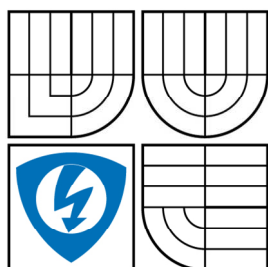


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH
TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

ODHAD ZÁKLADNÍ FREKVENCE ŘEČOVÉHO SIGNÁLU

ESTIMATION OF FUNDAMENTAL SPEECH FREQUENCY

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

TOMÁŠ RÁČEK

VEDOUcí PRÁCE
SUPERVISOR

Ing. MARTIN VONDRA, Ph.D.

LICENČNÍ SMLOUVA POSKYTOVANÁ K VÝKONU PRÁVA UŽÍT ŠKOLNÍ DÍLO

uzavřená mezi smluvními stranami:

1. Pan/paní

Jméno a příjmení: Tomáš Ráček

Bytem: Lieskové 423, 906 33 Cerová

Narozen/a (datum a místo): 6.6.1985 Myjava

(dále jen „autor“)

a

2. Vysoké učení technické v Brně

Fakulta elektrotechniky a komunikačních technologií

se sídlem Údolní 244/53, 602 00, Brno

jejímž jménem jedná na základě písemného pověření děkanem fakulty:

prof. Ing. Kamil Vrba, CSc

(dále jen „nabyvatel“)

Čl. 1

Specifikace školního díla

1. Předmětem této smlouvy je vysokoškolská kvalifikační práce (VŠKP):

- disertační práce
 - diplomová práce
 - bakalářská práce
 - jiná práce, jejíž druh je specifikován jako
- (dále jen VŠKP nebo dílo)

Název VŠKP: Odhad základní frekvence řečového signálu

Vedoucí/ školitel VŠKP: Ing. Martin Vondra, Ph. D.

Ústav: telekomunikací

VŠKP odevzdal autor nabyvateli v*:

tištěné formě – počet exemplářů 1

elektronické formě – počet exemplářů 1

* hodící se zaškrtněte

2. Autor prohlašuje, že vytvořil samostatnou vlastní tvůrčí činností dílo shora popsané a specifikované. Autor dále prohlašuje, že při zpracovávání díla se sám nedostal do rozporu s autorským zákonem a předpisy souvisejícími a že je dílo dílem původním.
3. Dílo je chráněno jako dílo dle autorského zákona v platném znění.
4. Autor potvrzuje, že listinná a elektronická verze díla je identická.

Článek 2

Udělení licenčního oprávnění

1. Autor touto smlouvou poskytuje nabyvateli oprávnění (licenci) k výkonu práva uvedené dílo nevýdělečně užít, archivovat a zpřístupnit ke studijním, výukovým a výzkumným účelům včetně pořizování výpisů, opisů a rozmnoženin.
2. Licence je poskytována celosvětově, pro celou dobu trvání autorských a majetkových práv k dílu.
3. Autor souhlasí se zveřejněním díla v databázi přístupné v mezinárodní síti
 - ihned po uzavření této smlouvy
 - 1 rok po uzavření této smlouvy
 - 3 roky po uzavření této smlouvy
 - 5 let po uzavření této smlouvy
 - 10 let po uzavření této smlouvy(z důvodu utajení v něm obsažených informací)
4. Nevýdělečné zveřejňování díla nabyvatelem v souladu s ustanovením § 47b zákona č. 111/1998 Sb., v platném znění, nevyžaduje licenci a nabyvatel je k němu povinen a oprávněn ze zákona.

Článek 3

Závěrečná ustanovení

1. Smlouva je sepsána ve třech vyhotoveních s platností originálu, přičemž po jednom vyhotovení obdrží autor a nabyvatel, další vyhotovení je vloženo do VŠKP.
2. Vztahy mezi smluvními stranami vzniklé a neupravené touto smlouvou se řídí autorským zákonem, občanským zákoníkem, vysokoškolským zákonem, zákonem o archivnictví, v platném znění a popř. dalšími právními předpisy.
3. Licenční smlouva byla uzavřena na základě svobodné a pravé vůle smluvních stran, s plným porozuměním jejímu textu i důsledkům, nikoliv v tísní a za nápadně nevýhodných podmínek.
4. Licenční smlouva nabývá platnosti a účinnosti dnem jejího podpisu oběma smluvními stranami.

V Brně dne:

.....
Nabyvatel

.....
Autor

ABSTRAKT

Bakalářská práce je zaměřena na studium algoritmů zabývajících se odhadem základní frekvence řečového signálu. První část je úvodem do problematiky řečového signálu a naznačuje co bude jádrem práce. Ve druhé části je vysvětlen charakter řečového signálu, proces jeho tvorby člověkem a modely na vytváření řeči. Kapitola č. 3 se zabývá zpracováním akustických signálů, do něž patří předzpracování, segmentace a aplikace Hammingova okénka na samotný akustický řečový signál. Další kapitola popisuje základní frekvenci řečového signálu jako fyzikální veličinu a její odvození ze základní periody. Dále vznik základní frekvence v řečových orgánech, rozsah hodnot pro různé lidi, vlastnosti, které se sebou nese a nakonec možnosti jejího využití. Kapitola č. 5 se věnuje základním metodám na určení základní frekvence řeči v časové, frekvenční a keprální oblasti. Kapitola č. 6 obsahuje popis metod, jenž se využívají v případech, kdy je řečový signál znehodnocen šumem. V následující části je popsán návrh a realizaci vybraného algoritmu. Dále jsou uvedeny výsledky jenž byly dosaženy tímto algoritmem a jsou porovnány s výsledky obyčejné metody autokorelace. Závěrečná kapitola shrnuje celou práci a je v ní pojednáno o pokračování, rozšíření, či zdokonalení algoritmu.

KLÍČOVÁ SLOVA

Řečový signál, základní frekvence řeči, číslicové zpracování signálů, banka filtrů, autokorelace

ABSTRACT

The Bachelor thesis focuses on algorithms with respect to estimation of fundamental speech frequency. First part is introduced to the questions of speech signals and the thesis at this point gives a clue what the core is going to be about. In the second part the nature of speech signal is explained, as well as the process of its creation by a person and models for speech generation. In the chapter 3 processing of acoustic signals are described, where pre-processing, segmentation and application of Hamming window on the same acoustic speech signal are included. The next chapter reports on pitch speech frequency signal as a physical magnitude and its derivation from the pitch period. Furthermore describes, fundamental frequency creation in speech organs, scale range for different people, properties that carries and finally possibilities of its usage. Chapter 5 deals with essential principles defining pitch speech frequency in time, frequency and cepstral domain. Chapter 6 contains description of principles, used in situations, where the speech signal is devalued by noise. In the next chapter author describes design and implementation of selected principle. Furthermore, author presents results that have been achieved with this specific principle and compares them to the results of ordinary autocorrelation principle. The final chapter summarises the thesis and discusses about possible further part, extension or improvement of the algorithm.

KEYWORDS

Speech signal, fundamental frequency of speech, digital processing signal, filter bank, autocorrelation

PROHLÁŠENÍ

Prohlašuji, že svoji bakalářskou práci na téma „ODHAD ZÁKLADNÍ FREKVENCE ŘEČOVÉHO SIGNÁLU“ 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 jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne

.....

podpis autora

PODĚKOVÁNÍ

Děkuji vedoucímu bakalářské práce Ing. Martinu Vondrovi, Ph. D. za velmi užitečnou metodickou pomoc a cenné rady při zpracování bakalářské práce.

V Brně dne

.....

Tomáš Ráček

Zoznam skratiek, symbolov a veličín

| | |
|-------------------------|--|
| F_0 | - základná frekvencia rečového signálu (fundamental frequency) |
| T_0 | - základná perióda (pitch period) |
| F_1, F_2, F_3, \dots | - formantové štruktúry (formanty) |
| ACF | - autokorelačná funkcia |
| PF | - pásmová filtrácia |
| $s(t)$ | - signál v časovej oblasti |
| $S(\omega)$ | - Fourierov obraz signálu $s(t)$ |
| ω | - uhlový kmitočet |
| $g(t)$ | - konvolúcia budiaceho signálu |
| $h(t)$ | - impulzná charakteristika hlasového traktu |
| $G(f)$ | - spektrum budiaceho signálu |
| $H(f)$ | - prenosová funkcia hlasového traktu |
| $S(f)$ | - spektrum rečového signálu |
| $s''(n)$ | - n-tá vzorka signálu po preemfázi |
| $s'(n)$ | - n-tá vzorka pôvodného signálu |
| λ | - penalty factor tzv. viazacia konštanta |
| $w(n)$ | - Hammingovo okienko |
| $x(n)$ | - pôvodná vzorka signálu |
| $y(n)$ | - novo získaná vzorka |
| FT | - Fourierova transformácia |
| DFT | - diskretná Fourierova transformácia |
| IDFT | - inverzná Fourierova transformácia |
| FFT | - rýchla Fourierova transformácia |
| STFT | - krátka Fourierova transformácia |
| STFTM | - modulová matica po aplikácii STFT |
| STFTP | - fázová matica po aplikácii STFT |
| L_m | - dĺžka hlasivkovej membrány |
| ZCR | - počet priechodov nulou (zero-crossing rate) |
| $R(k)$ | - ACF |
| $s(n)$ | - spojitý signál |
| k | - k-tá vzorka signálu |
| N | - dĺžka segmentu |
| HPS | - produkt harmonického spektra |
| $c_s(\tau)$ | - komplexné kepstrum signálu |
| RASTA | - metóda mapovania spektrogramu |
| SNR | - pomer signál/šum |
| $R(n)$ | - filter RASTA |
| HP | - horná prepust' |
| DP | - dolná prepust' |
| PP | - pásmová prepust' |
| $x[n]$ | - vstupný signál |
| H_0, H_1, \dots | - prenosové charakteristiky filtrov jednotlivých pásiem |
| $x_1[n], x_2[n], \dots$ | - výstupné signály jednotlivých pásiem |
| Mel(f) | - melovská frekvencia |
| IIR | - filtre s nekonečnou impulznou charakteristikou (Infinite Impulse Response) |
| f_c | - medzný/stredný kmitočet (cut-off/center frequency) |
| G | - zosilnenie priepustného/potlačenie nepriepustného pásma G (gain) |

| | |
|--|---|
| Q | - kvalita filtra (quality) |
| H(z) | - prenosová funkcia v rovine Z |
| b₀, b₁, b₂, a₁, a₂ | - koeficienty prenosovej funkcie IIR filtra |
| E_n | - Teager Energy operátor |
| f_h | - dolná frekvencia pásmovej prepusti |
| f_d | - horná frekvencia pásmovej prepusti |
| f_v | - vzorkovacia frekvencia |
| Nseg | - počet segmentov v signály |
| ACR | - autokorelácia jedného kanálu |
| SACR | - súčet autokorelácií všetkých kanálov |
| E | - energia signálu |
| P | - prahová hodnota |
| R | - referenčné hodnoty |

Obsah

| | |
|--|---------------|
| 1. ÚVOD | - 1 - |
| 2. CHARAKTER REČOVÉHO SIGNÁLU | - 2 - |
| 2.1 <i>Proces tvorby reči človekom</i> | <i>- 2 -</i> |
| 2.2 <i>Modely na vytváranie reči.....</i> | <i>- 3 -</i> |
| 2.2.1 <i>Elektronický model tvorby reči.....</i> | <i>- 4 -</i> |
| 2.2.2 <i>Akustický model hlasového traktu.....</i> | <i>- 5 -</i> |
| 3. SPRACOVANIE AKUSTICKÉHO SIGNÁLU | - 6 - |
| 3.1 <i>Základná charakteristika.....</i> | <i>- 6 -</i> |
| 3.2 <i>Predspracovanie signálu.....</i> | <i>- 6 -</i> |
| 3.2.1 <i>Analogové predspracovanie.....</i> | <i>- 6 -</i> |
| 3.2.2 <i>Digitalizácia.....</i> | <i>- 7 -</i> |
| 3.2.3 <i>Číslkové predspracovanie</i> | <i>- 7 -</i> |
| 3.3 <i>Metódy krátkodobej analýzy</i> | <i>- 8 -</i> |
| 3.3.1 <i>Spracovanie v časovej oblasti</i> | <i>- 8 -</i> |
| 3.3.2 <i>Spracovanie vo frekvenčnej oblasti.....</i> | <i>- 9 -</i> |
| 4. ZÁKLADNÁ FREKVENCIA REČOVÉHO SIGNÁLU | - 10 - |
| 4.1 <i>Vznik základnej frekvencie rečového signálu.....</i> | <i>- 10 -</i> |
| 4.2 <i>Hodnoty základnej frekvencie rečového signálu</i> | <i>- 10 -</i> |
| 4.3 <i>Vlastnosti základnej frekvencie rečového signálu.....</i> | <i>- 11 -</i> |
| 4.4 <i>Možnosti využitia základného tónu reči</i> | <i>- 12 -</i> |
| 5. METÓDY POUŽÍVANÉ NA URČENIE ZÁKLADNEJ FREKVENCIE REČOVÉHO SIGNÁLU | - 13 - |
| 5.1 <i>Metódy používané v časovej oblasti</i> | <i>- 13 -</i> |
| 5.1.1 <i>Detekcia počtu priechodov nulou (Zero-crossing rate ZCR)</i> | <i>- 14 -</i> |
| 5.1.2 <i>Detekcia počtu vrcholov (Peak rate)</i> | <i>- 15 -</i> |
| 5.1.3 <i>Autokorelačná funkcia (ACF).....</i> | <i>- 15 -</i> |
| 5.2 <i>Metódy používané vo frekvenčnej (spektrálnej) oblasti</i> | <i>- 16 -</i> |
| 5.2.1 <i>Metóda HPS (harmonic product spectrum).....</i> | <i>- 16 -</i> |
| 5.2.2 <i>Metóda maximálnej podobnosti (maximum likelihood)</i> | <i>- 17 -</i> |
| 5.3 <i>Metódy používané v kepstrálnej oblasti (Cepstrum)</i> | <i>- 18 -</i> |
| 5.2.2 <i>Kepstrálna analýza (cepstral analysis)</i> | <i>- 18 -</i> |
| 6. ROBUSTNÉ ALGORITMY NA ODHAD F_0 ODOLNÉ VOČI ŠUMU..... | - 20 - |
| 6.1 <i>Potlačenie šumu pozadia rečového signálu mapovaním spektrogramu metódou RASTA.....</i> | <i>- 20 -</i> |
| 6.1.1 <i>Filter RASTA.....</i> | <i>- 20 -</i> |
| 6.1.2 <i>Aplikácia metódy</i> | <i>- 20 -</i> |
| 6.1.3 <i>Nevýhody metódy.....</i> | <i>- 20 -</i> |
| 6.2 <i>Pásmová filtrácia</i> | <i>- 21 -</i> |
| 6.2.1 <i>Banky číslicových filtrov.....</i> | <i>- 21 -</i> |
| 6.2.1.1 <i>Bližšia špecifikácia</i> | <i>- 21 -</i> |
| 6.2.1.2 <i>Rozdelenie frekvenčného pásma.....</i> | <i>- 22 -</i> |
| 7. REALIZÁCIA ALGORITMU | - 23 - |
| 7.1 <i>Banka filtrov.....</i> | <i>- 23 -</i> |
| 7.1.1 <i>Rozloženie filtrov vo frekvenčnej oblasti</i> | <i>- 23 -</i> |
| 7.1.2 <i>Parametrické filtre.....</i> | <i>- 25 -</i> |
| 7.2 <i>Frekvenčne - časový modul</i> | <i>- 27 -</i> |
| 7.2.1 <i>Teager Energy Operator.....</i> | <i>- 27 -</i> |
| 7.2.2 <i>Pásmová filtrácia (63 Hz, 250 Hz).....</i> | <i>- 28 -</i> |
| 7.3 <i>Modul výpočtu autokorelácie pre jednotlivé kanály.....</i> | <i>- 29 -</i> |
| 7.3.1 <i>Segmentácia signálov</i> | <i>- 29 -</i> |
| 7.3.2 <i>Aplikácia autokorelácie</i> | <i>- 29 -</i> |
| 7.4 <i>Modul určenia základnej periódy T_0 a určenie znelosti/neznelosti segmentov.....</i> | <i>- 30 -</i> |
| 7.4.1 <i>Určenie základnej periódy T_0.....</i> | <i>- 30 -</i> |
| 8. DOSIAHNUTÉ VÝSLEDKY | - 33 - |
| 8.1 <i>Pomer signál / šum (SNR).....</i> | <i>- 33 -</i> |

| | |
|--|---------------|
| 8.2 Analýza chybovosti..... | - 33 - |
| 8.3 Zhodnotenie a stanovenie príčin chybovosti | - 34 - |
| 8.4 Možné vylepšenie algoritmu..... | - 34 - |
| 9. ZÁVER | - 35 - |
| ZOZNAM POUŽITEJ LITERATÚRY | - 36 - |
| ZOZNAM PRÍLOH..... | - 37 - |
| PRÍLOHA 1: HLAVNÝ PROGRAM | - 38 - |
| PRÍLOHA 2: CYKLUS NÁVRHU FILTROV PEAK | - 41 - |

Zoznam použitých obrázkov

| | |
|---|--------|
| OBR. 2.1: SCHEMATICKY ZJEDNODUŠENÝ MODEL HLASOVÉHO ÚSTROJENSTVA | - 2 - |
| OBR. 2.2: PRINCÍP TVORBY ZNELEJ HLÁSKY V ČASOVEJ A FREKVENČNEJ OBLASTI..... | - 4 - |
| OBR. 2.3: ELEKTRONICKÝ MODEL VYTVÁRANIA UMELEJ REČI | - 4 - |
| OBR. 2.4: AKUSTICKÝ MODEL HLASOVÉHO TRAKTU | - 5 - |
| OBR. 3.1: BLOKOVÝ DIAGRAM OPERÁCIÍ PREDSPRACOVANIA REČOVÝCH SIGNÁLOV | - 6 - |
| OBR. 3.2: POSTUP PARAMETRIZÁCIE REČOVÉHO SIGNÁLU..... | - 7 - |
| OBR. 4.1: ZÁVISLOSŤ MEDZI ZÁKLADNÝM KMITOČTOM REČI A DĹŽKOU HLASIVKOVEJ MEMBRÁNY..... | - 11 - |
| OBR. 4.2: ZNÁZORNENIE ZÁKLADNÉHO TÓNU REČI V ČASOVEJ (A) A KMITOČTOVEJ OBLASTI (B) | - 12 - |
| OBR. 5.1: PRÍKLAD ZOBRAZENIA REČOVÉHO SIGNÁLU V ČASE | - 13 - |
| OBR. 5.2: SIGNÁL $\sin(x)$ V ČASE | - 14 - |
| OBR. 5.3: SIGNÁL $0,9\sin(x)+0,25(20x)$ V ČASE | - 14 - |
| OBR. 5.4: PRINCÍP METÓDY HPS LOGARITMU | - 16 - |
| OBR. 5.5: PRINCÍP METÓDY MAXIMÁLNEJ PODOBNOSTI..... | - 17 - |
| OBR. 5.6: PRINCÍP SYSTÉMU NA URČENIE KEPSTRA REČOVÉHO SIGNÁLU | - 19 - |
| OBR. 5.7: LOGARITMICKÉ SPEKTRUM A K NEMU ODPOVEDAJÚCE KEPSTRUM ZNELÉHO SEGMENTU | - 19 - |
| OBR. 6.1: PRÍKLAD ROZKLADOVEJ BANKY FILTROV | - 21 - |
| OBR. 6.2: IDEÁLNA OKTÁVOVÁ BANKA FILTROV | - 22 - |
| OBR.7.1.: ZJEDNODUŠENÁ BLOKOVÁ SCHÉMA ALGORITMU NA ODHAD T_0 ZO ZAŠUMENÉHO REČOVÉHO SIGNÁLU | - 23 - |
| OBR. 7.2: MODULOVÁ FREKVENČNÁ CHARAKTERISTIKA BANKY FILTROV | - 24 - |
| OBR. 7.3: GRAF ZÁVISLOSTI MELOVSKEJ FREKVENCIE $Mel(f)$ NA FREKVENCII f | - 24 - |
| OBR. 7.4: SCHEMATICKÉ ZNÁZORNENIE FILTRA IIR 2. RÁDU POUŽÍVANÉHO V EKVALIZÉROCH | - 25 - |
| OBR. 7.5: FREKVENČNÁ ODOZVA FILTROV TYPU SHELIVING (A) DOLNEJ PREPUSTI (B) HORNEJ PREPUSTI | - 26 - |
| OBR. 7.6: FREKVENČNÁ ODOZVA FILTRU TYPU PEAK PÁSMOVEJ PREPUSTI..... | - 26 - |
| OBR. 7.7: ZOBRAZENIE NEZNELEJ HLÁSKY „s“ (MODROU) A K NEJ VYKRESLENÝ E_n (ČERVENOU) | - 28 - |
| OBR. 7.8: FREKVENČNÁ ODOZVA PÁSMOVEJ PREPUSTI 63-250 Hz | - 28 - |
| OBR. 7.9: PROCES SEGMENTÁCIE SIGNÁLU S DĹŽKOU SEGMENTU 512 A PREKRYTÍM 200 VZORKOU | - 29 - |
| OBR. 7.10: AUTOKORELÁCIA ZNELÉHO SEGMENTU DĹŽKY 512 VZORKOU | - 30 - |
| OBR. 8.1: ZNELÝ ÚSEK REČI ZAŠUMENÝ BIELYM ŠUMOM S HODNOTU SNR A.) 30 dB B.) 40 dB | - 33 - |

Zoznam použitých tabuliek

| | |
|---|--------|
| TAB. 7.1: KOEFICIENTY ČÍSLICOVÝCH FILTROV SHELIVING 2. RÁDU | - 26 - |
| TAB. 7.2: KOEFICIENTY ČÍSLICOVÝCH FILTROV PEAK 2. RÁDU | - 27 - |
| TAB. 8.1: HODNOTY CHYBOVOSTI ZÁKLADNEJ METÓDY ACF A METÓDY PF PRE SNR = 30 dB | - 33 - |
| TAB. 8.2: HODNOTY CHYBOVOSTI ZÁKLADNEJ METÓDY ACF A METÓDY PF PRE SNR = 40 dB | - 34 - |

1. Úvod

Rečový signál je akustický signál, ktorý človek používa na dorozumievanie medzi sebou. Je to súvislý, časovo premenný proces. Má svoje špecifické vlastnosti a vždy obsahuje určitú informáciu, ktorú je možné formulovať písmom. Rečový signál nám slúži tiež ako prostriedok na získanie informácií o rečníkovi. Môžeme sa z neho dozvedieť pohlavie, hrubý odhad veku, jazyk a tiež aj náladu, alebo zdravie človeka. V praxi sa využíva schopnosť rozpoznať podľa hlasu známu osobu. Aplikácia tejto schopnosti je v zabezpečovacích systémoch ale aj v hlasovom vytáčaní, ktorým disponuje väčšina novších mobilných telefónov. Na druhej strane je reč veľmi variabilná. Nikto nie je schopný vysloviť jedno slovo dva krát úplne presne. Myslí sa tým dodržanie toho istého prízvuku, výšky tónu, hlasitosti a rýchlosti hovorenia. Z toho vyplýva že ovládanie zložitých strojov so zvýšeným ohľadom na bezpečnosť hlasom, neprichádza do úvahy.

Detailnejší pohľad na rečový signál, jeho tvorbu a spracovanie nájdeme v [1] a [2].

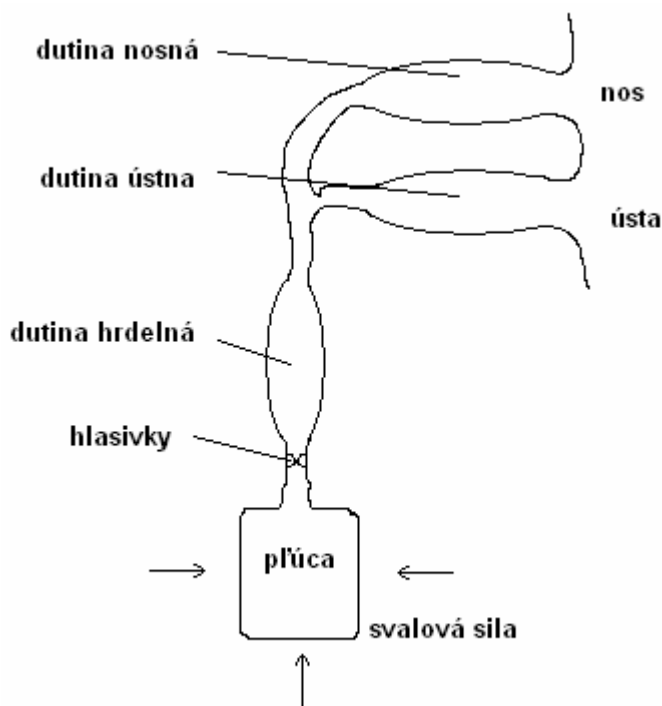
Hlavnou úlohou je navrhnúť a naprogramovať robustný algoritmus pre odhad základnej frekvencie rečového signálu, ktorý je poškodený hlukom a šumom pozadia. Malo by sa sústrediť hlavne na metódy pracujúce v spektrálnej (frekvenčnej) oblasti, ktoré rozdeľujú signál do viacerých frekvenčných pásiem a základnú periódu vyšetrujú v každom pásme zvlášť. Prvá časť (kapitola 2) sa zaoberá zjednodušeným modelom vytvárania reči a pre lepšie pochopenie sa prevádza do matematických funkcií. Kapitola 3 sa zaoberá spracovaním akustických signálov, do ktorého patrí pedspracovanie, segmentácia a aplikácia Hammingova okienka na samotný akustický rečový signál. V ďalšia časť (kapitola 4) je sústredená na vysvetlenie vlastného pojmu „základná frekvencia rečového signálu“, kde je rozobrané od čoho závisí, jej základné vlastnosti a ozrejmuje sa spojitosť na iné veličiny. Ďalšia časť bakalárskej práce (kapitola 5) je venovaná oboznámeniu so základnými metódami na odhad základnej frekvencie rečového signálu, uvádzajú sa rôzne oblasti ich použitia a vysvetľuje sa ich základný princíp. Nasledujúca časť (kapitola 6) je venovaná popisom robustných algoritmov odolných voči šumu pracujúcich vo frekvenčnej oblasti, konkrétne metóde RASTA (mapovanie spektrogramu) a pásmovej filtrácii PF, metóde využívajúcej rozklad signálu do viacerých pásiem pomocou banky filtrov . Samotný návrh a realizácia, čiže programovanie v prostredí MATLAB je popísané v kapitole 7, ako aj dôkladný popis jednotlivých blokov algoritmu. Predposledná časť (kapitola 8) sa zaoberá konfrontáciou dosiahnutých výsledkov môjho algoritmu pre rôzne pomery signál/šum, sú uvedené výsledky z analýzy chybovosti a porovnanie so základnou metódou odhadu základnej frekvencie pomocou autokorelácie. Závere v krátkosti rekapituluje dosiahnuté výsledky, obsah bakalárskej práce a analyzuje perspektívu spracovania rečových signálov v budúcnosti.

2. Charakter rečového signálu

2.1 Proces tvorby reči človekom

Tvorba reči v ľudskom hlasovom trakte je veľmi komplexný proces, ktorý vyžaduje presnú synchronizáciu mnohých svalov od pobrušnice a medzi rebrových svalov, ktoré regulujú dýchanie, po svaly v hlasivkách, v hrtane, v jazyku, v perách, svaly pohybujúce čeľusťami a ďalších.

Po fyzikálnej stránke ide o vzduchové kmity vytvárané ľudskými rečovými orgánmi. Princíp tvorby reči je veľmi podobný ako vytváranie tónov dychovými hudobnými nástrojmi. Fundamentálnym zdrojom tohto procesu sú pľúca a s nimi späté dýchacie svaly. Prúd vzduchu z tohto zdroja ďalej prechádza hlasivkovou štrbinou, ktorú obopínajú samotné hlasivky. Po ich rozkmitaní sa mení vzduch hnaný z pľúc na pravidelný budiaci signál. Takýto signál postupuje do nasledujúceho hlasového traktu čím je dutina hrdelná a potom dutina ústna alebo dutina nosná. Konečný rečový signál je v poslednom kroku formovaný jazykom v ústnej dutine a perami. Hrtan s hlasivkami vytvárajú buď periodický signál pre znelé hlásky alebo šum pre neznelé hlásky. To zapríčiňuje že, generovaný budiaci signál má už pri vstupe do hlasového traktu vždy iný spektrálny charakter. Schematicky zjednodušený fyziologický model hlavných častí hlasového ústrojenstva podieľajúceho sa na vytváraní reči je uvedený na (obr.2.1).



Obr. 2.1: Schematicky zjednodušený model hlasového ústrojenstva

2.2 Modely na vytváranie reči

Vytváranie umelej reči (syntéza reči) je historicky najstarší a dnes najviac zvládnutý proces v oblasti spracovania rečových signálov. Aby bolo možné technicky zvládnuť postup pri vytváraní reči, museli sa nájsť určité zjednodušené modely, v ktorých by ľudské rečové orgány boli nahradené technickými prostriedkami. Postupom času vznikli samostatné modely pre rečový vstup, čo je rozpoznávanie (analýza) a taktiež modely pre rečový výstup (syntéza).

Zjednodušene môžeme povedať, že celkový vplyv hlasového traktu na budiaci signál, je vlastne premenný akustický filter. Tento filter podľa vyslovovanej hlásky niektoré frekvencie tlmí a iné zosilňuje.

Pre ďalšie pokračovanie sa objasní pojem Fourierova transformácia (FT) (2.1). Fourierova transformácia je vyjadrenie časovo závislého signálu pomocou harmonických signálov, t.j. funkcií *sin* a *cos*, teda funkcie komplexnej exponenciály. Služi pre prevod signálu z časovej oblasti do oblasti frekvenčnej (spektrálnej). Matematicky predstavuje spektrálna funkcia $S(\omega)$ Fourierov obraz signálu $s(t)$, kde ω je uhlový kmitočet. Spätná Fourierova transformácia (2.2) je inverzné vyjadrenie FT, ktoré nám umožňuje, na základe spektrálnej funkcie $S(\omega)$ signálu $s(t)$, stanoviť časový priebeh signálu.

$$S(\omega) = \int_{-\infty}^{\infty} s(t) e^{-j\omega t} dt \quad (2.1)$$

[3]

$$s(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) e^{j\omega t} d\omega \quad (2.2)$$

[3]

V prípade, že by malo rečové ústrojenstvo konštantné parametre, bude generovaný rečový signál $s(t)$ v časovej oblasti daný konvolúciou budiaceho signálu $g(t)$, ktorý obsahuje **základný kmitočet** F_0 a impulznej charakteristiky hlasového traktu $h(t)$, ktorá je určená inverznou Fourierovou transformáciou prenosovej funkcie

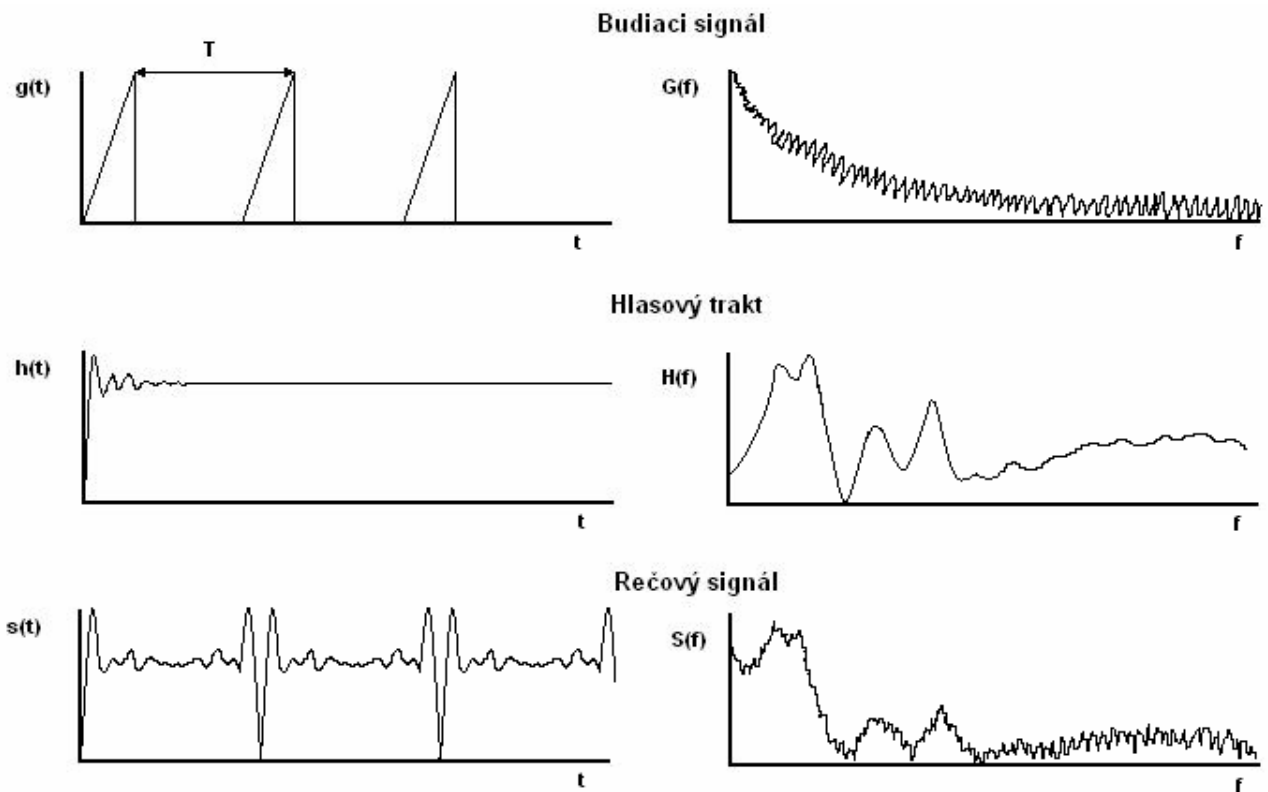
$$s(t) = g(t) * h(t) = \int_{-\infty}^{\infty} g(t) h(t - \tau) d\tau. \quad (2.3)$$

[1]

Po Fourierovej transformácii rečového signálu sa matematická operácia konvolúcie v časovej oblasti premení na operáciu súčinu vo frekvenčnej oblasti. Výsledkom súčinu spektra budiaceho signálu $G(f)$ a prenosovej funkcie $H(f)$ je potom spektrum rečového signálu $S(f)$

$$S(f) = G(f)H(f). \quad (2.4)$$

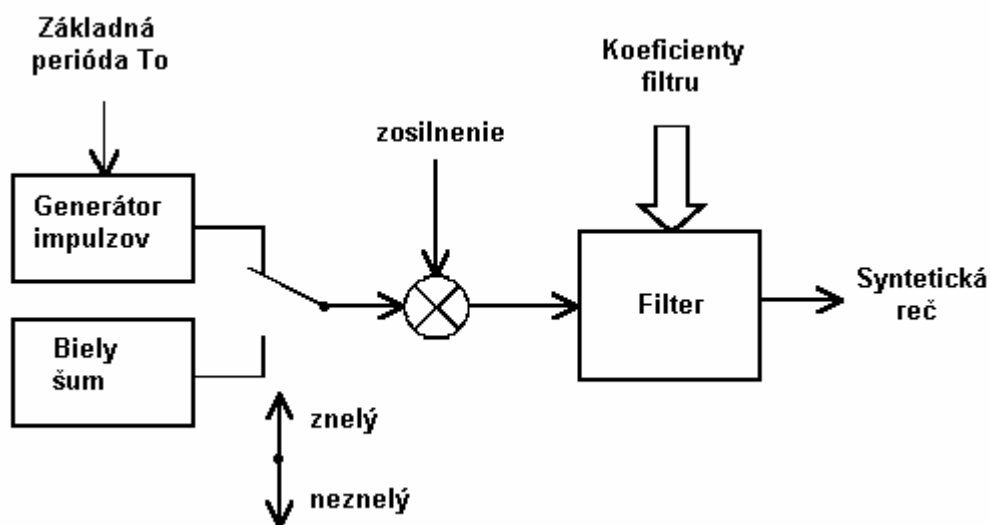
[1]



Obr. 2.2: Princíp tvorby znej hľásky v časovej a frekvenčnej oblasti

2.2.1 Elektronický model tvorby reči

Ak uvážime všetky doteraz spomenuté teoretické fakty a skutočnosti, môžeme model tvorby reči v elektronickej oblasti vytvoriť zo zdroja budenia a filtra. Vplyv vyžarovaných zvukov je možné modelovať zosilňovačom a samozrejme tiež zmenou parametrov filtra.



Obr. 2.3: Elektronický model vytvárania umelej reči

Tento uvedený model využíva iba dva typy budenia. Budiaci signál znelých hlások je možné si predstaviť ako periodický signál s pílovitým priebehom, (viz obr. 2.2), ktorý sme schopní vygenerovať impulzným generátorom. Neznelé hlásky majú šumový charakter. Ich energia je približne, ako u bieleho šumu, rovnomerne rozložená cez celé kmitočtové pásmo. Časový signál ma úplne nepravidelný priebeh bez akejkoľvek zjavnej periodicity. Budenie neznelých hlások je teda modelované šumovým generátorom.

Pre modelovanie kvalitnej reči nie je postačujúce jednoduché rozdelenie na znelé a neznelé hlásky. Prepínač budenia je potrebné nahradiť zmiešavacím potenciometrom a podiel oboch budení variabilne meniť podľa požiadavky kvality reči.

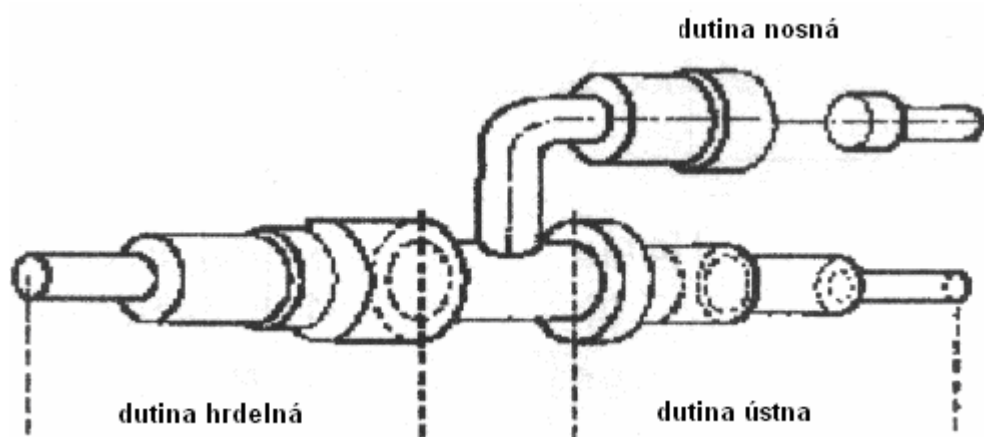
Celý systém hlasového traktu je považovaný za lineárny prenosový systém a pre jednoduchosť modelovaný ako jeden filter. Tento filter môže byť realizovaný niekoľkými rôznymi spôsobmi:

- kombináciou pásmových prepustí
- obvodmi so spínacími kapacitormi
- číslicovými filtermi s riadenými koeficientmi

2.2.2 Akustický model hlasového traktu

Existuje ešte tzv. akustický model. Ten vychádza z predstavy dutinového rezonátoru. Prenosovú funkciu hlasového traktu je možné vypočítať, len ak poznáme jeho tvar. Získaná funkcia potom odpovedá určitej fixnej polohe hlasového traktu, teda tvorbe určitej hlásky. Ak sa podarí namodelovať hlasový trakt pre rôzne hlásky, dostávame súbor deliacich modelov, ktoré predstavujú celkový akustický model vytvárania reči.

Aj keď vnútorný priestor hlasového traktu nie je súmerný vzhľadom k pozdĺžnej osi, môžeme si dovoliť bez väčšej ujmy na presnosti vytvoriť dutinový model v tvare sústredených valcov a s týmito tvarmi potom počítať v oblasti akustiky. Úplný valcový model hlasového traktu bez ohľadu na mierku je znázornený na (obr. 2.4).



Obr. 2.4: Akustický model hlasového traktu

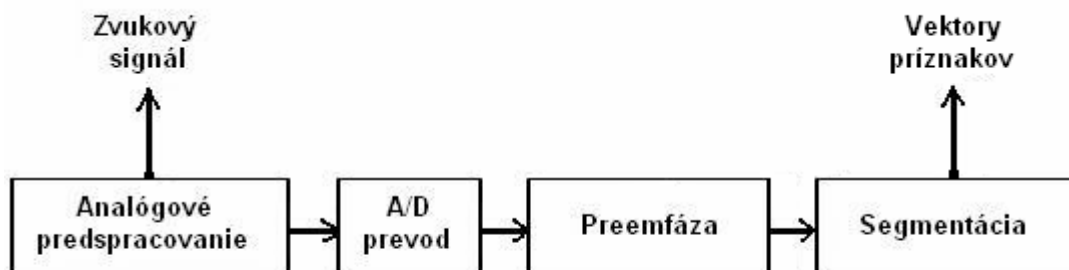
3. Spracovanie akustického signálu

3.1 Základná charakteristika

Základný predpoklad pre spracovanie rečového signálu je argument, ktorý hovorí, že vlastnosti rečového signálu sa v priebehu času menia pomaly. Z toho vyplýva možnosť spracovania jednotlivých krátkych úsekov (segmentov) dĺžky 10-35 ms (aplikácia metódy krátkodobej analýzy). Na týchto krátkych úsekoch sa rečový signál relatívne nemení a môžeme ho klasifikovať. Výsledkom krátkodobej analýzy sú vektory príznakov, popisujúce každý úsek reči. Väčšinou nosia informáciu o energetických a kmitočtových zmenách akustického signálu v čase a sú používané k parametrizácii. Hlavnou úlohou parametrizácie rečového signálu je snaha o odstránenie nadbytočnej (redundantnej) informácie, ktorá je vo veľkej miere zastúpená v rečovom signály. Parametrizácia sa robí tak, aby sa získali príznaky, ktoré budú čo najvernejšie reprezentovať originálny signál. Po príznakoch sa na druhú stranu požaduje, aby parametrizovaný signál bol pokiaľ možno čo najmenší. Samotná parametrizácia signálu je prvým krokom v celom reťazci krokov, ktoré je nutné urobiť pre zistenie základnej frekvencie.

3.2 Predspracovanie signálu

Predspracovanie signálu sa klasifikuje pred metódu krátkodobej analýzy. Na (obr.3.1) sú zobrazené všetky etapy predspracovania.



Obr. 3.1: Blokový diagram operácií predspracovania rečových signálov

3.2.1 Analógové predspracovanie

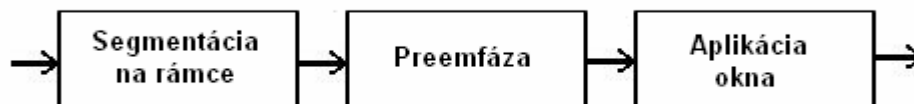
Analógové predspracovanie, t.j. manipulácia s rečovým signálom do tej doby, až bude prezentovaný sledom vzorkov, začína prevodom zmien akustického tlaku na elektrický signál. Takto získaný signál je v rozsahu niekoľkých mV a musí byť zosilnený pokiaľ možno bez šumu a s lineárnou kmitočtovou závislosťou v pracovnom pásme nasledujúcich stupňov spracovania. Zosilňovacia jednotka je realizovaná dvomi stupňami: predzosilňovač v priamej blízkosti mikrofónu a ďalším stupňom pred A/D prevodníkom. Analógový rečový signál je nutné obmedziť dolnou prepustou s ohľadom na následné vzorkovanie.

3.2.2 Digitalizácia

Pre spracovanie akustického signálu je potrebné ho previesť z analógového na číslicový tvar. Uvedený prevod sa nazýva digitalizácia a jej výsledkom je vyjadrenie spojitých analógových kmitov signálu postupnosť čísel, ktorá nadobúda hodnôt zo zvoleného konečného intervalu. Toho sa dosahuje vzorkovaním (8–22kHz) a kvantizáciou s rozlíšením 8-16 bitov. Digitalizácia sa v praxi implementuje analógovo-digitálnym prevodníkom (A/D).

3.2.3 Číslicové predspracovanie

Číslicové spracovanie sa väčšinou vykonáva podľa nasledujúceho postupu:



Obr. 3.2: Postup parametrizácie rečového signálu

- **Segmentácia na rámce:** Pri parametrizácii je signál väčšinou popísaný v inej než časovej oblasti. Obvykle sa využívajú charakteristiky signálu vychádzajúce z frekvenčnej oblasti. Vo frekvenčnej oblasti je možné signál pri rovnakej kvalite popísať menším počtom zložiek a ďalej sa tiež snažíme o odstránenie redundantnej (nadbytočnej) informácie. Vo frekvenčnej oblasti môžeme predpokladať, že parametre rečového signálu ostávajú v priebehu niekoľkých desiatok milisekúnd takmer konštantné (čo je spôsobené obmedzenou rýchlosťou premeny hlasového traktu človeka). Vzhľadom k veľkej dynamike rečového signálu sa tieto úseky (segmenty) volia v rozmedzí od 10 do 35 ms. Rámce sa väčšinou prekrývajú.
- **Preemfáza:** Preemfáza je metóda, ktorá slúži k zvýrazneniu vyšších kmitočtov v signály. Podstatná časť celkovej energie rečového signálu leží v kmitočtovom pásme pod hranicou 300 Hz, nič menej užitočné informácie sú obsiahnuté nad touto hranicou. Preto sa využíva preemfáza, ktorá sa realizuje pomocou jednoduchého číslicového filtra

$$s''(n) = s'(n) - \lambda \cdot s'(n-1), \quad (3.1)$$

[1]

kde $s''(n)$ je n-tá vzorka signálu po preemfázi a $s'(n)$ je n-tá vzorka pôvodného signálu. Viazacia konštanta λ sa bežne volí v rozsahu (0,90 – 0,98).

- **Aplikácia okna:** Ak sa rozdelí signál na segmenty, vzniká na ich okrajoch náhly prechod (orezanie), čo je nežiaduci jav pri ďalšom spracovaní (frekvenčná analýza). Odstránením nespojivosti na začiatku a na konci segmentov, sa dosiahne vyhladenie priebehu spektra.

Aplikovaním okienkovej funkcie, sa vykonáva potlačenie váhy vzorkov na začiatku a konci segmentov a to tak, že jednotlivé vzorky segmentov sa násobia váhovou alebo tiež okienkovou funkciou. Tvarov funkcií okna je celá rada, ale v oblasti spracovania reči dominuje Hammingov tvar funkcie okna. Pokiaľ sa označí funkcia pre Hammingovo okienko ako $w(n)$, $0 \leq n \leq N - 1$, kde N je opäť počet vzorkov v každom rámci, tak funkcia pre váženie jednotlivých segmentov bude mať tvar

$$w(n) = 0,54 + 0,46 \cdot \cos \left[\left(\frac{1}{2} \cdot N - n \right) \frac{2 \cdot \pi}{N} \right]. \quad (3.2)$$

[1]

Ak sa teda použije oknová funkcia uvedená vo vzťahu (3.2), n -tá vzorka signálu bude vyjadrená takto

$$y(n) = x(n) \cdot w(n), \quad (3.3)$$

[1]

kde $x(n)$ je pôvodná a $y(n)$ novo získaná vzorka signálu.

3.3 Metódy krátkodobej analýzy

Cieľom metódy krátkodobej analýzy je zníženie množstva redundantných dát. Toho sa dá dosiahnuť takýmto spôsobom. Signál nie je reprezentovaný vlastnými nasnímanými vzorkami, ale súborom príznakov získaných z danej vzorky. Snahou je, aby príznaky čo najlepšie popisovali priebeh signálu aj jeho vlastnosti, na ktoré sa v aplikáciách zameriava. Pre každý okruh úloh sú vhodné iné typy krátkodobej analýzy. Hlavnými znakmi výberu určitej analýzy je nenáročnosť na výpočet pri čo najväčšej účinnosti (presnosti) výpočtu.

Pre všetky typy metód je spoločná ich aplikácia len na krátky časový úsek signálu. Zo stacionarity akustického signálu práve z týchto krátkych úsekoch vyplýva nutnosť korektnej analýzy signálu len po malých častiach. V praxi sa teda robí segmentácia signálu na jednotlivé rámce a na ne sa potom aplikuje niektorá z metód krátkodobej analýzy. Pre každý rámec (segment) je získaní jeden príznak – skalárna hodnota, alebo vektor príznakov – akustický vektor, ktorý popisuje signál v konkrétnom rámci. Cieľom je, aby získaný popis bol jednoznačný pre každý typ signálu a čo najmenej ovplyvnený okolitým podmienkam. Musí byť čo najmenej závislý na množstve šumu v signály, na jeho intenzite a pri rozpoznávaní reči na rečníkovi.

Metódy krátkodobej analýzy sa delia do dvoch skupín a to podľa toho, či pracujú s akustickým signálom ako s časovou postupnosťou navzorkovaných hodnôt, alebo či reagujú skôr na jeho spektrálny charakter.

3.3.1 Spracovanie v časovej oblasti

Spracovanie v časovej oblasti vyžaduje najmenšie nároky na výpočet. Z parametrov získaných týmto spôsobom je možné vyčítať iba základné informácie a taktiež sa tieto metódy používajú pre zistenie klasifikácie reč/šum, znelosť/neznelosť reči, alebo v spojení s príznakmi vypočítanými z frekvenčnej oblasti. Medzi príznaky získané z časového priebehu patria napríklad krátkodobá

intenzita, krátkodobá funkcia stredného počtu priechodov signálu nulou alebo krátkodobá autokorelačná funkcia.

2.3.2 Spracovanie vo frekvenčnej oblasti

Hovorená reč je vo frekvenčnej oblasti reprezentovaná zastúpením jednotlivých frekvencií, svojim spektrom. Aj v tomto prípade sa môžu považovať v časovom úseku 10-30 ms za konštantné spektrálne charakteristiky reči. Z toho vyplýva, že sa jedná o krátkodobú spektrálnu analýzu. Základom väčšiny používaných metód spracovania vo frekvenčnej oblasti je Fourierova transformácia (2.1). Spektrum reči získané diskretnou Fourierovou transformáciou, však stále nesie veľké množstvo redundantných informácií. Navyše je korelované (frekvencie sú na sebe závislé) a obsahuje v sebe budenie hlasového traktu, preto sa spektrum pre klasifikáciu segmentov nepoužíva. Spektrum je obvykle rozdelené do niekoľkých pásiem, a spracovávané v každom pásme individuálne.

4. Základná frekvencia rečového signálu

4.1 Vznik základnej frekvencie rečového signálu

Ako už bolo spomenuté, vzduch hnaný pľúcami prechádza cez hlasivkovú štrbinu. Tlakom tohto vzduchu sa stiahnuté hlasivky stávajú pružnými a začnú kmitať. Frekvencia týchto kmitov je závislá najmä na tlaku vzduchu, ale aj na svalovom napätí hlasiviek. Táto frekvencia sa označuje ako základná frekvencia rečového signálu F_0 (fundamental speech frequency) a vyskytuje sa pri tvorení všetkých znelých zvukov, t.j. samohlásky a znelé spoluhlásky. Analogicky jej prevrátenou hodnotou bude perióda základného tónu reči T_0 (pitch period)

$$T_0 = \frac{1}{F_0}. \quad (4.1)$$

Základná frekvencia kmitov hlasiviek je považovaná za jeden z najdôležitejších parametrov pri spracovaní reči, ktorý určuje charakteristiku hovoriaceho, ale tiež výrazne ovplyvňuje informačný obsah vysloveného prejavu. Jeho stredná hodnota je človekom vnímaná ako výška hlasu. Taktiež rozpoznanie hovoriacich počúvaním, je založené veľkou mierou na základnom tóne. Na druhej strane nie je vhodné použiť základný tón ako jediný znak na rozpoznanie hovoriacich. Je to hlavne kvôli pomerne jednoduchému napodobeniu. Znalosť základného tónu sa využíva hlavne pri syntéze reči, ale v poslednej dobe tiež na potlačenie parazitného hluku superponovaného v užitočnej reči.

4.2 Hodnoty základnej frekvencie rečového signálu

Hodnota kmitočtu F_0 je u jednotlivých hovoriacich rôzna a závisí jednak od veku, ale aj od pohlavia. Väčšinou sa nachádza v pásme 50 - 400 Hz. Ako sa dá domyslieť, horná časť tohto rozsahu 300 - 400 Hz patrí detským hláskom. V stredných hodnotách sa nachádzajú ženské hlasy, u ktorých je priemerný kmitočet 210 Hz a najnižšie sú mužské hlasy s priemerným kmitočtom 120 Hz.

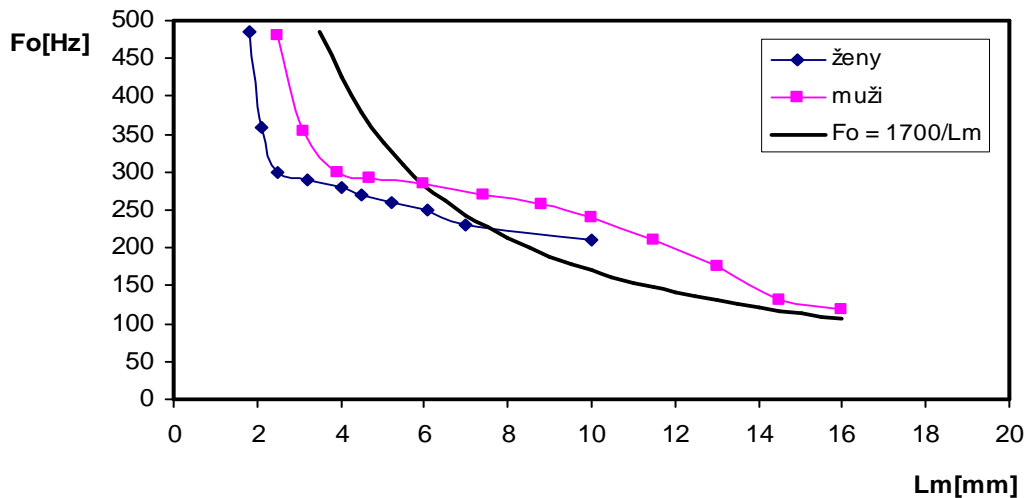
Prečo sa ale F_0 so zvyšujúcim sa vekom zmenšuje? Je to dané vývinom človeka a hlavne so zväčšujúcou sa dĺžkou hlasivkovej membrány L_m . Stredná hodnota základného tónu pokojnej reči (pri vyrovnanom pnutí v hlasivkách) je nepriamo úmerná dĺžke hlasivkovej membrány L_m . Pre nepriamu úmeru týchto veličín bol odvodený vzťah

$$F_0 = \frac{1700}{L_m} \quad [Hz; -, mm]. \quad (4.2)$$

[1]

Na (obr. 4.1) je možné sledovať aproximáciu mužskej aj ženskej krivky na hyperbolu podľa vzťahu (4.2). Jednotlivé body oboch kriviek udávajú aj vek, ktorý je v tomto grafe aj u mužov aj u žien v rozsahu 1 rok (krajná hodnota na ľavej strane) až 20 rokov (krajná hodnota na pravej strane). Na

vysvetlenie je uvedený názorný príklad: dospelý muž (20 rokov) má dĺžku hlasivkovej membrány $L_m = 16$ mm a k tomu odpovedajúcu základnú frekvenciu reči $F_0 = 120$ Hz.



Obr. 4.1: Závislosť medzi základným kmitočtom reči a dĺžkou hlasivkovej membrány.

Výška tónu v západných jazykoch (slovenčina, čeština, atď.) len nepriamo pomáha pri rozpoznávaní slov, avšak má tiež svoj význam. Prenáša dodatočné a užitočné informácie o tom, ktoré slová sú vo vete zdôrazňované, či je veta otázkou, oznamovacou alebo rozkazovacou vetou, ale aj rôzne pocity človeka.

4.3 Vlastnosti základnej frekvencie rečového signálu

Jadrom bakalárskej práce je samotný odhad základnej frekvencie F_0 a popísanie jednotlivých metód, ktoré sa pri tomto procese používajú. Ešte pred tým je potreba ozrejmiť určité vlastnosti rečového signálu, ktoré viac menej tento proces komplikujú. Na nesprávne určenie F_0 majú najväčší podiel tieto javy:

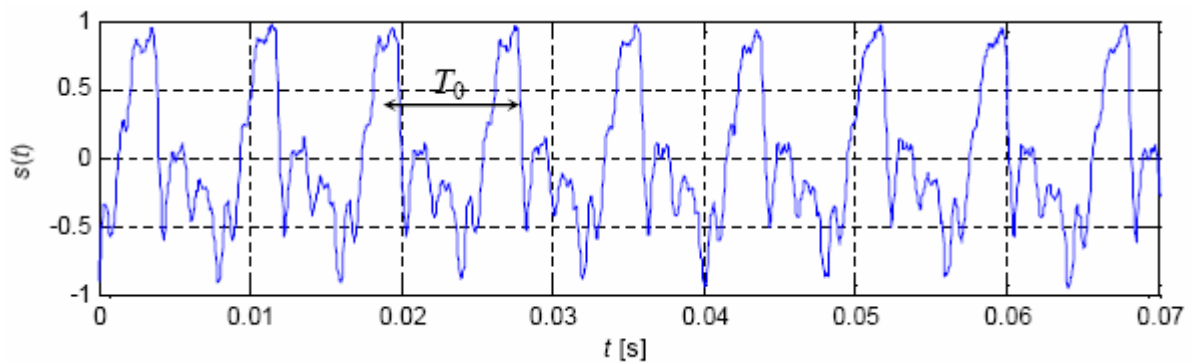
- Priebeh celého rečového signálu je nestály, t.j. môže sa značne meniť rovnako ako vo vnútri analyzovaného segmentu dĺžka periódy T_0 . Z toho vyplýva, že hlasivkami vytváraný signál nie je v žiadnom prípade sled periodických impulzov. To isté platí tiež pre tvar budiaceho signálu (vzduch hnaný z pľúc).
- Základná frekvencia F_0 sa mení s časom, často s každou hlasivkovou periódou.
- Ako bolo spomínané vyššie, základná frekvencia je obsiahnutá len v znelých úsekoch reči. Čisto znelé, alebo neznelé budenie sa vyskytuje len v zriedkavých a výnimočných prípadoch. Dokonca už aj pri analýze jedného segmentu signálu sa môže vyskytnúť prípad zmeny znelý/neznelý.
- Ďalší problém nastáva, ak má signál malú energiu. Je potom veľmi náročné navzájom oddeliť znelé a neznelé úseky.
- Budenie základnej frekvencie a formanty jednotlivých samohlások sa môžu navzájom tak značne ovplyvňovať, že je naozaj ťažké rozpoznať skutočnú periódu. Z toho vyplýva, že je možné aby vysoké kmitočty a najnižší formant ležali veľmi blízko seba, ba dokonca sa aj prekrývali.

- Pri určovaní F_0 z telefónneho signálu je obmedzenie len na telefónne pásmo (300 - 3400 Hz). To znamená, že v telefónnom signály sa často vyskytujú iba vyššie harmonické zložky základného tónu pôvodného signálu. Filtrácia za účelom získania F_0 by teda bola na nič.

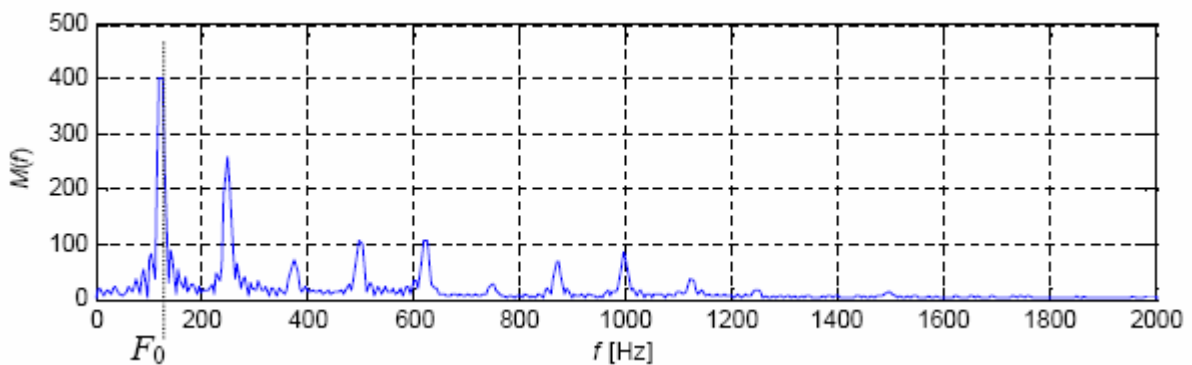
4.4 Možnosti využitia základného tónu reči

Priebeh základného tónu reči (popr. jeho stredná hodnota) nájde svoje využitie napríklad v týchto prípadoch:

- Kódovanie reči pomocou lineárnej predikcie, kde sa okrem LPC koeficientov a ďalších parametrov uchováva či prenáša práve základný tón reči.
- Syntéza reči a modelovanie prozódie.
- Analýza reči.
- Detekcia emočného stavu hovoriaceho.
- Identifikácia hovoriaceho.
- Detektory hlasovej aktivity, a ďalšie.



(a)



(b)

Obr. 4.2: Znázornenie základného tónu reči v časovej (a) a kmitočtovej oblasti (b)

5. Metódy používané na určenie základnej frekvencie rečového signálu

Odhad základnej frekvencie, niekedy tiež nazývaný extrakcia tónov, je populárna téma v oblasti výskumu. Ale aj po zhruba 50. rokoch štúdií, metódy naň používané stále nedosahujú požadované presnosti a odolnosti voči rôznym vplyvom. Síce za predpokladu, že je analyzovaný signál „čistý“, väčšina metód svoju úlohu splní takmer na výbornú. Ale ako náhle je signál zašumený, čo je reálny prípad, množstvo algoritmov stále nie je schopné splniť funkciu podľa očakávania.

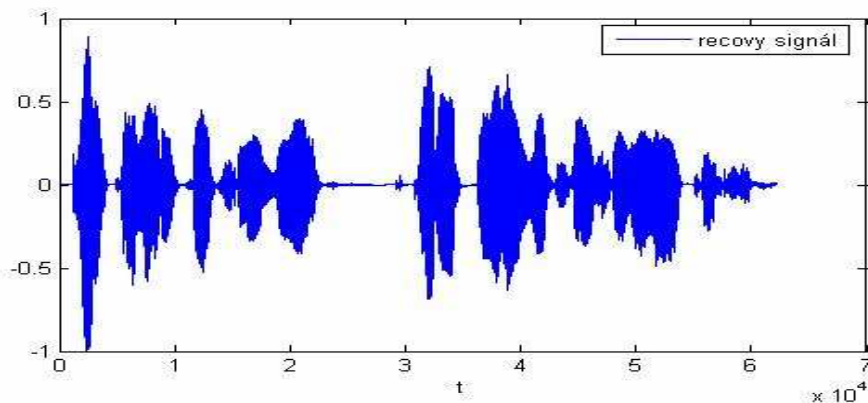
Základný problém je teda oddelenie základného tónu F_0 od ostatných nežiaducich zvukov. Ako bolo už spomenuté, ďalším problémom je fakt, že rečový signál obsahuje vo svojom frekvenčnom spektre takzvané formantové štruktúry (F_1, F_2, F_3, \dots). Ak sa berie na vedomie, že hodnota frekvencie základného tónu je rôzna pre rôznych rečníkov a že sa táto frekvencia môže aj pre jednotlivca meniť pri rozprávaní v pomere až 2:1, je zrejmé, že oddelenie základnej frekvencie nebude len tak jednoduché. Určite je možné počítať s tým, že by sa oddelila základná frekvencia priechodom rečového signálu vhodnou dolnou prepustňou. Problém je v tom, že v tomto prípade je ťažké odhadnúť univerzálnu hodnotu zlomovej frekvencie, ktorá by v každom danom okamžiku oddelila F_0 od jeho vyšších harmonických zložiek a od formantovej štruktúry.

V dnešnej dobe už sa používa veľa postupov a metód, ktoré sú cielené k čo najväčšej eliminácii nepriaznivého vplyvu formantov na spoľahlivosť a presnosť určenia základného tónu. Následne budú niektoré z nich predstavené a ukázané ako fungujú. Tieto metódy sa môžu rozdeliť do troch skupín, podľa toho v akej oblasti je ich využitie.

- Časová oblasť
- Frekvenčná (spektrálna) oblasť
- Kepstrálna oblasť

5.1 Metódy používané v časovej oblasti

Tieto metódy, alebo postupy pracujú tak, že sa snažia zistiť, ako často sa signál v čase opakuje. V literatúre rozoberajúcej tieto metódy sa uvádza toto. Ak je signál periodický, existujú odhadnuteľné časovo sa opakujúce časti, ktoré sú počítané a počet týchto častí za 1 sekundu, nepriamo súvisí s frekvenciou. Každá metóda z tejto oblasti je užitočná na rozličné druhy signálov, preto je dobré mať aspoň základné poznatky o čo najväčšom počte.



Obr. 5.1: Príklad zobrazenia rečového signálu v čase

5.1.1 Detekcia počtu priechodov nulou (Zero-crossing rate ZCR)

V navzorkovanom signály môže nastať priechod nulou za predpokladu, že majú dve susedné vzorky opačné znamienko. Formálne vyjadrenie ZCR je definované takto

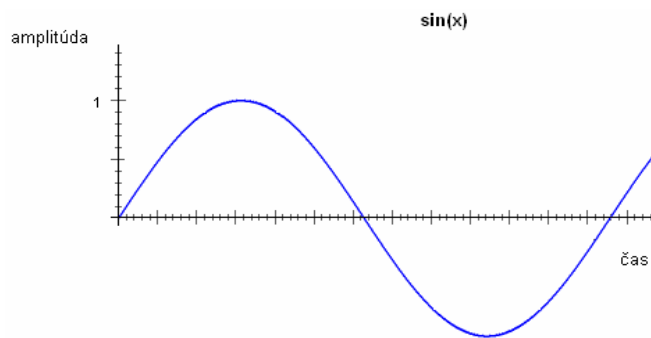
$$ZCR = \frac{1}{2} \sum_{n=1}^{N-1} |\text{sign } s(n) - \text{sign } s(n+1)|, \quad (5.1)$$

[1]

kde

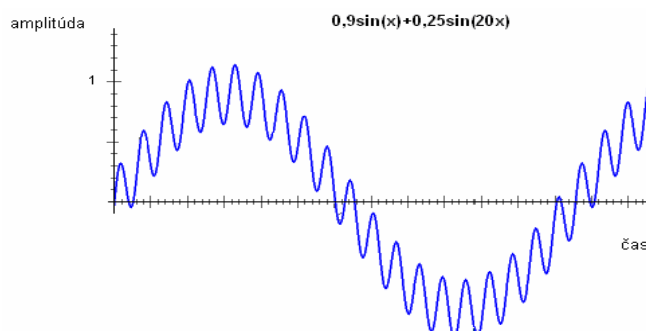
$$\begin{aligned} \text{sign } s(n) = +1 & \quad \text{ak} \quad s(n) > 0 \quad \text{alebo} \quad s(n) = 0 \quad \wedge \quad s(n+1) > 0 \\ \text{sign } s(n) = -1 & \quad \text{ak} \quad s(n) < 0 \quad \text{alebo} \quad s(n) = 0 \quad \wedge \quad s(n-1) < 0. \end{aligned}$$

Táto metóda býva veľmi často spochybňovaná. Meria ako často prejde signál cez nulovú úroveň za jednotku času. ZCR podáva informáciu o spektrálnom obsahu signálu. Základný kameň tejto metódy spočíva v tom, že dokáže priamo určiť hodnotu, koľko krát sa signál opaková za jednotku času. Ak je sila spektra sústredená okolo F_0 , potom prejde signál nulou dva krát za periódu, ako je naznačené na (obr. 5.2).



Obr. 5.2: Signál $\sin(x)$ v čase

Problém nastáva v momente, ak signál obsahuje vysokofrekvenčné zložky, ako je napríklad (obr. 5.3). V tom to prípade môže signál prejsť cez úroveň nuly viac ako dva krát a tým pádom je náročné odhadnúť periódu a z nej vypočítať frekvenciu.



Obr. 5.3: Signál $0,9\sin(x)+0,25\sin(20x)$ v čase

ZCR metóda môže byť vylepšená tým, že by sme z pôvodného signálu odfiltrovali vysokofrekvenčné zložky, ktoré sú zodpovedné za nepresnosť výsledkov. Filtrácia sa musí urobiť opatrne, pretože hrozí riziko odfiltrovania aj jednotlivých F_0 , ale na druhú stranu musí odfiltrovať čo najviac. Táto metóda sa pre nie príliš presvedčivé výsledky používa skôr na detekciu znelých a neznelých hlások. Znelé hlásky majú o mnoho nižšiu frekvenciu ako neznelé a tým aj menší počet prechodov nulovou hladinou.

5.1.2 Detekcia počtu vrcholov (Peak rate)

Princíp tejto metódy je založený na súčte počtu kladných vrcholov za sekundu. Teoreticky signál dosahuje maximálnej a minimálnej hodnoty v každej perióde. Stačí len jedna na vypočítanie týchto maximálnych, alebo minimálnych hodnôt a z nich určiť frekvenciu signálu. V reálnych prípadoch to ale vyzerá trochu inak. Musí sa použiť detekcia lokálnych maxím s cieľom nájsť miesto, kde je signál lokálne najväčší. Počet týchto lokálnych maxím za jednu sekundu určuje frekvenciu signálu, okrem prípadu, kedy perióda obsahuje viac ako jedno maximum.

5.1.3 Autokorelačná funkcia (ACF)

ACF je najdôležitejšou a najpoužívanejšou metódou analýzy rečového signálu v časovej oblasti. Táto funkcia má niektoré užitočné vlastnosti, ktoré indikujú energetické, ale aj frekvenčné charakteristiky rečového signálu.

Korelácia dvoch signálov je vlastne meranie ich podobnosti. Signáli sa porovnávajú v rôznych časových intervaloch a ich podobnosť sa určuje v každom intervale. Výsledkom korelácie je hodnota podobnosti ako funkcie časového posunu medzi začiatkami dvoch priebehov.

Autokorelačná funkcia je teda meranie podobnosti dvoch rovnakých signálov, iba že jeden z nich je o k vzorkou posunutý. Formálny zápis ACF pre nekonečne spojitý signál $s(n)$ je

$$R(k) = \sum_{n=-\infty}^{\infty} s(n)s(n+k). \quad (5.2)$$

[1]

V praxi sa ale používa ACF na konečne spojitý signáli. Je to dané segmentáciou signálu, ktorá signál rozdelí na krátke úseky. Je nutné dodať že, ak má táto funkcia obsahovať informáciu o veľkosti základnej periódy, musí byť dĺžka segmentu analyzovaného signálu väčšia než jedna perióda F_0 . Vzhľadom na to, že základná frekvencia hlasiviek u niektorých ľudí môže byť nižšia než 150 Hz, doporučuje sa voliť dĺžku mikrosegmentu v rozmedzí 20-40 ms. Potom sa môže vzorec (5.2) obmedziť na jeden segment dĺžky N a dostane tvar

$$R(k) = \sum_{n=0}^{N-1-k} s(n)s(n+k). \quad (5.3)$$

[2]

Pri určovaní základnej frekvencie rečového signálu pomocou autokorelačnej funkcie sa hľadajú jej maximálne hodnoty. Ak sa začne posúvať jeden signál voči pôvodnému smerom k polovici periódy, korelácia dostáva klesajúcu tendenciu až klesne na minimum. V tomto bode je presne polovica periódy. Ďalším posúvaním sa začína korelácia zväčšovať, až nadobudne svoje maximum v bode periódy.

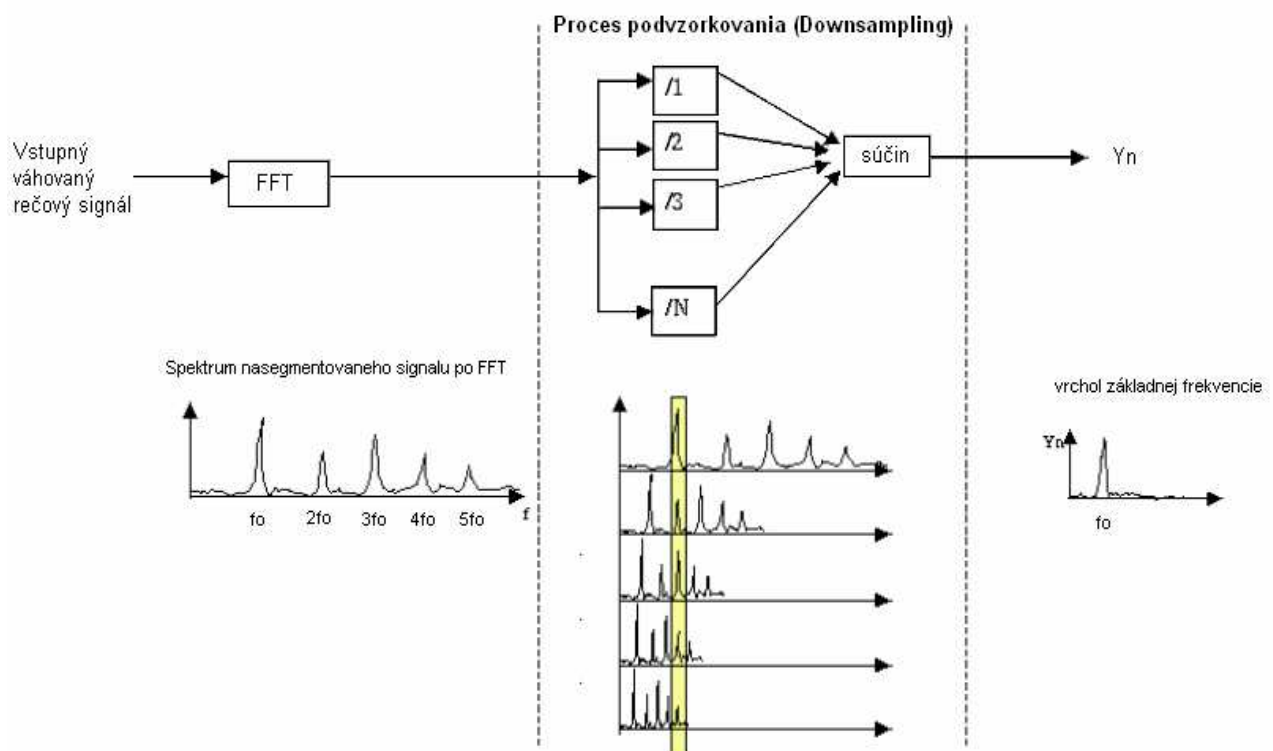
Ak bude analyzovaný signál periodický s periódou T potom ACF nadobúda maximálnych hodnôt práve pre $k=0, T, 2T, \dots$. ACF je vhodné používať pre nízke až stredné kmitočty. V závislosti na dĺžke signálu môže byť autokorelačná funkcia extrémne výpočetne náročná.

5.2 Metódy používané vo frekvenčnej (spektrálnej) oblasti

U metód v časovej oblasti bol signál rozložený v čase. To znamená že čas reprezentoval os x . Hľadala sa základná perióda T_0 analyzovaného signálu a až potom sa z nej podľa vzorca (4.1) vypočítala základná frekvencia F_0 . Podľa názvu je jasné, že u týchto metód je situácia dosť odlišná vzhľadom na to, že na ose x je rovno frekvencia čiže spektrum signálu. Na prevedenie pôvodného signálu rozloženého v čase do spektra, poslúži vyššie spomínaná Fourierova transformácia (2.1).

Metódy frekvenčnej oblasti majú väčšinou rovnaký postup. Analyzovaný signál sa nasegmentuje a potom je každý z týchto segmentov vynásobený tzv. váhovým okienkom. Jeho úlohou je vybrať po úsekoch príslušné vzorky signálu a prideliť im určitú váhu. Najčastejšie sa používa pravouhlé a Hammingovo okno. Po segmentácii a váhovaní príde na rad už spomínaná FT, ale k vôli krátkym segmentom špeciálna rýchla Fourierova transformácia (FFT). Ak je signál periodický, FT priamo ukáže vrcholy v násobkoch základnej frekvencie. Teraz sa bližšie priblíži funkcia jednotlivých metód.

5.2.1 Metóda HPS (harmonic product spectrum)

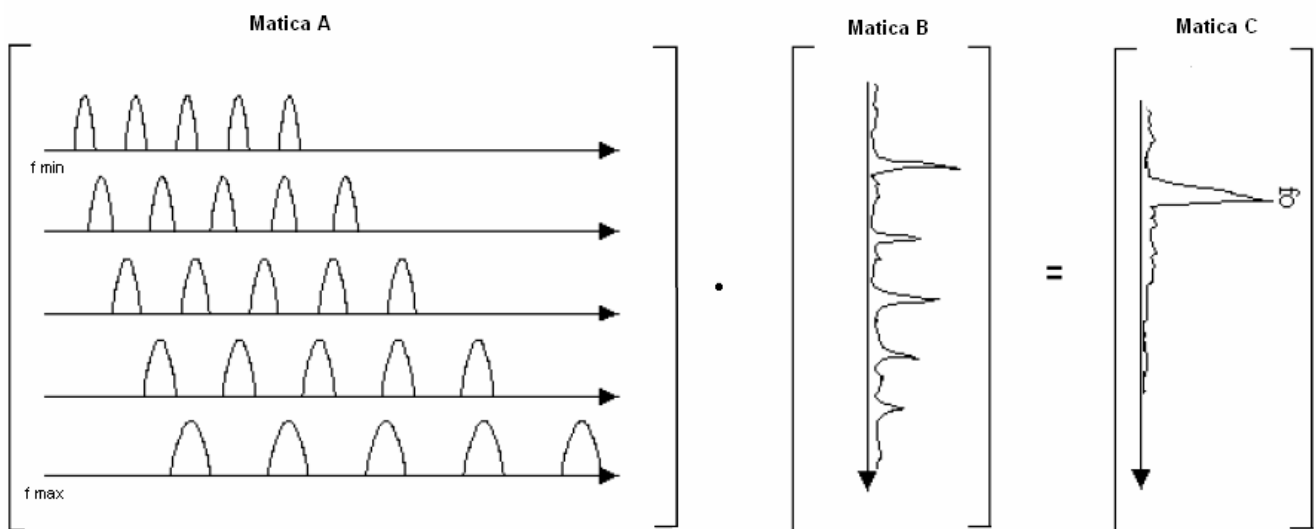


Obr. 5.4: Princíp metódy HPS logaritmu

Táto metóda má jednoduchú implementáciu. Princíp je založený na predpoklade, že rečový signál je výsledok filtrovania základnej frekvencie hlasovým traktom. Po segmentácii signálu a vynásobení váhovým oknom, je signál HPS algoritmom najprv prevedený z časovej do frekvenčnej oblasti pomocou rýchlej Fourierovej transformácie (FFT) a následne toto vzniknuté spektrum niekoľko krát podvzorkuje (downsampling). Ak sa podvzorkuje frekvenčné spektrum nasegmentovaného signálu na polovicu, tretinu, štvrtinu, atď., harmonické zložky sa zarovnajú na pozíciu základného tónu. To znamená že, prvý vrchol pôvodného spektra má tú istú pozíciu na frekvenčnej osi ako druhý vrchol spektra, ktoré sa podvzorkovalo na polovicu a jeho pozícia súhlasí s pozíciou tretieho vrcholu spektra podvzorkovaného tretinou. Samozrejme že cyklus pokračuje ďalej. Konečným krokom je súčin všetkých týchto rozdielnych spektier. Po vynásobení sa objaví jeden spoločný vrchol a ten určuje základnú frekvenciu F_0 . [5]

5.2.2 Metóda maximálnej podobnosti (maximum likelihood)

Veľmi často sa označuje ako Wienerovo riešenie. Hlavnou myšlienkou je u tejto metódy fakt, že spektrum pôvodného signálu, ktoré vznikne po segmentácii a aplikovaní rýchlej Fourierovej transformácie, je ďalej porovnávané (korelované) vo frekvenčnej oblasti s jedným ideálnym spektrom, ktoré je vybrané zo sady všetkých možných ideálnych spektier a najviac sa podobá na pôvodné spektrum. Najpodobnejšie ideálne spektrum je nazývané tiež vláčik okienok. Lepšie vysvetlenie ukazuje grafické zobrazenie tejto metódy (viz. obr. 5.5).



Obr. 5.5: Princíp metódy maximálnej podobnosti

Na (obr. 5.5) sú 3 matice. Matica A zobrazuje sadu všetkých možných ideálnych spektier (vláčikov) od minimálnej až po maximálnu frekvenciu. Tá je násobená (korelovaná) s maticou B, ktorá zastupuje spektrum pôvodného vstupného rečového signálu. Vo výsledku, matica C, je spektrum vzniknuté pri najväčšej hodnote korelácie matíc A a B, čiže 2. najpodobnejšie spektrá a toto spektrum vo výsledku vykreslí jeden vrchol, v ktorom sa nachádza základná frekvencia. [5]

5.3 Metódy používané v kepstrálnej oblasti (Cepstrum)

Výraz „kepstrum“ bol zavedený prehodením prvých 4. písmen v slove „spektrum“. Otočenie písmen vyjadruje inverzné spektrum, teda časovú funkciu. V tomto prípade ale nie je reč o bežnej časovej funkcii, lebo medzi obe Fourierove transformácie je vložená operácia logaritmovania. Podobne vznikli v anglickej odbornej literatúre ďalšie špeciálne výrazy:

| | |
|------------------|---------------------|
| spectrum | cepstrum |
| <i>frequency</i> | <i>quefrequency</i> |
| <i>filtering</i> | <i>liftering</i> |
| <i>harmonic</i> | <i>rahmonic</i> |

5.2.2 Kepstrálna analýza (cepstral analysis)

Kepstrálna analýza patrí do homomorfnej skupiny metód, teda do skupiny nelineárneho spracovania rečových signálov, ktoré využívajú zobecnený princíp superpozície. Tieto metódy sú najčastejšie aplikované na oddelenie (dekonvolúciu) signálov, ktoré vznikli konvolúciou alebo násobením dvoch či viacerých zložiek.

Ako bolo spomenuté v kapitole 2. hlasivky rozkmitané vzduchom z pľúc vytvárajú šumovú budiacu funkciu $g(n)$ a hlasový trakt s impulznou odozvou $h(n)$ túto funkciu moduluje. Konvolúciou $g(n)$ a $h(n)$ vzniká na výstupe hlasového traktu rečový signál $s(n)$, čo v prevedení do kmitočtovej oblasti odpovedá súčinu oboch funkcií po Fourierovej transformácii.

Jednoduchá metóda dekonvolúcie rečových signálov spočíva v aplikácii matematickej operácie logaritmovania súčinu spektrálnych funkcií

$$\log|S(f)| = \log[|G(f)||F(f)|] = \log|G(f)| + \log|H(f)|. \quad (5.4)$$

Logaritmované súčinové zložky sa po operácii logaritmovania premenia na súčet. Operácia logaritmovania mení amplitúdu spektra, ale nie jeho charakteristiku. Inverzná Fourierova transformácia transformuje zložky z frekvenčnej naspäť do časovej oblasti, kde existujú ďalej ako súčet, pretože ten ostáva po FT zachovaný

$$F^{-1}\{\log|S(f)|\} = F^{-1}\{\log|G(f)|\} + F^{-1}\{\log|H(f)|\}. \quad (5.5)$$

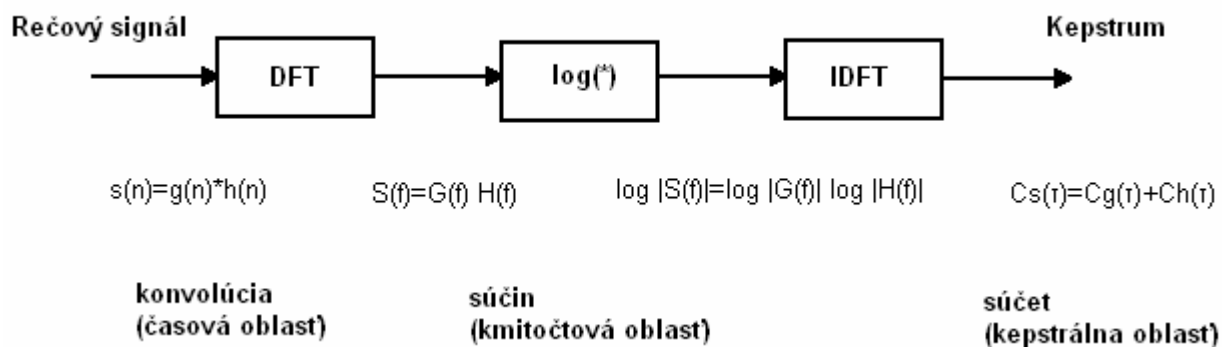
Po prevedení na časové funkcie je k dispozícii takýto tvar, v ktorom výstupný signál $c_s(\tau)$ je označovaný ako kepstrum signálu

$$c_s(\tau) = c_g(\tau) + c_h(\tau). \quad (5.6)$$

Zjednodušene nám komplexné kepstrum predstavuje inverznú Fourierovu transformáciu (FT) logaritmu Fourierovho obrazu vstupného signálu

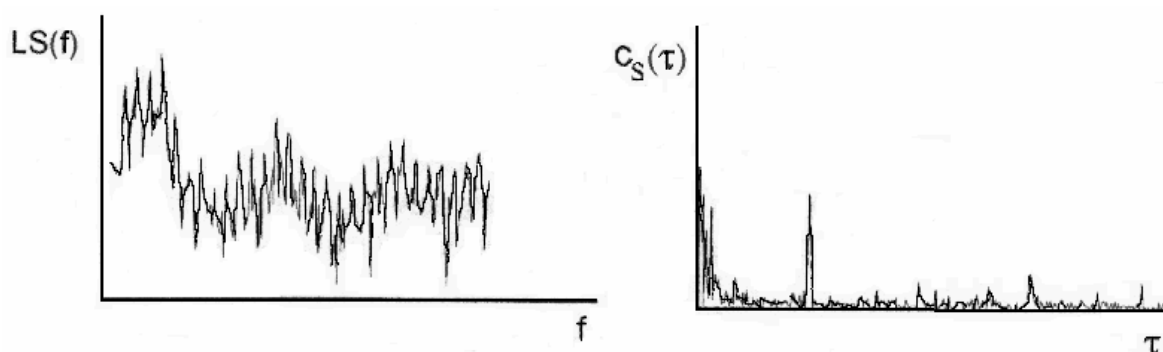
$$c_s(\tau) = F^{-1}\{\log|F\{s(n)\}|\}. \quad (5.7)$$

Poznámka: Vzťahy (5.4), (5.5), (5.6) a (5.7) sú požitité z literatúry [2] uvedenej na konci textu.



Obr. 5.6: Princíp systému na určenie kepstra rečového signálu

Na (obr. 5.7) je zobrazené kepstrum znelého úseku rečového signálu. Osobitná špička charakterizuje základnú periódu T_0 rečového signálu a nepravidelne zvlneý priebeh v okolí nulového bodu ukazuje vplyv formantov. Na určenie základnej frekvencie F_0 z kepstra sa teda hľadá poloha špičky T_0 , z ktorej sa podľa vzťahu (4.1) vypočíta F_0 .



Obr. 5.7: Logaritmicke spektrum a k nemu odpovedajúce kepstrum znelého segmentu

Metódy kepstrálnej analýzy v rôznych modifikáciách preukázali pomerne dobrú spoľahlivosť, aj po nespočetných výskumoch v oblasti spracovania rečového signálu, ako napríklad rozpoznávanie fonémou, overovanie hovoriacich, rozpoznávanie pohlavia, určovanie stresu a iných psychických stavov.

6. Robustné algoritmy na odhad F_0 odolné voči šumu

Je známe, že v prípade ak je rečový signál prekrytý hlukom, čo je v reálnom svete spracovaní rečových nahrávok bežná vec, sú klasické filtračné metódy nepoužiteľné. Pre tieto prípady už klasické metódy nebudú stačiť na požadovanú presnosť výsledkov. Musia sa použiť teda zložitejšie metódy, ktoré využívajú znalosť charakteru rušenia alebo priamo charakter rečového signálu.

6.1 Potlačenie šumu pozadia rečového signálu mapovaním spektrogramu metódou RASTA

Metóda RASTA je založená na modulačnom spektre akustického signálu. Po aplikácii transformácie STFT (short time Fourier transformation), nasleduje prevod komplexného spektrogramu na modulovú a fázovú maticu (STFTM a STFTP). Na základe následnej frekvenčnej analýzy časových trajektórií jednotlivých spektrálnych zložiek STFTM, premenlivých v čase, sa dopracuje k modulačnému kmitočtu. Ak sa skúmajú časové trajektórie celého spektra, hovorí sa o tzv. modulačnom spektre. Pretože modulačné spektrum reči má určité vlastnosti, aplikáciou filtra na modulačné spektrum sa dajú odfiltrovať akustické signály s odlišnými modulačnými kmitočtami aj v prípade, keď je rečový signál spektrálne mapovaný. Metóda RASTA využíva neustálu filtráciu časových trajektórií spektra, pričom charakter použitého filtra môže byť závislý na hodnote pomeru signálu k šumu (SNR).

6.1.1 Filter RASTA

Filter RASTA $R(n)$ je obvykle filter typu pásmovej prepusti PP s rôznymi frekvenčnými odozvami, ktoré potlačujú modulačné frekvencie pod 1 Hz a nad 16 Hz. Filtre v oblastiach nízkej spektrálnej energie sú typu dolnej prepusti DP, jednosmerný zisk filtrov je vysoký pri vysokom SNR časových trajektórií a nízky pri nízkom SNR.

6.1.2 Aplikácia metódy

Metóda RASTA teda aplikuje filtre typu pásmovej prepusti s konštantnými koeficientmi na časové trajektórie spektra. Tento prístup môže byť dosť efektívny v prípade, keď je spektrum hluku veľmi premenlivé v čase v porovnaní so spektrom rečového signálu.

RASTA môže poskytnúť dobré výsledky tiež v opačnom prípade, ak sa charakter spektra hluku vôbec nemení. Stály, nemeniaci sa hluk motora sa v spektrograme objaví ako jednosmerná zložka. Práve preto sú RASTA filtre charakteru pásmovej, a nie dolnej prepusti.

6.1.3 Nevýhody metódy

Problém nastane v prípade, keď sa zhoduje užitočné frekvenčné pásmo rečového signálu a hluku a keď sa podobá tiež charakter hluku a charakter reči. To znamená trvanie jednotlivých hlukov je podobné s trvaním hlások (s dobou zmeny hlasových orgánov), a preto sa pásmová prepust' stane neúčinnou. Toto je prípad, kedy technika RASTA nefunguje. [8]

6.2 Pásmová filtrácia PF

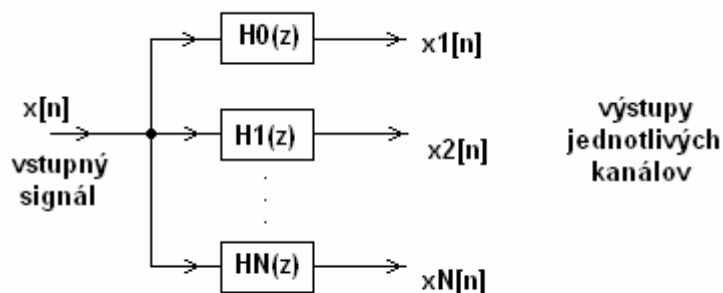
V mnohých prípadoch aplikácií spracovania rečového signálu je účelné vykonávať jeho pásmovú filtráciu bankou číslicových filtrov. Pásmové analyzátory reči sú pritom vedľa komunikačných systémov veľmi často používané v systémoch analýzy a rozpoznávaní reči. Vhodným výberom počtu a šírky pásiem je možné efektívne detegovať fonetickú štruktúru reči a využiť informácie obsiahnuté vo výstupných signáloch jednotlivých kanálov, napr. pre účinnú klasifikáciu fonémov (najmä samohlások), fonémových skupín alebo celých slov. Ľudské ucho spracováva reč podobným spôsobom ako je analýza bankou filtrov.

6.2.1 Banky číslicových filtrov

Banky číslicových filtrov predstavujú novú metódu efektívneho spracovávanía vybraných typov signálov, ako napríklad rečový alebo obrazový signál.

6.2.1.1 Bližšia špecifikácia

Banka filtrov je systém, ktorý umožňuje ľubovoľný digitálny signál rozdeliť do niekoľkých menších frekvenčných pásiem, ktoré je možné potom spracovávať jednotlivo. Spracovávanie v užšom frekvenčnom pásme prináša so sebou vyššiu výpočtovú efektívnosť, väčšiu presnosť a minimálnu redundanciu. To je výhodné pri redukcii šumu a rušenia, alebo zmeny parametrov zvuku (hudby, reči, atď.). Na (obr. 6.1) je znázornený princíp rozkladu signálu do jednotlivých pásiem z filtermi rôznych prenosových charakteristík. [4]



Obr. 6.1: Príklad rozkladovej banky filtrov

Banka filtrov sa obecné delí na dve časti:

- rozkladová (analyzujúca) časť banky filtrov
- rekonštrukčná (syntetizujúca) časť banky filtrov

Jednotlivé kanály analyzujúcej časti sú zložené väčšinou z pásmovej prepusti, ktorá vymedzí šírku frekvenčného pásma spracovávaného signálu. Signál ďalej prechádza cez nelineárny stupeň ktorý je realizovaný väčšinou dvojcestným usmerňovačom a neskôr je ešte filtrovaný dolnou prepust'ou DP. V niektorých prípadoch sa signál ešte podvzorkuje, pre skrátenie vstupnej postupnosti(v syntetizujúcej časti sa naspäť nadvzorkuje aby, bola rekonštrukcia plnohodnotná).

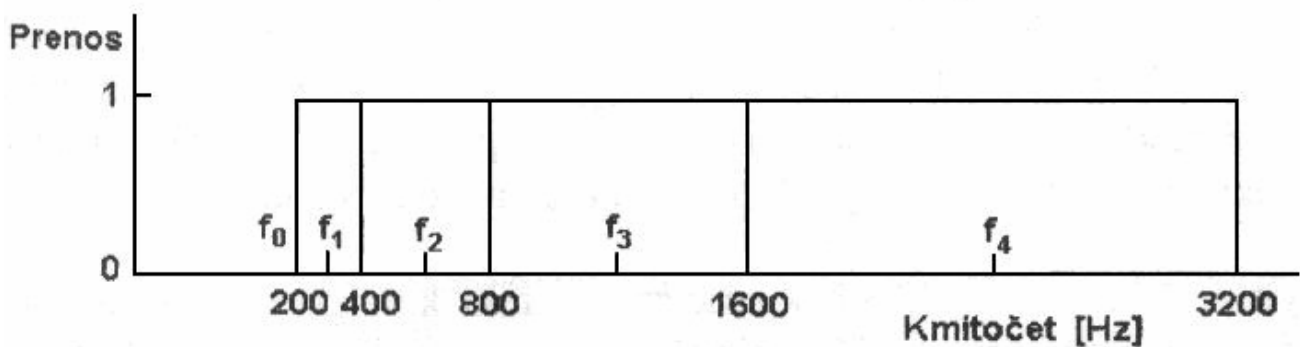
Na takto upravený signál je možné aplikovať jednotlivé metódy či už na vypočítanie energie v jednotlivých pásmach, alebo metódy na hľadanie základnej frekvencie F_0 .

Jednotlivé kanály syntetizujúcej časti obsahujú podobné prvky ako analyzujúca časť. Operácia nadvzorkovania, pokiaľ prebehlo podvzorkovanie a taktiež obsahujú filtre s rôznymi prenosovými charakteristikami, ktoré môžu slúžiť ako interpolačné filtre.

6.2.1.2 Rozdelenie frekvenčného pásma

Na prvé pásmo banky filtrov sa používa dolná prepust' DP alebo pásmová prepust' PP. Pri použití DP sa vychádza z 0 Hz. Pri použití pásmovej prepusti PP sa volí centrálny kmitočet filtra podľa rozloženia všetkých pásiem na frekvenčnej osi. Rozloženie pásiem môže byť v najjednoduchšom prípade lineárne s konštantnou šírkou pásiem. Vzhľadom na podobnosť s ľudským vnímaním zvukov, sa často používa banka filtrov s logaritmickým rozložením pásiem. V poslednej dobe sa dosť používa oktávové rozmiestnenie pásiem, ktoré je uvedené na (obr. 6.2), ale aj tretinooktávové či Melovské rozloženie pásiem.

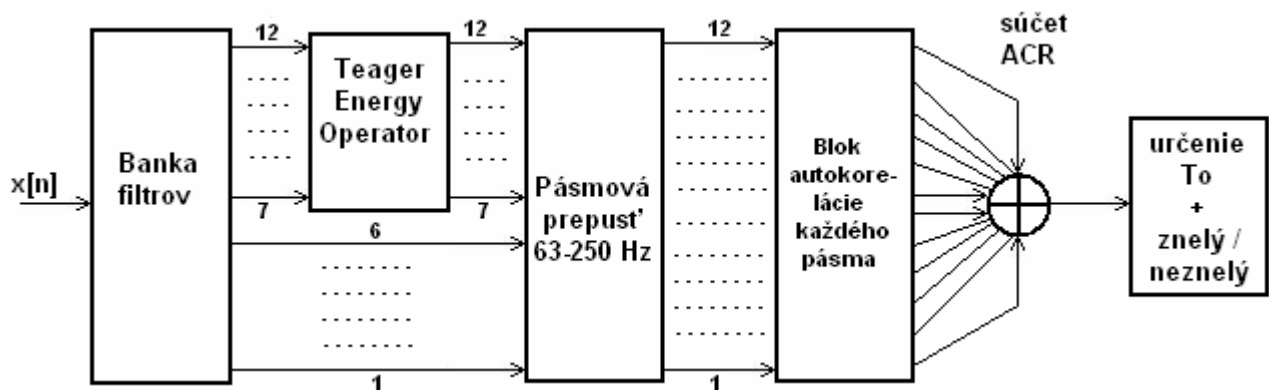
Kmitočtové pásma jednotlivých kanálov by mali na seba nadväzovať tak, aby výsledné spektrum malo hladkú obálku. Počet pásiem a ich rozloženie závisí na aplikácii pásmovej filtrácie. Pre rozpoznávanie reči sa obvykle používa 8-20 pásiem, pre prenos reči pomocou vokodéra 10-32 pásiem [2].



Obr. 6.2: Ideálna oktávová banka filtrov

7. Realizácia algoritmu

Realizácia robustného algoritmu sa začala programovaním banky číslicových filtrov. Rozsah frekvenčného pásma bol zvolený od 0 Hz do 10 kHz, čo je polovica vzorkovacej frekvencie testovaných súborov wav. Návrh pokračuje zaradením modulu operátora Teager energy pre vyššie a stredné pásma. Za ním nasleduje filter typu pásmová prepust', ktorý ale filtruje všetky pásma banky filtrov s rozsahom 63 Hz - 250 Hz. Ďalej je v algoritme umiestnený modul autokorelácie pre každé pásmo zvlášť. Po tomto procese nasleduje sčítanie jednotlivých matic autokorelácie, modul na určenie znelosti segmentov a následne určovanie základnej periódy T_0 . Bloková schéma algoritmu je na (obr.7.1). Celý algoritmus bol navrhovaný podľa článku „A pitch determination and voiced/unvoiced decision algorithm for noisy speech“ [11].



Obr.7.1.: Zjednodušená bloková schéma algoritmu na odhad T_0 v zašumenej reči

7.1 Banka filtrov

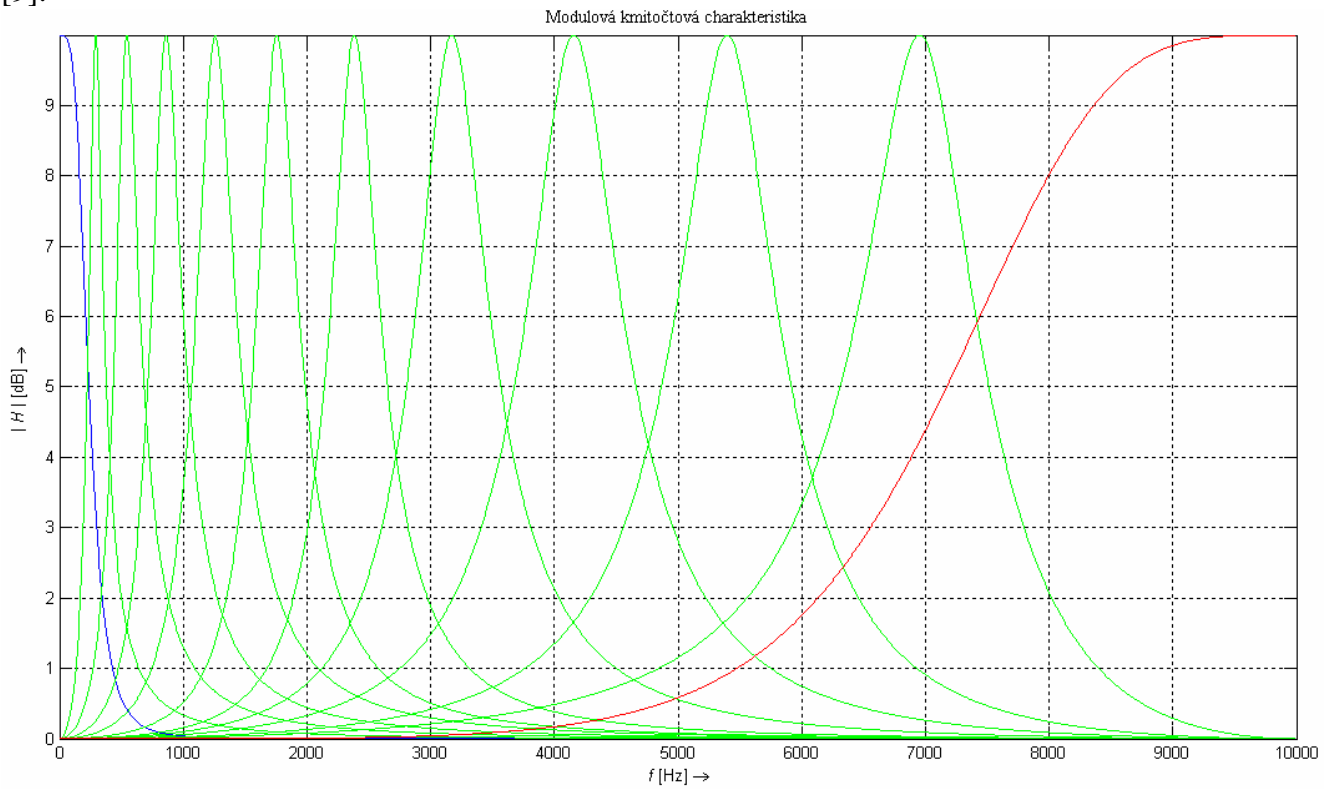
Základ banky filtrov bol použitý z programu 3-pásmový parametrický filter, z predmetu zaradeného do magisterského štúdia - Číslicové zpracování signálů (MCSI), ktorý som prerobil do požadovanej podoby. Banka filtrov je zložená z 12-tich filtrov v rozsahu 10 kHz. Použité sú parametrické filtre typu *shelving* pre 1. pásmo - dolná prepust' a 12. pásmo - horná prepust'. Pásma 2 - 11 sú pásmové prepusti a tieto parametrické filtre sú typu *peak*. Frekvenčná charakteristika banky filtrov je uvedená na (obr.7.2).

7.1.1 Rozloženie filtrov vo frekvenčnej oblasti

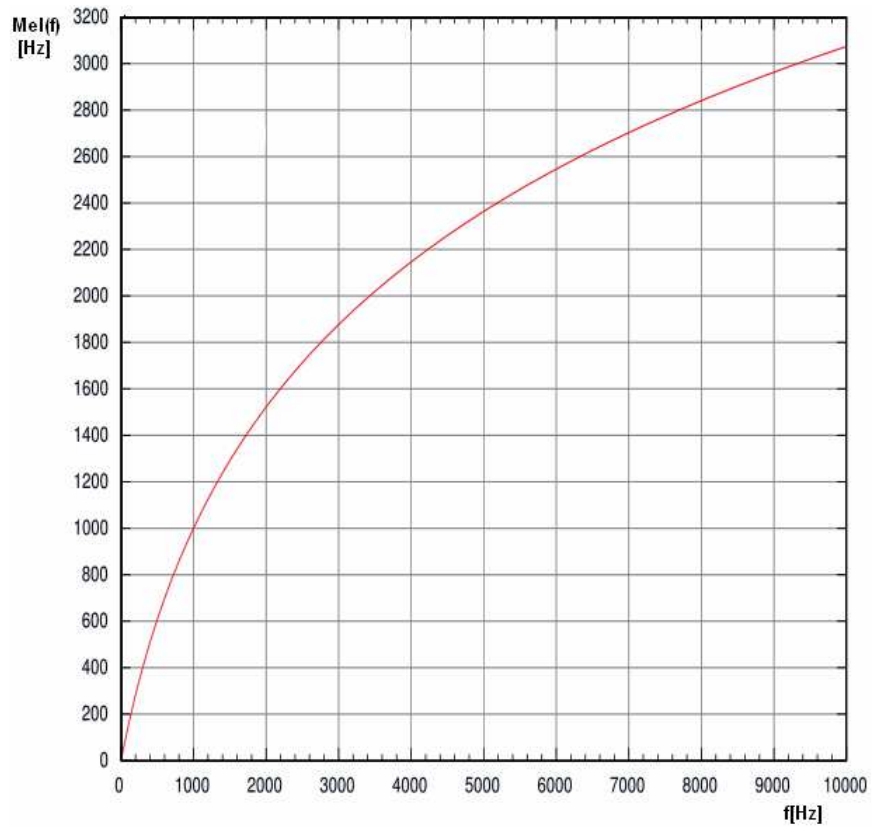
Pri návrhu rozloženia pásiem banky filtrov sa využíva skutočnosť, že človek vníma frekvencie zvuku približne v logaritmickú (melovskú) stupnici, (viz obr. 7.3). Prevodný vzťah medzi frekvenciou a melovskou frekvenciou je daný

$$Mel(f) = 1125 \cdot \ln\left(1 + \frac{f}{700}\right), \quad (7.1)$$

kde $Mel(f)$ je melovská frekvencia, f je pôvodná frekvencia a \ln je prirodzený logaritmus [9].



Obr. 7.2: Modulová frekvenčná charakteristika banky filtrov

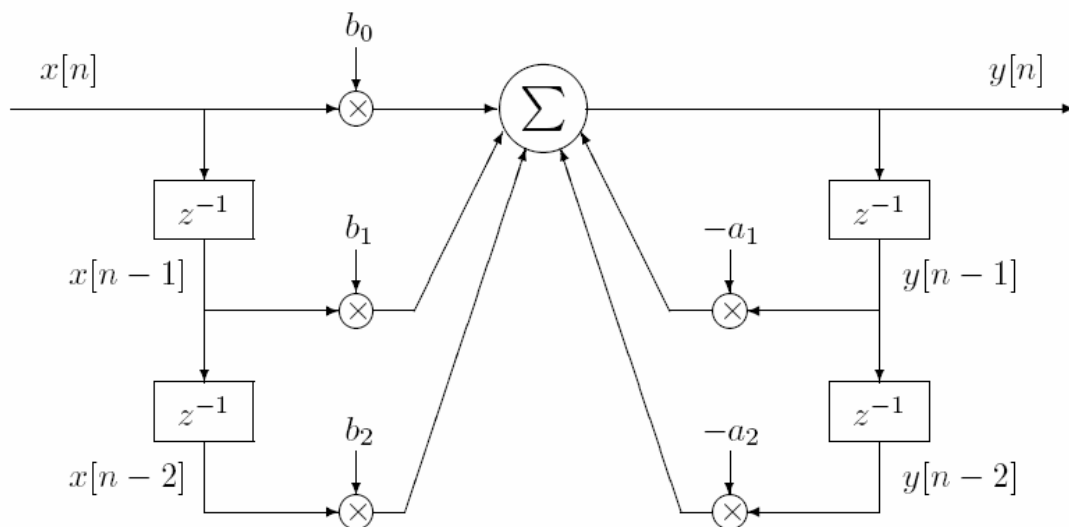


Obr. 7.3: Graf závislosti melovskej frekvencie $Mel(f)$ na frekvencii f

7.1.2 Parametrické filtre

Parametrické filtre sú navrhované na základe IIR filtrov 2. rádu. Schematické znázornenie je uvedené (viz. obr. 7.4). Parametrické filtre sa vyznačujú vlastnosťou plynulej zmeny svojich 3. parametrov:

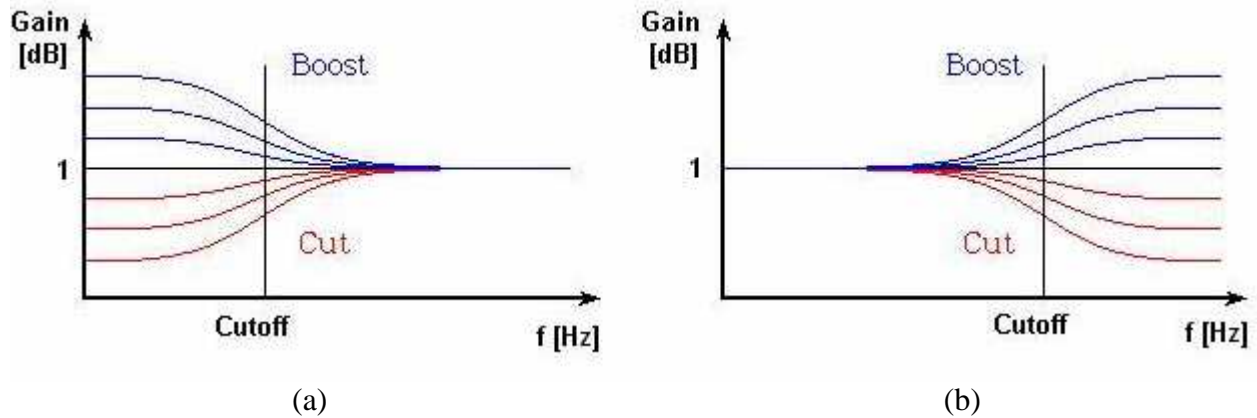
- medzný/stredný kmitočet f_c (cut-off/center frequency)
 - zosilnenie priepustného/potlačenie nepriepustného pásma G (gain)
 - kvalita filtra Q (quality):
- u filtrov typu pásmová prepust' a zádrž súvisí kvalita so strmosťou prechodného pásma a šírkou priepustného pásma,
 - u filtrov typu dolná a horná prepust' Sallen & Key súvisí kvalita so zosilnením filtra v okolí medzného kmitočtu (rezonancie),
 - u filtrov typu dolná a horná prepust' navrhnutých pomocou väčšiny aproximácií je možné zmeniť strmosť (šírku prechodného pásma) len skokovo zmenou rádu filtra [10].



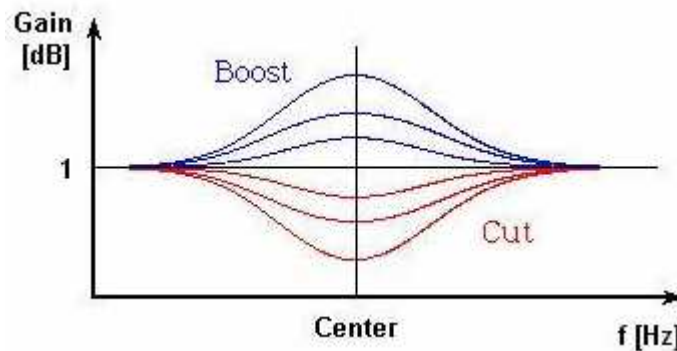
Obr. 7.4: Schematické znázornenie filtra IIR 2. rádu používaného v parametrických ekvalizéroch

Na (obr. 7.5) sú uvedené modulové kmitočtové charakteristiky dolnej a hornej prepusti pre filtre typu *shelving* a na (obr. 7.6) charakteristika pásmových prepustí typu *peak*. Prevod prenosovej funkcie analógového prototypu do roviny Z je vykonané pomocou bilinéárnej transformácie, z ktorej získame prenosovú funkciu v obecnom tvare

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}. \quad (7.2)$$



Obr. 7.5: Frekvenčná odozva filtrov typu shelving (a) dolnej prepusti (b) hornej prepusti



Obr. 7.6: Frekvenčná odozva filtru typu peak pásmovej prepusti

Koeficienty parametrický filtrov z prenosovej funkcie $H(z)$ sa počítajú podľa nasledovných vzťahov, ktoré sú (viz. tab. 7.1 a 7.2) [10].

| | b_0 | b_1 | b_2 | a_1 | a_2 |
|----------|---|--|---|--|---|
| LF boost | $\frac{1 + \sqrt{2V_0 m + V_0 m^2}}{1 + \sqrt{2m + m^2}}$ | $\frac{2(V_0 m^2 - 1)}{1 + \sqrt{2m + m^2}}$ | $\frac{1 - \sqrt{2V_0 m + V_0 m^2}}{1 + \sqrt{2m + m^2}}$ | $\frac{2(m^2 - 1)}{1 + \sqrt{2m + m^2}}$ | $\frac{1 - \sqrt{2m + m^2}}{1 + \sqrt{2m + m^2}}$ |
| LF cut | $\frac{1 + \sqrt{2m + m^2}}{1 + \sqrt{2V_0 m + V_0 m^2}}$ | $\frac{2(m^2 - 1)}{1 + \sqrt{2V_0 m + V_0 m^2}}$ | $\frac{1 - \sqrt{2m + m^2}}{1 + \sqrt{2V_0 m + V_0 m^2}}$ | $\frac{2(V_0 m^2 - 1)}{1 + \sqrt{2V_0 m + V_0 m^2}}$ | $\frac{1 - \sqrt{2V_0 m + V_0 m^2}}{1 + \sqrt{2V_0 m + V_0 m^2}}$ |
| HF boost | $\frac{V_0 + \sqrt{2V_0 m + m^2}}{1 + \sqrt{2m + m^2}}$ | $\frac{2(m^2 - V_0)}{1 + \sqrt{2m + m^2}}$ | $\frac{V_0 - \sqrt{2V_0 m + m^2}}{1 + \sqrt{2m + m^2}}$ | $\frac{2(m^2 - 1)}{1 + \sqrt{2m + m^2}}$ | $\frac{1 - \sqrt{2m + m^2}}{1 + \sqrt{2m + m^2}}$ |
| HF cut | $\frac{1 + \sqrt{2m + m^2}}{V_0 + \sqrt{2V_0 m + m^2}}$ | $\frac{2(m^2 - 1)}{V_0 + \sqrt{2V_0 m + m^2}}$ | $\frac{1 - \sqrt{2m + m^2}}{V_0 + \sqrt{2V_0 m + m^2}}$ | $\frac{2(m^2 - V_0)}{V_0 + \sqrt{2V_0 m + m^2}}$ | $\frac{V_0 - \sqrt{2V_0 m + m^2}}{V_0 + \sqrt{2V_0 m + m^2}}$ |

Tab. 7.1: Koeficienty číslicových filtrov shelving 2. rádu

kde $m = \text{tg}(\pi f_c / f_{vz})$, $V_0 = 10^{G/20}$ pre boost, $V_0 = 10^{-G/20}$ pre cut, $a_0 = 1$.

| | b_0 | b_1 | b_2 | a_1 | a_2 |
|------------|---|---|---|---|---|
| peak boost | $\frac{1 + \frac{V_0}{Q_\infty} m + m^2}{1 + \frac{1}{Q_\infty} m + m^2}$ | $\frac{2(m^2 - 1)}{1 + \frac{1}{Q_\infty} m + m^2}$ | $\frac{1 - \frac{V_0}{Q_\infty} m + m^2}{1 + \frac{1}{Q_\infty} m + m^2}$ | $\frac{2(m^2 - 1)}{1 + \frac{1}{Q_\infty} m + m^2}$ | $\frac{1 - \frac{1}{Q_\infty} m + m^2}{1 + \frac{1}{Q_\infty} m + m^2}$ |
| peak cut | $\frac{1 + \frac{1}{Q_\infty} m + m^2}{1 + \frac{V_0}{Q_\infty} m + m^2}$ | $\frac{2(m^2 - 1)}{1 + \frac{V_0}{Q_\infty} m + m^2}$ | $\frac{1 - \frac{1}{Q_\infty} m + m^2}{1 + \frac{V_0}{Q_\infty} m + m^2}$ | $\frac{2(m^2 - 1)}{1 + \frac{V_0}{Q_\infty} m + m^2}$ | $\frac{1 - \frac{V_0}{Q_\infty} m + m^2}{1 + \frac{V_0}{Q_\infty} m + m^2}$ |

Tab. 7.2: Koefficienty číslicových filtrov peak 2. rádu

7.2 Frekvenčne - časový modul

Obsahuje Teager Energy operátor a pásmovú prepust' s medznými frekvenciami 63 a 250 Hz.

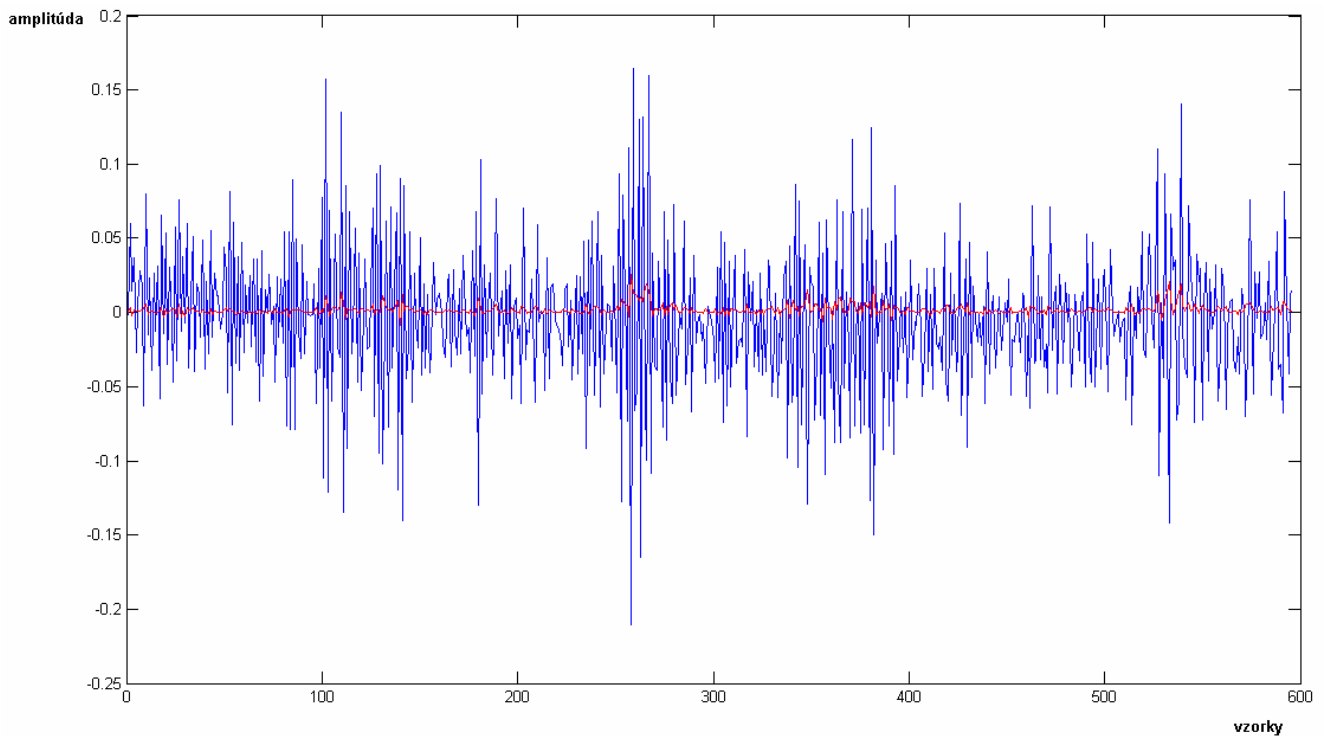
7.2.1 Teager Energy Operator

Teager energy operator E_n je výkonný nelineárny operátor, schopný extrahovať energiu signálu, založený na mechanických a fyzikálnych činiteľoch. Teager energy operator bol úspešne použitý v rôznych aplikáciách určených pre spracovanie rečového signálu ako nástroj na potlačenie šumu. Pre digitálny diskretný signál s_n je možné ho vyjadriť nasledujúcim vzťahom

$$E_n = s_n^2 - s_{n+1}s_{n-1}, \quad (7.2)$$

kde s_{n+1} je signál s_n posunutý o jednu vzorku dopredu a s_{n-1} dozadu. Na (obr. 7.7) je znázornený, signál písmena „s“ v čase a k nemu odpovedajúci priebeh po úprave Teager Energy operátorom. Je tu jasne viditeľný jeho účinok na neznelych, a zašumených úsekoch reči. V prostredí MATLAB sa realizuje nasledovne ako funkcia, ktorú si hlavný program po zavolaní spustí.

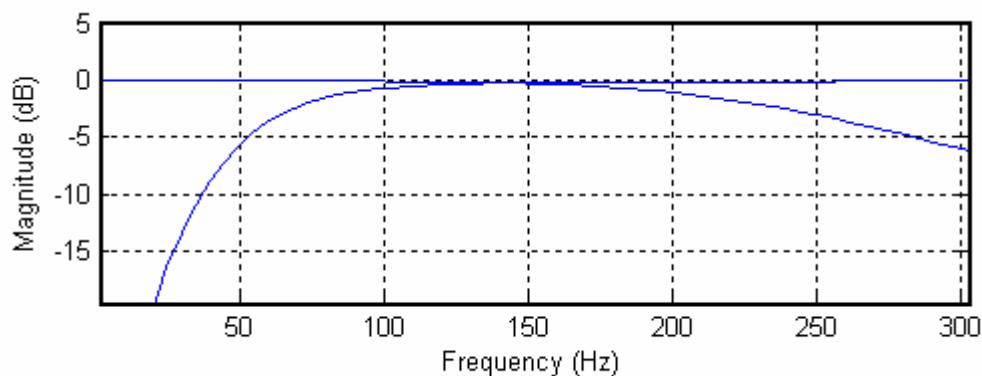
```
function Et = teagerEnergy(s)           % názov funkcie s vstupným a výstupným
                                        % parametrom
d=length(s);                           % výpočet dĺžky vstupného signálu
Et = zeros(d, 1);                       % vektor hodnôt do ktorých sa bude Et
                                        % ukladať
for n=2:d-1                             % cyklus pre výpočet Teager Energy operátor
    Et(n) = (s(n)^2)-(s(n+1)*s(n-1));
end
```



Obr. 7.7: Zobrazenie neznej hlásky „ s “ (modrou) a k nej vykreslený E_n (červenou)

7.2.2 Pásmová filtrácia (63 Hz, 250 Hz)

Pásmová prepust' 63 - 250 Hz je zložená z dvoch IIR filtrov s Butterworthovou aproximáciou (monotónne klesajúca funkcia). Skladá sa z dolnej prepusti 3. rádu s $f_d=250$ Hz a hornej prepusti 2. rádu s $f_h=63$ Hz. Tento filter má za úlohu potlačiť veľmi nízke frekvencie a odstrániť striedavú zložku vstupného signálu. Z (obr. 7.8) je viditeľné, že tieto filtre majú v medzných hodnotách frekvencie pokles o 3 dB.



Obr. 7.8: Frekvenčná odozva pásmovej prepusti 63-250 Hz

Realizácia v prostredí Matlab vyzerá nasledovne:

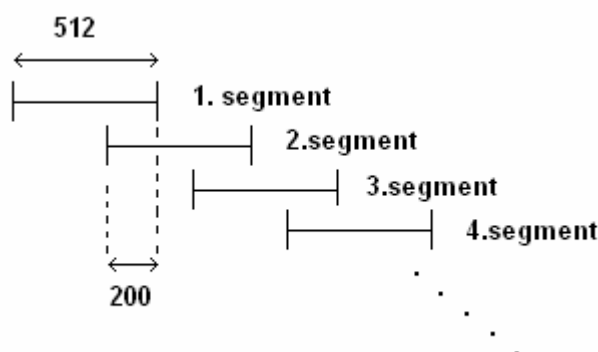
```
function BP = Bandpass (rec)      % názov funkcie s vstupným a výstupným parametrom
fs=20000;                       % vyzorkovacia frekvencia
fh=63;                           % medzný kmitočet filtra
fm=10000;
Wn=fh/fm;                        % normovaný medzný kmitočet - pokles o 3 dB
N=2;                              % rád filtra
[b,a]=butter(N,Wn,'high');       % funkcia na výpočet koeficientov
sig=rec;
out1=filter(b,a,sig);           % filtrácia vstupného signálu
fd=250;
Wn1=fd/fm;
N1=3;
[b1,a1]=butter(N1,Wn1);
out2=filter(b1,a1,out1);
BP=out2;
```

7.3 Modul výpočtu autokorelácie pre jednotlivé kanály

Tento modul má za úlohu aplikáciu autokorelácie na signál z každého pásma, ktoré vychádza z pásmovej prepusti. Tieto signály sa musia rozdeliť na menšie časti segmenty.

7.3.1 Segmentácia signálov

Podľa pokynov, ktoré boli prijaté aj s nahrávkami so vyzorkovacou frekvenciou $f_v = 20$ kHz na analýzu, sa mal signál nasegmentovať nasledovne. Dĺžka segmentu 25,6 ms a prekrytie dvoch nasledujúcich segmentov za sebou 10 ms. Pri vyzorkovacej frekvencii f_v bola dĺžka segmentu vypočítaná jednoduchým výpočtom, v ktorom sa vynásobí vyzorkovacia frekvencia s dĺžkou segmentu na 512 vzorkou s posunutím dĺžky 200 vzorkou.



Obr. 7.9: Proces segmentácie signálu s dĺžkou segmentu 512 a prekrytím 200 vzorkou

7.3.2 Aplikácia autokorelácie

Po segmentácii sa aplikuje autokorelácia ACR na každý segment zvlášť a to podľa vzťahu (5.3). Výstupom z tohto bloku je 12. matíc (12.pásiem) s rozmerom 512 (dĺžka segmentu) X Nseg - počet segmentov v signály. Súčtom týchto 12-tich matíc vznikne jedna pre celé frekvenčné pásmo.

Hľadanie základnej periódy sa aplikuje až na túto maticu v programe označenú SACR. V MATLABe sa funkcia autokorelácie zapíše nasledovne:

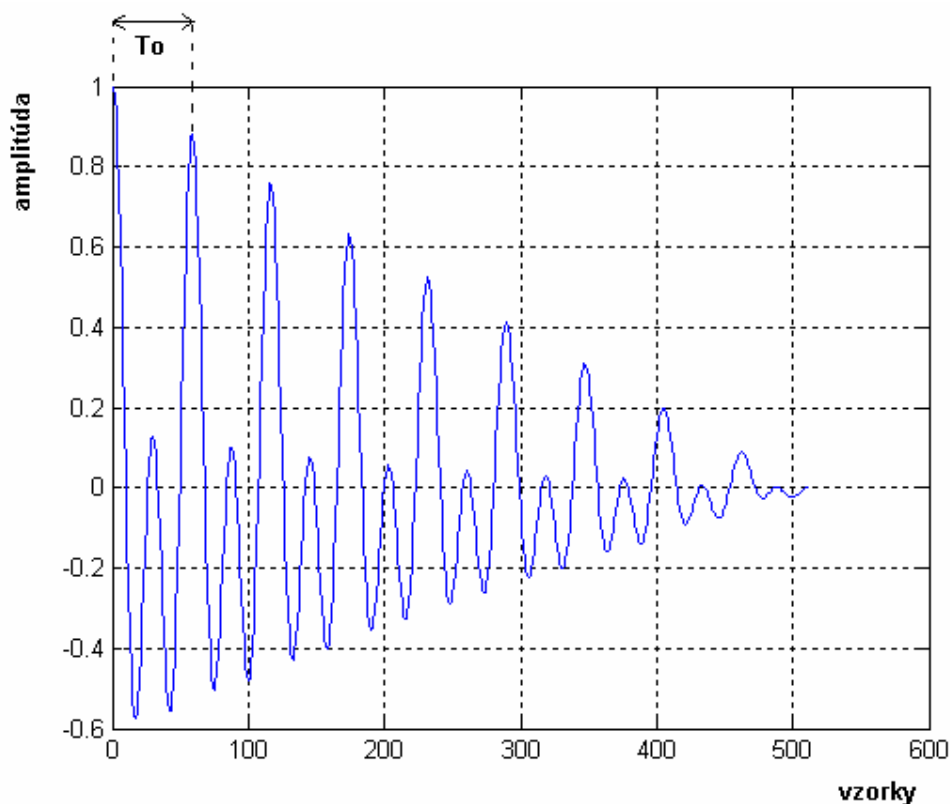
```
function ACR = AutoKorelace(rec); % funkcia autokorelácie  
  
rec = [zeros(156, 1); rec]; % synchronizácia výsledých hodnôt  $T_0$   
ACR = C_AutoCorr(rec, 512, 200); % autokorelácia v jazyku C
```

7.4 Modul určenia základnej periódy T_0 a určenie znelosti/neznelosti segmentov

Po funkcii autokorelácie sa hľadá základná perióda v každom segmente a z nej sa určuje nasledovným spôsobom znelosť/neznelosť segmentu. Nakoniec sa hodnoty T_0 algoritmu pásmovej filtrácie porovnávajú s referenčnými hodnotami, ktoré boli taktiež prijaté spolu s testovacími nahrávkami.

7.4.1 Určenie základnej periódy T_0

Na (obr. 7.10) je vykreslená autokorelácia jedného segmentu rečového signálu a naznačená základná perióda. Je z neho viditeľné, že určenie základnej periódy rečového signálu spočíva v nájdení lokálneho maxima funkcie v obmedzenom rozsahu. Pre naše nahrávky (mužské a ženské hlasy) je postačujúci frekvenčný rozsah 60 až 333 Hz. Pri vzorkovacej frekvencii $f_v = 20$ kHz vychádza rozsah vzorkou od 60 (333 Hz) po 333 (60 Hz).



Obr. 7.10: Autokorelácia znelého segmentu dĺžky 512 vzorkou

Realizácia určenia T_0 v prostredí MATLAB vyzerá nasledovne:

```
function chyba = PitchExtract (x,Nseg,SACR)

x = [zeros(156, 1); x];
ZCR = C_ZCrossing(x, 512, 200);           % počet prechodov nulou v C
for i=1:Nseg
    Td=60; Th=333;                        % určenie medzných kmitočtov
    [z,w]=max(SACR(Td:Th,i));             % hľadanie lokálneho maxima
    To(i)=(Td+w);
    pocatek = ((i-1)*200) + 1;
    konec = ((i-1)*200) + 512;
    seg = x(pocatek:konec);
    seg=seg';
    E = sum(seg.^2);                       % výpočet energie
    if (ZCR(i)>100)                         % určovanie znelý/neznelý zo ZCR
        To(i)=0;
    end
    if (E<0.005)                           % určovanie znelý/neznelý z E
        To(i)=0;
    end
    if (z<0.25)                             % určovanie znelý/neznelý z P
        To(i)=0;
    end
end
To=To';
ref = load('f5nw0000.pev', '-ascii'); % referenčné hodnoty To
R=ref;
chyba=0;
for i=1:Nseg                               % cyklus porovnávania To s R
    down=R(i)-5;
    up=R(i)+5;
    if (To(i)>up) || (To(i)<down)
        chyba=chyba+1;
    end
end
plot(To);                                  % vykreslenie To a R
hold on
plot(R, 'r');
hold off
title('Porovnanie hodnot R a To (Banky filtrov)')
xlabel('segmenty')
ylabel('To')
```

Tento algoritmus *Pitch Extract* má 3. hlavné úlohy:

- hľadanie lokálneho maxima funkcie autokorelácie v každom segmente
- určenie znelý/neznelý segment z 3. kritérií
- porovnanie hodnôt T_0 s referenčnými hodnotami R

1. Hľadanie lokálneho maxima sa v MATLABe vykonáva jednouchou funkciou $[z,w] = \max(a)$, kde a je vektor hodnôt vyšetrovanej funkcie, z je hodnota maxima na ose y a w je hodnota maxima na ose x .

2. Určenie znelosti segmentu je urobené nasledovne: najprv sa v danom segmente vypočíta počet prechodov nulou signálu podľa vzťahu (5.1). Ak je menší než zvolená hranica, hodnota T_0 pokračuje v cykle, ak je väčší, tak je $T_0 = 0$ t.j. neznelý. Potom sa zo signálu počíta energia (kvadrát) signálu. Ak je väčšia než zvolená hranica, hodnota T_0 pokračuje v cykle, ak je menšia tak je $T_0 = 0$. Nakoniec sa používa prahové rozhodnutie z hodnoty z . Postup je taký istý ako pri energii.

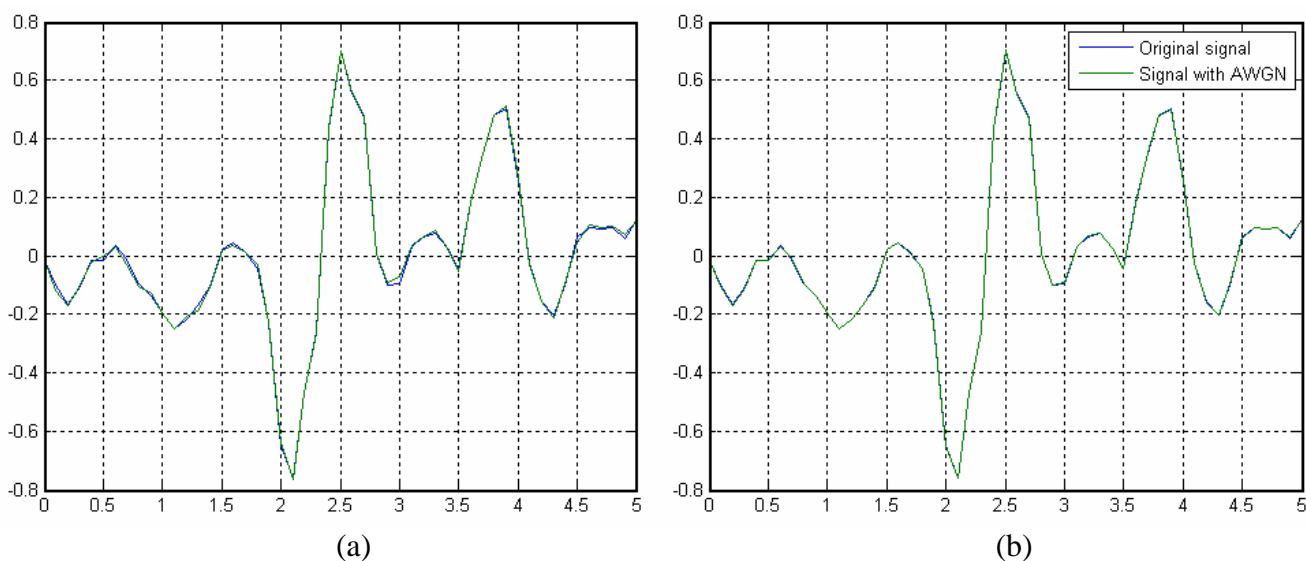
3. V tomto kroku sa porovnávajú hodnoty rečových signálov určené metódami autokorelácie a pásmovej filtrácie s referenčnými hodnotami R . Referenčné hodnoty R sú konkrétne čísla uložené v súbore s príponou `.pev`. Ich načítanie do MATLABu je vykonané pomocou funkcie **load('súbor.pev', '-ascii')**. Povolená akceptovateľná odchýlka bola stanovená na ± 5 . To znamená že poloha maxima funkcie v rozsahu 60 - 333 Hz sa mohla líšiť o ± 5 vzorkou na ose x . Každá odchýlka je ukladaná do premennej **chyba**, z ktorej hlavný program na konci vypočíta chybovosť metódy na celý signál.

Pre ľahšiu analýzu výsledkov a prípadné ladenie algoritmu sa na konci tejto funkcie nachádza vykreslenie hodnôt základnej periódy určené algoritmom s hodnotami referenčnými.

8. Dosiahnuté výsledky

8.1 Pomer signál / šum (SNR)

V tejto kapitole sú uvedené dosiahnuté výsledky analýzy chybovosti algoritmu pásmovej filtrácie PF a základnej metódy autokorelácie ACF pre dva rôzne pomery signálu k šumu. Jednak pre SNR = 30 dB a 40 dB. Graficky sú tieto pomery znázornené (viz. obr. 7.11). Zároveň sú výsledky PF a ACF porovnávané.



Obr. 8.1: Znelý úsek reči zašumený bielym šumom s hodnotu SNR (a) 30 dB (b) 40 dB

8.2 Analýza chybovosti

Testované nahrávky sú wav súbory vzorkované 20 kHz. Je medzi nimi 5 nahrávok s mužskými hlasmi a 5 nahrávok s ženskými hlasmi. Všetci rečníci rozprávajú rovnaký text s názvom „the north-wind story“.

| SNR = 30 dB | Chybovosť ACF [%] | | Chybovosť PF [%] | |
|-------------|-------------------|--------|------------------|--------|
| | muži | ženy | muži | ženy |
| 1 | 21,761 | 12,298 | 22,671 | 16,398 |
| 2 | 27,276 | 16,978 | 27,715 | 24,132 |
| 3 | 28,314 | 18,728 | 28,792 | 22,728 |
| 4 | 30,365 | 28,902 | 29,712 | 31,941 |
| 5 | 28,047 | 7,263 | 27,302 | 11,963 |

Tab. 8.1: Hodnoty chybovosti základnej metódy ACF a metódy PF pre SNR = 30 dB

| SNR = 40 dB | Chybovosť ACF [%] | | Chybovosť PF [%] | |
|-------------|-------------------|--------|------------------|--------|
| | muži | ženy | muži | ženy |
| 1 | 18,630 | 10,280 | 19,647 | 12,981 |
| 2 | 22,442 | 12,496 | 23,195 | 13,980 |
| 3 | 17,415 | 12,824 | 17,968 | 14,234 |
| 4 | 21,906 | 22,697 | 21,075 | 22,824 |
| 5 | 20,104 | 6,307 | 18,814 | 6,927 |

Tab. 8.2: Hodnoty chybovosti základnej metódy ACF a metódy PF pre SNR = 40 dB

Z (Tab. 8.1 a 8.2) je jasne vidieť, že navrhnutý a naprogramovaný algoritmus nefunguje dostatočne presne a spoľahlivo. Aj keď sú vo výsledkoch náznaky malého zlepšenia chybovosti, napr. v hodnotách s odstupom signál/šum 40 dB hodnota piatej mužskej nahrávky je 18,814, čo je zlepšenie o 1,3 %, vo väčšine prípadoch porovnania s ACF je algoritmus neúčinný.

8.3 Zhodnotenie a stanovenie príčin chybovosti

Z dosiahnutých výsledkov sa dá usúdiť že, obe metódy dosahovali pri nižších frekvenciách (mužské hlasy) menšiu presnosť než pri určovaní vyšších frekvencií (ženské hlasy), čo sa dalo čakať podľa teoretických predpokladov. Ďalej je jasne vidieť zvyšujúca sa presnosť oboch algoritmov po zvýšení hodnoty SNR.

Neúčinnosť algoritmu pásmovej filtrácie pri určovaní T_0 v rôznych rečových signáloch zašumených rôznym pomerom SNR bola spôsobená hlavne zlým návrhom banky filtrov a taktiež nezaradením bloku Channels Selection pre stredné a vyššie kmitočty uvedený v [11].

8.4 Možné vylepšenie algoritmu

Pre zvýšenie účinnosti PF algoritmu by sa mala banka filtrov navrhnuť s viacerými pásmami. Pre 20 kHz nahrávky aspoň 20 pásiem. Taktiež zvlnenie prenosovej charakteristiky banky filtrov bolo dosť veľké z dôvodu nízkej strmosti IIR filtrov 2. rádu typu *shelving* dolnej a hornej prepusti. Zmena strmosti u týchto filtrov sa dá urobiť len skokovo zmenou rádu filtra. Toho sa dá MATLABe dosiahnuť vynásobením prenosových funkcií dvoch filtrov pomocou funkcie $y = \text{conv}(a, b)$ na násobenie polynómov. Táto funkcia robí konvolúciu dvoch vektorov a pokiaľ sú tieto vektory koeficientmi polynómov, odpovedá ich konvolúcia násobeniu polynómov.

9. Záver

V bakalárskej práci boli oboznámené základné metódy odhadu základného tónu rečového signálu, ale aj s robustné metódy odolné voči šumu a jedna z nich bola navrhnutá a naprogramovaná v prostredí MATLAB. Z dosiahnutých výsledkov bolo vykonané hodnotenie chybovosti a porovnanie a výsledkov základnej metódy pomocou autokorelácie a metódy rozdelenia signálu do viacerých pásiem pomocou banky filtrov. Výsledky sú uvedené v tabuľkách.

Prvá časť sa venuje rečovému signálu, jeho vzniku a popisuje aj jeho základné vlastnosti. Ďalej je popísaný zjednodušený model vytvárania reči, ktorý je dôležitý na pochopenie vzniku rečového signálu. Ešte pred popisom základnej frekvencie je uvedený postup spracovania akustického signálu, ktorý je veľmi dôležitý pre následnú aplikáciu rôznych algoritmov, z ktorých sa získavajú požadované výsledky. Po oboznámení práca prechádza na výklad základnej frekvencie rečového signálu, jej závislosti na iných veličinách, popisuje aké sú základné problémy pri jej určovaní, jej vlastnosti a uvádza možnosti použitia. V kapitole 5. je uvedený podrobný rozbor jednotlivých základných metód na určovanie základnej frekvencie v časovej, frekvenčnej a keprálnej oblasti. Práca je zameraná hlavne na vysvetlenie ich základného princípu. Kapitola 6. obsahuje doplnenie teórie o dve robustné metódy odolné voči šumu a to o metódu RASTA a metódu rozloženia signálu do viacerých pásiem pomocou banky filtrov PF.

V časti realizácie je popísaný postup aký bol zvolený pri návrhu a realizácii metódy rozloženia signálu do viacerých pásiem pomocou banky filtrov, pracujúcej vo frekvenčnej oblasti v programovom prostredí MATLAB. Neskôr sú porovnávané výsledky odhadu T_0 naprogramovanej metódy so základnou metódou odhadu základnej frekvencie rečového signálu pomocou autokorelácie. Uvádza sa taktiež analýza chybovosti pre rôzne hodnoty SNR.

Aj keď je v dnešnej dobe modernej techniky už veľa komunikačných prostriedkov, rečový signál je stále ten základný od ktorého sa to všetko odvíja. Preto si myslím že, jeho skúmanie, spracovanie a vyhodnocovanie je pre ľudstvo veľmi dôležitá činnosť.

Od prvej analýzy rečového signálu už utieklo viac než polstoročie a vzniklo veľa algoritmov na odhad základnej frekvencie rečových signálov. Aj keď sa zdá táto oblasť spracovania rečového signálu už preskúmaná, nie je tomu tak. Rečový signál je veľmi variabilný a nestály a pre niektorých trochu tajuplný. Ako správny hudobník by som sa chcel aj naďalej venovať číslicovému spracovaniu akustických signálov.

Zoznam použitej literatúry

- [1] SIGMUND M.: Analýza řečových signálů. *VUT 2000*. ISBN 80-214-1783-8
- [2] PSUTKA J.: Komunikace s počítačem mluvenou řečí. *Academia Praha* 1995. ISBN 80-200-0203-0
- [3] ŠEBESTA V., SMÉKAL V.: Signály a soustavy. *VUT 2004*. ISBN 80-214-2434-6
- [4] SMÉKAL Z.: Číslicové zpracování signálů. Elektronická verze Skript *VUT 2007*
- [5] Pitch detection methods. [online]
Článek na stránke: <http://www-ccrma.stanford.edu/%7Eepdelac/154/m154paper.htm> >
- [6] VONDRA M., Kepstrální analýza řečového signálu. *elektrorevue* [online]. 2001 č. 48 [cit. 2001-07-11]. Článek na stránke: <http://www.elektrorevue.cz/clanky/01048/index.html> >
- [7] GERHARD D., Pitch Extraction and Fundamental Frequency: History and Current Techniques Technical Report. *University of Regina* [online] 2003 [cit. November 2003].
Článek na stránke: <http://www2.cs.uregina.ca/%7Egerhard/publications/TRdbg-Pitch.pdf> >
- [8] KÉPESI M., NAGY Z., Potlačení šumu pozadí řečového signálu mapováním spektrogramu. *elektrorevue* [online]. 2000 č. 37 [cit. 2000-06-10]. Článek na stránke: <http://www.elektrorevue.cz/clanky/00037/index.html> >
- [9] Mel scale
Článek na stránke: http://en.wikipedia.org/wiki/Mel_scale >
- [10] Jiří Schimmel, Návrh parametrických filtrů
Podklad pre cvičenie k predmetu - Číslicové zpracování signálů (MCSI) *VUT 2007*
- [11] ROUAT J., LIU Y.Ch., MORISSETE D.: A pitch determination and voiced/unvoiced decision algorithm for noisy speech. *Speech communication* 21, 1997, pp. 191-207.

Zoznam príloh

| | |
|--|----|
| PRÍLOHA 1: Hlavný program | 38 |
| PRÍLOHA 2: Cyklus návrhu filtrov <i>peak</i> | 42 |

Príloha 1: Hlavný program

```
% Návrh 12-pasmovej Mel-banky filtrov.
% x - vstupný signál
% fc - vektor medzných kmitočtov filtrov [LFS Peaks HFS]
% G - vektor zesilnenia filtra v dB LFS Peak HFS]
% Q - kvalita filtrov Peak
% fvz - vzorkovací kmitocet

function [chybovost chybovost1]= filterbank (fc, G, Q, fvz, x)

fc = [178 291 544 861 1258 1757 2384 3171 4158 5397 6952 7817];
G = [10 10 10];
Q = [3.3 3.8 4.6 5.3 6.9 7.5 6.2 5.5 5.3 3.8];
fvz = 20000;

home;
help(mfilename);
% ak nie je zadaný vstupný signál, otvorí sa dialog pre načítanie súboru
if (nargin<4),
    [filename, pathname] = uigetfile( ...
        {'*.wav', 'Zvukove soubory wav';...
        '.*', 'Vsechny soubory'},...
        'Vyber zvukovy soubor');
    [x, fvz, bits] = wavread( fullfile( pathname, filename));
end;
[N,channels] = size(x);
y=awgn(x,40); % pridanie bieleho šumu do vstupného signálu
x=y;

m = tan(pi*fc(1)/fvz); % výpočet koeficientov prenosovej funkcie
if (G(1)>0),
    V0 = 10^(G(1)/20); % low-frequency shelving boost
    den = (1 + sqrt(2)*m + m^2);
    b0 = (1 + sqrt(2*V0)*m + V0*m^2) / den;
    b1 = (2*(V0*m^2 - 1)) / den;
    b2 = (1 - sqrt(2*V0)*m + V0*m^2) / den;
    a1 = (2*(m^2 - 1)) / den;
    a2 = (1 - sqrt(2)*m + m^2) / den;
elseif (G(1)<0),
    V0 = 10^(-G(1)/20); % low-frequency shelving cut
    den = (1 + sqrt(2*V0)*m + V0*m^2);
    b0 = (1 + sqrt(2)*m + m^2) / den;
    b1 = (2*(m^2 - 1)) / den;
    b2 = (1 - sqrt(2)*m + m^2) / den;
    a1 = (2*(V0*m^2 - 1)) / den;
    a2 = (1 - sqrt(2*V0)*m + V0*m^2) / den;
else
    b0 = 1; % low-frequency shelving allpass
    b1 = 0;
    b2 = 0;
    a1 = 0;
    a2 = 0;
end;
blfs = [b0 b1 b2];
alfs = [1 a1 a2];

[bpeak,apeak] = peak(fc,G,Q,fvz);
```

```

m = tan(pi*fc(12)/fvz);
if (G(3)>0),
    V0 = 10^(G(3)/20);           %% high-frequency shelving boost
    den = (1 + sqrt(2)*m + m^2);
    b0 = (V0 + sqrt(2*V0)*m + m^2) / den;
    b1 = 2*(m^2 - V0) / den;
    b2 = (V0 - sqrt(2*V0)*m + m^2) / den;
    a1 = (2*(m^2-1)) / den;
    a2 = (1 - sqrt(2)*m + m^2) / den;
elseif (G(3)<0),

    V0 = 10^(-G(3)/20);         %% high-frequency shelving cut
    den = (V0 + sqrt(2*V0)*m + m^2);
    b0 = (1 + sqrt(2)*m + m^2) / den;
    b1 = (2*(m^2 - 1)) / den;
    b2 = (1 - sqrt(2)*m + m^2) / den;
    a1 = (2*(m^2 - V0)) / den;
    a2 = (V0 - sqrt(2*V0)*m + m^2) / den;
else
    b0 = 1;                     %% high-frequency shelving allpass
    b1 = 0;
    b2 = 0;
    a1 = 0;
    a2 = 0;
end;
bhfs = [b0 b1 b2];
ahfs = [1 a1 a2];

x1 = filter(blfs, alfs, x);     %% filtrácia kanálov a aplikácia výpočtu ACR
BP1 = Bandpass (x1);
ACR1 = AutoKorelace(BP1);
x2 = filter(bpeak(1:3), apeak(1:3), x1);
BP2 = Bandpass (x2);
ACR2 = AutoKorelace(BP2);
x3 = filter(bpeak(4:6), apeak(4:6), x2);
BP3 = Bandpass (x3);
ACR3 = AutoKorelace(BP3);
x4 = filter(bpeak(7:9), apeak(7:9), x3);
BP4 = Bandpass (x4);
ACR4 = AutoKorelace(BP4);
x5 = filter(bpeak(10:12), apeak(10:12), x4);
BP5 = Bandpass (x5);
ACR5 = AutoKorelace(BP5);
x6 = filter(bpeak(13:15), apeak(13:15), x5);
BP6 = Bandpass (x6);
ACR6 = AutoKorelace(BP6);
x7 = filter(bpeak(16:18), apeak(16:18), x6);   %% + aplikácia Teager Energy
Et7 = teagerEnergy(x7);
BP7= Bandpass (Et7);
ACR7 = AutoKorelace(BP7);
x8 = filter(bpeak(19:21), apeak(19:21), x7);
Et8 = teagerEnergy(x8);
BP8= Bandpass (Et8);
ACR8 = AutoKorelace(BP8);
x9 = filter(bpeak(22:24), apeak(22:24), x8);
Et9 = teagerEnergy(x9);
BP9 = Bandpass (Et9);
ACR9 = AutoKorelace(BP9);
x10 = filter(bpeak(25:27), apeak(25:27), x9);
Et10 = teagerEnergy(x10);
BP10 = Bandpass (Et10);

```

```

ACR10 = AutoKorelace(BP10);
x11 = filter(bpeak(28:30), apeak(28:30), x10);
Et11 = teagerEnergy(x11);
BP11 = Bandpass (Et11);
ACR11 = AutoKorelace(BP11);
x12 = filter(bhfs, ahfs, x11);
Et12 = teagerEnergy(x12);
BP12 = Bandpass (Et12);
ACR12 = AutoKorelace(BP12);

Nseg = 2+ floor((length(x) - 512)/200);           %% počet segmentov v signály
ACR = zeros(512, Nseg, 12);

ACR(:, :, 1) = ACR1;
ACR(:, :, 2) = ACR2;
ACR(:, :, 3) = ACR3;
ACR(:, :, 4) = ACR4;
ACR(:, :, 5) = ACR5;
ACR(:, :, 6) = ACR6;
ACR(:, :, 7) = ACR7;
ACR(:, :, 8) = ACR8;
ACR(:, :, 9) = ACR9;
ACR(:, :, 10) = ACR10;
ACR(:, :, 11) = ACR11;
ACR(:, :, 12) = ACR12;

SACR = sum(ACR, 3);                               %% súčet všetkých ACR

chyba = PitchExtract (x,Nseg,SACR);              % aplikácia funkcie na výpočet To PF
chybal = Autokorelaci(x,Nseg);                   %% aplikácia funkcie na výpočet To ACF
chybovost = (chyba/Nseg)*100                      %% výpočet chybovosti pre PF a ACF
chybovost1 = (chybal/Nseg)*100

```

Príloha 2: Cyklus návrhu filtrov *peak*

```
function [bpeak,apeak]=bandpass(fc,G,Q,fvz)    %%funkcia návrhu filtrov peak

j=1;
for i=2:11
m = tan( pi*fc(i)/fvz);
if (G(2)>0),
    V0 = 10^(G(2)/20);                %% frequency peak boost
    den = (1 + (1/Q(i-1))*m + m^2);
    b0 = (1 + (V0/Q(i-1))*m + m^2) / den;
    b1 = (2*(m^2 - 1)) / den;
    b2 = (1 - (V0/Q(i-1))*m + m^2) / den;
    a1 = (2*(m^2 - 1)) / den;
    a2 = (1 - (1/Q(i-1))*m + m^2) / den;
elseif (G(2)<0),
    V0 = 10^(-G(2)/20);                %% frequency peak cut
    den = (1 + (V0/Q(i-1))*m + m^2);
    b0 = (1 + (1/Q(i-1))*m + m^2) / den;
    b1 = (2*(m^2 - 1)) / den;
    b2 = (1 - (1/Q(i-1))*m + m^2) / den;
    a1 = (2*(m^2 - 1)) / den;
    a2 = (1 - (V0/Q(i-1))*m + m^2) / den;
else
    b0 = 1;                            %% peak allpass
    b1 = 0;
    b2 = 0;
    a1 = 0;
    a2 = 0;
end;
bpeak_pomoc(j) = b0; % uloženie hodnôt koeficientov do pomocnej premennej
bpeak_pomoc(j+1) = b1;
bpeak_pomoc(j+2) = b2;
apeak_pomoc(j) = 1;
apeak_pomoc(j+1) = a1;
apeak_pomoc(j+2) = a2;
j=j+3;
end

bpeak = bpeak_pomoc;
apeak = apeak_pomoc;
```