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

Tanuka Bhattacharjee[1], Jhansi Mallela[1], Yamini Belur[2], Nalini Atchayaram[3], Ravi Yadav[3], Pradeep Reddy[3], Dipanjan Gope[4], Prasanta Kumar Ghosh[1]

[1]SPIRE LAB, EE Dept., [4]ECE Dept., IISC, Bangalore, India
[2]Dept. of SPA and [3]Dept. of NEURO., NIMHANS, Bangalore, India

ICASSP 2021

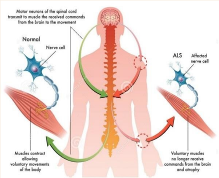
Overview

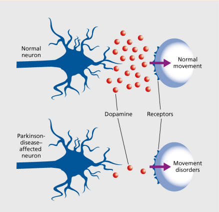
ALS and PD - Overview

Amyotrophic Lateral Sclerosis (ALS)[1] and **Parkinson's Disease (PD)**[2] are *incurable neuro-degenerative* disorders which affect *muscle movements*.

ALS

- Motor neurons degenerate.
- Brain loses its ability to initiate, control and coordinate voluntary muscle actions.





PD

- Dopaminergic neurons degenerate.
- Reduced level of dopamine leads to muscle movement related disorders.

1. https://www.als.org/understanding-als/what-is-als/
2. https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/

ALS and PD - Diagnosis

- Early detection and commencement of therapy can prolong the life expectancy and enhance the quality of life.
- **Unfortunately, no single blood or laboratory test can confirm ALS[1] or PD[2].**
- Diagnosis is based on subjective assessment of symptoms, medical histories, neurological and physical examinations[1,2].
  - ✗ Highly time expensive
  - ✗ Prone to subjective errors and biases
- **Accurate automated diagnostic tool is a need of the hour.**

1. https://www.ninds.nih.gov/disorders/patient-caregiver-education/fact-sheets/amyotrophic-lateral-sclerosis-als-fact-sheet
2. https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/

ALS and PD - Speech-based Markers

- **Dysarthria** is prevalent in both ALS and PD.
  - Early sign of ALS in ∼30% of the patients, with almost all patients developing it in later stages[1]
  - Experienced by ∼90% of PD patients[2]
- Speech functions including **articulation**, **respiration**, **phonation** and **prosody** are reported to get affected[3,4].
- Cues related to these speech components can act as potential bio-markers of ALS and PD.

1. Barbara Tomik and Roberto J Guiloff, "Dysarthria in Amyotrophic Lateral Sclerosis: a review," Amyotrophic Lateral Sclerosis, vol. 11, no. 1-2, pp. 4–15, 2010.

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. 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.

4. Serge Pinto et al., "Treatments for dysarthria in Parkinson's disease," The Lancet Neurology, vol. 3, no. 9, pp. 547–556, 2004.

Literature Survey

SPIRE LAB

Objective	Feature	Classifier
ALS/PD vs. Healthy classification	MFCC[1,2]	DNN[1], CNN-LSTM[2]
	Raw speech waveform[3]	CNN-BLSTM[3]
Classification & severity prediction of PD	MFCC, CSD, spectral dynamics, fundamental frequency variation[4]	Random Forest[4]

1. 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.

2. 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.

3. Jhansi Mallela et al., "Raw speech waveform based classification of patients with ALS, Parkinson's disease and healthy controls using CNN-BLSTM," in INTERSPEECH, 2020, pp. 4586–4590.

4. Taha Khan et al., "Assessing Parkinson's disease severity using speech analysis in non-native speakers," Computer Speech Language, vol. 61, pp. 101047, 2020.

Limitations

- **Models are highly expensive in terms of both run-time and memory requirements.**
  - Powerful computing resources are essential
  - Restricts practical deployment in systems with limited computational resource like mobile phones, general purpose computers
- **Models are mostly analyzed using clean speech recorded in controlled and noise-free laboratory environments.**
  - Presence of background noise in the speech data is inevitable in practice
  - Noise may lead to misclassification which could be fatal

Our Objective

- **To explore the robustness of different speech cues** against
  - the influence of background noise
  - the constraint of low complexity classifier
- **Focus on Pitch and MFCC**
  - MFCC is known to be suitable for the task at hand
  - Pitch is relatively unexplored but reported to get affected as a prosodic component of speech in these diseases
  - Feature fusion is not considered as it increases computational complexity

Overview

Dataset Description

- All speech data were collected at **National Institute of Mental Health and Neurosciences (NIMHANS)**, Bengaluru, India.

Table: Gender and age details of subjects

Condition	#Male	#Female	#Subjects	Age range (years)
ALS	38	21	59	36 - 75
PD	45	14	59	35 - 79
Healthy (HC)	44	16	60	22 - 53
Total	**127**	**51**	**178**	**22 - 79**

- Subjects had **six** different **native languages** - Bengali, Hindi, Kannada, Odiya, Tamil, and Telugu

Dataset Description

- **Audio Recorder:** Zoom H6 with XYH-6 stereo microphone capsule
- **Sampling frequency:** 44.1 kHz (downsampled to 16 kHz)
- **Speech task:** Spontaneous speech in native language on
  1. a festival you celebrate ($\sim$1 min)
  2. a place that you have recently visited ($\sim$1 min)
- **Total data duration**: 5.62 hours (considering all subjects)

Overview

Classification Tasks

1. ALS vs. HC
2. PD vs. HC

Features

Pitch

- Associated with prosody
- Captures speaking rate along with other prosodic features
- Since both ALS and PD lower the speaking rate of an individual, cues related to these diseases can be learned from pitch patterns in a data-driven manner

Specifications

- *Estimation algorithm used:* SWIPE[1] and PEFAC[2]
- *Feature vector dimension:* 1
- Estimated every 10 ms
- Pitch value for unvoiced/silence regions are set to 0

1. Arturo Camacho and John G Harris, "A sawtooth waveform inspired pitch estimator for speech and music," The Journal of the Acoustical Society of America, vol. 124, no. 3, pp. 1638–1652, 2008.

2. Sira Gonzalez and Mike Brookes, "A pitch estimation filter robust to high levels of noise (PEFAC)," in 19^{th} European Signal Processing Conference. IEEE, 2011, pp. 451–455.

Features

Mel-frequency cepstral coefficients (MFCC)

- Associated with spectral properties of speech
- Since muscle weakening in ALS and PD leads to improper vocal tract shape and alters spectral characteristics, cues indicative of this aspect of the impairment can be learned from MFCC in a data-driven manner

Specifications

- *Toolkit used:* KALDI[1]
- *Feature vector dimension:* 39 [13 MFCC + 13ΔMFCC + 13Δ^2MFCC]
- Estimated using 20 ms frame length and 10 ms overlap

1. Daniel Povey et al., "The Kaldi speech recognition toolkit," in Workshop on automatic speech recognition and understanding. IEEE Signal Processing Society, 2011.

Features - An Illustrative Example

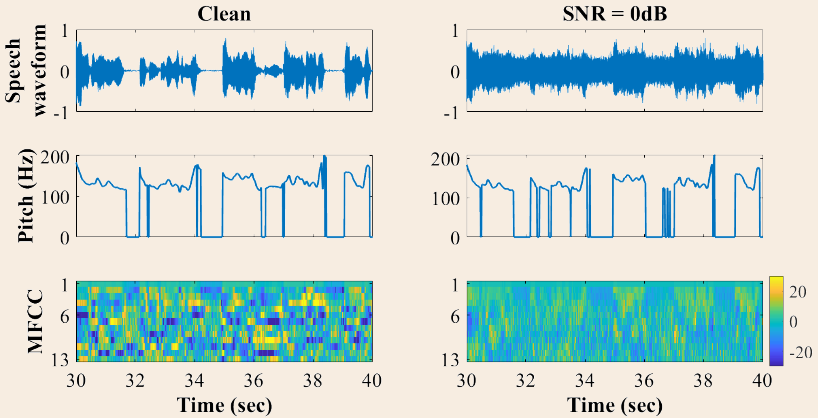
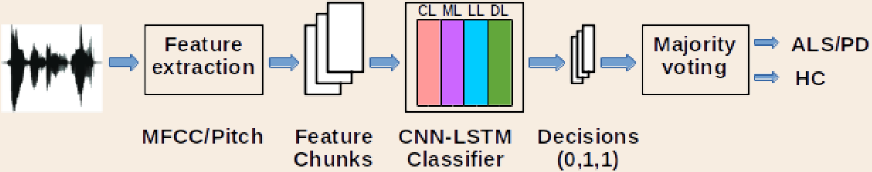


Figure: Illustration of pitch (SWIPE) and MFCC obtained from a 10 sec speech segment of an ALS patient under clean and 0 dB AWGN conditions

Classifier



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

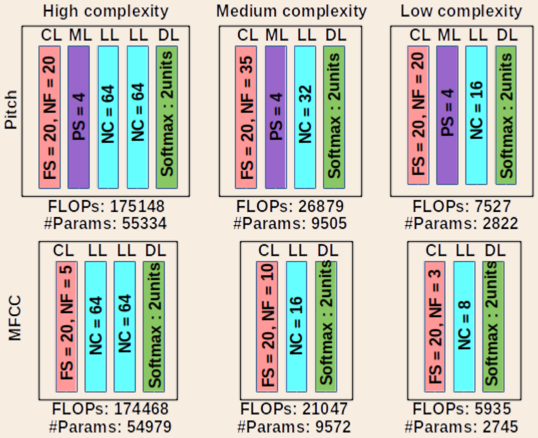
Jhansi Mallela et al., "Voice based classification of patients with Amyotrophic Lateral Sclerosis, Parkinson's disease and healthy controls with CNN-LSTM using transfer learning," in ICASSP, IEEE, 2020, pp. 6784–6788.

Classifier Complexity

- **Memory complexity**
  number of network parameters (#params)
- **Run-time complexity**
  number of floating point operations (FLOPs) needed by the network

We analyze CNN-LSTM models of three different levels of complexity - **Low, Medium and High**

Classifier Configuration



CL: CNN layer, ML: Maxpooling layer, LL: LSTM layer, DL: Dense layer, FS: Filter size, NF: Number of filters, PS: Pooling window size, NC: Number of LSTM cells

Activation for CL: *ReLU*
Activation for LL: *tanh*

Pitch
Medium Complexity Model

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

Noise Conditions

Noise: Additive White Gaussian Noise (AWGN)
SNR: 0, 5, 10 and 20 dB

Train-Test Settings

- **Matched:** Noise and SNR of the data used in training and testing the classifier are matched
- **Mismatched:** Classifier trained with clean data is used to test both clean and noisy test samples

Overview

Experimental Settings

- **Validation Protocol:**
  5-fold cross-validation
  (Each fold contains almost equal number of subjects from ALS/PD and HC classes)

- **Evaluation metrics:**
  - Classification accuracy
  - Wilcoxon signed rank test[1] at 5% significance level (i.e., $p < 0.05$) - to examine if the classification accuracies obtained using pitch and MFCC are significantly different across 5 folds

1. RF Woolson, "Wilcoxon signed-rank test," Wiley encyclopedia of clinical trials, pp. 1–3, 2007.

Overview

Matched Train - Test Results

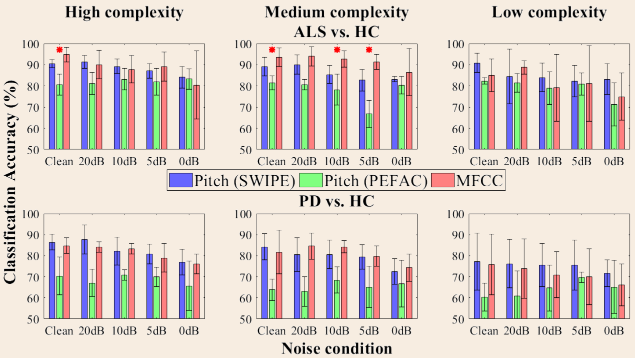


Figure: Mean classification accuracies (SD in error bar) for matched train-test condition
Here * indicates that the performance of pitch (SWIPE) and MFCC differ at 5% significance level

- 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 Results

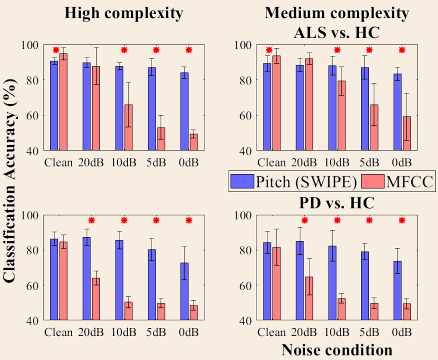


Figure: Mean classification accuracies (SD in error bar) for mismatched train-test condition
Here * indicates that the performance of pitch (SWIPE) and MFCC differ at 5% significance level

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

Overview

Key Takeaways

- Pitch is observed to provide similar level of distinctive information as MFCC in clean and matched train-test conditions.
- Pitch is found to be more noise robust in mismatched train-test condition.
- Pitch provides the 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
- Experimentation using denoising algorithms in both matched and mismatched cases

Acknowledgement

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

THANK YOU

Have Questions/Suggestions?
Write to us @ `spirelab.ee@iisc.ac.in`