



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

**ODHAD PROGRESE PARKINSONOVY NEMOCI POMOCÍ
AKUSTICKÉ ANALÝZY ŘEČI**

DEGREE OF PARKINSON'S DISEASE ESTIMATION BASED ON ACOUSTIC ANALYSIS OF SPEECH

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. Iveta Ustohalová

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Zoltán Galáž

BRNO 2016



Diplomová práce

magisterský navazující studijní obor **Telekomunikační a informační technika**

Ústav telekomunikací

Studentka: Bc. Iveta Ustohalová

ID: 146985

Ročník: 2

Akademický rok: 2015/16

NÁZEV TÉMATU:

Odhad progresu Parkinsonovy nemoci pomocí akustické analýzy řeči

POKYNY PRO VYPRACOVÁNÍ:

V současné době je klinické vyšetření pacienta s Parkinsonovou nemocí prováděno subjektivní kvantifikací aspektů doprovázejících toto onemocnění. Neurolog, klinický logoped a psycholog pacienta vyšetří a pomocí normovaných škál diagnostikují, hodnotí míru onemocnění nebo monitorují jeho progresi. Tento přístup má mnoho nevýhod, např. subjektivita, časová a finanční náročnost atd. Cílem této práce je implementace moderní neinvazivní metody analýzy Parkinsonovi s využitím analýzy řeči. V práci bude parametrizováno řečové cvičení fonace samohlásky /a/. Následně bude provedena analýza vypočtených parametrů se zaměřením na odhad progresu Parkinsonovi nemoci charakterizované vícero klinickými škálami. Uvedené postupy budou naprogramovány v jazycích MATLAB a C/C++.

DOPORUČENÁ LITERATURA:

[1] TSANAS, A.; LITTLE, M.; MCSHARRY, P.; RAMIG, L.: Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity, J R Soc Interface 8 (2010): s. 842-855.

[2] TSANAS, A.; LITTLE, M.; MCSHARRY, P.; RAMIG, L.: Accurate Telemonitoring of Parkinson's Disease Progression by Noninvasive Speech Tests. IEEE T Bio-Med Eng, roč. 57, č. 4, 2010: s. 884–893..

Termín zadání: 1.2.2016

Termín odevzdání: 25.5.2016

Vedoucí práce: Ing. Zoltán Galáž

Konzultant diplomové práce:

doc. Ing. Jiří Mišurec, CSc., předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

Výzkum popsany v této diplomové práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následku porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Tato diplomová práce se zabývá odhadem progresu Parkinsonovy nemoci pomocí akustické analýzy řeči. V jejím rámci byla prostudována problematika hypokinetické dysartrie jako důsledku Parkinsonovy nemoci a její vliv na mluvený projev. Dále byly podrobně prostudovány metody předzpracování řečového signálu a řečové parametry vhodné pro klasifikaci patologické řeči. Následně byl v prostředí MATLAB vytvořen systém pro odhad hodnoty UPDRS III škály, což je hodnota subjektivně určená lékařem, pomocí jednoho individuálního příznaku a poté byla pomocí algoritmu SFFS vybrána nejlepší kombinace příznaků, která odhaduje hodnotu UPDRS III s nejmenší chybou. Pozornost je taktéž věnována korelaci jednotlivých příznaků s UPDRS III škálou.

KLÍČOVÁ SLOVA

Neurologická onemocnění, Parkinsonova choroba, hypokinetická dysartrie, segmentace, příznaky, klasifikace, řečový parametr, řečové patologie, UPDRS III škála.

ABSTRACT

The diploma thesis deals with the non-invasive analysis of progression of Parkinson's disease using the acoustic analysis of speech. Hypokinetic dysarthria in connection with Parkinson's disease as well as speech parameters are described in this work. Speech parameters are sorted according to the speech component they affect. The work uses the phonation of vowels "a" speech task as the most commonly used speech task in the field of pathological speech processing, because of its resistance to demographic and linguistic characteristics of the speakers. Based on obtained knowledge, in MATLAB development environment were created system for UPDRS III scale estimation. The UPDRS III scale is based on subjective diagnosis given by the doctor. At first, one individual parameter is used for the UPDRS III scale value estimation. Then the feature selection using SFFS algorithm is applied to gain feature combination with minimal estimation error. Attention is also paid to correlation between individual symptoms and UPDRS III scale.

KEYWORDS

Neural disorders, Parkinson's Disease, hypokinetic dysarthria, segmentation, features, classification, speech parameter, speech pathology, UPDRS III range.

USTOHALOVÁ, I. *Odhad progresu Parkinsonovy nemoci pomocí akustické analýzy řeči*. Brno: Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií, 2016. 60 s. Vedoucí diplomové práce Ing. Zoltán Galáž.

PROHLÁŠENÍ

Prohlašuji, že svoji diplomovou práci na téma „Odhad progresu Parkinsonovy nemoci pomocí akustické analýzy řeči“ jsem vypracovala samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujícího autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne

.....

(podpis autora)

PODĚKOVÁNÍ

Ráda bych poděkovala vedoucímu mé diplomové práce, Ing. Zoltánu Galážovi, za ochotu a trpělivost s jakou vedl tuto práci a taktéž za pomoc a poskytnutí materiálů potřebných pro její realizaci. Dále bych pak chtěla poděkovat Ing. Štefanu Mišíkovi za podporu a cenné rady. Speciální poděkování patří mé rodině za neochvějnou víru v mé schopnosti.

V Brně dne

.....

(podpis autora)

OBSAH

Obsah	x
Seznam obrázků	xiii
Seznam tabulek	xiv
Úvod	15
1 Základní principy produkce řeči	16
1.1 Řečové ústrojí	16
1.1.1 Dechové ústrojí	17
1.1.2 Hlasové ústrojí	18
1.1.3 Artikulační ústrojí	18
2 Parkinsonova nemoc a Hypokinetická dysartrie	20
2.1 Patologická řeč	20
2.2 Parkinsonova nemoc	20
2.3 Hypokinetická dysartrie	21
2.3.1 Fonace	21
2.3.2 Artikulace	22
2.3.3 Prozodie	22
2.3.4 Plynulost řeči	22
2.3.5 UPDRS škála	22
3 Předzpracování signálu	23
3.1 Preemfázový číslicový filtr	23
3.2 Odstranění stejnosměrné složky	23
3.3 Segmentace	23
4 Řečové příznaky	25
4.1 Příznaky popisující fonaci	25
4.1.1 Základní frekvence	25
4.1.2 Jitter	26
4.1.3 PPE (Pitch Period Entropy)	26
4.1.4 Schimmer	26
4.1.5 Krátkodobá energie	27

4.1.6	LSTER (Low short-time energy ratio).....	27
4.1.7	TKEO (Teager-Kaiser Energy Operator)	27
4.1.8	MPSD (Median of Power Spectral Density).....	27
4.2	Příznaky popisující rezonanci (formanty).....	29
4.3	Příznaky založené na hodnocení kvality signálu	29
4.3.1	ZCR (zero-crossing rate)	29
4.3.2	HZCRR (high zero-crossing rate ratio)	30
4.3.3	HNR (harmonic-to-noise ratio).....	30
4.4	Extrakce vysokoúrovňových příznaků.....	30
5	Analýza řečových příznaků	31
5.1	Korelace	31
5.2	Strojové učení	32
5.2.1	Rozhodovací stromy	33
5.2.2	Cross-validace.....	33
5.3	Výběr příznaků	34
5.3.1	Metody výběru	34
6	Návrh systému	36
6.1	Vývojový diagram	36
6.2	Popis systému	37
7	Praktická část	38
7.1	Předzpracování vstupních dat	38
7.2	Předzpracování signálu	39
7.3	Parametrizace	39
7.3.1	Příznaky popisující fonaci	40
7.3.2	Příznaky popisující rezonanci (formanty).....	40
7.3.3	Příznaky založené na hodnocení kvality signálu	41
7.4	Korelace příznaků s UPDRS III.....	41
7.5	Regresní analýza	44
7.5.1	Individuální analýza.....	45
7.5.2	Selekce příznaků	47
8	Závěr	51
	Literatura	53

Seznam symbolů, veličin a zkratk	56
SEZNAM PŘÍLOH	58
A OBSAH PŘiloženého cd	59
B SEZNAM PARAMETRŮ POUŽITÝCH PRO REGRESNÍ ANALÝZU	60

SEZNAM OBRÁZKŮ

Obr. 1.1: Schématické zobrazení hlasového traktu člověka [21]	17
Obr. 3.1: Časový průběh (a) pravoúhlého okna a Hammingova okna a jejich modul spektra (b) [21].....	24
Obr. 5.1: Míra lineární závislosti dvou proměnných vyjádřená korelačním koeficientem r	32
Obr. 5.2: Schéma metody filter pro výběr příznaků	35
Obr. 5.3: Schéma metody SFFS pro výběr příznaků	35
Obr. 7.1: Graf znázorňující míru korelace příznaku min STE s UPDRS III.	43
Obr. 7.2: Graf znázorňující míru korelace příznaku min ZCR s UPDRS III.	44
Obr. 7.3: Vstupní nastavení použité při inicializaci funkce SFFS_reg.m.....	48
Obr. 7.4: Zobrazení výpisu funkce SFFS_reg.m při výpočtu nejvhodnější kombinace příznaků.....	48

SEZNAM TABULEK

Tab. 4.1: Hellwagův trojúhelník (klasifikace vokálů z hlediska polohy jazyka)	29
Tab. 7.1: Statistické parametry databáze PARCZ	38
Tab. 7.2: Seznam deseti příznaků s nejvyšší mírou korelace řazených podle hodnoty dosažené hladiny významnosti u Spearmanovy korelace.	42
Tab. 7.3: Seznam deseti příznaků s nejvyšší mírou korelace řazených podle hodnoty dosažené hladiny významnosti u Pearsonovy korelace.	42
Tab. 7.4: Seznam deseti příznaků, které při individuální analýze dosahují nejlepších výsledků odhadu UPDRS III škály.	46
Tab. 7.5: Příznaky vybrané metodou SFFS, výsledky regrese a korelace při použití těchto příznaků.....	50

ÚVOD

Jedním z nejzákladnějších dorozumívacích prostředků člověka je řeč. Umožňuje vyjadřovat nejen myšlenky, ale lze do ní promítnout i emoce či momentální náladu. Díky tomu si zachovává svoji nepostradatelnost i v dnešním světě moderních technologií. Je-li schopnost řeči narušena, dochází ke značnému snížení kvality života postižené osoby. Poruchy řeči neboli řečové patologie jsou důsledkem nesprávného fungování řečového ústrojí nebo některé jeho části. Příčiny těchto disfunkcí mohou být různé, počínaje vrozenými vadami, následky úrazu či pokročilým věkem řečníka, přes nemoci jiného než neurologického původu, jako jsou například nádory až po neurologická onemocnění. Mezi neurodegenerativní onemocnění patří také Parkinsonova choroba, jejímž vlivem na řeč se zabývá tato práce. Kromě projevů, které se manifestují na pohybovém aparátu postižené osoby (typický je například třes), existují i symptomy, které lze sledovat v mluveném projevu. Souhrnně jsou nazývány hypokinetická dysartrie. Při ní jsou v různé míře narušeny dílčí složky řeči, tj. artikulace, prozodie, plynulosti řeči, fonorespirace či faciokineze. V důsledku těchto poruch se řeč pacientů s Parkinsonovou nemocí stává méně srozumitelnou. Hlas je tichý, nevýrazný, může znít drsně či chraplavě. Typická je také jeho monotónnost a absence emocí. Pro pacienty je taktéž náročné hlas zeslabit nebo zesílit a velmi rychle se unaví. Často se vyskytují poruchy dýchání a zvýšená nazalita řeči. V současnosti je diagnostika Parkinsonovy nemoci a míra její progresese prováděna subjektivní kvantifikací aspektů doprovázejících toto onemocnění. Nevýhodou tohoto přístupu je subjektivita, kdy každý vyšetřující lékař vnímá pacientův projev mírně odlišně a také časová náročnost takového vyšetření. Je proto snaha nalézt a implementovat neinvazivní metody a postupy využívající analýzu řeči. Pro účely této práce bude použito řečové cvičení prodloužené fonace hlásky [a]. Jedná se o nejběžněji se vyskytující řečové cvičení, jehož velkou výhodou je nezávislost na demografických a lingvistických charakteristikách jednotlivých řečníků. Řečové cvičení bude parametrizováno a progres Parkinsonovy nemoci odhadnutá na základě těchto příznaků bude porovnána s klinickými daty pacientů, konkrétně s hodnotou UPDRS III, určenou subjektivně lékařem. Cílem práce je nalézt takovou sadu parametrů, která bude poskytovat co nejpřesnější odhad progres Parkinsonovy nemoci.

V rámci 1 kapitoly této práce je obecně popsána problematika zpracování řečového signálu. Kapitola 2 se zabývá Parkinsonovou nemocí, hypokinetickou dysartrií a jejími projevy v oblasti fonace, artikulace, prozodie a plynulosti řeči. V kapitole 3 je rozebráno předzpracování řečového signálu. Kapitola 4 je pak již zaměřena na Parametrizaci řeči. Jsou zde popsány jednotlivé sledované parametry v závislosti na řečových součástech, na které mají vliv. V kapitole 6 je vyobrazeno a popsáno schéma návrhu systému pro analýzu řečových dat. Kapitola 7 je zaměřena na popis praktické části této práce. Závěrečná 8 kapitola je věnována shrnutí dosažených výsledků práce.

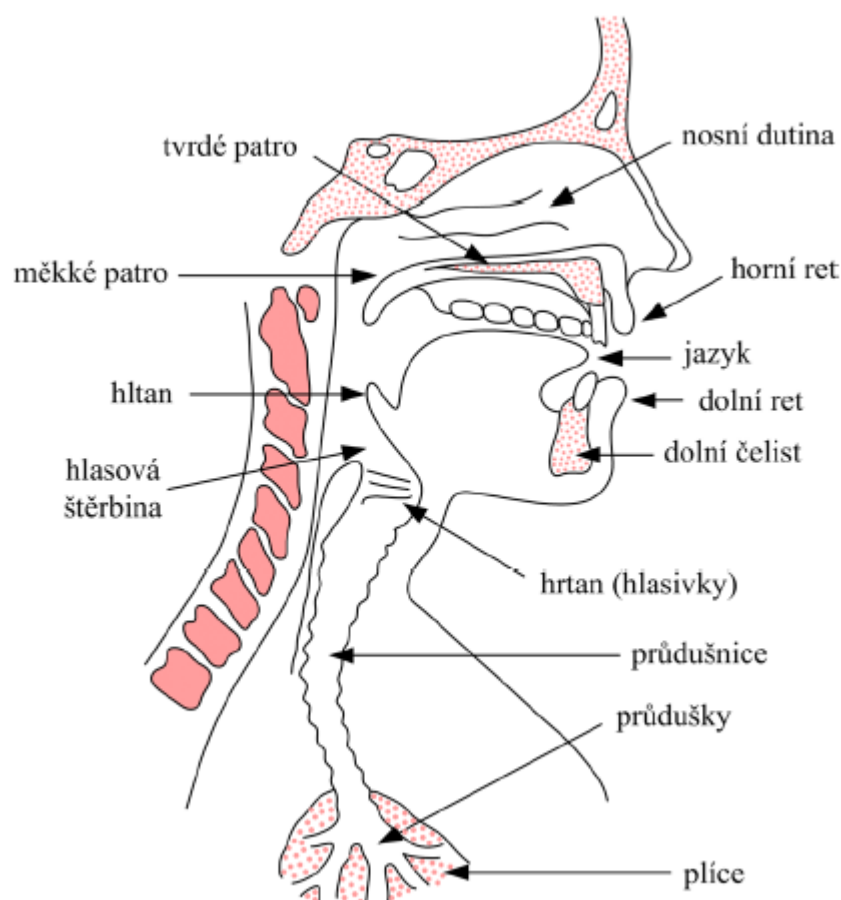
1 ZÁKLADNÍ PRINCIPY PRODUKCE ŘEČI

Řeč je jedinečným prostředkem pro komunikaci, který je vlastní pouze člověku. Vzájemnou komunikaci dvou nebo více lidí můžeme rozdělit do několika úrovní: jazykové, fyziologické a akustické. Mluvčí (ten který mluví) musí nejprve uspořádat informace do logického celku - myšlenky, jinými slovy se musí rozhodnout, co přesně chce sdělit. Dalším krokem je převod tohoto logického celku do jazykové formy. Mluvčí zvolí vhodná slova a fráze a podle gramatických pravidel daného jazyka je srovná do správného pořadí – věty, souvětí. Následuje převod na nervové impulsy, jež jsou přenášeny motorickými nervy až ke svalům mluvních orgánů (hlasivky, jazyk, čelist, rty apod.). Celý tento proces probíhá v mozku mluvčího. Na základě nervových impulzů se začnou pohybovat svaly a tento pohyb vede ke změnám akustického tlaku okolního vzduchu. Vzniká akustická vlna, jejímž prostřednictvím se informace šíří a také aktivuje sluchové ústrojí posluchače i řečníka. Nervové impulzy, vznikající transformací akustických podnětů ve sluchovém aparátu, jsou následně vedeny smyslovými nervy do mozku. Posluchač tak rozumí tomu, co chtěl mluvčí říci. Sluchové ústrojí mluvčího slouží mozku jako zpětná vazba, kdy dochází ke srovnávání kvality produkováných zvuků s tím, co chtěl mluvčí sdělit. Mozek tak může okamžitě reagovat a upravovat případné nesrovnalosti. Nejedná se pouze o význam slov, ale i jejich hlasitost nebo zabarvení hlasu vyjadřující emoce.

Pro přesný sluchový vjem je krom jiného důležitá artikulace neboli koordinovaná činnost mluvidel vedoucí k vytvoření zvuků. Každý jazyk má svůj charakteristický soubor artikulačních pohybů.

1.1 Řečové ústrojí

Na produkci řeči se podílí velké množství orgánů lidského těla. Mozek, který slouží jako řídicí jednotka celého systému je jen jednou z částí. Primární funkce těchto orgánů nemusí mít s řečí nic společného, avšak na jejím vzniku se významně podílí. Jednotlivé orgány hlasového traktu jsou znázorněny na .



Obr. 1.1: Schématické zobrazení hlasového traktu člověka [21]

1.1.1 Dechové ústrojí

Hlavním dýchacím svaelem je bránice. Odděluje hrudní dutinu od břišní a v důsledku jejího pohybu se mění objemu hrudního koše. Uplatňuje se především při nádechu. Při výdechu vzniká proud vzduchu, který vychází z plic a prochází průdušnicí přes hlasivky do nosní a ústní dutiny. Proud vzduchu dále pokračuje, přes jazyk, zuby a rty, ven ústy nebo nosními dírkami. Zatímco fyziologické dýchání je reflexní, při řeči je množství vzduchu i rytmus dýchání řízen z části vědomě. Poměr vdechu a výdechu je při klidném dýchání asi 2:3 (16krát – 20krát za minutu, na jeden nádech přibližně 0,5 litru vzduchu), při řeči se nádech výrazně zkracuje a výdech prodlužuje (1:7 až 1:12) a zvětšuje se objem vdechnutého vzduchu na přibližně 1,5 litru. Při řeči je nádech prováděn nosem i ústy. Mluvní projev je realizován při výdechu, který lze díky zásobnímu vzduchu v plicích podle potřeby prodloužit, a který zajišťuje trvalý tlak v hrtanu, což je důležité pro vznik hlasu. Nejvíce dechu spotřebují hlásky s nejvolnější ústní dutinou, tedy samohlásky (vokály, zvláště [á]), ze souhlásek [h], [f], [x] a sykavky.

1.1.2 Hlasové ústrojí

Základní hlasový tón vzniká činností hlasového (fonačního) ústrojí uloženého v hrtanu. V této fázi však ještě nemluvíme hlasu, ten vzniká až úpravami při průchodu hlasovým traktem. Základem hlasového ústrojí jsou párový hlasivkový sval, hlasivkový vaz a slizniční hlasivková řasa – souhrnně hlasivky. Vibrující hlasivky vytváří tlakové pulzy, které se střídavě šíří ústní nebo nosní dutinou. Právě při přerušování výdechového proudu vzduchu kmitajícími hlasivkami vzniká kvaziperiodický budící signál, který má tónový charakter. Vzniká-li budící signál vibracemi hlasivek, označujeme jej jako znělé buzení. Jsou jím tvořeny zvláště samohlásky a dále částečně některé další hlásky (znělé souhlásky). Kmitočet takového kvaziperiodického signálu se nazývá základní kmitočet F_0 (pitch frequency) a jemu odpovídající časový interval základní perioda (pitch period). Tento kvaziperiodický signál se nazývá základní tón (pitch). Druhou možností je buzení neznělé, označujeme tak stálý tok vzduchu šumového charakteru v úžinách hlasového traktu. Hlasivky při takovémto buzení nekmitají. Příkladem může být hláska [s].

Čím kratší jsou hlasivky, tím rychleji kmitají a základní tón je vyšší. Délka hlasivek u žen je přibližně 1,2 cm a u mužů 1,5 – 2,5 cm. Těmito anatomickými specifiky je pak dána obvyklá výška hlasu jedince. Stejně jako výška, závisí na anatomii i barva lidského hlasu. Hlas vycházející z hlasivek je dále modifikován při průchodu nadhrtanovými dutinami – rezonátory. Tyto dutiny jsou tři: hrdelní, nosní a ústní dutina. Dutina hrdelní se rozkládá bezprostředně nad hlasivkami. Její objem je ovlivňován pohybem kořene jazyka a činností svalů hrdla. Nosní dutina se jako rezonanční prostor uplatňuje jen u nosních hlásek (nazál). Při většině hlásek je průchod do nosní dutiny uzavřen měkkým patrem (velum) přitisknutým k zadní stěně dutiny. Dutina ústní je ve předu ohraničena rty a vzadu přechází do dutiny hrdelní. Na rozdíl od dutiny nosní je rezonanční prostor dutiny ústní využíván u každé hlásky. Objem dutiny se mění pohybem jazyka, rtů a čelistí. Aby lidská lebka nebyla příliš těžká, nachází se v ní velké množství dutin. Rezonance těchto dutin se také výrazně podílí na výsledné barvě hlasu, charakteristické pro daného jednotlivce. Úsek od hlasivek až o rty a nosní dírky nazýváme hlasový trakt.

1.1.3 Artikulační ústrojí

Posledním článkem v tvorbě řeči je artikulační ústrojí. Sestává z nadhrtanových dutin a artikulačních orgánů, jež jsou v těchto dutinách uloženy nebo je obklopují. Mezi nadhrtanové dutiny patří dutina hrdelní, ústní a nosní. Pomyslnou hranici mezi jednotlivými dutinami tvoří místo, kde se špička měkkého patra nejvíce přibližuje kořeni jazyka. Zatímco nadhrtanové dutiny se na tvorbě řeči podílejí pasivně, artikulační orgány, především jazyk, rty a měkké patro, ale i zuby, tvrdé patro nebo čelisti, se na tvorbě řeči podílejí aktivně. Svým pohybem mění velikost nadhrtanových dutin, čímž umožňují vytváření široké škály zvuků vlastních danému jazyku. Nejvýznamnějším artikulačním orgánem je jazyk, jehož pohybem je modifikován tvar

a objem celé ústní i hrdelní dutiny.

Při průchodu základního hlasového tónu nadhrtanovými dutinami dochází k rezonancím způsobujícím změny rozložení akustické energie v signálu. Ta se soustředí kolem určitých frekvencí, nazývaných formantové frekvence. Oblasti (frekvenční pásma), ve kterých dochází k nejvyšší koncentraci (největšímu zesílení) jsou označovány jako formanty. Značí se $F_1, F_2, \dots, F_n$, kde F_1 je formant s nejnižší frekvencí. Pokud vzduch proudí i nosními dutinami, je výsledná řeč ovlivňována také antirezonanční vlastností těchto dutin. Vznikají takzvané antiformanty, což jsou frekvenční oblasti, jejichž energie je potlačena. Výsledný zvuk je dán superponovanou hodnotou zvuku z ústní a nosní dutiny.

V závislosti na postavení artikulačních orgánů, které slouží jako překážka pro vydechovaný proud vzduchu, vznikají různé druhy šumu. Šum tvoří základ neznělé řeči, tedy souhlásek. Vzájemné postavení všech artikulačních orgánů se během řeči plynule mění a vzniká tak bezpočet konfigurací hlasového traktu, čím vznikají všechny možné hlásky. Hlasový trakt však nelze považovat za lineární systém. Závislost současného postavení hlasového traktu na postavení předcházejícím se nazývá koartikulace. Navenek se projevuje tím, že se akustické realizace stejného zvuku mění v závislosti na kontextu okolních zvuků [21][20].

2 PARKINSONOVA NEMOC A HPOKINETICKÁ DYSARTRIE

2.1 Patologická řeč

Patologická řeč je důsledkem špatné funkce řečového ústrojí nebo jeho částí. Zkoumána může být jak z hlediska poruch komunikačního projevu, tak z hlediska poruch funkce hlasu. Tyto poruchy mohou mít různý původ i rozdílné klinické příznaky, projevem je však vždy degradace kvality řečového projevu postižené osoby. Budeme-li hlas uvažovat jako synonymum pro vibrace hlasivek a tyto vibrace po průchodu oro-naso-laryngeálním traktem označovat jako řeč, je vhodné zavést pojem fonace označující kmitání hlasivek. Jako normofonické je označování kmitání hlasivek, které nejeví žádné známky anomálií. Jestliže hlasivky mluvčího nekmitají správně, mluvíme o takzvané dysfonii nebo v extrémním případě, kdy hlasivky nejsou schopny kmitat vůbec o afonii.

Titze klasifikoval patologie řeči z hlediska poruch do tříd: zánětlivá onemocnění (laryngitida, bronchitida, záškrt, atd.), systematické změny (dehydratace, vedlejší účinky medikace, hormonální změny, atd.), mechanické namáhání (polypy, vředy, granulomy, laryngocele, krvácení, atd.), povrchová podráždění (laryngitida, leukoplakie, gastroezofageální reflux, atd.), změny tkáně (rakovina hrtanu, keratóza, papilomy, cysty, atd.), neurologické a svalové změny (bilaterální a unilaterální paralýza hlasivek, spastická dysfonie, Parkinsonova nemoc, amyotrofická laterální skleróza, myotonická dystrofie, Huntingtonova choroba, myasthenia gravis, atd.) a abnormality svalstva (dysfonie, ventrikulární fonace, atd.) [9]. Tato práce se zaměřuje na poruchy funkce hlasu u pacientů postižených hypokinetickou dysartrií, která je důsledkem Parkinsonovy nemoci.

2.2 Parkinsonova nemoc

Parkinsonova nemoc (PN) je závažné idiopatické onemocnění, charakteristické progresivním úmrtím dopaminergních neuronů v oblasti substantia nigra – černé substance ve středním mozku. Poprvé byla popsána v roce 1817 Jamesem Parkinsonem v práci s názvem „An Essay on the Shaking Palsy“ [1]. S postupným úbytkem dopaminu, neurotransmiteru potřebného pro přenos vzruchu mezi jednotlivými nervovými buňkami - neurony, se začínají projevovat příznaky nemoci – svalová rigidita, bradykineze a klidový tremor. Schopnost pohybu je výrazně snížena a pacient musí i při běžných činnostech vynakládat značné úsilí. V krajním případě může dojít i k akinezi, kdy nelze pozorovat žádné známky hybnosti. Progrese PN není konstantní, zatím co v prvních stádiích postupuje nemoc velmi rychle, v pokročilejších fázích je její postup zpomaluje. K úbytku dopaminu dochází v průběhu života i u zdravých osob, ne

však tak výrazně jako u pacientů postižených Parkinsonovou chorobou. Onemocnění se projevuje především u starších osob, nejčastěji ve věku 60 let a výše. Má negativní vliv nejen na délku života pacientů, ale a to především, na jeho kvalitu. V současné době není známa příčina odumírání buněk produkujících dopamin ani způsob, jak Parkinsonovu chorobu zcela vyléčit. Léčba tak spočívá především v potlačování příznaků a snaze umožnit pacientům co možná nejvyšší život po co nejdelší dobu.

2.3 Hypokinetická dysartrie

Komplikace, které Parkinsonova nemoc přináší, nesouvisí pouze s tělesnou hybností, ale i se schopností komunikace. Hypokinetická dysartrie (HD) vzniká při poruchách funkce bazálních ganglií a jejich okruhů, což je i případ PN. Projevy jsou patrné především v oblastech fonace, artikulace, prozodie, plynulosti řeči a faciokineze. V rámci této práce jsou hodnoceny první čtyři zmíněné oblasti. Symptomy mohou kolísat v závislosti na medikaci a s ní spojené úrovni hybnosti, na psychickém stavu nebo denní době. Pacient s poruchou řeči může být okolím považován za méně inteligentního nebo dementního jedince se všemi konsekvencemi z toho vyplývajících [2]. Procento udávající počet parkinsoniků postižených HD, se v různých publikacích liší. Tato hodnota se pohybuje v rozmezí 60-90% [4] [5].

2.3.1 Fonace

První oblastí, kde se HD manifestuje je zpravidla fonace. Ztuhlost svalstva má u hlasivek za následek nedokonalé uzavření a redukované asymetrické nebo pomalé chvění. Turbulentní proudění vzduchu z plic je taktéž charakteristické výrazným kolísáním. Výsledkem je hlasový třes [5] [6] [7] [8]. Síla hlasu klesá u pacientů s PN daleko rychleji než u intaktních osob a hlasový rozsah je taktéž menší. Hlasový projev se proto jeví jako monotónní. U pacientů lze pozorovat hypofonii (sníženou intenzitu hlasu) a disfonii, vlivem které je hlas drsný, zastřený a chraplavý. Pro pacienta je taktéž problematické zesílení nebo zeslabení hlasu. Je-li pacient vyzván, aby hlas zesílil, jeho intenzita se krátkodobě zvýší, následně však klesá na původní hodnotu. Z hlediska řečového signálu lze u pacientů s PN pozorovat zvýšenou střední hodnotu frekvence základního tónu F_0 , a to jak u promluvy delších větných celků [10] [11] [12], tak u promluvy prodloužených vokálů [9]. Střední hodnota rovněž roste s prodlužující se dobou onemocnění, proto bývá tento parametr využíván k posuzování pokročilosti choroby [12]. Porucha resonance se projevuje ve formě hypernazality, která se v modulovém spektru řeči projevuje zdůrazněním některých formantových frekvencí.

Proces tvorby hlasu úzce souvisí s respirací. U pacientů s PN se projevují poruchy dýchání obstrukčního charakteru. Síla inspiračních i expiračních svalů se snižuje s progresí nemoci. Zkrácení obou fází dýchání nepříznivě ovlivňuje tvorbu hlasu, jeho kvalitu, sílu a délku promluvy. Pacienti také neefektivně hospodaří s dechem. Postižení náhle snižují či zvyšují tempo řeči, nevhodně se při řeči nadechují,

několikrát rychle opakují počátky slov a obtížně koordinují dýchání a mluvení [2] [16].

2.3.2 Artikulace

Porucha artikulace je dána nesprávnou funkcí řečového svalstva, přesněji artikulačních orgánů (rty, jazyk, hrtanová příklopka a hrtan). Studie uvádějí, že u artikulačního svalstva je snížen rozsah a rychlost pohybu. Slyšitelným projevem je setřelá, nedbalá, nepřesná výslovnost, která má vliv na srozumitelnost řeči. Nejvíce narušena je výslovnost souhlásek [2]. Je zřejmá zejména při vyslovení konsonant [16] [17] [18]. Nejkritičtější je pro pacienty s PN ploviva [k]. K určení kvality artikulace se využívají diadochokinetické úlohy (DDK), založené na opakování slabik vyžadujících rychlé změny v postavení artikulačních orgánů.

2.3.3 Prozodie

Jsou-li narušeny prozodické faktory řeči jako melodie, tempo a vhodné vytváření pauz, mluvíme o tzv. dysprozodii. U osob s PN se setkáváme s deficitem všech těchto faktorů. Řeč pacientů je monotónní a postrádá emoční projevy. Mezi další prozodické rysy patří například rytmus, intonace nebo přízvuk. Prozodie se částečně odvíjí také od fonace. K hodnocení kvality řeči se využívá suprasegmentárních rysů, mezi které patří frekvence základního tónu, intenzita a tempo řeči [2] [16].

2.3.4 Plynulost řeči

Poruchy plynulosti promluvy se projevují buďto jako hezitace – neúmyslné pauzy nebo palilalie – zrychlené opakování slov či slabik. Řeč může být zpomalená (bradyfemie) nebo zrychlená (tachyfemie). Tachyfemie zhoršuje artikulaci a má vliv na umístování pauz v mluveném projevu. Významně tak ovlivňuje celkovou srozumitelnost řeči [2].

2.3.5 UPDRS škála

UPDRS je zkratkou pro Unified Parkinson's Rating Scale, do češtiny je tento název překládán jako Jednotná stupnice pro hodnocení Parkinsonovy nemoci. Jedná se o zřejmě nejkomplexnější a v současnosti nejvíce využívané vyšetřovací schéma. UPDRS je tvořena dílčími škálami (I – VI), zaměřenými na různé skupiny symptomů PN. Dílčí škála UPDRS III slouží k posouzení motorických projevů a jejich vývoje a ke sledování efektu léčby [28].

3 PŘEDZPRACOVÁNÍ SIGNÁLU

Složitost analýzy lidské řeči spočívá v její neopakovatelnosti. Rozdíly nejsou jen mezi jednotlivými řečníky. Ani ten samý řečník nedokáže vyslovit jedno slovo dvakrát naprosto totožně. Negativní vliv na zpracování řečového signálu má také šum a rušení zapříčiněné hlukem okolního prostředí spolu se zkreslením vznikajícím při záznamu a přenosu řeči. Vhodné je proto vstupní signal ještě před samotným zpracováním upravit.

3.1 Preemfázový číslicový filtr

Podstatná část celkové energie řečového signálu je obsažena v kmitočtovém pásmu do 300 Hz, většina užitečné informace ovlivňující srozumitelnost se ale nachází na frekvencích nad touto hranicí. Uvažujeme-li kvantizační šum s přibližně rovnoměrným výkonostním spektrem, je zřejmé, že se jeho rušivý vliv výrazněji projeví právě u spektrálních složek na vyšších kmitočtech. U znělých hlásek navíc první formant obvykle energeticky převyšuje ostatní formanty. Vyrovnání kmitočtové charakteristiky dosáhneme použitím preemfázového číslicového filtru typu horní propust (FIR), jehož přenosová funkce má tvar

$$H(z) = 1 - a_1 z^{-1} \quad (3.1)$$

Koeficient filtru a_1 bývá obvykle v rozsahu 0,9 až 1.

3.2 Odstranění stejnosměrné složky

Řečový signál může obsahovat stejnosměrnou složku, ta ale nemá žádný praktický význam a naopak může způsobovat chyby při výpočtu některých příznaků, jako je počet průchodů nulou, energie signálu atd. Vhodné je proto její odstranění neboli tzv. ustředění. V praxi se provádí jednoduchým odečtením stejnosměrné složky (střední hodnoty) od signálu [20]. Známe-li předem celý signál $s[n]$, pak platí:

$$s'^{[n]} = \frac{1}{N} \sum_{n=0}^{N-1} s[n], \quad (3.2)$$

Kde $s'^{[n]}$ je výsledný signál po odečtení stejnosměrné složky. N je počet vzorků původního signálu a $s[n]$ jsou jednotlivé vzorky původního signálu.

3.3 Segmentace

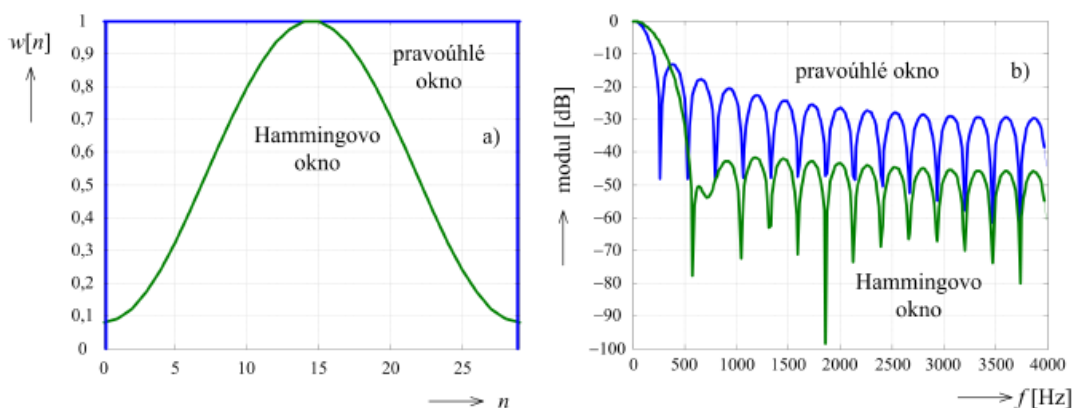
Řečový signál není stacionární a lze jej považovat za náhodný. Zpracování nestacionárního signálu je z principu velmi náročné, potřebujeme proto tento signál převést na signál stacionární. Toho dosáhneme rozdělením signálu na menší části – segmenty o délce N . Při volbě délky segmentu vycházíme z faktu, že artikulační orgány člověka vykazují určitou setrvačnost (10 -30 ms). Důležité je vyvážit požadavek na

neměnnost příznaků řeči v čase, což vyžaduje co nejkratší segment, s požadavkem na dostatečnou přesnost. Vhodná délka segmentu je tak okolo 20 ms (přibližně 160 vzorků při vzorkovací frekvenci 8 kHz). V rámci segmentu se vlastnosti řeči mění dostatečně pomalu a signál lze proto považovat za deterministický a lze jej zpracovávat metodami krátkodobé analýzy. K dělení signálu na segmenty se využívají časová okna, nejčastěji pravoúhlé a Hammingovo. Známe dva typy segmentace, segmentaci bez přesahu a segmentaci s přesahem. V praxi se častěji používá segmentace s přesahem, kdy se sousední segmenty z části překrývají. Běžně volí 50% překryv. Menší překrytí umožňuje větší rychlost signálu v čase a snižuje nároky na paměť. Nevýhodou je, že se hodnoty signálu v sousedních rámcích mohou skokově měnit. Větší překrytí skokovým změnám zamezí. Může při něm ale naopak docházet k přílišné podobnosti příznaků v sousedních rámcích a tím ke znesnadnění detekce daných příznaků. Počet rámců lze vypočítat podle vztahu:

$$N_{ram} = 1 + \left\lfloor \frac{N - I_{ram}}{S_{ram}} \right\rfloor, \quad (3.3)$$

kde N je délka signálu, který chceme segmentovat, I_{ram} je velikost rámce ve vzorcích a S_{ram} je velikost nepřekryté části rámce ve vzorcích.

Rámec vzniká násobením signálu a okénkové funkce v časové oblasti, což odpovídá konvoluci v kmitočtové oblasti. Dochází proto ke změně spektra signálu v rámci. Z grafu je patrné, že Hammingovo okno má širší hlavní lalok a z toho vyplývající horší kmitočtové rozlišení a širší přechodové pásmo než okno pravoúhlé. Jeho výhodou je naopak větší potlačení frekvencí za mezním kmitočtem, díky čemuž nedochází k prosakování spektrálních složek ze sousedních rámců. Vzhledem k těmto vlastnostem je Hammingovo okno při zpracování řečového signálu využíváno častěji.



Obr. 3.1: Časový průběh (a) pravoúhlého okna a Hammingova okna a jejich modul spektra (b) [21]

4 ŘEČOVÉ PŘÍZNAKY

Není reálné zkoumat všechny vlastnosti řečového signálu, proto volíme několik hodnot, tzv. příznaků (parametrů), které by měly co nejlépe popisovat vlastnosti řečového signálu, které chceme sledovat. Vhodné je zaměřit se na projevy řečových patologií a jejich negativní vliv na jednotlivé části řečového aparátu a jako příznaky volit hodnoty, které bývají danou patologií nejvíce ovlivněny, například poruchy funkce fonačního ústrojí, anomálie v artikulaci či intonaci. Tato proces se nazývá parametrizace a jejím výstupem může být skalární hodnota, vektor či matice. Příznaky je možné počítat buď z celého řečového signálu, nebo z krátkých úseků trvajících desítky až stovky milisekund. Tyto úseky jsou vytvářeny násobením řečového signálu váhovací posloupností a v rámci takového úseku je signál považován za přibližně stacionární. Proces segmentace signálu je blíže popsán v kapitole (sem dát odkaz na segmentaci).

Řečové příznaky lze obecně dělit na lokální, které jsou počítány přímo ze zdrojového signálu, vysokoúrovňové což jsou statistiky počítané z lokálních příznaků a příznaky globální, jež jsou kombinací lokálních a vysokoúrovňových příznaků.

V průběhu let bylo vyvinuto mnoho příznaků vhodných pro popis řečového signálu. V této práci budou uvedeny ty, které se v oblasti analýzy patologické řeči nejčastěji používají, popřípadě ty, které mají pro odhalování patologie velký potenciál.

4.1 Příznaky popisující fonaci

4.1.1 Základní frekvence

Pravděpodobně nejpoužívanějším příznakem popisujícím patologickou řeč je frekvence základního tónu F_0 spolu s příznaky popisujícími její změny v průběhu signálu – jitterem a schimmerem. Frekvence F_0 odpovídá základnímu hlasovému tónu mluvčího a je silně vázána na věk a pohlaví. U žen je všeobecně vyšší než u mužů a stejně tak je vyšší i u dětí oproti dospělým. Ve spektrální oblasti odpovídá základní (pitch) frekvence první harmonické složce signálu. Existuje pouze u znělých hlásek, tedy takových, které vznikají kvazy-periodickým buzením hlasivek. Vztah pro výpočet základní frekvence [23]:

$$F_0 = \frac{f_{vz}}{k_{min}}, \quad (4.1)$$

kde f_{vz} je vzorkovací kmitočet a k_{min} je pozice (ve vzorcích) prvního maxima autokorelační funkce. Možností výpočtu základní tónu je celá řada, tento vztah popisuje pouze jednu z nich. Základní frekvence má dobré uplatnění v diferenciaci zdravé a patologické řeči.

4.1.2 Jitter

Jedním z příznaků vyjadřujících variabilitu základní frekvence F_0 v čase je frekvenční nestabilita - jitter. Bývá označován také jako třes či chvění hlasivek. Je to rozdíl v délce dvou sousedních period hlasivkových pulzů v poměru k průměrné délce periody. Délka základní periody, což je vlastně převrácená hodnota základní frekvence, může vykazovat malé odchylky od průměru i v rámci velmi krátkého řečového signálu [20] [21]. Pro výpočet jitteru se používá vztah [21]:

$$J_g = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_g[i] - T_g[i-1]|}{\frac{1}{N} \sum_{i=1}^{N-1} T_g[i]}, \quad (4.2)$$

kde N je počet period hlasivkových pulzů, tj. Počet segmentů signálu, kde je pro každý segment vypočítána základní perioda a $T_g[i]$ je i -tá hodnota základní periody.

Jitter má mnoho variant, které se liší drobnými odlišnostmi ve způsobu výpočtu. Známý jsou například PPQ5 (five-point Pitch Perturbation Quotient), RAP (Relative Average Perturbation), $jitt_{loc}$ (průměrná absolutní odchylka mezi následujícími periodami dělená průměrnou periodou), $jitt_{abs}$ (průměrná absolutní odchylka mezi následujícími periodami), $jitt_{ddp}$ (průměrná absolutní odchylka mezi následujícími odchylkami sousedních glotálních period dělená průměrnou periodou) [15]. Jitter a jeho formy je vhodný především pro detekci vokálního tremoru. Jeho nevýhodou je, stejně jako u základní frekvence, velká závislost na pohlaví řečníka a u nejjednoduššího výpočtu jitteru také potřeba signálu s relativně velkým počtem period, který nemusí být k dispozici [15].

4.1.3 PPE (Pitch Period Entropy)

Zásadní nevýhodu předchozích příznaků, kterou je výrazná závislost na pohlaví řečníka, potlačuje PPE, kterou navrhli Little et. Al. Při výpočtu tohoto příznaku je uvažována logaritmická půltónová škála, inverzní filtrace a odhad entropie [24].

4.1.4 Schimmer

Kolísání úrovně hlasivkových pulzů – schimmer, je dalším příznakem vycházejícím ze základní frekvence F_0 . Zatím co jitter vyjadřuje změny periody, schimmer představuje změny amplitudy hlasivkových pulzů v rámci řečového signálu. Lze ho vypočítat pomocí vztahu:

$$S_g = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_g[i] - A_g[i-1]|}{\frac{1}{N} \sum_{i=1}^{N-1} A_g[i]}, \quad (4.3)$$

kde N je počet amplitud hlasivkových pulzů a $A_g[i]$ je i -tá hodnota základní amplitudy. Stejně jako u jitteru existuje i více druhů schimmeru.

4.1.5 Krátkodobá energie

Krátkodobá energie STE (Short-time energy) vyjadřuje průměrnou hodnotu energie řečového signálu v jednotlivých rámcích vzniklých segmentací signálu. Z teorie tvorby řeči je známo, že znělé hlásky obsahují více energie než neznělé hlásky, které mají šumový charakter a tím pádem obsahují i méně energie. Toho se využívá například při oddělování znělé části promluvy od neznělé nebo při automatickém oddělování segmentů ticha od segmentů řeči v tzv. detektoru řečové aktivity [20] [23]. Výpočet krátkodobé energie pro jeden segment řečového signálu se provádí podle vzorce:

$$STE = \sum_{k=0}^{N-1} s[k]^2, \quad (4.4)$$

kde $s[k]$ je vzorek diskrétního řečového signálu a N je počet vzorků segmentu.

4.1.6 LSTER (Low short-time energy ratio)

Udává počet segmentů, ve kterých je hodnota krátkodobé energie nižší než 0.5 násobek průměrné krátkodobé energie v sekundovém úseku segmentovaného signálu. Používá se k diferenciaci mezi řečovým a hudebním signálem. Řeč vykazuje větší odchylky v energii během úseků délky 10-30 ms než hudební signál. Existuje předpoklad, že v důsledku neschopnosti pacientů s PN udržet stejné množství vzduchu, procházející hrtanovými svaly, bude tento příznak dosahovat u nemocných řečníků vyšších hodnot. Pro výpočet LSTER platí vztah

$$\frac{1}{2N} \sum_{n=0}^{N-1} [sgn(0,5\bar{E} - E(n)) + 1], \quad (4.5)$$

kde E je krátkodobá energie segmentovaného signálu a N je počet segmentů signálu.

4.1.7 TKEO (Teager-Kaiser Energy Operator)

Jedná se o nelineární operátor, který poskytuje odhad okamžité frekvence a amplitudy z FM a AM modulace. Na rozdíl od STE, který pracuje se segmenty, je u zde energie vypočtena pro celý řečový signál. Výhodou oproti E je to, že bere v úvahu také frekvenci signálu. V diskrétním čase je TKEO definován jako

$$\psi[s(n)] = s(n)^2 - s(n+1)s(n-1), \quad (4.6)$$

kde $s(n)$ je vstupní řečový signál.

4.1.8 MPSD (Median of Power Spectral Density)

Medián spektrální výkonové hustoty udává výkon signálu v určitém frekvenčním pásmu. Pro výpočet odhadu výkonové spektrální hustoty může být použito několik různých metod. Jednou z nich je metoda periodogramu, definovaná jako

$$P_{xx}(e^{j\omega}) = \sum_{m=-(N-1)}^{N-1} r_{xx}[m] e^{-j\omega m}, \quad (4.7)$$

kde $r_{xx}[m]$ je autokorelační posloupnost, jejíž hodnotu získáme ze vztahu

$$r_{xx}[m] = \frac{1}{2N+1} \sum_{n=-N}^N x^*[n]x[n+m]. \quad (4.8)$$

Periodogram může být vyjádřen i pomocí DFT jako

$$P_{xx}(e^{j\omega}) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n]e^{-j\omega n} \right|^2 \quad (4.9)$$

$$= \frac{1}{N} |X(e^{j\omega})|^2. \quad (4.10)$$

Vstupní signál je nejprve rozdělen na segmenty, u kterých je následně provedena analýza výkonové spektrální hustoty. Daný postup odpovídá tzv. neparametrickým metodám odhadu PSD, které pro výpočet spektra zkoumaného signálu využívají algoritmus FFT. Předchozí vztah lze pak vyjádřit ve tvaru

$$P_{xx}(e^{j2\pi f k}) = \left| \sum_{n=0}^{N-1} x[n]e^{-jk\frac{2\pi}{N}n} \right|^2, \quad (4.11)$$

kde $k = 0, 1, \dots, N-1$. Periodogram je počítaný v N frekvenčních bodech f_k .

U Welchovy metody průměrování modifikovaných periodogramů se používá segmentace s překrýváním jednotlivých segmentů. Toto překrytí bývá z pravidla 50%. Analyzovaná data jsou rozdělena podle vztahu

$$s_i[n] = s_i[n + iD]. \quad (4.12)$$

Hodnota iD je počátečním bodem i -té posloupnosti $i = 0, 1, \dots, L-1$, kde L je počet posobě jdoucích segmentů. Periodogram modifikovaný podle Welchovy metody je definován jako

$$\tilde{P}_{xx}^{(i)}(e^{j2\pi f}) = \frac{1}{MU} \left| \sum_{n=0}^{M-1} x_i[n]w[n]e^{j2\pi f n} \right|^2, \quad (4.13)$$

kde $i = 0, 1, \dots, L-1$ a U je normalizační koeficient

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2[n]. \quad (4.14)$$

Odhad výkonové spektrální hustoty počítaný pomocí Welchovy metody je definovaný jako aritmetický průměr modifikovaných periodogramů

$$P_{xx}^w(e^{j2\pi f}) = \frac{1}{L} \sum_{i=0}^{L-1} \tilde{P}_{xx}^{(i)}(e^{j2\pi f}). \quad (4.15)$$

Pro zjištění středního výkonu v určitém pásmu zkoumaného signálu je třeba vypočítat medián výkonové spektrální hustoty podle vztahu

$$\text{med} \left(P_{xx}^w(e^{j2\pi f}) \right) = P_{xx}^w(e^{j2\pi f}) \left[\frac{n+1}{2} \right], \quad (4.16)$$

pro lichý počet vzorků a

$$\text{med} \left(P_{xx}^w(e^{j2\pi f}) \right) = \frac{1}{2} \left(P_{xx}^w(e^{j2\pi f}) \left[\frac{n}{2} \right] + P_{xx}^w(e^{j2\pi f}) \left[\frac{n}{2} + 1 \right] \right), \quad (4.17)$$

pro sudý počet vzorků výkonové spektrální hustoty.

4.2 Příznaky popisující rezonanci (formanty)

Jak již bylo zmíněno v 1.1.3, dochází při průchodu turbulentního proudění hlasovým traktem k rezonancím v dutině hrdelní, ústní a nosní. Frekvence, na nichž k těmto rezonancím dochází se nazývají formanty a značí se $F_1, F_2, \dots, F_n$. Aby byla řeč srozumitelná, musí signál obsahovat formanty alespoň tři [Mluvíme s počítačem česky]. Objem nadhrtanových dutin, a tím i frekvence jednotlivých formantů, ovlivňuje poloha artikulačních orgánů. Tento příznak je tedy vhodným kandidátem pro sledování hybnosti jazyka a celkové pohyblivosti mluvidel. Podle Hillebranda et al. jsou formanty F_1 a F_2 příznaky, které nejvíce vypovídají o produkci vokálů [15] a proto jsou spolu s příznaky z nich odvozenými používány k popisu patologické řeči. Značnou nevýhodou formantových frekvencí je jejich závislost na věku a pohlaví.

V roce 1780 definoval Ch. F. Hellwag ve svém díle „Dissertatio de formantione loquela“ závislost vokálů na poloze jazyka. Vytvořil tzv. Hellwagův trojúhelník, který mapuje vertikální a horizontální pohyb viz Tab. 4.1. Později byla také dokázána souvislost tohoto trojúhelníku s formantovými frekvencemi [15]. Vertikální pohyb jazyka má vliv na první formant, kdežto horizontální pohyb ovlivňuje formant druhý. Tyto vlastnosti umožňují podle obsahu trojúhelníku vyjádřit míru pohybu jazyka. Na tomto principu je založen příznak VSA (Vowel Speech Area) a jeho logaritmická forma $\ln VSA$.

Tab. 4.1: Hellwagův trojúhelník (klasifikace vokálů z hlediska polohy jazyka)

		$F_2 \rightarrow$			
		Zadní		Střední	Přední
$F_1 \uparrow$	Nízké	[a]			
	Středové	[o]		[e]	
	Vysoké	[u]		[i]	

4.3 Příznaky založené na hodnocení kvality signálu

Jedná se o výpočetně nenáročné příznaky, hodnotící energii signálu, úroveň šumu v signálu, periodicitu atd. Počítány jsou jak v časové, tak i ve frekvenční oblasti. Tato práce se bude zabývat příznaky ZCR, HZ-CRR a HNR které budou vysvětleny v jednotlivých podkapitolách.

4.3.1 ZCR (zero-crossing rate)

Tento příznak je počítán v časové oblasti a udává počet průchodů řečového signálu nulovou úrovní. Průchod nulou nastane, jestliže má amplituda dvou sousedních

vzorků zkoumaného signálu opačné znaménko. Velkou výhodou ZCR je nezávislost na energii signálu. Pro výpočet krátkodobé střední hodnoty počtu průchodů signálu nulovou úrovní se používá vztah

$$Z_{cr} = \sum_{k=0}^{N-1} |\text{sign}(x[k]) - \text{sign}(x[k-1])| / 2, \quad (4.18)$$

$$\text{sign}(x[k]) = \begin{cases} +1 & \text{ak } x[k] > 0 \text{ alebo } x[k] = 0 \wedge x[k-1] > 0, \\ -1 & \text{ak } x[k] < 0 \text{ alebo } x[k] = 0 \wedge x[k-1] < 0. \end{cases} \quad (4.19)$$

Platí, že $x[k]$ je vzorek signálu, k je pořadí vzorku v daném segmentu a N je počet vzorků v segmentu.

4.3.2 HZCRR (high zero-crossing rate ratio)

Je dán poměrem počtu segmentů, ve kterých je průchod nulou větší než 1,5 násobek průměru v jedné sekundě sledovaného signálu. Používá se pro diferenciaci řečového a hudebního signálu, kdy se předpokládá, že řečový signál je složen ze znělých a neznělých úseků, kdežto v hudebním signálu se neznělé úseky nevyskytují. Pro výpočet HZCRR platí vztah

$$HZCRR = \frac{1}{2N} \sum_{n=0}^{N-1} \left[\text{sgn} \left(ZRC(n) - 1,5 \overline{ZCR} \right) + 1 \right], \quad (4.20)$$

kde ZCR je počet průchodů nulou a N je počet segmentů signálu.

4.3.3 HNR (harmonic-to-noise ratio)

Příznak udávající poměr úrovně energie užitečného signálu a úrovně energie šumu. HNR vyjadřuje stupeň periodicity zkoumaného průběhu a je využíván k detekci řečové aktivity. Hodnotu příznaku lze zjistit ze vztahu

$$HNR = 10 \log_{10} \left(\frac{\sum_{i=1}^{N/2} |s_i|^2}{\sum_{i=1}^{N/2} |N_i|^2} \right), \quad (4.21)$$

kde $|s_i|$ je energie harmonického signálu a $|N_i|$ je odhad energie šumu [25].

4.4 Extrakce vysokoúrovňových příznaků

Řečové příznaky počítané přímo ze zdrojového signálu jsou označovány jako lokální příznaky. Pokud jsou příznaky počítány pro každý segment signálu zvlášť, jedná se o příznaky vysokoúrovňové. Jejich výstup je reprezentován vektorem či maticí hodnot, které je nutno v dalším kroku transformovat na skalární hodnotu tak, aby mohla být použita pro statistickou analýzu a klasifikaci. K tomuto účelu se většinou používá určitý druh statistiky. V rámci zpracování řečového signálu se nejčastěji používají tyto statistiky: střední hodnota, medián, rozptyl, směrodatná odchylka a percentily [15].

5 ANALÝZA ŘEČOVÝCH PŘÍZNAKŮ

5.1 Korelace

Při výzkumu bývá často prováděna korelace mezi klinickými a paraklinickými daty. Jsou sledovány například závislosti mezi řečovými příznaky a věkem pacienta, dobou trvání PN či různými škálami (UPDRS, FOG – Freezing of Gait, NMSS – Non-motor Symptoms Scale, LED – L-dopa Equivalent Daily dose), atd [15]. Korelace určuje míru statistické závislosti mezi dvěma náhodnými proměnnými X a Y . Dvě veličiny jsou korelované, jestliže určité hodnoty jedné veličiny mají tendenci se vyskytovat společně s určitými hodnotami druhé veličiny. Jedním z možných vztahů mezi dvěma proměnnými je lineární závislost, kterou je možno vyjádřit korelačními koeficienty. Existuje požadavek, aby absolutní hodnota míry statistické závislosti r ležela v uzavřeném intervalu od 0 po 1, a r může nabývat hodnot od -1 do +1, kdy platí:

- pro $r = -1$ (nepřímá závislost) – hodnota jedné proměnné klesá, zatímco hodnota druhé stoupá
- pro $r = +1$ (přímá závislost) – hodnoty obou proměnných stoupají a klesají zároveň
- pro $r = 0$ – mezi proměnnými neexistuje lineární vztah

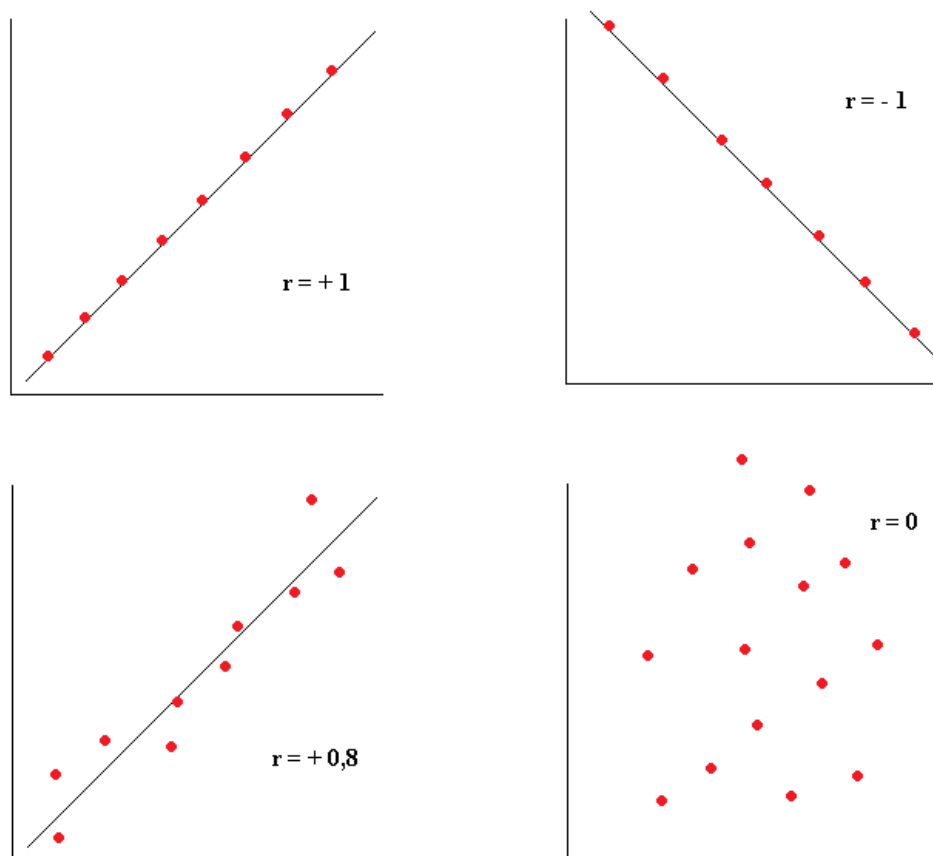
Neexistence lineárního vztahu však neznamená, že mezi hodnotami neexistuje žádná souvislost, pouze je jiného typu. Stejně tak statistická závislost není jistotou kauzality [29]. Různé typy lineární závislosti jsou znázorněny na Obr. 5.1. Spearmanova a Kendalova korelace jsou neparametrické, zatímco Pearsonova korelace je parametrická. Většinou se počítají jejich parciální formy, při kterých jsou odstraněny možné vlivy dalších veličin [15]. Vztah pro výpočet Pearsonovy korelace je:

$$r(X, Y)_{Pearson} = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}} = \frac{\sum_{i=1}^N [(x_i - \mu_X) \cdot (y_i - \mu_Y)]}{\sqrt{\sum_{i=1}^N (x_i - \mu_X)^2 \cdot \sum_{i=1}^N (y_i - \mu_Y)^2}} \quad (5.1)$$

Kde Cov značí kovariaci, N je počet vzorků (realizací) X a Y . Spearmanovu korelaci lze vypočítat podle vzorce:

$$r(X, Y)_{Spearman} = 1 - \frac{6 \sum_i (x_i - y_i)^2}{N \cdot (N^2 - 1)} \quad (5.2)$$

Dalším příznakem v souvislosti s korelací je hladina významnosti statistického testu (p -hodnota). Pokud je tato hodnota menší než 0,05 znamená to, že pravděpodobnost, že by pozorované závislosti vznikly čistě náhodou, je menší než 5%. [29].



Obr. 5.1: Míra lineární závislosti dvou proměnných vyjádřená korelačním koeficientem r

5.2 Strojové učení

Cílem strojového učení (Machine Learning) je vytvořit model, který dokáže na základě příznaků (features) X určit hodnotu cílové veličiny Y . Ta může být buďto kvantitativní (cena akcií na burze) nebo kvalitativní (odpověď systému pro analýzu rizik, jestli je danému klientovi vhodné poskytnout půjčku či nikoliv – tj. odpovědi ano/ne). Vstupní příznaky (atributy) mohou být také dvojího typu – kategoriální (diskrétní) a numerické (spojité). Učení modelu může probíhat s učitelem nebo bez učitele. Rozdíl je v tom, že při učení bez učitele nemáme žádnou cílovou veličinu Y , úkolem modelu je objektivní popis dat například klastrováním nebo určením závislostí. Princip strojového učení s učitelem je následující:

- Máme množinu trénovacích dat, u kterých známe jak příznaky X tak i hodnotu cílové veličiny Y
- Na základě těchto dat je vytváříme (trénujeme) model
- Trénování (větvení stromu, zvětšování jeho hloubky) je prováděno tak dlouho, dokud chyba predikce není dostatečně malá

Na trénovacích datech můžeme vytvořit model s prakticky libovolně malou chybou.

Proto chyba, s jakou model klasifikuje trénovací data, příliš nevypovídá o kvalitě tohoto modelu. Na testovacích datech takovýto model obvykle dosahuje velmi špatných výsledků. Tento jev se nazývá přeučení či přetrénování modelu (overfitting) a lze mu předejít tím, že data v první fázi rozdělíme na trénovací, testovací a validační. Parametry učícího se algoritmu se nastavují tak, aby se při natrénování pomocí trénovacích dat a ověření na testovacích datech, dosáhlo co nejlepších výsledků. Tím se ovšem nezávislost dat částečně vytrácí a pro objektivní posouzení přesnosti by měla existovat ještě data validační. Modely používané při strojovém učení jsou například neuronové sítě, algoritmus k -nejbližších sousedů, podpůrné vektory, lineární diskriminační analýza atd. Často používaným modelem jsou také rozhodovací stromy.

5.2.1 Rozhodovací stromy

Rozhodovací stromy (decision Trees), jinak nazývané taky klasifikační a regresní stromy jsou schopny na základě zkušeností (dat) z minulosti předpovědět budoucnost. Jejich výhodou je snadná interpretace, kterou zvládne i laik s minimem znalostí této problematiky. Nevýhodou je pak relativně malá míra přesnosti a z toho plynoucí nepřiliš uspokojivé výsledky. V závislosti na použitém algoritmu může mít každý uzel stromu, který není konečný, dvě nebo více větví. Pokud může algoritmus vytvářet více než dvě větve pro každý uzel, jedná se o N -ární stromy, kde N je počet větví vycházejících z každého uzlu. Speciálním případem N -árního stromu je strom binární (binary tree), kde z každého uzlu vycházejí právě dvě větve.

Vstupem je vektor atributů. Při rozhodování je postupováno od kořene (root) k listům stromu. Každý vnitřní uzel stromu reprezentuje podmínku (prediktor), každý list stromu reprezentuje rozhodnutí, tj. jednu z možných výstupních hodnot. Na konci je tak hodnotě každého vstupního atributu přiřazena výstupní hodnota. Výstupem klasifikačních stromů jsou nominální hodnoty (ano/ne) na rozdíl od regresních stromů, kde jsou výstupem hodnoty číselné.

Algoritmů schopných vytvářet rozhodovací stromy je celá řada – Huntův algoritmus, ID3, C4.5, SLIQ, SPRTINT, CART. V rámci této práce byl použit algoritmus CART, jehož princip je podrobně popsán například v [30].

5.2.2 Cross-validace

V ideálním případě se vstupní data rozdělí na data trénovací, data testovací a data validační. Toto je možné pouze v případě, kdy je dat dostatek, což není příliš běžné. V případě nedostatečného množství dat, je třeba použít některou z více sofistikovaných metod posouzení klasifikovaného modelu. Jednou z nich je cross-validace, která pracuje s daty, která slouží zároveň k trénování i testování. Data jsou nejprve rozdělena do n skupin. Nejčastěji je n rovno 10. Následně je 9 skupin použito pro učení algoritmu a zbývající desátá skupina pro testování. Postup se opakuje n -krát a každá skupina je právě jedenkrát použita jako testovací. Metody pro vytváření množin

(skupin) dat jsou následující:

- Lineární vzorkování (data jsou rozdělena do skupin tak jak jdou za sebou)
- Promíchané (schuffled) vzorkování (data jsou promíchána, jsou zahrnuty všechny vzorky a nikdy se žádný neopakuje)
- Vrstvené (stratified) vzorkování (data jsou promíchána, některé hodnoty se mohou opakovat a některé se dokonce pro trénování mohou vynechat – cílem je dosáhnout vyšší odolnosti proti proměnlivosti dat)

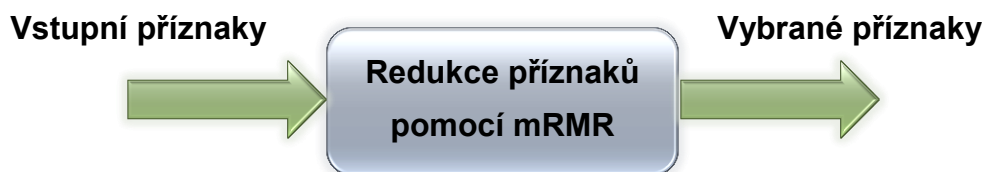
Speciálním případem cross-validace je leave-one-out, neboli vynechání jednoho ze vzorků dat. Jedná se o nejpřesnější možné zhodnocení, avšak je velmi časově náročné [31].

5.3 Výběr příznaků

Jak už bylo zmíněno v kapitole 25, při výběru příznaků vhodných pro odhad progresu PN je vhodné se zaměřit na ty příznaky, které nejlépe charakterizují kvalitativní popis dané nemoci (například dysfonie, hypofonie či disprosodie). Jinak řečeno, na oblasti, kde se její projevy nejčastěji manifestují. Jeden příznak nemusí být dostačující pro takovýto odhad a naopak v široké množině příznaků mohou být některé příznaky redundantní (navzájem korelované) či irelevantní. Velké množství příznaků může představovat problém i z pohledu strojového učení. Může vést k větší výpočetní náročnosti, menší efektivitě systému, přetrénování klasifikátoru apod. Proto je vhodné použít metodu nebo kombinaci několika metod s jejich pomocí je vybrána optimální skupina příznaků [21]. Tento proces se označuje jako selekce příznaků a dělí se na tzv. filter a wrapper metody [26].

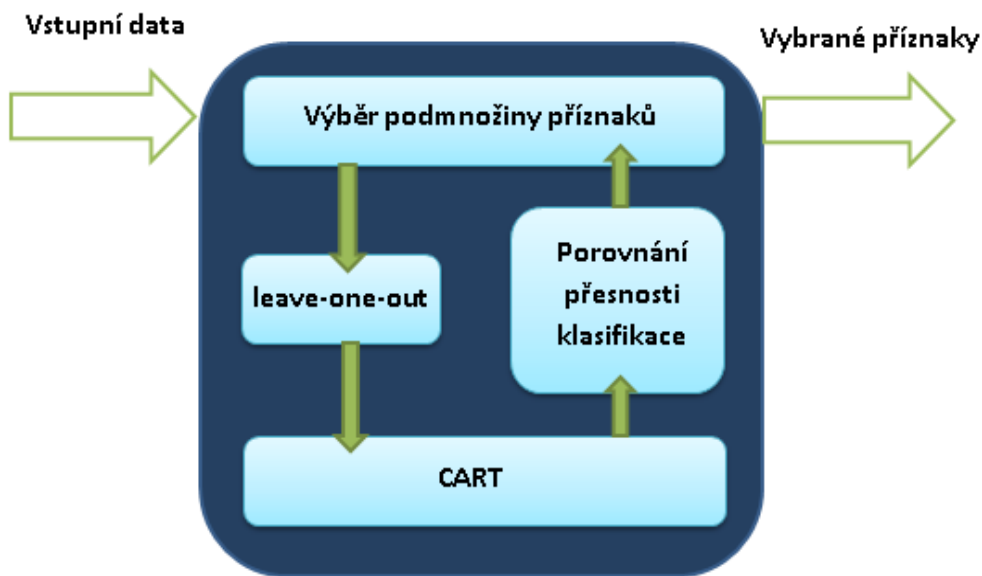
5.3.1 Metody výběru

Filter metody (na Obr. 5.2) vybírají podmnožinu příznaků zcela nezávisle na použitém klasifikátoru. Tyto metody jsou rychlé, avšak poskytují horší výsledky. Zástupcem je algoritmus mRMR (minimum Redundancy Maximum Relevance). Používá pro předvýběr ve chvíli, kdy je třeba ohodnotit úspěšnost velkého množství příznaků. Maximální relevance je dosahováno analýzou vzájemné informace mezi diskretizovanými příznaky a jejich třídami. Zároveň je hodnocena i vzájemná korelace mezi diskretizovanými příznaky, čímž se minimalizuje redundance. Výstupem algoritmu jsou příznaky seřazené podle jejich relevance a následně je vybráno prvních n příznaků. Schéma metody je na obrázku Podrobný princip funkce algoritmu mRMR lze nalézt v [27].



Obr. 5.2: Schéma metody filter pro výběr příznaků

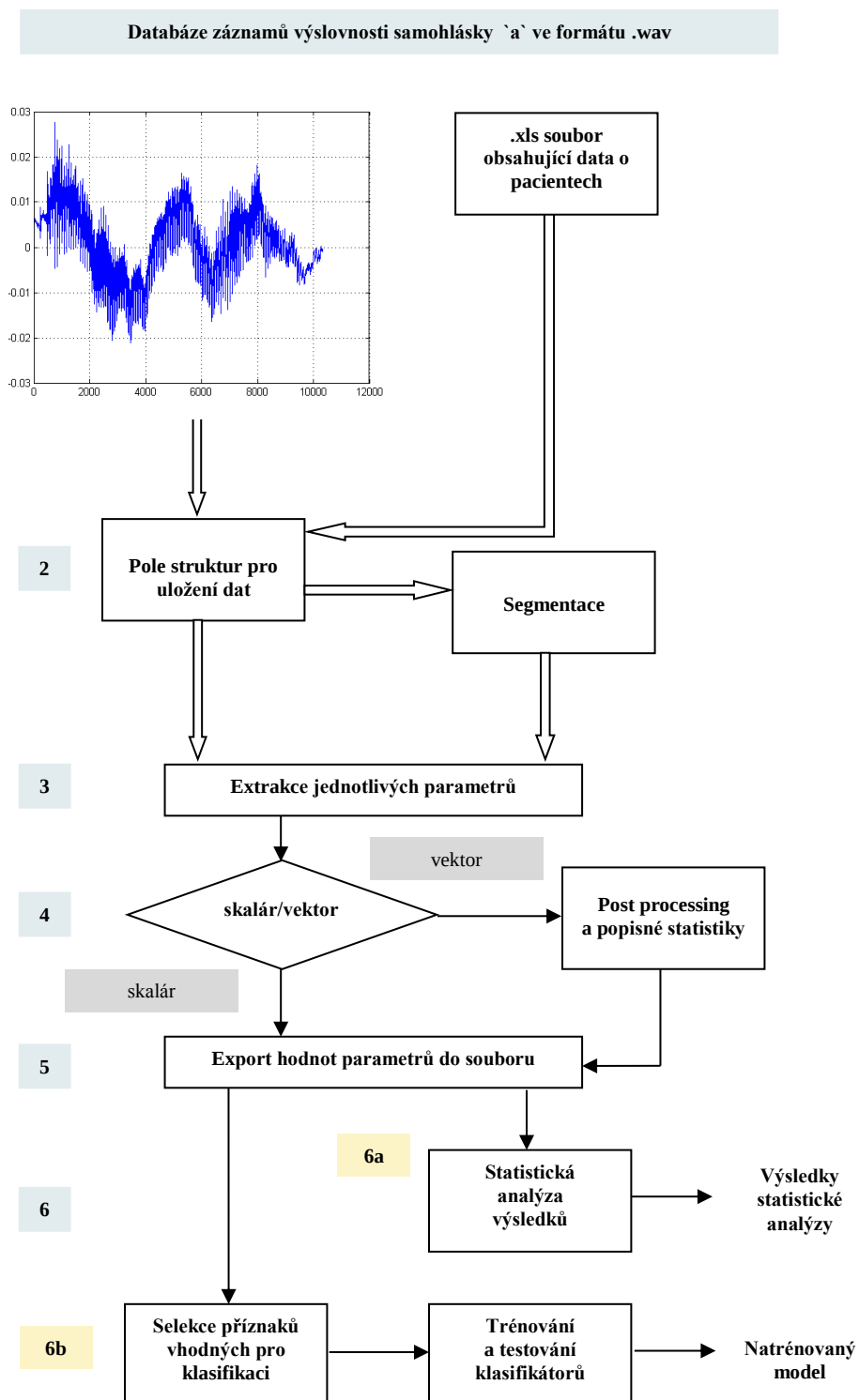
Další často používanou metodou výběru je algoritmus dopředné selekce příznaků SFFS (Sequential Forward Floating Search). Jedná se o plovoucí (floating) metodu, u které se rozměry vybrané podmnožiny příznaků nemění monotónně, ale jakoby „plave“ nahoru a dolů. Výhodou metody je, že při výběru optimálních příznaků rovnou používá klasifikátor, který bude následně použit i při klasifikaci. V prvním kroku je hodnocen každý příznak zvlášť, příznaky jsou ohodnoceny podle úspěšnosti v odhadu požadované hodnoty (v této práci například UPDRS III škály). Nejúspěšnější příznak je následně dáván do dvojice s dalšími příznaky a je zjišťováno, zda existuje nějaká dvojice, která má vyšší úspěšnost odhadu požadované hodnoty, než nejlepší parametr individuálně. Po nalezení nejúspěšnější dvojice příznaků je hledána trojice, čtveřice, atd. Hledání je prováděno do chvíle, kdy je nalezena skupina příznaků s nejvyšší úspěšností v odhadu požadované hodnoty. Zjednodušeně je princip algoritmu naznačen na Obr. 5.3.



Obr. 5.3: Schéma metody SFFS pro výběr příznaků

6 NÁVRH SYSTÉMU

6.1 Vývojový diagram



6.2 Popis systému

Navrhnutý systém se skládá z několika bloků realizovaných jako funkce v prostředí MATLAB (verze 8.6.0.267246 (R2015b, 32-bit)). Tyto funkce jsou postupně volány v hlavním skriptu. Schematicky jsou jejich vzájemný vztah a posloupnost znázorněny v kapitole 6.1. Na vstupu systému jsou načítány záznamy prodloužené fonace samohlásky [á] ve formátu *.wav. Z názvů nahrávek je zjištěno pohlaví a zdravotní stav řečníka (zdravá osoba/pacient s PN). Druhým vstupem je soubor ve formátu *.xlsx se záznamy o řečnících. Je zjištěno, zda jsou pro danou nahrávku k dispozici v souboru *.xlsx validní klinická data. Pokud ano, je pro pacienty s PN ze souboru získána hodnota UPDRS III škály, která bude dále sloužit pro trénování a testování odhadu progresu nemoci. Dále jsou extrahovány požadované informace – vzorkovací frekvence, samotný signál. V případě, že nejsou k dispozici informace o řečnících, je nahrávka považována za neúčinnou a je vyřazena z dalšího zpracování. Vstupní signál je v bloku pro předzpracování ustředěn, tj. Je odečtena jeho stejnosměrná složka a poté filtrován preemfázovým filter. Následuje segmentace signálu, kdy je původní signál rozdělen na menší části, tzv. segmenty. Toto dělení je prováděno násobením vstupního signálu váhovací posloupností. V této práci je jako váhovací posloupnost použito Hammingovo okno. Podrobnější popis segmentace viz kapitola 3.3. V následujícím bloku probíhá výpočet řečových příznaků, jež jsou podrobně popsány v kapitole 4. Dalším krokem je extrakce tzv. vysokoúrovňových příznaků. Z matic a vektorů jsou vypočteny statistické hodnoty jako medián, směrodatná odchylka, odhad střední hodnoty apod. Získané hodnoty příznaků jsou uloženy do matice příznaků a následně vyexportované do souboru vhodného pro použití navrhnutým klasifikačním modelem. Pro každý příznak je provedena Spearmanova a Pearsonova korelace se škálou UPDRS III a pro obě tyto korelace je vypočtena dosažená hladina významnosti. Následuje individuální analýza příznaků, kdy je každý příznak zvlášť použit pro učení algoritmu CART (Classification and Regression Trees). Hodnota UPDRS III odhadnutá pomocí příznaků, je porovnávána s hodnotou UPDRS III určenou subjektivně lékařem, která je pro každou nahrávku k dispozici. Je testována přesnost, s jakou je schopen odhadnout progresi PN nemoci ze vstupního záznamu, tato přesnost je vyjádřena pomocí hodnot MAE (Mean Absolute Error) a RMSE (Root Mean Square Error) a jejich směrodatných odchylek. Protože jeden příznak pro přesný odhad často nedostačuje a naopak v celé množině příznaků mohou některé z nich být redundantní či irelevantní, je použita metoda SFFS pro výběr optimální skupiny příznaků, která bude nejlépe charakterizovat kvalitativní popis projevů hypokinetické dysartrie v hlase. Výsledkem tohoto testování je soubor příznaků, při jejichž použití dosahuje navrhnutý systém nejvyšší přesnosti odhadu hodnoty UPDRS III škály. Výsledkem systému je natrénovaný model automatického odhadu progresu Parkinsonovy nemoci na základě analýzy řečového projevu.

7 PRAKTICKÁ ČÁST

V této kapitole bude popsán systém pro automatický odhad hodnoty UPDRS III u osob postižených Parkinsonovou nemocí, vytvořený na základě znalostí získaných z předchozích kapitol. Učení algoritmu i jeho testování probíhalo na nahrávkách řečového cvičení prodloužené fonace hlásky [a]. Nahrávky jsou součástí databáze PARCZ. Tato databáze řečových nahrávek byla poskytnuta Fakultní nemocnicí u sv. Anny v Brně. Její část s nahrávkami prodloužené fonace hlásky [a], která je v práci použita, obsahuje 152 řečníků. Jejich složení v databázi je naznačeno Tab. 7.1. Záznamy cvičení byly ve formátu *.wav se vzorkovací frekvencí 16 kHz. Zpracovávány byly pouze nahrávky pacientů postižených PD, z nichž bylo 40 žen a 59 mužů. Hodnoty UPDRS III pro tyto pacienty byly v rozsahu 3 až 55 u žen a 5 až 49 u mužů. Úlohou navrhnutého systému je pro každého řečníka odhadnout, na základě objektivních řečových příznaků, hodnotu UPDRS III škály a to s co největší přesností. Tyto hodnoty jsou porovnávány s hodnotami UPDRS III škály subjektivně určenými lékařem. V rámci této práce je hodnocena přesnost odhadu jak pomocí jednotlivých řečových příznaků, tak jejich kombinací.

Tab. 7.1: Statistické parametry databáze PARCZ

	Kontrolní		PN	
	Muži	Ženy	Muži	Ženy
Počet	26	26	59	41
Průměrná hodnota UPDRS III	—	—	27±10	22±14
Rozsah hodnot UPDRS III	—	—	5 - 49	3 - 55
Průměrný věk	66±9	62±9	66±9	69±8
Věkový rozsah	49 - 83	45 - 87	46 - 87	49 - 86

7.1 Předzpracování vstupních dat

Vstupem systému je databáze nahrávek řečového cvičení prodloužené fonace hlásky [a], prováděného jednotlivými pacienty a dále pak *.xlsx soubor s klinickými daty o těchto řečnících. V prvním kroku je pomocí souboru **gen_file_list.bat** vygenerován soubor **list.txt**, který obsahuje absolutní cesty ke všem záznamům v testovací databázi. Z absolutních cest jsou zjištěna jména nahrávek, což jsou kódy přiřazující danou nahrávku konkrétnímu pacientovi za současného zachování jeho anonymity. Každá z nahrávek nese název podle logiky:

- první písmeno značí klinický stav řečníka K – zdravý, P – pacient s PN

- první číslo indikuje pohlaví řečníka 1 – žena, 2- muž
- zbylé tři hodnoty jsou osobní číslo pacienta

Příkladem je “ P1086_7.1-2-a_1.wav ” – jedná se o pacienta ženského pohlaví. Následuje načtení souboru **patients_info_full.xls**. U každé nahrávky je porovnán její název, přesněji prvních 5 znaků značících ID pacienta s ID uvedenými u klinických údajů v souboru *.xls. Pokud je pro danou nahrávku nalezena shoda, je ze souboru zjištěna hodnota UPDRS III škály konkrétního pacienta a nahrávka je načtena do systému pomocí funkce **audioread**. Je zjištěna vzorkovací frekvence a všechny tyto hodnoty jsou uloženy do předpřipravené datové struktury **database_info**. Pokud pro danou nahrávku shoda nalezena není, je vyřazena z dalšího zpracování.

7.2 Předzpracování signálu

Extrahovaný signál je nejprve vhodné uvést do podoby vhodné pro další zpracování. To obnáší odstranění rušivé stejnosměrné složky a tzv., preemfázi, tj. vyrovnaní kmitočtové charakteristiky. Pro tento účel byly vytvořeny skripty **dccomp.m** – odstranění stejnosměrné složky a **preemphasis.m** – filtrace preemfázovým filtrem. Samotné odstranění stejnosměrné složky by nebylo nezbytně nutné, jelikož ta je z principu výrazně potlačena vlivem preemfázového filtru, což je filtr typu horní propust. Proces předzpracování signálu je podrobněji popsán v kapitole 3. Praat si zajišťuje předzpracování signálu pro výpočet příznaků sám.

V rámci předzpracování signálu je také provedena jeho segmentace za použití Hammingova váhovacího okna. Délka okna je volena 20 ms, což je doporučována délka okna pro segmentaci řečového signálu a překryv jednotlivých segmentů je 50%. Pro segmentaci signálu slouží script **segmentation.m**. Výstupem segmentace je matice ve tvaru:

[vzorky signálu x segmenty].

7.3 Parametrizace

Pro výpočet řečových příznaků je využíváno několik funkcí realizovaných v prostředí MATLAB. Pro výpočet některých příznaků byl použit program praatcon.exe, který je volně dostupný na [32]. Funkce pro výpočet příznaků jsou postupně volány v hlavním skriptu **main_function.m**. Vypočtené příznaky pro každého pacienta, jsou ukládány do struktury **database_info** a celá tato struktura je následně vyexportována do souboru **database_info.mat**. Celkově je v této práci počítáno 39 příznaků, které jsou blíže popsány v kapitole 4. Seznam a popis všech použitých příznaků viz příloha B. Schéma struktury **database_info** je následující:

```

[pohlaví] [ stav ] [ UPDRS III ] [1. příznak] [2. příznak] ...
[1. pacient] [ Ž/M ] [HC/PD] [ hodnota ] [1. příznak] [2. příznak] ...
[2. pacient] [ Ž/M ] [HC/PD] [ hodnota ] [1. příznak] [2. příznak] ...

. . . .
. . . .
. . . .

```

7.3.1 Příznaky popisující fonaci

Fonace je jednou z prvních oblastí, ve kterých je možno poruchy spojené s hypokinetickou dysartrií pozorovat. Mezi příznaky, které jsou s touto součástí tvorby řeči nejčastěji spojovány, patří frekvence základního tónu F_0 a její proměny v čase. Podrobný popis příznaků spojených s fonací viz kapitola 4.1. V práci jsou pro výpočet těchto příznaků použity tyto skripty:

- **energy.m** viz (4.1.5) – jedná se o funkci, která počítá hodnotu krátkodobé energie (STE) pro jednotlivé úseky signálu. Vstupem funkce je předzpracovaný a segmentovaný signál a jejím výstupem pak statistické hodnoty STE (střední hodnota, směrodatná odchylka, minimum a maximum).
- **tkeo.m** viz (4.1.7) – funkce počítající hodnoty Teager-Kaiserova energetického operátoru. Jedná se o odhad okamžité hodnoty frekvence a amplitudy. Na rozdíl od STE je vstupem celistvý předzpracovaný řečový signál.
- **median_psd.m** (4.1.8) – tato funkce počítá medián spektrální výkonové hustoty, který udává výkon signálu v určitém frekvenčním pásmu. Vstupem funkce je předzpracovaný a segmentovaný signál a jejím výstupem pak statistické hodnoty STE (střední hodnota, směrodatná odchylka, minimum a maximum).
- **praat_info.m** – skript, pomocí něhož je volán program **praatcon.exe**, který slouží k výpočtu základní frekvence kmitání hlasivek F_0 viz (4.1.1). Praat provádí předzpracování signálu a jeho segmentaci sám, proto je vstupem funkce původní signál a hodnota vzorkovací frekvence F_s . Kromě základního tónu F_0 počítá program i jeho statistické údaje (střední hodnotu, směrodatnou odchylku, minimum a maximum). Dále počítá jitter, schimmer a příznaky hodnotící kvalitu signálu - HNR a NHR.

7.3.2 Příznaky popisující rezonanci (formanty)

- **praat_formants.m** – skript volá program **praatcon.exe**, který vrací střední hodnotu, směrodatnou odchylku, minimum a maximum pro první, druhý a třetí formant viz (4.2). Z důvodů nedostatečné délky signálu, která je příčinou chyb ve výpočtu třetího formantu, nebyl tento parametr v práci použit.

7.3.3 Příznaky založené na hodnocení kvality signálu

- **zero_crossing_rate.m** viz (4.3.1) – vypočítá střední hodnotu, směrodatnou odchylku, minimum a maximum počtu průchodů signálu nulovou úrovní. Vstupem funkce je předzpracovaný a segmentovaný signál.
- příznaky **HNR hn** a **HNR nh** viz (4.3.3) jsou počítány pomocí výše zmíněného skriptu **praat_info.m**

7.4 Korelace příznaků s UPDRS III

Způsobem, který může značně vypovídat o vhodnosti použití jedné hodnoty pro odhad jiné hodnoty, je výpočet jejich korelace. V této práci je postupně prováděna korelace jednotlivých příznaků se škálou UPDRS III. Jedná se tedy o snahu uvést do souvislosti data subjektivní – UPDRS III škála určená lékařem a data čistě objektivní – hodnoty příznaků vypočtené z nahrávky řečového cvičení. Pro tento účel byl vytvořen skript **feature_analysis.m**, kde jsou ze souboru **feature_matrix.xlsx** získány hodnoty jednotlivých příznaků a hodnoty UPDRS III škály, které slouží jako vstup pro funkci **corr** implementovanou v prostředí MATLAB. Pro každý jednotlivý příznak je vypočtena jak lineární Pearsonova korelace, tak nelineární Spearmanova korelace. Pro každý typ korelace je navíc vypočtena hodnota hladiny významnosti (p-hodnota). Hodnoty vyexportovány do souboru **correlation_matrix.xlsx**.

V Tab. 7.2 je výpis deseti příznaků, které při Spearmanově korelaci r_{Spear} dosáhly statisticky nejrelevantnější korelace se škálou UPDRS III. Řazeny jsou podle dosažené hladiny významnosti p_{Spear} . Pouze minimální hodnota krátkodobé energie signálu (min STE) a minimální hodnota počtu průchodů signálu nulovou úrovní (min ZCR) mají hladinu významnosti nižší než je mezní hodnota 0,05. Pro tyto dva příznaky jsou vykresleny korelační diagramy viz Obr. 7.1 a Obr. 7.2.

Tab. 7.3 obsahuje výpis deseti příznaků, které při Pearsonově korelaci r_{Pears} dosáhly statisticky nejrelevantnější korelace se škálou UPDRS III. Řazeny jsou podle dosažené hladiny významnosti p_{Pears} . Hladinu významnosti nižší než hodnota 0,05 mají čtyři příznaky – min ZCR, min STE, střední hodnota Teager-Kaiserova energetického operátoru (mean TKEO) a minimální hodnota základního kmitočtu F_0 . Při porovnání obou tabulek je patrné, že se příznaky s nejvyšší mírou korelace pro oba typy téměř shodují. Je tedy možné říci, že s velkou pravděpodobností existuje lineární vztah mezi těmito příznaky a hodnotami UPDRS III škály.

Tab. 7.2: Seznam deseti příznaků s nejvyšší mírou korelace řazených podle hodnoty dosažené hladiny významnosti u Spearmanovy korelace.

příznak	r_{Spear}	p_{Spear}
min STE	-0.4146	0.0001
min ZCR	0,3355	0,0007
NHR nh	-0,1898	0,0598
mean F_0	-0,1851	0,0666
median F_0	-0,1728	0,0871
min F_0	-0,1680	0,0965
std F_1	0,1665	0,0995
mean F_1	-0,1566	0,1215
mean TKEO	-0,1463	0,1484
min F_2	-0,1427	0,1587

r_{Spear} – Spearmanův korelační koeficient, p_{Spear} – hladina významnosti r_{Spear}

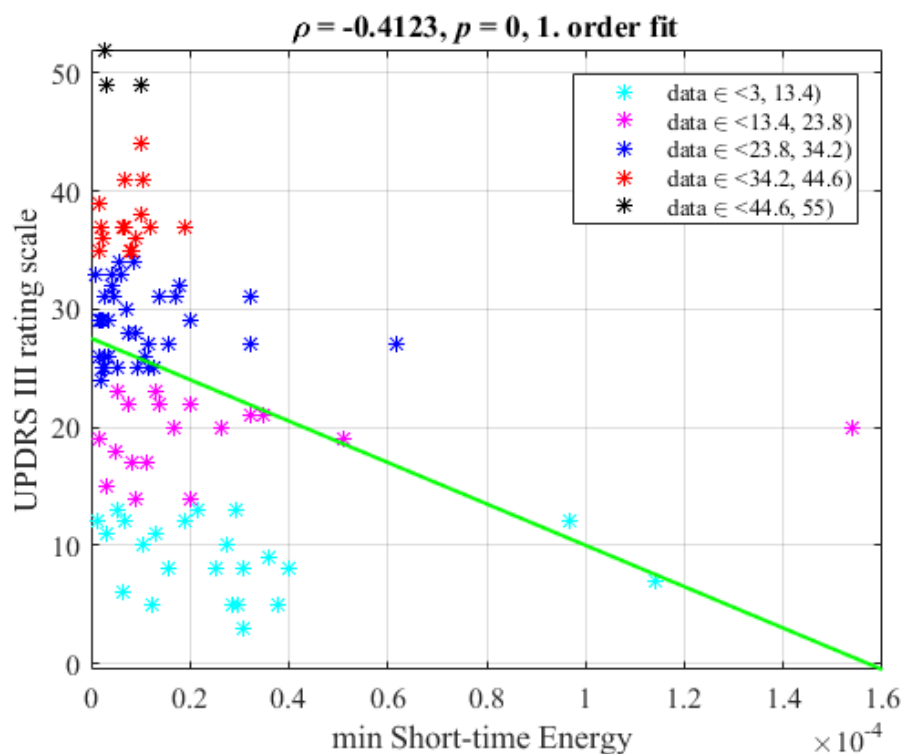
Tab. 7.3: Seznam deseti příznaků s nejvyšší mírou korelace řazených podle hodnoty dosažené hladiny významnosti u Pearsonovy korelace.

příznak	r_{Pears}	p_{Pears}
min ZCR	0,2866	0,0040
min STE	-0,2433	0,0152
mean TKEO	-0,2043	0,0425
min F_0	-0,2038	0,0431
mean F_0	-0,1964	0,0513
median F_0	-0,1870	0,0638
NHR nh	-0,1822	0,0711
max F_1	0,1777	0,0784
max F_0	-0,1586	0,1169
NHR hn	0,1570	0,1208

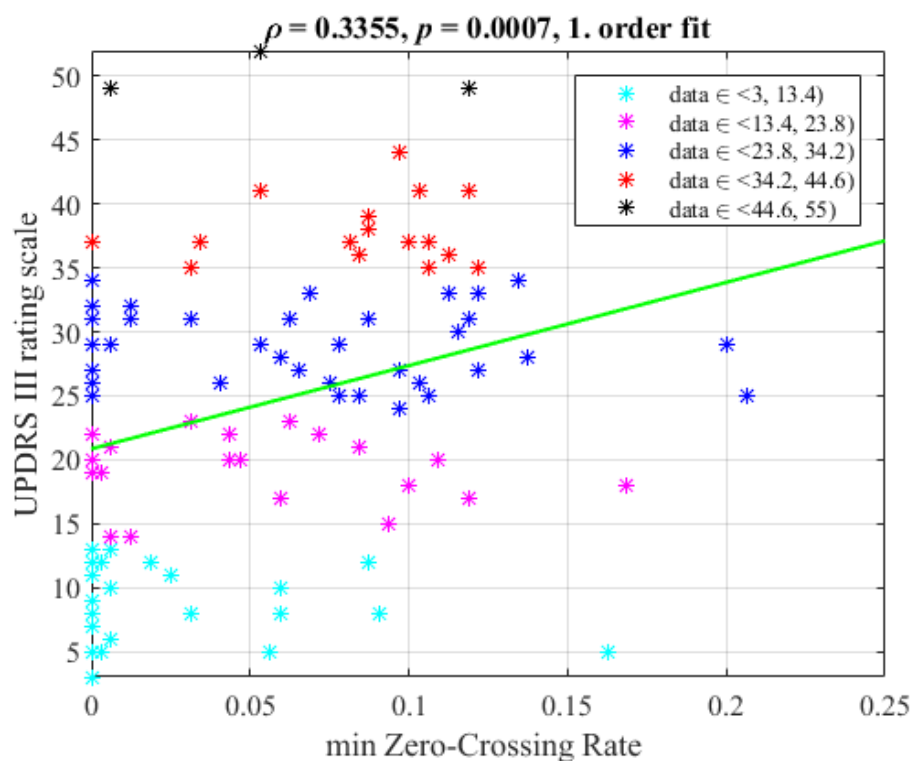
r_{Pears} – Pearsonův korelační koeficient, p_{Pears} – hladina významnosti r_{Pears}

Hodnoty korelačního koeficientu r_{Spear} i r_{Pears} pro min STE mají zápornou hodnotu, z čehož vyplývá, že se zvyšující se hodnotou UPDRS III škály a tedy se zhoršujícími se projevy HD klesá minimální hodnota krátkodobé energie. Tento závěr

se shoduje s teoretickými poznatky. Pacienti s PN nedokáží plně ovládat bránici a svaly hrudníku, které se podílejí na síle a kvalitě respirace. Důsledkem toho není v plicích dostatek vzduchu, výdech a jeho proudění přes hlasivky je při promluvě slabší, než u zdravého jedince a tím je i nižší energie signálu. Vliv má také nižší intenzita kmitání hlasivek. Klinicky je u parkinsoniků pozorována hypofonie - snížená intenzita hlasu. S energií signálu také souvisí střední hodnota Teager-Kaiserova energetického operátoru. Příznak min ZCR má kladnou hodnotu korelačního koeficientu. Minimální počet průchodů signálu nulovou úrovní roste s hodnotou UPDRS III škály. Tento parametr indikuje pravděpodobnost výskytu nadbytečného šumu v hlase, který se projevuje jako hlasová dysfonie. Více šumu pak znamená větší počet průchodů signálu nulovou úrovní. Základní tón hlasu F_0 vykazuje dobré výsledky při diferenciaci zdravé a dysartrické řeči. Korelační koeficienty pro tento příznak vychází záporné a tedy hodnota mean F_0 by měla s postupem onemocnění klesat. Příčinou může být, že při zvyšující se hodnotě UPDRS III a tedy zhoršujícím se zdravotním stavu pacienta, se vlivem svalové rigidity zhoršuje schopnost hlasivek periodicky kmitat. U pokročilejších stádií HD pak hlasivky kmitají pomaleji.



Obr. 7.1: Graf znázorňující míru korelace příznaku min STE s UPDRS III.



Obr. 7.2: Graf znázorňující míru korelace příznaku min ZCR s UPDRS III.

Korelační diagramy na Obr. 7.1 a Obr. 7.2 znázorňují korelaci mezi dvěma příznaky s hodnotou $p_{\text{Spear}} < 0,05$ a hodnotou UPDRS III škály, tedy mezi daty paraklinickými a klinickými, pomocí regresní křivky. K jejich vytvoření byl použit Software pro vizualizaci dyzartrické řeči, který je volně dostupný na webu www.splab.cz [33].

7.5 Regresní analýza

Požadovaným výstupem této práce je natrénovaný model, který je, na základě vstupních hodnot řečových příznaků, schopen určit (předpovědět) hodnotu UPDRS III škály řečníka s co největší přesností. Pokud bychom se dotazovali na to, zda je daný řečník pacient s PN nebo se jedná o zdravého člověka, použili bychom algoritmus založený na klasifikaci. Protože nás ale zajímá konkrétní hodnota z určité škály (hodnota UPDRS III škály se může pohybovat od 0 po 108), je třeba volit algoritmus regresní. Pro účely této práce je použit algoritmus CART, který je blíže popsán v kapitole 5.2.1. Na vstupu trénovacího algoritmu je množina 39 příznaků, jejich seznam a popis jsou uvedeny v příloze B. V první fázi je model trénován, validován a hodnocen pro každý příznak zvlášť. Dále vycházíme z předpokladu, že jeden příznak nestačí k odhadu výstupní hodnoty s uspokojivou přesností, je tedy třeba zvolit kombinaci více příznaků. V další fázi je proto použita metoda SFFS (Sequential Floating Feature Selection) pro výběr podmnožiny příznaků, jejichž kombinace dává

nejpřesnější odhad výstupní hodnoty. Model je znovu natrénován, tentokrát je na vstupu podmnožina příznaků vybraná pomocí SFFS, a následně je validován. Jelikož nebyl k dispozici dostatečný počet nahrávek k tomu, aby mohly být rozděleny na trénovací, testovací a validační data, je pro validaci se v obou případech používá metoda 10-fold cross-validace s pěti opakováními, která umožňuje efektivně ohodnotit kvalitu naučeného algoritmu na omezeném vzorku dat. Princip fungování cross-validace je podrobněji popsán v kapitole 5.2.2.

7.5.1 Individuální analýza

V rámci individuální analýzy je validován potenciál jednotlivých příznaků (individuálně) odhadovat hodnotu UPDRS III škály. Trénování modelu jednotlivými příznaky je stejně jako korelace řešeno ve skriptu **feature_analysis.m**. Vstupem jsou hodnoty příznaků pro 99 zkoumaných nahrávek. Celý proces trénování a cross-validace má pro každý příznak $N=5$ opakování s tím, že při 2. až 5. opakování je pořadí hodnot příznaků zpřeházeno pomocí vygenerovaného vektoru náhodných čísel. Dosáhneme tak větší robustnosti natrénovaného modelu a jeho menší závislosti na trénovacích datech.

Vstupní množinu 99 hodnot pro každý příznak rozdělíme na $K=10$ podmnožin – skupin. Protože počet nahrávek nemusí být (a v tomto případě není) dělitelný K beze zbytku, je vytvořeno $K-1$ skupin se stejným počtem hodnot a K -tá skupina obsahuje počet hodnot n , kdy:

$$n = (\text{počet nahrávek}) \bmod K.$$

Jedna skupina vždy slouží jako testovací a zbylých 9 jako trénovací.

K trénování modelu je použita funkce prostředí MATLAB **fitrtree**. Jedná se o binární regresní rozhodovací strom. Vstupními parametry jsou trénovací data a jim odpovídající známé hodnoty UPDRS III škály. Kvalita natrénovaného modelu je následně ověřena za použití funkce **predict**, jejímž vstupem je konkrétní natrénovaný model a testovací data. Pro každý příznak na konci existuje pět odhadnutých hodnot škály UPDRS III, ze kterých je vypočtena střední absolutní chyba odhadu, střední kvadratická chyba odhadu a jejich směrodatná odchylka. Ty pak slouží jako ukazatel míry úspěšnosti či přesnosti modelu při hodnocení neznámých dat. Dále je vypočtena hodnota Estimated Error, která udává procentuální úspěšnost modelu. Je nutno poznamenat, že i když má škála UPDRS III rozsah hodnot 0 až 108, v této práci jsou modely trénovány jen na rozsah 0 až 55. Data pro pacienty s klinicky zjištěnou hodnotou UPDRS III vyšší než 55 nebyly pro tuto práci k dispozici. Rozsah 0 až 55 je taktéž použit při výpočtu hodnoty EER.

Tab. 7.4: Seznam deseti příznaků, které při individuální analýze dosahují nejlepších výsledků odhadu UPDRS III škály.

příznak	MAE průměr±std	RMSE průměr±std	EER [%]
min ZCR	9,84±0,34	11,96±0,24	17,89±0,62
max F1	10,18±0,31	12,92±0,36	18,51±0,56
min STE	10,30±0,33	12,82±0,34	18,73±0,55
mean F2	10,35±0,46	12,86±0,36	18,82±0,84
jitter local	10,70±0,27	13,27±0,37	19,46±0,49
NHR nh	10,76±0,37	12,83±0,41	19,56±0,67
MPSD	10,89±0,47	13,81±0,55	19,80±0,85
max TKEO	11,09±0,50	13,70±0,37	20,17±0,91
std TKEO	11,10±0,41	14,29±0,45	20,19±0,75
min F1	11,16±0,59	14,13±0,81	20,30±1,07

MAE – střední absolutní chyba, RMSE – střední kvadratická chyba, std – směrodatná odchylka, EER – Estimation Error

Jak je patrné z Tab. 7.4, příznakem, který při individuálním odhadu UPDRS III škály dosahuje nejvyšší přesnosti, je minimální počet průchodů signálu nulovou úrovní (min ZCR). Chybně hodnotu UPDRS III určí jen v 17,89% případů. Čísla v tabulce mají jasnou vypovídající hodnotu, na dosažené výsledky se dá ale dívat i jinak a to z pohledu popisu konkrétních příznaků. V tabulce se objevuje pět příznaků popisujících fonaci, konkrétně jsou to:

- lokální kolísání základní periody T_0 , vyjadřující kmitočtovou nestabilitu signálu, bývá také označován jako třes či chvění hlasivek
- minimální hodnota krátkodobé energie, která vypovídá o intenzitě signálu, a z teorie víme, že u znělých hlásek, tedy hlásek u kterých hlasivky kvaziperiodicky kmitají, má větší hodnotu než u hlásek neznělých – šumových
- maximální hodnota a směrodatná odchylka Teager-Kaiserova energetického operátoru, který taktéž popisuje energii signálu, tentokrát však (na rozdíl od STE) nesegmentovaného
- Medián spektrální výkonové hustoty udávající výkon signálu ve zvoleném frekvenčním pásmu (závisí na vzorkovací frekvenci)

Dále tabulka obsahuje tři příznaky popisující rezonanci (formanty), konkrétně maximální a minimální hodnotu F_1 a střední hodnotu F_2 . První a druhý formant nejvíce vypovídají o produkci samohlásek. Z příznaků hodnotících kvalitu signálu, se mezi deseti příznaky s největší individuální přesností při odhadu UPDRS III škály objevují minimální hodnota počtu průchodů signálu nulovou úrovní a odstup signálu od šumu.

Pokud bychom podle těchto příznaků zpětně skládali řečový signál a dali ho do souvislosti s hodnotami korelace z předchozí kapitoly a známými klinickými projevy PN, vyznačoval by se vlivem kolísání základní periody špatnou srozumitelností samohlásek, o čemž svědčí i fakt, že jsou ovlivněny první a druhý formant. Dále by byl takovýto řečový projev nejspíš tichý, což souvisí s nízkou energií signálu a nízkým výkonem, které s postupem nemoci více a více klesají. Vlivem nedokonalého kmitání hlasivek, by takovýto signál obsahoval více šumu, u kterého je známo, že má vyšší hodnotu počtu průchodů signálu nulovou úrovní než signál znělý. Stejně tak lze s tímto jevem dát do souvislosti horší odstup signálu od šumu. Z tohoto příkladu vyplývají dva důležité poznatky. Zaprvé, jednotlivé příznaky je možné a při interpretaci výsledků důležité dát do kontextu s klinickými příznaky. Zadruhé, jednotlivé příznaky spolu mohou korelovat a dva a více příznaků tak mohou popisovat ten jistý projev dysartrické řeči.

7.5.2 Selektce příznaků

Výsledkem parametrizace je rozsáhlá množina příznaků, přičemž některé z nich mohou být redundantní nebo mají nedostatečný potenciál pro popis hypokinetické dysartrie. Pro výběr optimální skupiny příznaků, která bude při odhadu hodnot UPDRS III škály dosahovat největší přesnosti je v prostředí MATLAB implementována metoda selektce SFFS (Sequential Forward Floating Selection). Tato metoda vybírá skupiny příznaků a hodnotí jejich souvislost s analyzovanými daty. Výhodou je její přesnost, nevýhodou vyšší výpočetní náročnost. SFFS používá kombinaci klasifikátoru CART s cross-validací. Úspěšnost jednotlivých skupin příznaků je hodnocena podle hodnoty střední absolutní chyby. Předvýběr příznaků pomocí metody mRMR nebyl prováděn, jelikož množina 39 příznaků je sama o sobě dostatečně malá.

Funkce **SFFS_reg.m** implementující metodu selektce SFFS použitou pro výběr nejvhodnějších příznaků, je součástí Software pro sekvenční selekci příznaků, který je volně dostupný z [34]. Funkce je volána ve skriptu **multiple_features.m**, kde jsou také nastavovány její volitelné parametry, které budou podrobněji popsány dále. Vstupem funkce je matice hodnot *feat_matrix*, kdy jednotlivé sloupce představují příznaky a jednotlivé řádky hodnoty těchto příznaků. Hodnoty v této matici jsou normalizovány podle:

$$NHP = \frac{parametr - mean(parametr)}{std(parametr)}$$

Kde *NHP* je normalizovaná hodnota příznaku, *mean* je odhad střední hodnoty a *std* je odhad směrodatné odchylky. Dále pak hodnoty UPDRS III a volitelné vstupní parametry. Ty jsou voleny standardně stejně jako u individuální analýzy:

- 10-fold cross validace
- počet opakování: 5
- klasifikátor: algoritmus CART

```

INPUT VARIABLES
training matrix:      (OK): specified
training labels:      (OK): specified

OPTIONAL VARIABLES
Cross validations:    (OK): 5
K-fold                (OK): 10-fold
Classifier:           (OK): cart
Scoring method:       (OK): mae
Added features:       (OK): 3
Max. num. features:   (OK): 39 (AUTOMATIC)
Plot var. progress:   (OK): 0 (DEFAULT)

```

Obr. 7.3: Vstupní nastavení použité při inicializaci funkce `SFFS_reg.m`

Důležitý je parametr t , který zajišťuje, že při sestavování skupin příznaků nejsou nalezena pouze lokální maxima. Umožňuje ze skupiny zpětně některý vyloučit a objevit tak lepší kombinaci příznaků – globální maximum. Tento parametr je nastaven na hodnotu 3. Měřítkem, podle kterého funkce hodnotí úspěšnost odhadu při trénování modelu jedním příznakem a následně kombinací více příznaků, je střední absolutní chyba (MAE). Nastavení jednotlivých parametrů je parné z výpisu funkce `SFFS_reg.m` na Obr. 7.3. Výstup funkce je uložen do souboru **CART_multiple_features.mat**.

```

SFFS(0) : 1: SCR = 11.1931 +- 2.5594 MAE = 11.1931 +- 2.5594 RMSE = 13.981
SFFS(0) : 4: SCR = 10.9242 +- 2.4445 MAE = 10.9242 +- 2.4445 RMSE = 13.324
SFFS(0) : 7: SCR = 10.0469 +- 2.2519 MAE = 10.0469 +- 2.2519 RMSE = 12.283
SFFS(0) : 12: SCR = 9.625 +- 2.1914 MAE = 9.625 +- 2.1914 RMSE = 11.4883 +
SFFS(1) : 1: SCR = 11.268 +- 2.0901 MAE = 11.268 +- 2.0901 RMSE = 13.5414 .
SFFS(1) : 2: SCR = 10.6984 +- 2.982 MAE = 10.6984 +- 2.982 RMSE = 13.2224 .
SFFS(1) : 7: SCR = 10.2019 +- 2.2532 MAE = 10.2019 +- 2.2532 RMSE = 12.659
SFFS(1) : 9: SCR = 9.6474 +- 2.5352 MAE = 9.6474 +- 2.5352 RMSE = 12.2431 .
SFFS(2) : 1: SCR = 11.1913 +- 2.424 MAE = 11.1913 +- 2.424 RMSE = 13.9729 .
SFFS(2) : 2: SCR = 11.1367 +- 2.2323 MAE = 11.1367 +- 2.2323 RMSE = 13.982
SFFS(2) : 3: SCR = 10.8504 +- 2.3773 MAE = 10.8504 +- 2.3773 RMSE = 13.413
SFFS(2) : 4: SCR = 10.6138 +- 2.6375 MAE = 10.6138 +- 2.6375 RMSE = 13.159
SFFS(2) : 7: SCR = 9.9475 +- 2.2116 MAE = 9.9475 +- 2.2116 RMSE = 12.1982 .
SFFS(2) : 14: SCR = 9.625 +- 2.7944 MAE = 9.625 +- 2.7944 RMSE = 12.4712 +

```

Obr. 7.4: Zobrazení výpisu funkce `SFFS_reg.m` při výpočtu nejvhodnější kombinace příznaků

Metodou selekce SFFS byly ze 39 testovaných příznaků vybrány 4, při jejichž kombinaci dosahuje regresní model největší přesnosti odhadu. Seznam příznaků je v tabulce Tab. 7.5. Ve vybrané skupině je téměř rovnoměrné zastoupení všech skupin příznaků popsanych v kapitole 4. Dva příznaky – schimm localdb a schimm apq11 popisující fonaci, jeden příznak – max F_1 popisující rezonanci a taktéž jeden příznak – min ZCR hodnotící kvalitu signálu. Proces výpočtu je vidět na Obr. 7.4. V prvním kroku jsou vybrány nejlepší individuální příznaky (`SFFS(0)`), v dalších krocích (`SFFS(1)`, `SFFS(2)`,...) je testována přesnost systému při kombinování dvou, tří a více

příznaků. Z hodnot v Tab. 7.5 je jasné patrné, jak přidávání příznaků do konečné podmnožiny snižuje chybu odhadu UPDRS III. Chyba odhadu při použití všech čtyř příznaků je rovna $17,27 \pm 4,86$ %, při použití tří příznaků se už chyba zvyšuje na $18,34 \pm 4,34$ %, u dvou příznaků na $18,44 \pm 4,07$ % a při použití pouze prvního vybraného příznaku je chyba $18,71 \pm 3,79$ %.

Tab. 7.5: Příznaky vybrané metodou SFFS, výsledky regrese a korelace při použití těchto příznaků

Pořadí	Příznak	Kombinace	MAE $\pm$ std	RMSE $\pm$ std	EER $\pm$ std [%]	$r_{Spear}\pm$ std	$p_{Spear}\pm$ std
1	min ZCR	1	9,73 $\pm$ 1,97	11,66 $\pm$ 2,34	18,71 $\pm$ 3,79	0,27 $\pm$ 0,28	0,41 $\pm$ 0,30
2	schimm localdb	1,2	9,59 $\pm$ 2,11	11,56 $\pm$ 2,30	18,44 $\pm$ 4,07	0,39 $\pm$ 0,28	0,32 $\pm$ 0,30
3	max F1	1,2,3	9,54 $\pm$ 2,55	11,65 $\pm$ 2,44	18,34 $\pm$ 4,34	0,40 $\pm$ 0,33	0,29 $\pm$ 0,29
4	schimm apq11	1,2,3,4	8,98 $\pm$ 2,53	10,94 $\pm$ 2,95	17,27 $\pm$ 4,86	0,51 $\pm$ 0,25	0,21 $\pm$ 0,24

Kombinace – kombinace příznaků, MAE – střední absolutní chyba, RMSE – střední kvadratická chyba, std – směrodatná odchylka, EER – Estimation Error, r_{Spear} – Spearmanův korelační koeficient, p_{Spear} – hladina významnosti r_{Spear}

8 ZÁVĚR

Tato práce se zabývá odhadem progresu Parkinsonovy nemoci pomocí akustické analýzy řeči. V jejím rámci proto byla prostudována a v kapitole 1 popsána funkce řečového ústrojí člověka, jeho jednotlivých částí a jejich vlivu na výsledný hlas. Hypokinetická dysartrie u pacientů s Parkinsonovou nemocí je důsledkem poruchy jednotlivých složek řeči, fonace, artikulace, prozodie a plynulosti. Jednotlivé složky byly popsány v kapitole 2 a v kontextu hypokinetické dysartrie u nich bylo uvedeno, jak ovlivňují mluvený projev. V kapitole 4 byly podrobně rozebrány řečové příznaky. Jednotlivé příznaky byly rozděleny do skupin podle toho, jakou vlastnost (fonace, rezonance, kvalita signálu) řečového signálu popisují.

Cílem práce bylo na základě těchto teoretických znalostí navrhnout systém pro odhad progresu Parkinsonovy nemoci pomocí akustické analýzy řeči. Jedná se o neinvazivní metodu, která se od běžných klinických vyšetření, kdy diagnózu určuje lékař, liší především svojí objektivitou, nezávislostí na lidském faktoru, rychlostí a v neposlední řadě nízkými náklady. Její nevýhodou je značná závislost na správnosti implementace výpočetních metod. Velkým problémem je v dnešní době také nedostatek kvalitních a dostatečně rozsáhlých databází řečového projevu pacientů s Parkinsonovou nemocí. V této práci je použita databáze PARCZ, která byla poskytnuta Fakultní nemocnicí u sv. Anny v Brně. Konkrétně její část s nahrávkami prodloužené fonace hlásky [a], která obsahuje nahrávky 152 řečníků. V rámci práce byly zpracovávány pouze nahrávky pacientů s Parkinsonovou nemocí, kterých je celkem 100. Subjektivní hodnocení pokročilosti nemoci u těchto pacientů je reprezentováno hodnotou UPDRS III škály, kterou stanovuje lékař.

Praktická část práce byla realizována v prostředí MATLAB. Popisu návrhu systému se věnuje kapitola 6 a jeho praktická realizace je podrobně popsána v kapitole 7. Byly vytvořeny funkce pro předzpracování řečového signálu a pro výpočet některých řečových příznaků. Výpočet zbylých příznaků byl proveden pomocí volně dostupného programu Praat [32]. Prvním krokem je načtení jednotlivých nahrávek řečového cvičení do systému, jejich předzpracování a následná segmentace. Poté jsou extrahovány lokální příznaky a statistickými metodami vypočteny vysokoúrovňové příznaky. Celkový počet příznaků použitých v této práci je 39.

Pro všechny příznaky je počítána Spearmanova a Pearsonova korelace se škálou UPDRS III. Hodnoty korelačních koeficientů a především pak dosažená hladina významnosti těchto koeficientů dávají představu o tom, které řečové příznaky mají lineární souvislost (přímou nebo nepřímou) s hodnotami UPDRS III škály a lze u nich tedy očekávat dobré výsledky, při jejím odhadu. Prvních deset příznaků, které dosáhly statisticky nejrelevantnější korelace se škálou UPDRS III se pro Spearmanovu a

Pearsonovu korelaci prakticky neliší. U Spearmanovy korelace dosáhly hladiny významnosti nižší než hodnota 0,05 dva příznaky – min STE s hodnotou korelačního koeficientu -0,4146 na hladině významnosti 0,0001 a min ZCR s hodnotou 0,3355 na hladině významnosti 0,0007. U Pearsonovy korelace tuto podmínku splnily čtyři příznaky - min ZCR s hodnotou 0,2866 na hladině významnosti 0,0040, min STE s hodnotou korelačního koeficientu -0,2433 na hladině významnosti 0,0152, mean TKEO s hodnotou korelačního koeficientu -0,2043 na hladině významnosti 0,0425 a min F_0 s hodnotou korelačního koeficientu -0,2038 na hladině významnosti 0,0431.

V dalším kroku byl v rámci individuální analýzy validován potenciál jednotlivých příznaků (individuálně) odhadovat hodnotu UPDRS III škály. Při trénování modelu byla použita regresní analýza implementovaná pomocí algoritmu CART. Model byl vždy natrénován pomocí jednoho příznaku a k jeho ověření byla použita 10-fold cross-validace s pěti opakováními. Cross-validace byla použita z důvodu malého rozsahu databáze nahrávek. Nejlepších výsledků při individuální analýze dosáhl příznak min ZCR, schopný odhadnout hodnotu UPDRS III se 17,89% chybovostí. Mezi deseti nejúspěšnějšími parametry byly zastoupeny všechny typy řečových příznaků popsaných v kapitole 4. Příznaky popisující fonaci – jitter local, min STE, max a std TKEO a MPSD. Příznaky popisující rezonanci – max a min F_1 a mean F_2 . Příznaky hodnotící kvalitu signálu – HNR a výše zmíněný min ZCR.

Poslední část práce se zabývá výběrem optimální skupiny příznaků, která bude při odhadu UPDRS III škály dosahovat největší přesnosti. Za tímto účelem je v prostředí MATLAB implementována metoda selekce příznaků SFFS (Sequential Forward Floating Selection). Použitá funkce je součástí Software pro sekvenční selekci příznaků, který je volně šiřitelný [34]. Výstupem selekce jsou 4 příznaky – min ZCR, schimm localdb, max F_1 a schimm apq11. Regresní model natrénovaný pomocí těchto příznaků dosahuje při odhadu chybovosti $17,27 \pm 4,86\%$. Tato hodnota může být považována za nedostačující, avšak je třeba vzít v úvahu, že parametry v této práci jsou konvenční a lineární. Výhodou takovýchto parametrů je snadná interpretovatelnost, tedy možnost přesně popsat se kterou složkou řeči souvisí. Pro lepší popis hypokinetické dysartrie by bylo třeba použít nelineární dynamické příznaky, založené například na teorii chaosu, empirické modální dekompozici apod. Takovéto parametry však není možné klinicky interpretovat a z medicínského hlediska jsou tak v současné době nevhodné.

LITERATURA

- [1] PARKINSON, J. An essay on the shaking palsy; *J Neuropsychiatry Clin Neurosci*, 2002. 14(2): 223–236.
- [2] ZAMIŠKOVÁ, G., RESSNER, P., DLOUHÁ, J., ŠIGUTOVÁ, D. Poruchy řeči u Parkinsonovy nemoc. *Neurologie pro praxi*, 2010. 11(2): 112–116.
- [3] TITZE, I. R. *Principles of voice production*; Prentice Hall, 1994. ISBN 9780137178933.
- [4] HARTELIUS, L., SVENSSON, P. Speech and Swallowing Symptoms Associated with Parkinson's Disease and Multiple Sclerosis: a Survey. *Folia Phoniatr Logop*, 1994. ročník 46, č. 1, s. 9–17.
- [5] RAMIG, L. O., FOX, C., SAPIR, S. Speech treatment for Parkinson's disease. *Expert Rev Neurother*, 2008. ročník 8, č. 2. s. 297–309
- [6] GOBERMAN, A. M. Correlation between acoustic speech characteristics and non-speech motor performance in Parkinson Disease. *Med Sci Monit*, 2005. ročník 11, č. 3, s. CR109–116.
- [7] GOBERMAN, A. M., COELHO, C. Acoustic analysis of parkinsonian speech I: speech characteristics and L-Dopa therapy. *NeuroRehabilitation*, ročník 17, č. 3, 2002: s. 237–246
- [8] SHAO, J., MACCALLUM, J. K., ZHANG, Y., SPRECHER, A., JIANG, J. J. Acoustic analysis of the tremulous voice: Assessing the utility of the correlation dimension and perturbation parameters. *J Commun Disord*, ročník 43, 2010: s. 35–44.
- [9] DOYLE, P., RAADE, A., PIERRE, A. S., DESAI, S. Fundamental Frequency and Acoustic Variability Associated with Production of Sustained Vowels by Speakers with Hypokinetic Dysarthria. *Journal of Medical Speech-Language Pathology*, ročník 3, 1995: s. 41–50.
- [10] CANTER, G. J. Speech Characteristics of Patients with Parkinson's Disease: I. Intensity, Pitch, and Duration. *J Speech Hear Disord*, ročník 28, č. 3, 1963: s. 221–229.
- [11] HOLMES, R. J., OATES, J. M., PHZLAND, D. J., HUGHES, A. J. Voice Characteristics in the Progression of Parkinson's Disease. *Int J Lang Commun Disord*, ročník 35, č. 3, 2000: s. 407–418.
- [12] METTER, E. J., HANSON, W. R. Clinical and Acoustical Variability in Hypokinetic Dysarthria. *J Commun Disord*, ročník 19, č. 5, 1986: s. 347–366.

- [13] KOŠŤÁLOVÁ, M., BEDNARČÍK, J., MECHL, M., VOHÁŇKA, S., ŠNÁBL, I. *Multimediální výukový atlas poruch řeči a příbuzných kognitivních funkcí*. Brno: Masarykova univerzita 2004: 128.
- [14] LOEVE, R.J., WEBB, W. G. *Možek a řeč*. Praha: Portál 2009, s.372.
- [15] MEKYSKA, J. *Analýza řečových promluv pro IT diagnostiku neurologických onemocnění*: dizertační práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2014. 160 s.
- [16] LOGEMAN, J. A., FISCHER, H. B., BOSCHES, B., BLONSKY, E. Frequency and Cooccurrence of Vocal Tract Dysfunctions in the Speech of a Large Sample of Parkinson Patients. *J Speech Hear Disord*, ročník 43, č. 1, 1978: s. 47–57.
- [17] MCNEIL, M. R., ROSENBECK, J. C., ARONSON, A. E. The Dysarthrias: Physiology, Acoustics, Perception, Management. *Taylor & Francis Ltd*, 1985.
- [18] ZWIRNER, P., BARNES, G. J. Vocal Tract Steadiness: a Measure of Phonatory and Upper Airway Motor Control during Phonation in Dysarthria. *J Speech Hear Res*, ročník 35, č. 4, 1992: s. 761–768.
- [19] ČERNOCKÝ, J. *Zpracování řečových signálů*. Ústav počítačové grafiky a multimédií, FIT, VUT Brno, 2006.
- [20] PSUTKA, J. et al. *Mluvíme s počítačem česky*. Praha: ACADEMIA, 2006. 752 s. ISBN 80-200-1309-1.
- [21] SMÉKAL, Z. *Číslicové zpracování řeči (MZPR)*. Elektronická skripta pro magisterská studia, VUT Brno, 2010.
- [22] SMÉKAL, Z. *Číslicové zpracování signálů (MCSI)*. Elektronická skripta pro magisterská studia, Ústav telekomunikací, FEEC, VUT Brno, 2009.
- [23] SIGMUND, M. *Rozpoznávání řečových signálů: přednášky*. 1. vyd. Brno: VUT FEKT, ústav radioelektroniky, 2007, 122 s. ISBN 978-80-214-3526-1.
- [24] LITTLE, M., MCSHARRY, P., HUNTER, E., SPIELMAN, J., RAMIG, L. Suitability of Dysphonia Measurements for Telemonitoring of Parkinson's Disease. *IEEE T Bio-Med Eng*, ročník 56, č. 4, 2009, s. 1015–1022.
- [25] LU, L., ZHANG, H., JIANG, H. *Content analysis for audio classification and segmentation*. ISBN 10.1109/tsa.2002.804546.
- [26] KOHAVI, R., JOHN, G. H. Wrappers for feature subset selection. *Artificial Intelligence*, ročník 97, č. 1-2, (December 1997), s. 273-324.
- [27] PENG, H., LONG, F., DING, C. Feature Selection Based on Mutual Information: Criteria of Max-dependency, Max-relevance, and Min redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions*

on, ročník 27, č. 8, s. 1226–1238, Aug. 2005.

- [28] KOUTECKÁ, J. *Vývoj motorických funkcí u pacientů s Parkinsonovou chorobou při opakovaném vyšetření*. Katedra fyziologie a rehabilitace LF MU, 2008.
- [29] MEKYSKA, J., GALÁŽ, Z., MŽOUREK, Z., SMÉKAL, Z., REKTOROVÁ, I., ELIÁŠOVÁ, I., KOŠTÁLOVÁ, M., MRAČKOVÁ, M., BERANKOVA, D., FAÚNDEZ ZANUY, M., LOPEZ-DE-IPINA, K., ALONSO-HERNANDEZ, J. Assessing progress of Parkinson' s disease using acoustic analysis of phonation. In *2015 4th International Work Conference on Bioinspired Intelligence (IWOBI)*. 2015. s. 111-118. ISBN: 978-1-4673-7845-1.
- [30] CART Algorithm [online]. Dostupné z URL: <ftp://public.dhe.ibm.com/software/analytics/spss/support/Stats/Docs/Statistics/Algorithms/13.0/TREE-CART.pdf>
- [31] BURGET, R. *Teoretická informatika* [online]. Ústav telekomunikací, FEEC, VUT Brno, 20013. Dostupné z URL: <https://itmulti.cz/courses/get-file-fs?dir=archiveDir&realName=MTIN%2FSkripta+pro+p%C5%99edn%C3%A11%C5%A1ky%2FMTIN.pdf>
- [32] BOERSMA P., WEENINK D. Praat: doing phonetics by computer [Online]. [cit. 2015-05-08]. Dostupné z URL: <http://www.fon.hum.uva.nl/praat/>
- [33] MEKYSKA J., GALÁŽ Z., MŽOUREK Z., SMÉKAL Z., REKTOROVÁ I. Dysarthric speech visualisation tool [online]. Dostupné z URL: <http://splab.cz/en/download/software/software-pro-vizualizaci-dyarthricke-reci>
- [34] MEKYSKA J., GALÁŽ Z., MŽOUREK Z., SMÉKAL Z. Sequential floating feature selection tool [online]. Dostupné z URL: <http://splab.cz/en/download/software/software-pro-sekvencni-selekci-priznaku>

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

$\mathbb{R}$	množina reálných čísel
APQ3	three-point Amplitude Perturbation Quotient
APQ5	five-point Amplitude Perturbation Quotient
APQ11	11-point Amplitude Perturbation Quotient
DFT	diskrétní Fourierova transformace
E	Energie signálu
EER	Estimation Error $EER = \frac{MAE}{rozsah\ UPDSR\ III} * 100$
FFT	rychlá Fourierova transformace
HNR	Harmonic-to-noise Ratio
mRMR	Minimum Redundancy Maximum Relevance
MPSD	medián výkonové spektrální hustoty
MAE	střední absolutní chyba
NHR	Noise-to-Harmonic Ratio
PARCZ	Czech Parkinsonian Speech Database
PPQ5	five-point Pitch Perturbation Quotient
PSD	výkonová spektrální hustota
RMSE	střední kvadratická chyba
STD	Směrodatná odchylka
STE	krátkodobá energie
SFFS	Sequential Forward Floating Search – sekvenční dopředná selekce příznaků
TKEO	Teager-Kaiserův energetický operátor
UPDRS	Unified Parkinson's Disease Rating Scale – unifikovaná škála pro hodnocení Parkinsonovy nemoci
ZCR	počet průchodů signálu nulovou úrovní
α_1	koeficient preemfázového filtru

E	krátkodobá energie signálu
e	Eulerovo číslo $\left(e = \sum_{n=0}^{\infty} \frac{1}{n!} = 2.71828 \dots\right)$
F_0	základná frekvencia
f_{vz}	vzorkovacia frekvencia [Hz]
I	krátkodobá intenzita signálu
j	imaginárna jednotka
J	kolísání základní frekvence signálu
t	čas [s]
$W(e^{j\omega})$	spektrum váhovacích postupností
$w[n]$	váhovací postupnosť
$w[n - im]$	váhovací postupnosť posunutá o im vzorků
$X(e^{j\omega})$	spektrum vstupního signálu $x[n]$
$Y(e^{j\omega})$	spektrum segmentovaného signálu
π	Ludolphovo číslo $\left(\pi = 4 \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = 3.14159 \dots\right)$
ω	uhlová frekvencia [rad·s ⁻¹]
r_{Spear}	Spearmanův korelační koeficient
p_{Spear}	hladina významnosti Spearmanova korelačního koeficientu
r_{Pears}	Pearsonův korelační koeficient
p_{Pears}	hladina významnosti Pearsonova korelačního koeficientu

SEZNAM PŘÍLOH

A	OBSAH PŘILOŽENÉHO CD	59
B	SEZNAM PARAMETRŮ POUŽITÝCH PRO REGRESNÍ ANALÝZU	60

A OBSAH PŘILOŽENÉHO CD

Na přiloženém CD se nachází:

- Diplomová práce ve formátu .pdf
- Složka se zdrojovými soubory system_functions

B SEZNAM PARAMETRŮ POUŽITÝCH PRO REGRESNÍ ANALÝZU

příznak	jednotka	Popis příznaku
F_0 median	[Hz]	medián F_0
F_0 mean	[Hz]	střední hodnota F_0
F_0 std	[Hz]	směrodatná odchylka F_0
F_0 min	[Hz]	minimální hodnota F_0
F_0 max	[Hz]	maximální hodnota F_0
jitter local	[%]	lokální kolísání periody
jitter localabs	[%]	absolutní hodnota lokálního kolísání periody
jitter rap	[%]	průměrný rozdíl mezi periodou a průměrem dvou sousedních period
jitter ppq5	[%]	průměrný rozdíl mezi periodou a průměrem čtyř sousedních period
jitter ddp	[%]	průměrný rozdíl za sebou jdoucích rozdílů period
schimmer local	[%]	lokální kolísání amplitudy
schimmer localdb	[dB]	lokální kolísání amplitudy v dB
schimmer apq3	[%]	průměrný rozdíl mezi amplitudou a průměrem dvou sousedních amplitud
schimmer apq5	[%]	průměrný rozdíl mezi amplitudou a průměrem čtyř sousedních amplitud
schimmer apq11	[%]	průměrný rozdíl mezi amplitudou a průměrem deseti sousedních amplitud
schimmer dda	[%]	průměrný rozdíl za sebou jdoucích rozdílů amplitud
HNR nh	[—]	střední hodnota poměru signál-šum
HNR hn	[dB]	střední hodnota poměru signál-šum v dB
F_1 mean	[Hz]	střední hodnota F_1
F_1 std	[Hz]	směrodatná odchylka F_1
F_1 min	[Hz]	minimální hodnota F_1
F_1 max	[Hz]	maximální hodnota F_1
F_2 mean	[Hz]	střední hodnota F_2
F_2 std	[Hz]	směrodatná odchylka F_2

F_2 min	[Hz]	minimální hodnota F_2
F_3 max	[Hz]	maximální hodnota F_2
STE mean	[–]	medián energie
STE std	[–]	střední hodnota energie
STE min	[–]	minimální hodnota energie
STE max	[–]	maximální hodnota energie
MPSD	[–]	Medián spektrální výkonové hustoty
ZCR mean	[–]	střední hodnota počtu průchodů signálu nulovou úrovní
ZCR std	[–]	směrodatná odchylka počtu průchodů signálu nulovou úrovní
ZCR min	[–]	minimální hodnota počtu průchodů signálu nulovou úrovní
ZCR max	[–]	maximální hodnota počtu průchodů signálu nulovou úrovní
TKEO mean	[–]	střední hodnota Teager-Kaiserova operátoru energie
TKEO std	[–]	směrodatná odchylka Teager-Kaiserova energetického operátoru
TKEO min	[–]	minimální hodnota Teager-Kaiserova energetického operátoru
TKEO max	[–]	maximální hodnota Teager-Kaiserova energetického operátoru