MULTILINGUAL ANALYSIS OF HYPOKINETIC DYSARTHRIA IN PATIENTS WITH PARKINSON'S DISEASE

Daniel Kováč
Doctoral Degree Programme (1st year), FEEC BUT
E-mail: xkovac41@stud.feec.vutbr.cz

Supervised by: Jiří Mekyska
E-mail: mekyska@feec.vutbr.cz

Abstract: This article deals with the multilingual analysis of hypokinetic dysarthria (HD) in patients with Parkinson's disease (PD). The goal is to identify acoustic features that have high discrimination power and that are independent of the language of a speaker. The speech corpus contains 59 PD patients and 44 healthy controls (HC) speaking in Czech (cs) and American English (en-US). Based on non-parametric statistical tests and logistic regression, we observed the best discrimination power has the speech index of rhythmicity (extracted from a reading text) and harmonic-to-noise ratio (extracted from a sustained vowel). We were able to identify PD with 67% sensitivity and 79% specificity in the Czech corpus and with 78% sensitivity and 67% specificity in the English one. The performance of the model was significantly lower when combining both datasets, thus suggesting language plays a significant role during the automatic assessment of HD.

Keywords: Acoustic analysis, Parkinson's disease, hypokinetic dysarthria, classification

1 INTRODUCTION

Parkinson's disease (PD) is a chronic neurodegenerative disorder characterized by progressive degeneration of neurons in the substantia nigra pars compacta (black substance) [1]. The risk of developing PD increases with the age. Other significant factors are mainly genetic predisposition, traumatic brain injury and gender. It is also obvious that the number of patients increases with population and life expectancy [2]. The disease was first described in 1817 by the English surgeon James Parkinson [3]. Although science has progressed since this time, we still do not know the true cause of PD and are not able to cure it. However, it is possible to alleviate its course. The non-invasive way is for example, using levodopa, an amino acid, which is converted to dopamine in the brain. The riskier option is surgery [4]. Therefore, early detection and initiation of treatment have a major impact on the further course of the disease. Hoehn and Yahr defined five stages of PD, with speech impairments occurring already in the second early phase of the disease [5]. These speech impairments are caused by deteriorative phonation, articulation, prosody, and respiration and are collectively referred to as hypokinetic dysarthria (HD), which affect up to 90 % of patients with PD [6]. Acoustic analysis of speech and voice signal is one way to an automated diagnosis of PD. Many studies aim to find acoustic features with high discrimination power, but only a few are multilingual. There is no multilingual study dealing with Czech and American English. Finding features that have high discrimination power and are independent of language is key to the global automated diagnosis of PD based on acoustic analysis of speech.

2 METODOLOGY

Speech recording took place in the silent room using Audacity software and a condenser microphone with the cardioid polar pattern. The result was a dataset of mono signals with a 16 kHz sampling frequency obtained from each participant.

2.1 DATASET

The database contains subjects speaking Czech and American English. Table 1 represents the numbers of males and females in combination with language and diagnosis. The mean age of Czech speakers is 73.9 ± 15.4 (PD) and 66.1 ± 5.0 (HC). For American English, it is 67.7 ± 4.1 (PD) and 70.4 ± 8.6 (HC). The Czech participants were enrolled at the St. Annes University Hospital Brno and the American ones at the Department of Neurology, College of Medicine, University of Arizona. All subjects signed informed consent. The study was approved by local ethic committees.

Table 1: Demographic data.

	cs			en-US			all
	male	female	together	male	female	together	
PD	29	19	48	8	3	11	59
HC	9	17	26	5	13	18	44
together	38	36	74	13	16	29	103

2.2 SPEECH TASKS

To analyze all dimensions of HD, several speech tasks have been employed (see Table 2). Tasks are designed to examine phonation, articulation and prosody.

Table 2: Speech tasks

Label	Speech task	Description
TSK1	Monologue	Monolog, at least 90 s long without interruption of a clinician. The participants were instructed to speak about their hobbies, family, job, actual date activity, etc.
TSK2	Reading	Reading a short text. The patient could read the text for her-/himself in advance.
TSK3	Sustained phonation	Approximately 3-s (not longer than 5 s) sustained vowel of [a] at a comfortable pitch and loudness. Performed on one breath.
TSK4	Sustained phonation	Approximately 3-s (not longer than 5 s) sustained vowel of [i] at a comfortable pitch and loudness. Performed on one breath.
TSK5	Sustained phonation	Approximately 3-s (not longer than 5 s) sustained vowel of [u] at a comfortable pitch and loudness. Performed on one breath.
TSK6	Sustained phonation	Sustained phonation of [a] at a comfortable pitch and loudness as constant and long as possible, at least 5 s. Performed on one breath.
TSK7	Diadochokinetic task	Rapid steady [pa]-[ta]-[ka] syllables repetition as constant and long as possible, repeated at least 5 times. Performed on one breath.

2.3 ACOUSTIC FEATURES

Selected acoustic features were calculated from different speech tasks for each participant. A list of all parameters together with the disorders they investigate is given in Table 3. Acoustic features depend on sex and age, therefore these confounding factors were removed using linear regression.

2.4 STATISTICAL ANALYSIS

For testing the normal distribution of features, the Kolmogorov–Smirnov test was applied. Because many features did not fit the normal distribution, for subsequent statistical analyses, Mann–Whitney U test was used. The first test investigates the dependence of the feature on the disease (3 models: cs, en-US, all) and the second the dependence on the language (2 models: PD, HC). We additionally performed a multiple testing correction via the false discovery rate (FDR) approach.

2.5 MULTIVARIATE ANALYSIS

All acoustic features were used as input independent variables for training the machine learning models (3 models: cs, en-US, all). Logistic regression was chosen as the classifier. Training data was

Table 3: Acoustic features

Speech task	Acoustic feature	Specific disorder	Feature definition
			Phonation
TSK6	MPT	Airflow insufficiency	Maximum phonation time, aerodynamic efficiency of the vocal tract measured as the maximum duration of the prolonged vowel.
TSK3–6	relF0SD	Irregular pitch fluctuations	Standard deviation of fundamental frequency relative to its mean, variation in frequency of vocal fold vibration.
TSK3–6	Jitter (PPQ)	Microperturbations in frequency	Frequency perturbation, extent of variation of the voice range. Jitter is defined as the variability of the F0 of speech from one cycle to the next.
TSK3–6	Shimmer (APQ)	Microperturbations in amplitude	Amplitude perturbation, representing rough speech. Shimmer is defined as the sequence of maximum extent of the signal amplitude within each vocal cycle.
TSK3–6	HNR	Increased noise	Harmonics-to-noise ratio, the amount of noise in the speech signal, mainly due to incomplete vocal fold closure. HNR is defined as the amplitude of noise relative to tonal components in speech.
TSK3–6	DUV	Aperiodicity	Degree of unvoiced segments, the fraction of pitch frames marked as unvoiced.
TSK3–6	relF1SD, relF2SD	Tremor of jaw	Standard deviation of first (F1) and second (F2) formant relative to its mean. Formants are related to resonances of the oro-naso-pharyngeal tract and are modified by position of tongue and jaw.
			Articulation
TSK1–5	VAI	Decreased tongue movement	Vowel articulation index, based on formant centralization, defined as VAI = (F1a + F2i)/(F1i + F1u + F2a + F2u).
TSK1–2	relF1SD, relF2SD	Rigidity of tongue and jaw	Standard deviation of first (F1) and second (F2) formant relative to its mean.
TSK2	#lndmrk	Imprecise articulation	Number of speech landmarks representing local energy maxima characterized by harmonic power. Landmark patterns are identified by comparison between “coarse” and “fine” spectral detail.
TSK7	PR	Slow alternating motion rate	Pace rate, representing the number of syllable vocalizations per second. Considering first 30 syllables.
TSK7	COV	Instability of diadochokinetic pace	Coefficient of variation, defined as the ratio of the standard deviation of the duration of the fourth to tenth DDK cycles to the average duration of the first three cycles.
TSK7	RI	Instability of diadochokinetic pace	Rhythm instability, defined as sum of absolute deviations from a regression line modelling each DDK cycle duration, weighted to the total DDK performance time.
TSK7	PA	Acceleration of diadochokinetic pace	Pace acceleration, defined as PA = 100 x (avCycDur4–6 – avCycDur7–9) / avCycDur1–3, where avCycDurX–Y is average duration of cycles X–Y.
TSK7	RA	Acceleration of diadochokinetic pace	Rhythm acceleration, defined as gradient of regression line modelling DDK cycle durations (positive values mean acceleration).
			Prosody
TSK1–2	relSEOSD	Monoloudness	Speech loudness variation, defined as a standard deviation of intensity contour relative to its mean after removing silences exceeding 50 ms.
TSK1–2	relF0SD	Monopitch	Pitch variation, defined as a standard deviation of F0 contour relative to its mean.
TSK2	SPIR	Inappropriate silences	Number of pauses relative to total speech time after removing periods of silence lasting less than 50 ms.
TSK2	PPR	Higher proportion of silence time	Percentual pause ratio, defined as total duration of silences (longer than 50 ms)/total duration of speech.
TSK2	DurMED	Longer duration of silences	Median duration of silences longer than 50 ms.
TSK2	DurMAD	Higher variability of silence duration	Median absolute deviation of silence duration (longer than 50 ms).
TSK2	AR	Unnatural speech rate	Number of speech sounds produced per second after pauses longer than 50 ms were removed.
TSK2	NST	Higher proportion of silence time	Net speech time.
TSK2	TPT	Higher proportion of silence time	Total pause time (pauses longer than 50 ms were removed).
TSK2	TST	Higher proportion of silence time	Total speech time.
TSK2	EEVOL	Unstable loudness	Energy evolution, defined as the slope of intensity.

balanced by Synthetic Minority Oversampling Technique (SMOTE). Hyperparameter tuning focused on the best accuracy was led by grid search and 10-fold cross-validation with 10 repetitions. The leave-one-out method was used for validating the models.

3 RESULTS

3.1 STATISTICAL ANALYSIS

Significant combinations (level of significance $\alpha = 0.05$) of acoustic features and tasks are listed in Table 4. P-values of Test 1 are associated with the difference between PD and HC. P-values of Test 2 are linked with the differences between languages.

Table 4: Significant combinations of acoustic feature and speech task.

Speech task	Acoustic feature	Language	Test 1: p-value	Test 2: p-value	Change with PD
TSK2	SPIR	all	0.019	HC: 0.992 PD: 0.359	$\bar{x}$: ↓ Me: ↓ s: ↓
TSK3	mean HNR	all	0.019	HC: 0.185 PD: 0.102	$\bar{x}$: ↑ Me: ↑ s: ↓
TSK3	median HNR	all	0.019	HC: 0.136 PD: 0.087	$\bar{x}$: ↑ Me: ↑ s: ↓
TSK3	relF1SD	all	0.041	HC: 0.875 PD: 0.441	$\bar{x}$: ↓ Me: ↓ s: ↓
TSK3	relF2SD	all	0.019	HC: 0.986 PD: 0.990	$\bar{x}$: ↓ Me: ↓ s: ↑

Arrows represent the increase or decrease of mean $\bar{x}$, median Me and standard deviation s.

3.2 MULTIVARIATE ANALYSIS

The performance of the logistic regression model is given in Table 5 and evaluated by the following metrics: Area Under The Curve (AUC), Matthews Correlation Coefficient (MCC), Accuracy (ACC), Sensitivity (SEN) and Specificity (SPE).

Table 5: Results of logistic regression.

	threshold	AUC	MCC	ACC [%]	SEN [%]	SPE [%]
cs	0.51	0.68	0.46	73	67	79
en-US	0.66	0.73	0.45	72	78	67
all	0.51	0.53	0.34	67	64	69

ROC curves show how SEN and SPE vary by changing the classification threshold. The importance of features is given by the coefficients of logistic regression (see Figure 1).

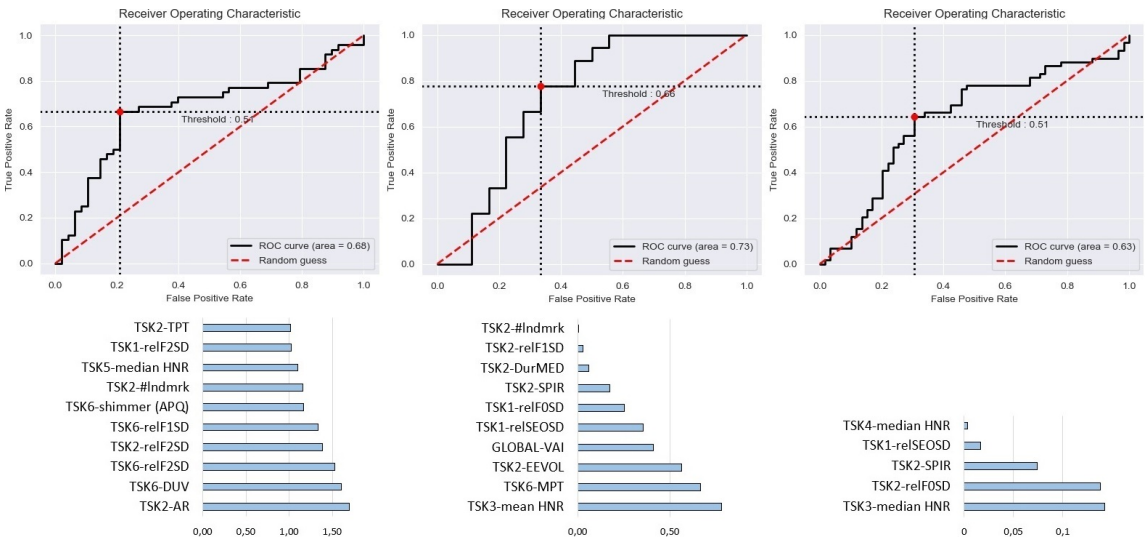

Figure 1: ROC curves and important features (cs: left, en-US: middle, all: right).

4 DISCUSSION

A total of 52 combinations of acoustic features and speech tasks were statistically tested. After correction of p-values, only five remained significant, and only for the model combining both languages. These features are: SPIR of TSK2 and mean HNR, median HNR, relF1SD, relF2SD of TSK3. Therefore, sustained phonation of vowel [a] and reading text played a significant role during the multilingual discrimination of PD and HC. According to the logistic regression results, it is obvious that for different languages, different features have high discrimination power. Five features are important for the model combining both languages including the earlier mentioned SPIR and median HNR. Because these two features are also significant according to statistical tests, they are highly promising for the global automated diagnosis of PD based on acoustic analysis of speech. Classification accuracy and overall performance of the machine learning model are higher when focusing on individual languages than both combined.

5 CONCLUSION

This article deals with the automated diagnosis of hypokinetic dysarthria and related Parkinson's disease focused on the dependence of acoustic features on language. Especially phonatory features and measures of pausing seem to play a significant role in this sense. To verify these features, it is desirable to expand the testing dataset, especially with other languages.

ACKNOWLEDGEMENT

This study has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 734718 (CoBeN), and from grant no. NU20-04-00294 of the Czech Ministry of Health.

REFERENCES

[1] HORNYKIEWICZ, O. Biochemical aspects of Parkinson's disease. Neurology [online]. 1998, 51(2, Supplement 2), S2-S9 [cit. 2020-11-17]. ISSN 0028-3878. Dostupné z: doi:10.1212/WNL.51.2_Suppl_2.S2

[2] DeMaagd G, Philip A. Parkinson's Disease and Its Management: Part 1: Disease Entity, Risk Factors, Pathophysiology, Clinical Presentation, and Diagnosis. P T. 2015 Aug;40(8):504-32. PMID: 26236139; PMCID: PMC4517533.

[3] PARKINSON, James. An Essay on the Shaking Palsy. The Journal of Neuropsychiatry and Clinical Neurosciences [online]. 2002, 14(2), 223-236 [cit. 2020-11-17]. ISSN 0895-0172. Dostupné z: doi:10.1176/jnp.14.2.223

[4] GOETZ, C. G. The History of Parkinson's Disease: Early Clinical Descriptions and Neurological Therapies. Cold Spring Harbor Perspectives in Medicine [online]. 2011, 1(1), a008862-a008862 [cit. 2020-11-17]. ISSN 2157-1422. Dostupné z: doi:10.1101/cshperspect.a008862

[5] OWNWARD, Emily. What Are the Stages of Parkinson's Disease? [online]. [cit. 2019-12-20]. Dostupné z: https://parkinsonsdisease.net/basics/stages/

[6] Ho AK, Iansek R, Marigliani C, Bradshaw JL, Gates S. Speech impairment in a large sample of patients with Parkinson's disease. Behav Neurol. 1999 Jan 1;11(3):131-7. PMID: 22387592.