Comparison of User Models Based on GMM-UBM and I-vectors for Speech, Handwriting, and Gait Assessment of Parkinson's Disease Patients

Juan Camilo Vasquez-Correa[1,2], Tobias Bocklet[3], Juan Rafael Orozco-Arroyave[1,2], and Elmar Nöth[1]

[1]Pattern Recognition Lab. Friedrich-Alexander University, Erlangen-Nürnberg, Germany.
[2]Faculty of Engineering. Universidad de Antioquia UdeA, Medellin, Colombia.
[3]Technische Hochschule Nürnberg, Germany.

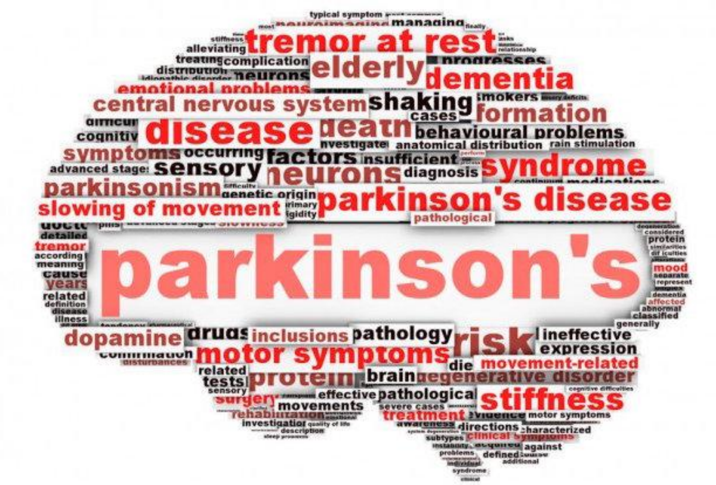
45th International Conference on Acoustics, Speech, and Signal Processing (ICASSP)
4th-8th May 2020

Background: Parkinson’s Disease (PD)

01 Second most prevalent neurological disorder worldwide.

02 Patients develop several motor and non-motor impairments.

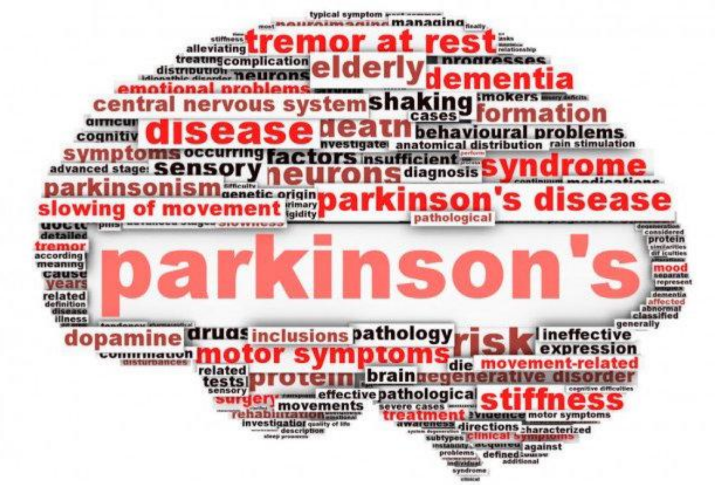
03 The disease affects all subsystems involved in motor activities.



Background: Parkinson’s Disease (PD)

Motor impairments:

- Bradykinesia
- Rigidity
- Resting tremor
- Gait deficits
- Dysarthria

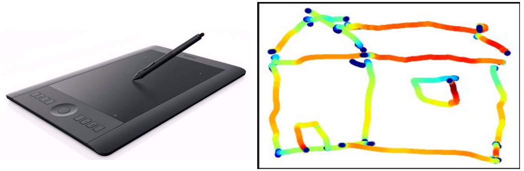


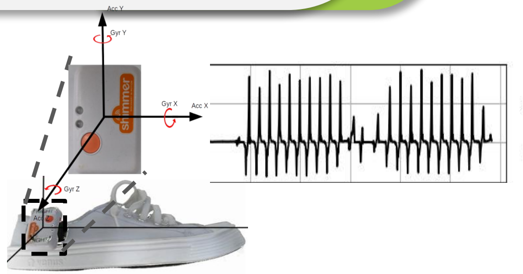
- Evaluated by neurologist experts according to the **MDS-UPDRS-III** scale

Background: Motivation

01 The evaluation of motor symptoms is crucial for clinicians to make decisions about the medication or therapy for the patients.

02 The analysis of signals such as gait, handwriting, and speech helps to assess the motor symptoms of patients.

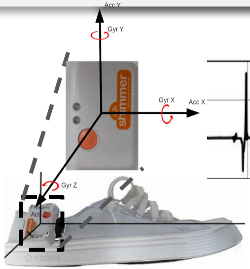


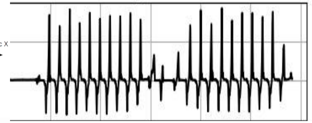


Background: Motivation

01 Most of related studies consider only one modality. Multimodal analyses with information from different sensors have not been extensively studied.

02 Previous studies [1] suggested that the combination of modalities also improved the accuracy in the prediction of the neurological state of the patients.





[1] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, et al., "Multi-view representation learning via GCCA for multimodal analysis of parkinson's disease," in *ICASSP, 2017, pp. 2966–2970.*

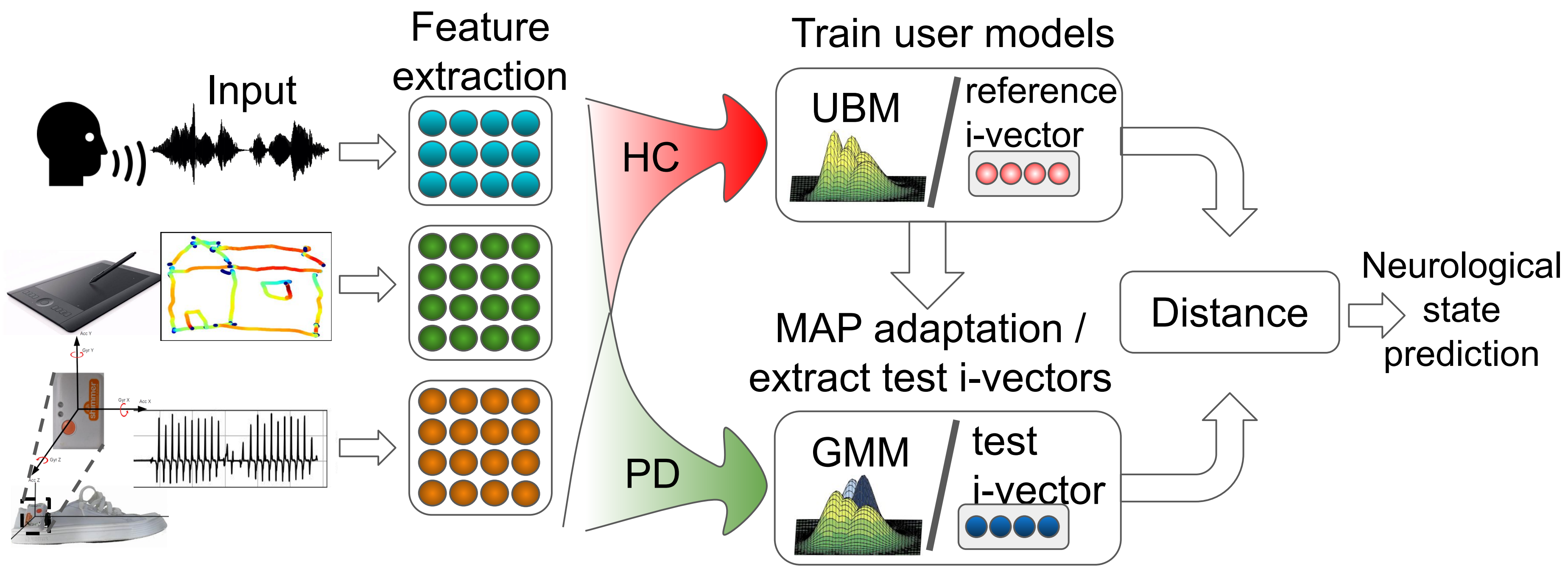
Proposed approach

01 Different features extracted from speech, handwriting, and gait to evaluate the neurological state of PD patients.

02 The prediction is performed with user models based on Gaussian mixture models - universal background models (GMM-UBM) and i-vectors.

03 This is one of the few studies for multimodal analysis of PD patients, and the first one that considers multimodal user models to evaluate the neurological state of the patients.

Methods


Input
Feature extraction
Train user models
HC
PD
UBM
reference i-vector
MAP adaptation / extract test i-vectors
GMM
test i-vector
Distance
Neurological state prediction

Methods: speech features

01 **Phonation**

- Designed to model abnormal patterns in the vocal fold vibration.
- Include: Jitter, shimmer, amplitude perturbation quotient, pitch perturbation quotient and the first and second derivatives of the fundamental frequency.
- Important to model the dysarthria severity of the patients [2].

[2] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, et al., "Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease," in *Journal of communication disorders, 76, 21-36, 2018.*

Methods: speech features

02 Articulation

- Designed to model the difficulties of the patients to start/stop the movement of the vocal folds.
- Based on the energy content in the transition between unvoiced and voiced segments (onset).
- Include: Energy distributed in 22 critical bands according to the Bark scale and 12 MFCCs with their respective first and second derivatives.
- Important to model the dysarthria severity of the patients [2].

[2] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, et al., "Towards an automatic evaluation of the dysarthria level of patients with Parkinson's disease," in *Journal of communication disorders, 76, 21-36, 2018.*

Methods: speech features

03 **Prosody**

- Designed to model monotonicity and monoloudness exhibited by the patients.
- The log-F0 and the log-energy contours of the voiced segments were approximated using a polynomials of order $P = 5$.
- Feature set formed with the coefficients of the polynomials.

Methods: speech features

04 **Phonological**

- Represented by a vector with interpretable information about the placement and manner of articulation.
- The phonological features are the posterior probability of a speech frame to belong to one or more phonological classes, *i.e., stop, nasal, vocalic, labial, etc.*
- 18 different phonological classes, estimated with a parallel bank of recurrent neural networks with Bi-GRU units [3].

[3] J. C. Vasquez-Correa, et al. Phonet: A Tool Based on Gated Recurrent Neural Networks to Extract Phonological Posteriors from Speech. *Interspeech 2019, 549-553.*

Methods: online handwriting features

01 **Kinematic**

- Based on the trajectory of the strokes in vertical, horizontal, radial, and angular positions [4].
- Velocity and acceleration of the strokes in the different axes, in addition to the pressure of the pen, the azimuth angle, the altitude angle, and their derivatives.
- Features based on the in-air movement before the participant put the pen on the tablet's surface.

[4] C. D. Rios-Urrego et al. Analysis and evaluation of handwriting in patients with parkinson's disease using kinematic, geometrical, and non-linear features, *Computer Methods and Programs in Biomedicine*, 173,43–52, 2019.

Methods: gait features

01 Spectral

- Designed to model the spectral wealth and the harmonic structure of the gait signals obtained from the inertial sensors.

- The feature set is formed with:
  - The energy content in 8 frequency bands from the continuous wavelet transform.
  - Three spectral centroids.
  - Energy in the in the 1st, 2nd, and 3rd quartiles of the spectrum.
  - Energy content in the locomotor band (0.5–3 Hz).
  - Energy content in the freeze band (3–8 Hz).
  - Freeze index.

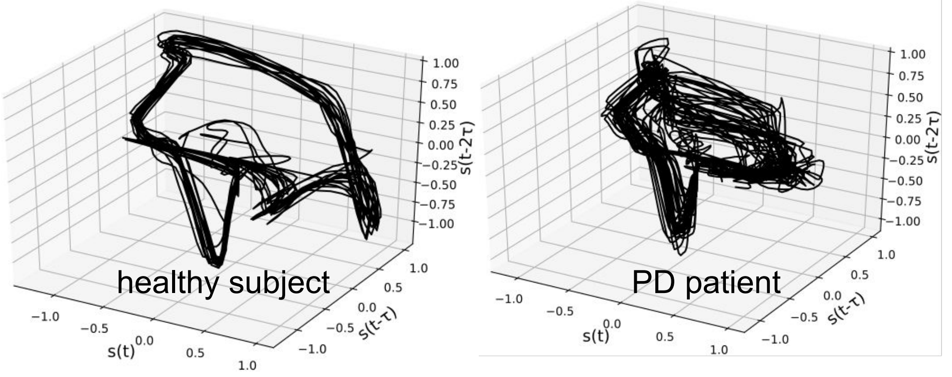
Methods: gait features

02 Non-linear dynamics

- Gait is a complex and non-linear activity that can be modeled with non-linear dynamics features based on reconstructed attractors from the Taken’s theorem.

- Designed to model complexity and stability of the walking process.

- The feature set includes:
  - Correlation dimension.
  - Largest Lyapunov exponent.
  - Hurst exponent.
  - Detrended fluctuation analysis.
  - Sampled entropy.
  - Lempel-Ziv complexity.



Methods: user models

01 GMM-UBM

- Speech, handwriting, or gait impairments of PD patients can be modeled by comparing a GMM adapted for individual patients with an universal model (UBM) created with data from healthy subjects.

- The model for each PD patient is derived from the UBM by adapting its parameters following a maximum a posteriori process.

- The neurological state of the patients is estimated by comparing the adapted model with the UBM using the *Bhattacharyya* distance.

Methods: user models

02 I-vectors

- I-vectors transform the original feature space into a low-dimensional representation called total variability space, which for speech signals models the inter- and intra-speaker variability, in addition to channel effects.

- We aim to capture changes in speech, handwriting, and gait due to the disease.

- I-vector extractor is trained with data from healthy subjects, and reference i-vectors are computed to represent healthy speech, handwriting, or gait.

- I-vectors extracted from PD patients are compared with reference i-vectors using the *cosine* distance.

Data: participants

Speech, handwriting, and gait data collected from:

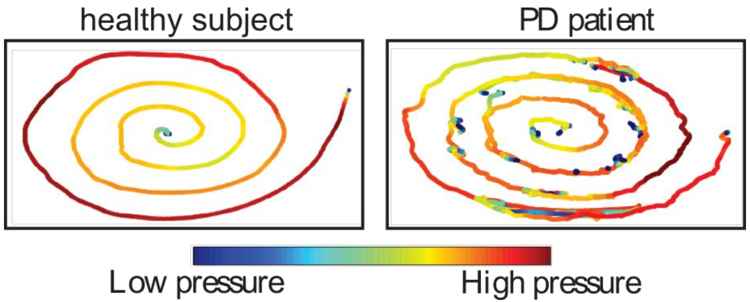
- 106 PD patients and 87 HC, age balanced.
- Colombian Spanish native speakers.
- Most patients in initial or intermediate stage of the disease.
  - MDS-UPDRS-III of 36.2 +/- 18.1.

	PD patients	Healthy subjects
Gender [Female/Male]	49 / 57	43 / 44
Age [Female/Male]	60.9 / 64.7	61.4 / 64.9
Time since diagnosis in years [Female/Male]	15.5 / 8.1	-
MDS-UPDRS-III	36.2 / 36.6	

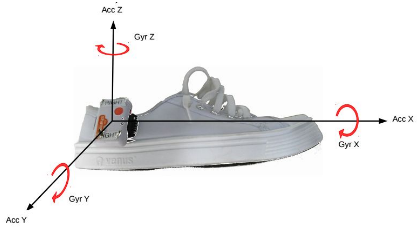
Data: signals

- Speech signals recorded with a sampling frequency of 16 kHz and 16-bit resolution.
  - Diadochokinetic exercises, read sentences, read text, and monologues.

- Online handwriting data collected with a sampling frequency of 180 Hz.
  - Six different signals: x-position, y-position, in-air movement, azimuth, altitude, and pressure.
  - Different writing and drawing shape tasks.



- Gait signals captured with 3D inertial sensors attached to the lateral heel of the shoe (100 Hz, 12-bit resolution).
  - 20 meters walking, heel toe tapping, among others.



Experiments and results

- Data from healthy subjects were used to train the UBMs and the i-vector extractors.
- For the GMM-UBM system, data from PD patients were used to adapt the UBMs into GMMs and the *Bhattacharya* distance is used to compare both models.
- For the i-vectors, a reference was created by averaging the i-vectors extracted from healthy subjects that have same gender and similar age as the patients (in a range of ± 2 years).
- I-vectors extracted from PD patients are compared to the reference i-vector using the *cosine* distance.
- The computed distances are correlated with the MDS-UPDRS-III score of the patients.

Results: User models from different modalities

Modality	Features	GMM-UBM	I-vector
Gait	Spectral	0.62	0.35
	Non-linear	0.31	0.38
Handwriting	Kinematic	0.26	0.35
Speech	Phonation	0.20	0.24
	Articulation	0.20	0.21
	Prosody	0.20	0.19
	Phonological	0.30	0.20

- Results in terms of the Spearman's correlation coefficient.

Results: User models from different modalities

Modality	Features	GMM-UBM	I-vector
Gait	Spectral	**0.62**	0.35
	Non-linear	0.31	0.38
Handwriting	Kinematic	0.26	0.35
Speech	Phonation	0.20	0.24
	Articulation	0.20	0.21
	Prosody	0.20	0.19
	Phonological	0.30	0.20

- Results in terms of the Spearman's correlation coefficient.
- A "strong" correlation (0.62) is obtained with the spectral features (gait analysis) modeled with the GMM-UBM system.

Results: User models from different modalities

Modality	Features	GMM-UBM	I-vector
Gait	Spectral	**0.62**	0.35
	Non-linear	0.31	0.38
Handwriting	Kinematic	0.26	**0.35**
Speech	Phonation	0.20	0.24
	Articulation	0.20	0.21
	Prosody	0.20	0.19
	Phonological	0.30	0.20

- Results in terms of the Spearman’s correlation coefficient.
- A “strong” correlation (0.62) is obtained with the spectral features (gait analysis) modeled with the GMM-UBM system.
- For handwriting features, “weak” correlations are obtained both with the GMM-UBM and the i-vector systems.

Results: User models from different modalities

Modality	Features	GMM-UBM	I-vector
Gait	Spectral	**0.62**	0.35
	Non-linear	0.31	0.38
Handwriting	Kinematic	0.26	**0.35**
Speech	Phonation	0.20	0.24
	Articulation	0.20	0.21
	Prosody	0.20	0.19
	Phonological	0.30	0.20

- Results in terms of the Spearman’s correlation coefficient.
- A “strong” correlation (0.619) is obtained with the spectral features (gait analysis) modeled with the GMM-UBM system.
- For handwriting features, “weak” correlations are obtained both with the GMM-UBM and the i-vector systems.
- Speech features are not robust enough to model the neurological state of the patients.

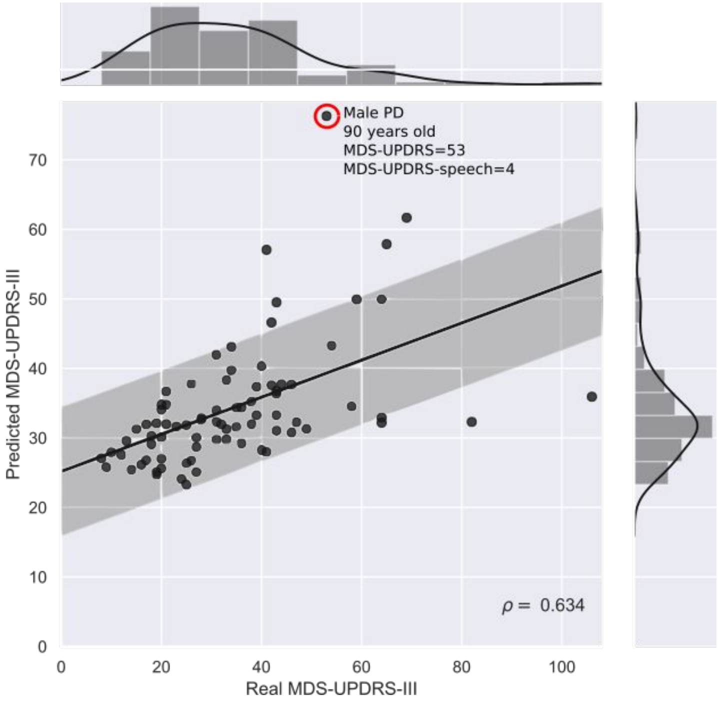
Results: Multimodal user models

	Spearman's correlation	MAE
Fusion of modalities	**0.64**	10.5

MAE: Mean absolute error

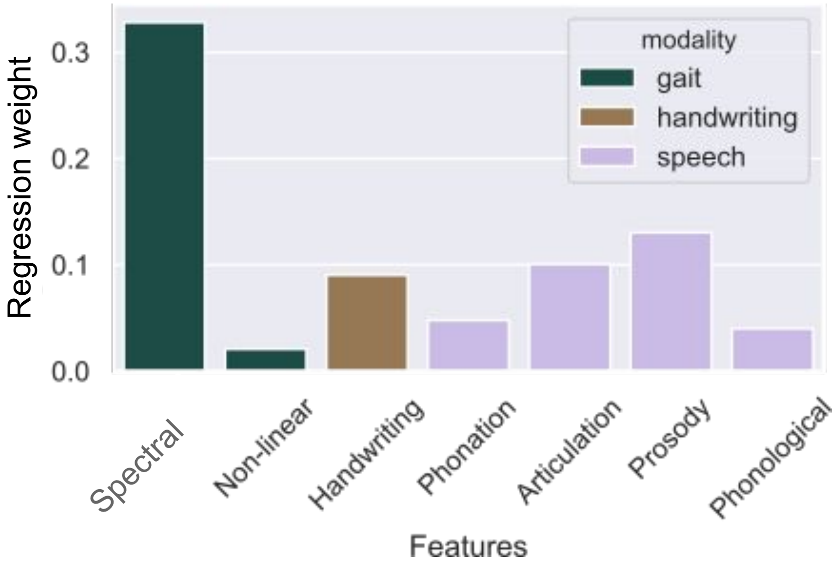
- The user models from the GMM-UBM systems were combined by concatenating the distance between the user model and the UBM for each feature set.
- A linear regression was trained with the matrix of distances to predict the MDS-UPDRS-III scale, following a leave one out cross-validation strategy.
- The Spearman's correlation increases by 2.4%, absolute with respect to the one obtained only with the gait features

Results: Multimodal user models



- Error in the prediction of the MDS-UPDRS-III of the patients.
- Most of the patients are in initial or intermediate state of the disease, and they were predicted with the same distribution.
- The outlayer corresponds to the oldest patient in the corpus.

Results: Multimodal user models



- Contribution of each feature set to the multimodal user model.
- Spectral gait features were the most important for the multimodal model, followed by prosody and articulation features.
- Handwriting features were less important than expected.

Conclusion

01 We compared GMM-UBM and i-vector systems to evaluate the neurological state of PD patients using speech, handwriting, and gait.

02 Gait features were the most accurate to model the general neurological state of the patients.

03 The combination of different bio-signals improved the correlation of the proposed method.

Further studies

01 Increase the data from healthy subjects, to develop better reference models

02 Additional features to model other aspects of PD symptoms, especially from handwriting signals.

03 Additional fusion methods can be considered to evaluate the neurological state of the patients.

Thank you for your attention, Questions?

Camilo Vasquez
Pattern Recognition Lab, Department of Computer Science,
Friedrich-Alexander University Erlangen-Nürnberg,
Erlangen, Germany

juan.vasquez@fau.de

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Marie Sklodowska-Curie grant agreement No 766287.