Length Analysis of Speech to be Recorded in the Recognition of Parkinson's Disease

Attila Zoltán Jenei[1], Dávid Sztahó[1]

[1] Department of Telecommunications and Media Informatics,
Budapest University of Technology and Economics,
Magyar tudósok körútja 2, 1117 Budapest, Hungary
jeneia@edu.bme.hu, sztaho.david@vik.bme.hu

Abstract: Parkinson's disease is an incurable neurodegenerative disease to the present clinical knowledge. It is diagnosed mostly by exclusion tests. Numerous studies have confirmed that speech can be promising to suspect the presence of the disease. On the other hand, just a few researches discuss the appropriate length of the speech sample or the contribution of parts of the full-length recordings in the classification. Hence, we partitioned each original recording into four shorter samples. We trained linear and radial basis function (rbf) kernel Support Vector Machine (SVM) models separately for original recordings, each partitioned group and all partitioned samples together. We found no significant difference between the results of the rbf kernel models. However, we obtained significantly better results with a portion of the entire speech using linear kernel models. In conclusion, even a shorter piece of a longer speech may be adequate for classification.

1 Introduction

Parkinson's disease (PD) is one of the most common neurodegenerative diseases described first by James Parkinson (Parkinson, 1817). The prevalence of PD is about 1-2 cases per 1000 worldwide. Nonetheless, cases in the 60+ population can exceed 100 per 1000 (Tysnes & Storstein, 2017). The aetiology of the disease is unknown however predisposing environmental and genetic factors may play a role in its development (Lindgren et al., 2005).

Pathologically, the emergence of PD is caused by the death of dopamine-producing neurons in the Substantia Nigra brain region. In addition, abnormal aggregation of alpha-synuclein protein (Lewy-bodies) is also observed (Simon et al., 2020). Dopamine is a neurotransmitter that serves as a messenger substance among nerve cells in the brain. It plays an important role in many everyday behaviours, including how we move, feel, or eat. It regulates movement and also supports the reward system (Wise, 2004).

The importance of PD connected researches is given by the fact that it is incurable according to current clinical knowledge. With therapy, medication, or deep brain stimulation, symptoms can be relieved and progression can be slowed. Therefore, it means a life-long procedure for the patient (Armstrong & Okun, 2020). As a result, it has significant cost implications for both the patient and the treating institution (Denisova et al., 2020).

In Parkinson's disease, both motor and non-motor symptoms may occur. Motor symptoms may include bradykinesia, muscle rigidity, resting tremor, and postural instability. The first three of these are considered to be the primary indicator of the disease. In addition, non-motor symptoms such as olfactory impairment, sleep dysfunction, and cognitive impairment may occur (Armstrong & Okun, 2020).

Because of these symptoms and the reasons listed above, it is crucial to recognize the disease as soon as possible. The appropriate therapy and medication could maintain the patient's quality of life and limits the progression of his disease. This is exacerbated by the fact that many other diseases exist with similar symptoms. Furthermore, many times it is not easy for a patient to get to a neurological examination. Because no clear diagnostic procedure is available to detect the disease, in most cases, the patient's history, laboratory tests, and other examinations can help rule out other diseases (Reichmann, 2010; Tolosa et al., 2006).

Several kinds of research focus on non-invasive modalities for the recognition of Parkinson's disease, such as imaging procedures, movement, drawing/handwriting, and speech analysis. From these, our present research focuses on text read aloud (bounded speech). Furthermore, there is a lot of research on recognizing Parkinson's disease from different types of speech types (e.g., persistent vowels, sentences, or spontaneous speech). However, there is less analysis on parts of longer speech recordings (detailed in Section 2). Within this scope, the effect of different parts of the read text on PD recognition is examined in this article. These results are also compared to the full-length recordings.

The structure of the article is the following: in Section 2 the literature related to the research is described, in Section 3 the methodology applied to the research is presented, in Section 4 the results are given, and in Section 5 conclusions are drawn from the results.

2 Related Work

Due to the reduced amount of dopamine, nerve conduction has a limited ability to function. Based on this, PD also affects the process of speech production (dysphonia) (Rusz et al., 2011). The voice of people with Parkinson's disease is typically low-volumed, with a tremor-like character, sudden stops, and starts may be present (Schulz & Grant, 2000).

There has been and there still is ongoing research into the use of these phenomena in speech. These are done with four types of speech databases: 1) persistent vowel, 2) pronunciation of words, syllables, 3) reading of bound texts, 4) spontaneous speech.

The significance of the sustained vowel was given by the fact that the formation of a vowel requires the active work of the muscles of the vocal cords. In Parkinson's disease, there is uncertainty/stuttering in muscle movement. One of the most commonly used vowels are /a/ (Tsanas et al., 2012), (Hemmerling & Sztahó, 2019), /e/ and /i/ (Vaiciukynas et al., 2017).

The use of syllables and words is widespread in the study of imprecise consonant formation (for example, the pronunciation of /pa-ta-ka/). This can be used to examine the sudden stops/starts of speech (Novotný et al., 2014).

One disadvantage of the former modalities is that they do not provide information about the continuous movement and functioning of the articulation. This is why the use of reading text and spontaneous speech has become widespread (Frid et al., 2014), (Kiss et al., 2018).

Based on these, it can be seen that many speech modalities and speech lengths are used in the recognition of Parkinson's disease. In the following research, several sustained vowels, words, and sentences were recorded and examined with the same model setting (Vadovsky & Paralic, 2017). With a Random Forest approach, 52.5%, 57.5% and 45% accuracy were achieved with sustained vowel /a/, /o/ and /u/, respectively. 67.2% of accuracy was obtained with words and 65% with sentences. The examinations were performed on 20 Parkinson's disease patients and 20 healthy subjects using 26 speech-based features.

In the next study, sustained /a/ vowel, syllable repetition, words, short sentences, longer read text, and short spontaneous speech were examined separately (Sztahó et al., 2019). The research was performed using several speech descriptors involving 55 people with Parkinson's disease and 33 healthy individuals. Many classification algorithms have been tested. Of these, the linear kernel support vector machine (SVM) results are: 72.4% (sustained /a/), 77.0% (syllable triplets), 77.7% (words), 81.0% (sentences), 73.4% (read longer text) and 83.3% (short spontaneous speech) accuracy. Using radial basis function (rbf) kernel with the same SVM, the accuracy values are the following: 77.0% (sustained /a/), 78.2% (syllable triplets), 78.8% (words), 72.6% (sentences), 83.5% (read longer text) and 89.3% (short spontaneous speech).

From these results, prominent accuracy can be achieved with longer speech samples. As a form of long text is usually given as a phonetically rich tale, such as *The North Wind and Sun* or *Rainbow Passage*. However, there is no evident research in the literature that would examine the effect of different parts of these read texts for recognizing PD. Namely, if a part of the read text would provide similar recognition performance as the full-text version then only that part is sufficient to record. This would speed up the real examination procedure and would not be burdensome for the patient.

For the present study, the following hypotheses can be made according to the recent studies: Any part of the read text can achieve similar performance as the whole text.

3 Methodology

Fig. 1 illustrates the stepwise approach of the examination. First, speech recordings are selected from the speech database. In this case, Hungarian Parkinson Speech Database (HPSD) is used. Secondly, the full-length recordings were partitioned into four parts using segmentation files. Then, speech acoustic features were extracted from the full text and partitioned recordings. Finally, these features were the input for the classifier which conduct the label prediction after training.

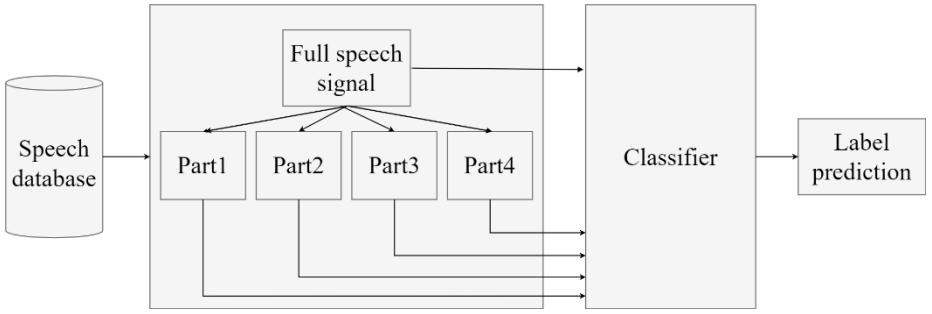

Fig. 1. Flowchart of the examination tasks: speech database, pre-processing (partitioning the full-length recordings), classification and prediction with the trained classifier.

3.1 Hungarian Parkinson Speech Database (HPSD)

79 recordings of patients with PD were selected from the HPSD. Speech acquisition was done in two health institutes in Budapest: Semmelweis University and Virányos Clinic. For the severity estimation, Hoehn & Yahr (H&Y) scale was used (Bhidayasiri & Tarsy, 2012). It defines non linearly progressive categories of motor functionality. Score 1 means unilateral involvement with minimal or no functional disability while 5 means the most severe case (bed or wheelchair bounded). Using such an estimator, 38 females (average age and standard deviation: 65±9.4) had an average of 2.8 (±1.1) and 41 males (average age and standard deviation: 64±9.3) had an average of 2.7 (±1.1) H&Y scores.

Speech of healthy population (HC) was also recorded using the same text and environmental properties. 41 females (average age and standard deviation: 53±15.9) and 38 males (average age and standard deviation: 52±16.4) were selected to balance the PD class sample size. Healthy subjects reported that they did not have any speech-related illnesses.

The text read was a tale called *The North Wind and the Sun*. External USB sound card (Terratec 6fire USB) with A/D converter was used to acquire recordings. A clip-on condenser microphone (Audio-Technika ATR3350) in a quiet office (or in a medical office) environment was applied. The recordings were stored in Pulse-code modulation (PCM) audio coding with a sampling frequency of 44.1 kHz and quantization of 16 bit. Subjects all consented to the use of their speech recordings for research purposes.

3.2 Speech Signal Pre-processing and Partitioning

The text of the entire tale contained 8 sentences, including also the title of the tale. This was chunked into 4 parts using Praat (Boersma & van Heuven, 2001) and annotation files available. This was done by examining the phonemes of the last word of the sentence before the place of cut. Since it was a bound text, these phonemes are in order and can be found and cuts can be performed. Using this method, four recordings were available for each person: 1) title and first sentence, 2) second and third sentences, 3) fourth longer sentence, and 4) fifth-seventh shorter sentences. Table 1 summarizes the average length and its standard deviation of recordings in seconds.

It should also be noted that no automatic chunking was made in the recording if the text was not read correctly (the phoneme set and order of the last words differed from the bound text ones). In that case, manual correction (manual chunk) was required.

After chunking the recordings, full and partitioned samples were resampled in 16 kHz sampling frequency and were normalized to the peak value in the signal. After that, the first 12 Mel Frequency Cepstral (MFC) coefficients were extracted from each sample using the Surfboard python library (Lenain et al., 2020). MFCCs were chosen according to the wide usage in the literature (Dasgupta et al., 2017; Godino-Llorente et al., 2017; Pompili et al., 2017). The mean, standard deviation, mean of the first derivate and standard deviation of the first derivate were calculated for the 12 MFCCs resulted in 48 features overall. Finally, the features were scaled between -1 and 1.

Table 1: The average length [sec.] and standard deviation of recordings separately for HC and PD classes.

	Full	**Part1**	**Part2**	**Part3**	**Part4**
HC	44.8±5.9	9.1±2.7	9.7±1.5	10.2±2.6	12.8±1.8
PD	59.1±23.4	13.3±4.0	10.6±3.6	15.6±4.9	14.1±4.5

3.3 Classification and model evaluation

The SVM classifier was chosen for this examination based on its widespread usage in speech-based researches (Pah et al., 2021; Sonawane & Sharma, 2021). SVM models were deployed in python (version 3.6) using *sklearn* machine learning module. Parameters were chosen as commonly used default values: C=1.0, gamma=$1/(n_{features} \cdot var_X)$ where $n_{features}$ is the number of features and var_X is the variance of the training data. For the experimentation, linear and rbf kernels were performed.

Six model training and testing scenarios were performed: (1) on full-length recordings, (2)-(5) on partitioned samples separately, (6) on partitioned samples together.

Speaker-wise leave-one-out cross-validation (LOOCV) was applied for model evaluation. With this method, samples of one speaker were grouped for testing and samples of the remaining speakers were grouped for training. This was repeated until each speaker in the dataset became test item. The performances were measured on the test items after the LOOCV. Speaker-wise separation was important for model (6), because multiple samples were available for each speaker. In this case, the majority

decision was resolved by having at least two positive samples (from four) to get a positive label for the test subject. To measure the performance of the models, sensitivity, specificity, accuracy, f1-score, and Matthews correlation were computed.

Furthermore, Mann–Whitney U non-parametric statistical test was performed to check the significance between the results of the models. With this test, we examined whether the performance of the model trained on the partitioned samples differ significantly from the performance of the model trained on the full recordings. Originally, the null hypothesis state that "*the two independent groups are homogeneous and have the same distribution*" (Nachar, 2008). So significant difference may be observed if the p-value of the statistics is lower than the significance level. So the alternative hypothesis can be stated as *the two independent groups are not homogeneous and have different distributions*. The significance level was set to 0.05, which is the most commonly used threshold in statistical studies.

4 Results

4.1 Result of Full-Length Recordings

For this experiment, SVM models with linear and rbf kernels were trained and tested with full-length recordings. The performance metrics can be seen in Table 2. The columns show the following metrics for both linear and rbf kernel classification: sensitivity ('sens'), specificity ('spec'), accuracy ('acc'), f1-score ('f1'), and Matthews correlation ('mc').

Table 2: Results of full-length recordings with linear and rbf kernels SVM models.

kernel	sens	spec	acc	f1	mc
linear	78.5%	79.7%	79.1%	79.0%	0.58
rbf	83.5%	84.8%	84.2%	84.1%	0.68

Based on Table 2, rbf kernel SVM achieved higher performance on all evaluation metrics. The average deviation on the metrics is 5.1%, while on the mc values it is 0.1. With an SVM with an rbf kernel, the algorithm recognized 8 speakers more correctly (true positive or true negative together) than using a linear kernel model.

4.2 Result of Partitioned Samples on Separate Models

For this examination, linear and rbf kernels SVM models were trained and tested with partitioned recordings separately. The performance metrics can be seen in Table 3. The columns represent the same metrics as in Table 2. *PartX* corresponds to the chunked segments (described in Section 3.2).

Based on Table 3, linear kernel models trained on part1 and part3 samples achieved the highest metric values (acc.: 86.1% and 88.0%, f1: 86.6% and 88.2%). Linear kernel SVM trained on part2 samples resulted in the lowest results (acc: 77.6%, f1: 77.7%). Using part4 samples with linear kernel SVM model, an intermediate result can be obtained (acc.: 81.6%, f1: 82.0%).

Table 3: Results of partitioned samples on separate SVM models with linear and rbf kernels.

kernel	samples	sens	spec	acc	f1	mc
linear	part1	89.9%	82.3%	86.1%	86.6%	0.72
	part2	78.1%	77.2%	77.6%	77.7%	0.55
	part3	89.9%	86.1%	88.0%	88.2%	0.76
	part4	83.5%	79.7%	81.6%	82.0%	0.63
rbf	part1	88.6%	83.5%	86.1%	86.4%	0.72
	part2	81.0%	83.5%	82.3%	82.1%	0.65
	part3	87.3%	84.8%	86.1%	86.2%	0.72
	part4	83.5%	83.5%	83.5%	83.5%	0.67

Compared to the whole recordings' results, linear models trained on the parts classified 11 (part1), 14 (part3) and 4 (part4) more samples correctly. In the case of part2, the model predicted 2 samples less correctly than the full-length linear model.

A similar tendency was earned with rbf kernel SVM with a narrower fluctuation in the results. Using part1 and part2 the highest performances were achieved (acc.: 86.1% and 86.1%, f1: 86.4% and 86.2%). With part2 samples, the lowest metrics were obtained (acc: 82.3%, f1: 82.1%) close to the result earned with part4 samples (acc.: 83.5%, f1: 83.5%). Compared to the whole recordings' results, the rbf models trained on parts classified 3 (part1 and part3) more samples correctly. In the case of part2 and part4, the models classified 3 (part2) and 1 (part4) fewer samples correctly than the full-length rbf model.

4.3 Result of Partitioned Samples Together in a Single Model

In this examination, all partitioned samples for each speaker were involved in the training and testing process. Overall, 628 samples' features of 157 subjects were used to train SVM models and one subject with four samples was used to test the performance in each cross-validation cycle. The results can be seen in Table 4. The metrics correspond to Table 3 and Table 2. MV (majority voting) means the aggregation of predictions per individuals and no MV means the results without speaker-wise aggregation.

Table 4: Results of partitioned samples together using linear and rbf kernel SVMs.

kernel	MV	sens	spec	acc	f1	mc
linear	MV	82.1%	81.3%	81.6%	81.5%	0.63
	no MV	82.6%	81.3%	82.0%	82.1%	0.64
rbf	MV	85.0%	85.9%	85.4%	85.5%	0.71
	no MV	83.5%	84.2%	83.9%	83.8%	0.68

Based on Table 4, the results of linear and rbf kernel SVM models are similar in accuracy, f1-score and Matthews correlation. The linear kernel SVM model achieved higher performance using multiple samples and majority voting than the original full-

length recordings model (Table 2). However, majority voting did not improve the results compared to the experiment without majority voting. With rbf kernel SVM, similar results were obtained with multiple samples and majority voting as in Table 2. In this case, a slight improvement can be realized with MV (from 83.9% to 85.4% (acc), from 83.8% to 85.5% (f1)).

The difference between the two results (no MV and with MV) is one speaker for the linear model and two speakers for rbf model. Furthermore, the samples clearly defined the category (HC or PD) for the individuals. This means that three or four of four-part samples were decided with the same class label. There was no indefinite decision (2 samples HC and 2 samples PD) neither for linear and nor rbf models. It also means that if the model categorized an individual to the wrong category, it did so for the most part and not because of an indecisive situation.

4.3 Significance Analysis Between Different Samples Trained Models

For each model, the output labels are given for the test elements as binary variables: 0 (sample is categorized to HC) and 1 (sample is categorized to PD). For all the test elements per model, a list of binary variables has resulted. That two lists can be compared to each other with an appropriate independent two-sample test. According to this, Mann–Whitney U non-parametric test was chosen to measure whether the models perform significantly to each other.

The p-values of corresponding models can be seen in Table 5 for linear kernel models and Table 6 for rbf kernel models. Notations part1, part2, part3 and part4 correspond to the partitioned samples' models, MV represents the majority version of all partitioned samples trained model, and full means the model trained on full-length recordings. An arbitrary row and column designate the p-value resulting from a comparison of the two models. Based on this property, the main diagonal would include the p-value of the same models, which is cleared out. Furthermore, the table is symmetric to the main diagonal. The p-values smaller from the significance level are highlighted.

Table 5: p-values of linear kernel models' performances.

	part1	part2	part3	part4	MV	full
part1	-	**0.026**	0.308	0.143	0.143	0.052
part2	**0.026**	-	**0.008**	0.189	0.201	0.376
part3	0.308	**0.008**	-	0.059	0.059	**0.017**
part4	0.143	0.189	0.059	-	0.500	0.286
MV	0.143	0.201	0.059	0.500	-	0.286
full	0.052	0.376	**0.017**	0.286	0.286	-

If the model trained and tested with full-length recordings is considered as a baseline (last row or last column). The model trained and tested with the third partitioned samples is different from this baseline significantly. This deviation is positive as the model on part3 achieved higher performance (acc.: 88.0%, f1: 88.2%) than the model on full-length recordings (acc.: 79.0%, f1: 79.1%). In contrast, a model trained on part2

obtained lower result (acc.: 77.6%, f1: 77.6%) significantly compared to the part1 (acc.: 86.1%, f1: 86.6%) and part3 (acc.: 88.0%, f1: 88.2%) partitioned samples based models.

According to Table 6, no significant difference can be observed between the models' performances. This implies that an arbitrary model can be deployed to recognize PD because it will result in the same performance metrics statistically as the others.

Table 6: p-values of rbf models' performances.

	part1	part2	part3	part4	MV	full
part1	-	0.178	0.500	0.266	0.437	0.318
part2	0.178	-	0.178	0.383	0.223	0.326
part3	0.500	0.178	-	0.266	0.437	0.318
part4	0.266	0.383	0.266	-	0.321	0.440
MV	0.437	0.223	0.437	0.321	-	0.378
full	0.318	0.326	0.318	0.440	0.378	-

5 Discussion and conclusion

Based on the literature, many speech modalities can be used for the recognition of Parkinson's disease. Each modality has its advantages and disadvantages. Much of the research uses the longer timespan read text. However, less research is focused on the optimal or efficient length of recording. Furthermore, it may also be questionable which part of a given read text is most significant in recognizing the disease.

Therefore, we aimed the following hypotheses with the present research: Can a similar classification performance be achieved with partitioned samples based models than with full-length recordings based models.

Based on these, we partitioned the nearly one-minute original recordings into four shorter samples using annotation files. Then we created independent SVM models for the original and fragmented recordings. 12 MFCCs were extracted as features for the models. Argument in using only 12 MFCC features is that it is a desirable approach to make our model feasible for real-time implementation. The less features decrease the processing time. However, it is worth exploring results with additional speech-related features (Sztahó et al, 2017), (Sztahó & Valálik, 2019), (López et al, 2019). For evaluation, leave-one-speaker-out cross-validation (LOOCV) was performed. We conducted the following experiments: 1) using full-length recordings to train and test SVM models (baselines), 2) using partitioned samples separately to train and test SVM models, 3) using partitioned samples together to train and test SVM models. In future experiments, it would be also beneficial to explore other technologies (for example deep learning) (Faiyaz et al., 2020), (Johri & Tripathi, 2019). Finally, Mann-Whitney U non-parametric test was applied to test whether there are significant differences between the models' performances.

Using full-length recordings, the rbf kernel SVM achieved better performance (acc.: 84.2%, f1: 84.1%) to classify PD samples than the linear kernel SVM (acc: 79.1%, f1: 79.0%). The results differences were about 5% for all descriptor metrics, and 0.1 for

the Matthews correlation. That implies these features may have some nonlinear behaviour which can improve the classification performance.

Applying partitioned samples separately, similar results were achieved with both linear and rbf kernel SVM models. The models with rbf kernel produced results in a narrower deviation along with different parts, while the deviation of models with linear kernel had a wider range. For both linear and rbf kernel models, part1 and part3 samples were able to obtain even higher metric values than the baselines had. Furthermore, the models with the lowest results (part2 for both rbf and linear kernel models) are almost as high as the full-length versions. The explanation may be that if there is a part of the full-length recording that would produce a worse result, it can also degrade the results of the full recording. Conversely, it can also be said that we have seen parts that perform better and also parts that perform worse with this fragmentation technique. Some sort of weighted average may be the result of the entire recordings. With this in mind, a procedure can be developed in which just the right amount of text and the right content needs to be read aloud, and the results remain appropriate as the full text would be read aloud.

It can be mentioned that the linear SVM outperformed the rbf SVM in the part3 case. This may be explained by the lack of parameter optimization. So the linear SVM had a better default setting for this problem than the rbf SVM. Alternatively, it is conceivable that some features showed a more linear character for that part of the given recording, which the linear SVM could take advantage of. In any case, this phenomenon requires further investigation.

If all the partitioned recordings were used together, the result of the linear kernel model (acc: 82.0%, f1: 82.1%) is improved compared to the full-length model (acc: 79.1%, f1:79.0%). With rbf kernel model, the same results were obtained with all samples (acc: 83.9%, f1:83.8%) than using full-length recordings (acc: 84.2%, f1: 84.1%). The metric values experienced with the use of majority voting resulted in almost the same results as without majority voting.

From the statistical examination, it can be concluded that using a linear kernel, a significant improvement can be achieved with the part3 samples compared to the full-length study (using the Mann-Whitney U test, with a significance level of 0.05). The study also points out that the results obtained on the partitioned samples even show significant differences between them (part 2 case). Moreover, non-significant p values indicate that homogeneous results can be obtained with various speech lengths. This thus confirms that even shorter recordings may be sufficient to detect PD. Statistical analysis of rbf kernel SVM models showed no significant results. This confirms the statement stated formerly.

In summary, the results indicated that a suitable piece of longer speech recording may be sufficient to obtain the same classification performance as with full-length recordings (with a significance score of 0.05). Moreover, even a piece can be found at certain model settings that can achieve significantly better results than the full-length recordings. According to this, a shorter piece would be enough to record from the patient to examine. This is also more convenient and less burdensome for the patient.

In the future, for a more detailed analysis on result differences obtained with different parts of the speech samples, a complete investigation is advisable on exactly how these parts differ from each other (by linguistic-acoustic content).

Acknowledgements

Project no. K128568 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K_18 funding scheme. The research was partly funded by the CELSA (CELSA/18/027) project titled: “Models of Pathological Speech for Diagnosis and Speech Recognition.

Bibliography

Armstrong, M. J., Okun, M. S.: Diagnosis and Treatment of Parkinson Disease: A Review. In: *JAMA*, *323*(6), pp. 548. (2020) https://doi.org/10.1001/jama.2019.22360

Bhidayasiri, R., Tarsy, D.: Parkinson’s Disease: Hoehn and Yahr Scale. In: R. Bhidayasiri & D. Tarsy (Eds.), *Movement Disorders: A Video Atlas: A Video Atlas* (pp. 4–5). Humana Press. (2012) https://doi.org/10.1007/978-1-60327-426-5_2

Boersma, P., van Heuven, V.: Praat, a system for doing phonetics by computer. In: *Glot International* 5(9/10), pp. 341–345. (2001) *Online* [Accessed: 3 Nov. 2021] https://www.fon.hum.uva.nl/paul/papers/speakUnspeakPraat_glot2001.pdf

Dasgupta, S., Harisudha, K., Masunda, S.: Voiceprint analysis for Parkinson’s disease using MFCC, GMM, and instance based learning and multilayer perceptron. In: *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, pp. 1679–1682. (2017) https://doi.org/10.1109/ICPCSI.2017.8391999

Denisova, I. A., Chubarova, T. V., Bogatova, I. E., Vartanov, S. A., Kucheryanu, V. G., Polterovich, V. M., Turdyeva, N. A., Shakleina, M. V.: Estimating economic efficiency of preclinical diagnostics of Parkinson’s disease with cost-utility approach. In: *Population and Economics*, *4*(3), pp. 111–127. (2020) https://doi.org/10.3897/popecon.4.e59949

Faiyaz, A., Danish, R., R., Iqra, N., Mumtaz, A.: An LSTM based Deep learning model for voice-based detection of Parkinson’s disease. In: International Journal of Advanced Science and Technology, 29(5s), pp. 337 - 343. (2020) Retrieved from http://sersc.org/journals/index.php/IJAST/article/view/7166

Frid, A., Safra, E., J., Hazan, H., Lokey, L., L., Hilu, D., Manevitz, L., Ramig, L., O., Sapir, S.: Computational Diagnosis of Parkinson’s Disease Directly from Natural Speech Using Machine Learning Techniques. In: 2014 IEEE International Conference on Software Science, Technology and Engineering, pp. 50–53. (2014) https://doi.org/10.1109/SWSTE.2014.17

Godino-Llorente, J. I., Shattuck-Hufnagel, S., Choi, J. Y., Moro-Velázquez, L., Gómez-García, J. A.: Towards the identification of Idiopathic Parkinson’s Disease from the speech. New articulatory kinetic biomarkers. In: *PLOS ONE*, *12*(12), pp. e0189583. (2017) https://doi.org/10.1371/journal.pone.0189583

Hemmerling, D., Sztahó, D.: Parkinson’s Disease Classification Based on Vowel Sound. In: *11th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA)*, pp. 29-32, Firenze, Italy, (2019) https://doi.org/10.36253/978-88-6453-961-4

Johri, A., Tripathi, A., K.: Parkinson Disease Detection Using Deep Neural Networks. In: 2019 Twelfth International Conference on Contemporary Computing (IC3), pp: 1-4. (2019) https://doi.org/10.1109/IC3.2019.8844941

Kiss, G., Takács, A., B., Sztahó, D., Vicsi, K.: Detection Possibilities of Depression and Parkinson’s disease Based on the Ratio of Transient Parts of the Speech. In: *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 165–168. (2018) https://doi.org/10.1109/CogInfoCom.2018.8639901

Lenain, R., Weston, J., Shivkumar, A., Fristed, E.: Surfboard: Audio Feature Extraction for

Modern Machine Learning. In: *Interspeech 2020*, pp. 2917–2921. (2020) https://doi.org/10.21437/Interspeech.2020-2879

Lindgren, P., von Campenhausen, S., Spottke, E., Siebert, U., Dodel, R.: Cost of Parkinson's disease in Europe. In: *European Journal of Neurology*, *12*(s1), pp. 68–73. (2005) https://doi.org/10.1111/j.1468-1331.2005.01197.x

López, J., V., E., Orozco-Arroyave, J., R., Gosztolya, G.: Assessing Parkinson's Disease from Speech Using Fisher Vectors. Proc. In: Interspeech 2019, pp. 3063-3067, (2019) https://doi.org/10.21437/Interspeech.2019-2217

Nachar, N.: The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. In: *Tutorials in Quantitative Methods for Psychology*, *4*(1), pp. 13–20. (2008) https://doi.org/10.20982/tqmp.04.1.p013

Novotný, M., Rusz, J., Čmejla, R., Růžička, E.: Automatic evaluation of articulatory disorders in Parkinson's disease. In: *IEEE Transactions on Audio, Speech and Language Processing*, *22*(9), pp. 1366–1378. (2014) https://doi.org/10.1109/TASLP.2014.2329734

Pah, N. D., Motin, M. A., Kumar, D. K.: Voice Analysis for Diagnosis and Monitoring Parkinson's Disease. In: S. P. Arjunan & D. K. Kumar (Eds.), *Techniques for Assessment of Parkinsonism for Diagnosis and Rehabilitation* (pp. 119–133). Springer Singapore. (2021) https://doi.org/10.1007/978-981-16-3056-9_8

Parkinson, J.: An essay on the shaking palsy. 1817 In: *The Journal of Neuropsychiatry and Clinical Neurosciences*, *14*(2). (2002) https://doi.org/10.1176/jnp.14.2.223

Pompili, A., Abad, A., Romano, P., Martins, I. P., Cardoso, R., Santos, H., Carvalho, J., Guimarães, I., Ferreira, J. J.: Automatic Detection of Parkinson's Disease: An Experimental Analysis of Common Speech Production Tasks Used for Diagnosis. In: K. Ekštein & V. Matoušek (Eds.), *Text, Speech, and Dialogue* (Vol. 10415, pp. 411–419). Springer International Publishing. (2017) https://doi.org/10.1007/978-3-319-64206-2_46

Reichmann, H.: Clinical Criteria for the Diagnosis of Parkinson's Disease. In: *Neurodegenerative Diseases*, *7*(5), pp: 284–290. (2010) https://doi.org/10.1159/000314478

Rusz, J., Cmejla, R., Ruzickova, H., Růžička, E.: Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. In: *The Journal of the Acoustical Society of America*, *129*, pp. 350–367. (2011) https://doi.org/10.1121/1.3514381

Schulz, G. M., Grant, M. K.: Effects of speech therapy and pharmacologic and surgical treatments on voice and speech in Parkinson's disease: A review of the literature. In: *Journal of Communication Disorders*, *33*(1), pp. 59–88. (2000) https://doi.org/10.1016/S0021-9924(99)00025-8

Simon, D. K., Tanner, C. M., Brundin, P.: Parkinson Disease Epidemiology, Pathology, Genetics, and Pathophysiology. In: *Clinics in Geriatric Medicine*, *36*(1), pp. 1–12. (2020) https://doi.org/10.1016/j.cger.2019.08.002

Sonawane, B., Sharma, P.: Speech-based solution to Parkinson's disease management. In: *Multimedia Tools and Applications*, *80*(19), pp. 29437–29451. (2021) https://doi.org/10.1007/s11042-021-11061-1

Sztahó, D., Valalik, I., Vicsi, K.: Parkinson's Disease Severity Estimation on Hungarian Speech Using Various Speech Tasks. In: *2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pp. 1–6. (2019) https://doi.org/10.1109/SPED.2019.8906277

Sztahó, D., Tulics, M. G., Vicsi, K., Valálik, I.: Automatic estimation of severity of Parkinson's disease based on speech rhythm related features. In: 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 11-16. (2017) https://doi.org/10.1109/CogInfoCom.2017.8268208

Sztahó, D., Valálik, I.: Speech Fluency Measurement of Patients with Parkinson's Disease by Forward-Backward Divergence Segmentation. In: 2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 295-300. (2019)

https://doi.org/10.1109/CogInfoCom47531.2019.9090001
Tolosa, E., Wenning, G., Poewe, W. (2006). The diagnosis of Parkinson's disease. *The Lancet Neurology*, *5*(1), pp. 75–86. https://doi.org/10.1016/S1474-4422(05)70285-4
Tsanas, A., Little, M., A., McSharry, P., E., Spielman, J., Ramig, L., O.: Novel Speech Signal Processing Algorithms for High-Accuracy Classification of Parkinson's Disease. In: IEEE Transactions on Biomedical Engineering, 59(5), pp. 1264–1271. (2012) https://doi.org/10.1109/TBME.2012.2183367
Tysnes, O.-B., Storstein, A. (2017). Epidemiology of Parkinson's disease. *Journal of Neural Transmission*, *124*(8), pp. 901–905. https://doi.org/10.1007/s00702-017-1686-y
Vadovsky, M., & Paralic, J.: Parkinson's disease patients classification based on the speech signals. In: *2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 321–326. (2017) https://doi.org/10.1109/SAMI.2017.7880326
Vaiciukynas, E., Verikas, A., Gelzinis, A., & Bacauskiene, M.: Detecting Parkinson's disease from sustained phonation and speech signals. In: *PLOS ONE*, *12*(10), pp. e0185613. (2017) https://doi.org/10.1371/journal.pone.0185613
Wise, R. A.: Dopamine, learning and motivation. In: *Nature Reviews Neuroscience*, *5*(6), pp. 483–494. (2004) https://doi.org/10.1038/nrn1406