Raw speech waveform based classification of patients with ALS, Parkinson's Disease and healthy controls using CNN-BLSTM

Jhansi Mallela[1], Aravind Illa[1], Yamini Belur[2], Nalini Atchayaram[3], Ravi Yadav[3], Prasanta Kumar Ghosh[1]

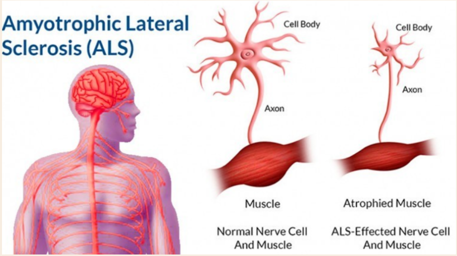
[1]SPIRE LAB, EE Dept., IISC, Bangalore, India
[23]Dept. of SPA and NEURO., NIMHANS, Bangalore, India

INTERSPEECH 2020

Overview

Amyotrophic Lateral Sclerosis (ALS)

- A motor neuron disorder
- Occurs due to gradual degeneration of motor neurons
- Neurons provide a communication link between the brain and voluntary muscles
- Due to the degeneration, there is a loss of muscle control



The ALS Association,"What is ALS?", May 2019.

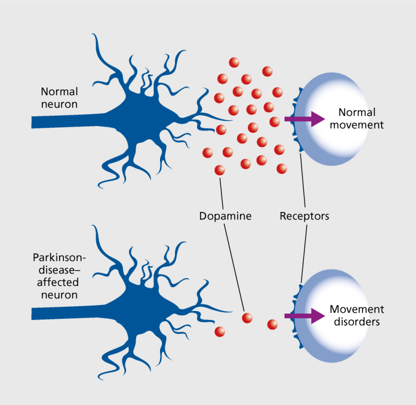
Symptoms of ALS

- Muscle stiffness
- A hard time in holding items
- Muscle cramps
- Swallowing problems
- Speech difficulties (slurred or slowness)

Mayo Clinic, 'Amyotrophic Lateral Sclerosis - Symptoms and causes', 6 August 2019

Parkinson's Disease (PD)

- A progressive brain disorder
- Occurs when Nerve cells / neurons in the brain get impaired and/or die
- These neurons produce Dopamine
- Reduced Dopamine levels in brain results in movement issues



Parkinson's Foundation, "What is Parkinson's? "

Symptoms of PD

- Tremors (hands, arms, legs)
- Stiffness in limbs and trunk
- Slowness in movements
- Difficulty in swallowing and chewing
- Speech difficulties (slurred or slowness)

Mayo Clinic, 'Parkinson's disease - Symptoms and causes', 30 June 2018

Life expectancy with ALS or PD

- ALS affected people[1,2]
  - 50% people - 3 or more years
  - 20% people - 5 or more years
  - 10% people - 10 or more years
- PD affected people[3]
  - Has increased from 9.4 (1967) to 14.6 years (2016) after diagnosis

1. C. Arthur, et al., "Projected increase in Amyotrophic Lateral Sclerosis from 2015 to 2040,"Nature communications, vol. 7, p. 12408, 2016

2. A. Nalini, et al., "Clinical characteristics and survival pattern of 1153 patientswith Amyotrophic Lateral Sclerosis: experience over 30 years fromIndia,"Journal of the Neurological Sciences, vol. 272, no. 1-2, pp. 60–70, 2008.

3. Golbe, Lawrence I., and Cristian E. Leyton. "Life expectancy in Parkinson disease." (2018): 991-992.

Diagnosis and Treatment for ALS and PD

- Currently no specific tests can confirm of having ALS or PD [1]
- Diagnosis is based on medical history (11 months) and a neurological examination
- No cure for either ALS or PD although there exists some treatment for managing their symptoms [1,2,3]

1. Veritas Neuro, "ALS vs Parkinson's - How Do These Conditions Differ?" Aug 2019. [Online]. Available: https://alstreatment.com/als-vs-parkinsons/
2. A. N. Lieberman, "Update on Parkinson disease." New York State Journal of Medicine, 1987.
3. J.-P. Julien, "ALS: Astrocytes move in as deadly neighbors," Nature Neuroscience, vol. 10, no. 5, pp. 535–537, 2007.

Motivation

- Main Objective : Automated methods for detection of ALS or PD could reduce diagnosis time
- Plan for the future : To develop a mobile application that helps in early detection, assists neurologist in diagnosis and to follow the progression of the disease using speech as a biomarker

Challenges

- Identifying speech cues that help in better detection
- Data collection from ALS and PD patients is often tedious making a large corpus development a challenging task - Automated methods require huge amount of data to train a classifier

Speech waveforms and spectrograms of ALS, PD, and Healthy controls

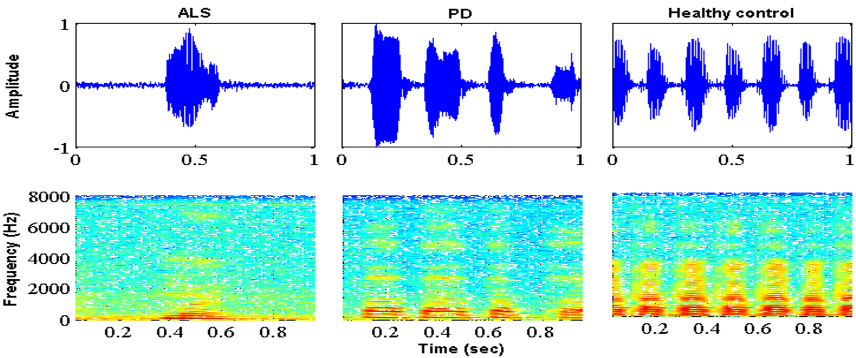


Figure: Subjects repeating syllable "pa"

Literature Survey

- Previous works studied the performance of various speech tasks in automatic classification between ALS and HC using SVM and DNN[1]. As seen, the spectro-temporal characteristics change depending on ALS/HC (and similarly for PD)

1. Suhas B.N., "Comparison of Speech Tasks and Recording Devices for Voice Based Automatic Classification of Healthy Subjects and Patients with Amyotrophic Lateral Sclerosis," in Proc. Interspeech 2019, pp. 4564–4568.

Literature Survey

- Performed classification tasks using MFCCs with CNN-LSTMs[1]. Although MFCCs perform well in speech recognition, speaker verification, and several other speech related problems these hand-crafted features (MFCCs) may not be optimal for classifying ALS, PD and healthy controls.

1. Jhansi M., 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 2020, pp. 6784–6788.

Goal of this work

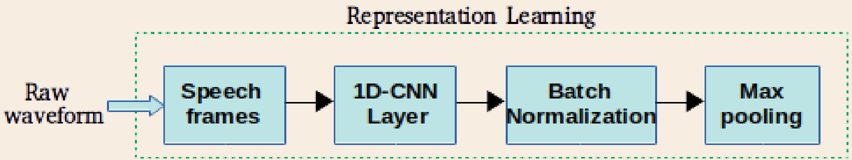
- To study the performance of using raw speech waveform for 3 different classification tasks (ALS vs HC), (PD vs HC), and (ALS vs PD) with 4 different speech tasks
- To study the performance of each classification task in two scenarios:
  - trained and tested in a task-specific manner,
  - trained on data pooled from all tasks, and test on each task separately
- To compare the classification performance using MFCCs and the proposed approach using raw speech

Overview

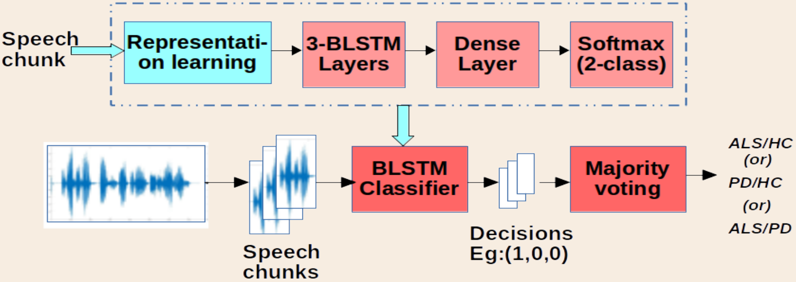
Learning representations from raw waveform and classification using CNN-BLSTM network

- The work presented here makes use of learning task-specific features from the raw waveform using an end-to-end network
- Unlike the handcrafted features (MFCCs) that are computed at frame level, representation learning using 1-D CNN might discriminate speech of ALS and PD from those of the healthy subjects in a better way
- The learned representations are fed to the BLSTM network for classification, since LSTM networks are well-suited to capture the temporal characteristics from the time series data

Representation learning



CNN-BLSTM architecture



Overview

Data collection

- Speech recordings collected at the National Institute of Mental Health and Neurosciences (NIMHANS), Bengaluru, India
- Recorder : Zoom H6 X/Y recorder
- Sampling frequency : 44.1 kHz

Dataset

- Number of subjects used in this work:
  - 60 ALS (30 Male, 30 Female)
  - 60 PD (34 Male, 26 Female)
  - 60 healthy control (HC) (30 Male, 30 Female)

Condition	Gender	Count	Age Range (Avg) in years
ALS	M	30	33 - 76 (58.60)
	F	30	38 - 75 (56.02)
PD	M	34	34 - 78 (58.22)
	F	26	36 - 74 (56.99)
HC	M	30	26 - 68 (44.21)
	F	30	31 - 65 (46.93)

Table: Subject count and age range for each condition - gender pair

Dataset

- Number of subjects used in this work: 60 ALS, 60 PD, and 60 HC
- Speech tasks: considering all subjects
  - Spontaneous speech (SPON) - 21 hours
  - Diadochokinetic rate (DIDK) - 22.42 hours
  - Sustained phoneme production (PHON) - 25.84 hours
  - Image Description (IMAG) - 25.22 hours

Overview

Experimental setup for proposed approach

- Proposed approach: CNN-BLSTM network using raw waveform as input
  - 5 fold cross validation (each fold consists of 12 ALS, 12 PD, 12 HC)
  - Features: Raw speech waveform downsampled to 8kHz (framed with window length of 20ms & shift of 10ms).
  - No. conv. filters or 'feature maps' $= 256/128/64$ (for which the val. loss is minimal)
  - Activation function : ReLUlog (softmax @ output)
  - Kernel : A 120×1 (represented by 1D Conv)
  - No. BLSTM layers : 3 (tanh as activation function)
  - No. units for each BLSTM layer : 150

Experimental setup for baseline

- Baseline: BLSTM network using MFCCs as input
  - 5 fold cross validation (each fold consists of 12 ALS, 12 PD, 12 HC)
  - Features:MFCC (computed with window length of 20ms & shift of 10ms)
  - No. BLSTM layers : 3 (tanh as activation function)
  - No. units for each BLSTM layer : 150
  - Activation function : softmax @ output

Evaluation metrics

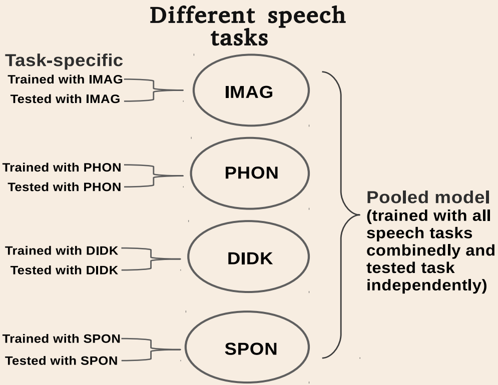
- **Classification accuracy**
- **p-value from sign rank test**[1] : performed using the five folds' classification accuracies of the baseline and the proposed approach where each test fold is again split into three sub folds since a minimum of five variables are required for this test.

1. Wilcoxon signed-rank,test, https://en.wikipedia.org/w/index.php?title=Wilcoxonsigned−ranktestoldid=972074830

Experiments

1. **2 class ALS vs HC**
2. 2 class PD vs HC
3. 2 class ALS vs PD

Experiments



ALS vs HC classification performance

Task-specific model				
Speech tasks	IMAG	PHON	DIDK	SPON
MFCC	90.14 (4.85)	87.67 (1.10)	93.63 (3.35)	96.51 (3.37)
Raw speech	97.31 (1.75)	89.47 (3.82)	96.39 (1.94)	96.16 (2.80)
Pooled model				
Speech tasks	IMAG	PHON	DIDK	SPON
MFCC	93.08 (1.2)	85.56 (5.0)	95.88 (2.2)	98.69 (1.9)
Raw speech	98.02 (1.75)	92.34 (1.43)	96.73 (2.45)	97.86 (2.38)

Table: Average classification accuracy (SD in brackets) of the proposed approach and baseline (MFCC). Blue entries indicate the cases where the proposed approach performs significantly($p < 0.05$) better than MFCC.

Experiments

1. 2 class ALS vs HC
2. **2 class PD vs HC**
3. 2 class ALS vs PD

PD vs HC classification performance

Task-specific model				
Speech tasks	IMAG	PHON	DIDK	SPON
MFCC	84.91 (2.13)	65.39 (2.92)	81.89 (5.32)	90.14 (2.92)
Raw speech	95.01 (2.17)	73.57 (5.05)	89.51 (2.20)	95.71 (4.10)
Pooled model				
Speech tasks	IMAG	PHON	DIDK	SPON
MFCC	91.51 (2.06)	70.25 (8.93)	89.87 (3.48)	95.34 (2.73)
Raw speech	97.28 (2.01)	82.18 (9.86)	93.08 (4.92)	97.41 (3.88)

Table: Average classification accuracy (SD in brackets) of the proposed approach and baseline (MFCC). Blue entries indicate the cases where the proposed approach performs significantly($p < 0.05$) better than MFCC.

Experiments

1. 2 class ALS vs HC
2. 2 class PD vs HC
3. **2 class ALS vs PD**

ALS vs PD classification performance

Task-specific model				
Speech tasks	IMAG	PHON	DIDK	SPON
MFCC	71.88 (3.08)	72.72 (2.91)	79.73 (3.90)	68.25 (10.06)
Raw speech	76.87 (8.20)	72.67 (5.76)	76.43 (2.75)	78.00 (7.84)
Pooled model				
Speech tasks	IMAG	PHON	DIDK	SPON
MFCC	74.56 (4.83)	66.54 (3.81)	78.71 (2.08)	73.47 (10.09)
Raw speech	78.28 (7.32)	69.33 (4.67)	82.09 (3.16)	82.36 (7.84)

Table: Average classification accuracy (SD in brackets) of the proposed approach and baseline (MFCC). Blue entries indicate the cases where the proposed approach performs significantly($p < 0.05$) better than MFCC.

Illustration of learned CNN filters

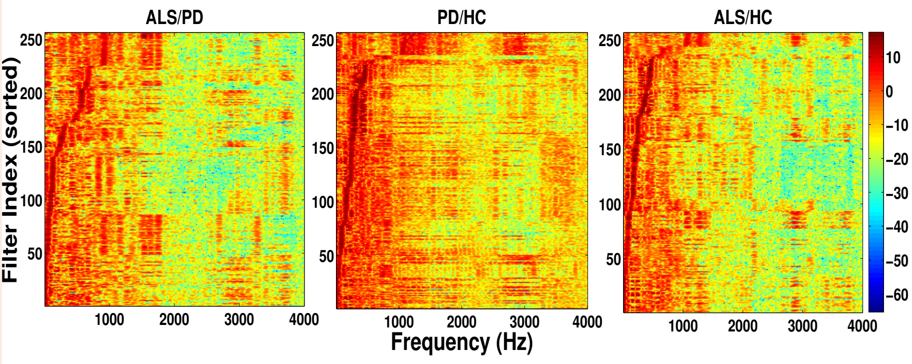


Figure: Magnitude response of 256 filters for ALS/PD, PD/HC, and ALS/HC

1-D CNN output

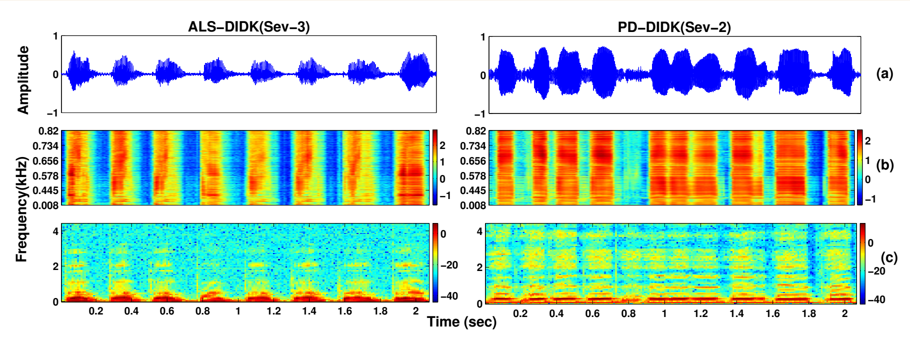


Figure: Illustration of /pa/ sequence spoken by ALS and PD patients using (a) speech waveform, (b) 1-D CNN output, and (c) spectrogram

Overview

Conclusions

- Task-specific+Raw wav outperforms Task-specific+MFCC across all classification tasks (ALS vs HC, PD vs HC, and ALS vs PD)
- Pooled+Raw wav in turn performs better than Task-specific+Raw wav or Pooled+MFCC in all classification tasks
- From the analysis of the learned CNN filter response, it is revealed that the filters are low pass in nature and the center frequencies lie below 800Hz, 500Hz, and 400Hz for ALS/PD, PD/HC, and ALS/HC, respectively

Key Takeaways

1. A comparison of the proposed approach with baseline acoustic features (MFCC) revealed that the proposed approach significantly performs better than baseline.
2. The proposed approach allows the classifier to extract features related to speech rate cues by enhancing vowels in low-frequency regions and help in accurate classification.

Future work

1. To investigate the proposed approach of classification for severity estimation of ALS and PD patients
2. To investigate the proposed approach incorporating the attention mechanism which enables of learning high quality features

Acknowledgement

- Authors thank Ms Renuka and Ms Hima Jyothi for their technical assistance in generating plots.
- 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 at `spirelab.ee@iisc.ac.in`