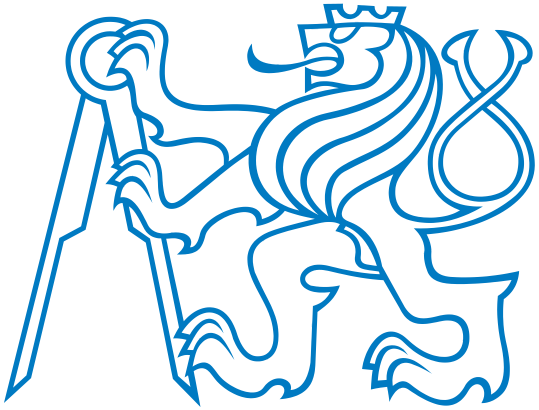
ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ
Fakulta Elektrotechnická



DIPLOMOVÁ PRÁCE

Automatická metoda hodnocení pauz v řeči u Parkinsonovy nemoci

2014

Jan Hlavnička

Poděkování

Na tomto místě bych chtěl poděkovat vedoucímu práce Ing. Janu Ruszovi, Ph.D. za jeho trpělivost, ochotu a vedení práce.

Prohlášení autora práce

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne ....

České vysoké učení technické v Praze
Fakulta elektrotechnická

Katedra teorie obvodů

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Jan H l a v n i č k a

Studijní program: Biomedicínské inženýrství a informatika (magisterský)

Obor: Biomedicínské inženýrství

Název tématu: Automatická metoda hodnocení pauz v řeči u Parkinsonovy nemoci

Pokyny pro vypracování:

1. Seznamte se s poruchami řeči u Parkinsonovy nemoci a možností detekce řečové aktivity.
2. Na základě dostupné literatury navrhněte algoritmus pro automatickou detekci pauz v řeči u Parkinsonovy nemoci. Pro návrh algoritmu využijte výpočetní prostředí MATLAB. Navržený algoritmus otestujte na vybraném vzorku řečových promluv zdravé populace a pacientů s výskytem Parkinsonovy nemoci. Výsledky porovnejte s konvenčním algoritmem pro detekci řečové aktivity u dysartrických pacientů.
3. Na bázi výstupu algoritmu navrhněte sadu vhodných řečových příznaků pro popis charakteristik dysartrie. Na dostupných datech proveďte experiment a pomocí navržených příznaků proveďte jednoduché statistické testy pro odlišení zdravých mluvčí od pacientů s Parkinsonovou nemocí.

Seznam odborné literatury:

[1] Duffy, J.R. (2005) Motor Speech Disorders and the Diagnosis of Neurologic Disease. 2nd ed., Mosby, New York.
[2] Skodda, S.; Schlegel, U. (2008) Speech rate and rhythm in Parkinson's disease. Mov Disord 23, 985-992.
[3] Rosen, K.; Murdoch, B.; Folker, J.; Vogel, A.; Cahill, L. et al. (2010) Automatic method of pause measurement for normal and dysarthric speech. Clin Linguist Phon 24, 141-154.
[4] Rusz, J.; Čmejla, R.; Růžičková, H.; Růžička, E. (2011) Quantitative acoustic measurements for characterisation of voice and speech disorders in early untreated Parkinson's disease. J Acoust Soc Am 129, 350-367.

Vedoucí diplomové práce: Ing. Jan Rusz, Ph.D.

Platnost zadání: do konce letního semestru 2014/2015

L.S.

prof. Ing. Pavel Sovka, CSc.
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 10. 1. 2014

Anotace

Tato práce si se zabývá metodou pro klasifikaci řečového signálu na třídy *řeč* a *pauza* u zdravé řeči a řeči postižené *hypokinetickou dysartrií* jako průvodní příznak Parkinsonovy nemoci. Účelem klasifikace je popsat fyziologické kvality řečového aparátu k produkci pauz. Pro ohodnocení této kvality byla navržena sada příznaků. Metoda předpokládá vyjádření řeči jako multimodální směsi normálních rozdělení parametrů výkonu, rozptylu autokorelační funkce, počtu průchodů nulou a melových kepstrálních koeficientů. Metoda spočívá v postupném odhadu parametrů jednotlivých složek *EM-algoritmem*. Pro zvýšení jakosti informace byly hranice segmentů určeny z hranic spektrálních změn *Autoregresního bayesovského detektoru změn*. Metoda je schopna detekovat respiraci jako samostatnou třídu řečového signálu. Metoda byla ohodnocena a porovnána s konvenční metodou a vykázala pro pauzy do $100ms$ o 20% vyšší úspěšnost než konvenční metoda a pro pauzy od $100ms$ o 10% vyšší úspěšnost než konvenční metoda, jak pro klasifikaci zdravé tak i dysartrické řeči. Nižší rychlost produkce pauz, pomalejší rytmus, zkracování pauz, zrychlování řeči i produkce pauz spolu s vyšším zastoupením respirace v pauzách byly nalezeny jako vhodné příznaky pro odlišení *hypokinetické dysartrie* od zdravé řeči.

This thesis aims on development of method for classification of the speech signal to *pause* and *speech* classes for a healthy speech as well as *hypokinetic dysarthria* in Parkinson's disease. The purpose of classification is to determine the physiological ability to produce pause using set of designed features. The proposed method suggests the speech as multimodal mixture including normal distributions of power of the signal, variance of the autocorrelation function, zero-crossings rate, and mel cepstral coefficients. The method sequentially estimates parameters of individual classes using the *EM-algorithm*. Segment boundaries were determined from the values of spectral changes obtained using the *Autoregressive bayesian changepoint detector* in order to improve the quality of classification. The method is able to detect respiration as a separate class of the speech signal. In comparison to conventional method for speech pause detection, our proposed algorithm reported 20% higher sucess rate for a pauses shorter than $50ms$ and 10% higher sucess rate fo a pauses longer than $100ms$. The lower rate of pause production ($p < 0.001$), slower rhythm, shortening pauses, accelerating the production of speech and pauses, along with a higher proportion of respiration in pauses are the possible features to differentiate *hypokinetic dysarthria* from healthy speech.

Seznam použitých zkratek

A Parametr rozptyl autokorelační funkce

AR Autoregresní model

E Parametr energie

f_0 Základní kmitočet hlasivek

KS Kontrolní skupina

MFCC Melovy kepstrální koeficienty

NHR Odstup harmonické a šumové složky signálu

P Parametr výkon

PN Parkinsonova nemoc

Rx Autokorelační funkce

SNR Odstup signál šum

SO Směrodatná odchylka

SPL Hladina hlasitosti

VAD Detektor řečové aktivity

Obsah

Kapitola 1

Úvod

1.1 Cíle práce

Tato diplomová práce si klade dva základní cíle, jejichž souvislost bude objasněna na následujících řádkách. Prvním cílem, jak již název tématu napovídá, bylo nalézt metodu pro ohodnocení pauz v řeči u Parkinsonovy nemoci. Ve stručnosti hledáme takový způsob klasifikace signálu na pauzu a řeč, který je schopný svým výstupem popsat řečovou aktivitu zatíženou projevy dysartrie u Parkinsonovy nemoci na fyziologické úrovni. Právě tím se tato metoda liší od konvenčních detektorů řečové aktivity, jež se snaží klasifikovat signál na řeč a pauzu pro účely přenosu informace. Maximální citlivost vůči signálu, který by mohl být projevem řečového aparátu, je žádaná vlastnost této metody oproti konvenčním detektorům řečové aktivity. Takový signál může postrádat informační hodnotu z hlediska řečového projevu a konvenční detektor řečové aktivity by jej vyhodnotil jako pauzu. Pro tuto diplomovou práci je takový signál naopak velice atraktivní, jelikož v něm jsou vyjádřeny drobné fyziologické odchylky funkce řečového aparátu, díky nimž můžeme splnit druhý základní cíl práce: návrh řečových příznaků pro popis charakteristik dysartrie u Parkinsonovy nemoci. Řečové příznaky tedy budou vycházet z časového průběhu klasifikace signálu na řeč a pauzu. Pro návrh příznaků vyjdeme z teoretických předpokladů pro projevy Parkinsonovy nemoci v řeči. Prezentovaná metoda umožňuje také detekci respirace, jejíž časový průběh také vezmeme v potaz a rozšíříme si naši paletu možných příznaků pro popis dysartrie o příznaky respirace.

Zmíněné dva cíle nesou pouze základní představu o náplni této práce. Tato práce samozřejme nesestává z popisu hledání těchto cílů, nýbrž je spíše interpretuje zasazené v širokém kontextu celé problematiky dysartrie. Poměrná část této diplomové spočívá právě v odhalení možných principů detekce pauz a její skloubení s projevy dysartrie. Abychom mohli posoudit prezentovaný algoritmus byl implementován také konvenční algoritmus pro detekci pauz v řeči postižené dysartrií. Výsledky obou těchto algoritmů porovnáme v ohodnocení na ručně klasifikované databázi zdravých a dysartrických mluvčí jak z hlediska výsledné úspešnosti, tak úspěšnosti na jednotlivých těchto skupinách mluvčí. Užitečnost navržených příznaků ověříme také na této databázi, ovšem pro na klasifikaci řečových signálů aplikujeme prezentovanou metodu. Výslednou kvalitu navržených příznaků ohodnotíme jednoduchými statistickými testy. Nyní vidíme, jak jsou oba základní cíle spolu úzce propojeny navzdory

zdánlivě odlišnému poli působnosti. Výsledná jakost navržených příznaků dysartrie totiž naprosto závisí na jakosti klasifikace navržené metody. Právě kvůli této vzájemnosti jsou oba cíle nedílnou součástí této diplomové práce.

1.2 Parkinsonova nemoc

Parkinsonova nemoc (dále jen *PN*) v idiopatické formě je neurologické onemocnění, jehož příčiny zatím nejsou známy. Dochází při něm k degeneraci dopaminergních neuronů v *substantia nigra - pars compacta*, jež vede k neurotransmiterové nerovnováze s nedostatkem dopaminu a zvýšenou hladinou acetylcholinu [1, 2]. Ta oslabuje funkci přenosu signálů v okruhu *basálních ganglií* [2], což zapřičiňuje charakteristické poruchy motoriky. Klinické symptomy se začínají projevovat při poklesu počtu dopaminergních neuronů v *substantia nigra* o $50-60\%$ a hladiny dopaminu ve *striatu* o $70-80\%$[1] [4, 5, 3]. V některých případech dochází k výskytu Lewyho tělísek v *substantia nigra* a v pozdějších fázích *PN* také v mozkové kůře [6].

Soudí se, že *PN* postihuje především věkovou skupinu starší 50 let. Některé studie uvádí střední nástupní věk okolo 60 let [7], nicméně odhaduje se, že $5-10\%$ pacientů postihne *PN* již v relativně mladém věku $21-40$ let [8] v závislosti především na genetických dispozicích[7]. *PN* je 2. nejčastějším neurodegenerativním onemocněním (po Alzheimerově chorobě [9]) s vysokou prevalencí okolo 0.3% v celé populaci [9] a 1.6% v populaci starší 65 let [10]. Odhaduje se, že v roce 2005 bylo diagnostikováno $4.1-4.6$ milionu pacientů, v roce 2030 má jejich počet vzrůst na $8.7-9.3$ milionu [11]. Pakliže nebude nalezena účinná léčba či prevence, bude narůstat procentuální zastoupení postižených *PN* v populaci - s ohledem na celkové stárnutí populace ve vyspělých zemích a s věkem následnou vyšší prevalencí [11, 10]. Doba dožití *PN* pacientů je značně individuální a bývá v rozmezí s nejnižší hranicí od $6-12$ let [12, 13, 9].

Účinná léčba *PN* zatím nebyla objevena. Terapie pacientů spočívá pouze ve zmírňování symptomů. Primárním cílem je vyrovnání neurotransmiterové nerovnováhy zvýšením hladiny dopaminu jeho metabolicky vhodnou formou *levodopa (L-DOPA)*, případně podáním *anticholinergika*. Další možností je *chirurgický zákrok* - užívaný především k odstranění ložisek tremoru (často lokalizovaný v *thalamu* nebo *globus pallidus* [14]) v případech, kdy medikace pozbývá účinnosti. [15] Pro potřeby tišení bolesti se užívá též *kraniální elektroterapie* [16].

Z hlediska postižení řízení svalového aparátu se *PN* projevuje *bradykinesií*[2], *rigiditou*[3] a *klidovým tremorem*[4]. Právě tyto pohybové příznaky jsou pro *PN* charakteristické a s progresí nemoci dochází k jejich výraznějším projevům. Mezi sekundární motorické příznaky lze zahrnout *hypomimii, dysartrii, úbytek posturálních reflexů, dysfagii, ptyalismus, mikrografii, charakteristickou parkinsonskou chůzi a postoj, dystomii, abnormální glaberální reflex,*

[1] podle jedné hypotézy je tento zajímavý nepoměr způsoben kauzálním zvýšením spotřeby dopaminu ve *striatu* [3], mechanismy *PN* jsou však stále předmětem výzkumů

[2] celkové zpomalení pohybu, u téměř poloviny pacientů *PN* postupně přechází v akinesii [17], v literatuře popisované jako „*freezing - zamrznutí*“

[3] stuhlost způsobená zvýšeným svalovým napětím, v této souvislosti bývá parkinsonismus popisován jako syndrom *hypokineticko-hypertonický*

[4] třes mimovolní, rytmický $4-6Hz$; projevuje se v klidové poloze, při pohybu mizí

blefarospasmus [17]. *PN* postihuje ne-motorický aparát *slábnutím kognitivních schopností*[5] *a neurobehaviorálními abnormalitami (obsedanstně-kompulsivní a impulsivní jednání), dysautonomiemi (ortostatická hypotense, konstipace, urogenitální dysfunkce), poruchami spánku, depresemi, apatií, anhedonií, únavou, halucinacemi* [17]. Pro včasnou diagnózu je nutné podchytit všechny klíčové projevy poruch motoriky v raném stádiu a k tomu může být vhodný právě řečový signál, tolik citlivý na přesnou souhru různých svalových skupin.

1.2.1 Postižení řečového aparátu

Proces řeči lze rozdělit do následujících vzájemně závislých dílčích subprocesů, z nichž každý může být *PN* individuálně ovlivněn [19]:

- **respirace** - postižení rigiditou dýchacího svalstva [20], snížení schopnosti volní kontroly dechu k syntaktickým, artikulačním a fonačním potřebám řečového projevu *PN* [21, 22]. Dostavuje se tachypnoe a soudí se, že *PN* ovlivňuje i kontrolu dechové automatiky - vliv mohou mít navíc i případné autonomní dysfunkce [23].

- **fonace** - zhoršení koordinace respiračních a laryngeálních svalů [21], tremor laryngeálních svalů hlasivkových chrupavek s výskytem fonačního tremoru [24], zhoršení schopnosti otevření hlasivek [25], nedovírání hlasivek [22], abnormální fáze uzavření hlasivek v průběhu fonace, nesymetrie fáze uzavření hlasivek [24], snížení poddajnosti hlasivek [14], celkově dysfonie [21, 26]

- **artikulace** - postižení především bradykinesií a rigiditou larynxu a pharinxu [27], bradykinesií jazyka a rtů [20] a rigiditou rtů [28].

- **prosodie** - komplexně do sebe zahrnuje respiraci, fonaci, artikulaci a navíc i neurologickou podstatu tvorby řeči - projev *PN* v hypokinetické dysartrii [29, 30].

Někteří autoři zmiňují respiraci jako úvodní poruchou řeči u *PN* [21] a vyšší náchylnost laryngeálního aparátu k patofyziologickým procesům *PN* než orálního aparátu [31, 32, 33]. Každopádně z výsledků většiny zde zmiňovaných studií lze usuzovat, že každý pacient může postupem *PN* vykazovat různý stupeň ovlivnění těchto subprocesů a projevovat tak individuální odchylku v řeči, která však bude u různých pacientů nabývat stejného charakteru typického pro pacienty *PN - hypokinetická dysartrie, monotónost projevu, zhoršení parametrů hlasu (jitter, shimmer, NHR, tremolo), snížení schopnosti artikulace a výsledné srozumitelnosti, snížení plynulosti řeči* [22, 32]. Uvádí se, že poruchy řeči se vyskytují již v raných stádiích. U neléčené *PN* u $70\% - 90\%$ postižených [32, 33].

Hypokinetická dysartrie

Vzhledem k vzájemnému ovlivnění subsystémů řeči, lze předpokládat, že prosodie, která do sebe zahrnuje příznaky všech ostatních subsystémů, bude nejčastěji postiženým subsystémem

[5] dle studie v průběhu 20 let rozvoje *PN* postihla demence 83% přeživších pacientů [18]

řeči. Vyústění všech řečových příznaků prosodie můžeme očekávat ve formě *hypokinetické dysartrie*. Odhaduje se, že k této poruše inklinuje až 90% pacientů[6] *PN* [34, 33, 32, 35, 36]. Hypokinetickou dysartrii lze vyjádřit v několika rozdílných rozměrech, především [37, 19, 38]:

- **hlasitost** - respirační subsystém, postižení *hypofonií*
- **intonaci** - fonační a respirační subsystém, postižení *monotónností*
- **rychlost a rytmus řeči** - zapojení respiračního, fonačního i artikulačního subsystému, typické zrychlování řeči na konci promluvy
- **produkce pauz** - účast respiračního, fonačního i artikulačního subsystému, produkce pauz se nepodřizuje obsahu promluvy - formě i kontextu

Jako opověď na medikaci může v souvislosti se zlepšením celkové motoriky pacienta docházet ke zlepšení fonace - rozšíření rozsahu intonace a zlepšení výslovnosti; kvalita hlasu a hlasitostní rozsah se mohou zlepšit bez korelace s vnějšími projevy rigydity a bradykinestie; zatímco plynulost řeči a časování fonace odpovídají na léčbu individuálně [34].

Produkce pauz v řečovém projevu *PN*-pacienta

Již James Parkinson si u některých *PN* pacientů ve své slavné *An Essay on the Shaking Palsy* povšiml charakteristicky postižené řeči, kterou popsal jako „*řeč zadrhávanou a přerušovanou*“ [39]. Je otázkou, jestli těmito příznaky nepopisoval právě postižení tvorby pauz, které se bude v této práci těšit velkému zájmu. Každopádně shrneme-li si předpokládaný vliv *PN* na řečový aparát, můžeme zkonstatovat, že od řečového signálu *PN* pacienta budeme očekávat [37]:

- **prodloužení pauz** vlivem zhoršeného řízení respirace a artikulace - *oddálení začátků slov* vlivem *bradykinesie jazyka* [22]
- **provedení formálně nevhodných pauz** vlivem nekontrolované respirace výskyt nekontrolovaných nádechů.
- **fonaci v průběhu formální pausy** vlivem zhoršené schopnosti otevření hlasivek - *krátké pausy nebudou dodrženy* nebo *budou znělé*
- **nekonstantní rychlost produkce pauz** - vyšší počet pauz na konci promluvy

Ve výsledném projevu se tedy zvýší četnot delších pauz, jak dokládají některé studie [22, 37, 14]. Z hlediska formálního členění řečového projevu by bylo vhodné klasifikovat pausy podle jejich funkce a přepokládané délky: Pro pausy uvnitř slov *(artikulační dělení)* se uvažuje v délce od $10ms$ [37]. Pro pausy mezi slovy - z fyziologického hlediska schopnosti uzavřít ústní dutinu „stop closure“ lze uvažovat délky od $50ms$ [14].

[6] tento údaj zanedbává vliv případné medikace

1.3 Detekce řečové aktivity

V přístupu k návrhu algoritmu pro detekci pauz lze zaujmout dva základní postoje: apriory předpokládáme vlastnosti řeči a tu detekujeme nebo předpokládáme vlastnosti pauzy a tu se pokusíme detekovat. Praktická konstrukce detektoru vyžaduje jistou kombinaci obou těchto přístupů, přičemž volba vlastností obou tříd odpovídá situaci i účelu jeho použití. Konvenční detektory řečové aktivity se zaměřují na takovou detekci, která minimalizuje redundantní složky signálu, tudíž i datový tok, a zároveň maximalizuje informační zisk, tedy řečovou aktivitu s ohledem na psychoakustiku. Tyto detektory bývají koncipovány s vyšší robustností vůči stacionárnímu i nestacionárnímu šumu a užívají poměrně složitá heuristická schémata rozhodování. Příkladnými zástupci jsou algoritmy detekce řečové aktivity standartizované *ITU (Mezinárodní telekomunikační unie)* a *ETSI (Evropský ústav pro telekomunikační normy)*. Algoritmus prezentovaný v této diplomové práci je oproti algoritmům pro detekci řečové aktivity zaměřen na postižení samotné funkce řečového aparátu z hlediska tvorby proluk v řeči. To vyžaduje odlišné přístupy k vlastnostem signálu i způsobu klasifikace. Neboť v případě signálů pacientů parkiunsonovi nemoci s nelze, s ohledem na všechny projevy související patologie řeči, vždy očekávat konkrétní vlastnosti řečového signálu a ani vlastnosti paus. Ty bývají často vyplněny hlasitou respirací, chrapotem a různými zvukovými artefakty (hlasité polikání, mlaskání). Pro uspokojení těchto potřeb není k dispozici žádný skutečně robustní detektor. Mnoho autorů pro tuto oblast výzkumu užívá časové náročnou ruční segmentaci. Někteří autoři vychází především z oscilogramu řečového signálu [37], jiní své rozhodování opírají o software pro analýzu řeči (TF32 [40]). Pro návrh metody lze do určité míry vyjít z principů detektorů řečové aktivity. Obsahem této kapitoly bude přiblížení základních parametrů popisu řečového signálu a pauz, jejich distribuce a možnosti rozhodování pro návrh metody.

1.3.1 Amplituda řečového signálu

Pokud si řečový signál představíme jako náhodný ergodický proces, můžeme se na něj podívat ze statistického hlediska a očekávat rozložení hustot pravděpodobnosti okamžitých amplitud signálu jaké je znázorněné na obrázku (1.1). Toto rozdělení lze aproximovat laplaceovým rozdělením dle vztahu (1.1). Kde μ značí střední hodnotu, je tedy lokačním parametrem, a b je parametrem měřítka. Pro řečový signál uvažujeme $\mu = 0$:

$$p(x; b, \mu) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}. \tag{1.1}$$

Dalším vhodným a často užívaným kandidátem pro aproximaci se jeví být gamma rozdělení určené vztahem (1.2). Kde α je parametrem tvaru, β parametrem měřítka a Γ gamma funkcí. Pro řeč se běžně volí tvar gamma rozdělení $\beta = 0.5$ [41]:

$$p(x; \alpha, \beta) = \frac{\beta^{-\alpha} x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}. \tag{1.2}$$

Pro detekci řečové aktivity lze předpokládat, že gamma rozdělení dobře aproximuje směs řečového signálu a signálu pauz. Naproti tomu laplaceovo rozdělení aproximuje dobře přede-

vším řečový signál [42]. V praxi se pro tyto situace spíše než okamžitou amplitudou popisuje signál jako její ortogonální transformace např. rychlou fourierovou transformací a diskrétní cosinovou transformací. Z jejich koeficientů se dále odhadují aktuální parametry rozdělení. Naproti tomu parametry distribucí pro řečový signál lze považovat v rámci promluvy za konstantní a pro jejich návrh lze vycházet ze standartních hodnot, které již byly předmětem mnohých studií. Samotné rozhodování může spočívat například v testování pravděpodobnostního rozdělení aktuálního vzorku signálu na rozdělení s parametry aproximujícími řečový signál např χ^2-testem [42]. Metoda detekce řečové aktivity založené na pravděpodobnostním rozložení signálu je vhodná především pro signály s delšími prolukami v řeči, které mohou být vyplněny například hudbou v pozadí. Tyto zarušené úseky mohou mít energii i harmonicitu podobnou lidskému hlasu avšak díky značně odlišnému pravděpodobnostnímu rozdělení je lze velmi snadno identifikovat. V případě signálů pro ohodnocení pauz budeme předpokládat vyšší SNR a pauzy vyplněné signálem šumového charakteru. Pro řečový signál lze v tomto případě předpokládat vyšší amplitudu. Problém nastává s rozpoznáním pauzy od řečového signálu s nízkou amplidudou reprezentovanou například frikativy. Na obrázku (1.2) můžeme pozorovat reálné rozložení hustot pravděpodobnosti na druhově rozmanitých typech normovaných úseků signálu s rozkmitem $< -1; 1 >$. Z distribuce je patrné, že rozložení hustot pravděpodobnosti delších řečových úseků v řádu jednotek sekund znázorněných na obrázku (1.1(černá)) a (1.2(zelená)) je dáno kombinací jasně odlišných distribucí harmonických (1.2(červená)) a šumových úseků (1.2(modrá, černá)). V signálu jsme tímto způsobem schopni odlišit znělé úseky od neznělých úseků, které se však svojí distribucí bohužel překrývají s distribucí šumu[7]. V případě delšího časového okna, které pokryje znělé i neznělé složky řeči by bylo lze možno použít tuto analýzu k detekci delších pauz (řádově od 200ms).

1.3.2 Energie a výkon řečového signálu

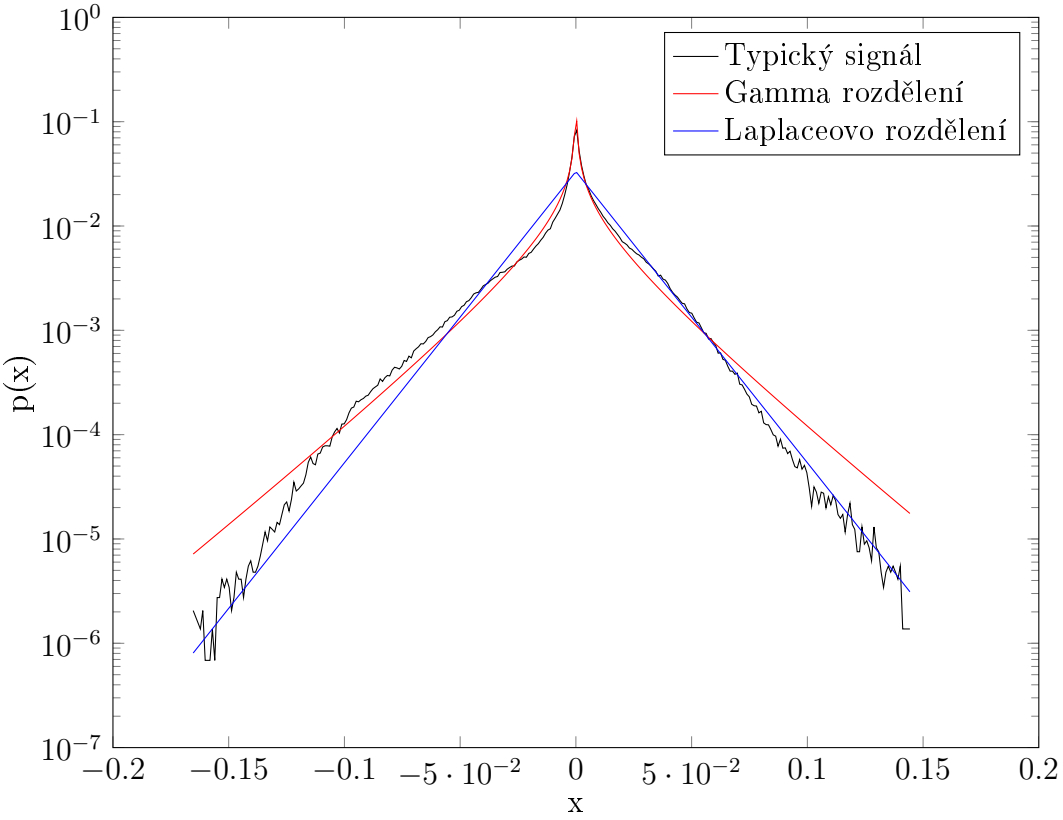
Energie signálu v diskrétním čase je definována podle vztahu (1.3). Energie je tedy sumou kvadrátu hodnot signálu a nese informaci o celém průběhu signálu. Z tohoto vztahu lze vyvodit energii pro úsek signálu podle vztahu (1.4):

$$E = \sum_{m=-\infty}^{\infty} x^2[m], \tag{1.3}$$

$$E_n = \sum_{m=n-N+1}^{n} x^2[m], \tag{1.4}$$

kde n značí úsek signálu a N jeho délku. S ohledem na různorodé způsoby tvorby fonémů řečovým aparátem je energie velice signifikantním parametrem pro úlohy klasifikace fonémů, rozpoznávání mluvčích a detekci řečové aktivity. Největší energii lze očekávat od vokalizovaných fonémů, především od samohlásek a nosovek. V případě znělých frikativů je energie

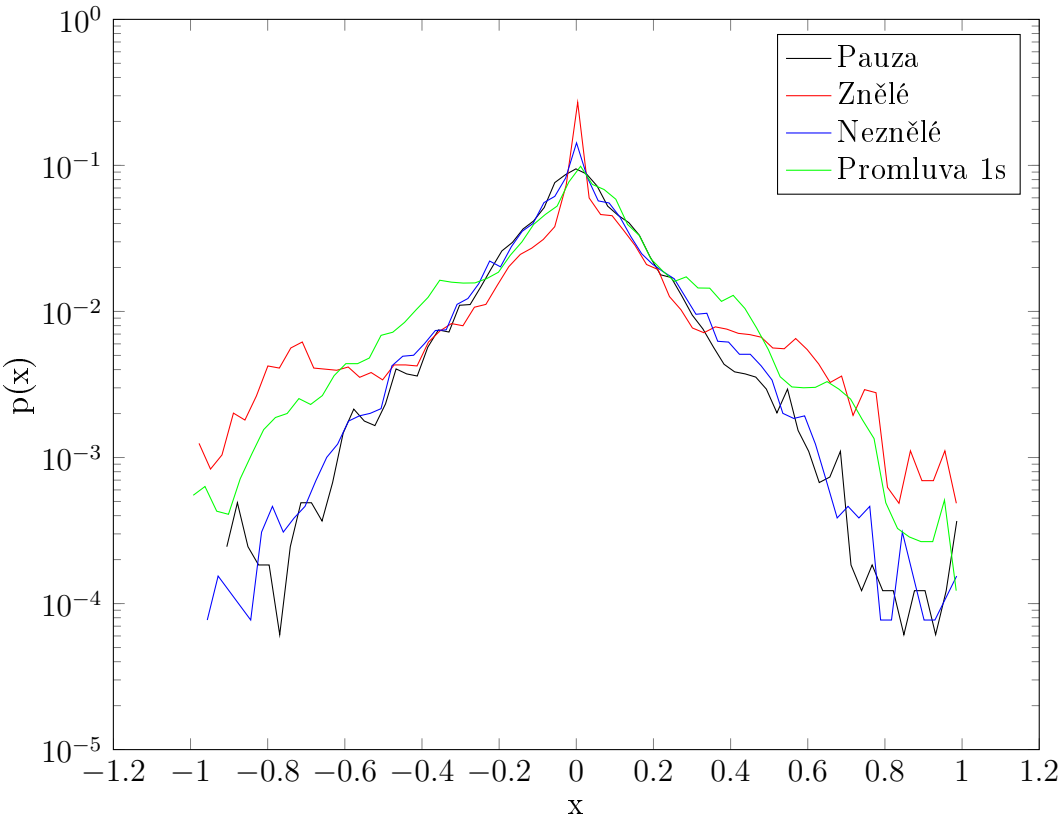
[7] při analýze distribuce koeficientů ortogonální transformace například diskrétní cosinové transformace dojdeme ke stejnému závěru



Obrázek 1.1: Hustota pravděpodobnosti amplitudy řečového signálu (černá) a její aproximace gamma rozdělením (červená) a laplaceovo rozdělením (modrá)

utlumena nižší excitací hlasivek a zmenšením rezonančního prostoru při sevřené ústní dutině pro potřeby vzniku turbulentního proudění, které je jejich charakteristickým znakem. Neznělé frikativy mají z fonémů nejnižší energii a při slabé artikulaci dochází ke značnému překryvu jejich energetického pásma s pásmem šumového pozadí. V úsecích pauz lze navíc očekávat výskyt respirace, jejíž energii nejsme schopni předem odhadnout. Energie neznělé respirace zasahuje do pásma neznělých frikativů. Znělá respirace patologické řeči se navíc může svojí energií blížit energii znělých frikativů. Energii celého řečového signálu si lze představit jako multimodální směs logaritmicko-normálních rozdělení ilustrovanou obrázkem (1.3). Jak je ze znázornění patrné, je energie rozlišujícím parametrem pro určení vokálů od neznělých frikativů. V případě signálů s dostatečným odstupen signálu od šumu lze vzít energii i jako výchozí parametr pro detekci řečové aktivity. Někteří autoři užívají energii dokonce jako primární parametr, který navíc klasifikují konstantním prahem [43] [8]. V této souvislosti je vhodné zavést veličinu výkon signálu, která je vyjádřena vztahem (1.5). Je zjevné, že se jedná o energii signálu vztaženou na jeho délku N. Opět můžeme délku signálu omezit na

[8] nutno podotknout, že efektivita takovéto klasifikace je silně zavislá na předem očakávané hodnotě odstupu signálu od šumu



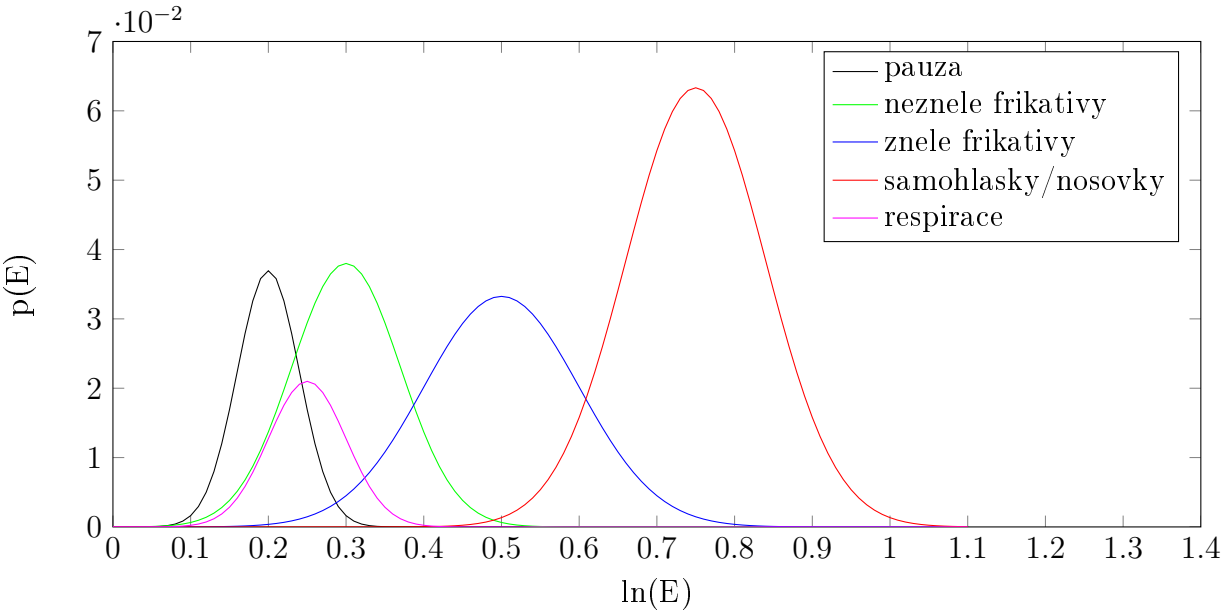
Obrázek 1.2: Hustota pravděpodobnosti amplitudy znělých (červená) a neznělých (modrá) úseků řečového signálu, úseku promluvy o délce 1s s výskytem znělých i neznělých fonací (zelená) a úsek pausy (černá)

úsek signálu n, jak znázorňuje vztah (1.6):

$$P = \lim_{N\to\infty} \frac{1}{N} \sum_{m=-\frac{N}{2}}^{\frac{N}{2}} x^2[m], \tag{1.5}$$

$$P_n = \frac{1}{N} \sum_{m=n-N+1}^{n} x^2[m]. \tag{1.6}$$

V případě totiž, že bychom energii klasifikovali konstantním prahem přes signály s odlišnou vzorkovací frekvencí nebo délkou segmentů, došlo by k naprostému selhání rozhodování. Pokud však budeme uvažovat v míře výkonu, osvobodíme se od této závislosti. A to nejen zavislosti mezi různými signály, ale i v rámci jednotlivých segmentů, užíváme-li při segmentaci proměnnou délku úseků. Obrázek (1.4) ilustruje hustotu pravděpodobnosti výkonu ručně segmentovaného signálu promluvy. Při pohledu na nesegmentovanou obálku signálu (zelená) si můžeme všimnout možné aproximace na bimodální směs normálních rozdělení. Parametry směsy lze samozřejmě velmi snadno odhadnout pomocí EM algoritmu a získali bychom



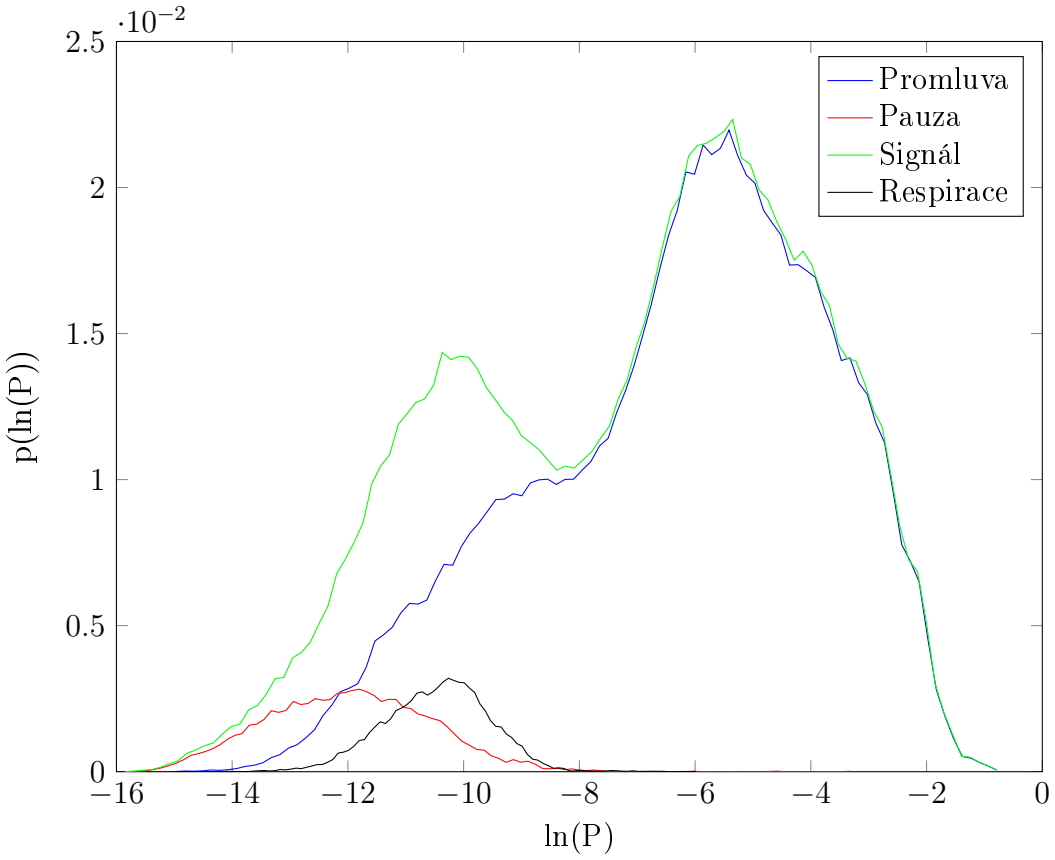
Obrázek 1.3: Hustota pravděpodobnosti energie řečového signálu reprezentovaná jako směs logaritmicko-normálních rozdělení pro složky samohlásky/nosovky (červená), znělé frikativy (modrá), neznělé frikativy (zelená), respiraci (fialová) a pausu (černá)

tak velice jednoduchý klasifikátor bez učitele. Nicméně při srovnání obrázků (1.3) a (1.4)) lze předpokládat vysokou chybu 1. a 2. druhu, zastoupenou především neznělými frikativy a respirací. Jednou z možností je upravit rozhodovací kritérium (viz. možnosti klasifikace) nebo modelovat směs více složkami. Užitečným se jeví popisovat energii jako bimodální směs pauzy a řeči pro jednotlivá frekvenční pásma, provést klasifikaci v rámci těchto pásem s ohledem na odstup signálu a šumu a celkové rozhodnutí o třídě určit hlasováním [44]. Obecně při větším odstupu řečové a šumové složky směsi bychom vždy dostáli zmenšení klasifikační chyby. Odstup složek je dán mody a rozptyly jednotlivých rozdělení. Při praktické aplikaci bychom chtěli mít složku pauzy s co nejnižším modem i rozptylem, o což se můžeme pokusit například pomocí *noise tracker*. Často se za příkladný vzorek šumového pozadí bere začátek nahrávky nebo úseky nahrávky s nízkou energií. Šum poté můžeme potlačit například *spectral subtraction* a z výsledného signálu odhadnout energetický práh [45]. V situacích, kdy předpokládáme proměnlivou hodnotu šumového pozadí, určuje se práh jako adaptivní kritérium z dlouhodobého pozorování. To můžeme realizovat pro ilustraci dle vztahů (1.7),(1.8):

$$E_p = \alpha E_d, \tag{1.7}$$

$$E_d^n = (1-p)E_d^{n-1} + pE, \tag{1.8}$$

kde E_p je hodnota prahu určená jako α násobek Ed (typicky $\alpha = 1.5$ [46]), aktuální hodnota E_d^n je určena z předešlé hodnoty E_d^{n-1}, energie aktuálního segmentu E a parametru $p \in (0.1)$, který určuje chování rovnice jako filtru 1. řádu [46, 47]. Pro zvýšení adaptability lze definovat parametr p jako proměnnou jejíž hodnotu určíme z poměru rozptylů předešlého a aktuálního



Obrázek 1.4: Hustota pravděpodobnosti výkonu ručně segmentovaného signálu promluvy (zelená) na její jednotlivé složky - řeč (modrá), respirace(černá) a pauza (červená)

segmentu $\beta = \frac{\sigma_{n-1}^2}{\sigma_n^2}$. Pro několik intervalů poměru β definujeme optimální hodnotu p [48]. Další variací na tento algoritmus je oddělit hodnotu prahu pro pausu a pro řečový signál a pro každou zvlášť provádět adaptaci přes odlišný parametr p, čímž lze předejít zákmitům na prahu rozhodování [48]. Poslední zajímavá možnost je definovat parametr p jako *scaling factor* λ podle vztahů (1.9) a (1.10):

$$E_p = (1 - \lambda)E_{max}^{n-1} + \lambda E_{min}^{n-1}, \tag{1.9}$$

$$\lambda = \frac{E_{max} - E_{min}}{E_{max}}, \tag{1.10}$$

kde E_{max}^{n-1} značí maximální a E_{min}^{n-1} minimální hodnotu předchozího segmentu. Z rovnice (1.10) je patrné, že *scaling factor* λ odprostí chování algoritmu od závislosti na předpokládaných mezích šumového pozadí [48].

1.3.3 Zero Crossing Rate

Vokalizované fonémy mají v řeči své nezastupitelné místo jak z hlediska vysoké četnosti tak i z hlediska informačního zisku. Jejich různorodost je určena možností mluvčího mě-

nit rezonanční vlastnosti ústní dutiny, která je buzena periodickými kmity hlasivek. Pokud bychom si kmity hlasivek zjednodušene představily jako harmonický signál a chtěli přibližně odhadnout jeho frekvenci, je nejsnažším řešením charakterizovat ho počtem průchodů jeho průběhu nulou. A právě k tomuto jednoduchému odhadu frekvence nám slouží parametr počet průchodů nulou *ZCR - zero crossing rate* definovaný vztahem (1.12), kde $sign(x[n])$ značí znaménkovou funkci vyjádřenou vtahem 1.14. *ZCR* je samozřejmě užitečný spíše pro úsek signálu ve vyjádření vztaženém na délku úseku jak naznačuje vzorec (1.13):

$$ZCR \propto \frac{2f_0}{f_s}, \tag{1.11}$$

$$ZCR = \sum_{m=-\infty}^{\infty} |sign(x[m]) - sign(x[m-1])|, \tag{1.12}$$

$$ZCR_n = \frac{1}{N} \sum_{m=n-N+1}^{n} |sign(x[m]) - sign(x[m-1])|, \tag{1.13}$$

$$sign(x[n]) = \begin{cases} 1, & x[n] \geq 0 \\ -1, & x[n] < 0 \end{cases}. \tag{1.14}$$

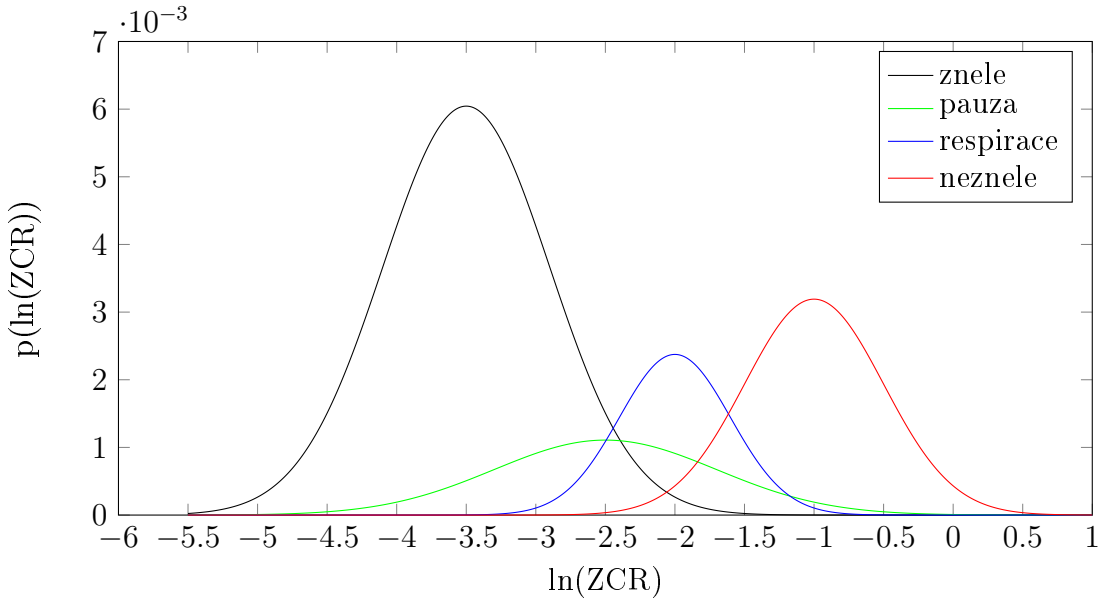
Mezi počtem průchodů nulou a základní frekvencí hlasivek f_0 platí přímá úměravztahu (1.11), nikoliv rovnost, protože znělý signál je tvořen množstvím vyšších harmonických, které mohou a většinou také mají vetší energii než základní frekvence f_0. V případě, že budeme mít signál neznělý, kupříkladu neznělý frikativ, odhadnutá hodnota *ZCR*, bude svojí hodnotou popisovat frekvence vyšších řádů, které bychom od znělého projevu neočekávali. *ZCR* pro signál pauzy se odvíjí od vlastností šumového pozadí a obvykle bývá vyšší a s větším rozptylem než pro znělou řeč. Hodnoty *ZCR* pro signál respirace nabývají hodnot nižších než neznělá řeč. Celou situaci lze popsat jako vícemodální směs logarytmicko-normálních rozdělení[9] znázorněnou pro ilustraci na obrázku (1.5). S ohledem na překryv jednotlivých složek směsi je nadmíru jasné, že samotná hodnota *ZCR* nám není schopna podat dostatek informací ke správné klasifikaci signálu na řeč a pauzu. Avšak pro klasifikaci na znělou a neznělou třídu je důležitým a často používaným parametrem. Vzhledem k vysoké citlivosti *ZCR* na stejnosměrnou složku i nízkofrekveční síťové rušení je velmi prozíravé v rámci preprocesingu tyto složky odstranit.

1.3.4 Autokorelace

Autokorelační funkce je nesmírně důležitým popisem vlastností řečového signálu. Obecně pro deterministický signál je definována dle vztahu (1.15). Pro stochastické signály můžeme její průběh odhadnout podle střední hodnoty (1.16):

$$R_x[k] = \sum_{n=-\infty}^{\infty} x[n]x[n+k], \tag{1.15}$$

[9] *ZCR* se běžně v logaritmickém měřítku nevyjadřuje, ale s ohledem na vysoký rozptyl *ZCR* pro neznělé složky je výhodné použít jeho škálu



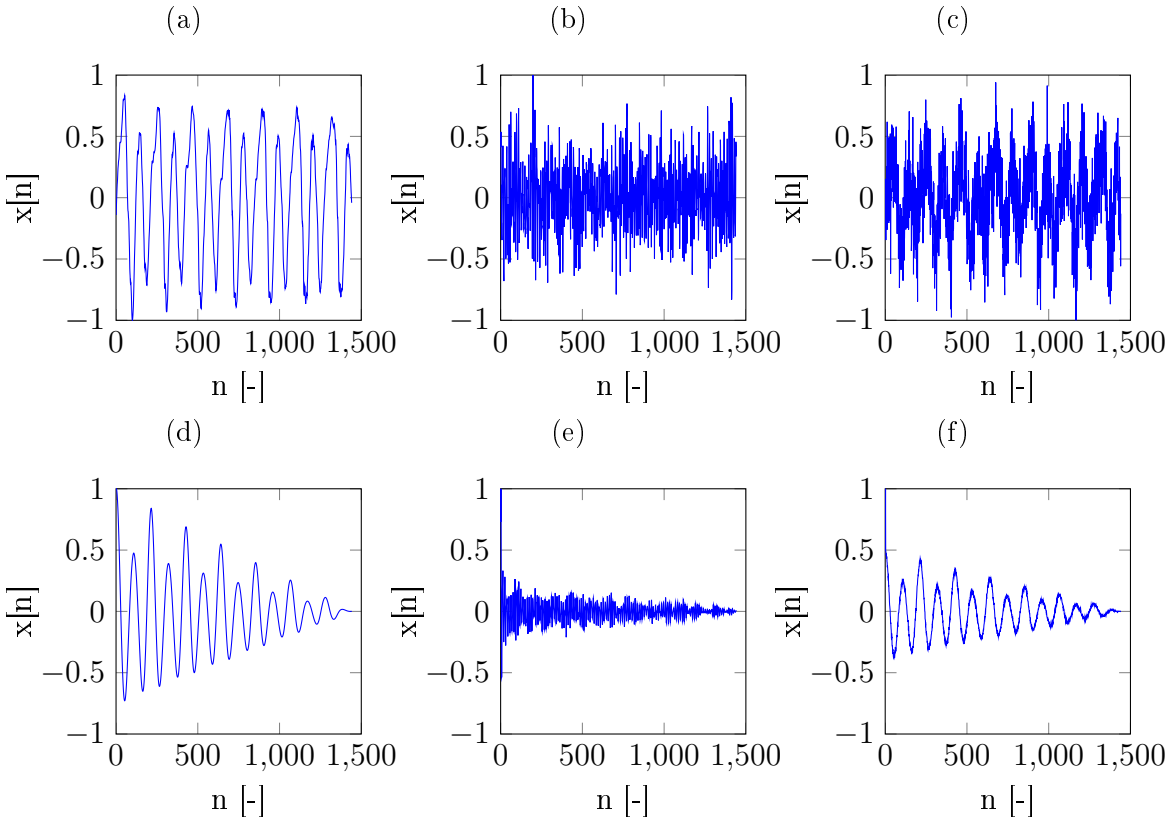
Obrázek 1.5: Hustota pravděpodobnosti ZCR pro znělé (černá) a neznělé (červená) složky řeči, respiraci (modrá) a pauzu(zelená)

$$R_x = \frac{1}{N}\sum_{n=1}^{N} x[n]x[n+k]. \tag{1.16}$$

Pro periodický signál s periodou p platí (1.17) - autokorelační funkce periodického signálu je také periodická o stejné periodě. Autokorelační funkce je sudá s maximem v $k = 0$, ve kterém vyjadřuje energii signálu či výkon signálu. Autokorelační funkce periodického signálu s periodou p nabývá svých lokálních maxim v $k = 0, \pm p, \pm 2p, ...$ nezávisle na fázi vstupního signálu:

$$R_x[k] = R_x[k+p]. \tag{1.17}$$

Tento výčet vlastností dává autokorelační funkci výsadní postavení pro odhad periodicity, periody, poměru signál šum a dalších parametrů sdělovacích a především také řečových signálů. Na obrázku (1.6) vidíme průběhy znělého signálu, neznělého signálu a znělého signálu se silnou složkou šumu a průběhy jejich normovaných autokorelačních funkcí. Z obrázku je jasně patrné, že v případě periodického signálu znělé řeči došlo ke zvýraznění základního harmonického průběhu a u neznělé řeči naopak k potlačení jeho šumového charakteru. Od ideálního šumu očekáváme, že bude korelovaný jen sám se sebou a jelikož je autokorelační funkce rovna součtu autokorelačních funkcí jednotlivých složek, dojde u zašuměného znělého signálu k potlačení jeho náhodné složky a zvýraznění harmonické složky. Je jasné, že těchto užitečných vlastností lze využít pro popis znělých a neznělých složek řeči. Výsledné průběhy autokorelačních funkcí by pro klasifikaci bylo vhodné popsat co nejjednodušším parametrem. To je možné například sumou (1.18) nebo odhadem střední hodnoty (1.19) kvadrátu autokorelačních koeficientů, čímž převedeme jejich záporné hodnoty do kladných souřadnic. Pro klasifikaci periodicity se nejužitečnějším zdá být vyjádřit autokorelační funkci rozptylem



Obrázek 1.6: Znázornění průběhů znělého signálu (a), neznělého signálu (b), znělého signálu s vysokým šumem (c) a průběhů jejich autokorelačních funkcí normovaných na jednotkový výkon (a→d),(b→e),(c→f)

σ^2 hodnot jejích koeficientů (1.20). V případě, že uvažujeme střední hodnotu autokorelační funkce blízkou nule $E(R_x) \to 0$ lze zjednodušeně odhadnout rozptyl autokorelačních koeficientů dle zmíněného vztahu (1.19):

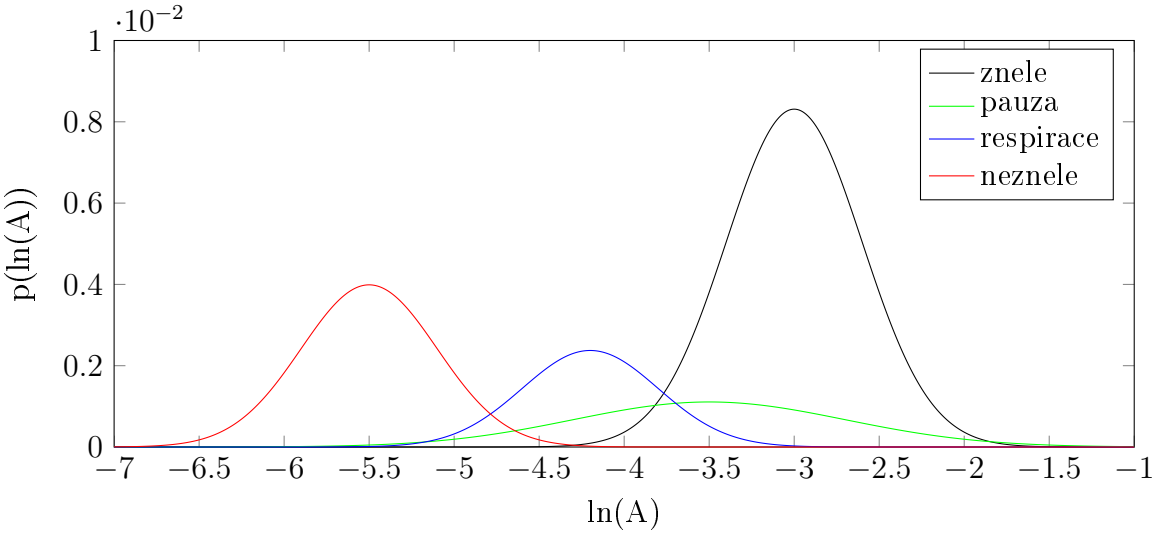
$$A = \sum_{k=1}^{N} R_x^2[k], \tag{1.18}$$

$$A = \frac{1}{N} \sum_{k=1}^{N} R_x^2[k], \tag{1.19}$$

$$A = \sigma^2(R_x[k]). \tag{1.20}$$

Parametrem A pro normovanou autokorelační funkci signálu můžeme řečový signál rozložit na vícemodální směs logaritmicko-normálních rozdělení ilustrovanou na obrázku (1.7). Nejvyšší rozptyl normované autokorelační funkce lze jistě nalézt především u znělé řeči a naopak nejnižší u neznělé řeči. Parametr A respirační složky uvažujeme u nižší než znělé a zároveň vyšší než neznělé, oblast výskytu je značně individuální, ale pohybuje se řádově v těchto mezích. Vlastnosti pauzy bohužel předem neznáme a i v případě jednoho signálu bývá rozptyl

hustoty pravděpodobnosti parametru A značně veliký. Tyto distribuce předurčují parametr A jako výchozí pro detekci znělých a neznělých složek řeči. Autokorelační funkci lze také aplikovat na signál rozdělený bankou filtrů *Subband Autocorrelation Function*, čímž lze velmi dobře popsat obálku spektra formantů pro zvýšení robustnosti vůči šumu při rozpoznávání slov [49] a *a* detekci řečové aktivity [50]



Obrázek 1.7: Hustota pravděpodobnosti výkonu parametru A normované autokorelační funkce pro znělé (černá) a neznělé (červená) složky řeči, respiraci (modrá) a pauzu (zelená)

1.3.5 Spektrum

Lidské vnímání zvuku je založené na frekvenční analýze zvuku a lidská řeč je právě tomu uzpůsobena, a proto využívá všech možností spektra pro přenos informace. V dlouhodobém spektru ilustrovaném na obrázku (1.8) je patrná převaha nižších frekvencí, což souvisí s vyšší energií vokalizovaných složek řeči. Lze rozeznat odlišnou artikulaci můžů i žen (srovnejme obrázek (1.8) modrá / červená), která je individuální i mezi jednotlivými mlučímu a sama o sobě může určit identitu mluvčího. Přes všechnu tuto přirozenou variabilitu je spektrum mocnou interpretací signálu pro klasifikaci na třídy pauza / řeč. Vlastnosti spektra řečového signálu užitečné pro detekci pauz lze shrnout následovně:

- **Entropie spektra** - popisuje organizovanost spektra. Řečové úseky a především výskyt formantů ve spektru indukuje nízkou entropii spektra, jejich absence naopak vysokou entropii spektra [51]. Toho lze využít k detekci koncových bodů při nízkém odstupu signál-šum a konečně také ke klasifikaci řečové aktivity [52]. Detektor založený na entopii spektra bude rubustní vůči nízkému odstupu signál-šum a především bude vykazovat vysokou odolnost vůči nestacionárnímu šumu [53]. Bohužel krátké pauzy patologické řeči (cca do $100ms$) velice často nesou informaci o formantech předchozího fonému, který sice již není vokalizován, ale turbulentní proudění sevřeného hrtanu může popsat prostor ústní dutiny podobnou spektrální obálkou a tudíž i nízkou entropií spektra.

- **Poměr výkonu v kmitočtových pásmech** - pokud si v signál rozdělíme na dvě kmitočtová pásma, z nichž jedno bude reprezentovat nízké kmitočty ($\approx 0 - 1kHz$) a druhé vysoké kmitočty ($\approx 6 - 8kHz$), lze poměrem jejich výkonů klasifikovat signál na souhlásku (těžiště výkonu v nižších kmitočtech), samohlásku (těžiště výkonu ve vyšších kmitočtech) a pauzu (oba výkony vyrovnané a nízké hodnoty). Prahové konstanty pro takovou klasifikaci bývají empiricky určené. Výsledný klasifikátor je jednoduchý na implementaci, jeho efektivita však velmi závisí na jakosti signálu.

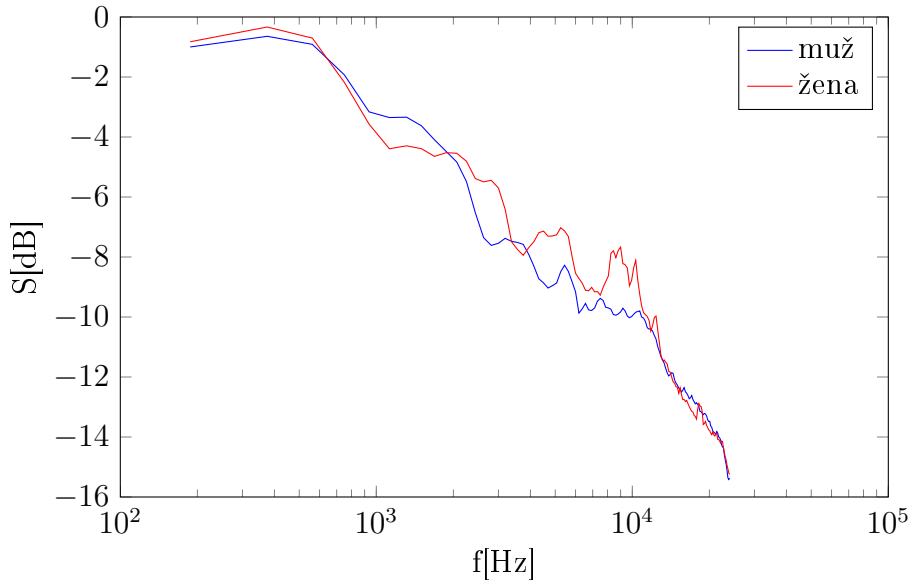
- **Podobnost spektra** - vzásadě se jedná o porovnání spektra neklasifikované signálu s roztříděnou bankou spekter. Ta může obsahovat jak exempláře řečových spekter, tak spekter různých šumových pozadí. Podobnost je určena spektrální vzdáleností, která může být pouhou euklidovskou vzdáleností nebo vzdáleností určenou *dynamic time warping*. Rozhodování probíhá buď jen na základě podobnosti spektra nebo jeho kombinací s dalšími příznaky (např. energie, *ZCR*) za použití komplexní rozhodovací logiky, kupříkladu fuzzy pravidel. Takový detektor je velmi robustní vůči výraznému a proměnnému šumovému pozadí [54].

- **Dynamika spektra** - lidská řeč je koncipována pro minimalizaci chybného sdělení tak, že za sebe řadí fonémy s výrazněji odlišným spektrem či budícím signálem. Lze tedy soudit, že úseky řečového signálu budou vykazovat vyšší variabilitu spektra v čase. To lze vyjádřit jako sníženou míru podobnosti spektra sousedících úseků signálu. Opět ji můžeme popsat pomocí prostých euklidovských vzdáleností či delta a delta-delta koeficienty melových frekvenčních kepstrálních koeficientů *MFCC*. Parametr dynamiky je velice robustní vůči šumovému pozadí [55] samozřejmě za předpokladu jeho stacionarity v rámci sousedících vzorků. Již z principu se nehodí pro detekci krátkých pauz, které mají délku srovnatelnou s fonémy.

1.3.6 Možnosti klasifikace

Pro samotný akt rozhodování do něhož spadá nejen rozdělení signálu do tříd, ale i vyhlazení rozhodování *decision smoothing*, využívají konvenční detektory mnoha různých, praxí ověřených algoritmů (bayesovské rozhodování, *SVM - support vector machine*, *HMM* - skryté markovovy modely, *MMC -maximum margin clustering*). Jak lze z předešlého popisu parametrů řečového sginálu vytušit, bude v této práci řečový signál pro účely rozhodování modelován jako multimodální gausovská směs. K odhadu parametrů jejích složek je praktické použít *EM*-algoritmus, který je součástí souboru standartních funkcí výpočetního prostředí MATLAB. Odhad směsí řečového signálu *EM*-algoritmem však vyžaduje specifický přístup k datům - vzhledem k tomu, že ne vždy předem známe jak počet složek směsi tak i jejich domnělé třídy. Najít vhodnou metodiku pro třídění signálu jako gausovské směsi je největším oříškem této rozhodovací procedury. Pro řečový signál budeme uvažovat následující předpoklady:

- Signály jednotlivých tříd považujeme za nezávislé náhodné vektory



Obrázek 1.8: Spektrální výkonová hustota promluvy o délce 30s pro jednoho zástupce mužského (modrá) a ženského (červená) pohlaví

- Signál pauzy předpokládáme jako směs stacionárního šumu produkovaného prostředím a řetězcem záznamu zvuku a nestacionárním šumem produkovaného samotným pacientem, zastoupený hlasitou respirací a neřečovými artefakty promluvy (např. mlaskání)

- Rozdělení jednotlivých parametrů budeme v rámci příslušné třídy aproximovat normálním rozdělením

- Pro vyhlazení rozhodování na základě kontextu budeme předpokládat, že jazyk promluvy spadá do indoevropské jazykové skupiny

Za těchto podmínek tedy budeme *EM*-algoritmem odhadovat parametry jednotlivých směsí, z nichž určíme posteriorní pravděpodobnosti příslušnosti k třídě. K vhodnému zvolení třídy se nám nabízí více řešení. Konvenční *VAD* provádí rozhodování na základě testování nulové hypotézy příslušnosti úseku signálu x k třídě (2.8, 2.8):

$$H_0 : x \in promluva, \tag{1.21}$$

$$H_1 : x \in pauza, \tag{1.22}$$

pomocí např. *LRT - likelihood ratio test* či *UMP - uniformly most powerful test*. Možný způsob rozhodování na základě znalosti parametrů bimodální gausovské směsy si ukážeme na následujícím příkladu. Mějme obecnou bimodální distribuci příznaku $P(X)$, kterou lze vyjádřit jako směs distribuce pauzy $P(X|pauza)$ a distribuce promluvy $P(X|promluva)$. Rozhodování mezi příslušnosti k třídě můžeme vyjádřit diskriminantem $D(X)$ (1.23). Dis-

kriminant lze definovat jako rozdíl věrohodností obou tříd $L(X)$ posunutých o práh T (1.24):

$$x \in \begin{cases} promluva, & D(x) \geq 0 \\ pauza, & D(x) < 0 \end{cases}, \tag{1.23}$$

$$D(X) = L(X) + T, \tag{1.24}$$

$$L(X) = \ln(\frac{P(X|promluva)}{P(X|pauza)}), \tag{1.25}$$

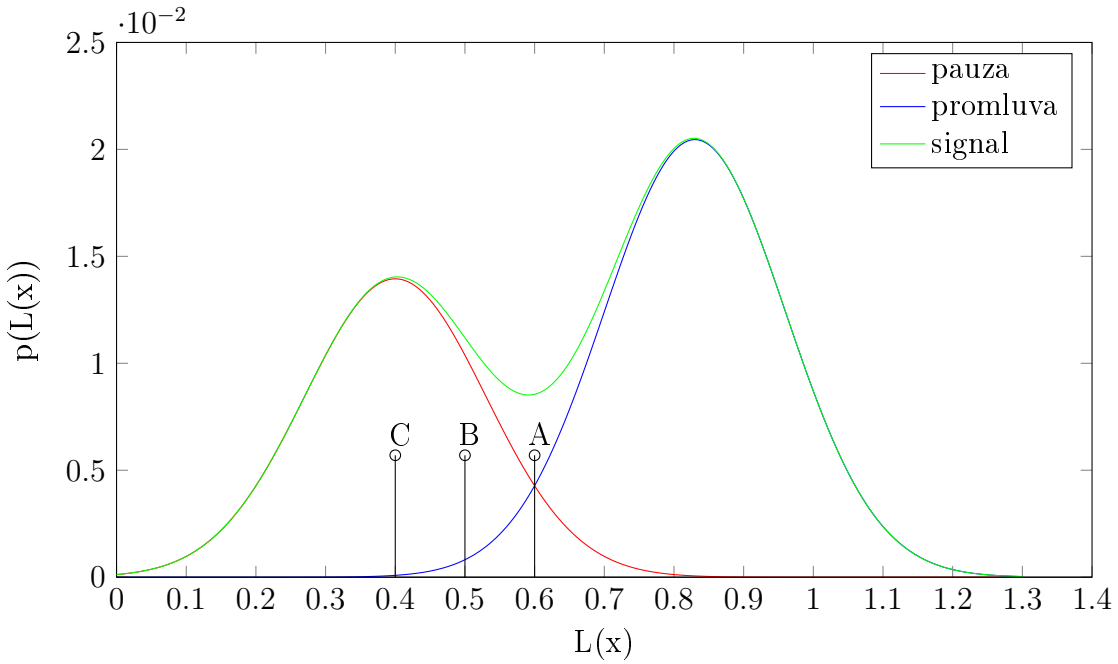
$$T = \ln(\frac{P(promluva)}{P(pauza)}), \tag{1.26}$$

$$T = \frac{P(promluva)}{P(pauza)}. \tag{1.27}$$

Pro optimální rozhodování se používá definice prahu dle (1.26), kterým v podstatě vážíme posteriorní pravděpodobnosti na pravděpodobnost výskytu třídy. Pokud by však cena za ztrátu užitečného signálu řeči byla vyšší než za přijetí redundantní pauzy, což u *VAD* pro sdělovací účely není překvapivé, jeví se být zajímavým a užitečným definovat práh například dle vztahu (1.27) [56]. Nabízí se také možnost rozhodovat se na základě minimálního prahové hodnoty k preferované třídy vztahem (1.28):

$$x \in promluva, P(x|promluva) > k, \tag{1.28}$$

to lze však aplikovat pouze za předpokladu dostatečného odstupu obou tříd. Tyto základní cesty rozhodování jsou ilustrovány na obrázku (1.9).



Obrázek 1.9: Hustota pravděpodobnosti bimodální směsy (zelená) normálního rozdělení pauzy (červená) a řeči (modrá) s vyznačenými rozhodovacími prahy: A) dle vztahu (1.26), B) dle vztahu (1.27), C) dle vztahu (1.28) pro $k \to 0$

Kapitola 2

Metody

2.1 Databáze

Pro analýzu příznaků i pro testování účinnosti metody budeme vycházet ze stejné databáze tvořené *kontrolní skupinou* (dále jen *KS*) a *skupinou PN*. Databáze je tvořena pouze rodilými mluvčími *českého jazyka*. Skupina *KS* je tvořena 23 mluvčími z nichž je 11 žen a 12 mužů. Skupina 22 *PN pacientů* sestává z 11 žen a 11 mužů. Věková skupina všech mluvčích je v rozsahu 50 až 80 let. Základní údaje o skupině *PN* uvádí tabulka 2.1. Pro popis byla zvolena standartní stupnice *Hoehn and Yahr* a *UPDRS III - Unified Parkinson's Disease Rating Scale*. Stupnice *Hoehn and Yahr* popisuje úroveň symptomů *PN* v pěti stupních. Stupeň 1 značí jednostranné příznaky onemocnění, stupeň 2 oboustranné postižení bez poruch rovnováhy, stupeň 3 mírné až středně těžké oboustranné postižení s postrální nestabilitou, stupeň 4 těžké postižení se schopností chůze a stání bez podpory, stupeň 5 imobilita - upoutání na lůžku nebo vozík, stání a pohyb pouze s asistencí. Nahrávky se pohybují v rozsahu $< 1; 2.5 >$, tedy v nižších stupních postižení *PN*. Stádia poruch motoriky jsou u skupiny popsána ve škále *UPDRS III*, jejíž rozsah je od pro $PN < 1; 108 >$. *UPDRS III* se zaměřuje na popis jednotlivých poruch motorik e.g. *tremor, rigidita, bradykinese, mimika, posturální stabilita* i pro nás důležitého postižení *řeči*. Postižení řeč je ve škále *UPDRS III* vystižena následovně: stupeň 0 v normě, stupeň 1 mírné ztráta exprese, dikce a hlasitosti, stupeň 2 monotóní, nezřetelná, přesto srozumitelná řeč, střední poškození, stupeň 3 výrazné poškození, těžko srozumitelný projev, stupeň 4 nesrozumitelná řeč.

Obsahem promluvy všech mluvčích je následující úryvek od spisovatele Karla Čapka: *„Když člověk poprvé vsadí do země sazeničku, chodí se na ni dívat třikrát denně: tak co, povyrostla už, nebo ne? I tají dech, naklání se nad ní, přitlačí trochu půdu u jejích kořínků, načechrává jí lístky a vůbec ji obtěžuje různým konáním, které považuje za užitečnou péči. A když se sazenička přesto ujme a roste jako z vody, tu člověk žasne nad tímto divem přírody, má pocit čehosi jako zázraku a považuje to vůbec za jeden ze svých největších osobních úspěchů.“* [57]

	Průměr	SO	Rozsah
Věk	65	9.7	48-82
Doba onemocnění	9.3	5.5	1-20
Hoehn und Yahr	2	0.49	1-2.5
UPDRS III	15.9	7.6	6-34
UPDRS III - řeč	0.72	0.7	0-2

Tabulka 2.1: Popis skupiny *PN* v databázi pro jednotlivé charakteristiky, průměrem, směrodatnou odchylkou *SO* a rozsahem hodnot

2.1.1 Ruční značení

Pro objektivní ohodnocení prezentovanéí metody a pro její srovnání s metodou konvenční bylo nutné vztáhnout výsledek na nezávisle klasifikované značky. Klasifikace byla provedena ručně pro třídy *pauza, respirace* podle stejných kritérií. Třída *pauz* byla navíc klasifikována na 3 stupně. Stupeň 1 obsahuje dlouhé formální pauzy mezi slovy a zahrnuje také *respiraci*. Stupeň 2 obsahuje krátké pauzy uvnitř slov typicky vzniká před artikulací *explosivu*. Stupeň 3 slouží pro zachycení krátkých pauz třídy 2, které nelze považovat za promluvu a v jejich signálu lze detekovat f_0 o velmi nízké energii. Tedy vykazují zapojení hlasového aparátu. Proto pro ohodnocení nebudeme uvažovat třídu 3, ale je vhodné ji mít označenu pro potřeby ladění algoritmu. Počáteční hranice pauzy pro znělé fonémy byla značena v časovém bodě takovém, kdy nad harmonickou složkou fonému převáží šumová složka. Tento bod zlomu lze nalézt jednak v oscilogramu signálu, jednak ve spektrogramu jako snížení energie formantů - zploštění spektra. Pokud po fonému následuje dlouhá pauza, pak mluvčí často ztiší hlas bez tohoto příznaku. V těchto případech byla hranice pauzy stanovena na nejzazší časové okolí s výskytem f_0. Za konec neznělých fonémů byla taková uvažována hranice ve spektru, za kterou již nebylo spektrum homogenní se spektrem předešlého fonému. Za koncovou hranici pauzy byl brán počátek fonému - pro *znělé fonémy* až ten úsek, který vykazoval harmonický signál v amplitudě a formanty ve spektru. *Hranice neznělý fonémů* v oscilogramu často splývají s hranicemi *pauzy*, proto byl počátek určen jako hranice ve spektrogramu, za níž lze zřetelně odlišit foném od pauzy. Ze všech výše uvedených kritérií byla vždy napřed dávána přednost informaci ze spektrogramu. Označeny byly pouze pauzy delší než $40ms$. Kratší pauzy bylo obtížné odhalit jak z hlediska jejich výskytu, tak i z hlediska jejich neostrých hranic. *Respirace* byla určována pouze ze spektogramu vždy v nejširších hranicích, za kterými ji lze ještě jednoznačně identifikovat.

2.2 Návrh metody

Připomeňme, že prezentovaná automatická metoda ohodnocení pauz v řeči u *PN* předpokládá vícemodální logaritmicko-normální rozdělení parametrů *rozptyl normované autokorelační funkce A, výkon signálu P a počet průchodů nulou ZCR* tříd *znělá řeč, neznělá řeč, pauza a respirace*. Pro popis spektrálních parametrů se jeví být užitečnou doménou *melovy kepstrální koeficienty*, jejichž charakteristika byla s ohledem na jejich multidimensionalitu

zařazena až do samostané podkapitoly o *prostorech parametrů*. K odhadu parametrů jednotlivých tříd využijeme *EM-algoritmu* a rozhodování o příslušnosti k třídě se bude řídit dle prahu (1.26) znázorněném na obrázku (1.9(a)) [1]. Nyní se zaměřme na způsob výpočtu samotných parametrů. Pro každý úsek signálu délky N byly parametry určeny následovně:

- **Výkon signálu** vychází z definice (1.4). Vstupní úsek signálu bude váhován *hammingovým* oknem, které potlačí hodnoty signálu na jeho začátku i konci a tím vlastně zredukuje redundantní informaci z rozhraní dvou segmentů [2].

- **Rozptyl normované autokorelační funkce** - výpočet byl uskutečněn jako odhad rozptylu normované autokorelační funkce dle vztahu (1.16). Pro vstupní úsek signálu byl váhován obdélníkovým oknem, aby nedošlo ke snížení citlivosti vlivem potlačení okrajové složky.

- **Počet průchodů nulou** určuje definice (1.13). Jelikož parametr vykazoval značný rozptyl hodnot pro skupinu znělých fonémů a u mnohých z mluvčí se tato skupina vyskytovala jako výrazně bimodální gausovská směs znělých samohlásek a znělých souhlásek, což vedlo k nepředvídatelné konvergenci *EM-algoritmu*, byla provedena kompenzace parametru v následující podobě: Počet průchodů nulou byl určen z průběhu normované autokorelační funkce vstupního úseku signálu váhovaném hammingovým oknem. Toto nekonvenční řešení má za následsledek snížení počtu průchodů nulou pro znělé souhlásky, jelikož autokorelační funkce potlačí jejich šumovou složku (pro ilustraci obrázek (1.6(f))). Obě výsledná rozložení znělých samohlásek i znělých souhlásek budou mít menší rozptyl i menší vzájemnou vzdálenost modů. Díky tomu může *EM-algoritmus* nejen lépe odhadnout znělé fonémy jako jeden celek, ale také lépe odlišit jejich hranice od ostatních tříd. Tato kompenzace bohužel ječtě zvyšuje citlivost počtu průchodů nulou vůči harmonickému rušení napájecí sítě. Proto bylo nutné v rámci *preprocesingu* tyto složky odstranit filtrem typu dolní propust s mezním kmitočtem $> 50Hz$.

- **Melovy kepstrální koeficienty** vypočteme jako diskrétní kosinovou transformaci logaritmu spektrální výkonové hustoty filtrované melovou bankou filtrů, o níž se blíže zmíníme v samostatné podkapitole o prostoru *MFCC*. Banku navrhneme s 24 filtry a pracovat budeme s prvými 8 melovými kepstrálními koeficienty. Vstupní úsek signálu budeme váhovat *hammingovým oknem*, které podobně jako pro předchozí parametry sníží váhu signálu na rozhraních, ale také především omezí prosakování spektra.

Při návrhu algoritmu bylo nutné položit si otázku: Které parametry a jak použít pro maximálně spolehlivý odhad pauz. Rozložení amplitudy bylo nutné zamítnout, jelikož sama o sobě nepodává dostatek informací k robustní klasifikaci. Navíc její případná kombinace s ostatními parametry by vedla k složité rozhodovací úloze. K řešení vedlo mnoho cest, které

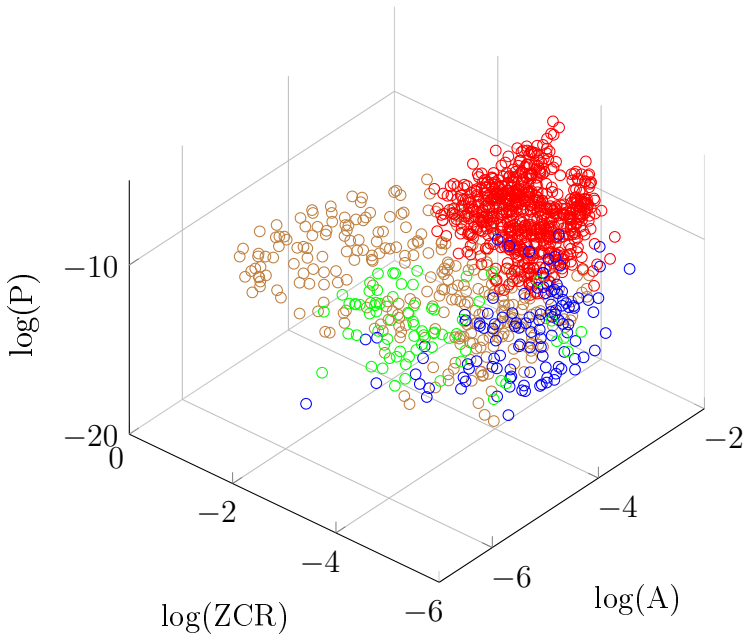
[1] rozhodování s prahem (1.26) projevuje svoji silnou stránku především v situací výrazného nepoměru a překryvu klasifikovaných rozdělení - tomu se v prezentované metodě pokusíme všemi možnými způsoby předejít a proto je toto rozhodování v naší metodě neužitečné

[2] to je nesmírně výhodné pokud užíváme pružnou segmentaci, na kterou přijde slovo později

se však v mnoha případech ukázaly jako nespolehlivé. Veškeré vhodné způsoby nakonec vyustily v konstrukci dvou základních parametrických prostorů: prostoru *(P, A, ZCR)* a prostoru *(MFCC)*. Jejich přiblížení bude náplní následujících dvou podkapitol.

2.2.1 Prostor parametrů

Vezmeme-li v potaz rozdělení parametrů (výkon signálu P, rozptyl autokorelační funkce A a počet průchodů nulou ZCR) v rámci tříd *(znělá řeč, neznělá řeč, pauza, respirace)* doložené ilustracemi (1.3,1.7,1.5) z předchozího oddílu, lze zkonstatovat, že jednotlivé parametry nesou vzájemě odlišné a zajímavé informace o jednotlivých třídách. Na základě jednoho parametru by klasifikace dosahovala značné chyby. Promítneme-li však všechny tři zmíněné parametry do společného prostoru, dojde k razantnímu zvýšení odstupu jednotlivých tříd. Celá situace je rozvržena na obrázku (2.1). Pro znělé fonémy uvažujeme nejvyšší výkon i



Obrázek 2.1: Znázornění skutečného řečového signálu v prostoru *(P, A, ZCR)* s barevně odlišenými třídami: pauza (modrá), dech (zelená), znělá řeč (červená), neznělá řeč (hnědá)

rozptyl autokorelační funkce a relativně nízký počet průchodů nulou. Pro pauzu dosahuje rozptyl autokorelační funkce i počet průchodů nulou podobných řádů. Respirace se od pauzy liší především vyšším počtem průchodů nulou. Tyto dvě skupiny mají ale zároveň mnohem nižší výkon vůči znělým fonémům. Neznělé fonémy mají naopak podobné umístění v parametrech rozptylu autokorelační funkce, ale vyšší počet průchodů nulou i výkon je znatelně prostorově diferencuje. Zaměřme se na obrázku (2.1) na shluk znělých fonémů (červená). Svojí hustotou je nejvýraznějším celkem ve zkoumaném prostoru, je silně izolovaný od svého okolí a obvykle tvořený dvěma základními shluky. Prvá jeho část je tvořena pouze četnými znělými samohláskami a druhá sestává ze všech ostatních znělých fonémů (e.g. znělé frikativy,

znělé explosivy). Samozřejmě se může přihodit, že některý neznělý frikativ bude těmito parametry náležet do této skupiny znělých fonémů. To se stane pro úseky signálu, kdy mluvčí již započal artikulovat neznělý frikativ a zároveň jeho hlasivky ještě nedotišily vokalizaci předešlého fonému. Tuto klasifikační odchylku zanedbáme, jelikož je obtížné rozsoudit, do které ze skupin by měla patřit a ponecháme jí tu třídu, do které náleží v rámci svého umístění v prostoru *(P, A,ZCR)*. Shluk neznělých fonémů (hnědá) má velmi vysoký rozptyl v našem prostoru. Zajímavá je jeho osamělá část s vysokým počtem průchodů nulou zastoupená hlavně sykavkami. Pro ostatní neznělé fonémy dochází k silnému překryvu zejména s parametry respirace. Polohu ani rozptyl množiny respirace nelze v tomto prostoru předem určit a velmi se liší mluvčí od mluvčího. Pokud bychom si odmysleli barevné odlišení skupin, tak vzhledem k překryvu našich tříd se mohou vyskytovat výrazné neklasifikovatelné shluky, zapřičiněné zejména respirací. Přes tuto variabilitu je prostor *(P, A, ZCR)* užitečnou interpretací řečového signálu.

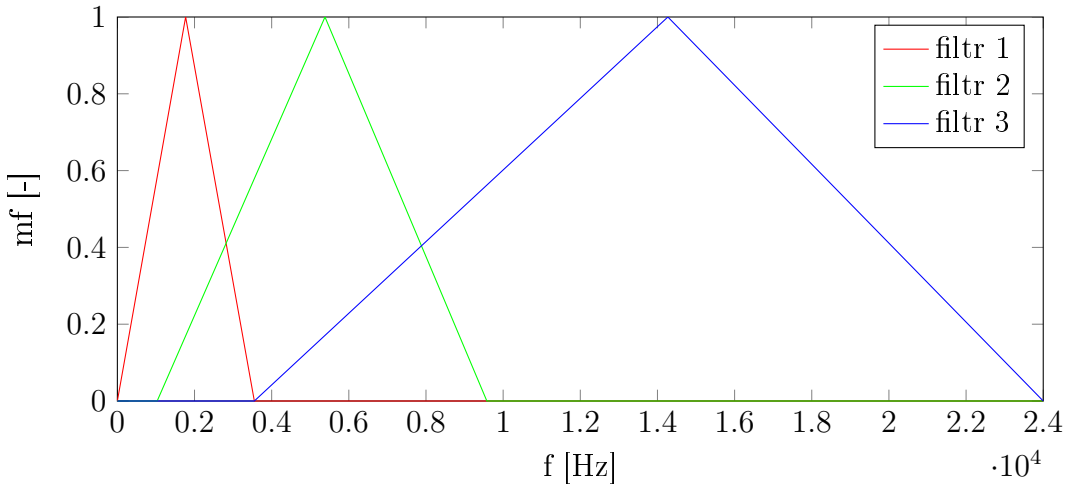
2.2.2 Prostor spektra

Na obrázku (2.2) vidíme průběh *melovy banky filrů* pro 3 filtry. Tento nízký počet uvažujme pouze usnadnění ilustrace, v praxi se používá obvykle vyšší počet e.g. 24. Všimněme si, že směrem k vyšším kmitočtům dochází k rozšiřování pásem jednotlivých filtrů. Jednotlivá pásma jsou definována podle *melovy frekvenční škály*, pro jejíž vyjádření se obvykle vychází ze vztahu 2.1:

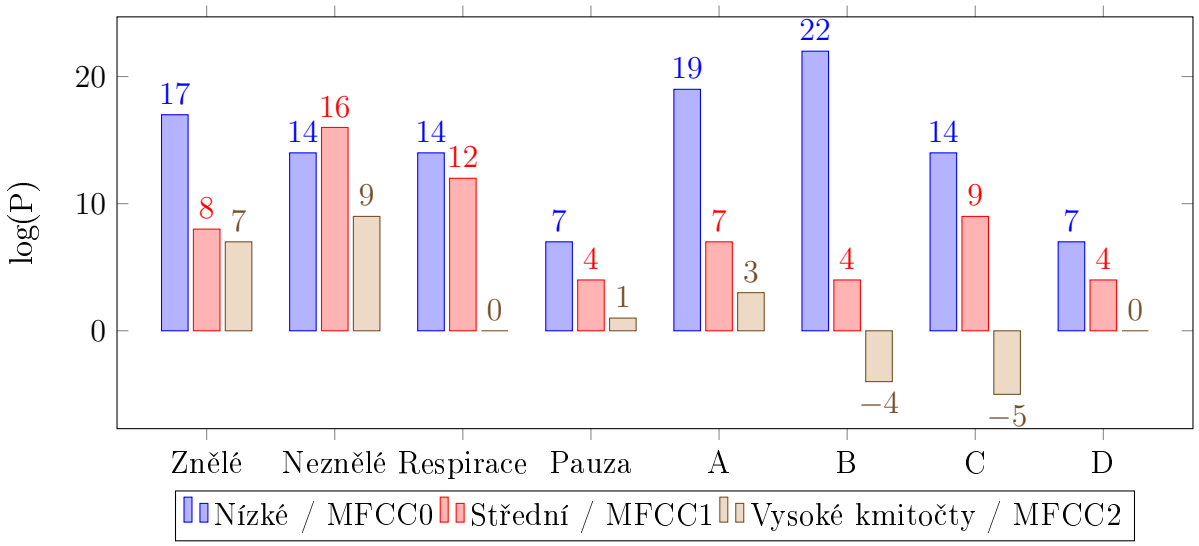
$$f_{mel} = 2595.log_{10}(1 + \frac{f}{700}), \tag{2.1}$$

kde f značí vstupní frekvenci v $[Hz]$a f_{mel} výstupní frekvenci v $[mel]$: Díky tomu, že je *melova škála* logaritmická, nemusíme nijak kompenzovat větší rozptyl parametrů u vysokofrekvenčních složek e.g. neznělých hlásek, jako tomu je například u parametru ZCR zmíněné v předchozím oddílu. Pokud bychom se rozhodli podívat se do prostoru $MFCC$ z hlediska různých tříd řečového signálu je nasnadě si je z hlediska různých tříd předem vhodně interpretovat. Na obrázku (2.3) vidíme jejich ukázkové spektrální vzory[3]. Prvá polovina sestává z výkonů určených ze spektrální výkonové hustoty po aplikaci melovy banky filtrů. Každá třída má tedy tři hodnoty výkonů ve frekvenčních pásmech ilustrovaných na obrázku (2.2). Jak již bylo zmíněno, jednotlivá pásma se citelně překrývají a jsou tudíž vzájemě silně korelována. Abychom je dekonvolovali, stačí pásma vyjádřit koeficienty *diskrétní cosinové transformace*, čímž přejdeme ze spektra do *kepstra*. Právě tyto kepstrální koeficienty jsou obsahem druhé poloviny grafu X ($A \rightarrow D$). $MFCC0$ si lze představit jako stejnosměrnou složku melovy banky. Pokud bychom se chtěly osvobodit od výkonu, bude to právě tento koeficient, na nejž zanevřeme. $MFCC1$ vyjadřuje poměr mezi vysokými a nízkými kmitočty. A nakonec $MFCC2$ vypovídá o zvlnění mezi jednotlivými pásmy - tedy o poloze středních kmitočtů vůči okolí. Z předpokladů zmíněných v parametrech řeči v předchozím oddílu, lze soudit, že pro velmi základní klasifikaci signálu nám stačí skutečně jen tyto prvé tři koeficienty. Pokud si naše $MFCC$ vzory tříd projikujeme do prostoru prvních třech koeficientů, jak znázorňuje

[3]pro zvýšení čitelnosti byl všem vzorům aditivně zvýšen výkon, což nijak nezkresluje dále popisované prostorové uspořádání shluků v prostoru $MFCC$



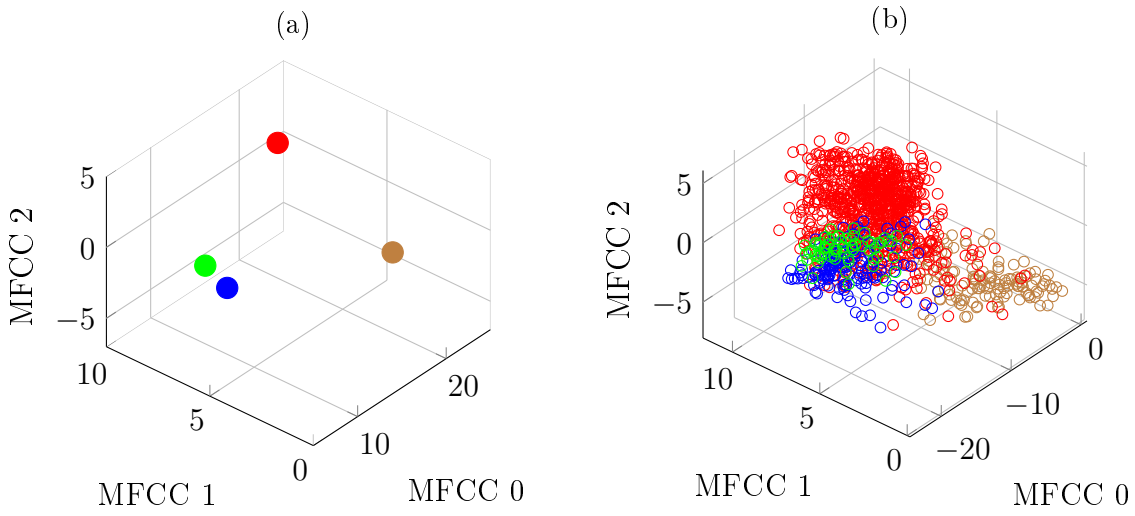
Obrázek 2.2: Ilustrace melovy banky filtrů pro 3 filtry: filtr 1 (červená), filtr 2 (zelená), filtr 3 (modrá)



Obrázek 2.3: Spektrální vzory a jejich *MFCC* obrazy: Znělé→ A, Neznělé→B, Respirace→C, Pauza→D

obrázek (2.4(a)), můžeme uzřít znatelné odlišnosti v poloze jednotlivých těžišť. Kolem těchto poloh lze očekávat shluky příslušných tříd reálného řečového signálu zobrazeném na obrázku (2.4(b)). Polohy těžišť se bohužel liší mluvčí od mluvčího. Jelikož *EM-algoritmus* je schopen odhalit shluky normálních rozdělení *bez učitele*, není tato rozmanitost těžišť na překážku. Velkým problémem je především možnost existence více shluků v rámci jedné třídy - typicky to bývá u znělých fonémů, kdy spektrální shluk na obrázku (2.4(b - červená)) bývá tvořen e.g. dvěma shluky se společným těžištěm a různým rozptylem. Všimněme si také, jak řídce zasahuje do shluku neznělých fonémů (hnědá). Tyto úseky signálu jsou neznělými frikativy,

jejichž spektrum má v pásmu vysokých kmitočtů vyšší energii a sdílí tak v prostoru prvních 3 *MFCC* polohu s neznělými frikativy. Obecně můžeme tento prostorový vzor vysledovat u všech řečových nahrávek. Berme na zřetel, že prvé tři koeficienty zde uvažujeme pro ilustraci. V implementaci algoritmu je počítáno s prvými 8 koeficienty z celkového počtu 24. Jelikož se nacházíme v *kepstru*, tak vzetím prvních 8 koeficientů vlastně provádíme homomorfní filtraci a získáváme informaci popisu spektra osmi pásmy. Protože jsou tato pásma určena ze 24 melových kepstrálních koeficientů, budou podávat odlišnou informaci, než kdyby byly určeny z melovy banky 8 filtrů. Vyšší počet koeficientů již nemá podstatný vliv na zamýšlenou klasifikaci, protože vyjadřuje vysokofrekvenční složku obálky spektra.

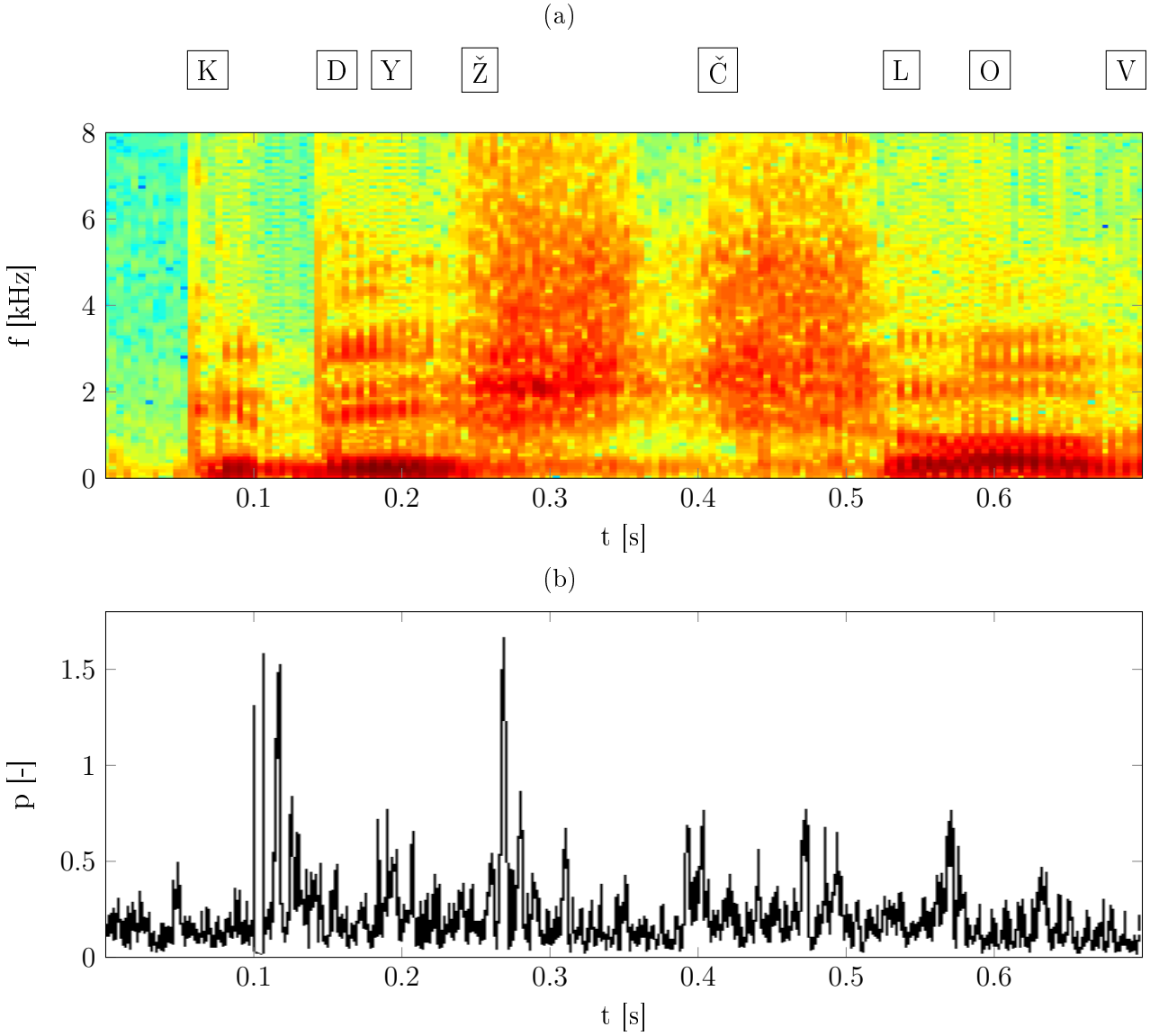


Obrázek 2.4: Znázornění prostorového vzoru (a) a reálného řečového signálu (b) v *MFCC* prostoru s barevně odlišenými třídami: pauza (modrá), dech (zelená), znělá řeč (červená); neznělá řeč (hnědá)

2.2.3 Pružná segmentace

Již při letmém průzkumu prostorů *(P, A, ZCR)* i *(MFCC)* se můžeme podivit, jak zajistit, aby měly jednotlivé třídy mezi sebou dostatečný odstup v prostoru parametrů. V praxi zpracování řeči se běžně počítá s úseky signálu segmentovanými oknem s konstantní délkou $\approx 10-30ms$ polohovaném klouzavě nebo bez překryvu. Při takovéto segmentaci lze s jistotou očekávat, že mnoho úseků bude svoji polohou překrývat dvě parametricky odlišné třídy a tento úsek se v prostoru parametrů zobrazí v poloze mezi hledané třídy. Klouzavé okno se běžně užívá pro tu výhodu, že podává hladký průběh parametrů. Vyhlazení může v některých případech ukrýt samotný hledaný shluk a *EM-algoritmus* odhadne parametry rozdělení které překrývá odlišné třídy. Tomu by bylo možné předejít takovým okénkováním, které by se svými hranicemi přizpůsobovalo hranicím tříd. Naše základní třídy *znělé fonémy, neznělé fonémy, pauza a respirace* se liší především ve spektru. Na obrázku (2.5(a)) můžeme už od pohledu odhadnou ideální hranice pro okénkování. K nalezení vhodných bodů se zdá být silným

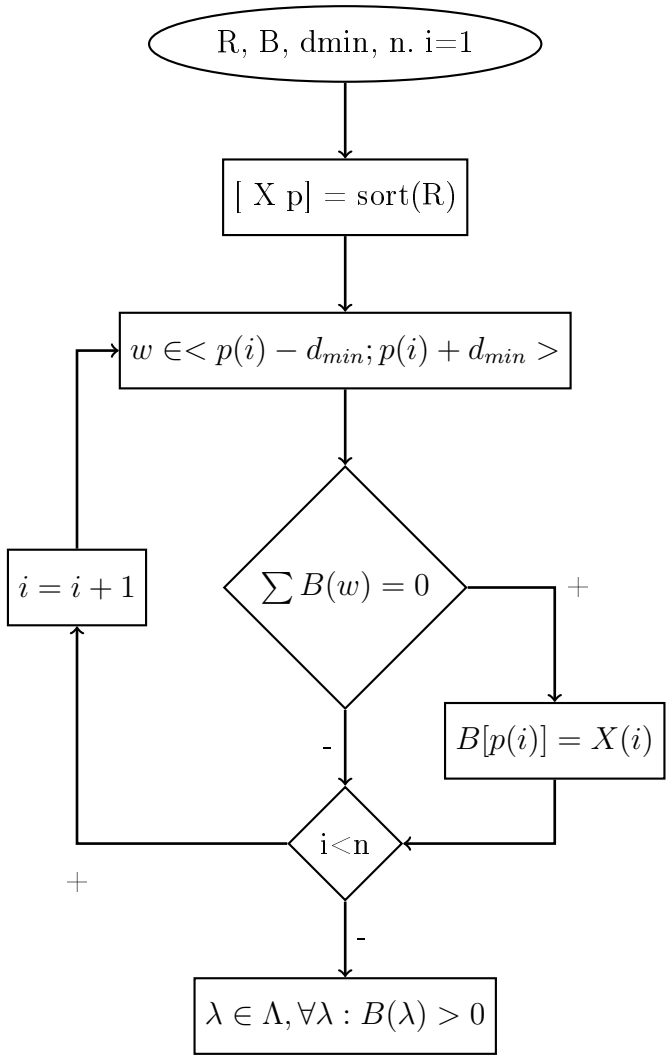
výchozím nástrojem *Bayesovský autoregresní detektor změn* [58]. Jeho výstupem je vektor posteriorních pravděpodobnost změn v signálu - pro zmíněný řečový signál je výstupní vektor vynesen na obrázku (2.5(b)). Při srovnání tohoto vektoru se spektrogramem jistě nalezneme ve vektoru naše myšlené ideální hranice jako *lokální maxima*. Na průběhu vektoru je také patrné, že význam těchto vrcholů je daný jejich relativnímu postavení vůči okolí a jejich osamocenost není směrodatná - viz. (2.5(b) $\approx 0.1s$). Jak získat z tohoto vektoru informaci o vhodných hranicích bude předmětem následující podkapitoly.



Obrázek 2.5: Ukázkový řečový signál reprezentovaný a) spektrogramem, b) vektorem posteriorních pravděpodobností *Autoregresivního bayesovského detektoru změn*

Algoritmus pružné segmentace

Mějme vektor posteriorních pravděpodobností změn spektra R o délce n, vektor hodnot segmentace B o délce n inicialisovaný nulami. Stanovme si minimální vzdálenost mezi hranicemi segmentů d_{min}, která určí minimální délku segmentu. Setříděním hodnot vektoru B sestupně (*sort*) přidělíme posteriorním pravděpodobnostem váhy úměrné jejich hodnotě. Sestupně roztříděným hodnotám X odpovídají jejich pozice p v původním vektoru. Budeme-li postupně s krokem i od nejvyšších hodnot X vkládat tyto hodnoty na původní pozice ovšem do vektoru B za podmínky, že v okolí w pozice p tohoto vektoru není ještě žádná vložená hodnota, naplníme vektor B právě jen těmi lokálně významnými vrcholy v okolí d_{min}. Pro přehlednost je algoritmus znázorněn na blokovém diagramu (2.6). Výsledná množina hranic



Obrázek 2.6: Vývojový diagram algoritmu pružné segmentace

Λ obsahuje pozice λ nenulových hodnot vektoru B. Jelikož je minimální vzdálenost d_{min} vzdáleností od jedné hranice, lze odvodit minimální délku segmentu L_{min} jako (2.2) a maxi-

mální délku segmentu L_{max} jako (2.3):

$$L_{min} = d_{min}, \tag{2.2}$$

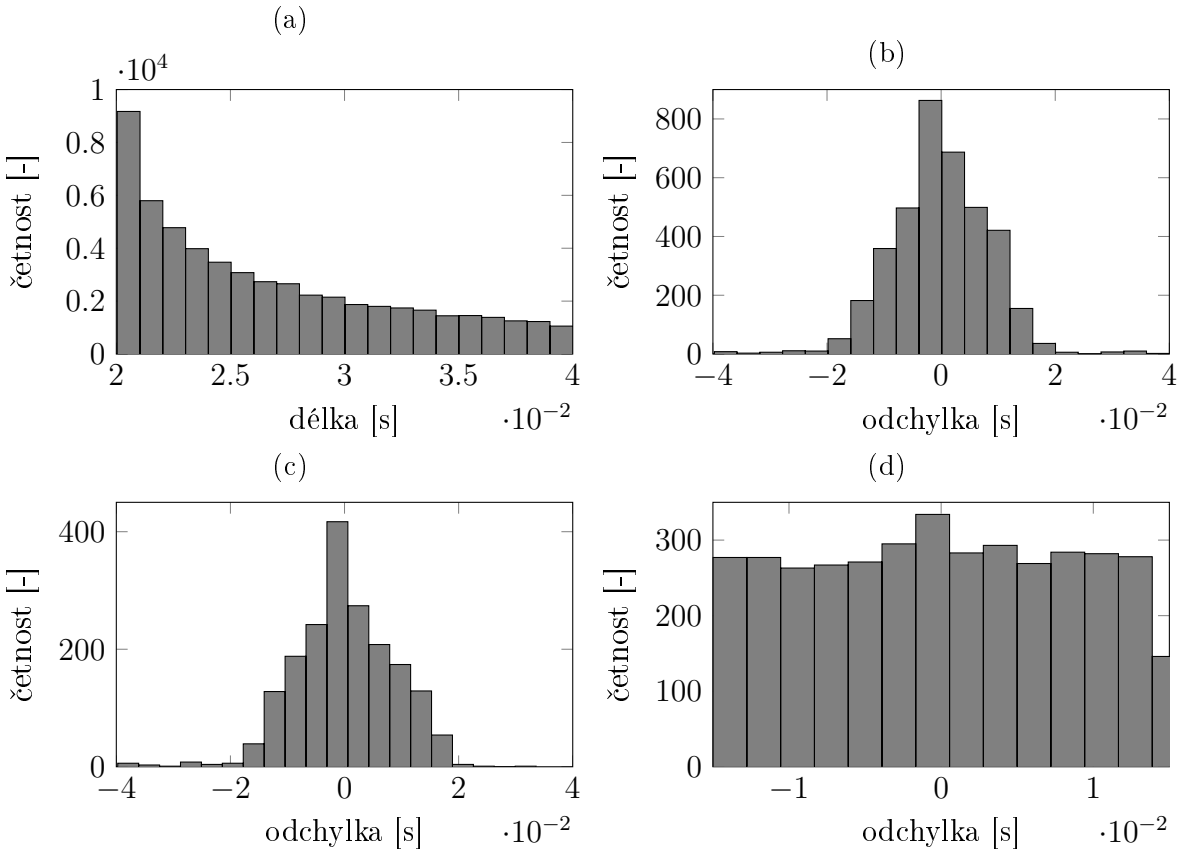
$$L_{max} = 2d_{min}. \tag{2.3}$$

Segmenty nebudou sice mít konstantní délku [4], ale rozsah délek se bude jistě pohybovat v rozsahu $< L_{min}; L_{max} >$, což je velice užitečné. Původní myšlenka, ze které tento algoritmus vyšel, bylo přiřadit do vektoru B maximální hodnotu R v okně konstantní délky bez překryvu. To však vedlo k neurčitě definované minimální délce segmentu L_{min}. Pro implementaci tohoto způsobu pro okénkování s překryvem bylo nutné zajistit zpětné odebrání vzorků nižších vah okolí překryvu z vektoru B. Výsledkem je tento algoritmus s pevně daným rozsahem délek.

Ohodnocení pružné segmentace

Předem víme rozsah délek segmentů. Vzhledem k tomu, že tímto algoritmem budeme úsekovat řečový signál, jehož vlastností je vysoká dynamika spektra, lze očekávat převahu kratších délek segmentů. Pro zjištění rozložení bylo tímto algoritmem nasegmentována databáze řečových signálů smíšené skupiny zdravých mluvčích i pacientů *PN*. Na obrázku (2.7)(a) vidíme histogram délek segmentů získaných ze signálů databáze. Podle předpokladů převažují kratší délky segmentů a jejich četnost se zvyšující se délkou klesá. Pokud bychom chtěli otestovat, jak moc se hranice blíží námi zamýšleným okrajům tříd, můžeme nalezené hranice otestovat na segmentované databázi pauz, která bude sloužit k ohodnocení celkové metody. Hranice pauz byly určeny ručně s rozhodováním na základě *spektra, výkonu, amplidudy, etc.*. Pro každou hranici pauzy v databázi nalezneme nejbližší hranici pružné segmentace a jejich vzdálenost zaznamenáme. Na histogramu (2.7)(b) lze pozorovat, že pro zvyšující se odchylku do kladných i záporných hodnot klesá i jejich četnost. Lze namítnout, že pro delší úseky také klesá četnost, což by mohlo být příčinou tohoto rozložení. Pokud se podíváme na histogram (a), tak pro hodnoty délky $0.03s$ a výše trend poklesu délek je již segmentů relativně nízký a pro ně bychom tedy měli obdržet odpovídající rovnoměrné rozložení podobné jako na histogramu (d). Pro každou délku segmentu přesahující $0.03s$ byly zaznamenány odchylky od ručních značek následovně: Pro každou hranici ruční segmentace byla přiřazena nejbližší nalezená hranice. Jestliže vzdálenost od této hranice k nejbližší další nalezené hranici v kladném i zápočném směru přesahovala $0.03s$, byla zaznamenána odchylka původní pozice od hranice ruční segmentace. Na obrázku (2.7(c)) je vynesen histogram těchto odchylek. K překvapení jsme získali rozložení, jehož četnost je opět klesající s rostorucí absolutní odchylkou. Na (2.7(d)) máme znázorněny odchylky pro segmenty s konstantní délkou okna bez překryvu $0.03s$. Můžeme pozorovat rovnoměrné rozložení v rozsahu $< -0.015; 0.015 >$. Pro všechny délky okna lze vysledovat rovnoměrné rozložení odchylek v rozsahu $< -\frac{L_w}{2}; \frac{L_w}{2} >$, kde L_w je délka okna. Porovnáním histogramů pružné segmentace s okénkováním bez překryvu lze ohodnotit kvalitu hranic jako výrazně vyšší. Dá se předpokládat, že algoritmus zvyšuje jakost informace v okně, jelikož přizpůsobuje jeho hranice tak, aby informace uvnitř okna byla

[4] tohoto požadavku nelze žádným způsobem dostát pro segmentaci s přizpůsobivými hranicemi



Obrázek 2.7: Histogramy pružné segmentace pro $d_{min} = 20ms$: a) délka segmentů, b) odchylka od ručních značek pauz, c) odchylka od ručních značek pauz pro úseky delší $30ms$, d) odchylka od ručních značek pauz pro okno bez překryvu s konstantní délkou $30ms$

homogenní. Z celé analýzy plyne, že algoritmus pružné segmentace je schopen nalézt hranice, které se blíží lidskému rozhodování o hranicích tříd.

Výsledné hranice

Z výstupního vektoru *Bayesovského autoregresního detektoru změn* znázorněném na obrázku (2.5(b)) jsme schopni vytěžit hranice signálu vyznačené svislými černými liniemi na obrázku (2.8). Jejich časové polohy samozřejmě závisí na zvoleném parametru d_{min}, který odpovídá za délky segmentů. Na obrázku (2.8(a)) jsou vyneseny značky pro $d_{min} = 30ms$. I při této vysoké hodnotě jsou hranice poměrně kvalitně umístěné, ale v jejich poloze je patrný kompromis mezi délkou segmentu a ideální polohou. Pro $d_{min} = 20ms$ lze získat již velice kvalitní hranice. Nicméně nehledě na subjektivní nastavení parametru d_{min} získáváme v obou případech užitečné informace o hranicích. Nemalou roli ve správném nalezení hranic hraje nastavení parametrů *Bayesovského autoregresního detektoru změn*. Pro popis řečových signálu je nejvhodnější zvolit řád *AR modelu* $5 - 7$ a délku okna $5 - 20ms$. Délka okna by měla do určité míry logicky korespondovat s parametrem d_{min}. Při praktickém nasazení se

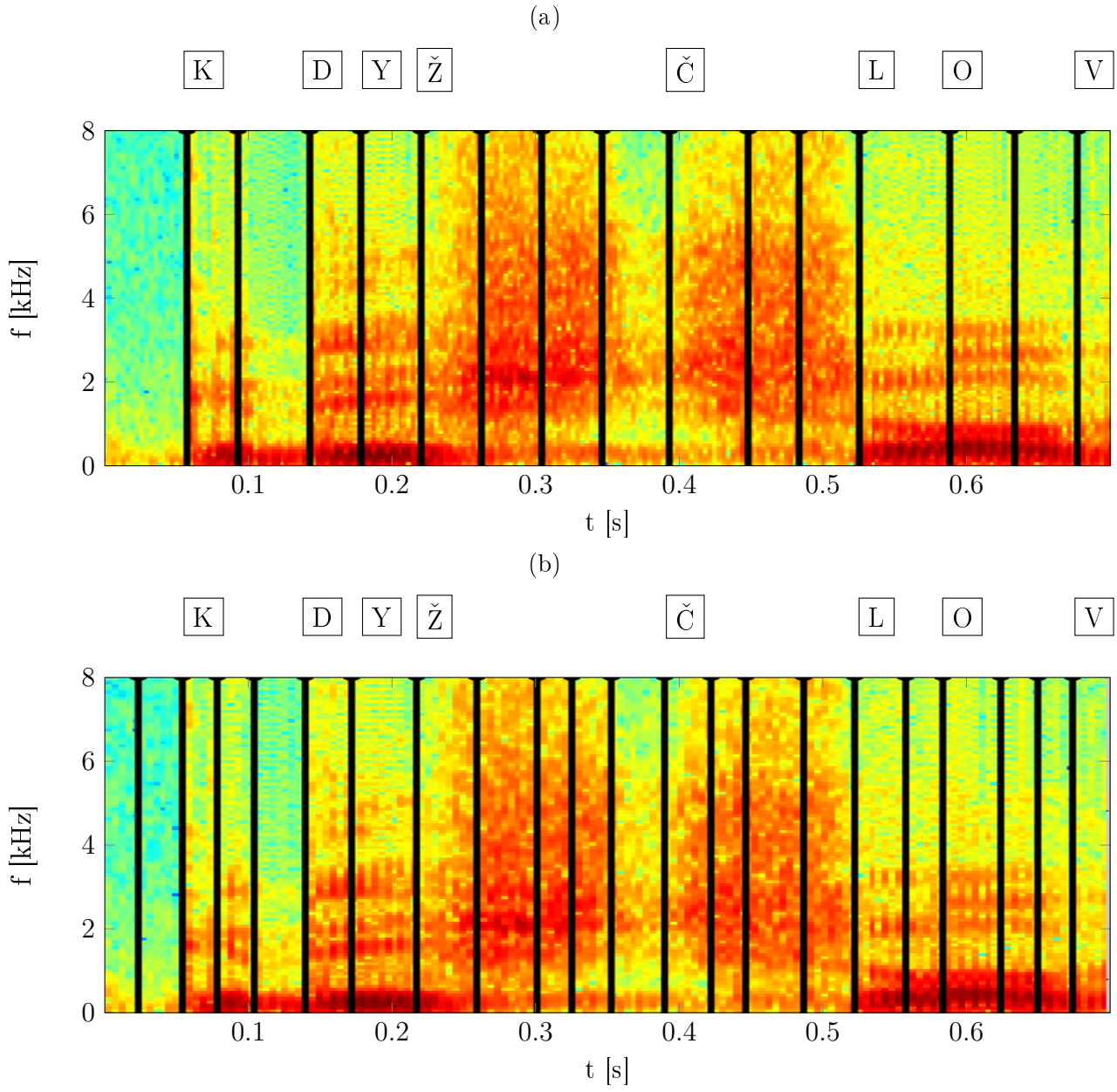
osvědčil řád *AR modelu* 7 a délka okna $10ms$.

2.2.4 Postup odhadování shluků

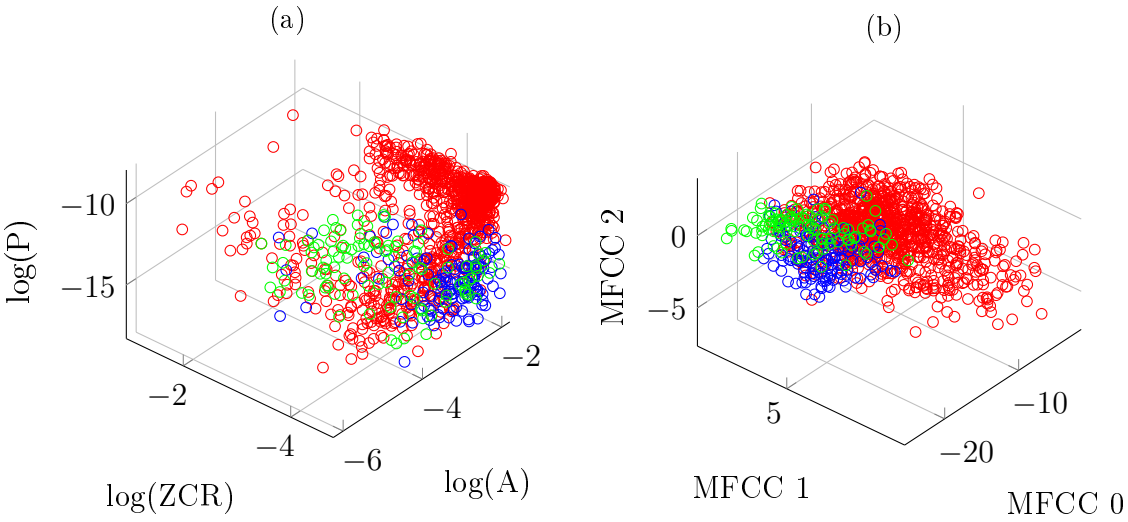
Abychom se v následujícím popisu metody neztratili, shrneme si výchozí podmínky. Máme řečový signál, který můžeme rozdělit oknem s konstantní délkou nebo pružnou segmentací. Jednotlivé úseky signálu si můžeme projikovat do dvou užitečných prostorů (*P, A, ZCR*) a (*MFCC*). Jednotlivé třídy *znělé fonémy, neznělé fonémy, pauza a respirace* budou v prostoru tvořit shluky, jejichž parametry budeme odhadovat pomocí *EM-algoritmu*. Cílem této kapitoly je najít, co nejrobustnější způsob odhadování těchto shluků. Pokud bychom chtěli všechny třídy určit přímým odhadem všech tříd v jednoz ze zmíněných prostorů, bylo by to velice elegantní řešení, bohužel však u některých signálů, zvláště pak těch patologické řeči, nelze předpokládat dostatečný odstup jednotlivých tříd v každém z prostorů, aby *EM-algoritmus* dokonvergoval právě k hledaným shlukům. Abychom zajistili dostatečnou robustnost, budou v prostorech hledány jednoznačně identifikovatelné shluky a ty budou následně ze signálu vyjmuty. Tím se v mnoha případech zbavíme nepřiměřeného překryvu, který by bránil správnému odhadu shluků. Vzhledem k tomu, že jednotlivé kroky odhadu na sebe navazují a případná chyba předchozího kroku se může velmi silně promítnout do výsledné detekce pauz, je nutné postupovat od nejspolehlivějších kroků po ty méně spolehlivé.

Odhad znělých fonémů

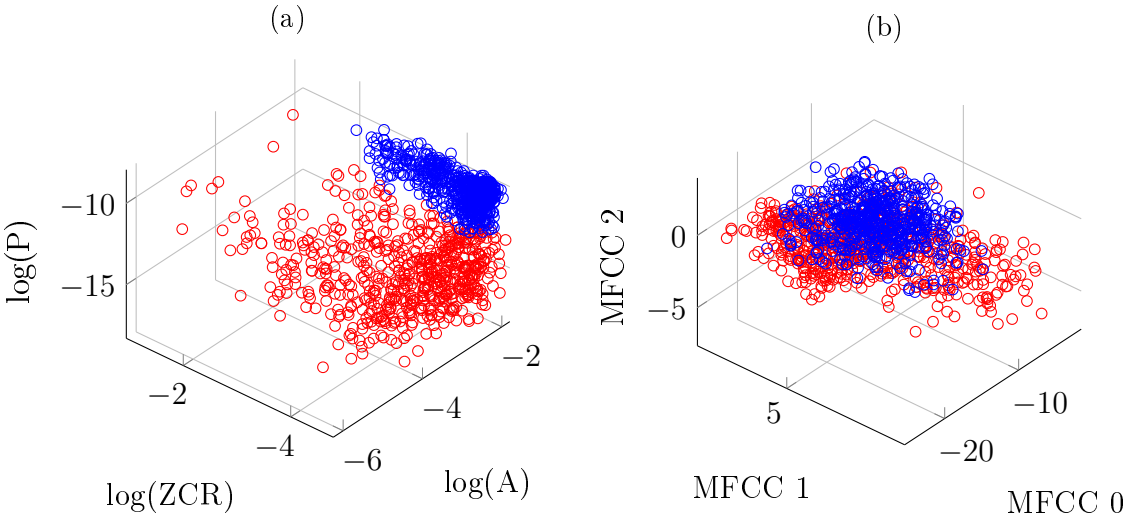
Skupina znělých fonémů má v obou prostorech význačné postavení jak z hlediska četnosti, tak i parametrů. Na obrázku 2.9 vidíme oba naše prostory s barevně odlišenými třídami *řeč, pauza, respirace*. V prostoru *MFCC* je zastoupen centrálním shlukem. Pokud bychom chtěli tuto skupinu odhadnout jako jeden shluk v tomto prostoru, podaří se to jen ve výjmečných situacích. Znělé fonémy mohou mít v případě e.g. nosovek či znělých explosivů poměrně ploché spektrum s energií soustředěnou v prvních několika harmonických. Taktéž znělé frikativy mají spektrum relativně ploché díky jejich šumové složce. U frikativů by to problém nebyl, jelikož je můžeme v prostoru *MFCC* najít ve shluku znělých frikativů. Problém je ve znělých fonémech, které se v *MFCC* prostoru prolínají s oblastí shluků *pauzy a respirace*. Další komplikací znělých fonémů je jejich rozmanitost ve spektru - v prostoru *MFCC* lze najít předem neodhadnutelné množství shluků. Právě neurčitost počtu hledaných shluků v prostoru*MFCC* jeho použití v tomto kroku znevýhodňuje. Řešení lze ovšem najít v prostoru *P, A, ZCR*, kde můžeme všechny znělé fonémy identifikovat v polohově diferencovaném shluku s nejvyšším výkonem. Jeho odhad *EM-algoritmem* pro 2 složkovou směs je ilustrován na obrázku (2.10(a-modrá)). Podíváme-li se na tuto odhadnutou skupinu v prostoru *MFCC* zjistíme, že je to skutečně ona problematická skupina znělých fonémů. Typicky obsahuje jak znělé hlásky, tak znělé souhlásky. Tuto skupinu vyjmeme a dále budeme pracovat pouze se zbytkem signálu.



Obrázek 2.8: Spektrogram řečového signálu s vyznačenými hranicemi pružné segmentace (svislé černé linie) pro parametr: a) $d_{min} = 30ms$ b) $d_{min} = 20ms$



Obrázek 2.9: Znázornění prostorů *P, A, ZCR* (a) a *MFCC* (b)
s barevně odlišenými třídami: pauza (modrá), dech (zelená), řeč (červená)



Obrázek 2.10: Znázornění prostorů a) *P, A, ZCR* s barevně odlišenými odhadnutými třídami 1 znělé fonémy (modrá), 2 ostatní fonémy (červená)
b) *MFCC* s barevně odlišenými třídami odhadnutými v prostoru *(P, A, ZCR)*, 1 znělé fonémy (modrá), 2 ostatní fonémy (červená)

Odhad neznělých fonémů

Podívejme se tedy na signál bez znělých fonémů v prostoru *MFCC*. Na obrázku (2.11(a)) vidíme barevně odlišené třídy signálu. Na první pohled jsme si vyjmutím usnadnili práci k odhadu ostatních skupin. Skupina neznělých frikativů (modrá), především sykavky, se v prostoru *MFCC* výrazně distancuje od zbylého shluku. Ten obsahuje *pauzu, respiraci , řečové artefakty a nevýrazně artikulované souhlásky*. Jelikož je poloha shluku *respirace* značně indi-

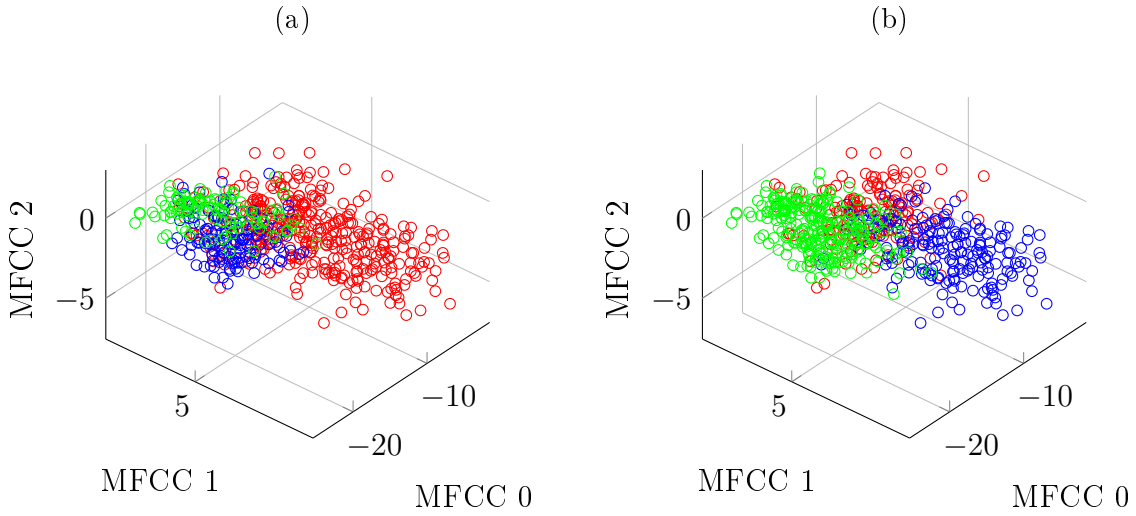
viduální, je výhodné odhadovat tento prostor pomocí 3 směsí *EM-algoritmu*. Prvá odhadnutá skupina bude zastoupena signálem *pauzy a případně respirace*, ve druhé se bude vyskytovat *slabě artikulované souhlásky a řečové artefakty* a ve třetí lze očekávat *neznělé frikativy*. O tom, vkteré skupině se bude vyskytovat *respirace* rozhoduje její znělost. Není výjimkou, že své těžiště nalezne ve skupině 2. Samozřejmě se některé úseky znělé respirace mohou zatoulat i do třetí skupiny. Proto je vhodné zavést vyhlazení rozhodování. Výskyt takto silného signálu *respirace* lze uvažovat uprostřed delších pauz. V indoevropské jazokové skupině se obvykle střídají znělé a neznělé úseky. Zřídka kdy nalezneme delší úseky řeči tvořené pouze konsonantami. Toho můžeme využít k vyhlazení rozhodování a rozsoudit skupinu 3 podle její vzdálenosti od nejbližšího znělého fonému, který už máme velice spolehlivě odhadnutý. Vhodná vzdálenost je v mezích $\approx 150 - 200ms$. Vyhlazení rozhodování probíhá následovně:

- Transformace úseků skupiny 3 na časový průběh
- Transformace úseků skupiny *znělých fonémů* na časový průběh
- Z obou časových průběhů nalezneme hrany
- Měříme vzdálenost náběžné hrany časového průběhu skupiny 3 k nejbližší doběžné hraně časového průběhu skupiny *znělých fonémů*
- Měříme vzdálenost doběžné hrany časového průběhu skupiny 3 k nejbližší náběžné hraně časového průběhu skupiny *znělých fonémů*
- Je-li jedna z těchto vzdáleností menší než kritická, pak příslušný úsek časového průběhu skupiny 3 přijmeme za *souhlásku*

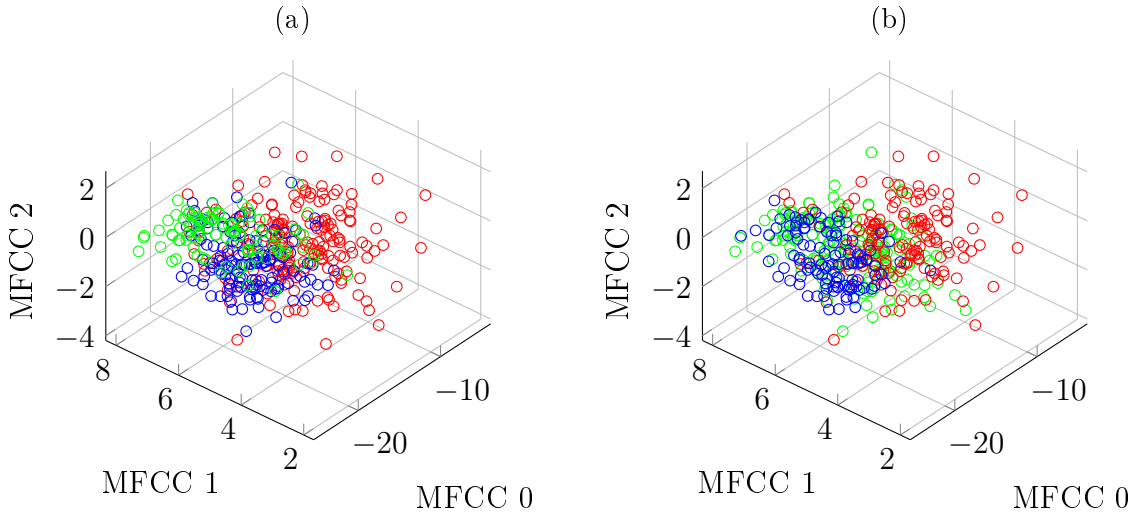
Takovým postupem pracujeme se skupinou 3 jako s časovými celky, což neomezuje délku fonémů jako v případě měření vzdálenosti jednotlivých segmentů skupiny 3.

Odhad pauz

Nyní můžeme skupinu *neznělých frikativů* vyjmout a prozkoumat zbylý signál. V něm nalezneme *respiraci, pauzu, neřečové artefakty (mlaskání, pohyby rtů), neznělé konsonanty (předopatrové a zadopatrové) a slabě artikulované frikativy*. Není výjímkou, aby se zde také nacházely nedetekované okraje znělých fonémů, které mají poměrně ploché spektrum zapřičiněné uzavíráním ústní dutiny před utišením hlasu. Nyní máme skupiny již velice očištěné a stačí odhadovat znovu tři složky. Na obrázku (2.12(a)) vidíme skutečné třídy *pauza, respirace,řeč* a jejich odhad (b). Za pauzu přijmeme skupinu s nejnižším těžištěm výkonu. Opět může dojít ke vniknutí respirace do skupiny 1. Tomu zamezíme jak jinak než vyhlazením rozhodování, o kterém již byla řeč. Místo samotné skupiny *znělých fonémů* užijeme skupinu sloučenou ze *znělých fonémů* a *neznělých frikativů* a tu budeme považovat za vztahový signál pro měření vzdálenosti. K této skupině provedeme rozhodování stejným postupem, jaký byl naznačen v předchozí podkapitole. Bohužel se může v některých případech ukázat nedostatečnost vyhlazení rozhodování. A to obzvláště pokud je mluvčí velmi dušný a projeví znělou respiraci ještě před kritickou pauzou. U takových pauz lze ovšem předpokládat, že budou mít poměrně



Obrázek 2.11: Znázornění prostorů *MFCC* s vyjmutými *znělými fonémy* s barevně odlišenými třídami: a) pauza (modrá), dech (zelená), řeč (červená), b) odhadnuté třídy: skupina 1 (modrá), skupina 2 (zelená), skupina 3 (červená)



Obrázek 2.12: Znázornění prostorů *MFCC* s vyjmutými *znělými fonémy* a *neznělými frikativy* s barevně odlišenými třídami: a) pauza (modrá), dech (zelená), řeč (červená), b) odhadnuté třídy: skupina 1 (modrá), skupina 2 (zelená), skupina 3 (červená)

velkou délku a lze ji tušit v prolukách hlasu delších 300ms. Nicméně pro zvýšení spolehlivosti je nutné myslet na to, že ona proluka v hlase není ve skutečnosti tolik dlouhou pauzou, ale může být prodloužena na svých okrajích neznělým frikativem. Proto si určíme následující kritéria vyhlazení rozhodování dlouhých pauz:

- Transformace úseků skupiny *znělých fonémů* na časový průběh
- Nalezneme proluky v časovém průběhu znělých fonémů delší než 300ms

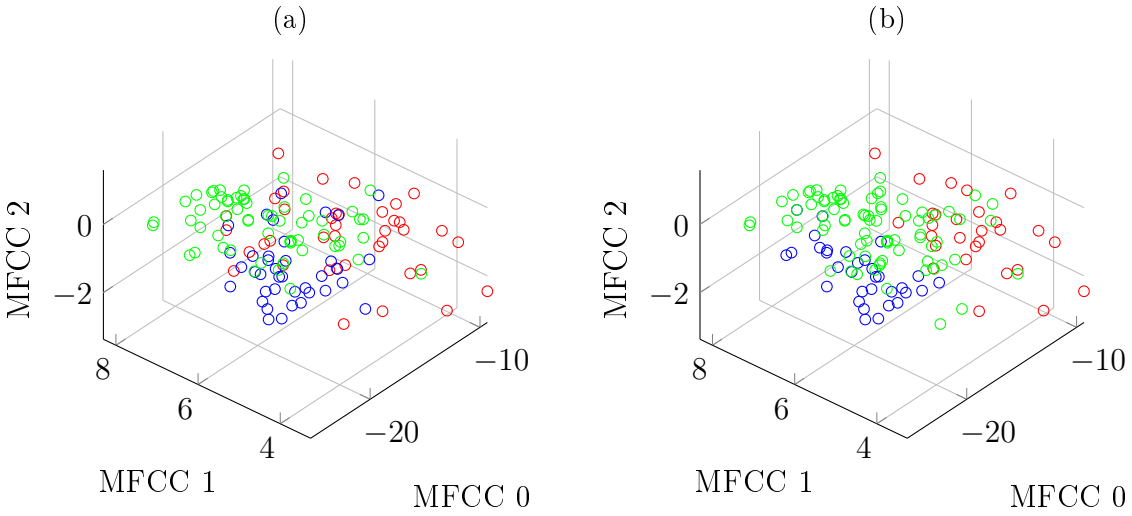
- Pro tyto proluky určíme jádro umístěné uvnitř pauzy v délce $60ms$ od obou okrajů
- Toto jádro přijmeme jako pauzu nehledě na její obsah

Odhad respirace

Ke konstrukci výchozího prostoru pro odhad respirace vyjdeme z již detekovaných pauz delších než $150ms$. Na obrázku 2.13(a) vidíme tyto dlouhé pauzy s vyznačenými třídami. Jelikož jsme se filtrací délek zbavili množství nadbytečných vzorků, můžeme přistoupit k přímému odhadu respirace v prostoru *MFCC*. Je velmi praktické odhadovat 3 složky, jelikož předem nevíme s jakou přesností byly výchozí pauzy detekovány. Složku 1, tedy tu složku s nejnižším těžištěm výkonu, přijmeme za pauzu a zbylé složky přijmeme za respiraci. To samo o sobě není výhodné řešení, jelikož se v prostoru často vyskytují e.g. okraje neznělých frikativů a také mnohdy detekujeme jen hlasité okraje respirace. Těchto chybových vzorků se zbavíme následujícím vyhlazením rozhodování:

- Transformace úseků respirace na časový průběh
- Proluky respirace kratší než $100ms$ přijmeme za respiraci
- Úseky respirace kratší než $50ms$ přijmeme za pauzu

Tato použitá kritéria aproximují skutečné očekávané časové parametry respirace a velmi zvýší účinnost výsledné detekce.



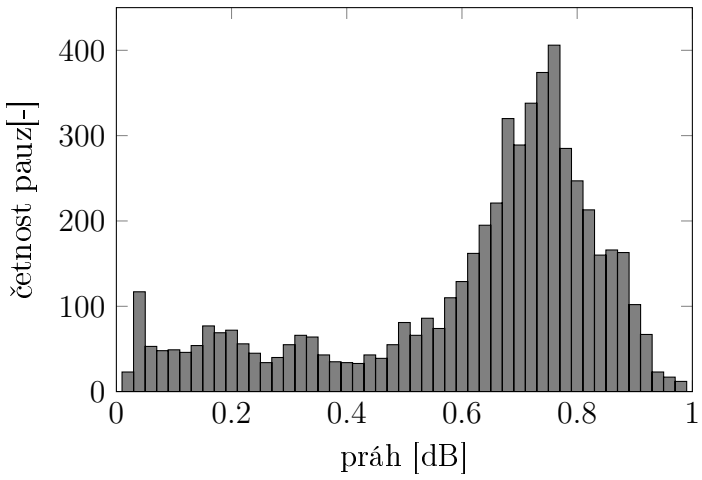
Obrázek 2.13: Znázornění prostorů *MFCC* pro nalezené pauzy $> 150ms$ s barevně odlišenými třídami: a) pauza (modrá), dech (zelená), řeč (červená), b) odhadnuté třídy: skupina 1 (modrá), skupina 2 (zelená), skupina 3 (červená)

2.3 Konvenční algoritmus detekce pauz

Algoritmus byl navržen pro detekci pauz v řečovém signálu normální a dysartrické řeči pro účely studie vlivu *Friedrichovy ataxie* na produkci pauz. Pro potřeby vyhodnocení pauz řečové dysartrie je toto jediný algoritmus, který neklasifikuje na základě empirického prahu. Proto byl vybrán jako srovnávací algoritmus k porovnání a ohodnocení funkce algoritmu prezentovaného v této práci. Výchozím parametrem algoritmu je výkonová obálka signálu o vzorkovací frekvenci $fs = 11.025kHz$ filtrované klouzavým průměrem řádu 100. Následně se pracuje s obálkou vyjádřenou jako *SPL*, tedy v logaritmuckém měřítku. Pro klasifikaci vychází algoritmus z kritérií pro detekci řečové aktivity popsaných Greenem [43]:

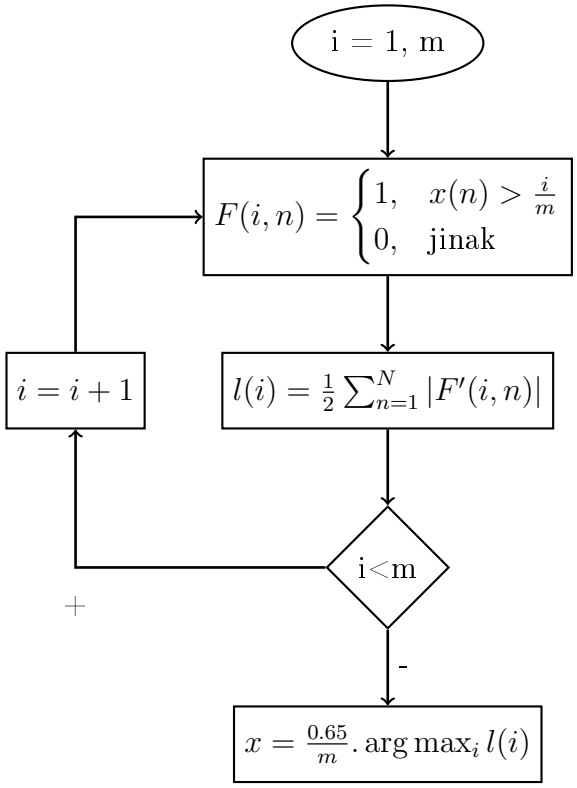
- minimální práh intenzity
- minimální délka pausy - $15 - 200ms$ [59, 43]
- minimální délka promluvy - $30 - 50ms$ [59, 43]

2.3.1 Metoda stanovení prahu



Obrázek 2.14: Histogram četnosti pauz pro úrovně typického řečového signálu českého mluvčího normalizovaného *SPL* (*sound pressure level - hladina hlasitosti*) na obor hodnot $< 0, 1 >$

Práh intenzity není jako v případě Greenovy metody určen empirickou konstantou, nýbž se stanovuje dynamicky v závislosti na signálu. Jeho určení vychází z předpokladu rozložení četnosti pauz v závistlosti na hodnotě prahu ilustrované na obrázku (2.14). Celý proces je znázorněn na vývojovém diagramu na obrázku (2.15), kde m značí počet počet (binů histogramu), i index prahu, $F(i, n)$ hodnotící funkci (promluva / pausa) signálu, N délku signálu, $l(i)$ celkový počet pauz v signálu a x výsledný práh intenzity. Pro jednotlivé prahy v celé škále hlasitosti vstupního signálu je z hodnotící funkce $F(i, n)$ stanoven počet pauz $l(i)$ pro konkrétní práh. Odhad prahu spočívá v postupném naprahování signálu

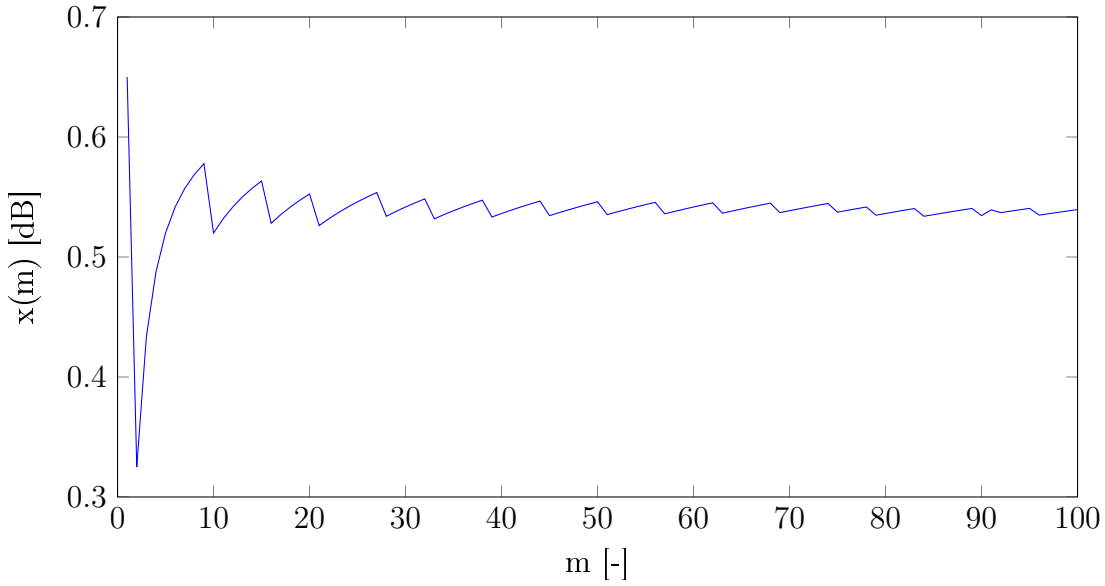


Obrázek 2.15: Diagram algoritmu pro stanovení prahu

v celém rozpětí výkonové obálky. S ohledem na dynamiku řeči, odpovídá každému prahu odlišný počet pauz. V rámci celé škály prahů předpokládá studie [59] rozložení znázorněné histogramem na obrázku (2.14). Hodnota klasifikačního prahu je určena jako 0.65 násobek modu tohoto histogramu $l(i)$. Šířka binů je doporučena na hodnotu $0.05dB$, ve studii však není udán rozsah hodnot SPL. Před samotnou implementací bude vhodné provést analýzu této metody stanovení prahu.

2.3.2 Analýza metody stanovení prahu

Na obrázku (2.16) je znázorněna nalezená hodnota prahu x v závislosti na počtu binů m pro jeden vstupní signál. Vstupní signál SPL byl normalizován na obor hodnot $< 0; 1 >$. Pro všechny signály lze nalézt závislost podobnou tomuto ukázkovému průběhu. Je patrné, že s vyším počtem binů hodnota prahu konverguje ke globálnímu optimu. Obecně lze říci, že pro počet binů menší než 100 dochází k výraznému rozptylu hodnot prahů, uvědomíme-li si rozsah hodnot $< 0; 1 >$. Počtu binů 100 odpovídá šířka $0.01dB$ a lze se domnívat nepracuje s normalizovanou obálkou. Pro implementaci algoritmu je vhodné použít co vyšší počet binů, minimálně větší než 100. V případě, že se blíže zamyslíme nad délku okna použitého filtru 100 ($9.1ms$) vzhledem k použité vzorkovací frekvenci $11.025kHz$ by mělo dojít k potlačení periodických signálů (f_0 hlasivek a $60Hz$ síťové rušení). Případné nízkofrekvenční rušení (zastoupené např. $50Hz$ síťového kmitočtu) však potlačeno nebude, což způsobí zkreslení



Obrázek 2.16: Závislost hodnoty nalezeného prahu x na počtu binů m histogramu pro normalizovanou obálku SPL na obor hodnot$< 0; 1 >$

odhadnutého počtu pauz $l(i)$ v celé škále prahů. Je tedy velmi vhodné provést vstupní filtraci původního signálu horní propustí pro potlačení tohoto nežádoucího pásma. Pokud má odhad prahu vycházet z modu nalezených pauz je namístě se zamyslet nad vlivem délky použitého filtru. Původní délka je zjevně určena k vyhlazení periodických energetických špiček signálu a má minimální vliv na potlačení dynamiky. Je otázkou jestli je vhodné uvažovat pauzu od délky $9.1ms$, kterou určuje řád filtru. Podívejme se na obrázek (2.17). Vidíme dva průběhy závislostí x_T^M střední hodnoty prahu definované vztahem (2.6), určené z jednotlivých prahů posunutých do jejich těžiště dle vztahu (2.4):

$$x_T = x - \frac{1}{M}\sum_{i=1}^{M} x_T(i), \tag{2.4}$$

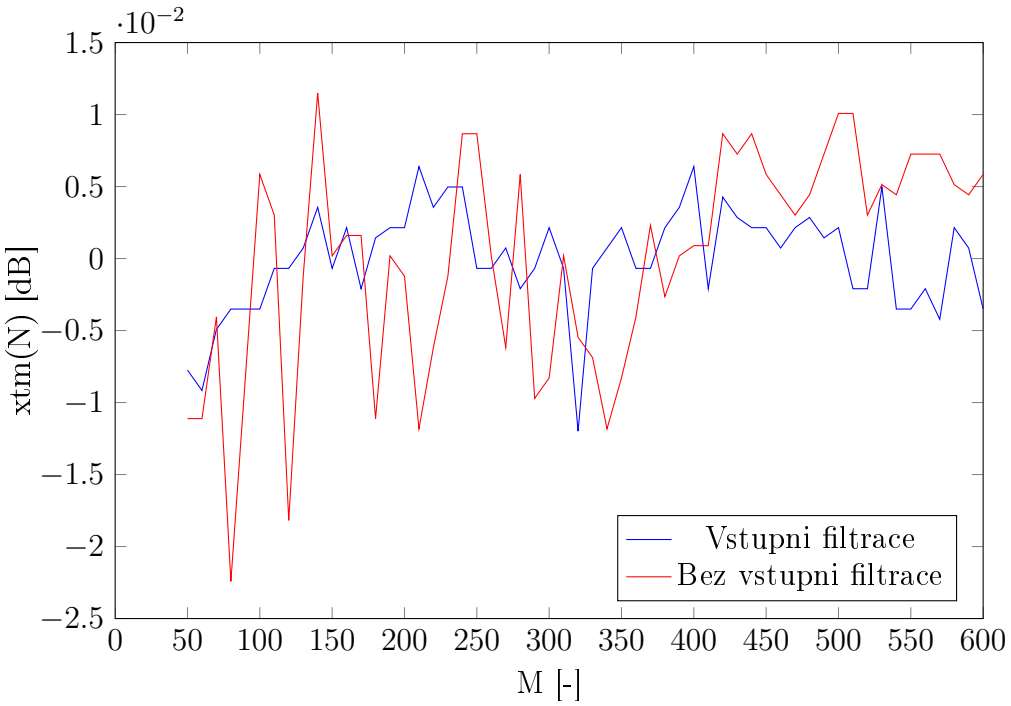
$$x_T^M = \frac{1}{N}\sum_{i=1}^{N} x_T(i). \tag{2.5}$$

První průběh (červená) patří signálu bez vstupního odstranění rušivého pásma. Z proměnlivého rozptylu hodnot prahu lze soudit, že jednotlivé prahy mají také velký rozptyl hodnot. Poměrně hladší průběh vidíme u druhého signál (modrá), který prošel *preprocessingem*. Data prahů byla určena pro signály z celé databáze délky N pro jednotlivé řády filtru M. Už při letmém porovnání průběhů musíme uznat, že vstupní *preprocessing* má významný vliv na stabilitu nalezeného prahu. Filtr řádu 100 neposkytuje dostatečnou filtraci pro konzistentní odhad prahu, což je patrné jak z obrázku (2.17), tak z průběhu rozptylů $\sigma^2(x_M)$ určených

vztahem (2.6):

$$\sigma^2(x_M) = \frac{1}{N}\sum_{i=1}^{N}(x_T(i) - x_T^M(i))^2. \tag{2.6}$$

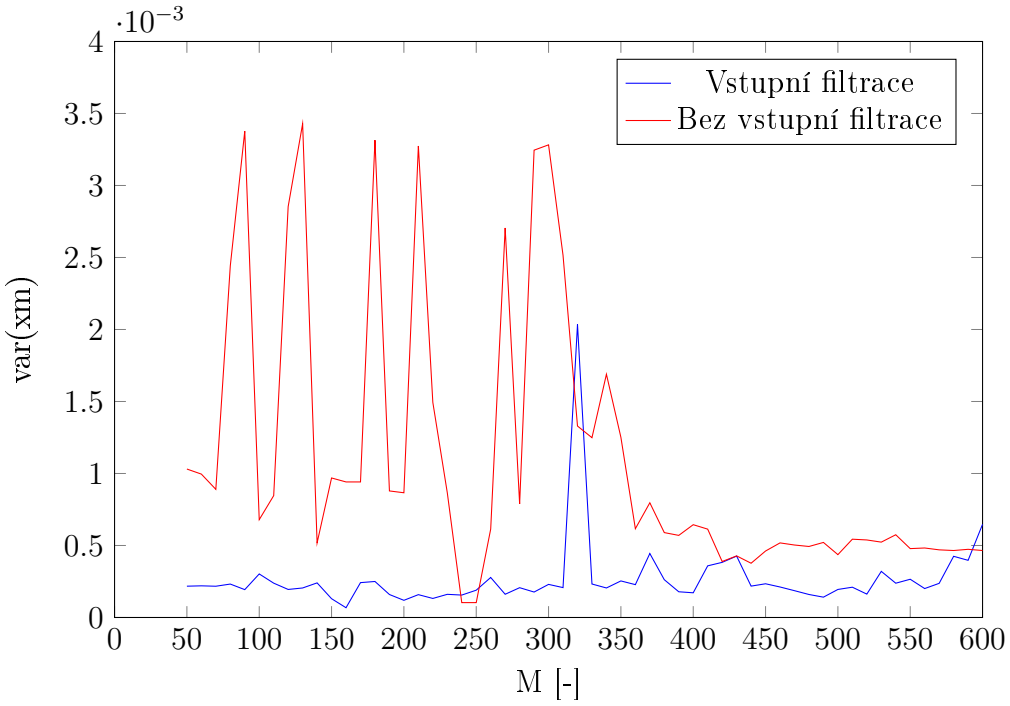
Průběhy jsou zobrazeny na obrázku (2.18). Na obrázku (2.18) je také čitelný rozdíl mezi rozptylem filtrovaného a nefiltrovaného vstupního signálu při změně řádu filtru. Filtrovaný vstupní signál nese jednoznačně kvalitnější informaci o produkci pauz. Za vhodný řád klouzavého průměru při vzorkovací frekvenci $11.025kHz$ se jeví být řád > 500, kterému odpovídá časová délka $> 40ms$. Takový filtr vyhladí obálku dostatečně spolehlivě a zvýší tak informaci o dynamice řeči jak pro filtrovaný tak i pro nefiltrovaný vstupní signál. Konstanta 0.65 byla empiricky zvolena tak, aby zachytila většinu konsonant a zároveň zachovala prahu odstup od případného signálu respirace. Výsledným prahem byla klasifikována výkonová obálka SPL a z ní odstraněny pausy $< 15ms$ a promluvy $< 30ms$ dle Greenových kritérií.



Obrázek 2.17: Závislost hodnoty x_M^T na řádu filtru M při filtrovaném (modrá) a nefiltrovaném (červená) vstupním signálu

2.3.3 Implementace

Algoritmus byl nejprve implementován přesně podle studie [59], tedy jako filtrovanou výkonovou obálku SPL ze signálu o vzorkovací frekvenci $f_s = 11025Hz$ s řádem filtru 100 a počtem binů 200. Jeho výsledky byly ale krajně neuspokující (nalezený práh se ve většině odhadů nacházel mimo efektivní rozsah), a proto byly provedeny úpravy metody, které vyšly z přede-



Obrázek 2.18: Závislost hodnoty rozptylu $\sigma^2(x_M)$ na řádu filtru M při filtrovaném (modrá) a nefiltrovaném (červená) vstupním signálu

šlé analýzy metody. Vstupní signál byl filtrován dolní propustí s mezním kmitočtem $100Hz$[5]. Při tomto *preprocessingu* není nutné měnit hodnotu řádu filtru 100. Nejvýše choulostivým bodem byla změna vzorkovací frekvence. Ukázalo se, že její snížení má pozitivní dopad na výslednou detekci. Nejlepších výsledků bylo dosaženo se vzorkovací frekvencí $8kHz$. Vysvětlení této závislosti určitě nespočívá v myšlence, že snížení vzorkovací frekvence má podobný efekt jako její ponechání v původní hodnotě za současného zvýšení řádu filtru. Vyšší řád filtru nezvyšuje účinnost detekce, pouze stabilitu prahu, kterou jsme zajistily *preprocessingem*. Celý systém odhadu prahu se choval velmi nestabilně v závislosti na vzorkovacím kmitočtu. Otázkou zůstává, jestli bude algoritmus podávat stejné výsledky i pro řečové signály v jiné kvalitě než v databázi, na kterou je naučen.

2.4 Navržené příznaky

Pro navržení příznaků produkce pauz vyjdeme z předpokladů shrnutých v kapitole 1. Minimální délku pauzy budeme uvažovat od $15ms$. Od této hranice po $50ms$ budeme hovořit o krátkých pauzách, jejichž výskyt je vedlejším produktem artikulace - typicky na rozhraní znělých a neznělých fonémů. Od $50ms$ do $100ms$ budeme uvažovat střední pauzy - ty mají význam jednak pro artikulaci explosivů, jednak pro zvýraznění kořenů složených slov. Pauzy

[5]hodnota není kritická, měla by se pohybovat v řádu $60 - 200Hz$, aby dostatečně potlačila síťové rušení a zároveň nesnižovala energii znělých fonémů

od $100ms$ berme za dlouhé pauzy pro formální člení textu a vykonání respirace. Prozkoumáme zastoupení pauz v promluvě parametrem pPS. Od tohoto parametru budeme očekávat vyšší zastoupení pauz v řeči PN. Dále prozkoumáme počty pauz počty jednotlivých délek pauz. Pro PN by počty jednotlivých délek pauz měly být nižší. U skupiny krátkých pauz lze očekávat, že budou v obou skupinách dodrženy v podobných řádech, jelikož jsou vyústěním artikulace. Střední a dlouhé pauzy by měly být u PN ochuzeny v počtu. Z průzkumu zastoupení délek pauz se pokusíme zjistit, jestli při *hypokinetické dysartrii* nedochází k obohacování jedné skupiny délek pauz na úkor jiné. Střední artikulační rychlostí se pokusíme vyjádřit pomalost řeči. Jelikož známe obsah promluvy, vztáhneme pouze počet slabik na dobu řečové aktivity. Dle studie [37] dochází v řeči PN ke zrychlování artikulace na úkor produkce pauz. Pro vyjádření artikulace v úseku promluvy bychom museli mít promluvu předem rozdělenu a znát přesný počet slabik této části. K tomuto problému můžeme přistoupit z jiné stránky a prozkoumat produkci pauz ve dvou polovinách délky promluvy pro jednotlivé délky pauz. Akceleraci artikulace na úkor produkce pauz popíšeme jako rozdíl rychlosti produkce pauz opět pro všechny pauzy i jejich jednotlivé skupiny délek. K výslednému posouzení parametru musíme přistoupit s ohledem na neznalost obsahu promluvy v těchto polovinách a tomu přizpůsobit soud o zrychlení či zpomalení řeči. Považujeme-li krátké pauzy za doprovodný jev artikulace, může být zajímavým parametrem poměr produkce krátkých pauz - tedy doba, do které bylo vyprodukováno $\frac{3}{4}$ krátkých pauz, vztažená na dobu řečové aktivity. Hranice poměru produkce je vhodné pro citlivý odhad zvolit vyšší než $\frac{1}{2}$, aby jí byla zahrnut i akcelerovaný úsek řeči. Z výsledků studie [37] můžeme soudit, že dojde ke zrychlení produkce pauz pro krátké pauzy a zpomalení produkce pauz pro dlouhé pauzy. Pro zrychlenou produkci pauz bychom tedy ony $\frac{3}{4}$ krátkých pauz měli detekovat v kratší době, kterou pro zobjektivnění porovnání navíc vztáhneme na dobu řečové aktivity. Takový závěr by mohl vést k úvaze o zrychlování artikulace. Řeč lze popsat z hlediska produkce pauz také jako rytmus, kterým je obsah promluvy členěn formálními pauzami. Rychlost rytmu můžeme určit jako střední hodnotu vzdáleností středů sousedních dlouhých pauz. Pro *hypokinetickou dysartrii* očekáváme tyto vzdálenosti signifikantně delší. Jestliže bychom chtěli vzít v potaz vliv zrychlení artikulace na akceleraci rytmu, můžeme podobně jako v případě parametru akcelerace produkce pauz rozdělit signál na dvě poloviny. Pro každou polovinu určíme délku rytmu všch pauz kratších než $50ms$. Minimální délku bude pro zvýšení přesnosti v tomto případě nutné zvolit výjmečně od $10ms$, tím získáme redundantní množství informací o rozhraních fonémů spíše než pauz. Vyvážíme tím však vychýlení naší střední hodnoty vzájemných vzdáleností způsobené dlouhými pauzami a získáme lepší představy o artikulačním rytmu. Rozdílem těchto rytmů vyjádříme akceleraci. Jelikož prezentovaný detektor je schopen určit také respiraci, pokusíme se i na tomto poli nalézt příznaky PN. Vyšší zastoupení respirace v pauzách by mohlo mít souvislost s příznaky PN - především rigiditou dýchacího svalstva. V kapitole 1 jsme vyslovili názor, že formální členění textu by mělo být u PN podřízeno respiraci. Pokud by tomu tak bylo, pak by se rytmus dechu měl promítnout do rytmu dlouhých pauz. Samotná rychlost rytmu dechu je velmi podřízena aktuální homeostáze mluvčího a jistě nebude příliš signifikantním parametrem. Pokud by však odchylky rytmu dechu vychylovaly rytmus pauz, mohli bychom tušit jejich příčinnou souvislost. Definujme si rytmicitu jako odchylku od rytmu - tedy rytmicitu pauz jako směrodatnou odchylku vzdáleností středů sousedních

Zkratka	Příznak	Definice
pPS	Poměr pauzy a řeči	Poměr celkové délky pauz a délky řečové aktivity
nP	Počet všech pauz	Celkový počet detekovaných pauz
nPk	Počet všech krátkých pauz	Celkový počet detekovaných pauz $< 50ms$
nPs	Počet středních pauz	Počet detekovaných pauz $< 50; 100>$ms
nPd	Počet všech dlouhých pauz	Počet detekovaných pauz $> 100ms$
zPk	Zastoupení krátkých pauz	Poměr počtu krátkých pauz a všech pauz
zPs	Zastoupení středních pauz	Poměr počtu středních pauz a všech pauz
zPd	Zastoupení dlouhých pauz	Poměr počtu dlouhých pauz a všech pauz
mVS	Střední artikulační rychlost	Počet slabik na dobu řečové aktivity
mVP	Střední rychlost produkce pauz	Počet všech pauz za dobu promluvy
mVPk	Střední produkce krátkých pauz	Počet pauz kratších $50ms$ za dobu promluvy
mVPs	Střední produkce středních pauz	Počet pauz délek $< 50; 100 > ms$ za dobu promluvy
mVPd	Střední produkce dlouhých pauz	Počet pauz delších $100ms$ za dobu promluvy
aVP	Akcelerace produkce pauz	Rozdíl rychlostí mVP v $1.\frac{1}{2}$ a $2.\frac{1}{2}$ promluvy
aVPk	Akcelerace produkce krátkých pauz	Rozdíl rychlostí mVPk v $1.\frac{1}{2}$ a $2.\frac{1}{2}$ promluvy
aVPs	Akcelerace produkce středních pauz	Rozdíl rychlostí mVPs v $1.\frac{1}{2}$ a $2.\frac{1}{2}$ promluvy
aVPd	Akcelerace produkce dlouhých pauz	Rozdíl rychlostí mVPd v $1.\frac{1}{2}$ a $2.\frac{1}{2}$ promluvy
ppP	Poměr produkce krátkých pauz	Poměrná doba vyprodukování $\frac{3}{4}$ krátkých pauz
srP	Střední rytmus pauz	Střední hodnota vzdáleností sousedních pauz $> 100ms$
arA	Akcelerace rytmu artikulace	Rozdíl parametru rP pro $1.\frac{1}{2}$ a $2.\frac{1}{2}$ promluvy
pRP	Zastoupení respirace	Poměr celkové délky respirace a dlouhých pauz
prR	Poměrná rytmicita respirace	Poměr směrodatné odchylky rytmu respirace a pauz

Tabulka 2.2: Navržené příznaky, jejich zkratky a definice

pauz delších než $100ms$ a rytmicitu respirace jako směrodatnou odchylku vzdáleností středů sousedních nádechů. Jestliže by dech byl v pauzách *PN* více zastoupen a rytmicita dechu by ovlivňovala rytmicitu dlouhých pauz, pak by poměr jejich rytmicit měl být blízký jedné. Všechny navržené příznaky jsou shrnuty v tabulce 2.2.

2.5 Statistika

2.5.1 Ohodnocení metody

Před určení způsobu ohodnocení je vhodné uvést v pořádek, jaké parametry od metody detekce pauz očekáváme. Předně nás zajímá zda byla pauza správně určena. Tedy jestli její hranice leží v určitém tolerančním pásmu. Toleranční pásmo bude ležet v okolí hodnotící značky v délce poloviny délky příslušné pauzy. Tím zajistíme delší toleranční pásmo pro dlouhé pauzy a kratší pro krátké pauzy. Příznaky *PN* budeme zkoumat i z hlediska délky pauz, proto druhý požadavek pro ohodnocení budeme klást právě na délku pauzy: zajímá nás jak dlouhá byla právě hodnocená pauza a jaké úspěšnosti jsme schopni dostát pro určité délky pauz. Následující ohodnocení tak stanoví účinnost nalezených pauz v závislosti na jejich délce. Pro lepší čitelnost výsledné informace budeme hodnotit kumulativní účinností v závistlosti na rostoucí délce. Zvolíme tedy práh minimální délky pauzy v posloupnosti

od 40ms[6] do $300ms$ a pro všechny pauzy větší než tento práh zjistíme účinnost detekce. Výslednou účinnost detekce můžeme definovat jako parametr s určený vztahem 2.7:

$$s = 100.\frac{\sum c(l)}{\sum v(l)}, \tag{2.7}$$

kde $c(l)$ zastupuje správně detekované hranice pauz delších než práh l. Pokud bylo detekováno více pauz delších prahu l než očekáváme, pak hodnota $v(l)$ bude zastupovat počet počet všech detekovaných pauz delších než práh l. Pokud však bylo detekováno méně pauz než očekáváme, pak bude hodnota $v(l)$ zastupovat očekávaný počet nalezených pauz. Hodnocení tak může spřísnit právě případný nadbytečný či podbytečný počet pauz do jednoho výsledku. Signál respirace je v mnoha případech velmi slabý a nepřesnost ručního značení vyvoláva potřebu širokého tolerančního pásma v délce odpovídající době nádechu. Prostá fltrace hodnotících a hodnocených značek *respirace* v širokém tolerančním pásmu však může velmi zkreslit výsledné hodnocení. Značky hodnocené totiž mohou ležet v tolerančním pásmu značek hodnotících a zároveň mít délku menší o dvojnásobnou velikost tolerance. Je velmi pravděpodobné, že taková hodnocená značka bude nad prahem délek a její hodnotící značka zároveň bude pod prahem délek. Ačkoliv byla správně nalezena, vyhodnotíme ji chybně. Takový precedens může nastat i v opačném případě krátké hodnocené značky v tolerančním pásmu. Každý úsek *respirace*, jejíž obě hranice byly kladně nalezeny, přiřadíme délku odpovídající jejímu mateřskému hodnotícímu úseku. Při filtraci délek značek tím předejdeme zmíněnému zkreslení hodnocení.

2.5.2 Ohodnocení příznaků

Pro porovnání odlišnosti našich dvou skupin - *zdravých mluvčí a pacientů PN* je nutné zvolit vhodnou statistiku a definovat nulovou hypotézu, kterou podrobíme statistickému testu a na jeho základě jeho výsledku ohodnotíme přístušný příznak. Navržené příznaky můžeme považovat za normální rozdělení popsané střední hodnotou a rozptylem. Skupiny *zdravých mluvčí* a skupinu *pacientů PN* budeme považovat za *nezávislé náhodné vektory* podléhající normálnímu rozdělení s různými středními hodnotami: μ_1 pro *zdravé mluvčí* a μ_2 pro *pacienty PN:*. Za těchto předpokladů definujme nulovou (2.8) a alternativní hypotézu (2.9):

$$H_0 : \mu_1 = \mu_2, \tag{2.8}$$

$$H_1 : \mu_1 \neq \mu_2. \tag{2.9}$$

Jako nulovou hypotézu H_0 tedy budeme testovat, zda navržené příznaky podléhaji rozdělení se stejnou střední hodnotou . V případě jejího zamítnutí zvolíme alternativní hypotézu H_1, která tvrdí, že skupiny v daném příznaku podléhají dvěma normálním rozdělením s odlišnou střední hodnotou. Za testovací statistiku byl zvolen *nepárový dvojvýběrový t-test*, kterým budeme hodnotit příznaky obou skupin popsaných odhadem střední hodnoty a rozptylu tohoto souboru. Výsledek testu budeme hodnotit podle hladiny významnosti. Hypotézu H_0 pro příslušný příznak zamítneme při výsledné hladině významnosti příslušného příznaku větší 5%.

[6] tato délka se odvíjí od minimální délky pauz ručních značek hodnotící databáze

Kapitola 3

Výsledky

3.1 Ohodnocení algoritmů

Na obrázku 3.1 máme ilustrovánu kvalitu detekce v závislosti na délce detekovaných pauz. Již při letmém porovnání průběhů poměrně nápadného neúspechu konvenční i prezentované metody pro krátké pauzy $< 200ms$. Je vhodné si uvědomit, že ohodnocení pro příslušnou délku určuje výsledná úspěšnost pro všechny pauzy delší tohoto kritéria. Chyby všech delších pauz se tedy kumulativně promítnou do hodnocení pauz kratších. Z tohoto pohledu se znovu podívejme na základní průběhy úspěšnosti obrázku 3.1(a) i jejich směrodatné odchylky (b). Problémem prezentovaného algoritmu nejsou krátké pauzy, nýbrž pauzy v rozmezí $100 - 200ms$, jimž odpovídá jak výrazný pokled úspěšnosti tak výrazný vrchol směrodatné odchylky úspěšnosti. S těmito pauzami mají problém obě srovnávané metody. Podstata problému detekce těchto pauz spočívá v podobnosti parametrů některých konsonant k parametrům šumu. Pokud jsou navíc slabě artikulovány, je velmi obtížné rozlišit tyto fonémy od skutečných pauz. Jejich chyba se tedy projevuje na obě strany - chybným přijetím pauzy i chybným zamítnutím pauzy. Pokud se v tomto kontextu znovu zamyslíme nad úspěšností detekce pauz kratších $100ms$, je zřejmé, že prezentovaná metoda je výrazně úspěšnější. U konveční metody můžeme pro tyto délky pozorovat neustávající pokles úspěšnosti. Tomu také odpovídá pokles směrodatné odchylky úspěšnosti. Pro vyšší délky u konvenční metody zjevně dochází k velkému rozptylu úspěšnosti. Ten lze přisoudit jednak situaci, kdy na problematickou pauzu $100 - 200ms$ navazuje dlouhá pauza a dojde ke zkreslení jedné hranice pauzy, jednak také snížené odolnosti vůči respiraci, která je svými parametry také podobná oné problematické skupině pauz. Navic lze respiraci očekávat právě v pauzách delších $150ms$, což celkovou detekci nesmírně komplikuje. Prezentovaný algoritmus vykazuje vůči této chybě vyšší odolnost, již lze pozorovat jak v úspěšnosti pro dlouhé pauzy tak v nižší směrodatné odchylce pro dlouhé pauzy. Porovnáním úspěšnosti na skupinách zdravých a dysartrických mluvčí lze soudit podobné chování pro obě skupiny u konvenční metody. Prezentovaná metoda pro jednotlivé skupiny mluvčí vykazuje jisté odlišnosti v úspěšnosti, které lze přiřknout rozdílné artikulaci fonému dysartrické skupiny. Nicméně při porovnání obou metod v tomto měřítku vyšla prezentovaná metoda jako úspěšnější.

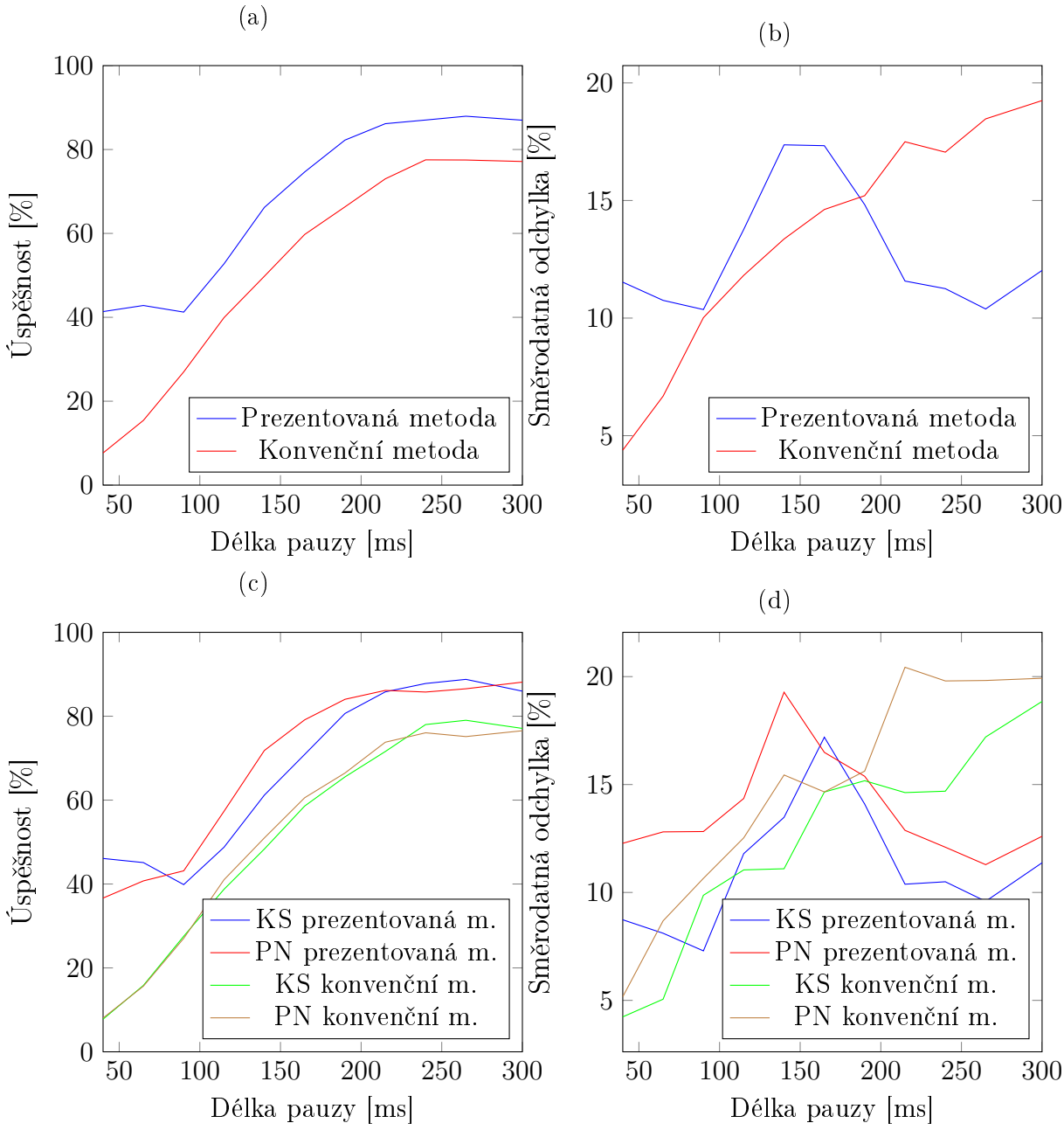
Pro detekci dechu bohužel není k dispozici potřebná srovnávací metoda, proto se mu-

síme smířit s ohodnocením bez porovnání výsledků. Na obrázku 3.2 se podívejme napřed na průměrnou úspěšnost (a). V celém hodnocení se pohybujeme v rozmezí $75 - 85\%$. Výrazným prvkem v průběhu je vrchol okolo $200ms$, který následně klesá. Tato délka odpovídá nejčetnější délce respirace[1], což se projeví ve vyšší pravděpodobnosti její detekce. Pro vyšší délky četnost detekovatelných respirací postupně klesá a navíc takto dlouhé nádechy mají poměrně neurčitě diferencované okraje. Tím lze vysvětlit nižší účinnost detekce i její vyšší směrodatnou odchylku, která od $200ms$ stoupá velice strmě. Dá se tedy předpokládat, že pro mnoho signálů jsme schopni respiraci detekovat od $200ms$ s přesností 85%, avšak některé případy tuto úspěšnost razantně snižují na směrodatnou odchylku 20% a úspěšnost 75%. Porovnáním výsledků pro zdravou a dysartrickou skupinu mluvčích máme výsledky podobné s rozdílem v nižší směrodatné odchylce dysartrické skupiny. U této skupiny lze předpokládat znělejší projev respirace, z čehož lze vyvodit tyto mírně stabilnější výsledky.

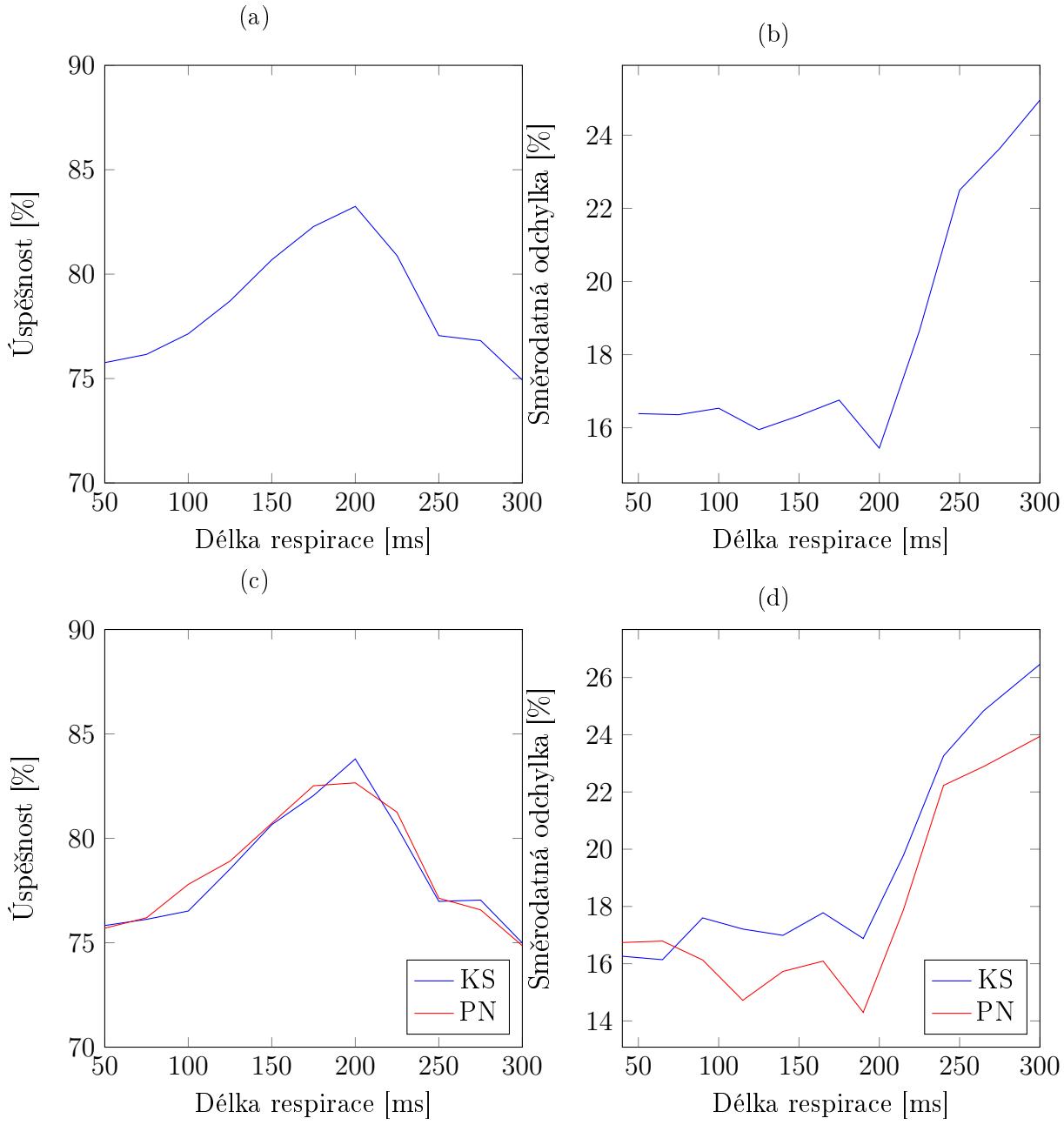
3.2 Ohodnocení příznaků

U 15 příznaků z celkového počtu 21 navržených příznaků jsme zamítli hypotézu o rovnosti středních hodnot předpokládaných normálních rozdělení příznaků pro obě skupiny mluvčích. Z těchto 15 příznaků dostálo 11 příznaků hladiny významnosti $p < 0.05$, dále 3 příznaky vykázaly hladinu $p < 0.01$ a jeden z příznaků hladinu $p < 0.001$. Nejsignifikantnějším příznakem pro odlišení dysartrické řeči PN se jeví být střední rychlost produkce pauz. Tento parametr vztažený na jednotlivé délkové skupiny vykazoval nižší hladinu významnosti. Zajímavé je, že ačkoliv tento příznak pro krátké pauzy projevil hladinu $p > 0.05$, pak z hlediska dynamiky produkce pauz jako akcelerace produkce pauz mezi prvou a druhou polovinou promluvy projevil hladinu $p < 0.05$. Je jasné, že na hypokinetickou dysartrii je nutné pohlížet jako na dynamickou poruchu produkce řeči. Tato dynamika se v příznací mnohem lépe projevila v příznaku akcelerace rytmu artikulace na hladině $p < 0.01$. Rychlost rytmu tentokráte dlouhých pauz dosáhla hladiny významnosti $p < 0.01$. Vyšší střední hodnota rychlosti rytmu pro skupinu PN svědčí o členění promluvy v delší celky a tedy o pomalejší produkci řeči. Hlubší průzkum rytmických aspektů hypokinetické dysartrie by si jistě dobudoucna zasloužil svoji pozornost. Příznak poměr pauzy a řeči poukazuje na vyšší vzácnost pauz v dysartrické řeči z hlediska celkové doby v promluvě. Z hlediska počtu pauz je tím příznakem celkový počet pauz, který pro skupiny středních pauz $< 50; 100 > ms$ nabyl hladiny významnosti $p < 0.01$. V celkovém počtu jsou signifikantně více zastoupeny krátké pauzy $< 50ms$. Z výsledku již zmiňovaných příznaků lze tušit souvislost s naopak nižším zastoupením středních pauz v dysartrické řeči. Oba hodnocené příznaky respirace vyšly na hladině významnosti < 0.05. Vyšší zastoupení respirace v dlouhých pauzách a nízké hodnoty poměru rytmicity respirace a dlouhých pauz uvádí v potaz souvislost mezi postižením řeči a respiračním aparátem PN. S ohledem na malou délku zkoumaných signálů vdatabázi není respirace zastoupena v dostatečném počtu vzorků na signál a pro delší nahrávky by se jistě projevily i další zajímavé příznaky, nicméně i přes nízký počet vzorků respirace v signálu se podařilo nalézt statisticky

[1] nejedná se o délku samotného nádechu, ale o délku nádechu, kterou lze jednoznačne identifikovat v signálu nahrávky



Obrázek 3.1: **a)** průměrná úspěšnost **prezentované** (modrá) a **konvenční** (červená) metody přes celou databázi **b)** směrodatné odchylky úspešnosti **prezentované** (modrá) a **konvenční** (červená) metody přes celou databázi **c)** průměrná úspešnost **prezentované** metody ve skupině KS (modrá) a skupině PN (červená) a pro **konvenční** metodu ve skupině KS (zelená) a skupině PN (hnědá) **d)** směrodatná odchylka úspešnosti **prezentované** metody ve skupině KS (modrá) a skupině PN (červená) a pro **konvenční** metodu ve skupině KS (zelená) a skupině PN (hnědá)



Obrázek 3.2: **a)** průměrná úspěšnost **detekce respirace** prezentované metody přes celou databázi. **b)** směrodatná odchylka úspěšnosti **detekce respirace** prezentované metody přes celou databázi **c)** průměrná úspěšnost **detekce respirace** prezentované metody ve skupině KS (modrá) a skupině PN (červená) **d)** směrodatná odchylka úspěšnosti **detekce respirace** prezentované metody ve skupině KS (modrá) a skupině PN (červená)

významné příznaky postižení respirace. V tabulce 3.1 máme shrnuté navržené příznaky. Pro skupiny zdravých *KS* i dysartrických mluvčí *PN* jsou zvlášť vyneseny dosažené průměrné

hodnoty i směrodatné odchylky. V posledním sloupci lze dohledat i dosaženou hladinu významnosti.

	KS		PN		
Zkratka příznaku	Průměr	SO	Průměr	SO	Hladina významnosti
pPS	0.49	0.09	0.43	0.11	p<0.05
nP	73.32	10.66	64.09	13.82	p<0.05
nPk	32.68	11.29	34.5	10.17	p=0.58
nPs	30.91	4.46	25.77	6.72	p<0.01
nPd	28.32	7.86	23.64	10.62	p=0.10
zPk	0.44	0.12	0.56	0.18	p<0.05
zPs	0.34	0.07	0.31	0.07	p=0.09
zPd	0.31	0.06	0.28	0.1	p=0.26
mVS	13.88	1.52	12.99	1.7	p=0.08
mVP	1.83	0.27	1.48	0.36	p<0.001
mVPk	1.01	0.35	1.05	0.35	p=0.7
mVPs	0.97	0.22	0.79	0.26	p<0.05
mVPd	0.87	0.2	0.69	0.24	p<0.05
aVP	-0.42	0.35	-0.2	0.33	p<0.05
aVPk	-0.33	0.35	-0.12	0.26	p<0.05
aVPs	-0.21	0.3	-0.09	0.27	p=0.17
aVPd	0.13	0.19	0.02	0.15	p<0.05
ppP	1.11	0.11	1.02	0.13	p<0.05
srP	1.19	0.28	1.57	0.57	p<0.01
arA	0.18	0.16	0.07	0.1	p<0.01
pRP	0.13	0.06	0.19	0.09	p<0.05
prR	2.49	1.77	1.57	0.96	p<0.05

Tabulka 3.1: Výsledné ohodnocení příznaků průměrem, směrodatnou odchylkou *SO* a hladinou významnosti pro skupiny *KS* a *PN*

Kapitola 4

Diskuze

Přínosy této práce jsou jednak metoda pro ohodnocení dysartrické řeči jednak sada příznaků pro popis dysartrické řeči. Zaměřme se nyní na prezentovanou metodu. Metoda podávala velmy dobré výsledky pro dlouhé pauzy $> 200ms$ a krátké pauzy $< 100ms$. Velkým jejím kladem byla odolnost vůči chybné klasifikaci respirace za řečový signál a navíc schopnost ji zvlášť detekovat jako samostatnou třídu. Pokud se blíže zamyslíme nad neduhem této metody, tedy pauzy $< 100; 200 > ms$, pak problém tkví ve schopnosti odlišit šum konsonant od šumového pozadí a šumu respirace. Tento problém byl řešen v oblasti spektra a dostál lepších výsledků než konvenční metoda, nicméně určitě by si dobudoucna zasloužil rozšířit jak parametrický prostor pro rozhodování tak i robustnější vyhlazení rozhodování založené především na kontextu. Jednou z možných cest pro zlepšení klasifikace by bylo rozklad signálu vlnkovou transformací či kombinace více prostorů pro popis této skupiny fonémů. Takový přístup by vedl k poměrně složité rozhodovací úloze, nicméně nezbytné, jelikož žádný samotný parametr či prostor více parametrů není schopen tento problém vyřešit. Zvýšení rozhodovací účinnosti pro tuto skupinu pauz by jistě zvýšilo účinnost detekce respirace, jelikož respiraci a signálem těchto pauz existuje velká podobnost v mnoha parametrech e.g. výkon, spektrum. Samotný detektor respirace byl schopen poměrně kvalitně rozlišit obsah dlouhých pauz na pauzu a respiraci, výsledek je však velmi podmíněn právě zmíněnou kvalitou detekce dlouhých pauz, které jsou zatíženy kvalitou detekce oné problematické skupiny fonémů. Budoucí rozvoj metody by se měl jistě ubírat směrem k vyřešení tohoto problému.

Podíváme-li se blíže na hladiny významnosti navržených příznaků, lze soudit, že pro odlišení dysartrické řeči můžeme považovat sníženou rychlost produkce pauz za nejsignifikantnější parametr, který v sobě zahrnuje počet pauz i délku promluvy, na kterou je vztažený. Vyjadřuje tedy i celkovou pomalost řeči *hypokinetické dysartrie*. Tuto pomalost jsme do určité míry popsali také parametrem střední rytmus pauz. Pro skupinu *PN* vyšly střední vzdálenosti mezi dlouhými pauzami významě delší, což vezmeme-li v potaz účel dlouhých pauz - tedy respirace a formální členění obsahu, vypovídá o pomalejším vedení promluvy a tedy *hypokinetické dysartrii*. Prozkoumáme-li blíže zastoupení respirace v dlouhých pauzách, pak skupina *PN* vykázala signifikantně větší výskyt respirace. Na takový výsledek může mít vliv rigidita dýchacího svalstva, ale také e.g. rigidita hrtanu a bradykineze řečového aparátu, neboť schopnost detekce respirace je podmíněna turbulentním prouděním v dýchacím apa-

rátu a vyšší výskyt detekovatelné respirace závisí jak na vyšším výskytu dechu tak i na jeho znělosti stejnou měrou. Za předpokladu vyššího zastoupení lze u *PN* předpokládat ovlivnění rytmu dlouhých pauz rytmem respirace. Podíváme-li se na rytmickou stránku promluvy i z hlediska zrychlení řeči, pak *PN* skupina projevila sklony ke zvýšení rychlosti artikulace zvýšenou produkcí krátkých pauz, o čemž svědčí statisticky nevýznamná rychlost pauz zároveň s významným zrychlením produkce krátkých pauz. K úvahám o zrychlování také přispívá výsledek akcelerace rytmu artikulace, jež vyšel významě nižší a znamená tedy vyšší zrychlení artikulace pro skupinu *PN*. Zrychlení produkce středních pauz $< 50; 100 > ms$ vyšlo statisticky nevýznamné. S ohledem na vyšší zastoupení krátkých pauz $< 50ms$ se lze přiklonit k doměnce o zrychlování řeči na úkor zkrácování pauz [37]. Pokud tedy k takovému zrychlování řeči dochází, pak by bylo zajímavé provést hlubší analýzu artikulačního zrychlení. Pokud by bylo zrychlení především na úkor zkracování pauz a ne na úkor zkracování samotné artikulace, mohli bychom velmi dobře posoudit postižení samotného artikulačního aparátu *PN*.

Příloha A

Obsah CD

/TEXT obsahem adresáře je elektronická verze diplomové práce *D.pdf*

/HELP zde naleznete dokumentaci k obsahu disku, jednotlivých adresářů i implementovaných funkcí v soubodu *dokumentace.pdf*

/METODA obsahuje dílčí hlavní skripty

/METODA/labels pracovní adresář se značkami klasifikace v souborech *name.mat*

/METODA/PATH pracovní adresář algoritmu

/METODA/PATH/bayes soubory výstupů *Bayesovského autoregresního detektoru změn* uložené v souborech *name.mat*

/METODA/PATH/database uchovává nahrávky mluvčích v souborech *name.wav* a ručních značek v souborech *name.txt*

/METODA/PATH/database/breath zde jsou shromážděny značeky respirace v souborech *name_breath.txt*

/METODA/PATH/m-scripts zde jsou uloženy pomocné funkce algoritmu

Každému mluvčímu v databázi přísluší unikátní jméno *name*, složené z návěstí, čísla a pořadového písmene nahrávky. *HC* v návěstí popisuje zdravého a *PN* dysartrického mluvčího.

Literatura

1. Vokurka, M., Hugo, J., et al. (2006) Velký lékařský slovník, 6.vydání, Jessenius, Maxdorf, 689.

2. Aosaki, T., Miura, M., Suzuki, T., Nishimura, K., and Masuda, M. (2010) Acetylcholine–dopamine balance hypothesis in the striatum: An update, Geriatrics and Gerontology International, Special Issue: Gerontology and Geriatrics Science: Gene to Longevity 10, 148–157.

3. Marsden, C. D., Schneider, S., Bhatia, K., and Donaldson, I. (2012) Marsden's Book of Movement Disorders, Oxford University Press, 176-177.

4. Fearnley, J. M. and Lees, A. J. (1991) Ageing and parkinson's disease: substantia nigra regional selectivity, Brain 114, 2283-2301.

5. Marsden, C. D. (1992) Parkinson's Disease, Postgraduate Medical Journal 68, 538-543.

6. Emreand, M., Aarsland, D., Brownand, R., Burn, D. J., Duyckaerts, C., Mizunoand, Y., Broe, G. A., Cummings, J., Dickson, D. W., Gauthier, S., Goldman, J., Goetz, C., Korczyn, A., Lees, A., Levy, R., Litvan, I., McKeith, I., Olanow, W., Poewe, W., Quinn, N., Sampaio, C., Tolosa, E., and Dubois, B. (2007) Clinical diagnostic criteria for dementia associated with parkinson's disease, Movement Disorders 22, 1689-1707.

7. Schechtman, E., Paleacu, D., and Inzelberg, R. (2002) Onset age of Parkinson Disease, American Journal of Medical Genetics 111, 459–460.

8. Golbe, L. I. (1991) Young onset parkinson's disease a clinical review, Neurology 41, 168-173.

9. Rajput, A. and Birdi, S. (1997) Epidemiology of Parkinson's disease, Parkinsonism and Related Disorders 3, 175–186.

10. deRijk, M. C., Tzourio, C., Breteler, M. M., Dartigues, J. F., Amaducci, L., Lopez-Pousa, S., Manubens-Bertran, J. M., Alpérovitch, A., and Rocca, W. A. (1997) Prevalence of parkinsonism and parkinson's disease in Europe: the EUROPARKINSON Collaborative Study. European Community Concerted Action on the Epidemiology of parkinson's disease, Journal of Neurology, Neurosurgery and Psychiatry 62, 10–15.

11. Dorsey, E. R., Constantinescu, R., Thompson, J. P., M.Biglan, K., Holloway, R. G., Kieburtz, K., Marshall, F. J., Ravina, B. M., Schifitto, G., Siderowf, A., and Tanner, C. M. (2006) Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030, Neurology 68, 384-386.

12. Rajput, A. H. (1992) Frequency and cause of parkinson's disease, Canadian Journal of Neurological Sciences 19, 103-107.

13. Willis, A. W., Schootman, M., Kunga, N., Evanoff, B. A., Perlmutter, J. S., and Racette, B. A. (2012) Predictors of Survival in Parkinson's Disease, Archives of Neurology 69, 601–607.

14. Goberman, A. M. (2005) Correlation between acoustic speech characteristics and non-speech motor performance in Parkinson disease, Medical Science Monitor 11, 109-116.

15. National Parkinson Foundation, I. N. C. (2013) Parkinson's Disease, www.parkinson.org, .

16. Rintala, D. H., Tan, G., Willson, P., Bryant, M. S., and Lai1, E. C. H. (2010) Feasibility of Using Cranial Electrotherapy Stimulation for Pain in Persons with Parkinson's Disease, Journal of Parkinson's Disease, ID 569154.

17. Jankovic, J. (2007) Parkinson's disease: clinical features and diagnosis, Journal of Neurology, Neurosurgery, and Psychiatry 79, 368-376.

18. Hely, M. A., Reid, W. G., Adena, M., Halliday, G. M., and Morris, J. G. (2008) The Sydney multicenter study of parkinson's disease: the inevitability of dementia at 20 years, Movement Disorders 23, 837-844.

19. Darley, F. L., Aronson, A. E., and Brown, J. R. (1975) Motor speech disorders, Saunders, Philadelphia, PA, 250.

20. Goberman, A. M. and Coelho, C. (2002) Acoustic analysis of Parkinsonian speech I:speech characteristics and L-dopa therapy, Neurorehabilitation 17, 237-246.

21. Critchley, E. M. R. (1981) Speech disorders of Parkinsonism: a rewiev, Journal of Neurology, Neurosurgery and Psychiatry 44, 751-758.

22. Rusz, J., Čmejla, R., Růžičková, H., and Růžička, E. (2011) Quantitative acoustic measurements for characterisation of voice and speech disorders in early untreated Parkinson's disease, The Journal of the Acoustical Society of America 129, 350-367.

23. Apps, M. C. P., Sheaff, P. C., Ingram, D. A., Kennard, C., and Empey, D. W. (1985) Respiration and sleep in parkinson's disease, Journal of Neurology, Neurosurgery and Psychiatry 48, 1240-1245.

24. Perez, K. S., Ramig, L. O., Smith, M. E., and Dromey, C. (1996) The Parkinson larynx: Tremor and videostroboscopic findings, Journal of Voice 10, 354–361.

25. Weismer, G. (2005) Articulatory characteristics of Parkinsonian dysarthria: segmental and phase level timing, spirantization, amd glottal-supraglottal coordination, The dysarthrias: Physiology, Acoustic, Perception, Management, edited by Mcneil M, Rosenbeck J, and Aronson A, Colege-Hill Press, San Diego, CA, 101-130.

26. Rusz, J., Čmejla, R., Růžičková, H., Růžička, E., Klempíř, J., Majerová, V., Picmausová, J., and Roth, J. (2011) Acoustic assesment of voice and speech disorders in Parkinsonian's disease through quick vocal test, Movement Disorders 26, 1951-1952.

27. Hunker, C., Abbs, J., and Barlow, S. (1982) The relationship between parkinsonian rigidity and hypokinestia in the orofacial system: a quantitative analysis, Neurology 32, 749-754.

28. Seibel, L., Barlow, S., Hammer, M., Prasad, S., and Pahwa, R. (2002) Stiffness of the human lips in parkinson's disease, Abstracts for the American Speech-Hearing-Language Association Atlanta, GA, -.

29. Hammen, V. L. and Yorkston, K. M. (1996) Speech and pause characteristics following speech rate reduction in hypokinetic dysarthria, Journal of Communication Disorders 29, 429-444.

30. Goberman, A. M. and McMillan, J. (2005) Relative speech timing in Parkinson's disease, Contemporary issues in communication's science and disorders 32, 22–29.

31. Ackermann, H. and Ziegler, W. (1991) Articulatory deficits in Parkinsonian dysarthria: an acoustic analysis, Journal of Neurology, Neurosurgery and Psychiatry 54, 1093-1098.

32. Ho, A. K., Iansek, R., Marigliani, C., Bradshaw, J., and Gates, S. (1998) Speech impairment in large sample of pacients with Parkinson's disease, Behavioural Neurology 11, 131-137.

33. Logermann, J. A., Fisher, H. B., Boshes, B., and Blonsky, E. R. (1978) Frequency and coocurence of vocal tract dysfunction in speech of large sampleof Parkinson patients, Journal of Speech and Hearing Disorders 43, 47-57.

34. Rusz, J., Čmejla, R., Růžičková, H., Růžička, E., Klempíř, J., Majerová, V., Picmausová, J., and Roth, J. (2012) Evaluation of speech impairment in early stages of Parkinson's disease: a prospective study with the role of pharmacotherapy, Journal of Neural Transmission 120, 319-329.

35. Canter, G. J. (1963) Speech characteristics of pacient with Parkinson's disease:I. Intensity, pitch and duration, Journal of Speech and Hearing Disorders 28, 221-229.

36. Darley, F. L., Aronson, A. E., and Brown, J. R. (1969) Differential diagnostic patterns of dysarthria, Journal of Speech Language and Hearing Research 12, 246-269.

37. Skodda, S. and Schlegel, U. (2008) Speech rate and rhythm in Parkinson's disease, Movement Disorders 23, 985-992.

38. Ackermann, H., Hertrich, I., Grodd, W., and Wildgruber, D. (2004) Gefühlen: Funktionell-neuroanatomische Grundlagen der Verarbeitzung affektiver Prosodie, Aktuelle Neurologie 31, 449-460.

39. Parkinson, J. (1817) An Essay on the Shaking Palsy, Sherwood, Neely and Jones, Paternoster Row, 11-15.

40. Tjaden, K. and Wilding, G. (2011) Speech and pause characteristics associated with voluntary rate reduction in Parkinson's disease and Multiple Sclerosis, Journal of Communication Disorders 44, 655-665.

41. Kim, C. and Stern, R. M. (2008) Robust Signal-to-Noise Ratio Estimation Based on Waveform Amplitude Distribution Analysis, INTERSPEECH, 2598-2601.

42. Gazor, S. and Zhang, W. (2003) Speech Probability Distribution, Signal Processing Letters, 2003 IEEE, 204-207.

43. Green, J. R., Beukelman, D. R., and Ball, L. J. (2004) Algorithmic Estimation of Pauses in Extended Speech Samples of Dysarthric and Typical Speech, Journal of medical speech-language pathology 12, 149 - 154.

44. Ying, D., Yan, Y., Dang, J., and Soong, F. K. (2011) Voice Activity Detection Based on an Unsupervised Learning Framework, Audio, Speech, and Language Processing, 2011 IEEE Transactions on, 2624-2633.

45. Kinnunen, T. and Rajan, P. (2013) A Practical, Self-Adaptive Voice Activity Detector For Speaker Verification With Noisy Telephone And Microphone Data, Acoustics, Speech, and Signal Processing (ICASSP), 2013 IEEE International Conference on, .

46. Pollak, P., Sovka, P., and Uhlir, J. (1993) Noise suppression system for a car, Eurospeech. Vol. 93, 1073-1076.

47. Harrison, W., Lim, J. S., and Singer, E. (1986) A new application of adaptive noise cancellation, Acoustics, Speech, and Signal Processing, 1986 IEEE Transactions on, 21-27.

48. Sakhnov, K., Verteletskaya, E., and Simak, B. (2009) Approach for Energy-Based Voice Detector with Adaptive Scaling Factor, IAENG International Journal of Computer Science, 36(4).

49. Kajita, S. and Itakura, F. (1994) Speech analysis and speech recognition using subband-autocorrelation analysis, Acoustics, Speech, and Signal Processing. ICASSP-94., 1994 IEEE International Conference on, 193-196.

50. Wu, B.-F. and Wang, K.-C. (2006) Voice Activity Detection Based on Auto-Correlation Function Using Wavelet Transform and Teager Energy Operator, Computational Linguistics and Chinese Language Processing 11, 87 - 100.

51. Misra, H., Bourlard, S. I. H., and Hermansky, H. (2004) Spectral Entropy as Speech Features For Speech Recognition, Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on, 193-196.

52. Shen, J. L., Hung, J. W., and Lee, L. S. (1998) Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments, ICSLP, 232-235.

53. Renevey, P. and Drygajlo, A. (2001) Entropy based voice activity detection in very noisy conditions, INTERSPEECH, 1887-1890.

54. Cavallaro, A., Beritelli, F., and Casale, S. (1998) A fuzzy logic-based speech detection algorithm for communications in noisy environments, Acoustics, Speech, and Signal Processing. Proceedings of the 1998 IEEE International Conference on, 565-568.

55. Haigh, J. A. and Mason, J. S. (1993) Robust voice activity detection using cepstral features, TENCON - Proceedings. Computer, Communication, Control and Power Engineering, 321 - 324.

56. Singh, R., Seltzer, M. L., Raj1, B., and Stern, R. M. (2001) Speech in noisy environments: robust automatic segmentation, feature extraction, and hypothesis combination, Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on (Vol. 1), 273-276.

57. Čapek, K. (1988) Spisy: Od člověka k člověku I, Československý spisovatel, 230.

58. Čmejla, R., Rusz, J., Bergl, P., and Vokřál, J. (2013) Bayesian changepoint detection for the automatic assessment of fluency, and articulatory disorders, Speech Communication 55, 178-189.

59. Rosen, K., Murdoch, B., Folker, J., Vogel, A., Cahill, L., Delatycki, M., and Corben, L. (2010) Automatic method of pause measurement for normal and dysarthric speech, Clinical Linguistics and Phonetics 24, 141 - 154.