B2M31ZRE - Zpracování řeči

Spektrální charakteristiky řečového signálu

Doc. Ing. Petr Pollák, CSc.

26. února 2023 - 21:23

Obsah přednášky

- **Spektrální reprezentace na bázi DFT**
  - Výpočet DFT řečového signálu (FFT, váhování, frekvenční rozlišení)
  - Banky filtrů
  - Preemfáze
- **Lineární prediktivní analýza - AR modelování**
  - Princip LPC, LPC spektrum, AR model
  - Algoritmy výpočtu (Levinson-Durbin, Burg)
- **Formantová analýza**
  - Formanty - definice a význam
  - Metody odhadu formantů

I. část

Spektrální charakteristiky na bázi DFT

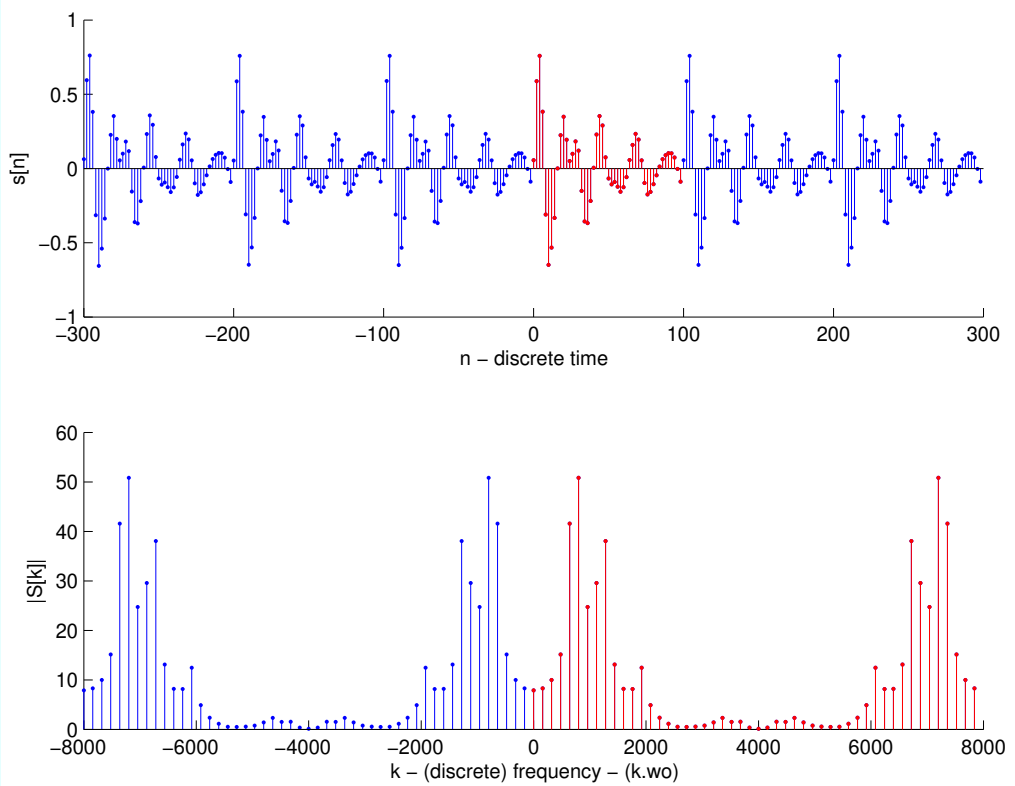
Diskrétní Fourierova Transformace (DFT) - základní vlastnosti

Přímá transformace - DFT

$$S[k] = \sum_{n=0}^{N-1} s[n] e^{-j\frac{2\pi}{N}kn}$$

Inverzní transf. - IDFT

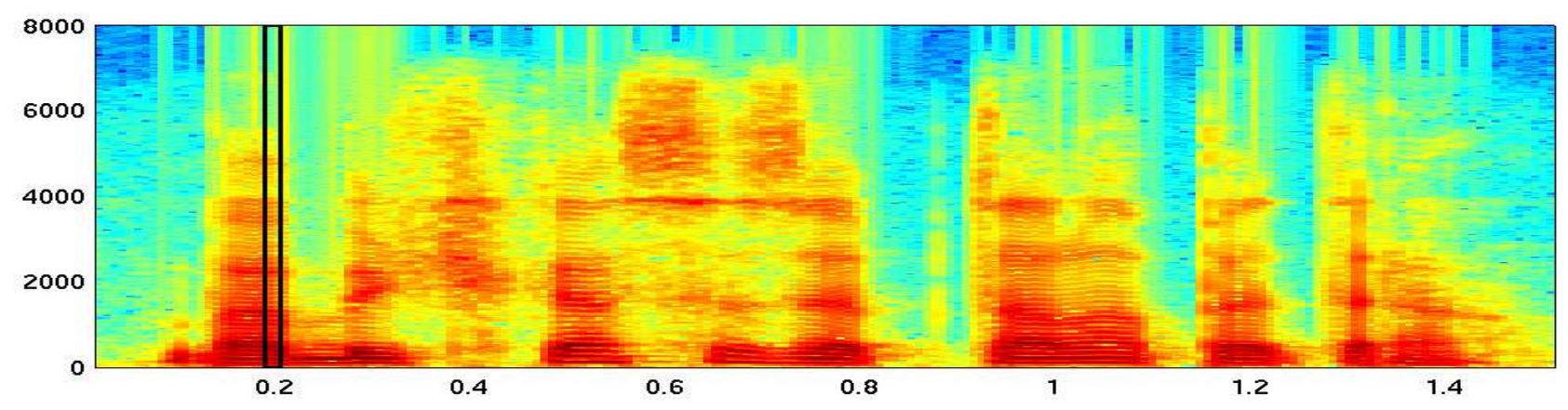
$$s[n] = \frac{1}{N} \sum_{k=0}^{N-1} S[k] e^{j\frac{2\pi}{N}kn}$$



- **transformace** : diskrétní signál konečné délky - diskrétní spektrum
- **spektrální rozlišení** : vzorky spektra jsou v rozsahu $0 \div f_s$ resp. $-\frac{f_s}{2} \div \frac{f_s}{2}$, tj. $\Delta f = \frac{f_s}{N}$
- **FFT**: rychlý algoritmus výpočtu ($N = 2^n$)

Přehled možností spektrální reprezentace promluvy

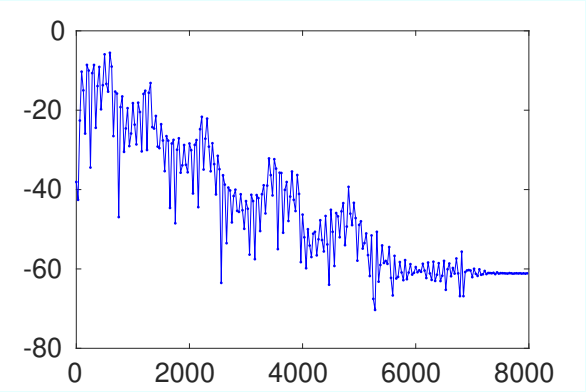
Spektrogram promluvy



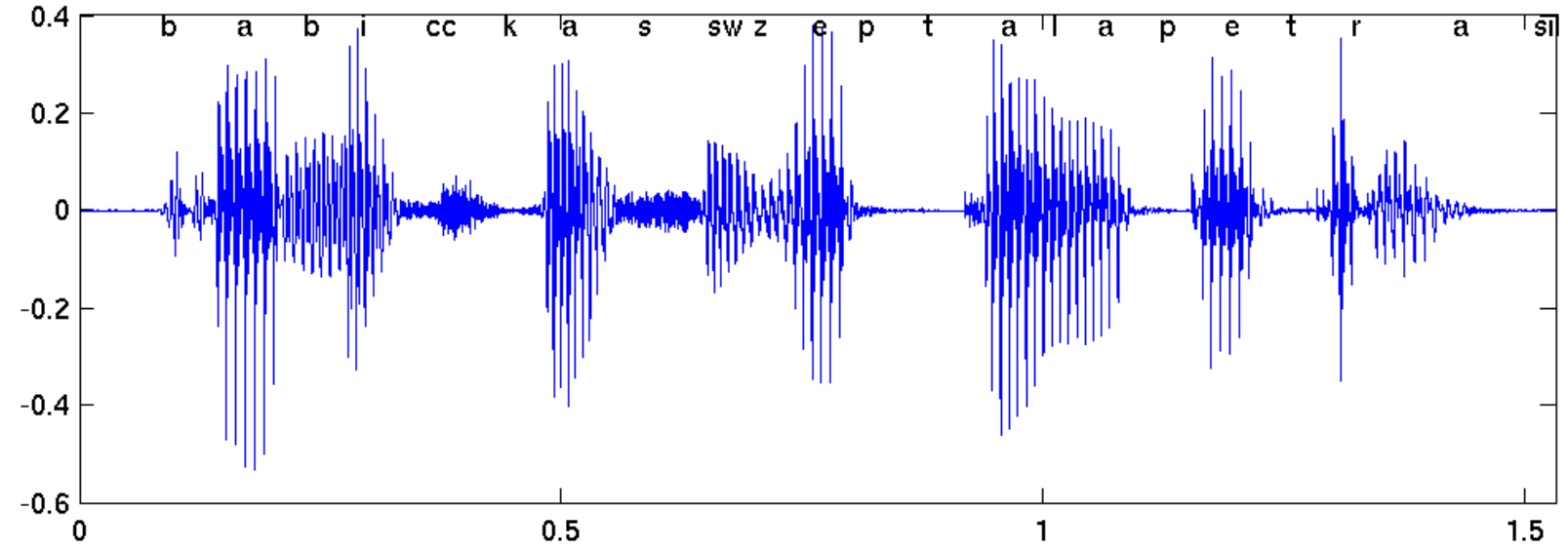
Dlouhodobé (průměrné) spektrum - **špatně, ztráta informace**

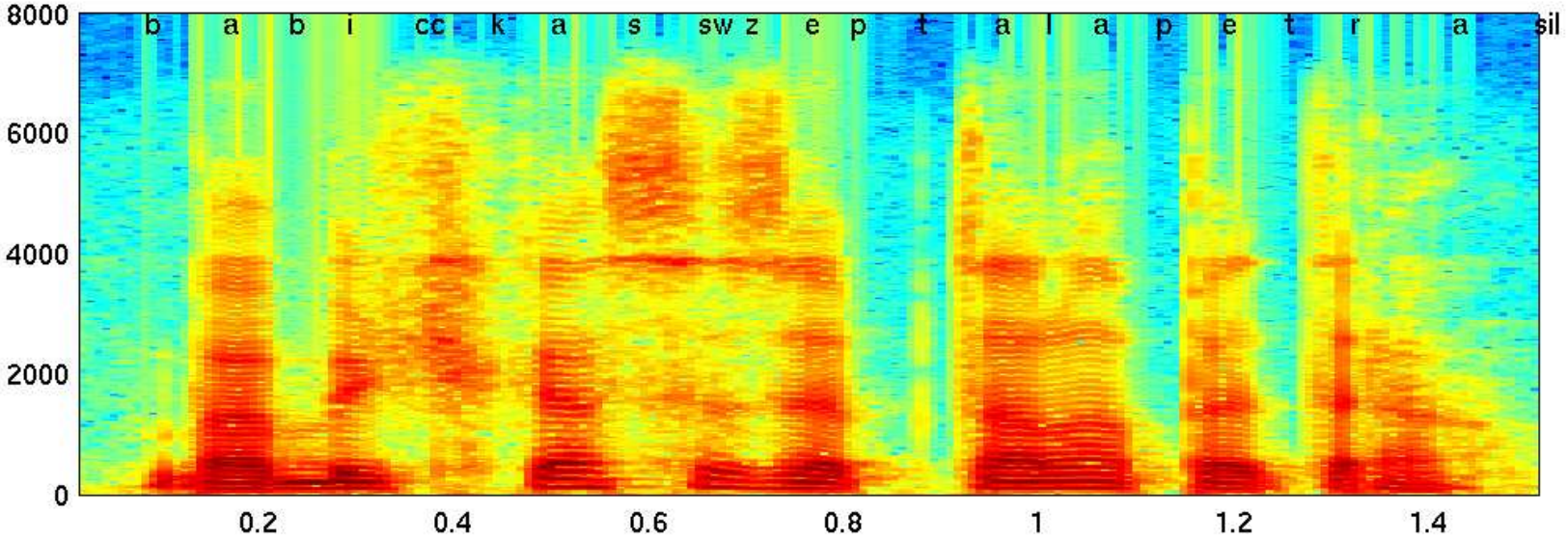
Spektrální reprezentace vybraného krátkodobého segmentu

DFT spektrum:



Časová a spektrální reprezentace promluvy





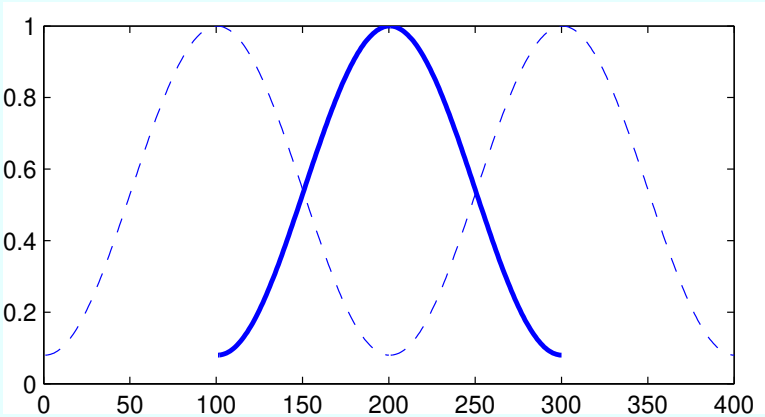
Spektrální analýza řeči na bázi DFT

Specifické vlastnosti:

- **řeč je obecně nestacionární signál** $\Rightarrow$ nutná segmentace a sledování vývoje krátkodobého spektra (spektrogram)
- **řeč je kvazistacionární**
  (tj. stacionární v krátkém časovém intervalu - cca 10-100 ms)
  $\Rightarrow$ 20-30 ms - typická délka krátkodobého segmentu
- **DFT spektrum je ovlivněno prosakováním**
  $\Rightarrow$ nutné váhování vhodným oknem (Hammingovo)
  $\Rightarrow$ nutná segmentace s překryvem (obvykle 50%)

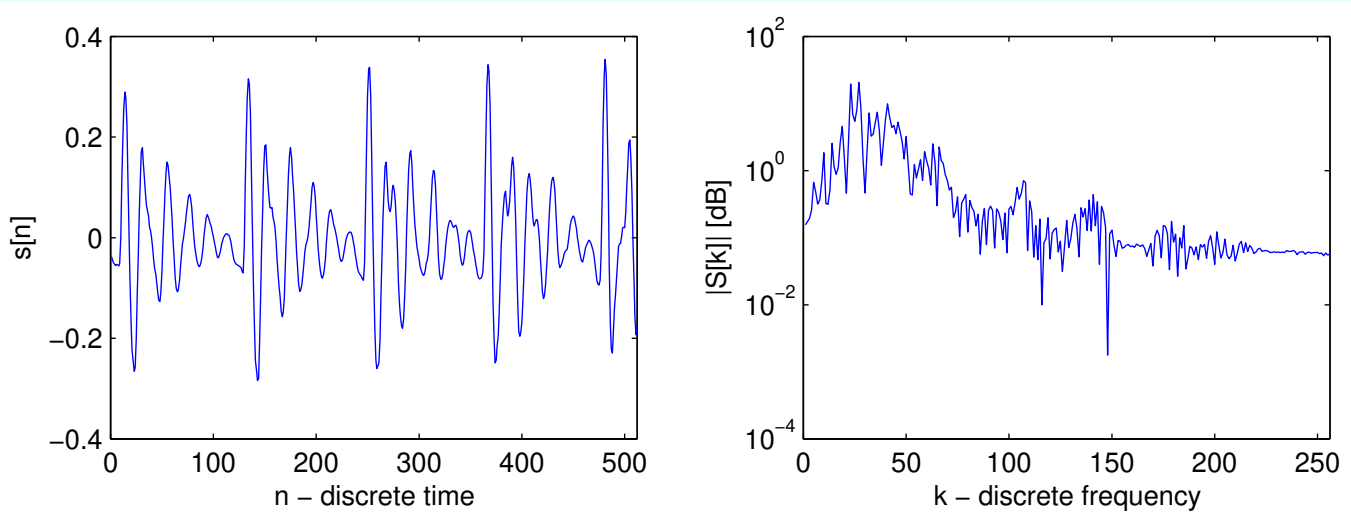
$$w[n] = 0,54 - 0,46 \cos \frac{2\pi n}{N}$$

$$\text{pro } 0 \leq n \leq N-1 \,.$$

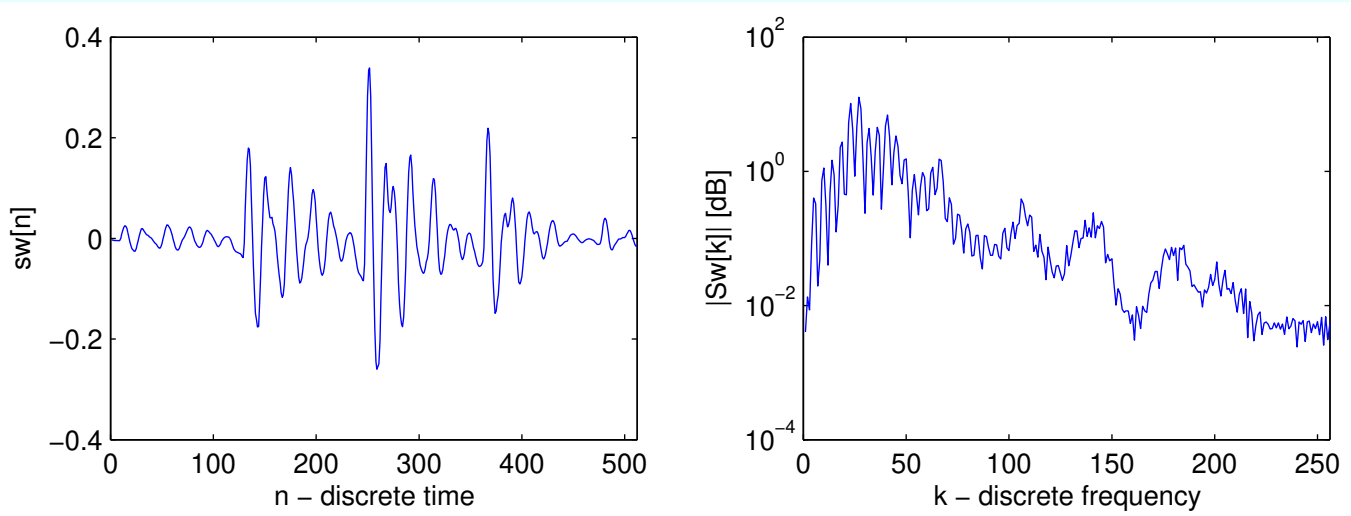


Vliv váhování na krátkodobé spektrum řečového signálu

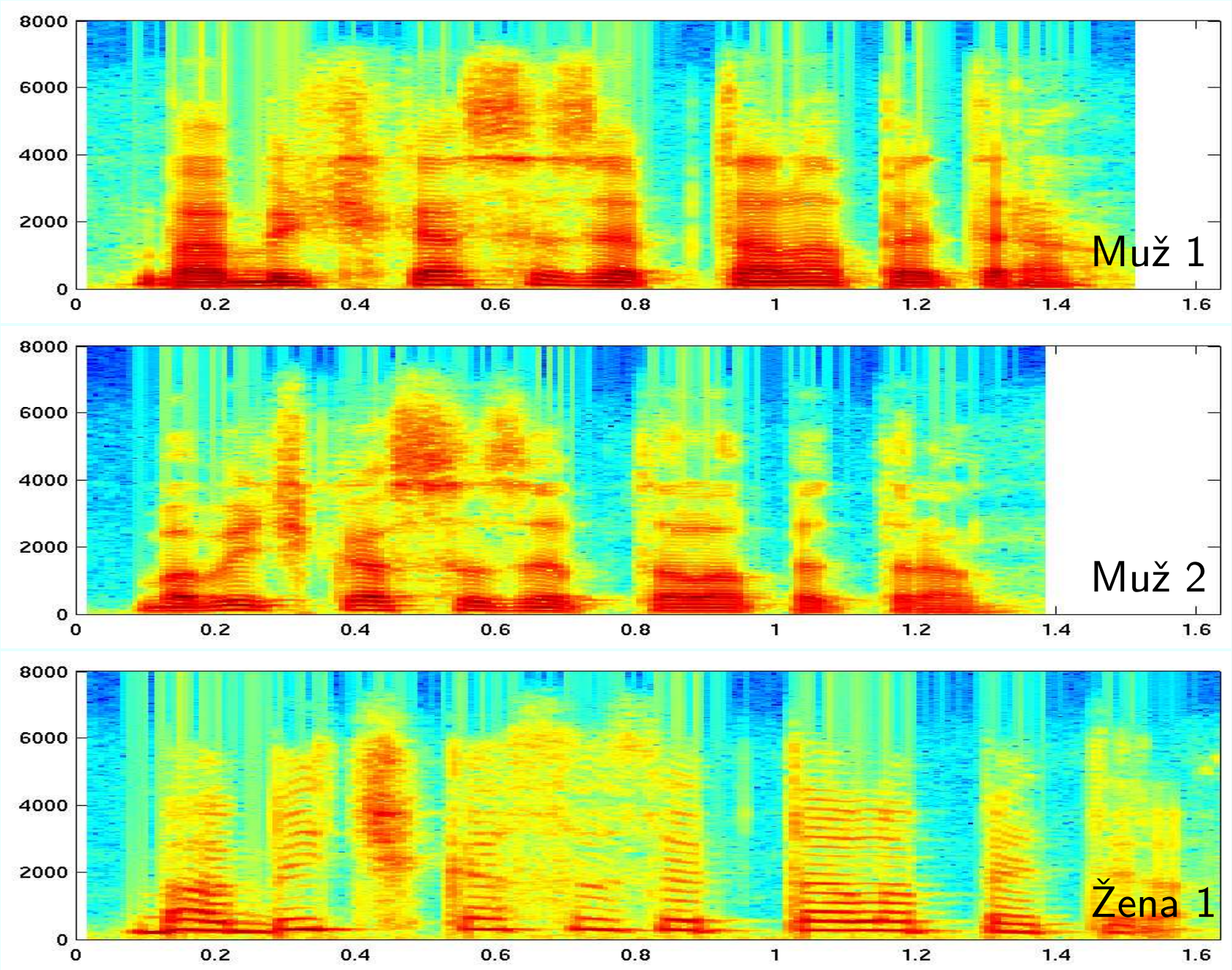
Spektrum neváhovaného segmentu - prosakování
(nízkoenergetické detaily ve spektru jsou maskovány)



Spektrum váhovaného segmentu - **minimalizace prosakování**
(viditelné nízkoenergetické detaily na vyšších kmitočtech)



Variabilita stejné promluvy - vliv f_o



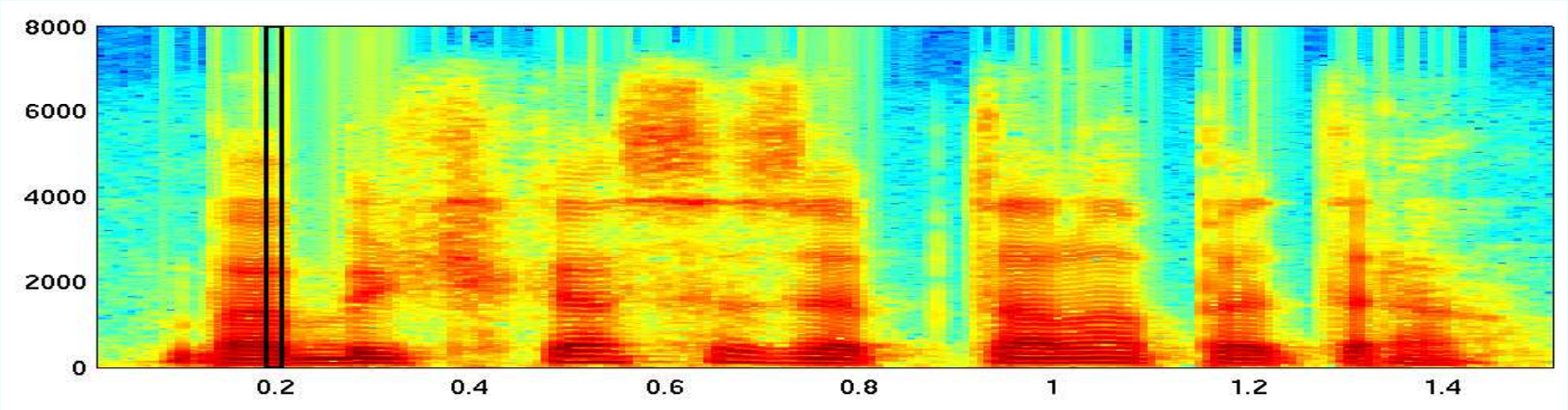
Vlastnosti krátkodobého spektra řeči na bázi DFT

- je možné rozlišit jednotlivé hlásky
- obsahuje informaci o periodicitě (dostatečně dlouhý segment)
- obsahuje náhodnou složku
- pro segmenty délky cca 30 ms - větší frekvenční rozlišení
  - + vhodné pro zvýrazňování, kódování, apod.
  - − pro rozpoznávání - nadbytečné množství informace

- vyhlazené spektrální charakteristiky
  - banky filtrů (nelineární frekvenční osa)
  - LPC
  - kepstrální analýza

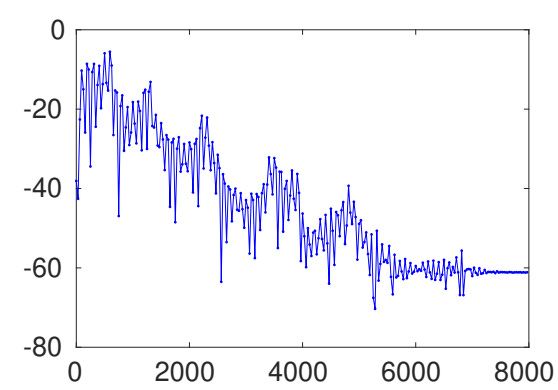
Přehled možností spektrální reprezentace promluvy

Spektrogram celé promluvy



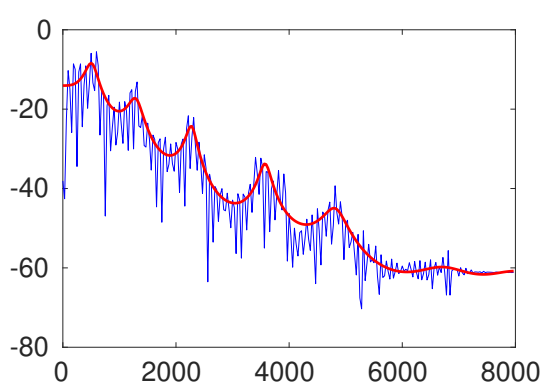
Spektrální reprezentace vybraného segmentu

DFT spektrum:



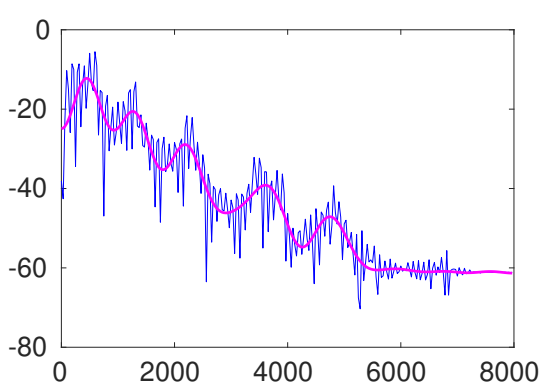
256 vzorků spektra
(amplitudové sp.)

LPC reprezentace:



$\approx$ 16 koeficientů a_k
(autoregresní koef.)

Kepstrální koeficienty:



$\approx$ 20 koeficientů c_n
(reálné kepstrum)

Banky filtrů ve spektrální analýze

Hlavní cíl $\rightarrow$ počítá se výkon (energie) ve zvolených pásmech

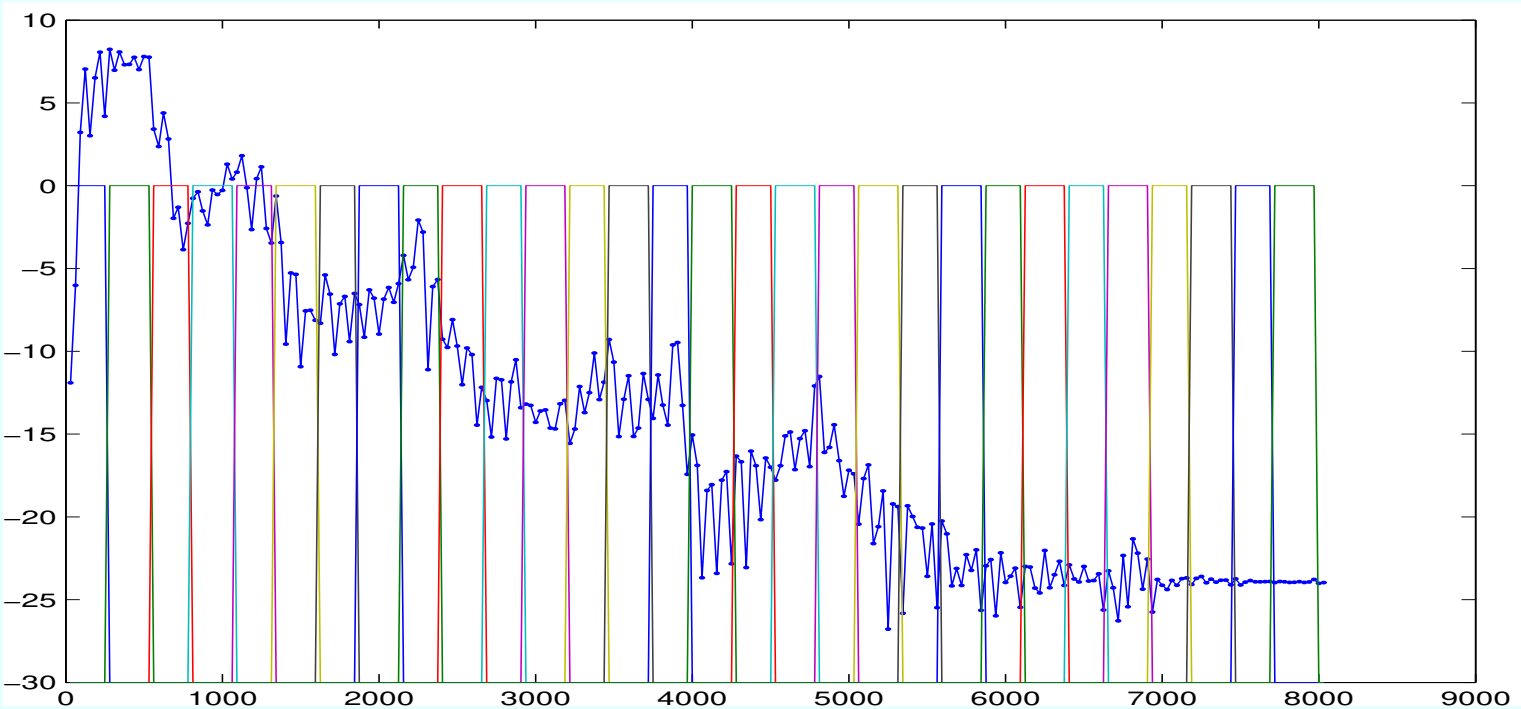
BF je realizovaná na bázi DFT

$\Rightarrow$ filtry jsou dány vahami DFT čar pro dané rozlišení (NDFT) a f_s

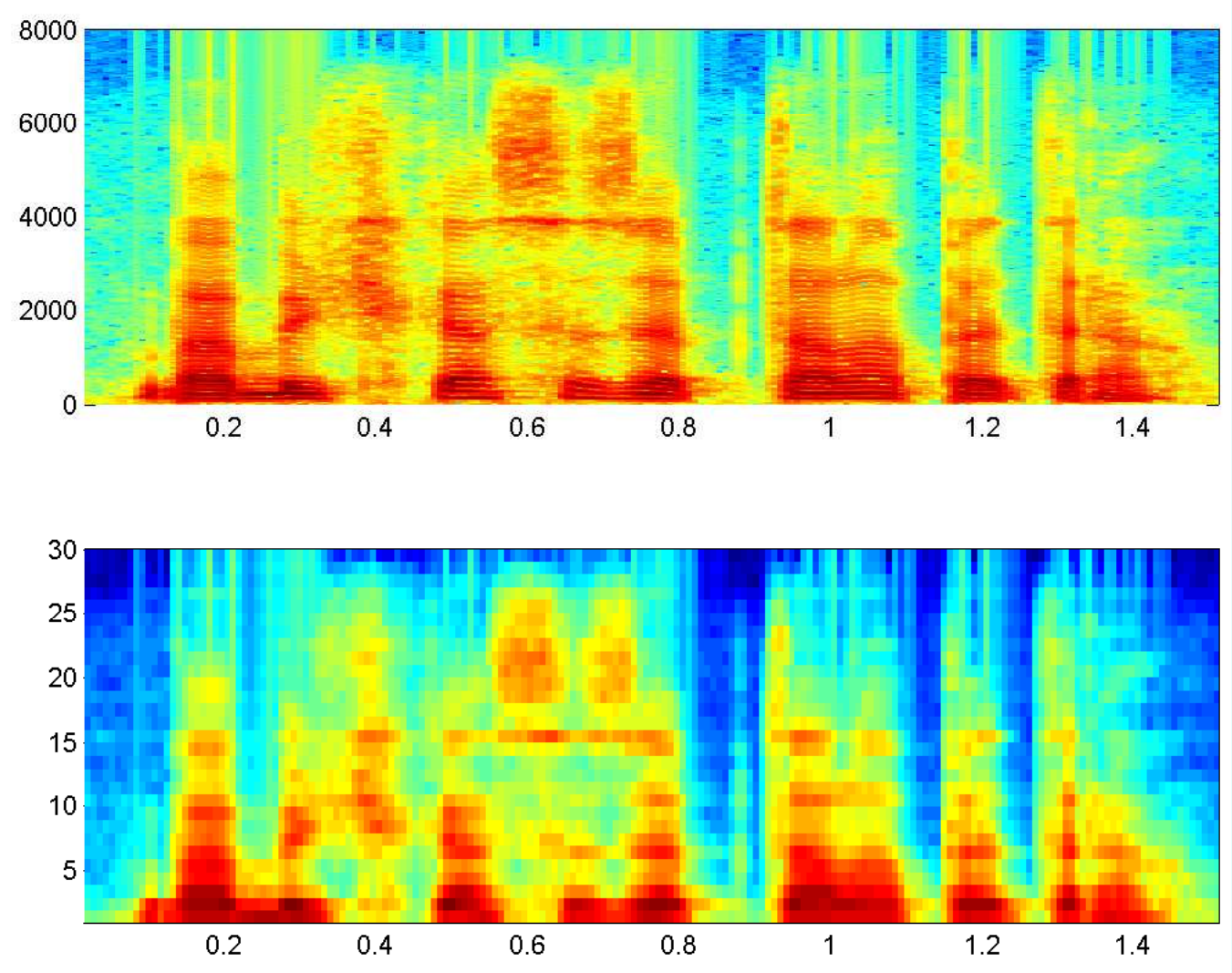
$$G_{mel}[j] = \sum_{k=0}^{N/2} |S[k]|^2 H_j[k] \quad \text{pro j} = 1, \ldots, \text{M}$$

M - počet pásem

- podle f_s, počtu bodů DFT a typu banky filtrů



Banky filtrů ve spektrální analýze



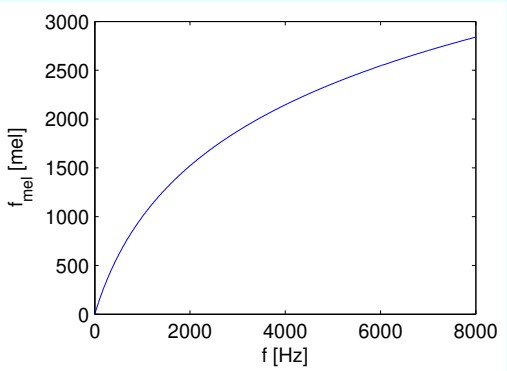
Lineární frekvenční osa - NEVÝHODA - hrubé rozlišení v DKP, jemné rozlišení v HKP
(neodpovídá vnímání frekvence)

Banka filtrů s melovskou nelineární frekvenční osou

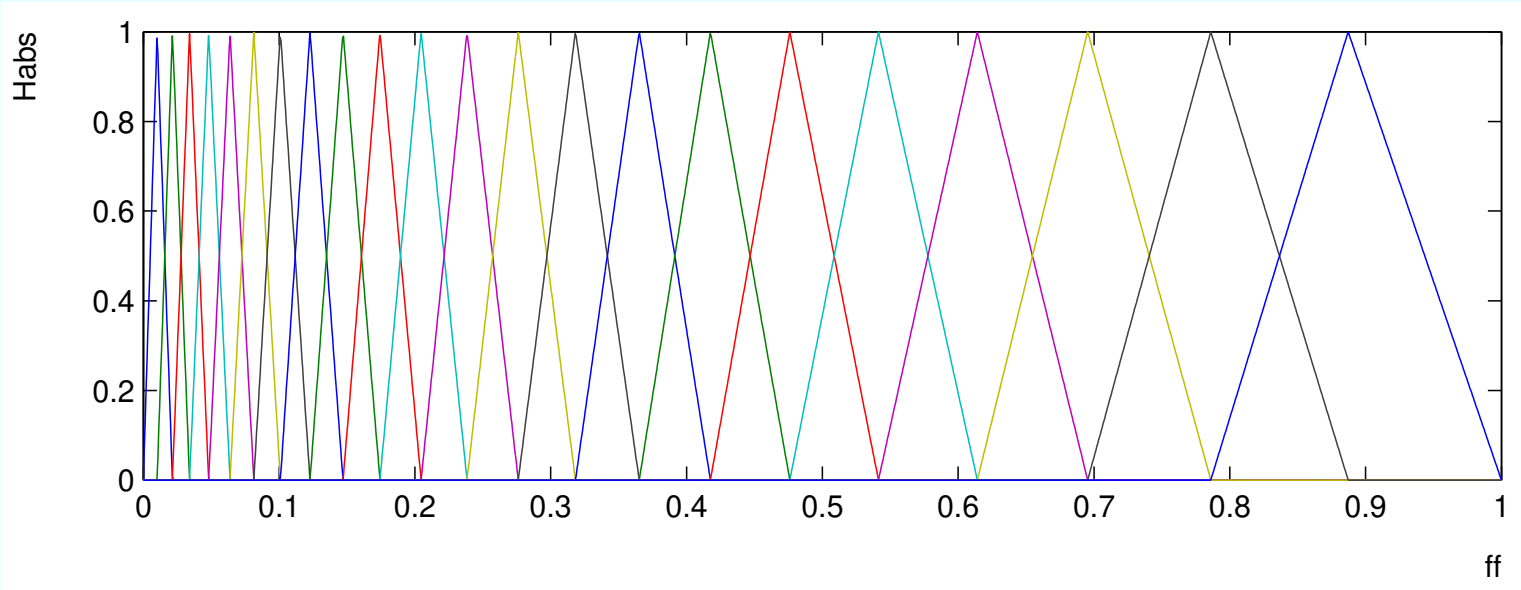
Nelineární zkreslení frekvenční osy - *melodická stupnice*

$$f_{mel} = \mathrm{Mel}(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

$$f = \mathrm{InvMel}(f_{mel}) = 700 \cdot (10^{\frac{f_{mel}}{2595}} - 1)$$



Trojúhelníková melovská BF (používaná pro výpočet MFCC)



Banka filtrů s melovskou nelineární frekvenční osou

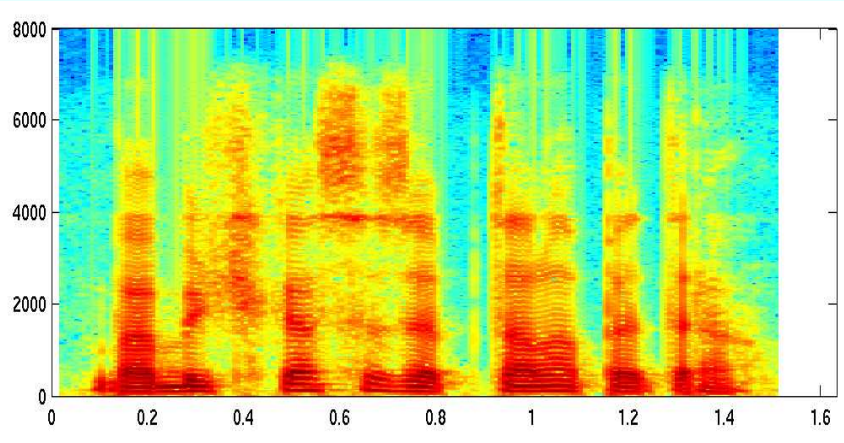
BF je opět realizovaná na bázi DFT

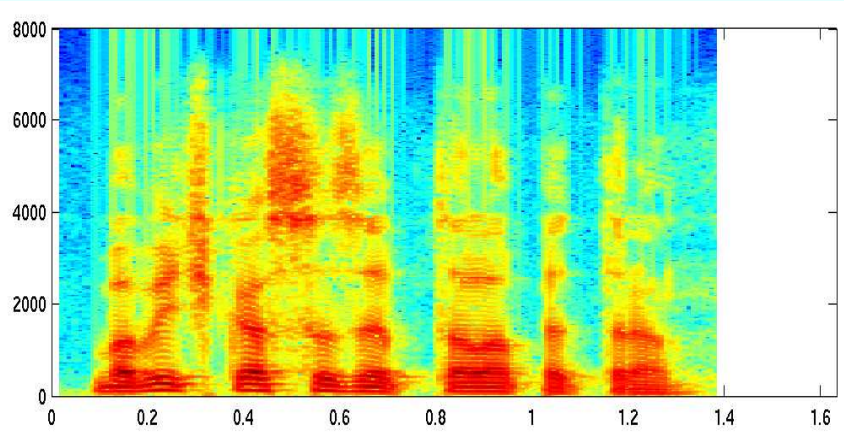
$\Rightarrow$ filtry jsou dány vahami DFT čar pro dané rozlišení (NDFT) a f_s
$\Rightarrow$ princip výpočtu je stejný pro všechny BF
$\Rightarrow$ pro jinou BF pouze jiné konkrétní váhy

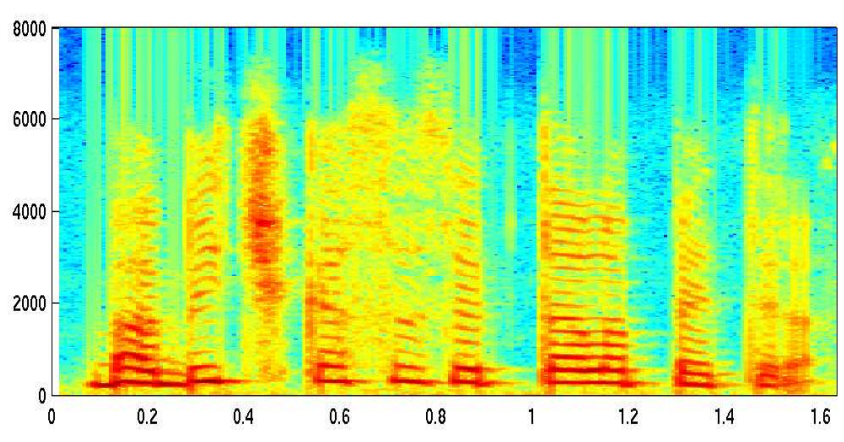
$$G_{mel}[j] = \sum_{k=0}^{N/2} |S[k]|^2 H_{mel,j}[k] \quad \text{pro j} = 1, \ldots, \text{M}$$

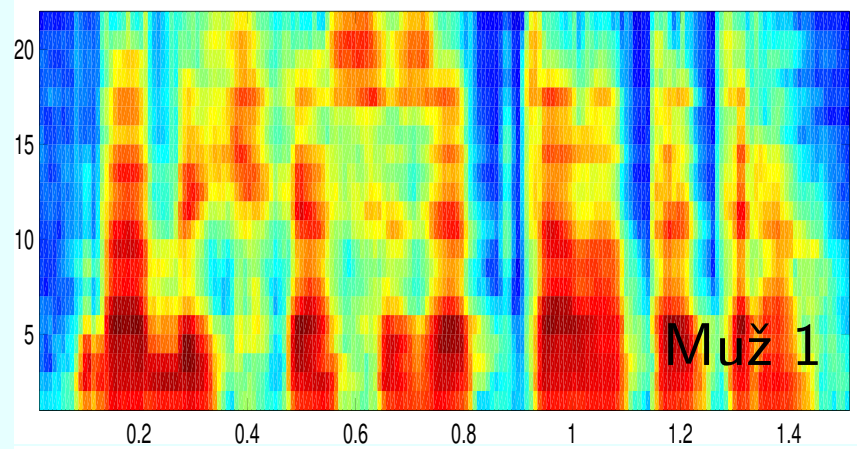
M - počet pásem ..... typické hodnoty 20-30 pásem
- podle f_s a počtu bodů DFT
- 22 pro $f_s = 8$ kHz a segment 25 ms
- 30 pro $f_s = 16$ kHz a segment 25 ms

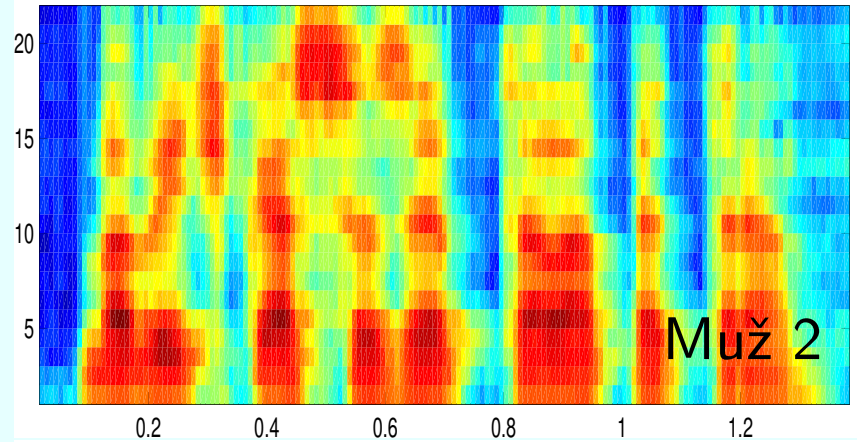
Variabilita promluvy v melovském spektrogramu

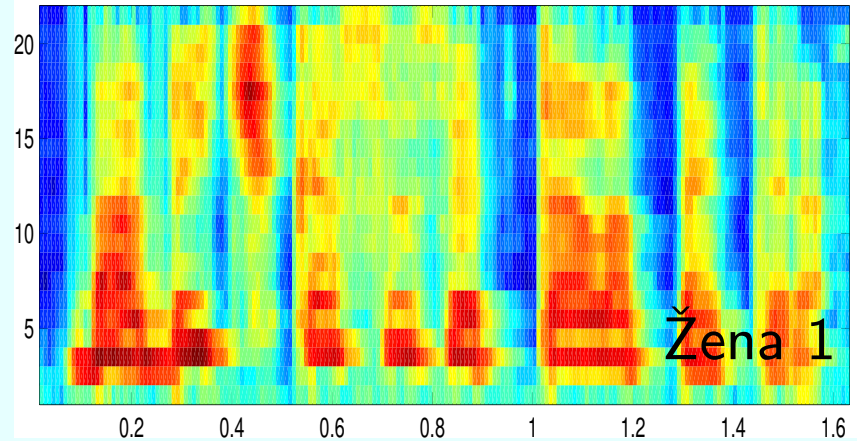








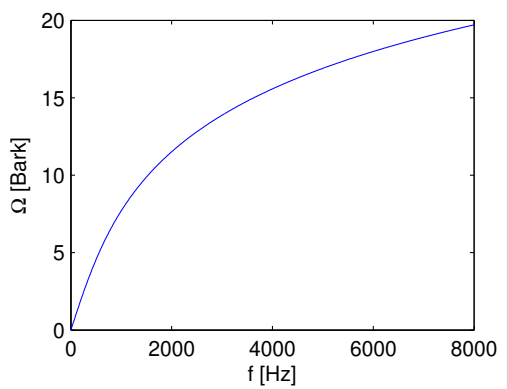




Banka filtrů s Barkovou nelineární frekvenční osou

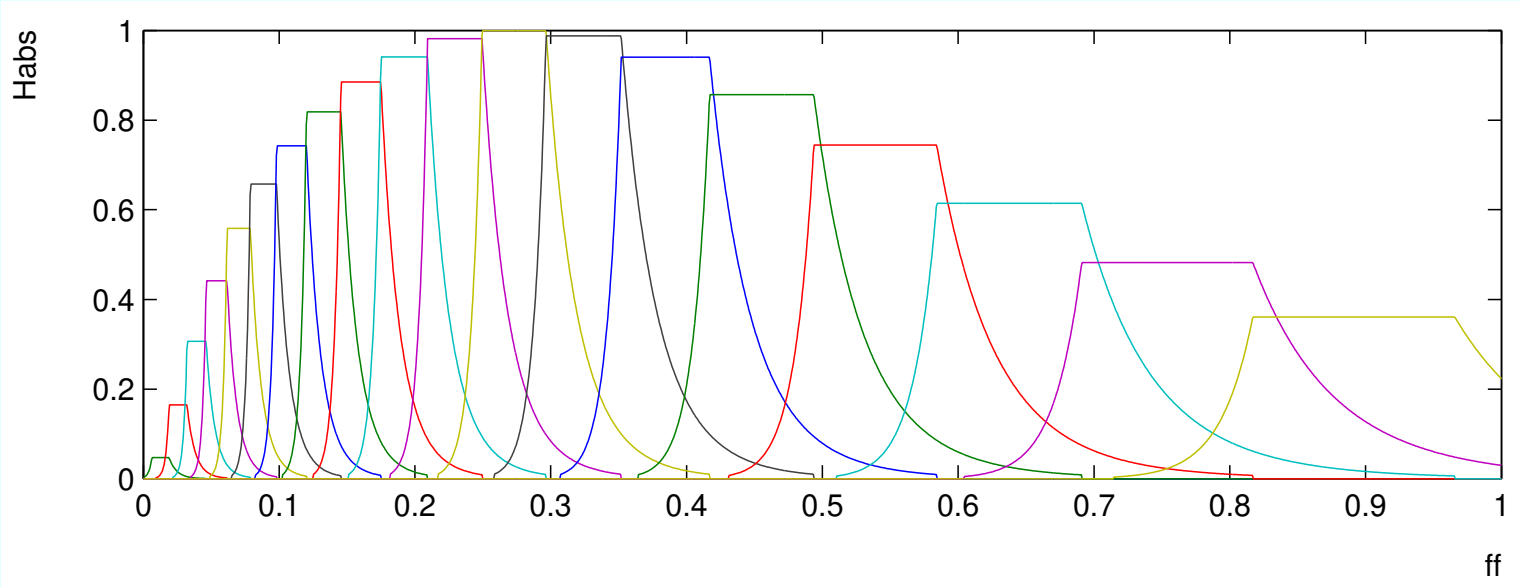
Barkova stupnice - definovaná na bázi kritických pásem slyšení

$$\Omega = \mathrm{Bark}(f) = 6 \ln\left(\frac{f}{600} + \sqrt{\left(\frac{f}{600}\right)^2 + 1}\right)$$



$$f = \mathrm{InvBark}(\Omega) = 600 \cdot \sinh\frac{\Omega}{6}$$

Banka filtrů ve výpočtu PLP kepstrálních koeficientů
(zahrnuje ještě křivky stejné hlasitosti a zákon slyšení)



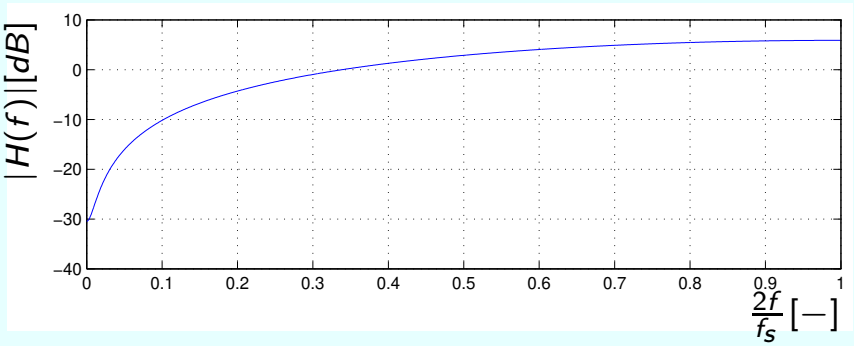
BF je realizovaná opět na bázi DFT (pro dané NDFT a f_s)

Preemfáze signálu - kompenzace útlumu vyšších kmitočtů

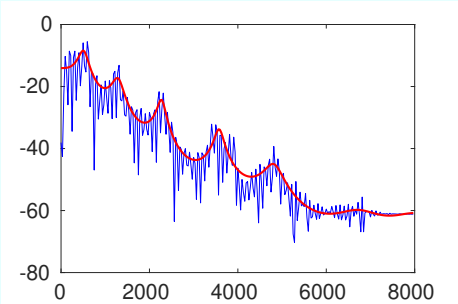
Sklon amplitudového spektra - vyšší kmitočty - nižší energie

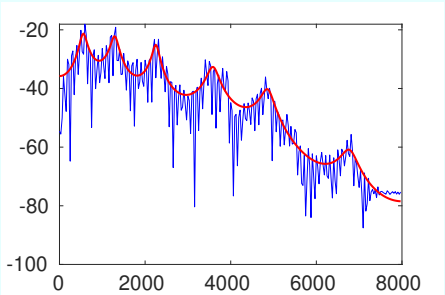
Preemfázový filtr (FIR 1.řád):

$$s'[n] = s[n] - m \cdot s[n-1]$$
$$(m = 0.97)$$

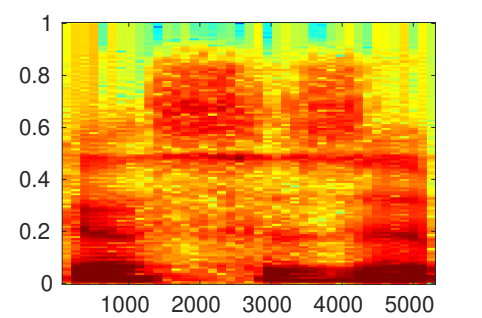


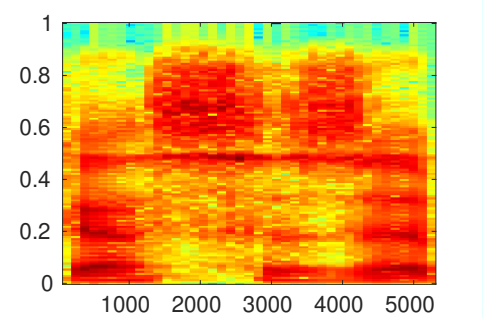
Vliv preemfáze v krátkodobém spektru (DFT a LPC)





Vliv preemfáze ve spektrogramu



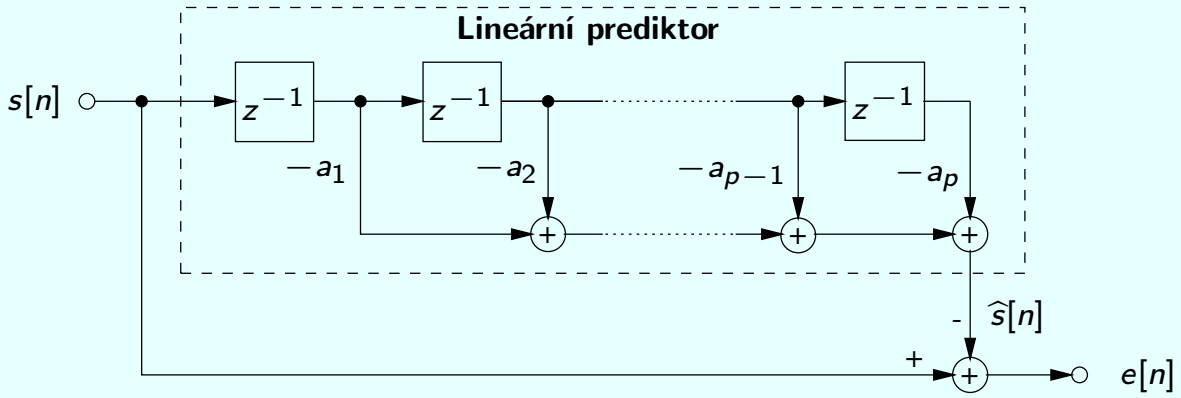


II. část

LPC, AR modelování

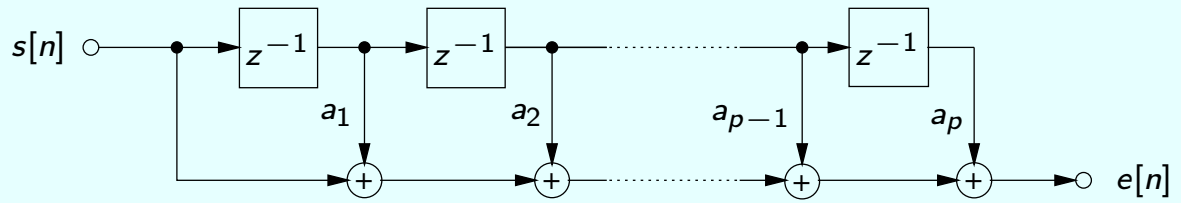
Lineární prediktivní analýza

Lineární predikce : $\widehat{s}[n] = -\sum_{k=1}^{p} a_k s[n-k]$.



Chybový signál (míra kvality prediktoru)

$$e[n] = s[n] - \widehat{s}[n] = s[n] + \sum_{k=1}^{p} a_k s[n-k] = \sum_{k=0}^{p} a_k s[n-k] \ .$$



Princip LPC analýzy

IDEA: přesnější predikce $\rightarrow$ nižší úroveň chybového signálu

Kritérium - výkon chybového signálu

$$J = E\left\{e^2[n]\right\}$$

Hledání koeficientů a_k $\equiv$Minimalizace chyby predikce
$\equiv$ hledání minima J, i.e.

$$\frac{\partial J}{\partial a_k} = 0\ , \quad \text{for} \quad k\ = 1, 2, ..., p \quad \Rightarrow \quad p \text{ lineárních rovnic}$$

Řešení a metody výpočtu (pro různé definice J):

- **autokorelační metoda** - nejčastěji používaný přístup
- Levinson-Durbinův algoritmus (rychlý výpočet autokor.met.)
- Burgův algoritmus - vychází z křížové struktury filtru

Autokorelační metoda, Yuleovy-Walkerovy rovnice

$$\begin{bmatrix} R[0] & R[1] & R[2] & \dots & R[p-1] \\ R[1] & R[0] & R[1] & & R[p-2] \\ R[2] & R[1] & R[0] & \ddots & R[p-3] \\ \vdots & & \ddots & \ddots & \vdots \\ R[p-1] & R[p-2] & R[p-3] & \dots & R[0] \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R[1] \\ R[2] \\ \vdots \\ \vdots \\ R[p] \end{bmatrix}$$

$R[k]$ .... autokorelační koeficienty analyzovaného signálu

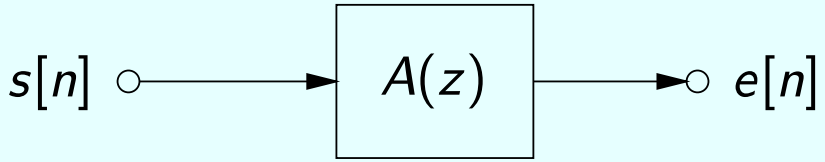
VÝSLEDEK:

a_k .... autoregresní koeficienty (AR model signálu)

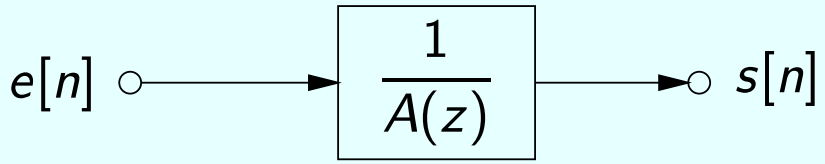
$P_p = R[0] + \sum_{k=1}^{p} a_k R[k]$ .... výkon chybového signálu

AR model signálu

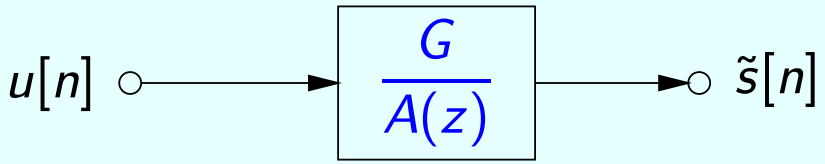
Dekorelační (analyzující) filtr : $A(z) = \sum_{k=0}^{p} a_k z^{-k}$



Syntéza se skutečným chybovým signálem (ideální případ)



Syntéza s umělým signálem s jednotkovým výkonem (AR model)
- G závisí na úrovni analyzovaného signálu ($G = \sqrt{P_p}$)



$$H(z) = \frac{G}{A(z)} = \frac{G}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_p z^{-p}}$$

- AR model je základ pro kódování řeči resp. formantovou syntézu

Spektrální vlastnosti AR modelu

Obecný popis AR modelu v Z-oblasti

$$\tilde{S}(z) = H(z) \cdot U(z)$$

Popis AR modelu ve frekvenční oblasti

$$S_{\tilde{s}}(e^{j\Theta}) = |H(e^{j\Theta})|^2 \cdot S_u(e^{j\Theta})$$

Vlastnosti a důsledky: - $S_u(e^{j\Theta})$ je ploché
$\rightarrow$ tvar $S_{\tilde{s}}(e^{j\Theta})$ je kompletně zahrnut v AR modelu

$$\Downarrow$$

LPC spektrum (pokud $S_u(e^{j\Theta}) = 1$) $\qquad S_{\tilde{S}}(e^{j\Theta}) = |H(e^{j\Theta})|^2$

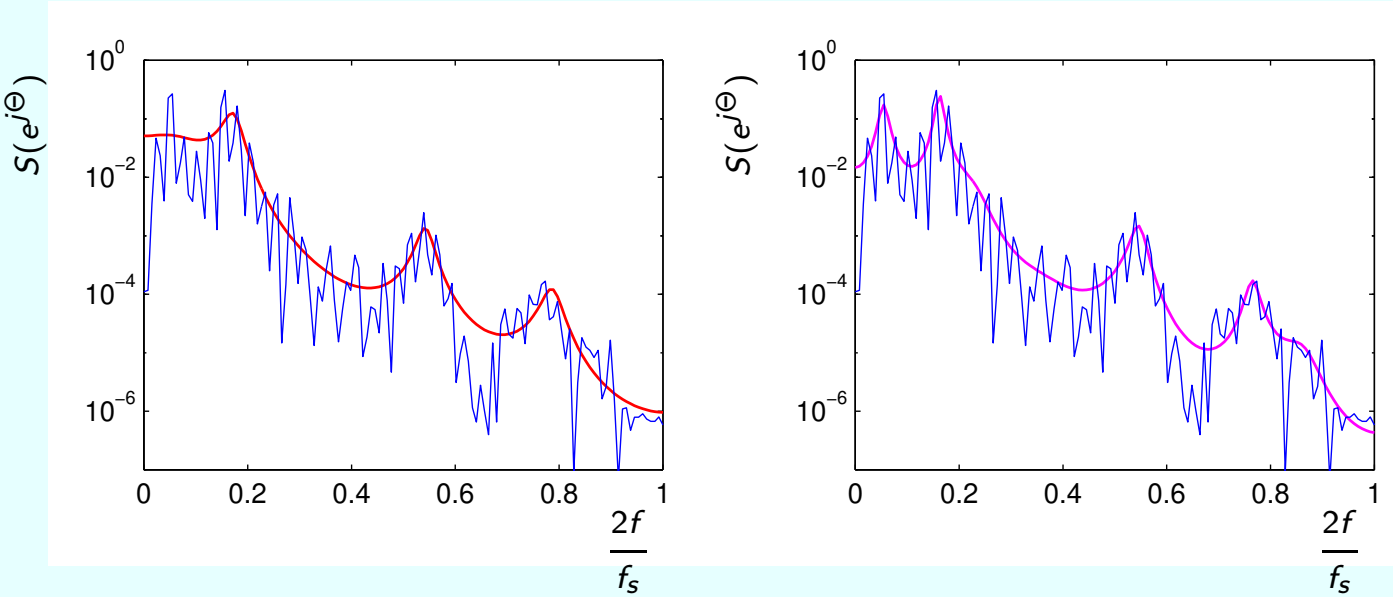
$$S_{\tilde{S}}(e^{j\Theta}) = \frac{G^2}{|A(e^{j\Theta})|^2} = \frac{G^2}{|1 + a_1 e^{-j\Theta} + a_2 e^{-j2\Theta} + \ldots + a_p e^{-jp\Theta}|^2}$$

$$\Downarrow$$

koeficienty a_k .... komprimovaná spektrální reprezentace

Srovnání LPC a DFT spektra

$$S_{\tilde{s}}(e^{j\Theta}) = |H(e^{j\Theta})|^2 \quad \approx \quad \frac{|S[k]|^2}{N}$$



- AR model: "all-pole" filtr, modeluje pouze špičky ve spektru (rezonátory v dutinách vokálního traktu)
- obecná špička = dvojice komplexně združených pólů
- reálný pól modeluje špičku v 0 nebo $f_s/2$
- vyšší řád AR modelu = více špiček v LPC spektru
  $\rightarrow$ typické hodnoty: $p = 10$ pro $f_s = 8$ kHz, $p = 16$ pro $f_s = 16$ kHz

Levinson-Durbinův algoritmus

Rychlý a **rekurentní** výpočet koeficientů a_k autokorelační metodou (vychází stále z celkové chyby predikce)

Inicializace:

$$P_0 = R[0] \quad a_1^{(1)} = k_1 = -\frac{R[1]}{R[0]}$$

$$P_1 = P_0 \cdot (1 - k_1^2)$$

Výpočet pro $m = 2, 3, \ldots, p$**:**

$$a_m^{(m)} = k_m = -\frac{R[m] + \sum_{j=1}^{m-1} a_j^{(m-1)} R[m-j]}{P_{m-1}}$$

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

$$P_m = P_{m-1} \cdot (1 - k_m^2)$$

Výsledek:

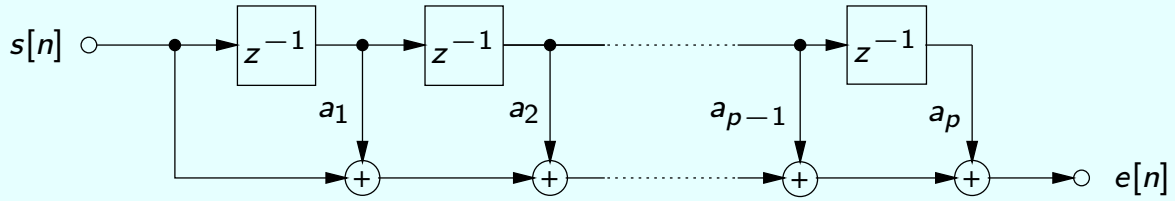
$$a_i = a_i^{(p)}, \quad i = 1, 2, \ldots, p$$

k_k .... koeficienty odrazu (křížová struktura filtru)

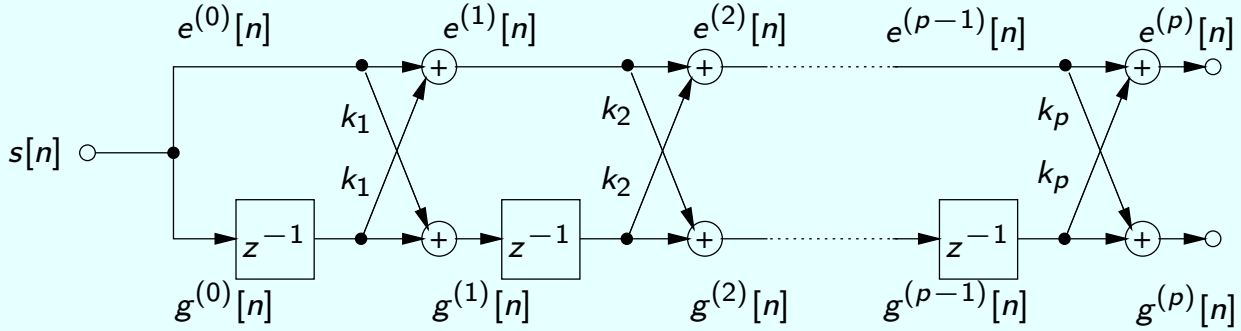
.... PARCOR koeficienty (partial correlation coefficients)

Křížová struktura při LPC analýze

Trasnverzální struktura analyzujícího FIR filtru:



Křížová struktura analyzujícího FIR filtru:



k_k ..... koeficienty odrazu, přepočet k_k vs. a_k - Levinsonova rekurze

Inicializace:

$$a_1^{(1)} = k_1$$

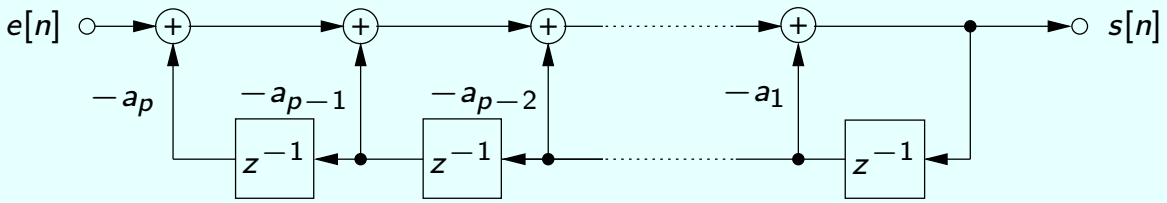
Výpočet pro $m = 2, 3, \ldots, p$**:**

$$a_m^{(m)} = k_m$$

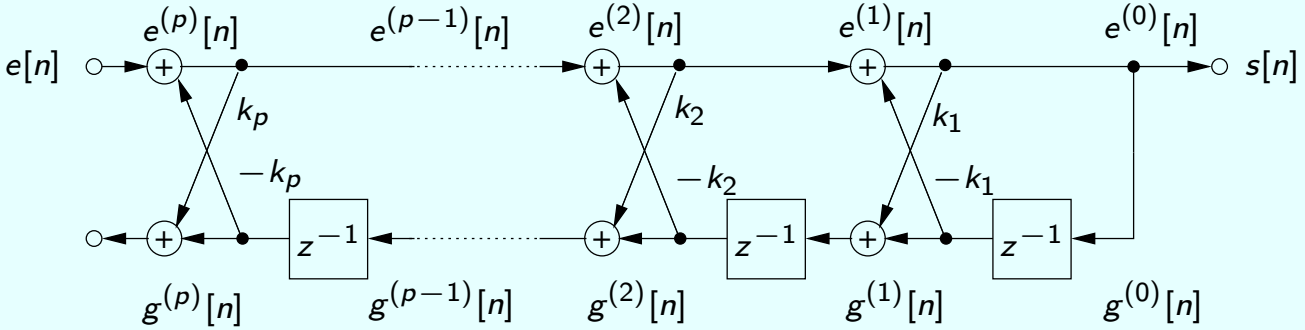
$$a_j^{(m)} = a_j^{(m-1)} + k_m a_{m-j}^{(m-1)}, \quad j = 1, 2, \ldots, m-1$$

AR model - standardní a křížová struktura

Trasnverzální struktura syntetizujícího all-pole IIR filtru:



Křížová struktura syntetizujícího all-pole IIR filtru:



Vlastnosti koeficientů odrazu k_k :

- stabilní syntetizující filtr pro $-1 < k_k < 1$
- více robustní pro malou variabilitu signálu než a_k ($\rightarrow$ vhodné příznaky)
- vhodné pro implementace (menší problém při kvantování)
- možné interpolace
- možný **přímý výpočet koeficientů odrazu** $\rightarrow$ Burgův algoritmus

Burgův algoritmus

Minimalizační kritérium (pro každou sekci křížové struktury):

$$J_m = \frac{1}{2}\sum_{n=0}^{N-1}\left[\left(e^{(m)}[n]\right)^2 + \left(g^{(m)}[n]\right)^2\right] \qquad \text{pro } m = 1,\ 2,\ \ldots,\ p\ .$$

Inicializace: $$e^{(0)}[n] = g^{(0)}[n] = s[n]$$

Výpočet pro $m = 1, 2, 3, \ldots, p$**:**

$$k_m = -\frac{2\cdot\sum_{n=m}^{N-1}\left(e^{(m-1)}[n]\cdot g^{(m-1)}[n-1]\right)}{\sum_{n=m}^{N-1}\left(e^{(m-1)}[n]\right)^2 + \sum_{n=m}^{N-1}\left(g^{(m-1)}[n-1]\right)^2}$$

Vždy platí $|k_m| < 1 \quad \longrightarrow \quad$ vždy stabilní řešení

$$e^{(m)}[n] = e^{(m-1)}[n] + k_m \cdot g^{(m-1)}[n-1], \quad n = 0,\ 1,\ \ldots,\ N-m$$

$$g^{(m)}[n] = g^{(m-1)}[n-1] + k_m \cdot e^{(m-1)}[n], \quad n = 0,\ 1,\ \ldots,\ N-m$$

Další výpočty: - autoregresní koeficienty a_k - Lev. rek., viz L.-D. alg.
- výkon chyby predikce P_k - viz L.-D. alg.

Výpočet AR koeficientů a LPC spektra v MATLABu

1 Výpočet parametrů AR modelu

- Autokorelační metoda (Yule-Walker) :

  funkce lpc ..... `[a, Ep] = lpc ( s, p ) ;`
  funkce aryule ..... `[a, Ep, rc] = aryule ( s, p ) ;`

- Burgův algoritmus :

  funkce arburg ..... `[a, Ep, rc] = arburg ( s, p ) ;`

  - s . . . analyzovaný signál
  - p . . . řád AR modelu
  - a . . . autoregresní koeficienty (včetně $a_0 = 1$)
  - Ep . . . výkon chybového signálu
  - rc . . . koeficienty odrazu

2 Výpočet LPC spektra

funkce freqz ..... `H = freqz ( sqrt(Ep), a, N ) ;`

- H . . . komplexní LPC spektrum
- N . . . počet bodů LPC spektra

III. část

Odhad formantů

Formanty - definice

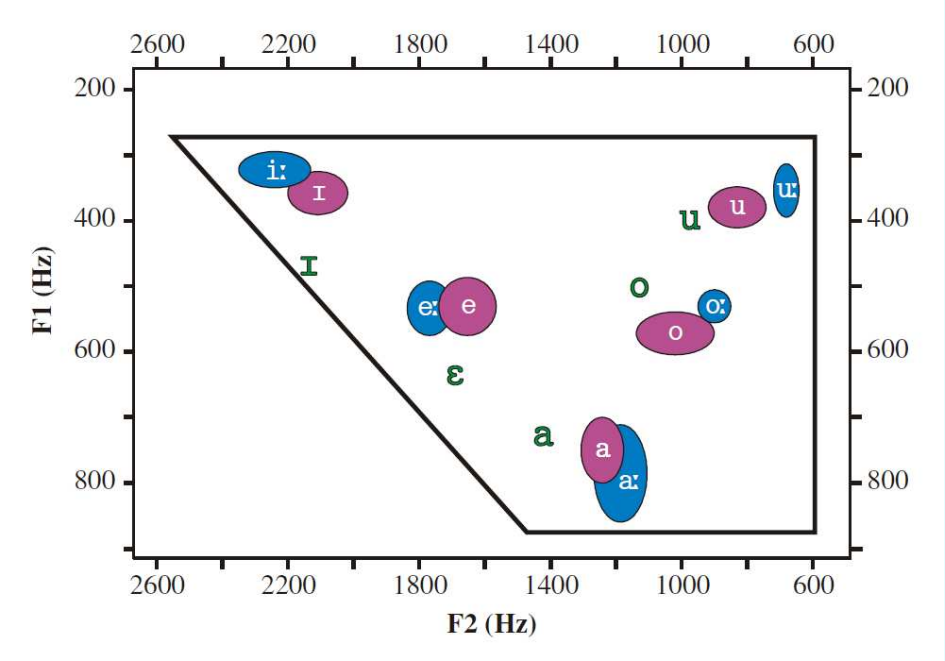
- **Formant (formantové frekvence)**
  $\rightarrow$ centrální kmitočty rezonátorů vokálního traktu

- významné špičky ve **VYHLAZENÉM krátkodobém spektru**
- významné formanty F1 - F4 v pásmu do 4 kHz
- F5 - méně významný (obtížně odhadnutelný formant)
- !! Nezaměňovat se základním tónem řeči f_0 !!
  (f_0 není detekovatelné ve vyhlazeném spektru)

- **Aplikace:**
  - elementární analýzy
  - formantová syntéza řeči
  - transformace hlasových charakteristik (Lombardův jev)

Významné formanty samohlásek

	I	E	A	O	U
F1	300 - 500	480 - 700	700 - 1100	500 - 700	300 - 500
F2	2000 - 2800	1560 - 2100	1100 - 1500	850 - 1200	600 - 1000
F3	2600 - 3500	2500 - 3000	2500 - 3000	2500 - 3000	2400 - 2900



	přední	střední	zadní
vysoké	i		u
středové	e		o
nízké		a	

Formanty - metody odhadu

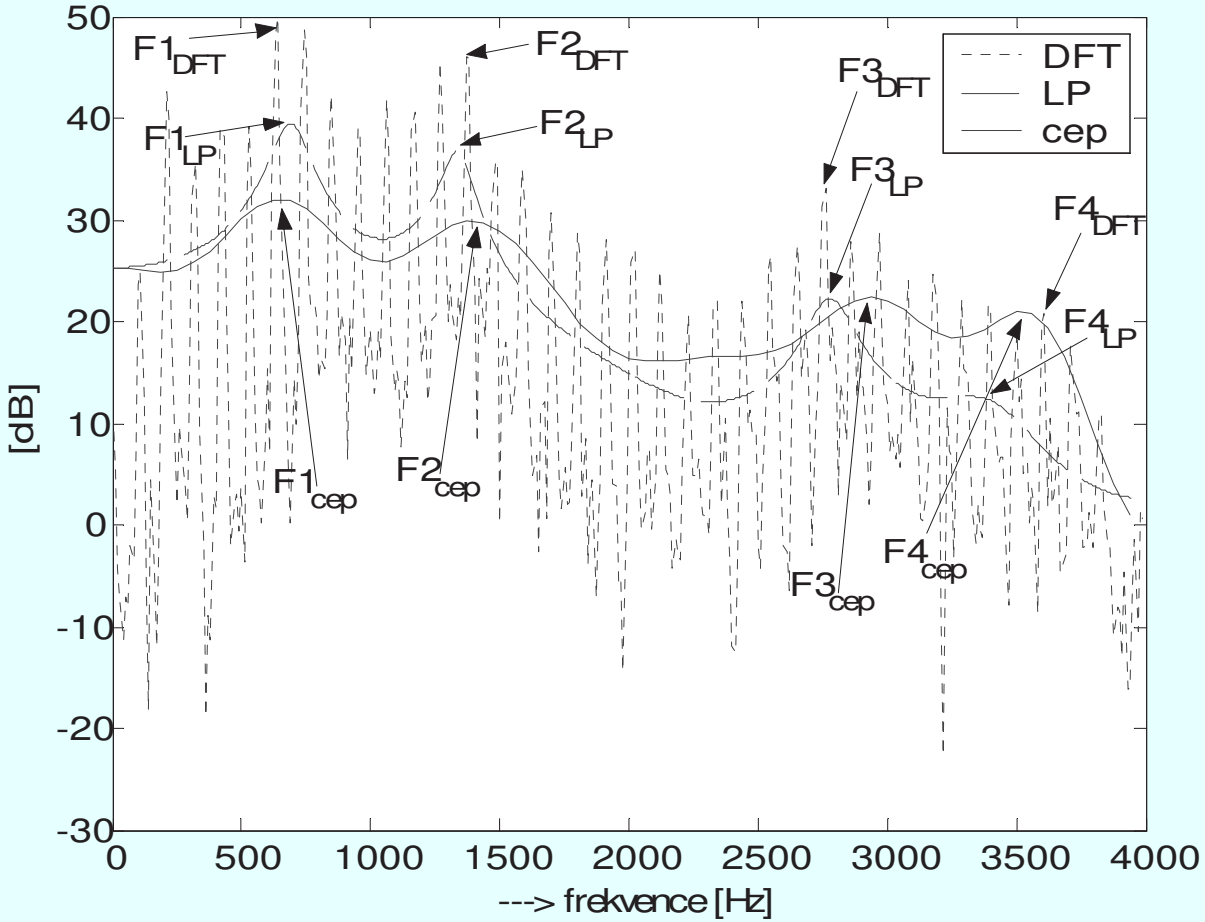
- **z vyhlazeného DFT spektra**
  - krátké okno, doplnění nulami, hledání maxima
  - nepřesné

- **Pomocí LPC**
  - LPC analýza - vyhlazené spektrum
  - špičky LPC spektra - rezonátory vokálního traktu
  - nejčastěji používaná technika

- **Pomocí kepstrální analýzy**
  - vyhlazené spektrum pomocí zkrácení reálného kepstra
  - hledání maxim

Formanty - krátkodobé spektrum

Hláska 'a' - formanty ve vyhlazeném a nevyhlazeném spektru



Formanty - Odhad na bázi LPC

- špičky LPC spektra - rezonátory = formanty
- špičky LPC spektra - určené **póly přenosové funkce** p_i

$$F_i = \frac{\arg\, p_i}{2\pi} \cdot f_s$$

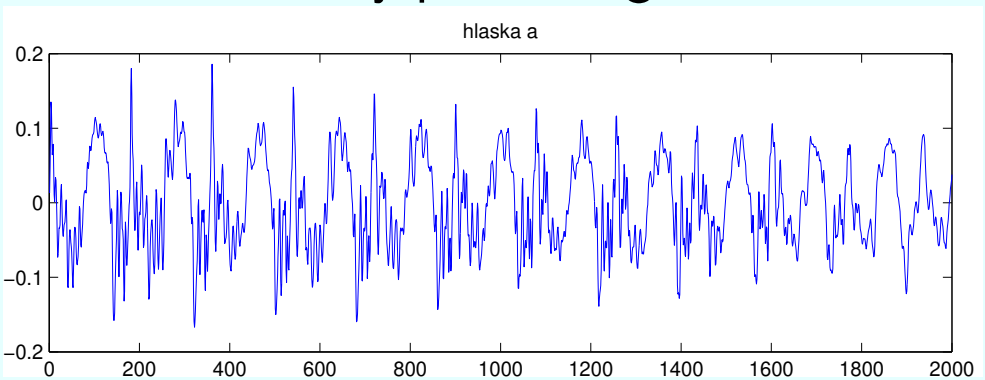
$$B_i = -\frac{\ln |p_i|}{\pi} \cdot f_s$$

- F_i - formantová frekvence (centrální kmitočet rezonátoru)
- B_i - šířka pásma formantu

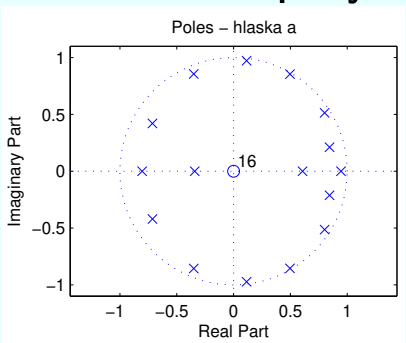
- Problémy:
  - obecně menší robustnost LPC analýzy (závislost na datech)
  - určení vhodného řádu (vliv přítomnosti šumu)
  - seřazení vypočítaných pólů (sledování stejného formantu)
  - vyřazení nadbytečného pólu (méně významné špičky)

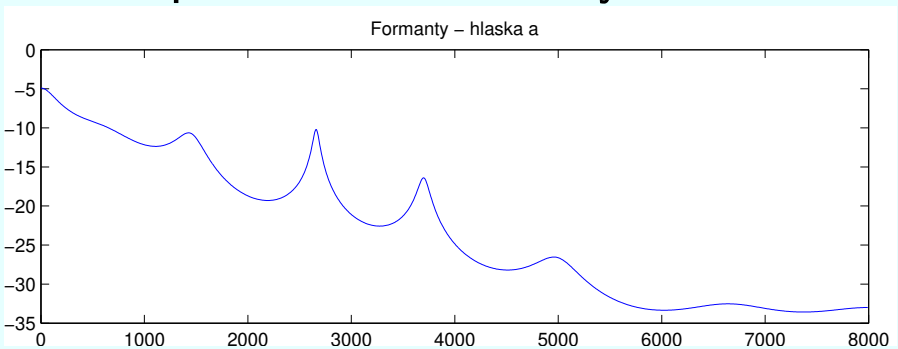
Odhad formantů na bázi LPC - příklad

Časový průběh signálu

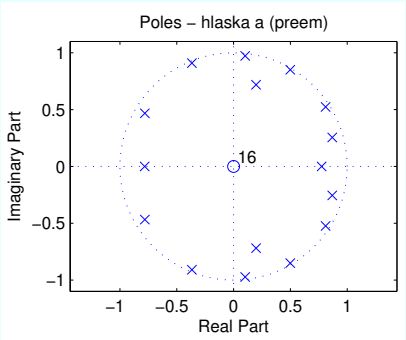


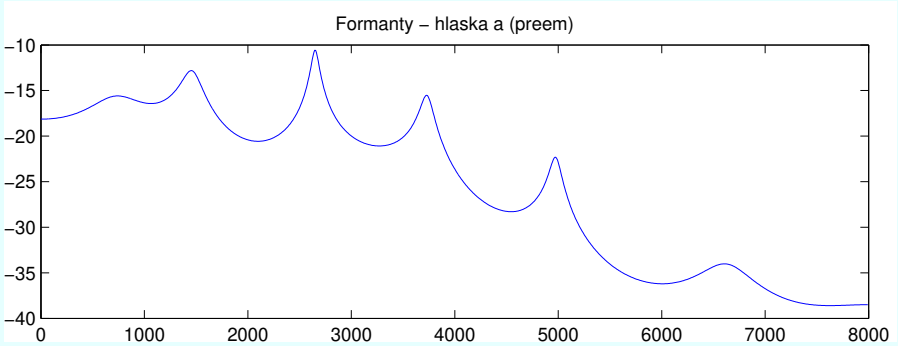
póly & LPC spektrum s formanty





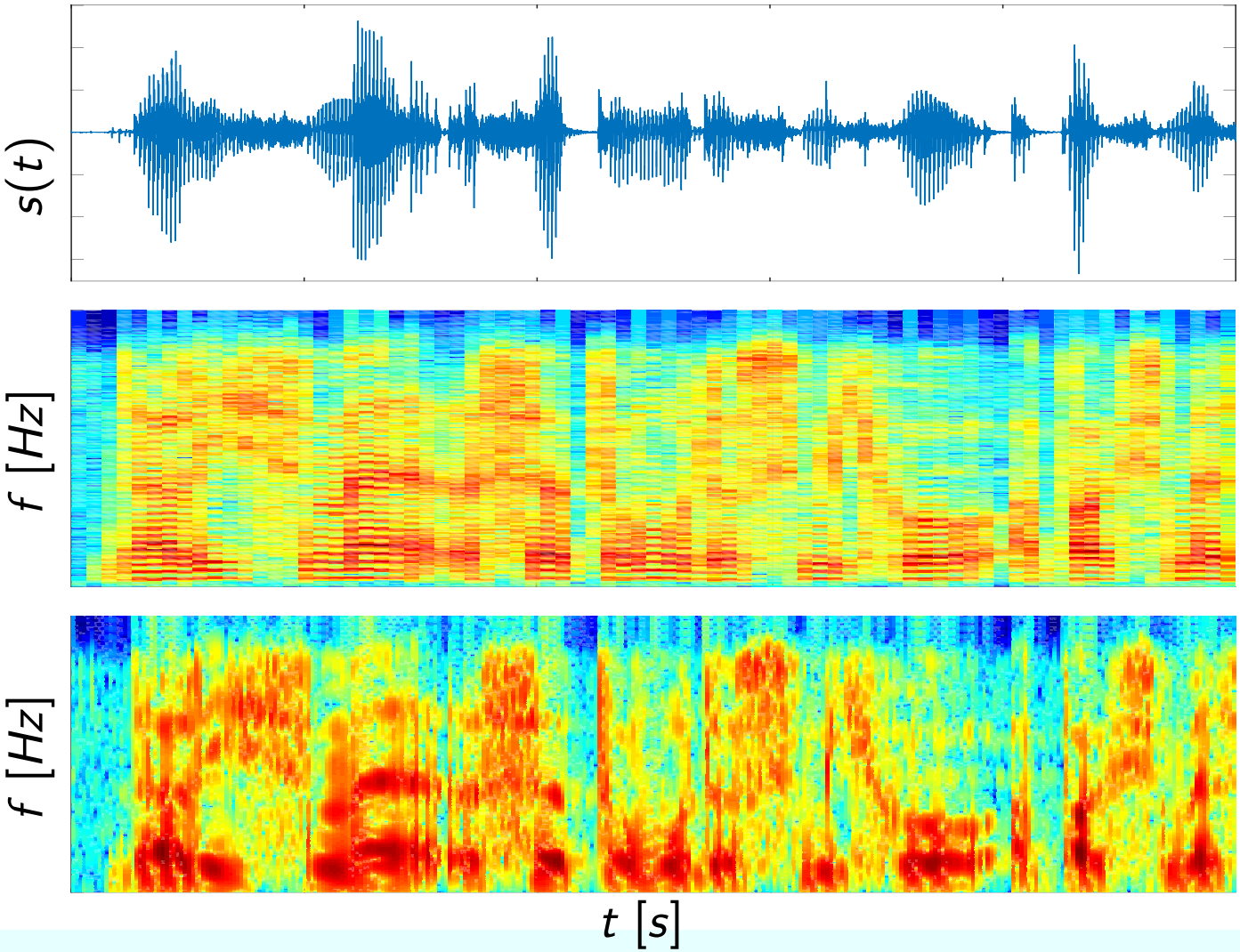
póly & LPC spektrum s formanty (preemfáze)





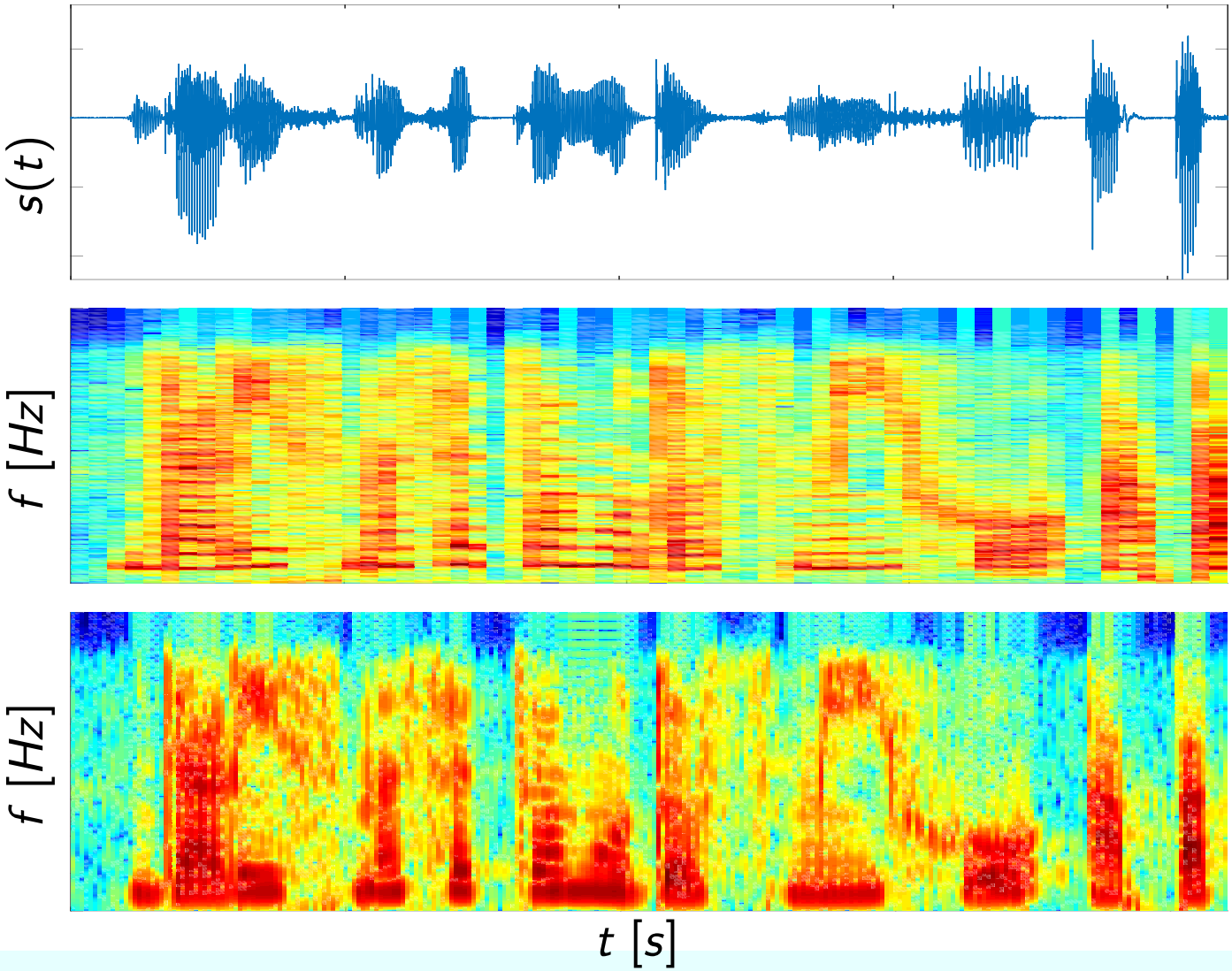
Formanty v DFT spektrogramu

Mužský hlas - dlouhé vs. krátké analyzující okno
(harmonické komponenty vs. vyhlazené spektrum)



Formanty v DFT spektrogramu

Ženský hlas - dlouhé vs. krátké analyzující okno
(harmonické komponenty vs. vyhlazené spektrum)



Děkuji za pozornost