Convolutional Neural Networks and a Transfer Learning Strategy to Classify Parkinson's Disease from Speech in Three Different Languages

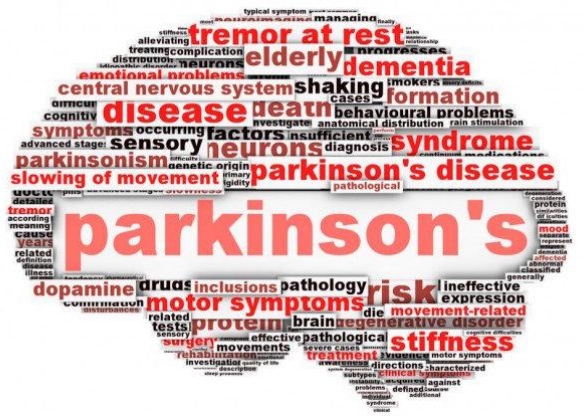
Juan Camilo Vasquez-Correa[1,2], Tomás Arias-Vergara[1,2,3], Cristian D. Rios-Urrego[2], Maria Schuster[3]
Jan Rusz[4], Juan Rafael Orozco-Arroyave[1,2], and Elmar Nöth[1]

[1]Pattern Recognition Lab. Friedrich-Alexander University, Erlangen-Nürnberg, Germany.
[2]Faculty of Engineering. Universidad de Antioquia UdeA, Medellin, Colombia.
[3]Ludwig-Maximilians-University, Munich, Germany.
[4]Czech Technical University in Prague, Prague, Czech Republic.

24th Iberoamerican Congress on Pattern Recognition (CIARP 2019)
Havana (Cuba), 28th-31st October 2019

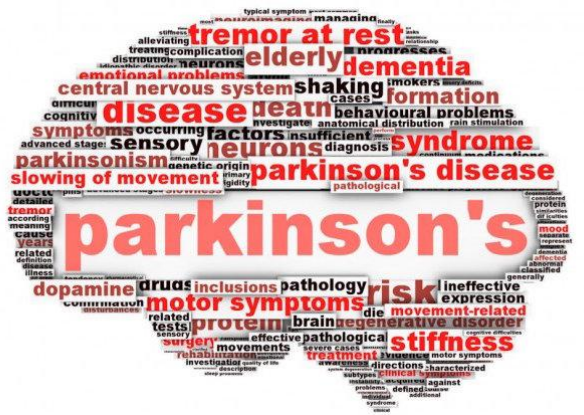
Background: Parkinson's Disease

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



Background: Parkinson's Disease

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



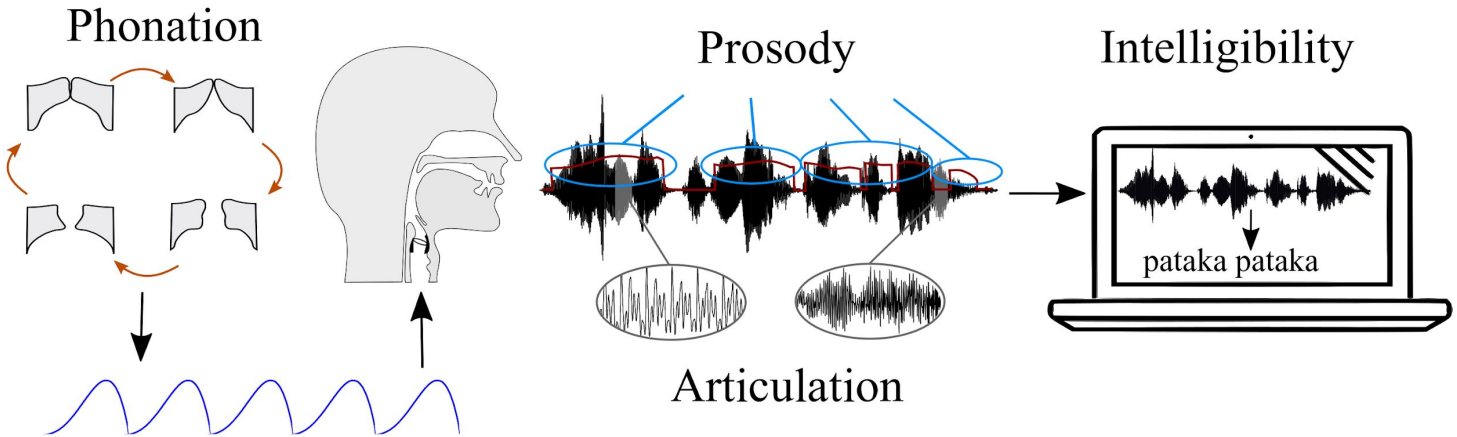
Motor impairments include:

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

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

Background: Parkinson's Disease

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



- Reduced loudness
- Monotonic speech
- Breathy voice

- Hoarse voice quality
- Imprecise articulation

Motivation

- Clinical observations in the speech of patients can be automatically measured to address two main aspects:
  1. To support the diagnosis of the disease.
  2. To predict the level of degradation of the speech of the patients.

Motivation

- Many studies in the literature to classify PD from speech are based on computing hand-crafted features and using classifiers such as support vector machines (SVMs) (Moro 2019, Arora 2019, Benba 2019).
- There is a growing interest in the research community to consider deep learning models in the assessment of the speech of PD patients (Vasquez-Correa 2018, Korzekwa 2019).
- However, in a language independent scenario the results are not satisfactory (accuracy<60%).
- The classification of PD from speech in different languages has to be carefully conducted to avoid bias towards the linguistic content present in each language.

Hypothesis

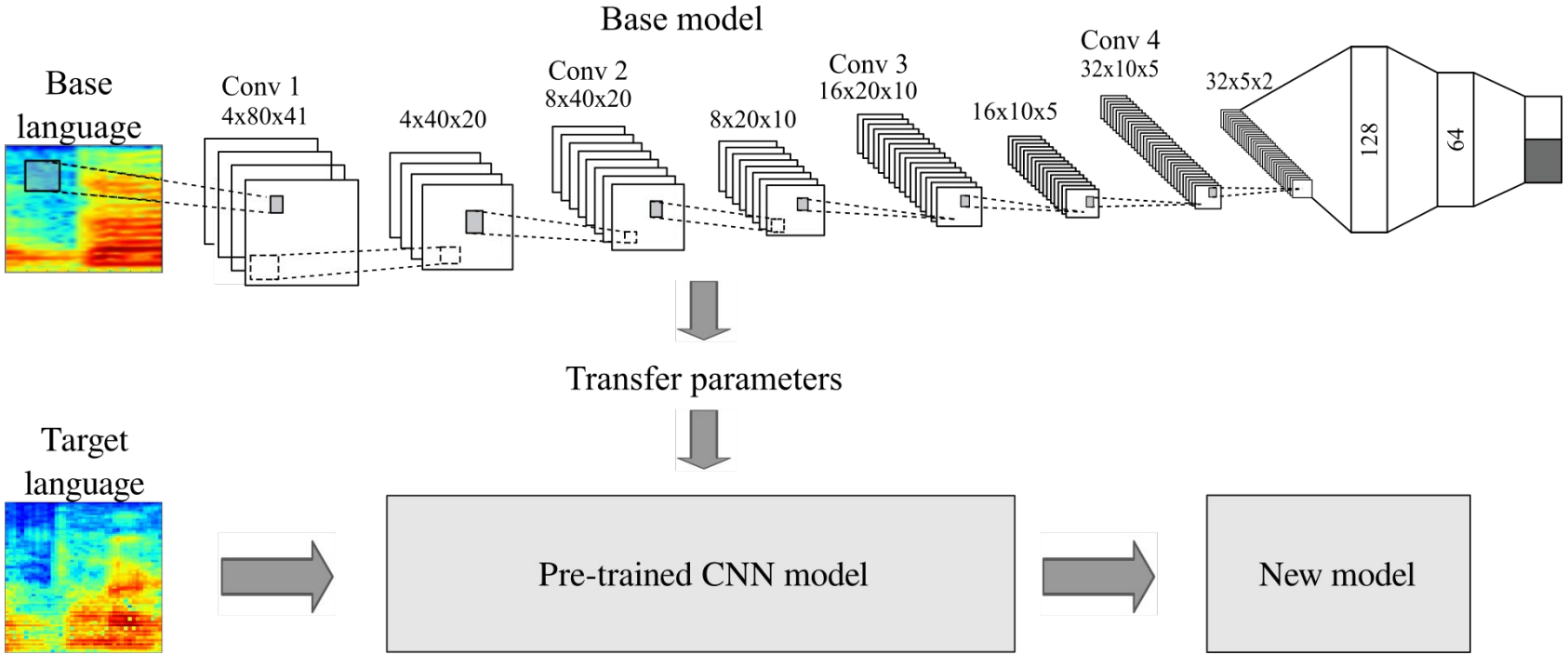
- The results in the classification of PD in different languages could be improved using a transfer learning strategy.
- We propose a methodology to classify PD via a transfer learning strategy to improve the accuracy in different languages.
- CNNs trained with utterances from one language are used to initialize a model to classify speech utterances from PD patients in a different language.

Data

- Spanish, German, and Czech utterances.
- Diadochokinetic exercises, read text, and a monologue.
- Age and gender balanced.

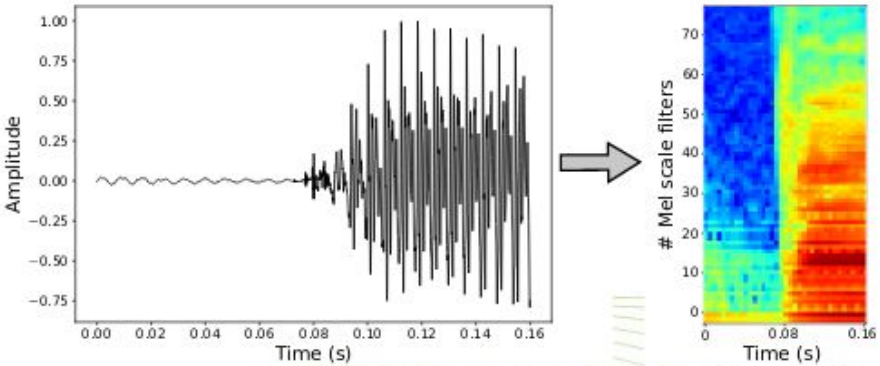
	Spanish		German		Czech	
	PD	**HC**	**PD**	**HC**	**PD**	**HC**
Subjects	50	50	86	86	50	50
Age	61.0	61.0	66.5	63.2	62.7	61.9
Time since diagnosis	10.7	-	7.0	-	6.7	-
MDS-UPDRS-III	37.7	-	22.2	-	19.8	-

Methods

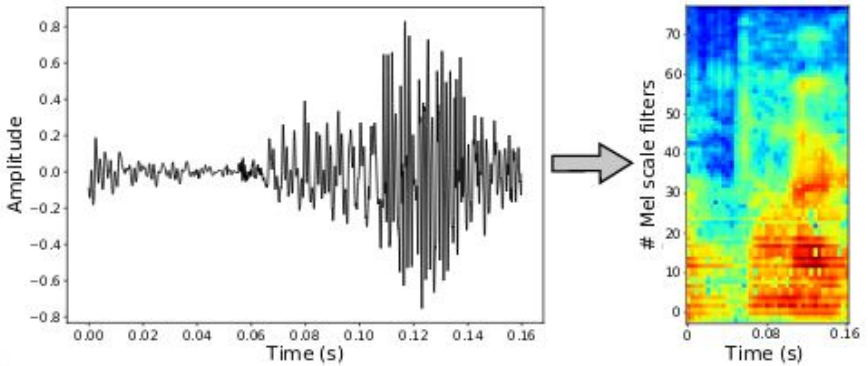

Base model
Base language
Conv 1
4x80x41
4x40x20
Conv 2
8x40x20
8x20x10
Conv 3
16x20x10
16x10x5
Conv 4
32x10x5
32x5x2
128
64
Transfer parameters
Target language
Pre-trained CNN model
New model

Feature extraction

Male HC
Age: 54

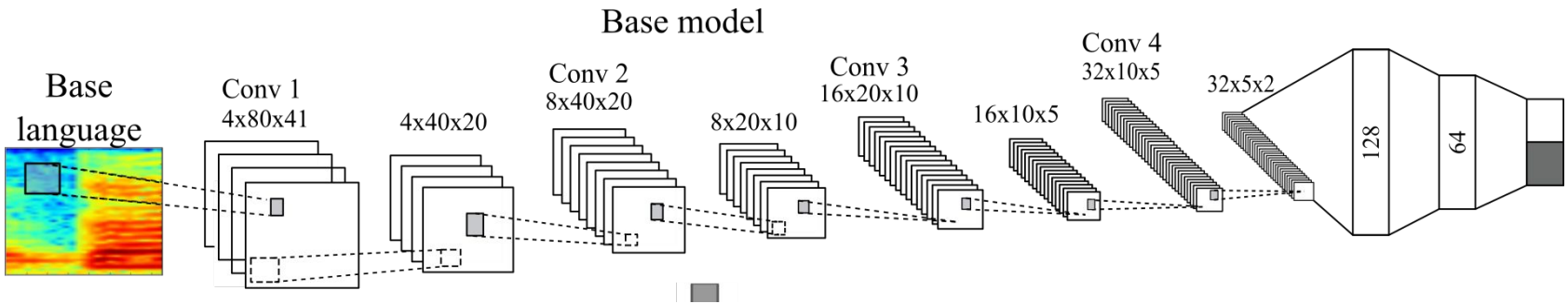


Male PD
Age: 48 MDS-UPDRS-III=9



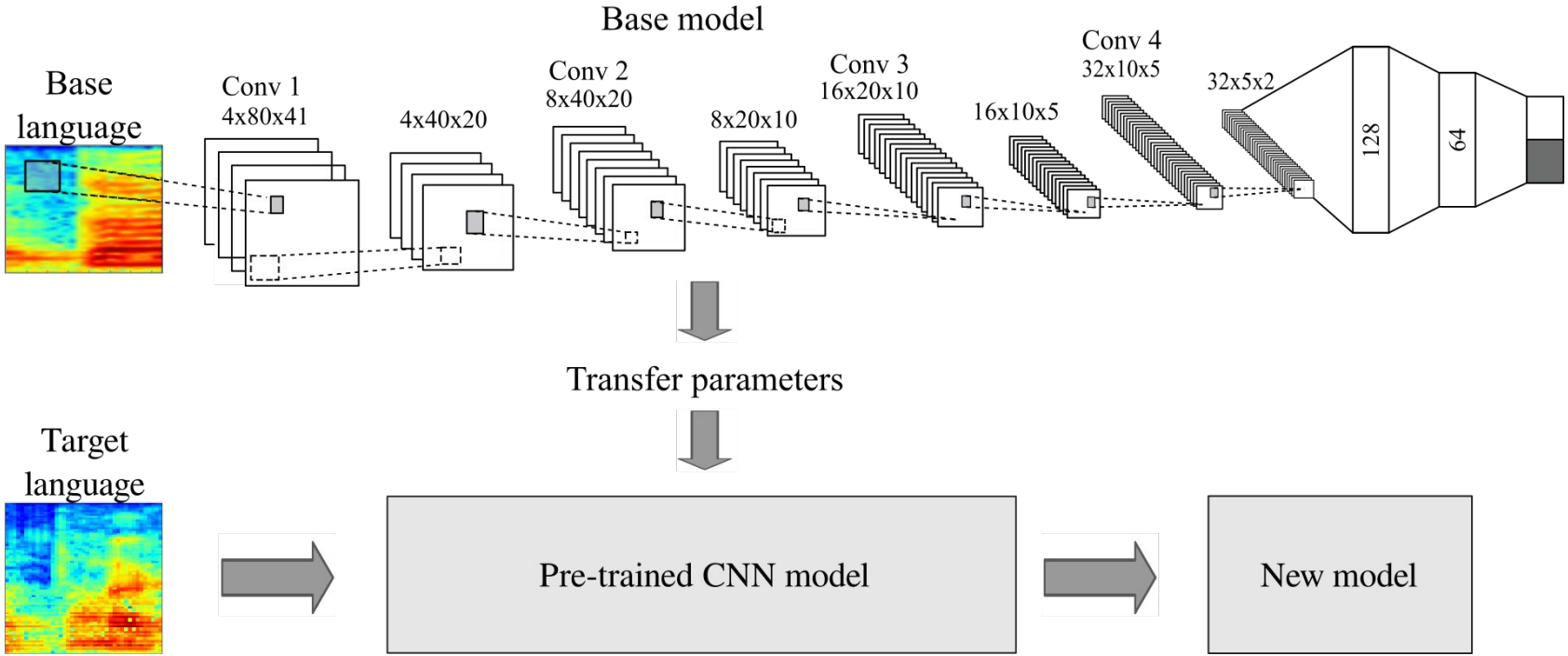
Classification

- A CNN is trained to process the Mel spectrograms extracted from the transitions



- Individual CNNs are trained with the utterance from each language

Transfer learning



Baseline

- Hand-crafted features extracted from the onset and offset transitions
  - 12 Mel frequency cepstral coefficients (MFCCs)
  - Log energy in 22 Bark bands
- Support vector machine (SVM) classifier with a Gaussian kernel.

Experiments and Results

Results obtained with the CNNs trained for each language individually.

Target language	Accuracy (%)	
	Baseline	Base CNN
Spanish	73.7 (13.0)	71.0 (15.9)
German	69.3 (9.9)	63.1 (11.7)
Czech	61.0 (12.5)	68.5 (14.1)

Experiments and Results

Results obtained with the pretrained CNNs in a different language, ie., Transfer learning (TL).

Target language	Accuracy (%)				
	Baseline	Base CNN	Pretrained CNN Spanish	Pretrained CNN German	Pretrained CNN Czech
Spanish	73.7 (13.0)	71.0 (15.9)	-	70.0 (12.5)	72.0 (13.1)
German	69.3 (9.9)	63.1 (11.7)	77.3 (11.3)	-	76.7 (7.9)
Czech	61.0 (12.5)	68.5 (14.1)	72.6 (13.9)	70.7 (14.5)	-

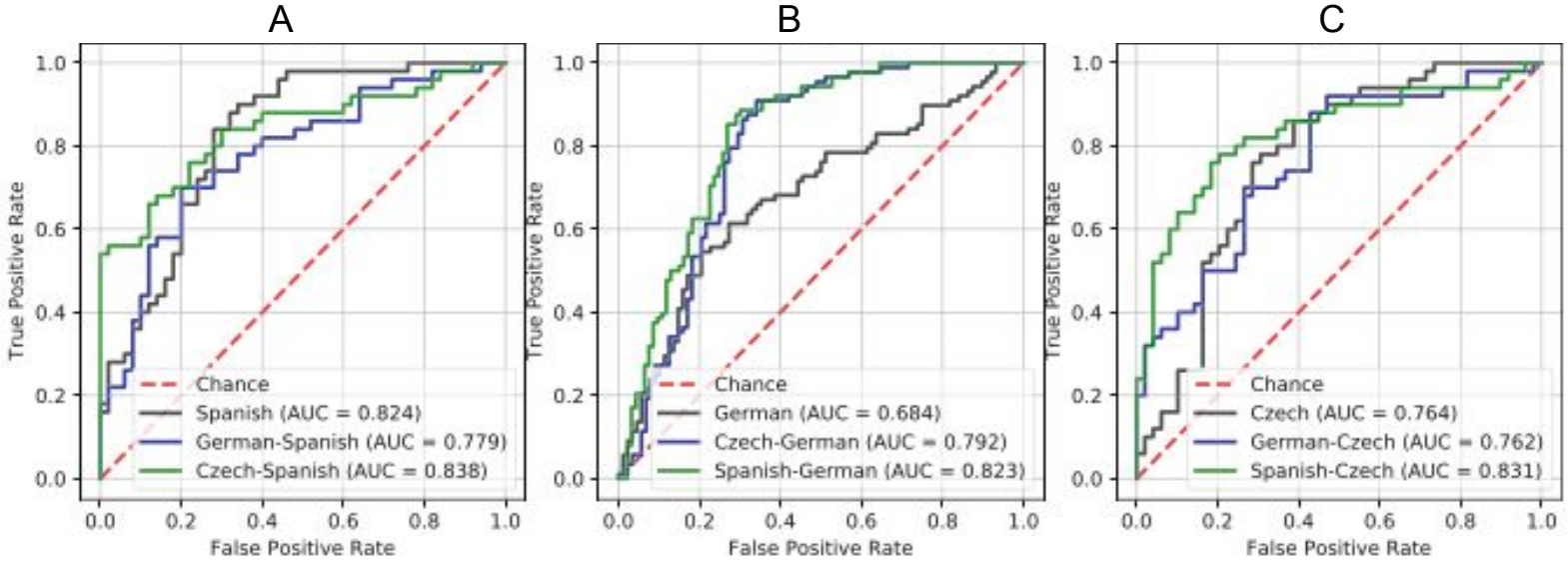
Experiments and Results

Results obtained with the pretrained CNNs in a different language, ie., Transfer learning (TL).

Target language	Accuracy (%)				
	Baseline	Base CNN	Pretrained CNN Spanish	Pretrained CNN German	Pretrained CNN Czech
Spanish	73.7 (13.0)	71.0 (15.9)	-	70.0 (12.5)	72.0 (13.1)
German	69.3 (9.9)	63.1 (11.7)	77.3 (11.3)	-	76.7 (7.9)
Czech	61.0 (12.5)	68.5 (14.1)	72.6 (13.9)	70.7 (14.5)	-

Experiments and Results

Results obtained with the pretrained CNNs in a different language, ie., Transfer learning (TL).



ROC curves for the transfer learning among languages, when the target language is **A**. Spanish, **B**. German, and **C**. Czech.

Conclusion

- Transfer learning strategy to classify PD from speech in three different languages: Spanish, German, and Czech.
- The transfer learning among languages aimed to improve the accuracy when the models are initialized with utterances from a different language than the one used for the test set.
- The transfer learning improved the classification of the German and Czech models, using a Spanish base model.
- Further experiment: to train base models with utterances from two languages instead of only one.
- Future study: transfer learning among diseases, i.e., training a base model to classify PD, and use it to initialize another one to classify other neurological diseases.

References

(Moro 2019) Moro-Velazquez, L., et al. (2019). A forced gaussians based methodology for the differential evaluation of Parkinson's Disease by means of speech processing. *Biomedical Signal Processing and Control*, *48*, 205-220.

(Arora 2019) Arora, S., Baghai-Ravary, L., & Tsanas, A. (2019). Developing a large scale population screening tool for the assessment of Parkinson's disease using telephone-quality voice. *The Journal of the Acoustical Society of America*, *145*(5), 2871-2884.

(Benba 2019) Benba, A., et al. (2019). Voice signal processing for detecting possible early signs of Parkinson's disease in patients with rapid eye movement sleep behavior disorder. *International Journal of Speech Technology*, *22*(1), 121-129.

(Vasquez-Correa 2018) Vásquez-Correa, J.C. et al. (2018). A Multitask Learning Approach to Assess the Dysarthria Severity in Patients with Parkinson's Disease. *Proc. Interspeech* (pp. 456-460).

(Korzekwa 2019) Korzekwa, D. et al. (2019). Interpretable Deep Learning Model for the Detection and Reconstruction of Dysarthric Speech. *Proc. Interspeech 2019*, 3890-3894.

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.