

AKUSTICKÉ LISTY

České akustické společnosti
www.czakustika.cz

ročník 15, číslo 2

červen 2009

Obsah

Formantová analýza a nové metody pro hodnocení míry artikulace u Parkinsonovy nemoci
Formant analysis and new methods for the assessment of articulation rate in Parkinson's disease

Jan Rusz, Roman Čmejla a Hana Růžičková

3

Akusticko-fonetické míry pro hodnocení neplynulosti čtených promluv
Acoustic-phonetic measures for evaluation of dysfluent speech

Petr Bergl, Libor Černý, Roman Čmejla a Miroslava Hrbková

9

ČESKÁ AKUSTICKÁ SPOLEČNOST

Formantová analýza a nové metody pro hodnocení míry artikulace u Parkinsonovy nemoci

Jan Rusz^{a,b}, Roman Čmejla^a a Hana Růžičková^c

^aKatedra teorie obvodů, ČVUT–FEL, Technická 2, 166 27 Praha 6

^bUniverzita Karlova v Praze, Laboratoř biokybernetiky a počítačové podpory výuky,

Ústav patologické fyziologie 1. LF UK, U nemocnice 5, 128 53 Praha 2

^cNeurologická klinika 1. LF UK a VFN, Kateřinská 30, 128 21 Praha 2

e-mail: [ruszj1;cmejla]@fel.cvut.cz, hruzi@email.cz

This study examined the measures of F_1 and F_2 formants which are the frequently used parameters examined in the acoustic voice analyses and the new methods for articulation rate measures. The first method is based on spectral distance change by using of Bayesian detector and the second is measure of F_2 periodicity in articulation. All measures were investigated in 16 male patients in early untreated stage of Parkinson disease (PD) and compared to 16 male members of the healthy control group (HC) of comparable age. Voice recordings consisted of at least 5 second prolongation of the vowels /a/, /e/, /i/, /o/, /u/ for formant analyses and fast repeatable syllable sequention for articulation rate measures. All calculations were performed with automatic scripts in MATLAB environment. The voice of PD groups demonstrated no significant difference in vowel space area. Present study also suggested that the PD patients' voices are characterized by lower distinct and periodicity of articulation.

1. Úvod

Parkinsonova nemoc (PN) je chronické neurodegenerativní onemocnění charakterizované progresivní ztrátou dopaminergních neuronů v jádrech mozkového kmene, nedostatkem dopaminu a postižením dalších neuromediátorových systémů mozku [1]. Vedle motorických příznaků, jako jsou klidový třes končetin, bradykineze, svalová rigidita a posturální poruchy, se u PN rozvíjejí nonmotorické projevy, mezi něž patří poruchy chování, sensorických, kognitivních a vegetativních funkcí. Hypokinetická dysartrie [2, 3] je nejčastěji popisovanou poruchou řeči u PN a projevuje se symptomy v orofaciální oblasti jako je hypomimie, hypokineze, bradykineze, akineze a dysdiachokineze. Dále sem patří rigidita a ojediněle i třes, deficit v respiraci, fonaci a fonetice.

V předchozích studiích bylo ukázáno [4, 5], že u PN dochází ke změnám v artikulační rychlosti, počtu pauz, intenzitě hlasitosti, základní frekvenci hlasu a amplitudovému a frekvenčnímu kolísání hlasu. PN pacienti dělají v čtených textech menší počet pauz, které jsou ovšem delší a mají sníženou artikulační rychlost. Při rychlé artikulaci klesá jejich intenzita hlasitosti a při čtení textu či samostatné promluvě dochází k většímu kolísání intenzity a ztrátě melodičnosti. Dále mají zmenšený rozsah hlasu při zpívané stupnici a vyslovování tázací či rozkazovací věty. Celkově tedy mají PN pacienti zvětšenou monotónnost hlasu a špatné vnímání důrazu. Při měření amplitudového a frekvenčního kolísání hlasu na hlásku /i/ dochází k větší nepravidelnosti kmitání hlasivek. U všech těchto měření bylo statisticky ověřeno, že dochází ke značným rozdílům mezi pacienty s výskytem PN a kontrolní skupinou (KS). První částí této studie je analýza formantů F_1 , F_2 pro samohlásky /a/, /e/, /i/, /o/, /u/ a stano-

vení vokalizace dané těmito samohláskami. Druhá část přináší zcela nové metody pro hodnocení zřetelnosti a pravidelnosti artikulace. Objektivní analýza hlasových nahrávek není důležitá pouze pro stanovení stupně dysfonie, ale také pro následné monitorování zlepšení stavů pacientů po nasazení léčby.

2. Metody

2.1. Pacienti

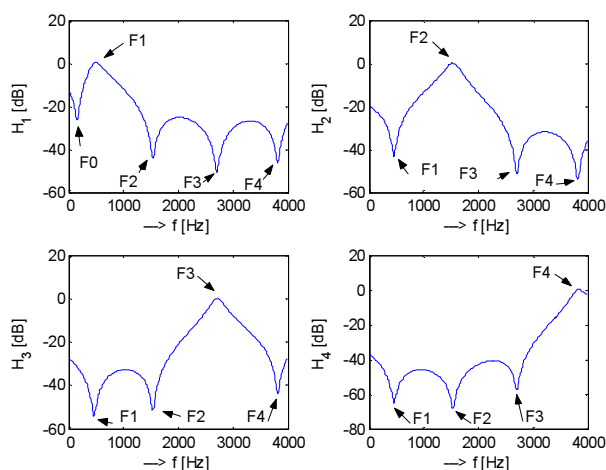
Pro tuto studii byla získána data od 16 českých mužských pacientů v raném stadiu PN [průměrný věk 63 ($\pm$ SD 11,8), délka trvání PN 2,3 ($\pm$ SD 1,4), stadium dle Hoehnové a Yahra 1–2], ve srovnání s 16 muži kontrolní skupiny (KS) [průměrný věk 61,8 ($\pm$ SD 13,3)]. U PN pacientů nebyla zahájena léčba pomocí levodopy ani jiných dopaminergních léčiv. Členové KS byli bez známek neurologických onemocnění a hlasových poruch.

2.2. Nahrávání promluvy

Nahrávky všech pacientů s PN byly pořízeny v odhlučněné místnosti (< 45 dB) Neurologické kliniky 1. lékařské fakulty Univerzity Karlovy v Praze ve vysoké kvalitě s vzorkovací frekvencí 48 kHz. Pacienti s PN byli požádáni o udržovanou fonaci samohlásek /a/, /e/, /i/, /o/, /u/, produkovanou při příjemné hlasitosti a výšce hlasu nejméně po dobu 5 sekund a o rychlé opakování kombinace slabik /pa-ta-ka/, přičemž všichni účastníci byli požádáni o opakování této promluvy a bylo pořízeno 38 vzorků od pacientů s PN a 52 vzorků od účastníků KS.

2.3. Akustické analýzy

Formanty Řeč je artikulována prostřednictvím hlasového ústrojí složeného z několika dutin, které mohou v určitých mezích měnit svůj tvar a objem. Do nich vstupují společně hlasivkový tón s výdechovým proudem vzduchu z plic vytvářející kvaziperiodický signál a aditivní šum. Jako formanty, resp. formantové frekvence jsou označovány ty frekvence, které odpovídají akustickým rezonančním frekvencím těchto dutin. Koncentraci akustické energie v určitých oblastech frekvenčního spektra v řečovém signálu lze velmi zřetelně sledovat na širokopásmových spektrogramech, kde se formanty zobrazují jako tmavé pásy. Při analýze řeči u pacientů s výskytem PN se ve velké většině případů zajímáme pouze o první dva formanty F_1 a F_2 [6–15].



Obrázek 1: Amplitudové charakteristiky čtyř adaptivních formantových filtrů

Pro tuto práci bylo využito automatického trasování formantů pomocí adaptivní banky filtrů v programovacím prostředí MATLAB. Výsledné hodnoty byly pro kontrolu ověřeny v analyzačním softwaru Praat [16]. Algoritmus podle [17, 18] vyhodnocuje formantové frekvence ve znělých segmentech na výstupu adaptivní banky čtyř formantových filtrů (obr. 1). Polohy pólů i nul přenosové funkce filtru se určují z aktuálních hodnot formantových frekvencí. Formantové frekvence se počítají na výstupu banky filtrů a aktualizují se s každým novým vzorkem pomocí LPC 2. řádu. Systém je doplněn o detekci znělý/neznělý a detektor žena/muž, neboť v neznělých a v nízkenergetických úsecích se hodnota formantů vyhodnocuje pomocí klouzavých průměrů. Paralelní výpočet hlasivkového tónu F_0 zajišťuje dostatečné potlačení jeho energie prvním formantovým filtrem, jehož nula se umístí na frekvenci F_0 (obr. 1). To umožňuje eliminovat nesprávný odhad formantové frekvence F_1 . Algoritmus vykazuje relativně vysokou odolnost vůči šumu.

Vokálníká oblast Artikulace je modifikace tvaru a police orgánu řeči (rozumíme jazyka) při tvorbě zvuku. Jed-

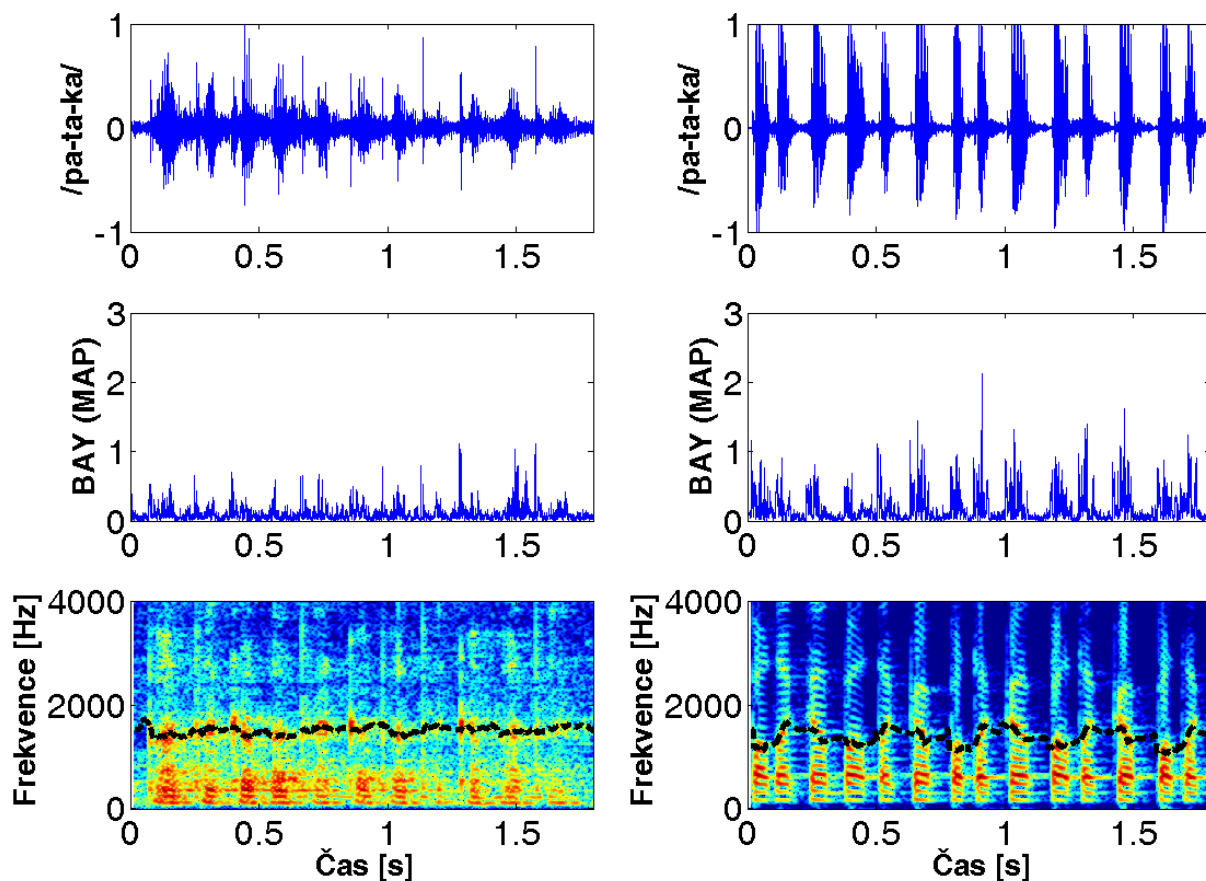
ním ze základních zkoumaných a měřených parametrů souvisejících s formantovou analýzou je určení tzv. vokálníké oblasti [6–14]. Ta je spočítána jako oblast pětistěny a zobrazena vykreslením závislosti $F_1 - F_2$ formantu pro rohové hlásky /a/, /e/, /i/, /o/, /u/ nebo též zjednodušená verze označována jako vokálníký trojúhelník zahrnující tři rohové hlásky /a/, /i/, /u/. Mezi další studie v této oblasti mohou patřit vzájemné poměry mezi formanty F_2/F_1 nebo také euklidovské vzdálenosti mezi jednotlivými rohovými hláskami [6, 7].

2.4. Periodicita F_2

Mezi častá měření patří tzv. měření tranzientů F_2 , často též udávaných jako sklon formantu F_2 [8–11, 13, 15]. Ty jsou reprezentovány jako rychlost pohybu jazyka od souhlásky k samohlásce a zpět. Další studie poukázaly na plošší F_2 při čtení textu či vyslovování vět [9, 15]. Tyto studie nás dovedly k nové metodě hodnocení artikulace, a to stanovení autokorelačního koeficientu $ACF\{F_2\}$ při rychlém opakování kombinace slabik /pa-ta-ka/ (obr. 2), které je často využíváno při hodnocení artikulární rychlosti diadochokineze DDK [9, 15, 19–21]. Takováto kombinace souhlásek rovnoměrně zatěžuje hlasový trakt od artikulace obouretné okluzivy /p/ přes předodásňové /t/ až po měkkopatrové /k/. Větší autokorelační koeficient znamená větší pravidelnost artikulace. Toto měření umožňuje automatické robustní hodnocení pravidelnosti a přesnosti při rychlé artikulaci.

Bayesovská detekce spektrální vzdálenosti

V případě tohoto měření byla využita opět zvuková nahrávka /pa-ta-ka/. Pro hodnocení bylo použito bayesovského autoregresního detektoru změn. Při odvození detektoru uvažujeme signál složený ze dvou úseků, popsány dvěma různými autoregresními modely, mezi nimiž hodnotíme velikost změny. Analytickým řešením Bayesova teorem získáme vztah, který je funkcí pouze analyzovaných dat, jejich délky a řádu autoregresních modelů [22]. Místo s největší změnou v signálu je pak určeno maximální pravděpodobností, kterou vypočítáme z uvedeného vztahu. Obsahuje-li však signál více změn, pak odvozený vztah nelze přímo použít. Tento předpoklad jediné změny bývá v praxi velmi omezujícím, neboť v řečovém signálu je vždy přítomno více změn. Nevýhodu však lze překonat výpočtem pravděpodobnosti v klouzavém okně pevné délky a použitím normování [23, 24]. Pravděpodobnost změny se pak počítá pro vzorek signálu ležícího uprostřed okna. Jinými slovy, výstup algoritmu je jakousi mírou nepodobnosti mezi signálem v levé a pravé polovině okna, ve kterém celý signál vzorek po vzorku procházíme. Pro detekci byl použit normovaný rekursivní bayesovský autoregresní detektor změn 6. řádu s délkou okna 512 vzorků. Vyšší hodnoty výstupního signálu jsou úměrně větší spektrální vzdálenosti dvou sousedních segmentů a představují větší zřetelnost artikulace. Ze signálů pořízených detektorem byly stanoveny koeficienty $BAY\mu$ (střední hodnota) a $BAY\sigma$ (směrodatná odchylka). Použitá metoda se uká-



Obrázek 2: Průběh promluvy /pa-ta-ka/ pacienta s výskytem PN (vlevo) a účastníka KS (vpravo). Průběh promluvy (nahore); průběh výstupu rekurzivního Bayesovského autoregresního detektoru změn – odhadu MAP (uprostřed); spektrogram promluvy, průběh F_2 je zobrazen černě (dole)

		/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/	/a/	/e/	/i/	/o/	/u/	Vokálníká
		F_1	F_1	F_1	F_1	F_1	F_2	F_2	F_2	F_2	F_2	F_2/F_1	F_2/F_1	F_2/F_1	F_2/F_1	F_2/F_1	oblast
		[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[Hz]	[–]	[–]	[–]	[–]	[–]	[Hz ²]
PN	μ	707	627	376	567	402	1144	1593	2150	940*	788	1,63	2,58	5,79	1,68	1,98	259161
	σ	107	61	44	85	55	118	175	139	83	85	0,18	0,46	0,70	0,19	0,25	89077
KS	μ	664	602	394	521	383	1117	1598	2216	875*	736	1,68	2,65	5,59	1,69	1,96	248493
	σ	57	52	48	47	43	94	161	154	32	47	0,20	0,33	0,68	0,13	0,21	68850

μ – střední hodnota; σ – směrodatná odchylka; * $p = 0,01$

Tabulka 1: Výsledky analýzy formantů a vokálníkové oblasti

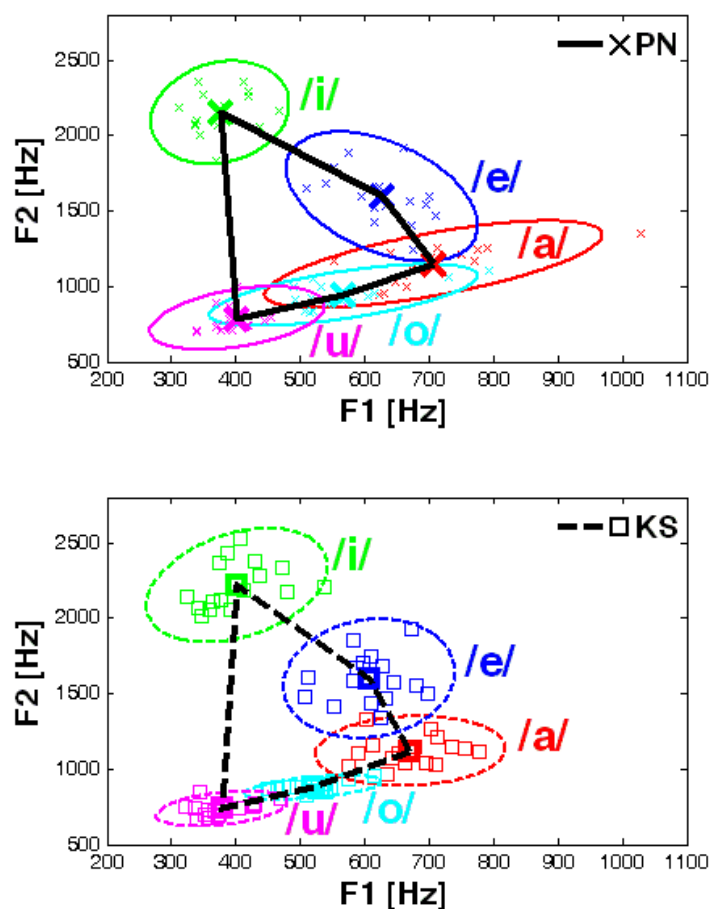
zala jako velmi vhodná pro automatické robustní hodnocení zřetelnosti artikulace (obrázek 2).

3. Výsledky

3.1. Formantová analýza

Pro všechna měření byla stanovena střední hodnota a směrodatná odchylka a výsledky byly ověřeny statistickými t -testy. V tabulce 1 jsou uvedeny hodnoty formantů pro jednotlivé hlásky, jejich vzájemné poměry a obsahy vokálníkové plochy. Obrázek 3 ilustruje klasifikace jednotlivých

formantů do skupin a vyznačuje střední hodnotu výsledné plochy pro obě skupiny PN a KS. Hlavním měřením těchto analýz je vokálníká oblast dana pěti základními samohláskami. Její rozměry jsou pro obě skupiny takřka shodné a rozdíl nedosáhl žádné statistické významnosti. Hodnoty jednotlivých formantů s výjimkou F_2 pro samohlásku /o/ taktéž nedosáhly statistické významnosti. Dalším měřením jsou vzdálenosti překrytí ploch u jednotlivých 95% pravděpodobnostních rozložení, které jsou spočteny ze střední hodnoty a rozptylu u každé samohlásky. Z obrázku 3 a tabulky 2 lze pozorovat, že překryvy elips u pacientů



Obrázek 3: Zobrazení vokálního pětistěnu středních hodnot formantů u pacientů s výskytem PN (nahore) a KS (dole)

s výskytem PN dosahují mnohem větší vzdálenosti než u účastníků KS. To může být způsobeno špatnou přesností artikulace. Vzdálenosti v tabulce 2 jsou určeny protnutím elipsy a hrany vokálního obsahu mezi dvojicí samohlásek. Záporná hodnota udává vzdálenost překryvu dvou sousedních oblastí, kladná hodnota udává vzdálenost mezi dvěma oblastmi.

Rozdíl

$$\Delta = \text{vzdálenost hranic}(KS) - \text{vzdálenost hranic}(PN)$$

představuje změnu v překrytí hranice oblastí mezi sousedními vokály. Kladné hodnoty změny překrytí Δ nám říkají, že ve všech případech došlo k nárůstu hranic vokálních oblastí sousedních samohlásek ve skupině PN, nejvíce pak mezi samohláskami /A–O/, kde dochází k největšímu artikulačnímu splnutí.

3.2. Metody pro hodnocení artikulace

Obě metody pro hodnocení artikulace statisticky potvrdily rozdíl mezi skupinami PN a KS. Z výsledků (tab. 3,

obr. 4) lze konstatovat, že u pacientů s výskytem PN dochází k výraznému snížení pravidelnosti a zřetelnosti artikulace, přičemž metody získané metodou bayesovské detekce byly statisticky významnější.

4. Diskuze

Zmenšená vokální oblast mezi samohláskami u dysartrických mluvčích vzhledem ke KS zdravých mluvčích je

Vzdálenost hranic [Hz]	I–E	E–A	A–O	O–U	U–I
PN	73,82	–165,39	–356,09	–217,18	841,17
KS	132,82	–91,92	–11,66	–49,74	1012,10
Δ	+59,00	+73,47	+344,43	+167,44	+170,93

Tabulka 2: Výsledky vzdáleností mezi hranicemi jednotlivých oblastí výskytu samohlásek

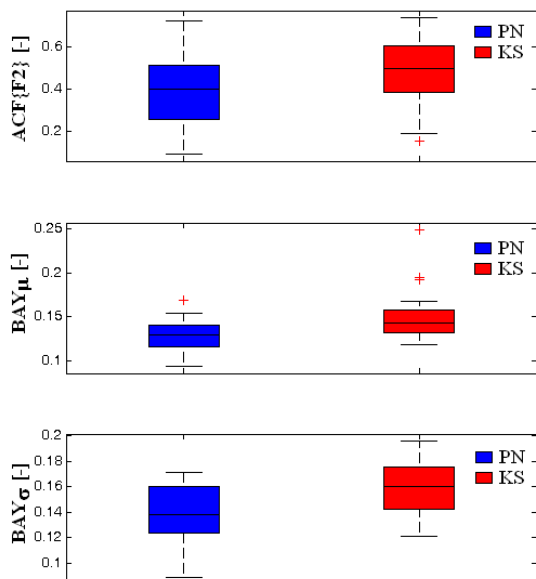
popisována v mnoha studiích [10–12, 25–28]. U pacientů s výskytem PN je udávána zmenšená vokální oblast oproti KS, avšak u tohoto rozdílu nebyla nalezena žádná statistická významnost [6, 8–11, 14]. V této studii bylo analyzováno 16 českých mužů v raném stadiu PN, byla stanovena jejich vokální oblast a formanty pro jednotlivé samohlásky. Toto měření potvrdilo, že mezi oběma skupinami nedochází ke statisticky významným posunům. Avšak v této studii došlo k pozorování, že oblasti výskytu jednotlivých samohlásek se více protínají u PN než u KS, což může být způsobeno zhoršenou přesností artikulace.

	PN		KS	
	μ	σ	μ	σ
ACF{F₂} [-]				
/pa-ta-ka/	0,3952*	0,1680	0,4790*	0,1446
BAY_{μ} [-]				
/pa-ta-ka/	0,1281**	0,0181	0,1465**	0,0223
BAY_{σ} [-]				
/pa-ta-ka/	0,1385**	0,0220	0,1594**	0,0199

* $p < 0,01$;

** $p < 0,001$.

Tabulka 3: Výsledky analýz pro metody hodnocení artikulace



Obrázek 4: Výsledky metod hodnocení artikulace u pacientů s výskytem PN a KS

Významných rozdílů mezi skupinami PN a KS bylo dosaženo při použití nových metod pro hodnocení artikulace.

Tato nová měření umožňují snadné a přesné hodnocení pravidelnosti a zřetelnosti artikulace.

5. Závěr

Studie ukázala, že u pacientů s PN oproti KS nevzniká významný posun formantů u samohlásek a nemají výrazně rozdílnou vokální oblast, ale může u nich docházet k zhoršené přesnosti artikulace. Nová měření prokázala, že u rychlého opakování kombinace slabik /pa-ta-ka/ dochází ke snížené pravidelnosti a přesnosti artikulace, a práce zároveň přináší nové robustní metody pro jejich vyhodnocení. Výsledky studie mohou být použity pro možnosti rozšíření objektivního hodnocení dysartrií nejen u PN. Umožňují tak pochopení změn vznikajících v řeči pacientů již v rané fázi PN, které se nedají zachytit pouhým poslechem, a dávají tak zpětnou vazbu pacientům či lékařům.

Poděkování

Tato práce je podporována z výzkumných záměrů „Transdisciplinární výzkum v oblasti biomedicínského inženýrství“ (MSM 6840770012), „Neuropsychiatrické aspekty neurodegenerativních onemocnění“ (MSM 0021620849) a (MSM 0021620806), grantů „Analýza a modelování biologických a řečových signálů“ (GAČR 102/08/H008) a Rozpoznávání mluvené řeči v reálných podmínkách (GAČR 102/08/0707), projektů „Non-motorické projevy Parkinsonovy nemoci“ (IGA MZČR NR9220) a (MSM 2C06031), a firmou Creative Connection, s. r. o.

Autoři děkují MUDr. Janě Picmausové, MUDr. Veronice Majerové, MUDr. Jiřímu Klempířovi, Ph.D. a doc. MUDr. Janu Rothovi, CSc. za poskytnutí souboru klinických dat a prof. MUDr. Evženu Růžickovi, DrSc. za koncepci a vedení klinické studie a za připomínky k rukopisu.

Reference

- [1] Hornykiewicz, O.: *Biochemical aspects of Parkinson's disease*, Neurology, 1998, 51(Suppl 2):2–9.
- [2] Canter, G. J.: *Speech characteristic of patients with Parkinson's disease. I. Intensity, pitch, and duration*, J. Speech Hear Disord, 1963, 28:221–229.
- [3] Darley, F. L., Aronson, A. E., Brown, J. R.: *Differential diagnostic patterns of dysarthria*, J. Speech Hear Res, 1969, 12:426–496.
- [4] Rusz, J., Čmejla, R.: *Analýza rychlosti řeči a intenzity u Parkinsonovy nemoci*, Akustické listy, 2008, 14(2–4):13–16.
- [5] Rusz, J., Čmejla, R., Růžicková, H.: *Analýza základní frekvence, amplitudového a frekvenčního kolísání hlasivek u Parkinsonovy nemoci*, Akustické listy, 2009, 15(1):13–18.

- [6] Goberman, A. M., Elmer, L. W.: *Acoustic analysis of clear versus conversational speech in individuals with Parkinson disease*, Journal of Communication Disorders, 2005, 38:215–230.
- [7] Sapir, S.: *Effect of Intensive Voice Treatment (the Lee Silverman Voice Treatment [LSVT]) on Vowel Articulation in Dysarthric Individuals With Idiopathic Parkinson Disease: Acoustic and Perceptual Findings*, Journal of Speech, Language, and Hearing Research, 2007, 50:899–912.
- [8] Goberman, A. M.: *Correlation between acoustic speech characteristics and non-speech motor tasks in Parkinson's disease*, Med. Sci. Monit, 2005, 11(3):109–116.
- [9] Goberman, A. M., Coelho, C.: *Acoustic analysis of Parkinsonian speech I: Speech characteristics and L-Dopa therapy*, NeuroRehabilitation, 2002, 237–246.
- [10] Tjaden, K.: *Anticipatory Coarticulation in Multiple Sclerosis and Parkinson's Disease*, Journal of Speech, Language, and Hearing Research, 2003, 46:990–1008.
- [11] Weismer, G., Jeng, J., Laures, J. S., et al.: *Acoustic and Intelligibility Characteristics of Sentence Production in Neurogenic Speech Disorders*, Folia Phoniatr Logop, 2001, 53:1–18.
- [12] Tjaden, K., Rivera, D., Wilding, G.: *Characteristics of the Lax Vowel Space in Dysarthria*, Journal of Speech, Language, and Hearing Research, 2005, 48:554–566.
- [13] Yunusova, Y., Weismer, G., Kent, R. D., Rusche, N. M.: *Breath-Group Intelligibility in Dysarthria: Characteristics and Underlying Correlates*, Journal of Speech, Language, and Hearing Research, 2005, 48:1294–1310.
- [14] Watson, P. J., Munson, B.: *Parkinson's disease and the effect of lexical factors on vowel articulation*, J. Acoust. Soc. Am., 2008, 124(5).
- [15] Rosen, K. M., Kent, R. D., Delaney, A. L.: *Parametric quantitative acoustic analysis of conversation produced by speakers with dysarthria and healthy speakers*, Journal of Speech, Language, and Hearing Research, 2006, 49:395–4.
- [16] Boersma, P., Weenink, D.: *Praat, a system for doing phonetics by computer*, Glot International, 2001, 5(9/10):341–345.
- [17] Mustafa, K., Bruce, I. C.: *Robust formant tracking for continuous speech with speaker variability*, IEEE Transactions on Speech and Audio Processing, 2006.
- [18] Mustafa, K.: *Robust formant tracking for continuous speech with speaker variability*, M.A.Sc. dissertation, Dept. Elect. and Comp. Eng. McMaster Univ., Hamilton, ON, Canada, 2003.
- [19] D'Alatri, L., Paludetti, G., Contarino, M. F., et al.: *Effects of Bilateral Subthalamic Nucleus Stimulation and Medication on Parkinsonian Speech Impairment*, J. Voice, 2006.
- [20] Tjaden, K., Watling, E.: *Characteristics of Diadochokinesis in Multiple Sclerosis and Parkinson's disease*, Folia Phoniatr et Logopaedica, 2003, 55:241–259.
- [21] Midi, I., Dogan, M., Koseoglu, M., et al.: *Voice abnormalities and their relation with motor dysfunction in Parkinson's disease*, Acta Neurol Scand, 2008, 117:26–34.
- [22] Ó Ruanaidh, J. J. K., Fitzgerald, W. J., *Numerical Bayesian methods applied to signal processing*. Springer-Verlag, New York, 1996.
- [23] Čmejla, R., Sovka, P.: *Family of Recursive Bayesian Detectors and Applications for Signal Segmentation*. In Inter-Noise, 2004, Proceedings, Prague: Czech Acoustic Society.
- [24] Čmejla, R., Sovka, P.: *Recursive Bayesian Autoregressive Change-point Detector for Sequential Signal Segmentation*, In EUSIPCO-2004 – Proceedings, Wien.
- [25] Weismer, G., Laures, J. S., Jeng, J.-Y., Kent, R. D., Kent, J. F.: *Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis*, Folia Phoniatr et Logopaedica, 2000, 52:201–219.
- [26] Ziegler, W., Cramon, D. R.: *Vowel distortion in traumatic dysarthria: Lip rounding versus tongue advancement*, Phonetica, 1983, 40:312–322.
- [27] Turner, G. S., Tjaden, K., Weismer, G.: *The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis*, Journal of Speech and Hearing Research, 1995, 38:1001–1013.
- [28] Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., Edwards, B. W.: *Lexical boundary error analysis in hypokinetic and ataxic dysarthria*, Journal of the Acoustical Society of America, 2000, 107:3415–3424.

Akusticko-fonetické míry pro hodnocení neplynulosti čtených promluv

Petr Bergl^a, Libor Černý^b, Roman Čmejla^a a Miroslava Hrbková^b

^aČVUT-FEL, Technická 2, 166 27 Praha 6

^bFoniatriká klinika 1. LF UK a VFN, Žitná 24, 120 00 Praha 2

Several methods for objective evaluation of dysfluent speech are introduced: silence to speech ratio; number of segments per minute; frequency of edges of power envelope; usage of levels in speech signals. These can be grouped into one combined parameter correlating strongly ($r = 0.73$) to subjective judgment made by two authors (the judgment divided all participants into 5 classes according to dysfluency severity). Higher correlation coefficient is achieved for parameters based on detection of spectral changes – e.g. $r = -0.77$ for Bayesian detector and $r = -0.8$ for number of phonemes in HTK. Output of the parameters can be clustered using discriminative analysis. A combination of all described parameters correctly classified 60 % of participants in set 1 (121 utterances in good audio quality) and 67 % of participants in set 2 (33 utterances in poor audio quality caused by additive interference etc.).

1. Úvod

Poruchy plynulosti řeči mohou mít mnoho příčin. Může se jednat o menší dispozice k rozmanitému mluvenému projevu, charakterové vlastnosti či výsledek emocionální rozladěnosti jedince. Kromě toho můžou být jejich důvodem různá onemocnění, např. koktavost. Koktavost (*balbuties*) se projevuje opakováním určitých hlásek či slabik (*repetice*), prodlužováním hlásek (*prolongace*), četnými pauzami apod. Koktaví si jsou své poruchy dobře vědomi, s tím spojený stres pak může vést až ke strachu z mluvení (*logofobie*).

Teorii o vzniku koktavosti je nespočet, s tím je spojeno i značné množství používaných terapeutických přístupů (až 250). Volba léčebného postupu tak není nikterak jednoduchou otázkou, správné posouzení tíže poruchy a výskytu příznaků je velmi důležité. Metoda, která by objektivně určila vážnost poruchy řeči, by byla velkým přínosem. Umožnila by zejména [1, 2]: 1) Určení tíže poruchy (jakousi „absolutní hodnotu“), 2) Hodnocení výsledků léčby, 3) Porovnání efektivnosti a účinnosti léčebného postupu s jiným léčebným postupem.

Pro posouzení vážnosti se dají použít různé škály, které ale mají základ v *subjektivním* posouzení výskytu *repetic* a *prolongací*. Tento příspěvek popisuje metodu, která na základě analýzy audio nahrávek *balbutiků* odhaduje tíži poruchy plynulosti řeči.

2. Databáze promluv

Základem výzkumu je databáze obsahující audio nahrávky od cca 160 mluvčích, z tohoto počtu je 12 kontrolních zdravých jedinců, zbytek je tvořen pacienty různého věku, s různou vážností poruchy plynulosti řeči. Databáze vznikla na Foniatriké klinice 1. LF UK a VFN. U většiny mluvčích byly zaznamenány jak čtené, tak volně formulované promluvy. Experimenty popsané v tomto článku se

zaměřují na čtený text, konkrétně úryvek (cca 75 slov) z Babičky od Boženy Němcové. Důležitým faktem je, že během roku 2008 byly všechny čtené promluvy posouzeny dvěma foniatry, kteří tíži poruchy plynulosti řeči popsali pomocí pětistupňové klasifikace (známkou v rozmezí 0 až 4) na základě relativní četnosti neplných slov c_{nepl} , konkrétně:

- 0 – žádné příznaky koktavosti, $c_{nepl} = 0 \%$
- 1 – *balbuties levis*, $0 \% < c_{nepl} \leq 5 \%$
- 2 – *balbuties gradus medius*, $5 \% < c_{nepl} \leq 20 \%$
- 3 – *balbuties gravis*, $20 \% < c_{nepl} \leq 60 \%$
- 4 – *balbuties gravis inaptus*, $c_{nepl} > 60 \%$ nebo promluva obsahuje dlouhé prefonační spazmy (bloky) v délce trvání 2 s a více.

Poznamenejme, že hodnocení zahrnuje jak neplynulosti způsobené koktavostí (*tony*, *klony*, *tonoklony* či prefonační spazmy), tak neplynulosti vyplývající z technických obtíží při čtení, prostého přeroknutí či projevů specifické poruchy čtení – *dyslexie*.

Pro každého mluvčího jsou tedy k dispozici dvě známky (jedna od každého lékaře), sloužící jako kontrolní pro navrhované algoritmy (viz dále).

Databáze byla rozdělena do dvou částí. První čítá 121 jedinců a obsahuje záznamy dobré technické kvality. Nejmladšímu mluvčímu bylo v době nahrávání 7 let a 5 měsíců, nejstaršímu pak 49 let a 5 měsíců. Ženy a dívky představují 23 % (28 ze 121) této skupiny, kterou budeme v dalším textu nazývat *soubor 1*. Druhá část databáze obsahuje 33 jedinců se záznamy horší technické kvality (*brum*, *kolísání zesílení* během nahrávky). Nejmladšímu mluvčímu bylo v době nahrávání 8 let a 6 měsíců, nejstaršímu pak 52 let a 4 měsíce. Ženy a dívky představují 12 %

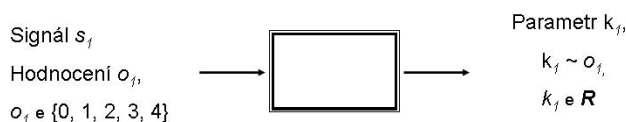
(4 ze 33) skupiny, kterou budeme v dalším textu označovat *soubor 2*. Počet jedinců se známkami 0 až 4 ukazuje v obou skupinách tabulka 1.

Množina	„0“	„1“	„2“	„3“	„4“	Celkem
Soubor 1	15	27	43	26	10	121
Soubor 2	1	13	8	8	3	33

Tabulka 1: Počet mluvčích v souborech

3. Popis parametrů

Prvním krokem při tvorbě celého systému je definování parametrů vypočtených na základě řečové nahrávky, přičemž hodnoty takovýchto parametrů by měly korelovat s hodnocením lékařů. Pro jeden signál toto schematicky zachycuje obr. 1. Konkrétní popis pak následuje v dalších odstavcích.



Obrázek 1: Základní princip analýzy

3.1. Poměr ticho – řeč

Základem tohoto parametru je detektor ticho – řeč (Voice Activity Detector – VAD), který je založen na klasifikátoru SVM (Support Vector Machine). Pro účely této práce stačí SVM chápat jako černou krabičku, na jejíž vstup přivedeme parametry krátkého úseku signálu (např. energii, počet průchodů nulou) a na jejímž výstupu obdržíme 0 či 1 (ticho vs. řeč). Konkrétně byly signály segmentovány s pevnou délkou okna $L = 50$ ms a překryvem $\frac{2}{3}L$. Pro každý segment byly určeny následující parametry: 1) počet průchodů nulou vztahený k délce segmentu; 2) logaritmus energie úseku; 3) první *lpc* koeficient; 4) logaritmus energie chyby predikce; 5) korelační koeficient; 6) první a druhý spektrální moment; 7) pět maxim segmentu, přičemž první je maximem v první pětině segmentu, druhé je maximem v druhé pětině segmentu atd. Pro 4 promluvy byl natrénován SVM klasifikátor. Úspěšnost na kontrolních datech (pátá promluva) byla přes 90 %. Tímto způsobem byly hodnoceny všechny promluvy ze souboru 1. Pro libovolný signál lze snadno určit celkovou délku trvání řeči t_r a celkovou dobu ticha t_t . Poměr ticho/řeč pak vyjádříme jako:

$$P_{t/r} = \frac{t_t}{t_r}. \quad (1)$$

Výsledky jsou dokumentovány v tabulce 2. Reprezentace je rozdělena na dvě části: 1) Neplynulost posouzená prvním hodnotitelem (levá část tabulky), 2) Neplynulost

posouzená druhým hodnotitelem (pravá část). Pro oba lékaře je uveden korelační koeficient (r) a výsledek Wilcoxonova znaménkového testu (W) pro hladinu významnosti $\alpha = 0,01$, který testuje hypotézu, že mediány mezi jednotlivými skupinami (tj. mezi 0–1, 1–2, 2–3, 3–4) jsou stejné. Např. zápis „0 0 1 0“ značí, že hypotézu o shodných mediánech zamítáme mezi daty pro známky 2 a 3. Volně přeloženo, zkoumaný parametr by mohl být užitečný pro rozoznávání, zda mluvčí patří do skupiny 2 nebo 3. Pro ideální parametr bychom tak měli vidět čtyři jedničky.

Detailednější výsledky nalezneme např. v [3]. Dokládají, že poměr ticho/řeč roste s vážností poruchy, což jistě není překvapující. Dá se předpokládat, že promluvy balbutiků obsahují více ticha. Lepšího výsledku bylo dosaženo pro neplynulost posouzenou prvním hodnotitelem – korelační koeficient je 0,69, Wilcoxonův test pak napovídá, že největším úskalím je rozpoznat mezi nultou a první skupinou a mezi třetí a čtvrtou.

Další jednoduchý parametr, který můžeme při znalosti rozložení ticho/řeč určit, je průměrný počet úseků (tj. ticha i řeči dohromady) za jednu minutu. Řeč balbutiků je charakteristická četnými pauzami a bloky, dá se proto očekávat, že s tíhou poruchy poroste průměrný počet úseků (detaily opět v [3]). Nicméně korelační koeficient (viz druhý řádek v tabulce 2) je poměrně nízký (0,44), výsledkem Wilcoxonova testu je pouze jedna jednička.

3.2. Obsazení úrovní intenzity

Pokud zdravý jedinec čte, poměrně rovnoměrně využívá hladiny intenzity zvuku, tj. při čtení nekřičí, na slova nedává přehnaný důraz apod. Oproti tomu koktaví často při vyslovování vyvíjejí větší snahu, kladou větší důraz na počátky slov apod. To by se mělo projevit rozdíly v obsazení jednotlivých úrovní intenzity (tj. v amplitudách vzorků řečového signálu).

Postup analýzy je následující. Signály zbavíme střední hodnoty a normujeme. Hodnoty vzorků se tak budou pohybovat od -1 do $+1$. Poté definujeme 200 intervalů, v nichž se mohou pohybovat, tj. od -1 do $+1$ s krokem 0,01. Pro každý interval spočteme jeho obsazení, tj. kolik vzorků do něj spadá. Výsledky vydělíme počtem vzorků (délkou signálu), obdržíme relativní četnosti. Podrobnější popis nalezneme v [4]. Hodnoty zdravých a koktavých jedinců se nejvýrazněji odlišují v „příhrádkách“ kolem hodnoty $-0,06$. Četnost v této příhradce by se tedy dala použít jako klasifikátor zdravých vs. koktavých jedinců. Jak silně koreluje s hodnocením lékařů, nalezneme v tabulce 2. Korelační koeficient je vyšší pro prvního hodnotitele ($-0,63$), naopak Wilcoxonův test vychází lépe pro druhého (dvě jedničky místo jedné). Naprosto stejně můžeme postupovat i pro pět příhrádek obklopujících hodnotu $+0,03$ (tedy téměř symetrické příhrádky jako v předchozím případě), přičemž výsledky se o mnoho neliší od těch pro příhrádky kolem $-0,06$, korelační koeficient mírně poklesl, výstup Wilcoxonova testu je stejný.

3.3. Analýza energetické obálky

Nyní se budeme zajímat o energetickou obálku promluv. Signál normujeme, zbavíme stejnosměrné složky a každý vzorek umocníme na druhou. Výsledné vzorky přivedeme na vstup integrátoru, čímž obdržíme energetickou obálku promluvy. Práh p , na jehož základě určíme náběžné hrany obálky, nastavíme na 0,1 (viz [3]). Průměrný počet náběžných hran spočteme pro všech 121 jedinců, podobně jako v předchozích případech. Korelační koeficient je $-0,59$ (tab. 2), již tradičně je vyšší (v absolutní hodnotě) pro prvního hodnotitele. Za zmínku stojí však zejména fakt, že poprvé vidíme kladný výsledek Wilcoxonova testu mezi nultou a první skupinou. Zajímavé je, že velmi podobné výsledky lze nalézt, pokud energetickou obálku nahradíme její derivací.

Parametr	1. hodnotitel		2. hodnotitel	
	r	W	r	W

Parametry popsané v kap. 3.1 až 3.4:

Poměr ticho/řeč	0,69	0 1 1 0	0,61	0 1 1 0
Počet úseků	0,44	0 1 0 0	0,46	0 1 0 0
Úrovně $-0,06$	$-0,63$	0 0 1 0	$-0,59$	0 1 1 0
Úrovně $+0,03$	$-0,62$	0 0 1 0	$-0,56$	0 1 1 0
Náb. hrany ener.	$-0,59$	1 0 0 1	$-0,56$	0 0 0 0
Souhrnný param.	0,73	0 1 1 1	0,66	0 1 1 0

Parametry popsané v kap. 3.5:

Počet max. Bay. <i>std</i> intervalu	$-0,76$ 0,46	0 1 1 0 1 1 1 0	$-0,77$ 0,40	0 1 1 0 0 1 1 0
Počet max. GLR <i>std</i> intervalu	$-0,76$ 0,43	0 1 1 1 1 1 1 0	$-0,75$ 0,36	0 1 1 0 0 1 1 0
Počet hl. HTK <i>std</i> intervalu	$-0,80$ 0,66	0 1 1 1 0 1 1 0	$-0,75$ 0,60	0 1 1 0 0 1 1 0

Parametry popsané v kap. 3.6:

<i>std</i> energie	0,67	0 1 1 0	0,64	0 1 1 0
<i>std</i> znělost	0,40	0 1 1 0	0,33	0 1 1 0

Tabulka 2: Korelační koeficient r a výsledky Wilcoxonova testu W (jednička značí zamítnutí hypotézy o rovnosti mediánů v sousedících třídách pro hladinu významnosti $\alpha = 0,01$)

3.4. Souhrn parametrů

V předchozích odstavcích jsme definovali několik jednoduchých parametrů. Nyní pro každý signál dáme tyto parametry „dohromady“ a signál budeme reprezentovat jediným číslem.

Nejdříve jednoznačně stanovme uvažované charakteristiky:

- poměr ticho/řeč;
- počet úseků (ticho i řeč) za minutu;
- obsazení úrovní intenzity, konkrétně hodnoty pro příhrádky kolem $-0,06$ a $0,03$;

- průměrný počet náběžných hran energetické obálky za jednu sekundu s prahem $p = 0,1$.

Výčet představuje pět čísel pro každý signál. Tato čísla transformujeme tak, aby jejich vysoká hodnota byla ukazatelem patologie. Tak je tomu např. u poměru ticho řeč, u obsazení úrovní intenzity je tomu naopak. Dále provedeme normalizaci do rozpětí od nuly do jedné. Poté velmi snadno vypočteme jakousi průměrnou známku, danou průměrem transformovaných a normovaných hodnot. Z tab. 2 je zjevné, že výsledný souhrnný parametr předstihuje všechny jednotlivé charakteristiky, korelační koeficient (pro neplynulost posouzenou prvním hodnotitelem) je $0,73$, ve výstupu Wilcoxonova testu nacházíme tři jedničky. Pro srovnání – z dosud uvedených parametrů byl nejlepší poměr ticho/řeč s korelačním koeficientem $0,69$ a dvěma jedničkami (rovněž pro prvního hodnotitele).

3.5. Parametry založené na detekci spektrálních změn

Výše popsané parametry mají jistá omezení. Mezi časté projevy koktavosti patří prolongace (velmi zjednodušeně řečeno se jedná o nepřirozeně prodloužené hlásky). Jednoduché parametry založené na VAD nemohou na prolongace poukázat, poněvadž je vyhodnotí (správně) jako řeč. Poměr ticho – řeč pak může být naopak nižší (což – jak víme – je ukazatelem plynulosti řeči). Pro prolongace je typická ustálenost ve spektrální oblasti. K jejich odhalení se tedy nabízí použít detektor spektrálních změn, např. bayesovský detektor popsáný v [5]. Jeho výstupem je křivka, jejíž lokální maxima označují okamžiky, kdy dochází ke změnám ve spektru signálu (např. hranice mezi hláskami). Průměrný počet těchto maxim za minutu se jeví jako velmi vhodný parametr pro popis neplynulosti řeči. Korelační koeficient je vysoký, $-0,77$ pro prvního hodnotitele a $-0,76$ pro druhého (viz tab. 2).

Každá dvě po sobě jdoucí maxima ve výstupu detektoru vymezují interval délky $I(j)$ (interval mezi dvěma spektrálními změnami). Na tomto základě vypočteme průměr jedenácti po sobě jdoucích intervalů $I(j), \dots, I(j+10)$. Poté se posuneme o interval „vpřed“, spočteme průměr z $I(j+1), \dots, I(j+11)$. Takto postupujeme pro celý signál, čímž obdržíme sadu průměrných hodnot. Směrodatná odchylka spočtená nad touto sadou hodnot představuje další parametr pro popis neplynulosti. Přestože korelační koeficient není vysoký (cca $0,4$), výstup Wilcoxonova testu je uspokojivý (viz tab. 2), zejména je příznivá jednička mezi nultou a první kategorií u prvního hodnotitele.

Pro popsání postupu samozřejmě můžeme využít i jiné spektrální detektory, např. GLR (General Likelihood Ratio, viz [6]) vzdálenost. Výsledky jsou velmi podobné, viz tab. 2.

Jak bylo výše řečeno, maximum ve výstupu detektorů odpovídá změně ve spektru signálu, např. hranici hlásek. Je tedy nasnadě pro analýzu promluv použít rozpoznávač řeči (nástroj HTK), jehož výstupem je odhad vyslovených hlásek a jejich pozice v signálu. Na tomto základě

bylo zkoumáno mnoho parametrů (délka skupin hlásek, jejich celkový počet, počet velmi dlouhých hlásek apod.). Nicméně jako nejužitečnější se zdá být průměrný počet všech hlásek za minutu a směrodatná odchylka průměru délek po sobě jdoucích hlásek, tedy parametry velmi podobné těm založeným na měření odlišnosti ve spektru. Tab. 2 ukazuje zatím nejvyšší korelační koeficient (−0,8 pro prvního hodnotitele), výstup Wilcoxonova testu je identický jako pro GLR detektor.

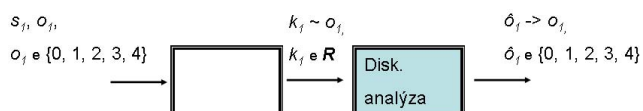
3.6. Parametry zkoumající pravidelnost promluvy

Pro neplynulé promluvy jsou typické změny tempa řeči, intervaly s běžnou rychlostí řeči se střídají s intervaly s velmi pomalou rychlostí. Tuto (ne)pravidelnost můžeme zkoumat na „nepravidelném výdeji“ energie promluvy. Použijeme následující postup: 1) Vzorky signálu umocníme na druhou. 2) Tyto hodnoty postupně přidáváme do akumulátoru. 3) Pokud je hodnota akumulátoru menší než práh, pak opakujeme krok 2, v opačném případě (přáh byl překročen) akumulátor vynulujeme a uložíme okamžik i (číslo vzorku), kdy došlo k překročení prahu. Pro úplnost dodáme, že práh se pro každého jedince určuje adaptivně na základě maxim energetické obálky signálu. Výsledkem celého postupu je sada indexů i , udávající místa překročení prahu. Pro každého mluvčího je možné zkoumat jejich průměrný počet, střední vzdálenost apod. Jako nejlepší se jeví směrodatná odchylka z délek intervalů vymezených okamžiky i .

Obdobně můžeme zkoumat pravidelnost znělosti promluvy. Opět použijeme akumulátor, jehož hodnotu inkrementujeme za každou pitch periodu signálu. Práh je tentokrát odvozen ze střední hodnoty F_0 mluvčího. Zajímavě se opět jeví směrodatná odchylka z délek intervalů vymezených okamžiky překročení prahu. Výsledky (pro energii i znělost) najdeme na posledních dvou řádcích tabulky 2.

4. Diskretizace parametrů

Dosud jsme diskutovali korelaci parametrů s hodnocením lékařů, přičemž výstupní hodnoty analýz patřily do oboru reálných čísel. Pro přímé porovnání algoritmu s lékaři je potřeba jeho výstup diskretizovat do hodnot 0, 1, 2, 3 a 4. To lze udělat mnoha způsoby, my jsme zvolili diskriminační analýzu (viz [7]). Principiálně tak do schématu analýzy přibude jeden krok, viz obr. 2.



Obrázek 2: Princip diskretizace s použitím diskriminační analýzy

Pro jednotlivé parametry je nejdříve nutné natrénovat v každé třídě střední hodnotu μ_k a rozptyl σ_k . K tomu po-

užijeme soubor 1, přičemž pro zařazení do třídy použijeme soud prvního hodnotitele. Po natrénování je možné klasifikovat libovolnou promluvu. To provedeme zpětně jak pro všech 121 mluvčích v souboru 1 (tedy uzavřený test), tak pro 33 jedinců v souboru 2. Úspěšnost klasifikace budeme posuzovat na základě odchylky:

$$\Delta = \sum_{j \in S} \min(|o_j - \hat{o}_j|), \quad (2)$$

kde o_j je hodnocení lékaře pro j -tou promluvu, $\hat{o}_j$ je odhad tíže neplynulosti, S značí soubor (1 nebo 2). Funkci $\min$ je potřeba chápat spíše symbolicky, značí, že do Δ zahrnujeme vždy menší odchylku od obou lékařů. Pokud např. první lékař hodnotil promluvu známkou 0, druhý známkou 1 a „náš“ odhad je 2, pak se Δ zvýší pouze o jedničku. Pro většinu algoritmů je odchylka v obou souborech uvedena v tabulce 3. Výsledky pro nejjednodušší akustické míry (jako poměr ticho – řeč) nejsou uvedeny pro soubor 2, zastupuje je pouze souhrnný parametr (šestý řádek tabulky).

Parametr	Δ_{s_1}	Δ_{s_2}
Poměr ticho/řeč	73	–
Počet úseků	126	–
Úrovně −0,06	104	–
Úrovně +0,03	101	–
Náb. hrany ener.	105	–
Souhrnný param.	55	28
Počet max. Bay. <i>std</i> intervalu	62 70	25 22
Počet max. GLR <i>std</i> intervalu	57 73	26 20
Počet hl. HTK <i>std</i> intervalu	63 55	21 21
<i>std</i> energie	72	27
<i>std</i> znělost	69	23
PRŮMĚR „21“	51	14

Tabulka 3: Odchylka Δ_{s_1} pro soubor 1 a Δ_{s_2} pro soubor 2

Právě tento parametr dosáhl velmi dobrého výsledku v souboru 1, celková odchylka $\Delta = 55$ představuje ve skupině 121 mluvčích průměrnou odchylku $\delta = 0,45$ na jedince. O něco horší je výsledek v souboru 2, celková $\Delta = 28$ odpovídá relativní $\delta = 0,85$. Výsledky složitějších měř (bayesovský detektor, GLR, HTK) jsou v souboru 1 podobné (δ je přibližně 0,5), nicméně v souboru 2 jsou o poznání lepší. Např. pro parametr „*std* intervalu GLR“ je δ pouze 0,6. Tato rozdílnost je pravděpodobně způsobena citlivostí jednoduchých měř (ze kterých se souhrnný parametr skládá) na horší technickou kvalitu dat v druhém souboru.

4.1. Kombinace parametrů

Množství popsanych měř vybízí k jejich kombinaci. Např. parametr daný průměrným počtem maxim bayesovského

detektoru se v souboru 1 mýlí pro 56 jedinců. Z těchto 56 je druhý bayesovský parametr (v tabulkách uváděný jako „*std* intervalu“) schopen správně ohodnotit 22. Asi nej-jednodušší způsob kombinace je určit aritmetický průměr z více měr a ten poté zaokrouhlit. Výsledek (hodnoty Δ) nalezneme v posledním řádku tab. 3. Odchylka v souboru 1 klesla na 51 (dosud nejlepší výsledek byl $\Delta = 55$ pro souhrnný parametr a *std* intervalu HTK). Ještě výraznější je přínos v souboru 2, kde Δ poklesla na 14, přičemž Δ pro žádnou z jednotlivých měr neklesla pod 20. Pro úplnost dodejme, že průměr byl počítán z celkem 21 měr (proto je v tabulce označen PRŮMĚR „21“), jedná se o míry popsané v tomto článku s různým nastavením. Např. pro detektory spektrálních změn byly uvažovány různé délky okna (od 30 do 120 ms), řády autoregresního modelu (od 2 do 8) apod.

Poněkud detailnější výsledky pro aritmetický průměr parametrů uvádí tabulka 4, konkrétně se jedná o počty jedinců s danou odchylkou. Např. tedy první řádek (odchylka 0) udává počet jedinců, pro které je odhad totožný s hodnocením lékařů – konkrétně 60 % v souboru 1 a 67 % v souboru 2. Důležité jsou velmi malé hodnoty pro odchylku 2 (2 jedinci v souboru 1 a 3 v souboru 2).

Odchylka	Soubor 1	Soubor 2
0	72 (60 %)	22 (67 %)
1	47 (39 %)	8 (24 %)
2	2 (1 %)	3 (9 %)

Tabulka 4: Počty jedinců s danou odchylkou pro oba soubory

5. Závěr

Seznámili jsme se s několika parametry pro objektivní měření plynulosti promluvy. K základním patří ty založené na VAD (jako poměr ticho – řeč, počet úseků), dále ty vyšetřující rovnoměrnost využití hladin intenzity signálu a zkoumající četnost hran energetické obálky. Velmi dobře se jeví souhrnný parametr vypočtený na jejich základě – dosahuje korelačního koeficientu až 0,73.

Ještě lepších výsledků dosahují parametry založené na detekci spektrálních změn – průměrný počet maxim ve výstupu bayesovského detektoru má korelační koeficient $-0,77$. Průměrný počet hlásek v rozpoznávací HTK pak dokonce $-0,80$.

Kromě průměrných hodnot se také vyplácí sledovat směrodatné odchylky určitých charakteristik. Např. směrodatná odchylka z intervalů vymezených maximy bayesovského detektoru dosáhla vynikajících výsledků ve Wilcoxonově testu, konkrétně se mediány liší mezi všemi třídami kromě třetí a čtvrté.

Pro diskretizaci výstupu algoritmů byla zvolena diskriminační analýza umožňující výpočet odchylek v odhadu neplynulosti. Za pozornost stojí výsledky obdržené spojením všech měr aritmetickým průměrem. Jeho výstup

správně ohodnotil 60 % jedinců v souboru 1 a 67 % jedinců v souboru 2. Příznivé je, že pouze u dvou jedinců v prvním souboru a tří v druhém byla odchylka v hodnocení 2. Odchylka 3 a víc nebyla vůbec pozorována. Pro srovnání: lékaři se v souboru 1 liší o 1 u 22 pacientů a o 2 u jednoho pacienta.

Modifikace	Počet max.		<i>std</i> inter.	
	Δ	δ	Δ	δ
Vliv způsobu trénování:				
Původní	62,0	0,51	70,0	0,58
„1 proti všem“	64,0	0,53	70,0	0,58
Vliv výpočtu odchylky (při strategii „1 proti všem“):				
Průměr hodnocení	76,0	0,63	80,0	0,66
Průměr odchylek	75,5	0,62	87,0	0,72
Omezený soubor	54,0	0,55	64,0	0,65

Tabulka 5: Vliv trénovací strategie a způsobu výpočtu chyby na celkovou odchylku Δ a průměrnou δ

Při hodnocení úspěšnosti celého postupu je však třeba brát zřetel na několik skutečností. Tou první je, že rozhodující jsou výsledky na souboru 1, který obsahuje známky dobré technické kvality. Pro něj prezentované výsledky jsou výstupem uzavřeného testu, nabízí se možnost použít strategii „1 proti všem“, tedy postup, kdy 120 jedinců slouží k trénování diskriminační analýzy a jeden jedinec jako testovací, přičemž se tento postup zopakuje pro všechny členy souboru 1. Celkovou odchylku Δ a průměrnou odchylku δ při použití této strategie uvádí tab. 5. Konkrétně se jedná o výsledky dvou měr založených na bayesovském detektoru (viz kap. 3.5). V levé části nalezneme výsledky pro průměrný počet maxim (označeno „Počet max.“), v pravé pak pro směrodatnou odchylku z intervalů vymezených maximy detektoru (označeno „*std* inter.“). První řádek tabulky (označen „Původní“) ukazuje odchylku při použití uzavřeného testu (tedy stejnou hodnotu jako v tab. 3), druhý řádek pro strategii „1 proti všem“. Je zřejmé, že došlo k mírnému zhoršení výsledků pro první míru (Δ vzrostla o 2), výsledky pro druhou míru zůstaly beze změny. Připomeňme, že pro trénování byl v obou případech použit soud prvního hodnotitele.

Další podstatnou skutečností je způsob výpočtu odchylky Δ . V předešlých odstavcích byla do Δ zahrnována vždy menší odchylka od obou lékařů (pokud např. první lékař hodnotil promluvu známkou 0, druhý známkou 1 a odhad je 2, pak se Δ zvýšila pouze o jedničku). Tento postup však může vést k jistému nadhodnocení úspěšnosti, metodicky správnější je odchylku počítat na základě průměru hodnocení lékařů. V tom případě je však dobré si uvědomit, že pro 22 jedinců v souboru 1 je průměrná známka necelé číslo (0,5, 1,5 apod.), tedy číslo, které nikdy nemůže být výsledkem popsáných algoritmů (výsledkem diskriminační analýzy je vždy celé číslo: 0, 1 apod.). Δ vypočtená na základě průměru pak pro všechny míry obsahuje jakýsi offset o hodnotě $22 \cdot 0,5 = 11$. Pro dvě míry je odchylka uvedena v tab. 5 (řádek označený „Průměr hodnocení“),

změnu Δ pak posoudíme porovnáním s řádkem označeným „1 proti všem“. Pro první míru vzrostla o 12, pro druhou o 10. Po odečtení výše zmíněného offsetu je tak zjevné, že výsledky se téměř nezměnily.

Samozřejmě existuje několik možných modifikací trénování a výpočtu odchylky. Dosud jsme vždy parametry diskriminační analýzy trénovali na základě soudu prvního hodnotitele (důvodem byly vyšší hodnoty korelačního koeficientu pro většinu měr). Postup je možné upravit tak, že trénování proběhne pro prvního hodnotitele a odchylka (při strategii „1 proti všem“) se vypočte na základě jeho soudu, tedy ne na základě průměru. Stejný postup se zopakuje pro druhého hodnotitele. Výsledkem jsou dvě odchylky, z nichž vypočteme průměr. Pro dvě bayesovské míry je tento průměr uveden v tab. 5 (řádek označený „Průměr odchylek“). Pro první z nich je výsledek příliš neliší od postupu v minulém odstavci (řádek „Průměr hodnocení“). Pro druhou míru je však rozdíl značný, Δ vzrostla o 7. Důvodem je vysoká partikulární odchylka pro druhého hodnotitele, která v tomto případě činila 94.

Další modifikace je velmi jednoduchá, hodnotitelé se neshodli pro 23 jedinců v souboru 1, po jejich vyloučení zbude 98 mluvčích s jednoznačnou známkou. Úvahy v předchozích odstavcích pak pozbývají smyslu, trénování i výpočet Δ je jednoznačný. Pro úplnost je odchylka vypočtená nad tímto omezeným souborem uvedena na posledním řádku tab. 5. Zdůrazněme, že vzhledem k nižšímu počtu účastníků je nutné při porovnávání s předešlými postupy hledět na průměrnou odchylku δ . Její hodnoty jsou nižší než při použití průměru hodnotitelů, ale vyšší než při použití „minima“.

Kromě aplikace výše popsaných úvah pro další míry by naše budoucí práce měla být zaměřena na kombinaci parametrů pomocí klasifikátoru SVM, analýzu robustnosti jednotlivých měr a na zkoumání vlivu zpětné vazby (Delayed Auditory Feedback – DAF) na plynulost řeči jedince. Soubor 1 (nebo jeho část) by měl být znovu posouzen oběma hodnotiteli, což by umožnilo vyhodnocení intraindividuální variability lékařů.

Poděkování

Tato práce byla podporována z výzkumného záměru „Transdisciplinární výzkum v oblasti biomedicínského inženýrství II“ MSM6840770012, z grantu „Analýza a modelování biomedicínských a řečových signálů“ GACR 102/08/H008 a „Rozpoznávání mluvené řeči v reálných podmínkách“ GACR 102/08/0707.

Reference

- [1] M. Laštovka, J. Vokřál, L. Černý, K. Radilová, M. Hrdličková, *Hodnocení tíže poruchy plynulosti řeči pomocí neuronových sítí*, Závěrečná zpráva grantu 237/1998/C/1. LF.
- [2] P. Howell, A. Hamilton, A. Kyriacopoulos, *Automatic detection of repetitions and prolongations in stuttered speech*, Speech Input/Output: Techniques and Applications, IEE Publications, 252–256, 1986.
- [3] P. Bergl, R. Čmejla, L. Černý, M. Hrbková, *Objective and Subjective Evaluation of Dysfluent Speech*. In Digital Technologies 2008 [CD-ROM]. Žilina: University of Žilina, Faculty of electrical engineering, 2008, vol. 1, ISBN 978-80-8070-953-2.
- [4] P. Bergl, *Akusticko-fonetické charakteristiky neplných promluv*. In Analýza a zpracování řečových a biologických signálů – Sborník prací 2007. Praha: ČVUT, 2007, díl 1, s. 7–12. ISBN 80-01-03940-3.
- [5] R. Čmejla, P. Sovka, *Audio Signal Segmentation Using Recursive Bayesian Change-point Detectors*. In Proceedings of the WSEAS International Conferences [CD-ROM]. New York: WSEAS Press, 2004, vol. 1, s. 1087–1091.
- [6] U. Appel, A. V. Brandt, *Adaptive Segmentation of Piecewise Stationary Time Series*, Information Sciences, 29:27–56, 1983.
- [7] R. Čmejla, P. Sovka, *Úvod do bayesovské klasifikace dat*, Akustické listy. 2002, roč. 8, č. 2, s. 3–10. ISSN 1212-4702.

