Computer Methods and Programs in Biomedicine 154 (2018) 89–97

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

A Diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of Parkinson's disease

David Montaña[a], Yolanda Campos-Roca[a], Carlos J. Pérez[b,*]

[a] Department of Computer and Communication Technologies, Universidad de Extremadura, Avda. de la Universidad s/n, Cáceres 10003, Spain
[b] Department of Mathematics, Universidad de Extremadura, Avda. de Elvas s/n, Badajoz 06006, Spain

ARTICLE INFO

Article history:
Received 25 April 2017
Revised 10 October 2017
Accepted 14 November 2017

Keywords:
Expert system
Acoustic features
Classification
Diadochokinesis (DDK)
Parkinson's disease (PD)
Speech disorders

ABSTRACT

Background and objective: A new expert system is proposed to discriminate healthy people from people with Parkinson's Disease (PD) in early stages by using Diadochokinesis tests.

Methods: The system is based on temporal and spectral features extracted from the Voice Onset Time (VOT) segments of /ka/ syllables, whose boundaries are delimited by a novel algorithm. For comparison purposes, the approach is applied also to /pa/ and /ta/ syllables. In order to develop and validate the system, a voice recording database composed of 27 individuals diagnosed with PD and 27 healthy controls has been collected. This database reflects an average disease stage of 1.85 ± 0.55 according to Hoehn and Yahr scale. System design is based on feature extraction, feature selection and Support Vector Machine learning.

Results: The novel VOT algorithm, based on a simple and computationally efficient approach, demonstrates accurate estimation of VOT boundaries on /ka/ syllables for both healthy and PD-affected speakers. The PD detection approach based on /k/ plosive consonant achieves the highest discrimination capability (92.2% using 10-fold cross-validation and 94.4% in the case of leave-one-out method) in comparison to the corresponding versions based on the other two plosives (/p/ and /t/).

Conclusion: A high accuracy has been obtained on a database with a lower average disease stage than previous articulatory databases presented in the literature.

1. Introduction

Expert systems have been used in almost every field of medicine, being diagnosis the dominant decision-making issue. A survey on expert systems for diagnosis support in the field of neurology is presented in [1]. The key element of an expert system is the knowledge base. Complex areas in medicine require extensive knowledge that may be extracted from clinical datasets [2–4]. The primary aim of this research is to design an expert system for early detection of Parkinson's disease (PD).

PD is the second most common neurodegenerative disorder after Alzheimer's disease. According to the Parkinson's Disease Foundation, an estimated 7 to 10 million people worldwide are living with this medical condition. This disease is a chronic neurodegenerative disorder caused by the progressive degeneration and death of dopaminergic neurons, that play a key role in coordinating the movement at level of muscle tone.

Voice and speech, as dependent on laryngeal, respiratory and articulatory functions, may also be affected in patients with PD [5]. Acoustic analysis on recorded speech signals can help to detect subtle abnormalities in speech that may not be perceptible to listeners [6]. Some authors have considered measures extracted from speech recordings and machine learning techniques to discriminate healthy people from those with PD [7–12]. These techniques have a great potential to establish efficient biomarkers that may help neurologists in their diagnoses or allow primary care physicians to refer the patient to a neurology unit.

The medical literature describes numerous advantages that may be associated with early intervention in PD [13]. Besides medical treatment, PD patients should have access to other type of services including physiotherapy or speech and language therapy. Therefore, successfully addressing early diagnoses of people with PD is a key

* Corresponding author.
E-mail addresses: dmongut@unex.es (D. Montaña), ycampos@unex.es (Y. Campos-Roca), carper@unex.es (C.J. Pérez).

https://doi.org/10.1016/j.cmpb.2017.11.010
0169-2607/

issue to improve the patients' quality of life. However, it is estimated that 20% of people with PD remain undiagnosed [14].

An important drawback to test the effectiveness of a detection methodology in an early-stage scenario is the scarcity of data. Therefore, one of the primary goals of this investigation was to build a new voice recording database. In order to check a system for early diagnosis of PD, patients between 1 and maximum 2.5 of the Hoehn and Yahr (H&Y) scale have been considered [15]. This corresponds to cases from very mild to mild stages. To the best of the authors' knowledge, the experimental results shown have been obtained on a database with lower average disease stage (1.85 ± 0.55 according to H&Y scale) in comparison to previous articulatory datasets reported in the literature.

Note that the goal of this work is not to track progression of concrete symptoms or search for a correlation between speech impairment and H&Y stage, but to predict disease presence in an automatic way as early as possible. Stage 1 (H&Y) represents the earliest definable stage of PD with current diagnostic. So if a computer-aided system is able to predict that the disease is present (even in the case of patients at low H&Y stages), this means that this system is able to detect the disease in an automatic way, without the need of a neurologist, with the consequent cost saving. Once the disease is detected by a general practitioner, the patient should be forwarded to the neurology unit for further tracking of the disease.

There are several speaking tasks that could be used to evaluate voice disorders in PD based on the extraction of phonation, articulation or prosody features. Most of the approaches for automatic detection of PD from speech are based on sustained vowel phonations, where the speaker attempts to produce a vowel sound as steady as possible in terms of amplitude and fundamental frequency. This type of vocal task enables the measurement of dysphonic aspects of speech. By using features extracted from sustained vowels, accuracy rates between 73.5% to 100% have been reported depending on the feature selection and classification techniques based on the dataset provided in [7] (see [11]). However, it is important to note that these approaches lead to overoptimistic estimations of the accuracy rates, since they are based on replicated measurements (6 voice recordings per subject) and the dependent nature of the observations has not been taken into account [16,17]. Peréz et al. [12] demonstrate the first classification approach for PD detection that takes into account the underlying within-subject dependence of the replicated recordings by using the dataset provided in [7]. In this case the accuracy rate was 90.4%. In spite of this, it is necessary to highlight the key role that this dataset has played in the development of this research line, allowing the investigation with new linear and nonlinear features.

Articulatory difficulties represent an important manifestation of speech disorders in PD. This fact motivates the search for features extracted from speaking tasks involving quick movements of the articulators. Diadochokinesis (DDK) tests are one of the most common tools to evaluate articulatory impairments in both research and clinical assessment contexts. DDK tasks typically measure the subject's ability to repeat a consonant - vowel (C-V) combination with bilabial, alveolar, and velar places of articulation, quickly and at a rhythmic timing. Subjects are asked to repeat a combination of the three-syllable train, for example, /pa/-/ta/-/ka/, as fast as possible.

In [18], the authors presented an approach to discriminate PD from healthy controls (HCs) based on features extracted from DDK utterances. This study was based on 24 individuals diagnosed with PD and 22 HCs, all of them Czech native speakers. The task was repeated twice per speaker resulting in the acquisition of 80 utterances in total. The within-subject dependence of the utterances was not taken into account. The approach was based on 13 features representing six different articulatory aspects of speech: vowel quality, coordination of laryngeal and supralaryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing. The authors achieved best success rates of 87.1% (using 10-fold cross-validation) and 88.4% (with the leave-one-out (LOO) method) on a database reflecting disease stages of 2.2 ± 0.5 (H&Y scale).

In [17], the authors performed discrimination of PD from HCs based on phonation, articulation and prosody, by using different speaking tasks (including rapid syllable repetition). Three different languages (Spanish, German and Czech) were considered. The reported accuracy was 99% for Spanish in the case of features extracted from the unvoiced segments of DDK utterances. These segments were modeled by using 12 Mel Frequency Cepstrum Coefficients (MFCCs) and the energy measured over 25 bands based on the Bark scale. However, it is necessary to remark that the patients had a mean H&Y stage of 2.3 ± 0.8, including patients in advanced stages. In the case of Czech language, where the mean stage is slightly lower 2.2 ± 0.5, the reported accuracy when using the unvoiced segments is reduced to 93.1%. In general, the lower the disease stage, the more challenging the diagnosis task is, since the speech impairment is less severe.

DDK utterances cover three types of syllables (/pa/, /ta/ and /ka/), composed of two regions with completely different characteristics (plosive consonant and vowel segments). Some previous studies in clinical assessment contexts have shown irregular articulation of velar stops by speakers with PD. In [19], the authors report imprecise velar contact in PD, after an investigation based on real-time dynamic magnetic resonance imaging. Hammer et al. [20] point out that velopharyngeal control may be impaired in PD. In [21], the authors indicate that syllable /ka/ is more impaired than /pa/ and /ta/. This allows to hypothesize that automatic diagnosis based on /k/ segments should provide better performance than in the case of /p/ or /t/ segments. However, to the best of the authors' knowledge, a validation of this hypothesis in an automatic detection scenario has not been reported. In [17] and [18], the authors extract features from different unvoiced segments (/p/, /t/ and /k/), but there is no distinction between the three plosives, that is, it is not considered which one would perform best in the development of a classification approach based on a single plosive. The focus only on one type of syllable saves computational effort since it avoids feature extraction tasks on the other types of syllables. Here three different classification experiments on the three plosive segments have been performed and the results are comparatively analyzed for an early-stage PD database.

Voice Onset Time (VOT) is defined as the duration of the part of the syllable (/pa/, /ta/ or /ka/, in this work) between initial burst and vowel onset. Since all the acoustic features are extracted from VOT segments, an accurate estimation of VOT is necessary. This accuracy must be also guaranteed in the case of dysarthric speech. A simple and intuitive algorithm is proposed, that provides accurate results both for healthy and dysarthric speakers.

Using the proposed segmentation algorithm, several features have been considered that may be sensitive to possible articulatory deficits due to dysarthria. Both temporal and spectral features based on VOT segments have been considered. The former group is based on time durations [6,18], whereas the latter one includes MFCC-based features [22] and spectral moments [23]. In a different context, but also related to plosive consonants, Lee et al. [24] demonstrate accurate discrimination of articulation place based on a combination of different types of features, including temporal and spectral ones, and in particular MFCC-based features.

Spectral moments measure the shape of the energy distribution in the spectrum. In the context of PD, spectral moments have been applied to the analysis of long-time average spectra from a standard reading sample in [25] and have been used to describe fricatives occurring in the word initial position of a reading passage in