Automatic detection of Parkinson's disease in running speech spoken in three different languages

J. R. Orozco-Arroyave[a)]
Faculty of Engineering, Universidad de Antioquia, Calle 67 Número 53-108, Medellín 1226, Colombia

F. Hönig
Pattern Recognition Lab, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Martensstraße 3, Erlangen 91058, Germany

J. D. Arias-Londoño and J. F. Vargas-Bonilla
Faculty of Engineering, Universidad de Antioquia, Calle 67 Número 53-108, Medellín 1226, Colombia

K. Daqrouq
Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 22254, Saudi Arabia

S. Skodda
Department of Neurology, Knappschaftskrankenhaus, Ruhr-University, In der Schornau 23-25, Bochum D-44892, Germany

J. Rusz
Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic

E. Nöth[b)]
Pattern Recognition Lab, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Martensstraße 3, Erlangen 91058, Germany

(Received 5 February 2015; revised 17 October 2015; accepted 27 December 2015; published online 22 January 2016)

The aim of this study is the analysis of continuous speech signals of people with Parkinson's disease (PD) considering recordings in different languages (Spanish, German, and Czech). A method for the characterization of the speech signals, based on the automatic segmentation of utterances into voiced and unvoiced frames, is addressed here. The energy content of the unvoiced sounds is modeled using 12 Mel-frequency cepstral coefficients and 25 bands scaled according to the Bark scale. Four speech tasks comprising isolated words, rapid repetition of the syllables /pa/-/ta/-/ka/, sentences, and read texts are evaluated. The method proves to be more accurate than classical approaches in the automatic classification of speech of people with PD and healthy controls. The accuracies range from 85% to 99% depending on the language and the speech task. Cross-language experiments are also performed confirming the robustness and generalization capability of the method, with accuracies ranging from 60% to 99%. This work comprises a step forward for the development of computer aided tools for the automatic assessment of dysarthric speech signals in multiple languages. © *2016 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4939739]

[CYE] Pages: 481–500

I. INTRODUCTION

Parkinson's disease (PD) is a neurological disorder that affects functions of the basal ganglia and it is characterized by the progressive loss of dopaminergic neurons in the substantia nigra of the midbrain.[1] It is estimated that PD affects 1% to 2% of the people older than 65.[2] The problems induced by PD include motor deficits such as bradykinesia, rigidity, postural instability, and resting tremor. Non-motor deficits include negative effects on the sensory system, sleep, behavior, cognition, and emotion.[3] The neurological state of patients with PD is clinically evaluated using the unified Parkinson's disease rating scale (UPDRS)[4] and Hoehn & Yahr staging scale.[5] These scales consider motor and non-motor symptoms; however, the evaluation of speech represents just one item. 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. These impairments are called *hypokinetic dysarthria*.[3,6]

There are several types of dysarthria that appear due to different neurological disorders and the research community has shown interest in characterizing them in order to support the diagnosis process. Darley *et al.*[7] presented a comprehensive study of different types of dysarthria with origins in seven neurological conditions including bulbar palsy, pseudobulbar palsy, amyotrophic lateral sclerosis, cerebellar

[a)]Also at: Pattern Recognition Lab, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Martensstraβe 3, Erlangen 91058, Germany. Electronic mail: rafael.orozco@i5.informatik.uni-erlangen.de
[b)]Also at: Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 22254, Saudi Arabia.

lesions, Parkinsonism, dystonia, and choreoathetosis. The authors considered thirty patients in each group (thirty-two with Parkinsonism). Most of the participants read a standard paragraph with all English phonemes. In some cases, a sample of conversational speech was used, and in a very few cases, it was necessary to ask the speaker to repeat sentences spoken by the examiner. The speech samples were perceptually assessed by three experts who considered a total of thirty-eight speech dimensions to evaluate the recordings. The ratings are given on a seven point scale of severity (1 representing normal speech and 7 representing very severe deviation from normal). The agreement among judges was evaluated considering a total of 150 patients and 37 of the dimensions. The judges agreed perfectly or up to one point on the scale in 84% of the evaluations. Two hundred and five of the 212 patients exhibited imprecise articulation, 180 showed irregular speech rate and reduced intelligibility. Monotonicity was observed in 177 patients, and 174 exhibited a harsh voice. The authors conclude that it is possible to differentiate among different types of dysarthria and the observed occurrence or co-occurrence of the analyzed speech dimensions can be used as diagnostic aids to identify different neurological disorders. This work is highly relevant to the study of dysarthric speech signals because it provides a very detailed and comprehensive revision of the characteristics of different types of dysarthria. There are other studies in the literature that analyze several phenomena in dysarthric speech. For instance, Green *et al.*[8] considered the duration of pauses and speech timing in recordings of ten patients with amyotrophic lateral sclerosis (ALS) and 10 healthy controls (HCs). Each participant read a 60-word paragraph and the authors observed that pauses are significantly longer and more variable in ALS speakers than in HCs. Automatic and manual measures were compared and no significant differences were observed, thus the authors conclude that the automatic approach could be suitable to extract and analyze pauses from continuous speech signals of different speech impairments. Similarly, Wang *et al.*[9] analyzed the suitability of several automatic measurements calculated from recordings of the rapid repetitions of syllables such as /pə/ and /kə/ [diadochokinetic evaluation (DDK)], to aid the clinical diagnosis processes. A total of 21 individuals with ataxic dysarthria were considered and the set of measures included in the analysis comprises, among others, average DDK period, average DDK rate, and average DDK peak intensity. The features are calculated using the diadochokinetic rate analysis protocol of the Kay-PENTAX motor speech profile. The automatic analyses are compared with respect to manual measures. Strong correlations between these two approaches are found, indicating that DDK analysis could be suitable to assess dysarthric speech; however, the relatively small number of speakers does not allow strong conclusions. Another contribution in the automatic assessment of dysarthric speech signals is presented by Paja *et al.*[10] The authors considered a set of 765 isolated words uttered by ten speakers with spastic dysarthria [from the universal access (UA-Speech) audio-visual database[11]] and applied several acoustic and prosodic features to model the speech signals. The authors perform automatic discrimination between two levels of dysarthria (mid-to-low vs mid-to-high) and also the prediction of the intelligibility level. The authors report accuracies of up to 95% in the binary classification experiments, and Pearson's correlations (r) of up to 0.96 between the original dysarthria levels and the predicted ones. This study shows the suitability of automatic speech analysis to evaluate intelligibility in dysarthric speech; however, it is important to highlight that the validation strategy addressed in this work, which consisted of a randomized bootstrap with 15-folds, can lead to highly optimistic results and biased conclusions because the speaker independence is not satisfied.

This paper is focused on the automatic discrimination of speech of people with hypokinetic dysarthria due to PD and healthy speakers. The study of PD is particularly relevant because it is the second most prevalent neurological disorder, affecting more than 4×10^6 people worldwide.[12] Additionally, PD has significant impact in the social, psychological, and physical interaction of patients. According to the Royal College of Physicians, in order to relieve such impact, in addition to the pharmacological treatment, PD patients should have access to a set of services and therapies including specialized nursing care, physiotherapy, and *speech and language therapy*;[13] however, it is estimated that only 3% to 4% of PD patients receive speech therapy.[14]

Medical therapies and surgery procedures such as deep brain stimulation have shown significant improvements in motor functions of patients with PD,[13] however, their impact on speech production remains unclear.[14,15] As the evaluation of speech of people with PD is performed non-objectively (perceptually) by clinicians, there is a general interest in the research community to develop accurate and robust methodologies to objectively assess the speech of PD patients.[16,17]

The impact of PD on the speaking skills of the patients can be characterized by three principal "dimensions" of speech: *phonation, articulation,* and *prosody*.[18] Phonation is defined as the vibration of vocal folds to produce sound, articulation comprises the modification of the position, stress, and shape of organs and tissues involved in speech production, and prosody is the variation of loudness, pitch, and timing to produce natural speech.[19]

From the clinical point of view, phonation problems are related to vocal fold bowing and incomplete closing of vocal folds.[20] Articulation deficits are manifested as reduced amplitude and velocity of the articulatory movements of lips, jaw, and tongue,[16] and prosody impairments are manifested as monopitch, monoloudness, and changes in speech rate and pauses,[18] and difficulties to express emotions through speech.[21]

The evaluation of different aspects or characteristics of speech with discriminative criteria is important to quantify and to understand their role in the automatic classification of PD patients and HCs from both the clinical and the engineering points of view. This paper considers contributions from the engineering side by reviewing different studies that address the problem of automatic classification of speech of people with PD and HCs using techniques based on statistics and/or machine learning. The link with the clinics is considered in this paper by reviewing methods in the literature that

analyze phonation, articulation, and/or prosody in speech. Additionally, the methodology presented here to characterize the speech signals is motivated by previous clinical and neurological observations.

In the following, several contributions performed from the engineering and machine learning points of view are presented to understand different clinical aspects of dysarthric speech signals.

Little *et al.*[22] applied different standard and non-standard *phonation* measures to discriminate between people with PD and HCs. The methodology applied by the authors is based on several characteristics of speech mainly used to detect dysphonia. The authors calculated a total of seventeen features, including different standard measures and others which are based on the nonlinear content of speech signal. After a systematic search through all pairs of features, those that are highly correlated (with a correlation coefficient greater than 0.95) were excluded. The final subset was composed of ten features, six standard and four nonlinear. The standard features were calculated using the software PRAAT.[23] The set comprised two versions of jitter (absolute and average absolute difference between cycles), the amplitude perturbation quotient, shimmer (calculated as the average absolute difference between the amplitudes of consecutive periods), harmonics to noise ratio (HNR), and noise to harmonics ratio (NHR). The nonlinear analysis included recurrence period density entropy (RPDE),[24] detrended fluctuation analysis (DFA),[24] correlation dimension,[25] and the pitch period entropy (PPE), which is a novel measure of dysphonia introduced in Ref. 22. Additionally, the authors performed a second round of exhaustive search through this set of ten features, and the resulting subset of features included HNR, RPDE, DFA, and PPE. The discriminative capability of this reduced set of features was tested by the authors by performing an automatic classification of 23 patients with PD and 8 HCs. All of the participants uttered the English vowel /ʌ/ in a sustained manner. The reported accuracy considering the subset with four measures is 91%.

Sapir *et al.*[26] evaluated the *articulation* capability of 38 patients with PD and 14 HCs considering different spectral features such as vowel space area (VSA), formant centralization ratio (FCR), natural logarithm of VSA, and the quotient F_{2i}/F_{2u}. F_{2i} and F_{2u} are the values of the second formants extracted from the vowels /i/ and /u/, respectively. All of the participants were English native speakers, and they were asked to repeat three sentences several times per day during at least 2 or 3 days before and after receiving voice treatment based on the Lee Silverman voice treatment (LSVT).[27] The set of three sentences includes (1) "the blue spot is on the key," (2) "the potato stew is in the pot," and (3) "the stew pot is packed with peas." The vowels /ʌ/, /i/, and /u/ were extracted from the recordings to perform the measurements. According to the reported results, FCR and F_{2i}/F_{2u} are highly correlated ($r = -0.90$) and both can differentiate between dysarthric and non-dysarthric speech signals.

Skodda *et al.*[15] measured different *prosodic* features on four sentences uttered by 138 patients with PD and 50 age matched HCs, all of the participants were German native speakers. The calculated features are based on estimations of the fundamental frequency of speech (F_0) performed using the standard software PRAAT.[23] The set of measures includes mean value of F_0, the standard deviation of F_0 (F_0SD) in Hz, and the difference of F_0SD calculated from the first and the fourth sentences. Additionally, the analysis of speech rate was performed by measuring the length of each syllable and each pause, respectively, based on the spectrogram of the sound pressure signal, and the net speech rate (NSR) was measured in syllables per second related to the net speech time in milliseconds. Further, the authors introduced the concept of *articulatory acceleration* as the difference between the NSR of the first and the fourth sentences. The authors performed several statistical tests to consider information from all of the measures and recordings. According to the results, the variation of F_0 is lower in PD patients than in HCs. The authors also observed that there is a correlation between several PD symptoms and prosody variables, such as the number of pauses in speech. The *articulation* ability of people with PD is also analyzed and the authors intended to reveal possible correlations among vowel articulation, global motor performance, and the stage of disease. A total of 68 patients with PD and 32 HCs were included in the study. The participants read a text and the values of the first two formants (F_1 and F_2) were measured from the vowels /a/, /i/, and /u/. The articulation analysis performed was based on measures of the triangular vowel space area (tVSA) and vowel articulation index (VAI). The authors performed several statistical tests and concluded that VAI in PD patients is significantly reduced compared to HC. Additionally, they indicate that tVSA is only reduced in male PD speakers. No correlations were found between vowel articulation and the extent of the disease.

Rusz *et al.*[18] considered recordings from a total of 46 participants (23 with PD and 23 HCs). Voice recordings comprised six different tasks including (1) isolated vowels pronounced in a sustained manner, (2) rapid repetition of /pa/-/ta/-/ka/ syllables, also called DDK evaluation, (3) read text of 136 words, (4) one monologue of at least 90 s, (5) read sentences, and (6) rhythmically read text of 34 words (8 rhymes followed by an example given by the examiner). These speech tasks were characterized considering three dimensions of speech: *phonation, articulation,* and *prosody*. Phonation features were evaluated on the sustained vowels and the set of measures includes the variation of F_0, different versions of jitter and shimmer, and noise content quantified through HNR and NHR. The evaluation of articulation was mostly performed considering the DDK task, and the features include the number of vocalizations of /pa/-/ta/-/ka/ per second, the ability to maintain a constant rate of C-V combinations in the pronunciation of /pa/-/ta/-/ka/, and different spectral-based measures of energy. Additionally, the authors included the vowel space area (VSA) measured from the sustained phonation of the vowels /a/, /i/, and /u/.[26]

The prosody evaluation was performed considering reading texts, sentences, and the monologue. The set of prosody features includes variation of F_0, percent pause time, articulation rate, number of pauses, standard deviation of the intensity, and the ability to reproduce perceived rhythm. The authors concluded that 78% of the patients evidenced speech

problems; articulation was the second most affected dimension of speech while prosody was the most affected even in the initial stage of the disease. They also found that the variation of the fundamental frequency measured on the monologues and emotional sentences contained very useful information for separating HCs from PD speakers.

Tsanas *et al.*[28] evaluated *phonation* of people with PD considering 132 measures from sustained phonations of the English vowel /ʌ/. A total of 263 speech samples were recorded from 43 subjects (33 with PD and 10 HCs). The set of measures included different estimations of jitter and shimmer, different variants of noise measures, Mel frequency cepstral coefficients (MFCCs), and nonlinear measures.[29] The authors applied four different feature selection techniques to find the best subset of features that separates between phonations of PD patients and HCs. They followed a tenfold cross validation (CV) strategy. The feature selection process was applied to the training sets to avoid overfitting. The final subset had selected features comprised of a total of ten measures, which were selected applying a voting scheme. Two different classification strategies were compared: random forest (RF) and support vector machine (SVM) with Gaussian kernel. The 263 phonations were split into two subsets: a training subset with 90% of the data (237 phonations), and a testing subset with the remaining 10% of the data (26 phonations). The process was repeated 100 times randomly permuting the subsets prior to splitting into training and testing. Errors over the 100 repetitions were averaged. The authors reported performances from 94.4% to 98.6%, depending on the feature selection technique.

According to the validation methodologies for automatic classification systems, training and testing subsets must be separated during the entire experiment and the validation process must be speaker-independent to avoid bias and to find more realistic results.[30] Note that the database used by Tsanas *et al.*[28] contains 263 phonations from 43 subjects, i.e., each subject repeated the phonation several times, thus speaker independence is not guaranteed in the experiments because all of the recordings are mixed into training and testing subsets. This methodological issue can lead to optimistic results and possible biased conclusions. In particular, since the target (the detection of PD) is constant per speaker, there is a chance for the system to decide by recognizing the speaker rather than recognizing the pathology.

Bocklet *et al.*[31] performed automatic classification of speech from patients with PD and HCs considering three different strategies to model the speech signal, *articulation, prosody,* and *phonation,* along with a set of 1582 acoustic features extracted using the OPENSMILE toolkit.[32] Articulation modeling was performed using the 13 MFCCs and their first and second order derivatives, forming a feature vector with 39 components per voice frame. Feature vectors were modeled using Gaussian mixture models (GMMs), such that one GMM was created for each speaker by the universal background modeling (GMM-UBM) technique. The GMM was created using a total of 128 Gaussians trained on the whole training set using the expectation-maximization (EM) algorithm. The means of the UBM were adapted by relevance maximum *a posteriori* (MAP) adaptation to find specific mixtures per speaker. Finally, the means of each Gaussian were used as speaker-specific features, forming 4992-dimensional (128×39) feature vectors per speaker. Prosodic modeling is performed using measures derived from F_0, energy, duration, pauses, jitter, and shimmer.[33] Feature vectors were formed computing mean, minimum, maximum, and standard deviation of a total of 73 features per voiced segment (292 dimensional). Phonation modeling was based on the estimation of physical parameters of the glottis. The two-mass vocal fold model was used with the aim of finding physically meaningful parameters.[34] A total of nine features were derived from the model. Stevens[34] presented the mathematical description of such parameters. The experiments included utterances of 176 German native speakers, 88 with PD and 88 HCs. The set of speech tasks comprises spontaneous speech, read text, read sentences, isolated words, sustained vowels, and the repetition of the syllable /pa/. The results were reported in terms of the correct classification per class and of unweighted average recall (UA). The highest classification rate was reached considering only articulation models (MFCCs and GMM-UBM). The recognition rate of PD patients was 86.5%, evaluating only the read sentences, while the highest UA was 81.9% when all of the tasks were combined.

The highest recognition achieved of phonations from people with PD (specificity) was 94.3%.

Phonation of PD patients was evaluated by Orozco-Arroyave *et al.*[35] through nonlinear dynamic features. The authors considered a group with 40 participants (20 with PD and 20 age-matched HCs). All of them uttered the five Spanish vowels (/a/, /e/, /i/, /o/, and /u/). The set of features included correlation dimension, largest Lyapunov exponent, Lempel-Ziv complexity, Hurst exponent, RPDE, DFA, approximate entropy, approximate entropy with Gaussian kernel, sample entropy, and sample entropy with Gaussian kernel. Accuracies ranging from 70.2% to 76.8% (depending on the vowel) were reported. The highest accuracy was obtained with the vowel /i/. The combination of all phonations did not improve the performance of the system. This work allowed the authors to determine the real contribution of nonlinear features separating PD patients and HCs. According to the results, more measures, such as HNR, NHR, jitter, and shimmer, need to be added to the set described by Tsanas *et al.*[28] to achieve higher accuracies when only sustained vowels are evaluated.

Bayestehtashk *et al.*[36] evaluated the speech of 168 patients with PD. The speech tasks considered included (1) sustained phonations of the English vowel /aː/, (2) DDK evaluation, and (3) reading text. The aim of the authors was to perform an automatic evaluation of the neurological state of the patients through speech. The set of features was comprised of a total of 1582 measures calculated using the OPENSMILE toolkit.[32] The accuracy of the model was tested using three different regression techniques to evaluate the severity of the disease according to the motor section of the UPDRS scale.[4] The authors report that ridge regression performs better than lasso and support vector regression. According to the results, features extracted from the reading texts are the most effective and robust to quantify the extent

of the disease. The mean absolute error obtained with respect to the motor section of the UPDRS scale is about 5.5, with a baseline of 8.0. The authors followed a leave-one-out cross-validation strategy to optimize the parameters and to measure the performance of the system. The authors claim that further work is required to present the information to clinicians in a useful and interpretable manner. Additionally, they conclude that different speech characteristics such as imprecise articulation, short rushes of speech and language impairments are still not modeled in the literature of PD. Apparently the authors were not aware of the study reported by Chenausky *et al.*,[37] where several articulation phenomena were modeled. In the study the authors considered a total of ten patients with Parkinson's disease and twelve healthy speakers. All the PD patients underwent a deep-brain-stimulation surgery (DBS) and their speech was recorded both on-stim and off-stim, i.e., with the electrical stimulator turned on and turned off. The participants were asked to produce rapid repetitions of the syllables /pa/ and /ka/. The authors studied several articulation phenomena in those speech tasks including syllable rate and syllable-length variability, syllable length patterning, vowel fraction, voice onset time (VOT) variability, and stop consonant spirantization. According to their findings, these articulation-based measures are suitable to assess speech-related improvements after the DBS surgery. This study provides a set of suitable measures that describe the articulation capability of PD patients. The features presented in this study describe articulatory deficits mainly from the duration point of view, e.g., syllable duration, vowel fraction, and VOT, and motivate further research on articulatory features considering energy-based measures calculated upon unvoiced frames.

Rusz *et al.*[38] considered a group with 20 early PD patients and 15 HCs (Czech native speakers). The authors analyzed *vowel articulation* across different speaking tasks including sustained phonations of the vowels /a/, /i/, and /u/, sentence repetition, reading text, and monologue. The set of features was comprised of measures of the first (F_1) and the second formant (F_2), VAI, VSA, and the quotient F_{2i}/F_{2u}. The authors claim that sustained phonations are not appropriate to evaluate vowel articulation in PD patients, while monologue is the most sensitive task to differentiate between PD patients and HC. The results indicate that it is possible to separate between PD patients and HCs with classification scores of about 80% when different articulation measures (i.e., VSA and F_{2i}/F_{2u}) are applied on the monologue.

Recently, Tsanas *et al.*[17] analyzed the impact of LSVT (Ref. 27) in the speech therapy of patients with PD. The authors measured a total of 309 dysphonia features to assess whether a sustained phonation is "acceptable" or "unacceptable" according to the clinical criteria of six experts. The system was evaluated on 126 phonations of the vowel /ʌ/ uttered by 14 PD patients. The LOGO (fit locally and think globally) feature selection algorithm was applied to find the most discriminant subset of features. The subset was selected following a tenfold CV strategy. The feature selection process was repeated 100 times on the training sets to avoid overfitting. The final subset of features was formed following a voting scheme. RF and SVM were used to discriminate between "acceptable" and "unacceptable" phonations. The authors reported a classification score of 90% considering a subset of features with 10 measures. Although the system was tested following a CV strategy with 10 folds, note that in this study the speaker independence is not guaranteed, leading to optimistic results.

Novotný *et al.*[39] presented a study where different articulatory deficits in speech of people with PD were modeled. The authors considered a total of 46 speakers, 24 of them with PD (20 male and 4 female). The group of HCs includes 15 males and 7 females. All participants (PD and HCs) had no history of speech therapy. The speech task performed by the speakers consisted of the rapid repetition of the syllables /pa-ta-ka/. The task was repeated twice per speaker. No limits in the number of repetitions were imposed. The authors calculated 13 features to describe six different articulatory aspects of speech, including vowel quality, coordination of laryngeal and supralaryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing. The authors reported a classification result of 88% in separating speech signals of PD patients and HCs. The results reported in this study confirm previous observations made by other authors who reported imprecise articulation as the most predominant characteristic of PD-related dysarthria. These results represent a step forward in the automatic evaluation of articulation in PD speech, not only because they were obtained automatically, but also because the evaluation is performed with a discriminative criterion, which allows the analysis of accuracy, specificity, and sensitivity of the method. The drawback of this study is that it was performed with a relatively small number of participants, thus further experiments considering more patients are required in order to obtain more conclusive results.

In the same year, Orozco-Arroyave *et al.*[40] performed automatic classification of speech signals from people with PD and HCs considering three different languages: German, Czech, and Spanish. The set of recordings considered in the three languages includes (1) 6 words uttered by 176 German native speakers (88 with PD and 88 HCs), (2) 13 words spoken by a total of 100 Spanish native speakers from Colombia (50 with PD and 50 HCs), and (3) the rapid repetition of the syllables set /pa/-/ta/-/ka/ (DDK analysis), which was uttered by 42 Czech speakers as well as by the Colombian and German ones. The authors presented a method based on the systematic separation of voiced and unvoiced segments of speech. The characterization and classification processes were performed considering each kind of segment separately. For voiced sounds, the authors calculated 12 MFCCs, three different noise measures, and the first two formants, while the unvoiced sounds were modeled using 12 MFCCs and the energy measured over 25 bands scaled according to the Bark scale.[41] The authors reported results from the voiced and unvoiced sounds separately. For the case of unvoiced sounds, the maximum reported accuracies obtained with Spanish and German words were 99% and 96%, respectively. For the case of /pa/-/ta/-/ka/, accuracies of 97% for German and Czech were reported, while for Spanish, 99% was reached. The highest accuracy reported using voiced

sounds on Spanish data was 84% with the word "petaka." For the case of German recordings, the highest accuracy obtained with voiced sounds was 73% with the word "perlenkettenschachtel." The results with voiced sounds of /pa/-/ta/-/ka/ were 90%, 69%, and 80% for Czech, German, and Spanish, respectively.

Note that the data considered in that work included recordings of the three languages but only with DDK analysis and isolated words. Although this task allows the assessment of different articulators, namely, lips, tongue, and velum, these recordings do not correspond to continuous speech and do not contain articulatory and prosody information of each particular language. The analysis of continuous speech signals in different languages has not been addressed in the literature.

From the reviewed literature it is possible to identify different aspects in the evolution of the speech processing to model dysarthric signals.

(1) The phonation dimension of speech has been widely covered and analyzed considering different sets of features including analysis of stability, periodicity, noise content, nonlinear structure, spectral wealth, and others.
(2) Articulation has also been addressed in different papers; however, most of them were focused on vowel articulation. Thus, considering that PD patients develop problems in the correct pronunciation of stop and voiceless consonants,[42] further research is required to model consonant sounds, unvoiced frames, and other speech units that require the control of different muscles and limbs involved in the speech production process. Such uncontrolled production of consonants affects communication skills of PD patients and has additional impact on the prosody and fluency of their speech.[36] There are studies on quantitative analyses of articulation in speech of PD patients, however, more research is required to analyze specific phenomena observed during the production of consonants, especially unvoiced sounds. A couple of appropriate and insightful papers that motivated this research were presented by Chenausky[37] and Stevens.[43] Additionally, further research is also required to develop computer aided tools that help clinicians, speech therapists, and patients to evaluate and improve their performance during the therapy and to detect problems in the pronunciation of specific sounds.[36]
(3) Prosodic characteristics provide information about speech rate, pause, intonation, and general communication skills of people. These characteristics must be included in the evaluation of people with PD for a better understanding of the impact of the disease on speech.[44]
(4) One aspect that has not been widely addressed in the literature is the reliability of characterization and classification methods to assess speech of people with PD in different languages. The main challenge of such an analysis is the need for databases with recordings of different languages.

Furthermore, from the clinical and neurological points of view, it has been observed that people suffering from dysarthria (most PD patients) develop problems controlling the vagus and hypoglossal nerves, inducing problems pronouncing consonants that require pressure build-up in the mouth and lingual movements, respectively.[45] The most serious pronunciation problems occur mainly in the plosives /p/, /t/, /k/, /b/, /d/, and /g/, due to the developed impairments to control nerves and muscles involved in the movement of different articulatory organs, such as the lips, tongue tip, center of the tongue, tongue base, jaw, epiglottis, and larynx.[46]

Notwithstanding the evidence reported by clinicians, the research community has been mainly focused on modeling voiced frames. One possible reason is that the vocal folds comprise the most important subset of muscles and tissues involved in speech production. The scientists have modeled their movements and the glottal source accurately, however, as highlighted in several works,[42,45,46] for the case of dysarthric speech, there is also important information in the frames where the vocal folds should not vibrate. The modeling of such loss of control to produce these kind of frames (unvoiced sounds) should improve the modeling of dysarthric speech signals such as those produced by patients with PD.

The contributions of this paper include a simple, useful, and robust methodology to classify between speech of people with PD and HCs. The method consists of modeling the energy content of the unvoiced sounds in different speech recordings. Four speech tasks are considered: (1) isolated words, (2) DDK evaluation, (3) sentences, and (4) reading text. The speech tasks were uttered in three different languages: Czech, German, and Spanish (spoken in Colombia). As the recordings from each language were captured separately, the robustness and the validity of the methods presented here are tested not only in three languages, but also in different technical conditions, i.e., different microphones, sound cards, sampling frequencies, noise conditions, etc. Besides the experiments with recordings of the three languages, the method is validated through cross-language tests, i.e., the system is trained with one language and tested in another one. Experiments with all of the six possible combinations of the three languages for training and testing are performed, yielding promising results and opening the possibility to design computer aided tools to evaluate speech of people with PD in different languages. To the best of our knowledge, this is the first paper that addresses the problem of automatic classification of PD and HC speakers including continuous speech uttered in three different languages and also cross-language experiments.

The rest of the paper is organized as follows. Section II presents the details of all of the methods applied in this paper, including the methodology that is proposed here to characterize speech of people with PD and a brief description of the speech tasks. In Sec. III, details of the three databases with recordings in Spanish, German, and Czech are presented. Section IV includes details of the experiments and the obtained results. Section V includes the discussion about the evaluated speakers and the results obtained in the experiments. Finally, Sec. VI includes conclusions derived from this paper and shows potential applications and the limitations of the proposed method.

II. METHODS

The methodology of this study is comprised of three main stages: (i) preprocessing, (ii) speech modeling, and (iii) classification. The first stage consists of manual and automatic segmentations of the utterances. Manual segmentation is performed to remove the silence at the beginning and end of each recording; the automatic segmentation consists of the estimation of voiced regions to separate each utterance into voiced and unvoiced segments. Pauses and unvoiced frames shorter than 40 ms are excluded from the recordings. The second stage includes four different approaches for speech modeling:

(a) The utterances are recorded without the voiced/unvoiced (v/uv) separation, only pauses are removed, and feature vectors formed by MFCCs are modeled using the GMM-UBM approach.[31]
(b) Prosody analysis is performed on the voiced frames using different measures including those extracted from F_0, energies, duration, and pauses.[33]
(c) Noise content, MFCCs, and formant measures are extracted from voiced segments.
(d) The unvoiced segments are characterized by MFCCs and the energy of the signal distributed in 25 Bark bands, namely, Bark band energies (BBEs).[41]

The third stage of the methodology consists of the decision on whether a recording belongs to a speaker with PD or a HC; this decision is computed using a radial basis SVM with parameters γ and C. The described methodology is summarized in Fig. 1.

A. Preprocessing

Silences in the recordings were removed manually from the beginning and the end of each voice register. Then, the recordings were automatically segmented using the software PRAAT.[23] Before the estimation of features, the speech frames were windowed using Hamming windows of 40 ms with 20 ms of overlap.

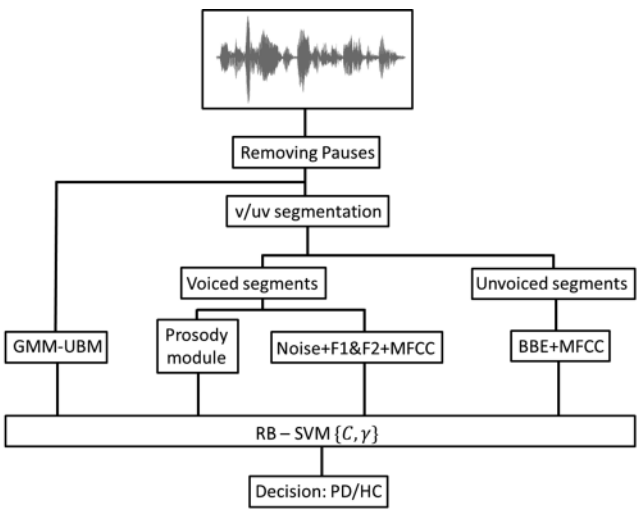

FIG. 1. Proposed methodology to address the experiments.

B. Speech modeling

1. Modeling based on MFCC-GMM supervectors

This modeling was performed following the same approach presented by Bocklet *et al.*[31] Hamming windows with 25 ms and time shifts of 10 ms were applied to the speech signal. A total of 13 MFCCs were taken (including C_0). MFCCs are standard features for speech recognition. They are a coarse representation of the short-time spectrum and have been shown to be appropriate to model irregular movements in the vocal tract.[47] MFCCs have been applied to model several speech pathologies such as dysphonia,[47] hypernasality,[48] and dysarthria.[18] The feature vector was formed with first and second order derivatives of the MFCCs (39-dimensional). Afterwards, the feature vector was modeled using the GMM-UBM strategy, i.e., a class-independent GMM with 128 Gaussians was trained on the whole data set by means of the EM algorithm. The mean values of the UBM were adapted by relevance MAP adaptation, finding speaker- and speech task-specific GMMs. The means of the Gaussians were then used as speaker- and task-specific features, forming 4992-dimensional ($128 \times 39 = 4992$) feature vectors.

2. Prosody analysis

Prosodic features were computed using the Erlangen prosody module.[33] For the sake of comparisons with the approaches presented in Secs. II B 3 and II B 4, the same segments with voiced frames are considered to be processed by the prosody module.

The set of features extracted with the prosody module comprises a total of 64 features. Seventeen of them are based on the utterances duration, 28 are based on the F_0 contour, and 18 are based on energy measures. The duration-based subset includes, among others, measures of the number of voiced frames, average duration of voiced frames, maximal length, and fraction of voiced frames. The subset with F_0-based features includes measures of the mean squared error (MSE) measured relative to the regression curve, and the regression coefficient of the F_0 contour within a frame, mean value of F_0, minimum and maximum of F_0, its value in the onset and offset, its temporal variation (jitter), its variation in amplitude (shimmer), and others. Referring to the energy-based features, this subset includes measures of the absolute value of energy within the words, maximum and minimum values of energy, MSE of the normalized energy curve relative to the regression curve, position of the maximum energy, and others. The detailed description of the features developed in the prosody module is presented by Zeißler *et al.*[33]

The features were grouped into one feature vector and four functionals were calculated: mean value (m), standard deviation (std), kurtosis (k) and skewness (sk), forming a 256-dimensional ($64 \times 4 = 256$) feature vector per recording.

3. Noise content, formant measures, and cepstral analysis of voiced frames

Voiced frames are characterized considering a set with 17 features. Three measures of noise content including HNR, normalized noise energy, and glottal to noise excitation ratio

along with the first two formants in Hz ($F1$ and $F2$), and 12 MFCCs. Voiced frames shorter than 40 ms were excluded from the analysis.

The four functionals (*m, std, k,* and *sk*) were also calculated from the measures, forming a 68-dimensional feature vector per recording.

4. Cepstral analysis and energy content of unvoiced frames

The main hypothesis of this paper is the existence of discriminant information in unvoiced frames to discriminate between people with PD and HC. One of the cues to state this hypothesis is that patients with PD develop problems pronouncing consonant sounds in the right moment due to their lack of control of different articulators like tongue, lips, and jaw.[45] Additionally, their mispronunciation problems are also related to impaired control of respiratory and laryngeal muscles, inducing the lack of intratracheal pressure while producing speech.[49] In order to model these articulatory and respiratory dysfunctions, the energy content of the unvoiced frames is modeled using 12 MFCCs and 25 BBEs.[41]

The four functionals (*m, std, k,* and *sk*) were also calculated here, forming feature vectors with 148 components ($37 \times 4 = 148$).

5. Classification and validation

The classification was performed with a radial basis SVM, with margin parameter C and a Gaussian kernel with parameter γ. The parameters C and γ were optimized through a grid-search with $1 < C < 10^4$ and $1 < \gamma < 10^3$. The selection criterion was based on the obtained accuracy on test data. A tenfold cross-validation strategy was employed for experiments with German and Spanish recordings, in which training and testing subsets never contained the same speakers. In the case of the Czech database, a leave-one-speaker-out strategy was followed.[40]

It is important to note that the folds are randomly assembled with the constraint of the balance of age and gender of the speakers in German and Spanish data. Additionally, the speaker independence is guaranteed during the training and testing. Thus although the selection criteria of the SVM parameters can lead to slightly optimistic accuracy estimates, the bias effect is minimal.

An SVM is used here due to its validated success in similar works related to automatic detection of pathological speech signals.[25,50]

C. Speech tasks

The database of each language includes different utterances distributed into four speech tasks, including (1) reading text, (2) sets of sentences (six uttered in Spanish, five in German, and three in Czech), (3) diadochokinetic evaluation through the rapid repetition of the syllables /pa/-/ta/-/ka/, and (4) sets of isolated words (13 uttered in Spanish, 6 in German, and 11 in Czech). Further details with the texts of each speech task are provided in the Appendix.

The texts evaluated on each language are balanced. The participants were asked to read the texts at their normal intonation and speech rate. The average duration in seconds for the recordings of patients and controls are, respectively, 18.6 ± 6.3 and 17.7 ± 3.8 for Spanish, 46.0 ± 11.3 and 46.4 ± 7.4 for German, and 37.8 ± 6.2 and 38.2 ± 4.4 for Czech. The sets of sentences uttered in Spanish and German are simple from the syntactic and lexical points of view. The three Czech sentences differ only in some words among them. The DDK evaluation allows the assessment of the capability of PD patients to do the correct occlusion of the oral cavity, performed by the lips in the case of /pa/, by the tongue in the case of /ta/, and by the velum in the case of /ka/.[51] Note that these recordings would even allow the evaluation of speech signals independent of native language of the patients. Isolated words are included because in a computational tool, the therapist or patient can determine more accurately which kind of articulatory movements are being evaluated with a particular word or set of words.

III. EXPERIMENTAL SETUP

A. The data

1. Spanish

This database contains speech recordings of 50 patients with PD and 50 HCs sampled at 44.1 KHz with 16 resolution-bits. These recordings were captured in noise controlled conditions, in a sound proof booth. All of the speakers are balanced by gender and age. The age of the 25 male patients ranges from 33 to 77 (mean 62.2 ± 11.2) and the age of the 25 female patients ranges from 44 to 75 (mean 60.1 ± 7.8). For the case of the HCs, the age of the 25 men ranges from 31 to 86 (mean 61.2 ± 11.3) and the age of the 25 women ranges from 43 to 76 (mean 60.7 ± 7.7). All of the patients were diagnosed and labeled by neurologist experts. The labels of their neurological evaluation were assigned according to the UPDRS-III and Hoehn & Yahr scales,[5] with mean values of 36.7 ± 18.7 and 2.3 ± 0.8, respectively. The average duration of the disease prior to recording (in years) was 10.7 ± 9.2. The speech samples were recorded with the patients in the ON-state, i.e., no more than 3 h after the morning medication. None of the people in the HC group has a history of symptoms related to Parkinson's disease or any other kind of neurological disorder. Further details of this database are provided by Orozco-Arroyave *et al.*[52]

2. German

This corpus consists of 176 German native speakers. The set of patients includes 88 persons (47 men and 41 women). The age of male patients ranges from 44 to 82 (mean 66.7 ± 8.4), while the age of the female patients ranges from 42 to 84 (mean 66.2 ± 9.7). The HC group contains 88 speakers (44 men, 44 women). The age of the men ranges from 26 to 83 (mean 63.8 ± 12.7), and the age of the women is from 54 to 79 (mean 62.6 ± 15.2). The mean values of the neurological evaluation performed on all of the patients according to the UPDRS-III and Hoehn & Yahr

scales are 22.7 ± 10.9 and 2.4 ± 0.6, respectively. The average duration of the disease prior to recording (in years) is 7.1 ± 5.8. The speech samples were also recorded with the patients in the ON-state. The voice signals were sampled at 16 kHz with 16 resolution-bits. Skodda *et al.*[15] describe this corpus in more detail.

3. Czech

A total of 36 Czech native speakers were recorded (all were men), 20 of them were diagnosed with idiopathic PD and their age ranges from 41 to 60 (mean 61 ± 12). The age of the HC speakers range from 36 to 80 (mean 61.8 ± 13.3). The mean values of the neurological evaluation of the patients, according to the UPDRS-III and Hoehn & Yahr scales, are 17.9 ± 7.3 and 2.2 ± 0.5, respectively. All of the patients included in this database were newly diagnosed with PD, and none of them had been medicated before or during the recording session. The voice signals were sampled at 48 KHz with 16 resolution-bits. The average duration of the disease prior to recording (in years) is 2.4 ± 1.7. Since the Czech participants were diagnosed with PD in the same moment of the recording session, this disease duration was obtained as a self report of patients according to the occurrence of the first motor impairment symptoms. Further details of this database are described by Rusz *et al.*[38]

Figure 2 shows the age distribution of the participants from the three databases, the distribution of the UPDRS-III values, and the distribution of time after PD diagnosis.

IV. EXPERIMENTS AND RESULTS

The utterances of each speech task are evaluated independently per language. Additionally, cross-language experiments are performed following a two-step strategy, i.e., (1) the system is trained with recordings of one language and tested on the other ones, and (2) subsets of the target language are included in the training set (another language) and excluded from the test set incrementally (from 10% to 80%) while maintaining strict separation between the list of speakers in the train and test sets. With this incremental procedure it is possible to observe the evolution of the system accuracy while more samples from a second language are added to the training stage.

The results obtained on each experiment are presented in the following Secs. IV A–IV E and are discussed in terms of the area under the ROC curve (AUC) values, allowing objective comparisons among the different systems.[30]

A. Results on reading texts

Results obtained with each text read in the three languages are presented in Table I. The features applied on voiced segments show AUC values ranging from 0.78 to 0.85. These results are consistent with previous observations, indicating both the presence of noise and the articulatory problems of people with PD evaluated using reading texts.[53] Since the AUC values obtained with the prosodic modeling are 0.79, 0.83, and 0.76 on Spanish, German, and Czech recordings, respectively, impairments in the speech rate, intonation, and general prosodic features are also evidenced in the recordings of the three databases. Results obtained with the GMM-UBM modeling approach are around 0.80 in the three databases, indicating that this acoustic modeling can also be used to screen speech impairments in PD patients. The accuracies obtained with German data are consistent with previous studies[31] where a similar approach is addressed and general accuracies of 81.9% are reported using the same data set. Regarding the results obtained with the proposed modeling on unvoiced frames, note that in all of the languages this approach exhibited the highest AUC values: 0.99, 0.93, and 0.85 for Spanish, German, and Czech

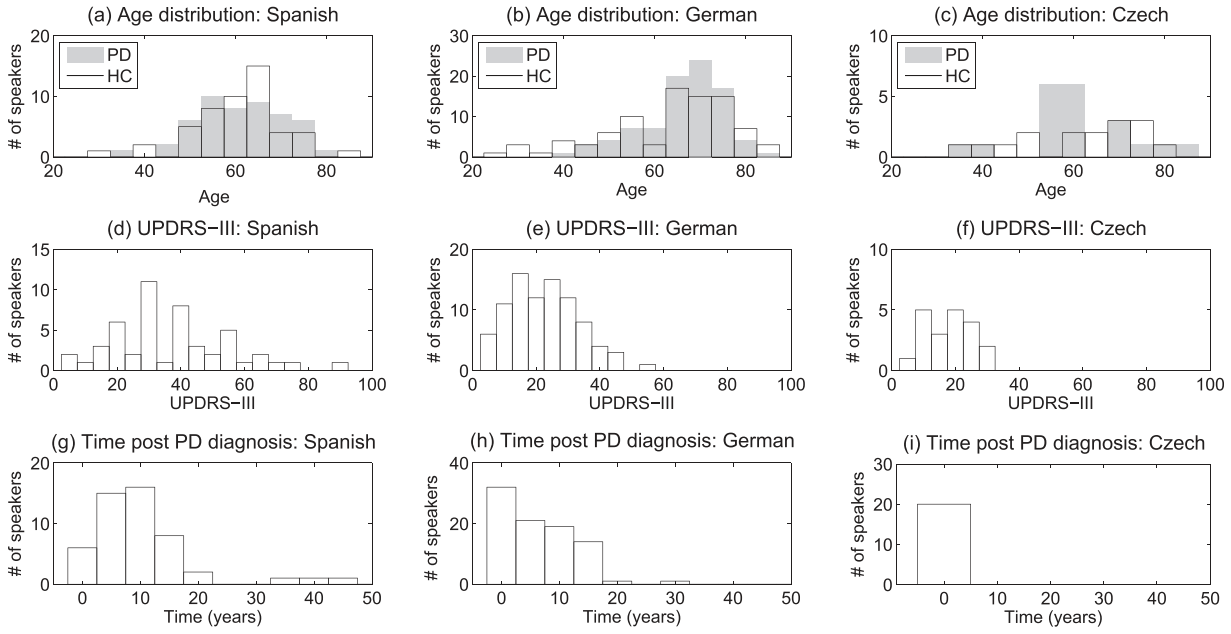

FIG. 2. Age, UPDRS-III, and time after PD diagnosis distribution for the three databases.

TABLE I. Results obtained with read texts of the three languages and modeled using the four feature sets studied in this paper.

		Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Spanish	Noise+F1&F2+MFCC	82 ± 10.3	88 ± 13.9	76 ± 24.6	0.84
	Prosody	77 ± 12.5	86 ± 9.7	68 ± 21.5	0.79
	GMM-UBM	82 ± 13.2	86 ± 13.5	78 ± 28.9	0.78
	Unvoiced	**97 ± 4.8**	**98 ± 6.3**	**96 ± 8.4**	**0.99**
German	Noise+F1&F2+MFCC	78.4 ± 5.2	76.1 ± 14.3	80.9 ± 12.8	0.78
	Prosody	83.9 ± 8.1	77.1 ± 12.2	90.8 ± 7.2	0.83
	GMM-UBM	78.9 ± 9.7	70.6 ± 15.0	87.4 ± 11.9	0.80
	Unvoiced	**94.3 ± 3.9**	**95.4 ± 7.9**	**93.3 ± 7.8**	**0.93**
Czech	Noise+F1&F2+MFCC	78.3 ± 25.3	100 ± 0.0	56.7 ± 50.6	0.85
	Prosody	78.7 ± 25.3	82.7 ± 38.7	74.7 ± 44.0	0.76
	GMM-UBM	80.6 ± 24.4	69.4 ± 46.7	86.1 ± 35.1	0.83
	Unvoiced	**85.0 ± 23.4**	**76.7 ± 43.6**	**93.3 ± 18.9**	**0.85**

recordings, respectively. These results show that there is discriminant information in unvoiced sounds, and it can be extracted using energy-based features.

Results are summarized in Fig. 3, which includes ROC curves of the results obtained from the texts read in the three languages. Note that the best performance is shown by the system that is based on the modeling of the energy content of the unvoiced segments.

Further experiments were performed in order to understand which of the features included here to characterize the voiced and unvoiced frames are giving the information about the presence or absence of the disease. Different combinations of features were tested. First, the feature matrices obtained from voiced and unvoiced segments were merged, and second, MFCCs and BBEs calculated from unvoiced segments and from utterances without the v/uv segmentation were tested separately. As the obtained results in all of the additional experiments ranged below or around the same accuracies compared to those obtained with MFCCs and BBEs calculated on the unvoiced segments, we decided to perform all of the remaining experiments (with sentences, words, DDK analysis, and cross-language) considering the same characterization approaches.

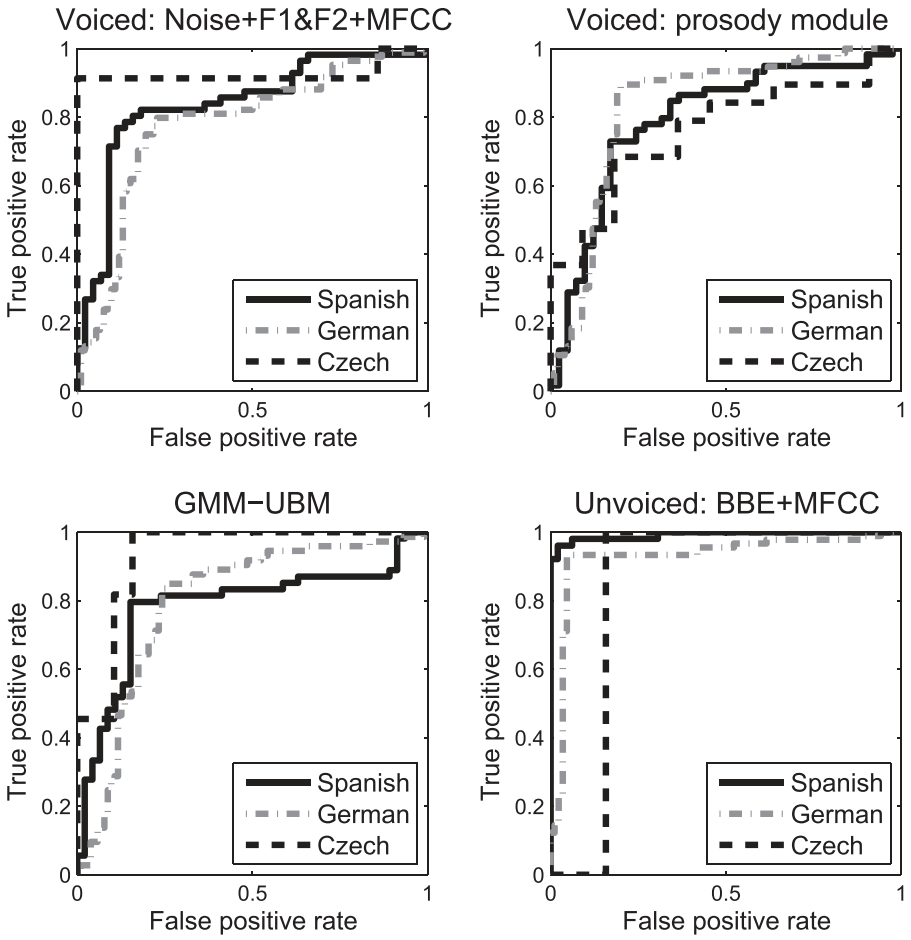

FIG. 3. ROC curves obtained from read texts modeled using the four different characterization approaches studied in this paper.

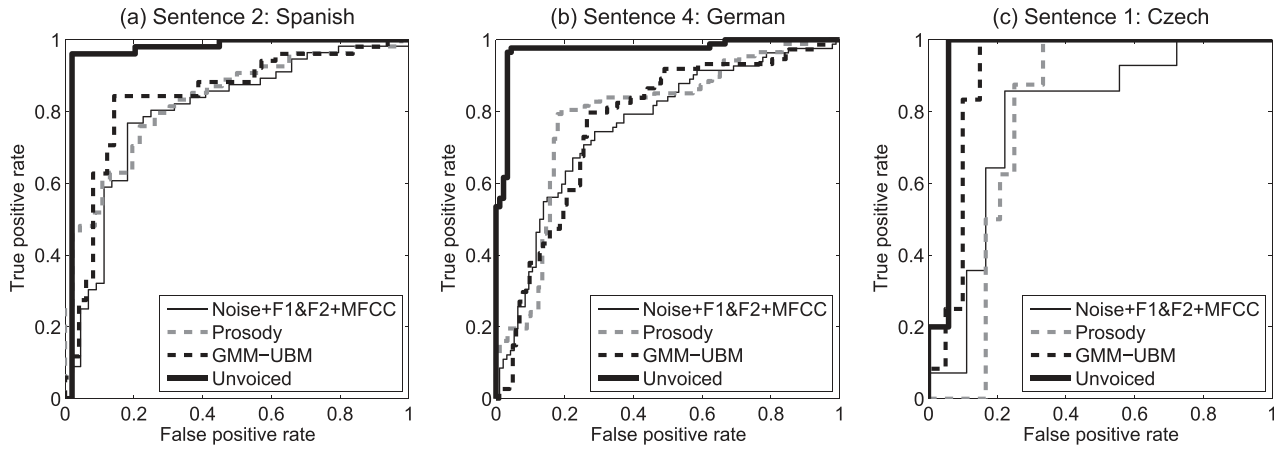

FIG. 4. ROC curves of different sentences uttered in the three languages and modeled with the four approaches studied in this paper.

B. Results on sentences

The ROC curves obtained with sentence 2 in Spanish (*"Los libros nuevos no caben en la mesa de la oficina"*), sentence 4 in German (*"Das Fest war sehr gut vorbereitet"*), and sentence 1 in Czech (*"Kolik máte ted u sebe asi peněz?"*) are depicted in Fig. 4.

The results reported here suggest that the characterization approach proposed in this paper is better than others typically addressed in the literature. The methodology proves to be robust and highly accurate, since it is tested on continuous speech signals and on different technical conditions and contexts, i.e., with utterances recorded independently in three different languages.

Further details with the results obtained on each sentence are provided in the Appendix.

C. Results on DDK evaluation

The results obtained with the DDK evaluation performed on the three languages are shown in Table II. Note that highest AUC values are again reached with the characterization approach based on unvoiced segments. The values are above 0.95 in all three languages, suggesting that the method is robust and accurate for discriminating between PD and HC speakers from recordings of the rapid repetition of syllables with stop consonants. With the other three approaches, the AUC values are below 0.90, except for the Czech utterances characterized with the GMM-UBM approach where results reach 0.93. These results are summarized in Fig. 5.

D. Results on words

Although the methodology proposed in this paper is conceived to be applied in continuous speech signals, several experiments with isolated words were also performed in order to evaluate its robustness and accuracy on specific syllabic groups. A total of 31 isolated words are evaluated here; 13 in Spanish, 6 in German, 12 in Czech. This set includes the same Spanish and German words previously evaluated by Orozco-Arroyave *et al.*[40] In this paper, Czech words and two additional characterization approaches are included.

The details with the results obtained on each word are presented in the Appendix. In general, the results suggest that the method proposed in this paper also performs better on isolated words. The AUC values obtained in 23 of the 31 words spoken in the three languages are above 0.90, indicating that the method is also robust for the automatic classification of PD and HC speakers from isolated words. However, these results on words should be taken carefully because the analysis of isolated words reveal limitations on the proposed method. For instance, in utterances where the

TABLE II. Results of the DDK evaluation with recordings in Spanish, German, and Czech.

		Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Spanish	Noise+F1F2+MFCC	80 ± 9.4	90 ± 14.1	70 ± 19.4	0.82
	Prosody	80 ± 6.7	88 ± 13.9	72 ± 13.9	0.84
	GMM-UBM	82 ± 9.2	96 ± 8.4	68 ± 21.5	0.84
	Unvoiced	**99 ± 3.2**	**99 ± 0.0**	**98 ± 6.3**	**0.99**
German	Noise + F1F2+MFCC	69.8 ± 9.5	61.7 ± 24.9	77.2 ± 15.1	0.68
	Prosody	73.2 ± 11.4	75.8 ± 15.9	70.6 ± 11.9	0.72
	GMM-UBM	70.9 ± 8.3	64.3 ± 19.8	77.1 ± 15.7	0.70
	Unvoiced	**97.8 ± 2.9**	**98.9 ± 3.5**	**96.5 ± 5.6**	**0.98**
Czech	Noise+F1F2+MFCC	81.1 ± 24.7	79.3 ± 41.8	82.9 ± 38.3	0.65
	Prosody	84.6 ± 23.9	71.4 ± 46.9	97.9 ± 28.1	0.83
	GMM-UBM	86.9 ± 22.0	97.2 ± 16.7	69.4 ± 46.7	0.93
	Unvoiced	**93.6 ± 16**	**99.3 ± 2.7**	**87.9 ± 31.4**	**0.96**

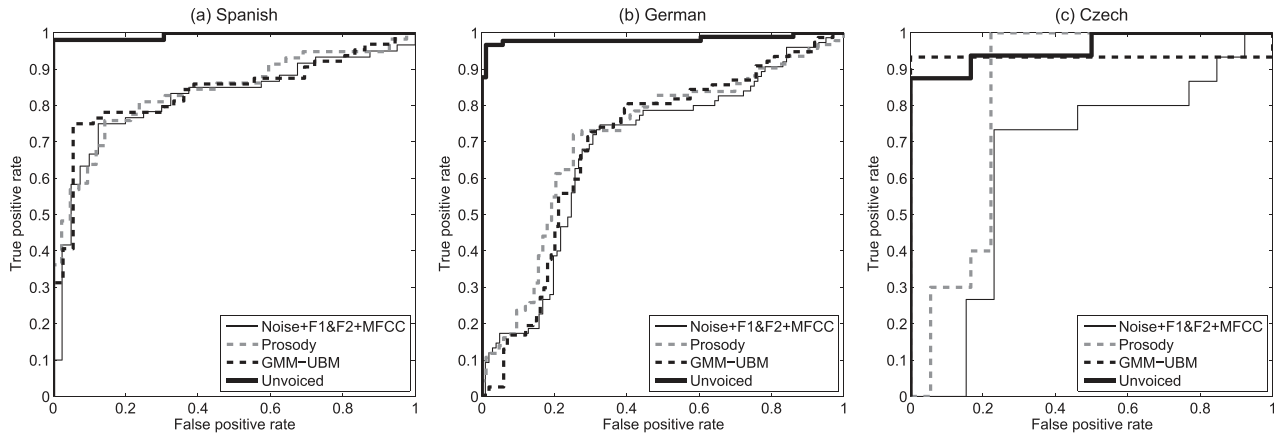

FIG. 5. ROC curves obtained with DDK evaluation in Spanish, German, and Czech. The results of the four modeling approaches studied in this paper are included.

unvoiced segments are not long enough to provide several windows of 40 ms, it is not possible to calculate the statistics of the features, or the estimates of statistics are not stable.

E. Results on cross-language experiments

The generalization capability of the proposed approach is tested through several cross-language experiments performed considering read texts and the DDK evaluation on each language. To address these experiments, all of the recordings were re-sampled to 16 kHz. Additionally, the cepstral mean subtraction process is applied in order to perform a channel normalization, avoiding possible bias introduced by the microphones and sound cards.

The experiments consisted on training the system with recordings of one language and testing with recordings of another one. Additionally, the improvement of the accuracy is analyzed when moving portions of the data in the target language to the data in the training set. The recordings of the target language are included in the test set and excluded from the training set to avoid bias. The results are summarized in Figs. 6 and 7. Note that the performance of the system improves from 60% to 99% depending on the task, the added fraction of the target language, and the combination of the training and test sets.

Figure 6 shows the results on read texts. When the system is trained with Spanish recordings and tested on German (part a), only 30% of the German recordings are required to be moved from the test set to the training set to reach accuracies of 90%. The resulting training set contains 152 recordings, 66% of them correspond to Spanish and the remaining 34% correspond to German. Conversely, the accuracies obtained when testing on the Czech set are above 80% when 50% of the test recordings are added to the train set and excluded from the test set (part a). This 50% of the Czech data represents 15% of the resulting training set when the train language is Spanish and 9% when the train language is German. When the system is trained with the German data and tested on Spanish, 50% of the test set needs to be added to the training set to reach accuracies above 90%. Note that the added data represent 22% of the resulting training set. Similarly, when the system is trained with the Spanish recordings and tested on Czech, at least 80% of the Czech samples need to be added to the train set to obtain accuracies of around 80% (part b). In this case those additional recordings represent 22% of the resulting training set. Finally, when the system is trained with Czech recordings and tested on Spanish or German the behavior is similar, i.e., the accuracy begins at 60% and increases incrementally up to 90% when recordings of the target language are added to the test set and excluded from the train sets (part c).

The results on the rapid repetition of the syllables /pa/-/ta/-/ka/ (DDK evaluation) are shown in Fig. 7. Note that when the system is trained with Spanish and tested on German, it reaches accuracies of 90% when adding only 20% of the test recordings, which means that 35% of the training set is formed with recordings of the target language (those recordings are excluded from the test subset). When 20% of the Czech recordings are moved to the test set (Spanish), the system reaches around 80% accuracy (part a).

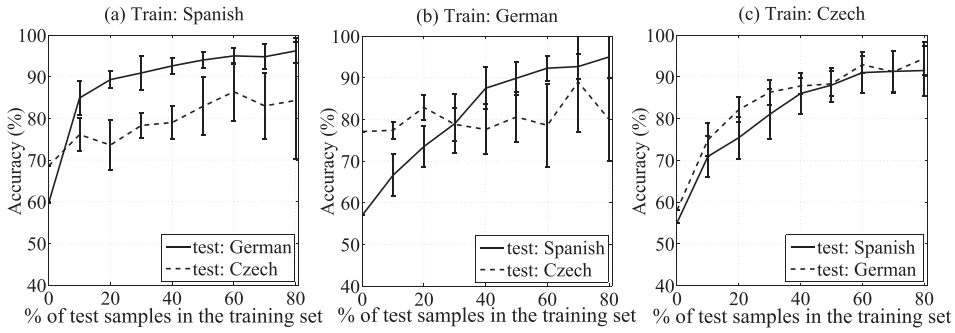

FIG. 6. ROC curves obtained with cross-language experiments from read texts.

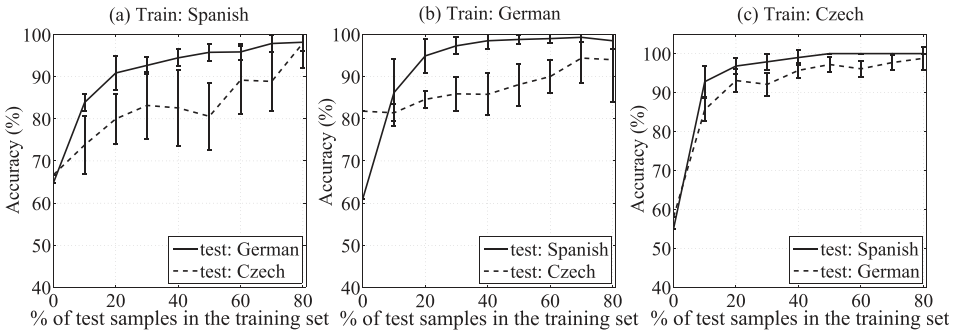

FIG. 7. ROC curves obtained with cross-language experiments from the DDK evaluation.

Note also that when the system is trained with German and tested on Spanish (part b), it can reach accuracies of 95% adding 20% of the test recordings, which represents 10% of the resulting training set. On the other hand, when the system is tested on the Czech recordings, it needs more than 60% of the test data to reach accuracies above 90%. Finally, if the system is trained with the Czech recordings and tested on Spanish or German (part c), it only needs to move 20% of the test set to the train set to reach accuracies of about 95%. Further details with the values of accuracy, sensitivity, specificity, and AUC are provided in the Appendix.

V. DISCUSSION

A. The patients

Three different data sets were used in this work. Considering that each data set was built independently, the most important biological aspects of the participants should be discussed in order to analyze their influence on the obtained results.

From parts a, b, and c in Fig. 2 it is possible to observe the balance in the age of the participants of the three databases. Regarding the neurological state of the participants, parts d, e, and f of Fig. 2 show that Czech patients are in early to middle stage of the disease (they were diagnosed in the same session where the recording took place) with UPDRS-III values ranging from 5 to 32. German patients exhibited UPDRS-III values around 23, ranging from 5 to 55, while Colombian patients have UPDRS-III values distributed from 5 to 92. As the impairments of speech increase with the extent of the disease,[53] one can expect that the speech of Colombian speakers is more compromised than the speech of German and Czech participants. This point can explain in some way the highest accuracies obtained with recordings uttered in Spanish. Another important aspect to be highlighted is that Czech patients were in the OFF state, while German and Colombian patients were in the ON-state. Theoretically, patients in the OFF-state should perform worse than patients in the ON-state, however, our experiments do not show this. It is worth noting that the impact of the medication on the speech of PD patients is not clear yet.[14,15]

Finally, Parts g, h, and i in Fig. 2 indicate the time after PD diagnosis. In Czech patients this time is 0 years, while German and Colombian participants were diagnosed about 7 and 10 years ago, respectively. Since PD is a progressive disorder, time after its diagnosis can be considered as another cue to state that Colombian patients were in a more advanced stage compared to the German and Czech participants.

B. The results

1. Results in read texts

The robustness of the proposed approach is validated in texts read in three different languages, with data sets recorded independently, and with different technical conditions, e.g., sampling frequency, microphones, and sound cards, among others. The approach proposed in this paper yields the highest accuracies in the three databases. Spanish and German recordings of the read texts exhibited accuracies of 97% and 94.3%, respectively. Note that the difference of the accuracies in Spanish and German is less than 3 percentage points, while results in Czech are around 85%.

As pointed out in the discussion about the patients, this difference can be explained by the fact that Czech patients were all in the early to middle stages of the disease, none of them had been diagnosed before the recording session.

Regarding the results obtained with the other three characterization approaches, the prosody module exhibited accuracies above 83% on German recordings, while the classical approach, based on noise measures, the first two formants, and the MFCCs, reaches accuracies of around 82% in Spanish. The best approach on Czech samples was that based on the GMM-UBM modeling.

According to the results obtained with the unvoiced features, it seems like the hypokinetic dysarthria suffered by PD patients is being modeled more accurately by the proposed characterization approach.

2. Results in sentences

The robustness of the proposed method is also validated in the experiments performed with sentences in Spanish and German; however, this behavior changes with Czech recordings. Three sentences were evaluated on this language, and the main difference among them was a couple of words and syllables. Highest accuracies are obtained with the prosody module on second and third sentences.

3. Results in DDK evaluation

This task shows accuracies above 90% in the three languages when the proposed approach was applied. For Spanish and German, the results are 99% and 97.8%, respectively, while for Czech recordings, the accuracy reaches 93.6%, which is one of the highest obtained along the evaluated speech tasks. Czech patients were in early to middle stages of the disease, while German and Spanish patients ranged from

middle to advanced stages. Results indicates that the proposed approach is able to perform automatic detection of PD in early and middle stages of the disease.

4. Results in isolated words

Specific movements in the vocal tract can be assessed through isolated words. In Spanish, the word "*campana*" shows a very high classification score. Since the production of this word requires the movement of the lips in the phone /p/ and the velum in the phone /mp/, the nasal-bilabial combination /mp/ allows the assessment of the movement in the lips and the velum. A similar analysis can be done when the German word "*Toilettenpapier*" is evaluated. This word also yields high classification results and its production also requires a nasal-bilabial combination (/np/).

The set of Czech words also showed high classification results. For instance, the production of the word *kuká* allows the evaluation of the velar movement to produce the phone /k/ and the tongue movement to produce the vowels /u/ and /a/.

The results presented in this paper seem to show that the evaluation of isolated words is worthwhile for the assessment of specific movements in the vocal tract. Further experiments are required to draw stronger conclusions, i.e., grouping words with similar syllabic groups or grouping syllables that require similar movements for their production,[54] which may lead to methods for screening specific articulatory movements in the vocal tract.[43]

5. Results in the cross-language experiments

The results discussed above show the robustness of the proposed method in different databases with speech samples recorded in different technical conditions. Additionally, the generalization capability of the method is tested by crossing the three databases. The results in the cross-language experiments show that incrementally adding samples from the test language into the training set quickly helps to improve the performance of the system. In general, the accuracies range from 60% to 100% when recordings of the language that is going to be tested are moved from testing and added to training.

According to the results in read texts, higher accuracies are reached when Spanish recordings are used for training and German samples for testing. In the same way, when German recordings are used for training and Spanish samples are used for testing, the highest accuracies are reached. When Czech recordings are used for training the behavior on both test sets (Spanish and German) is similar, and at least 60% of the test samples are required in the training set to reach accuracies around 90%.

The results on the DDK evaluation show that Czech samples are the most appropriate to be used in training because only 20% of the test samples (either Spanish or German) are required in training to reach accuracies above 90%. When tests are performed on Czech recordings, the behavior is similar with any training set (German or Spanish), requiring at least 60% of the test samples in training to reach accuracies of about 90%. The need for such a high number of recordings from the target language data could be explained by the fact that Czech patients were all in early to middle stages of the disease and thus, probably experience less impact on their articulatory capability. This apparent contradiction needs to be analyzed in more detail in further research. On the other hand, the Czech patients were all in the OFF-state, so articulatory deficits should have been reflected more clearly in the experiments. Further experiments comparing speech of early PD patients with advanced PD patients in the ON- and OFF-states could be interesting.

VI. CONCLUSIONS

An innovative and robust methodology for the characterization of continuous speech of people with Parkinson's disease is presented. The methodology is based on the automatic segmentation of voiced and unvoiced segments, voiced being defined as those frames where the vocal folds vibrate and unvoiced frames as those where vocal folds do not vibrate. The energy content of the unvoiced sounds is modeled using 12 MFCCs and 25 BBEs.

The method is tested on different speech tasks performed by speakers in Spanish, German, and Czech. The recordings contain four speech tasks, including read texts, sentences, isolated words, and the rapid repetition of the syllables /pa/-/ta/-/ka/.

The proposed approach is directly compared with other "standard" approaches classically used for speech modeling, such as (1) noise measures, MFCCs, and vocal formants extracted from voiced segments, (2) MFCCs extracted from the utterances without pauses and modeled using a GMM-UBM strategy, and (3) different prosodic features extracted with the Erlangen prosody module.

According to the results, our method proves to be more accurate than the classical approaches, reaching accuracies that range from 85% to 95% (depending on the language, severity of the disease, and the speech task) in the automatic classification of speech of people with PD and HCs.

Czech patients were in early to middle stages of the disease, while German and Spanish patients were ranging from middle to advanced stages. This indicates that the proposed approach is able to perform automatic detection of PD in early and middle stages of the disease.

The data of each language were recorded using different microphones, sound cards, sampling frequencies, noise conditions, etc., indicating that the method is also robust against different technical conditions, and is a promising alternative for future implementations of computer aided tools to perform the automatic evaluation of dysarthric speech signals.

The recordings of read texts and the DDK evaluations of the three languages are also evaluated on cross-language experiments, validating the robustness and reliability of the method. The generalization capability of the method is evidenced in both tasks, read texts and DDK, thus it can be stated that it can be used to screen information of speech in continuous speech signals and in particular, articulatory exercises like the DDK evaluation.

DDK evaluation seems to be more appropriate than the read texts to evaluate Parkinsonian speech signals in cross-language tests. This could be due to its simplicity to pronounce and its ability to make the speaker produce specific movements in the vocal tract.

From the results presented here it is possible to address further experiments in read texts and sentences. For instance, grouping syllables that require the movement of the same articulators, assessing the capability of a patient to move particular parts of the vocal tract.

The method presented in this paper has several limitations. For instance, when it is applied to isolated words and the unvoiced segments are not long enough to contain several windows of 40 ms length, is not possible to calculate the statistics of the features or the estimates are not stable. Another limitation of this study is that we only considered speech recordings of PD patients and HCs. We did not include patients with other type of neurological diseases, thus the suitability of the methods presented here is only demonstrated in hypokinetic dysarthria due to Parkinson's disease but not from any other neurological disorder.

The method suggests the possibility to address further analysis of disordered speech considering the borders between voiced and unvoiced sounds, making possible the evaluation of specific movements of the articulators or tissues in the vocal tract.

The method presented here seems to be a very promising alternative for the development of computer aided tools for the accurate evaluation of different speech disorders that affect the movement of several articulators during the speech production process.

ACKNOWLEDGMENTS

J.R.O.-A. was supported by grants of COLCIENCIAS through the call No. 528 "generación del bicentenario 2011." This work was also financed by COLCIENCIAS through project No. 111556933858. The research leading to these results has received funding from the Hessen Agentur, Grant Nos. 397/13-36 (ASSIST 1) and 463/15-05 (ASSIST 2). The authors express thanks to CODI at Universidad de Antioquia for its support through "estrategia de sostenibilidad 2014-2015 de la Universidad de Antioquia." This project was also funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, under Grant No. 9-135-1434-HiCi. The authors, therefore, acknowledge with thanks DSR technical and financial support.

APPENDIX

1. Reading texts

Spanish: *"Ayer fui al médico. Qué le pasa? Me preguntó. Yo le dije: Ay doctor! Donde pongo el dedo me duele. Tiene la uña rota? Sí. Pues ya sabemos qué es. Deje su cheque a la salida."*

German: *"Schildkröteninvasion: Von einer gewaltigen, von den Behörden geschützten Invasion wird zur Zeit die Golf- und Pazifikküste Mexikos heimgesucht: Wie alljährlich im Juni kommen Hunderttausende von Schildkröten aus dem Meer, um an Land ihre Eier abzulegen. Allein in der Nähe von Tampico wurden etwa* 5000 *Schildkröten beobachtet. Insgesamt wird in den kommenden Wochen mit einer Invasion von mehr als einer halben Million Schildkröten gerechnet. Die mexikanischen Behörden lassen die Legeplätze sorgfältig bewachen, um den Diebstahl von Eiern zu verhindern und ausreichend Schildkrötennachwuchs sicherzustellen."*

Czech: *"Když člověk po prvé vsadí do země sazeničku, chodí se na ni dívat třikrát denně: takco, povyrostla už nebo ne? I tají dech, naklání se nad ní přitlačí trochu půdu u jejích kořínků, načechrává jí lístky a vůbec ji obtěžuje různým konáním, které považuje za užitečnou péči. A když se sazenička přesto ujme a roste jako z vody, tu člověk žasne nad tímto divem přírody, má pocit čehosi jako zázraku a považuje to za jeden ze svých největších úspěchů."*

2. Sentences

Spanish:

(1) *Laura sube al tren que pasa.*
(2) *Los libros nuevos no caben en la mesa de la oficina.*
(3) *Luisa Rey compra el colchón duro que tanto le gusta.*
(4) *Mi casa tiene tres cuartos.*
(5) *Omar, que vive cerca, trajo miel.*
(6) *Rosita Niño, que pinta bien, donó sus cuadros ayer.*

TABLE III. Results obtained from sentences spoken in Spanish.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
		Sentence 1		
Noise+F1&F2+MFCC	74 ± 12.7	84 ± 18.4	64 ± 32.4	0.69
Prosody	81 ± 9.90	88 ± 16.9	74 ± 26.8	0.81
GMM-UBM	86 ± 11.7	90 ± 14.1	82 ± 17.5	0.86
Unvoiced	**92 ± 13.2**	**94 ± 9.70**	**90 ± 19.4**	**0.93**
		Sentence 2		
Noise+F1&F2+MFCC	78 ± 9.20	84 ± 15.8	72 ± 16.9	0.80
Prosody	76 ± 10.7	80 ± 23.1	72 ± 16.9	0.78
GMM-UBM	85 ± 8.50	86 ± 13.5	84 ± 18.4	0.84
Unvoiced	**97 ± 4.80**	**98 ± 6.30**	**96 ± 8.40**	**0.97**
		Sentence 3		
Noise+F1&F2+MFCC	81 ± 11.9	88 ± 13.9	74 ± 18.9	0.83
Prosody	84 ± 12.6	84 ± 26.3	84 ± 15.8	0.83
GMM-UBM	84 ± 10.8	82 ± 14.8	86 ± 16.5	0.83
Unvoiced	**94 ± 6.90**	**94 ± 9.70**	**94 ± 13.5**	**0.95**
		Sentence 4		
Noise+F1&F2+MFCC	78 ± 11.4	78 ± 17.5	78 ± 17.5	0.79
Prosody	73 ± 8.20	88 ± 13.9	58 ± 22.0	0.73
GMM-UBM	86 ± 12.7	80 ± 18.9	92 ± 10.3	0.86
Unvoiced	**90 ± 9.42**	**90 ± 14.1**	**90 ± 14.1**	**0.90**
		Sentence 5		
Noise+F1&F2+MFCC	77 ± 11.6	80 ± 13.3	74 ± 21.2	0.81
Prosody	**78 ± 7.90**	**94 ± 13.5**	**62 ± 22.0**	**0.87**
GMM-UBM	81 ± 11.0	78 ± 23.9	84 ± 12.7	0.83
Unvoiced	81 ± 7.40	84 ± 18.4	78 ± 14.8	0.82
		Sentence 6		
Noise+F1&F2+MFCC	79 ± 11.0	84 ± 12.7	74 ± 21.2	0.82
Prosody	77 ± 12.5	86 ± 9.70	68 ± 25.3	0.82
GMM-UBM	88 ± 13.2	88 ± 13.8	88 ± 19.3	0.89
Unvoiced	**90 ± 9.40**	**92 ± 10.3**	**88 ± 13.9**	**0.91**

TABLE IV. Results obtained from sentences spoken in German.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
		Sentence 1		
Noise+F1&F2+MFCC	71.6 ± 7.1	71.7 ± 13.2	71.8 ± 10.4	0.70
Prosody	72.5 ± 12.1	73.8 ± 16.4	71.4 ± 17.4	0.71
GMM-UBM	70.3 ± 10.1	57.1 ± 22.3	84.3 ± 11.9	0.69
Unvoiced	**93.1 ± 5.3**	**91.9 ± 7.7**	**94.2 ± 8.5**	**0.94**
		Sentence 2		
Noise+F1&F2+MFCC	72.1 ± 4.4	60.9 ± 17.2	83.1 ± 13.3	0.73
Prosody	76.1 ± 7.6	77.4 ± 14.7	75 ± 11.7	0.76
GMM-UBM	71.5 ± 6.5	60.0 ± 15.7	82.8 ± 16.2	0.71
Unvoiced	**85.8 ± 6.2**	**83.8 ± 13.9**	**87.4 ± 14.1**	**0.86**
		Sentence 3		
Noise+F1&F2+MFCC	77.4 ± 10.1	77.7 ± 12.1	77.8 ± 15.6	0.75
Prosody	81.7 ± 6.1	86.1 ± 11.1	76.9 ± 12.7	0.84
GMM-UBM	72.2 ± 7.9	67.1 ± 10.9	77.1 ± 15.7	0.72
Unvoiced	**96.1 ± 5.4**	**95.4 ± 5.9**	**96.7 ± 7.5**	**0.96**
		Sentence 4		
Noise+F1&F2+MFCC	72.7 ± 9.3	69.4 ± 12.9	76.1 ± 12.2	0.76
Prosody	80.1 ± 6.3	78.9 ± 19.2	80.7 ± 14.9	0.79
GMM-UBM	74.9 ± 7.9	67.2 ± 15.9	82.9 ± 12.1	0.77
Unvoiced	**96.7 ± 5.9**	**95.6 ± 7.7**	**97.8 ± 7**	**0.97**
		Sentence 5		
Noise+F1&F2+MFCC	78.4 ± 8.9	80.6 ± 10.8	76.3 ± 15.9	0.77
Prosody	81.2 ± 8.2	78.3 ± 11.5	84.3 ± 9.3	0.81
GMM-UBM	74.4 ± 6.7	65.9 ± 11.5	83.1 ± 11	0.73
Unvoiced	**94.3 ± 5.4**	**97.6 ± 4.9**	**90.8 ± 10.6**	**0.95**

TABLE V. Results obtained from sentences spoken in Czech.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
		Sentence 1		
Noise+F1&F2+MFCC	77.5 ± 25.4	71.3 ± 45.8	83.8 ± 37.2	0.79
Prosody	71.9 ± 24.2	51.9 ± 50.9	91.9 ± 23.9	0.79
GMM-UBM	77.3 ± 24.9	66.7 ± 47.8	94.4 ± 23.2	0.88
Unvoiced	**93.1 ± 17.5**	**88.8 ± 32.3**	**97.5 ± 10**	**0.95**
		Sentence 2		
Noise+F1&F2+MFCC	77.8 ± 25.4	77.5 ± 42.8	78.1 ± 41.7	0.78
Prosody	**89.7 ± 20.3**	**84.4 ± 36.8**	**95 ± 18.4**	**0.89**
GMM-UBM	81.6 ± 24.1	66.7 ± 47.8	86.1 ± 35.1	0.83
Unvoiced	86.3 ± 22.6	79.4 ± 41.2	93.1 ± 19.3	0.80
		Sentence 3		
Noise+F1&F2+MFCC	81.6 ± 24.7	94.4 ± 20.9	68.8 ± 47.1	0.91
Prosody	**94.4 ± 16.2**	**93.8 ± 25.0**	**95.0 ± 20.0**	**0.93**
GMM-UBM	78.9 ± 24.7	91.7 ± 28.0	66.7 ± 47.8	0.83
Unvoiced	85.6 ± 23.2	87.5 ± 34.2	83.8 ± 37.4	0.86

TABLE VI. Results obtained from the isolated words of the Spanish data.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
		Atleta		
Noise+F1F2+MFCC	82 ± 9.2	86 ± 13.5	78 ± 19.9	0.79
Prosody	83 ± 11.6	96 ± 8.4	70 ± 21.6	0.87
GMM-UBM	76 ± 11.7	92 ± 13.9	60 ± 26.7	0.80
Unvoiced	**99 ± 3.2**	**98 ± 6.3**	**99 ± 0.0**	**0.99**
		Campana		
Noise+F1F2+MFCC	73 ± 11.6	86 ± 13.5	60 ± 23,1	0.76
Prosody	74 ± 10.8	80 ± 13.3	68 ± 28.6	0.76
GMM-UBM	70 ± 11.5	86 ± 21.2	54 ± 31.3	0.69
Unvoiced	**99 ± 3.2**	**98 ± 6.3**	**99 ± 0.0**	**0.99**
		Gato		
Noise+F1F2+MFCC	76 ± 15.1	84 ± 15.8	68 ± 16.9	0.76
Prosody	76 ± 12.6	70 ± 14.1	82 ± 19.9	0.80
GMM-UBM	86 ± 9.7	90 ± 10.5	82 ± 14.8	0.86
Unvoiced	**98 ± 6.3**	**98 ± 6.3**	**98 ± 6.3**	**0.98**
		Petaka		
Noise+F1F2+MFCC	84 ± 10.8	88 ± 16.9	80 ± 16.3	0.82
Prosody	81 ± 9.9	86 ± 13.5	76 ± 20.7	0.82
GMM-UBM	82 ± 10.3	96 ± 8.4	68 ± 16.9	0.87
Unvoiced	**97 ± 4.8**	**96 ± 8.4**	**98 ± 6.3**	**0.98**
		Braso		
Noise+F1F2+MFCC	75 ± 8.5	86 ± 13.5	64 ± 27.9	0.74
Prosody	72 ± 13.2	82 ± 17.5	62 ± 23.9	0.74
GMM-UBM	70 ± 11.5	68 ± 35.5	72 ± 31.6	0.70
Unvoiced	**96 ± 8.4**	**99 ± 0.0**	**92 ± 16.9**	**0.98**
		Caucho		
Noise+F1F2+MFCC	80 ± 16.3	86 ± 13.5	74 ± 25	0.83
Prosody	73 ± 8.2	78 ± 14.8	68 ± 19.3	0.75
GMM-UBM	79 ± 11.9	88 ± 16.8	70 ± 25.4	0.80
Unvoiced	**96 ± 5.1**	**92 ± 10.3**	**99 ± 0.0**	**0.95**
		Presa		
Noise+F1F2+MFCC	81 ± 8.8	80 ± 13.3	82 ± 19.9	0.81
Prosody	73 ± 12.5	78 ± 22	68 ± 25.3	0.72
GMM-UBM	75 ± 9.7	82 ± 11.4	68 ± 21.5	0.72
Unvoiced	**95 ± 9.7**	**92 ± 13.9**	**98 ± 6.3**	**0.94**
		Apto		
Noise+F1F2+MFCC	78 ± 13.2	80 ± 16.3	76 ± 18.4	0.78
Prosody	77 ± 14.2	80 ± 16.3	74 ± 16.5	0.77
GMM-UBM	77 ± 11.6	80 ± 16.3	74 ± 21.2	0.73
Unvoiced	**95 ± 7.1**	**98 ± 6.3**	**92 ± 13.9**	**0.95**
		Flecha		
Noise+F1F2+MFCC	76 ± 11.7	76 ± 26.3	76 ± 27.9	0.76
Prosody	78 ± 10.3	78 ± 17.5	78 ± 19.9	0.78
GMM-UBM	81 ± 9.9	86 ± 18.9	76 ± 15.8	0.78
Unvoiced	**94 ± 6.9**	**98 ± 6.3**	**90 ± 14.1**	**0.93**
		Trato		
Noise+F1F2+MFCC	77 ± 6.8	90 ± 14.1	64 ± 22.7	0.83
Prosody	78 ± 10.3	76 ± 18.4	80 ± 13.3	0.79
GMM-UBM	72 ± 12.3	76 ± 24.6	68 ± 25.3	0.75
Unvoiced	**94 ± 6.9**	**99 ± 0.0**	**88 ± 13.9**	**0.95**

TABLE VI. (*Continued*.)

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
		Coco		
Noise+F1F2+MFCC	76 ± 11.7	74 ± 31	78 ± 14.8	0.69
Prosody	83 ± 11.6	82 ± 17.5	84 ± 18.4	0.80
GMM-UBM	78 ± 12.3	90 ± 10.5	66 ± 21.2	0.81
Unvoiced	**93 ± 8.2**	**98 ± 6.3**	**88 ± 16.9**	**0.94**
		Plato		
Noise+F1F2+MFCC	69 ± 5.7	74 ± 18.9	64 ± 22.7	0.64
Prosody	72 ± 10.3	78 ± 17.5	66 ± 23.2	0.76
GMM-UBM	75 ± 9.7	88 ± 10.3	62 ± 19.9	0.75
Unvoiced	**88 ± 13.2**	**92 ± 16.9**	**84 ± 18.3**	**0.92**
		Pato		
Noise+F1F2+MFCC	76 ± 8.4	86 ± 13.5	66 ± 16.5	0.75
Prosody	84 ± 8.4	86 ± 13.5	82 ± 17.5	0.82
GMM-UBM	77 ± 10.6	92 ± 10.3	62 ± 23.9	0.79
Unvoiced	**84 ± 8.4**	**90 ± 10.5**	**78 ± 14.8**	**0.83**

TABLE VII. Results obtained from the isolated words of the German data.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUROC
		Bahnhofsvorsteher		
Noise + F1F2 + MFCC	72.2 ± 11.3	66.9 ± 14.9	77.6 ± 17.3	0.72
Prosody	86.9 ± 8.6	82.8 ± 14.6	90.6 ± 12.5	0.91
GMM-UBM	71.7 ± 7.7	65.0 ± 18.1	78.2 ± 9.0	0.74
Unvoiced	**96.6 ± 2.9**	**95.6 ± 5.7**	**97.6 ± 4.9**	**0.97**
		Rettungsschwimmer		
Noise+F1F2+MFCC	68.7 ± 8.8	58.9 ± 24.5	78.8 ± 18.4	0.66
Prosody	76.1 ± 10.9	77.5 ± 12.9	74.9 ± 18.2	0.75
GMM-UBM	68.2 ± 8.3	62.6 ± 19.8	73.6 ± 11.3	0.71
Unvoiced	**95.9 ± 4.8**	**93.1 ± 8.4**	**98.8 ± 3.9**	**0.96**
		Toilettenpapier		
Noise+F1F2+MFCC	70.9 ± 9.6	69.2 ± 15.2	72.8 ± 12.9	0.70
Prosody	73.1 ± 9.5	69.1 ± 9.6	77.1 ± 13.9	0.75
GMM-UBM	74.9 ± 9.5	70.6 ± 18.1	79.6 ± 23	0.72
Unvoiced	**94.8 ± 5.1**	**97.6 ± 4.9**	**91.9 ± 7.7**	**0.95**
		Bundesgerichtshof		
Noise+F1F2+MFCC	61.9 ± 8.3	33.1 ± 16.5	90.9 ± 10.3	0.65
Prosody	75.6 ± 13.6	61.4 ± 23	90.0 ± 11.1	0.78
GMM-UBM	74.9 ± 13.2	69.4 ± 20.9	80.8 ± 20.4	0.74
Unvoiced	**93.7 ± 1.9**	**94.4 ± 5.9**	**93.2 ± 7.9**	**0.95**
		Bedienungsanleitung		
Noise+F1F2+MFCC	67.5 ± 8.6	65 ± 28.2	69.3 ± 21.9	0.61
Prosody	88.2 ± 7.1	85.3 ± 10.5	90.9 ± 7.1	0.90
GMM-UBM	73.2 ± 10.4	54.4 ± 21.1	91.9 ± 5.6	0.74
Unvoiced	**89.8 ± 5.1**	**80.1 ± 11.8**	**98.8 ± 3.9**	**0.92**
		Perlenkettenschachtel		
Noise+F1F2+MFCC	73.3 ± 11.6	66.5 ± 21.3	79.7 ± 14.7	0.70
Prosody	79.6 ± 10.7	76.4 ± 14.3	83.1 ± 16.9	0.77
GMM-UBM	71.1 ± 8.9	57.2 ± 22.8	85.3 ± 14.1	0.73
Unvoiced	**84.1 ± 8.7**	**89.9 ± 8.2**	**78.2 ± 14.8**	**0.85**

TABLE VIII. Results obtained from the isolated words of the Czech data.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
		pepa		
Noise+F1F2+MFCC	86.6 ± 22.7	73.1 ± 45.4	100 ± 0	0.90
Prosody	76.3 ± 25.7	62.5 ± 49.8	90.0 ± 30.2	0.81
GMM-UBM	82.1 ± 23.9	69.4 ± 46.7	86.1 ± 35.1	0.90
Unvoiced	**91.6 ± 18.9**	**84.4 ± 36.6**	**98.8 ± 35**	**0.92**
		fouká		
Noise+F1F2+MFCC	70.6 ± 24.8	61.9 ± 49.9	79.4 ± 40.5	0.67
Prosody	83.8 ± 23.8	92.5 ± 26.8	75 ± 43.8	0.86
GMM-UBM	80.6 ± 24.4	80.6 ± 40.1	66.7 ± 47.8	0.88
Unvoiced	**95.9 ± 13.9**	**99 ± 0.0**	**91.9 ± 27.8**	**0.96**
		sada		
Noise + F1F2+MFCC	85.3 ± 23.1	98.1 ± 7.5	72.5 ± 45.3	0.86
Prosody	84.4 ± 23.6	96.9 ± 12.5	71.9 ± 45.9	0.79
GMM-UBM	85.4 ± 22.7	80.6 ± 40.1	77.8 ± 42.1	0.82
Unvoiced	**94.7 ± 15.7**	**89.4 ± 31.4**	**99 ± 0.0**	**0.94**
		tiká		
Noise + F1F2 + MFCC	78.8 ± 25.4	83.8 ± 37.7	73.8 ± 44.5	0.71
Prosody	88.8 ± 21.1	93.1 ± 25.9	84.4 ± 36.2	0.87
GMM-UBM	89.2 ± 20.6	80.6 ± 40.1	94.4 ± 23.2	0.86
Unvoiced	**90 ± 20.4**	**98.8 ± 5**	**81.3 ± 39.9**	**0.90**
		kuká		
Noise + F1F2 + MFCC	77.2 ± 25.4	76.9 ± 43.4	77.5 ± 41.8	0.68
Prosody module	85.6 ± 23.2	80.6 ± 40.4	90.6 ± 29.3	0.79
GMM-UBM	74.1 ± 25	72.2 ± 45.5	83.3 ± 37.8	0.78
Unvoiced	**96.6 ± 12.9**	**93.8 ± 25**	**99.4 ± 2.5**	**0.96**
		chata		
Noise + F1F2 + MFCC	67.2 ± 24.2	80 ± 39.9	54.4 ± 50	0.73
Prosody module	80 ± 24.9	79.4 ± 40.4	80.6 ± 39	0.82
GMM-UBM	86.4 ± 22.3	75 ± 43.9	97.2 ± 16.7	0.86
Unvoiced	**84.7 ± 23.3**	**86.9 ± 34.2**	**82.5 ± 37.4**	**0.84**
		tči		
Noise + F1F2 + MFCC	78.7 ± 25.3	72 ± 45.9	85.3 ± 35.6	0.74
Prosody module	80 ± 25.2	74.7 ± 44.6	85.3 ± 35.1	0.83
GMM-UBM	89.4 ± 20.5	91.7 ± 28	97.22 ± 16.7	0.94
Unvoiced	**90.3 ± 20.3**	**92.7 ± 26.8**	**88 ± 33**	**0.89**
		vzhůru		
Noise + F1F2 + MFCC	68.4 ± 24.8	57.5 ± 50.1	79.4 ± 40.9	0.68
Prosody	81.9 ± 24.6	83.1 ± 38.5	80.6 ± 39.9	0.84
GMM-UBM	78.7 ± 24.7	66.7 ± 47.8	69.4 ± 46.7	0.72
Unvoiced	**89.1 ± 21.2**	**78.8 ± 42.1**	**99.4 ± 2.5**	**0.88**
		sdružit		
Noise + F1F2 + MFCC	79 ± 25.4	80 ± 41.4	78 ± 42.4	0.79
Prosody module	72 ± 25.3	87.3 ± 34.3	56.7 ± 50.6	0.71
GMM-UBM	83.7 ± 23.5	77.8 ± 42.1	94.4 ± 23.2	0.84
Unvoiced	**89.7 ± 20.7**	**94.7 ± 20.7**	**84.7 ± 36.2**	**0.90**
		funkční		
Noise + F1F2+MFCC	80.6 ± 24.9	90.6 ± 28.9	70.6 ± 46.8	0.86
Prosody module	80.9 ± 24.8	93.8 ± 25	68.1 ± 47.5	0.91
GMM-UBM	86.9 ± 22	97.2 ± 16.7	77.8 ± 42.1	0.88
Unvoiced	**96.3 ± 13.4**	**93.8 ± 25**	**98.8 ± 25**	**0.95**

TABLE VIII. (*Continued.*)

	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
	cukrářství			
Noise+F1F2+MFCC	80.3 ± 24.9	73.1 ± 45.5	87.5 ± 30.6	0.88
Prosody	76.9 ± 25.6	78.8 ± 41.6	75 ± 44.5	0.80
GMM-UBM	83.3 ± 23.6	77.8 ± 42.1	91.7 ± 28	0.84
Unvoiced	**86.9 ± 22.5**	**86.3 ± 35.1**	**87.5 ± 33.6**	**0.86**
	vstříc			
Noise + F1F2 + MFCC	80.9 ± 24.5	81.3 ± 38.9	80.6 ± 37.9	0.84
Prosody	79.1 ± 25.3	75 ± 44.7	83.1 ± 38.1	0.87
GMM-UBM	87.8 ± 21.5	77.8 ± 42.2	88.9 ± 31.9	0.85
Unvoiced	**86.6 ± 22.7**	**81.3 ± 40.3**	**91.9 ± 25.9**	**0.87**

German:

(1) *Peter und Paul essen gerne Pudding*.
(2) *Das Fest war sehr gut vorbereitet*.
(3) *Seit seiner Hochzeit hat er sich sehr verändert*.
(4) *Im Inhaltsverzeichnis stand nichts über Lindenblütentee*.
(5) *Der Kerzenständer fiel gemeinsam mit der Blumenvase auf den Plattenspieler*.

Czech:

(1) *Kolik máte ted u sebe asi peněz*?
(2) *Kolikpak máte tedka u sebe asi peněz*?
(3) *Kolikpak máte tedka u sebe asi tak peněz*?

3. Words

Spanish: *atleta, campana, gato, petaka, braso, caucho, presa, apto, flecha, trato, coco, plato, pato*.

German: *Bahnhofsvorsteher, Rettungsschwimmer, Toilettenpapier, Bundesgerichtshof, Bedienungsanleitung, Perlenkettenschachtel*.

Czech: *pepa, fouká, sada, tiká, kuká, chata, tči, vzhůru, sdružit, funkční, cukrářství, vstříc*.

4. Tables with results

Results in read texts: Tables III, IV, and V in this appendix include results obtained in sentences spoken in Spanish, German, and Czech, respectively. Note that in Spanish and German recordings the best results are obtained

TABLE IX. Details of the results obtained in cross-language experiments. Accuracy (Acc.), sensitivity (Sens.), specificity (Spec.), and area under the ROC curve (AUC).

Train / Test	Spanish				German				Czech			
	Acc.(%)	Sens.(%)	Spec.(%)	AUC	Acc.(%)	Sens.(%)	Spec.(%)	AUC	Acc.(%)	Sens.(%)	Spec.(%)	AUC
Spanish + 0% train	—	—	—	—	57	82	32	56	55	80	30	55
German + 0% train	60	42	77	58	—	—	—	—	58	57	60	59
Czech + 0% train	69	47	85	68	77	53	95	78	—	—	—	—
Spanish + 10% train	—	—	—	—	67	68	66	68	71	74	68	71
German + 10% train	85	84	85	85	—	—	—	—	75	79	71	75
Czech + 10% train	76	59	89	77	78	52	96	81	—	—	—	—
Spanish + 20% train	—	—	—	—	74	77	70	74	75	86	65	77
German + 20% train	89	89	90	89	—	—	—	—	82	84	81	82
Czech + 20% train	74	52	90	74	83	86	80	82	—	—	—	—
Spanish + 30% train	—	—	—	—	79	78	80	80	81	77	85	80
German + 30% train	91	88	94	91	—	—	—	—	86	89	83	86
Czech + 30% train	78	64	89	76	79	61	91	81	—	—	—	—
Spanish + 40% train	—	—	—	—	88	85	90	88	86	88	84	84
German + 40% train	93	92	94	94	—	—	—	—	88	91	85	88
Czech + 40% train	79	64	90	80	78	58	93	81	—	—	—	—
Spanish + 50% train	—	—	—	—	90	90	90	90	88	90	86	88
German + 50% train	94	93	95	94	—	—	—	—	88	90	86	89
Czech + 50% train	83	73	90	81	81	60	95	83	—	—	—	—
Spanish + 60% train	—	—	—	—	92	94	91	93	91	92	91	92
German + 60% train	95	93	97	95	—	—	—	—	93	94	91	94
Czech + 60% train	86	77	94	87	79	58	94	81	—	—	—	—
Spanish + 70% train	—	—	—	—	93	91	94	94	91	89	93	91
German + 70% train	95	94	96	95	—	—	—	—	91	91	92	92
Czech + 70% train	83	73	90	86	89	78	97	89	—	—	—	—
Spanish + 80% train	—	—	—	—	95	94	96	95	92	92	91	91
German + 80% train	96	95	97	97	—	—	—	—	94	94	95	95
Czech + 80% train	84	67	98	86	80	60	95	84	—	—	—	—

with the unvoiced characterization approach proposed in this paper. In two of the three Czech sentences the results obtained with the prosody module were higher.

It can be observed from Table III that in general, results obtained in the Spanish sentences modeled with the other three approaches (Noise+F1&F2+MFCC, prosody, and GMM-UBM) are below 0.90 of AUC, with values mostly around 0.80. For the German sentences, the results with these approaches are below 0.80, except for two cases, sentences 3 and 5, where the prosody module reached AUC of 0.84 and 0.81, respectively. In the Czech sentences the results are slightly different. In the sentence 1 the obtained AUC value is 0.95 using the unvoiced features, while the GMM-UBM approach gives 0.90, and the other two approaches are below 0.80. In the sentences 2 and 3 the highest AUC values are obtained with the prosody module (0.89 and 0.93, respectively). The other methods exhibited AUC values below 0.90, except for the "Noise+F1&F2+MFCC" approach, which reaches 0.91 in the third sentence (the highest result obtained with this approach).

Results in isolated words: A total of 31 isolated words are evaluated on this work. Tables VI, VII, and VIII include results obtained with 13 words uttered in Spanish, 6 in German, and 12 in Czech, respectively.

Results in cross-language experiments: Table IX includes results obtained in the cross-language experiments. A portion of the target language is included in the training set incrementally, beginning with 0% up to 80% in steps of 10%. Note that the accuracy of the system improves very quickly, indicating that it has a good generalization capability that can be used to deploy computer-aided tools for the automatic assessment of speech of people with Parkinson's disease in different languages.

[1]O. Hornykiewicz, "Biochemical aspects of Parkinson's disease," Neurology **51**(2), S2–S9 (1998).
[2]M. C. de Rijk, "Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts," Neurology **54**, 21–23 (2000).
[3]J. A. Logemann, H. B. Fisher, B. Boshes, and E. R. Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients," J. Speech Hear. Disord. **43**, 47–57 (1978).
[4]Movement Disorder Society, "State of the art review the Unified Parkinson's Disease Rating Scale (UPDRS): Status and recommendations," Movement Disord. **18**(7), 738–750 (2003).
[5]C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, N. Giladi, R. G. Holloway, C. G. Moore, G. K. Wenning, M. D. Yahr, and L. Seidl, "Movement disorder society task force report on the Hoehn and Yahr staging scale: Status and recommendations," Movement Disord. **19**(9), 1020–1028 (2004).
[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," Behav. Neurol. **11**, 131–137 (1999).
[7]F. L. Darley, A. E. Aronson, and J. R. Brown, "Differential diagnostic patterns of dysarthria," J. Speech Hear. Res. **12**(2), 246–269 (1969).
[8]J. R. Green, D. R. Beukelman, and L. J. Ball, "Algorithmic estimation of pauses in extened speech samples of dysarthric and typical speech," J. Med. Speech-Lang. Pathol. **12**(4), 149–154 (2004).
[9]Y. T. Wang, R. D. Kent, J. R. Duffy, and J. E. Thomas, "Analysis of diadochokinesis in ataxic dysarthria using the motor speech profile program," Folia Phoniatr. Logop. **61**, 1–11 (2009).
[10]M. O. Paja and T. Falk, "Automated dysarthria severity classification for improved objective intelligibility assessment of spastic dysarthric speech," in *Proceedings of the 13th Annual Conference of the International Speech Communication Association (INTERSPEECH)* (2012), pp. 62–65.
[11]K. Heejin, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T. Huang, K. Watkin, and S. Frame, "Dysarthric speech database for universal access research," in *Proceedings of the 9th Annual Conference of the International Speech Communication Association (INTERSPEECH)* (2008), pp. 4–7.
[12]C. M. Tanner, M. Brandabur, and E. R. Dorsey, "Parkinson disease: A global view," Parkinson Rep. **Spring**, 9–11 (2008).
[13]P. F. Worth, "How to treat Parkinson's disease in 2013," Clin. Med. **13**(1), 93–96 (2013).
[14]L. O. Ramig, C. Fox, and S. Sapir, "Speech treatment for Parkinson's disease," Expert Rev. Neurother. **8**(2), 297–309 (2008).
[15]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 **25**(4), 199–205 (2011).
[16]S. Skodda, W. Visser, and U. Schlegel, "Vowel articulation in Parkinson's diease," J. Voice **25**(4), 467–472 (2011).
[17]A. Tsanas, M. A. Little, C. Fox, and L. O. Ramig, "Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease," IEEE Trans. Neural Syst. Rehab. Eng. **22**(1), 181–190 (2014).
[18]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. Acoust. Soc. Am. **129**(1), 350–367 (2011).
[19]A. M. Goberman, "Correlation between acoustic speech characteristics and non-speech motor performance in Parkinson disease," Med. Sci. Monitor **11**(3), CR109–CR116 (2005).
[20]K. S. Perez, L. O. Ramig, M. E. Smith, and C. Dromery, "The Parkinson larynx: Tremor and videostroboscopic findings," J. Voice **10**(4), 354–361 (1996).
[21]J. Möbes, G. Joppich, F. Stiebritz, R. Dengler, and C. Schröder, "Emotional speech in Parkinson's disease," Movement Disord. **23**(6), 824–829 (2008).
[22]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. Bio-Medical Eng. **56**(4), 1015–1022 (2009).
[23]P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," Glot Int. **5**(9/10), 341–345 (2001).
[24]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 **6**, 23 (2007).
[25]J. R. Orozco-Arroyave, J. F. Vargas-Bonilla, J. D. Arias-Londoño, S. Murillo-Rendón, G. Castellanos-Domínguez, and J. F. Garcés, "Nonlinear dynamics for hypernasality detection in spanish vowels and words," Cogn. Comput. **5**(4), 448–457 (2013).
[26]S. Sapir, L. O. Raming, J. L. Spielman, and C. Fox, "Formant centralization ratio (FCR): A proposal for a new acoustic measure of dysarthric speech," J. Speech Lang. Hear. Res. **53**(1), 114–125 (2010).
[27]S. Sapir, J. L. Spielman, L. O. Ramig, B. Story, and C. Fox, "Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson's disease: Acoustic and perceptual findings," J. Speech Lang. Hear. Res. **50**, 899–912 (2007).
[28]A. Tsanas, M. A. Little, P. E. Mcsharry, J. L. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," IEEE Trans. Bio-Med. Eng. **59**(5), 1264–1271 (2012).
[29]A. Tsanas, M. Little, P. E. McSharry, and L. O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity," J. R. Soc. Interface **8**(59), 842–855 (2010).
[30]N. Sáenz-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. Sign. Process. Control **1**, 120–128 (2006).
[31]T. Bocklet, S. Steidl, E. Nöth, and S. Skodda, "Automatic evaluation of Parkinson's speech—Acoustic, prosodic and voice related cues," in *Proceedings of the 14th Annual Conference of the International Speech Communication Association (INTER-SPEECH)* (2013), pp. 1149–1153.
[32]F. Eyben, M. Wöllmer, and B. Shuller, "Opensmile—The Munich versatile and fast open-source audio feature extractor," in *Proceedings of the ACM Multimedia* (2010), pp. 1459–1462.
[33]V. Zeiβler, J. Adelhardt, A. Batliner, C. Frank, E. Nöth, R. Shi, and H. Niemann, "The prosody module," in *SmartKom: Foundations of Multimodal Dialogue Systems*, Cognitive Technologies Series (Springer, Berlin, 2006), pp. 139–152.

[34]K. N. Stevens, *Acoustic Phonetics* (MIT Press, Cambridge, 1998).
[35]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," Lect. Notes Comput. Sci. **7911**, 112–119 (2013).
[36]A. Bayestehtashk, M. Asgari, I. Shafran, and J. Mcnames, "Fully automated assessment of the severity of Parkinson's disease from speech," Comput. Speech Lang. **29**(1), 172–185 (2015).
[37]K. Chenausky, J. Macauslan, and R. Goldhor, "Acoustic analysis of PD speech," Parkinson's Dis. **2011**, 435232 (2011).
[38]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. Am. **134**(3), 2171–2181 (2013).
[39]M. Novotný, J. Rusz, R. Cmejla, and E. Ruzicka, "Automatic evaluation of articulatory disorders in Parkinsons disease," IEEE/ACM Trans. Audio Speech Lang. Process. **22**(9), 1366–1378 (2014).
[40]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 *Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH)* (2014), pp. 1473–1577.
[41]E. Zwicker and E. Terhardt, "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. **68**(5), 1523–1525 (1980).
[42]A. M. Goberman and M. Blomgren, "Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease," J. Voice **22**(2), 178–191 (2008).
[43]K. N. Stevens, "Toward a model for lexical access based on acoustic landmarks and distinctive features—Consonants modeling," J. Acoust. Soc. Am. **111**(4), 1872–1891 (2002).
[44]S. Skodda and U. Schlegel, "Speech rate and rhythm in Parkinson's disease," Movement Disord. **23**(7), 985–992 (2008).
[45]J. R. Duffy, "Motor speech disorders: Clues to neurologic diagnosis," in *Parkinson's Disease and Movement Disorders: Diagnosis and Treatment Guidelines for the Practicing Physician* (Humana Press, Clifton, NJ, 2000), pp. 35–53.
[46]J. A. Logemann and H. B. Fisher, "Vocal tract control in Parkinson's disease: Phonetic feature analysis of misarticulations," J. Speech Hear. Disord. **46**(4), 348–452 (1981).
[47]J. I. 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. Bio-Med. Eng. **53**(10), 1943–1953 (2006).
[48]A. Maier, 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. **51**(5), 425–437 (2009).
[49]E. M. Critchley, "Speech disorders of parkinsonism: A review," J. Neurol. Neurosurg. Psychiatry **44**(9), 751–758 (1981).
[50]J. Rusz, R. Cmejla, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka, "Acoustic assessment of voice and speech disorders in Parkinson's disease through quick vocal test," Movement Disord. **26**(10), 1951–1952 (2011).
[51]H. Ackermann, I. Hertrich, and T. Hehr, "Oral diadochokinesis in neurological dysarthrias," Folia Phoniatr. Logopaed. **47**, 15–23 (1995).
[52]J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. González-Rátiva, and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proceedings of the 9th Language Resources and Evaluation Conference (LREC)* (2014), pp. 342–347.
[53]S. Skodda, W. Grönheit, N. Mancinelli, and U. Schlegel, "Progression of voice and speech impairment in the course of Parkinson's disease: A longitudinal study," Parkinson's Dis. **2013**, 1–8 (2013).
[54]H. Ackermann and W. Ziegler, "Articulatory deficits in Parkinsonian dysarthria: An acoustic analysis," J. Neurol. Neurosurg. Psychiatry **54**(12), 1093–1098 (1991).