A6M33BIO - Biometrie

Biometrické metody založené na rozpoznávání hlasu I

Doc. Ing. Petr Pollák, CSc.

13. prosince 2018 - 16:24

Obsah přednášky

- **Úvod**
  - Aplikace a principy hlasové biometrické verifikace
  - Základní popis vzniku řeči (hlasu)
- **Řečové charakteristiky a příznaky pro identifikaci**
  - Základní frekvence řeči
  - Spektrální charakteristiky
  - Formanty
  - Expertní identifikace, spektrografické metody
  - Kepstrum, kepstrální vzdálenost

I. část

Identifikace řečníka (základní principy, variabilita hlasu)

Aplikační oblasti úlohy rozpoznávání mluvčího

1) přesná identifikace totožnosti mluvčího

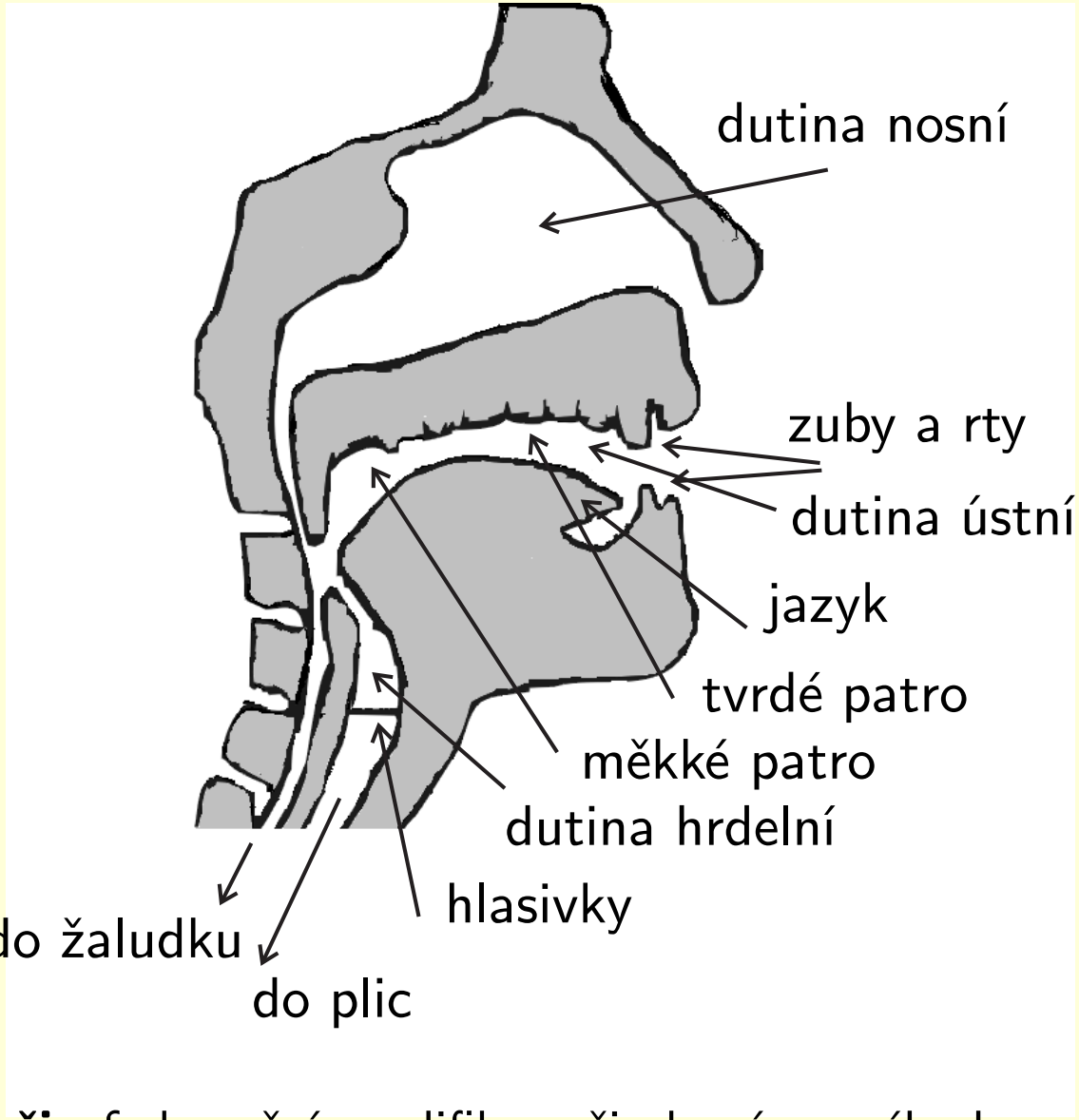
- kriminalistická a soudní praxe - forenzní aplikace
  (dosud subjektivní fonetická a lingvistická analýza)
- identifikace pro přístup k zabezpečeným systémům
  (bankovní účty, přístup do chráněných objektů/systémů,
  bezpečnostní kontroly, apod.)

2) identifikace mluvčího s největší podobností hlasu

- *př. - identifikace volajících v call-centrech*
- *komplexní rozpoznávače řeči*
  (LVCSR - diktovací systémy, transkripční systémy pro přepis
  rozhlasových/TV zpravodajství)
  - modely pro konkrétního mluvčího
  - skupinové modely
  - modely závislé na pohlaví mluvčího

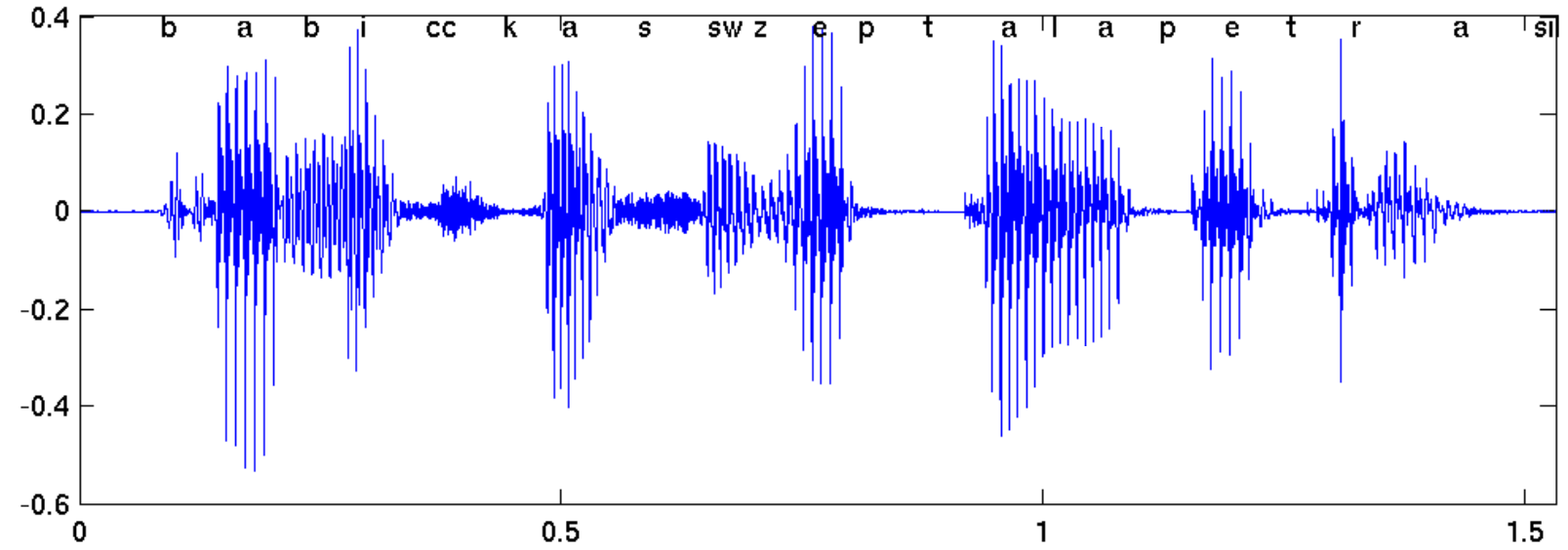
Produkce řeči - fyziologie hlasového ústrojí

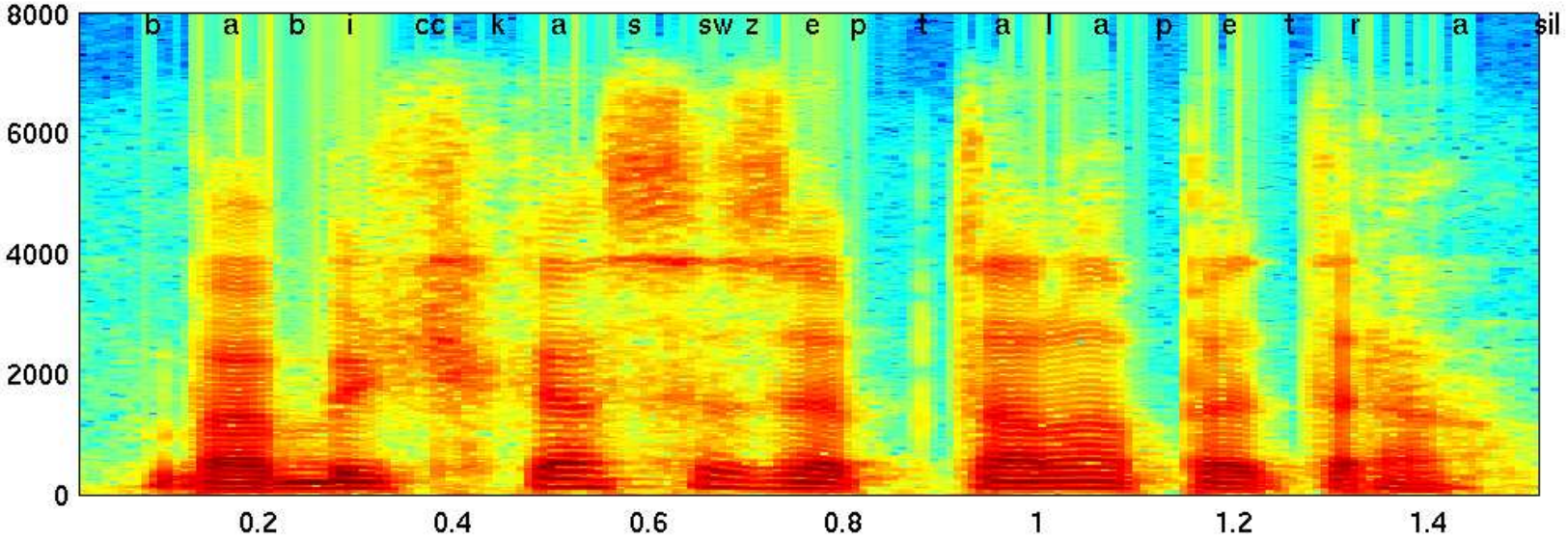
Artikulační orgány hlasového ústrojí člověka



Produkce řeči : frekvenční modifikace širokopásmového buzení proudem vzduchu průchodem dutinami (rezonátory) hlasového ústrojí

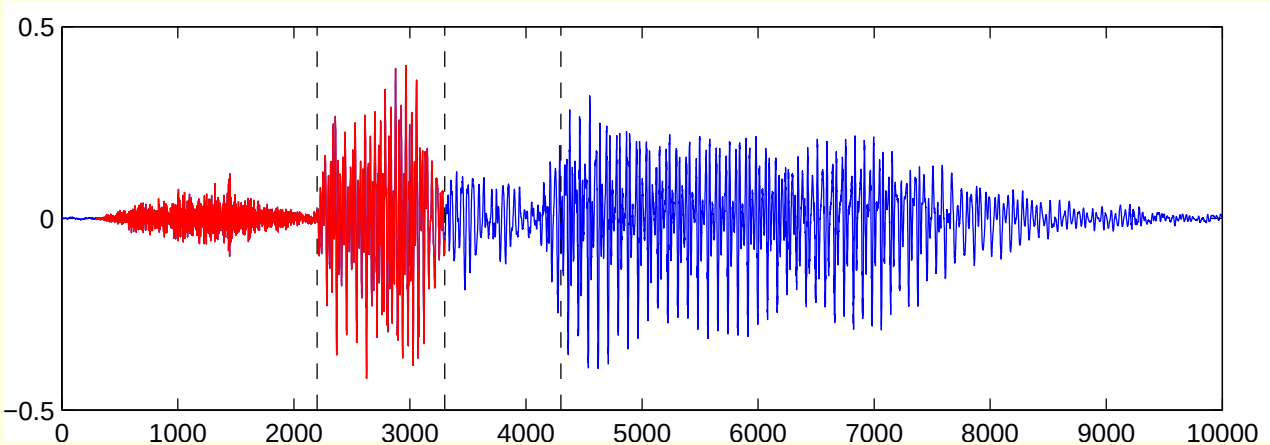
Časový a spektrální obsah řeči



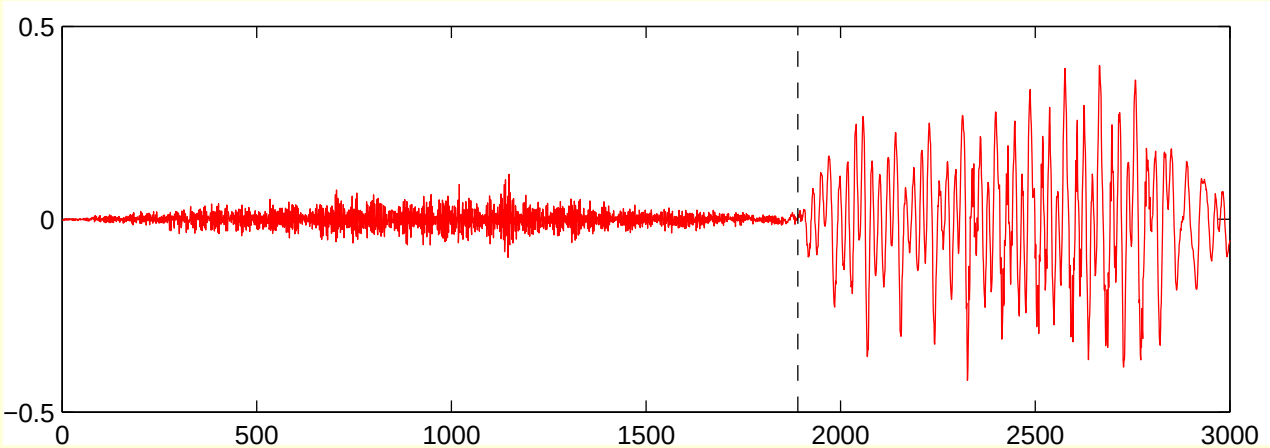


Řečové hlásky v časové oblasti

Slovo **“šedý”**



Slabika **“še”**

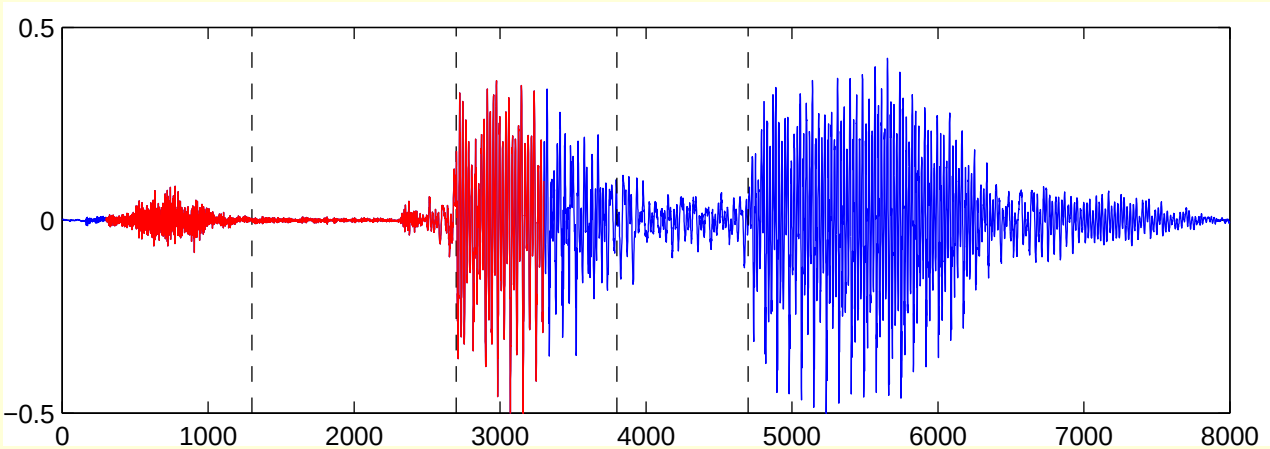


Hláska **“š”** ... neznělá, šumový charakter

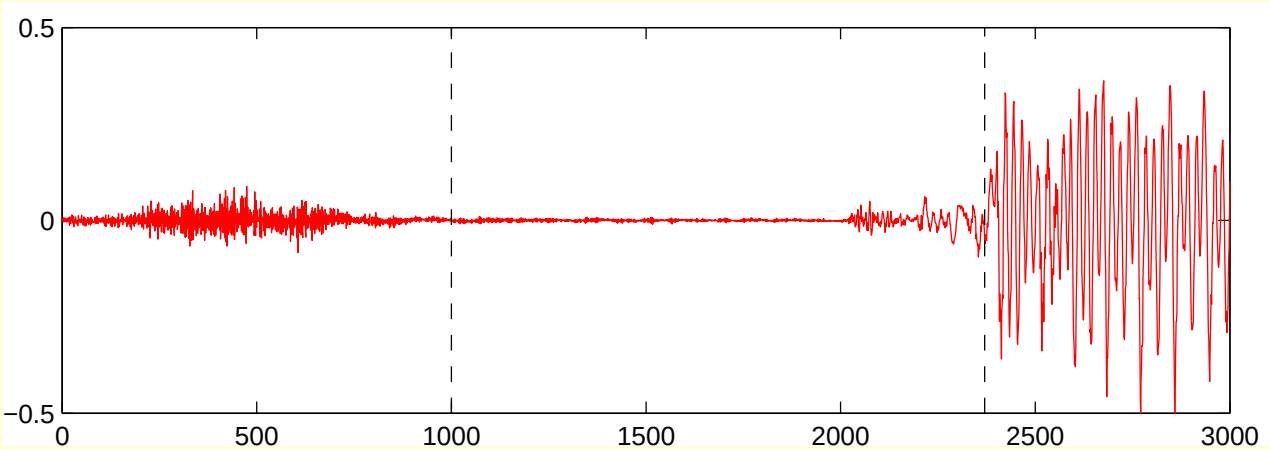
Hláska **“e”** ... znělá, periodický charakter (harmonická struktura)

Řečové hlásky v časové oblasti

Slovo **“čtyři”**



Slabika **“čty”**

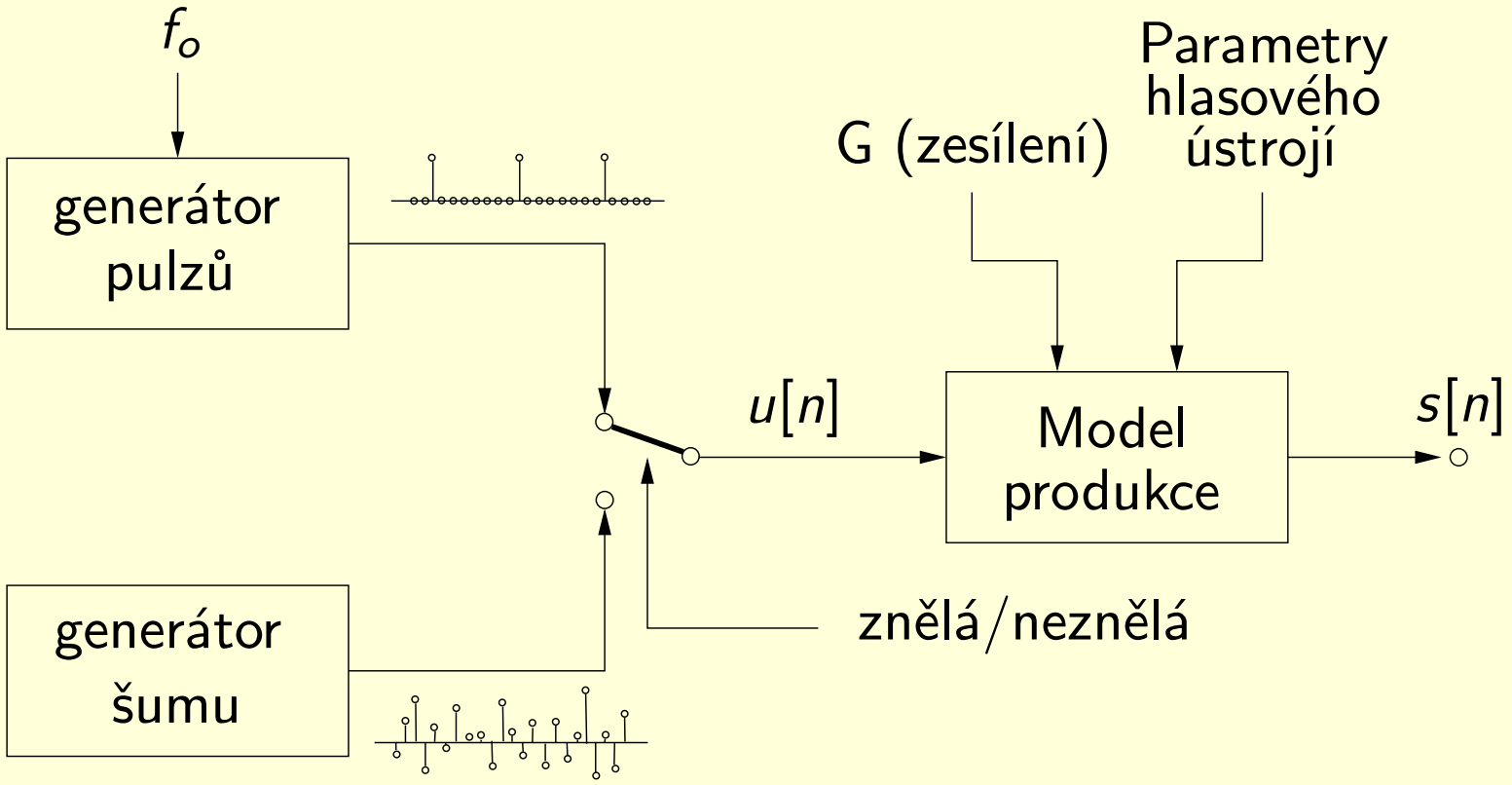


Hláska **“č”** ... neznělá, šumový charakter

Hláska **“t”** ... plozivní, okluze (závěr) + exploze

Hláska **“y”** ... znělá, periodický charakter (harmonická struktura)

Produkce řeči - signálový model vzniku řeči



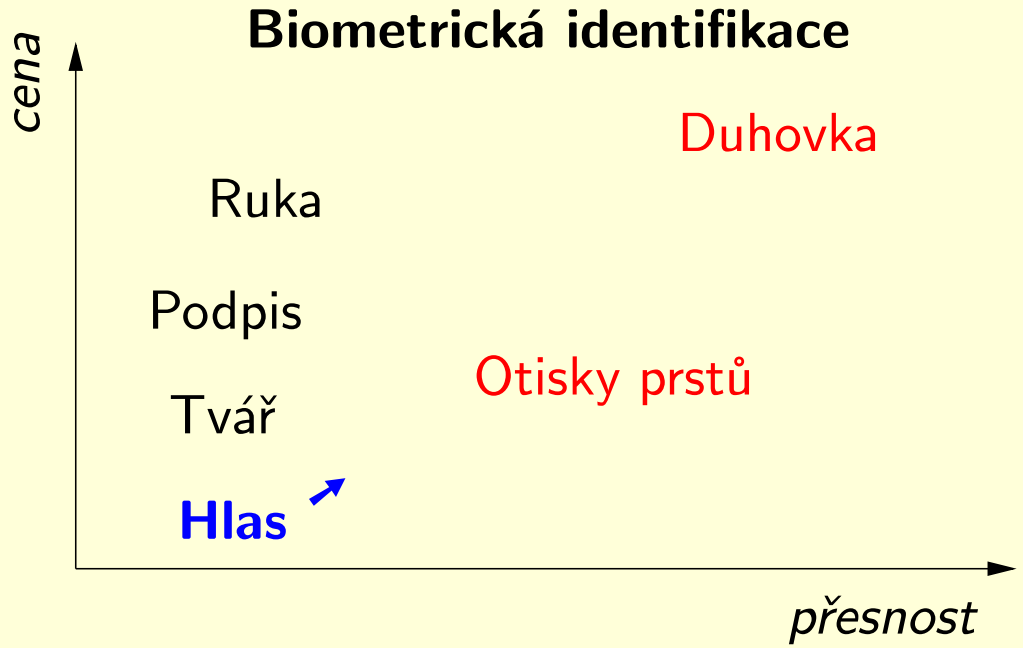
Závislost na mluvčím (rozdíly v rozměrech hlasového ústrojí):

1. vokální trakt (barva hlasu)
   - souvislost s anatomií rezonátorů (dutin) hlasového ústrojí
   - AR model, snadná identifikace parametrů (LPC)
2. generování pulzů (výška hlasu) - daná vlastnostmi hlasivek

Variabilita řeči

Odlišnost mluvčích :

- originalita hlasu - výška a barva
- originalita stylu - různá doba trvání hlásek
- obecná variabilita jednotlivých realizací - PROBLÉM
- možnost napodobení hlasu - PROBLÉM



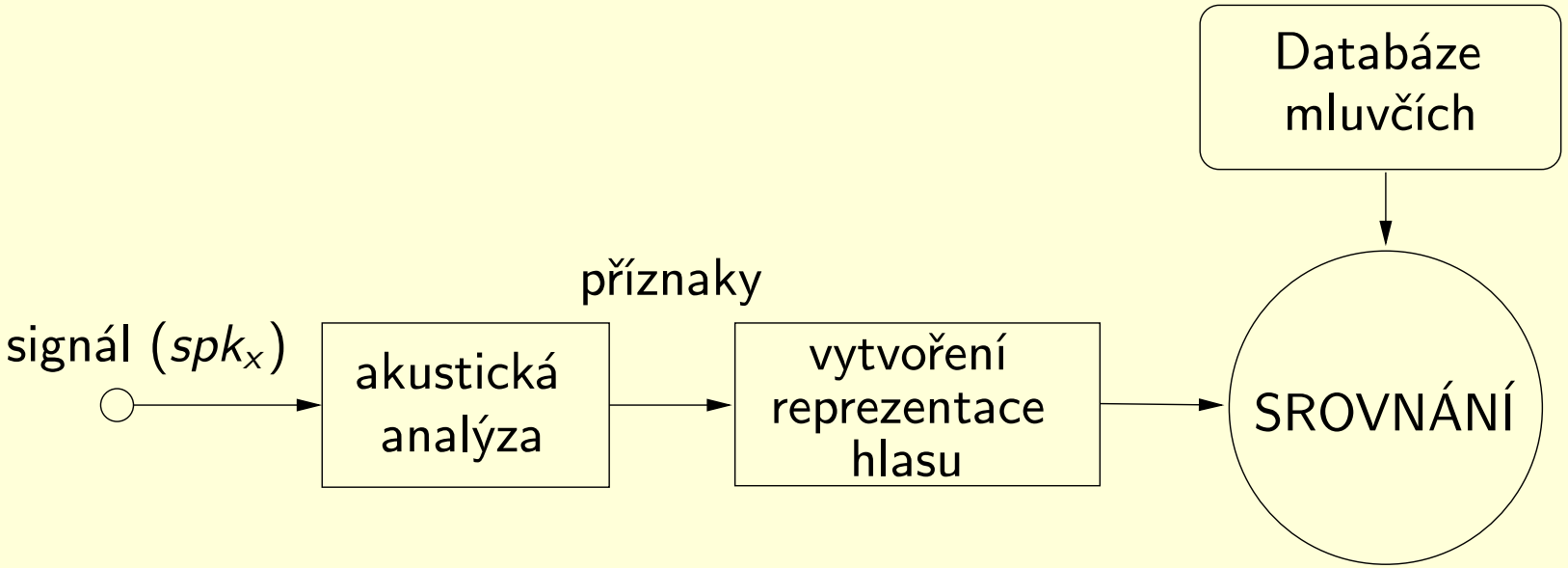
Motivace pro použití hlasové identifikace

- přirozenost komunikace, relativně jednodušší realizace
- jediná volba při dostupnosti pouze hlasového záznamu

Možnosti identifikace mluvčího

- historické klasické přístupy :
  - expertní rozhodování (fonetici, lingvisté)

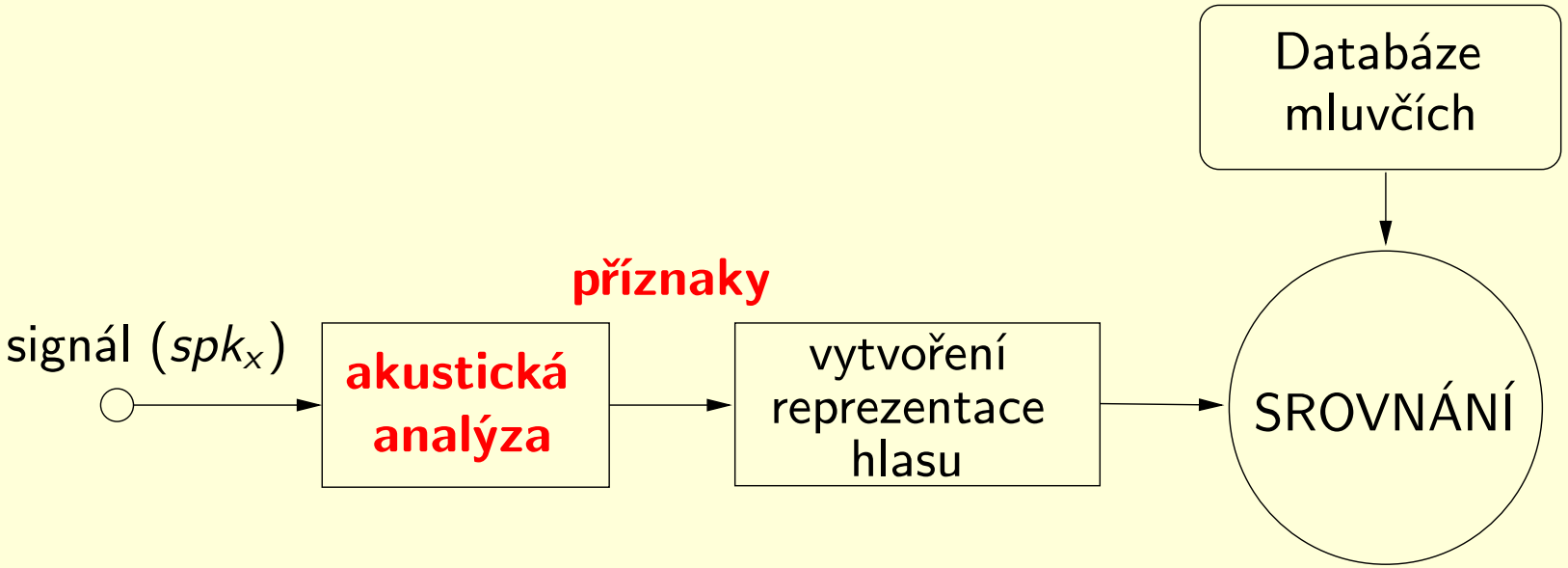
- moderní automatizované metody :
  - rozpoznávání mluvčího



Možnosti identifikace mluvčího

- historické klasické přístupy :
  - expertní rozhodování (fonetici, lingvisté)

- moderní automatizované metody :
  - rozpoznávání mluvčího



III. část

Řečové charakteristiky a možnosti využití pro identifikaci

Řečové příznaky pro hlasovou identifikaci

Obecné požadavky pro příznaky resp. systémy identifikace

- **vysoká variabilita pro různé mluvčí**
- **nízká variabilita pro jednoho mluvčího**
  (možné vlivy - aktuální stav, nálada, stres, hluk, styl promluvy)

- snadný a efektivní výpočet
- odolnost vůči šumu a zkreslení (výše zmiňované jevy)
- odolnost proti imitaci hlasu

$$\Downarrow$$

Vnitřní charakteristiky - související s vytvářením řeči

Získané charakteristiky - souvisejí s dynamikou pohybů hlasového traktu (dané prostředím)

Řečové příznaky pro hlasovou identifikaci - PROBLÉMY

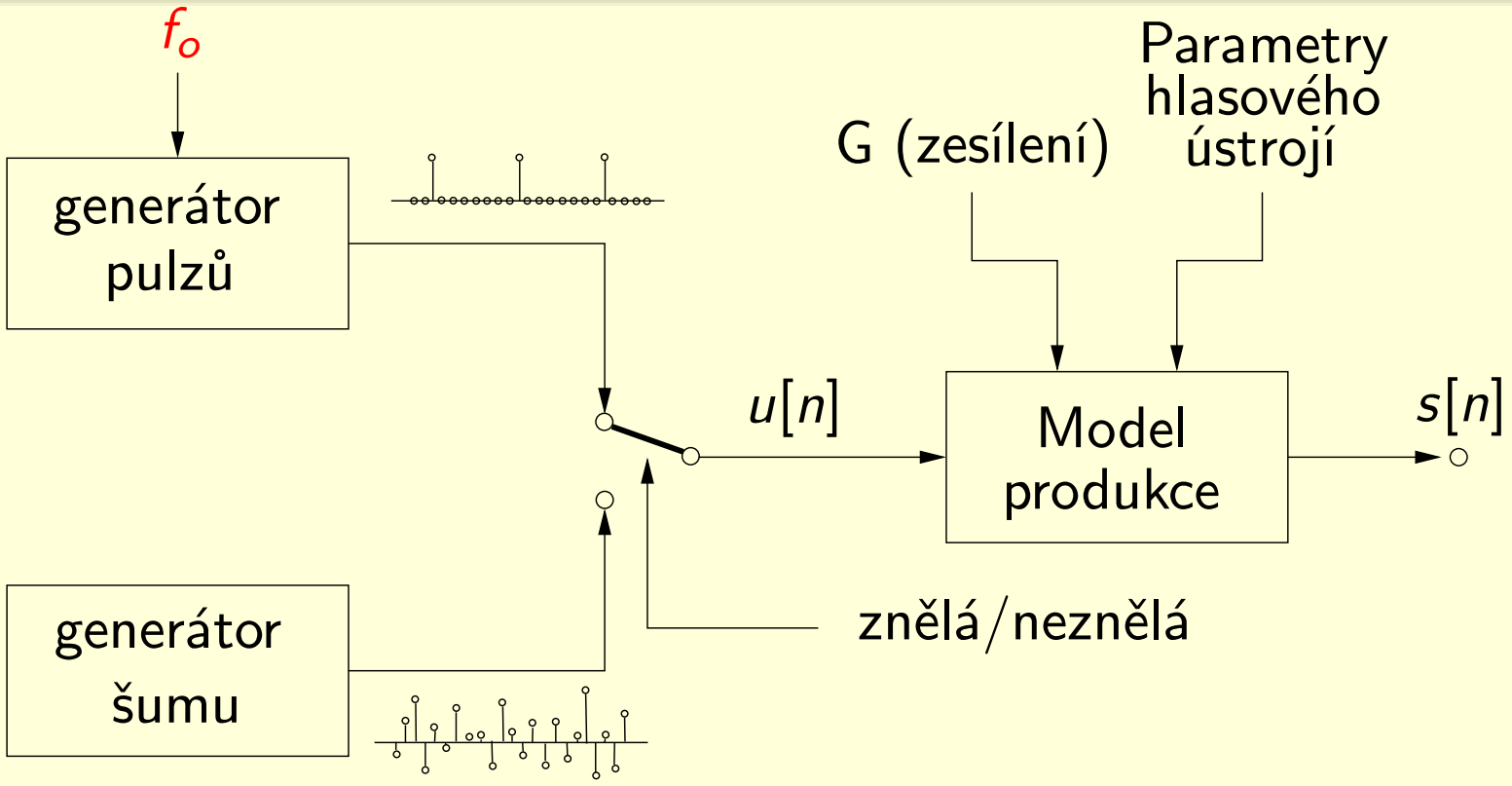
Vnitřní charakteristiky řečníka

- dané fyzikálními rozměry hlasového ústrojí
  (lze obtížně napodobit)
- ovlivnitelné zdravotním stavem
  (např. nosní dutina : neměnné rozměry při artikulaci,
  mírné nachlazení = zásadní změna)

Získané charakteristiky řečníka

- styl mluvy (časování, intonace, hrubost, živost, síla, srozumitelnost)
  $\rightarrow$ jako celek komplexní charakteristika řečníka
  (používáno člověkem při přirozené identifikaci)
- nemusí být snadno modelovatelné různými modely
- způsob řeči lze snadno napodobit

Základní tón řeči



- základní frekvence $f_o = 1/T_o$
- pro znělé hlásky s harmonickou strukturou
- souvisí s kmitáním hlasivek
- hodnota f_o je ovlivněna vlastnostmi hlasivek (pružnost, hmotnost, délka)
  $\rightarrow$ hrubá charakteristika mluvčího

Odhad základního tónu řeči

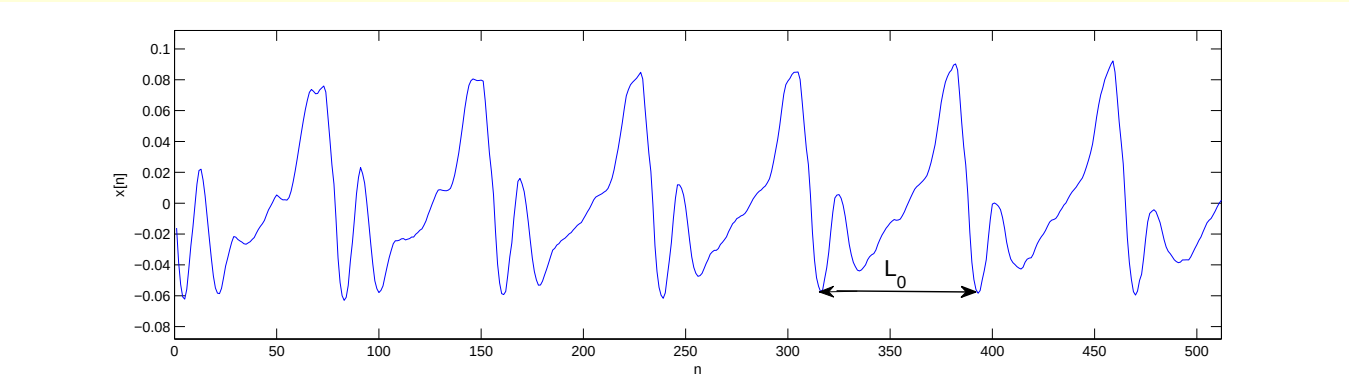
f_o ..... základní tón (frekvence) řeči ..... $f_o = \frac{1}{T_o}$

T_o (L_o) ..... základní perioda (v sekundách vs. ve vzorcích)

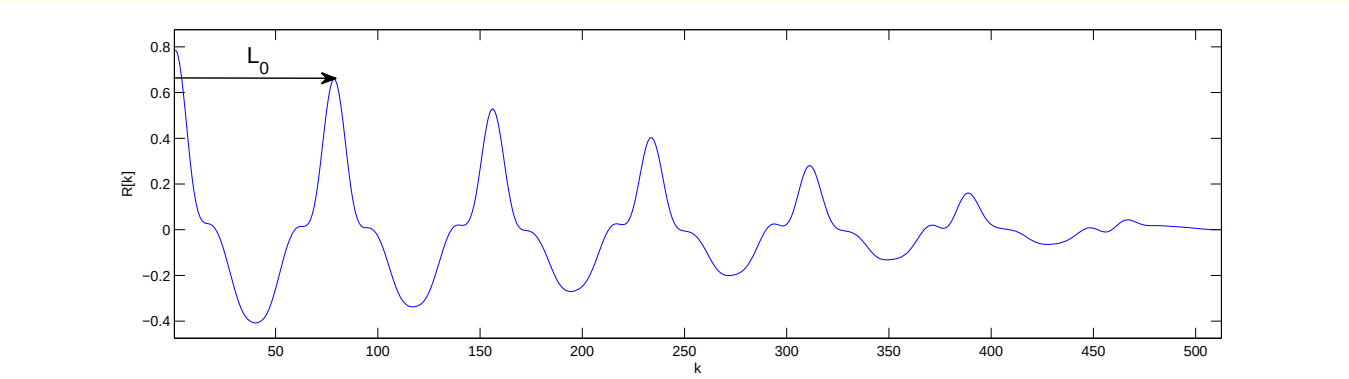
Nejčastější metoda odhadu - na bázi autokorelační funkce

(hledání postranního maxima autokorelační funkce)

segment signálu

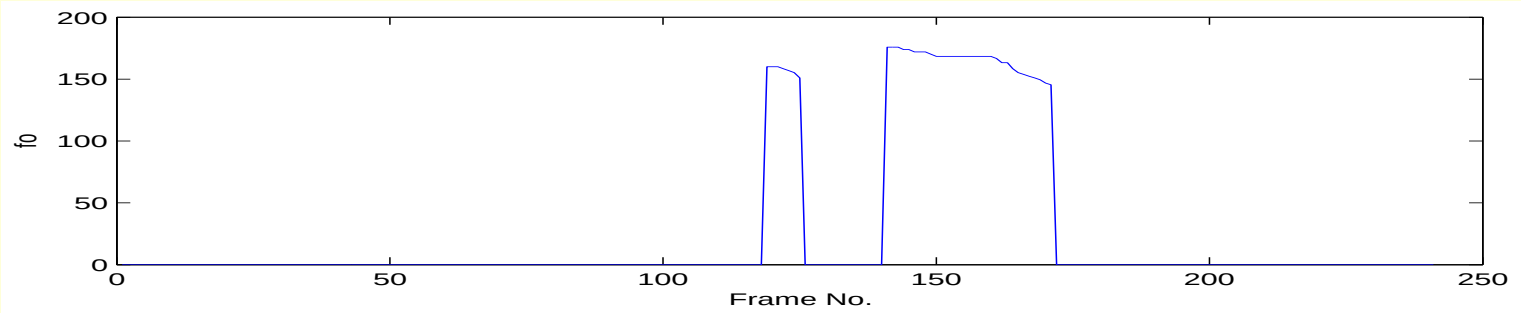


odhad autokorelační funkce

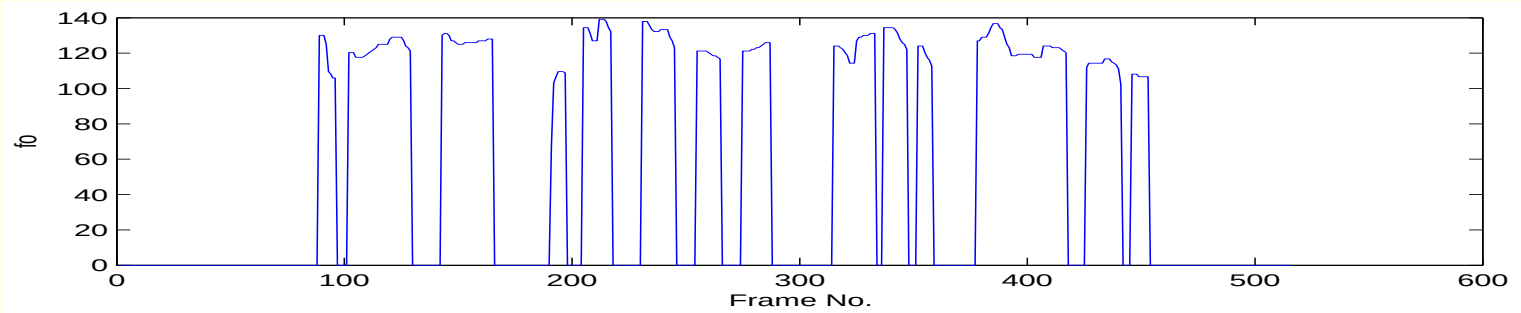


Průběh základního tónu v promluvě

Krátká promluva - slovo



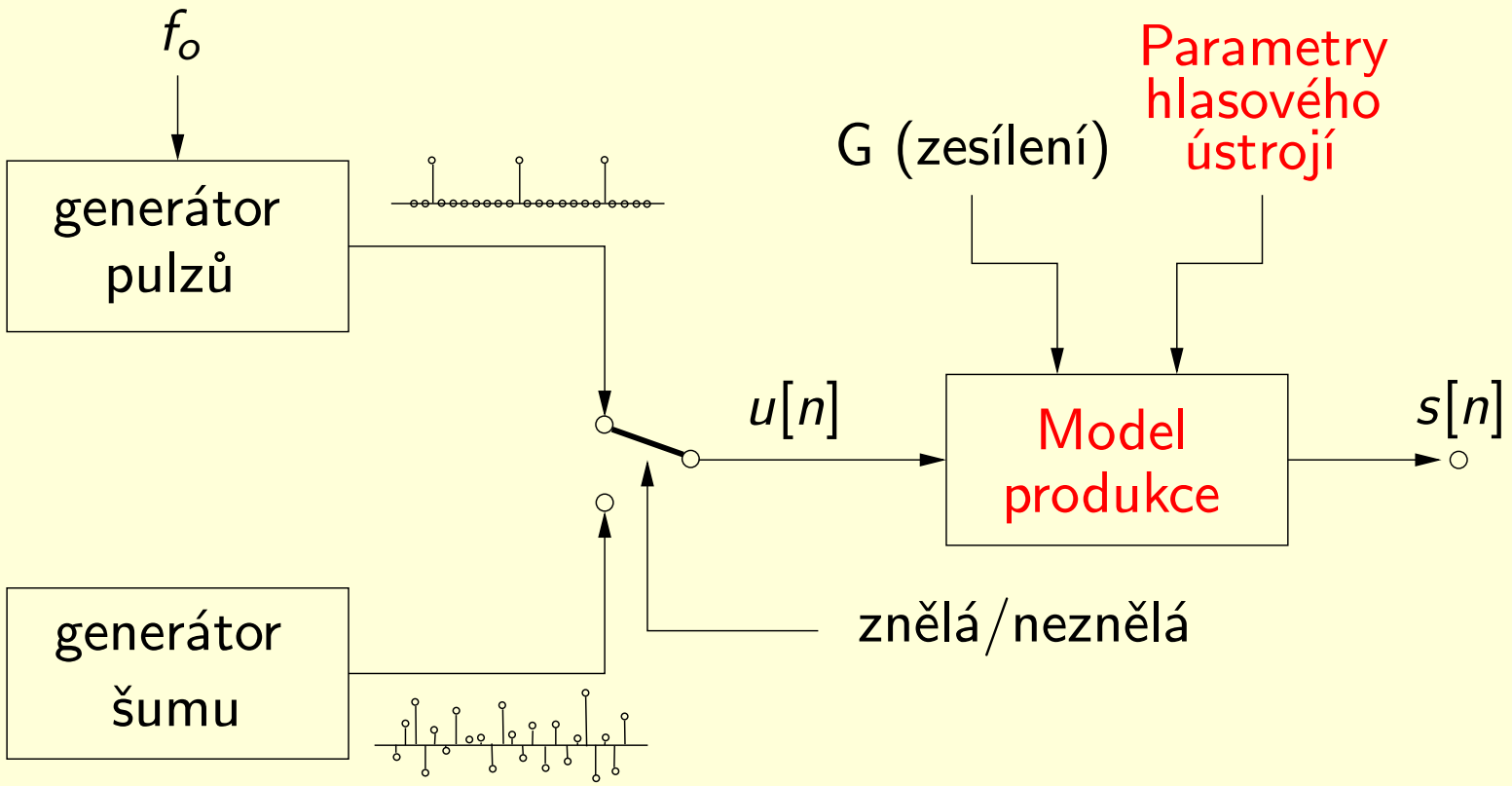
Delší promluva - věta



Průběh f_o v promluvě $\rightarrow$ získaná (naučená) charakteristika

Průměrná hodnota f_o $\rightarrow$ vnitřní charakteristika (výška hlasu)

Spektrální charakteristiky řeči



- spektrální charakteristiky souvisí s vokálním traktem
- otázka vhodné reprezentace pro identifikaci

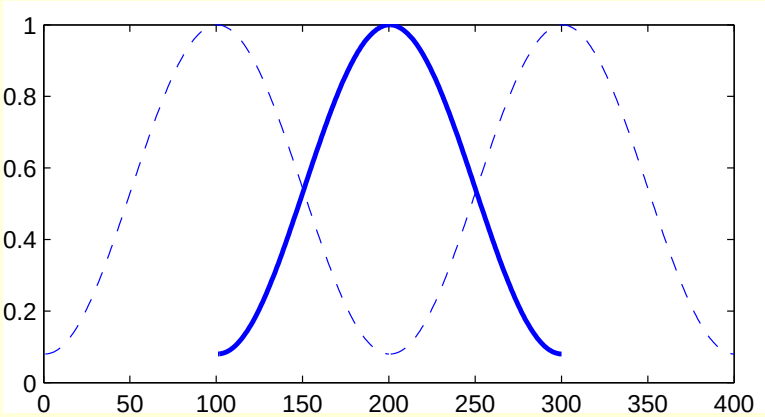
Spektrální reprezentace řeči na bázi DFT

Odhad spektra na bázi DFT:

- **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%)

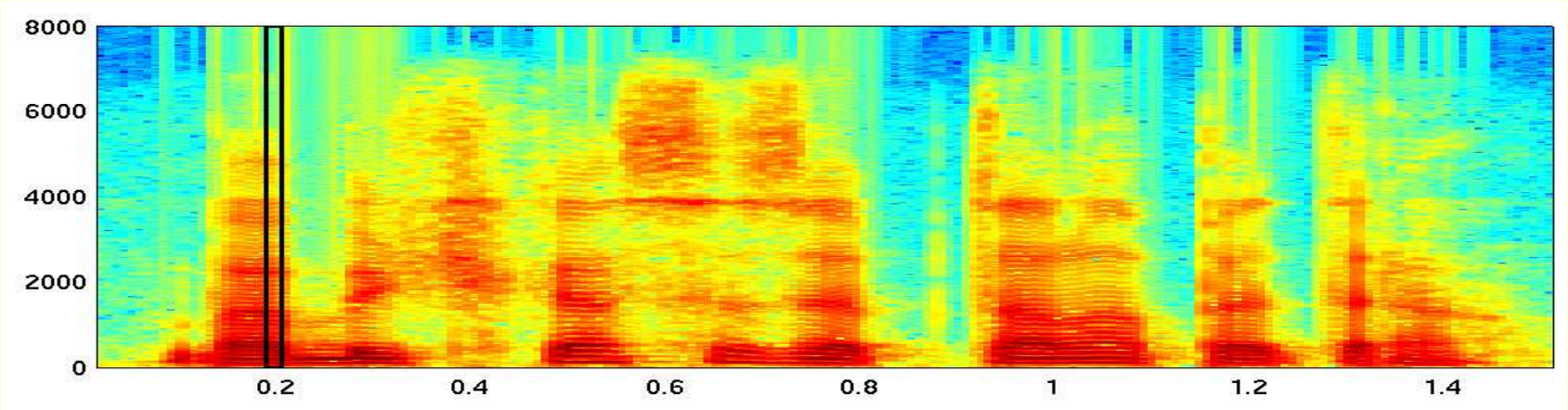
$$w[n] = 0,54 - 0,46 \cos \frac{2\pi n}{N}$$

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



Přehled možností spektrální reprezentace promluvy

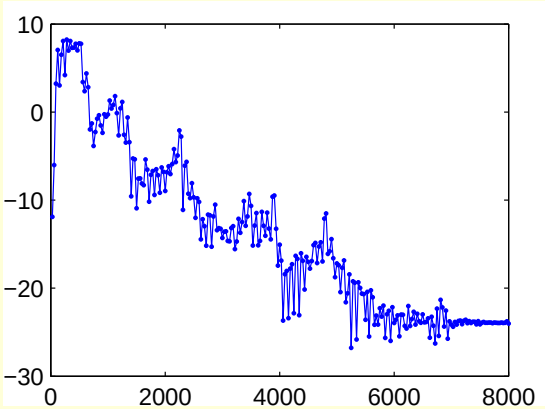
Spektrogram celé promluvy



Spektrální reprezentace vybraného segmentu

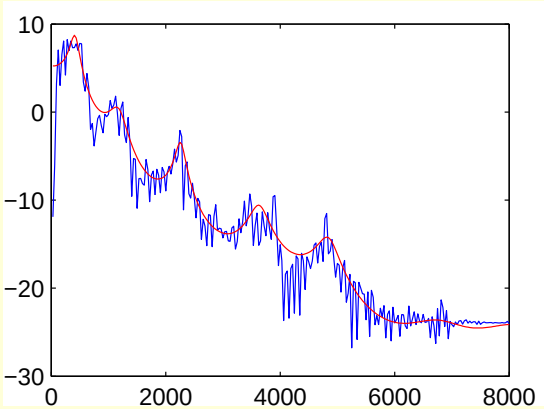
Sklon amplitudového spektra - vyšší kmitočty - nižší energie

DFT spektrum:



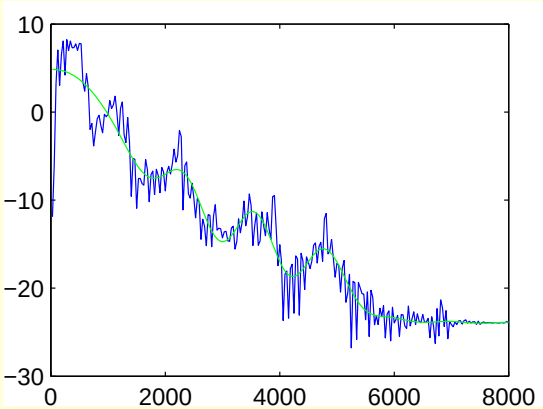
256 vzorků spektra
(amplitudové sp.)

LPC reprezentace:



16 koeficientů a_k
(autoregresní koef.)

Kepstrální koeficienty:



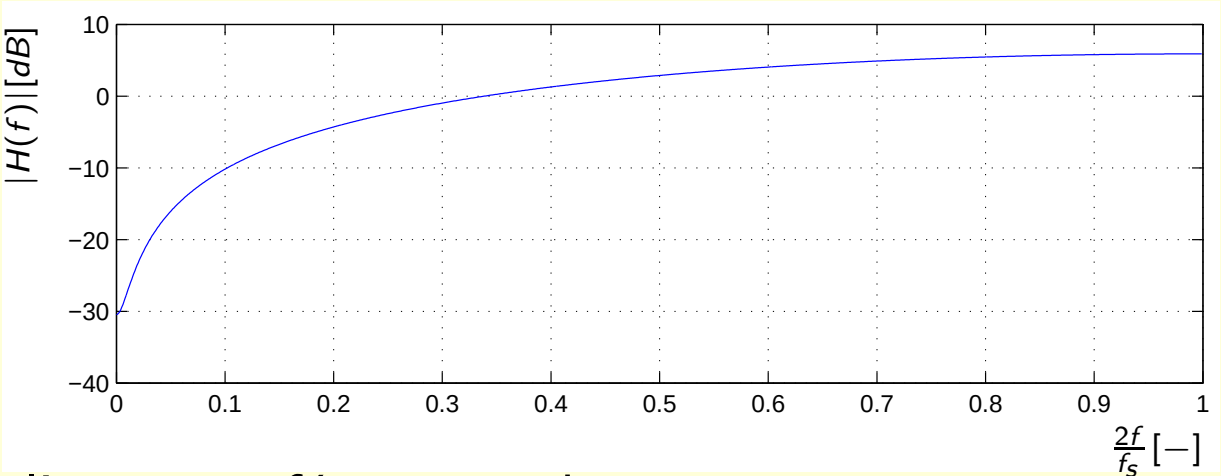
16 koeficientů c_n
(reálné kepstrum)

Preemfáze signálu

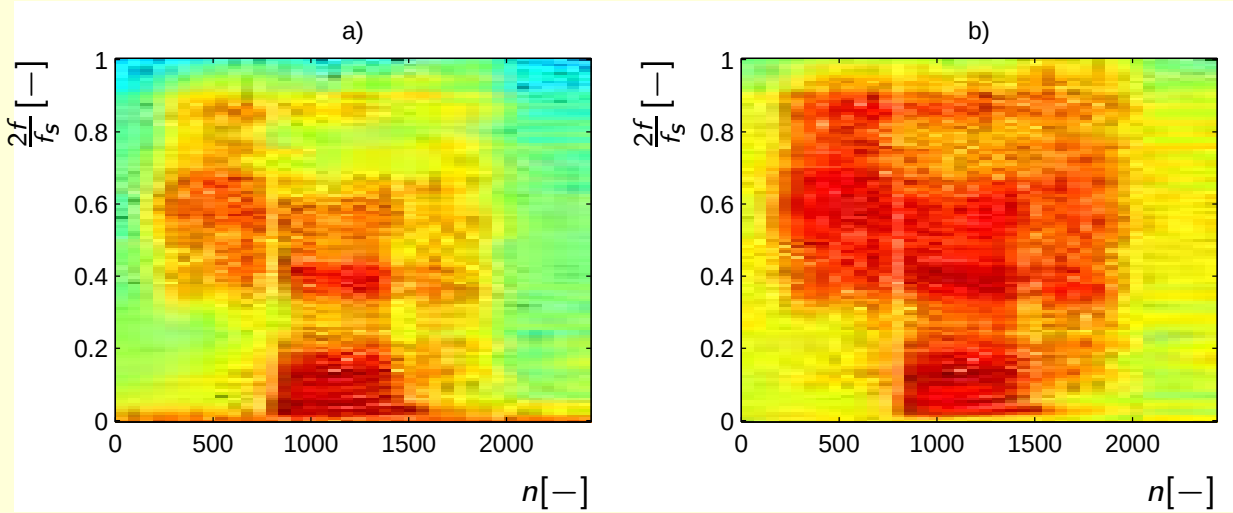
Kompenzace sklonu amlitudového spektra

$$s'[n] = s[n] - m \cdot s[n-1], \qquad m = 0,97$$

Frekvenční charakteristika preemfázového filtru

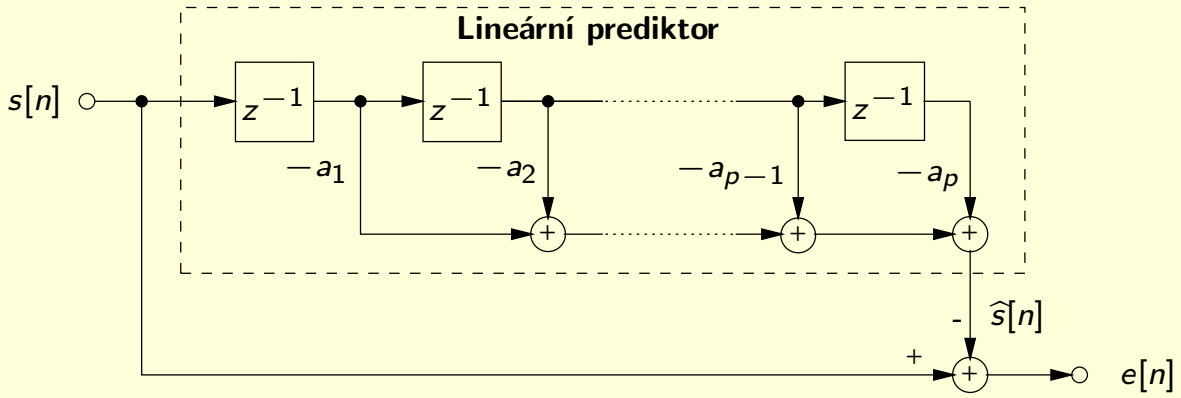


Ilustrace vlivu preemfáze ve spektrogramu



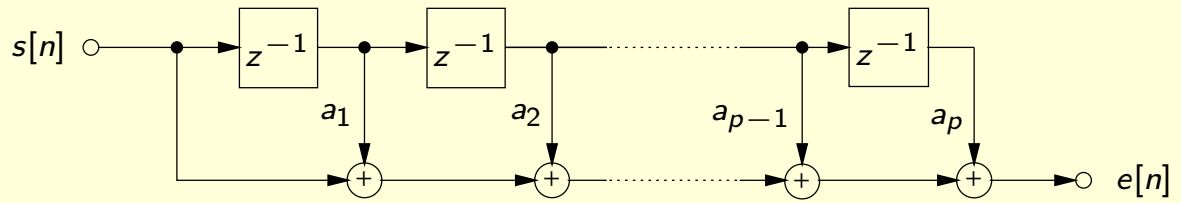
Lineární prediktivní analýza

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

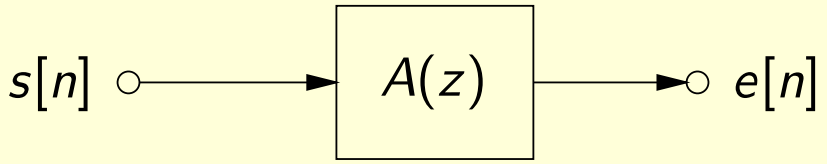
VÝSLEDEK:

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

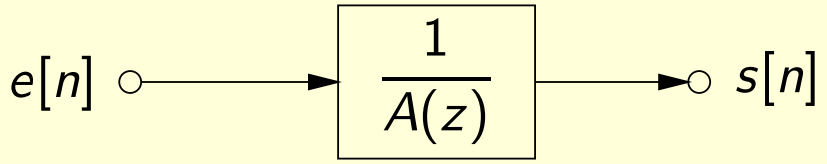
$P_p = R[0] + \sum_{k=1}^{p} a_k R[k]$ .... výkon chybového signálu

AR model signálu

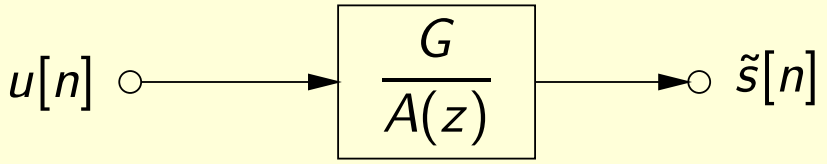
Dekorelační (analyzující) filtr : $A(z) = \sum_{k=0}^{p} a_k z^{-k}$



Syntéza se skutečným chybovým signálem (ideální případ)



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



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 $S_u(e^{j\Theta}) = 1$)

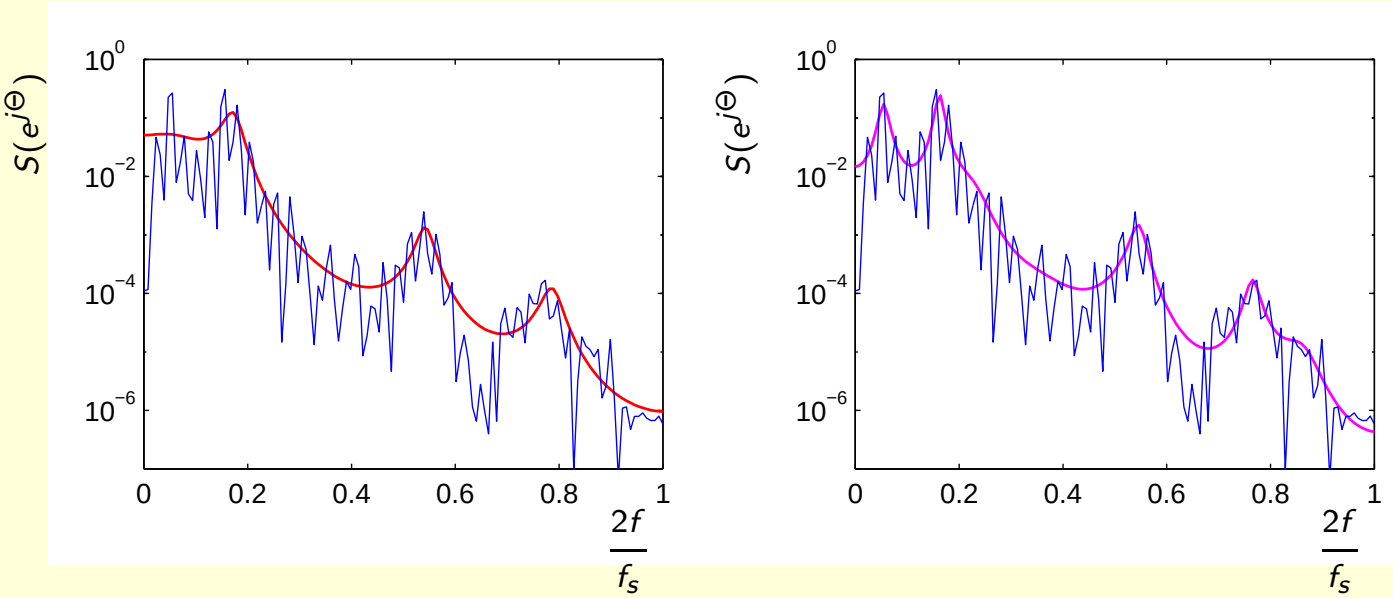
$$S_{\tilde{S}}(e^{j\Theta}) = |H(e^{j\Theta})|^2 = \frac{G^2}{|A(e^{j\Theta})|^2}$$

$$\Downarrow$$

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
- 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

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)

$\downarrow$

Souvislost s fyziologií vokálního traktu = vhodný vnitřní příznak
(formantové frekvence jsou nepřímo úměrné délce vok. traktu)

$$F_i = \frac{(2i-1)\cdot c}{4 \cdot VTL}$$

$\downarrow$

Pro rozlišení mluvčích - vzdálenost sousedních formantů

Formanty - Odhad na bázi LPC

- špičky LPC spektra - rezonátory = formanty
- F_i - formantová frekvence (centrální kmitočet rezonátoru)
- B_i - šířka pásma formantu
- špičky LPC spektra - určené **póly přenosové funkce** p_i

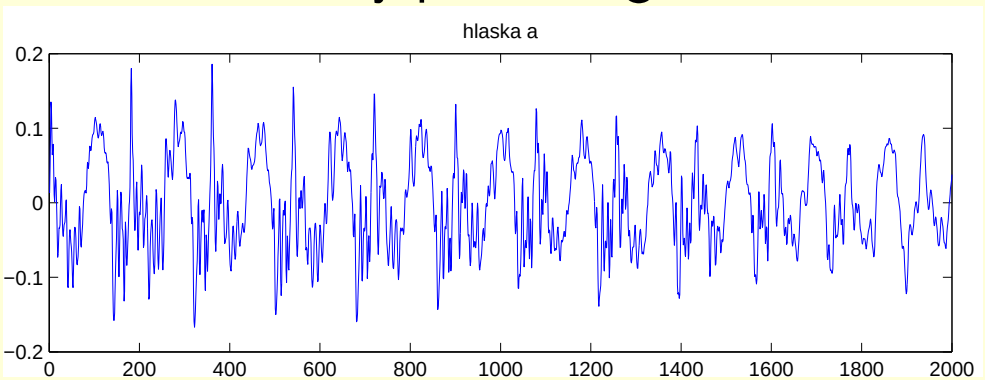
$$F_i = f_s \cdot \arg\, p_i / 2\pi$$

$$B_i = -f_s \cdot \ln |p_i| / 2\pi$$

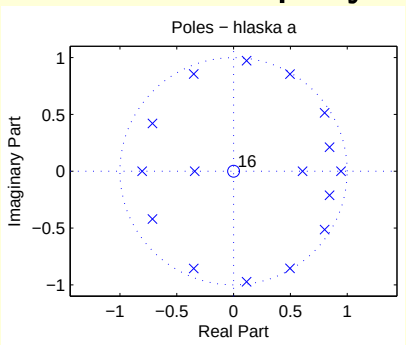
- Problémy:
  - obecně menší robustnost LPC analýzy (závislost na datech)
  - určení vhodného řádu (vliv přítomnosti šumu)
  - seřazení vypočítaných pólů (sledování stejného formantu)
  - vyřazení nadbytečného pólu (méně významné špičky)

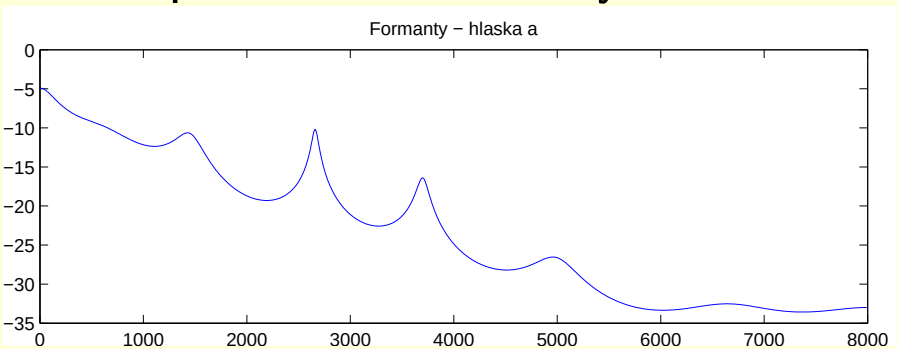
Odhad formantů na bázi LPC - příklad

Časový průběh signálu

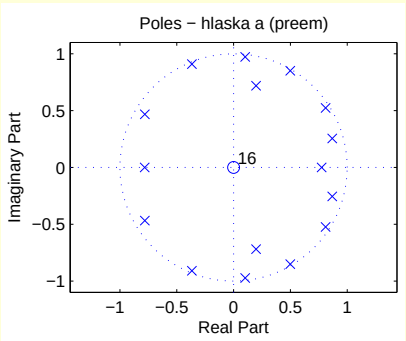


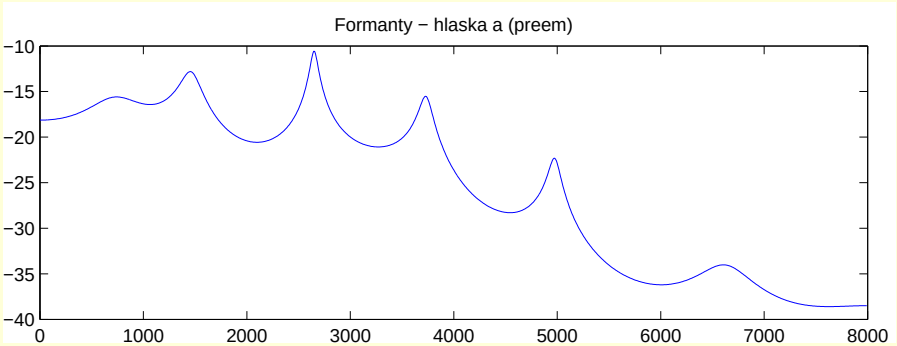
póly & LPC spektrum s formanty





póly & LPC spektrum s formanty (preemfáze)





Speciální příznaky pro rozpoznávání mluvčího

- F2 v “n”
- F3 v “u”
- F2 v “i”
- délka trvání “k”

- ..... *obecnější formulace* .....
- hodnota formantu ve vybrané hlásce
- šířka pásma vybraného formantu ve vybrané hlásce
- směrnice poklesu formantu ve vybrané hlásce
- Průběh F0 ve vybrané větě (slově)
- průměrná hodnota F0 ve větě (slově)
- ..... apod. .....

Kepstrum - definice a základní vlastnosti

Základní definice pomocí Z-tranformace

$$\hat{c}[n] = \mathcal{Z}^{-1}\{\ln \mathcal{Z}\{x[n]\}\}$$

Přímý výpočet pomocí DFT

$$c_k[n] = \mathcal{IDFT}\{\ln \mathcal{DFT}\{x[n]\}\} \quad \ldots \text{komplexní kepstrum}$$

$$c_r[n] = \mathcal{IDFT}\{\ln |\mathcal{DFT}\{x[n]\}|\} \quad \ldots \text{reálné kepstrum}$$

$$c_v[n] = \mathcal{IDFT}\{\ln |\frac{1}{N}\mathcal{DFT}\{x[n]\}|^2\} \quad \ldots \text{výkonové kepstrum}$$

Vlastnosti:

- $\hat{c}[n]$ .... nekonečně dlouhé, rychle ubývá k nule
- $c_k[n]$ .... konečně dlouhé, nesymetrické, informace o fázi
- $c_r[n]$ .... konečně dlouhé, symetrické, inf. o ampl. spektru
- $c_v[n]$ .... oproti $c_r[n]$ se liší pouze měřítkem a hodnotou $c[0]$
- ve všech případech vždy reálné hodnoty

Kepstrum - definice a základní názvosloví

Základní slovní přesmyčka: **spektrum** vs. **kepstrum**

Další vybrané přesmyčky:

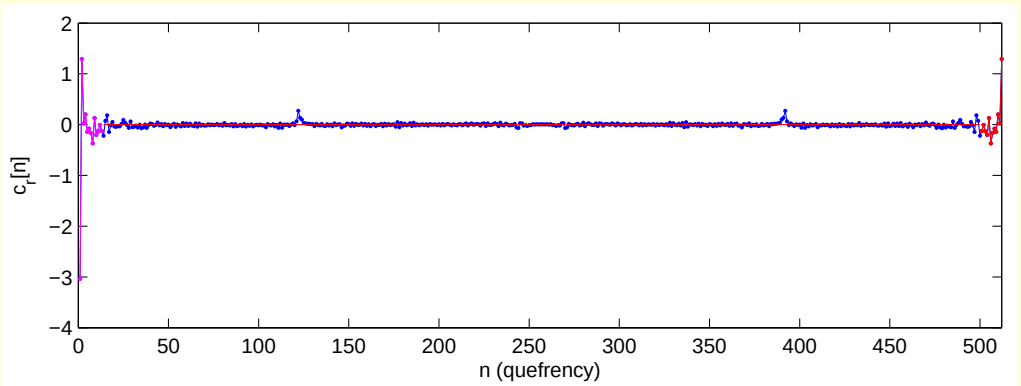
- kvefrence (frekvence) - základní proměnná kepstra [čas]
- liftr (filtr)
- liftrace (filtrace) = modifikace kepstra
  (váhování, oříznutí, zkracování)
- krátko-dobý liftr (dolno-frekvenční filtr)
- dlouho-dobý liftr (horno-frekvenční filtr)
- gamnituda (magnituda, amplituda)
- .....

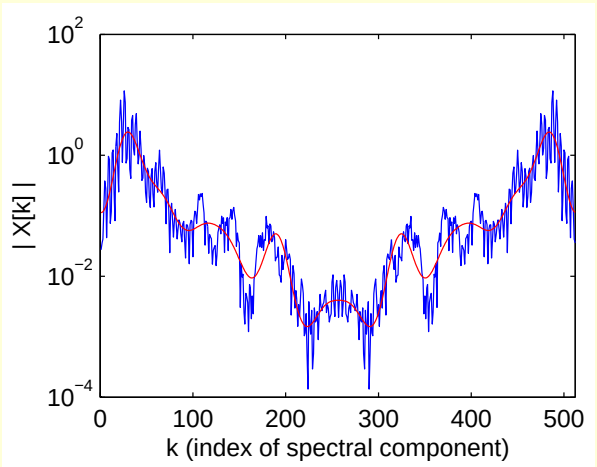
Vlastnosti reálného DFT kepstra

$$c_r[n] = \mathcal{IDFT}\{\ln|\mathcal{DFT}\{x[n]\}|\}$$

Vlastnosti:

- DFT kepstrum - numerický výpočet (period. a symetr.)
- První část - hlavní informace o tvaru amplitudového spektra
  tj. spektrum neperiodické složky signálu
  tj. spektrální obálka (vyhlazené spektrum)





Vyhlazené spektrum: $\overline{|X[k]|} = e^{DFT\{c_n \cdot w_n\}}$

LPC kepstrum signálu

Výchozí veličiny: parametry AR modelu - a_k, $G = \sqrt{E_p}$

$$c_0 = \ln G$$

$$c_n = -a_n - \frac{1}{n}\sum_{k=1}^{n-1}(n-k)a_k c_{n-k}, \text{ pro } n = 1, 2, ..., p,$$

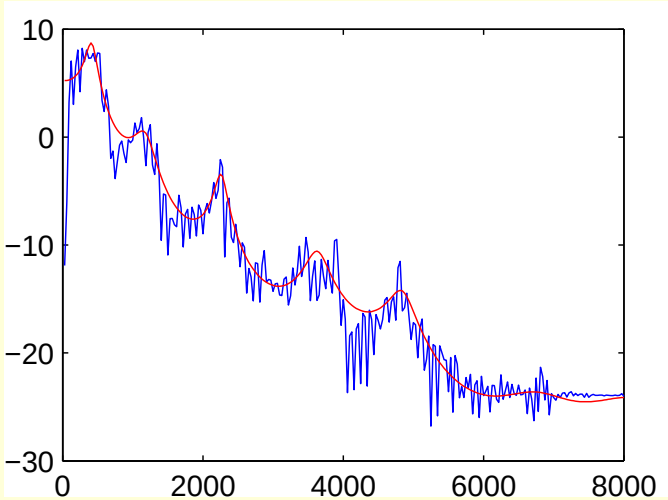
$$c_n = -\frac{1}{n}\sum_{k=1}^{p}(n-k)a_k c_{n-k}, \text{ pro } n = p+1, p+2, ...$$

Vlastnosti:

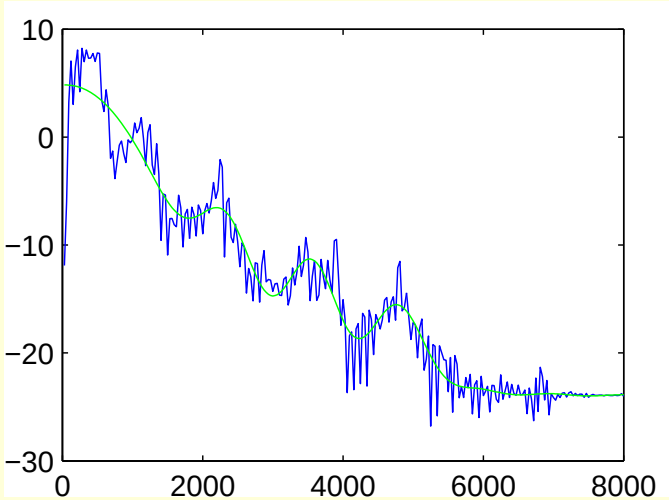
- Koeficienty Taylorova rozvoje $\ln|H(z)|$ (inverzní Z-transf.)
- Nekonečně dlouhé, první hodnoty opět nejvýznamnější
- Lze spočítat rekurentně, neobsahuje náhodnou složku
- Tvar spektra kopíruje LPC spektrum

Kepstrální analýza pro zpracování řeči

LPC spektrum:



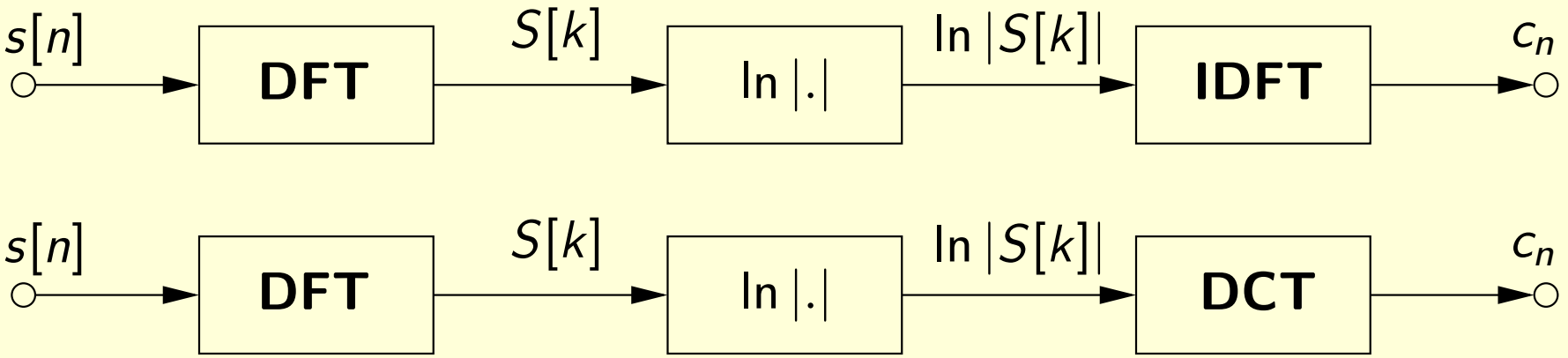
Vyhlazený odhad z reálného kepstra:



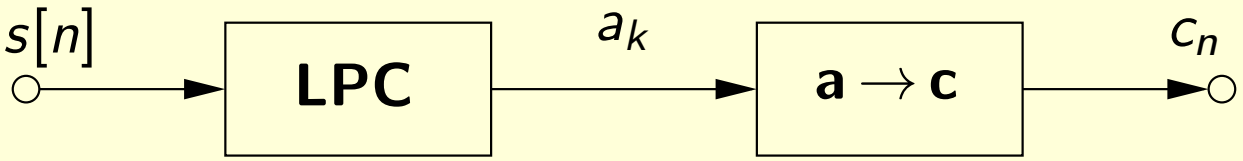
- První koeficienty nesou hlavní informaci o tvaru amplitudového spektra (12-20 kepstrálních koeficientů)
- **kepstra podobných segmentů tvoří shluky**
  $\Longrightarrow$ **použití jako příznaky pro rozpoznávání**
- **kepstrum nese informaci o vokálním traktu**
  $\Longrightarrow$ **použití pro automatickou hlasovou identifikaci**

DFT a LPC kepstrum - bloková schémata výpočtu

Blokové schéma výpočtu kepstrálních koeficientů pomocí DFT



Blokové schéma výpočtu LPC kepstrálních koeficientů



Mel-kepstrum - melodická frekvenční stupnice

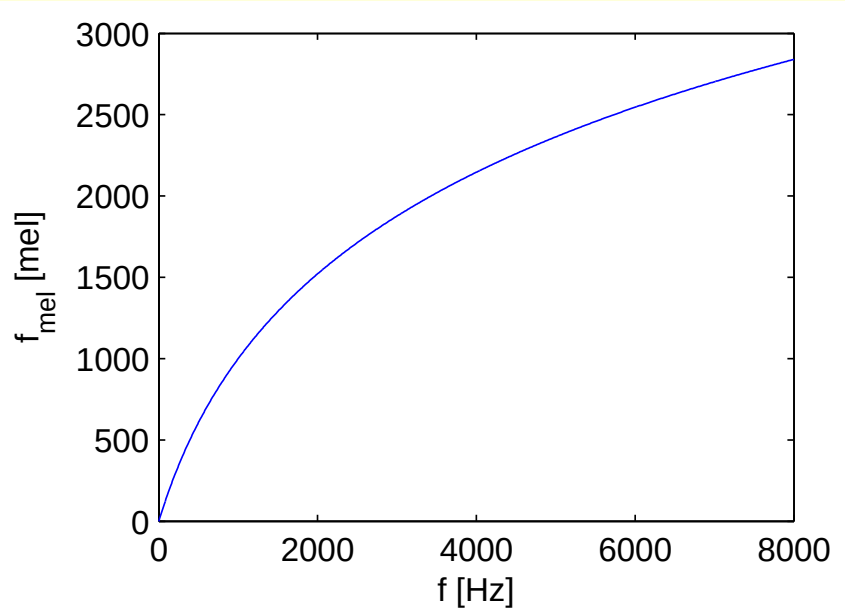
Vhodnější výpočet kepstra :

$\approx$ **modelování nelinearity vnímání frekvence lidským sluchem**

Nelineární zkreslení frekvenční osy - *melodická stupnice*

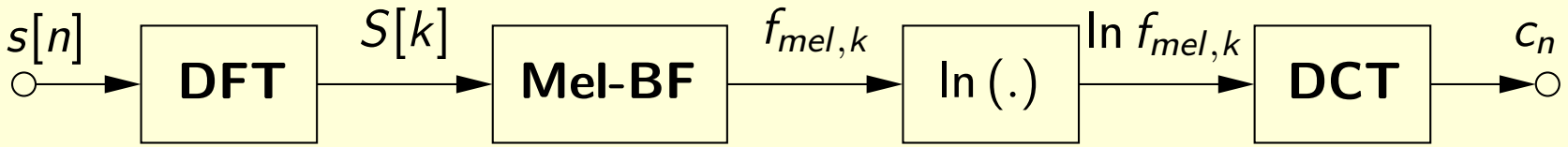
$$f_{mel} = \mathrm{Mel}\,(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

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



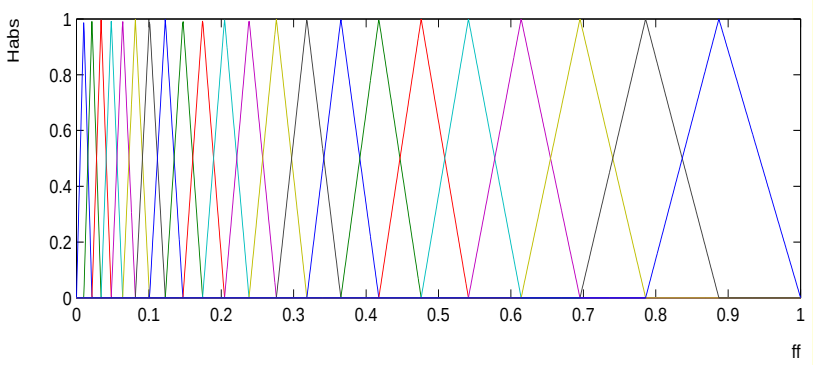
MFCC - Melovské kepstrální koeficienty

Blokové schéma výpočtu mel-kepstrálních koeficientů:



Výpočet energie v jednom pásmu

$$g_j = \ln \sum_{k=0}^{N/2} |S[k]|^2 H_{mel,j}[k] \,.$$



Výpočet kepstra pomocí DCT

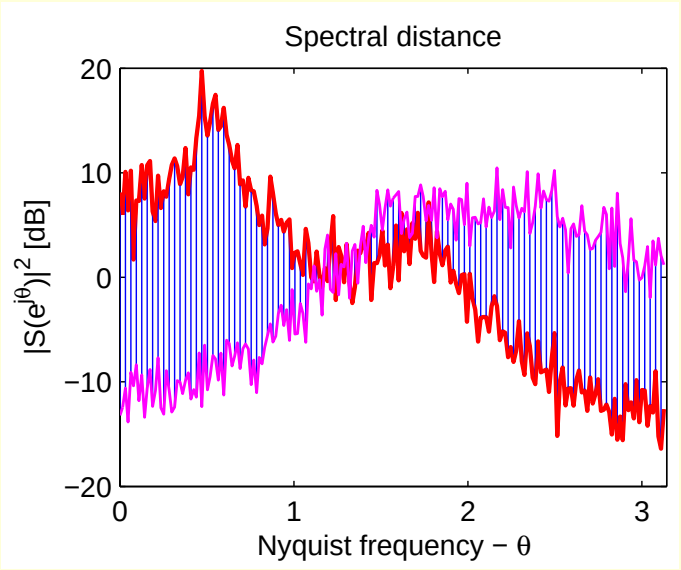
$$c_i = \sqrt{\frac{2}{P}} \sum_{j=1}^{P} g_j \cos\left(\frac{\pi i}{P}(j - 0.5)\right)$$

- **nejrozšířenější příznaky používané v ASR**
- nyní hodně používané i pro **hlasovou identifikaci** (v úlohách textově nezávislého rozpoznávání řečníka)

Spektrální a kepstrální vzdálenost

Spektrální vzdálenost (L_2-norma)

$$L_2 = \int_{-\pi}^{\pi} \ln \frac{|S_1(e^{j\theta})|^2}{|S_2(e^{j\theta})|^2} d\theta$$



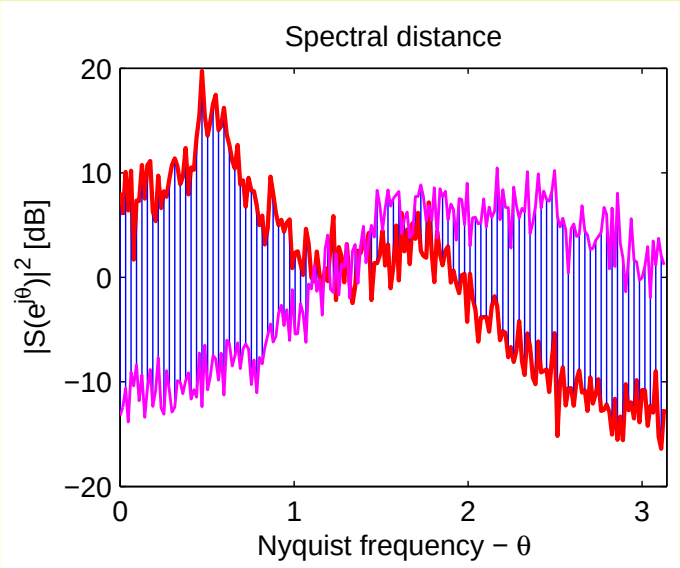
Spektrální vzdálenost na bázi L_2**-normy**

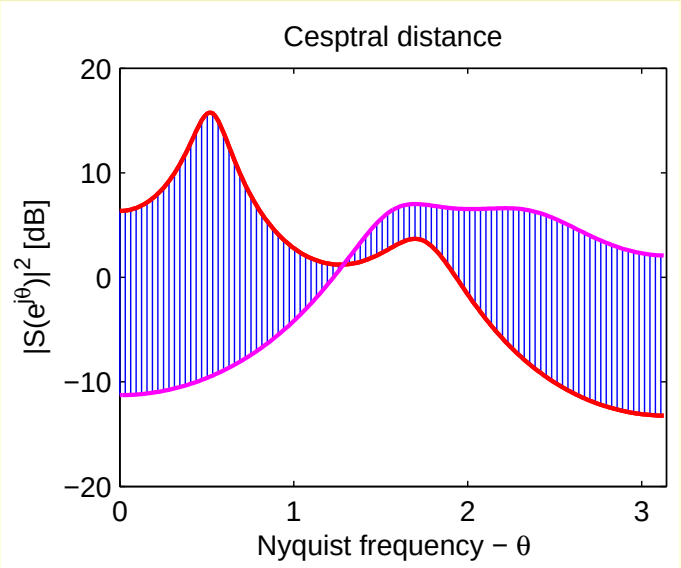
$\rightarrow$ kvantifikuje plochu ohraničnou dvěma spektry (křivkami)

Kepstrální vzdálenost

Kepstrální vzdálenost

$$\mathrm{CD} = \sqrt{(c_s[0] - c_x[0])^2 + 2\sum_{k=1}^{L}(c_s[k] - c_x[k])^2}$$





- CD aproximuje spektrální vzdálenost na bázi L_2-normy
- používá první kepstrální koeficienty
- vzdálenost je vypočítána ze spektrální obálky (tj. z vyhlazených spekter)

Různé definice kepstrální vzdálenosti

- Euklidovská vzdálenost: $\mathrm{CD} = \sqrt{\sum_{k=0}^{L}(c_s[k] - c_x[k])^2}$
- Euklidovská vzdálenost bez $c[0]$: $\mathrm{CD} = \sqrt{\sum_{k=1}^{L}(c_s[k] - c_x[k])^2}$
- kvadrát Euklidovské vzdálenosti bez $c[0]$: $\mathrm{CD} = \sum_{k=1}^{L}(c_s[k] - c_x[k])^2$
- vážená (liftrovaná) kepstrální vzdálenost: $\mathrm{CD} = \sum_{k=1}^{L}(L_k c_s[k] - L_k c_x[k])^2$

- Vždy kvantifikace rozdílů ve spektru
- Varianty - různá citlivost a různé měřítko

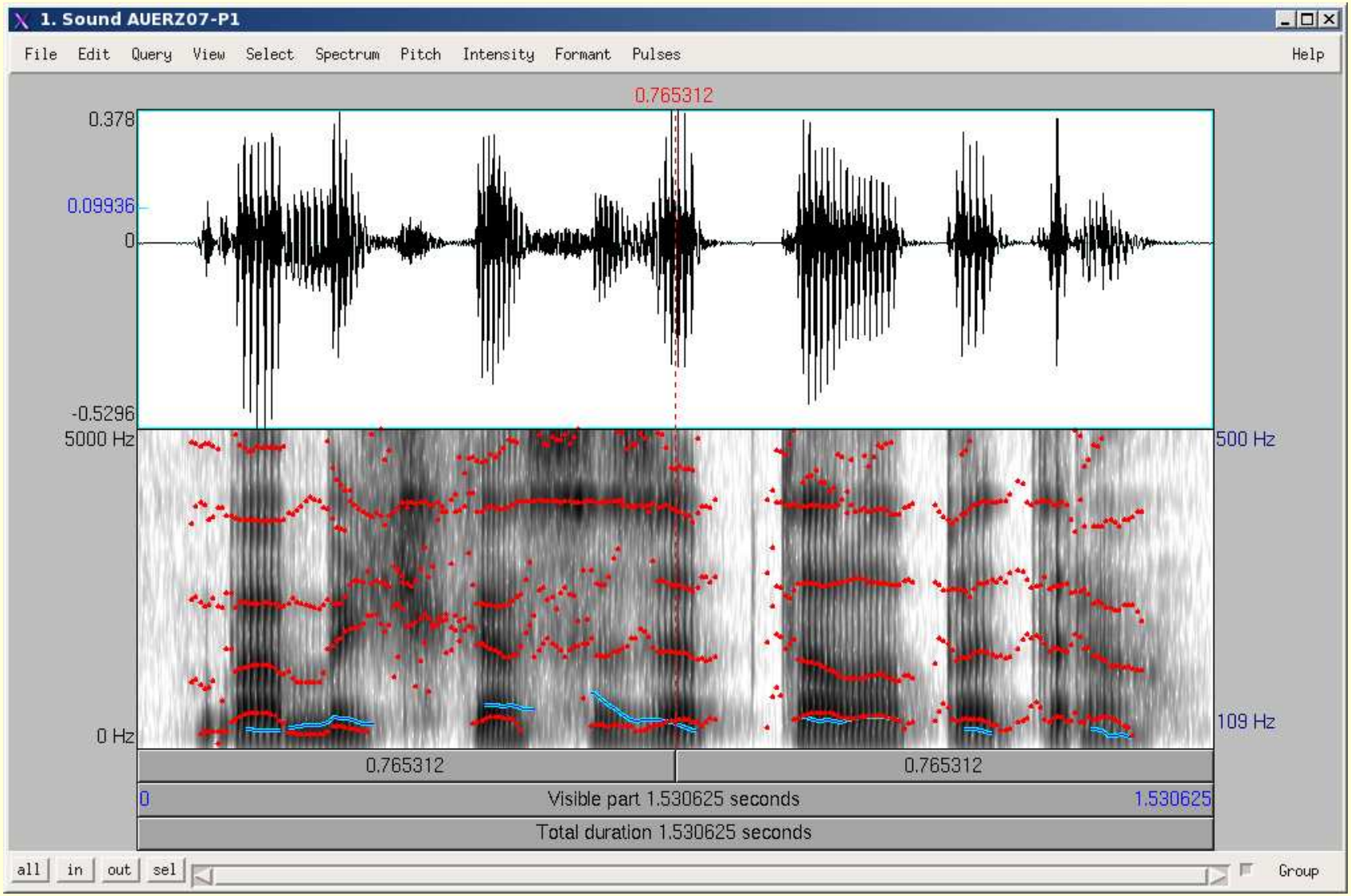
Expertní metody při rozpoznávání mluvčího

- **Forenzní lingvistika a fonetika**
  - sledování osobitých rysů projevu řečníka
  - zaměření na artikulační zvláštnosti
  - typické vedení melodie řeči (intonace)
  - většinou na bázi poslechu

- **Spektrografické metody**
  - Využívají možnost zobrazení diskutovaných hlasových charakteristik (spektrogramů, průběhu f_o, trajektorií formantů, apod.)
  - řešeno opět na expertní bázi
  - detaily realizace vybraných hlásek

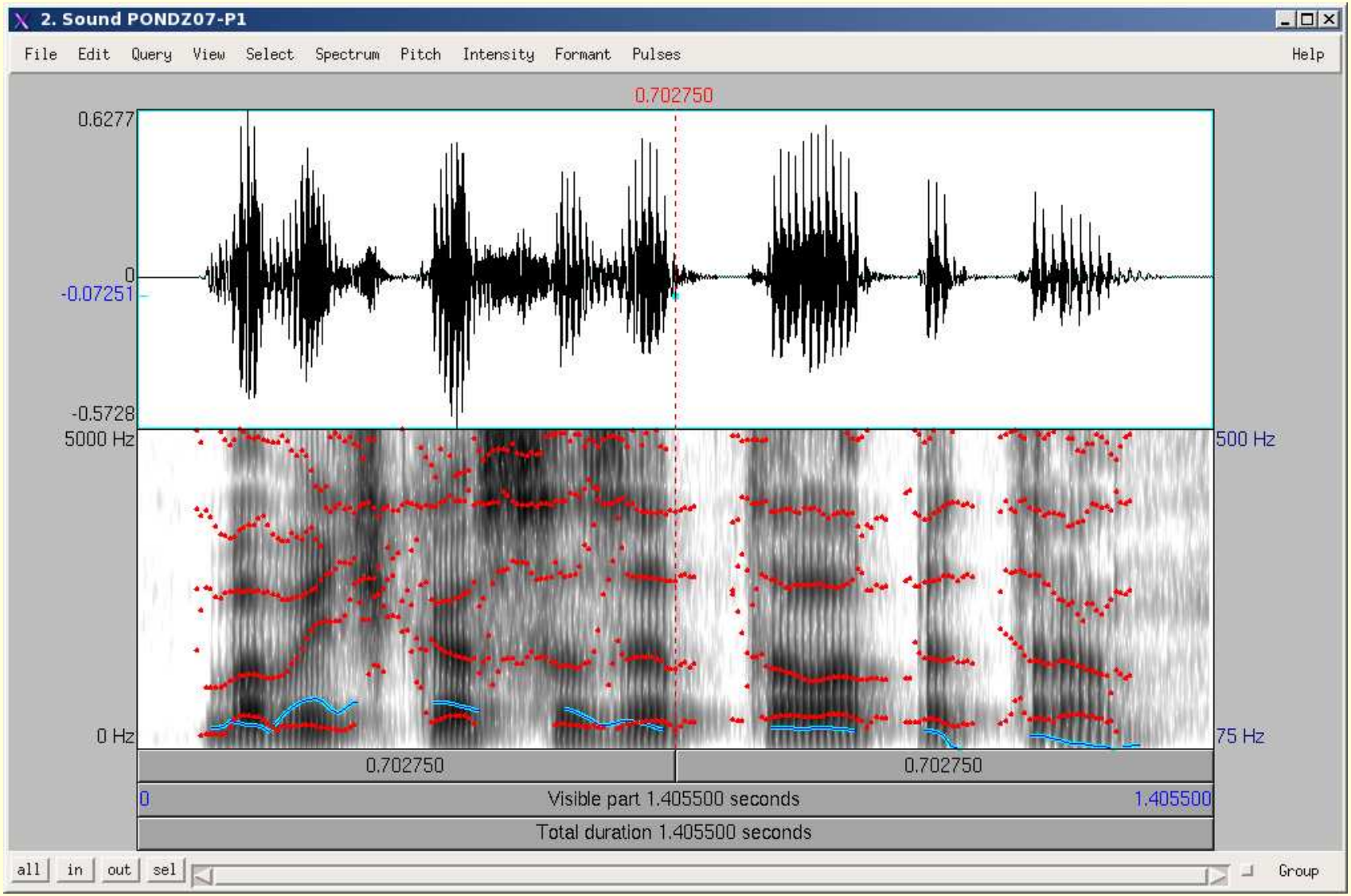
Formanty & základní tón - odhad Praat

Muž 1



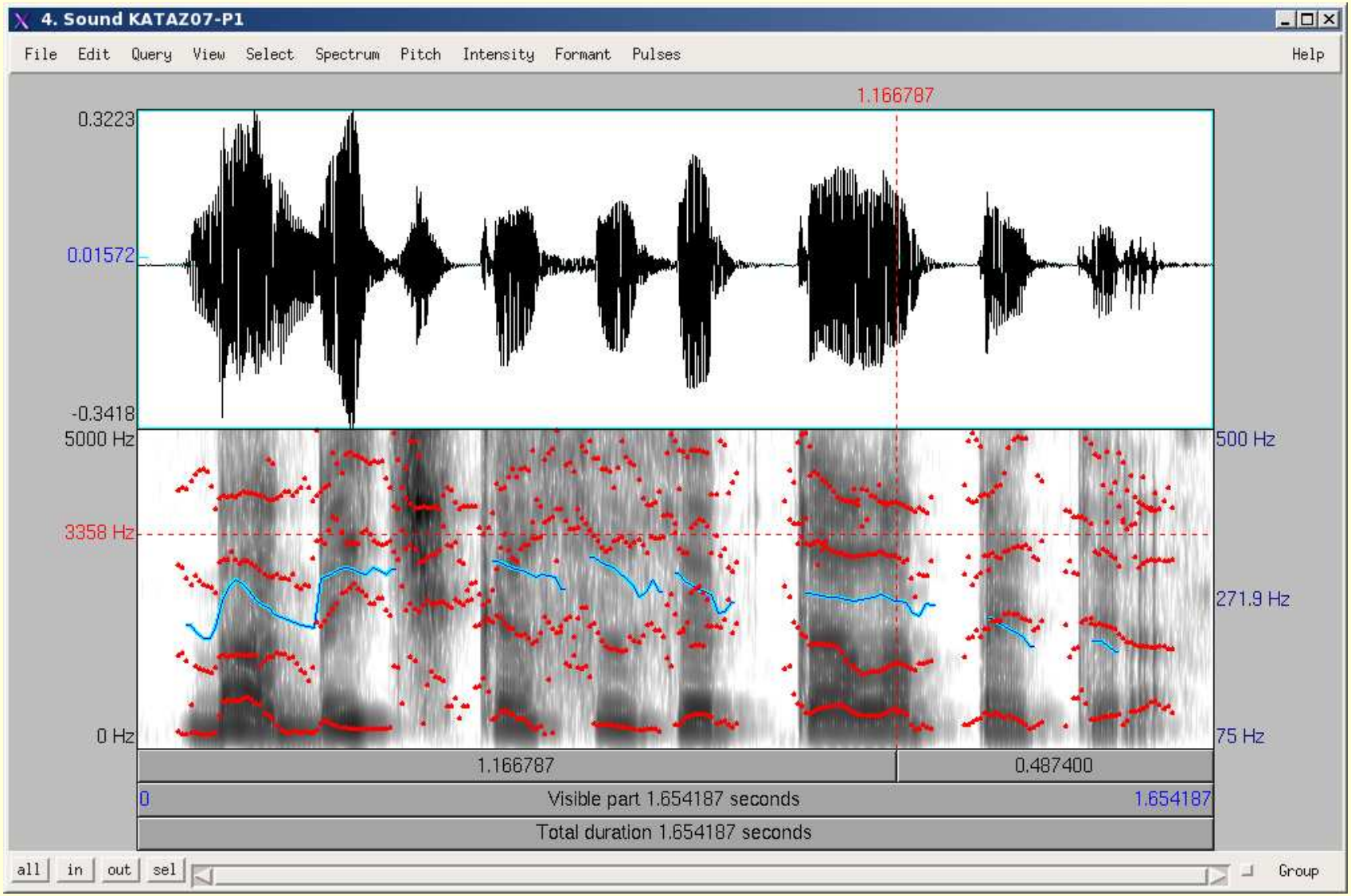
Formanty & základní tón - odhad Praat

Muž 2



Formanty & základní tón - odhad Praat

Žena 1



Děkuji vám za pozornost !