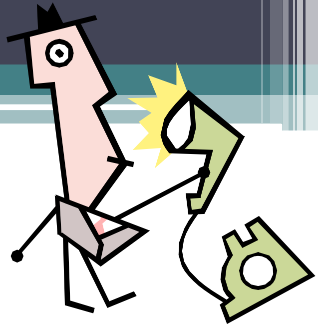
stanislav ondáš

Spektrálne charakteristiky signálu

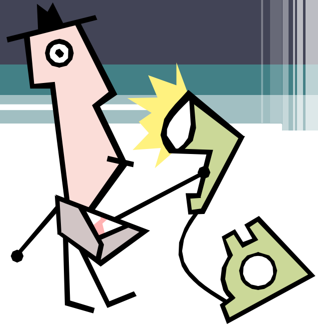
Euklidovská vzdialenosť

Podľa J. Uhlíř a kol.: Technologie hlasových komunikací



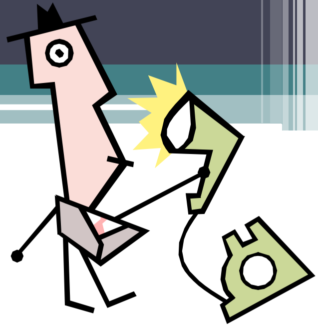
Obsah

- Spektrálna analýza na báze DFT
- LPC spektrálna analýza
- Dlhodobá spektrálna analýza
- Kepstrálna analýza
- Koherenčné charakteristiky viackanálových signálov
- Momentové charakteristiky



Rozdelenie

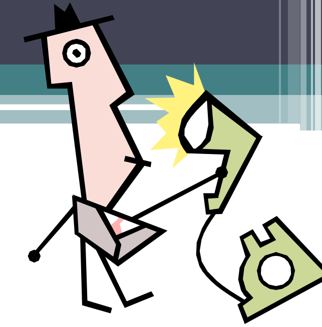
- Metódy spektrálnej analýzy pre spracovanie reči sa zvyčajne delia podľa dvoch hľadísk:
- Podľa metódy spracovania
  - Klasické metódy používajúce DFT
  - Parametrické metódy
    - Pre reč: lineárna predikčná analýza - LP
- Podľa dĺžky spracovávaného signálu
  - Krátkodobá analýza
  - Dlhodobá analýza



Spektrálna analýza na báze **DFT**

- definované *Fourierovou transformáciou*
- Spektrum diskrétneho signálu je *spojité* a *periodické* s periódou ekvivalentnou F_{vz}.
- DFT spektrum spočítané z N vzoriek signálu potom obsahuje N vzoriek jednej periódy spojitého diskrétneho signálu (ak neurčíme ináč).
- Pri analýze reči je významné hlavne amplitúdové spektrum |X[k]|.
- Dôležitá charakteristika – *výkonová spektrálna hustota*.

Spektrálna analýza na báze **LPC**

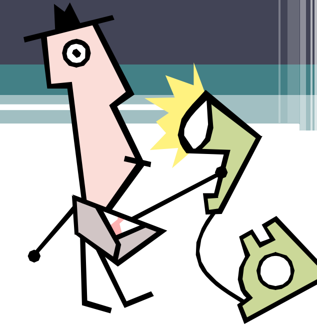


- je štandardným nástrojom analýzy rečového signálu.
- Predpokladajme, že reč je modelovaná *autoregresným (AR) modelom*,
  - kde budúcu vzorku je možné predikovať lineárnou kombináciou *p* minulých vzoriek (teda lineárnym prediktorom *p*-tého rádu).
  - Cieľom analýzy je **určiť koeficienty tejto lineárnej kombinácie** tak, aby energia chyby predikcie bola minimálna:

(1)
$$e[n] = x[n] - \tilde{x}[n] = x[n] + \sum_{k=1}^{p} a_k x[n-k]$$

- **Parametrický popis analyzovaného signálu potom predstavujú koeficienty a_k nazývané *autoregresné koeficienty signálu*.**
- Je možné ukázať, že minimalizácia chyby predikcie vedie k dekorelovanosti vzoriek chybového signálu, ktorý má potom charakter bieleho šumu.
- Rovnica 1 je diferenčnou rovnicou FIR filtra s prenosovou funkciou A(z).

LPC spektrálna analýza



- Ak má chybový signál charakter bieleho šumu, znamená to, že spektrálne vlastnosti signálu x[n] sú určené frekvenčnou charakteristikou filtra → **syntetizujúci filter AR modelu**
  - Jeho prenosová funkcia je inverzná k prenosovej funkcii A(z).
  - Spektrum analyzovaného signálu x[n] je modelované frekv. charakteristikou prenosovej funkcie:

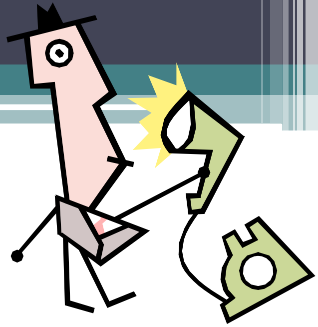
$$H(z) = \frac{G}{A(z)} = \frac{G}{1 + \sum_{k=1}^{p} a_k . z^{-k}}$$

- **LPC spektrum** (=vyhladený odhad spektrálnej výkonovej hustoty) je definované:

$$S_x(e^{j\theta}) = \left|H(e^{j\theta})\right|^2 = \frac{G^2}{\left|A(e^{j\theta})\right|^2}$$

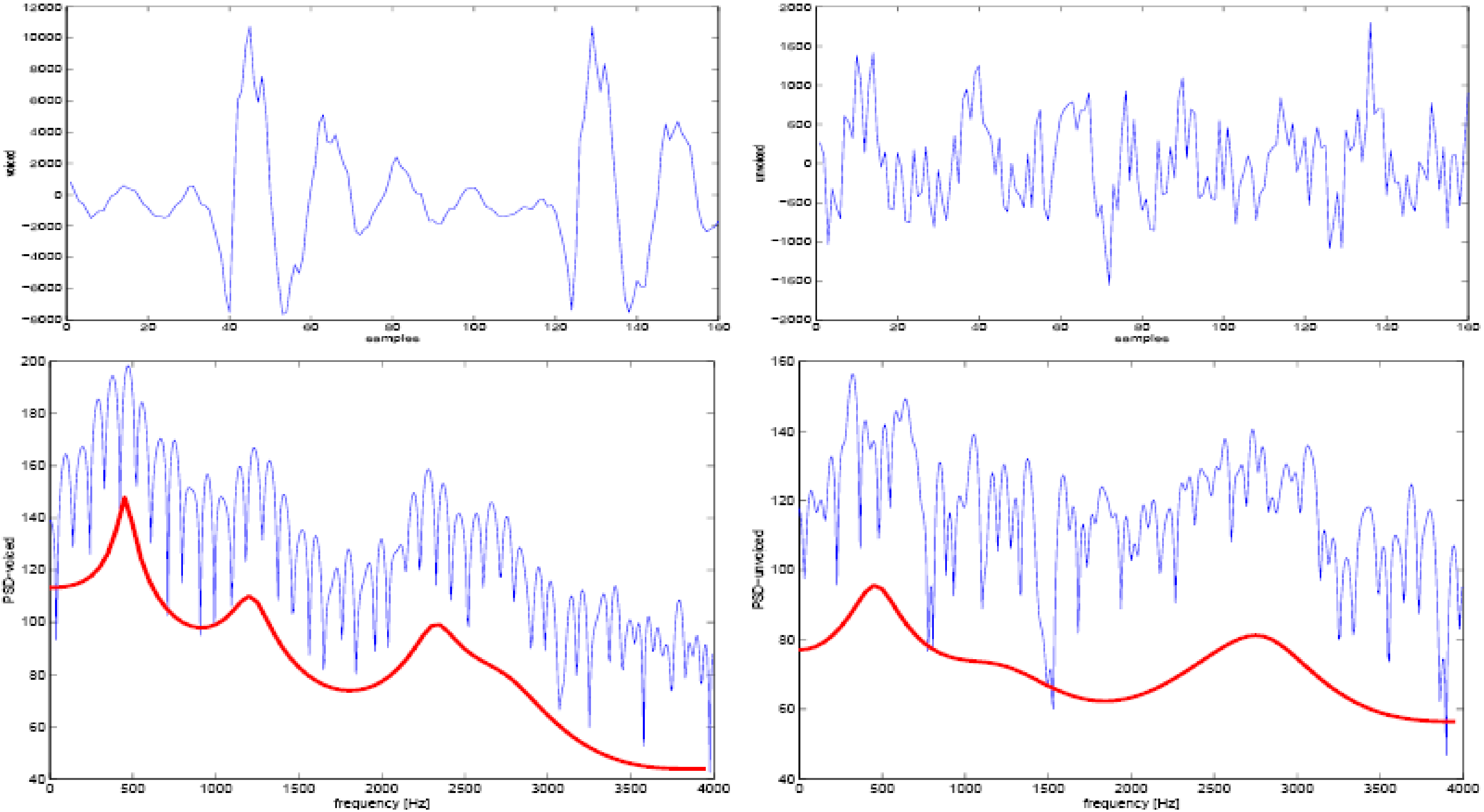
Diskrétne LPC spektrum S_x[k] s N vzorkami dostaneme dosadaním za $\theta = k.f_s / N$

LPC spektrálna analýza

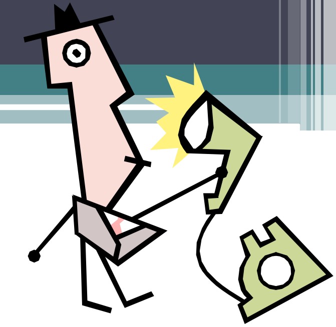


- Jadro analýzy na báze LP spočíva vo výpočte parametrov AR modelu a_k a zisku G pomocou autokorelačnej matice signálu => Levinson-Durbin.
- Rád modelu sa volí podľa počtu formantov, ktoré chceme pozorovať. Typické hodnoty sú p=12 (maximálne 6 formantov, pri f_{vz} = 8kHz).
- LPC spektrum predstavuje vyhladený odhad spektra analyzovaného signálu.
- Oproti DFT spektru je LPC spektrum vždy hladké, lebo:
  - Nie je v ňom obsiahnutá informácia o periodicite! Táto informácia je v chybovom signáli, ktorého tvar charakterizuje typ budenia.

Príklad: odhad spektr. hustoty výkonu pomocou DFT a LPC na znelom a neznelom rámci



Dlhodobá spektrálna analýza

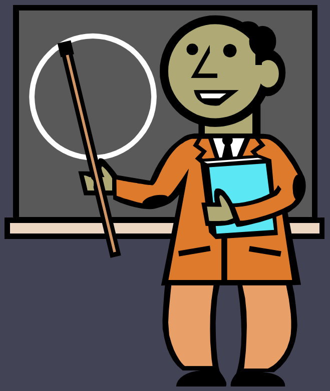


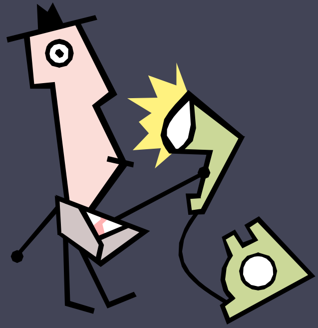
- Dlhodobá analýza – analýza takého úseku signálu, kde sa už prejavuje nestacionarita reči
  - Periodogram alebo spektrálna hustota výkonu (získaný Welchovou metódou)

- Dlhodobá analýza je nevhodná pre analýzu spektrálnych charakteristík reči!

- Ale je možné ju použiť pre získanie kvalitatívnej informácie o rozdiele spektra reči a šumového pozadia.

- Pre sledovanie vývoja reči sa používa prevažne krátkodobé amplitúdové spektrum usporiadané do ***spektrogramu***.

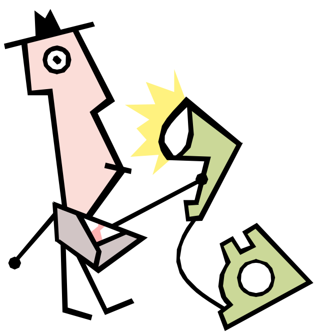




Kepstrálna analýza

Problém oddelenia budenia a vplyvu artik. traktu

- pre oddelenie budenia a vplyvu artikulačného traktu (modifikácie) – v kódovaní sa s nimi pracuje lepšie oddelene, pri rozpoznávaní reči, budenie zahadzujeme úplne (budenie je príliš závislé na rečníkovi, nálade a pod.)
- **1. možnosť:** odfiltrujeme frekvencie pod 400Hz (max. hlasivkova frekv. je cca 400Hz) a zbavime sa základného tónu ...
  - nebude to fungovať, lebo:
    - násobky základného tónu sú „rozsiate“ po celom spektre.
    - mohli by sme prísť o prvý formant
    - telefónne pásmo začína na hodnote 300Hz.
    - odstránenie informácie o výške hlasu hovoriaceho
  - potrebujeme lepší spôsob



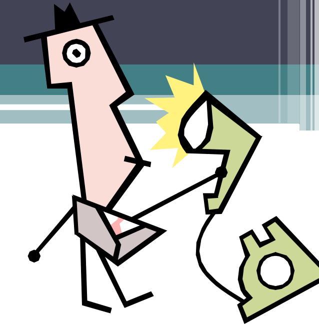
CEPSTRUM

- Problém! Budenie `g(t)` je konvoluované s impulznou odozvou filtra:

$$s(t) = g(t) \star h(t) = \int_{-\infty}^{+\infty} g(\tau)h(t-\tau)d\tau,$$

- teda vo frekv. oblasti súčin: $S(f) = G(f)H(f).$
- Ani v jednej oblasti nejde tieto dve zložky dobre oddeliť.
- Riešenie: ***nelinearita***, ktorá prevedie súčin na súčet

Kepstrálna analýza

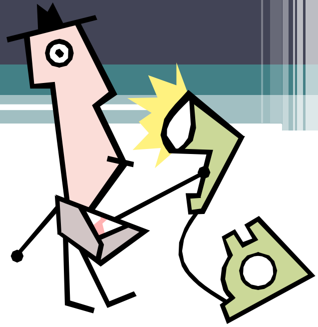


- Inverzná Fourierová transformácia logaritmu komplexného spektra.
- Ak nepotrebujeme fázovú informáciu stačí nám amplitúdové spektrum.

$$c_n = IDFT\{\ln|X[k]|\} = \sum_{k=0}^{N-1} \ln|X[k]| e^{jkn\frac{2\pi}{N}}$$

- Koeficienty c_n nazývame (reálnym) kepstrom signálu
- Použitie logaritmu umožňuje vykonať dekonvolúciu vstupného signálu, ak je jedna z jeho zložiek periodická postupnosť impulzov => možnosť použiť na meranie periodicity
- Kepstrálna transformácia prevádza pôvodné vzorky (periodické a neperiodické) signálu x[n] na ich súčet.
  - Ak je neperiodická zložka (sústredená okolo nulových kvefrencií) silno utlmená, je možné ju od periodickej zložky oddeliť jednoduchým výberom (liftráciou) koeficentov c_n.
- Výber koeficientov ovplyvňuje tvar spektra signálu. Pri výbere malého počtu (napr. 10) prvých kepstrálnych koeficientov získame aproximáciu spektra signálu (vyhladené spektrum).

Kepstrálna analýza

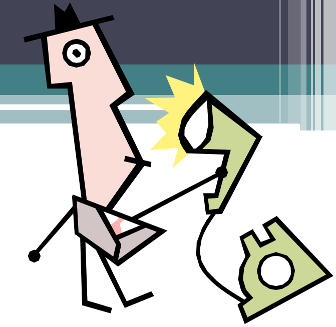


- Je možné ju použiť aj pre výpočet euklidovskej vzdialenosti.
- Spojenie kepstrálneho popisu signálu a nelineárnej frekvenčnej osi (výsledkom môžu byť MFCC koeficienty) umožňuje aproximáciu fyziológie počutia a je odolnejšie voči šumom ako AR koeficienty.
- Reálne *kepstrum* je možné získať aj pomocou AR koeficientov, ak rozvinieme prirodzený logaritmus zo vzťahu pre výpočet kepstra do Taylorovho radu a vykonáme deriváciu rádu aj logaritmu podľa premennej *z* získame rekurentný vzťah pre koeficienty a_n a c_n:

$$c_0 = \ln G$$

$$c_n = -a_n - \frac{1}{n}\sum_{k=1}^{n-1}(n-k)a_k c_{n-k}$$

- Ak vypočítame kepstrum z AR koeficientov, dostaneme vyhladené LPC kepstrum.



Definícia kepstra

$$\ln G(f) = \sum_{n=-\infty}^{+\infty} c(n) e^{-j2\pi f n}$$

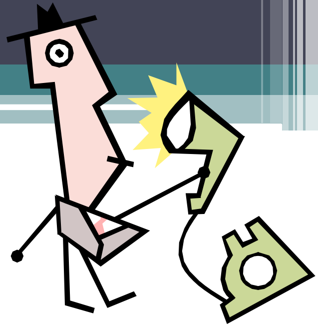
- `G(f)` je *spektrum budenia*
- Hodnoty `c(n)` sú *kepstrálne koeficienty*.
- Nakoľko je `G(f)` párna funkcia, sú `c(n)` **reálne** a platí: `c(n) = c(-n)`
- Suma v rovnici je definíciou **DFT**, preto môžeme `c(n)` vypočítať ako:

$$c(n) = \mathcal{F}^{-1}\left[\ln G(f)\right]$$

Terminológia:

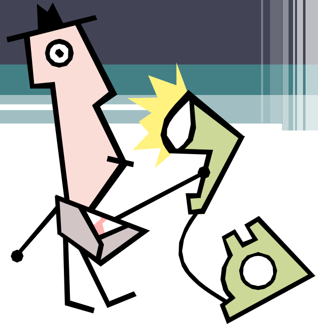
spektrum	$\rightarrow$	***cepstrum***
frekvency	$\rightarrow$	***quefrency***
filtring	$\rightarrow$	***liftring***

Mel-frequency cepstrum - MFCC

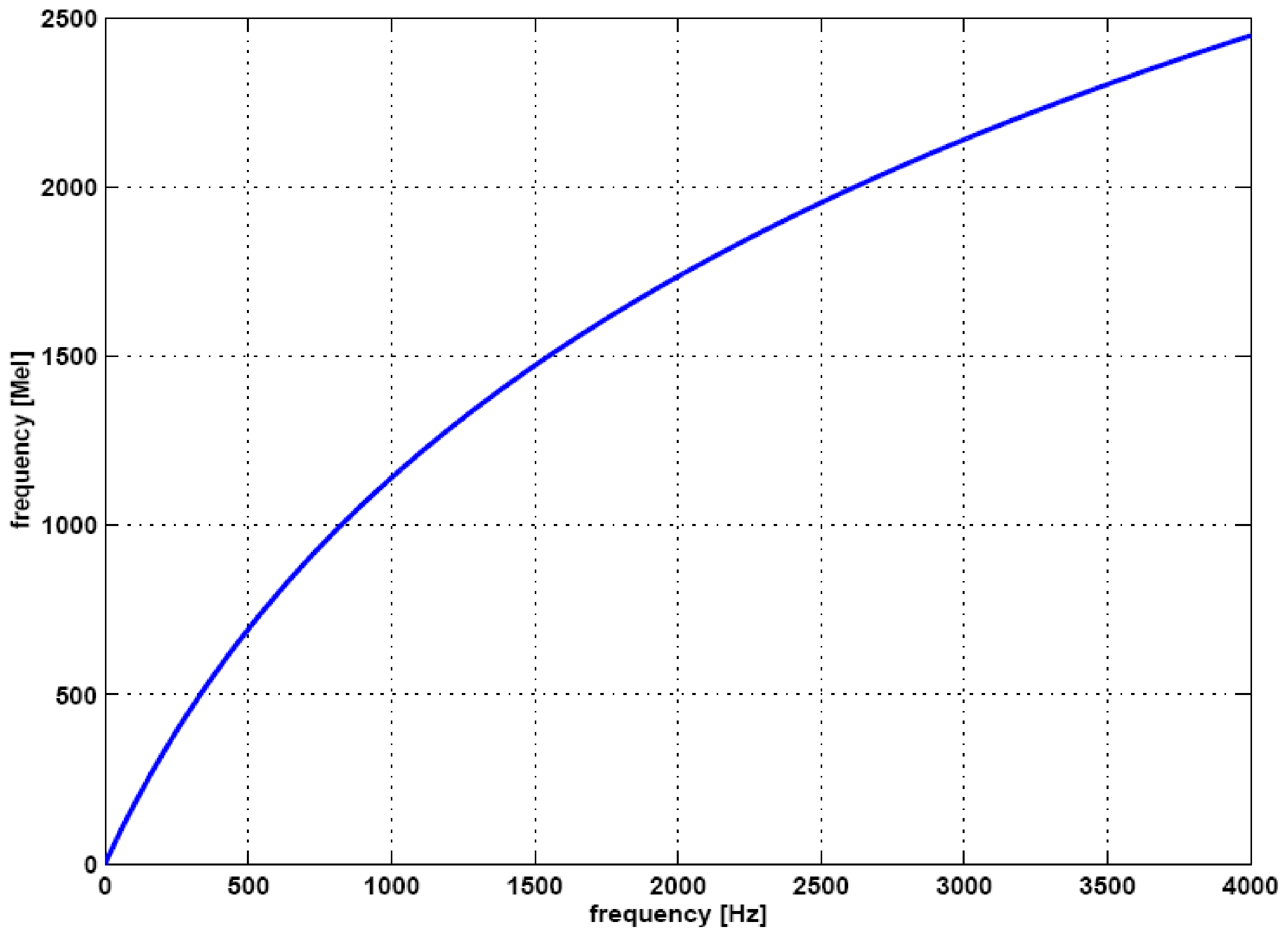


- DFT má všade **rovnomerné frekvenčné rozlíšenie**
- *ľudské ucho* má na nízkych frekvenciách väčšie rozlíšenie než na vysokých - nelinearita
- Napr. Pre účely rozpoznávania reči chceme **priblížiť kepstrum akustickým vlastnostiam ucha**.
- Ako na to?
  - na frekvenčnú os rozmiestnime **nelineárne filtre**, meriame energiu na ich výstupe a použijeme ju namiesto DFT pri výpočte kepstra.
  - **frekvenčnú os** môžeme **nelineárne upraviť** a na upravenú os potom filtre rozmiestniť rovnomerne

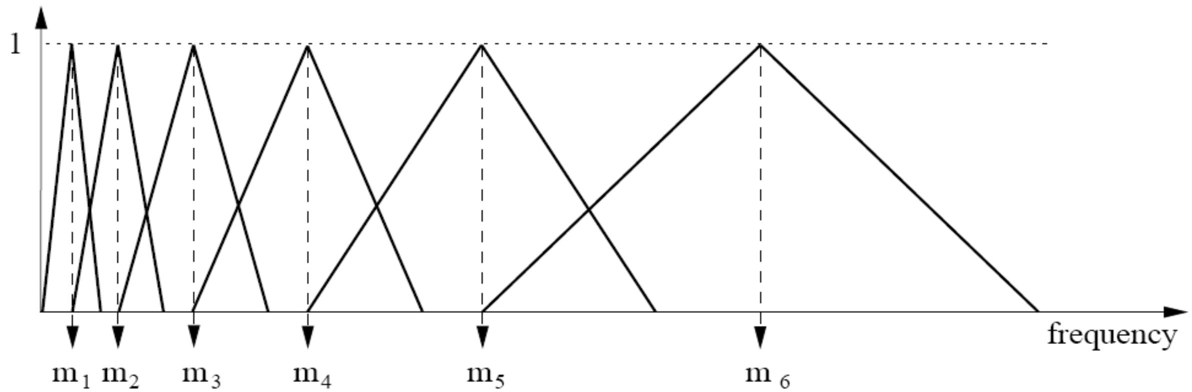
Používaná **nelineárna úprava** využíva prevod *Hz* $\rightarrow$ *Mely*:

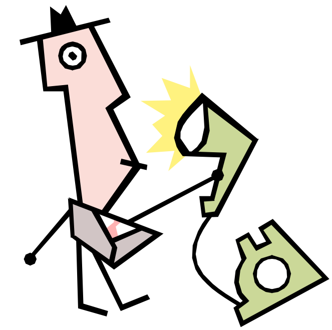


$$F_{Mel} = 2959\log_{10}(1 + \frac{F_{Hz}}{700})$$



- Lineárne rozmiestnenie filtrov na Mel-ovej osi má za následok nelineárne rozmiestnenie na štandardnej frekvenčnej osi v Hz.





Výpočet energie:

1. Skonštruujeme banku filtrov, vstupný signál filtrujeme v časovej oblasti a počítame energiu:

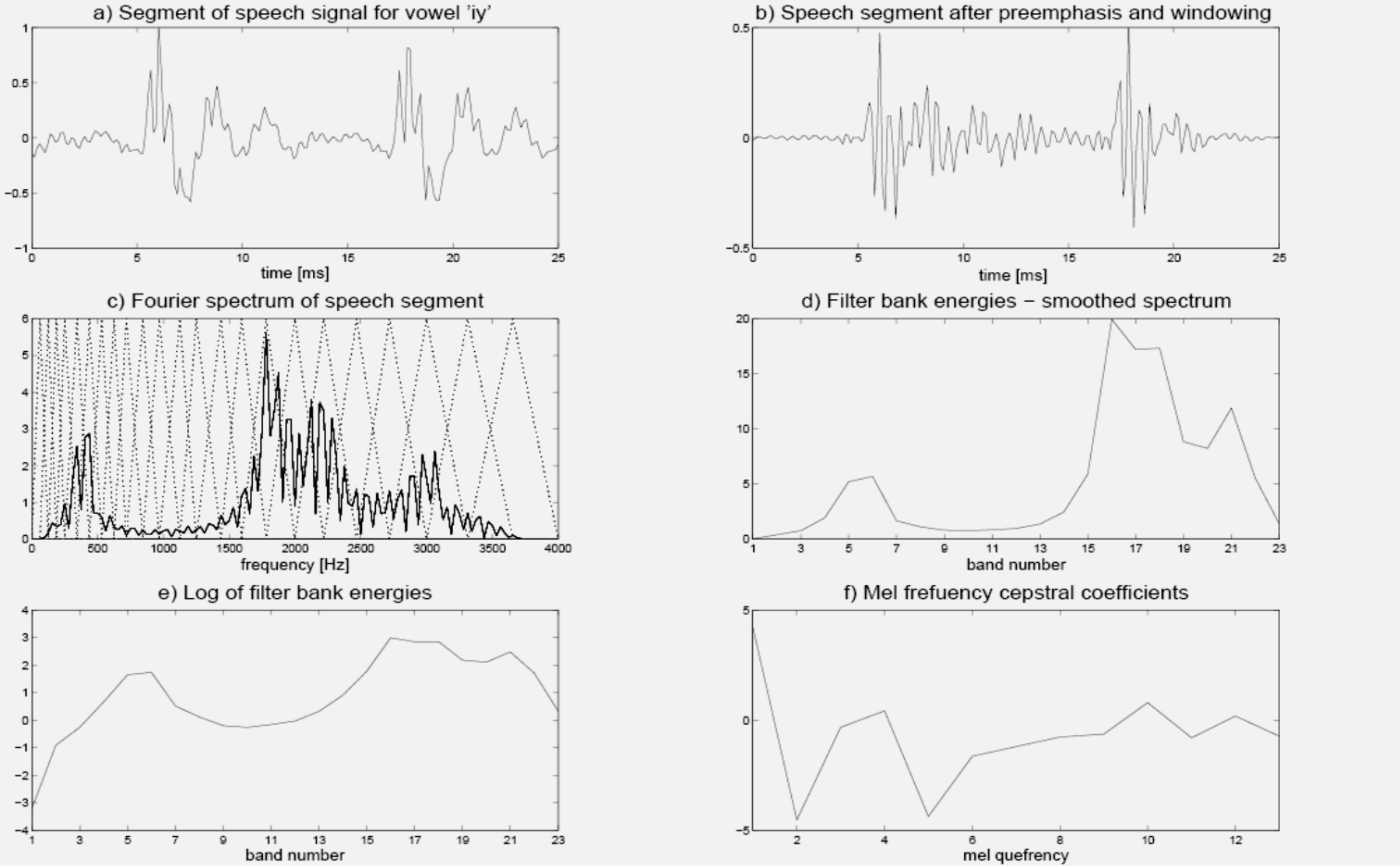
   $\sum_n s_i^2(n)$ ... je to veľmi komplikované!

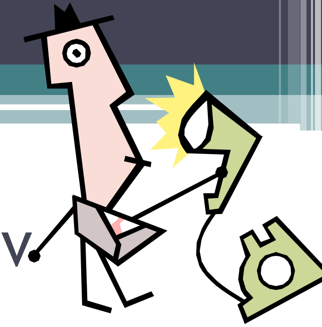
2. Vykonáme DFT, umocníme, vynásobíme trojuholníkovým oknom a spočítame.

Spätnú FT môžeme realizovať pomocou diskrétnej kosínusovej transformácie (DCT):

$$c_{mf}(n) = \sum_{i=1}^{K} \log m_k \cos\left[n(k-0.5)\frac{\pi}{K}\right]$$

Mel-frekvenčné cepstrálne koeficienty (MFCC)


a) Segment of speech signal for vowel 'iy'
time [ms]
b) Speech segment after preemphasis and windowing
time [ms]
c) Fourier spectrum of speech segment
frequency [Hz]
d) Filter bank energies − smoothed spectrum
band number
e) Log of filter bank energies
band number
f) Mel frefuency cepstral coefficients
mel quefrency



Koherenčné charakteristiky viackanálových signálov

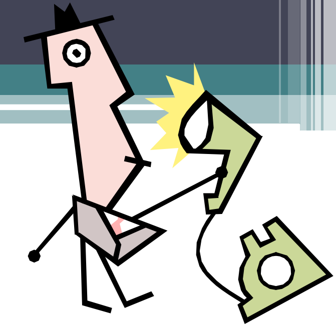
- *Koherenčná funkcia pre dvojicu signálov rozširuje informácie o vzájomnej korelovanosti medzi analyzovanými signálmi.*
- Na rozdiel od korelácie *je koherencia medzi dvoma signálmi definovaná vo frekvenčnej oblasti a vyjadruje mieru korelovanosti medzi jednotlivými signálmi pre jednotlivé frekvenčné zložky*.
- Často používanou charakteristikou je *kvadrát modulu koherenčnej funkcie* označovaný ako *MSC* (Magnitude Squared Coherence):

$$\gamma_{xy}^2[k] = \frac{\left|S_{xy}[k]\right|^2}{S_x[k]S_y[k]}$$

- Kde S_x[k] a S_y[k] sú odhady spektrálnej výkonovej hustoty signálov x[n] a y[n] a S_{xy}[k] je odhad vzájomnej spektrálnej výkonovej hustoty:

$$S_{xy}[k] = \frac{1}{M}\sum_{i=0}^{M-1}\frac{1}{N}X_i^*[k]Y_i[k]$$

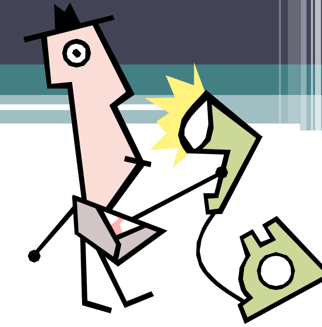
Nutné vychádzať z vyhladených odhadov spektra!!!



Koherenčné charakteristiky viackanálových signálov

- Výhody:
  - Koherenčná funkcia nie je veľmi citlivá na posun signálov, ak je menší ako dĺžka analyzovaného segmentu
  - Koherencia poskytuje informáciu o korelovanosti jednotlivých fr. zložiek.
- Ak rečový signál vychádza z jedného zdroja (úst, reproduktora) a je snímaný vo viacerých kanáloch je silne korelovaný v celom frekv. rozsahu a hodnota MSC sa blíži k 1.
- Naproti tomu, ak šum pozadia nevychádza z jedného zdroja (odrazy reči, hluk v miestnosti, zvuky automobilu) koherencia je podstatne nižšia. => vhodnosť koherenčnej charakteristiky pre detekciu reči v šumovom pozadí.

Momentové charakteristiky



- *<u>Prvý spektrálny moment</u>* je váhovaným priemerom spektrálnych čiar:

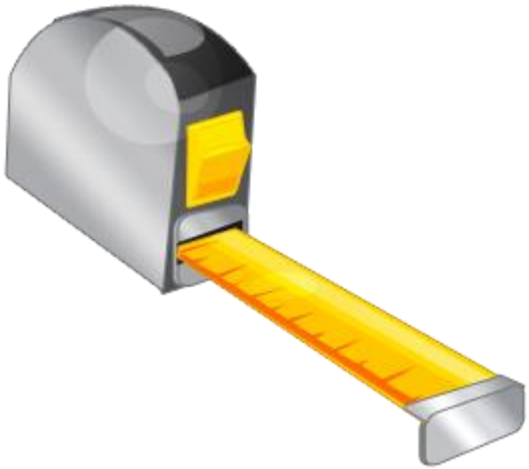
$$mom_1 = \frac{\sum_{k=1}^{N/2} k\left|X[k]\right|}{\sum_{k=1}^{N/2}\left|X[k]\right|}$$

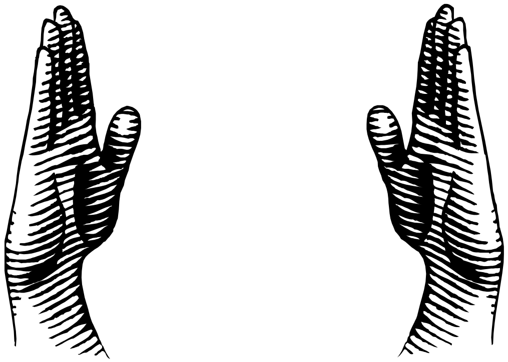
- Predstavuje **ťažisko rozloženia amplitúd jednotlivých spektrálnych zložiek**.

- *<u>Druhý spektrálny moment</u>* je ekvivalentný statickému rozptylu a **je indikátorom rozprestretia spektra**:

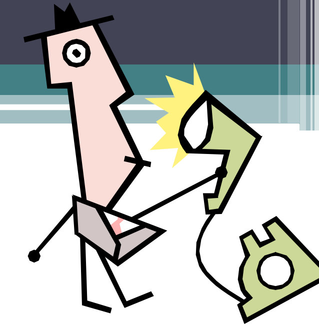
$$mom_2 = \sqrt{\frac{\sum_{k=1}^{N/2} k^2\left|X[k]\right|}{\sum_{k=1}^{N/2}\left|X[k]\right|} - (mom_1)^2}$$

- Vyššie hodnoty majú tie hlásky, ktorých energia je viac rozprestretá cez celé spektrum.





Euklidovská vzdialenosť



Euklidovská vzdialenosť

- Pre klasifikáciu reči (hlások) do fonetických tried sa používajú rôzne druhy parametrov, ktoré vytvárajú parametrické vektory s rôznou dĺžkou (LPCC, MFCC).

- Ak máme fonetické triedy popísané množinou M parametrov, potom najjednoduchším rozhodnutím o zaradení pozorovaných dát do týchto tried je využitie euklidovskej vzdialenosti:

$$d = \sqrt{\sum_{m=1}^{M} \left(r[m] - o[m]\right)^2}$$

- Kde *r*[m] je *M* parametrov (vektor) referenčného vzoru

- A *o*[m] je *M* parametrov (vektor) pozorovaného úseku signálu