

Classification of ALS patients based on acoustic analysis of sustained vowel phonations

M. Vashkevich¹, Yu. Rushkevich²

Abstract

Amyotrophic lateral sclerosis (ALS) is incurable neurological disorder with rapidly progressive course. Common early symptoms of ALS are difficulty in swallowing and speech. However, early acoustic manifestation of speech and voice symptoms is very variable, that making their detection very challenging, both by human specialists and automatic systems. This study presents an approach to voice assessment for automatic system that separates healthy people from patients with ALS. In particular, this work focus on analysing of sustain phonation of vowels /a/ and /i/ to perform automatic classification of ALS patients. A wide range of acoustic features such as MFCC, formants, jitter, shimmer, vibrato, PPE, GNE, HNR, etc. were analysed. We also proposed a new set of acoustic features for characterizing harmonic structure of the vowels. Calculation of these features is based on pitch synchronized voice analysis. A linear discriminant analysis (LDA) was used to classify the phonation produced by patients with ALS and those by healthy individuals. Several algorithms of feature selection were tested to find optimal feature subset for LDA model. The study's experiments show that the most successful LDA model based on 32 features picked out by LASSO feature selection algorithm attains 99.7% accuracy with 99.3% sensitivity and 99.9% specificity. Among the classifiers with a small number of features, we can highlight LDA model with 5 features, which has 89.0% accuracy (87.5% sensitivity and 90.4% specificity).

Keywords: Voice pathology detection, amyotrophic lateral sclerosis, ALS, Acoustic analysis, voice quality

1. Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease involving the upper and lower motor neurons. There are two main forms of ALS which differ by onset: spinal form (first symptoms manifest in the arms and

¹Department of Computer Engineering, Belarusian State University of Informatics and Radioelectronics, 6 P.Brovky str., 220013, Minsk, Belarus

²Republican Research and Clinical Center of Neurology and Neurosurgery, 24 F. Skoriny Skoriny str., 220114, Minsk, Belarus

legs) and bulbar form (voice and/or swallowing difficulties are often the first symptoms). Progressive bulbar motor impairment due to ALS leads to deterioration in speech and swallowing function [1]. The abnormalities in speech production, phonation and articulation due to neurological disorders is referred to as *dysarthria*. Dysarthria develops in more than 80% of affected by ALS individuals at some point during the disease's course [2]. Currently the diagnosis of ALS is based on clinical observations of upper and lower motor neuron damage in the absence of other causes. Due to the lack of clinical diagnostic markers of ALS, the pathway to correct diagnosis in average takes 12 months [3].

During the last years objective evaluation of voice and speech has gained popularity as a means of detecting early signs of neurological diseases [4, 5, 6]. It can be explained by the fact that speech is accomplished through complex articulatory movements, requires precise coordination and timing and therefore is very sensitive to violations in the peripheral or central nervous system [7, 8]. Recent studies suggested that acoustic voice and speech analysis might provide useful biomarkers for diagnosis and remote monitoring of ALS patients [9, 10]. The advantage of using voice/speech signals is the capability of using smartphone or tablet for recording patients at home conditions without the logistical difficulty in a clinical environment [4, 11].

The main goal of this work is automatic detection of ALS patients with or without bulbar disorders (i.e. classification of healthy controls vs. patients with ALS) based on sustained vowel phonation (SVP) test. Our long-term aim is to build automated system for classification of neuromotor degenerative disorders based on analysis of SVP test. Therefore, we consider the problem of binary classification of the voice recording to be belonging to ALS patient or healthy person as a first step toward this aim. We chose sustained vowel phonation test among different diagnostic speech tasks due to its simplicity and wide spreading in medical practice. Besides all, recent research shows [12] that using SVP test it is possible to detect persons with Parkinson's disease. This give us hope that this test can be effective for ALS detection.

Sustained phonation is a common speech task used to evaluate the health of the phonatory speech subsystem [5]. By using SVP test the following characteristics of voice can be assessed: pitch, loudness, resonance, strain, breathiness, hoarseness, roughness, tremor, etc [4, 8, 13]. However, it can be argued that some of the vocal abnormalities in continuous speech might not be captured by use of sustained vowels, but the analysis of continuous speech is much more complex due to articulatory and other linguistic confounds [13]. One more argument is that the use of sustained vowels is commonplace in clinical practice [14]. Besides all this, early study [15] had been reported that abnormal acoustic parameters of the voice were demonstrated in ALS subjects with perceptually normal vocal quality on sustained phonation. Also in [16] it was reported that glottic narrowing due to vocal cord dysfunction (that can be assessed using SVP test) is one of ALS symptoms.

SVP test is widely used for detecting and diagnosing of different neurological diseases such as Parkinson's, Alzheimer, Dystonia and others [4, 5]. For example, it has been shown in [12] that classifier based on the features extracted form

SVP test allows one to discriminate Parkinson’s disease subjects from healthy controls (HC) with almost 99% overall classification accuracy. However, there are few studies dedicated to the detection of the ALS based on SVP test. In [8] SVP was used along with the other speech tests for dysarthria classification. Sustained phonation also was used for assessing laryngeal subsystem within a comprehensive speech assessment battery in [17]. But in the majority of prior works *running speech* test that consist in reading of specially-designed passage was used for ALS detection [9, 11, 18, 19]. In [10] rapid repetition of syllable (pa/ta/ka), which is often referred to as *diadochokinetic task* (DDK) was used for automatic ALS detection. Some studies use kinematic sensors to model articulation for ALS detection [20], however this approach is invasive in nature and less attractive compared to non-invasive speech test.

The purpose of this work is to investigate the possibility of designing a classifier for detection of patients with ALS based on the sustained phonation test. Traditionally, vowel /a/ is used in SVP test, however, in our study along with /a/ we have used vowel /i/. This decision is based on preliminary results of works [21, 22, 23], that provide evidence that information contained in these vowels might allow to obtain classifier with high performance. This work is based on the analysis of the sustained phonation of vowels /a/ and /i/, in contrast to other studies that extract vowels from running speech tests (see e.g. [22, 23]).

The remainder of the paper is organized as follows; Section 2 provides information about methods of acoustic analysis used for feature extraction. The voice data used in this study along with various methods of feature selection, classification and validation are presented in section 3. In section 4 we present the results of our findings and discuss the interpretation of them. Section 5 provides conclusion on the work.

2. Acoustic analysis

Bulbar system that is affected by ALS is considered as a part of the larger speech production network and comprises of four distinct subsystems [1]: respiratory, phonatory, articulatory, and resonatory. In this short review of acoustic features, we indicate which subsystem is described by each feature.

2.1. Perturbation measurements

2.1.1. Jitter

Jitter (i.e. frequency/period perturbation) is the measure of variability of fundamental period from one cycle to the next. As far as jitter estimates short-term variations it can not be accounted to voluntary changes in F0. Therefore jitter is intended to provide an index of the stability of the phonatory subsystem. High level of jitter results from diminished neuromotor and aerodynamic control [14]. The jitter has been used as an indicator of the voice quality that characterizes the severity of dysphonia [24]. In this study we have used following popular jitter measures [25]:

1) local jitter (J_{loc}) that is defined as average difference between consecutive periods, divided by the average period:

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

where $T_0(i)$ is the duration of i -th fundamental period in seconds, N is the number of extracted periods;

2) period perturbation quotient (J_{ppq}) to quantify the variability of pitch period evaluated in L consecutive cycles:

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

In this work, we used the parameter L equal to 3, 5 and 55.

2.1.2. Shimmer

Shimmer is an amplitude perturbation measure that characterize the extent of variation of expiratory flow during the phonation. This feature can be considered as characteristic of the respiratory subsystem. Basic shimmer measure (S_{loc}) is defined as average absolute difference between the amplitude of consecutive periods, divided by the average amplitude:

$$S_{loc} = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A(i) - A(i+1)|}{\frac{1}{N} \sum_{i=1}^N A(i)} \times 100, \quad (3)$$

where $A(i)$ is the amplitude of the i -th pitch period.

S_{loc} fall under influence of long-term changes in vocal intensity [14]. To eliminate the effects of amplitude “drift” and get a truer index of underlying shimmer it has been suggested to measure amplitude perturbation quotient (APQ) [25]. APQ quantify whether the amplitude of pitch period duration is smooth over L consecutive cycles:

$$S_{apqL} = \frac{\frac{1}{N-L+1} \sum_{i=1+\frac{L-1}{2}}^{N-\frac{L-1}{2}} \left| A(i) - \frac{1}{L} \sum_{k=i-\frac{L-1}{2}}^{i+\frac{L-1}{2}} A(k) \right|}{\frac{1}{N} \sum_{i=1}^N A(i)} \times 100, \quad (4)$$

Typically the parameter L takes value 3, 5, 11 or 55 [5, 6, 26]. We used all of those options in our study.

2.1.3. Directional perturbation factor

Directional perturbation factor (DFP) is a measure of perturbation that ignores the magnitude of period perturbation: it depends on the number of times that frequency changes shift direction [14]. The DFP calculation consists of two steps. At the first step the difference between adjacent fundamental periods is calculated:

$$\Delta T_0(i) = T_0(i) - T_0(i - 1). \quad (5)$$

At the second step the number of sign changes in sequence of $\Delta T_0(i)$ is calculated:

$$N_{\Delta\pm} = \frac{1}{2} \sum_{i=2}^N |\text{sign}(\Delta T_0(i)) - \text{sign}(\Delta T_0(i - 1))|.$$

Finally, DFP parameter is obtained as follows:

$$\text{DFP} = \frac{N_{\Delta\pm}}{N} \times 100, \quad (6)$$

where N is the total number of fundamental periods.

2.2. Noise measurements

The existence of noise energy, broadly understood as that outside of harmonic components during sustained phonation, is the result of incomplete closure of the vocal folds during the phonation, indicative of an interruption of the morphology of the larynx [26]. We used two different noise measurements: harmonic-to-noise ratio (HNR) [27] and glottal-to noise excitation ratio (GNE).

2.2.1. HNR

The HNR measures the ratio between periodic (or harmonic) component and non periodic (or noise) component of the voice signal. Sonorant and harmonic voices are characterized by high HNR values. A low HNR denotes that voice comprise increased amount of noise. For calculation of HNR we used mathematical background presented in [27]. At the beginning, for a voice signal a normalized autocorrelation function $AC_V(\tau)$ is calculated. Then, the first local maximum outside 0 (with corresponded lag τ_{max}) is searched. The normalized autocorrelation $AC_V(\tau_{max})$ represents the relative power of the periodic component of the signal (while full power $AC_V(0) = 1$). Finally, HNR is calculated as

$$\text{HNR} = 10 \log_{10} \frac{AC_V(\tau_{max})}{1 - AC_V(\tau_{max})}. \quad (7)$$

2.2.2. GNE

GNE measures the amount of excitation in voice due to the vibration of the vocal folds relative to the excitation noise due to the turbulence in the vocal tract [28]. The GNE is often associated with the breathiness [8, 29] and therefore can be considered as characteristics of phonatory subsystem.

Calculation of the GNE is based on the correlation between Hilbert envelopes of three different frequency channels [30]. Since full band signal simultaneously excited by a single glottis closure the envelopes in all channels have the same shape. This leads to high correlation between envelopes. However, in case of turbulent signals a narrowband noise is excited in each frequency channel. These narrow band noise signals are uncorrelated. Thus, interband correlation can be used to measure the amount of turbulence in a signal.

Calculation of the GNE factor is consist in the following steps:

1. Down sampling the signal to 8 kHz.
2. Divide signal into 30 ms overlapping frames with 10 ms hop size. For each frame execute steps 3–7.
3. Inverse filtering of the signal by calculating the linear prediction error signal, using a predictor of 10-th order estimated by the autocorrelation method [31] with Hamming window.
4. Calculating the Hilbert envelopes of three different frequency bands with 1000 Hz bandwidth and central frequencies at 500, 1500 and 2500 Hz.
5. Calculating the cross correlation function between every pair of envelopes for which the difference of their center frequency is equal or greater than half the bandwidth.
6. Pick the maximum of each correlation function.
7. The GNE for the current frame is equal to the maximum of the maximums obtained in step 6.
8. Compute the mean value of GNE and its standard deviation.

2.3. Spectral parameters

2.3.1. MFCC

Mel-Frequency Cepstral Coefficients (MFCCs) is the most widely used feature in speech-related applications such as speaker identification and recognition. Moreover, recent studies have shown promising results on the identification of voice pathology with MFCCs [4, 10, 12, 32]. MFCCs can detect subtle changes in the motion of the articulators (tongue, lips), which are known to be affected in many neurological diseases [12]. They have been used for detecting of hypernasality due to the velopharyngeal insufficiency in [33]. In [28] the usage of MFCCs is argued by its ability of modelling changes in the speech spectrum, especially around the first two formants (F1 and F2), where most of the energy of the signal is concentrated. The work [32] showed that MFCCs have an inherent ability to model an irregular movement of the vocal folds, or a lack of closure due to a change in the properties of the tissue covering vocal folds. Therefore MFCCs can be considered as parameters describing both resonatory and articulatory subsystems.

MFCC parameters [32, 31] are obtained by applying discrete cosine transform over the logarithm of the energy calculated in several mel-frequency bands:

$$\text{MFCC}(m) = \sum_{k=1}^M \ln S(k) \cos \left[m(k - 0.5) \frac{\pi}{M} \right]. \quad (8)$$

where M is the number of uniform frequency bands in the mel scale, $m = 1, 2, \dots, L$, and L is the order of the MFCC coefficients. The energy of frequency bands are calculated using N -point magnitude spectrum $X(j)$ of the frame of the voice signal:

$$S(k) = \sum_{j=1}^N W_k(j)X(j), \quad k = 1, 2, \dots, M, \quad (9)$$

where $W_k(j)$ is the triangular weighting function [31] associated with k -th band.

In our study we used $L = 12$ MFCC parameters that computed within windows of 25-ms length and 10-ms time shift. Magnitude spectrum $X(j)$ is calculated in the range [0; 4000] Hz and averaged within $M = 20$ uniform mel-frequency bands (see (9)). The first (Δ) derivatives of MFCC have also been calculated since they provide information about the dynamics of the time-variation in MFCC parameters. A priori, these features can be considered as significant because the disorder lowers stability of the voice signal [32]; therefore larger time-variations of the parameters may be expected in ALS voice relative to normal voice.

Because the MFCCs are a timeseries, we averaged the MFCCs across the time domain to consolidate them to a single set of coefficients. Finally 12 MFCCs and 12 Δ MFCCs are evaluated for each voice recording.

2.3.2. Formants

Changes of formant frequencies during the vowel phonation due to dysarthria have been reported in many studies [21, 34, 35, 36, 37]. The most frequently reported abnormalities of vowel production include: centralization of formant frequencies [38], reduction of the vowels space area [36], and abnormal formant frequencies for high vowels and front vowels [22, 34]. In [37] it was shown that in patients with ALS measurement of the F2 slope (or F2 transition) is correlated with overall speech intelligibility score. Also features derived from statistics of the first (F1) and second (F2) formant frequencies (and their trajectories) have shown good performance for predicting speaking rate decline in ALS [39]. Though SVP test cannot reflect the dynamics of formant frequency trajectories, we still can use the values of formant frequencies as source of information. In [23, 40] it was shown that the value of F2 for vowel /i/ appears to be a good feature for discriminating between patients with ALS and healthy control group. In this study we use second formant of vowel /i/ ($F2_i$) and Euclidean distance (convergence) between the vowels /i/ and /a/:

$$F2_{conv} = |F2_i - F2_a|. \quad (10)$$

Study [21] have shown that convergence of the F2 of vowels /i/ and /a/ is much stronger in speakers with dysarthria due to ALS, than in healthy speakers. Both features ($F2_i$ and $F2_{conv}$) are prove to be a highly informative for ALS detection using running speech test [23].

2.3.3. Distance between the spectral envelopes of the vowels

In [22] it was suggested to use distance between the spectral envelopes of the vowels /a/ and /i/ to quantify the amount of articulatory undershoot. The joint analysis of envelopes of vowels /a/ and /i/ of persons with ALS have revealed an increased similarity between the shapes of these envelopes. The similarity between the envelopes is assessed using l_1 -norm distance metric

$$d_1(E_i, E_a) = \frac{1}{P} \sum_{k=1}^P |E_i(k) - E_a(k)|, \quad (11)$$

where $E_i(k)$ is envelope of the vowel /i/, $E_a(k)$ is envelope of the vowel /a/, P is the number of points in frequency domain. The spectral envelopes of the vowels were estimated using all-pole modelling [31]. An example of vowel envelopes from healthy individual are shown in figure 1,a. A typical example of envelopes with a high degree of similarity is given in figure 1,b.

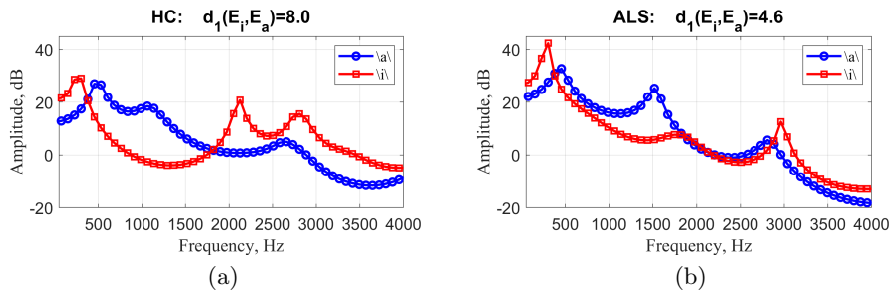


Figure 1: Envelopes of vowels /a/ and /i/: (a) healthy speaker; (b) ALS patient

2.4. F0 contour based parameters

2.4.1. Phonatory frequency range

Phonatory frequency range (PFR) is defined as semitone difference between lowest ($F0_{low}$) and highest ($F0_{high}$) fundamental frequencies [26]:

$$PFR = 12 \frac{\log_{10}(F0_{high}/F0_{low})}{\log_{10} 2}. \quad (12)$$

This parameter measures the degree of variability in fundamental frequency contour and characterizes the functioning of the phonatory subsystem.

2.4.2. Pitch period entropy

Pitch period entropy (PPE) is a highly informative feature proposed in [41] to assess the degree of loss of control over the stationary voice pitch during sustain phonation (due to Parkinson's disease). We have used this measure in our study since the ALS also affects the ability to control the stability of voice pitch.

The calculation of PPE is based on the following observations: 1) the healthy voice has natural pitch variation characterized by smooth vibrato or microtremor [14, 41]; and 2) speakers with naturally high-pitch voices have much larger vibrato and microtremor than speakers with low-pitch voices. PPE measurement takes into account both these factors. The natural smooth variations is removed prior to measuring the extent of such variations (first factor) and pitch transformation to perceptually-relevant, logarithmic semitone scale is applied (second factor). The algorithm of PPE calculation used in this study is given below.

1. Estimation of $F0(m)$ contour with 5 ms time step using IRAPT algorithm [42];
2. Transformation of $F0(m)$ contour to semitone scale:

$$p(m) = 12 \frac{\log_{10}(F0(m)/f_{\text{low}})}{\log_{10} 2}, \quad (13)$$

where f_{low} is lower octave band limit, calculated considering that mean value of pitch correspond to center of this octave:

$$f_{\text{low}} = \text{mean}(F0)/\sqrt{2}.$$

3. Applying whitening filter to $p(m)$ signal to remove healthy, smooth variation:

$$r(m) = \sum_{i=0}^M a_i p(m-i), \quad a_0 = 1, \quad (14)$$

where a_i is linear prediction coefficients (LPC) estimated using covariance method [31], M is the predictor order. We used $M = 2$;

4. Calculation of discrete probability distribution of occurrence of relative semitone variations $P(r)$ by computing normalized histogram in $N = 31$ equal-sized bins r_i ($i = 1, 2, \dots, N$) in the range from -1.5 to 1.5 ;
5. Calculation the entropy distribution $P(r)$ obtained on previous step:

$$\text{PPE} = - \sum_{i=1}^N P(r_i) \log_2 P(r_i), \quad (15)$$

The larger the measure of entropy, the more the observed variations exceed the natural level of variation of the fundamental frequency in a healthy voice. The fig. 2 give an example that illustrates the process of calculation of PPE measure.

Figure 2 shows that semitone pitch sequence $p(t)$ of healthy voice is quite stable and has signs of small regular vibrato. After eliminating this healthy vibrato with whitening filter, the distribution of residuals $r(t)$ shows strong peak at zero. This leads to small value of entropy. On the contrary, for ALS voice the semitone pitch sequence has significant irregular variation, the distribution of residuals is spread over a wider range as a result the larger value of entropy is obtained.

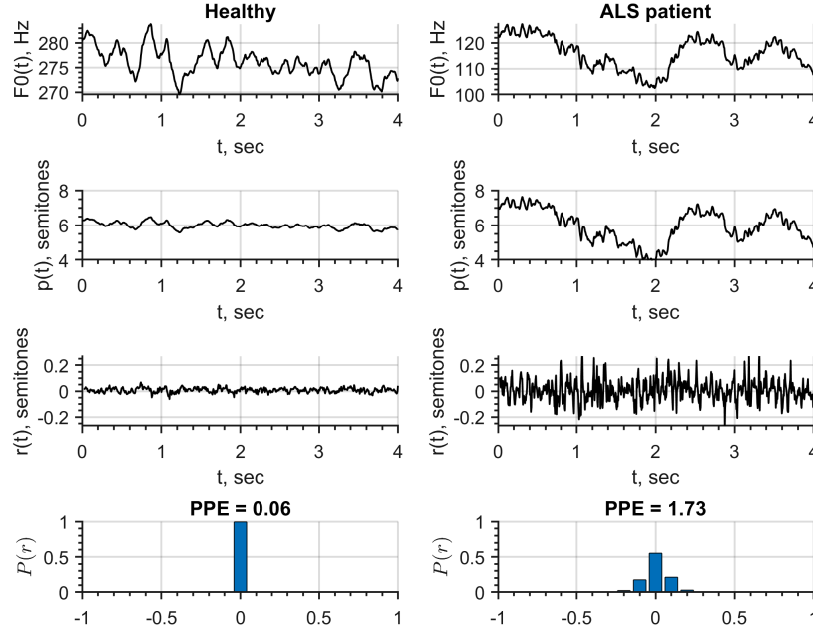


Figure 2: Details of PPE calculation, left column: healthy subject, right column: ALS patient. Rows from the top: extracted F0, pitch $p(t)$ in semitone scale, residual signal $r(t)$ after spectral whitening filter, probability densities $P(r)$ of residual pitch period r

2.4.3. Tremor (vibrato) analysis

Vocal tremor is involuntary quasi-sinusoidal modulation in energy and F0 contour appeared during sustained phonation [43]. In our study we consider only the modulation in F0 contour. Some authors distinguish wow (oscillation of 1-2 Hz), tremor (oscillation of 2-10 Hz) and flutter (oscillation of 10-20 Hz) [34]. An example of vowel phonation for a patient with a rapid tremor (or flutter) is given in figure 3,b (the voice is taken from the database used in experiments).

An essential distortion can be seen when compared spectrogram of a ALS patient (figure 3,b) with spectrogram of a normal subject (figure 3,a). In particular, in figure 3 narrowband spectrograms are shown (long 84 ms analysis window have been used for their calculation). Thus it can be seen substantial changes in harmonics behaviour. Normal voice shows stable harmonics with low variation, while harmonics of pathological voice exhibiting high frequency quasi-sinusoidal modulations.

In [43] in order to characterize the tremor the average spectra of F0 contour is analysed in frequency band from 3 to 25 Hz. However, as reported in [44] the most essential frequency peaks of person with ALS lies within the range 6 to 12

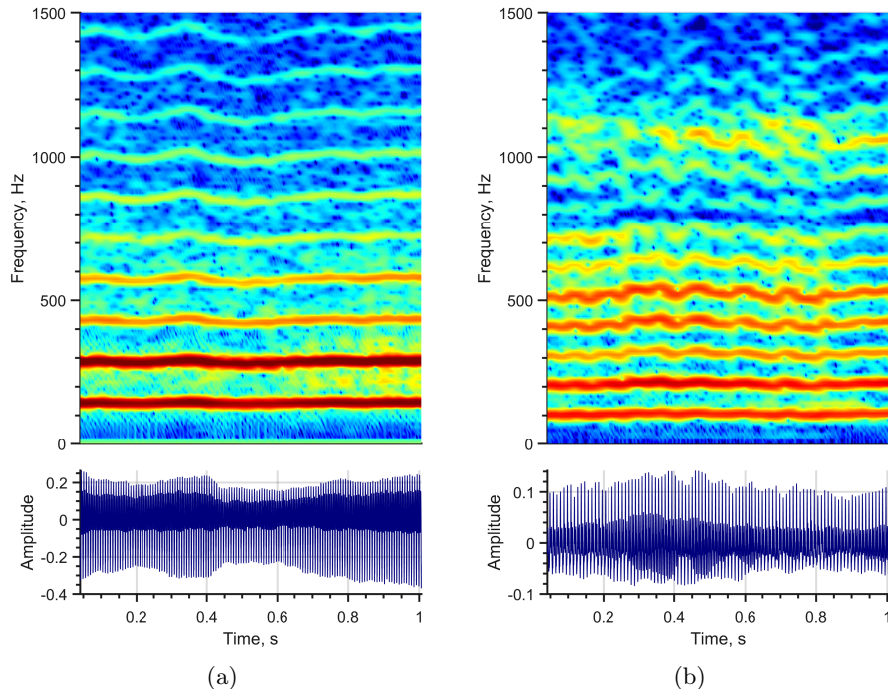


Figure 3: Time-frequency representation of vowel phonation /a/: (a) speaker from HC group (men, 60 years old); (b) ALS patient (men, 67 years old, subject code in voice base 039)

Hz. It seems that sum of the amplitudes of spectral components in frequency band [6, 12] Hz could be a good feature for detection of ALS voices. However, normal voices also have inherent modulations (some times called vibrato) in range 5 to 8 Hz [45]. Thus vibrato frequency bands of healthy and ALS voices are overlapped. So, for obtaining a new feature, that characterizes the extent of pathological modulations in F0 contour we decide to analyse the amplitudes of spectral components in range from 9 to 14 Hz. The obtained feature is referred to as *pathological vibrato index* (PVI) and presented in [46]. The algorithm for PVI calculation is given below

1. Estimation of $F0(m)$ contour with 5 ms time step using IRAPT algorithm [42];
2. Normalization of F0 contour:

$$F0'(m) = \frac{F0(m)}{\text{mean}(F0)}; \quad (16)$$

3. Bandpass filtering of $F0'(m)$ using 3-th order Butterworth IIR with pass band [9, 14] Hz;
4. Amplitude spectrum $A_{F0}(f)$ estimation using Welch's method with windows of 1 sec length and 95% overlap;

5. Calculation of pathological vibrato index:

$$PVI = \sum_{f \in [9, 14] \text{ Hz}} A_{F_0}(f). \quad (17)$$

Figure 4 shows the steps of the PVI calculation for a typical normal and pathological case. It can be seen that frequency components of amplitude spectrum $A_{F_0}(f)$ in the range from 9 to 14 Hz are significantly higher for the ALS voice than for a healthy voice.

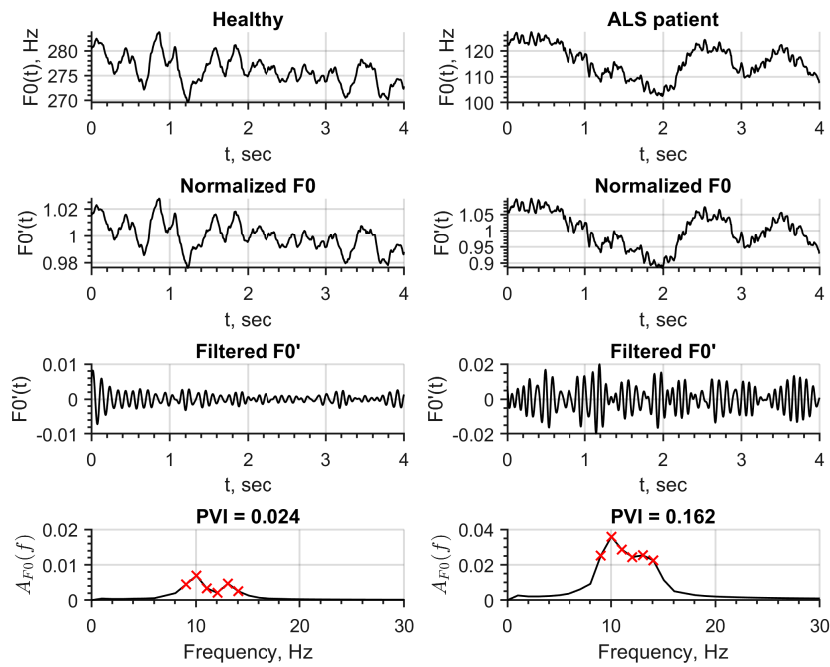


Figure 4: Left column: normal case, right column: pathological case. Rows from the top: Extracted F0, normalized F0 contour, IIR filtered F0 contour, amplitude spectrum $A_{F_0}(f)$, amplitudes used for PVI calculation are indicated by red x-marks

2.4.4. Analysis of the harmonic structure of the vowels

Harmonic structure of sustained vowel has been recognized as an important and informative feature for voice pathology identification [8, 47]. Incomplete glottal closure during phonation, which allows the air to escape, is one of the factors that makes voice more breathy. In particular, for vowel /a/ this produces a disturbance of harmonic structure: amplitude of first harmonic (H1) becomes higher than the second (H2) [47].

One of the important aspect of voice quality is stability of harmonics structure during the phonation process. Evaluation of harmonics structure can be considered as feature for description the excitation source (a driving force for voice production). The difficulty in estimation of harmonic parameters is that they depend on the fundamental frequency F0. In this study we have used voice analysis based on fixed number of fundamental periods (alternatively it can be considered as pitch synchronized voiced analysis). We focused on extracting mean and standard deviation (SD) of the first eight harmonics of the vowels. Given a voice signal $s(n)$ the analysis process can be summarized in the following steps.

1. Split $s(n)$ into fundamental periods using waveform matching method with phase constrain [46].
2. Divide $s(n)$ into N_f overlapping frames that containing N_c fundamental periods with one period overlap. For each frame $s^{(i)}(n)$, $i = 1, \dots, N_f$ execute steps 3–5.
3. Interpolate $s^{(i)}(n)$ into $I \times N_c$ equidistant time points: $s^{(i)}(n) \rightarrow \widehat{s}^{(i)}(m)$.
4. Apply Hamming window $h(m)$ to interpolated frame and compute discrete Fourier Transform (DFT): $\widehat{S}^{(i)}(k) = \text{DFT}[\widehat{s}^{(i)}(m)h(m)]$.
5. Extract harmonic amplitudes:

$$h_p(i) = |\widehat{S}^{(i)}(p \times I)| \quad p = 1, 2 \dots 8.$$

6. Scale the harmonic amplitudes as

$$\widetilde{H}_p(i) = 20 \log_{10} \left(\frac{h_p(i)}{\max_{p \in [1, 8], i \in [1, N_f]} \{h_p(i)\}} \right).$$

7. Compute mean and SD for scaled harmonic amplitudes

$$\text{Hp}^\mu = \text{E}\{\widetilde{H}_p\}, \quad \text{Hp}^\sigma = \sqrt{\text{E}\{(\widetilde{H}_p - \text{Hp}^\mu)^2\}}$$

8. Compute additional feature – inverse of the sum of absolute value of Hp^μ and Hp^σ :

$$\text{RelHp} = \frac{1}{|\text{Hp}^\mu| + \text{Hp}^\sigma}. \quad (18)$$

The intuition behind the feature (18) is that strong and stable harmonic should have low scaled amplitude $|\text{Hp}^\mu|$ and low deviation Hp^σ and therefore high value of RelHp .

In this study the following parameters of the procedure were used: $N_c = 8$ and $I = 512$.

3. Experiments

3.1. Database

Voice database³ used in this study was collected in Republican Research and Clinical Center of Neurology and Neurosurgery (Minsk, Belarus). It consists of

³The database available online at https://github.com/Mak-Sim/Minsk2020_ALS_database

128 sustained vowel phonations (64 of vowel /a/ and 64 of vowel /i/) from 64 speakers, 31 of which were diagnosed with ALS. Each speaker was asked to produce sustained phonation of vowels /a/ and /i/ at a comfortable pitch and loudness as constant and long as possible. It can be seen that voice database is almost balanced and contains 48% of pathological voices and 52% of healthy voices.

The age of the 17 male patients ranges from 40 to 69 (mean 61.1 ± 7.7) and the age of the 14 female patients ranges from 39 to 70 (mean 57.3 ± 7.8). For the case of healthy controls (HC), the age of the 13 men ranges from 34 to 80 (mean 50.2 ± 13.8) and the age of the 20 females ranges from 37 to 68 (mean 56.1 ± 9.7). The samples were recorded at 44.1 kHz using different smartphones with a regular headsets and stored as 16 bit uncompressed PCM files. Average duration of the records in the HC group was 3.7 ± 1.5 s, and in ALS group 4.1 ± 2.0 s. The detailed information about ALS patients is presented in table 1. All the patients were judged by the neurologist (the second author) to have presence of bulbar motor changes in speech (last column of the table 1).

3.2. Aggregation of feature set and its statistical survey

For each vowel used in SVP test 64 features are extracted (see figure 5). These features include the following groups (the number of parameters in each group is indicated in parentheses): jitter (4), shimmer (5), DPF(1), HNR (1), GNE (mean and SD), PFR(1), PPE(1), PVI (1), Hp^μ (8), Hp^σ (8), RelHp (8), MFCC (12), Δ MFCC (12). We also used three additional parameters

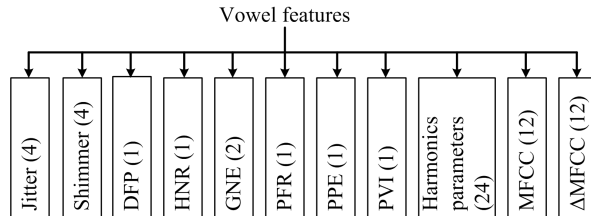


Figure 5: Features extracted from SVP test of one vowel

$d_1(E_a, E_i)$, $F2_{conv}$ and $F2_i$ (extra feature for vowel /i/). Thus the total number of features used in this study was 131 (64 for vowel /a/, 64+1 for /i/ and 2 joint parameters). In most cases we have used lower subscript to indicate the vowel for which feature was calculated. For example, $H2_i^\sigma$ is SD of 2nd harmonic of vowel /i/ phonation.

In order to get initial understanding of the statistical properties of the features, we computed the Pearson correlation coefficient $r(\mathbf{x}, \mathbf{y})$, where the vector \mathbf{x} contains the values of a single feature for all phonations, and \mathbf{y} is the associate labels (“0” for healthy subject, “1” – for ALS patient).

Table 1: ALS participants clinical records

Subject code	Sex	Age	Time from ALS onset (months)	Bulbar/ spinal onset	Presence of the bulbar signs
008	M	67	28	bulbar	yes
020	F	57	35	spinal	no
021	F	55	15	spinal	yes
022	F	70	11	bulbar	yes
024	M	66	16	spinal	no
025	M	51	7	spinal	no
027	M	57	18	bulbar	yes
028	M	58	5	spinal	yes
031	M	67	6	spinal	yes
032	M	61	19	spinal	yes
039	M	67	12	bulbar	yes
042	M	67	22	spinal	yes
046	F	50	12	spinal	yes
048	F	63	22	bulbar	yes
052	F	62	36	spinal	no
055	M	61	11	spinal	yes
058	M	58	9	bulbar	yes
062	M	57	23	bulbar	yes
064	M	57	58	spinal	yes
068	M	40	11	bulbar	yes
072	F	64	10	spinal	yes
076	M	68	12	bulbar	yes
078	F	64	12	bulbar	yes
080	F	63	20	bulbar	yes
084	F	55	33	bulbar	yes
092	F	39	57	spinal	no
094	F	55	14	spinal	no
096	F	52	14	spinal	yes
098	M	68	37	spinal	yes
100	M	68	16	bulbar	yes
102	F	53	25	spinal	no

3.3. Feature selection

It is known that reducing the number of features often improves the model’s predictive power. Also the reduced feature subset give better insight into the problem via analysis of the most predictive features [48].

In this study we used four efficient feature selection (FS) approaches: 1) maximization of quality of variation (QoV) [49], 2) Relief [50] 3) least absolute shrinkage and selection operator (LASSO) [51], 4) RelieFF [52]. Maximization of QoV is a noise-resistant method for feature selection based on order statistics. The basic notion of this method is *class impurity* – characteristic that is calculated for each feature based on its order statistics. The *quality of variation* of a feature is inverse of the average impurity of all the classes along the feature. This method allows one to rank all features according to the QoV criterion. It has been show that QoV method performs well when the available training data is small or not much bigger compared to the dimensionality of feature vector [49]. LASSO is a linear regression based technique that minimizes the residual sum

of squares subject to the absolute value of the coefficient being less than a constant. This leads to some coefficients that are shrunk to zero, which in essence means that feature associated with those coefficients are eliminated. In order to rank the features using LASSO we repeat its computation with different values of regularization parameter λ in order to track the order in which features are eliminated. The first eliminated feature is considered as least informative while the last as the most relevant. The key idea of Relief is to estimate features according to how well their values distinguish among the instances that are near to each other. Original Relief algorithm estimates relevance of feature for a given instance by analysis closest neighbors: one from the same class (nearest hit) and one from the opposite class (nearest miss). Advanced version RelieFF extends this idea to k nearest neighbors. Overall, all four feature selection algorithms have shown promising results in machine learning application.

3.4. Classification

For a binary classification between normal and pathological classes, linear discriminant analysis (LDA) with Fisher criterion was used [53]. The basic idea of LDA consists in searching for such a direction \mathbf{w} in the feature space, that the projection of all training vectors onto it minimizes the within-class variation and maximizes the between-class variation:

$$\mathbf{w} = \arg \max_{\mathbf{w}} \frac{\mathbf{w} \mathbf{S}_B \mathbf{w}^T}{\mathbf{w} \mathbf{S}_W \mathbf{w}^T}, \quad (19)$$

where \mathbf{S}_B – between class scatter matrix and \mathbf{S}_W – within class scatter matrix. In turn these matrices are calculated as follows

$$\mathbf{S}_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T, \quad (20)$$

$$\mathbf{S}_W = \sum_{j=1}^2 \sum_{\mathbf{x}} (\mathbf{x} - \mu_j)(\mathbf{x} - \mu_j)^T, \quad (21)$$

where \mathbf{x} – feature vectors from training set, μ_1 – mean value of feature vector for healthy people and μ_2 – mean value of feature vector for people with ALS. The solution of (19) can be found via the generalized eigenvalue problem

$$\mathbf{S}_B \mathbf{w} = \lambda_m \mathbf{S}_W \mathbf{w}, \quad (22)$$

where the eigenvector associated with maximum eigenvalue λ_m gives the projection basis. Classification function of LDA is formulated as follows

$$f(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \mathbf{x} \rangle + b), \quad (23)$$

where b is a bias. In the experiments, the value of b was chosen in a such way that the number of correctly detected positive and negative instance in the training set was equal. More detailed description of LDA can be found in [53].

3.5. Classifier Validation

The goal of validation is to estimate of the generalization performance of the classification based on the selected set of features, when presented new (previously unseen) data. Most studies use cross-validation to achieve this goal [12, 28, 47].

In this work we used k -fold stratified cross validation (CV) method [54], with k equal to 8. According to this method at the beginning of the CV process dataset randomly permuted and then splits into eight equal subsets (folds) (s_1 - s_8), the folds are stratified so that they contain approximately the same proportions of labels as original dataset. At first iteration classifier is trained using subsets s_1 - s_7 , while testing is conducted using s_8 subset. Then training is repeated using s_2 - s_8 subsets, and classifier tested using s_1 subset, and so on. After 8 iteration whole dataset is labelled using eight classifiers. This process was repeated a total of 40 times. The classification performance is evaluated in terms of the mean and standard deviation of the accuracy on the test set across all folds.

Accuracy, sensitivity, and specificity were used in this study to measure the classification performance. Accuracy is the overall probability of correctly classified instance over the total number of instances. Sensitivity is the probability of correctly classified ALS patients given all ALS samples and specificity is probability of classified HC given all HC samples. Accuracy, sensitivity, and specificity are calculated as follows:

$$\begin{aligned} Acc &= \frac{TP + TN}{TP + FP + FN + TN} \\ Sens &= \frac{TP}{TP + FN} \\ Spec &= \frac{TN}{TN + FP} \end{aligned}$$

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

4. Results

4.1. Preliminary statistical survey

Table 2 presents the several features most strongly associated with the labels in dataset, sorted by the absolute value of the correlation coefficient [53]. We used label “0” for healthy controls and “1” for people with ALS. Thus, positive correlation coefficient suggest that the feature takes, in general, larger value for ALS voices. All of the listed features exhibit statistically significant correlation ($p < 0.05$).

According to the table 2 the most relevant features are d_1 and $MFCC_i(2)$. The distance between spectral envelopes of the vowels /a/ and /i/ (d_1) has

Table 2: Statistical analysis of the acoustic features

Feature	Correlation coefficient	Difference between groups
d_1	-0.456	$p < 0.0002$
$\text{MFCC}_i(2)$	-0.446	$p < 0.0003$
PVI_a	0.422	$p < 0.0006$
PPE_a	0.418	$p < 0.0006$
$F2_{conv}$	-0.390	$p < 0.002$
RelH7_a	-0.381	$p < 0.002$
$\text{MFCC}_i(6)$	0.371	$p < 0.003$
$J_{ppq55}^{(a)}$	0.361	$p < 0.004$
PVI_i	0.351	$p < 0.005$
RelH1_i	-0.347	$p < 0.005$
PFR_a	0.346	$p < 0.006$
H8_a^μ	-0.335	$p < 0.007$
GNE_a^μ	-0.324	$p < 0.01$
$\Delta\text{MFCC}_i(6)$	0.321	$p < 0.01$
$\mathcal{S}_{apq11}^{(i)}$	0.311	$p < 0.02$
$F2_i$	-0.302	$p < 0.02$
RelH1_a	-0.285	$p < 0.03$
GNE_i^σ	0.282	$p < 0.03$
H4_a^σ	0.282	$p < 0.03$
$\text{MFCC}_i(8)$	0.273	$p < 0.03$
$\text{MFCC}_a(11)$	0.250	$p < 0.05$

the strongest correlation with the labels in the dataset. The negative sign of its correlation coefficient means that the smaller distance d_1 , the more likely that voice belongs to the category of ALS patients. $\text{MFCC}_i(2)$ has almost the same strong correlation as spectral distance d_1 . It is well known that, low-order MFCC describes the spectral envelope of the sound, therefore it can be concluded that patients with ALS usually have significant changes in spectral envelope of the vowel /i/.

Parameters PVI_a and PPE_a (third and fourth rows in the table 2) also have high correlation with the labels in the dataset. This fact indicates that, as a result of neuromotor disorders in patients with ALS, oscillations uncharacteristic for healthy people appear in the F0 contour, which lead to increasing of PVI_a and PPE_a . It is interesting that along with PVI_a parameter PVI_i (ninth row) is also presented in table 2 and has high correlation coefficient, while PPE_i does not show a statistically significant correlation ($p > 0.14$). This may indicate that PVI less depends on type of analyzing vowel than PPE and better reflects changes associated with a decrease in the control over fundamental frequency in patients with ALS.

Five features (RelH7_a , RelH1_i , H8_a^μ , RelH1_a and H4_a^σ), among twenty-one

of those listed in Table 2, relate to parameters that describe harmonic structure of the vowel. This suggests that these parameters could be useful for accurate voice classification.

We also estimated distribution of several features listed in table 2 using Gaussian kernel density method to characterize their statistical properties. Figure 6,a shows the distribution of distance between spectral envelopes of the vowels /a/ and /i/ (first row in table 2). As expected, on average this feature has lower value for ALS patients than for healthy subjects.

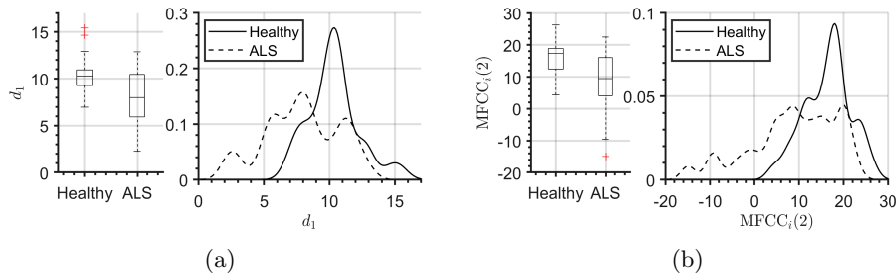


Figure 6: Box plot and probability densities of (a) $d_1(E_a, E_i)$; (b) $MFCC_i(2)$

The distribution of 2nd MFCC of vowel /i/ that has a strong correlation with labels in the dataset is given in figure 6,b. As stated above the differences in $MFCC_i(2)$ indicate changes in the spectral envelope of the vowel /i/ in patients with ALS. Among the others, this can be seen from the changes of the second formant frequency of the vowel /i/. From table 2 we see that a lower value of $F2_i$ is typical for patients with ALS. This observation is consistent with previous findings in this area [21, 40]. Let us consider scatter plot of the pairs of $F2_i$ and $MFCC_i(2)$ for healthy and pathological voices (see figure 7). It can be seen that for healthy voices $F2_i$ and $MFCC_i(2)$ are weakly correlated (i.e. they not set out along slanting line). In contrast, for the voices of patients with ALS, it can be seen that $F2_i$ and $MFCC_i(2)$ are strongly correlated (points are grouped along slanting line). Thus high relevance of the $MFCC_i(2)$ is likely caused by the fact that it reflects the changes in second formant of vowel /i/ in patients with ALS.

Figure 8,a-b illustrate distributions of PPE_a and PVI_a features. Both of them characterize the excess of variability in a pitch contour and have high correlation coefficients (3-rd and 4-th rows of table 2). Comparing boxplots of the PPE_a and PVI_a parameters, we can see that the first quartile of PPE_a for pathological voices is located at the level of the median of the PPE_a for healthy voices. In turn, the first quartile of PVI_a for pathological voices exceeds the third quartile of PVI_a for healthy voices. This indicates that PVI_a has stronger discriminatory power than PPE_a .

Examples of features that characterized harmonic structure of voice are given in figure 8,c-d. Both features are strongly associated with the labels in the dataset. As expected, distributions figure 8,c indicate that $RelH1_i$ has tending to have higher value for healthy voices. Figure 8,d shows that 8-th harmonic of

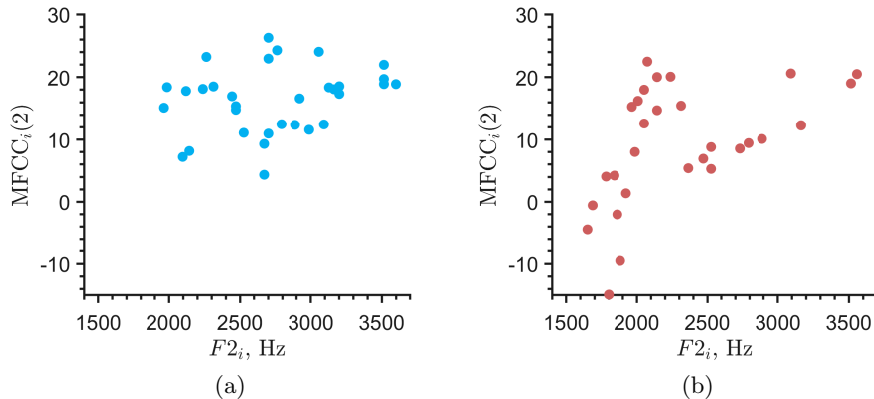


Figure 7: Scatter plots of pairs of $F2_i$ and $MFCC_i(2)$ showing low correlation for healthy voices (a) and high correlation for ALS voices (b)

vowel /a/ has lower mean value in ALS group.

Figure 8,e-f illustrates the distributions of the MFCC and delta MFCC that have high correlation with labels in the dataset (rows 7 and 14 in table 2). Boxplot in figure 8,f shows that $\Delta MFCC_i(6)$ have almost symmetrical distribution with median greater than zero for ALS voices, while for healthy voices this parameter have asymmetrical distribution with near zero median.

The presented findings give tentative confidence that we can expect good results for the classification problem of this study.

4.2. Classification results and discussion

In our experiments we computed the accuracy (see section 3.5) of LDA classifier using different number of features selected by the four FS algorithms described in section 3.3. Figure 9 shows the obtained results.

The analysis of figure 9 shows that performance of all FS algorithm is quite similar while the number of features N is less than 6. However, for $N > 6$ LASSO demonstrates significantly better performance in comparison with other approaches. Possible explanation of this fact is that mathematical principles of LASSO regression are in accordance with the discriminant function (23) of the LDA classifier.

The optimal size of the feature vector is equal to 43 and it was achieved using the LASSO approach. The accuracy obtained in this case is 97%. This result considerably outperforms the others. For example, the best accuracy of LDA classifier with QoV FS algorithm is 79% and was achieved using 4 features. The best results obtained with ReliefF and Relief algorithms are even smaller – 76% and 72%, accordingly.

It is always desirable to have a classifier with a low number of features, therefore we applied backward-step selection procedure [48] to reduce the number of features picked out by FS algorithms. The backward-step selection starts with

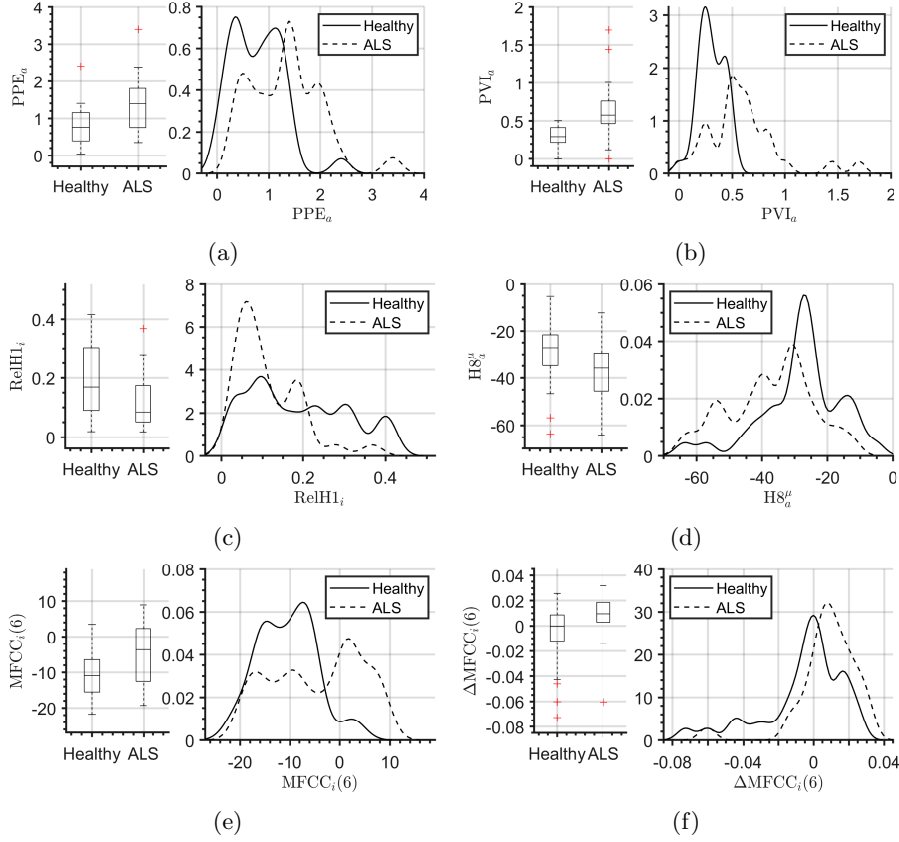


Figure 8: Box plot and probability densities of : (a) PPE_a ; (b) PVI_a ; (c) $RelHI_i$; (d) $H8_a^\mu$; (e) $MFCC_i(6)$; (f) $\Delta MFCC_i(6)$

LDA model that used best feature subset found by FS algorithm, and sequentially deletes the feature that has low (or negative) impact on the fit. The result of the described feature selection process is summarized in table 3.

Result shown in table 3 demonstrate that the best accuracy for LDA classifier is obtained using feature selected by the LASSO algorithm. Also it can be noted that backward-stepwise selection (BSS) is effective in reducing the number of features and increasing the accuracy of classifier. The most noticeable result, in this regard, is increasing the accuracy of LDA model with feature subset selected using RelieFF algorithm by 7 %, while reducing the number of features by 9. Nevertheless, feature subsets found by QoV, Relief and RelieFF algorithms with application of BSS procedure give the resulting accuracy considerably lower compared to feature subset selected using LASSO algorithms.

Table 4 lists final subsets of features selected using FS algorithms (with application of BSS). The obtained accuracy, sensitivity and specificity for each cases are also given in table 4.

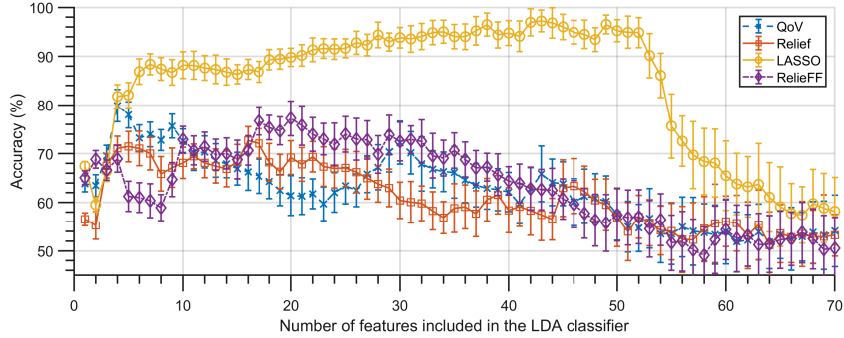


Figure 9: Classification accuracy with confidence interval (one standard deviation around the quoted mean accuracy). The results obtained using different feature selection algorithm. For ReliefFF algorithm adjustable parameter $k = 11$ was used.

The analysis of tables 2 and 4 leads to the logical question: why statistically significant feature d_1 was not selected by LASSO FS algorithm? Detailed analysis have revealed that d_1 and $\text{MFCC}_i(2)$ have strong correlation ($r = 0.54$ with $p < 1.0 \cdot 10^{-5}$), thus, MFCCs already contain information possessed in d_1 feature. Another question: why such significant features like $F2_{conv}$ and $F2_i$ were not selected by neither algorithm? First of all these features are highly correlated ($r = 0.85$ with $p < 10^{-18}$), thus the location of $F2_i$ is more relevant rather than its proximity to $F2_a$. Furthermore, $F2_i$ is strongly correlated with $\text{MFCC}_i(2)$ and $\text{MFCC}_i(6)$ ($r = 0.48$ and $r = -0.44$, accordingly), therefore the information about $F2_i$ can be passed to classifier with any of these parameters. LASSO and QoV algorithms have selected $\text{MFCC}_i(4)$ that contained information about $F2_i$ location, while ReliefFF algorithm selected $\text{MFCC}_i(6)$ for this purpose. A visual example of interplay between $F2_i$ and $\text{MFCC}_i(6)$ is given in figure 10.

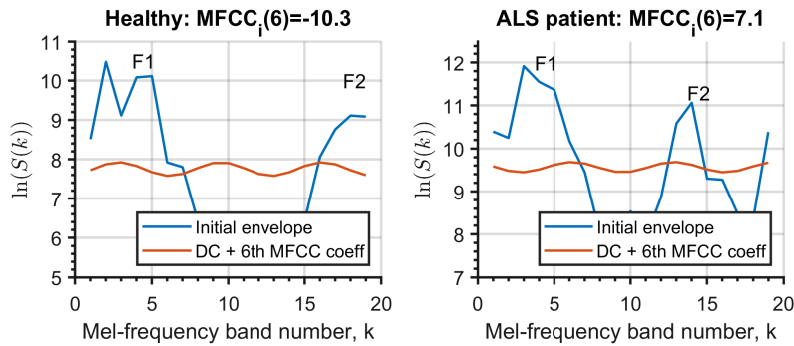


Figure 10: Interplay between features $F2_i$ and $\text{MFCC}_i(6)$

Figure 10 shows the estimation of spectral envelopes computed during MFCC

Table 3: Classifiers accuracy obtained using different feature selection (FS) algorithms. The resulting number of features is given in parentheses.

FS algorithm	Accuracy with initial subset	Accuracy after backward-stepwise selection
QoV	$79.5 \pm 3.5\%$ (4)	$79.5 \pm 3.5\%$ (4)
Relief	$72.5 \pm 3.4\%$ (16)	$80.3 \pm 2.3\%$ (5)
LASSO	$97.0 \pm 2.4\%$ (43)	$99.7 \pm 0.6\%$ (32)
RelieFF	$75.9 \pm 4.2\%$ (20)	$82.9 \pm 2.8\%$ (11)

calculation and partial reconstruction of envelopes using DC component and 6-th MFCC coefficient. It can be seen that the voice of ALS patient is characterized by reduced frequency of second formant. As a result, projection onto 6th basis function of discrete cosine transform which is used in MFCC calculation is changing sign (if we compare HC and ALS voices).

The result of our study confirm the findings of [12], where MFCC are also found to be highly informative features for Parkinson’s disease detection. However unlike [12] we give interpretation that MFCC reflect changes in second formant of vowel /i/ for ALS patients. Also it should be noted that proposed features extracted using harmonic analysis of the vowels are essential for obtaining good classifier. For example, among the 32 features selected by LASSO, ten describe the harmonic structure. Among the rest features: 9 MFCCs describe the envelopes of vowels (6 for /a/ and 3 for /i/), 7 delta MFCCs reflect variability of vowels envelopes, the GNE parameters gives information about noise content of the voice and PVI describes the changes in vibrato. Another interesting observation is that in subset of features selected by LASSO 19 are related to vowel /a/ and 13 to vowel /i/. It means that information contained in phonation /a/ is relevant and necessary for gaining high classification accuracy. It is interesting that traditional measures such as jitter, shimmer and HNR are out of table 4. This suggests that PVI, MFCC and harmonic structure parameters have greater predictive power for distinguishing between HC and patients with ALS.

Surely, that main goal of classification is most accurate detection of the ALS patients voices. In this regard, LDA model with 32 features and 99% accuracy is a significant result. However, there is reason to believe that this feature set is quite specific for our voice database. More relevant information about parameters that are most important for ALS detection can be derived by an-

Table 4: Selected feature subsets and classification accuracy

FS algorithm	Features	Accuracy results (%)
QoV ($N = 4$)	$PVI_a, PFR_a, d_1, MFCC_i(2)$	$Acc = 79.5 \pm 3.5$ $Sens = 75.3 \pm 3.4$ $Spec = 83.5 \pm 5.9$
Relief ($N = 5$)	$PVI_i, d_1, MFCC_i(4), MFCC_i(9), \Delta MFCC_a(3)$	$Acc = 80.3 \pm 2.3$ $Sens = 68.1 \pm 3.4$ $Spec = 91.7 \pm 2.3$
LASSO ($N = 32$)	$PVI_i, H2_a^\mu, H4_a^\mu, H3_a^\sigma, RelH1_a, RelH3_a, RelH4_a, RelH6_a, RelH8_a, RelH1_i, RelH3_i, MFCC_a(1), MFCC_a(4), MFCC_a(7), MFCC_a(10), MFCC_a(11), MFCC_a(12), \Delta MFCC_a(5), \Delta MFCC_a(9), \Delta MFCC_a(11), MFCC_i(2), MFCC_i(4), MFCC_i(8), MFCC_i(9), \Delta MFCC_i(1), \Delta MFCC_i(9), \Delta MFCC_i(10), \Delta MFCC_i(12), GNE_a^\sigma, GNE_i^\mu, GNE_i^\sigma, DPF_a$	$Acc = \mathbf{99.7 \pm 0.6}$ $Sens = \mathbf{99.3 \pm 1.4}$ $Spec = \mathbf{99.9 \pm 0.5}$
RelieFF ($N = 11$)	$PVI_i, H3_a^\sigma, H4_a^\sigma, H1_i^\sigma, RelH1_i, RelH3_a, MFCC_a(11), MFCC_i(6), \Delta MFCC_a(1), \Delta MFCC_a(3), GNE_a^\sigma,$	$Acc = 82.9 \pm 2.8$ $Sens = 78.0 \pm 4.4$ $Spec = 87.6 \pm 2.6$
Low order model		
10 best LASSO features +BSS ($N = 5$)	$PVI_i, MFCC_i(2), MFCC_i(9), MFCC_a(8), MFCC_a(10),$	$Acc = 89.0 \pm 2.5$ $Sens = 87.5 \pm 2.9$ $Spec = 90.4 \pm 3.3$

alyzing high-performance LDA models with small number of features. Table 4 shows that LDA models obtained using the QoV and RelieFF feature selection algorithms have a small number of features, however they have quite low performance. To find a model with higher performance we took LDA classifier with 10 best feature picked out using LASSO algorithm, which has accuracy $87.6 \pm 2.6\%$ ($Sens = 90.5 \pm 4.1\%$, $Spec = 84.8 \pm 3.0\%$) and applied back-step selection procedure. As a result LDA model with five features ($MFCC_i(2)$, $MFCC_i(9)$, PVI_i , $MFCC_a(8)$, $MFCC_a(10)$) was obtained, which has accuracy 89.0% (see last row of table 4). Thus, it can be concluded that the most important information for detecting the ALS patients' voices is contained in the spectral envelopes of sounds /a/ and /i/ (MFCC parameters), as well as in the vibrato changes (PVI).

Table 5: Comparison with other studies

	Norel [9]	Spangler [10]	An [11]	Present
Feature	Extracted with Open-SMILE toolkit	Fractal jitter, MFCC, RPDE + articulatory data	filterbank energies + its deltas	MFCC, Harmonic parameters, PVI
Total number of features	<i>for male 1 for female 15</i>	17	120	32
Classifier	linear SVM	Extreme Gradient Boosting	CNN	LDA
Verification	Leave-five-subject-out CV	Leave-one-subject-out CV	Leave-one-subject-pair-out CV	8-fold CV
Database	133 speakers (67 ALS, 66 HC), running speech	83 speakers (49 ALS, 34 HC), DDK test	26 speakers (13 ALS, 13 HC), running speech	64 speakers (31 ALS, 33 HC), SVP test
Reported performance	<i>for male Acc=79% Sens=76% Spec=70% for female Acc=83% Sens=78% Spec=88%</i>	Acc=90.2% Sens=94.2% Spec=85.1%	Acc=76.2% Sens=71.5% Spec=80.9%	Acc=99.7% Sens=99.3% Spec=99.9%

Table 5 presents comparison of the present work with recent similar studies. The purpose of those works was to discriminate between healthy people and ALS patients. The main differences between these studies concern speech tasks, classification approaches, features and verification methods. The most closest result was obtained in [10]. However, in [10] along with voice recording articulatory data was used. In table 5 two different performance results are given for study [9] because it uses sex-specific features for classifiers to take into account differences in the vocal tracts of males and females. Study [11] presents results of two type: sample-level and person level classification. The second type is obtained based on sample voting. In table 5 we compare only sample-level classifiers. However, even person-level classifier based on 5 samples [11] has accuracy 90.8%, sensitivity 85.6% and specificity 94.9%. Therefore the obtained result with near 99% of accuracy, sensitivity and specificity based on LDA classifier can be considered as an essential improvement over the previous results.

4.3. Additional experiment: early ALS detection

The following additional experiment has been performed in order to determine validity of LDA models with features extracted from SVP test for early ALS detection problem. From ALS patients were chosen 12 that having been diagnosed less than one year before recordings (see table 1). So, the reduced dataset included 45 speakers (33 HC + 12 ALS).

Using the reduced dataset, we performed feature selection procedures and optimization of feature set as described above. However, in contrast to experiments presented in previous sections, we used leave-one-subject-out (LOSO) cross-validation procedure to evaluate the performance of classifiers. [48, 53]. In fact, the LOSO method is a k -fold CV procedure, with k equal to the size of the dataset. We used LOSO in order to bring closer the size of the samples on which the classifiers are trained in sections 4.2 and 4.3. In section 4.2, where the 8-fold CV was used, the LDA classifier model was trained on 54 samples, in this section, using the LOSO CV method, the classifier is trained on 44 samples.

Table 6: Early ALS detection: selected feature subsets and classification accuracy

FS algorithm	Features	Accuracy results (%)
QoV + BSS ($N = 5$)	$H3_a^\sigma$, $H5_i^\mu$, $H6_i^\mu$, $RelH6_i$, $MFCC_i(6)$	$Acc = 84.4 \pm 5.4$ $Sens = 75.0 \pm 6.5$ $Spec = 87.9 \pm 4.9$
RelieFF ($N = 5$)	$MFCC_a(8)$, $MFCC_a(11)$, $MFCC_i(2)$, $MFCC_i(6)$, PFR_a	$Acc = 93.3 \pm 3.7$ $Sens = 83.3 \pm 5.6$ $Spec = \mathbf{97.0 \pm 2.6}$
LASSO ($N = 12$)	d_1 , PFR_a , $H7_i^\sigma$, $RelH6_a$, $MFCC_a(6)$, $MFCC_a(8)$, $MFCC_i(2)$, $MFCC_i(3)$, $MFCC_i(6)$, $MFCC_i(9)$, $\Delta MFCC_i(6)$, $\Delta MFCC_i(12)$,	$Acc = \mathbf{95.6 \pm 3.1}$ $Sens = \mathbf{91.7 \pm 4.1}$ $Spec = \mathbf{97.0 \pm 2.6}$

LDA model with 5 features and above 80% accuracy has been obtained using QoV feature selection algorithm with BSS procedure (see table 6). Best LDA model obtained using Relief algorithm with BSS procedure has 39 features and 100% accuracy. The same accuracy is achieved by the LDA model using 28 features selected by LASSO algorithm. However, these feature sets (unless they legitimacy) are too specifically fit to our database. We believe that more relevant conclusions can be derived by analyzing models with feature sets of limited size. For example, LDA model trained on the first 5 features selected by RelieFF algorithm has 93,3% accuracy (see table 6). Furthermore, among the LDA models with a small number of features we can highlight one that has 95,6% accuracy and trained on the first 12 features selected by the LASSO algorithm.

Analyzing the features contained in the table 6, we can draw the following conclusions. In all feature sets $\text{MFCC}_i(6)$ is present, its relevance is discussed in previous sections (for example, see figure 10). Four out of five features selected by RelieFF algorithm are also included in feature set picked out by LASSO algorithm. This indicates their high significance for early ALS detection. Feature set obtained using the RelieFF algorithm shows that valuable information for early ALS detection is contained in spectral envelopes of the vowels /a/ and /i/ (this information is concentrated in parameters $\text{MFCC}_a(8)$, $\text{MFCC}_a(11)$ and $\text{MFCC}_i(2)$, $\text{MFCC}_i(6)$). Parameter PFR_a , which indicates the degree of fundamental frequency variation, is also important for early ALS detection. It should be noted that neither of the feature sets contains parameters PVI and PPE, the significance of which was revealed in the previous experiment. This means that changes in the vibrato are not related to the early diagnosis of the ALS, but rather characteristic of later stages of the disease.

5. Conclusion

In this study we investigate the possibility of designing linear classifier for discriminate ALS patients from healthy controls using acoustical sustained vowels /a/ and /i/ phonation tests. A large set of features was analysed. LDA classifier with 99.7% accuracy (99.3% sensitivity, 99.9% specificity) was obtained based on 32 features determined by LASSO feature selection algorithm. We also obtained the LDA model with only 5 features that has 89.0% accuracy (87.5% sensitivity, 90.4% specificity). We found that the most important information for detecting the ALS patients' voices is contained in the spectral envelopes of sounds /a/ and /i/ (MFCC parameters), as well as in the vibrato changes (PVI). Like in [10] traditional jitter measures were found not to have a high importance. We also carried out experiment to determine validity of LDA models with features extracted from SVP test for early ALS detection problem. Our results show that it is possible to obtain LDA model with 93.3% accuracy (83.3% sensitivity, 97.0% specificity) based on only 5 features determined by RelieFF algorithm. We can also draw the conclusion that valuable information for early ALS detection is contained in spectral envelopes of the vowels /a/ and /i/ (MFCC parameters). We also found that the selected feature sets did not contain the PVI and PPE parameters. This means that changes in the vibrato are not related to the early diagnosis of the ALS, but rather characteristic of the later stages of the disease. It should be noted that the data for this study was collected using smartphone with regular headset. Therefore we can assert that proposed approach is tolerant to non-professional recording condition.

Acknowledgements

The authors thank the anonymous reviewers for their useful comments.

References

- [1] J. R. Green, Y. Yunusova, M. S. Kuruvilla, J. Wang, G. L. Pattee, L. Synhorst, L. Zinman, J. D. Berry, Bulbar and speech motor assessment in ALS: challenges and future directions, *Amyotrophic lateral sclerosis and frontotemporal degeneration* 14 (7-8) (2013) 494–500. doi:10.3109/21678421.2013.817585.
- [2] J. R. Duffy, *Motor speech disorders: Substrates, differential diagnosis, and management*, Elsevier Health Sciences, 2013.
- [3] Y. Iwasaki, K. Ikeda, M. Kinoshita, The diagnostic pathway in amyotrophic lateral sclerosis, *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders* 2 (3) (2001) 123–126. doi:10.1080/146608201753275571.
- [4] A. Benba, A. Jilbab, A. Hammouch, Discriminating between patients with Parkinson’s and neurological diseases using cepstral analysis, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24 (10) (2016) 1100–1108. doi:10.1109/TNSRE.2016.2533582.
- [5] J. Rusz, R. Cmejla, H. Ruzickova, E. Ruzicka, Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson’s disease, *The Journal of the Acoustical Society of America* 129 (1) (2011) 350–367. doi:10.1121/1.3514381.
- [6] J. R. Orozco-Arroyave, F. Honig, J. D. Arias-Londoño, J. F. Vargas-Bonilla, K. Daqrouq, S. Skodda, J. Rusz, E. Noth, Automatic detection of parkinson’s disease in running speech spoken in three different languages, *The Journal of the Acoustical Society of America* 139 (1) (2016) 481–500.
- [7] P. Gomez-Vilda, A. R. M. Londral, V. Rodellar-Biarge, J. M. Ferrandez-Vicente, M. de Carvalho, Monitoring amyotrophic lateral sclerosis by biomechanical modeling of speech production, *Neurocomputing* 151 (2015) 130–138. doi:10.1016/j.neucom.2014.07.074.
- [8] E. Castillo Guerra, 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. doi:10.1109/IEMBS.2003.1280248.
- [9] R. Norel, M. Pietrowicz, C. Agurto, S. Rishoni, G. Cecchi, Detection of amyotrophic lateral sclerosis (ALS) via acoustic analysis, in: *Proc. Interspeech*, 2018, pp. 377–381. doi:10.21437/Interspeech.2018-2389.
- [10] T. Spangler, N. V. Vinodchandran, A. Samal, 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. doi:10.1109/BHI.2017.7897299.

- [11] K. An, M. Kim, K. Teplansky, J. Green, T. Campbell, Y. Yunusova, D. Heitzman, J. Wang, Automatic early detection of amyotrophic lateral sclerosis from intelligible speech using convolutional neural networks, in: Proc. of Interspeech 2018, 2018, pp. 1913–1917. doi:10.21437/Interspeech.2018-2496.
- [12] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, L. O. Ramig, Novel speech signal processing algorithms for high-accuracy classification of Parkinson’s disease, IEEE Transactions on Biomedical Engineering 59 (5) (2012) 1264–1271. doi:10.1109/TBME.2012.2183367.
- [13] J. Gómez-García, L. Moro-Velázquez, J. Godino-Llorente, On the design of automatic voice condition analysis systems. part i: Review of concepts and an insight to the state of the art, Biomedical Signal Processing and Control 51 (2019) 181–199. doi:10.1016/j.bspc.2018.12.024.
- [14] R. J. Baken, R. F. Orlikoff, Clinical measurement of speech and voice, 2nd Edition, Singular Thomson Learning, 2000.
- [15] A. K. Silbergleit, A. F. Johnson, B. H. Jacobson, Acoustic analysis of voice in individuals with amyotrophic lateral sclerosis and perceptually normal vocal quality, Journal of Voice 11 (2) (1997) 222 – 231. doi:10.1016/S0892-1997(97)80081-1.
- [16] M. M. van der Graaff, W. Grolman, E. J. Westermann, H. C. Boogaardt, H. Koelman, A. J. van der Kooi, M. A. Tijssen, M. de Visser, Vocal Cord Dysfunction in Amyotrophic Lateral Sclerosis, JAMA Neurology 66 (11) (2009) 1329–1333. doi:10.1001/archneurol.2009.250.
- [17] Y. Yunusova, J. S. Rosenthal, J. R. Green, S. Shellikeri, P. Rong, J. Wang, 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, 2013, p. 217–220.
- [18] M. V. Mujumdar, 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. doi:10.1109/ICASSP.2010.5495563.
- [19] A. Illa, D. Patel, B. Yaminiy, S. Meera, N. Shivashankar, P.-K. Veeramaniz, S. Vengalilz, S. N. K. Polavarapuz, A. Naliniz, P. K. Ghosh, Comparison of speech tasks for automatic classification of patients with amyotrophic lateral sclerosis and healthy subjects, in: Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 6014–6018. doi:10.1109/ICASSP.2018.8461836.
- [20] A. Bandini, J. Green, L. Zinman, 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. doi:10.21437/Interspeech.2017-478.

- [21] J. Lee, M. A. Littlejohn, Z. Simmons, Acoustic and tongue kinematic vowel space in speakers with and without dysarthria, *International Journal of Speech-Language Pathology* 19 (2) (2017) 195–204. doi:10.1080/17549507.2016.1193899.
- [22] M. Vashkevich, E. Azarov, A. Petrovsky, 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. doi:10.23919/SPA.2018.8563414.
- [23] M. Vashkevich, A. Gvozдович, Y. Rushkevich, Detection of bulbar dysfunction in als patients based on running speech test, in: *International Conference on Pattern Recognition and Information Processing*, Springer, 2019, pp. 192–204. doi:10.1007/978-3-030-35430-5_16.
- [24] I. C. Miller, 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)*, 2013, pp. 171–175.
- [25] H. Kasuya, S. Ebihara, T. Chiba, T. Konno, Characteristics of pitch period and amplitude perturbations in speech of patients with laryngeal cancer, *Electronics and Communications in Japan* 65 (5) (1982) 11–19. doi:10.1002/ecja.4410650503.
- [26] R. J. Moran, R. B. Reilly, P. de Chazal, P. D. Lacy, Telephony-based voice pathology assessment using automated speech analysis, *IEEE Transactions on Biomedical Engineering* 53 (3) (2006) 468–477. doi:10.1109/TBME.2005.869776.
- [27] P. Boersma, Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, in: *Proceedings of the institute of phonetic sciences*, Vol. 17, 1993, pp. 97–110.
- [28] J. R. Orozco-Arroyave, E. A. Belalcazar-Bolaños, J. D. Arias-Londono, J. F. Vargas-Bonilla, S. Skodda, J. Ruzs, K. Daqrouq, F. Honig, E. Noth, Characterization methods for the detection of multiple voice disorders: Neurological, functional, and laryngeal diseases, *IEEE Journal of Biomedical and Health Informatics* 19 (6) (2015) 1820–1828. doi:10.1109/JBHI.2015.2467375.
- [29] S. N. Awan, *Instrumental Analysis of Phonation*, John Wiley and Sons, Ltd, 2009, Ch. 21, pp. 344–359. doi:10.1002/9781444301007.ch21.
- [30] D. Michaelis, T. Gramss, H. W. Strube, Glottal-to-noise excitation ratio—a new measure for describing pathological voices, *Acta Acustica united with Acustica* 83 (4) (1997) 700–706.

- [31] X. Huang, A. Acero, H.-W. Hon, R. Foreword By-Reddy, Spoken language processing: A guide to theory, algorithm, and system development, Prentice hall PTR, 2001.
- [32] J. I. Godino-Llorente, P. Gomez-Vilda, M. Blanco-Velasco, Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters, *IEEE transactions on biomedical engineering* 53 (10) (2006) 1943–1953. doi:10.1109/TBME.2006.871883.
- [33] A. K. Dubey, S. R. M. Prasanna, S. Dandapat, Pitch-adaptive front-end feature for hypernasality detection, in: *Proc. Interspeech*, 2018, pp. 372–376. doi:10.21437/Interspeech.2018-1251.
- [34] R. D. Kent, G. Weismer, J. F. Kent, H. K. Vorperian, J. R. Duffy, Acoustic studies of dysarthric speech: Methods, progress, and potential, *Journal of Communication Disorders* 32 (3) (1999) 141–186.
- [35] B. Tomik, R. J. Guiloff, Dysarthria in amyotrophic lateral sclerosis: A review, *Amyotrophic Lateral Sclerosis* 11 (1-2) (2010) 4–15. doi:10.3109/17482960802379004.
- [36] G. S. Turner, K. Tjaden, G. Weismer, The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis, *Journal of Speech and Hearing Research* 38 (5) (1995) 1001–1013. doi:10.1044/jshr.3805.1001.
- [37] G. Weismer, R. Martin, R. D. Kent, J. F. Kent, Formant trajectory characteristics of males with amyotrophic lateral sclerosis, *The Journal of the Acoustical Society of America* 91 (2) (1992) 1085–1098. doi:10.1121/1.402635.
- [38] K. L. Lansford, J. M. Liss, Vowel acoustics in dysarthria: Speech disorder diagnosis and classification, *Journal of Speech, Language, and Hearing Research* 57 (2014) 57–67.
- [39] R. L. Horwitz-Martin, T. F. Quatieri, A. C. Lammert, J. R. Williamson, Y. Yunusova, E. Godoy, D. D. Mehta, 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*, 2016, pp. 1205–1209. doi:10.21437/Interspeech.2016-403.
- [40] J. Lee, E. Dickey, Z. Simmons, Vowel-specific intelligibility and acoustic patterns in individuals with dysarthria secondary to amyotrophic lateral sclerosis, *Journal of Speech, Language, and Hearing Research* 62 (1) (2019) 34–59. doi:10.1044/2018_JSLHR-S-17-0357.
- [41] M. Little, P. McSharry, E. Hunter, J. Spielman, L. Ramig, Suitability of dysphonia measurements for telemonitoring of parkinson’s disease, *Nature Precedings* (2008) 1–1doi:10.1038/npre.2008.2298.1.

- [42] E. Azarov, M. Vashkevich, A. A. Petrovsky, Instantaneous pitch estimation based on RAPT framework, in: Proc. of the 20th European Signal Processing Conference (EUSIPCO), 2012, pp. 2787–2791.
- [43] J. Peplinski, V. Berisha, J. Liss, S. Hahn, J. Shefner, S. Rutkove, K. Qi, K. Shelton, Objective assessment of vocal tremor, in: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 6386–6390. doi:10.1109/ICASSP.2019.8682995.
- [44] A. E. Aronson, W. S. Winholtz, L. O. Ramig, S. R. Silber, Rapid voice tremor, or “flutter,” in amyotrophic lateral sclerosis, *Annals of Otolology, Rhinology & Laryngology* 101 (6) (1992) 511–518. doi:10.1177/000348949210100612.
- [45] T. Nakano, M. Goto, 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.
- [46] M. Vashkevich, A. Petrovsky, Y. Rushkevich, Bulbar ALS detection based on analysis of voice perturbation and vibrato, in: Proc. of Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2019, pp. 267–272. doi:10.23919/SPA.2019.8936657.
- [47] H. Cordeiro, C. Meneses, Low band continuous speech system for voice pathologies identification, in: Proc. of the Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2018, pp. 315–320. doi:10.23919/SPA.2018.8563393.
- [48] P. Flach, *Machine Learning: The art and science of algorithms that make sense of data*, Cambridge University Press, United Kingdom, 2012.
- [49] R. Liu, D. F. Gillie, Feature selection using order statistics, in: Proc. of International Conference on Pattern Recognition and Information Processing (PRIP), 2011, pp. 195–199.
- [50] K. Kira, L. Rendell, A practical approach to feature selection., in: Proc. 9th Int. Conf. Mach. Learn., 1992, pp. 249–256.
- [51] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the royal statistical society* 58 (1994) 267–288.
- [52] I. Kononenko, E. Simec, M. Robnik-Sikonja, Overcoming the myopia of inductive learning algorithms with relieff, *Applied Intelligence* 7 (1) (1997) 39–55. doi:10.1023/A:1008280620621.
- [53] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA, 2001.
- [54] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proc. of International Joint Conference on Artificial Intelligence, 1995, pp. 1137–1143.