Bulbar ALS Detection Based on Analysis of Voice Perturbation and Vibrato

*Maxim Vashkevich[1], Yuliya Rushkevich[2], Alexander Petrovsky[*1]*

[1]Belarusian State University of Informatics and Radioelectronics, Minsk, Belarus

[2]Republican research and clinical center of neurology and neurosurgery, Minsk, Belarus

[*] Died from the ALS disease on 14 March 2019, aged 66.

Aim of the work

Goal

Development of feature extraction methods for detection of pathological changes in speech for the early diagnosis bulbar form of Amyotrophic Lateral Sclerosis (ALS).

Actual problems

1) early diagnosis of ALS;
2) monitoring of the ALS progression;
3) optimization of the efficacy of medicinal treatment of ALS

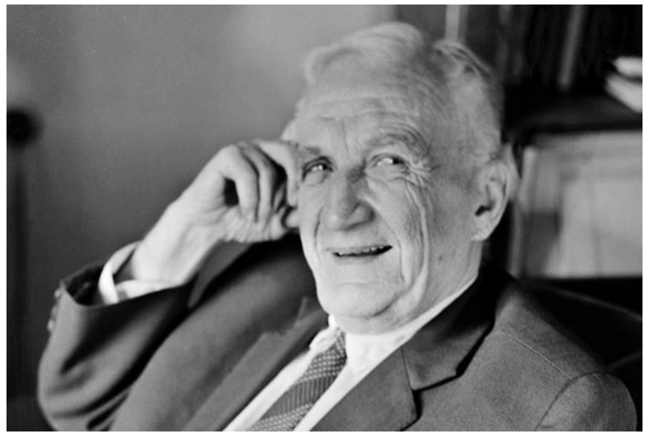
M. Zedong – first chairman of the People's Republic of China;
D. Shostakovich – Russian composer and pianist;
I. Tamm – Soviet physicist (Nobel Prize, 1958);
K. Nowak – Polish football player

Famous persons suffering from ALS

D.Shostakovich (1906-1975)

Mao Zedong (1901-1976)

Ihor Tamm (1895-1971)



Krzysztof Nowak (1975–2005)

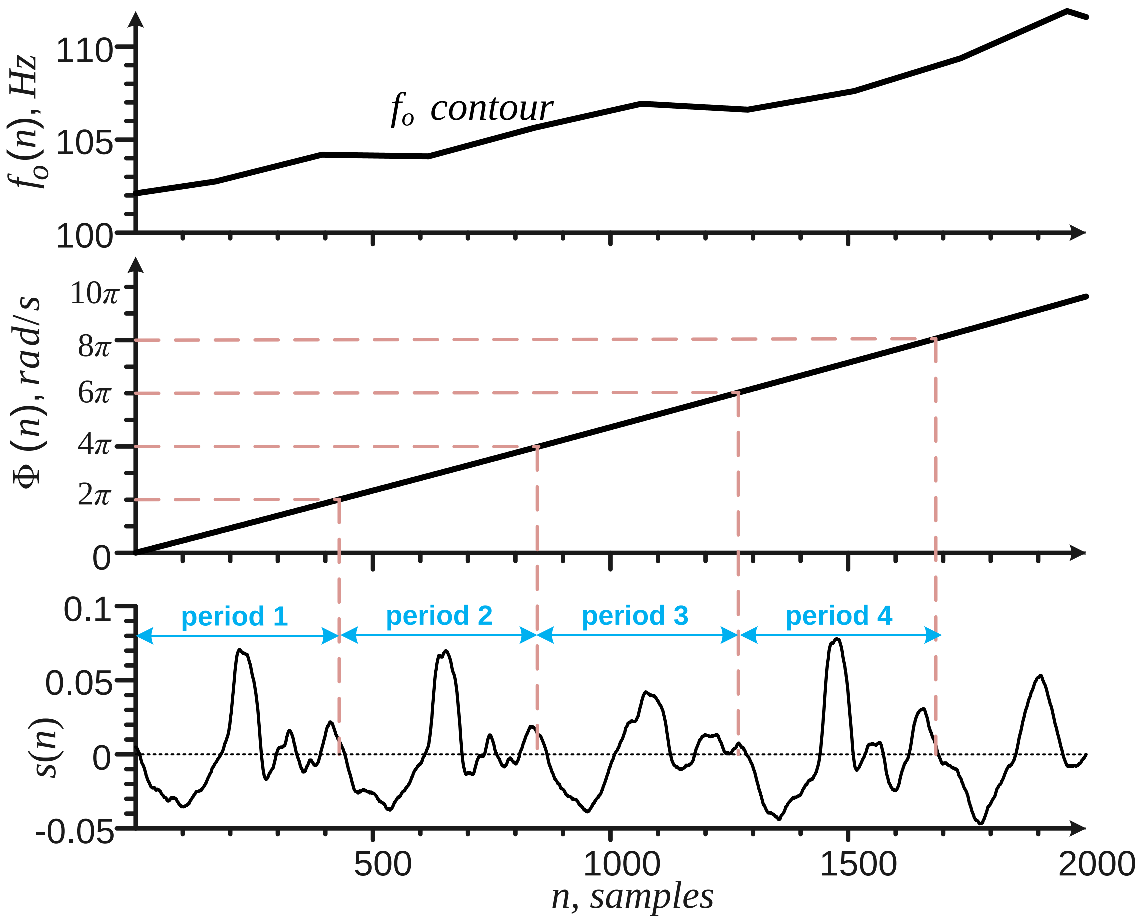
1. Voice perturbation analysis

We *verified the suitability* of the ***sustain vowel phonation*** test for ALS detection.

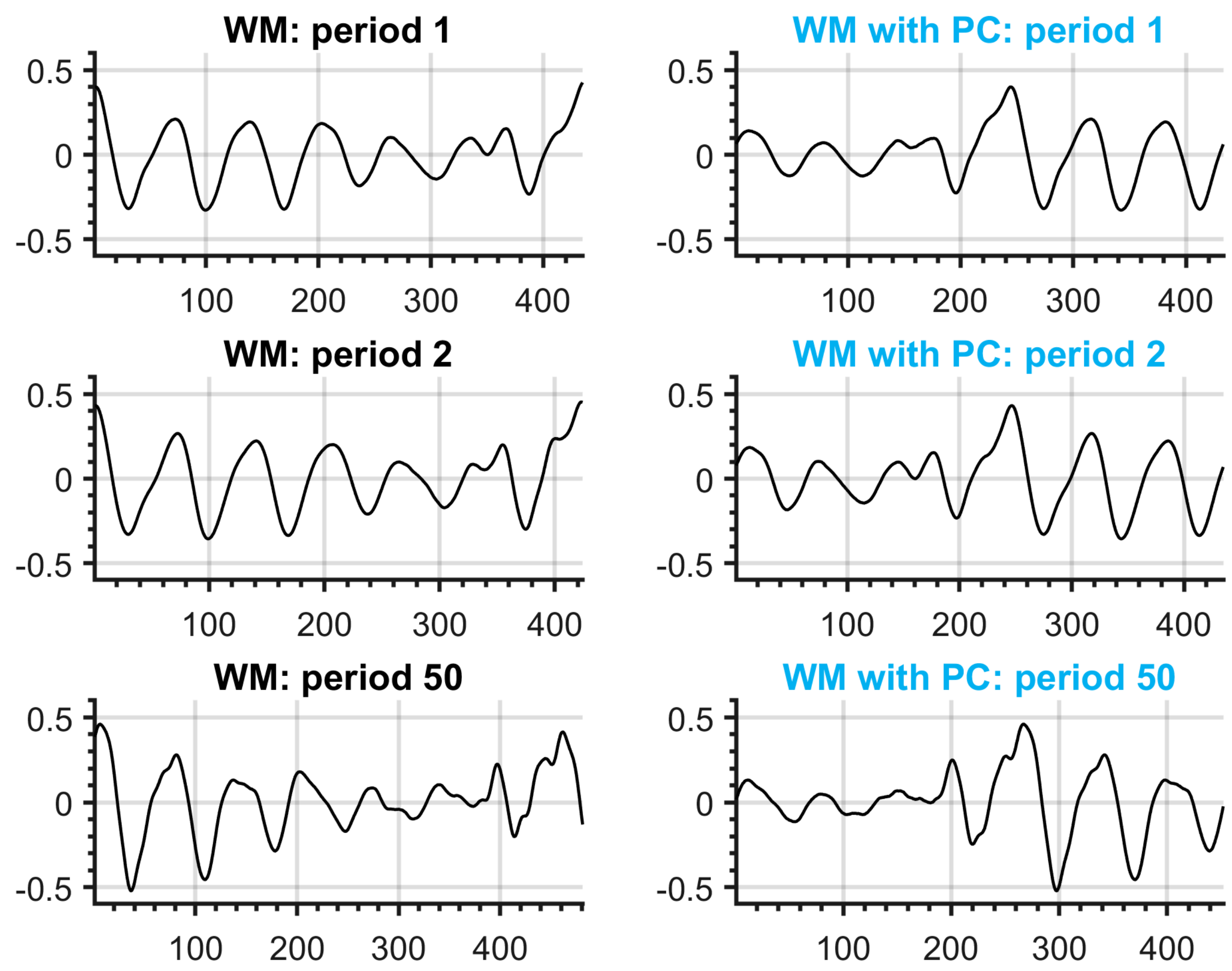
Waveform matching (WM) with phase constrain

In WM entire waveshapes are matching across adjacent cycles. The drawback of WM is that an error in detecting one period will affect all subsequent periods.

Proposition: *Use phase* $\Phi(n) = \sum_{k=1}^{n} \omega_o(k)$, *where* $\omega_o(k) = 2\pi f_o(k)/F_s$ *as a* ***reference signal*** *for period detection.*



WM with phase constrain: *extracted cycles are more synchronized compared to conventional WM.*



2. Amplitude and frequency perturbation

Jitter and Shimmer

Jitter *– estimates short term involuntary changes in* f_o.

Shimmer *– measure of short-term amplitude instability.*

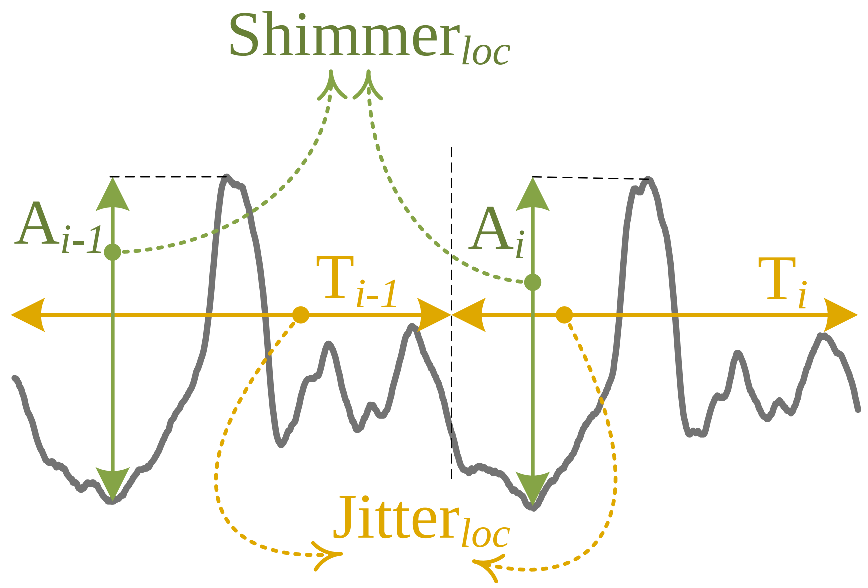
$$J_{loc} = J_1 = \frac{\text{mean}\,(|T_i - T_{i-1}|)}{\text{mean}\,(T_i)}$$

$$S_{loc} = S_1 = \frac{\text{mean}\,(|A_i - A_{i-1}|)}{\text{mean}\,(A_i)}$$

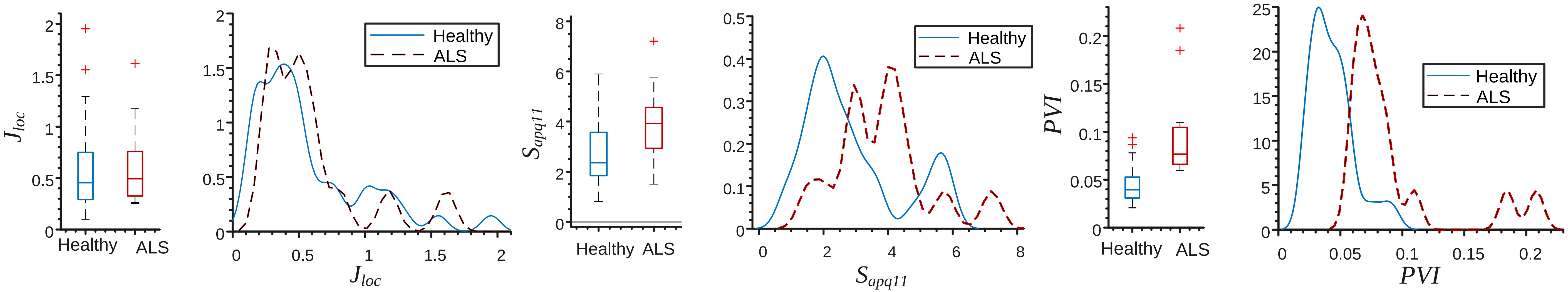
$$J_{rap} = J_3 = \frac{\text{mean}(T_i - \text{mean}([T_{i+1}, T_i, T_{i-1}]))}{\text{mean}(T_i)}$$

$$S_{apq3} = S_3 = \frac{\text{mean}(A_i - \text{mean}([A_{i+1}, A_i, A_{i-1}]))}{\text{mean}(A_i)}$$

...



Feature statistics



3. Pathology vibrato index (PVI)

Vibrato is a rapid, and regular fluctuation of the f_o that arises during sustained vowel phonation.

***Observation**: for healthy voices vibrato lies in range of 5-8 Hz, while for ALS patients characterized by presence of high-frequency components in 9-14 Hz range*

***Proposition**: In this study we use the following method of estimating pathological vibrato index (PVI)*

1) *Estimation and normalization of $f_o(m)$ contour:*

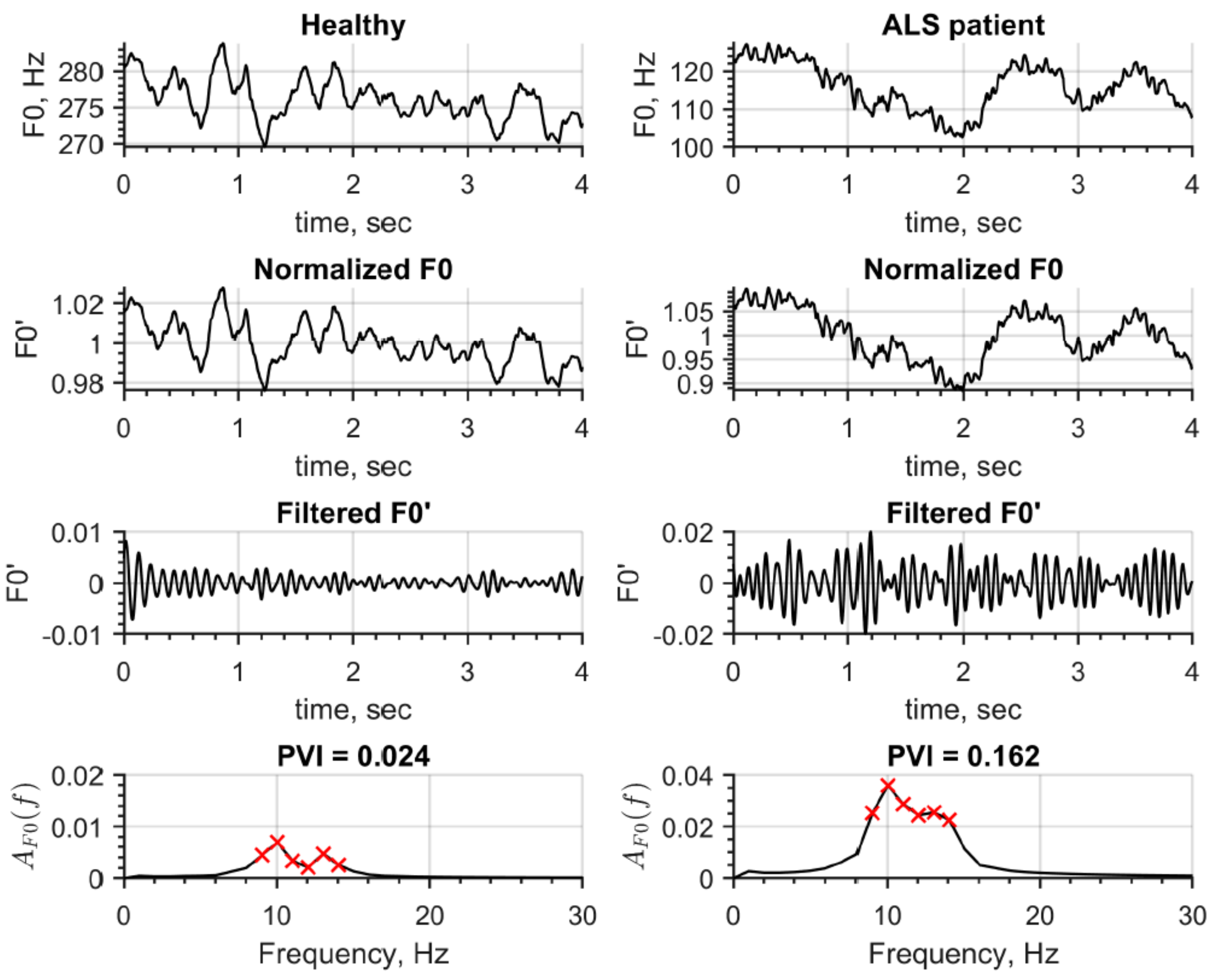
$$f_o'(m) = \frac{f_o(m)}{mean(f_o)}$$

2) *Filtering of $f_o'(m)$ using IIR filter with pass band [9; 14] Hz;*

3) *Amplitude spectrum $A_{f_o'}(f)$ estimation;*

4) *Calculation of **pathological vibrato index**:*

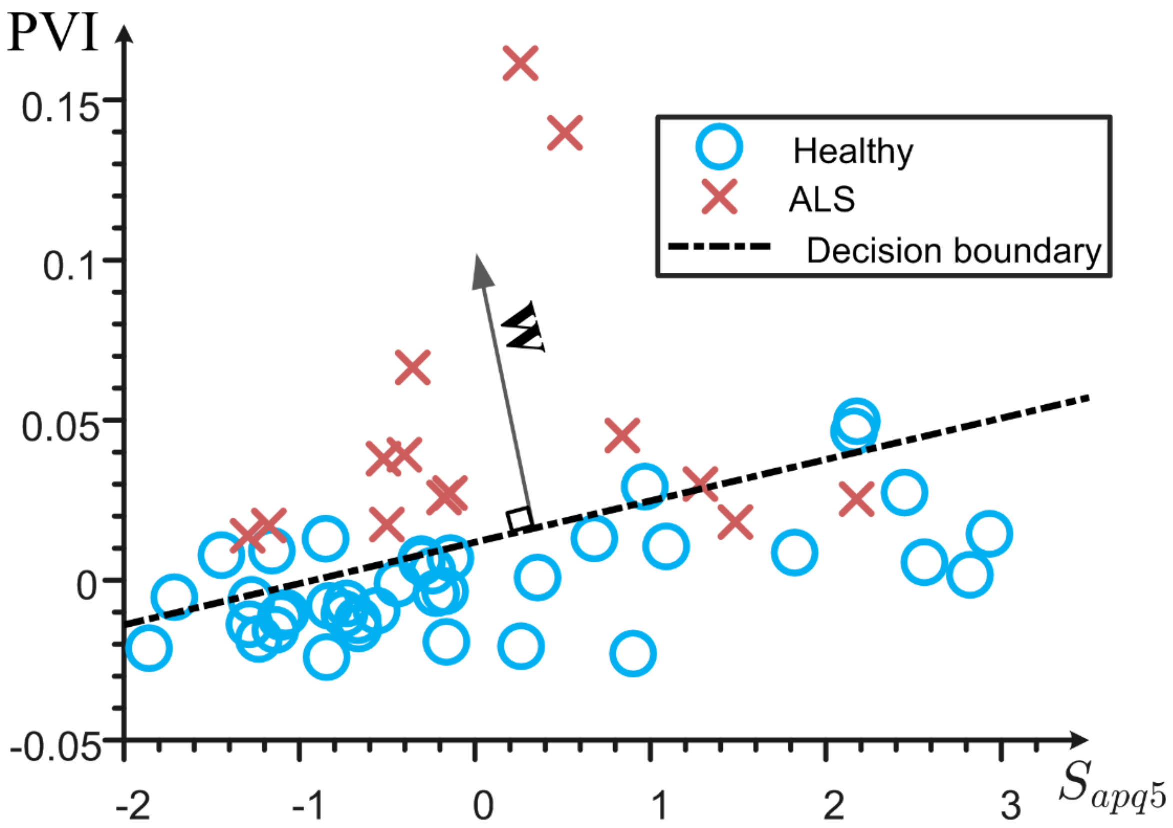
$$PVI = \sum_{f \in [9;\, 14]\ Hz} A_{f_o}(f)$$



4. Classification

Linear discriminant analysis (LDA)

Basic idea of **LDA**: selection of hyperplane ($\boldsymbol{w}$) in the feature space, such that the projection onto it *minimizes the within-class variation and maximizes the between-class variation*.

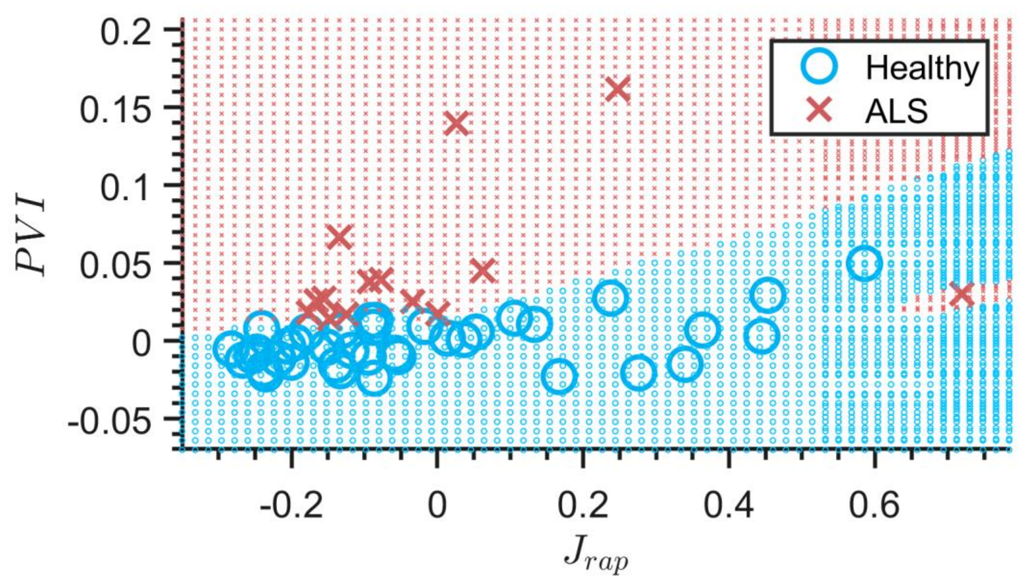


k-Nearest Neighbors (kNN)

kNN approach: to classify new sample K nearest points $\mathbf{x}^{+}_{1\ldots K}$ from positive class and K nearest points from negative class were determined. Then label is assigned based on *distances weighting*:

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

where $d(x, y)$ – Mahalanobis distance.



Cross-validation

Classification experiments are performed using k-fold cross-validation (CV) method. CV process was repeated 40 times, then mean and standard deviation values for the performance metrics of classifier were calculated.

5. Dataset description

– ***Number of speakers***: 54

Category	Male	Female	Total	Mean age	SD (age)
Healthy	23	16	39	41.9	16.3
ALS	6	9	15	57.7	9.0

All the participants were asked to produce the sustained vowel /a/ at a comfortable pitch and loudness as constant and long as possible.

Age effect elimination*: linear regression technique was applied to remove age effect using the data of the healthy group. The correction was applied to the data of healthy speakers and ALS patients.*

– ***Equipment***: The samples recorded at 44.1 kHz using smartphone with a standard headset and stored as 16 bit uncompressed PCM files.

Voice database is available in public GitHub repository:

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

6. Experiments

Classification performance measures

$$\text{Acc} = \frac{TP+TN}{TP+FP+FN+TN} \qquad \text{Sens} = \frac{TP}{TP+FN}$$

$$\text{Spec} = \frac{TN}{TN+FP} \qquad \text{R}_{\text{avg}} = \frac{1}{2}(\text{Sens} + \text{Spec})$$

TP – true positive, TN – true negative,
FP – false positive, FN – false negative

LDA classifier

Features	R_{avg}	Acc	Sens	Spec
	PRAAT features			
$[S_5\ S_{11}]$	83.1	81.1	83.0	83.1
$[J_3\ J_5\ S_1\ S_5\ S_{11}]$	81.4	82.8	78.3	84.6
	Feature based on WM with PC			
$[J_1\ J_5\ S_3\ S_{11}]$	86.0	86.4	85.0	87.0
$[J_5\ S_1\ S_5\ S_{11}]$	84.9	85.0	84.7	85.1
	Feature based on WM with PC + PVI			
$[S_1\ S_3\ S_{11}$ PVI $]$	**89.5**	**90.7**	**86.7**	**92.2**
$[J_5\ S_1\ S_3\ S_{11}$ *PVI* $]$	88.0	89.0	85.8	90.2

KNN classifier

Features	R_{avg}	Acc	Sens	Spec
	PRAAT features			
$[J_3\ S_1\ S_3\ S_5]$	76.0	84.6	56.8	95.3
$[J_1\ J_3\ S_1\ S_{11}]$	75.3	85.2	53.2	97.5
	Feature based on WM with PC			
$[J_1\ J_3\ J_5\ S_5\ S_{11}]$	81.0	87.3	66.8	95.1
$[J_3\ J_5\ S_5\ S_{11}]$	80.9	86.1	69.3	92.6
	Feature based on WM with PC + PVI			
$[J_1\ J_5$ PVI$]$	**86.9**	**91.6**	76.3	**97.5**
$[J_3$ PVI$]$	86.5	90.5	**77.7**	95.4