. W. Srube, "Acoustic ‘Breathiness Measures' in the description of pathologic voices," in *Proc. Acoust., Speech, Signal Process. ICASSP'98*, Seattle, WA, 1998, vol. 2, pp. 937–940.
[9] B. Boyanov and S. Hadjitodorov, "Acoustic analysis of pathological voices: A voice analysis system for screening of laryngeal diseases," *IEEE Eng. Med. Biol. Mag.*, vol. 16, pp. 74–82, Jul./Aug. 1997.
[10] E. J. Wallen and J. H. L. Hansen, "A screening test for speech pathology assessment using objetive quality measures," in *Proc. Int. Conf. Spoken Lang. Process.*, Philadelphia, PA, Oct. 1996, vol. 2, pp. 776–779.
[11] M. Rosa, J. C. Pereira, and A. C. P. L. F. Carvalho, "Evaluation of neural classifiers using statistic methods for identification of Laryngeal pathologies," in *Proc. 5th Brazilian Symp. Neural Netw.*, Dec. 1998, vol. 1, pp. 220–225.
[12] S. Hadjitodorov and P. Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening," *Med. Eng. Phys.*, vol. 24, pp. 419–429, 2002.
[13] S. Ayache, M. Ouaknine, P. Dejonkere, P. Prindere, and A. Giovanni, "Experimental study of the effects of surface mucus viscosity on the Glottic cycle," *J. Voice*, vol. 18, 2004.
[14] H. M. Teager and S. M. Teager, "Evidence of nonlinear sound production mechanisms in the vocal tract," in *Proc. NATO Adv. Study Inst. Speech Production Modeling*, Bonas, France, 1989, pp. 241–26.
[15] D. A. Berry, H. Herzel, I. R. Titze, and K. Krische, "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," *J. Acoust. Soc. Amer.*, vol. 95, pp. 3595–3604, Jun. 1994.
[16] J. J. Jian, Y. Zhang, and J. Stern, "Modeling of chaotic vibrations in symmetric vocal folds," *J. Acoust. Soc. Amer.*, vol. 110, pp. 2120–2128, Jun. 2001.
[17] I. Steinecke and H. Herzel, "Bifurcations in an asymmetric vocal-fold model," *J. Acoust. Soc. Amer.*, vol. 97, pp. 1879–1884, Mar. 1995.
[18] J. B. Alonso, J. de León, I. Alonso, and M. A. Ferrer, "Automatic detection of pathologies in the voice by hos based parameters," *EURASIP J. Appl. Signal Process.*, vol. 2001, pp. 275–284, 2001.
[19] B. Boyanov, S. Hedjitodorov, and T. Ivanov, "Analysis of voiced speech by means of Bispectrum," *Electron. Lett.*, vol. 27, pp. 2267–2268, 1991.
[20] L. Gavidia-Ceballos, J. H. L. Hansen, and J. F. Kaiser, "Vocal fold pathology assessment using AM autocorrelation analysis of the Teager energy operator," in *Proc. 4th Int. Conf. Spoken Lang., ICSLP'96*, 1996, vol. 2, pp. 757–760.
[21] J. H. L. Hansen, L. Gavidia-Ceballos, and J. F. Kaiser, "A nonlinear operator-based speech feature analysis method with application to vocal fold pathology assessment," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 3, pp. 300–313, Mar. 1998.
[22] D. A. Cairns and J. H. L. Hansen, "A noninvasive technique for detecting hypernasal speech using a nonlinear operator," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 35–44, Jan. 1996.
[23] D. A. Cairns, J. H. L. Hansen, and J. F. Kaiser, "Recent advances in hypernasal speech detection using the nonlinear Teager energy operator," in *Proc. Int. Conf. Spoken Lang. Process., ICSLP'96*, 1996, vol. 2, pp. 780–783.
[24] J. B. Alonso, F. Díaz-de-María, C. M. Travieso, and M. A. Ferrer, "Using nonlinear features for voice disorder detection," in *Proc. 3rd Int. Conf. Nonlinear Speech Process.*, Barcelona, Spain, Apr. 2005, pp. 94–106.
[25] P. Maragos, A. G. Dimakis, and I. Kokkinos, "Some advances in nonlinear speech modeling using modulations, fractals, and chaos," in *Proc. 14th Int. Conf. Digital Signal Process.*, 2002, vol. 1, pp. 325–332.
[26] P. Yu, M. Ouaknine, J. Revis, and A. Giovanni, "Objective voice analysis for dysphonic patients: A multiparametric protocol including acoustic and aerodynamic measurements," *J. Voice*, vol. 15, pp. 529–542, 2001.
[27] Y. Zhang and J. J. Jiang, "Nonlinear dynamic analysis in signal typing of pathological human voices," *Electron. Lett.*, vol. 39, pp. 1021–1023, Jun. 2003.
[28] J. J. Jiang and Y. Zhang, "Nonlinear dynamic analysis of speech from pathological subjects," *Electron. Lett.*, vol. 38, pp. 294–295, March 2002.
[29] A. Aaccardo, F. Fabbro, and E. Mumolo, "Analysis of normal and pathological voices via short-time fractal dimension," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Oct./Nov. 1992, vol. 14, pp. 1270–1271.
[30] M. Banbrook and S. Mc Laughlin, "Dynamical modelling of vowel sounds as a synthesis tool," in *Proc. 4th Int. Conf. Spoken Lang. Process.*, Oct. 1996, vol. 3, pp. 1981–1984.
[31] M. A. Little, P. E. McSharry, I. M. Moroz, and S. J. Roberts, "Stroboscopic method for detecting complex dynamics in disordered speech," *IEEE Trans. Biomed. Eng.*, vol. 1, pp. 1–6, Jan. 2006.
[32] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1997.
[33] G. L. Baker and J. P. Gollub, *Chaotic Dynamics: An Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 1990.
[34] F. Takens, "Detecting strange attractors in turbulence," in *Lecture Notes in Math.*. New York: Springer, 1981, vol. 898, pp. 366–381.
[35] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," *Phys. Rev. A*, vol. 33, p. 1134, 1986.
[36] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, pp. 3403–3411, 1992.
[37] M. B. Kennel and H. Abarbanel, "False neighbours and false strands: A reliable minimum embedding dimension algorithm," *Phys. Rev. E*, vol. 66, 2002.
[38] R. Hegger, H. Kantz, and T. Schreiber, "Practical implementation of nonlinear time series methods: The TISEAN package," *Chaos*, vol. 9, pp. 413–435, Jun. 1999.
[39] J. Koreman and M. Pützer, "Finding correlates of vocal fold adduction deficiencies," in *PHONUS*, W. J. Barry and J. Koreman, Eds. Saarbrucken: Inst. of Phonetics, Univ. of the Saarland, 1997, vol. 3, pp. 155–178.
[40] Massachusetts Eye and Ear Infirmary, Voice Disorders Database, Version 1.03. Kay Elemetrics Corp., Lincoln Park, NJ, 1994, CD-ROM.
[41] V. Parsa and D. G. Jamieson, "Identification of pathological voices using glottal noise measures," *J. Speech Lang. Hearing Res.*, vol. 43, pp. 469–485, Apr. 2000.
[42] J. Theiler, "Lacunarity in a best estimator of fractal dimension," *Phys. Lett. A 135*, vol. 133, pp. 195–200, 1988.
[43] N. Saénz-Lechón, J. I. Godino-LLorente, V. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomed. Signal Process. Control*, vol. 1, no. 2, pp. 120–128, Apr. 2006.
[44] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer eural networks by choosing initial values of the adaptive weights," in *Proc. Int. Joint Conf. Neural Netw.*, 1990, vol. 3, pp. 21–26.
[45] M. Hirano, *Clinical Examination of Voice*. New York: Springer Verlag, 1981.
[46] T. Nawka, L. C. Anders, and J. Wendler, "Die auditive Beurteilung heiserer Stimmen nach dem RBH System," *Sprache Stimme Gehör*, vol. 18, pp. 130–133, 1994.
[47] L. Gu, J. G. Harris, R. Shrivastav, and C. Sapienza, "Disordered speech assessment using automatic methods based on quantitative measures," *EURASIP J. Appl. Signal Process.*, vol. 9, pp. 1400–1409, 2005.
[48] V. Parsa and D. G. Jamieson, "Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech," *J. Speech Lang. Hearing Res.*, vol. 44, pp. 327–339, 2001.
[49] R. Linder, A. E. Albers, M. Hess, S. J. Pöppl, and R. Schönweiler, "Artificial neural network-based classification to screen for dysphonia using psychoacoustic scaling of acoustic voice features," *J Voice*, vol. 22, pp. 155–163, Mar. 2008.

Patricia Henríquez was born in Las Palmas de Gran Canaria, Spain. She received the Telecommunication Engineering degree in 2006 from the University of Las Palmas de Gran Canaria, Las Palmas, Spain. She is currently pursuing the Ph.D. degree from the University of Las Palmas de Gran Canaria.

Her research areas include nonlinear processing, speech processing, and pattern recognition.

Jesús B. Alonso received the B.Sc. degree in telecommunication engineering and the Ph.D. degree, both from the University of Las Palmas de Gran Canaria (ULPGC), Las Palmas, Spain, in 2001 and 2006, respectively.

He has been an Associate Professor at ULPGC since 2002, teaching subjects on signal processing. His research interests lie in the fields of pattern recognition, data mining, and speech recognition.

Miguel A. Ferrer was born in Spain in 1965. He received the M.Sc. degree in telecommunications and Ph.D., both from the Universidad Politécnica de Madrid, Madrid, Spain, in 1988 and 1994, respectively.

He is an Associate Professor at the Universidad de Las Palmas de Gran Canaria, Las Palmas, Spain, where he has taught since 1990 and heads the Digital Signal Processing Group. His research interests lie in the fields of biometrics and audio quality evaluation.

He is a member of the IEEE Carnahan Conference on Security Technology Advisory Committee.

Carlos M. Travieso received the M.Sc. degree in telecommunication engineering from the Polytechnic University of Catalonia (UPC), Catalonia. Spain, in 1997 and the Ph.D. degree from the University of Las Palmas de Gran Canaria (ULPGC), Las Palmas, Spain, in 2002.

He has been an Assistant Professor with ULPGC since 2001, teaching subjects on signal processing. His research lines are biometrics, classification system, environmental intelligence, and data mining. He has worked on International and Spanish Research Projects, with papers published in international journals and conferences.

Dr. Travieso is an Image Processing Technical IASTED Committee Member.

Juan I. Godino-Llorente was born in Madrid, Spain. He received the M.Sc. degree in communications engineering and the Ph.D. degree with honors in computer science from the Universidad Politécnica de Madrid, Madrid, Spain, in 1996 and 2002, respectively.

Currently, he is an Associate Professor at the Universidad Politécnica de Madrid. Also, he is the Head of the Circuits and Systems Engineering Department, which belongs to the same university. His main research areas are in the field of biomedical signal and image processing, with applications to voice disorders and ECG signal processing.

Fernando Díaz-de-María received the Telecommunication Engineering degree and the Ph.D. degree from the Polytechnic University of Madrid, Madrid, Spain, in 1991 and 1996, respectively.

From 1991 to 1995, he was a Lecturer at the Universidad de Cantabria, Santander, Spain. From 1995 to 1996, he was an Assistant Professor at the Universidad Politécnica de Madrid. Since 1996, he has been an Associate Professor in the Department of Signal Processing and Communications, Universidad Carlos III de Madrid, Madrid, Spain. Since 1997, he has held several offices in both his department and the university. His primary research interests include robust speech processing, video coding, and multimedia indexing. He has led numerous projects and contracts in the mentioned fields. He is a coauthor of several papers in prestigious international journals, two chapters in international books, and quite a few papers in revised national and international conferences.