

Kódování řeči II.

Jan Černocký ÚPGM FIT VUT Brno, cernocky@fit.vutbr.cz

FIT VUT Brno

Plán

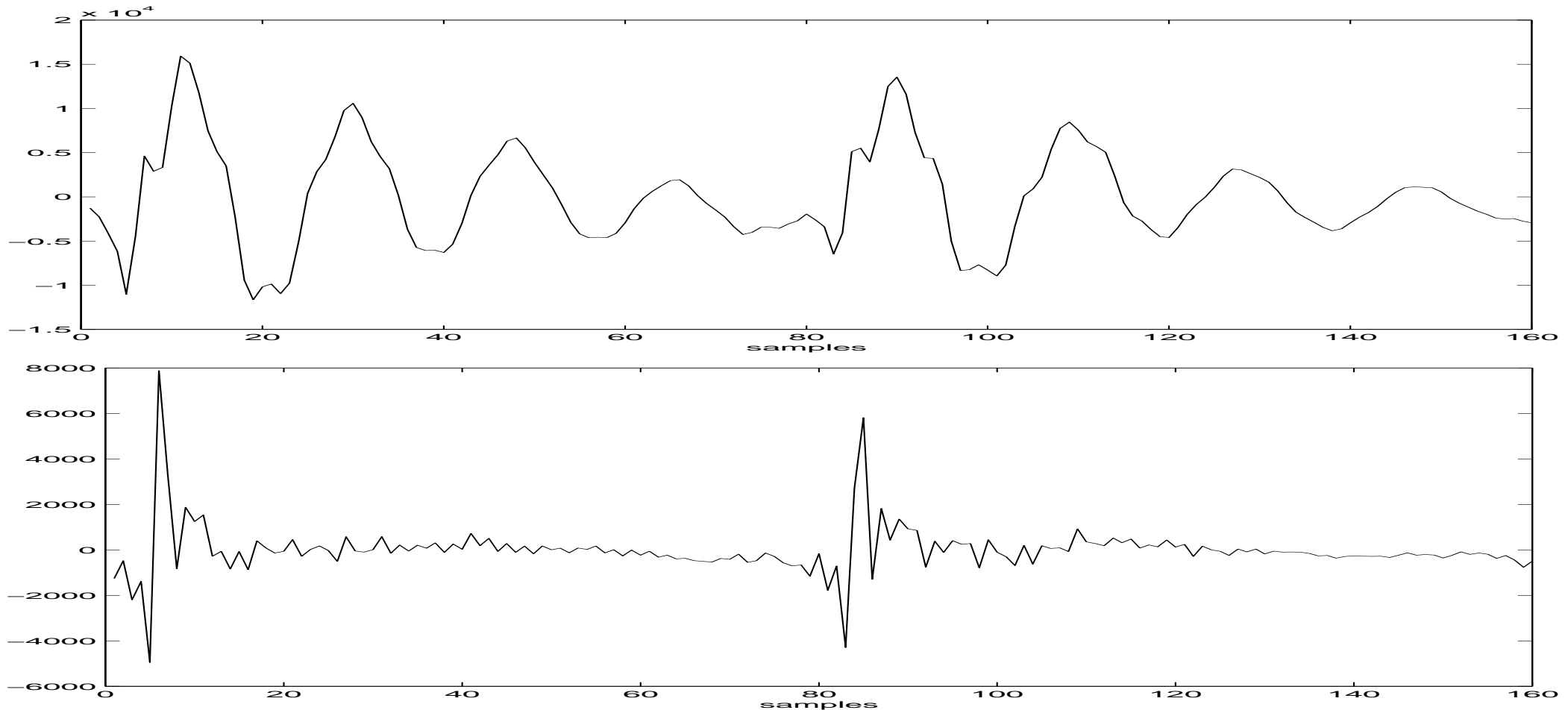
- Triky používané v kódování – dlouhodobý prediktor, analýza syntézou, perceptuální filtr.
- GSM full rate – RPE-LTP
- CELP
- GSM enhanced full rate – ACELP

KÓDOVÁNÍ BUZENÍ

- LPC dosahuje nízkých bitových toků, ale kvalita je velmi špatná (zní “plechově”).
- způsobeno velmi jednoduchým kódováním buzení (pouze voiced/unvoiced), zatímco u lidí jsou přítomny obě složky.
- v moderních (hybridních) kodérech je buzení věnována velká pozornost.
- následující slajdy: několik triků pro kódování buzení.

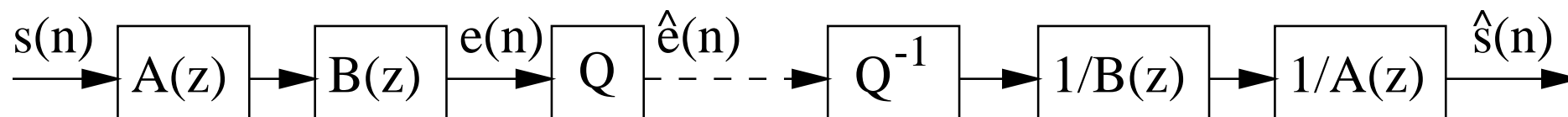
Dlouhodobý prediktor – long term predictor (LTP)

Víme, že chybový signál LPC má charakter šumu, ale pouze krátkodobě. Pro znělé hlásky je signál v delším časovém horizontu korelovan v po periodách základního tónu:



⇒ dlouhodobý prediktor (LTP) s přenosovou funkcí $-bz^{-L}$. Předpovídá vzorek $s(n)$ z minulého vzorku $s(n - L)$ (L je perioda základního tónu ve vzorcích – lag).

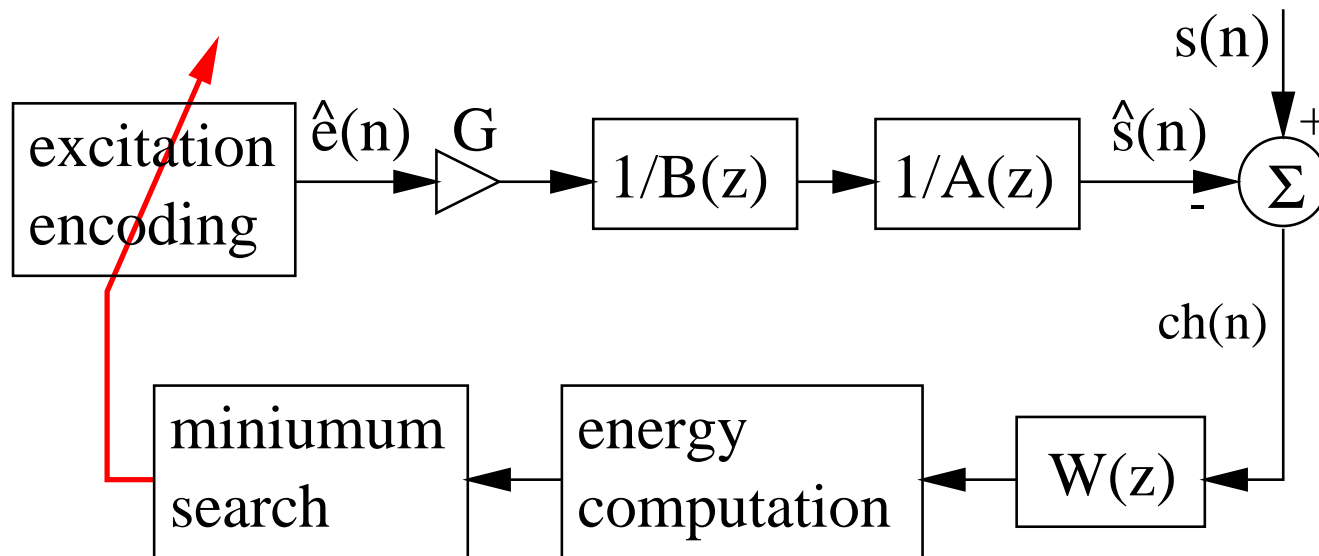
Chybový signál: $e(n) = s(n) - \hat{s}(n) = s(n) - [-bs(n - L)] = s(n) + bs(n - L)$,
takže $B(z) = 1 + bz^{-L}$.



Analýza syntézou

nejsme schopni najít analytický “vzoreček” pro vyhledání optimálního buzení \Rightarrow uzavřená smyčka (closed-loop).

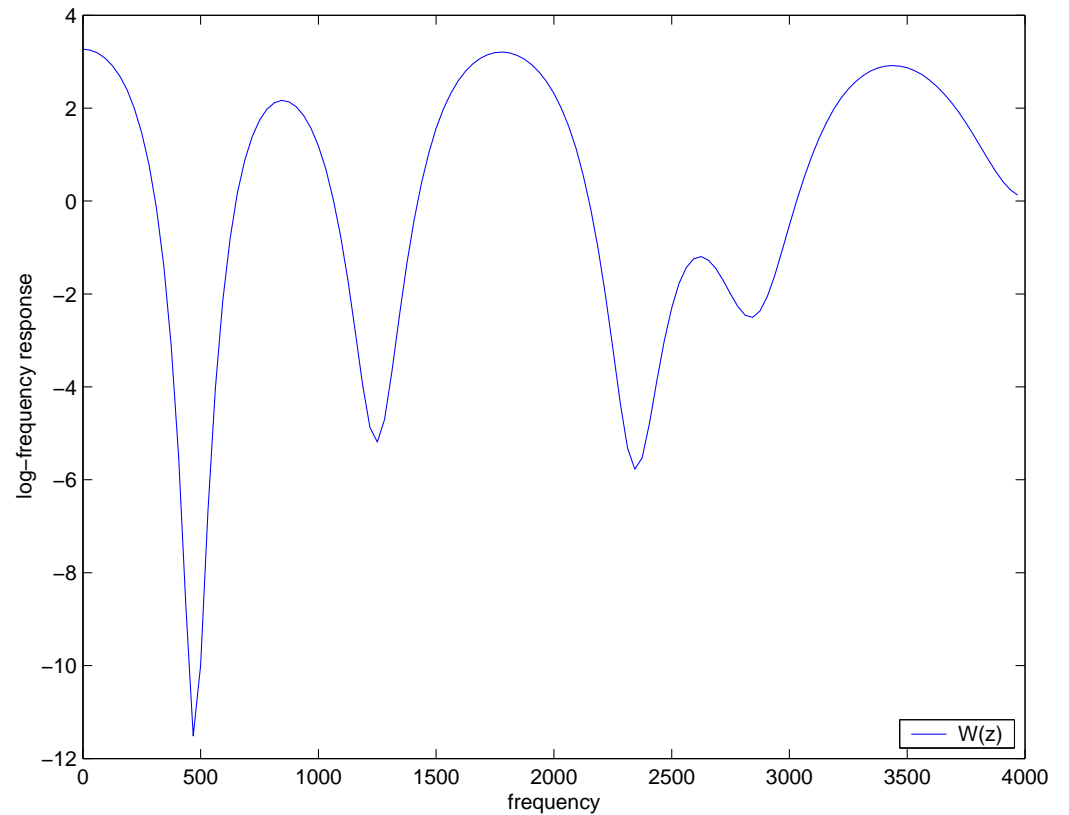
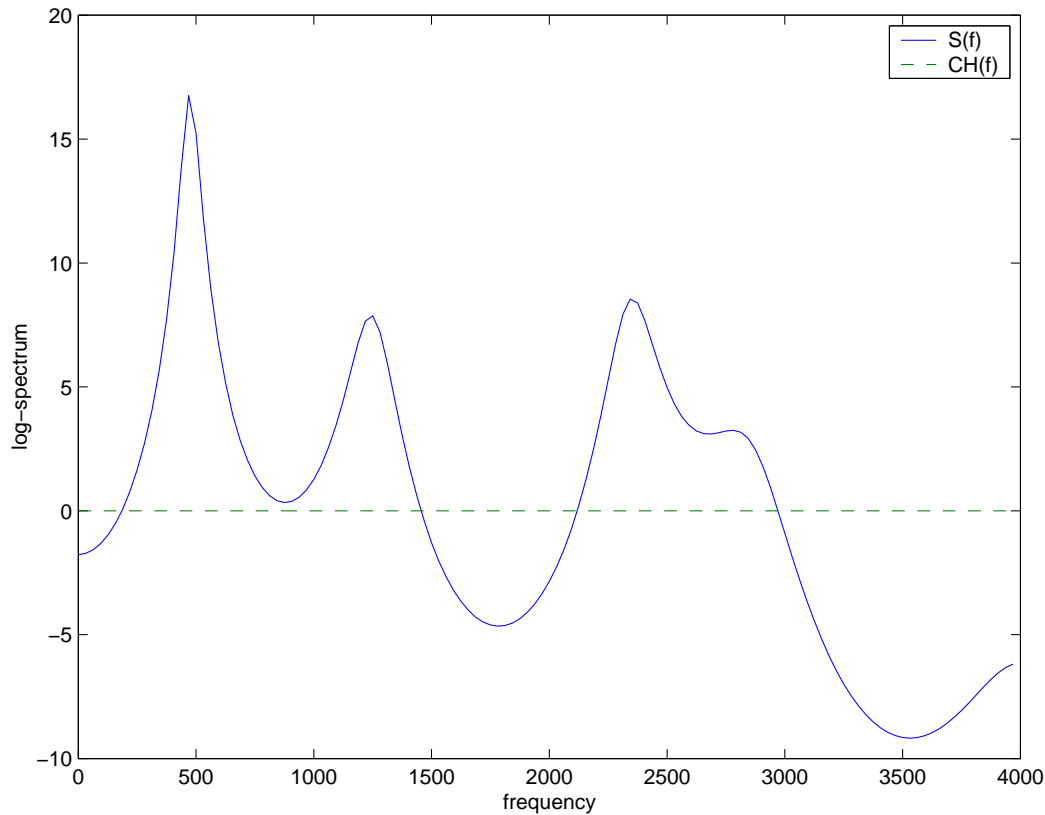
Pro odlišení od buzení $e(n)$ je rozdílový signál mezi dekódovaným a originálním (chyba) označen jako $ch(n)$.



Perceptuální filtr $W(z)$

v closed-loop srovnáváme syntetizovaný signál s originálem. PF – trik, který přibližuje kódování lidskému slyšení.

Srovnání spektra chyby se spektrem řeči:



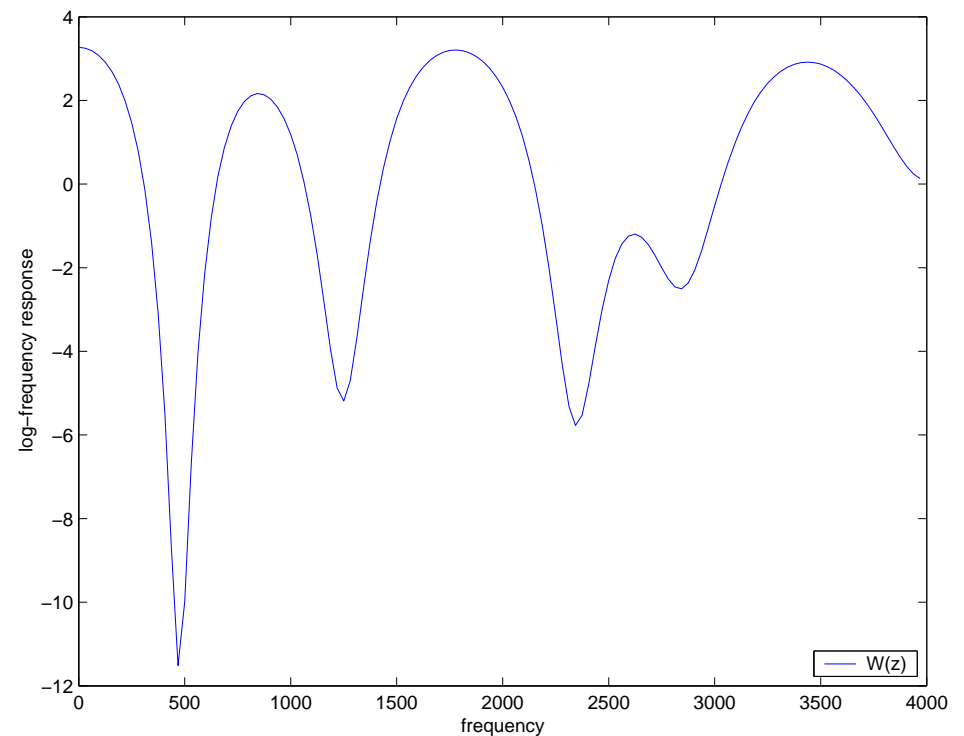
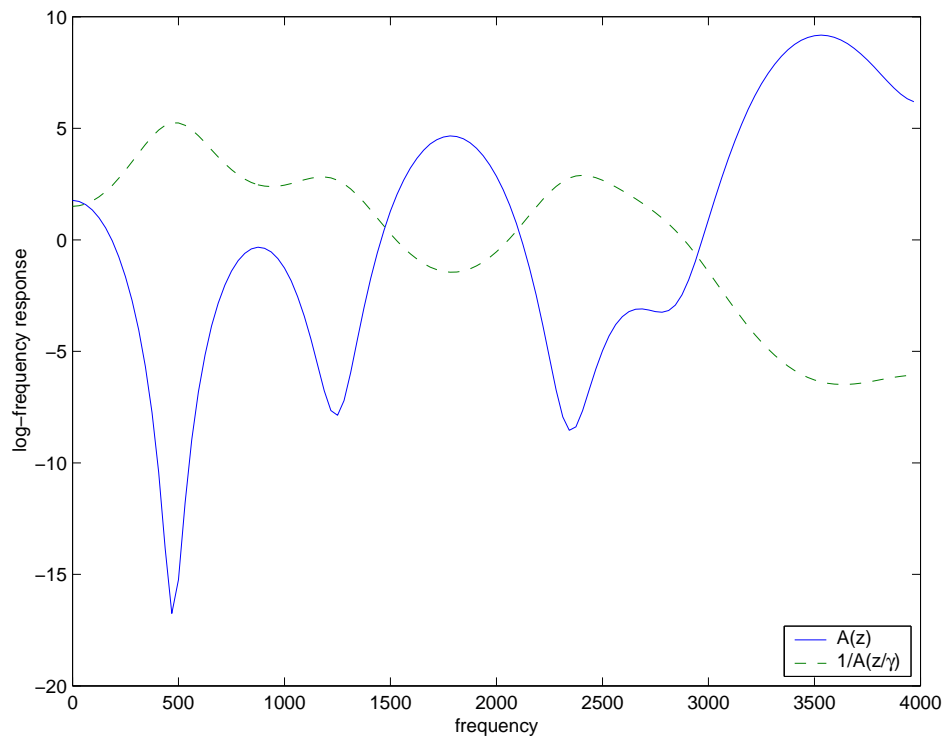
“V oblastech, kde jsou formanty, může být chyba jaká chce, protože ji stejně nebudu slyšet. Pojďme se spíše zaměřit na oblasti mezi formanty, kde má řečový signál malou energii”.

⇒ filtr, který bude chybový signál tlumit v místech formantů (tvrdíme, že tam není důležitý) a naopak posilovat v “údolích”.

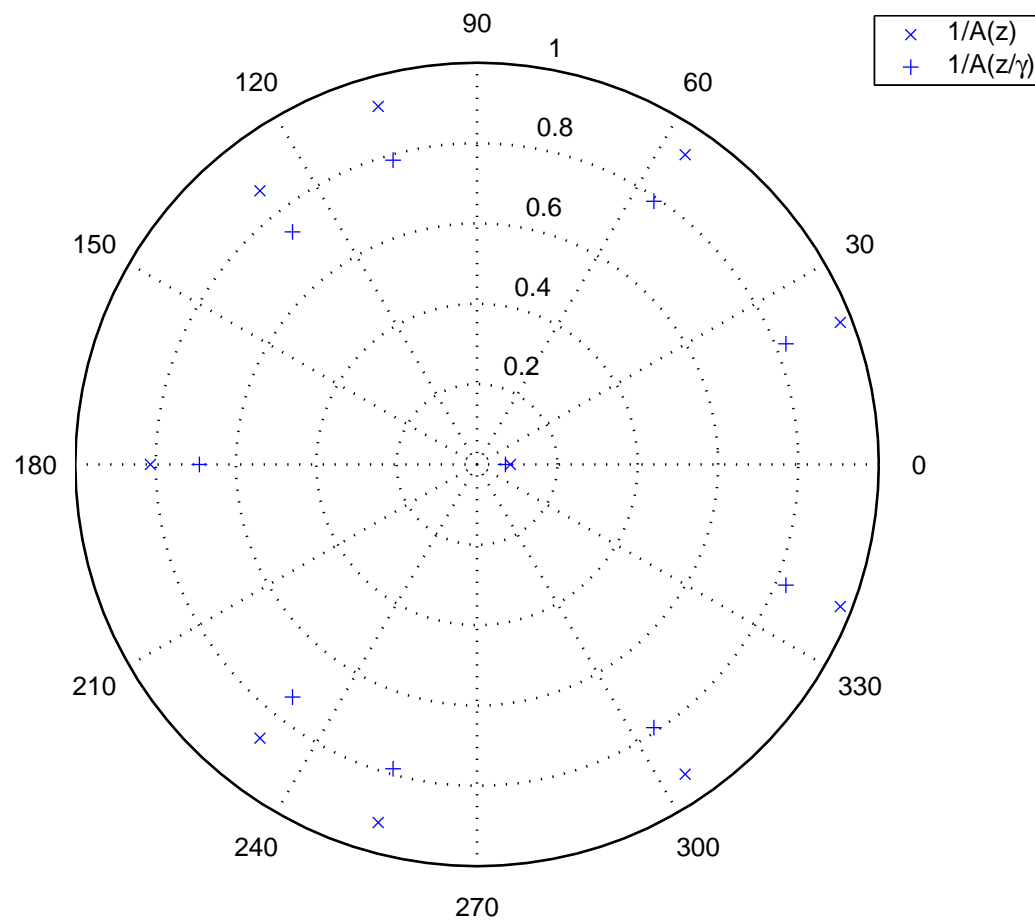
$$W(z) = \frac{A(z)}{A(z/\gamma)}, \quad \text{kde } \gamma \in [0.8, 0.9] \quad (1)$$

Jak to funguje ?

- čítateľ $A(z)$ (je to vlastně “inverzní filtr LPC”) má charakteristiku inverzí k vyhlazenému spektru řeči.
- Filtr $\frac{1}{A(z/\gamma)}$ má podobnou charakteristiku jako $\frac{1}{A(z)}$ (vyhlazené spektrum řeči), ale kvůli “tahání pólů směrem ke středu jednotkové kružnice” nejsou jeho špičky tak ostré.
- Když se to vynásobí, dostaneme PF $W(f)$ s požadovanou charakteristikou.



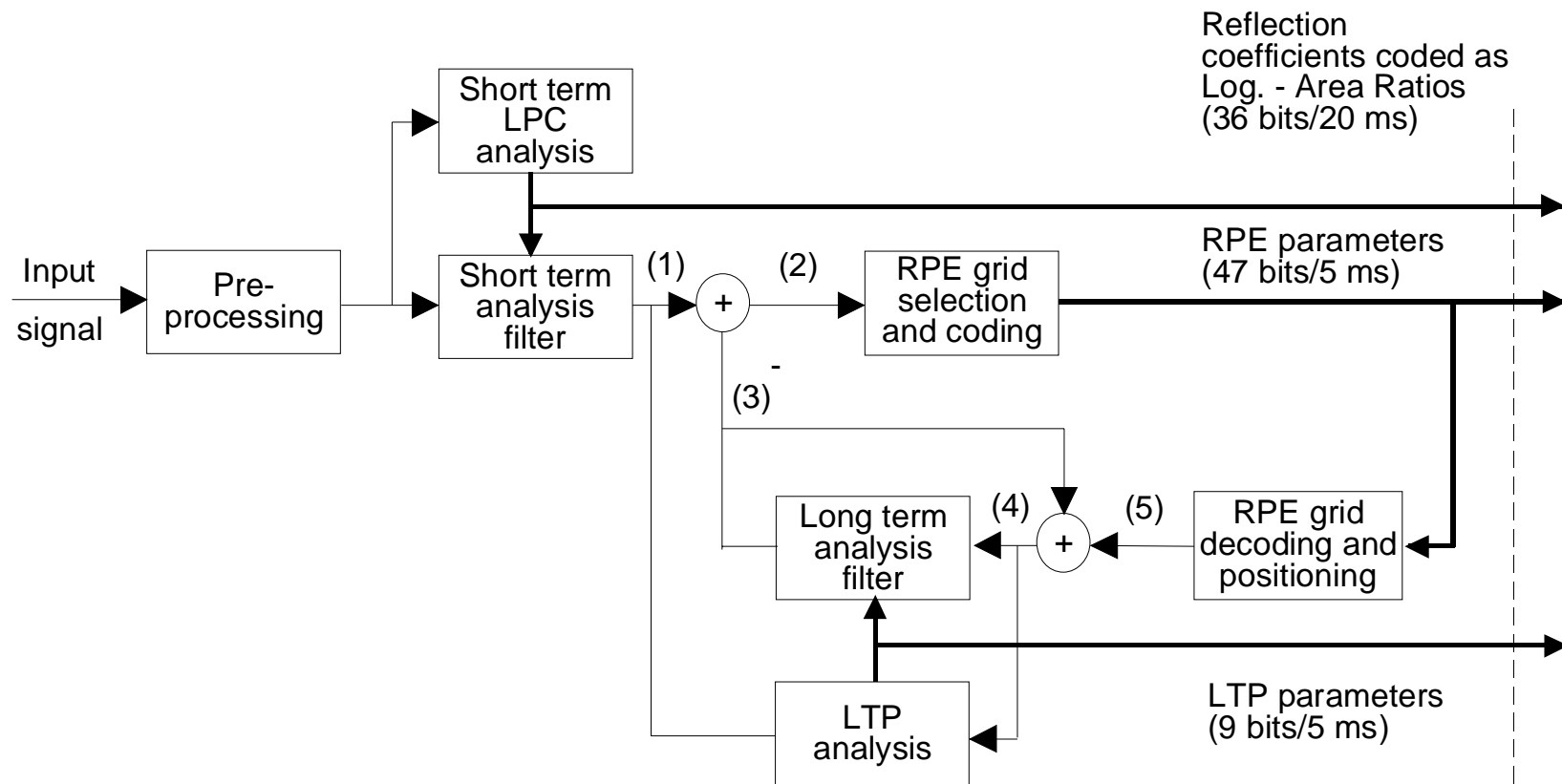
póly přenosové funkce: $\frac{1}{A(z)}$ a $\frac{1}{A(z/\gamma)}$:



Kódování buzení v kratších rámcích

Zatímco LP analýza probíhá v rámcích “obvyklé” délky (běžně 20 ms – 160 vzorků pro $F_s = 8 \text{ kHz}$), buzení je často kódováno v kratších rámcích - typicky 40 vzorků.

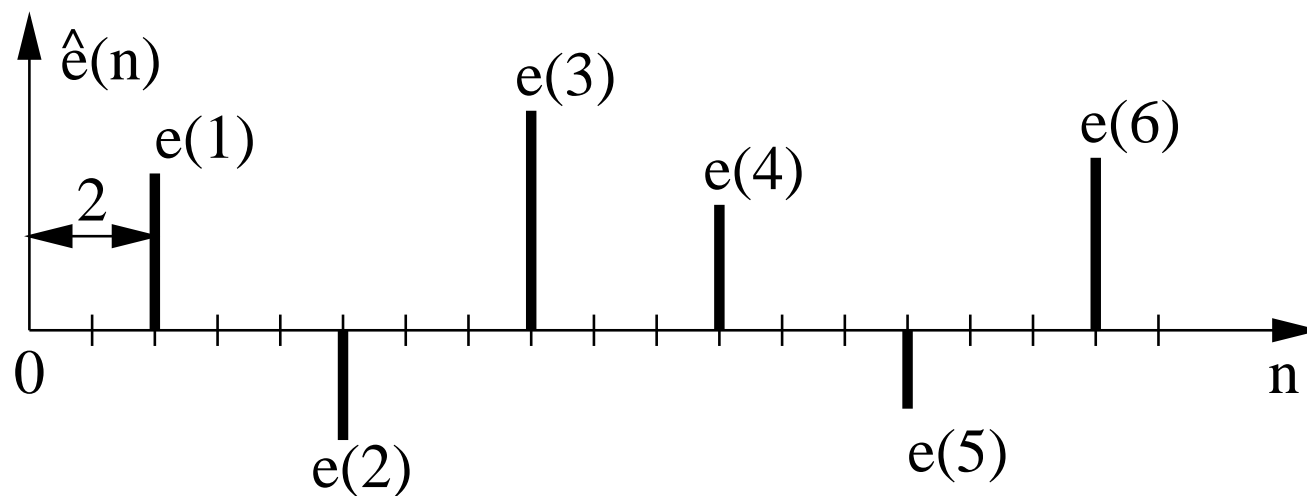
KODÉR I: RPE-LTP



- (1) Short term residual
- (2) Long term residual (40 samples)
- (3) Short term residual estimate (40 samples)
- (4) Reconstructed short term residual (40 samples)
- (5) Quantized long term residual (40 samples)

To
radio
subsystem

- Regular-Pulse Excitation, Long Term Prediction, GSM full-rate ETSI 06.10
- Krátkodobá analýza (v rámcích 20 ms 160 vzorků), koeficienty LPC filtru převedeny na 8 LAR.
- dlouhodobá analýza (LTP) - v rámcích 5 ms 40 vzorků) - lag a gain.
- Buzení kódováno v rámcích o 40 vzorcích kódováno tak, že je chybový signál podvzorkován s faktorem 3 (14,13,13), a je kvantována pouze poloha prvního impulsu (0,1,2,3(!))
- velikosti jednotlivých impulsů jsou kvantovány pomocí APCM.



- výsledkem je rámeček o 260 bitech \times 50 = 13 kbit/s.
- více viz norma 06.10, ke stažení z <http://pda.etsi.org>

Dekodér RPE-LTP

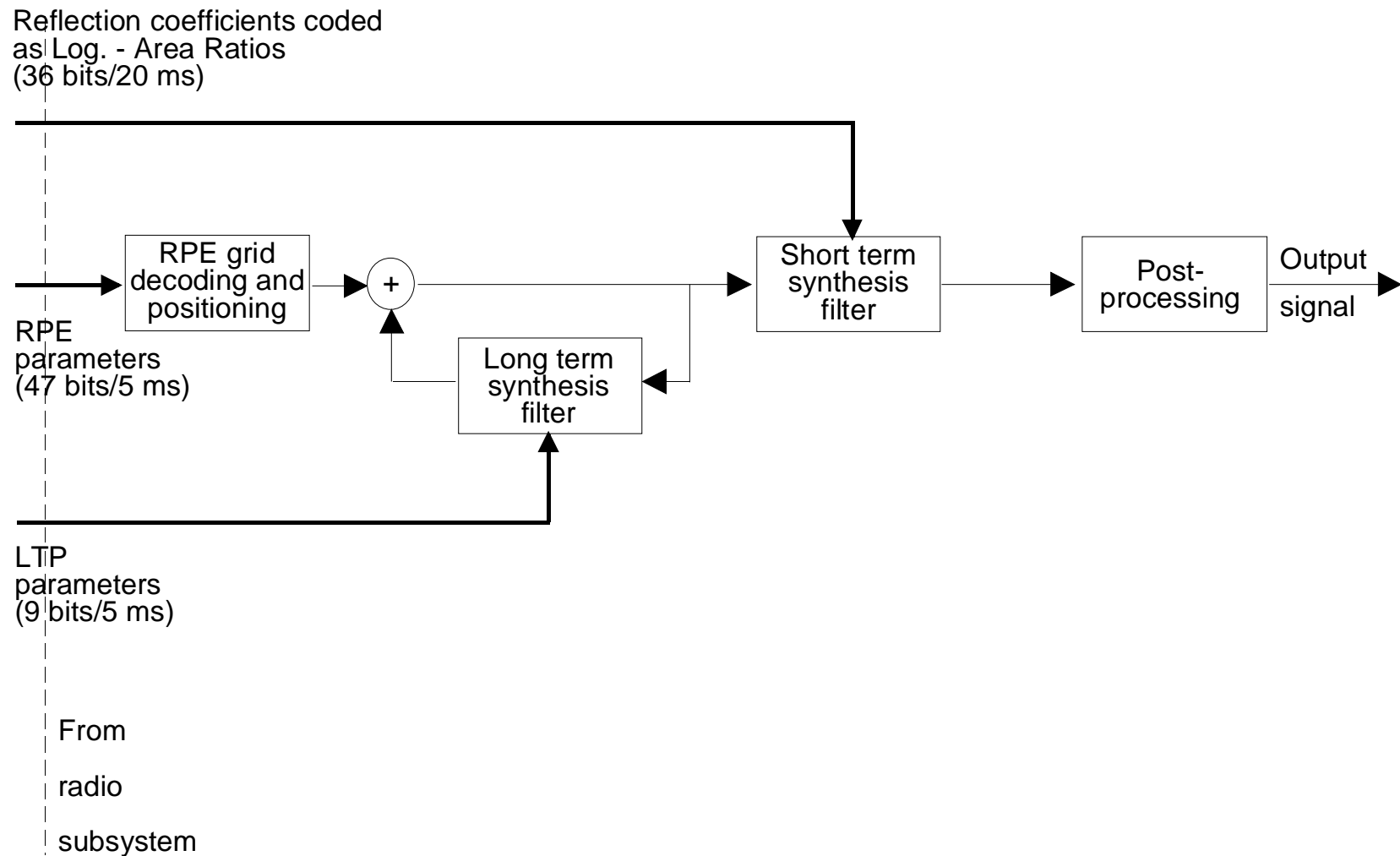
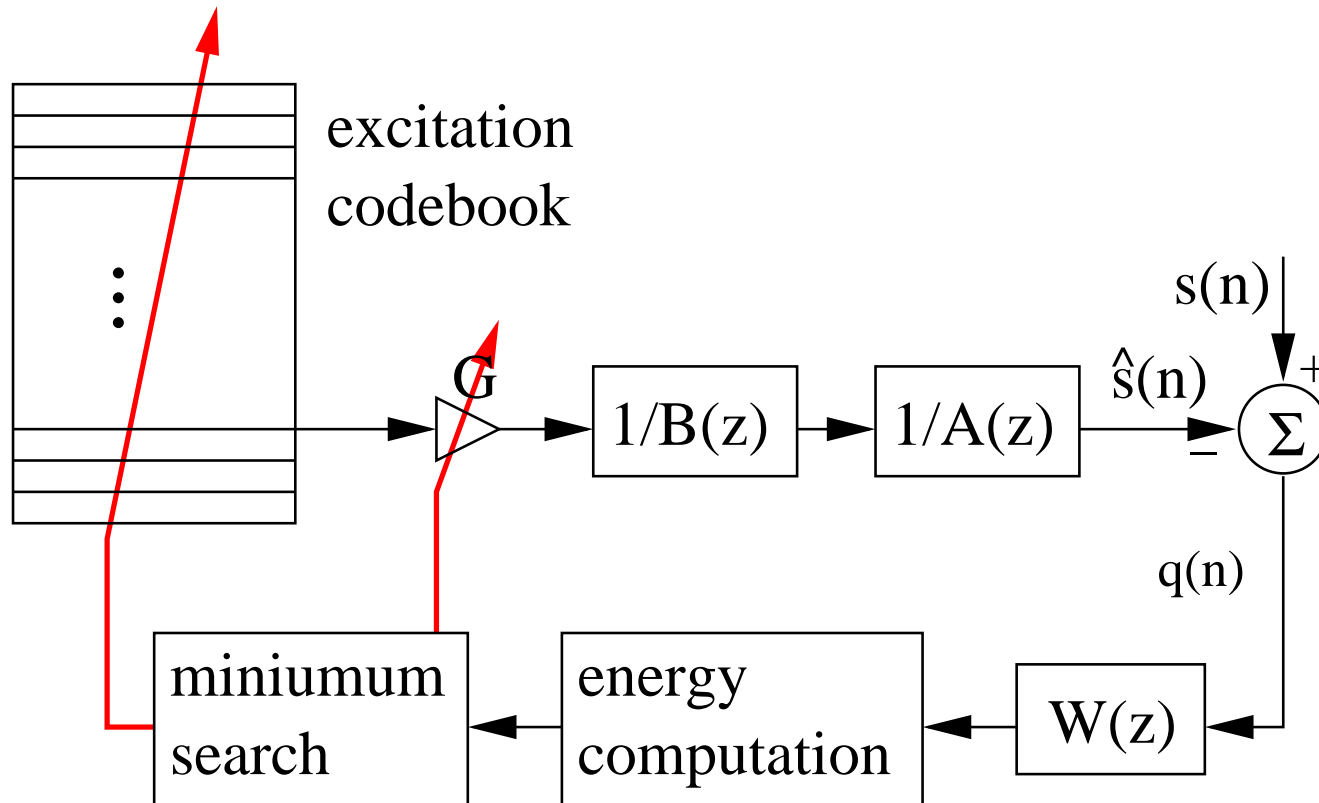


Figure 1.2: Simplified block diagram of the RPE - LTP decoder

CELP – Codebook-Excited Linear Prediction

Buzení se kóduje pomocí **kódových knih** - **codebooků**.

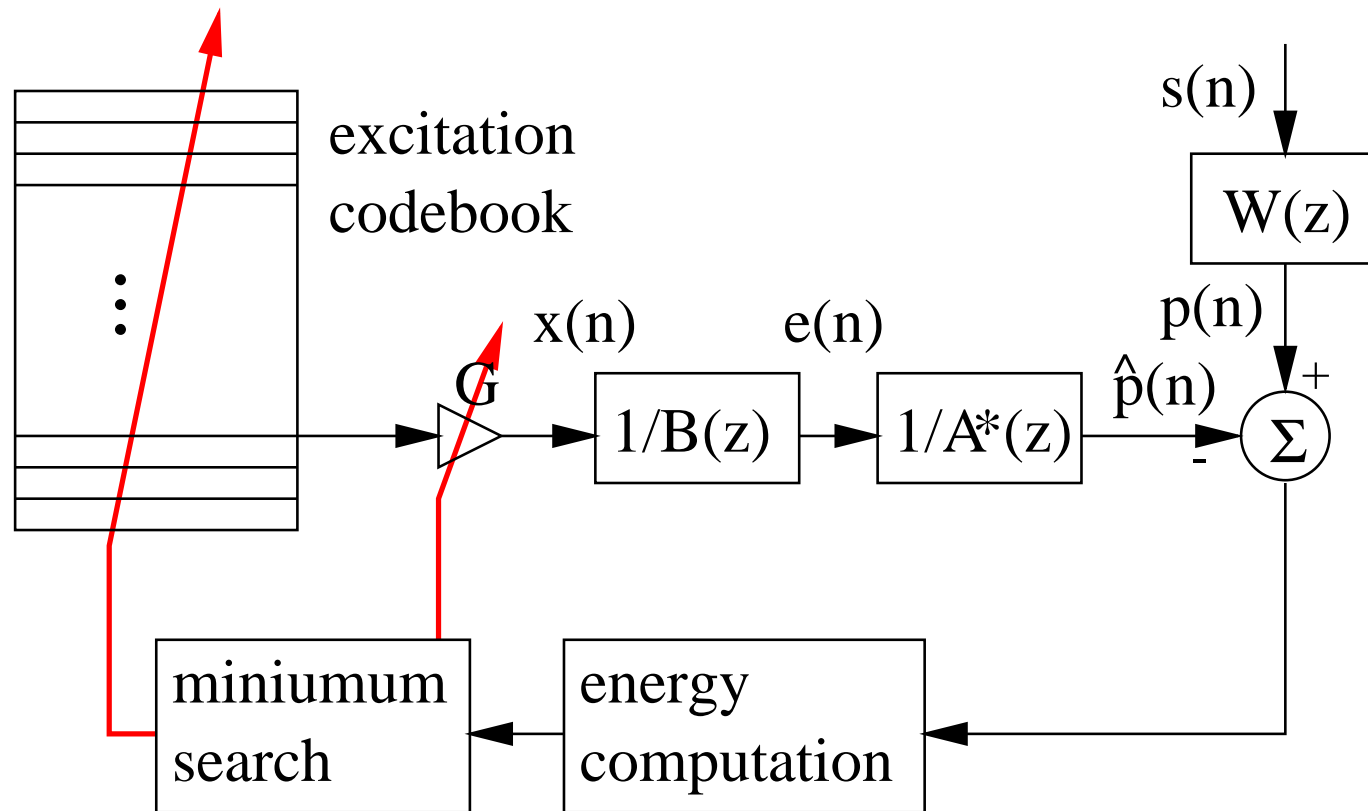
Základní struktura s perceptuálním filtrem:



... Každý testovaný signál musí být filtrován $W(z) = \frac{A(z)}{A^*(z)}$ — moc práce !

Perceptuální filtr na vstupu

Filtrování je lineární operace, takže $W(z)$ můžeme přesunout do obou větví: na vstup a do signálu za sekvencí $\frac{1}{B(z)} - \frac{1}{A(z)}$. Můžeme zjednodušit na: $\frac{1}{A(z)} \frac{A(z)}{A^*(z)} = \frac{1}{A^*(z)}$ — toto je nový filtr, který musíme použít za $\frac{1}{B(z)}$.



Odstranění odezvy filtru naprázdno.

Filter $\frac{1}{A^*(z)}$ neodpovídá jen na buzení, které se zuřivě snažíme najít, ale má i vlastní paměť (příspěvek od minulých rámců). Označíme jeho impulsní odezvu $h(i)$:

$$\hat{p}(n) = \underbrace{\sum_{i=0}^{n-1} h(i)e(n-i)}_{\text{tento rámeček}} + \underbrace{\sum_{i=n}^{\infty} h(i)e(n-i)}_{\text{minulé rámečky } \hat{p}_0(n)} \quad (2)$$

Odezva od minulých rámců nezávisí na tom, co teď hledáme – můžeme ji spočítat jen jednou a odečíst od porovnávaného signálu (vstup filtrovaný $W(z)$), dostaneme:

$$\hat{p}(n) - \hat{p}_0(n) = \sum_{i=0}^{n-1} h(i)e(n-i) \quad (3)$$

Další postup: vezmeme v úvahu long-term predictor:

$$\frac{1}{B(z)} = \frac{1}{1 - bz^{-M}}, \quad (4)$$

kde M je optimální lag, můžeme zapsat $e(n)$ jako

$$e(n) = x(n) + be(n - M), \quad (5)$$

v rovnici pro filtrování dostaneme místo $e(n - i)$:

$$\hat{p}(n) - \hat{p}_0(n) = \sum_{i=0}^{n-1} h(i)[x(n - i) + be(n - M - i)]. \quad (6)$$

to dále rozložíme:

$$\hat{p}(n) - \hat{p}_0(n) = \sum_{i=0}^{n-1} h(i)x(n - i) + b \sum_{i=0}^{n-1} h(i)e(n - M - i), \quad (7)$$

protože filtrování je lineární operace. Druhý výraz je ve skutečnosti **minulé buzení** filtrované $\frac{1}{A^*(z)}$ a násobené LTP gainem b . Namísto LTP si v CELP kodéru můžeme představit druhou kódovou knihu, která bude obsahovat minulá buzení. Zde si ji představujeme jako skutečnou kódovou knihu, s řádky $e(n - M)$ (nebo $e^{(M)}$ ve vektorové notaci), v reálných aplikacích je to prostě kus zpožděného budícího signálu.

CELP rovnice

$$\hat{p}(n) - \hat{p}_0(n) = \sum_{i=0}^{n-1} h(i)e(n-i) = \sum_{i=0}^{n-1} h(i) \left[g^{(j)} u^{(j)}(n-i) + be(n-i-M) \right] = \hat{p}_2(n) + \hat{p}_1(n), \quad (8)$$

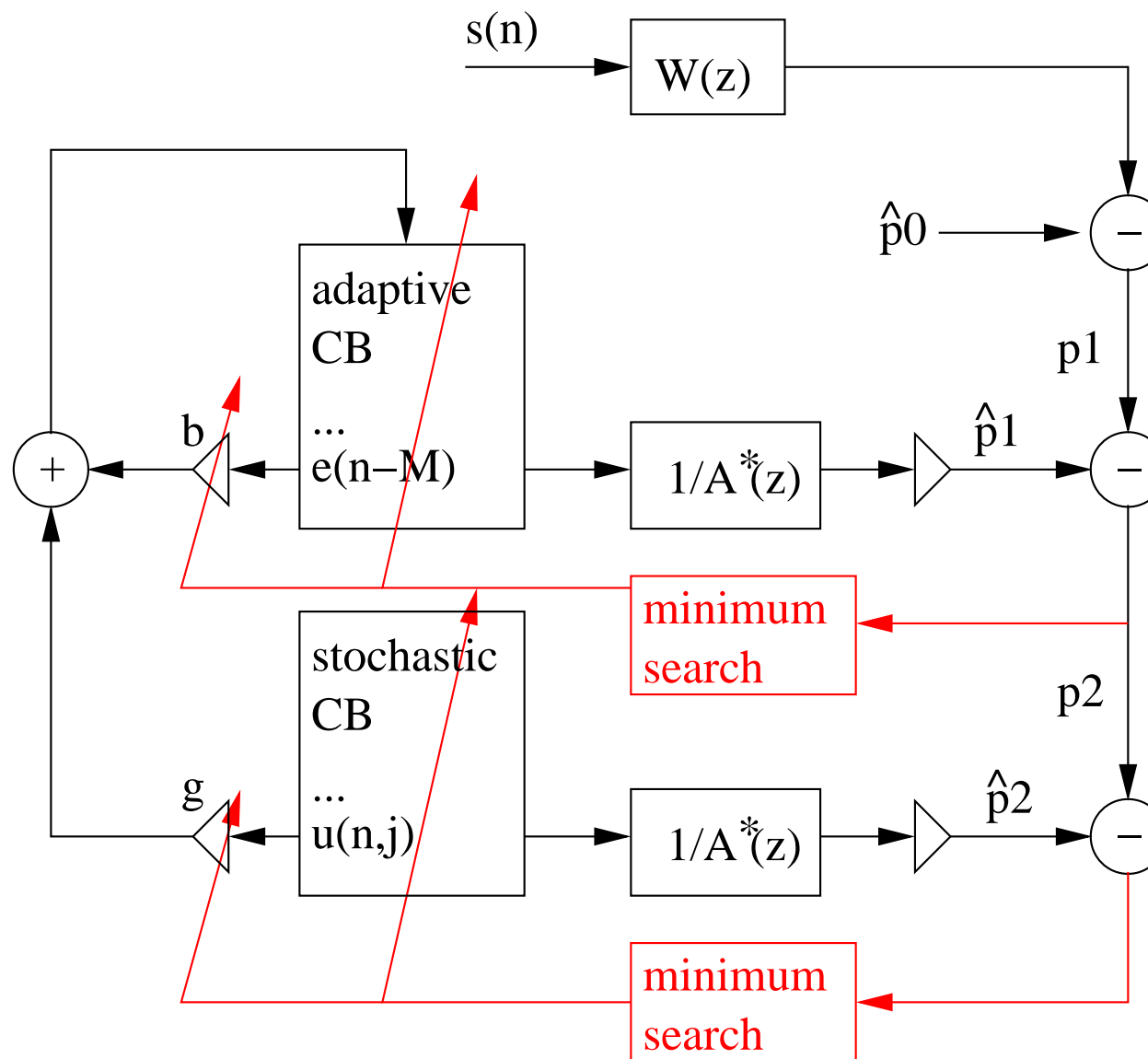
Úkol:

- musíme najít gain pro stochastickou kódovou knihu g
- nejlepší vektor ze stochastické kódové knihy $\mathbf{u}^{(j)}$
- gain adaptivní kódové knihy b
- nejlepší vektor z adaptivní kódové knihy $\mathbf{e}^{(M)}$.

Teoreticky bychom měli zkoušet všechny kombinace \Rightarrow ne ! Suboptimální procedura:

- nejprve nalezneme, $\hat{p}_1(n)$ hledáním v adaptivní kódové knize. Výsledkem je lag M a gain b .
- Odečteme $\hat{p}_1(n)$ od $p_1(n)$ a dostaneme “chybový signál druhé generace”, který by nám měla vygenerovat stochastická kódová kniha.
- Najdeme $\hat{p}_2(n)$ hledáním ve stoch. kódové knize. Výsledkem je index j a gain g .
- Na konci se obvykle jen re-optimalizují gainy (příspěvky obou kódových knih).

Finální struktura CELP:



CELP rovnice – pouze pro tvrdáky – nepovinné...

Searching the adaptive codebook

We want to minimize the error E_1 that is (on one frame):

$$E_1 = \sum_{N=0}^N (p_1(n) - \hat{p}_1(n))^2. \quad (9)$$

We can rewrite in more convenient vector notation and develop \hat{p}_1 into the gain g and excitation signal $q(n - M)$, where $q(n)$ is the filtered version of $e(n - M)$:

$$q(n) = \sum_{i=0}^{n-1} h(i)e(n - i - M) \quad (10)$$

When we multiply this signal by the gain b , we will get the signal

$$\hat{p}_1(n) = bq(n). \quad (11)$$

Our goal is to find the minimum energy:

$$E_1 = \|\mathbf{p}_1 - \hat{\mathbf{p}}_1\|^2 = \|\mathbf{p}_1 - b\mathbf{q}^{(M)}\|^2. \quad (12)$$

which we can rewrite using inner products of involved vectors:

$$E_1 = \langle \mathbf{p}_1 - b\mathbf{q}^{(M)}, \mathbf{p}_1 - b\mathbf{q}^{(M)} \rangle = \langle \mathbf{p}_1, \mathbf{p}_1 \rangle - 2b \langle \mathbf{p}_1, \mathbf{q}^{(M)} \rangle + b^2 \langle \mathbf{q}^{(M)}, \mathbf{q}^{(M)} \rangle. \quad (13)$$

We have to begin by finding an optimal gain for each M by a derivation with respect to b (the derivation must be zero for minimum energy):

$$\frac{\delta}{\delta b} \left[\langle \mathbf{p}_1, \mathbf{p}_1 \rangle - 2b \langle \mathbf{p}_1, \mathbf{q}^{(M)} \rangle + b^2 \langle \mathbf{q}^{(M)}, \mathbf{q}^{(M)} \rangle \right] = 0 \quad (14)$$

from here, we can directly write the result for lag M :

$$b^{(M)} = \frac{\langle \mathbf{p}_1, \mathbf{q}^{(M)} \rangle}{\langle \mathbf{q}^{(M)}, \mathbf{q}^{(M)} \rangle} \quad (15)$$

We need now to determine the optimal lag. We will substitute found $b^{(M)}$ in Eq. 13 and we obtain: the minimization of

$$\min \left[\langle \mathbf{p}_1, \mathbf{p}_1 \rangle - \frac{2 \langle \mathbf{p}_1, \mathbf{q}^{(M)} \rangle \langle \mathbf{p}_1, \mathbf{q}^{(M)} \rangle}{\langle \mathbf{q}^{(M)}, \mathbf{q}^{(M)} \rangle} + \frac{\langle \mathbf{p}_1, \mathbf{q}^{(M)} \rangle^2 \langle \mathbf{q}^{(M)}, \mathbf{q}^{(M)} \rangle}{\langle \mathbf{q}^{(M)}, \mathbf{q}^{(M)} \rangle^2} \right]. \quad (16)$$

we can drop the first term, as it does not influence the minimization and convert minimization into the maximization of:

$$M = \arg \max \frac{\langle \mathbf{p}_1, \mathbf{q}^{(M)} \rangle^2}{\langle \mathbf{q}^{(M)}, \mathbf{q}^{(M)} \rangle} \quad (17)$$

Searching the stochastic codebook

First, all codebook vectors $\mathbf{u}^{(j)}(n)$ must be filtered by the filter $\frac{1}{A^*(z)} \longrightarrow \mathbf{q}^{(j)}$. Then, we have to search $\hat{\mathbf{p}}_2(n)$ minimizing the energy of error:

$$E_2 = \|\mathbf{p}_2 - \hat{\mathbf{p}}_2\|^2 = \|\mathbf{p}_2 - g\mathbf{q}^{(j)}\|^2$$

Solution:

- $\langle \mathbf{p}_2 - g\mathbf{q}^{(j)}, \mathbf{q}^{(j)} \rangle = 0$

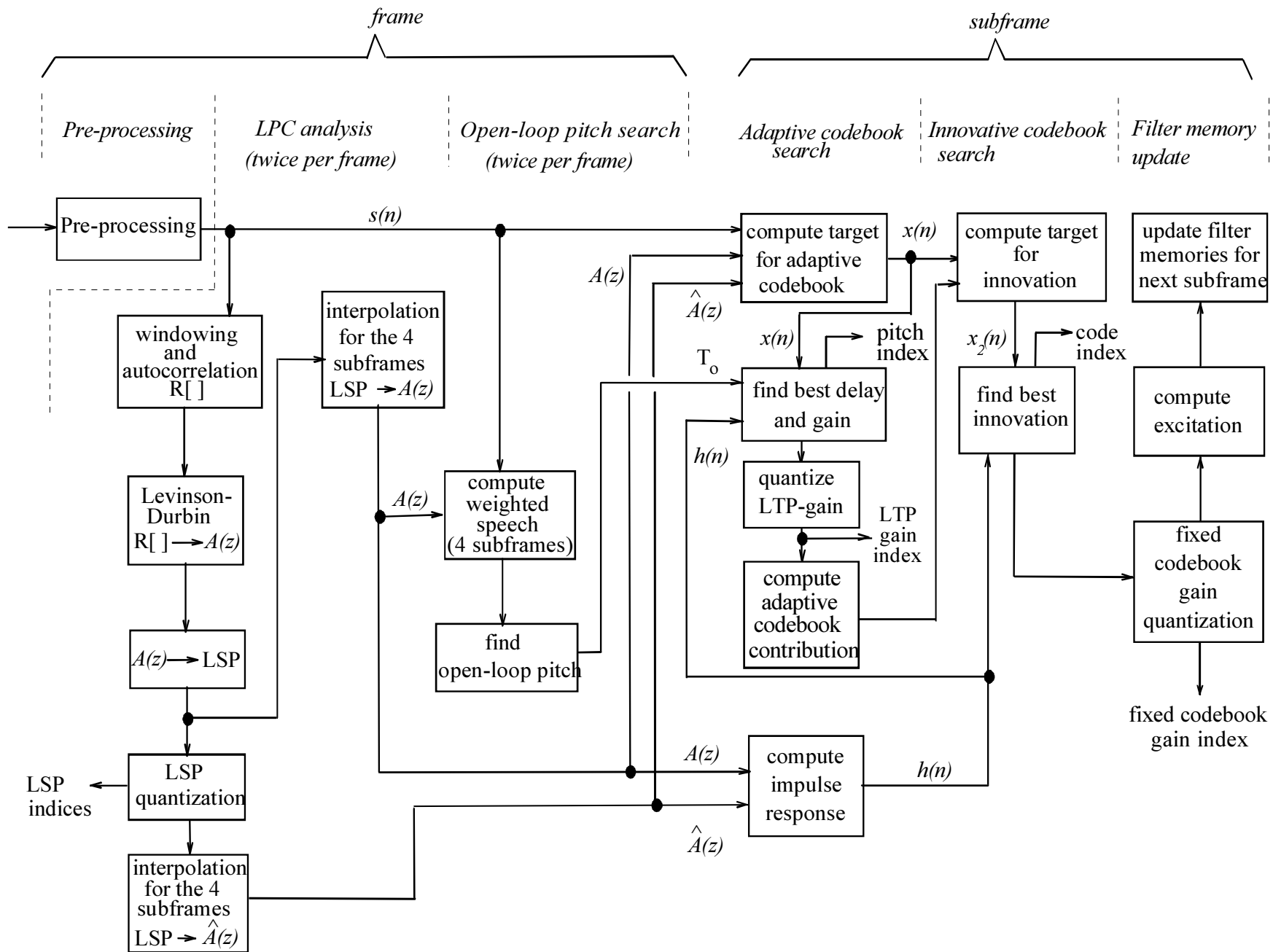
$$g = \frac{\langle \mathbf{p}_2, \mathbf{q}^{(j)} \rangle}{\langle \mathbf{q}^{(j)}, \mathbf{q}^{(j)} \rangle}.$$

- $\min E_2 = \min \langle \mathbf{p}_2 - g\mathbf{q}^{(j)}, \mathbf{p}_2 \rangle = \|\mathbf{p}_2\|^2 - \frac{\langle \mathbf{p}_2, \mathbf{q}^{(j)} \rangle^2}{\langle \mathbf{q}^{(j)}, \mathbf{q}^{(j)} \rangle}.$

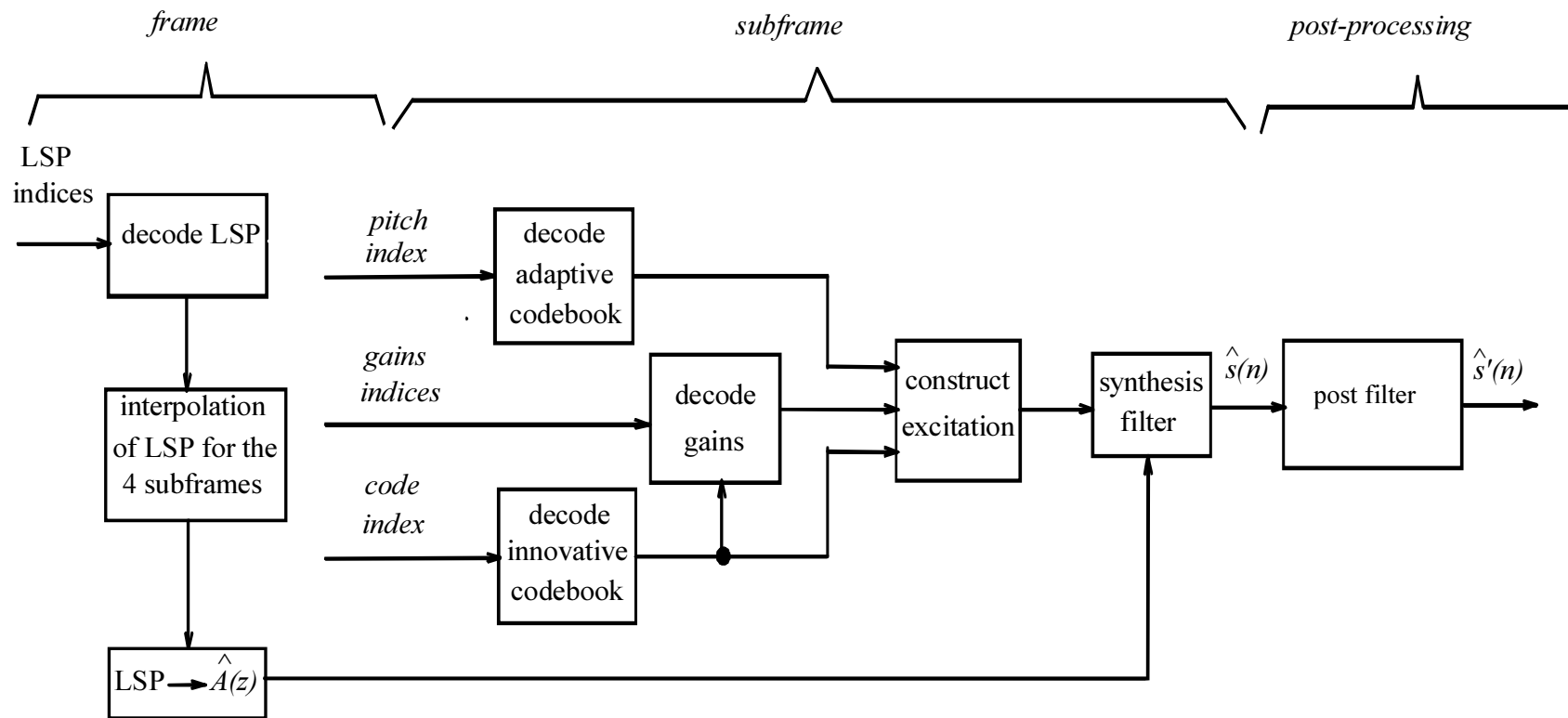
$$j = \arg \max \frac{\langle \mathbf{p}_2, \mathbf{q}^{(j)} \rangle^2}{\langle \mathbf{q}^{(j)}, \mathbf{q}^{(j)} \rangle}.$$

Příklad CELP kodéru ACELP – GSM EFR

- klasický CELP s “inteligentní kódovou knihou”.
- Algebraic Codebook Excited Linear Prediction GSM Enhanced Full-rate ETSI 06.60



- opět rámce 20 ms (160 vzorků)
- Krátkodobý prediktor – 10 koeficientů a_i ve dvou sub-rámcích, převedeny na line-spectral pairs, ze dvou sub-rámců společně kvantovány pomocí split-matrix quantization (SMQ).
- 4 sub-rámce po 40 vzorcích (5 ms) pro buzení.
- Odhad lagu nejprve open-loop, potom closed-loop okolo hrubého odhadu, fractional pitch s rozlišením $1/6$ vzorku.
- stochastický codebook: algebraická kódová kniha - může obsahovat pouze 10 nenulových impulsů, které mohou být pouze +1 nebo -1 \Rightarrow rychlé vyhledávání (rychlé korelace - pouze sčítání, ne násobení), atd.
- 244 bitů na rámeček $\times 50 = 12.2$ kbit/s.
- více viz norma 06.60, ke stažení z <http://pda.etsi.org>
- dekodér. . .



K dalšímu čtení

- Andreas Spanias (Arizona University):
<http://www.eas.asu.edu/~spanias>
v sekci Publications/Tutorial Papers je ke stažení výborná přehledová práce: “Speech Coding: A Tutorial Review”, z níž byla část otištěna v Proceedings of the IEEE, Oct. 1994.
- na téže stránce v sekci Software/Tools/Demo - Matlab Speech Coding Simulations naleznete software pro FS1015, FS1016, RPE-LTP a další. Příjemné hraní.
- Normy ETSI pro mobilní telefony jsou zdarma ke stažení z:
<http://pda.etsi.org/pda/queryform.asp>
Jako klíčová slova můžete zadat např “gsm half rate speech”. K mnoha normám jsou k dispozici také zdrojové kódy kodérů v jazyce C.