Effect of Noise and Model Complexity on Detection of Amyotrophic Lateral Sclerosis and Parkinson's Disease using Pitch and MFCC

Tanuka B⋆, Jhansi M⋆, Yamini BK†, Nalini A‡, Ravi Y‡, Pradeep R‡, Dipanjan G⋆⋆, Prasanta Kumar G⋆

⋆EE Dept., ⋆⋆ECE Dept., Indian Institute of Science, Bangalore 560012, India

†Dept. of SPA , ‡Dept. of NEURO., National Institute of Mental Health and Neurosciences, Bangalore 560012, India

Introduction

- **Dysarthria** due to **Amyotrophic Lateral Sclerosis (ALS)** and **Parkinson's Disease (PD)** impacts **articulation**, **respiration**, **phonation** and **prosody** in an individual's speech.
- Complex classifiers, especially deep neural networks, exploit speech cues for detection of ALS and PD.

Limitations of Existing Works:

- Models are highly expensive in terms of run-time and memory requirements.
- Models are mostly analyzed using clean speech recorded in controlled and noise-free laboratory environments.

Objective:

- To explore the **robustness of prosody (pitch) and articulation (MFCC) cues** against background noise and model complexity for **ALS vs. Healthy (HC)** and **PD vs. HC** classification

Dataset

- Speech data were collected at NIMHANS, Bengaluru, India.

Subjects	ALS: 38 M, 21 F; PD: 45 M, 14 F; HC: 44 M, 16 F
Recording device	Zoom H6 with XYH-6 capsule
Sampling frequency	44.1 kHz (downsampled to 16 kHz)
Speech task	Spontaneous speech in native language on 1. a festival you celebrate (~1 min) 2. a tourist place that you have visited (~1 min)
Languages	Bengali, Hindi, Kannada, Odiya, Tamil, and Telugu
Total duration	5.62 hours

Speech Features

- **Pitch (1D)** captures speaking rate which is lowered in ALS and PD.
  Algorithm: *SWIPE, PEFAC*
- **MFCC (39D)** captures spectral properties which are altered in ALS and PD due to improper vocal tract shape.
  Toolkit: *KALDI*

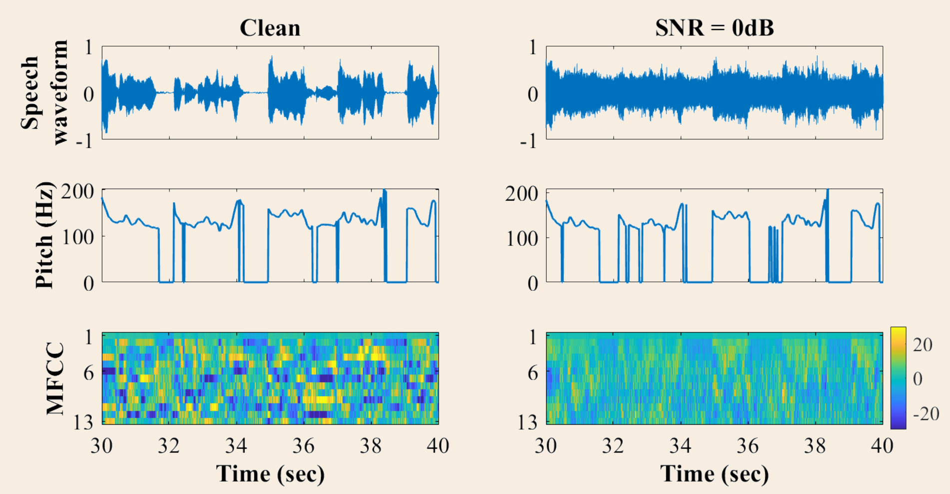
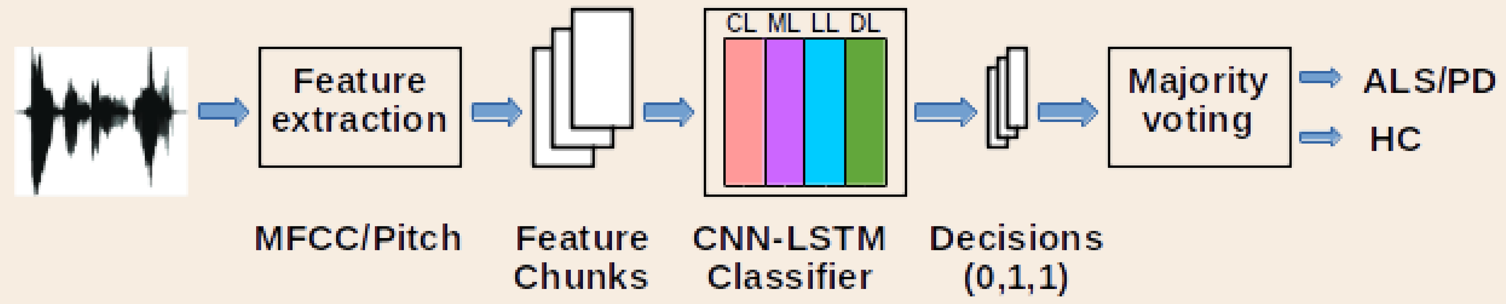


Figure: Illustration of pitch (SWIPE) and MFCC obtained from a 10 sec speech segment of an ALS patient under clean and 0 dB AWGN conditions

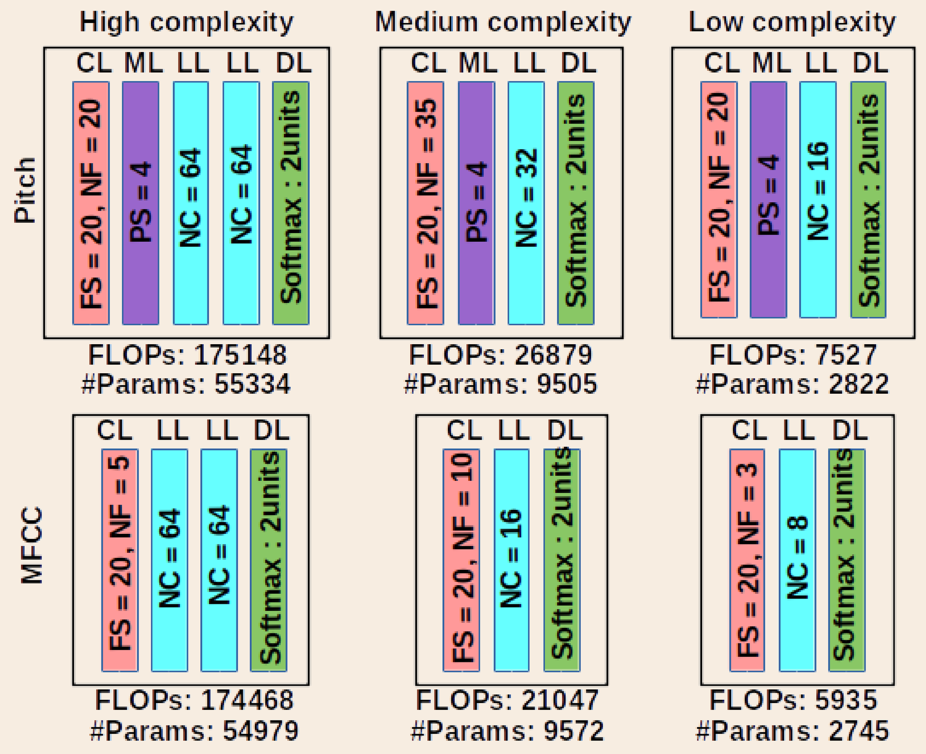
ALS/PD vs. HC Classification

- **CNN-LSTM classifier**:



CL: CNN layer, ML: Maxpooling layer, LL: LSTM layer, DL: Dense layer

- **Classifier configuration for varied complexity**:



#Params: Memory complexity
FLOPs: Run-time complexity
Activation for CL: *ReLU*
Activation for LL: *tanh*

Pitch - Medium Complexity

Layer	#Params
CL	735
ML	-
LL	8704
DL	66
Total	**9505**

FS: Filter size, NF: #filters, PS: Pooling window size, NC: #LSTM cells

Experimental Settings

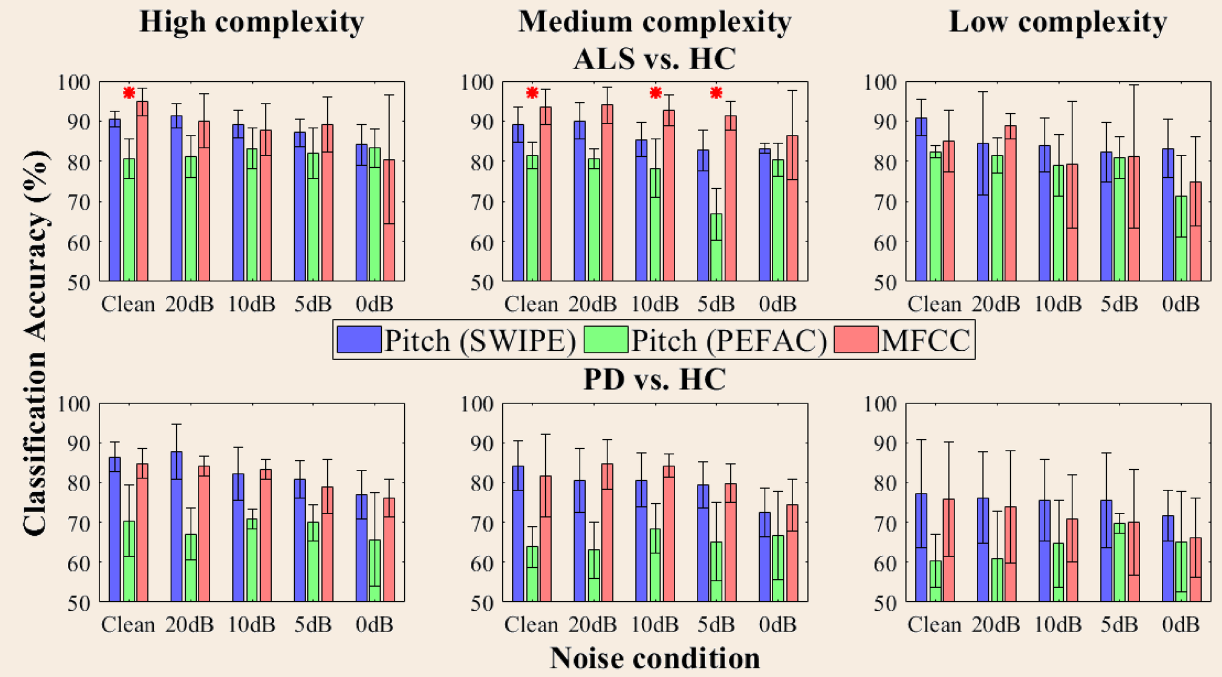
- **Noise condition:** To simulate noisy speech, **AWGN** is added to each utterance at SNRs of **0, 5, 10**, and **20 dB**.
- **Train-test settings**:
  Matched : Noise and SNR of data used in training and testing are matched.
  Mismatched : Noisy and clean speech are tested using classifier trained with clean data.
- **Validation protocol**: 5 fold cross-validation
- **Evaluation metrics** : 1. Classification accuracy,
  2. Wilcoxon signed rank test at 5% significance level

References

- [1] Lavoisier Leite Neto and Ana Carolina Constantini, "Dysarthria and quality of life in patients with Amyotrophic Lateral Sclerosis," Revista CEFAC, vol. 19, no. 5, pp. 664–673, 2017.
- [2] G. Moya-Galé and E. S. Levy, "Parkinson's disease-associated dysarthria: prevalence, impact and management strategies," Research and Reviews in Parkinsonism, vol. 9, pp. 9–16, 2019.
- [3] Jhansi Mallela et al., "Voice based classification of patients with Amyotrophic Lateral Sclerosis, Parkinson's disease and healthy controls with CNN-LSTM using transfer learning," in ICASSP, IEEE, 2020, pp. 6784–6788.
- [4] Suhas BN et al., "Comparison of speech tasks and recording devices for voice based automatic classification of healthy subjects and patients with Amyotrophic Lateral Sclerosis," in INTERSPEECH, 2019, pp. 4564–4568.

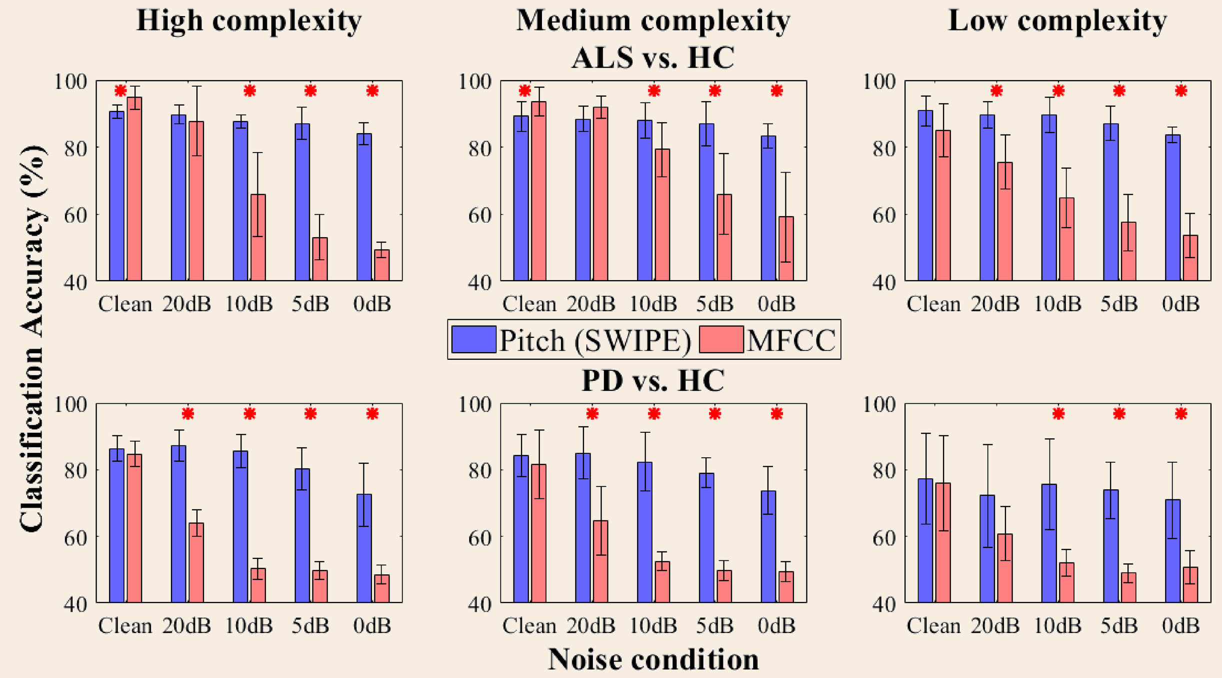
Results & Discussion

- **Matched Train - Test**:



- SWIPE outperforms PEFAC.
- Pitch is as informative as MFCC, mainly for low complexity classifiers.
- Pitch based classifiers are more consistent across folds.

- **Mismatched Train - Test**:



- Performance of pitch is mostly unchanged with decreasing SNR.
- Performance using MFCC drops drastically.
- Pitch is more robust to unseen SNR conditions.

Figures illustrate mean classification accuracy with SD in error bar. Here * indicates that the performance of pitch (SWIPE) and MFCC differ at 5% significance level.

Conclusion

- **Key Takeaways**:
  - Pitch provides similar level of distinctive information as MFCC in clean and matched conditions.
  - Pitch is more noise robust in mismatched condition.
  - Pitch provides classifiers with better generalization ability to unseen SNR conditions.
- **Future work** :
  - To examine the noise robustness of different speech features in various additive noise conditions as well as real noisy recordings
  - To experiment using denoising algorithms in matched and mismatched cases

Acknowledgement: Authors thank the **Department of Science and Technology, Govt. of India** for their support in this work.

tanukab@iisc.ac.in, jhansimallela5@gmail.com, prasantg@iisc.ac.in