

Automatické hodnocení pauz v řeči u Parkinsonovy nemoci

J. Hlavnička, J. Ruzs, R. Čmejla

Katedra teorie obvodů, Fakulta elektrotechnická, ČVUT v Praze

Abstrakt

Studie se zabývá metodou pro hodnocení pauz v řeči zdravé a řeči postižené vlivem Parkinsonovy nemoci. Realizace metody spočívala v návrhu klasifikátoru řeč / pauza a sady příznaků pro popis fyziologické schopnosti produkce pauz. Klasifikátor je založen na postupné separaci složek řeči *EM-algoritmem*. Klasifikátor je schopen rozlišit hlásky / konsonanty / pauzu / respiraci jako samostatné třídy. Informační zisk klasifikátoru byl zvýšen přízpusobením hranic okének poloze spektrálních změn. Ze sady 22 příznaků projevilo 15 příznaků statistickou významnost ($p < 0.05$) k odlišení parkinsonického a zdravého mluvčího. Známé příznaky poukazující na pomalost a současné zrychlování řeči u Parkinsonovy nemoci se podařilo vyjádřit automatickou metodou. Nově byly objeveny příznaky vyššího ovlivnění pauz respirací a vyšší zastoupení respirace v pauzách.

1 Úvod

Parkinsonova nemoc (*PN*) je po Alzheimerově nemoci druhým nejrozšířenějším neurodegenerativním onemocněním. Jeho prevalence 0.3% pro celou populaci [1] vzrůstá s věkem např. od 65 let na 1.6% [2]. S ohledem na stárnutí populace ve vyspělých zemích se očekává nárůst počtu pacientů *PN* z 4.1–4.6 milionu pacientů odhadovaných k roku 2005 až na 8.7–9.3 milionu pacientů předvídaných k roku 2030 [3]. *PN* omezuje motoriku i celkovou kvalitu života pacienta, který se postupně stává zátěží pro své okolí. Dobudoucná tedy *PN* představuje velký socio-ekonomický problém.

Diagnostika i monitoring *PN* vyžaduje přítomnost pacienta ve zdravotnickém zařízení. Telemedicínské aplikace pro posouzení *PN* jsou obvykle založeny na detekci zjevných motorických příznaků ze signálů akcelerometru či EMG. Tyto symptomy se však začínají projevovat až při 50–60% postižení klíčového mozkového centra [4, 5]. Náklady na snímací zařízení, nutnost odborné asistence či zátěž pacienta patří mezi další nevýhody přímého snímání motorických projevů. Oproti tomu pořízení řečového signálu je neinvazivní, časově i technicky nenáročné. Současně s tím lidská řeč jako komplexní činnost reflektuje nejen duševní tvorbu řeči, ale také schopnost ovládat jemnou motoriku řečového aparátu a vyžaduje i správnou funkci všech jeho součástí. Každá porucha řečového ústrojí se tak projeví typickou poruchou řeči. To činí z řeči kvalitní zdroj dat pro budoucí výzkum i aplikace nejen *PN*, ale i např. Huntingtonovy nemoci [6], Alzheimerovy nemoci [7], roztroušené sklerózy [8], schizofrenie [9] atd.

Význam řečových poruch u *PN* pro např. brzkou diagnostiku [10, 11] či monitoring progresu nemoci [12, 13] a odpovědi na medikaci [14, 15] již byl předmětem studií. Hodnocení postižení řeči se však opírá o časově náročnou ruční segmentaci řečového signálu. Dalším logickým krokem do budoucna je tedy hodnocení automatizovat. Prezentovaná studie si klade za cíl navrhnout automatickou metodu pro hodnocení produkce pauz. Metoda zahrnuje robustní klasifikátor řeč / pauza a sadu příznaků. Na rozdíl od běžných detektorů řečové aktivity prezentovaný klasifikátor posuzuje řeč a pauzu nikoliv na úrovni vnímání nýbrž na úrovni jejich skutečné fyziologické produkce. Liší se tedy mnohem vyšší citlivostí. Klasifikátory pro tyto účely obvykle spočívají v prahování výkonové obálky empirickou konstantou [16], která je však závislá na jakosti signálu. Pro automatickou klasifikaci byl publikován pouze jeden algoritmus [17], jež klasifikuje na základě odhadu této konstanty. Sada příznaků byla navržena pro odlišení zdravé řeči a řeči postižené *PN* na základě výstupu klasifikátoru.

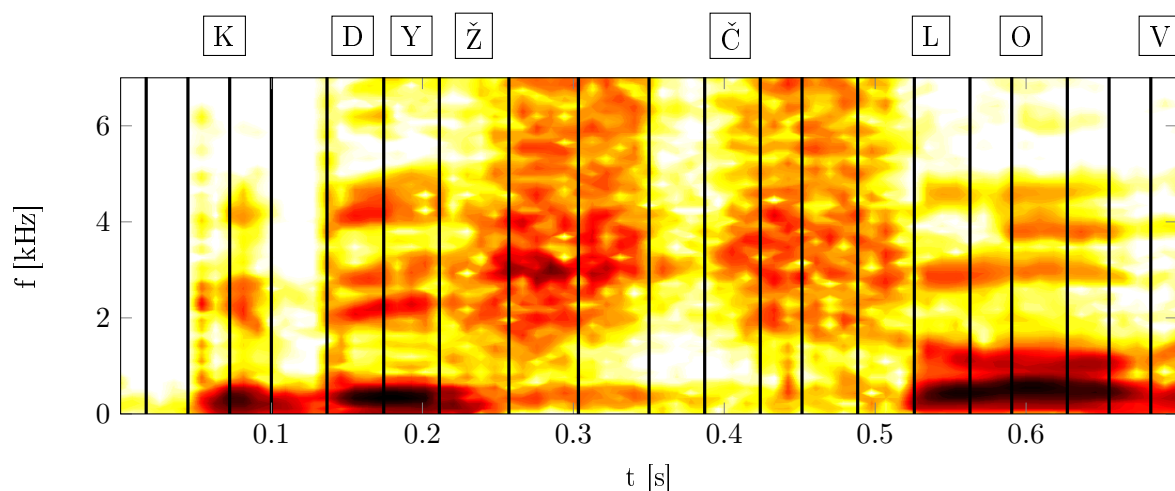
2 Metoda

Databáze mluvicích sestávala z kontrolní skupiny 23 zdravých mluvcí *KS* a skupiny 22 pacientů *PN*. Obě skupiny byly pohlavím i věkově vyvážené. Všichni mluvcí spadají do věkové skupiny 50 – 80 let s průměrným věkem 65 let a směrodatnou odchylkou 9.7 roku pro skupinu *PN* a s průměrným věkem 64 let a směrodatnou odchylkou 10 let pro skupinu *KS*. Obsahem všech promluv byl standartizovaný čtený text o délce 80 slov. Všechny zvukové nahrávky byly ručně segmentovány na třídy *řeč / pauza*. Hranice segmentů byly stanoveny ze spektrogramu jako hranice změn ve spektru.

Prezentovaná metoda předpokládá *multimodální logaritmicko-normální rozdělení* parametrů: výkon signálu (P), počet průchodů nulou (ZCR), rozptyl autokorelační funkce (ACR) a melových kepstrálních koeficientů ($MFCC$). Pro klasifikaci budeme uvažovat třídy: *znělá řeč*, *neznělá řeč*, *respirace*, *pauza*. K odhadu parametrů normálního rozdělení těchto tříd bylo použito *EM-algoritmu*.

Abychom zvýšili jakost informací za současného snížení dimenze dat, nebylo použito konvenčního okénkování s konstantní délkou okna nýbrž pružné segmentace. Hranice okének stanoví automatický algoritmus z výstupu *Autoregresního bayesovského detektoru změn* [18] jako nejvýznamější spektrální změny v okolí o délce d_{min} , jak je ilustrováno na obrázku (1). Délky okének mají proměnnou délku $d \in \langle d_{min}; 2 \cdot d_{min} \rangle$. Odchytky hranic pružné segmentace od hranic ručně segmentovaných vykazovaly napříč celou databází normální rozdělení. Okénkování s konstantní délkou okna oproti tomu vykazuje vždy rovnoměrné rozdělení. Hranice pružné segmentace se tedy blíží lidskému rozhodování o hranicích segmentů. Řečový signál vzorkovaný na $24kHz$ parametrizujeme v rámci okének pružné segmentace. Znělé složky řeči odhadneme v prostoru (P, ZCR, ACR) , který nejlépe popisuje harmonicitu signálu. Ostatní třídy mají charakter šumu a lze je výstižněji vyjádřit v prostoru $MFCC$. Samotná klasifikace je založena na postupné separaci složek řeči. Nejprve jsou separovány znělé složky řeči, poté neznělé, respirace a pauzy. Konsonanty jsou velmi podobné respiraci, nicméně v indoevropské jazykové skupině se vyskytují v těsné blízkosti hlásek cca do $200ms$. Na tomto principu spočívá vyhlazení rozhodování konsonant. Pro porovnání účinnosti jsme implementovali též konvenční klasifikátor *řeč / pauza* navržený pro klasifikaci dysartrické řeči [17].

K vyjádření fyziologických kvalit řečového projevu jsme navrhli sadu 22 příznaků. Při řádově odlišných délkách pauz se uplatňují odlišné orgány řečového aparátu (např. krátké pauzy - rty a jazyk, střední pauzy - též hlasivky, dlouhé pauzy - navíc i respirační aparát). Pro pauzy jsme proto definovali následující časové hranice: Minimální délku pauzy budeme uvažovat od



Obrázek 1: Spektrogram řečového signálu s vyznačenými hranicemi pružné segmentace (svislé černé linie) pro parametr $d_{min} = 20ms$

15ms. Od této hranice po 50ms budeme hovořit o krátkých pauzách, jejichž výskyt je vedlejším produktem artikulace - typicky na rozhraní znělých a neznělých fonémů. Od 50ms do 100ms budeme uvažovat střední pauzy - ty mají význam jednak pro artikulaci explosív, jednak pro zvýraznění kořenů složených slov. Pauzy od 100ms berme za dlouhé pauzy pro formální členění textu a vykonání respirace. Celá sada zahrnuje příznaky:

- **Základní:** Počet všech pauz (nP), počet všech krátkých pauz (nPk), počet středních pauz (nPs), počet všech dlouhých pauz (nPd) a střední artikulační rychlost (mVS).
- **Poměrná zastoupení:** Poměr pauzy a řeči (pPS), zastoupení krátkých pauz (zPk), zastoupení středních pauz (zPs), zastoupení dlouhých pauz (zPd) a zastoupení respirace (pRP).
- **Rychlost produkce:** Střední rytmus pauz (srP), poměrná rytmicita respirace (prR), střední rychlost produkce pauz (mVP), střední produkce krátkých pauz ($mVPk$), střední produkce středních pauz ($mVPs$) a střední produkce dlouhých pauz ($mVPd$).
- **Dynamiku produkce:** Poměr produkce krátkých pauz (ppP), akcelerace rytmu artikulace (arA), akcelerace produkce pauz (aVP), akcelerace produkce krátkých pauz ($aVPk$), akcelerace produkce středních pauz ($aVPs$) a akcelerace produkce dlouhých pauz ($aVPd$).

Délka pauzy je pro sadu příznaků velmi důležitým údajem, proto ohodnocení klasifikátoru bere v potaz délku nalezených pauz podle definice parametru $s(l)$ určené vztahem (1):

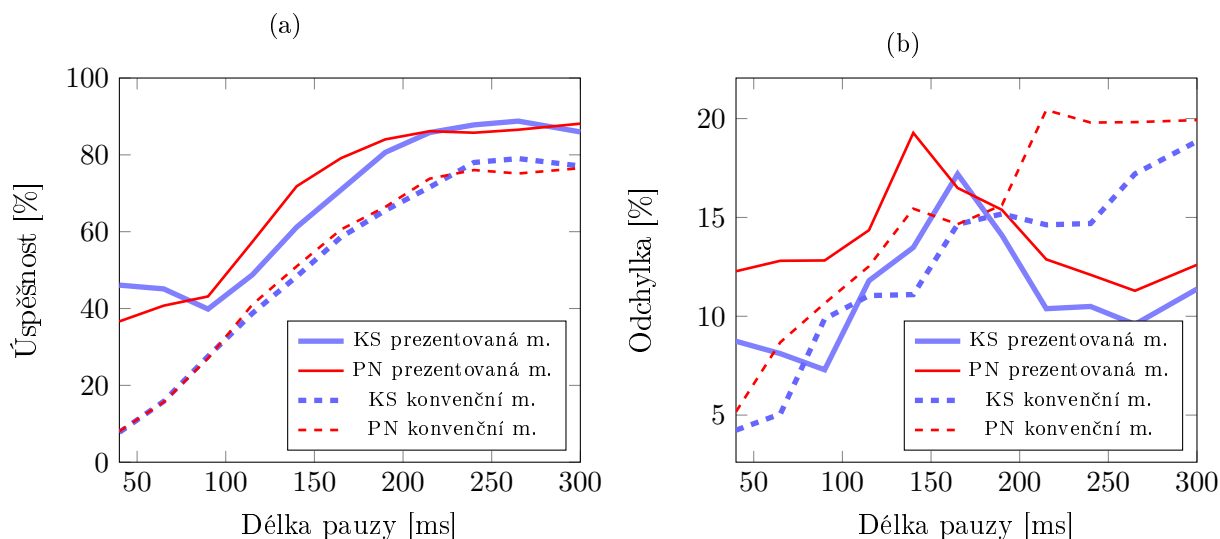
$$s(l) = 100 \cdot \frac{\sum c(l)}{\sum v(l)}, \quad (1)$$

kde $c(l)$ zastupuje správně detekované hranice pauz delších než práh l . Pokud bylo detekováno více pauz delších prahu l než očekáváme, pak hodnota $v(l)$ bude zastupovat počet všech detekovaných pauz delších než práh l . Pokud však bylo detekováno méně pauz než očekáváme, pak bude hodnota $v(l)$ zastupovat očekávaný počet nalezených pauz. Hodnocení tak může spřísnit právě případný nadbytečný či podbytečný počet pauz do jednoho výsledku. Pro lepší čitelnost výsledné informace budeme hodnotit kumulativní účinností v závislosti na rostoucí délce. Příznaky byly hodnoceny nepárovým dvojitým t-testem.

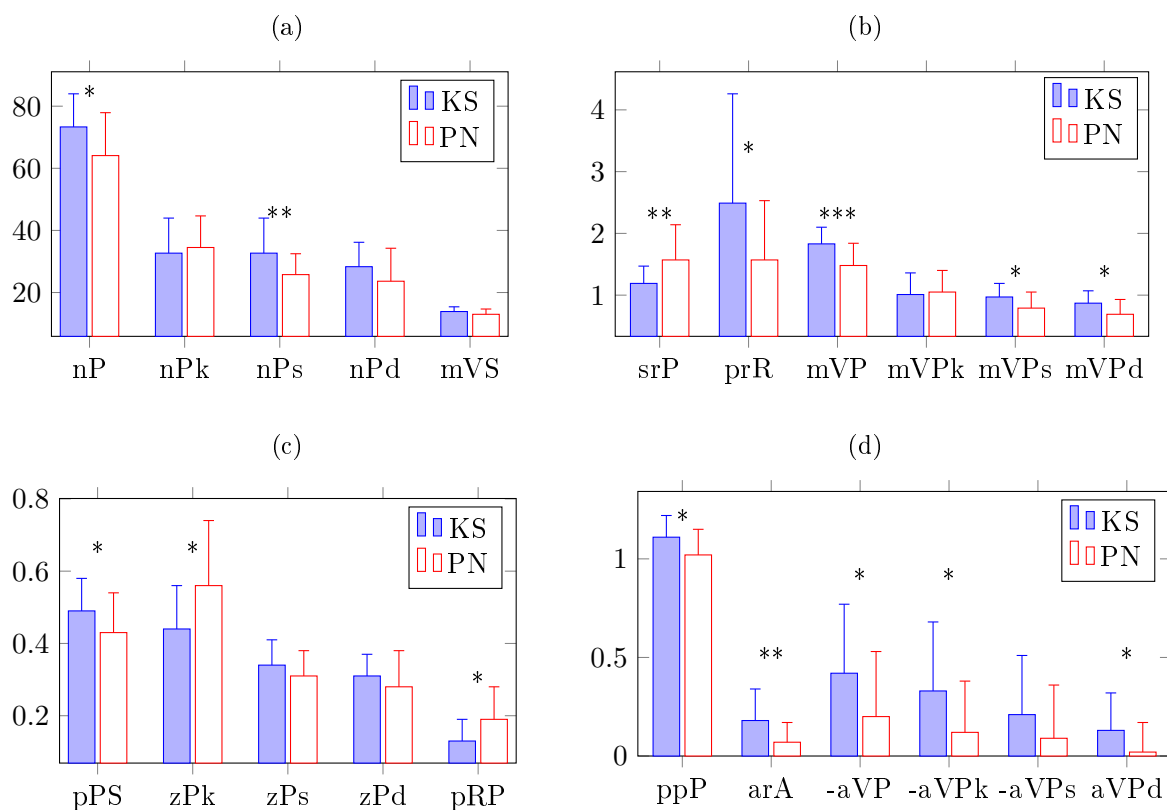
3 Výsledky

Klasifikátor vykázal pro pauzy od 100ms účinnost ($> 50\%$) a pro pauzy od 200ms účinnost ($> 80\%$). Při porovnání s konvenčním algoritmem [17] prospěl klasifikátor pro pauzy ($< 100ms$) o 20% vyšší účinností a pro pauzy ($> 100ms$) o 10% vyšší účinností. Skupina KS i PN zatížila obě hodnocené metody stejnou chybou ilustrovanou na obrázku (2). Z grafů je patrný především výrazný nárůst chybovosti pauz délek $< 100; 200 > ms$. Příčinou je značná podobnost parametrů některých konsonant s parametry pauz. Respirace byla detekována s účinností 75 – 85% se směrodatnou odchylkou 16 – 25%. Testované promluvy obsahovaly ≈ 6 nádechů a každá chyba detekce se tedy projeví silným propadem úspěšnosti.

Ze sady 22 příznaků projevilo 15 příznaků statistickou významnost ($p < 0.05$). Výsledky jednotlivých příznaků ilustruje obrázek (3). Nejsignifikantnějším parametrem pro odlišení řeči PN jse zdá být snížení rychlosti produkce pauz ($p < 0.001$). Ze skupiny základních charakteristik projevilo nejvyšší významnost ($p < 0.01$) nižší počet středních pauz u PN . V sadě rychlostních charakteristik obstály všechny příznaky významnosti ($p < 0.05$) vyjma střední produkce krátkých pauz. Střední rytmus pauz dosáhl hladiny ($p < 0.01$). Z hlediska poměrných zastoupení nelze přijmout za významné ($p < 0.05$) zastoupení středních a dlouhých pauz. Ve skupině dynamických charakteristik nedostala významnosti ($p < 0.05$) pouze akcelerace produkce středních pauz.



Obrázek 2: **a)** průměrná úspěšnost **prezentované** metody ve skupině *KS* (silná modrá) a skupině *PN* (tenká červená) a pro **konvenční** metodu ve skupině *KS* (přerušovaná silná modrá) a skupině *PN* (přerušovaná tenká červená)
b) směrodatná odchylka úspěšnosti **prezentované** metody ve skupině *KS* (silná modrá) a skupině *PN* (tenká červená) a pro **konvenční** metodu ve skupině *KS* (přerušovaná silná modrá) a skupině *PN* (přerušovaná tenká červená)



Obrázek 3: Výsledné ohodnocení příznaků skupin *KS* (plná modrá) a *PN* (červená bez výplně):
a) základní charakteristiky **b)** rychlost produkce **c)** poměrná zastoupení **d)** dynamika produkce
t-test: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4 Diskuze

Významě nižší snížení rychlosti produkce pauz u *PN* spolu se statisticky významným nižším počtem středních pauz a pomalejším rytmem pauz poukazuje na nižší rychlost řeči u *PN*. Významně nižší hodnota příznaku akcelerace rytmu artikulace vypovídá o typickém zrychlování řeči u *PN*. Tomu lze přisoudit také zrychlení celkové produkce a produkce krátkých a dlouhých pauz. Přihlédneme-li také k vyššímu zastoupení krátkých pauz a nižšímu zastoupení středních pauz, děje se tak pravděpodobně zrychlování na úkor středních pauz. Tento poznatek včetně ostatních příznaků již byly předmětem studií [19]. V této studii se však všechny příznaky podařilo nově vyjádřit a ohodnotit plně automatickou metodou. Studie také zkoumala vliv respirace a bylo zjištěno vyšší zastoupení respirace a vyšší ovlivnění rytmu pauz respirací u *PN*. To vzhledem k nesignifikantnímu počtu dlouhých pauz i zastoupení dlouhých pauz svědčí o postižení respiračního aparátu u *PN*.

5 Závěr

Přínosy studie spočívají nejen ve zautomatizování hodnocení produkce pauz. Kromě již známých příznaků produkce pauz u *PN* (nižší rychlost produkce pauz, nižší rychlost artikulace a zrychlování řeči) byly nalezeny i příznaky vyššího ovlivnění pauz respirací a vyšší zastoupení respirace v pauzách u *PN*. Klasifikátor je schopen rozlišit *hlásky / konsonanty / pauzu / respiraci* jako samostatné třídy, což do budoucna rozšiřuje možnou sadu příznaků. Další uplatnění metody lze očekávat i pro hodnocení produkce pauz u Huntingtonovy nemoci, která je z hlediska řečové poruchy protipólem *PN*. Metoda prezentovaná ve studii byla kompletně implementována ve výpočetním prostředí MATLAB®.

Poděkování

Tento výzkum byl podpořen projekty GAČR 102/12/2230 a SGS 12/185/OHK4/3T/13.

Reference

- [1] Rajput, A. and Birdi, S. (1997) Epidemiology of Parkinson's disease, *Parkinsonism and Related Disorders* 3, 175–186.
- [2] deRijk, M. C., Tzourio, C., Breteler, M. M., Dartigues, J. F., Amaducci, L., Lopez-Pousa, S., Manubens-Bertran, J. M., Alpérovitch, A., and Rocca, W. A. (1997) Prevalence of parkinsonism and parkinson's disease in Europe: the EUROPARKINSON Collaborative Study. *European Community Concerted Action on the Epidemiology of parkinson's disease, Journal of Neurology, Neurosurgery and Psychiatry* 62, 10–15.
- [3] Dorsey, E. R., Constantinescu, R., Thompson, J. P., M. Biglan, K., Holloway, R. G., Kieburtz, K., Marshall, F. J., Ravina, B. M., Schifitto, G., Siderowf, A., and Tanner, C. M. (2006) Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030, *Neurology* 68, 384-386.
- [4] Fearnley, J. M. and Lees, A. J. (1991) Ageing and parkinson's disease: substantia nigra regional selectivity, *Brain* 114, 2283-2301.
- [5] Marsden, C. D. (1992) Parkinson's Disease, *Postgraduate Medical Journal* 68, 538-543.
- [6] Rusz, J., Klempíř, J., Baborová, E., Tykalová, T., Majerová, V., Čmejla, R., Růžička, E., and Roth, J. (2013) Objective Acoustic Quantification of Phonatory Dysfunction in Huntington's Disease, *PLOS ONE*, 8 (6).
- [7] Ballard, K. J., Savage, S., Leyton, C. E., Vogel, A. P., Hornberger, M., et al. (2014) Logopenic and Nonfluent Variants of Primary Progressive Aphasia Are Differentiated by Acoustic Measures of Speech Production, *PLOS ONE*, 9(2).

- [8] Tjaden, K. and Wilding, G. (2011) Speech and pause characteristics associated with voluntary rate reduction in Parkinson's disease and Multiple Sclerosis, *Journal of Communication Disorders* 44, 655-665.
- [9] Nicolson, R., Lenane, M., Singaracharlu, S., Malaspina, D., Giedd, J. N., Hamburger, S. D., Gochman, P., Bedwell, J., Thaker, G. K., Fernandez, T., and Wudarsky, M. (2000) Premorbid Speech and Language Impairments in Childhood-Onset Schizophrenia: Association With Risk Factors, *The American Journal of Psychiatry*, 157(5).
- [10] Harel, B. T., Cannizzaro, M. S., Cohen, H., and Reilly N, a. S. P. J. (2004) Acoustic characteristics of Parkinsonian speech: A potential biomarker of early disease progression and treatment, *Journal of Neurolinguistics* 17, 439-453.
- [11] Ruzs, J., Čmejla, R., Růžičková, H., and Růžička, E. (2011) Quantitative acoustic measurements for characterisation of voice and speech disorders in early untreated Parkinson's disease, *The Journal of the Acoustical Society of America* 129, 350-367.
- [12] Maetzler, W., Liepelt, I., and Berg, D. (2009) Progression of Parkinson's disease in the clinical phase: potential markers, *The Lancet Neurology* 8, 1158-1171.
- [13] Skodda, S., R, H., and Schlegel, U. (2009) Progression of dysprosody in Parkinson's disease over time—a longitudinal study, *Movement Disorders* 24, 716-722.
- [14] Ho, A. K., Iannsek, R., Marigliani, C., Bradshaw, J., and Gates, S. (1998) Speech impairment in large sample of patients with Parkinson's disease, *Behavioural Neurology* 11, 131-137.
- [15] Ruzs, J., Čmejla, R., Růžičková, H., Růžička, E., Klempfř, J., Majerová, V., Picmausová, J., and Roth, J. (2012) Evaluation of speech impairment in early stages of Parkinson's disease: a prospective study with the role of pharmacotherapy, *Journal of Neural Transmission* 120, 319-329.
- [16] Green, J. R., Beukelman, D. R., and Ball, L. J. (2004) Algorithmic Estimation of Pauses in Extended Speech Samples of Dysarthric and Typical Speech, *Journal of medical speech-language pathology* 12, 149 - 154.
- [17] Rosen, K., Murdoch, B., Folker, J., Vogel, A., Cahill, L., Delatycki, M., and Corben, L. (2010) Automatic method of pause measurement for normal and dysarthric speech, *Clinical Linguistics and Phonetics* 24, 141 - 154.
- [18] Čmejla, R., Ruzs, J., Bergl, P., and Vokřál, J. (2013) Bayesian changepoint detection for the automatic assessment of fluency, and articulatory disorders, *Speech Communication* 55, 178-189.
- [19] Skodda, S. and Schlegel, U. (2008) Speech rate and rhythm in Parkinson's disease, *Movement Disorders* 23, 985-992.