

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

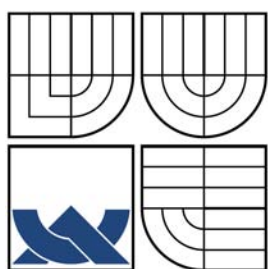
MODERNÍ METODY KÓDOVÁNÍ ŘEČI PRO PŘENOS PO
TELEKOMUNIKAČNÍ SÍTI

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

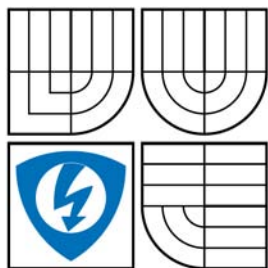
AUTOR PRÁCE
AUTHOR

MAREK ZAJÍČEK

BRNO 2009



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

MODERNÍ METODY KÓDOVÁNÍ ŘEČI PRO PŘENOS PO TELEKOMUNIKAČNÍ SÍTI

MODERN METHODS OF SPEECH CODING FOR TELECOMMUNICATION
NETWORK TRANSMISSION

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

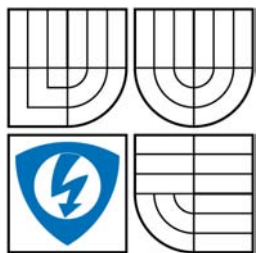
AUTOR PRÁCE
AUTHOR

MAREK ZAJÍČEK

VEDOUcí PRÁCE
SUPERVISOR

PROF. ING. ZDENĚK SMÉKAL, CSC.

BRNO 2009



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Marek Zajíček

ID: 70361

Ročník: 3

Akademický rok: 2008/2009

NÁZEV TÉMATU:

Moderní metody kódování řeči pro přenos po telekomunikační síti

POKYNY PRO VYPRACOVÁNÍ:

Pro přenos po mobilní síti se používají moderní postupy kódování řeči využívající lineární predikci. Jsou využívány dva systémy kódování s proměnou přenosovou rychlostí typu AMR-WB (Adaptive Multi-Rate Wide-Band) a VMR-WB (Variable Multi-Rate Wide-Band). Cílem bakalářské práce je uvést popis základních kódovacích principů, podrobně popsat vlastnosti obou systému, jejich výhody, nevýhody a způsob použití. Vybrané algoritmy je nutné ověřit v programu Matlab.

DOPORUČENÁ LITERATURA:

- [1] BESSETTE, B. et al.: The Adaptive Multirate Wideband Speech Codec (AMR-WB). In IEEE Trans. on Speech and Audio Processing, Vol. 10, No.8, November 2002, pp 620-363.
- [2] AHMADI, S., JELINEK, M.: On the Architecture, Operation, and Applications of VMR-WB: The New CDMA2000 Wideband Speech Coding Standard. In IEEE Communication Magazine, May 2006, pp.74-81.

Termín zadání: 9.2.2009

Termín odevzdání: 2.6.2009

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

prof. Ing. Kamil Vrba, 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 § 152 trestního zákona č. 140/1961 Sb.

ABSTRAKT

Tato bakalářská práce pojednává o použití kodeku AMR-WB (Adaptive Multi-Rate Wide-Band) v telekomunikačních sítích, zejména však v mobilních sítích a VoIP technologiích. Postupně popisuje základní principy kódování hlasu a poté jsou podrobně rozebrány vlastnosti tohoto systému, zejména jeho výhody, nevýhody a možné způsoby použití v praxi. Tato bakalářská práce byla zpracována pomocí aplikace $\text{\LaTeX}$.

KLÍČOVÁ SLOVA

komprese zvuku, vzorkovací frekvence, lineární predikce, širokopásmový hovor, algebraická kódová kniha, algoritmus

ABSTRACT

This bachelor's thesis describes the usage of AMR-WB (Adaptive Multi-Rate Wide-Band) codec in telecommunication networks, especially in mobile networks and VoIP technologies. There are also described basic principles of coding and resolved characteristics of this system in detail, especially advantages, disadvantages and the way of practical usage of codecs. This bachelor's thesis was processed by using $\text{\LaTeX}$ application.

KEYWORDS

voice compression, sampling frequency, linear prediction, wide-band speech, algebraic codebook, algorithm

Zajíček M. *Moderní metody kódování řeči pro přenos po telekomunikační síti*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2009. Počet stran 57. Vedoucí bakalářské práce: prof. Ing. Zdeněk Smékal, CSc. .

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Moderní metody kódování řeči pro přenos po telekomunikační síti“ 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)

Tímto bych chtěl poděkovat vedoucímu práce panu prof. Ing. Zdeňku Smékalovi, CSc. za ochotu, trpělivost a cenné rady při vypracování této bakalářské práce. Dále bych chtěl poděkovat za morální podporu svým příbuzným a přátelům a mým spolupracovníkům za pochopení při řešení problematiky této práce.

OBSAH

Úvod	13
1 Kódování řeči	14
1.1 Úvod	14
1.2 Metody kódování	15
1.2.1 Kódování vzorků časového průběhu hovorového signálu	15
1.2.2 Kódování parametrů zdroje hovorového signálu	17
1.2.3 Hybridní metody kódování	17
2 Moderní metody kódování řeči	19
2.1 Kodek AMR-WB (Adaptive Multi-Rate Wide-Band)	19
2.1.1 Úvod	19
2.1.2 Vývoj kódování řeči v systémech GSM a WCDMA	20
2.1.3 Algoritmický popis kodéru	22
2.1.4 Algoritmický popis dekodéru	34
2.1.5 Výpočetní náročnost kodeku	38
2.1.6 Adaptivní provoz AMR-WB v GSM kanálech	38
2.1.7 Detekce hlasové aktivity a funkce nesouvislého přenosu	38
2.1.8 Kvalita hovoru	39
3 Simulace v aplikaci Simulink	41
3.0.9 Úvod	41
3.0.10 Simulace modulu pre-processing	41
3.0.11 Simulace LPC analýzy	43
3.0.12 Návrh modulu horní propusti	45
3.0.13 Porovnání kódovacích technik PCM a CELP v prostředí Si- mulink	47
4 Závěr	51
Literatura	52
Seznam symbolů, veličin a zkratk	53
A Bloková schémata AMR-WB	55
B Obsah CD	57
B.0.14 Elektronické dokumenty	57
B.0.15 Zdrojové kódy aplikace MATLAB	57

B.0.16 Zdrojové kódy aplikace Simulink	57
--	----

SEZNAM OBRÁZKŮ

1.1	Blokové schéma PCM modulace.	16
1.2	Principiální schéma DPCM modulace.	16
3.1	Blokové schéma modulu podvzorkování a filtrace.	41
3.2	Blokové schéma modulu předzpracování kodeku AMR-WB.	42
3.3	Frekvenční průběh vstupního signálu.	42
3.4	Frekvenční průběh předzpracovaného signálu.	43
3.5	Blokové schéma analýzy lineární predikce.	44
3.6	Grafické zobrazení výstupu LP analýzy.	44
3.7	Zobrazení nulových bodů a pólů na jednotkové kružnici.	46
3.8	Zobrazení impulsní charakteristiky.	46
3.9	Zobrazení modulové a fázové frekvenční charakteristiky.	47
3.10	Blokové schéma zobrazení průběhů signálů u kódování PCM.	48
3.11	Časový průběh vstupního originálního signálu u kódování PCM.	48
3.12	Časový průběh výstupního dekódovaného signálu u kódování PCM.	48
3.13	Blokové schéma zobrazení průběhů signálů u kódování CELP.	49
3.14	Časový průběh vstupního originálního signálu kódování CELP.	49
3.15	Časový průběh výstupního dekódovaného signálu kódování CELP.	49
A.1	Blokové schéma AMR-WB ACELP kodéru.	55
A.2	Blokové schéma AMR-WB ACELP dekodéru.	56

SEZNAM TABULEK

1.1	Porovnání přenosových rychlostí dle typu kodeku.	17
2.1	Srovnání kodeků dle MOS a komprese.	21
2.2	Přiřazení algebraických kódových knih k bitovým rychlostem.	23
2.3	Bitová alokace pro jednotlivé módy kodeku AMR-WB.	24

ÚVOD

Tato bakalářská práce je zaměřena na moderní metody kódování řeči v telekomunikačních a konvergovaných sítích, konkrétně v mobilních sítích třetí generace a VoIP hlasových přenosech pomocí širokopásmového kódování.

Úvodní část je věnována seznámení se s problematikou kódování řeči, jejími základními principy a historickým vývojem. Budou popsány v současné době používané, ale i již méně efektivní a svým způsobem zastaralé metody komprese hlasu a obrazu. V dalších částech bude podrobně rozebrána problematika kódování metodou AMR-WB, jednotlivé její části včetně matematických definic a také výhody, nevýhody a praktické způsoby použití této metody.

Tato práce je zaměřena na analýzu teoretických poznatků a jejich podrobný rozbor, sloužící k pochopení problematiky výše uvedeného tématu. Na základě toho bude určena oblast možného praktického použití, přičemž bude brán zřetel na vlastnosti tohoto kodeku. Vybrané části budou poté prakticky ověřeny pomocí aplikace MATLAB v části věnované simulaci v prostředí Simulink.

1 KÓDOVÁNÍ ŘEČI

1.1 Úvod

Vlastnosti hovorového signálu

Základní tón řeči tzv. pitch periode je důležitým parametrem při analýze řečového signálu. Kmitočet hlasivek F_0 je charakterizován základním tónem lidského hlasu (pitch periode), který tvoří základ znělých zvuků (tj. samohlásek a znělých souhlásek). Kmitočet základního tónu je různý u dětí, dospělých, mužů i žen, pohybuje se většinou v rozmezí 80 až 400 Hz.

Sdělení zprostředkované řečovým signálem je diskrétní, tzn. může být vyjádřeno ve tvaru posloupnosti konečného počtu symbolů. Každý jazyk má vlastní množinu těchto symbolů - hlásek, většinou 30 až 50. Hlasy řeči můžeme rozdělit na znělé (n, e, ...), neznělé (š, č, ...) a jejich kombinace. Znělá hláska představuje kvaziperiodický průběh signálu, neznělá pak signál podobný šumu. Energie znělých hlásek je větší než neznělých. Krátký časový úsek znělé hlásky můžeme charakterizovat její formantovou strukturou. Formant je vlastně rezonance, která představuje spektrální obal řečového signálu.

Jemná harmonická struktura představuje chvění hlasivek. Lidská řeč je souvislý časově proměnný proces, z toho plyne i náročnost popisu lidské řeči a jejího modelování. Hovorový signál snímáme mikrofonom, který převádí modulovaný proud vzduchu na elektrický signál.

Dle [7] jsou požadavky na kódování:

- co nejmenší počet bitů
- co největší kvalita
- co nejmenší zpoždění
- co největší odolnost proti chybám
- co nejmenší výpočetní náročnost

Dělení dle digitálního zpracování řeči:

- kódování vzorků časového průběhu hovorového signálu - WC (Wave Coding).
Mezi tyto postupy patří PCM (Pulse Code Modulation), DM (Delta Modulation), ADM (Adaptive Delta Modulation), DPCM (Diferencial Pulse Code Modulation), ADPCM (Adaptive Diferencial Pulse Code Modulation).

- kódování parametrů zdroje hovorového signálu - SC (Source Coding). Mezi základní metody patří vokodérové metody spočívající na principu lineární predikce.
- hybridní metody kódování. Do této kategorie patří kodeky založené na adaptivních predikčních metodách APC (Adaptive Prediction Coding), kodeky s multiimpulsním MPE-LP či regulárním buzením RPE-LP (Regular Pulse Excitation Long Prediction) nebo kodeky využívající vektorové kvantizace CELP (Code Excited Linear Prediction, LD-CELP. Dále mezi hybridní kodéry patří metody založené na složkovém kódování nebo adaptivním transformačním kódování ATC (Adaptive Transform Coding) využívající rychlé diskrétní transformace (DFT, DCT). Cílem kompresních metod je dosáhnout co nejnižší přenosové rychlosti při zachování dobré srozumitelnosti, barvy hlasu mluvčího a přiměřené výpočetní náročnosti použitých algoritmů.

1.2 Metody kódování

1.2.1 Kódování vzorků časového průběhu hovorového signálu

PCM - pulsní kódová modulace

PCM kódování je tvořeno třemi kroky:

Vzorkování - hodnoty spojitého analogového signálu u_x se odečítají v diskrétním čase. Vzorkovací kmitočet vychází z Shannon-Kotelnikova vzorkovacího teorému, kdy vzorkovací kmitočet musí mít minimálně dvojnásobnou velikost než kmitočet vzorkovaný.

$$F_s > 2F_m \quad (1.1)$$

Kvantování - ke každému vzorku získanému ve výše uvedeném kroku se přiřadí jedna z možných diskrétních úrovní. U tohoto kódování je důležité rozložení kvantovacích hladin. Nejjednodušší je lineární (rovnoměrné) kvantování, které se používá při kódování zvuku u CD nebo zvukových wav souborů. Rozložení těchto hladin však není přijatelné při kódování malých signálů a řeči. Proto je výhodnější rozložit tyto kvantovací hladiny nerovnoměrně: hustější jsou určeny pro malé amplitudy a řidší pak pro velké amplitudy. Tento princip využívá logaritmická PCM, která používá dva druhy aproximací:

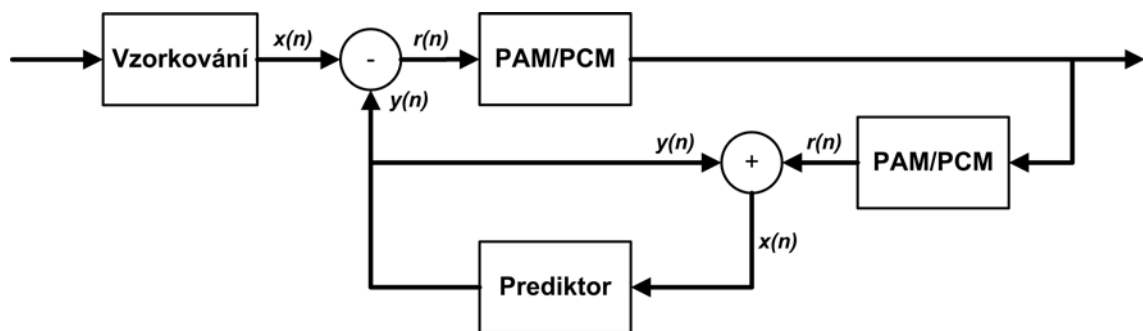
- A-law - používaná v Evropě
- μ -law - používaná v USA

Kódování - diskrétní hodnoty vzorků jsou reprezentovány pomocí n -bitových čísel. n se vybírá jako mocnina čísla 2 podle toho, jaké rozlišení požadujeme. PCM je oproti klasickým analogovým metodám odolnější vůči rušení přenášeného digitálního signálu a je schopná mnohonásobné obnovy při přenosu digitálního signálu. Její nevýhodou jsou větší nároky na šířku kmitočtového pásma a též vznik kvantizačního zkreslení v důsledku rozdílu mezi skutečnou hodnotou vzorku a výslednou kvantovanou hodnotou, přiřazenou danému vzorku. Při využití telekomunikačního kanálu s šířkou 4 kHz a vzorkovacím kmitočtu 8 kHz je přenosová rychlost pro jeden telefonní kanál 64 kbit/s. Tento postup odpovídá lineárnímu kvantování.



Obr. 1.1: Blokové schéma PCM modulace.

DPCM - diferenční pulsní kódová modulace



Obr. 1.2: Principiální schéma DPCM modulace.

Jestliže existuje závislost mezi jednotlivými vzorky, je zde možnost pokusit se odhadnout hodnotu současného vzorku z několika předcházejících. Toho se využívá u systému DPCM, kdy jsou kódovány rozdíly mezi okamžitou hodnotou vzorku signálu v daném vzorkovacím okamžiku a hodnotou predikovanou z předchozích vzorků. Pokud je odhad dobrý, pak je pro zakódování rozdílu zapotřebí nižšího počtu bitů, což vede k redukci přenosové rychlosti a tedy i množství přenášených dat se snižuje. DPCM se využívá především při kódování obrazu.

Standard	Algoritmus	Bitová rychlost kbit/s	Šířka pásma	Zpoždění
G.711	PCM	48, 56, 64	3 kHz	<< 1ms
G.722	ADPCM	48, 56, 64	7 kHz	< 2ms
G.723.1	MP-MLQ/ACELP	5,3 6,5	3 kHz	67 až 97 ms
G.728	LD-CELP	16	3 kHz	<< 2ms
G.729	ACELP	8	3 kHz	25 až 35 ms
G.729.A	ACELP	8	3 kHz	25 až 35 ms

Tab. 1.1: Porovnání přenosových rychlostí dle typu kodeku.

ADPCM - Adaptivní diferenční pulsní kódová modulace

Tento způsob kódování vychází z DPCM. Je vylepšen tak, že srovnávací průběh je vytvářen adaptivně a přizpůsobuje se konkrétním statistickým parametrům řeči. Výsledkem je ještě menší dynamický rozsah než v případě DPCM a tedy opět potřebujeme méně bitů k zakódování vzorku. Toto kódování je použito v normě G.721, kdy je logaritmická PCM komprimována na poloviční hodnotu tedy 32 kbit/s.

1.2.2 Kódování parametrů zdroje hovorového signálu

LP - lineární predikce

Při zdrojovém kódování se zdroj (lidský vokální trakt) modeluje elektronickým zařízením – vokodérem (voice coder). Ve vysílací části vokodéru se na základě hovorového signálu analyzují a kódují (pomocí tzv. lineární predikce) parametry zdroje tzv. deskriptory, které se po té přenášejí na přijímací stranu, kde jsou použity k řízení hovorového syntezátoru. Syntezátor je v podstatě filtr jehož parametry a buzení se mění na základě analyzovaného zdroje. Filtr se označuje jako LPF (linear prediction filter). Tento způsob je účinnější než kódování tvaru vlny, ale kvalita signálu je menší.

RELp - lineární predikce s regulárním buzením

Jedná se o typ kodéru s kódováním zbytkového rozdílového signálu po lineární predikci, který umožňuje zlepšení kvality hlasu u syntézy s lineární predikcí. Pro buzení filtru syntezátoru na přijímací straně se používá přenesený (zakódovaný) zbytkový signál mezi originálním signálem a syntetickým signálem.

1.2.3 Hybridní metody kódování

Abychom dosáhli lepší kvality hlasu u syntézy na základě lineární predikce, je budící signál vybírán tak, aby syntetizovaný signál byl co nejblíže původnímu hovorovému

signálu.

CELP - lineární predikce s buzením kódové knihy

Jedná se o lineárně predikční kodeky s vektorovým kvantováním budícího signálu. Jejich přenosová rychlost se pohybuje od 4 do 8 kbit/s. V pamětech kodéru a dekodéru jsou uloženy v kódových knihách soubory možných posloupností vstupního signálu. Je-li v kódové knize nalezena posloupnost odpovídající budícímu signálu, je příslušná adresa budící posloupnosti binárně přenesena do přijímače. Přijímač pak podle přijaté adresy generuje na základě shodné kódové knihy příslušnou budící posloupnost pro LP hlasový syntezátor. Kvalita řeči u CELP kodérů může být zlepšena následnou filtrací syntetizované řečové posloupnosti s cílem zdůraznit strukturu řeči z hlediska formantů a dominantních kmitočtů.

LD-CELP - lineární predikce s buzením kódové knihy s krátkou dobou zpoždění

Tento způsob kódování je specifický krátkou dobou zpoždění a je založen na zpětnovazebním řízením adaptace s kódovacím zpožděním do 2 ms a přenosovou rychlostí 16 kbit/s. Principiálně se od kodeků CELP liší v řešení vlastního hlasového syntezátoru a způsobu jeho řízení. U kodeků CELP se přenášejí v časovém multiplexu lineárně predikční koeficienty vyčíslené v kodéru krátkodobou (STP) a dlouhodobou analýzou (LTP), které slouží v přijímači k periodickému nastavování LPC filtrů hlasového syntezátoru. Naproti tomu mezi kodérem a dekodérem LD-CELP se tato data nepřenášejí a jak kodér, tak dekodér jsou řízeny zpětnovazebně. Nastavení budící úrovně se provádí pro každý blok vzorků. V LD-CELP dekodéru se nastavení úrovně a koeficienty LPC filtru nastavují obdobně jako u kodéru. Převodník kódu převádí blok vzorků odpovídající vektoru syntetického signálu na odpovídající PCM kód podle kompresní charakteristiky A nebo μ .

ACELP - lineární predikce s buzením algebraické kódové knihy

Algoritmus, který je používán kodérem ke generování pravděpodobnostní součásti buzení, kde budící vektor obsahuje malý počet nenulových pulsů s předdefinovanými prokládanými sadami pozic. Pulsy jsou fixovány s amplitudou od 1 do -1 a každý puls má sadu možných pozic odlišných od ostatních pulsů. Tato sada je prokládána a budící kód je identifikovaný pozicemi jeho nenulových pulsů. Vyhledávání kódové knihy je v podstatě vyhledávání optimální pozice nenulových bodů.

Praktické využití je v kodeku GSM Enhanced Full Rate ETSI 06.60.

2 MODERNÍ METODY KÓDOVÁNÍ ŘEČI

2.1 Kodek AMR-WB (Adaptive Multi-Rate Wide-Band)

2.1.1 Úvod

Současné mobilní sítě druhé a třetí generace pracují s úzkou šířkou hovorového pásma omezeného na 300-3400 Hz. Tyto bezdrátové systémy se neustále vyvíjejí, nejdříve byly používány hlavně pro přenos hlasu a postupně přecházely také k používání přenosu dat, multimedií a vysokorychlostních datových služeb. Zavedení rozšířené hovorové šířky pásma od 50 do 7000 Hz umožňuje podstatně zlepšenou hovorovou kvalitu a přirozenost hlasu. Ve srovnání s úzkým hovorovým pásmem přispívá nízkofrekvenční rozšíření od 50 do 200 Hz k lepšímu vykreslení spektra v okolí prvního formantu (základního tónu). Vysokofrekvenční rozšíření od 3,4 kHz do 7 kHz poskytuje lepší vykreslení neznělých hlásek a proto lepší srozumitelnost. Celková šířka pásma od 50 do 7000 Hz zlepšuje srozumitelnost, dává pocit přirozené řeči a jednodušší rozpoznání na výstupu reproduktoru.

Současné pokusy v hovorovém kódování prováděly širokopásmové kódování jež bylo uskutečnitelné v přenosových rychlostech aplikovatelných v mobilních komunikacích. Od roku 1999 sdružení 3GPP a ETSI uskutečňovalo vývoj a standardizaci širokopásmového hovorového kodeku pro WCDMA 3G a GSM systémy. Vlastnímu zahájení standardizace předcházela fáze studie proveditelnosti širokopásmového kódování. Po téměř dvou letech intenzivního vývoje a dvou fázích výběru kodeků, byl organizacemi 3GPP/ETSI vybrán v prosinci 2000 algoritmus širokopásmového kodeku. Specifikace hovorového kodeku byly dokončeny a schváleny v březnu 2001.

3GPP/ETSI širokopásmový kodek je adaptivní kodek, který je schopen provozu s velkým množstvím hovorových kódovacích přenosových rychlostí od 6,6 do 23,85 kbit/s. Kodek byl nazván jako Adaptive Multi-Rate Wideband (AMR-WB) kodek.

AMR-WB kodek zahrnuje sadu pevných přenosových rychlostí a kanál režimů kodeku, detektor hlasové aktivity (VAD), funkcionalitu nesouvislého přenosu (DTX) pro GSM a funkcionalitu zdrojové regulace přenosové rychlosti (SCR) pro 3G sítě, vnitropásmovou signalizaci přenosu režimů kodeku a spojovací přizpůsobení pro ovládání výběru režimů. AMR-WB kodek upravuje přiřazení přenosové rychlosti mezi hovorové a kanálové kódování s tím, že optimalizuje kvalitu hovoru dle směřovatných podmínek rádiového kanálu. Ačkoli AMR-WB poskytuje kvalitu hlasu přes nynější úzkopásmové standardy, je také velmi odolný vůči přenosovým chybám

způsobených vícerychlostními operacemi a přizpůsobením provozním podmínkám kanálu. Přizpůsobení je založeno na podobných principech jako v dříve standardizovaném AMR kodeku (označovaného také jako AMR-NB).

AMR-WB kodek byl vyvinutý pro použití v několika aplikacích: GSM full-rate kanál, obvodově spínané kanály s osmistavovou fázovou modulací (8 PSK) pro GSM EDGE radiové přístupové sítě (GERAN), kanál pro 3G univerzální pozemní radiovou přístupovou síť (UTRAN), a také pro paketové aplikace pro přenos hlasu přes internetový protokol (VoIP).

V průběhu standardizace AMR-WB organizací 3GPP, probíhal zároveň v organizaci ITU-T proces pro standardizaci širokopásmového hovorového kodeku s přenosovými rychlostmi přibližně 16 kbit/s. Vyvíjený standard širokopásmového kódování zahrnoval široký rozsah aplikací zahrnující: ISDN širokopásmovou telefonii, audio/video telekonference, VoIP a IP videokonference, hlasovou poštu, audio chat, rozhlas, a hlasový streaming. Velkým přínosem byla shledána také vzájemná slučitelnost s 3G mobilními systémy. Na základě výsledků testů AMR-WB kodeku v organizaci 3GPP, ITU-T schválilo tento kodek pro účast ve výběrovém řízení pro výběr standardu širokopásmového kódování.

V červenci 2001 byl na základě výsledků testů oznámen výběr AMR-WB kodeku, což znamenalo jediný harmonizovaný širokopásmový kodek pro GSM, WCDMA 3G a ITU-T. Tato skutečnost eliminuje potřebu transkódování a snadnější implementaci širokopásmových hlasových aplikací a služeb napříč širokým spektrem sdělovacích systémů a platforem. Schválení kodeku AMR-WB touto organizací bylo velmi významné, neboť poprvé je stejný kodek použit jak pro bezdrátové přenosy, tak pro klasické přenosy metalickými vedeními.

2.1.2 Vývoj kódování řeči v systémech GSM a WCDMA

Full-Rate (FR) kodek s přenosovou rychlostí 13 kbit/s byl první hlasový kodek definovaný pro technologii GSM. Standardizován byl v roce 1989 a je používán v GSM Full-Rate kanálu dodnes. Hrubá (celková) bitová rychlost je 22,8 kbit/s a je výchozím kodekem pro hlasové služby v GSM. Half-Rate (HR) kodek s přenosovou rychlostí 5,6 kbit/s byl standardizován v roce 1995 pro provoz šetřící kapacitu kanálu v Half-Rate kanálu s hrubou bitovou rychlostí 11,4 kbit/s.

HR kodek poskytuje stejnou úroveň kvality řeči (MOS) jako FR kodek, s výjimkou zbytkového šumu v pozadí. V tandemu (dvě kódování hovoru od mobilní stanice k mobilní stanici), mívá komprese výkon poněkud nižší.

Kodek zvaný Enhanced Full-Rate (EFR) byl první kodek poskytující hovorovou kvalitu srovnatelnou s klasickými (metalickými) telekomunikačními sítěmi. Oproti předchozím dvěma typům kodeků přinesl značné zlepšení kvality přenášeného hlasu.

Typ komprese	Přenosová rychlost [kbit/s]	MOS	Zpoždění [ms]
PCM (G.711)	64	4,4	0,75
ADPCM (G.726)	32	4,2	1
LD-CELP (G.728)	16	3,61	2
RPE-LTP (GSM 06.10)	13	3,5	20
CS-ACELP (G.729A)	8	4,2	15
ACELP (G.723.1)	5,3	3,5	35

Tab. 2.1: Srovnání kodeků dle MOS a komprese.

EFR poskytuje hovorovou kvalitu jako klasické sítě přes celé spektrum typických radiových podmínek při odstupu interferencí k nosné 10 dB. Stejná kvalita přenosu jako u metalických vedení byla požadována jakmile se systém GSM začal stále více používat v komunikačním prostředí, kde začal konkurovat pevným linkám nebo bezdrátovým systémům (např. NMT). Konkurenceschopnost spočívala v tom, že systém GSM dokázal poskytovat hovorovou kvalitu stejně jako pevná vedení, která jsou odolná v podmínkách typického použití, například vůči šumu v pozadí nebo přenosovým chybám.

EFR kodek ještě zanechal místo pro zlepšení. Především výkon v náročných podmínkách, kdy dochází k chybám v kanálu, může být zlepšen použitím rozdílné bitové alokace pro hovor a kanálové kódování. Dalším vývojem v GSM ve zlepšení hlasové kvality byla standardizace AMR-NB kodeku v roce 1999. Tento kodek nabízí hlavní zlepšení oproti EFR v chybové odolnosti ve FR kanálu přizpůsobením přenosové rychlosti řeči a kanálového kódování závisující na běžných kanálových podmínkách. v GSM FR kanálu, AMR kodek rozšiřuje provozní oblast v EFR od cca $C/I \geq 10$ dB do $C/I \geq 4-7$ dB. Přepínáním do provozu v HR kanálu během dobrých kanálových podmínek, dává AMR kodek také zisk kanálové kapacity přes EFR při zachování vysoké kvality hlasu. Proto byl tento kodek přijat jako výchozí kodek pro WCDMA 3G systémy a byl společně vyvíjen společnostmi Ericsson, Siemens a Nokia.

AMR-WB kodek je nejnovějším hlasovým kodekem standardizovaným pro GSM a WCDMA 3G systémy. Zatímco všechny předchozí kodeky v mobilních komunikacích pracovaly s úzkým audio pásmem limitovaným od 200-3400 Hz, AMR-WB kodek rozšiřuje toto pásmo na 50-7000 Hz a přináší značné zlepšení kvality. Širokopásmový kodek pracuje na devíti přenosových rychlostech hovorového kódování mezi 6,6 až 23,85 kbit/s. Stejně jako ostatní GSM a WCDMA 3G kodeky, má také AMR-WB kodek zdrojově závislý mód s nízkou přenosovou rychlostí pro kódování šumu pozadí.

Standardizaci tohoto kodeku, která byla zahájena v polovině roku 1999, před-

cházel fáze studie proveditelnosti. Výsledky ukázaly, že cíl je uskutečnitelný a tudíž mohla být standardizace zahájena. Tyto práce byly uskutečněny jako společné úsilí organizace ETSI a 3GPP dosáhnout vývoje širokopásmového kódování pro oba mobilní telekomunikační systémy 2G a 3G.

Po zahájení standardizace byly definovány podrobné požadavky na hovorovou kvalitu a návrh kodeku. Výběr AMR-WB kodeku byl pak zajištěn soutěžním postupem, jež se skládal ze dvou částí: kvalifikační fáze na jaře 2000 a výběrová fáze v období od června až do října 2000.

Ve výběrové fázi byli kandidáti důkladně testováni v šesti nezávislých testovacích laboratořích. Testování bylo koordinováno mezinárodně a v pěti jazycích. Každý pokus v testech byl proveden ve dvou jazycích s cílem nebýt zaujatý vůči jednotlivému jazyku. Testy zahrnovaly hovor s a bez šumu pozadí, chyby kanálu, přizpůsobení módu a také provoz zdrojově řízené přenosové rychlosti. Na základě výsledků a technických detailech navrhovaného kodeku byl v prosinci 2000 vybrán organizacemi ETSI/3GPPP kodek od společnosti Nokia/VoiceAge. Poté byly dokončeny specifikace hovorového kodeku a ty byly schváleny v březnu 2001.

Po schválení specifikací kodeku se konala fáze charakteristiky. V průběhu této fáze byl AMR-WB kodek podroben různým reálným provozním podmínkám.

2.1.3 Algoritmický popis kodéru

AMR-WB kodek, jehož blokové schéma je zobrazeno na Obr. A.1, je založen na technologii lineární predikce buzené algebraickým kódem (ACELP). Tato technologie byla velmi úspěšná v kódování hovorových signálů telefonního pásma a několik ACELP standardů bylo použito v širokém rozsahu aplikací zahrnujících digitální celulární aplikace a VoIP technologie.

Přestože technologie ACELP poskytuje dobrou výkonnost v oblasti úzkopásmových signálů, určité obtíže nastávají, když je použito telefonní pásmo optimalizované ACELP modelem pro širokopásmovou řeč. ACELP často spotřebuje nejvíce svých kódovacích bitů v nízkofrekvenční oblasti, která má obvykle větší obsah energie, vyplývající z výstupního signálu dolní propusti. K překonání tohoto problému musí být modifikován váhovací filtr za účelem přizpůsobení širokopásmových signálů. Dále obsahy základního tónu ve spektru hlasových segmentů v širokopásmových signálech nedosahují přes celý rozsah spektra a vyjadřující množství ukazuje více variant oproti úzkopásmovým signálům. Je tedy důležité zlepšit řízenou analýzu základního tónu k lepšímu přizpůsobení variant v hlasové úrovni. Ze stejných důvodů je také důležité zlepšit techniky rozšíření periodicity v dekodéru. Další důležitá záležitost jež vyvstává v kódování širokopásmových signálů, je potřeba použít velmi velké budící

kódové knihy. Proto se tedy účinné struktury kódové knihy, jež požadují minimální paměť a mohou být rychle vyhledány, stávají zásadními.

Popis kodeku

AMR-WB kodek sestává z devíti módů kódování řeči s bitovými rychlostmi 23,85; 23,05; 19,85; 18,25; 15,85; 14,25; 12,65; 8,85 a 6,6 kbit/s. AMR-WB zahrnuje také mód šumu v pozadí, který je určen k použití v provozu nesouvislého přenosu (DTX) v GSM a jako zdrojově závislý mód s nízkou bitovou rychlostí pro kódování šumu v pozadí v jiných systémech. Mód 12,65 kbit/s a výše nabízí širokopásmovou řeč vysoké kvality a dva nižší módy 8,85 a 6,6 kbit/s jsou určeny pro dočasné použití během náročných podmínek radiového kanálu nebo během přetížení sítě.

Blokové schéma kodéru a dekodéru je zobrazeno na Obr. A.1 a Obr. A.2. AMR-WB kodek pracuje s vzorkovací frekvencí 16 kHz a kódování je prováděno v blocích po 20 ms. Dvě frekvenční pásma 50-6400 Hz a 6400-7000 Hz jsou kódovány odděleně za účelem snížení výpočetní náročnosti a soustředit bitovou alokaci do nejdůležitějšího frekvenčního pásma.

Vstupní signál je podvzorkován na 12,8 kHz a je předzpracován použitím horní propusti a filtrem předběžné korekce ve tvaru: $P(z) = 1 - \mu z^{-1}$, kde $\mu = 0,68$. Algoritmus ACELP je poté použit na podvzorkovaný a předzpracovaný signál. Analýza lineární predikce je prováděna jedenkrát za 20 ms. Sada parametrů LP je konvertována do párů imitančního spektra a vektorově kvantována použitím vícestupňového kvantování s 46 bity. Hovorový rámec je rozdělen do čtyř subrámců po 5 ms. Parametry adaptivní a fixní kódové knihy jsou přenášeny každým subrámcem. Interval základního tónu je kódován devíti bity v lichých subrámcích a relativně kódován šesti bity v sudých subrámcích. Jeden bit na rámec je použit pro určení LP u posledního buzení. Zisk algebraické kódové knihy a základní tón jsou společně kvantovány použitím sedmi bitů na rámec.

Algebraická kódová kniha [b]	36	44	52	64	72	88
Bitová rychlost [kbit/s]	12,65	14,25	15,85	18,25	19,85	23,05

Tab. 2.2: Přiřazení algebraických kódových knih k bitovým rychlostem.

Mód s bitovou rychlostí 23,85 kbit/s se liší od módu s bitovou rychlostí 23,05 kbit/s v použití čtyř bitů na subrámec ke kódování vysokopásmového zisku. Mód s bitovou rychlostí 8,85 kbit/s používá 46 bitů pro parametry LP, 8 bitů pro interval základního tónu v lichých subrámcích a 5 bitů v sudých subrámcích, 20 bitů pro algebraickou kódovou knihu a 6 bitů pro dva zisky v každém subrámcu. Mód s bitovou rychlostí 6,6 kbit/s používá 36 bitů pro parametry LP, 8 bitů pro interval

základního tónu v prvním subrámcí a 5 bitů v ostatních subrámcích, 12 bitů pro algebraickou kódovou knihu a 6 bitů pro dva zisky v každém subrámcí.

Vyšší frekvenční pásmo (6400-7000 Hz) je rekonstruováno v dekodéru použitím parametrů nižšího pásma a náhodného buzení. Nejsou přenášeny žádné informace o vyšším pásmu kromě módu s bitovou rychlostí 23,85 kbit/s, kde je přenášén širokopásmový zisk použitím 4 bitů na rámec. V ostatních módech je zisk vyššího pásma nastaven dle nižšího pásma použitím hlasové informace. Spektrum vyššího pásma je rekonstruováno použitím filtru širokopásmové LP generované z nízkopásmového filtru LP.

Parametr	Mód kodeku [kbit/s]								
	6,6	8,85	12,65	14,25	19,85	18,25	19,85	23,05	23,85
VAD znak	1	1	1	1	1	1	1	1	1
LTP znak	0	0	4	4	4	4	4	4	4
ISP	36	46	46	46	46	46	46	46	46
Zpoždění zákl. tónu	23	26	30	30	30	30	30	30	30
Algebraický kód	48	80	144	176	208	256	288	352	352
Zisk	24	24	28	28	28	28	28	28	28
HP energie	0	0	0	0	0	0	0	0	16
Celkem rámec	132	177	253	285	317	365	397	461	477

Tab. 2.3: Bitová alokace pro jednotlivé módy kodeku AMR-WB.

Percepční váhování

V kodérech typu analýza-syntézou jsou optimální základní tón a parametry inovace hledány minimalizující střední kvadratickou chybou mezi vstupní řečí a syntetickou řečí v percepčně váhované oblasti.

Tradičně je percepční váhování prováděno pomocí váhovacího filtru:

$$W'(z) = \frac{A'\left(\frac{z}{\gamma_1}\right)}{A'\left(\frac{z}{\gamma_2}\right)}, \text{ přičemž } 0 < \gamma_2 < \gamma_1 \leq 1 \quad (2.1)$$

kde $W'(z)$ je filtr LP a γ_1 a γ_2 jsou faktory, jež řídí míru váhování omezeného na $0 < \gamma_2 < \gamma_1 \leq 1$. To může ukazovat, že šum kódování (u kterého se předpokládá, že má bílé spektrum) je váhován přenosovou funkcí $\frac{1}{W'(z)}$, která je inverzní k přenosové funkci percepčního váhovacího filtru. Přenosová funkce $\frac{1}{W'(z)}$ vykazuje formantovou strukturu vstupního řečového signálu. Maskovací schopnost lidského

ucha je využívána zabezpečením chyby kvantování tak, že má více energie ve formantových oblastech, kde bude maskován silnou signálovou energií nacházející se v těchto oblastech.

Tento tradiční váhový filtr pracuje dobře se signály telefonního pásma, kde přenosová funkce $W(z)$ nevykazuje silné spektrální vychýlení. Nicméně to není vyhovující pro účinné váhování širokopásmových signálů přímo k vlastním omezením $W(z)$ v modelování formantové struktury a požadovaného spektrálního vychýlení současně. Spektrální vychýlení je zřetelnější v širokopásmových signálech, protože mají široký dynamický rozsah mezi nízkými a vysokými frekvencemi. To bylo vyřešeno v minulosti přidáním vychylovacího filtru do $W(z)$ za účelem odděleně řídit vychýlení a formantové váhování širokopásmového vstupního signálu.

Novým řešením tohoto problému je uvedení předzkreslujícího filtru tvaru: $P(z) = 1 - \mu z^{-1}$ na vstupu, počítat filtr LP $A(z)$ založeného na předzkreslené řeči $s(n)$ a použít modifikovaný filtr $W(z)$ fixací jeho průměrné úrovně. Percepční váhovací filtr může být například definován jako:

$$W(z) = \frac{A\left(\frac{z}{\gamma_1}\right)}{(1 - \gamma_2 z^{-1})}, \text{ přičemž } 0 < \gamma_2 < \gamma_1 \leq 1 \quad (2.2)$$

Vyšší řád filtru může být použit ve jmenovateli. Tato struktura podstatně potlačuje vazbu formantového váhování z odchylky.

Stojí za povšimnutí, že $A(z)$ je vypočten na základě předzkresleného řečového signálu $s(n)$, odchylka filtru $\frac{1}{A\left(\frac{z}{\gamma_1}\right)}$ je méně výrazný v porovnání, kdy je $A(z)$ vypočten na základě originální řeči. Jelikož je snímací korekce prováděna v dekodéru použitím filtru majícím přenosovou funkci $\frac{1}{P(z)}$, chyba kódování je tvarována filtrem $W^1(z)P^{-1}(z)$, kde je γ_1 nastavena na stejnou hodnotu jako μ . Toto je typický případ, kdy je spektrum chyby kvantování zformováno filtrem, jehož přenosová funkce je $\frac{1}{A\left(\frac{z}{\gamma_1}\right)}$ a $A(z)$. Subjektivní poslech ukázal, že tato struktura je pro dosažení zabezpečení chyby kombinací předzkresleného a modifikovaného váhovacího filtrování velmi efektivní u kódování širokopásmových signálů. Dále předzkreslující filtr redukuje dynamický rozsah vstupního hovorového signálu, který je pak vhodnější pro implementaci řadové čárky. Bez předzkreslení je analýza LP obtížná k implementaci řadové čárky používající jednoduchý a přesný výpočet.

Analýza základního tónu

V AMR-WB kodeku je vyhledávání základního tónu složeno ze tří stupňů. V prvním stupni je interval základního tónu otevřené smyčky T_0 očekáván každých 10 ms na

základě váhovaného řečového signálu filtrovaného dolní propustí.

Ve druhém stupni je prováděno hledání základního tónu uzavřené smyčky pro celočíselné intervaly základního tónu a očekávaného intervalu základního tónu otevřené smyčky T_0 v rozsahu ± 7 vzorků, které významně zjednodušují vyhledávací proceduru. Jakmile je nalezen optimální celočíselný interval základního tónu ve druhém stupni, třetí stupeň vyhledávání prochází přes frakce kolem kterých je optimální hodnota celého čísla. Interval základního tónu je omezen v rozsahu [34, 231].

Pro první a třetí subrámeček s bitovými rychlostmi 12.65 kbit/s a vyšší je frakční interval základního tónu použit s rozlišeními $1/4$ a $1/2$ v rozsazích [34, 127 $3/4$] a [128, 159 $1/2$]. Celočíselné rozlišení je použito v rozsahu [160, 231]. Pro druhý a čtvrtý subrámeček je vždy použito rozlišení $1/4$ ve vyhledávání omezeného v rozsahu $[-8, 7 \frac{3}{4}]$ kolem intervalu základního tónu pro přechozí subrámeček.

Vyhledávání intervalu základního tónu uzavřené smyčky je prováděno minimalizováním střední kvadratické chyby mezi originální a syntetizovanou řečí. Toho je dosaženo maximalizováním:

$$C_k = \frac{\sum_{n=0}^{N-1} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{N-1} y_k(n)y_k(n)}} \quad (2.3)$$

kde N je velikost subrámečku, $x(n)$ je cílový signál a $y_k(n)$ poslední buzení v intervalu k , filtrovaného přes váhovací filtr syntézy $\frac{W(z)}{A(z)}$. $W(z)$ je váhovací filtr, jež je popsán v předchozí sekci a $\frac{1}{A(z)}$ je filtr syntézy s kvantovanými koeficienty. Signál $y_k(n)$ je vypočten konvolucí posledního buzení v intervalu k , $u_k(n)$ s impulsní odezvou váhovacího filtru syntézy $h(n)$. Pro intervaly základního tónu menší než velikost subrámečku, je poslední buzení rozšířeno zbytkovým signálem LP. Zajímavostí je, že rozsah vyhledávání je limitován cca odhadem základního tónu otevřené smyčky T_0 .

Nejdříve je určen optimální celočíselný interval základního tónu a části tzv. frakce od $-3/4$ do $3/4$ s krokem cca $1/4$, jež jsou prohledány. Frakční vyhledávání základního tónu je prováděno interpolací normalizované korelace (2.2) a hledání jeho maxim. Nejprve je určen frakční interval základního tónu, interpolován posledním buzením a $v'(n)$ je vypočten interpolací signálu posledního buzení $u(n)$ v daném základním tónu a frakci. Interpolace je provedena s pomocí dvou FIR filtrů složených z Hammingových okénkových *sinc* funkcí, jeden pro interpolaci výrazu (2.2) s filtrem 35. řádu a druhý pro interpolaci posledního buzení s filtrem 127. řádu. Filtry mají mezní kmitočet v 6 kHz, což značí, že interpolační filtry vykazují frekvenční odezvu dolní propustí.

Pro zvýšení predikce základního tónu v širokopásmových signálech je použit frekvenčně závislý prediktor základního tónu. Toto je důležité v širokopásmových signálech, poněvadž se periodicitu nutně nerozšiřuje přes celé spektrum.

Je-li prediktor základního tónu reprezentován filtrem:

$$W(z) = \frac{1}{(1 - bz^{-T})}, \quad (2.4)$$

který je oprávněným předpokladem pro interval základního tónu $T \geq N$, filtr základního tónu vykazuje harmonickou strukturu přes veškeré frekvenční rozsahy s harmonickou frekvencí související s $\frac{1}{T}$. V širokopásmových signálech tato struktura není účinná, poněvadž harmonická struktura typicky nepokrývá úplné spektrum až do 7 kHz. Harmonická struktura trvá pouze do jisté frekvence závisující na hovorovém segmentu. Za účelem dosáhnout účinné reprezentace za přispění základního tónu v hlasových segmentech širokopásmové řeči, filtr predikce základního tónu potřebuje mít flexibilitu proměnné sumy periodicity ve vysoké frekvenci.

Metoda, která dosahuje účinného modelování harmonické struktury řečového spektra v širokopásmových signálech umožňuje, že několik forem dolních propustí je aplikováno na poslední buzení a je vybrána dolní propust s nejvyšším ziskem predikce.

Pokud je použito dílčí rozeznání základního tónu, použití dolní propusti v interpolačních filtrech umožní lepší rozeznání základního tónu. Poté je třetí stupeň vyhledávání základního tónu, v kterém jsou testovány zlomky kolem vybraného celočíselného intervalu základního tónu, zopakován pro několik interpolačních filtrů majících rozdílné charakteristiky dolní propusti, frakci a index filtru, které jsou vybrány pro maximalizaci kritéria vyhledávání.

Kompletního vyhledávání ve třech stupních k určení optimálního frakčního vektoru základního tónu je dosaženo pomocí pouze jednoho interpolačního filtru (obvykle filtr s širší frekvenční odezvou). Optimální podoba dolní propusti může být stanovena v závěru pomocí předem určených filtrů k vybrání vektoru v_T základního tónu a výběru filtru, jež minimalizuje chybu predikce základního tónu.

V AMR-WB kodeku je alokován jeden bit na subrámeček charakterizující dolní propust použitou pro tvarování buzení adaptivní kódové knihy. Pokud je dolní propust vyřazena, buzení adaptivní kódové knihy se stává jednoduchým $v(n) = v'(n)$. Neboli FIR filtr druhého řádu $B_L(z) = b_L(0) + b_L(1) + b_L(2)z^{-1}$ je použit pro filtrování adaptivní kódové knihy plynoucí ze vztahu:

$$v(n) = \sum_{i=1}^1 b_L(i+1) v'(n+i) \quad (2.5)$$

Koeficienty filtru jsou vybrány tak, že $b_L(0) = b_L(2) = 0,18$ a $b_L(1) = 0,64$. Tento filtr druhého řádu dává praktický kompromis mezi obtížností a daným vysokofrekvenčním útlumem. Zisk adaptivní kódové knihy je pak nalezen jako:

$$g_p = \frac{\sum_{n=0}^{N-1} x(n) y(n)}{\sqrt{\sum_{n=0}^{N-1} y(n) y(n)}}, \text{ pro } 0_p \leq 1, 2 \quad (2.6)$$

kde $y_n = v_n \cdot h_n$ je filtrovaný vektor adaptivní kódové knihy. Aby se předešlo potenciální nestabilitě v dekodéru v případě kanálových chyb, zisk adaptivní kódové knihy g_p je shora ohraničen hodnotou 0,95, v případě, že zisky adaptivní kódové knihy byly v předchozích subrámcích malé a filtry LP předchozích subrámců byly blízko nestabilního stavu.

Struktura a vyhledávání algebraické fixní knihy

V širokopásmovém hovorovém kódování jsou potřeba velmi rozsáhlé kódové knihy za účelem garance vysoké subjektivní kvality. Složení kódové knihy je založeno na metodě ISPP. [3] Kódový vektor obsahuje 64 pozic, jež jsou rozděleny do čtyř prokládaných stop, s 16 pozicemi v každé stopě. Různé kódové knihy s různými rychlostmi jsou vytvořeny umístěním určitého čísla označených impulzů ve stopách od 1 do 6 impulzů na stopu. Index kódové knihy nebo kódového slova představuje impulzní pozice a znaky v každé stopě. Není tedy potřeba žádná paměť kódové knihy, protože vektor buzení v dekodéru může být vytvořen pomocí informace obsažené v jeho indexu bez vyhledávacích tabulek.

Pokud je umístěn jeden označený impuls v každé stopě, pozice impulsu je kódována čtyřmi bity a jeho znak je kódován jedním bitem. Výsledkem je pak 20 bitová kódová kniha. Pokud jsou umístěny dva označené impulsy v každé stopě, dvě pozice impulsu jsou kódovány osmi bity a jejich odpovídající znaky mohou být kódovány pouze jedním bitem využívajícího impulsní uspořádání. Proto je tedy celkem $4 \cdot (4 + 4 + 1) = 36$ bitů požadovaných k upřesnění pozic impulsů znaků pro toto partikulární složení algebraické kódové knihy. Další složení kódové knihy může být navrženo umístěním 3, 4, 5 nebo 6 impulzů v každé stopě.

Důležitá vlastnost použité kódové knihy je, že je dynamická a skládá se z algebraické kódové knihy následované adaptivním prefilterem $F(z)$, který rozšiřuje specifické spektrální součásti pro zlepšení subjektivní hovorové kvality. V AMR-WB kodeku se $F(z)$ skládá ze dvou filtrů zapojených do kaskády, rozšíření periodicity filtru je:

$$F_1(z) = \frac{1}{1 - 0,85z^{-T}} \quad (2.7)$$

a filtr odchylky je ve tvaru:

$$F_2(z) = (1 - \beta_T z^{-1}) \quad (2.8)$$

Koeficient β_T je vztažen k vyjádření předchozího subbrámce a je omezen intervalem $[0; 0,5]$ a T je celočíselná část intervalu základního tónu. Vyhledávání kódové knihy je provedeno v algebraické oblasti kombinováním filtru $F(z)$ s váženou syntézou filtru před vyhledáváním kódové knihy. Toto je dáno konvolucí impulsní odezvy vážené syntézy filtru s impulsní odezvou $F(z)$.

Algebraická kódová kniha je hledána minimalizováním střední kvadratické chyby mezi váženou vstupní řečí a váženou syntetizovanou řečí. Cílový signál použitý ve vyhledávání uzavřené smyčky základního tónu je aktualizován odčítáním přínosu kódové knihy.

Následně je signál $h(n)$ označen impulsní odezvou váhovacího filtru syntézy spojeného s prefilterem $F(z)$. Matice $\mathbf{H}$ je definována jako nižší tzv. trojhranná Toeplitzova konvoluční matice s diagonálou $h(0)$ a nižšími diagonálami $h(1), \dots, h(N-1)$ [1]. Použitím této symboliky $d = H^t x_2$ je korelace mezi cílovým signálem $x_2(n)$ a impulsní odezvou $h(n)$, přičemž $\phi = H^t H$ je matice korelací $h(n)$. Korelace $\mathbf{d}$ je také uvedena jako zpětný filtrovaný cílový vektor a odpovídající signál označen $d(n)$.

Jestliže je c_k algebraický kódový vektor s indexem k , pak je algebraická kódová kniha vyhledávána maximalizací vyhledávacího kritéria:

$$Q_k = \frac{(x_2^t H c_k)^2}{c_k^t H^t H c_k} = \frac{(d^t c_k)^2}{c_k^t \phi c_k} = \frac{(C_k)^2}{E_k} \quad (2.9)$$

Vektor d a matice ϕ jsou obvykle vypočteny před vyhledáváním kódové knihy.

Algebraická struktura kódové knihy dovoluje velmi rychlé vyhledávací procedury, protože vektor c_k obsahuje pouze několik nenulových impulzů. Korelace v čitateli je pak:

$$C = \sum_{i=0}^{N_p-1} a_i d(p_i) \quad (2.10)$$

kde p_i je pozice i -tého impulzu, a_i je jeho znak a N_p je počet pulzů v c_k . Energie ve jmenovateli E_k je dána:

$$E = \sum_{i=0}^{N_p-1} \phi(p_i, p_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} a_i a_j \phi(p_i p_j) \quad (2.11)$$

K usnadnění vyhledávací procedury jsou předem určené impulzní amplitudy založené na referenčním signálu $b(n)$. Tento signál je dán poměrem impulzní amplitudy, znak impulzu v pozici i je nastaven tak, že se rovná znaku referenčního signálu v dané pozici. Pak je zde referenční signál $b(n)$ definován jako:

$$b(n) = \sqrt{\frac{E_d}{E_r}} r(n) + \alpha d(n), \quad (2.12)$$

kde $E_d = d^T d$ je energie signálu $d(n)$. Podobně $E_r = r^T r$ je energie zbytkového signálu $r(n)$ vyplývajícího z predikce základního tónu. Součinitel velikosti α řídí závislost referenčního signálu na $d(n)$ a je snížen zatímco počet budících impulzů je zvýšen. Například hodnota α je nastavena na 1 při bitové rychlosti 12,65 kbit/s a redukován na 0,5 při bitové rychlosti 23,05 kbit/s.

K zjednodušení vyhledávání jsou signál $d(n)$ a matice ϕ upraveny k připojení předvolených znaků. Pokud $s_b(n)$ udá znak $b(n)$, tak modifikovaný signál $d'(n)$ je dán:

$$d'(n) = s_b(n) d(n), \dots, n = 0, \dots, N-1 \quad (2.13)$$

a upravená autokorelační matice:

$$\phi'(i, j) = s_b(i) s_b(j) \phi(i, j) \quad i = 0, \dots, N-1; j = i, \dots, N-1 \quad (2.14)$$

čitatel kritéria vyhledávání je teď jednoduchý:

$$C_k = \sum_{i=0}^{N_p-1} d'(i) \quad (2.15)$$

a tvar jmenovatele:

$$E_k = \sum_{i=0}^{N_p-1} \phi'(p_i, p_i) + 2 \sum_{i=0}^{N_p-2} \sum_{j=i+1}^{N_p-1} \phi'(p_i p_j) \quad (2.16)$$

Cílem vyhledávání je určit kódový vektor s nejlepší sadou N_p impulsních pozic amplitud impulzů, jež byly vybrány a popsány výše. Základní kritérium výběru je maximalizace Q_k .

Pro zmenšení obtížnosti je použita rychlá procedura známá jako hloubkové stromové vyhledávání, kterým jsou najednou určeny impulzní pozice N_m impulzy. Přesněji, N_p dostupné impulzy jsou rozděleny do M neprázdných podskupin N_m impulzů v tomto pořadí: $N_1 + N_2 \dots + N_m \dots N_M = N_p$. Specifická volba pozic pro první $J = N_1 + N_2 \dots + N_{m-1}$ impulzy je volána cestou úrovně m nebo trasou délky J . Základní kritérium pro trasu impulsních pozic J je poměr $Q_k(J)$, kdy jsou pouze zvažovány významné J impulzy.

Vyhledávání začíná s podskupinou č.1 a postupuje s následující podskupinou dle stromové struktury, takže podskupina m je vyhledávána v m -té úrovni stromu. Účelem vyhledávání v úrovni 1 je uvažovat N_1 impulzy podskupiny č.1 a jejich platné pozice pro určení jednoho nebo několika kandidátských tras délky N_1 . Tyto kandidující trasy jsou stromové uzly v úrovni 1. Trasa v každém koncovém uzlu úrovně $m - 1$ je rozšířena na délku $N_1 + N_2 \dots + N_m$ v úrovni m s ohledem na N_m nové impulzy a jejich platné pozice. Jedna nebo několik kandidujících rozšířených tras jsou určeny ke zřízení uzlů úrovně m . Nejlepší kódový vektor odpovídá trase délky N_p , která maximalizuje kritérium $Q_k(N_p)$ s ohledem na všechny uzly úrovně M . [literatura]

Je zde použit speciální tvar procesu hloubkového stromového vyhledávání, ve kterém jsou najednou prohledávány dva impulzy, ($N_m = 2$), a tyto dva impulzy náleží k následujícím drahám. Dále je vyhledávání prováděno tak, že pouze část korelační matice ϕ je přepočtena a uložena. Toto redukuje spotřebu paměti, protože plná matice ϕ požaduje N slov ($64 \times 64 = 4096$ slov). Předem uložená část představuje korelace impulsní odezvy odpovídající možným impulzním pozicím v následujících drahách, rovněž korelace odpovídající $\phi(j, j)$ pro $j = 0, \dots, N - 1$ tj. úhlopříčných základních prvků matice ϕ .

Ke snížení obtížnosti jsou testovány limitované počty možných pozic prvního impulzu, zatímco je prováděno testování možných kombinací dvou impulzů. V případě velkého počtu impulzů, je několik impulzů ve vyšších úrovních vyhledávacího stromu fixováno. Pro zhotovení inteligentního odhadu na možných pozicích byl zvažován první impuls $b(n)$, který je použit jako impulzně-poziční pravděpodobnost odhadu založeném na charakteru hovorových signálů. Hodnota $b(n)$ charakterizuje pravděpodobnost impulzu zabírajícího pozici n v nejlepším kódovém vektoru, který je vyhledáván.

Vyhledávací postupy jsou pro všechny módy podobné. Dva impulzy jsou vyhledávány najednou a tyto impulzy vždy odpovídají sousledným drahám. Před vyhledáním pozic je znak impulsu v potenciální pozici n a je provedeno nastavení znaku impulsu $b(n)$ v dané pozici. Poté jsou upravený signál $d'(n)$ a korelace $\phi'(n, n)$ vypočteny včetně předem určených znaků.

Pro první dva impulzy (první stromová úroveň) je dán čítel kritéria Q_k vztahem:

$$C = d'(p_0) + d'(p_1) \quad (2.17)$$

a jmenovatel:

$$E = \phi'(p_0, p_0) + \phi'(p_1, p_1) + 2\phi'(p_0, p_1), \quad (2.18)$$

kde korelace $\phi'(p_i, p_j)$ byla modifikována k zahrnutí předem určených znaků na pozicích p_i a p_j .

Pro následné úrovně čítel a jmenovatel jsou aktualizovány přidáním podílu dvou nových impulzů. Zamýšlené dva nové impulzy jsou vyhledávány v jisté stromové úrovni s pozicemi p_k a p_{k+1} z dvou po sobě jdoucích drah, aktualizované hodnoty C a E jsou dány:

$$C = C + d'(p_k) + d'(p_{k+1}) \quad (2.19)$$

$$E = E + \phi'(p_k, p_k) + \phi'(p_{k+1}, p_{k+1}) + 2\phi'(p_k, p_{k+1}) + 2_{hv}(p_k) + 2_{hv}(p_{k+1}), \quad (2.20)$$

kde $R_{hv}(m)$ znamená korelaci mezi impulzní odezvou $h(n)$ a

$$v_h(n) = \sum_{i=0}^{k-1} h(n - p_i) \quad (2.21)$$

$$R_{hv}(m) = \sum_{n=m}^{N-1} h(n) v_h(n-m) \quad (2.22)$$

Stojí za povšimnutí, že $v_h(n)$ představuje zpoždění impulzní odezvy v dříve určených pozicích. V každé stromové úrovni jsou hodnoty $R_{hv}(m)$ vypočteny přímo pro všechny možné pozice v každé ze dvou testovaných drah. Pouze korelace $\phi'(m_k, m_{k+1})$ odpovídající impulzním pozicím ve dvou po sobě jdoucích drahách je třeba uložit (4·16·16 slov) spolu s korelacemi $\phi'(m_k, m_k)$ odpovídající úhlopříčné matici ϕ (64 slov). Takto požadavek paměti v existující algebraické struktuře je 1088 slov namísto 64·64=4096 slov.

Jako příklad je zde uvedeno vyhledávání 36-bitové kódové knihy použité u módu s rychlostí 12,65 kbit/s. V této kódové knize jsou umístěny dva impulzy v každé dráze dávající celkově 8 pulzů na subrámeček o délce 64. Dva impulzy odpovídající vždy s následujícími drahami jsou vyhledávány najednou. To jest dva impulzy k vyhledávání jsou v drahách T_0 a T_1 , T_1 a T_2 , T_2 a T_3 nebo T_3 a T_0 . V tomto případě má strom čtyři úrovně. V první úrovni je impulz P_0 přiřazen k dráze T_0 a impulz P_1 k dráze T_1 . V této úrovni není prováděno žádné vyhledávání a pozice dvou impulzů jsou nastaveny k dvěma maximům $b(n)$ v každé dráze. V druhé úrovni je impulz P_2 přiřazen k dráze T_2 a impulz P_3 k dráze T_3 . Čtyři pozice pro impulz P_2 jsou testovány proti všem 16 pozicím impulzu P_3 . Čtyři testované pozice impulzu P_2 jsou určeny na základě maxima $b(n)$ v dráze. V třetí úrovni je impulz P_4 přiřazen k dráze T_1 a impulz P_5 k dráze T_2 . Osm pozic pro impulz P_4 je testováno proti všem 16 pozicím impulzu P_5 . Podobně v předchozí úrovni je osm testovaných pozic impulzu P_4 určeno na základě maxima $b(n)$ v dráze. Ve čtvrté úrovni je impulz P_6 přiřazen k dráze T_3 a impulz P_7 k dráze T_0 . Osm pozic pro impulz P_6 je testováno proti všem 16 pozicím impulzu P_7 . Proto je celkový počet testovaných kombinací 4·16+8·16+8·16=320. Celý proces je opakován čtyřikrát přiřazením impulzů do různých drah.

Další příklad je mód s rychlostí 15,85 kbit/s, kde jsou tři impulzy umístěny v každé ze tří drah pro celkem 12 impulzů. Tři impulzy v dráze mohou být kódovány 13-ti bity, jež používají 52-bitovou kódovou knihu. Existuje šest úrovní v stromovém vyhledávání, kterými jsou vyhledávány dva impulzy v každé úrovni. V prvních dvou úrovních jsou čtyři impulzy nastaveny na maxima $b(n)$. V následujících čtyřech úrovních jsou počty vzájemných testovaných kombinací následující: 4·16, 6·16, 8·16 a 16·16. Jsou použity čtyři iterace dávající celkem 4·26·16=1664 kombinací.

Kvantování zisku

Ziskem adaptivní kódové knihy a fixní kódové knihy jsou vektor kvantovaný použitím 6-bitové kódové knihy pro módy 8,85 a 6,60 kbit/s a 7-bitové kódové knihy pro

všechny ostatní módy. Kvantování zisku fixní kódové knihy používá tzv. "moving-average" predikci s konstantními koeficienty. Čtvrtý řád MA predikce je proveden energií v logaritmické oblasti. Predikce a vyhledávání kódové knihy jsou podobné těm, jež jsou použity u GSM EFR kodeku nebo G.729.

2.1.4 Algoritmický popis dekodéru

Funkce dekodéru, viz. blokové schéma na obr. A.2, se skládá z dekódování přenosových parametrů (VAD - znaku, parametrů LP, vektoru adaptivní kódové knihy, zisku adaptivní kódové knihy, vektoru pevné kódové knihy, zisku pevné kódové knihy a vysokopásmového zisku) a provedení syntézy k získání rekonstruované řeči. Rekonstruovaná řeč je následně zpracována a nadvzorkována. Nakonec je vysokopásmový signál generován na frekvenční pásmo od 6 do 7 kHz. Do dekodéru bylo rovněž přidáno několik funkcí následného zpracování pro zlepšení výkonu širokopásmových signálů.

Dekódování, zpracování buzení a řečová syntéza

Přijaté indexy ISP kvantování jsou použity k rekonstrukci kvantovaného ISP vektoru. Čtyři interpolované ISP vektory jsou vypočteny a konvertovány pomocí LP, která je použita pro syntézu řeči v subrámcích. V každém subrámcích je použit přijatý index adaptivní kódové knihy k nalezení celého čísla a dílčích částí intervalu základního tónu. Buzení adaptivní kódové knihy $v'(n)$ je zjištěno interpolací posledního buzení $u(n)$ v intervalu základního tónu.

Přijatý index algebraické kódové knihy je použit k získání pozic a znaků budících pulsů a též pro nalezení buzení algebraické kódové knihy $c(n)$. Jestliže je celočíselná část intervalu základního tónu menší než subrámeček o velikosti 64, jsou zisky adaptivní a fixní kódové knihy dekódovány a celkové buzení je pak dáno vztahem:

$$u(n) = \hat{g}_p v(n) + \hat{g}_c c(n) \quad (2.23)$$

Před syntézou řeči je provedeno zpracování budících složek za účelem zlepšení subjektivní výkonnosti kodeku.

Zvýraznění signálu ze šumu

Nelineární technika vyhlazení zisku je aplikována na zisk fixní kódové knihy $\hat{g}_c$ pro zlepšení buzení při šumu. Podle stability je zisk fixní kódové knihy vyhlazen za

účelem redukce kolísání energie buzení při stálých signálech. To zlepšuje subjektivní hovorovou kvalitu při stálém šumu v pozadí.

Voicing subbrámce je odhadován pomocí vztahu:

$$\lambda = \frac{E_c}{(E_v + E_c)}$$

, kde E_v je energie buzení adaptivní kódové knihy a E_c je energie buzení fixní kódové knihy. Ustálenost rámce je odhadnuta podle filtrů LP pomocí rozdílu ISP parametrů mezi současným a předchozím rámcem. Faktor ustálenosti θ je omezen intervalem $[0, 1]$ a jeho hodnota inklinuje k souladu se stálými rámci. Nakonec je vypočten faktor vyhlazení zisku, jež je dán vztahem:

$$\sigma = \lambda\theta \quad (2.24)$$

Hodnota σ inklinuje k souladu s nehlasovými a pevnými signály jako je např. pevný šum v pozadí. Pro čistě hlasové nebo nestálé signály se hodnota σ blíží k nule.

Nelineární vyhlazení používá zisk předešlého subbrámce s referenční hodnotou $\hat{g}_c^r$ pro výpočet upraveného zisku $\hat{g}_c^m$. Pokud je $\hat{g}_c \geq \hat{g}_c^r$, upravený zisk $\hat{g}_c^m$ je nastaven 1,5 dB pod hodnotu $\hat{g}_c$, přičemž je zdola omezen k referenční hodnotě. Pokud není splněna výše uvedená podmínka, je $\hat{g}_c^m$ nastaven 1,5 dB nad hodnotu $\hat{g}_c$, přičemž je zhora omezen k referenční hodnotě. Nakonec je zisk fixní kódové knihy vyhlazen pomocí vyhlazovacího faktoru a upravený zisk lze pak vyjádřit vztahem:

$$\hat{g}_c = \sigma \hat{g}_c^m + (1 - \sigma) \hat{g}_c \quad (2.25)$$

Takže ve stálých nehlasových subbrámcích se zisk fixní kódové knihy použitý v syntéze blíží k hodnotě upraveného zisku $\hat{g}_c^m$.

Zvýraznění základního tónu

Toto vylepšení umožňuje lepší subjektivní hovorovou kvalitu filtrováním signálu buzení fixní kódové knihy $c(n)$. To se provádí pomocí inovačního filtru, jehož frekvenční charakteristika zdůrazňuje vysoké frekvence a redukuje nízké frekvence. Koeficienty filtru jsou dány vztahem:

$$F_E(z) = -f_E z + 1 - f_E z^{-1} \quad (2.26)$$

Tyto koeficienty souvisí s pravidelností signálu $f_e = 0,25\lambda$ jsou obnovovány jednou za rámec. Tento filtr poskytuje maximální zdokonalení v nehlasových signálech. S použitím filtru zlepšení základního tónu je pak buzení fixní kódové knihy:

$$c_E(n) = c(n) - f_E(c(n+1) + c(n-1)) \quad (2.27)$$

následně je pak celkové buzení:

$$\hat{u}(n) = \hat{g}_p v(n) + \hat{g}_c c_E(n) \quad (2.28)$$

Syntéza řeči

Rekonstruovaná řeč pro subrámec o velikosti 64 je získána filtrováním zpracovaného celkového buzení pomocí syntézy filtru LP. Platí, že:

$$\hat{s}(n) = \hat{u}(n) - \sum_{i=1}^{16} \hat{a}_i \hat{s}(n-i), n = 0, 1, \dots, 63 \quad (2.29)$$

kde $\hat{a}_i$ je i -tý koeficient interpolovaného filtru LP. Syntetická řeč $\hat{s}(n)$ pak prochází přes adaptivní zpracování. K tomu, aby byla udržena synchronizace s kódérem, je paměť adaptivní kódové knihy aktualizovaná pomocí buzení $u(n)$ bez zpracování.

Nadvzorkování signálu a generování vysokých frekvencí

Syntetický signál je filtrován horní propustí, která eliminuje nežádoucí nízkofrekvenční složky. Poté jsou u signálu utlumeny vysokofrekvenční složky pomocí filtru:

$$\frac{1}{P(z)} = \frac{1}{(1 - 0,68z^{-1})}$$

Nakonec je signál nadvzorkován k získání syntézy nižšího pásma $\hat{s}_L(n)$ při vzorkovacím kmitočtu 16 kHz.

Procedura generování vysokých frekvencí je použita pro zaplnění frekvenčního pásma v rozmezí od 6,4 do 7 kHz. Vysokofrekvenční složka je generována vyplněním horní části spektra bílým šumem, jež je vhodně přizpůsoben v oblasti buzení. Ten je poté převeden do řečové oblasti jeho tvarováním za pomoci filtru, který je odvozen ze stejného LP filtru syntézy, který je použit pro syntetizující podvzorkovaný signál.

Hodnota vysokopásmového buzení $u_H(n)$ je získána ze signálu bílého šumu $w(n)$ pomocí vztahu:

$$u_H(n) = \frac{\hat{g}_H w(n)}{\sqrt{\sum_{k=0}^{63} w^2(k)}}, \quad (2.30)$$

kde $\hat{g}_H$ je vysokofrekvenční zisk a τ je odchylka signálu syntézy. V módu s přenosovou rychlostí 23,85 kbit/s je hodnota $\hat{g}_H$ dekodována z přijatého indexu zisku. V ostatních módech je tato hodnota odhadnutá pro každý subrámeček s pomocí hlasové informace. Výpočet vysokofrekvenčního zisku:

$$\hat{g}_H = w_S(1 - \tau) + 1,25(1 - w_S)(1 - \tau), \quad (2.31)$$

kde v případě aktivované funkce VAD je $w_S=1$, jinak je $w_S=0$.

Činitel odchylky je začleněn do výpočtu zisku, jelikož bere v úvahu vysokofrekvenční složky syntetického signálu. V hlasových částech, kde je přítomno méně energie ve vysokých frekvencích se hodnota τ blíží k 1, vyplývající z nižšího vysokofrekvenčního zisku. Tato skutečnost redukuje energii generovaného šumu v hlasových částech. Vysokopásmový filtr LP syntézy je získán pro každý subrámeček z nízkopásmového filtru LP syntézy pomocí vztahu:

$$\hat{A}_H(z) = \hat{A}\left(\frac{12,8z}{16}\right), \quad (2.32)$$

kde $\hat{A}(z)$ je interpolovaný filtr LP syntézy, který byl odhadnutý pro vzorkovací kmitočet 12,8 kHz. Tento filtr je nyní použit pro signál o frekvenci 16 kHz. Toho lze efektivně využít k tomu, že pásmo 5,1-5,6 kHz v oblasti 12,8 kHz je promítnuto do pásma 6,4-7 kHz v oblasti 16 kHz.

Signál vysokopásmové syntézy $\hat{s}_H(n)$ je vypočten filtrováním signálu $u_H(n)$ přes vysokopásmový filtr LP syntézy. Následně je kmitočtově omezen pomocí FIR pásmové propusti s rozsahem od 6,4-7 kHz. Poslední operací je přidání signálu vysokopásmové syntézy $\hat{s}_H(n)$ k signálu nízkopásmové syntézy $\hat{s}_L(n)$ k vytvoření syntetického řečového signálu.

2.1.5 Výpočetní náročnost kodeku

AMR-WB kodek je definován v aritmetice s pevnou řádovou čárkou používající sadu základních operací definovaných organizací 3GPP/ETSI. Výpočetní náročnost AMR-WB hovorového kodeku je 38,9 WMOPS (Weighted Million Operations Per Second) a tento odhad zahrnuje též VAD, DTX a CNG (Comfort Noise Generation) funkce. Porovnání výpočetní náročnosti kodeku AMR-WB a AMR-NB je uvedeno v tabulce č.3.

2.1.6 Adaptivní provoz AMR-WB v GSM kanálech

AMR-WB má vysokou nespojitost bitových rychlostí mezi 6,6 až 23,85 kbit/s. Pro GSM kanály to dává možnost maximalizovat hovorovou kvalitu přizpůsobením bitových rychlostí pro zvýšení odolnosti vůči přenosovým chybám. Pro neadaptivní 3G UTRAN kanály používající rychlé výkonové řízení si operátoři mohou vybrat vhodné bitové rychlosti k provedení optimálního kompromisu mezi hlasovou kvalitou a kapacitou sítě. Proces linkového přizpůsobení v GSM však nese potřebu měření kvality kanálu a vybrání nejvhodnějšího módu podle běžných podmínek kanálu. Vnitropásmová signalizace (400 bit/s ve FR kanálu) přenáší oba požadované módy pro opačnou linku (na základě měření kvality kanálu) a aktivní mód kodeku přední linky přes bezdrátové rozhraní k přijímací straně.

2.1.7 Detekce hlasové aktivity a funkce nesouvislého přenosu

AMR-WB kodek obsahuje detekci hlasové aktivity tzv. VAD. Tato funkce umožňuje kodeku přepnutí do módu s nižší bitovou rychlostí pro kódování šumu v pozadí. Tato funkce šetří energii v mobilní stanici a také redukuje úroveň celkové interference při bezdrátovém přenosu. Funkce VAD počítá Boolean VAD rozhodnutí pro hlasový rámec každých 20 ms. Toto vyhodnocení je založeno na dělení řečového signálu ve frekvenci 6,4 kHz do dvanácti subpásem a výpočtu úrovně signálu v každém pásmu. Funkce tónové detekce je použita pro zjištění výskytu signalizačního tónu, hlasové řeči nebo ostatních silně periodických signálů. Tato funkce je založena na normalizovaných ziscích základního tónu otevřené smyčky, které jsou vypočteny analýzou základního tónu řečového kodéru. Funkce tónové detekce generuje tónový indikátor.

Úroveň šumu v pozadí je odhadována v každém frekvenčním pásmu na základě VAD rozhodnutí, signálové stacionarity a tónového indikátoru. Střední VAD rozhodnutí je pak počítáno porovnáním vstupního SNR (Signal-to-noise ratio) k adaptivní prahové hodnotě. Přizpůsobení prahové hodnoty je založeno na šumu a odhadu

dlouhodobé úrovně řeči. Konečné VAD rozhodnutí je vypočteno přidáním zbytku periody ke střední hodnotě VAD rozhodnutí

V systému GSM je VAD rozhodnutí použito s pomocí DTX (Discontinuous Transmission) funkce, která umožňuje radiovému vysílači vypnutí v průběhu hovorových prodlev. V průběhu hovorových prodlev je na přijímací straně generován syntetický šum, podobný šumu v pozadí na vysílací straně. Tento syntetický komfortní šum je vytvořen přenosovými parametry popisujícími šum v pozadí v běžné rychlosti v průběhu hovorových prodlev. Tato skutečnost umožňuje přizpůsobit komfortní šum vůči změnám šumu v pozadí.

Analýza komfortního šumu je uskutečněna použitím vstupního signálu o frekvenci 12,8 kHz. Algoritmus nejdříve stanoví váhovaný průměr spektrálních parametrů a průměr energie logaritmického signálu. Parametry komfortního šumu jsou kódovány do speciálního rámce zvaného SID (Silence Descriptor) pro přenos k přijímané straně.

Procedura vytváření komfortního šumu generuje pseudonáhodný šumový vektor, který je přizpůsoben pomocí zisku buzení vypočteného z energie logaritmického signálu. Přizpůsobený šum je pak filtrován přes syntézový filtr, který je generován z váhovaného průměru spektrálních parametrů. Poté je nadvzorkován zpět na frekvenci 16 kHz a vyšší pásmo je generováno v dekodéru.

2.1.8 Kvalita hovoru

Kvalita řeči a její hodnocení

Testování QoS (Quality of Service) je jednou z klíčových rolí v moderních telekomunikačních sítích a význam těchto testování roste se zvyšující se komplikovaností a komplexností telekomunikačních sítí, kde přenosový řetězec zahrnuje stále více přenosových technologií (tzv. konvergence sítí) [8]. Na hodnocení kvality řeči přenosem telekomunikační sítí, resp. po zpracování řečovým kodekem, se používají dvě základní hodnocení, a to objektivní a subjektivní. Takový parametr je jedním z mála obecně měřitelných, pro porovnání odlišných přenosových zařízení, který je ve své podstatě nejbližší z pohledu jednotlivých koncových uživatelů [8].

PSQM - Perceptual Speech Quality Measure

Intruzivní algoritmus pro výpočet PSQM byl navrhnut v roce 1993 a jedná se vlastně o adaptaci verze PAQM (), jež je optimalizována pro telefonní řečové signály. Důvodem bylo zjištění, že psychoakustický efekt maskování se zdá jiný, pokud porovnáme vjem řeči a vjem hudby. Jednou z příčin této skutečnosti je, že lidský mozek

si dokáže připomínat tzv. referenční řečové zvuky častěji z každodenního života, než zvuky hudební.

V PSQM je šumové rušení přímo závislé na kvalitě kódované řeči. Kromě modelování vjemu, metoda PSQM využívá také kognitivní model, což vede k vysoké korelaci mezi subjektivními a objektivními měřeními.

PESQ - Perceptual Estimation of Speech Quality

PESQ je objektivní měřicí nástroj, který odhaduje výsledky subjektivních poslechových testů na telefonních systémech. Používá algoritmus, jež slouží k výpočtu kvality přenosového signálu na základě modelu lidského vjemu. Metoda PESQ používá smyslový model k porovnání originálního, nezpracovaného signálu se znehodnoceným signálem ze sítě nebo síťového prvku. Výsledná hodnota kvality je analogická k subjektivní MOS (Mean Opinion Score), jež je měřena pomocí panel testů dle doporučení ITU-T P.800. Hodnoty PESQ jsou kalibrovány pomocí rozsáhlé databáze subjektivních testů.

Kvalita hovoru AMR-WB kodeku

AMR-WB umožňuje mobilní komunikaci s hovorovou kvalitou, která podstatně převyšuje kvalitu poskytovanou existujícími mobilními sítěmi druhé a třetí generace. V typických provozních podmínkách poskytuje AMR-WB vyšší kvalitu než všechny ostatní GSM kodeky. Dokonce i za velmi špatných podmínek radiových kanálů tento kodek ještě nabízí srovnatelnou kvalitu jako AMR-NB kodek a dalece převyšuje kvalitu GSM kodeků s pevnou přenosovou rychlostí.

AMR-WB poskytuje vysokou nespojitost bitové přenosové rychlosti, což umožňuje jeho použití v mnoha aplikacích v 2G a 3G sítích. Vysoká hovorová kvalita předurčuje tento kodek pro širokopásmové hlasové aplikace u služeb provozovaných na metalických vedeních, což dokázalo jeho schválení organizací ITU-T.

3 SIMULACE V APLIKACI SIMULINK

3.0.9 Úvod

Předešlý rozbor současného stavu kódovacích technik v moderních telekomunikačních sítích byl zaměřen na širokopásmový kodek AMR-WB. Jeho podrobný rozbor poskytuje detailnější popis této, v současné době nejperspektivnější techniky kódování řeči s ohledem na vyžadovanou šířku pásma. Současně byl též brán zřetel na míru hodnocení kódované řeči lidským sluchem a možné vysokorychlostní přenosy.

Z předchozí analýzy vyplývá, že další budoucí vývoj bude směřovat k metodám založeným na technologii kódování ACELP. Jedná se zejména o kódovací techniky AMR-WB. Proto byla zaměřena pozornost na detailní popis jednotlivých kroků uvedených algoritimů, matematický rozbor jednotlivých částí a konečně na praktickou realizaci v programovém prostředí aplikace MATLAB, konkrétně v simulačním prostředí Simulink. Byla použita softwarová verze 7.1.0.246 (R14) Service Pack 3. Všechna následující bloková schémata byla vytvořena pomocí programu MS Visio, neboť aplikace Simulink neumožňuje export blokových schémat do souboru obrazového formátu.

3.0.10 Simulace modulu pre-processing

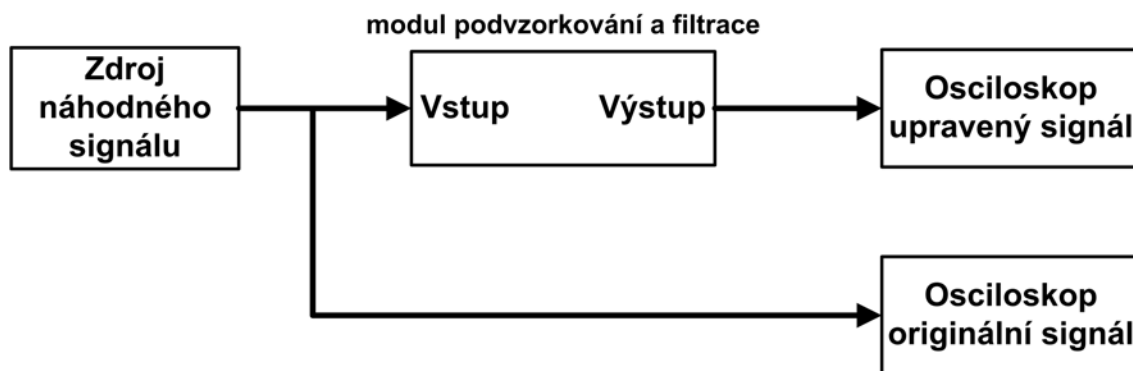
V první části vlastní práce jsem se zabýval simulací vstupního modulu pre-processing neboli předzpracování hovorového signálu, který upravuje signál před vstupem do ACELP kodéru.

Nejdříve jsem vytvořil podmodul Down-sampling (Obr. 3.1), v kterém jsem sestavil blokové schéma skládající se z následujících částí: modul nadvzorkování, modul FIR dolní propusti a poslední člen byl modul podvzorkování. Koeficienty modulů



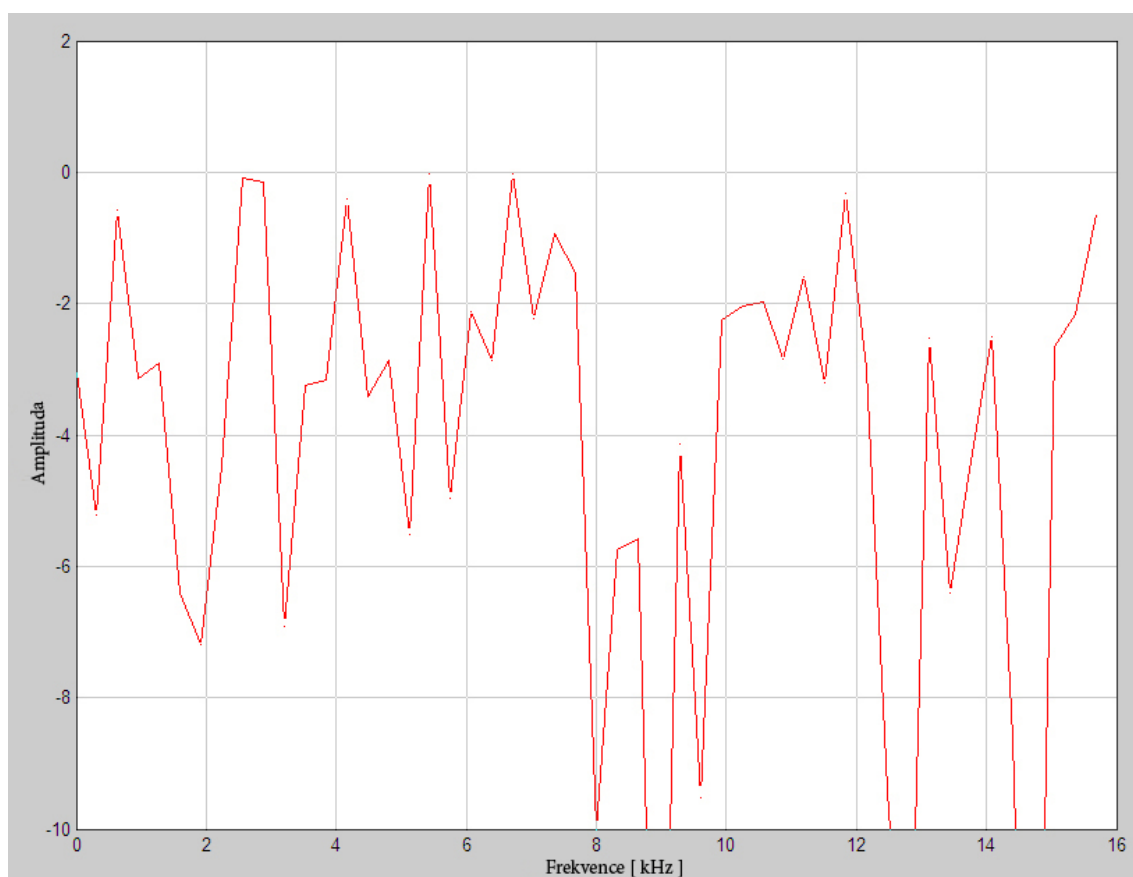
Obr. 3.1: Blokové schéma modulu podvzorkování a filtrace.

nadvzorkování (L) a podvzorkování (M) jsem vypočetl pomocí funkce *gcd*, vzorkovacího a mezního kmitočtu. Nastavení parametrů filtru jsem provedl v dialogovém okně modulu filtru. Takto vytvořený podmodul jsem vložil do hlavního blokového schématu (obr. 3.2). Zde v této simulaci jsem pak porovnával originální řečový signál a signál po předzpracování.

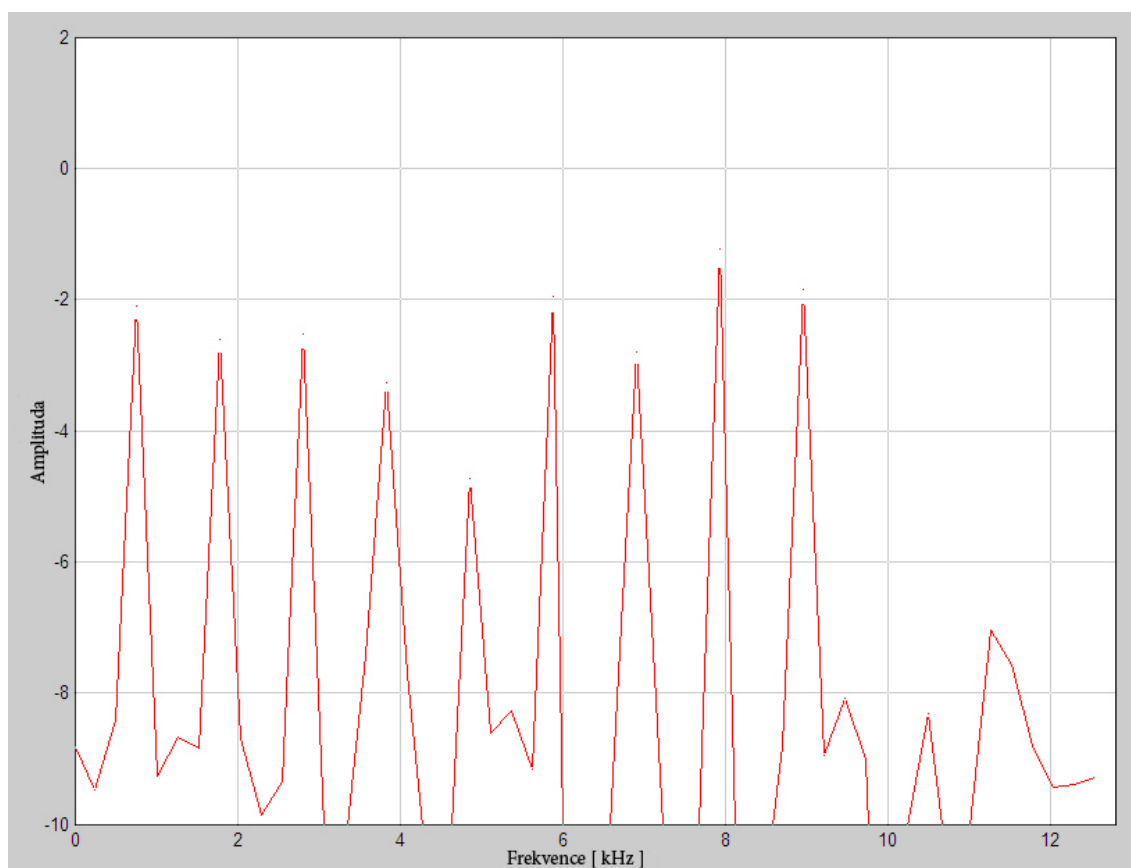


Obr. 3.2: Blokové schéma modulu předzpracování kodeku AMR-WB.

Jako zdroj signálu jsem použil modul náhodného signálu Random Source a k získání hodnot z frekvenčního pásma jsem použil moduly Vektor scope. Všechny parametry simulace a modulů byly nastaveny dle části 2.1.3.



Obr. 3.3: Frekvenční průběh vstupního signálu.



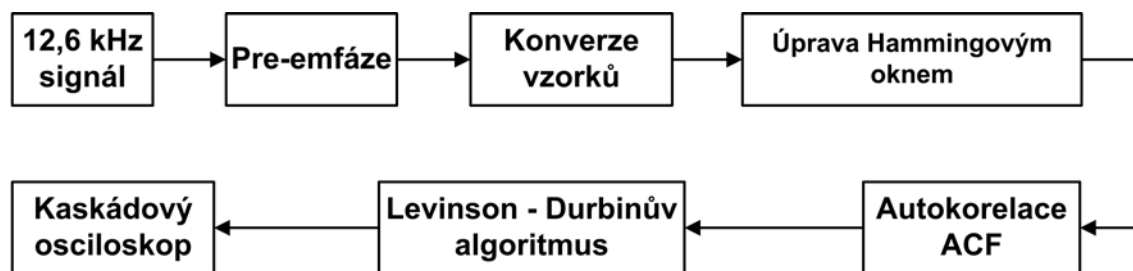
Obr. 3.4: Frekvenční průběh předzpracovaného signálu.

3.0.11 Simulace LPC analýzy

V této části jsem se soustředil na část ACELP kodéru, konkrétně na LP analýzu (viz. blokové schéma na obr. 3.5), jež je jednou ze součástí bloku analýzy rámce. V první řadě bylo třeba upravit vstupní signál, jež má mít po podvzorkování frekvenci 12,8 kHz. Toto jsem, pro méně komplikované řešení, provedl použitím wav souboru, jehož vzorkovací frekvence byla dle výše uvedeného požadavku. Poté byl signál upraven modulem pre-emfáze, jehož účelem je zvýraznit vyšší frekvence. Je zde použit jednoduchý filtr prvního řádu v tomto tvaru:

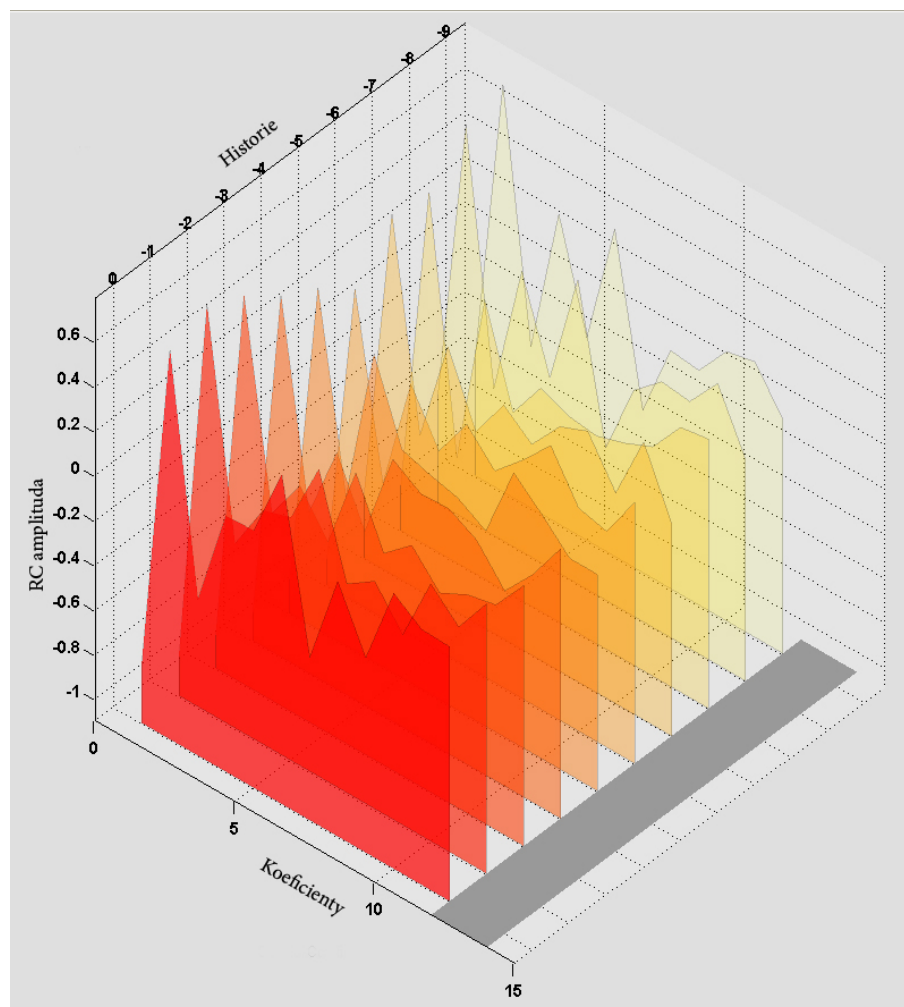
$$H(z) = 1 - 0,68z^{-1}$$

Dalším modulem ve kterém je signál upraven, je modul segmentace na rámce. Zde je řečový signál rozdělen na rámce o délce 20 ms a zároveň se stává stacionárním, což je nutné pro metodu odhadu parametrů. Funkce váhové posloupnosti Hammingova okna, kterou je tento signál následně upraven, výrazně ovlivní vlastnosti filtru jako je strmost v přechodovém pásmu nebo útlum v nepropustném pásmu. V našem případě např. bude dosaženo vrcholu postranního laloku v úrovni -41 dB.



Obr. 3.5: Blokové schéma analýzy lineární predikce.

Autokorelační funkce umožní nalezení koeficientů jedenáctého řádu. Modul Levinson-Durbin poté vypočítá z autokorelačních koeficientů parametry řečového signálu. Na výstupu je pro zobrazení parametrů neboli reflexních koeficientů použit kaskádový osciloskop. Tyto koeficienty jsou poté aplikovány v perceptuálním filtru, kdy počet koeficientů určuje řád i přenosovou funkci tohoto filtru.



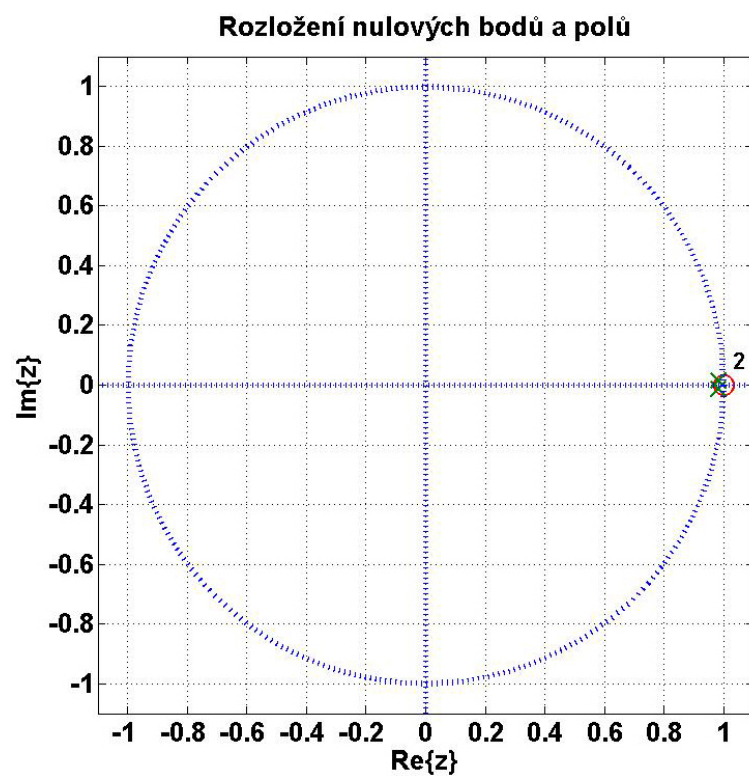
Obr. 3.6: Grafické zobrazení výstupu LP analýzy.

3.0.12 Návrh modulu horní propusti

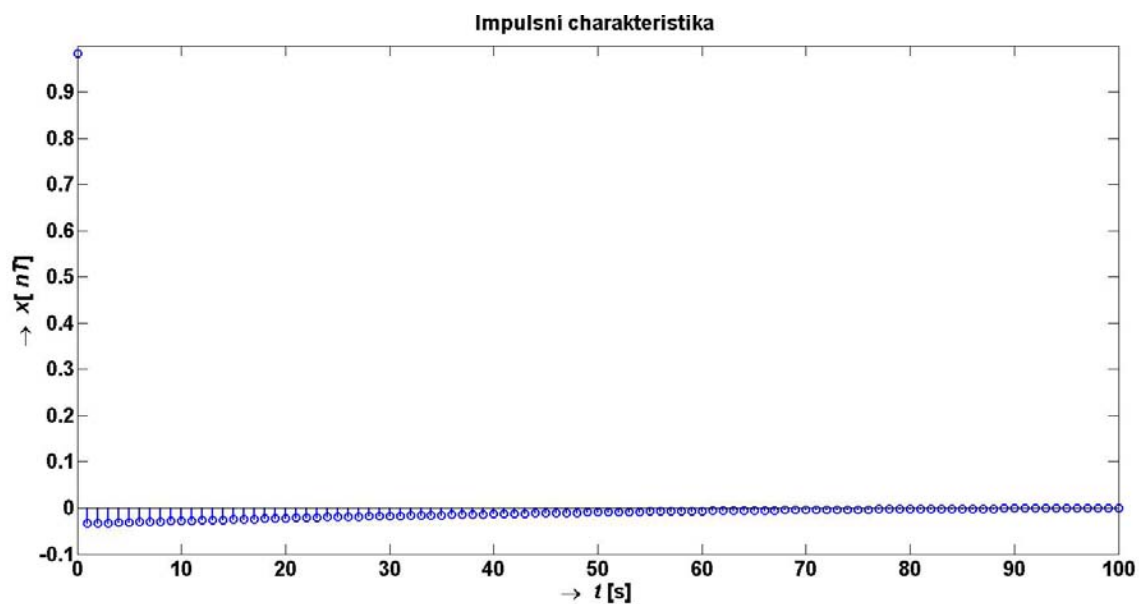
Tato část je zaměřena na návrh filtru typu horní propust, která je použita u tohoto kodeku k potlačení nežádoucích nízkofrekvenčních složek. Jeho přenosová funkce je dle [4]:

$$H(z) = \frac{0,982910156 - 1,965820313z^{-1} + 0,982910156z^{-2}}{1 - 1,965820313z^{-1} + 0,966308593z^{-2}} \quad (3.1)$$

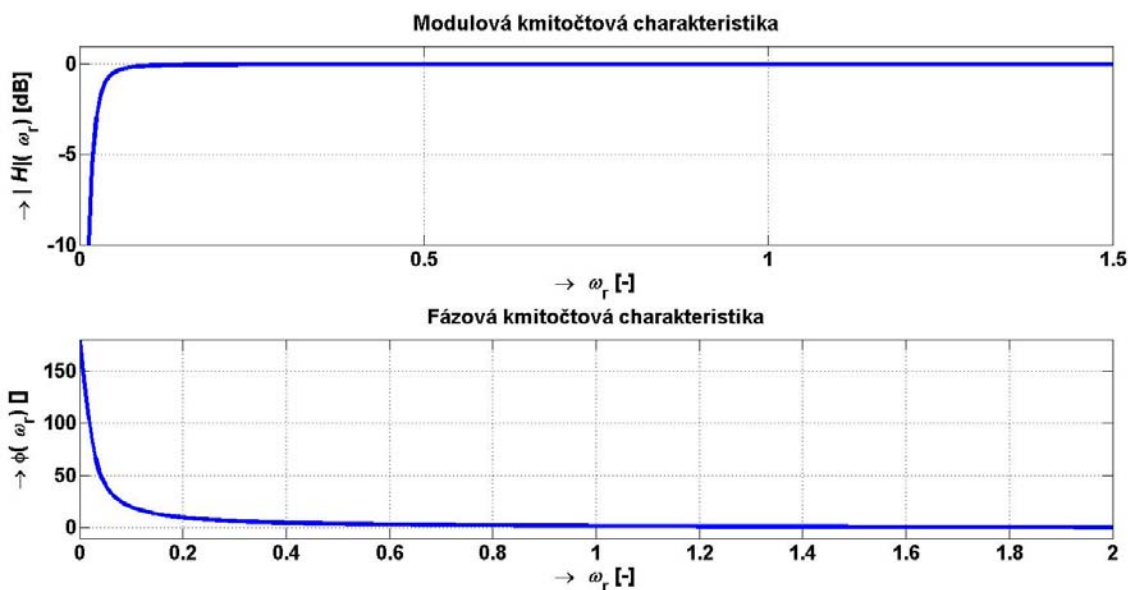
přičemž mezní kmitočet tohoto filtru je 50 Hz. Výpočet a vykreslení jednotlivých charakteristik jsem provedl pomocí funkcí v aplikaci Matlab pro to určených. Prvním krokem bylo zjištění rozložení nulových bodů a pólů pomocí funkce *zplane*. V našem případě se jedná o filtr IIR, jehož přenosová funkce má nulové body i póly. To, že se jedná o IIR filtr dokazuje také nelineární fázová charakteristika, jež je typická pro tento druh filtru. U toho filtru je navíc nutné ověřit stabilitu, jež spočívá v tom, že póly leží uvnitř jednotkové kružnice. V našem případě se jedná o stabilní filtr s póly poblíž okraje kružnice. Další krok spočíval ve výpočtu impulsní charakteristiky pomocí funkce *impz*, pomocí které zjistíme odezvu filtru na jednotkový impuls. Pro IIR filtr je typická nekonečná impulsní charakteristika viz. [9]. Průběh kmitočtové charakteristiky, jež udává přenos harmonického signálu v ustáleném stavu, jsem zjistil pomocí funkce *freqz*.



Obr. 3.7: Zobrazení nulových bodů a pólů na jednotkové kružnici.



Obr. 3.8: Zobrazení impulsní charakteristiky.



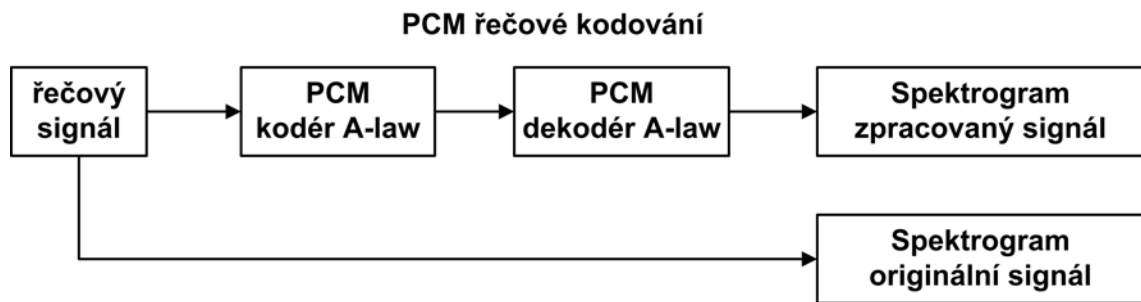
Obr. 3.9: Zobrazení modulové a fázové frekvenční charakteristiky.

3.0.13 Porovnání kódovacích technik PCM a CELP v prostředí Simulink

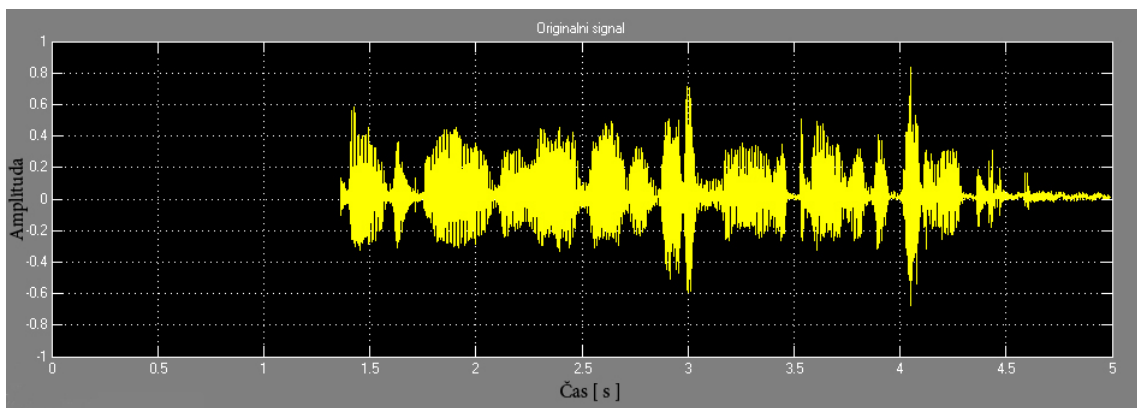
V této části byly porovnávány kódovací techniky, jež se používají při kódování řeči:

PCM kódování

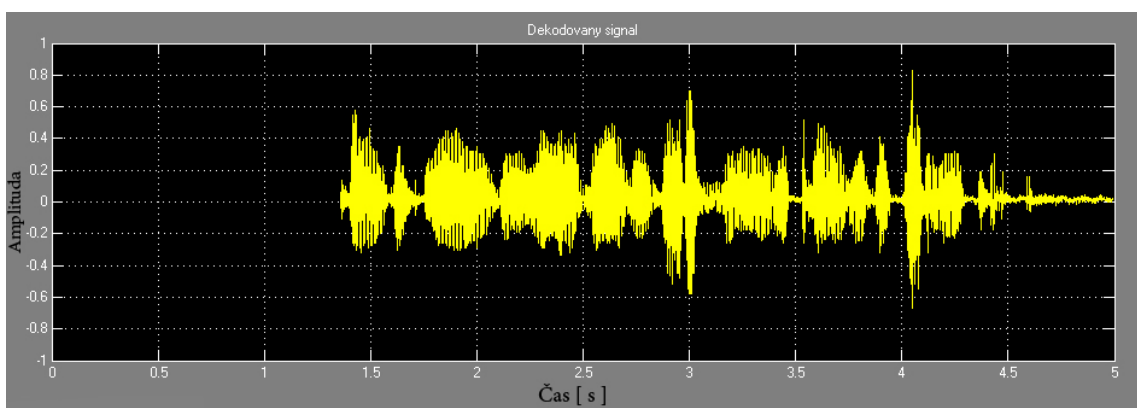
V modelu aplikace Simulink jsem vytvořil blokové schéma PCM kódování (G.711). Vstupní signál bylo třeba nejdříve převést na jiný datový typ, čehož jsem docílil pomocí modulu pro konverzi datového typu. Kodérem byl poté kódován řečový signál z knihovny Signal Processing a pro kódování byla použita logaritmická PCM evropské normy A-law. Výstupní přenesený dekódovaný signál bylo nutné opět převést na jiný datový typ. Před vstup osciloskopu byl ještě vložen modul konverze rámce na skalární vzorky a na osciloskopu byl tento signál porovnáván s originálním řečovým signálem. Z těchto průběhů je zřejmé, že signál byl přenesen s minimálním zkreslením, ale za cenu větších nároků na šířku kmitočtového pásma.



Obr. 3.10: Blokové schéma zobrazení průběhů signálů u kódování PCM.

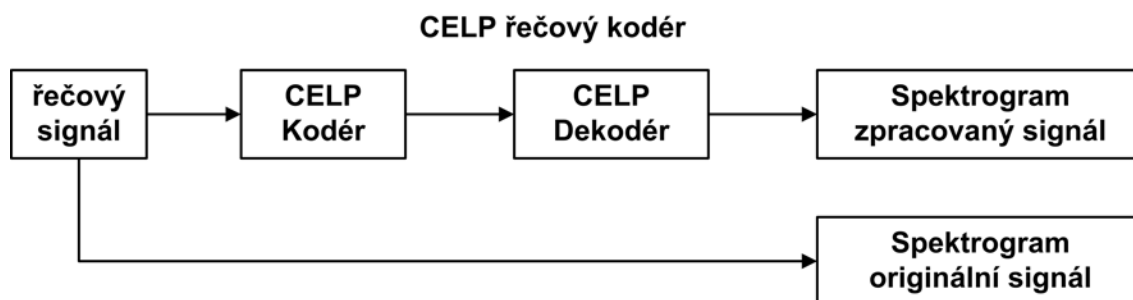


Obr. 3.11: Časový průběh vstupního originálního signálu u kódování PCM.

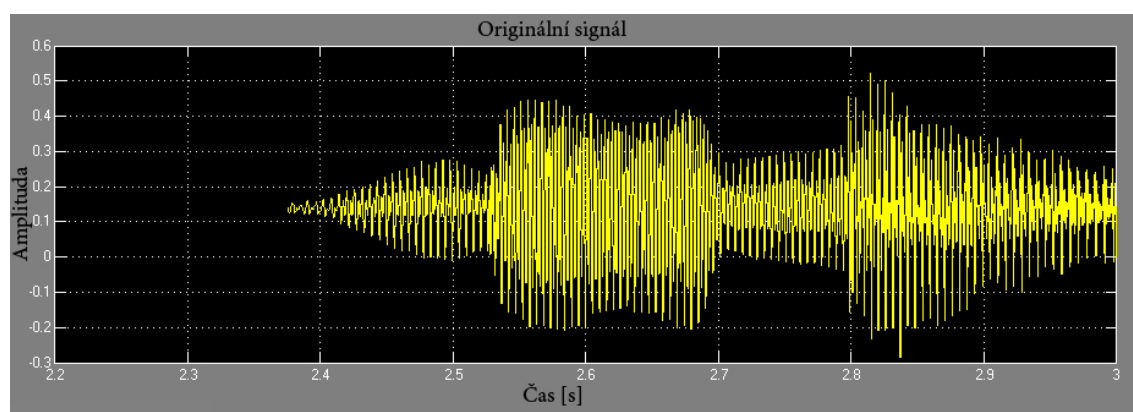


Obr. 3.12: Časový průběh výstupního dekódovaného signálu u kódování PCM.

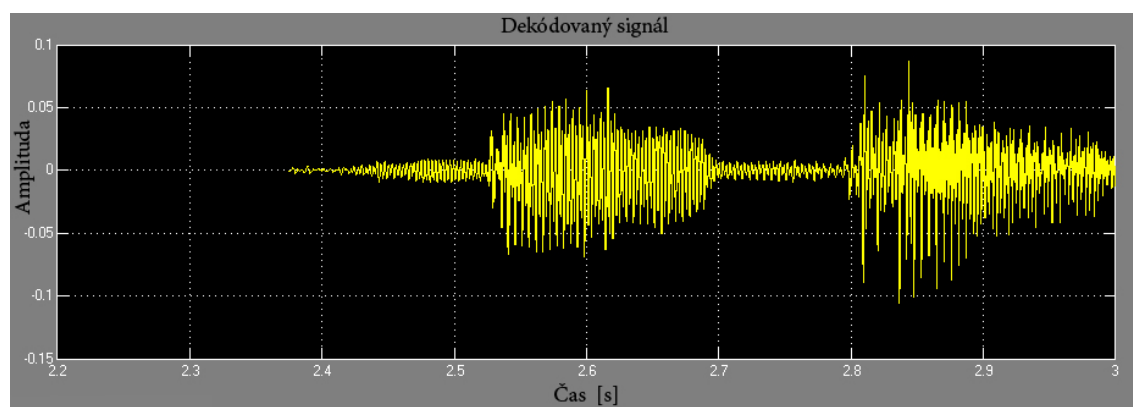
CELP kódování



Obr. 3.13: Blokové schéma zobrazení průběhů signálů u kódování CELP.



Obr. 3.14: Časový průběh vstupního originálního signálu kódování CELP.



Obr. 3.15: Časový průběh výstupního dekódovaného signálu kódování CELP.

V této simulaci jsem opět porovnával vstupní a výstupní signál, tentokrát za použití kódovací metody CELP. Použil jsem moduly CELP kodér a CELP dekodér z knihovny Simulink, za něž jsem opět zařadil modul konverze rámce na skalární

vzorky, přičemž vstupní a výstupní signály byly zobrazovány na osciloskopu. Z časových průběhů a subjektivního poslechu je patrné, že v oblastech s nízkou energií zní výstupní řeč poněkud strojově, což je způsobeno zokrouhlováním při vektorovém kvantování v procesu kódování.

4 ZÁVĚR

V této bakalářské práci jsem shrnul současně používané kódovací techniky v pevných i mobilních telekomunikačních sítích. Z tohoto rozboru je zřejmé, že s rozvojem polovodičových technologií v druhé polovině 20. století bylo možné vyvíjet dokonalejší a výkonnější přenosové prostředky. Takto získaná přenosová kapacita umožnila použití dokonalejších kódovacích technik, které se vyznačují aplikací složitých algoritmů a tím spojené výpočetní náročnosti.

Zaměřil jsem se zejména na perspektivní metodu AMR-WB (Adaptive Multi-Rate Wide-Band), která navazuje na úspěšnou aplikaci úzkopásmového kodeku AMR-NB v mobilních sítích druhé generace. Obě tyto metody používají kódování pomocí lineární predikce s buzením algebraické kódové knihy neboli ACELP kódování. Jeho podrobný rozbor poskytuje detailnější popis této, v současné době nejperspektivnější techniky kódování řeči s ohledem na vyžadovanou šířku pásma. Současně byl též brán zřetel na míru hodnocení kódované řeči lidským sluchem a možné vysokorychlostní přenosy.

Z předchozí analýzy vyplývá, že další budoucí vývoj bude směřovat k metodám založeným na technologii kódování ACELP. Proto byla zaměřena pozornost na detailní popis jednotlivých kroků uvedených algoritmů, matematický rozbor jednotlivých částí a konečně na praktickou realizaci v programovém prostředí aplikace MATLAB, konkrétně v simulačním prostředí Simulink.

Ze srovnání se standardními technikami používanými v současnosti vyplývá, že výše uvedený kodek AMR-WB umožňuje dosažení lepší srozumitelnosti při větších kódovacích poměrech, využití širšího přenosového pásma a dosahuje vyšší kvality přenášeného hovorového signálu. Značnou výhodou se jeví též fakt, že tento kodek lze aplikovat jak u bezdrátových sítí, tak i pro metalická vedení. V době konvergence telekomunikačních a počítačových (datových) sítí je ideální pro nasazení např. pro VoIP nebo multimediální aplikace.

Určitou nevýhodou může být vyšší výpočetní náročnost, jež však nemusí znamenat závažný problém, neboť dramatický vývoj číslicových signálových procesorů v posledních několika letech umožní realizaci těchto metod za poměrně nízké ceny.

Praktická implementace tohoto kódování bude řešena v případné diplomové práci navazující na toto téma.

LITERATURA

- [1] SMÉKAL, Z. *Číslicové zpracování signálů*, Skriptum MSCI, Ústav telekomunikací FEKT VUT v Brně, 2008.
- [2] SMÉKAL, Z. *Číslicové zpracování řeči*, Skriptum MZPR, Ústav telekomunikací FEKT VUT v Brně, 2008.
- [3] SALAMI, R. *The Adaptive Multi-Rate Wideband Speech Codec (AMR-WB)*, Skriptum, VoiceAge Corporation, Montreal, Canada.
- [4] SASSAN, A. *Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB)*, 2005.
- [5] VALIN, J.M. *Speex: A Free Codec For Free Speech*, CSIRO ICT Centre, Marsfield, Australia, 2002.
- [6] ŠIMÁK, B. *Principy zpracování hlasu v klasické a IP telefonii*, ČVUT FEL v Praze, 2004.
- [7] ČERNOCKÝ, J. *Zpracování řečových signálů*, Ústav počítačové grafiky a multimédií FIT VUT v Brně, 2006.
- [8] HOLUB, J. – ŠMÍD, R. – KOLOMAZNÍK, J. *Hodnocení hlasových přenosů v telekomunikačních sítích*, Sdělovací technika, 2001.
- [9] SYSEL, P. *Číslicové filtry*, Skriptum, Ústav telekomunikací FEKT VUT v Brně, 2004.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

3G The 3rd Generation

3GPP The 3rd Generation Partnership Project

ACELP Algebraic Code Excited Linear Prediction

ADM Adaptive Delta Modulation

ADPCM Adaptive Diferencial Pulse Code Modulation

AMR-WB Adaptive Multi-Rate Wide Band

AMR-NB Adaptive Multi-Rate Narrow Band

APC Adaptive Prediction Coding

ATC Adaptive Transform Coding

C/I Carrier to Interference Ratio

CELP Code Excited Linear Prediction

CNG Comfort Noise Generation

DM Delta Modulation

DPCM Diferencial Pulse Code Modulation

DSP Digital Signal Processing

DTX Discontinuous Transmission

EFR Enhanced Full Rate

ETSI The European Telecommunications Standards Institute

FR Full Rate

GSM Global System for Mobile Communications

HR Half Rate

ISPP Interleaved single-pulse permutation

LP Linear Prediction

LPF Linear Prediction Filter

LD-CELP Low Delay Code Excited Linear Prediction

MOS Mean Opinion Score

PCM Pulse Code Modulation

PESQ Perceptual Estimation of Speech Quality

PSQM Perceptual Speech Quality Measure

QR Quarter Rate

QoS Quality of Service

RELP Residually Excited Linear Prediction

RPE-LP Regular Pulse Excitation Long Prediction

SCR Source Controlled Rate

SC Source Coding

SID Silence Descriptor

SNR Signal-to-Noise Ratio

VAD Voice Activity Detector

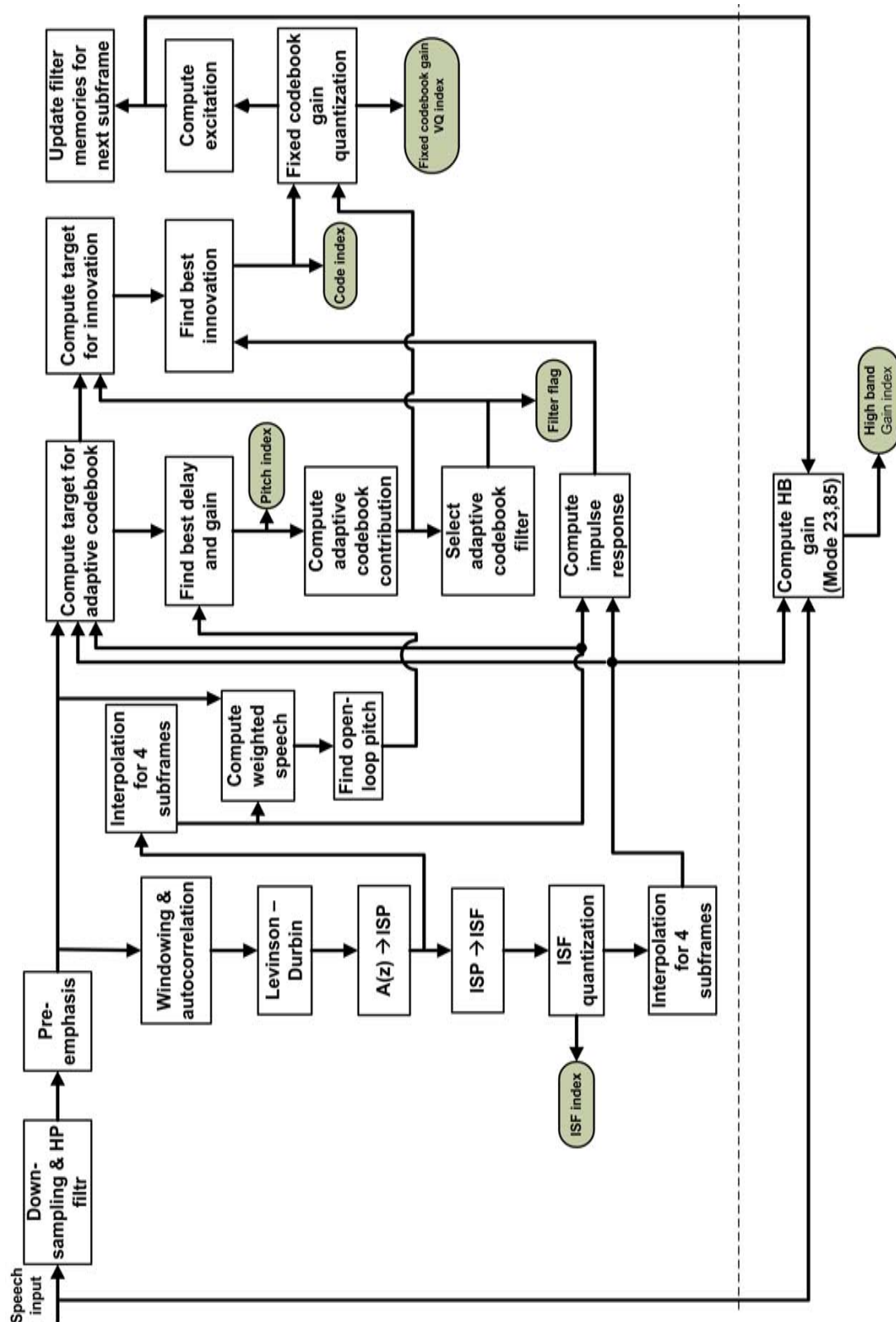
VMR-WB Variable Multi-Rate Wide Band

VoIP Voice over Internet Protocol

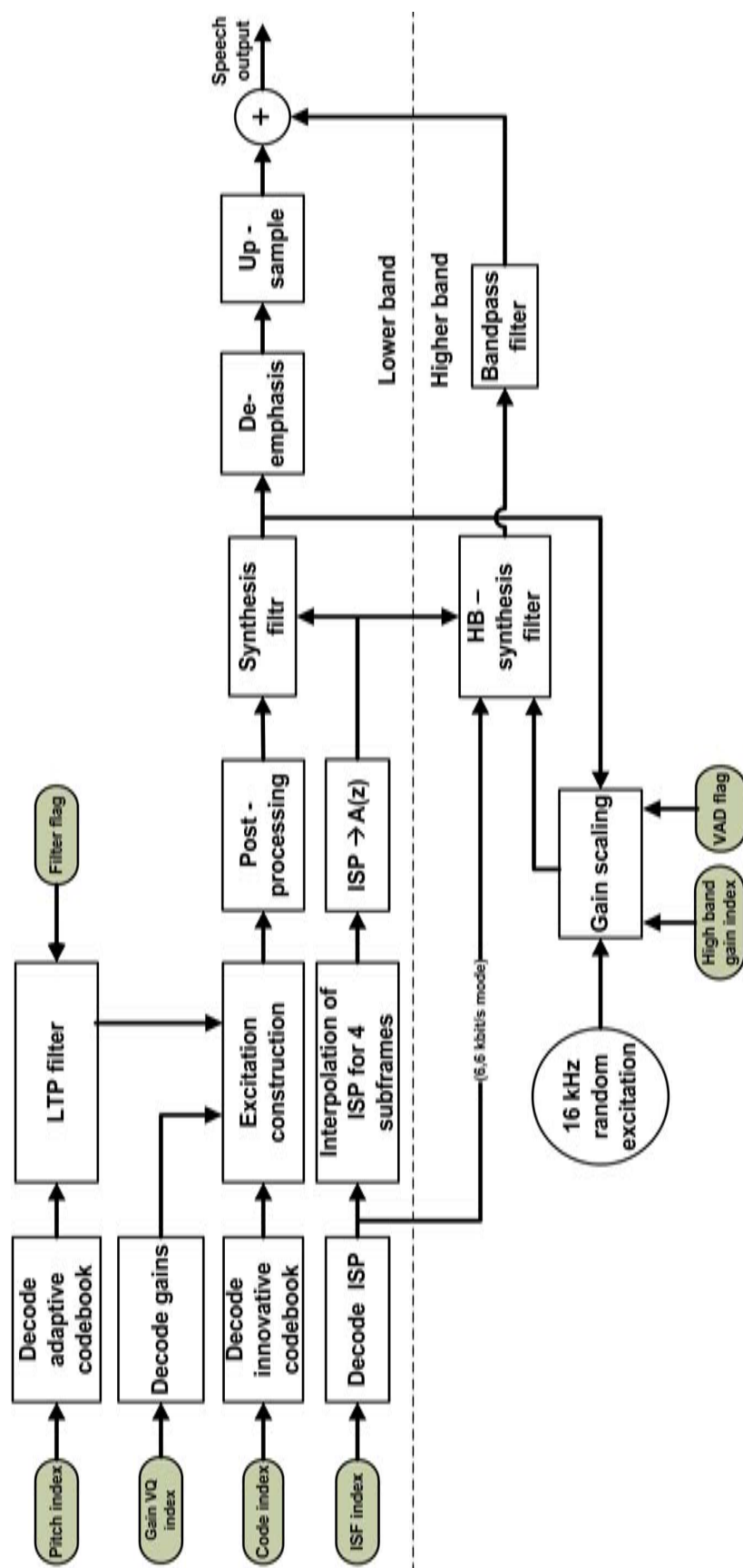
WC Wave Coding

WDCMA Wideband Code Division Multiple Access

A BLOKOVÁ SCHÉMATA AMR-WB



Obr. A.1: Blokové schéma AMR-WB ACELP kodéru.



Obr. A.2: Blokové schéma AMR-WB ACELP dekodéru.

B OBSAH CD

B.0.14 Elektronické dokumenty

`bachelor.pdf` - kompletní bakalářská práce v elektronické podobě

`zadani.pdf` - zadání bakalářské práce

B.0.15 Zdrojové kódy aplikace MATLAB

`navrh-HP.m` - návrh horní propusti

B.0.16 Zdrojové kódy aplikace Simulink

`Downsampling.mdl` - předzpracování signálu

`lpc-analysis.mdl` - LPC analýza

`pcmencoding.mdl` - PCM kódování

`celpencoding.mdl` - CELP kódování