

Bulbar ALS Detection Based on Analysis of Voice Perturbation and Vibrato

Maxim Vashkevich*, Alexander Petrovsky*

Yuliya Rushkevich†

*Belarusian State University of Informatics and Radioelectronics
vashkevich@bsuir.by

†Republican Research and Clinical Center of Neurology and Neurosurgery

Abstract—On average the lack of biological markers causes a one year diagnostic delay to detect amyotrophic lateral sclerosis (ALS). To improve the diagnostic process an automatic voice assessment based on acoustic analysis can be used. The purpose of this work was to verify the suitability of the sustain vowel phonation test for automatic detection of patients with ALS. We proposed enhanced procedure for separation of voice signal into fundamental periods that requires for calculation of perturbation measurements (such as jitter and shimmer). Also we proposed method for quantitative assessment of pathological vibrato manifestations in sustain vowel phonation. The study's experiments show that using the proposed acoustic analysis methods, the classifier based on linear discriminant analysis attains 90.7% accuracy with 86.7% sensitivity and 92.2% specificity.

Index Terms—voice pathology identification, amyotrophic lateral sclerosis, voice assessment.

I. INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is neurodegenerative incurable disease, approximately 50% of patients with ALS die within 30 months of symptom onset [1]. There is no biological marker for ALS and the diagnosis is made based on constellation of clinical observations. On average, it takes over a year to arrive at the diagnosis [2]. Using an acoustic analysis of voice and speech is a promising way to improve the process of ALS detection and monitoring of disease progression [3]. It becomes possible since bulbar motor changes (i.e., difficulty with speech or swallowing) are the first symptoms in approximately 30% of persons with ALS [4]. Almost all people with ALS display bulbar symptoms at later stages. It is important that abnormal acoustic parameters of voice were also demonstrated in ALS subjects with perceptually normal vocal quality on sustained phonation [5].

The need for development voice and speech diagnostic tools has motivated recent work on elaboration of special acoustic analysis methods. One important aspect of current research related to neurological disease detection (such as Parkinson's and ALS) is that they targeted to use smartphones and tablets to record the voice using standard microphone in the different home environments [6, 7]. Advancement in this direction can lead to creation of easy to use self- and tele-monitoring and disease detection means.

In previous studies different approaches and speaking tasks are used for ALS detection. One general approach that tends to solve the problem of differential diagnosis consists in classification of dysarthria type [8, 9]. The main difficulty of this approach is that it requires the collection of representative

dataset for all types of dysarthria. More simple approach to ALS detection is based on using dataset containing pathological and normal voice/speech samples [4, 7, 10, 11]. In several studies detection of ALS is performed using kinematic sensors [4, 12] to model articulation and measure prosodic elements such as vowel duration or speaking rate [2]. *Running speech* test was used in [7, 10, 11] as a basis for ALS detection. In [10] representational learning approach based on convolutional neural networks (CNN) was applied for ALS detection. As a low-level features filterbank energies with its 1st and 2nd derivatives were used. The best sample-level accuracy was obtained by time-domain CNN (71.6% sensitivity and 80.9% specificity). In [7] huge feature set obtained with openSMILE toolkit was analysed, the mel-frequency cepstral coefficients (MFCCs) and features based on RASTA filtering were found the most informative. However, the support vector machine (SVM) classifier based on linear kernel presented in [7] has moderate accuracy (79% for males and 83% for females). In [11] joint analysis of the vowels /a/ and /i/ extracted from running speech test was used for ALS detection. Classifier based on linear discriminant analysis (LDA) reported in [11] has 88% accuracy (90.5% sensitivity and 84.6% specificity). A *diadochokinetic task* (rapid repetitions of syllables) was used for automatic detection of ALS in [4]. Fractal jitter introduced in [4] along with MFCCs and articulatory features allows to obtain classifier based on extreme gradient boosting (XGBoost) with 90.2% accuracy (94.5% sensitivity and 85.1% specificity). Our work extends prior efforts in following aspects, it 1) verifies the suitability of the most simple *sustain vowel phonation* test (to the best knowledge of the authors there is no works dedicated to ALS detection based on sustained phonation test), 2) adapts acoustic analysis methods to work with non-studio voice recordings obtained using smartphone, 3) proposes new acoustic feature called pathological vibrato index (PVI) that considerably increases the detection accuracy.

II. VOICE PERTURBATION ANALYSIS

The perturbation measures jitter and shimmer and their variants are often use to assess the voice function [13]. All of them require preliminary segmentation of the voice signal into fundamental periods. This problem is very similar to f_0 extraction but has its own specifics.

A. F_0 extraction based on waveform matching

In perturbation analysis it is conventional to use event detection methods (peak-picking or waveform matching) for f_0 extraction [13, 14]. The main reason of this is that objective of perturbation analysis is cycle-to-cycle variation of vocal fold vibration and therefore short-term average methods (when several cycles are analysed) are not applicable. In waveform matching (WM) method entire waveshapes are matching across adjacent cycles. The difference between periods is decided by a least square difference between adjacent cycles [14]. Currently waveform matching is concerned as the method that allows to obtain good results for voice perturbation analysis [15]. In the widely used PRAAT voice analysis system there is an implementation of the WM method [15, 16].

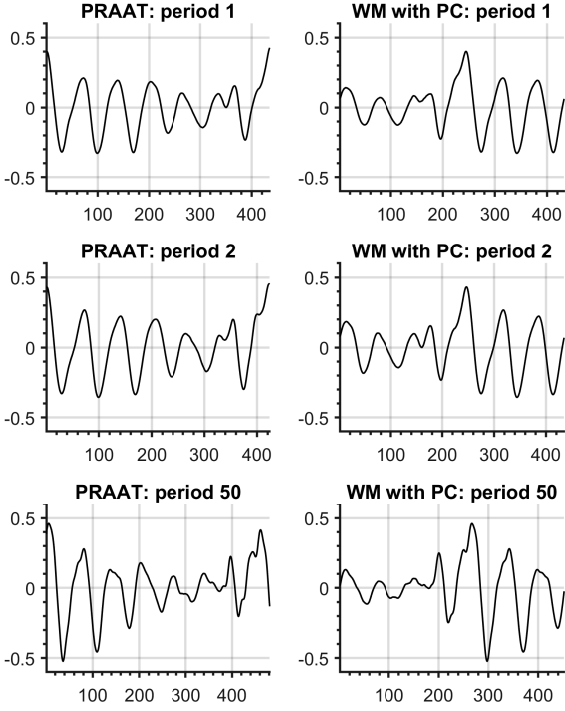


Fig. 1: Phonatory cycles of pathological voice extracted using PRAAT and WM with phase constrain

Waveform matching method exploit the fact that current period is closely related to previous period. The drawback of this approach is that if an error appears in detecting current period it will affect all following periods. In practice it results in phase shift if we compare cycles distant one from another. The example of this phenomenon can be seen in Fig. 1. Period #1 is almost the same as period #2, however if we compare it with period #50 the phase shift can be seen. The risk of errors in the WM method increases in the case of using non-studio recordings obtained using smartphone, that can contain environmental noise fluctuation. In this paper we propose a method aimed at eliminating these drawbacks.

B. Waveform matching with phase constrain

The proposed procedure starts by pitch estimation $f_o(n)$ using any general purpose method. The estimated f_o contour is used as a “reference” signal for further period detection. In this work we used IRAPT estimator [17] on this stage. The first period is detected using cumulative sum

$$\Phi(n) = \sum_{k=1}^n \omega_o(k), \quad (1)$$

where

$$\omega_o(k) = 2\pi f_o(k)/F_s,$$

where $f_o(k)$ – fundamental frequency contour in Hz, F_s – sampling frequency. Starting with $n = 1$ and incrementing n

function (1) is calculated until $\Phi(n)$ exceeds 2π , after that first period is set equal to $n - 1$.

The following periods are detected using two-stage procedure: 1) detecting approximate period boundary using phase function (1) and 2) refining of boundary such that mean absolute error between two adjacent waveforms is minimal. The Matlab-implementation of the proposed method is given below.

```
function T0 = WM_phase_const(s, w0)
% Initialization
Ln = length(s);
Phi_n = 0; % phase from eq.(1)
n = 0; % sample index
Nc = 1; % current cycle number

% Searching first period
while (Phi_n < 2*pi)
    n = n + 1;
    Phi_n = Phi_n + w0(n);
end

T0(Nc) = n - 1;
T0_prev = s(1:T0(Nc));
cur_T0 = 1; % length of current period
Phi_n = Phi_n - 2*pi;

% Search cycle
while (n < (Ln - T0(Nc)))
    if (Phi_n > 2*pi)
        cur_T0 = cur_T0 - 1;
        % find better offset
        offset = round(cur_T0*0.1);
        b = -offset:offset; % bias
        for k=1:length(b)
            T0_cur = s(n-1-T0(Nc)+1+b(k):n-1+b(k));
            err(k) = mean(abs(T0_prev - T0_cur));
        end
        [~,min_ind] = min(err);
        n = n - 1 + b(min_ind);
        Nc = Nc + 1;
        T0(Nc) = cur_T0 + b(min_ind);
        T0_prev = s(n-T0(Nc)+1:n);

        cur_T0 = 0; % Reset running variables
        Phi_n = 0;
    else
        n = n + 1;
        Phi_n = Phi_n + w0(n);
        cur_T0 = cur_T0 + 1;
    end
end
end
```

An example of application of WM method with phase constrain to pathological voice analysis is given in Fig. 1. It can be seen that cycles extracted using proposed method are more phase synchronised compared to conventional WM.

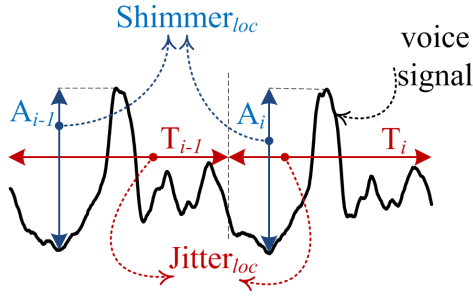


Fig. 2: Jitter and Shimmer concepts illustration

C. Amplitude and frequency perturbation

1) *Jitter*: Jitter (frequency or period perturbation) is the variability of the fundamental period from one cycle to the next (see Fig. 2). As far as jitter estimates short-term variation it can not be accounted to voluntary changes in f_o . Therefore jitter is intended to provide an index of the stability of the phonatory system. High level of jitter results from diminished neuromotor and aerodynamic control [13]. The jitter has been used as an indicator of the voice quality that characterizes the severity of dysphonia [18]. Mathematically jitter (J_{loc}) is defined as average difference between consecutive periods, divided by the average period:

$$J_{loc} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_0(i) - T_0(i+1)|}{\frac{1}{N} \sum_{i=1}^N T_0(i)} \times 100, \quad (2)$$

where N is the number of extracted periods.

Jitter can be also estimated as relative average perturbation (RAP) to quantify whether the period duration is smooth over three adjacent cycles [13]:

$$J_{rap} = \frac{\frac{1}{N-1} \sum_{i=2}^{N-2} \left| T_0(i) - \frac{1}{3} \sum_{k=i-1}^{i+1} T_0(k) \right|}{\frac{1}{N} \sum_{i=1}^N T_0(i)} \times 100. \quad (3)$$

In this study period perturbation quotient (PPQ) is also used to quantify the variability of the of the pitch period evaluated in five consecutive cycles:

$$J_{ppq5} = \frac{\frac{1}{N-4} \sum_{i=3}^{N-2} \left| T_0(i) - \frac{1}{5} \sum_{k=i-2}^{i+2} T_0(k) \right|}{\frac{1}{N} \sum_{i=1}^N T_0(i)} \times 100. \quad (4)$$

2) *Shimmer*: Shimmer is a measure that characterizes the extent of variation of expiratory flow during the phonation. Basic shimmer measure (S_{loc}) is defined as average absolute difference between the amplitude of consecutive periods, di-

vided by the average amplitude [13]:

$$S_{loc} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A(i) - A(i+1)|}{\frac{1}{N} \sum_{i=1}^N A(i)} \times 100, \quad (5)$$

where $A(i)$ is the amplitude of the i -th period. S_{loc} fall under influence of long-term changes in vocal intensity. To eliminate the effects of amplitude “drift” in order to get a truer index of underlying shimmer it has been suggested to measure amplitude perturbation quotient (APQ):

$$S_{apqL} = \frac{\frac{1}{N-L+1} \sum_{i=1+\frac{L-1}{2}}^{N-\frac{L-1}{2}} \left| A(i) - \frac{1}{L} \sum_{k=i-\frac{L-1}{2}}^{i+\frac{L-1}{2}} A(k) \right|}{\frac{1}{N} \sum_{i=1}^N A(i)} \times 100. \quad (6)$$

where $A(i)$ is the amplitude of the i -th period. The most commonly used window size L are 3, 5 and 11 [19].

D. Vibrato analysis

Vibrato is a rapid, and regular fluctuation of the f_o that arises during sustained vowel phonation. A direct comparison ensures that there is a significant difference between the vibrato frequency in patients with ALS and healthy controls.

The estimation of the extent of pathological changes in vibrato is based on the observation that for healthy voices vibrato lies in range of 5-8 Hz [20] while for ALS patients characterized by presence of high-frequency components in 9-14 Hz range [21]. In this study we use the following method of estimating *pathological vibrato index* (PVI):

- 1) Estimation of $f_o(m)$ contour with 5 ms time step using IRAPT algorithm [17];
- 2) Normalization of f_o contour:

$$f'_o(m) = \frac{f_o(m)}{\text{mean}(f_o)};$$

- 3) Bandpass filtering of $f'_o(m)$ using 3-th order Butterworth IIR with pass band [9, 14] Hz;
- 4) Amplitude spectrum $A_{f_o}(f)$ estimation using Welch’s method with windows of 1 sec length and 95% overlap;
- 5) Calculation of pathological vibrato index:

$$PVI = \sum_{f \in [9, 14] \text{ Hz}} A_{f_o}(f). \quad (7)$$

Fig. 3 shows the steps of the PVI calculation for a healthy speaker and ALS patient with moderate bulbar dysfunction.

III. CLASSIFICATION

For discriminating between the two classes of normal and pathological cases two widely spread machine-learning algorithms: linear discriminant analysis (LDA) and k-Nearest Neighbors (k-NN) were used.

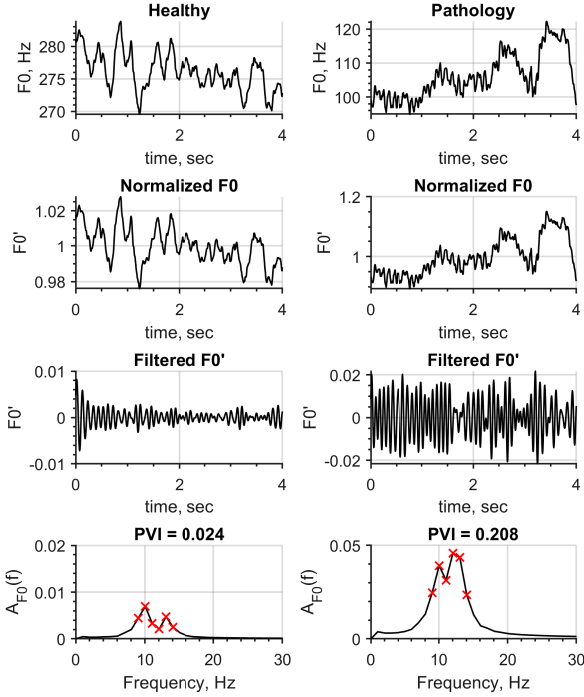


Fig. 3: Left column: normal case, right column: pathological case. Rows from the top: Extracted f_o , normalized f_o contour, IIR filtered f_o contour, amplitude spectrum $A_{f_o}(f)$, amplitudes used for PVI calculation are indicated by red x-marks

A. Linear discriminant analysis

The LDA is an approach which uses linear hyperplane (decision surface) to separate the dataset into two classes [22]. In LDA the following classification function is used

$$f(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \mathbf{x} \rangle + b),$$

where \mathbf{x} – feature vector, b – bias and \mathbf{w} – weight vector that defines decision surface and is chosen according to Fisher criterion to maximize the quotient

$$J(\mathbf{w}) = \frac{(\mu_{\mathbf{w}}^+ - \mu_{\mathbf{w}}^-)^2}{(\sigma_{\mathbf{w}}^+)^2 + (\sigma_{\mathbf{w}}^-)^2},$$

where $\mu_{\mathbf{w}}^+$ and $\mu_{\mathbf{w}}^-$ are the mean of the projections of the positive and negative samples onto direction \mathbf{w} , $\sigma_{\mathbf{w}}^+$ and $\sigma_{\mathbf{w}}^-$ are the corresponding standard deviations.

B. k-NN classifier

The idea of k-NN classification consists in assigning to a new data sample the label by taking into account K nearest samples in the training set. In this work the following approach was used: for new data sample K nearest points $\mathbf{x}_{1...K}^+$ form positive class and K nearest points $\mathbf{x}_{1...K}^-$ from negative class of training set were determined. Then simple voting technique based on *distances weighting* [23] is applied to assign the label

to the input sample \mathbf{x} :

$$f(\mathbf{x}) = \text{sign}\left(\sum_{k=1}^K \frac{1}{d(\mathbf{x}, \mathbf{x}_k^+)} + \sum_{k=1}^K \frac{-1}{d(\mathbf{x}, \mathbf{x}_k^-)}\right).$$

In this work we used $K = 3$. For finding nearest neighbors a Mahalanobis distance was used:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \Sigma^{-1} (\mathbf{x} - \mathbf{y})},$$

where Σ^{-1} is inversion of covariance matrix calculated from training set. The choice of Mahalanobis distance is explained by the fact that components of feature vector \mathbf{x} have non-zero correlation.

C. Cross-validation process

All classification experiments are performed using k -fold cross-validation (CV) method [23]. Procedure starts with random mixing of the dataset, then it is divided into k equal parts and during the following k rounds one of the parts used as a test set and the remaining as training set. We used $k=7$. After all rounds (at each round new classifier is trained) the dataset appears to be separated into two classes and accuracy of the ALS detector was assessed. CV process was repeated 40 times, then mean and standard deviation values for the performance metrics of classifier were calculated.

In order to measure the performance of the classifiers accuracy, sensitivity, specificity and averaged recall were calculated [23]:

$$\begin{aligned} \text{Acc} &= \frac{TP + TN}{TP + FP + FN + TN} \\ \text{Sens} &= \frac{TP}{TP + FN} \\ \text{Spec} &= \frac{TN}{TN + FP} \\ \text{R}_{\text{avg}} &= \frac{1}{2} (\text{Sens} + \text{Spec}) \end{aligned} \quad (8)$$

where TP , TN , FP , FN – the number of true positive, true negative, false positive and false negative results of detection, respectively. In this case, positive means a prediction that the voice sample is produced by a speaker with ALS.

IV. EXPERIMENTS

A. Dataset description

The voice data used in this study was collected in Republican Research and Clinical Center of Neurology and Neurosurgery (Minsk, Belarus). A total of 54 speakers were recorded, with 39 healthy speakers (23 males, 16 females) and 15 ALS patients with signs of bulbar dysfunction (6 males, 9 females). The average age in the healthy group was 41.9 years (SD 16.3, Min 18, Max 82) and the average age in the ALS group was 57.7 years (SD 9.0, Min 40, Max 70). All the participants were asked to produce the sustained vowel /a/ at a comfortable pitch and loudness as constant and long as possible. The phonation was performed on one breath. The voice signals were recorded using smartphone with a headset (sample rate 44.1 kHz) and stored as 16 bit uncompressed

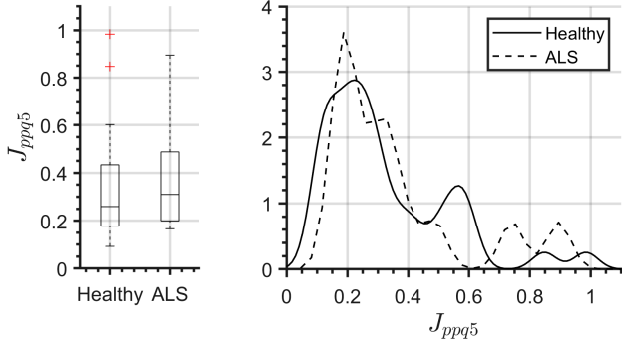


Fig. 4: Box plot and probability densities of jitter measure

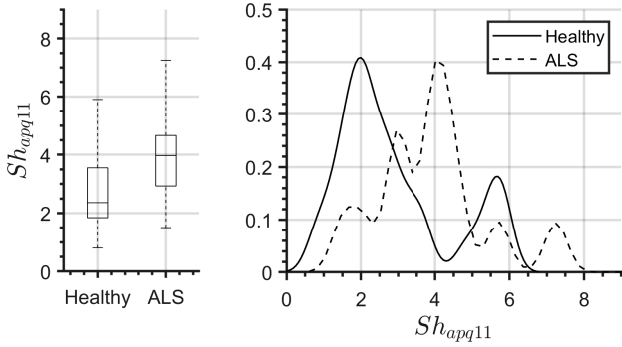


Fig. 5: Box plot and probability densities of shimmer measure

PCM files. The records were manually edited to remove the beginning and ending of each utterance, removing the onset and offset effects in these parts of each utterance. Average duration of the samples is 4.1 s.

Given that mean age for healthy group was about 17 years younger than ALS patients, we applied linear regression technique to remove age effect using the data of the healthy group. The correction was applied to the data of healthy speakers and ALS patients.

The voice database and Matlab tools used for the voice analysis are available in public GitHub repository¹.

B. Feature statistics

In order to visualize the parameters of voice in HC and ALS groups we calculate statistics for several features. Figure 4 shows the box plot and distribution estimated using the Gaussian kernel density method for jitter measure. It can be seen that in ALS voices does not observed significant increasing of jitter compared to HC.

Fig. 5 represents the box plot and kernel density estimation for shimmer measure that shows significant difference between healthy and ALS groups. In Fig. 6 the same statistics for pathology vibrato index (PVI) is given. The PVI shows good separation capability. A visual analysis of similar graphs for jitter parameters did not revealed a significant difference between the ALS patients voices compared to HC.

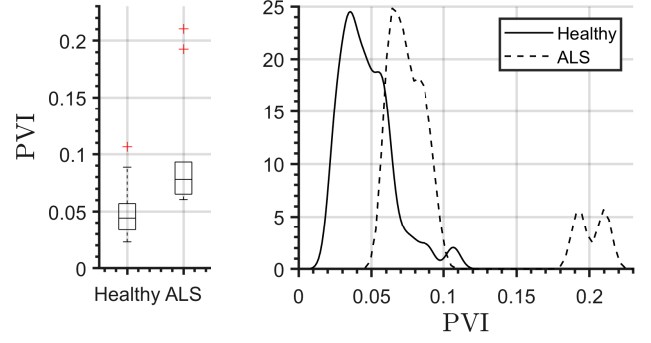


Fig. 6: Box plot and probability densities of PVI

C. Classification results and discussion

We performed ALS classification for voice samples using following three feature sets: 1) shimmerers/jitters extracted using PRAAT [16]; 2) shimmerers/jitters extracted using the proposed method; 3) the same as in previous case but with PVI.

Since the dimensionality D of the feature sets was low ($D=7$ for 1st and 2nd and $D=8$ for 3rd) we have performed exhaustive search through all possible feature subset to find the best one.

In the tables I and II for we used following notation: J_1 for J_{loc} , J_3 for J_{rap} , J_5 for J_{ppq} , S_1 for S_{loc} , S_L for S_{apqL} .

Table I shows the results of classification based on LDA². For each feature set we show three best feature subsets. The results were ranked according to R_{avg} measure because our goal was to get classifier with both high specificity and sensitivity. It can be seen that best results are obtained for third feature set, that contained PVI.

TABLE I: LDA based classification results

Features	R_{avg} (%)	Acc (%)	Sens (%)	Spec (%)
<i>PRAAT features</i>				
$[S_5 S_{11}]$	83.1	81.1 ± 1.9	83.0 ± 5.6	83.1 ± 1.4
$[J_3 J_5 S_1 S_5 S_{11}]$	81.4	82.8 ± 2.0	78.3 ± 4.7	84.6 ± 2.6
$[J_1 J_5 S_1 S_5 S_{11}]$	81.2	83.0 ± 7.8	77.3 ± 5.8	85.1 ± 2.8
<i>Features based on WM with PC</i>				
$[J_1 J_5 S_3 S_{11}]$	86.0	86.4 ± 2.3	85.0 ± 3.3	87.0 ± 2.6
$[J_5 S_1 S_3 S_{11}]$	84.9	85.0 ± 1.9	84.7 ± 3.8	85.1 ± 1.9
$[J_3 J_5 S_5 S_{11}]$	84.4	86.5 ± 1.6	79.8 ± 1.8	89.0 ± 2.0
<i>Features based on WM with PC + PVI</i>				
$[S_1 S_3 S_{11} PVI]$	89.5	90.7 ± 1.7	86.7 ± 0.1	92.2 ± 2.3
$[J_5 S_1 S_3 S_{11} PVI]$	88.0	89.0 ± 2.2	85.8 ± 2.7	90.2 ± 3.0
$[J_1 S_1 S_3 S_{11} PVI]$	88.0	88.9 ± 2.2	86.0 ± 2.0	90.0 ± 2.8

The classifiers based on the PRAAT features have lower characteristics than classifiers that used jitters and shimmerers extracted with proposed method.

Table II shows the results of classification based on k-NN approach.

²Accuracy, sensitivity and specificity are given in the form mean \pm standard deviation and are out of sample computed using 7-fold cross validation with 40 iteration. Averaged recall is calculated using (8) by taking mean sensitivity and specificity.

¹<https://github.com/Mak-Sim/Troparion/tree/master/SPA2019>

TABLE II: kNN based classification results

Features	R _{avg} (%)	Acc (%)	Sens (%)	Spec (%)
<i>PRAAT features</i>				
[J ₃ S ₁ S ₃ S ₅]	76.0	84.6±2.2	56.8±4.8	95.3±2.1
[J ₃ S ₁ S ₁ S ₃]	75.8	84.2±1.7	56.8±4.3	94.7±2.0
[J ₁ J ₃ S ₁ S ₁₁]	75.3	85.2±1.0	53.2±1.1	97.5±1.5
<i>Features based on WM with PC</i>				
[J ₁ J ₃ J ₅ S ₅ S ₁₁]	81.0	87.3±2.1	66.8±4.9	95.1±2.2
[J ₃ J ₅ S ₅ S ₁₁]	80.9	86.1±2.7	69.3±6.9	92.6±2.8
[J ₃ J ₅ S ₁ S ₃ S ₁₁]	80.2	86.3±2.4	66.7±6.4	93.8±2.2
<i>Features based on WM with PC + PVI</i>				
[J ₁ J ₅ PVI]	86.9	91.6±2.3	76.3±5.8	97.5±1.7
[J ₃ PVI]	86.5	90.5±1.5	77.7±5.1	95.4±1.0
[J ₁ PVI]	86.3	90.3±1.1	77.2±3.3	95.4±1.3

A common feature of presented k-NN classifiers is that their specificity is much greater than sensitivity. Though the accuracy of the best k-NN classifier (91.6%) is greater than best LDA classifier (90.7%), the average recall of k-NN classifier (86.9%) is considerably lower than in best LDA classifier (89.5%). In general, the low sensitivity of the k-NN classifier makes them less attractive for practical use.

V. CONCLUSIONS

This study presented improved voice analysis method for perturbation parameters estimation and new feature (PVI) for detecting bulbar dysfunction in ALS speakers. PVI makes quantitative assessment of pathological changes of vibrato. Comparing data in tables I and II clearly indicates that adding PVI to voice perturbation parameters significantly improves the classification accuracy. The best achieved result is 90.7% accuracy (86.7% sensitivity and 92.2% specificity).

The presented voice analysis methods may also be useful in discerning different levels of ALS severity. Future work will include other dysarthric-related disease to exclude a potential bias caused by common feature of dysarthric speech and focus on specific detection of ALS disease.

REFERENCES

- [1] C. Ingre, P. M. Roos, F. Piehl, and F. Kamel, "Risk factors for amyotrophic lateral sclerosis," *Clinical epidemiology*, vol. 7, pp. 181–193, 2015.
- [2] Y. Yunusova, J. S. Rosenthal, J. R. Green, S. Shellikeri, P. Rong, J. Wang, and L. H. Zinman, "Detection of bulbar ALS using a comprehensive speech assessment battery," in *Proc. of the International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, 12 2013, p. 217–220.
- [3] R. L. Horwitz-Martin, T. F. Quatieri, A. C. Lammert, J. R. Williamson, Y. Yunusova, E. Godoy, D. D. Mehta, and J. R. Green, "Relation of automatically extracted formant trajectories with intelligibility loss and speaking rate decline in amyotrophic lateral sclerosis," in *Proc. of Interspeech*, 2016, pp. 1215–1219.
- [4] T. Spangler, N. V. Vinodchandran, A. Samal, and J. R. Green, "Fractal features for automatic detection of dysarthria," in *Proc. of IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2017, pp. 437–440.
- [5] B. Tomik and R. J. Guiloff, "Room reverberation effects in hearing aid feedback cancellation," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 1-2, pp. 4–15, 2010.
- [6] A. Benba, A. Jilbab, and A. Hammouch, "Discriminating between patients with parkinson's and neurological diseases using cepstral analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 10, pp. 1100–1108, 2016.
- [7] R. Noirel, M. Pietrowicz, C. Agurto, S. Rishoni, and G. Cecchi, "Detection of amyotrophic lateral sclerosis (ALS) via acoustic analysis," in *Proc. of Interspeech*, 2018, pp. 377–381.
- [8] E. Castillo Guerra and D. F. Lovey, "A modern approach to dysarthria classification," in *Proc. of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEMBS)*, vol. 3, 2003, pp. 2257–2260.
- [9] M. V. Mujumdar and R. F. Kubichek, "Design of a dysarthria classifier using global statistics of speech features," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2010, pp. 582–585.
- [10] K. An, M. Kim, K. Teplansky, J. Green, T. Campbell, Y. Yunusova, D. Heitzman, and J. Wang, "Automatic early detection of amyotrophic lateral sclerosis from intelligible speech using convolutional neural networks," in *Proc. of Interspeech*, 2018, pp. 1913–1917.
- [11] M. Vashkevich, E. Azarov, A. Petrovsky, and Y. Rushkevich, "Features extraction for the automatic detection of ALS disease from acoustic speech signals," in *Proc. of Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, 2018, pp. 321–326.
- [12] A. Bandini, J. Green, L. Zinman, and Y. Yunusova, "Classification of bulbar als from kinematic features of the jaw and lips: Towards computer-mediated assessment," in *Proc. of Interspeech*, 2017, pp. 1819–1823.
- [13] R. J. Bakem and R. F. Orlikoff, *Clinical measurement of speech and voice, 2nd Edition*. Singular Thomson Learning, 2000.
- [14] I. R. Titze and H. Liang, "Comparison of F0 extraction methods for high-precision voice perturbation measurements," *Journal of Speech and Hearing Research*, vol. 36, no. 6, pp. 1120–1133, 1993.
- [15] P. Boersma, "Should jitter be measured by peak picking or by waveform matching," *Folia Phoniatrica et logopaedica*, vol. 61, no. 2, pp. 305–308, 2009.
- [16] —, "PRAAT, a system for doing phonetics by computer," *Glott International*, vol. 5, no. 9/10, pp. 341–345, 2002.
- [17] E. Azarov, M. Vashkevich, and A. Petrovsky, "Instantaneous pitch estimation based on RAPT framework," in *Proc. of the 20th European Signal Processing Conference (EUSIPCO)*, Aug 2012, pp. 2787–2791.
- [18] I. C. Miller and M. Moerman, "Voice therapy assistant: a useful tool to facilitate therapy in dysphonic patients."

- in *Proc. of the 8th International Workshop: Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, Dec. 2013.
- [19] J. R. Orozco-Arroyave, F. Honig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, and E. Nöth, “Automatic detection of parkinson’s disease in running speech spoken in three different languages,” *The Journal of the Acoustical Society of America*, vol. 139, no. 1, pp. 481–500, 2016.
- [20] T. Nakano, M. Goto, and Y. Hiraga, “An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features,” in *Proc. of Interspeech*, 2006, pp. 1706–1709.
- [21] A. E. Aronson, W. S. Winholtz, L. O. Ramig, and S. R. Silber, “Rapid voice tremor, or “flutter,” in amyotrophic lateral sclerosis,” *Annals of Otology, Rhinology & Laryngology*, vol. 101, no. 6, pp. 511–518, 1992.
- [22] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
- [23] P. Flach, *Machine Learning: The art and science of algorithms that make sense of data*. United Kingdom: Cambridge University Press, 2012.