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

ABSTRACT

In this work, we consider the task of acoustic and articulatory feature based automatic classification of Amyotrophic Lateral Sclerosis (ALS) patients and healthy subjects using speech tasks. In particular, we compare the roles of different types of speech tasks, namely rehearsed speech, spontaneous speech and repeated words for this purpose. Simultaneous articulatory and speech data were recorded from 8 healthy controls and 8 ALS patients using AG501 for the classification experiments. In addition to typical acoustic and articulatory features, new articulatory features are proposed for classification. As classifiers, both Deep Neural Networks (DNN) and Support Vector Machines (SVM) are examined. Classification experiments reveal that the proposed articulatory features outperform other acoustic and articulatory features using both DNN and SVM classifier. However, SVM performs better than DNN classifier using the proposed feature. Among three different speech tasks considered, the rehearsed speech was found to provide the highest F-score of 1, followed by an F-score of 0.92 when both repeated words and spontaneous speech are used for classification.

Index Terms— Amyotrophic Lateral Sclerosis, Support Vector Machine, Deep Neural Networks, Electro Magnetic Articulography, Articulatory Kinematic Features

1. INTRODUCTION

ALS is a rapid and progressive neurodegenerative disease that mainly involves the degeneration of both upper and lower motor neurons [1] responsible for controlling voluntary muscle movements like chewing, walking, breathing and talking. It affects the speech motor functions of patients, thus causing dysarthria [2]. No single test can provide a definitive diagnosis of ALS [3]. The diagnosis is done using the revised El Escorial criteria [4]. It has been reported that the median time for diagnosis amounts to 14 months [5, 6]. Thus, timely diagnosis and assessment of ALS are crucial owing to such delays before a definitive diagnosis is reached. At present, ALS Functional Rating Scale-Revised (ALSFRS-R) [7] is used for monitoring the progression of ALS.

Speech production decline is among the earliest indicators of bulbar motor involvement due to ALS [8, 9]. Bulbar system has been considered to be a part of the four speech subsystems viz. respiratory, phonatory, articulatory, and resonatory [10, 11, 12]. Standardized diagnostic procedures are not available for bulbar dysfunction assessment in the case of ALS. Green et al. [8] have attempted bulbar ALS detection with the help of physiological measures of the four speech subsystems, and suggested the usage of speech motor performance, in particular, tongue movement speed for early detection as well as monitoring the progress of ALS. A number of works have reported changes in speech characteristics of ALS patients. For example, there is a reduction in formant transition slopes [13, 14, 15] and dispersion of the vowel space area [13, 14, 16, 17]. Moreover, the rate of change in the second formant is correlated with perceptual severity of dysarthria [15, 18]. Different structures of the articulatory subsystem (e.g., the lips, tongue, and jaw) are affected at different times during the progression of the disease. The tongue has been observed to be affected earlier and to a greater extent than the jaw and the lips [19]. Presumably, this non-uniform rate of deterioration leads to compensatory interactions between the articulators (e.g., tongue and jaw). Early movement studies revealed evidence supporting this notion and showed a decrease in the size of tongue movements but exaggerated jaw movements during speech tasks [20, 21]. Mefferd et al. [22] investigated lip and jaw movements to look for articulatory pattern inconsistencies in talkers with mild ALS with respect to speaking rates thereby trying to understand the speech rate decline during the early stages of ALS. Recently, the automatic detection of ALS from speech acoustics and articulatory samples using machine learning techniques has been attempted by Wang et al [23].

As different articulators are affected to varying degrees at different stages of the disease, they may cause degradation of speech to different amounts depending on the types of speech. In this work, we consider three types of speech tasks viz. rehearsed speech, spontaneous speech and repeated words. Each of these speech tasks have a different amount of cognitive load. During spontaneous speech, subjects can speak at their own pace and have a control on the content of the speech. On the contrary, repeated words and rehearsed speech require the subjects to remember what they need to speak. In fact, the amount of text the subject has to remember is more for rehearsed speech than for repeated words. These differences in cognitive load can, in turn, influence the articulatory gestures to different degrees. One of the goals in this work is to experimentally examine which speech task could be more suitable for the classification between ALS patients and healthy subjects. The classification experiment is conducted by using articulatory and acoustic data of 16 subjects, eight ALS patients and eight healthy subjects. The acoustic data in synchronism with articulatory data was recorded using Electromagnetic Articulograph (EMA). We experiment with the acoustic features as well as different features from kinematics of articulatory movements. ALS primarily affects the motor functioning which results in a reduction in articulator kinematics, such as maximum speed and maximum range, along with the duration between onset and offset of movement for an utterance [24]. We propose new

articulatory features to capture this information about articulatory kinematics which could be helpful in providing significant information for ALS classification. Classification experiments in a four-fold cross validation setup reveal that the proposed articulatory feature along with the SVM classifier results in the highest classification accuracy of 100% for rehearsed speech followed by that for spontaneous speech and repeated words.

2. DATASET

For the experiments in this work, speech and articulatory movement data was collected using the AG501 recording facility at Speech Kinematics Lab, Department of Speech Pathology and Audiology Lab, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore, India. Healthy subjects and patients were recruited from NIMHANS Hospital. Prior to data collection, an informed consent was obtained from each subject. The data collection was approved by the ethics committee of NIMHANS. All patients exhibited bulbar involvement in at least one region of the speech system (e.g., voice, soft palate, tongue, and/or face). The patients were native speakers of Kannada, and the recordings of all speech stimuli were also done in the same language. Proper care was taken to ensure age and gender balance within and across patient and control groups. Eight patients (5 males and 3 females) diagnosed with ALS by the Neurologist at NIMHANS, comprised the patient group (**P**). The healthy controls group (**C**) comprised of 4 males and 4 female participants. The details of age and gender of subjects, along with speech score on ALSFRS-R (in case of patients) are tabulated in Table 1. Participants of healthy controls group had no history of significant health, cognitive, or sensory problems or a history of other neurology conditions.

The articulatory movement data was recorded with an Electromagnetic Articulograph, AG501 [25]. In this study, articulatory motion data was obtained at a sampling rate of 250 Hz from the eight AG501 sensors placed at different articulators. The placement of the sensors closely matches with the recommended optimal sensor placement in [26]. In the recordings, we excluded the placement of the sensor on the velum to avoid discomfort of subjects while speaking. Out of the eight sensors, two are placed at the back of the two ears to act as a reference for head correction. The remaining six sensors are used to record the articulatory movements in the mid-sagittal plane. Three sensors are attached on the articulators outside the oral cavity (Upper Lip (UL), Lower Lip (LL) and Jaw), and the remaining three sensors are attached inside the oral cavity (Tongue Tip (TT), Tongue Body (TB) and Tongue Dorsum (TD)). The sensor movements along the midsagittal plane are captured by the X (horizontal) and Y (vertical) co-ordinates of the positional data provided by AG501, which is used in the present study. The horizontal and vertical movements of six sensors attached to the articulators result in 12-dimensional articulatory features namely, UL_x, UL_y, LL_x, LL_y, Jaw_x, Jaw_y, TT_x, TT_y, TB_x, TB_y, TD_x, TD_y. Thus, we obtain a 12-dimensional articulatory feature vector representing sensor positions along with the simultaneous audio recorded at a sampling rate of 48 kHz for each speech task.

Informal and formal assessment measures are used for assessment of articulation. The former are, in general, authentic means of articulation assessment [27] as they provide a relatively accurate portrayal of the natural speech of the subject. Both informal and formal assessment measures were carried out for both patients and healthy subjects. Informal assessment measures include 'rehearsed speech' (Task #1) and 'spontaneous speech' (Task #2), whereas formal assessment measures include 'repetition of words' (Task #3) from the Kannada Articulation test by Babu et al [28]. Each task was repeated 2-3 times. In *Task #1*, with the clinician's help, subject would rehearse a two-sentence template, "*My name is X. I am now in Bengaluru.*" and recite the same (X is subject's name and changes from subject to subject). Each subject performed the rehearsed speech task for six times in a row. *Task #2* is considered to be a major factor in deciding the necessity and benefits of treatment [29]. Participants in this case would produce a monologue to elicit a natural speech output. Two separate recordings of spontaneous speech were performed. In Task #3, the clinician would say a word in Kannada which would then be repeated by the subject. The words were chosen from a set of nine Kannada words viz. **Topi** (*Hat*), **Karnataka**, **Pustaka** (*Book*), **Pen**, **Alilu** (*Squirrel*), **Ili** (*Rat*), **Ungura** (*Ring*), **Chappali** (*Slipper*), **Kitaki** (*Window*). The number of times each word was chosen varied across subjects. Words in Task #3 are used to assess specific vowels and consonants, in specified positions at word level [28]. Task #1 has longer utterances than words whereas, Task #2 had the longest recordings. The recordings for speech tasks were annotated manually. Following removal of silence and noise segments, the range of duration of each task are reported in Table 2 for, both, ALS patients and healthy subjects.

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

Table 1: Details of the patients and the healthy subjects used in this work

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

Table 2: Range of durations of different speech tasks

3. ARTICULATORY AND ACOUSTIC FEATURES

The horizontal and vertical movements of six sensors attached to the articulators results in 12-dimensional articulatory features namely, UL_x, UL_y, LL_x, LL_y, Jaw_x, Jaw_y, TT_x, TT_y, TB_x, TB_y, TD_x, TD_y. Till now, to the best of our knowledge, in all the previous works, kinematics of articulators are derived considering horizontal and vertical directions separately. In [24], maximum speed and range of movements were calculated for kinematics of articulators by considering only the vertical direction. In [23], the statistical features of articulators are extracted from openSMILE [30] as univariate functions and do not exploit the relation between horizontal and vertical movements. For ALS classification, in this work we propose kinematics features by considering both horizontal and vertical directions.

The kinematics of each articulator in the mid-sagittal plane are computed as follows. Let $p_i^d(n)$ be the raw position in a given direction ($d \in \{X, Y\}$) of the i^{th} sensor at the n^{th} sample. The velocity $v_i^d(n)$ and acceleration $a_i^d(n)$ at the n^{th} sample in a particular direction d are computed as follows:

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

Form the above, we propose magnitude of velocity and acceleration for each articulator as follows:

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

where, $i \in \{UL, LL, Jaw, TT, TB, TD\}$. Different articulatory features based on eq. 1 and 2 are defined, namely *VE*, *AE*, *VAE*, *svaE* as shown in Table 3. We also consider the acoustic features, namely 13-dimensional Mel frequency cepstral coefficients (MFCC) with their velocity (Deltas) and acceleration (Delta-Deltas) coefficients, thus resulting in a 39-dimensional vector for a window size of 20 ms with a frame shift of 10 ms [31]. In addition to MFCCs, a combination of MFCCs and other articulatory features is used for classification (denoted by *MFCC* and *All* respectively in Table 3).

4. EXPERIMENTS AND RESULTS

4.1. Experimental setup

The following pre-processing step is performed before extracting features from the 12-dimensional articulatory data from EMA. As articulatory trajectory is slowly varying in nature [32], the articulatory trajectory is passed through a low pass filter with a cut-off frequency of 25 Hz in order to avoid effects of measurement error which is present in the form of high frequency noise in the data. Articulatory movement data is further down-sampled from 250 Hz to 100 Hz to obtain frame synchronized MFCCs. Since cues for ALS

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

Table 3: Articulatory and Acoustic low-level features

detection belong to para-linguistic information present in speech, we assume that they are encoded over a long-term. Hence, we extract supra-segmental features from the low-level features [33]. The supra-segmental features considered are the mean and standard deviations (SD), computed for every 0.8 sec (80 frames of low level features) with a shift of 0.2 sec. Note that the dimension for the supra-segmental features becomes twice of those listed in Table 3. The abbreviations in Table 3 are used to indicate supra-segmental features for rest of the paper.

The classification of ALS is done using the DNN and SVM, where training and testing are done based on supra-segmental features. For a DNN with L layers, a supra-segmental feature vector $\mathbf{x}$ is given as input to the first layer. Given the weight matrix $\mathbf{W}_l$, and hidden bias $\mathbf{b}_l$, the output of the l^{th} hidden layer $\mathbf{v}_l$ is given by,

$$\mathbf{v}_l(\mathbf{x}) = \tanh(\mathbf{u}_l(\mathbf{x})), l = 2, 3, \ldots, L-1 \quad (3)$$

where,

$$\mathbf{u}_l(\mathbf{x}) = \mathbf{W}_l \mathbf{v}_{l-1}(\mathbf{x}) + \mathbf{b}_l, \quad (4)$$

The last layer of the DNN has two units. Soft-max function was used as an activation which maps the outputs to the probability vector. The DNN was trained using the cross-entropy as a metric with Adam optimizer [34]. The number of layers L is set to four (including visible layers), and we choose 256 hidden units in each layer. The implementation was done using Keras library [35]. The SVM classifier is trained using *libsvm* package [36]. Radial basis function is chosen as the kernel function for the SVM classifier. γ and C parameters in SVM were optimized on the training set using a cross-validation setup.

4.2. Results and Discussion

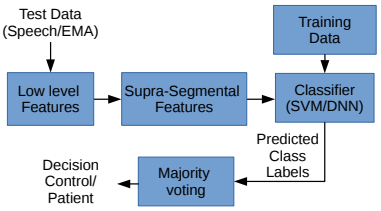
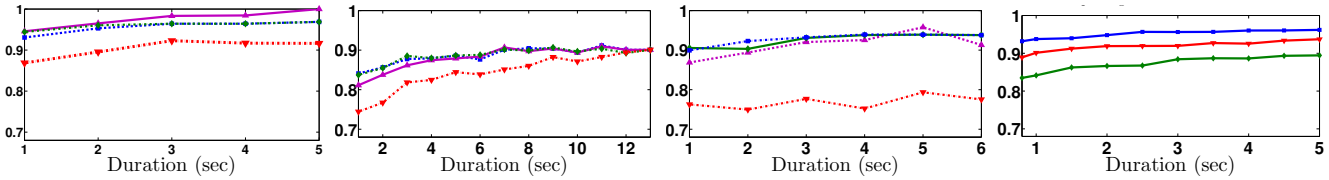


Fig. 1: Block diagram illustrating the steps in classification

All classification experiments (for each feature listed in Table 3) are performed in a four fold cross-validation setup where, in every fold, two patients and two healthy subjects are chosen as the test set and the remaining subjects as the training set in a round-robin manner. From the available training data, 15% is used as the development set for choosing the parameters of the classifiers. Fig. 1 illustrates the steps of the classification experiment. Initially, silence removed test recordings are used for low-level feature computation, and then supra-segmental features are computed from low-level features. The classifier (SVM/DNN) trained using the training data, is used to obtain decision on every supra-segmental feature in a given test utterance. To get a robust decision from the classifier, the predicted class decisions from all supra-segmental features are combined using majority voting to obtain one decision for a given test recording. To evaluate the classification performance using different tasks and also the features sets, we use F-score as a measure. For the evaluation of F-score, we considered patients as positives and healthy subjects as negatives. With a fixed supra-segmental duration of 0.8 sec, we first present the classification performance by majority voting on the whole utterance. We also examine how the classification performance changes when the entire recording is not available, rather, initial parts of the recordings of varying duration are used for classification. This helps in quantifying a minimum test recording duration needed to obtain a desired classification performance. Finally, we also examine the classification performance by varying the supra-segmental duration.

Utterance level classification: The majority voting is done at an utterance level, and one decision is made for the whole utterance. F-score is computed across all utterances of the test subjects present in a given fold. Classification results in terms of an F-score obtained for SVM and DNN are summarized in Tables 4 and 5, respectively. Both tables report the F-score, averaged across all folds, with SD in bracket.

From the average F-score in Table 4 and 5, it is clear that SVM performs either equally or better than the DNN classifier for different features. It is also seen that the proposed articulatory features (*AE*, *VE*, *VAE*) result in a better classification performance compared to that of *svaE* and *MFCC*. Among the proposed articulatory features, *VAE* performs the best using both classifiers. In particular, the average F-score using *VAE* turns out to be 0.95 for rehearsed speech



(a) Rehearsed Speech (b) Spontaneous Speech (c) Repeated words (d) Supra-Segmetal Analysis

Fig. 2: a,b,c): F-score for SVM based classification ('●' MFCC (▼), '●' VE (■), '●' AE (▲), '●' VAE (♦)). d): Supra-segmental Duration Analysis ('●' Task #1 (■), '●' Task #2 (♦),'●' Task #3 (▼))

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)**

Table 4: F-score using SVM 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)**

Table 5: F-score using DNN classifier and different features used

followed by 0.92 for, both, spontaneous speech and repeated words. Interestingly, the proposed *VE* feature performs the best using SVM classifier with an F-score of 1 using rehearsed speech. *All* features resulted in an F-score of 0.9 which is higher than those using *svaE* and *MFCC* but lower than that using *VAE*. This could be due to the fact that *All* features have a dimension of 174, while *VAE* features have a dimension of 24 for the same amount of training data. This, in turn, could impact the trained classifier and, hence, the classification performance.

Classification using test recording of different durations: In this setup, for a given duration level, we split each utterance into segments. A decision of control/patient classification is done segment-wise using majority voting. F-score is computed across all segments of the utterances of the test subjects in a given fold. We choose the minimum segment duration to be 1 sec which is close to supra-segmental feature duration (0.8 sec). The duration levels are varied with an increment of 1 sec to a maximum of 5 sec (rehearsed), 13 sec (spontaneous), 6 sec (repeated words) based on the minimum durations available in the database for each task. Fig. 2 (a,b,c) shows the F-score plots using SVM for different duration values. Consistent with Table 4, the *AE*, *VE* and *VAE* features perform better than the rest of the features for all chosen durations. We observe that the performance of the SVM classifier in spontaneous speech is less using all the features compared to the rehearsed speech and repeated words. The classifier performance increases monotonically as the duration level increases for all tasks. Similar trend in performance is also observed in the case of DNN.

Classification with varying duration of supra-segmental features: In addition to computing supra-segmental features with a fixed duration of 0.8 seconds, we also experiment with different supra-segmental duration. So, supra-segmental features are computed on every interval (I) ranging from 0.8 to 5 seconds in steps of s seconds ($s = 0.25 \times I$). In this experiment, we do not perform majority voting, instead F-score is computed using decisions on all supra-segmental features from testing set. The average F-score plots using feature *VAE* and SVM classifier for all three tasks are shown in Fig. 2 (d). From Fig. 2 (d), we can observe that the F-score monotonically increases with the duration of supra-segmental feature for all tasks, and saturates after 3 seconds. In order to get an insight into discriminative performance of each component of the *VAE* feature, we compute Fisher Discriminative Ratio (FDR) [37] for supra-segmental features computed at 0.8 sec duration. We found that the SD of *AE* for TT performed the best among all tasks. Fig. 3 shows the histogram of the SD of *AE* for TT for both ALS patients and healthy subjects for all three tasks. The FDR for each task is shown at the top of each plot. From Fig. 2 and 3 it can be noted that, consistently, Task #1 performs the best among the three tasks, and Task #3 performs better than Task #2.

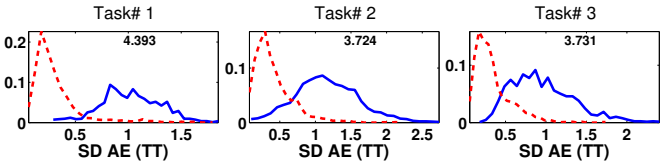


Fig. 3: Histogram of best component among *VAE* with supra-segmental feature computed at 0.8 sec (ALS patients (...) and healthy subjects (—))

5. CONCLUSIONS AND FUTURE WORK

In this work, we study three different speech tasks for classifying ALS patients and healthy control. Experiments are conducted on 16 subjects, 8 controls and 8 patients. From the classification results using SVM and DNN, we observe that the rehearsed speech performs better than the spontaneous and repeated words. We also propose new articulatory features for the classification task. Experiments using the three speech tasks show that the proposed kinematic features consistently perform the best among all the feature sets considered. It should be noted that the experiments in this work are conducted using Kannada subjects, and the consistency of the conclusions across different languages needs further investigation. Given the robustness of the proposed kinematic features, estimating these from acoustics for classification task would be interesting. Developing a subject independent Acoustic-to-Articulatory Inversion (AAI) model, which could help to estimate articulatory features for ALS classification is a part of our future work.

Acknowledgement: Authors thank G. Nisha Meenakshi and Ms. Preetie Shetty (NIMHANS) 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.

6. 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.

[3] Amyotrophic Lateral Sclerosis (ALS) Fact Sheet, "https://www.ninds.nih.gov/disorders/patient-caregiver-education/fact-sheets/amyotrophic-lateral-sclerosis-als-fact-sheet," September 2014.

[4] Benjamin Rix Brooks, Robert G Miller, Michael Swash, and Theodore L Munsat, "El escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, vol. 1, no. 5, pp. 293–299, 2000.

[5] Matthew C Kiernan, Steve Vucic, Benjamin C Cheah, Martin R Turner, Andrew Eisen, Orla Hardiman, James R Burrell, and Margaret C Zoing, "Amyotrophic lateral sclerosis," *The Lancet*, vol. 377, no. 9769, pp. 942 – 955, 2011.

[6] A Chio, "Isis survey: an international study on the diagnostic process and its implications in amyotrophic lateral sclerosis," *Journal of neurology*, vol. 246, pp. III1–III5, 1999.

[7] Jesse M Cedarbaum, Nancy Stambler, Errol Malta, Cynthia Fuller, Dana Hilt, Barbara Thurmond, Arline Nakanishi, Bdnf Als Study Group, 1A complete listing of the BDNF Study Group, et al., "The alsfrs-r: a revised als functional rating scale that incorporates assessments of respiratory function," *Journal of the neurological sciences*, vol. 169, no. 1, pp. 13–21, 1999.

[8] Jordan R. Green, Yana Yunusova, Mili S. Kuruvilla, Jun Wang, Gary L. Pattee, Lori Synhorst, Lorne Zinman, and James D. Berry, "Bulbar and speech motor assessment in als: Challenges and future directions," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 14, no. 7-8, pp. 494–500, 2013.

[9] Y. Yunusova, J. R. Green, L. Greenwoode, J.Wang, G. Pattee, and L. Zinman, "Tongue movements and their acoustic consequences in ALS," *Folia Phoniatrica et Logopaedica*, vol. 64, pp. 94–102, 2012.

[10] Steven M Barlow and James H Abbs, "Force transducers for the evaluation of labial, lingual, and mandibular motor impairments," *Journal of Speech, Language, and Hearing Research*, vol. 26, no. 4, pp. 616–621, 1983.

[11] Bruce E Murdoch, *Dysarthria: a physiological approach to assessment and treatment*, Nelson Thornes, 1998.

[12] R Netsell, WK Lotz, and SM Barlow, "A speech physiology examination for individuals with dysarthria," *Recent advances in clinical dysarthria*, pp. 3–37, 1989.

[13] Greg S. Turner, Kris Tjaden, and Gary Weismer, "The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research*, vol. 38, no. 5, pp. 1001–1013, 1995.

[14] Gary Weismer, Ruth Martin, Ray D. Kent, and Jane F. Kent, "Formant trajectory characteristics of males with amyotrophic lateral sclerosis," *The Journal of the Acoustical Society of America*, vol. 91, no. 2, pp. 1085–1098, 1992.

[15] Moira Mulligan, Joseph Carpenter, Joanne Riddel, Maureen Kenny Delaney, Gary Badger, Patricia Krusinski, and Rup Tandan, "Intelligibility and the acoustic characteristics of speech in Amyotrophic Lateral Sclerosis (ALS)," *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 3, pp. 496–503, 1994.

[16] BK Yamini, N. Shivashankar, and A. Nalini, "Vowel space area in patients with Amyotrophic Lateral Sclerosis," *Amyotrophic Lateral Sclerosis*, vol. 9, Supp.1, pp. 118, 2008.

[17] Greg S. Turner and Kris Tjaden, "Acoustic differences between content and function words in amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research*, vol. 43, no. 3, pp. 769–781, 2000.

[18] RD Kent, R Netsell, and LL Bauer, "Cineradiography assessment of articulatory mobility in the dysarthrias," *Journal of Speech and Hearing Disorders*, vol. 40, no. 4, pp. 467–480, 1975.

[19] R. DePaul and B. R. Brooks, "Multiple orofacial indices in amyotrophic lateral sclerosis.," *Journal of Speech & Hearing Research*, vol. 36, no. 6, pp. 1158–1167, 1993.

[20] Hajime Hirose, Shigeru Kiritani, and Masayuki Sawashima, "Velocity of articulatory movements in normal and dysarthric subjects," *Folia Phoniatrica et Logopaedica*, vol. 34, no. 4, pp. 210–215, 02 1982.

[21] Yana Yunusova, Gary Weismer, John R. Westbury, and Mary J. Lindstrom, "Articulatory movements during vowels in speakers with dysarthria and healthy controls," *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 3, pp. 596–611, 2008.

[22] A. S. Mefferd, J. R. Green, and G. L. Pattee, "Speaking rate effects on articulatory pattern variability in talkers with mild ALS," in *Clinical Linguistics & Phonetics*, 2014, vol. 28, pp. 799–811.

[23] Jun Wang, Kothalkar Prasanna V., Beiming Cao, and Daragh Heitzman, "Towards automatic detection of amyotrophic lateral sclerosis from speech acoustic and articulatory samples," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2016, vol. 08-12-September-2016, pp. 1195–1199.

[24] Mili Kuruvilla-Dugdale and Miguel Chuquilin-Arista, "An investigation of clear speech effects on articulatory kinematics in talkers with ALS," in *Clinical Linguistics & Phonetics*, 2017, vol. 31, pp. 725–742.

[25] 3D Electromagnetic Articulograph, "http://www.articulograph.de/," .

[26] Ashok Kumar Pattem, Aravind Illa, Amber Afshan, and Prasanta Kumar Ghosh, "Optimal sensor placement in electromagnetic articulography recording for speech production study," *Computer Speech & Language*, vol. 47, pp. 157–174, 2018.

[27] Renee Fabus and Felicia Gironda, *A Guide to Clinical Assessment and Professional Report Writing in Speech-Language Pathology*, Delmar, 2011.

[28] RM Babu, N Ratna, and R Bettagiri, "Test of articulation in Kannada," *The JAIISH*, vol. 3, pp. 7–19, 1972.

[29] Jacqueline Bauman-Waengler, *Articulatory and Phonological Impairments: A Clinical Focus*, Pearson, 2012.

[30] Florian Eyben, Martin Wöllmer, and Björn Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1459–1462.

[31] Steve J Young and SJ Young, *The HTK hidden Markov model toolkit: Design and philosophy*, University of Cambridge, Department of Engineering, 1993.

[32] Prasanta Kumar Ghosh and Shrikanth Narayanan, "A generalized smoothness criterion for acoustic-to-articulatory inversion," *The Journal of the Acoustical Society of America*, vol. 128, no. 4, pp. 2162–2172, 2010.

[33] Björn Schuller, Stefan Steidl, Anton Batliner, Felix Burkhardt, Laurence Devillers, Christian MüLler, and Shrikanth Narayanan, "Paralinguistics in speech and language—state-of-the-art and the challenge," *Computer Speech & Language*, vol. 27, no. 1, pp. 4–39, 2013.

[34] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," in *CoRR*, 2014, vol. abs/1412.6980.

[35] François Chollet et al., "Keras," https://github.com/fchollet/keras, 2015.

[36] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[37] Andrew R Webb, *Statistical pattern recognition*, John Wiley & Sons, 2003.