Available online at www.sciencedirect.com
ScienceDirect
Computer Speech and Language xxx (2013) xxx–xxx
COMPUTER SPEECH AND LANGUAGE
www.elsevier.com/locate/csl

Fully automated assessment of the severity of Parkinson's disease from speech☆

Alireza Bayestehtashk [a], Meysam Asgari [a], Izhak Shafran [a,*], James McNames [b]

[a] *Center for Spoken Language Understanding, Oregon Health & Science University (OHSU), United States*
[b] *Electrical and Computer Engineering, Portland State University, United States*

Received 21 June 2013; received in revised form 3 December 2013; accepted 13 December 2013

Abstract

For several decades now, there has been sporadic interest in automatically characterizing the speech impairment due to Parkinson's disease (PD). Most early studies were confined to quantifying a few speech features that were easy to compute. More recent studies have adopted a machine learning approach where a large number of potential features are extracted and the models are learned automatically from the data. In the same vein, here we characterize the disease using a relatively large cohort of 168 subjects, collected from multiple (three) clinics. We elicited speech using three tasks – the sustained phonation task, the diadochokinetic task and a reading task, all within a time budget of 4 min, prompted by a portable device. From these recordings, we extracted 1582 features for each subject using openSMILE, a standard feature extraction tool. We compared the effectiveness of three strategies for learning a regularized regression and find that ridge regression performs better than lasso and support vector regression for our task. We refine the feature extraction to capture pitch-related cues, including jitter and shimmer, more accurately using a time-varying harmonic model of speech. Our results show that the severity of the disease can be inferred from speech with a mean absolute error of about 5.5, explaining 61% of the variance and consistently well-above chance across all clinics. Of the three speech elicitation tasks, we find that the reading task is significantly better at capturing cues than diadochokinetic or sustained phonation task. In all, we have demonstrated that the data collection and inference can be fully automated, and the results show that speech-based assessment has promising practical application in PD. The techniques reported here are more widely applicable to other paralinguistic tasks in clinical domain.

Keywords: Parkinson's disease; Pitch estimation; Jitter; Shimmer

1. Introduction and motivation

Parkinson's disease (PD), which is characterized by tremors and impaired muscular co-ordinations, currently has no cure and hence screening for early detection and monitoring its progression are important tools for managing the disease in the growing population of the elderly. The disease is associated with low levels of dopamine in the brain and the symptoms are managed by artificially increasing amounts of dopamine with drugs (e.g., L-dopa) and in severe cases by electrically stimulating specific regions in the mid-brain. The severity of the disease is typically assessed in a

☆ This paper has been recommended for acceptance by Dr. S. Narayanan.
* Corresponding author. Tel.: +1 503 748 1158; fax: +1 503 748 1603.
E-mail address: zakshafran@gmail.com (I. Shafran).

0885-2308/$ – see front matter © 2013 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.csl.2013.12.001

clinic with a battery of tests – the Unified Parkinson's Disease Rating Scale (UPDRS) – consisting of clinician-scored motor evaluations and self evaluation of the activities of daily life including speech, swallowing, handwriting, dressing, hygiene, falling, salivating, turning in bed, walking, and cutting food. The UPDRS scores range from 0 to 176, with 0 corresponding to a healthy state and 176 to a severe affliction (MDSTF, 2011). The assessment is time-consuming and is performed by trained medical personnel, which becomes burdensome when, for example, frequent re-assessment is required to fine-tune dosage of drugs or the parameters of the electrical pulse train in deep brain stimulations. Not surprisingly, there has been a growing interest in creating tools and methods for alternative home-based assessments of this disease. Easier methods of assessment can play a crucial role in screening for early detection of PD, the second most common neurodegenerative disease in US.

Since speech production involves complex motor coordination, the disease exhibits tell tale symptoms which are well-known to speech pathologists, although the exact pathophysiological cause remains unclear. For several decades now, researchers have been interested in measuring these symptoms in speech more objectively with the hope of augmenting or simplifying the assessment. Speech tasks can be administered remotely, avoiding the need for driving to the clinic, which can be challenging for those with severe PD-related motor tremors. Speech can be elicited, recorded and analyzed automatically relatively easily at much lower cost than in-person clinical assessment. Furthermore, speech-based assessment can monitor changes objectively over time more accurately.

While there has been considerable interest in analyzing speech in PD, spanning about five decades, only recently it has attracted the attention of computational speech researchers. Early studies, reported in clinical journals, employed relatively simple analysis of speech samples. Here, we set the context of this work by reviewing a sampling of previous work in Section 2. As evident from this review, previous studies have several limitations. With a few exceptions, most studies have been conducted on relatively smaller cohorts, recruited from a single clinic and often narrowly focused on characterizing pathology related to production of vowels. Data collected from a single clinic can suffer from bias due to the subjective nature of clinical assessments.

In this article, we investigate the accuracy of automatically inferring the severity of PD from speech samples in a relatively large cohort collected from multiple clinics. The data collection is described in detail in Section 3. One of the aims of this paper is to investigate the utility of current speech processing and machine learning techniques with publicly available tools for inferring the severity of Parkinson's disease. In Section 4, we extract a number of potential speech features using standard speech processing algorithms and apply several machine learning algorithms to predict the clinical ratings from the speech features. Standard pitch detection algorithms do not have the necessary time-frequency resolution to capture the fine tremors observed in PD. We recently developed a pitch estimation algorithm that addresses this problem, which incidentally won the 2013 Interspeech Challenge on detecting and diagnosing Autism Spectral Disorders (Asgari and Shafran, 2013). We describe our method and our evaluation on PD in Section 5. Finally, we summarize the contributions of this paper.

2. Brief review of speech in PD

The earliest work on measuring speech abnormalities objectively in PD can be traced back to Canter's dissertation. Taking advantage of then newly available instruments to measure pitch using a direct-writing oscilloscope (Sanborn, Model 450) and vocal intensity using a high-speed level recorder (Bruel and Kjaer, Model 2304), Canter compared speech from 17 patients, who were off medication for 48 h, with 17 age-matched controls (Canter, 1963, 1965a,b). PD subjects exhibited higher median pitch and lower range than the controls. They lacked the necessary control to generate soft sounds. At the other end of the scale, they had lower intensity of maximal loudness. They were able to sustain phonation for markedly shorter duration, about 50% of the control. In articulation, PD subjects were slower and plosives lacked precision, often confused with fricatives. Canter also noted that the rate of speech and intelligibility were significantly different from the controls. Most of his conclusions, even though deduced from manual measurements from plots, have been subsequently confirmed by measurements with more sophisticated instruments, although with small number of subjects (Titze, 1994; Darley et al., 1975). One exceptionally large sample study collected speech from about 200 PD subjects (Logemann et al., 1978). They confirmed Canters findings and in addition observed characteristic types of misarticulations (e.g., back-of-tongue, tongue tip, lips). The above mentioned studies were conducted by clinical researchers who compiled the features with manual measurements and perceptual ratings.

There have been very few studies on employing automatic speech processing to classify PD subjects from controls or inferring the severity of the diseases. Here we describe a few representative studies. Guerra and Lovely attempted to

Please cite this article in press as: Bayestehtashk, A., et al., Fully automated assessment of the severity of Parkinson's disease from speech. Comput. Speech Lang. (2013), http://dx.doi.org/10.1016/j.csl.2013.12.001