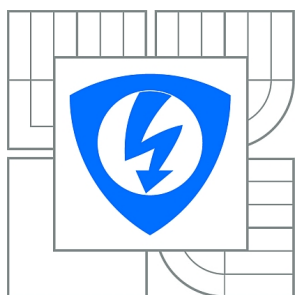


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

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 MOBILNÍ SÍTI

MODERN METHODS OF SPEECH CODING FOR MOBILE NETWORK TRANSMISSION

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

VOJTĚCH KRÁSNÝ

VEDOUCÍ PRÁCE

SUPERVISOR

prof. Ing. ZDENĚK SMÉKAL, CSc.

BRNO 2013



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor
Teleinformatika

Student: Vojtěch Krásný

ID: 125254

Ročník: 3

Akademický rok: 2012/2013

NÁZEV TÉMATU:

Moderní metody kódování řeči pro přenos po mobilní 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 systémů, jejich výhody, nevýhody a způsob použití. Vybrané algoritmy je nutné ověřit v Matlabu.

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 NMew CDMA2000 Wideband Speech Coding Standard. In IEEE Communication Magazine, May 2006, pp.74-81.

Termín zadání: 11.2.2013

Termín odevzdání: 5.6.2013

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

Konzultanti bakalářské práce:

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 části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Tato bakalářská práce se věnuje popisu metod kódování řečového signálu používaných především v 2G a 3G mobilních sítích, popisuje jejich vývoj a zaměřuje se na detailnější popis v současné době nejperspektivnějšího způsobu kódování založeného na technologii ACELP, kodeku AMR-WB.

KLÍČOVÁ SLOVA

2G, 3G, algoritmus, amr-wb, kodek, metody, mobilní

ABSTRACT

This bachelor's thesis describes speech's signals coding techniques used especially in 2G and 3G cellular networks, describes their evolution and aims in great detail at today's most perspective way of coding based on ACELP technology, codec AMR-WB.)

KEYWORDS

2G, 3G, algorithm, amr-wb, codec, mobile, technique

KRÁSNÝ, Vojtěch *Moderní metody kódování řeči pro přenos po mobilní síti*: bakalářská práce. BRNO: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2013. 43 s. Vedoucí práce byl prof. Ing. Zdeněk Smékal, CSc.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Moderní metody kódování řeči pro přenos po mobilní 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/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

BRNO

.....

(podpis autora)

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu Ing. Zdeňku Smékalovi, CSc za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

BRNO

.....

(podpis autora)



Faculty of Electrical Engineering
and Communication
Brno University of Technology
Purkynova 118, CZ-61200 Brno
Czech Republic
<http://www.six.feec.vutbr.cz>

PODĚKOVÁNÍ

Výzkum popsany v této bakalářské práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

BRNO

.....

(podpis autora)



EVROPSKÁ UNIE
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ
INVESTICE DO VAŠÍ BUDOUCNOSTI



OP Výzkum a vývoj
pro inovace

OBSAH

Úvod	8
1 Kódování řečového signálu	9
1.1 Řečový signál	9
1.2 Základní pojmy a druhy kódování	9
1.2.1 Základní pojmy	10
1.2.2 Rozdělení kodérů	10
2 Kodér typu AMR-WB	14
2.1 Úvod	14
2.2 Evoluce kódování řeči v GSM a WCDMA 3G sítích	15
2.3 Popis algoritmu hlasového kodéru	16
2.4 Princip kodeku AMR-WB	17
2.5 Percepční váhování	18
2.6 Analýza základního tónu	20
2.7 Struktura a vyhledávání v algebraické fixní kódové knize	22
2.8 Algoritmický popis řečového dekodéru	28
2.9 Složitost kodeku	32
2.10 Adaptivní provoz v GSM kanále	32
2.11 Detekce hlasové aktivity a přerušované vysílání	33
2.12 Kvalita řeči	34
3 Simulace vybraných bloků	35
3.1 Úvod	35
3.2 Simulace předzpracování a LPC analýzy	35
4 Závěr	38
Literatura	39
Seznam symbolů, veličin a zkratk	41
A Obsah CD	43

ÚVOD

Tato práce se věnuje popisu metod kódování řečového signálu používaných především v 2G mobilních sítích, popisuje jejich vývoj a zaměřuje se na detailnější popis v současné době nejperspektivnějšího způsobu kódování založeného na technologii ACELP.

Hlavní částí práce je rozbor širokopásmového kodeku AMR-WB, který poskytuje výrazně lepší kvalitu přenášené řeči oproti používaným kodekům, při současně nízké bitové rychlosti, která se mění podle kvality spojení a zatížení sítě.

V závěrečné části práce bude nasimulována část algoritmu kodeku zahrnující předzpracování signálu a LPC analýzu.

1 KÓDOVÁNÍ ŘEČOVÉHO SIGNÁLU

1.1 Řečový signál

Mezi základní způsoby lidské komunikace patří verbální projev, reprezentovaný řečí v podobě různých jazyků. Řečový signál je tvořen soustavou orgánů, z nichž každý plní konkrétní funkci. Základem jsou plíce, které dodávají proud vzduchu řečovým orgánům, které tvoří soustava hlasivky, hrdelní dutina, ústní dutina, nosní dutina, měkké a tvrdé patro a také zuby. Proud vzduchu z plic při průchodu hlasivkami způsobí jejich rozkmitání, které vyvolá periodický zvuk. Kmitočet tohoto zvuku závisí na tlaku vzduchu a napětí hlasivek a je označován termínem základní tón řeči. Tento kmitočet se u lidí liší podle věku a pohlaví a pohybuje se obvykle v intervalu 150 až 400Hz.

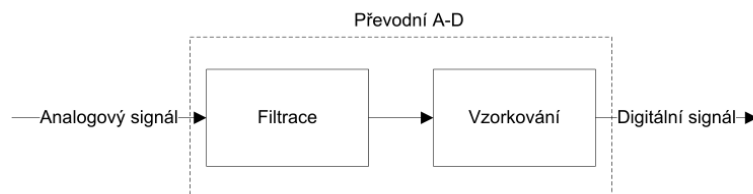
Lidská řeč je diskrétní signál, který lze vyjádřit jako posloupnost konečným počtem hlásek. Tento počet je v každém jazyce různý, ale obvykle se pohybuje v rozmezí 30 až 50 fonémů. Řečový signál je snímán mikrofonom, kde je modulovaný proud vzduchu převáděn na elektrický signál.[1]

Při zpracování hlasového záznamu se používá kódování, aby bylo dosaženo níže uvedených parametrů[2]:

- minimální přenosová rychlost
- co nejmenší zkreslení původního hlasového signálu
- minimální výpočetní náročnost a algoritmické zpoždění
- univerzálnost použití - nezávislost na jazyku, okolním ruchu, poruchách při přenosu

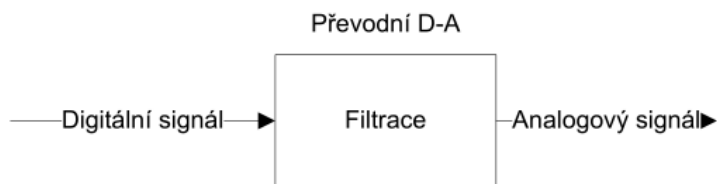
1.2 Základní pojmy a druhy kódování

ADC - analogově digitální převodník (analog to digital converter)



Obr. 1.1: Schéma AD převodníku[3]

DAC - digitálně analogový převodník (digital to analog converter)



Obr. 1.2: Schéma DA převodníku[3]

1.2.1 Základní pojmy

Vzorkování - Sbírání hodnot signálu v daných časových intervalech. Vzorkovací kmitočet musí splňovat Shannonův teorém, který říká, že přesná rekonstrukce spojitého, frekvenčně omezeného signálu je možná pouze v případě, že je vzorkovací frekvence alespoň dvojnásobkem nejvyšší harmonické složky vzorkovaného signálu.

Kvantování - Jedná se o diskretizaci hodnot signálu a přiřazení jedné diskrétní úrovně každému vzorku. Tento proces je vždy ztrátový a nevratný. Jsou-li kvantizační hladiny od sebe rozmístěny rovnoměrně, nazýváme je ekvidistantní. V některých případech, typicky u malých a hovorových signálů, je výhodnější použít logaritmické kvantování, kdy je více hladin umístěno v oblastech s menší amplitudou a méně hladin v oblastech s větší amplitudou. Rozdíl mezi diskrétní hodnotou úrovně porovnaný se zdrojovým spojitým signálem jsou označovány jako kvantizační šum.[1, 3]

1.2.2 Rozdělení kodérů

Kódování	Vlastnosti
Kódování tvaru vlny (Waveform coders)	kódování vzorek po vzorku v časové oblasti, vysoká kvalita, velký bitový tok vhodný pro řeč i hudbu
Parametrické (zdrojové) kódování	složitější než kodéry tvaru vlny, nízká bitová rychlost, vhodné pouze pro řeč - syntetická reprodukce
Hybridní kodéry	využívají kombinace předešlých vlastností

Tab. 1.1: Přehled kódování a jejich vlastností[3]

Druh kódování	Kodek
Kódování tvaru vlny (Waveform coders)	PCM (Pulse Code Modulation) DPCM (Difference Pulse Code Modulation) ADPCM (Adaptive Difference Pulse Code Modulation)
Parametrické (zdrojové) kódování	LPC (Linear Predictive Coding) - LPC-10
Hybridní kodéry	CELP

Tab. 1.2: Přehled kodeků podle druhu kódování[3]

Kódování tvaru vlny

PCM - pulzní kódová modulace

Jedná se o nejjednodušší kodér tvaru vlny. Před vstupem do kodéru je signál filtrován filtrem typu dolní propust a poté je vzorkován. V telekomunikacích používá vzorkovací kmitočet 8 kHz a získané vzorky přiřazuje k nejbližší kvantovací hladině a podle té jim je přiřazen binární kód. Počet bitů kódového slova a počet kvantovacích hladin určuje vztah:

$$N_C = 2^N \quad (1.1)$$

kde N_C označuje počet kvantovacích hladin a N je počet bitů na jeden symbol. Při 8 bitovém kvantování a 8 kHz vzorkovací frekvenci dává datový tok 64 kbit/s, základní kapacitu hovorového kanálu v digitálních sítích.

Pokud je vzdálenost jednotlivých kvantovacích úrovní stejná, pak hovoříme o lineární PCM, u které je třeba pro minimalizaci kvantizační chyby použít více kvantizačních úrovní, čímž se zvýší přenosová rychlost a požadavky na kapacitu přenosového kanálu.

Tento jev lze u hovorového signálu omezit použitím nelineární PCM, která nepoužívá stejně velké kvantizační kroky a jejich velikost se s rostoucí amplitudou vstupního signálu zvětšuje. Pro úpravu hodnot signálu je použit kompresor na vysílací straně, který zesiluje vzorky malých hodnot a zeslabuje signály vysokých hodnot. Na přijímací straně je umístěn expandor, který dělá opak a vrací vzorkům původní rozsah.

Pro PCM jsou dle ITU-T definovány dva mírně odlišné kvantizéry. Logaritmický kvantizér používaný v Severní Americe a Japonsku nese označení μ -Law, Evropská norma nese označení A-Law. Rozdíl mezi severoamerickou a evropskou normou spočívá v rozdílném počtu bitů využitých pro přenos signálu, kdy severoamerická norma používá 7 bitů a evropská 8.[3]

DPCM - diferenční pulzní kódová modulace

Kodek DPCM vychází z PCM a byl publikován v roce 1952. Systém využívá podobnosti mezi jednotlivými vzorky a kóduje pouze rozdíl mezi kvantovaným vzorkem a jeho predikovanou hodnotou. To umožňuje snížit počet bitů potřebných k zakódování signálu a množství přenášených dat. Využití nachází především při kódování obrazu.[1, 3]

ADPCM - adaptivní diferenční pulzní kódová modulace

ADPCM rozšiřuje DPCM a zvyšuje účinnost prediktivního kódování adaptací systému na základě parametrů vstupního signálu. Adaptován může být kvantizátor, prediktor, nebo obojí zároveň. Kodekem ADPCM lze zakódovat řečový signál rychlostí 32 kb/s ve kvalitě srovnatelné s PCM 8 bit s přenosovou rychlostí 64 kb/s. ADPCM je využito u kodeků řeči G.721, G.726 a G.727.[1, 3]

Parametrické(zdrojové) kódování hovorových signálů

LPC - lineární predikce

Pro kódování pomocí zdrojového kódování je třeba vytvořit vhodný model hlasového traktu, který bude odpovídat způsobu generování hlasového signálu člověkem. Lidskou řeč lze chápat jako odezvu na vydechaný proud vzduchu z plic.

Příkladem parametrického systému je tzv. vokodér (voice coder). Vstupní signál je ve vokodéru nejprve digitalizován a segmentován na úseky dlouhé 10 až 30 ms, které jsou dále považovány z hlediska popisu za stacionární, a pro každý segment zvlášť je určen soubor parametrů (deskriptorů). Pro vyjádření hodnot koeficientů se nejčastěji používá metoda založená na lineární predikci LPC (Linear Prediction Coding). Stanovené deskriptory popisují tyto parametry:

- znělost/neznělost konkrétního segmentu
- perioda základního tónu
- zisk G
- koeficienty syntetizujícího filtru

Parametry stanovené analýzou jsou multiplexovány v čase a jsou přenášeny v digitální podobě na přijímací stranu. V syntetizátoru je zvuk vytvářen na základě buzení periodickou posloupností v případě znělých segmentů, v případě neznělých segmentů je zvuk generován náhodnou posloupností s rovnoměrným spektrem. Po úpravě buzení podle hodnoty zisku G vstupuje signál do syntetizujícího filtru na jehož výstupu je hovorový signál.[1, 3]

Kvalitu generovaného zvuku určují vlastnosti hlasového syntetizátoru. Vokodéry dosahují velmi nízké přenosové rychlosti cca 2,5 kbit/s, ale jejich nevýhodou je syntetický zvuk.

Hybridní kodéry

Hybridní kodéry využívají předností předchozích dvou typů kodérů. Součástí kodéru je i dekodér vytvářející syntetizovaný signál, který se odečítá od vstupního signálu. Tato odchylka se minimalizuje na základě smyslového (perceptuálního) vnímání. [10]

CELP - Code-Excitation Linear Prediction

Jedná se o kombinaci LPC a vektorového kvantování. Z jednotlivých segmentů se získávají potřebné budící parametry a ty se porovnávají s kódovou knihou, která obsahuje vzorové budící posloupnosti. Při porovnání je vybrána budící posloupnost s nejlepší shodou a dále se odesílá pouze její adresa v kódové knize. Příjímač poté podle přijaté adresy generuje příslušnou budící posloupnost pro hlasový syntetizátor.

ACELP - Algebraic Code-Excitation Linear Prediction

Používá algebraickou strukturu kódové knihovny a umožňuje rychlé vyhledávání v rozsáhlých knihovnách. V současné době je nejrozšířenějším algoritmem používaným v mobilních systémech. Algoritmus je použit kodérem k vytvoření náhodných součástí buzení, kde budící vektor obsahuje několik nenulových impulzů s předdefinovanými skupinami pozic. Amplituda každého impulzu je v intervalu $[1, -1]$ a každý impulz může nabývat skupinu pozic, odlišnou od pozic ostatních impulzů. Skupiny pozic jsou prokládané a budící kód je identifikován na základě pozic nenulových impulzů. Vyhledávání v kódové knize je prakticky vyhledávání optimálních pozic nenulových impulzů. [11]

2 KODÉR TYPU AMR-WB

2.1 Úvod

Současné sítě druhé a třetí generace používané v mobilních komunikačních systémech pracují s šířkou pásma omezenou na 200-3400 Hz. S vývojem mobilních systémů ustupují do pozadí telefonní hlasové služby ve prospěch multimediálních služeb a vysokorychlostních datových přenosů. Rozšíření šířky pásma na 50-7000 Hz pro hovorový signál nabízí podstatné zvýšení kvality a věrnosti řeči hovorového signálu. Při porovnání se současnými sítěmi využívajícími úzké telefonní pásmo, nabízí rozšíření pásma od 50 do 200 Hz přirozenější zvuk a rozšíření pásma z 3400 na 7000 Hz poskytuje hlasovému signálu větší srozumitelnost a přirozenost projevu. Celková šířka pásma 50-7000 Hz mimo výše jmenované přispívá k jednoduššímu rozeznání mluvčích a pocitu transparentní komunikace.

Nedávné pokroky v oblasti kódování řeči umožnily širokopásmové kódování v přenosových rychlostech použitelných pro mobilní komunikaci. V roce 1999 začala skupina 3GPP (3rd Generation Partnership Project) společně s ETSI (European Telecommunication Standards Institute) vyvíjet standard širokopásmového kodeku pro WCDMA 3G a GSM systémy. Standardizaci předcházely dva roky během kterých byla testována proveditelnost a bylo vybíráno ze dvou kodeků, z nichž byl v prosinci roku 2000 vybrán širokopásmový. Specifikace kodeku byla schválena a dokončena v březnu 2001. Výsledný širokopásmový kodek je adaptabilní - schopný pracovat s přenosovými rychlostmi 6,6 až 23,85 kb/s. Dále je označován jako Adaptive Multirate Wideband (AMR-WB).[4]

Kodek AMR-WB obsahuje soubor pevných přenosových rychlostí a režimů kódování kanálu, detektor hlasové aktivity (Voice Activity Detector), funkci přerušovaného vysílání pro použití v GSM sítích (DTX), funkci ovládající bitrate zdroje (SCR) v sítích 3G a vnitropásmovou signalizaci a ovládání pro výběr vhodného režimu v závislosti na podmínkách přenosu. Kodek AMR-WB rozděluje šířku pásma mezi hovorové a kanálové kódování podle převládajících přenosových podmínek. I když kodek poskytuje lepší kvalitu hovoru než současné úzkopásmové standardy, tak je současně velmi odolný vůči chybám vzniklým při přenosu následkem přepínání mezi různými rychlostmi a přízpusobením podmínkám radiového kanálu. Kodek AMR-WB byl vyvinut pro využití v následujících sítích: GSM-FR, GERAN, UTRAN a také pro využití v paketových sítích pro přenos přes VoIP protokol.

V průběhu standardizace kodeku AMR-WB organizací 3GPP, probíhalo v Mezinárodní telekomunikační unii (ITU) řízení o standardizaci širokopásmového kodeku s přenosovou rychlostí kolem 16 kb/s. Navrhovaný standard měl být vhodný pro

široký rozsah použití zahrnující ISDN širokopásmovou telefonii a videokonference, VoIP aplikace umožňující videokonference, hlasovou poštu, chat a hlasové vysílání.

Na základě výsledků testů kodeku AMR-WB provedených skupinou 3GPP, zařadila Mezinárodní telekomunikační unie tento kodek do výběrového řízení o budoucím standardu. Významným faktorem ovlivňujícím šanci tohoto kodeku byla možnost součinnosti s 3G sítěmi. Tento kodek byl v červenci 2001 vybrán a byl uznán jako jediný harmonizovaný širokopásmový kodek pro GSM a WCDMA 3G. Následně byl v lednu 2002 schválen a označen jako doporučení ITU-T G.722.2.[5]

Přijetí AMR-WB Mezinárodní telekomunikační unií mělo velký význam, protože to bylo poprvé, kdy byl použit jednotný kodek pro bezdrátové i drátové sítě. Tímto odpadla nutnost překódování a byla usnadněna implementace širokopásmových hlasových služeb na širokou škálu komunikačních systémů a platforem.[4]

2.2 Evoluce kódování řeči v GSM a WCDMA 3G sítích

Prvním používaným hlasovým kodekem v GSM sítích byl 13 kb/s Full-Rate kodek, který byl standardizován v roce 1989 a používá se v GSM-FR (Full rate) kanálech, kde činí jeho bitová rychlost na fyzické vrstvě 22,8 kb/s. Tento kodek je standardním kodekem používaných v sítích GSM. V roce 1995 byl standardizován Half-Rate (HR) kodek pracující s bitovou rychlostí na fyzické vrstvě 11,4 kb/s. Tento kodek poskytuje stejnou kvalitu řeči jako FR kodek, ale při použití v tandemu (volání mobilní telefon - mobilní telefon) a při šumu na pozadí podává horší výkon.

Prvním kodekem který nabízel kvalitu hovoru srovnatelnou s drátovými telefony byl EFR (Enhanced Full-Rate), který přinesl podstatné zvýšení kvality oproti FR a HR kodekům. EFR nabízí tuto kvalitu při všech možných rádiových podmínkách až do poměru signál/šum 10 dB[6]. Tato kvalita byla vyžadována kvůli rostoucímu použití v podmínkách, kde byly obvykle používány pevné linky nebo bezdrátové telefony. Proto musel být standard GSM odolný vůči okolnímu hluku a chybám přenosu. Kodek EFR byl prvně standardizován v roce 1995 pro systém PCS 1900 (Personal Communication Service) založený na systému GSM, do kterého byl v roce 1996 začleněn. Kodek EFR byl vyvinut společně firmou Nokia a University of Sherbrooke.

Po uvedení kodeku EFR bylo stále co zlepšovat. Zejména v náročných podmínkách, kdy dochází k chybám v kanálu, lze zlepšit podávaný výkon změnou bitové alokace pro řeč a kanálové kódování. Stejně jako GSM-HR nebyl již schopen poskytovat požadovanou kvalitu. Dalším vývojovým stupněm standardizovaným v roce 1999 byl kodek AMR-NB, který opět dále posunul kvalitu hovoru v GSM sítích[4].

Tento kodek nabízí oproti EFR výrazné zvýšení odolnosti ve Full-rate kanálu, kdy upravuje řečové a kanálové kódování podle rádiových podmínek a dokáže udržet kvalitu srovnatelnou s drátovými telefony až do odstupů signálu od šumu 4-7 dB. Dobrou kvalitu hovoru dokáže při silném signálu udržet i při použití half-rate kanálu a tím ušetřit část přenosové kapacity. Kodek AMR zvládá pracovat s osmi bitovými rychlostmi od 4,75 do 12,2 kb/s. Kodek byl v roce 1999 skupinou 3GPP přijat jako výchozí kodek pro WCDMA 3G sítě a zároveň přijal EFR jako jeden ze svých módů. Byl vyvíjen společně firmami Ericsson, Nokia a Siemens. [4]

Zatím posledním kodekem, standardizovaným v roce 2001, je AMR-WB (Adaptive Multi-Rate Wideband), vyvíjený společnostmi Nokia a VoiceAge a určený pro GSM a WCDMA 3G systémy. Oproti všem předchozím kodekům rozšiřuje šířku pásma řečového signálu na 50-7000 Hz a tím výrazně zvyšuje kvalitu hovoru. Pracuje s devíti bitovými rychlostmi kódování řeči od 6,6 do 23,85 kb/s a stejně jako ostatní GSM a WCDMA 3G kodeky využívá pro kódování šumu v pozadí zdrojově závislý mód. [4, 7, 8]

2.3 Popis algoritmu hlasového kodéru

Kodek AMR-WB je založen na technologii ACELP (Algebraic code-excited linear prediction), technologii lineární predikce buzené algebraickým kódem[9]. Tato technologie je velmi rozšířená a je použita v mnoha standardech kódování hlasových signálů v mobilních sítích a VoIP, například v kodecích 3GPP AMR (TS 26.090),ETSI EFR (TS 06.60), NA-TDMA IS-641, NA-CDMA IS-127, ITU-T G.729 a G.723.1.

I když technologie ACELP poskytuje dobré výsledky při použití s úzkopásmovými signály, je potřeba jí pro dosažení dobrých výsledků s širokopásmovými signály přidat některé vlastnosti. Klasický model algoritmu ACELP využije obvykle nejvíce kódovacích bitů pro nízkofrekvenční oblast s větší energií, což vede k signálu oříznutému dolní propustí. K překonání tohoto problému je třeba modifikovat váhovací filtr k použití s širokopásmovými signály. Kromě toho rozsah základních tónů v hlasových segmentech širokopásmových signálů nepokrývá celý rozsah použitého spektra, z čehož plyne potřeba více variant oproti úzkopásmovým signálům. Proto je důležité vylepšit analýzu základního tónu pro přizpůsobení algoritmu více úrovním. Ze stejného důvodu je také důležité zlepšit techniky rozšíření periodicity v dekodéru. Dalším problémem, který vyvstává při kódování širokopásmových signálů, je nutnost použití rozsáhlých kódových knih. Proto je potřeba aby jejich struktura byla efektivní, nenáročná na paměť a bylo možné ji velmi rychle prohledávat.[4]

2.4 Princip kodeku AMR-WB

Kodek AMR-WB zahrnuje devět módů kódování řeči s přenosovými rychlostmi 23,85, 23,05, 19,85, 18,25, 15,85, 14,25, 12,65, 8,85 a 6,6 kb/s. Kodek obsahuje mód pro kódování šumu na pozadí, který je navržen pro provoz v módu přerušovaného vysílání (DTX) v GSM sítích a jako nízkobitratový zdrojově závislý mód pro kódování šumu na pozadí v ostatních systémech. Přenosová rychlost v sítích GSM v tomto módu činí 1,75 kb/s.

Módy od 12,65 kb/s jsou používány pro vysoce kvalitní širokopásmovou řeč. Naproti tomu dva nejnižší módy 8,85 a 6,6 kb/s jsou navrženy pouze pro dočasné použití při obtížných radiových podmínkách nebo zahlcení sítě.

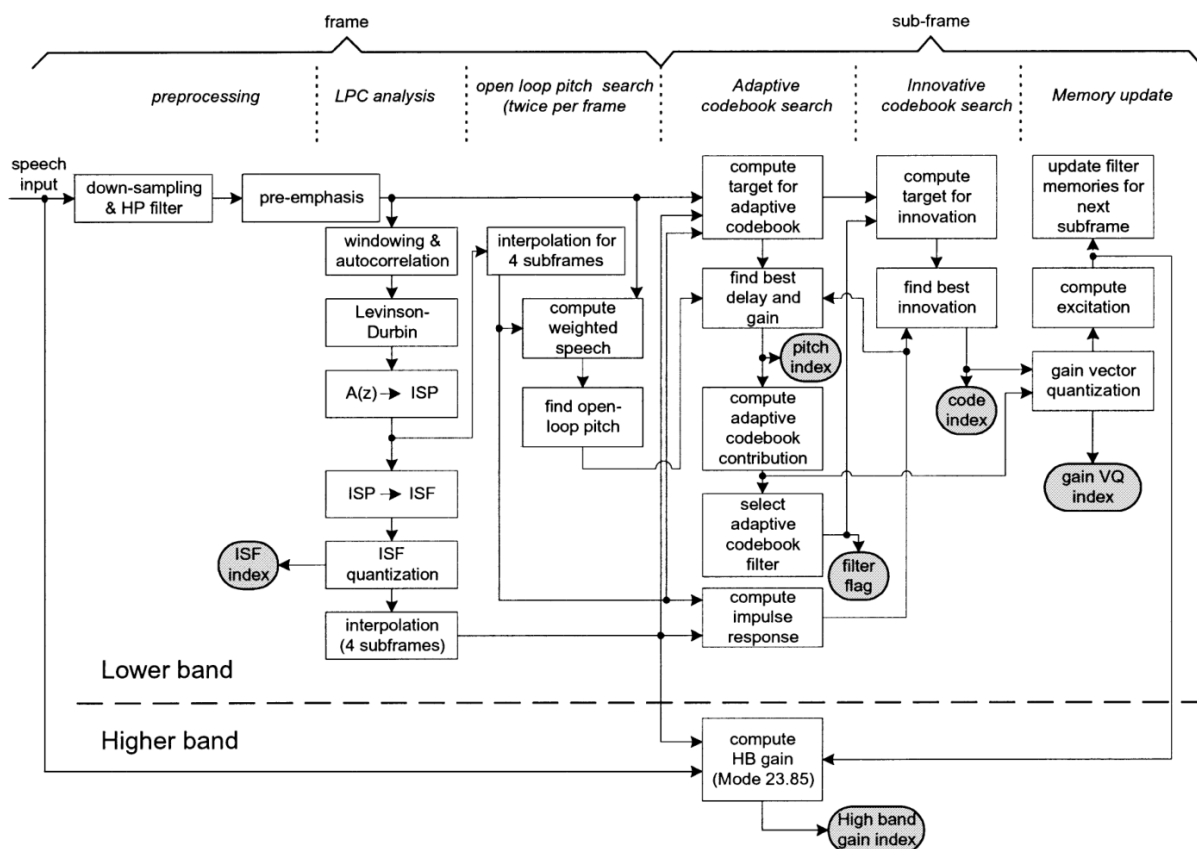
Proces kódování a dekódování zobrazují bloková schémata na obrázku 2.1 a 2.2. Alokace bitů v jednotlivých módech kodeku AMR-WB je znázorněna v tabulce 2.2. Kodek pracuje se vzorkovací frekvencí 16 kHz a kóduje signál po 20 ms blocích. Kódovaný signál je rozdělen na dvě frekvenční pásma, 50-6400 Hz a 6400 až 7000 Hz pro snížení složitosti a pro zaměření bitové alokace pro nejdůležitější frekvence.

Vstupní signál je podvzorkován na 12,8 kHz a předpřipraven použitím hornopropustního a preemfázového filtru ve tvaru $P(z) = 1 - \mu z^{-1}$, kde $\mu = 0,68$. Algoritmus ACELP je aplikován na předzpracovaný a podvzorkovaný signál, kde je jedenkrát za 20 ms provedena analýza lineární predikce. Z ní získaný soubor parametrů je konvertován do párů imitanciálního spektra a vektorově kvantován za použití rozdělovacího víceúrovňového 46 bitového kvantování. Hlasový rámec je rozdělen do 4 podrámců po 5 ms. Parametry adaptivní a pevné kódové knihy jsou přenášeny každým podrámcem. Interval základního tónu je kódován 9 bity v lichých podrámcích a 6 bity relativně kódováno v sudých podrámcích. K určení filtru dolní propusti u posledního buzení je použit jeden bit pro každý podrámec. Základní tón a algebraická kódová kniha jsou kvantizovány společně 7 bity na subrámec. Velikost algebraické kódové knihy je různá podle použitého módu kodeku, viz tabulka 2.1.

Bitová rychlost [b/s]	12,65	14,25	15,85	18,25	19,85	23,05
Velikost algebr. kódové knihy [b]	36	44	52	64	72	88

Tab. 2.1: Velikost kódové knihy v závislosti na módu kodeku[4]

Mód 23,85 kb/s se odlišuje módu 23,05 kb/s tím, že na každý podrámec používá 4 bity na kódování vysokopásmového zisku. Mód 8,85 kb/s používá 46 bitů pro parametry lineární predikce (LP), 8 a 5 bitů pro interval základního tónu pro lichý, respektive sudý podrámec, 20 bitů pro algebraickou kódovou knihu a 6 bitů pro zisk v každém podrámcu. Mód 6,6 kb/s používá 36 bitů pro parametry lineární predikce (LP), 8 bitů pro interval základního tónu v prvním podrámcu a 5 v dalších



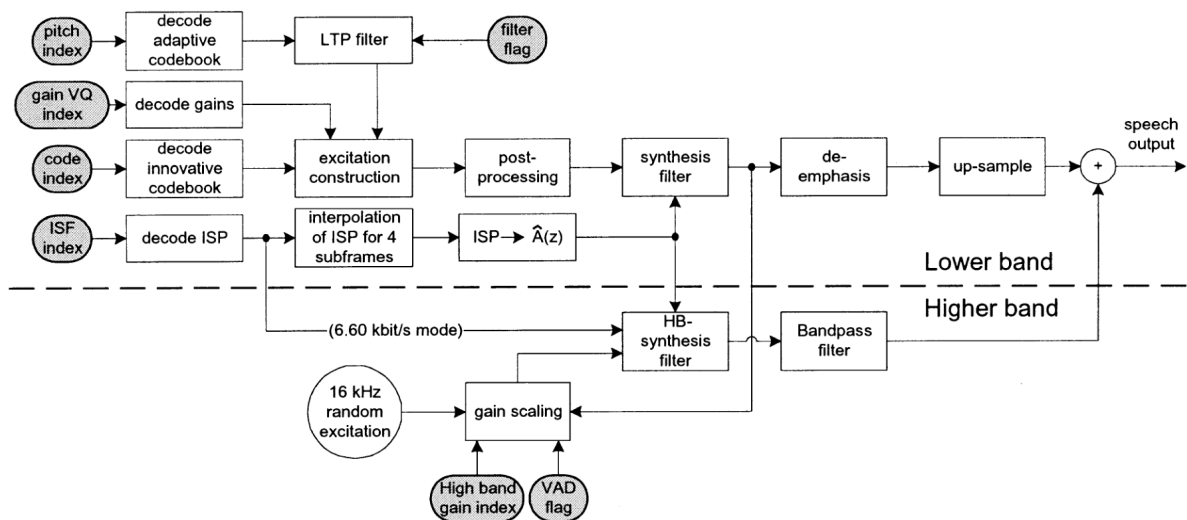
Obr. 2.1: Blokový diagram kodéru ACELP AMR-WB[4]

podrámcích. Dále používá 12 bitů pro algebraickou kódovou knihu a 6 bitů pro oba zisky v každém podrámcí. Vyšší frekvenční pásmo (6400-7000 Hz) je rekonstruováno v dekodéru z parametrů nízkého pásma a náhodného buzení. Nejsou přenášeny žádné informace o vyšším frekvenčním pásmu, kromě režimu 23,85 kb/s, kdy je zisk vyššího pásma přenášen pomocí 4 bitů na podrámeček. U ostatních módů je zisk vyššího pásma nastaven relativně k nižšímu pásmu podle hlasových informací. Spektrum vysokého pásma je rekonstruováno prostřednictvím širokopásmového filtru typu dolní propust generovaného z filtru typu dolní propust pro nízké pásmo.[4]

2.5 Percepční váhování

U kodérů typu analýza po syntéze (Analysis by synthesis), je optimální základní tón a změna parametrů vyhledávána minimalizováním střední kvadratické chyby mezi vstupní řečí a řečí synteticky vytvořenou v percepčně váhované oblasti.

Percepční váhování je obvykle provedeno použitím váhovacího filtru



Obr. 2.2: Blokový diagram dekodéru ACELP AMR-WB[4]

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

kde $A'(z)$ je lineární predikční filtr a γ_1 a γ_2 jsou faktory ovlivňující míru percepčního váhování v mezích $0 < \gamma_2 < \gamma_1 \leq 1$. To může ukazovat na to, že kódovací šum (u kterého je předpokládáno bílé spektrum) je váhován přenosovou funkcí $1/W'(z)$, která je inverzní k přenosové funkci percepčního váhovacího filtru. Přenosová funkce $1/W'(z)$ vykazuje některé formantové struktury vstupního řečového signálu. To znamená, že maskovací vlastnost lidského sluchu je využívána pro tvarování kvantizační chyby tak, že má větší energii ve formantových oblastech, kde je maskován silnější signálovou energií.[4]

Klasický percepční váhovací filtr pracuje skvěle se signály telefonního pásma, kde přenosová funkce $W(z)$ nevykazuje silné spektrální vychýlení, avšak není vhodný pro efektivní váhování širokopásmových signálů díky vlastním omezením $W(z)$ v modelování formantové struktury a současně požadovanému spektrálnímu vychýlení. V širokopásmových signálech je spektrální vychýlení výraznější kvůli širokému dynamickému rozsahu mezi nízkými a vysokými kmitočty. V minulosti bylo navrženo přidat vychylovací filtr do $W(z)$ pro dosažení separátního ovládní vychýlení a formantového váhování širokopásmového signálu.[12]

Novým řešením tohoto problému je použití preemfázového filtru ve tvaru $P(z) = 1 - \mu z^{-1}$ na vstupu, vypočítání LP filtru $A(z)$ po preemfázové úpravě řeči $s(n)$ a použití upraveného filtru $W(z)$. Percepční váhovací filtr může mít například tento

Parametr	Mód kodeku kb/s								
	6,60	8,85	12,65	14,25	15,85	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
Interval 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
Vys. pás. energie	0	0	0	0	0	0	0	0	16
Celkem na rámeček	132	177	253	285	317	365	397	461	477

Tab. 2.2: Bitová alokace v módech AMR-WB[4]

tvar, ve kterém je podstatně odděleno formantové váhování a spektrální vychýlení:

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

Stojí za povšimnutí, že v případě kdy je funkce $A(z)$ vypočítána podle signálu $s(n)$ před preemfázovou úpravou, je vychýlení filru $1/A(z/\gamma_1)$ méně výrazné než v případě, kdy je $A(z)$ vypočítáno z původního hlasového signálu.

Protože je deem fáze provedena v dekodéru filtrem s přenosovou funkcí $1/P(z)$, je kódovací chyba tvarována filtrem $W^{-1}(z)P^{-1}(z)$. V typickém případě, kdy se γ_2 ve vztahu 2.2 rovná μ , je kvantizační chyba tvarována filtrem s přenosovou funkcí $1/A(z/\gamma_1)$, kde je $A(z)$ vypočítáno ze signálu po úpravě preemfáze. Subjektivním poslechem bylo zjištěno, že tato konstrukce použitá k tvarování chyby sestávající se z preemfázové úpravy signálu a upraveného váhového filtrování, je pro kódování širokopásmových signálů velmi účinná. Navíc preemfázový filtr snižuje dynamický rozsah vstupního řečového signálu, čímž se signál stává vhodnějším pro celočíselné implementace.[4]

2.6 Analýza základního tónu

Hledání základního tónu se u kodeku AMR-WB skládá ze 3 kroků. V prvním kroku je interval otevřené smyčky základního tónu T_O odhadován každých 10 ms podle decimálně váhovaného signálu filtrovaného dolní propustí.

Ve druhém kroku je hledání uzavřené smyčky prováděno pro celočíselné intervaly kolem odhadnuté hodnoty T_O v rozsahu ± 7 vzorků, což znatelně zjednodušuje hledání. V okamžiku, kdy je nalezen optimální celočíselný interval, je ve třetím kroku

prohledáván zlomek kolem celočíselné hodnoty. Interval základního tónu je omezen v intervalu [34, 231].

Pro první a třetí podrámec při bitové rychlosti 12,65 kb/s a vyšší, je zlomkový interval základního tónu využíván s rozlišením 1/4 a 1/2 v intervalu [34, 127(3/4)] a [128, 159(1/2)]. V rozsahu [160, 231] je použito celočíselné rozlišení. Pro druhý a čtvrtý podrámec je při hledání v rozmezí [-8, 7(3/4)] vždy použito rozlišení 1/4 kolem celočíselného intervalu nalezeného pro předcházející podrámec.

Vyhledávání uzavřené smyčky je prováděno minimalizováním střední kvadratické chyby mezi původním signálem a syntetickou řečí. Toho je dosaženo maximalizací:

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

kde N je velikost podrámece, $x(n)$ je výsledný signál a $y_k(n)$ je poslední buzení v intervalu k filtrované váhovacím filtrem $W(z)/\hat{A}(z)$. Signál $y_k(n)$ je vypočítán konvolucí posledních buzení v intervalu k , $u_k(n)$ s impulzní odezvou váhovacího syntézního filtru $h(n)$. Pro intervaly základního tónu menší než je velikost podrámece, je poslední buzení rozšířeno zbytkovým LP signálem. Je dobré připomenout, že rozsah vyhledávání je omezen kolem odhadu základního tónu otevřené smyčky (T_O), jak je uvedeno výše.[4]

Jakmile je nalezen optimální celočíselný interval základního tónu, jsou zlomky v intervalu $[-3/4, 3/4]$ kolem něj prohledávány s krokem 1/4. Zlomkové vyhledávání základního tónu je prováděno interpolováním normalizované korelace ve vztahu 2.3 a hledáním maxim. Když je určen zlomkový interval základního tónu, tak je interpolován poslední budící signál $u(n)$ při daném intervalu a zlomku a z něj je vypočítáno poslední interpolované buzení $v'(n)$. Interpolace je provedena použitím dvou FIR filtrů složených z funkce sinc ohraničené Hammingovým oknem, jeden pro interpolování výrazu 2.3 s filtrem 35. řádu a druhý pro interpolaci posledního buzení s filtrem 127. řádu. Tyto filtry mají mezní kmitočet (pokles o 3 dB) 6000 Hz, což znamená že interpolační filtry mají frekvenční odezvu dolní propusti.[4]

Pro rozšíření předpovědi periody základního tónu u širokopásmových signálů je třeba použít frekvenčně závislý prediktor, protože periodičita ne vždy pokryje celé spektrum.

Pokud je prediktor základního tónu reprezentován filtrem $1/(1 - bz^{-T})$, což je platný předpoklad pro periody základního tónu splňující $T \geq N$, pak filtr základního tónu projevuje harmonickou strukturu přes celý frekvenční rozsah s harmonickou frekvencí vztaženou k $1/T$. U širokopásmových signálů není tato struktura efektivní, protože harmonická struktura obvykle nepokrývá celé spektrum až do 7 kHz. Harmonická struktura existuje pouze do určité frekvence v závislosti na hlasovém

segmentu. K tomu, aby bylo dosaženo efektivní reprezentace podílů základních tónů v hlasových segmentech širokopásmové řeči, je třeba, aby filtr predikce základního tónu měl možnost měnit hodnotu periodicity na vysoké frekvenci.

Nová metoda, která účinně modeluje harmonické struktury spektra širokopásmových hlasových signálů, využívá různé formy dolních propustí aplikovaných na poslední buzení, přičemž je vybrána dolní propust s největším predikčním ziskem.

Když je použito zlomkové rozlišení základního tónu, je možné filtr s dolní propustí začlenit do interpolačního filtru k dosažení lepšího rozpoznání základního tónu. V tom případě je třetí krok hledání základního tónu (při kterém jsou testovány zlomky kolem celočíselné hodnoty) opakován pro několik interpolačních filtrů s různou charakteristikou dolní propusti, frakčním a filtračním indexem.

Účelněji se však jeví využití pouze jednoho interpolačního filtru (obvykle filtru s širší frekvenční odezvou) ke zjištění optimálního zlomkového intervalu základního tónu. Optimální tvar dolní propusti lze nakonec stanovit výběrem z testovaných filtrů, které jsou aplikovány na vybraný vektor V_T kódových knih základních tónů. Vybraný filtr minimalizuje chybu předpovědi základního tónu.

V kodeku AMR-WB je pro každý podrámeček alokován jeden bit pro popis dolní propusti filtru určeného pro tvarování buzení adaptivní kódové knihy. Když je filtr typu dolní propust zakázán, buzení adaptivní kódové knihy odpovídá $v(n) = v'(n)$. Jinak je pro filtrování adaptivní kódové knihy použit FIR filtr druhého řádu, který popisuje vztah $B_L(z) = b_L(0)z + b_L(1) + b_L(2)z^{-1}$. Výsledkem tohoto filtrování je:

$$v(n) = \sum_{i=-1}^1 b_L(i+1)v'(n+i) \quad (2.4)$$

Koeficienty filtru jsou zvoleny takto: $b_L(0) = b_L(2) = 0, 18$ a $b_L(1) = 0, 64$. Tento filtr druhého řádu je kompromisem mezi složitostí a požadovaným vysokofrekvenčním útlumem. Zisk adaptivní kódové knihy je pak nalezen dle vztahu:

$$g_p = \frac{\sum_{n=0}^{N-1} x(n)y(n)}{\sum_{n=0}^{N-1} y(n)y(n)} \quad 0 \leq g_p \leq 1, 2 \quad (2.5)$$

kde $y(n) = v(n)*h(n)$ je filtrovaný vektor adaptivní kódové knihy. Aby bylo dosaženo stability kodéru i v případě chyb v přenosovém kanále, je zisk adaptivní kódové knihy omezen shora hodnotou 0,95.

2.7 Struktura a vyhledávání v algebraické fixní kódové knize

Pro kódování širokopásmového řečového signálu jsou vyžadovány velké kódové knihy až do velikosti 88 bitů pro zajištění vysoké subjektivní kvality hlasu, které je třeba

efektivně prohledávat.

Velikost podrámcce je 64 vzorků, což odpovídá času 5 ms resp 12,8 kHz vzorkovací frekvenci. Struktura kódové knihy je založena na způsobu ISPP (interleaved single-pulse permutation)

Stopa	Pozice impulsu v podrámcí															
T_0	0	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60
T_1	1	5	9	13	17	21	25	29	33	37	41	45	49	53	57	61
T_2	2	6	10	14	18	22	26	30	34	38	42	46	50	54	58	62
T_3	3	7	11	15	19	23	27	31	35	39	43	47	51	55	59	63

Tab. 2.3: Pozice impulsů v algebraické kódové knize pro jednotlivé stopy[4]

Jak je vidět v tabulce 2.3, je 64 pozic kódového vektoru rozděleno do 4 stop s prokládanými pozicemi po 16 v každé stopě. Rozdílné kódové knihy jsou vytvořené pro různé rychlosti rozdělením určitého počtu impulsů do stop, od 1 do 6 pro každou stopu. Index kódové knihy nebo kódového slova představuje pozice impulsu a znaku v každé stopě. Tím pádem není potřeba žádná paměť kódové knihy, protože budící vektor v dekodéru může být vytvořen z informací v indexu bez vyhledávacích tabulek.

Je-li jeden označený impuls umístěný v každé stopě, je jeho pozice kódována 4 bity a jeho znak 1 bitem, a výsledkem je 20 bitová kódová kniha. Jsou-li v každé stopě umístěny 2 impulsy, je jejich pozice kódována 8 bity, ale odpovídající znaky jsou kódovány pouze 1 bitem díky využití impulsního uspořádání. Důsledkem toho je, že je potřeba $4 * (4 + 4 + 1) = 36$ bitů pro určení pozic impulsů a znaků pro toto partikulární algebraické uspořádání kódové knihy. Další uspořádání kódové knihy mohou být tvořena 3, 4, 5 nebo 6 impulsy v každé stopě.

Důležitou vlastností použité kódové knihy je její dynamičnost skládající se z algebraické kódové knihy následované adaptivním pre-filtrem $F(z)$, který rozšiřuje zvláštní spektrální součásti pro dosažení lepší subjektivní kvality. V kodeku AMR-WB se pre-filtr $F(z)$ skládá z kaskády dvou filtrů. Pro filtr pro rozšíření periodicity platí vztah:

$$F_1(z) = \frac{1}{1 - 0,85z^{-T}} \quad (2.6)$$

a pro filtr odchylky platí vztah:

$$F_2(z) = 1 - \beta_T z^{-1} \quad (2.7)$$

Koeficient β je vztažen k vyjádření předchozího poddrámce a je ohraničen intervalem $[0, 0, 5]$ a T je celočíselná část intervalu základního tónu. Prohledávání kódové knihy je prováděno v algebraické oblasti při použití filtru $F(z)$ a filtru s váženou syntézou před prohledáním kódové knihy. To je dáno konvolucí impulzní odezvy filtru s váženou syntézou a impulzní odezvou filtru $F(z)$. Výsledný signál vzniklý konvolucí je dále označen jako $h(n)$.

Algebraická kódová kniha je prohledávána minimalizováním střední kvadratické chyby mezi váženou vstupní řečí a váženou syntetickou řečí. Cílový signál použitý při vyhledávání uzavřené smyčky základního tónu je aktualizován odečtem příspěvku kódové knihy.[4]

Kódový vektor s označením z_k s indexem k (o délce N) odpovídá signálu $z_k(n) = c_k(n) * h(n) = \sum_{i=0}^n c_k(i)h(n-i)$. Prohledávání algebraické kódové knihy je provedeno maximalizováním vyhledávacího kritéria:

$$Q_k = \frac{(x_2^t z_k)^2}{z_k^t z_k} \quad (2.8)$$

Matice H je definována jako dolní trojúhelníková Toeplitzova konvoluční matice velikosti $N \times N$ s diagonálou $h(0)$ a nižšími diagonálami $h(1), \dots, h(N-1)$, kde $z_k = Hc_k$. Vztah $d = H^t x_2$ označuje korelaci mezi cílovým signálem $x_2(n)$ a impulzní odezvou $h(n)$, zatímco $\Phi = H'H$ označuje matici korelací $h(n)$. Korelace d je rovněž uvedena jako zpětně filtrovaný cílový vektor a odpovídající signál je označen jako $d(n)$. Proto lze zapsat vyhledávací kritéria vztahem:

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

Vektor d a matice Φ jsou obvykle vypočítány před prohledáním kódové knihy. Algebraická struktura kódové knihy umožňuje velmi rychlé vyhledávání, protože rozdílový vektor c_k obsahuje pouze několik nenulových pulzů. Korelace mezi čitatelem ve vztahu 2.9 odpovídá vztahu:

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

kde p_i je pozice i -tého pulzu, a_i je jeho znak a N_p je počet pulzů ve vektoru c_k . Energie ve jmenovateli vztahu 2.9 je určena vzorcem:

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

Pro zjednodušení vyhledávání jsou amplitudy impulzů předurčeny na základě referenčního signálu $b(n)$. U tohoto signálu je znak impulzu na pozici i nastaven stejně jako znak referenčního signálu na stejné pozici. Pak je referenční signál $b(n)$ definován takto:

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

kde $E_D = d^T d$ je energie signálu $d(n)$ a energie zbytkového signálu $r(n)$ vyplývajícího z predikce základního tónu je označena $E_R = r^T r$. Koeficient α ovládá závislost referenčního signálu na $d(n)$ a s rostoucím počtem budících impulsů se snižuje. Při bitové rychlosti 12,65 kb/s je koeficient α roven 1, ale při rychlosti 23,05 kb/s je jeho hodnota snížena na 0,5.[4]

Pro zjednodušení hledání jsou signál $d(n)$ a matice Φ upraveny tak, aby začlenily přednastavené znaky. Značí-li $s_b(n)$ znak signálu $b(n)$, pak pro modifikovaný signál $d'(n)$ platí:

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

a pro upravenou autokorelační matici Φ'

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

Čitatel vyhledávacího kritéria Q_k má tvar

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

a jmenovatel

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

Cílem vyhledávání je určení kódových vektorů s nejlepší sadou impulsních pozic N_p za předpokladu, že byly vybrány postupem popsáním výše. Základním kritériem výběru je maximalizace Q_k .

Pro zjednodušení je použita rychlá procedura známá jako „prohledávání do hloubky“, která určí najednou pozice N_m impulsů. Dostupné impulsy N_p jsou rozděleny do M neprázdných podmnožin o N_m impulzech tak, že $N_1 + N_2 \dots + N_m = N_p$. Konkrétní volba pozic pro první $J = N_1 + N_2 \dots + N_{m-1}$ impulzy je nazývána m-tá

úroveň cesty, nebo délka cesty J . Základním kritériem pro cestu impulzních pozic J je poměr $Q_k(J)$, kdy jsou brány v úvahu pouze relevantní impulzy.

Hledání začíná v podmnožině č. 1 a pokračuje v následující množině podle stromové struktury tak, že podmnožina m je hledána na m -té úrovni stromu. Účelem hledání na první úrovni je vzít v úvahu impulzy v podmnožině 1 a jejich platné pozice použít pro určení jedné nebo více možných cest délky N_1 . Tyto možné cesty jsou uzly stromu na první úrovni. Na každém konečném uzlu úrovně $m - 1$ je cesta rozšířena do délky $N_1 + N_2 \dots + N_n$ na úrovni m díky novým impulzům N_m a jejich pozicím. Jedna nebo více možných prodloužených cest jsou určeny k tomu, aby vytvořily uzly m -té úrovně. Nejlepší kódový vektor odpovídá cestě délky N_p , která maximalizuje kritérium $Q_k(N_p)$ s ohledem na všechny uzly M -té úrovně.[4]

V tomto případě je použit zvláštní druh „prohledávání do hloubky“ který hledá 2 impulzy najednou ($N_m = 2$) a tyto dva impulzy náleží dvěma po sobě jdoucím cestám. Pro snížení množství použité paměti je dále vyhledávání prováděno tak, že je vypočtena a uložena pouze část korelační matice Φ . Celá matice Φ vyžaduje paměť $N \times N$ slov ($64 \times 64 = 4096$). Předuložená část zastupuje korelace impulzní odezvy odpovídající pozicím potenciálního impulsu v po sobě jdoucích cestách, stejně jako korelace odpovídající $\phi(j, j)$ pro $j = 0, \dots, N - 1$, což jsou prvky na diagonále matice Φ .

Pro zjednodušení testování možných kombinací dvou impulsů je testováno omezené množství možných pozic prvního impulsu. Je-li testováno velké množství impulsů, pak jsou některé impulzy na vyšších úrovních vyhledávacího stromu fixovány. Pro vytvoření inteligentního odhadu potenciálních pozic prvních pulzů je referenční signál $b(n)$ definovaný v 2.12 použit pro odhad pravděpodobnosti impulzní pozice založený na charakteru hlasových signálů. Jeho hodnota označuje pravděpodobnost obsazení pozice n v nejlepším hledaném kódovém vektoru.[4]

Postup vyhledávání je u všech módů podobný. Vyhledávány jsou dva impulzy najednou, které odpovídají po sobě jdoucím cestám. To znamená, že hledané impulzy jsou na cestách T_0 a T_1 , T_1 a T_2 , T_2 a T_3 , nebo T_3 a T_0 .

Dříve než je vyhledána pozice, je znak impulsu na potenciální pozici n nastaven na místo znaku signálu $b(n)$ na odpovídající pozici. Dále je vypočítán modifikovaný signál $d'(n)$ a korelace $\phi'(n, n)$ podle vztahu uvedených výše za použití předem určených znaků.

U prvních dvou impulsů (první úroveň stromu) je čísel vyhledávacího kritéria Q_k popsán vztahem:

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

a jmenovatel vztahem:

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

kde byly korelace $\phi'(p_i, p_j)$ upraveny tak, aby zahrnovaly předem vybrané znaky na pozicích p_i a p_j .

Pro následující úrovně je čitatel a jmenovatel upraven přidáním příspěvků dvou nových impulsů. Za předpokladu, že jsou nalezeny dva nové impulzy z dvou po sobě jdoucích cest na určité úrovni stromu na pozicích p_k a p_{k+1} , pak aktualizované hodnoty C a E jsou dány vztahem:

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

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

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

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

Výsledkem je

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

Stojí za povšimnutí, že $v_h(n)$ je příspěvek již určených impulsních odezev na již určených pozicích. Na každé úrovni stromu jsou hodnoty $R_{hv}(m)$ vypočítány pro všechny možné pozice pro obě testované cesty. Z rovnice 2.20 je patrné, že uložení samotné korelace $\phi'(m_k, m_{k+1})$, odpovídající pozici impulsů ve dvou po sobě jdoucích cestách je možné uložit do paměti o velikosti ($4 \times 16 \times 16$ slov). Spolu s korelací $\phi'(m_k, m_k)$ odpovídající úhlopříčce matice Φ (velikost 64 slov), je třeba paměť o velikosti 1088 slov, místo $64 \times 64 = 4096$ slov.[4]

Jako příklad bude prohledávání ukázáno na 36-bitové kódové knize používané v režimu 12,65 kb/s. V použité kódové knize jsou v každé stopě umístěny dva impulzy, což je dohromady 8 impulsů na jeden podrámeček o délce 64. Dva impulzy odpovídající po sobě jdoucím drahám jsou vyhledávány současně. To znamená, že jsou vyhledávány impulzy v drahách T_0 a T_1 , T_1 a T_2 , T_2 a T_3 a T_3 a T_0 . V tomto případě má strom čtyři úrovně. V první úrovni je impuls P_0 přiřazen stopě T_0 a impuls P_1 je přiřazen stopě T_1 . Na této úrovni není prováděno vyhledávání a pozice obou impulsů jsou nastaveny jako maximum signálu $b(n)$ v každé stopě. Na druhé úrovni je impuls P_2 přiřazen stopě T_2 a impuls P_3 stopě T_3 . Čtyři pozice pro impuls P_2 jsou testovány proti všem 16 pozicím impulsu P_3 . Tyto čtyři testované pozice pro impuls P_2 jsou určeny na základě maxima signálu $b(n)$ ve stopě. Na třetí úrovni je impuls P_4 přiřazen stopě T_1 a impuls P_5 stopě T_2 . Osm pozic pro impuls P_4 je testováno proti všem 16 pozicím impulsu P_5 . Podobně jako na předchozí úrovni je osm testovaných pozic impulsu P_4 určeno podle maxima signálu $b(n)$ ve stopě. Ve

čtvrtě úrovni je impuls P_6 přiřazen stopě T_3 a impuls P_7 stopě T_0 . Osm pozic pro impuls P_6 je testováno proti všem 16 pozicím impulsu P_7 . Celkový počet testovaných kombinací je pak $4 \times 16 + 8 \times 16 + 8 \times 16 = 320$. Celý tento proces je čtyřikrát zopakován přiřazením impulsů do jiných stop. Při druhém opakování jsou impulsy P_0 a P_7 přiřazeny stopám $T_1, T_2, T_3, T_0, T_2, T_3, T_0$ a T_1 . Tím je prozkoumáno celkem $4 \times 320 = 1280$ kombinací pozic.[4]

Při použití módu s přenosovou rychlostí 15,85 kb/s jsou v každé ze čtyř stop umístěny 3 impulsy. Tři impulsy v každé stopě lze kódovat 13 bity (3 pozice a 1 znak a další dva znaky lze odvodit z řazení impulsů), které dávají dohromady 52-bitovou kódovou knihu. V stromové struktuře je 6 úrovní a v každé z nich jsou vyhledávány dva impulsy. V prvních dvou úrovních jsou impulsy nastaveny na maximum signálu $b(n)$. V následujících čtyřech úrovních je následující počet testovaných kombinací $4 \times 16, 6 \times 16, 8 \times 16$ a 8×16 . Při použití čtyř opakování je celkový počet $4 \times 26 \times 16 = 1664$ kombinací.

Kvantizace zisku

Zisk fixní a adaptivní kódové knihy je vektor kvantizovaný 6-bitovou kódovou knihou při použití módů 6,60 a 8,85 kb/s a 7-bitovou knihou při použití všech ostatních módů. Kvantování zisku fixní kódové knihy využívá metodu zvanou „moving-average prediction“ s konstantními koeficienty. Predikce a vyhledávání kódové knihy je podobné jako u kodeků GSM EFR a G.729.

2.8 Algoritmický popis řečového dekodéru

Funkce dekodéru spočívá v dekodování přenášených parametrů (VAD znak, LP parametry, index a zisk adaptivní kódové knihy, vektory a zisky fixní kódové knihy a vysokopásmový zisk) a provedení syntézy k získání rekonstruované řeči. Rekonstruovaný řečový signál je zpracován a nadzorkován. Vysokofrekvenční signál je generován na frekvencích 6,4 - 7 kHz. Postup zpracování signálu je zobrazen na obrázku 2.2.

Dekódování, zpracování buzení a syntéza řeči

Přijaté indexy kvantizace ISP jsou použity k rekonstrukci vektoru ISP. Čtyři interpolované vektory ISP (odpovídající čtyřem poddrámcům) jsou vypočítány a použity pro syntézu řeči v každém poddrámcu. Poté je v každém poddrámcu použit index přijaté adaptivní kódové knihy k nalezení celočíselné a zlomkové části intervalu základního tónu. Buzení adaptivní kódové knihy $v'(n)$ je nalezeno pomocí posledního buzení $v(n)$ v intervalu základního tónu.[4]

Index přijaté algebraické kódové knihy je použit pro získání pozic a znaků budících impulzů a k získání buzení algebraické kódové knihy $c(n)$. Je-li celočíselná část intervalu základního tónu menší než velikost podrámcce(64), je použita procedura která filtruje $c(n)$ filtrem $F(z)$ popsáním výše. Poté je zisk adaptivní a fixní kódové knihy dekódován a celkové buzení popisuje rovnice:

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

Pro zlepšení subjektivního vjemu rekonstruovaného signálu je před syntézou provedeno zpracování budících složek. Procedury použité při zpracování budícího signálu jsou uvedeny níže.

Protirozptylové zpracování u režimů 6,60 a 8,85 kb/s

Zpracování je použito na buzení fixní kódové knihy $c(n)$ pro snížení množství vjemových vad vycházejících z fixní algebraické kódové knihy s pouhými pár nenulovými vzorky v podrámcí. Protirozptylové zpracování zahrnuje kruhovou konvoluci buzení fixní kódové knihy s impulsní odezvou. Podle rozsahu nutných úprav je vybrána jedna ze tří uložených impulzních odezev, které jsou vybrány adaptivně podle zisků adaptivní a fixní kódové knihy.

Oddělení šumu

Pro zlepšení buzení na šumu je použita technika nelineárního vyhlazování zisku, která je aplikována na zisk fixní kódové knihy \hat{g}_c . Zisk fixní kódové knihy je vyhlazen podle stability a znělosti pro snížení fluktuace energie buzení při stálých signálech, což vede k subjektivnímu zvýšení kvality řeči při stálém šumu na pozadí.

Znělost podrámcce je odhadována takto:

$$\lambda = \frac{E_c}{(E_v + E_c)} \quad (2.24)$$

kde E_v a E_c jsou energie buzení adaptivní, resp. fixní kódové knihy. Činitel znělosti nabývá podle definice hodnot $[0, 1]$. Hodnot blízkých nule dosahuje u čistě znělých podrámců a hodnot kolem jedné u neznělých podrámců. Ustálenost rámcce je odhadována podle filtrů LP pomocí rozdílu parametrů ISP mezi současným a předcházejícím rámcem. Rozdíl parametrů ISP mezi současným a předchozím rámcem určuje vztah:

$$D_s = \sum_{i=1}^{p-1} (isp_i^{(n)} - isp_i^{(n-1)})^2 \quad (2.25)$$

kde $p = 16$ je řád LP filtru. Vzdálenost parametrů ISP je přiřazena stabilizačnímu činitele θ v rozsahu od 0 do 1 při použití vztahu:

$$\theta = 1,25 - \frac{D_s}{400000}, \quad 0 \leq \theta \leq 1. \quad (2.26)$$

Nakonec je vypočítán činitel vyhlazení zisku σ :

$$\sigma = \lambda\theta \quad (2.27)$$

U neznělých a stacionárních signálů jako je stálý šum na pozadí, nabývá σ hodnoty blízké 1 a u znělých a nestacionárních signálů se hodnota blíží 0.

Nelineární vyhlazování používá zisk předcházejícího podrámcce jako referenční hodnotu \hat{g}_c^r pro výpočet modifikovaného zisku \hat{g}_c^m . Je-li $\hat{g}_c \geq \hat{g}_c^r$, pak je modifikovaný zisk \hat{g}_c^m nastaven na hodnotu 1,5 dB pod \hat{g}_c a zároveň je zdola omezen referenční hodnotou. V opačném případě je hodnota \hat{g}_c^m nastavena 1,5 dB nad \hat{g}_c , kde je shora omezena referenční hodnotou. Nakonec je upravený zisk fixní kódové knihy vyjádřen vztahem:

$$\hat{g}_c = \sigma\hat{g}_c^m + (1 - \sigma)\hat{g}_c \quad (2.28)$$

Zvýraznění základního tónu

Zvýraznění základního tónu zvyšuje subjektivní kvalitu řeči pomocí filtrování buzení fixní kódové knihy $c(n)$ filtrem, jehož frekvenční odezva zvýrazňuje vysoké frekvence a tlumí nízké. Koeficienty tohoto filtru jsou určeny vztahem:

$$F_E(z) = -f_E z + 1 - f_E z^{-1} \quad (2.29)$$

Koeficienty jsou vztaženy k periodicitě signálu $f_E = 0,25\lambda$ s činitelem znělosti ze vztahu 2.24. Tyto koeficienty jsou obnovovány jednou za rámeček. Použitý filtr vykazuje největší zvýraznění u neznělých řečových signálů.

Při použití filtru získá buzení fixní kódové knihy tento tvar:

$$C_E(n) = c(n) - f_E(c(n+1) + c(n-1)) \quad (2.30)$$

a v důsledku toho je celé buzení popsáno vztahem:

$$\hat{u}(n) = \hat{g}_p v(n) + \hat{g}_c c_E(n). \quad (2.31)$$

Syntéza řeči

Rekonstruovaná řeč v podrámcce velikosti 64 je získána filtrováním zpracovaného celkového buzení přes LP syntézni filtr. To je popsáno vztahem:

$$\hat{s}(n) = \hat{u}(n) - \sum_{i=1}^{16} \hat{a}_i \hat{s}(n-i) \quad n = 0, \dots, 63 \quad (2.32)$$

kde \hat{a}_i a je i -tý koeficient interpolovaného filtru LP. Syntetizovaná řeč $\hat{s}(n)$ je poté adaptivně zpracována. Pro zachování synchronizace s kódérem je paměť adaptivní kódové knihy aktualizována buzením $u(n)$ bez zpracování.[4]

Nadvzorkování a vytváření vysokofrekvenčního signálu

Syntetizovaný signál je filtrován horní propustí, která zabrání proniknutí nechtěných nízkofrekvenčních složek. Poté je signál utlumen filtrem

$$\frac{1}{P(z)} = \frac{1}{(1 - 0,68z^{-1})} \quad (2.33)$$

Poté je signál nadvzorkován za účelem získání syntézy dolního pásma $\hat{s}_L(n)$ při vzorkovacím kmitočtu 16 kHz.

Vysoké frekvence jsou vytvářeny pro zaplnění pásma od 6,4 do 7 kHz. Vysokofrekvenční složka je vytvářena tak, že horní část spektra je vyplněna bílým šumem, který je vhodně škálován v budící oblasti. Tato složka je poté převedena do řečové oblasti pomocí filtru odvozeného od syntézního LP filtru použitého pro syntézu podvzorkovaného signálu.

Vysokopásmové buzení $u_H(n)$ je ze signálu bílého šumu $w(n)$ získáno podle vztahu

$$u_H(n) = \frac{\hat{g}_H w(n)}{\sqrt{\sum_{k=0}^{63} w^2(k)}} \quad (2.34)$$

kde \hat{g}_H je vysokofrekvenční zisk. Při použití módu 23,85 kb/s je zisk \hat{g}_H dekódován z přijatého indexu zisku. V ostatních módech je zisk \hat{g}_H odhadován pro každý podrámeček podle informace o znělosti. Odchylka signálu syntézy je popsána vztahem:

$$\tau = \frac{\sum_{n=0}^{63} \hat{s}(n)\hat{s}(n-1)}{\sum_{n=0}^{63} \hat{s}^2(n)} \quad 0 \leq \tau \leq 1 - 2\lambda \quad (2.35)$$

kde λ je činitel znělosti nabývající hodnot [0, 1, 1, 0]. Pak je vysokofrekvenční zisk \hat{g}_H vypočítán jako

$$\hat{g}_H = w_S(1 - \tau) + 1,25(1 - w_S)(1 - \tau) \quad (2.36)$$

kde w_S má hodnotu 1 v případě, že je aktivní funkce VAD, a hodnotu 0, když je funkce VAD vypnuta. Činitel odchylky je zahrnut ve výpočtu zisku kvůli zohlednění vysokofrekvenční složky syntetizovaného signálu. Ve znělých segmentech, kde je přítomno méně energie ve vysokofrekvenční složce, se τ blíží hodnotě 1, což vede k nižšímu zisku vysokofrekvenční složky. To snižuje energii generovaného šumu v těchto segmentech.[4]

Vysokopásmový filtr LP syntézy $1/\hat{A}_H(z)$ je získán pro každý podrámec z nízkopásmového LP syntézního filtru použitím vztahu

$$\hat{A}_H(z) = \hat{A}\left(\frac{12,8z}{16}\right) \quad (2.37)$$

kde $1/\hat{A}(z)$ je interpolovaný filtr syntézy LP, který byl odhadnut pro vzorkovací frekvenci 12,8 kHz, ale je používán pro 16 kHz. To znamená, že pásmo 5,1-5,6 kHz se vzorkovací frekvencí 12,8 kHz je přiřazeno pásmu 6,4-7,0 kHz se vzorkovací frekvencí 16 kHz.

Vysokopásmový syntetizovaný signál $\hat{s}_H(n)$ je vypočítán filtrováním vysokopásmového buzení $u_H(n)$ filtrem $1/\hat{A}_H(z)$ a omezením pásma na 6,4-7 kHz FIR pásmovou propustí. Posledním krokem vedoucím k syntetické řeči je přidání vysokopásmového syntetizovaného signálu \hat{s}_H k nízkopásmovému signálu $\hat{s}_L(z)$. [4]

2.9 Složitost kodeku

Kodek AMR-WB je definován jako sada základních operátorů v aritmetice s pevnou desetinnou čárkou definovaný skupinou 3GPP/ETSI. Základní operátory kodeku jsou implementací DSP(Digital signal processing) instrukcí do jazyka C, což umožňuje hrubý odhad výpočetní složitosti algoritmu. Výpočetní složitost kodeku AMR-WB činí včetně funkcí VAD, DTX a CNG 38,9 WMOPS (Weighted Million Operations Per Second). Tato složitost je uváděna jako největší možná, ve většině případů je nižší. Porovnání složitosti s kodekem AMR-NB je uvedeno v tabulce 2.4

Součást kodeku	AMR-WB	AMR-NB
	Výpočetní náročnost [WMOPS]	
Řečový kodér	31,1	14,2
Řečový dekodér	7,8	2,6
Celkem	38,9	16,8
DATA RAM [tisíce slov, 16-bit]	6,5	5,3
DATA RAM [tisíce slov, 16-bit]	9,9	14,6
PROGRAM ROM [počet ETSI operací]	3889	4851

Tab. 2.4: Porovnání výpočetní náročnosti kodeku AMR-WB a AMR-NB[4]

2.10 Adaptivní provoz v GSM kanále

Kodek AMR-WB nabízí celou řadu přenosových rychlostí od 6,6 do 23,85 kb/s. To umožňuje maximalizovat kvalitu přenášené řeči díky přizpůsobení bitové rychlosti

pro zvýšení odolnosti proti chybám při přenosu. U neadaptivních 3G UTRAN kanálů s rychlým výkonovým řízením mohou operátoři zvolit vhodnou bitovou rychlost, která bude kompromisem mezi kvalitou řečového signálu a zatížením sítě. Adaptace spojení a zvolení vhodného módu vyžaduje měření kvality kanálu a je také nutné vzít v úvahu vytížení sítě. Zvolení vhodného módu je řešeno prostřednictvím vnitropásmové signalizace.

2.11 Detekce hlasové aktivity a přerušované vysílání

Kodek AMR-WB obsahuje detekci hlasové aktivity (VAD), která umožňuje využívat pro šum na pozadí režim s nižší bitovou rychlostí. To šetří energii mobilní stanice a také snižuje úroveň zarušení okolí. VAD vyhodnocuje aktivitu pro každý 20 ms rámec a vrací booleovskou hodnotu. Rozhodnutí VAD je založeno na rozdělení řečového signálu v pásmu $[0, 6,4 \text{ kHz}]$ na 12 podpásem, ve kterých je odděleně počítána úroveň signálu. Funkce tónové detekce je používána k označení přítomnosti signálního tónu, znělé řeči, nebo jiného silně periodického signálu. Funkce je založena na normalizovaných ziscích základního tónu otevřené smyčky, které jsou vypočítány analýzou základního tónu řečového kodéru a generuje znak indikující přítomnost tónu.

V každém frekvenčním pásmu je odhadována úroveň šumu na pozadí podle VAD rozhodnutí, stálosti signálu a znaku indikujícího přítomnost tónu. Střední VAD rozhodnutí je vypočítáno porovnáním SNR a adaptivní prahové hodnoty, která je založena na šumu a dlouhodobém odhadu úrovně řečového signálu. Konečné VAD rozhodnutí vznikne přidáním délky hangoveru ke střednímu VAD rozhodnutí.

V sítích GSM je VAD rozhodování využíváno pro funkci DTX, která umožňuje vypnout vysílač po většinu času během prodlev mezi řeči. Během těchto prodlev je na přijímací straně generován syntetický šum podobný šumu na vysílací straně. Tento syntetický šum je produkován na základě přenesených parametrů šumu na vysílací straně přenesených běžnou rychlostí.

Analýza komfortního šumu je prováděna na vstupním signálu se vzorkovací frekvencí 12,8 kHz. Algoritmus nejprve určí vážený průměr spektrálních parametrů a průměrnou energii logaritmického signálu. Průměr těchto parametrů z posledních osmi rámců poskytuje informaci o úrovni a spektru šumu na pozadí. Parametry komfortního šumu jsou kódovány do speciálního rámce nazvaného Silence Descriptor (SID), ve kterém jsou přenášeny k přijímací straně.[4]

Při vytváření komfortního šumu se vytváří pseudonáhodný šumový vektor škálovaný ziskem buzení, které je vypočítáno z energie logaritmického signálu. Tento šum

je filtrován syntézním filtrem vytvořeným z váženého průměru spektrálních parametrů. U nestacionárního šumu na pozadí je použit dithering pro úpravu spektrálních parametrů i zisku buzení. Syntetizovaný signál je nadvzorkován zpět na 16 kHz a jeho vyšší pásmo je generováno stejně jako v řečovém dekodéru.

2.12 Kvalita řeči

Kvalita řeči při použití vyšších módů kodeku AMR-WB (23,85, 23,05, 19,85, 18,25, 15,85 a 14,25 kb/s) je lepší nebo alespoň stejná, v porovnání s širokopásmovým kodekem definovaným ITU-T pod označení G.722 s bitovou rychlostí 64 kb/s. Kvalita řeči u módu 12,65 a 8,85 kb/s je minimálně stejně dobrá, jako při použití G.722 s bitovou rychlostí 56 resp. 48 kb/s.[4]

3 SIMULACE VYBRANÝCH BLOKŮ

3.1 Úvod

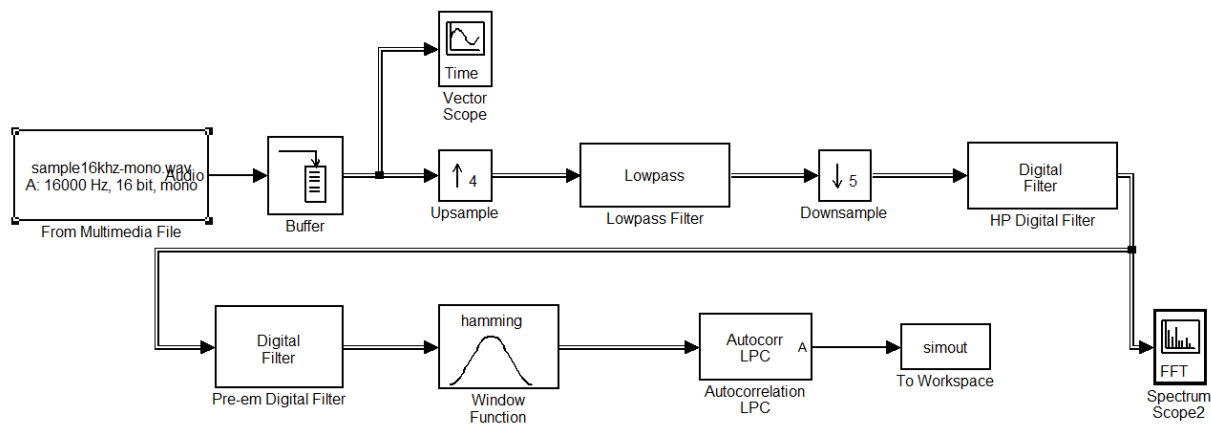
Předchozí část práce byla věnována teoretickému rozboru širokopásmového kodeku AMR-WB založeného na technologii ACELP. Tato část práce je zaměřena na praktickou ukázkou části algoritmu kodeku, resp. její simulaci v programu Matlab-Simulink. Vybrané bloky budou popsány a ve vybraných mezikrocích bude zobrazen náhled průběhu signálu.

3.2 Simulace předzpracování a LPC analýzy

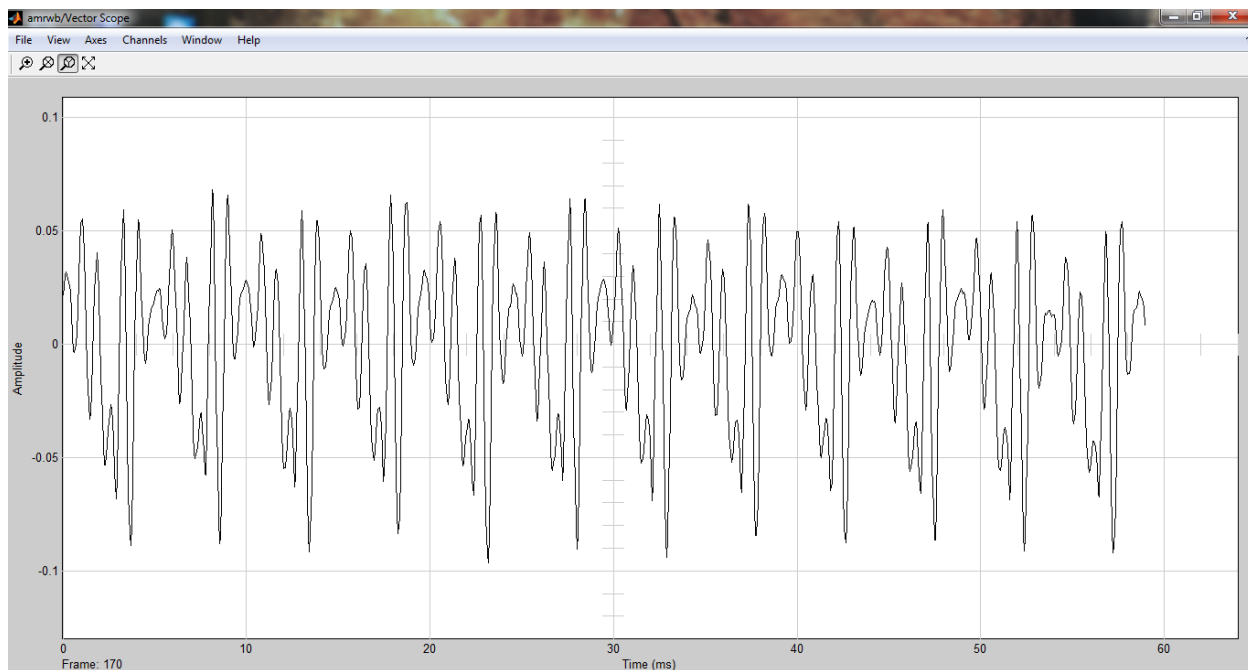
Úvodní část algoritmu kodeku se stará o předpřípravu řečového signálu a o simulaci LPC analýzy. Vstupní signál má vzorkovací kmitočet 16 kHz a jako první prochází přes blok bufferu, který zajišťuje překrývání jednotlivých rámců. V dalším bloku je signál nadvzorkován na čtyřnásobek. Následně projde filtrem typu dolní propust s mezním kmitočtem 6,4 kHz a v dalším bloku je signál podvzorkován na 12,8 kHz. Poté je pro potlačení nízkých kmitočtů signál filtrován horní propustí, která je popsána vztahem podle [13]:

$$H_{h_1}(z) = \frac{0,989502 - 1,979004z^{-1} + 0,989502z^{-2}}{1 - 1,978882z^{-1} + 0,9799126z^{-2}} \quad (3.1)$$

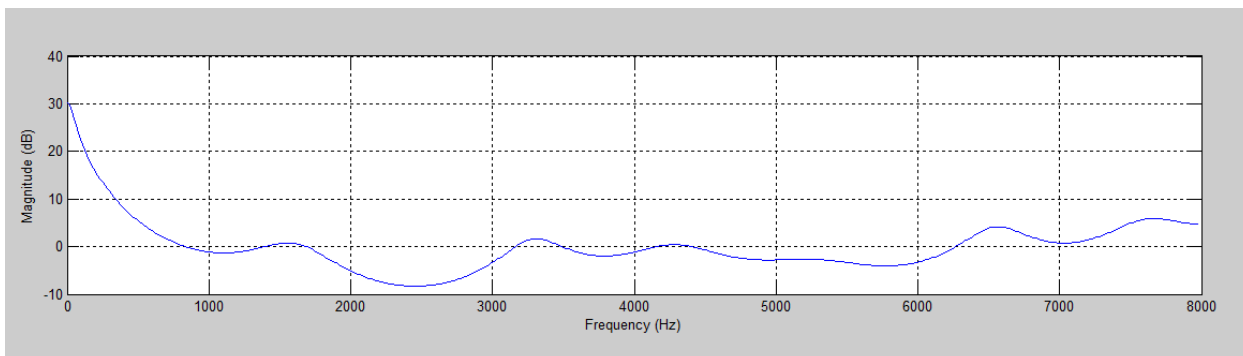
Následuje další filtr typu horní propust, který má za úkol zdůraznit vysoké kmitočty. Tento filtr popisuje rovnice $H_{pre-emph}(z) = 1 - 0,68z^{-1}$, podle [13]. Dále jsou vytvořeny řečové segmenty pomocí váhové posloupnosti Hammingova okna, na které je použita autokorelační funkce. Získané koeficienty jsou použity k výpočtu algoritmu Levinson-Durbin, jehož výsledkem jsou ISP parametry řečového signálu [13]. Průběhy signálu snímané jednotlivými sondami znázorňují následující obrázky.



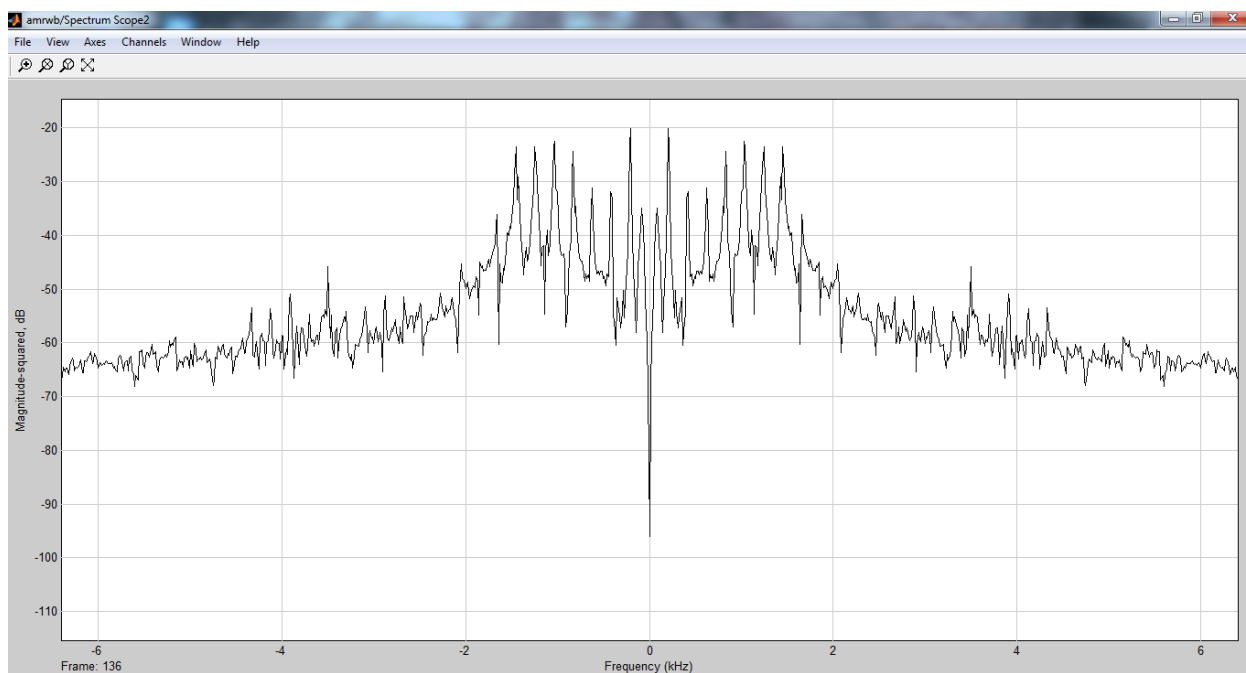
Obr. 3.1: Zapojení bloků pro předzpracování a LPC analýzu



Obr. 3.2: Časový průběh vstupního signálu



Obr. 3.3: Kmitočtová odezva filtru, lokální maxima znázorňují formanty



Obr. 3.4: Spektrum předzpracovaného signálu

4 ZÁVĚR

V úvodu bakalářské práce jsem popsal některé historické kódovací metody používané v mobilních sítích druhé a třetí generace. Postupným vývojem těchto algoritmů dochází k optimalizaci využití šířky pásma za použití složitějších kódovacích algoritmů s dříve nerealizovatelně vysokou výpočetní náročností.

V práci jsem popsal a v Matlabu ověřil principy kodeku AMR-WB, který oproti předchozím kodekům umožňuje přenášet širší pásmo řečového signálu i za horších rádiových podmínek a tak přispět k lepšímu vjemu a snadnější identifikaci mluvčího. Ze současného vývoje je patrné, že je výhodnější přenášet pouhé parametry řečového signálu místo celých vzorků, což je umožněno díky vysokému výpočetnímu výkonu a nízké ceně současných mobilních zařízení. S velmi rychle rostoucím počtem zařízení se tato optimalizace jeví jako nutná pro další vývoj, neboť kapacita přenosových cest je omezená a poptávka po datových službách vyžadujících volnou kapacitu sítě prudce stoupá. Nezanedbatelnou výhodou kodeku AMR-WB je i jeho univerzálnost, kdy může být použit v mobilních i pevných sítích a lze ho použít i pro stále více rozšířenou VoIP komunikaci.

LITERATURA

- [1] ŠIMÁK, B. *Principy zpracování hlasu v klasické a IP telefonii* [online]. Praha: České vysoké učení technické v Praze, Fakulta elektrotechnická, Katedra telekomunikační techniky, [cit. 2013-02-12]. Dostupné z URL: <<http://www.itpoint.cz/voip/?i=principy-zpracovani-hlasu-13>>.
- [2] JELÍNEK, M. *Moderní komprese řeči* Université de Sherbrooke, Kanada, 2004 [cit. 2013-02-12].
- [3] ČÍKA, P. *Multimediální služby* Brno: Vysoké učení technické v Brně, 2012. ISBN: 978-80-214-4443- 0. (cs).
- [4] BESSETTE, B.; SALAMI, R.; LEFEBVRE, R.; JELÍNEK, M.; ROTOLAPUKKILA, J.; VAINIO, J.; MIKKOLA, H.; JARVINEN, K. *The adaptive multi-rate wideband speech codec (AMR-WB)* Speech and Audio Processing, IEEE Transactions on , vol.10, no.8, pp.620,636, Nov 2002 (en).
- [5] ITU. *Wideband Coding of Speech at Around 16 kbit/s Using Adaptive Multi-Rate Wideband (AMR-WB)* Geneva, ITU-T Recommend.G.722.2, 2002.
- [6] JÄRVINEN, K.; VAINIO, J.; KAPANEN, P.; HONKANEN, T.; HAAVISTO, P.; SALAMI, R.; LAFLAMME, C.; ADOUL, J. *GSM enhanced full rate codec* IEEE Int.Conf.Acoustics, Speech, Signal Processing(ICASSP), Munich, Germany, Apr. 20–24, 1997, pp. 771–774.
- [7] 3GPP. *AMR Wideband Speech Codec; General Description*, 3GPP TS 26.171.
- [8] 3GPP. *AMR Wideband Speech Codec; Source Controlled Rate Operation*, 3GPP TS 26.193.
- [9] SALAMI, R.; LAFLAMME, C.; ADOUL, J.; MASSALOUX, D. *A toll quality 8 kb/s speech codec for the personal communications system (PCS)* IEEE Trans. Vehicular Technol., vol. 43, pp. 808–816, Aug. 1994.
- [10] HANUS, S. *Rádiové a mobilní komunikace* Skripta FEKT VUT v Brně. Brno, 2009.
- [11] AHMADI, S. *Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB), Service Options 62 and 63 for Spread Spectrum Systems* TIA1016-A March 18, 2005

- [12] ORDENTLICH, E.; SHOYAM Y. *Low-delaycode-excitedlinearpredictive coding of wideband speech at 32 kbps*. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Toronto, ON, Canada, May14–17,1991, pp. 9–12.
- [13] 3GPP. *Speech codec speech processing functions; Adaptive Multi-Rate – Wideband (AMR-WB) speech codec; Transcoding functions. TS 26.190* 3rd Generation Partnership Project (3GPP), March 2001.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

3G Third generation of mobile telecommunications technology

3GPP 3rd Generation Partnership Project

μ -law Quantizer μ -law

A-law Quantizer A-law

ACELP Algebraic Code Excited Linear Prediction

ADC Analog to Digital Converter

ADPCM Adaptive Differential Pulse Code Modulation

AMR-NB Adaptive Multi-Rate Narrowband

AMR-WB Adaptive Multi-Rate Wideband

CELP Code Excited Linear Prediction

CNG Comfort Noise Generation

DAC Digital to Analog Converter

DPCM Differential Pulse Code Modulation

DSP Digital signal processing

DTX Discontinuous Transmission

EFR Enhanced Full Rate

ETSI European Telecommunications Standards Institute

FR Full Rate

GERAN GSM EDGE Radio Access Network

GSM Global System for Mobile

HR Half Rate

ISDN Integrated Services Digital Network

ISP Immittance Spectral Pairs

ISPP Interleaved Single-Pulse Permutation

ITU International Telecommunication Union

LP Linear Prediction

LPC Linear Predictive Coding

LPF Linear Prediction Filter

PCM Pulse Code Modulation

PCS Personal Communications Service

SID Silence Descriptor

SCR Source Controlled Rate

SNR Signal to Noise Ratio

UTRAN UMTS Terrestrial Radio Access Network

VAD Voice Activity Detector

VC Voice Coder

VoIP Voice over Internet Protocol

WCDMA Wideband Code Division Multiple Access

WMOPS Weighted Million Operations Per Second

A OBSAH CD

readme.txt	soubor s nápovědou
amr-wb	složka obsahující zdrojové soubory
bakalářská práce	složka obsahující text bakalářské práce