Exploring the Role of Fricatives in Classifying Healthy Subjects and Patients with Amyotrophic Lateral Sclerosis and Parkinson's Disease

Tanuka Bhattacharjee[1], Yamini Belur[2], Atchayaram Nalini[3], Ravi Yadav[3], Prasanta Kumar Ghosh[1]

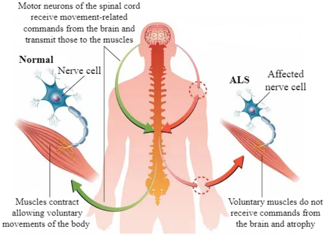
[1]SPIRE LAB, EE Dept., IISc, Bangalore, India
[2]Dept. of SPA and [3]Dept. of NEURO., NIMHANS, Bangalore, India

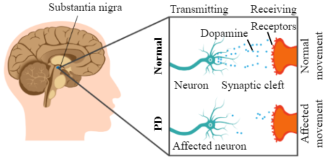
ICASSP 2023

Overview

ALS and PD

- **Amyotrophic Lateral Sclerosis (ALS)** and **Parkinson's Disease (PD)** are **incurable** and **progressive neuro-degenerative** diseases affecting **muscle movements**.
- **Dysarthria** is prevalent in both diseases.
- Speech functions including phonation, articulation, and respiration, are reported to get affected.





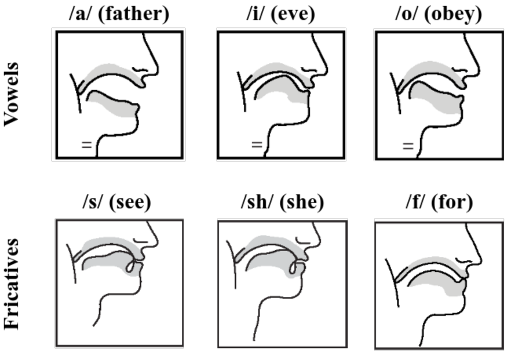
1. Lavoisier Leite and Ana Carolina Constantini, "Dysarthria and quality of life in patients with Amyotrophic Lateral Sclerosis," Revista CEFAC, vol. 19, pp. 664–673, 2017.
2. Serge Pinto et al., "Treatments for dysarthria in Parkinson's disease," The Lancet Neurology, vol. 3, no. 9, pp. 547–556, 2004.

SPP for Dysarthria

- **Sustained Phoneme Production (SPP)** tasks are commonly used in clinical assessment of dysarthria.
    - Simple task
    - Easy to administer
    - Can assess all the required sub-systems of speech, e.g. phonation, articulation, and respiration
- Sustained utterances of different types of phonemes can be examined.
    - Vowels
    - Fricatives

J. 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, pp. 6784–6788, 2020.

Vocal Tract Configurations for Vowels and Fricatives



Physiological mechanisms of uttering vowels and fricatives being different, the impact of dysarthria on their productions may also vary significantly.

Our Objective

- To analyze the **relative utility** of different **sustained fricatives (SFs), as compared to sustained vowels (SVs),** in SPP task based **automatic ALS/PD vs. healthy control (HC)** classification

Literature

Phoneme	Features	Classifier
Vowels /a/, /e/, /i/, /o/, /u/	dynamic articulation transition features[1], STFT[2], MFCC[3], tunable Q-factor wavelet coefficients[4]	SVM[2,4], RF[4], BLSTM[1]
Vowels + Fricatives /a/, /i/, /o/, /u/, /æ/, /s/, /sh/, /f/	MFCC[5], log mel spectrograms[6], 1D-CNN based features from raw speech[7]	SVM[5], 2D-CNN[6], BLSTM[7]

1. C. Quan et al., "A deep learning based method for Parkinson's disease detection using dynamic features of speech," IEEE Access, vol. 9, pp. 10239–10252, 2021.

2. B. Karan et al., "Non-negative matrix factorization-based time-frequency feature extraction of voice signal for Parkinson's disease prediction," Computer Speech Language, vol. 69, pp. 101216, 2021.

3. M. Vashkevich and Y. Rushkevich, "Classification of ALS patients based on acoustic analysis of sustained vowel phonations," Biomedical Signal Processing and Control, vol. 65, pp. 102350, 2021.

4. C. Sakar et al., "A comparative analysis of speech signal processing algorithms for Parkinson's disease classification and the use of the tunable Q-factor wavelet transform," Applied Soft Computing, vol. 74, pp. 255–263, 2019.

5. BN Suhas 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, pp. 4564–4568, 2019.

6. BN Suhas et al., "Speech task based automatic classification of ALS and Parkinson's disease and their severity using log mel spectrograms," in SPCOM, IEEE, pp. 1–5, 2020.

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

Overview

Dataset Description

- **Place of data collection:**
  - National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India
- **Speech task:**
  - Sustained utterances of
    - Vowels - /a/, /i/, /o/
    - Voiceless fricatives - /s/, /sh/, /f/
  - 1-3 utterances per phoneme per subject

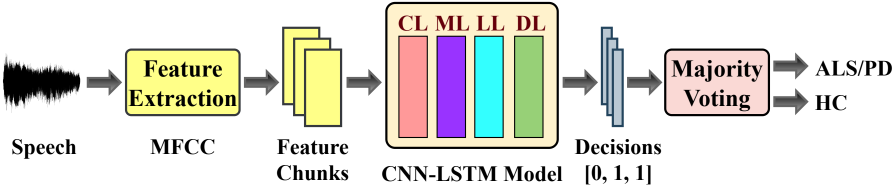
Table: Subject and utterance details

Condition	#M:#F	Age range (years)	#Utterances	Mean (SD) of utterance duration (sec)
ALS	25:10	36 - 70	526	3.30 (2.36)
PD	25:10	45 - 73	528	4.09 (2.53)
HC	25:10	35 - 62	507	5.06 (2.04)

- Data were arranged in 5-fold cross-validation setup with disjoint subjects in the 5 groups.

Overview

Approach



CL: 1D-CNN layer, **ML:** Maxpooling layer, **LL:** LSTM layer, **DL:** Dense layer

J. 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, pp. 6784–6788, 2020.

SFs vs. SVs

Table: Mean classification accuracies in % (SD in bracket) obtained using MFCC of different sustained phonemes

	Phonemes	ALS vs. HC	PD vs. HC
Vowels	/a/	62.88 (7.91)	55.97 (9.89)
	/i/	**78.42 (10.03)**	**72.85 (12.04)**
	/o/	68.40 (5.47)	51.78 (8.73)
	Overall	**69.90**	**60.20**
Fricatives	/s/	76.90 (7.86)	65.37 (7.84)
	/sh/	77.47 (7.56)	66.66 (9.40)
	/f/	72.44 (6.24)	64.70 (10.43)
	Overall	**75.60**	**65.58**

- Fricatives achieve higher mean classification accuracies than /a/ and /o/, though /i/ outperforms all.
- /sh/ achieves the highest mean performance among the fricatives.

SFs vs. SVs

Table: Mean classification accuracies in % (SD in bracket) obtained using MFCC of different sustained phonemes

	Phonemes	ALS vs. HC	PD vs. HC
Vowels	/a/	62.88 (7.91)	55.97 (9.89)
	/i/	**78.42 (10.03)**	**72.85 (12.04)**
	/o/	68.40 (5.47)	51.78 (8.73)
	Overall	**69.90**	**60.20**
Fricatives	/s/	76.90 (7.86)	65.37 (7.84)
	/sh/	77.47 (7.56)	66.66 (9.40)
	/f/	72.44 (6.24)	64.70 (10.43)
	Overall	**75.60**	**65.58**

- Patients seem to find it difficult to form constrictions while producing fricatives, or to proximally position the tongue and palate while uttering /i/.

Spectral Characteristics

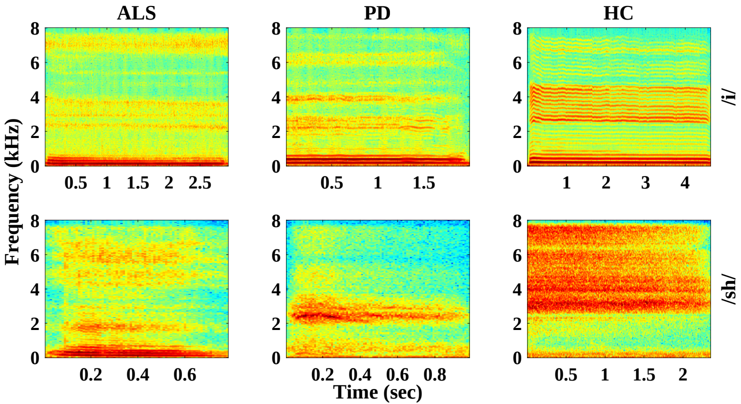


Figure: Spectrograms of sustained utterances of vowel /i/ and fricative /sh/

Unwanted Voicing of Voiceless Fricatives

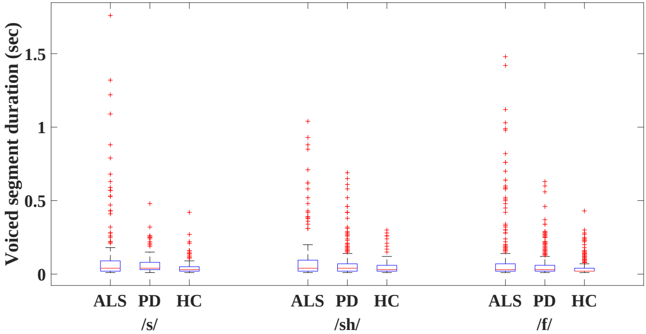
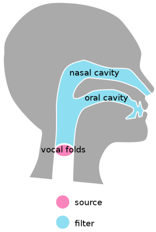
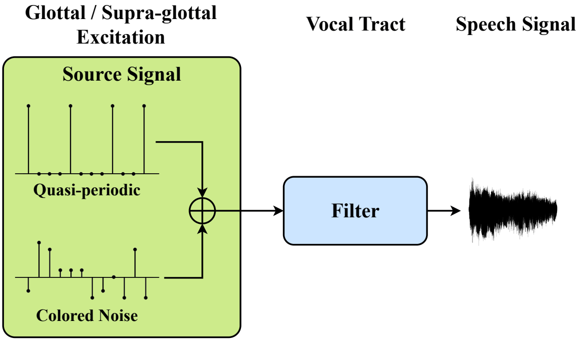


Figure: Distributions of durations of voiced segments detected in different sustained fricative utterances produced by ALS, PD and HC subjects

- Sustained fricatives obtained from ALS and PD subjects have longer voiced segments (at 1% significance level as per Wilcoxon ranksum test) than the sustained fricatives produced by HCs.

Source - Filter Analysis

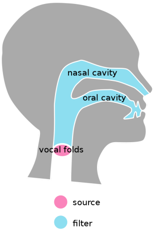


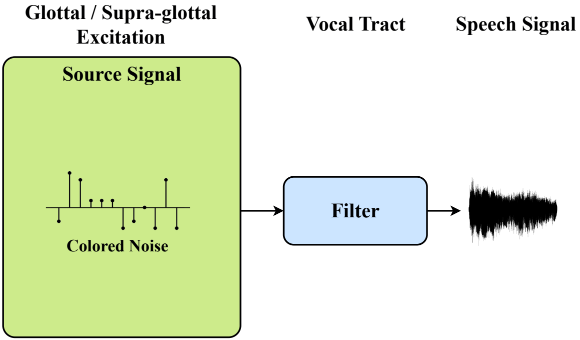


Vowel Production Model

G. Fant, Acoustic theory of speech production. Walter de Gruyter, no. 2, 1970.

Source - Filter Analysis



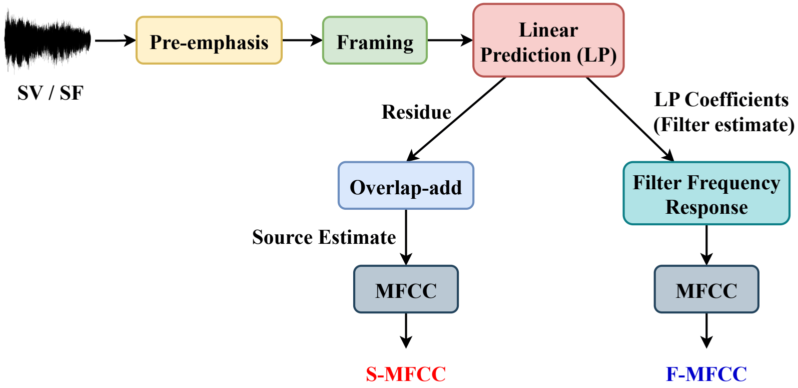


Voiceless Fricative Production Model

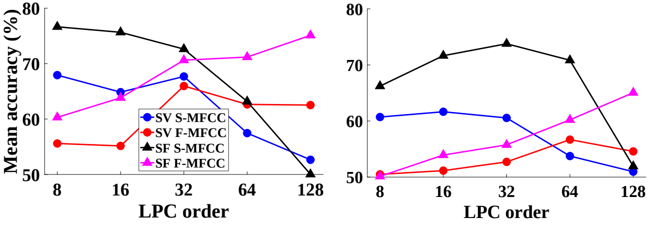
G. Fant, Acoustic theory of speech production. Walter de Gruyter, no. 2, 1970.

Source - Filter Analysis

Source - Filter Estimation Method



Source - Filter Analysis



(a) ALS vs. HC (b) PD vs. HC

Figure: Mean classification accuracies (in %) over all SVs and those over all SFs obtained using S-MFCC and F-MFCC estimated with varying LPC orders

- S-MFCC and F-MFCC of SFs outperform those of SVs at most LPC orders.
- At lower LPC orders, S-MFCC outperforms F-MFCC, while F-MFCC achieves better performance at higher LPC orders.
- At high LPC orders, more detailed structures are captured in the filter estimate and the source estimate becomes nearly white.

Fusion

Table: Mean classification accuracies in % (SD in bracket) obtained using intra- and inter-phoneme decision-level fusion

	Fusion scheme	ALS vs. HC	PD vs. HC
Intra	/i/+/i/+/i/	81.83 (13.35)	**80.03 (11.96)**
	/s/+/s/+/s/	80.04 (8.58)	70.05 (13.19)
	/sh/+/sh/+/sh/	79.95 (8.90)	66.15 (11.36)
Inter	/i/+/s/+/sh/ (Distinct model)	82.02 (8.31)	75.67 (7.58)
	/i/+/s/+/sh/ (Pooled model)	**83.35 (5.93)**	72.65 (9.63)
No Fusion	/i/	78.42 (10.03)	72.85 (12.04)
	/s/	76.90 (7.86)	65.37 (7.84)
	/sh/	77.47 (7.56)	66.66 (9.40)

- Intra-phoneme fusion outperforms the single utterances in most cases.
  - Nature of cues vary in different utterances of a single phoneme.

Fusion

Table: Mean classification accuracies in % (SD in bracket) obtained using intra- and inter-phoneme decision-level fusion

	Fusion scheme	ALS vs. HC	PD vs. HC
Intra	/i/+/i/+/i/	81.83 (13.35)	**80.03 (11.96)**
	/s/+/s/+/s/	80.04 (8.58)	70.05 (13.19)
	/sh/+/sh/+/sh/	79.95 (8.90)	66.15 (11.36)
Inter	/i/+/s/+/sh/ (Distinct model)	82.02 (8.31)	75.67 (7.58)
	/i/+/s/+/sh/ (Pooled model)	**83.35 (5.93)**	72.65 (9.63)
No Fusion	/i/	78.42 (10.03)	72.85 (12.04)
	/s/	76.90 (7.86)	65.37 (7.84)
	/sh/	77.47 (7.56)	66.66 (9.40)

- Inter-phoneme fusion using the pooled model achieves the highest mean ALS vs. HC classification accuracy.
  - Cues present in different phoneme utterances are complementary in nature.

Fusion

Table: Mean classification accuracies in % (SD in bracket) obtained using intra- and inter-phoneme decision-level fusion

	Fusion scheme	ALS vs. HC	PD vs. HC
Intra	**/i/+/i/+/i/**	81.83 (13.35)	**80.03 (11.96)**
	/s/+/s/+/s/	80.04 (8.58)	70.05 (13.19)
	/sh/+/sh/+/sh/	79.95 (8.90)	66.15 (11.36)
Inter	**/i/+/s/+/sh/ (Distinct model)**	82.02 (8.31)	75.67 (7.58)
	/i/+/s/+/sh/ (Pooled model)	**83.35 (5.93)**	72.65 (9.63)
No Fusion	**/i/**	78.42 (10.03)	72.85 (12.04)
	/s/	76.90 (7.86)	65.37 (7.84)
	/sh/	77.47 (7.56)	66.66 (9.40)

- However, inter-phoneme fusion could not outperform intra-phoneme fusion of /i/ for PD vs. HC classification.

Overview

Key Takeaways

- Phonemes involving constrictions in the vocal tract (fricatives) or even close placement of tongue and palate (/i/) are found to be better differentiators than the relatively more open ones.
- Different phonemes are observed to capture complementary cues making inter-phoneme fusion the best choice for ALS vs. HC classification.
- However, the same is not empirically true for PD vs. HC case.

Future Work

- To derive some quantifying measures of proximity of pairs of articulators from the speech signals
- To use those measures directly for performing ALS/PD vs. HC classifications

THANK YOU

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