MULTI-VIEW REPRESENTATION LEARNING VIA GCCA FOR MULTIMODAL ANALYSIS OF PARKINSON'S DISEASE

J. C. Vásquez-Correa[1,2], *J. R. Orozco-Arroyave*[1,2], *R. Arora*[3], *E. Nöth*[2]
N. Dehak[3], *H. Christensen*[4], *F. Rudzicz*[5] *T. Bocklet*[6], *M. Cernak*[7],
H. Chinaei[5], *J. Hannink*[2], *Phani Sankar Nidadavolu*[3], *M. Yancheva*[5], *A. Vann*[8], *N. Vogler*[9]

[1]University of Antioquia UdeA, Colombia.[2]Friedrich Alexander Universität, Germany.
[3]Johns Hopkins University, USA., [4]University of Sheffield, UK., [5]University of Toronto, Canada
[6]Intel, Germany. [7]Idiap Research Institute, Switzerland.
[8]Stanford University, USA; [9]University of California-Irvine, USA.

⋆corresponding: jcamilo.vasquez@udea.edu.co

ABSTRACT

Information from different bio–signals such as speech, handwriting, and gait have been used to monitor the state of Parkinson's disease (PD) patients, however, all the multimodal bio–signals may not always be available. We propose a method based on multi-view representation learning via generalized canonical correlation analysis (GCCA) for learning a representation of features extracted from handwriting and gait that can be used as a complement to speech–based features. Three different problems are addressed: classification of PD patients vs. healthy controls, prediction of the neurological state of PD patients according to the UPDRS score, and the prediction of a modified version of the Frenchay dysarthria assessment (m-FDA). According to the results, the proposed approach is suitable to improve the results in the addressed problems, specially in the prediction of the UPDRS, and m-FDA scores.

Index Terms— Parkinson's disease, Multi-view learning, GCCA, Speech processing, Handwriting processing, Gait processing, UPDRS, Frenchay dysarthria assessment.

1. INTRODUCTION

Parkinson's disease (PD) is a neurological disorder characterized by the progressive loss of dopaminergic neurons in the midbrain producing several motor and non-motor impairments [1]. The motor symptoms include, among others, bradykinesia, rigidity, resting tremor, micrographia, and different speech impairments. The progression of the disease in the motor capabilities is currently evaluated with the third section of the movement disorder society, unified Parkinson's disease rating scale (MDS-UPDRS-III) [2], which is assigned by neurologist experts. The scale contains several items to evaluate the motor capabilities of the patients such as finger tapping, gait, speech, and facial expression. For simplicity, in the rest of the paper, the scale will be referred as UPDRS.

Several studies have analyzed different bio–signals such as speech, gait, and handwriting to monitor the state of the PD patients. In [3] the authors investigated the suitability of features extracted from sustained vowel phonations to evaluate whether the speech is "acceptable" or "unacceptable" in order to assess the rehabilitation treatment of PD patients. The authors reported accuracies close to 90% on a dataset with utterances from 14 PD patients. In [4] the authors evaluated the classification of emotional speech of 5 PD patients using different acoustic features, and reported an accuracy of 65.5%. In the "2015 computational paralinguistic challenge (ComParE)" [5] one of the sub-challenges was to predict the UPDRS score of PD patients. Utterances from 61 PD patients were used, and a Spearman's correlation coefficient of 0.39 was obtained as baseline in the test set. The winner of the challenge [6] grouped the patients tasks automatically and used deep neural networks and Gaussian processes to predict the UPDRS score. They reported a correlation coefficient of 0.65. In [7], the authors analyzed features related to articulation and intelligibility to predict the UPDRS score of 50 PD patients, and reported a Spearman's correlation of up to 0.72 in a similar scheme as part of the ComParE 2015 challenge. For the gait analysis, in [8], the authors analyzed the influence of gait features to classify PD vs. healthy controls (HC) subjects. The gait signals are obtained from accelerometers and gyroscopes attached to the shoes of the patients [9]. The authors reported an accuracy of up to 81% for the detection of PD. In [10], the authors used several inertial sensors attached to the lower and higher limbs with the aim of predicting the UPDRS score of PD patients. They computed features related to the stance time, the length of the stride, and the velocity of each step, and reported a correlation coefficient of 0.60 in a dataset formed with signals from 34 PD patients. For handwriting analysis, in [11], the authors assessed whether handwriting measures extracted from the pen trajectory can be used to classify PD vs. HC participants. The

computed features include the velocity and the acceleration of the pen, and the time and the size of the drawing. The authors performed statistical tests and found significant differences in the features extracted from the HC and PD subjects. In [12] the authors used features related to the velocity, acceleration and jerk of the pen, the on-surface time, and the average pressure to classify PD vs. HC subjects. The authors reported an accuracy of up to 81% in a database with drawings of 37 PD patients and 38 HC.

To the best of our knowledge, the different modalities (speech, gait and handwriting) have not been analyzed together to monitor the state of PD patients so far. The main reason is because the information from the three modalities cannot be guaranteed to always be available at both training and run-time. In this study, a multimodal monitoring of PD patients using features from speech, handwriting, and gait is performed. As the information from multiple modalities is not always available, we apply a method for multi-view learning based on the generalized canonical correlation analysis (GCCA). This method allows the transformation and representation of features from different modalities into a different feature space, where only one modality is available [13, 14].

Three experiments are performed with the proposed approach: (1) classification of PD vs. HC, (2) the prediction of the UPDRS score of the patients, and (3) the prediction of a perceptual scale designed to assess only the speech impairments of the speakers. The perceptual scale is a modification of the Frenchay dysarthria assessment (FDA) [15]. According to the results, gait is the most suitable modality to predict the UPDRS score, and the use of GCCA improves the results obtained in the three performed experiments.

2. METHODS

2.1. Feature extraction

Features from Speech– Two different feature sets are considered for speech analysis: The first one is formed with articulation–based features, and includes 86 features such as the energy content in the Bark scale in the transition from voiced to unvoiced segments (22 features), and from unvoiced to voiced segments (22 features) [7]. The feature set is completed with the first and second formant frequencies, and 12 MFCC with their derivatives. The extracted features are grouped and four functionals are computed (mean, standard deviation, skewness, and kurtosis), forming a 344-dimensional feature vector per utterance. The second feature set from speech contains prosody–based features computed with the Erlangen prosody module [16], using voiced segments as speech unit. The set of features comprises a total of 95 features. 19 of them are based on duration and include among others the number and the length of voiced frames, and duration of pauses. 36 of the features are based on the F_0 contour, including the mean, standard deviation, jitter, and others. The energy-based features include measures of the energy within the voiced frames, shimmer, position of the maximum energy, and others. The features are grouped into one feature vector and four functionals are also computed: mean, standard deviation, maximum, and minimum, forming a 380-dimensional feature vector per utterance.

Features from Handwriting– A total of 21 features are considered. They are based on kinematics of the x and y position, and the pressure of the pen. The feature set includes the speed of the stroke, velocity, acceleration, and jerk of the trajectory, and the average pressure of the pen.

Features from Gait– This feature set is formed with 12 bio-mechanical features including the swing, stance and stride times, the length of the stride, the maximum toe clearance, the gait velocity, the cadence, the rotation angle at stance, and the number of recognized strides. The mean and standard deviation of the features per each stride are computed, forming a 24-dimensional feature vector.

2.2. Multi-view representation learning

The multi-view learning is performed using GCCA, with the aim of obtaining a feature embedding that represents the maximally correlated projection from the multimodal information and the speech respectively. This projected feature space can be used even when the multimodal information is not available. Let $X_j \in \mathbb{R}^{N\times d_j}$ be the mean centered feature matrix from the modality j, d_j the number of features of the modality j, N the number of subjects in the data, and k the number of components of the representation matrix. The GCCA process can be expressed according to the following optimization problem:
Find $G \in \mathbb{R}^{N\times k}$ and $U_j \in \mathbb{R}^{d_j\times k}$ that solve:

$$\underset{G,U_j}{\arg\min} \sum_{j=1}^{J} \|G - X_j U_j\|_F^2 \text{ s.t. } G^T G = I \qquad (1)$$

G is our original representation matrix, and U_j corresponds to the transformation matrix of the modality j. In the optimization problem, a projection matrix $\tilde{P}_j \in \mathbb{R}^{N\times N}$ is defined according to Equation 2, which is regularized by adding the parameter r_j before doing the inversion for numerical stability [14].

$$\tilde{P}_j = X_j(X_j^T X_j + r_j I)^{-1} X_j^T \qquad (2)$$

The parameters r_j are optimized in a range from 10^{-8} to 10^0 into powers of 10, The other parameter to be optimized is the number of components k of the representation matrix ($k \in \{5, 10, 15, 20\}$). The optimization is performed using a KNN regressor following a 6–fold cross-validation on the training set. Finally, we obtain a representation matrix based on the multimodal information, which can be stacked to the features from speech that are available in the test data.

2.3. Classification and regression

Three tasks are performed to analyze the suitability of the proposed method: (1) The classification of PD vs. HC subjects, (2) the prediction of the UPDRS score of the PD patients, and (3) the prediction of the m-FDA score to evaluate the level of dysarthria of the participants. For the classification we use an SVM with a Gaussian kernel. The parameters C and γ are optimized in a grid search, with selection criterion based on the accuracy obtained in the train set ($C \in \{10^{-5}, 10^{-4}, \dots 10^{4}\}$ and $\gamma \in \{10^{-6}, 10^{-5}, \dots 10^{2}\}$). For the regression problems we use an SVR with an ε-insensitive loss function and a linear kernel. The parameters of the regressor C and ε are optimized in a grid-search with $C \in \{10^{-4}, 10^{-3}, \dots, 100\}$ and $\varepsilon \in \{10^{-4}, 10^{-3}, \dots, 10, 20\}$. The performance is evaluated using the Spearman's correlation coefficient between the predicted values and the real scores. For the three tasks a leave-one-subject-out (LOSO) cross-validation is performed.

3. DATA

3.1. Training Data

The training data contains recordings from speech in Spanish, online handwriting and gait from 30 PD patients labeled according to the UPDRS score. The speech signals were recorded with a sampling frequency of 44.1 kHz and 16-bit resolution, using an omnidirectional microphone, and a portable sound-proof booth. The speech tasks include the repetition of /pa-ta-ka/, a read text, and a monologue. The handwriting data consists of online drawings that are obtained with a tablet Wacom cintiq 13-HD [17] with a sampling frequency of 180 Hz. The tablet captures six different signals: x-position, y-position, in-air movement, azimuth, altitude, and pressure. The patients performed 14 tasks including among others draw a cube, their ID number, their name, Rey-Osterrieth figure [18], and a spiral. The acquisition of the gait signals was performed using the embedded gait analysis using intelligent technology (eGaIT) platform [9]. The system consists of inertial sensors (three axes gyroscopes and accelerometers) attached to the lateral heel of a shoe [8]. Data from both foot was captured with sampling rate of 102 Hz. The tasks performed by the patients include among others 20 meters walking with a stop at 10 meters (2×10), and 40 meters walking with a stop every 10 meters (4×10).

3.2. Test Data

Spanish– We consider the PC-GITA database [19], which contains speech utterances from 50 PD and 50 HC Colombian native speakers, recorded in a sound-proof booth with a sampling frequency of 44.1 kHz and with the same microphone used in the training data. The patients were labeled also according to the UPDRS score by the same neurologist expert than the training data.

German– The German data contains recordings from 88 PD and 88 HC subjects. The speakers perform different speech tasks, including the repetition of /pa-ta-ka/, a read text, and a monologue [20].

Czech– The Czech data is formed with recordings from 20 PD patients and 15 HC. The patients were newly diagnosed with PD, and none of them had been medicated before or during the recording session. The speech tasks performed by the speakers include also the repetition of /pa-ta-ka/, a read text and a monologue [20].

3.3. Modified Frenchay dysarthria assessment

Additionally to the UPDRS score, the training and test subjects in Spanish were labeled by three phoniatricians according to a modified version of the FDA [15]. The main aim was to evaluate only the speech impairments that the PD patients develop. The original version of the FDA needs the patient to be with the examiner. We introduced a modified version that considers only the speech recordings and evaluates 13 items including among others the movements of the lips, larinx, palate and tongue, the respiration, and the intelligibility. The evaluation of each item ranges from 0 to 4, for a total range from 0 to 52 (0 normal, and 52 completely dysarthric). The three phoniatricians agreed in the first ten evaluations, and then performed the evaluation of the other recordings. The inter-rater reliability among the labelers is 0.75. The median among the labels of the three evaluators was considered as the label of the speaker.

4. RESULTS AND DISCUSSION

Table 1 shows the results obtained with each modality separately for the prediction of the UPDRS and the m-FDA scores. The highest correlation w.r.t the UPDRS is obtained with gait (0.72), followed by handwriting, and speech. The results predicting the m-FDA are higher than those obtained predicting the UPDRS using only the speech modality. Considering these results we decided to choose the articulation–based features in /pa-ta-ka/ and the prosody–based features in the read text to train the GCCA using the speech modality. For the gait and the handwriting modalities, we selected the features extracted from the 4×10 both foot and the cube respectively, to train the GCCA approach. Table 2 contains the results for the test data where only the speech information is available. These results are used as baseline for the GCCA. The classification accuracy ranges from 69% to 82%, depending on the language and the feature set. The highest correlation predicting the UPDRS score is obtained with the prosody–based features in the read text (0.41). The results predicting the m-FDA score range from 0.39 to 0.67, where the best result is obtained with the repetition of /pa-ta-ka/ using the articulation–based features. After training the GCCA, the learnt features are stacked to the baseline features. The results with these

Features and tasks	# of Feat.	Pediction of UPDRS	Prediction of m-FDA
Speech Modality			
Art.–/pa-ta-ka/	344	-0.33	0.40
Art.–monologue	344	-0.39	0.19
Art.–read text	344	-0.19	0.13
Pros.–monologue	380	-0.23	0.22
Pros.–read text	380	-0.11	-0.31
Gait Modality			
4x10–left	24	0.68	0.49
4x10–right	24	0.66	0.32
4x10–both	24	0.72	0.42
2x10–left	24	0.72	0.39
2x10–right	24	0.59	0.42
2x10–both	24	0.66	0.38
Handwriting Modality			
Cube	21	0.48	-0.18
ID	21	0.47	0.25
Name	21	0.20	0.20
Digits	21	0.32	0.36
Rey-Osterrieth	21	0.44	-0.22
Spiral	21	0.12	-0.22

Table 1. Baseline results for the multimodal database. Art: articulation, Pros: prosody

Features and tasks	# of Feat.	Classif. Acc.	Pediction of UPDRS	Prediction of m-FDA
Speech Modality in Spanish				
Art.–/pa-ta-ka/	344	77%	0.34	0.67
Art.–monologue	344	70%	0.32	0.39
Art.–read text	344	78%	0.28	0.56
Pros.–monologue	380	69%	-0.43	0.41
Pros.–read text	380	69%	0.41	0.39
Speech Modality in German				
Art.–/pa-ta-ka/	344	70%	0.11	-
Art.–monologue	344	73%	0.01	-
Art.–read text	344	79%	0.03	-
Pros.–monologue	380	76%	-0.69	-
Pros.–read text	380	77%	0.31	-
Speech Modality in Czech				
Art.–/pa-ta-ka/	344	82%	0.29	-
Art.–monologue	344	77%	-0.51	-
Art.–read text	344	80%	-0.59	-
Pros.–monologue	380	69%	0.00	-
Pros.–read text	380	80%	0.59	-

Table 2. Baseline results for the test databases. Art: articulation, Pros: prosody.

new feature sets are shown in Table 3. Improvements in a range from 1% to 3% are observed in the classification task, depending on the language and on the feature set. For the prediction of the UPDRS score, high improvements are observed in Spanish and Czech, which indicates that the GCCA–based transformed features from gait and handwriting are representing additional information that help to predict the neurological state of the patients. Finally, for the prediction of the m-FDA score in the Spanish database, most of the results are improved using the GCCA, specially those obtained with the

Features and tasks	# of Feat.	Classif. Acc.	Pediction of UPDRS	Prediction of m-FDA
Speech Modality in Spanish				
Art.–/pa-ta-ka/	364	**78%**	**0.40**	**0.72**
Art.–monologue	364	73%	0.30	0.40
Art.–read text	364	78%	**0.39**	0.59
Pros.–monologue	400	70%	0.14	0.40
Pros.–read text	400	71%	0.41	0.39
Speech Modality in German				
Art.–/pa-ta-ka/	364	71%	0.14	-
Art.–monologue	364	74%	-0.03	-
Art.–read text	364	76%	0.03	-
Pros.–monologue	400	76%	-0.69	-
Pros.–read text	400	76%	0.40	-
Speech Modality in Czech				
Art.–/pa-ta-ka/	364	82%	**0.46**	-
Art.–monologue	364	77%	0.12	-
Art.–read text	364	80%	-0.59	-
Pros.–monologue	400	69%	**0.51**	-
Pros.–read text	400	77%	0.60	-

Table 3. Results for the test databases using GCCA. Art: articulation, Pros: prosody.

articulation–based features in /pa-ta-ka/.

5. CONCLUSION

A method based on GCCA is applied to map features from three modalities (speech, gait, and handwriting) into a different dataset that contains only features from one modality (speech). The suitability of this approach is tested in three databases for PD analysis in different languages. Three problems are addressed: the classification of PD vs. HC, the prediction of the UPDRS score, and the prediction of the dysarthric level of the patients (m-FDA score). An improvement in the performance of the three tasks is observed, specially in the prediction of the UPDRS score in Spanish and Czech, and the prediction of the m-FDA score in Spanish. These results indicate that the proposed approach is suitable to map the features from other modalities that are not always available, providing additional information for the PD analysis, including the cases when the language of the test data is different. For the separately analysis of each modality, the best results predicting the UPDRS score of PD patients is obtained with gait, followed by handwriting, and speech. Further experiments may be performed with new features from gait and handwriting with the aim of improving the results.

6. ACKNOWLEDGMENTS

This work was started at 2016 Jelinek Memorial Summer Workshop on Speech and Language Technologies, which was supported by JHU via DARPA LORELEI, Microsoft, Amazon, Google, and Facebook. Thanks also to COLCIENCIAS project # 111556933858, and CODI from University of Antioquia.

7. REFERENCES

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

[2] C. G. Goetz and 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. A. Little, C. Fox, and L. O. Ramig, "Objective automatic assessment of rehabilitative speech treatment in Parkinson's disease," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 181–190, 2014.

[4] S. Zhao, F. Rudzicz, L. G Carvalho, C. Márquez-Chin, and S. Livingstone, "Automatic detection of expressed emotion in Parkinson's disease," in *39th International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, 2014, pp. 4813–4817.

[5] 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, Parkinsons & eating condition," in *16th Annual Conference of the Speech and Communication Association (INTERSPEECH)*, 2015, pp. 478–482.

[6] T. Grósz, R. Busa-Fekete, G. Gosztolya, and L. Tóth, "Assessing the degree of nativeness and Parkinsons condition using Gaussian processes and deep rectifier neural networks," in *16th Annual Conference of the Speech and Communication Association (INTERSPEECH)*, 2015, pp. 1339–1343.

[7] J. R. Orozco-Arroyave, J. C. Vásquez-Correa, F. Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, and E. Nöth, "Towards an automatic monitoring of the neurological state of the Parkinson's patients from speech," in *41st International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, 2016.

[8] J. Klucken, J. Barth, P. Kugler, J. Schlachetzki, T. Henze, F. Marxreiter, Z. Kohl, R. Steidl, J. Hornegger, and B. Eskofier, "Unbiased and mobile gait analysis detects motor impairment in Parkinson's disease," *PloS one*, vol. 8, no. 2, pp. e56956, 2013.

[9] "eGaIT - embedded Gait analysis using Intelligent Technology, `http://www.egait.de/`," 2011.

[10] F. Parisi, G. Ferrari, M. Giuberti, L. Contin, V. Cimolin, C. Azzaro, G. Albani, and A. Mauro, "Body-sensor-network-based kinematic characterization and comparative outlook of UPDRS scoring in leg agility, sit-to-stand, and gait tasks in Parkinson's disease," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 6, pp. 1777–1793, 2015.

[11] E. J. Smits, A. J. Tolonen, L. Cluitmans, M. van Gils, B. A. Conway, R. C. Zietsma, K. L. Leenders, and N. M. Maurits, "Standardized handwriting to assess bradykinesia, micrographia and tremor in Parkinson's disease," *PloS one*, vol. 9, no. 5, pp. e97614, 2014.

[12] P. Drotár, J. Mekyska, I. Rektorová, L. Masarová, Z. Smékal, and M. Faundez-Zanuy, "Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease," *Artificial intelligence in Medicine*, vol. 67, pp. 39–46, 2016.

[13] S. Bharadwaj, R. Arora, K. Livescu, and M. Hasegawa-Johnson, "Multiview acoustic feature learning using articulatory measurements," in *Intl. Workshop on Stat. Machine Learning for Speech Recognition*, 2012.

[14] P. Rastogi, B. Van Durme, and R. Arora, "Multiview LSA: Representation learning via generalized CCA," in *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, 2015, pp. 556–566.

[15] P Enderby and R. Palmer, *FDA-2: Frenchay Dysarthria Assessment*, P. Education, 2nd edition, 2008.

[16] V. Zeißler, J. Adelhardt, A. Batliner, C. Frank, E. Nöth, R. P. Shi, and H. Niemann, "The prosody module," in *SmartKom: foundations of multimodal dialogue systems*, pp. 139–152. Springer, 2006.

[17] "Cintiq 13 hd graphic pen `http://www.wacom.com/en-us/products/pen-displays/cintiq-13-hd`,".

[18] M. Shin, S. Park, S. Park, S. Seol, and J. Kwon, "Clinical and empirical applications of the Rey–Osterrieth complex figure test," *Nature Protocols*, vol. 1, no. 2, pp. 892–899, 2006.

[19] 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 *Language Resources and Evaluation Conference, (LREC)*, 2014, pp. 342–347.

[20] 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 different languages," *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 481–500, 2016.