Speech task based automatic classification of ALS and Parkinson's Disease and their severity using log Mel spectrograms

Suhas BN[1], Jhansi Mallela[2], Aravind Illa[2], Yamini Belur[3], Nalini Atchayaram[4], Ravi Yadav[4], Prasanta Kumar Ghosh[2]

[1]Penn State University,[2]SPIRE LAB, Indian Institute of Science,
[34]Dept. of SPA and Neurology, NIMHANS

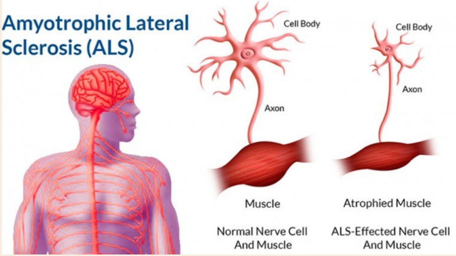
Paper Presentation

July 2020

Overview

Amyotrophic Lateral Sclerosis (ALS)

- A motor neuron disorder
- Neurons : communication link
- Gradual degeneration of motor neurons
- Loss of muscle control



Morris, Jerry. "Amyotrophic lateral sclerosis (ALS) and related motor neuron diseases: an overview." The Neurodiagnostic Journal 55.3 (2015): 180-194.

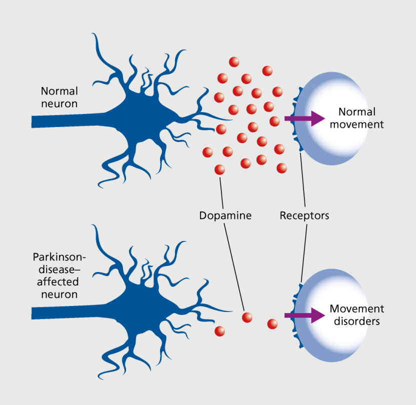
Symptoms of ALS

- 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. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/amyotrophic-lateral-sclerosis/symptoms-causes/syc-20354022

Parkinson's Disease (PD)

- A progressive brain disorder
- Impairment of nerve cells
- Neurons : Dopamine production
- Movement issues



"Parkinson's Disease." National Institute on Aging, U.S. Department of Health and Human Services, May 2017. [Online]. Available: www.nia.nih.gov/health/parkinsons-disease.

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. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/parkinsons-disease/symptoms-causes/syc-20376055

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 that can confirm of having ALS or PD [1]
- Diagnosis
- 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, Plans and Challenges

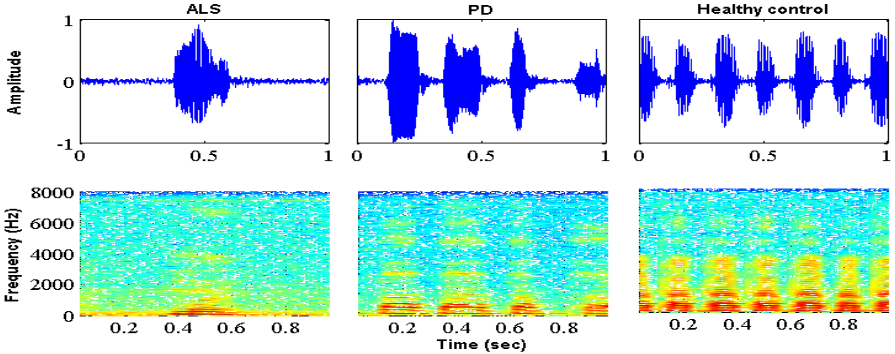
Motivation

- Objective : Automated methods
- What's new? : *Methodology + Severity classification*
- Future Plan: Supplementing diagnosis

Challenges

- Identifying speech cues
- Access for people of different socio-economic backgrounds
- Data collection

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



Literature Survey

- 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)
- CNNs for identifying ALS patients[2] using two 1-D convolution networks - (one each for time and frequency respectively) for filter bank features (MFBE)

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," inProc. Interspeech 2019, 2019, pp. 4564–4568.
2. An, KwangHoon, et al. "Automatic Early Detection of Amyotrophic Lateral Sclerosis from Intelligible Speech Using Convolutional Neural Networks." Interspeech. 2018.

Goal of this work

- Performance of log Mel Spectrograms in 3 class ALS/PD/HC, 5 class ALS/3 class PD severity detection
- Propose a 2D CNN approach that incorporates both Time-Frequency (TF) plane which helps us to model temporal and harmonic structures of audio signals

Overview

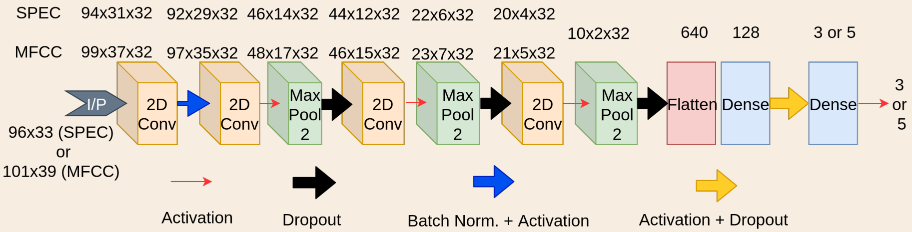
2D CNN and Feature dimensions

- The work makes use of TF plane as a whole and utilizes a 2-dimensional convolutional network for log Mel spectrograms (SPEC) and Mel frequency cepstral coefficients (MFCC)
- Due to the TF plane, the feature is better understood through a 2D CNN (finer modelling of temporal-harmonic structures)

1. SPEC dimension of 96×33 with Melbins[1] $= 96$, and a audio length of 1 second represented by 33 frames
2. MFCC dimension of 101×39 with audio length represented by 101 frames and feature dimension of 39

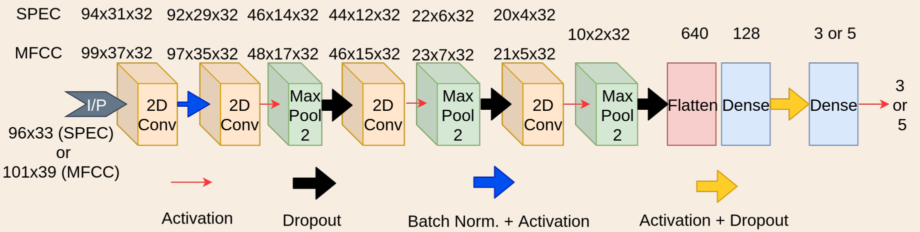
[1] For the same architecture, smaller number of Melbins (24 and 48) led to lower accuracies (0.6-0.75) due to a lower resolution while a higher value of 128 gave similar accuracies when compared to 96

2D CNN Architecture



- No. conv. filters or 'feature maps' $= 32$
- Kernel : A 3×3 (represented by 2D Conv)
- Convolution layer of size $(h \times w \times d)$ learns ' d ' features of size $h \times w$.
- Size of pooling area : 2×2 (represented by Max Pool 2)

2D CNN architecture (contd.)



- Activation function : ReLU (softmax @ output)
- Optimum conv. layer dropout is $= 0.5$; dense layer dropout $= 0.6$
- Loss function : Categorical cross-entropy
- Optimizer : Adadelta

Overview

Data collection

- Collected from National Institute of Mental Health and Neurosciences (NIMHANS), Bengaluru, India
- Recorders : Apple iPhone 7 (IPH), Motorola G5 Plus (MOT), Xiaomi Redmi 4 (XIA), Zoom H6 X/Y recorder (ZOO) and Dell XPS 15 laptop (LAP)
- 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 (30 Male, 30 Female)
  - 60 PD (34 Male, 26 Female)
  - 60 healthy control (HC) (30 Male, 30 Female)

- Speech tasks: considering all subjects across all devices
  - Spontaneous speech (SPON) - 21 hours
  - Diadochokinetic rate (DIDK) - 22.42 hours
  - Sustained phoneme production (PHON) - 25.84 hours
  - Image Description (IMAG) - 25.22 hours

Severity Ratings

- Provided by five speech language pathologists (SLP) from NIMHANS
- Inter-rater reliability [1] has been calculated using the Fleiss' kappa (κ)
  - ALS subjects, $\kappa = 0.9017$ (Almost perfect agreement)
  - PD subjects, $\kappa = 0.6995$ (Substantial agreement)

ALSFRS-R for Speech		UPDRS-III for Speech	
Finding	**Sev**	**Finding**	**Sev**
Normal	4	Normal	0
Detectable speech disturbance	3	Slight loss of expression, diction and/or volume.	1
Intelligible with repeating	2	Monotone, slurred but understandable; moderately impaired.	2
Speech combined with nonvocal communications	1	Marked impairment, difficult to understand.	3
Loss of useful speech	0	Unintelligible.	4

Table: ALSFRS-R (for ALS) and UPDRS-III (for PD) scales used for rating the subjects

1. Zapf, A., Castell, S., Morawietz, L. et al. Measuring inter-rater reliability for nominal data – which coefficients and confidence intervals are appropriate?. BMC Med Res Methodol 16, 93 (2016). https://doi.org/10.1186/s12874-016-0200-9

Overview

Experimental setup

- 5 fold cross validation (each fold consists of 12 ALS, 12 PD, 12 HC)
- Proposed approach: 2D CNN : Uses ReLU activation function (except for softmax before output)
  - Features: SPEC and MFCC (computed for window length of 20ms & shift of 10ms, analysis window of 1s).
- Baseline: SVM and DNN for ALS/PD/HC
  - Features: MFCC (suprasegmental features on 1s analysis window)
  - Kernel function in SVM: Radial basis function
  - DNN: 2-hidden layers with 128/256/512 units in each layer (for which the val. loss is minimized) and output layer with three units (ALS/PD/HC) and softmax activation

Evaluation : AUC-ROC Characteristic curves

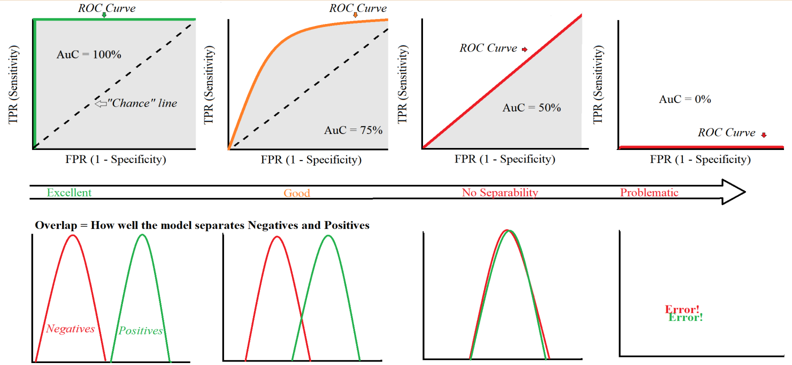


Figure: AUC-ROC curves for 2 class classification with True Positive Rate (TPR) vs False Positive Rate (FPR) on the Y & X axis

$\uparrow$ AUC $\Leftrightarrow$ better ability of the the model at distinguishing between classes.

1. S. Narkhede, "Understanding AUC - ROC Curve," May 2019. [Online]. Available: https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

Experiments

Three sets of classification experiments are carried out:

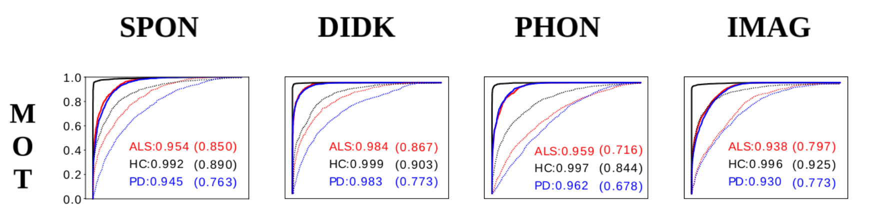
1. **3 class ALS vs PD vs HC**
2. 5 class ALS severity classification
3. 3 class PD severity classification

3 class ALS vs PD vs HC

Speech Task/Device		MOT	ZOO	IPH	XIA	LAP
SPON	SPEC	**0.86 (0.01)**	0.85 (0.01)	0.85 (0.01)	0.84 (0.01)	0.84 (0.02)
	MFCC	0.67 (0.04)	0.68 (0.04)	0.67 (0.08)	**0.68 (0.03)**	0.64 (0.01)
DIDK	SPEC	**0.93 (0.01)**	0.90 (0.01)	0.92 (0.02)	0.90 (0.01)	0.89 (0.01)
	MFCC	**0.74 (0.04)**	0.72 (0.05)	0.73 (0.06)	0.73 (0.01)	0.75 (0.03)
PHON	SPEC	**0.89 (0.0)**	0.80 (0.01)	0.86 (0.01)	0.80 (0.01)	0.83 (0.01)
	MFCC	**0.72 (0.04)**	0.70 (0.05)	0.70 (0.05)	0.68 (0.02)	0.67 (0.07)
IMAG	SPEC	0.84 (0.01)	0.82 (0.01)	**0.86 (0.00)**	0.80 (0.01)	0.81 (0.01)
	MFCC	0.72 (0.01)	**0.77 (0.04)**	0.75 (0.03)	0.66 (0.03)	0.72 (0.01)

Table: Comparison of accuracy (with SD in brackets) across 5 folds for different tasks and devices between MFCC baseline and SPEC

AUC-ROC for 3 class ALS/PD/HC



- Tasks (columns) & devices (rows) with TPR vs FPR on the Y & X axis
- Bold lines : SPEC
  Dashed lines : MFCC.
- AUC scores for MFCC are in brackets next to the SPEC values for reference

Experiments

Three sets of classification experiments are carried out:

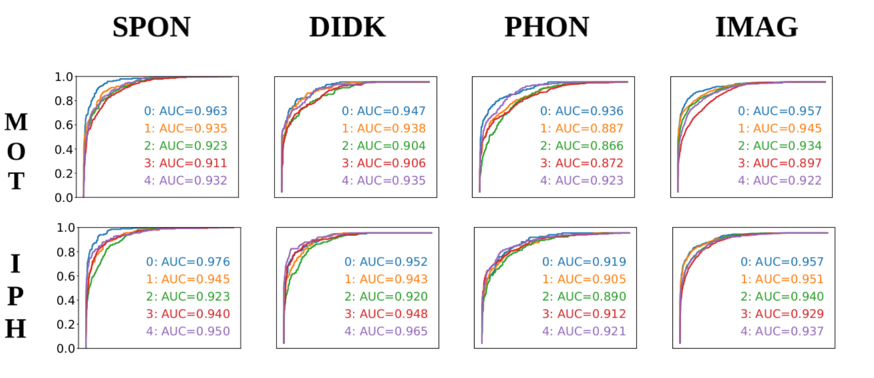
1. 3 class ALS vs PD vs HC
2. **5 class ALS severity classification**
3. 3 class PD severity classification

5 class ALS severity classification

Speech Task/Device		MOT	ZOO	IPH	XIA	LAP
SPON	SPEC	0.76 (0.08)	0.73 (0.09)	**0.80 (0.09)**	0.74 (0.04)	0.74 (0.06)
	MFCC	0.70 (0.09)	0.62 (0.01)	0.72 (0.01)	0.61 (0.05)	0.66 (0.07)
DIDK	SPEC	0.74 (0.01)	0.77 (0.07)	**0.79 (0.01)**	0.78 (0.01)	0.76 (0.01)
	MFCC	0.64 (0.01)	0.67 (0.08)	0.70 (0.01)	0.65 (0.01)	0.65 (0.02)
PHON	SPEC	0.70 (0.07)	0.68 (0.01)	**0.72 (0.01)**	0.71 (0.08)	0.70 (0.08)
	MFCC	0.64 (0.08)	0.55 (0.01)	0.65 (0.01)	0.60 (0.09)	0.58 (0.09)
IMAG	SPEC	**0.77 (0.05)**	0.75 (0.06)	0.74 (0.07)	0.74 (0.08)	0.74 (0.01)
	MFCC	0.67 (0.06)	0.66 (0.08)	0.72 (0.08)	0.70 (0.09)	0.68 (0.02)

Table: Comparison of accuracy (with SD in brackets) across 5 folds for different tasks and devices between SPEC and MFCC for 5 class ALS Severity classification

AUC-ROC for 5 class ALS Severity



Tasks (columns) & devices (rows) with TPR vs FPR on the Y & X axis

Severity: 0 : Severe (loss of useful speech) and

4 : Normal speech (but has ALS)

Experiments

Three sets of classification experiments are carried out:

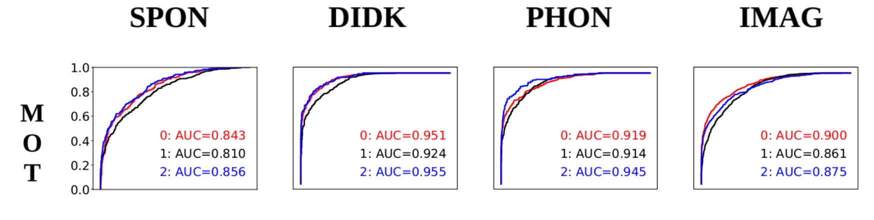
1. 3 class ALS vs PD vs HC
2. 5 class ALS severity classification
3. **3 class PD severity classification**

3 class PD severity classification

Speech Task/Device		MOT	ZOO	IPH	XIA	LAP
SPON	SPEC	0.75 (0.01)	0.76 (0.01)	0.76 (0.01)	0.76 (0.01)	**0.77 (0.06)**
	MFCC	0.65 (0.04)	0.68 (0.01)	0.69 (0.03)	0.63 (0.01)	0.61 (0.01)
DIDK	SPEC	**0.87 (0.01)**	0.87 (0.01)	0.85 (0.01)	0.85 (0.00)	0.86 (0.01)
	MFCC	0.80 (0.01)	0.82 (0.01)	0.84 (0.02)	0.80 (0.01)	0.79 (0.01)
PHON	SPEC	**0.85 (0.01)**	0.81 (0.00)	0.82 (0.01)	0.79 (0.01)	0.77 (0.01)
	MFCC	0.68 (0.04)	0.75 (0.05)	0.62 (0.08)	0.66 (0.09)	0.70 (0.01)
IMAG	SPEC	**0.81 (0.00)**	0.79 (0.01)	0.81 (0.01)	0.76 (0.01)	0.77 (0.01)
	MFCC	0.70 (0.05)	0.74 (0.04)	0.75 (0.02)	0.70 (0.01)	0.67 (0.03)

Table: Comparison of accuracy (with SD in brackets) across 5 folds for different tasks and devices between SPEC and MFCC for 3 class PD Severity classification

AUC-ROC for 3 class PD Severity



Tasks (columns) & devices (rows) with TPR vs FPR on the Y & X axis

Severity:

0: Normal speech (but has PD)

2: Monotone, slurred but understandable; moderately impaired

Overview

Summary

- 2D CNN+SPEC outperforms 2D CNN+MFCC across all speech task-device pairs
- 2D CNN+MFCC in turn performs better than SVM+MFCC or DNN+MFCC for ALS/PD/HC classification
- 5 class ALS Severity : Across all severity-speech task-device combinations, the max(min) AUC score is 0.976(0.866) - good separability between severity classes
- 3 class PD Severity : Across all severity-speech task-device combinations, the max(min) AUC score is 0.951(0.807) - (which may further improve with an increase in Fleiss' kappa (κ))

Key Takeaways

1. SPEC features performed better than MFCC and regardless of the recording device used, similar accuracies were obtained
2. Severity classification (to the best of our knowledge) of a neurological disease such as ALS or PD having not been attempted earlier shows good promise in identifying the condition and also its severity at an earlier stage

References

1. S. N. Awan, N. Roy, D. Zhang, and S. M. Cohen, "Validation of the cepstral spectral index of dysphonia (CSID) as a screening tool for voice disorders: Development of clinical cutoff scores," Journal of Voice, vol. 30, no. 2, pp. 130–144, 2016.
2. A. J. Hughes, S. E. Daniel, L. Kilford, and A. J. Lees, "Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases." Journal of Neurology, Neurosurgery & Psychiatry, vol. 55, no. 3, pp. 181–184, 1992.
3. K. C. Arthur, A. Calvo, T. R. Price, J. T. Geiger, A. Chio, and B. J. Traynor, "Projected increase in Amyotrophic Lateral Sclerosis from 2015 to 2040," Nature communications, vol. 7, p. 12408, 2016.
4. A. Nalini, K. Thennarasu, M. Gourie-Devi, S. Shenoy, and K. Dinakar, "Clinical characteristics and survival pattern of 1153 patients with Amyotrophic Lateral Sclerosis: experience over 30 years from India," Journal of the Neurological Sciences, vol. 272, no. 1-2, pp. 60–70, 2008.
5. A. N. Lieberman, "Update on Parkinson disease." New York State Journal of Medicine, 1987.
6. C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel et al., "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results," Movement disorders: official journal of the Movement Disorder Society, vol. 23, no. 15, pp. 2129–2170, 2008.

Acknowledgement

This work is supported by the Pratiksha Trust and the Department of Science and Technology (DST), Government of India.

THANK YOU

Have Questions/Suggestions?

Write to us at spirelab.ee@iisc.ac.in