Automatic Intelligibility Assessment of Parkinson's Disease with Diadochokinetic Exercises

L.F. Parra-Gallego[1], T. Arias-Vergara[1,2], J.C. Vásquez-Correa[1,2], N. Garcia-Ospina[1], J.R. Orozco-Arroyave[1,2], and E. Nöth[2]

[1] Faculty of engineering. Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia
[2] Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
Corresponding author: lfelipe.parra@udea.edu.co

Abstract. This paper presents preliminary results for the analysis of intelligibility in the speech of Parkinson's Disease (PD) patients. An automatic speech recognition system is used to compute the word error rate (WER), the Levenshtein distance, and the similitude based dynamic time warping. The corpus of the speech recognizer is formed with speech recordings of three Diadochokinetic speech tasks: /pa-ta-ka/, /pa-ka-ta/, and /pe-ta-ka/. The data consist of 50 PD patients and 50 Healthy Controls. According to the results, the recognition error is lower for the healthy speakers (WER=2.70%) respect to the PD patients (WER=11.3%).

Keywords: Parkinson's disease, Intelligibility, Speech processing, Automatic Speech Recognition

1 Introduction

Parkinson's disease (PD) is a neurological disorder caused by the degeneration of neurons within the brain, resulting in the progressive loss of dopamine in the substantia nigra of the midbrain [1]. The primary motor symptoms include resting tremor, slowness of movement , postural instability, rigidity and several dimensions of speech are affected including phonation, articulation, prosody, and intelligibility [2, 3]. These deficits reduce the communication ability of PD patients and make their normal interaction with other people difficult. The most outstanding disturbance is consonant imprecision [4]. In recent years, the research community has been developing systems that can diagnose and monitor PD in an objective and non-obtrusive way. As the speech production is one of the most complex processes in the brain, speech signals have been a focus in this research. Diadochokinetic (DDK) tasks are commonly included in speech assessment protocols as they have shown to be useful in the differential diagnosis of dysarthria and even neurological disease. The sequential motion rate

diadochokinetic exercises involve the rapid repetition of syllable sequences, e.g., /pa-ta-ka/ [5]. There are different studies that considers DDK exercises analysis to detect speech problems in PD patients. In [6] the authors present a portable system for the automatic recognition of the syllables /pa-ta-ka/. The proposed approach consists of a tablet and a headset to capture the speech signals. The system was trained using speech recordings from two group of speakers: patients with traumatic brain injuries and PD patients. The automatic recognition of /pa-ta-ka/ is performed in the mobile device using an Automatic Speech Recognition (ASR)-based system. Speech impairments are assessed using the syllable error rate (SER). Similarly, in [7] the authors presented an iOS application that integrated a speech recognition system for the analysis of intelligibility problems in PD patients. The corpus of the speech recognizer is formed with recordings of the syllables /pa/-/ta/-/ka/. The syllable error rate was computed in two group of speakers: healthy speakers and PD patients. According to the results, the SER was lower for the healthy group (SER=1.34%) respect to the patients (SER=1.67%), however, the analysis of intelligibility is limited to one feature. In [8], the authors modeled different articulatory deficits in PD patients in the rapid repetition of the syllables /pa-ta-ka/, and reported an accuracy of 88% discriminating between PD patients and Healthy Controls (HC). In this work is presented a ASR system trained with speech recordings of three DDK exercises: /pa-ta-ka/, /pa-ka-ta/, and /pe-ta-ka/. Additionally, three different features based on the word error rate (WER), the Levenshtein distance (LD), and the similitude based dynamic time warping (sDTW) are considered. Furthermore, articulation and intelligibility features are considered to train a classifier based on the support vector machine (SVM) approach. Our main hypothesis is that the patients has more difficulties to produce certain sound during speech, thus, it is possible to detect those problems considering the DDK analysis.

2 Materials and Methods

2.1 Data

The PC-GITA database is considered for this study [9]. The data contain speech utterances from 100 (50 PD, 50 HC) Colombian native speakers balanced in age and gender. Speech signals were captured in a soundproof booth with a sampling frequency of 44100 Hz and 16 bits resolution. All patients were diagnosed by an expert neurologist a labeled according to the MDS-UPDRS-III (Movement Disorder Society-Unified Parkinson's Disease Rating Scale). Additional information from the participants is shown in Table 1. Different DDK exercises were evaluated in this study to assess the intelligibility of the patients. DDK tasks are based on the rapid repetition of syllables that require the use of different speech articulators such as lips, larynx, or palate. These exercises increase motor and cognitive activities in the patients, which make them suitable to assess the neurological state and the speech impairments of the patients [10]. The DDK exercises used in this study consist of the rapid repetition of the syllables /pa-ta-ka/, /pa-ka-ta/, and /pe-ta-ka/. There is a special interest in the assessment

Table 1. Information of the participants from this study

	PD patients		HC subjects	
	male	**female**	**male**	**female**
Number of subjects	25	25	25	25
Age ($\mu \pm \sigma$)	61.3±11.4	60.7±7.3	60.5±11.6	61.4±7.0
Range of age	33-81	49-75	31-86	49-76
Duration of the disease ($\mu \pm \sigma$)	8.7±5.8	12.6±11.6	-	-
MDS-UDRS-III ($\mu \pm \sigma$)	37.8±22.1	37.6±14.1	-	-

of the phonemes included in these exercises to measure co-articulatory impairments in different muscles of the vocal tract. For instance, the phoneme /p/ is performed by pressing the lips together, the phoneme /t/ is produced by the interaction between the tongue and the bone obstruction behind the upper front tooth, while the phoneme /k/ is produced by the interaction between the soft palate and back of the tongue.

2.2 Feature extraction

Articulation–These features are designed to model changes in the position of the tongue, lips, velum, and other articulators involved in the speech production process. The articulation impairment of the patients are modeled by extracting features from the voiced to unvoiced segment (offset) transitions, the unvoiced to the voiced segment (onset) transitions, and voiced segments. The set of features include the energy content distributed in 22 Bark bands, 12 Mel-frequency cepstral coefficients (MFCC) with their first and second derivatives, and the first two formant frequencies (F1 and F2) with their first and second derivatives. The total number of descriptors corresponds to 122. Four functionals are also computed, obtaining a 488-dimensional feature-vector per utterance. The complete description of the articulation features is available in [11], and the code is freely available [3].

Intelligibility– Intelligibility is related to the capability of a person to be understood by another person or by a system. This speech dimension causes loss of the communication abilities of the patients, producing social isolation especially at advanced stages of the disease [12, 13]. Although extensively reported, impairments in speech intelligibility of PD patients have been analyzed through perceived intelligibility. We perform the intelligibility analysis based on an ASR, extracting several features to compare the recognized sentence with the real utterance produced by the speakers. The ASR is trained using the Kaldi framework [14]. The corpus of the ASR considers the words pataka, petaka, and pakata; and several variations found in the recordings, e.g., bataka, patakam, badaga, among others. The acoustic model is created using a hidden Markov model (HMM), where each state corresponds to a phoneme. The HMM was

[3] `https://github.com/jcvasquezc/DisVoice`

trained with MFCCs with their first and second derivatives. In addition, a language model based on tri-grams was considered. An scheme of the ASR system is shown in Figure 1. Specific ASRs are trained per DDK exercise. Then, we consider the transcrciptions obtained from the ASR to compute several features to assess the intelligibility impairments of the PD patients. The computed features include the WER, the LD, and the similitude based dynamic time warping sDTW, proposed previously [11].

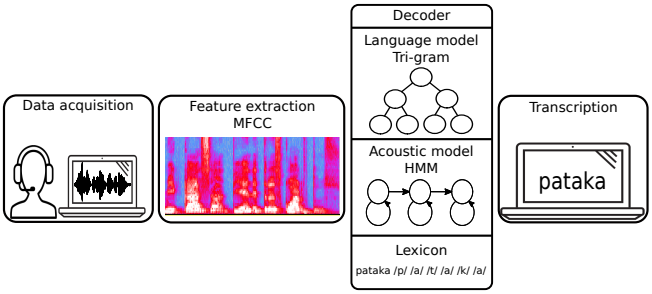

Fig. 1. Scheme of the automatic speech recognition system

The WER is the most common measure to evaluate the performance of ASR systems. We have the hypothesis that PD patients will produce more errors in the ASR than the HC subjects, producing larger WER. This feature is computed using the transcriptions generated by the ASR and the original transcription, according to Equation 1, where S is the number of substitutions, B is the number of deletions, I is the number of insertions, and N is the total of words in the original transcription.

$$\text{WER} = \frac{S + B + I}{N} \tag{1}$$

The LD is a metric used to measure the difference between text sequences. The LD betwwen to text strings is obtained as the minimum number of single-character edits required to change one string into the other one. The main difference between WER and LD is that WER is computed at word level, while LD is computed as character-level, which may provide a more accurate estimation of the ASR evaluation. Finally the sDTW was introduced to analyse differences between two time–series that differ in time and number of samples. The dynamic time warping performs a time–alignment between two text sequences, then the euclidean distance is computed between the predicted and the original transcriptions. The distance measure is transformed into a similarity score using Equation 2. If the original and recognized transcriptions are the same, the DTW_{dist} will be zero, and the sDTW will be 1.

$$\text{sDTW} = \frac{1}{1 + \text{DTW\_distance}} \tag{2}$$

2.3 Classification

The decision whether an speech utterance is from a PD patient or a HC subject is performed with a SVM with a Gaussian kernel with margin parameter C and bandwidth of the kernel γ. The parameters were optimized in a grid search with $10^{-3} < \mathrm{C} < 10^{4}$ and $10^{-6} < \gamma < 10^{3}$. A 10-fold speaker independence cross-validation strategy is performed, where eight folds are used to train the SVM, one to optmize the hyper-parameters, and one for test.

3 Experiments and results

3.1 Intelligibility analysis

The WER, LD, and sDTW were computed per speakers (PD and HC), using the ASR system described in Section 2.2. Table 2 shows the obtained results for both patients and healthy speakers. It can be observed that the intelligibility of the patients is more affected than the HC. In particular, the WER and LD were lower in the HC group (WER= 2.70%; LD= 11.3×10^{-2}), compared to the other speech tasks. For the case of the sDTW, the highest distance was obtained for the HC (sDWT= 79.1×10^{-2}), which means that there is higher similarity between the real and the estimated (ASR) transcriptions. These results support the hypothesis that the patients have more difficulties than the healthy speakers to produce rapid alternating sounds during the DDK exercises.

Table 2. Intelligibility features. ***WER****: Word Error Rate.* ***LD****: Levenshtein Distance.* ***sDTW****: Dynamic Time Warping.* ***HC****: Healthy Controls.* ***PD****: Parkinson's Diseases.*

	WER (%)		LD (10^{-2})		sDTW (10^{-2})	
Task	HC	PD	HC	PD	HC	PD
/pa-ka-ta/	4.90	11.7	1.71	9.92	76.3	69.3
/pa-ta-ka/	2.70	11.3	0.71	7.38	79.1	70.5
/pe-ta-ka/	2.90	7.10	1.26	2.97	78.8	73.8

3.2 Evaluation system

In order to evaluate the speech impairment of the patients, intelligibility and articulation features were extracted from each speech task and the combination of the three DDK exercises (Fusion). Additionally, both set of features were merged in order to test the suitability of the method to improved the detection speech problems in PD patients. Table 3 shows the performance of the SVM in terms of the accuracy (ACC), sensibility (SEN), specificity (SPE), and the area under the ROC curve (AUC). In biomedical applications the AUC is interpreted as follows: AUC < 0.70 indicates poor performance, $0.70 \leq$ AUC < 0.80 is fair,

Table 3. Performance of the SVM. ***ACC****: Accuracy.* ***SEN****: Sensibility.* ***SPE****: Specificity.* ***AUC****: Area Under the ROC Curve.* ***Art+Int****: Articulation and Intelligibility features.* ***Fusion****: Combination of the DDK speech tasks.*

Task	Feature	ACC (%)	SEN (%)	SPE (%)	AUC
/pa-ka-ta/	Articulation	75	81	73	0.78
	Intelligibility	56	53	64	0.54
	Art+Int	76	82	75	0.78
/pa-ta-ka/	Articulation	67	68	67	0.76
	Intelligibility	59	70	61	0.59
	Art+Int	73	74	74	0.78
/pe-ta-ka/	Articulation	68	72	69	0.74
	Intelligibility	61	68	63	0.61
	Art+Int	67	72	69	0.75
Fusion	Articulation	76	77	78	0.84
	Intelligibility	63	68	67	0.61
	Art+Int	72	72	75	0.82

$0.80 \leq \text{AUC} < 0.90$ is good, and $0.90 \leq \text{AUC} < 1$ is excellent [15]. Table 3 shows the results obtained for each speech tasks. The performance of the classifier is fair for the three DDK exercises (/pa-ka-ta/= 0.78; /pa-ta-ka/= 0.76; /pe-ta-ka/= 0.74). When the articulation features are extracted from the combined DDKs the performance improved from fair to good (AUC= 0.84). When the intelligibility features are considered for training, the performance of the SVM is poor in all cases (/pa-ka-ta/= 0.54; /pa-ta-ka/= 0.59; /pe-ta-ka/= 0.61; Fusion= 0.61). Figure 2 shows the ROC curves for each set of features on Fusion task. Figure 3 shows the distribution of the PD patients according to the WER and LD for the word /pa-ta-ka/. It can be observed that the PD patients are more dispersed than the healthy speakers, however, both group of speakers are very close to each other, which explain the low performance obtained for the intelligibility features. Additional to the articulation and intelligibility features, the SVM was trained considering the combination of both set of features. From Table 3 it can be observed that the accuracies obtained are not significantly different from those obtained when only the articulation features are considered for training. These results are expected, since the DDK exercises are designed to assess articulation problems of the patients. Additionally, the corpus used to train the ASR is limited to the words /pa-ta-ka/, /pa-ka-ta/, and /pe-ta-ka/, thus, different miss-pronunciations produced by the speakers are not included, e.g., /ba-ta-ka/,/pa-pa-ka/,.

4 Conclusions

In this work is presented a methodology to model the intelligibility problems in the speech of PD patients. The ASR was trained considering speech recordings of the rapid repetition of the words /pa-ta-ka/, /pa-ka-ta/, and /pe-ta-ka/. According to the results, the recognition error in the ASR is lower in the healthy group

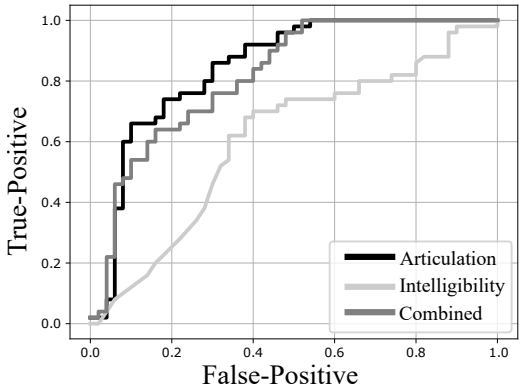

Fig. 2. ROC curve considering the Fusion of the three DDKs.

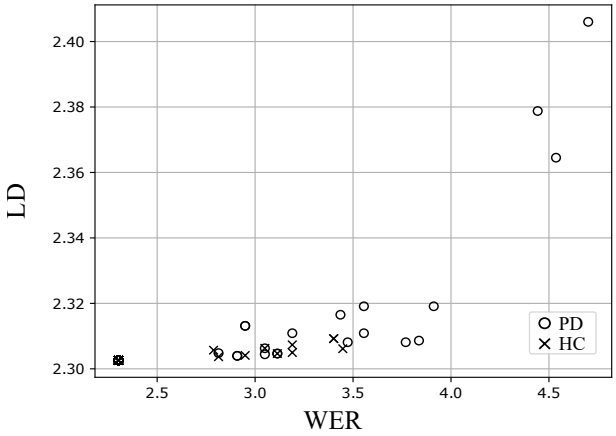

Fig. 3. Distribution of PD patients (O) and healthy speakers (X) considering two intelligibility features (LD vs. WER).

respect to the patients, which, indicates that the patients have more problems to produce certain sounds during speech. Additionally, articulation, intelligibility, and the combination of both set of features were considered to train a SVM. In this case, the best results were obtained when the articulation features were considered to train the classifier. Furthermore, the results improved when the three DDK exercises were combined to train the SVM. This can be explained considering that the DDKs are exercises designed to detect articulation problems in pathological speech. Also, note that the corpus used to detect intelligibility problems is limited, thus, future work should include sentences, read texts, and monologues in order to model the intelligibility problem of the patients.

Acknowledgments

This work was financed by CODI from University of Antioquia by the grant Number PRV16-2-01 and 2015–7683.Also the authors thanks to the Training Network on Automatic Processing of PAthological Speech (TAPAS) funded by the Horizon 2020 programme of the European Commission. Tomás Arias-Vergara is under grants of Convocatoria Doctorado Nacional-785 financed by COLCIENCIAS.

References

1. Hornykiewicz, O.: Biochemical aspects of Parkinson's disease. Neurology **51**(2 Suppl 2) (1998) S2–S9
2. Ho, A.K., et al.: Speech impairment in a large sample of patients with parkinson's disease. Behavioural neurology **11**(3) (1999) 131–137
3. Darley, F.L., et al.: Differential diagnostic patterns of dysarthria. Journal of Speech, Language, and Hearing Research **12**(2) (1969) 246–269
4. Logemann, J.A., et al.: Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of Parkinson patients. Journal of Speech and Hearing Disorders **43**(1) (1978) 47–57
5. Tjaden, K., et al.: Characteristics of diadochokinesis in multiple sclerosis and Parkinson's disease. Folia Phoniatrica et Logopaedica **55**(5) (2003) 241–259
6. Tao, F., et al.: A portable automatic pa-ta-ka syllable detection system to derive biomarkers for neurological disorders. In: INTERSPEECH. (2016) 362–366
7. Montaña, D., et al.: A diadochokinesis-based expert system considering articulatory features of plosive consonants for early detection of parkinson's disease. Computer methods and programs in biomedicine **154** (2018) 89–97
8. Novotnỳ, M., et al.: Automatic evaluation of articulatory disorders in Parkinson's disease. IEEE/ACM Trans. on Audio, Speech and Language Processing **22**(9) (2014) 1366–1378
9. Orozco-Arroyave, J.R., et al.: New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. In: Language Resources and Evaluation Conference, (LREC). (2014) 342–347
10. Yaruss, J.S., et al.: Evaluating rate, accuracy, and fluency of young children's diadochokinetic productions: a preliminary investigation. Journal of Fluency Disorders **27**(1) (2002) 65–86
11. Orozco-Arroyave, J.R., et al.: Neurospeech: An open-source software for Parkinson's speech analysis. Digital Signal Processing (In press) (2017)
12. De Letter, M., et al.: The effects of levodopa on word intelligibility in parkinson's disease. Journal of Communication Disorders **38**(3) (2005) 187–196
13. Miller, N., et al.: Prevalence and pattern of perceived intelligibility changes in parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry **78**(11) (2007) 1188–1190
14. Povey, D., et al.: The kaldi speech recognition toolkit. In: IEEE 2011 workshop on automatic speech recognition and understanding. (2011) 1–4
15. Swets, J.A., et al.: Psychological science can improve diagnostic decisions. Psychological science in the public interest **1**(1) (2000) 1–26