

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

SHLUKOVÁ ANALÝZA V OBLASTI ČÍSLICOVÉHO ZPRACOVÁNÍ
PATOLOGICKÉ ŘEČI

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

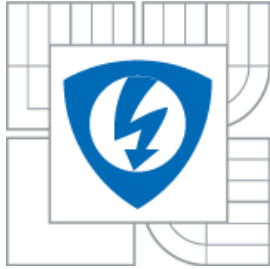
AUTOR PRÁCE
AUTHOR

KAREL ČAPEK

BRNO 2015



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATION

SHLUKOVÁ ANALÝZA V OBLASTI ČÍSLICOVÉHO ZPRACOVÁNÍ PATOLOGICKÉ ŘEČI

CLUSTER ANALYSIS IN THE FIELD OF PATHOLOGICAL SPEECH SIGNAL PROCESSING

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

KAREL ČAPEK

VEDOUCÍ PRÁCE
SUPERVISOR

Ing. ZOLTÁN GALÁŽ

BRNO 2015



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor

Teleinformatika

Student: Karel Čapek

ID: 145980

Ročník: 3

Akademický rok: 2014/2015

NÁZEV TÉMATU:

Shluková analýza v oblasti číslicového zpracování patologické řeči

POKYNY PRO VYPRACOVÁNÍ:

Patologický hlas se vyznačuje nesprávnou funkcí hlasových a artikulačních orgánů. Existuje mnoho patologií řeči. Všeobecně známé dělení patologií na základě poruch hlasového ústrojí od Titzeho: zánětlivé nemoci (laryngitida, bronchitida, atd.), systematické změny (dehydratace, hormonální změny atd.), mechanické namáhání (polypy, vředy, atd.), povrchové podráždění (gastroezofageální reflux, atd.), změny tkáně (rakovina hrtanu, cysty, atd.), neurologické a svalové změny (Parkinsonova choroba, amyotrofická laterální skleróza, atd.) a abnormality svalstva (dysfonie, ventrikulární fonace). V rámci této práce bude proveden rozbor jednotlivých patologií řeči a následně bude vytvořen nový systém dělení těchto patologií, založený na algoritmech tzv. shlukové analýzy. Tyto algoritmy budou naprogramované v jazycích MATLAB, Java nebo C / C ++ a budou testovány na databázi MEEI (The Massachusetts Eye and Ear Infirmary database).

DOPORUČENÁ LITERATURA:

[1] PSUTKA, J. et. al. Mluvíme s počítačem česky. 1. vydání. Praha: Academia, 2006. 752 s. ISBN 80-200-1309-0

[2] SMÉKAL, Z. Číslicové zpracování řeči (MZPR). Elektronická skripta pro magisterská studia, Ústav telekomunikací, FEEC, VUT Brno, 2009.

[3] JIRSÍK, Václav, HRÁČEK, Petr. Neuronové systémy, expertní systémy a rozpoznávání řeči. Brno : [s.n.], 2000. 106 s. Skriptum FEKT VUT v Brně.

Termín zadání: 9.2.2015

Termín odevzdání: 2.6.2015

Vedoucí práce: Ing. Zoltán Galáž

Konzultanti 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

Výzkum popsáný v této bakalářské práci byl realizovaný v laboratořích podpořených projektem Centrum sensorických, informačních a komunikačních systémů (SIX); registrační číslo CZ.1.05/2.1.00/03.0072, operačního programu Výzkum a vývoj pro inovace.

ABSTRAKT

Tato bakalářská práce se zabývá výpočtem parametrů kvantifikujících degradaci řečové produkce způsobenou přítomností určité řečové patologie a následným rozřazením uvažovaných patologií do skupin pomocí algoritmu k-means. Účelem bylo nalezení skupin patologií, které navzdory možnému rozdílnému původu v podobné míře postihují fonační a artikulační schopnosti řečníka a znehodnocují kvalitu řeči jako takové. V práci bylo využito fonaci hlásky „a“, jako nejběžněji se vyskytujícího řečového cvičení, kvůli své odolnosti vůči demografickým a lingvistickým charakteristikám jednotlivých řečníků. Dále je na parametry uplatněna předběžná analýza, která má za úkol vybrat parametry nejlépe charakterizující degradaci řečové promluvy. Nakonec jsou tyto vybrané parametry použity pro nalezení výsledných dělení patologií pomocí algoritmu k-means.

KLÍČOVÁ SLOVA

Shluková analýza, řečový parametr, zpracování řeči, patologická řeč, předzpracování, tvorba řeči, řečové ústrojí, předběžná analýza, k-means

ABSTRACT

The bachelor thesis deals with the calculation of speech features that quantifies the degradation of speech production caused by the presence of certain speech pathology and the subsequent classification of considered speech pathologies into several groups using the k-means algorithm. The purpose was to find the groups of pathologies that in spite of possible differences in the origin do affect phonation and articulation skills of the speakers and damage the quality of speech. The work uses the phonation of vowels "a" speech task as the most commonly used speech task in the field of pathological speech processing, because of its resistance to demographic and linguistic characteristics of the speakers. Furthermore, the preliminary analysis was applied to the features in order to select the features to best characterize the degradation of speech production. Finally, the selected features were used to find the resulting groups of pathologies using k-means algorithm.

KEYWORDS

Cluster analysis, parameter speech, speech processing, speech pathology, pre-processing, creation of speech, speech tract, pre-analysis, k-means

ČAPEK, K. *Shluková analýza v oblasti číslicového zpracování patologické řeči*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2015. 55 s. Vedoucím bakalářské práce byl Ing. Zoltán Galáž.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma Shluková analýza v oblasti číslicového zpracování patologické řeči 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 zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne

.....

(podpis autora)

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu mé bakalářské práce, panu Ing. Zoltánu Galážovi, za odborné vedení, konzultace, trpělivost a poskytnutí cenných rad a materiálů potřebných pro realizaci této bakalářské práce.

V Brně dne

.....

(podpis autora)

OBSAH

| | |
|---|-----------|
| Seznam obrázků | 13 |
| Seznam tabulek | 14 |
| Úvod | 15 |
| 1 Úvod do zpracování řečového signálu | 17 |
| 1.1 Tvorba řeči..... | 17 |
| 1.2 Řečové ústrojí | 17 |
| 1.2.1 Dechové ústrojí | 18 |
| 1.2.2 Hlasové ústrojí | 18 |
| 1.2.3 Artikulační ústrojí | 19 |
| 1.3 Matematický model generování řeči..... | 19 |
| 2 Patologie řečového projevu | 21 |
| 2.1 Patologická řeč..... | 21 |
| 2.2 Patologie řečového projevu a řečového ústrojí..... | 21 |
| 2.2.1 Fonace | 21 |
| 2.2.2 Artikulace..... | 22 |
| 2.2.3 Prozodie | 22 |
| 2.2.4 Plynulost řeči | 22 |
| 3 Parametry hodnotící patologie řečového projevu | 24 |
| 3.1 Předzpracování řečového projevu..... | 24 |
| 3.1.1 Preemfáze..... | 24 |
| 3.1.2 Ustředění | 25 |
| 3.1.3 Segmentace | 25 |
| 3.2 Parametry hodnotící fonaci | 26 |
| 3.2.1 Zakladní frekvence | 26 |
| 3.2.2 Jitter | 27 |
| 3.2.3 Shimmer..... | 27 |
| 3.3 Parametry hodnotící energii řečového signálu..... | 27 |

| | |
|---|-----------|
| 3.3.1 STE (Short-time energy)..... | 27 |
| 3.3.2 LSTER (Low short-time energy ratio)..... | 28 |
| 3.3.3 TKEO (Teager-Kaiser energy operator)..... | 28 |
| 3.3.4 MPSD (Median of power spectral density)..... | 28 |
| 3.4 Parametry hodnotící kvalitu řeči..... | 29 |
| 3.4.1 ZCR (Zero-crossing rate)..... | 29 |
| 3.4.2 HZCRR (High zero-crossing rate ratio)..... | 29 |
| 3.4.3 HNR (harmonic-to-noise ratio)..... | 30 |
| 3.5 Parametry hodnotící pohyblivost artikulačních orgánů..... | 30 |
| 3.6 Parametry hodnotící tempo řeči..... | 31 |
| 3.7 Vysokourovňové parametry..... | 31 |
| 4 Shluková analýza | 32 |
| 4.1 Kohenova samoorganizační mapa..... | 32 |
| 4.1.1 Učení neuronové sítě..... | 32 |
| 4.2 K-means..... | 33 |
| 4.2.1 Popis algoritmu..... | 33 |
| 5 Návrh systému | 34 |
| 5.1 Blokové schéma..... | 34 |
| 5.2 Popis systému..... | 35 |
| 6 Realizace systému | 36 |
| 6.1 Databáze řečových záznamů..... | 36 |
| 6.2 Použité parametry..... | 36 |
| 6.3 Načtení záznamů z databáze..... | 38 |
| 6.4 Výpočet parametrů..... | 39 |
| 6.4.1 Následné zpracování..... | 39 |
| 6.5 Předběžná analýza..... | 40 |
| 6.6 Klasifikace k-means..... | 42 |
| 6.7 Konečné výsledky..... | 43 |
| 7 Závěr | 45 |

| | |
|---|-----------|
| Použitá literatura | 47 |
| Seznam symbolů, veličin a zkratk | 48 |
| A Obsah přiloženého DVD | 50 |
| B Tabulka patologií obsažených v MEEI databázi | 51 |
| C Tabulky výsledků | 52 |

SEZNAM OBRÁZKŮ

| | |
|---|----|
| Obr. 1.1: Hlasový trakt člověka. | 18 |
| Obr. 1.2: Matematický model generování řeči. | 20 |
| Obr. 3.1: Zobrazení prvního, druhého a třetího formantu pro samohlásku „a“ | 30 |
| Obr. 6.1: Grafy hustoty pravděpodobnosti pro tři nejlépe korelující parametry (byla uvažována Spearmanova korelace mezi hodnotami parametrů zdravých a patologických řečníků a jejich klinickým stavem, tj. zdravý/nemocní) vypočtené pomocí jádrových odhadů s použitím Gaussova jádra. První řádek značí parametry založené na sledování variability základního tónu a intenzity řeči. Druhý řádek je vyhrazen pro parametry sledující energetické změny ve spektru sledovaného signálu. Třetí řádek je vyhrazen pro parametry pracující na základě sledování pohyblivosti artikulačních orgánů. | 42 |

SEZNAM TABULEK

| | |
|--|----|
| Tab. 6.1: Demografické informace vypočteny z MEEI databáze | 36 |
| Tab. 6.2: Tabulka předběžné analýzy šesti nejlepších parametrů seřazených sestupně podle hodnot Spermenova korelačního koeficientu (ρ). Tabulka obsahuje analýzu pro případ analýzy zdravé a patologické řeči a také analýzy pouze řeči patologické. | 41 |
| Tab. 6.3: Separace zdravé řeči pomocí k-means | 43 |
| Tab. 6.4: Zařazení patologií do skupin podle podobnosti vypočtených výsledků pro druhý scénář. | 44 |
| Tab. B.1: Informace o patologiích. U parametrů, u kterých není uveden český název nebyl tento název dohledán..... | 51 |
| Tab. C.1: Tabulka vypočtených hodnot pro první scénář 1.část. Zkratka sk. znamená skupinu, do které byla daná patologie zařazena. Výskyt potom procentuálně vyjadřuje četnost zařazení do dané skupiny. Jestliže patologie nedosáhla alespoň 60% není zařazena do žádné ze skupin..... | 52 |
| Tab. C.2: Tabulka vypočtených hodnot pro první scénář 2.část. Zkratka sk. znamená skupinu, do které byla daná patologie zařazena. Výskyt potom procentuálně vyjadřuje četnost zařazení do dané skupiny. Jestliže patologie nedosáhla alespoň 60% není zařazena do žádné ze skupin..... | 53 |
| Tab. C.3: Tabulka vypočtených hodnot pro druhý scénář. Zkratka sk. znamená skupinu, do které byla daná patologie zařazena. Výskyt potom procentuálně vyjadřuje četnost zařazení do dané skupiny. Jestliže patologie nedosáhla alespoň 60% není zařazena do žádné ze skupin. | 54 |
| Tab. C.4: Zařazení patologií do skupin podle podobnosti vypočtených výsledků pro první scénář. Není zde uvedeno rozdělení ve skupině hodnocení variability základního tónu. V této skupině byly všechny patologie zařazeny do stejné skupiny viz Tab. B.1a Tab. B.2. | 55 |

ÚVOD

Patologie¹ řeči neboli choroba řeči je název pro určitou poruchu (dysfunkci) řečové produkce v důsledku nesprávného fungování řečového ústrojí. Existuje mnoho podob těchto tzv. patologií. Tyto patologie můžou mít různý původ, jako je např. výskyt určitého neurologického onemocnění, dále mohou být následkem úrazu, nemoci jiného než neurologického původu, vrozené vady či jednoduše stářím konkrétního řečníka. V důsledku přítomnosti poruchy řečové produkce je těmto řečníkům špatně rozumět, jejich řečový projev se stává méně srozumitelným, mohou koktat nebo mluvit hlubším či jemnějším hlasem. Častým projevem poškození řečového aparátu je přítomnost nadměrného množství šumu v hlase, špatné dýchání, zvýšená nazalita řeči, špatná artikulace atd. Tyto klinické projevy je možné analyzovat, tj. pro řeč každého člověka je možné vypočítat mnoho parametrů kvantifikujících kvalitu řečového projevu, neboli ohodnotit do jaké míry je jejich řečový projev normální (bez poškození). V této práci budou tyto řečové parametry počítány a dále zpracovávány za úkolem tzv. kvantifikace dysfunkce řečového aparátu. Cílem práce je následná analýza řečových parametrů popisujících řečový projev z hlediska funkčnosti hlasivek, artikulačních orgánů a celkové kvality řečového signálu. Ve výstupu je analýza směřována k návrhu nového dělení patologií řeči na základě klinických projevů s využitím algoritmů tzv. shlukové analýzy.

V této práci je tedy snaha o rozřazení různých patologií do skupin, které spolu souvisí ne na základě původu ale na základě jejich působení na konkrétní části řečového ústrojí. Analytický náhled právě na tento aspekt řečové degradace by v budoucnosti mohl být vodítkem při tzv. diferenční analýze onemocnění vedoucích k degradaci řečové promluvy. Tato práce je prvotní studií, která dává do souvislosti poškození řečové produkce různého původu čistě z hlediska jejich manifestace na řečovém projevu. Pro tyto účely je použita databáze řečových záznamů MEEI (Massachusetts Eye and Ear Infirmary). Každému řečníkovi v databázi je v práci vypočítáno několik parametrů. Je zde skupina parametrů hodnotících variabilitu základního hlasového tónu, skupina parametrů popisujících pohyblivost artikulačních orgánů neboli schopnost artikulace a také skupina parametrů hodnotících úroveň šumu v hlase a tím nepřímo samotnou kvalitu řeči. Dalším krokem je tzv. předběžná analýza. Jejím úkolem je najít reprezentanty těchto skupin parametrů, které nejlépe korelují s konkrétním klinickým stavem mluvčího (patologická řeč/zdravá řeč). Tyto parametry jsou dále uvažovány jako vhodné kandidáty pro následné zpracování za účelem rozdělení patologií obsažených v databázi do několika skupin pomocí shlukové analýzy. Takto vypočtené skupiny patologií by podle předpokladů učiněných v této práci měli mít společné vlastnosti (rysy) v rámci poškození řečové produkce. V práci je také učiněná analýza a následné určení

¹patologie se v práci používá ve smyslu nezdravý, porušený

skupin patologií na základě kombinace nejlepších zástupců všech zvolených skupin parametrů společně. Výsledkem této analýzy je skupina patologií, které v přibližně stejném měřítku poškozují kvalitu řeči po stránce jak artikulační, fonační a také samotné kvality hlasu.

V první kapitole bakalářské práce je stručně popsána tvorba řeči. Jsou zde popsány jednotlivé ústrojí, která se na tvorbě řeči podílejí a také matematický model generování řeči, který zjednodušeně popisuje tvorbu řeči v lidském těle a je také základem značné míry tzv. lineárních řečových parametrů. V druhé kapitole jsou zmíněny patologie řeči a také oblasti řečového ústrojí, na kterých se jednotlivé patologie mohou projevit. Ve třetí kapitole jsou popsány parametry, které mohou hodnotit jednotlivé patologie řečového projevu. Jsou zde vybrané parametry, na kterých by mohly být dobře patrné rozdíly určitého poškození řečové produkce. Čtvrtá kapitola pojednává o metodě rozřazení řečových signálů do skupin podle již vypočtených parametrů ze třetí kapitoly. V páté kapitole je navrženo blokové schéma navrženého systému analýzy řečové promluvy a je zde popsáno, co jednotlivé bloky vykonávají. V šesté kapitole se nachází samotná realizace systému. Je zde popsána databáze MEEI (Massachusetts Eye and Ear Infirmary), z které jsou použity hlasy řečníků k této práci. Jsou zde také popsány jednotlivé skripty a soubory realizované v prostředí MATLAB. Tato kapitola také poskytuje přehled vypočtených dělení.

1 ÚVOD DO ZPRACOVÁNÍ ŘEČOVÉHO SIGNÁLU

Řeč je nejstarší prostředek, který slouží ke komunikaci. Je to nejpřirozenější a nejčastěji používaná forma komunikace. Tuto schopnost dorozumívat se pomocí promyšlených slov a frází mají pouze lidské bytosti. Pro člověka není většinou tato činnost složitá. Člověk je schopen si své myšlenky v krátké době promyslet a poté je sdělit jednomu či více posluchačům. Pro posluchače zase není nijak složité tyto informace poslouchat a zároveň je vnímat. Když se tyto role mluvčího a posluchače, ať už dvou či více lidí, střídají, jedná se o rozhovor. Výhodou této formy komunikace je, že člověk se může při rozhovoru věnovat i jiným činnostem např. chůzi, řízení atd. Tuto schopnost vytvářet a poslouchat řeč si lidé osvojí většinou už jako děti [6].

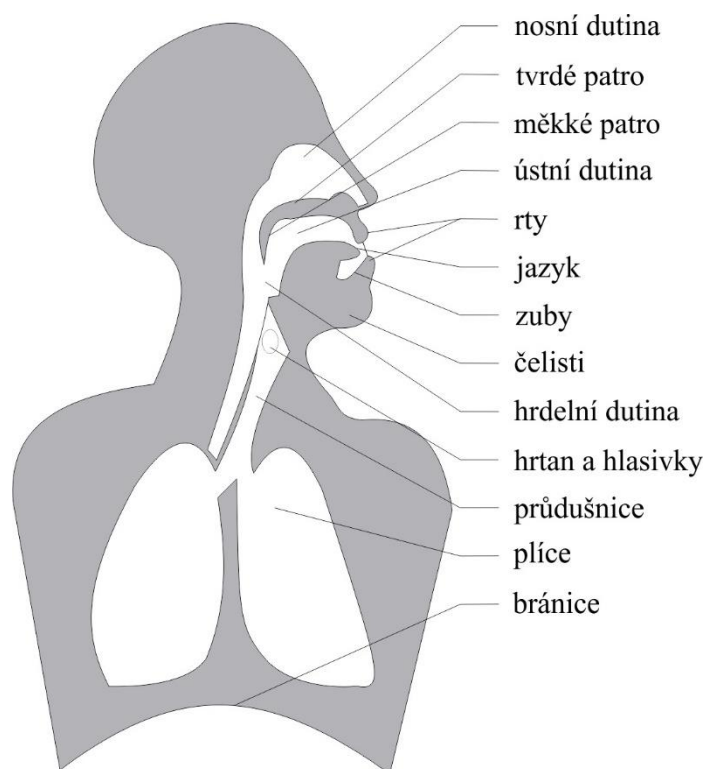
1.1 Tvorba řeči

V této práci se vyskytují dva pojmy a těmi jsou řeč a hlas. Tyto dva pojmy se často zaměňují. Titze definuje pojem řeč a hlas tak, že v širším smyslu je hlas synonymem pro řeč a v užším smyslu jsou to dva rozdílné pojmy. Hlas definuje jako vibraci hlasivek a řeč jako takovou definuje jako interakci proudění vzduchu z dechového ústrojí upravenou hlasivkami a do konečné podoby dotvořenou ostatními řečovými ústojími, jako jsou např. artikulační orgány neboli mluvidla. Na základě interakce proudění vzduchu z dechového ústrojí a hlasivkových svalů je možné řeč definovat jako znělou a neznělou. Znělá řeč vzniká vibracemi hlasivkových štěrbin. Touto vibrací hlasivky vytváří kvázy-periodické impulzy, které jsou základem znělé řeči. Neznělá řeč vzniká turbulentním prouděním vzduchu skrze hlasivky, které vytváří základ řeči neznělé. Pro účely této práce budou pojmy hlas a řeč považované za synonyma [10].

Řeč je akustický signál šířící se elastickým prostředím v lidském uchem slyšitelných frekvencích. Na tvorbě řeči se v lidském těle podílejí tzv. řečové orgány. Konkrétní anatomie a funkce jednotlivých orgánů jsou popsány v kapitole 1.2.

1.2 Řečové ústrojí

Řečové ústrojí neboli hlasový trakt člověka je skupina orgánů, jejichž primární úkoly jsou v lidském těle různé. Jedná se například o dýchání a přijímání potravy. Společně se však tyto orgány účastní také na procesu utváření řeči. Toto ústrojí se skládá ze tří základních částí. Těmito částmi jsou dechové, hlasové a artikulační ústrojí, které jsou popsány níže v této kapitole. Popis jednotlivých orgánů hlasového traktu ukazuje Obr. 1.1



Obr. 1.1: Hlasový trakt člověka.

1.2.1 Dechové ústrojí

Dechové ústrojí vytváří zdroj energie pro řeč. Je umístěno v hrudním koši. Je tvořeno plicemi a přívodní dýchací cestou. Při nádechu dochází k pohybu vzduchu směrem do plic. Při výdechu tato nahromaděná energie slouží jako základní materiál pro tvorbu řeči. Výdechový proud vzduchu je odváděn průdušnicí přes hlasové a artikulační ústrojí, kde dochází k jeho další modifikaci [6].

1.2.2 Hlasové ústrojí

Je to část řečového ústrojí, ve kterém dochází k samotnému vytváření hlasu. Je uloženo v hrtanu, který je s plicemi spojen průdušnicí. V hrtanové dutině se nacházejí hlasivky, které se podílejí na samotném vytváření hlasu. V hlasivkách se nachází hlasivková štěrbinu. Jestliže člověk mlčí, hlasivky drží hlasivkovou štěrbinu otevřenou a může jí procházet vzduch potřebný k dýchání. Toto se nazývá klidové postavení hlasivek. Při tvorbě hlasu se hlasivky uzavřou a jsou v tzv. hlasovém (fonačním) postavení. Při fonačním postavení se proud vzduchu postaví do cesty překážka vytvořená hmotou hlasivek, tím vznikají periodické kmity, které vytvářejí zvukové pulsy, označující se jako základní (hlasivkový) tón. Frekvence jeho kmitání se označuje jako F_0 a nazývá se fundamentální frekvence nebo taky frekvence základního hlasivkového tónu. Frekvence F_0 je u každého člověka různá. Muži mají v průměru 132 Hz, ženy 223 Hz a u dětí se tato hodnota pohybuje v rozmezí 200 až 600 Hz [6].

Pro tvorbu řeči se využívá jak klidové tak fonační postavení hlasivek. Fonační postavení se využívá pro tvorbu znělých zvuků. To jsou samohlásky a znělé souhlásky. Naopak v klidovém postavení hlasivek se vytvářejí neznělé zvuky. Tyto zvuky nemají základní hlasivkový tón a vznikají až modifikací v nadhrtanových dutinách [6].

1.2.3 Artikulační ústrojí

Artikulační ústrojí se skládá z nadhrtanových dutin a artikulačních orgánů. Artikulační orgány (artikulátory) tvoří pohyblivou tzv. aktivní část, která se na tvorbě řeči podílí svým pohybem. Patří mezi ně hlavně jazyk, rty a měkké patro. Dalšími artikulátory jsou zuby, čelisti, tvrdé patro a hrtan. Do nadhrtanových dutin patří dutiny hrdelní, ústní a nosní. Tyto dutiny se na tvorbě řeči podílejí pasivně [6].

Když základní tón prochází přes nadhrtanové dutiny dochází vlivem rezonance k soustředění akustické energie kolem určitých frekvencí. Tyto oblasti kde se energie koncentruje se nazývají formáty. Značí se jako F_1, \dots, F_n . Pokud se do vytváření řeči zapojí i dutina nosní, dochází k potlačení některých tzv. antiformančních frekvencí [6].

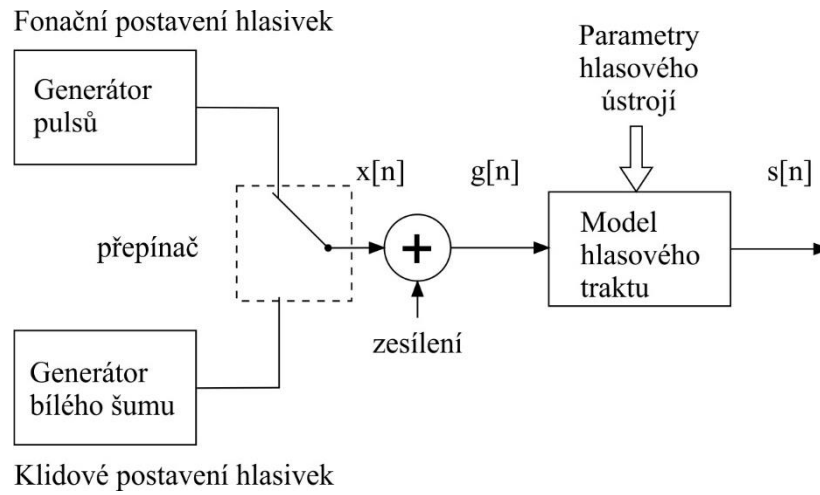
Na výstupu artikulačního ústrojí může vznikat řeč znělá či neznělá, více o znělosti řeči, viz kapitola 1.2.2. Při znělé řeči je kvázy-periodické kmitání hlasivkových svalů, které produkují hlasové pulzy upravené konkrétním postavením artikulačních orgánů. Při neznělé řeči je šumový signál produkovaný hlasivkami dotvořený do výsledné podoby neznělé řeči pomocí překážek tvořených artikulátory, jedná se např. o tzv. plozivý, explozivý (p, t, k, b) [6].

1.3 Matematický model generování řeči

Model generování řeči je matematický model, popisující a zjednodušující tvorbu řeči v lidském těle. Tento model se skládá z několika bloků, které aproximují práci řečového ústrojí. Zjednodušený model generování řeči je zobrazen na obrázku Obr. 1.2. Tento model se stal základem mnohých metod zpracování řečového signálu a skládá se z následujících bloků:

- **Generátor jednotkových impulzů** – tento blok simuluje práci dechového a hlasového ústrojí při tvorbě znělé řeči, kdy je vzduch vydechovaný plicemi upravený do podoby kvázy-periodických impulzů pomocí vibrování hlasivek a konkrétní frekvenci viz kapitola 3.2.1.
- **Generátor bílého šumu** – tento blok se používá na aproximaci funkce dechového a hlasového ústrojí při tvorbě neznělé řeči, kdy je vydechovaný proud vzduchu upravený do podoby turbulentního proudění na výstupu hlasivek, které jsou při neznělé řeči otevřené
- **Model hlasového traktu** – je to blok, ve kterém se signál na výstupu konkrétního generátoru upraví do podoby výsledné řeči. Tento blok simuluje práci artikulačního ústrojí (postavení artikulátorů, rezonanci v ústní dutině,

práci měkkého patra atd.). Je v praxi realizovaný číslicovým filtrem a je vytvářen pro každou hlásku.



Obr. 1.2: Matematický model generování řeči.

2 PATOLOGIE ŘEČOVÉHO PROJEVU

2.1 Patologická řeč

Patologie řeči může být zkoumána z hlediska jak komunikačního projevu, tak z hlediska funkce hlasu. V této práci je zkoumána jako porucha funkce hlasu. Porucha, u níž určité části řečového ústrojí nepracují správně, je označována jako patologie řeči. Tato porucha může mít různý původ a také různé klinické příznaky, které se projeví degradací kvality řečového projevu postižené osoby.

Titze klasifikoval patologie řeči z hlediska poruch do tříd: zánětlivá onemocnění (laryngitida, bronchitida, záškrt, atd.), systematické změny (dehydratace, vedlejší účinky medikace, hormonální změny, atd.), mechanické namáhání (polypy, vředy, granulomy, laryngocele, krvácení, atd.), povrchová podráždění (laryngitida, leukoplakie, gastroezofageální reflux, atd.), změny tkáně (rakovina hrtanu, keratóza, papilomy, cysty, atd.), neurologické a svalové změny (bilaterální a unilaterální paralýza hlasivek, spastická dysfonie, Parkinsonova nemoc, amyotrofická laterální skleróza, myotonická dystrofie, Huntingtonova choroba, myasthenia gravis, atd.) a abnormality svalstva (dysfonie, ventrikulární fonace, atd.) [10].

2.2 Patologie řečového projevu a řečového ústrojí

Jednotlivé patologie řečového projevu se projevují špatnou funkcí řečového ústrojí nebo jeho částí. Konkrétně se řečové patologie často projevují poškozením fonace, to znamená, že je napadnuté hlasové ústrojí, tj. hlasivky a poškozením artikulace, kdy se patologie projevuje nesprávnou funkcí artikulačního ústrojí jako například jazyka, rtů, atd. Cílem této práce je výzkum podobností klinických příznaků jednotlivých patologií různého druhu a původu na degradaci řečového projevu. Pomocí tzv. paraklinických metod, využívajících neinvazivní metody zkoumání projevů různých patologií, jako je např. Parkinsonova nemoc, s využitím řečového signálu bude v práci snaha o vytvoření nového dělení patologií pomocí metod tzv. shlukové analýzy.

V následujících podkapitolách jsou stručně zmíněné oblasti tvorby a percepce řeči a klinické příznaky, které jsou patrné při určité patologii řeči na konkrétní část řečového ústrojí, která je patologií napadená.

2.2.1 Fonace

Fonací můžeme popsat tvorbu hlasu v hlasovém ústrojí. Konkrétně se pod pojmem fonace ve vědecké sféře rozumí kmitání hlasivek, tedy tvorba hlasu jako takového. Úpravou signálu produkovaného hlasovým ústrojím vzniká řeč. V případě, že je toto hlasové, respektive fonační, ústrojí poškozené, například působením určité choroby nebo

zraněním, nepracuje toto ústrojí správně a ve výsledku je možné tuto řeč považovat za patologickou. Řečník s určitým poškozením fonace je označován jako tzv. *dysfonický* a naopak řečník, kterému hlasivky kmitají správně se označuje jako *normofonický*.

Hlasivky v důsledku působení určité patologie nejsou například schopné pracovat v úplném rozsahu, nebo se úplně uzavřít, což může způsobit výskyt šumu v hlase. Svalstvo hlasivek bývá ztuhlé a proudění vzduchu z plic výrazně kolísá a to má za následek třesot hlasu. V tzv. patologické řeči, kdy tato patologie napadá fonační ústrojí, je často patrný pokles intenzity řeči. Takto napadnutá řeč se označuje jako *hypofonická*.

2.2.2 Artikulace

Tato porucha je dána nesprávnou funkcí artikulačního svalstva a projevuje se nesprávnou funkcí aktivních artikulačních orgánů. Mezi ně patří např. jazyk, rty a měkké patro. Lidé s degradovanou funkčností artikulačních orgánů nejsou schopni v plném rozsahu využít artikulační orgány k dotvoření řečového signálu do podoby srozumitelné posluchači. Mezi často sledované parametry patří např. sledování pohyblivosti jazyka, tzv. nazility řečového projevu. Některé patologie jako například Parkinsonova choroba způsobují tzv. *hypernazalitu*, která je způsobena nesprávnou funkcí měkkého patra. Tyto parametry se často sledují na cvičeních, které vyžadují vyšší míru artikulace. Jsou to tzv. *Diadocho-kynetické úlohy* (DDK).

2.2.3 Prozodie

Pacienti, kteří trpí určitými patologiemi řečového projevu, jako je např. Parkinsonova nemoc, také často nesprávně nakládají s melodií řeči. Intonace řečového projevu je poškozená a takto poškozený řečový projev je nazýván jako *dysprozodický*. Tyto vady jsou detekovatelné parametry, které sledují tzv. prozodické vlastnosti řeči. Jedná se o tzv. suprasegmentální² parametry.

Prozodie je částečně odvozená od fonace³. Kromě základních prozodických parametrů jakými jsou frekvence základního tónu a intenzita řeči patří mezi prozodické jevy například práce s dechem. Prozodie řeči je velmi častým nástrojem pro systémy diagnostiky patologické řeči, a proto je také předpoklad, že parametry sledující prozodii se v rámci této práce projeví jako signifikantní diskriminátor patologické a zdravé řeči.

2.2.4 Plynulost řeči

Pacienti postižení určitou řečovou patologií mohou neúmyslně vytvářet pauzy

² Suprasegmentální značí, že popisují jevy nad úrovní jednotlivých segmentů

³ Jednotlivé parametry častokrát spadají do více než jedné kategorie. Např. parametry sledující chování hlasivek (základní tón, atd.) se v rámci této práce dají zařadit mezi fonační i prozodické parametry. Ve všeobecnosti je toto dělení spíše orientační.

mezi slovy nebo opakovat jednotlivé hlásky, slabiky či celá slova. Tato vada se nazývá hezitace promluvy. Některé patologie, jako je např. Parkinsonova nemoc se projevují tím, že pacienti během řeči zrychleně opakují slova či jednotlivé slabiky (*palilalie*). Některé patologie dále způsobují to, že pacienti během promluvy náhle zpomalují (*bradyfemie*) nebo zrychlují řeč (*tachyfemie*).

3 PARAMETRY HODNOTÍCÍ PATOLOGIE ŘEČOVÉHO PROJEVU

V následující kapitole je popsáno zpracování řečového signálu. Pro účely diagnostiky jednotlivých patologií řečového projevu je potřebné řečový signál tzv. parametrizovat. Ve všeobecnosti to znamená, že z průběhů záznamů řečového projevu jsou vypočtené jisté hodnoty, které nazýváme parametry. Tyto parametry by měly co nejlépe popisovat vlastnosti řečového signálu, které jimi sledujeme. Při diagnostice patologické řeči se často pracuje s parametry hodnotící postižené části řečového aparátu, jako jsou např. parametry sledující práci fonačního ústrojí, parametry, které se zaměřují na anomálie artikulačního ústrojí, parametry sledující intonaci řeči a další. Předpokladem pro takovéto zpracování řeči je, že jednotlivé patologie poškozují řečové orgány a ty následně vykazují nesprávnou funkci, kterou se snažíme zachytit.

V následujících podkapitolách jsou stručně zmíněné parametry, které jsou vhodnými kandidáty pro zařazení do navrhovaného systému klasifikace řečových patologií. Dále jsou tu zmíněné i metody tzv. předzpracování řečového signálu, které jsou důležitou částí systému a jsou ve všeobecnosti využívány k úpravě řečového signálu do podoby, která je parametrizovatelná. V poslední podkapitole je zmíněný výpočet tzv. vysokoúrovňových parametrů.

3.1 Předzpracování řečového projevu

Lidská řeč a s ní spjatý řečový signál jsou velmi rozmanité. Mezi jednotlivými řečníky jsou velké rozdíly i ve výslovnosti stejného slova. Nikdo také není schopný vyslovit jedno slovo vždy naprosto stejně. Na řečový signál mají negativní vliv různá okolní rušení, okolní zvuky, ale také zkreslení při zpracování signálu. Proto je vhodné tento vstupní řečový signál před samotným zpracováním ještě upravit [7], [10].

3.1.1 Preemfáze

Preemfáze se v číslicovém zpracování řečového signálu používá na kompenzaci působení kmitočtové charakteristiky prostředí na daný řečový signál, která potlačí vyšší kmitočtové složky spektra řečového signálu (cca. od 2 kHz a vyšší), které jsou pro srozumitelnost řečového signálu významné. Jedná se tedy o zdůraznění vyšších frekvenčních složek řečového signálu za účelem vyrovnání kmitočtového spektra řeči. Tyto pro srozumitelnost nepříznivé vlivy prostředí na vstupní řečový signál se omezují použitím preemfázového číslicového filtru, který je realizován filtrem typu horní propust (HP) [2]. Přenosová funkce preemfázového filtru má tvar:

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

kde a_1 je tzv. koeficient filtru a běžně nabývá hodnot 0,9 až 1.

3.1.2 Ustředění

Řečový signál často obsahuje tzv. stejnosměrnou složku. Pro účely zpracování řečového signálu za účelem jeho parametrizace je však tato složka nepodstatná, v některých případech až nežádoucí, kdy její přítomnost může dokonce způsobit určení chybné hodnoty parametru. Odstranění stejnosměrné složky neboli tzv. ustředění je tedy pro účely této práce vhodné. V praxi je ustředění často vykonávané jednoduchým odečtením stejnosměrné složky (střední hodnoty) od signálu [2]. Pokud celý signál známe v čase výpočtu, co bude i případ této práce, můžeme použít zjednodušený výpočet off-line:

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

kde $s'^{[n]}$ je výsledný signál s odečtenou stejnosměrnou složkou, N je počet vzorků původního filtrovaného signálu a $s[n]$ jsou jednotlivé vzorky původního signálu.

3.1.3 Segmentace

Je rozdělení řečového signálu před dalším zpracováním na menší úseky tzv. rámce. Řečový signál jako takový není stacionární a dá se považovat za náhodný. Proto se signál rozděluje na již zmíněné rámce, na kterých už se dá považovat za kvazy-stacionární a je možné ho dále zpracovávat metodami, které uvažují deterministický průběh vstupního signálu. Délka rámce se volí přibližně 20-25 ms (160-200 vzorků pro vzorkovací kmitočet 8000 Hz), aby byla dostatečně malá a průběh signálu jednoho segmentu bylo možné považovat za stacionární a zároveň dostatečně velká, aby bylo možné přesně určit požadované parametry [2].

V praxi se často používá tzv. segmentace s překryvem, tedy segmentace, kdy se sousední rámce z části překrývají (běžně se při segmentaci používá překrytí sousedních segmentů 50%). Čím menší bude překrytí sousedních segmentů, tím větší bude rychlost posunu signálu v čase a tím menší nároky na paměť. Velké překrytí zase zabezpečí hladký průběh parametrů. Výsledné parametry mohou být však rámec od rámce podobné a to není dobré z hlediska detekce daných parametrů a jejich rozpoznávání [2]. Počet rámců s překrytím je možné vypočítat vztahem:

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

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

Segmentace neboli dělení signálu na rámce se provádí tzv. okénkovou funkcí. Při zpracování řeči se nejvíce využívá pravoúhlé, nebo Hammingovo okno [8]. Pro výpočet pravoúhleho okna platí vztah [8]:

$$w[n] = 1, \quad \text{pro } n = 0, 1, \dots, N - 1 \quad (3.4)$$

$$w[n] = 0, \quad \text{pro ostatní } n$$

Ze vztahů pro výpočet funkce pravouhlého okna je možné vidět, že pravouhlé okno zanechává na okrajích segmentů ostré přechody, které mohou v spektrální oblasti způsobovat tzv. „prosakování“ spektrálních složek do sousedních segmentů. Proto se v praxi častěji používá Hammingova okno, které průběh segmentovaného signálu v čase na jeho okrajích potlačuje a tím zabraňuje vzniku spektrálních nehomogenit. Pro výpočet funkce Hammingova okna platí vztah [8]:

$$w[n] = 0,54 - 0,46 \cos \left[n \frac{2\pi}{N} \right], \quad \text{pro } n = 0, 1, \dots, N - 1 \quad (3.5)$$

$$w[n] = 0, \quad \text{pro ostatní } n$$

V bakalářské práci bylo použito hammingovo okno o šířce $0,02 * f_{vz}$ vzorků. Jinými slovy, bylo uvažováno okno o délce 20 milisekund. Takováto délka okna se v praxi používá nejčastěji a souvisí s maximální rychlostí změny polohy řečových orgánů. Při časovém okně 20 milisekund se uvažuje, že signál v segmentu bude dostatečně krátký aby se už dal považovat za kvazy-stacionární.

3.2 Parametry hodnotící fonaci

3.2.1 Zakladní frekvence

Neboli frekvence základního hlasivkového tónu F_0 . Odpovídá výšce hlasu, jakou vnímá posluchač. Ve spektrální oblasti základní frekvence, nebo tzv. „pitch“, odpovídá frekvenci první harmonické složky signálu. Je obsažena u hlásek, které jsou buzené kvazy-periodickým kmitáním hlasivkových svalů. Tento parametr je velmi často využíván při zpracování řečového signálu a při sledování intonace řeči, a jeho diferenciální formy, jako je např. „jitter“, velmi dobře sledují melodii řečového projevu.

V rámci teorie působení patologií řečového projevu je tento parametr a jeho formy velmi vhodným kandidátem na diferencii zdravé a patologické řeči. Tato frekvence je u dětí a dospělých různá a roli hraje také pohlaví mluvčího. Vztah pro výpočet základní frekvence [7]:

$$F_0 = \frac{f_{vz}}{k_{\min}}, \quad (3.6)$$

kde f_{vz} je vzorkovací kmitočet a k_{\min} je pozice (ve vzorcích) prvního maxima autokorelační funkce. Tento vztah popisuje pouze jednu možnost detekce základního hlasivkového tónu. Existuje řada postupů využívajících metody detekce v časové oblasti a ve spektrální oblasti. Více viz [8].

3.2.2 Jitter

Frekvenční nestabilita. Perioda jednotlivých pulzů základního hlasivkového tónu se může lišit i v rámci velmi krátkého úseku řečového signálu. Je to kolísání délky základní periody (převrácené hodnoty základní frekvence). Zjistíme ho jako rozdíl v délce dvou sousedních period hlasivkových pulzů dělený průměrnou délkou periody. Označuje se též jako třes nebo chvění hlasivek [6], [8]. Vztah pro výpočet jitteru [8]:

$$J_g = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |T_g[i] - T_g[i-1]|}{\frac{1}{N} \sum_{i=1}^{N-1} T_g[i]}, \quad (3.7)$$

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

3.2.3 Shimmer

Amplitudová nestabilita. Jak již bylo zmíněno v podkapitole 3.2.2 ohledně základní periody hlasového tónu, tak i amplituda hlasivkových pulzů se může lišit i v rámci velmi krátkého úseku řečového signálu. Je to tedy kolísání amplitudy hlasivkových impulzů. Zjistíme ho jako rozdíl velikosti dvou sousedních pulzů dělený průměrnou velikostí těchto pulzů. Označuje se také jako vibrace nebo kolísání pulzů. Vztah pro výpočet shimmeru je definován jako [8]:

$$S_g = \frac{\frac{1}{N-1} \sum_{i=1}^{N-1} |A_g[i] - A_g[i-1]|}{\frac{1}{N} \sum_{i=1}^{N-1} A_g[i]}, \quad (3.8)$$

kde N je počet amplitud hlasivkových pulzů a $A_g[i]$ je i -tá hodnota amplitudy.

3.3 Parametry hodnotící energii řečového signálu

3.3.1 STE (Short-time energy)

Krátkodobá energie je parametr popisující energii segmentované řeči, tj. tímto parametrem se vyjádří energie řečového signálu v jednotlivých segmentech. Krátkodobá energie slouží například k rozdělení znělých a neznělých hlásek, kdy z teorie tvorby řeči je známo, že znělé hlásky obsahují více energie. Naopak neznělé hlásky, které se často nazývají šumové hlásky, obsahují mohem méně energie. Tento parametr se dá také použít k rozpoznání přechodu mezi znělým a neznělým úsekem signálu, čehož je využíváno v jednoduchých rozpoznávacích znělosti segmentu, který může být použit např. při tzv. detektoru řečové aktivity [7]. Vztah pro výpočet krátkodobé energie jednoho segmentu řečového signálu je počítán jako [7]:

$$STE = \sum_{k=0}^{N-1} s[k]^2, \quad (3.9)$$

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

řečový signál tedy STE představuje vektor energií pro jednotlivé segmenty.

3.3.2 LSTER (Low short-time energy ratio)

Je poměr počtu segmentů, ve kterých je krátkodobá energie nižší jako 0,5 násobek průměrné krátkodobé energie v sekundovém úseku segmentovaného signálu. Tento parametr se používá pro diskriminaci mezi hudebním a řečovým signálem, při předpokladu, že v řečovém signálu se vyskytuje více segmentů obsahujících tiché části promluvy, (tj. pauzy), než je tichých segmentů v hudebním signálu. Z tohoto předpokladu vyplývá, že parametr LSTER vykazuje vyšších hodnot právě pro řečový signál [1]. Vztah pro výpočet parametru LSTER [1]:

$$LSTER = \frac{1}{2N} \sum_{n=0}^{N-1} [\text{sgn}(0,5\bar{E} - E(n)) + 1], \quad (3.10)$$

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

3.3.3 TKEO (Teager-Kaiser energy operator)

Je nelineární operátor, který může poskytnout odhad okamžité frekvence a amplitudy z FM a AM modulace. Na rozdíl od STE popsaného v podkapitole 3.3.1 je zde energie počítána pro celý řečový signál. Pro diskrétní čas je tedy TKEO definován jako:

$$\psi[s(n)] = s(n)^2 - s(n+1)s(n-1), \quad (3.11)$$

kde $s(n)$ je vstupní řečový signál.

3.3.4 MPSD (Median of power spectral density)

Medián výkonové spektrální hustoty, je parametr určující výkon signálu v určitém frekvenčním pásmu. Odhad výkonové spektrální hustoty může být vypočtený pomocí více metod. Jednou z metod je použití tzv. periodogramu.

V praxi je signál nejprve rozdělen na segmenty. Tyto segmenty jsou následně podrobené analýze výkonové spektrální hustoty (PSD). Tento postup odpovídá tzv. neparametrickým metodám výpočtu PSD. V této práci bude použita Welchova metoda průměrování modifikovaných periodogramů. Při této metodě se používá segmentace s překrýváním jednotlivých segmentů. Analyzovaná data jsou tedy rozdělena podle [9]:

$$x_i[n] = x[n + iD], \quad (3.12)$$

kde hodnota iD je počátečním bodem i -té posloupnosti a $i = 0, 1, \dots, L - 1$. L je počet po sobě jdoucích segmentů. V praxi se nejčasteji používá přesahu 50%. Modifikovaný periodogram podle Welchovi metody pak vypadá takto [9]:

$$\tilde{P}_{xx}^{(i)}(e^{j2\pi f}) = \frac{1}{MU} \left| \sum_{n=0}^{M-1} x_i[n] w[n] e^{j2\pi f n} \right|^2, \quad (3.13)$$

kde $i = 0, 1, \dots, L - 1$ a U je normalizační koeficient [9]:

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2[n]. \quad (3.14)$$

Odhad výkonové spektrální hustoty počítané pomocí Welchovi metody je pak průměr modifikovaných periodogramů daný vztahem [9]:

$$P_{xx}^w(e^{j2\pi f}) = \frac{1}{L} \sum_{i=0}^{L-1} \tilde{P}_{xx}^{(i)}(e^{j2\pi f}). \quad (3.15)$$

V práci bude výkonová spektrální hustota použita pro výpočet jejího mediánu. Tato metoda je použita ke zjištění středního výkonu v určitém pásmu zkoumaného průběhu. Výpočet mediánu PSD je:

$$\text{med}\left(P_{xx}^w(e^{j2\pi f})\right) = P_{xx}^w(e^{j2\pi f}) \left[\frac{n+1}{2}\right], \quad (3.16)$$

jestliže počet PSD bude lichý bude platit:

$$\text{med}\left(P_{xx}^w(e^{j2\pi f})\right) = \frac{1}{2} \left(P_{xx}^w(e^{j2\pi f}) \left[\frac{n}{2}\right] + P_{xx}^w(e^{j2\pi f}) \left[\frac{n}{2} + 1\right] \right). \quad (3.17)$$

3.4 Parametry hodnotící kvalitu řeči

3.4.1 ZCR (Zero-crossing rate)

Je parametr, který popisuje počet průchodu řečového signálu nulovou úrovní. Zjednodušeně si to lze představit tak, že přechod nulovou úrovní nastane, když dva sousední vzorky zkoumaného signálu mají různé znaménko. Tento parametr navíc dává i omezenou informaci o kmitočtu signálu. Jeho velikou výhodou je nezávislost na energii signálu. Vztah pro výpočet počtu průchodu nulou v jednom segmentu řečového signálu je vyjádřen jako [7]:

$$Z_{cr} = \sum_{k=0}^{N-1} |\text{sign}(s[k]) - \text{sign}(s[k-1])| / 2, \quad (3.18)$$

$$\text{sign}(s[k]) = \begin{cases} +1 & \text{ak } s[k] > 0 \text{ nebo } s[k] = 0 \wedge s[k-1] > 0, \\ -1 & \text{ak } s[k] < 0 \text{ nebo } s[k] = 0 \wedge s[k-1] < 0. \end{cases} \quad (3.19)$$

kde $s[k]$ jsou vzorky segmentovaného signálu, k představuje pořadí vzorků v segmentu a N je počet vzorků v jednom segmentu.

3.4.2 HZCRR (High zero-crossing rate ratio)

Je to poměr počtu segmentů, ve kterých je průchod nulou (ZCR) větší jako 1,5 násobek průměrného ZCR v jedné sekundě daného signálu. Tento parametr se také používá pro diskriminaci mezi hudebním a řečovým signálem, při předpokladu, že řečový signál se skládá z kombinace znělých a neznělých úseků, které se však v hudebním signálu nevyskytují. Vztah pro výpočet HZCRR [4]:

$$HZCRR = \frac{1}{2N} \sum_{n=0}^{N-1} [\text{sgn}(ZRC(n) - 1,5\overline{ZCR}) + 1], \quad (3.20)$$

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

3.4.3 HNR (harmonic-to-noise ratio)

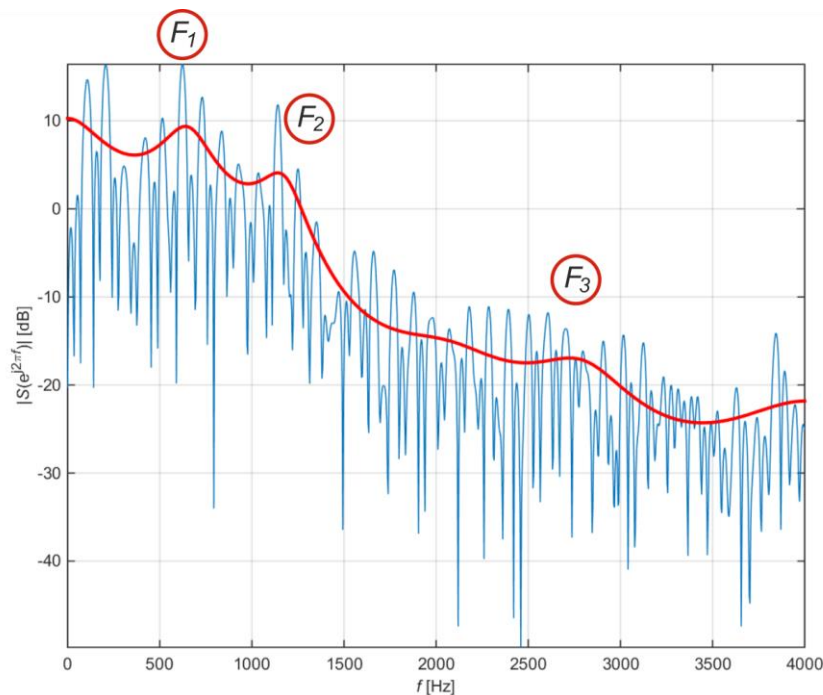
Odstup harmonické od šumu. Tento parametr porovnává úroveň energie harmonického signálu a úroveň energie šumu. Při práci s řečovými signály se tento parametr používá k detekci řečové aktivity. Tento parametr vyjadřuje stupeň periodicity zkoumaného průběhu a je počítán jako.

$$HNR = 10 \log_{10} \left(\frac{\sum_{i=1}^{N/2} |S_i|^2}{\sum_{i=1}^{N/2} |N_i|^2} \right), \quad (3.21)$$

kde $|S_i|$ je energie harmonického signálu a $|N_i|$ je odhad energie šumu.

3.5 Parametry hodnotící pohyblivost artikulačních orgánů

Parametry vyjadřující schopnost artikulace jsou především formanty. Formanty jsou frekvenční oblasti s vyšší koncentrací akustické energie vznikající v důsledku průchodu hlasivkového tónu zejména dutinou hrdelní a ustní a následnou rezonací v těchto dutinách. Lze je definovat jako „vrcholky“ frekvenčního spektra daného zvuku. Jsou závislé na konfiguraci řečových orgánů a na rozměrech hlasového traktu. Pro syntézu srozumitelné řeči je třeba modelovat alespoň tři formanty [6].



Obr. 3.1: Zobrazení prvního, druhého a třetího formantu pro samohlásku „a“.

Na Obr. 3.1 je zobrazena červeně lineární predikční analýza (LPC) pro samohlásku „a“. LPC je metoda zpracování řečového signálu. Základem této metody je předpoklad, že každý vzorek signálu je možné dopředu určit (predikovat) pomocí predikčního řádu p . Vztah pro výpočet predikčního řádu [9]:

$$p = \frac{f_{vz}}{1000} + 4, \quad (3.22)$$

kde f_{vz} je vzorkovací frekvence. V této práci se počítá se vzorkovací frekvencí 50, 25 nebo 10 kHz podle jednotlivého řečníka.

Pomocí predikčních koeficientů je vytvořen filtr, který má na vstup přiveden vstupní signál a na výstupu je bílý šum. Invertováním tohoto procesu vznikne tzv. syntezující filtr sloužící pro vytvoření náhodného procesu s charakterem bílého šumu. Vztah pro výpočet tohoto filtru [9]:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + [\sum_{i=1}^p a_p[i]z^{-i}z^{-i}]}. \quad (3.23)$$

3.6 Parametry hodnotící tempo řeči

Parametry, které vyjadřují tempo řeči jsou založeny na předpokladech působení patologií řeči na jednotlivá řečová ústrojí, více viz kapitolu 2.2. Tempo řeči se dá vyjádřit různými parametry. Těmito parametry jsou například TPT (Total Pause Time), které určuje celkové trvání pauz. Dobu trvání čistě řečového signálu udává NST (Net Speech Time). TST (Total Speech Time) je pak doba trvání celého řečového signálu včetně pauz a tichých míst. Parametr určující absolutní tempo řeči je TSR (Total Speech Rate) a tempo pouze čistého řečového signálu udává NSRA (Net Speech Rate).

3.7 Vysokourovňové parametry

Některé řečové parametry jsou počítány pro každý segment řečového signálu zvlášť. Tím se vytvoří matice popřípadě vektor hodnot příslušející jednomu parametru pro daná řečová cvičení. Aby bylo možné s těmito hodnotami dále pracovat je třeba transformovat je na skalární hodnotu. K tomu se používá metoda aplikace statistiky. Statistické hodnoty, které jsou vypočteny z řečových parametrů, se označují jako vysokourovňové parametry. Nejčastějšími statistickými hodnotami používanými pro tento účel jsou například střední hodnota, medián, rozptyl atd.

4 SHLUKOVÁ ANALÝZA

Shluková analýza neboli *clustering* je vícerozměrná statistická analýza. Používá se ke klasifikaci objektů. Tato analýza třídí objekty do skupin tak, že objekty patřící do jedné skupiny jsou si podobnější než objekty z různých skupin. Shluková analýza je metoda, která ke klasifikaci vstupních objektů nepotřebuje žádné počáteční klasifikátory. Tato metoda bývá označována jako učení bez učitele [5].

4.1 Kohenova samoorganizační mapa

Je to metoda založená na umělých neuronových sítích. Cílem této metody je rozdělení vícerozměrných dat do dvourozměrného prostoru. Rozdělení má zajistit, aby podobné výsledky z vícerozměrného prostoru byly v mapě co nejbliže u sebe. Tato síť využívá metodu učení bez učitele. To znamená, že nepotřebuje žádné předpokládané výstupy a je schopna se sama přizpůsobit [3].

4.1.1 Učení neuronové sítě

Při procesu učení se neuronová síť snaží uspořádat neurony do určitých oblastí, aby byly schopny klasifikovat předložená vstupní data. Proces, jak už bylo řečeno, je realizován bez učitele tzv. autonomně. Proces probíhá iterativně. V každém kroku dochází k adaptaci vah, která je realizována tak, že porovnává vstupní vzory a vektory, jež jsou v každém neuronu Kohenové mapy. V případě nalezení neuronu, který nejlépe odpovídá vstupnímu neuronu, jsou jeho váhy a vektory upraveny. Toto platí i pro neurony v nejbližším okolí. Kohenova mapa je tedy optimalizována postupně, aby co nejlépe odpovídala vstupnímu prostoru dat. Všechny neuronové sítě, které jsou nejprve uspořádány náhodně, dostávají v průběhu učení konkrétní tvar, který je reprezentací vstupních dat [3]. Popis postupu učení ukazuje těchto následujících 6 fází:

1. **Inicializace** – v prvním kroku algoritmu je provedeno počáteční nastavení vah neuronové sítě w_{ij} , $0 \leq i \leq N - 1$, $0 \leq j \leq M - 1$ pro všechny spoje z N do M výstupních neuronů na malé počáteční hodnoty. Nastavení parametru učení η blíží se hodnotě 1.
2. **Předložení vzoru** – v tomto kroku jde o tzv. předložení nového vzoru na vstupy neuronové sítě $X(t) = \{x_0(t), x_1(t), \dots, x_{N-1}(t)\}$.
3. **Výpočet vzdáleností vzorů** – zde jsou vypočteny vzdálenosti mezi d_j mezi předloženým vzorem a výstupními neurony j podle vztahu:

$$d_j = \sum_{i=0}^{N-1} [x_i(t) - w_{ij}(t)]^2, \quad (4.1)$$

kde $x_i(t)$ jsou elementy vstupního vzoru $X(t)$ a $w_{ij}(t)$ jsou váhy mezi i -tým vstupem a j -tým neuronem.

4. **Výběr nejbližšího neuronu** – v tomto kroku algoritmus provádí výběr vstupního neuronu j^* , který nejlépe splňuje podmínku. Odpovídá tedy nepodobnějšímu neuronu:

$$d_{j^*} = \min(d_j) \quad (4.2)$$

5. **Přizpůsobení vah** – zde algoritmus Kohonenovi síti provádí přizpůsobení vah pro neuron j^* a jeho okolí podle následujícího vzoru

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)[x_i(t) - w_{ij}(t)], \quad (4.3)$$

kde parametr j označuje všechny neurony ležící v okolí, i jsou vstupy neuronové sítě a $\eta(t)$ označuje tzv. parametr učení.

6. **Ukončení procesu** – pokud nejsou vyčerpány všechny vzory, pokračuje algoritmus k bodu č.2, v opačném případě je síť naučená a algoritmus končí.

Navzdory svému rozsáhlému použití v mnoha aplikacích využívajících shlukové analýzy tento algoritmus do konečného řešení bakalářské práce integrován nebyl. Nahradila ho metoda tzv. „clusterování“ K-mean, která se v oblasti spracování řečového signálu používá velice často a dosahuje velice dobrých výsledků.

4.2 K-means

Úkolem této metody je rozdělit množinu vektorů n do k podmnožin a to tak, aby byla suma vzdáleností jednotlivých vektorů od středu příslušné podmnožiny minimalizovaná. Počet podmnožin k je u algoritmu k-means volen na začátku procesu. Jedná se o iterační algoritmus, který vytváří shluky vektorů a z nich dále vypočítává jejich geometrické středy, zvané centroidy. Centroid je charakteristický tím, že má od všech vektorů v daném shluku minimální vzdálenost. Pro účely bakalářské práce byla tato metoda zvolena jako referenční metoda pro shlukovou analýzu patologické řeči [2].

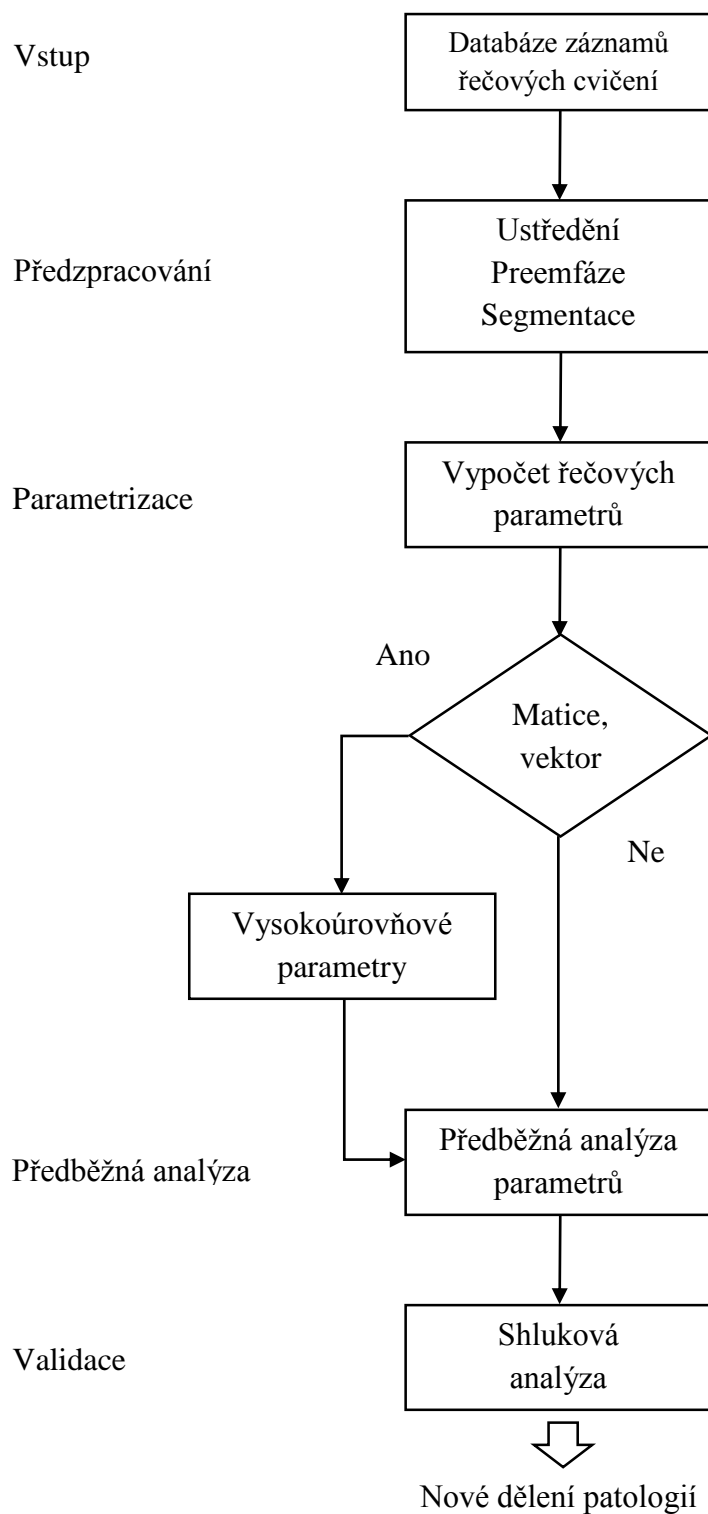
4.2.1 Popis algoritmu

Tento algoritmus je ideální při výběru do malého počtu shluků z velkého počtu vstupních informací. V tomto směru je to jeden z nejpoužívanějších algoritmů. Algoritmus pracuje v následujících krocích:

1. **Zvolení středů** – algoritmus k-means si na začátku běhu nejprve zvolí (náhodně) výchozí pozice jednotlivých středů tzv. centroidů.
2. **Přiřazení k nejbližšímu středu** – poté na základě vzdálenosti se každý ze vstupních vektorů přiřadí k příslušnému shluku, jehož centroid má nejmenší vzdálenost od uvažovaného vektoru.
3. **Výpočet středů** – následně u každé podmnožiny vypočítá nový centroid, který je geometrickým středem podmnožiny.
4. **Návrat ke kroku 2.** – návrat ke kroku 2.

5 NÁVRH SYSTÉMU

5.1 Blokové schéma



5.2 Popis systému

Navrhnutý systém klasifikace řečových patologií na základě analýzy řečových cvičení se skládá z několika bloků, které jsou vyobrazeny v kapitole 5.1. Funkce těchto bloků slouží k úpravám jednotlivých řečových cvičení, parametrizaci a následné klasifikaci pomocí algoritmů tzv. shlukové analýzy. Jedná se o tyto bloky:

- **Vstupní data** – první blok je tvořen vstupními daty. Jedná se tedy o záznamy jednotlivých řečových cvičení (signálů) z vybrané databáze záznamů. Následně se tyto signály dostávají do bloku předzpracování. Tyto tzv. surová data jsou v konkrétní realizaci bakalářské práce načítány pomocí jistých pravidel (hodnocení počtu pacientů dané patologie). Více viz kapitola 6.3.
- **Předzpracování** – v bloku předzpracování řečových záznamů dochází k ustředění, preemfázi a segmentaci signálu, více o předzpracování řečových signálů, viz kapitolu 3.1. V této fázi se takto upravené záznamy cvičení dají použít pro výpočet řečových parametrů (parametrizace). Popis aplikace předzpracování v rámci realizace bakalářské práce se nachází v kapitole 6.4.
- **Parametrizace** – v této části dochází k výpočtu řečových parametrů, které jsou popsány v kapitole 3. Jejich použití vychází přímo ze zjednodušeného modelu generování řeči, viz kapitola 1.3. Další blok, který je považován za součást parametrizace je tzv. rozhodovací blok, který na základě dimenze parametru určí, zdali je nutné tento parametr dále zpracovávat v bloku, který aplikuje statistickou analýzu na parametry tvořeny vektory nebo maticemi. Tyto parametry následně upraví do podoby skalárních hodnot ve tvaru popsaných statistických hodnot těchto parametrů, jako jsou například střední hodnota, medián, atd. více popisuje kapitola 3.7. Pokud je tedy hodnota parametru ve tvaru skalární hodnoty, jde tento parametr přímo do dalšího bloku klasifikace. Pokud je ale hodnota vektor nebo matice, jde tento parametr do bloku extrakce vysokoúrovňových parametrů. Více o konkrétní realizaci parametrizace, viz kapitola 6.4.
- **Předběžná analýza** – v tomto bloku jsou vypočtené parametry analyzovány pomocí Spearmanova korelačního koeficientu a společné informace za účelem rozhodnutí o tom, které parametry jsou dále vybrány do procesu klasifikace patologické řeči. Více o tomto procesu pojednává kapitola 6.5.
- **Klasifikace** – v bloku klasifikace probíhá shluková analýza, která je blíže popsána v kapitole 4. Výsledkem této analýzy je zařazení jednotlivých patologií v databázi řečových promluv (samohláska „a“) do skupin podle společných klinických projevů. Ve výstupu jsou tedy patologie shlukovány podle toho v jakém rozsahu poškozují jednotlivá řečová ústrojí (fonační, artikulační, atd). Více o tomto procesu je možné najít v kapitole 6.6.

6 REALIZACE SYSTÉMU

6.1 Databáze řečových záznamů

Pro účely této bakalářské práce byla použita databáze MEEI (The Massachusetts Eye and Ear Infirmary Database), která již obsahuje nahrané hlasy velké sady různých řečníků uskupených v této databázi. MEEI databáze obsahuje celkem 53 zdravých a 657 nemocných řečníků, kteří jsou postiženi různými typy patologií (např. spastická dysfonie, konverzní dysfonie, hyperfunkce, atd.). Nahrávka každého řečníka obsahuje 12 vteřinový text (tzv. „The Rainbow Passage“) a také prodlouženou fonaci samohlásky „a“ s výslovností jako ve slově „father“. Pro účely této práce je použito právě toto řečové cvičení, tj. fonace samohlásky „a“. Nahrávky v databázi MEEI jsou vzorkovány s frekvencí $f_{vz} = 50$ kHz, $f_{vz} = 25$ kHz nebo se vzorkovací frekvencí $f_{vz} = 10$ kHz. Kompletní demografické informace k této databázi jsou shrnuty v Tab. 6.1. Tato databáze obsahuje celkem 200 patologií. V této práci byly použity pouze patologie, které obsahovaly alespoň 6 řečníků, ostatní by neměly z hlediska statistiky smysl. Vybrané patologie ukazuje tabulka Tab. B.1.

Tab. 6.1: Demografické informace vypočteny z MEEI databáze.

| Řečníci | Počet řečníků | | Věkový rozsah | | Průměrný věk | | std věku | |
|---------|---------------|------|---------------|-------|--------------|-------|----------|------|
| | Muži | Ženy | Muži | Ženy | Muži | Ženy | Muži | Ženy |
| Zdraví | 21 | 32 | 26-59 | 22-52 | 38,81 | 34,16 | 8,49 | 7,87 |
| Nemocní | 70 | 103 | 26-58 | 21-51 | 41,70 | 37,59 | 9,40 | 8,19 |

std představuje směrodatnou odchylku („standard deviation“)

6.2 Použité parametry

V rámci bakalářské práce byly použity řečové parametry podrobně popsané v kapitole 3. Na začátku implementace byly všechny tyto parametry uvažovány jako stejně vhodné a bylo předpokládáno, že dokážou rozlišit zdravou řeč a patologickou řeč s požadovanou přesností a následně rozlišit patologickou řeč do skupin podle závažnosti dané patologie. Tyto parametry je možné rozdělit do 3 skupin a to následovně:

- 1. Hodnocení variability základního tónu** – jedná se o parametry hodnotící poškození periodicity vibračního charakteru hlasivkových svalů. Častým jevem při poškození hlasivkového ústrojí bývá také ztuhlost hlasivkových svalů, popřípadě jejich omezená pohyblivost. Všechny vzpomínané faktory zásadním způsobem ovlivňují fonační schopnosti takto postižených řečníků. Nejčastěji se poškození fonačního ústrojí sleduje pomocí parametrů hodnotících variabilitu

základního hlasivkového tónu a variabilitu intenzity hlasu. V bakalářské práci byly použity následující parametry:

- a. Parametry hodnotící statistické vlastnosti základního hlasivkového tónu (F0) při analýze řečového projevu (fonaci samohlásky „a“), konkrétní popis jednotlivých parametrů je možné najít v kapitole 3.2.1 a v kapitole 3.2.2. Konkrétně se jedná o parametry:
 - Střední hodnota F0 (F0 avg⁴)
 - Směrodatná odchylka F0 (F0 std)
 - Maximální hodnota F0 (F0 max)
 - Minimální hodnota F0 (F0 min)
 - Jitter
- b. Parametry hodnotící statistické vlastnosti intenzity hlasu (A0) při analýze řečového projevu (fonaci samohlásky „a“), konkrétní popis jednotlivých parametrů je možné najít v kapitole 3.2.3. V bakalářské práci byl použit parametr Shimmer.

2. Hodnocení úrovně šumu v hlase – jedná se o parametr popisující úroveň nežádoucího šumu v uvažovaném řečovém signálu. Šum se v hlase může objevit například jako následek neúplného uzavření hlasivek, kdy i při klidovém stavu (stavu bez fonace) kvůli této vadě proudí skrze hlasivky vířivý proud vzduchu. Jedná se o tzv. hlasivkový tremor. V bakalářské práci byly použity následující parametry:

- a. Parametry hodnotící statistické vlastnosti změny energie ve spektru uvažovaného signálu. Energetická změna je sledována v jednotlivých segmentech. V této práci jsou uvažovány parametry jako krátkodobá energie (STE), dále také Teager-Kaiserův energetický operátor (TKEO) a medián výkonové spektrální hustoty (MPSD). Podrobný popis všech těchto parametrů se nachází v kapitolách 3.3.1 až 3.3.4. Vypočteny byly parametry:
 - Střední hodnota STE (STE avg)
 - Směrodatná odchylka STE (STE std)
 - Maximální hodnota STE (STE max)
 - Minimální hodnota STE (STE min)
 - Teager-Kaiserův energetický operátor
 - Střední hodnota MPSD (MPSD avg)
 - Směrodatná odchylka MPSD (MPSD std)
 - Maximální hodnota MPSD (MPSD max)
 - Minimální hodnota MPSD (MPSD min)

⁴ avg v kontextu této práce vyznačuje tzv. aritmetický průměr („average“); výpočet průměru je roven střední hodnotě daných parametrů. Tj. F0 avg vyznačuje střední hodnotu základního hlasivkového tónu za čas fonace samohlásky „a“. Průměr se počítá přes jednotlivé segmenty řečového cvičení.

- b. Parametry hodnotící poměr signálu od šumu HNR (harmonic-to-noise ratio) nebo odstup šumu od užitečného signálu NHR (noise-to-harmonic ratio). Bližší popis těchto parametrů je možné najít v kapitole 3.4.3 V práci jsou použity konkrétně dva parametry tohoto typu a to:
 - HNR
 - NHR
 - c. Parametry, které nepřímo měří kvalitu signálu a jejich statistické vlastnosti. Jedná se o parametry fungující na principu, kdy se pomocí počtu přechodu signálu nulovou úrovní rozhoduje o znělosti, popřípadě neznělosti úseku řečového signálu. Jsou to parametry typu ZCR (zero-crossing rate) nebo HZCR (high zero-crossing rate ratio). Konkrétní popis těchto parametrů se nachází v kapitolách 3.4.1, popřípadě 3.4.2. V práci byly počítány tyto parametry:
 - Střední hodnota ZCR (ZCR avg)
 - Směrodatná odchylka ZCR (ZCR std)
 - Maximální hodnota ZCR (ZCR max)
 - Minimální hodnota ZCR (ZCR min)
 - Střední hodnota HZCR (HZCR avg)
 - Směrodatná odchylka HZCR (HZCR std)
 - Maximální hodnota HZCR (HZCR max)
 - Minimální hodnota HZCR (HZCR min)
- 3. Hodnocení pohyblivosti artikulačních orgánů** – jedná se o parametry, které popisují pohyblivost artikulačních orgánů a tím pádem srozumitelnost řeči. Tyto parametry jsou jedny z nejčastěji používaných parametrů při analýze řečového projevu. V bakalářské práci byly použity parametry hodnotící polohu prvních tří formantových kmitočtů (1. form. – 3. form.) a jejich statistické vlastnosti (avg, std, avg, max, mix), viz předchozí parametry.

6.3 Načtení záznamů z databáze

Vstupní hodnoty (informace z databáze MEEI) jednotlivých řečníků jsou nejprve načteny do programového prostředí MATLAB, aby mohly být dále zpracovávány. Soubor **inspect_database.m** načte soubory *.NSP, ve kterých jsou uloženy nahrávky samohlásek „a“ všech řečníků z databáze. Tyto záznamy jsou následně použity pro výpočet řečových parametrů. Spolu s konkrétním průběhem řečové nahrávky (tzv. „waveform“) jsou v bakalářské práci ukládány také následující informace o nahrávce, popř. řečníkovi: název patologie, vzorkovací frekvence (tyto jsou nutné pro následující zpracování). Z těchto záznamů je vytvořen soubor **MEEI_records.mat**. Tento soubor obsahuje celkem 31 skupin patologických řečí a jednu skupinu zdravých řečníků, kde u každého záznamu (jednotlivých řečníků) jsou ukládány všechny vzpomínané informace.

6.4 Výpočet parametrů

Pro výpočet řečových parametrů byl v rámci bakalářské práce vytvořen soubor **feature_extraction.m**, který načte databázi **MEEI_records.mat** a z ní postupně počítá jednotlivé řečové parametry. Vypočtené parametry jsou následně ukládány do tabulky **feat_matrix.mat**. K výpočtu je u některých parametrů použit volně dostupný program **praatcon.exe** [11]. Tento program je volně dostupný na <http://www.praat.org> (Phonetic Sciences, University of Amsterdam, The Netherlands). Jednotlivé parametry, které jsou v bakalářské práci počítány jsou uvedeny zde:

- **praat_info.m** – tento skript volá program **praatcon.exe**, který vypočítá parametr F0 (základní frekvence kmitání hlasivek) viz, kapitola 3.2.1. Počítá jeho střední hodnotu, standardní odchylku, minimum a maximum. Dále vypočítá Jitter viz kapitola 3.2.2, Shimmer viz podkapitola 3.2.3, HNR a NHR viz 3.4.3.
- **praat_formants.m** – skript volá program **praatcon.exe**, který vrací střední hodnotu, standardní odchylku, minimum a maximum pro první, druhý a třetí formant, viz kapitola 3.5.
- **energy.m** – skript vypočítá střední hodnotu, standardní odchylku, minimum a maximum krátkodobé energie, viz kapitola 3.3.1.
- **zero_crossing_rate.m** – vypočítá střední hodnotu, standardní odchylku, minimum a maximum počtu průchodů signálu nulovou úrovní viz kapitola 3.4.1.
- **tkeo.m** – tento skript počítá střední hodnotu, standardní odchylku, minimum a maximum Teager-Kaisrova energetického operátoru, viz kapitola 3.3.3.
- **mean_psd.m** – skript pro výpočet střední hodnoty výkonové spektrální hustoty signálu, viz kapitola 3.3.4.
- **high_zero_crossing_rate.m** – tento skript počítá poměr počtů segmentů, ve kterých je průchod nulovou úrovní větší než 1,5 násobek průměrného průchodu nulovou úrovní v jedné sekundě signálu, viz kapitola 3.4.2. S tímto parametrem se však v této práci nepočítá. Parametr nešel z důvodu příliš krátkého trvání nahrávky (fonace samohlásky „a“) vypočítat.
- **low_short_time_energy.m** – tento skript vypočítá poměr segmentů, ve kterých je krátkodobá energie nižší než 0,5 násobek průměrné krátkodobé energie v sekundovém signálu viz kapitola 3.3.2. S tímto parametrem se stejně jako s předchozím uvedeným nepočítá z důvodu krátkého trvání nahrávky.

6.4.1 Následné zpracování

Soubor **post_proc.m** odstraní sloupce z **feat_matrix.mat**, ve kterých se nachází prvky, které nejsou číselné, jsou nekonečné, nebo jsou komplexními čísly. Tento skript odstraní sloupce obsahující střední hodnotu a směrodatnou odchylku třetího formantu, protože se v nich nacházely nečíselné hodnoty.

6.5 Předběžná analýza

Předběžná analýza byla v bakalářské práci zvolena jako prvotní analýza vypočtených řečových parametrů. Jejím účelem bylo vzít všechny řečové parametry (pro všechny tři uvažované skupiny parametrů, blíže viz kapitola 3) a prozkoumat, které z těchto parametrů dokážou nejlépe rozlišit zdravou a patologickou řeč. Na základě této úvahy: „parametr, který dokáže s dostatečnou úspěšností rozlišit tyto dvě skupiny je vhodným kandidátem na následné zpracování pomocí shlukové analýzy“ byl analyzován první scénář, kde byl pomocí vytvořeného skriptu **pre_analysis_1.m** počítán Spearmanův korelační koeficient a tzv. společná informace. Tyto ukazatele statistické závislosti dvou veličin jsou popsány níže. Na základě výpočtu těchto hodnot byly vybrány parametry pro následné zpracování pomocí algoritmu k-means. Výsledky analýzy prvního scénáře je možné najít v Tab. 6.2. V rámci předběžné analýzy byl uvažován i druhý scénář, kde nebylo se zdravou řečí již počítáno. Cílem této analýzy byl výpočet Spearmanova korelačního koeficientu a společné informace pro jednotlivé patologie zvlášť. Konkrétní postup je popsán níže. Výsledky předběžné analýzy pro druhý scénář jsou zhrnuty v Tab. 6.2.

Spearmanův korelační koeficient (ρ) je bezrozměrné číslo udávající statistickou závislost mezi dvěma veličinami. Nabývá hodnot -1 až 1. Čím větší číslo, tím je větší závislost mezi těmito dvěma vstupními veličinami. Vstupními parametry jsou vypočtené parametry a čísla skupin, do které daná patologie patří. V prvním případě bylo zdravé řeči přiřazeno číslo 1 a patologické řeči číslo 2, a hledala se korelace mezi těmito dvěma skupinami. V druhém případě, bez zdravé řeči, bylo každé skupině patologií přiřazeno jiné číslo a hledala se vzájemná korelace mezi všemi skupinami. Tento koeficient byl počítán, aby ukázal, které parametry se nejlépe hodí na oddělení zdravé a patologické řeči.

Společná informace („mutual information“, tj. MI) je bezrozměrná veličina. Udává míru množství informace jedné náhodné veličiny vůči druhé náhodné veličině. Vstupními parametry jsou opět vypočtené parametry a čísla skupin, do které daná patologie patří tak jako u Spearmanova korelačního koeficientu. Je to poměr mezi sdruženou pravděpodobností a součinu marginálních pravděpodobností [12]:

$$MI(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}, \quad (6.1)$$

kde $p(x, y)$ je sdružená pravděpodobnost a $p(x)$ a $p(y)$ jsou marginální pravděpodobnosti. Z Tab. 6.2 při uvažování patologické a zdravé řeči vyplývá, které parametry by měly nejlépe separovat zdravou a patologickou řeč v následné shlukové analýze. Pro těchto devět parametrů byla následně vypočítána hustota pravděpodobnosti. Znázorňuje ji Obr. 6.1. Pro výpočet hustoty pravděpodobnosti je použit soubor **density_estim.m**. Pro zobrazení hustoty pravděpodobnosti byly vybrány vždy první tři zástupci z tabulky.

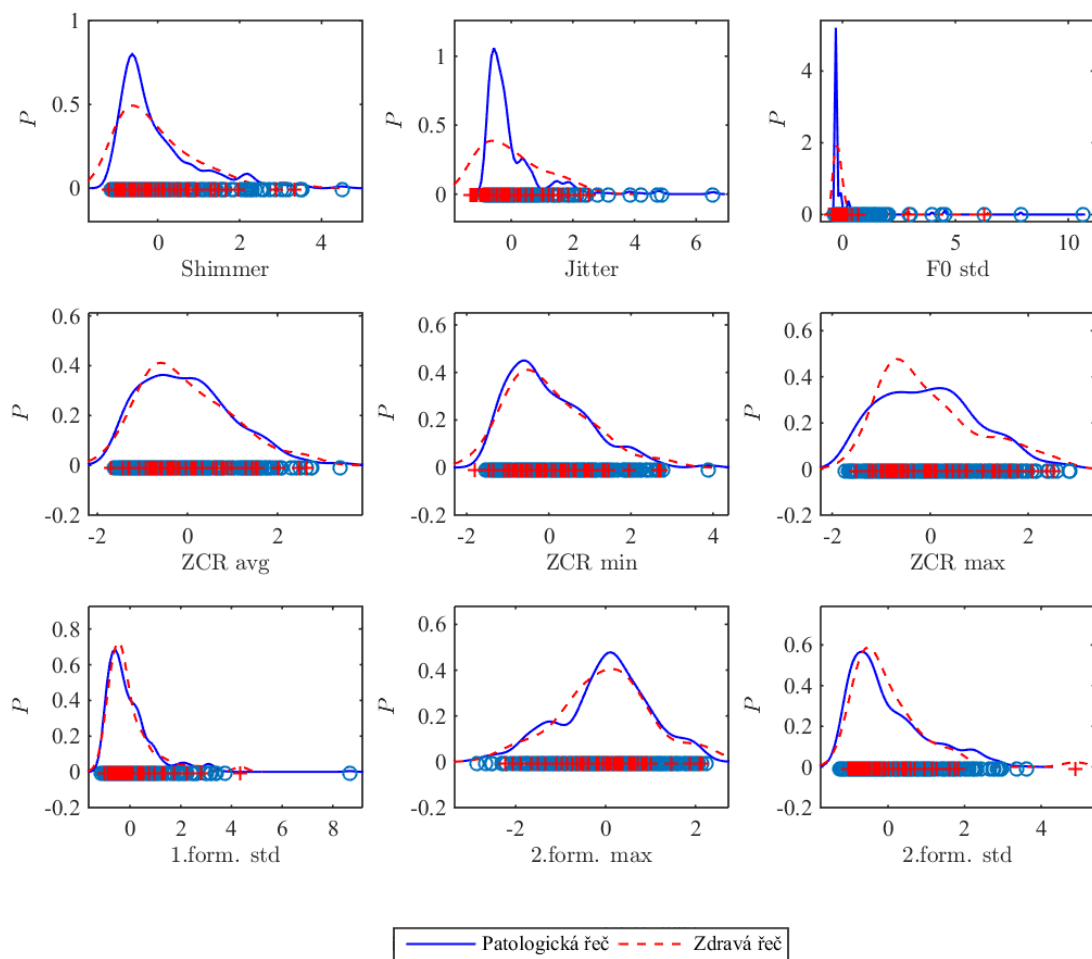
Tab. 6.2: Tabulka předběžné analýzy šesti nejlepších parametrů seřazených sestupně podle hodnot Spearmanova korelačního koeficientu (ρ). Tabulka obsahuje analýzu pro případ analýzy zdravé a patologické řeči a také analýzu pouze řeči patologické.

| Předběžná analýza při uvažování patologické i zdravé řeči | | | | | | | | |
|---|---------|--------|---------------------|--------|--------|----------------------|--------|--------|
| variabilita základního tónu | | | úroveň šumu v hlase | | | schopnost artikulace | | |
| Parametr | ρ | MI | Parametr | ρ | MI | Parametr | ρ | MI |
| Shimmer | 0,3296 | 0,2706 | ZCR avg | 0,3739 | 0,3008 | 1. form. std | 0,2559 | 0,3008 |
| Jitter | 0,3236 | 0,2975 | ZCR min | 0,3709 | 0,2502 | 2. form. max | 0,1701 | 0,3008 |
| F0 std | 0,0557 | 0,2827 | ZCR max | 0,3654 | 0,2636 | 2. form. std | 0,1525 | 0,3008 |
| F0 max | -0,1679 | 0,3008 | TKEO std | 0,3413 | 0,3008 | 1. form. max | 0,1292 | 0,3008 |
| F0 avg | -0,1874 | 0,3008 | TKEO max | 0,3266 | 0,3008 | 2. form. avg | 0,0844 | 0,3008 |
| F0 min | -0,1988 | 0,3008 | ZCR std | 0,3233 | 0,3008 | 3. form. min | 0,0782 | 0,2928 |
| Předběžná analýza při uvažování pouze patologické řeči | | | | | | | | |
| variabilita základního tónu | | | úroveň šumu v hlase | | | schopnost artikulace | | |
| Parametr | ρ | MI | Parametr | ρ | MI | Parametr | ρ | MI |
| F0 std | 0,0110 | 0,2706 | TKEO avg | 0,0541 | 2,4907 | 3. form. min | 0,0630 | 2,2561 |
| F0 max | 0,0098 | 0,2975 | STE min | 0,0523 | 2,4907 | 1. form. max | 0,0483 | 2,4907 |
| Shimmer | 0,0049 | 0,2827 | TKEO std | 0,0479 | 2,4907 | 3. form. max | 0,0465 | 2,2613 |
| F0 avg | -0,0018 | 0,3008 | TKEO max | 0,0462 | 2,4907 | 1. form. avg | 0,0314 | 2,4907 |
| Jitter | -0,0022 | 0,3008 | MPSD | 0,0437 | 2,4712 | 2. form. max | 0,0281 | 2,4907 |
| F0 min | -0,0098 | 0,3008 | ZCR min | 0,0432 | 1,8171 | 2. form. min | 0,0266 | 2,4907 |

ρ značí Spearmanův korelační koeficient; MI značí společnou informaci (mutual information)

Na základě Obr. 6.1 bylo v rámci bakalářské práce usouzeno, že vypočtené parametry pro patologickou a zdravou řeč jsou statisticky rozdílné a je tehdy možné tyto parametry použít pro diferencování mezi těmito dvěma skupinami řečníků. Z grafu hustoty pravděpodobnosti je vidět, že rozložení hodnot pro zdravé a patologické řečníky se liší a to zejména u první skupiny parametrů, je teda možné usuzovat, že sledování variability základního tónu řeči a intenzity, tj. úrovní tzv. dysfonie je dobrým indikátorem případné řečové vady. Z grafu a Tab. 6.2 je také vidět, že pozorováním prvního formantu, který je zároveň dominantním pro hlásku „a“, viz kapitola 3.5 je možné oddělit patologickou a zdravou řeč. Pro parametry založené na odhadu kvality hlasu vyšel nejlépe parametr sledující počet průchodu signálu nulovou úrovní.

Na základě těchto pozorování byly parametry uvedené v Tab. 6.2 vybrány jako vhodné pro následující zpracování pomocí algoritmu k-means, sloužícímu pro nalezení nového dělení patologií na základě právě uvažovaných parametrů. Samotný proces aplikace shlukové analýzy je popsán v následující kapitole.



Obr. 6.1: Grafy hustoty pravděpodobnosti pro tři nejlépe korelující parametry (byla uvažována Spearmanova korelace mezi hodnotami parametrů zdravých a patologických řečníků a jejich klinickým stavem, tj. zdravý/nemocní) vypočtené pomocí jádrových odhadů s použitím Gaussova jádra. První řádek značí parametry založené na sledování variability základního tónu a intenzity řeči. Druhý řádek je vyhrazen pro parametry sledující energetické změny ve spektru sledovaného signálu. Třetí řádek je vyhrazen pro parametry pracující na základě sledování pohyblivosti artikulačních orgánů.

6.6 Klasifikace k-means

Soubory `k_means_distribution.m` a `iteration_distribution.m` jsou vytvořené k určení, které parametry nejlépe dokaží rozřadit patologickou a zdravou řeč do odlišných skupin. k-means (viz kapitola 4.2) postupně rozřazoval patologickou a zdravou řeč do tří, čtyř a pěti skupin. Celkově se zde objevuje 53 zdravých řečníků. Procentuální rozdělení ukazuje Tab. 6.3.

Tab. 6.3: Separace zdravé řeči pomocí k-means

| parametr | Počet (k) skupin algoritmu k-means | | | | | |
|------------|------------------------------------|-------|-------|-------|-------|-------|
| | k = 3 | ACC | k = 4 | ACC | k = 5 | ACC |
| Jitter | 53 | 100% | 53 | 100% | 53 | 100% |
| ZCR avg | 53 | 100% | 53 | 100% | 53 | 100% |
| 1.form std | 50 | 94,3% | 50 | 94,3% | 47 | 88,7% |

ACC představuje úspěšnost klasifikace

Z první skupiny hodnocení variability základního tónu vyšel nejlépe parametr Jitter. Z druhé skupiny hodnocení úrovně šumu v hlase vyšel nejlépe parametr ZCR avg. Ze třetí skupiny hodnocení pohyblivosti artikulačních orgánů byl nejlepší parametr 1.form std. Rozřazení zobrazuje Tab. 6.3.

6.7 Konečné výsledky

Na základě klasifikace k-means byly vybrány tři nejlepší parametry Jitter, ZCR avg a 1.form std. Tyto parametry měly i nejlepší výsledky v předběžné analýze (viz kapitola 6.5), což vede k tomu, že by mohly nejlépe reprezentovat své skupiny.

Soubor **k_means.m** slouží ke konečnému rozdělení patologických řečí do tří skupin. Vstupem do tohoto souboru jsou již zmiňované tři nejlepší parametry. Při těchto výpočtech už se neuvažuje zdravá řeč. Vstupem do tohoto souboru je tedy pouze řeč patologická. Očekávaným výstupem mělo být zařazení co největšího počtu, nejlépe všech, řečníků do jedné ze tří skupin. Tři skupiny byly uvažovány z důvodu nejlepšího separování zdravé a patologické řeči při použití právě tří skupin.

Pro výstupní výsledky byly zhotoveny dva scénáře. V prvním se patologie rozřazovaly na tři skupiny v rámci každé ze skupin, kterými byly hodnocení variability základního tónu, hodnocení úrovně šumu v hlase a hodnocení pohyblivosti artikulačních orgánů. Použitý soubor je **k_means.m**. Následné rozdělení pak vykonává soubor **sorting_pathology.m**. Výsledky ukazuje Tab. C.1 a Tab. C.2. Patologie, které měly v jedné ze tří skupin alespoň 60% zastoupení by se daly z hlediska konkrétního parametru zastupujícího danou dysfunkci řečového ústrojí považovat za podobné. V první skupině hodnocení variability základního tónu se téměř všechny patologie rozřadily do stejné skupiny (proto také nejsou v tabulce Tab. C.4 uvedeny, jejich zařazení je však vidět v Tab. C.1 a Tab. C.2). Odstraněny byly pouze patologie *bowing* a *ventricular compression (full)* a to kvůli nedosažení 60% zastoupení. Ostatní rozřazení ukazuje Tab. C.4.

Druhý scénář už nepočítá s každou skupinou parametrů zvlášť. Vstupem jsou všechny tři skupiny dohromady jako jedna matice. K tomuto výpočtu je použit soubor **k_means_matrix**. Setřídění potom provádí soubor **sorting_pathology_matrix.m**. Výsledky ukazuje Tab. C.3. Opět byly odstraněny patologie, které nedosahovali alespoň 60% zastoupení v některé ze skupin, a byla vytvořena Tab. 6.4, která ukazuje, které patologie by mohly mít společné vlastnosti v rámci poškození řečové produkce.

Tab. 6.4: Zařazení patologií do skupin podle podobnosti vypočtených výsledků pro druhý scénář.

| 1.skupina | 2.skupina | 3.skupina |
|-----------|-----------|------------------------------------|
| | | A-P squeezing (mild) |
| | | A-P squeezing (moderate) |
| | | A-P squeezing (severe) |
| | | adductor spasmodic dysphonia |
| | | conversion dysphonia |
| | | gastric reflux |
| | | interarytenoid hyperplasia |
| | | polypoid degeneration (Reinke's) |
| | | post thyroplasty |
| | | ventricular compression (mild) |
| | | ventricular compression (moderate) |
| | | vocal fold edema |
| | | vocal fold polyp |

Všechny patologie patřící do stejné skupiny by tedy mohly mít podobné klinické příznaky dané skupiny. Pro hodnocení variability základního tónu to je jitter. Pro hodnocení úrovně šumu v hlase je to průměrná hodnota počtu průchodu nulovou úrovní (ZCR avg). A pro hodnocení pohyblivosti artikulačních orgánů je to směrodatná odchylka prvního formantu (1.form. std). Tyto vady však nemusí mít stejné příčiny. Mohou být způsobené jak nemocí, pooperačními potížemi tak i například rakovinou. Tato práce se zabývá pouze společnými rysy v rámci poškození řečové produkce.

7 ZÁVĚR

Cílem bakalářské práce byla analýza řečového projevu pacientů s různou formou řečových patologií za účelem analýzy s využitím algoritmů tzv. shlukové analýzy. Výsledkem této analýzy je nalezení patologií řeči, které navzdory různému původu mají společné klinické příznaky ve formě degradace řečové produkce. Pro tyto účely byly v rámci bakalářské práce nastudovány základní poznatky o tvorbě hlasu v tzv. hlasovém ústrojí a také uložení těchto orgánů v hlasovém traktu. Dále byla popsána tvorba řeči z fyziologického hlediska a byl zde uveden matematický model tvorby řeči, který je základem velkého množství řečových parametrů. V práci byla dále rozebrána problematika patologické řeči. Byly zde zmíněné i některé parametry, které hodnotí projevy patologií řečového projevu na konkrétních částech řečového ústrojí. Nastudována byla také problematika shlukové analýzy, která byla využita k rozdělení řečových patologií na základě řečových parametrů popisujících klinické příznaky těchto patologií, které se projevují degradovanou funkčností konkrétní části řečového aparátu.

Pro načtení databáze byl vytvořen soubor, který načítá potřebná data (fonaci hlásky „a“) z databáze MEEI a zároveň vylučuje patologie, které kvůli malému počtu zastoupení nebylo vhodné uvažovat. Z načtených údajů jsou vypočteny parametry popsané v kapitole 3. Na všechny tyto parametry je uplatněna předběžná analýza. Tato analýza je vytvořena pro separaci zdravé a patologické řeči. Očekávaným výstupem bylo najít parametry, které při uvažování zdravé a patologické řeči měly největší Spearmanův korelační koeficient a tedy nejlépe reprezentovaly poškození řečové produkce, jinými slovy nejlépe kvantifikovaly zdravou a patologickou řeč. Byla také provedena k-means klasifikace, která měla potvrdit tvrzení o nejlepších parametrech. Následně s těmito parametry bylo počítáno jako s parametry, které nejlépe reprezentují svoji skupinu popisující poškození dané části řečového ústrojí. Tyto parametry byly vstupem do konečného k-means algoritmu. Očekávaným výstupem bylo rozřazení patologií do skupin dle jejich společných rysů v rámci poškození řečové produkce.

Z Tab. 6.2 je vidět, které parametry se k separaci zdravé a patologické řeči nejlépe hodí. Jako referenční parametry byly zvoleny ty, které měly nejvyšší hodnotu Spearmanova korelačního koeficientu. Tvrzení o nejlepší separaci potvrdila také následná k-means klasifikace jak ukazuje Tab. 6.3. Pro konečné rozřazení bylo znovu využito algoritmu k-means. V prvním scénáři byly tyto tři parametry s nejlepšími výsledky, které ukazuje Tab. 6.3 algoritmem rozřazeny do tří skupin viz Tab. C.1 a Tab. C.2 v rámci dané skupiny hodnotící konkrétní poškození řečového projevu. Z výsledků se dá uvažovat, že parametry zařazené do stejné skupiny viz Tab. C.4 by mohly mít podobné klinické příznaky na konkrétní části řečového ústrojí, které daná skupina reprezentuje. V druhém scénáři byla vstupem do algoritmu k-means matice obsahující tři nejlepší parametry jako celek. Výsledky jsou vidět v Tab. C.3. Rozřazení patologií do skupin potom ukazuje Tab. 6.4.

Z výsledků shlukové analýzy je možné uvažovat skupiny patologií, které navzdory rozdílnému původu mají podobné příznaky ve formě degradace určité části řečové produkce. V rámci bakalářské práce byly tedy vytvořeny skupiny patologií, které podobným způsobem poškozují fonační ústrojí, tj. samotné kmitání hlasivek, dále artikulační schopnosti a samotnou kvalitu řeči, reprezentovanou hladinou šumu v řečovém signálu. Byla také nalezena skupina patologií, která má společnou manifestaci pro všechny tři uvažované atributy řeči.

Tato práce by mohla být v budoucnu rozšířena o bližší analýzu jednotlivých patologií. Využilo by se dělení uvedené v této práci a na základě uvedených výsledků provedla hlubší analýza fonace, artikulace, kvality řeči pomocí robustnějšího parametrizačního aparátu. Výsledkem by byl komplexnější pohled na jednotlivé manifestace těchto patologií na řečovém ústrojí. Bylo by také možné prozkoumání jenom několika vybraných patologií, které mají podobný klinický dopad na řečové schopnosti a hledání parametrů, které mezi nimi dokáží nejlépe diferencovat. Tento přístup by mohl vést k systému včasné analýzy konkrétní vady řeči, která je často svázána s výskytem jisté nemoci, kupříkladu Parkinsonovi nemoci, kdy je v raném stádiu nemoci těžké určit konkrétní diagnózu.

POUŽITÁ LITERATURA

- [1] ALEXANDRE-CORTIZO, E., M. ROSA-ZURERA a F. LOPEZ-FERRERAS. *Application of Fisher Linear Discriminant Analysis to Speech/Music Classification*. ISBN 10.1109/eurcon.2005.1630291.
- [2] ČERNOCKÝ, J. *Zpracování řečových signálů*. Ústav počítačové grafiky a multimédií, FIT, VUT Brno, 2006.
- [3] JIRSÍK, Václav, HRÁČEK, Petr. *Neuronové sítě, expertní systémy a rozpoznávání řeči*. Brno : [s.n.], 2000. 106 s. Skriptum FEKT VUT v Brně.
- [4] LU, Lie, Hong-Jiang ZHANG a Hao JIANG. *Content analysis for audio classification and segmentation*. ISBN 10.1109/tsa.2002.804546.
- [5] LUKASOVÁ, A a M ŠARMANOVÁ. *Metody shlukové analýza*. Praha, 1985, 210 s.
- [6] PSUTKA, Josef, et al. *Mluvíme s počítačem česky*. Praha: ACADEMIA, 2006. 752 s. ISBN 80-200-1309-1.
- [7] SIGMUND, Milan. *Rozpoznávání řečových signálů: přednášky*. 1. vyd. Brno: VUT FEKT, ústav radioelektroniky, 2007, 122 s. ISBN 978-80-214-3526-1.
- [8] SMÉKAL, Z.: *Číslicové zpracování řeči (MZPR)*. Elektronická skripta pro magisterská studia, VUT Brno, 2010.
- [9] SMÉKAL, Z. *Číslicové zpracování signálů (MCSI)*. Elektronická skripta pro magisterská studia, Ústav telekomunikací, FEEC, VUT Brno, 2009.
- [10] TITZE, I.R.: *Principles of voice production*. Prentice Hall, 1994, ISBN 9780137178933.
- [11] P. Boersma, D. Weenink, Praat: Doing Phonetics by Computer (Version 5.3.10), 2012.
- [12] Cover, T. M.; Thomas, J. A. *Elements of information theory*. 1991, ISBN 0-471-06259-6.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

| | |
|-----------|--|
| DDK | Diadocho-kynetické úlohy |
| HNR | odstup harmonické od šumu |
| HP | horní propust |
| HZCRR | poměr vysokých průchodů nulovou úrovní |
| LSTER | poměr nízké krátkodobé energie |
| MI | vzájemná informace |
| MPSD | medián výkonové spektrální hustoty |
| NSRA | tempo čistě řečového signálu |
| NST | trvání čistě řečového signálu |
| PSD | výkonová spektrální hustota |
| STE | krátkodobá energie |
| TKEO | Teagero-Kaiserův energetický operátor |
| TSR | celkové tempo řeči |
| TST | celkové trvání řečového signálu |
| TPT | celkové trvání pauz |
| ZCR | počet průchodů signálu nulovou úrovní |
| a_1 | koefficient preemfázového filtru |
| $A_g[i]$ | i -tá hodnota amplitudy |
| E | krátkodobá energie |
| F_0 | základní hlasivkový tón |
| f_{vz} | vzorkovací frekvence |
| $H(z)$ | přenosová funkce preemfázového filtru |
| l_{ram} | velikost rámce ve vzorcích |
| J_g | jitter |
| k_{min} | pozice prvního maxima |
| N | počet vzorků signálu |
| N_{ram} | velikost nepřekryté části rámce |
| $ N_i $ | energie harmonického signálu |

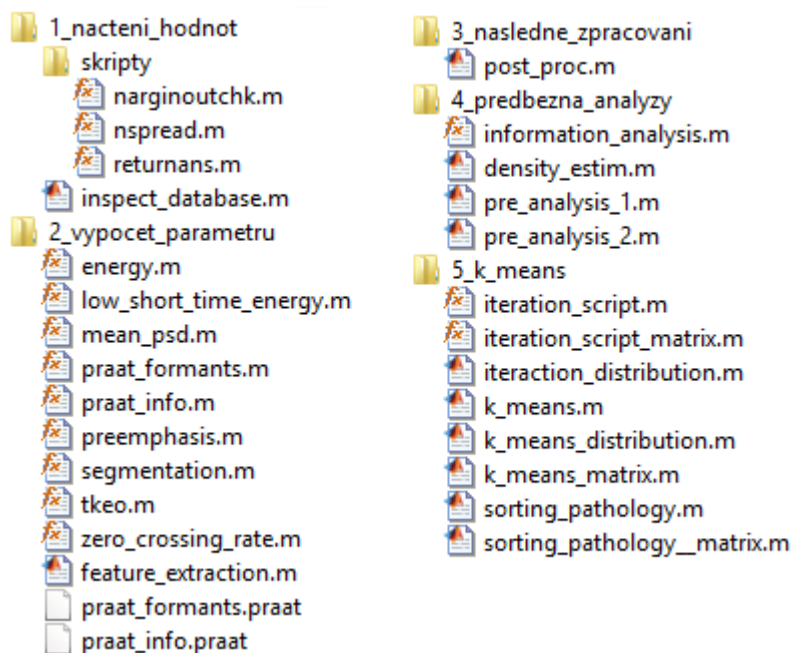
| | |
|------------------|---|
| $s[k]$ | vzorek diskrétního řečového signálu |
| $s[n]$ | řečový signál |
| $s'[n]$ | řečový signál po ustředění |
| S_g | shimmer |
| S_{ram} | velikost nepřekryté části rámce ve vzorcích |
| $ S_i $ | odhad energie šumu |
| $T_g[i]$ | je i -tá hodnota základní periody |
| U | normalizační koeficient modifikovaného periodogramu |
| $w[n]$ | váhovací posloupnost |

A OBSAH PŘILOŽENÉHO DVD

Příložené DVD obsahuje soubory vytvořené v rámci bakalářské práce. Jsou zde *.mat soubory obsahující jednotlivé skripty a funkce.

Databáze záznamů jednotlivých řečníků a DVD nejsou, protože se jedná o důvěrná data, ke kterým není volný přístup. Ve smyslu tohoto nařízení nejsou na příložením DVD ani vypočtené soubory *.mat obsahující segmentované záznamy a vypočtené parametry. Obsah DVD je tedy následovný:

- Složka: Matlab – zde jsou všechny *.m soubory použité v práci:



- Soubor: bakalářská práce.pdf

B TABULKA PATOLOGIÍ OBSAŽENÝCH V MEEI DATABÁZI

Tab. B.1: Informace o patologiích. U parametrů, u kterých není uveden český název nebyl tento název dohledán.

| Původní název | Česky překlad | Počet řečníků |
|------------------------------------|------------------------------|---------------|
| A-P squeezing | | 52 |
| A-P squeezing (mild) | | 56 |
| A-P squeezing (moderate) | | 37 |
| A-P squeezing (severe) | | 19 |
| adductor spasmodic dysphonia | křečovitá dysfónie | 19 |
| bowing | | 11 |
| conversion dysphonia | dysfónie | 13 |
| cyst | cysta | 7 |
| erythema | erytém | 14 |
| gastric reflux | žaludeční reflux | 48 |
| granulation tissue | granulační tkáň | 6 |
| hyperfunction | hyperfunkce | 263 |
| interarytenoid hyperplasia | interarytenoidní hyperplázie | 9 |
| keratosis / leukoplakia | keratóza / leukoplaktie | 26 |
| laryngeal web | hrtanová blána | 7 |
| paralysis | paralýza | 64 |
| paresis | částečná paralýza | 18 |
| polypoid degeneration (Reinke's) | polypoidní degenerace | 24 |
| post laryngoplasty | post laryngoplastie | 6 |
| post thyroplasty | post tyroplasie | 9 |
| presbyphonia | presbyopie | 13 |
| scarring | zjizvení | 15 |
| ventricular compression | ventrikulární komprese | 16 |
| ventricular compression (full) | úplná vent. komprese | 8 |
| ventricular compression (mild) | mírná vent. komprese | 36 |
| ventricular compression (moderate) | tlumená vent. komprese | 22 |
| ventricular compression (severe) | těžká vent. komprese | 23 |
| vocal fold edema | hlasivkový opuch | 44 |
| vocal fold polyp | polyp hlasivek | 20 |
| vocal nodules | hlasivkové výrostky | 19 |
| vocal tremor | hlasivkový třes | 14 |
| normal voice | zdravá řeč | 53 |

C TABULKY VÝSLEDKŮ

Tab. C.1: Tabulka vypočtených hodnot pro první scénář 1.část. Zkratka sk. znamená skupinu, do které byla daná patologie zařazena. Výskyt potom procentuálně vyjadřuje četnost zařazení do dané skupiny. Jestliže patologie nedosáhla alespoň 60% není zařazena do žádné ze skupin.

| k-means rozdělení | Hodnocení variability základního tónu | | | | | Hodnocení úrovně šumu v hlase | | | | | Hodnocení pohyblivosti artikulačních orgánů | | | | |
|------------------------------|---------------------------------------|----|---|-----|--------|-------------------------------|-----|----|-----|--------|---|----|----|-----|--------|
| | 1 | 2 | 3 | sk. | výskyt | 1 | 2 | 3 | sk. | výskyt | 1 | 2 | 3 | sk. | výskyt |
| A-P squeezing | 41 | 11 | 0 | 1 | 78,8% | 17 | 28 | 7 | - | - | 28 | 20 | 4 | - | - |
| A-P squeezing (mild) | 46 | 9 | 1 | 1 | 82,1% | 22 | 29 | 5 | - | - | 34 | 19 | 3 | 1 | 60,7% |
| A-P squeezing (moderate) | 36 | 1 | 0 | 1 | 97,3% | 17 | 14 | 6 | - | - | 24 | 9 | 4 | 1 | 64,9% |
| A-P squeezing (severe) | 16 | 2 | 1 | 1 | 84,2% | 8 | 8 | 3 | - | - | 12 | 7 | 0 | 1 | 63,2% |
| adductor spasmodic dysphonia | 15 | 3 | 1 | 1 | 78,9% | 7 | 11 | 1 | - | - | 14 | 4 | 1 | 1 | 73,7% |
| bowing | 6 | 5 | 0 | - | - | 7 | 2 | 2 | 1 | 63,6% | 6 | 4 | 1 | - | - |
| conversion dysphonia | 8 | 3 | 2 | 1 | 61,5% | 5 | 6 | 2 | - | - | 8 | 5 | 0 | 1 | 61,5% |
| cyst | 6 | 1 | 0 | 1 | 85,7% | 2 | 4 | 1 | - | - | 3 | 3 | 1 | - | - |
| erythema | 12 | 2 | 0 | 1 | 85,7% | 4 | 7 | 3 | - | - | 6 | 5 | 3 | - | - |
| gastric reflux | 42 | 5 | 1 | 1 | 87,5% | 18 | 25 | 5 | - | - | 32 | 15 | 1 | 1 | 66,7% |
| granulation tissue | 5 | 1 | 0 | 1 | 83,3% | 1 | 4 | 1 | 2 | 66,7% | 3 | 2 | 1 | - | - |
| hyperfunction | 221 | 36 | 6 | 1 | 84,0% | 100 | 120 | 43 | - | - | 157 | 88 | 18 | - | - |
| interarytenoid hyperplasia | 8 | 1 | 0 | 1 | 88,9% | 4 | 4 | 1 | - | - | 6 | 2 | 1 | 1 | 66,7% |
| keratosis / leukoplakia | 23 | 2 | 1 | 1 | 88,5% | 11 | 9 | 6 | - | - | 14 | 9 | 3 | - | - |
| laryngeal web | 6 | 1 | 0 | 1 | 85,7% | 3 | 1 | 3 | - | - | 3 | 4 | 0 | - | - |
| paralysis | 44 | 18 | 2 | 1 | 68,8% | 30 | 22 | 12 | - | - | 34 | 24 | 6 | - | - |
| paresis | 15 | 3 | 0 | 1 | 83,3% | 9 | 7 | 2 | - | - | 7 | 6 | 5 | - | - |

Tab. C.2: Tabulka vypočtených hodnot pro první scénář 2.část. Zkratka sk. znamená skupinu, do které byla daná patologie zařazena. Výskyt potom procentuálně vyjadřuje četnost zařazení do dané skupiny. Jestliže patologie nedosáhla alespoň 60% není zařazena do žádné ze skupin.

| k-means rozdělení | Hodnocení variability základního tónu | | | | | Hodnocení úrovně šumu v hlase | | | | | Hodnocení pohyblivosti artikulačních orgánů | | | | |
|------------------------------------|---------------------------------------|---|---|-----|--------|-------------------------------|----|---|-----|--------|---|----|---|-----|--------|
| | 1 | 2 | 3 | sk. | výskyt | 1 | 2 | 3 | sk. | výskyt | 1 | 2 | 3 | sk. | výskyt |
| polypoid degeneration (Reinke's) | 21 | 3 | 0 | 1 | 87,5% | 11 | 12 | 1 | - | - | 15 | 6 | 3 | 1 | 62,5% |
| post laryngoplasty | 5 | 1 | 0 | 1 | 83,3% | 1 | 5 | 0 | 2 | 83,3% | 2 | 3 | 1 | - | - |
| post thyroplasty | 7 | 2 | 0 | 1 | 77,8% | 4 | 3 | 2 | - | - | 7 | 2 | 0 | 1 | 77,8% |
| presbyphonia | 10 | 3 | 0 | 1 | 76,9% | 7 | 5 | 1 | - | - | 7 | 6 | 0 | - | - |
| scarring | 12 | 3 | 0 | 1 | 80,0% | 6 | 6 | 3 | - | - | 8 | 7 | 0 | - | - |
| ventricular compression | 13 | 3 | 0 | 1 | 81,3% | 7 | 7 | 2 | - | - | 9 | 6 | 1 | - | - |
| ventricular compression (full) | 4 | 2 | 2 | - | - | 5 | 3 | 0 | 1 | 62,5% | 3 | 4 | 1 | - | - |
| ventricular compression (mild) | 31 | 4 | 1 | 1 | 86,1% | 21 | 13 | 2 | - | - | 24 | 7 | 5 | 1 | 66,7% |
| ventricular compression (moderate) | 21 | 1 | 0 | 1 | 95,5% | 8 | 10 | 4 | - | - | 14 | 6 | 2 | 1 | 63,6% |
| ventricular compression (severe) | 18 | 5 | 0 | 1 | 78,3% | 10 | 7 | 6 | - | - | 13 | 9 | 1 | - | - |
| vocal fold edema | 39 | 5 | 0 | 1 | 88,6% | 14 | 25 | 5 | - | - | 31 | 12 | 1 | 1 | 70,5% |
| vocal fold polyp | 15 | 4 | 1 | 1 | 75,0% | 8 | 6 | 6 | - | - | 14 | 4 | 2 | 1 | 70,0% |
| vocal nodules | 19 | 0 | 0 | 1 | 100,0% | 5 | 13 | 1 | 2 | 68,4% | 10 | 7 | 2 | - | - |
| vocal tremor | 11 | 3 | 1 | 1 | 73,3% | 5 | 9 | 1 | 2 | 60,0% | 7 | 7 | 1 | - | - |

Tab. C.3: Tabulka vypočtených hodnot pro druhý scénář. Zkratka sk. znamená skupinu, do které byla daná patologie zařazena. Výskyt potom procentuálně vyjadřuje četnost zařazení do dané skupiny. Jestliže patologie nedosáhla alespoň 60% není zařazena do žádné ze skupin.

| k-means rozdělení | 1 | 2 | 3 | skupina | výskyt |
|------------------------------------|----|----|-----|---------|--------|
| A-P squeezing | 4 | 17 | 31 | - | - |
| A-P squeezing (mild) | 4 | 16 | 36 | 3 | 64,3% |
| A-P squeezing (moderate) | 4 | 14 | 19 | - | - |
| A-P squeezing (severe) | 0 | 8 | 11 | - | - |
| adductor spasmodic dysphonia | 1 | 4 | 14 | 3 | 73,7% |
| bowing | 1 | 6 | 4 | - | - |
| conversion dysphonia | 0 | 4 | 9 | 3 | 69,2% |
| cyst | 1 | 3 | 3 | - | - |
| erythema | 3 | 5 | 6 | - | - |
| gastric reflux | 1 | 12 | 35 | 3 | 72,3% |
| granulation tissue | 1 | 1 | 4 | 3 | 66,7% |
| hyperfunction | 23 | 89 | 151 | - | - |
| interarytenoid hyperplasia | 1 | 2 | 6 | 3 | 66,7% |
| keratosis / leukoplakia | 3 | 10 | 13 | - | - |
| laryngeal web | 1 | 4 | 2 | - | - |
| paralysis | 7 | 31 | 26 | - | - |
| paresis | 7 | 3 | 8 | - | - |
| polypoid degeneration (Reinke's) | 4 | 6 | 14 | - | - |
| post laryngoplasty | 2 | 1 | 3 | - | - |
| post thyroplasty | 2 | 4 | 3 | - | - |
| presbyphonia | 0 | 5 | 8 | 3 | 61,5% |
| scarring | 0 | 8 | 7 | - | - |
| ventricular compression | 1 | 6 | 9 | - | - |
| ventricular compression (full) | 1 | 4 | 3 | - | - |
| ventricular compression (mild) | 4 | 12 | 20 | - | - |
| ventricular compression (moderate) | 2 | 8 | 12 | - | - |
| ventricular compression (severe) | 2 | 9 | 12 | - | - |
| vocal fold edema | 1 | 10 | 33 | 3 | 75,0% |
| vocal fold polyp | 2 | 10 | 8 | - | - |
| vocal nodules | 2 | 4 | 13 | 3 | 68,4% |
| vocal tremor | 0 | 5 | 10 | 3 | 66,7% |

Tab. C.4: Zařazení patologií do skupin podle podobnosti vypočtených výsledků pro první scénář. Není zde uvedeno rozdělení ve skupině hodnocení variability základního tónu. V této skupině byly všechny patologie zařazeny do stejné skupiny viz Tab. C.1a Tab. C.2.

| Hodnocení úrovně šumu v hlase | | | Hodnocení pohyblivosti artikulačních orgánů | | |
|-------------------------------|--------------------|-----------|---|------------|-----------|
| 1. skupina | 2. skupina | 3.skupina | 1.skupina | 2. skupina | 3.skupina |
| bowing | granulation tissue | | A-P squeezing (mild) | | |
| ventricu. comp. (full) | post laryngoplasty | | A-P squeezing (moderate) | | |
| | vocal nodules | | A-P squeezing (severe) | | |
| | vocal tremor | | adductor spasmodic dysphonia | | |
| | | | conversion dysphonia | | |
| | | | gastric reflux | | |
| | | | interarytenoid hyperplasia | | |
| | | | polypoid degeneration (Reinke's) | | |
| | | | post thyroplasty | | |
| | | | ventricular compression (mild) | | |
| | | | ventricular compression (moderate) | | |
| | | | vocal fold edema | | |
| | | | vocal fold polyp | | |