Proceedings of the World Congress on Electrical Engineering and Computer Systems and Science (EECSS 2015)
Barcelona, Spain, July 13-14, 2015
Paper No. 303

Parkinson's Disease Detection Based on a Heterogeneous Acoustic Database

Y. Campos-Roca, D. Montaña
University of Extremadura, Department of Computer and Communication Technologies
Avda. de la Universidad, s/n, Cáceres, Spain
ycampos@unex.es; dmongut@unex.es

C. J. Pérez
University of Extremadura, Department of Mathematics
Avda. de la Universidad, s/n, Cáceres, Spain
carper@unex.es

Abstract A database composed of 50 individuals suffering from PD and 50 healthy controls has been collected to develop and validate an automatic diagnosis system. 46 acoustic features were initially extracted, which were reduced after feature selection. Repeated measurements taken on the same individual have been averaged before being assigned to subject, avoiding the usual practice of considering measurements within the same subject as independent. A classification experiment has been conducted based on support vector machines. Performance has been estimated by using k-fold cross-validation for different kernel functions. According to the obtained results, PD can be detected with an accuracy of 85% (sensitivity $83,74\%$ and specificity $87,48\%$) based on recordings of the sustained vowel /a/. These results are remarkable, since the PD patients were under treatment and the effects of the disease were decreased.

Keywords: Acoustic features, Parkinson's disease (PD), Speech disorders, Support vector machines (SVM).

1 Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease, affecting one in every 100 persons above the age of 65 years. PD may be difficult to diagnose and investigate in its early stages. Symptoms typically begin when the loss of dopamine reaches a critical point, typically when 50 to 80% of dopamine neurons have died. Up to 90% of PD patients show vocal disorders and vocal impairment is most likely one of the earliest signs of the disease (Duffy (2005)). Therefore, speech analysis is a useful tool for detecting and monitoring the disease. These procedures have the advantage over other traditional techniques of being non-invasive, objective, inexpensive and easy to self-use so that they can be used for telediagnosis and telemonitoring.

Little et al. (2009) considered both linear and non-linear measures to discriminate healthy people from those with PD. Hariharan et al. (2014) compare their proposals with the ones from fifteen previously published papers that use the data in Little et al. (2009). The results are overly optimistic because their test and training sets have significant overlap, with same speakers contributing large number of samples to both. The Parkinson Voice Initiative[1] has largely contributed to the development of this research line.

In this work, a database has been collected comprised of 50 PD subjects (in different stages of the disease and under prescribed medication) and 50 healthy controls. Repeated measurements taken on the same subject

[1] http://www.parkinsonsvoice.org/

have been averaged before being assigned to subject. Individuals with PD still experience fluctuations in speech performance when taking medication, however, the effect is lower (Goberman (2005)). This paper shows that automatic detection of PD from medicated individuals is possible and opens the possibility of increasing this database to track PD progression.

The paper is organized as follows. In Section 2, the methodology is described, including the main information on participants, speech recordings and feature extraction. Section 3 deals with the classification results. Finally, the conclusion is presented in Section 4.

2 Methodology

2.1 Participants

A total of 100 Spanish native speakers participated in the study. 50 of them were diagnosed with PD. Their mean age ($\pm$ standard deviation) was 69.6 ± 8.4 years and the stage of disease according to the Hoehn and Yahr scale ranged from 1-4 (disability scale comprised of stages 1 through 5, where 5 is the most severe). In addition, 50 healthy control subjects with no history of neurological or communication disorders were considered. Their mean age was 66.3 ± 10.1 years.

The research protocol was approved by the Bioethical Committee from the University of Extremadura. All subjects signed an informed consent. The PD patients participating in this study were members of the Regional Association for Parkinsons Disease from Extremadura (Spain).

2.2 Vocal task and speech recording

The vocal task was the sustained phonation of /a/ vowel at comfortable pitch and loudness, as constant as possible. This phonation had to be kept for at least 5 seconds and on one breath. The task was repeated three times per individual.

The speech recordings were performed using a portable computer with an external sound card (TASCAM US322) and a headband microphone (AKG 520) featuring a cardiod pattern. The digital recording was performed at a sampling rate of 44.1 KHz and a resolution of 16 bits/sample by using Audacity software (release 2.0.5).

2.3 Feature extraction

The study is based on 46 acoustic features, which can be classified into five families: pitch local perturbation measures, amplitude local perturbation measures, noise features, spectral envelope measures and nonlinear ones.

Five pitch local perturbation measures were obtained: jitter relative, jitter absolute, jitter RAP (Relative Average Perturbation), jitter PPQ (Pitch Perturbation Quotient) and jitter DDP (Average absolute difference of differences between cycles, divided by the average period)(see, e.g., Baken & Orlikoff (2000)). Jitter features were extracted by using a waveform matching pitch detection algorithm (PDA) consisting in two steps: 1) Calculation of the rough fundamental period length by using the normalized auto-correlation function over 80 ms frames and, 2) Second application of the auto-correlation function on segments with a length equal to twice the mean value of the fundamental periods roughly estimated before, after removing gross pitch errors (halving and doubling) and unvoiced frames.

The five amplitude perturbation measures are: shimmer local, shimmer dB, APQ3 (3-point Amplitude Perturbation Quotient), APQ5 (5-point Amplitude Perturbation Quotient) and APQ11 (11-point Amplitude Perturbation Quotient) (see, e.g., Baken & Orlikoff (2000)).

As noise features Harmonic-to-Noise Ratio (HNR) measures and Glottal-to-Noise Excitation (GNE) are considered. HNR is a measure of the level of tonal components relative to noise present in speech. There are many variants of HNR (based on time-domain or frequency-domain approaches). In this work we have used

five different HNR features, corresponding to different frequency bandwidths: 0-500 Hz, 0-1500 Hz, 0-2500 Hz, 0-3500 Hz and 0-3800 Hz. Individual HNR values for each frame are extracted by using the VoiceSauce toolbox freely available (Shue et al. (2010)). To the authors' knowledge, the use of subband HNR measures for PD detection has not been reported. GNE attempts to quantify the amount of voice excitation by vocal-fold oscillations versus excitation by turbulent noise. GNE measure was extracted by following the steps proposed in Michaelis et al. (1997).

Mel Frequency Cepstral Coefficients (MFCCs) are related to the speech spectral envelope, which depends on articulator position. PD is known to affect also articulation, therefore this type of coefficients are promising features to track PD progression (Tsanas et al. (2011)). By the use of MFCCs, it should be possible to detect slight misplacement of the articulators. Delta MFCC features are time derivatives of MFCCs, so they can be used to detect subtle changes in articulator positions (due to tremor) when performing a sustained phonation. In this work, both types of coefficients have been used to characterize PD. The length of the feature vector has been chosen to be 26 (13 MFCCs plus 13 Delta coefficients).

The existence of non-linear phenomena in the production process of the speech signal has been theoretically and experimentally established. Nonlinear features are of particular relevance to clinical practice, because severe dysphonic pathological voices are precisely the ones that are most likely to present highly nonlinear and random phenomena, whereas healthy voices are closer to the linear source-filter model. The following nonlinear features have been considered in this work: RPDE (Recurrence Period Density Entropy, Little et al. (2007)), DFA (Detrended Fluctuation Analysis, Little et al. (2007)) and PPE (Pitch Period Entropy, Little et al. (2009)).

Repeated measurements taken on the same individual have been averaged before being assigned to subject.

3 Results

Two hundred repetitions of k-fold cross-fold validation ($k = 5, 10$) were used to estimate the classifier performance. A sequential feature selection approach has been used to search for the optimal feature subset. The performance of SVM is evaluated based on four different kernel functions namely linear, quadratic, radial basis function (RBF), polynomial (order 3) as well as MLP (multilayer perceptron). Classifier performance was measured using overall accuracy, sensitivity and specificity. The results are shown in Table 1. The best performance has been obtained by using a linear kernel function.

Table 1: Classifier performance.

Validation method	SVM kernel	Accuracy (%)	Sensitivity (%)	Specificity (%)
10-fold	Linear	85,70	83,96	87,44
10-fold	Quadratic	76,71	74,88	78,54
10-fold	RBF	84,17	78,94	89,39
10-fold	POLY, order=3	78,58	72,29	84,86
10-fold	MLP	72,25	71,11	73,39
5-fold	Linear	83,66	81,51	85,81
5-fold	Quadratic	74,72	73,33	76,10
5-fold	RBF	81,32	74,19	88,45
5-fold	POLY, order=3	77,09	69,95	84,23
5-fold	MLP	71,57	70,04	73,10

In the 10-fold cross-validation experiment, an overall accuracy of $85,7\%$ is achieved and the 95% Confidence Interval (CI) is $85,41 - 85,99$. Balanced sensitivity ($83,96\%$, with 95% CI $83,54 - 84,37$) and

specificity ($87,44\%$, with 95% CI $87,05 - 87,83$) are achieved. These results are remarkable, since the PD patients were under prescribed medication and the effects of the disease were decreased.

4 Conclusion

An automatic system for detection of PD from speech signals has been developed. A new database has been created and used to validate the proposed approach. The system is able to accurately discriminate between healthy individuals and PD subjects, even when PD patients are under prescribed medication. The collection of a heterogeneous database (including PD individuals in different severity stages) allows for future work on PD progression.

Acknowledgments

Thanks to the anonymous participants and to Rosa María Muñoz and some Bachelor's students for carrying out the voice recordings and providing information from the patients. We are also grateful to the *Asociación Regional de Parkinson de Extremadura* and *Confederación Española de Personas con Discapacidad Física y Orgánica* for providing support in the experiment development. The authors also thank some other Bachelor's students for some programming tasks.

This research has been supported by *Ministerio de Economía y Competitividad*, Spain (Projects MTM2011-28983-C03-02 and MTM2014-56949-C3-3-R), *Gobierno de Extremadura*, Spain (Projects GR15052 and GR15106) and *European Union* (European Regional Development Funds). Funding for a PhD grant comes from Fundación Tatiana Pérez de Guzmán el Bueno (Spain).

References

Baken, J. & Orlikoff, R. F. (2000), *Clinical Measurement of Speech and Voice*, Singular Thomson Learning.

Duffy, J. R. (2005), *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management*, Elsevier.

Goberman, A. (2005), 'Correlation between acoustic speech characteristics and non-speech motor performance in Parkinson disease', *Medical Science Monitor* **11**(3), 109–116.

Hariharan, M., Polat, K. & Sindhu, R. (2014), 'A new hybrid intelligent system for accurate detection of Parkinson's disease', *Computer Methods and Programs in Biomedicine* **113**(3), 904–913.

Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J. & Ramig, L. O. (2009), 'Suitability of dysphonia measurements for telemonitoring of Parkinson's disease', *IEEE Transactions on Biomedical Engineering* **56**(4), 1015–1022.

Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A. E. & Moroz, I. M. (2007), 'Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection', *BioMedical Engineering OnLine* **6**(23), 1–19.

Michaelis, D., Gramss, T. & Strube, H. W. (1997), 'Glottal-to-noise excitation ratio - a new measure for describing pathological voices', *Acustica/Acta acustica* **83**, 700–706.

Shue, Y. L., Keating, P., Vicenik, C. & Yu, K. (2010), 'Voicesauce: A program for voice analysis', *Energy* **1**(H2), H1–A1.

Tsanas, A., Little, M. A., McSharry, P. E. & Ramig, L. O. (2011), 'Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity', *The Royal Society Interface* **8**(59), 842–855.