Speech Communication 101 (2018) 11–25

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Speaker models for monitoring Parkinson's disease progression considering different communication channels and acoustic conditions

T. Arias-Vergara[a,b], J.C. Vásquez-Correa[a,b], J.R. Orozco-Arroyave[*,a,b], E. Nöth[b]

[a] GITA research group, Faculty of Engineering, Universidad de Antioquia, Medellín, Colombia
[b] Pattern Recognition Lab., Friedrich-Alexander-Universität, Erlangen, Nürnberg, Germany

ARTICLE INFO

Keywords:
Speech disorders
GMM–UBM
i–Vectors
Parkinson's disease
Dysarthria
Speaker models
Longitudinal analysis

ABSTRACT

Symptoms of Parkinson's disease vary from patient to patient. Additionally, the progression of those symptoms also differs among patients. Most of the studies on the analysis of speech of people with Parkinson's disease do not consider such an individual variation. This paper presents a methodology for the automatic and individual monitoring of speech disorders developed by PD patients. The neurological state and dysarthria level of the patients are evaluated. The proposed system is based on individual speaker models which are created for each patient. Two different models are evaluated, the classical GMM–UBM and the i–vectors approach. These two methods are compared with respect to a baseline found with a traditional Support Vector Regressor. Different speech aspects (phonation, articulation, and prosody) are considered to model recordings of spontaneous speech and a read text. A multi-aspect coefficient is proposed with the aim of incorporating information from all of these speech aspects into a single measure. Two different scenarios are considered to assess a set with seven PD patients: (1) the longitudinal test set which consists of speech recordings captured in five recording sessions distributed from 2012 to 2016, and (2) the at-home test set which consists of speech recordings captured in the home of the same seven patients during 4 months (one day per month, four times per day). The UBM is trained with the recordings of 100 speakers (50 with Parkinson's disease and 50 healthy speakers) captured with controlled acoustic conditions and a professional audio-setting. With the aim of evaluating the suitability of the proposed approaches and the possibility of extending this kind of systems to remotely assess the speech of the patients, a total of five different communication channels (sound-proof booth, Skype®, Hangouts®, mobile phone, and land-line) are considered to train and test the system. Due to the reduced number of recording sessions in the longitudinal test set, the experiments that involved this set are evaluated with the Pearson's correlation. The experiments with the at-home test set are evaluated with the Spearman's correlation. The results estimating the dysarthria level of the patients in the at-home test set indicate a correlation of 0.55 with a modified version of the Frenchay Dysarthria Assessment scale when the GMM-UBM model is applied upon the Skype® recordings. The results in the longitudinal test set indicate a correlation of 0.77 using a model based on i-vectors with recordings captured in the sound-proof-booth. The evaluation of the neurological state of the patients in the longitudinal test set shows correlations of up to 0.55 with the Movement Disorder Society - Unified Parkinson's Disease Rating Scale also using models based on i-vectors created with Skype® recordings. These results suggest that the i–vector approach is suitable when the acoustic conditions among recording sessions differ (longitudinal test set). The GMM-UBM approach seems to be more suitable when the acoustic conditions do not change a lot among recording sessions (at-home test set). Particularly, the best results were obtained with the Skype® calls, which can be explained due to several preprocessing stages that this codec applies to the audio signals. In general, the results suggest that the proposed approaches are suitable for tele-monitoring the dysarthria level and the neurological state of PD patients.

* Corresponding author at: GITA research group, Faculty of Engineering, Universidad de Antioquia, Medellín, Colombia.
E-mail address: rafael.orozco@udea.edu.co (J.R. Orozco-Arroyave).

https://doi.org/10.1016/j.specom.2018.05.007
Received 29 June 2017; Received in revised form 10 May 2018; Accepted 28 May 2018
Available online 31 May 2018
0167-6393/

1. Introduction

1.1. Motivation

People suffering from PD are characterized by the progressive loss of dopaminergic neurons in the midbrain (Hornykiewicz, 1998). PD symptoms include tremor, slow movement, lack of coordination, and speech impairments (Ho et al., 1999; Darley et al., 1969). Currently, neurologists rely on medical history, physical and neurological examinations to assess the patients. This procedure has two main limitations: (i) it is not objective (the evaluation depends on the doctor's criterion and expertise), and (ii) due to the motor disability of PD patients, to visit a hospital to perform medical screenings and/or assessments is expensive and difficult (Theodoros et al., 2006). Besides such difficulties, the symptoms progress differently among patients, thus it is important to monitor their symptoms individually (per patient) and over long periods of time. Such a monitoring is not feasible if the patient is required to visit the doctor to every screening. The most suitable methods to perform continuous monitoring of the symptoms are based on computer-aided tools. These methods have captured the attention of the research community because they are objective, easy to use, and reproducible. Speech signals are one of the most suitable ways to capture information about the neurological state of PD patients (Tsanas et al., 2010; Skodda et al., 2013; Orozco-Arroyave et al., 2016a). Studies reported in the state-of-the-art about assessing the neurological state of PD patients from speech signals always consider situations where the acoustic conditions are relatively controlled, i.e., quiet rooms, good/expensive microphones, and direct connection to the recording device. Additionally, the state-of-the-art is mainly based on classical methods to model speech signals, i.e., measurements are extracted from the speech signal and regression methods are used to assess the neurological state of the patient. This paper presents a methodology for the individual monitoring of speech impairments developed by PD patients during the disease progression. The proposed approach overcomes the state-of-the-art in several aspects: (i) the method is based on individual models, which are based on Gaussian Mixture Models – Universal Background Models (GMM–UBM), thus the system performance is adapted to the speech of each patient, (ii) different communication channels are considered including land-lines, mobile phones, Internet-based systems (Skype® and Hangouts®), and traditional recordings performed during a medical appointment. The proposed approach is also tested on two kinds of recordings: (i) signals captured during several recording sessions distributed from 2012 to 2016, and (ii) signals captured in 16 sessions performed in the houses of several patients during 4 months (one day per month, every two hours and during 8 h). The use of these two recording sets make the experiments reported in this paper highly original and novel, thus we consider that this work is a significant contribution to the development of computer-aided tools to monitor the progression of PD.

1.2. Parkinson's disease: evaluation and monitoring

1.2.1. Neurological evaluation

There is no standard test to diagnose PD. Doctors rely on the clinical history and physical examinations to assess patients. There are several tests to evaluate the disease severity. One of the most widely used is the Movement Disorder Society - Unified Parkinson's Disease Rating Scale (MDS-UPDRS). This scale is divided into four sections: Section 1 comprises non-motor experiences (13 items), Section 2 includes motor activities of daily living (13 items), Section 3 evaluates motor capabilities (33 items), and Section 4 considers motor complications (6 items) (Goetz et al., 2008). Although the scale has a total of 65 items, speech is only considered in one of them.

1.2.2. Dysarthria level assessment

There are several scales and clinical methods to evaluate dysarthric speech. One of them is the Frenchay Dysarthria Assessment–2 (FDA–2) (Enderby and Palmer, 2008). The original version of the FDA–2 considers several factors that are affected in people suffering from dysarthria, such as reflexes, respiration, lips movement, palate movement, laryngeal capability, tongue posture/movement, intelligibility, and others. The FDA–2 requires the patient to visit the examiner, which is not possible in most cases when people suffering from PD are considered. Bering this in mind, it was necessary to develop a modified version of the FDA (m–FDA), which can be administered based on speech signals previously recorded, thus the patient is not required to visit the clinician to be evaluated (Cernak et al., 2017). The m–FDA considers several aspects of speech: respiration, lips movement, palate/velum movement, larynx, tongue, monotonicity, and intelligibility. Speech impairments are evaluated in a total of 13 items and each of them ranges from 0 (normal or completely healthy) to 4 (very impaired), thus the total score of the scale ranges from 0 to 52.

1.2.3. Assessment of the neurological state from speech

In recent years the research community has been interested in developing methods to assess the neurological state of PD patients from speech. One of the reasons to look for such an aim is to reduce treatment and monitoring costs and another reason is to develop objective tools/systems that help clinicians in the assessment and screening of the patients. In Asgari and Shafran (2010) the authors proposed a methodology to assess the UPDRS-III score from speech recordings of 82 subjects. The participants were asked to perform three speech tasks including the sustained phonation of the vowel /a/, the rapid repetition of the syllables (/pa/-/ta/-/ka/), and the reading of three standard texts. The set of features extracted from the speech recordings include pitch, spectral entropy, 13 cepstral coefficients, the number and duration of voiced and unvoiced frames, jitter, shimmer, Harmonic to Noise Ratio (HNR), and the ratio of energy in the first and second harmonics. The set of features was computed separately for each speech task. The UPDRS scores were obtained using two Support Vector Regressor (SVR)-based approaches: (1) ϵ-SVR and (2) ν-SVR. Additionally, different kernels were used to train the SVRs including polynomial, radial basis function, and sigmoid functions. The authors reported that it is possible to estimate the UPDRS-III with a Mean Absolute Error (MAE) of 5.66 using an ε-SVR with a cubic polynomial kernel. Later in Bayestehtashk et al. (2015) the authors compared three regression techniques to assess the UPDRS scores including ridge regression, Least Absolute Shrinkage and Selection Operator (LASSO) regression, and linear SVR. Speech recordings of 168 patients were collected in a single recording session. Besides the features described in Asgari and Shafran (2010), the authors added information extracted with the openSMILE toolkit (Eyben et al., 2010). The authors reported that the neurological state of the patients can be assessed with a MAE of 5.5 considering only PD patients in the training process, however, due to the lack of longitudinal data, it is not clear whether the proposed approach is suitable to track the neurological state of each patient. Furthermore, the results are presented only in terms of the MAE, which only makes sense when there is a baseline to compare the performance of the models. Besides, in the INTERSPEECH 2015 Computational Paralinguistic Challenge (ComParE 2015) our team participated in the organization of the Parkinson's Condition sub-challenge, where the task of neurological state evaluation of PD patients from speech was addressed (Schuller et al., 2015). Recordings of the 50 patients (25 male, 25 female) included in the PC-GITA database (Orozco-Arroyave et al., 2014) were considered to form the train and development subsets. The test set included a total of 11 patients recorded in non-controlled noise conditions, i.e., not using a sound-proof booth and a professional audio setting. A total of 42 speech tasks were considered. The neurological state of the patients was assessed by a neurologist expert according to the motor section of the MDS-UPDRS (MDS-UPDRS-III). The winners of the challenge reported a Spearman's correlation coefficient of 0.65 between the real MDS-UPDRS-III scores and the estimated values. The authors developed a