A Multitask Learning Approach to Assess the Dysarthria Severity in Parkinson's Patients

Juan Camilo Vásquez-Correa[1,2], Tomás Arias-Vergara[1,2,3] Juan Rafael Orozco-Arroyave[1,2], and Elmar Nöth[1,2]
[1] Faculty of Engineering, University of Antioquia, Medellín, Colombia
[2] Pattern Recognition Lab, Friedrich-Alexander University of Erlangen-Nürnberg
[3]Ludwig-Maximilians-University, Munich, Germany

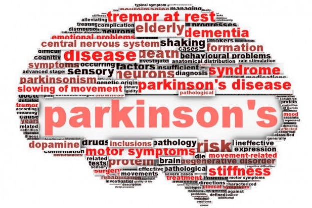
September 1, 2018

Introduction: Parkinson's Disease (PD)

- Second most prevalent neurological disorder worldwide.
- Patients develop several motor and non-motor impairments.
- Speech impairments are one of the earliest manifestations.

Introduction: Parkinson's Disease (PD)

- Second most prevalent neurological disorder worldwide.
- Patients develop several motor and non-motor impairments.
- Speech impairments are one of the earliest manifestations
- The neurological condition of the patients can be assessed using the MDS-UPDRS scale.
- Only one of the 33 items of the scale is related to speech.

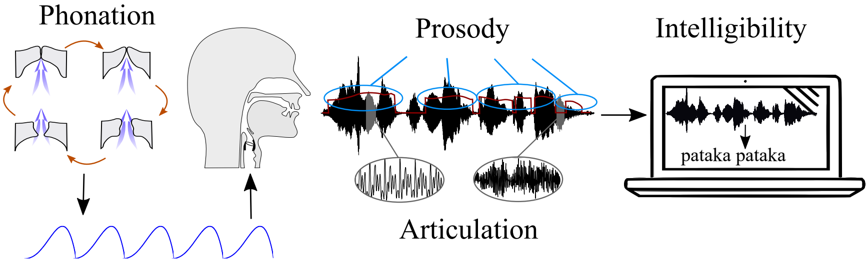


Introduction: Speech impairments

- Reduced loudness
- Monotonic speech
- Monoloudness
- Reduced stress
- Breathy voice
- Hoarse voice quality
- Imprecise articulation

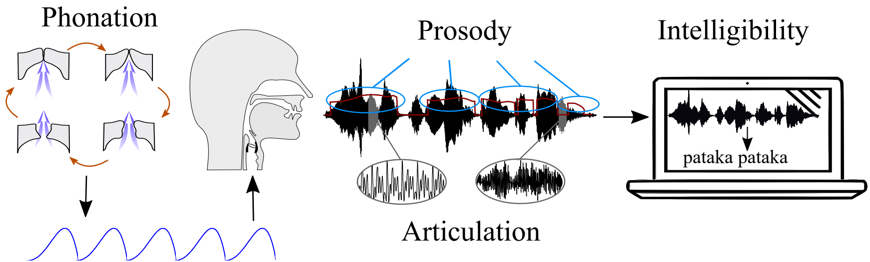
Introduction: Speech impairments

Speech impairments in PD patients: **hypokinetic dysarthria**



Introduction: Speech impairments

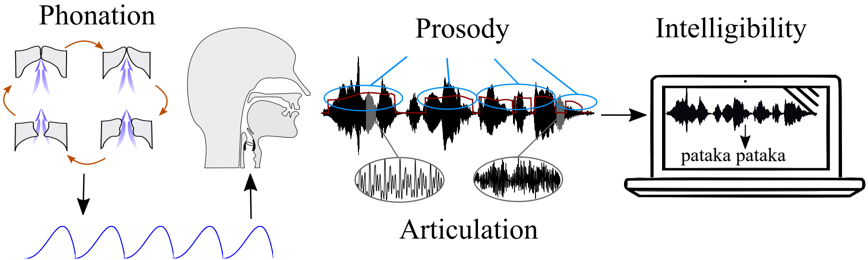
Speech impairments in PD patients: **hypokinetic dysarthria**



Phonation: bowing and inadequate closure of vocal folds.
Phonation is mainly characterized by perturbation features and noise measures.

Introduction: Speech impairments

Speech impairments in PD patients: **hypokinetic dysarthria**

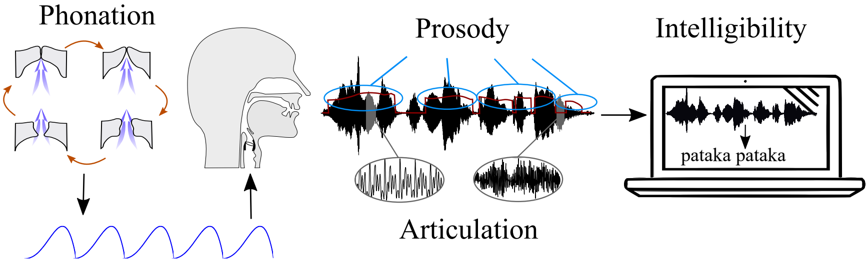


Articulation: reduced amplitude and velocity in the movement of articulators.

Articulation is mainly characterized by features related to formant frequencies, voiced onset time, energy content in transitions, among others.

Introduction: Speech impairments

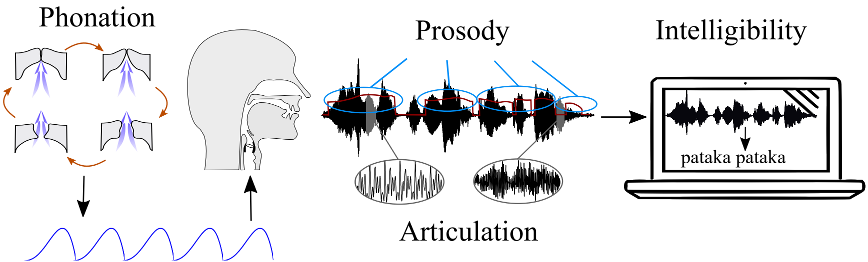
Speech impairments in PD patients: **hypokinetic dysarthria**



Prosody: manifested as monotonocity, monoloudness, and changes in speech rate and pauses.
Prosody is mainly characterized by features related to fundamental frequency, energy, and duration.

Introduction: Speech impairments

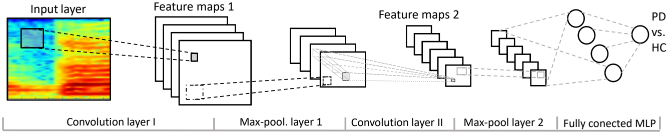
Speech impairments in PD patients: **hypokinetic dysarthria**



Intelligibility: capacity to be understood by other person or by a system.
Intelligibility is mainly characterized by word error rate in a speech recognition system.

Introduction: Motivation

- There is already known success of classical feature extraction and machine learning approaches.
- However, deep learning methods have been successfully implemented recently in pathological speech assessment tasks, including PD.
  - Interspeech 2015 computational paralinguistic challenge (ComParE)[a].
  - Articulation model based on convolutional neural networks (CNNs)[b]



[a] B. Schuller, S. Steidl, et al. (2015). "The INTERSPEECH 2015 computational paralinguistics challenge: Nativeness, Parkinson's & eating condition". In: *Proceedings of INTERSPEECH*, pp. 478–482.

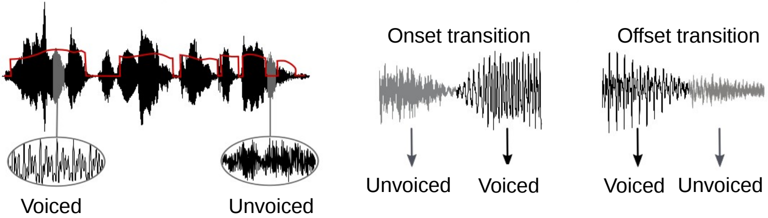
[b] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, and E. Nöth (2017). "Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson's Disease". In: *Proceedings of INTERSPEECH*, pp. 314–318.

Introduction: Motivation

- There is already known success of classical feature extraction and machine learning approaches.
- However, deep learning methods have been successfully implemented recently in pathological speech processing, including PD.
- Most of the studies consider only one specific task to evaluate the speech of PD patients e.g., to classify PD patients vs. healthy subjects.
- A multitask learning scheme offers the possibility to evaluate several deficits simultaneously.
  - Breathing capacity.
  - Intelligibility.
  - Larynx movement capacity.
  - Tongue movement capacity.

Introduction: Hypothesis

- *PD patients have difficulties to begin and to stop the vocal fold vibration, and such difficulties can be observed on speech signals by modeling the transitions between voiced and unvoiced sounds*

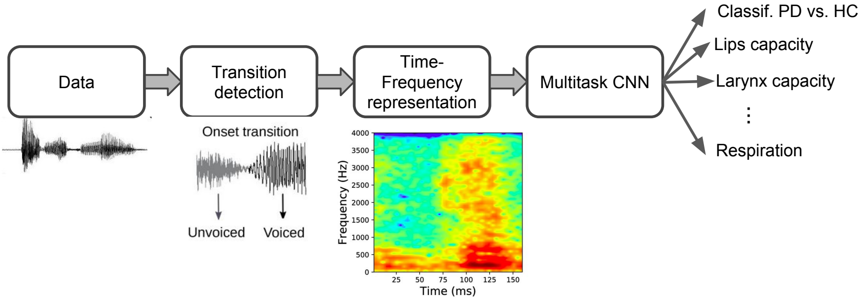


- *A multitask learning strategy combined with the transitions assessment gives us a suitable tool to assess several speech impairment of the patients, improving also the generalization in the learning process.*

Introduction: Aims

- A multitask learning scheme based on CNNs to assess the severity of different speech aspects that are impaired in PD patients.
- A total of eleven tasks are considered.
  - Classification of PD patients and HC subjects.
  - Evaluation of the neurological state of the patients.
  - Evaluation of the dysarthria severity of the patients.
  - Respiration capability
  - Larynx movement capacity
  - Lips movement capacity
  - among others...

Materials and Methods



Materials and Methods: Data

- 50 patients. Most in early to mid-stages of the disease.
- 50 healthy subjects.
- Balanced in age and gender.
- Spanish native speakers (Colombian).
- Diadochokinetic exercises (rapid repetition of syllables).

Materials and Methods: Data

- 50 patients. Most of them in early to mid-stages of the disease.
- 50 healthy subjects.
- Balanced in age and gender.
- Spanish native speakers (Colombian).
- Diadochokinetic exercises (rapid repetition of syllables).
- Patients were labeled according to the MDS-UPDRS score.
- All participants were labeled according to the modified Frenchay dysartrhia assessment (m-FDA) scale[a]

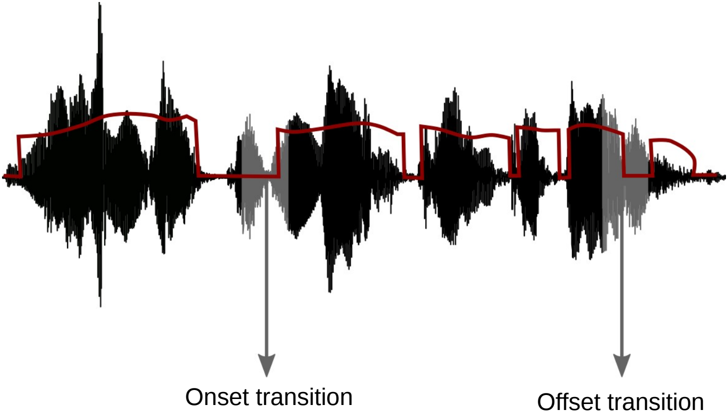
[a] J. C. Vásquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth (2018). "Towards an Automatic Evaluation of the Dysarthria Level of Patients with Parkinson's Disease". In: *Journal of Communication Disorders* 76, pp. 21–36.

Materials and Methods: Data

Table: m-FDA scale

Aspect	m-FDA items
Respiration	1) Duration of respiration 2) Respiratory capacity.
Lips	3) Strength of closing the lips. 4) General capacity to control the lips.
Palate/Velum	5) Nasal escape. 6) Velar movement.
Laryngeal	7) Phonatory capacity in vowels. 8) Phonatory capacity in continuous speech. 9) Effort to produce speech.
Tongue	10) Velocity to move the tongue in /pa-ta-ka/. 11) Velocity to move the tongue in /ta/.
Intelligibility	12) General intelligibility.
Monotonicity	13) Monotonicity and intonation.

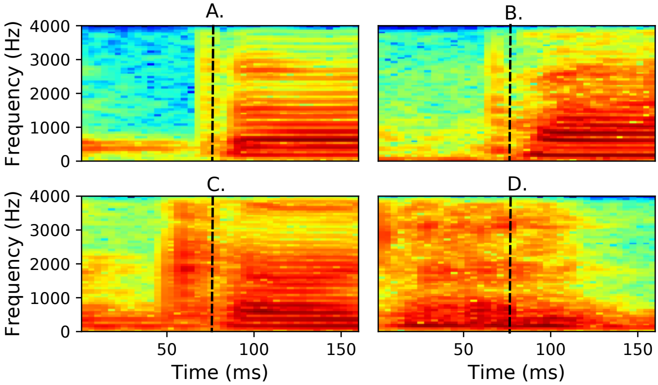
Methods: Transitions detection



Onset and offset are detected according to the presence of the fundamental frequency[1]

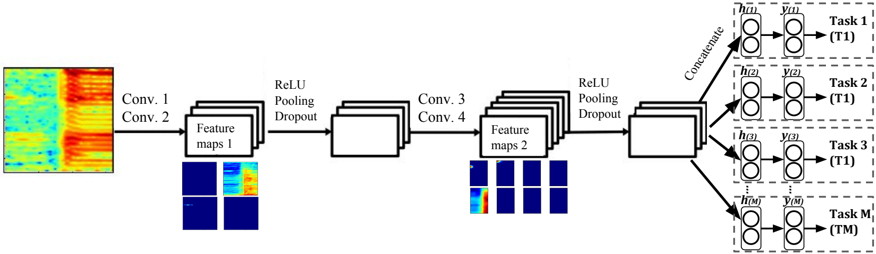
[1] J. R Orozco-Arroyave, J. C. Vásquez-Correa, et al. (2018). "NeuroSpeech: An open-source software for Parkinson's speech analysis". In: *Digital Signal Processing* 77, pp. 207–221.

Methods: Time-frequency representations



STFT of an onset produced by: **A)** HC subject; **B)** PD patient in low state of the disease. **C)** PD patient in intermediate state and **D)** PD patient in severe state. All figures correspond to the syllable /ka/.

Methods: Multitask CNN



The loss function in a multitask strategy is a linear combination of the individual loss functions for each task.

For two tasks

$$L(\Theta) = \gamma L_1(\Theta) + (1-\gamma)L_2(\Theta) \qquad (1)$$

Any number of tasks

$$L(\Theta) = \sum_i \gamma_i L_i(\Theta) \qquad (2)$$

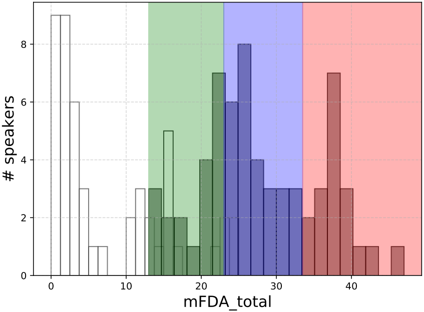
Experiments

Table: Description of the Tasks considered for the multitask learning approach

Task	Description	N. classes
Task 1.	PD. vs. HC	2
Task 2.	Total MDS-UPDRS-III	4
Task 3.	speech item MDS-UPDRS-III	4
Task 4.	Total m-FDA	4
Task 5.	m-FDA Respiration aspect	4
Task 6.	m-FDA Lips movement aspect	4
Task 7.	m-FDA Palate movement aspect	4
Task 8.	m-FDA Larynx movement aspect	4
Task 9.	m-FDA monotonicity aspect	3
Task 10.	m-FDA Tongue aspect	4
Task 11.	m-FDA Intelligibility aspect	3

Experiments

Distribution of the m-FDA scale into four classes.



Multitask CNNs are trained with information of the onset and offset transitions, and the results are compared to those obtained by training single CNNs per task.

Validation

- 5 fold cross-validation: 3 for training, 1 to optimize hyper-parameters, and 1 for test.

Results: Comparison multitask vs. single task learning schemes

Results in terms of the Unweighted average recall (UAR %).

Task	N. classes	Multitask onset	Singletask onset	Multitask offset	Singletask offset
PD vs HC	2	85.0	86.0	89.0	79.0
Total MDS-UPDRS-III	4	55.2	41.0	38.8	41.5
MDS-UPDRS-speech	4	51.7	38.3	47.0	33.6
Total m-FDA	4	43.3	43.8	40.3	42.9
m-FDA respiration	4	44.7	41.2	37.6	42.4
m-FDA lips	4	51.4	49.0	31.1	33.3
m-FDA palate	4	37.6	33.6	31.1	34.5
m-FDA larynx	4	43.2	44.4	42.6	34.9
m-FDA monotonicity	3	49.7	59.6	50.3	32.8
m-FDA tongue	4	43.1	42.6	53.8	40.7
m-FDA intelligibility	3	57.8	67.5	68.2	67.4
Average		**51.1**	**49.8**	**48.2**	**43.9**

Results: Comparison multitask vs. single task learning schemes

Results in terms of the Unweighted average recall (UAR %).

Task	N. classes	Multitask onset	Singletask onset	Multitask offset	Singletask offset
PD vs HC	2	85.0	86.0	89.0	79.0
Total MDS-UPDRS-III	4	55.2	41.0	38.8	41.5
MDS-UPDRS-speech	4	51.7	38.3	47.0	33.6
Total m-FDA	4	43.3	43.8	40.3	42.9
m-FDA respiration	4	44.7	41.2	37.6	42.4
m-FDA lips	4	51.4	49.0	31.1	33.3
m-FDA palate	4	37.6	33.6	31.1	34.5
m-FDA larynx	4	43.2	44.4	42.6	34.9
m-FDA monotonicity	3	49.7	59.6	50.3	32.8
m-FDA tongue	4	43.1	42.6	53.8	40.7
m-FDA intelligibility	3	57.8	67.5	68.2	67.4
Average		**51.1**	**49.8**	**48.2**	**43.9**

- Articulation related tasks e.g., m-FDA larynx and tongue are those that provide the largest improvements in the multitask scheme.
- More labeled data is need to improve the results for the multi-class tasks.

Results: effect of the weight hyper-parameter of the loss function

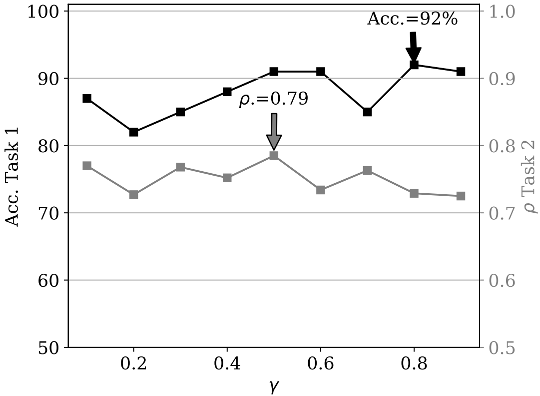


Figure: Results when the loss function of the CNN includes the classification of PD vs. HC subjects (Task 1) and the prediction of the total m-FDA score (Task 2).
γ: weight hyper-parameter for the loss function.
ρ: Spearman's correlation coefficient.

Conclusion

- A multitask learning scheme based on CNNs is proposed to assess the severity of different speech impairments that appear in PD patients.
- The results indicate that it is more accurate to train a CNN in a multitasks learning scheme rather than to train individual CNNs to learn tasks for each speech deficit.
- The most representative tasks in the multitask learning where those related to the articulation dimension of the speech.

Conclusion

- An additional improvement in the results might be obtained if only those tasks related to the articulation capabilities are used in the multi-task learning framework
- Other models might be considered in further experiments to improve the results in the other tasks not related to the articulation impairments such as respiration, monotonicity, and intelligibility.

References I

Schuller, B., S. Steidl, et al. (2015). “The INTERSPEECH 2015 computational paralinguistics challenge: Nativeness, Parkinson’s & eating condition”. In: *Proceedings of INTERSPEECH*, pp. 478–482.

Vásquez-Correa, J. C., J. R. Orozco-Arroyave, and E. Nöth (2017). “Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson’s Disease”. In: *Proceedings of INTERSPEECH*, pp. 314–318.

J. C. Vásquez-Correa, J. R. Orozco-Arroyave, T. Bocklet, and E. Nöth (2018). “Towards an Automatic Evaluation of the Dysarthria Level of Patients with Parkinson’s Disease”. In: *Journal of Communication Disorders* 76, pp. 21–36.

Orozco-Arroyave, J. R, J. C. Vásquez-Correa, et al. (2018). “NeuroSpeech: An open-source software for Parkinson’s speech analysis”. In: *Digital Signal Processing* 77, pp. 207–221.

Thanks for attending.

Any questions?

juan.vasquez@fau.de

www5.cs.fau.de/en/our-team/vasquez-camilo

TAPAS

Training Network on Automatic Processing of PAthological Speech (TAPAS)
Horizon 2020 Marie Sklodowska-Curie Actions Initial Training Network European Training Network (MSCA-ITN-ETN) project.

A Multitask Learning Approach to Assess the Dysarthria Severity in Parkinson's Patients

Juan Camilo Vásquez-Correa[1,2], Tomás Arias-Vergara[1,2,3] Juan Rafael Orozco-Arroyave[1,2], and Elmar Nöth[1,2]

[1] Faculty of Engineering, University of Antioquia, Medellín, Colombia

[2] Pattern Recognition Lab, Friedrich-Alexander University of Erlangen-Nürnberg

[3]Ludwig-Maximilians-University, Munich, Germany

September 1, 2018

Appendix: full results

Table: Results obtained for a multitask learning approach to classify eleven tasks related to speech impairments of PD patients. **ACC:** Accuracy (%), **UAR**: Unweighted average recall (%). The bold UARs correpond to the best result obtained per task, for onset and offset

Task	N. classes	Multitask onset		Single task onset		Multitask offset		Single task offset	
		ACC.	UAR	ACC.	UAR	ACC.	UAR	ACC.	UAR
PD vs HC	2	85.0±10.8	85.0	86.0±2.7	**86.0**	89.0±7.7	**89.0**	79.0±6.7	79.0
Total MDS-UPDRS-III	4	55.4±9.4	**55.2**	51.2±8.1	41.0	55.5±11.4	38.8	52.0±10.5	**41.5**
MDS-UPDRS-speech	4	57.8±11.8	**51.7**	50.4±10.6	38.3	56.8±14.4	**47.0**	54.2±9.1	33.6
Total m-FDA	4	45.2±6.7	43.3	46.8±7.8	**43.8**	44.3±8.4	40.3	43.0±3.8	**42.9**
m-FDA respiration	4	40.7±4.2	**44.7**	42.8±1.1	41.2	40.8±15.2	37.6	44.3±11.9	**42.4**
m-FDA lips	4	54.3±6.3	**51.4**	49.3±4.2	49.0	43.8±3.3	31.1	41.7±7.7	**33.3**
m-FDA palate	4	43.6±2.4	**37.6**	41.4±5.3	33.6	39.8±14.2	31.1	39.7±5.8	**34.5**
m-FDA larynx	4	46.2±8.5	43.2	44.5±5.7	**44.4**	43.4±6.6	**42.6**	35.9±10.6	34.9
m-FDA monotonicity	3	49.6±10.1	49.7	50.1±11.5	**59.6**	50.6±3.2	**50.3**	44.4±9.8	32.8
m-FDA tongue	4	43.9±4.2	**43.1**	48.8±9.9	42.6	54.3±6.9	**53.8**	39.5±4.3	40.7
m-FDA intelligibility	3	68.4±6.5	57.8	70.0±8.6	**67.5**	69.5±6.3	**68.2**	69.5±6.3	67.4
Average		**53.6**	**51.1**	**52.9**	**49.8**	**53.5**	**48.2**	**49.4**	**43.9**

Appendix: full results 2

Table: Comparison between multitask learning and single learning for the classification of PD vs. HC subject and the prediction of the m-FDA score. **Dev.**: results for the development set, **Test**: results for the test set.

Task	Metric	Multitask			Single task		
		Dev.	Test	γ	Dev	Test	γ
PD vs. HC	ACC.	92.0	80.0	0.8	89.0	74.0	1
Total m-FDA	ρ	0.79	0.58	0.5	0.71	0.54	0