Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson's Disease

Juan Camilo Vásquez-Correa[1,2]
Juan Rafael Orozco-Arroyave[1,2], Elmar Nöth[2]

[1]GITA research group, University of Antioquia UdeA.
[2]Pattern recognition Lab. Friedrich Alexander Universität. Erlangen-Nürnberg.
jcamilo.vasquez@udea.edu.co

18th INTERSPEECH, 2017

Outline

Outline

Introduction: Parkinson's Disease

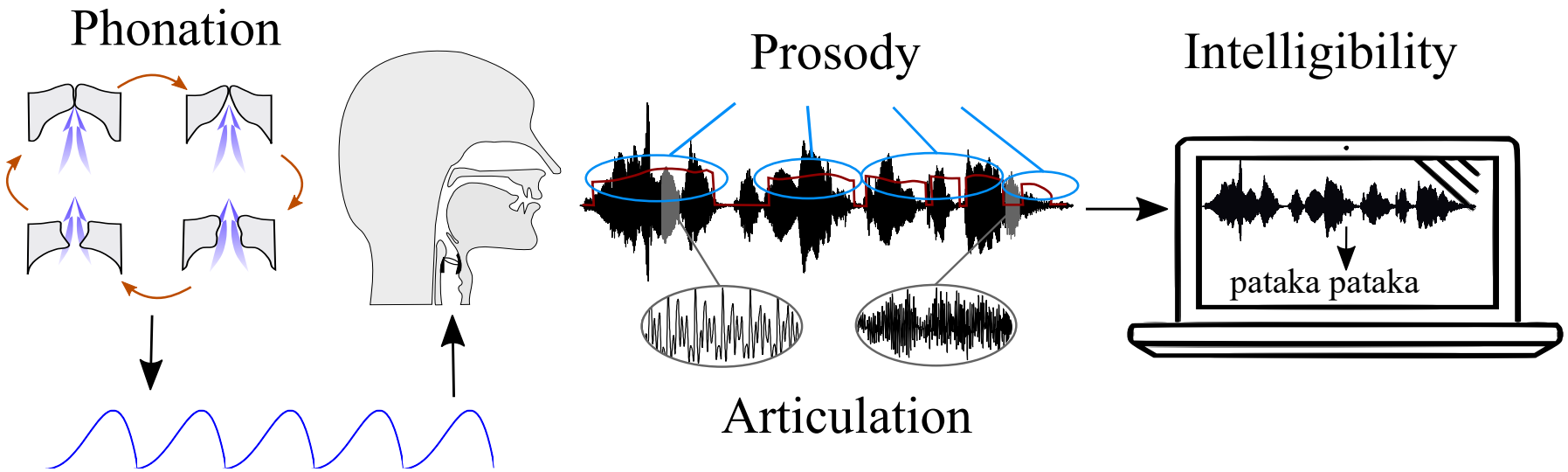
UNIVERSIDAD DE ANTIOQUIA

- Second most prevalent neurological disorder worldwide.
- Patients develop several motor and non-motor impairments. (O. Hornykiewicz 1998).
- Speech impairments are one of the earliest manifestations.

Introduction: Speech impairments

UNIVERSIDAD DE ANTIOQUIA

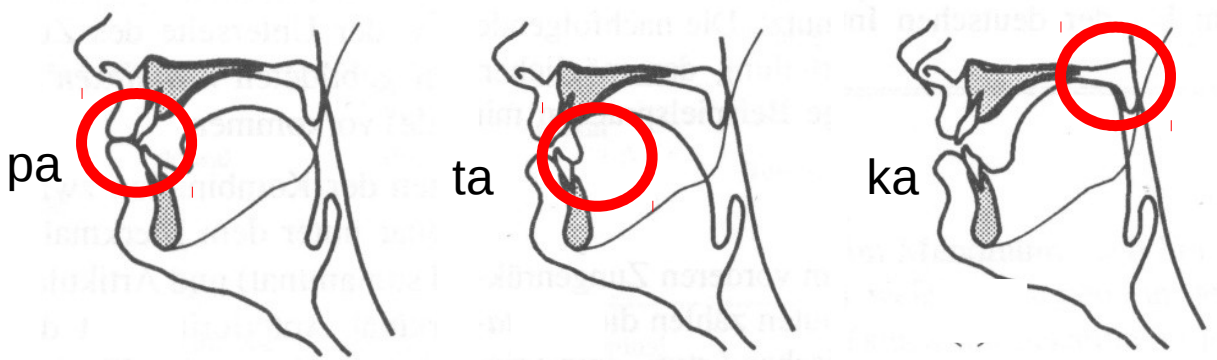
Speech impairments in PD patients: **hypokinetic dysarthria**



Introduction: Imprecise articulation

UNIVERSIDAD DE ANTIOQUIA

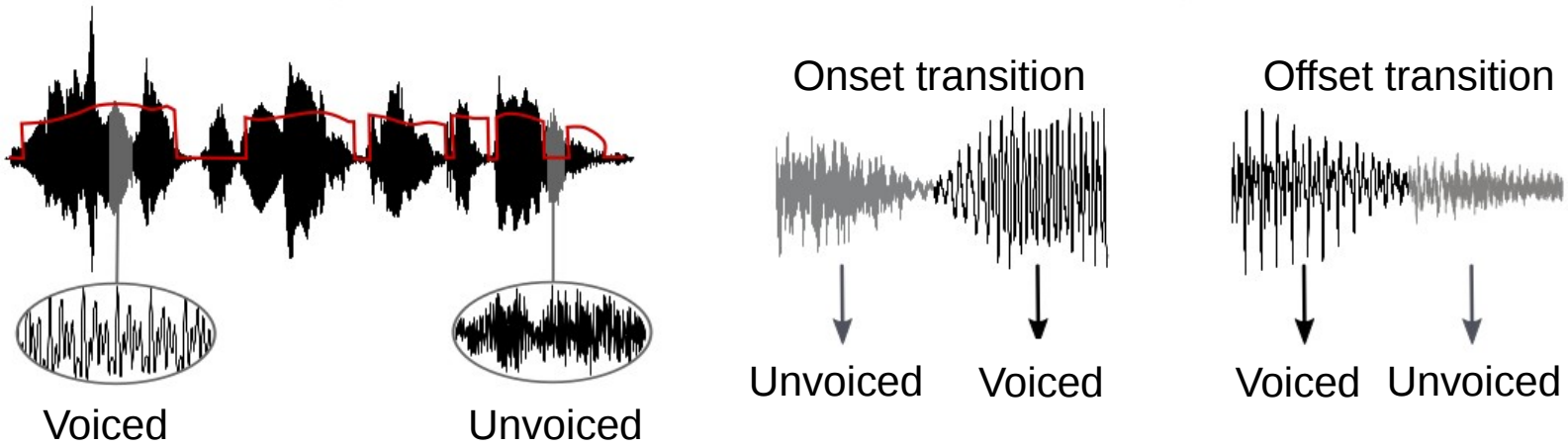
- One of the most deviant speech dimensions in PD.
- Reduced velocity of lip, tongue, and jaw movements.
- Strong indication of the literature statement: **imprecise consonants caused by reduced range of movements of articulators**



Introduction: Hypothesis

UNIVERSIDAD DE ANTIOQUIA

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

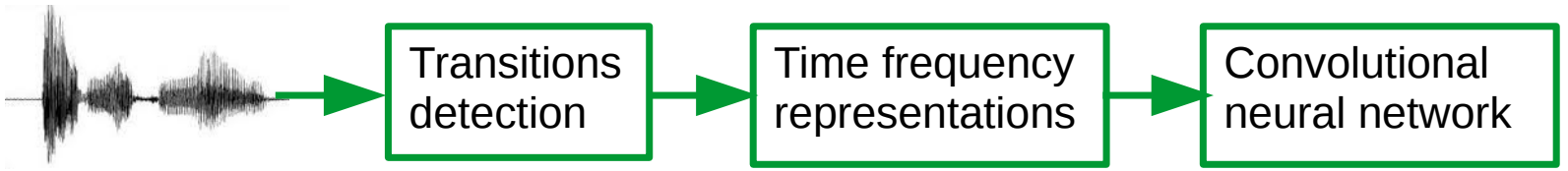


Introduction: Aims

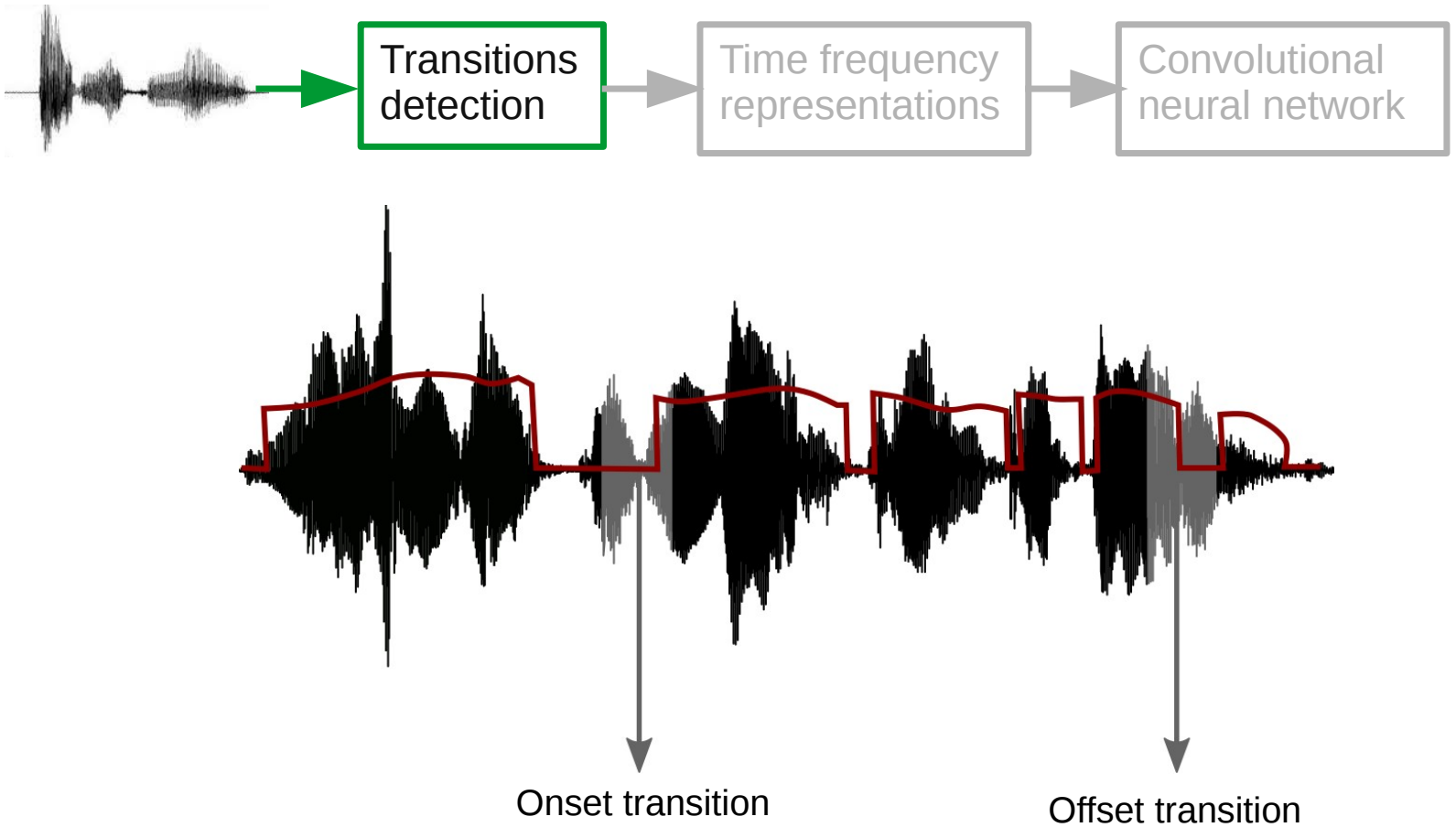
- To model the time-frequency (TF) information provided by the onset and offset transitions: short-time Fourier transform (STFT) and continuous wavelet transform (CWT).
- To “learn” features from time-frequency representations: convolutional neural network (CNN).
- Why TF and feature-learning? both have been successfully used in several paralinguistics tasks: emotion, deception, depression, and others.

Outline

Methods

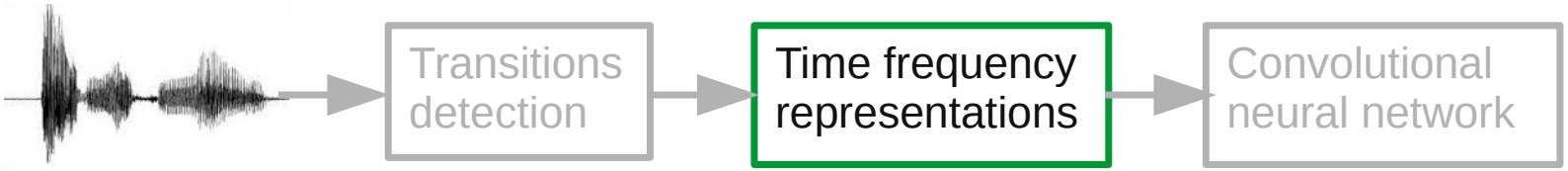

Transitions detection
Time frequency representations
Convolutional neural network

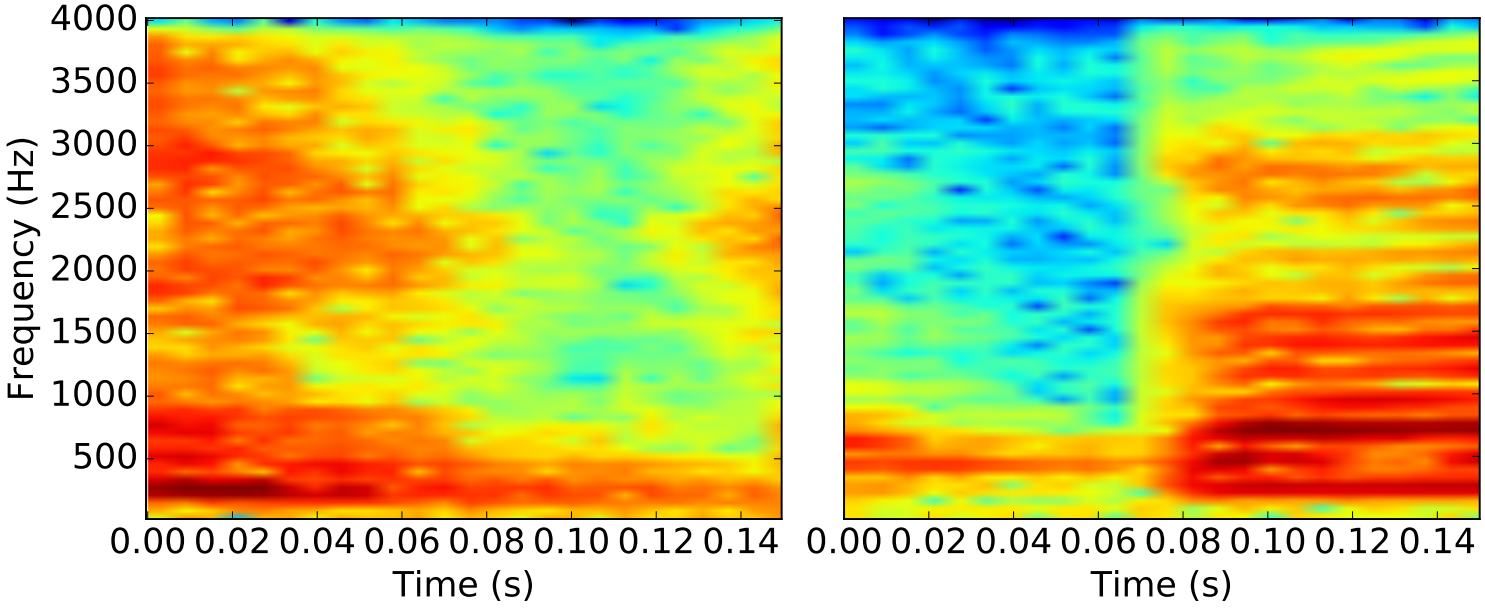
Methods: Transitions detection



Onset and offset are detected according to the presence of the fundamental frequency.

Methods: Time-frequency representation



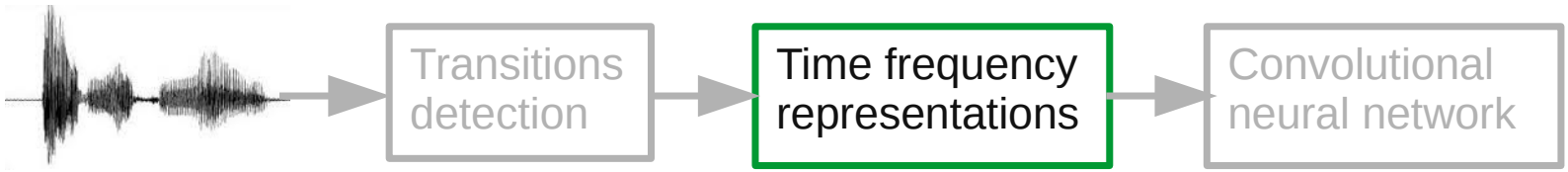


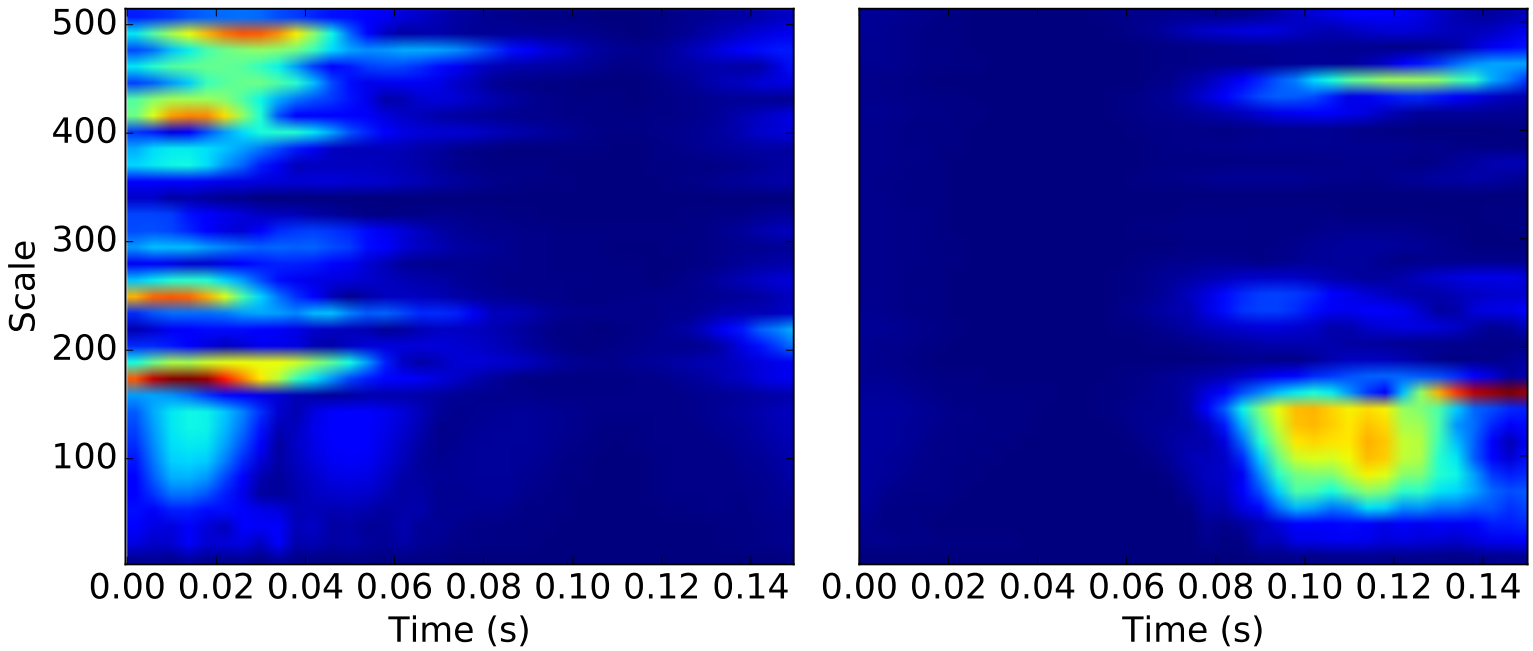
STFT of onset for a PD patient (left) and a HC subject (right)

Play PD

Play HC

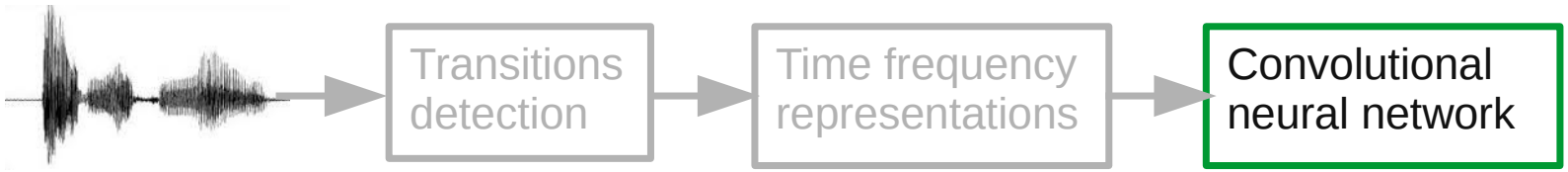
Methods: Time-frequency representation

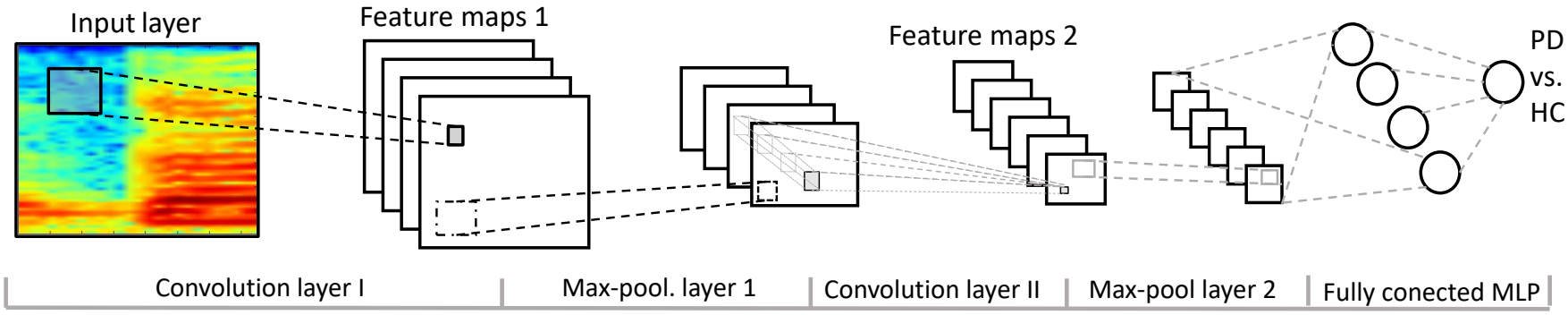




CWT of onset for a PD patient (left) and a HC subject (right)

Methods: Convolutional neural network





CNN learns high–level representations from the low–level raw data

Outline

Data

- Three databases with recordings in three languages: Spanish, German, and Czech.
- Diadochokinetic exercises, isolated sentences, read texts, and monologues.

Data

Language	Description
Spanish	50 Patients and 50 Healthy controls. Balanced in age (60 years old) and gender. Patients in middle state of the disease.
German	88 Patients and 88 Healthy controls. Balanced in age (64 years old). patients in low and middle state of the disease.
Czech	20 Patients and 15 Healthy controls. All male speakers. Patients diagnosed during recording session.

Table: Databases

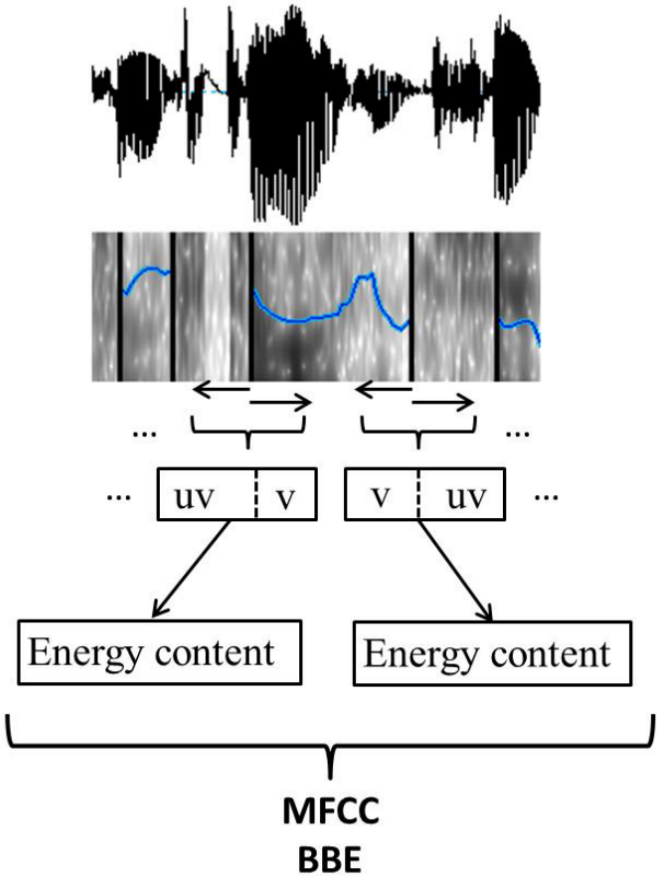
Experiments and validation

- Classification of PD patients vs. HC subjects in the same language.
  - 10 fold cross-validation: 8 for training, 1 to optimize hyper-parameters, and 1 for test.
- Cross-language classification.
  - One language used for train and validation and other language used for test.

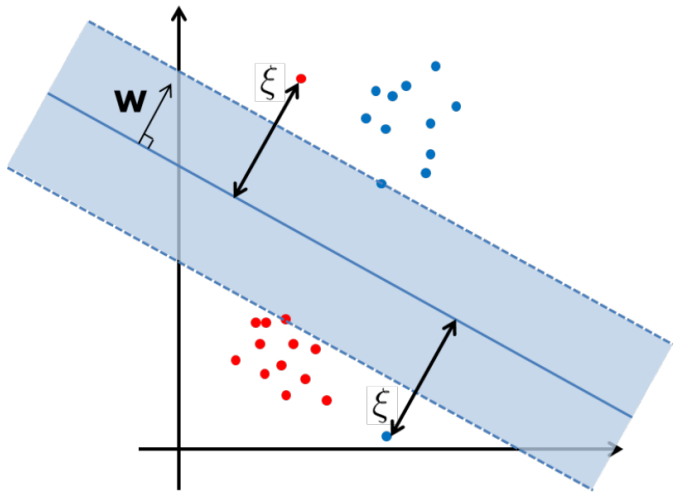
Experiments and validation

UNIVERSIDAD DE ANTIOQUIA

- Results are compared respect to previous studies[1].



Support vector machine



[1] Juan Camilo Vásquez-Correa et al. "Effect of acoustic conditions on algorithms to detect Parkinson's disease from speech". In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP),*. 2017, pp. 5065–5069.

Outline

Results: same language for train and test

UNIVERSIDAD DE ANTIOQUIA

TFR	Onset	Offset	Onset+Offset
Spanish			
CNN-STFT	85.3	81.6	85.9
CNN-CWT	84.2	81.8	85.2
Baseline	69.3	69.6	71.6
German			
CNN-STFT	70.3	68.0	75.0
CNN-CWT	68.0	66.9	70.5
Baseline	72.7	70.9	74.0
Czech			
CNN-STFT	77.9	80.4	84.4
CNN-CWT	89.2	87.7	89.4
Baseline	75.3	74.4	78.8

Results: same language for train and test

UNIVERSIDAD DE ANTIOQUIA

TFR	Onset	Offset	Onset+Offset
Spanish			
CNN-STFT	85.3	81.6	85.9
CNN-CWT	84.2	81.8	85.2
Baseline	69.3	69.6	71.6
German			
CNN-STFT	70.3	68.0	75.0
CNN-CWT	68.0	66.9	70.5
Baseline	72.7	70.9	74.0
Czech			
CNN-STFT	77.9	80.4	84.4
CNN-CWT	89.2	87.7	89.4
Baseline	75.3	74.4	78.8

Results: same language for train and test

UNIVERSIDAD DE ANTIOQUIA

TFR	Onset	Offset	Onset+Offset
Spanish			
CNN-STFT	85.3	81.6	85.9
CNN-CWT	84.2	81.8	85.2
Baseline	69.3	69.6	71.6
German			
CNN-STFT	70.3	68.0	75.0
CNN-CWT	68.0	66.9	70.5
Baseline	72.7	70.9	74.0
Czech			
CNN-STFT	77.9	80.4	84.4
CNN-CWT	89.2	87.7	89.4
Baseline	75.3	74.4	78.8

Results: same language for train and test

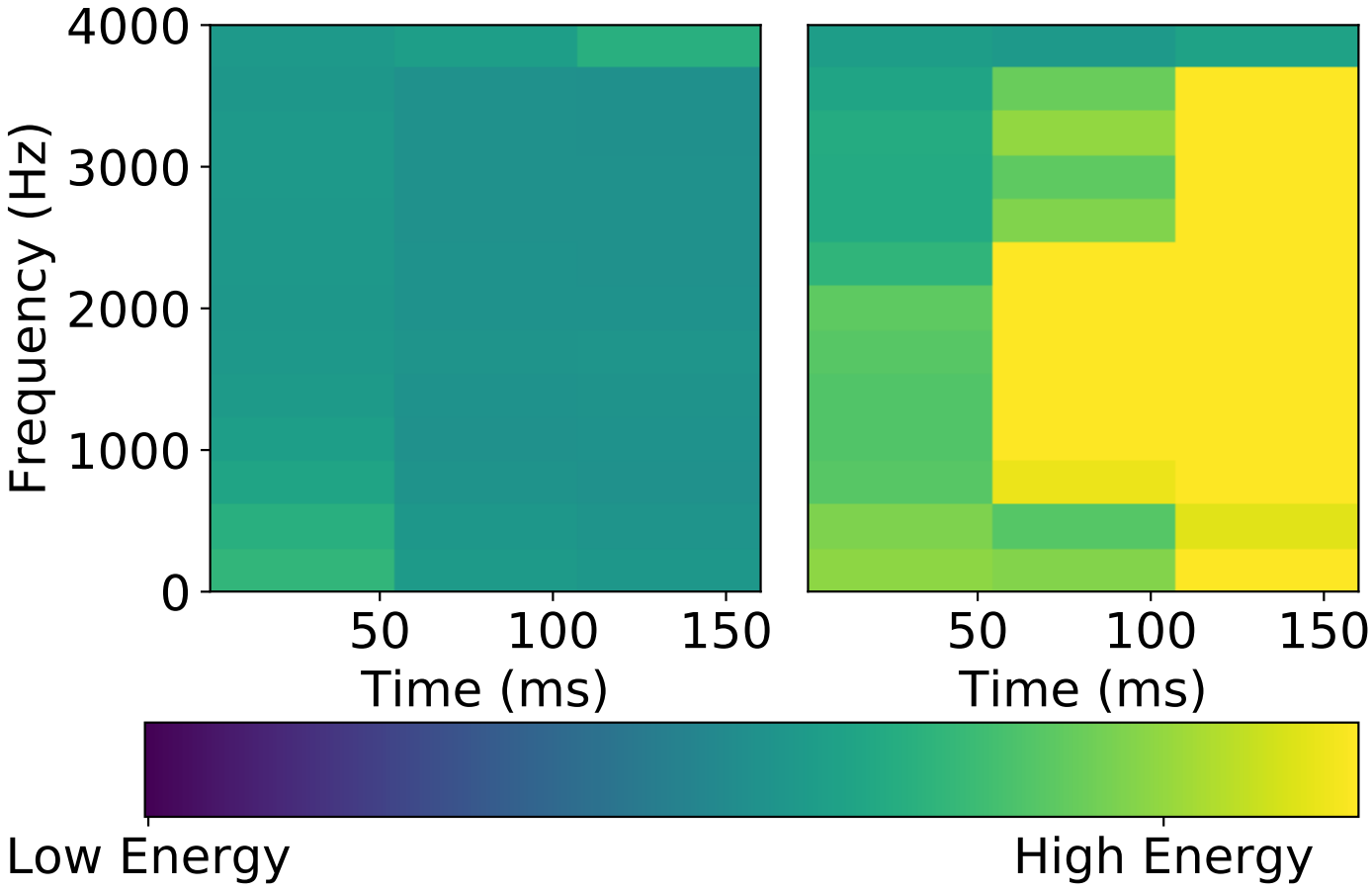


Figure: Output of the CNN after the last max–pool layer: PD patient (left) and a HC speaker (right)

Results: same language for train and test

Speech tasks	Spanish	German	Czech
read text	85.0	70.3	88.5
monologue	85.6	70.3	89.1
/pa-ta-ka/	85.4	70.7	89.2

Results: different language for train and test

Test Lang.	TFR	onset	offset	onset+offset
Train with Spanish				
German	CNN-STFT	51.7	50.2	54.7
German	Baseline	53.7	55.0	54.1
Czech	CNN-CWT	55.2	55.4	57.9
Czech	Baseline	60.3	57.4	60.4
Train with German				
Spanish	CNN-STFT	58.0	55.7	55.8
Spanish	Baseline	53.5	53.5	53.6
Czech	CNN-STFT	53.0	52.4	53.0
Czech	Baseline	50.9	51.7	52.6
Train with Czech				
Spanish	CNN-CWT	53.8	56.3	56.7
Spanish	Baseline	53.4	51.6	52.4
German	CNN-STFT	54.0	51.8	54.0
German	Baseline	51.2	51.0	50.7

Results: different language for train and test

Test Lang.	TFR	onset	offset	onset+offset
Train with Spanish				
German	CNN-STFT	51.7	50.2	54.7
German	Baseline	53.7	55.0	54.1
Czech	CNN-CWT	55.2	55.4	57.9
Czech	Baseline	60.3	57.4	60.4
Train with German				
Spanish	CNN-STFT	58.0	55.7	55.8
Spanish	Baseline	53.5	53.5	53.6
Czech	CNN-STFT	53.0	52.4	53.0
Czech	Baseline	50.9	51.7	52.6
Train with Czech				
Spanish	CNN-CWT	53.8	56.3	56.7
Spanish	Baseline	53.4	51.6	52.4
German	CNN-STFT	54.0	51.8	54.0
German	Baseline	51.2	51.0	50.7

Results: different language for train and test

UNIVERSIDAD DE ANTIOQUIA

Test Lang.	TFR	onset	offset	onset+offset
Train with Spanish				
German	CNN-STFT	51.7	50.2	54.7
German	Baseline	53.7	55.0	54.1
Czech	CNN-CWT	55.2	55.4	57.9
Czech	Baseline	60.3	57.4	60.4
Train with German				
Spanish	CNN-STFT	58.0	55.7	55.8
Spanish	Baseline	53.5	53.5	53.6
Czech	CNN-STFT	53.0	52.4	53.0
Czech	Baseline	50.9	51.7	52.6
Train with Czech				
Spanish	CNN-CWT	53.8	56.3	56.7
Spanish	Baseline	53.4	51.6	52.4
German	CNN-STFT	54.0	51.8	54.0
German	Baseline	51.2	51.0	50.7

Outline

Conclusion

- A deep learning approach is proposed to model articulation impairments of PD patients.
- Voiced-Unvoiced transitions are modeled with CNNs using STFT and CWT.

Conclusion

- The proposed method is able to classify PD patients and HC subjects and improves the baseline when the language used for train and test is the same.
- Additional approaches should be proposed when the train and test language are different.
- Recurrent neural networks and other architectures may be considered to assess co-articulation.
- Deep learning approaches trained with phonation, articulation, and prosody information may be addressed to evaluate specific speech impairments.

Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson's Disease

Juan Camilo Vásquez-Correa[1,2]
Juan Rafael Orozco-Arroyave[1,2], Elmar Nöth[2]

[1]GITA research group, University of Antioquia UdeA.
[2]Pattern recognition Lab. Friedrich Alexander Universität. Erlangen-Nürnberg.
jcamilo.vasquez@udea.edu.co

18th INTERSPEECH, 2017