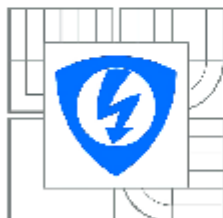




VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ**
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

ZVÝRAZŇOVÁNÍ DEGRADOVANÉ ŘEČI POMOCÍ ODSTRANĚNÍ DISONANTNÍCH SLOŽEK

SPEECH ENHANCEMENT USING CANCELLING OF DISSONANT COMPONENTS

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

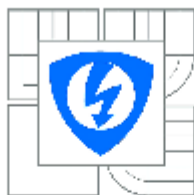
AUTOR PRÁCE
AUTHOR

RADIM STUDENÝ

VEDOUCÍ PRÁCE
SUPERVISOR

prof. Ing. ZDENĚK SMÉKAL, CSc.

BRNO 2015



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Radim Studený
Ročník: 3

ID: 154878
Akademický rok: 2014/2015

NÁZEV TÉMATU:

Zvýrazňování degradované řeči pomocí odstranění disonantních složek

POKYNY PRO VYPRACOVÁNÍ:

Zvýraznění jedno-kanalového řečového signálu, který je znehodnocen šumem a rušením, je náročná úloha, protože máme k dispozici pouze jeden záznam této směsi bez znalosti intenzity a směru, odkud řeč i rušení přichází. V praxi jsou využívány různé typy metod. Cílem práce je využít znalostí ze zpracování hudebního signálu a číslicovou filtrací odstranit kmitočtová pásma, která projevují disonanci vůči základnímu tónu řeči. Výsledky by měly být ověřeny percepčními metodami zjištění kvality a srozumitelnosti řeči.

DOPORUČENÁ LITERATURA:

- [1] SMÉKAL, Z.: Číslicové zpracování signálu (MCSI). Elektronické učební texty pro magisterské studium, VUT Brno, 2009.
- [2] PSUTKA, J., MÜLLER, L., MATOUŠEK, J., RADOVÁ, V.: Mluvíme s počítačem česky. Academia, Praha 2006. ISBN 80-2100-1309-1
- [3] SYROVÝ V.: Hudební akustika. Akademie múzických umění, Praha 2003. ISBN 80-7331-901-2

Termín zadání: 9.2.2015

Termín odevzdání: 2.6.2015

Vedoucí práce: prof. Ing. Zdeněk Smékal, CSc.

Konzultanti bakalářské práce:

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

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Tato práce si klade za cíl odstranit rušivé vlivy z řečového signálu a tím zvýšit srozumitelnost, kvalitu degradovaného signálu a odstup od šumu. Nejčastější typy rušení mohou být hluk ulice, vítr přicházející na mikrofon, cizí rozhovory na pozadí nebo puštěná hudba. Metoda popsaná v téhle práci filtruje frekvenční pásma signálu, která jsou vzhledem k základnímu tónu řeči disonantní. Konkrétně pro referenční tón C to jsou tóny F#, B a C#.

KLÍČOVÁ SLOVA

Základní tón řeči, disonantní frekvence, zvýrazňování degradovaného signálu

ABSTRACT

This work is supposed to remove interferences from speech signal and so increase its clarity, quality of degraded signal and signal-to-noise ratio. The most common sources of interference could be street noise, wind coming on a microfon, speech on the background or music. The metod described in this work remove frequency bands of a signal, which in relation to the fundamental frequency of a speech are disonant. Be specific, to reference tone C there are F#, B a C# tones.

KEYWORDS

Fundamental frequency of a speech, disonant frequency, enhancing degraded signal

STUDENÝ, R. *Zvýrazňování degradované řeči pomocí odstranění disonantních složek*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2015. 31 s. Vedoucí bakalářské práce prof. Ing. Zdeněk Smékal, CSc..

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma Zvýrazňování degradované řeči pomocí odstranění disonantních složek jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne

.....

(podpis autora)

PODĚKOVÁNÍ

Děkuji vedoucímu bakalářské práce prof. ing. Zdeněk Smékal, CSc. za účinnou metodickou, pedagogickou a odbornou pomoc a další cenné rady při zpracování mé bakalářské práce.

V Brně dne

.....

(podpis autora)

OBSAH

Obsah	vi
Úvod	1
1 Řeč	2
1.1 Proces tvorby řeč	2
1.2 Rozdělení hlásek	2
1.2.1 Samohlásky	2
1.2.2 Souhlásky	2
1.3 Základní tón řeči	3
2 Hudba	3
2.1 Barva tónu	3
2.2 Konsonance, disonance	4
3 Nalezení základního tónu řeči	6
3.1 Segmentace	6
3.2 Metody	7
3.2.1 Detekce v časové oblasti – Autokorelace	7
3.2.2 Detekce v časové oblasti – centrální klipování	8
3.2.3 Detekce v časové oblasti – Inverzní filtrace	10
3.2.4 Detekce v kmitočtové oblasti – Spektrální metoda	11
3.2.5 Detekce v kepstru – Kepstrální metoda	12
3.3 Výběr metody	12
Výběr	13
4 Návrh metody filtrace	14
4.1 Segmentace	14
4.2 Prahování	15
4.3 Normalizace	17
4.4 Autokorelace	18
4.5 Filtrace	20
4.5.1 1. Filtrovací větev	20
4.5.2 2. Filtrovací větev	21

4.5.3	Sloučení 1. a 2. filtrovací větve	23
4.6	Sloučení segmentů	24
5	Testování metody	26
6	Závěr	28
	literatura	29
	Seznam zkratk	30
	Seznam příloh	31

ÚVOD

Cílem této práce je zlepšení výkonu systémů hlasové komunikace, kdy tyto signály jsou znehodnoceny rušením nebo hlukem. Za zlepšení považujeme snížení účinků rušení a tím zvýšit výkon daných systémů, zvýšit odstup signálu od šumu a tím i zvýšit srozumitelnost.

Předpokládá se, že takhle upravený signál umožní přesnější zpracování signálu v dalších aplikacích, které mohou být například rozpoznávání řeči - slov, slovních spojení či plynulé řeči, rozpoznávání mluvčích, případně rozpoznávání specifických jevů, jako třeba diagnostika řečových orgánů.

Potřeba zvýraznění řečových signálů je potřeba v různých situacích, kdy signál pochází z hlučného prostředí. Komunikace probíhá mezi lidmi, nebo mezi lidmi a strojem. První část práce ve stručnosti popisuje proces vzniku řeči a základní fyzikální vlastnosti, aby mohla být problematika zpracována po technické stránce. Důležité je rozdělení řeči z fonetického hlediska na znělé a neznělé hlásky. Tyto skupiny lze popsat skupinou příznaků, které jsou důležité pro jejich zpracování. Důležitý je zejména jejich vztah k základnímu tónu řeči, který je zásadní pro tuhle práci.

Druhá část se zabývá zvukem po hudební stránce. Popisuje vztah mezi tónem a jeho barvou a souvislost barvy tónu s tím jak jsou jednotlivé složky barvy libozvučné či nelibozvučné v souzvuku s tónem základním. Dochází zde k propojení hudebního a technického pohledu na zvukové signály.

Třetí část uvádí jakými metodami lze získat základní tón řeči a popisuje jejich vlastnosti. V závěru této části se zabýváme volbou nejvhodnější metody pro naši aplikaci.

V závěrečné kapitole je ukázán návrh metody pro filtraci disonantních složek řečového signálu. Popisuje funkci jednotlivých částí, které jsou nezbytné pro správnou funkci algoritmu a uvádí případné problémy, které mohou při zpracování signálu nastat.

1 ŘEČ

1.1 Proces tvorby řeč

Princip tvorby řeči je proces, jehož hnací silou je proud vzduchu dodávaný plícemi. Proud vzduchu prochází hlasivkovou štěrbinou, která je obklopena hlasivkami. Hlasivky jsou rozkmitávány proudem vzduchu z plic, čímž tvoří pravidelný budící signál.

Budící signály dále procházejí dutinou hrdelní, dutinou ústní a dutinou nosní (hlasový trakt), kde jsou zpracovány pohybem řečových orgánů (artikulace).

Hrtan a hlasivky tvoří buď periodický signál, který je typický pro znělé hlásky nebo šumový signál pro neznělé hlásky. [1]

1.2 Rozdělení hlásek

Z fonetického hlediska se dělí čeština na samohláskový a souhláskový systém.

1.2.1 Samohlásky

- Všechny samohlásky jsou znělé
- Téměř periodický signál
- Obsažen základní tón
- Vysoká energie signálu

1.2.2 Souhlásky

Neznělé

- Šumový signál bez známek periodicity
- Neobsahují základní tón
- Nízká energie signálu rozdělena téměř rovnoměrně přes kmitočtové pásmo

Znělé

- Kombinace periodického a šumového signálu
- Obsažen základní tón
- Vysoká energie signálu

[1]

1.3 Základní tón řeči

Základní tón řeči je významný parametr signálu v kmitočtové oblasti a uplatňuje se v mnoha aplikacích zpracovávající řečový signál. Jeho průběh se jeví jako melodie řeči. Základní kmitočtový rozsah se uvádí mezi 60 a 400 Hz. Tenhle kmitočet se liší u osob různého pohlaví a také u osob různého věku.

Obecně lze říct, že nejnižší kmitočet mají muži a to 80-160Hz. U žen se pohybuje v rozmezí 150-300Hz a u dětí mezi 200-600Hz. S rostoucím věkem však základní frekvence klesá.

Při normální promluvě se základní tón pohybuje v rozmezí jedné oktávy.[5]

2 HUDBA

Hudební teorie definuje základní vlastnosti zvuku jako:

- Délka trvání
- Síla
- Výška tónu
- Barva tónu

2.1 Barva tónu

Z výšky tónu a jeho barvy budeme vycházet. Každý tón má pro naše ucho zcela charakteristický zvuk. Zdroj zvuku generuje kromě základního tónu (určuje výšku) další alikvótní tóny (jeho vyšší harmonické). To jakou má tón barvu určuje právě počet přítomných vyšších harmonických a jejich intenzita. To, jakou má tón barvu nám umožňuje rozeznat dva rozdílné hudební nástroje, stejně tak jako dva rozdílné mluvčí. Barva tónu je daná rozdílným způsobem vzniku zvuku v různých rezonátorech (hrdelní, ústní a nosní dutina), které se liší svou velikostí, tvarem či materiálem. [4]

Frekvence tónů v závislosti na základním tónu v intervalu jedné oktávy můžeme definovat jako:

$$F_k = F_0 \cdot 2^{(k/12)}, k = 0, 1, \dots, 12 \quad (1)$$

Kde F_0 je frekvence základního tónu a F_k je frekvence k-tého půltónu (číslo 12 je zvoleno z toho důvodu, že oktáva obsahuje 12 půltónů)

Nutno podotknout, že vzorec (1) vychází z temperovaného ladění. Namísto složitějšího přirozeného ladění je použito temperované, jelikož odstraňuje jeho nevýhody. Jeho

myšlenka spočívá v konsonanci (libozvučnosti) tónu a tónu dvojnásobné frekvence. Tenhle interval je označován jako oktáva. Základní interval jedné oktávy je následně rozdělen na 12 pravidelných intervalů tak, aby poměry dvou sousedních frekvencí byly vždy stejné. Z toho vyplývá, že tento poměr musí být dvanáctá odmocnina ze dvou, to je $\sqrt[12]{2} = 1,05946$. Číslo dvanáct bylo zvoleno, protože takové ladění dobře aproximuje přirozené ladění a jeho složitost se nezvyšuje. Temperované ladění je tedy nedokonalé, ale většina lidí nedokáže rozeznat rozdíl a tudíž nebudou vnímat takovou hudbu jako rozladěnou. [2]

Interval je definován jako poměr frekvence daného půltónu k frekvenci základního tónu. Zlomky

$$F_k/F_0, \text{ pro } k = 0, 1, \dots, 12 \quad (2)$$

kde „k“ je číslo půltónu představují jednotlivé intervaly v jedné oktávě. Jsou to 1/1, 135/128, 9/8, 6/5, 5/4, 4/3, 45/32, 3/2, 8/5, 27/16, 9/5, 15/8, 2/1.

Když je interval jednodušší vzorek, potom interval jako akord tónů je stabilnější a proto taky libozvučnější. Naopak, když je zlomek složitý, stabilita intervalu je menší a tudíž se jeví jako disonantní. [4]

2.2 Konsonance, disonance

Konsonance (libozvučnost) a disonance (nelibozvučnost) nejsou přesně rozděleny. Můžeme je však v rámci jedné diferencované stupnice rozdělit, kde na jednom konci stupnice se nacházejí stabilní intervaly a vedou k nestabilitě na druhém konci stupnice. Takto rozdělené intervaly lze rozřadit do čtyř skupin.

- Ideální (úplný) souzvuk
Prima 1/1, oktáva 2/1, kvinta 3/2, kvarta 4/3
- Nedokonalý (neúplný) souzvuk
Velká tercie 5/4, velká sexta 5/3, malá tercie a malá sexta 8/5
- Nedokonalá (neúplná) disonance
Malá septima 9/5, velká sekunda 9/8
- Dokonalá (úplná) disonance
Malá sekunda 135/138, zvětšená kvarta 45/32, velká septima 15/8

[4]

0	C	Prima	1/1	++
1	C#	Malá sekunda	135/138	--
2	D	Velká sekunda	9/8	-
3	D#	Malá tercie	6/5	+
4	E	Velká tercie	5/4	+
5	F	Čistá kvarta	4/3	++
6	F#	Zvětšená kvarta	45/32	--
7	G	Čistá kvinta	3/2	++
8	G#	Malá sexta	8/5	+
9	A	Velká sexta	5/3	+
10	B (A#)	Malá septima	9/5	-
11	H	Septima	15/8	--
12	C'	Oktáva	2/1	++

Tabulka 2-1: Přehledová tabulka pro půltóny „k“, jim přiřazen název tónu, hudební interval vztáhnutý k základnímu tónu C, interval vyjádřený zlomkem, ++ vyjadřuje úplný souzvuk, +vyjadřuje neúplný souzvuk, - neúplná disonance a -- kompletní disonance

Z pohledu, jak zvuk vnímáme je hudební interval definován jako konsonantní, když se jeho zvuk jeví jako příjemný. Disonantní interval se jeví jako nepříjemný či drsný.

V závislosti k základnímu tónu frekvence půltónů, které s ním tvoří souzvuk ve všech oktávách jsou vyjádřeny vztahem:

$$F_d = F_0 * 2^{(n+(k/12))}, n = 0,1,...,7, k = 0,1,...,12 \quad (3)$$

kde F_0 je frekvence základního tónu, „n“ je číslo oktávy a „k“ je číslo půltónu. Když vezmeme tón C jako referenční tón, potom jeho disonanční půltóny budou B, F# a C# ve všech oktávách. Pro „k“ tedy bude platit $k = \{1,6,11\}$. [2]

3 NALEZENÍ ZÁKLADNÍHO TÓNU ŘEČI

Jak již bylo řečeno, základní tón řečového signálu vyjadřuje kmitání hlasivek při vytváření znělých úseků řeči. Jeho nalezení je tedy stěžejní pro filtrování disonančních frekvencí. V případě jeho nesprávného určení se totiž může stát, že budou filtrovány nesprávné části spektra, což může snížit srozumitelnost již tak degradovaného signálu. Hlavní problémy při určování hodnot základního tónu vyplývají z následujících jevů:

- Hlasivkami vytvářený signál není vždy přesně periodický. Perioda se proto může uvnitř analyzovaného segmentu měnit, stejně jako celkový průběh signálu
- S čistě znělým či neznělým buzením se setkáváme pouze ve výjimečných případech. V jednom segmentu se proto můžou vyskytnout jak znělé, tak neznělé části
- Při nízké energii signálu se obtížněji oddělují znělé a neznělé úseky
- Při nízkém odstupu užitečného signálu od šumu může šum ovlivňovat hodnotu nalezeného tónu a v extrémních případech může dokonce znemožnit jeho nalezení

Je proto důležité nalézt metodu, která ve znělých úsecích řeči spolehlivě najde základní tón s co největší přesností. Zároveň však musí být metoda dostatečně odolná vůči degradování signálu šumem. [1]

3.1 Segmentace

Vzhledem ke své povaze je řečový signál téměř výhradně zpracováván metodami krátkodobé analýzy. Tyhle metody využívají kvazistacionární povahy signálu. To je předpoklad, že signál se v čase mění pomalu. Proto je signál rozdělen do úseků, které musí splňovat následující. Musí být dostatečně krátké, aby byla zajištěna stacionarita signálu, ale zároveň dostatečně dlouhé, aby zachytily signály s větší vlnovou délkou, tedy nízké frekvence přibližně od 50-60 Hz (rozsah základního tónu 50-400 Hz).

Oba požadavky jsou splněny pro úseky řeči dlouhé 10-30ms. Pro nahrávky s vzorkovacím kmitočtem 22050 Hz o délce 20ms bude délka segmentu 442 vzorků. Spousta metod včetně autokorelační, která pracuje v časové oblasti využívají rychlé Fourierovy transformace (FFT – Fast Fourier Transformation), která předpokládá vstupní segment o délce 2^N , kde N je přirozené číslo. Nejbližší počet vzorků splňující tenhle požadavek je 512. [5]

3.2 Metody

Existuje řada metod, které zjišťují základní tón s různou přesností, s různou odolností vůči šumu a v některých aplikacích neméně důležitými nároky na výpočetní výkon.

Metody lze rozdělit podle oblasti výpočtu na:

- Detekce v časové oblasti
- Detekce v kmitočtové oblasti
- Detekce v kepru

3.2.1 Detekce v časové oblasti – Autokorelace

Autokorelace je jedna z nejdůležitějších metod analýzy v časové oblasti. Pro diskrétní systémy je definovaná vztahem

$$R(k) = \sum_{n=-\infty}^{+\infty} s(n)s(n+k), \quad (4)$$

kterým je posuzována shoda signálu se stejným signálem posunutým o „k“ vzorků. Posloupnost $R(k)$ tedy vyjadřuje míru podobnosti v rámci jednoho signálu.

Autokorelační funkce vykazuje následující vlastnosti:

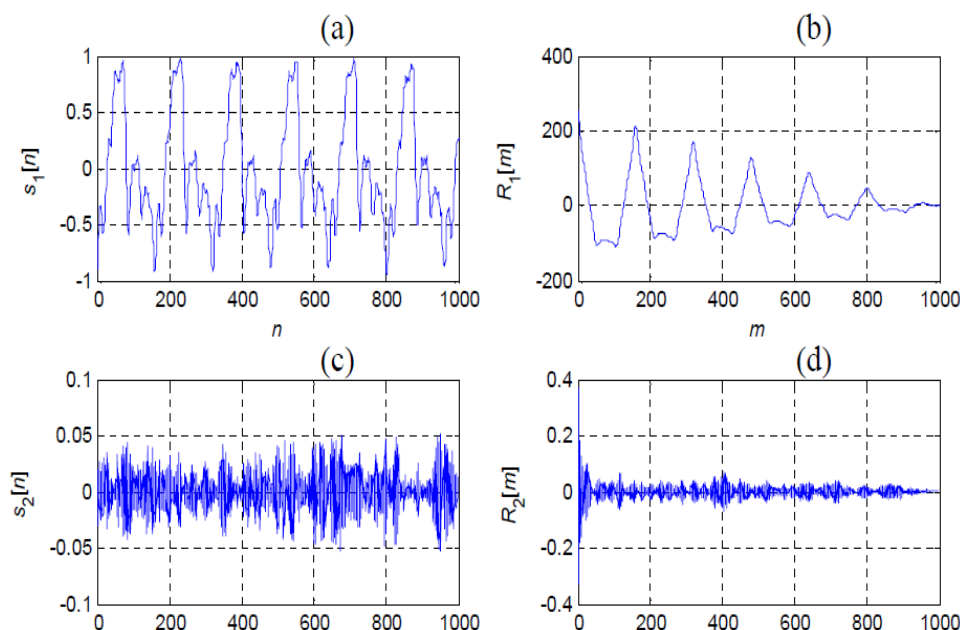
- Periodické signály jsou periodické se stejnou periodou
- Maximální hodnota je v bodě $k = 0$
- Energie signálu je rovna maximu hodnoty $R(0)$

Obrázek představuje časový průběh úseku znělé hlásky „a“ $s_1[n]$ a jeho jednostrannou autokorelační funkci $R_1[m]$ a časový průběh úseku neznělé hlásky „s“ $s_2[n]$ a jeho jednostrannou autokorelační funkci $R_2[m]$. Z obrázku je patrný rozdíl průběhů autokorelační funkce obou segmentů.

K určení základního tónu musíme najít první vrchol $R[k]$ následující po $R[0]$. Hodnotu základního tónu zjistíme ze vztahu

$$F_0 = \frac{f_{vz}}{k} \text{ [Hz]}, \quad (5)$$

kde f_{vz} je vzorkovací frekvence a „k“ je k-tý vzorek segmentu odpovídající maximu pro první vrchol $R[k]$. [5]



Obrázek 3-1: Časový průběh úseku znělé hlásky „a“ $s_1[n]$ (a) a jeho jednostrannou autokorelační funkci $R_1[m]$ (b) a časový průběh úseku neznělé hlásky „s“ $s_2[n]$ (c) a jeho jednostrannou autokorelační funkci $R_2[m]$ (d)- obrázek převzat z [5]

Existují metody, které přímo vychází z výpočtu autokorelační funkce a různě ji vylepšují. Takovým příkladem je metoda centrálního klipování a metoda inverzní filtrace.

3.2.2 Detekce v časové oblasti – centrální klipování

Princip metody vychází z poznatku, že pro stanovení kmitočtu základního tónu stačí znát pouze špičky v průběhu řeči. Střední část signálu mezi prahovými hodnotami P je odříznuta a signálu zůstanou pouze špičky přesahující prahovou hodnotu.

Velikost signálu v čase kolísá, proto nemůžou být prahové hodnoty P stanoveny pevně pro celý signál. Prahové hodnoty jsou zásadně voleny symetricky, to znamená, že $P = P_+ = -P_-$. Pro zjištění prahové hodnoty j -tého segmentu potřebujeme znát maxima sousedních segmentů MAX_{j-1} a MAX_{j+1} . Hodnota P se pak zjistí ze vztahu [1,5]

$$P = k(MAX_{j-1}, MAX_{j+1}), \quad (6)$$

Kde k je redukční faktor obvykle nabývající hodnotu $k = 0,8$.

Zbylé špičky po klipování jsou dále normovány na jednotkovou úroveň. Normovaný signál $\tilde{s}(n)$ bude nabývat pouze tří hodnot :

- $+1$ pro $s(n) > P_+$
- 0 pro $P_- \leq s(n) \leq P_+$
- -1 pro $s(n) < P_-$

Kmitočet základního tónu určíme následně z autokorelační funkce $\tilde{R}(k)$ signálu

$\tilde{s}(n)$

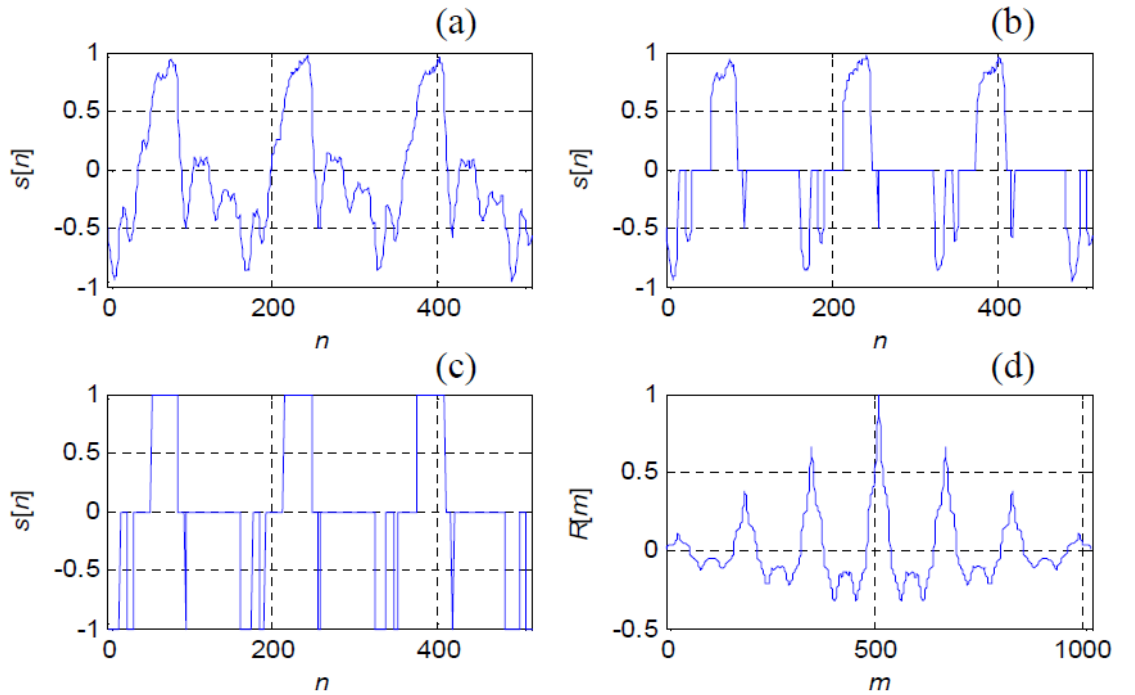
Značná výhoda spočívá v tom, že znělost či neznělost hlásek nemusíme testovat zvláštní metodou, ale vyjdeme z faktu, že znělé hlásky mají výrazně větší energii signálu než neznělé.

Za tohoto předpokladu stačí kromě $\tilde{R}(k)$ zjistit $\tilde{R}(0)$ a porovnáme

$$\tilde{R}(k)_{\max} < \alpha \tilde{R}(0) \text{ pro neznělý úsek} \quad (7)$$

$$\tilde{R}(k)_{\max} \geq \alpha \tilde{R}(0) \text{ pro znělý úsek} \quad (8)$$

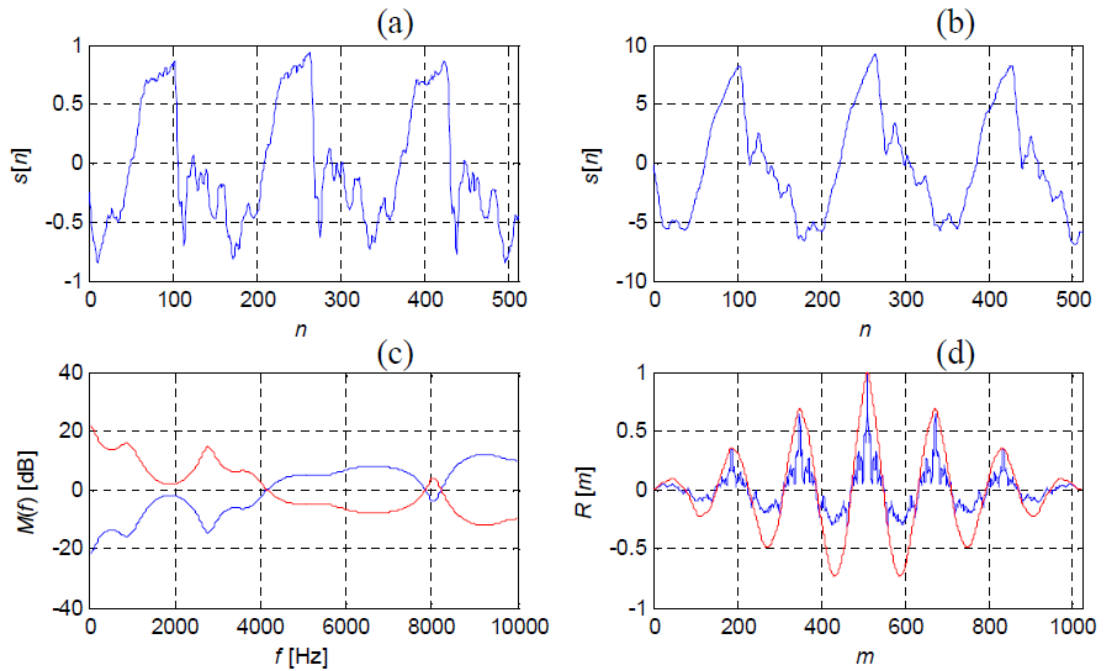
kde α je empirická konstanta s hodnotou $\alpha = 0,3$.



Obrázek 3-2: Postup výpočtu F_0 metodou centrálního klipování, vstupní segment (a), vstupní segment po prahování (b), vstupní segment po klipování (c), oboustranná autokorelační funkce klipovaného signálu (d)- obrázek převzat z [5]

3.2.3 Detekce v časové oblasti – Inverzní filtrace

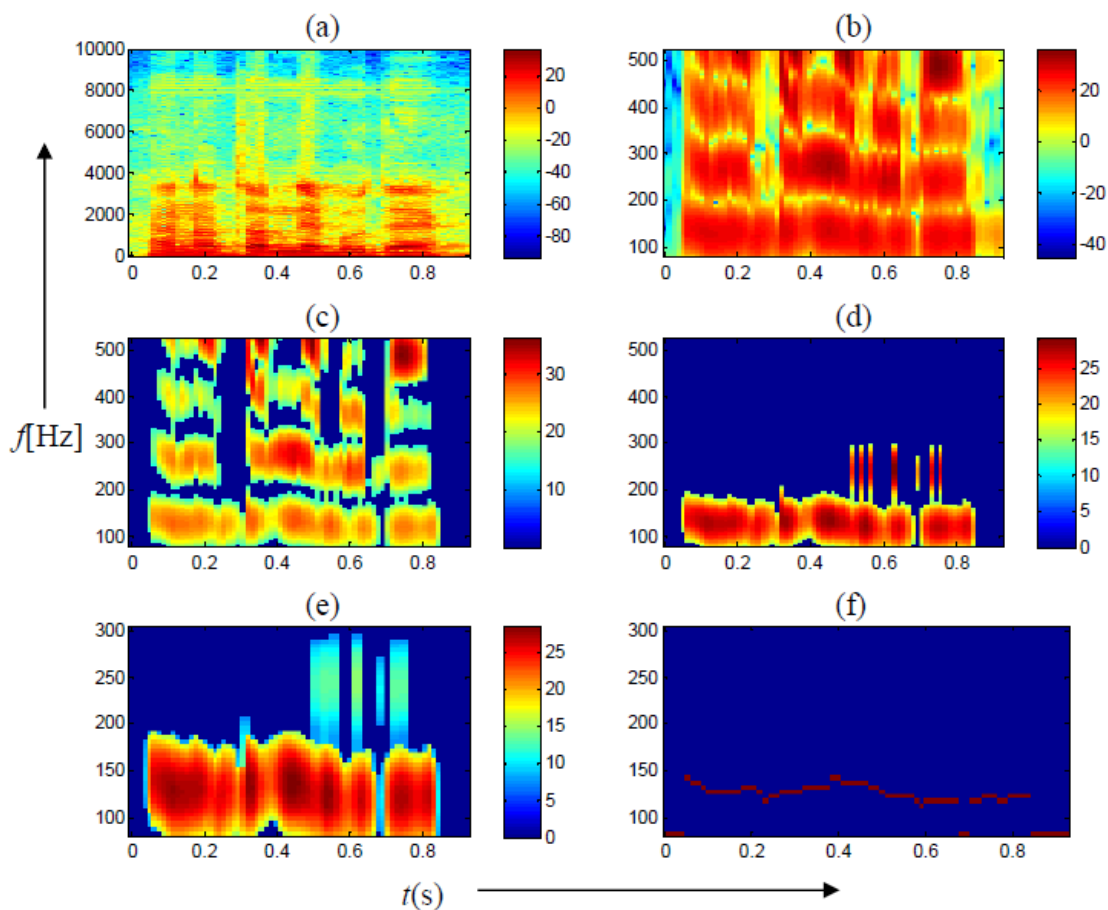
Při tvorbě řeči dochází k průchodu budícího signálu hlasovým traktem (viz. kapitola 1). Parametry hlasového traktu tedy jeho impulsní charakteristiku a tím i jeho kmitočtovou charakteristiku lze odhadnout lineární predikční analýzou označovanou jako LPC. Za předpokladu, že hlasový trakt tvoří filtr s přenosovou funkcí $A(z)$, kterou získáme transformací Z impulsní charakteristiky $a(n)$ hlasového traktu získanou LPC analýzou. Tím získáme budící signál generovaný hlasivkami, který ještě není zpracován hlasovým traktem, tím pádem není znehodnocen formanty (rezonanční frekvence hlasového ústrojí) [5]



Obrázek 3-3: Vstupní segment (a), segment řeči po filtraci inverzním filtrem (b), odhad kmitočtové charakteristiky hlasového traktu (červená) a jeho inverzní systém (modrá) (c) oboustranná autokorelační funkce půvoho (modrá) a filtrovaného segmentu řeči (červená) (d)- obrázek převzat z [5]

3.2.4 Detekce v kmitočtové oblasti – Spektrální metoda

Metoda je založena na výpočtu spektrogramu řečového signálu pomocí krátkodobé Fourierovy transformace (STFT). Při zpracování se díváme na spektrogram jako na obraz, proto se provádí zpracování samotného obrazu. [5]



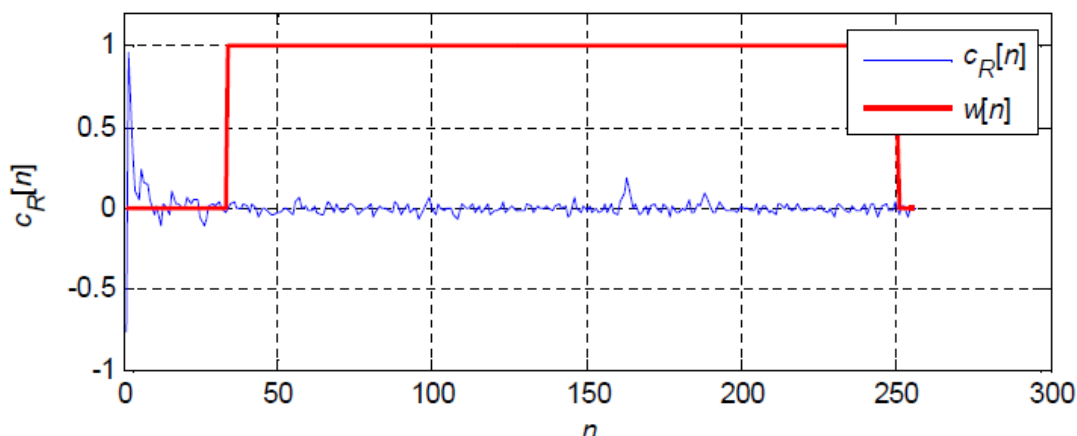
Obrázek 3-4: Spektrogram řečového signálu (a), omezení kmitočtu na hodnoty 80-500 Hz (b), prahování empirickým prahem (c), výběr pouze první harmonické (d), aplikace dvourozměrného mediánového filtru na obrázek (d), tím se odstraní nepotřebné hodnoty obrázek (e), výběr pouze maximální hodnoty pro jednotlivé časové momenty (f)- obrázek převzat z [5]

3.2.5 Detekce v kepstru – Kepstrální metoda

Tato metoda používá reálné kepstrum řečového signálu pro detekci základního tónu řeči. Reálné kepstrum $c(n)$ je definováno jako reálná část inverzní Fourierovy transformace IFFT přirozeného logaritmu modulu Fourierovy transformace FFT vstupního signálu.

$$c_R[n] = \text{Re}\{\text{IFFT}(\ln|\text{FFT}(s[n])|)\} \quad (10)$$

Kepstrum na nižších indexech n (od 1 do 16) reprezentuje vlastnosti hlasového ústrojí. Na vyšších je zachycena excitační informace. V případě znělých hlásek základní tón.



Obrázek 3-5: Kepstrum řečového signálu s omezovacím okénkem – obrázek převzat z [5]

Omezovací okno se používá k oddělení vlastností hlasového traktu od excitační informace. Omezení je provedeno k omezení oblasti hledání maxima. Vychází to z faktu, že kmitočet základního tónu řeči se pohybuje od 60 Hz do 400 Hz.

V reálném kepstru po vynásobení oknem hledáme maximum v podobě špičky.

3.3 Výběr metody

Na začátku kapitoly bylo zmíněno, že správné určení základního tónu je zcela zásadní pro správnou funkci filtrace disonančních složek signálu a to i signálu degradovaného šumem. V některých aplikacích taky může být kladen důraz na co nejmenší výpočetní náročnost.

Výňatek z [5]:

„Odhad F_0 pomocí autokorelační funkce je jednoduchá a výpočtově nenáročná metoda, její základní podobu lze vylepšit klipováním vstupního signálu (metoda centrálního klipování) která mírně zrychluje výpočet autokorelační funkce, protože počítá pouze se třemi hodnotami 1, -1 a 0, navíc částečně potlačuje vliv formantů při výpočtu, na druhé straně zvětšuje náchylnost na šum, protože může dojít k úplnému klipování daného segmentu vlivem špatného odhadu hodnoty prahu pro daný segment. Metoda inverzní filtrace též používá autokorelační funkci pro výpočet, její velkou výhodou je téměř úplné potlačení vlivu formantů, naproti tomu je výpočtově náročnější než

předchozí dvě metody a má větší náchylnost na šum. V případě malých hodnot odstupů signálu od šumu dochází k nesprávnému odhadu parametrů hlasového traktu a tím k znehodnocení celého výsledku. Kepstrální metoda je výpočetně nenáročná, navíc není náchylná na vliv formantů. Její velkou nevýhodou je velká náchylnost na šum, to je způsobeno výpočtem logaritmu spektra, který vede k zesílení menších složek představující šum. Spektrální metoda má velkou výhodu v tom, že v sobě zahrnuje proces určení charakteru daného segmentu řeči, tj. zda se jedná o znělý či neznělý úsek, navíc je velice odolná vůči šumu. Její nevýhodou je větší výpočetní náročnost než předchozí metody a má nejmenší přesnost.“

Hledisko porovnání	AKF	Cent. klip.	Inv. filtrace	kepstrální	Spektrální
Přesnost	3	1	2	4	5
Odolnost vůči šumu	2	3	4	5	1
Výpočetní náročnost	3	2	4	1	5
Největší výhoda	Jednoduchost	Přesnost	Potlačení vlivu formantů	Rychlost	Malá citlivost na šum
Největší nevýhoda	Citlivost na vliv formantů	-	Výpočetní náročnost	Velká citlivost na šum	Malá přesnost

Pozn.: Čísla 1 až 5 představují pořadí jednotlivých metod, (1-nejlepší, 5-nejhorší)

Obrázek 3-6: Shrnutí vlastností metod odhadu základního tónu řeči. Obrázek převzat z [5]

Výběr

Na základě zmíněných informací se jeví metoda centrálního klipování jako nejvhodnější. Ze všech zmíněných metod je nejpřesnější a výpočtově nenáročná a proto je vhodná i pro aplikace, kde je výpočetní náročnost významné kritérium. S výhodou může být využita autokorelační funkce, která je v metodě obsažena, k zjištění zda je segment řeči znělý či neznělý (viz. metoda centrálního klipování). Rozhodovací funkce znělosti může být eventuálně doplněna o funkci počítající počet průchodů signálu nulou ve snaze zpřesnění výsledku.

Jako nevýhodná se jeví skutečnost, že metoda má menší odolnost vůči šumu, tím pádem bude méně vhodná pro signály silně degradované šumem, ne-li úplně nevhodná.

4 NÁVRH METODY FILTRACE

- Segmentace
 - Zjištění znělosti / neznělosti segmentu
 - Výpočet základního tónu řeči
- Filtrace

4.1 Segmentace

Segmentace je úvodní fáze této metody. Za úkol má rozdělit řečový signál na úseky a to tak, aby zajistila stacionaritu zkoumaného signálu v rámci jednoho segmentu a zároveň aby zaručila bezchybné změření potřebných parametrů. Délku segmentu tedy budeme brát $t = 20\text{ms}$. [5] Tato hodnota odpovídá $n=882$ vzorkům při vzorkovací frekvenci audio signálu $F_{vz} = 44100$ Hz podle vzorce

$$t = \frac{n}{F_{vz}} \quad (11)$$

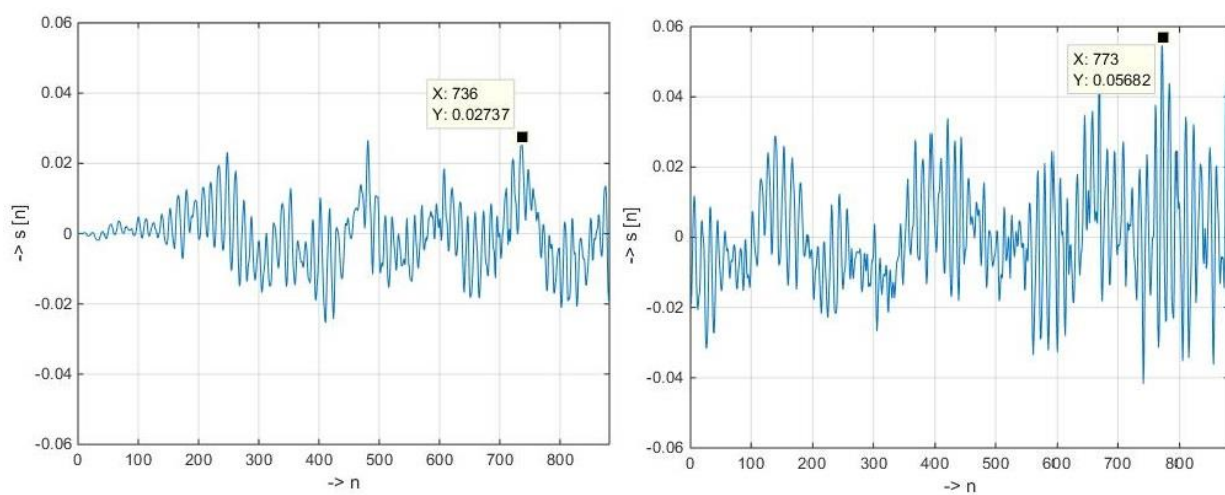
a zároveň zaručí možnost zjištění spodní hranice frekvenčního rozsahu základního tónu řeči $F_0 = 50$ Hz při korelaci podle vzorce (5).

Nejdříve se do paměti načte řečová nahrávka za pomoci funkce `audioread` a následně je segmentovaná pomocí funkce `buffer`. Jednotlivé segmenty se z poloviny překrývají.

```
[xin,fs]=audioread('vseCoJsikdysiVlastnil.wav');  
seg = 882;  
prekryti = 441;  
y_in = buffer(xin,seg,prekryti);
```

4.2 Prahování

Vzhledem k tomu, že řečový signál v čase výrazně kolísá, nemůžeme zvolit pevnou hodnotu "P" pro celý signál. Nejprve zjistíme maxima sousedních segmentů. Pro segment "i" zjistíme MAX_{i-1} a MAX_{i+1} . [1,5]



Obrázek 4-1: Segment i-1 a segment i+1

Následně si i-tý segment umístíme do počátku soustavy souřadnic. Hodnoty x-ových souřadnic pro i-1 segment budou tedy x01 a pro i+1 segment x03.

```
x01 = x01 - seg + překrytí;  
x03 = x03 + seg - překrytí;
```

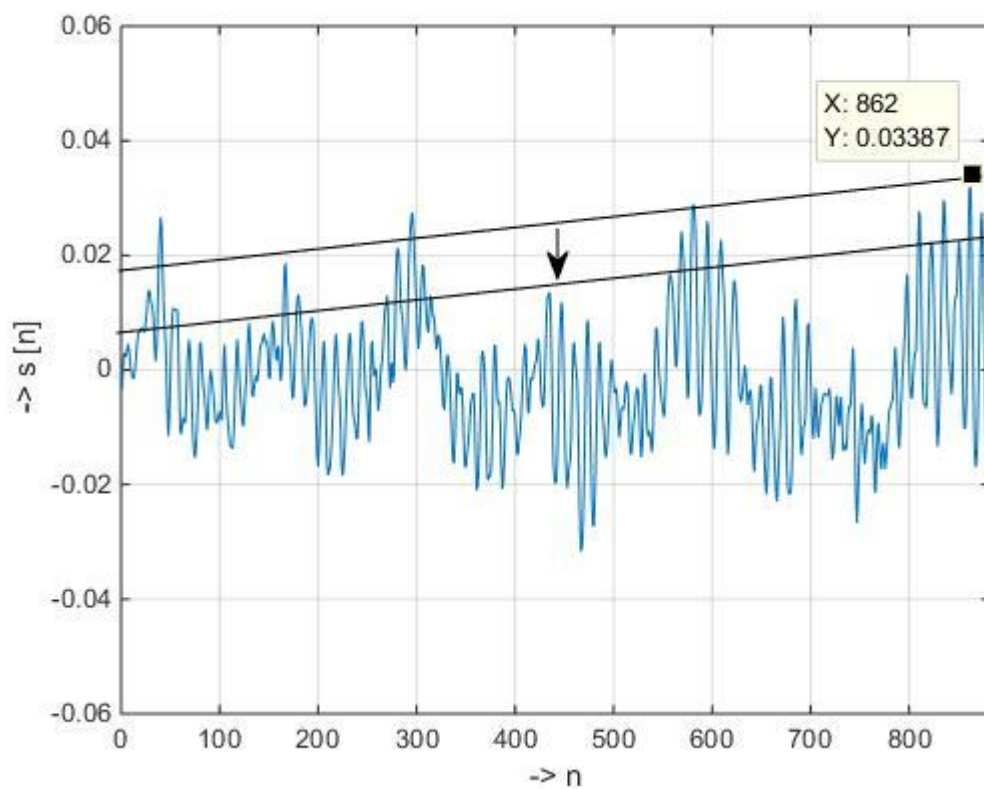
Z hodnot souřadnic MAX_{i-1} a MAX_{i+1} vypočítáme směrový vektor $\vec{u}$. Pomocí vektorového vyjádření přímky

$$\mathbf{x} = \mathbf{x}_a + t \cdot \mathbf{u}_1 \quad (12)$$

$$y = y_a + t \cdot u_2 \quad (13)$$

$$y = y_a + \left(\frac{x - x_a}{u_1} \right) \cdot u_2 \quad (14)$$

,kde $A[x_a, y_a]$ jsou souřadnice Max_i , vypočítáme přímku procházející maximem i-tého segmentu, kterou následně vynásobíme redukčním faktorem $k = 0,8$ a tím získáme hodnotu prahu P_+ pro kladné hodnoty.



Obrázek 4-2: Ukázka prahu P_+

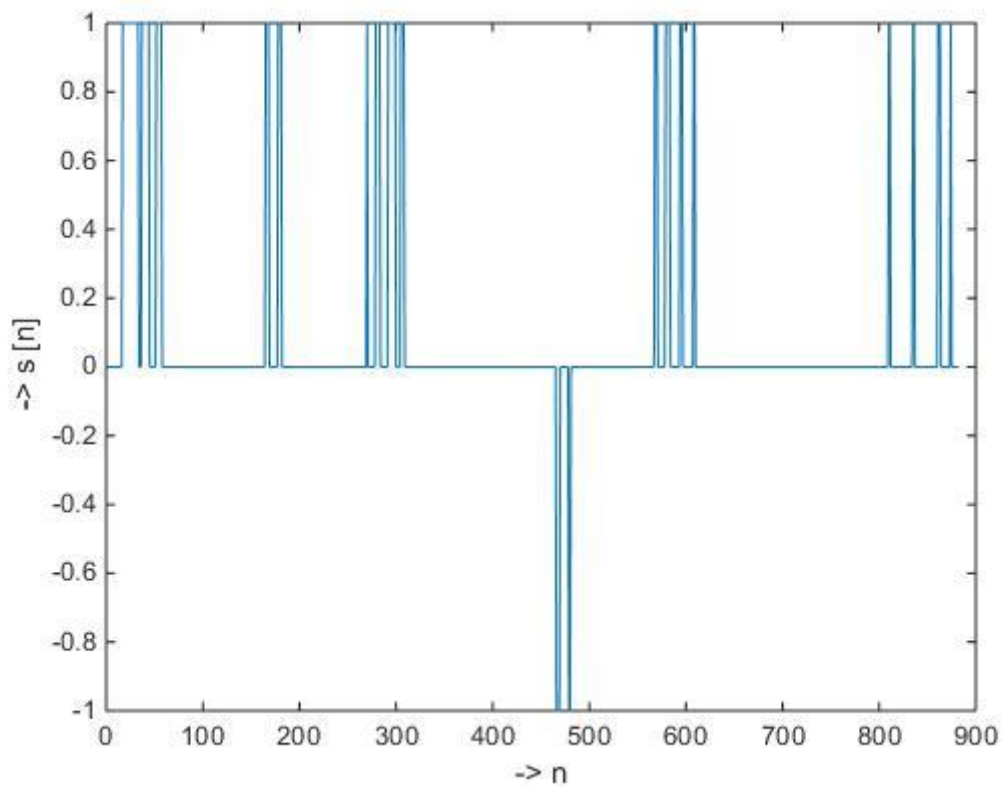
Obdobným způsobem se vypočítá práh P_- pro záporné hodnoty, akorát se místo maxim hledají minima signálu. Prahy jsou nesymetrické a to z důvodu přizpůsobení se průběhu signálu v časové oblasti.

4.3 Normalizace

Po prahování se hodnoty signálu, které přesáhly úroveň prahu “P” znormalizují na jednotkovou úroveň a hodnotám signálu, které ji nedosáhly, nebo se jí rovnají je přiřazena hodnota 0. Normovaný signál $\tilde{s}(n)$ bude nabývat tedy pouze tří hodnot :

- +1 pro $s(n) > P_+$
- 0 pro $P_- \leq s(n) \leq P_+$
- -1 pro $s(n) < P_-$

Funkce $s(n)$ je vstupní signál a “P” je hodnota prahu. [1]



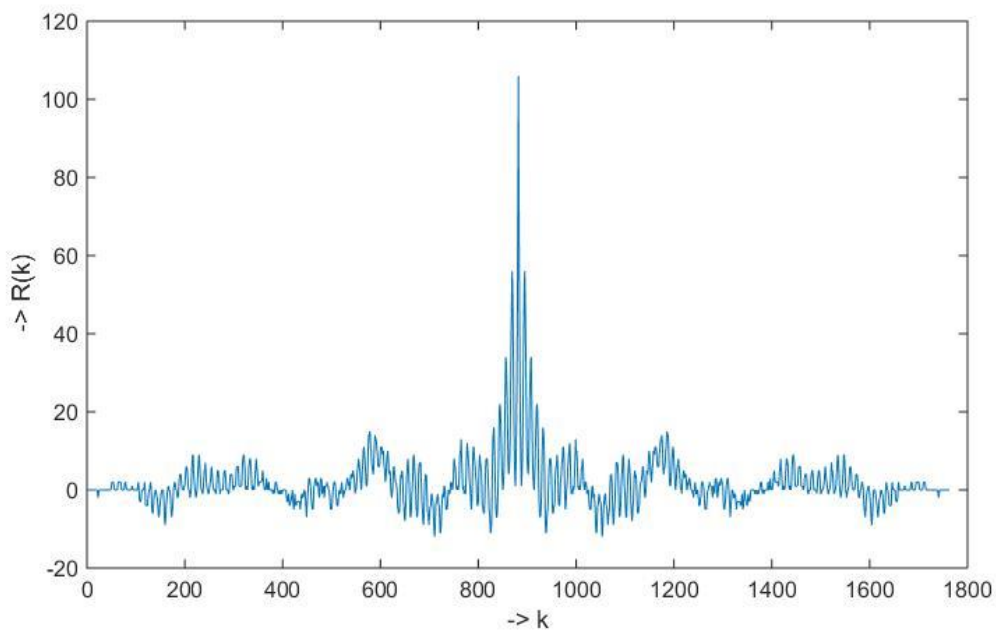
Obrázek 4-3: Průběh signálu segmentu po normování

4.4 Autokorelace

Autokorelační funkce zde neslouží pouze k zjištění základního tónu řeči, ale můžeme ji zároveň použít k posouzení, zda je daný segment znělý či neznělý a tím snížit výpočetní náročnost metody, jelikož nemusíme zařazovat další algoritmy rozhodující o znělosti.

Autokorelační funkce signálu $\tilde{s}(n)$ bude mít tvar:

$$R(k) = \sum_{n=-\infty}^{N-k} \tilde{s}(n)\tilde{s}(n+k), k = 0, 1, \dots, N-1 \quad (4)$$



Obrázek 4-4: Autokorelační funkce pro segment “i”

Z autokorelační funkce zjistíme hodnotu maxima, na základě kterého budeme následně testovat znělost segmentu.

Nyní upravíme autokorelační funkci tak, abychom byli schopni zjistit hodnotu druhého maxima.

Mějme vzorec:

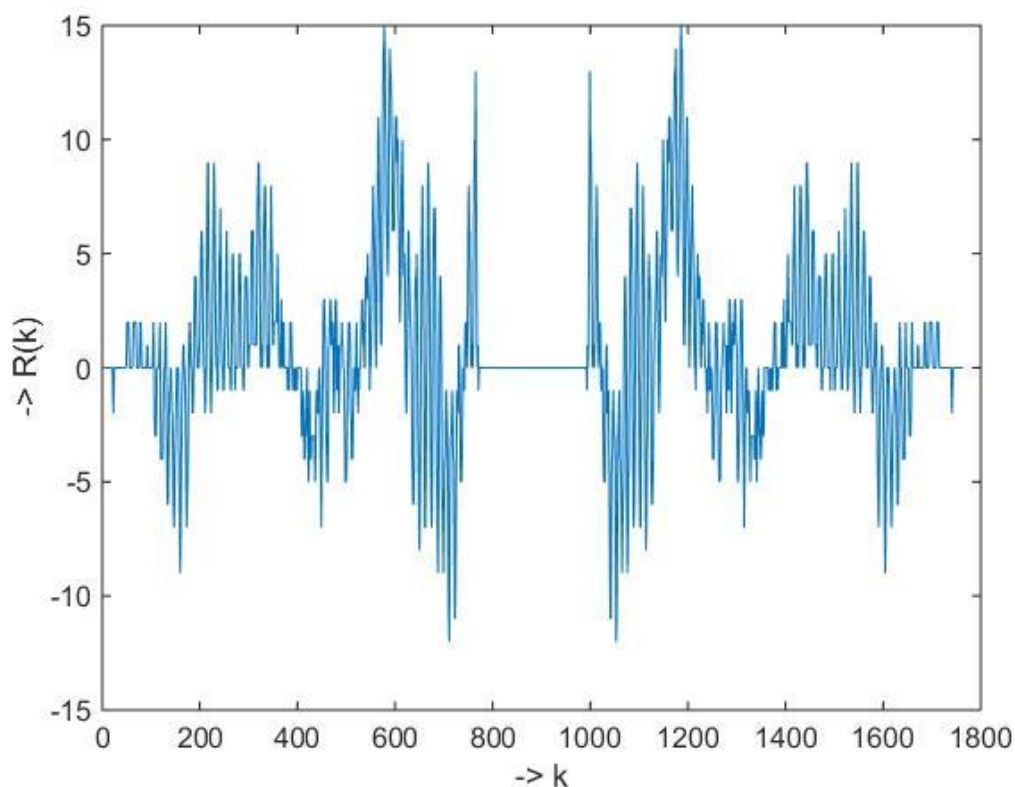
$$F_0 = \frac{f_{vz}}{k} \text{ [Hz]}, \quad (5)$$

kterým se zjišťuje základní tón F_0 , kde f_{vz} je vzorkovací kmitočet a "k" je k-tý vzorek segmentu. Jeho úpravou a dosazením krajních hodnot rozsahu (50 a 400Hz) za F_0 dostáváme tvary rovnic:

$$k = \frac{f_{vz}}{50} \quad (15)$$

$$k = \frac{f_{vz}}{400} \quad (16)$$

Dosažením vzorkovacího kmitočtu 44100Hz dostáváme hodnoty 882 a 110,25. Budeme tedy uvažovat pouze hodnoty $k = 110$ až 882 od středu grafu oboustranné autokorelační funkce.



Obrázek 4-5: Graf oboustranné autokorelační funkce pro výpočet základního tónu řeči

Po nalezení druhého maxima k -tého vzorku otestujeme znělost segmentu. Při testování znělosti je porovnávána hodnota druhého maxima $\tilde{R}(k)$ s hodnotou prvního maxima $\tilde{R}(882)$.

$$\tilde{R}(k)_{\max} < \alpha \tilde{R}(882) \text{ pro neznělý úsek} \quad (7)$$

$$\tilde{R}(k)_{\max} \geq \alpha \tilde{R}(882) \text{ pro znělý úsek} \quad (8)$$

$\tilde{R}(882)$ je hodnota energie signálu a " α " je empirická konstanta ($\alpha = 0,3$).

Pro znělé úseky se bude počítat frekvence základního tónu podle rovnice (5). [1] Pro neznělé úseky se bere hodnota F0 posledního znělého segmentu.

4.5 Filtrace

Při návrhu filtrace se vycházelo z pokusu, při kterém byl puštěn sinusový signál tónu C a k němu byl puštěn sinusový signál tónu Cis, tedy disharmonický interval (sekunda), který se budeme snažit potlačit. Pokus byl proveden ve všech oktávách rozsahu řečového signálu. Při poměru signálů 1:1 byly slyšet zřetelné zázněje. Při postupném zvyšování odstupu signálů se nejdříve začal uplatňovat jev frekvenčního maskování a to přibližně při odstupu signálů o 10-15dB, ale stále byly přítomny zřetelné zázněje. Zázněje přestaly být výrazným rušivým elementem až při odstupu signálů kolem 20dB a větším.

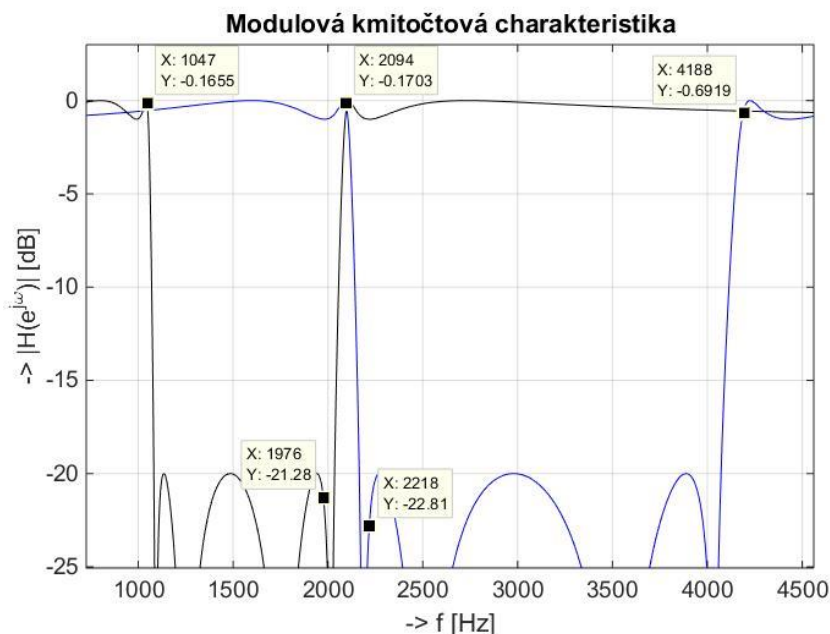
Vzhledem k tomu, že filtrování bude probíhat ve dvou větvích, které se budou následně sčítat je hodnota nepropustného pásma filtrů volena minimálně 20dB.

Při dalším pokusu se testovala spolehlivost zjišťování základního tónu řeči. Porovnávána byla testovací nahrávka řeči, ke které byl přimíchán bílý šum s poměrem signál/šum 1:1. Ukázalo se, že metoda centrálního klipování a autokorelace už není schopna bezpečně rozeznat základní tón z důvodu klipování šumu a není proto třeba zvyšovat útlum nepropustného pásma filtrů. Disharmonické frekvence budou tedy frekvenčně maskovány blízkými frekvencemi v propustném pásmu filtrů.

Jako typ filtrů byly zvoleny IIR filtry z důvodu výrazně nižších řádů, při zachování frekvenční odezvy. Nevýhoda spočívá v nelinearitě fázové charakteristiky. Lidský sluch však není citlivý na fázi signálu a jednotlivé větve filtrovaného signálu by se neměly výrazně ovlivňovat při dodržení útlumu v nepropustném pásmu alespoň 20dB. Dále je třeba při návrhu filtrů brát v úvahu jejich stabilitu.

4.5.1 1. Filtrovací větev

V 1. filtrovací větvi jsou navrženy filtry typu pásmová zadrž a to tak, aby v propustném pásmu zůstala pouze základní frekvence a její oktávy.



Obrázek 4-7: Modulová kmitočtová charakteristika filtrů 1. větve, ukázka pro 2 oktávy

K výpočtu mezních frekvencí filtrů bylo využito rovnice

$$F_d = F_0 * 2^{(n+(k/12))}, n = 0,1,...,7, k = 0,1,...,12 \quad (3)$$

kde F_0 je frekvence základního tónu, „n“ je číslo oktávy a „k“ je číslo půltónu.

K filtraci byla použita eliptická aproximace, která vykazuje zvlnění v propustném i nepropustném pásmu, což v této konkrétní aplikaci není na škodu. Využito je vysoké strmosti filtru a nízkého řádu funkce.[3] Na obrázku (4-7) je zobrazena modulová kmitočtová charakteristika filtrů pro dvě oktávy. Jsou v ní vyznačeny body, které odpovídají referenčnímu tónu C a jeho oktávám v propustném pásmu a jeho disonantním tónům Cis a H v nepropustném pásmu.

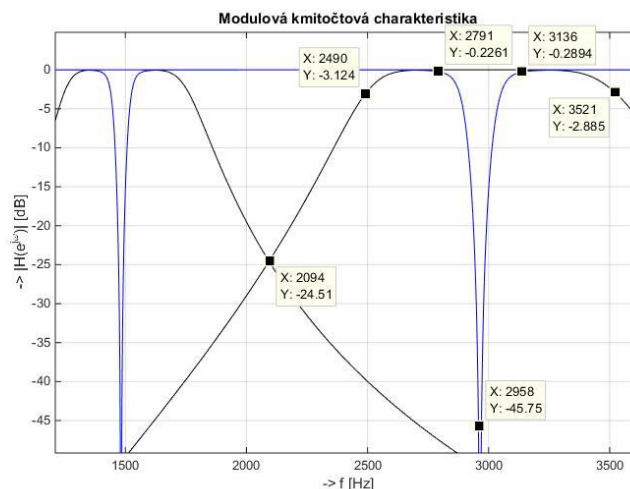
K odhadu řádu filtru byla použita funkce „ellipord“ a stabilita filtru byla testována funkcí „isstable“. Z důvodů stability filtrů byl na nižších frekvencích volen 3. řád filtru. Na vyšších frekvencích byl volen 4. řád.

```
[b, a] = ellip( 3,2,20, Wn, 'stop' );
x_filt = filtfilt(b,a,x_filt);
```

Funkce „ellip“ slouží k výpočtu koeficientů filtru eliptickou aproximací a následně je segment „x_filt“ filtrován.

4.5.2 2. Filtrovací větev

V 2. filtrovací větvi jsou použity dva druhy filtrů. Jeden druh typu pásmová propust, který k referenčnímu tónu C ponechává tóny Dis až A. Tyto tóny jsou s referenčním tónem C v konsonanci kromě tónu Fis. Tón Fis je následně odstraněn filtrem typu pásmová zadrž.



Obrázek 4-8: Modulová kmitočtová charakteristika filtrů 2. filtrovací větve, ukázka pro 2 oktávy

K výpočtu mezních frekvencí filtrů bylo využito rovnice

$$F_d = F_0 * 2^{(n+(k/12))}, n = 0,1,...,7, k = 0,1,...,12 \quad (3)$$

kde F_0 je frekvence základního tónu, „n“ je číslo oktávy a „k“ je číslo půltónu.

Uvedené filtry byly navrženy Butterworthovou aproximací, požadavek na filtry byl, aby měly nezvlněnou kmitočtovou charakteristiku v propustném pásmu a využívají monotónně klesající modulovou funkci.[3] Na obrázku (4-8) je zobrazena modulová kmitočtová charakteristika filtrů pro dvě oktávy. Vyznačené body ukazují konsonantní tóny Dis až F a G až A v propustném pásmu a disonantní tón Fis v nepropustném pásmu.

V místě křížení filtrů pásmových propustí by za podmínky souhlasné fáze signálů došlo k součtu a tedy ke zdvihu o +6dB. Hodnota v nepropustném pásmu by se teoreticky mohla dostat až na hodnotu útlumu -18,5dB, což by mohlo nepatrně ovlivnit signál 1. filtrovací větve. Filtry se totiž kříží v místě základního tónu, referenčního C.

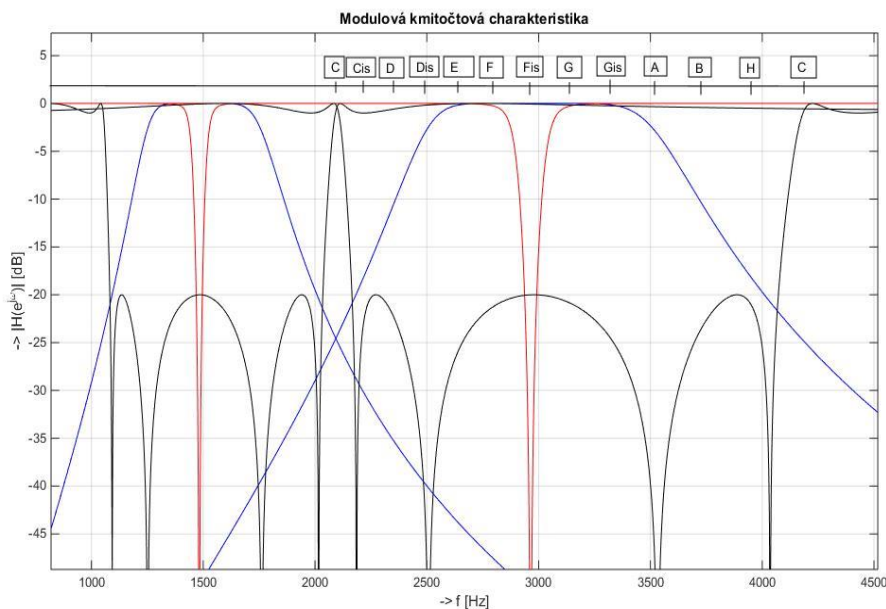
K odhadu řádu filtru byla použita funkce „butterord“ a stabilita filtru byla testována funkcí „isstable“. Z důvodů stability filtrů byl volen řád filtrů pásmová propust od třetího řádu na nižších frekvencích až po šestý řád na vyšších frekvencích. Filtr typu pásmová zadrž byl zvolen druhého řádu.

```
[b, a] = butter( 4, [fp1/(fs/2) fp2/(fs/2)], 'bandpass');  
x_filt2=x_filt2+filtfilt(b,a,x_filt1);
```

Funkce „butter“ slouží k výpočtu koeficientů filtru Butterworthovou aproximací a následně je segment „x_filt1“ filtrován.

4.5.3 Sloučení 1. a 2. filtrovací větve

Po filtraci se sečtou filtrované segmenty „x_filt“ z 1. filtrovací větve a „x_filt2“ z 2. filtrovací větve.



Obrázek 4-9: Ukázka modulové kmitočtové charakteristiky filtrů 1. a 2. filtrovací větve s rozložením tónů pro jednu oktávu

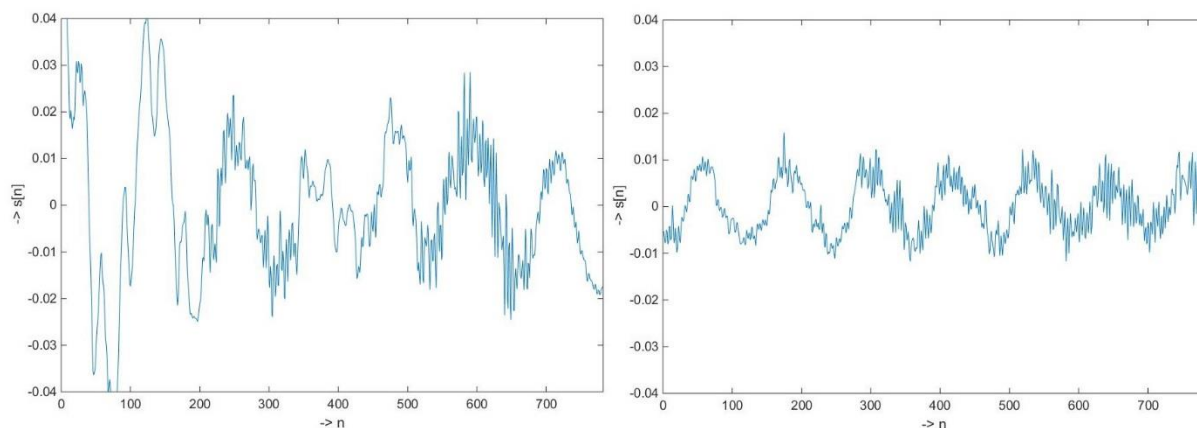
0	C	Prima	1/1	++
1	Cis	Malá sekunda	135/138	--
2	D	Velká sekunda	9/8	-
3	Dis	Malá tercie	6/5	+
4	E	Velká tercie	5/4	+
5	F	Čistá kvarta	4/3	++
6	Fis	Zvětšená kvarta	45/32	--
7	G	Čistá kvinta	3/2	++
8	Gis	Malá sexta	8/5	+
9	A	Velká sexta	5/3	+
10	B (A#)	Malá septima	9/5	-
11	H	Septima	15/8	--
12	C'	Oktáva	2/1	++

Tabulka 4-1: Přehledová tabulka pro půltóny „k“, jim přiřazen název tónu, hudební interval vztáhnutý k základnímu tónu C, interval vyjádřený zlomkem, ++ vyjadřuje úplný souzvuk, +vyjadřuje neúplný souzvuk, - neúplná disonance a -- kompletní disonance

Při návrhu filtrace byly použity IIR filtry s různou aproximační funkcí za účelem optimalizace filtrování disonančních složek signálu. Jak je vidět z grafu, konsonantní tóny jsou filtrovány s nulovou hodnotou útlumu až na Dis a A, které mají hodnotu útlumu přibližně -3dB. Naopak tóny s kompletní disonancí jsou filtrovány s útlumem téměř -20dB a více. Tóny s neúplnou disonancí jsou potlačeny o přibližně -10dB.

4.6 Sloučení segmentů

Po filtrování jednotlivých segmentů je potřeba jednotlivé segmenty sloučit do výsledného signálu. IIR filtry mají nelineární fázovou charakteristiku. Nelinearita se projeví tak, že jednotlivé segmenty budou vůči sobě posunuty ve fázi. To způsobí, že signál navazujících segmentů bez vzájemného překrytí, bude v místě navázání nespojitý.

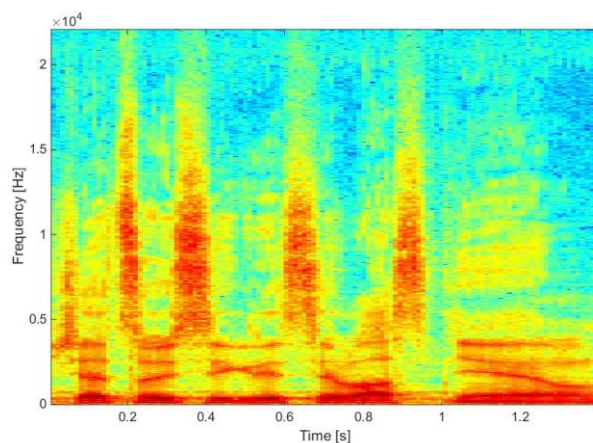


Obrázek 4-10: Ukázka nespojitosti signálu v místě navázání segmentů

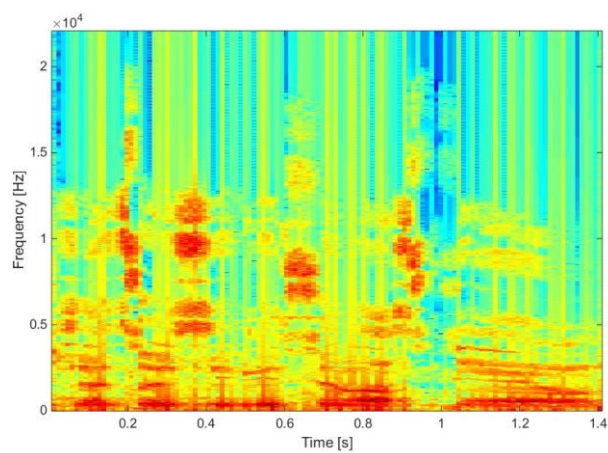
Skoková změna signálu při spojení segmentů vyvolá ve frekvenčním spektru vyšší harmonické složky, které při poslechu můžeme identifikovat jako výrazné praskání a lupání. Proto je výhodné ponechat překrytí sousedních segmentů a použít funkci, v hudebních editorech známou jako „cross-fade“. Jde o to, že v místě překrytí je navazující segment postupně zesilován, zatímco současný segment je postupně zeslabován. V této metodě byl použit lineární cross-fade, kdy hlasitost segmentů roste a klesá lineárně po dobu trvání překrytí.

V následujících spektrogramech lze vidět rozdíl.

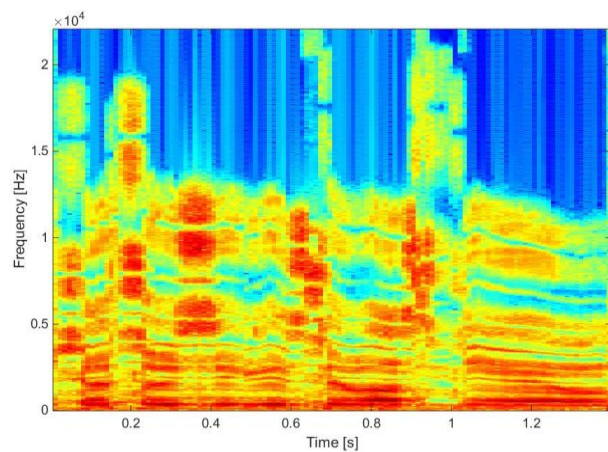
a)



b)



c)



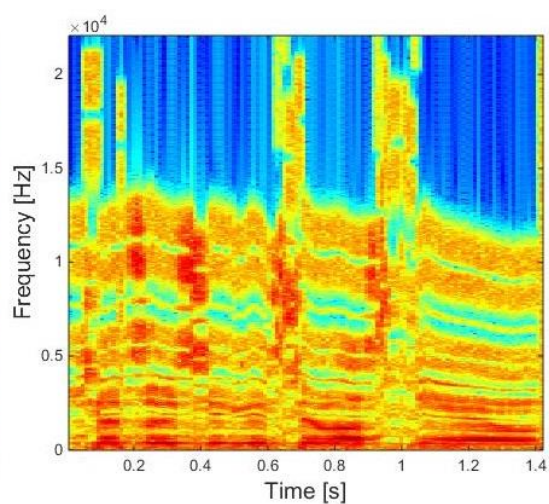
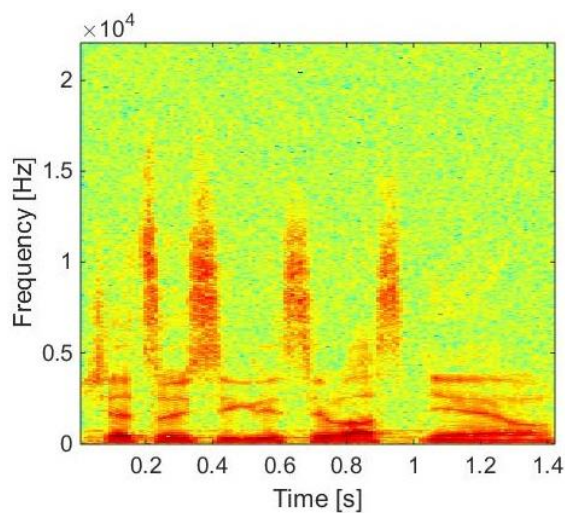
Obrázek 4-11: a) Originální řečová, nefiltrovaná nahrávka bez rušení, b) Filtrovaná nahrávka bez lineárního cross-fadeu, c) Filtrovaná nahrávka s lineárním cross-fadem

Porovnáním spektrogramů a) a b) můžeme zároveň vidět vliv adaptivní filtrace na originální signál.

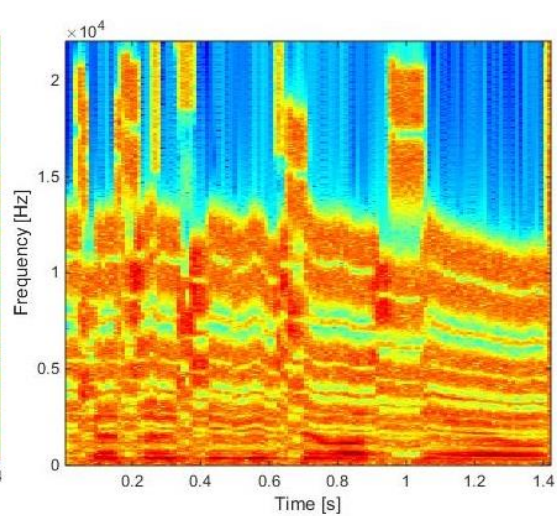
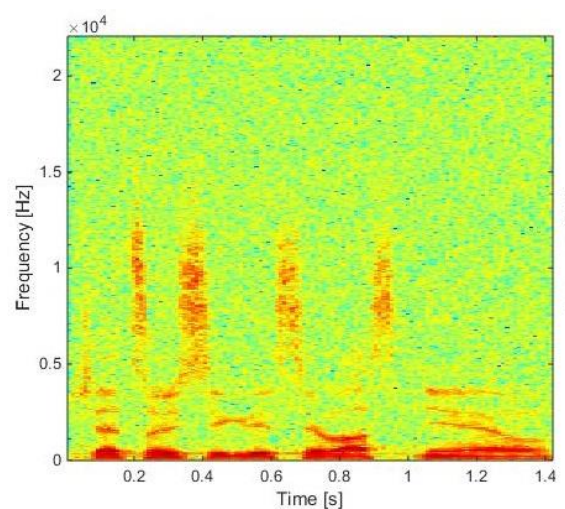
5 TESTOVÁNÍ METODY

Pro testování metody byl zvolen krátký testovací vzorek řeči, ke kterému byl následně přimíchán bílý šum. Byly zvoleny hodnoty odstupu signálu od šumu 20,10,5 a 0 dB. Rozdíl mezi degradovanou a filtrovanou nahrávkou můžeme pozorovat na následujících spektrogramech.

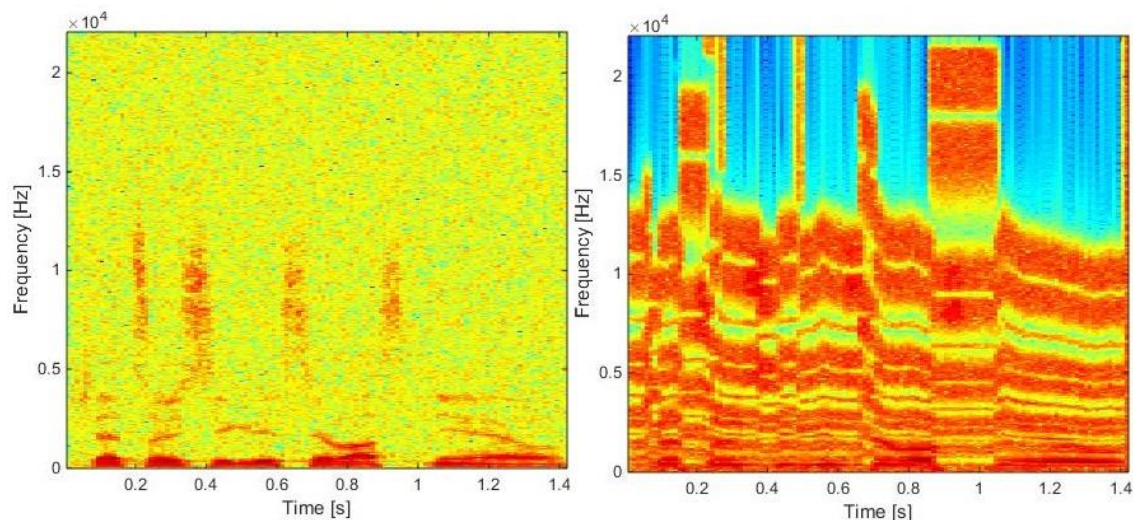
a)



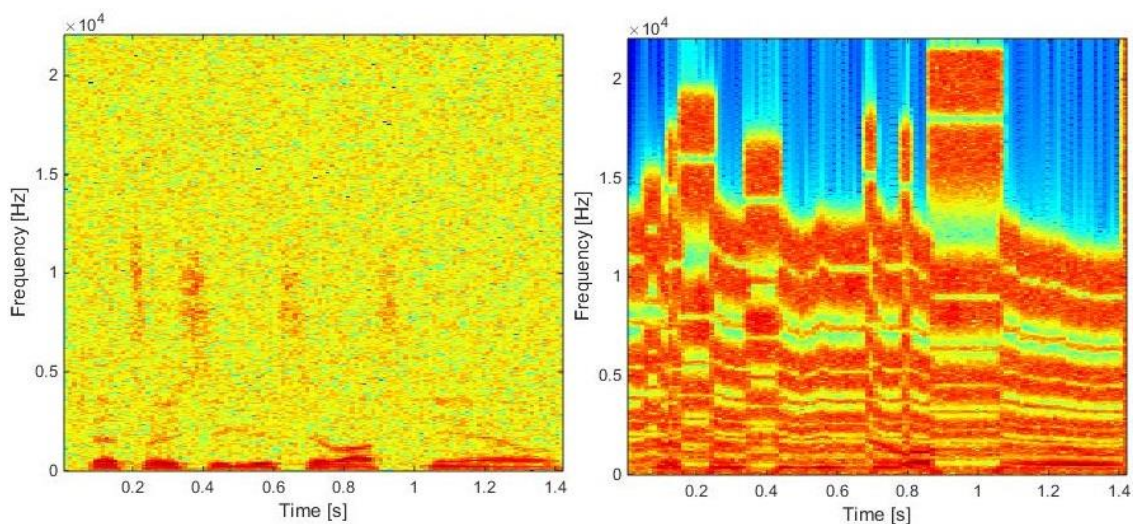
b)



c)



d)



Obrázek 5-1: a) filtrace nahrávky s odstupem signálu od šumu 20dB, b) odstup 10dB, c) odstup 5dB, d) odstup 0dB

Při odstupu signálu od šumu 20 a 10 dB jsou filtrované nahrávky srozumitelné a v průběhu základního tónu řeči jeví minimální odchylky oproti originální nahrávce. Při odstupu 5 a méně dB se začíná projevovat chybovost v určování základního tónu řeči. Neznělé hlásky, především sykavky, jejichž energie je výrazně nižší než znělých hlásek[1], začínají být maskovány šumem v pozadí a snižuje se srozumitelnost řeči. I když průběh znělých úseků řeči je podobný, jako u nahrávek s větším odstupem signálu od šumu, rozladování filtrů v kombinaci s neznělými úseky, které mají na poslech šumový charakter, působí jako rušivý element a srozumitelnost řeči se značně snižuje.

6 ZÁVĚR

Tato práce si kladla za cíl zvýšit kvalitu a srozumitelnost řeči v nahrávkách degradovaných šumem pomocí odstranění disonantních složek signálu. Návrh a testování aplikace probíhalo v prostředí matlab.

Stěžejní pro celou aplikaci bylo určit základní tón řeči. Správnost určení můžeme ovlivnit volbou délky jednotlivých segmentů signálu a redukčním faktorem „k“ při určování prahu pro klipování. Neméně důležitý je i koeficient „ α “ při testování znělosti segmentů.

Při testování metody jsme zjistili, že aplikace je schopna správně určit základní tón přibližně do -5dB odstupu signálu od šumu. Při vyšší hodnotě je šum klipován a mění průběh autokorelační funkce. Tím je celá aplikace limitovaná.

Filtrovanou nahrávku provádí jev, který by se dal označit za zpívající šum. Filtry jsou laděny základním tónem řeči, tedy jeho melodií. Tuto melodii následně slyšíme ve formě melodického šumu na pozadí nahrávky. Při vyšší úrovni šumu je tato melodie rušivým elementem a v okamžik, kdy jsou neznělé hlásky, převážně sykavky, tímto šumem maskované, stává se nahrávka nesrozumitelná.

Při návrhu byl brán ohled i na výpočetní náročnost metody. Byly voleny filtry IIR s nízkými řády funkcí (2-8) a různé aproximace pro optimální filtraci disonantních složek. Zamýšlený útlum v nepropustném pásmu filtrů byl -20dB, což by mělo stačit k účinnému odstranění disonantních složek i při odstupu signálu od šumu 0dB, za předpokladu zjištění správného základního tónu řeči. Zvolená metoda je přesto výpočetně náročná a proto zpracování nahrávky trvá až 10x déle než je její délka. Navržená metoda je tedy nepoužitelná pro aplikace v reálném čase.

Aplikace, za předpokladu obměny, by mohla najít uplatnění v hudební produkci, při mixování skladeb. Zpěv by sloužil jako zdroj základního tónu a přiměřená filtrace uplatněna na celou skladbu by mohla píseň „zharmonizovat“. Podobně by se dalo základního tónu využít pro zakomponování zpěvu do instrumentální části skladby.

LITERATURA

- [1] SIGMUND, Milan. *Analýza řečových signálů*. Vyd. 1. Brno: Fakulta elektrotechniky, 2000, 86 s. ISBN 80-214-1783-8. LÁČÍK, J., LUKEŠ, Z., ŠMÍD, P., RAIDA, Z. LESTARI, A.A., YAROVY, A.G., LIGTHART, L.P. Numerical and experimental analysis of circular-end wire bow-tie antennas over a lossy ground. *IEEE Transactions on Antennas and Propagation*. 2003, vol. 52, no. 1, p. 26–35.
- [2] SADIL, Jiří. *Zvýrazňování degradované řeči pomocí odstranění disonantních složek* [online]. 2012 [cit. 2014-12-16]. 46 l. Chip [online]. Praha: Vogel Publishing, 2010 – [cit. 6. ledna 2010]. Dostupné na www: <http://www.chip.cz/>.
- [3] SMÉKAL, Zdeněk. *Číslíkové filtry* [online]. 1. vyd. Brno: VUT, 1993 [cit. 2014-12-16]. 136 s. ISBN 80-214-0500-7.
- [4] MILIVOJEVIC, Zoran a Dragisa BALANESKOVIC. *Enhancement of the perceptive quality of the noisy speech signal by using of DFF-FBC algorithm* [online]. [cit. 2014-12-16]. ISBN 10.2298/fuee0903391m.
- [5] Ing. HichamAtassi. *Metody detekce základního tónu řeči 2008/4 – 21.1.2008 Elektro Revue* Ustav telekomunikaci, FEKT VUT, Purkyňova 118, Brno

SEZNAM ZKRATEK

DFT Discrete Fourier Transform, diskrétní fourierova transformace

FFT Fast Fourier Transformation, rychlá fourierova transformace

IFFT Inverse Fast Fourier Transformation, inverzní rychlá fourierova transformace

IIR Infinite Impulse Response, filtry s nekonečnou impulsní charakteristikou

FIR Finite Impulse Response, filtry s konečnou impulsní charakteristikou

SEZNAM PŘÍLOH NA CD

Na CD jsou kromě elektronické verze práce

- 1) Aplikace v matlabu
- 2) Řečové nahrávky použité při testech