Convolutional Neural Network to Model Articulation Impairments in Patients with Parkinson's Disease by J. C. Vasquez-Correa et al.

Review by Manila Kodali

1. Motivation

- Speech impairments are one of the earliest manifestations in patients with Parkinson's disease.
- Particularly, articulation deficits related to the capability of the speaker to start/stop the vibration of the vocal folds have been observed in the patients.
- Those difficulties can be assessed by modeling the transitions between voiced and unvoiced segments from speech

Baseline

- Introduced a method to model difficulties observed in PD patients to start/stop the vibration of vocal folds [2].
- The method consists of detecting the transitions from voiced to unvoiced (v-uv), i.e., offset, and from unvoiced to voiced (uv-v), i.e., onset in the speech recording [2].
- Then the energy content in frequency bands separated according to the Bark scale is computed [2].

1.1 Contribution

- A robust strategy to model the articulatory deficits related to the starting or stopping vibration of the vocal folds is proposed in this study

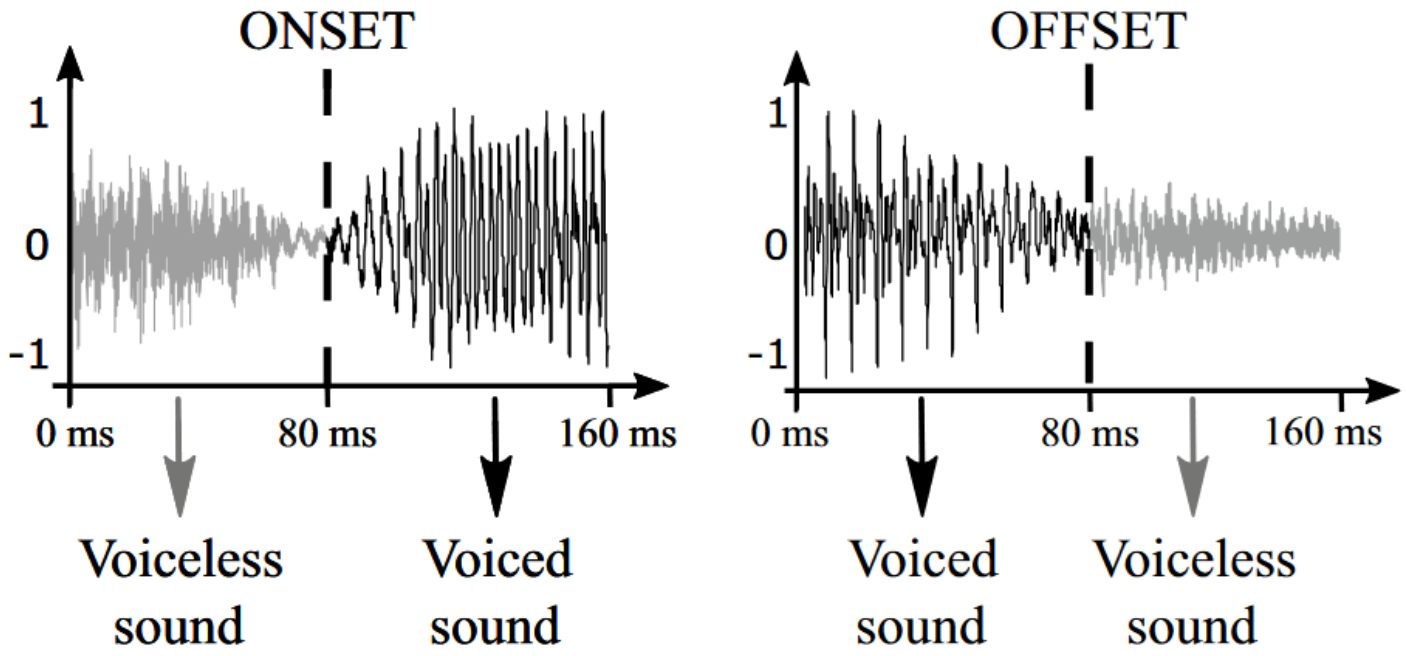
The proposed approach overcomes the state-of-the-art in several aspects:

- Accuracies of up to 89% are obtained for the classification of Parkinson's patients vs. healthy speakers
- Relative to previous studies robust modeling of the speech impairments in patients with neuro–degenerative disorders

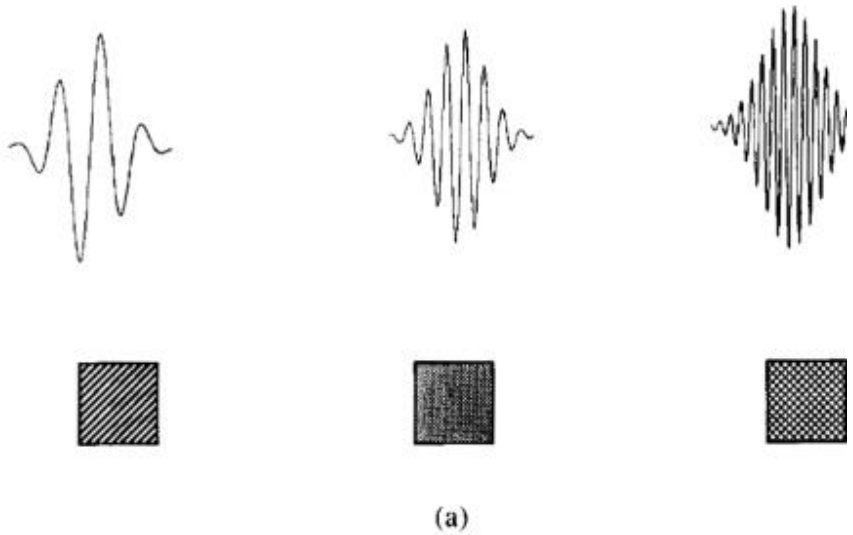
Methods

- The detection of the onset and offset transitions,
- The computation of the time–frequency representations : the short time Fourier transform, and the continuous wavelet transform
- The feature learning classification using the CNN.

Onset and Offset detection



Short time Fourier transform (STFT)



- Fig.Basis functions and time-frequency resolution of the Short Time Fourier Transform(STFT), from [1] . (a) Basis functions, (b) Coverage of time-frequency plane.

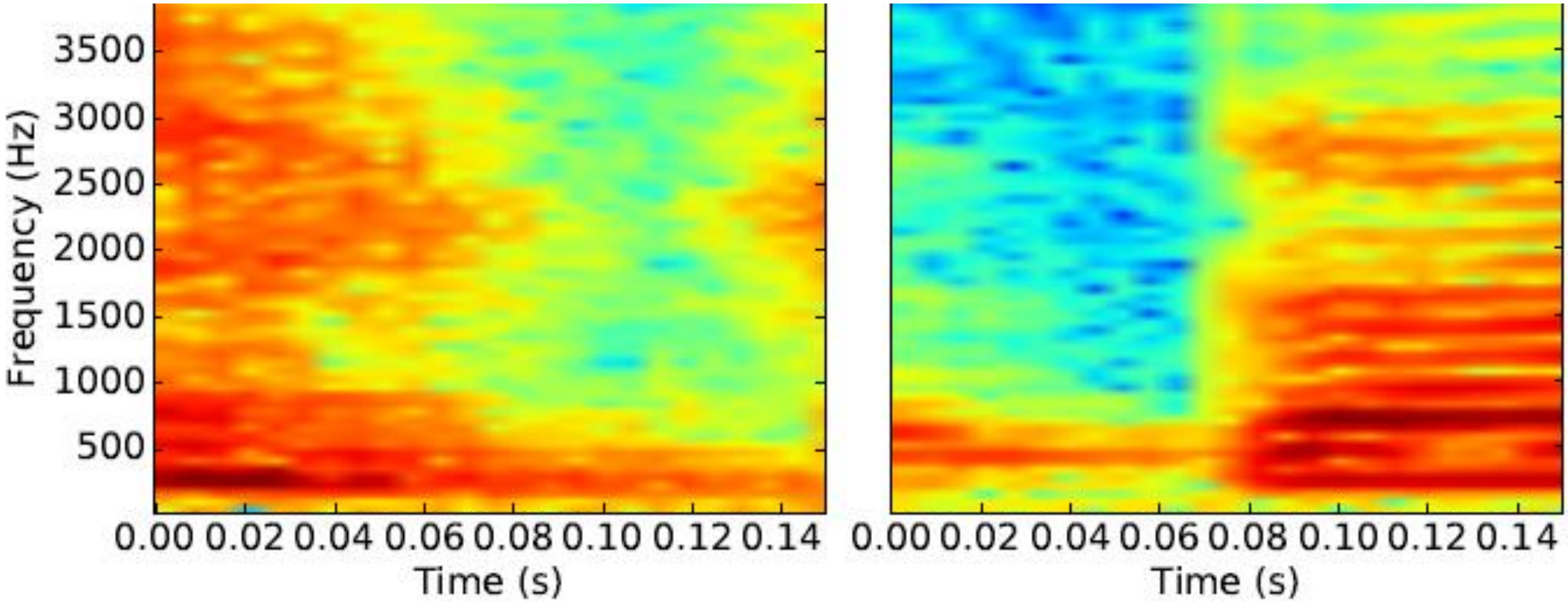
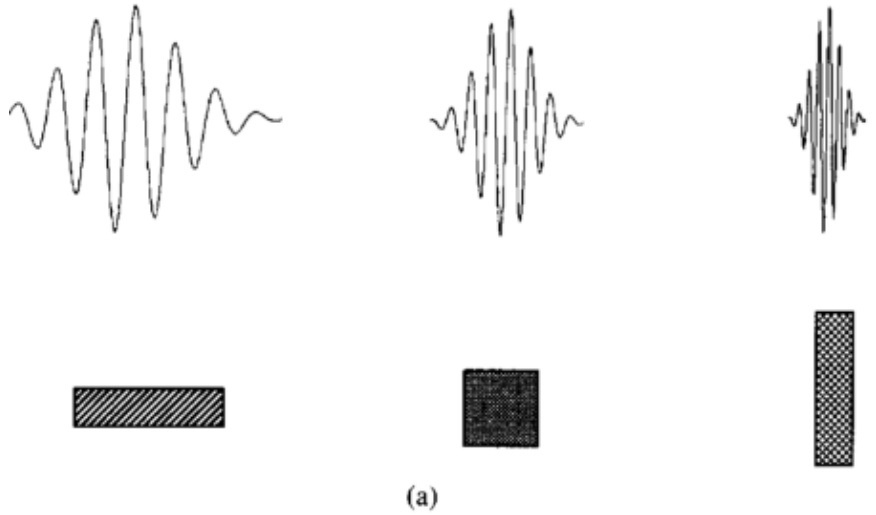


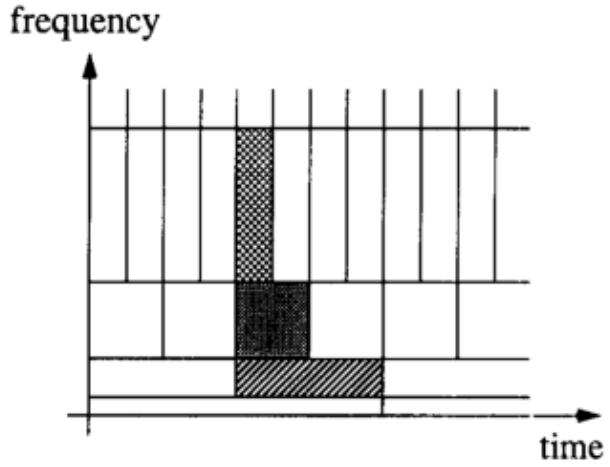
Figure 1: *STFT of the onset for a PD patient (left) and a HC speaker (right) when they pronounce the syllable /ta/*

Continuous wavelet transform (CWT)

- Basis functions and time-frequency resolution of the Wavelet Transform, from [1]. (a) Basis functions, (b) Coverage of time-frequency plane



(a)



(b)

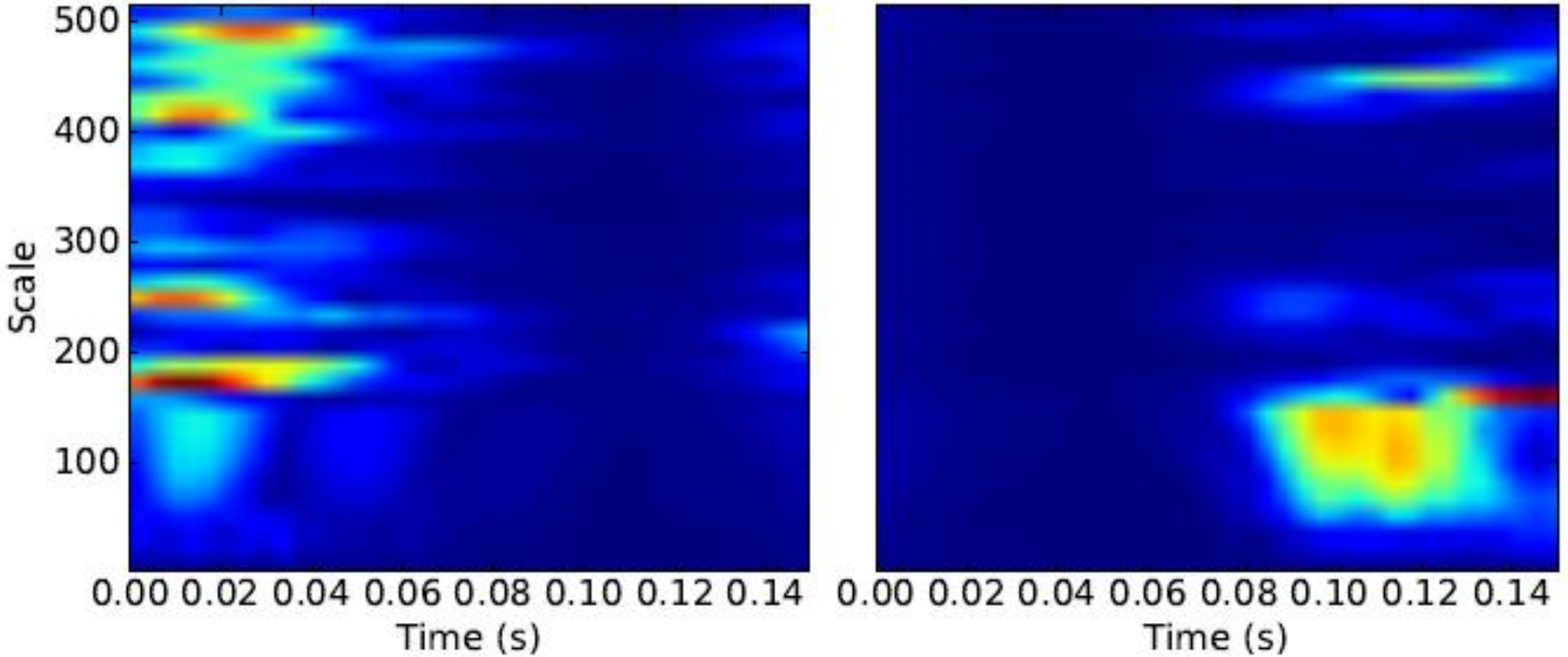


Figure 2: *CWT of the onset for a PD patient (left) and a HC speaker (right) when they pronounce the syllable /ta/*

Architecture of CNN

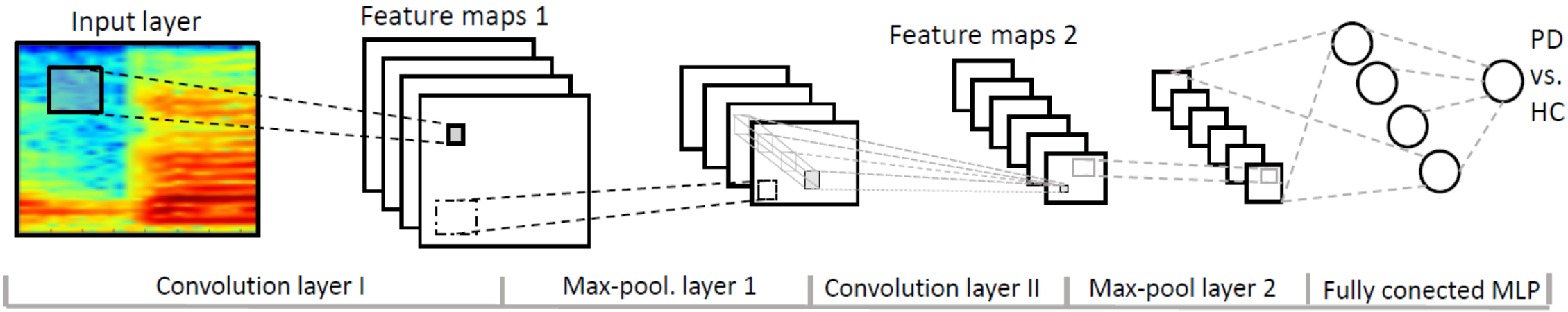


Figure 3: *Architecture of the convolutional neural network implemmented for this study*

Dataset

- Spanish
- German
- Czech

Results- Classification in the same language

Table 2: *Accuracies (%) for classification of PD patients vs. HC speakers in three different languages*

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

Results- Accuracies obtained with the transitions
from read text, monologue, and the rapid repetition of /pa-taka/
in Spanish, German, and Czech

Table 3: *Individual accuracies (%) for monologues, read texts, and the repetition of /pa-ta-ka/ in the three different languages*

Speech Task	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- Cross-language classification

Table 4: *Accuracies (%) for classification of PD patients vs. HC speakers in three different languages when the train and the target languages are different*

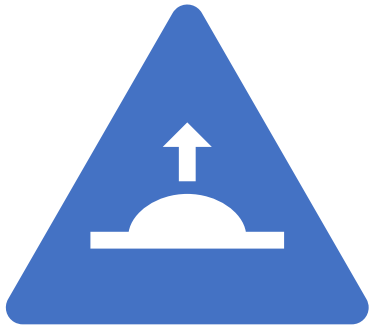
Test Lang.	TFR	onset	offset	onset+offset
Train with Spanish				
German	STFT	51.7	50.2	54.7
German	CWT	50.8	50.3	50.6
German	Baseline	53.7	55.0	54.1
Czech	STFT	53.0	55.0	51.7
Czech	CWT	55.2	55.4	57.9
Czech	Baseline	60.3	57.4	60.4
Train with German				
Spanish	STFT	58.0	55.7	55.8
Spanish	CWT	51.5	51.3	50.8
Spanish	Baseline	53.5	53.5	53.6
Czech	STFT	53.0	52.4	53.0
Czech	CWT	53.1	51.7	52.5
Czech	Baseline	50.9	51.7	52.6
train with Czech				
Spanish	STFT	55.1	51.2	50.5
Spanish	CWT	53.8	56.3	56.7
Spanish	Baseline	53.4	51.6	52.4
German	STFT	54.0	51.8	54.0
German	CWT	50.8	50.2	50.6
German	Baseline	51.2	51.0	50.7

References

1. Vetterli, Martin, and Cormac Herley. "Wavelets and filter banks: Theory and design." *IEEE transactions on signal processing* 40.ARTICLE (1992): 2207-2232.
2. Orozco-Arroyave, Juan Rafael, et al. "Towards an automatic monitoring of the neurological state of Parkinson's patients from speech." *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016.

Questions

Write briefly about STFT and CWT



What are onset and offset transitions and how they are detected ?