Speaker models for monitoring Parkinson's disease progression considering different communication channels and acoustic conditions by T. Arias-Vergara et al.

Review by Manila Kodali

Contents

Part-1

Motivation

Part-2

Dataset & Methods

Part-3

Results

1. Motivation

- People suffering from PD are characterized by the progressive loss of dopaminergic neurons in the midbrain.[1]
- Rely on medical history, physical and neurological examinations to assess the patients.
- The symptoms progress differently among patients.
- The most suitable methods to perform continuous monitoring of the symptoms are based on computer-aided tools.
- Assessing the neurological state of PD patients from speech signals always consider situations where the acoustic conditions are relatively controlled, i.e., quiet rooms, good/expensive microphones, and direct connection to the recording device.[2]

1.1 Contribution

- The proposed approach overcomes the state-of-the-art in several aspects:

1. The method is based on individual models
2. Different communication channels

- Tested on two kinds of recordings:

1. Longitudinal test
2. At-home test

2. Dataset

- Three datasets are considered in this study, one is used for train the models and the other two sets are considered to test.
- Two speech tasks : (1) a monologue and (2) the reading text
- Training Set : PC-GITA.[3]

Table 1
Description of the training set. **PD patients**: Parkinson's disease patients. **HC**: healthy controls.

	PD patients		Healthy speakers	
	male	female	male	female
Number of speakers	22	22	25	25
Age [years] (mean ± standard deviation)	61.3 ± 12.3	61.9 ± 7.3	60.5 ± 11.4	61.4 ± 6.9
Range of age [years]	33–81	49–75	31–86	49–76
Disease duration [years] (mean ± standard deviation)	9.2 ± 6.0	13.0 ± 12.0		
Range of disease duration [years]	0.4–20	1–43		
m-FDA (mean ± standard deviation)	31.2 ± 8.1	32.0 ± 10.1	7.6 ± 7.3	5.1 ± 9.1
Range of m-FDA	17–41	13–51	0–29	0–25
MDS-UPDRS-III (mean ± standard deviation)	40.7 ± 21.5	37.5 ± 15.2		
Range of the MDS-UPDRS-III scores	9–92	19–71		
Average duration of the monologues (in seconds)	47.2 ± 26.4	41.5 ± 20.6	43.1 ± 30.9	54.4 ± 27.3
Average duration of the read texts (in seconds)	18.6 ± 5.9	18.6 ± 6.9	17.5 ± 3.2	18.3 ± 4.2

2.1 Test set

Table 3
Dysarthria scores of the at-home test set. Hi, i ∈{1, 2, .., 16}: m–FDA scores of the sixteen recording sessions.

			m-FDA (At-home)															
Patients (Pi)	Age	Gender	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16
P1	70	M	25	25	25	23	21	27	27	27	27	27	27	27	22	21	22	22
P2	57	M	37	38	35	35	35	38	35	37	37	27	39	37	36	36	37	39
P3	67	M	23	23	23	6	23	14	12	12	12	17	22	23	28	22	16	16
P4	59	F	33	34	34	34	33	33	33	33	33	34	36	36	41	41	41	41
P5	56	F	27	25	25	25	31	29	29	29	29	29	31	31	39	39	37	39
P6	52	F	13	13	13	13	13	13	13	13	15	15	15	15	16	14	14	14
P7	61	M	23	24	24	23	26	26	25	25	26	26	26	26	26	25	24	24

Table 2
General information of patients in the longitudinal test set. LSi: ith longitudinal session (LSi,i∈{1, 2, ...,5}).

			MDS-UPDRS-III					m-FDA (longitudinal)				
Patients (Pi)	Age	Gender	LS1	LS2	LS3	LS4	LS5	LS1	LS2	LS3	LS4	LS5
P1	70	M	14	25	–	7	15	37	22	18	23	31
P2	57	M	–	58	–	63	51	–	34	25	34	35
P3	67	M	28	19	–	13	24	31	15	17	16	23
P4	59	F	41	35	–	33	33	29	39	24	21	40
P5	56	F	29	26	–	26	30	23	26	16	16	14
P6	52	F	38	49	–	44	45	14	20	1	12	15
P7	61	M	6	8	–	24	21	21	36	12	13	17

2.2 Non-controlled acoustic conditions

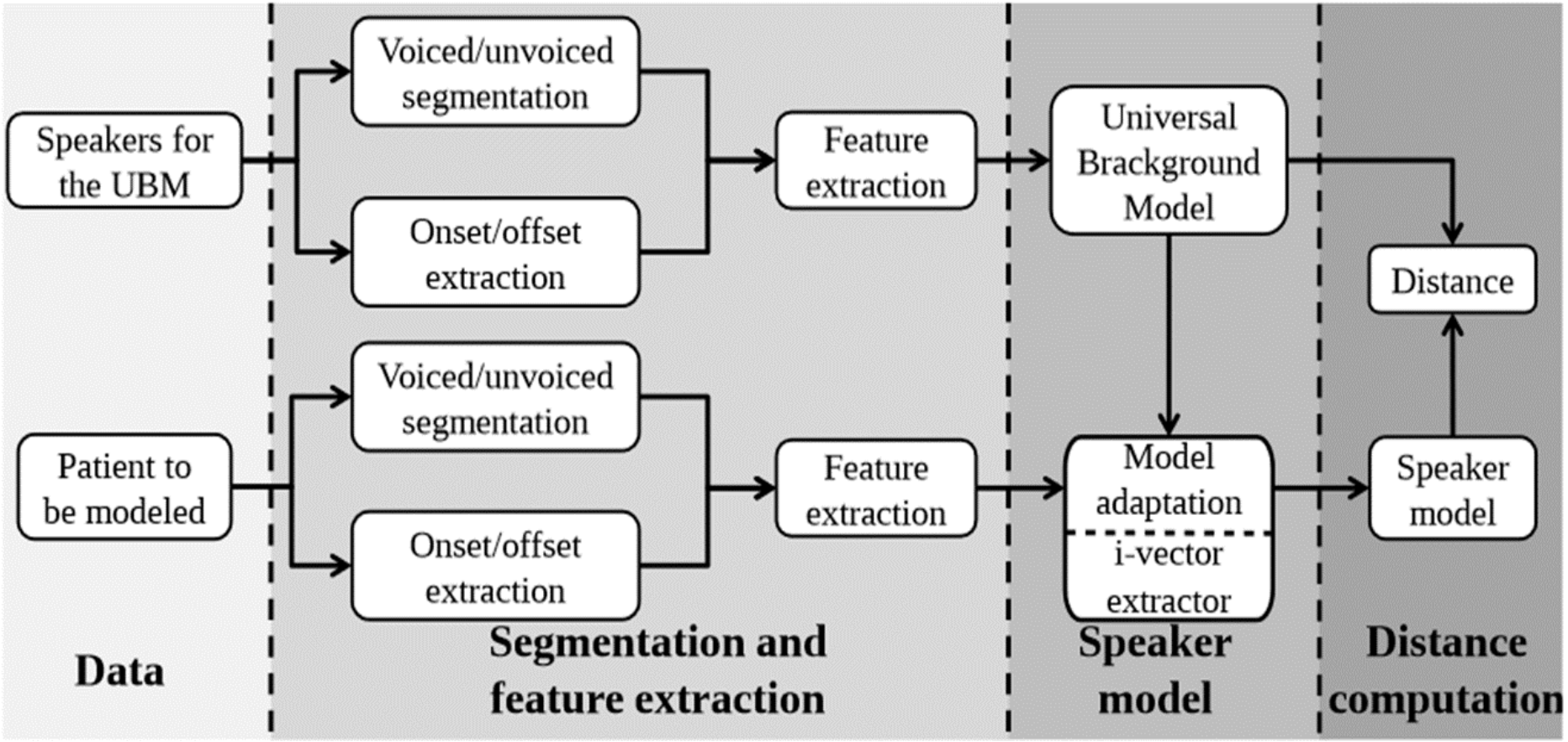
- It is necessary to test its suitability in more realistic conditions.
- Approaches more robust to different acoustic conditions.

Table 4

Transmission rates (kbps) for the five channels considered in this study.

Channel	Mobile	Landline	Skype*	Hangouts*	Original
Transmission rate (kbps)	6.60–23.85	56	6–40	6–510	256

2.3 Method


Speakers for the UBM
Voiced/unvoiced segmentation
Onset/offset extraction
Feature extraction
Universal Brackground Model
Distance
Patient to be modeled
Voiced/unvoiced segmentation
Onset/offset extraction
Feature extraction
Model adaptation
i-vector extractor
Speaker model
Data
Segmentation and feature extraction
Speaker model
Distance computation

2.3.1 Segmentation

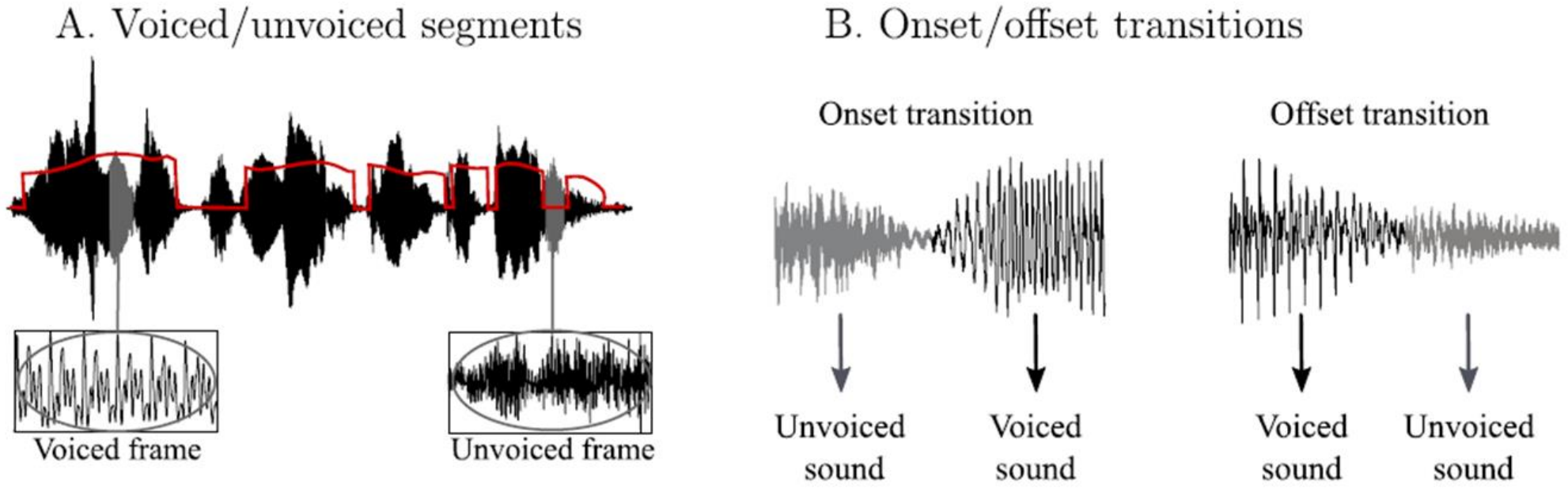


Fig. 3. (A) Pitch contour (red line) and voiced/unvoiced short time windows extracted from a speech signal. (B) Onset and offset transition frames. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3.2 Features

Phonation features

- Include temporal and amplitude variations of the pitch period.
- The first and second derivatives of the pitch contour are considered temporal variability of fundamental frequency

Prosodic features

- Pitch and energy-based features extracted from the voiced segments
- The set of features is computed based on the methodology presented in Dehak et al. (2007).

Articulation features

- Considering information from onset/offset transitions
- The set of features includes 12 Mel-Frequency Cepstral Coefficients (MFCCs)
- Log energy- 22 Bark Bands

2.4 Regression model

- The estimation ($y\hat{}$)a is measured with an ε-insensitive loss function $L(y, y\hat{})$, which ensures the existence of the global minimum, and it is computed with Eq. (1).

$$L(y,\hat{y}))=\begin{cases}0 & \text{if } |y-\hat{y}|\le\varepsilon\\ |y-\hat{y}|-\varepsilon & \text{otherwise}\end{cases} \quad (1)$$

- The feature vectors x are mapped into a m-dimensional feature space using a linear kernel g(x). The estimated values $y\hat{}$, with weights ω, and bias b, are estimated using Eq. (2).

$$\hat{y}=\sum_{j=1}^{m}\omega_j g_j(\boldsymbol{x})+b \quad (2)$$

2.5 Speaker model – GMM-UBM

- GMMs are parametric probabilistic models represented as a weighted sum of M Gaussian densities
- For a D-dimensional feature vector x a GMM is defined as:

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

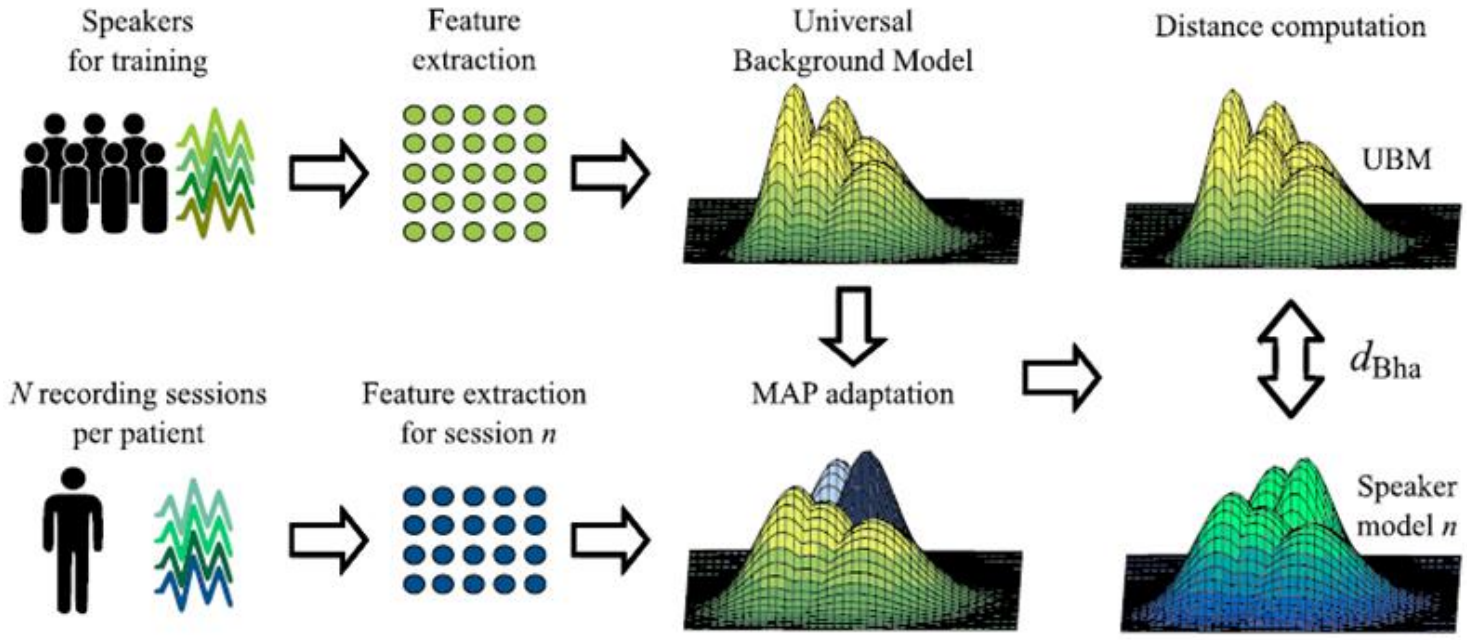


Fig. 4. Speaker modeling. PD progression in N recording sessions per patient: $n \in \{1, 2, 3, \ldots, N\}$.

2.5.1 Speaker model – i-vectors

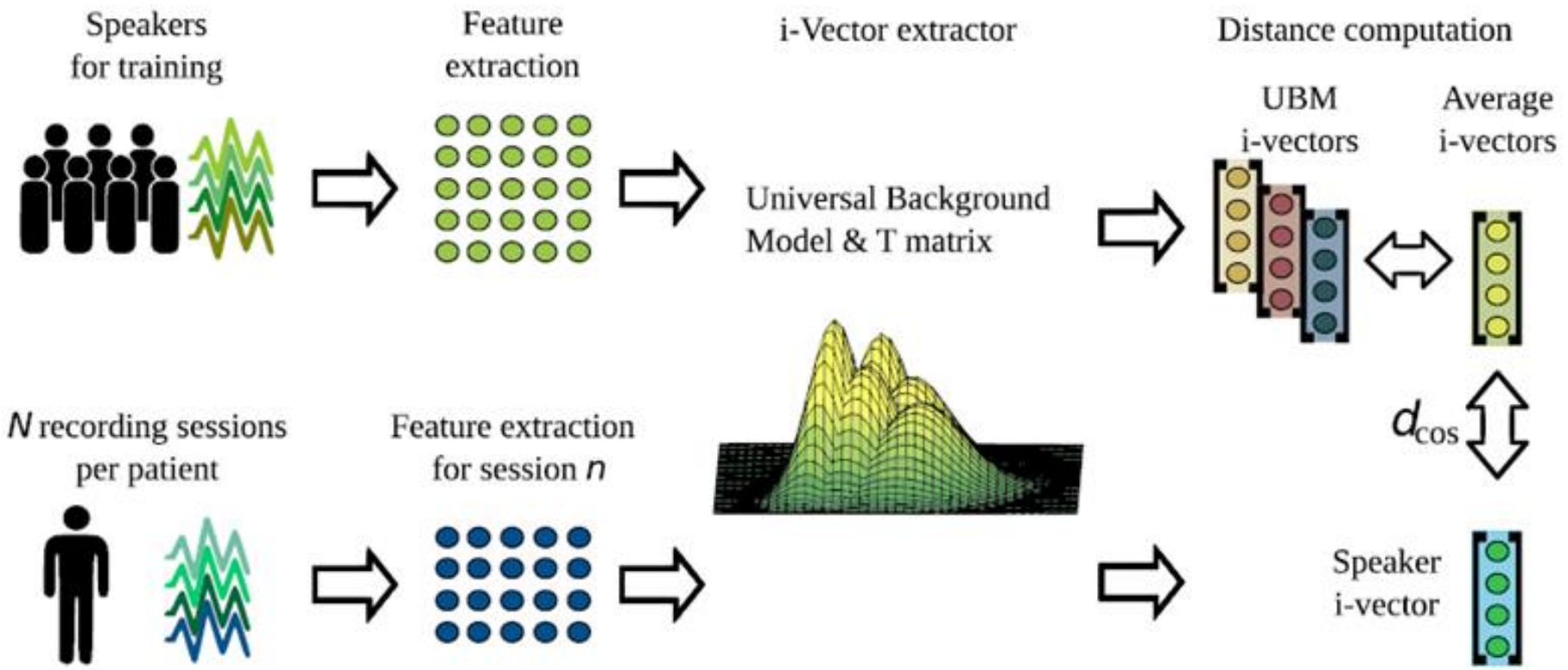


Fig. 5. Speaker modeling. PD progression in N recording sessions per patient: $n \in \{1, 2, 3, ..., N\}$.

2.6 Distances transformed to similarity measures

Similarity measures:

The computed distances per speaker model are transformed into similarity measures using Eq. (12)

$$s_i = 1 - d_i \tag{12}$$

- where di, i ∈ {1, 2, 3, .., 7} are the distances computed per speaker model

Multi-aspect coefficient ξ :

$$\xi_i = \frac{1}{1 + \alpha \mathrm{phon}_i + \beta \mathrm{pro}_i + \theta \mathrm{art}_i} \tag{13}$$

- where phoni, proi, and arti are the distances corresponding to the phonation, prosody, and articulation aspects, respectively for the patient i. α, β, and θ are the weights of each aspect

3. Disease Progression

- Severity get worse over the time.
- Impairment progresses with the disease.
- Identify changes in the speech of the patient over the time.
- Compute the distance between the UBM and the speaker model

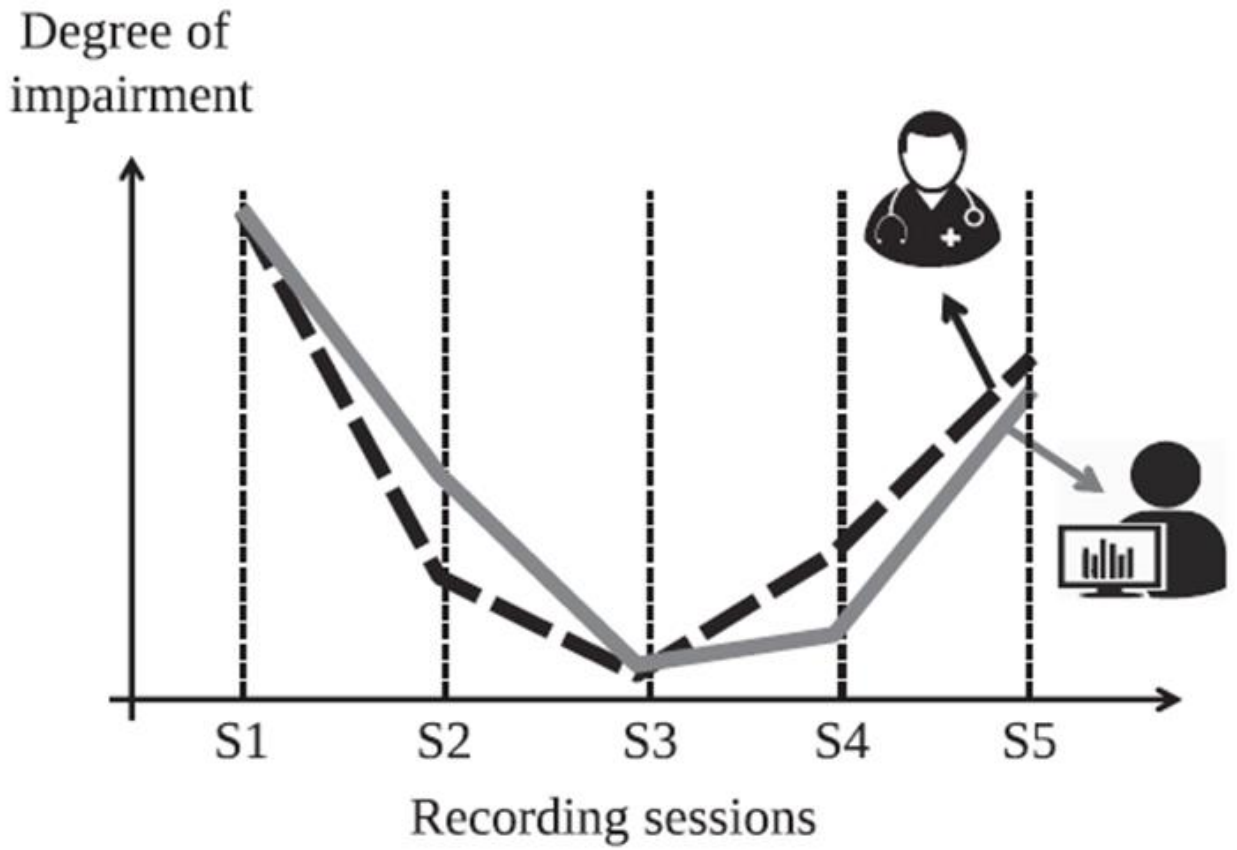


Fig. 6. Graphical representation of the progression of PD for patient 1. The dotted black line represents the progression of the disease according to the clinical score and the continuous gray line represents the progression obtained with the speaker models.

3.1 Results: At-home test

Table 8
Spearman's correlation coefficient (ρ) between the multi-aspect coefficient ξ and m–FDA per patient in the at-home test set (Pi). AVG: Average correlation per communication channel. MSE: Average Mean Squared Error.

Model	Channel	P1	P2	P3	P4	P5	P6	P7	AVG	MSE
	Original	0.46	−0.49	0.18	−0.35	−0.01	0.26	0.12	0.02	1.85
	Skype*	0.39	0.21	−0.20	−0.29	0.61	−0.07	0.20	0.12	1.72
SVR	Mobile	0.82	−0.01	−0.09	−0.37	0.37	0.10	0.37	0.17	1.99
	Landline	−0.08	−0.03	0.16	−0.15	0.07	0.23	−0.12	0.01	1.47
	Hangouts*	0.30	−0.15	−0.29	−0.18	0.05	−0.00	−0.06	−0.05	2.14
	Original	0.62	0.44	0.22	0.31	0.86	0.44	0.39	0.47	1.07
	Skype*	0.76	0.54	0.19	0.46	0.86	0.48	0.54	0.55	0.89
GMM–UBM	Mobile	0.61	0.25	0.24	0.67	0.77	0.29	0.26	0.44	1.26
	Landline	0.73	0.57	0.06	0.40	0.87	0.56	0.47	0.51	1.00
	Hangouts*	0.70	0.49	0.23	0.50	0.45	0.66	0.30	0.48	1.22
	Original	0.63	0.53	0.12	0.46	0.14	0.48	0.30	0.38	1.14
	Skype*	0.26	0.00	0.33	0.67	0.58	0.34	0.61	0.40	1.26
i–vectors	Mobile	0.54	0.24	0.36	0.41	0.77	0.31	0.27	0.41	1.13
	Landline	0.68	0.07	0.22	0.63	0.46	0.49	0.23	0.38	1.40
	Hangouts*	0.59	0.32	0.28	0.45	0.66	0.39	0.34	0.43	1.04

3.2 Results: Dysarthria assessment (Longitudinal test)

Table 12
Pearson's correlation coefficient (ρ) between the multi-aspect coefficient ξ and m–FDA per patient in the longitudinal test set (Pi). AVG: Average correlation per communication channel. MSE: Average Mean Squared Error.

Model	Channel	P1	P2	P3	P4	P5	P6	P7	AVG	MSE
SVR	Original	−0.74	−0.57	−0.95	0.46	−0.50	−0.29	0.13	−0.35	2.70
	Skype*	0.89	−0.94	−0.63	−0.21	−0.54	−0.09	−0.19	−0.24	2.49
	Mobile	−0.08	0.52	0.26	−0.64	0.30	0.36	−0.42	0.04	1.91
	Landline	−0.57	−0.02	−0.79	0.21	−0.18	−0.56	0.21	−0.24	2.49
	Hangouts*	−0.50	0.23	−0.48	−0.91	−0.07	0.43	−0.38	−0.24	2.48
GMM–UBM	Original	0.85	0.76	0.74	0.26	0.95	0.85	0.36	0.68	0.64
	Skype*	0.80	0.55	0.55	0.58	0.29	0.65	0.70	0.59	0.82
	Mobile	0.55	0.79	0.16	0.75	0.79	0.75	0.76	0.65	0.79
	Landline	0.75	0.90	0.40	0.53	0.85	0.91	0.63	0.71	0.58
	Hangouts*	0.82	0.60	0.89	0.51	0.86	0.63	0.15	0.64	0.73
i–vectors	Original	0.81	0.94	0.88	0.65	0.96	0.39	0.75	0.77	0.47
	Skype*	0.73	0.80	0.96	0.53	0.87	0.82	0.50	0.74	0.52
	Mobile	0.68	0.43	0.44	0.97	0.88	0.81	0.55	0.68	0.64
	Landline	0.51	0.24	0.34	0.85	0.79	0.60	0.81	0.59	0.81
	Hangouts*	0.49	0.47	0.53	0.89	0.84	0.13	0.67	0.54	0.93

3.2.1 Results: Dysarthria assessment (Longitudinal test)

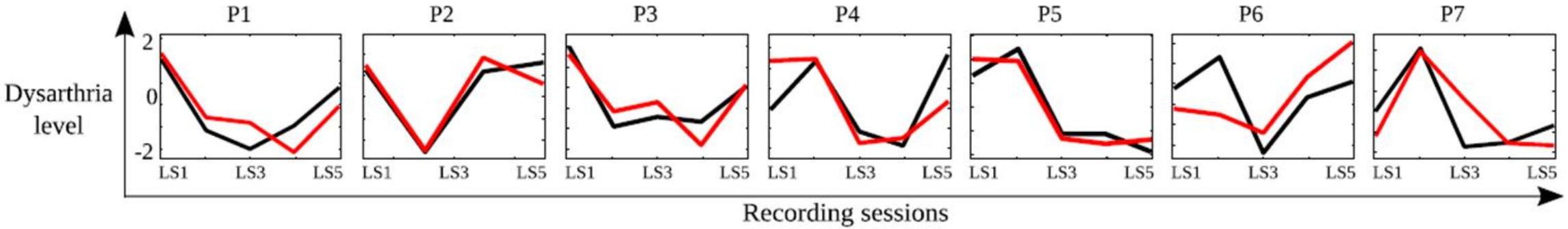


Fig. 8. Curves of the dysarthria level per patient (Pi) in the longitudinal test set. Comparison of the m-FDA scores estimated using i–vectors with the original recordings (red lines) and the original m–FDA values assigned by the phoniatricians (black lines). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3 Results: Neurological assessment (Longitudinal test)

- Results are not good as dysarthria level.
- Why?

Table 13
Pearson's correlation coefficients (r) estimated between ξ calculated using i–vectors and MDS-UPDRS-III per patient in the longitudinal test set (Pi). AVG: Average correlation per communication channel. MSE: Average Mean Squared Error.

Channel	P1	P2	P3	P4	P5	P6	P7	AVG	MSE
Original	0.31	−0.85	0.93	0.40	−0.35	0.65	0.08	0.17	1.38
Skype*	0.70	0.99	0.93	0.54	0.28	−0.03	0.41	0.55	0.89
Mobile	0.57	−0.77	0.94	−0.03	−0.57	0.63	−0.98	−0.03	1.98
Landline	0.82	0.20	0.69	−0.37	0.25	−0.33	−0.99	0.04	1.68
Hangouts*	0.88	0.28	0.49	0.42	−0.15	0.05	−0.77	0.17	1.36

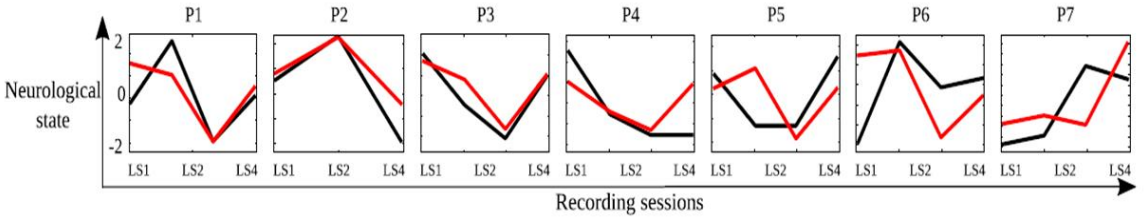


Fig. 9. Curves of the neurological level per patient (Pi). Comparison of the MDS-UPDRS-III scores estimated using i–vectors with the recordings of the Skype* calls (red lines) and original MDS-UPDRS-III values assigned by the neurologist expert (black lines). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

1. Hornykiewicz, O., 1998. Biochemical aspects of Parkinson's disease. Neurology 51 (2), S2–S9.

2. Tsanas, A., Little, M., McSharry, P.E., Ramig, L., 2010. Accurate telemonitoring of Parkinson's disease progression by noninvasive speech tests. IEEE Trans. Biomed.Eng. 57 (4), 884–893.

3. Orozco-Arroyave, J.R., Arias-Londoño, J.D., Vargas-Bonilla, J.F., Gonzalez-Rátiva, M.C., Nöth, E., 2014. New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. Proceedings of the 9th International Conference on Language Resources and Evaluation. pp. 342–347.

Questions

01

Explain briefly about multi-aspect coefficient and state its performance over individual distance-based similarity measures.

02

Which feature performed well and why?

03

Suggest the speaker models for acoustic and non-acoustic conditions.