Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal Diseases

Juan Rafael Orozco-Arroyave, Elkyn Alexander Belalcazar-Bolaños, Julián David Arias-Londoño, Jesús Francisco Vargas-Bonilla, Sabine Skodda, Jan Rusz, Khaled Daqrouq, Florian Hönig, and Elmar Nöth

***Abstract*—This paper evaluates the accuracy of different characterization methods for the automatic detection of multiple speech disorders. The speech impairments considered include dysphonia in people with Parkinson's disease (PD), dysphonia diagnosed in patients with different laryngeal pathologies (LP), and hypernasality in children with cleft lip and palate (CLP). Four different methods are applied to analyze the voice signals including noise content measures, spectral-cepstral modeling, nonlinear features, and measurements to quantify the stability of the fundamental frequency. These measures are tested in six databases: three with recordings of PD patients, two with patients with LP, and one with children with CLP. The abnormal vibration of the vocal folds observed in PD patients and in people with LP is modeled using the stability measures with accuracies ranging from 81% to 99% depending on the pathology. The spectral-cepstral features are used in this paper to model the voice spectrum with special emphasis around the first two formants. These measures exhibit accuracies ranging from 95% to 99% in the automatic detection of hypernasal voices, which confirms the presence of changes in the speech spectrum due to hypernasality. Noise measures suitably discriminate between dysphonic and healthy voices in both databases with speakers suffering from LP. The results obtained in this study suggest that it is not suitable to use every kind of features to model all of the voice pathologies; conversely, it is necessary to study the physiology of each impairment to choose the most appropriate set of features.**

***Index Terms*—Hypernasality, laryngeal pathologies (LP), noise measures, nonlinear behavior, Parkinson's disease (PD), periodicity, spectral-cepstral modeling, stability.**

Manuscript received April 10, 2015; revised July 9, 2015; accepted August 7, 2015. Date of publication; date of current version. This work was supported by COLCIENCIAS through Project 111556933858, by the Deanship of Scientific Research (DSR), King Abdulaziz University, under Grant 9-135-1434-HiCi, and by CODI, "estrategia de sostenibilidad 2014-2015 from Universidad de Antioquia." The work of J. R. Orozco-Arroyave was supported by COLCIENCIAS under grants of "Convocatoria 528 para estudios de doctorado en Colombia, generación del bicentenario, 2011." The work of J. Rusz was supported by the Czech Science Foundation (GACR 102/12/2230).

J. R. Orozco-Arroyave is with the Faculty of Engineering, Universidad de Antioquia UdeA, 1226 Medellín, Colombiaand also with the Pattern Recognition Laboratory, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany (e-mail: rafael.orozco@i5.informatik.uni-erlangen.de).

E. A. Belalcazar-Bolaños, J. D. Arias-Londoño, and J. F. Vargas-Bonilla are with the Faculty of Engineering, Universidad de Antioquia UdeA, 1226 Medellín, Colombia (e-mail: elkyn.belalcazar@udea.edu.co; julian.ariasl@udea.edu.co; jesus.vargas@udea.edu.co).

S. Skodda is with the Department of Neurology, Knappschaftskrankenhaus, Ruhr University Bochum, 44801 Bochum, Germany (e-mail: Sabine.Skodda@kk-bochum.de).

J. Rusz is with the Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, 16636 Praha, Czech Republic (e-mail: ruszjan@fel.cvut.cz).

K. Daqrouq is with the Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 22254, Saudi Arabia (e-mail: haleddaq@yahoo.com).

F. Hönig is with the Pattern Recognition Laboratory, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany (e-mail: hoenig@informatik.uni-erlangen.de).

E. Nöth is with the Pattern Recognition Laboratory, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany, and also with the Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 22254, Saudi Arabia (e-mail: noeth@informatik.uni-erlangen.de).

Digital Object Identifier 10.1109/JBHI.2015.2467375

I. Introduction

DIFFERENT challenges have been addressed in automatic speech processing including intelligibility assessment, speaker verification/identification, and recognition of paralinguistic phenomena in speech such as emotions and pathologies. One of the aims of pathological speech processing is the development of computer-aided tools, enabling the objective assessment of voice. The studies found in the literature consider many different characteristics of speech including spectral and cepstral modeling, perturbation measurements [such as jitter, shimmer, amplitude perturbation quotient (APQ), and pitch perturbation quotient (PPQ)], noise content measures, prosodic features, and nonlinear behavior [1]. Typically, all these measurements are merged in the same representation space in order to obtain high recognition rates; however, the interpretation of those results is not completely clear. With the aim of advancing the interpretation and analysis of different voice pathologies using different characterization methods, this paper considers voice registers of six databases with recordings of sustained phonations of speakers with pathologies of three different origins: *laryngeal*, *functional*, and *neurological*. The symptoms in the voice of patients with laryngeal pathologies (LP) are mainly related to breathy voice, hoarseness, and abnormal vibration of the vocal folds due the presence of polyps and/or nodules [2]. On the other hand, one of the voice pathologies with functional origins is hypernasality, which is the most common feature in the voice of patients with cleft lip and palate (CLP). This pathology causes the patient to produce voice with excess of nasalization, which results from the inappropriate control of the velum, generating abnormal resonances in the vocal and nasal cavities [3]. The impaired laryngeal function has been clinically observed in these patients [4]. Regarding the neurological disorders, Parkinson's disease (PD) is one of the most common. The voice of PD patients is characterized among others, by excess of tremor, reduced loudness, monotonicity, hoarseness [5], [6].

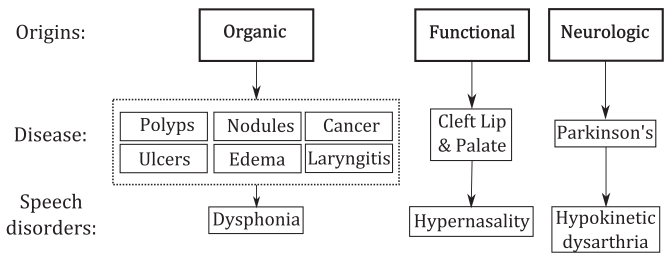

Fig. 1. Speech pathologies grouped according to their origin.

Since the characteristics and symptoms of the pathologies described above are mostly different, it is worth to address different strategies to model each kind of pathology. In this paper, we consider four different characterization approaches to model these pathologies separately; thus, the paper is a step forward to show which characterization approach is able to reflect characteristics of specific vocal pathologies. The sets of considered features include measurements to quantify the noise content in the phonations, features to model the voice spectrum and the cepstrum, measures to assess the stability and periodicity of the vocal sounds, and the nonlinear behavior of the signals. Each set of features addresses specific symptoms or phenomena of vocal pathologies, enabling us to perform more detailed analyses regarding the voice production process. These sets of features are used in this paper to perform the automatic discrimination of pathological and healthy speakers. The accuracies obtained in the experiments are used to analyze which feature set is most suitable to model each pathology.

Fig. 1 summarizes the diseases and the resulting speech pathologies studied in this paper. The characteristics of each speech disorder are explained below.

Neurological diseases can generate several speech impairments in the patients. PD is a common neurological disorder that affects about 2% of the people older than 65 years [7]. It is characterized by the loss of dopaminergic neurons in the substantia nigra of the midbrain and its main symptoms include resting tremor, bradykinesia, rigidity, and postural instability [8]. According to the literature, the majority of patients with PD feature some voice and speech impairments including reduced loudness, monopitch, monoloudness, reduced stress, breathy, hoarse voice quality, and imprecise articulation [5], [6]. As we are only considering sustained phonation of vowels here, this study mainly addresses the symptoms related with PD dysphonia. The analysis of voice signals from PD patients has considered several techniques including perturbation measures [9], energy content [10], nonlinear dynamics (NLD) [11], and combination of several techniques [12]. Table I includes some of the studies that have considered recordings of people with PD.

Functional pathologies are associated with problems controlling different limbs and/or muscles involved in the speech production process. CLP is one of the most prevalent craniofacial malformations and it generates different functional problems in the vocal tract. About 1 in every 1000 children born with CLP [19]. The speech of children with CLP is disordered even after surgery and shows abnormal characteristics such as hypernasality or hyponasality, glottal stops, backing, and weakening of consonants [20]. According to previous studies, hypernasality appears in approximately 90% of patients with CLP [20]. Hypernasality has been evaluated typically using acoustic analyses including perturbation and noise measures, spectral characteristics, and NLD features. In [21], the authors show the suitability of noise measures and Mel-frequency cepstral coefficients (MFCC) for hypernasality detection in the five Spanish vowels. In [22], different pronunciation and articulation features along with MFCC are used to evaluate speech disorders in recordings of German children with CLP. In [23], the authors improve the resolution of the speech spectrum using the modified group delay functions (MGDF) to find a peak located in 250 Hz. They found that this spectral peak has a higher intensity in hypernasal voices than in healthy ones. Additionally, according to [24], velopharyngeal insufficiency or incompetence suffered by CLP patients causes them to make compensatory movements in the vocal tract, generating glottal stops and abnormal vibration of the glottis. Such vibration can produce nonlinear behavior in the vocal tract of CLP speakers [25], which can be modeled using different NLD features [21]–[26]. A summary of the database, features, and highest accuracy reported in several studies that have addressed the automatic detection of hypernasality in speech signals are provided in Table II.

TABLE I
SELECTION OF STUDIES CONSIDERING PD

	Database	Features	Highest Accuracy (%)
[12]	Private - English	Stability, noise, and NLD	97.7
[13]	Private - Czech[a]	Acoustic, prosodic, and two-mass model features	90.5
[14]	Private - Czech	Acoustic, prosodic, and spectral features	80.0
[15]	Private - Spanish [a]	NLD	76.8
[16]	Private - German[a]	Acoustic, prosodic modeling, glottal excitation	81.9
[17]	Private - Spanish[a]*	Energy of unvoiced sounds	99.0
	Private - German[a]	Energy of unvoiced sounds	97.0
	Private - Czech[a]	Energy of unvoiced sounds	97.0
[10]	Private - Spanish[a]	Spectral and cepstral measures	90.0

[a] Considered in this paper.
*PC-GITA [18].

TABLE II
SELECTION OF STUDIES CONSIDERING HYPERNASALITY

	Database	Features	Highest Accuracy (%)
[27]	Private	TEO	94.4
[23]	Private[b]	MGDF	100
[28]	Private	Prosodic, MFCC and TEO	75.8
[29]	Private[b]	High order LPC	100
[21]	Private[a]	Acoustic and NLD	90.6
[26]	Private[a]	NLD	92.1

[a] Considered in this paper.
[b] Non repaired CLP i.e., patients without surgery.
TEO: Teager energy operator, LPC: linear prediction coefficients.

LP are characterized by increase of mass, lack of closure, and changes in elasticity of the vocal folds. The most common disorders in speech induced by LP include, among others, dysphonia,

TABLE III
SELECTION OF STUDIES CONSIDERING LARYNGEAL PATHOLOGIES

	Database	Features	Highest Accuracy (%)
[32]	Private	Stability, noise, and cepstrum	87.7
[33]	MEEI	Pitch, stability, and noise	96.5
[30]	MEEI	Stability and noise	96.1
[34]	MEEI	Spectral and energy	93.2
[35]	Private	Spectral	92.0
[36]	MEEI	MFCC	94.0
[37]	MEEI	Measures from MDVP[a] and MFCC	98.3
[38]	MEEI	Music information retrieval features	92.1
[39]	MEEI	Acoustic and noise	92.3
[40]	MEEI	Noise and MFCC	96.7
	Private[b]	MFCC	82.1
[41]	MEEI	NLD	98.2

[a] Multi-Dimensional Voice Program (MDVP).
[b] Considered in this paper.
MEEI: Massachusetts Eye and Ear Infirmary.

breathiness, and hoarseness [2]. Several approaches have been developed for the screening of LP and most of the applied parameters are based on long-term signal analysis, i.e., sustained vowels [30]. This analysis is based on averaging local perturbations and it can be divided into three categories: amplitude perturbation, frequency perturbation, and noise measures. The noise analysis has shown to be suitable for detecting voice disorders, since most of the pathological voices contain noise in some extent. The literature covers different noise measurements in voice such as: signal-to-noise ratio, harmonics-to-noise ratio (HNR), normalized noise energy (NNE), voice turbulence index (VTI), soft phonation index (SPI), and glottal to noise excitation ratio (GNE). In voice signals with LP, the nonlinear behavior is associated with an abnormal vocal fold collision, increased pressure-flow in the glottis, and stress–strain curves of vocal fold tissue [31]. A selection of studies that have considered recordings of people with LP is provided in Table III. The summary includes data of the database, features, and the highest accuracy obtained in the automatic discrimination of pathological and healthy speakers.

With the aim of showing which sets of features are more suitable to discriminate between healthy speakers and people with different kind of pathologies, four characterization approaches are tested on six databases with recordings of healthy speakers and people suffering from diseases with several origins including laryngeal, functional, and neurological. Different aspects in the voice are modeled including noise content, periodicity, spectral-cepstral features, and nonlinear behavior.

II. METHODOLOGY

Fig. 2 depicts the methodology applied in this study. The stages of the process are explained in the next sections.

A. Preprocessing and Characterization

The recordings are considered in short-time frames analysis using Hamming windows with different lengths depending on the estimated features. The length and time shift is described when each feature set is introduced. After the windowing process, several features are extracted from the voice frames.

Noise measures: The presence of noise in speech is defined as the existence of glottal noise in the signal during the phonation due to an incomplete closure of the vocal folds [42]. A set with six measures is calculated in order to perform a detailed characterization of the noise content in the voice signals. The voice recordings are windowed with Hamming windows of 40-ms length and 20-ms time shift. The set of measures includes HNR, which is a measure of the ratio between the harmonic energy of the signal and its noise content [42]; cepstral version of HNR; VTI, which allows the estimation of the turbulence components in voice that appear due to the incomplete abduction of the vocal folds [42]; SPI, to evaluate the poorness of high-frequency harmonic components [43]; NNE, to measure the energy of the noise in the voice relative to the total energy of the signal [42]; and GNE, which measures the amount of excitation in voice due to the vibration of the vocal folds relative to the excitation noise due to the turbulences in the vocal tract [44].

Periodicity and stability of voice: These properties refer to the ability to generate constant airflow during the production of sustained vowels [45]. In this feature set, two different windowing lengths are used. For one group of measures, windows of 40-ms length and 20-ms time shift are used. This group includes the variation of the pitch period amplitude (shimmer) and the variation of the cycle-to-cycle pitch period (jitter). For the second group of measures, windows of 150-ms length with 75-ms time-shift are used. This length ensures enough pitch periods to calculate several stability measures including relative average perturbation (to quantify the difference between period-to-period in a phonation, and for evaluating whether the period duration is smooth over three adjacent cycles [30]), PPQ (to quantify the variability of the pitch period evaluated in five consecutive cycles), and APQ (to calculate the average difference between the amplitude of five preceding and successive pitch periods).

Spectral-Cepstral modeling: The aim of modeling the speech spectrum in the spectral or cepstral domains is to assess the ability of the speaker to generate periodic movements of the vocal folds, i.e., with a lot of harmonic components (or rahmonics in the Cepstral domain). This property is also called "spectral wealth" in the literature [45]. Features associated with the spectral and cepstral domains are included in this paper with aim of modeling changes in the speech spectrum, especially around the first two formants (F_1 and F_2), where most of the energy of the signal is concentrated. The features are computed within windows of 40-ms length and 20-ms time shift. The set of features includes 11 MFCCs, which are a smooth representation of voice spectrum that considers the human auditory perception [36]; a total of 28 linear predictive coefficients (LPC) are calculated and 14 formants are extracted from the LPC spectrum of the voice signals [46]; the amplitude and frequency values of F_1 and F_2 are calculated from the MGDF-based speech spectrum, and finally, the ratio between their amplitude values is also included. Further details of the features extracted from the MGDF can be found in [23].

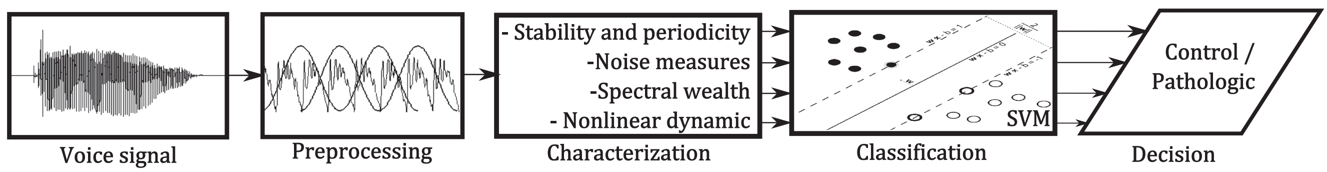

Fig. 2. Methodology addressed to discriminate between pathological and healthy voice signals.

Nonlinear behavior: This is associated with several phenomena observed in voice signals including nonlinear pressure-flow in the glottis, nonlinear stress–strain curves of vocal fold tissues, and nonlinearities observed in the vocal fold collision [31]. The compensatory movements in different muscles and limbs involved in the speech production process can also lead to a nonlinear behavior in the speech signal. This phenomenon appears mainly when the speaker realizes that he is speaking inappropriately and then tries to correct the "errors" [24]. To estimate the nonlinear features, windowing with 55-ms length and 27.5-ms time shift is used. The set of measures evaluated in this paper includes correlation dimension (D_2), which is a measure of the space dimensionality occupied by the points in the reconstructed attractor [47]; largest Lyapunov exponent, defined as the average divergence rate of neighbor trajectories in the attractor [47]; Lempel–Ziv complexity, which allows us to estimate the randomness of a voice signal [48]; Hurst exponent (H), which measures possible long-term dependences in a time series [47]; and four entropy measurements are also included to estimate the uncertainty of the signal [49]. Additionally, assuming that the voice signals have two components, deterministic and stochastic, these two components are modeled with the recurrence period density entropy and the detrended fluctuation analysis, respectively [50].

The spectrum of voice signals produced by people with speech impairments is modified with respect to the spectrum of healthy voices. Those modifications can be reflected by additional peaks and/or valleys in the spectrum. This multicomponent structure of the pathological voice spectrum can be modeled using the Teager Energy Operator (TEO), which is a nonlinear feature that is used in speech processing to model amplitude modulations in speech. Two versions of TEO are applied in this study. One version assumes that the voice signal can be represented as the sum of two uncorrelated components allowing the detection of additional components introduced due to the presence of speech impairments [27]. The other version of the TEO is based on the fast Fourier transform. This measure shows to be more effective in modeling the energy content of voice signals and seems to be sensitive to specific events like the initial and final points of fricatives and plosives [51]. We included this measure in this paper because of its robustness to model the energy content of the voice signals. The differences between the TEO contours estimated from healthy and pathological signals are quantified in a total of four measurements. The set of features includes correlation coefficient [27], Euclidean and logarithmic distances, and the area under the absolute value of the TEO contour [52].

B. Classification

A Gaussian kernel support vector machine (SVM) with soft margin is used to discriminate between pathological and healthy speakers. The margin parameter C and the bandwidth of the Gaussian kernel γ are optimized through a grid-search with $10^{-3} \leq C \leq 10^4$ and $10^{-1} \leq \gamma \leq 10^3$. The SVM is trained following a cross-validation (CV) strategy, i.e., one portion of the data is chosen for training the model and the remaining is used for testing it. The procedure is repeated several times to compute the performance of the system and its confidence interval. As the distribution of the train and test subsets depends mainly on the size of the database, the CV procedure is explained in Section III-B, after introducing the details of the databases considered in this study. The selection criterion of C and γ was based on the accuracy obtained in the test set, which could lead to slightly optimistic results, but as there are only two parameters to be optimized, the bias effect is minimal. The SVM is used here due to its validated success in similar studies that address the problem of the automatic detection of pathological speech signals [10], [26], [53], [54].

III. Experimental Setup

A. Corpus of Speakers

1) Neurological Disorder: PD: Three different databases with speakers of three languages are considered in this paper.

Spanish recordings: Recordings of the PC-GITA database are considered [18]. It contains recordings of the five Spanish vowels uttered by 100 speakers, 50 patients with PD and 50 healthy controls, all of them balanced in age and gender. The age of the 25 male patients ranges from 33 to 77 (mean 62.2 $\pm$ 11.2) and the age of the 25 female patients ranges from 44 to 75 years (mean 60.1 $\pm$ 7.8). For the case of HC, the age of the 25 men ranges from 31 to 86 years (mean 61.2 $\pm$ 11.3) and the age of the 25 women ranges from 43 to 76 (mean 60.7 $\pm$ 7.7). All of the participants were recorded in a sound-proof booth at Clínica Noel of Medellín in Colombia, using a dynamic omnidirectional microphone and professional audio card. The sampling frequency of the recordings was 44.1 kHz with 16 resolution bits. All of the patients were recorded in on-state, i.e., no more than 3 h after their morning medication. The patients were diagnosed by a neurologist expert. The mean values of their evaluation according to the motor section (the third one) of the unified Parkinson's disease rating scale (UPDRS-III) and Hoehn & Yahr (H&Y) [55] scale are 36.7 $\pm$ 18.7 and 2.3 $\pm$ 0.8, respectively. These two scales are considered as the global standard for assessing the neurological state of PD patients. The UPDRS-III only considers motor aspects of the patients and its value ranges from 0 to 132, while H&Y ranges from 1 to 4 (lower values indicate earlier stages of the disease). PC-GITA can be obtained through the first author of this paper and under a nondisclosure agreement.

German recordings: This corpus consists of 176 German native speakers: 88 patients with PD (47 men and 41 women) and 88 healthy controls (44 men, 44 women). The age of male patients ranges between 44 and 82 years (mean 66.7 $\pm$ 8.4), while the age of the female patients ranges from 42 to 84 years (mean 66.2 $\pm$ 9.7). Regarding the control group, the age of men ranges from 26 to 83 years (mean 63.8 $\pm$ 12.7), and the age of the women is between 54 and 79 years (mean 62.6 $\pm$ 15.2). The mean values of the neurological evaluation performed on all of the patients according to the UPDRS-III and H&Y scales are 22.7 $\pm$ 10.9 and 2.4 $\pm$ 0.6, respectively. The speech samples were also recorded with the patients in on-state. The voice signals were sampled at 16 kHz with 16 resolution bits. This database was collected in the Knappschaftskrankenhaus of Bochum in Germany. Only the sustained phonation of the German vowel /a/ is considered in this paper. Further details of the data can be found in [56].

Czech recordings: This database contains sustained phonations of the vowel /i/ pronounced by 42 Czech native speakers: 21 with PD and 21 healthy controls (all of them are male). The age of the patients ranges from 37 to 83 years (mean 62.2 $\pm$ 11.0), and the age of the healthy speakers ranges from 36 to 80 years (mean 57.2 $\pm$ 13.0). The mean values of their evaluation according to the UPDRS-III and H&Y scales are 17.9 $\pm$ 7.4 and 2.2 $\pm$ 0.5, respectively. None of the patients had been medicated at the recording session. The speech data were recorded in the General University Hospital in Prague, Czech Republic. The voice signals were sampled at 48 kHz with 16 resolution bits. Further details of this database can be found in [54].

2) Functional Disease: CLP: The database was collected by *Grupo de Procesamiento y Reconocimiento de Seales (GPRS)* from the *Universidad Nacional de Colombia, branch Manizales* [21]. This database includes phonations of the five Spanish vowels uttered by children with repaired-CLP, i.e., the children were already under post-operative speech therapy. A total of 130 children with CLP and 108 healthy controls are considered. All of the CLP patients were evaluated by a phoniatrician and diagnosed with hypernasal speech. The age of the speakers in both groups (pathological and healthy) ranged from 5 to 15 years (mean 10). The voice recordings were captured with a sampling frequency of 44.1 kHz and 16 bits resolution.

3) Laryngeal Pathologies: LP: Two databases are considered in this paper. The first one was collected by *the Massachusetts Eye and Ear Infirmary (MEEI) Voice & Speech Lab* [57]. This database contains voice registers recorded at different sampling frequencies. For the experiments addressed in this study, all of the recordings were downsampled to 25 kHz with 16 bits resolution. The voice registers consist of sustained phonations of the English vowel /a/ uttered by 173 patients with different voice disorders including laryngeal cancer, polyps, nodules, edema, among others. The control group includes a total of 53 healthy speakers. The mean age of the participants of both groups is 37 years [33].

The second database with recordings of people suffering from LP was collected by *Universidad Politécnica de Madrid (UPM)* [58]. It contains sustained phonations of the Spanish vowel /a/. The pathological voice set includes recordings of 200 speakers, 74 male and 126 female, with ages ranging from 11 to 76 years (mean 36). As in the database described above, different LP are included. The set of healthy speakers includes recordings of 199 participants: 87 males and 112 females, with ages ranging from 16 to 70 years (range 36). The sampling frequency is 50 Hz with 16 bits resolution.

All of the speakers in the databases used in this study were evaluated by a phoniatrician, and only those participants that showed speech impairments associated with each disease were included from each database.

B. Experiments

The features extracted with each of the four characterization approaches introduced in Section II are used to form four feature vectors per voice window. Four functionals are calculated from each feature vector across the analysis windows: mean value, standard deviation, kurtosis, and skewness. The set of noise features contains a total of six measures; thus, a 24-D features vector is formed per recording ($6 \times 4 = 24$). For the analysis of periodicity and stability of the voice registers, a total of five measures are calculated, forming a 20-D features vector per voice register ($5 \times 4 = 20$). The spectral-cepstral modeling is performed with a set with 30 measures; thus, a 120-D features vector is formed per recording ($30 \times 4 = 120$). With respect to the nonlinear behavior, it is evaluated considering a set with 18 features, forming a 72-D feature vector per recording ($18 \times 4 = 72$).

The SVM is trained following a CV strategy, i.e., one portion of the data is considered for training the SVM and the rest is for testing the resulting model. The process is repeated until having tested all of the speakers. In order to estimate the standard deviation of the recognition rates, at least ten speakers are included in the test sets; therefore, we are using fourfold CV on the Czech data and for Spanish and German a tenfold CV strategy is followed, which means to include ten and sixteen speakers in the test sets, respectively.

C. Results and Discussion

This section includes the results obtained with the four sets of features applied on the databases described above. The general performance of the models presented here is mainly discussed in terms of accuracy (Acc). Sensitivity (Sens) and specificity (Spec) are also included to show the capability of each model to correctly detect pathological and healthy speakers, respectively. Additionally, the area under the receiver operating characteristic (ROC) curve (AUC) is included in order to show the results more compactly [59].

Table IV shows the results obtained with the noise measures evaluated on each database. The highest accuracy is obtained

TABLE IV
RESULTS OBTAINED WITH NOISE MEASURES

			Acc %	Spec %	Sens %	AUC
LP	English	/a/	97 ± 4	97 ± 4	95 ± 9	0.93
	Spanish	/a/	80 ± 8	75 ± 14	84 ± 12	0.84
		/a/	87 ± 10	88 ± 12	86 ± 11	0.87
		/e/	89 ± 8	89 ± 8	88 ± 13	0.90
CLP	Spanish	/i/	92 ± 8	92 ± 10	91 ± 13	0.92
		/o/	87 ± 7	83 ± 10	91 ± 14	0.90
		/u/	83 ± 11	87 ± 13	78 ± 29	0.86
	German	/a/	71 ± 7	74 ± 18	67 ± 15	0.69
	Czech	/i/	82 ± 9	77 ± 16	88 ± 15	0.84
		/a/	77 ± 7	79 ± 14	75 ± 12	0.77
PD		/e/	75 ± 6	72 ± 14	77 ± 16	0.74
	Spanish	/i/	77 ± 8	75 ± 10	79 ± 16	0.79
		/o/	74 ± 8	71 ± 9	76 ± 14	0.75
		/u/	72 ± 8	69 ± 18	75 ± 18	0.76

TABLE V
RESULTS OBTAINED WITH SPECTRAL-CEPSTRAL FEATURES

			Acc %	Spec %	Sens %	AUC
LP	English	/a/	95 ± 5	97 ± 5	91 ± 12	0.93
	Spanish	/a/	78 ± 7	74 ± 16	83 ± 5	0.79
		/a/	97 ± 5	96 ± 6	98 ± 4	0.97
		/e/	99 ± 3	97 ± 5	100 ± 0	0.99
CLP	Spanish	/i/	98 ± 3	99 ± 5	97 ± 6	0.97
		/o/	95 ± 6	95 ± 8	95 ± 7	0.95
		/u/	97 ± 4	97 ± 5	98 ± 6	0.98
	German	/a/	66 ± 6	62 ± 21	70 ± 17	0.66
	Czech	/i/	76 ± 11	85 ± 18	67 ± 23	0.76
PD		/a/	69 ± 8	65 ± 19	73 ± 14	0.67
		/e/	72 ± 9	72 ± 19	73 ± 14	0.72
	Spanish	/i/	67 ± 10	64 ± 20	69 ± 15	0.69
		/o/	75 ± 8	73 ± 16	78 ± 12	0.78
		/u/	71 ± 7	72 ± 14	69 ± 11	0.73

TABLE VI
RESULTS OBTAINED WITH NONLINEAR BEHAVIOR FEATURES

			Acc %	Spec %	Sens %	AUC
LP	English	/a/	95 ± 4	95 ± 5	92 ± 4	0.93
	Spanish	/a/	82 ± 6	76 ± 13	89 ± 8	0.86
		/a/	96 ± 5	95 ± 7	97 ± 7	0.96
		/e/	97 ± 5	96 ± 8	98 ± 6	0.97
CLP	Spanish	/i/	94 ± 5	94 ± 6	93 ± 8	0.93
		/o/	93 ± 5	92 ± 9	93 ± 8	0.93
		/u/	92 ± 8	90 ± 14	96 ± 11	0.92
	German	/a/	72 ± 11	82 ± 17	62 ± 29	0.77
	Czech	/i/	81 ± 11	75 ± 18	87 ± 13	0.81
		/a/	78 ± 7	79 ± 11	78 ± 12	0.77
PD		/e/	76 ± 7	84 ± 13	68 ± 16	0.74
	Spanish	/i/	72 ± 11	71 ± 28	72 ± 17	0.74
		/o/	73 ± 8	78 ± 20	68 ± 21	0.72
		/u/	79 ± 9	80 ± 16	79 ± 20	0.78

TABLE VII
RESULTS OBTAINED WITH STABILITY AND PERIODICITY FEATURES

			Acc %	Spec %	Sens %	AUC
LP	English	/ah/	99 ± 4	98 ± 5	100 ± 0	0.99
	Spanish	/a/	82 ± 11	86 ± 11	79 ± 21	0.79
		/a/	98 ± 4	99 ± 3	97 ± 6	0.98
		/e/	97 ± 3	97 ± 4	96 ± 6	0.96
CLP	Spanish	/i/	99 ± 3	100 ± 0	98 ± 6	0.99
		/o/	99 ± 3	100 ± 0	98 ± 6	0.99
		/u/	98 ± 3	97 ± 5	98 ± 5	0.97
	German	/a/	87 ± 17	83 ± 32	91 ± 16	0.89
	Czech	/i/	92 ± 7	97 ± 7	87 ± 14	0.91
		/a/	91 ± 6	92 ± 7	91 ± 12	0.92
PD		/e/	81 ± 6	85 ± 13	77 ± 15	0.83
	Spanish	/i/	84 ± 10	82 ± 20	86 ± 18	0.83
		/o/	86 ± 9	85 ± 12	87 ± 15	0.90
		/u/	86 ± 7	89 ± 15	83 ± 13	0.85

with recordings of native English speakers suffering from LP. For the functional pathologies, children with CLP exhibit accuracies above 82% in all of the vowels. This result indicates that although the presence of glottal noise in the voice of children with CLP has not been widely documented so far in the literature, this phenomenon deserves to be studied with detail. On the other hand, with the recordings of PD patients the results are around 77% and the highest accuracy is obtained with the Czech vowel /i/. Note that these results are consistent with the clinical observations describing that LP induce dysphonia problems in the speakers and increase the noise level of their voice. These results are also consistent with previous studies on PD patients where the suitability of dysphonia measures to detect PD is evaluated [9].

Table V shows the results obtained with the features associated with the spectral-cepstral modeling. The results obtained with the CLP database are above 95% in all of the Spanish vowels, and with the vowel /e/, the accuracy is 99%. These results exceed those presented in [21], where the highest accuracy was 94% considering the same database and following the same CV strategy. This increase is probably caused by the set of spectral features considered in this paper which are designed to improve the spectral resolution in the low frequency zone, especially around 250 Hz, which is mostly affected in hypernasal speech signals [23]. The results obtained from the recordings of patients with LP are high but below those obtained with the features that model the noise content in voice signals. With respect to the results obtained from recordings of PD patients, note that most of them are around 70%, indicating that the spectral-cepstral measurements included in this paper are not suitable to discriminate between PD speakers and healthy controls.

The results with nonlinear features are shown in Table VI. The highest accuracies are obtained with the LP and CLP databases, in accordance with previous studies where the suitability of nonlinear features in speakers with LP and CLP is shown [60], [26]. For the recordings of people with PD, note that most of the results are below 80%, which is consistent with previous studies where the set of nonlinear features has to be merged with other measurements such as shimmer and HNR to achieve accuracies above 80% [12], [61].

The results obtained with the stability and periodicity measures are presented in Table VII. Note that most of the accuracies in the three pathologies are higher than those obtained with other feature sets. These results reflect the laryngeal problems previously observed by clinicians in children with CLP [3], [4]. Most of these impairments appear due to their velopharyngeal

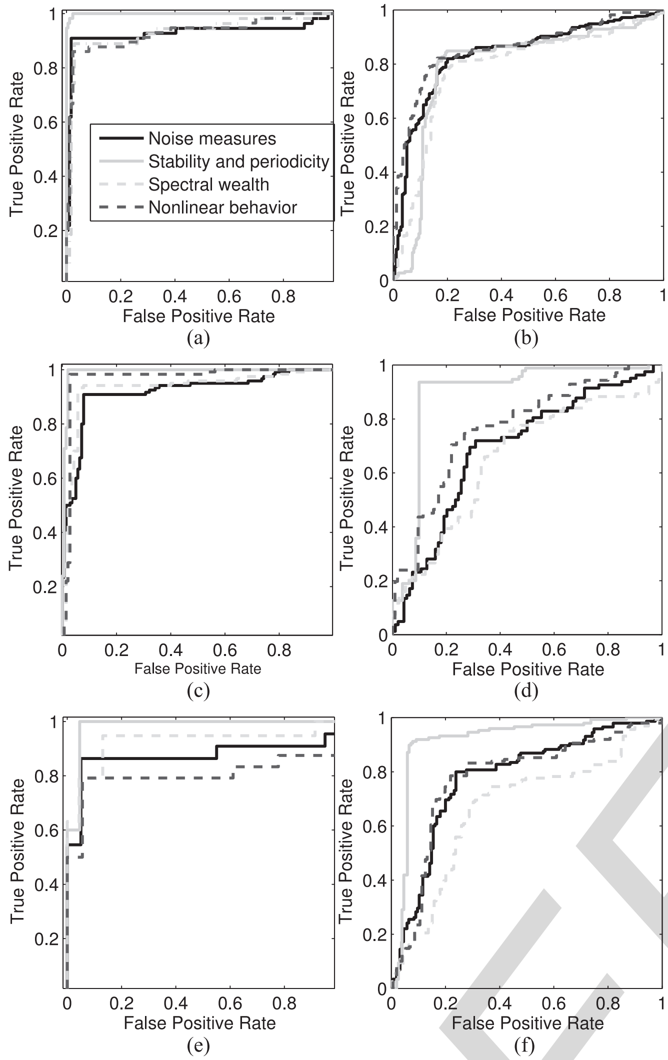

Fig. 3. ROC curves obtained with each set of features on each database.

insufficiency [24]. The results with the LP recordings are also consistent with previous studies where the abnormal vibration of vocal folds is observed in patients with LP [2]. With respect to the results in the PD databases, the set of features included to model periodicity and stability of sustained phonations were shown to suitably model irregular vibration of the vocal folds in PD patients.

Fig. 3 includes the ROC curves obtained in most of the experiments. This figure allows us to show the result more detailed. Each curve compares the results obtained with the four modeling approaches presented in this paper. For the CLP and PC-GITA databases, only the results with the vowel /a/ are shown.

The results obtained in this paper suggest that noise features are suitable to model voice signals of people with LP due to the presence of polyps, nodules, or laryngeal cancer. These measures are also suitable for assessing hypernasal speech, which can be likely explained by laryngeal problems that have been clinically observed in CLP patients [4]. The spectral-cepstral modeling applied in this paper is mainly focused on characterizing the voice spectrum around the first two formants. According to the results, that seems to be suitable for modeling the nasal formants and antiformants that appear around F_1 and F_2 in hypernasal voices. We are aware of the fact that the hypernasal spectrum is modified in a wider spectral range and that is the main motivation for including the high-order LPC and the MFCCs in the set of features. The spectral-cepstral modeling approach is also suitable for the automatic discrimination of voice recordings from people with LP and HC. The results suggest that there are alterations in the harmonics of these voice signals that are being modeled with this approach. Regarding the NLD features, they show to be suitable to model voices from people with LP and CLP. Although this result is consistent with previous studies with LP [62] and CLP [21] patients, further research is required in order to enable stronger conclusions regarding the interpretation of NLD features. The periodicity and stability measures seem to be the most suitable features to assess the three pathologies considered in this paper. This result in PD patients is explained due to their problems to control the vocal fold vibration, and in LP patients due to their problems to perform a complete closure of vocal folds. In CLP patients, it can be explained by the compensatory movements in the vocal tract [25]; however, considering that those movements are mainly manifested in continuous speech, further research is required to validate this observation in sustained phonations.

IV. Conclusion

Four groups of features describing different aspects of voices have been studied: noise content, stability and periodicity, spectral-cepstral modeling, and nonlinear behavior. The capability of each group of features to discriminate between pathological and healthy voices is tested in a total of six databases which contain recordings of patients with several pathologies including laryngeal (dysphonia due to polyps, nodules, cancer, among others diseases), neurological (dysphonia due to PD), and functional (hypernasality due to CLP).

The results obtained with the stability and periodicity features indicate that these measurements are suitable to discriminate between healthy speakers and people with different kind of pathologies. For CLP and LP, the accuracies are around 98%. For PD, sustained vowels of three different languages are tested and the accuracies range between 81% and 98% depending on the pronounced vowel and the language. Moreover, the alterations in the voice spectrum of hypernasal signals are accurately modeled by the spectral-cepstral approach presented in this paper, with accuracies above 95% in the five Spanish vowels. Additionally, the results obtained with these features in the LP databases show that the frequency zone around the first two formants, and the harmonic structure of the signals are also modified in voices with dysphonia. Conversely, this phenomenon is not clearly observed in the recordings of PD patients, who can also exhibit dysphonic voices but mainly due to problems to control the vibration of the vocal folds but not due to the presence of polyps, nodules, or tumors. It seems like the voice spectrum, around the first two formants and its harmonic structure are not equally affected in all of the diseases. Although there are some studies that report the presence of hypernasality in the voice of PD patients, the prevalence of such impairment in PD is still unclear. The noise content of the voice signals does not show high accuracies in the automatic discrimination of pathological and healthy speakers. Notwithstanding these features have been widely used in the literature to model different pathologies in voice, it seems that the group of noise measures included in this study, which is actually quite comprehensive, is not suitable to model noise in voice signals (especially from Parkinson's patients). Regarding the results obtained with the nonlinear behavior features, the accuracies indicate that those measurements are suitable to discriminate between pathological and healthy speakers; however, further research in this topic is necessary to find more conclusive and interpretable results.

According to the analyses performed in this paper, before characterizing the voice recordings, it is useful to understand the details of the pathology considering its origin and the organs or tissues involved in the disease and in the speech production process. After these analyses, the characterization will not be blind and will allow an appropriate selection of the measurements to be applied. For instance, if the pathology is closely related to the stability of vocal fold vibration, the periodicity features could be the most appropriate, but if the problem is related to hoarseness, the best option should be noise content features or spectral and cepstral modeling. In any case, an informed selection of the techniques that are applied to model the voice signal could help the speech therapist and the clinician to make more accurate decisions regarding the therapy and/or the treatment for the patients.

Finally, there are several aspects and limitations regarding the methods presented in this paper that need to be discussed. For instance, the results could improve if a different classification technique is applied, e.g., deep neural networks. This study only considers recordings of sustained phonations; although these analyses are relevant, experiments with continuous speech signals would allow studying other phenomena in speech related with articulation and/or prosody. The experiments performed with PD patients only considered the motor impairments reflected in the voice production, but cognitive and mood problems that are also associated with this disease were not studied here. The computation of statistical functionals (mean value, standard deviation, skewness, and kurtosis) could lead to lose temporal aspects in the voice recordings; however, as only sustained phonations are considered here, the quasi-stationarity can be assumed and the information lost is minimal. The diversity of languages is only considered in two of the three diseases studied in this paper, which limits the comparability of the results regarding language differences in different pathologies. For future work, we expect to perform several experiments with continuous speech signals of different pathologies, considering different characterization approaches, also grouped according to the phenomena that are being modeled. Additionally, the sizes of the databases need to be increased to assess the generalization capability of the models.

References

[1] A. Bayestehtashk, M. Asgari, I. Shafran, and J. McNames, "Fully automated assessment of the severity of Parkinson's disease from speech," *Comput. Speech Lang.*, vol. 29, no. 1, pp. 172–185, 2015.

[2] J. Hillenbrand and R. A. Houde, "Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech," *J. Speech Hearing Res.*, vol. 39, no. 2, pp. 311–321, 1996.

[3] G. Henningsson, D. P. Kuehn, D. Sell, T. Sweeney, J. E. Trost-Cardamone, and T. L. Whitehill, "Universal parameters for reporting speech outcomes in individuals with cleft palate," *Cleft Palate-Craniofacial J.*, vol. 45, no. 1, pp. 1–17, 2008.

[4] L. L. D'Antonio, H. R. Muntz, M. A. Province, and J. L. Marsh, "Laryngeal/voice findings in patients with velopharyngeal dysfunction," *Laryngoscope*, vol. 98, no. 4, pp. 432–438, 1988.

[5] J. Logemann, H. Fisher, B. Boshes, and E. Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients," *J. Speech Hearing Disorders*, vol. 43, pp. 47–57, 1978.

[6] A. K. Ho, R. Jansek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of people with Parkinson's disease," *Behavioral Neurol.*, vol. 11, pp. 131–137, 1998.

[7] M. de Rijk, L. Launer, K. Berger, M. Breteler, J. Dartigues, M. Baldereschi, L. Fratiglioni, A. Lobo, J. Martinez-Lage, C. Trenkwalder, and A. Hofman, "Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group," *Neurology*, vol. 54, no. 11 Suppl 5, pp. S21–S23, 2000.

[8] F. Darley, A. Aronson, and J. Brown, "Differential diagnosis patterns of dysarthria," *Motor Speech Disorders*. Philadelphia, PA, USA: Saunders, 1975.

[9] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1015–1022, Apr. 2009.

[10] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Spectral and cepstral analyses for Parkinson's disease detection in Spanish vowels and words," *J. Expert Syst.*, pp. 1–10, 2015, doi: 10.1111/exsy.12106, in press.

[11] Y. Zhang, J. Jiang, and D. A. Rahn, "Studying vocal fold vibrations in Parkinson's disease with a nonlinear model," *Chaos*, vol. 15, p. 33903, 2005.

[12] A. Tsanas, M. Little, P. McSharry, and L. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1264–1271, May 2012.

[13] T. Bocklet, E. Nöth, G. Stemmer, H. Ruzickova, and J. Rusz, "Detection of persons with Parkinson's disease by acoustic, vocal, and prosodic analysis," in *Proc. IEEE Workshop Autom. Speech Recog. Understanding*, 2011, pp. 478–483.

[14] J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka, "Imprecise vowel articulation as a potential early marker of Parkinson's disease: Effect of speaking task," *J. Acoust. Soc. Amer.*, vol. 134, no. 3, pp. 2171–2181, 2013.

[15] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Analysis of speech from people with Parkinson's disease through nonlinear dynamics," *Lecture Notes Comput. Sci.*, vol. 7911, pp. 112–119, 2013.

[16] T. Bocklet, S. Steidl, E. Nöth, and S. Skodda, "Automatic evaluation of Parkinson's speech—Acoustic, prosodic and voice related cues," in *Proc. 15th Annu. Conf. Int. Speech Commun. Assoc.*, 2013, pp. 1149–1153.

[17] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, and E. Nöth, "Automatic detection of Parkinson's disease from words uttered in three different languages," in *Proc. 16th Annu. Conf. Int. Speech Commun. Assoc.*, 2014, pp. 1473–1577.

[18] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. González-Rátiva, and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," presented at the 9th Int. Conf. Lang. Resources Eval., Reykjavik, Iceland, 2014.

[19] E. B. Hook, "Congenital malformations worldwide: A report from the international clearinghouse for birth defect monitoring systems," *Amer. J. Human Genetics*, vol. 51, no. 4, p. 919, 1992.

[20] D. Sell, P. Grunwell, S. Mildinhall, T. Murphy, T. Cornish, D. Bearn, W. Shaw, J. Murray, A. Williams, and J. Sandy, "Cleft lip and palate care in the United Kingdom—The Clinical Standards Advisory Group (CSAG) study. Part 3: Speech outcomes," *Cleft Palate-Craniofacial J.*, vol. 38, pp. 30–37, 2001.

[21] J. R. Orozco-Arroyave, S. Murillo-Rendón, A. Álvarez-Meza, J. D. Arias-Londoño, E. Delgado-Trejos, J. F. Vargas-Bonilla, and C. G. Castellanos-Domínguez, "Automatic selection of acoustic and non-linear dynamic

features in voice signals for hypernasality detection," in *Proc. 12th Annu. Conf. Int. Speech Commun. Assoc.*, 2011, pp. 529–532.

[22] A. Maier, F. Hönig, T. Bocklet, E. Nöth, F. Stelzle, E. Nkenke, and M. Schuster, "Automatic detection of articulation disorders in children with cleft lip and palate," *J. Acoust. Soc. Amer.*, vol. 126, no. 5, pp. 2589–2602, Nov. 2009.

[23] P. Vijayalakshmi, M. R. Reddy, and D. O'Shaughnessy, "Acoustic analysis and detection of hypernasality using a group delay function," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 68–71, Apr. 2007.

[24] K. Golding, *Therapy Techniques for Cleft Palate Speech and Related Disorders*. San Diego, CA, USA: Singular Thomson Learning, 2001.

[25] A. Giovanni, M. Ouaknine, R. Guelfucci, T. Yu, M. Zanaret, and J. Triglia, "Nonlinear behavior of vocal fold vibration: the role of coupling between the vocal folds," *J. Voice*, vol. 13, no. 4, pp. 456–476, 1999.

[26] J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, J. D. Arias-Londoño, S. Murillo-Rendón, C. G. Castellanos-Domínguez, and J. F. Garcés-Rodríguez, "Nonlinear dynamics for hypernasality detection in Spanish vowels and words," *Cognitive Comput.*, vol. 5, no. 4, pp. 448–457, 2013.

[27] D. A. Cairns, J. H. L. Hansen, and J. E. Riski, "A noninvasive technique for detecting hypernasal speech using a nonlinear operator," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 1, pp. 35–45, Jan. 1996.

[28] A. Maier, F. Hönig, C. Hacker, M. Shuster, and E. Nöth, "Automatic evaluation of characteristic speech disorders in children with cleft lip and palate," in *Proc. 9th Annu. Conf. Int. Speech Commun. Assoc.*, 2008, pp. 1757–1760.

[29] P. Vijayalakshmi, T. Nagarajan, and V. Jayanthan Ra, "Selective pole modification-based technique for the analysis and detection of hypernasality," in *Proc. IEEE TENCON*, 2009, pp. 1–5.

[30] S. Hadjitodorov and P. Mitev, "A computer system for acoustic analysis of pathological voices and laryngeal diseases screening," *Med. Eng. Phys.*, vol. 24, no. 6, pp. 419–429, 2002.

[31] H. Herzel, D. Berry, and I. R. Titze, "Analysis of vocal disorders with methods from nonlinear dynamics," *J. Speech Hearing Res.*, vol. 37, no. 2, pp. 1008–1019, 1994.

[32] S. Feijoo and C. Hernández-Espinosa, "Short term stability measures for the evaluation of vocal quality," *J. Speech Hearing Res.*, vol. 33, pp. 324–334, 1990.

[33] V. Parsa and D. Jamieson, "Identification of pathological voices using glottal noise measures," *J. Speech, Lang. Hearing Res.*, vol. 43, no. 2, pp. 469–485, 2000.

[34] K. Umapathy, S. Krishnan, and V. Parsa, "Discrimination of pathological voices using a time-frequency approach," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp. 421–430, Mar. 2005.

[35] R. T. Ritchings, M. McGillion, and C. J. Moore, "Pathological voice quality assessment using artificial neural networks," *Med. Eng. Phys.*, vol. 24, nos. 7/8, pp. 561–564, 2002.

[36] J. Godino-Llorente, P. Gómez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1943–1953, Oct. 2006.

[37] A. Dibazar, S. Narayanan, and T. Berger, "Feature analysis for automatic detection of pathological speech," in *Proc. 24th Annu. Conf. Annu. Fall Meeting Biomed. Eng. Soc. EMBS/BMES Conf.*, 2002, pp. 182–183.

[38] T. Dubuisson, T. Dutoit, B. Gosselin, and M. Remacle, "On the use of the correlation between acoustic descriptors for the normal/pathological voices discrimination," *EURASIP J. Adv. Signal Process.*, vol. 2009, p. 19, 2009.

[39] J. Goddard, G. Schlotthauer, M. E. Torres, and H. L. Rufiner, "Dimensionality reduction for visualization of normal and pathological speech data," *Biomed. Signal Processi. Control*, vol. 4, no. 3, pp. 194–201, 2009.

[40] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Domínguez, "An improved method for voice pathology detection by means of a HMM-based feature space transformation," *Pattern Recog.*, vol. 43, no. 9, pp. 3100–3112, 2010.

[41] C. M. Travieso-González, J. B. Alonso-Hernández, J. R. Orozco-Arroyave, J. Solé-Casals, and E. Gallego-Jutglá, "Automatic detection of laryngeal pathologies in running speech based on the HMM transformation of the nonlinear dynamics," *Lecture Notes Comput. Sci.*, vol. 7911, pp. 136–143, 2013.

[42] H. Kasuya, S. Ogawa, K. Mashima, 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.

[43] D. Deliyski, "Acoustic model and evaluation of pathological voice production," in *Proc. 3rd Conf. Speech Commun. Technol.*, 1993, pp. 1969–1972.

[44] D. Michaelis, T. Gramss, and H. Strube, "Glottal-to-noise excitation ratio—A new measure for describing pathological voices," *Acustica/Acta Acustica*, vol. 83, pp. 700–706, 1997.

[45] J. B. Alonso-Hernández, J. de Leon, I. Alonso, and M. F. Ferrer-Ballester, "Automatic detection of pathologies in the voice by HOS based parameters," *EURASIP J. Adv. Signal Process.*, vol. 2001, no. 4, pp. 275–284, 2001.

[46] L. R. Rabiner and R. W. Schafer, *Introduction to Digital Speech Processing (Foundations and Trends in Speech Processing)*. Hanover, MA, USA: Now Publishers, 2007, pp. 1–194.

[47] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[48] F. Kaspar and H. Shuster, "Easily calculable measure for complexity of spatiotemporal patterns," *Phys. Rev. A*, vol. 36, no. 2, pp. 842–848, 1987.

[49] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and C. G. Castellanos-Domínguez, "Automatic detection of pathological voices using complexity measures, noise parameters, and Mel-cepstral coefficients," *IEEE Trans. Bio. Eng.*, vol. 58, no. 2, pp. 370–279, Feb. 2011.

[50] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomed. Eng. Online*, vol. 6, p. 23, 2007.

[51] G. Ying, C. Mitchell, and L. Jamieson, "Endpoint detection of isolated utterances based on a modified Teager energy measurement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1993, vol. 2, pp. 732–735.

[52] E. Belalcazar-Bolaños, J. R. Orozco-Arroyave, J. Vargas-Bonilla, J. D. Arias-Londoño, C. G. Castellanos-Domínguez, and E. Nöth, "New cues in low-frequency of speech for automatic detection of Parkinson's disease," *Lecture Notes Comput. Sci.*, vol. 7930, pp. 283–292, 2013.

[53] A. Marier, T. Haderlein, U. Eysholdt, F. Rosanowski, A. Batliner, M. Schuster, and E. Nöth, "PEAKS—A system for the automatic evaluation of voice and speech disorders," *Speech Commun.*, vol. 51, no. 5, pp. 425–437, 2009.

[54] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease," *J. Acoustical Soc. Amer.*, vol. 129, pp. 350–367, 2011.

[55] C. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. Stebbins, C. Counsell, N. Giladi, R. Holloway, C. Moore, G. Wenning, M. Yahr, and L. Seidl, "Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations," *Movement Disorders*, vol. 19, no. 9, pp. 1020–1028, 2004.

[56] S. Skodda, W. Grönheit, and U. Schlegel, "Intonation and speech rate in parkinson's Disease: General and dynamic aspects and responsiveness to Levodopa admission," *J. Voice*, vol. 25, no. 4, pp. 199–205, 2011.

[57] Corporate Author, *Voice disorders database, Version 1.0.3 [CD-ROM]*. Lincoln Park, NJ, USA: Kay Elemetrics Corp, 1994.

[58] J. Godino-Llorente, V. Osma-Ruiz, N. Sáenz-Lechón, I. Cobeta-Marco, T. Gonzales-Herranz, and C. Ramírez-Calvo, "Acoustic analysis of voice using WPCVox: A comparative study with multi-dimensional voice program," *Eur. Arch. Otorhinolaryngol.*, vol. 265, no. 4, pp. 465–476, 2008.

[59] N. Sáenz-Lechón, J. I. Godino-Llorente, J. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomed. Signal Process. Control*, vol. 1, pp. 120–128, 2006.

[60] J. Jiang, Y. Zhang, and C. McGilligan, "Chaos in voice, from modeling to measurement," *J. Voice*, vol. 1, no. 20, pp. 2–17, 2006.

[61] D. Rahn, M. Chou, J. J. Jiang, and Y. Zhang, "Phonatory impairment in Parkinson's disease: Evidence from nonlinear dynamic analysis and perturbation analysis," *J. Voice*, vol. 21, pp. 64–71, 2007.

[62] P. Henríquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. D. de María, "Characterization of healthy and pathological voice through measures based on nonlinear dynamics," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1186–1195, Aug. 2009.

Authors' photographs and biographies not available at the time of publication.