VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

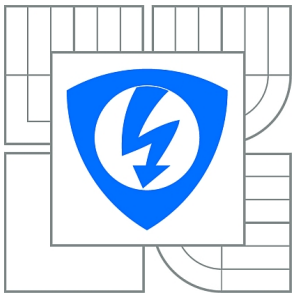
DIAGNÓZA PARKINSONOVY CHOROBY Z ŘEČOVÉHO SIGNÁLU

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE Bc. MICHAL KARÁSEK
AUTHOR

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

DEPARTMENT OF TELECOMMUNICATIONS

DIAGNÓZA PARKINSONOVY CHOROBY Z ŘEČOVÉHO SIGNÁLU

PARKINSON DISEASE DIAGNOSIS USING SPEECH SIGNAL ANALYSIS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

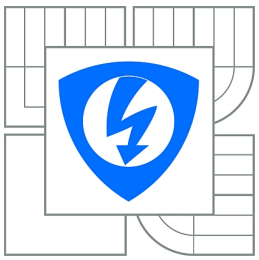
AUTOR PRÁCE — Bc. MICHAL KARÁSEK

AUTHOR

VEDOUCÍ PRÁCE — Ing. JIŘÍ MEKYSKA

SUPERVISOR

BRNO 2011



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky a komunikačních technologií

Ústav telekomunikací

Diplomová práce

magisterský navazující studijní obor
Telekomunikační a informační technika

Student:	Bc. Michal Karásek	***ID:***	70104
Ročník:	2	***Akademický rok:***	2010/2011

NÁZEV TÉMATU:

Diagnóza Parkinsonovy choroby z řečového signálu

POKYNY PRO VYPRACOVÁNÍ:

Cílem diplomové práce je rozbor aktuální problematiky analýzy řečových signálů, především pro účely diagnózy Parkinsonovy choroby, a návrh a implementace systému automatické diagnózy v prostředí MATLAB. Systém by měl využívat různé řečové příznaky a klasifikátor GMM. V případě, že nebude existovat dostačující množství trénovacích nahrávek, bude nad daty alespoň provedena analýza rozptylu (ANOVA).

DOPORUČENÁ LITERATURA:

[1] PSUTKA, Josef, et al. Mluvíme s počítačem česky. Praha: Academia, 2006. 752 s. ISBN 80-200-1309-1.
[2] HUANG, Xuedong; ACERO, Alex; HON, Hsiao-Wuen. Spoken Language Processing a Guide to Theory, Algorithm and System Development. Upper Saddle River: Prentice Hall PTR, 2001. 980 s. ISBN 0-13-022616-5.
[3] F. QUATIERI, Thomas. Discrete-Time Speech Signal Processing: Principles and Practice. Upper Saddle River: Prentice Hall PTR, 2001. 816 s. ISBN 978-0132429429.

Termín zadání:	7.2.2011	***Termín odevzdání:***	26.5.2011

Vedoucí práce: Ing. Jiří Mekyska

prof. Ing. Kamil Vrba, CSc.
Předseda oborové rady

UPOZORNĚNÍ:

ABSTRAKT

Práce se zabývá rozpoznáním Parkynsonovy choroby z řečového signálu. V první části poukazuje na základy řečových signálů a řečových signálů u pacientů postižených Parkinsonovou chorobou. Dále popisuje problematiku zpracování řečových signálů, základní příznaky používané k diagnóze Parkinsonovy choroby (např. VAI, VSA, FCR, VOT atd.) a redukci těchto příznaků. Další část je zaměřena na blokové schéma programu pro diagnózu Parkinsonovy choroby. Hlavním cílem této práce je porovnání dvou metod výběru příznaků (mRMR a SFFS). Pro klasifikaci byly vybrány dvě rozdílné metody. První metodou je klasifikace kNN a druhou metodou klasifikace jsou Gaussovy smýšené modely (GMM).

KLÍČOVÁ SLOVA

Řeč, analýza řečového signálu, formant, preemfáze, segmentace, příznak, sekvenční dopředný plovoucí výběr, mRMR, Gaussovy smíšené modely, kNN.

ABSTRACT

The thesis deals with the recognition of Parkinson's disease from the speech signal. The first part refers to the principles of speech signals and speech signals by patients suffering from Parkinson's disease. Further, it continues to describe the issues of speech signals processing, basic symptoms used for diagnosis of Parkinson's disease (e. g. VAI, VSA, FCR, VOT etc.) and reduction of these symptoms. The next part focuses on a block diagram of the program for the diagnosis of Parkinson's disease. The main objective of this thesis is comparison of two methods of feature selection (mRMR and SFFS). For classification have selected two different methods were used. The first method is classification kNN and second method of classification is Gaussian mixture model (GMM).

KEYWORDS

Speech, speech signal analysis, formant, pre-emphasis, segmentation, symptom, sequential floating forward selection, mRMR, Gaussian mixture models, kNN.

KARÁSEK, Michal *Diagnóza Parkinsonovy choroby z řečového signálu*: diplomová práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2011. 50 s. Vedoucí práce byl Ing. Jiří Mekyska

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Diagnóza Parkinsonovy choroby z řečového signálu“ jsem vypracoval 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 neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

Brno ....

(podpis autora)

PODĚKOVÁNÍ

Moje poděkování patří Ing. Jiřímu Mekyskovi za jeho odborné konzultace, poskytnuté materiály a podklady, které mi pro účely diplomové práce obstaral.

V Brně dne..... .....

podpis autora

OBSAH

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

ÚVOD

Tato diplomová práce se zabývá zpracováním řečových signálů a následným využitím pro diagnózu Parkinsonovy choroby. Rozpoznání této choroby závisí na jejím stádiu. V diplomové práci se rozpoznává choroba už v ranném stádiu a to z řečového signálu.

Tato choroba byla poprvé popsána britským lékařem Jamesem Parkinsonem roku 1817. Známa byla však už od středověku. Výskyt Parkinsonovy choroby se projevuje u lidí nad 50 let. Jsou však i případy, kdy se choroba projeví i u mladších lidí. V České republice trpí touto nemocí kolem 15000 lidí. Je zapříčiněna nadměrnou ztrátou nervových buněk, které produkují v mozku neurotransmiter dopamin. Je to látka, která reguluje činnost určité části mozku.

V současné době neexistuje na Parkinsonovu chorobu žádný lék, pouze prostředky, které mírní průběh nemoci a proto je snaha o zjištění příznaků co nejdříve. Takovým lékem, který tlumí tuto nemoc je L-dopa (dopamin).

Cílem práce „Diagnóza Parkinsonovy choroby z řečového signálu“ je vybrat vhodné příznaky sloužící k diagnóze Parkinsonovy choroby, pomocí různých metod pro výběr příznaků. Dalším cílem bylo porovnání různých metod sloužících ke klasifikaci. Veškeré algoritmy jsou implementovány v prostředí Matlab.

1 ŘEČOVÉ SIGNÁLY

1.1 Tvorba řeči

Jedním z nejsložitějších procesů, které probíhají v lidském těle je řeč. Je to nejpoužívanější prostředek komunikace mezi lidmi.

Velmi důležitým faktorem ve tvorbě řeči je mozek. Jeho řečová centra a s nimi spojené funkce se většinou nacházejí v jedné mozkové hemisféře. U praváků je to obvykle v levé hemisféře a u leváků v pravé. Tady mluvčí vytváří vhodná slova a slovní spojení, které složí do srozumitelné podoby daného jazyka. A odtud pomocí pohybových nervů, přenáší impulzy do svalů mluvících orgánů. Tyto svaly se pohybují takovým způsobem, že vytváří změny akustického tlaku vzduchu a informace se šíří pomocí akustické vlny až k posluchači, který tuto informaci zachytí svým sluchovým ústrojím. V tomto ústrojí se akustický tlak transformuje na nervové impulzy, které se převádějí pomocí smyslových nervů až do mozku posluchače [19].

1.2 Hlasový trakt

Je tvořen jednotlivými řečovými orgány. Můžeme jej rozdělit na tři základní ústrojí:

1.2.1 Dechové ústrojí (respirační)

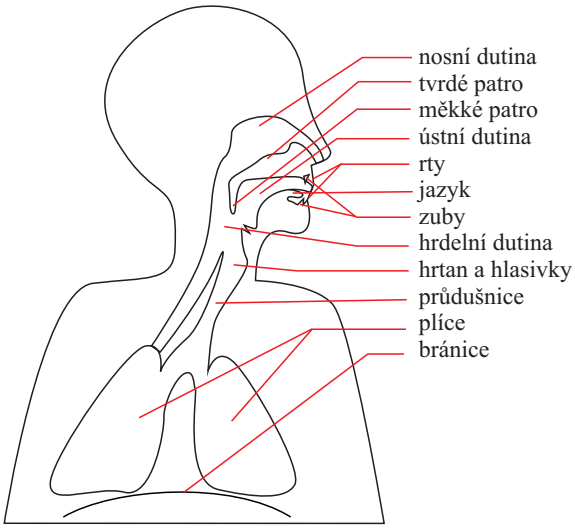
Primárně slouží k základní životní funkci – fyziologickému dýchání, ale zároveň je hlavním zdrojem energie pro řeč. Při nádechu se vzduch nažene do plic a vytváří základní materiál pro tvorbu řeči. Při řeči se množství nadechnutého vzduchu a zároveň i rytmus dýchání řídí zčásti vědomě, zatímco u fyziologického dýchání je nevědomé (reflexivní). Mluvní projev se realizuje při výdechu z plic, kde proud vzduchu prochází průdušnicí, hrtanem a nadhrtanovými dutinami. Tady se výdechový vzduch modifikuje. Díky zásobnímu vzduchu v plicích a trvání výdechu můžeme mluvní projev prodloužit. Síla výdechu má za následek to, jak bude hlasové ústrojí fungovat, jaký bude mít vliv na sílu hlasu a z části i na jeho výšku. Opětovný nádech vytváří nový materiál pro tvorbu řeči [11].

1.2.2 Hlasové ústrojí (fonační)

Hlasové ústrojí se nachází v hrtanu, který se nachází přibližně ve střední části krku nad průdušnicí, která spojuje hrtan s plícemi.

Stěna hrtanu je tvořena hrtanovými chrupavkami a vstup do hrtanu uzavírá hrtanová záklopka, která je kontrolována mozkem. Část hrtanu, která vytváří hlas se

nazývá hlasivky (glotis). Hlasivky jsou tvořeny dvěma hlasovými řasami, mezi nimiž je hlasová štěrbina. Ty se nacházejí v místě jeho nejužšího průchodu. Hlasové řasy se skládají z dvou tenkých vazů, které se otevírají a zavírají, když jimi proudí vzduch a jejich zadní konec je spojen s párem pohyblivých hlasových chrupavek. Přední konec je pevně ukotven ve štítné chrupavce (část Adamova jablka). Hlasivkové chrupavky mění polohu a tím se mění velikost hlasové štěrbiny. Když člověk mlčí, pak je hlasová štěrbina odkryta a vzduch může bez odporu proudit. Při polykání je hlasová štěrbina zcela uzavřená. Během řeči se hlasová štěrbina zúží, vzduch z plic je vypuzen ven přes hlasivky a tím nastane vibrace hlasivek. Tomuto procesu se říká fonace. Hlasivky se při tom střídavě otevírají a prudce uzavírají a tím se vzduchový proud „rozdrobí“ tak, že se kvantum řidšího a hustšího vzduchu téměř pravidelně střídá. Tento periodický proud vzduchových pulzů se nazývá základní hlasivkový tón. Výška hlasu závisí na délce a napětí hlasivek. Přirozené zbarvení a hloubka hlasu závisí na velikosti a tvaru hrdla, úst a nosu [11], [19].



Obr. 1.1: Hlasové ústrojí člověka.

1.2.3 Hláskotvorné ústrojí (artikulační)

Ústa jsou dalším aspektem pro vytvoření určitého zvuku. Například pro vytvoření souhlásek [d] nebo [k], je zapotřebí aby jazyk přerušil proud vzduchu přicházejícího z hrtanu. Naopak samohlásky [a] nebo [u] toto přerušení nepotřebují ale potřebují určitou polohu jazyka, zubů a rtů. Aby se přeměnili jednoduché zvuky, tvořené hlasivkami, na srozumitelná slova je zapotřebí, kromě již zmíněných rtů a jazyka, také měkké patro a prostory umožňující rezonanci. To může být například celá ústní a nosní dutina, hltan a v menší míře se uplatní i hrudní dutina. Koordinaci mezi

jednotlivými strukturami, podílející se na tvorbě řeči, je dosaženo díky velkému počtu drobných svalů, které spolu velmi rychle spolupracují [11].

1.3 Spektrogram řeči

Je to časově frekvenční reprezentace řečového signálu. Řeč je nestacionární a proto usilujeme o to aby znázornění spektra řeči bylo v čase. Na vodorovné ose vynášíme čas, na svislé ose udáváme frekvenci a stupně šedi (nebo barevné rozlišení) udává energii. Spektrogramy zobrazují pouze modul. Stejně jako u modulových spekter je informace o fázi ignorována [11].

Rozlišujeme dva typy spektrogramů: širokopásmový a úzkopásmový.

1.3.1 Širokopásmový spektrogram

Je základním nástrojem pro spektrální analýzu signálů. Používá krátká váhová okénka. V časové oblasti zobrazují detaily, ale nejsou schopny zobrazit příliš jemné frekvenční detaily. Spíše než vlastní frekvence zobrazují obálku signálu, díky tomu můžeme sledovat vývoj jednotlivých formantů (viz. 1.4) v čase.

1.3.2 Úzkopásmový spektrogram

Používá se méně. Používá se delší váhové okénko. Tyto spektrogramy zobrazují detaily ve frekvenční oblasti a díky tomu z nich lze vyčíst jednotlivé harmonické frekvence, které odpovídají frekvenci základnímu hlasivkovému tónu. Naopak nezobrazují detaily v časové oblasti.

Pokud není žádáno, aby se objevovali parazitní horizontální nebo vertikální čáry, lze použít kompromis („středněpásmový“ spektrogram). V tom případě odpovídá délka okénka přibližně dvoj až trojnásobku lokální periody základního hlasivkového tónu.

1.4 Formanty

Jsou to oblasti koncentrace (zesílení) akustické energie, které vznikají v důsledku průchodu základního hlasivkového tónu nadhrtanovými dutinami a následné rezonance v těchto dutinách. Jednotlivé formanty označujeme čísly od nejnižší frekvence po nejvyšší ($F_1, F_2, \ldots, F_n$). Pokud se zahrne nosní dutina do procesu vytváření řeči, dochází vlivem jejích antirezonančních vlastností k potlačení některých frekvencí. To jsou tzv. antiformanty a označují se ($A_1, A_2, \ldots, A_n$) [11].

1.5 Suprasegmentální rysy (prozodie)

Jsou to takové vlastnosti řečového systému, které souvisí zejména s frekvencí základního hlasového tónu (výškou hlasu), intenzitou (hlasitostí) a časováním. Pod časováním si můžeme představit rytmus a rychlost řeči. Suprasegmentální rysy se vztahují především k delším úsekům řeči. Mohou to být např. slabiky, slova, celé věty, nebo i delší promluvy. Jednotlivé výše zmíněné vlastnosti jsou spolu úzce spjaty (např. časové členění řeči je výrazně ovlivňováno melodií a podobně) [11].

2 ŘEČ PACIENTŮ POSTIŽENÝCH PARKINSONOVOU CHOROBOU

Tato choroba je způsobena ztrátou neurotransmiteru dopaminu, která způsobuje ztuhnutí svalů, nemožnost uvedení svalů do pohybu, třesem a pomalým pohybem svalstva. Postižení řeči se projevuje u zhruba 70% pacientů postižených Parkinsonovou chorobou a je doprovázeno hypokinetickou disartrií [5], [18].

Poruchy řeči u těchto pacientů jsou časté. Zpočátku se může hlas projevovat tlumeně s malými výchylkami v intenzitě a intonace monotónně. Dále je také možná porucha řeči, která se projevuje rychlím nahuštěním slov, „zaseknutím se“ uprostřed věty (nevhodné mlčení), přerušením řeči uprostřed slova a vyslovením malého počtu hlásek na jeden nádech. Tyto aspekty se obecně nazývají „problém s časováním“. Někdy se projevuje chraplavé zabarvení hlasu a artikulace. V pokročilejších případech se mohou objevit i poruchy řeči připomínající koktání. Tím pádem může být řeč nesrozumitelná [5].

Fonace dosahuje u pacientů s Parkinsonovou chorobou vyšších hodnot (střední hodnota F_0). To může způsobovat ztuhlost hlasivek.

Intonace je ovlivněna tím, že pacienti mluví monotónně a s nevýrazným přízvukem. Sledování probíhá pomocí směrodatné odchylky.

Artikulace (práce s mluvidly) pacientům s Parkinsonovou chorobou dělá ve většině případů značné problémy. Je sledována pomocí formantových kmitočtů. Ty se používají u řady příznaků (viz. 5). Pacienti mohou mít špatnou promlouvu jak vokálů tak konsonantů [5], [15], [18].

3 ZPRACOVÁNÍ ŘEČOVÝCH SIGNÁLŮ

Zpracování řečových signálů není zdaleka jednoduchou záležitostí. Důvodem k tomu je, že lidská řeč je dosti různorodá. Touto různorodostí se může rozumět například: rychlost promluvy, výška základního tónu, hlasitost, nebo také rozdíl mezi jednotlivými osobami, které řečový signál podávají.

V současné době se řečový signál výhradně zpracovává číslicově. Ale ještě před samotným použitím je vhodné tento signál upravit, jelikož při nahrávání se mohou vyskytnout negativní vlivy. Ty můžou být způsobeny okolním hlukem, šumem, různými rušivými elementy, může se projevit zkreslení při záznamu a přenosu na médium. Na těchto zkresleních se mohou podílet kmitočtové charakteristiky mikrofonů, zesilovačů, ekvalizérů atd. Proto, když je požadován kvalitní zvukový příjem, je důležité mít kvalitní studiový mikrofon (kondenzátorový) a nahrávku pořizovat v bezodrazové komoře [11], [19].

3.1 Preemfáze

U řečového signálu dochází k tomu, že se v kmitočtovém spektru úroveň spektrálních složek, se zvyšující se frekvencí, snižuje. Proto k tomu, abychom vyrovnali kmitočtové spektrum (zdůraznili amplitudy na vyšších frekvencích), použijeme preemfázový číslicový filtr typu horní propust (FIR), jehož diferenciální rovnice je [19]:

$$y[n] = x[n] - a_1 x[n-1]. \tag{3.1}$$

A odpovídající přenosová funkce má tvar:

$$H(z) = 1 - a_1 z^{-1}, \tag{3.2}$$

kde koeficient a_1 nabývá hodnot od 0,9 do 1.

3.2 Ustředění

Neboli také potlačení stejnosměrné složky. Tato složka, při zpracování řeči, může být velmi nežádoucí. Proto její střední hodnotu jednoduše odečteme od signálu:

$$s'[n] = s[n] - a[n], \tag{3.3}$$

kde $s'[n]$ je řečový signál s odečtenou průměrnou hodnotou $a[n]$. Pokud známe celý řečový signál, můžeme střední hodnotu odhadnout off-line:

$$a[n] = \frac{1}{N} \sum_{n=0}^{N-1} s[n]. \tag{3.4}$$

3.3 Segmentace

Úkolem segmentace je rozdělit řečový signál na kratší části o takové délce, aby byl pro metody odhadu parametrů stacionární,tzn. že statistické charakteristiky nejsou závislé na posunutí počátku časové osy. Tyto části by na druhou stranu měly být dostatečně velké, aby bylo možné přesně odhadnout parametry.

Při výběru signálu do rámců se používají nejčastěji tyto okénkové funkce:

- **Hammingovo okno** – utlumuje signál na okrajích [11]:

$$w[n] = \begin{cases} 0,54 - 0,46\cos(\frac{2\pi n}{l_{\text{ram}}-1}) & \text{pro} \quad 0 \leq n \leq l_{\text{ram}} - 1 \\ 0 & \text{jinde} \end{cases} \tag{3.5}$$

- **Pravoúhlé okno** – signál ponechává beze změny [11]:

$$w[n] = \begin{cases} 1 & \text{pro} \quad 0 \leq n \leq l_{ram} - 1 \\ 0 & \text{jinde} \end{cases} \tag{3.6}$$

4 SEGMENTÁLNÍ PŘÍZNAKY

4.1 Lineární prediktivní analýza

K nejefektivnějším metodám analýzy akustických signálů patří LPC (Linear Predictive Coding), neboli AR (Auto–Regressive) modeling. Tato metoda je často používaná, protože je rychlá, jednoduchá a dokáže efektivně odhadnout hlavní parametry řečového traktu. Metoda LPC je založena na předpovídání aktuálního vzorku $\tilde{s}[n]$ z lineární kombinace předchozích p vzorků [11]:

$$\tilde{s}[n] = \sum_{k=1}^{p} a_k s[n-k], \tag{4.1}$$

kde a_k jsou lineární predikční koeficienty.
Chyba lineární predikce:

$$e[n] = s[n] - \tilde{s}[n] = s[n] - \sum_{k=1}^{p} a_k s[n-k]. \tag{4.2}$$

Přenosová funkce analyzujícího filtru:

$$A(z) = \frac{E(z)}{S(z)} = 1 + \sum_{k=1}^{p} a_k z^{-k}, \tag{4.3}$$

kde $E(z)$ je Z-transformace $e[n]$.

4.2 Perceptivní lineární predikční analýza

Lineární prediktivní analýza, jak již bylo zmíněno, je velmi efektivní při popisu spektálních vlastností řečového signálu. Avšak tento prostředek špatně reprezentuje řečový signál, jak jej vnímá lidký sluchový systém, protože člověk vnímá zvuky nelineárně. Také není do této metody zahrnuto maskování zvuků a kritická pásma spektrální citlivosti. Z těchto důvodů byla navržena nová analýza profesorem Hynkem Heřmanským nazývaná PLP (Perceptual Linear Predictive). Na rozdíl od analýzy LPC se používá [6], [11]:

- kritické pásmo spektrální citlivosti,
- křivky stejné hlasitosti,
- závislost mezi intenzitou zvuku a jeho vnímanou hlasitostí.

4.3 Melovské kepstrální koeficienty

MFCC (Mel Frequency Cepstral Coefficients) jsou podobně jako u PLP navrženy podle vlastností lidského sluchového systému. Pomocí banky trojúhelníkových pásmových filtrů s nelineárním kmitočtovým rozložením, se Melovské kepstrální koeficienty snaží napodobyt nelineární vnímání frekvencí.

Melovské měřítko lze aproximovat následujícím vztahem [11]:

$$f_m = 2595 \log_{10} \left(1 + \frac{f}{700}\right), \tag{4.4}$$

kde f je frekvence v lineární škále a f_m je odpovídající frekvence v melovské škále.

Trojúhelníkové filtry jsou rozloženy přes celé frekvenční pásmo od nuly do Nyquistova kmitočtu. Pokud existují frekvenční oblasti, které neobsahují užitečnou energii signálu, jsme schopny omezit toto přenášené pásmo.

Hlavní rozdíl MFCC oproti PLP je, že se nepoužívají křivky stejné hlasitosti a nezohledňuje se závislost mezi intenzitou zvuku a jeho vnímanou hlasitostí [2], [6], [11].

4.4 Lineární predikční kepstrální koeficienty

LPCC (Linear Prediction Cepstral Coefficients) byly mnoho let běžně používany v mnoha aplikacích rozpoznávačů řeči. Pomocí kepstrálních koeficientů můžeme popsat lineární systém, kterým se modeluje hlasový trakt.

Kepstrální koeficienty jsou obecně dekorelované, díky tomu se používají v systémech rozpoznávání řeči. To je velká výhoda tohoto systému, protože v rozpoznávačích řeči založených na skrytých Markovových modelech, si vystačíme s vektory rozptylů (neuvažujeme plné kovarianční matice) [2], [11].

5 PŘÍZNAKY POUŽÍVANÉ K DIAGNÓZE PARKINSONOVY CHOROBY

Klasické příznaky jako MFCC, PLP,... se při diagnóze Parkinsonovy choroby moc nepoužívají. Spíše se využívá příznaků zmíněných níže, protože lépe popisují tvorbu řeči, práci s mluvidly atd. Kromě těchto příznaků se také používají suprasegmentální příznaky, což je kmitočet základního tónu F_0, intenzita a tempo (viz. 1.5).

5.1 Vowel articulation index (VAI)

Artikulační index samohlásek se získává ze tří samohlásek [a], [i] a [u]. Výpočet tohoto indexu je popsán následujícím vzorcem [18]:

$$VAI = (F_{2\mathrm{i}} + F_{1\mathrm{a}})/(F_{1\mathrm{i}} + F_{1\mathrm{u}} + F_{2\mathrm{u}} + F_{2\mathrm{a}}), \tag{5.1}$$

kde F_1 a F_2 jsou první a druhé formantové frekvence tří výše uvedených samohlásek. Tyto formantové frekvence se většinou průměrují z několika (např. deseti) samohlásek definovaných z řečového úkolu (předem definovaný text).

5.2 Vowel space area (VSA)

VSA je obvykle postavena na euklidovské vzdálenosti mezi F_1 a F_2 souřadnicemi samohlásek [u], [i], [a] – **trojúhelníkový VSA**, nebo samohlásek [u], [i], [a], [e] – **čtyřúhelníkový VSA**. Vzniklá plocha mezi jednotlivými samohláskami je u pacientů s Parkinsonovou chorobou menší než u zdravých lidí. To má za následek, že těmto pacientům je hůře rozumět [5].

Trojúheníkový VSA může být matematicky vyjádřen následujícím vzorcem [15]:

$$VSA = ABS((F_{1\mathrm{i}} \cdot (F_{2\mathrm{a}} - F_{2\mathrm{u}}) + F_{1\mathrm{a}} \cdot (F_{2\mathrm{u}} - F_{2\mathrm{i}}) + F_{1\mathrm{u}} \cdot (F_{2\mathrm{i}} - F_{2\mathrm{a}}))/2), \tag{5.2}$$

kde ABS je absolutní hodnota.

5.3 Formant centralization ratio (FCR)

Tento příznak pracuje podobně jako VSA, tj. jak je schopen řečník od sebe odlišit jednotlivé vokály. Od VSA se však odlišujě tím, že zavádí normalizaci. To znamená, že není tak citlivý na pohlaví mluvčího (smaže rozdíly mezi muži, ženami a dětmi). FCR je možné vypočítat následovně [15]:

$$FCR = (F_{2\mathrm{u}} + F_{2\mathrm{a}} + F_{1\mathrm{i}} + F_{1\mathrm{u}})/(F_{2\mathrm{i}} + F_{1\mathrm{a}}). \tag{5.3}$$

5.4 Voice onset time (VOT)

VOT je doba mezi počátkem plozivy a počátkem samohlásky. Tato doba je u jednotlivých slabik různá. Například slabika [pa] má tuto dobu menší než slabika [ta], přičemž tato doba se u pacientů postižených Parkinsonovou chorobou v různých literaturách liší. Není přesně specifikováno, jestli bude vyšší, nebo nižší.

VOT ratio – VOT normalizován vůči tempu řeči [4].

Dále se pro diagnózu Parkinsonovy choroby používají příznaky: střední hodnota F_0, jitter, směrodatná odchylka F_0, energie, shimmer, articulation rate, pause ratio, phonatory onset and offset, net speech rate, TSR (Total Speech Rate), F_0VR (F_0 Variation Range), ATRI (Amplitude Tremor Intensity Index), FTRI (Frequency Tremor Intensity Index), ATF (Amplitude Tremor Frequency), FFTF (Fundamental Frequency Tremor Frequency), ISD (Inter-pause Speech Duration), SPIR (Speech Index of Rhythmicity),...

6 VÝBĚR PŘÍZNAKŮ

Náklady na měření i čas klasifikace se zvyšují se vzrůstajícím počtem příznaků. Častokrát se stává, že zvyšující se množství příznaků nevede k lepší klasifikaci. Může dojít i k tomu, že úspěšnost klasifikace se snižuje. Z tohoto důvodu se snažíme vybrat pro klasifikaci pouze ty příznaky, které mají největší informační přínos. Při tom rozlišujeme zda se jedná o selekci, či extrakci příznaků.

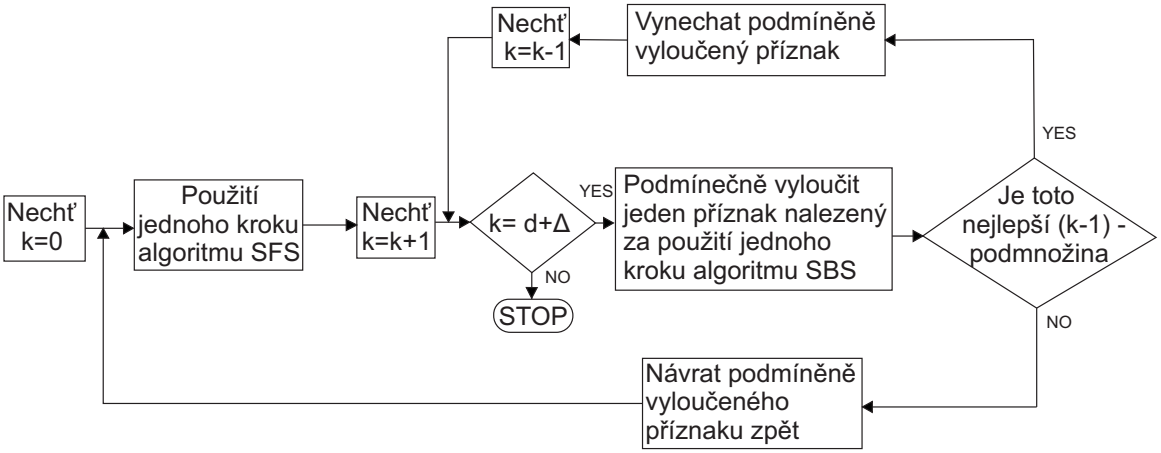
Selekce je výběr nejvhodnějších příznaků z celkového množství příznaků původních.

Extrakce slouží ke změně původních příznaků, za použití funkce závislé na všech těchto příznacích, na menší počet nových příznaků se stejným informačním přínosem.

6.1 Sequential Floating Forward Selection (SFFS)

SFFS je jednou z metod pro výběr příznaků. Tato metoda má velkou výhodu v tom, že používá klasifikátor pro výběr optimálních příznaků. Ten se poté použije pro klasifikaci. Hledání v opačném směru je známé jako „sekvenční zpětný plovoucí výběr“ (Sequential Backward Floating Selection – SBFS). Oba tyto algoritmy se celkově nazývají floating methods (plovoucí metody). Tento název je odvozen od toho, že dimenzionalita v každém kroku se nemění monotónně, ale „plave“ nahoru a dolů. Plovoucí metody hledání byly v nezávislých studiích vyhodnoceny jako nejefektnější suboptimální algoritmy výběru příznaků [12].

Sekvenční dopředný plovoucí výběr pracuje takovým způsobem, že v prvním kroku vezme samostatně každý příznak a u konkrétních dat se přiřadí pořadí podle úspěšnosti jejího rozpoznání. V dalším kroku se k nejúspěšnějšímu příznaku přidávají do dvojice další příznaky a opět se zjišťuje, jestli neexistuje dvojice s vyšší úspěšností, než je hodnota pro samostatný nejlepší příznak. Jakmile se nalezne nejlepší dvojice příznaků, přidávají se k této dvojici další příznaky a tím vznikne trojice atd. Toto hledání probíhá tak dlouho dokud se nenajde skupina příznaků, která bude mít největší úspěšnost v rozpoznání konkrétních dat. Princip je patrný z obrázku 6.1.



Obr. 6.1: Zjednodušené schéma algoritmu SFFS.

6.2 minimum Redundancy Maximum Relevance Feature Selection (mRMR)

Maximální relevance je hledání funkce vyhovující vzorci 6.1, která odpovídá $D(S, c)$ ze vzorce 6.2. Z hlediska vzájemné výměny informací, je účelem výběru příznaků najít sadu funkcí S, se střední hodnotou všech vzájemně vyměněných informací mezi funkcí x_i a třídou c [9]:

$$\max\ D(S, c),\ D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c), \tag{6.1}$$

kde $I(x_i; c)$ je největší vzájemná výměna informací v třídě c, což odráží největší závislost na cílové třídě.

$$\max\ D(S, c),\ D = I(x_i, i = 1, \ldots, m; c). \tag{6.2}$$

Je pravděpodobné, že vybrané funkce podle maximální relevance by mohly být bohaté na redundanci (závislost mezi těmito funkcemi by mohla být velká). Když jsou dva prvky na sobě velmi závislé, tak při odstranění jednoho z nich se odlišnost tříd moc nezmění. Vzorec pro minimální redundanci[9]:

$$\min\ R(S),\ R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i; x_j). \tag{6.3}$$

Kritérium, které kombinuje vzorce 6.1 a 6.3 se nazývá „minimální redundance a maximální relevance“ definovaná následujícím vztahem[9]:

$$\max\phi(D, R), \phi = D - R. \tag{6.4}$$

6.3 Míra geometrické oddělitelnosti

Tato metoda pracuje na principu ohodnocení jednotlivých příznaků na základě jejich rozptylu a vzdálenosti mezi třídami. Poté záleží jen na výběru příznaků, které budou použity dále v klasifikaci.

Kritérium míry geometrické oddělitelnosti $Q(x_i)$ vyjádřuje kvalitu příznaku x_i pomocí sledování rozložení hodnot příznaků v příznakovém prostoru. Pokud se prvky jedné třídy vyskytují v okolí střední hodnoty a zároveň se střední hodnoty jednotlivých tříd co nejvíce liší, považuje se příznak za kvalitní.

Kvadrát vzdálenosti mezi středními hodnotami ($\underline{\mu}_u$ a $\underline{\mu}_v$) tříd u a v [17]:

$$D_{v,u}^2 = (\underline{\mu}_v - \underline{\mu}_u)^2. \tag{6.5}$$

Aritmetická střední hodnota vzdáleností mezi všemi třídami je určena podle [17]:

$$D^2 = \frac{1}{V(V-1)} \sum_{v=1}^{V} \sum_{u=1}^{V} D_{v,u}^2, \tag{6.6}$$

kde V je celkový počet tříd. Kvadrát rozptylu třídy v okolo střední hodnoty je určený [17]:

$$S_v^2 = (\underline{x} - \underline{\mu}_v)^2, \tag{6.7}$$

kde $\underline{x}$ je vektor příznaků. Aritmetickou střední hodnotu určíme ze vztahu [17]:

$$S^2 = \frac{1}{V} \sum_{v=1}^{V} S_v^2. \tag{6.8}$$

Geometrická oddělitelnost (separabilita) tříd v příznakovém prostoru se vypočítá podle vztahu [17]:

$$Q(.) = \frac{S^2}{S^2 + D^2}, \quad 0 \leq Q(.) \leq 1. \tag{6.9}$$

Jakmile příznak x_i vykazuje malé rozdíly v rámci své třídy a naopak velké rozdíly mezi třídami, pak míra oddělitelnosti $Q(x_i)$ dosahuje malých hodnot (přibližuje se k nule). Velké hodnoty (blížící se jedné) naopak ukazují, že příznaky jsou nevhodné pro rozpoznávání z důvodu velkých rozptilů těchto hodnot [17].

7 KLASIFIKÁTORY

Pomocí klasifikace se rozhoduje, zda zkoumaný objekt (člověk) patří do určité skupiny. V našem případě, zda se jedná o nemocného či zdravého člověka. Konkrétně se tato diplomová práce zaměřuje na dva druhy klasifikátorů (GMM a kNN).

7.1 Metoda nejbližších sousedů (kNN)

Z anglického „K-nearest neighbor“. Jedná se o jeden z nejzákladnějších a nejjednodušších klasifikátorů. Častokrát se používá pro první klasifikaci studie, kde je malá nebo žádná předchozí znalost o distribuci dat.

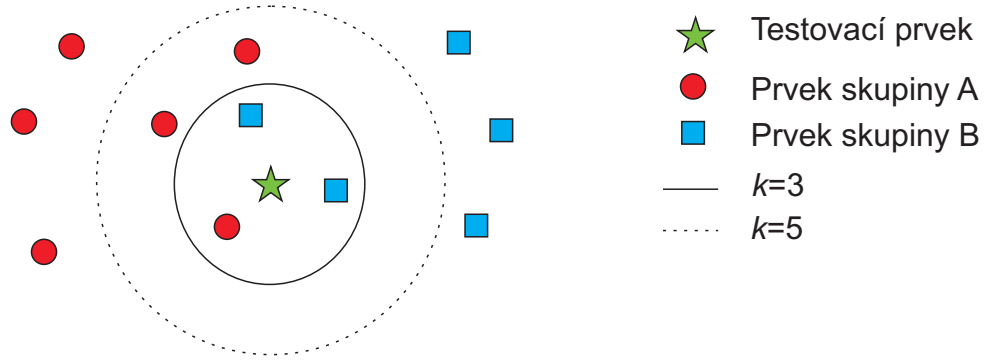
Metoda nejbližších sousedů je založena na Euclidovské vzdálenosti mezi zkušebním vzorkem a natrénovanými vzorky. Dotazovaný prvek se při klasifikaci umístí do konkrétního místa N-rozměrného prostoru, kde se nacházejí již natrénované vzorky jednotlivých množin a nalezne se k nejbližších sousedů. Většinou se k volí jako liché číslo, aby se zabránilo přiřazení do jednotlivých množin stejné množství bodů. Když je $k = 1$ pak se kontrolní bod přiřadí do skupiny k nejbližšímu bodu.

Euclidovská vzdálenost mezi vzorkem x_i a x_l $l = (1, 2, 3, \ldots, n)$, je definována jako [16]:

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \ldots + (x_{ip} - x_{lp})^2}, \tag{7.1}$$

kde x_i je vstupní vzorek funkce p $(x_{i1}, x_{i2}, \ldots, x_{ip})$, n bude celkový počet vzorků $(i = 1, 2, 3, \ldots, n)$ a p dimenze vektoru příznaků $(j = 1, 2, 3, \ldots, n)$.

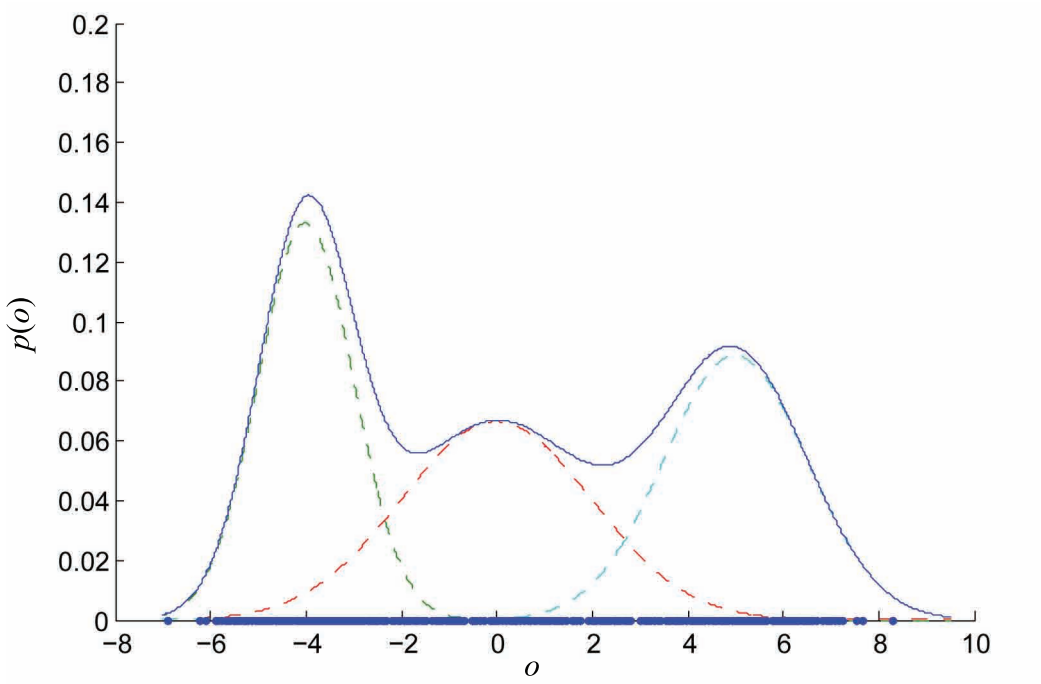
Jak můžeme vidět z obrázku 7.1, záleží velmi na volbě k. Když bylo zvoleno $k = 3$, pak byl testovací bod přiřazen ke skupině B (modré čtverečky) a při volbě $k = 5$ byl tento bod přiřazen do skupiny A (červená kolečka).



Obr. 7.1: Klasifikace pomocí metody kNN.

7.2 Gaussovy smíšené modely (GMM)

Gaussovy smíšené modely jsou jednou z metod, které využívají statistické rozpoznávání vzorů. Tyto metody pracují na stejném základu a to takovém, že některé statistické vlastnosti mohou být podobné u modelů stejných tříd. U GMM se pomocí směsí Gaussových funkcí modelují jednotlivé třídy příznaků. GMM parametry jsou odhadnuty z trénovacích dat pomocí iteračního EM (Expectation-Maximalization) algoritmu [11].



Obr. 7.2: Příklad GMM v jednorozměrné dimenzi.

Váženou lineární kombinací normálních rozdělení jednotlivých tříd (smíšené Gaussovy modely) můžeme popsat rozdělení pravděpodobnosti příznakových vektorů. To můžeme vyjádřit následující rovnicí [11]:

$$p(o|\lambda) = \sum_{i=1}^{M} w_i p_i(o), \tag{7.2}$$

kde M je počet Gaussových funkcí, $w_i, i = 1, \ldots, M$, jsou váhy jednotlivých složek, které vyhovují podmínce:

$$\sum_{i=1}^{M} w_i = 1 \tag{7.3}$$

a $p_i(o), i = 1, \ldots, M$ jsou hustoty pravděpodobností jednotlivých složek viz. [11]:

$$p_i(o) = \frac{1}{(2\pi)^{n/2}|\mathbf{C_i}|^{1/2}} \exp\left(\frac{1}{2}(o-\mu_i)^T(\mathbf{C_i})^{-1}(o-\mu_i)\right), \tag{7.4}$$

kde n je dimenze příznakových vektorů. Dále obsahuje n-rozměrnou normální hustotu pravděpodobnosti se střední hodnotou μ_i a kovarianční maticí $\mathbf{C_i}$.

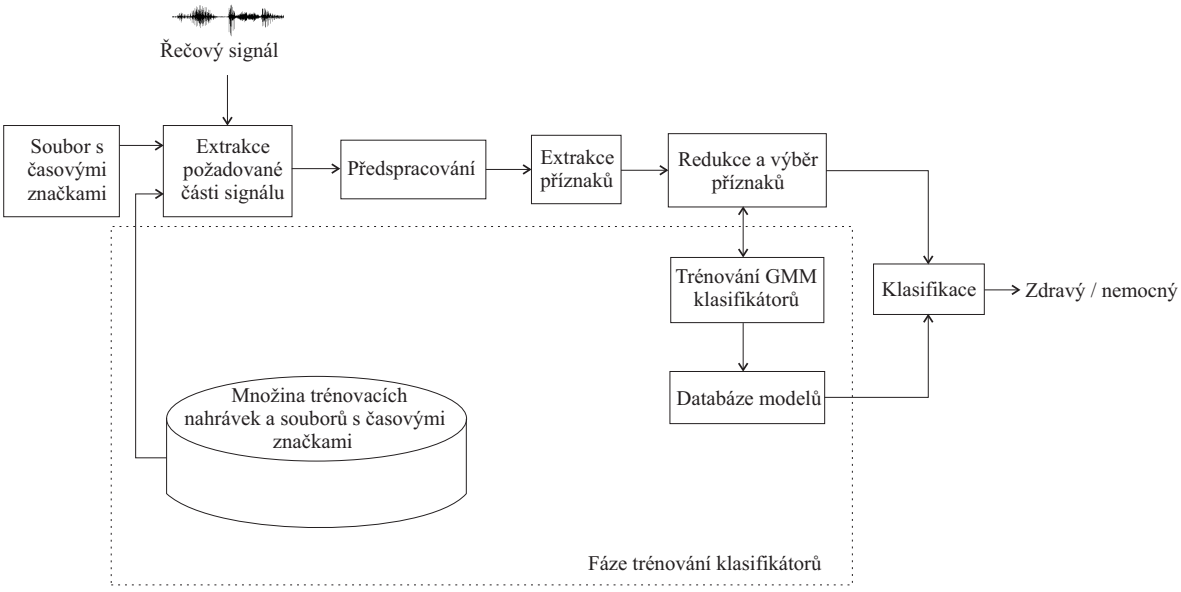
Kompletní Gaussovy smíšené modely jsou parametrizované vektory středních hodnot, kovarianční matice a váhy, které se smíchají ze všech hustot komponent. Tyto parametry jsou kolektivně reprezentovány notací:

$$\lambda = \{w_i, \mu_i, \mathbf{C_i}\} \quad i = 1, \ldots, M. \tag{7.5}$$

Existuje několik variant GMM uvedených v 7.5. Kovarianční matice $\mathbf{C_i}$ může být plnohodnotná, nebo diagonálně omezená. Kromě toho mohou být parametry mezi Gaussovými modely sdílené, a sice pomocí společné kovarianční matice.

8 ROZPOZNÁVÁNÍ PARKINSONOVY CHOROBY

Program pro rozpoznání Parkinsonovy choroby bude pracovat na principu, jenž zobrazuje následující blokové schéma (8.1).



Obr. 8.1: Blokové schéma programu pro rozpoznání Parkinsonovy choroby.

Jak je vidět z tohoto schématu, program bude pracovat ve dvou režimech a to v režimu trénování a v diagnostickém režimu.

8.1 Režim trénování

Proto aby se mohly porovnávat nahrávky od pacientů postižených Parkinsonovou chorobou a nahrávky aktuálního pacienta, musí se nejprve natrénovat GMM klasifikátor.

Blok „Množina trénovacích nahrávek a souborů s časovými značkami“ obsahuje jenotlivé nahrávky pacientů s Parkinsonovou chorobou. U konkrétní nahrávky se v bloku „Extrakce požadované části signálu“ vybere část řečového signálu, ve které se bude s největší pravděpodobností vyskytovat příznak choroby. Tato část signálu se v bloku „Předspracování“ upraví pomocí ustředění (odstranění stejnosměrné složky), preemfáze (horní propust typu FIR) a segmentace. U „Extrakce příznaků“ se v takto upraveném signálu vypočítají jednotlivé příznaky a poté se vyberou pro „Trénování GMM klasifikátorů“ (výpočet podle vzorce 7.4 a 7.2). Následně se uloží do „databáze modelů“.

8.2 Režim diagnózy

Režim diagnózy vyhodnocuje s jakou pravděpodobností se nahrávka nového pacienta (u kterého chceme zjistit zda je postižený Parkinsonovou chorobou) shoduje s nahrávkami postižených pacientů Parkinsonovou chorobou.

Blok „Soubor s časovými značkami“ slouží k tomu, aby v řečovém signálu rozpoznal konkrétní úsek (např. souhlásku [a]). Poté se tento úsek vybere a opět se provede předzpracování sígnálu (ustředění, preemfáze, segmentace), náselduje výpočet příznaků a redukce (výběr) příznaků. Blok „Klasifikace“ vyhodnotí, který z modelů s největší pravděpodobností patří vektoru příznaků (zda se jedná o zdravého či nemocného člověka).

9 ŘEŠENÍ DIPLOMOVÉ PRÁCE

Tato diplomová práce vychází ze spolupráce s I. neurologickou klinikou ve Fakultní nemocnici u sv. Anny v Brně. Tady se vytvářejí jednotlivé nahrávky od pacientů s Parkinsonovou chorobou a zdravých osob. Tyto nahrávky obsahují přednes pacientů jednotlivých samohlásek, různých slov, slovních spojení a vět, následně utříděných do sekcí.

Jednotlivé příznaky, kterými se diplomová práce zabývá, jsou popsány v tabulce 9.1.

Tab. 9.1: Popis jednotlivých příznaků

název příznaku	popis
VSA	vokální oblast hlasového traktu
lnVSA	přirozený logaritmus formantů před výpočtem VSA
FCR	centralizační poměr formantů
VAI	artikulační index samohlásek
F2i/F2u	podíl F2i/F2u
F0VR	rozdíl mezi minimem a maximem kmitočtu základního tónu
F0_median	medián kmitočtu základního tónu
F0_mean	střední hodnota kmitočtu základního tónu
F0_std	směrodatná odchylka kmitočtu základního tónu
F0_min	minimální hodnota kmitočtu základního tónu
F0_max	maximální hodnota kmitočtu základního tónu
relF0VR	podíl (F0VR/F0_mean)*100
relF0SD	podíl (F0_std/F0_mean)*100 (v procentech)
voicing_frac	podíl znělých úseků v řečovém signálu
jitter_local	lokální kolísání periody
jitter_localabs	kolísání periody (střední absolutní rozdíl po sobě jdoucích intervalů)
jitter_rap	jitter (Relative Average Perturbation)
jitter_ppq5	jitter (five-point Period Perturbation Quotient)
jitter_ddp	jitter (Difference of Differences of Periods)
hnr_aut	příznak popisující harmonicitu
hnr_nh	příznak popisující harmonicitu
hnr_hn	příznak popisující harmonicitu
F1_mean	střední hodnota prvního formantu
F1_var	rozptyl prvního formantu
F1_std	směrodatná odchylka prvního formantu
F1_max	maximální hodnota prvního formantu
F1_min	minimální hodnota prvního formantu
F1_med	medián prvního formantu
F1_max_min	rozdíl maximální a minimální hodnoty prvního formantu
F1b_mean	střední hodnota šířky pásma prvního formantu
F1b_var	rozptyl šířky pásma prvního formantu
F1b_std	směrodatná odchylka šířky pásma prvního formantu
F1b_max	maximální hodnota šířky pásma prvního formantu
F1b_min	minimální hodnota šířky pásma prvního fotmantu
F1b_med	medián šířky pásma prvního formantu
F1b_max_min	rozdíl maximální a minimální hodnoty šířky pásma prvního formantu
F2_mean	střední hodnota druhého formantu
F2_var	rozptyl druhého formantu
F2_std	směrodatná odchylka druhého formantu
F2_max	maximální hodnota druhého formantu
F2_min	minimální hodnota druhého formantu
F2_med	medián druhého formantu
F2_max_min	rozdíl maximální a minimální hodnoty druhého formantu
F2b_mean	střední hodnota šířky pásma druhého formantu
F2b_var	rozptyl šířky pásma druhého formantu
F2b_std	směrodatná odchylka šířky pásma druhého formantu
F2b_max	maximální hodnota šířky pásma druhého formantu
F2b_min	minimální hodnota šířky pásma druhého fotmantu
F2b_med	medián šířky pásma druhého formantu
F2b_max_min	rozdíl maximální a minimální hodnoty šířky pásma druhého formantu

Nejprve je nutno uvést, že Fakultní nemocnice u sv. Anny v Brně prozatím disponuje velmi malou databází řečových signálů a zvláště pak databází zdravých řečníků. Proto byla stávající databáze rozšířena o nahrávky kontrolních řečníků, kteří byli zaznamenáni v nahrávacím studiu VUT v Brně, aby splnila účel této diplomové práce. Výsledky měření tedy mohou být zkresleny, a to především kvůli různému prostředí nahrávání pacientů (odlišných nastaveních mikrofonu, hlasitosti,. . .) a různých věkových skupin těchto pacientů. Dále se tato diplomová práce zabývá pouze úseky nahrávek se samohláskami vyslovenými muži, z důvodu malé databáze řečníků ženského pohlaví a neúplností databaze nahrávek.

9.1 Databáze příznaků

Databáze jednotlivých příznaků je reprezentována tabulkovým souborem typu *.xls. Tato databáze je vytvořena (vypočtené příznaky) z databáze nahrávek poskytnutých Fakultní nemocnicí u sv. Anny. Výpočet příznaků z databáze nahrávek nebyl součástí této diplomové práce.

Rozlišení řečníků v databázi příznaků:

- P1 – řečník ženského pohlaví s Parkinsonovou chorobou
- P2 – řečník mužského pohlaví s Parkinsonovou chorobou
- K1 – řečník ženského pohlaví kontrolní skupiny
- K2 (K0) – řečník ženského pohlaví kontrolní skupiny

Za tímto rozlišovacím parametrem následuje identifikační číslo pacienta, popřípadě obsahuje ještě příjmení osoby.

Databáze příznaků, kterou se budeme zabývat obsahuje 11 mužských řečníků s Parkinsonovou chorobou a 40 mužských kontrolních řečníků. Ženskými řečníky se diplomová práce nezabývá, z již zmíněného důvodu, nedostatku kontrolních řečníků.

Dále bude řešena pouze ta část s artikulací samohlásek. Ta se dělí na labely podle následujícího plánu:

1. krátké samohlásky
    - 7.1_1_a – [a]
    - 7.1_1_e – [e]
    - 7.1_1_i – [i]
    - 7.1_1_o – [o]
    - 7.1_1_u – [u]
2. dlouh0 samohlásky
    - 7.1_2_a – [á]
    - 7.1_2_e – [é]
    - 7.1_2_i – [í]

- 7.1_2_o – [ó]
- 7.1_2_u – [ú]

3. dlouhé samohlásky vysloveny co nejhlasitěji
   - 7.1_3_a – [á]
   - 7.1_3_e – [é]
   - 7.1_3_i – [í]
   - 7.1_3_o – [ó]
   - 7.1_3_u – [ú]
4. dlouhé samohlásky vysloveny co nejtišeji (ne šepot)
   - 7.1_4_a – [á]
   - 7.1_4_e – [é]
   - 7.1_4_i – [í]
   - 7.1_4_o – [ó]
   - 7.1_4_u – [ú]
5. globální příznaky používané pro diagnózu Parkinsonovy choroby
   - ParkinsonPr

9.2 Výpočet míry geometrické oddělitelnosti

V první fázi z celkové tabulky příznaků extrahujeme pouze mužské (zdravé a nemocné) řečníky. Dále provedeme výpočet míry geometrické oddělitelnosti, pro každý příznak.

Příklad výpočtu kvality příznaku $F0\_mean$ samohlásky [a]:
Třídy máme pouze dvě: jednotlivci s Parkinsonovou chorobou a kontrolní (zdraví) jednotlivci. Nejprve si vypočteme střední hodnoty příznaku v obou třídách. Podle vztahu 6.5 vypočítáme kvadrát vzdálenosti mezi jejich středními hodnotami.

$$D^2_{p,k} = (\underline{\mu}_p - \underline{\mu}_k)^2 = 128{,}1040 - 118{,}1334 = 99{,}4129$$

Vztah 6.6 nemusíme uvažovat, protože počítá se všemi možnými kombinacemi tříd, v našem případě jsou třídy pouze dvě, přičemž kombinace tříd p–k a k–p se považují za jednu a tu samou kombinaci. Z toho vyplývá, že kombinace je jen jedna. Dále jsou vypočítány kvadráty rozptylu podle vztahu 6.7.

$$S^2_p = (\underline{x} - \underline{\mu}_p)^2 = 394{,}8414$$

$$S^2_k = (\underline{x} - \underline{\mu}_k)^2 = 154{,}5657$$

Poté z jednotlivých získaných hodnot vypočítáme aritmetickou střední hodnotu podle vztahu 6.8.

$$S^2 = \frac{1}{V}\sum_{v=1}^{V} S_v^2 = \frac{1}{2}\sum_{v=1}^{2} S_v^2 = 274{,}7036$$

Pro určení kvality zvoleného příznaku, dosadíme předchozí vypočítané hodnoty do vzorce 6.9.

$$Q(F0\_\text{mean}) = \frac{S^2}{S^2 + D^2} = \frac{274{,}7036}{274{,}7036 + 99{,}4129} = 0{,}7343$$

Jednotlivé vypočítané hodnoty jsou zapsány v tabulkách 9.2 a 9.3.

Jak již bylo dříve řečeno, čím více se hodnota míry geometrické oddělitelnosti blíží hodnotě 0, jedná se o vhodný příznak. Pokud se hodnota blíží k 1, tak hodnoty obou tříd se navzájem prolínají a způsobují chaos.

Z tabulky 9.3 můžeme vyzdvihnout z hlediska kvality příznaky: voicing_frac, F1_std (oba v labelu 7.1_1_a); F1_var, F1_std, F1_max, F1_max_min (v labelu 7.1_1_u); voicing_frac, hnr_hn, F1_std(label 7.1_2_a); F1b_med (label 7.1_2_e); jitter_local, jitter_localabs (label 7.1_2_o); hnr_hn, F1_max (label 7.1_2_u); F1_std (label 7.1_3_a); jitter_ddp, jitter_rap, voicing_frac (label 7.1_3_e); jitter_ddp, jitter_rap (label 7.1_3_i); voicing_frac (label 7.1_3_o); F1b_std, F1b_max, hnr_hn (label 7.1_3_u).

Podle této metody se jako nejlepší příznak jeví voicing_frac u dlouhé samohlásky [o] vyslovené co nejhlasitěji.

Tabulka 9.2 udává globální příznaky používané při diagnóze Parkinsonovy choroby. Z těchto příznaků je nejlepší příznak F2i/F2u, ale jeho hodnota Q vychází oproti ostatním příznakům z tabulky 9.3 na průměrné úrovni. U ostatních těchto příznaků se hodnoty výrazně blíží 1, což udává že tyto příznaky nejsou vhodné k diagnóze. To může být způsobeno příliš malou databází a tudíž špatným rozlišením mezi třídami zdravých a nemocných pacientů.

Tab. 9.2: Míra geometrické oddělitelnosti globálních příznaků používaných k diagnóze PCH

	Parkinson_pr
VSA	0,871017
lnVSA	0,889456
FCR	0,69414
VAI	0,674558
F2i/F2u	0,463018

Tab. 9.3: Míra geometrické oddělitelnosti

	7.1_1_a	7.1_1_e	7.1_1_i	7.1_1_o	7.1_1_u	7.1_2_a	7.1_2_e	7.1_2_i	7.1_2_o	7.1_2_u	7.1_3_a	7.1_3_e	7.1_3_i	7.1_3_o	7.1_3_u	7.1_4_a	7.1_4_e	7.1_4_i	7.1_4_o	7.1_4_u
F0_median	0,7343	0,7410	0,8039	0,8505	0,9087	0,7640	0,7191	0,6917	0,6707	0,9775	0,9917	0,9794	0,9979	0,9868	0,7958	0,3308	0,3566	0,4087	0,3918	0,3543
F0_mean	0,7147	0,8470	0,8510	0,5727	0,7121	0,7839	0,7229	0,7029	0,6970	0,9719	0,7963	0,9807	0,9990	0,9749	0,7504	0,2312	0,3064	0,3844	0,4041	0,3764
F0_std	0,9031	0,9960	0,9883	0,6763	0,6747	0,7061	0,6332	0,7229	0,5989	0,5301	0,5447	0,9101	0,8345	0,9961	0,5274	0,2856	0,5041	0,4655	0,5437	0,6592
F0_min	0,8789	0,9531	0,9313	0,9365	0,9849	0,9723	0,9673	0,9817	0,9896	0,9529	0,9988	0,9611	0,9310	0,9278	0,4227	0,9623	0,8044	0,7940	0,7427	0,7103
F0_max	0,8325	0,9841	0,9227	0,6098	0,6043	0,7463	0,6340	0,7259	0,7232	0,8127	0,4564	0,9916	0,9857	0,9699	0,9985	0,2094	0,4488	0,4051	0,5588	0,4780
F0_max_min	0,8978	0,9990	0,9765	0,6571	0,6385	0,6442	0,6221	0,6097	0,7153	0,6401	0,4988	0,8441	0,8629	0,9983	0,5488	0,2482	0,4985	0,5037	0,6341	0,6356
relF0VR	0,9151	0,9873	0,9837	0,6480	0,6457	0,6731	0,6178	0,6528	0,7512	0,6323	0,5294	0,7546	0,8051	0,9952	0,4005	0,2650	0,5276	0,6039	0,6528	0,7638
relF0SD	0,9202	1,0000	0,9901	0,6630	0,6726	0,7318	0,6374	0,7599	0,6380	0,5406	0,5747	0,8355	0,7982	0,9915	0,3956	0,2788	0,5408	0,5767	0,5878	0,7976
voicing_frac	0,2703	0,4352	0,3757	0,3997	0,3146	0,2660	0,6766	0,3385	0,5499	0,3829	0,1705	0,1994	0,2266	0,0516	0,2223	0,4289	0,4300	0,2966	0,2145	0,3428
jitter_local	0,5198	0,4570	0,4020	0,5453	0,8048	0,4959	0,6427	0,4583	0,2336	0,3382	0,2979	0,2245	0,1642	0,2427	0,2367	0,4377	0,3652	0,2449	0,2965	0,4653
jitter_localabs	0,6228	0,5433	0,4814	0,7006	0,8933	0,6264	0,7031	0,5069	0,2560	0,4082	0,3622	0,3433	0,2852	0,3255	0,2635	0,5380	0,4270	0,2997	0,3807	0,5016
jitter_rap	0,4147	0,4038	0,3355	0,4483	0,6468	0,5367	0,5944	0,4532	0,2978	0,2786	0,3086	0,1982	0,1503	0,2351	0,2320	0,3962	0,3348	0,1866	0,3200	0,4636
jitter_ppq5	0,4895	0,4443	0,3932	0,4543	0,5523	0,3932	0,3448	0,5057	0,3474	0,3673	0,3407	0,2319	0,2034	0,2609	0,3015	0,3871	0,4141	0,2074	0,2975	0,4422
jitter_ddp	0,4146	0,4038	0,3354	0,4484	0,6467	0,5362	0,3245	0,4533	0,2978	0,2785	0,3086	0,1978	0,1502	0,2352	0,2321	0,3962	0,3349	0,1866	0,3200	0,4659
hnr_aut	0,6823	0,6621	0,3978	0,4817	0,4270	0,2703	0,5334	0,4107	0,4384	0,3124	0,3794	0,6254	0,3965	0,2587	0,3368	0,4096	0,2202	0,3607	0,3194	0,3919
hnr_nh	0,7291	0,6264	0,4169	0,5094	0,4366	0,3474	0,6039	0,4093	0,5088	0,3647	0,3588	0,6654	0,3383	0,2823	0,3784	0,4599	0,2362	0,3583	0,3481	0,4303
hnr_hn	0,5755	0,7053	0,4942	0,3865	0,4517	0,2135	0,3487	0,3777	0,2276	0,1710	0,2950	0,4499	0,6543	0,2201	0,1544	0,3304	0,2067	0,3886	0,2415	0,3070
F1_mean	0,9952	0,7708	0,7523	0,9991	0,3430	0,9601	0,9536	0,5567	0,8579	0,6184	0,8000	0,7686	0,7199	0,5387	0,6044	0,9574	0,8264	0,7099	0,8698	0,6807
F1_var	0,3934	0,4762	0,8935	0,6380	0,2894	0,3590	0,4950	0,7504	0,5863	0,3758	0,2424	0,7074	0,5868	0,3175	0,3896	0,4252	0,6581	0,5203	0,3753	0,4203
F1_std	0,2999	0,5575	0,7806	0,4680	0,2387	0,2544	0,5435	0,7524	0,4236	0,2828	0,1662	0,7669	0,6180	0,2386	0,2966	0,3574	0,6428	0,4573	0,3297	0,3774
F1_max	0,4594	0,9067	0,9998	0,5448	0,2337	0,4577	0,7653	0,8797	0,5531	0,2307	0,4607	0,9987	0,8440	0,4613	0,2814	0,4593	0,9213	0,9608	0,3197	0,3894
F1_min	0,8275	0,7281	0,6492	0,4849	0,6555	0,3093	0,8242	0,4673	0,3520	0,6076	0,3460	0,9570	0,8891	0,2449	0,3469	0,8566	0,8110	0,5310	0,3609	0,7032
F1_med	0,9335	0,8971	0,9125	0,9638	0,9799	0,8688	0,9810	0,6910	0,7813	0,9182	0,5466	0,5792	0,9407	0,4404	0,9757	1,0000	0,9125	0,8051	0,7360	0,9073
F1_max_min	0,4220	0,8400	0,6905	0,4888	0,2403	0,3396	0,7560	0,7719	0,4572	0,2406	0,2030	0,9984	0,8296	0,3218	0,2762	0,4775	0,9302	0,5446	0,2892	0,3953
F1b_mean	0,5363	0,6864	0,8523	0,3782	0,3622	0,5515	0,5253	0,9183	0,5804	0,5575	0,2073	0,3433	0,7568	0,3627	0,7462	0,5804	0,6066	0,9091	0,3384	0,6575
F1b_var	0,9008	0,8160	0,9969	0,9371	0,5390	0,7495	0,8077	0,9999	0,9902	0,8546	0,8191	0,6256	0,8612	0,7431	0,1857	0,7464	0,9750	0,8762	0,5516	0,9458
F1b_std	0,8478	0,8385	0,9736	0,6426	0,3958	0,5941	0,8508	0,9977	0,9633	0,7196	0,6407	0,6559	0,6782	0,4323	0,1324	0,9220	0,9120	0,6926	0,3676	0,7261
F1b_max	0,8773	0,8630	0,9950	0,7262	0,4563	0,5537	0,9953	0,9710	0,9989	0,8904	0,7950	0,8810	0,6982	0,6537	0,1530	0,9790	0,9805	0,5070	0,3277	0,6260
F1b_min	0,9899	0,5512	0,3546	0,9982	0,8635	0,9987	0,4489	0,4802	0,8536	0,9480	0,8761	0,2830	0,5824	0,7066	0,9636	0,8101	0,3996	0,6622	0,9188	0,9791
F1b_med	0,4564	0,5970	0,6015	0,5633	0,5498	0,5991	0,2923	0,7605	0,8413	0,8660	0,5680	0,3501	0,9084	0,9996	0,9142	0,4190	0,4960	0,9202	0,5438	0,7656
F1b_max_min	0,8769	0,7806	0,9998	0,7291	0,4553	0,5512	0,9556	0,9828	0,9993	0,8863	0,7899	0,9315	0,6682	0,6467	0,1492	0,9660	0,9991	0,4882	0,3372	0,6291
F2_mean	0,7206	0,6419	0,9784	0,5154	0,4313	0,9079	0,7942	0,9268	0,5562	0,5419	0,9236	0,7417	0,9973	0,9964	0,9178	0,5326	0,6875	0,9971	0,5439	0,9388
F2_var	0,6586	0,7565	0,5279	0,9996	0,9790	0,6987	0,7210	0,9298	0,9882	0,9852	0,4008	0,7428	0,8755	0,3868	0,5964	0,6835	0,8006	0,3902	0,9163	0,9967
F2_std	0,6568	0,7032	0,4199	0,9820	0,9985	0,7694	0,6547	0,8691	0,9917	0,9731	0,3397	0,6935	0,8571	0,3249	0,4154	0,5778	0,7998	0,4168	0,9815	0,9114
F2_max	0,5802	0,7056	0,6496	0,7390	0,3847	0,7698	0,7167	0,9807	0,4819	0,4460	0,4389	0,9810	0,9957	0,4735	0,5100	0,9992	0,3228	0,4134	0,5013	0,3153
F2_min	0,9975	0,7811	0,4539	1,0000	0,5970	0,9184	0,8978	0,6034	0,8419	0,6192	0,7221	0,9179	0,7686	0,3773	0,7671	0,4500	0,6617	0,6235	0,7314	0,8767
F2_med	0,7016	0,8523	0,9771	0,5467	0,5677	0,8556	0,8706	0,9668	0,5000	0,7990	0,7747	0,9769	0,9692	0,9986	0,9640	0,6509	0,8046	0,9860	0,6133	0,9983
F2_max_min	0,8566	0,8279	0,4162	0,9103	0,9672	0,8147	0,7860	0,8307	0,9958	0,9983	0,4322	0,9309	0,9302	0,3228	0,4032	0,5998	0,8109	0,3659	0,9819	0,6014
F2b_mean	0,9999	0,6954	0,9486	0,9222	0,9687	0,9194	0,9916	0,9830	0,9020	0,9892	0,8279	0,9875	0,9897	0,4357	0,3501	0,7025	0,7877	0,9179	0,8316	0,9828
F2b_var	0,8664	0,7769	0,9975	0,8143	0,7842	0,9759	0,7365	0,9833	0,8909	0,6938	0,9355	0,6325	0,9045	0,7814	0,6589	0,5185	1,0000	0,8618	0,5768	0,5670
F2b_std	0,8140	0,5747	0,9511	0,8910	0,7846	0,9945	0,5807	0,9841	0,9213	0,7896	0,9870	0,5631	0,8691	0,6191	0,5036	0,5240	0,9059	0,7218	0,6512	0,6163
F2b_max	0,5862	0,6080	0,9740	0,9979	0,9210	0,9596	0,6552	0,9646	0,9856	0,9105	0,9996	0,5710	0,9068	0,8139	0,7557	0,7820	0,8345	0,6745	0,9624	0,7724
F2b_min	0,9326	0,9718	0,8911	0,9759	0,7806	0,6618	0,3862	0,9950	0,9903	0,6360	0,9379	0,5678	0,7350	0,9249	0,9016	0,8743	0,9474	0,9977	0,9761	0,4220
F2b_med	0,9426	0,8258	0,9759	0,9733	0,5868	0,9660	0,9417	0,9969	0,9177	0,8973	0,7390	0,8759	0,9828	0,3916	0,5410	0,8787	0,7962	0,9783	0,9760	0,8353
F2b_max_min	0,5724	0,5882	0,9665	0,9988	0,9003	0,9455	0,6012	0,9658	0,9877	0,8751	0,9985	0,5183	0,8842	0,8080	0,7355	0,7998	0,8166	0,6787	0,9596	0,6772

9.3 Výběr příznaků mRMR

V této části byl použit výběr příznaků metodou „minimum-redundancy maximum-relevancy“, která je součástí The mutual information computing toolboxu [10] v programu Matlab, který vypočítal pomocí vzájemné informace 30 nejlepších příznaků z celkových 455.

9.3.1 Trénovací data

Jako trénovací databáze je použita matice 30 nejlepších příznaků vybraných pomocí metody mRMR od devíti pacientů s Parkinsonovou chorobou a devíti kontrolních mluvčích. Omezení této databáze je pouze na mužské jedince.

9.3.2 Testovací data

Testovací databáze obsahuje dva kontrolní řečníky a dva řečníky s Parkinsonovou chorobou. Při dalším trénování byl vyměněn jeden mluvčí z trénovací množiny za jednoho mluvčího z testovací množiny.

9.3.3 Selekce metodou mRMR a klasifikace pomocí kNN

Následně byl proveden výpočet pomocí klasifikátoru kNN, pro zjištění úspěšnosti klasifikace vybraných příznaků z metody mRMR.

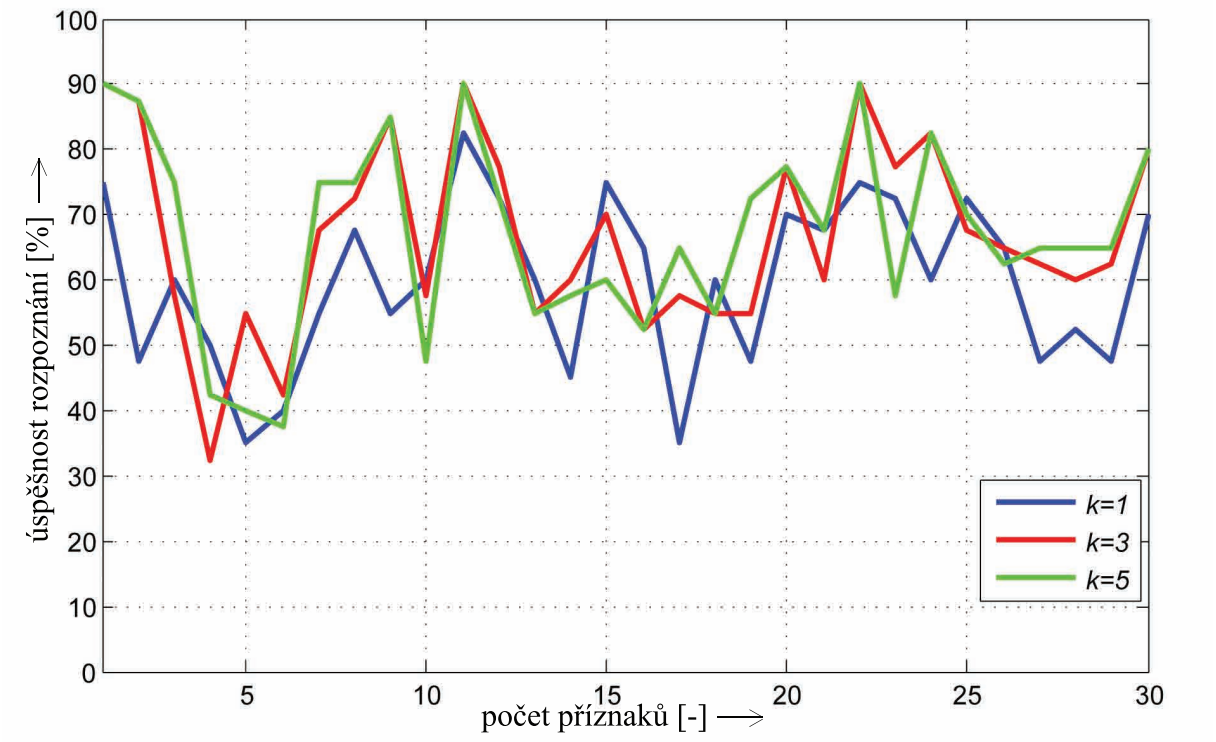
U klasifikátoru kNN s nastaveným paremetrem na k=3 znamená, že konkrétní testovací jedinec se přiřadil ke skupině (Parkinson, kontrolní), která měla nejblíže této testované hodnoty více bodů. Kde k je většinou voleno jako liché číslo aby nedošlo ke shodě. Proto také jsou vyrovnány jednotlivé trénovací množiny, aby nedošlo k chybě přiřazováním do množiny s více hodnotami.

Tabulka 9.4 ukazuje odchylky jednotlivých procentuálních úspěšností mezi různými nastaveními klasifikátoru kNN. Konkrétně na k=1, k=3 a k=5. Tyto rozdíly byly následně vyneseny do grafu 9.1. Z tohoto grafu lze vidět, že při použití k=1, což je přiřazení testovacího jedince k množině, jejíž hodnota je nejblíže k hodnotě testované, je úspěšnost menší a výkyvy větší než u k=3 a k=5.

Při porovnávání klasifikátoru kNN s klasifikátorem GMM, bylo nastaveno k=3. Výsledné hodnoty byly zapisovány do tabulky 9.5. Klasifikátor kNN je oproti GMM výpočetně a tím i časově méně náročnější.

Tab. 9.4: Procentuální úspěšnost přiřazení testovacího jedince ke správné množině klasifikátorem kNN

ID příznaku	název příznaku	label příznaku	úspěšnost k=1 [%]	úspěšnost k=3 [%]	úspěšnost k=5 [%]
12	'jitter_rap'	'7.1_1_a'	75	90	90
189	'voicing_frac'	'7.1_1_u'	47,5	87,5	87,5
414	'voicing_frac'	'7.1_2_u'	60	57,5	75
99	'voicing_frac'	'7.1_1_i'	50	32,5	42,5
369	'voicing_frac'	'7.1_2_o'	35	55	40
324	'voicing_frac'	'7.1_2_i'	40	42,5	37,5
279	'voicing_frac'	'7.1_2_e'	55	67,5	75
9	'voicing_frac'	'7.1_1_a'	67,5	72,5	75
54	'voicing_frac'	'7.1_1_e'	55	85	85
144	'voicing_frac'	'7.1_1_o'	60	57,5	47,5
234	'voicing_frac'	'7.1_2_a'	82,5	90	90
103	'jitter_ppq5'	'7.1_1_i'	72,5	77,5	72,5
237	'jitter_rap'	'7.1_2_a'	60	55	55
328	'jitter_ppq5'	'7.1_2_i'	45	60	57,5
145	'jitter_local'	'7.1_1_o'	75	70	60
327	'jitter_rap'	'7.1_2_i'	65	52,5	52,5
238	'jitter_ppq5'	'7.1_2_a'	35	57,5	65
239	'jitter_ddp'	'7.1_2_a'	60	55	55
418	'jitter_ppq5'	'7.1_2_u'	47,5	55	72,5
282	'jitter_rap'	'7.1_2_e'	70	77,5	77,5
373	'jitter_ppq5'	'7.1_2_o'	67,5	60	67,5
14	'jitter_ddp'	'7.1_1_a'	75	90	90
148	'jitter_ppq5'	'7.1_1_o'	72,5	77,5	57,5
10	'jitter_local'	'7.1_1_a'	60	82,5	82,5
58	'jitter_ppq5'	'7.1_1_e'	72,5	67,5	70
147	'jitter_rap'	'7.1_1_o'	65	65	62,5
192	'jitter_rap'	'7.1_1_u'	47,5	62,5	65
280	'jitter_local'	'7.1_2_e'	52,5	60	65
194	'jitter_ddp'	'7.1_1_u'	47,5	62,5	65
372	'jitter_rap'	'7.1_2_o'	70	80	80



Obr. 9.1: Úspěšnost rozpoznání jednotlivých příznaků klasifikátorem kNN s různým nastavením k.

9.3.4 Selekce metodou mRMR a klasifikace pomocí GMM

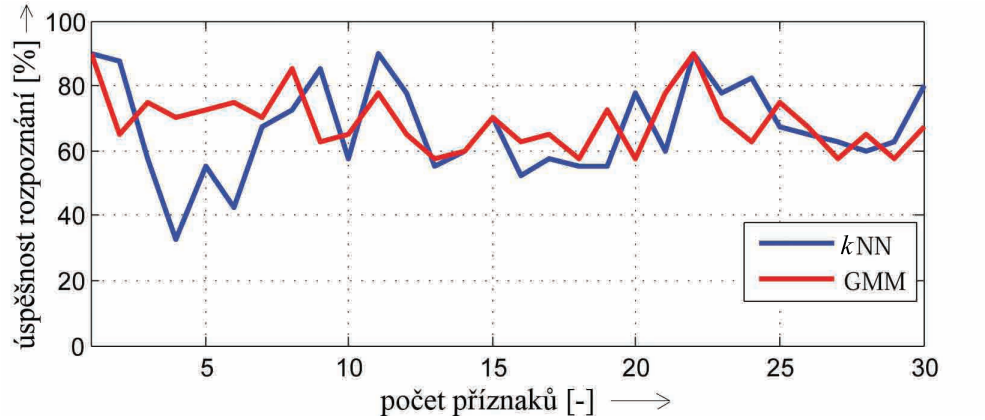
Ke klasifikaci byl použit GMM toolbox [8].

Postup výběru příznaků je podobný tomu uvedenému v kap. 9.3.3 s tím rozdílem že se použije klasifikátor GMM s diagonální kovarianční maticí. U tohoto druhu klasifikace je ještě před samotným přiřazováním testovacího jedince k množině, nutné trénování klasifikátoru. Počet Gaussových směsí byl nastaven na 1 a počet iterací během trénování byl zvolen 50. Výsledné údaje jsou zapsány v tabulce 9.5 a následně vyneseny do grafu 9.2.

Z dosažených výsledků vyplývá, že použitím GMM klasifikátoru je úspěšnost rozpoznání jedince vyšší a rovnoměrnější než při použití klasifikárotu kNN. Dále lze z tabulky vyčíst, že nejlepší příznaky vybrané metodou mRMR jsou typu „voicing“ a „jitter“.

Tab. 9.5: Procentuální úspěšnost přiřazení testovacího jedince ke správné množině klasifikátorem kNN a GMM

ID příznaku	název příznaku	label příznaku	úspěšnost přiřazení (kNN) [%]	úspěšnost přiřazení (GMM) [%]
12	'jitter_rap'	'7.1_1_a'	90	90
189	'voicing_frac'	'7.1_1_u'	87,5	65
414	'voicing_frac'	'7.1_2_u'	57,5	75
99	'voicing_frac'	'7.1_1_i'	32,5	70
369	'voicing_frac'	'7.1_2_o'	55	72,5
324	'voicing_frac'	'7.1_2_i'	42,5	75
279	'voicing_frac'	'7.1_2_e'	67,5	70
9	'voicing_frac'	'7.1_1_a'	72,5	85
54	'voicing_frac'	'7.1_1_e'	85	62,5
144	'voicing_frac'	'7.1_1_o'	57,5	65
234	'voicing_frac'	'7.1_2_a'	90	77,5
103	'jitter_ppq5'	'7.1_1_i'	77,5	65
237	'jitter_rap'	'7.1_2_a'	55	57,5
328	'jitter_ppq5'	'7.1_2_i'	60	60
145	'jitter_local'	'7.1_1_o'	70	70
327	'jitter_rap'	'7.1_2_i'	52,5	62,5
238	'jitter_ppq5'	'7.1_2_a'	57,5	65
239	'jitter_ddp'	'7.1_2_a'	55	57,5
418	'jitter_ppq5'	'7.1_2_u'	55	72,5
282	'jitter_rap'	'7.1_2_e'	77,5	57,5
373	'jitter_ppq5'	'7.1_2_o'	60	77,5
14	'jitter_ddp'	'7.1_1_a'	90	90
148	'jitter_ppq5'	'7.1_1_o'	77,5	70
10	'jitter_local'	'7.1_1_a'	82,5	62,5
58	'jitter_ppq5'	'7.1_1_e'	67,5	75
147	'jitter_rap'	'7.1_1_o'	65	67,5
192	'jitter_rap'	'7.1_1_u'	62,5	57,5
280	'jitter_local'	'7.1_2_e'	60	65
194	'jitter_ddp'	'7.1_1_u'	62,5	57,5
372	'jitter_rap'	'7.1_2_o'	80	67,5



Obr. 9.2: Závislost úspěšnosti rozpoznání jednotlivých příznaků pomocí různých klasifikátorů.

9.4 Výběr příznaků metodou SFFS

Na rozdíl od předchozího výběru příznaků (mRMR), kde se pomocí klasifikátorů pouze ověřovala úspěšnost, předem vybraných příznaků, se u sekvenčního dopřednho plovoucího výběru používá jednotlivých klasifikátorů pro výběr těchto příznaků.

9.4.1 Trénovací data

Jako trénovací databáze je použita matice všech 455 příznaků, která se složí ze všech částí s promluvou jednotlivých samohlásek (labelů) a globálních příznaků sloužících pro rozpoznánvání Parkinsonovy choroby (VSA, lnVSA, FCR, VAI, F2i/F2u), od devíti pacientů s Parkinsonovou chorobou a devíti kontrolních mluvčích. Omezení této databáze je pouze na mužské jedince.

9.4.2 Testovací data

Testovací databáze obsahuje dva kontrolní řečníky a dva řečníky s Parkinsonovou chorobou. Při dalším trénování byl vyměněn jeden mluvčí z trénovací množiny za jednoho mluvčího z testovací množiny.

V prvním kroku SFFS se projdou postupně všechny příznaky (455) jeden po druhém a vypočítá se úspěšnost přiřazení testovacích řečníků ke správné množině. Odtud se poté vybere příznak s nejlepší úspěšností, tento příznak se odstraní z trénovací databáze. V dalším kroku se postupně k nejúspěšnějšímu příznaku přiřazují další příznaky do dvojice. Nyní se zjišťuje nejúspěšnější dvojice. V dalším kroku trojice a tak dále až do počtu příznaků kterou potřebujeme. Pro tyto účely je maximální počet sloučených příznaků 30.

9.4.3 SFFS s klasifikátorem kNN

Klaslifikátor kNN je nastaven na k=3. Výsledné hodnoty a příznaky jsou zapsány v tabulce 9.6, kde „krok“ značí konkrétní krok v SFFS, a příznak který byl vybrán v tomto kroku jako nejúspěšnější a přiřazen k předchozím vybraným příznakům.

Z tabulky 9.6 je možné zjistit zásadní problém této metody výběru příznaků. Když se hodnota nejvyšší úspešnosti v konkrétním kroku SFFS nachází u dvou nebo více příznaků, vybere se hned první příznak s touto nejvyšší hodnotou. To způsobí, že jako první příznaky jsou vybrané příznaky z prvních labelů, jak můžeme vidět v tabulce 9.6. Tento problém je způsoben hlavně tím, že je příliš malá databáze mluvčích a možnost, že se vyskytnou dvě stejné procentuální úspěšnosti je velmi vysoká.

9.4.4 SFFS s klasifikátorem GMM

Výběr příznaků SFFS s klasifikátorem GMM je výpočetně náročnější než SFFS s kNN. GMM je nutné před testováním natrénovat zatímco kNN trénování nepotřebuje. Čím větší je krok SFFS tím je delší trénování. Proto je pro tuto metodu vybráno pouze 20 kroků výběru příznaků. Jednotlivé kroky jsou zapsány v tabulce 9.7. V tabulce se od trojice příznaků dále objevuje úspěšnost rozpoznání 100%. Toto je způsobené opět malou databází řečníků.

Při pokusu klasifikace se dvěmi testovacími osobami (časová náročnost je na únosné míře), kde proběhl algoritmus SFFS až do konce, se při vysokém počtu sloučených příznaků úspěšnost začala snižovat. Při sloučení všech příznaků byla úspěšnost na nejnižší úrovni. Z toho plyne, že není vždy ideální používat velké množství příznaků pro diagnózu. Je lepší použít méně příznaků, ale zato kvalitnějších.

9.4.5 První krok SFFS

První krok SFFS vybírá každý příznak samostatně jeden po druhém a zapisuje jeho úspěšnost přiřazení. Tabulky 9.9 a 9.10 zaznamenávají deset nejlepších příznaků podle klasifikace kNN a GMM. Z hodnot v těchto tabulkách je vidět, že úspěšnost je vyšší u metody klasifikace kNN. Při výpočtu aritmetického průměru všech hodnot úspěšností 1. kroku SFFS, dosahuje klasifikace kNN hodnoty 60,56 % a klasifikace GMM hodnoty 55,57 %. V tomto experimentu tedy byla klasifikace kNN úspěšnější než GMM klasifikace.

V tabulce 9.8 je uvedena procentuální úspěšnost klasifikací kNN a GMM, u globálních příznaků používaných k diagnóze Parkinsonovy choroby. Z výsledků je možné vidět, že tato úspěšnost je velmi malá. Pouze u příznaku „F2i/F2u“ a „lnVSA“ je úspěšnost na přijatelné úrovni. V porovnání klasifikací kNN a GMM je u těchto

příznaků úspěšnější GMM, ale příznaky „VSA“, „FCR“, „VAI“ jsou prakticky nepoužitelné. Přitom by tyto příznaky na základě literatury měli dosahovat nejlepších hodnot. To může být způsobeno příliš malou databází řečníků.

Tab. 9.6: Příznaky vybrané v jednotlivých krocích SFFS s klasifikátorem kNN

krok	ID příznaku	název příznaku	název labelu	úspěšnost [%]
1	83	'F2_max_min'	'7.1_1_e'	92,5
2	1	'F0_median'	'7.1_1_a'	92,5
3	2	'F0_mean'	'7.1_1_a'	92,5
4	3	'F0_std'	'7.1_1_a'	92,5
5	4	'F0_min'	'7.1_1_a'	92,5
6	5	'F0_max'	'7.1_1_a'	92,5
7	6	'F0_max_min'	'7.1_1_a'	92,5
8	7	'relF0VR'	'7.1_1_a'	92,5
9	8	'relF0SD'	'7.1_1_a'	92,5
10	9	'voicing_frac'	'7.1_1_a'	92,5
11	10	'jitter_local'	'7.1_1_a'	92,5
12	11	'jitter_localabs'	'7.1_1_a'	92,5
13	12	'jitter_rap'	'7.1_1_a'	92,5
14	13	'jitter_ppq5'	'7.1_1_a'	92,5
15	14	'jitter_ddp'	'7.1_1_a'	92,5
16	15	'hnr_aut'	'7.1_1_a'	92,5
17	16	'hnr_nh'	'7.1_1_a'	92,5
18	17	'hnr_hn'	'7.1_1_a'	92,5
19	46	'F0_median'	'7.1_1_e'	92,5
20	47	'F0_mean'	'7.1_1_e'	92,5
21	48	'F0_std'	'7.1_1_e'	92,5
22	49	'F0_min'	'7.1_1_e'	92,5
23	50	'F0_max'	'7.1_1_e'	92,5
24	51	'F0_max_min'	'7.1_1_e'	92,5
25	52	'relF0VR'	'7.1_1_e'	92,5
26	53	'relF0SD'	'7.1_1_e'	92,5
27	54	'voicing_frac'	'7.1_1_e'	92,5
28	55	'jitter_local'	'7.1_1_e'	92,5
29	56	'jitter_localabs'	'7.1_1_e'	92,5
30	57	'jitter_rap'	'7.1_1_e'	92,5

Tab. 9.7: Příznaky vybrané v jednotlivých krocích SFFS s klasifikátorem GMM

krok	ID příznaku	název příznaku	název labelu	úspěsnost [%]
1	180	'F2b_max_min'	'7.1_1_o'	92,5
2	422	'hnr_hn'	'7.1_2_u'	95
3	32	'F2_mean'	'7.1_1_a'	100
4	40	'F2b_var'	'7.1_1_a'	100
5	96	'F0_max_min'	'7.1_1_i'	100
6	42	'F2b_max'	'7.1_1_a'	100
7	23	'F1_med'	'7.1_1_a'	100
8	2	'F0_mean'	'7.1_1_a'	100
9	10	'jitter_local'	'7.1_1_a'	100
10	3	'F0_std'	'7.1_1_a'	100
11	1	'F0_median'	'7.1_1_a'	100
12	11	'jitter_localabs'	'7.1_1_a'	100
13	18	'F1_mean'	'7.1_1_a'	100
14	5	'F0_max'	'7.1_1_a'	100
15	13	'jitter_ppq5'	'7.1_1_a'	100
16	7	'relF0VR'	'7.1_1_a'	100
17	24	'F1_max_min'	'7.1_1_a'	100
18	12	'jitter_rap'	'7.1_1_a'	100
19	8	'relF0SD'	'7.1_1_a'	100
20	14	'jitter_ddp'	'7.1_1_a'	100

Tab. 9.8: Úspěšnost rozpoznání příznaků sloužících k diagnóze Parkinsonovy choroby

ID příznaku	název příznaku	úspěšnost kNN [%]	úspěšnost GMM [%]
451	VSA	40	32,5
452	lnVSA	42,5	82,5
453	FCR	40	22,5
454	VAI	40	22,5
455	F2i/F2u	65	82,5

Tab. 9.9: 10 nejlepších příznaků vybraných 1.krokem SFFS pomocí klasifikátoru kNN

pořadí	ID příznaku	název příznaku	label	úspěšnost [%]
1.	83	F2_max_min	7.1_1_e	92,5
2.	154	F1_var	7.1_1_o	92,5
3.	377	hnr_hn	7.1_2_o	92,5
4.	422	hnr_hn	7.1_2_u	92,5
5.	12	jitter_rap	7.1_1_a	90
6.	15	hnr_aut	7.1_1_a	90
7.	234	voicing_frac	7.1_2_a	90
8.	244	F1_var	7.1_2_a	90
9.	245	F1_std	7.1_2_a	90
10.	300	F1b_med	7.1_2_e	90

Tab. 9.10: 10 nejlepších příznaků vybraných 1.krokem SFFS pomocí klasifikátoru GMM

pořadí	ID příznaku	název příznaku	label	úspěšnost [%]
1.	180	F2b_max_min	7.1_1_o	92,5
2.	219	F2b_mean	7.1_1_u	92,5
3.	14	jitter_ddp	7.1_1_a	90
4.	44	F2b_med	7.1_1_a	90
5.	114	F1_max_min	7.1_1_i	90
6.	128	F2_max_min	7.1_1_i	90
7.	167	F2_mean	7.1_1_o	90
8.	177	F2b_max	7.1_1_o	90
9.	216	F2_min	7.1_1_o	90
10.	250	F1b_mean	7.1_1_u	90

10 ZÁVĚR

Tato práce se zaměřuje na rozpoznávání Parkinsonovy choroby z řečového signálu. K těmto účelům se využívá databáze nahrávek pacientů postižených Parkinsonovou chorobou, která byla zaznamenána na pracovišti I. neurologické kliniky ve Fakultní nemocnici u sv. Anny. V úvodních kapitolách jsou základní informace o tvorbě řeči, z jakých ústrojí se skládá hlasový trakt, odlišnosti, které má řeč pacientů s Parkinsonovou chorobou a stručný popis zpracování řečových signálů, kam patří: ustředění, preemfáze a segmentace.

V kapitole 4 jsou popsány příznaky LPC, MFCC, PLP, LPCC a jejich výhody. Avšak příznaky, které jsou stěžejní pro diagnózu Parkinsonovy choroby (např. VAI, VSA, VOT, FCR,...), jsou popsány v kapitole 5.

V další části je vysvětlen princip míry geometrické oddělitelnosti. Dále jsou popsány metody pro výběr příznaků mRMR a SFFS. Poslední teoretická kapitola 7 je zaměřena na metody využívající statistické rozpoznávání vzorů a to metodu Gaussových smíšených modelů a metodu kNN.

Kapitola 8 obsahuje blokové schéma návrhu programu pro diagnózu Parkinsonovy choroby z řečového signálu a popis jednotlivých bloků.

V kapitole 9 je řešena praktická část diplomové práce. Nejprve je vypočítána míra geometrické oddělitelnosti, která vyšla nejlépe pro příznak vyjadřující podíl znělých úseků v řečovém signálu u dlouhé samohlásky [ó]. Nejlepší hodnoty u globálních příznaků používaných k diagnóze Parkinsonovy choroby vyšel nejlépe u příznaku „F2i/F2u“. U ostatních těchto příznaků vycházejí hodnoty nepříznivě, ale to může být způsobeno příliš malou databází řečníků.

Následující praktická část by se dala rozdělit do dvou částí, podle metody pro výběr příznaků. První část je zaměřena na výběr příznaků mRMR, který vybral 30 nejlepších příznaků. Úspěšnost těchto příznaků byla ověřována pomocí klasifikátorů kNN a GMM. Při tomto ověřování byla optimální metoda GMM. U kNN byl navíc proveden experiment s různým nastavením k, kde k=1 měla menší úspěšnost a velké výkyvy, k=3 a k=5 bylo téměř na stejné úrovni. V dalších částech bylo používáno k=3.

Druhá část byla zaměřena na výběr příznaků SFFS s klasifikátorem kNN a SFFS s klasifikátorem GMM. Při výběru touto metodou byl zjištěn zásadní problém, způsobený malou databází řečníků, při stejných úspěšnostech dvou příznaků, vybere tato metoda první s těchto příznaků. To znamená, že u malých databází řečníků je velká pravděpodobnost stejného procentuálního úspěchu a tím pádem vybírání příznaků z prvních labelů. Úspěšnější metodou byla GMM. Při průchodu 1. kroku SFFS algoritmu byl nejúspěšnější příznak rozdílu maxima a minima šířky pásma druhého formantu. Úspěšnost rozpoznání byla v prvním kroku vyšší u metody kNN, kde

průměrná hodnota byla 60,56%. U klasifikátoru GMM byl průměr 55,57%.

Pro diagnózu Parkinsonovy choroby bych doporučil příznaky vyjadřující podíl znělých úseků v řečovém signálu, směrodatnou odchylku prvního formantu, jitter_rap, jitter_ppq5, příznak určující rozdíl maximální a minimální hodnoty druhého formantu a příznak F2i/F2u. Tyto příznaky se v předchozích měření jevily jako nejúspěšnější pro diagnózu Parkinsonovy choroby. Reálná úspěšnost, při použití klasifikátorů kNN a GMM u „nejlepších“ příznaků, se pohybuje okolo hodnoty 90%. Dále bych pro diagnózu Parkinsonovy choroby doporučil výběr příznaků metodou SFFS za použití klasifikátoru GMM. Při experimentech dosahoval velmi dobrých výsledků, ale jeho hlavní nevýhodou je jeho časová náročnost.

Jako další postup práce by byla tvorba rozsáhlejší databáze řečových signálů, která by zaručovala dosažení lepších výsledků při klasifikaci. Dále by bylo vhodné otestovat více metod pro výběr příznaků, použití dalších klasifikátorů a použití nových příznaků. Dále by bylo vhodné porovnat rozdíly při diagnóze řečníků ženského a mužského pohlaví.

LITERATURA

[1] BISHOP, Ch. M. *Pattern Recognition and Machine Learning*. Cambridge: Springer Science+Business Media, 2006. 729 s. ISBN 0-387-31073-8.

[2] Černocký, J. *Zpracování řečových signálů*. Brno: FIT VUT v Brně, 2006. 129 s.

[3] DUDA, R. O., HART, P. E., STORK, D. G. *Pattern Classification*. New York: John Wiley & Sons Ltd, 2001. 637 s. ISBN 0-471-05669-3.

[4] FISCHER, E., GOBERMAN, A. M. *Voice onset time in Parkinson disease*. Journal of Communication Disorders. 2010, 43.1, s. 21–34, [cit. 2010-11-09]. Dostupný z WWW: <http://dx.doi.org/10.1016/j.jcomdis.2009.07.004>.

[5] GOBERMAN, A. M., ELMER, L. W. *Acoustic analysis of clear versus conversational speech in individuals with Parkinson disease*. Journal of Communication Disorders. 2010, 38.3, s. 215–30, [cit. 2010-11-09]. Dostupný z WWW: <http://dx.doi.org/10.1016/j.jcomdis.2004.10.001>.

[6] HUANG, X., ACERO, A., HON, H. *Spoken language processing: A Guide to Theory, Algorithm, and System Development*. New Jersey: Prentice–Hall, 2001. 980 s. ISBN 0-13-022616-5.

[7] LEVINSON, S. E. *Mathematical Models for Speech Technology*. Illinois: John Wiley & Sons Ltd, 2005. 257 s. ISBN 0-470-84407-8.

[8] NABNEY, I. T. *Copyright (c) Ian T Nabney (1996-2001).*

[9] PENG, H., LONG, F., DING, CH. *Feature Selection Based on Mutual Information:Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy*. IEEE Transaction on pattern analysis and machine intelligence,Vol. 27, No. 8, pp.1226-1238, 2005.

[10] PENG, H., et al. *mRMR (minimum-redundancy maximum-relevancy) feature selection method*. 2005 and Ding & Peng, 2005, 2003.

[11] PSUTKA, Josef, et al. *Mluvíme s počítačem česky*. Praha: ACADEMIA, 2006. 752 s. ISBN 80-200-1309-1.

[12] PUDIL, P., NOVOVIČOVÁ, J., KITTLER, J. *Floating Search Methods in Feature Selection*. New Jersey: Pattern Recognition Letters, 1994, 15, 1119–1125.

[13] RABINER, L. R., JUANG, B.–H. *Fundamentals of speech recognition*. New Jersey: Pattern Recognition Letters, 1993. 497 s. ISBN 0-13-285826-6.

[14] RABINER, L. R., SCHAFER, R. W. *Digital Processing of Speech Signals*. New Jersey: Prentice–Hall signal processing series, 1978. 512 s. ISBN 0-13-213603-1.

[15] SAPIR, S., RAMIG, L. O., SPIELMAN, J. L., FOX, C. *Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech*. J Speech Lang Hear Res. 2010, 53, s. 114–25, [cit. 2010-11-09]. Dostupný z WWW: <http://dx.doi.org/10.1044/1092-4388(2009/08-0184)>.

[16] *Scholarpedia* [online]. 2009 [cit. 2011-24-04]. Dostupný z WWW: <http://scholarpedia.org/article/K-nearest_neighbor>.

[17] SIGMUND, Milan. *Analýza řečových signálů: Přednášky*. Brno: MJ servis, spol. s r.o., 2000. 86 s. ISBN 80-214-1783-8.

[18] SKODDA, S., VISSER, W., SCHLEGEL, U. *Short- and long-term dopaminergic effects on dysarthria in early Parkinson's disease*. J Neural Transm. 2010, 117, s. 197–205, [cit. 2010-11-09]. Dostupný z WWW: <http://dx.doi.org/10.1007/s00702-009-0351-5>.

[19] SMÉKAL, Z. *Číslicové zpracování řeci: skripta*. Brno: Ústav telekomunikací FEKT VUT v Brně, 2009. 123 s.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

AR	Auto–Regressive
ATF	Amplitude Tremor Frequency
ATRI	Amplitude Tremor Intensity Index
F_0VR	F_0 Variation Range
FCR	Formant Centralization Ratio
FIR	Finite Impulse Response
FTRI	Frequency Tremor Intensity Index
GMM	Gaussian Mixture Models
ISD	Inter-pause Speech Duration
kNN	k-nearest neighbor algorithm
LPC	Linear Predictive Coding
LPCC	Linear Prediction Cepstral Coefficients
MFCC	Mel Frequency Cepstral Coefficients
PLP	Perceptual Linear Predictive
SBFS	Sequential Backward Floating Selection
SFFS	Sequential Floating Forward Selection
SPIR	Speech Index of Rhythmicity
TSR	Total Speech Rate
VAI	Vowel Articulation Index
VOT	Voice Onset Time
VSA	Vowel Space Area
a_k	lineární predikční koeficienty
$A(z)$	přenosová funkce analyzujícího filtru
$\mathbf{C_i}$	kovarianční matice

D^2	aritmetická střední hodnota vzdáleností mezi všemi třídami
$D^2_{v,u}$	kvadrát vzdálenosti středních hodnot dvou tříd
$e[n]$	chyba lineární predikce
$E(z)$	Z-transformace $e[n]$
f	frekvence
F_0	střední hodnota
f_m	frekvence v melovské škále
l_{ram}	konstantní délka rámce
$p_i(o)$	hustoty pravděpodobností jednotlivých složek
$Q(.)$	míra geometrické oddělitelnosti tříd
R	počet tříd
S^2	aritmetická střední hodnota
S^2_v	kvadrát rozptilu třídy v okolo střední hodnoty
$s[n]$	řečový signál
$s'[n]$	řečový signál bez stejnosměrné složky
$\tilde{s}[n]$	aktuální vzorek
w_i	váhy jednotlivých složek
$w[n]$	okénková funkce
$\underline{x}$	vektor příznaků
μ_i	střední hodnota hustoty pravděpodobnosti
$\underline{\mu}_k$	střední hodnota třídy kontrolních řečníků
$\underline{\mu}_p$	střední hodnota třídy řečníků s Parkinsonovou chorobou
$\underline{\mu}_v$	střední hodnota třídy v
λ	koeficient preemfáze

A PŘÍLOHA

Přiložené CD obsahuje:

- diplomovou práci ve rormátu pdf
- adresáře s jednotlivými částmi programu
  - Mira oddelitelnosti Q
  - mRMR GMM
  - mRMR KNN
  - SFFS GMM
  - SFFS KNN

Spouštění programu

Pro rozdělení souboru „Priznaky.xls“ na kontrolní muže a ženy, muže a ženy s Parkinsonovou chorobou, slouží skript Rozdeleni.m umístěný v adresáři „Mira oddelitelnosti Q“. Míra oddělitelnosti se vypočítá pomocí skriptu mira_oddelitelnosti.m.

K výběru příznaků metodou mRMR a klasifikace pomocí kNN i GMM slouží skript DEMO.m v konkrétních aresářích.

Pro výběr příznaků metodou SFFS pomocí klasifikace kNN i GMM slouží skript SFFS.m v konkrétních aresářích.

Jednotlivé dílčí programy byly testovány na programu Matlab R2008a s operačním systémem Windows 7 Professional (x86).