Applied Soft Computing 62 (2018) 649–666

Contents lists available at ScienceDirect

Applied Soft Computing

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

Analysis of speaker recognition methodologies and the influence of kinetic changes to automatically detect Parkinson's Disease

Laureano Moro-Velázquez [a,*], Jorge Andrés Gómez-García [a], Juan Ignacio Godino-Llorente [a], Jesús Villalba [b], Juan Rafael Orozco-Arroyave [c,d], Najim Dehak [b]

[a] Center for Biomedical Technology, Universidad Politécnica de Madrid, Campus Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain
[b] ECE, Johns Hopkins University, Homewood Campus, 21218 Baltimore, MD, USA
[c] Universidad de Antioquia, 050010 Medellín, Colombia
[d] Pattern Recognition Lab, University of Erlangen, 91058 Erlangen, Germany

A R T I C L E I N F O

Article history:
Received 28 December 2016
Received in revised form 18 August 2017
Accepted 1 November 2017
Available online 23 November 2017

Keywords:
Parkinson's Disease
Speech
GMM-UBM
iVectors
Speaker recognition techniques

A B S T R A C T

The diagnosis of Parkinson's Disease is a challenging task which might be supported by new tools to objectively evaluate the presence of deviations in patient's motor capabilities.

To this respect, the dysarthric nature of patient's speech has been exploited in several works to detect the presence of this disease, but none of them has deeply studied the use of state-of-the-art speaker recognition techniques for this task.

In this paper, two classification schemes (GMM-UBM and *i-Vectors*-GPLDA) are employed separately with several parameterization techniques, namely PLP, MFCC and LPC. Additionally, the influence of the kinetic changes, described by their derivatives, is analysed.

With the proposed methodology, an accuracy of 87% with an AUC of 0.93 is obtained in the optimal configuration. These results are comparable to those obtained in other works employing speech for Parkinson's Disease detection and confirm that the selected speaker recognition techniques are a solid baseline to compare with future works. Results suggest that Rasta-PLP is the most reliable parameterization for the proposed task among all the tested features while the two employed classification schemes perform similarly. Additionally, results confirm that kinetic changes provide a substantial performance improvement in Parkinson's Disease automatic detection systems and should be considered in the future.

1. Introduction

The second most prevalent neurodegenerative disease, Parkinson's Disease (PD), is usually diagnosed on the basis of the observation of motor *cardinal signs* [1] and other non-motor indicators (physiological and cognitive manifestations) which are employed in the *clinical diagnosis*. Despite neuropathological diagnosis during autopsy is considered as the gold standard, some studies demonstrate that following the usual clinical diagnosis criteria it is possible to obtain 90% of accuracy in final judgement, but the average detection time to reach this accuracy is 2.9 years [2]. To monitor the progress of the disease, specialists often employ the Unified Parkinson's Disease Rating Scale (UPDRS)[1] [3] or the Hoehn and Yahr (H&Y) scale[2] [4] which include objective and non-objective assessments. In this regard, new technologies could accelerate the diagnosis process and provide a more objective monitoring of the affection.

Since PD affects the coordination of movements, it is reasonable to hypothesize that the assessment of the patients' performance during a complex motor task might be employed for diagnosis purposes. Speech production, an ability that is almost universal, might be affected by PD since it involves complex and very precise movements, but on spite of being good candidate for PD detection and evaluation, its capabilities have not been deeply exploited yet.

* Corresponding author.
E-mail address: laureano.moro@upm.es (L. Moro-Velázquez).

[1] UPRDS comprises four main parts: I, mentation, behaviour and mood; II, activities of daily living; III motor; IV complications. The rating of this scale is obtained through interviews with the patient and clinical observations. The best possible score for each part is 0 whereas the worst one depends on the part. The global UPDRS value can range between 0 and 147, where the larger the value, the higher the affection of the disease.

[2] The H&Y scale comprises several levels whose values can range from 1 to 5 in which 1 implies that the patient has low or no functional disabilities and 5 that patient is totally dependent.

https://doi.org/10.1016/j.asoc.2017.11.001

It is well known that the neurodegenerative processes associated to the disorder cause *hypokinetic dysarthria*, thus producing a reduction of loudness and articulation amplitude, slowing down the speech sometimes and principally reducing intelligibility [5–9]. Literature evidences the influence of PD on speech from early to advanced stages although it is mainly perceived in mild to advanced phases [10–12]. In this regard, it is expected that new methods of automatic assessment employing voice and speech can be used to detect the signs that are not perceived in the first stages of the disease but which could provide relevant information.

There are several studies and approaches using speech or voice to find biomarkers of the presence of PD or to assess its severity. Most of the literature can be divided into four groups depending on the analysed aspect: *phonatory*, *articulatory*, *prosodic* and *linguistic*. The *phonatory* studies are related to the glottal source and resonant structures of the vocal tract. Works based on *articulatory* and *prosodic* aspects are more abundant and diverse as there exist more analysis possibilities and since the influences of PD in articulation and prosody seem to be more evident [12]. The works within these groups are based on syllable rate analysis, or the processing of certain segments of the speech to obtain indexes correlated with the disease. Concerning the *prosodic* works, studies are mainly focused in the paralinguistic features such as pitch variation or the manifestation of emotions among others [13–17]. Finally, the studies related to deviations in the *linguistic* domain examine the vocabulary, phrase construction and the existence of word repetitions. Some representative works within this group are found in [18–21]. The speech material used in each case is a differentiating factor of the four groups. In the phonatory analysis, the most advisable acoustic material is sustained vowels while in the other three groups, running speech is needed. Specifically, in articulatory analysis, diadochokinetic (DDK)[3] speech can be valuable in addition to the other running speech materials such as spontaneous speech or reading text.

The present study can be framed into both the *articulatory* and the *phonatory* groups attending to the type of analysis that is employed and the acoustic material used.

Going into detail about some *articulatory* relevant works, studies as [23] indicate that speech processing can produce powerful indicators of imprecise consonant articulation in PD-related dysarthria. Authors perform an analysis of DDK tasks (/pa-ta-ka/) in a database of 24 PD patients and 22 controls, providing 88% of efficiency on separating PD from controls. In this study all the utterances are subdivided automatically into different representative segments to analyse articulation. Only 13 features are obtained by performing measurements on these segments, each feature describing a different articulatory trait of speech. Its main drawbacks are the use of a small database which is sex unbalanced and the use of only DDK utterances, limiting the possible articulatory combinations. Other works such as [24] employ frequency features, namely Mel Frequency Cepstrum Coefficients (MFCC) and Band Bark Energies (BBE) from running speech, and other features obtained after the segmentation of specific regions, providing good results with three corpora. However, in this case the results are too optimistic due to an over-fitting of the model, since it was optimized during training.

Equally, there are *articulatory* studies more focused on the fluctuations of the voice onset, offset and break segments during running speech, which are considered to be crucial in the evaluation of voice quality. For instance, in [25,26] it is evidenced that the parkinsonian speech has lower values of relative fundamental frequency, which is the ratio between the fundamental frequency in the cycles of a vowel before or after a voiceless consonant and the typical fundamental frequency during the utterance. The main drawback of these two works is that the databases are unbalanced in sex, which could bias some conclusions. Other studies perform the tracking of vowel formants and VSA during articulation, including onset and offset, with heterogeneous results [27–29]. As formants reflect the position of the tongue, a reduction of the articulation ranges could subsequently limit the frequency ranges of the formants. In [30], a comparison of PD detection techniques is performed using the acoustic material extracted from sustained vowels, sentence repetitions, reading passages and monologues. An accuracy of 80% is achieved using vowels extracted from monologues, providing enhanced results compared to utilizing sustained vowels. The main drawback of this study is the use of a small and unbalanced database (20 patients and 15 controls).

In any case, these and many other works such as [8,9,31,32] evidence that articulation perturbations introduced by PD can provide reliable information about the presence of the disorder.

Respecting the *phonatory* works, sustained vowels are expected to generate simpler acoustic structures that might be easier to analyse. Some works demonstrate that it is possible to detect the influence of PD on the vocal folds vibration by reason of the presence of noise and other perturbations caused by incomplete closure [33], abnormal phase closure and phase asymmetry or vocal tremor [34]. Likewise, some works like [35–37] use dysphonia measures including noise or frequency and amplitude perturbations to assess the severity of PD in telemonitoring scenarios achieving good results. A major drawback, though, is that recordings are done using portable and different equipments introducing noise and variability in the databases which could bias the system.

Finally, other authors employ a combination of techniques as in [38]. In this case, phonatory, prosodic and articulatory features are used jointly, providing results of 80% of accuracy in PD detection.

Although there are many approaches using voice and speech as acoustic materials to detect and assess PD, as far as the authors of this study know, no work has analysed thoughtfully the use of state-of-the-art speaker recognition techniques for this task. Two major classification schemes in this field are *Gaussian Mixture Model – Universal Background Model* (GMM-UBM) [39] and *i-Vectors* [40] which are usually employed in combination with phonatory and articulatory information of the speaker. In this study several PD automatic detectors are analysed using GMM-UBM and *i-Vectors* in combination with different parameterizations and speech tasks.

The paper is organized as follows: Section 2 summarizes the main guidelines of this study. Section 3 develops the theoretical background about the different parameterizations and classification techniques. Section 4 introduces the experimental setup and describes the databases used in this study. Section 5 presents the obtained results. Lastly, Section 6 presents the discussions and 7, the conclusions and future work.

2. Overview and contribution

The present work performs a thorough study about the influence of the different parameters and configurations of state-of-the-art speaker recognition techniques for the detection of PD. Mainly, different combinations of acoustic material, parameterization and classification schemes are analysed separately to identify the strengths and weaknesses of each one in PD detection.

As it is depicted in Fig. 1, speech materials can be a sustained vowel, a DDK task or two different sentences. Three families of

[3] DDK tests consist in the repetition of words or syllables starting in a calm syllable rate which is increased until the speaker reaches her/his limit rate. Literature reports significant differences between controls and PD patients in several measures over DDK tasks [22]. DDK is of great interest because this task implies alternating articulatory movements where the employment of plosive syllables and different points of articulation promotes a good scenario in which changes in velocity of articulation can facilitate PD detection or assessment.