

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

LPC KÓDOVÁNÍ ŘEČI

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

ONDŘEJ ZAPLETAL

BRNO 2008



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND
COMMUNICATION
DEPARTMENT OF TELECOMMUNICATIONS

LPC KÓDOVÁNÍ ŘEČI LPC SPEECH CODING

BAKALÁŘSKÁ PRÁCE
BACHELOR'S THESIS

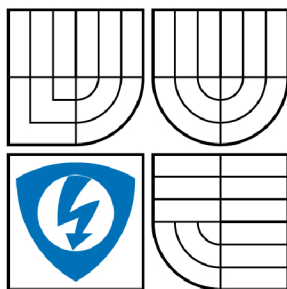
AUTOR PRÁCE
AUTHOR

ONDŘEJ ZAPLETAL

VEDOUCÍ PRÁCE
SUPERVISOR

MGR. PAVEL RAJMIC, PH.D.

BRNO 2008



VYSOKÉ UČENÍ
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

Ústav telekomunikací

Bakalářská práce

bakalářský studijní obor

Teleinformatika

Student: Zapletal Ondřej

ID: 78368

Ročník: 3

Akademický rok: 2007/2008

NÁZEV TÉMATU:

LPC kódování řeči

POKYNY PRO VYPRACOVÁNÍ:

Nastudujte základy metody LPC, která je účinná a často používaná technika pro kódování řeči. Svě znalosti dejte do souvislosti s matematickými tzv. ARMA modely. Simulujte v Matlabu řetězec zpracování řečového signálu od vzorkování až po vyslání po přenosovém kanále. Zhodnoťte výhody a nevýhody stávajících přístupů u dnes používaných řečových kodérů.

DOPORUČENÁ LITERATURA:

[1] DELLER, J.R., HANSEN, J.H.L., PROAKIS, J.G: Discrete-Time Processing of Speech Signals (IEEE Press Classic Reissue). Wiley-IEEE Press; Reprint edition, 1999. ISBN 0780353862

[2] DAVIS, G.M.: Noise reduction in speech applications. Davis, G.M., editor, CRC Press LLC, 2002. ISBN 0-8493-0949-2

Termín zadání: 11.2.2008

Termín odevzdání: 4.6.2008

Vedoucí práce: Mgr. Pavel Rajmic, Ph.D.

prof. Ing. Kamil Vrba, CSc.

předseda oborové rady

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

LICENČNÍ SMLOUVA

POSKYTOVANÁ K VÝKONU PRÁVA UŽÍT ŠKOLNÍ DÍLO

uzavřená mezi smluvními stranami:

1. Pan/paní

Jméno a příjmení: Ondřej Zapletal
Bytem: Došlíkova 2849/34, 63600, Brno - Židenice
Narozen/a (datum a místo): 17.5.1986, Brno

(dále jen "autor")

a

2. Vysoké učení technické v Brně

Fakulta elektrotechniky a komunikačních technologií
se sídlem Údolní 244/53, 60200 Brno 2
jejímž jménem jedná na základě písemného pověření děkanem fakulty:
prof. Ing. Kamil Vrba, CSc.

(dále jen "nabyvatel")

Článek 1

Specifikace školního díla

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

- disertační práce
- diplomová práce
- bakalářská práce

jiná práce, jejíž druh je specifikován jako

(dále jen VŠKP nebo dílo)

Název VŠKP: LPC kódování řeči
Vedoucí/školicel VŠKP: Mgr. Pavel Rajmic, Ph.D.
Ústav: Ústav telekomunikací
Datum obhajoby VŠKP:

VŠKP odevzdal autor nabyvateli v:

- tištěné formě - počet exemplářů 1
- elektronické formě - počet exemplářů 1

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

Článek 2
Udělení licenčního oprávnění

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

Článek 3
Závěrečná ustanovení

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

V Brně dne:

.....

Nabyvatel

.....

Autor

ABSTRAKT

Náplní práce „LPC kódování řeči“ je studie problematiky této metody parametrického zdrojového kódování, vysvětlení matematických postupů, které se při ní užívají (lineární predikce, autokorelace, Levinson-Durbinův algoritmus, převod do formy vhodné pro přenos, Čebyševova polynomiální metoda hledání kořenů) a seznámení se s významem a aplikacemi této metody v reálných hovorových kodérech. Úkolem samostatného projektu této práce je popis a simulace jednoduchého řečového kodéru na bázi LPC, který převádí reálný řečový signál na bitový tok, ve kterém jsou obsaženy všechny důležité parametry pro jeho zpětnou rekonstrukci (LSF koeficienty, perioda základního tónu, úroveň buzení, detekce znělosti – metoda AMDF). Součástí práce je i diskuse o současně používaných řečových kodérech.

KLÍČOVÁ SLOVA

signál, řeč, kódování, lineární, predikce, autokorelace, formant, hlas, filtr, vokodér, spektrum, šum

ABSTRACT

The contents of the thesis "LPC speech coding" are studies of this method of a parametric source coding, explanation of mathematical procedures that are used in it (linear prediction, autocorrelation, Levinson-Durbin algorithm, transfer to a form suitable for transmission, Chebyshev root searching polynomial method) and acquaintance with the signification and application of that method in real speech encoders. The task of the original project of this thesis is a description and simulation of a simple speech encoder based on LPC, which transforms a real speech signal into a bit flow, which contains all of the significant parameters for its backward reconstruction (LSF coefficients, pitch period, excitation level, voice detection – AMDF method). One part of this thesis is a discussion about currently used speech encoders.

KEYWORDS

signal, speech, coding, linear, prediction, autocorrelation, formant, voice, filter, vocoder, spectrum, noise

ZAPLETAL, O. *LPC kódování řeči*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2008. 44 s. Vedoucí bakalářské práce Mgr. Pavel Rajmic, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „LPC kódování řeči“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

V Brně dne

.....

(podpis autora)

PODĚKOVÁNÍ

Děkuji vedoucímu této bakalářské práce Mgr. Pavlu Rajmicovi, Ph.D. za jeho ochotu shánět potřebné materiály pro tuto práci, za poskytnuté konzultace a za informace ohledně formální úpravy dokumentu.

V Brně dne

.....

(podpis autora)

OBSAH

Úvod	11
1 Lineární analýza a predikce signálu	12
1.1 Tvorba řeči a její model	12
1.2 Lineární predikce	14
1.2.1 ARMA modely	14
1.2.2 Obecný lineární prediktor	15
1.2.3 Minimalizace chyby lineární predikce	16
1.3 Výpočet LP koeficientů segmentu	17
1.3.1 Autokorelace a autokorelační matice segmentu	17
1.3.2 Výkon chyby predikce	17
1.3.3 Levinson-Durbinův algoritmus	18
1.4 Vliv LPC kódování na spektrum řečového signálu	20
1.5 Ukázka LP koeficientů a průběhů rekonstruovaných signálů	20
1.6 Přenos LP koeficientů ve formě spektrálních párů (LSP) či frekvencí (LSF)	22
2 Vokodéry	25
2.1 Typy vokodérů	25
2.2 LPC vokodér	27
3 Realizace LPC vokodéru	29
3.1 Rozbor činnosti jednotlivých částí simulovaného kodéru	29
3.2 Pravidla pro návrh dekodéru	34
3.3 Výsledky simulace	36
3.4 Možná vylepšení tohoto LPC kodéru	36
3.5 Porovnání se současně využívanými hovorovými kodéry	37
4 Závěr	39
Literatura	40
Seznam symbolů, veličin a zkratk	42
A CD se simulací řečového kodéru	44

SEZNAM OBRÁZKŮ

1.1	Tvorba řeči a její model. Obrázek převzat z [2]	12
1.2	Model pro tvorbu řečového signálu	13
1.3	Model obecného lineárního řečového prediktoru [2]	15
1.4	Závislost výkonu chyby lineární predikce na řádu predikce	18
1.5	Spektrum získané pomocí LPC	20
1.6	Časový průběhu původního a predikovaného signálu hlásky „e“ . . .	21
1.7	Časový průběhu původního a predikovaného signálu hlásky „es“ . . .	22
1.8	Poloha kořenů na jednotkové kružnici pro LSP	23
2.1	Blokové schéma vysílací části LPC vokodéru	27
2.2	Blokové schéma přijímací části LPC vokodéru	28
3.1	Kvantování posledních čtyřech LSF koeficientů	31
3.2	Průběh funkce AMDF pro znělý segment	32
3.3	Bitový rámec pro jeden segment	34

ÚVOD

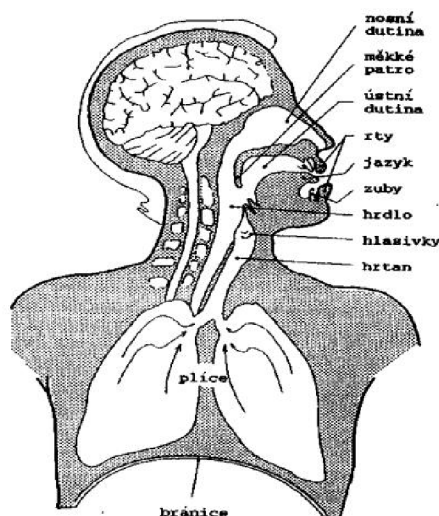
Lineární prediktivní kódování řeči (*Linear Predictive Coding* – LPC) je velmi účinná a dnes často používaná metoda pro kódování řečového signálu. Patří mezi parametrická zdrojová kódování, tedy ze signálu jsou získávány jeho charakteristické parametry, které nesou informaci potřebnou pro jeho zpětnou rekonstrukci. Vychází z toho, že následující vzorky signálu se dají určit ze vzorků předchozích dle lineární závislosti. Princip kódování spočívá v tom, že reálný řečový signál je navzorkován a rozdělen na krátké segmenty, ze kterých jsou následně vypočítány tzv. LP koeficienty, které pak pro onen signál reprezentují jeho kmitočtové vlastnosti. Tyto koeficienty pak slouží jako koeficienty rekonstrukčního filtru hlasového traktu na přijímací straně. U každého segmentu se zjišťuje více parametrů nezbytných pro jeho rekonstrukci, např. perioda základního tónu, které jsou společně s LP koeficienty vyslány skrze komunikační kanál k přijímači. Zařízení, které celou tuto analýzu provádí a pak signál zpětně rekonstruuje, se nazývá vokodér. Vokodérům je v této práci věnována celá kapitola.

Praktickým úkolem této práce je simulace kodéru na bázi LPC kódování. V programu Matlab budou simulovány všechny operace, které jsou se signálem prováděny za účelem získání parametrů pro jeho rekonstrukci. Projekt je zaměřen zejména na získání LP koeficientů a jejich převod na formu vhodnou pro kvantování a přenos kanálem – LSP (*Line Spectral Pairs*), LSF (*Line Spectral Frequencies*). Vstupem do tohoto kodéru bude krátká řečová nahrávka, výstupem pak bitový tok zakódovaných parametrů.

1 LINEÁRNÍ ANALÝZA A PREDIKCE SIGNÁLU

1.1 Tvorba řeči a její model

Pro pochopení samotného LPC kódování je nejprve nutné se seznámit s principem tvorby hlásek a modely jednotlivých partií lidského těla, které se na jejich tvorbě podílejí. Celá cesta tvorby řeči je obrázku 1.1:

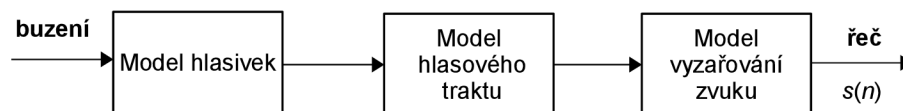


Obr. 1.1: Tvorba řeči a její model. Obrázek převzat z [2]

Princip tvorby hlásek spočívá v tom, že dýchací svaly (zejména bránice) vyženou z plic vzduch, který pak postupně prochází jednotlivými částmi hlasového traktu (hrtan, hlasivky, hrdlo, měkké patro, ústní a nosní dutina, jazyk, zuby, rty), přičemž každá tato část má jistý podíl na výsledném znění hlásky. Hlásky máme **znělé** (*a, e, í, o, u, y, b, d, g, v, z, . . .*) a **neznělé** (*p, t, k, f, s, c, ch, č, x, . . .*). U znělých hlásek se rozeznávají hlasivky, kdežto u neznělých nikoliv. Můžeme si to v praxi otestovat tak, že si při vyslovování znělé hlásky přiložíme ruku na krk do oblasti hlasivek a pak jasně ucítíme, že před samostatným vyslovením znělé hlásky se rozezní - začnou vibrovat. Vibrace vzniká tím, že pod tlakem proudu vzduchu vycházejícího z plic dochází k modulaci tohoto proudu hlasivkami. Kmitočet hlasivek je charakterizován základním tónem lidského hlasu, který má frekvenci o tzv. **pitch periodě**, což lze chápat jako kmitočet základního tónu, který nám tvoří základ znělých hlásek. Tento kmitočet se u každého člověka liší a určuje výšku hlasu jisté osoby. Úkolem kódování řeči je i to, aby byl tento kmitočet dodržen a abychom po následné rekonstrukci poznali „kdo to mluví“. Průběh signálu u neznělých hlásek pak zastupuje **bílý šum**

(tj. náhodný signál, který je nositelem stejného výkonu ve stejně širokých pásmech, tedy jeho spektrální hustota výkonu je konstantní v celém uvažovaném pásmu)[2].

Při LPC kódování je tvorba hlásek reprezentována modelem na obrázku 1.2.



Obr. 1.2: Model pro tvorbu řečového signálu

Celou tuto soustavu budeme nazývat **artikulační trakt**. Každá z částí tohoto modelu je reprezentována jako digitální filtr s jistou přenosovou funkcí. [2]

Model hlasivek:

Je reprezentován jako dolní propust druhého řádu s lomovou frekvencí okolo 100 Hz a přenosovou funkcí

$$G(z) = \frac{1}{[1 - e^{-cT_s} z^{-1}]^2}. \quad (1.1)$$

Model hlasového traktu:

Je reprezentován jako kaskáda k dvojpólových rezonátorů odpovídajících k formantům. **Formanty** jsou vyšší harmonické složky lidského hlasu vznikající při artikulaci řeči rezonancí dutiny hrdelní, ústní a nosní, jsou zásadními parametry, které dávají informaci o barvě hlasu konkrétního člověka.

$$V(z) = \frac{1}{\prod_{i=1}^K [1 - 2e^{-\alpha_i T_s} \cos \beta_i T_s z^{-1} + e^{-2\alpha_i T_s} z^{-2}]}, \quad (1.2)$$

kde F_i , B_i , α_i a β_i jsou parametry jednotlivých formantů.

Model vyzařování zvuku:

Je reprezentován horní propustí prvního řádu s přenosovou funkcí

$$L(z) = 1 - z^{-1}. \quad (1.3)$$

Celková přenosová funkce tohoto modelu pak odpovídá

$$\begin{aligned} H(z) &= G(z)L(z)V(z) = \\ &= \frac{1 - z^{-1}}{(1 - e^{-cT_s} z^{-1})^2 \prod_{i=1}^K [1 - 2e^{-\alpha_i T_s} \cos \beta_i T_s z^{-1} + e^{-2\alpha_i T_s} z^{-2}]}. \end{aligned} \quad (1.4)$$

Protože člen $cT_s \rightarrow 0$ [2], je možné tento poměrně složitý vztah zjednodušit a obecně shrnout celkovou přenosovou funkci celého modelu na

$$H(z) = \frac{1}{1 + \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)}, \quad (1.5)$$

což odpovídá čistému IIR filtru (s nekonečnou impulzní charakteristikou), protože obsahuje jen póly. Pro vyjádření $A(z)$ používáme polynom

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_p z^{-p}, \quad (1.6)$$

kde řád polynomu $p = 2k + 1$ je dán počtem formantů k . Nejčastěji volíme $p = 10$ při vzorkovací frekvenci $F_s = 8$ kHz. Znamená to, že jsou zjišťovány koeficienty deseti předchozích vzorků. S vyšší F_s pak bývá navyšován i p . [2]

1.2 Lineární predikce

Lineární predikce je využívána pro odhad následujících vzorků signálu ze vzorků předešlých. Realizuje se to tak, že se signál rozdělí na segmenty délky nejčastěji 25 ms a každý takový segment je následně analyzován. Výsledkem takovéto analýzy jsou koeficienty prediktoru, který pak umí další vzorky předpovědět. U lineární predikce předpokládáme, že každý vzorek je lineární funkcí vzorků předešlých, tedy pro p předešlých vzorků signálu $s(n)$ to vypadá takto:

$$s(n) = a_1 s(n-1) + a_2 s(n-2) + \dots + a_p s(n-p) - e(n), \quad (1.7)$$

kde a_1, a_2, \dots, a_p jsou koeficienty vzorků a $e(n)$ je chyba lineární predikce.[2]

1.2.1 ARMA modely

Lineární predikce vychází z tzv. ARMA modelů, které si nyní budeme specifikovat. Tyto modely se vyznačují velkou přesností predikce, ale je poměrně obtížné najít jejich parametry. ARMA vznikl sloučením dvou dílčích modelů MA a AR, které vypadají následovně:

MA ... model klouzavých průměrů; je to v podstatě lineární kombinace minulých chyb. Matematický zápis modelu MA řádu q vypadá takto:

$$\text{MA}(q)_t = b_0 - b_1 e_{t-1} - b_2 e_{t-2} - \dots - b_q e_{t-q} + e_t, \quad (1.8)$$

kde t je čas v jistém okamžiku, b_0, b_1, \dots, b_q jsou parametry modelu, $e_{t-1}, \dots, e_{t-q}, e_t$ je bílý šum – náhodná chyba, která do systému vstupuje. Filtr $\text{MA}(q)_t$ je FIR filtr řádu q , který obsahuje jen nuly a má přenosovou funkci

$$H(z) = B(z). \quad (1.9)$$

AR ... autoregresní model; dá se popsat jako lineární kombinace vlivů minulých hodnot. Matematický zápis modelu AR řádu p vypadá takto:

$$\text{AR}(p)_t = a_0 - a_1 y_{t-1} - a_2 y_{t-2} - \dots - a_p y_{t-p} + e_t, \quad (1.10)$$

kde a_0, a_1, \dots, a_p jsou autoregresní koeficienty a $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ jsou hodnoty minulých vzorků zkoumané veličiny, e_t je opět náhodná chyba. Filtr $\text{AR}(p)_t$ je IIR filtr řádu p , který obsahuje jen póly a má přenosovou funkci

$$H(z) = \frac{1}{A(z)}. \quad (1.11)$$

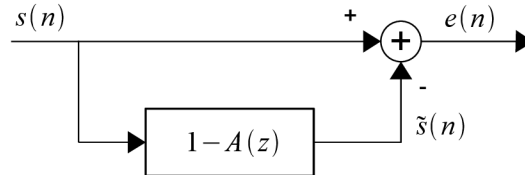
Ve výsledku je pak ARMA filtr IIR obsahující nuly i póly s přenosovou funkcí

$$H(z) = \frac{B(z)}{A(z)}. \quad (1.12)$$

ARMA modelu je využito jako výchozího modelu pro odhad vlastních LP koeficientů $a_i^{(p)}$. Pro jejich správné určení je operováno jak s předchozími vzorky signálu, tak s chybou, která je určena z chyb předchozích. [1]

1.2.2 Obecný lineární prediktor

Model obecného lineárního prediktoru je zakreslen na obrázku 1.3:



Obr. 1.3: Model obecného lineárního řečového prediktoru [2]

Signál $s(n)$ je analyzovaný řečový signál, $\tilde{s}(n)$ je predikovaná hodnota signálu a $e(n) = s(n) - \tilde{s}(n)$ je chyba predikce (chybový signál). Predikovaná hodnota signálu $\tilde{s}(n)$ je zde lineární kombinací p předchozích vzorků a je předpovědí skutečného vzorku $s(n)$:

$$\tilde{s}(n) = - \sum_{i=1}^p a_i s(n-i). \quad (1.13)$$

Vzniká zde ona chyba predikce $e(n)$ jako rozdíl mezi skutečnou a předpovězenou hodnotou signálu. Tu je potřeba mít co nejmenší a právě minimalizaci této chyby se věnuje následující kapitola. [2]

1.2.3 Minimalizace chyby lineární predikce

Vychází se ze vztahu pro tzv. nenormalizovanou chybu predikce[2]:

$$E = \sum_n e^2(n), \quad (1.14)$$

kde $e(n)$ je chybový signál vzniklý v důsledku predikce signálu. Vztah (1.14) je nuntné minimalizovat. My známe podobu signálu $s(n)$, neznáme ovšem velikosti koeficientů a_i , které budou naši proměnnou. Vyjádřeme si hodnotu chyby signálu pomocí těchto parametrů [2]:

$$e(n) = s(n) - \tilde{s}_n = s(n) + \sum_{i=1}^p a_i s(n-i). \quad (1.15)$$

Minimum nalezneme pomocí parciálních derivací funkce (1.14) podle každého a_i , které položíme rovny nule. Ve výsledku pak vznikne soustava rovnic (1.21), která nás bude posléze zajímat [2].

$$\frac{\partial}{\partial a_j} \left\{ \sum_n \left[s(n) + \sum_{i=1}^p a_i s(n-i) \right]^2 \right\} = 0 \quad (1.16)$$

$$\sum_n s(n) + \sum_{i=1}^p a_i s(n-i) \sum_n s(n-i) s(n-j) = 0 \quad (1.17)$$

$$\sum_n 2 \left[s(n) + \sum_{i=1}^p a_i s(n-i) \right] s(n-j) = 0. \quad (1.18)$$

Označme si

$$\sum_n s(n-i) s(n-j) = \phi(i, j), \quad (1.19)$$

pak

$$\sum_{i=1}^p a_i \phi(i, j) = -\phi(0, j) \quad \text{pro } 1 \leq j \leq p. \quad (1.20)$$

Vztah (1.20) pak odpovídá soustavě lineárních rovnic, která vypadá následovně:

$$\begin{aligned} \phi(1, 1)a_1 + \phi(2, 1)a_2 + \cdots + \phi(p, 1)a_p &= -\phi(0, 1) \\ \phi(1, 2)a_1 + \phi(2, 2)a_2 + \cdots + \phi(p, 2)a_p &= -\phi(0, 2) \\ &\vdots \\ \phi(1, p)a_1 + \phi(2, p)a_2 + \cdots + \phi(p, p)a_p &= -\phi(0, p) \end{aligned} \quad (1.21)$$

1.3 Výpočet LP koeficientů segmentu

1.3.1 Autokorelace a autokorelační matice segmentu

Důležitým krokem je určit koeficienty $\phi(i, j)$ ze soustavy (1.21) tak, aby měly reálný význam a byly co nejjednodušší. Nabízejí se dvě metody řešení:

- kovariační metoda
- korelační metoda.

Rozhodující je to, jak známe signál vně zkoumaného segmentu. U kovariační metody signál vně segmentu neznáme a řešení by vedlo k soustavě lineárních rovnic a výsledkem by byl nestabilní filtr – složité a bez smysluplného výsledku. Naopak u korelační metody známe signál vně segmentu – je nulový, tedy porovnávané vzorky mají vlastnost $\phi(i, j) = \phi(i + \text{konst.}, j + \text{konst.})$ a na diagonálách budou stejné hodnoty, což velmi usnadní výpočet. Protože je vždy k dispozici jen signál v jednom segmentu, provádíme autokorelaci (korelaci signálu se sebou samotným) a $\phi(i, j)$ budou tedy autokorelační koeficienty [2].

Autokorelace je operace se signálem, kdy zkoumáme podobnost signálu samotného se sebou. Signál posouváme o k jeho vzorků a hledáme *autokorelační koeficienty* $R(k)$ dle vztahu:

$$R(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} s(n)s(n+k), \quad (1.22)$$

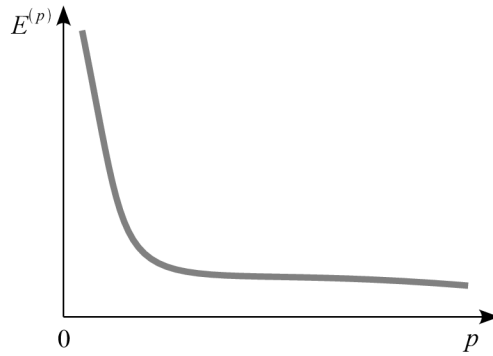
kde $R(k)$ je autokorelace v bodě k , $s(n)$ je analyzovaný signál a N je délka analyzovaného segmentu. Pro analyzovaný signál pak vznikne soustava rovnic, kde neznámé budou koeficienty a_1 až a_p a jejich známé koeficienty budou ony vypočtené autokorelační koeficienty $R(k)$ [2, 4]:

$$\begin{aligned} R(0)a_1 + R(1)a_2 + \cdots + R(p-1)a_p &= -R(1) \\ R(1)a_1 + R(0)a_2 + \cdots + R(p-2)a_p &= -R(2) \\ &\vdots \\ R(p-1)a_1 + R(p-2)a_2 + \cdots + R(0)a_p &= -R(p) \end{aligned} \quad (1.23)$$

1.3.2 Výkon chyby predikce

Výkon (energie) chyby lineární predikce je dán vztahem

$$E = \sum_{n=0}^{N+p-1} e^2(n) = R(0) + \sum_{n=0}^p a_i R(i). \quad (1.24)$$



Obr. 1.4: Závislost výkonu chyby lineární predikce na řádu predikce

Tohoto vztahu bude využito při aplikaci Levinson-Durbinovy metody pro hledání LPC koeficientů. Typický průběh závislosti výkonu chyby predikce E na velikosti řádu predikce p je vykreslen na obrázku 1.4. Je vidět, že na charakteristice je jasný ohyb. Vhodný řád prediktoru volíme právě v oblasti tohoto ohybu, protože při dalším zvyšování řádu prediktoru už výkon chyby predikce klesá jen velmi pozvolna a spíše je systém zatěžován více výpočty. Praktické určení takového optimálního řádu prediktoru bylo součástí semestrálního projektu v zimním semestru třetího ročníku.

Při vlastním kódování dojde vlivem výkonu chyby predikce ke snížení úrovně původního signálu $s(n)$, bude tedy potřeba ho zesílit, abychom se dostali na původní úroveň signálu. V případě, že budící signál má tzv. normovanou energii rovnu 1 (např. bílý šum s rozptylem 1), pak zesílení filtru pro dosažení požadované úrovně musí být [2]

$$G^2 = \frac{E}{N} = \frac{1}{N} \left[R(0) + \sum_{i=1}^p a_i R(i) \right]. \quad (1.25)$$

1.3.3 Levinson-Durbinův algoritmus

Matice autokorelačních koeficientů \mathbf{R} ze soustavy (1.23) je symetrická Töplitzova matice (všechny prvky na diagonálách jsou stejné), proto je možné využít k řešení této soustavy *Levinson-Durbinův algoritmus*. Tento algoritmus má následující zápis [2]:

$$E^{(0)} = R(0) \quad (1.26)$$

$$k_i = - \left[R(i) + \sum_{j=1}^{i-1} a_j^{i-1} R(i-j) \right] / E^{(i-1)} \quad (1.27)$$

$$a_i^{(i)} = k_i \quad (1.28)$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-j)} \dots \text{ pro } 1 \leq j \leq i-1 \quad (1.29)$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)}. \quad (1.30)$$

Aplikací tohoto algoritmu na soustavu (1.23) získáme soubor všech hledaných koeficientů $a_i = a_i^{(p)}$ tak, že postupně zvyšujeme řád prediktoru p . Pro přehled si tento soubor ukažme [2]:

$$\begin{array}{cccccc} a_1^{(1)} & a_1^{(2)} & a_1^{(3)} & \dots & a_1^{(p)} & \\ & a_2^{(2)} & a_2^{(3)} & \dots & a_2^{(p)} & \\ & & a_3^{(3)} & \dots & a_3^{(p)} & \\ & & & \ddots & \vdots & \\ & & & & a_p^{(p)} & \end{array} \quad (1.31)$$

Pro získání takovýchto koeficientů se postupuje v algoritmu po následujících krocích [2, 7]:

1. vypočteme $a_1^{(1)}$:

$$E_0 = R(0) \quad (1.32)$$

$$k_1 = -\frac{R(1)}{R(0)} \quad (1.33)$$

$$a_1^{(1)} = k_1 \quad (1.34)$$

$$E_1 = E_0(1 - k^2) \quad (1.35)$$

2. zvyšujeme řád: $m = 2, 3, \dots, p$:

$$k_m = -\frac{R(m) + \sum_{j=1}^{m-1} a_j^{(m-1)} R(m-j)}{E_{m-1}} \quad (1.36)$$

$$a_m^{(m)} = k_m \quad (1.37)$$

$$a_j^{(m)} = a_j^{(m-1)} + k_m a_{m-j}^{(m-1)}, \quad j = 1, 2, \dots, m-1 \quad (1.38)$$

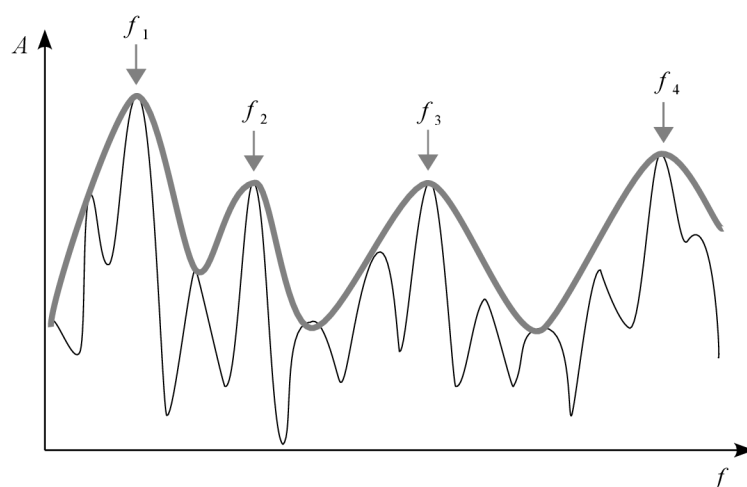
$$E_m = E_{m-1}(1 - k_m^2) \quad (1.39)$$

3. sestavíme matici koeficientů $a_i = a_i^{(p)}$, $i = 1, 2, \dots, p$, ($a_0 = 1$).

Tyto zjištěné koeficienty určují přenosovou funkci filtru hlasového traktu (1.6). Mají v sobě zakódovanou informaci o formantových frekvencích daného segmentu řeči. Jejich využití pro rezezenraci řeči pro přenos kanálem způsobuje až stonásobné zmenšení objemu přenášených dat oproti přenosu samotných vzorků signálu.

1.4 Vliv LPC kódování na spektrum řečového signálu

Jak bylo již zmíněno, úkolem lineární analýzy je najít formantové frekvence řečového signálu. Počet takto nalezených frekvencí je stejný jako řád prediktoru p . Na obrázku 1.5 je vykresleno jak to může vypadat. Spektrum lidské řeči obsahuje spoustu frekvenčních složek a proto je tolik zvlněno. Lidské ucho ovšem není natolik citlivé, aby se muselo spektrum řeči určovat co nejpřesněji. LPC analýza vytvoří spektru jakousi jeho obálku obsahující ty nejnütnější formantové frekvence – zobrazí se jen spektrální vrcholky (f_1 až f_4). [2, 5]



Obr. 1.5: Spektrum získané pomocí LPC

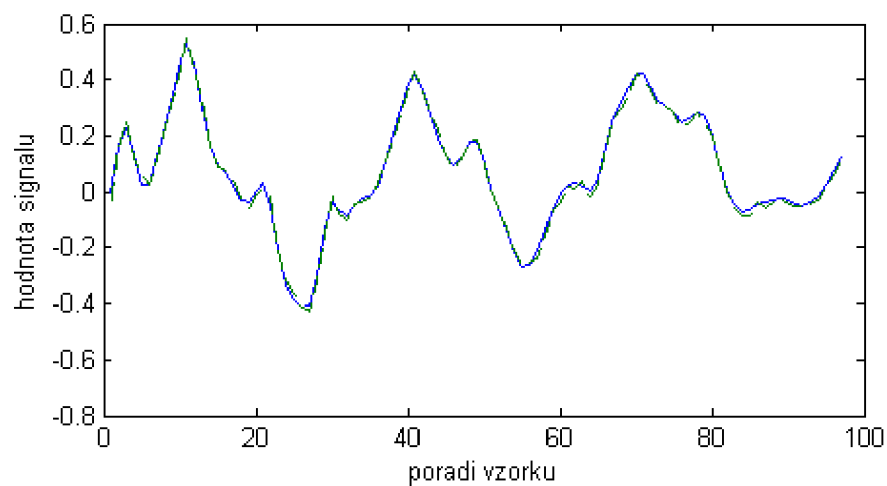
1.5 Ukázka LP koeficientů a průběhů rekonstruovaných signálů

Zde je ukázka experimentálně určených LP koeficientů a hodnot výkonu predikce E pro kódování hlásky „e“ při užití řádu predikce od 1 do 12 (tabulka 1.1). Všimněme si také vývoje hodnoty výkonu predikce na použitém řádu predikce (E klesá s rostoucím p , viz. obrázek 1.5). Hodnoty byly změřeny v rámci zpracování semestrálního projektu LPC kódování řeči užitím funkce `lpc` v prostředí Matlab.

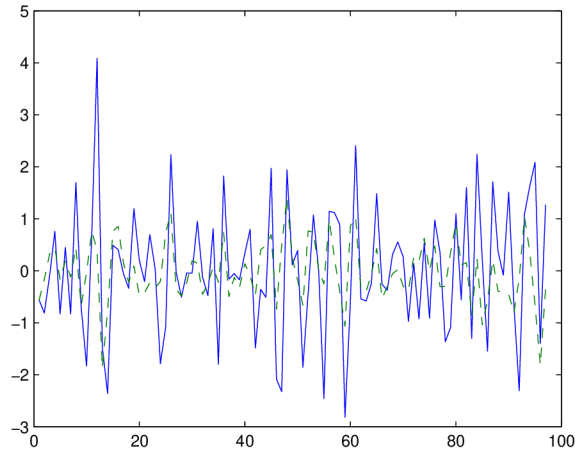
Dále jsou zde ukázky průběhů původního i rekonstruovaného signálu pro onu hlásku „e“ (obrázek 1.6) a pro neznělou hlásku „es“ (obrázek 1.7). Jsou zde patrné rozdíly v průběhu obou hlásek – znělá hlásky má téměř periodický průběh a neznělá má charakter náhodného šumu a velký počet průchodů nulou. Byl použit řád predikce 4.

Tab. 1.1: LP koeficienty a výkon chyby predikce

řád LPC	ukázka prvních 5 LP koeficientů	E
1	-0,9416	48,4913
2	-1,5763 0,6741	26,4581
3	-1,9581 1,5670 -0,5665	17,9679
4	-2,3413 2,6270 -1,8911 0,6765	9,7452
5	-2,7440 3,7526 -3,4548 2,0701 -0,5952	6,9290
6	-3,0492 4,8140 -5,2261 3,9941 -2,0021	4,6372
7	-3,0571 4,8449 -5,2877 4,0748 -2,0764	4,6361
8	-3,0613 4,9981 -5,8562 5,1903 -3,5240	4,2881
9	-3,0200 4,8695 -5,5715 4,6584 -2,7406	4,1907
10	-2,9776 4,8167 -5,5989 4,9397 -3,5096	3,8603
11	-2,9652 4,7858 -5,5468 4,8661 -3,4073	3,8528
12	-2,9561 4,8163 -5,6459 5,0569 -3,7033	3,6901



Obr. 1.6: Časový průběhu původního a predikovaného signálu hlásky „e“



Obr. 1.7: Časový průběhu původního a predikovaného signálu hlásky „es“

1.6 Přenos LP koeficientů ve formě spektrálních párů (LSP) či frekvencí (LSF)

V komunikačních systémech jsou jen málokdy přenášeny přímo LP koeficienty. Tyto koeficienty se špatně kvantují, přičemž je třeba je přenášet s velkou přesností. Bylo tedy třeba najít nějakou vhodnou alternativu pro jednodušší přenos LP koeficientů. Technologie LSP (*Line Spectral Pairs*) se objevila v 80. letech 20. století, kdy se objevila možnost přenášet koeficienty ve formě čárových spektrálních párů. Hlavním důvodem je snadnější kvantování těchto koeficientů. Byla vynalezena jako reprezentace inverzního filtru v rovině z [4]:

$$A(z) = 1 - \sum_{i=1}^p a_i z^{-i}. \quad (1.40)$$

Tento výtah byl rozložen do dvou polynomů řádu $p + 1$:

$$P(z) = A(z) + z^{-(p+1)}A(z^{-1}) \quad (1.41)$$

$$Q(z) = A(z) - z^{-(p+1)}A(z^{-1}), \quad (1.42)$$

$A(z)$ se zpětně vypočítá jako aritmetický průměr hodnot $P(z)$ a $Q(z)$ [4]:

$$A(z) = \frac{P(z) + Q(z)}{2}. \quad (1.43)$$

$P(z)$ je symetrický a $Q(z)$ je antisymetrický polynom. Toho se využívá při výpočtu kořenů těchto polynomů. $P(z)$ má vždy kořen $z = -1$ ($\omega = \pi$) a $Q(z)$ pak kořen $z = 1$ ($\omega = 0$). Tyto kořeny ale nebudeme potřebovat, protože v praxi znamenají nulovou a maximální frekvenci v uvažovaném spektru, což v našem případě

nemá smysl. Zbavíme se jich takovým způsobem, že vytvoříme nové polynomy, které vzniknou z původních takto:

$$P'(z) = \frac{P(z)}{1+z^{-1}} \quad (1.44)$$

$$Q'(z) = \frac{Q(z)}{1-z^{-1}}. \quad (1.45)$$

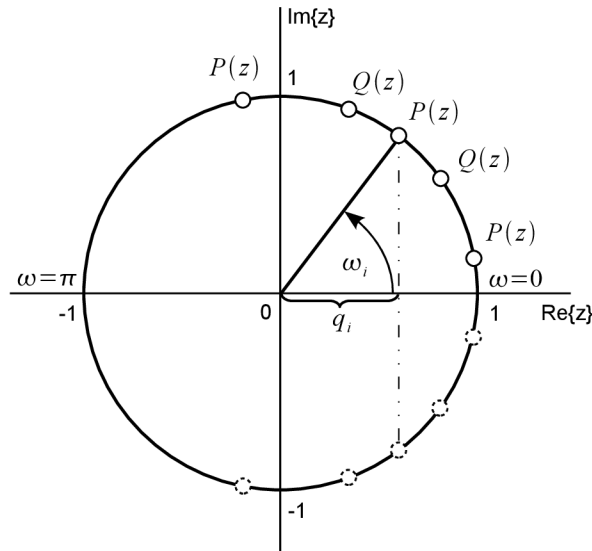
Tyto polynomy se dají rozložit na součin tak, že je potom snadné spočítat jejich kořeny.

$$P'(z) = \prod_{i=1,3,\dots,p-1} (1 - 2q_i z^{-1} + z^{-2}) \quad (1.46)$$

$$Q'(z) = \prod_{i=2,4,\dots,p} (1 - 2q_i z^{-1} + z^{-2}) \quad (1.47)$$

Kořeny q_i se nazývají LSP koeficienty nabývají hodnot od -1 do 1 . Jejich vztah k LSF (*Line spectral Frequencies*) koeficientům (frekvencím ω_i) je $q_i = \cos(\omega_i)$. Platí pro ně $0 < \omega_1 < \omega_2 < \dots < \omega_p < \pi$. [7]

Pro každé q_i jsou uvažovány dva komplexně sdružené kořeny (nuly) ležící na jednotkové kružnici v rovině z , nás bude zajímat jen její horní polovina, kde poloha nuly reprezentuje určitou frekvenci v rozsahu $< 0; \pi > = < 0; \frac{f_{vz}}{2} >$. Dochází zde k postupnému střídání kořenů z polynomů $P(z)$ a $Q(z)$. Celá situace je přehledně znázorněna na obrázku 1.8 [5].



Obr. 1.8: Poloha kořenů na jednotkové kružnici pro LSP

Získání LSP koeficientů q_i z LP koeficientů je poměrně složitá záležitost. Řeší se následujícím postupem [7]:

Abychom se vyvarovali obecných a nepřehledných vztahů, budeme v dalších výpočtech pracovat s řádem predikce $p = 10$, který se uvažuje u většiny dnešních kodérů a uijeme ho i při realizaci našeho kodéru. Polynomy $P'(z)$ a $Q'(z)$ jsou symetrické a stačí určit jen prvních pět tzv. koeficientů polynomu pomocí rekurzivních vztahů

$$\begin{aligned} \text{pro } P'(z): \quad f_1(i+1) &= a_{1+i} + a_{10-i} - f_1(i), \quad i = 0, \dots, 4, \\ \text{pro } Q'(z): \quad f_2(i+1) &= a_{1+i} - a_{10-i} + f_2(i), \quad i = 0, \dots, 4, \end{aligned} \quad (1.48)$$

přičemž $f_1(0) = f_2(0) = 1$. Pro další práci je potřeba definovat polynom $F(\omega)$, který vznikne z polynomů $P'(z)$ a $Q'(z)$ po substituci $z = e^{j\omega}$ a užívá se při Čebyševově polynomiální metodě pro nalezení LSP koeficientů:

$$F(\omega) = 2e^{-j5\omega}C(x), \quad (1.49)$$

kde

$$C(x) = T_5(x) + f(1)T_4(x) + f(2)T_3(x) + f(3)T_2(x) + f(4)T_1(x) + \frac{f(5)}{2}, \quad (1.50)$$

kde $T_m(x) = \cos(m\omega)$ je m -tý řád Čebyševova polynomu a $f(i), i = 1, \dots, 5$ jsou koeficienty obou polynomů ze vztahů (1.48). Čebyševova polynomiální metoda vyhledávání kořenů pak spočívá v tom, že pro postupně se zvyšující ω od 0 do π po velmi jemných krocích se zkoumá znaménko polynomu $F(\omega)$ a v bodech jeho změny se nachází kořeny ω_i . Každý z polynomů $P'(z)$, $Q'(z)$ má na tomto intervalu 9 kořenů, tedy celkem získáme 18 hodnot. Vlastností těchto polynomů je, že 4 kořeny jsou shodné pro oba polynomy, které nemá smysl uvažovat kvůli jejich vzájemnému vyrušení při zpětném výpočtu LP koeficientů dle vztahu (1.43). Po vyřazení těchto čtyř dvojic shodných kořenů získáváme 10 výsledných kořenů, LSF koeficientů ω_i . Ve výsledném vektoru je seřazujeme dle velikosti od nejmenšího po největší, jejich matematický význam je formantová f_i frekvence, která se zjistí dle vztahu

$$f_i = \frac{\omega_i}{2\pi} f_{vz}. \quad (1.51)$$

Tyto LSF koeficienty se nadále kvantují a kódují pro přenos, jsou nositeli informace o parametrech filtru hlasového traktu v dekodéru.[7]

2 VOKODÉRY

Vokodéry, nebo-li zdrojové hovorové kodéry, jsou zařízení, která provádí kompletní kódování řeči se všemi jejími parametry. V dnešní době je prakticky ve všech typech vokodérů základem LPC kodér, který má za úkol určit parametry barvy hlasu. Ostatní jeho části zjišťují např. periodu základního tónu T_0 , úroveň buzení G , detekují znělou–neznělou hlásku či mezeru, kvantují a multiplexují hodnoty jednotlivých parametrů atd.

Vokodér se skládá z vysílací a přijímací části. Do vysílací části (kodéru) vstupuje řečový signál, který je následně navzorkován, rozsegmentován a z jednotlivých segmentů jsou dále zjišťovány všechny parametry, které se na jeho výstupu multiplexují do bitového toku, který je následně upraven kanálovým kodérem pro přenos a vpuštěn do komunikačního kanálu. Na druhé straně je přijímací část (dekodér), který má za úkol příchozí bitový tok z vysílače dekodovat a ze zjištěných parametrů opět rekonstruovat onen řečový signál. Rekonstruovaná řeč má syntetický charakter, protože nevzniká obnovou původního signálu, ale syntézou z jednotlivých parametrů signálu. Srozumitelnost dosahuje 80 až 85 % při přenosové rychlosti okolo 4 kbit/s. Nejčastěji se vokodéry užívají se pro kódování telefonních hovorů a řečových záznamů. Nejdou použít pro obecné zvukové signály jako např. hudební záznamy, protože ty obsahují nejen řeč a tato metoda by v tomto případě nebyla účinná.[5]

2.1 Typy vokodérů

V současné době existuje celá řada hovorových kodérů a jejich modifikací. Pro různé výstupní bitové rychlosti se dosahuje různé kvality kódování a užívá se různých pomocných metod. Vokodéry jsou rozděleny do následujících skupin [5]:

LOW RATE vokodéry Jde o vokodéry s výstupní bitovou rychlostí do 2,4 kbit/s. Jsou to ty nejjednodušší kodéry pracující na základě mála vlastností řeči. Patří mezi ně:

- Formantové vokodéry – k rekonstrukci řeči užívají formantové frekvence
- Kanálové vokodéry – kodéry obsahují banku filtrů, skrze kterou je signál rozdělen do několika kmitočtových pásem a pak se pracuje se signálem v každém pásmu (kanále) zvlášť
- Kepstrální vokodéry – užívají vlastností kepstrální funkce signálu, ta se získává pomocí DFT

- LPC vokodéry – užívají LPC analýzy, rekonstrukce pomocí IIR filtru

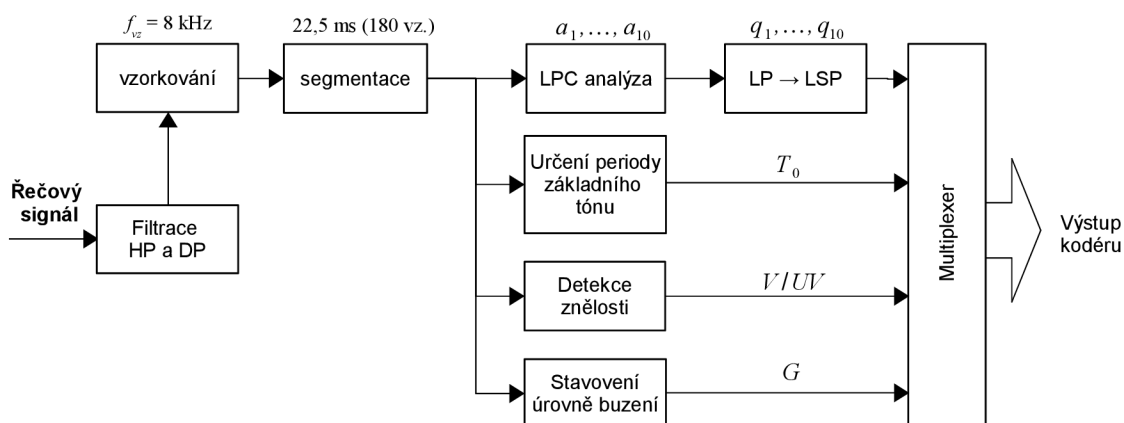
MEDIUM RATE vokodéry Sem patří hlavně vokodéry založené na LPC kombinované s nějakou další metodou pro zvýšení kvality. Výstupní bitové rychlosti se pohybují od 2,4 do 12 kbit/s. Uvedme si některé jejich typy podle názvu metody, kterou používají[5]:

- CELP (*Code-Excited Linear Prediction*) vokodéry – užívají vektorové kvantizace koeficientů (rozdílové vektory slouží jako ukazatele na kombinace parametrů uložených v kódové knize)
- RELP (*Residual Excited LP*) vokodéry – starší verze CELP, přenáší se přímo zbytkový signál, dnes se téměř nepoužívá
- VSELP (*Vector Sum Excited LP*) vokodéry – užívá dvou kódových knih, databáze pro vnámavostní váhování chyby, aktualizace LP koeficientů po 5 ms
- Multipulzní LPC vokodéry – obsahuje multipulzní budící generátor, funguje za základě analýza-syntéza

HIGH RATE vokodéry V tomto případě se jedná již o hovorové kodéry, které kódují řeč ve vysoké kvalitě za cenu větších nároků na přenosovou rychlost, ale s méně náročnou kompresní metodou. Výstupní bitové rychlosti se pohybují nejčastěji mezi od 12 a 32 kbit/s [5].

- RPE-LTP vokodéry – zde je základem LPC analýza kombinovaná s regulárním impulzním buzením a dlouhodobou predikcí, užívají se v síti GSM, 13 kbit/s [6]
- CVSD (*Continuously Variable Slope Delta Modulation*) vokodéry – užívá delta modulaci upravenou pro kódování řeči (s rozdílnou velikostí kroku), 9,6 až 128 kbit/s
- ADPCM (*Adaptive Differential Pulse Code Modulation*) – neopírá se o vlastnosti řeči, v tomto případě jde o účinnou kódovací metodu vzorkovaného signálu, 32 kbit/s

Pro srovnání je bitová rychlost řečového signálu navzorkovaného kmitočtem 8 kHz 104 kbit/s. Je tedy vidět obrovské snížení objemu potřebných dat pro přenos řečového signálu při zachování akceptovatelné kvality.



Obr. 2.1: Blokové schéma vysílací části LPC vokodéru

2.2 LPC vokodér

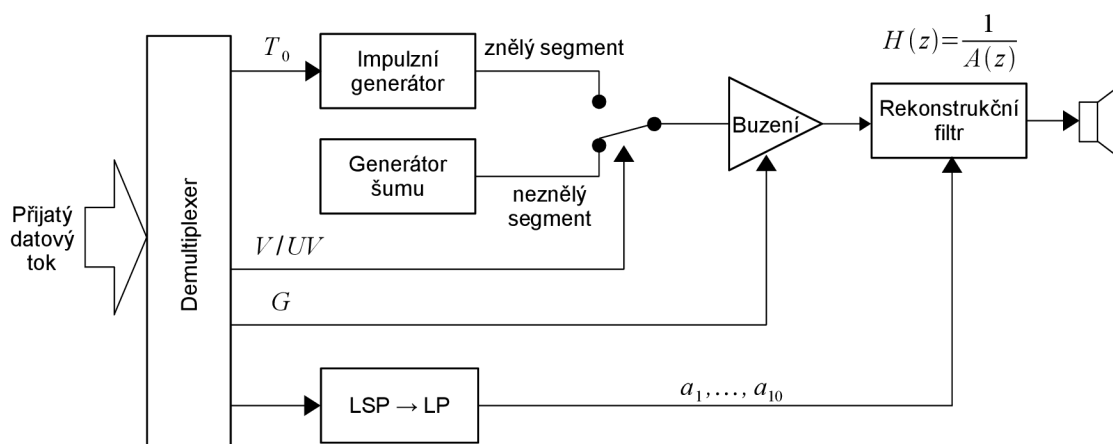
Pro účely této práce je nejnmutnější seznámit se se základním vokodérem na bázi LPC. Blokové schéma jeho kodéru je na obrázku 2.1 [6]. Jeho funkce je následující:

Do kodéru vstupuje reálný řečový signál, např. z mikrfonu. Tento signál je třeba digitalizovat, proto je navzorkován kmitočtem 8 kHz. Uvažujeme tedy, že se jedná o kodér pro telefonní hovory. Dále je signál segmentován po 180 vzorcích, což odpovídá délce segmentu 22,5 ms [4]. Poté se pro každý takový segment provádí analýza jistého jeho parametru. Jsou to:

- **Stanovení LP koeficientů** – pomocí autokorelace segmentu a aplikace Levinson-Durbinova algoritmu dostaneme LP koeficienty a_1, \dots, a_{10} . Užívá se řád prediktoru $p = 10$, stejně jako drtivě většině ostatních vokodérů na bázi LPC. Tyto koeficienty jsou dále převedeny na LSP(LSF) koeficienty.
- **Zjištění periody základního tónu T_0 (pitch)** – provádí se nejčastěji autokorelační metodou nebo keprální analýzou.
- **Detekce znělosti** – tzv. V/UV bit, který určuje o jaký druh segmentu se jedná. Metod pro určení znělosti je spousta a jsou většinou složité. Nejjednodušší způsob je zjištění počtu průchodů signálu nulou v rámci jednoho segmentu.
- **Zjištění úrovně buzení G** – v tomto případě se určí jednoduše pomocí výpočtu efektivní hodnoty signálu v segmentu.

Výstupní bitová rychlost závisí hlavně na počtu kvantovacích bitů pro LSP(LSF) koeficienty. Nejběžnější rychlost je v tomto případě 2,4 kbit/s.

Aby byl tento LPC vokodér kompletní, je třeba si ještě popsat a zobrazit podobu jeho dekodéru – obrázek 2.2 [6].



Obr. 2.2: Blokové schéma přijímací části LPC vokodéru

Příchozí bitový tok je demultiplexován a jsou získány všechny parametry z kodéru. Nejdůležitější je rozpoznání znělosti-neznělosti. Podle hodnoty V/UV bitu je přepínán spínač mezi impulzním generátorem, který v případě znělé hlásky vysílá impulzy vzdálené od sebe o T_0 a šumovým generátorem, který je zdrojem náhodného (bílého) šumu v případě neznělé hlásky. Dále je signál zesílen buzením o velikosti G a prochází rekonstrukčním filtrem o přenosové funkci $H(z)$ (dle vztahu 1.6), kde se pro každý segment mění $A(z)$ podle LP koeficientů, které jsou vypočteny z přijatých LSP (LSF) koeficientů. Filtr dodá zvuku potřebné formantové frekvence a určí tím jeho barvu. [6]

3 REALIZACE LPC VOKODÉRU

Hlavním cílem této práce je simulace řetězce zpracování řečového signálu od vzorkování až po vyslání po přenosovém kanále. Simulována bude práce vysílčí části jednoduchého LPC vokodéru dle schématu na obrázku 2.1. Vstupními daty budou krátké řečové nahrávky ve formátu `.wav` a výstupem bude bitový tok, který bude obsahovat zakódované všechny důležité parametry signálu, především LP koeficienty převedené do LSF. K realizaci bude použit program Matlab, verze 7.5.0.342 (R2007b).

3.1 Rozbor činnosti jednotlivých částí simulovaného kodéru

FILTRACE HP A DP

Vstupní zvukový signál načteme pomocí funkce `wavread`. Výstupem bude posloupnost vzorků s hodnotami od -1 do 1 nakvantovaných lineárně s přesností na čtyři desetinná místa. Načtený zvuk může obsahovat nežádoucí nízkofrekvenční složky, které je třeba pro další analýzu odstranit. Užijeme k tomu horní propust (HP) s mezní frekvencí 60 Hz. Protože bude použita vzorkovací frekvence $f_{vz} = 8$ kHz, je nutné mít splněnou Nyquistovu podmínku pro vstupní zvukový signál, proto jej přefiltrujeme dolní propustí s mezní frekvencí 4 kHz. Kombinaci HP a DP místo pásmové propusti (PP) užijeme v této simulaci proto, že PP s takovými parametry by měla několikanásobně vyšší řád.

VZORKOVÁNÍ

Přefiltrovaný zvukový signál bude převzorkován z původní vzorkovací frekvence na 8 kHz.

SEGMENTACE

Posloupnost vzorků se rozdělí na segmenty po 180 vzorcích, což odpovídá době trvání segmentu $22,5$ ms. Tento údaj byl vzat z popisu LPC-10 vokodéru [4].

LPC ANALÝZA

Úkolem LPC analýzy je stanovit koeficienty a_1, \dots, a_{10} rekonstrukčního filtru. Pro každý segment bude provedena autokorelace dle vzathu (1.22) a získané autokorelační koeficienty budou užity pro Levinson-Durbinův algoritmus (1.30) pro vyřešení matice (1.23), ze které získáme ony LP koeficienty.

PŘEVOD LP KOEFICIENTŮ NA LSF

Pro lepší kvantování je potřeba přepočítat LP koeficienty na LSP či LSF. Pro tuto

simulaci jsem vybral LSF, protože se dají snadněji kvantovat lineárně v určitých intervalech. Nalezeny budou pomocí aplikace Čebyševovy polynomiální metody pro hledání kořenů při užití 1024 kroků mezi 0 a π . Výsledných deset LSF koeficientů je potřeba nakvantovat. Nejúčinnější je vektorová kvantizace, kdy se určí vektor rozdílů koeficientů, který je ukazatelem na jisté místo v kódové knize. Tu ovšem k dispozici nemáme a navíc se používá až u CELP vokodérů [10]. Je tedy třeba navrhnout způsob kvantování, který bude co možná nejúčinnější pro nakvantování všech LSF koeficientů do vyhrazených 60 bitů. Vyjděme z následujících faktů:

- koeficienty reprezentují lineárně rozložené frekvence, tedy čím jsou nižší, je třeba je kvantovat přesněji (opírá se o vlastnosti vnímání zvuku – logaritmicky, stejný tón větší výšky má dvojnásobnou frekvenci)
- experimentálně zjištěné maximální rozdíly mezi dvěma sousedními LSF koeficienty byly vždy menší než 1, tedy uvažujme maximální možný rozdíl $\pi/2$
- poslední čtyři koeficienty měly vždy hodnoty v intervalu $< \pi/2; \pi >$.

Po uvědomění si těchto faktů jsem navrhnul způsob kvantování LSF koeficientů popsany v tabulce 3.1.

Tab. 3.1: Kvantování LSF koeficientů

LSF koef.	bitů	co se kvantuje:
ω_1	8	jeho vlastní hodnota (max. $\pi/2$)
ω_2	8	vzdálenost od předchozího koeficientu (max. $\pi/2$)
ω_3	7	vzdálenost od předchozího koeficientu (max. $\pi/2$)
ω_4	7	vzdálenost od předchozího koeficientu (max. $\pi/2$)
ω_5	6	vzdálenost od předchozího koeficientu (max. $\pi/2$)
ω_6	6	vzdálenost od předchozího koeficientu (max. $\pi/2$)
ω_7	5	poměr d_{ω_i}/d_π
ω_8	5	poměr d_{ω_i}/d_π
ω_9	4	poměr d_{ω_i}/d_π
ω_{10}	4	poměr d_{ω_i}/d_π
celkem tedy 60 bitů		

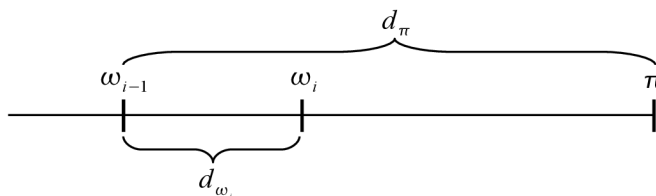
K tabulce 3.1 bych ještě doplnil pár důležitých poznámek:

Omezením maximálního možného rozdílu dvou sousedních koeficientů na $\pi/2$ získáme „bit k dobru“ a počet kvantizačních kroků se tím zdvojnásobí. Například ω_1 , je kvantován osmi bity, tedy jeden krok je původně $\frac{1}{2^8}\pi = \frac{1}{256}\pi$ a po uvážení výše uvedené skutečnosti ho zjemníme na $\frac{1}{512}\pi$. Dopouštíme se tedy maximální chyby

$\pm \frac{1}{1024}\pi$. Až do nakvantování šestého bitu se tedy můžeme dopustit maximální chyby

$$\left(\frac{1}{1024} + \frac{1}{1024} + \frac{1}{512} + \frac{1}{512} + \frac{1}{256} + \frac{1}{256} \right) \pi = \frac{7}{512}\pi \quad (3.1)$$

$\doteq 1,37\%$ z π , což odpovídá ± 55 Hz. Pro ω_6 uvažujeme frekvence přes 2000 Hz, kde lidský sluch je již málo citlivý na změny v řádech desítek Hz (max. 55 Hz). Aby nedošlo při sečítání rodílů mezi frekvencemi k překročení hodnoty π , je třeba u posledních čtyř koeficientů „brzdit“ pomocí kvantování poměru vzdálenosti od předchozího koeficientu d_{ω_i} a vzdálenosti od předchozího koeficientu k π d_π jak je znázorněno na obrázku 3.1, kde vzdálenost d_π reprezentuje 2^n kroků, což udá měřítko pro nakvantování d_{ω_i} pro příslušné ω_i .



Obr. 3.1: Kvantování posledních čtyřech LSF koeficientů

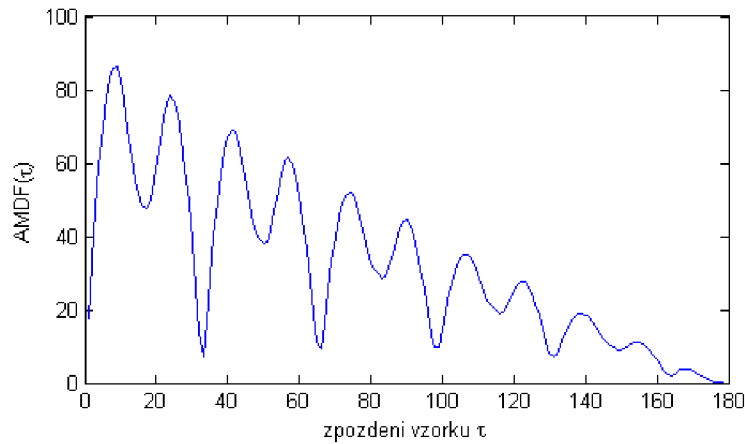
ZJIŠTĚNÍ PERIODY ZÁKLADNÍHO TÓNU

Pro zjištění T_0 byla vybrána jako nejvhodnější metoda AMDF (*Average Magnitude Difference Function*), kde se užívá vztahu [13]

$$AMDF(\tau) = \sum_{n=0}^{N-1} |s(n) - s(n + \tau)|, \quad (3.2)$$

kde $AMDF(\tau)$ je průměrný rozdíl funkce AMDF v bodě τ , N je délka segmentu (počet vzorků) a τ je zpoždění segmentu. Následně je potřeba nalézt takové zpoždění, při kterém je $AMDF(\tau)$ minimální (pro které zpoždění je signál nejpodobnější). Hledané T_0 je potom nejmenší takové τ . Protože τ bude v tomto případě počet vzorků, na dobu periody ho převedeme dle vztahu $T_0 = \frac{\tau}{f_{vz}} = \frac{\tau}{8000}$ s. Typický průběh funkce $AMDF(\tau)$ je vykreslen na obrázku 3.2. Hledaná perioda T_0 pak odpovídá vzdálenosti těchto prvního takového minima od nuly a dále se vyskytují tato minima po stejných intervalech.

Pro správné kvantování hodnoty T_0 bylo třeba nejprve zjistit, v jakých mezích se tato hodnota pohybuje. Po testování přibližně třiceti různých hlasových nahrávek (s různými hláskami od mužů i žen) jsme zjistili rozsah T_0 od 80 Hz do 400 Hz po přidání patřičné tolerance. Odpovídá to zpožděním vzorků o 20 až 100 τ . Algoritmus pro vyhledávání těchto koeficientů, který jsem testováním prověřil za účinný, stojí na dvou pravidlech:



Obr. 3.2: Průběh funkce AMDF pro znělý segment

- Minima jsou hledána ve stejně širokých intervalech o šířce 20τ , protože se mohou opakovat a nás zajímá první takové minimum
- „Pravidlo 50%“: Každé takové minimum má hodnotu maximálně 50% hodnoty aritmetického průměru hodnot předešlých funkce $AMDF(\tau)$
- „Pravidlo 1,8“: Kolem každého minima vyovujícího podmínce 50% je prohledáváno, zde se nenachází v okolí širokém $1,8\tau$ minimum, které je 1,8-krát menší než ono nalezené minimum, což by znamenalo, že minimum nalezené v předchozí podmínce nebude platné pro určení T_0 .

Výše uvedené tři podmínky jsou výsledkem experimentování, jehož cílem pro mne bylo nalezení vhodného algoritmu pro spolehlivou detekci periody základního tónu. V případě změny některého z nich (např. zvýšení z 50 na 60%) již tato metoda detekovala nesmyslná minima při evidentně neznělém segmentu nebo jiné minimum na vyšší frekvenci, která odpovídá vyššímu kmitočtu obsaženém v signálu ale není kmitočtem základního tónu.

Kvantování T_0 bude provedeno lineárně šesti bity. Číslo, které se bude převádět na binární dostaneme dle vztahu

$$T_{0_bity} = \text{int} \left(\frac{\tau - \tau_{min}}{\tau_{max} - \tau_{min}} \cdot 63 \right) = \text{int} \left(\frac{\tau - 20}{80} \cdot 63 \right). \quad (3.3)$$

kde „int“ znamená *integer*, čímž je myšlena celčíselná část výrazu v závorce.

ZJIŠTĚNÍ ÚROVNĚ BUZENÍ

Zjišťována je efektivní hodnota signálu v každém segmentu dle vztahu

$$G = \sqrt{\frac{1}{N} \sum_{n=1}^N s^2(n)}. \quad (3.4)$$

Výsledek je kvantován pěti bity. Protože signál nabývá hodnot od -1 do 1 , bude jeho efektivní hodnota ležet v intervalu $\langle 0; 1 \rangle$. Protože výsledky simulace ukázaly, že hodnota G jen málokdy přesáhne $0,5$, je třeba tomu přizpůsobit i kvantování. Bylo také třeba zvolit jemnější měřítko pro nejmenší hodnoty, protože lineární rozdělení po celém intervalu by znamenalo zatišení některých smysluplných segmentů. Kvantovat tedy budeme způsobem uvedeným v taulce, kde je G kvantována ve třech intervalech, v každém z nich lineárně.

Tab. 3.2: Kvantování úrovně buzení signálu

G	úrovní	binárně	vztah pro výpočet bin. čísla
$\langle 0,5; 1 \rangle$	8	11000 až 11111	$\text{int}((G - 0,5) \cdot 16) + 24$
$\langle 0,25; 0,5 \rangle$	8	10000 až 10111	$\text{int}((G - 0,25) \cdot 32) + 16$
$\langle 0; 0,25 \rangle$	16	00000 až 01111	$\text{int}(G \cdot 64)$
celkem 32 úrovní (5 bitů)			

DETEKCE ZNĚLOSTI

V tomto případě budou zkoumány vlastosti tří parametrů:

- Detekovatelnost pitch periody

Metoda AMDF pro detekci pitch periody buď detekovala nebo nedetekovala T_0 . Podle toho je hláska znělá či neznělá.

- Normalizovaný autokorelační koeficient signálu v segmentu[4]:

Ten se určuje dle vztahu

$$r_1 = \frac{\sum_{n=1}^N s(n)s(n-1)}{\sum_{n=1}^N y^2(n)}. \quad (3.5)$$

Experimentálně bylo zjištěno, že každý segment s $r_1 < 0,25$ je neznělý (vypovídá to o tom, že se příliš neshoduje sám se sebou).

- Počet průchodů signálu „nulou“[4]:

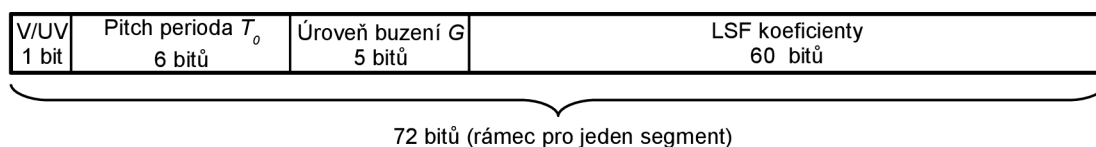
„Nulou“ je v tomto případě myšlen medián z hodnot signálu v segmentu. Segment bude označen jako neznělý v případě, že počet takových průchodů bude větší než 40% z maximálního možného počtu průchodůpočet vzorků v segmentu. Daná mez byla opět zjištěna experimentálně.

Pro každý segment budou testovány pro všechny tři uvedené parametry a každému z nich bude přiřazena určitá váha. Není-li detekována T_0 , pak je automaticky T_0 nastavena na 0 a o neznělosti segmentu pak není pochyb (váha 50%). Pro splnění

podmínek pro neznělost má koeficient r_1 váhu 20 % a počet průchodů mediánem segmentu pak 30 %. Je-li pro daný segment splněno alespoň 50 % pro splnění neznělosti, pak je daný segment označen za neznělý a V/UV bit se nastaví na 1. Na začátku analýzy je každý segment považován za znělý a V/UV bit nastaven na hodnotu 0.

MULTIPLEXOVÁNÍ

Teď už zbývá jen seřadit jednotlivé bity za sebe tak, aby mohly vytvořit rámce s parametry pro každý segment. Jeden rámec bude mít velikost 72 bitů a jeho struktura je zobrazena na obrázku 3.3.



Obr. 3.3: Bitový rámec pro jeden segment

Výsledná bitová rychlost simulovaného kodéru bude 3,2 kbit/s. Bylo tedy dosaženo kompresního poměru $\frac{104 \text{ kbit/s}}{3,2 \text{ kbit/s}} = 32,5 : 1$.

3.2 Pravidla pro návrh dekodéru

Budeme uvažovat dekodér podle obrázku 2.2. Je ho třeba synchronizovat tak, aby uměl správně pracovat s rámci přijatými z kodéru. Každý 72 bitový rámec je demultiplexován a jsou získány čtyři bloky parametrů:

1. V/UV bit

V/UV bit = 0: přepnutí na impulzní generátor (znělá hláska)

V/UV bit = 1: přepnutí na generátor šumu (neznělá hláska)

2. Následujících 6 bitů je kóduje T_0

Samotnou hodnotu T_0 získáme přepočtením binárního čísla na dekadické a užitím vztahu:

$$T_0 = \frac{T_{0_bity}}{63}(\tau_{max} - \tau_{min}) + \tau_{min} \quad (3.6)$$

Hodnota T_0 je užitečná jen v případě znělé hlásky a má význam vzálenosti impulzů od sebe v impulzním generátoru. Pro detekovanou neznělou hlásku je hodnota T_{0_bity} automaticky nastavena na nulu.

3. Následuje 5 bitů pro stanovení velikosti buzení G

Tentokrát je potřeba hodnotu přepočítat na jistý koeficient, který bude udávat zesílení zesilovače, který bude zesilovat nebo utlumovat generovaný šum či impulz. Algoritmus získávání hodnot zesilovacího koeficientu z bitů je zřejmý po inverzi kvantovacího mechanismu z tabulky 3.2. Vlastní hodnota koeficientu závisí na tom, jaké budou normalizované hodnoty rozptylu šumu z generátoru či velikosti impulzů.

Při přijaté hodnotě $G = 0$ je segment považován za tichý a impulzy či šum budou ztlumeny.

4. LSF koeficienty

Zbytek rámce (60 bitů) reprezentuje LSF koeficienty a ty je třeba převést na LP koeficienty pro určení přenosové funkce rekonstrukčního filtru hlasového traktu dle vztahu (1.6). Algoritmus převodu bitů na koeficienty ω_i se bude řídit pravidly pro kvantizaci z tabulky 3.1. Dále je třeba převést nalezené koeficienty na LSP koeficienty dle vztahu $q_i = \cos(\omega_i)$ a provést s nimi následující operace [7]:

$$f_1(i) = -2_{q_{2i-1}} f_1(i-1) + 2f_1(i-2), \quad i = 1, \dots, 5, \quad (3.7)$$

kde je pro každé $j = (i-1), \dots, 1$ počítáno

$$f_1^{(i)}(j) = f_1^{(i-1)}(j) - 2_{q_{2i-1}} f_1^{(i-1)}(j-1) + f_1^{(i-1)}(j-2), \quad (3.8)$$

přičemž $f_1(0) = f_1(-1) = 0$. Koeficienty $f_2(i)$ se vypočítají stejným způsobem, akorát místo q_{2i-1} užitíme sudé koeficienty q_{2i} . Máme nalezené koeficienty vztahů $P'(z)$, $Q'(z)$, které potřebujeme vynásobit výrazy $1 + z^{-1}$ a $1 - z^{-1}$ abychom dostali koeficienty pro $P(z)$, $Q(z)$:

$$\begin{aligned} \text{pro } P(z) : \quad g_1(i) &= f_1(i) + f_1(i-1), \quad i = 1, \dots, 5, \\ \text{pro } Q(z) : \quad g_2(i) &= f_2(i) + f_2(i-1), \quad i = 1, \dots, 5, \end{aligned} \quad (3.9)$$

Konečně je možno spočítat samotné LP koeficienty:

$$a_i = \begin{cases} 0,5g_1(i) + 0,5g_2(i) & i = 1, \dots, 5 \\ 0,5g_1(11-i) - 0,5g_2(11-i) & i = 6, \dots, 10 \end{cases} \quad (3.10)$$

Vztah (3.10) koresponduje se vztahem (1.43) pro zpětný výpočet LP koeficientů.

3.3 Výsledky simulace

Výsledkem této práce jsou dva programy. Ten první s názvem „**kódování**“ splňuje úkol samostatného projektu této práce – převádí nahrávky ve formátu `.wav` na bitový tok, tedy pole jedniček a nul, které má počet řádků = počtu segmentů v nahrávce, 1 řádek = 72 bitů (viz. obrázek 3.3). Druhý program má název „**segmenty**“; ten prochází nahrávku segment po segmentu, zobrazuje jeho průběh a ukazuje LP koeficienty, periodu základního tónu, normalizovaný autokorelační koeficient, počet průchodů mediánem a úroveň buzení signálu v onom segmentu. Oba tyto programy jsou k dispozici na přiloženém CD, návod na jejich správné použití je v příloze této práce. Na CD jsou k dispozici i nahrávky, na kterých je možno si programy otestovat.

Programy bylo třeba testovat na nejrůznějších nahrávkách za účelem správné funkčnosti. Nejvíce času zabralo pozorování správné detekce znělosti-neznělosti pomocí AMDF. Program označil za znělý každý segment, který byl znělý v celé jeho délce. Přechodné děje, velmi tiché úseky a evidentně neperiodické děje označil za neznělé segmenty. Tvorba vlastní simulace přinesla spoustu dalších poznatků a možností jak simulaci samotnou zdokonalit. Důležité bylo, že výsledky získané pomocí zmíněných matematických postupů se shodovali s výsledky získanými pomocí vestavěných funkcí v Matlabu. Tedy:

- Hodnoty LP koeficientů naprosto odpovídaly hodnotám získaným při zadání stejných vstupních parametrů funkci `lpc` (ta vypočítá LP koeficienty pro zvolený vektor čísel)
- Hodnoty vypočtených LSF koeficientů se blížily hodnotám získaným užití funkce `poly2lsp` (ta převádí koeficienty polynomu na LSF). Úplně se s nimi neshodovaly proto, že jsme nastavili krok při Čebyševově metodě hledání kořenů „jen“ na 1024, což vzhledem k jejich následné kvantizaci stačí; větší přesnosti bychom dosáhli při několikanásobně vyšším počtu kroků, což by mělo za následek výrazné zpomalení programu.

3.4 Možná vylepšení tohoto LPC kodéru

Na tento kodér se dá aplikovat spousta dalších vylepšení aniž by se změnila podstata jeho činnosti. Např. při segmentaci se v lepších kodérech užívá **okénkování segmentu** (nejčastěji Hammingovým a Blackmannovým okénkem) s přesahem 50 % do sousedního segmentu. Zlepší se tím homogenost průběhu signálu v segmentu[5].

U **neznělých hlásek** není ani třeba všech deset koeficientů vypočítávat. Vzhledem k velice chaotickému průběhu signálu u těchto hlásek stačí vypočítat **jen první čtyři koeficienty**, abychom se dostali na požadovanou kvalitu[5].

Kromě znělých a neznělých segmentů existují také **tiché segmenty** (bude-li buzení menší než určitá hodnota, pak se dá daný segment klasifikovat jako mezera) a **přechodné segmenty** (např. náběh na znělou hlásku, či přechod ze znělé na neznělou). Tyto údaje by vyžadovaly další parametry do rámce[6].

Existuje plno dokonalejších metod pro určení periody základního tónu (pomocí filtru **LTP** – *Long-Term Prediction*) a detekce znělosti-neznělosti (**keprstrální analýza**, krátkodobá analýza autokorelační funkce). Tyto metody jsou ovšem i složitější a jejich rozbor by byl obsáhlý a nad rámec účelu této práce[6, 5].

Existují také algoritmy, které umožňují neposílat určité segmenty, například při dlouze vyslovované hlásece „éé“ by se poslalo mnoho stejných či velmi podobných rámců. V tomto případě se **některé rámce mohou vynechat** a ty, co jsou poslány, nesou informaci o možné vzniklé chybě a v dekodéru se pak ty chybějící dopočítávají. Tím se výrazně sníží bitová rychlost na úkor výpočetní náročnosti[5].

Dají se využít i **psychoakustické modely pro váhování a kvantování** zkoumaných parametrů.

3.5 Porovnání se současně využívanými hovorovými kodéry

Samotné LPC kódování se používá od konce 60. let 20. století a dodnes je „srdcem“ řečových kodérů. Na samotné této metodě nebylo co zdokonalovat – má za úkol zjistit parametry rekonstrukčního filtru hlasového traktu. Zdokonalovalo se tedy vše možné okolo za účelem kvalitnější zpětné rekonstrukce řeči po přidání dalších vlastností signálu.

V dnešní době se využívají hlavně kódové knihy a modely vytvořené dlouhodobými výzkumy vlastností parametrů řeči. Přišlo se na to, že se spousta kombinací různých parametrů opakuje nebo má mezi sebou určitou závislost, nebo také že některé kombinace se v řeči vůbec nevyskytují a nemusí být vůbec uvažovány. Toto umožnilo zvýšení efektivity kódování, protože dnes nebývají do kanálu vyslány přímo hodnoty parametrů řeči, ale pouze odkazy na jejich jistou kombinaci v kódové knize (vektorová kvantizace). Kromě pevných kódových knih existují i adaptivní kódové knihy, které si může sám kodér upravovat. Dále se pracuje s chybou predikce $e(n)$, která bývá zkoumána nejčastěji autokorelací a výsledky autokorelace jsou opět pomocí odkazů zaslány k dekodéru a napomáhá to k minimalizaci chyby predikce při rekonstrukci. Dále se zkoumají parametry pro segmenty různých délek, protože na zkoumání některých vlastností jsou segmenty okolo 20–30 ms příliš krátké (třeba pro LTP) a někdy jsou příliš dlouhé (např. změny buzení při přechodném ději). Kromě rámců pro segment pak tím pádem vznikají i subrámce s vybranými parametry,

které jsou zasílány každých 5 ms. Často se užívá i predikce buzení a pitch periody dle stanovených predikčních metod. Bitové rychlosti výstupních dat zůstávají nejčastěji v rozmezí 8–16 kbit/s, přičemž vzhledem k možnostem přenosové techniky ji nemá význam snižovat, na druhou stranu její zvýšení by nevedlo ke zvýšení kvality dekódované řeči; ta se zvýší spíše přidáním nového modulu do kodéru. [10, 5, 8]

Doporučení pro řečové kodeky standardizuje ITU (*International Telecommunication Union*) pod označeními ITU-T G.7xx. Příklady doporučení[10, 7]:

- **G.723.1** — ACELP (*Algebraic Code-Excited Linear Prediction*) - 5,3 kbit/s a MP-MLQ (*Multipulse Maximum Likelihood Quantization*) – 6,3 kbit/s s použitím ve videotelefonech v klasických telefonních sítích
- **G.726** — ADPCM (*Adaptive Differential Pulse Code Modulation*) – 40, 32, 24, a 16 kbit/s
- **G.729** — CS-ACELP (*Conjugate Structure Algebraic Code-Excited Linear Prediction*) — 8 kbit/s

Příklady dnes užívaných hovorových kodeků[9, 11]:

- **GSM - Full Rate** — první generace kodeku RPE-LTP pro mobilní síť GSM (ETSI GSM 06.10) — 13 kbit/s (+ 9,8 kbit/s zabezpečení FEC (*Forward Error Correction*))
- **GSM - Half Rate** — kodek s poloviční rychlostí VSELP pro síť GSM (ETSI GSM 06.20) – 5,6 kbit/s
- **IS-127** — druhá generace mobilních sítí CDMA v USA používající kodek RCELP - 8, 4, nebo 2 kbit/s

4 ZÁVĚR

V této práci byla popsána a vysvětlena účinná a často používaná metoda kódování řečového signálu zvaná lineární prediktivní kódování řeči (LPC). Tato metoda je založena na tom, že pro krátké úseky řeči se dá signál předpovídat ze vzorků předěšlých dle lineární závislosti. Samotný mechanismus kódování spočívá v tom, že je řečový signál rozdělen na krátké segmenty, nejčastěji 20–30 ms, a pro každý takový segment jsou užitím autokorelace segmentu a Levinson-Durbinova algoritmu vypočteny tzv. LP koeficienty, které pro daný segment určují jeho frekvenční vlastnosti. LPC kódování kooperuje s dalšími vlastnostmi řeči jako je periodický průběh signálu u znělých hlásek, charakter signálu jako bílého šumu u neznělých hlásek, základní tón hlasu a formantové frekvence. LP koeficienty pak slouží jako koeficienty pro přenosovou funkci rekonstrukčního filtru IIR v dekodéru řeči. LPC kódování se využívá v řadě zdrojových hovorových kodérů, tzv. vokodérů. Zde jsou dále určeny parametry řečového signálu jako perioda základního tónu, úroveň buzení, detekce znělé či neznělé hlásky a samotné LP koeficienty jsou přenášeny ve formě spektrálních párů (LSP) či frekvencí (LSF).

Úkolem samostatného projektu v rámci této práce byla simulace kodéru na bázi LPC pomocí programu Matlab, kdy pro každý segment byly určovány výše uvedené parametry a následně byly multiplexovány je do bitového toku. Tento program má název „kódování“ a je v něm užito všech metod uvedených v práci pro získání bitového toku, který signál reprezentuje tak, aby se dal ve srovnatelné kvalitě rekonstruovat. Díky takovému kodéru bylo možno snížit přenosovou rychlost ze 104 kbit/s na 3,2 kbit/s, což odpovídá kompresnímu poměru 32,5:1. Důraz byl kladen hlavně na kvantování LSF koeficientů, kdy bylo zapotřebí navrhnout co nejefektivnější kvantování těchto koeficientů tak, aby celekově zabíraly 60 bitů v rámci jednoho segmentu. Bitový tok byl vysílán po rámcích dlouhých 72 bitů pro segmenty délky 22,5 ms. Princip činnosti tohoto kodéru je v textu detailně popsán. Matematické operace nutné k realizaci onoho programu byly převzaty z literatury, některé metody (např. vyhledávání minima funkce AMDF při detekci pitch periody, kvantování LSF koeficientů, určení mezí pro detekci neznělosti) jsem si vytvořil a ověřil sám na základě vlastních pokusů. K tomuto programu jsem ještě přidal program „segmenty“, který prochází nahrávku segment po segmentu, zobrazuje jeho průběh a ukazuje LP koeficienty, periodu základního tónu, normalizovaný autokorelační koeficient, počet průchodů mediánem a úroveň buzení v onom segmentu. Oba tyto programy byly pro jejich správnou funkčnost testovány pro asi tři desítky krátkých zvukových nahrávek, které jsou společně s oběma funkčními programy k dispozici na přiloženém CD.

LITERATURA

- [1] BROCKWELL, P. J. , DAVIS, R. A. : *Time Series: Theory and Methods (Springer Series and Statistics)*. 3rd edition, Springer Science + Business Media, LLC. New York, USA. ISBN 0-387-97429-5
- [2] ČERNOCKÝ, J. : *Číslicové zpracování řeči: LPC*. Brno: Vysoké učení technické v Brně, Fakulta informačních technologií, 2003. Dostupné z URL: <http://www.fit.vutbr.cz/~cernocky/speech/pred/04_lpc/04_lpc.pdf>.
- [3] DAVIS, G.M. : *Noise reduction in speech applications*. Davis, G.M., editor, CRC Press LLC, 2002. ISBN 0-8493-0949-2
- [4] DELLER, J. R. , HANSEN, J. H. L. , PROAKIS, J.G: *Discrete-Time Processing of Speech Signals (IEEE Press Classic Reissue)*. Wiley-IEEE Press; Reprint edition, 1999. ISBN 0780353862
- [5] GOLD, B., MORGAN, N. : *Speech and audio signal processing: Processing and perception of speech and music*. John Wiley & Sons, Inc., New York 2000, ISBN 0-471-35154-7
- [6] HANUS, S. : *Bezdrátové mobilní komunikace*. Skriptum k předmětu. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2005.
- [7] INTERNATIONAL TELECOMMUNICATION UNION: *General Aspects of Digital Transmission Systems*. ITU-T recommendation G.729, 1996.
- [8] MORGAN, N. : *Audio Signal Processing in Humans and Machines*. University of California - Berkeley, International Computer Science Institute USA, 1995.
- [9] SMÉKAL, Z., VRBA, K., MALENOVSKÝ, V., SYSEL, P. : *Optimalizace ANSI-C kódu realizujícího kodek GSM Full Rate pro signálový procesor TMS320C6400*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2003.
- [10] SMÉKAL, Z., VRBA, K., MALENOVSKÝ, V., SYSEL, P. : *Optimalizace ANSI-C kódu realizujícího kodek G.723.1 pro signálový procesor TMS320C6400*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2003.
- [11] VRBA, K., SMÉKAL, Z., SYSEL, P. : *Optimalizace ANSI-C kódu realizujícího GSM kodér pro signálový procesor TMS320C6000*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a informatiky, 2001.

- [12] The MathWorks, Inc., *Matlab* [počítačový program]. Ver. 7.5.0.342 (R2007b), 2007. *Nápověda k funkcím lpc, poly2lfs, plot, wavread.*
- [13] YING, G.S., JAMEISON, L.H., MICHELL, C.D.: *A Probablistic Approach To AMDF Pitch Detection*. Newark: University of Delaware, Applied Science and Engineering Laboratories, USA. Dostupné z URL:
<<http://www.asel.udel.edu/icslp/cdrom/vol2/218/a218.pdf>>.

SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

LPC	Linear Predictive Coding – Lineární prediktivní kódování
IIR	Infinite Impulse Response – Nekkonečná impulzní charakteristika
LSP	Line Spectral Pairs – Čárové spektrální páry
LSF	Line Spectral Frequencies – Čárové spektrální frekvence
AR	Autoregressive – Autoregresní
MA	Moving-Average – Pohyblivý průměr
DFT	Discrete Fourier Transformation – Diskrétní Fouriérova transformace
CELP	Code-Excited Linear Prediction
RPE	Regular Pulse Excitation – Regulární impulzní buzení
LTP	Long-Term Prediction – Dlouhodobá predikce
ADPCM	Adaptive Differential Pulse Code Modulation
AMDF	Average Magnitude Difference Function
ITU	International Telecommunication Union
CDMA	Code Division Multiple Access – Přístup kódovým multiplexem
$G(z)$	Přenosová funkce modelu hlasivek
$V(z)$	Přenosová funkce modelu hlasového traktu
$L(z)$	Přenosová funkce modelu vyzařování zvuku
$H(z)$	Přenosová funkce modelu artikulačního traktu
T_s	Doba trvání jednoho vzorku
p	Řád predikce
$A(z)$	Polynom stupně p obsahující LP koeficienty
$a_i^{(p)}$	LP koeficienty predikce řádu p
$s(n)$	Původní řečový signál
$\tilde{s}(n)$	Predikovaný řečový signál

$e(n)$	Chyba predikce
$\mathbf{MA}(q)_t$	Model klouzavých průměrů řádu q
b_q	Charakteristické parametry modelu $\mathbf{MA}(q)_t$
e_t	Náhodná chyba
$\mathbf{AR}(p)_t$	Autoregresní model řádu p
y_t	Hodnoty vzorků zkoumané veličiny metodou $\mathbf{AR}(p)_t$
E	Výkon (energie) chyby predikce
$\phi(i, j)$	Soustava rovnic s autokorelačními koeficienty
$R(k)$	Autokorelace signálu v bodě k
N	Počet vzorků v segmentu
G	Úroveň buzení signálu
$P(z)$	Symetrický polynom pro hledání LSP koeficientů
$Q(z)$	Antisymetrický polynom pro hledání LSP koeficientů
q_i	LSP koeficienty
ω_i	LSF koeficienty
$f_1(i)$	Pomocné koeficienty pro polynom $P(z)$
$f_2(i)$	Pomocné koeficienty pro polynom $Q(z)$
$F(\omega)$	Polynom zkoumaný při Čebyševově polynomiální metodě
$C(x)$	Čebyševův polynom
$T_m(x)$	Člen určující řád Čebyševova polynomu
f_{vz}	Vzorkovací frekvence
T_0	Perioda základního tónu (tzv. pitch perioda)
V/UV	Symbol označující znělost-neznělost hlásky
r_1	Normalizovaný autokorelační koeficient
int	Integer – Celočíslná část výsledku

A CD SE SIMULACÍ ŘEČOVÉHO KODÉRU

Náplní přiloženého CD jsou dva programy (`kodovani.m`, `segmenty.m`), elektronická verze této BP (`BP_LPC_kodovani.pdf`) a adresář `zvuky`, ve kterém jsou uloženy krátké řečové nahrávky ve formátu `.wav`, které byly vybrány z databáze nahrávek na UTKO FEKT VUT v Brně.

Pokyny pro práci s programy:

V obou případech je potřeba ke spuštění program Matlab. V tom je třeba nastavit „Current Directory“ na CD či DVD jednotku kde ono CD máte, nebo adresář, kam jste jeho obsah zkopírovali. Charakteristika programů:

1. **Kódování** – `kodovani.m`

Program se spouští voláním funkce `bits = kodovani('<nazev_zvuku>')`. Jeho výstupem je pole `bits` obsahující výsledný bitový tok po provedení všech zmíněných operací pro celou nahrávku. Jeden segment je zobrazen vždy do jednoho řádku. Výstupem funkce je také obrázek průběhu vyfiltrovaného signálu v celé nahrávce.

2. **Segmenty** – `segmenty.m`

Spouští se voláním funkce `[pocet delka] = segmenty('<nazev_zvuku>')`. Program ukáže na obrázku průběh signálu v segmentu a v „Command Window“ zobrazí tyto parametry segmentu:

`f0_____n0_____r1_____G_____ =`

(frekvence základního tónu [Hz], podíl průchodů mediánem signálu v segmentu [%], normalizovaný autokorelační koeficient pro segment [-], úroveň buzení signálu v segmentu [-])

`A=`

(Deset LP koeficientů pro signál v segmentu)

Funkce obsahuje příkaz `pause`, tedy čeká na stisk libovolné klávesy pro přechod na další segment. Po vyhodnocení posledního segmentu nahrávky ukáže funkce své výstupní hodnoty – `pocet` = počet segmentů v nahrávce; `delka` = délka nahrávky v milisekundách.

Zde jsou uvedeny všechny možné parametry `<nazev_zvuku>` oddělené čárkami: `a1, a2, be1, be2, ce1, ce1z, de1, deset1, dve, e1, e2, ef1, ef1z, es1, es1z, hhzh, jedna1, jedna1z, nula1, nula1z, o1, o1z, o2, te1, te2, tri1, tri1z, u1`.