K128568 Final Report

We implemented the project plan to the maximum despite the fact that the COVID-19 pandemic situation has hampered our work to some extent.
A significant part of the work on which research is based, the collection of speech material from healthy and sick people, has encountered serious difficulties in the second and third year of the project. We had to record most of the pathology databases in clinics and hospitals, which we only had access to in the summer months, and where the use of masks is still mandatory. (Recording the speech in masks would have been pointless as the mask distorts the parameters of speech.)
That is why we requested and received an extension of the project duration from 3 years to 4 years.
By the end of the fourth year, we have fully completed the research work that we set out as a goal in the project's work plan. We recorded the voices of speech pathology and healthy speakers in the planned number and quality, prepared the databases and made them available to researchers. We examined the effect of changes in the vocal organs and the changes in certain neural processes controlling the motor function of voice production on the measurable parameters of the speech product. We lay the foundations for a new group of automatic diagnostic procedures, voice-based diagnostics. Finally, an application, an automatic diagnostic system was created, which is able to evaluate a speaker's voice sample, a prototype for assisting medical staff. This system is currently capable of estimating the probability of three types of voice disorders: depression, dysphonia and Parkinson's disease in English and Hungarian, based on a short fixed text. The overall balanced accuracy is 81.1% for the four-class classification. The performance of the model was evaluated in a 10-fold nested cross-validation setup using the samples available in the datasets.
Furthermore, two of our colleagues working on the project, Gabor Kiss and Miklós Gabriel Tulics, successfully defended their PhD dissertations on two different sub-topics of the project:
Gabor Kiss: Investigation of acoustic-phonetic characteristics of depressed speech. The date of his defence was in June 2020. [1]
Miklós Gabriel Tulics: Automatic Classification of Dysphonia [9]. The date of his defence was in February 2021. [2]

WORK PACKAGES:

1. Construction of databases

1.1. Recordings

The collection and the preparation of databases started in the first year, and as I wrote it in the introduction, we had some difficulties in the second and third year because of the pandemic situation. After all, in the 4th year, we were able to prepare all three pathological and the corresponding healthy speech databases.
Three different types of pathological speech databases were constructed. These are the followings:
1. **Laryngological pathology-speech database** (LAPASDA)
2. **Depression-speech database** (DEPISDA)
3. **Speech database for patients with Parkinson's disease** (PASDA)

In order to compare healthy and pathological speech, speech material from healthy speakers is also needed, with a similar distribution of age as the pathological databases. That's why we created a fourth database: the "**Healthy elderly database**".

All four databases were created under the same technical conditions: sound recordings were made with a Monacor ACM-100 close-up microphone and an Audio-Technika ATR3350 clip-on microphone, using an external USB sound card, with a sampling frequency of 16 kHz and 16-bit quantization. Although the speech databases contain text materials corresponding to the specific disease type, the reading of the phonetically balanced story "The North Wind and the Sun" is included in all datasets. This will make it possible to examine the possibilities of joint classification of diseases. Scales indicating the severity of the disease, the patient's gender, age, smoking, and other diseases affecting voice production were assigned to the speech material. The databases completed so far are described below:

1. **LAPASDA database**

The LAPASDA database is a continuous speech database showing laryngeal disorders. The recordings were made at the Head and Neck Surgery Department of the National Oncology Institute, at the time of appointment, under the guidance of specialist Dr. Krisztina Mészáros. We recorded recordings from a total of 200 speakers. To characterize the severity of voice distortion, the medical expert assigned a value according to the widely used 3-dimensional RBH severity scale to each voice recording.

2. **DEPISDA database**

The database contains speech recordings of people suffering from depression of varying severity. The recordings were made under the supervision of Dr. Lajos Simon, chief physician of the Department of Psychiatry and Psychotherapy of Semmelweis University. In addition to the read tale, the database also contains continuous spontaneous speech (conversation between doctor and patient). We recorded recordings from 200 speakers. We assigned values according to the BDI scale used in medical practice to characterize the severity of depression, which were given based on the evaluation of the form filled out by the patients.

3. **PASDA database**

It contains the speech of patients with Parkinson's disease. The recordings were made at the Virányos Clinic, with the help of neurosurgeon Dr. István Valálik. In addition to the read fairy tale, it also contains special material of sustained sounds, syllables, and words. We recorded recordings from 100 speakers. Each audio recording was assigned the severity of the patient's Parkinson's disease, according to the UPDRS scale used in medical practice.

4. **Healthy Elderly database**

It contains audio recordings of healthy, mainly elderly people (over 60). We have currently recorded recordings from 200 speakers.

1.2. Segmentation and labeling of databases

In the course of our research, a detailed acoustic analysis of speech sounds belonging to different phonetic classes is necessary. Therefore, we have to prepare speech-level segmentation and labeling of the databases. The work is quite time-consuming, but we have developed automatic segmentation procedures [30]. Thus, we have to correct the errors of automatic segmentation in the future.

RESULTS:
The finalized and anonymized datasets are available at the following url: http://lsa.tmit.bme.hu/fundings/pathological_ai.html

2. Parameter reduction

It is valid for all works in Sections 2, 3, 4 and 5 of this final report that we were able to carry out the experiments step by step on the currently existing databases, constantly expanding and improving our models.

Statistical examination was prepared on existing databases to study the characteristic acoustic parameters of three types of pathological speech separately, for vocal disorders, for depression and for Parkinson's disease. Correlation between the judgment of doctors and numerical features of acoustic-phonetic parameters were examined. Each patient's record comes up with a set of multi-dimensional vector, from which the most characteristic parameters were selected separately for the three types of distorted speech. The selection of optimal parameters by multivariate statistical methods were prepared in order to reduce the number of variables: principal component analysis (PCA), linear discriminant analysis (LDA), etc. At the end an extended version of optimal parameter sets for all types of pathological speech were selected.

RESULTS:
Final version of optimal parameter sets for 3 types of pathological speech. Test results of different classification and regression scenarios were published [3][4][5].

3. Classification, regression

Two-class and multiple-class classifications were carried out to detect vocal disorders, depression and Parkinson's disease, using Support Vector Machines (SVM) and different type of deep neural networks, i-vector and x-vector classifiers. The different multidimensional acoustic-phonetic parameters were the input vectors of the classifiers respectively [6][7][11][12][13][14][22][23]. The different of classifiers were compared [29].

Methods for the automatic assessment of severity of the three types of illness were developed also using different types of regression methods. On the base of these experiments, it is clear that automatic estimation of the severity of the tree type of illness is possible [3][4][5][8][9] [10][24][25][26][27].
Based on the existing databases, the different classification and regression methods were compared.
Gábor Kiss's PhD dissertation gives a nice summary of the results obtained for depressed speech [8]:

- Using the SVM machine learning procedure, speech samples from depressed and healthy speakers can be distinguished from each other, with 88% accuracy in case of Hungarian.
- Using an SVR machine learning procedure, the severity of a speaker's depressed state can be estimated. The root mean square error achieved (RMSE) is 6.3 while the mean error is 5.1 (MAE) for Hungarian.

Miklós Tulics's PhD dissertation gives a nice summary of the results obtained for dysphonic speech [9].

- Binary classification of dysphonic and healthy voices is possible for Hungarian with 88%accuracy applying a Fully-Connected Deep Neural Network.
- Automatic estimation of the severity of dysphonia is possible with Support Vector Regression (linear kernel), reaching 0.85 Pearson correlation and 0.46 RMSE.

A nice summary of the results obtained for Parkinson's disease are presented in [28].

But in practice all of these diseases may occur simultaneously among the patients. Therefore, we focused on separating more (4 or 6) different disease classes and subclasses simultaneously, using multi-class classification method. Examined disease types are: speech samples of depression, Parkinson's disease, structural morphological alteration of vocal organs [12][15][16][17][18][19].

Various models are built: estimating the probability of each disease in a joint model in one step and estimating the severity of each disease by separately trained model per disease.
Using multiple delay scales and different classification methods, the best overall accuracy was 81.1 % in 4-class classification tasks [17]. This is a remarkable achievement, especially because there are few speech samples per class of illness available for the training. This suggests that there are indeed correlation differences in the time domain signals of the measured features due to the articulation abnormalities of the examined 4 disease types.

RESULTS:
Two-class and multi-class classification, regression results are published in conference papers and scientific journals, and in two PhD dissertations.

4. The examination of language dependency

Language dependence was examined in the case of dysphonic speech, and depressed speech since we were able to obtain a dysphonic foreign language speech database. Thus cross-lingual experiments of dysphonic voice detection and dysphonia severity level estimation was carried out using Hungarian and a Dutch speech. Various acoustic features were calculated on the entire speech samples and phoneme level.
It was found that cross-lingual detection of dysphonic speech is indeed possible with acceptable generalization ability and features calculated on phoneme-level parts of speech can improve the results [20].
Depression read speech materials were gathered in the Hungarian and Italian languages from both healthy people and patients diagnosed with different degrees of depression. In addition, we received a depression database in German language. By statistical examination it was found that there are many parameters in the speech of depressed people that show significant differences compared to a healthy reference group. Moreover, most of those parameters behave similarly in other languages such as in Italian and German. Furthermore, we found that it is possible creating different language-specific models using the same feature extraction method. It can be established that it is possible to create models valid for several languages, or even cross-language models with acceptable performance [21].

RESULTS:
Optimal multi-lingual parameter set for all types of pathological speech, and test results of different multi-lingual classifications and regressions are published in conference papers.

5. The development on an automatic diagnostic system and audio recording protocol for laryngeal pathological speech, depression and Parkinson's disease.

We put the results of chapter 4 into practice, and an application, a decision support tool has been developed that is able to assess a voice sample according to three different voice disorders: depression, Parkinson's disease and dysphonic speech (Fig.1) Affection probability of each disorder is analysed along with their severity estimation. Although the acoustic models (support vector machine and regression models) are trained on Hungarian voice samples, English samples can also be utilized for assessment, due to the fact, that processing algorithms are language independent (for European languages).
The process of decision making is the following. The input sample (live or pre-recorded) is normalized and segmented to phonemes. The feature extractor calculates acoustic features necessary to analyse the sample. These features are the fed into four decision units. The joint decision unit (first unit) calculates the probability of each disorder by Support Vector Machine

model using linear kernel function (result is displayed as pie chart (see Figure 2) Beside the joint probability decision, the severity of each disorder is also estimated (additional three decision units) and displayed in the application by support vector regression (with linear kernel).

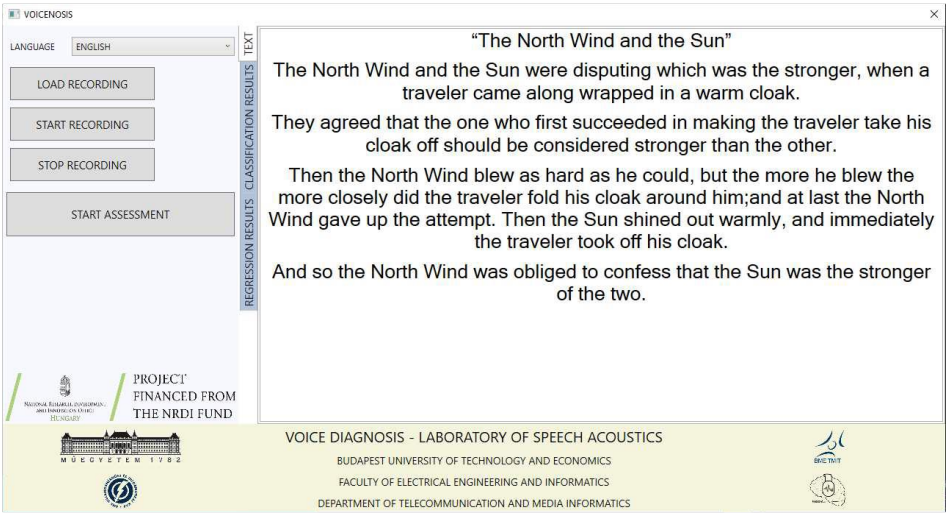

Figure 1: User interface - recording.

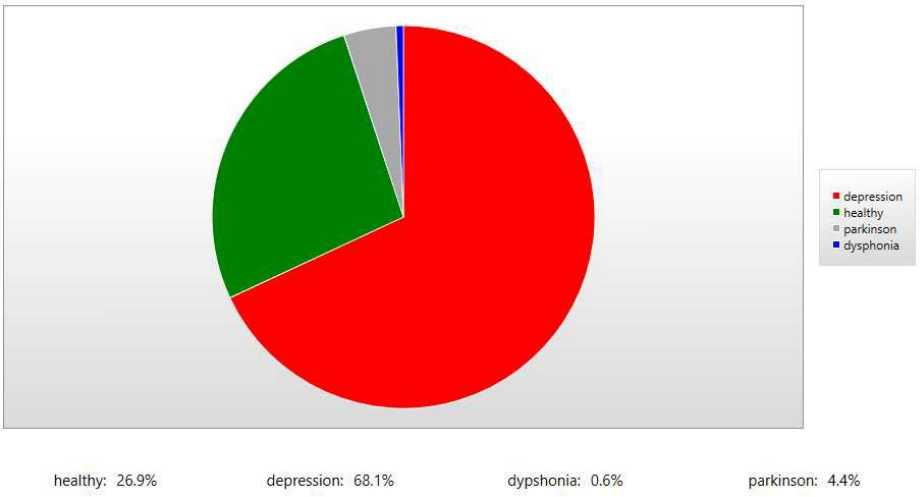

Figure 2: User interface – joint decision

The results are displayed as pie chart for probabilities and separate severity scores.
The input of the application is a read text with a fixed linguistic content (the phonetically balanced story "The North Wind and the Sun"). It is possible to load a pre-recorded voice sample or create a live recording.
The models in the application are evaluated in a 10-fold nested cross-validation setup. The overall balanced accuracy is 81.1%.
The estimation of the severity of each disorder is evaluated by RMSE and Pearson correlation metrics. Separate models are trained for each given disease category to estimate its severity. For depression, dysphonia and Parkinson disease, the normalized RMSE values are 0.59, 0.82, 0.67 respectively and the Pearson correlation values are 0.59, 0.82, 0.67.

This application was presented in September at the most important international scientific conferences on speech [17].

RESULT:
Decision support tool what is able to assess a voice sample according to three different voice disorders: depression, Parkinson's disease and dysphonic speech. The program can be downloaded here: https://lsa.tmit.bme.hu/files/voicenosis.zip
(All you have to do is unpack it somewhere and start voicenosis.exe. The program can be tested. Here are test samples: https://lsa.tmit.bme.hu/files/voicenosis_test_samples.zip, although the system can also be tested by reading the text directly.)

Publications:

[1] Kiss Gábor: A depressziós beszéd akusztikai-fonetikai jellemzőinek vizsgálata, PhD dissertation, BME Doctoral School of Informatics (2020)

[2] Tulics Miklós Gábriel: Automatic Classification of Dysphonia, PhD dissertation, BME Doctoral School of Electrical Engineering (2021)

[3] Dávid Sztahó, Anett Orosz, István Valálik
Articulation correctness measurement of Parkinson's disease using low resource-intensitve segmentation methods
11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2020) pp. 121-124.

[4] Pašic Azra, Kiss Gábor, Sztahó, Dávid
A depresszió hang alapú felismerésének optimalizációja hangfelvétel hossz alapján
XVI. Magyar Számítógépes Nyelvészeti Konferencia ,Szeged, Hungary (SZTE Informatikai Intézet) (2020) pp. 83-92. 10 p.

[5] Gábor Kiss, Attila Zoltán Jenei
Investigation of the Accuracy of Depression Prediction Based on Speech Processing
43rd International Conference on Telecommunications and Signal Processing (TSP 2020) pp. 129-132.

[6] Daria Hemmerling, Dávid Sztahó
Parkinson's Disease Classification Based on Vowel Sound
MAVEBA 2019 (11th International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications) pp. 29-32.

[7] Miklós Gábriel Tulics, György Szaszák, Krisztina Mészáros, Klára Vicsi
Using ASR Posterior Probability and Acoustic Features for Voice Disorder Classification
11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2020) ID: 25 , 5 p.

[8] Attila Zoltán Jenei, Gábor Kiss
Possibilities of Recognizing Depression with Convolutional Networks Applied in Correlation Structure
43rd International Conference on Telecommunications and Signal Processing (TSP 2020) pp. 101-104.

[9] Jenei Attila Zoltán, Kiss Gábor
Depresszió detektálása korrelációs struktúrán alkalmazott konvolúciós hálók segítségével
In: Berend, Gábor; Gosztolya, Gábor; Vincze, Veronika (szerk.) XVI. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, Hungary (SZTE Informatikai Intézet) 2020 pp.59-71.

[10] Attila Zoltán Jenei, Gábor Kiss
Severity Estimation of Depression Using Convolutional Neural Network
PERIODICA POLYTECHNICA-ELECTRICAL ENGINEERING AND COMPUTER SCIENCE 65 (3) 227-234 (2021)

[11] Dávid Sztahó, Gábor Kiss, Miklós Gábriel Tulics
Deep Learning Solution for Pathological Voice Detection using LSTM-based Autoencoder Hybrid with Multi-Task Learning
International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSIGNALS 2021) pp. 135-141

[12] Gábor Kiss, Miklós Gábriel Tulics, Attila Zoltán Jenei, Dávid Sztahó
Analysis of Cross Disorder Severity Prediction Problems Based on Speech Features
International Workshop Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2021), Firenze, Italy, pp. 71-74.

[13] Tulics Miklós Gábriel
A diszfónia és automatikus felismerése
Markó Alexandra (szerk.): Tanulmányok a beszédtudomány alkalmazásainak köréből (ELTE Eötvös Kiadó) 2021, pp. 35-63.

[14] Kiss Gábor
A depresszió automatikus becslése a beszéd akusztikai-fonetikai jellemzői alapján
Markó Alexandra (szerk.): Tanulmányok a beszédtudomány alkalmazásainak köréből (ELTE Eötvös Kiadó) 2021, pp. 65-86.

[15] Péter Rozmán, Dávid Sztahó, Gábor Kiss, Attila Zoltán Jenei
Automatic recognition of depression and Parkinson's disease by LSTM networks using sample chunking
12th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2021) pp. 163-166

[16] Attila Zoltán Jenei, Gábor Kiss, Miklós Gábriel Tulics, Dávid Sztahó
Separation of Several Illnesses Using Correlation Structures with Convolutional Neural Networks
ACTA POLYTECHNICA HUNGARICA 18(7) 47-66 (2021) IF: 1.711

[17] Gábor Kiss, Dávid Sztahó, Miklós Gábriel Tulics
Application for detecting depression, Parkinson's disease and dysphonic speech
Interspeech 2021, Brno, Czech Republic, pp. 956-957 (Tell&Show)

[18] Attila Zoltán Jenei, Gábor Kiss, Dávid Sztahó
Detection of Speech Related Disorders by Pre-Trained Embedding Models Extracted Biomarkers
24th International Conference on Speech and Computer (SPECOM 2022) Gurugram, India, accepted

[19] Dávid Sztahó, Gábor Kiss, Miklós Gábriel Tulics, Bálint Hajduska-Dér, Klára Vicsi
Automatic discrimination of several types of speech pathologies
10th Conference on Speech Technology and Human-Computer Dialogue) SpeD 2019, Timișoara, Romania
Paper: 119

[20] Dávid Sztahó, Miklós Gábriel Tulics, Jinzi Qi, Hugo Van hamme, Klára Vicsi
Cross-lingual Detection of Dysphonic Speech for Dutch and Hungarian Datasets
15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSIGNALS 2022) pp. 215-220

[21] Gábor Kiss
Investigation of speech-based language-independent possibilities of depression recognition
International Conference on Telecommunications and Signal Processing (TSP 2022) pp. 226-229

[22] Miklós Gábriel Tulics, György Szaszák, Krisztina Mészáros, Klára Vicsi
Artificial Neural Network and SVM based Voice Disorder Classification
10th IEEE International Conference on Cognitive InfoCommunications (CogInfoCom 2019) Naples, Italy pp. 307-311 (Paper No. 054-PID6115443)

[23] Miklós Gábriel Tulics, Lívia Judit Lavati, Krisztina Mészáros, Klára Vicsi
Possibilities for the automatic classification of functional and organic dysphonia
10th Conference on Speech Technology and Human-Computer Dialogue, (SpeD 2019) Timișoara, Romania)
Paper: 116

[24] Miklós Gábriel Tulics, Klára Vicsi
The automatic assessment of the severity of dysphonia
INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY Vol.22. Issue 2 pp.341-350 (2019)

[25] Gábor Kiss, Dávid Sztahó, Klára Vicsi
Depression State Assessment: Application for detection of depression by speech
InterSpeech 2019, pp. 966-967

[26] Dávid Sztahó, István Valálik, Klára Vicsi
Parkinson's Disease Severity Estimation on Hungarian Speech Using Various Speech Tasks
10th Conference on Speech Technology and Human-Computer Dialogue, (SpeD 2019) Timișoara, Romania)
Paper: 112

[27] Dávid Sztahó, István Valálik
Speech Fluency Measurement of Patients with Parkinson's Disease by Forward-Backward Divergence Segmentation
10th IEEE International Conference on Cognitive InfoCommunications (CogInfoCom 2019) Naples, Italy pp. 295-299 (Paper No. 052-PID6123495)

[28] Bálint Hajduska-Dér, Gábor Kiss, Dávid Sztahó, Klára Vicsi, Lajos Simon
The applicability of the Beck Depression Inventory and Hamilton Depression Scale in the automatic recognition of depression based on speech signal processing
Frontiers in Psychiatry, 2022 p. 12 (Open Access)

[29] Dávid Sztahó, Attila Zoltán Jenei, István Valálik, Klára Vicsi
The Effect of Speech Fragmentation and Audio Encodings on Automatic Parkinson's Disease Recognition
J. Biomedical Science and Engineering 15(1) 6-25, 2022 (Open Access)

[30] Kiss, G., Sztahó, D., Vicsi, K. "Language independent automatic speech segmentation into phoneme-like units on the base of acoustic distinctive features." 4th IEEE International Conference on Cognitive Infocommunications - CogInfoCom 2013. pp. 579- 582. 2013