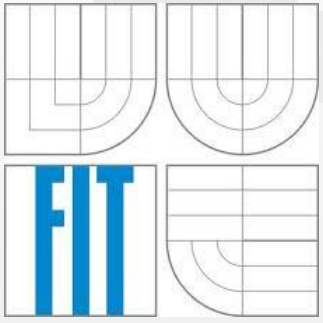
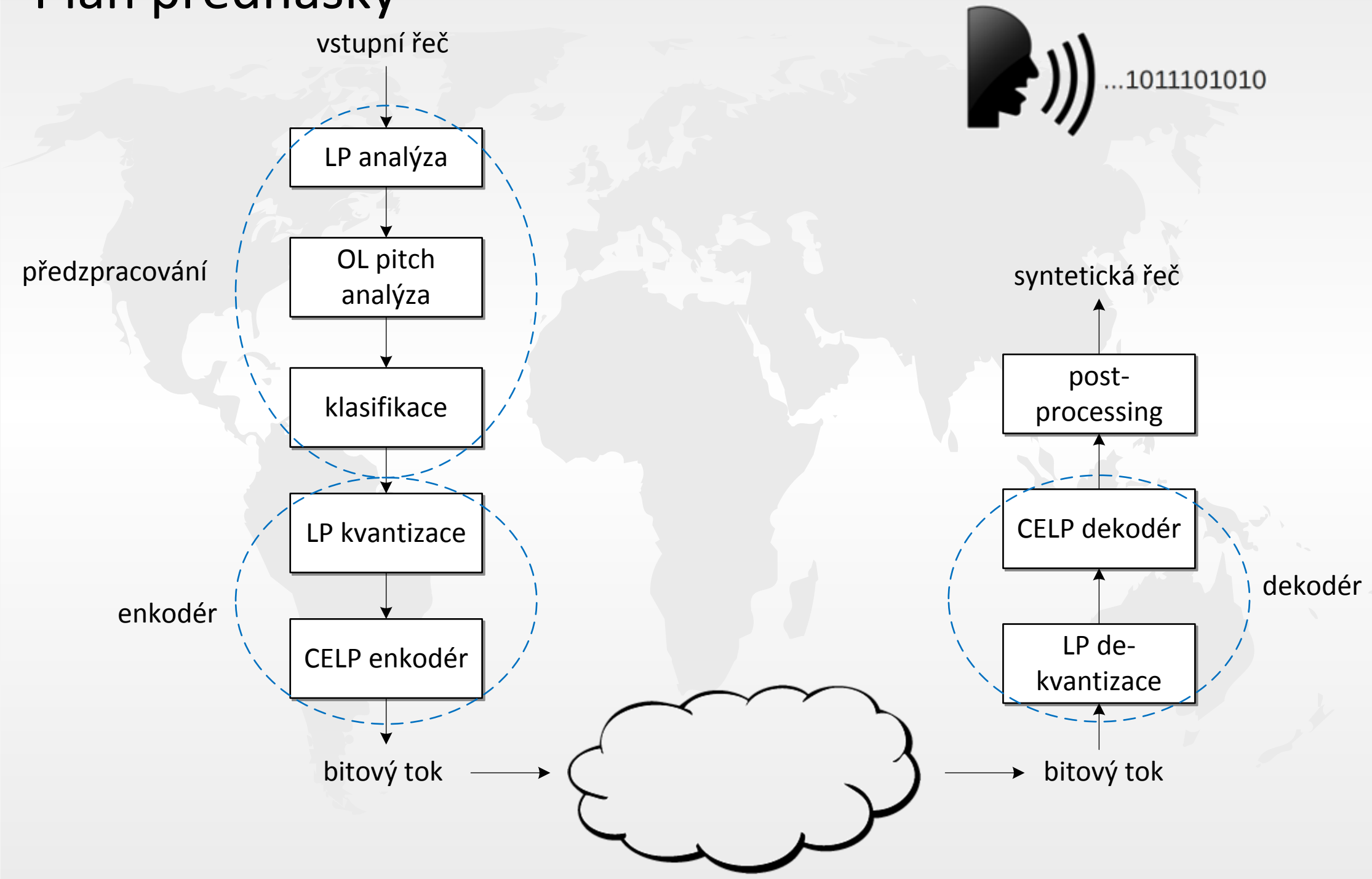
ZRE - Kódování řeči II.

CELP

Vladimír Malenovský, ÚPGM FIT VUT Brno

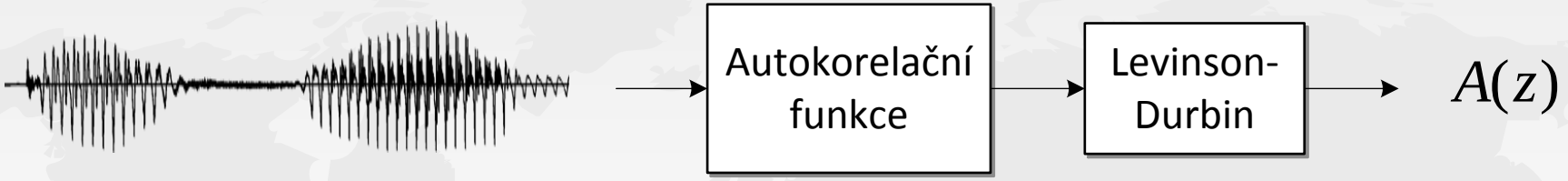


Plán přednášky



LP analýza/syntéza

LP analýza



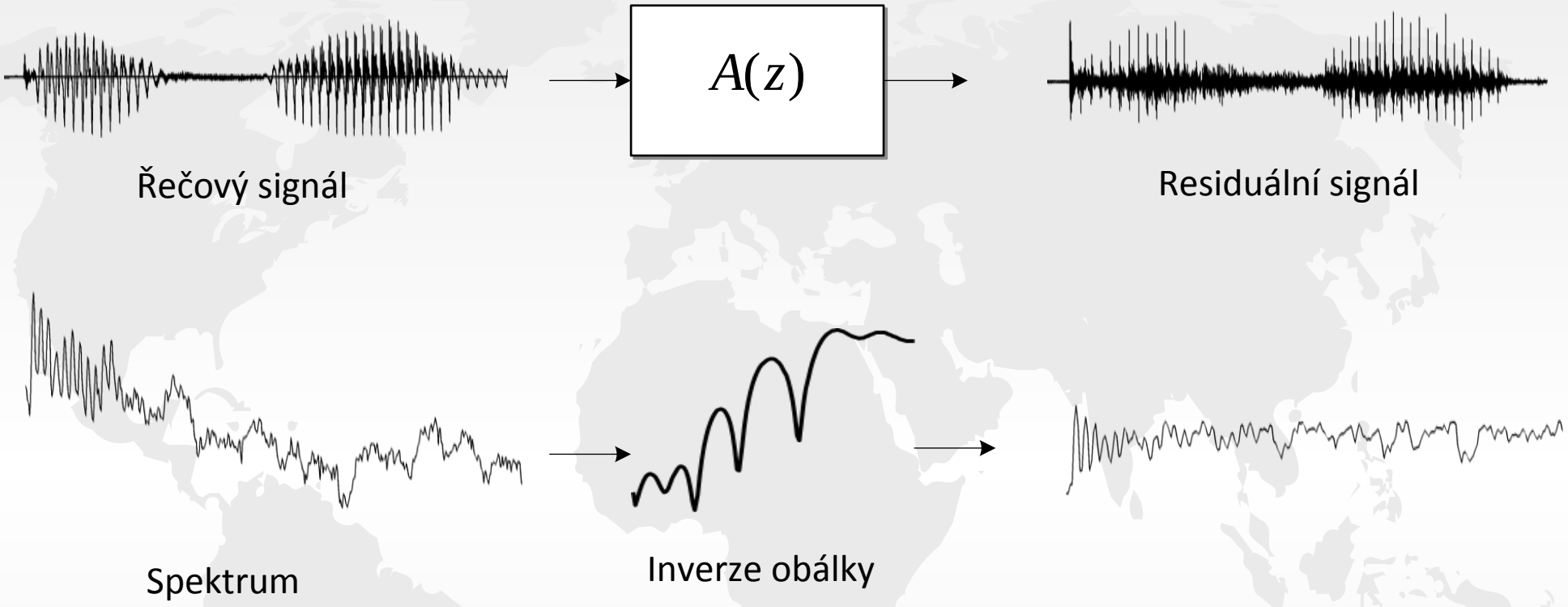
Řečový signál

Spektrum

Obálka

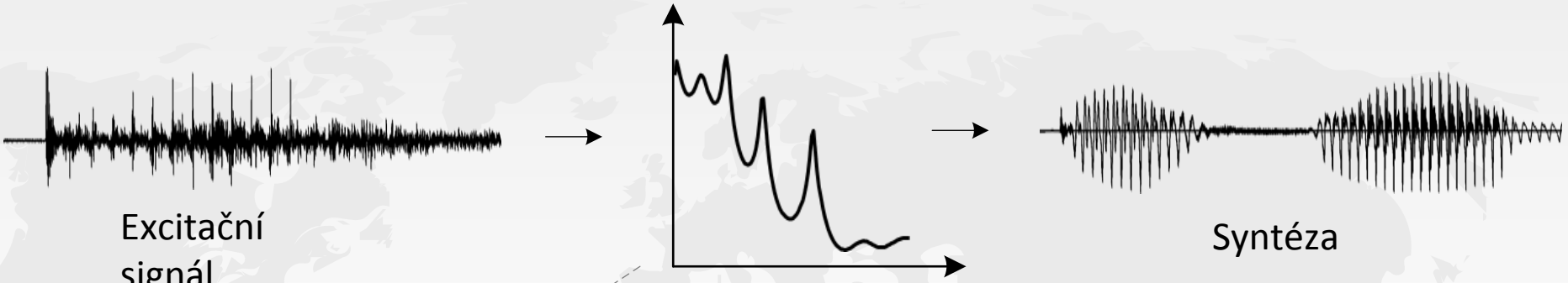
- LP analýza se provádí na krátkých úsecích vstupního signálu (5-30ms)
- Koeficienty filtru získané na základě LP analýzy se kódují v podobě LSP/LSF parametrů
- Poloha pólů filtru 1/A(z) odpovídá formantům ve spektru
- Řád filtru P=15-20 postačuje bohatě pro řeč či šum, ne však např. pro hudbu
- Drtivá většina řečových kodeků používaných na světě využívá LP analýzu
- applet na http://web.mit.edu/6.302/www/pz/
- nebo na http://www.falstad.com/dfilter/index.html

LP analýza



- Filtrace A(z) odstraňuje z řečového signálu jeho „obálku“ a tím zbavuje spektrum formantů
- Reziduální (excitační) signál je dále kódován

LP syntéza

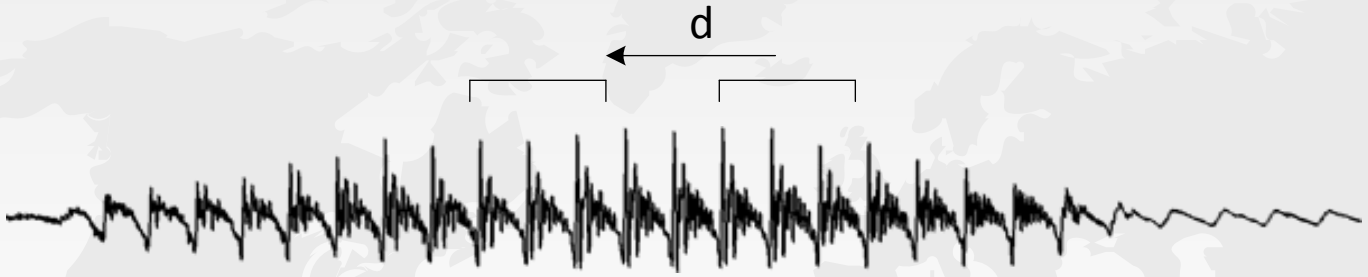


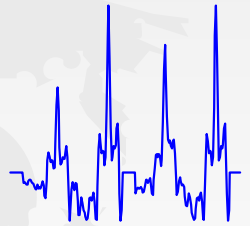
$$\frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^{P} a_i z^{-i}}$$

- Koeficienty LP filtru se kódují nejčastěji pomocí MSVQ.
- Pro filtr 16.řádu je zapotřebí cca 30 bitů/20ms, aby nedošlo ke zkreslení signálu.
- Filtr LP syntézy je IIR filtr, takže má vlastní paměť, což je několik posledních vzorků z minulé syntézy
- Vzhledem k charakteru filtru může dojít k jeho nestabilitě a „explozi“ syntézy

OL pitch analýza

OL pitch analýza





korelační funkce

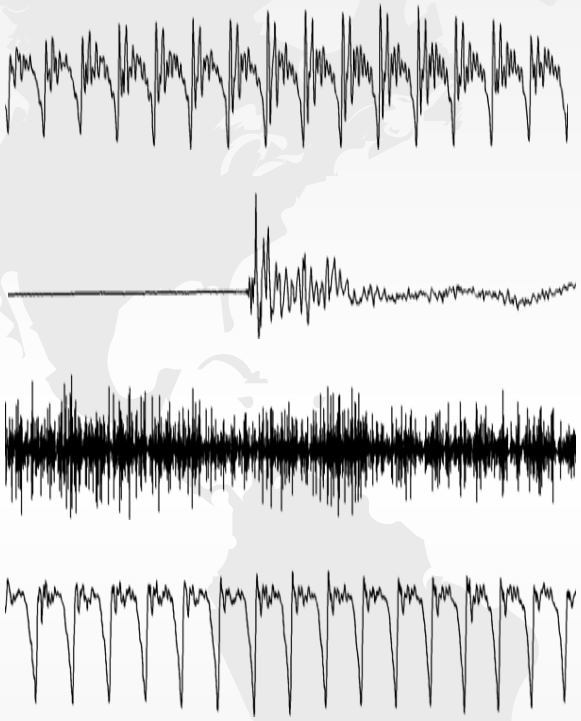
$$C_{norm}(d) = \frac{\sum_{n=0}^{L} s(n)s(n-d)}{\sqrt{\sum_{n=0}^{L} s^2(n) \sum_{n=0}^{L} s^2(n-d)}},$$

- korelační funkci počítáme pro d = <20;230>
- extremely low voice https://www.youtube.com/watch?v=AaPtiFO-NLc
- vybereme první „velké“ maximum, to prohlásíme za OL pitch
- snažíme se vyhnout násobkům OL pitch
- zapamatujeme si hodnotu $C(d_{max})$, to prohlásíme za OL voicing

Klasifikace

Typy excitačních signálů

řečový signál



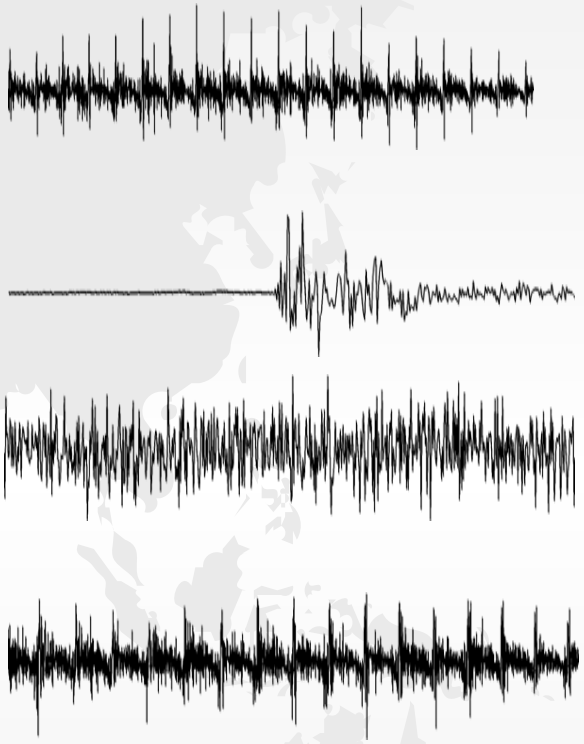
VOKÁLY (znělé úseky)
a, e, i, o, u

PLOZÍVY
t, d, k, g

FRIKATIVY (neznělé úseky)
s, z, š, h, ch, v, f

NEPÁROVÉ KONSONANTY
m, n, j, l, r

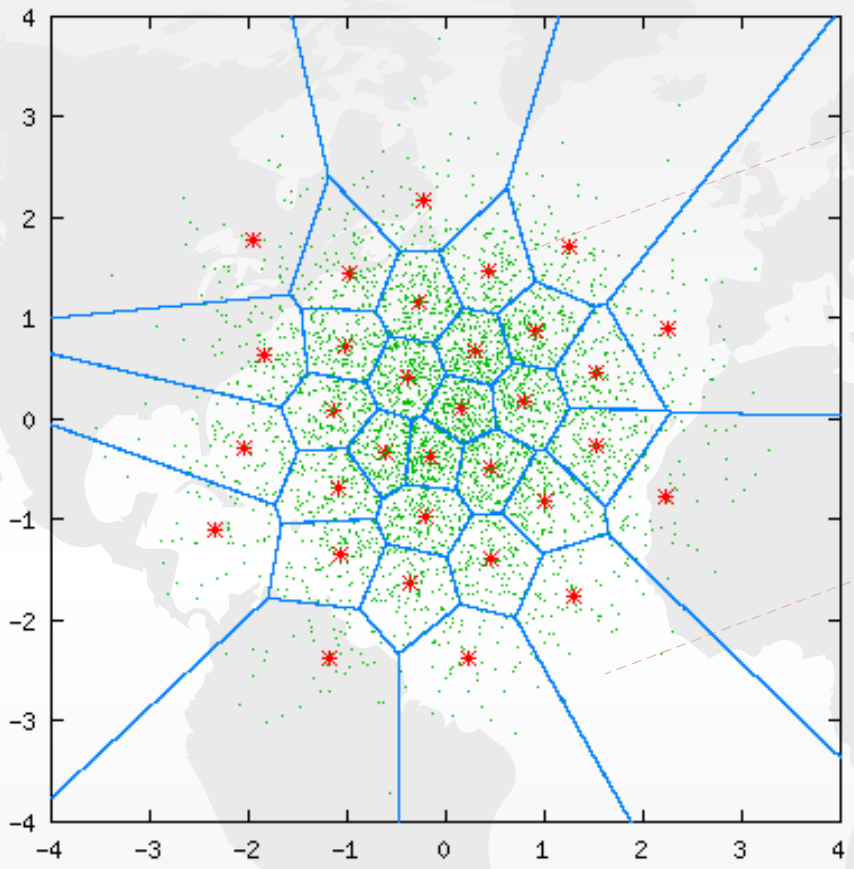
excitace



- klasifikace signálů na ZNĚLÉ, NEZNĚLÉ a OSTATNÍ
- výběr vhodného modelu pro kódování excitačního signálu

LP kvantizace

LSF kvantizace



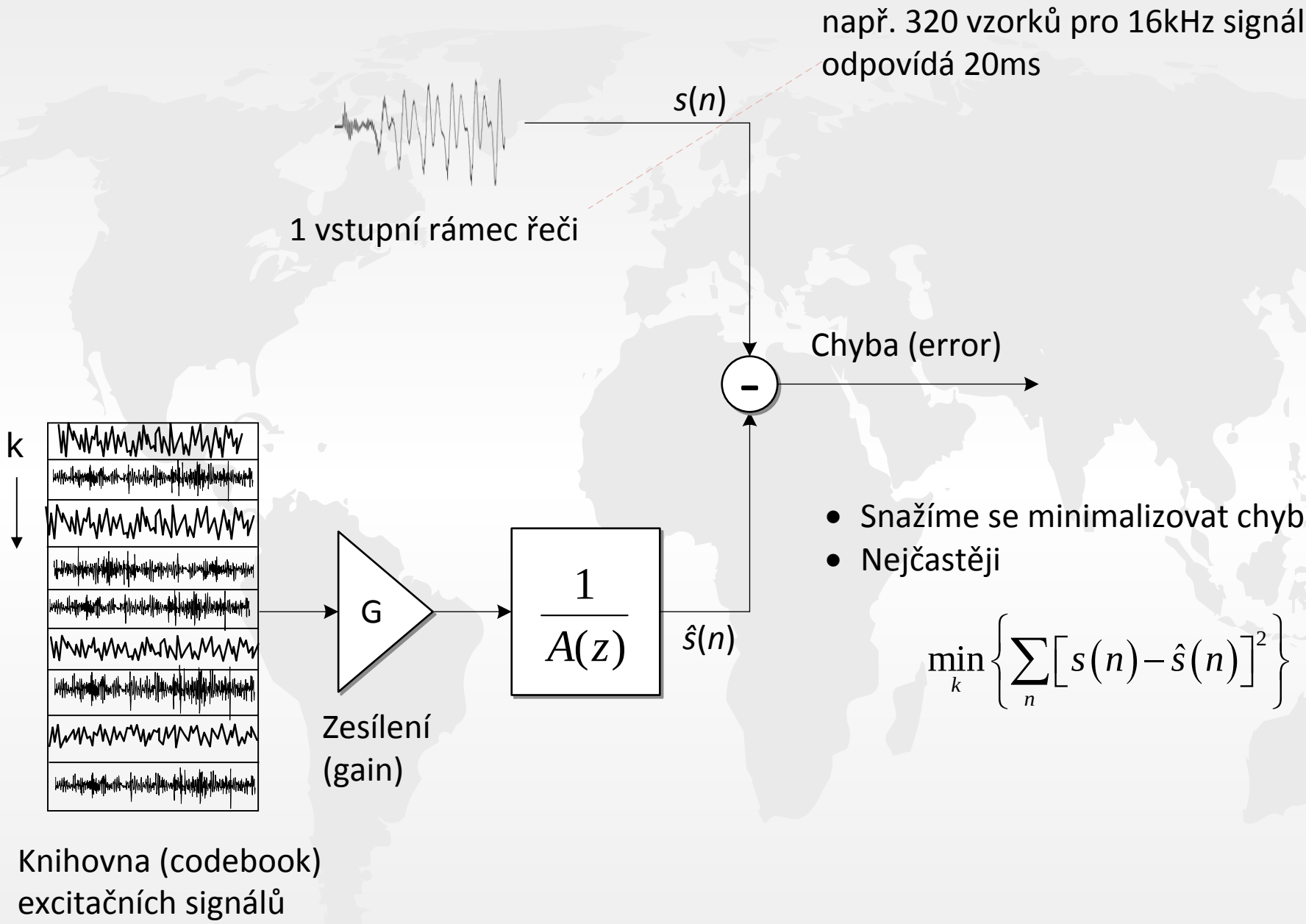
kódové slovo (codeword)

Voronoi region

- LP parametry se převádí na LSP parametry a ty se převádí na LSF parametry, které se kvantizují
- Vektorová kvantizace (MSVQ), k-means algorithm
- Animace VQ na http://www.data-compression.com/vqanim.shtml
- 20-30 bitů na jeden 16-ti dimenzionální LSF vektor


CELP

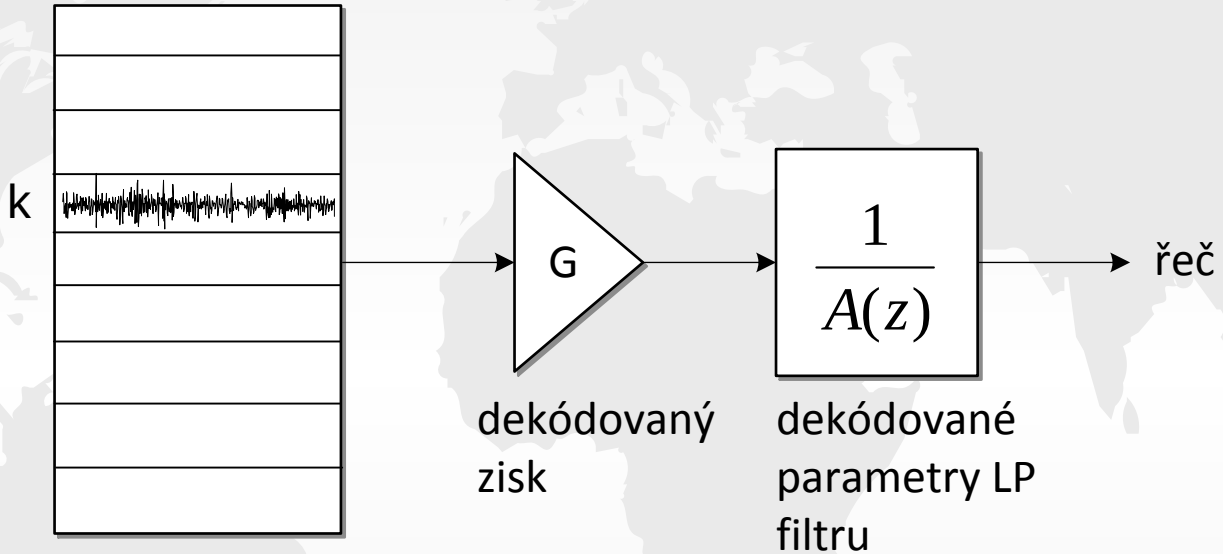
Kódování excitačního signálu



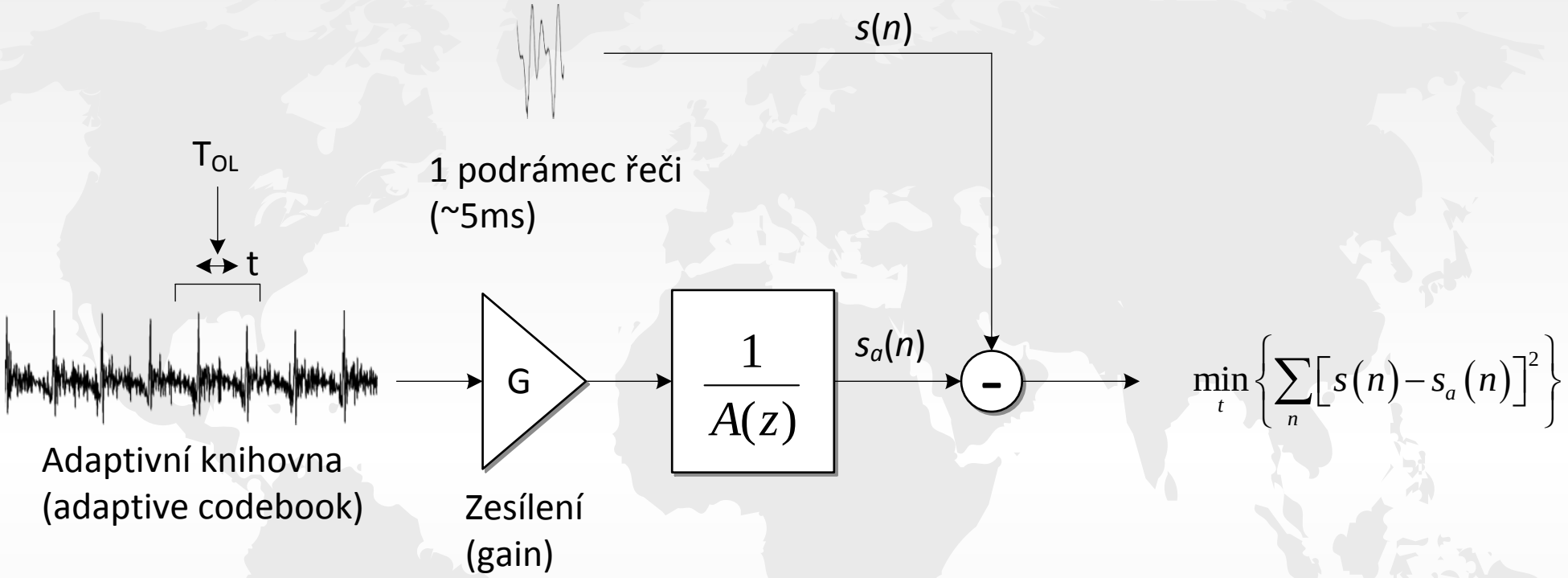
- Snažíme se minimalizovat chybu
- Nejčastěji

$$\min_k \left\{ \sum_n \left[s(n) - \hat{s}(n) \right]^2 \right\}$$

Dekódování excitačního signálu

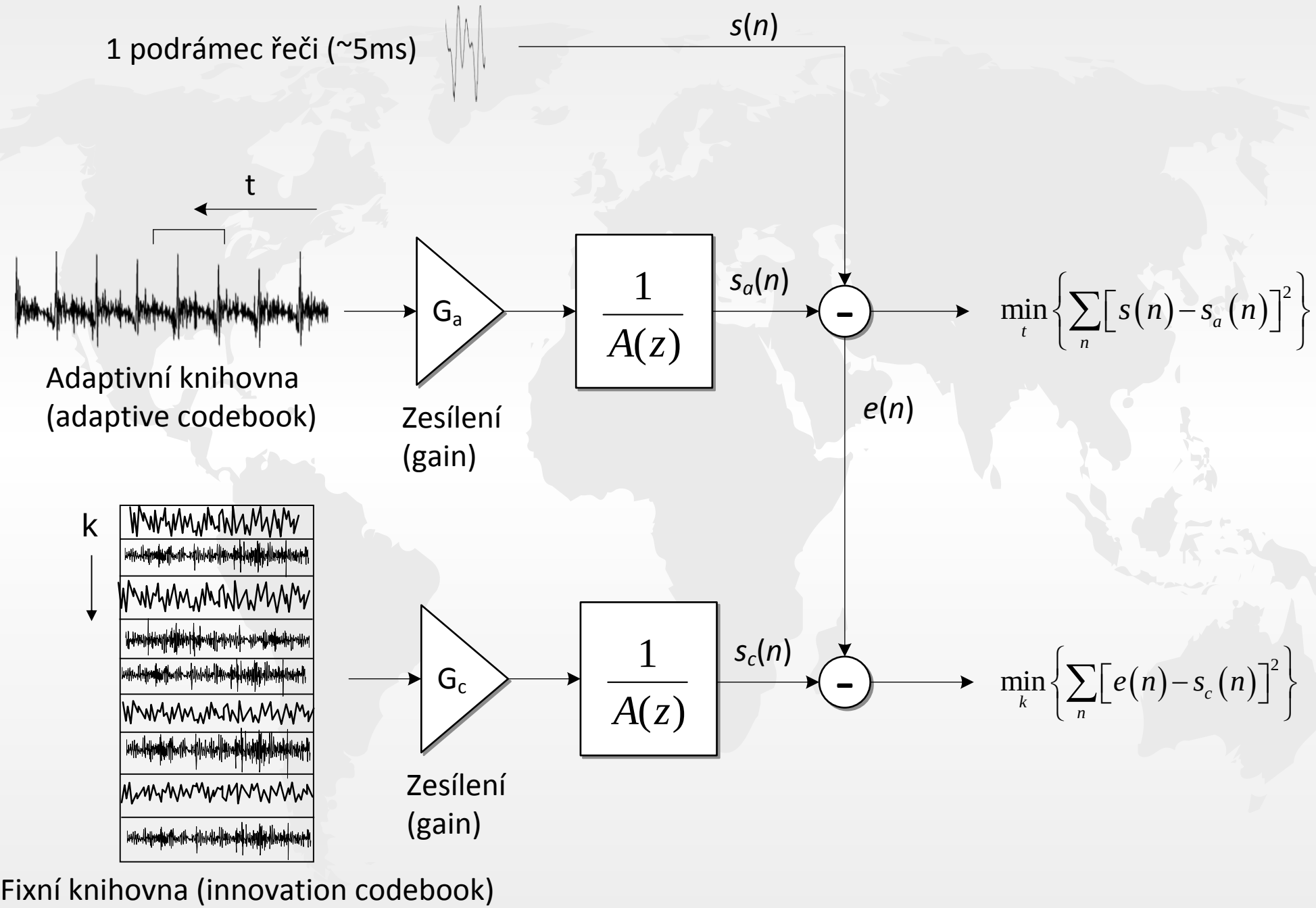


Zavedení adaptivní knihovny a podrámců

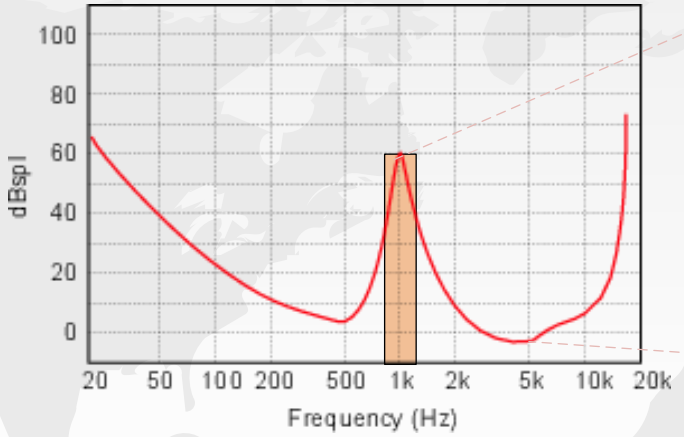


- adaptivní knihovna je v podstatě minulá excitace
- hledá se nejlepší úsek minulé excitace, který by mohl reprezentovat svým tvarem současný podrámec řeči
- korelace a minimalizace kvadrátu chyby okolo T_{OL} (open-loop pitch)
- estimace zesílení
- prohledávání adaptivní knihovny se dělá podle residuálního signálu

Zavedení adaptivní knihovny a podrámců

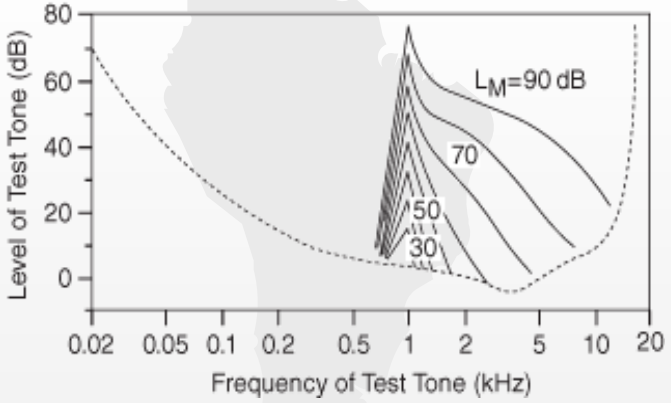


Maskování kvantizačního šumu



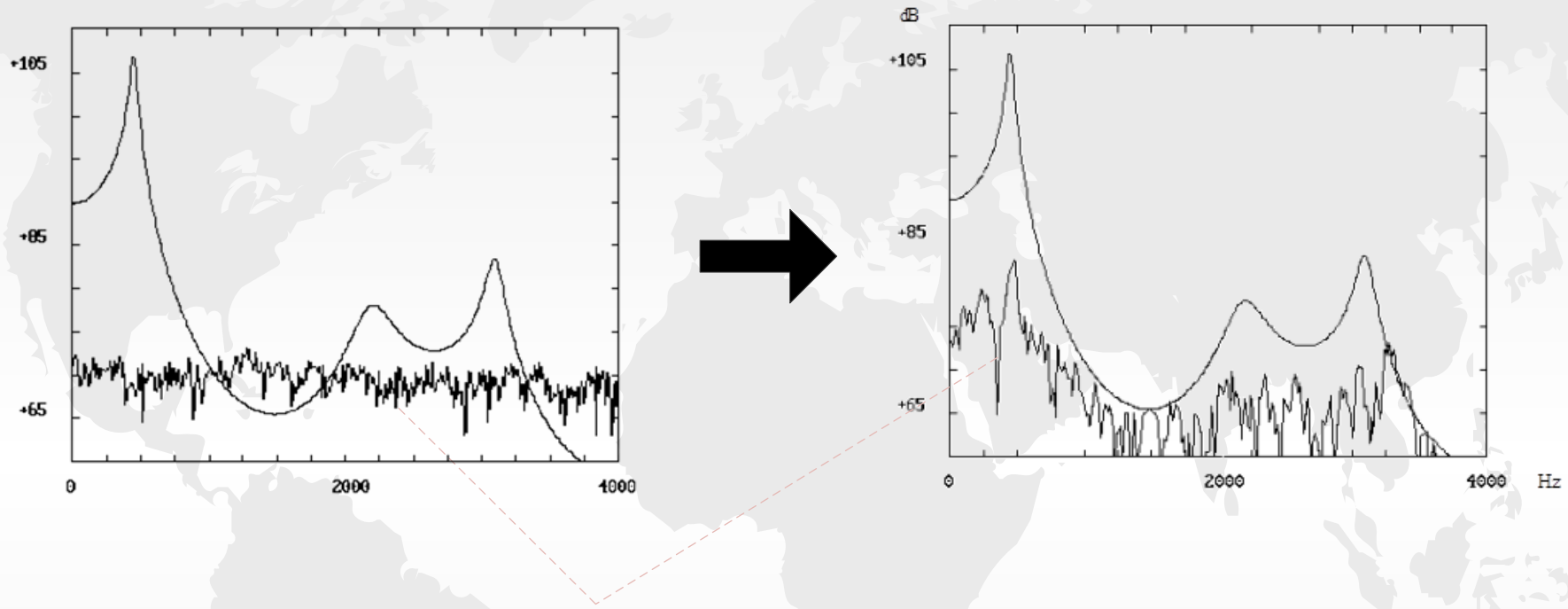
maskovací šum o centralní frekvenci 1kHz, kritické šírce pásma 400Hz a úrovni 60 dBSPL

práh slyšitelnosti



- tóny v blízkosti silného tónu jsou maskovány
- spektrální komponenty s úrovní pod prahem slyšitelnosti není třeba kódovat
- lze tolerovat vyšší úroveň kvantizačního šumu v blízkosti silných tónů, např. formantů
- demo na https://www.youtube.com/watch?v=k6DVywW5NR4

Maskování kvantizačního šumu

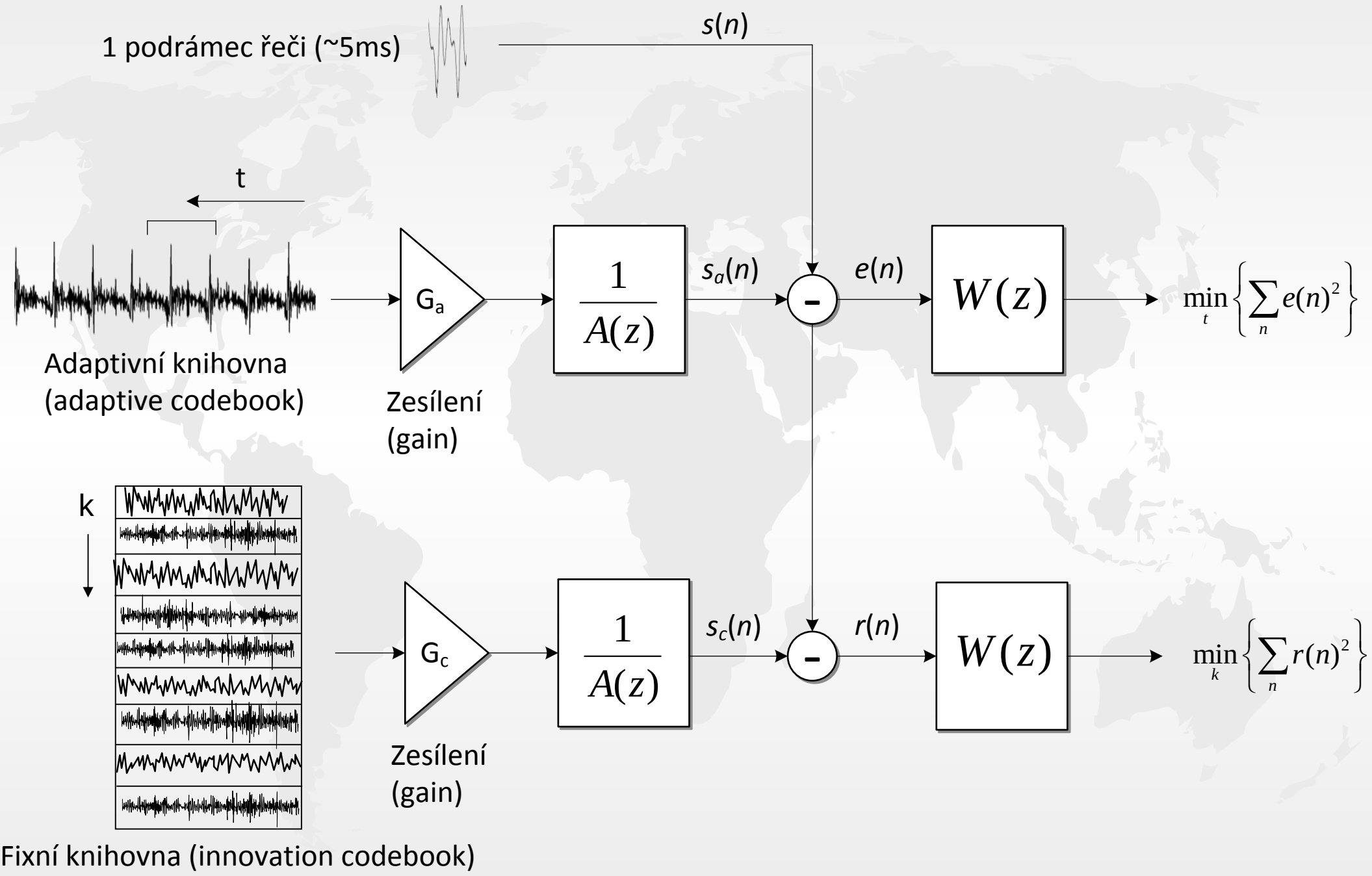


kvantizační šum

$$W(z) = \frac{A(z/\gamma_1)}{1-\beta z^{-1}} \qquad 0 < \gamma_1 \leq 1 \qquad 0 < \beta \leq 1$$

perceptuální filtr

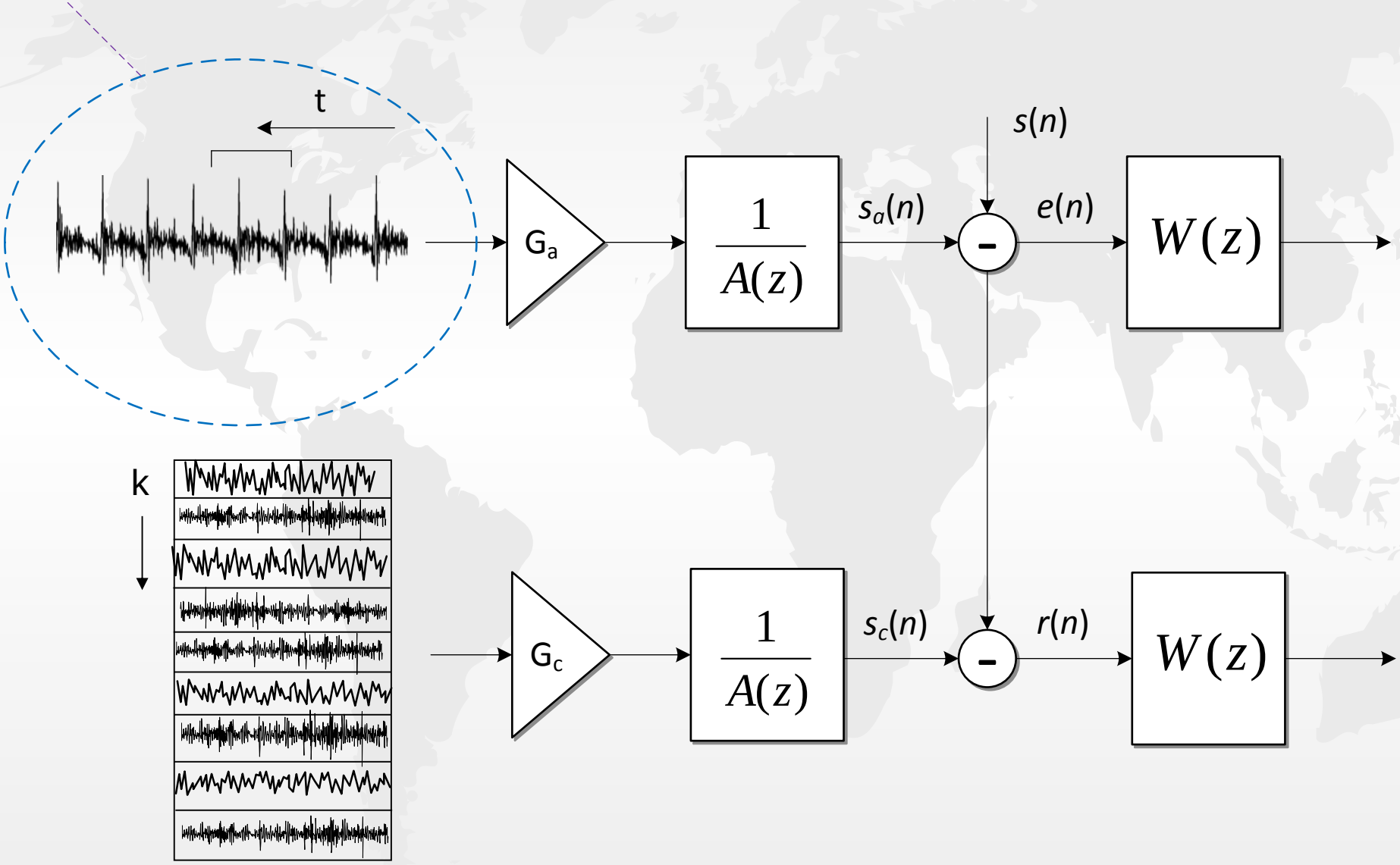
Zavedení perceptuálního filtru



Výpočetní náročnost

Adaptivní knihovna: (7-9 bitů, t.j. 128 – 512 vektorů)

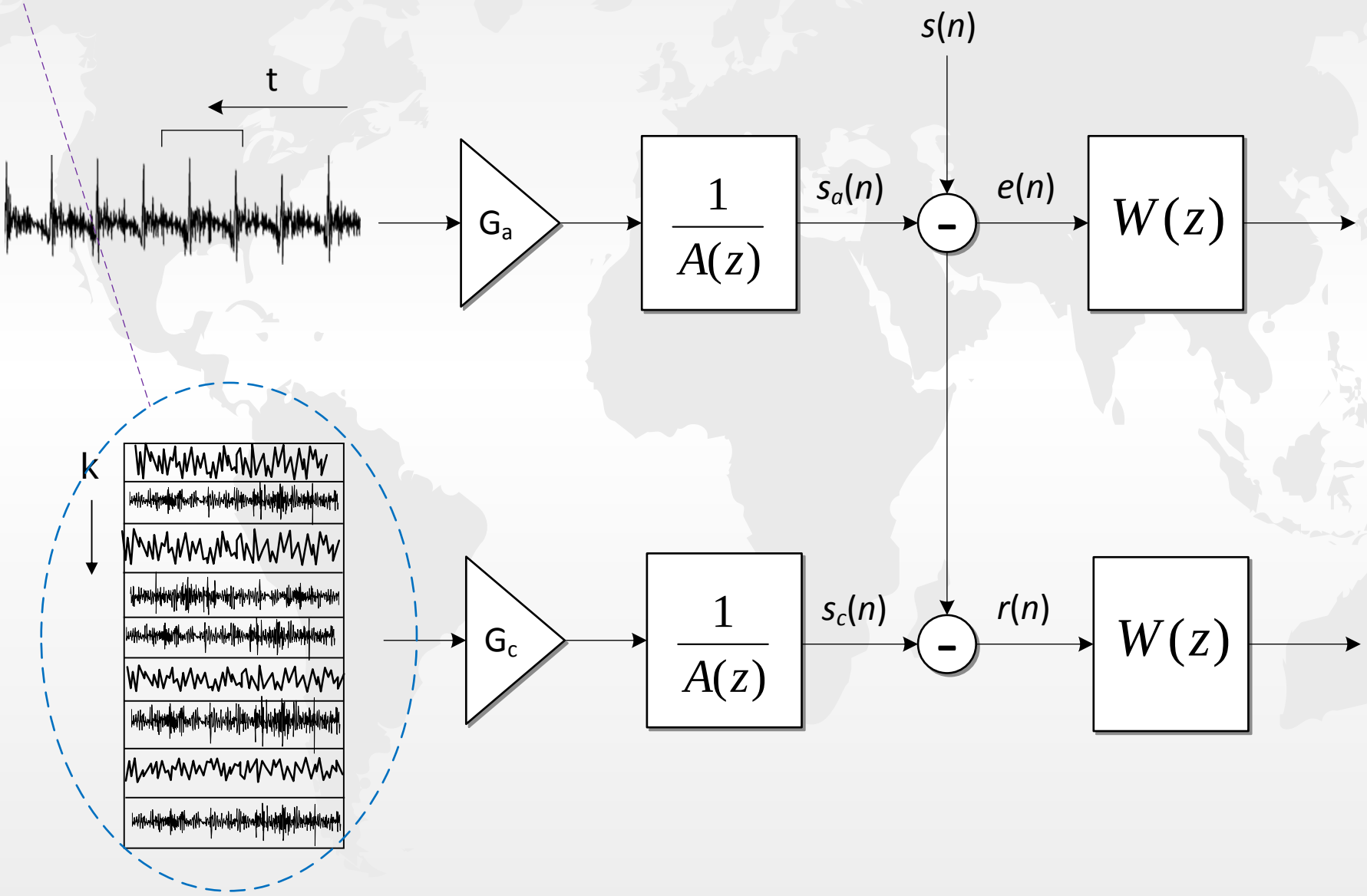
(estimace periodicity předem (OL pitch analysis), prohledávání knihovny kolem této hodnoty)



Výpočetní náročnost

Fixní knihovna: (10-88 bitů, t.j. 1024 – 3.10^{26} vektorů)

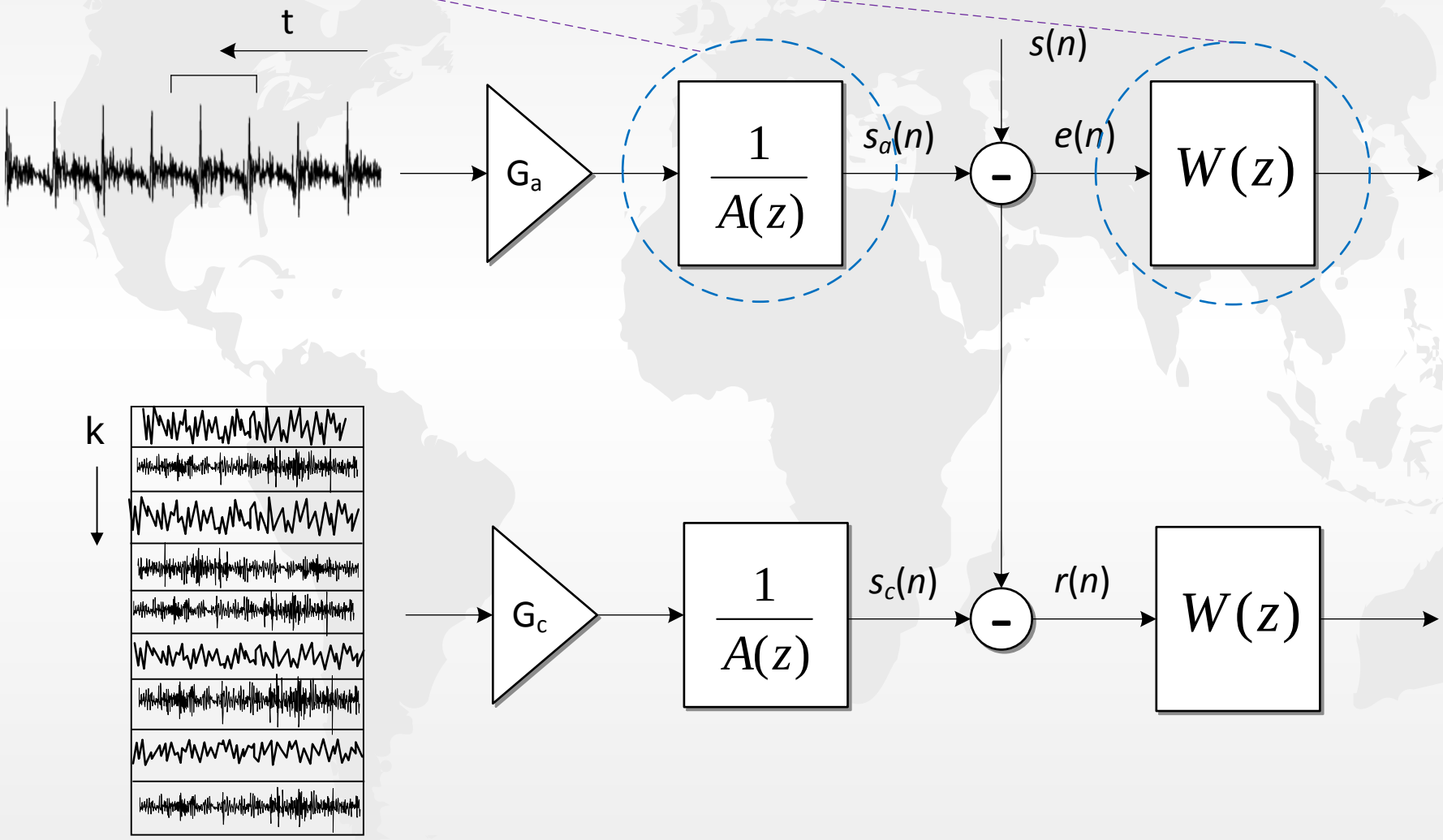
(vnucení jednoduchých struktur – pouze několik pulzů na stopu, omezený počet pozic pulzů, znaménka)



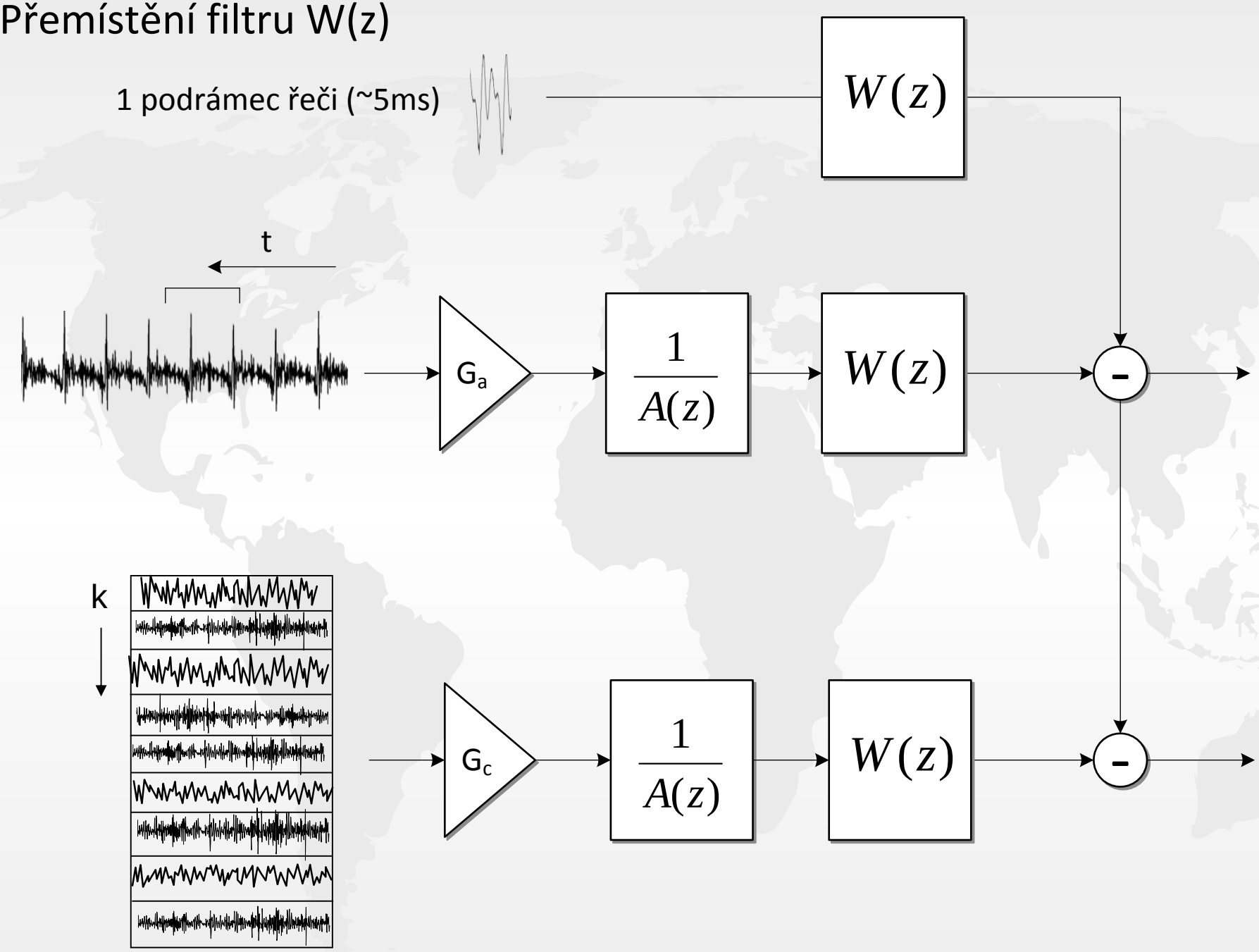
Výpočetní náročnost

Filtrace: (pro každý codevector nutno provést konvoluci s filtry 16.řádu)

(nahrazení filtrů 1/A(z) a W(z) jejich impulzní odezvou) – viz dále

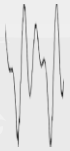


Přemístění filtru W(z)

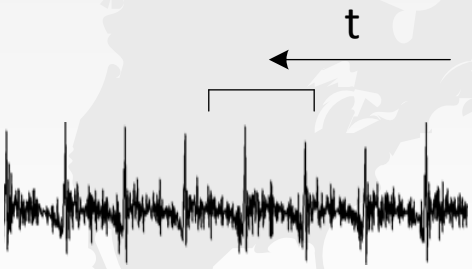


Odečtení ZIR

1 podrámec řeči (~5ms)



$W(z)$

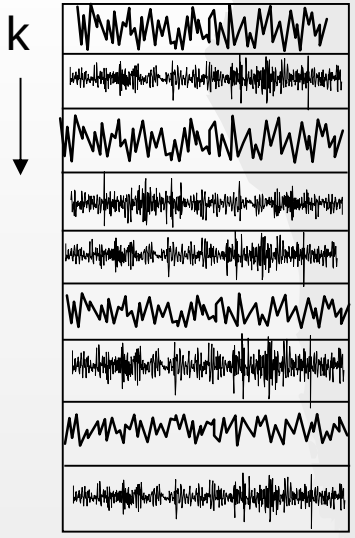


G_a

$\frac{1}{A(z)}$

$W(z)$

koeficienty filtrů se mění,
ale co s pamětmi?

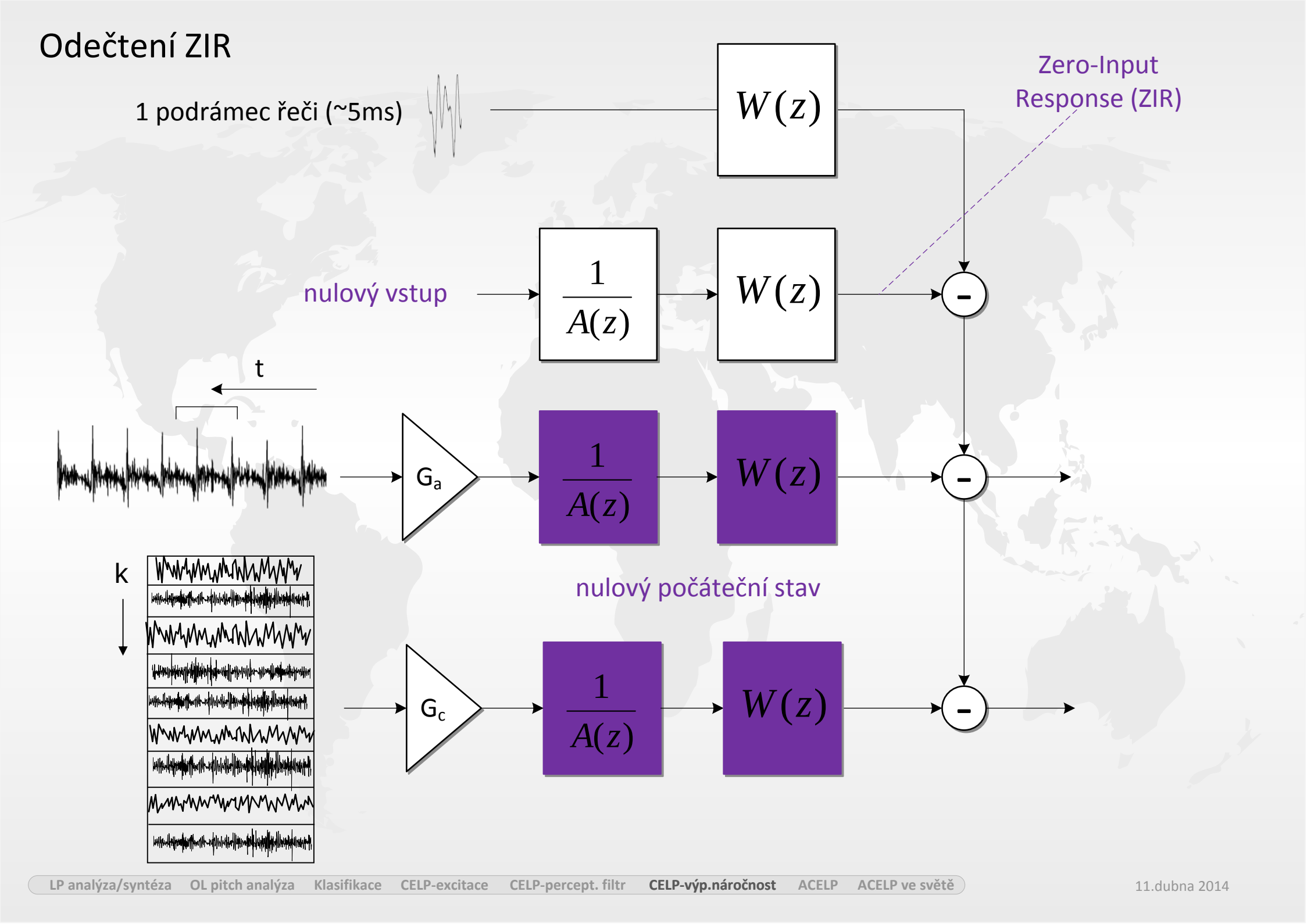


G_c

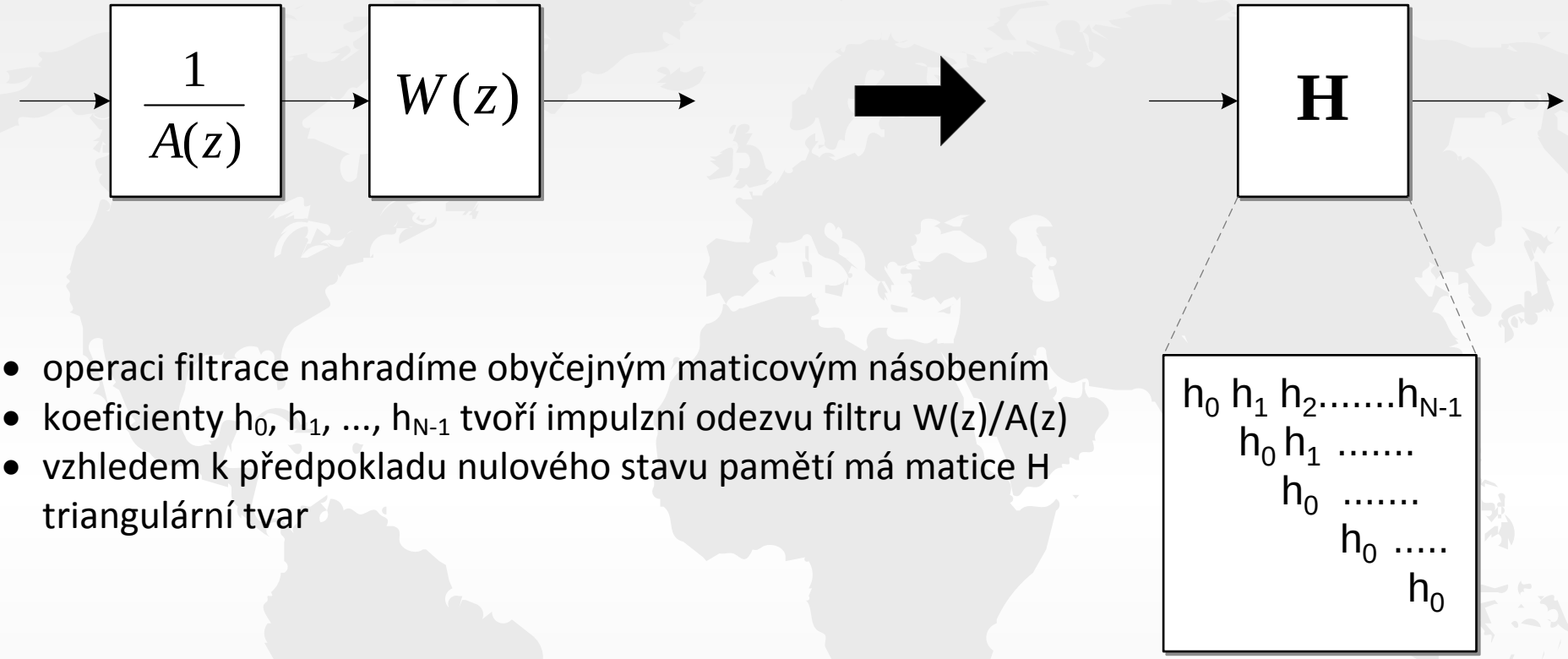
$\frac{1}{A(z)}$

$W(z)$

Odečtení ZIR

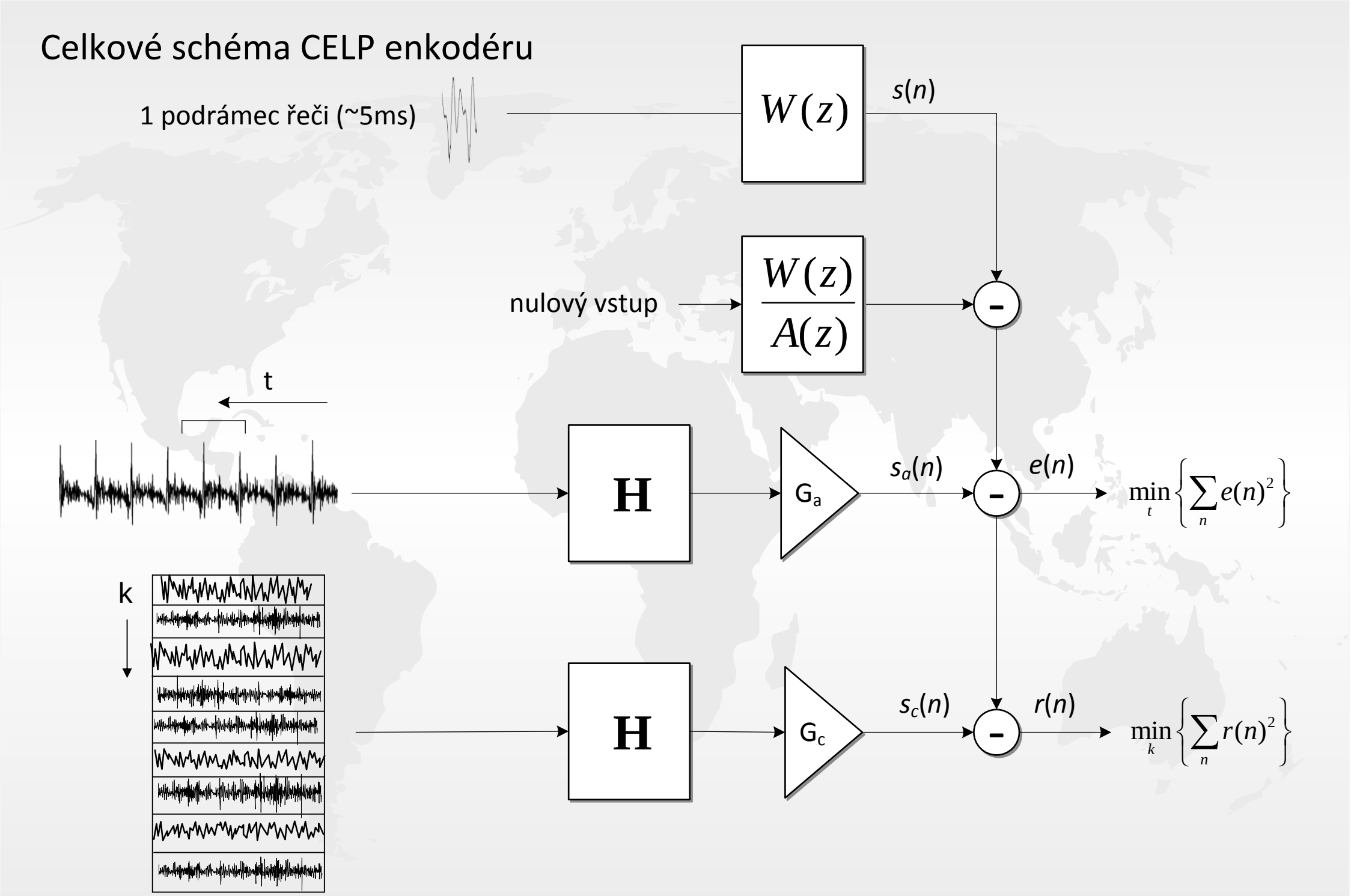


Nahrazení filtrů impulzní odezvou

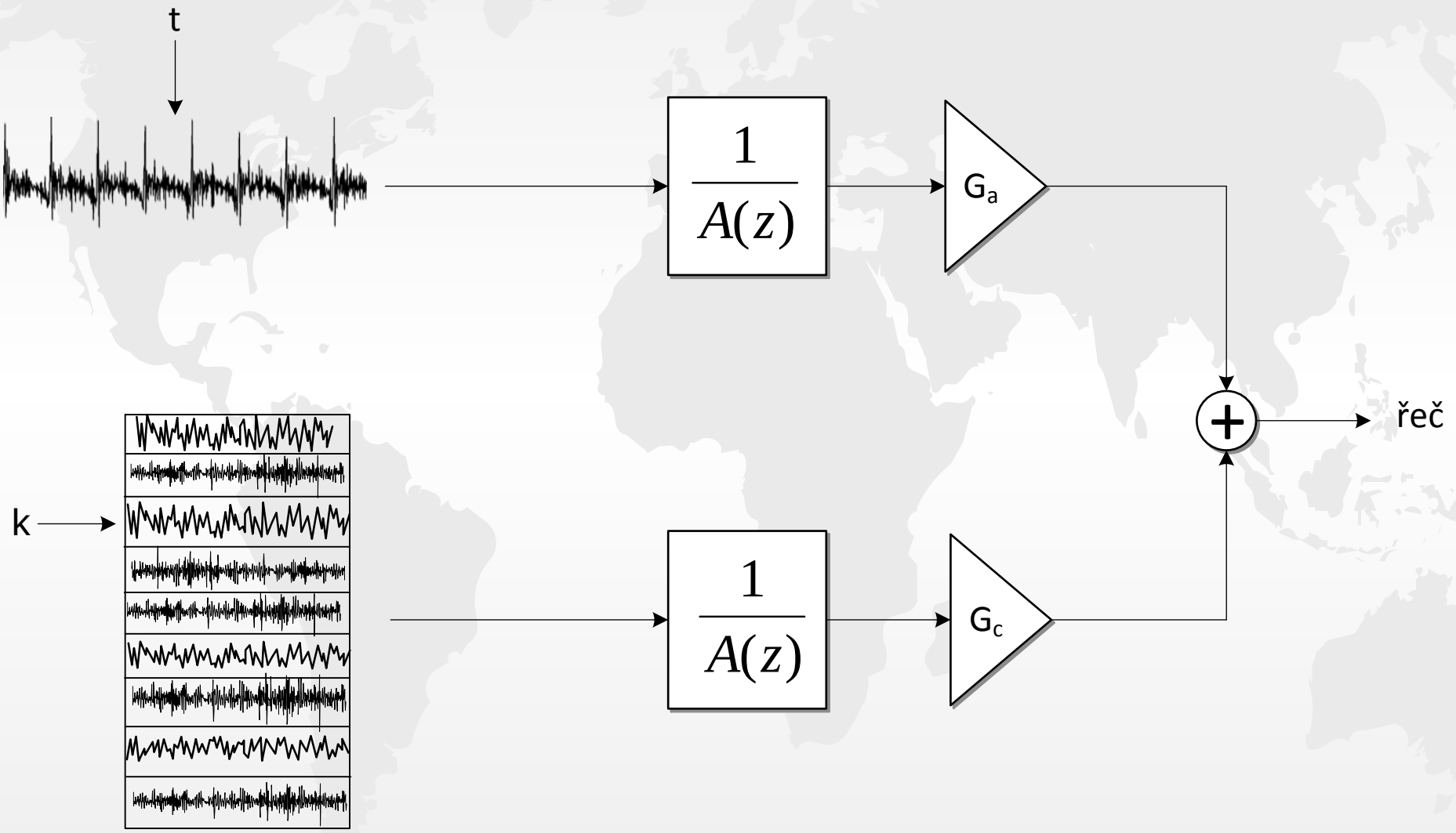


- operaci filtrace nahradíme obyčejným maticovým násobením
- koeficienty h_0, h_1, ..., h_{N-1} tvoří impulzní odezvu filtru W(z)/A(z)
- vzhledem k předpokladu nulového stavu pamětí má matice H triangulární tvar

Celkové schéma CELP enkodéru



Celkové schéma CELP dekodéru



ACELP

ACELP

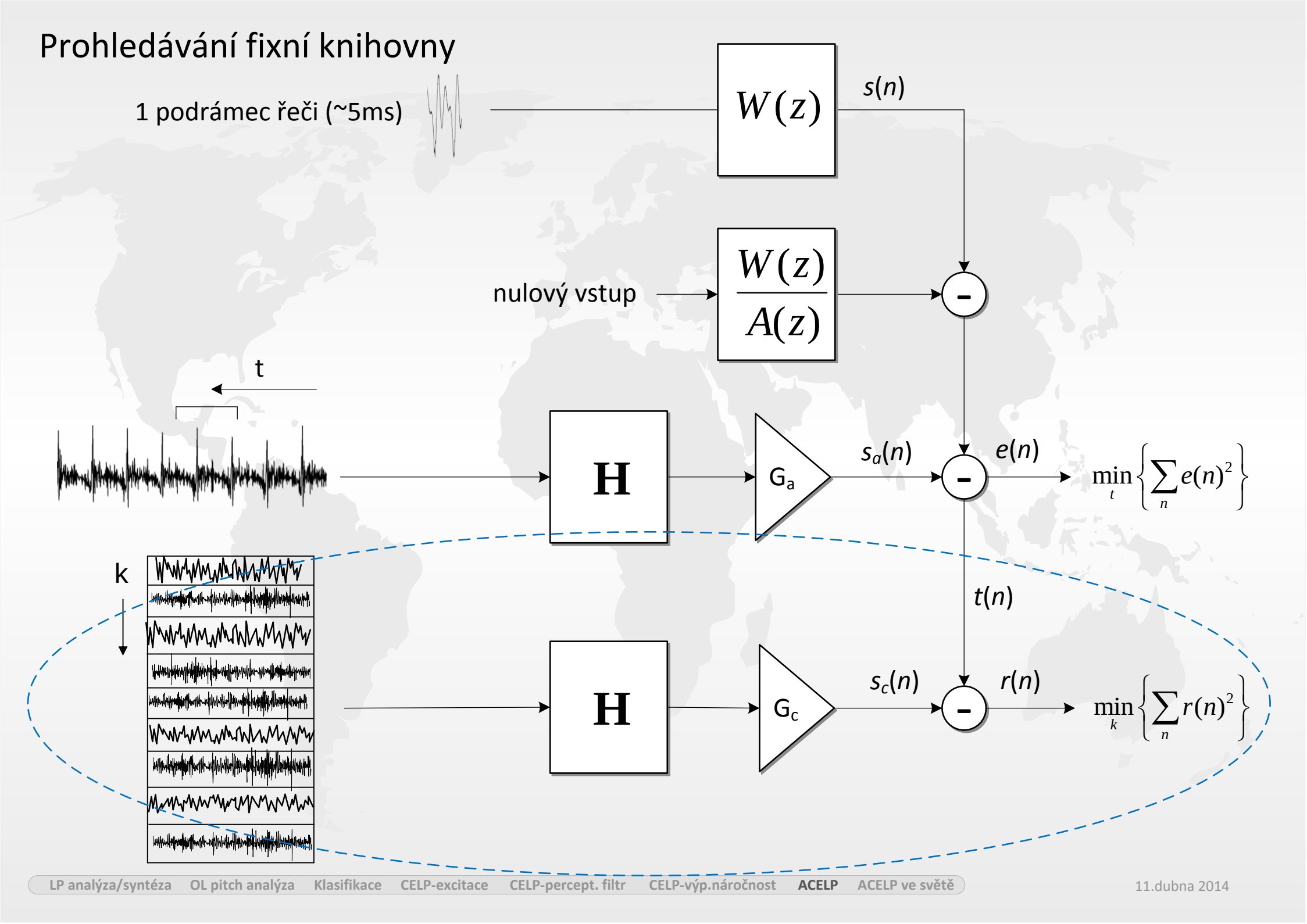
- ACELP® je patentovanou technologií VoiceAge Corp. a Université de Sherbrooke, CANADA
- vyvinuto v roce 1989 (Jean-Pierre Adoul, Claude Laflamme, Redwan Salami, Bruno Bessette)

UNIVERSITÉ DE SHERBROOKE

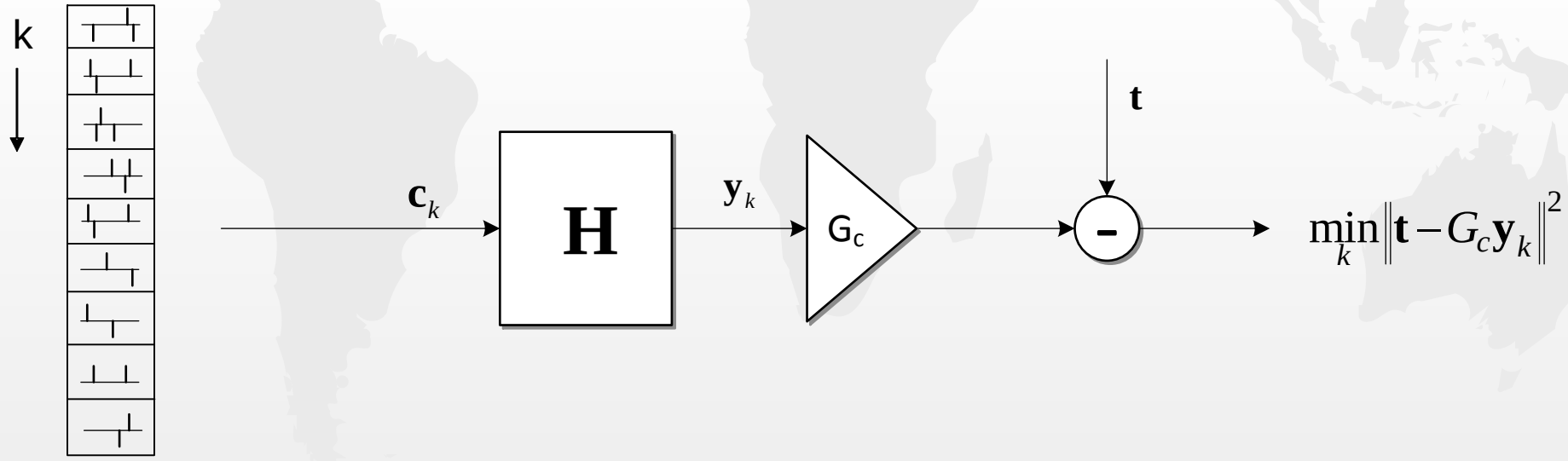
VoiceAge®
www.voiceage.com
The World's Premier Supplier of Speech and Audio Codecs

- kouzlo ACELPu spočívá v tom, že dokáže nahradit „obří“ fixní knihovnu signálů jednoduchou knihovnou s algebraickou strukturou, kde je jen několik málo pulzů v přesně definovaných pozicích a tím zredukovat paměťovou a výpočetní náročnost
- technologii ACELP využívá cca
  - 2,4 miliard uživatelů mobilních telefonů na celém světě
  - 35 milionů uživatelů přehrávačů MP3
  - 500 milionů uživatelů internetových přehrávačů RealPlayer nebo MediaPlayer

Prohledávání fixní knihovny



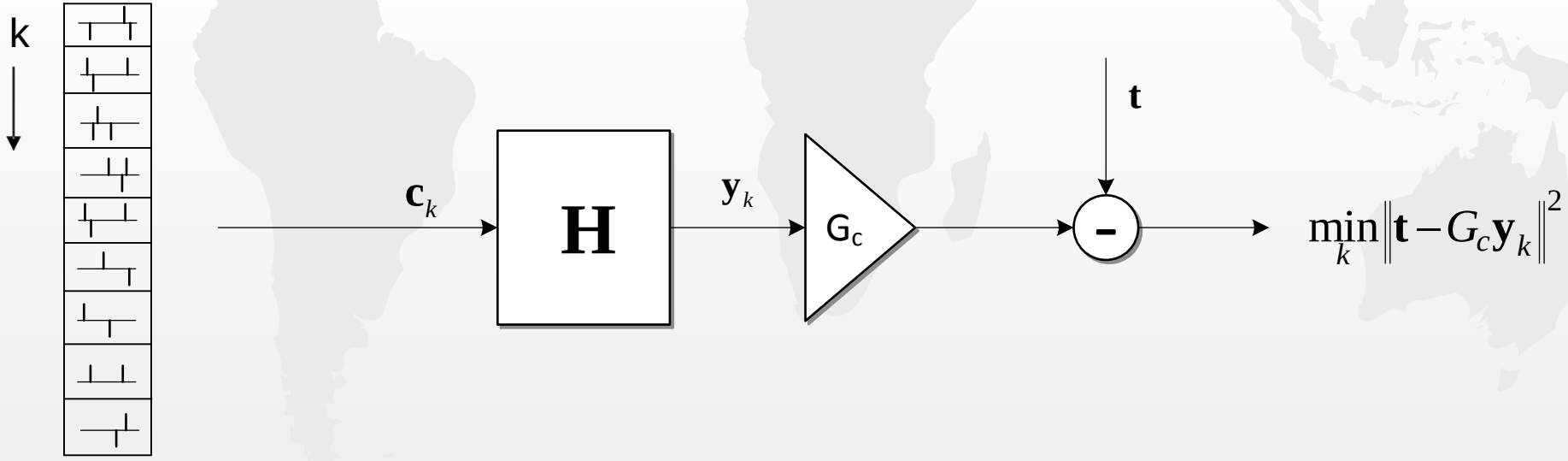
Zavedení algebraické knihovny



algebraická knihovna (až 80 bitů)

Prohledávání fixní knihovny

$$\min_k \left\| \mathbf{t} - G_c \mathbf{y}_k \right\|^2 \xrightarrow{\frac{\partial}{\partial G_c}}$$



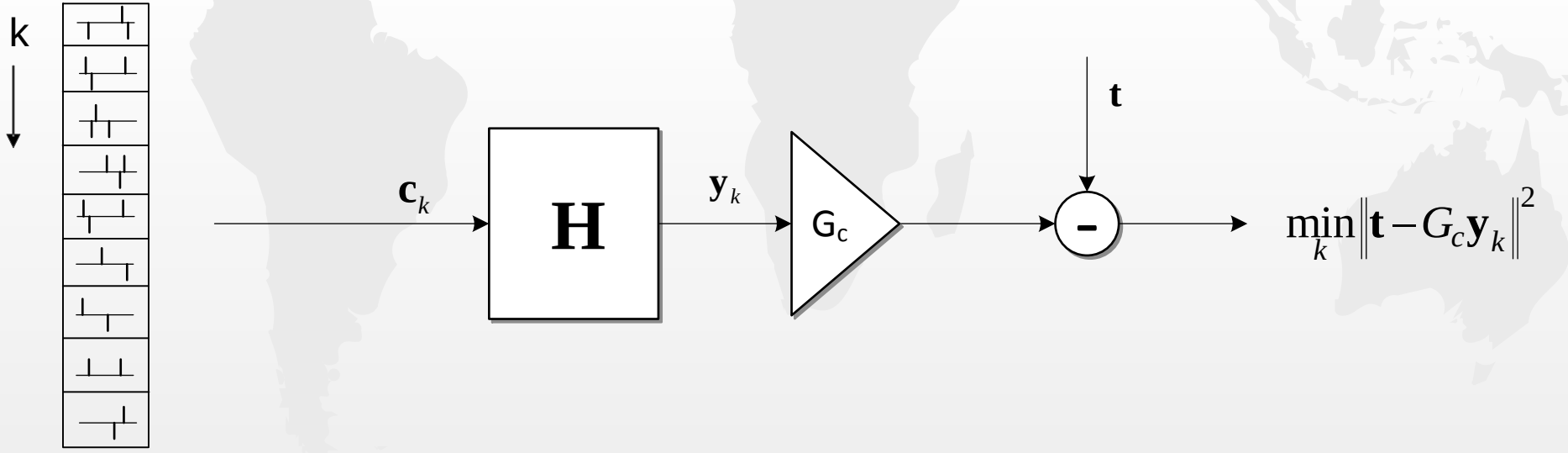
algebraická knihovna (až 80 bitů)

Prohledávání algebraické knihovny

$$\min_k \left\| \mathbf{t} - G_c \mathbf{y}_k \right\|^2 \xrightarrow{\frac{\partial}{\partial G_c}} \max_k \frac{\mathbf{t}^T . \mathbf{y}_k}{\mathbf{y}_k^T . \mathbf{y}_k}$$

korelace mezi cílovým (target) vektorem a testovaným vektorem

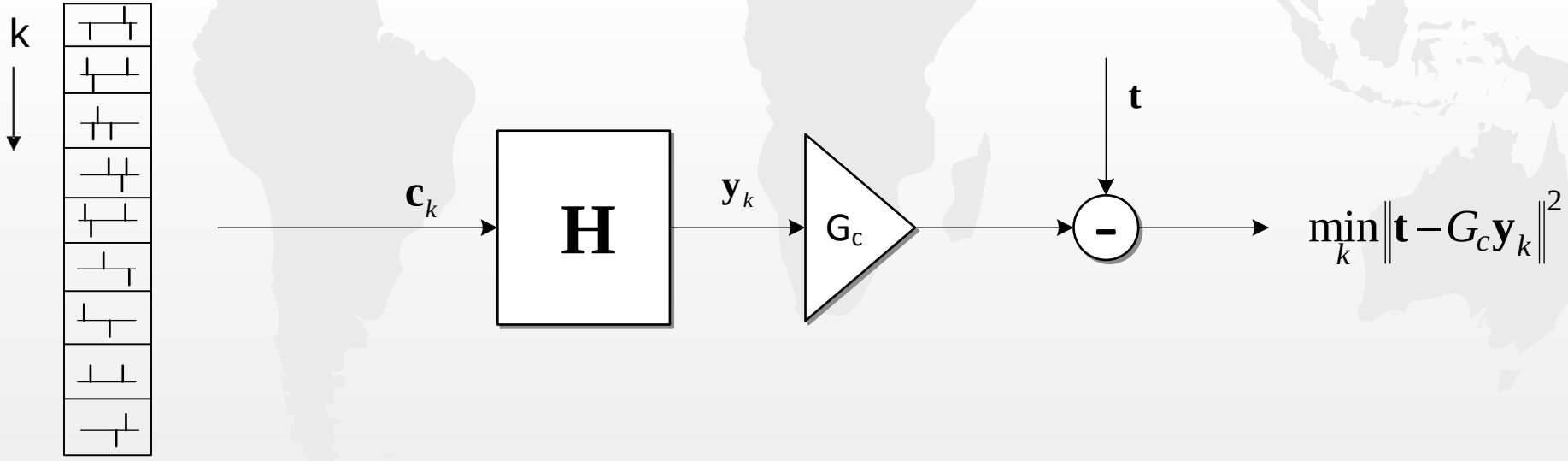
energie testovaného vektoru



algebraická knihovna (až 80 bitů)

Prohledávání algebraické knihovny

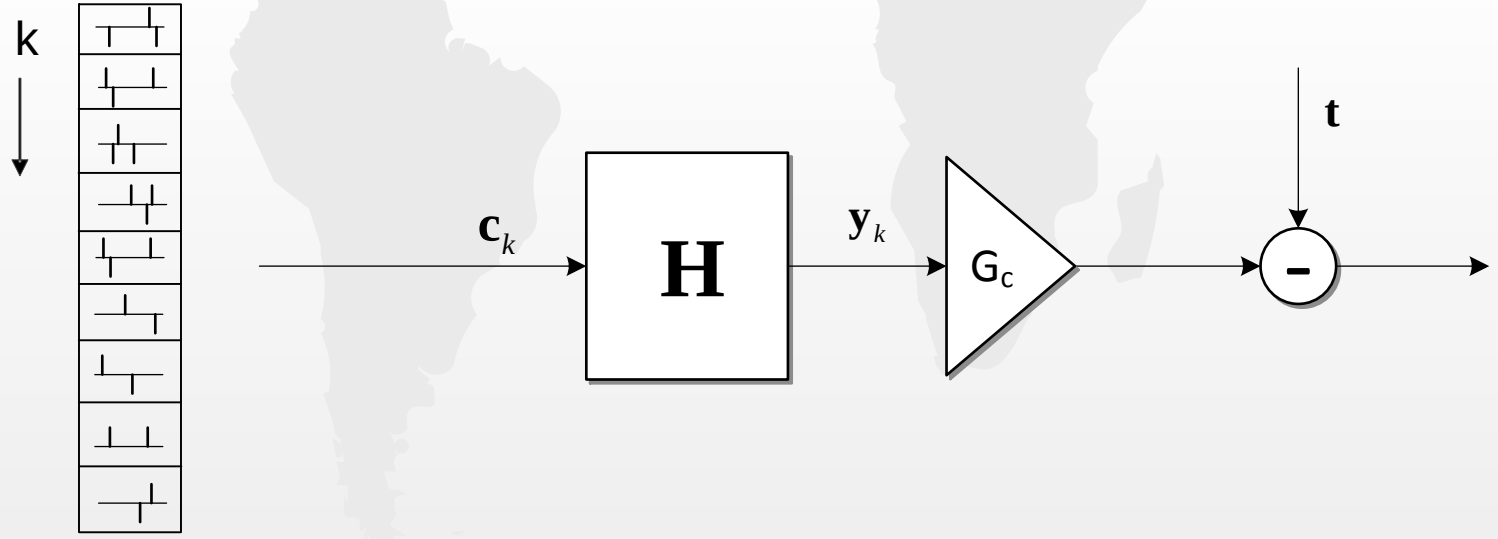
$$\min_k \left\| \mathbf{t} - G_c \mathbf{y}_k \right\|^2 \xrightarrow{\frac{\partial}{\partial G_c}} \max_k \frac{\left(\mathbf{t}^T . \mathbf{y}_k \right)^2}{\mathbf{y}_k^T . \mathbf{y}_k} \longrightarrow \max_k \frac{\left(\mathbf{t}^T . \mathbf{H} . \mathbf{c}_k \right)^2}{\mathbf{c}_k^T . \mathbf{H}^T . \mathbf{H} . \mathbf{c}_k}$$



algebraická knihovna (až 80 bitů)

Prohledávání algebraické knihovny

$$\min_k \left\| \mathbf{t} - G_c \mathbf{y}_k \right\|^2 \xrightarrow{\frac{\partial}{\partial G_c}} \max_k \frac{\left(\mathbf{t}^T . \mathbf{y}_k \right)^2}{\mathbf{y}_k^T . \mathbf{y}_k} \rightarrow \max_k \frac{\left(\mathbf{t}^T . \mathbf{H} . \mathbf{c}_k \right)^2}{\mathbf{c}_k^T . \mathbf{H}^T . \mathbf{H} . \mathbf{c}_k} \rightarrow \max_k \frac{\left(\mathbf{d}^T . \mathbf{c}_k \right)^2}{\mathbf{c}_k^T . \mathbf{\Phi} . \mathbf{c}_k}$$



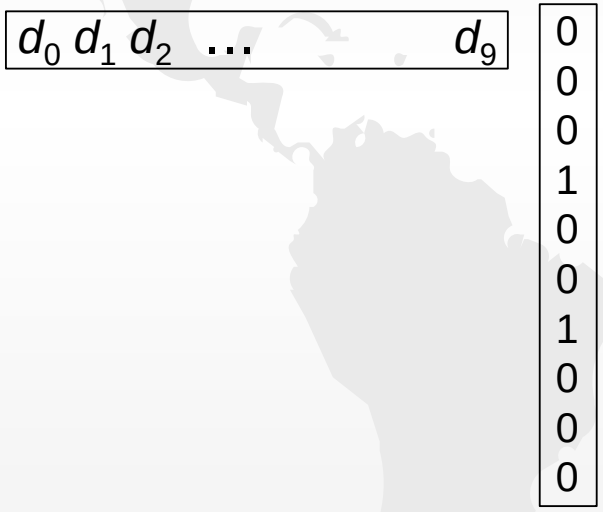
algebraická knihovna (až 80 bitů)

Prohledávání algebraické knihovny

$$\max_k \frac{\left(\mathbf{d}^T.\mathbf{c}_k\right)^2}{\mathbf{c}_k^T.\mathbf{\Phi}.\mathbf{c}_k}$$

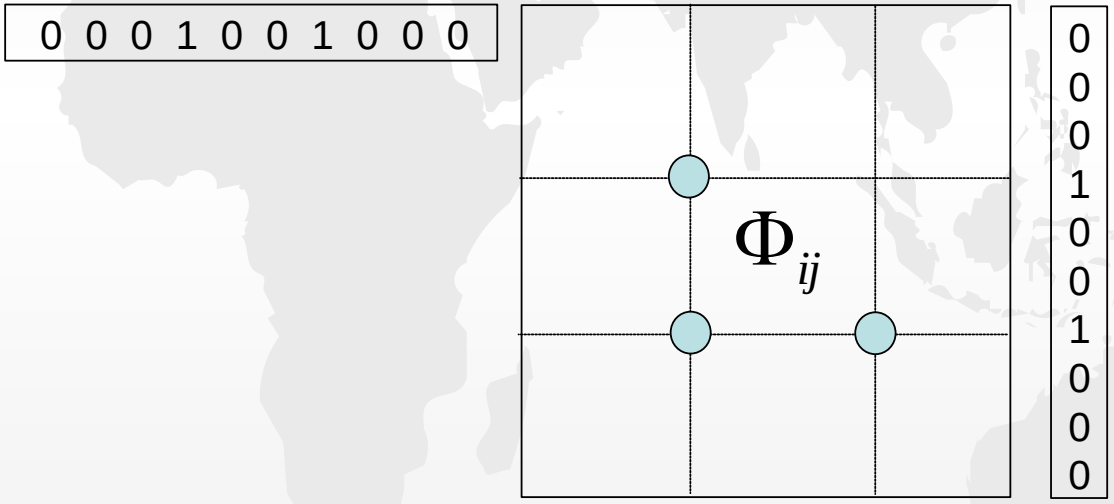
Lze prohledávat rychle, pokud $\mathbf{c}_k$ obsahuje jen velmi málo nenulových prvků s hodnotami +1 nebo -1

$\mathbf{d}^T.\mathbf{c}_k$



$= d_3 + d_6$

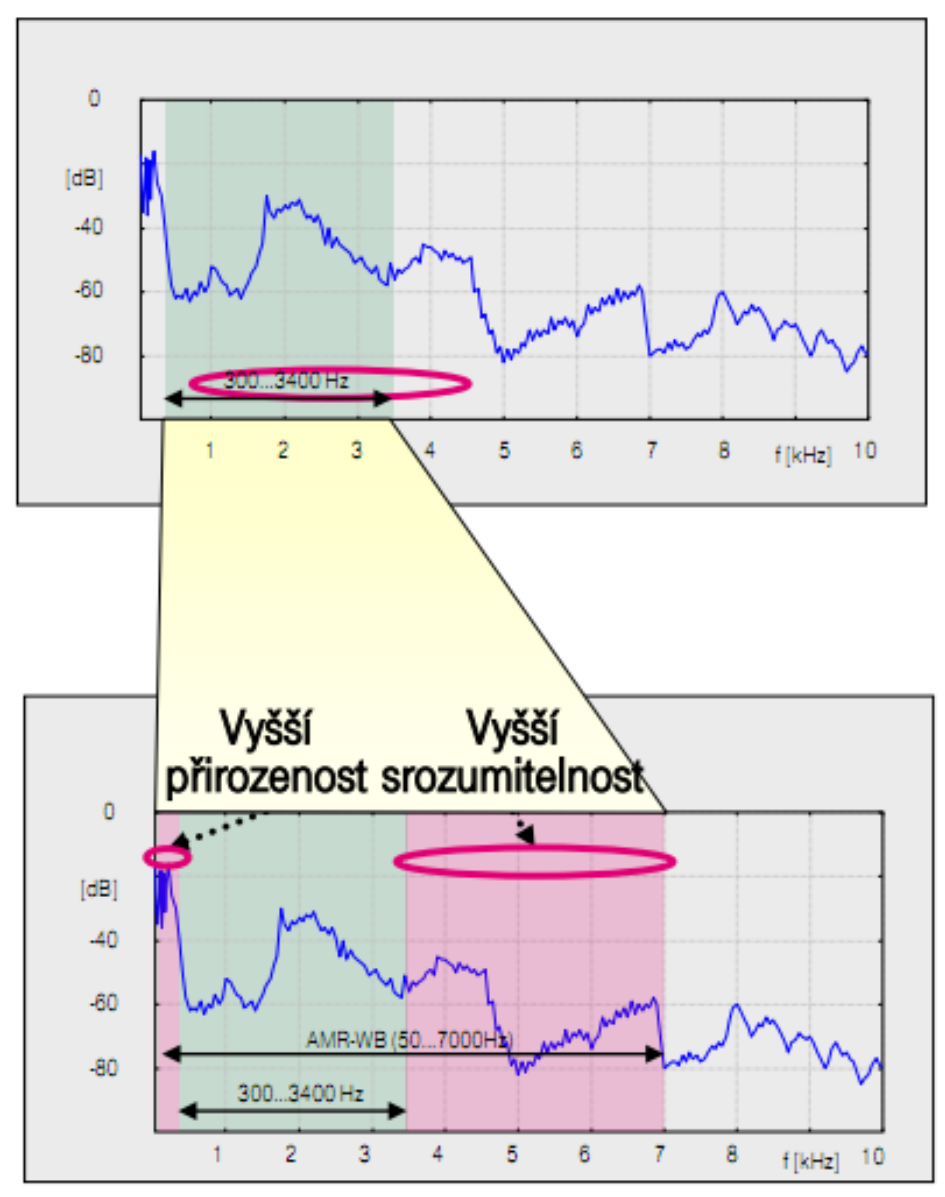
$\mathbf{c}_k^T.\mathbf{\Phi}.\mathbf{c}_k$



$= \Phi_{3,3} + \Phi_{6,6} + 2\Phi_{3,6}$

ACELP ve světě

Od AMR-NB k AMR-WB (HD VOICE)



- HD voice demo na https://www.youtube.com/watch?v=Y4bb3b9PiRg

Technologie ACELP v mezinárodních standardech