

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATION

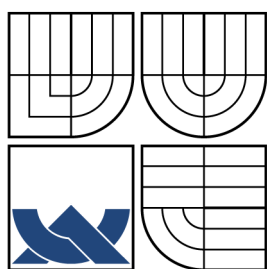
HLASOVÉ KODÉRY PRO NÍZKÉ PŘENOSOVÉ RYCHLOSTI

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

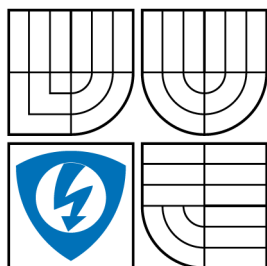
AUTOR PRÁCE
AUTHOR

BC. JAKUB LEITNER

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 TELECOMMUNICATION

HLASOVÉ KODÉRY PRO NÍZKÉ PŘENOSOVÉ RYCHLOSTI LOW BIT RATE VOICE ENCODERS

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

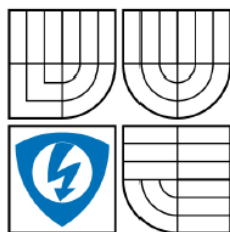
AUTOR PRÁCE
AUTHOR

BC. JAKUB LEITNER

VEDOUCÍ PRÁCE
SUPERVISOR

ING. RADIM PUST

BRNO 2009



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Diplomová práce

magisterský navazující studijní obor
Telekomunikační a informační technika

Student: Bc. Jakub Leitner
Ročník: 2

ID: 85282
Akademický rok: 2008/2009

NÁZEV TÉMATU:

Hlasové kodéry pro nízké přenosové rychlosti

POKYNY PRO VYPRACOVÁNÍ:

Cílem diplomové práce je vyhledat a popsat kodéry a vokodéry určené pro přenos hlasu, zaměřit se hlavně na typy vhodné pro komunikační kanály s velmi nízkými přenosovými rychlostmi. Následně uvedené kodéry a vokodéry mezi sebou vzájemně porovnat pomocí vhodně zvolených kritérií. Dále vytvořit simulaci kodéru a vokodéru v prostředí programu Matlab.

DOPORUČENÁ LITERATURA:

- [1] GOLD, Ben, MORGAN, Nelson. Speech and Audio Signal Processing : Processing and Perception of Speech and Music. [s.l.] : Wiley, 1999. 560 s. ISBN 978-0-471-35154-2.
[2] KARBAN, Pavel. Výpočty a simulace v programech Matlab a Simulink. Brno : Computer Press, 2006. 220 s. ISBN 80-251-1301-9.

Termín zadání: 9.2.2009

Termín odevzdání: 26.5.2009

Vedoucí práce: Ing. Radim Pust

prof. Ing. Kamil Vrba, CSc.
Předseda oborové rady

UPOZORNĚNÍ:

Autor diplomové práce nesmí při vytváření diplomové 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

Diplomová práce se zabývá problematikou kodérů a vokodérů používaných ve zpracování řečového signálu. Cílem práce je vytvořit ucelený přehled koderů a vokodérů včetně popisu jejich vlastností, v druhé části pak simulace algoritmů a metod pro zpracování řeči v programu Matlab Simulink. Simulovány byly základní metody pro zpracování řeči v časové oblasti a parametrický LPC vokodér. V modelu LPC vokodéru je implementováno několik algoritmů pro získání příznaků neboli parametrů segmentu řečového signálu. Jsou to: metoda pro určení znělosti segmentu řečového signálu, LPC analýza, odhad frekvence základního tónu řeči. Výstupem je tedy parametrizovaný signál, z něhož je možné na přijímací straně řečový signál obnovit. V příloze na konci práce je uvedena tabulka názvů kodérů, případně číslo normy ve které je kódér standardizován, a jejich vlastnosti. Druhou přílohou je přehled metod zpracování řečového signálu.

KLÍČOVÁ SLOVA

řeč, hlas, kodér, vokodér

ABSTRACT

The final thesis deals with coders and voice coders used in speech signal processing. The aim is to create an integral overview of coders and voice coders including a description of their properties, in the second part of the thesis a simulation of algorithms and methods of speech processing is performed in Matlab Simulink program. The basic methods of speech processing and a parametric LPC voice coder were simulated in time domain. In the LPC voice coder model there are implemented the algorithms for obtaining speech segment parameters. These are the algorithm for classification of voiced and unvoiced speech segment, LPC analysis and pitch detection. The output is a parametric signal that enables a receiver to synthesize a speech signal. The appendix 1 contains a list of names of coders or standard numbers of coders and their properties, the appendix 2 includes an overview of speech processing methods.

KEYWORDS

speech, voice, encoder, vocoder

Leitner J. *Hlasové kodéry pro nízké přenosové rychlosti*. Brno: VUT Brno. Fakulta elektrotechniky a komunikačních technologií. Ústav telekomunikací, 2009. 72 s., 10 s. příloh. Diplomová práce. Vedoucí práce byl Ing. Radim Pust.

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Hlasové kodéry pro nízké přenosové rychlosti“ jsem vypracoval samostatně pod vedením vedoucího diplomové 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 diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne

.....

(podpis autora)

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu své diplomové práce panu Ing. Radimu Pustovi za cenné rady, metodické pokyny a čas, který věnoval mé diplomové práci a za jeho osobní přístup při konzultacích.

V Brně dne

.....
(podpis autora)

OBSAH

Úvod	10
1 Řečový signál	11
1.1 Hlasové ústrojí	11
1.1.1 Hlásky	13
1.1.2 Vědy	13
2 Digitalizace řečového signálu	15
2.1 Zdrojové kódování	15
2.1.1 Kódování v časové oblasti	16
2.1.2 Parametrické kódování zdroje	20
2.1.3 Vektorové kvantování	23
2.1.4 Hybridní kódování zdroje	23
3 Kodéry řečového signálu	27
3.1 Kvalita přenosu a zpracování řeči	27
3.2 Vlastnosti kodérů	29
3.3 Kodéry	29
3.3.1 ITU-G.711	29
3.3.2 ITU-G.718	30
3.3.3 ITU-G.722	31
3.3.4 ITU-G.723.1	32
3.3.5 ITU-G.726	33
3.3.6 ITU-G.727	33
3.3.7 ITU-G.728	34
3.3.8 ITU-G.729	35
3.3.9 GSM	36
3.3.10 VMR-WB	39
3.3.11 ISAC, ILBC	39
3.3.12 TETRA	39
3.3.13 AMBE	40
3.3.14 Speex	40
3.3.15 DSS	40
3.3.16 HVXC	41
4 Modelování v programu Matlab	
Simulink	42

5 Závěr	54
Literatura	56
Seznam symbolů, veličin a zkratk	58
Seznam příloh	61
A Obsah elektronického média	62
B Přehled	63
B.1 Přehled vlastností kodérů a vokodérů	63
B.2 Rozdělení metod zpracování řečového signálu	64
C Modely v programu Simulink	65
C.1 Model pro simulaci metod zpracování řečového signálu v časové oblasti	65
C.2 Model DM	66
C.3 Model DM - průběh signálů	66
C.4 Model PCM	67
C.5 Model PCM - průběh signálů	67
C.6 Model DPCM	68
C.7 Model DPCM - průběh signálů	68
C.8 Model ADPCM	69
C.9 Model ADPCM - průběh signálů	69
C.10 Model LPC vokodéru	70
C.11 Subsystem pro výpočet frekvence základního tónu	71
C.12 Subsystem pro určení znělosti segmentu signálu	72

SEZNAM OBRÁZKŮ

1.1	Hlasový trakt	11
1.2	Schéma hlasového traktu	12
2.1	Blokové schéma PCM	16
2.2	Blokové schéma DM	17
2.3	Blokové schéma ADM	18
2.4	Blokové schéma DPCM	19
2.5	Blokové schéma ADPCM	20
2.6	Analyzátor vokodéru	21
2.7	Syntezátor vokodéru	22
3.1	Kompresní charakteristika A-law	30
3.2	Kompresní charakteristika u-law	31
4.1	Frekvenční charakteristika filtrů emfáze	43
4.2	Průběh energie a koeficientu počtu průchodu signálu nulou, binární signál určující příznak znělý/neznělý úsek	47
4.3	Průběh frekvence základního tónu	49
4.4	Autokorelace signálu a omezení autokorelace	50
4.5	Průběh budícího signálu s frekvencí 160 Hz	50
4.6	Časový průběh originálního a syntetizovaného řečového signálu	52
4.7	Spektrogram originálního a syntetizovaného řečového signálu	53
B.1	Přehled vlastností kodérů a vokodérů	63
B.2	Rozdělení metod zpracování řečového signálu	64
C.1	Obecný model pro simulaci	65
C.2	Model delta modulace	66
C.3	Průběh signálů delta modulace	66
C.4	Model PCM	67
C.5	Průběh signálů pulzní kódové modulace	67
C.6	Model diferenční PCM	68
C.7	Průběh signálů diferenční pulzní kódové modulace	68
C.8	Model ADPCM	69
C.9	Průběh signálů adaptivní diferenční pulzní kódové modulace	69
C.10	Model LPC vokodéru	70
C.11	Subsystem pro výpočet frekvence základního tónu	71
C.12	Subsystem pro výpočet frekvence základního tónu	72

ÚVOD

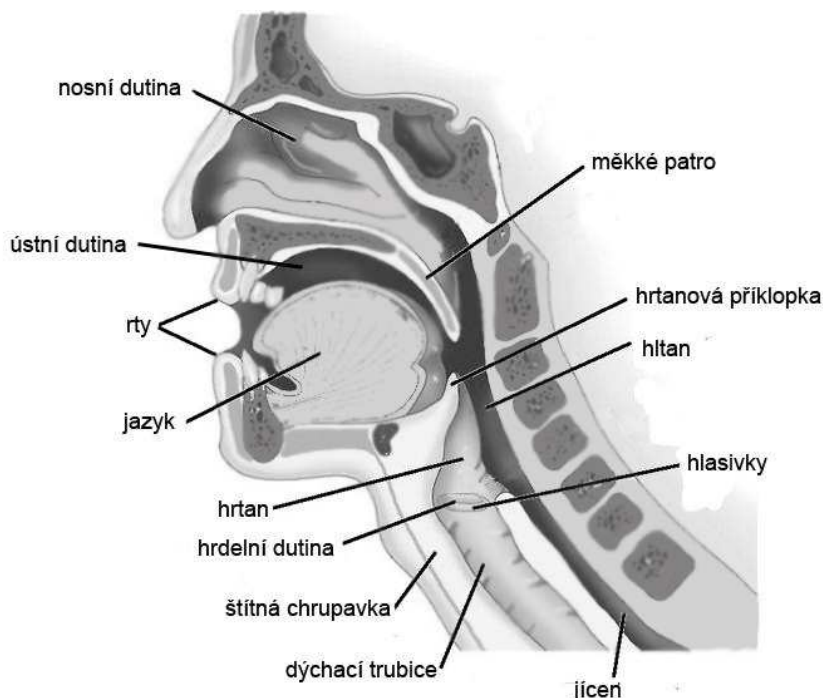
Diplomová práce se zabývá problematikou kodérů a vokodérů pro zpracování řečového signálu. Lidé mají jako jediní tvorové na Zemi výsadu mluvit. Tvorba řeči je v teoretické rovině v podstatě jednoduchý proces. Proud vydechovaného vzduchu rozkmitává hlasivky, které vytváří hlásky a ty jsou pak dále pomocí ústní dutiny a artikulačních orgánů dotvářeny. Avšak popis tvorby řeči ve fyzikální rovině není již tak triviální, tudíž ani zpracování řečového signálu není jednoduchým procesem. Jako „referenční“ metoda zpracování řečového signálu se používá PCM, používaná v klasických telefonních sítích, s výstupní bitovou rychlostí 64 kbit/s. Pod touto bitovou rychlostí se pohybují výstupní bitové rychlosti ostatních kodérů, ať už jsou založeny na jakékoliv metodě zpracování signálu. První kapitola se zabývá hlasovým ústrojím, tvorbou řeči, dělením hlásek a krátce také rozděluje vědy, které se řečí zabývají, ať už z jakéhokoli hlediska. V druhé kapitole s názvem Digitalizace řečového signálu byly shrnuty metody pro zpracování řeči a byly rozděleny do čtyř podkapitol. První část je věnována zpracování signálu v časové oblasti, zde jsou popsány základní metody digitalizace a zpracování signálu. Ve druhé podkapitole s názvem Parametrické kódování se práce věnuje popisu modelu vokálního traktu, jeho parametrizaci a následnému využití ve zpracování řečového signálu. V podkapitole Vektorové kvantování je obecný popis, jakým způsobem je možné využít techniky kódových knih, a poslední část Hybridní kódování zdroje popisuje algoritmy, které využívají ke zpracování signálu kombinace některých metod uváděných v předchozích kapitolách. Třetí kapitola s názvem Kodéry řečového signálu začíná dvěma částmi, ve kterých je popsána metodika subjektivního hodnocení kvality řečového signálu po obnovení respektive přenosu hlasu telekomunikační sítí a parametry kodérů, na jejichž základě je možné srovnávat je z technického hlediska. Následuje část s výčtem konkrétních standardů, kodérů a vokodérů, včetně jejich popisu. Další kapitola nazvaná Modelování v programu Matlab Simulink se již věnuje samotné simulaci a popisu modelu v programu Simulink. Namodelovány jsou základní metody zpracování signálu v časové oblasti, DM, PCM, DPCM a ADPCM a LPC vokodér, který je základem většiny moderních vokodérů pro zpracování hlasu. Algoritmy pro zpracování v časové doméně jsou implementovány do subsystému v obecném modelu pro jejich testování. LPC vokodér je tvořen samostatným modelem, který obsahuje subsystémy s dílčími algoritmy pro parametrizaci řeči. Průběhy konkrétních signálů v modelech je možno vidět v přílohách diplomové práce nebo v kapitole Modelování v programu Matlab Simulink.

1 ŘEČOVÝ SIGNÁL

Řeč je nejčastější zvukový projev člověka sloužící zpravidla ke komunikaci s druhou osobou, v dnešní době také se strojem. Každá řeč je složena ze slov, která tvoří lexikální systém, jednodušeji slovník, daného jazyka. Slova se skládají z hlásek, jež tvoří tzv. fonémický systém. Slova nemůžeme klasifikovat jako pouhé dorozumívací signály odvozující svůj význam pouze z dané situace, nýbrž jsou to symboly s kulturně historickým významem. Jedno slovo pak může mít v různých situacích rozdílné významy. Řazení a spojování slov není libovolné a řídí se jistými pravidly, gramatikou daného jazyka.

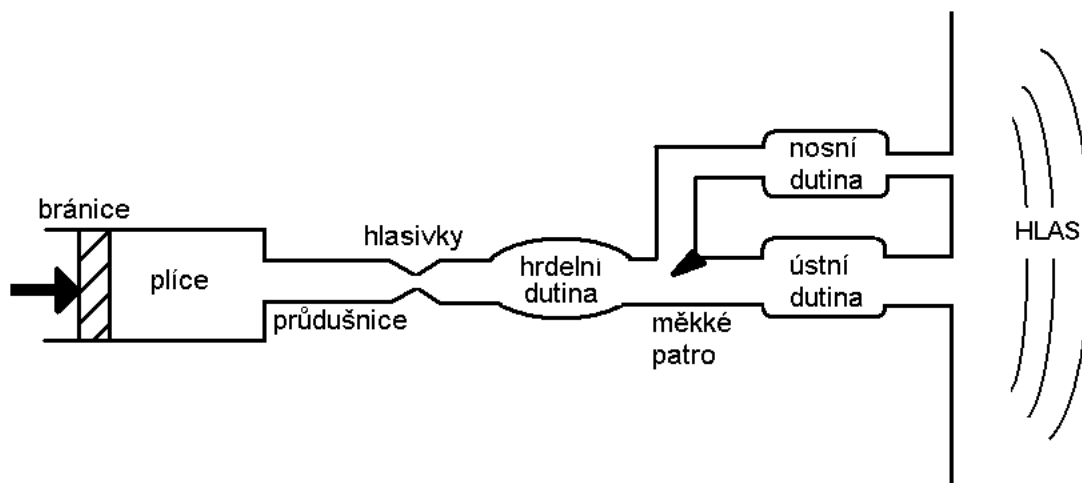
1.1 Hlasové ústrojí

Hlasový, hovorový nebo také řečový signál vzniká díky několika orgánům lidského těla. Základní hlas vytváří orgán zvaný hlasivky. Hlasivky jsou uloženy v hrtanu a tvoří je párový hlasivkový sval, hlasivkový vaz a slizniční hlasivková řasa, napjaté mezi konvicovitou a štítnou chrupavkou. Vzduch, který prochází při výdechu hlasivkovou štěrbinou mezi nimi, může rozechvívát napjaté hlasivkové řasy. Tak vzniká základní hlas (základní tón, kmitočet f_0). Pohybuje se v rozmezí od 150–400 Hz. Základní tón se uplatňuje u tvorby hlásek znělých. Těsně nad hlasivkami se na-



Obr. 1.1: Hlasový trakt

chází dutina hrdelní. Ta za pomoci svalů hrdla a kořene jazyka mění svůj objem a uplatňuje se při tvorbě samohlásek. Nosní dutina se podílí na tvorbě minima hlásek, (typicky m, n, ň). Naopak ústní dutina je rezonanční prostor, který se podílí na tvorbě každé hlásky. Její objem se mění pomocí neméně důležitých částí hlasového traktu, měkkého patra, tvrdého patra, jazyka a zubů. Tyto poslední články vokálního traktu se nazývají artikulační a jejich prioritním úkolem je zajistit srozumitelnost vyslovované hlásky respektive slova.



Obr. 1.2: Schéma hlasového traktu

Celý hlasový trakt je vybuzen proudícím vzduchem z plic, které roztahuje a smršťuje svalovošlachová blána neboli bránice. Poměr vdechu a výdechu i objem vdechovaného a vydechovaného vzduchu se mění podle emočního a fyzického stavu mluvčího. Dýchání je proces reflexní, ale při řeči se z části mění na proces vědomý a to z toho důvodu, že člověk mění tempo výdechů a nádechů nepravidelně a používá tzv. přídechy.

Aby mohl být hlasový trakt lépe popisován, bylo vytvořeno jednoduché schéma hlasového traktu 1.2, v kterém jsou zakresleny všechny části podílející se na tvorbě hlasu. Vynechány jsou části artikulační, jazyk, zuby a rty 1.1.

1.1.1 Hlášky

Všechny hlásky můžeme rozdělit na dvě základní skupiny:

- Samohlásky – při tvorbě samohlásek má hlasový trakt za úkol nechat procházet proud vzduchu volně. Hlasivky pak svým kmitáním generují základní tón f_0 a rezonance v dutinách hlasového traktu vytváří další vyšší tóny, které nazýváme formanty. Ty se značí $f_1, f_2, \dots, f_n$ a dodávají řeči její barvu a sluchový dojem. Vzájemný poměr formantů f_1 a f_2 je důležitý pro rozlišování jednotlivých samohlásek.
- Souhlásky – na rozdíl od samohlásek se objevuje v akustickém spektru souhlásek šum. Šum je vytvářen třením proudu vzduchu o překážky hlasového traktu. Typickými překážkami hlasového traktu jsou jazyk, zuby nebo rty. Jinak je můžeme nazvat také artikulačními orgány.

Dělení souhlásek podle typu vytvořené překážky:

- Závěrové, ražené – před vznikem souhlásky dochází k přerušení proudu vzduchu a jeho náhlému uvolnění, vybuzení je pak tvořeno krátkodobým šumem tvaru krátkého pulsu – b, p, m, n, d, t, ě, ě, ř, g, k.
- Úžinové, třené – zúžením výdechové cesty vzniká v místě zúžení třecí šum, který navazuje na předchozí hlásky – f, v, s, z, š, ž, j, ch, h, l, r, ř.
- Polozávěrové, polotřené – souhlásky tohoto typu jsou tvořeny kombinací předchozích případů. To znamená, že budící signál je krátce přerušen, pak následuje rychlé uvolnění proudu vzduchu a následně je vytvořena zúžená část – c, č, dz, dž.

A dále pak rozdělení na:

- Znělé – kmitající hlasivky dávají znělým souhláskám periodický charakter.
- Neznělé – hlasivky nekmitají, jsou buď otevřené, přivřené nebo se mohou rychle otevřít či zavřít podle souhlásky, která má být vytvořena.

1.1.2 Vědy

Vědy zabývající se řečí:

1. Humanitní vědy

- Fonetika – jazykovědná disciplína, zkoumající základní výrazový plán jazyka, zvukovou realizaci lidské řeči, fyziologii řeči.

- fyziologicko - artikulační - popisuje zvukovou stránku jazyka z hlediska mluvčího, tvorbu jednotlivých hlásek, jejich spojování
- akustická - popis z hlediska posluchače, vnímání zvukových signálů
- funkční - popis složek zvuku z hlediska jejich přítomnosti a nadbytečnosti.
- Fonologie – jazykovědná disciplína, zabývající se strukturou plánu jazyka, vztahy mezi zvuky, významotvorné prvky. Základní jednotkou fonologie je foném, který má v jazyce úlohu významotvornou, proto jeho změna může měnit význam slova.
- Prosodie – věda zabývající se zvukovou stránkou jazyka, melodií, délkou hlásek, přízvukem ve slovech.
- Gramatika – tvoří soubor pravidel pro syntézu, tvorbu vět.
- Lexikologie – zkoumá slovní zásobu jazyka, vztahy mezi slovy.

2. Přírodní vědy

- Fyziologie – zkoumá mechanické, fyzické a biochemické procesy živých organismů, tedy i funkci hlasového traktu, a napomáhá tak vytváření jeho modelů.
- Akustika – studuje fyzikální mechanismy tvorby, vnímání a charakter akustického, tedy i řečového signálu.

3. Technické vědy

- Zpracování signálu – modelování, zpracování, parametrizace, spektrální analýza, identifikace kódování, teorie informace, syntéza řeči.

2 DIGITALIZACE ŘEČOVÉHO SIGNÁLU

Jakákoliv informace, která má být zaznamenána, uchována na libovolném médiu či přenesena komunikačními systémy, musí být nejdříve převedena na elektrický signál. K tomu účelu slouží převodníky mechanických, optických, zvukových a jiných veličin na veličiny elektrické, nejčastěji napětí. Vytvořený elektrický signál, obvykle v analogové podobě, se převádí na signál digitální v analogově-digitálním převodníku. [3].

Výše zmíněné skutečnosti platí samozřejmě i pro řečový signál. Řečový signál má sice svá specifika, ale lze na něj aplikovat i obecné metody zpracování signálů. I když v dnešní době jsou kapacity paměťových médií zdá se někdy nevyčerpatelné, neplatí to bohužel o přírodním bohatství jako je frekvenční spektrum. Proto se stále hledají a vyvíjí specializované metody pro zpracování řečového signálu, které umožňují vyšší kompresi a tím i snižují požadavky na kapacitu přenosového kanálu. Klasickými metodami zpracování signálu aplikovanými na řečový signál však není možné dosáhnout tak výrazné úspory jako metodami specializovanými. Úroveň komprese se vyjadřuje veličinou zvanou kompresní poměr a je dán poměrem mezi vstupní a výstupní velikostí souboru.

2.1 Zdrojové kódování

Úkolem zdrojového kodéru je odstranit popřípadě alespoň snížit redundantní a irrelevantní informace signálu. Proces zdrojového kódování je nutný především při zpracování řečových (ale i zvukových a obrazových) signálů, které jsou vysoce redundantní a obsahují také značné množství irrelevantní informace.

- Redundance – nadbytečnost, je definována jako množství dat, které lze při přenosu nebo uložení vynechat, aniž by byla ztracena užitečná informace. Odstranění redundance signálu je proces nevratný.
- Irrelevance – zbytečnost, je definována jako nepodstatná složka signálu, kterou není nutno přenášet z toho důvodu, že příjemce není schopen tuto složku vnímat. Tudíž z pohledu příjemce nemá žádný vliv na charakter informace. Odstranění irrelevance signálu je proces nevratný.

Pro zdrojové kódování řečových signálů (maximální frekvence je 4 kHz) se používají kodéry, které je možné dělit do tří základních skupin. První z nich jsou kodéry tvarového průběhu (Waveform Coders), realizující tzv. zdrojové kódování tvaru vlny. Jsou konstruovány tak, aby se časový průběh analogového signálu na výstupu dekodéru co nejvíce shodoval s časovým průběhem analogového signálu na vstupu

kodéru. Druhou skupinu tvoří vokodéry (Vocoders, Voice Coders), realizující tzv. parametrické zdrojové kódování. Při tomto způsobu kódování není přenášen původní signál, ale pouze jeho charakteristické parametry, vytvořené na základě analýzy původního signálu. Na přijímací straně je pomocí těchto parametrů řízen syntezátor hovorových signálů. Reprodukovaný signál má výrazný syntetický charakter. Třetí skupinu tvoří kodéry hybridní, které vznikají kombinací kodérů obou předchozích skupin a vhodným způsobem spojují jejich přednosti. Realizují hybridní zdrojové kódování. [3].

2.1.1 Kódování v časové oblasti

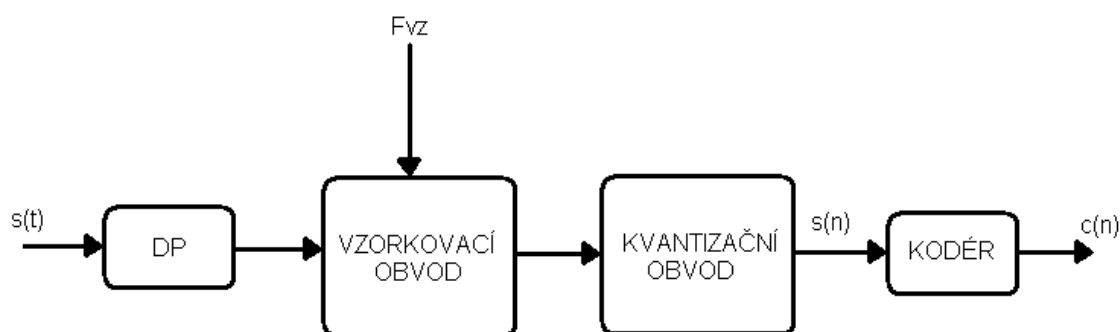
Pulse Code Modulation (PCM)

PCM je nejjednodušší způsob převodu analogového signálu na signál digitální. Byla vytvořena H.A. Reevesem v roce 1938. Převod A/D probíhá ve třech krocích: vzorkování, kvantování a kódování C.4.

Vzorkování: v diskrétním čase je odebírán vzorek ze spojitého signálu, s tou podmínkou, že je splněn tzv. Nyquistův teorém $f_{vz} = 2 \cdot f_{max}$ ¹, f_{max} je pak maximální možnou frekvencí, kterou je možno navzorkovat bez aliasingu. Proto je ještě před vzorkováním signál frekvenčně omezen dolní propustí.

Kvantování: každý odebraný vzorek může nabývat hodnoty z oboru reálných čísel, proto je nutné určit diskrétní úrovně, ke kterým bude hodnota vzorku zaokrouhlena. Počet úrovní m a počet bitů kódového slova N je dán vztahem $m = 2^N$.

Kódování: každému kvantovanému vzorku je přiřazen binární kód, kódové slovo. Pro analogový řečový signál v rozsahu 300 – 3400 Hz, je podle vzorkovacího teorému



Obr. 2.1: Blokové schéma PCM

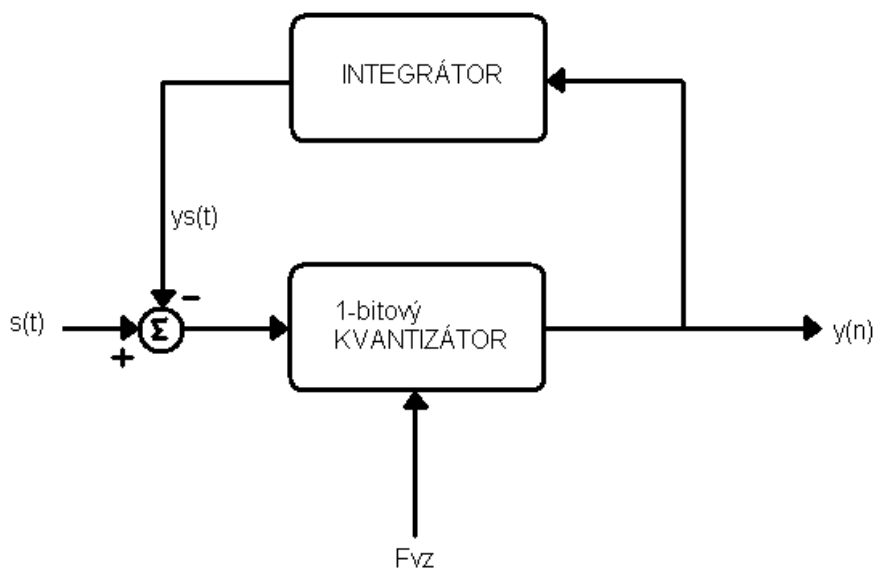
zvolena vzorkovací frekvence 8 kHz. Počet kvantovacích hladin je určen dynamikou

¹<www.fit.vutbr.cz/study/courses/ISS/public/pred/vzork/vzork.pdf>

signálu. Hodnota každého vzorku je zaokrouhlena na jednu z 256 kvantovacích hladin. Přenosová rychlost digitálního signálu na výstupu je tedy 8000 vzorků krát 8 bitů = 64000 bitů za sekundu. Bitová rychlost 64 kbit/s je tedy základní přenosová rychlost pro přenos hlasového signálu a můžeme ji považovat jako referenční hodnotu pro srovnávání dalších způsobů zdrojového kódování hlasu. Tento postup popisuje lineární kvantování. Aby byla zajištěna rovnoměrná hodnota kvantizačního zkreslení v celém pracovním rozsahu kodéru, je vhodné použít nelineárního rozložení kvantizačních stupňů. Realizace nelineárního rozložení spočívá ve změně dynamiky signálu. Na vysílací straně je použit kompresor, tzn., že malé hodnoty vzorků jsou zesíleny a velké hodnoty vzorků jsou zeslabeny. V přijímači je pak použit expandor, ve kterém probíhá inverzní proces.

Delta Modulation (DM)

Delta modulace umožňuje převádět signál na posloupnost logických symbolů 1 a -1, popřípadě 0 a 1. Místo hodnoty vzorku nebo difference oproti předcházejícímu se zaznamenává, respektive přenáší jen informace o tom, zda má aktuální vzorek nižší nebo vyšší hodnotu ve srovnání se vzorkem předcházejícím. Přesnost kódování musí být zajištěna dostatečnou vzorkovací frekvencí, a to tak, aby největší možná difference mezi vzorky byla blízká velikosti kvantizačního kroku. Vzorkovací frekvence je tedy několikrát vyšší než požadovaná $f_{vz} = 2 \cdot f_{max}$.



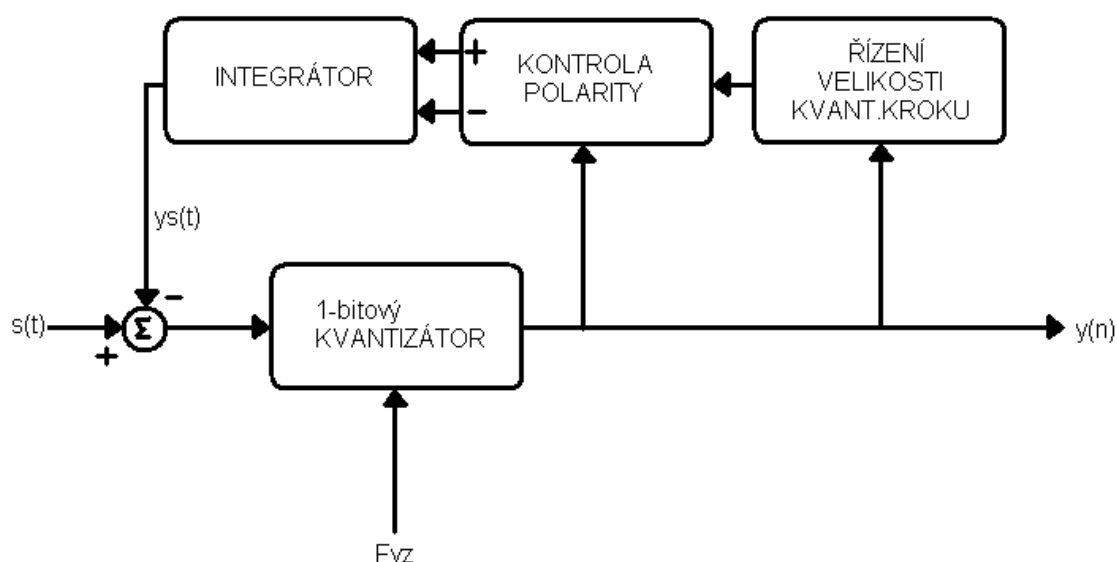
Obr. 2.2: Blokové schéma DM

Nevýhodou delta modulace je především zkreslení přetížením strmosti. Toto zkreslení se projevuje tím, že rekonstruovaný signál nesleduje skokové změny veli-

kosti vstupního signálu. Dalším problémem je tzv. zkreslení - granulační šum. Vzniká při kódování konstantního signálu, kdy rekonstruovaný signál střídavě mění v každé vzorkovací periodě svoji hodnotu o velikost kvantizačního kroku kolem původního skutečného průběhu. Pokud je pro libovolnou úroveň vstupního signálu kvantizační krok konstantní, nazývá se DM jako DM lineární C.2. Protože každý vzorek při DM je kódován pouze jedním bitem, je možné konstatovat, že bitová rychlost odpovídá vzorkovací frekvenci.

Adaptive Delta Modulation (ADM)

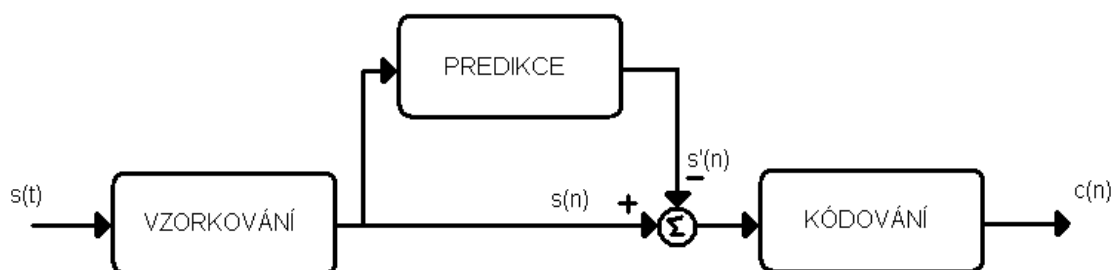
Adaptivní delta modulace používá proměnný kvantizační krok, proto je možné touto metodou zpracovávat signály s vyšší dynamikou než klasickou delta modulací. Kvantizační krok závisí na velikosti předchozího kvantizačního kroku a na hodnotě několika předchozích vzorků výstupního signálu. Velikost kvantizačního kroku je řízena buď sledováním změn amplitudy vstupního signálu nebo sledováním přetížení kodéru, kdy se na výstupu objevují stále stejné symboly. Pokud je kodér přetížen, pak řídicí blok kodéru zvětšuje velikost kvantizačního kroku. Po kolika symbolech stejné polarity zareaguje řídicí blok kodéru na přetížení, je dáno použitou metodou. V okamžiku, kdy se na výstupu opět objeví symboly opačné polarity, řídicí blok kodéru opět začne snižovat velikost kvantizačního kroku.



Obr. 2.3: Blokové schéma ADM

Differential Pulse Code Modulation (DPCM)

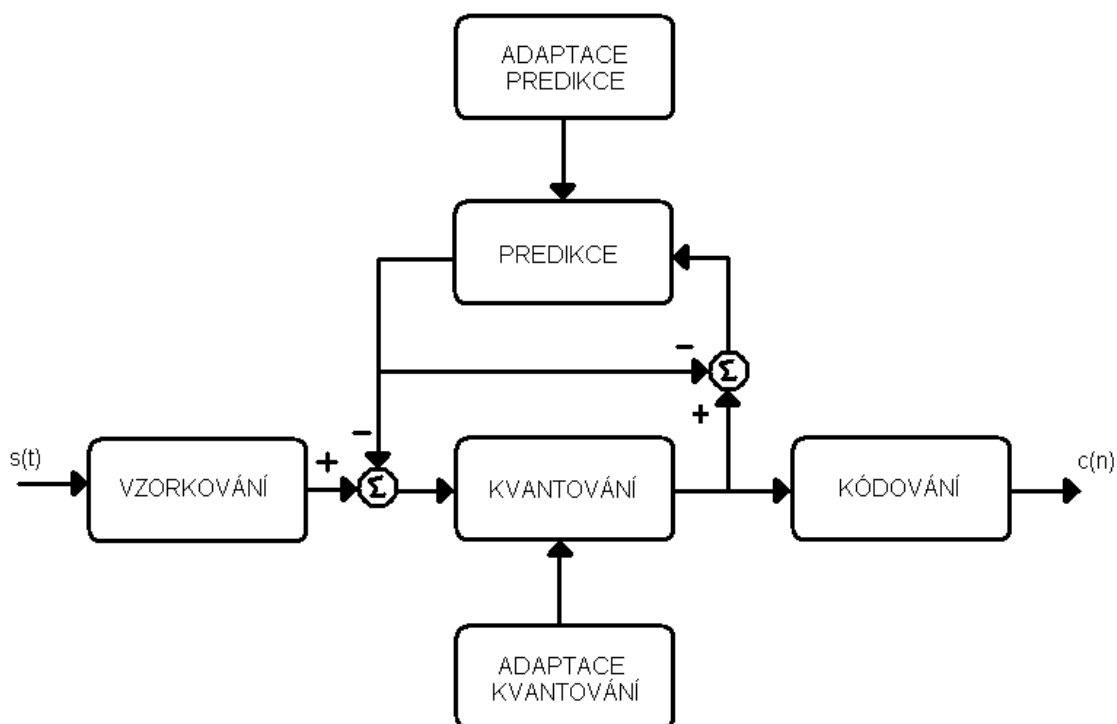
DPCM je modifikací PCM a jejím autorem je C. Cutler, který tuto metodu publikoval v roce 1956. DPCM kóduje rozdíly mezi okamžitou hodnotou vzorku signálu v dané vzorkovací periodě a hodnotou predikovanou z předchozích vzorků. Z vlastností signálů se totiž předpokládá (na základě korelace mezi vzorky), že rozdíl mezi současným vzorkem a hodnotou vzorku, predikovanou z několika předchozích, je menší než celková velikost současného vzorku. Z toho vyplývá, že k zakódování je třeba menší počet bitů a tím samozřejmě dochází k redukci výstupního bitového toku. Predikovaná hodnota vzorku slouží pouze pro výpočet difference, nadále pak představuje redundantní složku signálu a nepřenáší se. Blokové schéma je uvedeno na obrázku 2.4.



Obr. 2.4: Blokové schéma DPCM

Adaptive Differential Pulse Code Modulation (ADPCM)

Hovorové signály jsou svou podstatou nestacionární, takže jejich střední výkon a autokorelační funkce jsou proměnné s časem. Předchozí metody k tomuto faktu nepřihlíží. ADPCM je založeno na stejném principu jako DPCM s tím rozdílem, že místo lineárního kvantování a predikce je použito adaptivní kvantování a adaptivní predikci. Algoritmy používané k adaptaci jsou konstruovány jednoduše, protože změny průběhu řečového signálu jsou relativně pomalé. Velikost kvantovacího kroku je při adaptivním kvantování funkcí času, okamžiku vzorkování. Na rozdíl od ADM, kdy se kvantizační krok mění podle strmosti vstupního analogového signálu, se u ADPCM mění kvantizační krok za účelem redukce dynamického rozsahu kvantizačního šumu. Blokové schéma kodéru ADPCM znázorňuje obr. 2.5. Řídící signály pro adaptaci kvantování a adaptaci predikce jsou generovány na základě sledovaných parametrů výstupního a vstupního signálu.

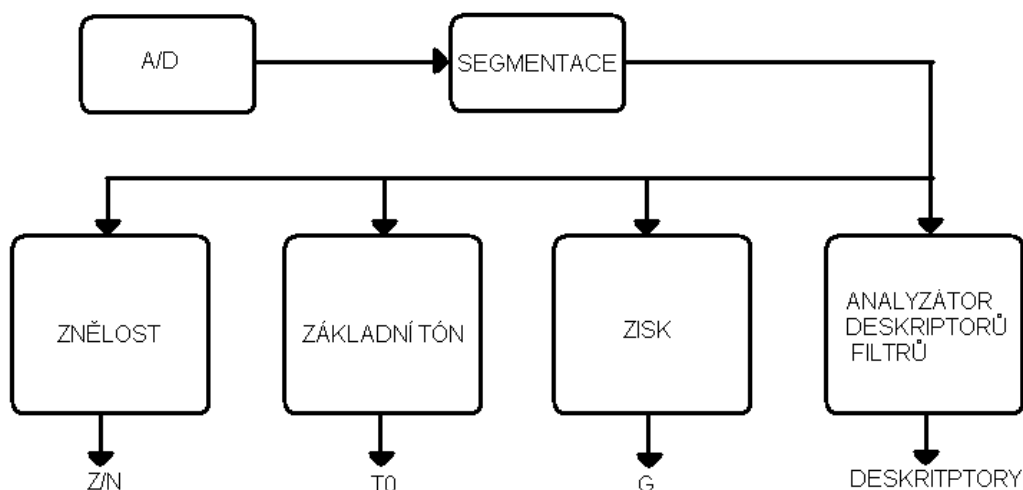


Obr. 2.5: Blokové schéma ADPCM

2.1.2 Parametrické kódování zdroje

Na základě znalosti fyziologického modelu hlasového traktu byl vytvořen jeho elektrický model určený pro syntézu řeči v přijímači. Při parametrickém kódování zdroje proto není přenášén původní řečový signál, ale pouze jeho charakteristické parametry, které jsou ve vysílači výsledkem analýzy vstupního řečového signálu, viz obr. 2.6. V přijímači je pomocí těchto parametrů řízen právě syntezátor řečového signálu. Charakteristické parametry řečového signálu tak neobsahují redundantní informace, což vede k velmi výraznému snížení bitové rychlosti. Snížení bitové rychlosti má samozřejmě i negativní stránku, tou je syntetický charakter reprodukováného signálu. Proto zní hlas nepřírozeně. Základem všech vokodérů je model pro syntézu řeči, viz obr. 2.7, který je schopen na základě charakteristických parametrů generovat řečový signál. Tato funkce vychází z poznatků, že řečový signál se skládá ze znělých a neznělých hlásek a mezer. Všechny hlásky jsou vytvářeny jako odezva filtru na vhodně zvolený excitační signál. Excitačním signálem je u generování znělých hlásek jejich základní kmitočet, který je tvořen periodickým sledem impulsů. Základní kmitočet se může měnit nejen s mluvčím, ale i u jednoho mluvčího. U neznělých hlásek je excitačním signálem šum, respektive náhodný signál, který má charakter bílého šumu. Filtr je samozřejmě proměnný, a tak je možné vytvářet odpovídající prezentaci znělých či neznělých hlásek. Vokodér využívající lineární predikci zpracovává

rámce většinou o velikosti kolem 20 ms. Výsledkem analýzy signálu jsou následující parametry: koeficienty filtru, které jsou důležité pro syntézu signálu, parametr, zda je hláska znělá či neznělá a základní kmitočet hlásky. Všechny tyto parametry popisují vlastnosti daného rámce a pak jsou již vysílány komunikačním kanálem. V dekodéru probíhá samozřejmě inverzní proces.



Obr. 2.6: Analyzátor vokodéru

Popis modelu zdroje

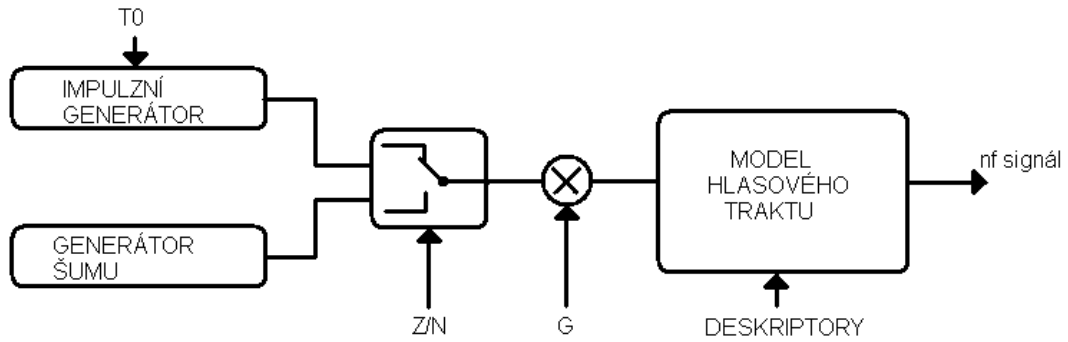
Model zdroje lze popsat například přenosovou funkcí. Popis celku by byl poněkud složitý, proto je vhodné popisovat dílčí části modelu. Model buzení hlasového traktu, model hlasivek, model hlasového traktu a model vyzařování zvuku.

Model buzení hlasového traktu

Model využívá pro znělé zvuky sled impulsů s periodou T_0 a pro neznělé zvuky náhodný šum s plochým spektrem. Buzení je nenulové v časových okamžicích rovných násobku vzorkovací periody T , tj. $T_0 = bT$ Koeficient b značí, kolikrát je perioda základního tónu delší než perioda vzorkování. Odpovídající z-obraz budící funkce za předpokladu normované periody vzorkování je

$$U(z) = G \sum_{n=0}^{\infty} (z^{-b})^n = \frac{G}{1 - z^{-b}} \quad . \quad (2.1)$$

G odpovídá velikosti vybuzení vzhledem k normované jednotkové hodnotě. Pro neznělé úseky hovorového signálu je využit generátor šumu s rovnoměrným spektrem. [9]



Obr. 2.7: Syntezátor vokodéru

Model hlasivek

Reprezentuje rychlost kmitání, hmotnost a pružnost hlasivek, velikost plochy štěrbin y hlasivek a lze jej systémově nahradit dolní propustí druhého stupně. Hraniční kmitočty filtru se pohybuje okolo 100 Hz. Přenosovou funkci lze uvažovat ve tvaru

$$H_G(z) = \frac{1}{[1 - \exp(-cT)z^{-1}]^2} \quad , \quad (2.2)$$

kde c představuje neznámý parametr filtru a T odpovídá periodě vzorkování. [9]

Model hlasového traktu

Respektuje vlastnosti a parametry hrdelní, ústní a nosní dutiny, které jsou u každého mluvčího rozdílné a způsobují tedy odlišnosti i ve výslovnosti stejného zvuku. Lze jej modelovat kaskádou dvoupólových rezonátorů. Rezonanční frekvence odpovídají kmitočtům jednotlivých formantů. Hlasový trakt lze tedy popsat přenosovou funkcí dle vztahu (2.2), která obsahuje pouze póly. Pro přesnější popis by měl model obsahovat i nulové body.

$$H_V(z) = \frac{1}{\prod_{i=1}^K [1 - 2 \exp(-\alpha_i T) \cos(\beta_i T) z^{-1} + \exp(-2\alpha_i T) z^{-2}]} \quad , \quad (2.3)$$

kde K představuje počet formantů. V případě češtiny vystačíme se 3 až 4 formanty. Parametry α_i a β_i lze určit na základě znalosti kmitočtů formantů F_i a šířky rezonanční křivky B_i rezonátoru 2. řádu $F_i = \beta_i/2\pi$ a $B_i = \alpha_i/\pi$. [9]

Model vyzařování zvuku

lze popsat přenosovou funkcí ve tvaru

$$H_L(z) = 1 - z^{-1}. \quad (2.4)$$

Celkový přenos systému charakterizujícího vokální trakt je dán součinem dílčích přenosových funkcí.

$$H(z) = H_G(z)H_V(z)H_L(z) = \quad (2.5)$$

$$= \frac{1 - z^{-1}}{[1 - \exp(-cT)z^{-1}]^2 \prod_{i=1}^K [1 - 2 \exp(-\alpha_i T) \cos(\beta_i T) z^{-1} + \exp(-2\alpha_i T) z^{-2}]} \quad (2.6)$$

2.1.3 Vektorové kvantování

Vektorové kvantování je velmi často používaná metoda ve zpracování řeči pomocí kodeků. Vzorky signálu jsou kvantovány po blocích. Z jednotlivých nekvantovaných vzorků signálu vytvoříme vektor. Dále existuje konečná množina vektorů, která se nazývá kódová kniha. Vektorové kvantování je pak proces, při kterém je vektoru signálu přiřazen vektor z kódové knihy, který je mu nejvíce podobný. Z hlediska přenosu informace nemusíme přenášet do syntetizátoru vektor signálu, ale pouze polohu podobného vektoru v kódové knize. To platí pouze za předpokladu, že na přijímací straně máme k dispozici stejnou kódovou knihu. Na základě příjmu přeneseného indexu přijímačem je vybrána odpovídající hodnota vektoru z kódové knihy.

Vektorový kvantizér musí být navržen tak, aby výsledné výstupní zkreslení bylo minimální při požadované přenosové rychlosti, nebo aby pro stanovenou hodnotu zkreslení bylo dosaženo co nejnížší přenosové rychlosti. Je třeba zvolit optimální strukturu kódové knihy a použít takovou strategii vektorové kvantizace, aby složitost výpočtu bylo možno řešit v reálném čase a realizace byla ekonomicky únosná. Vektorové kvantování je možno použít nejen pro vzorky hovorového signálu, ale i pro koeficienty lineární predikce či další deskriptory při parametrickém kódování řečového signálu.[9] Této vlastnosti se samozřejmě často využívá v různých modifikacích kodeků. Používané kódové knihy se dělí do několika skupin. S plným prohledáváním, s binárním prohledáváním, kaskádní kódová kniha, dělené kódové knihy a adaptivní kódové knihy.

2.1.4 Hybridní kódování zdroje

Hybridní zdrojové kodéry v sobě spojují přednosti obou předchozích typů kódování. Stejně jako u vokodérů se i u hybridních kodérů vytváří výstupní řečový signál jako odezva syntetizujícího filtru. Generování excitačních signálů však není založeno na

dělení znělých a neznělých hlásek a nezavádí se pojem základní kmitočet. U hybridních kodérů se naopak zavádí pojem tzv. multiimpulsní excitace. Multiimpulsní excitace je realizována úzkými impulsy ve velmi krátkém časovém úseku. Amplitudy a časové polohy impulsů jsou určovány tak, aby byla minimalizována perceptuální chyba mezi syntetizovanou a originální řečí. Základní částí hybridního kodéru je časově proměnný syntetizační filtr, který krátkodobě modeluje spektrální obálku průběhu řečového signálu. Tento filtr je označován jako krátkodobý korelační filtr. Do kaskády s ním se však v mnoha případech zařazuje dlouhodobý korelační filtr, který je schopen zjemňovat spektrální obálku řečového signálu. Syntetizační filtr je pak řízen excitačním generátorem a generuje excitační sekvenci. Dohromady tvoří v podstatě dekodér, který je implementován v kodéru na vysílací straně. Excitační generátor je řízen rozdílem mezi originálním a syntetizovaným řečovým signálem, zpracovávaným v blocích vážení a minimalizace chyby. Použitím multiimpulsní excitace vede k vysoké účinnosti tohoto typu kodérů, při zachování velmi dobré kvality reprodukovatého řečového signálu.

Residually Excited Linear Prediction (RELP)

RELP vychází z analýzy segmentovaných úseků hovorového signálu a používá složitější metody kódování zbytkového signálu. Při optimálním určení a nastavení koeficientů predikce v lineárním predikčním filtru vysílací části má zbytkový signál rovnoměrné spektrum. U znělé hlásky je to diskrétní spektrum s periodou základního tónu, u neznělé hlásky pak spektrum spojitě. Na základě tohoto poznatku lze omezit kmitočtové spektrum dolní propustí. Pro zajištění rovnoměrnosti kmitočtového spektra se používá dlouhodobá lineární predikce. Jelikož u RELP kodérů dochází k omezení kmitočtového pásma diferenčního signálu, můžeme provést decimaci vzorkovacího kmitočtu. Na přijímací straně dochází k rozšíření kmitočtového pásma. Obnova vyšších kmitočtů zbytkového signálu se provádí zároveň s interpolací, tj. zvýšením redukovatého vzorkovacího kmitočtu zbytkového signálu na původní hodnoty. Obnoveným zbytkovým signálem se pak budí lineární predikční filtry hlasového syntezátoru. RELP kodek pracující na přenosové rychlosti 9,6 kbit/s poskytuje dobrou telefonní kvalitu.[9]

Multipulse Excited Linear Prediction (MELP)

Kodéry založené na metodě MELP generují budící sekvenci složenou z určitého počtu nestejnoměrně rozložených pulsů. Během analýzy jsou jak amplituda, tak umístění pulsů určeny jedním pulsem v daném okamžiku tak, aby byly vážené kvadratické chyby minimalizovány. MELP algoritmus užívá kolem 5 pulsů každých 5 ms.

Kódování buzení v MELP je náročnější než u klasických lineárních vokodérů, neboť je kódována jak amplituda, tak umístění pulsů. [9]

Regular Pulse Excitation Linear Prediction (RPELP)

RPELP stejně jako MELP využívá budící sekvenci složenou z několika pulsů. Hlavní rozdíl mezi MELP a RPELP spočívá v předem daném rozložení pulsů. U RPELP kodéru je pozice těchto pulsů určena specifikovaným umístěním prvních pulsů uvnitř rámce a mezerami mezi nenulovými pulsy. Posloupnost budících impulsů je periodická a pro daný analyzovaný úsek je vybrána vždy jedna z několika možných vzájemně posunutých budících posloupností. Není nutno počteně stanovovat polohu každého impulsu. Počítají se pouze velikosti impulsů na základě minimalizace vážené střední kvadratické chyby. Počet pulsů u RPELP bývá mezi 10 a 13 na segment délky 5 ms a jejich amplitudy jsou určeny řešením soustavy lineárních rovnic. Mezery mezi pulsy nemusejí být kódovány, jelikož počet pulsů v rámci je pevný.[9]

Code Excited Linear Prediction (CELP)

CELP jsou lineárně predikční kodéry 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 v kódové knize binárně přenesena do přijímače. Přijímač pak podle přijaté adresy generuje příslušnou budící posloupnost pro LP hlasový syntezátor. Ačkoli parametry dlouhodobé predikce LTP mohou být určeny přímým postupem, zpětnovazební přístup vede k mnohem vyšší kvalitě řeči. Kvalita řeči u CELP kodérů může být zlepšena následnou filtrací syntetizované řečové posloupnosti s cílem zdůraznit strukturu řeči z hlediska formantů a základních kmitočtů. Jedna z nevýhod původního CELP algoritmu je velká výpočetní náročnost prohledávání kódových knih. [9]

Low Delay Code Excited Linear Prediction (LD-CELP)

Způsob kódování CELP s krátkou dobou zpoždění je založen na zpětnovazebním řízením adaptace s kódovacím zpožděním do 2 ms a přenosovou rychlostí 16 kbit/s. Po převodu digitalizovaného signálu z osmimístného nelineárního PCM kódu do lineárního vícemístného kódu v převodníku kódu jsou vzorky kódovaného signálu ve vyrovnávací paměti vzorků segmentovány. Délka segmentů je 0,625 ms. Jsou tvořeny vždy pěti po sobě následujícími vzorky kódovaného signálu. Koncepce LD-CELP se od kodeků MPELP, RPELP nebo CELP liší v řešení vlastního hlasového syntezátoru

a způsobu jeho řízení. U jmenovaných kodeků se přenášejí v časovém multiplexu lineárně predikční koeficienty vyčíslené v kodéru krátkodobou a dlouhodobou predikční analýzou, které slouží v přijímači k periodickému nastavování lineárně predikčního filtru hlasového syntezátoru. Naproti tomu mezi kodérem a dekodérem LD-CELP se tato data nepřenášejí a jak kodér, tak dekodér jsou řízeny zpětnovazebně. Po skončení vektorového kvantování daného bloku vzorků je zde vybraný vektor budícího signálu znovu přiveden na vstup hlasového syntezátoru kodéru. Ten při svém průchodu hlasovým syntezátorem uloží vybrané, pro kódování daného bloku optimální, predikční koeficienty v blocích zpětnovazebního řízení úrovně budícího signálu a nastavení koeficientů lineárně predikčního filtru do jejich pamětí. Tím připraví hlasový syntezátor pro kódování následujícího vektoru syntetického signálu. Koeficienty predikčního filtru hlasového syntezátoru a bloku normování úrovně jsou nastavovány zpětnovazebně na základě kvantovaného signálu, respektive normování budícího signálu. Nastavení budící úrovně se provádí pro každý blok vzorků. Koeficienty predikčního filtru se nastavují vždy po osmi vektorech (perioda 5 ms). V LD-CELP dekodéru se koeficienty nastavují obdobně jako u kodéru. Převodník kódu převádí dané složky vektoru syntetického signálu na odpovídající PCM kód.[9]

3 KODÉRY ŘEČOVÉHO SIGNÁLU

3.1 Kvalita přenosu a zpracování řeči

Aby hlas jednoho účastníka hovoru doputoval přes telekomunikační síť k druhému, musí být jistým způsobem zpracován a přenesen. V současné době již nejde o klasický telefonní přenos, ale přenášená informace prochází několika různými druhy sítí. Jaké aspekty mají tedy vliv na kvalitu řeči? V klasických telefonních sítích má na kvalitu přenosu a zpracování řeči vliv hlavně kvalita telefonního přístroje, akustické echo, interference mezi linkami, šum, přeslechy. V mobilních sítích hraje hlavní roli úroveň signálu v radiovém prostředí, protože data přenášená mezi komunikující mobilní a základnovou stanicí jsou velmi náchylná na chyby v komunikačním kanále. Chyby mohou být způsobeny například interferencemi, impulzním rušením atd. Na této komunikační cestě je pak důležité použití opravných kódů. O paketových sítích v oblasti přenosu hlasu se dnes mluví ve všech směrech a zdá se, že není důvod proč danou oblast nadále nerozvíjet. A jak mohou paketové sítě ovlivnit vlastnosti hlasových dat? Jedná se především o ztrátovost paketů, správné pořadí doručení paketů nebo celkové zpoždění doručení paketů. To, co je primárním krokem k dosažení kvalitní budoucí interpretace řečového signálu, je samozřejmě vhodné digitální zpracování. Samotné metody digitálního zpracování jsou uvedeny v kapitole Digitalizace řečového signálu a konkrétní příklady kodérů v textu této kapitoly níže.

Měřicí metody kvality řeči můžeme rozdělit do dvou základních skupin: subjektivní a objektivní metody měření.

Subjektivní metody měření – tato metoda je založena na statistickém vyhodnocování ohodnocených vzorků přeneseného řečového signálu konkrétní skupinou osob. Úkolem osob je vyplnit dotazník (uvedený ve standardu ITU-P.82) a tím se subjektivně vyjádřit k danému testovanému úseku přeneseného řečového signálu. Při tomto průzkumu musí být dodrženo několik základních podmínek podle ITU Rec. P.800. Mezi ně patří specifikace nahrávacího studia, snímacího a záznamového zařízení, prostorové umístění mikrofону a řečníka, použití větrného mikrofonního filtru, množina nahrávek musí obsahovat jak ženskou, tak mužskou řeč. Přesně danými podmínkami je specifikována i poslechová místnost, její minimální objem a kvalita reproduktorů. Účastníci poslechových testů se nesmí orientovat v této problematice, nesmí znát namluvené věty apod. Jelikož jde o statistický průzkum, který je prováděn na rozmanité a početné skupině osob (vybrané podle zvolených kritérií reprezentativního statistického souboru), jedná se o časově a finančně velmi náročný výzkum. Tyto zápory jsou však nahrazeny vahou výsledků testování. Výsledkem testování jsou pak konkrétní hodnoty MOS.

Konverzační testy: jsou určeny k simulaci situací, se kterými se mohou setkat

uživatelé telekomunikačních koncových zařízení. Při použití těchto testů je nutné zvolit vhodné podmínky, testovací množinu subjektů, korektní a přesné vedení testu. Princip testu spočívá v umístění dvou testovacích subjektů do dvou zvukotěsných místností, které jsou blízko kontrolní místnosti. Objem místností nesmí být menší než dvacet metrů krychlových. Hladina šumu v obou místnostech musí být udržována na co nejnižší hladině. Oba testovací subjekty vedou mezi sebou telefonní rozhovor a hodnotí kvalitu přenosu hovorového signálu. Během hovoru však obsluha mění v kontrolní místnosti podmínky testu. Škála pro ohodnocení kvality přeneseného hovorového signálu je v případě konverzačních testů pětistupňová (1 = nedostatečný, 2 = špatný, 3 = přijatelný, 4 = dobrý, 5 = výborný).

Poslechové testy: výsledky poslechových testů nejsou stejně realistické, jako testy konverzační. Omezení jsou proto v určitých ohledech méně přísná. Znamená to, že je nutné sledovat, kontrolovat a specifikovat mnoho parametrů, které jsou při konverzačních testech automaticky v rovnováze. Proces testování je však jednodušší než u testů konverzačních. Princip testu je následující: testovací subjekt je umístěn do zvukotěsné místnosti, která má objem v rozmezí 30 až 120 m³. Šum musí být udržován na nejnižší možné hladině. Testovacímu subjektu jsou do sluchátka přehrávány testované řečové signály a subjekt hodnotí jejich kvalitu pomocí jedné z metod uvedených dále. Poslechových testů pro hodnocení je několik, jejich vyčerpávající popis lze najít v [4].

Objektivní metody měření – u těchto metod se kvalita přeneseného signálu hodnotí statisticky z matematických modelů, které se snaží modelovat sluchový aparát člověka. Pro tyto metody hraje fakt, že nejsou časově ani finančně náročné. Bohužel v tomto případě určování MOS se nejedá o reálnou hodnotu, ale pouze o její odhad, který je získán z několika konkrétních parametrů měření. Zde pak záleží na kvalitě zpracování matematického modelu. Mezi metody objektivního ohodnocení kvality přenosu hovorového signálu telekomunikačními sítěmi patří metoda PESQ (Perceptual Evaluation of Speech Quality). Tato metoda kombinuje techniku časového zarovnání známou z metody PAMS (Perceptual Speech Quality Measurement) s přesným modelováním fyziologie sluchu metody PSQM (Perceptual Speech Quality Measurement). Je určena nejen pro testování kvality přenosu hlasovými vokodéry, lze ji využít i k testování celého okruhu.

Vztah mezi hodnotou PESQ a faktorem MOS: při porovnávání a zpracování hodnot se nevychází přímo z hodnot PESQ, jelikož se v telekomunikacích daleko častěji udávají hodnoty vzešlé ze subjektivních testů, tedy MOS. Hodnoty PESQ leží v intervalu 0,5 až 4,5, zatímco hodnoty MOS pro poslechové metody jsou v intervalu od 1,02 do 4,56. V dokumentu [5] lze najít mapovací funkce mezi těmito hodnotami. [1].

3.2 Vlastnosti kodérů

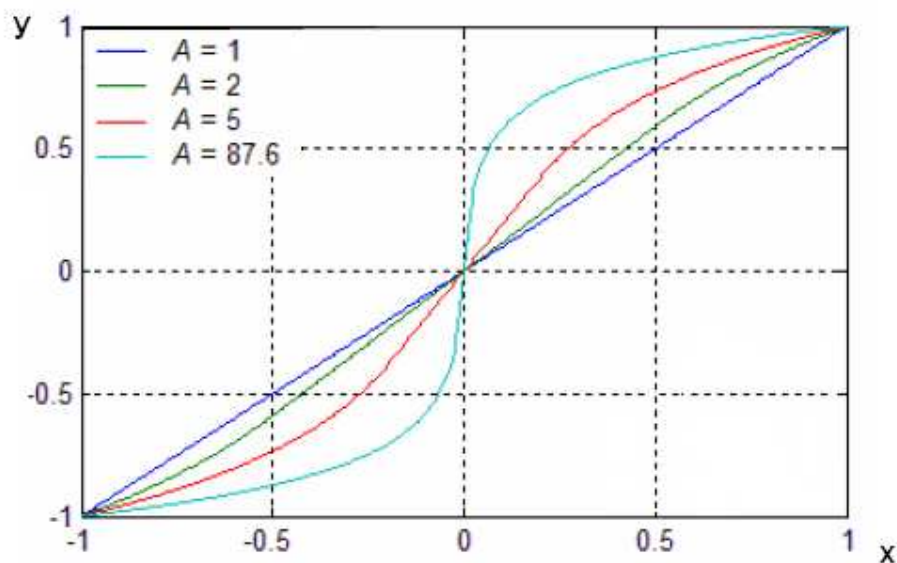
Aby bylo možné porovnávat kodeky mezi sebou, je nutné určit parametry, které mají dostatečnou informační hodnotu. Základním parametrem je samozřejmě výstupní bitová rychlost kodeku. Čím nižší bitová rychlost, tím lépe pro přenosový kanál, ovšem ne vždy pro kvalitu rekonstruovaného řečového signálu. Důležitá je také šířka pásma, kterou je schopen kodér zpracovat. Většinou se kodeky dělí na úzkopásmové (zpracovávají signál do 3,5 kHz) a širokopásmové (do 7 kHz). Dalším sledovaným parametrem je algoritmické zpoždění. Je většinou udáváno v ms a čím nižší zpoždění, tím lépe. Řádově by se mělo zpoždění pohybovat kolem 30 – 40 ms, horní hranice by neměla přesáhnout 100 ms. Důležitým aspektem je také typ použitého kódování, od něj se pak odvíjí veškeré další parametry, včetně výše uvedených. Parametr výpočetní složitost udává náročnost na procesor. U zpracování řečového signálu na PC tento parametr nemusí být směrodatný, ale u mobilních terminálů ovlivňuje přímo spotřebovanou energii a zbytečně zatěžuje baterii přístroje. Obdobné je to s parametrem paměťová náročnost. Všechny dosud uvedené parametry jsou technického rázu. Pro lidské ucho však není důležitá paměťová nebo procesorová náročnost, ale subjektivní pocit z rekonstruovaného signálu. Subjektivně určenou kvalitu lze vyjádřit pomocí některé metody hodnocení kvality signálu, viz výše. Souhrn vlastností konkrétních kodérů je uveden v příloze B.1. Přehled vlastností není zcela kompletní a to především z důvodu nedostupnosti některých informací.

3.3 Kodéry

3.3.1 ITU–G.711

Standard ITU–G.711, který byl zveřejněn v roce 1972, je základním kamenem kódování hlasového signálu. Používá se výhradně v klasických telefonních sítích. Analogový signál je podle tohoto standardu převáděn do digitální podoby pomocí logaritmické PCM a výsledný digitální signál je bez dalších úprav přenášen telefonní sítí. Hodnota dostatečné vzorkovací frekvence je stanovena na 8 kHz, hloubka rozlišení je 8 bitů. Z toho lze jednoduše vypočítat, že výstupní bitová rychlost takto zpracovaného signálu je 64 kbit/s. Norma však definuje dvě kompresní charakteristiky. A-law, obr. 3.1, používanou v Evropě a Austrálii a μ -law, obr. 3.2, používanou v Americe a Japonsku.

Princip komprese spočívá v nerovnoměrném rozložení kvantovacích hladin, jejich četnost klesá s rostoucími hodnotami úrovně signálu. Vychází se totiž z charakteru řečového signálu, kdy se předpokládá menší pravděpodobnost výskytu vysokých

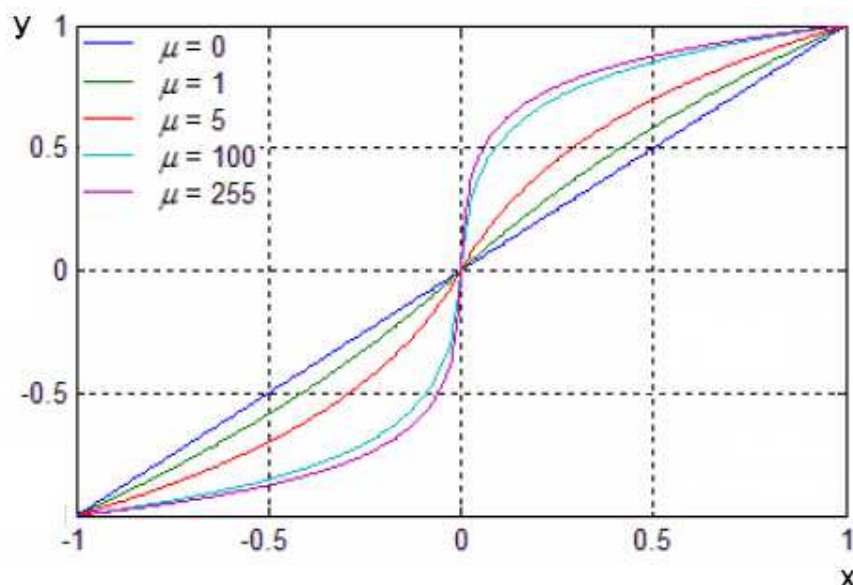


Obr. 3.1: Kompresní charakteristika A-law

hodnot úrovně řečového signálu. Původně 16 bitové vzorky jsou oříznuty na 13 bitové, které jsou pak převedeny danou kompresní charakteristikou na 8 bitové. Rozdíl mezi oběma charakteristikami je v kompresním poměru. A-law převádí na 8 bitové vzorky, kdežto μ -law na vzorky 7 bitové. Tento rozdíl je dán odlišnou strukturou telefonních sítí v Evropě (Austrálii) a Americe (Japonsku). V Evropě je pro přenos hlasu určených právě 8 bitů a signalizace se přenáší ve zvláštním kanále. Základní linka obsahuje 32 kanálů. V Americe je pro přenos hlasu určeno jen 7 bitů, osmý je používán jako signalizační a základní linka obsahuje jen 24 kanálů. Pokud dojde k použití obecně jakékoliv komprese, dá se předpokládat, že na druhé straně přenosového kanálu je nutné použít dekompresi. Na straně příjemce je použita inverzní charakteristika a dochází k expanzi signálu do původního průběhu.

3.3.2 ITU–G.718

Standard ITU–G.718 je v platnosti od 13. června roku 2008 a popisuje úzkopásmový (3,4 kHz) i širokopásmový kodek (7 kHz) pro zpracování řeči a audio signálu. Struktura kodeku je navržena tak, že může být používán v pevných, mobilních, i bezdrátových sítích a to prostřednictvím IP protokolu. V kodeku je implementována funkce škálovatelnosti, což zajišťuje dostatečnou přizpůsobivost v IP při přenosu řečového signálu a zároveň tak umožňuje řídit výstupní bitové rychlosti podle fyzického stavu linky, po které komunikace probíhá. Kodér disponuje rychlostmi 8, 12, 16, 24 a 32 kbit/s. Úzkopásmový signál se vzorkuje frekvencí 8 kHz nebo 16 kHz, široko-



Obr. 3.2: Kompresní charakteristika u-law

pásmový signál je nutné vzorkovat pouze 16 kHz, aby nedocházelo k aliasingu. Pro bitové rychlosti 8 až 12 kbit/s může být použit úzkopásmový mód, jinak je možné použít širokopásmové zpracování pro všechny výstupní bitové rychlosti. Kodek generuje na výstupu rámce o délce 20 ms, jeho algoritmické zpoždění je 43 ms. To je však možné snížit uvedením kodeku do módu nízkého zpoždění na 33 ms, za cenu snížení bitové rychlosti. Výhodou tohoto standardu je, že definuje i zpětnou kompatibilitu se standardem G.722, G.722.1 a G.722.2. Zpětná kompatibilita pak umožňuje kódování a dekódování v módech kodeků dle G.722. Celý kodek vychází z technologie CELP a pro zajištění kvality i při nízkých bitových rychlostech je použita diskretní kosinová transformace (DCT). Na jeho vývoji se podílelo několik celosvětových firem z oblasti telekomunikací a elektrotechniky, například Nokia, Motorola, Texas Instruments, Ericsson.[13]

3.3.3 ITU–G.722

G.722 je standard, zveřejněný v roce 1988, který je založen na zpracování řečového signálu pomocí SB-ADPCM (Sub Band ADPCM). G.722 můžeme zařadit do skupiny širokopásmových kodeků, protože šířka pásma, kterou zpracovává, je 7 kHz. Používá vzorkovací frekvenci 16 kHz, která je dvakrát vyšší než vzorkovací frekvence u klasického G.711. Z toho vyplývá, že díky rozdělení signálu do dvou pásem pomocí zrcadlového filtru, aplikaci ADPCM na každé pásmo zvlášť a vyšší vzorkovací frekvenci, se dosahuje slušných výsledků kvality obnoveného řečového signálu na při-

jímající straně. A to jak po stránce zvukové – věrnosti hlasu, tak i po stránce srozumitelnosti. Hodnoty bitových rychlostí, které kodér podporuje jsou 48, 56 a 64 kbit/s. Takovéto bitové rychlosti jsou samozřejmě pro účely například mobilních sítí naprosto nepoužitelné, avšak vyhovující jsou pro aplikace v technologiích Voice over IP (VoIP). Je zcela jasné, že pokud bude hovor zakódovaný pomocí G.722, zůstane jeho vysoká kvalita zachována pouze při přenosu v datových sítích. Jakmile by byl hovor směrován do klasické telefonní sítě, byl by překódován na klasické PCM a tím pádem znehodnocen.

Další modifikace G.722 jsou označeny ITU–G.722.1 (Siren7) a ITU–G.722.2.

G.722.1 je standard definující zpracování zvuku včetně řečového signálu, zveřejněný roku 1999. Řadí se také mezi širokopásmové kodeky a stejně jako základní kodér používá vzorkovací frekvenci 16 kHz. U tohoto kodéru jsou definované dvě rychlosti, 24 a 32 kbit/s. G.722.1 je v podstatě standardizovaný kodek PT716 od firmy Polycom, která je předním výrobcem videokonferenčních systémů. Jeho nástupcem je pak PT716plus a Siren14, který umožňuje zpracování šířky pásma FM rozhlasu, tj. 14 kHz, při zachování bitových rychlostí. Zpracování dvojnásobné šířky pásma bylo přidáno do standardu teprve v roce 2005 jako příloha C.

G.722.2 nebo-li také AMR–WB, je kodek používající technologii ACELP standardizovaný na přelomu let 2001 - 2002. Tato modifikace patří opět do skupiny širokopásmových kodeků a může se pyšnit hned několika různými bitovými rychlostmi: 23,85, 23,05, 19,85, 18,25, 15,85, 14,25, 12,65, 8,85 a 6,6 kbit/s při vzorkovací frekvenci 16 kHz. Tři nejnižší bitové rychlosti jsou primárně určené pro přenos hlasu v mobilních sítích třetí generace, kompatibilní je však i s druhou (dvouapůltou) generací mobilních sítí. Základní rychlostí běžně používanou v těchto sítích je 12,65 kbit/s, nižší rychlosti se používají v případech, kdy je komunikační kanál zúžený. Bitové rychlosti vyšší než 14,25 kbit/s jsou vhodné pro přenos hudby nebo hlasu při videokonferencích v dostatečně širokém komunikačním kanálu. Avšak jejich použití není vyloučeno v mobilních sítích pro multimediální zprávy a streamování audia.[14]

3.3.4 ITU–G.723.1

Doporučení G.723.1 bylo schváleno 2006 a popisuje kodek pro multimediální komunikaci s výstupními bitovými toky 5,3 a 6,3 kbit/s. Oba bitové toky jsou povinnou částí kodéru a dekodéru. G.723.1 umožňuje přepínání mezi oběma bitovými toky při každé změně rámce. Kodek byl především používán při videokonferencích založených na dnes již zastaralém protokolu H.324, který používal pro komunikaci běžnou telefonní síť a modem s rychlostí nejméně 28,8 kbit/s. G.723 specifikuje kódování použité pro kompresi řeči a dalších audiosignálů v multimediálních službách

na velmi nízké bitové rychlosti. Kodér je optimalizován pro reprezentaci řeči, jiné audiosignály je sice možné kodekem zakódovat, ale jejich reprezentace není zdaleka dostačující. Kodér používá lineární predikci, pro vyšší bitový tok se používá metoda MP–MLQ (Multipulse Maximum Likelihood Quantization) a pro nižší bitový tok metoda ACELP. Velikost rámců byla stanovena na 30 ms plus rezerva pro budoucí použití 7,5 ms. Z toho je možné odvodit, že algoritmické zpoždění je 37,5 ms. Kodér pracuje s 16-ti bitovými vzorky navzorkovanými frekvencí 8 kHz. Kódovací metoda je založena na lineární predikci typu analýza–syntéza a chybový signál je minimalizován na základě perceptuálního (smyslového) vnímání člověka. Vstupní vzorky jsou rozděleny do skupin po 240 vzorcích, což odpovídá času 30 ms. Každý rámeček je filtrován horní propustí pro odstranění stejnosměrné složky a dále rozdělen do 4 subrámců s délkou 60 vzorků. Pro každý subrámeček je provedena predikční analýza a jsou spočteny koeficienty LPC filtru. Tyto koeficienty jsou určeny z autokorelačních koeficientů pomocí Levinson–Durbinova algoritmu². LPC koeficienty slouží pro krátkodobý perceptuální váhovací filtr, kterým filtrujeme každý rámeček, abychom získali perceptuálně váhovaný řečový signál. Následuje převod LPC koeficientů na LSP koeficienty (Line Spectral Pairs), které se kvantují pomocí 8 bitů a spolu s některými dalšími parametry se odešlou komunikačním kanálem. Na přijímací straně se samozřejmě postupuje opačným způsobem.[11]

3.3.5 ITU–G.726

Standard G.726 vznikl v roce 1990 a v roce 2005 byl obnoven sloučením dvou starších standardů – G.721, který používal bitovou rychlost 32 kbit/s a G.723, který používal 24 a 40 kbit/s. Nově byla přidána bitová rychlost 16 kbit/s. G.726 pracuje stejně jako oba předchozí s metodou ADPCM a vzorkovací frekvencí 8 kHz.

3.3.6 ITU–G.727

G.727 je standardizovaný kodek s využitím Embedded ADPCM, volně přeloženo kodek s vloženým ADPCM algoritmem. Definuje převod signálu z nelineární PCM o konstantním datovém toku 64 kbit/s na ADPCM s proměnlivým bitovým tokem. K tomu jsou používány 2, 3, 4 nebo 5 bitové lineární kvantizátory. Počet bitů použitých pro vyjádření rozdílu je dán použitou bitovou rychlostí (16, 24, 32 nebo 40 kbit). Vložené ADPCM algoritmy uvedené v tomto standardu jsou rozšířením ADPCM algoritmů definovaných v G.726. Tyto algoritmy jsou určeny pro používání v systémech s paketizací řeči a systémy jsou definovány ve standardu G.764. Protokol paketizace řeči je schopen pracovat s datovými kodéry, které mohou adaptivně

²www.comtel.cz/files/download.php?id=3370

přizpůsobovat bitovou rychlost podle fyzického stavu přenosového média. Standard umožňuje, aby byla snížena bitová rychlost kdekoli v síti bez nutnosti spolupráce vysílače a přijímače. G.727 nepatří do skupiny ADPCM algoritmů G.726 a proto před jeho použitím musí být zajištěno odeslání informací mezi kodérem a dekodérem. Nastavení ADPCM kodéru se řídí pomocí (x,y) párů. Proměnná x se odkazuje na hodnotu parametru Feed Forward, y na hodnotu parametru Feed Back. FF využívá nadstavbové i základní bity, FB pouze základní. Nadstavbové a základní bity jsou v podstatě parametry výstupního signálu. Kombinacemi (x,y) vznikají číselné páry, které definují chování kodéru. Standard pak poskytuje kódové rychlosti 40, 32, 24 a 16 kbit/s a základní rychlosti 32, 24 a 16 kbit/s, čemu odpovídají následující číselné páry $(5,2)$, $(4,2)$, $(3,2)$, $(2,2)$; $(5,3)$, $(4,3)$, $(3,3)$; $(5,4)$, $(4,4)$. V kodéru dochází po konverzi signálu z PCM k odečtení vzorku signálu a vzorku ze zpětnovazební smyčky. Adaptivní kvantizér pak používá 2, 3, 4 nebo 5 bitová čísla k vyjádření rozdílu vzorků signálu, které jsou pak přenášeny do dekodéru. Při přenosu mají základní bity garantované doručení (při bezchybném přenosu), bity nadstavbové mohou být ořezány z důvodu potřebného snížení bitové rychlosti. Kodek je možné používat v širokopásmové IP telefonii, audio–video konferencích a například ve wifi telefonech.[12]

3.3.7 ITU–G.728

Standard G.728 byl schválen 1. září 1992, doplnění bylo provedeno v roce 2006 a popisuje kodek pro kódování řeči za použití metody kódování s krátkou dobou zpoždění LD–CELP. Metoda je založena na zpětnovazebním řízení adaptace s malým kódovacím zpožděním a přenosovou rychlostí 16 kbit/s. Toto doporučení používá systém s lineárním prediktivním kódováním, jenž spočívá v analýze hovorového signálu, ze kterého se odvozují charakteristické parametry hlasového signálu. Algoritmus LD–CELP používá zpětné přizpůsobení prediktoru a zisku a dosahuje zpoždění algoritmu 0,625 ms. Je přenášén pouze index z kódové knihy. Koeficienty prediktoru jsou aktualizovány během predikční analýzy z předchozích kvantovaných vzorků řeči. Váhový filtr je aktualizován užitím predikční analýzy nekvantizované řeči. Po konverzi z A-law nebo μ -law PCM do lineární PCM je vstupní signál rozdělen do vektorů po 5 vzorcích. Pro každý vstupní blok kodér prochází 1024 možných kódových vektorů, uložených v kódové knize. Vektory pak prochází přes řízený zesilovač a vstupují do syntetického filtru, který krátkodobě modeluje spektrální obálku průběhu hovorového signálu. Vektor z kódové knihy se nazývá budícím vektorem. Kódová kniha spolu se syntetickým filtrem tvoří vlastně dekodér, který je koncipován téměř stejně jako na straně dekodéru. Velikost excitačního signálu je řízena rozdílovým signálem mezi originální a syntetizovanou řečí, jenž je zpracováván v blocích vážení a mi-

nimalizace chyby. Optimalizační postup se zde realizuje v uzavřené smyčce, která nutí syntetický filtr generovat řečový signál, jež se co nejméně liší od signálu originálního. To potom vede k velké efektivitě kodérů tohoto typu, které potom při relativně malých bitových rychlostech poskytují velmi dobrou kvalitu reprodukováné řeči. Z výsledných 1024 možných kvantizovaných vektorů signálu vybere kodér ten, který má minimální frekvenčně váhovanou střední kvadratickou odchylku měřenou s ohledem na vektor vstupního signálu. Desetibitový index z kódové knihy odpovídající nejlepšímu vektoru z kódové knihy je přenesen do dekodéru.[11]

3.3.8 ITU–G.729

G.729 bylo schváleno 19. března 1996 a popisuje kódování řeči s výstupním bitovým tokem 8 kbit/s použitím metody CS-ACELP (Conjugate Structure ACELP). CS-ACELP kodér pracuje se vzorkovací frekvencí 8 kHz, 16 bitovým rozlišením a rámci velikosti 10 ms. Každý 10 ms rámec řečového signálu je analyzován a po získání parametrů modelu (koeficienty LPC, indexy slov z pevné a adaptivní kódové knihy, zisk) jsou tyto parametry kódovány a přenášeny komunikačním kanálem.

V dekodéru jsou tyto parametry použity k obnovení budícího signálu a koeficientů syntetického filtru. Řečový signál je pak rekonstruován filtrací excitačního vektoru přes syntetický filtr. Kvalita řečového signálu je pak ještě zvýšena postfiltrem. Velkou nevýhodou je stejně jako u G.728 velký výpočetní výkon potřebný k prohledání celé kódové knihy a nalezení nejlepšího budícího vektoru.

Kodér je navržen pro práci se vstupním signálem, který byl získán z analogového signálu navzorkováním frekvencí 8 kHz s kvantizací pomocí 16-ti bitů. Vstupní signál je filtrován horní propustí s $f_m = 140$ Hz a normován v bloku předzpracování signálu. Vhodné je samozřejmě omezení spodní části spektra horní propustí, protože řeč je srozumitelná i bez těchto frekvenčních složek, navíc se v tomto pásmu často vyskytují různé druhy rušení. Například dýchání mluvčích či rušení z elektrické sítě o frekvenci 50 Hz. Výstupní signál tohoto bloku slouží jako vstupní signál pro všechny další výpočty.

Nejprve kodér vypočítá autokorelační koeficienty, z nichž se pomocí Levinson–Durbinova algoritmu [10] vypočtou koeficienty LPC filtru pro každý rámec. Tyto koeficienty jsou transformovány na spektrální páry (Linear Spectral Pairs), které mají vyšší kódovací účinnost. Ty jsou dále kvantovány použitím prediktivní dvoufázové vektorové kvantizace použitím 18 bitů. V jednom jsou tedy přenášeny koeficienty LSP, paritní bit, index pevné a adaptivní kódové knihy, znaménka pevné kódové knihy a úroveň signálu. Přesné bitové rozložení těchto parametrů ve výsledném bitovém toku specifikuje doporučení G.729. Přenosovým kanálem je přenesen ten index z kódové knihy, pro který má originál a rekonstruovaná řeč minimální od-

chylku. Toho se dosáhne filtrováním chybového signálu ve váhovacím perceptuálním filtru, jehož koeficienty jsou odvozeny z nekvantovaných koeficientů LPC.[11]

3.3.9 GSM

Při počátečním výběru typu zdrojového kódování pro použití v sítích GSM bylo předloženo dvacet návrhů z devíti zemí Evropy. Z nich však byly vybrány čtyři, které nejlépe vyhověly požadavkům na kvalitu výsledného zakódovaného hovorového signálu. Testování kódovacích algoritmů bylo prováděno na několika jazycích. Důraz byl kladen na jejich rozdílný charakter, proto se testy prováděly pro těchto sedm jazyků: angličtina, němčina, francouzština, finština, japonština, španělština a hindština. Testy probíhaly při různých hodnotách úrovně signálu, bitové chybovosti a hladiny šumu. Kvalita reprodukováného signálu byla samozřejmě hodnocena pomocí MOS.[8]

- RPE–LPC, MOS = 3,54, bitová rychlost je 14,77 kbit/s, počet matematických operací je 1,5 mil/s – Philips (Německo)
- MPE–LTP, MOS = 3,27, bitová rychlost je 13,20 kbit/s, počet matematickým operací je 4,9 mil/s – IBM (Francie)
- SBC–ADPCM, MOS = 3,14, bitová rychlost je 13,00 kbit/s, počet matematických operací je 4,9 mil/s – Ellemtel (Švédsko)
- SBC–ADPCM, MOS = 2,92, bitová rychlost je 15,00 kbit/s, počet matematických operací je 4,9 mil/s – Telecom Research (Velká Británie)

Výsledkem byla nakonec kombinace řešení od firmy Philips a IBM. Vznikl tak kodek s označením RPE–LTP (Regular Pulse Excitation Long Term Prediction). Obvody kodéru a dekodéru jsou realizovány signálovými procesory a lze je podle funkce rozdělit na tyto hlavní bloky:

- předzpracování signálu
- analýza LPC
- krátkodobá analýza a filtrace
- kódování RPE a analýza LTP

GSM Full Rate

GSM-FR je specifikován ve standardu ETSI 06.10, který byl zveřejněn na začátku devadesátých let a je založen na RPE-LTP. Ve své době představoval velmi dobrý kompromis mezi výpočetní složitostí a kvalitou reprodukováného řečového signálu.

V systému GSM je počáteční zpracování podobné jako v ostatních telekomunikačních systémech. Řečový signál je vzorkován frekvencí 8 kHz, hodnota každého vzorku je vyjádřena 13 bity. Výsledný PCM bitový tok je tedy 104 kbit/s. Tento digitalizovaný signál již přichází na vstup kodeku RPE-LTP. V prvním bloku dochází k předzpracování signálu. Dochází zde k odstranění stejnosměrné složky a k preemfázi, což je jednoduše řečeno zvýraznění vyšších kmitočtů. Lineární prediktor je v podstatě adaptivní FIR filtr, který mění své koeficienty v čase, podle procházejícího signálu, a predikuje hodnoty následující. Do prediktoru vstupuje 160 vzorků signálu, což je jeden segment signálu. Chybový signál, který dostáváme na výstupu, je rozdíl predikované a skutečné hodnoty signálu. Chybový signál by neměl mít žádnou závislost na signálu vstupním a jeho charakter by se měl blížit bílému šumu. LPC prediktor je někdy označován jako bělicí filtr. Pokud by byl prediktor ideální a chybový signál měl skutečně charakter bílého šumu, stačilo by přenášet na stranu dekodéru pouze koeficienty filtru. Dekodér by pak generoval bílý šum, který by procházel syntetizujícím filtrem a generoval by řečový signál. To je bohužel jen ideální případ, proto je v praxi nutné přenášet i chybový signál. Výstupem tohoto bloku je tedy 160 vzorků signálu a 13 koeficientů prediktoru, které jsou vyjádřeny pomocí 36 bitů. Dalším krokem je již LTP analýza a RPE kódování. V prvním kroku je 160 vzorků signálu rozděleno paralelně na čtyři bloky po čtyřiceti vzorcích. Paralelně znamená, že první blok obsahuje 1., 5., 9. a 157. vzorek, druhý 2., 6., 10. a 158. vzorek atd. K dalšímu zpracování je vybrán blok, který obsahuje signál s nejvyšší úrovní. Vzorky excitačního signálu jsou pak vyjádřeny pomocí 47 bitů. LTP analýza a filtrace poté vypočítá dalších 9 bitů. Jedná se o koeficienty filtru, jenž predikují průběh celého chybového signálu, tedy nejen vybraného bloku, ale i ostatních tří. Protože byl vybrán ke zpracování pouze jeden blok, je nutné ho čtyřikrát zopakovat, aby byl zachován segment o délce 20 ms. Výsledkem je tedy 36 bitů LPC analýzy, 188 bitů excitačního signálu, 36 bitů LTP analýzy. Dohromady pak 260 bitů. Jinými slovy bitová rychlost 13 kbit/s.[3]

GSM Half Rate

GSM-HR je specifikován ve standardu ETSI 06.20, který na začátku devadesátých let následoval těsně po ETSI 06.10. Kodek používá algoritmus VSELP (Vector sum excited linear prediction), který patří do rodiny algoritmů CELP. Má poloviční bitovou rychlost než GSM-FR, to znamená 6,5 kbit/s. Bohužel kvalita hlasu je poněkud

nízká. V podstatě je možné 8 timeslotů využít pro 16 hovorů. Mobilní operátoři vsadili spíše na kvalitu hlasu, proto se kodek takřka nepoužívá.

GSM Enhanced Full Rate

Tento kodek je schopen zajistit vyšší kvalitu řeči při zachování bitové rychlosti GSM–FR. To ovšem platí pouze v místech s dobrým signálem. Pokud nemá signál dostatečnou úroveň, má přenosový kanál zvýšenou chybovost a tím pádem se GSM–EFR chová stejně jako GSM–FR. Vyšší kvalita reprodukováného signálu je dána tím, že kodek používá metodu ACELP.

GSM Adaptive Multi Rate

GSM–AMR je technologie, která v dnešní době představuje špičku v přenosu řečového signálu v sítích GSM. Je definována ve standardu ETSI 06.90. Základní myšlenkou je implementace proměnlivého datového toku do EFR. Důvodem adaptivní změny bitových rychlostí je možnost značné úspory přenosové kapacity. Jde o to, že digitálnímu systému je ve své podstatě jedno jestli přenáší ticho nebo hlas. Proto je na místě v době, kdy není co vysílat, snížit výstupní bitový tok nebo vyslat pouze informaci o tom, že není co vysílat. AMR může používat osm bitových rychlostí 12,2, 10,2, 7,95, 7,40, 6,70, 5,90, 5,15 a 4,75 kbit/s. Mezi těmito rychlostmi může kóder přepínat každých 20 ms. Tato doba totiž odpovídá délce jednoho rámce. Používá technologii Multi Rate ACELP. AMR je zpětně kompatibilní s EFR pouze při bitové rychlosti 12,2 kbit/s. AMR zavádí také dvě nové funkce, které přispívají k výborné úrovni tohoto kodeku. První funkcí je tzv. DTX – Discontinuous Transmission, která vypíná vysílání signálu, pokud uživatel mobilní stanice nehovoří. Pokud bychom se nad tím zamysleli, je očividné, že při telefonním hovoru je minimálně v polovině času přenášeno ticho. Tím se šetří energie v baterii mobilního terminálu, kapacita sítě a možnost vzniku interferencí. Funkce by měla být efektivní na delším časovém úseku. Druhou funkcí je VAD – Voice Activity Detection, která je schopná určit, který z rámců obsahuje řečový signál a který ne. VAD je oproti DTX efektivnější na krátkých časových úsecích. Pokud se nevysílá signál, mělo by být ve sluchátku ticho. To však působí velmi nepřírozeně a proto funkce CNI – Comfort Noise Insertion posílá do sluchátka přirozený šum pozadí. Ten ale není nutné přenášet. S kodekem AMR se počítá i v sítích UMTS, takže se dá předpokládat, že jeho vývoj bude dále pokračovat.[7]

3.3.10 VMR-WB

VMR-WB (Variable Multi Rate Wide Band) je řečový kodek, který je používán pro širokopásmové telefony a multimediální streaming. Byl standardizován v červenci 2004. Primárně je nasazen v cdma sítích (cdmaOne nebo cdma2000). Byl vyvíjen firmami Nokia a VoiceAge a jeho základem je doporučení ITU-G.722.2, AMR-WB. Dokáže zpracovat širokopásmové (7 kHz) a úzkopásmové (3,4 kHz) signály, pracuje s velikostí rámců 20 ms. VMR-WB může pracovat ve čtyřech režimech. Premium, Standard, Economy a interoperabilní s AMR-WB. Těmto módům odpovídají průměrné bitové rychlosti 12,8, 10,5, 8,1 a 13,3 kbit/s. Módy 0, 1 a 2 jsou specifické právě pro cdma systémy. Síťově řízená bitová rychlost je důležitá pro optimální provoz cdma systémů. Cílem řízení VBR je zajistit co nejlepší kvalitu hovoru při průměrné bitové rychlosti přenosu. Poskytuje tak dobrý kompromis mezi kvalitou hovoru a kapacitou sítě.

3.3.11 ISAC, ILBC

Jedním z kodeků, které používají VoIP softwarové aplikace pro zpracování hlasu, je Internet Speech Audio Codec. ISAC pochází z dílny firmy Global IP Solutions. Kodek je širokopásmový a signály zpracovává vzorkovací frekvencí 16 kHz. Disponuje variabilní přenosovou rychlostí v rozmezí od 10 kbit/s do 32 kbit/s. Na výstupu generuje podle bitové rychlosti rámce o velikost 30 ms až 60 ms a jeho algoritmické zpoždění je dáno velikostí generovaného rámce plus maximálně 3 ms. Tento kodek je licencován, takže případná implementace je zpoplatněna. Druhý kodek, často používaný ve VoIP aplikacích, je Internet Low Bit Rate Codec. Tento kodek je také od firmy Global IP Solutions. Pracuje se vzorkovací frekvencí 8 kHz a 16 bitovým rozlišením. Má dvě pevně stanovené rychlosti 15,2 kbit/s a 13,33 kbit/s, kterým odpovídá velikost rámců 20 ms a 30 ms. Tento kodek není licencován, avšak komerční využití jeho zdrojového kódu je zpoplatněno. Dva výše uvedené kodeky jsou používány ve Skypu, Google talk nebo Yahoo! Messenger.

3.3.12 TETRA

Tento kodek je používán v satelitních systémech poskytujících telekomunikační a datové služby po celé planetě. TETRA kodek je založen na několika základních blocích, stejně jako většina řečových kodeků. Základním blokem je lineární predikce, která probíhá nad rámcí dlouhými 30 ms. Koeficienty LPC jsou pak převáděny na lineární spektrální páry kvůli aplikaci kvantování. Koeficienty jsou kvantovány a vyjádřeny pomocí 26 bitů. Kodek používá ACELP technologii, která umožňuje lepší

přístupnost, indexaci a robustnost v kódové knize. Budící vektor, který je v knize vyhledáván, je samozřejmě dán rozdílem originálního a predikovaného signálu. Každý 30 ms rámec, který kodek vyprodukuje, je vyjádřen pomocí 137 bitů, což ve výsledku dává výstupní bitový tok 4,5 kbit/s.

3.3.13 AMBE

Advanced Multi Band Excitation je kodek vyvíjený firmou Digital Voice Systems. Pracuje s výstupními bitovými rychlostmi v rozmezí 2 až 9,6 kbit/s, vzorkovací frekvencí 8 kHz a šířkou pásma do 3 kHz. Délka zpraovávaných rámců je 20 ms. I tento kodek pracuje s technologií kódových knih. Jeho nástupcem je Improved MBE. Oba tyto kodeky jsou využívány v satelitních komunikačních systémech. Nejznámějšími satelitními systémy jsou pravděpodobně Inmarsat a Iridium. Kodeky jsou chráněny tvrdou licenční politikou a poplatky za jejich použití se pohybují v řádech desetitisíců dolarů.

3.3.14 Speex

Speex je kodek, který je určen pro VoIP a streaming audio signálu. Není omezen žádnou licencí a patří do multimediálního kontejneru ogg. Speex je založen na kompresním postupu CELP, disponuje širokým spektrem výstupních bitových rychlostí od 2 kbit/s do 44 kbit/s, používá tři vzorkovací frekvence podle módu. Frekvence 8, 16 a 32 kHz odpovídají módům úzkopásmový, širokopásmový a ultra-šírokopásmový. Speex má i několik důležitých funkcí. Variable Bit Rate je funkce, která umožňuje měnit výstupní bitový tok, podle charakteru vstupního signálu popřípadě podle aktuálního fyzického stavu linky. Stejně jako u některých předchozích kodeků se i ve Speex používá funkce VAD, DTX a funkce CNI. Funkce zvyšují kvalitu a komfort hovoru. Algoritmické zpoždění se pohybuje podle aktuálního módu od 30 ms do 40 ms.

3.3.15 DSS

Digital Speech Standard je řečový kodek vyvinutý firmami Olympus, Philips a Grundig v roce 1994, standardizovaný v roce 1997. Jeho primární využití bylo pro digitální nahrávací systémy (diktafony). Výsledný soubor může být opatřen hlavičkou, včetně jména autora nahrávky, mluvčího a jinými údaji. Je to v podstatě adekvátní varianta MP3 pro řeč.

3.3.16 HVXC

Harmonic Vector Excitation Coding. Kodek standardizovaný v normě MPEG-4, který pracuje s LPC a kódovou knihou. Je využíván v DRM rozhlasovém vysílání.

4 MODELOVÁNÍ V PROGRAMU MATLAB SIMULINK

V simulačním programu Matlab Simulink byly vytvořeny modely některých základních metod používaných ke zpracování hlasových signálů. Tyto základní metody jsou popsány v kapitole Kódování v časové oblasti. K simulacím byly použity dva druhy signálů. Jedním byl signál sinusového průběhu a druhým řečový signál v souboru formátu .wav, mezi kterými je možné v modelu ručně přepínat. Čas simulace je nastaven na 0,01 s a to kvůli zobrazení průběhu vzorkování signálu. V programu byl také vytvořen model parametrického zpracování řečového signálu. Jedná se o model LPC vokodéru, který je základem ve většině moderních vokodérů, které jsou používány v běžně dostupných komerčních zařízeních.

Delta modulátor

Delta modulace je jednoduchá a při správném nastavení i relativně účinná metoda zdrojového zpracování signálu v časové oblasti. Vstupní signál $x(t)$ je porovnáván s pomocným signálem $s(t)$ v komparátoru a výstupní signál $e(t)$ je určen podle následujícího pravidla: $e(t) = 1$ platí pro $s(t) < x(t)$, $e(t) = -1$ platí pro $s(t) \geq x(t)$. Výstupní signál je vzorkován tak, že výsledkem jsou úzké pulzy o výšce 1 nebo -1. Dále se pak signál vede do integrátoru, kde se vytváří schodovitá funkce podle vztahu

$$s(t) = \frac{\Delta s}{T} \int_0^t y(t) dt \quad . \quad (4.1)$$

Δs je v tomto případě velikost schodu. Problémem je u této metody přetížení kodéru, ke kterému dochází v případě, že přírůstek vstupního signálu za jednu vzorkovací periodu je větší než velikost schodu. Prakticky to znamená, že vzorkovací perioda musí být u delta modulace až dvanáctkrát kratší než u klasické PCM. Konkrétně v tomto modelu je nastavena vzorkovací frekvence na 48 kHz. Vstupní signál je omezen filtrem s mezní frekvencí 4 kHz. Amplituda sinusového signálu byla zvolena tak, aby korespondovala s úrovní signálu ve .wav souboru.

Pulzní kódová modulace

PCM je nejjednodušší základní metodou, která je používána ve zpracování signálů. Model se skládá ze vzorkovacího bloku, bloku generování vzorkovacího signálu a kvantizéru. Stejně jako u delta modulace je signál omezen dolní propustí s mezním kmitočtem 4 kHz. V tomto případě je vzorkovací frekvence je nastavena na 8 kHz, což je zcela dostatečná hodnota pro vzorkování řečového signálu.

Diferenční pulzní kódová modulace

Model je nastaven stejně jako PCM. V DPCM modelu je použita pro jednoduchost pouze „predikce“ prvního řádu. V praxi se však používá predikce vyšších řádů.

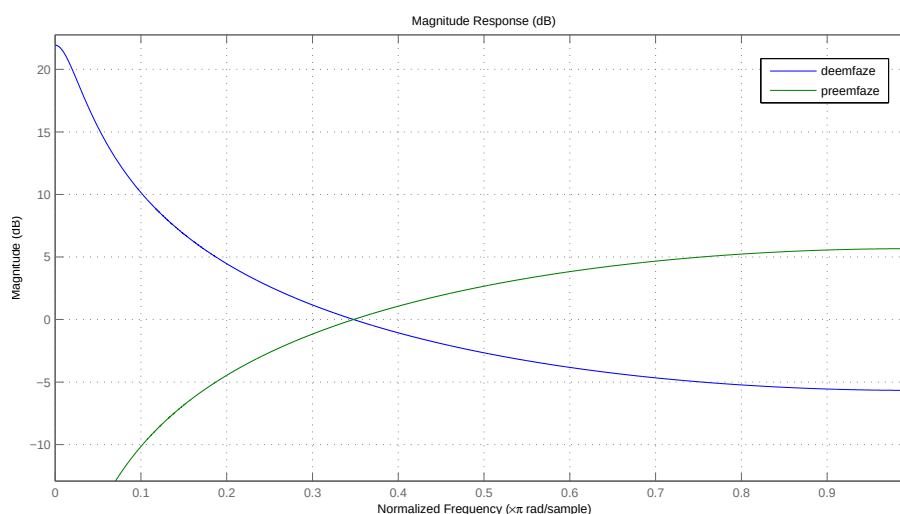
Adaptivní diferenční pulzní kódová modulace

V modelu byl použit blok ADPCM, který lze najít v knihovnách Matlab Simulink. Konkrétně v knihovně dspwvfmcd.lib. Jeho nastavení bylo provedeno na základě testování konkrétních hodnot.

LPC vokodér

Při realizaci LPC vokodéru bylo nutné modelovat několik základních algoritmů, jejichž výsledkem jsou požadované parametry řečového signálu. Na straně příjemce tyto parametry slouží k syntéze řečového signálu. Základními parametry, které je nutno určit, jsou znělost či neznělost segmentu řeči, frekvence základního tónu, energie segmentu a koeficienty syntetizujícího filtru. Řečový signál je nejprve rozdělen do segmentů, které jsou 10 až 30 ms dlouhé, při vzorkovací frekvenci 8 kHz to znamená segmenty o velikost 80 až 240 vzorků. Po segmentaci jsou se segmentem signálu prováděny následující operace.

Preemfáze – jedná se o filtr typu FIR a jeho koeficienty se nastavují na hodnoty $a = 1$, $b = 1 - \alpha$. Hodnoty α se pohybují v rozmezí 0,9 - 0,95. Filtr má za úkol zvýraznit vyšší kmitočty signálu. Inverzní filtr k preemfázi je použit po obnově signálu na přijímací straně, operace filtrace je označována jako deemfáze. Charakteristiky obou filtrů jsou uvedeny na obrázku 4.1.



Obr. 4.1: Frekvenční charakteristika filtrů emfáze

LPC analýza – lineární predikční analýza je základním blokem parametrického zpracování řečového signálu. Model hlasového traktu, který se snažíme při parametrizaci popsat, lze vyjádřit například následující rovnicí:

$$H_z = \frac{c_0}{d_0 + d_1 z^{-1} + d_2 z^{-2} + d_3 z^{-3} + \dots + d_N z^{-N}} = \frac{G}{1 + \sum_{i=1}^{N-1} a_p[i] z^{-1}} \quad (4.2)$$

Koeficienty $a_p[i]$ jsou výsledkem lineární predikční analýzy. Tyto koeficienty jsou pak použity v syntetizujícím filtru k syntéze signálu. Přesnější popis lineární predikční analýzy je možné najít v [10].

Určení F_0 – frekvence základního tónu řeči je základním parametrem řečového signálu v kmitočtové oblasti. Jeho úroveň není konstantní. Průběh frekvence základního tónu charakterizuje melodii řeči. Při řeči se hodnota frekvence pohybuje v rozmezí jedné oktávy (u mužů přibližně mezi 80 - 160 s průměrem 130 Hz, u žen přibližně mezi 150 - 300 s průměrem 220 Hz a u dětí v rozmezí 200 - 600 Hz). Pro detekci frekvence základního tónu byla zvolena metoda centrálního klipování. Metoda pracuje následovně:

- pro jednotlivé segmenty signálu je vypočítán práh, protože úroveň signálu kolísá a tudíž nejde určit pevnou hodnotu prahu po dobu trvání celého signálu. Hodnota i-tého prahu se vypočítá podle následujícího vzorce

$$P_i = k \cdot \min(MAX_{i-1}, MAX_{i+1}), \quad (4.3)$$

kde $MAX_{i\pm 1}$ je maximum sousedních segmentů, k je redukční faktor, jehož obvyklá hodnota je 0,8.

- získaný signál je po prahování znormalizován na jednotkovou velikost, vznikne tím signál nabývající pouze tří hodnot 1, 0 a -1.
- ze signálu s úrovněmi 1, 0 a -1 se frekvence základního tónu řeči určí z autokorelační funkce signálu
- frekvence základního tónu nás však zajímá pouze u znělých segmentů, proto její konečnou hodnotu vypočítáme podle vzorce

$$F_0 = f_{vz}/k. \quad (4.4)$$

V tomto případě značí k hodnotu indexu maxima autokorelační funkce signálu znělého segmentu. [2]

Určení znělosti segmentu – znělost segmentu lze určit na základě charakteristických příznaků pro znělý či neznělý úsek. Znělý úsek řeči se vyznačuje tím, že má vysokou úroveň energie (myšleno oproti úseku neznělému) a za předpokladu, že se ve znělém úseku jedná o periodický signál, lze říci, že počet průchodů signálu nulovou hodnotou je výrazně nižší než u úseku neznělého. Signál v neznělém úseku má charakter šumu a proto lze konstatovat, že počet průchodů signálu nulovou hodnotou je několikrát vyšší než u úseku znělého. energii signálu a počet průchodů signálu nulovou hodnotou lze vyjádřit následujícími dvěma rovnicemi

$$E = \frac{1}{N} \sum_{n=0}^{N-1} (x[n])^2 \quad , \quad (4.5)$$

$$ZCR = \sum_{n=0}^{N-1} (|sgn((s)[n] - (s)[n-1])|) \quad . \quad (4.6)$$

Pro přísnější třídění segmentů je vhodné nastavit podmínky tak, aby segment splňoval nastavená kritéria současně. To znamená, že jeho energie bude vyšší než nastavený práh a počet průchodů signálu nulovou hodnotou bude nižší než nastavený práh. Pokud je splněna tato podmínka, je příznak segmentu označen jako znělý. Tento příznak segmentu je kódován logickým binárním signálem 1,0 a záleží na realizaci, zda je použita kladná či záporná logika. Přístup pomocí porovnání energie a počtu průchodů signálu nulou je sice výpočetně nejjednodušší, ale ne ve všech případech funguje tak, jak by měl. Popis dalších metod pro určení znělosti je možné najít na různých webových stránkách.³

K obnovení signálu na přijímací straně pak dochází následující způsobem: syntetizující filtr, jehož přenosová charakteristika je dána predikčními koeficienty LPC, je buzen střídavě pulzním signálem o frekvenci F_0 a generátorem šumu. Přepínání mezi těmito dvěma budícími signály je zajištěno parametrem znělý nebo neznělý segment. Následuje filtrace pomocí filtru deemfáze.

Konkrétní popis jednotlivých subsystémů modelu:

1. **Zdroj řečového signálu** - jako zdroj řečového signálu je použita nahrávka hlasu o celkové délce trvání 7 sekund, se vzorkovací frekvencí 8 kHz a bitovým rozlišením 16 bitů na vzorek.
2. **Buffer** - obstarává segmentaci signálu na segmenty o volitelné délce. Pro simulaci je použita délka rámců 128 vzorků, což při vzorkovací frekvenci 8 kHz odpovídá 16 ms. Protože řada algoritmů ve zpracování signálu používá rychlou Fourierovu transformaci, je vhodné volit délku segmentů 2^N to znamená například 64,128,256 vzorků.

³ <<http://noel.feld.cvut.cz/vyu/prj-czs-asi/prasek/>>

3. **Preemfáze** - je metoda pro zvýraznění vyšších kmitočtů. Konkrétní popis filtru viz. výše. Koeficient a byl nastaven na hodnotu 0,92.
4. **Hammingovo okénko** - provádí váhování segmentu okénkem. V modelu je použito váhování Hammingovým okénkem, které utlumí signál na okrajích segmentu a brání tak rušivým přechodným jevům. Váhování okénkem bylo již v podstatě jednou provedeno a to použitím buferu, ale u buferu se jedná o pravoúhlé okénko, které není pro zpracování pomocí LPC analýzy vhodné. Hammingovo okénko délky N je definováno následující rovnicí.

$$w(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N}, \text{ kde } 0 \leq n < N. \quad (4.7)$$

5. **LPC analýza** - Pro analýzu byl použit blok programu Simulink Autocorrelation LPC. Jeho výstupem mohou být tři veličiny. Pro náš účel byly použity koeficienty A , lineární predikční koeficienty, které jsou následně použity přímo pro syntetizační filtr. Počet koeficientů je vždy o jeden více než je řád predikce. Optimální řád predikce je možno určit jednoduchým výpočtem podle vzorce

$$p = \left\lfloor \frac{f_{vz}}{1000} \right\rfloor + 2, \quad (4.8)$$

kde p je řád predikce a f_{vz} je vzorkovací frekvence zpracovávaného signálu.

Optimální řád predikce v modelu je tedy:

$$p = \left\lfloor \frac{8000}{1000} \right\rfloor + 2 = 10. \quad (4.9)$$

A koeficientů je tedy 11 a první koeficient je 1. Pokud bychom se neřídili vzorcem pro optimální řád predikce, mohlo by dojít k následujícím stavům.

- při nižší hodnotě řádu predikce než je řád optimální dojde k nežádoucímu splývání formantů.
- při hodnotě řádu predikce vyšší než 25 až 30 se projevuje nežádoucí modulace spektra základním tónem.

Výstupem simulinkového bloku mohou být ještě koeficienty K (koeficienty odrazu) a predikční chyba P . Avšak ani jeden z těchto výstupů není třeba při simulaci používat. V praktické realizaci by však bylo vhodné použít koeficientů K , které mají tu vlastnost, že jejich hodnoty nabývají maximálně hodnoty 1, což je pro zpracování signálů ideální. Koeficienty K a A lze mezi sebou přepočítat.⁴

⁴<http://noel.feld.cvut.cz/vyu/asi/cv_asi_00_LPC3.html>

6. **Určení znělosti segmentu** - popis algoritmu na jehož základě se rozhodne, zda je segment znělý či neznělý, je uveden výše. Nastavení prahových úrovní musí být provedeno poněkud „nepřívětivým“ způsobem. V našem případě máme k dispozici zdrojový řečový signál, proto je vhodné si průběh veličin, na základě kterých se rozhoduje, zobrazit a práh určit odhadem. Tato metoda je primitivní a dala by se označit jako pokus – omyl. Obrázek 4.2 zobrazuje průběh energie signálu a koeficient počtu průchodů signálu nulou pro řečový signál, následně z těchto parametrů odvozený binární signál určující znělost segmentu.



Obr. 4.2: Průběh energie a koeficientu počtu průchodu signálu nulou, binární signál určující příznak znělý/neznělý úsek

Výpočet těchto veličin zajišťují funkce *energie.m* a *zcr.m*, které jsou v modelu použity formou bloků. Za nimi následují rozhodovací bloky Relay, v nichž je nastaven rozhodovací práh. Na základě zobrazení energie E a koeficientu počtu průchodů signálu nulou, byly rozhodovací prahy nastaveny na hodnoty $E = 0,002$ a $zcrh = 0,2$. Na vstupu subsystému je sice signál normalizován, ale i tak nelze prahy nastavit na pevně, z důvodu rozmanitosti možných vstupních signálů. Pokud signál splňoval podmínku, že energie segmentu je vyšší než stanovený práh a koeficient průchodu nulou je nižší než stanovený práh, pak se na výstupu subsystému objevil logický signál s úrovní 1. V jiném případě to byl logický signál s úrovní 0. O dodržení podmínky a generování výstupního signálu se stará blok logické operace AND.

7. **Určení frekvence základního tónu** - pro určení frekvence základního tónu řeči je v modelu použita metoda centrálního klipování, která je založená na autokorelaci signálu. Popis funkce algoritmu centrálního klipování je uveden výše. Na vstup subsystému přichází segment signálu $x(i)$ a je nutné pro něj určit práh pro další operace se segmentem. Pro určení prahu je nutné znát ještě segmenty $x(i-1)$ a $x(i+1)$. Pro aktuální segment je vypočítán práh následovně: z maximální hodnoty vedlejších segmentů je určena menší a ta je vynásobena konstantou $k=0,8$. Blok zpoždění ve větvi pro určení prahu zajišťuje získání maxim z vedlejších segmentů. Se segmentem, pro který je práh určován, se nepracuje, ale do výpočtu je zařazen až v následujícím cyklu pro další segment. Blok zpoždění v přímé větvi má za úkol aktuální segment zpoždit o dva takty, kdy na výstupu určení prahu bude hodnota jemu určená. Na začátku simulace se jistě v modelu dopouštíme chyby, protože počítáme s nulovými hodnotami. Chyba se však projeví pouze na prvních dvou respektive třech segmentech. To je maximálně 48 ms ze začátku řečového signálu. Pro lidské ucho je to velmi krátká doba na adaptaci, proto chybu nezaznamená. Blok Dead zone dynamic provádí prahování segmentu podle aktuální hodnoty prahu. Prahování musí být provedeno jak v kladných tak záporných hodnotách signálu, proto je hodnota prahu pro záporné hodnoty násobena -1. Výstupní signál je pak blokem Sign převeden na signál o úrovních 1, 0, -1. Kladný vstupní signál je roven 1, záporný -1, nula zůstává nulou. Následuje blok Autokorelace, který provede se segmentem signálu operaci autokorelace. Výstupem autokorelace, respektive korelace signálu samého s sebou, je obvykle signál délky $2 \cdot N - 1$, což znamená, že bychom museli výstupní signál omezit na původní délku segmentu od maxima korelace. Výstupem tohoto bloku je ale signál délky N , což je pro náš případ ideální situace. Následuje omezení délky segmentu, a to z důvodu hledání maxima autokorelační funkce pouze v rozmezí námi hledaných hodnot frekvence základního tónu. Konkrétně $f_{(min)} = 80\text{Hz}$ až $f_{(max)} = 400\text{ Hz}$. Nové hranice a tím i novou délku segmentu lze určit podle vzorců

$$L_{min} = \frac{f_{vz}}{f_{max}} \quad , \quad (4.10)$$

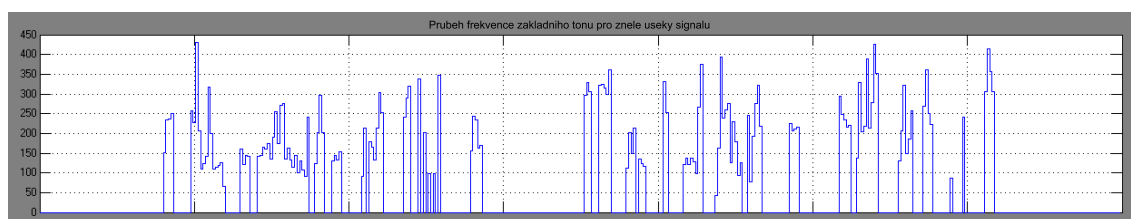
$$L_{max} = \frac{f_{vz}}{f_{min}} \quad . \quad (4.11)$$

Samotné omezení segmentu provádí funkce *omezeni.m*. Na obrázku 4.4 je vidět první maximum autokorelační funkce na indexu $R_k = 60$. Omezením s hraničními hodnotami indexu $R_k = 20$ a $R_k = 100$ se posune na index $R_k = 40$. Z toho důvodu je nutné pro správný výpočet přičíst k indexu hodnotu L_{min} .

Přičtení provádí až výpočetní blok Vypocet F_0 . Dalším v pořadí je blok, který hledá právě maximum autokorelační funkce, respektive námi požadovaný index. Blok převodu typu dat je vložen pouze z toho důvodu, že výstupní index je číslo typu int32 a k výpočtu požadujeme číslo typu double. Blok terminator zabraňuje výpisu chybových hlášení Simulinku o nepřipojeném portu. Blok funkce Vypocet F_0 provádí výpočet hodnoty frekvence základního tónu pomocí vzorce

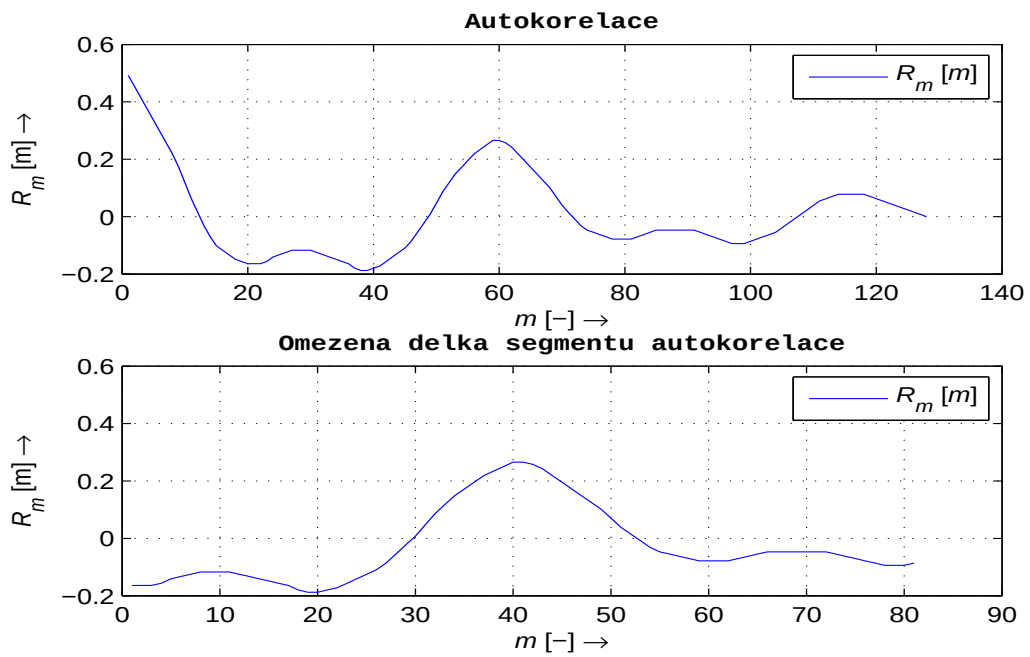
$$F_0 = \frac{f_{vz}}{R_k} \quad . \quad (4.12)$$

Průběh frekvence základního tónu je možné filtrovat dolní propustí o mezním kmitočtu 40–60 Hz. Tato možnost je využita a navíc je možné použít filtraci mediánovým filtrem. Ten je řazen před dolní propust. Na obrázku 4.3 je možné vidět průběh frekvence získaný na výstupu subsystémového bloku.

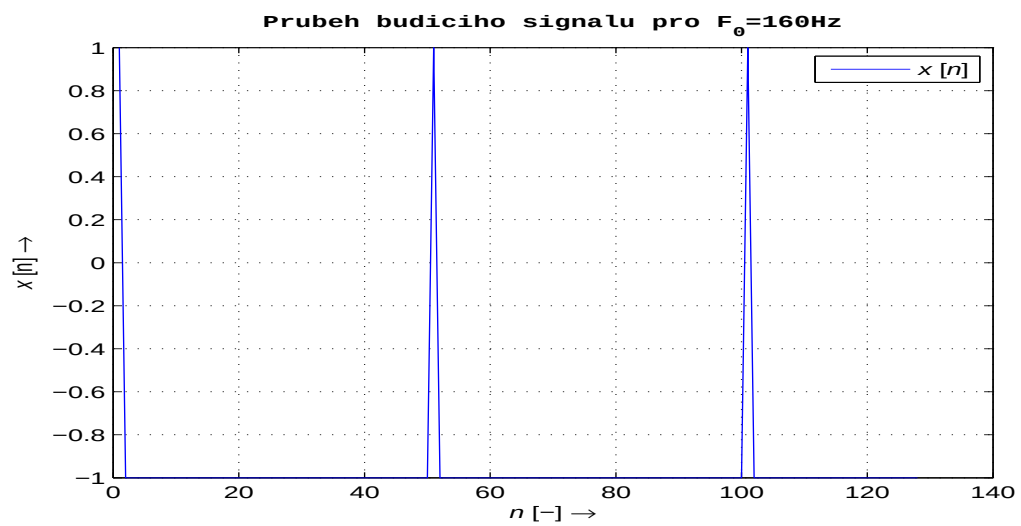


Obr. 4.3: Průběh frekvence základního tónu

8. **Generátor budících impulsů** - generování budících impulsů obstarává funkce *gener.m*. Vstupním parametrem je hodnota frekvence základního tónu a na jejím základě je pak vygenerován 128 vzorků dlouhý segment s impulzy. Jedná se o obdélníkový signál, který má střihu jedno procento. Na obrázku 4.5 je vidět segment s budícími impulzy o frekvenci $F_0=160\text{Hz}$.
9. **Generátor šumu** - pro generování budícího šumového signálu je použit blok Simulinku Gaussian Noise Generator. Nastavení parametrů generátoru bylo provedeno na základě subjektivního posouzení syntetické řeči. Střední hodnota šumu je nastavena na nulu, variabilita na 2. Výstupem generátoru je segment dlouhý 128 vzorků.
10. **Switch** - zajišťuje přepínání generátorů budících signálů na základě logického binárního signálu. Pokud se na rozhodovacím vstupu switchu objeví 1, znamená to, že k buzení má být použit generátor impulsů, protože se jedná o znělou hlásku. V opačném případě, tedy při hodnotě logického signálu 0, je vstup přepnut na generátor šumu.



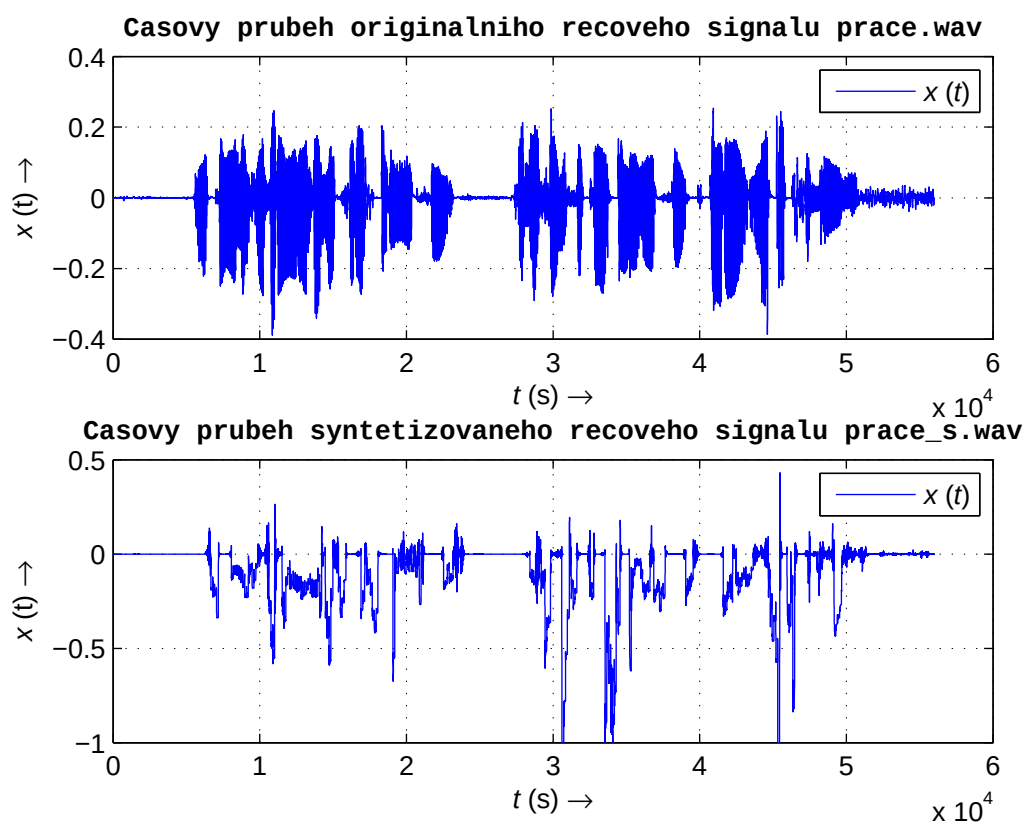
Obr. 4.4: Autokorelace signálu a omezení autokorelace



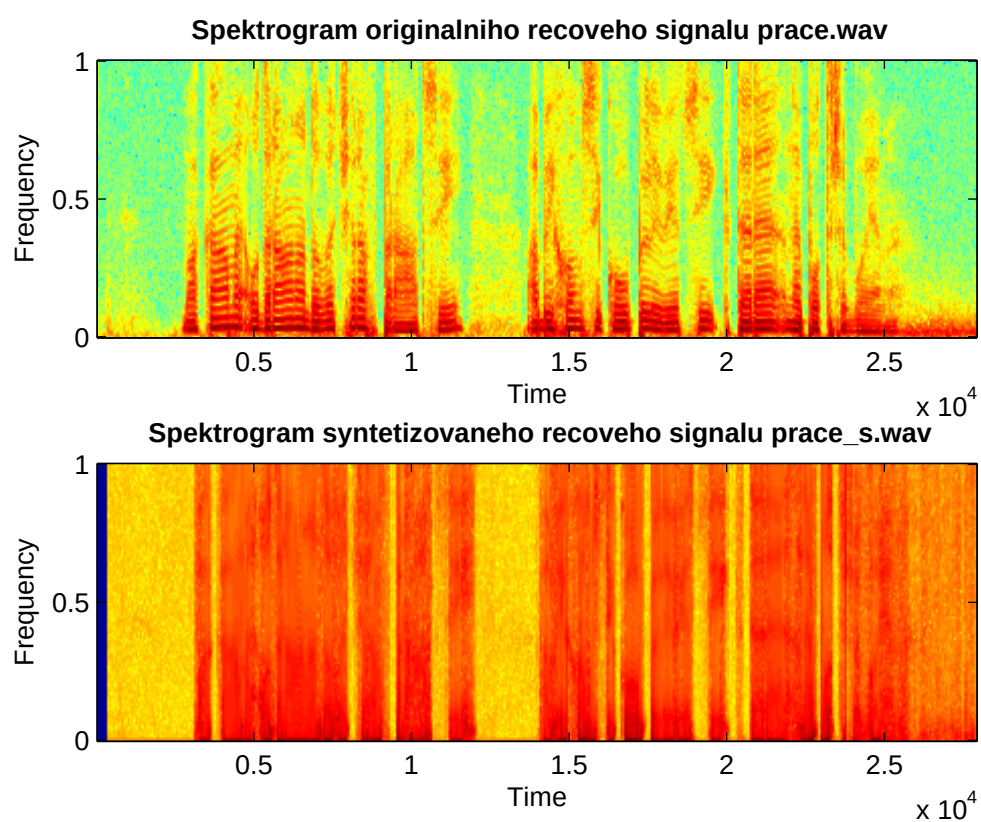
Obr. 4.5: Průběh budícího signálu s frekvencí 160 Hz

11. **Product** - násobení segmentu vygenerovaného signálu úrovní energie, která byla pro původní segment vypočítaná v subsystému Určení znělosti segmentu. Násobení segmentu hodnotou energie by mělo spolu s průběhem frekvence základního tónu zachovat prozodii řeči.
12. **Syntetizující filtr** - syntetizující filtr je realizován pomocí bloku Simulinku, kterým je možno navrhnout digitální filtr. Tento blok je, co se týče nastavení, velmi variabilní. V našem modelu je použito následující nastavení: blok má nastaveny dva vstupy - jeden pro koeficienty, které jsou výstupem LPC analýzy, druhý pro zdroj budícího signálu. Filtr je typu IIR, pouze s póly, zvolena je možnost odebrání prvního koeficientu ze struktury filtru, a to z toho důvodu, že první koeficient A je vždy roven jedné. Sníží se tím výpočtová náročnost o jeden řád. Obnovování koeficientů filtru je nastaveno po jednom rámci, což je odpovídající zpracování řeči po segmentech v našem modelovém případě.
13. **Deemfáze** - filtr je inverzní oproti filtru preemfáze, koeficient musí být nastavený na stejnou hodnotu jako u preemfáze, to znamená na 0,92. V modelu není filtr zařazen, protože zesiloval nízké kmitočty do té míry, že srozumitelnost klesla na velmi nízkou úroveň.
14. **Unbuffer** - blok by měl provádět rekonstrukci segmentů do jednoho vektoru, avšak pro přehrání zvukového výstupu ani pro uložení do souboru .wav není jeho použití potřeba, a to z důvodu nepoužití přesahu rámců v bloku Buffer.
15. **Gain** - blok provádí zesílení signálu. Gain je nastaven na 5, aby hlasitost signálu byla zhruba stejná jako originál.
16. **Zvukový výstup** - blok slouží pro poslech výstupního řečového signálu, odesílá digitální signál ke zpracování zvukové kartě. Ta jej převádí na analogový signál, který se objeví na výstupu zvukové karty pro elektroakustický měnič.
17. **Uložení řečového signálu** - blok provádí uložení syntetizovaného řečového signálu do souboru .wav v 16 bitovém rozlišení.
18. **Delay** - bloky zajišťují zpoždění přenosu parametrů tak, aby byly správně přiřazeny k segmentu, který je aktuálně zpracováván.

Časový průběh originálního a syntetizovaného signálu je zobrazen na obrázku 4.6, spektrogramy signálů na obrázku 4.7. Již podle prvního pohledu na tyto průběhy je jasné, že originální a syntetizovaný signál mají dosti rozdílný charakter. Rozdíl je samozřejmě slyšet v nahrávkách `prace.wav` a `prace_s.wav`. Originální signál je čistý, plný a srozumitelný. Syntetizovaný signál je strohý, monotónní a plný parazitních zvuků. Důvody tohoto rozdílu jsou jednoduché. Primitivní klasifikace znělosti segmentu řeči, stejný budící signál pro všechny znělé úseky, méně přesné určení frekvence základního tónu. Popis možného řešení těchto problémů je uveden v závěru práce.



Obr. 4.6: Časový průběh originálního a syntetizovaného řečového signálu



Obr. 4.7: Spektrogram originálního a syntetizovaného řečového signálu

5 ZÁVĚR

Úkolem práce byl souhrn a rekapitulace již známých standardů a norem definujících algoritmy pro zpracování řečového signálu a simulace některých z nich. Ve třetí kapitole je tedy souhrn standardů a norem, čímž je splněn první cíl práce. Přínosem práce a zároveň splněním dalšího cíle je pak modelování algoritmů v programu Simulink. Popis modelování je možné najít ve čtvrté kapitole Modelování v programu Matlab Simulink. Srovnání standardů je provedeno pomocí tabulky, kterou lze nalézt v příloze B.1. Na začátku je nutno podotknout, že moderní vokodéry jsou již tak složité, že jejich simulace a popis není v rámci diplomové práce možný jak časově, tak rozsahem textu. Diplomová práce by se musela zabývat pouze jedním z těchto kodérů (vokodérů), a i tak by byl rozsah práce značný. Simulace algoritmů pro zpracování signálu v časové doméně přináší náhled na metody, které jsou základním učivem v předmětech zpracování signálů, získané průběhy signálů demonstrují klady a zápory těchto algoritmů, jako jsou nedostatečně rychlá adaptace nebo nedostatečné rozlišení úrovnových změn signálů. V modelu LPC vokodéru je použito několik dílčích algoritmů pro získání parametrů řečového signálu, které by v praktické realizaci byly zakódovány a odeslány komunikačním kanálem. Kanálové či linkové kódování jsou v modelu zanedbány. Řečový signál je samozřejmě zpracováván po segmentech, konkrétně délky 128 vzorků, a pro každý segment jsou určeny následující příznaky: frekvence základního tónu, znělost/neznělost, energie a lineární predikční koeficienty. Na základě těchto příznaků je na přijímací straně řeč obnovována. Jelikož se jedná pouze o základní LPC vokodér, je nutné podotknout, že řeč má značně syntetický charakter a její srozumitelnost je na nízké úrovni. LPC vokodér sám o sobě tedy neposkytuje kvalitní výsledky i přesto, že je při simulaci použit pouze čistý řečový signál. Úroveň kvality je dána několika aspekty. Za prvé je to v našem modelu nepoužitý překryv rámců. Překryv nebyl použit z důvodu nejasné funkčnosti bloku Unbuffer, který má zajišťovat zpětnou rekonstrukci signálu ze segmentů do jednoho vektoru. Problém nebyl vyřešen ani po konzultacích se zkušenějšími osobami. Za druhé se na kvalitě řeči podílí budící signál. V modelu je použit obdélníkový signál s frekvencí základního tónu a generátor šumu. Šumový generátor je použit vždy pro buzení filtru pro syntézu sykavek. Budící signál filtru pro syntézu znělých úseků je použit impulzový signál, jedná se například o syntézu a, e ,o. Problém nastává s buzením třených nebo explozivních hlásek jako jsou f, v, b, p. Budící signál je pak složitější, bere v potaz umístění impulzů na časové ose signálu a také jejich amplitudu, případně se jedná o součty signálů. Pro zlepšení kvality je vhodné použít kódovou knihu budících signálů, v které je respektováno například rozmístění a amplituda impulzů v budícím signálu. Za třetí je to určování znělých úseků signálu, které je funkční, ale v moderních kodérech není omezeno pouze na dvě úrovně

znělý - neznělý. Příznak znělosti pak může nabývat několika úrovní, například znělý, poloznělý, neznělý a podobně. K buzení se pak využívá například směsi šumového signálu a impulzního signálu. Simulace LPC vokodéru potvrzuje tedy fakt, že parametrické zpracování řeči není opravdu jednoduché a skýtá mnohá úskalí. Hodnocení a porovnání kodérů (vokodérů) bylo dalším cílem práce. Dle zadání je nutné se zaměřit na ty s nízkými výstupními bitovými rychlostmi. Mezi špičkové vokodéry je nutno zařadit vokodéry typu ACELP. Konkrétně ITU-G.722.2 a GSM AMR, které jsou díky širokému spektru výstupních bitových rychlostí vhodné pro přenosové kanály, ve kterých je nutná adaptace výstupní rychlosti kvůli proměnnému rušení v přenosovém kanálu (typicky mobilní sítě). Navíc se s jejich použitím již počítá do budoucna v nových generacích mobilních a hlasových sítích. Na přední příčky lze zařadit i ISAC a ILBC, kodéry určené pro hlasové internetové aplikace jako je Skype nebo Google Talk. Výstupní bitová rychlost se pohybuje okolo 10 kbit/s což je vyhovující i pro běžného uživatele s linkou od 128 kbit/s. Pro shrnutí tedy: v poměru subjektivní vjem / výpočetní náročnost (a cena), s ohledem na dnešní možnosti výpočetních kapacit a napájecích (akumulátorových) zdrojů, je možné konstatovat, že na předních místech se drží kodéry jejichž základem je LPC vokodér. Konkrétně se jedná o vokodéry využívající knihoven budících signálů v rozličných modifikacích, CELP, ACELP, RELP, a kodéry používané v internetových aplikacích. Implementace složitějších kodérů do zařízení běžné denní potřeby je samozřejmě otázkou peněz. Záleží na konkrétních firmách, zda najdou peníze na vývoj a realizaci takovýchto zařízení. V tomto případě je trefné srovnání s evolucí, silnější přežije.

LITERATURA

- [1] AKSAMÍT, J. *Přenos komprimovaných řečových signálů*, [online]. Praha: konference Audio Technologies and Processing, ČVUT, 2005. Dostupné z URL: <<http://www.radio.feld.cvut.cz/AES/atp2005/proc/atp05p08.pdf>>.
- [2] ATASSI, H. *Metody detekce základního tónu řeči*, [online]. Brno: Elektrorevue, 2008. 17 s. Dostupné z URL: <<http://www.elektrorevue.cz/cz/clanky/zpracovani-signalu/10/metody-detekce-zakladniho-tonu- reci/>>.
- [3] HANUS, S. *Bezdrátové a mobilní komunikace*, 1. dotisk. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky, Purkyňova 118, 2003. 135 s. ISBN 80-214-1833-8.
- [4] THE INTERNATIONAL TELEGRAPH AND TELEPHONE CONSULTATIVE COMMITTEE *Recommendation P.800*, Methods for subjective determination of transmission quality. ITU-T 1996.
- [5] THE INTERNATIONAL TELEGRAPH AND TELEPHONE CONSULTATIVE COMMITTEE *Recommendation P.862*, Amendment 1: Revised Annex A: Source code for the reference implementation and conformance tests. ITU-T 2003.
- [6] KARBAN, P. *Výpočty a simulace v programech Matlab a Simulink*, Brno : Computer Press, 2001. 220 s. ISBN 80-251-1301-9.
- [7] NOVOTNÝ, V. *Komunikační prostředky mobilních sítí*, Elektronické skriptum VUT, Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, Purkyňova 118, 2006. 92 s.
- [8] Novák, A. *Zpracování řečového signálu v systému GSM*, [online]. Praha: Semestrální práce předmětu Mobilní komunikace ČVUT, 2005. 4 s. [cit. 2005-4-27]. Dostupné z URL: <http://www.radio.feld.cvut.cz/personal/mikulak/MK/MK05_semestralky/MK_novaka4.pdf>.
- [9] PRCHAL, J., ŠIMÁK, B. *Digitální zpracování signálů v telekomunikacích*, Skriptum ČVUT. Praha: Vydavatelství ČVUT, 2001. 220 s. ISBN 80-01-02149.
- [10] Smékal, Z. *Číslíkové zpracování signálů : skripta k předmětu MCSI*, Elektronické skriptum VUT. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, Purkyňova 118, 2007. 188 s.

- [11] ŠKORPIL, V., PROCHÁZKA, T. *Kompresa dat u ISDN*, [online]. Brno: Elektrorevue, 2002. 10 s. Dostupné z URL: <<http://www.elektrorevue.cz/clanky/02068/index.htm>>.
- [12] THE INTERNATIONAL TELEGRAPH AND TELEPHONE CONSULTATIVE COMMITTEE *Recommendation G.727*, [online]. 1990. Dostupné z URL: <<http://www.itu.int/rec/T-REC-G.727-199012-I/en>>.
- [13] THE INTERNATIONAL TELEGRAPH AND TELEPHONE CONSULTATIVE COMMITTEE *Work Programme-Recommendation G.718*, [online]. 2008. Dostupné z URL: <http://www.itu.int/ITU-T/workprog/wp_search.aspx?isn_task=344>.
- [14] Voice Age Corporation *AMR-WB*, [online]. 2008.[cit.2002-12-19]. Dostupné z URL: <<http://www.voiceage.com/amrwb.php#standards>>.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

ACELP Adaptive Code Excited Linear Prediction

ADM Adaptive Delta Modulation – adaptivní delta modulace

ADPCM Adaptive Differential Pulse Code Modulation – adaptivní diferenční kódová delta modulace

AMBE Advanced Multi Band Excitation

AMR–WB Adaptive Multi Rate Wide Band

A-law mezinárodní standard pro kompresi řeči, používaný v Evropě

CDMA Code Division Multiple Access – přístupová metoda mobilních sítí

CELP Code Excited Linear Prediction

CNI Comfort Noise Insertion – funkce cíleného šumu ve sluchátku

CS–CELP Conjugate Structure Algebraic Code-Excited Linear Prediction

DCT Discrete Cosine Transform – diskrétní kosinová transformace

DM Delta Modulation – delta modulace

DPCM Differential Pulse Code Modulation – diferenční pulzní kódová modulace

DSP Digital Signal Processing – číslicové zpracování signálů

DSS Digital Speech Standard

DTX Discontinuous Transmission – funkce řídicí výstup kodeku

FM Frequency Modulation – frekvenční modulace

GSM Global System Mobile – celosvětový standard mobilních telefoních sítí

GSM–AMR Adaptive Multi Rate – mód kodeku používaný pro zpracování hlasu v GSM, plná bitová rychlost

GSM–EFR Enhanced Full Rate – mód kodeku používaný pro zpracování hlasu v GSM, plná bitová rychlost

GSM–FR Full Rate – mód kodeku používaný pro zpracování hlasu v GSM, plná bitová rychlost

GSM–HR Half Rate – mód kodeku používaný pro zpracování hlasu v GSM, poloviční bitová rychlost

HVXC Harmonic Vector Excitation Coding

ILBC Internet Low Bit Rate Codec

IMBE Improved Multi Band Excitation

IP Internet Protocol – internetový protokol

ISAC Internet Speech Audio Codec

ITU International Telecommunication Union – mezinárodní telekomunikační unie

LD–CELP Low Delay Code Excited Linear Prediction

LPC Linear Prediction Coding – lineární predikční kódování

LSP Line Spectral Pairs – koeficienty, spektrální páry

LTP Long Term Prediction

MELP Multipulse Excited Linear Prediction

MOS Mean Opinion Score

MP–MLQ Multipulse Maximum Likelihood Quantization

MPE–LTP Multi Pulse Excited Long Term Prediction – návrh kodeku pro zpracování hlasu GSM

μ -law mezinárodní standard pro kompresi řeči, používaný v Americe

PCM Pulse Code Modulation – pulzní kódová modulace

RELP Residually Excited Linear Prediction

RPELP Regular Pulse Excitation Linear Prediction

RPE–LPC Regular pulse excited-Linear Predictive Coder – návrh kodeku pro zpracování hlasu GSM

RPE–LTP Regular Pulse Excitation Long Term Prediction – kodek používaný pro zpracování hlasu v GSM

SBC–APCM Sub Band Code Adaptive Pulse Code Modulation – návrh kodeku pro zpracování hlasu GSM

VAD Voice Activity Detection – funkce skenující vstup kodeku

VBR Variable Bit Rate – proměnný bitový tok

VMR-WB Variable Multi Rate Wide Band

VoIP Voice over Internet Protocol – technologie pro přenos hlasu datovými sítěmi

VSELP Vector sum excited linear prediction

SEZNAM PŘÍLOH

A	Obsah elektronického média	62
B	Přehled	63
B.1	Přehled vlastností kodérů a vokodérů	63
B.2	Rozdělení metod zpracování řečového signálu	64
C	Modely v programu Simulink	65
C.1	Model pro simulaci metod zpracování řečového signálu v časové oblasti	65
C.2	Model DM	66
C.3	Model DM - průběh signálů	66
C.4	Model PCM	67
C.5	Model PCM - průběh signálů	67
C.6	Model DPCM	68
C.7	Model DPCM - průběh signálů	68
C.8	Model ADPCM	69
C.9	Model ADPCM - průběh signálů	69
C.10	Model LPC vokodéru	70
C.11	Subsystem pro výpočet frekvence základního tónu	71
C.12	Subsystem pro určení znělosti segmentu signálu	72

A OBSAH ELEKTRONICKÉHO MÉDIA

1. Elektronická verze diplomové práce v souboru Hlasové kodéry pro nízké přenosové rychlosti.pdf
2. Model algoritmu DM v programu Matlab Simulink verze 7.5.0 (R2007b) v souboru dm.mdl
3. Model algoritmu PCM v programu Matlab Simulink verze 7.5.0 (R2007b) v souboru pcm.mdl
4. Model algoritmu DPCM v programu Matlab Simulink verze 7.5.0 (R2007b) v souboru dpcm.mdl
5. Model algoritmu ADPCM v programu Matlab Simulink verze 7.5.0 (R2007b) v souboru adpcm.mdl
6. Model LPC vokodéru v programu Matlab Simulink verze 7.5.0 (R2007b) v souboru LPC_vokoder.mdl
7. Skripty v programu Matlab Simulink verze 7.5.0 (R2007b) ve formátu souboru .m
8. Funkce v programu Matlab Simulink verze 7.5.0 (R2007b) ve formátu souboru .m
9. Nahrávky řečového signálu v souboru prace.wav a prace_s.wav, vzorkovací frekvence 8 kHz, bitové rozlišení 16 bitů
10. Signál v souboru test.wav, vzorkovací frekvence 44100 kHz, bitové rozlišení 16 bitů
11. Přehled vlastností kodérů řečového signálu v souboru prehled_kodery.pdf
12. Přehled metod zpracování řečového signálu v souboru strom.pdf

B PŘEHLED

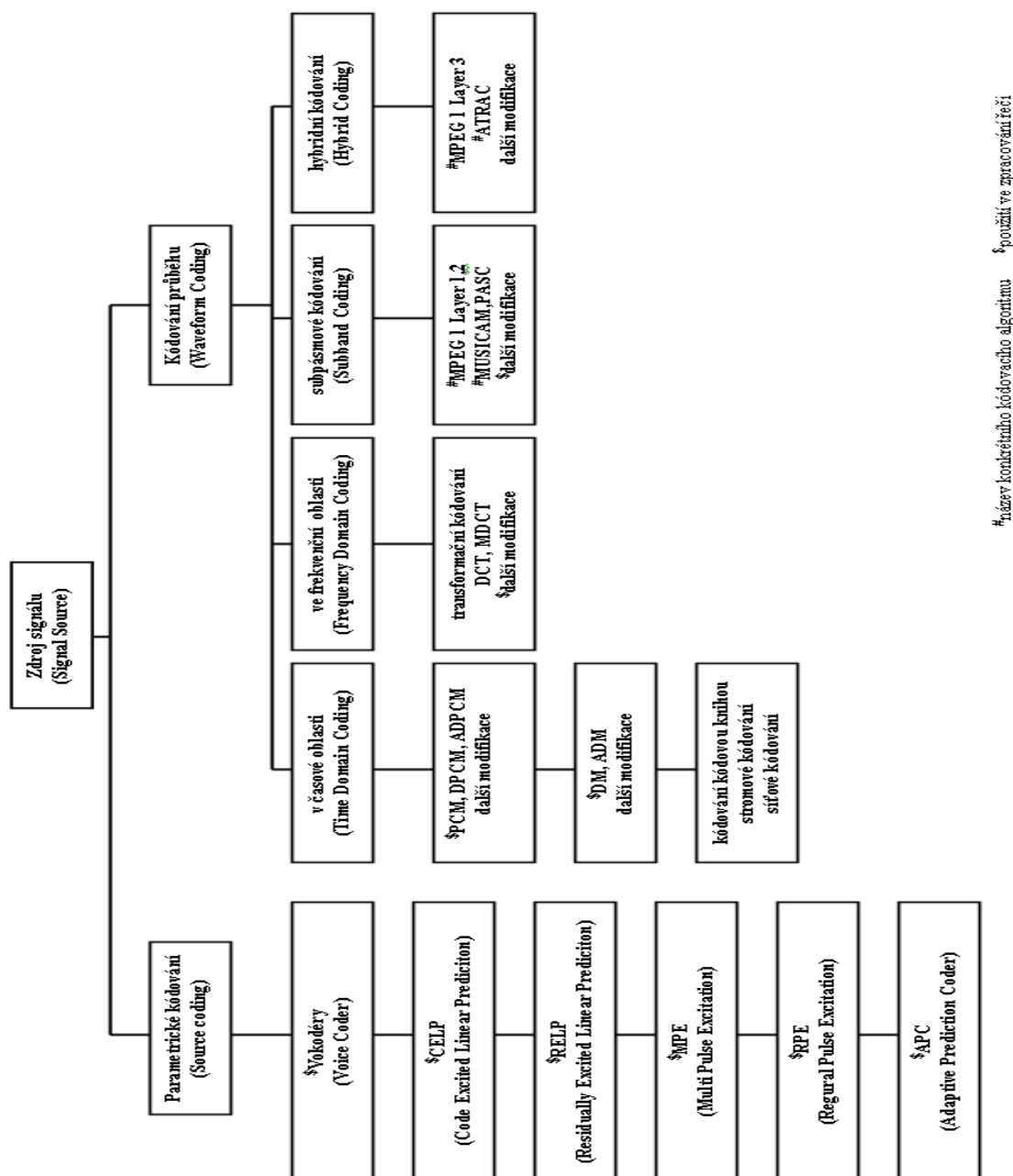
B.1 Přehled vlastností kodérů a vokodérů

STANDARD	TYP KÓDOVÁNÍ	BITOVÁ RYCHLOST	Fvz	VSTUPNÍ SIGNÁL ¹	MOS ²	VAD ³	ZPOŽDĚNÍ ALGORITMU	VELIKOST RÁMCE	LICENCE ⁴	POUŽITÍ	ROK ZVEŘEJNĚNÍ (DODATEK)
	-	kbit/s	kHz	-	-	-	ms	ms	-	-	-
ITU-G.711	PCM	64	8	úzkopásmový	4,3	ne	0,125	0,125	ne	telefonní síť	1972
ITU-G.718	CELP	8/12/16/24/32	8/16	úzkopásmový/ šírokopásmový	-	-	43(33)	20	-	-	2008
ITU-G.722	SB-ADPCM	48/56/64	16	šírokopásmový	-	ne	1,5	0,125	ne	Videokonference	1988
ITU-G.722.1	transformační kódování	24/32	16	šírokopásmový	-	ne	20	20	ano	Videokonference/ VoIP	1999
ITU-G.722.2	ACELP	6,6/8,85/12,65/ 14,25/15,85/18,25/ 19,85/23,05/23,85	16	šírokopásmový	-	ano	5	20	ano	Videokonference/ 3G	2001
ITU-G.723.1	MP-MLQ, ACELP	5,3/6,3	8	úzkopásmový	3,9/3,65	ne	37,5	30(37,5)	ano	videokonference H.324	2006
ITU-G.726	ADPCM	16/24/32/40	8	úzkopásmový	2,0/3,2/ 4,0/4,2	ne	0,125	-	ne	-	1990(2005)
ITU-G.727	Embedded ADPCM	16/24/32	8	úzkopásmový	-	ne	0,125	-	-	VoIP	1990(2005)
ITU-G.728	LD-CELP	16	8	úzkopásmový	4	ne	0,625	-	ne	VoIP	1992(2006)
ITU-G.729	CS-ACELP	8	8	úzkopásmový	4	ne	15	10	ano	-	1996
GSM FR	RPE-LTP	13	8	úzkopásmový	3,7	ne	20	20	ne	GSM	1989
GSM HR	VSELP(CELP)	6,5	8	úzkopásmový	3,5	ne	24,375	10	ne	GSM	1994
GSM EFR	ACELP	13	-	úzkopásmový	4	ne	20	20	-	GSM	1995
GSM AMR	MR-ACELP	4,75/6,15/6,90/6,70/ 7,40/7,95/10,2/12,2	12	úzkopásmový	3,79	ano	20	-	ano	GSM/UMTS	1998
VMR WB	SB-ADPCM	8,1/10,5/12,8/13,3	8/16	úzkopásmový/ šírokopásmový	-	-	-	20	-	cdmaOne/ cdma2000	-
ISAC	-	10 až 32	16	úzkopásmový	3,5	-	33,63	30, 60	-	Skype, Google talk, Yahoo! Messenger	2004
ILBC	-	13,33/15,2	8/16	úzkopásmový/ šírokopásmový	3,5	-	23,33	20, 30	ne	Skype, Google talk, Yahoo! Messenger	2004
TETRA	ACELP	4,5	-	-	-	-	-	30	ne	satelitní systémy	1994
AMBE	CELP	2 až 9,6	8	úzkopásmový	-	-	-	20	ano	Inmarsat, Iridium	1990
SPEEX	CELP	2 až 44	8/16/32	úzkopásmový/ šírokopásmový	-	ano	30, 40	30, 40	ne	VoIP/ streaming audio	

1 - úzkopásmový (3,5kHz)/šírokopásmový(7kHz)
2 - Mean Opinion Score - stupnice subjektivního hodnocení kvality kódér
3 - Voice Activity Detector (detektor hlasové aktivity)
4 - ne- licence vypršela nebo je free/ ano - licence má platnost do určitého data nebo licenci vlastní autor, který umožňuje volné použití algoritmu

Obr. B.1: Přehled vlastností kodérů a vokodérů

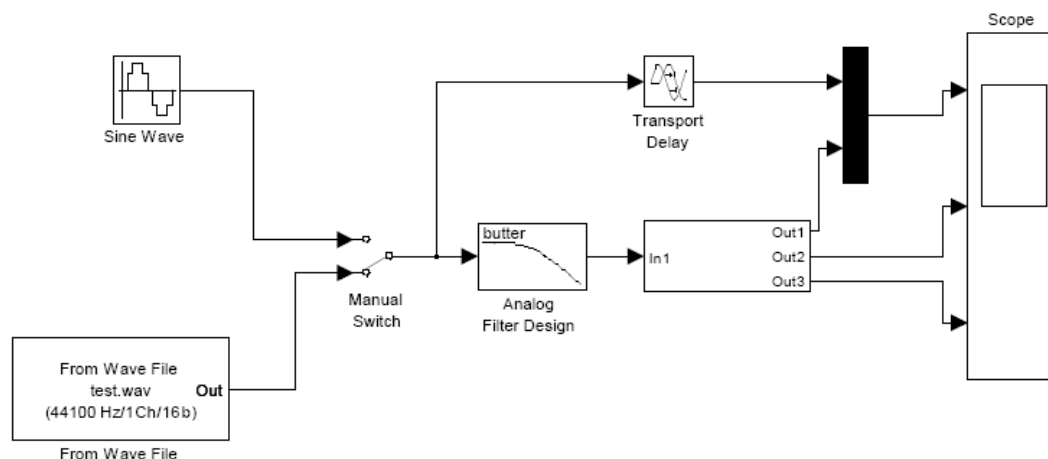
B.2 Rozdělení metod zpracování řečového signálu



Obr. B.2: Rozdělení metod zpracování řečového signálu

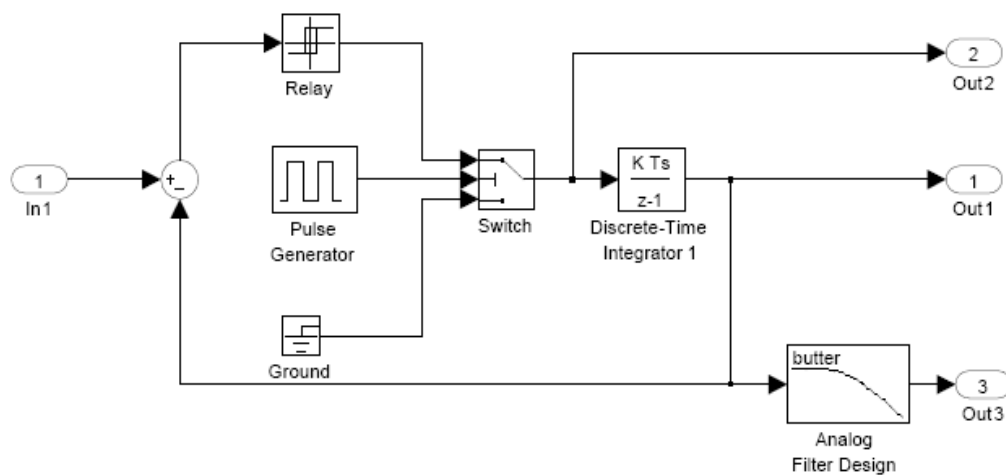
C MODELY V PROGRAMU SIMULINK

C.1 Model pro simulaci metod zpracování řečového signálu v časové oblasti



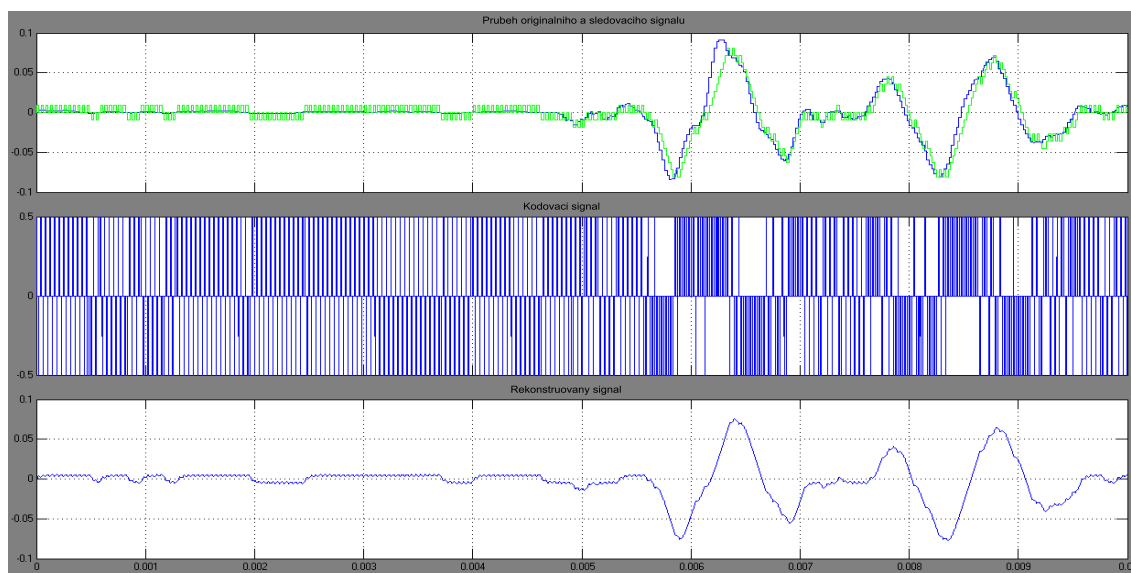
Obr. C.1: Obecný model pro simulaci

C.2 Model DM



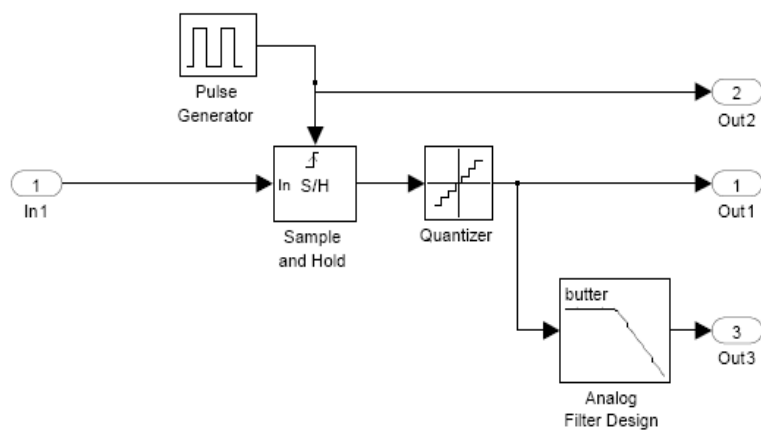
Obr. C.2: Model delta modulace

C.3 Model DM - průběh signálů



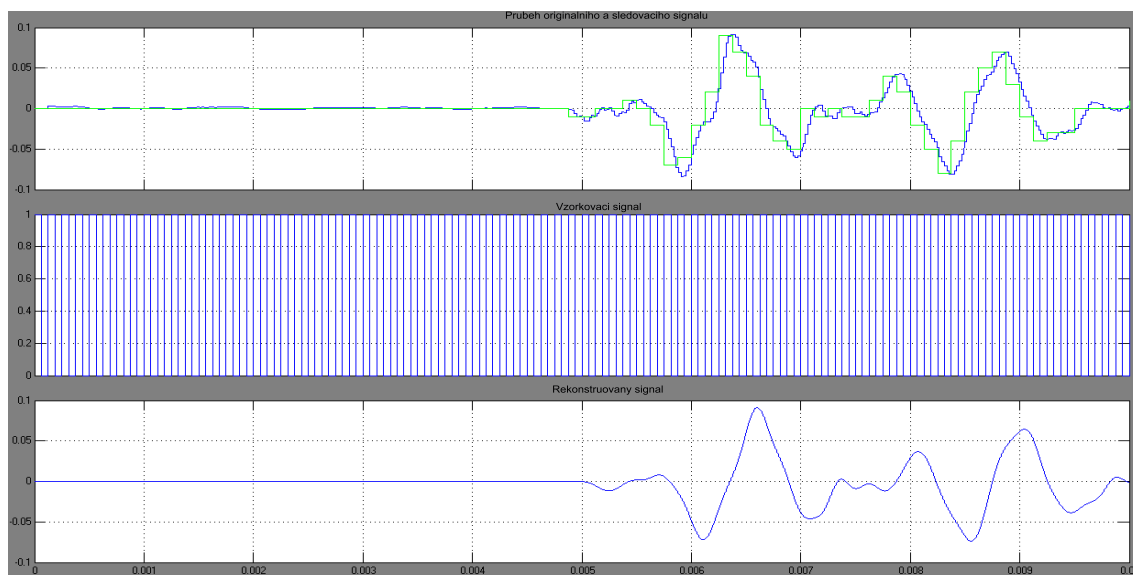
Obr. C.3: Průběh signálů delta modulace

C.4 Model PCM



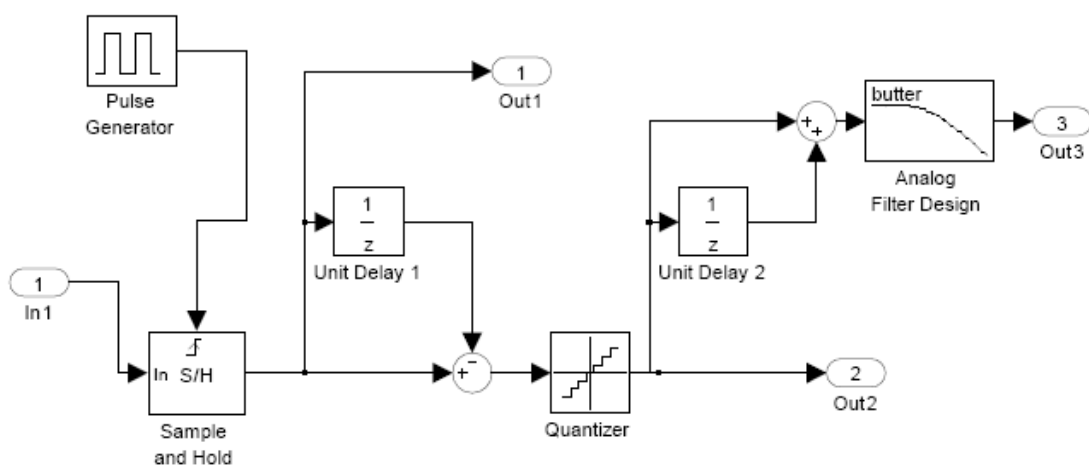
Obr. C.4: Model PCM

C.5 Model PCM - průběh signálů



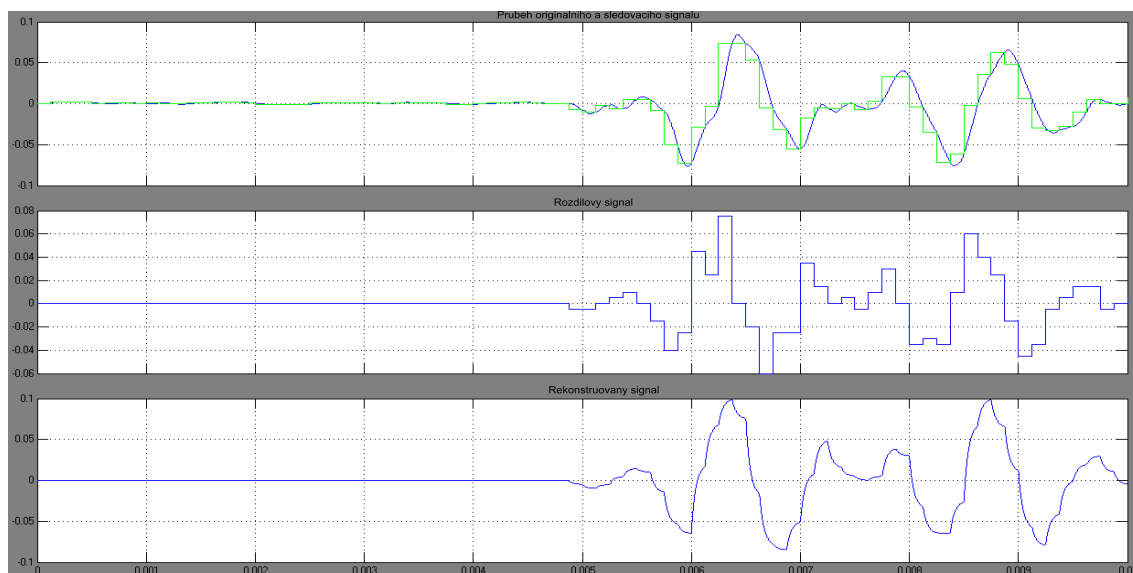
Obr. C.5: Průběh signálů pulzní kódové modulace

C.6 Model DPCM



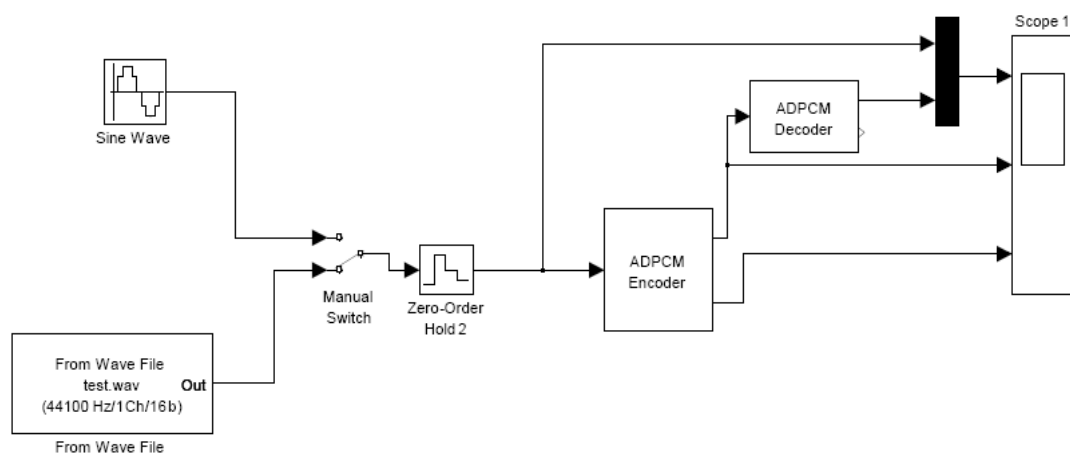
Obr. C.6: Model diferenční PCM

C.7 Model DPCM - průběh signálů



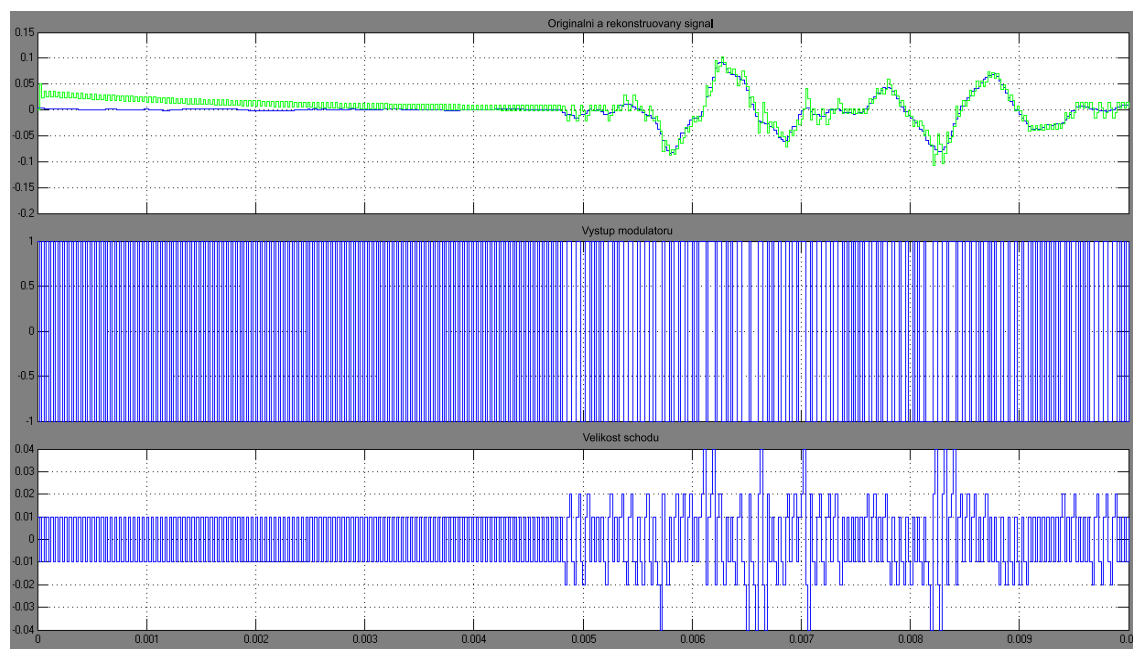
Obr. C.7: Průběh signálů diferenční pulzní kódové modulace

C.8 Model ADPCM



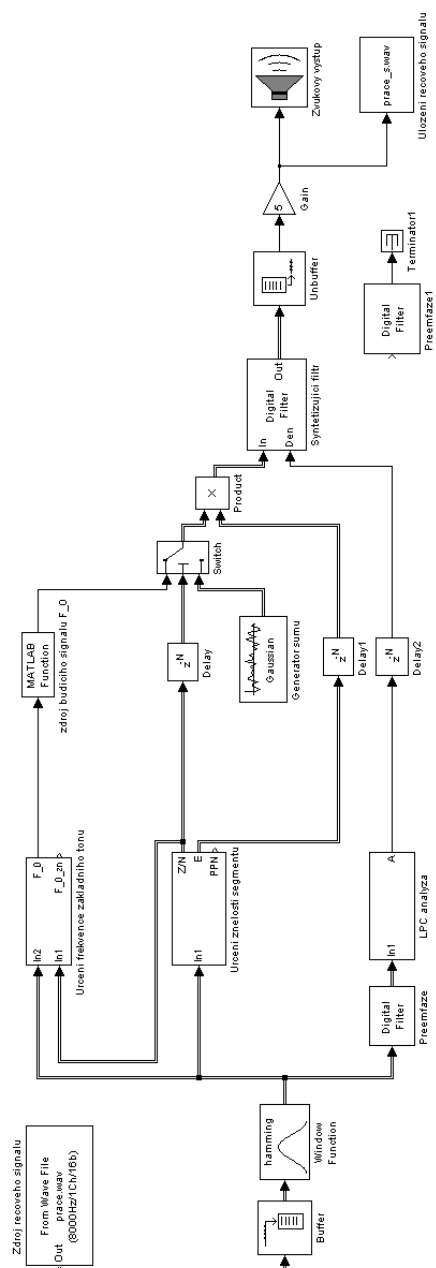
Obr. C.8: Model ADPCM

C.9 Model ADPCM - průběh signálů



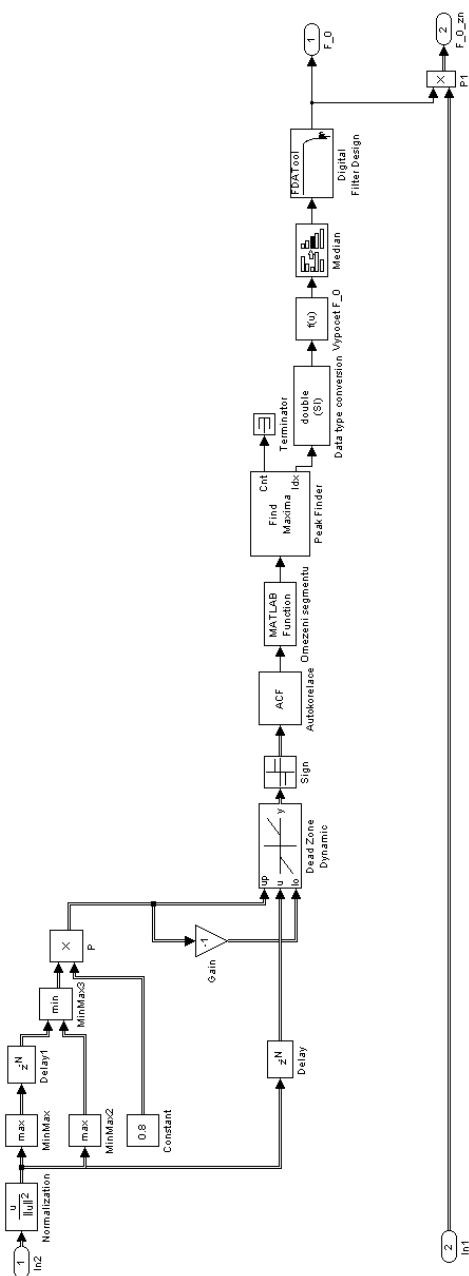
Obr. C.9: Průběh signálů adaptivní diferenční pulzní kódové modulace

C.10 Model LPC vokodéru

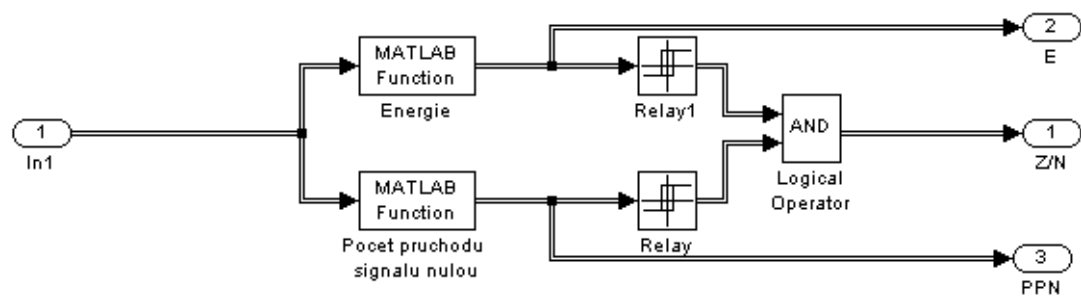


Obr. C.10: Model LPC vokodéru

Obr. C.11: Subsystem pro výpočet frekvence základního tónu



C.12 Subsystem pro určení znělosti segmentu signálu



Obr. C.12: Subsystem pro výpočet frekvence základního tónu