Exploring Syllable Discriminability during Diadochokinetic Task with Increasing Dysarthria Severity for Patients with Amyotrophic Lateral Sclerosis

Neelesh Samptur[1], Tanuka Bhattacharjee[2], Anirudh Chakravarty K[1], Seena Vengalil[3], Yamini Belur[3], Atchayaram Nalini[3], **Prasanta Kumar Ghosh[2]**

[1]CSE Dept., PES University, Bangalore, India
[2]SPIRE LAB, EE Dept., Indian Institute of Science, Bangalore, India
[3]National Institute of Mental Health and Neurosciences, Bangalore, India

INTERSPEECH 2024

Overview

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.

Amyotrophic Lateral Sclerosis (ALS)

- **Clinically**, different **speech tasks** are used for **diagnosis and assessment** of dysarthria prevalent in ALS.
  - Sustained phonation
  - Diadochokinetic task
  - Spontaneous speech
  - Read speech
  - Image description

Amyotrophic Lateral Sclerosis (ALS)

- **Clinically**, different **speech tasks** are used for **diagnosis and assessment** of dysarthria prevalent in ALS.
  - Sustained phonation
  - **Diadochokinetic task**
  - Spontaneous speech
  - Read speech
  - Image description

Diadochokinetic (DDK) Task

- Examines how **quickly** and **accurately** one can **repeat**, without any interruption, a series of **monosyllabic targets like 'pa-pa-pa'** or **tri-syllabic targets like 'pataka'**[1]

1. B. Tomik et al., "Dysarthria in Amyotrophic Lateral Sclerosis: a review," Amyotrophic Lateral Sclerosis, vol. 11, no. 1-2, pp. 4–15, 2010.

Effect of ALS on DDK Tasks

- **Reduced DDK rate**[1]
  - Dysarthria due to ALS restricts the speed of movements of lips, jaw, tongue, and velum.
- **Compromised syllable discriminability**[2]
  - ALS patients often perform incomplete closures while uttering stop consonants.
  - Consonant-to-vowel formant transitions get impaired.

1. B. Yamini et al., "Measures of maximum performance of speech-related tasks in patients with Amyotrophic Lateral Sclerosis," Amyotrophic Lateral Sclerosis, vol. 9, 2008.

2. R. D. Kent et al., "Speech deterioration in Amyotrophic Lateral Sclerosis: A case study," Journal of Speech, Language, and Hearing Research, vol. 34, no. 6, pp. 1269–1275, 1991.

Effect of ALS on DDK Tasks

- **Reduced DDK rate**[1]
  - Dysarthria due to ALS restricts the speed of movements of lips, jaw, tongue, and velum.
- **Compromised syllable discriminability**[2]
  - ALS patients often perform incomplete closures while uttering stop consonants.
  - Consonant-to-vowel formant transitions get impaired.

The **extent** to which the syllable discriminability gets compromised at **different dysarthria severity levels** for ALS is **not well understood**.

1. B. Yamini et al., "Measures of maximum performance of speech-related tasks in patients with Amyotrophic Lateral Sclerosis," Amyotrophic Lateral Sclerosis, vol. 9, 2008.

2. R. D. Kent et al., "Speech deterioration in Amyotrophic Lateral Sclerosis: A case study," Journal of Speech, Language, and Hearing Research, vol. 34, no. 6, pp. 1269–1275, 1991.

Our Objective

- **To analyze the degree of discriminability among /pa/, /ta/, and /ka/ syllables produced during the tri-syllabic DDK 'pataka' task, at varied severity levels of ALS-induced dysarthria**
- **Task** - 3-class classification of /pa/, /ta/ and /ka/ sounds
  - To be done **at different dysarthria severity levels**
  - **Automatic classification** performances are to be compared against **manual classification** accuracies

Literature

- Several **studies** have been reported on **discriminability among certain vowels and fricatives** with increasing dysarthria severity for ALS patients.
  - **Discriminability of different vowels and different fricatives**, as perceived by both human and machine, **decline** drastically **with increasing dysarthria severity**.[1]
  - **Vowel space area reduces** making it difficult to discriminate between vowels.[2]
  - Patients often add **unwanted voicing** to **voiceless fricatives** making them sound like their respective voiced counterparts.[3]

1. C. V. T. Kumar et al., "Classification of multi-class vowels and fricatives from patients having Amyotrophic Lateral Sclerosis with varied levels of dysarthria severity," in Proc. Interspeech, 2023, pp. 146–150.

2. B. Yamini et al., "Vowel space area in patients with Amyotrophic Lateral Sclerosis," Amyotrophic Lateral Sclerosis, vol. 9, no. 1, pp. 118–119, 2008.

3. T. Bhattacharjee et al., "Exploring the role of fricatives in classifying healthy subjects and patients with Amyotrophic Lateral Sclerosis and Parkinson's Disease," in IEEE ICASSP 2023, pp. 1–5.

Literature

- Discriminability among **syllables** like /pa/, /ta/, and /ka/ remains relatively **unexplored**.
  - Tao et al.[1] have analysed **automatic speech recognition on DDK 'pataka' sequences** for patients with traumatic brain injuries and Parkinson's disease, but **not for ALS**.

1. F. Tao et al., "A portable automatic PA-TA-KA syllable detection system to derive biomarkers for neurological disorders," in Proc. Interspeech, 2016, pp. 362–366.

Overview

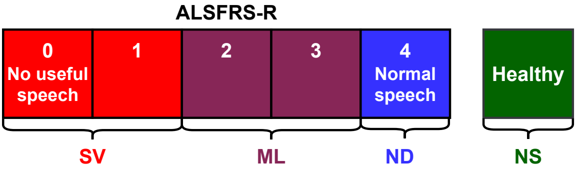
Data Collection Details

- **Place of data collection:**
  - National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India
- **Speech task:**
  - Take a deep breath and keep repeating the tri-syllabic sequence ‘pataka’ as fast as possible.
  - 1-3 such trials 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	1	2	3	4
No useful speech				Normal speech

Subject Grouping



Group	Description	ALSFRS-R	#Subjects	#Male:#Female	Age range (years)
SV	**Severe**	0,1	30	18:12	30-73
ML	**Mild**	2,3	35	24:11	28-70
ND	**ALS without dysarthria**	4	35	22:13	28-70
NS	**Healthy**	-	35	18:17	31-55

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

Data Preprocessing - Syllable Segmentation

Two-phase semi-automatic method

Phase 1 - Automatic

- Obtain the upper peak envelope of the speech waveform.
- Low-pass filter the envelope at 15 Hz cutoff frequency to make it smooth.
- Locate the prominent local minima of the smooth envelope.
- Consider the speech segment between two consecutive minima as one syllable.
- Cyclically label the identified syllables as /pa/, /ta/, and /ka/.

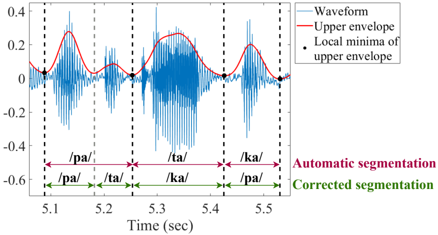


Figure: Automatic and manually corrected syllable segmentations

Phase 2 - Manually correct all erroneous segmentations

Syllable Corpus Statistics

Table: Number of utterances of different syllables

Syllable	SV	ML	ND	NS
/pa/	294	730	1417	1349
/ta/	287	621	1323	1215
/ka/	267	716	1405	1346

Table: Mean (SD) of durations (in sec) of utterances of different syllables

Syllable	SV	ML	ND	NS
/pa/	0.33 (0.16)	0.17 (0.06)	0.13 (0.03)	0.11 (0.03)
/ta/	0.32 (0.17)	0.18 (0.08)	0.12 (0.03)	0.11 (0.03)
/ka/	0.38 (0.15)	0.25 (0.12)	0.17 (0.05)	0.15 (0.03)

Overview

Manual Classification

- **Listening tests** performed through a **Web app**
- **Test set** -
  - **360 syllable** utterances
    (4 severity groups × 15 subjects × 3 syllables × 2 utterances)
- **Classification strategy** -
  - 3 human listeners classified each utterance as one of /pa/, /ta/, or /ka/, along with a confidence score in the range of [0, 100].
  - **A response is considered as correct if** it matches with the ground truth and the corresponding confidence score is above 40.

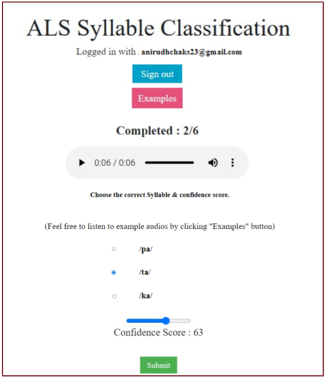


Figure: Illustrative screenshot of the interface

Manual Classification

- Each listener was presented with
  - 90 unique syllable utterances
  - Almost equal number of utterances from each severity group
  - Random 10 repeated utterances (2 or 3 per severity group)
  - Refreshing music after every 5 utterances
  - Access to example healthy utterances at the beginning and throughout the assessment
  - Option to play a test audio as many times as needed before submitting the response

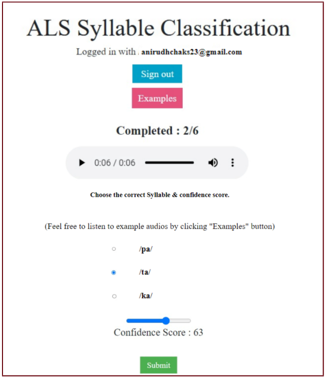


Figure: Illustrative screenshot of the interface

Manual Classification

- **27 listeners** participated.
  - 19 Male + 8 Female
  - Age range: 17-54 years
  - Native languages: Hindi, Kannada, Malayalam, Tamil, Tulu, Urdu and Telugu
- **Listener selection criteria**
  1. **Accuracy on healthy utterances** $\geq 80\%$
  2. **Consistency on repeated utterances** $\geq 80\%$
- 12 out of the 27 participating listeners were selected.

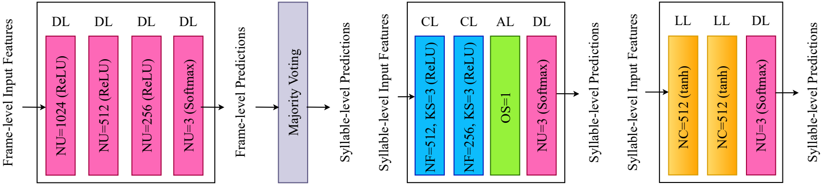
Automatic Classification

- **Features**

	MFCC	Self-supervised speech representations				
		wav2vec 2.0	**Hubert**	**Tera**	**NPC**	**Decoar**
Stride (ms)	5	20	20	10	10	10
Dimension	36	768	768	768	512	2048

Automatic Classification

Classifiers



(a) **DNN** (b) **CNN** (c) **LSTM**

Figure: Architectures of different classifiers

Here, DL: Dense layer, CL: 1D-Convolutional layer, AL: Adaptive average pooling layer, LL: LSTM layer, NU: Number of units in DL, NF: Number of filters, KS: Kernel size, OS: Output size of AL, NC: Number of LSTM cells

Evaluation Protocol

SPIRE LAB

Phase - 1:

Evaluation of automatic classifiers through 5-fold cross-validation separately for each severity group

- Subjects of each severity group are equally and randomly distributed among the 5 folds.
- For any severity group, at every iteration,
  - One of the folds acts as the test set.
  - Remaining folds are used together as the train set.
  - Random 4 subjects from the train set are chosen to form the validation set.
- Same fold structure is maintained across all automatic classifiers.

Evaluation Protocol

SPIRE LAB

Phase - 2:

Evaluation of automatic classifiers on the same dataset as used for the manual listening tests, and comparison with manual classification performance

- For every severity group,
  - Utterances selected for the manual listening task are used as the test set for the automatic classifiers.
  - Remaining subjects are used to form the train set.
  - Random 3 subjects from the train set are used for validation.
- Automatic and manual classification accuracies are compared using the **Wilcoxon signed-rank test** at 1% significance level.
  - 30 random subsets, each containing 30 utterances, are formed from the test set for every severity group.

Overview

Phase - 1: Automatic Classification

Table: Mean classification accuracies in % (SD in bracket) over 5-fold cross-validation obtained using different automatic classification methods for different severity groups

Feature	DNN				CNN				LSTM			
	NS	ND	ML	SV	NS	ND	ML	SV	NS	ND	ML	SV
MFCC	83.69 (5.67)	74.82 (11.31)	61.12 (11.11)	50.27 (6.77)	59.07 (6.81)	50.92 (10.95)	45.38 (6.26)	42.70 (6.46)	76.89 (13.15)	74.74 (11.01)	63.12 (7.96)	48.68 (7.28)
NPC	45.66 (7.95)	59.97 (20.35)	55.94 (11.69)	48.59 (3.49)	47.79 (5.59)	43.19 (4.74)	38.41 (3.68)	35.19 (5.23)	60.47 (17.97)	60.67 (8.17)	58.80 (6.21)	45.83 (6.39)
TERA	83.91 (5.23)	90.10 (5.20)	67.80 (9.89)	61.53 (8.92)	57.57 (5.98)	54.43 (7.89)	55.64 (5.59)	40.65 (3.07)	83.32 (5.44)	90.82 (3.10)	80.64 (3.25)	47.83 (4.86)
HuB-ERT	**99.03 (0.40)**	**99.46 (0.47)**	**97.03 (1.45)**	79.72 (8.68)	94.06 (6.59)	89.25 (4.25)	78.54 (4.29)	54.12 (5.11)	97.98 (2.37)	98.56 (1.44)	94.52 (3.09)	**79.75 (10.47)**
Wav2-Vec2.0	98.13 (0.57)	97.29 (1.43)	91.21 (6.12)	71.14 (6.82)	64.49 (14.06)	61.37 (7.25)	46.31 (5.43)	38.02 (1.37)	96.23 (1.54)	94.40 (1.64)	86.40 (4.51)	68.11 (14.22)
DeC-oAR	85.53 (7.84)	93.17 (3.47)	74.05 (5.43)	53.01 (6.04)	50.53 (2.86)	49.16 (3.55)	38.84 (3.48)	36.96 (3.33)	85.99 (10.03)	72.07 (19.47)	61.88 (4.21)	48.74 (5.90)

- Classification accuracies generally decline with increasing severity.

Phase - 1: Automatic Classification

Table: Mean classification accuracies in % (SD in bracket) over 5-fold cross-validation obtained using different automatic classification methods for different severity groups

Feature	DNN				CNN				LSTM			
	NS	ND	ML	SV	NS	ND	ML	SV	NS	ND	ML	SV
MFCC	83.69 (5.67)	74.82 (11.31)	61.12 (11.11)	50.27 (6.77)	59.07 (6.81)	50.92 (10.95)	45.38 (6.26)	42.70 (6.46)	76.89 (13.15)	74.74 (11.01)	63.12 (7.96)	48.68 (7.28)
NPC	45.66 (7.95)	59.97 (20.35)	55.94 (11.69)	48.59 (3.49)	47.79 (5.59)	43.19 (4.74)	38.41 (3.68)	35.19 (5.23)	60.47 (17.97)	60.67 (8.17)	58.80 (6.21)	45.83 (6.39)
TERA	83.91 (5.23)	90.10 (5.20)	67.80 (9.89)	61.53 (8.92)	57.57 (5.98)	54.43 (7.89)	55.64 (5.59)	40.65 (3.07)	83.32 (5.44)	90.82 (3.10)	80.64 (3.25)	47.83 (4.86)
HuB-ERT	**99.03 (0.40)**	**99.46 (0.47)**	**97.03 (1.45)**	79.72 (8.68)	94.06 (6.59)	89.25 (4.25)	78.54 (4.29)	54.12 (5.11)	97.98 (2.37)	98.56 (1.44)	94.52 (3.09)	**79.75 (10.47)**
Wav2-Vec2.0	98.13 (0.57)	97.29 (1.43)	91.21 (6.12)	71.14 (6.82)	64.49 (14.06)	61.37 (7.25)	46.31 (5.43)	38.02 (1.37)	96.23 (1.54)	94.40 (1.64)	86.40 (4.51)	68.11 (14.22)
DeC-oAR	85.53 (7.84)	93.17 (3.47)	74.05 (5.43)	53.01 (6.04)	50.53 (2.86)	49.16 (3.55)	38.84 (3.48)	36.96 (3.33)	85.99 (10.03)	72.07 (19.47)	61.88 (4.21)	48.74 (5.90)

- HuBERT-based features with DNN classifier achieve the best classification performance for NS, ND, and ML groups.

Phase - 1: Automatic Classification

Table: Mean classification accuracies in % (SD in bracket) over 5-fold cross-validation obtained using different automatic classification methods for different severity groups

Feature	DNN				CNN				LSTM			
	NS	ND	ML	SV	NS	ND	ML	SV	NS	ND	ML	SV
MFCC	83.69 (5.67)	74.82 (11.31)	61.12 (11.11)	50.27 (6.77)	59.07 (6.81)	50.92 (10.95)	45.38 (6.26)	42.70 (6.46)	76.89 (13.15)	74.74 (11.01)	63.12 (7.96)	48.68 (7.28)
NPC	45.66 (7.95)	59.97 (20.35)	55.94 (11.69)	48.59 (3.49)	47.79 (5.59)	43.19 (4.74)	38.41 (3.68)	35.19 (5.23)	60.47 (17.97)	60.67 (8.17)	58.80 (6.21)	45.83 (6.39)
TERA	83.91 (5.23)	90.10 (5.20)	67.80 (9.89)	61.53 (8.92)	57.57 (5.98)	54.43 (7.89)	55.64 (5.59)	40.65 (3.07)	83.32 (5.44)	90.82 (3.10)	80.64 (3.25)	47.83 (4.86)
HuB-ERT	**99.03 (0.40)**	**99.46 (0.47)**	**97.03 (1.45)**	79.72 (8.68)	94.06 (6.59)	89.25 (4.25)	78.54 (4.29)	54.12 (5.11)	97.98 (2.37)	98.56 (1.44)	94.52 (3.09)	**79.75 (10.47)**
Wav2-Vec2.0	98.13 (0.57)	97.29 (1.43)	91.21 (6.12)	71.14 (6.82)	64.49 (14.06)	61.37 (7.25)	46.31 (5.43)	38.02 (1.37)	96.23 (1.54)	94.40 (1.64)	86.40 (4.51)	68.11 (14.22)
DeC-oAR	85.53 (7.84)	93.17 (3.47)	74.05 (5.43)	53.01 (6.04)	50.53 (2.86)	49.16 (3.55)	38.84 (3.48)	36.96 (3.33)	85.99 (10.03)	72.07 (19.47)	61.88 (4.21)	48.74 (5.90)

- For SV group, HuBERT-based features with LSTM classifier perform the best.

Phase - 2: Automatic vs Manual

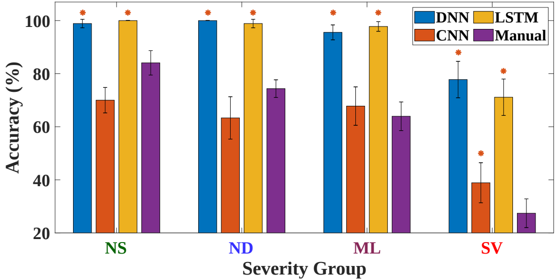


Figure: Mean automatic (using HuBERT features) and manual classification accuracies (SD in error bar) obtained on the manual listening test set for different severity groups; at each severity, * indicates the automatic methods which significantly outperform humans at 1% significance level

- DNN and LSTM significantly outperform the manual classification approach at all severity levels.

Phase - 2: Automatic vs Manual

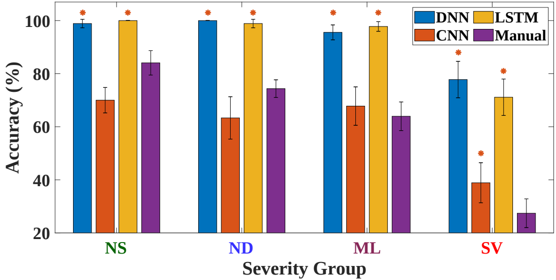


Figure: Mean automatic (using HuBERT features) and manual classification accuracies (SD in error bar) obtained on the manual listening test set for different severity groups; at each severity, * indicates the automatic methods which significantly outperform humans at 1% significance level

- The performance of CNN is significantly better than the manual classification performance only for the SV group.

Phase - 2: Automatic vs Manual

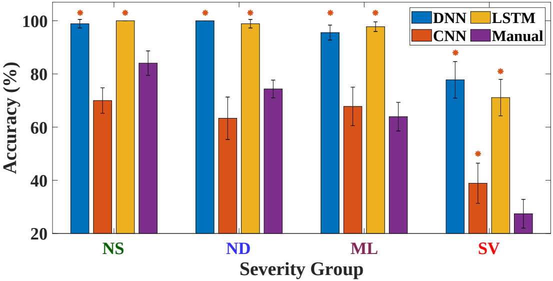


Figure: Mean automatic (using HuBERT features) and manual classification accuracies (SD in error bar) obtained on the manual listening test set for different severity groups; at each severity, * indicates the automatic methods which significantly outperform humans at 1% significance level

- The drops in the mean accuracies from NS to SV are significantly less for automatic DNN (21.11%) and LSTM (28.89%) methods than that for manual classification (56.66%).

Phase - 2: Automatic vs Manual

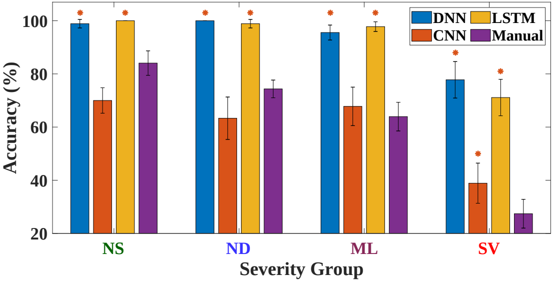


Figure: Mean automatic (using HuBERT features) and manual classification accuracies (SD in error bar) obtained on the manual listening test set for different severity groups; at each severity, * indicates the automatic methods which significantly outperform humans at 1% significance level

- Though humans may fail to perceive the differences among these syllables with increasing dysarthria severity, distinct cues persist in the syllables which data-driven models can capture.

Phase - 2: Confusion Matrices

	/pa/	/ta/	/ka/	CS < 40
/pa/	0.91 1	0.01 0	0 0	0.08 -
/ta/	0.02 0	0.84 1	0.06 0	0.08 -
/ka/	0.12 0	0.06 0	0.77 1	0.06 -

(a) **NS**

	/pa/	/ta/	/ka/	CS < 40
/pa/	0.83 1	0.02 0	0.01 0	0.13 -
/ta/	0.1 0	0.73 1	0.04 0	0.12 -
/ka/	0.11 0	0.02 0	0.66 1	0.20 -

(b) **ND**

	/pa/	/ta/	/ka/	CS < 40
/pa/	0.8 0.97	0.03 0.03	0 0	0.17 -
/ta/	0.11 0	0.52 0.97	0.10 0.03	0.26 -
/ka/	0.08 0	0.14 0	0.6 1	0.18 -

(c) **ML**

	/pa/	/ta/	/ka/	CS < 40
/pa/	0.36 0.93	0.01 0.03	0.11 0.03	0.52 -
/ta/	0.08 0.13	0.15 0.73	0.16 0.13	0.62 -
/ka/	0.18 0.2	0.08 0.13	0.32 0.67	0.42 -

(d) **SV**

Figure: Confusion matrices obtained on the manual listening test set of different severity groups using manual (in red) and the best-performing automatic (in blue) classification methods; here CS: confidence score

- Humans could identify /pa/ the best at all severity levels.
- They confused /ka/ the most for NS and ND but /ta/ for ML and SV.

Phase - 2: Confusion Matrices

	/pa/	/ta/	/ka/	CS < 40
/pa/	0.91 1	0.01 0	0 0	0.08 -
/ta/	0.02 0	0.84 1	0.06 0	0.08 -
/ka/	0.12 0	0.06 0	0.77 1	0.06 -

(e) **NS**

	/pa/	/ta/	/ka/	CS < 40
/pa/	0.83 1	0.02 0	0.01 0	0.13 -
/ta/	0.1 0	0.73 1	0.04 0	0.12 -
/ka/	0.11 0	0.02 0	0.66 1	0.20 -

(f) **ND**

	/pa/	/ta/	/ka/	CS < 40
/pa/	0.8 0.97	0.03 0.03	0 0	0.17 -
/ta/	0.11 0	0.52 0.97	0.10 0.03	0.26 -
/ka/	0.08 0	0.14 0	0.6 1	0.18 -

(g) **ML**

	/pa/	/ta/	/ka/	CS < 40
/pa/	0.36 0.93	0.01 0.03	0.11 0.03	0.52 -
/ta/	0.08 0.13	0.15 0.73	0.16 0.13	0.62 -
/ka/	0.18 0.2	0.08 0.13	0.32 0.67	0.42 -

(h) **SV**

Figure: Confusion matrices obtained on the manual listening test set of different severity groups using manual (in red) and the best-performing automatic (in blue) classification methods; here CS: confidence score

- The best performing automatic method faces the highest confusion in the case of /ka/, followed by /ta/, for SV.

Overview

Key-Takeaways

- **Automatic methods** are found to **outperform humans** in classifying /pa/, /ta/, and /ka/ syllables **at all severity levels** of ALS-induced dysarthria.
- The **drop** in the mean classification **accuracy** from NS to SV is significantly **less for automatic methods** than that for humans.
- Discriminative **acoustic cues** seem to **persist** among the syllables, which automatic methods capture.
- Thus, these **syllables** can be explored further as potential choices of **voice commands** for **automatic voice assistants**, even for the most severe patients.

Future Work

- To analyze the **discriminability** of different **voiced stops** like /b/, /d/, /g/ 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 the Department of Science and Technology (DST), Govt of India, for their support in this work.

THANK YOU

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