Automatic Detection of Parkinson's Disease from Continuous Speech Recorded in Non-Controlled Noise Conditions

J. C. Vásquez-Correa[1], *T. Arias-Vergara*[1], *J. R. Orozco-Arroyave*[1,2], *J. F. Vargas-Bonilla*[1], *J. D. Arias-Londoño*[1], *and E. Nöth*[2]

[1] Faculty of Engineering. Universidad de Antioquia UdeA, Medellín, Colombia.
[2]Pattern Recognition Lab, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany

Corresponding author: rafael.orozco@i5.informatik.uni-erlangen.de

Abstract

Automatic classification of Parkinson's disease (PD) speakers and healthy controls (HC) is performed considering speech recordings collected in non-controlled noise conditions. The speech tasks include six sentences and a read text. The recording is performed using an open source portable device and a commercial microphone. A speech enhancement (SE) technique is applied to improve the quality of the signals. Voiced and unvoiced frames are segmented from the speech tasks and characterized separately. The discrimination of speakers with PD and HC is performed using a support vector machine with soft margin. The results indicate that it is possible to discriminate between PD and HC speakers using recordings collected in non-controlled noise conditions. The accuracies obtained with the voiced features range from 64% to 86%. For unvoiced features the accuracies range from 78% to 99%. The SE algorithm improves the accuracies of the unvoiced frames in up to 11 percentage points, while the accuracies decrease in the voiced frames when the SE algorithm is applied. This work is a step forward to the development of portable devices to assess the speech of people with PD.

Index Terms: Parkinson's disease, speech, voiced/unvoiced, speech enhancement, non-controlled noise conditions.

1. Introduction

PD is a neurological disorder characterized by the progressive loss of dopaminergic neurons in the substantia nigra of the midbrain [1]. Voice impairments appear in about 90% of people with PD. The set of symptoms observed in the speech of PD patients include reduced loudness, monotonous pitch, and misarticulation, among others. Although the high prevalence of PD and its negative effects in speech, only from 3% to 4% of the patients receive speech therapy [2]. The motor skills of the patients with PD are impaired, thus visiting a hospital to perform medical screenings and/or assessments is not a straightforward task for them [3]. The research community has shown interest developing computer-aided tools to perform screenings from speech. The main goals of such tools are (a) to spare patients moving from their home to the hospital to perform routine screenings, and (b) to raise early alerts to patients and doctors that allow them make assertive and timely decisions regarding pharmacological treatments and/or therapies. Several studies have been focused on such computational approach for telemonitoring of patients with PD through speech. In [4] the authors developed a portable device to assess speech of people with PD. The device evaluates the speech signal and provides bio-feedback to the patients in real time. It is equipped with an LCD monitor and generates a report with different measurements to assess the phonatory and articulatory capabilities of the patient. Although the device is able to calculate different acoustic features on the vowels /a/, /i/, and /u/, more speech tasks need to be included to perform more complete screenings. In [5] the authors present a portable device to assess the speech therapy of patients with PD. The device provides bio-feedback of the speech volume levels. A signal tone is sent to the patient if the vocal intensity falls below an adjustable threshold. This device operates with a microphone stuck to the neck of the patient, so it could be considered invasive in some way by several users.This device was updated later in [6], where the authors included visual feedback. However, the new version still requires the microphone stuck to the neck. In [7] a device equipped with an accelerometer to measure the skin vibrations and a microphone to record the speech signals is presented. The device calculates three voice parameters: fundamental frequency, energy, and sound pressure level. The feedback is given via vibrations from the device to the patients. When the device vibrates, the patient is aware that is speaking incorrectly. In [8] a portable device to record speech signals from a patient under monitoring is presented. The signals are acquired using a microphone stuck to the neck of the patient. The aim of this device is to identify different voice disorders through the fundamental frequency and sound pressure level estimated from sustained phonations of the vowel /a/ and read texts. In [9] a methodology for the automatic detection of PD using speech recordings captured in controlled noise conditions, i.e. in a sound-proof booth and with a professional audio card and microphone, is presented. The authors report an accuracies of up to 99% considering recordings of several isolated words. Another study reporting accuracies above 98% is presented in [10]. A total of 263 voice samples were recorded from 43 subjects. The experiments were performed following a 10-folds cross-validation strategy, but the speaker independence was not fulfilled, leading to optimistic and probably biased results.

According to the reviewed literature, the devices are mainly focused on the analysis of sustained phonations to estimate fundamental frequency and its variability, the sound pressure level, and the frequency of the formants. Further work is required to develop portable devices for the analysis of speech of people with PD considering running speech signals recorded in non-controlled noise conditions. The aim of this work is to apply the methodology presented in [9] using speech signals recorded in real-world conditions, i.e. without using the sound-proof booth and using the device presented in [4] to record the signals. The methodology is tested on a set with six isolated sentences and on a read text. The device will be accessible to the community

through a web page, allowing other people to do similar speech assessments and improve the current version of the system.

The rest of the paper is organized as follows: Section 2 includes details of the methodology presented in the paper. Section 3 includes the description of the data and the experiments. Section 4 shows the results obtained in the different experiments, and finally in Section 5, the conclusions derived from this work are provided.

2. Methods

2.1. Preprocessing

The recordings considered in this work were captured in non-controlled noise conditions, thus the speech enhancement (SE) procedure presented in [11] is applied. The noisy signal is represented by two components: the clean signal, i.e the target, and the noise. The method consists of finding a linear minimum mean-square error estimator of the target signal $\widehat{s} \in \Re^{k\times 1}$, which is defined as $\widehat{s} = H\mathbf{x}$, where $\mathbf{x} \in \Re^{k\times 1}$ is a vector formed by k samples of the noisy signal and $H \in \Re^{k\times k}$ is a linear estimation matrix that is optimized to "model" the noise in $\mathbf{x}$. The optimum H is obtained from a matrix that diagonalizes the covariance matrices of the target signal and the noise simultaneously [12]. This method was chosen due to its good performance in previous studies [11]. For the sake of comparisons, the recordings without SE are also considered here. Possible bias introduced by the channel are eliminated by means of mean cepstral subtraction [13].

2.2. Voiced/Unvoiced characterization

Voiced and unvoiced frames are detected and grouped separately. Hamming windows with 20 ms length and time shift of 10 ms are applied on all of the segmented frames. The features estimated for voiced frames include the variation of the fundamental frequency of speech (F_0), jitter, shimmer, *log* energy per window, and 12 MFCCs. For unvoiced frames, the set of features includes 12 MFCCs, and the *log* energy of the signal distributed in 25 Bark bands, as in [9]. The features from voiced and unvoiced frames are grouped separately into feature vectors. Four functionals of these vectors are calculated: mean value, standard deviation, kurtosis, and skewness. Forming 55- and 148-dimensional feature vectors for the voiced and unvoiced frames, respectively.

2.3. Classification

The decision whether a recording is from a PD patient or from a healthy person is taken by means of a radial basis support vector machine (SVM) with margin parameter C and a Gaussian kernel with parameter γ. C and γ are optimized through a grid-search up to powers of ten with $10^{-1} < C < 10^4$ and $1 < \gamma < 10^3$. The selection criteria was based on the obtained accuracy on test data. This approach can lead to slightly optimistic accuracy estimates, but considering that only two parameters are optimized, the bias effect should be minimal. Due to the low number of speakers, the SVM is tested following a leave-one-speaker-out cross-validation (LOSO-CV) strategy. An SVM is used here due to its validated success in similar works related to automatic detection of pathological speech signals [9, 14, 15].

The stages of the methodology are detailed in Figure 1.

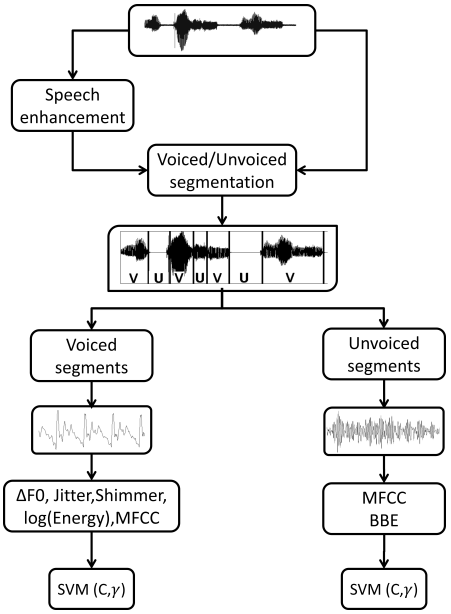

Figure 1: *Methodology*

3. Experimental setup

3.1. Data

14 patients with PD and 14 HC (7 female and 7 male in each group) were recorded with a sampling frequency of 44.1 KHz and 16 bits of resolution. The age of the PD patients ranges from 51 to 71 (mean 61.64 ± 6.43), and the age of the HC ranges from 50 to 78 (mean 63.29 ± 10.43). All patients were diagnosed by a neurologist expert. The speech signals were recorded in a normal room, under non-controlled noise conditions. This database was collected by the GITA research group, from Universidad de Antioquia, in Medellín, Colombia. Table 1 provides detailed information of the recorded patients including age, gender, time after the PD diagnosis, and their neurological state according to the MDS-UPDRS-III scale, which is the motor sub-scale of the full MDS-UPDRS evaluation [16].

Table 1: *Detailed information of the speakers.* ***t****: time post PD diagnosis in years.* ***PD****: Parkinson's disease.* ***HC****: Healthy controls.* ***UPDRS****: Unified Parkinson's disease rating scale.*

PD				HC	
AGE	**GENDER**	**UPDRS**	**t**	**AGE**	**GENDER**
71	F	27	1	78	F
70	M	62	4	78	M
69	M	22	2	78	M
67	M	43	14	75	M
66	M	22	5	70	F
66	M	30	8	62	M
61	F	33	16	61	M
60	M	15	9	60	M
58	F	44	38	59	F
57	M	79	5	55	M
56	F	18	14	54	F
56	F	30	44	54	F
55	F	30	8	52	F
51	F	49	42	50	F

3.2. Speech tasks

A set with six sentences and a read text with 36 words are considered. These speech tasks comprise a subset of the tasks presented in [17]. The details are provided in Table 2.

Table 2: *Speech tasks (ST)*

ST	Texts
1	Mi casa tiene tres cuartos.
2	Omar, que vive cerca, trajo miel.
3	Laura sube al tren que pasa.
4	Los libros nuevos no caben en la mesa de la oficina.
5	Rosita Niño, que pinta bien, donó sus cuadros ayer.
6	Luisa Rey compra el colchón duro que tanto le gusta.
7	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.

3.3. Technical specifications of the device

The device is based on a board with a micro sized open development platform called *ODROID-U2*. It has an ARM Cortex-A9 quad core processor with 2 GB of RAM memory and its operation frequency is 1.7 GHz. The board is equipped with a micro SD card port to store the operating system (OS) and the data. The OS running on the board is *Ubuntu 12.10*. The algorithms used to record speech were written in Python. Additionally, the device has a 7" LCD monitor used to give visual feedback to the patient. A wireless keyboard is used to type the data of the user. The audio signal is captured using a h250 Logitech headset. The *ODROID-U2* includes an audio codec *MAX98090* which operates with up to 24 bits. Figure 2 illustrates the technical characteristics of the device.

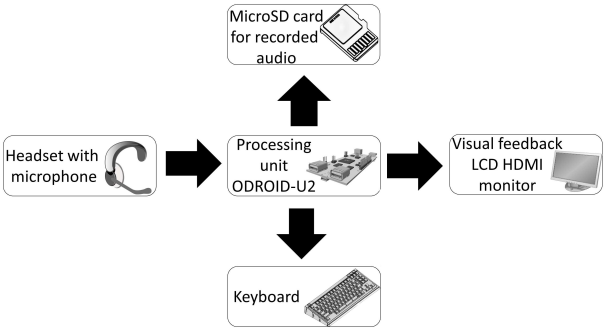

Figure 2: Technical characteristics of the device

4. Experiments and results

The experiments are divided into two parts. First, the signals are considered in their original version, i.e. without the SE procedure. Second, the SE algorithm is applied. The speech tasks are evaluated separately. The results are presented in terms of accuracy, specificity, and sensitivity. Accuracy is the general performance of the system, while specificity and sensitivity indicate the capability of the system to detect pathological and healthy speakers, respectively. The results can be reported more compactly using the receiver operating characteristic (ROC) curve. Typically the area under the ROC curve (AUC) is used as a measure of the general performance of the binary classification systems. These statistics are commonly used to evaluate the performance of medical systems [18]. Table 3 contains the results obtained using features calculated upon the voiced segments. The accuracies range from 71% to 86% when the signals are considered without the SE procedure. When the recordings are processed considering the SE procedure the accuracies range from 64% to 82%. Although the best results in voiced features are not obtained with the enhanced signals, note that the accuracies increased in three of the speech tasks after applying the SE algorithm. Note also that the highest accuracies with the voiced features are not obtained when the SE is applied. Even, there are several cases e.g., speech tasks 2, 4, 6, and 7, where the accuracy decreased after applying the SE algorithm. This result is similar to those obtained in [11], which motivate us to do a detailed analysis of such behavior in future experiments.

Table 3: *Results for features estimated from voiced segments.* ***ST****: Speech task numbered according to Table 2.* ***Sig****: Signal.* ***Acc (%)****: Accuracy.* ***Sens (%)****: Sensitivity.* ***Spec (%)****: Specificity.* ***AUC****: Area under the ROC curve. Orig: Signal without SE. SE: Signal with SE.*

ST	Sig	Acc	Sens	Spec	AUC
1	Orig	71 ± 26	92 ± 27	50 ± 52	0.78
	SE	82 ± 25	71 ± 47	93 ± 26	0.94
2	Orig	75 ± 26	79 ± 43	71 ± 47	0.78
	SE	64 ± 36	47 ± 57	57 ± 51	0.71
3	Orig	71 ± 26	64 ± 49	86 ± 36	0.77
	SE	79 ± 25	100 ± 0	57 ± 51	0.85
4	Orig	$\boxed{86 \pm 31}$	92 ± 27	79 ± 43	0.89
	SE	79 ± 25	79 ± 43	79 ± 43	0.80
5	Orig	79 ± 25	86 ± 36	71 ± 47	0.84
	SE	82 ± 25	79 ± 43	86 ± 36	0.80
6	Orig	$\boxed{86 \pm 23}$	71 ± 47	100 ± 0	0.86
	SE	75 ± 26	71 ± 47	79 ± 43	0.75
7	Orig	79 ± 25	100 ± 0	57 ± 51	0.84
	SE	71 ± 26	100 ± 0	43 ± 51	0.76

Table 4 contains the results obtained using the features calculated upon the unvoiced segments. The accuracies range from 78% to 99% when the original signals are considered. The results obtained when the SE algorithm is applied range from 91% to 99%. Note that in six of the seven speech tasks the accuracy increased when the SE procedure is applied. The only exception was the second speech task, where the accuracy decreased from 95% to 91%. Note also that, for the read texts, with the enhanced and with the original speech recordings the accuracies achieved 99%. This result can be likely explained because the read texts contain more variety of words, syllables, and accents.

The best results with both feature sets (voiced and unvoiced) in sentences were obtained with the sixth speech task. In order to show such results more compactly, Figures 3 and 4 include the ROC curves obtained with this sentence characterized with voiced and unvoiced features, respectively. For the sake of comparison, the curves obtained with and without SE are included in both figures. Note that the accuracy obtained with the voiced features decreases when the SE algorithm is applied. Conversely, the accuracy of unvoiced features increases when the SE algorithm is applied.

Table 4: *Results for features estimated from unvoiced segments.* ***ST****: Speech task numbered according to Table 2.* ***Sig****: Signal.* ***Acc (%)****: Accuracy.* ***Sens (%)****: Sensitivity.* ***Spec (%)****: Specificity.* ***AUC****: Area under the ROC curve. Orig: Signal without SE. SE: Signal with SE.*

ST	Sig	Acc	Sens	Spec	AUC
1	Orig	92 ± 19	96 ± 13	87 ± 34	0.96
	SE	93 ± 17	92 ± 25	95 ± 17	0.96
2	Orig	94 ± 15	91 ± 27	98 ± 8	0.96
	SE	91 ± 20	100 ± 0	81 ± 40	0.91
3	Orig	86 ± 23	100 ± 0	72 ± 47	0.87
	SE	97 ± 12	100 ± 0	94 ± 24	0.95
4	Orig	93 ± 18	100 ± 0	86 ± 36	0.97
	SE	94 ± 16	100 ± 0	89 ± 33	0.97
5	Orig	78 ± 25	91 ± 26	65 ± 49	0.83
	SE	90 ± 20	92 ± 23	87 ± 32	0.94
6	Orig	86 ± 23	100 ± 0	72 ± 47	0.86
	SE	97 ± 12	100 ± 0	94 ± 24	0.98
7	Orig	99 ± 3	100 ± 0	99 ± 5	0.99
	SE	$\boxed{99 \pm 1}$	100 ± 0	99 ± 1	0.99

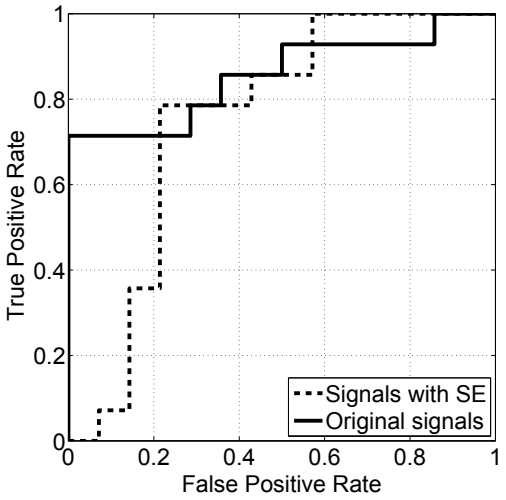

Figure 3: *ROC curves obtained from the voiced frames of the sixth speech task with and without the SE method*

5. Conclusions

A method to discriminate between PD speakers and HC considering speech samples recorded in non-controlled noise conditions is presented here. A total of six sentences and a read text uttered by 28 speakers (14 with PD and 14 HC) were evaluated. A SE algorithm is applied to improve the quality of the recordings. The method to discriminate speakers with PD and HC consists on the characterization of voiced and unvoiced frames separately. The highest accuracies obtained with the voiced frames range from 64% to 86%, while the results with unvoiced frames range from 78% to 99%. The SE algorithm improves the accuracy obtained with the unvoiced frames in up to 11 percentage points. The opposite effect is observed in the results with the voiced frames, where the accuracy decreases in about 11 percentage points when the SE algorithm is applied in some of the speech tasks. This result motivates us to continue studying the influence of different SE algorithms in different environments and speech tasks, in order to state which technique is more convenient according to each task and characterization method. The incorporation of prosody features into the framework of this device, such that allow the evaluation of speech characteristics related to timing, duration, and speech rate, is also planned for the near future.

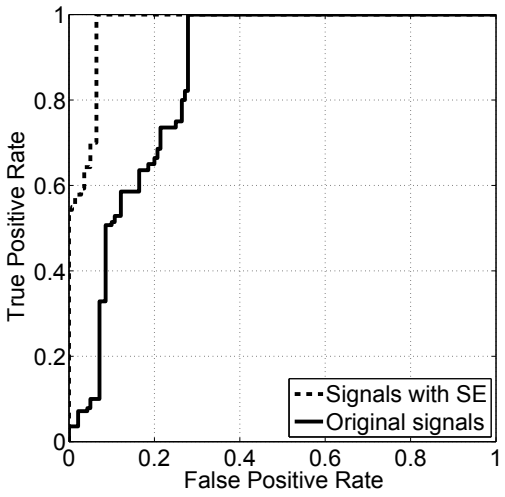

Figure 4: *ROC curves obtained from the unvoiced frames of the sixth speech task with and without the SE method*

Finally, as the recordings were collected using an open source platform and a commercial headset, this work is a step forward to the development of portable devices to assess the speech of people with PD. The platform presented here will be publicly accessible in the near future, allowing the community to improve the system functionalities. The use of this device to follow the speech therapy of PD patients and to assess their neurological state is also expected in the future.

6. Acknowledgement

Tomás Arias-Vergara is granted by the program of young researchers and innovators 2015, financed by COLCIENCIAS. Juan Rafael Orozco-Arroyave is under grants of Convocatoria 528 para estudios de doctorado en Colombia 2011 financed by COLCIENCIAS. 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 work is partially funded also by COLCIENCIAS through the project Nº 111556933858.

7. References

[1] O. Hornykiewicz, "Biochemical aspects of Parkinson's disease," *Neurology*, vol. 51, no. 2 Suppl 2, pp. S2–S9, 1998.

[2] M. Trail, C. Fox, L. O. Ramig, S. Sapir, J. Howard, and E. C. Lai, "Speech treatment for Parkinson's disease," *NeuroRehabilitation*, vol. 20, no. 3, pp. 205–221, 2005.

[3] D. G. Theodoros, G. Constantinescu, T. G. Russell, E. C. Ward, S. J. Wilson, and R. Wootton, "Treating the speech disorder in Parkinson's disease online," *Journal of Telemedicine and Telecare*, vol. 12, no. suppl 3, pp. 88–91, 2006.

[4] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "New computer aided device for real time analysis of speech of people with Parkinson's disease," *Revista Facultad de Ingeniería Universidad de Antioquia*, vol. 1, no. 72, pp. 87–103, 2014.

[5] J. E. Zicker, W. J. Tompkins, R. T. Rubow, and J. H. Abbs, "A portable microprocessor-based biofeedback training device," *IEEE Transactions on Biomedical Engineering*, vol. BME-27, no. 9, pp. 509–515, Sept 1980.

[6] R. Rubow and E. Swift, "A microcomputer-based wearable biofeedback device to improve transfer of treatment in Parkinsonian dysarthria," *Journal of Speech and Hearing Disorders*, vol. 50, no. 2, pp. 178–185, 1985.

[7] M. Wirebrand, "Real-time monitoring of voice characteristics using accelerometer and microphone measurements," Master's thesis, Linkping University, Linkping, Sweden., 2011.

[8] A. Carullo, A. Vallan, and A. Astolfi, "Design issues for a portable vocal analyzer," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 5, pp. 1084–1093, May 2013.

[9] 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 14th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Singapore, 2014, pp. 1473–1577.

[10] A. Tsanas, M. A. Little, P. E. Mcsharry, J. Spielman, and L. O. Ramig, "Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[11] J. C. Vásquez-Correa, N. Garcia, J. F. Vargas-Bonilla, J. R. Orozco-Arroyave, J. D. Arias-Londoño, and M. O. L. Quintero, "Evaluation of wavelet measures on automatic detection of emotion in noisy and telephony speech signals," in *Proceedings of the 48th International Carnahan Conference on Security Technology (ICCST)*, Rome, Italy, Oct 2014, pp. 1–6.

[12] Y. Hu and P. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 4, pp. 334–341, 2003.

[13] M. Westphal, "The use of cepstral means in conversational speech recognition," in *Proceedings of the 5th European Conference on Speech Communication and Technology (EUROSPEECH)*, Rhodes, Greece, 1997, pp. 1143–1146.

[14] 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 Disorders*, vol. 26, no. 10, pp. 1951–1952, 2011.

[15] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, and E. Nöth, "Spectral and cepstral analyses for Parkinson's disease detection in Spanish vowels and words," *Expert Systems*, pp. 1–10, 2015, to appear.

[16] C. G. Goetz and et al., "Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, 2008.

[17] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rátiva, and E. Nöth, "New spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland, may 2014.

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