

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE

Brno, 2016

Radoslav Kárník



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

**DE-IDENTIFIKACE ŘEČNÍKŮ POSTIŽENÝCH
HYPOKINETICKOU DYSARTRIÍ**

DE-IDENTIFICATION OF SPEAKERS WITH HYPOKINETIC DYSARTHRIA

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Radoslav Kárník

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Jiří Mekyska, Ph.D.

BRNO 2016



Bakalářská práce

bakalářský studijní obor **Audio inženýrství**
Ústav telekomunikací

Student: Radoslav Kárník

ID: 152699

Ročník: 3

Akademický rok: 2015/16

NÁZEV TÉMATU:

De-identifikace řečníků postižených hypokinetickou dysartrií

POKYNY PRO VYPRACOVÁNÍ:

U 90 % pacientů s Parkinsonovou nemocí se vyskytuje porucha motorické realizace řeči, která se označuje jako hypokinetická dysartrie. Tento druh dysartrie se projevuje především v oblastech fonace, artikulace, prozodie, plynulosti řeči a faciokineze. Z hlediska ochrany osobních informací řečníka není možné řečové korpusy volně šířit. Proto je v poslední době pozornost věnována metodám de-identifikace. V rámci této práce budou analyzovány různé druhy metody de-identifikace řečníků a bude navržen a implementován systém, který by pacienta de-identifikoval a zároveň zachoval v signálu informaci nutnou k diagnóze hypokinetické dysartrie.

DOPORUČENÁ LITERATURA:

[1] RUSZ, J.; CMEJLA, R.; RUZICKOVA, H.; RUZICKA, E.: Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. J Acoust Soc Am, roč. 129, č. 1, 2011: s. 350–367.

[2] POBAR, M.; IPSIC, I.: Online speaker de-identification using voice transformation. in Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on, s. 1264-1267.

Termín zadání: 1.2.2016

Termín odevzdání: 1.6.2016

Vedoucí práce: Ing. Jiří Mekyska, Ph.D.

Konzultant bakalářské práce:

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

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské 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.

ABSTRAKT

Tato práce se zabývá návrhem a realizací systému, který provádí de-identifikace řečových nahrávek pacientů postižených Parkinsonovou nemocí. V práci jsou popsány příčiny a projevy Parkinsonovy nemoci a vlivy hypokinetické dysartrie na řečový projev pacientů. Část práce je věnována řečovým příznakům popisujícím prozodii, podle pomcí kterých se dá hypokinetická dysartrie diagnostikovat z řeči. Dále se zabývá způsoby de-identifikace řeči a systémem na evaluaci výsledků pomocí rozeznávání řečníků a pacientů. De-identifikační systém využívá Normalizaci vokálního traktu (VTLN), evaluační systém využívá Gaussovy smíšené modely (GMM). Na testování byla využita databáze PARCZ, která obsahuje nahrávky řečových cvičení pacientů postižených Parkinsonovou nemocí a kontrolních řečníků

KLÍČOVÁ SLOVA

De-identifikace řeči, Gaussovy smíšené modely, hypokinetická dysartrie, normalizace vokálního traktu, ochrana osobních informací, Parkinsonova nemoc, zpracování řeči.

ABSTRACT

This paper discusses design and implementation of a system that performs de-identification of speech recordings of patients suffering from Parkinson's disease. The paper describes causes and symptoms of Parkinson's disease and effects of hypokinetic dysarthria on speech. Part of the paper is devoted to speech features that can be used for diagnosing hypokinetic dysarthria from speech. It also describes ways of speech de-identification and system for evaluating results using recognition of speakers and patients. De-identification system uses vocal tract length normalization (VTLN) and evaluating system uses Gaussian mixture models (GMM). PARCZ database was used for testing. It contains recordings of speech of patients affected by Parkinson's disease and control speakers.

KEYWORDS

Gaussian mixture models, hypokinetic dysarthria, Parkinson's disease, protection of personal information, speech de-identification, speech processing, vocal tract length normalization.

KÁRNÍK, Radoslav *De-identifikace řečníků postižených hypokinetickou dysartrií: bakalářská práce*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2016. 49 s. Vedoucí práce byl Ing. Jiří Mekyska, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „De-identifikace řečníků postižených hypokinetickou dysartrií“ jsem vypracoval samostatně pod vedením vedoucího bakalářské 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 bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících 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.

Brno

.....

podpis autora

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu Ing. Jiřímu Mekyskovi, Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Brno

.....

podpis autora



Faculty of Electrical Engineering
and Communication
Brno University of Technology
Purkynova 118, CZ-61200 Brno
Czech Republic
<http://www.six.feec.vutbr.cz>

PODĚKOVÁNÍ

Výzkum popsany v této bakalářské 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.

Brno

.....

podpis autora



EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI



OBSAH

Úvod	11
1 Parkinsonova nemoc a hypokinetická dysartrie	12
1.1 Parkinsonova nemoc	12
1.1.1 Příčiny nemoci	12
1.1.2 Projevy nemoci	12
1.2 Hypokinetická dysartrie	13
1.2.1 Plynulost řeči	13
1.2.2 Fonace	13
1.2.3 Artikulace	14
1.2.4 Prozodie	14
2 Vybrané řečové příznaky	15
2.1 Základní tón	15
2.2 Energie	16
2.3 Tempo	17
2.4 Pauzování	17
2.5 Formanty	17
2.6 Melovské frekvenční kepstrální koeficienty	18
2.7 Lineární predikční koeficienty	18
2.8 Doba navození řeči	19
3 Metody de-identifikace řeči	20
3.1 Normalizace vokálního traktu	20
3.1.1 Transformační funkce	20
3.1.2 Warpovací faktor	21
3.2 Změna prozodie	21
3.3 DROPSY	22
4 Evaluace metod de-identifikace řeči	23
4.1 Gaussův smíšený model	23
4.2 Fonetický SID systém	24
5 Návrh systému	25
6 Databáze nahrávek	26

7 Implementace de-identifikačního systému	28
7.1 Postup	28
7.1.1 Načtení a segmentace	28
7.1.2 Transformace segmentů	28
7.1.3 Spojení segmentů	29
8 Implementace evaluačního systému	30
8.1 Rozeznání řečníků	30
8.1.1 Parametrizace	30
8.1.2 Klasifikace	30
8.2 Rozeznání pacientů	31
8.2.1 Parametrizace	31
8.2.2 Klasifikace	31
9 Výsledky práce	32
9.1 Rozeznání řečníků	32
9.2 Identifikace Parkinsonovy nemoci	33
9.2.1 Klasifikace	33
9.2.2 Zpracování výsledků	33
9.3 Vyhodnocení výsledků	41
10 Závěr	42
Literatura	43
Seznam symbolů, veličin a zkratk	46
Seznam příloh	48
A Obsah přiloženého CD	49

SEZNAM OBRÁZKŮ

2.1	Formantové frekvence ve spektru signálu řečového projevu.	18
2.2	Příklad VOT v časovém průběhu signálu řečového projevu.	19
4.1	GMM systém.	24
5.1	Koncept de-identifikačního systému.	25
7.1	Příklad detekce špiček a segmentace v Matlabu.	29

SEZNAM TABULEK

6.1	Statistické údaje databáze PARCZ.	26
7.1	Použité hodnoty bortícího faktoru. α	29
9.1	Výsledky rozpoznání řečníků.	32
9.2	Příklad výsledků klasifikace cvičení.	33
9.3	Výsledky rozpoznání pacientů – referenční nezměněné soubory.	35
9.4	Výsledky rozpoznání pacientů – soubory po de-identifikaci podle asymetrické bortící funkce.	36
9.5	Výsledky rozpoznání pacientů – soubory po de-identifikaci podle symetrické bortící funkce.	37
9.6	Výsledky rozpoznání pacientů – soubory po de-identifikaci podle výkonové bortící funkce.	38
9.7	Výsledky rozpoznání pacientů – soubory po de-identifikaci podle kvadratické bortící funkce.	39
9.8	Výsledky rozpoznání pacientů – soubory po de-identifikaci podle lineární bortící funkce.	40
9.9	Výsledky testování výsledků klasifikace T-testem.	41

ÚVOD

Práce se zabývá de-identifikací řečového projevu pacientů postižených hypokinetickou dysartrií, což je porucha realizace řeči, která se projevuje u většiny pacientů postižených Parkinsonovou nemocí.

V dnešní době lze automatizovat velké množství lékařských úkonů – a to hlavně diagnózu různých nemocí. V tomto případě se jedná o diagnózu Parkinsonovy choroby z řeči pacientů. Aby se tato metoda mohla vylepšovat, je potřeba velké množství testovacích řečníků. Nahrávky řečníků jsou v databázích různých nemocnic, jejich šíření je však nemožné z důvodu ochrany osobních údajů. Výzkumné týmy tak nemohou zdokonalovat a testovat své systémy na větším počtu vzorků a jsou často odkázáni pouze na menší databáze.

Cílem práce je proto najít způsob, díky kterému by bylo šíření těchto databází možné. Toho se dá docílit pomocí de-identifikace řeči nahraných pacientů. Musíme však klást důraz na zachování důležitých řečových příznaků, které se používají k identifikaci nemoci.

První část práce se věnuje popisu hypokinetické dysartrie a Parkinsonovy nemoci včetně projevů této nemoci, ty jsou popsány v první kapitole. Další kapitola pojednává o řečových příznacích, které popisují prozodii a díky kterým můžeme sledovat dopad nemoci na řeč pacientů. Třetí kapitola popisuje způsobech de-identifikace řeči, hlavně o VTLN, tedy normalizaci vokálního traktu, což je způsob, který byl v práci použit. Čtvrtá kapitola se věnuje způsobům evaluace výsledků de-identifikace a zaměřuje se na Gaussovy smíšené modely. Další kapitola je věnována návrhu systému, který provede de-identifikaci a klasifikuje její úspěšnost. Tento systém využívá normalizaci vokálního traktu k de-identifikaci a Gaussovy smíšené modely k evaluaci výsledků. Následuje kapitola popisující použitou databázi řečových cvičení.

Druhá část práce je věnována popisu implementace systému v prostředí Matlab a vyhodnocení výsledků. Těch bylo dosaženo porovnáním přesností klasifikace řečníků a pacientů před de-identifikací a po de-identifikaci.

1 PARKINSONOVA NEMOC A HYPOKINETICKÁ DYSARTRIE

Tato kapitola se obecně věnuje příznakům a projevům Parkinsonovy nemoci a hypokinetické dysartrie, tedy vadě řeči, která se vyskytuje u pacientů s Parkinsonovou nemocí. Dále popisuje, jakým způsobem se hypokinetická dysartrie projevuje v mluveném projevu pacientů.

1.1 Parkinsonova nemoc

Parkinsonova nemoc je chronické, postupující onemocnění centrální nervové soustavy. Jedná se o onemocnění motorického systému, vyznačuje se tremorem a narušením svalové koordinace.

I když se nemoc vyskytuje od starověku, poprvé byla popsána britským lékařem Jamesem Parkinsonem roku 1817.

1.1.1 Příčiny nemoci

Choroba je důsledkem vady dopaminergního systému. V části mozku nazývané černá substance (substantia nigra) dochází k úbytku dopaminergních neuronů, které spojují černou substanci s částí šedé hmoty (corpus striatum). Tyto neurony zajišťují tvorbu dopaminu, neurotransmiteru, který slouží k přenosu signálu mezi nervovými buňkami. Dopaminergní systém se podílí na modulaci celkového chování organismu, zejména v souvislosti s motivací a emocemi a reguluje funkci motorických neuronů v bazálních gangliích a účastní se regulace systému hypotalamu a hypofýzy [4].

Nedostatek dopaminu ovlivňuje chování kritických nervových buněk způsobem, že pacient postupně není schopen ovládat a kontrolovat svůj pohyb.

1.1.2 Projevy nemoci

Mezi tři nejvýznamnější klinické příznaky nemoci patří akineze, rigidita a tremor. Někteří pacienti mohou mít problémy s pamětí, změnami nálady a depresemi. Běžné jsou také poruchy spánku [2].

Tremor je nejčastější a nejlépe rozpoznatelný symptom nemoci. Nejvíce se projevuje na prstech, rukách, hlavě, rtech, bradě a čelisti. Frekvence chvění se pohybuje okolo 4 až 6 Hz. Tremor je nejsilnější v klidové poloze daných svalů a mizí s pohybem a během spánku [9].

Akineze se projevuje zpomalením pohybu a vyskytuje se přibližně u poloviny pacientů. Nejčastěji postihuje nohy při chůzi. Většinou se projevuje jako náhlá a krátkodobá neschopnost pohybu [9].

Rigidita je charakterizována zvýšením odporu pohybu. Pacient také může přijít o výraz v obličeji, tedy přestat mrkat a projevovat emoce. Zvýšené pnutí svalů na zádech vede ke shrbenému postoji pacienta. Shrbený postoj je také způsoben sníženou funkcí posturálního reflexu [2] [9]. Rigidita také způsobuje hypokinetickou dysartrii, tedy poruchu řeči, kterou často trpí lidé postižení Parkinsonovou nemocí.

1.2 Hypokinetická dysartrie

Název hypokinetická je složen ze dvou slov s řeckým původem, a to hypo (snížený) a kinetic (pohyb). Dysartrie způsobuje poruchu artikulace. Jedná se tedy o poruchu řeči způsobenou sníženou pohyblivostí řečového ústrojí v důsledku špatné funkce bazálních ganglií.

Nejčastěji se spojuje s Parkinsonovou nemocí, může být ale příznakem jiné choroby nebo mechanického poškození mozku. Projevuje se hlavně ve změně barvy hlasu, artikulaci a prozodii. Řeč je monotónní a tichá [5].

1.2.1 Plynulost řeči

Pacienti mají problémy s plynulostí řeči, nejčastěji se objevuje palilalie a hezitace.

Palilalie je porucha slovního projevu charakterizována chorobným opakováním frázi a slov, často na konci věty. Někdy je pacient vyjadřuje s rostoucí prudkostí a klesající hlasitostí hlasu [3].

Při hezitaci pacient dělá nezáměrné pauzy ve slovním projevu.

1.2.2 Fonace

Jednou z nejvýrazněji ovlivněných oblastí hlasového projevu je fonace. Ztuhlost pacientova hlasového ústrojí vede například k hlasovému tremoru, lehkému zvýšení základního tónu a zmenšení rozsahu intenzity řeči. Řeč je také chraptivá, což je způsobeno neúplným zavřením hlasivek. Vzduch tak proniká dál a tvoří turbulentní zvuky.

Hlasový tremor se objevuje v pokročilých stádiích nemoci, může při něm kolísat intenzita a frekvence hlasového projevu.

Základní tón je u pacientů přibližně o 20 Hz vyšší než u zdravých jedinců stejného věku. Nejedná se ale o velmi častý jev a proto se nepovažuje za symptom, podle kterého by se jednoznačně diagnostikovala hypokinetická dysartrie [5]. Mnohem lépe

rozpoznatelný rozdíl je ve změně rozsahu frekvence řeči, ten je u pacientů mnohem menší. Opět se objevuje spíše u závažnějších případů nemoci.

Rozsah intenzity řeči (hypofonie) patří k lépe rozpoznatelným symptomům, pacienti mluví monotónně s malým dynamickým rozsahem a slabě. Tento jev je patrný při vyslovování delších hlásek a vět.

1.2.3 Artikulace

U pacientů se projevují problémy se špatnou a nepřesnou artikulací. Příčinou je opět ztuhlost a omezený rozsah pohybu svalů hlasového ústrojí, přesněji mluvidel. Hlavně jazyka, čelistí a rtů.

Jedním z příznaků je nepřesnost, tedy neschopnost úplně dokončit hlásku anebo udržet hlásku po dostatečnou dobu. Mezi slovy můžeme místo pomlky slyšet slabý šum. Ten je způsoben neúplným uzavřením mluvidel před začátkem dalšího slova. Tento šum celkově snižuje srozumitelnost a kontrast mezi slovy a pomlkami.

U pacientů se nenašla žádná změna rychlosti řeči oproti zdravým jedincům. Mají však problém s udržením tempa. Při porovnání první a poslední věty v delším textu bylo vyzorováno, že pacienti téměř vždycky zrychlují [20].

1.2.4 Prozodie

Prozodie je melodie řeči. Mezi nejdůležitější faktory výpovědi patří přízvuk a intonace.

Přízvuk je ovlivněn změnou frekvence základního tónu, hlasitostí a délkou hlásek ve slově. Díky těmto faktorům může řečník přidat či ubrat slovu na důležitosti. Intonace používá změnu výšky tónu a přízvuku ke komunikaci. Můžeme tak rozlišit například věty tázací, rozkazovací, přací a oznamovací.

Dodání prozodie k projevu vyžaduje koordinaci fonace, respirace (dýchání), rezonance a artikulace. Proto se pro pacienty postižené Parkinsonovou nemocí nejedná o lehký úkol. Koordinace těchto faktorů musí být přesná, jinak bude prozodie znít abnormálně. Hypokinetická dysartrie tak může prozodii ovlivnit mnoha způsoby.

To vede k monotónní řeči v oblasti frekvence a intenzity, nepravidelné změny výšky hlasu, náhlé změny hlasitosti a prodloužení intervalů mezi slabikami a slovy [7].

2 VYBRANÉ ŘEČOVÉ PŘÍZNAKY

Kapitola se zabývá příznaky, které se používají ve zpracování řečových signálů a mohou být ovlivněny Parkinsonovou nemocí.

Měření se provádějí na nahraném hlasovém projevu pacientů. Mezi taková měření patří frekvence F_0 popisující fundamentální frekvenci hlasového projevu. Dále pak *jitter* a *shimmer*, tyto dva parametry se používají pro měření mikro nestabilit vibrací hlasivek. Nepravidelné vibrace způsobují náhodné modulace signálu, ty ovlivňují amplitudu (*shimmer*) a rozložení harmonických složek ve spektru a jejich periodu (*jitter*) Dále sledujeme například tvorbu pauz, tempo řeči, HNR (Harmonics to Noise Ratio) anebo dobu navození řeči (Voice Onset Time).

K posudku artikulace řečníka se používají například diadochokinetické testy (DDK). Při těchto testech se měří, jak přesně je řečník schopný vyslovovat rychle se měnící zvuky s plovivými a samohláskami.

2.1 Základní tón

U základního tónu (F_0) měříme jeho výšku, rozsah a jeho kolísání. Využití výšky je však spekulativní kvůli individuálním rozdílům jedinců, jako je pohlaví, či věk.

U pacientů s vyšším hlasem se při měření v Hz projevuje větší změna než u pacientů s hlubším hlasem. Proto se kolísání vyjadřuje v půltónech anebo častěji v procentech (relativní odchylka) [18]. Kolísání základního tónu se označuje F_0SD . I když je řeč pacientů postižených HD při vyslovování delších větných úseků monotónní, při prodloužených vokálech se projevuje kolísání frekvence z důvodu neschopnosti udržet svaly hrtanu ve stabilní poloze (kap. 1.2).

Jedním z parametrů, který sledujeme je F_0VR (F_0 Variation Range), který vyjadřuje rozdíl mezi maximální (F_{0_max}) a minimální (F_{0_min}) frekvencí, je vyjádřen vzorcem [14]:

$$F_0VR = F_{0_max} - F_{0_min} \text{ [Hz]}. \quad (2.1)$$

Tento parametr vede k výpočtu dalších parametrů: relativního kolísání ($relF_0VR$) a relativní směrodatné odchylky ($relF_0SD$). K jejich výpočtu slouží tyto rovnice [14]:

$$relF_0SD = \frac{\text{std}(F_0)}{\text{mean}(F_0)} \cdot 100 \text{ [%]} \quad (2.2)$$

$$relF_0VR = \frac{F_0VR}{\text{mean}(F_0)} \cdot 100 \text{ [%]}. \quad (2.3)$$

Výpočet F_0 může být založen na autokorelační metodě. *Jitter* představuje proměnlivost F_0 mezi periodami signálu řečového projevu. Jeho hodnota je vyjádřena v pro-

centech a je dána rovnicí [6]:

$$\text{jitter}(\text{relative}) = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad [\%], \quad (2.4)$$

kde T_i je délka periody a N je počet period.

Dalšími používanými parametry jsou například RAP (Relative Average Perturbation) a PPQ (five-point Pitch Perturbation) [6].

2.2 Energie

U pacientů pozorujeme hypofonii, tedy sníženou intenzitu hlasu. Také se projevují nedostatky v rozsahu a kolísání intenzity.

Kolísání je možné vyjádřit pomocí *shimmeru*, který představuje proměnlivost amplitud v každé periodě vibrace hlasivek. Vypočítá se podobně, jako *jitter* [6]:

$$\text{shimmer}(\text{relative}) = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad [\%], \quad (2.5)$$

kde A_i je peak-to-peak amplituda a N je počet period.

Využívá se také TKEO (Teager-Kaiser Energy Operator) (2.6), jedná se o nelineární operátor sledující energii. Měří okamžité změny energie signálů, které jsou tvořeny jednou, v čase proměnnou frekvencí.

$$\psi_c[x(t)] = (x'(t))^2 - x(t)x''(t), \quad (2.6)$$

kde $x(t)$ je spojitý signál a $x'(t) = \frac{dx(t)}{dt}$.

Také je možné sledovat ΔE a ΔA , tedy rozdíl minimální a maximální hodnoty energie a úrovně [14]. U tohoto měření musí být mikrofon kalibrován a v konstantní vzdálenosti od testovaného jedince.

Další parametry jsou například MPSD (Median of Power Spectral Density), LSTER (Low Short-Time Energy Ratio) [15] a HNR. HNR vyjadřuje poměr mezi energií harmonických složek řeči a šumu, používá se k určení kvality hlasu. Můžeme ho vypočítat ze vztahu:

$$\text{HNR} = 10 \log \left(\frac{E_p}{E_n} \right) \quad [dB], \quad (2.7)$$

kde E_p je energie harmonických složek a E_n energie šumu.

2.3 Tempo

S rostoucí vážností dysartrie by se tempo mělo snižovat, i když se na tom mnohé studie neshodují. Tempo artikulace je dáno průměrným počtem vyslovených slabik za sekundu.

Čas čistě řečového signálu se označuje NST (Net Speech Time). Do času se nepočítají pauzy, NST se tak vypočítá jako rozdíl TST (Total Speech Time) a TPT (Total Pause Time). Vydělením těchto parametrů počtem slabik v daném čase můžeme vyjádřit absolutní tempo TSR (Total Speech Rate) a tempo čistého řečového signálu NSR (Net Speech Rate). Oba parametry jsou vyjádřeny ve slabikách za sekundu [14].

Pro vypočítání těchto parametrů musíme přesně vyznačit části signálu, ve kterých se nachází řeč a rozlišit je od pauz. Proto se používá VAD (Voice Activity Detection), česky detektor řečové aktivity. VAD obsahuje mechanismus, který rozhoduje, zda se v daném místě signálu jedná o řeč, či nikoliv. K tomu se používá například metoda PARADE (Periodic to Aperiodic component Ratio based Activity DEtection) – detekce řeči pomocí poměru harmonických složek a šumu, tedy periodických a neperiodických složek [8].

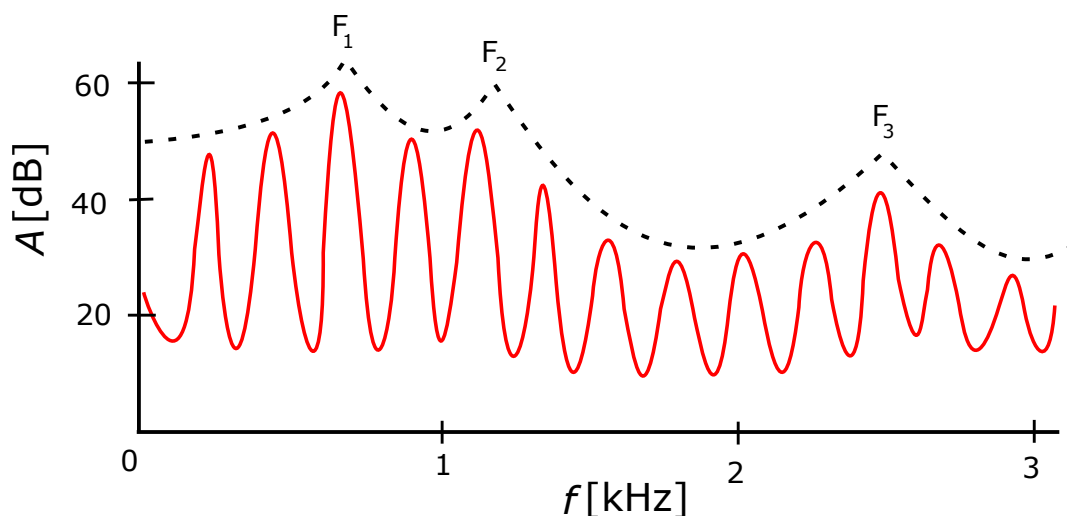
2.4 Pauzování

U pacientů dochází k přerušování mluvy a prodlužováním pauz (kap. 1.2.1). Proto jsou pauzy v mluveném projevu dalším sledovaným rysem. Charakteristika zahrnuje počet pauz, střední dobu pauz a procento gramaticky vhodných pauz (například mezi větami).

Hlavním parametrem je PPR (Procentual Pause Ratio), který se vypočítá pomocí příznaků z minulé podkapitoly podílem TPT a TST. PPR je vyjádřeno v procentech. Dalšími parametry jsou ISD (Interpause Speech Duration), příznak vyjadřující dobu mezi pauzami v sekundách, a SPIR (SPeech Index of Rhythmicity), který odpovídá počtu mezislovních pauz za minutu [14].

2.5 Formanty

Jazyk je jedním z hlavních orgánů podílejícím se na tvorbě řeči. Pohyb jazyka mění objem a tvar ústní dutiny a tím tvoří formantové oblasti ve spektru řečového signálu. Jeho ztuhlost způsobená Parkinsonovou nemocí proto může mít vliv na frekvenci těchto formantů.



Obr. 2.1: Formantové frekvence ve spektru signálu řečového projevu.

První formantová frekvence F_1 je závislá na výšce jazyka, druhá frekvence F_2 závisí na jeho strmosti při vyslovování [16]. Formantové kmitočty jsou závislé na věku, pohlaví a tvaru dutin. Kvůli tomu není možné porovnávat nenormalizované formanty mezi více jedinci. Proto se používají poměry mezi stejnými formanty při jiných samohláskách, například F_{2i}/F_{2u} [14].

2.6 Melovské frekvenční keprální koeficienty

MFCC jsou parametry, při jejichž výpočtu je prováděn výpočet reálného kepru řečového signálu, který je ve spektru podroben nelineární transformaci kmitočtové osy. Byly to první parametry, které braly v úvahu nelineární a maskovací vlastnosti lidského slyšení [21].

Metoda využívá zpracování signálu v časové i ve frekvenční oblasti. Popisuje přitom krátkodobé komplexní spektrum signálu. Výsledkem této metody jsou vektory čísel pro každý segment řečového signálu, které nesou informaci o dynamickém průběhu řeči. Důležitá část procesu MFCC, je pásmová filtrace frekvenčního spektra signálu. Tato filtrace je reprezentovaná bankou filtrů, nejčastěji trojúhelníkového tvaru, lineárně rozmístěných na tzv. melovské frekvenční škále [10].

2.7 Lineární predikční koeficienty

Lineární predikční koeficienty jsou vypočítány pomocí lineárního prediktivního kódování, což je jedna z metod analýzy řečového signálu, která odhaduje aktuální

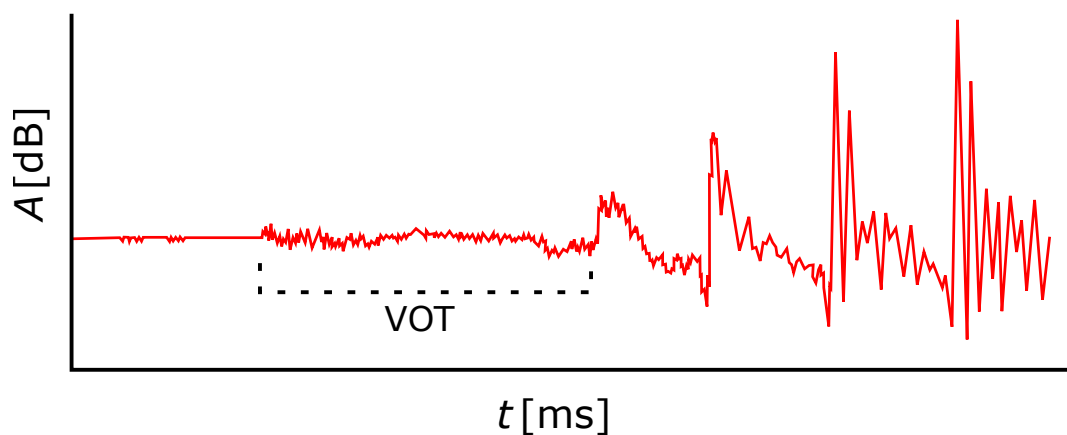
vzorek signálů z předešlých vzorků. Vychází tedy z předpokladu, že je možné popsat aktuální vzorek lineární kombinací předchozích vzorků. Pro lineární predikční analýzu vzorku $\tilde{s}[n]$ platí vztah:

$$\tilde{s}[n] = - \sum_{i=1}^p a_p[i] s[n-i] , \quad (2.8)$$

kde koeficienty $a_p[i]$ se označují jako lineární predikční koeficienty, proměnná p jako řád lineární predikce [13].

2.8 Doba navození řeči

Jelikož Hypokinetická Dysartrie ovlivňuje pohyblivost, může vést k nedokonalé synchronizaci svalů používaných k řečovému projevu. K posouzení této nepřesnosti se používá parametr VOT (Voice Onset Time). Ten je definován jako doba mezi uvolněním plovivní souhlásky a začátkem vokálu. Používá se také VOT ratio, které vypočítáme podílem VOT a délky slabiky [16].



Obr. 2.2: Příklad VOT v časovém průběhu signálu řečového projevu.

3 METODY DE-IDENTIFIKACE ŘEČI

De-identifikace slouží ke změně řeči jedince za účelem skrytí jeho identity, většinou kvůli ochraně osobních údajů.

V našem případě se jedná o databázi pacientů postižených Parkinsonovou nemocí. Aby bylo možné databázi šířit, je nutné skrýt identitu těchto jedinců. Každá metoda však jistým způsobem ovlivňuje určité prozodické příznaky důležité k identifikaci Parkinsonovy nemoci. Proto je nutné najít způsob, jak u těchto vzorků provést de-identifikaci a nepřijít o cenná data.

3.1 Normalizace vokálního traktu

VTLN (Vocal Tract Length Normalisation) je metoda, která se běžně používá pro normalizaci hlasového traktu řečníků. A to například k rozeznání hlasu anebo de-identifikaci.

Zakládá se na normalizaci formantových kmitočtů, které jsou u každého řečníka specifické díky rozdílné fyziologické stavbě (například ženy je mají přibližně o 20 % výš než muži). Tyto formantové kmitočty vznikají při průchodu rezonančním prostorem tvořeným hrdlem a ústní dutinou, popřípadě dutinou nosní. Jinými slovy tedy můžeme říct, že VTLN kompenzuje projevy způsobené různou délkou vokálního traktu tak, že přesune frekvenční osu řečníka, aby se jeho pozice formantů blížily pozicím průměrného řečníka. Komponenty zahrnuté v této technice jsou následující [19]:

- transformační (warpovací) funkce
- bortící (warpovací) faktor α

3.1.1 Transformační funkce

Transformace frekvenční osy spočívá v jejím nelineárním natažení (popřípadě smrštění), odborně nazývaném borcení (warping). Transformační funkce je monotónní funkce, která vrací hodnoty mezi 0 a ω_m . Její tvar závisí na bortícím faktoru α .

Podle [22] může mít warpovací funkce $F_\alpha(\omega) = \tilde{\omega}$ různé tvary, například:

Symetrická:

$$g(\omega, \alpha) = \begin{cases} \alpha\omega, & \omega \leq \omega_0 \\ \alpha\omega_0 + \frac{\pi - \alpha\omega_0}{\pi - \omega_0}(\omega - \omega_0), & \omega > \omega_0 \end{cases}, \quad \omega_0 = \begin{cases} \frac{7\pi}{8}, & \alpha \leq 1 \\ \frac{7\pi}{8\alpha}, & \alpha > 1 \end{cases}. \quad (3.1)$$

Asymetrická:

$$g(\omega, \alpha) = \begin{cases} \alpha\omega, & \omega \leq \omega_0 \\ \alpha\omega_0 + \frac{\pi - \alpha\omega_0}{\pi - \omega_0}(\omega - \omega_0), & \omega > \omega_0 \end{cases}, \quad \omega_0 = \frac{7\pi}{8}. \quad (3.2)$$

Kvadratická:

$$g(\omega, \alpha) = \omega + \alpha \left(\left(\frac{\omega}{\pi} \right) - \left(\frac{\omega}{\pi} \right)^2 \right). \quad (3.3)$$

Výkonová:

$$g(\omega, \alpha) = \pi \left(\frac{\omega}{\pi} \right)^\alpha. \quad (3.4)$$

Bilineární:

$$g(\omega, \alpha) = \left| -i \frac{z - \alpha}{1 - \alpha z} \right|, \quad z = e^{i\omega}. \quad (3.5)$$

Pro tyto funkce platí, že původní frekvenční osa je transformována na stejný interval $\tilde{\omega} \in \langle 0, \omega_m \rangle$ a má dva fixované body $g(0, \alpha) = 0$ a $g(\omega_m, \alpha) = \omega_m$.

3.1.2 Warpovací faktor

Jedna z metod, pomocí které můžeme zjistit warpovací faktor α pro normalizaci vokálního traktu, je maximum-likelihood (ML)-based optimization criterion, tedy kritérium založené na největší pravděpodobnosti. Optimální warpovací faktor pro daného řečníka lze nalézt pomocí vzorce:

$$\hat{\alpha}_s = \arg \max P(\mathbf{O}_\alpha^s | \lambda, \mathbf{W}^s), \quad (3.6)$$

kde $\hat{\alpha}_s$ je optimální warpovací faktor s -tého řečníka, \mathbf{O}^s je optimalizační kritérium, \mathbf{W}^s je kritériu odpovídající soubor přepisů a λ je model daného přepisu [23]. Faktor také můžeme vypočítat pomocí parametrické metody. Zde se α odhaduje pomocí předpokládané polohy formantů.

Warpovací faktor je obvykle vybrán z tabulky dostupných hodnot. Tato technika se nazývá grid search.

3.2 Změna prozodie

De-identifikace může být dosaženo také změnou prozodických příznaků. Měnit se dá hlavně kmitočet základního tónu, intenzita a tempo řeči.

Změna frekvence základního tónu se skládá ze dvou kroků – roztažení (případně smrštění) signálů a jeho převzorkování. Nejdříve se změní hustota a délka signálu podle faktoru roztažení α . Touto změnou se změní počet vzorků původního signálu

M na počet $M' = \alpha \cdot M$. Roztažením docílíme hlubší frekvence, smrštěním se kmitočet zvýší. Změněný signál se poté musí převzorkovat, aby byla zachována jeho původní délka [17].

3.3 DROPSY

DROPSY (Diphone RecOgnition and sPeech SYnthesis) je systém, který pomocí difonového rozpoznání řeči (Diphone speech recognition) ze vstupního řečového signálu obstará fonetický přepis, který poté pomocí syntézy převede na řeč jiného řečníka. Tento systém je jazykově závislý, textově nezávislý a může být použit v reálném čase. Používá se tak hlavně na internetu, kdy je potřeba zamezit úniku osobních informací – například ve zdravotnických a právních internetových poradnách a při používání osobních virtuálních asistentů (Siri od společnosti Apple a Cortana od Microsoftu) [12].

De-identifikace, docílená tímto systémem je díky doslovnému přepisu a následnému vytvoření zcela nového umělého řečového signálu na nejvyšší možné úrovni. Nedochovejí se totiž absolutně žádné řečové příznaky původního řečníka. Tím je ale DROPSY metoda pro tuto práci nepoužitelná.

4 EVALUACE METOD DE-IDENTIFIKACE ŘEČI

Po provedení de-identifikace je potřeba její účinnost ověřit identifikačním systémem.

Identifikace řečníků (SID – Speaker IDentification) je proces zpracování řečového signálu a následného automatického rozpoznání osoby, které řeč patří. V našem případě systém rozhodne, zda se transformovaný řečový signál podobá původnímu signálu před procesem de-identifikace. Systémy k identifikaci řečníků můžeme rozdělit na ty, které jsou závislé na textu a na ty, které na něm závislé nejsou.

- **Závislé na textu** – tyto systémy se používají v případech, kdy jsou fráze k rozeznání pevně dané a předem známé.
- **Nezávislé na textu** – při používání těchto systému nezáleží na tom, co řečník říká. Proto fráze, které jsou použity k natrénování systému, mohou být jiné než fráze testované.

4.1 Gaussův smíšený model

GMM (Gaussian Mixture Model) (obr. 4.1) je klasifikátor vybraný v této práci. Potřebuje ale poměrně velké množství dat. Je nezávislý na textu. Během učící fáze se pro každého řečníka vytvoří D -rozměrná funkce parametrů X , využívající pravděpodobnostní funkci:

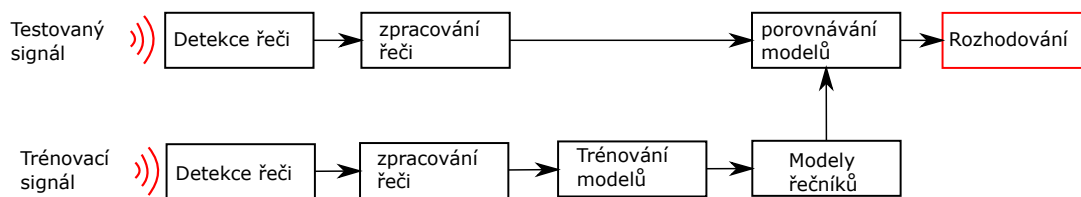
$$L\left(\frac{\lambda}{X}\right) = P\left(\frac{X}{\lambda}\right) = \prod_{i=1}^D P\left(\frac{x_i}{\lambda}\right). \quad (4.1)$$

K výpočtu parametrů $\lambda = \{\alpha_i, \mu_i, \Sigma_i\}$ se používá odhad největší pravděpodobnosti. Tím se vytvoří model pro každého řečníka. μ_i je vektor středních hodnot, Σ_i kovarianční matice a α_i vážící koeficient.

Ve fázi identifikace se vyvolá D -rozměrná funkce parametrů X a využije se funkce (4.2) k vyhodnocení podobnosti s modelem každého řečníka [24].

$$P\left(\frac{X}{\lambda}\right) = \sum_{i=1}^M a_i b_i(X) \quad (4.2)$$
$$b_i(X) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2} (X - \mu_i)' \Sigma_i^{-1} (X - \mu_i)\right]$$

Prvním krokem po detekci řečového signálu je zpracování řeči, kdy se ze signálu extrahují vhodné příznaky. Vektor příznaků neznámého řečníka se dosadí do rovnice 4.2 a hledá se model s největší pravděpodobností. Při rozhodování potom systém rozhodne, zda se jedná o stejného řečníka, případně s jakou jistotou.



Obr. 4.1: GMM systém.

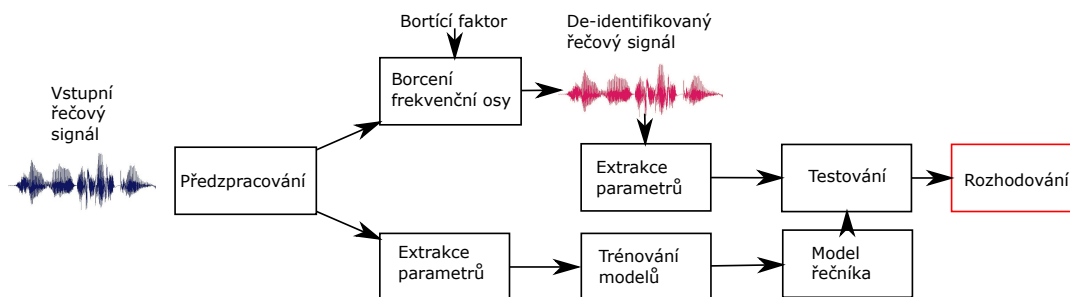
4.2 Fonetický SID systém

Základní myšlenka, se kterou fonetická identifikace řečníků pracuje, je využití statistického modelu řečnickovy výslovnosti, který je natrénován z fonetických sekvencí řečnickovy výpovědi.

Důvod používání tohoto systému je ten, že fonetické sekvence zachycují jedinečnou výslovnost mluvčího – a to včetně složitějších příznaků, které jednodušší systémy nepoužívají. Mezi tyto příznaky patří například výstřednosti ve výslovnosti, využití méně běžných slov a styl mluvení řečníka [11].

5 NÁVRH SYSTÉMU

Systém je tvořen ze dvou částí: první část provede de-identifikaci původního řečového signálu pomocí VTLN a druhá část následně porovná úspěšnosti určení řečníků a pacientů před de-identifikací a po de-identifikací. Podle toho bude možné určit úspěšnost de-identifikace.



Obr. 5.1: Koncept de-identifikačního systému.

Systém se skládá z těchto částí:

- Předzpracování – při předzpracování se normalizuje úroveň hlasitosti nahrávek a sjednotí vzorkovací kmitočet. Také se odstraní případná střední hodnota, a pokud je nahrávka vícekanálová, převede se do jednoho kanálu.
- Extrakce parametrů – signál se rozdělí na rámce (ty se násobí Hammingovým oknem) z každého rámce je poté možné vypočítat parametry, které se zapisují do matice či vektoru. Z matic a vektorů potom mohou být převedeny na skalární parametry.
- Borcení frekvenční osy – při VTLN se frekvenční osa bortí podle faktoru α , jehož hodnota bude subjektivně vybrána testováním tak, aby nebylo možné rozeznat řečníky.
- Po VTLN se změněný řečový signál porovná s modelem, který se natrénoval z původního nezměněného řečového signálu.
- V posledním kroku systém vždy rozhodne, o jakého řečníka se jedná a posoudí, zda se jedná o pacienta, nebo kontrolního řečníka. Díky tomu bude dále možné porovnat tyto výsledky s výsledky testování nezměněných zvukových nahrávek.

6 DATABÁZE NAHRÁVEK

Budeme pracovat s databází PARCZ (Czech Parkinsonian Speech Database), nahanou ve Fakultní nemocnici u sv. Anny v Brně. Databáze obsahuje 154 řečníků, z tohoto počtu je 101 pacientů postižených Parkinsonovou nemocí. Zbytek jsou kontrolní řečníci, kteří netrpí žádným neurologickým onemocněním ani poruchou řeči. Mezi pacienty je 60 mužů a 41 žen, mezi kontrolními řečníky 26 mužů a 27 žen.

Tab. 6.1: Statistické údaje databáze PARCZ.

Řečníci	Počet		Průměrný věk		Věkový rozsah		STD věku	
	Muži	Ženy	Muži	Ženy	Muži	Ženy	Muži	Ženy
Zdraví	26	26	65,58	61,81	49 - 83	45 - 87	8,9	9,05
Nemocní	60	40	66,28	68,98	46 - 87	49 - 86	8,63	7,74

Protokol, z jehož nahrávání se tvoří databáze, obsahuje množství úloh, které testují všechny aspekty řeči týkající se hypokinetické dysartrie. My se zaměříme na cvičení s větami. Z nich je možné zjistit intonaci, tempo řeči, přízvuk a různé prozodické příznaky. Vybraná cvičení jsou:

8.1 – Udržení rytmu v rytmickém celku – řečník přečte básně s přednesem.

*Chcete vidět velký lov?
Budu lovit v džungli slov.
Osedlám si Picassa,
Chytím básně do lasa.*

8.2 – Základní větné intonační vzorce – řečník přečte větu v tázacím, rozkazovacím a oznamovacím tvaru.

Prostřete k obědu.

8.3 – Přemísťování kontrastního přízvuku – řečník přečte větu postupně s důrazem na slovech okapu, špinavá a voda.

Z okapu teče špinavá rezavá voda.

8.4 – Intonační variabilita – řečník přečte věty různými způsoby.

*Teď musíš být chvíli trpělivý, než to dokončíme.
Tak dáš mi už konečně pokoj!
Už mě to nebaví, dej mi už konečně pokoj!
Tak co, jak to dopadlo?*

9.2 a 9.2_I – Srozumitelnost předřikávaných vět – řečník opakuje věty složité na výslovnost.

*Do čtvrt hodiny tam byla smršť.
Prohovořte to s ním dopodrobna.
Při ústupu pluku duní bubny.
Kuchařští učni nejsou jak zlatníci.
Celý večer se učí sčítat.*

Předřikávané věty ze cvičení 9.2 se později opakují ve cvičení 9.2_I z důvodu sledování vlivu únavy na řeč.

Záznam probíhal v místnosti s hlukem prostředí menším než 30 dB a kvůli potlačení dozvuků byly do místnosti nainstalovány akustické paravany. K nahrávání byl použit kvalitní kondenzátorový mikrofon a zvuková karta umožňující záznam s vzorkovacím kmitočtem 48 kHz.

7 IMPLEMENTACE DE-IDENTIFIKAČNÍHO SYSTÉMU

Tato kapitola popisuje způsob implementace de-identifikačního systému v programu Matlab. Systém načte zvukový soubor a provede de-identifikaci pomocí VTLN za použití jedné z pěti funkcí a zvoleného bortícího faktoru α .

Tímto způsobem byla transformována celá databáze podle všech pěti bortících funkcí. Vzniklo tak dalších pět databází s transformovanými nahrávkami řečníků.

7.1 Postup

Část programu provádějící de-identifikaci se skládá z načtení, segmentace, transformace segmentů a opětovného složení segmentů na zvukový soubor [1].

7.1.1 Načtení a segmentace

Program načte zvolený zvukový soubor a normalizuje ho, zároveň je zaznamenán i vzorkovací kmitočet načteného zvukového souboru.

I když se většinou při analýze řečového signálu používá rozdělení do stejně dlouhých segmentů, pro potřeby VTLN je pro co nejlepší kvalitu výsledku nutné použít dělení podle periodicity řeči. K tomu je použita funkce, která detekuje špičky v řeči a označuje jejich polohu (tzv. *pitch marking*).

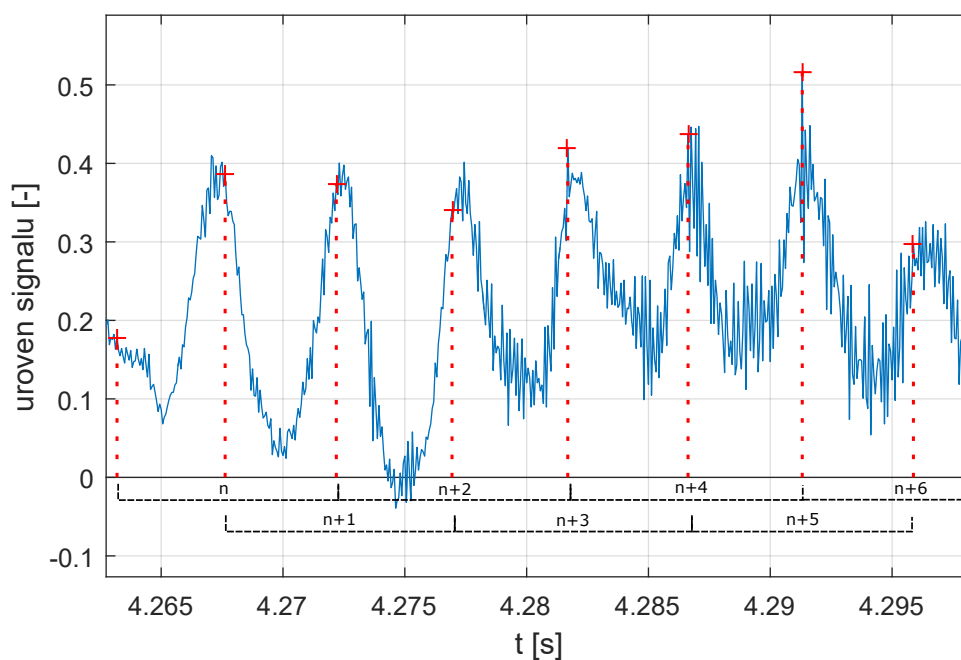
Následně je signál podle špiček rozdělen na segmenty (obr. 7.1). Sousední segmenty se vždy polovinou překrývají.

7.1.2 Transformace segmentů

Aby mohlo dojít k transformaci, musí být všechny segmenty převedeny pomocí FFT do frekvenční oblasti.

Následně se spektrum každého segmentu bortí podle vybrané bortící funkce (podkapitola 3.1.1) a faktoru α . Tento faktor byl pro každou funkci vybrán podle testování pěti posluchačů, přičemž měl každý za úkol rozeznat, od kterého z pěti vybraných netransformovaných řečníků pochází vybraná transformovaná věta. Hodnota faktoru α byla postupně měněna až do té doby, kdy posluchači nebyli schopní správně určit řečníka. Použité hodnoty α jsou zobrazeny v tabulce 7.1.

Hodnoty spektra jsou poté interpolovány. Nakonec se segmenty opět převedou do časové oblasti pomocí IFFT.



Obr. 7.1: Příklad detekce špiček a segmentace v Matlabu.

Tab. 7.1: Použité hodnoty bortícího faktoru. α

bortící funkce	asymetrická	symetrická	výkonová	kvadratická	bilineární
hodnota faktoru α	1,15	1,15	1,1	1,1	0,15

Při testování byla hodnota α měněna s krokem 0,05.

7.1.3 Spojení segmentů

V posledním kroku se segmenty spojí pomocí PSOLA (Pitch Synchronous Overlap and Add). Jedná se o metodu, která se používá v syntéze řeči. Výsledná řeč se skládá z překrývajících segmentů. Každý ze segmentů se vynásobí Hammingovým oknem a s určitým překryvem se přičte k předchozímu segmentu.

Pokud se určí jiná hodnota překryvu, než hodnota, podle které se řeč dělila, dojde ke změně prozodie. V této práci je však PSOLA použita pouze ke spojení segmentů, nejde nám tedy o její využití pro změnu řeči. Tímto spojením vznikne zvukový soubor, který je uložen se stejnou vzorkovací frekvencí.

8 IMPLEMENTACE EVALUAČNÍHO SYSTÉMU

Další část se věnuje implementaci evaluačního systému. Ten se skládá z programu na rozeznání řečníků a programu na rozeznání pacientů postižených Parkinsonovou nemocí.

8.1 Rozeznání řečníků

Po de-identifikaci řečníků je potřeba zjistit, jakým způsobem se změní úspěšnost rozeznávání řečníků. K tomu byl použit program pracující na základě GMM (podkapitola 4.1). Tato kapitola popisuje způsob implementace tohoto klasifikačního programu v prostředí Matlab.

8.1.1 Parametrizace

Každá nahrávka byla rozdělena na rámce, které byly vynásobeny Hammingovým oknem. Z každého okna byly vypočítány tyto parametry:

- F_0 (frekvence základního tónu)
- Krátkodobá energie
- Poloha prvních tří formantů a šířka formantových pásem
- LPC (lineární predikční koeficienty)
- MFCC (melovské frekvenční keprální koeficienty)

Vznikne tak matice, ve které se počet sloupců rovná počtu segmentů a počet řádků počtu parametrů. Z každého řádku je vypočítáno minimum, maximum, medián, aritmetický průměr, směrodatná odchylka a rozsah. Tyto operace jsou využity i u TKEO (Teagerův-Kaiserův energetický operátor), který byl vypočítán z celé nahrávky. Tyto hodnoty jsou dále postupně dosazeny do vektoru příznaků.

8.1.2 Klasifikace

Při kvalifikaci je vždy vybrán jeden transformovaný soubor (teď již vektor příznaků) od každého řečníka a z ostatních netransformovaných vět je pro každého řečníka vytvořen model.

Program poté porovná vybrané věty s modely a vybere řečníka s největší pravděpodobností.

8.2 Rozeznání pacientů

Abychom mohli posoudit, zda má de-identifikace dopad na klasifikaci nemoci, byl navrhnut systém na klasifikaci hypokinetické dysartrie. Systém opět využívá GMM s odlišným způsobem klasifikace, vytváření modelů a testování.

8.2.1 Parametrizace

Jelikož VTLN může ovlivnit určité parametry víc než ty ostatní, parametrizace je provedena čtyřmi způsoby. Každý z nich pracuje pouze s vybranými parametry:

- **1. způsob:** všechny výše popsání parametry
- **2. způsob:** F_0 , krátkodobá energie, TKEO
- **3. způsob:** MFCC
- **4. způsob:** LPC, formanty

Pro každý soubor byly tedy vytvořeny čtyři vektory příznaků, každý z nich jiným způsobem parametrizace.

8.2.2 Klasifikace

Tentokrát je databáze rozdělena podle cvičení. Ze cvičení je vybrán jeden testovací soubor a z ostatních souborů jsou natrénovány dva modely – model pacientů a model kontrolních řečníků. Vybraný soubor je porovnán s těmito dvěma modely a program podle větší pravděpodobnosti rozhodne, zda se jedná o pacienta, nebo kontrolního řečníka.

9 VÝSLEDKY PRÁCE

Tato kapitola popisuje výsledky, kterých bylo dosaženo použitím systému popsaného v předešlé kapitole. Kapitola je tak zaměřená na porovnání výsledků, kterých bylo dosaženo před de-identifikací a po de-identifikaci nahrávek řečových cvičení.

9.1 Rozeznání řečníků

Testování řečníku bylo pro každou databázi opakováno desetkrát. Z každého měření byla vypočítána přesnost klasifikace (ACC):

$$ACC = \frac{\text{počet správně určených}}{\text{počet správně určených} + \text{počet špatně určených}} \cdot 100 \text{ [%]}. \quad (9.1)$$

Z těchto deseti měření byla dále vypočítána průměrná úspěšnost. Výsledky testování jsou zobrazeny v tabulce 9.1.

Tab. 9.1: Výsledky rozpoznání řečníků.

test	Transformační funkce					
	bez transformace	asymetrická	symetrická	výkonová	kvadratická	bilineární
	ACC [%]	ACC [%]	ACC [%]	ACC [%]	ACC [%]	ACC [%]
1.	77,92	12,99	16,88	14,94	5,84	7,79
2.	77,92	13,64	18,83	10,39	6,49	8,44
3.	79,22	10,39	17,53	14,29	3,90	3,90
4.	76,62	14,94	20,13	14,29	3,25	5,84
5.	74,03	12,34	15,58	16,88	4,55	5,19
6.	77,27	13,64	22,08	12,34	7,14	7,14
7.	77,27	12,99	16,23	14,29	3,90	5,19
8.	77,27	9,74	12,99	12,34	6,49	5,84
9.	70,78	7,14	20,78	14,94	1,95	3,90
10.	73,38	9,09	17,53	14,94	3,90	6,49
průměr	76,17	11,69	17,86	13,96	4,74	5,97

ACC - úspěšnost klasifikace.

Z tabulky je patrné, že průměrná úspěšnost rozpoznání řečníku před de-identifikací byla 76,17 %, po de-identifikaci se zmenšila na 4,74–17,86 %. Program měl největší průměrnou úspěšnost při použití asymetrické transformací funkce a nejmenší úspěšnost při použití kvadratické transformací funkce. Dobrý je také výsledek klasifikace souborů de-identifikovaných pomocí bilineární transformací funkce.

9.2 Identifikace Parkinsonovy nemoci

9.2.1 Klasifikace

Klasifikace pacientů probíhala po jednotlivých řečových cvičeních – a to pro všechny řečníky a všechna cvičení pro čtyři způsoby parametrizace. Takto bylo testováno pět dříve vytvořených databází s transformovanými zvukovými soubory i referenční netransformovaná databáze.

Pro každé cvičení nám tak vznikla tabulka (tab 9.2), přičemž hodnota 1 znamená správnou klasifikaci, 0 špatnou klasifikaci a 2 absenci nahrávky daného řečníka v daném cvičení.

Tab. 9.2: Příklad výsledků klasifikace cvičení.

řečník	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	...154
výsledek klasifikace	0	1	1	1	1	1	1	1	1	1	0	2	1	1	1	2	1	1	...

9.2.2 Zpracování výsledků

Dále byla pro každé cvičení vypočítána přesnost klasifikace (ACC), senzitivita (SN), specificita (SP) a Matthewsův korelační koeficient (MCC).

Přesnost vyjadřuje, v kolika procentech případů klasifikační program zhodnotil stav řečníka správně. Senzitivita testu vyjadřuje úspěšnost, s níž test zachytil přítomnost nemoci, specificita naopak vyjadřuje schopnost testu rozpoznat případy, u nichž se nemoc nevyskytuje. MCC se používá k měření kvality binární klasifikace a nabývá hodnot od -1 do 1. Pokud se MCC rovná 1, jedná se o bezchybnou klasifikaci.

Pokud se rovná 0, jedná se o přesnost klasifikace, která se rovná náhodným výsledkům. Jestliže se rovná -1, jedná se o naprostý rozpor mezi klasifikací a predikcí. Všechny tyto testy používají následující čtyři stavy:

- **TP:** správně určený výskyt nemoci
- **TN:** správně určený kontrolní řečník
- **FP:** omylně určený výskyt nemoci
- **FN:** omylně určený kontrolní řečník.

Přesnost:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100 \text{ [%]}. \quad (9.2)$$

Senzitivita:

$$SN = \frac{TP}{TP + FN} \cdot 100 \text{ [%]}. \quad (9.3)$$

Specificita:

$$SP = \frac{TN}{TN + FP} \cdot 100 \text{ [%]}. \quad (9.4)$$

Matthewsův korelační koeficient:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \text{ [-]}. \quad (9.5)$$

Tyto výsledky se nachází v tabulkách 9.3 až 9.8.

Tab. 9.3: Výsledky rozpoznání pacientů – referenční nezměněné soubory.

cv.	Způsob parametrizace															
	1.				2.				3.				4.			
	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]
1.	61,84	62,63	60,38	0,220	65,13	68,69	58,49	0,264	58,55	60,61	54,72	0,147	46,71	42,42	54,72	-0,027
2.	66,89	73,74	53,85	0,273	69,54	72,73	63,46	0,351	58,94	63,64	50,00	0,132	43,71	49,49	32,69	-0,170
3.	64,47	63,64	66,04	0,283	65,13	66,67	62,26	0,278	61,84	63,64	58,49	0,212	45,39	29,29	75,47	0,051
4.	60,26	62,63	55,77	0,176	63,58	66,67	57,69	0,235	60,93	62,63	57,69	0,194	49,67	51,52	46,15	-0,022
5.	64,63	67,01	60,00	0,259	63,27	67,01	56,00	0,222	63,95	61,86	68,00	0,283	52,38	39,18	78,00	0,173
6.	63,33	61,00	68,00	0,274	67,33	68,00	66,00	0,323	55,33	55,00	56,00	0,104	52,00	38,00	80,00	0,182
7.	67,55	67,00	68,63	0,339	66,89	68,00	64,71	0,312	64,90	64,00	66,67	0,291	53,64	41,00	78,43	0,193
8.	64,71	68,00	58,49	0,256	66,67	71,00	58,49	0,287	55,56	53,00	60,38	0,127	50,33	57,00	37,74	-0,051
9.	59,48	65,00	49,06	0,137	63,40	68,00	54,72	0,221	59,48	62,00	54,72	0,160	50,33	51,00	49,06	0,001
10.	63,40	64,00	62,26	0,251	66,01	67,00	64,15	0,299	60,13	57,00	66,04	0,219	51,63	47,00	60,38	0,071
11.	64,43	68,04	57,69	0,249	65,10	71,13	53,85	0,246	58,39	50,52	73,08	0,228	61,07	72,16	40,38	0,128
12.	62,91	64,29	60,38	0,237	65,56	70,41	56,60	0,264	55,63	56,12	54,72	0,104	51,66	55,10	45,28	0,004
13.	65,56	68,37	60,38	0,278	68,87	72,45	62,26	0,339	60,26	63,27	54,72	0,173	53,64	51,02	58,49	0,091
14.	70,63	72,63	66,67	0,378	69,23	71,58	64,58	0,348	69,23	65,26	77,08	0,400	60,14	73,68	33,33	0,073
15.	72,11	71,88	72,55	0,426	72,79	73,96	70,59	0,430	63,27	69,79	50,98	0,204	60,54	55,21	70,59	0,246
16.	64,24	63,27	66,04	0,280	60,93	58,16	66,04	0,231	64,24	59,18	73,58	0,313	59,60	65,31	49,06	0,140
17.	66,21	64,89	68,63	0,321	63,45	63,83	62,75	0,255	61,38	56,38	70,59	0,258	51,72	46,81	60,78	0,073
18.	62,76	68,82	51,92	0,204	59,31	61,29	55,77	0,165	68,97	69,89	67,31	0,360	49,66	58,06	34,62	-0,072
19.	60,71	62,92	56,86	0,192	62,14	64,04	58,82	0,222	55,00	50,56	62,75	0,129	65,71	76,40	47,06	0,242
20.	65,73	65,22	66,67	0,306	65,03	68,48	58,82	0,266	55,24	48,91	66,67	0,151	50,35	38,04	72,55	0,107
21.	58,62	54,84	65,38	0,194	60,69	63,44	55,77	0,186	52,41	43,01	69,23	0,121	51,72	54,84	46,15	0,010

cv. – číslo cvičení, ACC – úspěšnost klasifikace, SN – senzitivita, SP – Specificita, MCC – Matthewsův korelační koeficient

1. způsob parametrizace: všechny parametry

2. způsob parametrizace: F_0 , krátkodobá energie, TKEO

3. způsob parametrizace: MFCC

4. způsob parametrizace: LPC, formanty

Tab. 9.4: Výsledky rozpoznání pacientů – soubory po de-identifikaci podle **asymetrické** bortící funkce.

cv.	Způsob parametrizace															
	1.				2.				3.				4.			
	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]
1.	60,53	60,61	60,38	0,200	63,82	67,68	56,60	0,235	55,92	56,57	54,72	0,108	40,79	24,24	71,70	-0,044
2.	61,59	64,65	55,77	0,196	60,26	64,65	51,92	0,160	61,59	66,67	51,92	0,181	52,32	55,56	46,15	0,016
3.	65,79	65,66	66,04	0,303	65,13	65,66	64,15	0,286	61,84	59,60	66,04	0,244	46,05	29,29	77,36	0,071
4.	57,62	58,59	55,77	0,137	60,93	63,64	55,77	0,186	52,32	51,52	53,85	0,051	51,66	46,46	61,54	0,077
5.	65,99	68,04	62,00	0,288	63,27	65,98	58,00	0,230	57,14	57,73	56,00	0,130	50,34	32,99	84,00	0,181
6.	62,00	60,00	66,00	0,245	64,67	65,00	64,00	0,275	58,67	57,00	62,00	0,179	46,67	33,00	74,00	0,072
7.	64,24	63,00	66,67	0,281	65,56	62,00	72,55	0,327	59,60	60,00	58,82	0,179	48,34	35,00	74,51	0,097
8.	58,17	58,00	58,49	0,157	60,78	64,00	54,72	0,180	57,52	58,00	56,60	0,139	47,06	45,00	50,94	-0,039
9.	60,78	58,00	66,04	0,229	62,09	62,00	62,26	0,232	61,44	61,00	62,26	0,222	49,02	39,00	67,92	0,068
10.	66,67	71,00	58,49	0,287	62,09	64,00	58,49	0,216	60,13	61,00	58,49	0,186	46,41	42,00	54,72	-0,032
11.	59,06	62,89	51,92	0,143	65,77	69,07	59,62	0,278	61,07	60,82	61,54	0,214	62,42	73,20	42,31	0,158
12.	60,93	65,31	52,83	0,176	64,24	69,39	54,72	0,236	54,97	55,10	54,72	0,094	60,93	58,16	66,04	0,231
13.	64,24	67,35	58,49	0,250	68,87	68,37	69,81	0,366	60,26	63,27	54,72	0,173	57,62	55,10	62,26	0,166
14.	64,34	65,26	62,50	0,264	68,53	70,53	64,58	0,337	55,24	51,58	62,50	0,133	49,65	50,53	47,92	-0,015
15.	70,07	73,96	62,75	0,359	67,35	68,75	64,71	0,322	62,59	61,46	64,71	0,249	60,54	54,17	72,55	0,256
16.	62,25	61,22	64,15	0,242	63,58	61,22	67,92	0,278	57,62	58,16	56,60	0,141	53,64	54,08	52,83	0,066
17.	60,69	60,64	60,78	0,205	57,93	60,64	52,94	0,131	59,31	58,51	60,78	0,184	46,90	36,17	66,67	0,028
18.	63,45	70,97	50,00	0,209	60,69	64,52	53,85	0,178	60,00	65,59	50,00	0,153	49,66	43,01	61,54	0,044
19.	62,14	64,04	58,82	0,222	60,71	64,04	54,90	0,184	61,43	55,06	72,55	0,267	59,29	68,54	43,14	0,117
20.	58,74	57,61	60,78	0,176	55,24	55,43	54,90	0,099	53,85	46,74	66,67	0,130	52,45	41,30	72,55	0,138
21.	62,07	59,14	67,31	0,254	56,55	55,91	57,69	0,131	55,17	49,46	65,38	0,143	60,69	70,97	42,31	0,135

cv. – číslo cvičení, ACC – úspěšnost klasifikace, SN – senzitivita, SP – Specificita, MCC – Matthewsův korelační koeficient

1. způsob parametrizace: všechny parametry

2. způsob parametrizace: F_0 , krátkodobá energie, TKEO

3. způsob parametrizace: MFCC

4. způsob parametrizace: LPC, formanty

Tab. 9.5: Výsledky rozpoznání pacientů – soubory po de-identifikaci podle **symetrické** bortící funkce.

cv.	Způsob parametrizace															
	1.				2.				3.				4.			
	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]
1.	61,84	60,61	64,15	0,236	61,18	65,66	52,83	0,179	57,24	61,62	49,06	0,103	43,42	29,29	69,81	-0,009
2.	60,93	63,64	55,77	0,186	62,91	64,65	59,62	0,232	64,24	69,70	53,85	0,230	43,05	37,37	53,85	-0,085
3.	67,11	65,66	69,81	0,339	64,47	64,65	64,15	0,276	66,45	67,68	64,15	0,306	48,68	31,31	81,13	0,134
4.	62,91	63,64	61,54	0,240	60,26	62,63	55,77	0,176	57,62	59,60	53,85	0,128	50,99	48,48	55,77	0,040
5.	65,31	67,01	62,00	0,278	65,31	67,01	62,00	0,278	64,63	65,98	62,00	0,267	49,66	31,96	84,00	0,171
6.	64,67	64,00	66,00	0,284	63,33	65,00	60,00	0,238	63,33	64,00	62,00	0,247	47,33	35,00	72,00	0,070
7.	69,54	69,00	70,59	0,377	66,23	64,00	70,59	0,328	64,90	66,00	62,75	0,274	52,98	43,00	72,55	0,152
8.	62,09	63,00	60,38	0,224	62,75	68,00	52,83	0,203	60,78	62,00	58,49	0,196	50,33	54,00	43,40	-0,025
9.	59,48	60,00	58,49	0,176	63,40	63,00	64,15	0,259	60,13	66,00	49,06	0,147	50,33	43,00	64,15	0,069
10.	60,78	65,00	52,83	0,172	60,13	60,00	60,38	0,194	56,21	60,00	49,06	0,087	47,06	50,00	41,51	-0,081
11.	62,42	63,92	59,62	0,226	61,74	68,04	50,00	0,177	61,74	62,89	59,62	0,216	65,77	80,41	38,46	0,205
12.	67,55	70,41	62,26	0,317	68,87	74,49	58,49	0,326	61,59	70,41	45,28	0,157	56,29	53,06	62,26	0,146
13.	64,24	67,35	58,49	0,250	66,23	67,35	64,15	0,303	62,91	64,29	60,38	0,237	64,24	70,41	52,83	0,229
14.	67,83	69,47	64,58	0,326	67,83	68,42	66,67	0,334	62,24	64,21	58,33	0,215	54,55	63,16	37,50	0,006
15.	72,11	72,92	70,59	0,419	70,07	71,88	66,67	0,372	63,95	70,83	50,98	0,215	59,18	53,13	70,59	0,227
16.	61,59	59,18	66,04	0,241	64,24	63,27	66,04	0,280	61,59	59,18	66,04	0,241	61,59	69,39	47,17	0,164
17.	66,90	67,02	66,67	0,323	58,62	59,57	56,86	0,157	59,31	55,32	66,67	0,210	53,79	44,68	70,59	0,149
18.	57,24	61,29	50,00	0,109	60,69	66,67	50,00	0,164	58,62	63,44	50,00	0,131	56,55	59,14	51,92	0,107
19.	60,00	61,80	56,86	0,181	60,71	62,92	56,86	0,192	55,71	51,69	62,75	0,139	66,43	79,78	43,14	0,244
20.	59,44	60,87	56,86	0,171	60,14	63,04	54,90	0,173	59,44	59,78	58,82	0,179	47,55	33,70	72,55	0,064
21.	62,07	60,22	65,38	0,246	53,10	53,76	51,92	0,055	60,69	53,76	73,08	0,259	60,69	70,97	42,31	0,135

cv. – číslo cvičení, ACC – úspěšnost klasifikace, SN – senzitivita, SP – Specificita, MCC – Matthewsův korelační koeficient

1. způsob parametrizace: všechny parametry

2. způsob parametrizace: F_0 , krátkodobá energie, TKEO

3. způsob parametrizace: MFCC

4. způsob parametrizace: LPC, formanty

Tab. 9.6: Výsledky rozpoznání pacientů – soubory po de-identifikaci podle **výkonové** bortící funkce.

cv.	Způsob parametrizace															
	1.				2.				3.				4.			
	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]
1.	66,45	70,71	58,49	0,285	67,11	70,71	60,38	0,302	61,84	67,68	50,94	0,182	42,76	35,35	56,60	-0,079
2.	62,91	66,67	55,77	0,217	66,23	69,70	59,62	0,284	59,60	64,65	50,00	0,142	51,66	54,55	46,15	0,007
3.	65,13	68,69	58,49	0,264	68,42	69,70	66,04	0,344	63,16	71,72	47,17	0,189	51,32	36,36	79,25	0,161
4.	61,59	65,66	53,85	0,188	64,90	65,66	63,46	0,278	54,97	61,62	42,31	0,038	54,97	52,53	59,62	0,115
5.	66,67	68,04	64,00	0,307	65,31	65,98	64,00	0,286	57,14	58,76	54,00	0,121	57,14	44,33	82,00	0,261
6.	62,00	62,00	62,00	0,227	65,33	64,00	68,00	0,302	60,67	66,00	50,00	0,154	46,00	29,00	80,00	0,097
7.	66,89	69,00	62,75	0,304	71,52	75,00	64,71	0,386	57,62	60,00	52,94	0,123	54,97	46,00	72,55	0,179
8.	54,25	60,00	43,40	0,033	63,40	70,00	50,94	0,206	53,59	58,00	45,28	0,032	52,29	58,00	41,51	-0,005
9.	58,17	60,00	54,72	0,141	62,09	66,00	54,72	0,200	56,21	62,00	45,28	0,071	49,67	38,00	71,70	0,097
10.	64,71	70,00	54,72	0,242	56,86	59,00	52,83	0,113	58,17	59,00	56,60	0,149	41,83	41,00	43,40	-0,149
11.	61,74	65,98	53,85	0,192	62,42	68,04	51,92	0,195	56,38	64,95	40,38	0,053	59,73	69,07	42,31	0,114
12.	56,95	60,20	50,94	0,107	62,25	64,29	58,49	0,219	54,97	64,29	37,74	0,020	55,63	58,16	50,94	0,087
13.	66,23	71,43	56,60	0,275	71,52	73,47	67,92	0,402	61,59	66,33	52,83	0,186	51,66	38,78	75,47	0,144
14.	64,34	64,21	64,58	0,273	65,73	64,21	68,75	0,312	60,84	61,05	60,42	0,203	50,35	51,58	47,92	-0,005
15.	68,03	68,75	66,67	0,340	65,99	66,67	64,71	0,301	62,59	71,88	45,10	0,171	59,86	55,21	68,63	0,227
16.	64,90	66,33	62,26	0,275	66,89	68,37	64,15	0,313	55,63	62,24	43,40	0,055	58,28	64,29	47,17	0,112
17.	64,83	64,89	64,71	0,284	60,69	62,77	56,86	0,189	63,45	65,96	58,82	0,239	55,17	45,74	72,55	0,179
18.	57,93	65,59	44,23	0,097	63,45	62,37	65,38	0,266	56,55	62,37	46,15	0,083	58,62	67,74	42,31	0,100
19.	63,57	67,42	56,86	0,237	60,00	64,04	52,94	0,166	57,86	51,69	68,63	0,197	60,00	75,28	33,33	0,092
20.	58,74	61,96	52,94	0,144	62,24	64,13	58,82	0,221	58,04	58,70	56,86	0,149	45,45	31,52	70,59	0,022
21.	62,76	63,44	61,54	0,241	56,55	55,91	57,69	0,131	54,48	59,14	46,15	0,051	65,52	77,42	44,23	0,226

cv. – číslo cvičení, ACC – úspěšnost klasifikace, SN – senzitivita, SP – Specificita, MCC – Matthewsův korelační koeficient

1. způsob parametrizace: všechny parametry

2. způsob parametrizace: F_0 , krátkodobá energie, TKEO

3. způsob parametrizace: MFCC

4. způsob parametrizace: LPC, formanty

Tab. 9.7: Výsledky rozpoznání pacientů – soubory po de-identifikaci podle **kvadratické** bortící funkce.

cv.	Způsob parametrizace															
	1.				2.				3.				4.			
	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]
1.	66,45	69,70	60,38	0,292	65,79	66,67	64,15	0,296	57,24	60,61	50,94	0,111	51,97	46,46	62,26	0,084
2.	64,90	68,69	57,69	0,255	69,54	72,73	63,46	0,351	53,64	56,57	48,08	0,044	53,64	48,48	63,46	0,114
3.	62,50	61,62	64,15	0,246	65,13	63,64	67,92	0,301	62,50	65,66	56,60	0,215	47,37	30,30	79,25	0,103
4.	61,59	62,63	59,62	0,212	63,58	64,65	61,54	0,250	61,59	64,65	55,77	0,196	53,64	57,58	46,15	0,036
5.	62,59	64,95	58,00	0,220	65,99	68,04	62,00	0,288	59,86	58,76	62,00	0,197	49,66	35,05	78,00	0,134
6.	64,67	65,00	64,00	0,275	64,00	63,00	66,00	0,274	60,00	60,00	60,00	0,189	49,33	33,00	82,00	0,157
7.	72,19	71,00	74,51	0,433	67,55	68,00	66,67	0,331	70,86	72,00	68,63	0,390	53,64	39,00	82,35	0,217
8.	58,82	60,00	56,60	0,159	57,52	59,00	54,72	0,131	52,94	57,00	45,28	0,022	56,21	63,00	43,40	0,062
9.	62,09	64,00	58,49	0,216	60,78	64,00	54,72	0,180	61,44	64,00	56,60	0,198	50,98	50,00	52,83	0,027
10.	63,40	68,00	54,72	0,221	65,36	69,00	58,49	0,266	54,25	56,00	50,94	0,066	47,06	38,00	64,15	0,021
11.	63,09	68,04	53,85	0,213	65,10	70,10	55,77	0,253	62,42	62,89	61,54	0,234	59,06	74,23	30,77	0,053
12.	65,56	70,41	56,60	0,264	60,93	65,31	52,83	0,176	63,58	66,33	58,49	0,240	59,60	51,02	75,47	0,256
13.	66,23	66,33	66,04	0,310	67,55	70,41	62,26	0,317	66,89	65,31	69,81	0,336	54,30	41,84	77,36	0,192
14.	72,73	73,68	70,83	0,427	64,34	65,26	62,50	0,264	66,43	73,68	52,08	0,255	55,94	65,26	37,50	0,027
15.	65,31	66,67	62,75	0,282	63,27	66,67	56,86	0,227	62,59	75,00	39,22	0,148	58,50	50,00	74,51	0,237
16.	64,90	66,33	62,26	0,275	63,58	64,29	62,26	0,255	66,89	66,33	67,92	0,328	58,94	55,10	66,04	0,202
17.	70,34	73,40	64,71	0,371	60,69	61,70	58,82	0,197	60,00	54,26	70,59	0,238	56,55	48,94	70,59	0,189
18.	60,00	63,44	53,85	0,168	57,93	61,29	51,92	0,128	66,90	66,67	67,31	0,327	52,41	47,31	61,54	0,085
19.	61,43	64,04	56,86	0,203	60,71	62,92	56,86	0,192	60,00	60,67	58,82	0,188	55,71	65,17	39,22	0,044
20.	58,04	56,52	60,78	0,166	59,44	57,61	62,75	0,195	55,94	47,83	70,59	0,179	50,35	39,13	70,59	0,097
21.	59,31	56,99	63,46	0,196	60,69	62,37	57,69	0,193	59,31	52,69	71,15	0,230	53,10	47,31	63,46	0,104

cv. – číslo cvičení, ACC – úspěšnost klasifikace, SN – senzitivita, SP – Specificita, MCC – Matthewsův korelační koeficient

1. způsob parametrizace: všechny parametry

2. způsob parametrizace: F_0 , krátkodobá energie, TKEO

3. způsob parametrizace: MFCC

4. způsob parametrizace: LPC, formanty

Tab. 9.8: Výsledky rozpoznání pacientů – soubory po de-identifikaci podle **bilineární** bortící funkce.

cv.	Způsob parametrizace															
	1.				2.				3.				4.			
	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]	ACC [%]	SN [%]	SP [%]	MCC [-]
1.	67,76	67,68	67,92	0,341	69,08	72,73	62,26	0,341	62,50	63,64	60,38	0,230	51,97	44,44	66,04	0,102
2.	66,23	71,72	55,77	0,269	64,90	67,68	59,62	0,263	60,93	63,64	55,77	0,186	52,98	46,46	65,38	0,114
3.	66,45	64,65	69,81	0,329	65,79	65,66	66,04	0,303	67,11	68,69	64,15	0,316	50,66	36,36	77,36	0,141
4.	58,94	60,61	55,77	0,156	66,23	65,66	67,31	0,314	56,29	59,60	50,00	0,092	55,63	58,59	50,00	0,082
5.	64,63	67,01	60,00	0,259	64,63	64,95	64,00	0,276	60,54	62,89	56,00	0,180	50,34	36,08	78,00	0,144
6.	67,33	63,00	76,00	0,368	61,33	63,00	58,00	0,199	59,33	60,00	58,00	0,170	50,00	34,00	82,00	0,167
7.	69,54	69,00	70,59	0,377	66,89	66,00	68,63	0,329	66,23	69,00	60,78	0,286	54,30	40,00	82,35	0,226
8.	59,48	63,00	52,83	0,152	57,52	60,00	52,83	0,123	58,82	62,00	52,83	0,142	52,29	57,00	43,40	0,004
9.	64,05	67,00	58,49	0,246	60,78	63,00	56,60	0,188	62,75	66,00	56,60	0,218	52,94	51,00	56,60	0,072
10.	62,75	64,00	60,38	0,233	67,32	72,00	58,49	0,298	60,13	58,00	64,15	0,211	42,48	35,00	56,60	-0,082
11.	63,09	67,01	55,77	0,221	64,43	69,07	55,77	0,242	55,70	59,79	48,08	0,076	57,05	72,16	28,85	0,011
12.	63,58	69,39	52,83	0,218	61,59	64,29	56,60	0,201	62,91	67,35	54,72	0,215	57,62	47,96	75,47	0,229
13.	64,90	66,33	62,26	0,275	67,55	71,43	60,38	0,310	67,55	67,35	67,92	0,338	52,98	38,78	79,25	0,184
14.	67,83	66,32	70,83	0,352	62,94	62,11	64,58	0,253	66,43	73,68	52,08	0,255	57,34	66,32	39,58	0,058
15.	67,35	68,75	64,71	0,322	64,63	67,71	58,82	0,256	63,27	73,96	43,14	0,174	64,63	60,42	72,55	0,314
16.	63,58	63,27	64,15	0,262	64,90	66,33	62,26	0,275	62,91	64,29	60,38	0,237	57,62	48,98	73,58	0,219
17.	65,52	67,02	62,75	0,287	60,69	62,77	56,86	0,189	62,07	58,51	68,63	0,259	53,79	45,74	68,63	0,140
18.	64,14	68,82	55,77	0,241	58,62	61,29	53,85	0,146	61,38	63,44	57,69	0,204	48,28	45,16	53,85	-0,010
19.	59,29	60,67	56,86	0,169	62,86	64,04	60,78	0,240	62,86	64,04	60,78	0,240	58,57	67,42	43,14	0,106
20.	60,14	59,78	60,78	0,197	60,84	58,70	64,71	0,224	56,64	50,00	68,63	0,180	51,75	41,30	70,59	0,118
21.	59,31	59,14	59,62	0,180	61,38	63,44	57,69	0,204	57,93	49,46	73,08	0,219	55,86	54,84	57,69	0,120

cv. – číslo cvičení, ACC – úspěšnost klasifikace, SN – senzitivita, SP – Specificita, MCC – Matthewsův korelační koeficient

1. způsob parametrizace: všechny parametry

2. způsob parametrizace: F_0 , krátkodobá energie, TKEO

3. způsob parametrizace: MFCC

4. způsob parametrizace: LPC, formanty

Aby bylo možné určit míru vlivu transformace na příznaky kvantifikující Parkinsonovu nemoc, přesnosti všech výsledných měření po transformaci zvukových nahrávek řečníků (tabulka 9.4 až 9.8) byly pomocí t-testu porovnány s referenčními výsledky, tedy přesností klasifikace před transformací (tabulka 9.3).

T-test je párový test, který umožňuje ověřit hypotézu, zda dvě normální rozdělení mající stejný rozptyl, z nichž pocházejí dva výběry, mají stejné střední hodnoty [22]. Pro tento účel byla použita funkce `ttest` v prostředí Matlab. V tabulce 9.9 jsou zobrazeny výsledné parametry p , které byly opět počítány pro všechny čtyři způsoby parametrizace. V tabulce jsou také zvýrazněny hodnoty, pro které platí $p > 0,05$. Ty totiž neodporují nulové hypotéze a daná data tedy teoreticky nejsou ovlivněna de-identifikací.

Jelikož se t-test používá pouze pro hodnoty s normálním rozdělením, data byla pomocí Lillieforstova testu otestována, zda normálnímu rozdělení skutečně odpovídají.

Tab. 9.9: Výsledky testování výsledků klasifikace T-testem.

způsob parametrizace	bortící funkce				
	asymetrická	symetrická	výkonová	kvadratická	bilineární
1.	0,0161	0,312	0,094	0,789	0,739
2.	0,0023	0,003	0,251	0,015	0,055
3.	0,133	0,381	0,095	0,269	0,125
4.	0,451	0,363	0,694	0,448	0,417

Podle výsledků testování má na příznaky Parkinsonovy nemoci nejmenší vliv bortící funkce výkonová a bilineární. Nejlepších výsledků také bylo dosaženo při použití třetího a čtvrtého způsobu parametrizace. Naopak největší vliv vykazují zvukové soubory po použití asymetrické bortící funkce a druhého způsobu parametrizace.

9.3 Vyhodnocení výsledků

Úkolem této práce je navrhnout systém, který pacienta de-identifikuje a zároveň zachová v signálu informaci nutnou k diagnóze hypokinetické dysartrie. Proto byly porovnány výsledky klasifikace řečníků a pacientů.

Pro de-identifikaci hypokinetické dysartrie se nejvíce osvědčila funkce výkonová a bilineární. Při klasifikaci řečníků vycházela nejmenší úspěšnost při použití kvadratické funkce, případně bilineární funkce, u které byla úspěšnost rozeznání pacientů vyšší pouze o 1,23%. Z výsledků těchto testů můžeme určit funkci, která si vedla dobře v obou případech. Nejlepší výsledky měla funkce kvadratická, dobře si vedla i funkce bilineární.

10 ZÁVĚR

Cílem práce bylo najít způsob, díky kterému by bylo možné volně šířit databáze řečníků postižených hypokinetickou dysartrií. Toho se dá docílit pomocí de-identifikace řeči nahraných pacientů. Musí však být kladen důraz na zachování důležitých řečových příznaků, které jsou používány k diagnostice nemoci.

Na začátku práce byla popsána hypokinetická dysartrie a Parkinsonova nemoc. Poté jsme rozebrali klinické příznaky ovlivňující řečový projev pacientů a způsoby, jak se dají tyto příznaky identifikovat v řečových nahrávkách.

Dále jsou rozebrány typy de-identifikace řeči a systémy potřebné k evaluaci výsledků (identifikační systémy). V další kapitole byl navržen systém, který provádí de-identifikaci řeči a následně výsledky porovnává s původním řečovým signálem a rozhoduje o úrovni de-identifikace. V následující kapitole je popsána databáze zvukových souborů, která byla v práci použita.

Následující část práce tedy pojednává o systému, který má za úkol de-identifikovat tyto soubory a následně provést evaluaci výsledků. Samotná de-identifikace byla prováděna pomocí normalizace vokálního traktu. Použito bylo pět bortících funkcí, přičemž byl pro každou zvolen jeden bortící faktor. Tento faktor byl vždy zvolen na základě poslechu a porovnávání de-identifikovaných a původních nahrávek pěti posluchači.

Následná evaluace byla prováděna na základě Gaussových smíšených modelů (GMM). Systém vyhodnocoval úroveň de-identifikace pomocí rozpoznávání řečníků. Dále pomocí stejné metody klasifikoval změnu řečových příznaků poukazujících na hypokinetickou dysartrii. Při vyhodnocení výsledků byla vybrána funkce, při jejímž použití měl systém klasifikace řečníků nejmenší úspěšnost. Funkce však zároveň musela mít co nejmenší vliv na příznaky popisující hypokinetickou dysartrii v řečovém signálu. Proto byl učiněn kompromis a byly vybrány dvě bortící funkce – kvadratická, která si vedla nejlépe v de-identifikaci a funkce bilineární, která měla dobré výsledky v obou případech.

Aby bylo možné bortící funkce otestovat podrobněji, bylo by nutné každou z nich otestovat pro různé hodnoty bortícího faktoru. Jeho hodnota má totiž vliv na míru změny frekvenčního spektra de-identifikovaných zvukových souborů. Pro větší přesnost klasifikace by také bylo velmi vhodné použít větší množství parametrů při parametrizaci nahrávek – a to hlavně při klasifikaci pacientů.

LITERATURA

- [1] ANDĚL, Jiří. *Základy matematické statistiky*. Praha: Matfyzpress, 2005. ISBN 80-867-3240-1.
- [2] ARMSTRONG, Richard A. Visual signs and symptoms of Parkinson's disease. *Clinical and Experimental Optometry*. 2008, extbf91(2), 129-138. DOI: 10.1111/j.1444-0938.2007.00211.x. ISSN 0816-4622. Dostupné také z: <http://doi.wiley.com/10.1111/j.1444-0938.2007.00211.x>
- [3] BORSEL, John Van a John A. TETNOWSKI. Fluency disorders in genetic syndromes. *Journal of Fluency Disorders*. 2007, extbf32(4), 279-296. DOI: 10.1016/j.jfludis.2007.07.002. ISSN 0094730x. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/S0094730X07000411>
- [4] CAMMACK, Richard et al. *Oxford dictionary of biochemistry and molecular biology*. Rev. ed. Oxford: Oxford University Press, 2006, s. 720. ISBN 9781613441138.
- [5] DUFFY, Joseph R. *Motor speech disorders: substrates, differential diagnosis, and management*. Third edition. St. Louis, Missouri: Elsevier, 2013, 187 - 215. ISBN 9780323072007.
- [6] FARRÚS, Mireia, Javier MIREIA a Pascual EJARQUE. *Jitter and Shimmer Measurements for Speaker Recognition* [online]. Barcelona, 2007 [cit. 2015-11-11]. Dostupné z: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.160.9924>
- [7] FREED, Donald B. *Motor speech disorders: diagnosis and treatment*. 2nd ed. Clifton Park, NY: Delmar, Cengage Learning, 2012, xvii, 302 p. ISBN 11-111-3827-3.
- [8] ISHIZUKA, Kentaro, Tomohiro NAKATANI, Masakiyo FUJIMOTO a Noboru MIYAZAKI. Noise robust voice activity detection based on periodic to aperiodic component ratio. *Speech Communication*. 2010, extbf52(1), 41-60. DOI: 10.1016/j.specom.2009.08.003. ISSN 01676393. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/S0167639309001277>
- [9] JANKOVIC, J. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery*. 2008, extbf79(4), 368-376. DOI: 10.1136/jnnp.2007.131045. ISSN 0022-3050. Dostupné také z: <http://jnnp.bmj.com/cgi/doi/10.1136/jnnp.2007.131045>

- [10] JAROLÍN, Milan. *Melovské frekvenční keprální koeficienty MFCC*. Plzeň, 2012. Bakalářská práce. Západočeská univerzita v Plzni.
- [11] JIN, Qin, Tanja SCHULTZ a AlexWaibel. *Phonetic Speaker Identification* [online]. Denver, CO, 2002 [cit. 2015-11-25]. Dostupné z: <http://www.cs.cmu.edu/~tanja/Papers/ICSLP02-qjin.pdf>. CarnegieMellonUniversity.
- [12] JUSTIN, Tadej, Vitomir ŠTRUC, Simon DOBŘÍŠEK, Boštjan VESNICER, Ivo IPŠIČ a France MIHELIČ. *Speaker De-identification using Diphone Recognition and Speech Synthesis* [online]. Ljubljana, 2015 [cit. 2015-12-06]. Dostupné z: <http://ieeexplore.ieee.org.ezproxy.lib.vutbr.cz/stamp/stamp.jsp?tp=&arnumber=7285021>. Faculty of Electrical Engineering, University of Ljubljana.
- [13] MEKYSKA, Jiří. *Číslíkové zpracování řeči: Lineární predikční analýza řeči*. Brno: Ústav telekomunikací, FEKT, VUT.
- [14] MEKYSKA, Jiri, Eva JANOUSOVA, Pedro GOMEZ-VILDA, et al. Robust and complex approach of pathological speech signal analysis. *Neurocomputing*. 2015, extbf2015(167), 94-111. DOI: 10.1016/j.neucom.2015.02.085. ISSN 09252312. Dostupné také z: <http://linkinghub.elsevier.com/retrieve/pii/S0925231215007304>
- [15] MEKYSKA, Jiri, Irena REKTOROVA a Zdenek SMEKAL. Selection of optimal parameters for automatic analysis of speech disorders in Parkinson's disease. *2011 34th International Conference on Telecommunications and Signal Processing (TSP)*. IEEE, 2011, , 408-412. DOI: 10.1109/TSP.2011.6043700. ISBN 978-1-4577-1410-8. Dostupné také z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6043700>
- [16] NOVOTNY, Michal, Jan RUSZ, Roman CMEJLA a Evzen RUZICKA. Automatic Evaluation of Articulatory Disorders in Parkinson's Disease. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2014, extbf22(9), 1366-1378. DOI: 10.1109/TASLP.2014.2329734. ISSN 2329-9290. Dostupné také z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6827910>
- [17] RAGHUNATHAN, Anusha. *Evaluation of Intelligibility and Speaker Similarity of Voice transformation*. Kuntucky, 2011. Diplomová práce. University of Kentucky.

- [18] RUSZ, J., R. CMEJLA, H. RUZICKOVA a E. RUZICKA. Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *The Journal of the Acoustical Society of America*. 2011, extbf129(1), 350-. DOI: 10.1121/1.3514381. ISSN 00014966. Dostupné také z: <http://scitation.aip.org/content/asa/journal/jasa/129/1/10.1121/1.3514381>
- [19] SAHEER, Lakshmi, John DINES a Philip N. GARNER. Vocal Tract Length Normalization for Statistical Parametric Speech Synthesis. *IEEE Transactions on Audio, Speech, and Language Processing*. 2012, extbf20(7), 2134-2148. DOI: 10.1109/TASL.2012.2198058. ISSN 1558-7916. Dostupné také z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6196182>
- [20] SKODDA, Sabine, Heiko RINSCHKE a Uwe SCHLEGEL. Progression of dysprosody in Parkinson's disease over time-A longitudinal study. *Movement Disorders*. 2009, extbf24(5), 716-722. DOI: 10.1002/mds.22430. ISSN 08853185. Dostupné také z: <http://doi.wiley.com/10.1002/mds.22430>
- [21] SMÉKAL, Zdeněk. *Číslíkové zpracování řeči: (MZPR)*. Brno, 2010. Skripta. Vysoké učení technické v Brně.
- [22] STADNICZUK, Denis, Gregor BAUCKMANN a David SUENDERMANNOEFT. *An Open-Source Octave Toolbox for VTLN-Based Voice Conversion* [online]. Stuttgart, Germany: Baden-Wuerttemberg Cooperative State University [cit. 2016-05-05]. Dostupné z: <http://suendermann.com/su/pdf/gsc12013.pdf>
- [23] ZAJÍC, Zbyněk. *Adaptace akustického modelu v úloze s malým množstvím adaptčních dat*. Plzeň, 2012. Disertační práce. Fakulta aplikovaných věd Západočeské Univerzity v Plzni, katedra kybernetiky.
- [24] ZHOU, Yuhuan, Jinming WANG a Xiongwei ZHANG. Research on Adaptive Speaker Identification Based on GMM. *2009 International Forum on Computer Science-Technology and Applications*. IEEE, 2009, , 330-332. DOI: 10.1109/IFCSTA.2009.203. ISBN 978-1-4244-5422-8. Dostupné také z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5384570>

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

ACC	přesnost
DDK	diadochokinetické testy
DROPSY	diphone recognition and speech synthesis
F_0	základní frekvence
FFT	fast Fourier transform - rychlá Fourierova transformace
FN	false negative – omylně určený negativní výsledek
FP	false positive – omylně určený pozitivní výsledek
GMM	Gaussian mixture model - Gaussův smíšený model
HD	hypokinetická dysartrie
HNR	harmonics to noise ratio - poměr harmonických frekvencí a šumu
IFFT	inverse fast Fourier transform - inverzní rychlá Fourierova transformace
ISD	interpause speech duration
LPC	linear predictive coding – lineární predikční kódování
LSTER	low short-time energy ratio
MFCC	mel frequency cepstral coefficient - melovské frekvenční keprální koeficienty
MSPD	median of power spectral density
NST	net speech time
PARADE	periodic to aperiodic component ratio based activity detection
PARCZ	Czech Parkinsonian Speech Database
PN	Parkinsonova nemoc
PPQ	five-point pitch perturbation
PPR	procentual pause ratio
PSOLA	pitch synchronous overlap and add

RAP	relative average perturbation
SID	speaker identification
SN	senzitivita
SP	specificita
SPIR	speech index of rhythmicity
STS	total speech time
VTLN	vocal tract length normalization – normalizace vokálního traktu
TKEO	Teager-Kaiser Energy Operator – Teagerův-Kaiserův energetický operátor
TN	true negative – správně určený negativní výsledek
TP	true positive – správně určený pozitivní výsledek
TSR	total speech rate
VAD	voice activity detection
VOT	voice onset time - doba navození řeči

SEZNAM PŘÍLOH

A Obsah přiloženého CD

49

A OBSAH PŘILOŽENÉHO CD

Přiložené CD obsahuje:

- Radoslav Kárník - bakalářská práce.pdf - elektronická verze bakalářské práce
- Implementace - složka se zdrojovými soubory programu

Složka Implementace obsahuje všechny použité skripty. Skript `Program.m` spustí zvolený druh de-identifikace u zvoleného souboru. Skript `parametrizace.m` obsahuje kód použitý k parametrizaci souborů, `test_r.m` byl použit ke klasifikaci řečníků a `test_p.m` ke klasifikaci nemocí.

Většina skriptů vyžaduje přítomnost vybraných řečových cvičení z databáze PARCZ ve složce s danými skripty. Tato databáze však není volně šiřitelná, proto se na CD nenachází.