

## PRINCIPY ZPRACOVÁNÍ HLASU V KLASICKÉ A IP TELEFONII

**Doc. Ing. Boris ŠIMÁK, CSc.**

pracoviště: ČVUT FEL, Katedra telekomunikační techniky; mail: simak@feld.cvut.cz

**Abstrakt:** Tento příspěvek si klade za cíl seznámit čtenáře se základními technickými přístupy ke kódování hlasového signálu v komunikačních aplikacích. Potřeba vysokého stupně komprese hlasových signálů a přitom rostoucí požadavky na kvalitu hlasového signálu vede k využívání řady kompresních postupů, některé z nich jsou standardizovány ve tvaru konkrétních doporučení ITU-T nebo ETSI či ISO norem. Používané principy jsou již dlouhou dobu známé, ale teprve v posledních letech díky technologickému pokroku jsou složité a výpočetně náročné algoritmy realizovatelné ať již specializovanými integrovanými obvody či programově s využitím signálových procesorů. Z možných kompresních technik jsou popsány stručně postupy využívané při přenosu hovoru prostřednictvím internet protokolu.

### 1 Hovorový signál a jeho vlastnosti

Zdrojem hovorového signálu jsou řečové orgány, které se skládají z hlasivek, hrdelní dutiny, ústní dutiny, nosní dutiny, měkkého a tvrdého patra, zubů a jazyka. Zdrojem buzení této soustavy jsou plíce ve spolupráci s dýchacími svaly. Vlivem tlaku proudu vzduchu vycházejícího z plic dochází k jeho modulaci hlasivkami. Kmitočet závisí na tlaku vzduchu na svalovém napětí hlasivek. Kmitočet hlasivek je charakterizován *základním tónem* lidského hlasu (pitch periode)  $F_0$ , který tvoří základ znělých zvuků (tj. samohlásek a znělých souhlásek). Kmitočet základního tónu je různý u dětí, dospělých, mužů i žen, pohybuje se většinou v rozmezí 150 až 400 Hz. V klidu je štěrbina hlasivek otevřena a proud vzduchu volně prochází hlasivkami. Vytvářený zvuk je po průchodu hlasivkami formován ústní dutinou a je vyřazován do volného prostoru.

Sdělení zprostředkované řečovým signálem je diskrétní, tzn. může být vyjádřeno ve tvaru posloupnosti konečného počtu symbolů. Každý jazyk má vlastní množinu těchto symbolů - fonemů, většinou 30 až 50.

Hlasy řeči dále můžeme rozdělit na znělé (n, e, ...) a neznělé (š, č, ...) a jejich kombinace. Znělá hláska představuje kvaziperiodický průběh signálu, neznělá pak signál podobný šumu. Navíc energie znělých hlásek je větší než neznělých. Krátký časový úsek znělé hlásky můžeme charakterizovat její jemnou a formantovou strukturou. Formantem označujeme tón tvořící akustický základ hlásky. Ten vlastně představuje spektrální obal řečového signálu. Jemná harmonická struktura představuje chvění hlasivek.

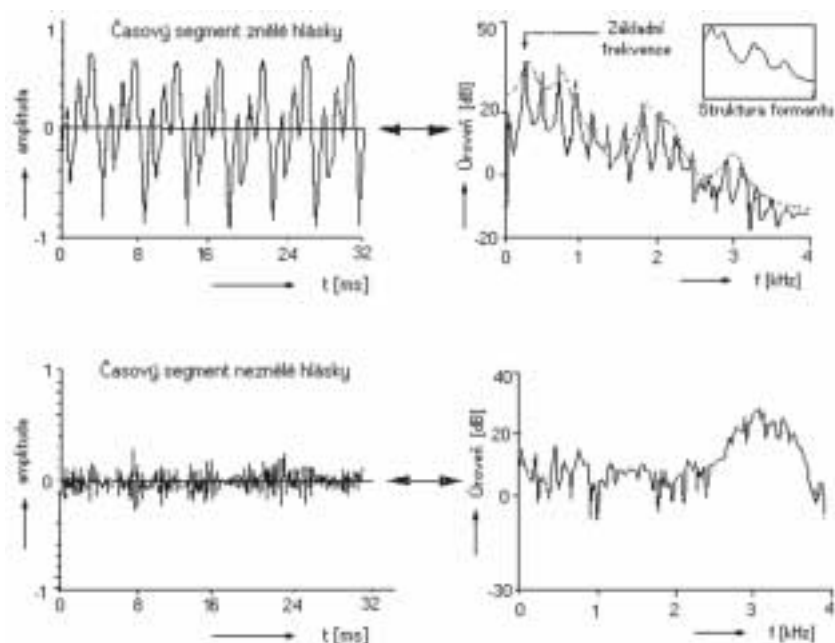
Lidská řeč je souvislý časově proměnný proces, z toho plyne i náročnost popisu lidské řeči a jejího modelování. Příklad časového a kmitočtového průběhu znělého a neznělého segmentu hovorového signálu je na obr. 1.

Hovorový signál snímáme mikrofonom, který převádí modulovaný proud vzduchu na elektrický signál. Z hlediska digitálního zpracování řeči vycházíme nejčastěji z:

*kódování vzorků časového průběhu* hovorového signálu, tj. kódování tvaru vlny ('Waveform Coding'). Mezi tyto postupy patří pulsně kódová modulace PCM, delta modulace DM, adaptivní delta modulace ADM, diferenciální pulsně kódová modulace DPCM, adaptivní diferenciální pulsně kódová modulace ADPCM,

*kódování parametrů zdroje* hovorového signálu, tj. parametrického kódování ('Source Coding'). Mezi základní metody patří vokodérové metody spočívající na principu lineární predikce.

*hybridních metod kódování.* Do této kategorie patří kodeky založené na adaptivních predikčních metodách APC, kodeky s multiimpulsním MPE-LP či regulárním buzením RPE-LP nebo kodeky využívající vektorové kvantizace (CELP, LD-CELP). Dále mezi hybridní kodéry patří metody založené na složkovém kódování (SBC-Sub-Band Coding) nebo adaptivním transformačním kódování (ATC – Adaptive Transform Coding) využívající rychlé diskrétní transformace (DFT, DCT) [3].



Obr. 1: Časový a kmitočtový průběh znělého a neznělého segmentu hovorového signálu

Cílem kompresních metod je dosáhnout co nejnižší přenosové rychlosti při zachování dobré srozumitelnosti, barvy hlasu mluvčího a přiměřené výpočetní náročnosti použitých algoritmů.

## 2 Kódování tvaru vlny

Zpracování vychází z přeměny hovorového signálu ve tvaru elektrického signálu. Tento signál je vzorkován, kvantován a následně kódován.

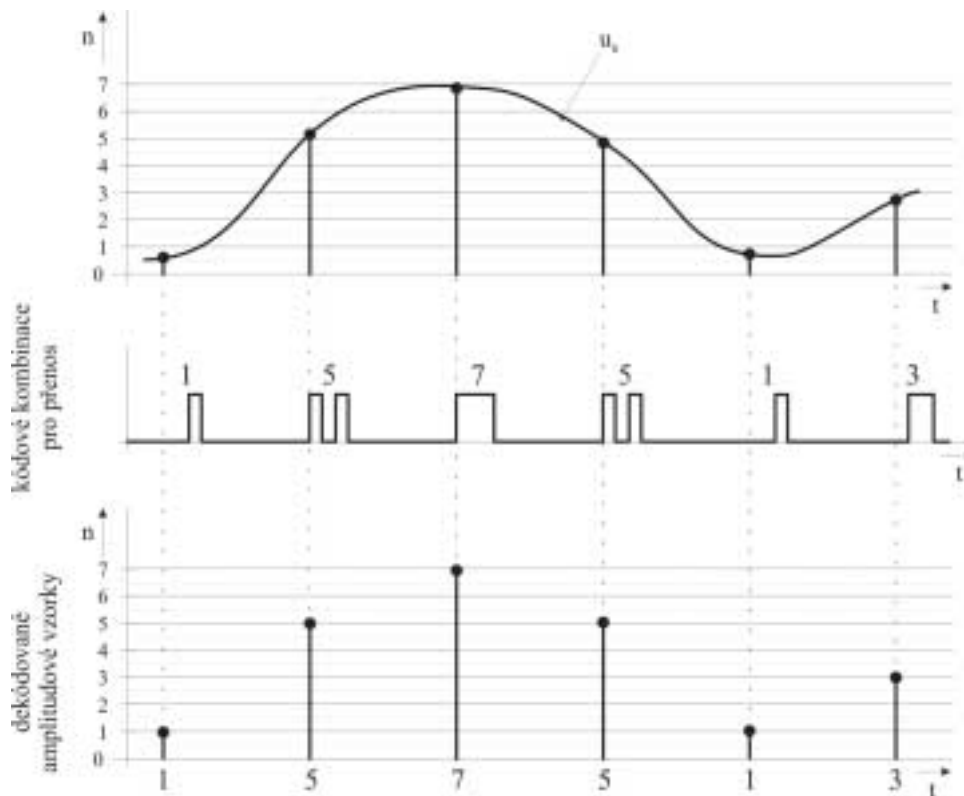
### 2.1 Pulsně kódová modulace PCM (Pulse Code Modulation)

PCM kódování sestává ze tří kroků:

Vzorkování - hodnoty spojitého analogového signálu  $u_x$  se odečítají v diskretním čase. Vzorkovací kmitočet vychází z Shannonova-Kotělnikova vzorkovacího teorému [3].

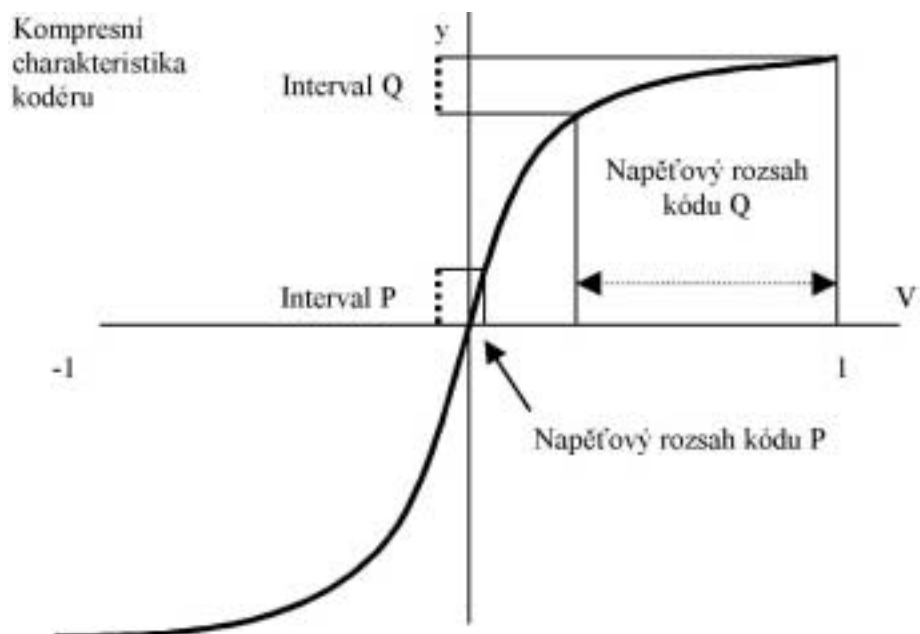
Kvantování – ke každému vzorku získanému v předchozím kroku se přiřadí jedna z možných diskretních úrovní.

Kódování - diskretní hodnoty vzorků jsou reprezentovány pomocí  $n$ -bitových čísel.  $n$  se vybírá jako mocnina čísla 2 podle toho, jaké rozlišení požadujeme, viz obr. 2. V porovnání s analogovými metodami přenosu má PCM výhodu ve větší odolnosti přenášeného digitálního signálu vůči rušení a v možnosti mnohonásobné obnovy při přenosu digitálního signálu. Nevýhodou je větší náročnost na šířku kmitočtového pásma a vznik tzv. kvantizačního zkreslení daného rozdílem mezi skutečnou hodnotou vzorku a výslednou kvantovanou hodnotou, která byla vzorku přiřazena. Při využití telekomunikačního kanálu s šířkou 4 kHz a vzorkovacím kmitočtu 8kHz je přenosová rychlost pro jeden telefonní kanál 64 kbit/s. Tento postup odpovídá lineárnímu kvantování.



Obr. 2: Princip kódování PCM

Pro zajištění rovnoměrné hodnoty odstupe kvantizačního zkreslení v pracovním rozsahu kodéru PCM je možno využít nelineárního rozložení kvantizačních stupňů. Obdobného výsledku dosáhneme kompresí dynamiky přenášeného signálu na vysílací straně a expanzí dynamiky na straně přijímací. Malé hodnoty vzorků se na vysílací straně zesílí a velké zeslabí. Na přijímací straně proběhne inverzní proces. Zakódování se pak realizuje kodérem s lineárním kvantováním. Příklad kompresní charakteristiky je na obr. 3 (velikosti rozsahů P a Q se transformují na intervaly P a Q).

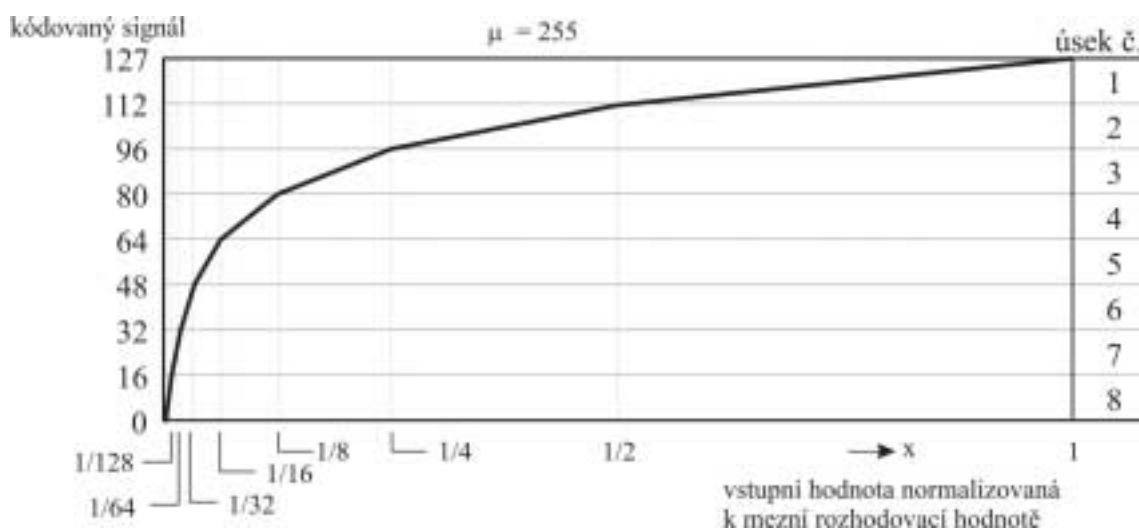


Obr. 3: Příklad kompresní charakteristiky

Z hlediska realizovatelnosti se spojitá charakteristika nahrazuje charakteristikou složenou z lineárních lomených úseků. Na obr. 4 je nakreslena odpovídající charakteristika pro kladné hodnoty signálu. Charakteristika je složená z osmi lomených úseků. Při použití kódování čtyřmi bity je každý segment rozdělen do šestnácti dílčích hodnot, celkově pro kladné i záporné hodnoty signálu tak získáme rozlišení na 256 stupňů, což odpovídá kódování osmi bity. Signál, který je v praxi kódován s vyšším počtem bitů se převádí na kódování osmi bity tak, že:

- 1 bit určuje polaritu signálu,
- 3 bity určují příslušný segment,
- 4 bity kódují velikost kvantizační úrovně v daném segmentu.

Tím získáme rozlišení odpovídající 13 bitové lineární kvantizaci. Pro kompresní charakteristiky se používá název charakteristika typu A (využívaná především v Evropě) nebo  $\mu$  (používaná zejména v USA).



Obr. 4: Kompresní charakteristika  $\mu$

## 2.2 Diferenciální pulsně kódová modulace DPCM (Differential Pulse Code Modulation)

Diferenciální PCM je modifikací PCM kódování a její princip byl publikován v roce 1952. Systém s DPCM kóduje rozdíly mezi okamžitou hodnotou vzorku signálu v daném vzorkovacím okamžiku a hodnotou predikovanou z předchozích vzorků. K zakódování rozdílu je zapotřebí nižšího počtu bitů. To vede k redukci přenosové rychlosti a tedy i množství přenášených dat se snižuje. DPCM se využívá především při kódování obrazu.

## 2.3 Adaptivní diferenciální pulsně kódová modulace ADPCM (Adaptive Differential Pulse Code Modulation)

Tento způsob kódování vychází z DPCM. Je vylepšen tak, že srovnávací průběh je vytvářen adaptivně a přizpůsobuje se konkrétním statistickým parametrům řeči. Výsledkem je ještě menší dynamický rozsah než v případě DPCM a tedy opět potřebujeme méně bitů k zakódování vzorku.

## 3 Zdrojové kódování hovorových signálů

### 3.1 Model zdroje

Pro účely syntézy hovorového signálu je nutno vytvořit vhodný model syntezátoru odpovídající způsobu generování hovorového signálu člověkem. Tento model můžeme popsat např. přenosovou funkcí. Model vychází z obr. 5, na kterém nalezneme zjednodušený model produkce řeči. Fyziologický model uvedený na obr. 5 a) můžeme

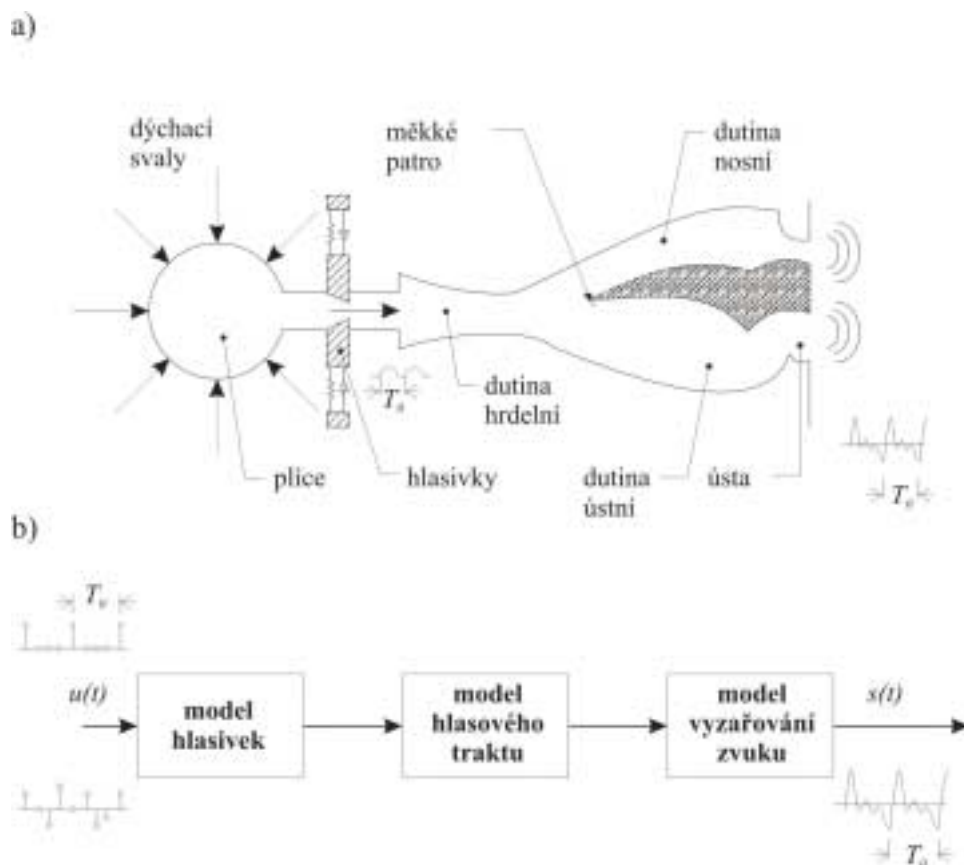
nahradit kaskádním zapojením dílčích bloků odpovídajících modelu hlasivek, modelu hlasového traktu a modelu vyzářování zvuku podle schéma na obr. 5 b).

Celkový přenos systému modelujícího vokální trakt je dán součinem dílčích přenosových funkcí, které obsahují pouze póly, viz [2]. Výslednou přenosovou funkci  $H(z)$  můžeme uvést v polynomiálním tvaru (1)

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

**Model buzení řečového systému** využívá pro znělé zvuky sled impulsů s periodou odpovídající časové periodě základního tónu  $T_0$  a pro neznělé zvuky náhodný šum s plochým spektrem. Odpovídající z-obraz budící funkce za předpokladu normované periody vzorkování je  $U(z)$ . Z-obraz signálu na výstupu modelu je dán vztahem (2)

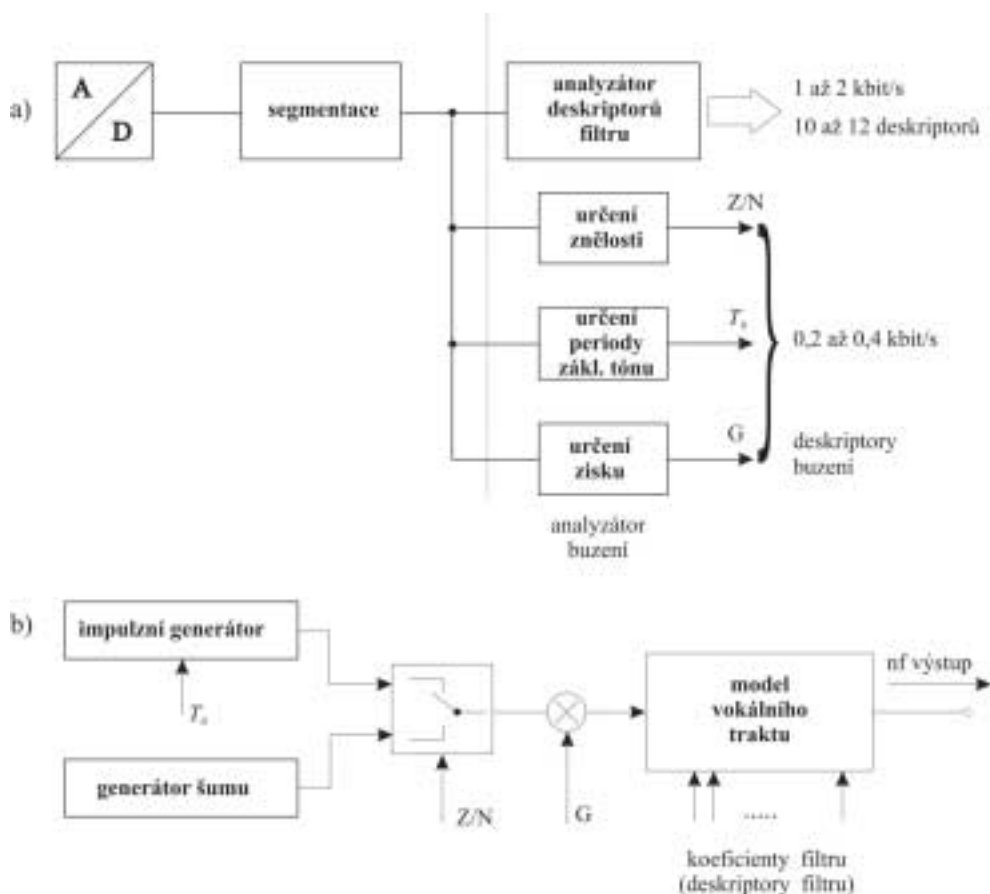
$$S(z) = H(z)U(z) \quad (2)$$



Obr. 5: Model produkce řečového signálu a) a jeho náhradní blokové schéma b)

### 3.2 Princip vokodéru

Klasickým parametrickým systémem je tzv. *vokodér* (Voice Coder). Vysílací část systému v procesu parametrizace hovorového signálu (analýza) vyhodnocuje cca 10 až 30 ms dlouhé časové úseky hovorového signálu (segmenty), které je možno považovat z hlediska jejich popisu za stacionární a vyčísľuje soubory parametrů - deskriptorů pro každý časový segment (znělý/neznělý úsek, periodu základního tónu, zisk  $G$ , koeficienty syntetizujícího filtru), které charakterizují daný segment. Blokové schéma vysílací části vokodéru je uvedeno na obr. 6 a). Stanovené deskriptory jsou v časovém multiplexu přenášeny v digitálním tvaru na přijímací stranu, kde může být generován hovorový signál na základě modelu 6 b).



Obr. 6: Blokové schéma vysílací části vokodéru a), blokové schéma syntezátoru vokodéru b)

Zvuk je v syntezátoru vytvářen na základě buzení periodickou posloupností pro znělé segmenty hovorového signálu a náhodnou posloupností s rovnoměrným spektrem (bílý šum) pro neznělé segmenty. Po úpravě buzení na základě hodnoty zisku  $G$  vstupuje budící signál do syntetizujícího filtru reprezentujícího model vokálního traktu a na výstupu modelu je produkován hovorový signál – srozumitelná řeč. Koeficienty jsou nastaveny na hodnotu nalezenou při analýze hovorového signálu ve vysílací části vokodéru.

Rozhodujícím faktorem pro kvalitu reprodukce přijímaného hovorového signálu jsou vlastnosti hlasového syntezátoru. Vokodéry dosahují nízké přenosové rychlosti (cca 2,5 kbit/s), výstupem syntezátoru je ovšem nepřírozený zvuk. Tyto nevýhody vokodéru se odstraňují v modernějších kodérech např. vhodnějšími způsoby buzení nebo kódováním zbytkového signálu, který získáme na vysílací straně průchodem hovorového signálu inverzním filtrem LPC.

Přenosová rychlost komprimovaného signálu je pak dána počtem přenášených deskriptorů a jejich vyjádřením v bitech. Vztah pro přenosovou rychlost je dán vztahem (3)

$$v_p = F_r \cdot \sum_{i=1}^{N_d} N_{kdi} \quad (3)$$

- kde  $F_r$  je opakovací kmitočet rámce  
 $N_d$  je počet přenášených deskriptorů  
 $N_{kdi}$  je počet bitů, použitých pro dvojkové vyjádření hodnoty  $i$ -tého deskriptoru.

Deskriptory jsou získávány analýzou hovorového signálu. Existuje řada možností jak získat popis vstupního hovorového signálu na základě jeho analýzy.

V praxi se nejvíce používají postupy založené na *lineární predikci* (LP). Koeficienty predikce jsou získávány pomocí minimalizace střední kvadratické odchylky mezi skutečnými a předpověděnými hodnotami vzorků řešením soustavy lineárních rovnic. Charakterizují spolu s dalšími parametry daný segment. Často používaný postup nalezení predikčních koeficientů je tzv. Levinsonova-Durbinova rekurzívní metoda [3]. Tyto koeficienty jsou využívány k realizaci filtru v syntezátoru.

Další redukce přenosové rychlosti lze dosáhnout např. technikou *vektorového kvantování*.

#### 4 Využití lineární predikce v kodérech hovorového signálu

Lineární predikce je založena na předpovědi následující hodnoty na základě lineárního závislosti výpočtem z předchozích hodnot. U lineárního prediktoru  $p$ -tého řádu je aktuální hodnota vzorku získána lineární kombinací posledních  $p$  vzorků tj.

$$s(n) = \sum_{i=1}^p a_i s(n-i), \quad (4)$$

kde  $s(n)$  je předpovídaný vzorek procesu,  $a_i$  jsou koeficienty predikčního filtru. Vokální trakt je modelován přenosovou funkcí  $H(z)$  obsahující samé póly podle vztahu (5).

$$H(z) = \frac{G}{1 - \sum_{j=1}^p a_j z^{-j}} = \frac{G}{A(z)} = \frac{S(z)}{X(z)} \quad (5)$$

Ta odpovídá kaskádně řazeným dvojpólovým rezonátorům reprezentujícím hlasivkový model. Na základě buzení  $GX(z)$  je produkována srozumitelná řeč ve formě obrazu v  $z$ -rovině  $S(z)$ , viz vztah (6).

$$S(z) = \frac{G}{A(z)} X(z), \quad (6)$$

kde  $G$  je zisk a jmenovatel  $A(z)$  je dán vztahem (7)

$$A(z) = 1 - \sum_{j=1}^p a_j z^{-j}. \quad (7)$$

Koeficienty  $a_j$  ve vztahu (7) jsou získány lineární predikcí. Systémové a budící parametry jsou určeny z konečné množiny řečových vzorků.

Vztah (6) můžeme v časové oblasti vyjádřit diferenční rovnicí (8)

$$s(n) = Gx(n) + \sum_{i=1}^p a_i s(n-i). \quad (8)$$

Chyba predikce  $e(n)$  je dána vztahem (9)

$$e(n) = s(n) - \sum_{i=1}^p a_i s(n-i), \quad (9)$$

z čehož plyne, že buzení syntetizujícího filtru odpovídá *chybový signál*, viz (10):

$$e(n) = G(n)x(n) = u(n). \quad (10)$$

Tento chybový signál získáme na vysílací straně průchodem hovorového signálu *inverzním predikčním filtrem* LP s přenosovou funkcí podle rovnice (11)

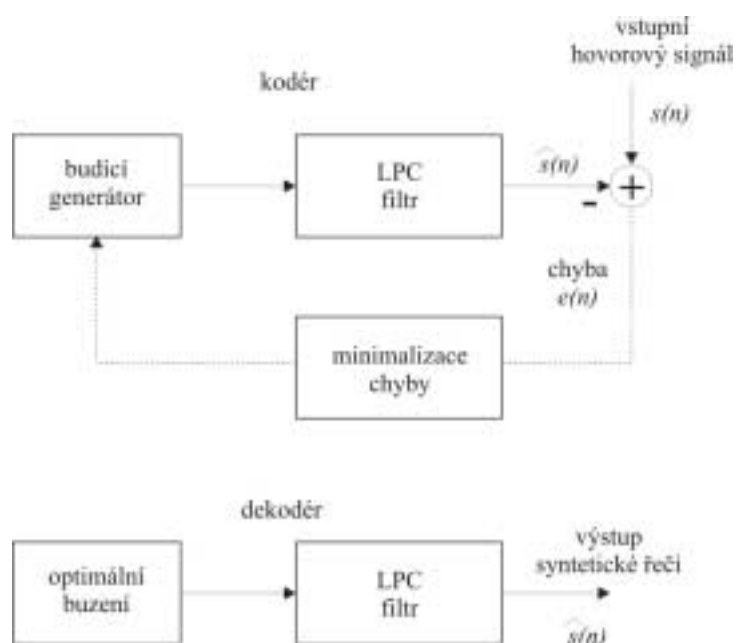
$$H^{-1}(z) = 1 - \sum_{j=1}^p a_j z^{-j} = A(z). \quad (11)$$

Obraz budící posloupnosti (chybového signálu) je

$$U(z) = H^{-1}(z)S(z). \quad (12)$$

## 5 Kodeky s počítačovou syntézou budícího signálu

Další postupy zlepšení kvality kodérů hovorového signálu vycházejí z možností digitálního zpracování signálů. Cílem je nalézt vhodnou budící posloupnost hlasového syntezátoru modelujícího vokální trakt, aby vytvořená umělá řeč se co nejvíce blížila původní kódované řeči. Tyto postupy jsou označovány obecným názvem analýza na základě syntézy (Analysis by Synthesis - Abs). Zjednodušené blokové schéma principu funkce kodéru a dekodéru hovorového signálu s digitálním generátorem budícího signálu je uvedeno na obr. 7.



Obr. 7: Blokové schéma kodéru a dekodéru hovorového signálu na základě metody Abs

Hlasový syntezátor (LPC filtr), který je součástí kodéru, generuje posloupnost vzorků hovorového signálu (daného segmentu)  $\hat{s}(n)$  jako odezvu na budící posloupnost. Tyto vzorky jsou porovnávány v rozdílovém členu s originálními vzorky (segment stejné délky) hovorového signálu. Hledá se pro daný hovorový signál taková budící posloupnost, aby chybový signál  $e(n)$  na výstupu rozdílového členu byl minimální podle určitého kritéria. Nalezená optimální budící posloupnost je vhodným způsobem zakódována a spolu s ostatními deskriptory (koeficienty LPC filtru) přenesena do přijímače, kde budí predikční filtr reprezentující vokální trakt a na jehož výstupu získáváme syntetickou řeč.

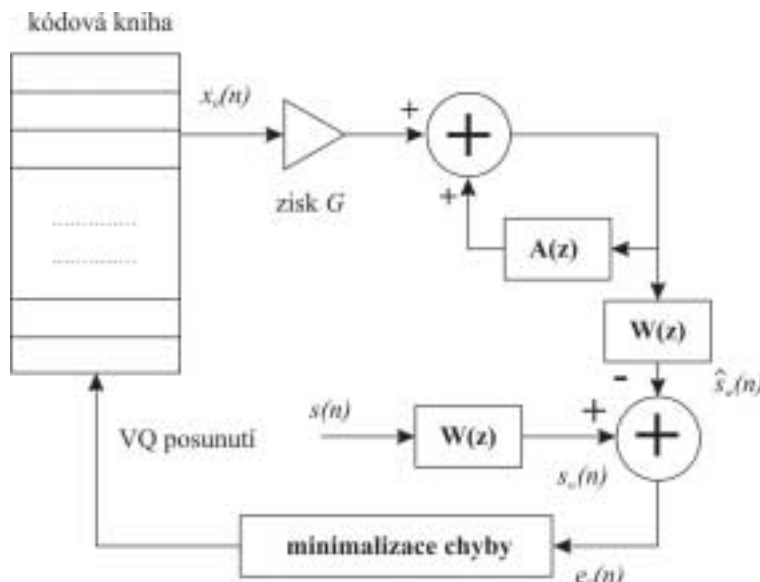
Chybový signál na výstupu rozdílového členu kodéru bývá upraven percepčním *váhovým filtrem*  $W(f)$  tak, aby odpovídal lidskému vnímání hovorového signálu. Často bývá tento filtr zařazen do vstupů rozdílového členu místo na výstup, viz obr. 8.

Podle způsobu vytváření budící posloupnosti můžeme kodeky nazývat kodeky s *multipulsním buzením* (MPE – Multi-Pulse Excitation) nebo regulárním (RPE – Regular-Pulse Excitation) buzením. V prvním případě se hledají vhodné amplitudy a velikosti budících impulsů, v druhém případě je k dispozici několik budících posloupností s pevně danými parametry budících impulsů a hledá se optimální vzájemná poloha budících posloupností. Je-li budící posloupnost generována na základě vektorové kvantizace hovoříme o *kodecích CELP* (Code - Excited Linear Prediction). Existuje řada modifikačních postupů, podrobnější informace lze nalézt např. v [3].



## 5.1 CELP

Jedná se o lineárně predikční kodeky s vektorovým kvantováním budícího signálu. Jejich přenosová rychlost se pohybuje od 4 do 8 kbit/s. Blokové schéma kodéru CELP je na obr. 8 a odpovídá základní koncepci AbS metody.



Obr. 8: Kodér CELP

V pamětech kodéru a dekodéru jsou uloženy v kódových knihách soubory možných posloupností vstupního signálu. Je-li v kódové knize nalezena posloupnost odpovídající budícímu signálu, je příslušná adresa budící posloupnosti (tj. vektoru uloženého v kódové knize) binárně přenesena do přijímače. Přijímač pak podle přijaté adresy generuje na základě shodné kódové knihy příslušnou budící posloupnost pro LP hlasový syntezátor. Optimální vektor je určen užitím kritéria minima střední kvadratické chyby. Vstupní vzorky i syntetizovaný signál jsou upraveny váhovým filtrem  $W(z)$ . Kvalita řeči u CELP kodérů může být zlepšena následnou filtrací syntetizované řečové posloupnosti s cílem zdůraznit strukturu řeči z hlediska formantů a dominantních kmitočtů.

Jedna z nevýhod původního CELP algoritmu je velká výpočetní náročnost prohledávání kódových knih. Mnoho CELP algoritmů vyžaduje procesory schopné zpracovat více než 20 MIPS a prostor pro kódovou knihu kolem 40 kbyťů. Z toho vyplývá snaha o vývoj vhodných prohledávacích algoritmů. Kódovací zpoždění je až 35 ms a omezuje použití CELP kodeků pro přenosy na delší vzdálenosti a jejich vícenásobné užití.

## 5.2 LD-CELP

Způsob kódování s krátkou dobou zpoždění LD-CELP (LD – *Low Delay*) podle doporučení ITU-T G.728 je založen na zpětnovazebním řízení adaptace s kódovacím zpožděním do 2ms a přenosovou rychlostí 16 kbit/s. Koncepce LD-CELP se od kodeků CELP liší v řešení vlastního hlasového syntezátoru a způsobu jeho řízení. U kodeků CELP se přenášejí v časovém multiplexu lineárně predikční koeficienty vyčíslené v kodéru krátkodobou (STP) a dlouhodobou analýzou (LTP), které slouží v přijímači k periodickému nastavování LPC filtrů hlasového syntezátoru. Naproti tomu mezi kodérem a dekodérem LD-CELP se tato data nepřenášejí a jak kodér, tak dekodér jsou řízeny zpětnovazebně. Nastavení budící úrovně se provádí pro každý blok vzorků. V LD-CELP dekodéru se nastavení úrovně a koeficienty LPC filtru nastavují obdobně jako u kodéru. Převodník kódu převádí blok vzorků odpovídající vektoru syntetického signálu na odpovídající PCM kód podle kompresní charakteristiky  $A$  nebo  $\mu$ .

## 6 Vlastnosti kodeků

Z důvodu porovnání kodeků mezi sebou, je potřeba stanovit parametry, pomocí kterých je budeme hodnotit. Nejčastěji se sledují tyto parametry:

**Bitrate [kbit/s]** – přenosová rychlost potřebná k plynulému přenosu zakódovaných dat.

**Typ kódování** - způsob jakým je signál kódován.

**Zpoždění [ms]** - při kódování signálu dochází ke zpoždění, příliš velké zpoždění (cca 200 ms) působí problémy při komunikaci, čím menší zpoždění kodek způsobuje, tím lépe.

**Výpočetní složitost [MIPS – million instructions per second]** – udává náročnost na výkon procesoru.

**Paměťová náročnost [kByte]** - velikost paměťového prostoru nutného pro realizaci kodeku.

**MOS - [Mean Opinion Score]**. Kritérium k hodnocení kvality řečového signálu. MOS je určeno číselnou stupnicí, kde číslo 1 znamená nejhorší kvalitu, 5 představuje kvalitu nejlepší. Kvalitu vždy posuzuje skupina lidí z jejichž názoru se potom statisticky určí MOS.

## 7 Přehled používaných standardů

**Kodek G.711** je základní kodek, který se používá i v klasické telefonní síti. Pro převod analogového signálu na digitální je použito kódování PCM, které se bez dalších úprav přenáší. Analogový signál se vzorkuje s frekvencí 8000 vzorků za vteřinu, každý vzorek má 8 bitů, z toho vyplývá, že potřebná šířka pásma pro přenos je 64 kbit/s. Při vzorkování signálu se používá charakteristika typu A (A-zákon, A-Law) nebo typu  $\mu$  ( $\mu$ -zákon,  $\mu$ -Law) ke kompresi/expanzi signálu. Tento kodek je základní, měla by ho podporovat všechna zařízení pro Voice over IP. Kvalita přenášeného hlasu je totožná s kvalitou hlasu při běžném telefonním hovoru, MOS je přibližně 4.1. Šestnácti bitové vzorky signálu jsou oříznuty na třináct bitů, které jsou následně kódovány na 8 bitů na základě definované kompresní charakteristiky.

**G.722** používá kódování ADPCM. Potřebná přenosovou rychlost je 16 kbit/s, tedy kodek signál komprimuje v poměru 4:1. Signál je rozdělen na 2 kmitočtová pásma, která jsou kódována pomocí ADPCM samostatně.

**Kodek G.723.1** Tento kodek používá buď kódování MP-MLQ (Multi-pulse Maximum Likelihood Quantization) nebo ACELP (Algebraic-code-excitation). První typ kódování vyžaduje přenosovou rychlost 6.3 kbit/s, druhý typ 5.3 kbit/s. Zpoždění kodeku je 30 ms a MOS hodnota je 3.9 při použití kódování MP-MLQ a 3.65 při použití ACELP.

**Kodek G.726** - tento kodek používá kódování ADPCM, potřebná přenosovou rychlost je 16, 24, 32 a 40 kbit/s. Kodek může zpracovávat bloky různé délky podle toho, jak velké zpoždění je požadováno. Vzorkovací frekvence je 8 kHz.

Standard	Algoritmus	MIPS	Přenosová rychlost [kbit/s]	MOS
G.711	PCM	---	64	4,1
G.723.1	MP-MLQ	16	6,3	3,9
G.723.1	ACELP	20	5,3	3,65
G.726	ADPCM	1	32	3,85
G.728	LD-CELP	30	16	3,61
G.729A	CS-ACELP	11	8	3,7
G.729	CS-ACELP	20	8	3,92
GSM 06.10	RPE-LP	10	13	3,5

Tab. 1: Porovnání výkonnosti kodeků

**Kodek G.728** používá kódování LD-CELP (Low Delay Code Excited Linear Prediction). Potřebná přenosovou rychlost je 16 kbit/s, MOS je přibližně 3.9 a zpoždění asi 30 ms.

**Kodek G729** - použité kódování je CS-ACELP (Conjugate Structure Algebraic Code Excited Linear Prediction). Potřebná přenosovou rychlost je 8 kbit/s, kvalita je podobná jako u ADPCM pro rychlost 32kbit/s.

**GSM** Tento způsob komprese řeči se používá v GSM sítích. Požadovaná přenosovou rychlost je 13 kbit/s. GSM je rychlejší než metody založené na slovníku (CELP), ale i tak je GSM komprese relativně náročná na výkon procesoru. Zpoždění kodeku je přibližně 20 ms.

Z tabulky 1 je patrné výhradní postavení kodeků G.729 a G.723.1, které poskytují nejlepší kompresní poměr vzhledem k dosažené kvalitě.

## **8 Realizace kodeků**

Kodeky jsou realizovány ve formě specializovaných obvodů často společně s dalšími funkčními bloky na čipu. V řadě případů je ale výhodnější realizace kodeků ve formě signálového procesoru, kde komprimační a dekomprimační postupy jsou realizovány programově s využitím instrukčního souboru daného signálového procesoru. Tento přístup je velice progresivní a umožňuje snadnou modifikaci a implementaci nových verzí kodeků. Představu o výpočetní náročnosti kompresních postupů si lze udělat na základě údajů uvedených v tab. 1 ve třetím sloupci, charakterizujících hodnoty potřebného početního výkonu v milionech instrukcí vykonaných za jednu sekundu (MIPS).

## **9 Závěr**

Příspěvek podává stručný pohled na nejvíce používané technické principy při realizaci kodeků pro přenos hlasu. V moderních kodecích se mimo prezentované postupy využívá možností komprese v kmitočtové rovině, komprese s užitím diskrétních transformací (FFT, DCT, waveletové transformace), bitové alokace na základě využití psychoakustického modelu a maskovacího efektu apod.

## **Literatura**

[1] Kondoz, A., M.: *Digital Speech. Coding for Low Bit Rate Communications Systems*. John Wiley & Sons, New York, 1996.

[2] Škop, M. a kol.: *Digitální telekomunikační technika. I. Díl, Digitální zpracování signálů*. TTC Marconi, 1996.

[3] Prchal, J. - Šimák, B.: *Digitální zpracování signálů v telekomunikacích*. ČVUT v Praze, 2000.