TOWARDS AN AUTOMATIC MONITORING OF THE NEUROLOGICAL STATE OF PARKINSON'S PATIENTS FROM SPEECH

J.R. Orozco-Arroyave[1,2], J.C. Vásquez-Correa[1], F. Hönig[2], J.D. Arias-Londoño[1], J.F. Vargas-Bonilla[1], S. Skodda[3], J. Rusz[4], and E. Nöth[2]

[1]Faculty of Engineering, Universidad de Antioquia, Medellín, Colombia
[2]Pattern Recognition Lab, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany
[3]Department of Neurology, Knappschaftskrankenhaus, Ruhr-University Bochum, Germany
[4]Faculty of Electrical Engineering, Czech Technical University, Prague, Czech Republic

Corresponding author: rafael.orozco@i5.informatik.uni-erlangen.de

ABSTRACT

The suitability of articulation measures and speech intelligibility is evaluated to estimate the neurological state of patients with Parkinson's disease (PD). A set of measures recently introduced to model the articulatory capability of PD patients is considered. Additionally, the speech intelligibility in terms of the word accuracy obtained from the Google® speech recognizer is included. Recordings of patients in three different languages are considered: Spanish, German, and Czech. Additionally, the proposed approach is tested on data recently used in the INTERSPEECH 2015 Computational Paralinguistics Challenge. According to the results, it is possible to estimate the neurological state of PD patients from speech with a Spearman's correlation of up to 0.72 with respect to the evaluations performed by neurologist experts.

Index Terms— Parkinson's disease, articulation, speech, intelligibility, UPDRS

1. INTRODUCTION

Parkinson's disease (PD) is a neurological disorder that alters the function of the basal ganglia in the midbrain, affecting motor and non-motor abilities of patients [1]. Motor deficits include bradykinesia, rigidity, postural instability, and resting tremor, and non-motor impairments include negative effects on the sensory system, sleep, cognition, and emotions [2]. The standard method to evaluate and rate the neurological state of Parkinson's patients is based on the unified Parkinson's disease rating scale (UPDRS) [3]. This scale includes a subset of items to assess different motor activities such as movement of the legs and arms, finger tapping, and others. This evaluation only includes one item about speech; however, it has already been observed in the literature that the loss of control in the motor activity causes different voice and speech impairments in the majority of PD patients. The most common symptoms in the speech of PD patients include reduced loudness, monopitch, monoloudness, reduced stress, breathy and hoarse voice quality, and imprecise articulation [2]. The objective evaluation of speech in PD patients is a topic that deserves the attention of the research community. Several studies in the literature have described the speech impairments of PD patients in terms of phonation, articulation, and prosody [4, 5, 6, 7]. Phonation is related with the capability of the speaker to make the vocal folds vibrate to produce vocal sounds, articulation is related with the modification of the position, stress, and shape of several limbs and muscles to produce the speech, and prosody reflects variation of loudness, pitch, and timing to produce natural speech [8]. Along with these three aspects of speech, intelligibility is also deteriorated in PD patients causing loss of communication abilities and social isolation specially at advanced stages of the disease [9]. There are studies reporting a close relation of speech degradation and PD symptoms [10]. This relation can be explained by the fact that there are many muscles and limbs involved in voiced production; therefore speech is a promising source of information for assessing the neurological state of PD patients. Indeed the research community has been highly interested in studying the speech of people with PD in order to develop computer aided systems to assist the screening and monitoring of PD patients from speech. In this paper we are focused on studying the correlation between articulation models and intelligibility with the neurological state of PD patients according to the motor section of the UPDRS, i.e., UPDRS-III.

Several studies in the literature have analyzed and described characteristics of PD speech with the aim of discriminating PD patients and healthy controls (HC) or to evaluate the correlation among different speech characteristics and the severity of the disease. For instance in [5] the authors studied possible correlations among vowel articulation, global motor performance, and the stage of the disease. The authors considered a total of 68 patients and 32 HC and performed several statistical tests. They concluded that the vowel articulation in-

dex (VAI) is significantly reduced in PD speakers; however, no significant correlations between vowel articulation and the extent of the disease were found. In [11] different articulatory deficits observed in PD speech are modeled. The study considered a total of 46 Czech native speakers, 24 of them with PD. Each speaker performed the rapid repetition of the syllables /pa-ta-ka/. The authors calculated 13 features to describe six different articulatory aspects of speech: vowel quality, coordination of laryngeal and supra-laryngeal activity, precision of consonant articulation, tongue movement, occlusion weakening, and speech timing. The authors reported a classification result of 88% in separating speech signals of PD patients and HC. The results confirm previous observations where the imprecise articulation is reported as the most predominant characteristic of PD-related dysarthria. With respect to the problems of intelligibility in the speech of PD patients, several clinicians have observed those impairments over several years in a large sample of patients [2]. After these observations were reported, several studies have addressed perceptual analyses of speech intelligibility in PD patients. In [12] the authors analyze possible correlations between the perceived intelligibility (in terms of correctly pronounced words) and the severity of the disease. According to their findings, there is no correlation between the perceived intelligibility and the neurological state of the patients. A similar result is reported in [13] where a total of 125 PD patients and the same number of age matched HC are evaluated. The authors investigate the correlation between their speech intelligibility (also in terms of correctly pronounced words) perceived by listeners unfamiliar with dysarthric speech and the disease severity. The authors conclude that although speech intelligibility is significantly reduced in people with PD, no correlations were found between perceived intelligibility and disease severity. Although extensively reported, the impairments in speech intelligibility of PD patients only have been analyzed subjectively through perceived intelligibility, which is non-objective and time-consuming; it seems worthwhile to develop methods to automatically assess speech intelligibility in people with PD. A straightforward way to automatically evaluate speech intelligibility consists of computing the word accuracy (WA) of a speech recognizer with respect to a known text. This approach was proposed and successfully tested in [14] considering the speech of children with cleft lip and palate.

In this paper we evaluate the suitability of the articulatory model introduced in [15] and the speech intelligibility in terms of WA to estimate the neurological state of PD patients according to the motor section of the UPDRS scale. Three groups of patients are considered, Spanish, German, and Czech native speakers. Besides the experiments with different languages, the sets of speakers (train, development, and test) introduced in the recent version of the INTERSPEECH 2015 Computational Paralinguistics Challenge (ComParE) [16] are used to evaluate the suitability of the proposed models on recordings of a publicly accessible dataset.

2. PATIENTS AND SPEECH TASKS

Spanish: The set of 50 patients included in the PC-GITA database [17] is considered here (25 male and 25 female). The age of the speakers ranges from 33 to 77 years (mean 61.1 $\pm$ 9.6). The participants were recorded in a sound-proof booth at Clínica Noel of Medellín in Colombia, using a dynamic omni-directional microphone and a professional audio card. The speakers were asked to pronounce a total of 21 isolated words, 6 sentences, one text with 36 words, and a monologue (average duration: 45 $\pm$ 23 seconds). The sampling frequency of the recordings was 44.1 kHz with 16-bit resolution. All of the patients were recorded in ON-state, i.e., no more than three hours after their morning medication. The patients were labeled by a neurologist expert; the mean value of their evaluation according to the MDS-UPDRS-III scale (max value 132) is 36.7 $\pm$ 18.7; the average time post PD diagnosis was 10.7 $\pm$ 9.2 years.
German: A total of 88 native German speakers (47 male and 41 female) were recorded in a quiet room at the Knappschaftskrankenhaus of Bochum in Germany [5]. The participants performed four speech tasks: 6 isolated words, 5 sentences, one text with 81 words, and a monologue (average duration: 33 $\pm$ 8 seconds). The age of the speakers ranged between 42 and 84 years (mean 66.4 $\pm$ 8.9). The sampling frequency of the recordings was 16 kHz with 16-bit resolution. As in the Spanish data, the patients were in ON-state during the recording session; the mean value of their neurological evaluation according to the UPDRS-III (max value 108) scale is 22.7 $\pm$ 10.9; the average time post PD diagnosis was 7.1 $\pm$ 5.8 years.
Czech: A total of 20 native Czech speakers were considered (all of them male). The patients were newly diagnosed with PD, and none of them had been medicated before or during the recording session. The participants were recorded in the General University Hospital in Prague, Czech Republic, and the speech tasks considered here include 7 isolated words, 3 sentences, one text with 80 words, and a monologue (average duration: 115 $\pm$ 56 seconds). The recordings were sampled at 48 kHz with 16-bit resolution. The age of the speakers ranges between 34 and 83 years (mean 61 $\pm$ 12); the mean value of their neurological evaluation according to the UPDRS-III scale is 17.9 $\pm$ 7.4. Further details of this database can be found in [18].

3. SPEECH MODELING AND NEUROLOGICAL STATE ESTIMATION

3.1. Articulation model

In [15] we recently introduced a new method to model difficulties observed in PD patients to start/stop the vibration of vocal folds. The method consists of detecting the onset (from unvoiced to voiced) and offset (from voiced to unvoiced) tran-

sitions in the speech recording. In the borders between voiced and unvoiced sounds a frame of 40 ms to each side (80 ms in total) is taken. Cepstral mean subtraction is performed over the recordings excluding initial and final silences. The energy content of those transitions is calculated in terms of 12 mel-frequency cepstral coefficients (MFCCs) and 25 Bark band energies (BBE). Four functionals (mean value, standard deviation, kurtosis, and skewness) are calculated per speech recording, forming 148-dimensional feature vectors to represent each voice sample. As the method showed to be highly accurate and robust to discriminate between people with PD and HC in three languages, here we want to evaluate its suitability to find correlations with the neurological state of the patients.

3.2. Speech intelligibility

The speech impairments observed in PD patients generate loss of speech intelligibility. As already introduced in [14], WA can be a good descriptor of speech intelligibility in people with pathological voice. In [14] the speech recognizer was trained with speech of German native speakers and a unigram language model was used to increase the influence of the acoustic model on the recognizer. In this study we want to test the suitability of an off-the-shelf speech recognizer, thus we decided to use the well known and publicly accessible ASR system developed by Google Inc®. In particular, the speech API v2 is used (https://www.google.com/intl/es/chrome/demos/speech.html). It is a cloud-based ASR system that can only be used off-the-shelf, i.e., we did not have access to the parameters of the platform, thus we only made the requests and analyzed the results. The recordings are encoded and converted to the Free Lossless Audio Codec (FLAC) format with a sampling frequency of 16 kHz. The resulting file is sent to the Google® servers using the HTTP protocol. The only parameter we could adjust was the language. We choose "es-CO" for the Colombian Spanish, "de-DE" for German, and "cs" for Czech. The Google® speech API converts the input signal into a word chain. WA was computed for the set of words together. For sentences and texts it was calculated for each item.

3.3. Neurological state estimation

The disease severity is estimated using a linear support vector regressor (SVR) with ε-insensitive loss function. The parameters of the regressor C and ε are optimized in a grid search with $C \in [10^{-4}, 10^{-3}, \ldots, 10]$ and $\varepsilon \in [1, 10, 20, 30]$. The system is evaluated using Spearman's (ρ) correlation between the predicted values and the UPDRS labels.

4. EXPERIMENTS AND RESULTS

4.1. Experiments with cross-validation

The articulation features proposed in [15] and the WA extracted from the speech tasks are considered for the regression model. The optimization of the SVR parameters in the experiments with Spanish and German data is performed following a 10-fold cross-validation (CV) strategy. For the case of the Czech data the optimization is performed following a leave-one-speaker-out CV (LOSO-CV) strategy. The speaker independence condition is obeyed in all the experiments and in all cases the optimization criterion was the highest correlation in test which can lead to slightly optimistic estimates but the bias is low due to the low number of parameters.

Table 1. ρ values obtained with the articulation models and WA. "all": corresponds to the fusion of the feature vectors of all speech tasks evaluated with the intelligibility measures.

	words	sentences	read text	monologue
Energy of the onset transitions				
Spanish	0.44	0.49	0.44	0.56
German	0.36	0.28	0.41	0.55
Czech	0.29	0.25	0.21	0.23
Energy of the offset transitions				
Spanish	0.46	0.46	0.53	0.74
German	0.37	0.24	0.36	0.31
Czech	0.24	0.18	0.35	0.35
Early fusion of onset and offset transitions				
Spanish	0.47	0.60	0.40	0.72
German	0.39	0.24	0.38	0.40
Czech	0.30	0.26	0.28	0.15
Intelligibility measures (WA)				
	words	sentences	read text	all
Spanish	0.39	0.20	0.07	0.49
German	0.14	0.18	0.19	0.31
Czech	0.22	0.16	0.15	0.25

The Spanish data exhibit the highest correlations which can be explained by the fact that these patients have been suffering from Parkinson's disease for more years than the others, also their average neurological score and its variability are higher than in the German and Czech patients. We did not use the monologues to compute WA because it is not realistic to transcribe those recordings during a medical evaluation. The highest correlations with WA are obtained when all individual intelligibility measures are fused. We are aware of the fact that the results indicated in Table 1 are slightly optimistic due to the optimization criterion. However, this experiment gives us a good idea of how far we can go with the models. In order to evaluate the suitability of the proposed approach in a publicly accessible dataset, the method is tested with the data recently introduced in the INTERSPEECH 2015 Computational Paralinguistics Challenge (ComParE) [16].

4.2. Experiments with the ComParE 2015 subsets

For the ComParE 2015 challenge the 50 patients of the PC-GITA database [17] were partitioned into train and development subsets. The test set was formed from 11 additional patients that were recorded in non-controlled noise conditions. The MDS-UPDRS-III labels of the speakers in the test subset were assigned by the same neurologist of the other 50 patients. A total 42 speech tasks were considered for the competition [16], but for this paper we are only considering 29 of them (21 words, 6 sentences, the read text, and the monologue). In the challenge it was not allowed to manually group speech tasks performed by the same person. The winners of the challenge grouped the tasks automatically and showed that grouping speech tasks leads to better results [19]. As we consider that in real conditions the neurologist has access to the patient's identity, we decided to manually group the 29 speech tasks per speaker to perform the experiments in the same way as in Section 4.1. Table 2 includes the ρ values obtained on the development and test subsets with the same approaches of the previous section. The C and ε values optimized on development are also provided.

Table 2. ρ values obtained with the articulation models and WA on the ComParE 2015 subsets. C and ε are the parameters of the SVR, "dev": development, "sent": sentences, "text": read text, "monol": monologue, "all": fusion of the feature vectors of all speech tasks modeled with WA.

Energy of the onset transitions				
	words	sent	text	monol
$C;\varepsilon$	$10;10^{-1}$	$10;10^{-2}$	$10;10^{-2}$	$20;10^{-4}$
train/dev	0.28	0.44	0.48	0.42
train/test	0.33	0.63	0.23	0.39
Energy of the offset transitions				
$C;\varepsilon$	$10;10^{-4}$	$10;10^{-1}$	$10;10$	$10;10^{-4}$
train/dev	0.24	0.23	0.23	0.62
train/test	0.22	0.52	-0.24	0.12
Early fusion of onset and offset transitions				
$C;\varepsilon$	$10;10^{-2}$	$10;10^{-4}$	$15;10^{-3}$	$20;10^{-4}$
train/dev	0.34	0.39	0.29	0.51
train/test	0.36	0.53	0.15	0.43
Intelligibility measures (WA)				
	words	sent	text	all
$C;\varepsilon$	$5;10^{-1}$	$5;10^{-1}$	$5;10^{-1}$	$5;10^{-4}$
train/dev	0.40	0.22	0.28	0.44
train/test	0.56	0.31	0.51	0.69

The most stable results are obtained with the energies of the onset transitions and with the WA measures. Also the highest ρ values are obtained with these two approaches, i.e., 0.63 with onsets of sentences and 0.69 with WA measured over all of the speech tasks.

4.3. Experiments with ranks of the predictions

Bearing in mind that the real-world application of this technology is the estimation of the neurological state of PD patients to monitor their disease progress, we decided to perform a third experiment considering the rank assigned by the proposed approaches as features, to correlate them with the actual MDS-UPDRS-III labels. The ranks of the predicted values obtained with the energy of the onset transitions and with the WA on each speech task were considered as independent features. The average rank obtained on each speech task with each characterization approach is computed and correlated with the MDS-UPDRS-III scores; no further optimization was performed regarding the SVR parameters. The results obtained with WA on the development and test subsets are 0.49 and 0.59, respectively; the results obtained with the onsets on development and test are 0.50 and 0.69, respectively. When the ranks obtained with all the speech tasks and the two approaches are combined, the results on development and test are 0.51 and 0.72, respectively, indicating that these two approaches are slightly complementary.

5. DISCUSSION AND CONCLUSIONS

We showed that the energy content of the onset transitions are suitable for classification of people with PD and HC [15] as well as for predicting the neurological state of the patients. In general, the correlation values obtained with the Spanish data are higher than with the other languages. A likely explanation is that this group of speakers has higher UPDRS values and higher variability, which makes it easier to predict the neurological state. According to the results, an off-the-shelf speech recognizer can be the basis for an intelligibility-based test which could be suitable for telemonitoring PD patients. The results obtained in this paper are slightly better than those reported by the winners of the ComParE 2015. We are aware of the fact that we are using additional information that was not allowed to be used in the competition, namely the identity of the speakers; however, we wanted to test the presented methods in real-world conditions, where the neurologist has access to the patients' ID. This result indicates that the proposed approach, considering the articulation model and the WA, is suitable to monitor the disease progression. We are currently collecting data to study the suitability of our models to perform longitudinal analyses.

Acknowledgments

J.R. Orozco-Arroyave is under grants of "Convocatoria Nº 528, generación del bicentenario COLCIENCIAS 2011". This work was also financed by COLCIENCIAS project Nº 111556933858. The research leading to these results has also received funding from the Hessen Agentur, grant numbers 397/13-36 (ASSIST 1) and 463/15-05 (ASSIST 2).

6. REFERENCES

[1] O. Hornykiewicz, "Biochemical aspects of Parkinson's disease," *Neurology*, vol. 51, no. 2, pp. S2–S9, 1998.

[2] J.A. Logemann, H.B. Fisher, B. Boshes, and E.R. Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients," *Journal of Speech and Hearing Disorders*, vol. 43, pp. 47–57, 1978.

[3] C. G. Goetz et al., "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129–2170, 2008.

[4] J. R. Orozco-Arroyave, E. A. Belalcázar-Bolaños, J. D. Arias-Londoño, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, F. Hönig, K. Daqrouq, and E. Nöth, "Chracterization methods for the detection of multiple voice disorders: neurological, functional, and organic diseases," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1820–1828, 2015.

[5] S. Skodda, W. Visser, and U. Schlegel, "Vowel articulation in parkinson's diease," *Journal of Voice*, vol. 25, no. 4, pp. 467–472, 2011, Erratum in Journal of Voice. 2012 Mar;25(2):267-8.

[6] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated parkinson's disease," *Journal of the Acoustical Society of America*, vol. 129, no. 1, pp. 350–367, 2011.

[7] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, and E. Nöth, "Automatic Detection of Parkinson's Disease in Running Speech Spoken in Three Dierent Languages," *Journal of the Acoustical Society of America*, vol. In press, pp. 1–19, 2016.

[8] A. M. Goberman and E.W. Lawrence, "Acoustic analysis of clear versus conversational speech in individuals with Parkinson disease," *Journal of Communication Disorders*, vol. 38, no. 3, pp. 215–230, 2005.

[9] National Parkionson Foundation, *Guidelines for speech-language therapy in Parkinson's disease*, Nederlandse Vereniging voor Lodopedie en Foniatrie, Nijmegen, 2010.

[10] S. Pinto, C. Ozsancak, E. Tripoliti, S. Thobois, P. Limousin-dowsey, and P. Auzou, "Review treatments for dysarthria in Parkinson's disease," *The Lancet Neurology*, vol. 3, no. September, pp. 547–556, 2004.

[11] M. Novotný, J. Rusz, R. Čmejla, and E. Růžička, "Automatic evaluation of articulatory aisorders in Parkinsons disease," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 9, pp. 1366–1378, 2014.

[12] M. De Letter, P. Santens, and J. Van Borsel, "The effects of levodopa on word intelligibility in Parkinson's disease," *Journal of Communication Disorders*, vol. 38, no. 3, pp. 187–196, 2005.

[13] Nick Miller, Liesl Allcock, Diana Jones, Emma Noble, Anthony J Hildreth, and David J Burn, "Prevalence and pattern of perceived intelligibility changes in Parkinson's disease," *Journal of Neurology, Neurosurgery, and Psychiatry*, vol. 78, no. 11, pp. 1188–1190, 2007.

[14] M. Schuster, A. Maier, T. Haderlein, E. Nkenke, U. Wohlleben, F. Rosanowski, U. Eysholdt, and E. Nöth, "Evaluation of speech intelligibility for children with cleft lip and palate by means of automatic speech recognition," *International Journal of Pediatric Otorhinolaryngology*, vol. 70, pp. 1741–1747, 2006.

[15] J. R. Orozco-Arroyave, F. Hönig, J. D. Arias-Londo no, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, and E. Nöth, "Voiced/unvoiced transitions in speech as a potential bio-marker to detect Parkinson's disease," in *Proceedings of the 16th INTERSPEECH*, Dresden, Germany, 2015, pp. 95–99.

[16] B. Schuller, S. Steidl, A. Batliner, S. Hantke, F. Hönig, J. R. Orozco-Arroyave, E. Nöth, Y. Zhang, and F. Weninger, "The INTERSPEECH 2015 Computational Paralinguistics Challenge: nativeness, Parkinson's & eating condition," in *Proceedings of the 16th INTERSPEECH*, Dresden, Germany, 2015, pp. 478–482.

[17] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. González-Rátiva, and E. Nöth, "New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proceedings of the 9th LREC*, Reykjavik, Iceland, 2014, pp. 342–347.

[18] J. Rusz, R. Cmejla, T. Tykalova, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka, "Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task," *The Journal of the Acoustical Society of America*, vol. 134, no. 3, pp. 2171–2181, 2013.

[19] T. Grósz, R. Busa-Fekete, G. Gosztolya, and L. Tóth, "Assessing the degree of nativeness and Parkinson's condition using Gaussian processes and deep rectifier neural networks," in *Proceedings of the 16th INTERSPEECH*, Dresden, Germany, 2015, pp. 919–923.