



## SPEECH AND TREMOR TESTER - MONITORING OF NEURODEGENERATIVE DISEASES USING SMARTPHONE TECHNOLOGY

Maurycy CHRONOWSKI\*, Maciej KLACZYŃSKI\*, Małgorzata DEC-ĆWIEK\*\*,  
Karolina POREBSKA\*\*, Katarzyna SAWCZYŃSKA\*\*

\* AGH University of Science and Technology, Faculty of Mechanical Engineering and Robotics  
Adama Mickiewicza 30, 30-059 Kraków, Poland,

e-mail: [mchronow@student.agh.edu.pl](mailto:mchronow@student.agh.edu.pl), [maciej.klaczynski@agh.edu.pl](mailto:maciej.klaczynski@agh.edu.pl)

\*\*Department of Neurology, Jagiellonian University, Collegium Medicum  
Botaniczna 3, 31-503 Kraków, Poland,

e-mail: [malgorzata.dec-cwiek@uj.edu.pl](mailto:malgorzata.dec-cwiek@uj.edu.pl), [karolina.porebska@uj.edu.pl](mailto:karolina.porebska@uj.edu.pl),  
[katarzyna.sawczynska@doctoral.uj.edu.pl](mailto:katarzyna.sawczynska@doctoral.uj.edu.pl)

### Abstract

One of the most frequently diagnosed neurodegenerative disorders, along with Alzheimer's disease, is Parkinson's disease. It is a slowly progressing disease of the central nervous system that affects parts of the brain which are responsible for one's motor functions. Despite the frequency of its occurrence among the elderly population, there has not yet been established a universal approach towards its certain diagnostics *ante mortem*. The study presents a pilot experiment regarding the assessment of the usefulness of simultaneous processing and analysis of speech signal and hand tremor accelerations for patient's screening and monitoring of the progress in healing, using the data acquired with a mid-range Android smartphone. During the study, a mobile device of this kind was used to record the patients of the Department of Neurology, University Hospital of the Jagiellonian University in Kraków and a control group of healthy persons over the age of 50. The samples were then analysed and an attempt towards classification was made using statistical methods and machine learning techniques (PCA, SVM, LDA). It was shown that even for a limited population, the classifier reaches about 85% accuracy. Another topic discussed in the study is the possibility of implementing a fully automated mobile system for the monitoring of the disease's progression. Propositions of further research were also drawn.

Keywords: Parkinson's disease, speech analysis, hand tremor, smartphone, diagnostics, machine learning

### TESTER MOWY I DRŻENIA – MONITOROWANIE PRZEBIEGU CHOROÓB NEURODEGENERACYJNYCH Z WYKORZYSTANIEM SMARTFONA

#### Streszczenie

Jednym z najczęściej diagnozowanych zaburzeń neurodegeneracyjnych, obok choroby Alzheimera, jest choroba Parkinsona. To wolno postępująca choroba zwyrodnieniowa ośrodkowego układu nerwowego, która zajmuje obszary mózgu odpowiedzialne za motorykę. Pomimo powszechności choroby wśród osób starszych, do tej pory nie została opisana uniwersalna metoda jej pewnego zdiagnozowania. Praca przedstawia pilotażowe badanie dotyczące określenia przydatności i możliwości wykorzystania metod jednoczesnego przetwarzania i analizy sygnału mowy oraz sygnału przyspieszenia drgań kończyny górnej w kontekście badań przesiewowych lub obiektywnego monitorowania postępu leczenia chorób neurodegeneracyjnych, z wykorzystaniem danych pozyskanych za pomocą średniej klasy smartfona z systemem Android. W ramach badania wykonano za pomocą urządzenia mobilnego nagrania pacjentów Oddziału Neurologii Szpitala Uniwersyteckiego w Krakowie ze zdiagnozowaną chorobą Parkinsona oraz osób zdrowych powyżej 50 roku życia. Próbkę poddano analizie i wstępnej klasyfikacji z wykorzystaniem metod statystycznych oraz technik uczenia maszynowego (PCA, SVM, LDA). Pokazano, że skuteczność klasyfikacji już dla niewielkiej populacji sięga około 85%. W pracy omówiono również możliwość implementacji w pełni automatycznego systemu mobilnego monitorowania przebiegu choroby, a także przedstawiono propozycję dalszych badań w tym kierunku.

Słowa kluczowe: choroba Parkinsona, analiza mowy, drżenie, smartfon, diagnostyka, uczenie maszynowe

## 1. INTRODUCTION

With ageing societies and global life expectancy constantly increasing, new challenges arise. Age-

related diseases became a significant risk and it is predicted that the global elderly population is going to more than double over the next three decades [21]. In the area of neurodegenerative disorders,

Alzheimer's and Parkinson's diseases show to be the most frequently occurring [8].

### 1.1. Parkinson's disease

This paper covers a study on a group of Parkinson's disease (PD) patients. It is a progressive disorder of the nervous system that affects one's motor functions. Its symptoms include resting tremor with the frequency of about 3-6 Hz [16, 6], bradykinesia and muscle rigidity. Parkinson's disease diagnostics are challenging and a universal method for identifying patients has not yet been established. Moreover, with the rate of recognition around 80%, many PD patients are being erroneously distinguished, making the disease overdiagnosed on a large scale [12].

Several rating methods can be used for describing severity and progression of PD. In this study, Hoehn and Yahr scale (HY) was used. It is a scale that was first proposed in 1967 and is widely used also today, mainly for its easiness of application. Although being less precise than e.g. UPDRS scale, HY is a convenient way to quickly describe the patient's state [13], which is sufficient for the purpose of this study. The HY scale rates the patient's condition using numbers from 1 to 5, where 1 means the mildest symptoms and 5 is used when the patient is confined to a bed or a wheelchair.

### 1.2. Motivation

For reasons presented above, new methods for the diagnostics of PD and other neurodegenerative diseases ought to be found. The following study presents an approach towards using a mid-range smartphone with its built-in microphone and sensors to diagnose PD patients against the healthy population.

Numerous studies over the past two decades have shown that there is a significant potential in using speech signal for PD diagnostics, with classification accuracy reaching almost 100% [1, 2, 3, 6]. The influence of PD on human speech is widely documented. Its symptoms include voice tremor and stuttering, and the progression of the disease may render one's speech to be completely unintelligible. In case of the tremor, it is less researched in terms of automated diagnostics, but some approaches were already made, e.g. by using touch pad to measure disorders in writing [4] or by using mobile devices to monitor gait freeze [5] and hand vibrations with the phone attached to the hand [15].

Currently produced smartphones are equipped not only with a microphone, but also a whole set of sensors that could be used also for the purposes of diagnostics. It is also shown in global statistics that the number of smartphone users around the world is increasing, with over 3 billion people using them nowadays [19]. Around 72% of these devices run on the Android OS, and almost every remaining one is iOS [18]. That is why an approach to use

both audio and acceleration signals recorded using a mid-range Android device, is proposed in this study. If accurate models for speech and tremor diagnostics based on the smartphone input would be obtained, they could be then used for remote monitoring of the disease. Such systems could benefit both doctors and patients in the long term.

## 2. EXPERIMENT OUTLINE

The conducted study consisted of several steps:

- I. First, a mobile application for Android OS was developed for the purpose of data acquisition. It was designed to record audio and 3-axis acceleration data. The application was written in Kotlin language in Android Studio environment.
- II. A population of persons diagnosed with PD was recruited from the patients of the Department of Neurology, University Hospital of the Jagiellonian University. The stage of disease's was assessed using HY scale, as mentioned earlier. A control group was gathered from volunteers over the age of 50 with no PD symptoms diagnosed or visible. The control group is referred to as HP (healthy population) further on in the paper.
- III. The phonation and kinetic tests were recorded with the participants. For the phonation part, each participant was reading out loud a set of polish syllables and sentences [9]. The kinetic part involved a set of simple arm movements. Both tests are described in more detail in the following chapters.
- IV. Recordings were segmented using audio processing software. Both signals were analysed using scripts written in Python 3.7 programming language with open-source packages including: NumPy, SciPy, scikit-learn and statsmodels. Seaborn and matplotlib packages were used for the visualisation of the data.
- V. Classification was performed using principal component analysis (PCA), support-vector machines (SVM) and linear discriminant analysis (LDA).
- VI. The classifiers were compared in terms of accuracy and performance. Conclusions were drawn with the propositions of further research.

The aim of the experiment was to show the possible potential of using smartphones for the diagnostics and monitoring of neurodegenerative diseases and assess the limitations of such approach. It is also a pilot study in terms of using both acoustic and vibration signals simultaneously for the purpose of PD diagnostics.

## 3. DATA COLLECTION

The testing procedure with description of the tasks, as well as format, quality and amount of the data gathered during the study are presented below.

### 3.1. Android application and file format

For the purpose of this research, the application was used only to record the data. The possible implementation of the classifier in an end-to-end application is discussed later in the paper. The interface of the application was designed to be as simple as possible, allowing the researcher to input the patients ID code and use REC/STOP button to start or finalise the recording. Application screen is shown in Figure 1.

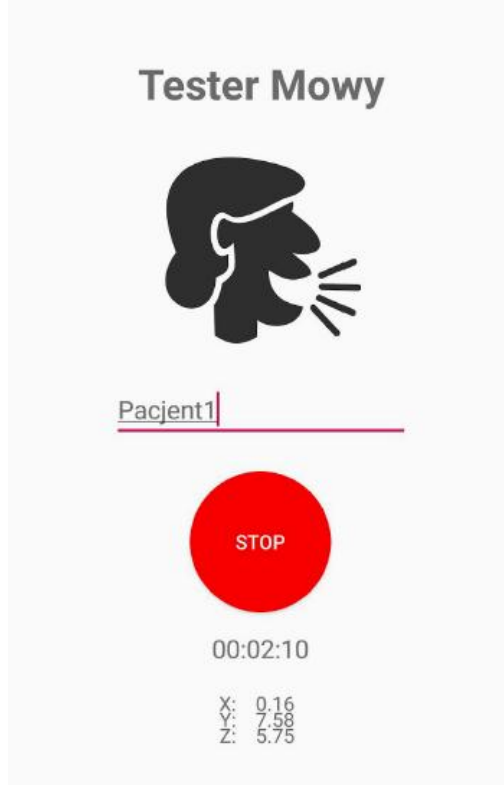


Fig. 1. Screenshot of the application's main screen.

The application also shows indicators for the accelerations on each axis, so the researcher is able to check whether the sensors are working properly. Sound and vibration data are recorded simultaneously, so the audible effects may be later correlated with the corresponding changes in accelerations.

The audio signal is stored in 16-bit PCM .wav format with sampling frequency of 44.1 kHz and 1411 kb/s bitrate. One channel of the recording is the phone's main microphone input; the other one comes from the noise cancelling microphone. Noise cancellation in smartphones is explained in section 1.4. Accelerations are written into a .txt file in three columns, each one representing one of the three-dimensional axes. A timestamp is placed at the beginning and at the end of each file, allowing the calculation of the sampling frequency. The application always tries to use the fastest sampling possible on a given device. For the model used (Huawei P8) this sampling frequency was 102 Hz.

Audio recordings are referred to as 'audio samples' and acceleration recordings as 'tremor samples' further in the paper.

### 3.2. Testing procedure

Each participant was recorded individually, under the supervision of one of the researchers. The phonation test consisted of reading a set of phrases, which were presented on a laptop screen. Before the recording, participants were instructed about the tasks and also size of the font on the screen was adjusted, so that the text was easily readable. The main focus of this part was to record sustained phonation of vowels /a/, /e/, /i/, /u/, which proved to give good results in previous studies [3,9]. Each of the sustained vowels was repeated three times. Aside from that, a set of simple syllables, words and sentences in Polish language was recorded for further analysis.

The kinetic test consisted of two tasks, one aiming to measure the intention tremor and the second one focused on the resting tremor. The intention task involved holding the device in one hand with both arms extended frontally in the air. Then, participants were asked to move the hand holding the device towards the head, next interchange this movement with the other arm, and then repeat the whole sequence one more time. For the resting part, participants were dropping both hands onto their laps and trying to hold them steady for 10 seconds.

Most of the times, participants were holding the device in one of their hands throughout the whole procedure. In some cases however, it was impossible for patients with severe resting tremor to hold it steady during the phonation part. When this was the case, the phone was placed on the table in front of the patient and an adequate note was written in the research log.

### 3.3. Amount of the collected data

During the course of the research, the total number of 49 recordings was made, including data as listed below.

- 37 proper PD patient audio samples
- 24 proper PD patient tremor samples
- 10 proper HP audio samples
- 10 proper HP tremor samples
- 1 noisy audio sample
- 1 repeated audio sample

The disproportion in the audio and tremor samples in PD population stems from the fact that some of the audio was recorded earlier, before the vibration recorder got implemented into the application. One audio sample was rejected, as it turned out that the patient was patting directly on the microphone case throughout the recording, resulting in a noisy signal. Also, a pair of recordings turned out to come from the same patient, who mistakenly volunteered for the experiment twice. In this case, only one of these two samples was used.

### 3.4. Quality of the collected data

Each of the recordings was made using the same Huawei P8 smartphone in order to keep a unified distribution of the data. The environment for the recordings included various locations in the hospital and at the university. Therefore, not all of the samples are perfectly free from background noise.

Modern smartphones in most cases use advanced systems for noise cancellation during calls and speech recognition [20]. One of such improvements was also implemented in this research. The underlying idea is that smartphones are nowadays equipped with a set of two MEMS microphones, one of them being located at the bottom of the device, close to the speaker's mouth, and the other one pointing outside, in the opposite direction. Therefore, the bottom microphone captures more of the actual speech, and the rear one is focused more on the environmental sounds. Two signals recorded from such a pair can be then subtracted from each other, cancelling out some of the unwanted sound. An example of this noise cancelling procedure is presented on a fragment of signal in Figures 2 and 3.

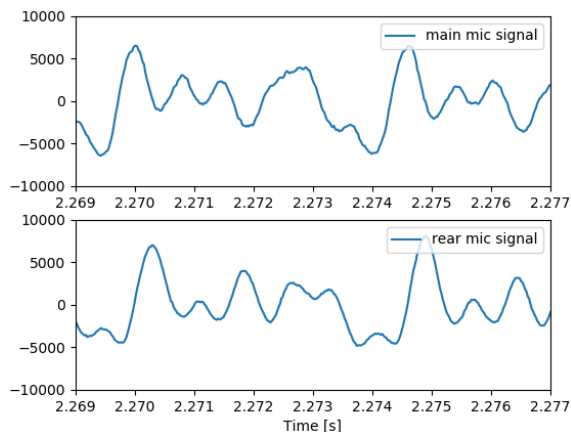


Fig. 2. Fragment of the same audio clip recorded with main microphone (upper figure) and rear microphone (lower figure).

Subtracting the rear microphone signal from the main one, provides a signal where higher frequency noise becomes attenuated compared to the stronger peaks visible in the main microphone time series. The result is presented in Figure 3.

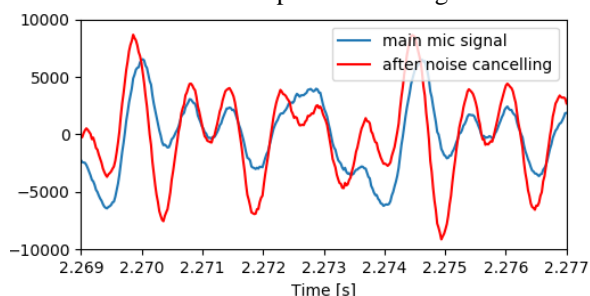


Fig. 3. The same audio clip as shown in Figure 2, before and after subtracting the rear microphone signal.

It is important to note that participants were instructed to hold the device in front of their face in a usual manner. If it was not possible for a person to hold the phone during the phonation test, it was then placed on the table in front of the participant, as described in the testing procedure. So, even though the main microphone was not always located directly next to the speaker's mouth, it was still possible to take the advantage of the noise cancellation presented above.

## 4. METHODS OF ANALYSIS

### 4.1. Feature extraction

Audio samples were first segmented by hand using a DAW<sup>1</sup> software. For the purpose of this study, a set of sustained phonation vowels was extracted from each sample. There were 4 vowels recorded, each repeated 3 times, resulting in 12 audio files for every person. Only the middle 80% of the signal (so just the clear phonation period) was taken for analysis. Tremor samples were segmented in a similar manner, by marking starting and ending points of every test.

Then, a set of speech parameters was extracted to construct a dataset for the classifier. The parameters are listed below:

- Jitter: a parameter describing the variability of the fundamental frequency (F0) during phonation. First, autocorrelation function of the signal was calculated for analysing periodicity. Then, through the analysis of strongest peaks in the autocorrelation series corresponding to frequencies from the range of 50 to 500 Hz, the mean derivation from the average F0 was calculated. The use of the autocorrelation function in this case made it not necessary to use a narrow bandpass filter for finding the value of the fundamental frequency. The obtained value of F0 was also added to the output dataset.

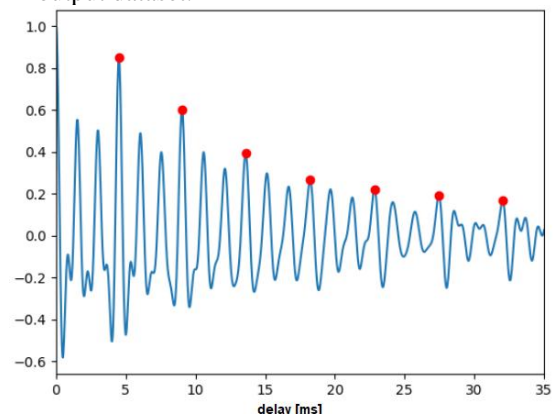


Fig. 4. An example of peak analysis of the autocorrelation function for determining periodicity. The red dots represent peaks corresponding to the fundamental frequency.

<sup>1</sup> Digital Audio Workstation

- Shimmer: a parameter describing the variability of the fundamental frequency amplitude during phonation. For calculating this parameter, audio samples were first filtered with a bandpass FIR filter with the passband between  $\pm 75\%$  of the F0 obtained in the previous step.
- MFCC (mel-frequency cepstral coefficients): a set of coefficients used widely in speech recognition systems. To calculate MFCC, the preemphasis is first applied to the signal to improve SNR<sup>2</sup> in high frequency bands. Signal is then windowed using a Hamming window with length of 25 ms, which was proved to be successful for analysing human speech [14,11]. Next, the squared magnitude of the Fourier transform is taken, a mel filter bank is applied, and the logarithm of this spectrum is transformed using the discrete cosine transform. This results in a cepstrum (inverse of spectrum) in a quefrequency domain, with values of this cepstrum being MFCC. In this study, 13 MFCC parameters were calculated for every time window and they were then averaged over the audio sample to represent the sustained phonation.
- Spectral moments: parameters used in machine diagnostics [10] and in some speech recognition systems to determine phonemes [7,11]. They describe the shape of the spectrum. Moment of  $m$ -th order is calculated according to the equation:

$$M(m) = \sum_{k=0}^{\infty} |G(k)| [f_k]^m \quad (1)$$

where  $G$  is the discrete spectrum of the signal and  $f_k$  is the centre frequency of the  $k$ -th frequency band. Moment  $M_0$  is used for normalising the higher-order spectral moments. Normalised moment  $M_{Inorm}$  may be interpreted as the 'centre of mass' of the spectrum (weighted average frequency) and is therefore used to calculate the normalised central moments of higher order. These may be then used to calculate the parameter called *kurtosis*, which is obtained by dividing 4-th normalised central moment by a square of the 2-nd. In this study spectral moments were used to:

- 1) extract more parameters from audio samples ( $M_{Inorm}$  and *kurtosis* were used)
- 2) describe tremor samples with  $M_{Inorm}$ , *kurtosis* and 3-rd normalised central moment, referred to as *skewness* of the spectrum.

For calculations of the spectral moments in tremor samples, the net acceleration was taken, i.e. the Euclidean norm was calculated from XYZ accelerations. The mean was also subtracted from the net value in order to remove the gravitational constant from the signal. Spectral moments were extracted from the spectral density of the net acceleration signal, calculated as the Fourier transform of the autocorrelation function. Example

time series of the intentional tremor tests are presented in Figure 5, and example spectral densities of the resting tremor tests are shown in Figure 6. Spectrums were analysed in the range from 0 to 20 Hz, as the tremor was expected to reside in low frequency bands [16]. The device's sampling frequency of 102 Hz was sufficient for this purpose, meeting the requirements of the Nyquist-Shannon theorem.

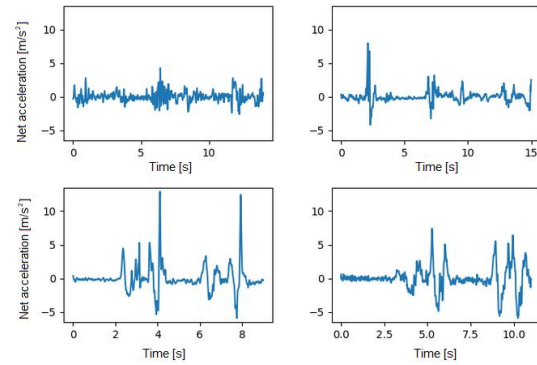


Fig. 5. A comparison of net accelerations during the intention tremor tests for: PD patients (upper row); HP (lower row).

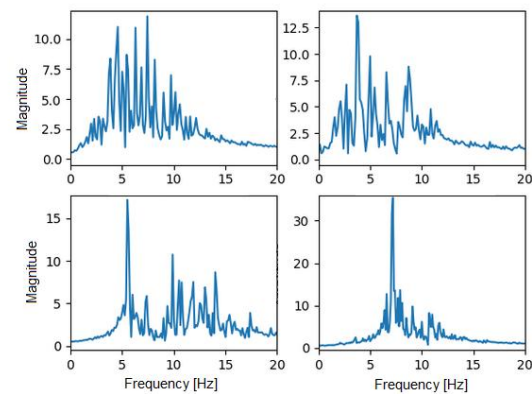


Fig. 6. A comparison of spectral densities of net accelerations during the resting tremor tests for: PD patients (upper row); HP (lower row).

## 4.2. Classification

A labelled dataset was constructed with the parameters presented above. Each example in the dataset was a vector of 78 input features (72 from audio samples and 6 from tremor samples). The examples were labelled in two ways: first, using 'PD' and 'HP' labels to represent PD patients and healthy population, and second one, using the HY rating (0 was used for HP). Three different classifiers were tested on the data, as listed below.

- Principal component analysis (PCA): a method used for feature reduction. Input features are transformed into a set of orthogonal, uncorrelated features. From these new features the ones with highest variance may be chosen.
- Support-vector machines (SVM): a supervised machine learning model which aims to find a hyperplane separating two classes

<sup>2</sup> Signal-to-noise ratio



in the data, based on a given number of support vectors and allowed cost margin. With the use of the ‘kernel trick’ it can be used to classify data with non-linear distribution.

- Linear discriminant analysis (LDA): a classifier based on the Fisher’s linear discriminant. It finds a linear combination of features that separates their classes possibly best. The number of transformed features is equal to the number of classes subtracted by one.

To properly train and test machine learning algorithms a sufficient number of training examples is necessary. However, even with limited data, it might be beneficial to implement these classification methods to assess the possible potential of the method.

For testing the classifiers three datasets were prepared. First one (labelled: A) consisting only of audio samples, second one (B) built only from examples where both audio and tremor tests were recorded properly, and third one (C) where the missing tremor data was filled with the mean value of a given parameter for the whole class.

A 2-component PCA was used to visualise the data on a 2D plane. SVM classifier was tested both on the whole feature set, as well as on the PCA-reduced data, which an approach that proved to give accurate results in other studies [17]. LDA transform was also used as another feature reduction and classification technique. Therefore 3 different classifiers were tested on each dataset (A, B, C):

- 1) SVM without feature reduction
- 2) SVM with 2-component PCA
- 3) LDA.

Before fitting each classifier, the data was also normalised to common mean and variance.

## 5. RESULTS

As described in the previous chapter, three datasets were used to train three classifiers each. For the SVM, the cost parameter was set to 10, as it gave best results in tests. Accuracy values presented in tables below were calculated as mean results of k-fold cross validation with the size of each fold set to 6 examples. The FNR parameter stands for False Negative Rate and was calculated as the ratio of the sum of false negatives (i.e. classifier predictions that the example comes from a healthy person, whereas it was a PD patient) to the total number of all positives in the train/test dataset. False negatives are an important issue considering the nature of the problem, which is also discussed later in the comments. Train accuracy describes how well the classifier fits to the training set; while the test accuracy shows how it performs on the data it hasn’t seen before (in other words: how it generalises the problem).

Table 1. Train accuracy and FNR scores obtained using cross validation.

Dataset / Classifier	Train accuracy [%]	Train FNR [%]
A / SVM	100.0 ±0.0	0.0 ±0.0
A / PCA + SVM	87.5 ±1.5	5.0 ±1.5
A / LDA	98.5 ±1.7	0.8 ±1.4
B / SVM	100.0 ±0.0	0.0 ±0.0
B / PCA + SVM	82.9 ±3.8	11.0 ±4.9
B / LDA	95.3 ±3.3	4.1 ±3.2
C / SVM	100.0 ±0.0	0.0 ±0.0
C / PCA + SVM	89.0 ±3.3	7.0 ±2.7
C / LDA	98.5 ±2.1	1.6 ±2.2

Table 2. Test accuracy and FNR scores obtained using cross validation.

Dataset / Classifier	Test accuracy [%]	Test FNR [%]
A / SVM	85.4 ±15.5	7.3 ±12.8
A / PCA + SVM	81.2 ±17.6	2.1 ± 5.5
A / LDA	70.4 ±10.6	21.2 ±18.1
B / SVM	73.6 ± 7.5	8.9 ±13.1
B / PCA + SVM	68.1 ±10.1	5.6 ±12.4
B / LDA	81.9 ±14.0	2.8 ± 6.2
C / SVM	78.3 ±14.7	12.3 ±14.0
C / PCA + SVM	82.1 ±19.3	4.6 ± 8.0
C / LDA	76.2 ±15.0	15.0 ±12.4

As it was shown in presented tables, a linear SVM classifier fits well to the data, but struggles to generalise to the test sets, scoring up to 85.4% of accuracy. Adding tremor data to the set does not improve performance in most cases, which suggests that either a different method of feature extraction should be used, or a separate model has to be trained. This is important, as it was shown previously in Figures 5 and 6 that the tremor data carries some distinct features that should be helpful for the classifier in discriminating samples. The use of PCA to reduce feature set before training the SVM classifier did not improve the accuracy, however, it reduced the FNR score, which also needs to be credited. The LDA classifier fits well to the data, but its test accuracy is worse than that of the SVM. It also should be noted that due to the small size of the population, resulting in a small number of training examples, the classifiers are vulnerable against outliers and adding new data points.

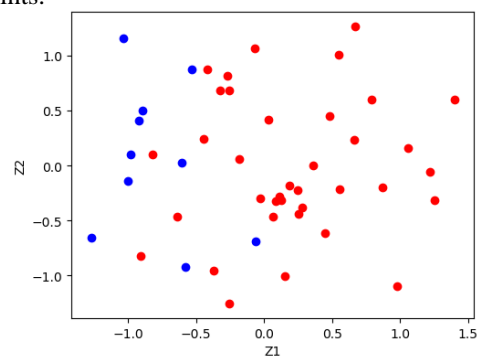


Fig. 7. Dataset A after PCA reduction to 2 components. Blue dots mark HP; red dots represent PD patients.

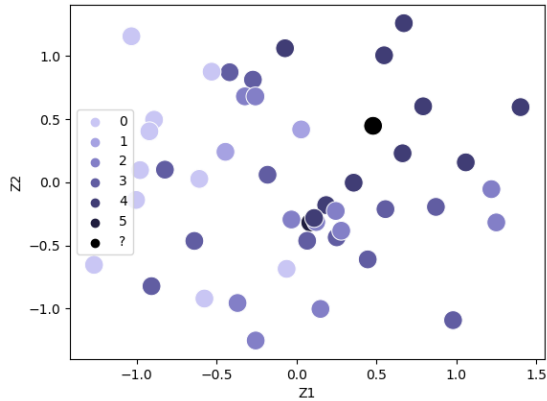


Fig. 8. Dataset A after PCA reduction to 2 components. Colours of the dots represent the HY rating, with 0 standing for HP.

It is worth noting that in Figure 8 darker markings, which correspond to higher stages of PD in the HY scale, have a slight tendency of shifting towards higher values of the extracted principal components. This suggests a possibility of using this approach to not only diagnose the disease, but also assess its severity. However, more balanced data with a representation of each stage ought to be used to train a multi-class classifier.

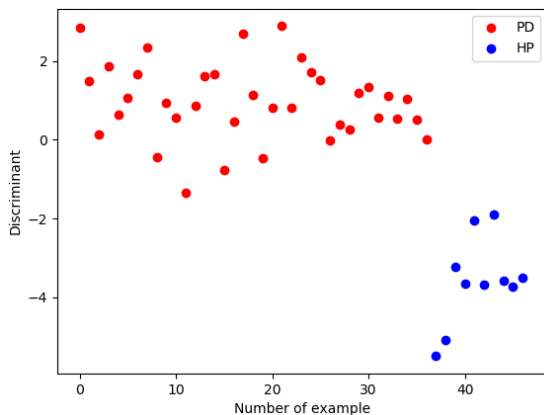


Fig. 9. Dataset A after LDA linear feature reduction.

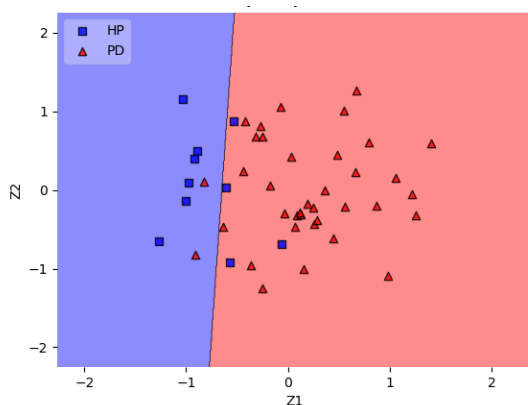


Fig. 10. Example of SVM decision boundaries for dataset A.

## 6. COMMENTS

The accuracy achieved using the presented approach is obviously not yet sufficient for proper medical diagnostics. However, even with this limited amount of data, some tendencies are already visible, and the SVM classifier reaching 85% accuracy in tests draws a promising start for the future research. It can be assumed that with a larger and more balanced dataset, the classification accuracy will increase. Low FNR score using the SVM approach is also a good sign, as for the system of telemedical diagnostics, it is better if errors turn out to be false positives, rather than false negatives (i.e. it is generally better to overdiagnose and suggest the user to consult his doctor, rather than miss an actual case of the disease). Also, a subtle trend can be found in the PCA-reduced data, which suggests that this approach might be used to diagnose the severity of the disease, and not just its presence, if provided a balanced dataset.

Regarding data acquisition, both audio and tremor signals recorded using a smartphone are of sufficient quality for time and spectral analysis. Even despite the fact that the recordings were taken in a not acoustic-insulated environment, the SNR remains on an acceptable level. On the contrary, there are many things to consider regarding the application interface and the testing procedure. Current approach relied on manual data segmentation, which is very time consuming and might be imprecise in some cases. An improvement in the interface allowing the researcher to input segmentation markings during the recording could help to solve this issue. Another thing worth commenting is that some patients required larger font of the text for the reading part, as their vision was partially impaired. An alternative way of presenting test tasks should be provided (e.g. using a text-to-speech system).

Future research should focus on acquiring more data to form a bigger and more balanced dataset (as classifiers are also vulnerable to high bias due to unbalanced data) and using it to provide a more end-to-end machine learning approach. The remaining speech data should also be used to feed the classifier.

## 7. CONCLUSIONS

During the experiment, speech and hand tremor tests were conducted on a group of Parkinson's disease patients in different stages of the disease's development and a control group of healthy persons over the age 50. The tests were recorded using a mid-range Android smartphone produced by a popular manufacturer. The recordings were segmented by hand and a set of features was extracted, describing the sustained phonation fragments of vowels /a/, /e/, /i/, /u/ and recorded intention and resting tremor tests. An attempt to classify the data using PCA, SVM and LDA

techniques was made. The classifiers scored up to 85.7% of test accuracy.

The experiment has shown that there are many challenges that might stand in the way of a fully automated disease monitoring system, such as data segmentation and an interface that's accessible for the elderly patients. Achieved classification results are far from optimal, however, they draw promising prospects if a bigger, more balanced dataset was provided and the model was trained to take full advantage of the speech samples and tremor data. The quality of the data recorded using the provided smartphone is sufficient for feature extraction and analysis. Therefore, it seems possible to one day implement a fully remote system of disease monitoring, which was the main question to answer at this point of the study.

### ETHICAL SUPERVISION

The research was approved by The Bioethics Committee of the Jagiellonian University (review no. 1072.6120.271.2019 of 21 Nov 2019).

### ACKNOWLEDGEMENTS

The work was created as part of research project of Department of Mechanics and Vibroacoustics no. 16.16.130.942/AGH and CMUJ in Kraków.

### REFERENCES

- Ali L, Zhu C, Zhou M, Liu Y. Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection. *Expert Systems with Applications*. 2019;137:22–8. <https://doi.org/10.1016/j.eswa.2019.06.052>
- Almeida JS, Rebouças Filho PP, Carneiro T, Wei W, Damaševičius R, Maskeliūnas R, et al. Detecting Parkinson's disease with sustained phonation and speech signals using machine learning techniques. *Pattern Recognition Letters*. 2019;125:55–62. <https://doi.org/10.1016/j.patrec.2019.04.005>
- Benba A, Jilbab A, Hammouch A. Voice assessments for detecting patients with Parkinson's diseases using PCA and NPCA. *Int J Speech Technol*. 2016;19(4):743–54. <https://doi.org/10.1007/s10772-016-9367-z>
- Bhattacharjee M, Bandyopadhyay D. Flexible Paper Touchpad for Parkinson's Hand Tremor Detection. *Sensors and Actuators A: Physical*, 2019. 294:164–72. <https://doi.org/10.1016/j.sna.2019.05.006>
- Capecchi M, Pepa L, Verdini F, Ceravolo M. A smartphone-based architecture to detect and quantify freezing of gait in Parkinson's disease. *Gait & Posture*, 2016.50. <https://doi.org/10.1016/j.gaitpost.2016.08.018>
- Gillivan-Murphy P, Miller N, Carding P. Voice Tremor in Parkinson's Disease: An Acoustic Study. *J Voice*. 2019, Jul;33(4):526–35. <https://doi.org/10.1016/j.jvoice.2017.12.010>
- Gottsmann F, Harwardt C. Investigating Robustness of Spectral Moments on Normal- and High-Effort Speech. *Interspeech*, 2011
- Han Z, Tian R, Ren P, Zhou W, Wang P, Luo M, et al. Parkinson's disease and Alzheimer's disease: a Mendelian randomization study. *BMC Med Genet*. 2018,19. <https://doi.org/10.1186/s12881-018-0721-7>
- Kłaczyński M. Vibroacoustic methods in diagnosis of selected laryngeal diseases, *Journal of Vibroengineering*. 2015; 17(4):2089-2098.
- Kłaczyński M., Wszolek T., Artificial Intelligence and Learning Systems Methods in Supporting Long-Term Acoustic Climate Monitoring, *Acta Physica Polonica A*, 2013; 123(6):1024-1028.
- Konior M, Kłaczyński M, Wszolek W. Reduction of speech signal deformation in patients after nasal septum surgery (septoplasty). *Acta Physica Polonica A*, vol. 119, issue 6A:1000-1004, 2011
- Marshall VL, Reininger CB, Marquardt M, Patterson J, Hadley DM, Oertel WH, et al. Parkinson's disease is overdiagnosed clinically at baseline in diagnostically uncertain cases: a 3-year European multicenter study with repeat [123I]FP-CIT SPECT. *Mov Disord*. 2009;24(4):500–8. <https://doi.org/10.1002/mds.22108>
- Martinez-Martin P, Skorvanek M, Rojo-Abuin JM, Gregova Z, Stebbins GT, Goetz CG. Validation study of the hoehn and yahr scale included in the MDS-UPDRS. *Movement Disorders*. 2018; 33(4):651–2. <https://doi.org/10.1002/mds.27242>
- Paliwal KK, Lyons JG, Wójcicki KK. Preference for 20–40 ms window duration in speech analysis. *4th International Conference on Signal Processing and Communication Systems*, 2010:1–4. <https://doi.org/10.1109/ICSPCS.2010.5709770>
- Printy BP, Renken LM, Herrmann JP, Lee I, Johnson B, Knight E, et al. Smartphone application for classification of motor impairment severity in Parkinson's disease. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2014: 2686–2689. <https://doi.org/10.1109/EMBC.2014.6944176>
- Reich SG. Does This Patient Have Parkinson Disease or Essential Tremor? *Clinics in Geriatric Medicine*, 2020;36(1):25–34. <https://doi.org/10.1016/j.cger.2019.09.015>
- Shahbakhti M, Taherifar D, Zareei Z. Combination of PCA and SVM for diagnosis of Parkinson's disease. *2nd International Conference on Advances in Biomedical Engineering*. 2013: 137–40. <https://doi.org/10.1109/ICABME.2013.6648866>
- StatCounter, website: <https://gs.statcounter.com/os-market-share/mobile/worldwide> (visited: 2020-04-29)
- Statista, website: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> (visited:2020-04-29)
- TechRadar, website: <https://www.techradar.com/news/phone-and-communications/mobile-phones/background-noise-reduction-one-of-your-smartphone-s-greatest-tools-1229667> (visited: 2020-04-29)
- UN Report, *World Population Ageing*, 2019.

Received 2020-02-04

Accepted 2020-05-12

Available online 2020-05-13





**Maurycy CHRONOWSKI**, B.Sc. Eng. in the field of Acoustic Engineering, student of Computer Science with specialisation in systems modelling and intelligent data analysis at the AGH University of Science and Technology. An enthusiast of modern technologies, who is passionate about

applications of computer intelligence in healthcare, medicine and research



**Maciej KŁACZYŃSKI** Ph.D.D.Sc. Eng. Prof. AGH, works at Department of Mechanics and Vibroacoustics in AGH University of Science and Technology in Krakow. His research is focused on measurement, signal processing and pattern recognition methods of vibroacoustic signals

applied in medicine, technology and environmental monitoring. Author of over one hundred scientific publications and conferences papers. Member of European Acoustics Association (EAA), Polish Acoustical Society (PTA) and Polish Society of Technical Diagnostics (PTDT).



**Malgorzata DEC-ĆWIEK** MD, PhD, graduated from Medical School at the Jagiellonian University in Krakow. She has a position of a consultant neurologist in the Department of Neurology at the Jagiellonian University Medical College. She specializes in

movement disorders. Specifically, she has been involved in neuromodulation procedures (deep brain stimulation, spinal cord stimulation) for several years. Her research is focused on Parkinson's disease, other parkinsonisms and dystonia. Member of Polish Neurological Society and International Movement Disorder Society.



**Karolina POREBSKA** MD Ph.D., neurologist, works in Neurology Department of University Hospital and in Jagiellonian University Collegium Medicum in Kraków. During studies she had practice in

Jerusalem, Arkhangelsk and was Erasmus student in Rome. Her research is focused on neurodegenerative diseases particularly on Parkinson's disease, dystonia and Alzheimer's disease. She is a member of movement disorders team and a specialist in treating advanced Parkinson's disease with Deep Brain Stimulation (DBS) and infusion therapies (Duodopa and Apomorphine). She has also experience in treating dystonia and spasticity with injections of botulinum toxin under ultrasound guidance.



**Katarzyna SAWCZYŃSKA**, MD, Graduate of Jagiellonian University Medical College (2017). Resident doctor in the Neurology Clinic of the University Hospital in Kraków and PhD student in Jagiellonian University Medical College.