**ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE**

**Fakulta Elektrotechnická**

**Katedra radioelektroniky**

# Realizace algoritmu pro segmentaci promluv pacientů trpících Huntingtonovou nemocí

# Design of algorithm for segmentation of speech utterances in patients with Huntington's disease

Diplomová práce

Studijní program: Komunikace, Multimédia a Elektronika
Studijní obor: Multimediální technika

Vedoucí práce: Ing. Jan Rusz, Ph.D.

**Bc. Jakub Pospíšil**

**Praha 2015**

## Čestné prohlášení

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 4. 5. 2015 ……

Bc. Jakub Pospíšil

## Poděkování

Na tomto místě bych rád poděkoval především Ing. Michalovi Novotnému za jeho čas, spolupráci, vstřícný přístup a cenné rady, které mi poskytoval v průběhu zpracování celé diplomové práce.

České vysoké učení technické v Praze
Fakulta elektrotechnická

katedra radioelektroniky

# ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: **Jakub Pospíšil**

Studijní program: Komunikace, multimédia a elektronika
Obor: Multimediální technika

Název tématu: **Realizace algoritmu pro segmentaci promluv pacientů trpících Huntingtonovou nemocí**

Pokyny pro vypracování:

1. Seznamte se s poruchami řeči u Huntingtonovy nemoci a metodami zpracování řečové aktivity.
2. Na základě dostupné literatury navrhněte parametrizaci diadochokinetických (DDK) promluv (rychlé opakování slabik /pa/-/ta/-/ka/) pacientů trpících Huntingtonovou nemocí.
3. Navrženou parametrizaci využijte v návrhu algoritmu segmentace DDK promluv. Návrh algoritmu realizujte ve výpočetním prostředí MATLAB.
4. Vytvořte ruční reference a navržený algoritmus otestujte na vybraném vzorku promluv zdravé populace a pacientů s výskytem Huntingtonovy nemoci. Výsledky porovnejte s konvenčním algoritmem pro segmentaci DDK promluv u pacientů s Parkinsonovou nemocí.
5. Do algoritmu implementujte hodnocení vhodných řečových příznaků pro popis charakteristik dysartrie a proveďte jednoduché statistické testy pro odlišení zdravých mluvčích od pacientů s Huntingtonovou nemocí.

Seznam odborné literatury:

[1] Walker, F. O.: Huntington's disease, Lancet, 369, pp. 218-228, 2007.
[2] Duffy, J. R.: Motor Speech Disorders: Substrates, Differential Diagnosis and Management, 2nd ed., Mosby, New York, 2005.
[3] Rusz J, Klempíř J, Tykalová T, Baborová E, Čmejla R, Růžička E, Roth J.: Characteristics and occurence of speech impairment in Huntington's disease: possible influence of antipsychotic medication. J. Neural. Transm. 121, pp. 1529-1539, 2014.
[4] Novotný, M., Rusz, J.,Čmejla, R., and Růžička, E.: Automatic evaluation of articulatory disorders in Parkinson's disease. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 22, 1366-1378, 2014.

Vedoucí: Ing. Jan Rusz, Ph.D.

Platnost zadání: do konce letního semestru 2015/2016


doc. Mgr. Petr Páta, Ph.D.
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 10. 2. 2015

**Anotace:**

Tato diplomová práce se zabývá problematikou diagnostiky hypokinetiské dysartrie jako prvotní příznak Huntingtonovi nemoci (HN). K vyhodnocování kvality řečového aparátu jsou používány řečové diadochokinetické (DDK) úlohy, založené na rychlém opakování slabik /pa/-/ta/-/ka/. Hlavním tématem je realizace algoritmu pro segmentaci patologických promluv pacientů trpící touto nemocí. Metoda předpokládá, že řeč obsahující explozívy, vokály a části bez řečové aktivity lze považovat za multimodální směs normálních rozdělení parametrů počtu průchodů nulou, spektrální entropie, vlnové transformace či rozptylu autokorelační funkce, tzv. směs Gaussovských rozdělení (Gaussian Mixture Model – GMM). Pro klasifikaci parametrů je využito GMM-algoritmu. Výstupem algoritmu jsou hranice jednotlivých explozív, vokálů a částí bez řečové aktivity. Dále jsou segmenty hodnoceny vhodnými řečovými příznaky, podle kterých jsou odlišeny nahrávky zdravého člověka od pacienta s HN.

**Klíčová slova:**

Huntington, GMM, dysartrie, /pa/-/ta/-/ka/, MATLAB.

**Summary:**

This thesis deals with problem of diagnosis hypokinetic dysarthria in Huntington's disease (HD). To evaluate the quality of the speech apparatus are used speech diadochokinetic (DDK) tasks, based on the repetition of syllables /pa/-/ta/-/ka/. Main aim is implementation of the algorithm for segmentation pathological utterances of patients suffering from HD. The method assumes that speech comprising plosives, vocals and parts without speech activity is considered to be multi-modal mixture normal distribution of zero-crossing rate, spectral entropy, wavelet transform, or variance of autocorrelation function, ie. Gaussian Mixture Model – GMM. The method sequentially estimates parameters of individual classes using the GMM-algorithm. Plosives, vocals, and parts without speech activity boundaries are outputs of the algorithm. Segmented utterances are evaluated appropriate speech symptoms. These symptoms distinguish records of healthy people form patients with HD.

**Index Terms:**

Huntington, GMM, dysarthria, /pa/-/ta/-/ka/, MATLAB.

# Obsah

# Seznam použitých zkratek

BSCD .. Bayesův detektor změn (Beyesian Step Changepoint Detector)
CSM .... spektrální moment konsonant (Consonant Spectral Moment)
CST ..... trend spektra konsonant (Consonant Spectral Trend)
DDK .... diadochokinetická úloha
DFT ..... diskrétní Fourierova transformace (Discrete Fourier Transform)
DP ..... dolní propust
E..... energie signálu
fs ..... vzorkovací kmitočet
GMM ... model Gaussovských směsí (Gaussian Mixture Model)
HC..... kontrolní skupina zdravých lidí (Healthy Control)
HD..... skupina pacientů s Huntingtonovou nemocí (Huntington Disease)
HE..... Hilbertova obálka (Hilbert Envelope)
HN ..... Huntingtonova nemoc
HP..... horní propust
Hx..... spektrální entropie určitého kmitočtového pásma (x – 1 až 15)
IB..... počáteční exploze (Initial Burst)
IDFT .... inverzní DFT
LPC ..... lineární prediktivní kódování (Linear Predictive Coding)
MFCC.. Melovské frekvenční kepstrální koeficienty (Mel-Freq. Cepst. Coef.)
O ..... okluze (Occlusion)
P..... výkon signálu
PDF..... hustota pravděpodobnosti (Probability Density Dunction)
PLP ..... percepční predikční koeficienty (Perceptual Linear Prediction)
RASTA RASTA koeficienty (Relative Spectra PLP)
SE ..... spektrální entropie (Spectral Entropi)
SNR .... odstup signálu od šumu (Signal-to-Noise Ratio)
STFT ... krátkodobá Fourierova transformace (Short Time Fourier Transform)
VO..... nástup vokálu (Vowel Onset)
VOT..... doba nástupu řečové aktivity (Voice Onset Time)
VSQ .... autokorelační poměr (Vowel Similarity Quotient)
WT ..... vlnková transformace (Wavelet Transform)
ZCR..... počet průchodů nulou (Zero Crossing Rate)

# 1 Úvod

## 1.1 Huntingtonova nemoc

Huntingtonova nemoc (dále HN) je autosomálně dominantně dědičné, neurodegenerativní onemocnění centrálního nervového systému. Nemoc obvykle vypuká u dospělých lidí ve středním věku, ale symptomy se mohou projevit kdykoliv v době od raného dětství přibližně do 80 let. Narušený centrální nervový systém má za následek abnormální svalové pohyby nebo nepravidelnosti v řečovém systému. Zjednodušeně řečeno, pacient trpící HN nedokáže „včas zastavit“ pohyby při chůzi či gestikulaci, jeho kroky jsou nepravidelně dlouhé a celková chůze může připomínat kulhání. Charakteristické jsou také rytmické nebo neočekávané rychlé nebo pomalé nedobrovolné pohyby. Psychické potíže spojené s nemocí způsobují, že lidé trpící HN mají přibližně o 5-10% zvýšené sklony k sebevraždám, tyto sklony se navíc mění v průběhu onemocnění a v prediagnostickém stupni stoupá tento poměr až na 22% [1].

První zmínky o nemoci pochází již ze 14. století, kdy byla známá pouze jako „dancing mania“. Od té doby se nemocí zabývalo několik lékařů, kteří definovali původ nemoci, její dědičnou formu a v roce 1872 to byl George Huntington, kdo zásadně přispěl ke klinickému popisu nemoci. To vedlo k pojmenování choroby na *Huntingtonova nemoc*. Během dalších několika desetiletí se po celém světě začaly objevovat záznamy o HN a její rané formě. V padesátých letech dvacátého století byla objasněna struktura DNA odpovědná za vznik nemoci. Znalost genetického původu HN, poskytuje možnost zkoumat vývoj neurodegenerativních onemocnění od jejich počátku. V současné době se tedy lékaři zaměřují zejména na molekulární mechanismy nemoci. Nicméně stále neexistuje účinná léčba a všechny farmakologické terapie jsou zaměřeny pouze na zmírnění projevů onemocnění. [1]

Před nástupem prvních projevů nemoci předchází období, kdy pacient není nemocí nijak viditelně ovlivněn. S tzv. zdravým obdobím je v poslední fázi spojen predaignostický stupeň, kdy pacient začne pozorovat drobné změny chování, změny kognitivních funkcí nebo slabé poruchy koordinace. Pacient mnohdy tyto příznaky nepovažuje za významné a proto k diagnóze HN často dohází až po výrazném rozvinutí projevů nemoci.

Prediagnostický stupeň se postupně promění ve stupeň diagnózy. V tuto chvíli bývají motorické schopnosti pacienta znatelně postiženy a začínají být zřetelné choreatické pohyby. Mezi nejčastější projevy HN patří především chorea,

dysartrie, dystonie, poruchy koordinace, pokles kognitivních funkcí a změny chování.

Pro choreu jsou charakteristické prudké nekontrolované náhodné pohyby různých částí těla. Velmi podobnou poruchou je dystonie. Jedná se také o mimovolné svalové stahy, které způsobují, stejně jako u chorey, abnormální pohyby a pozice těla [2]. Z kognitivních schopností jsou oslabeny především schopnost učit se, schopnost improvizovat v nepředvídatelné situaci, schopnost plánovat nebo kontrolovat, nicméně například dlouhodobá paměť zůstává neporušená.

Jeden z dalších projevů nemoci je neschopnost udržovat dobrovolné svalové kontrakce na konstantní úrovni. Pacient v tomto případě nedokáže např. pevně sevřít dlaň a udržet v ní předměty. Tento jev je nazýván pohybová imperzistence někdy známá jako *milkmaid's grip*. Pohybová imperzistence je zcela nezávislá na choree [1]. Všechny projevy HN se s postupem času dále zhoršují.

V poslední fázi jsou poruchy pohybových i kognitivních funkcí velmi těžké a často vedou ke smrti způsobené vdechnutím tekutin při pití, deprivací, nekontrolovaným pádem nebo dysfagií (potíže při polykání), které vede k nechutenství s následky podvýživy pacienta. Typická doba mezi diagnózou a smrtí je přibližně 20 let [1].

Ačkoliv jsou motorické obtíže v běžném životě pacienta velmi omezující, z hlediska diagnostiky nemoci poskytují velké množství užitečné informace.

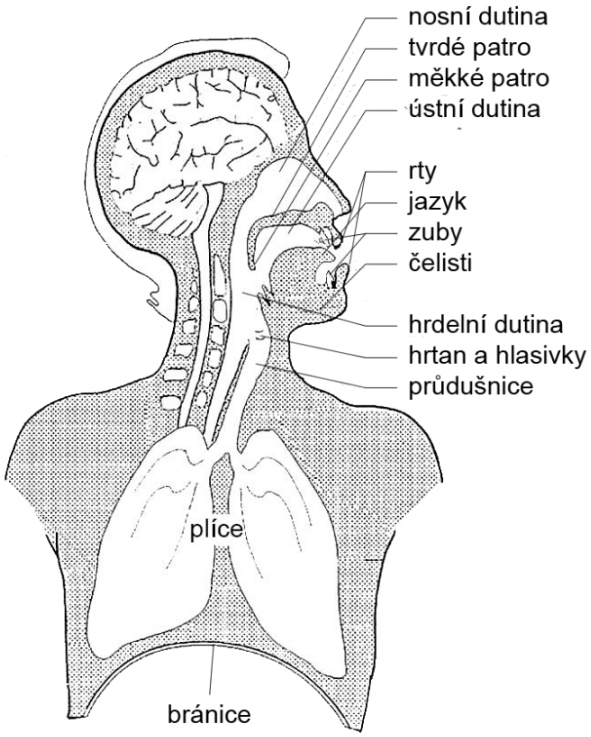
Tato práce se zabývá diagnostikou symptomu dysartrie, která se u pacientů s HN objevuje již v prediagnostickém stupni nemoci a může být prvním symptomem degenerativní nemoci.

## 1.2 Dysartrie

Lidská řeč je produkovaná prohánění vzduchu z plic přes hrdlo ovládané koordinovanými pohyby hrdelních svalů, hltan a dutinu ústní (Obrázek 1.1). Dysartrie patří do rodiny poruch řeči, způsobené narušením neuromuskulární kontroly řečových mechanizmů. Jednou z těchto neuromuskulárních patologií ústících v dysartrii je Huntingtonova nemoc. Tyto poruchy mohou být způsobeny poškozením centrálního nervového systému nebo periferního nervového systému společně s neuromuskulární ploténkou. Dysartrie postihuje všechny aspekty řeči zahrnující artikulaci, fonaci, prozódii a dýchání. Artikulace akusticky ovlivňuje zejména rezonanční vlastnosti hlasového traktu. Fonace jako parametr popisující kmitání hlasivek se projevuje ve znělých částech promluv jako narušení hlasu. Prozódie popisuje vytváření důrazů v řeči. Dýchání poukazuje na schopnost regulovat výdechový proud potřebný pro tvoření hlásek [3].

Na základě povahy řečové poruchy lze klasifikovat o jaký typ dysartrie se jedná. Jednotlivé typy dysartrie:

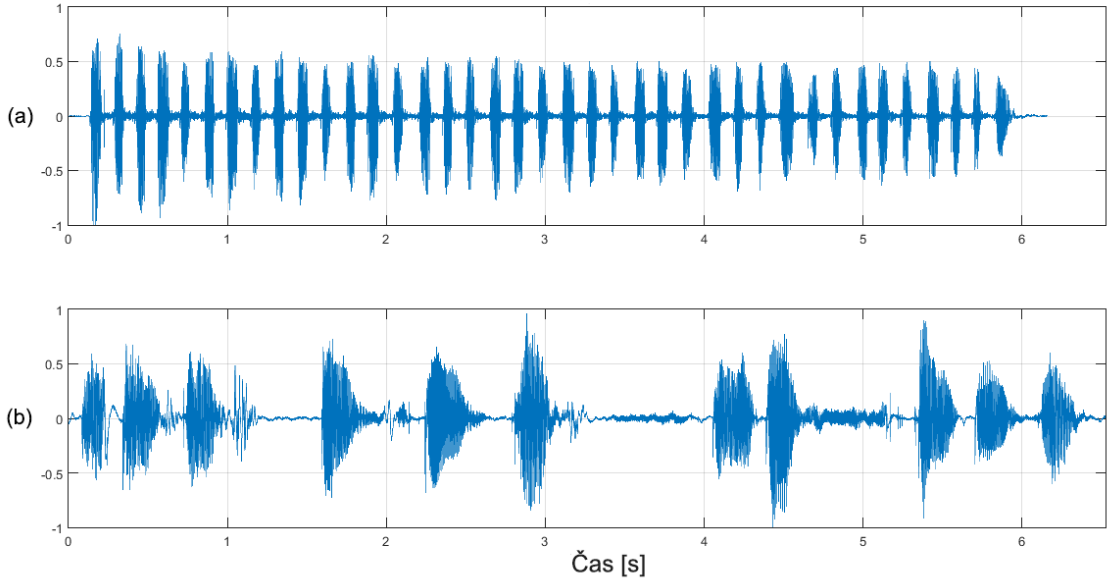
- Spastická (centrální) dysartrie (Spastic Dysarthria)
- Ataxická (cerebelární) dysartrie (Ataxic Dysarthria)
- Flacidní (periferní) dysartrie (Flaccid Dysarthria)
- Hypokinetická (extrapyramidová) dysartrie (Hypokinetic Dysarthria)
    - Parkinsonova nemoc
- Hyperkinetická (extrapyramidová) dysartrie (Hyperkinetic Dysarthria)
    - Huntingtonova nemoc




Obrázek 1.1 Řečový trakt člověka [4]

### *1.2.1 Hyperkinetická dysartrie*

U Huntingtonovy nemoci se vyskytuje hyperkinetická dysartrie, která postihuje 90% pacientů trpících HN. V řeči pacienta se v případě hyperkinetické dysartrie objevují abnormality jako je neudržením rytmu, protahování slabik, slyšitelné nádechy nebo neočekávané zvyšování hlasu v průběhu věty. Problémy s řečí jsou způsobeny především špatnou artikulací často násobenou nesprávnou koordinací dýchání [3].




Obrázek 1.2 Nahrávka (a) zdravého člověka a (b) pacienta s diagnostikovanou HN

## 1.3 Hodnocení dysartrie

Abychom dokázali charakterizovat motorické řečové poruchy, je důležité správně ohodnotit symptomy jednotlivých poruch. Existuje mnoho způsobů, které mohou být zařazeny do dvou základních skupin: percepční a instrumentální metody. Každá z těchto skupin metod má svoje klady i zápory. Přístupy hodnocení z jedné skupiny mohou být přesnější v různých případech abnormalit v jednotlivých částech řečového systému a naopak. Joseph R. Duffy ve své knize uvádí, že nejlepší variantou jsou kombinace obou skupin metod, jak percepční, tak instrumentální [2].

### *1.3.1 Využívané metody*

U percepčních metod je k hodnocení vždy potřeba zkušeného pozorovatele, který se díky sluchovým, zrakovým nebo taktilním vjemům stává platným diagnostickým nástrojem. V počátcích sedmdesátých let se Frederic L. Darley, Arnold E. Aronson a Joe R. Brown zasloužili o sluchově-percepční hodnocení dysartrie, které je používáno mnoha lékaři a vědeckými pracovníky, kteří se zajímají o akustické i fyziologické poruchy řeči [2]. I přes to, že se výzkum více zaměřuje na sluchově-percepční hodnocení, zrakové a taktilní pozorování mají při diagnostice veliký význam a nesmí být úplně přehlíženy. Přestože percepční hodnocení hraje v současnosti klíčovou roli, v případě dysartrie jako komplexu více poruch, může být pro lékaře velmi obtížné přímo hodnotit postižení jednotlivých aspektů řeči.

Další skupinou metod jsou metody instrumentální. Ačkoliv tyto metody velice pomohli k popisu a porozumění motorických řečových poruch, nejsou u lékařů příliš rozšířené. Profesor Duffy se domnívá, že je to způsobeno nedostatkem obecně uznávaných norem, dat pro řečové úlohy a parametrů pro instrumentální měření. Lékaři navíc nemají s instrumentálními metodami zkušenosti [2]. Instrumentální metody se mohou dělit do tří základních skupin:

- Akustické metody

  Tyto metody jsou velmi blízké metodám sluchově-percepčním. Základem akustických metod je řečový signál a jeho parametry, jako je periodicita, energie a další krátkodobé frekvenční i časové charakteristiky. Parametry jsou dále zpracovány a vyhodnocovány a promluvy jsou analyzovány například z pohledu obsahu chvění v hlase, přerušování řeči, změn tempa, proměnlivosti délky základní periody nebo hlasitosti, artikulace a podobně. Akustické metody jsou oblíbené, především díky jednoduchosti ovládání pro lékaře, dostupnosti a účinnosti. Elementární analýza řeči lze provádět pouhým zobrazením signálu. V dnešní době jsou nejčastěji používány metriky měřící rytmus řeči, založené na měření délky vokálů a konsonant. Tato práce spadá do skupiny akustických metod.

- Fyziologické metody

  Předchozí metody se zabývaly především akustickým výstupem řečového traktu. Fyziologická metoda se posouvá více směrem ke zdroji aktivit, který řídí a kontroluje samotnou řeč. Je zaměřena hlavně na hodnocení ovládání svalů a souvislosti mezi centrální a periferní nervovou biomechanickou aktivitou. Mezi nejčastěji používané nástroje patří magnetická rezonance (MRI), pozitronová emisní tomografie (PET) nebo elektroencefalografie (EEG).

- Metody založené na vizuálním hodnocení funkce řečového traktu

  Základem těchto metod je pozorování horní části řečového traktu (rty, jazyk, hlasivky, hrdlo a průdušnice) během mluvení. Nejčastěji používané metody jsou doprovázeny videofluoroskopií (vyšetření polykacího traktu), nosní endoskopií, nebo laryngoskopií (vyšetření hrtanu).

### 1.3.2 Řečové úlohy

Abychom mohli provádět jak percepční, tak i akustické metody hodnocení, je potřeba předem připravit řečové úlohy, které bude pacient během vyšetření vykonávat. Tyto úlohy můžeme rozdělit do pěti kategorií [5]:

- Spontánní řeč

  Pacientovi je např. položena otázka *„Jak jste trávil včerejší dopoledne?“* a je na něm, jak dlouho a o čem bude mluvit.

- Opakování

  Po pacientovy je požadováno aby opakoval sadu slabik, např. /*pa*/-/*pa*/-/*pa*/, /*pa*/-/*ta*/-/*ka*/. Toto opakování je známé jako diadochokinetická (DDK) úloha.

- Čtení připraveného textu

  Pro tyto účely jsou připraveny foneticky bohaté věty, které pacient předčítá.

- Udržení rytmu

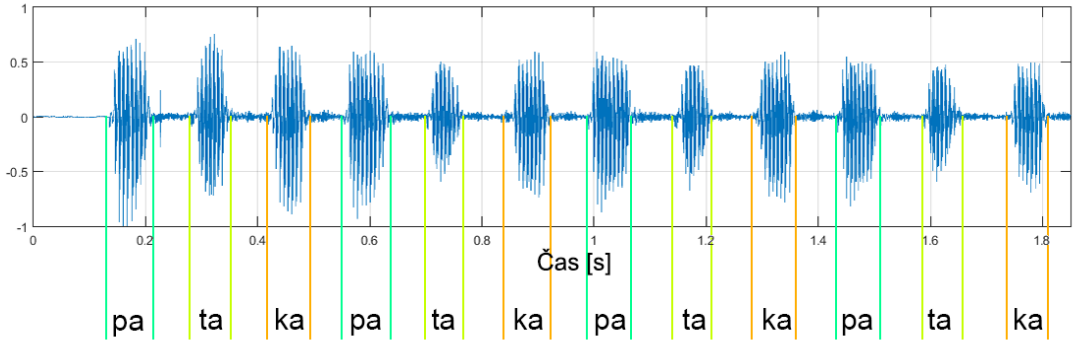
  Na začátku tohoto cvičení lékař zapne metronom, pacient ve stejném rytmu opakuje např. jednu slabiku /*pa*/. Metronom je poté vypnut a pacient se snaží co nejdéle udržet stejný rytmus.

- Prodloužená fonace

  Při této úloze je pacient žádán o co nejdelší udržení hlásky *„á“*. Pozoruje se poté kolísání kmitočtu hlasivek.

#### 1.3.2.1 Diadochokinetické úlohy DDK (Diadochokinetic Tasks)

K hodnocení hyperkinetické dysartrie se mimo jiné používá řečová diadochokinetické (DDK) úloha. Ta je založená na co nejrychlejším opakování slabik /*pa*/-/*ta*/-/*ka*/. Tato úloha je postavená tak, aby rovnoměrně zatěžovala hlasový trakt. A to díky artikulaci obouretné okluzivy /*p*/, předodásňového /*t*/ a měkkopatrového /*k*/. DDK úlohy jsou obvykle hodnoceny dvěma parametry. Průměrná rychlost jako počet slabik za sekundu a pravidelnost, která měří míru variance mezi jednotlivými slabikami [6]. Navíc DDK úlohy ukazují výrazný potenciál pro potřeby hodnocení artikulačních obtíží právě u různých dysartrických profilů.




Obrázek 1.3 DDK úloha: /pa/-/ta/-/ka/

## 1.4 Cíle práce

Cílem této práce je na základě získaných znalostí o promluvách pacientů trpících Huntingtonovou nemocí navrhnout algoritmus pro segmentaci jednotlivých událostí v nahrávkách diadochokinetických (DDK) úloh. V souvislosti s tímto úkolem je třeba navrhnout efektivní parametrizaci signálu a následnou segmentaci s využitím dynamického prahování. Návrh celého algoritmu bude realizován ve výpočetním prostředí MATLAB. Za účelem ověření funkčnosti návrhu je nutné provést experiment pro porovnání automatických detekcí s ručně nalezenými značkami. Výsledky navrženého algoritmu budou poté porovnány s výsledky konvenčních přístupů zabývajících se problematikou segmentace záznamů mluvčích s Parkinsonovou nemocí. Pro segmentované promluvy je potřeba navrhnout hodnocení vhodných řečových příznaků pro popis charakteristik dysartrie spojené se statistickými testy pro odlišení zdravých mluvčích od pacientů s Huntingtonovou nemocí.

# 2 Metodika

## 2.1 Data

### *2.1.1 Mluvčí*

Databáze představující Huntingtonovu nemoc (HD, z anglického Huntington's Disease) obsahovala 77 nahrávek 40 mluvčích (20 mužů a 20 žen), u kterých byla diagnostikována HN. Průměrný věk skupiny lidí s HN byl 48,6±13,4 let. U všech účastníků byla HN stanovena specialistou, který pro vyhodnocení použil Unified Huntington's Disease Rating Scale (UHDRS) [7]. Výsledky UHDRS pro tuto skupiny vycházely 26,9±11,6. Všech 77 nahrávek bylo původně pořízeno pro část předchozí studie [8].

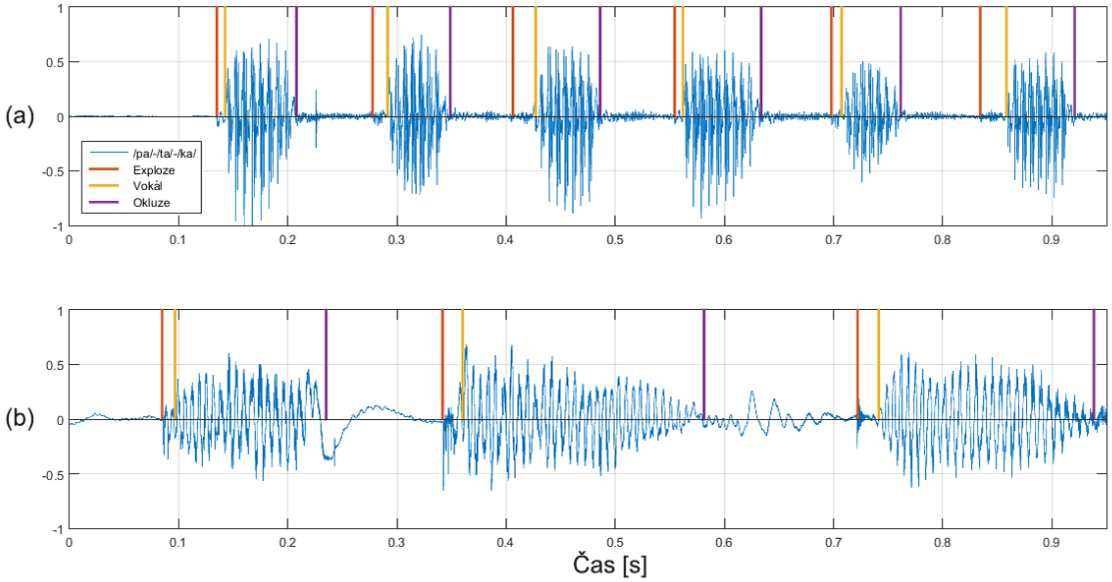
Kontrolní skupina zdravých lidí (HC, z anglického Healthy Control) obsahovala 80 nahrávek 22 mluvčích (15 mužů a 7 žen) s průměrným věkem 58,7±4,6 let.

### *2.1.2 Nahrávání*

Nahrávky byly pořízeny v tiché místnosti s nízkou úrovní okolního hluku pomocí kondenzátorového mikrofonu ve vzdálenosti asi 5 cm od pacientových úst. Data byla zaznamenána s vzorkovací frekvencí 48 kHz a kvantována 16-bitovým AD převodníkem. Všechny promluvy byly zaznamenány ve zkušební místnosti v rámci neurologického oddělení za přítomnosti patologa, který po mluvčích požadoval opakování slabik /*pa*/-/*ta*/-/*ka*/ konstantní rychlostí a tak rychle, jak to bylo možné. Nebyl stanoven žádný časový limit a účastníkům bylo umožněno úkol opakovat.

### *2.1.3 Manuální segmentace*

Aby bylo možné vyhodnotit úspěšnost algoritmu, u všech slabik musí být nejprve ručně určeny referenční pozice počátku exploze (IB – Initial Burst), nástupu samohlásky (VO – Vowel Onset) a okluze (O – Occlusion). Ruční segmentace dysartrických promluv může být v mnoha případech velmi náročný úkol, proto se celá manuální segmentace držela dvou pravidel, která jsou v souladu se stanovenými pokyny [9]. Za prvé, v případě nalezení více explozí je jako počáteční exploze souhlásky zvolena první z nich [10]. Za druhé, začátek vokálu byl definován na základě přítomnosti základního kmitočtu a prvních dvou formantových kmitočtů [11].




Obrázek 2.1 Manuální segmentace. (a) zdravý člověk (b) pacient s HN

## 2.2 Parametrizace signálu

Aby byl jakýkoliv automatický řečový detektor schopný klasifikovat řečový signál, musí k němu být nejprve stanoveny určité parametry, podle kterých je dále segmentován. Tento krok se nazývá „parametrizace signálu“. Moderní řečové detektory mohou k signálu přistupovat ve třech doménách:

- Časová oblast
  - Energetické charakteristiky
  - Počet průchodů nulou (kapitola 2.2.3)
  - Autokorelační charakteristiky signálu
- Kmitočtová (spektrální) oblast
  - Spektrální analýza na základě DFT (Discrete Fourier Transform)
  - LPC spektrální analýza (Linear Predictive Coding)
- Kepstrální oblast
  - Kepstrální analýza

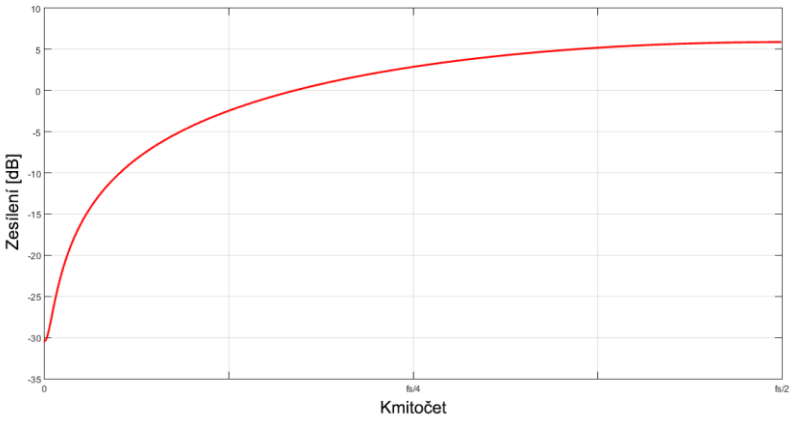
### 2.2.1 Předzpracování signálu

Základem pro výpočty parametrů je samotný diskrétní signál. V kapitole 2.1.2 je uvedeno, že surové nahrávky byly vzorkovány kmitočtem 48 kHz. Pro účely parametrizace je každý řečový signál převzorkován dostatečným kmitočtem 20 kHz. Stejný vzorkovací kmitočet byl použit i v předchozí práci [12].

Při výstupu akustické vlny z artikulačního ústrojí do volného prostředí dochází k útlumu intenzity zvuku pro vyšší kmitočty. Amplituda signálu klesá o 20 dB na dekádu kmitočtu. Jelikož je u určitých hlásek je důležitá informace ukryta právě ve vyšších kmitočtech, je potřeba tento útlum kompenzovat. Tato kompenzace se provádí filtrem 1. řádu zesilujícím vyšší kmitočty. Tento filtr se také nazývá *preemfázový filtr*. Je dán vztahem:

$$s'[n] = s[n] - m \cdot s[n-1] \qquad (2.1)$$

kde $m$ je koeficient preemfáze, který se volí v rozsahu 0,9 – 1 (typická hodnota je 0,97) [13].




Obrázek 2.2 Přenosová charakteristiky preemfázového filtru pro $m$ = 0,97

Pro výpočty krátkodobých charakteristik signálu je nejprve potřeba celý signál rozdělit na segmenty, ze kterých jsou parametry počítány. Pokud není uvedeno jinak, základní délka každého segmentu je 10 ms a jeho posun je roven 2 ms. Vybrané segmenty nejsou dále skládány, proto není potřeba používat váhovací okno.

V následujících kapitolách se seznámíme s vybranými parametry signálu, které byly dále použity ke klasifikaci nahraných promluv.

### 2.2.2 Krátkodobý výkon signálu

Krátkodobý výkon signálu je skládán z výkonů posouvaných segmentů. Počítaný výkon $P_k$ je dán vztahem:

$$P_k = \frac{1}{N} \cdot \sum_{n=1}^{N} |s[n] \cdot w[n]|^2 \tag{2.2}$$

kde $s[n]$ je vzorek segmentu v čase $n$, $w[n]$ je Hammingovo okno o stejné jako je segment $N$ je počet vzorků použitých pro výpočet výkonu (při vzorkovacím kmitočtu 20 kHz je $N$ rovno 200 vzorkům) [4].

### 2.2.3 Počet průchodů nulou (ZCR - Zero Crossing Rate)

Tento parametr můžeme chápat jako charakteristiku v časové oblasti, popisující spektrální vlastnosti signálu. Jelikož je řečový signál považován za širokopásmový, je parametr *počet průchodů nulou* vhodnější než spektrální analýza, která předpokládá, že analyzovaný signál bude úzkopásmový.

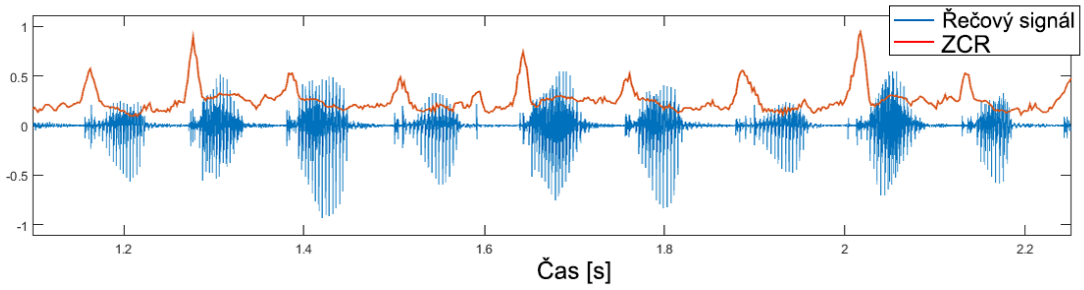
Krátkodobá funkce je stejně jako krátkodobý výkon skládána z výpočtů počtu průchodu nulou jednotlivých segmentů tímto vztahem [4]:

$$Z_k = \sum_{n=-\infty}^{\infty} |sgn(s[n]) - sgn(s[k-1])| \tag{2.3}$$

kde

$$sgn(s[n]) = \begin{cases} 1, & pro \quad s[n] \geq 0 \\ -1, & pro \quad s[n] < 0 \end{cases} \tag{2.4}$$

Pokud bude signál neznělý, v případě konsonant, hodnota ZCR bude popisovat kmitočty obsažené ve vyšší části spektra. Naopak u signálů znělých, v případě vokálů, očekáváme hodnoty ZCR nižší. Části signálu bez řečové aktivity vykazují šumový charakter, ZCR obvykle bývá vyšší s větším rozptylem, než u vokálů. Jelikož je tento parametr velmi citlivý na stejnosměrnou složku, je potřeba ji před samotným výpočtem odstranit.




Obrázek 2.3 Řečový signál s parametrem Počet průchodů nulou

### 2.2.4 Zbytkový signál lineární predikce (LP Residual)

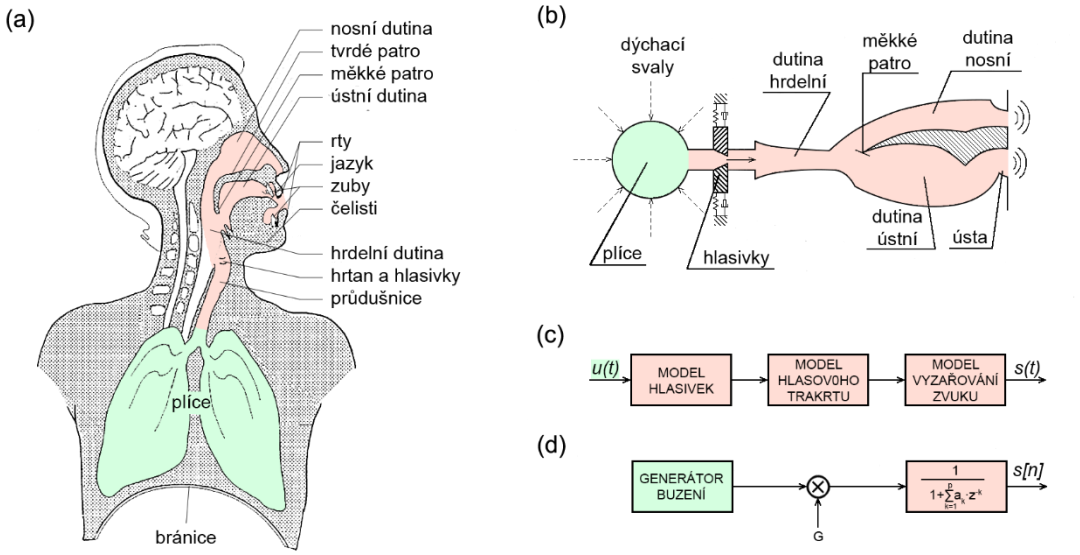
Řeč je produkovaná časově proměnným systémem hlasového traktu, kde se neznělý vzduch vytlačovaný z plic mění na znělý energeticky se měnící zvuk. Systém hlasového traktu lze popsat pomocí analýzy řečového signálu metodou lineární predikce (LPC – z anglického Linear Predictive Coding). LPC analýza předpokládá, že každý $n$-tý vzorek signálu $s[n]$ je možné popsat lineární kombinací $p$ předchozích vzorků a buzením $u[n]$.

$$s[n] = -\sum_{k=1}^{p} a_k s[n-k] + Gu[n] \tag{2.5}$$

kde $p$ je řád autoregresního (AR) modelu a $G$ je koeficient zesílení. Přenosová funkce modelu $H(z)$ *pak* vypadá takto:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{A(z)} = \frac{G}{1 + \sum_{k=1}^{p} a_k \cdot z^{-k}} \tag{2.6}$$

Z přenosové funkce jsou poté zajímavé koeficienty číslicového filtru $a_k$ a koeficient zesílení $G$ [4].




Obrázek 2.4 (a) fyziologický model (b) zjednodušený fyziologický model (c) blokové schéma modelu (d) AR model produkce řeči [4]

Ke stanovení nejvhodnější kombinaci koeficientů $a_k$ a zesílení $G$ lze využít metody nejmenších čtverců. Při analýze signálu ovšem členy $Gu[n]$ v rovnici (2.5) neznáme a vzniká tak chyba predikce $e[n]$ (také známou jako zbytkový signál lineární predikce, dále LP residuum) mezi skutečnou $s[n]$ a předpovězenou $\hat{s}[n]$ hodnotou. LP analýza tuto chybu predikce $e[n]$ minimalizuje [4].

$$e[n] = s[n] - \hat{s}[n] = s[n] + \sum_{k=1}^{p} a_k s[n-k] \tag{2.7}$$

Analyzovaný signál má poté podobu:

$$s[n] = -\sum_{k=1}^{p} a_k s[n-k] + e[n] \tag{2.8}$$

Kde $e[n]$ představuje buzení celého řečového traktu.

Charakteristiky časově proměnného buzení zahrnují změny z neznělých na znělé hlásky, úroveň energie hlasu a také periodicitu. Některé tyto charakteristiky mohou pomoci k detekci vokálů.

Řečový signál vzorkovaný frekvencí 8 kHz je zpracován po jednotlivých blocích o velikosti 20 ms s 10 ms posuvem [14]. Z každého bloku jsou pomocí LP analýzy desátého řádu vypočítány koeficienty $a_k$. Z těchto koeficientů je sestaven časově

proměnný inverzní filtr. Informace o buzení jsou získány filtrací řečového signálu sestaveným inverzním filtrem ve formě LP residua. Z residua je pomocí Hilbertovy transformace vypočítaná obálka $h_e[n]$ (dále HE – z anglického Hilbert envelope). Výpočet obálky je definován takto:

$$h_e[n] = \sqrt{e^2[n] + e_h^2[n]} \tag{2.9}$$

kde $e_h[n]$ je Hilbertova transformace $e[n]$ [14].

$$e_h[n] = IDFT\{E_h(k)\} \tag{2.10}$$

$IDFT\{E_h(k)\}$ je operátor zpětné diskrétní Fourierovy transformace.

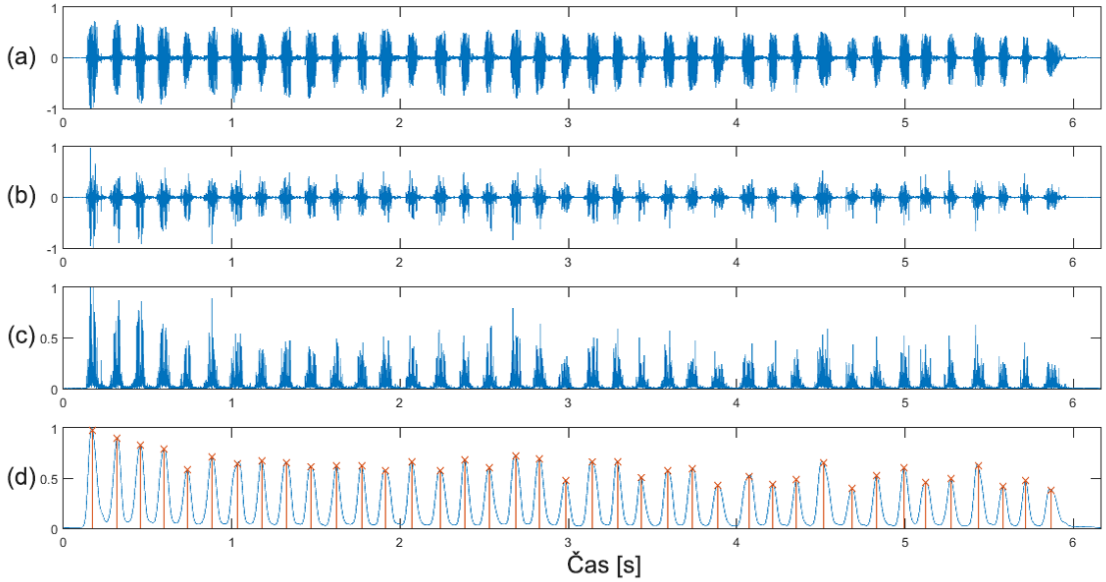
$$E_h(k) = \begin{cases} -j \cdot E[k], & k = 0,1,\dots,\left(\frac{N}{2}\right) - 1 \\ j \cdot E[k], & k = \left(\frac{N}{2}\right), \left(\frac{N}{2}\right) + 1, \dots, (N-1) \end{cases} \tag{2.11}$$

$E_h(k)$ představuje diskrétní spektrum LP residua, spočítané diskrétní Fourierovou transformací $DFT\{e[n]\}$ z $N$ vzorků [14].

$$E[k] = DFT\{e[n]\} \tag{2.12}$$

V HE jsou uchovány informace jak periodicitě, tak informace o energii. Pro tuto práci je zajímavý především parametr energie, proto je potřeba obálku vyhladit. Vyhlazení je provedeno konvolucí signálu s Hammingovým oknem $w[n]$ o délce 50 ms.

$$h_{e\_smoo}[n] = h_e[n] * w[n] = \sum_{k=-\infty}^{\infty} h_e[n] \cdot w[k-n] \tag{2.13}$$




Obrázek 2.5 (a) řečový signál (b) LP residuum (c) HE z LP residua (d) vyhlazená HE s kandidáty na slabiku.

### 2.2.5 Spektrální entropie (SE)

Mnoho moderních automatických detektorů řeči používá k popisu řečového signálu kepstrální koeficienty získané ze spektra počítaného krátkodobou Fourierovou transformací (STFT – z anglického Short-Time Fourier Transform). Nejčastěji používané jsou Melovské kepstrální koeficienty (MFCC), percepční predikční koeficienty (PLP) nebo RASTA koeficienty (z anglického Relative Spectra). V práci z IDIAP Research Institute [15] autoři spekulují o množství informace nesené v kepstrálních koeficientech vůči množství informace nesené v STFT spektru a zabývají se myšlenkou zachycení dalších informací z STFT spektra pomocí entropie.

U znělých hlásek jsou ve vyhlazeném spektru zřetelné formantové kmitočty a entropie takového spektra je nízká. U hlásek neznělých je vyhlazené STFT spektrum více ploché, tudíž entropie by měla být vyšší. Díky tomu lze spektrální entropii použít na detekci znělých a neznělých hlásek. Ve výše uvedeném článku [15] rozšířili tuto myšlenku o rozdělení vyhlazeného STFT spektra na kmitočtové sub-kanály v poměrech 1/1, 1/2, 1/3, 1/4, a 1/5 (viz Obrázek 2.6). To umožní zaměření se na část spektra obsahující pouze formantové kmitočty.

Před výpočtem samotné entropie je potřeba upravit kmitočtové spektrum do tvaru funkce hustoty pravděpodobnosti (PDF), v níž je suma všech prvků rovna jedné.

$$s_x[k] = \frac{S_x[k]}{\sum_{k=1}^{N} X[k]}, \qquad pro\ k = 1, \dots, N \tag{2.14}$$

Kde $S_x[k]$ je energie $k$-tého vzorku vyhlazeného odhadu spektrální výkonové hustoty, $N$ je počet vzorků, ze kterých je vyhlazené spektrum počítáno a $x[k]$ je výsledná PDF, která je dále použita k výpočtu entropie.

Vyhlazené spektrum $S_x[k]$ je modelováno frekvenční charakteristikou přenosové funkce $H(z)$ z rovnice (2.6) a je definováno jako

$$S_x(e^{j\theta}) = \frac{G^2}{|A(e^{j\theta})|^2} \tag{2.15}$$

Diskrétní vyhlazené spektrum $S_x[k]$ získáme dosazením $\theta = \frac{k \cdot f_s}{N}$ [13].

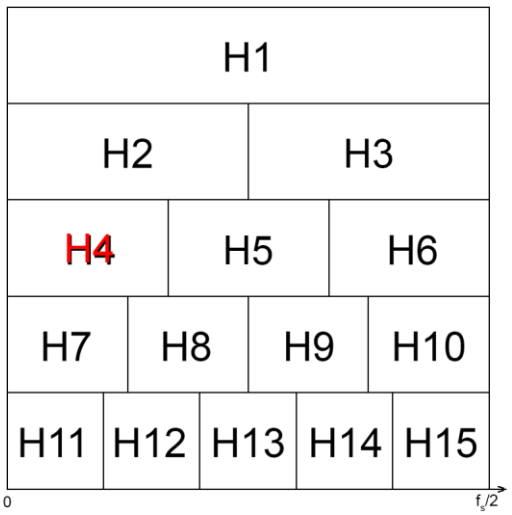
Pro každý segment signálu je poté entropie počítána takto:

$$H = -\sum_{k=1}^{N} s_x[k] \cdot \log_2 s_x[k] \tag{2.16}$$

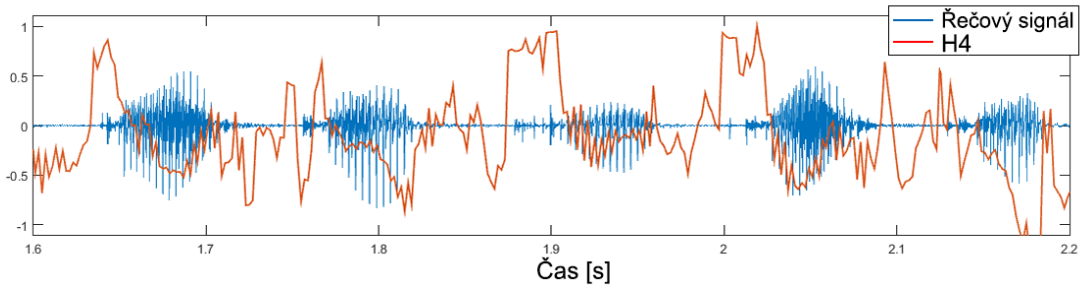
Toto je výpočet entropie pro celou šířku spektra. Pokud chceme počítat entropii pouze ve vybrané části spektra, v rovnici (2.14) zvolíme meze např.:

$$k = 1, \dots, \frac{1}{3}N$$

pro část spektra H4 (viz Obrázek 2.6), které je ilustrováno i v Obrázek 2.7.




Obrázek 2.6 Rozdělení spektra pro výpočet entropie




Obrázek 2.7 Řečový signál s vybranou spektrální entropií (H4).

### 2.2.6 Vlnková transformace (WT – Wavelet Transform)

Fourierova transformace poskytuje pouze informaci o jednotlivých kmitočtech obsažených v signále, nenese však informaci o jejich poloze v čase, je tedy vhodná pro popis stacionárních signálů. Řeč stacionární není, proto je zde potřeba uvažovat o jiném popisu signálu. Vlnková transformace je jedna z několika transformací, které nám dávají časově-kmitočtový popis signálu. Hlavní myšlenkou vlnkové transformace je vhodná změna šířky okna segmentujícího signál tak, aby bylo nastaveno správné rozlišení v časové i kmitočtové oblasti. Tomuto oknu se říká mateřská vlnka (funkce). Se zkracující se vlnkou získáváme sice vyšší rozlišení v časové oblasti, ovšem rozlišení v kmitočtové oblasti je nižší (viz Obrázek 2.9)

Mateřská vlnka ve spojitém čase má tvar:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \cdot \psi\left(\frac{t-\tau}{s}\right) \tag{2.17}$$

kde parametr $s$ (dilatace) je měřítko, kterým je možné měnit délku vlnky a $\tau$ (translace) je poloha, kterou se mění umístění vlnky v čase [16].

Vlnková transformace se pak počítá takto:

$$W(\tau, s) = \int_{-\infty}^{\infty} s(t) \cdot \frac{1}{\sqrt{s}} \cdot \psi^*\left(\frac{t-\tau}{s}\right) dt \tag{2.18}$$

Výsledkem je poté dvourozměrná funkce $W(\tau, s)$, která je často nazývána scalogram nebo vlnková mapa.

Diskrétní vlnková transformaci dostaneme vhodnou závislostí parametrů $s$ a $\tau$.

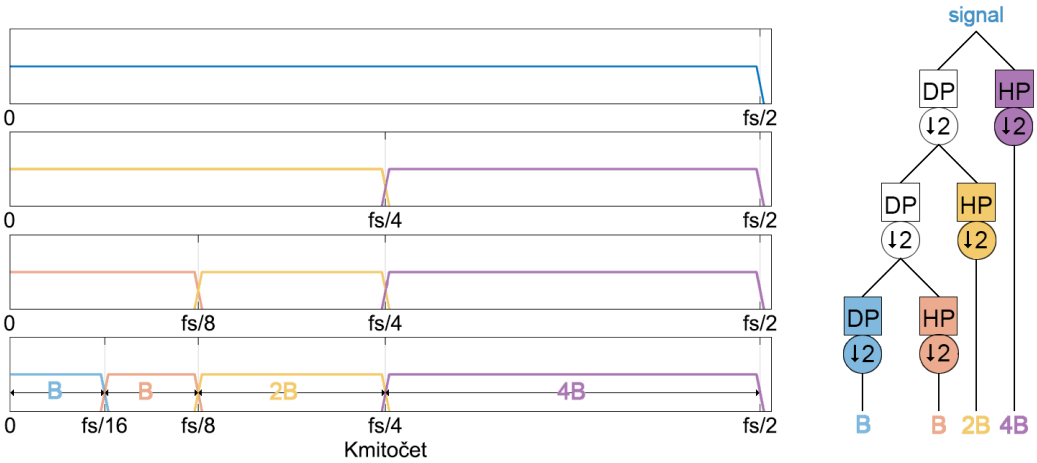
$$s = 2^p, \qquad \tau = 2^p \cdot k \tag{2.19}$$

Diskrétní mateřská vlnka má potom tvar:

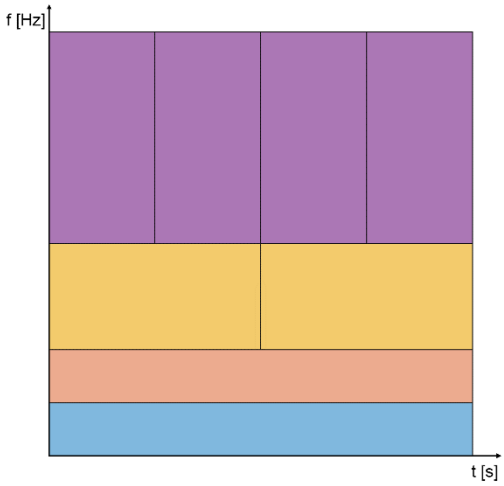
$$\psi_{k,p}[n] = \frac{1}{\sqrt{2^p}} \cdot \psi\left[\frac{n - 2^p \cdot k}{2^p}\right] \tag{2.20}$$

kde $p$ odpovídá měřítku a $k$ poloze [16].

Vlnková funkce $\psi$ se chová jako pásmová propust filtrující vstupní signál. V každém kroku je zachovaná horní část kmitočtového pásma (výstup z hornopropustného filtru – HP) a spodní část pásma (výstup dolnopropustného filtru – DP) je dále filtrována (viz Obrázek 2.8)




Obrázek 2.8 Frekvenční pohled na diskrétní vlnkovou transformaci

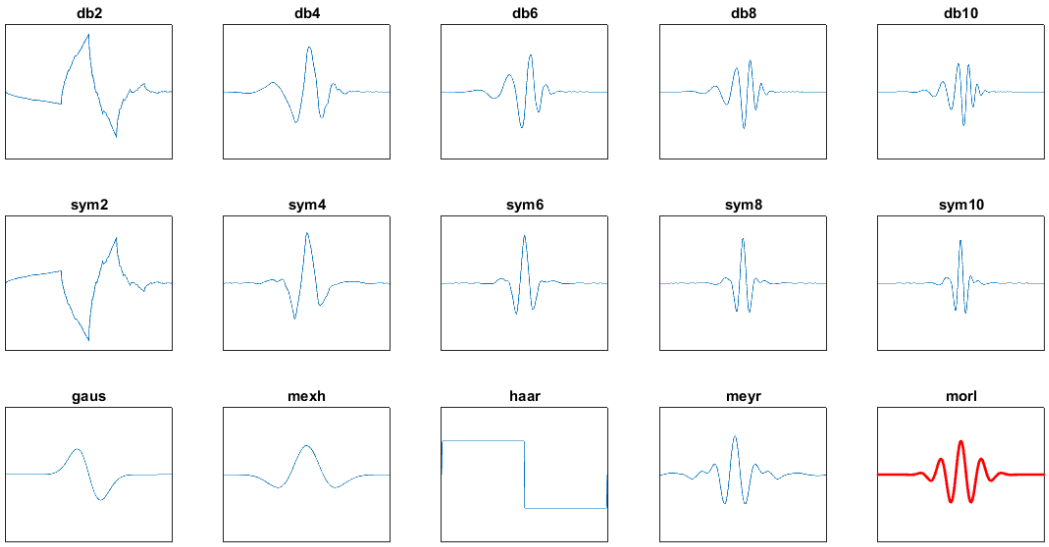



Obrázek 2.9 Rozložení vlnkové mapy (scalogram)

Vlnkovou transformaci si můžeme také představit jako konvoluci signálu s mateřskou vlnkou. Je potom zřejmé, že tato transformace ukazuje, ve kterých časových okamžicích je signál podobný určité vlnce. Mateřských funkcí je k dispozici celá řada. Od obyčejné vlnky tvaru Gaussovy funkce, přes vlnku Daubechies, Mexican Hat, Mayerovu nebo Haarovu vlnku a mnoho dalších [17]. V následujícím obrázku jsou znázorněny vybrané vlnky, mezi nimiž je i červeně vyznačená Morletova vlnka použitá pro účely parametrizace. Morletova vlnka je použita díky svému tvaru, který se nejblíže podobá tvaru signálu znělého „a“, tudíž bude dobře popisovat vokály v DDK promluvách.

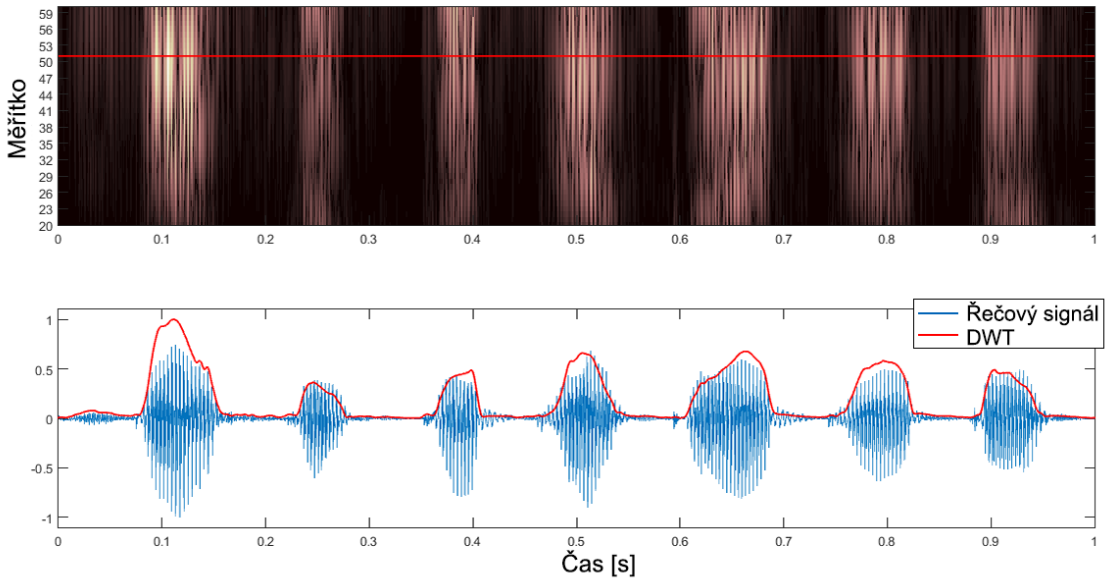
Matematický zápis Morletovy vlnky vypadá takto [16]:

$$\psi(x) = a \cdot e^{-\frac{1}{2}x^2}(\cos(5 \cdot x) + j \sin(5 \cdot x)) \tag{2.21}$$




Obrázek 2.10 Vybrané mateřské vlnky.
Nahoře: Daubechies (5 vybraných řádů).
Uprostřed: Symlet (5 vybraných řádů).
Dole: Gaussova vlnka, Mexican Hat, Haarova, Mayerova a Morletova vlnka.

Pro účely parametrizace je potřeba z vlnkové mapy vybrat pouze jeden časový průběh, který bude nejlépe popisovat výskyt vokálů v promluvě. S rostoucí energií ve vlnkové mapě roste i „podobnost“ signálu s mateřskou vlnkou. Proto je z mapy vybrán vždy ten časový průběh (měřítko vlnky), který má nejvyšší energii (viz Obrázek 2.11)




Obrázek 2.11 Scalogram a vybraný časový průběh s nejvyšší energií (podobností vlnce)

## 2.3 Dynamické prahování pomocí GMM

Model Gaussovských směsí (GMM – z anglického Gaussian Mixture Model) je jeden z nejpoužívanějších statistický popisů měřených dat. Jedná se o multimodální popis pravděpodobností, který se skládá z více Gaussových rozložení, často se překrývající jedna přes druhou [13].

Úkolem dynamického prahování pomocí GMM je nalezení hranice mezi jednotlivými křivkami Gaussovských hustot pravděpodobností a tím vytvořit klasifikační třídy (vokál, konsonant a část signálu bez řečové aktivity). Křivky jsou sestavovány z hustot pravděpodobností výše uvedených parametrů, které nesou informace o jednotlivých třídách.

K samotné klasifikaci tříd podle vstupních parametrů je v této práci využita knihovna MATLABu a to Statistics and Machine Learning Toolbox, která nabízí funkci *fitgmdist*. Ta vytvoří ze vstupních pozorování model Gaussovských směsí, ve kterém jsou mimo jiné uloženy střední hodnoty, rozptyly a počet nalezených komponent. Nejprve je nutné definovat prostor parametrů, který bude pro funkci *fitgmdist* představovat sadu pozorování.

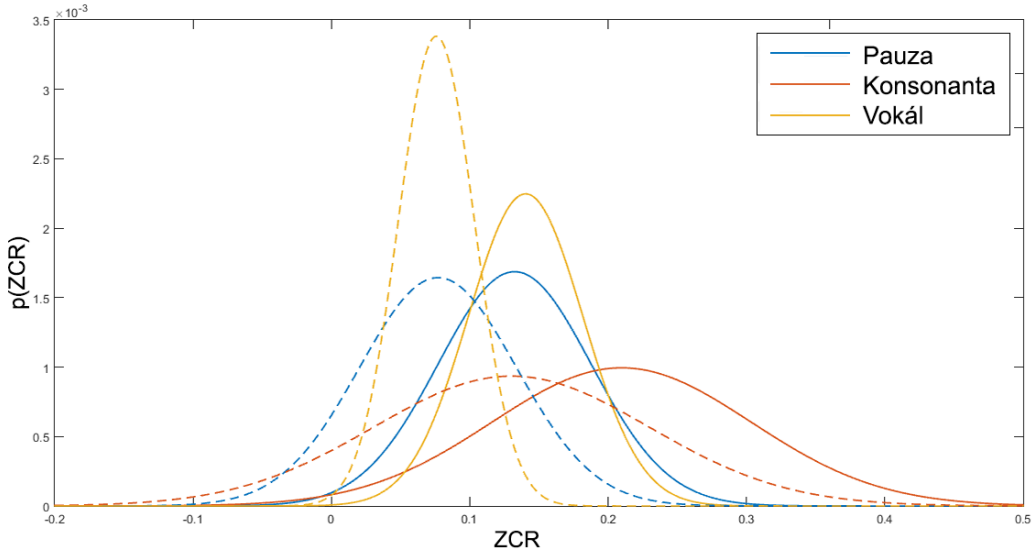
### *2.3.1 Prostor parametrů*

V následující kapitole se podíváme na Gaussovské rozložení jednotlivých tříd pro vybrané parametry.

### 2.3.1.1 Počet průchodů nulou

Když se podíváme na Obrázek 2.12, je zřejmé, že konsonanty mají vyšší ZCR než vokály. V případě části signálu bez řečové aktivity jsou očekávány hodnoty ZCR přibližně stejné jako v případě konsonant. Na Obrázek 2.12 je ale vidět, že hodnoty klesají téměř na úroveň vokálů, které jsou charakteristické znělostí a tím i nižší hodnotou ZCR.

Pokles hodnot ZCR u části signálu bez řečové aktivity může být způsoben aditivní stejnosměrnou složkou, nebo síťovým rušením. Tyto nedostatky je třeba ze signálu eliminovat již při preprocesingu.



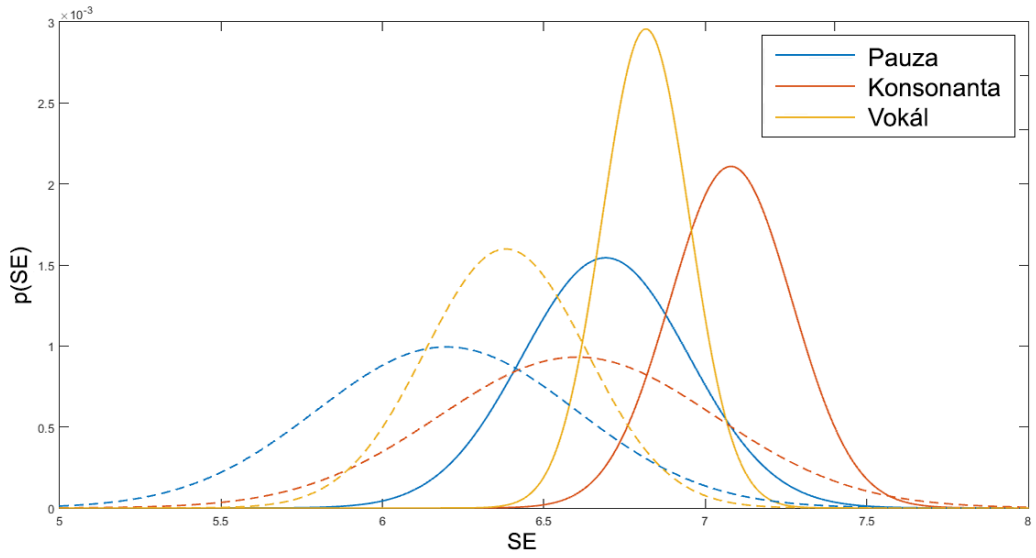

Obrázek 2.12 Hustota pravděpodobnosti ZCR pro signál bez řečové aktivity (modře), konsonanty (červeně) a vokály (žlutě). Zdraví lidé (plnou čarou), pacienti s HN (čárkovanou čarou)

### 2.3.1.2 Spektrální entropie

Nedostatky ZCR by však měla eliminovat spektrální entropie.

Jak je uvedeno v kapitole 2.2.5, spektrální entropie je počítána v jednom ze 14 sub-pásem (myšleno ze spektra od stejnosměrné složky do poloviny vzorkovacího kmitočtu $f_s = 20kHz$). Na tomto místě je vhodné zaměřit se na část spektra obsahující formantové kmitočty. Tedy přibližně v pásu od 300 Hz až do 3500 Hz [13]. Když se podíváme na výsledné rozložení Gaussovské směsi pro spektrální entropii počítanou v pásu od stejnosměrné složky do jedné třetiny spektra (od 0 Hz do 3333 Hz), je zřejmé, že v této části spektra se pro pasáže bez řečové aktivity objevují nežádoucí složky, které snižují entropii až pod úroveň

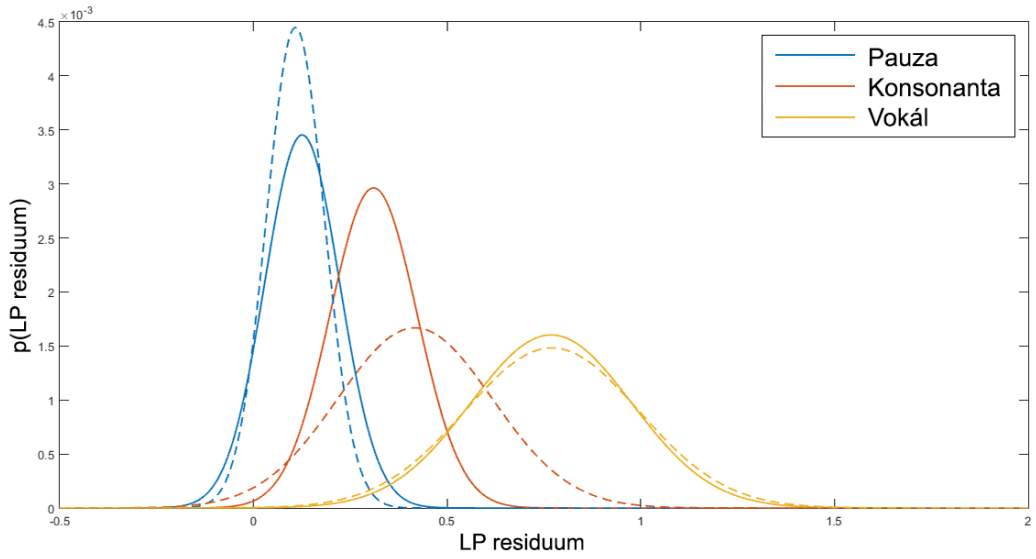
vokálů (Obrázek 2.13). Nicméně zvolit toto pásmo je výhodné především pro oddělení konsonant od vokálů. Porovnávaná pásma (viz Obrázek 2.6).




Obrázek 2.13 Hustota pravděpodobnosti SE (v pásmu H4) pro signál bez řečové aktivity (modře), konsonanty (červeně) a vokály (žlutě). Zdraví lidé (plnou čarou), pacienti s HN (čárkovanou čarou)

### *2.3.1.3 LP residuum*

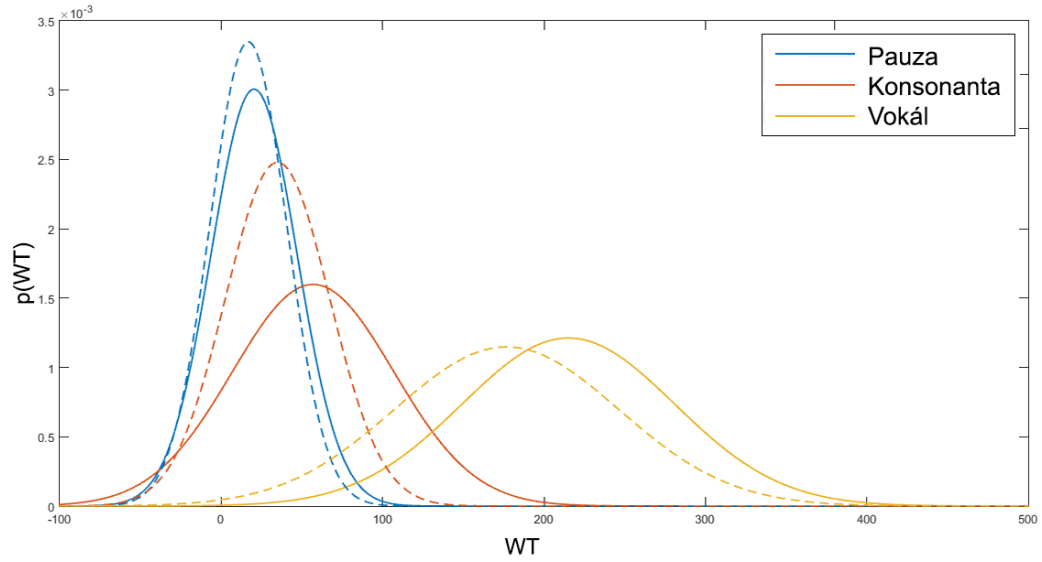
LP residuum dává informaci o buzení hlasového traktu. Je tedy zřejmé, že signálu části bez řečové aktivity budou nabývat hodnot blízkých nule. Rozptyl těchto hodnot je způsoben např. respiracemi během řeči. Nejvyšších hodnot nabývají znělé vokály, u kterých je vzduch tlačený z plic zesilován rezonátorem (hlasivkami). Širší rozptyl u vokálů je dán především tím, že ke konci nahrávky mluvčímu „dochází dech“ a energie celkového projevu má klesající trend. U neznělých konsonant předpokládáme hodnoty LP residua mezi hodnotami ticha a vokálů. U jedinců s HN jsou hodnoty PL residua u konsonant vyšší. Je to způsobeno dysartrií spojenou se špatnou koordinací dýchání.




Obrázek 2.14 Hustota pravděpodobnosti LP residua pro signál bez řečové aktivity (modře), konsonanty (červeně) a vokály (žlutě). Zdraví lidé (plnou čarou), pacienti s HN (čárkovanou čarou)

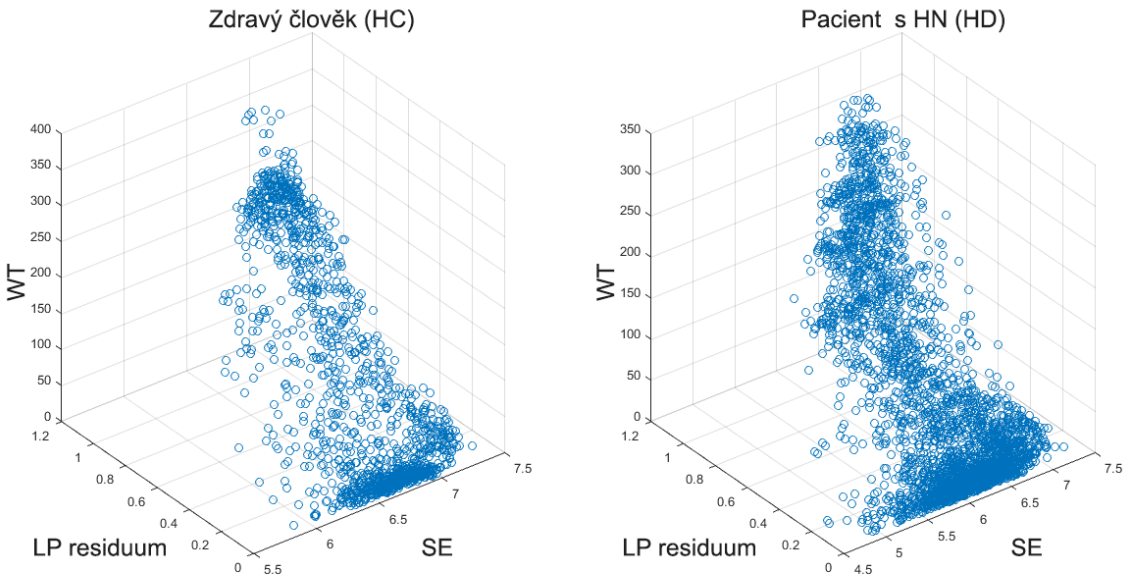
### 2.3.1.4 Vlnková transformace

Vlnková transformace má za úkol parametricky oddělit vokály od zbytku signálu. Části signálu bez řečové aktivity a konsonanty vykazují šumový charakter, který s Morletovou vlnkou vykazují velmi nízkou míru korelace. Vokály jsou na opak s vlnkou téměř totožné. Rozptyl tohoto parametru u vokálů je způsoben především nestálostí hlasivek a tím způsobeným kolísáním základního kmitočtu rezonátoru a tím i kolísání korelace s Morletovou vlnkou. (Obrázek 2.15)




Obrázek 2.15 Hustota pravděpodobnosti vlnkové transformace pro signál bez řečové aktivity (modře), konsonanty (červeně) a vokály (žlutě). Zdraví lidé (plnou čarou), pacienti s HN (čárkovanou čarou)

Pokud by byla klasifikace založena pouze na jednom z parametrů, dosahovala by značných chyb. Proto je k rozhodování použito tří parametrů, které společně zvýší odstup jednotlivých tříd v prostoru. (viz Obrázek 2.16) Tyto tři parametry představují sadu pozorování, které budou sloužit k sestavení modelu Gaussovských směsí.




Obrázek 2.16 Prostor parametrů. WT - Vlnková transformace, LP residuum, SE - spektrální entropie. Zdravý člověk (vlevo), pacient s HN (vpravo).

Za pomoci takto definovaného prostoru parametru je nyní možné sestavit GMM model. Odhadování parametrů modelu je v případě *fitgmdist* optimalizováno pomocí iterativního EM-algoritmu (Expectation-Maximization algorithm). Ten slouží k získání maximálně věrohodného odhadu v případě, kdy máme omezené množství pozorování. EM-algoritmus odhaduje parametry z jednotlivých směsí a z nich poté určuje posteriorní pravděpodobnosti příslušnosti do jednotlivých tříd. Způsoby rozhodování EM-algoritmu možné ilustrovat na bimodální Gaussovské směsi, která obsahuje znělý vokál a neznělé konsonanty společně s tichem. Nechtě $P(X)$ je obecná bimodální Gaussovská směs složená ze dvou distribucí $P(X|vokál)$ a $P(X|neznělé)$. Rozhodovací úroveň o tom, do které třídy vzorek patří, vyjadřuje diskriminant $D(X)$. Ten lze definovat jako rozdíl věrohodností obou tříd $L(X)$ posunutý o $T$.

$$D(X) = L(X) + T \tag{2.22}$$

$$L(X) = ln\left(\frac{P(X|vokál)}{P(X|neznělé)}\right) \tag{2.23}$$

$$T = ln\left(\frac{P(vokál)}{P(neznělé)}\right) \tag{2.24}$$

$$T = \frac{P(vokál)}{P(neznělé)} \tag{2.25}$$

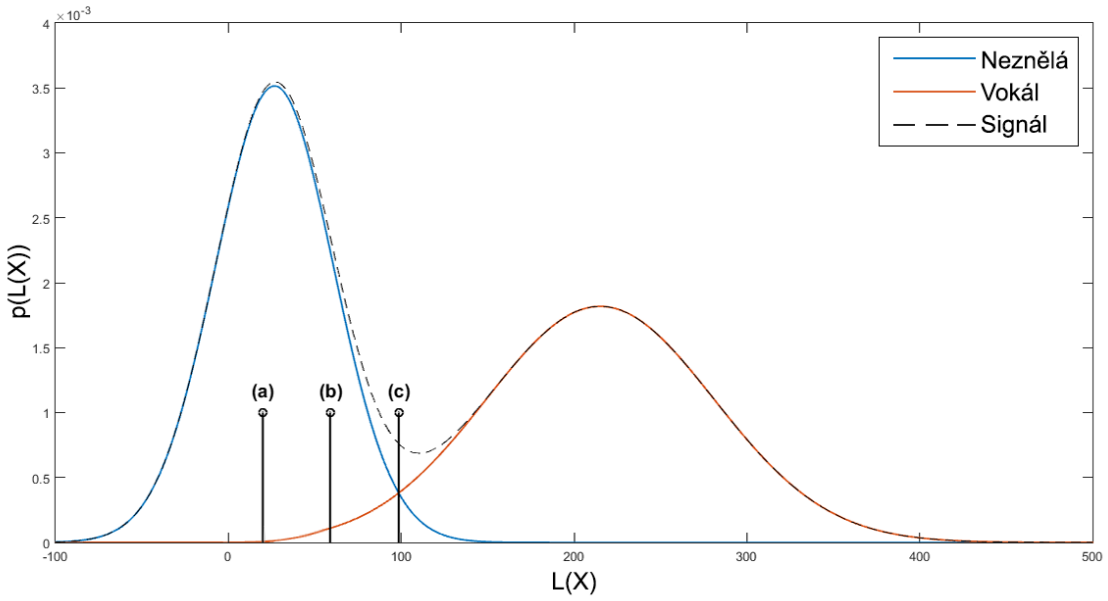
Rozhodovat se zda vzorek $x$ patří do jedné ze dvou tříd lze dvěma způsoby. O příslušnosti do třídy vyjádří diskriminant $D(x)$:

$$x = \begin{cases} vokál, & D(x) \geq 0 \\ neznělé, & D(x) < 0 \end{cases} \tag{2.26}$$

V tomto případě se obvykle používá definice prahu podle (2.24) [18].

Další možností je rozhodování na základě minimální hodnoty prahu $k$:

$$x = vokál, \qquad P(x|vokál) > k \tag{2.27}$$




Obrázek 2.17 Hustota pravděpodobnosti bimodální směsi. Signál (černě), vokál (červeně) a neznělá část signálu (modře).
Prahy: (a) práh ze vztahu (2.27), (b) práh ze vztahu (2.25), (c) práh ze vztahu (2.24).

Před zahájením samotného iterativního EM-algoritmu proběhne ve funkci *fitgmdist* počáteční inicializace hodnot požadovaných pro každou komponentu Gaussovských směsí (střední hodnoty, kovariance a poměr směsí. Tato inicializace využívá k-means++ algoritmus, který má následující postup:

1. Nastaví rovnoměrné rozdělení pravděpodobností jednotlivých komponent.

$$p_i = \frac{1}{k}, \qquad pro\ i = 1, \dots, k \tag{2.28}$$

2. Kovarianční matici zvolí tak aby byla diagonální a identická, kde

$$\sigma_i = diag(a_1, a_2, \dots, a_k) \text{ a } a_j = var(X_j) \tag{2.29}$$

3. Ze všech bodů pozorování $\mathrm{X}$ rovnoměrně zvolí první střed $\mu_1$ komponent.

4. Pro zvolení středu $j$:
   a. Vypočte Mahalanobisovu vzdálenost z každého pozorování pro každý střed a přiřadí každé pozorování k nejbližšímu středu.
   b. Pro $\mathrm{m} = 1, \dots, \mathrm{n}$ a $\mathrm{p} = 1, \dots, \mathrm{j} - 1$ náhodně vybere střed $j$ z pozorování $\mathrm{X}$ s pravděpodobností:

$$p = \frac{d^2(x_m, \mu_p)}{\sum_h d^2(x_h, \mu_p)}, \qquad x_h \in M_p \tag{2.30}$$

   kde $d(x_m, \mu_p)$ je vzdálenost mezi pozorováním $m$ a středem $\mu_p$. $M_p$ je potom množina všech pozorování blízkých ke středu $\mu_p$ a $x_m$ do této množiny spadají.
   Takto se zvolí další střed s pravděpodobností úměrnou vzdálenosti k nejbližšímu středu, který je již zvolený.

5. Krok 4. se opakuje, dokud není nalezeno $k$ středů [17].

Mahalanobisova vzdálenost počítá vzdálenost mezi bodem pozorování a pravděpodobnostním rozdělením. Mějme pozorování $x = (x_1, x_2, \dots, x_N)^T$ se středními hodnotami $\mu = (\mu_1, \mu_2, \dots, \mu_N)^T$ a Kovarianční maticí $S$. Mahalanobisova vzdálenost je definovaná takto:

$$d_M(x) = \sqrt{(x - \mu)^T \cdot S^{-1} \cdot (x - \mu)} \tag{2.31}$$

Rozdělování do klastrů pomocí GMM je často používaná metoda. Klastry jsou vytvářeny vybíráním komponent, což maximalizuje posteriorní pravděpodobnost. GMM, podobně jako metoda k-means, využívá iterativního algoritmu, který se přibližuje k lokálnímu minimu. Modelování Gaussovských směsí bývá přesnější především ve chvíli, kdy jsou hledané shluky rozdílně veliké [17].

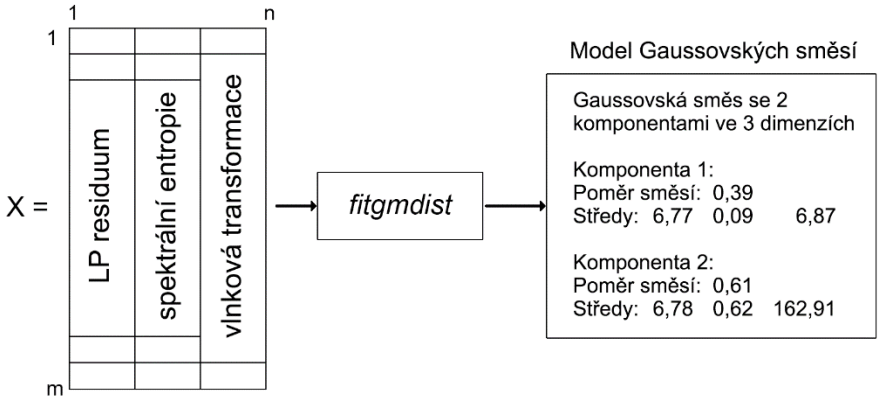
### 2.3.2 Detekce nástupu vokálů (VO – Vowel Onset)

Když máme definované dynamické prahování pomocí GMM, můžeme se pustit do detekce vokálů.

Na začátku navrhovaného algoritmu proběhne načtení DDK promluvy spojené s jeho úpravou v podobě filtrace preemfázovým filtrem, odečtení stejnosměrné složky a normalizací celého signálu.

Poté jsou vypočítána potřebná pozorování pro stanovení prahů pomocí dynamického prahování GMM. Pro detekci vokálů byly zvoleny charakteristiky LP residuum, Vlnková transformace a Spektrální entropie (H4). Jelikož jsou všechny tyto charakteristiky energeticky závislé, není vhodné stanovovat GM model pro celý signál najednou. A to kvůli klesajícímu trendu energie v průběhu času u mnoha DDK nahrávek. Z těchto důvodů je signál rozdělen na segmenty pomocí klouzavého okna o délce jedné sekundy, které je posouváno o 0,6 sekundy. Tento posun byl stanoven empiricky.

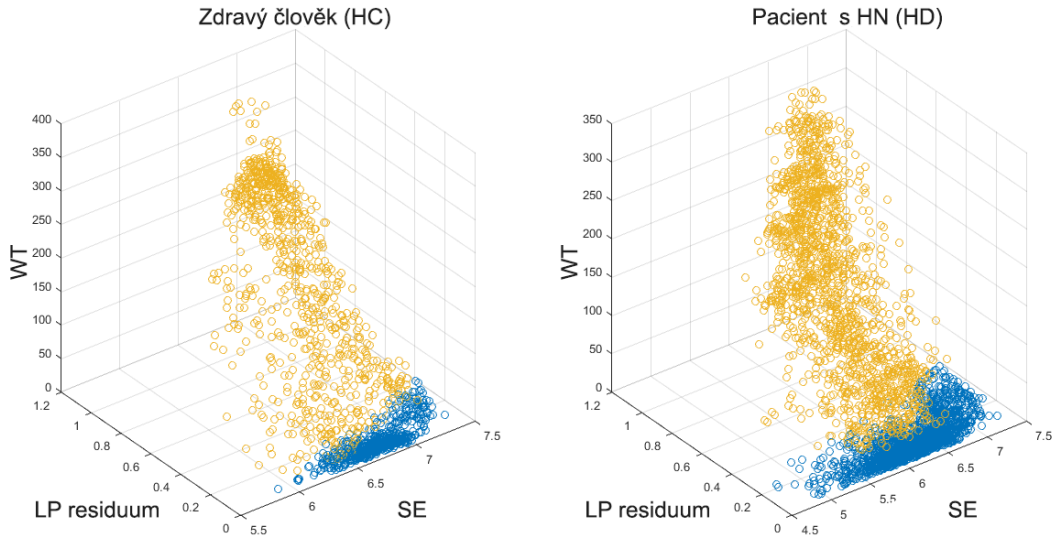
Vypočítaná pozorování, pro každý jeden segment, jsou poskládaná do matice $X$, která je dále zpracovaná funkcí *fitgmdist* (viz kapitola 2.3). Výsledkem je pak model Gaussovských směsí s danými středními hodnotami, rozptyly a poměrem směsí (viz Obrázek 2.18). Pomocí těchto parametrů dokážeme klastrovat nejen matici $X$, ale i originální signál, ze kterého byla pozorování počítána. Jednotlivé segmenty s indexy klastrů jsou poté skládány zpět do plné délky původního řečového signálu.




Obrázek 2.18 Matice *X* s *m* pozorování *n* parametrů (vlevo). Model Gaussovských směsí s dvěma komponentami (vpravo).

Protože jsou modely Gaussovských směsí každého segmentu jedinečné, často dochází k rozdílnému označení obou směsí. V jednom segmentu je vokálu přiřazen index 0 a neznělé části index 1, v následujícím segmentu to může být právě naopak. Z toho důvodu je nutné před složením segmentů do jednoho celku

vždy definovat, který index patří vokálu a který patří neznělé pauze. K tomu slouží porovnání výkonů jednotlivých částí signálu s indexem 0 a částí s indexem 1. Část s vyšším výkonem je považovaná za signál obsahující vokál.




Obrázek 2.19 Ukázka dvou Gausovských směsí, získaných z GMM. Index 0 pro vokály (žlutě), index 1 pro neznělé části signálu (modře)

Nyní jsou správně stanoveny indexy, u kterých ale na hranicích dochází k zákmitům (viz Obrázek 2.20 (a)). Tento nepříjemný jev je eliminován binární dilatací a erozí. Dilatace binárního signálu $A(x)$ jádrem $B(x)$ je definována:

$$(A \oplus B)(x) = max\{A(x - x') + B(x')\}, \qquad (x') \in D_B \qquad (2.32)$$

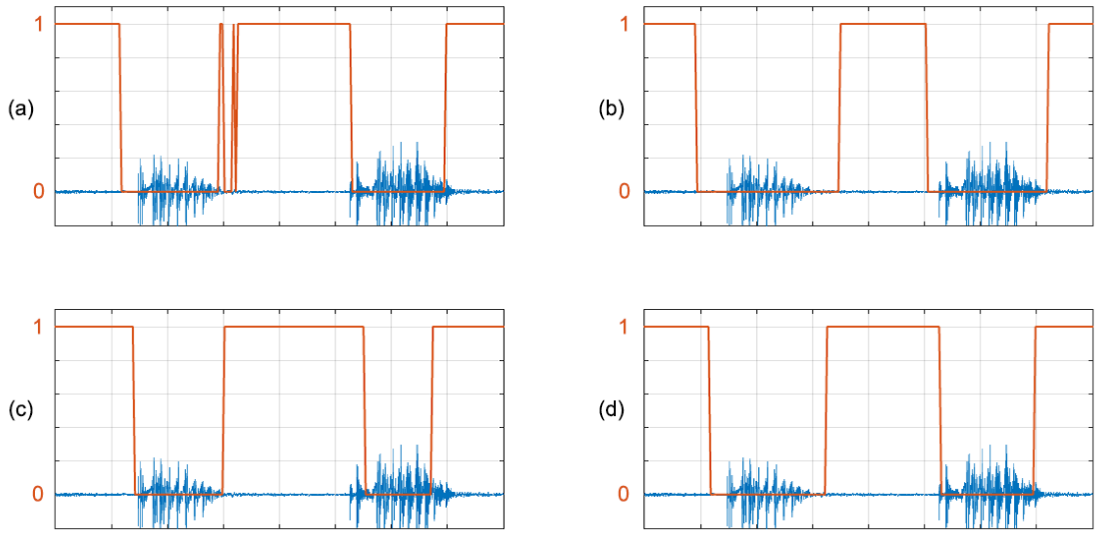
kde $D_B$ je oblast prvku $B$ a $A(x)$ je mimo oblast signálu předpokládáno rovno $-\infty$ [17]. Dilatovaný signál poté vyrovná zákmity z indexu 1 (Obrázek 2.20 (c)).

Binární eroze signálu $A(x)$ jádrem $B(x)$ je potom definována takto:

$$(A \ominus B)(x) = min\{A(x - x') - B(x')\}, \qquad (x') \in D_B \qquad (2.33)$$

kde $D_B$ je oblast prvku $B$ a $A(x)$ je mimo oblast signálu předpokládáno rovno $+\infty$ [17]. Eroze naopak od dilatace vyrovnává zákmity z indexu 0 (Obrázek 2.20 (b)).

Při vyhlazování zákmitů je nejprve aplikovaná binární eroze, pro zachování co největší části detekovaných vokálů, poté je signál s indexy dilatován s dvojnásobně dlouhým jádrem oproti erozi a nakonec je vše znovu prohnáno binární erozí. Tyto tři kroky jsou znázorněny na Obrázek 2.20.




Obrázek 2.20 Znázornění binární dilatace a binární eroze. (a) originální indexy, (b) erodované indexy, (c) dilatované indexy, (d) podruhé erodované indexy (na původní velikost)

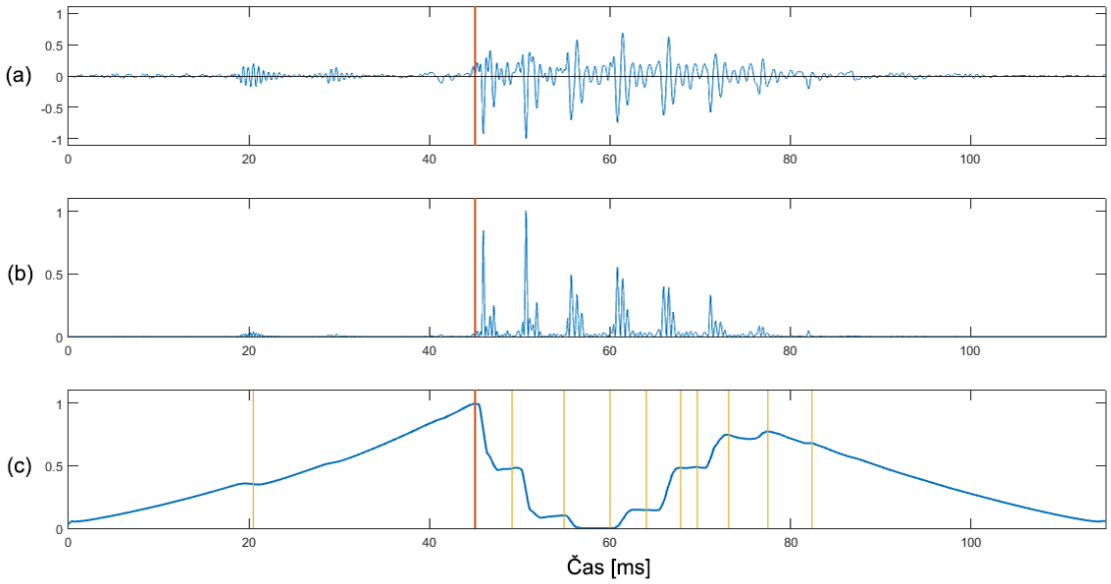
Samotné detekce počátku vokálů jsou poté nalezeny pomocí detektoru hran.

$$\delta[n] = diff(idx[n]) \tag{2.34}$$

kde funkce $diff(x)$ představuje diferenci mezi dvěma sousedními body signálu $idx[n]$. Ten reprezentuje indexy vypočítané pomocí GMM. Pozice začátku vokálu je potom stanovena jako vzorek $n$, který splňuje následující podmínku:

$$\delta[n] = -1 \tag{2.35}$$

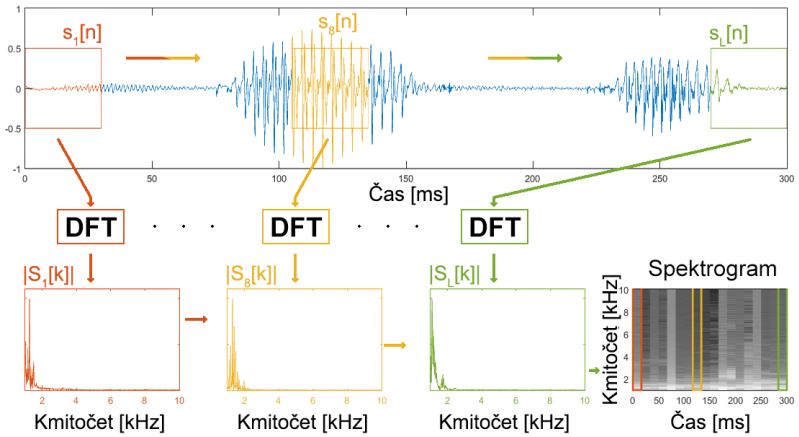
Na předchozím obrázku je patrné, že tato samotná metoda nesplňuje požadovanou přesnost na detekci vokálů. Z těchto důvodů je pro zpřesnění použit Bayesův detektor změn (BSCD – Bayesian Step Changepoint Detector) použitý i v předešlé práci [12]. BSCD předpokládá, že analyzovaný signál se skládá ze dvou konstantních hodnot a počítá posteriorní pravděpodobnost změn v signále s využitím Bayesovské marginalizace. BSCD svým charakterem vyznačuje hranici mezi dvěma odlišnými signály. Rozhodování o počátku vokálu je založeno na rozdílu vzdáleností detekovaných špiček v BSCD (Obrázek 2.21).




Obrázek 2.21 Detekce VO pomocí Bayesovského detektoru změn.
(a) původní signál, (b) umocněný signál (signál$^2$), (c) BSCD s detekovanými špičkami.

### *2.3.3 Detekce počátku exploze (IB – Initial Burst)*

Ve chvíli, kdy jsou stanoveny hranice VO, je možné zaměřit se na detekci exploziv, které se v případě DDK promluv nachází před detekovaným vokálem. Detekce exploziv je založená na frekvenční analýze, konkrétně analýze v podobě spektrogramu. Je to dvojrozměrná reprezentace signálu, kde v prvním rozměru je čas a druhou dimenzí je kmitočet. Pro výpočty spektrogramu se využívá tzv. krátkodobá Fourierova transformace (STFT – Short Time Fourier Transform). Spektrogram je na zpracováván po segmentech o délce 0,5 ms s polovičním překryvem.




Obrázek 2.22 Ilustrace výpočtu spektrogramu.

Poté je vypočítaná filtrační matice $T$, která slouží k odfiltrování zanedbatelných hodnot ve spektrogramu. Filtrační matice je definována následujícím vztahem:

$$T(i, 1 \dots N) = 0{,}75 \cdot \frac{1}{N} \sum_{n=1}^{N} P(i, n) \tag{2.36}$$

kde $i$ představuje index každého kmitočtového binu, $N$ je celkový počet časových binů a $P$ je matice spektrální výkonové hustoty [12].

Spektrogram je maticí $T$ filtrován následujícím způsobem:

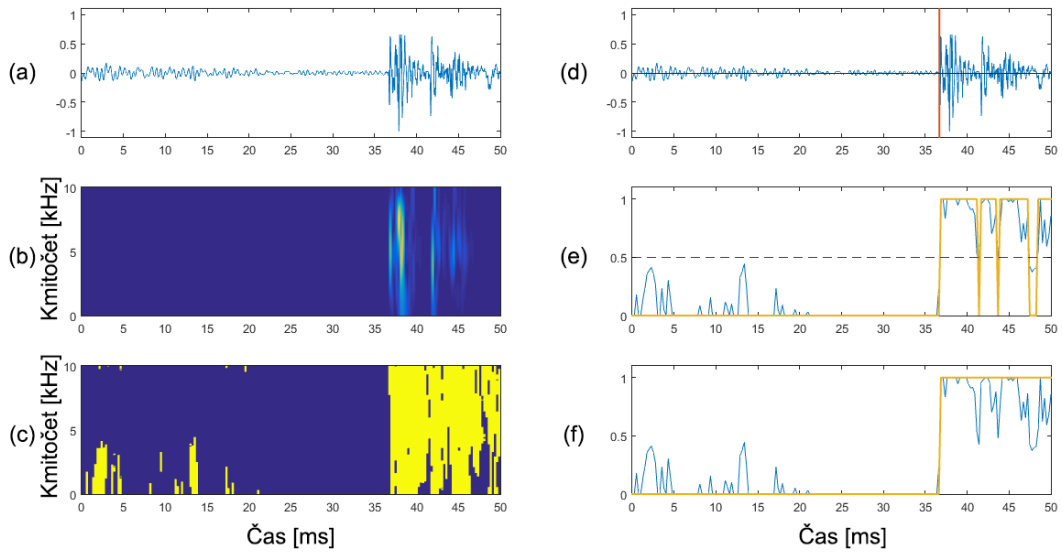
$$P_{filtrovaný}(i, n) = \begin{cases} 1, & P(i, n) \geq T(i, n) \\ 0, & P(i, n) < T(i, n) \end{cases} \tag{2.37}$$

Z takto vyfiltrovaného spektrogramu je dále vypočítána obálka. Ta je definována součtem všech hodnot ve sloupci matice:

$$P_{obálka}(n) = \sum_{i=1}^{I} P_{filtrovaný}(i, n) \tag{2.38}$$

V této obálce jsou dále hledané hodnoty vyšší, než 50% viz Obrázek 2.23 (e).

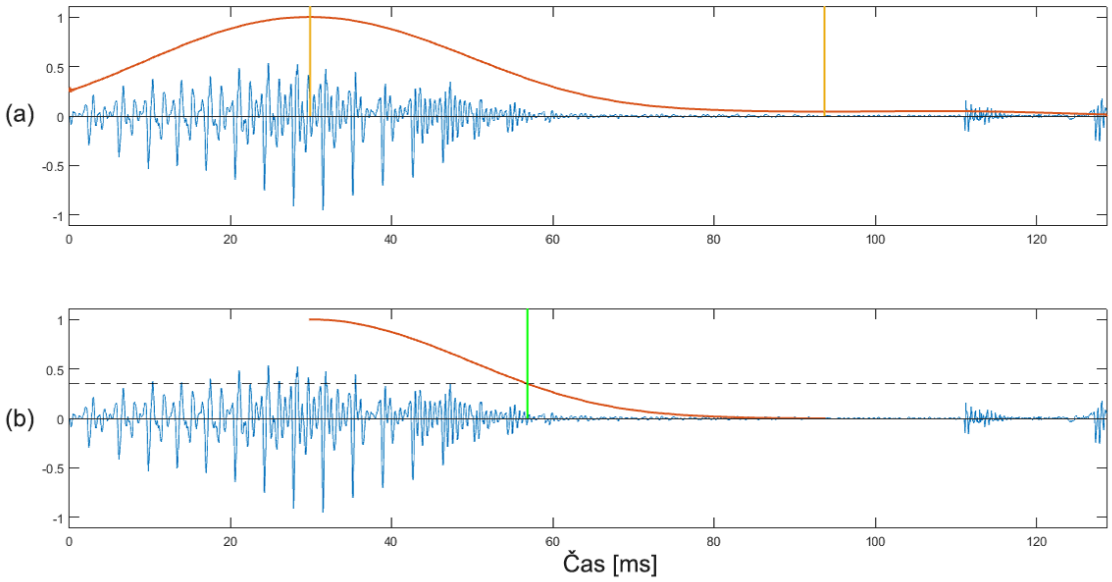
Z takto upravené obálky je možné pomocí detektoru hran nalézt pozici IB. Jak je ale vidět na Obrázek 2.23 (e), hodnoty obálky v místě exploze kolísají. Pro odstranění tohoto nepříjemného jevu využijeme binární eroze a binární dilatace, zmíněné v předchozí kapitole 2.3.2. Nyní je obálka vyhlazená a pomocí detektoru hran nalezneme pozici IB.




Obrázek 2.23 Detekce exploze. (a), (d) analyzovaný signál s detekovaným IB (červeně). (b) Spektrogram, (c) filtrovaný spektrogram, (e) obálka (modře) s vyznačenými hodnotami (>50%) (žlutě), (f) vyhlazená obálka (>50%) (žlutě)

### 2.3.4 Detekce okluzí (O – Occlusion)

Nyní jsou již stanoveny VO a je možné se pustit do hledání okluzí. Pro tyto účely je znovu využita parametrizace a to konkrétně LP residuum. V prvním kroku je z řečového signálu vyříznut segment mezi dvěma po sobě jdoucími VO a z tohoto segmentu je vypočítáno LP residuum. Jak je uvedeno v kapitole 2.2.4, residuum vypovídá o buzení řečového traktu a tím nejlépe vypovídá o ukončení řečové aktivity. Po výpočtu LP residua je v něm nalezeno první maximum vpravo od VO a následující první minimum. Dále je tato část residua mezi nalezeným maximem a minimem normována od nuly do jedné, což umožňuje přistupovat k němu procentuálně. Hranice okluze je poté stanovena podle pevně daného prahu 35% poklesu energie znormovaného LP residua. Hodnota 35% byla stanovena experimentálně.




Obrázek 2.24 Detekce okluzí pomocí LP residua. (a) signál (modře), LP residuum (červeně), nalezené první maximum a minimum (žlutě). (b) Znormovaná část residua (červeně), hranice 35% (čárkovaně), deteekovaná okluze (zeleně).

### 2.3.5 Akustické příznaky

Pro zhodnocení dopadu Huntingtonovy nemoci na řečový projev pacienta je zvoleno 11 charakteristik, představující 5 aspektů řeči. Charakteristiky popisují *Kvalitu řeči, Koordinaci činnosti hlasivek, Přesnost artikulace konsonant, Oslabení okluzí* a *Časování promluvy.* Všechny tyto charakteristiky jsou uvedeny v Tabulka 1. Vzhledem k rozdílným spektrálním charakteristikám konsonant /p/, /t/ a /k/ a jejich následujících vokálů, je jejich popis přesnosti artikulace počítána

na různých typech slabik (obouretné /pa/, předodásňové /ta/ a měkkopatrové /ka/) odděleně. Přesnost artikulace konsonant probíhala také separátně. Z těchto důvodů se celkový počet měření rozšířil na 21 [9].

#### 2.3.5.1 Kvalita řeči

Jak je uvedeno v kapitole 1.2, u pacientů s HN jsou, mimo jiné, postiženy také hrtanové svaly. Narušení těchto svalů může mít za následek snížení schopnosti udržet laryngální svaly ve stabilní poloze. To vede ke chvění hlasu, kolísání základního kmitočtu hlasivek nebo obecně zašumění řeči. V této je použit parametr Kvocient Autokorelační funkce prvních 30 ms od nástupu vokálu ($VSQ_{30}$ – AutoCorrelation Quotient), který je definován jako první autokorelační koeficient a odhad schopnosti produkovat stabilně hlasitý tón. Motivace prozkoumávat prvních 30 ms je založena na studii o artikulaci samohlásek u pacientů s Parkinsonovou nemocí [19].

#### 2.3.5.2 Koordinace činnosti hlasivek

Poruchy svalové činnosti u pacientů s HN ovlivňují koordinaci skupiny hlasivkových svalů. K vyhodnocení míry dopadu HN na koordinaci skupiny hrtanových svalů byl použit parametr doba nástupu vokálu (VOT) spolu s jeho směrodatnou odchylkou. Parametr VOT, definovaný jako doba mezi počáteční explozí a nástupem vokálu, byl motivován z předpokladu, že tato řečová aktivita je spojena s artikulací (vydání konsonant ovlivňuje rozkmitání hlasivek při nástupu vokálu) [9].

#### 2.3.5.3 Přesnost artikulace konsonant

Snaha o udržení konstantní rychlosti opakování slabik při DDK úloze může vést ke snížení artikulačních schopností. To se může projevit jako proudní vzduchu přes nedostatečně nastavené artikulátory stejně tak jako úbytek energie při počátečních explozích. Pro správné posouzení dopadu HN na nepřesné nastavení artikulátorů jsou použity parametry *Spektrální moment konsonant (CSM – Consonant Spectral Moment)* a *Útlum spektra konsonant (CST – Consonant Spectral Trend)*. CSM je definováno jako první spektrální moment popisující těžiště energie obsaženého v celém kmitočtovém rozsahu Fourierova spektra. CST je počítáno jako sklon přímky získané z Fourierova spektra v určitém kmitočtovém pásmu. Pro zdůraznění různých spektrálních charakteristik /p/, /t/ a /k/ jsou zvolena různá kmitočtová pásma. Pro konsonanty /p/ je zvoleno pásmo [2500, 3000] Hz pro /t/ [2000, 3000] Hz a pro /k/ [1500, 2500] Hz [9].

#### 2.3.5.4 Oslabení okluzí

Snížení pohybu artikulačních svalů může vést k úniku turbulentního proudění vzduchu v průběhu mezery mezi slabikami. To má za následek zvýšení šumu v mezerách bez řečové aktivity [20]. K popisu šumu obsaženého v mezeře je počítán parametr *Odstup signálu od šumu* (*SNR – Signal to Noise Ratio*) definovaný jako:

$$SNR = 10 \cdot \log_{10}\left(\frac{P_S}{P_N}\right) \tag{2.39}$$

kde $P_S$ představuje výkon signálu a $P_N$ výkon šumu (mezery mezi slabikami).

#### 2.3.5.5 Časování promluvy

Narušení pohybového aparátu neovlivňuje pouze konkrétní svalové skupiny zvlášť. Může také narušit všechny aspekty časování řeči. Proto jsou k vyhodnocení dopadu HN na časování promluvy navrženy další čtyři parametry. První parametr *DDK rychlost* zkoumá celkovou rychlost DDK promluv. Je definován jako počet slabik za jednu sekundu a je počítán jako počet IB za dobu celé promluvy. Druhý parametr *DDK tempo* odhaduje poměr tichých částí během DDK úlohy. Je definován jako průměrná hodnota mezer obsažených v celé DDK promluvě. *DDK tempo* má souvislost s *DDK rychlost*, poskytuje informaci o poměru délky trvání řeči/délky trvání ticha. Třetí parametr *DDK kolísání* odráží schopnost mluvčího udržovat stálý rytmus během DDK úlohy. Je počítán jako směrodatná odchylka trvání tichých mezer v promluvě [9]. Posledním parametrem je průměrná délka trvání vokálů (VO) plus její směrodatná odchylka. Průměr délky vokálů je počítán přes celou promluvu.

Tabulka 1 Definice artikulačních vlastností

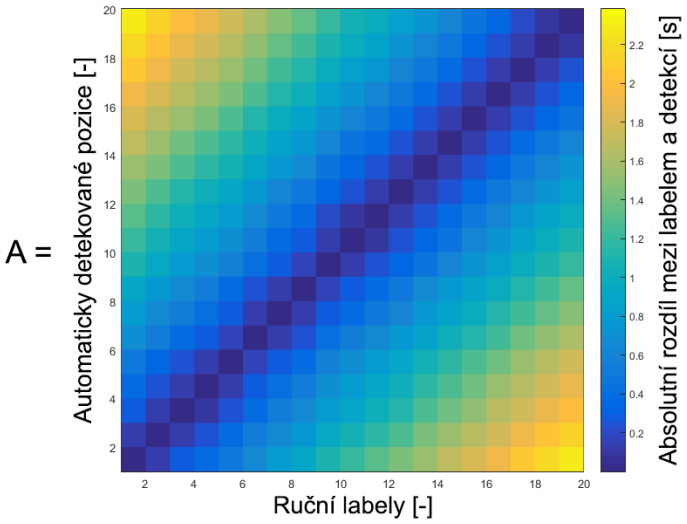
| Název | Interval | Definice parametru |
|---|---|---|
| ***Kvalita řeči*** | | |
| VSQ $_{30}$ | prvních 30 ms po nástupu vokálu | Autokorelační poměr, definující schopnost udržet stabilně hlasitý tón (Vowel Similarity Quotient) |
| ***Koordinace činnosti hlasivek*** | | |
| VOT | počáteční exploze až nástup vokálu | Doba nástupu slabiky, definující délku konsonant (Voice Onset Time) |
| σ VOT | počáteční exploze až nástup vokálu | Směrodatná odchylka délky konsonant |
| ***Přesnost artikulace konsonant*** | | |
| CSM | počáteční exploze až nástup vokálu | První spektrální moment konsonant (Consonant Spectral Moment) |
| CST | počáteční exploze až nástup vokálu | Útlum spektra konsonant, počítaný v definovaném intervalu (Consonant Spectral Trend) |
| ***Oslabení okluzí*** | | |
| SNR | nástup vokálu až okluze (harmonický signál) porovnáván s okluzí až následující počáteční exploze (signál bez řečové aktivity) | Odstup signálu od šumu, reprezentující amplitudu znělé části k části signálu obsahující šum (Signal-to-noise Ratio) |
| ***Časování promluvy*** | | |
| DDK rychlost | celá promluva | Diadochokinetická rychlost, počet slabik za sekundu |
| DDK tempo | okluze až následující počáteční exploze | Diadochokinetické tempo, střední hodnota tichých pauz mezi slabikami |
| DDK kolísání | okluze až následující počáteční exploze | Diadochokinetická nestabilita, míra nestability tichých pauz mezi slabikami |
| VO | nástup vokálu až okluze | Průměrná délka vokálu (Vowel Onset) |
| σ VO | nástup vokálu až okluze | Směrodatná odchylka od průměrné délky vokálů |

## 2.4 Statistika

### *2.4.1 Hodnocení algoritmu*

Před tím, než budou vyhodnoceny příznaky segmentů DDK promluv pro jednotlivé skupiny (HC – kontrolní skupina, HD – skupina pacientů s HN), je potřeba uvést přesnost navrženého algoritmu, s jakou detekuje hranice počátku explozí, nástupu vokálů a okluzí. Přesnost algoritmu je modelována kumulativní distribucí absolutních rozdílů referenčními, ručně stanovenými, labely a pozicemi detekovanými navrženým algoritmem. Pro každou část slabiky (počátek exploze, nástup vokálu a okluzí) je pak kumulativní distribuce počítána zvlášť.

Automatickému detektoru mnohdy některá slabika unikne a počet detekovaných hranic (IB, VO i O) je v tom případě menší než počet ručně stanovených labelů. Naopak v ojedinělých případech detektor nalezne falešné detekce, což může počty vyrovnat. Z těchto důvodů jsou absolutní rozdíly mezi referencí a detekovanými pozicemi hledány v rozdílové matici $A$ (viz Obrázek 2.25). V případě rozdílů počtu labelů a detekcí jsou spočítány falešné i chybějící detekce.




Obrázek 2.25 Matice rozdílů.

Na obrázku je vidět, že v tomto případě je počet detekcí roven počtu referenčních labelů. Je zřejmé, že všech dvacet automatických detekcí odpovídá dvaceti referenčním labelům (minima rozdílů jsou na diagonále). Přesnost na tomto obrázku ovšem není zřetelná.

Nyní jsou nalezeny absolutní rozdíly mezi detekcemi a labely. Dalším krokem je výpočet kumulativní distribuce $CD_i$. Do samotného výpočtu byla zahrnuta i chyba

detektoru, kdy nebyla v blízkosti ručního labelu detekovaná žádná hranice. Výpočet distribuce pak vypadá takto:

$$CD_i = 100 \cdot \frac{\sum_{n=1}^{N} r[n]}{cl}, \qquad r[n] = \begin{cases} 1, & r[n] \leq i \\ 0, & r[n] > i \end{cases} \tag{2.40}$$

kde $r[n]$ jsou nalezené minimální rozdíly mezi detekcí a labelem o celkovém počtu $N$, $cl$ je celkový počet labelů a $i = 1, \ldots, 20$ je práh, pod kterým jsou hledané rozdíly.

### 2.4.2 Hodnocení příznaků

Aby bylo možné zhodnotit rozdíly mezi jednotlivými skupinami mluvčích (HC a HD), proběhlo měření výše uvedených charakteristik pro každou nahrávku z obou skupin. Pokud nahrávka obsahovala více než 7 trojic slabik /pa/-/ta/-/ka/ (celkem 21 slabik), byly nadbytečné slabiky zahozeny. Ze všech naměřených charakteristik byly poté vypočítány střední hodnoty s jejich směrodatnými odchylkami, zvlášť pro skupinu HC i HD. Pro posouzení rozdílů mezi skupinami byl použit *dvouvýběrový t-test*. Pro zvýšení přesnosti výsledků měření je dopočítán parametr *velikost účinku CES* (z anglického Cohen's Effect Size). Ten udává velikost rozdílu mezi jednotlivými skupinami.

Výpočet t-testu vypadá následovně:

$$t = \frac{\mu_x - \mu_y}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}} \tag{2.41}$$

kde $\mu_x$ a $\mu_y$ jsou střední hodnoty vzorků, $\sigma_x$ a $\sigma_y$ jsou jejich směrodatné odchylky a $m, n$ jsou celkové počty jednotlivých vzorků [17].

Cohenova velikost účinku je počítána podle:

$$CES = \frac{\mu_x - \mu_y}{\sqrt{0{,}5 \cdot \left(\sigma_x^2 + \sigma_y^2\right)}} \tag{2.42}$$

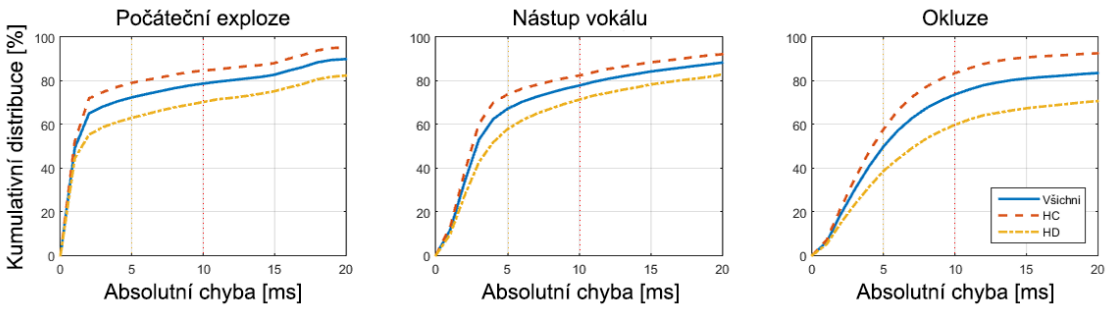
# 3 Výsledky

## 3.1 Hodnocení algoritmu

Na začátku byla zhodnocena celá skupina (HC i HD). Ta obsahovala 8568 slabik. Další dvě distribuce se poté počítaly ze dvou skupin (HC – 5021 slabik a HD – 3547 slabik). Pro zvolení procentuální úspěšnosti byl zvolen práh a 5 ms, po vzoru předchozí studie [12]. Obrázek 3.1 ukazuje kumulativní sumu reprezentující absolutní rozdíl mezi referenčními ručně nalezenými labely a automaticky detekovanými pozicemi počátečních explozí (vlevo), nástupů vokálů (uprostřed) a okluzí (vpravo). Vzhledem ke zvolenému prahu absolutního rozdílu mezi referencí a automatickými detekcemi 5 ms je úspěšnost navrhovaného algoritmu rovna 79,01% pro počáteční explozi, 73,73% pro nástup vokálů a 57,84% pro okluzi. Tyto výsledky odpovídají skupině HC (na obrázku červeně). Pro skupinu pacientů s Huntingtonovou nemocí (HD) je úspěšnost se stejným prahem rovna 62,98% pro počáteční explozi, 57,82% pro nástup vokálů a 44,12% pro okluzi (na obrázku žlutě). Pro všechny mluvčí jsou poté výsledky úspěšnosti následující: 72,37% pro počáteční explozi, pro nástup vokálů 67,15% a pro okluze 49,92% (na obrázku modře). Pokud by byl v úvahu brán práh 10 ms, přesnost detekcí vzroste v řádu jednotek procent (viz Tabulka 2).

Tabulka 2 Přesnost automatických detekcí vzhledem k ručním referencím.

| **Skupina** | **HC** | | **HD** | | **Vše** | |
|---|---|---|---|---|---|---|
| **Práh** | ***5 ms*** | ***10 ms*** | ***5 ms*** | ***10 ms*** | ***5 ms*** | ***10 ms*** |
| **Exploze** | 79,01% | 84,64% | 62,98% | 70,23% | 72,37% | 78,68% |
| **Vokál** | 73,73% | 82,35% | 57,82% | 71,27% | 67,15% | 77,77% |
| **Okluze** | 57,84% | 83,51% | 44,12% | 59,77% | 49,92% | 73,68% |




Obrázek 3.1 Kumulativní distribuce absolutních rozdílů mezi detekcemi a referencí.

Téměř všechny počítané parametry jsou energeticky závislé, proto je v některých případech těžké rozeznat slabiku s velmi nízkou energií od okolního šumu. S tím jsou spojené i chyby detektoru při náhlém úbytku energie u jedné slabiky v energeticky silném projevu. Takovému „šepotu“ je mnohdy velmi energeticky podobná náhlá respirace mluvčího a dochází tak k falešné detekci. Tento fakt však dobře eliminuje prostor vlnkové transformace. Díky tomu nejsou respirace v pauzách falešně detekovány jako řečová aktivita.

## 3.2 Statistické porovnání HC a HD

Tabulka 3 Přehled výsledků

| # | parametr | HC | | HD | | CES† |
|---|---|---|---|---|---|---|
| | | μ | σ | μ | σ | |
| ***Kvalita řeči*** | | | | | | |
| 1 | VSQ $_{30}$ [-] | 0,37 | 0,09 | 0,40 | 0,11 | 0,28 |
| ***Koordinace činnosti hlasivek*** | | | | | | |
| 2 | VOT $_{vše}$ [ms] | 6,64 | 1,30 | 7,89 | 1,30 | 0,96*** |
| 3 | VOT $_{pa}$ [ms] | 5,34 | 1,36 | 7,05 | 2,09 | 0,97*** |
| 4 | VOT $_{ta}$ [ms] | 6,51 | 1,80 | 7,62 | 1,91 | 0,59*** |
| 5 | VOT $_{ka}$ [ms] | 8,06 | 2,49 | 8,99 | 2,33 | 0,39** |
| 6 | σ VOT $_{vše}$ [ms] | 2,80 | 0,96 | 3,72 | 0,96 | 0,96*** |
| 7 | *σ* VOT $_{pa}$ [ms] | 1,82 | 1,04 | 2,73 | 1,58 | 0,68*** |
| 8 | *σ* VOT $_{ta}$ [ms] | 1,98 | 1,25 | 2,82 | 1,54 | 0,59*** |
| 9 | *σ* VOT $_{ka}$ [ms] | 2,36 | 1,51 | 3,87 | 1,59 | 0,97*** |
| ***Přesnost artikulace konsonant*** | | | | | | |
| 10 | CSM $_{pa}$ [kHz] | 2,91 | 0,42 | 2,86 | 0,46 | 0,11 |
| 11 | CSM $_{ta}$ [kHz] | 2,56 | 0,37 | 2,69 | 0,51 | 0,29* |
| 12 | CSM $_{ka}$ [kHz] | 2,47 | 0,43 | 2,75 | 0,48 | 0,61*** |
| 13 | CST $_{pa}$ [rad·10$^{-6}$] | -6,02 | 10,32 | -2,21 | 5,55 | 0,46** |
| 14 | CST $_{ta}$ [rad·10$^{-6}$] | -3,98 | 5,58 | -3,62 | 5,35 | 0,06 |
| 15 | CST $_{ka}$ [rad·10$^{-6}$] | -3,45 | 4,24 | -3,07 | 4,07 | 0,09 |
| ***Oslabení okluzí*** | | | | | | |
| 16 | SNR [dB] | 7,27 | 2,35 | 15,39 | 4,24 | 2,37*** |
| ***Časování promluvy*** | | | | | | |
| 17 | DDK rychlost [slabika/s] | 17,95 | 2,12 | 11,69 | 3,72 | 2,07*** |
| 18 | DDK tempo [ms] | 30,37 | 6,11 | 68,14 | 47,73 | 1,11*** |
| 19 | DDK kolísání [ms] | 7,56 | 4,58 | 30,53 | 24,72 | 1,29*** |
| 20 | VO [ms] | 21,09 | 3,22 | 26,51 | 5,93 | 1,13*** |
| 21 | σ VO [ms] | 4,23 | 2,22 | 9,61 | 4,92 | 1,41*** |

† Přesnost výsledku je označena hvězdičkou
*) p<0,05 **) p<0,01 a ***) p<0,001.

Tabulka 3 obsahuje stručný přehled všech charakteristiky jednotlivých měření z kapitoly 2.4.2 obsahující střední hodnoty a směrodatné odchylky hodnot a jejich velikosti účinku. Jsou zde rozlišeny i jednotlivé skupiny mluvčích (HC a HD).

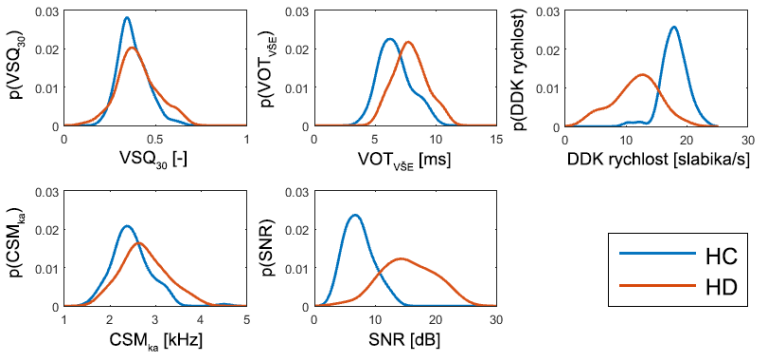
V rámci kvality řeči vychází VSQ $_{30}$ u pacientů s Huntingtonovou nemocí takto: ($t(195) = 1,97$, CI = [$-9,40E^{-5}$, $5,25E^{-2}$], $p < 0,05$). Z těchto výsledků i z Tabulka 3Tabulka 1 je zřejmé, že tento parametr pro účely separace zdravých a nemocných lidí neobstál.

Při měření koordinace ovládání hlasivek vychází VOT $_{vše}$ ($t(195) = 6,56$, CI = [$8,73E^{-1}$; 1,62], $p < 0,001$) pro charakteristiku VOT $_{pa}$ ($t(195) = 6,95$, CI = [1,23, 2,20], $p < 0,001$), VOT $_{ta}$ ($t(195) = 4,11$, CI = [$5,75E^{-01}$, 1,64], $p < 0,001$) a VOT $_{ka}$ ($t(195) = 2,62$, CI = [$2,32E^{-01}$, 1,63], $p = 0,009$). Charakteristiky VOT ukázaly, že podávají vynikající výsledky při vyhodnocování HN. Je to zřejmé i v Tabulka 3, kde, až na případ VOT $_{ka}$, vykazují vysokou CES. Směrodatné odchylky charakteristik VOT pak vychází ($t(195) = 6,57$, CI = [$6,45E^{-1}$, 1,20], $p < 0,001$) pro σ VOT $_{vše}$, pro σ VOT $_{pa}$ ($t(195) = 4,89$, CI = [$5,45E^{-1}$, 1,28], $p < 0,001$), σ VOT $_{ta}$ ($t(195) = 4,20$, CI = [$4,46E^{-1}$, 1,24], $p < 0,001$) a σ VOT $_{ka}$ ($t(195) = 6,72$, CI = [1,07, 1,95], $p < 0,001$).

Výsledky v rámci přesnosti artikulace konsonant, jak vidím už v Tabulka 3, se nesešly s velkým úspěchem. Pro první spektrální mement konsonant CSM nejlépe charakterizuje pacienty s HN parametr CSM $_{ka}$ ($t(195) = 4,21$, CI = [$1,47E^{2}$, $4,07E^{2}$], $p < 0,001$), hned za ním je CSM $_{ta}$ ($t(195) = 2,10$, CI = [7,96, $2,56E^{2}$], $p = 0,03$) a v poslední řadě je CSM $_{pa}$ ($t(195) = -0,78$, CI = [$-1,75E^{2}$, $7,58E^{1}$], $p = 0,4$). Útlum spektra konsonant také nepřináší silné výsledky. Konkrétně CST $_{pa}$ ($t(195) = 2,97$, CI = [$1,28E^{-6}$, $6,33E^{-6}$], $p = 0,003$), CST $_{ta}$ ($t(195) = 0,44$, CI = [$-1,23E^{-6}$, $1,94E^{-6}$], $p = 0,6$) a CST $_{ka}$ ($t(195) = 0,61$, CI = [$-8,28E^{-7}$, $1,58E^{-6}$], $p = 0,5$).

Odstup signálu od šumu, popisující chování artikulačních svalů, svými výsledky dobře ukazuje rozdíl mezi zdravými lidmi a lidmi s HN ($t(195) = 17,27$, CI = [7,20, 9,05], $p < 0,001$).

Poslední sada parametrů stejně jako SNR dokáže rozdělit naše dvě porovnávané skupiny. DDK rychlost ukazuje, že pacienti s HN nedokáží opakovat slabiky tak rychle jako lidé zdraví ($t(195) = -15,05$, CI = [-7,08, -5,44], $p < 0,001$), stejně tak DDK tempo ($t(195) = 8,57$, CI = [$2,91E^{1}$, $4,65E^{1}$], $p < 0,0001$) i měření DDK kolísání ($t(195) = 9,93$, CI = [$1,84E^{1}$, $2,75E^{1}$], $p < 0,001$). Nakonec i měření délky vokálů se ukazuje jako dobrý rozhodovací nástroj. VO ($t(195) = 8,28$, CI = [4,13, 6,71], $p < 0,001$) a σ VO ($t(195) = 10,45$, CI = [4,37, 6,40], $p < 0,001$).

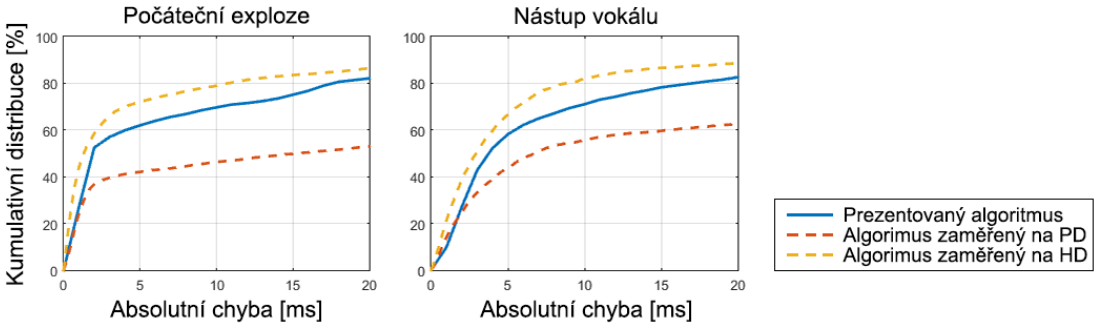



Obrázek 3.2 Hustoty pravděpodobnosti pěti reprezentativních charakteristik.

## 3.3 Porovnání s konvenční metodou

Výsledky algoritmu navrženého v této práci (na Obrázek 3.3 modře) jsou porovnány s konvenčním algoritmem pro segmentaci DDK promluv pacientů s Parkinsonovou nemocí [9] (na Obrázek 3.3 červeně). Na první pohled je zřejmé, že algoritmus navržený v této práci dosahuje výrazně lepších výsledků. Při stanovení prahu 5 ms dosahuje prezentovaný algoritmus vyšší přesnosti o 20% v případě počátečních explozí a o 15% v případě nástupu vokálů.

Porovnání výsledků je rozšířeno o další algoritmus, do kterého byla zapojena i část této diplomová práce. Jedná se o metodu založenou na detekci vokálů pomocí LP residua (viz *Automatic detection of voice onset time in dysarthric speech* [12]). V tomto případě oba algoritmy dosahuji srovnatelné úspěšnosti, kdy skóre algoritmu založeného na dynamickém prahování bývá o jednotky procent nižší.




Obrázek 3.3 Porovnání tří algoritmů. Kumulativní distribuce absolutního rozdílu mezi detekovanou a referenční hodnotou. Výsledky pro počáteční explozi a nástup vokálu.

# 4 Diskuze

Hlavním přínosem této práce je metoda automatického vyhodnocování dysartrické řeči u pacientů s Huntingtonovou nemocí. Segmentace nahrávek na jednotlivé slabiky s detekcemi explozí a vokálů umožňuje následné statistické měření charakteristik popisující kvalitu řeči.

Tato práce je primárně zaměřená na pacienty mluvící česky, na kterých byl navržený algoritmus vyvíjen i otestován. Cizojazyčné nahrávky mohou zásadně snížit přesnost celé metody. Další limitací pro správný chod algoritmu jsou DDK úlohy ve tvaru /pa/-/ta/-/ka/. Prezentovaná metoda předpokládá opakování těchto slabik, tedy opakování sekvence „ticho – konsonanta – vokál“. V případě detekce hlásek ve větách by tento algoritmus s vysokou pravděpodobnost neobstál. Pokud by byla DDK úloha /pa/-/ta/-/ka/ nahrazena úlohou podobného rázu, např. /pa/-/pa/-/pa/, lze předpokládat, že úspěšnost detekcí bude srovnatelná. Tento experiment však nebyl realizován.

Procentuální úspěšnost detekce explozí i okluzí je závislá na úspěšnosti detekce vokálů, která v algoritmu probíhá jako první a určuje tím i úspěšnost celého algoritmu. S ohledem na pozvolné ubývání energie v okluzích je velmi obtížným úkolem najít přesnou hranici okluze i při ruční segmentaci. Především z těchto důvodů dosahuje navržený algoritmu v této oblasti nižší přesnost.

Při porovnání prezentovaného algoritmu s metodou založenou na LP residuu [12] je skóre sice o jednotky procent nižší, nicméně segmentace pomocí dynamického prahování je robustnější z hlediska falešných detekcí respirací, které jsou v projevech pacientů s HN časté.

Do budoucna by tento algoritmus mohl být rozšířen o některé další rozměry parametrického prostoru pro rozhodování, tak i zpřesnění v rozhodování první správné změny u Bayesovského detektoru změn (u vokálů) či ve spektrogramu (u konsonant).

Z výsledků navržených parametrů popisující kvalitu řeči můžeme vydedukovat, že pro správnou klasifikaci dysartrické řeči dosahuje nejlepších výsledků skupina charakteristik zaměřených na *Časování promluvy* společně s *Koordinací činnosti hlasivek*. Jak lze předpokládat, u pacientů s HN při DDK úlohách klesá rychlost opakování slabik (z 18 slabik/s u HC na 12 slabik/s u HD) a délky pauz mezi slabikami se stávají nestabilními (ze 7 ms u HC na 30 ms u HD). Stejně tak u *Koordinace činnosti hlasivek* délka konsonant (VOT) se prodlužuje a má nestabilní délku v průběhu jedné promluvy. Všechny tyto předpoklady byly porovnáním skupiny zdravých lidí s pacienty s HN potvrzeny a můžeme tedy potvrdit, že navržený algoritmus pro vybraná měření dokáže klasifikovat hyperkinetickou dysartrii u pacientů s HN.

# 5 Závěr

Hlavními cíli této práce bylo navrhnout algoritmus pro segmentaci patologických promluv pacientů trpících Huntingtonovou nemocí společně s ohodnocením vhodných řečových příznaků pro popis charakteristik dysartrické řeči. Metoda dynamického prahování pomocí GMM s vybraným prostorem parametrů dosáhla velmi dobrých výsledků. Díky tomu bylo možné na segmentovaných promluvách provézt statistická měření, která úspěšně rozlišila zdravé mluvčí od pacientů s diagnostikovanou Huntingtonovou nemocí. Hlavních cílů práce tedy bylo dosaženo.

Navržený algoritmus dosahuje přesnosti 63% při detekci počátečních explozí, 58% při hledání hranice nástupu vokálů a 44% při detekci okluzí. Tyto výsledky splňují hranici absolutního rozdílu do 5 ms. Při zvýšení prahu na 10 ms výsledky vzrostou o jednotky procent (IB - 70%, VO - 71% a O - 60%).

Klasifikace promluv do dvou skupin HN a HC na základě řečových příznaků nejlépe dopadla pro skupiny charakteristik zaměřených na *Časování promluvy* společně s *Koordinací činnosti hlasivek*. Zde kromě měření délky počátečního velární /*k*/, kde se klasifikátor dopouštěl chyby menší než 5%, vyšly všechny charakteristiky ze zmiňovaných skupin s věrohodností větší než 99,9%.

# Seznam obrázků



# Seznam tabulek

# Seznam použité literatury


[1] F. O. Walker, “Huntington's disease,” *The Lancet,* vol. 369, no. 9557, pp. 218-228, 2007.

[2] J. R. DUFFY, Motor speech disorders: substrates, differential diagnosis, and management., New York: Mosby, 2013.

[3] R. A. Macdonell and R. Holmes, “Motor Speech and Swallowing Disorders,” in *Neurology and Clinical Neuroscience*, Elsevier Inc., 2007, pp. 155-170.

[4] J. Psutka, L. Müller, J. Matoušek a V. Eadová, Mluvíme s počítačem česky, Praha: ACADEMIA, 2006.

[5] D. Kempler a D. Van Lancker, „Effect of Speech Task on Intelligibility in Dysarthria: A Case Study of Parkinson’s Disease,“ *Brain and Language,* č. 80, p. 449–464, 2002.

[6] J. Rusz, R. Čmejla a H. Růžičková, „Hodnocení důrazu, emocí, rytmu, artikulační rychlosti a pravidelnosti u Parkinsonovy nemoci,“ v *Mezinárodní konference Technical Computing Bratislava 2010*, Bratislava, 2010.

[7] H. S. Group, “Unified Huntingto’s Disease Ratin Scale: reliability and consistency,” *Movment Disorders,* no. 11, pp. 136-142, 1996.

[8] J. Rusz, J. Klempíř, T. Tykalová, E. Baborová, R. Čmejla, E. Růžička and J. Roth, “Characteristics and occurrence of speech impairment in Huntington's disease: possible influence of antipsychotic medication,” *Journal of Neural Transmission,* vol. 121, no. 12, pp. 1529-1539, 2014.

[9] M. Novotný, J. Rusz, R. Čmejla a E. Růžička, „Automatic evaluation of articulatory disorders in Parkinson's disease,“ *The IEEE/ACM Transactions on Audio, Speech, and Language Processing,* sv. 22, č. 9, pp. 1366-1378, 2014.

[10] Y. Wang, R. D. Kent, J. R. Duffy, J. E. Thomas and G. Weismer, “Alternating motion rate as an index of speech motor disorder in traumatic brain injury.,” *Clinical Linguistics & Phonetics,* vol. 18, no. 1, pp. 57-84, 2004.

[11] L. E. Volaitis a J. L. Miller, „Phonetic prototypes: influence of place of articulation and speaking rate on the internal structure of voicing categories,“ *The Journal of the Acoustical Society of America,* sv. 92, pp. 723-735, 1992.

[12] M. Novotný, J. Pospíšil, R. Čmejla and J. Rusz, “Automatic detection of voice onset time in dysarthric speech,” in *IEEE/ICASSP*, Brisbone, Australia, 2015.

[13] J. Uhlíř, P. Sovka, P. Pollák, V. Hanžl a R. Čmejla, Technologie hlasových komunikací, Praha: Nakladatelství ČVUT, 2007.

[14] S. Prasanna, B. Reddy and P. Krishnamoorthy, “Vowel Onset Point Detection Using Source, Spectral Peaks, and Modulation Spectrum Energies,” *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 17, no. 4, pp. 556 - 565, 2009.

[15] H. MISRA, S. IKBAL, H. BOURLARD and H. HERMANSKY, “Spectral entropy based feature for robust ASR,” in *IEEE/ICASSP*, Montreal, Quebec, Canada, 2004.

[16] R. Šmíd, „Úvod do vlnkové transformace,“ ČVUT , katedra měření, Praha, 2001.

[17] I. The MathWorks, „MathWorks®,“ The MathWorks, Inc., 1994-2015. [Online]. Available: http://www.mathworks.com/. [Přístup získán 10 Listopad 2014].

[18] R. Singh, M. L. Seltzer, B. Raj a R. M. Stern, „Speech in noisy environments: robust automatic segmentation, feature extraction, and hypothesis combination, Acoustics, Speech, and Signal Processing,“ v *IEEE/ICASSP*, Salt Lake City, 2001.

[19] S. Sapir, L. Ramig, J. Spielman a C. Fox, „Formant centralization ratio: A proposal for a new acousticmeasure of dysarthric speech,“ *Journal of Speech Language and Hearing Research,* č. 53, p. 114–125, 2010.

[20] D. Duez, “Acoustic analysis of occlusive weakening in parkinsonian French speech,” in *International Congress Phonetic Sciences*, Saarbrücken, 2007.

# Publikace autora

NOVOTNÝ, Michal, Jakub Pospíšil, Roman Čmejla, and Jan Rusz, *Automatic Detection of Voice Onset Time in Dysarthria* in Proc. IEEE/ICASSP, Brisbane, Australia, 2015, (in press).

# Příloha I.

## Obsah CD

**\TEXT**

Elektronická verze diplomové práce *Diplomová práce.pdf*

**\METODA\database\labels**

Ruční značky v souborech *name.mat*

**\METODA\database\OUT**

Výstupy algoritmu *detektorDDK.m* v souborech *name.mat*

**\METODA\database\records**

Nahrávky mluvčích v souborech *name.wav*

**\METODA\Detektor**

Obsahuje hlavní skript *detektrorDDK.m* spolu s testovacím skriptem *TEST_DDK.m*

**\METODA\Detektor\funkce**

Zde jsou uloženy všechny potřebné funkce k běhu skriptu *detektroDDK.m*

**\METODA\Porovnani**

Skripty *POROVNANI.m* a *CUMDIST.m* pro porovnání přesnosti detektoru s ručními značkami. Společně se skriptem pro vykreslení výsledků *ZOBRAZ.m*

**\METODA\Porovnani\funkce**

Pomocné funkce pro běh skriptu *POROVNANI.m*

**\METODA\Statistika**

Skript *Features.m* pro výpočet charakteristik popisující kvalitu řeči. *Stat.m* pro výpočet statistických výsledků.

**\METODA\Statistika\funkce**

Pomocné funkce pro výpočet charakteristik.