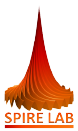
Classification of multi-class vowels and fricatives from patients having Amyotrophic Lateral Sclerosis with varied levels of dysarthria severity

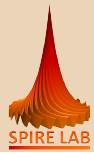
Chowdam Venkata Thirumala Kumar[1], **Tanuka Bhattacharjee**[1], Yamini Belur[2], Atchayaram Nalini[3], Ravi Yadav[3], Prasanta Kumar Ghosh[1]

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



INTERSPEECH 2023

Overview



Discriminability of Sounds

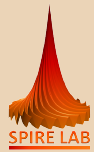
- **Acoustic characteristics** of individual speech sounds help in **discriminating** them from one another.
- Level of discrimination can change due to
  - Background noise
  - Reverberation
  - Cross-talk
  - Impaired speech production, etc.

Discriminability of Sounds

- **Acoustic characteristics** of individual speech sounds help in **discriminating** them from one another.
- Level of discrimination can change due to
  - Background noise
  - Reverberation
  - Cross-talk
  - **Impaired speech production -**

    Dysarthria due to Amyotrophic Lateral Sclerosis (ALS)

Amyotrophic Lateral Sclerosis (ALS)



- **Incurable** and **progressive neuro-degenerative** disease affecting **muscle movements**[1]
- Speech musculature get severely affected leading to **Dysarthria**
  - Affects **articulation**, **phonation**, **prosody**, **respiration** and **resonance**[2]
  - Can adversely impact the discriminability of different sounds
  - Form and extent of different impairments vary with the degree of **severity**

1. https://www.als.org/understanding-als/what-is-als/

2. 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.

Our Objective

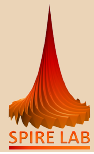
- **To analyze the discriminability of different vowel and fricative sounds with increasing severity of ALS-induced dysarthria**
- **Tasks** - Classification of
  1. 4 sustained vowels - /a/, /i/, /o/ and /u/
  2. 3 sustained fricatives - /s/, /sh/ and /f/
- Both tasks are to be done **at different dysarthria severity levels**
- **Automatic classification** performances are to be compared against **manual classification** accuracies

Literature

Vowel	**Analysis**	1. Vowel height dimension is frequently misidentified due to limited tongue height control. 2. Vowel contrasts reduce in severe patients.
	Manual Classification	1. /u/ has less vowel intelligibility than /a/, /i/, and /o/ in control group but not in severe dysarthric group; /i/ is observed to have declined intelligibility with an increase in dysarthria severity.**
	Automatic Classification	-
Fricative	**Analysis**	1. Articulatory differences are observed between fricatives produced by speakers with ALS and healthy controls. 2. Place of articulation gets affected for lingual fricatives in men with ALS. 3. Unwanted voicing is observed in the voiceless fricative /s/.
	Manual Classification	-
	Automatic Classification	-

** J. Lee et al., "Vowel-specific intelligibility and acoustic patterns in individuals with dysarthria secondary to Amyotrophic Lateral Sclerosis," Journal of Speech, Language, and Hearing Research, vol. 62, no. 1, pp. 34–59, 2019.

Overview



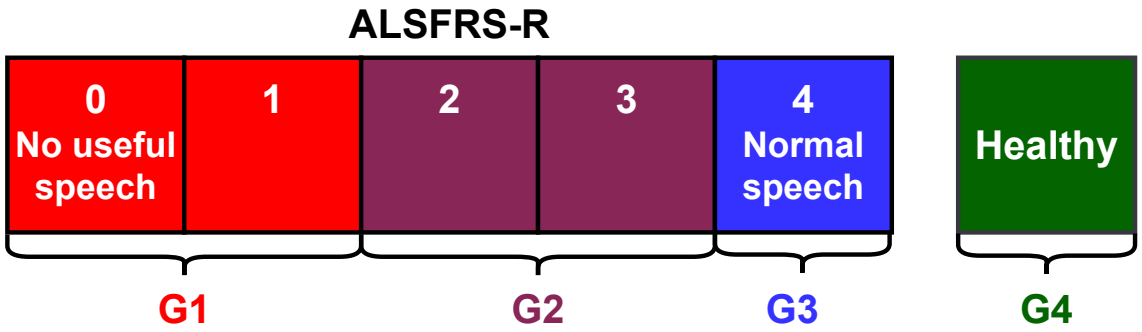
Data Collection Details

- **Place of data collection:**
  - National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India
- **Speech task:**
  - Sustained utterances of /a/, /i/, /o/, /u/, /s/, /sh/ and /f/
  - 1-3 utterances per phoneme per subject
- **Dysarthria severity rating:**
  - As per the 5-point speech component of ALSFRS-R scale
  - Mode of the ratings given by 3 Speech-Language Pathologists

ALSFRS-R

0 No useful speech	1	2	3	4 Normal speech

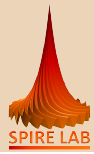
Subject Grouping



Group	Description	ALSFRS-R	#Subjects	#Male:#Female	Age range (years)
G1	Severe	0,1	39	22:17	23-81
G2	Mild	2,3	40	26:14	
G3	ALS without dysarthria	4	40	25:15	
G4	Healthy	-	40	20:20	22-55

- Subjects had **five** different **native languages** - Bengali, Hindi, Tamil, Telugu and Kannada.

Overview



Automatic Classification

DNN-based classification

Table: Features

Details	MFCC	Self-supervised speech representations						
		Wav2vec	wav2vec 2.0	Hubert	Hubert large	Tera	NPC	Decoar 2.0
Stride (ms)	10	10	20	20	20	10	10	10
Dimension	36	512	768	768	1024	768	512	768

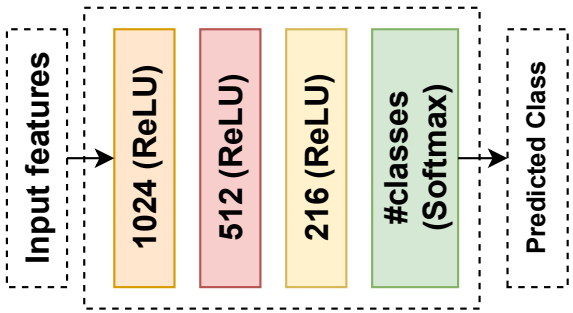


Figure: DNN-based classifier architecture

CNN-based classifier with mel-spectrograms as input[1]

1. C. K. Dewa, "Javanese vowels sound classification with convolutional neural network," in International Seminar on Intelligent Technology and Its Applications (ISITIA), 2016, pp. 123–128.

Manual Classification

- Listening tests performed through a Web app
- **Test set** -
  - **320 vowel** utterances
    (4 severity groups × 4 vowels × 20 subjects)
  - **240 fricative** utterances
    (4 severity groups × 3 fricatives × 20 subjects)
- **Classification strategy** -
  - 3 human listeners classified each utterance along with a confidence score in the range of [0, 100].
  - **Final decision** -
    (1) Mode of the 3 decisions, or
    (2) Decision with high confidence score if all the 3 decisions are different.

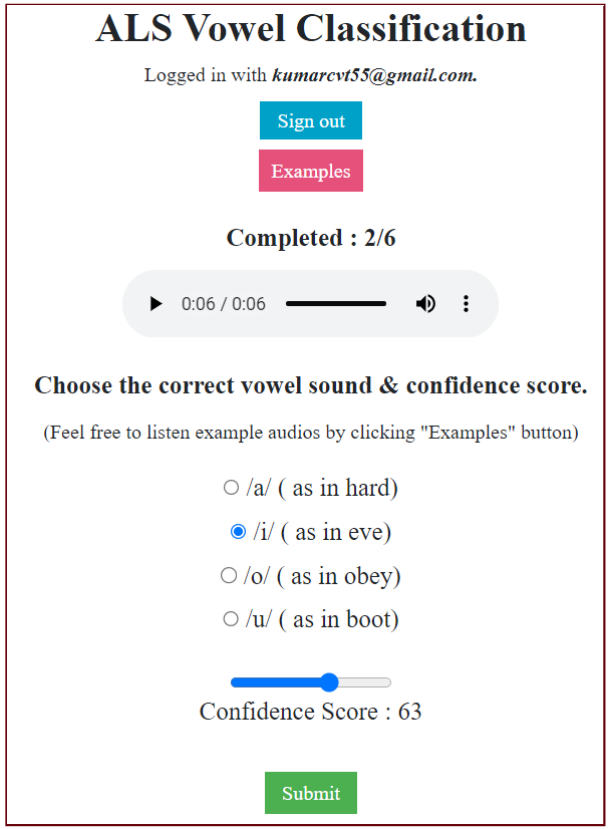


Figure: Illustrative screenshot of the interface

Manual Classification

- Each listener was presented with
  - 41-52 vowel and 31-40 fricative utterances
  - Almost equal number of utterances (9-12) from each severity group
  - Random 8 repeated utterances for both vowels and fricatives
  - Refreshing music after every 5 utterances
  - Access to example healthy utterances at the beginning and throughout the assessment

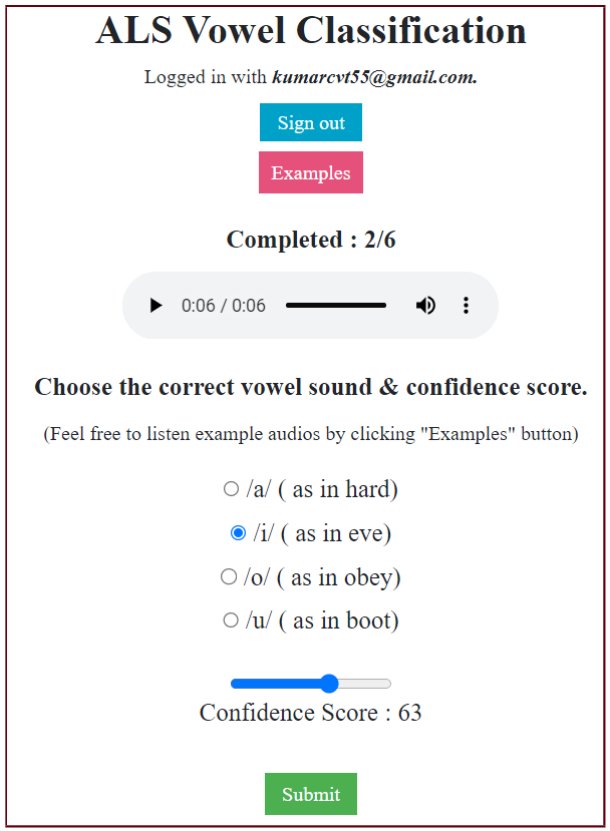


Figure: Illustrative screenshot of the interface

Manual Classification

- **Listener selection criteria**

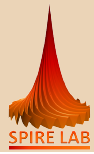
	Accuracy on Healthy utterances	Consistency on repeated utterances
Vowels	$\geq 75\%$	$\geq 75\%$
Fricatives	$\geq 60\%$	$\geq 60\%$

- 20 out of the 44 participating listeners were selected.
- Listeners had native languages spanning over Bengali, Hindi, Kannada, Malayalam, Tamil, and Telugu.

Evaluation Protocol

- **Phase - 1**
  - Evaluation of automatic classifiers through 5-fold cross-validation separately for each severity group for both vowels and fricatives
- **Phase - 2**
  - Evaluation of automatic classifiers on the same dataset as used for the manual listening tests, and comparison with manual classification performance

Overview



Phase - 1: Vowels

Table: Mean (SD) of vowel classification accuracies in % over 5-fold cross-validation obtained using different automatic classification methods

	DNN								
	MFCC	**wav2vec**	**wav2vec 2.0**	**Hubert**	**Hubert large**	**Tera**	**NPC**	**Decoar 2.0**	**CNN**
G1	33.82 (5.24)	50.17 (1.79)	49.06 (7.18)	46.34 (12.36)	43.85 (6.58)	34.57 (3.05)	24.8 (4.76)	28.34 (4.58)	**55.43 (14.34)**
G2	38.67 (4.54)	64.97 (6.38)	64.59 (7.41)	59.56 (15.19)	64.2 (6.25)	43.58 (4.71)	25.44 (2.67)	29.95 (4.68)	**66.21 (8.01)**
G3	42.58 (6.29)	66.58 (5.75)	69.16 (6.66)	**74.21 (5.69)**	68.54 (6.00)	54.45 (5.09)	29.52 (4.63)	44.31 (3.50)	70.63 (6.58)
G4	41.11 (2.21)	73.44 (5.25)	72.69 (7.11)	**78.39 (7.87)**	73.43 (5.05)	52.19 (6.87)	29.87 (4.63)	42.58 (4.69)	77.34 (4.20)

G1:Severe — **G4**:Healthy

- Classification performance declines with an increase in severity for most speech representations.

Phase - 1: Fricatives

Table: Mean (SD) of fricative classification accuracies in % over 5-fold cross-validation obtained using different automatic classification methods

	DNN								
	MFCC	**wav2vec**	**wav2vec 2.0**	**Hubert**	**Hubert large**	**Tera**	**NPC**	**Decoar 2.0**	**CNN**
G1	31.75 (6.57)	41.88 (7.57)	34.04 (13.24)	40.53 (5.44)	**44.38 (11.07)**	37.64 (6.10)	30.11 (7.32)	33.05 (2.75)	32.36 (6.01)
G2	36.83 (7.52)	57.81 (11.19)	54.61 (8.21)	46.84 (11.38)	43.26 (8.86)	44.93 (5.04)	33.46 (33.46)	37.87 (5.90)	**63.93 (6.31)**
G3	40.49 (5.71)	67.37 (7.75)	54.49 (4.20)	58.27 (6.17)	48.85 (10.27)	45.19 (6.72)	35.76 (2.95)	48.09 (7.12)	**70.36 (6.65)**
G4	43.86 (6.36)	73.01 (4.65)	55.2 (5.55)	65.33 (13.42)	58.11 (12.42)	41.33 (7.90)	35.4 (5.29)	48.17 (6.26)	**73.6 (8.12)**

G1:Severe — **G4**:Healthy

Phase - 2: Automatic vs Manual

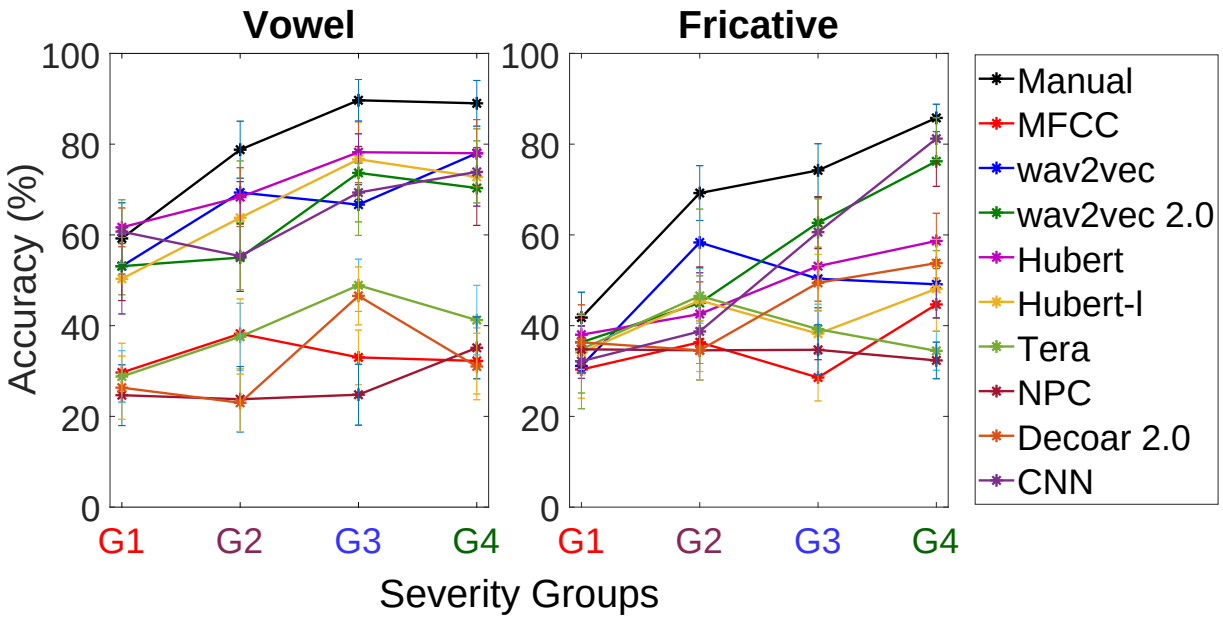


Figure: Mean automatic and manual classification accuracies (SD in error bar) for vowels and fricatives for different severity groups evaluated on the manual test set (**G1**:Severe — **G4**:Healthy)

- Manual classification is better than the automatic ones for all severity groups except the highest severity case for vowels.
- Differences between the automatic and manual classification performances gradually reduce with an increase in severity.

Phase - 2: Confusion Matrices

G1 (Severe)

	/a/	/i/	/o/	/u/
/a/	20,14	0,0	0,3	0,3
/i/	7,1	9,10	0,1	2,6
/o/	9,2	0,0	8,14	3,4
/u/	4,3	1,1	5,5	9,10

	/s/	/sh/	/f/
/s/	16,5	2,8	1,6
/sh/	8,9	4,7	8,4
/f/	12,2	3,6	4,11

G2

	/a/	/i/	/o/	/u/
/a/	20,17	0,0	0,2	0,1
/i/	3,0	16,16	1,1	0,3
/o/	2,2	0,0	16,14	1,3
/u/	0,0	0,3	9,8	11,9

	/s/	/sh/	/f/
/s/	17,15	2,1	1,3
/sh/	5,7	8,5	7,8
/f/	3,2	1,3	16,15

G3

	/a/	/i/	/o/	/u/
/a/	20,19	0,0	0,1	0,0
/i/	0,0	20,19	0,0	0,1
/o/	1,3	0,0	17,10	2,7
/u/	0,0	0,0	5,5	15,15

	/s/	/sh/	/f/
/s/	15,15	3,1	1,3
/sh/	0,9	15,7	5,4
/f/	0,3	5,1	14,15

G4 (Healthy)

	/a/	/i/	/o/	/u/
/a/	20,19	0,0	0,1	0,0
/i/	0,0	20,18	0,0	0,2
/o/	1,1	0,0	18,12	1,7
/u/	0,2	0,4	6,1	14,13

	/s/	/sh/	/f/
/s/	18,16	0,2	1,2
/sh/	0,5	13,14	7,1
/f/	0,2	0,0	29,17

Figure: Confusion matrices for vowels and fricatives using **manual (in black)** and the **best-performing automatic (in red)** classification

Language-wise Analysis

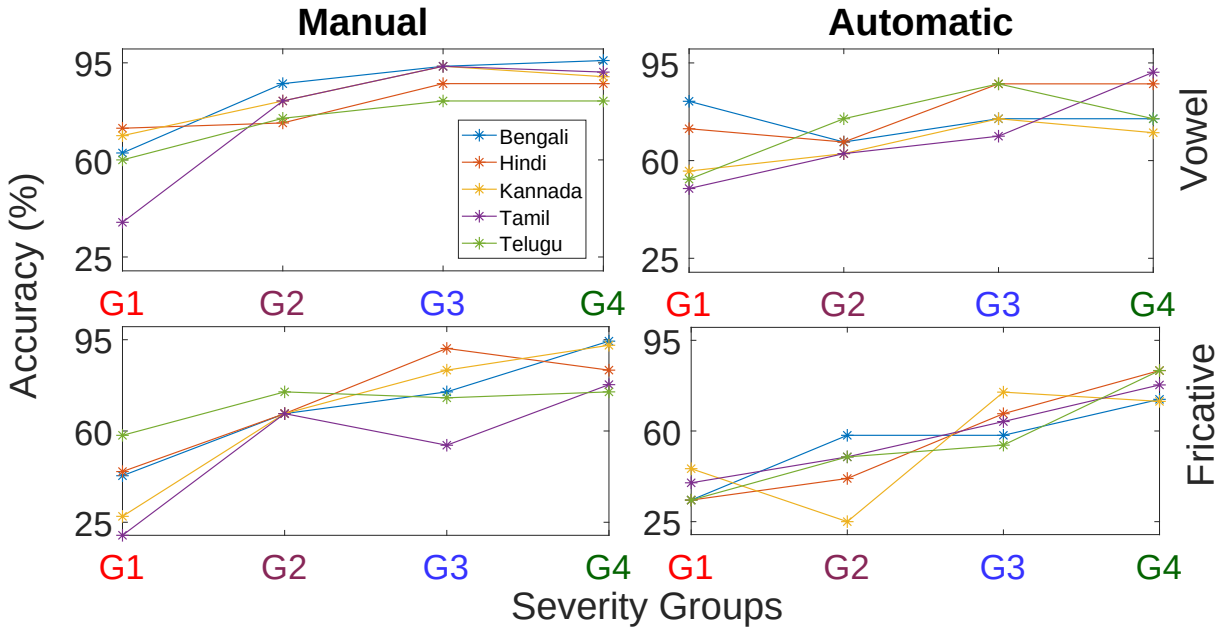


Figure: Language-wise accuracies of automatic and manual classification of vowels and fricatives for different severity groups (**G1**:Severe — **G4**:Healthy)

- No language-specific pattern is observed.

Overview

Key-Takeaways

- Both automatic and manual classification performances decline with an increase in severity.
- Manual classification is always better than automatic classification except for the highest severity case of vowels.
- The performance gap between manual and automatic classification reduces with an increase in severity level.

Future Work

- To analyze the discriminability of different syllables with increasing severity of ALS-induced dysarthria
- To visualize the changes in the overall acoustic space with increasing severity of ALS-induced dysarthria

Acknowledgement

- Authors thank Seena Vengalil and M S Keertipriya for their valuable assistance during data collection.
- Authors also 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