Gender–dependent GMM–UBM for tracking Parkinson's disease progression from speech

T. Arias-Vergara[1], J.C. Vasquez-Correa[1], J.R. Orozco-Arroyave[1,2], J.F. Vargas-Bonilla[1], T. Haderlein[2], 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
Email: tomas.arias@udea.edu.co

Abstract

Parkinson's disease (PD) severity is evaluated by neurologist experts by means of several tests. One of them is the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS–UPDRS). The main hypothesis is that changes in the speech of PD patients reflect changes in their neurological state. In this study we use the Gaussian Mixture Model–Universal Background Model approach to track the disease progression per speaker. Speech recordings from 62 PD patients were captured from 2012 to 2015 in three recording sessions. The validation of the models is performed with recordings of 7 patients (3 male and 4 female). The models were trained using speech recordings from male and female patients separately. According to the results, it is possible to track the disease progression with a Pearson's correlation of up to 0.88 for males and 0.53 for females.

Keywords: Parkinson's disease, Disease progression, User modeling, Gaussian Mixture Model, MDS–UPDRS–III.

1 Introduction

Parkinson's disease (PD) is a neurodegenerative disorder characterized by the progressive loss of dopaminergic neurons in the midbrain [1]. PD symptoms include tremor, rigidity, slowed movements and speech impairments [2]. The severity of PD varies among the patients, i.e., the progression of the disease and the symptoms experienced for some patients varies from person to person. The neurological state of PD patients is evaluated with the Movement Disorder Society–Unified Parkinson's Disease Rating Scale (MDS–UPDRS). This is a perceptual scale used to assess the motor and non-motor capabilities of PD patients [3]. The complete scale is divided into four parts. In this study we consider only the third part of the MDS–UPDRS (MDS–UPDRS–III) because it evaluates the motor abilities of the patients. The scale includes only one item to evaluate speech impairments; however, speech disorders affect the majority of PD patients [4]. Since PD severity is evaluated by neurologist experts according to their own clinical criterion, the inter–expert–variability of the test could be high. Thus, it is important to develop computer aided systems to support the clinical diagnosis and to assess the disease progression objectively. There are studies focused on monitoring the disease progression from speech over the time. In [5] voice signals captured in two different recordings sessions are considered. The speech is perceptually evaluated considering four terms: voice, articulation, prosody, and fluency. The authors also correlated the perceptual speech scores with the speech item of the UPDRS. The prediction of the disease severity according to the UPDRS is presented in [6]. In that study, speech recordings were collected once per week during six months. The authors modeled speech extracting several acoustic measures. The prediction of the UPDRS score was possible using a Classification And Regression Trees (CARTs) approach. The authors do not guarantee speaker independence in the validation process. Other work has focused on the assessment of the disease severity predicting the UPDRS/MDS–UPDRS motor score [7–9]. In these studies the speech recordings are captured once per patient.

In this paper we propose a methodology to track the PD progression from speech signals collected in three recording sessions. The disease progression is assessed individually following a user–modeling approach. A separate analysis of female and male patients is also considered with the aim to analyze the gender–dependence of the proposed approach. In this study, a Gaussian Mixture Model adapted from a Universal Background Model (GMM–UBM) is considered for modeling the progression of the disease. A subset of the patients recorded was considered for the adaptation process.

The rest of the paper is organized as follows: Section 2 contains the data description and methods for modeling. Section 3 contains the results. Section 4 describes the conclusion derived from this study.

2 Methods and materials

Speech recordings from 62 patients were collected in three recording sessions within a period of three years. A subset of seven patients participated in the three recording sessions, and they are considered for tracking the disease progression using individually-adapted models. One patient is selected to be modeled. The remaining group of speakers used to train the UBM is selected depending on the extracted patient (male or female). We consider the UBM as the baseline to assess the disease progression according to its distance to the adapted model. Three different UBMs are trained for each group of speakers (males and females): (1) with recordings of the PD patients, (2) with speech of healthy speakers, and (3) with both groups of speakers. The models are built with several features extracted from the voiced (v) and unvoiced (uv) segments of the speech signals. The final model per speaker consists of three single models, one per recording session. The disease progression is evaluated calculating the distance between the background model and the speaker model. Finally, the correlation between the distance measures estimated for each recording session and the three neurological scores is calculated. The process is summarized in Figure 1 and further details are provided in the following subsections.

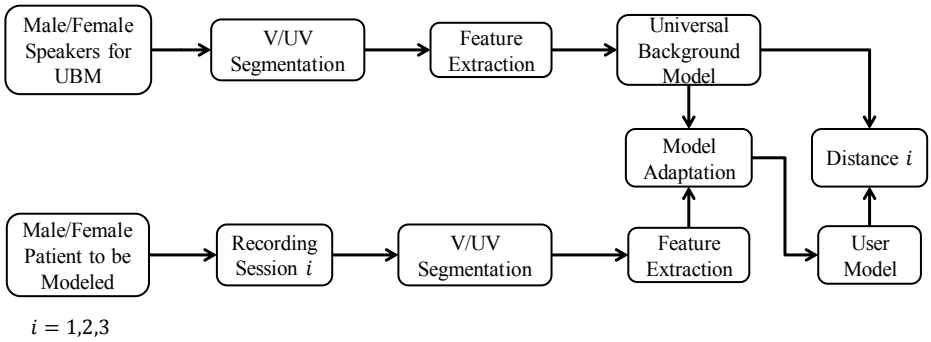

Figure 1: Proposed methodology

2.1 Data description

Speech recordings from 34 males and 28 females were collected from 2012 to 2015. A professional audio setting was used for the first two sessions, and the third session was recorded with the device presented in [10]. All of the patients from the three sessions were diagnosed by a neurologist expert according to the MDS–UPDRS–III [3]. Only 3 of the 34 male patients (MP) participated in the three recording sessions. For the females only 4 patients are present in all sessions. A Healthy Control (HC) group is also considered. Speech recordings from 31 males and 31 females were captured. None of the participants in the HC group has a history of symptoms related to PD or any other kind of movement disorder. Each subject in the HC group was recorded one time. All the participants of the tests followed the set of speech tasks presented in [11]. In this study, only the reading of a phonetically balanced text with 36 words was considered.

Table 1: *Distribution of patients recorded in all sessions.* ***MP*** **i** *($i \in \{1,2,3\}$): Male Patients.* ***FP*** **i** *($i \in \{1,2,3,4\}$): Female Patients.* ***Session*** **i** *($i \in \{1,2,3\}$): MDS–UPDRS–III scores obtained on each recording session.*

Patient	Age	Session 1	Session 2	Session 3
MP1	64	28	19	13
MP2	59	6	8	24
MP3	68	14	25	7
FP1	55	29	26	26
FP2	51	38	49	44
FP3	57	41	35	33
FP4	56	43	10	19

2.2 Voiced/unvoiced characterization

Voiced and unvoiced segments are extracted and grouped separately to characterize the read text task. Hamming windowing with 20 ms length and time shift of 10 ms is applied. The set of features extracted from the voiced frames include the jitter, the shimmer, and 12 Mel–Frequency Cepstral Coefficients (MFCCs). For unvoiced frames the set of features include 12 MFCCs and the log energy of the signal distributed in 25 Bark bands. To compensate for the channel acoustic condition, cepstral mean subtraction is applied.

2.3 Gaussian Mixture Model–Universal Background Model

We assess the disease progression from speech by modeling the speakers from Table 1. User models are obtained using GMM–UBMs. The GMM approach allows to represent the distribution of arbitrary probabilistic densities. For this reason, in speech processing such an approach is used to represent the feature vectors from one speaker. When several speakers are considered for training, the model is called UBM. GMMs are defined as parametric probabilistic models represented as a linear combination of M Gaussian densities. For a D-dimensional feature vector $\boldsymbol{x}$, a GMM is defined as

$$p(\boldsymbol{x}|\lambda) = \sum_{i=1}^{M} w_i p_i(\boldsymbol{x}) \quad (1)$$

The GMM is parametrized by the mixture weights ω_i, a D−×1 mean vector $\boldsymbol{\mu_i}$, and a D×D covariance matrix $\boldsymbol{\Sigma_i}$ [12]. The parameters of the Gaussian mixtures can be denoted as $\lambda = (\omega_i, \boldsymbol{\mu_i}, \boldsymbol{\Sigma_i})$ and the Gaussian densities as

$$p_i(\boldsymbol{x}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma_i}|^{1/2}} exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_i})^T \boldsymbol{\Sigma_i}^{-1}(\boldsymbol{x}-\boldsymbol{\mu_i})\} \quad (2)$$

In this study, the UBM was trained with a number of Gaussians that ranges from 2 to 1024 in 2^n steps with $n \in \{1,2,3...,10\}$. One patient is selected to be modeled. The remaining speakers are used to train the UBM. Depending on the selected patient, the UBM is trained including only male or female speakers. Next, the parameters of the UBM are updated and adapted with the feature vector of the selected patient. The adaptation is performed using the Maximum A Posteriori (MAP) rule. Then, we compute the distance between the UBM and the adapted model. Three adaptations (one per recording session) are performed for each patient. The resulting user model contains 3 distance values.

2.4 Distance computation

The Bhattacharyya distance measures the dissimilarity between two probabilistic distributions. We use equation 3 to calculate the distance between the UBM $(\widehat{\boldsymbol{\omega}}, \widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}})$ and the adapted models $(\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ [13].

$$\mathrm{d_{Bha}} = \frac{1}{8}\sum_{i=1}^{M}\left\{(\widehat{\boldsymbol{\mu_i}} - \boldsymbol{\mu_i})^T\left[\frac{\widehat{\boldsymbol{\Sigma_i}}+\boldsymbol{\Sigma_i}}{2}\right](\widehat{\boldsymbol{\mu_i}} - \boldsymbol{\mu_i})\right\}^{-1} + \frac{1}{2}\sum_{i=1}^{M}\left[\ln\frac{\frac{\widehat{\boldsymbol{\Sigma_i}}+\boldsymbol{\Sigma_i}}{2}}{\sqrt{|\widehat{\boldsymbol{\Sigma_i}}||\boldsymbol{\Sigma_i}|}}\right] - \omega_{Bha} \quad (3)$$

Here $\omega_{Bha} = \frac{1}{2}\sum_{i=1}^{M}\ln(\widehat{\omega_i}\omega_i)$ is the mixture weight measure.

Table 2: *Pearson's correlation between the predicted scores and the real MDS–UPDRS–III.* ***Seg****: Voiced/unvoiced segments.* ***MPi/FPj*** *($i \in \{1,2,3\}$), ($j \in \{1,2,3,4\}$): Pearson's correlation between the predicted scores and the real MDS–UPDRS–III score per patient.* ***Avg****: Average value of the correlations per patient.*

		Male Speakers				Female Speakers				
Training set	**Seg**	**MP1**	**MP2**	**MP3**	**Avg**	**FP1**	**FP2**	**FP3**	**FP4**	**Avg**
PD	V	0,99	0,99	0,71	0,90	-0,50	0,99	0,97	-0,25	0,30
	UV	-0,33	0,52	0,98	0,39	-0,36	0,98	0,45	0,56	0,40
HC	V	0,99	0,96	0,76	0,90	-0,50	0,99	0,84	0,96	0,57
	UV	-0,67	0,51	0,77	0,20	0,01	0,24	-0,28	0,07	0,01
PDHC	V	0,99	0,99	0,76	0,91	-0,58	0,99	-0,10	0,34	0,16
	UV	0,65	0,53	0,99	0,72	-0,20	0,87	0,88	0,18	0,43

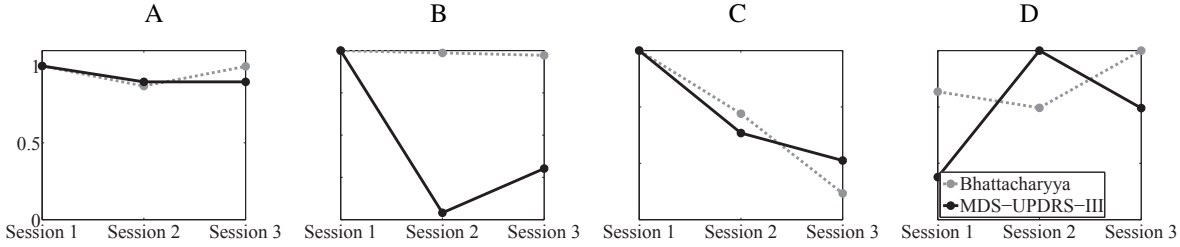

Figure 2: *Normalized scores for the female patients considering the Bhattacharyya distance (Grey dotted line) and the MDS-UPDRS-III labels (Black solid line) using features from voiced segments. (A) FP1, (B) FP2, (C) FP3, (D) FP4*

2.5 Regression model

The disease severity according to the MDS–UPDRS–III is estimated using a linear Support Vector Regressor (SVR). This method is applied to validate that the features implemented are suitable to assess the neurological state of a patient. The prediction ($\widehat{y}$) is measured with the ε–insensitive loss function $L(y,\widehat{y})$, which ensures the existence of the global minimum, and it is computed with Equation 4.

$$L(y,\widehat{y}) = \begin{cases} 0 & \text{if } |y-\widehat{y}| \leq \varepsilon \\ |y-\widehat{y}| - \varepsilon & \text{otherwise} \end{cases} \quad (4)$$

The parameters of the regressor C and ε, are optimized in the training set in a grid search with $C \in \{10^{-4}, 10^{-3}, 10^{-2}, ..., 100\}$ and $\varepsilon \in \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 20\}$. The performance is evaluated using the Pearson's correlation coefficient r between the predicted values and the MDS–UPDRS–III labels.

3 Experiments and results

3.1 Validation of voiced and unvoiced features

The suitability of the features to predict the MDS–UPDRS–III scores is evaluated using an SVR following a leave-one-speaker-out cross-validation strategy. Three different training sets were used to predict the MDS–UPDRS–III scores for males and females. The train sets are formed with: (1) only patients (SVR–PD), (2) only healthy speakers (SVR–HC), and (3) the combination of PD patients and HCs (SVR–PDHC). Additionally, voiced and unvoiced features were used to train each model separately. In general, the features from voiced segments produce the highest correlations for male speakers. A correlation of up to $r = 0.91$ is obtained for males. For the female speakers, the highest correlation is obtained with features from voiced segments ($r = 0.57$). Additionally, a correlation of up to $r = 0.72$ is obtained for male speakers when only features from unvoiced segments are included in the SVR–PDHC training set. However, for female speakers only correlations of up to $r = 0.43$ are obtained using features from unvoiced segments. This results can be explained considering the small data set used for validation i.e. 3 male and 4 female. Additionally, Table 2 shows that FP4 has a strong variation in the correlations. Note also that FP4 in Table 1 has the highest variation in the MDS–UPDRS–III which affects the correlations obtained using the SVR approach.

3.2 Experiments with GMM–UBM

The same groups of speakers used to train the SVR are used to train the UBMs (UBM–PD, UBM–HC, and UBM–PDHC). Then, individual GMMs are adapted for each patient. The highest correlations are obtained using unvoiced features both for males and females. For males, a Pearson's correlation of up to $r = 0.73$ is obtained training the UBM–PDHC. For the case of the females, a correlation of up to $r = 0.53$ is obtained for the UBM–PD. The more accurate modeled speakers are MP1 (UBM: $r = 0.93$, SVR: $r = 0.99$) for males and FP2 for females (UBM: $r = 0.80$, SVR: $r = 0.99$). Note that in Table 3 MP2, FP1, and FP4 have the lowest performance in the models. These results are explained considering that MP2 and FP4 have strong variations in the MDS–UPDRS–III scores and FP1 has almost no variation. The best individual results when HC and PD speakers are considered for training are shown in Figure 2 and Figure 3. The x-axis of the figures represents the recording session and the y-axis represents the normalized value of the Bhattacharyya distance. The normalization is performed with respect to the maximum value of each vector (MDS-UPDRS-III for black solid lines and distances for dotted gray lines). This procedure is only with the aim of depicting comparable curves (MDS-UPDRS-III and the distances) in the same picture.

Table 3: *Pearson's correlation between d_{Bha} and the real MDS–UPDRS–III.**Seg**: Voiced/Unvoiced segments. **MPi/FPj** ($i \in \{1,2,3\}$), ($j \in \{1,2,3,4\}$): Pearson's correlation between d_{Bha} and the real MDS–UPDRS–III score per patient. **Avg**: Average value of the correlations per patient.*

		Male speakers				Female Speakers				
UBM	**Seg**	**MP1**	**MP2**	**MP3**	**Avg**	**FP1**	**FP2**	**FP3**	**FP4**	**Avg**
PD	V	0,90	-0,79	0,44	0,18	0,51	0,14	0,68	-0,05	0,32
	UV	0,96	-0,91	0,78	0,28	0,61	0,65	0,90	-0,05	0,53
HC	V	0,80	-1,00	0,89	0,23	-0,98	0,80	0,90	-0,28	0,11
	UV	0,99	0,94	0,72	0,88	-0,51	0,39	0,42	0,48	0,19
PDHC	V	0,93	-0,99	0,82	0,25	0,52	0,70	0,94	-0,22	0,48
	UV	0,95	0,45	0,78	0,73	-0,82	0,23	0,61	-1,00	-0,24

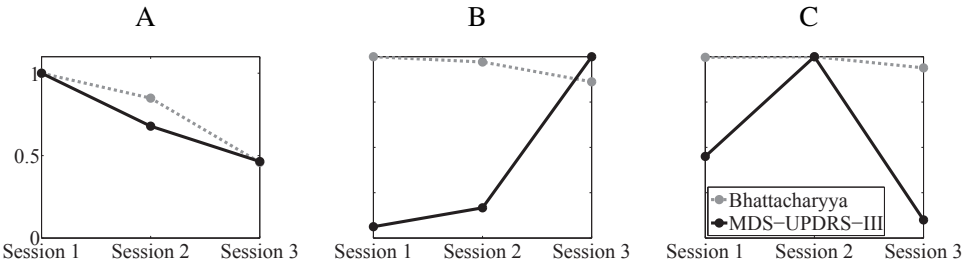

Figure 3: *Normalized scores for the male patients considering the Bhattacharyya distance (Grey dotted line) and the MDS-UPDRS-III labels (Black solid line) using features from voiced segments. (A) MP1, (B) MP2, (C) MP3*

4 Conclusions

A method to assess the progression of PD from speech of male and female speakers is presented. The method allows to model the disease progression of each patient considering individually adapted models using a GMM–UBM approach. The adapted models are used for tracking each patient's neurological state considering speech signals captured over the time. For each group of speakers (males and females), three different UBMs were trained: PD patients, healthy controls, and the combination of both. The Bhattacharyya distance (one per recording session) between the adapted models and the trained UBM was computed for each patient. The Pearson's correlation between the computed distances and the real MDS–UPDRS–III scores was estimated. According to the results for the GMM–UBM approach, the highest correlation between the Bhattacharrya distance and the MDS–UPDRS–III score for males is $r = 0.88$ (average value) training with the HC group. For females the highest average correlation is $r = 0.53$ when the UBM is trained using only the PD group. In the case of the SVR approach the best results obtained for males was approximately $r = 0.90$ using features from the voiced segments in the three groups of speakers. For the females the best results were obtained in the HC group ($r = 0.57$). The mismatch in the results can be explained considering the small data set used for training and validation. Moreover, the variations in the MDS–UPDRS–III affects the performance of the models. Including more people for training the UBMs could increase the performance of the user models, since the GMM–UBM approach performs better with more data. Currently, the data collection is still ongoing in order to improve the number of patients and recording sessions, thus in the near future we will be able to validate this approach with a relatively high number of PD speakers.

5 Acknowledgements

This work was financed by COLCIENCIAS through the project Nº 111556933858.

References

[1] O. Hornykiewicz, "Biochemical aspects of Parkinson's disease," *Neurology*, vol. 51, no. 2 Suppl 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, no. 1, pp. 47–57, 1978.

[3] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel, *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] M. Trail, C. Fox, L. Ramig, S. Sapir, J. Howard, and E. C. Lai, "Speech treatment for Parkinson's disease," *NeuroRehabilitation*, vol. 20, no. 3, pp. 205–221, 2005.

[5] S. Skodda, W. Grönheit, N. Mancinelli, and U. Schlegel, "Progression of voice and speech impairment in the course of Parkinson's disease: a longitudinal study," *Parkinson's Disease*, vol. 2013, 2013. Art. ID 389195.

[6] A. Tsanas, M. Little, P. E. McSharry, and L. Ramig, "Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests," *IEEE Transactions on, Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, 2010.

[7] A. Bayestehtashk, M. Asgari, I. Shafran, and J. McNames, "Fully automated assessment of the severity of Parkinson's disease from speech ," *Computer Speech and Language*, vol. 29, no. 1, pp. 172–185, 2015.

[8] 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 Interspeech*, pp. 478–482, 2015.

[9] T. Grósz, R. Busa-Fekete, G. Gosztolya, and L. Tóth, "As-

sessing the Degree of Nativeness and Parkinson's Condition Using Gaussian Processes and Deep Rectifier Neural Networks," in *Sixteenth Annual Conference of the International Speech Communication Association*, pp. 919–923, 2015.

[10] J. Vásquez-Correa, T. Arias-Vergara, J. Orozco-Arroyave, J. Vargas-Bonilla, J. Arias-Londoño, and E. Nöth, "Automatic Detection of Parkinson's Disease from Continuous Speech Recorded in Non-Controlled Noise Conditions," in *Sixteenth Annual Conference of the International Speech Communication Association*, pp. 105–109, 2015.

[11] J. R. Orozco-Arroyave, J. D. Arias-Londoño, J. F. Vargas-Bonilla, M. C. Gonzalez-Rátiva, and E. Nöth, "New Spanish Speech Corpus Database for the Analysis of People Suffering from Parkinson's Disease," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pp. 342–347, 2014.

[12] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital signal processing*, vol. 10, no. 1, pp. 19–41, 2000.

[13] C. H. You, K. A. Lee, and H. Li, "GMM-SVM kernel with a Bhattacharyya-based distance for speaker recognition," *IEEE Transactions on, Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1300–1312, 2010.