INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Parkinson's disease progression assessment from speech using GMM-UBM

T. Arias-Vergara[1], *J.C. Vasquez-Correa*[1], *J.R. Orozco-Arroyave*[1,2], *J.F. Vargas-Bonilla*[1], *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
rafael.orozco@udea.edu.co

Abstract

The Gaussian Mixture Model Universal Background Model (GMM-UBM) approach is used to assess the Parkinson's disease (PD) progression per speaker. The disease progression is assessed individually per patient following a user modeling-approach. Voiced and unvoiced segments are extracted and grouped separately to train the models. Additionally, the Bhattacharyya distance is used to estimate the difference between the UBM and the user model. Speech recordings from 62 PD patients (34 male and 28 female) were captured from 2012 to 2015 in four recording sessions. The validation of the models is performed with recordings of 7 patients. All of the patients were diagnosed by a neurologist expert according to the MDS-UPDRS-III scale. The features used to model the speech of the patients are validated by doing a regression based on a Support Vector Regressor (SVR). According to the results, it is possible to track the disease progression with a Pearson's correlation of up to 0.60 with respect to the MDS-UPDRS-III labels.
Index Terms: Speech disorders, GMM-UBM, Parkinson's disease, user modeling.

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, slowed movement, lack of coordination, and several speech disorders. The severity and progression of PD varies among patients. Currently the disease 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) [2]. This is a perceptual scale used to assess motor and non-motor abilities of PD patients. The total MDS-UPDRS scale is divided into four sections. In this study we consider only the third section (MDS-UPDRS-III) because it evaluates the motor capabilities of the patients which are highly affected by the disease. The scale has a total of 33 items to evaluate different motor abilities and only one of them considers speech; however the speech production process involves more than 100 muscles and limbs, thus it makes sense to model motor capabilities from speech considering different aspects such as stability in the vocal folds vibration, energy content, articulatory capability, among others. On the other hand PD severity is evaluated according to a clinical criterion and the inter-variability of the MDS-UPDRS score could be high. Therefore, it is necessary to develop computer aided systems to support the clinical diagnosis and to assess the disease progression objectively. There are many studies focused on the detection and tracking of PD from speech. In [3] the authors presented a methodology to predict the disease progression from speech signals. The UPDRS scores were predicted with the Classification And Regression Trees (CARTs) approach. A total of 42 PD patients were captured once per week during six months. Neurologist experts evaluated the patients three times during the study. The authors report a Mean Absolute Error (MAE) of 7.5 points in the predictions of the total value of the UPDRS scale. Further, the UPDRS-III was predicted with a MAE of 6 points. The novelty of the method is the disease severity assessment from speech. However, the speaker independence condition is not guaranteed in the validation process. Thus, the reported results are highly optimistic and biased. The progression of speech impairments in a longitudinal study is presented in [4]. The speech of 80 PD patients was recorded from 2002 to 2012 in two different recording sessions. The time between the first and second session ranged from 12 to 88 months. A control group of 60 healthy persons was also considered. The participants were asked to read a text and to produce a sustained phonation of the vowel /a/. In both sessions the patients were assessed by neurologists experts according to the UPDRS-III. The audio signals were perceptually evaluated and the authors report significant Pearson's correlations (r) between the speech item of the UPDRS scale and the perceptual speech score. The evaluation was performed in four aspects of speech (voice, articulation, prosody, and fluency). For voice, the correlation was $r = 0.406$, $P < 0.0001$, for articulation it was $r = 0.364$, $P = 0.0004$, for prosody it was $r = 0.383$, $P = 0.0002$, and for fluency, it was $r = 0.472$, $P < 0.0001$.

There are also studies considering speech signals recorded only in one session to predict the neurological state of PD patients. In [5] the authors proposed a methodology to predict the UPDRS-III score. Speech recordings of 82 subjects were collected. Three different speech tasks were considered: sustained phonation of the vowel /a/, the repetition of syllables, and the reading of three different texts. The authors reported that it is possible to predict the UPDRS-III with a MAE of 5.66 using an ε-SVR with a cubic polynomial kernel. Recently in the INTERSPEECH 2015 Computational Paralinguistic Challenge (ComParE 2015) there was a Parkinson's Condition subchallenge that addressed the task of neurological state prediction of PD patients from speech [6]. Recordings of the 50 patients included in the PC-GITA database [7] were considered to form the train and development subsets. The test set included a total of 11 new patients recorded in non-controlled noise conditions, i.e., not using a sound-proof booth. A total of 42 speech tasks were considered. The neurological state of the patients was assessed by a neurologist expert according to the MDS-UPDRS-III subscale. The winners of the challenge reported a Spearman's correlation of 0.65 between the real MDS-UPDRS-III scores and the predicted values [8]. Note that most of the studies in the literature are focused on predicting the neurological state of groups of PD patients from speech recordings. In this paper we introduce a methodology to track the disease progression of PD patients comparing their neurological state

http://dx.doi.org/10.21437/Interspeech.2016-1122

with themselves, i.e., we are proposing an approach for individual modeling of the disease progression that can be adapted to each patient individually. The proposed method is based on the Gaussian Mixture Model Universal Background Model (GMM-UBM) approach. The GMM-UBM systems are commonly used in speaker recognition due to their capability of representing a large class of sample distributions from which single-speaker models are obtained [9]. As a result, the main hypothesis is that the changes in the voice of PD patients can be track from the individual speaker models. Three different version of the universal background model are considered: (1) with recordings of a total of 61 PD patients, (2) with 62 healthy speakers, and (3) with both groups of speakers. The speech data is formed with recordings of four sessions. Some of the participants were recorded twice, but most of them participated in three of the four recording sessions. Seven of the patients were recorded in all of the sessions, and their recordings were considered to evaluate their disease progression, i.e., they were considered for the adaptation process. To the best of our knowledge, this is the first study focused on modeling the disease progression considering individually-adapted models for tracking the neurological state of PD patients from speech over the time.

The rest of the paper is organized as follows: Section 2 includes details of the data and methodology. Section 3 describes the experiments and results. Section 4 provide conclusions derived from this work.

2. Methods and materials

Speech recordings from several patients were collected in four recording sessions during a period of three years. A subset of seven patients participated in the four recording sessions, thus their speech signals are considered to make the adaptation of the background model. Each patient that participates in the adaptation process is excluded from the UBM. The background model is considered as the baseline to assess the disease progression according to its distance to the adapted speaker. Three background models are considered, i.e., with recordings of the PD patients, with the healthy speakers, and 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 four 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 four neurological scores is calculated. The process is summarized in Figure 1 and further details are provided in the following subsections.

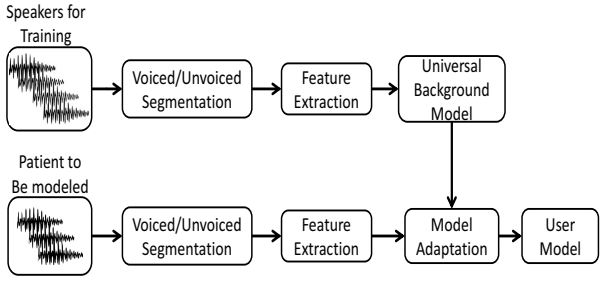

Figure 1: General methodology

2.1. Data description

Speech recordings from 62 PD patients (34 males and 28 females) were collected in a total of four recording sessions distributed between 2012 to 2015. Seven of these 62 speakers participated in all of the four sessions. A professional audio setting was used for the first two sessions, and the other sessions were recorded in non-controlled acoustic conditions using the device presented in [10]. All of the patients from the first, second and fourth sessions were diagnosed by a neurologist expert according to the MDS-UPDRS-III [2]. The labels of the third recording session were not available, thus a linear interpolation was used to obtain the corresponding score. The age, gender, and MDS-UPDRS-III scores of the patients obtained in each session are provided in Table 1 (the age was collected during the first recording session). A table with the information of all patients can be found online[1]. The set of healthy control (HC) speakers is formed with recordings from 62 persons. 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 once. All the participants of the tests followed the set of speech tasks presented in [7]. In this study only the read text was used.

Table 1: *Distribution of patients recorded in all sessions.* ***Session*** **i** *($i \in \{1, 2, 4\}$): MDS-UPDRS-III scores obtained on each recording session. Session 3 corresponds to the Linearly interpolated MDS-UPDRS-III score.*

Patient	Age	Gender	Session 1	Session 2	Session 3	Session 4
P1	64	M	28	19	16	13
P2	59	M	6	8	16	24
P3	55	F	29	26	26	26
P4	51	F	38	49	47	44
P5	57	F	41	35	34	33
P6	56	F	43	10	15	19
P7	68	M	14	25	16	7

2.2. Voiced/Unvoiced characterization

Voiced and unvoiced segments are extracted from the recordings. Hamming windowing with 20 ms length and a time-shift of 10 ms is applied. Cepstral Mean Subtraction (CMS) is applied to reduce possible bias introduced due to the channel conditions. For voiced frames the set of features includes jitter, shimmer, and 12 Mel–Frequency Cepstral Coefficients (MFCCs), forming a 14-dimensional feature vector. The unvoiced frames are modeled computing 12 MFCCs and the log energy of the signal distributed in 25 Bark bands, forming a 37-dimensional feature vector.

2.3. User modeling

We use the GMM-UBM approach to assess the disease progression per speaker. GMM-UBM based systems are capable of representing arbitrary probabilistic densities. In speech processing these models are used to represent the distribution of feature vectors extracted from several speakers. We referred to the resulting trained model as UBM. GMMs are parametric probabilistic models represented as a weighted sum of M Gaussian densities. For a D-dimensional feature vector $\boldsymbol{x}$ a GMM is

[1] https://www5.cs.fau.de/en/our-team/orozco-rafael/projects/IS-2016-Tomas

defined as

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

The Gaussian densities $p_i(\boldsymbol{x})$ are parametrized by the mixture weights ω_i, a D×1 mean vector $\boldsymbol{\mu_i}$, and a D×D covariance matrix $\boldsymbol{\Sigma_i}$ [11]. The parameters of the density models 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 number of Gaussians used to train the UBM ranges from 2 to 1024 in steps of 2^n ($n \in \{1, 2, 3..., 10\}$). User models are generated for the patients described in Table 1. One patient is extracted to be modeled. The remaining speakers are used to train the UBM. The trained model is adapted using the Maximum A Posteriori (MAP) approach. Then, we compute the distance between the UBM and the adapted model. Four adaptations (one per recording session) are performed for each patient, thus the resulting user model contains 4 distance values. The details of the procedure are depicted in Figure 2.

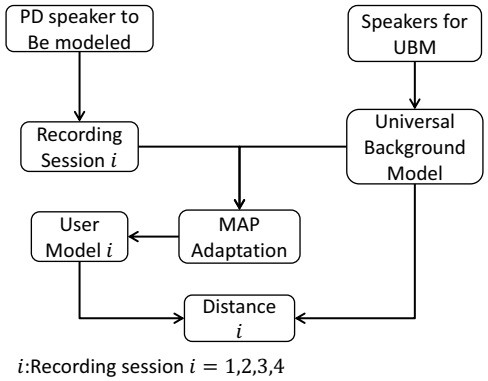

Figure 2: User modeling methodology.

2.4. Distance computation

The Bhattacharyya distance measures the change 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 ($\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$) [12].

$$\mathrm{d_{Bha}} = \mu_{\mathrm{Bha}} + \Sigma_{\mathrm{Bha}} \quad (3)$$

The first term is the mean statistical measure and is calculated using Equation 4.

$$\mu_{\mathrm{Bha}} = \frac{1}{8}\sum_{i=1}^{M}\left\{(\widehat{\mu}_i - \mu_i)^T \left[\frac{\widehat{\Sigma}_i + \Sigma_i}{2}\right]^{-1}(\mu_i - \widehat{\mu}_i)\right\} \quad (4)$$

The second term is the covariance statistical measure and is defined as

$$\Sigma_{\mathrm{Bha}} = \frac{1}{2}\sum_{i=1}^{M}\left[\ln\frac{\frac{\widehat{\Sigma}_i+\Sigma_i}{2}}{\sqrt{\left|\widehat{\Sigma}_i\right|\left|\Sigma_i\right|}}\right] - \omega_{\mathrm{Bha}} \quad (5)$$

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

2.5. Regression model

The disease severity according to the MDS-UPDRS-III (y) is estimated using a linear support vector regressor (SVR). 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 6.

$$L(y,\widehat{y})) = \begin{cases} 0 & \text{if } |y-\widehat{y}| \leq \varepsilon \\ |y-\widehat{y}| - \varepsilon & \text{otherwise} \end{cases} \quad (6)$$

The parameters of the regressor C and ε are optimized 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 a ε-SVR. Three different groups of speakers were considered as training sets: patients (SVR-PD), controls (SVR-HC), and patients with controls (SVR-PDHC). The optimization is performed following a leave-one-speaker-out cross-validation (LOSO-CV) strategy. Table 2 shows the Pearson's correlation obtained per patient. The highest correlations were obtained for the unvoiced segments (SVR-PD: $r = 0.53$; SVR-HC: $r = 0.46$; SVR-PDHC: $r = 0.77$). Note that when the training process is performed including patients and controls we obtained the highest result ($r = 0.77$). For voiced segments the correlation is higher when healthy controls are considered for training ($r = 0.43$). In this case when PD and HC speakers are grouped the correlation also improved ($r = 0.63$).

Table 2: *Pearson's correlation between the predicted scores and the real MDS-UPDRS-III values.* ***Seg****: Voiced (V)/Unvoiced (UV) segments.* ***Pi*** *($i \in \{1, 2, \ldots, 7\}$): Pearson's correlation between predicted scores and the real MDS-UPDRS-III per patient.* ***Avg****: Average value of the correlation scores per patient.*

Training set	Seg	P1	P2	P3	P4	P5	P6	P7	Avg
PD	V	0.99	0.69	−0.62	0.38	−0.03	0.62	0.62	0.38
	UV	0.68	0.97	0.13	0.99	−0.32	0.38	0.87	0.53
HC	V	0.99	0.85	−0.33	0.38	−0.04	0.56	0.61	0.43
	UV	0.91	0.54	0.38	0.99	0.04	0.25	0.10	0.46
PDHC	V	0.99	0.83	0.61	0.55	−0.03	0.84	0.62	0.63
	UV	0.99	0.85	0.83	0.95	0.42	0.51	0.87	0.77

3.2. Experiments with the GMM-UBM approach

Three different UBMs were trained using feature vectors from patients (UBM-PD), controls (UBM-HC), and patients combined with controls (UBM-PDHC). Each UBM was trained using voiced and unvoiced frames separately. Table 3 shows the computed Pearson's correlation between the $\mathrm{d_{Bha}}$ values and the MDS-UPDRS-III labels. Note that the highest correlation is obtained when the models are trained using unvoiced segments (UBM-PD: $r = 0.44$, UBM-HC: $r = 0.29$, UBM-PDHC: $r = 0.60$). According to the results, the correlation decreases when the model is trained using only HC speakers. However, when patients and controls are combined the correlation increased ($r = 0.60$). This result can be likely explained

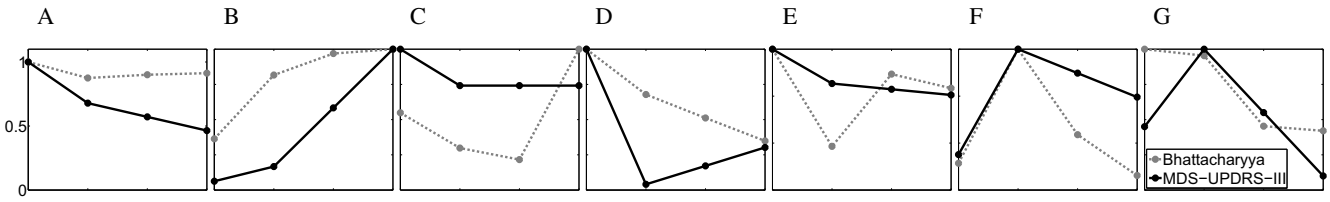

Figure 3: *Normalized scores for each patient considering the Bhattacharyya distance (Grey dotted line) and the MDS-UPDRS-III labels (Black solid line) using features from unvoiced segments. (A) P1, (B) P2, (C) P3, (D) P4, (E) P5, (F) P6, (G) P7.*

because when HC and PD speakers are combined the number of participants increases, making the background more suitable to perform the adaptation. We hypothesize that this is not observed when only HC or PD speakers are included in the UBM because the reduced number of speakers. The user models are also tested considering only the non-interpolated MDS-UPDRS-III scores (S1, S2, and S4). The results are comparable to those reported in Table 3 and can be observed in the online document[1]. Further experiments with more speakers/recordings are required to validate this hypothesis.

Table 3: *Pearson's correlation between d_{Bha} and the real MDS-UPDRS-III.* ***UBM****: Group of speakers used to train the UBM model.* ***M****: Number of Gaussians used to train the model.* ***Seg****: Voiced/Unvoiced segments.* ***Pi*** *($i \in \{1, 2, \ldots, 7\}$): Pearson's correlation between d_{Bha} and the real MDS-UPDRS-III score per patient.* ***Avg****: Average value of the correlations per patient.*

UBM	M	Seg	P1	P2	P3	P4	P5	P6	P7	Avg
PD	1024	V	0.88	−0.83	0.01	0.41	0.58	0.31	0.79	0.31
	256	UV	0.95	−0.49	0.99	−0.33	0.84	0.65	0.48	0.44
HC	512	V	0.92	−0.99	0.91	0.95	0.70	−0.87	0.15	0.25
	16	UV	0.67	−0.81	0.64	0.34	0.40	0.14	0.68	0.29
PDHC	128	V	0.87	0.01	0.90	0.97	0.24	−0.99	0.54	0.36
	64	UV	0.78	0.79	0.06	0.73	0.50	0.74	0.59	0.60

In general, the results obtained with the SVR show that the features used to predict the disease progression are suitable to be used for the GMM-UBM modeling, i.e., with that set of features it is possible to obtain predictions of the MDS-UPDRS-III scores highly correlated with the real labels. For voiced and unvoiced segments there is an improvement in the correlations when the models are trained with patients and controls. Additionally, the best results are obtained for the unvoiced segments. Unvoiced segments improve the correlation per speaker when the models (UBM,SVR) are trained combining patients and controls. The best results when HC and PD speakers are considered in train are shown in Figure 3. The x-axis represents the recording session and the y-axis represents the normalized values of the Bhattacharyya distance and the real MDS-UPDRS-III score. 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. The distances computed from each user model represent the progression of the disease. Note that the trend of Bhattacharrya distances follows the trend of the neurological state of the patients. This behavior can be observed clearly in Figures 3.A, 3.B, and 3.F which are the patients with higher correlation. Additional to the Bhattacharyya measure, we calculated the Euclidean, Cosine, City Block, and Chebyshev distances. We considered the mean vectors of the speaker model and the UBM for these tests. The higher correlation was obtained with the Cosine distance for the UBM-PD trained with voiced segments ($r = 0.63$). However, when the UBM is trained with the combination of patients and controls the results are not satisfactory. This experiments were performed only to verify whether these classical measures are suitable in this particular problem.

4. Conclusions

A methodology to assess Parkinson's disease progression from speech using the GMM-UBM approach is presented. Voiced and unvoiced segments are extracted and grouped separately. The method allows the assessment of the disease progression of a specific patient, i.e., modeling his/her disease progression considering individually-adapted models for tracking his/her neurological state over the time. Three different UBMs are trained: Parkinson's patients, healthy speakers, and the combination of both. The Bhattacharyya distance between the speaker models and the UBM was computed. One distance value per recording session was calculated per patient. The Pearson's correlation between the Bhattacharrya distances and the MDS-UPDRS-III labels was calculated. The highest correlation values are obtained when PD and HC speakers are combined in the UBM, indicating that it is worth to include information from control people to improve the results of the predictions. The validity of the proposed approach to predict the neurological state of the patients is evaluated considering the same set of features to correlate the real values of the MDS-UPDRS-III scores with the predicted ones. According to the results of the proposed approach, when HC and PD speakers are included in the background model, the distance between the UBM and the model of the adapted speaker correlates with the disease progression (average value of $r = 0.60$). This result indicates that the prediction of the neurological state of PD patients can be improved using information of healthy speakers in the training process. To the best of our knowledge this is the first contribution considering a method to track the neurological state of individual PD patients over the time. This paper is a step forward in the development of computer aided tools for the continuous and unobtrusive monitoring of people with Parkinson's disease. 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.

6. References

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

[2] 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.

[3] 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.

[4] 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.

[5] M. Asgari and I. Shafran, "Extracting Cues from Speech for Predicting Severity of Parkinson's Disease," in *IEEE International Workshop on, MLSP*. IEEE, 2010, pp. 462–467.

[6] 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 *Proceeding of the Sixteenth Annual Conference of the International Speech Communication Association*, 2015, pp. 478–482.

[7] 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*, 2014, pp. 342–347.

[8] 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 *Proceeding of the Sixteenth Annual Conference of the International Speech Communication Association*, 2015, pp. 919–923.

[9] D. Reynolds, *Gaussian Mixture Models*. Boston, MA: Springer US, 2009, pp. 659–663.

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

[11] 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.

[12] 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.