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[2], Ravi Yadav[2], Prasanta Kumar Ghosh[1]

[1]SPIRE LAB, Electrical Engineering Department, Indian Institute of Science, Bengaluru, India

[2]National Institute of Mental Health and Neurosciences, Bengaluru, India

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.

Sustained Phoneme Production (SPP) Task

- **SPP tasks** are commonly used in **clinical assessment of dysarthria**.
  - Simple and easy to administer
  - Can assess all the required sub-systems of speech
  - Different types of phonemes are examined, e.g., **vowels** and **fricatives**
- Can be a **potential task** for speech-based **automatic diagnosis** of ALS & PD.

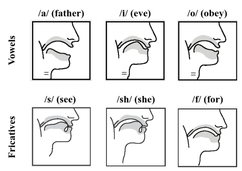
Our Objective

To analyze the relative utility of sustained fricatives (SFs), w.r.t. sustained vowels (SVs), in SPP task based classification of ALS/PD vs. Healthy (HC)

Motivation

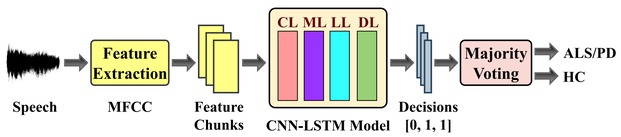
Physiological mechanisms of uttering vowels and fricatives are different.

- **Vowels:** Open vocal tract acting as a resonance chamber
- **Fricatives:** Narrow constrictions in the vocal tract leading to frication



The impacts of dysarthria on vowels and fricatives may differ significantly.

Classification Approach



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

Experimental Details

- **Dataset**
  - **Place of data collection:** NIMHANS, Bengaluru, India
  - **Subjects:** 35 (25M + 10F) from each of ALS, PD and HC groups (Every subject gave an informed consent.)
  - **Speech task:** Sustained utterances of vowels /a/, /i/, /o/ and fricatives /s/, /sh/, /f/
  - **Total #utterances:** 526 (ALS), 528 (PD), 507 (HC)
  - **Mean (SD) of utterance length (sec):** 3.30 (2.36) (ALS), 4.09 (2.53) (PD), 5.06 (2.04) (HC)
  - **Recording device:** Zoom H6 with XYH-6 capsule
  - **Sampling frequency:** 44.1 kHz (downsampled to 16 kHz)
- **Validation Protocol:** 5-fold cross-validation at subject level

Performance of Fricatives vs. Vowels

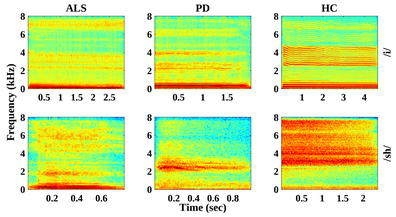
Mean classification accuracies in % (SD in bracket) obtained using different sustained phonemes

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

Some Spectral Differences

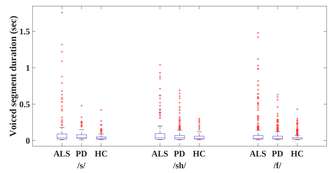
- For vowels (e.g., /i/),
  - **Less evident harmonic structure** of fundamental frequency in dysarthric utterances than HC utterances
  - **Lower formant energies** in dysarthric utterances than healthy ones
- For voiceless fricatives (e.g., /sh/),
  - **Less prominent high frequency content** in dysarthric spectrograms than HC ones
  - **Unwanted voicing** in dysarthric utterances



Illustrative narrowband spectrograms

Unwanted Voicing of Voiceless Fricatives

- Plot shows the distributions of the durations of voiced segments in fricative utterances.
- **Dysarthric fricatives have longer voiced segments** (at 1% significance level as per Wilcoxon ranksum test) than those produced by HCs.



Effect of Fusion

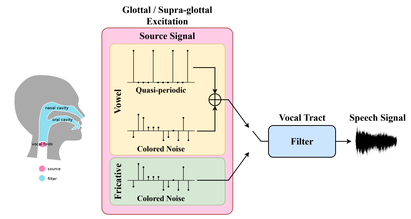
Mean (SD) of classification accuracies in % obtained using intra- and inter- phoneme decision-level fusion

	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)

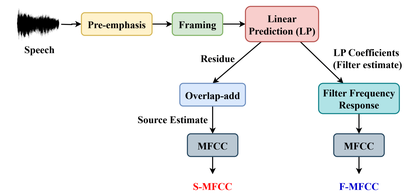
- Intra-phoneme fusion outperforms the single utterances in most cases.
  - Nature of cues vary in different utterances of a single phoneme.
- Inter-phoneme fusion using the pooled model achieves the highest mean ALS vs. HC classification accuracy.
  - Cues present in different phoneme utterances are complementary in nature.
- However, inter-phoneme fusion could not outperform intra-phoneme fusion of /i/ for PD vs. HC classification.

Source - Filter Analysis

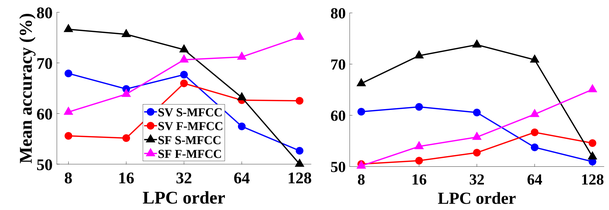
Source - Filter Model



Estimation Method



Relative Performance



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

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.

Conclusion

- **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

References

1. Jhansi Mallela, Aravind Illa, BN Suhas, Sathvik Udupa, Yamini Belur, Nalini Atchayaram, Ravi Yadav, Pradeep Reddy, Dipanjan Gope, andPrasanta Kumar Ghosh, "Voice based classification of patients with Amyotrophic Lateral Sclerosis, Parkinson's disease and healthy controls with CNN-LSTM using transfer learning," in International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 6784–6788.
2. Gunnar Fant, Acoustic theory of speech production, Walter de Gruyter, 1970.

Acknowledgement: We thank Navaneetha G and Agniv Chatterjee for their valuable assistance in data preparation. We also thank the Department of Science and Technology (DST), Govt. of India for supporting this work.

tanukab@iisc.ac.in, prasantg@iisc.ac.in