INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Automatic Classification of Hypokinetic and Hyperkinetic Dysarthria based on GMM-Supervectors

C. D. Rios-Urrego[1], *J. Rusz*[2], *E. Nöth*[3], *J. R. Orozco-Arroyave*[1,3]

[1]GITA Lab, Faculty of Engineering, University of Antioquia, Medellín, Colombia
[2]Department of Circuit Theory, Czech Technical University in Prague, Prague, Czech Republic
[3]Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

cdavid.rios@udea.edu.co

Abstract

Hypokinetic and hyperkinetic dysarthria are motor speech disorders that appear in patients with Parkinson's and Huntington's disease, respectively. They are caused due to progressive lesions or alterations in the basal ganglia. In particular, Huntington's disease (HD) is known to be more invasive and difficult to treat than Parkinson's disease (PD), producing more aggressive motor and cognitive alterations. Since speech production requires the movement and control of many different muscles and limbs, it constitutes a highly complex motor activity that may reflect relevant aspects of the patient's health state. This paper proposes the discrimination between patients with PD, HD, and healthy controls (HC) based on different speech dimensions. Speaker models based on Gaussian-mixture model supervectors are created with the features extracted from each speech dimension. The results suggest that it is possible to distinguish between PD and HD patients using the supervectors-based approach.

Index Terms: Parkinson's disease, Huntington's disease, Pathological speech, Articulation, Phonation, Prosody.

1. Introduction

Parkinson's and Huntington's diseases have captured the attention of the research community since many years ago. Both diseases cause different motor and non-motor impairments, contributing to a significant decrease in the quality of life of patients [1]. On the one hand, Parkinson's disease (PD) is characterized by symptoms such as resting tremor, bradykinesia, rigidity and freezing of gait [2]. Most of PD patients develop several speech deficits which are grouped and called hypokinetic dysarthria. Dysarthric speech appears as the result of losing the control of the muscles and limbs involved in the speech production process. Typical characteristics of hypokinetic dysarthria include monoloudness, reduced voice quality, monotonicity, imprecise pronunciation of consonants and vowels, lack of fluency, voice tremor, and other characteristics [3, 4]. On the other hand, Huntington's disease (HD) produces involuntary movements or chorea, cognitive deficits, dystonia, and rigidity that appear even in patients in early stages of the disease [5]. HD patients develop hyperkinetic dysarthria, which appears primarily as a consequence of chorea. The most relevant deficits in speech include phonatory dysfunction, unpredictable interruptions of articulation, and abnormal prosody [6].

Therefore, PD and HD could provide a theoretical model for the evaluation of speech patterns connected with hypokinetic and hyperkinetic dysarthria, which are often counteractive. This might be helpful in situations like estimation of effect of levodopa-induced dyskinesia that may lead to hyperkinetic speech patterns in PD [7] as well as effect of pallidal deep brain stimulation on speech in patients with dystonia that may both improve hyperkinetic but aggravate hypokinetic speech aspects as a negative side-effect of stimulation [8]. However, the scientific community has lees explored the classification between hypokinetic and hyperkinetic dysarthrias. For example, in [9], the authors analyzed the effect of both diseases in the initiation, planning, and production of speech. 12 PD patients, 12 HD patients, and 12 HC subjects were evaluated. The authors extracted different prosody features and concluded that the most discriminating ones for HD are the ones that model changes in syllable duration, and the duration of pauses in the sentences. For PD patients, only the duration of the sentences was altered. Prosody impairments were also studied in [10], where the authors considered a set with 7 PD patients, 5 HD patients, and 12 HC subjects. Several acoustic features were extracted including duration, intensity, and durational accent. The authors concluded that HD patients presented a reduction in the duration, tone, and volume of their voice, while for PD patients there was a slight decrease in the duration. The authors in [11] introduced an automated method for the analysis of vocal tremor in multiple neurological diseases. The authors included 240 participants divided into 9 groups of pathologies among which there were 40 PD and 20 HD patients. The authors observed that 65% of the HD patients showed abnormal vocal tremor while only 20% of the PD patients showed the pattern. Finally, in [12] the authors proposed guidelines for speech recording and acoustic analyses in dysarthrias. They analyzed data from 50 HC subjects, 30 PD patients, and 30 HD patients. The authors demonstrated that the hyperkinetic dysarthria group had more affected speech dimensions compared with the HCs than the hypokinetic speakers.

In this work, we studied hypokinetic and hyperkinetic dysarthria in native Czech patients with PD and HD. Three different speech dimensions were evaluated: articulation, phonation, and prosody. The analyses are based on Gaussian mixture model GMM [13] supervectors, which are created for each speech dimension. Different Universal Background Model (UBM) were generated using German and Spanish corpora. Finally, each supervector and their combination were used for two classification scenarios: PD vs. HD, and PD vs. HD vs. HC. We also tried to associate the abnormal patterns observed in the speech of each subject's group with different types of dysarthria. As far as we know, this is one of the first studies that addresses the topic of classifying between PD and HD subjects considering different speech dimensions.

2. Data

2.1. Recordings considered to train the UBM

Two Parkinson's databases were considered. The first one is PC-GITA [14] which contains recordings of 50 PD patients and

10.21437/Interspeech.2023-2146

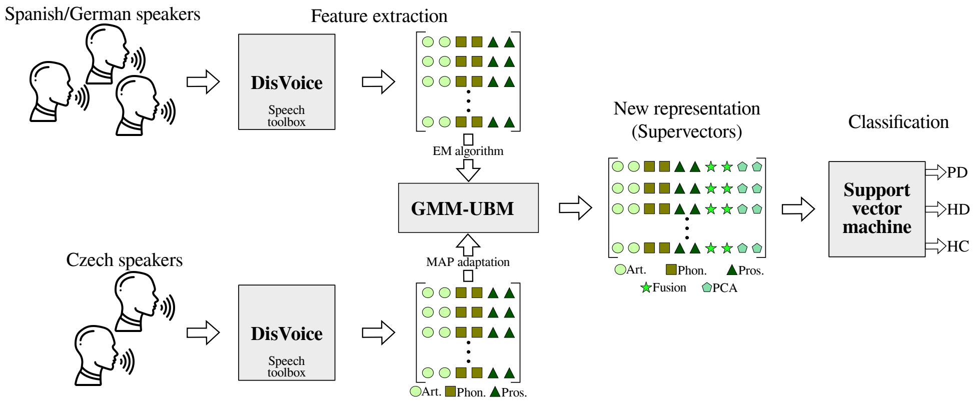

Figure 1: *General methodology addressed in this study to model the speech of patients with neurodegenerative disorders using GMM supervectors created with information extracted from articulation (**Art.**), phonation (**Phon.**), and prosody (**Pros.**) features. **Fusion**: Fusion of supervectors from articulation, phonation, and prosody. **PCA.**: Principal component analysis computed from the fusion supervector. **PD**: Parkinson's disease. **HD**: Huntington's disease. **HC**: Healthy controls. **MAP**: Maximum a posteriori.*

50 HC subjects, all of them native speaker of Colombian Spanish. The corpus is balanced in age, gender, and education level. The second corpus is composed with recordings of 88 PD patients and 88 HC subjects balanced in gender. All speakers in this dataset were German native speakers [15]. Besides the Parkinson's databases, two more corpora with a large number of utterances were considered for UBM training. These corpora are mainly used for the training of speech recognition systems. The first database is called CIEMPIESS and consists of 17 hours of FM podcasts in Mexican Spanish [16]. The data contain 16717 audio files recorded at a sampling frequency of 16 kHz. The second database was the Verbmobil corpus [17], which consists of speech recordings from 586 German native speakers with a total of 29 hours of dialogues. The data contain 11714 audio files recorded at a sampling frequency of 16 kHz.

2.2. Recordings considered to create the GMM supervectors

Two different corpora were used in this study to evaluate the proposed approach. Both consist of recordings of Czech speakers. One includes 50 PD patients and 50 HC subjects [18]. Each patient used to create the supervectors were evaluated by a neurologist expert according to the third section of the Unified Parkinson's Disease Rating Scale (UPDRS-III). The other corpus includes recordings of 40 HD patients and 40 HC subjects [19], the Huntington's patients were evaluated according to the Unified Huntington's Disease Rating Scale (UHDRS) [20]. Both corpora were collected in a quiet room with using a head-mounted condenser microphone (Beyerdynamic Opus 55) placed approximately 5 cm from the corner of the subject's mouth. None of the healthy participants had a history of neurological or communication disorders. Two speech tasks were considered for this study: the rapid repetition of the syllables /pa-ta-ka/ and a monologue (the participants were requested to talk about their daily routines). Each signal was down-sampled to 16kHz. Table 1 summarizes the demographic information about the speakers.

Table 1: *Demographic information of the speakers. [F/M]: Female/Male. Time since diagnosis and age are given in years. Values reported in terms of mean ± standard deviation.*

	Patients	**HC subjects**	**PD/HD vs. HC**
	PD Czech		
Gender [F/M]	20/30	20/30	$^{*}p$=0.94
Age [F/M]	60.1±9/65.3±10	63.5±11/60.3±12	$^{**}p$=0.32
Range of age [F/M]	41–72/43–82	40–79/41–77	
Time since diagnosis [F/M]	6.8±5/6.7±5		
UPDRS-III [F/M]	18.1±10/21.4±12		
Speech item (UPDRS-III) [F/M]	0.7±0.6/0.9±0.5		
	HD Czech		
Gender [F/M]	20/20	20/20	$^{*}p$=1.00
Age [F/M]	49.5±14.1/47.7±12.2	50.1±13.9/48.3±12.3	$^{**}p$=0.43
Range of age [F/M]	27–69/23–67	27–69/26–70	
UHDRS [F/M]	27.1±10.7/26.8±12.7		
Speech item (UHDRS) [F/M]	0.7±0.5/0.9±0.3		

* p–value calculated through Chi–square test.
** p–value calculated through Mann-Whitney U test.

3. Methods

The methodology addressed in this work consists of four main stages: (1) the training of the UBM from the Spanish and German databases using dynamic features of articulation, phonation and prosody. (2) Adaptation of each speaker with PD, HD, and HC in Czech using MAP adaptation. (3) New representations called supervectors are built using the vector of means and the diagonal of covariance. (4) Classification of the Czech subjects is performed using a support vector machine (SVM) classifier. Two classification scenarios are considered: PD vs. HD, and PD vs. HD vs. HC. This methodology is summarized in Figure 1. Details of each method are presented below.

3.1. Feature extraction

Articulation, phonation, and prosody features were extracted to model different deficits that appear in the speech of subjects suffering from motor speech disorders like those derived from PD or HD. The features are extracted using the DisVoice toolkit[1]. Details of each feature set are presented below.

Articulation: This speech dimension evaluates the ability of a speaker to control the movement of the articulators to a correct position, at the correct time, and with the appropriate duration and energy while producing speech. In this work, the transition from unvoiced to voiced segments (onset) was considered as the way to evaluate the difficulties of the speaker to start the vibration of the vocal folds [21]. Onset segments were detected based on the presence of the fundamental frequency (F_0). After onset detection, 40 ms were taken to the left and to

[1] https://disvoice.readthedocs.io/en/latest/

the right of the border, forming segments with 80 ms length. A total of 58 features were extracted from the transition segments including the energy content in 22 critical bands distributed according to the Bark scale, and 12 MFCCs with their first and second derivatives [22].

Phonation: This speech dimension aims to model the ability of a speaker to produce air in the lungs to produce voiced or unvoiced sounds. We focused mainly on the production of voiced sounds with the aim to model the capability of the subjects to control the vocal fold vibration. The phonation feature set was formed with seven measures computed over voiced segments of the speech signal: (1-2) the first and second F_0 derivative, (3) shimmer, (4) jitter, (5-6) amplitude and pitch perturbation quotients, namely APQ and PPQ, respectively, and (7) log energy per frame as a measure of loudness. Additional information about the computation of phonation features is presented in [23]

Prosody: These measures intend to model changes in intonation, timing, and loudness. A total of 13 prosody features was extracted upon each voiced segment including the duration of the segment, the coefficients of a 5-degree polynomial that models the F_0 contour and also the coefficients of a 5-degree Lagrange polynomial that models the energy contour. Additional information can be found at [24].

3.2. Gaussian Mixture Models - Universal Background Models

The dynamics of the features described in the previous subsections was modeled by following the GMM-UBM framework. GMMs are probability models representing a population from a combination of Gaussian probability distributions. For a D-dimensional feature vector $\boldsymbol{x}$, the mixture density used for the likelihood function for M Gaussian is defined as $p(\boldsymbol{x}|\lambda) = \sum_{i=1}^{M} w_i p_i(\boldsymbol{x})$, where $p_i(\boldsymbol{x})$ corresponds to a Gaussian density weighted by w_i such that it satisfies the constraint $\sum_{i=1}^{M} w_i = 1$. In addition, each p_i distribution is composed of a mean vector $[\boldsymbol{\mu}_i]_{D\times 1}$ and a covariance matrix $[\boldsymbol{\Sigma}_i]_{D\times D}$. The set of parameters for the density model are denoted as $\lambda = \{w_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$, where $i = 1, ..., M$.

EM algorithm: The parameter set λ of the maximum likelihood function can be estimated using the Expectation Maximization (EM) algorithm [25] which iteratively re-defines the parameters and increases the likelihood of the estimated model for the observed feature vectors; that is, for iterations k and $k+1$, $p(\boldsymbol{X}|\lambda^{(k+1)}) > p(\boldsymbol{X}|\lambda^{(k)})$, where $\boldsymbol{X}$ is a matrix with the group of features $\boldsymbol{x}$ extracted from each participant in the database [13].

Maximum A-posteriori Adaptation: The parameters that model each speaker were derived from an adaptation process denoted as maximum a-posteriori (MAP) [26]. Unlike using only GMM and the EM algorithm, the main idea of the MAP adaptation is to derive parameter updates from the UBM which is considered as a robust and well-trained basis. This provides a closer coupling between each model and the UBM [13].

Supervectors: A GMM supervector can be considered as a representation in smaller-dimensional vectors after adaptation from the UBM, this allows mapping from a dynamic representation for each utterance to a global static representation. For this work, the GMM supervector is created by stacking the means $\boldsymbol{\mu}_i'$ and the diagonal of the covariance matrix $\boldsymbol{\Sigma}_i'$ of the mixture components.

4. Experiments and results

For the classification stage, we used an SVM to discriminate the different groups of subjects. The parameter was varied as $C \in \{0.001, 0.005, 0.01, \cdots, 100, 500, 1000\}$. Similarly, the bandwidth of the kernel (γ_k) was varied as $\gamma_k \in \{0.0001, 0.001, \cdots, 1000\}$. Each experiment was trained and evaluated following a stratified k-fold cross-validation strategy with 10 folds. The process was repeated 10 times for a better generalization of the results. In addition, the optimization of the number of Gaussian components M in the UBM was based on the accuracy in test where $M \in \{2, 4, 8, 16, 32, 64, 128\}$, resulting in supervectors of size $M \times 2 \times (58 + 7 + 13)$.

Two different experiments were performed for this study, all of them considering only Czech speakers. A bi-class problem (PD vs. HD) and one tri-class problem (PD vs. HD vs. HC) were included. The speaker adaptation was based per speaker on the UBMs created with the recordings of the Colombian (Spanish) and German speakers. UBMs created with only samples of HC subjects and also with the combination of PD patients and HC were considered. In addition, we considered creating another 3 UBM models with a large number of recordings in Spanish and German, namely CIEMPIESS and Verbmobil, respectively. Finally, different GMM supervectors are obtained from each UBM including those based on articulation, phonation, and prosody features separately. Additionally, two other schemes were evaluated: the fusion of the 3 speech dimensions and dimensionality reduction of the fusion using a principal component analysis (PCA) with 90% of the cumulative variance.

4.1. Bi-class classification (PD vs. HD)

Table 2 shows the overall results of the classification between PD vs. HD patients. Accuracies are reported in terms of the unweighted average recall (UAR). The best result was obtained for the monologue with an UAR of 86.2% and its adaptation was obtained using an UBM trained with the full German database (including patients and controls). For the /pa-ta-ka/ task, the best result was obtained with the UBM trained with the complete Spanish database with an UAR of 81.6%. From this result, we can conclude that it is possible to differentiate between these two diseases due to two main reasons: in the case of HD, there are involuntary and rapid movements, while in PD, there is rigidity in the muscles which causes a reduction in voice quality. Previous studies such as [27, 28] have shown that the phonatory ability of patients with PD and HD is largely impaired as the neurodegenerative disease progresses. It was possible to verify that this dimension is fundamental for the discrimination of these diseases, specifically when continuous speech tasks are evaluated. The highest UAR obtained with the /pa-ta-ka/ task was 81.6% when the three speech dimensions are combined, indicating that the three of them are relevant and actually they are complementary.

4.2. Multi-class classification (PD vs. HD vs. HC)

For this experiment, controls from the 2 Czech databases were merged to be evaluated with respect to the PD and HD patients (a Kruskal-Wallis test between the two subgroups with HC subjects was performed to discard any possible bias due to acoustic conditions). A one-vs-rest SVM was used in this case to perform the tri-class classification. Table 3 shows the results obtained in this experiment. It is possible to observe that, as in the previous experiments, the best results were obtained with the

Table 2: *Classification of PD vs. HD with each speech dimension and their fusion.* ***UAR****: Unweighted Average Recall,* ***M****: Number of Gaussian components. Values reported in terms of mean ± standard deviation.*

	Monologue								Pataka							
UBM	**Articulation**		**Phonation**		**Prosody**		**Fusion**	**PCA**	**Articulation**		**Phonation**		**Prosody**		**Fusion**	**PCA**
	M	**UAR (%)**	**M**	**UAR (%)**	**M**	**UAR (%)**	**UAR (%)**	**UAR (%)**	**M**	**UAR (%)**	**M**	**UAR (%)**	**M**	**UAR (%)**	**UAR (%)**	**UAR (%)**
German (HC)	4	77.9±2.4	64	83.2±2.9	2	72.1±1.6	81.8±2.4	61.8±3.0	2	68.6±2.9	32	71.4±2.0	2	74.9±3.6	75.2±2.9	57.9±3.0
Spanish (HC)	4	75.9±2.3	64	82.8±1.8	2	69.1±3.3	79.9±3.2	48.1±4.8	4	73.1±2.2	16	68.1±3.4	2	69.7±2.2	78.2±2.2	54.7±2.8
German-Spanish (HC)	4	75.7±2.5	64	83.1±1.6	4	71.1±2.5	81.2±3.0	56.3±3.4	2	68.8±4.0	8	67.1±2.8	2	68.4±1.8	77.7±2.8	52.0±3.3
German (HC+PD)	4	74.8±3.0	128	**86.2±1.8**	4	73.9±2.0	84.1±1.7	65.2±1.8	2	66.6±2.3	16	69.3±4.3	2	68.1±3.7	75.2±1.5	47.2±3.9
Spanish (HC+PD)	4	75.2±1.9	64	83.7±2.1	2	69.1±2.0	83.2±2.2	67.9±2.9	4	70.6±2.7	32	73.2±2.0	2	67.0±3.4	**81.6±1.3**	66.7±2.2
German-Spanish (HC+PD)	4	77.7±1.7	64	83.6±1.8	2	70.4±2.3	82.2±1.7	62.1±2.5	4	70.9±2.0	16	69.4±3.3	2	75.4±2.2	78.2±2.5	59.4±3.6
CIEMPIESS	2	73.0±3.8	32	81.8±2.9	2	71.1±1.6	81.8±2.0	54.4±3.0	–	–	–	–	–	–	–	–
Verbmobil	4	75.4±2.9	64	82.8±2.1	4	72.1±1.7	82.8±1.1	63.9±3.1	–	–	–	–	–	–	–	–
CIEMPIESS+Verbmobil	4	74.8±2.5	64	79.9±3.8	2	71.2±2.1	80.0±2.9	63.0±1.4	–	–	–	–	–	–	–	–
Average	–	75.6±2.6	–	83.0±2.3	–	71.1±2.1	81.9±2.2	60.3±2.9	–	69.8±2.7	–	69.8±3.0	–	70.6±2.8	77.7±2.2	56.3±3.1

Table 3: *Classification of PD vs. HD vs. HC with each speech dimension and their fusion.* ***UAR****: Unweighted Average Recall,* ***M****: Number of Gaussian components. Values reported in terms of mean ± standard deviation.*

	Monologue								Pataka							
UBM	**Articulation**		**Phonation**		**Prosody**		**Fusion**	**PCA**	**Articulation**		**Phonation**		**Prosody**		**Fusion**	**PCA**
	M	**UAR (%)**	**M**	**UAR (%)**	**M**	**UAR (%)**	**UAR (%)**	**UAR (%)**	**M**	**UAR (%)**	**M**	**UAR (%)**	**M**	**UAR (%)**	**UAR (%)**	**UAR (%)**
German (HC)	4	67.1±1.5	64	63.7±2.6	2	60.9±1.8	**71.0±1.4**	48.2±2.3	2	52.5±2.2	32	61.9±2.2	2	54.3±1.8	61.0±2.1	41.5±1.4
Spanish (HC)	4	65.0±2.4	64	62.8±2.2	4	61.2±3.2	67.9±1.6	35.1±2.3	2	56.7±1.3	8	58.7±1.0	2	59.4±2.5	64.0±1.6	43.3±1.5
German-Spanish (HC)	2	63.0±3.0	64	64.1±0.2	4	58.5±1.3	67.1±2.6	38.9±2.4	2	55.9±2.0	8	58.3±2.7	2	57.3±3.1	**64.6±1.5**	40.0±1.6
German (HC+PD)	4	65.1±1.6	128	64.9±1.7	2	57.7±2.2	67.7±2.6	49.7±2.9	2	52.9±1.2	16	61.1±2.2	2	55.0±1.7	59.6±2.1	39.8±1.5
Spanish (HC+PD)	4	66.3±2.2	128	68.1±1.9	4	56.2±2.4	70.0±1.4	47.6±2.2	2	57.8±1.4	8	60.3±1.8	2	55.6±2.6	63.0±1.4	45.7±1.1
German-Spanish (HC+PD)	4	67.6±1.1	32	66.0±1.2	4	58.3±2.4	70.4±1.9	37.3±5.6	2	55.6±1.3	8	60.2±1.8	2	55.2±4.2	62.4±2.4	44.2±1.9
CIEMPIESS	2	61.0±1.2	32	64.9±2.6	4	57.6±2.4	67.1±1.8	33.5±0.9	–	–	–	–	–	–	–	–
Verbmobil	4	62.7±1.6	64	62.5±1.3	2	57.3±2.0	71.0±2.6	43.7±3.1	–	–	–	–	–	–	–	–
CIEMPIESS+Verbmobil	2	60.9±3.2	32	67.9±1.3	2	61.0±2.6	70.1±1.0	37.2±4.5	–	–	–	–	–	–	–	–
Average	–	64.3±2.0	–	65.0±1.7	–	58.7±2.3	69.2±1.9	41.2±2.9	–	55.2±1.6	–	60.1±2.0	–	56.1±2.7	62.4±1.9	42.4±1.5

fusion of the three speech dimensions. The highest UAR for the monologue was 71% with the UBM built with the HC subjects of the German database. The same result was obtained with the UBM based on the Verbmobil database (one of the few cases in which the use of larger UBM resulted in high UAR). For the /pa-ta-ka/ task, the best result was also obtained with the combination of the three dimensions with an UAR of 64.6%. Therefore, it is possible to conclude that these speech dimensions are complementary, so it is necessary to include all of them to obtain higher classification accuracies when discriminating both pathologies from the healthy population. When we analyzed each speech dimension, we could observe that the phonation is the most discriminative one with average UARs of 65% and 60.1% for monologue and /pa-ta-ka/ task, respectively. This is consistent with the above mentioned where it was pointed out that the phonatory ability of patients is largely impaired as the neurodegenerative disease progresses.

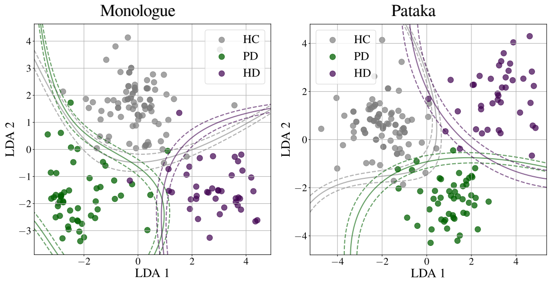

Figure 2: *Visualization of the samples after applying LDA.*

Figure 2 shows a representation of each group (only the best results of Table 3 are considered) created by concatenating the 3 supervectors of articulation, phonation, and prosody, and performing a Linear Discriminant Analysis (LDA) to reduce the dimension of the resulting matrix to 2 dimensions (LDA1 - LDA2). Notice that there are three clusters clearly distinguishable in both figures. Kruskal-Wallis tests over the 2 final dimensions showed that exists a significant difference between the medians of each population. Notice also that the confusion between the HD and the other speakers is minimal, while there is more overlap between the HC subjects and PD patients, this is likely due to the fact that most PD patients who participated in this study were in an early stage of the disease.

5. Conclusions

In this paper, we created GMM-Supervectors with features extracted from three speech dimensions (articulation, phonation, and prosody) and their fusion to perform two classification scenarios: PD vs. HD; and PD vs. HD vs. HC. In this work, it was possible to observe and associate that the phonation dimension is fundamental for the discrimination between both pathologies, especially in continuous speech. For other scenarios, such as rapid /pa-ta-ka/ repetition, we observed that the fusion of the three dimensions obtained the best results, both in the bi-class and multi-class classification. This allows concluding that the three dimensions are relevant and actually they are complementary. Another interesting pattern observed in the experiments presented in this study is that the use of larger datasets to create the UBMs does not result in better results. No clear patterns were observed regarding using different languages in the UBM to create the GMM-Supervectors. We consider that the paper has some limitations, such as data privacy and using other classifiers to evaluate both pathologies. However, we believe that since this study is a good baseline for other methodologies like those based on deep neural networks. In this regard, our future research will include the use of convolutional neural networks and transfer learning between languages and pathologies.

6. Acknowledgment

This study was funded by UdeA grant # ES92210001-IAPFI23-1-01 and National Institute for Neurological Research (Programme EXCELES, ID Project No. LX22NPO5107 - Funded by the European Union – Next Generation EU).

7. References

[1] A. Schapira *et al.*, "Slowing of neurodegeneration in Parkinson's disease and huntington's disease: future therapeutic perspectives," *The Lancet*, vol. 384, no. 9942, pp. 545–555, 2014.

[2] J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *Journal of neurology, neurosurgery & psychiatry*, vol. 79, no. 4, pp. 368–376, 2008.

[3] S. Pinto *et al.*, "Treatments for dysarthria in Parkinson's disease," *The Lancet Neurology*, vol. 3, no. 9, pp. 547–556, 2004.

[4] J. Logemann *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*, vol. 43, no. 1, pp. 47–57, 1978.

[5] C. Saldert *et al.*, "Comprehension of complex discourse in different stages of Huntington's disease," *International journal of language & communication disorders*, vol. 45, no. 6, pp. 656–669, 2010.

[6] L. Hartelius *et al.*, "Speech disorders in mild and moderate Huntington disease: Results of dysarthria assessments of 19 individuals," *Journal of Medical Speech-Language Pathology*, vol. 11, no. 1, pp. 1–15, 2003.

[7] F. Cavallieri *et al.*, "Dopaminergic treatment effects on dysarthric speech: acoustic analysis in a cohort of patients with advanced Parkinson's disease," *Frontiers in Neurology*, vol. 11, p. 616062, 2021.

[8] J. Rusz *et al.*, "Dualistic effect of pallidal deep brain stimulation on motor speech disorders in dystonia," *Brain Stimulation*, vol. 11, no. 4, pp. 896–903, 2018.

[9] C. Ludlow, N. Connor, and C. Bassich, "Speech timing in Parkinson's and Huntington's disease," *Brain and language*, vol. 32, no. 2, pp. 195–214, 1987.

[10] I. Hertrich and H. Ackermann, "Acoustic analysis of speech prosody in Huntington's and Parkinson's disease: a preliminary report," *Clinical linguistics & phonetics*, vol. 7, no. 4, pp. 285–297, 1993.

[11] J. Hlavnička *et al.*, "Characterizing vocal tremor in progressive neurological diseases via automated acoustic analyses," *Clinical Neurophysiology*, 2020.

[12] J. Rusz *et al.*, "Guidelines for speech recording and acoustic analyses in dysarthrias of movement disorders," *Movement Disorders*, vol. 36, no. 4, pp. 803–814, 2021.

[13] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital signal processing*, vol. 10, no. 1-3, pp. 19–41, 2000.

[14] J. R. Orozco-Arroyave *et al.*, "New spanish speech corpus database for the analysis of people suffering from Parkinson's disease," in *Proceedings of LREC*, 2014, pp. 342–347.

[15] S. Skodda, W. Visser, and U. Schlegel, "Vowel articulation in Parkinson's disease," *Journal of Voice*, vol. 25, no. 4, pp. 467–472, 2011.

[16] C. D. Mena and A. Camacho, "Ciempiess: A new open-sourced mexican spanish radio corpus," in *Proceedings of LREC*, 2014, pp. 371–375.

[17] W. Wahlster, *Verbmobil: foundations of speech-to-speech translation*. Springer Science & Business Media, 2013.

[18] J. Rusz, *Detecting speech disorders in early Parkinson's disease by acoustic analysis*. Habilitation thesis, Czech Technical University in Prague, 2018.

[19] J. Rusz *et al.*, "Characteristics and occurrence of speech impairment in Huntington's disease: possible influence of antipsychotic medication," *Journal of Neural Transmission*, vol. 121, no. 12, pp. 1529–1539, 2014.

[20] K. Kieburtz *et al.*, "Unified Huntington's disease rating scale: reliability and consistency," *Neurology*, vol. 11, no. 2, pp. 136–142, 2001.

[21] J. R. Orozco-Arroyave, *Analysis of speech of people with Parkinson's disease*. Logos-Verlag, 2016, vol. 41.

[22] J. R. Orozco-Arroyave *et al.*, "Neurospeech: An open-source software for Parkinson's speech analysis," *Digital Signal Processing*, vol. 77, pp. 207–221, 2018.

[23] J. C. Vásquez-Correa *et al.*, "Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease," *Journal of communication disorders*, vol. 76, pp. 21–36, 2018.

[24] N. Dehak, P. Dumouchel, and P. Kenny, "Modeling prosodic features with joint factor analysis for speaker verification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 7, pp. 2095–2103, 2007.

[25] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[26] J.-L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains," *IEEE transactions on speech and audio processing*, vol. 2, no. 2, pp. 291–298, 1994.

[27] J. Rusz *et al.*, "Objective acoustic quantification of phonatory dysfunction in Huntington's disease," *PLoS One*, vol. 8, no. 6, p. e65881, 2013.

[28] T. Arias-Vergara, J. C. Vásquez-Correa, and J. R. Orozco-Arroyave, "Parkinson's disease and aging: analysis of their effect in phonation and articulation of speech," *Cognitive Computation*, vol. 9, no. 6, pp. 731–748, 2017.