Comparison of speech tasks for automatic classification of patients with amyotrophic lateral sclerosis and healthy subjects

Aravind Illa[*], *Deep Patel*[*], *BK Yamini*[†], *Meera SS*[†], *Shivashankar N*[†], *Preethish-Kumar Veeramani*[‡], *Seena Vengalil*[‡], *Kiran Polavarapu*[‡], *Saraswati Nashi*[‡], *Atchayaram Nalini*[‡], *Prasanta Kumar Ghosh*[*]

[*]Department of Electrical Engineering, Indian Institute of Science, Bangalore, India-560 012,

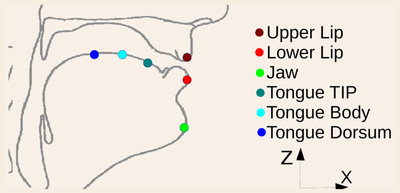
[†]Department of Speech Pathology and Audiology, [‡]Department of Neurology, National Institute of Mental Health and Neurosciences, Bangalore, India-560 029

Introduction

- **ALS**: Rapid and progressive **neuro degenerative disease** that mainly involves the degeneration of motor neurons responsible for controlling voluntary muscle movements like chewing, walking, breathing and talking [1,2].
- ALS also affects **the speech motor functions** of patients, thus causing dysarthria.
- Patients suffering from ALS have an average survival of 2 to 4 years.
- Annual incidence of ALS: Worldwide-1.9/100,000, India-1/100,000.
- **Motivation**: To develop an health care application to detect ALS at an early stage.
- **Objectives of the work**:
  - To automatically classify patients with ALS and healthy subjects using speech tasks, viz. rehearsed speech, spontaneous speech, and repeated words.
  - To experimentally examine which speech task would be more suitable for the classification between ALS patients and healthy subjects.

Data Collection

- Articulatory movement data recorder: → **EMA AG501**.
- **Six** sensors are connected: UL-upper lip, LL-lower lip, Jaw-jaw, TT-tongue tip, TB-tongue body, TD-tongue dorsum.



- **Speech Tasks**: 'Rehearsed speech' (Task #1), 'Spontaneous speech' (Task #2) and 'Repetition of words' (Task #3)
  - **Task #1**:"*My name is X. I am now in Bengaluru.*" (X is the subject's name). Repetition of stimuli upto 6 times.
  - **Task #2**: Participants in this case would produce a monologue to elicit a natural speech output. Repetition of monologue - 2 times.
  - **Task #3**: A set of nine Kannada words viz. **Topi** (*Hat*), **Karnataka**, **Pustaka** (*Book*), **Pen**, **Alilu** (*Squirrel*), **Ili** (*Rat*), **Ungura** (*Ring*), **Chappali** (*Slipper*), **Kitaki** (*Window*).
- Details of the patients and the healthy subjects used in this work

Subject ID	C01	C02	C03	C04	C05	C06	C07	C08
Gender	M	M	M	M	F	F	F	F
Age	35	70	48	70	45	47	60	47
Subject ID	P01	P02	P03	P04	P05	P06	P07	P08
Gender	M	M	M	M	M	F	F	F
Age	38	75	54	49	65	58	54	57
ALSFRS-R	3	3	3	3	2	3	2	3

- Range of durations of different speech tasks

Task #	**Speech Task**	**Range (C)**	**Range (P)**
1	Rehearsed Speech	6.2-19.9 sec	9.8-69 sec
2	Spontaneous Speech	14.7-64.83 sec	14-82.8 sec
3	Repeated Words	0.5-0.7 sec	0.9-1.4 sec

Articulatory and Acoustic Features

- From the **six** sensors, we obtain **12**-dimensional articulatory features namely, UL_x, UL_z, LL_x, LL_z, Jaw_x, Jaw_z, TT_x, TT_z, TB_x, TB_z, TD_x, TD_z.
- We propose kinematics features by considering both horizontal (x) and vertical (z) directions as follows:

$$v_i^d(n) = p_i^d(n) - p_i^d(n-1); a_i^d(n) = v_i^d(n) - v_i^d(n-1) \quad (1)$$

$$V_i(n) = \left|\sqrt{(v_i^x(n))^2 + (v_i^z(n))^2}\right|$$

$$A_i(n) = \left|\sqrt{(a_i^x(n))^2 + (a_i^z(n))^2}\right| \quad (2)$$

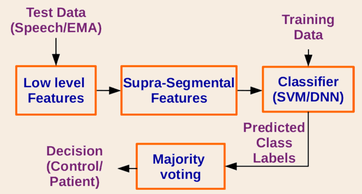
where, $p_i^d(n)$ is the position of the i^{th} sensor at the n^{th} sample, ($d \in \{x, z\}$) and $i \in \{UL, LL, Jaw, TT, TB, TD\}$.

- Articulatory and Acoustic low-level features

Feature (Dimension)	**Description**
VE (6)	*Velocity (eq. 2) of six EMA sensors*
AE (6)	*Acceleration (eq. 2) of six EMA sensors*
VAE (12)	*VE and AE together*
svaE (24)	*static, velocity and acceleration of EMA (eq. 1)*
MFCC (39)	*MFCC from acoustics*
All (87)	*combining VE, AE, VAE, svaE, and MFCC*

Experimental setup

- Four fold cross-validation setup.
- Supra-segmental features are the mean and standard deviations (SD), computed for every 0.8sec (80 frames of low level features) with a shift of 0.2sec.
- Block diagram illustrating the steps in classification



- For DNN we choose 3-hidden layer with 256 units in each layer and output layer with two units and soft-max activation.
- For SVM, radial basis function is chosen as the kernel function.
- Evaluation metric: F-score.

References

1. M. C. Kiernan, S. Vucic, B. C. Cheah, M. R. Turner, A. Eisen, O. Hardiman, J. R. Burrell, and M. C. Zoing, "Amyotrophic lateral sclerosis," in The Lancet, 2011, vol. 377, pp. 942–955.
2. S. E. Langmore and M. Lehman, "The orofacial deficit and dysarthria in ALS," Journal of Speech and Hearing Research, vol. 37, pp. 28–37, 1994.

Results & Discussion

Utterance level classification:

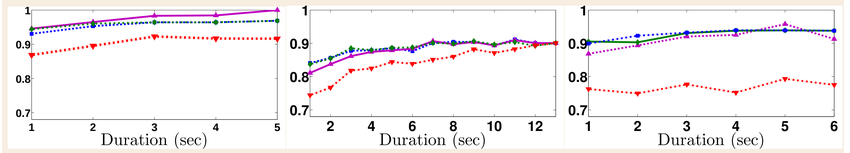
F-score using **SVM** classifier and different features used

Task #	**All**	**svaE**	**MFCC**	**AE**	**VE**	**VAE**
1	.95(.1)	.89(.21)	.89(.21)	.95(.1)	**1(0)**	.95(.1)
2	.88(.16)	.85(.13)	.86(.2)	.89(.15)	**.92(.17)**	**.92(.17)**
3	.88(.16)	.88(.16)	.83(.11)	**.92(.17)**	.88(.16)	**.92(.17)**
Avg	.90 (.14)	.87 (.17)	.86 (.18)	.92 (.14)	**.93 (.11)**	**.93 (.14)**

F-score using **DNN** classifier and different features used

Task #	**All**	**svaE**	**MFCC**	**AE**	**VE**	**VAE**
1	.92(.09)	.89(.08)	.9(.11)	**.95(.1)**	.92(.09)	**.95(.1)**
2	.88(.15)	.68(.28)	.79(.25)	.84(.14)	.73(.12)	**.92(.16)**
3	**.92(.17)**	.88(.16)	.79(.17)	**.92(.17)**	**.92(.17)**	**.92(.17)**
Avg	.90 (.14)	.82 (.17)	.83 (.18)	.90 (.13)	.86 (.13)	**.93 (.14)**

Effect of test recording durations on classification performance:



(a) Rehearsed Speech (b) Spontaneous Speech (c) Repeated words

Figure : F-score for SVM based classification ('•' MFCC (▼), '•' VE (■), '•' AE (▲), '•' VAE (♦)).

Fisher Discriminative Ratio and histogram of best component among VAE:

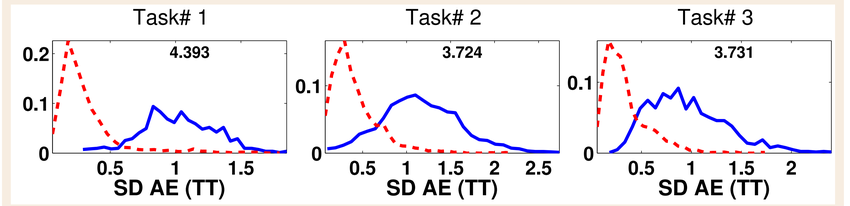


Figure : Histogram of *VAE* with supra-segmental feature computed at 0.8 sec (ALS patients(. . .) & healthy subjects(—))

Conclusion

- Best stimuli : **Rehearsed speech**
- Best feature set: Proposed **kinematic features of articulators**.
- Future work : Experimenting with data from other Indian languages.

Acknowledgement: Authors thank **G. Nisha Meenakshi** and **Ms. Preetie Shetty** for helping out with the recordings, **Babita Behera** for the manual annotations of all recordings, and **all the subjects** who participated for the study. Authors thank the **Pratiksha Trust** for their support.

aravindi@iisc.ac.in, deeppatel@iisc.ac.in,yaminihk@gmail.com, prasantg@iisc.ac.in