INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Comparison of Speech Tasks and Recording Devices for Voice Based Automatic Classification of Healthy Subjects and Patients with Amyotrophic Lateral Sclerosis

Suhas B N[⋆], *Deep Patel*[⋆], *Nithin Rao*[‡], *Yamini Belur*[⊙], *Pradeep Reddy*[∧], *Nalini Atchayaram*[∧], *Ravi Yadav*[∧], *Dipanjan Gope*[•], *Prasanta Kumar Ghosh*[⋆]

[⋆]EE Department, Indian Institute of Science, Bengaluru 560012, India
[‡]EE Department, University of Southern California, Los Angeles, CA 90089, USA
[⊙]Department of Speech Pathology and Audiology, NIMHANS, Bengaluru 560029, India
[∧]Department of Neurology, NIMHANS, Bengaluru 560029, India
[•]ECE Department, Indian Institute of Science, Bengaluru 560012, India

bnsuhas96@gmail.com, prasantg@iisc.ac.in

Abstract

We consider the task of speech based automatic classification of patients with amyotrophic lateral sclerosis (ALS) and healthy subjects. The role of different speech tasks and recording devices on classification accuracy is examined. Sustained phoneme production (PHON), diadochokinetic task (DDK) and spontaneous speech (SPON) have been used as speech tasks. The chosen five recording devices include a high quality microphone and built-in smartphone microphones at various price ranges. Experiments are performed using speech data from 25 ALS patients and 25 healthy subjects using support vector machines and deep neural networks as classifiers and suprasegmental features based on mel frequency cepstral coefficients. Results reveal that DDK consistently performs better than SPON and PHON across all devices for discriminating ALS patients and healthy subjects. Considering DDK, the best classification accuracy of 92.2% is obtained using a high quality microphone but the accuracy drops if there is a mismatch between the microphones for training and test. However, a classifier trained with recordings from all devices together performs more uniformly across all devices. The findings from this study could aid in determining the choice of the task and device in developing an assistive tool for detection and monitoring of ALS.

Index Terms: Amyotrophic lateral sclerosis, support vector machines, deep neural networks

1. Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive neuro degenerative disorder causing upper and lower motor neuron degeneration. Patients suffering from ALS have an average survival of 2 to 4 years with a worldwide annual incidence of about 1.9 per 100,000 [1, 2] and a median diagnosis time of 14 months [3]. Only 5-10% of all patients survive beyond 10 years [4]. In India, ALS has a prevalence rate of 4/100,000 with an annual incidence of 1/100,000 and a male to female ratio of 5:7 [2]. Currently, Revised El Escorial criteria is used for the diagnosis of ALS [5], whereas for the monitoring of progress of the disease, ALS Functional Rating Scale-Revised (ALSFRS-R) i s used [6]. Patients with ALS experience symptoms of progressive muscle atrophy and weakness leading to problems including dysphagia, dyspnea, orthopnea and dysarthria [4]. Dysarthria in case of ALS patients occurs frequently with increasing severity as the disease progresses [7, 8]. About 30% of all ALS patients experience dysarthria as the first symptom [9, 10]. Often, the assessment of speech impairment is done based on clinician's auditory perception which is subjective. These judgements maybe inconsistent [11]. Therefore, automated methods for early detection of speech impairment due to ALS could avoid clinicians' subjectivity in diagnosis of the disease and reduce diagnosis time. The speech impairment due to ALS is caused by the muscle disorders which, in turn, affect the speech articulators. There have been attempts to use Electromyography (EMG) to assess neuromuscular disorder [12], and perform automatic classification using features extracted from EMG signal [13, 14]. The rate of articulatory movement of ALS patients have been studied [15] [16], and were found to be lower than those of healthy subjects. On the contrary, there are few works that use impact of ALS on voice and use voice cues to perform automatic classification of ALS patients. Kent et al. [17, 8] studied the relationship between speech intelligibility on a single word identification test using average second-formant (F2) slope and found that F2 slope index is an useful acoustic measure of speech proficiency in ALS. Another study by Kent et al. [18], showed that the most disruptive phonetic features in speech, impaired by ALS, involve phonatory function, place and manner of articulation for lingual consonants and regulation of tongue height for vowels suggesting their potential use as an index of bulbar muscle impairment in ALS. Tomik et al. [19] studied the most significantly affected vowels for ALS patients in order to detect and monitor the progression of the disease based on the acoustic analysis of specific sounds. Gomez et al. [20] used running speech segments to infer articulation kinematics to detect early symptoms and monitor the evolution of the ALS. Yamini et al. [21] observed a reduction in the vowel space area in case of bulbar ALS patients compared to that of healthy controls. Using syllable rate and maximum phonation duration, Yamini et al. [22] also found that diadochokinetic rate and phonation tasks are efficient ways to discern between healthy subjects and ALS patients. Pedro et al. [23] proposed a speech articulation biomechanical model to assess the state and progress of ALS. Taylor et al. [24] attempted automatic classification of ALS patients based on fractal analysis and using diadochokinetic (DDK) rates as speech tasks. Different speech based studies for ALS have used a variety of speech tasks. Kent et al. [18] used different vowels and fricatives as tasks. Green et al. [25] used read speech using bamboo passage, sustained vowel and repeated words. Dif-

http://dx.doi.org/10.21437/Interspeech.2019-1285

Table 1: *ALSFRS-R Score versus Age for all subjects*

ALSFRS-R Score	0	1	2	3	4	All ALS	All Healthy
Total subjects	5	5	5	5	5	25	25
Mean Age	52	59.2	59.4	53.2	60.2	56.8	51.8
Std. Dev.	8.2	12.1	11.0	6.5	11.6	9.9	7.7

Table 2: *Language wise distribution for all subjects*

Language	Bengali	Hindi	Kannada	Odiya	Tamil	Telugu
ALS Patient count	5	5	5	3	3	4
Healthy count	5	5	5	3	3	4

ferent tasks have been used in various studies in the past but there are no investigations on relative role of each task for automatically classifying healthy subjects and patients with ALS. The ALS patients considered in this work come from different parts of India. The work presented here is part of a project that aims to develop a smartphone application for Indian population that can detect and monitor the degree of ALS for providing treatment at an early stage. It is important to determine speech task that has required discriminatory power for achieving good classification accuracy as well as is suitable given the diversity in terms of different vernaculars being spoken in the context of Indian demographic. In this work, we have chosen spontaneous speech (SPON), diadochokinetic rate (DDK) and sustained phoneme production (PHON) as the three speech tasks. Due to different socio-economic backgrounds in India, there is a great variety in the phones used by the target users. This, in turn, requires investigation of the robustness of classifiers across different recording devices. For this reason, we have experimented with five recording devices: Apple iPhone 7 (referred to as IPH), Moto G5 Plus (MOT), Xiaomi Redmi 4 (XIA), Zoom H-6 recorder with XYH-6 X/Y capsule high-quality unidirectional microphone [26] (ZOO) and Dell XPS 15 Laptop (LAP). The smartphones have been chosen such that they represent popular brands at various price ranges. Recordings from 50 subjects (25 controls and 25 ALS patients) are used for comparing the classification accuracy across all speech tasks and recording devices.

2. Dataset

For all experiments in this work, speech data is considered from 25 male patients and 25 male healthy subjects. All patients had been recruited from National Institute of Mental Health and Neurosciences (NIMHANS), Bengaluru, India. The data collection has been approved by the ethics committee of NIMHANS and informed consent forms were signed by the subjects prior to the data collection. All patients included in this study were confirmed as having ALS by Neurologists at NIMHANS as per the El Escorial criteria. The details of age are provided in Table 1 for patients based on each ALSFRS-R score as well as for healthy subjects. The native languages of patients & controls are provided in Table 2. The selected subjects are matched for age, gender and language for uniformity.

The recording setup using five devices (IPH, MOT, XIA, ZOO, LAP) is done with the subject at a distance of 2 feet from the recording devices. For all devices, speech data was recorded at a sampling rate of 44.1kHz. Although read speech was used as a stimulus in previous works, we chose SPON due to poor literacy level of a few patients.

In PHON, subjects were instructed and demonstrated to produce a sustained phoneme of five vowels, namely, /a/, /i/, /o/, /u/, /æ/, and three fricatives, namely, /s/, /sh/, and /f/. Subjects

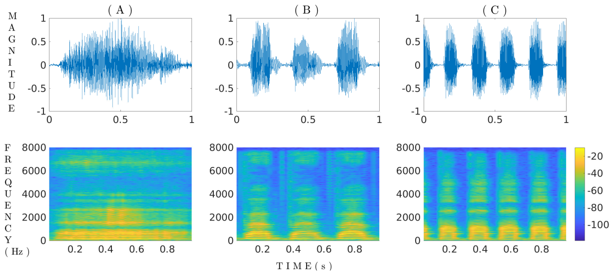

Figure 1: *Waveforms and spectrograms of recordings from of ALS patients and a healthy control uttering 'pa' for a duration of 1 second. (A) a patient with ALSFRS-R score of 0, (B) a patient with score of 2, (C) a healthy subject.*

were asked to do this for 5 seconds at a comfortable pitch and loudness level, after taking a deep breath. The same process was repeated three times in succession for each of the vowels and fricatives. Vowel prolongation is a task which isolates the respiratory-phonatory system for speech [27]. The fricative prolongation requires the respiratory-articulatory competence. The total duration of recording for PHON is 7.9 hours considering all fifty subjects across all devices.

DDK is used for assessing speed and regularity of rapid and repetitive articulatory movements. DDK consists of two parts: (a) Alternating Motion Rates (AMRs), which include rapid repetition of monosyllabic targets-'pa','ta','ka' & (b) Sequential Motion Rates (SMRs) measure the ability of articulators to move quickly and in a proper sequence from one articulatory position to another [27]. SMR were captured through the syllabic targets such as 'pataka' and 'badaga'. Thus, DDK is used to measure articulatory speed and precision in the movements of jaw, lips, anterior and posterior tongue, phonatory support, adequacy of velopharyngeal closure, and respiratory support for sustaining the task. We expect the corresponding speech recordings to reflect such characteristics, cues which could be used to discriminate ALS patients from healthy subjects. Subjects were asked to repeat the target production for three trials for a duration of upto 5 seconds. The total duration of recording for this task is 5.36 hours for all subjects across all devices. A spectrogram in Figure 1 depicts three speech recordings from ALS patients and healthy subjects repeating 'pa' for a duration of 1s. It is observed that the healthy person (C) is able to utter over five 'pa' in a second while a patient with ALSFRS-R score of 2 (B) is able to repeat 'pa' three times in a second. A patient with ALSFRS-R score of 0 (A) utters 'pa' only one time in a second. The articulatory precision was most affected in patient A than in patient B.

In SPON, monologue was elicited wherein, the subjects were instructed to spontaneously talk about a festival celebration and a recent place of visit. For this, preparation time of a few minutes was given to the subjects before they could start speaking. These monologues are spoken in their native language thus eliciting a natural response. SPON is an informal assessment measure but has a good representation of the natural speech of a subject, thus making it an useful task for assessing a subject's articulation [28]. It is also useful in evaluating an integrated function of all components in speech production (respiration, phonation, articulation, resonance, and prosody) [28]. The total duration of recording for this task is 7 hours considering all fifty subjects across all devices.

For all experiments, the begin and end time for each task were noted down separately using which the speech segments of interest were obtained from the entire recording.

3. ALS vs Healthy classification

The ALS patient and healthy subject classification consists of training and test phases. The first step in both phases is computation of the acoustic features from the speech recording. The acoustic features are the 12-dimensional MFCC (excluding energy coefficient) along with their velocity and acceleration co-efficient, resulting in a 36-dimensional feature vector computed using a window size of 20 ms and a frame shift of 10 ms [29]. The speech recordings are downsampled from 44.1kHz to 16kHz before computing MFCC features. Cepstral mean variance normalization (CMVN) has been applied to the raw MFCCs. These are referred to as low-level features. For classification experiments, we extract suprasegmental features from the low level features, and use these higher-level, long-term features as input to the classifier instead of providing the low-level, short-term frame-based spectral features. This is because paralinguistic information, such as ALS disease condition, could be embedded in subtle cues present in long-term features and this, in turn, can increase the performance of the classifier [30, 31]. The suprasegmental features are the mean, median, and standard deviation (SD) of each MFCC computed for an analysis window of N_w seconds with a shift of N_{sh} seconds. Thus, the dimensions of the suprasegmental feature vectors become three times that of the original MFCC feature vector ($108xN_{aw}$), where N_{aw} is the number of suprasegmental analysis windows for each recording. In the training phase, the suprasegmental features obtained from every analysis window along with their class labels (ALS & Healthy) are used to train the classifier (model). In this work, we use two models for classification - support vector machine (SVM) and deep neural network (DNN). $N_w = \{0.5, 0.8, 1, 2, 3\}$s with $N_{sh} = 0.1$s have been used for the analysis.

4. Experimental Setup

For classification, a five-fold cross-validation setup is used. Five groups, each with ten subjects are formed. Subjects in each group is chosen such that they are balanced in all aspects as mentioned earlier. It is ensured that the subjects belonging to both healthy population and ALS patients are equally present in each group. In every group, five ALS patients are chosen in a way that there is equal representation of ALSFRS-R scores.

In each fold, four groups are used for training and the remaining group is used as the test set in a round robin fashion. 15% of the training data has been used as a validation set. The SVM classifier with radial basis kernel has been trained using the libsvm package [32]. Optimal values of the soft margin constant (C) and width of the Gaussian kernel (γ) are selected by maximizing the performance on the validation set. For DNN, the optimal choice of the activation function (AF) corresponding to each hidden layer, number of hidden layers (HL), and number of neurons (NN) in each hidden layer are determined by the validation loss. The parameters that result in the least validation loss are chosen for the experiments. The candidate AFs, HLs, and NNs for which the validation loss is minimized are {'sigmoid', 'tanh', 'relu'}, {1, 2, 3}, and {64, 128, 256, 512} respectively. The optimal DNN architecture was found to have 'tanh' as the AF irrespective of the speech task and the recording device chosen.

The trained model is then used to obtain a decision on every suprasegmental feature for the test data. The utterance level decision is obtained by majority voting on the decisions using the suprasegmental features. Training of the DNN has been done using cross-entropy as the loss function with Adam optimizer [33]. Keras library has been used for the implementation. Performance of the automatic classification is determined by the classification accuracy which is computed as the number of test utterances for which the decision from the classifier matches the ground truth class label.

Table 3: *Average accuracy (standard deviation in brackets) for SVM and DNN classifiers for N_w = 0.8s. The bold entries indicate higher classification accuracy between the two classifiers for every speech task and recording device combination*

Device	Classifier	SPON (%)	DDK (%)	PHON (%)
MOT	SVM	**81.84 (7.51)**	**90.40 (9.10)**	79.88 (3.23)
	DNN	81.79 (7.67)	88.80 (11.01)	**80.13 (3.62)**
ZOO	SVM	79.79 (0.47)	90.40 (6.23)	**82.98 (4.04)**
	DNN	**83.84 (9.60)**	**92.20 (4.71)**	77.90 (9.35)
IPH	SVM	79.74 (6.41)	**89.20 (7.16)**	**80.93 (3.44)**
	DNN	**79.79 (7.09)**	88.40 (7.13)	78.45 (9.54)
XIA	SVM	**84.89 (4.87)**	**87.60 (3.85)**	78.38 (6.55)
	DNN	82.89 (11.97)	86.40 (5.37)	**78.57 (7.12)**
LAP	SVM	83.89 (5.34)	**88.80 (8.56)**	78.80 (2.03)
	DNN	**87.95 (9.70)**	87.60 (8.17)	**81.15 (5.62)**
Avg	SVM	82.03 (4.92)	**89.28 (6.98)**	**80.19 (3.86)**
	DNN	**83.25 (9.20)**	88.68 (7.28)	79.24 (7.05)

5. Results and Discussion

Table 3 shows the utterance level classification accuracies averaged across all folds using SVM and DNN classifiers separately using $N_w = 0.8$s. The number in the bracket indicates the S.D of the accuracies across all folds. The last two rows (indicated by 'Avg') in Table 3 report accuracies averaged across all recording devices for each speech task.

From the 'Avg' accuracies, it is clear that the highest classification accuracy is achieved in DDK task using both SVM and DNN classifiers. In particular, SVM performs better than DNN classifier by an average accuracy of 0.6%. When DDK is used as a speech task, SVM performs better than DNN in case of all devices except ZOO where SVM yields an average classification accuracy of 90.40% while DNN achieves a classification accuracy of 92.20%. Across all devices, the highest SVM-based average classification accuracy of 90.40% is obtained using MOT and ZOO in the case of DDK task. This suggests that these devices are superior than the remaining three in terms of preserving cues for healthy subjects and ALS patients classification. In SPON, the highest classification accuracy of 87.95% is obtained using LAP. However, it is still lower than the accuracy (88.80%) obtained using DDK task in LAP. This indicates more discriminatory power of DDK compared to SPON.

When averaged over all devices (Avg case), unlike DDK task, DNN performs 1.22% better than SVM. However, considering Avg case, SVM performs 0.95% better than DNN classifier in the case of PHON task. The Avg accuracy of 80.19% in PHON task using SVM is 9.09% lower than by using DDK task suggesting the superiority of the DDK task for healthy subject and ALS patient discrimination. Ranking the device wise performance for the three tasks, it is observed that LAP performs the best (87.95%) among all devices for SPON followed by XIA, ZOO, MOT and IPH (79.79%). For DDK, ZOO performs the best among all devices (92.20%) followed by MOT, IPH, LAP and XIA (87.60 %). For PHON, ZOO (82.98 %) performs the best followed by LAP, IPH, MOT and XIA (78.57%).

The range of classification accuracy is 8.16% for SPON task, 4.6% for DDK and 4.41% for PHON task. This range highlights the robustness of the DDK task since it consistently achieves the highest accuracy across devices with a small variation. From the data, Table 4 is constructed where a device was ranked based on its relative performance. Rank 1 (5) indicates the best (worst) performing device. Total score is calculated by adding the ranks among all tasks. A lower total indicates better performance.

In order to investigate which features among suprasegmental features contribute the most to the classification accuracy, an experiment is conducted where a subset ($3xN_{aw}$) of suprasegmental features is considered. Each such ($3xN_{aw}$) consists of the mean, median and SD (MMSD hereon) of one MFCC. This was repeated 36 times for each dimension of the MFCC. The MMSD of the 2^{nd}, 14^{th} and 26^{th} dimension of MFCC yield the highest accuracy of 71.2%, 84.4% and 84.4% among the twelve static, twelve velocity and twelve acceleration coefficients respectively. It is to be noted that 2^{nd}, 14^{th} and 26^{th} dimensions of MFCC correspond to the 3^{rd} discrete cosine transform (DCT) basis function.

Table 4: *Rank and score of each device for different speech tasks*

Device	MOT	ZOO	IPH	XIA	LAP
Rank using SPON task	4	3	5	2	1
Rank using DDK task	2	1	3	5	4
Rank using PHON task	4	1	3	5	2
Total score	10	5	11	12	7

Microphone characteristics vary from device to device and have varying performance under different stimuli. To check which device has robust characteristics, we compare the performance of one device model on recordings from other devices as test set. Table 5 shows the accuracies for DDK task for N_w = 0.8s averaged across all folds for each device model against test data of all devices (including matched case). Apart from XIA (where test data of ZOO secured highest accuracy), the matched case of model and test data perform the best among test devices.

Table 5: *Performance comparison using device model on all devices. Bold entries indicate higher classification accuracy for each device model and italics indicate the best averaged performance for a device model*

Model	Test (Accuracy(%) and SD)					Avg
	MOT	ZOO	IPH	XIA	LAP	
MOT	**90.40** **(9.10)**	90.00 (7.35)	88.40 (5.37)	85.84 (7.87)	88.80 (8.67)	88.69 (7.67)
ZOO	89.20 (4.60)	**90.40** **(6.23)**	87.20 (3.63)	76.80 (4.60)	87.64 (2.94)	86.25 (4.40)
IPH	**89.20** **(9.44)**	**89.20** **(7.16)**	**89.20** **(7.16)**	82.80 (6.26)	88.00 (8.12)	87.68 (7.63)
XIA	86.40 (10.62)	**89.60** **(7.40)**	88.80 (4.60)	87.60 (3.85)	86.40 (9.10)	87.76 (7.11)
LAP	87.20 (10.64)	87.20 (10.06)	85.20 (7.56)	78.00 (8.12)	**88.80** **(8.56)**	85.28 (8.99)
ALL	89.20 (8.79)	91.20 (9.12)	90.00 (6.32)	87.60 (5.55)	87.60 (8.65)	***89.12*** ***(7.69)***

A combined model with training data taken equally from all devices is built (referred to as ALL in Table 5). It is observed that the performance of any model on XIA drops when compared to their matched case (except XIA). Considering the average accuracy across different devices for test, it turns out that MOT based model is the most robust model (next to ALL that achieves the highest averaged accuracy of 89.12%).

Observing the range of performance of a device as test, it is seen that MOT (86.4 to 90.4%), ZOO (87.2 to 90.4%), IPH (85.2 to 89.2%) and LAP (86.4 to 88.8%) show minimal variation in performance across models (2.4 to 4%) as opposed to XIA (76.8 to 87.6%) with a range of 9.8%.

To check if there is a change in classification accuracy for different choices of N_w, the classification experiments are repeated for $N_w = \{0.5,1,2,3\}$s. Figure 2 shows the device wise performance for each speech task with varying N_w. The trend observed in Table 3 is seen here with DNN performing better than SVM for SPON task (except XIA). For DDK, SVM performs better than DNN (except IPH). SVM performs better for PHON in the case of MOT, ZOO and IPH while DNN performs better in XIA and LAP. In SPON, it is observed that for all devices, the accuracy rises from 0.5s to 1s and then either decreases or remains the same. In DDK, the accuracy reaches a maximum at N_w = 0.8s for MOT, ZOO and LAP while it is second best for XIA and LAP, for which N_w = 0.5s has a higher accuracy. For PHON, the accuracy reaches the maximum at N_w = 1s (except MOT) while N_w = 0.8s was second best. Although N_w=1s yields the best accuracy among all choices of N_w for most of the task and device combinations, N_w=0.8s with DDK task and ZOO device achieves the best performance among all combinations.

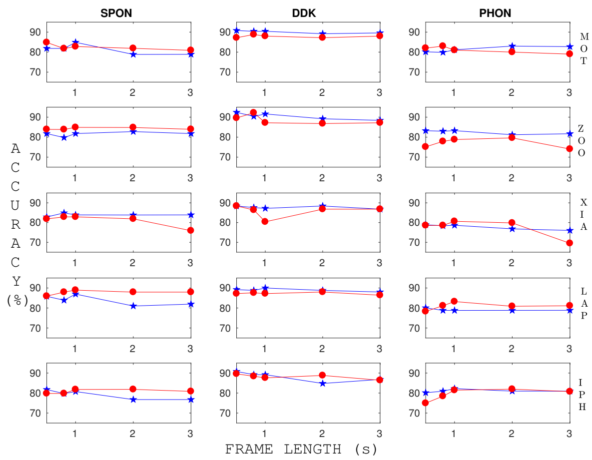

Figure 2: *Classification accuracy by varying N_w. ⋆ SVM, • DNN. Each column and row correspond to one speech task and recording device.*

6. Conclusions

In this work, we have performed a comparative study on speech based classification of healthy subjects and ALS patients using different speech tasks and five recording devices. The experiments show that DDK task consistently performs better than other tasks for discriminating ALS patients and healthy population. While the high quality microphone achieves the highest classification accuracy of 92.2%, the accuracy lies in the range 87.6%-90.4% using rest of the recording devices, that include three smartphones and one laptop. It is observed that when a classifier is built using recordings from all devices, classification accuracy on smartphone recordings improves. The findings from this study could aid in determining the choice of the task and device in developing an assistive tool for detection and monitoring of ALS.

7. Acknowledgement

We thank the Department of Science & Technology, Government of India and the Pratiksha Trust for their support.

8. References

[1] Karissa C Arthur, Andrea Calvo, T Ryan Price, Joshua T Geiger, Adriano Chio, and Bryan J Traynor, "Projected increase in Amyotrophic Lateral Sclerosis from 2015 to 2040," *Nature communications*, vol. 7, pp. 12408, 2016.

[2] A Nalini, K Thennarasu, M Gourie-Devi, Sandhya Shenoy, and Dinkar Kulshreshtha, "Clinical characteristics and survival pattern of 1153 patients with Amyotrophic Lateral Sclerosis: experience over 30 years from India," *Journal of the Neurological Sciences*, vol. 272, no. 1-2, pp. 60–70, 2008.

[3] 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.

[4] Adriano Chiò, Giancarlo Logroscino, Orla Hardiman, Robert Swingler, Douglas Mitchell, Ettore Beghi, Bryan G Traynor, Eurals Consortium, et al., "Prognostic factors in ALS: a critical review," *Amyotrophic Lateral Sclerosis*, vol. 10, no. 5-6, pp. 310–323, 2009.

[5] 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.

[6] Jesse M Cedarbaum, Nancy Stambler, Errol Malta, Cynthia Fuller, Dana Hilt, Barbara Thurmond, Arline Nakanishi, and BDNF ALS Study Group (Phase III), "The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function," *Journal of the neurological sciences*, vol. 169, no. 1-2, pp. 13–21, 1999.

[7] RD Kent, Robert L Sufit, John C Rosenbek, Jane F Kent, Gary Weismer, Ruth E Martin, and Benjamin R Brooks, "Speech deterioration in Amyotrophic Lateral Sclerosis: a case study," *Journal of Speech, Language, and Hearing Research*, vol. 34, no. 6, pp. 1269–1275, 1991.

[8] Aravind Illa, Deep Patel, BK Yamini, N Shivashankar, Preethish-Kumar Veeramani, Kiran Polavarapui, Saraswati Nashi, Atchayaram Nalini, and Prasanta Kumar Ghosh, "Comparison of speech tasks for automatic classification of patients with amyotrophic lateral sclerosis and healthy subjects," in *Proceedings ICASSP*. IEEE, 2018, pp. 6014–6018.

[9] Lanny J Haverkamp, Vicki Appel, and Stanley H Appel, "Natural history of Amyotrophic Lateral Sclerosis in a database population validation of a scoring system and a model for survival prediction," *Brain*, vol. 118, no. 3, pp. 707–719, 1995.

[10] Barbara Tomik and Roberto J Guiloff, "Dysarthria in Amyotrophic Lateral Sclerosis: a review," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 1-2, pp. 4–15, 2010.

[11] RD Kent, "Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders," *Am. Journal of Speech-Language Pathology*, vol. 5, no. 3, pp. 7–23, 1996.

[12] Fritz Buchthal, Christian Guld, and Poul Rosenfalck, "Action potential parameters in normal human muscle and their dependence on physical variables.," *Acta physiologica scandinavica*, vol. 32, no. 2-3, pp. 200–218, 1954.

[13] Vipin K Mishra, Varun Bajaj, and Anil Kumar, "Classification of normal, ALS, and myopathy EMG signals using ELM classifier," in *Proceedings AEEICB*. IEEE, 2016, pp. 455–459.

[14] Constantinos S Pattichis and Andreas G Elia, "Autoregressive and cepstral analyses of motor unit action potentials," *Medical Engineering & Physics*, vol. 21, no. 6-7, pp. 405–419, 1999.

[15] Hajime Hirose, Shigeru Kiritani, and Masayuki Sawashima, "Patterns of dysarthric movement in patients with Amyotrophic Lateral Sclerosis and pseudobulbar palsy," *Folia Phoniatrica et Logopaedica*, vol. 34, no. 2, pp. 106–112, 1982.

[16] Jordan R Green, Christopher A Moore, and Kevin J Reilly, "The sequential development of jaw and lip control for speech," *Journal of Speech, Language, and Hearing Research*, vol. 45, no. 1, pp. 66–79, 2002.

[17] RD Kent, JF Kent, G Weismer, RE Martin, RL Sufit, BR Brooks, and JC Rosenbek, "Relationships between speech intelligibility and the slope of second-formant transitions in dysarthric subjects," *Clinical Linguistics & Phonetics*, vol. 3, no. 4, pp. 347–358, 1989.

[18] RD Kent, Jane F Kent, Gary Weismer, Robert L Sufit, John C Rosenbek, Ruth E Martin, and Benjamin R Brooks, "Impairment of speech intelligibility in men with Amyotrophic Lateral Sclerosis," *Journal of Speech and Hearing Disorders*, vol. 55, no. 4, pp. 721–728, 1990.

[19] Barbara Tomik, Jerzy Krupinski, Lidia Glodzik-Sobanska, Maria Bala-Slodowska, Wieslaw Wszolek, Monika Kusiak, and Anna Lechwacka, "Acoustic analysis of dysarthria profile in ALS patients," *Journal of the neurological sciences*, vol. 169, no. 1-2, pp. 35–42, 1999.

[20] PP Gomez, D Palacios, A Gomez, V Rodellar, and AR Londral, "Articulation acoustic kinematics in ALS speech," in *Proceedings IWOBI*. IEEE, 2017, pp. 1–6.

[21] B. Yamini, N. Shivashankar, and A. Nalini, "Vowel space area in patients with amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis*, vol. 9, no. 1, 2008.

[22] B.K. Yamini and A. Nalini, "Measures of maximum performance of speech-related tasks in patients with amyotrophic lateral sclerosis," *Amyotrophic Lateral Sclerosis*, vol. 9, 2008.

[23] Pedro Gómez-Vilda, Ana Rita M Londral, Victoria Rodellar-Biarge, José M Ferrández-Vicente, and Mamede de Carvalho, "Monitoring amyotrophic lateral sclerosis by biomechanical modeling of speech production," *Neurocomputing*, vol. 151, pp. 130–138, 2015.

[24] Taylor Spangler, NV Vinodchandran, Ashok Samal, and Jordan R Green, "Fractal features for automatic detection of dysarthria," in *EMBS International Conference on Biomedical Health Informatics (BHI)*. IEEE, 2017, pp. 437–440.

[25] 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.

[26] "Zoom h6 capsules," https://www.zoom-na.com/products/field-video-recording///field-recording/h6-handy-recorder/capsulesxy, 2016.

[27] Joseph R Duffy and Keith A Josephs, "The diagnosis and understanding of apraxia of speech: Why including neurodegenerative etiologies may be important," *Journal of Speech, Language, and Hearing Research*, 2012.

[28] Cyndi Stein-Rubin and Renee Fabus, *A guide to clinical assessment and professional report writing in speech-language pathology*, Nelson Education, 2011.

[29] Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlicek, Yanmin Qian, Petr Schwarz, et al., "The kaldi speech recognition toolkit," in *Proceedings ASRU*. IEEE, 2011.

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

[31] Joseph P Campbell, Douglas A Reynolds, and Robert B Dunn, "Fusing high-and low-level features for speaker recognition," in *Eighth European Conference on Speech Communication and Technology*, 2003.

[32] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.

[33] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.