A6M33BIO - Biometrie

Biometrické metody založené na rozpoznávání hlasu II

Doc. Ing. Petr Pollák, CSc.

4. prosince 2012 - 11:59

Obsah přednášky

- **Kepstrální příznaky**
  - Definice kepstra signálu, základní vlastnosti
  - DFT a LPC kepstrum
  - MFCC, mel-frekvenční kepstrum
  - kepstrální vzdálenost
- **Reprezentace řečníka a algoritmy klasifikace**
  - Časové funkce (DTW)
  - Kódová kniha (VQ)
  - Statistický model (GMM)
- **Příklady systémů verifikace**

I. část

Kepstrální příznaky

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)

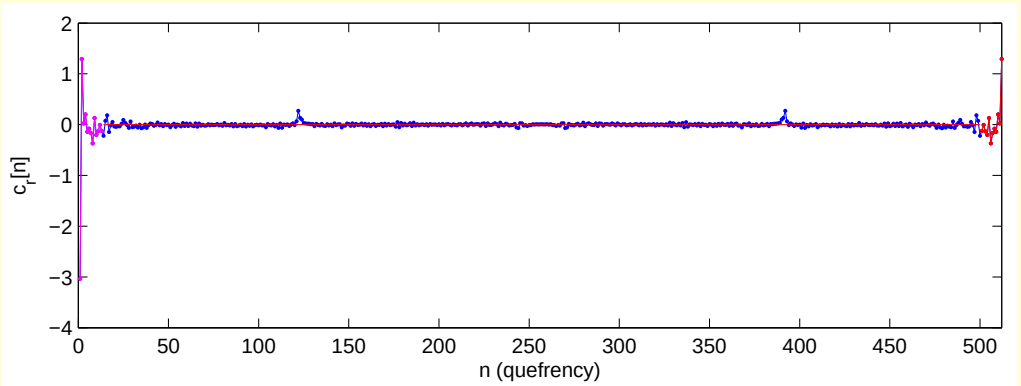
- gamnituda (magnituda, amplituda)
- .....

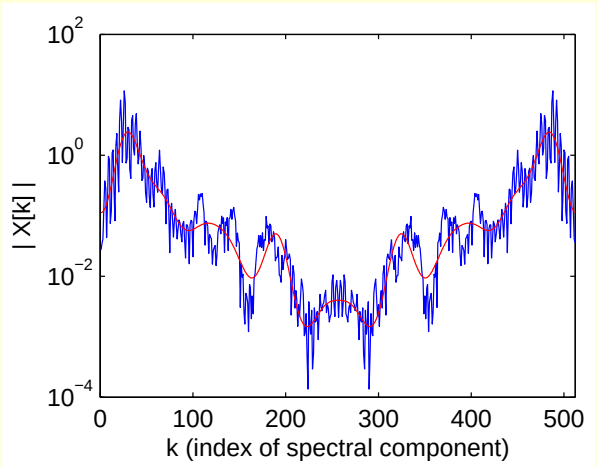
Vlastnosti reálného DFT kepstra

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

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, $G = \sqrt{E_p}$

$$c_0 = \ln G$$

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

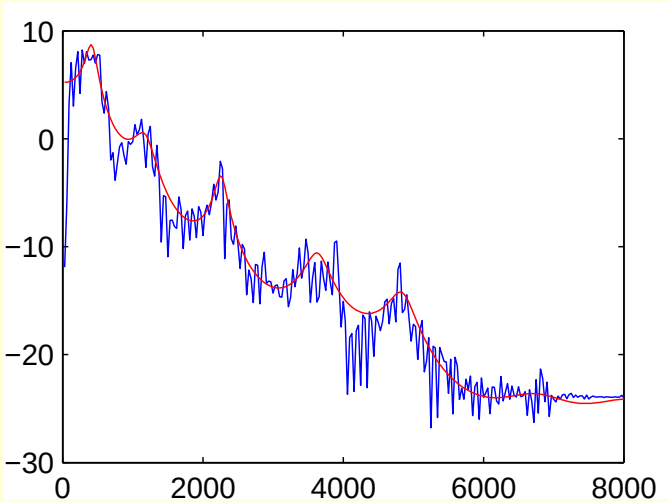
$$c_n = -\frac{1}{n}\sum_{k=1}^{n-1}(n-k)a_k c_{n-k}, \text{ pro } n = p+1, p+2, ...$$

Vlastnosti:

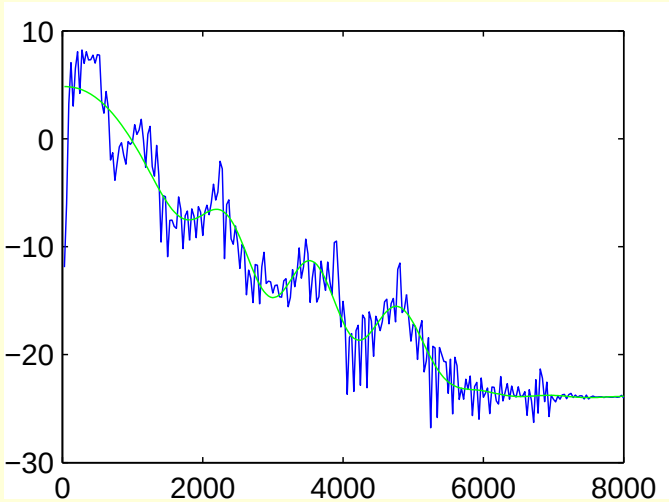
- Koeficienty Taylorova rozvoje $\ln|H(z)|$ (inverzní Z-transf.)
- Nekonečně dlouhé, první hodnoty opět nejvýznamnější
- Lze spočítat rekurentně, neobsahuje náhodnou složku
- Tvar spektra kopíruje LPC spektrum

Kepstrální analýza pro zpracování řeči

LPC spektrum:



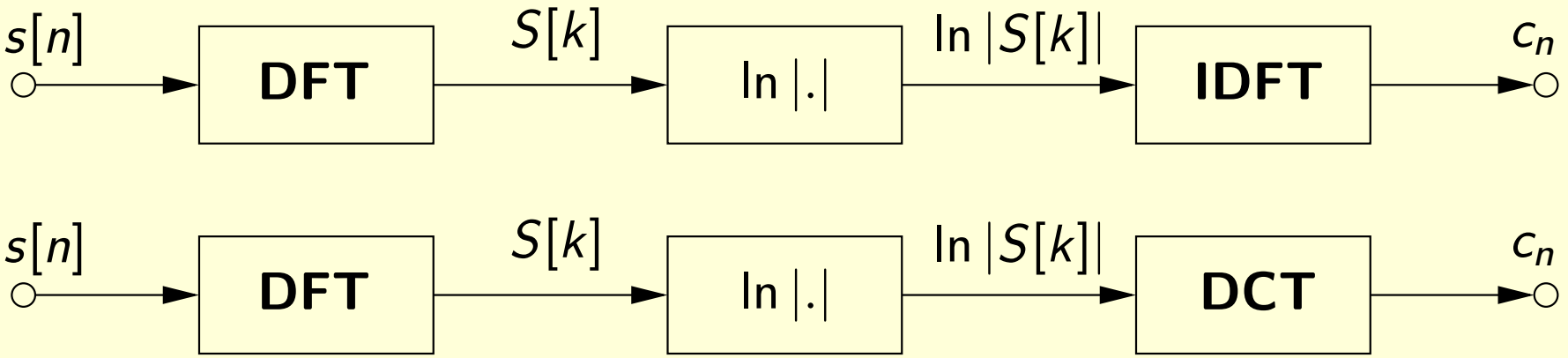
Vyhlazený odhad z reálného kepstra:



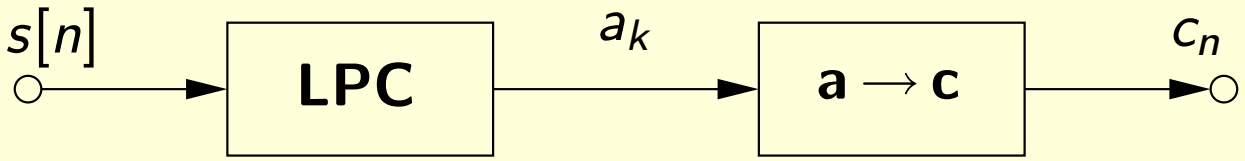
- První koeficienty nesou hlavní informaci o tvaru amplitudového spektra
  - typicky 12 kepstrálních koeficientů
- **kepstra podobných segmentů tvoří shluky (!!!)**
  $\Longrightarrow$ **použití jako příznaky pro rozpoznávání**
- IDFT lze nahradit DCT (zajímavé pro reálné implementace)

DFT a LPC kepstrum - bloková schémata výpočtu

Blokové schéma výpočtu kepstrálních koeficientů pomocí DFT



Blokové schéma výpočtu LPC kepstrálních koeficientů



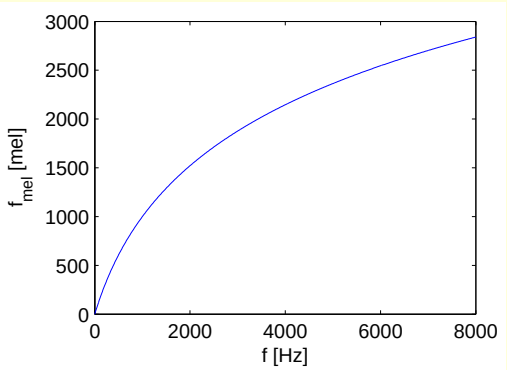
Melodická a Barkova frekvenční stupnice

Nelineární frekvenční stupnice
$\approx$ **modelování nelinearity vnímání frekvence lidským sluchem**

Nelineární zkreslení frekvenční osy - *melodická stupnice*

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

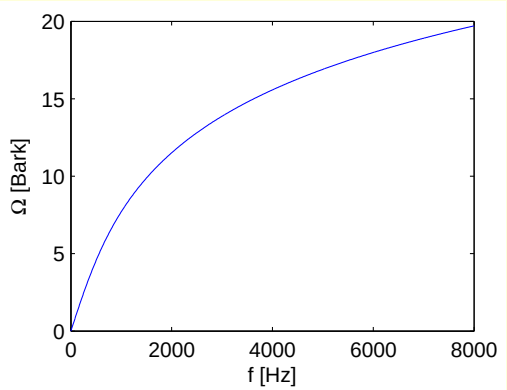
$$f = \mathrm{InvMel}(f_{mel}) = 700 \cdot \left(10^{\frac{f_{mel}}{2595}} - 1\right)$$



Barkova stupnice - definovaná na bázi kritických pásem slyšení

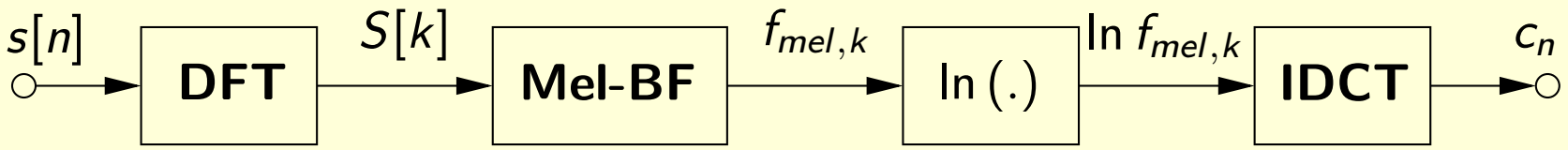
$$\Omega = \mathrm{Bark}(f) = 6 \ln\left(\frac{f}{600} + \sqrt{\left(\frac{f}{600}\right)^2 + 1}\right)$$

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



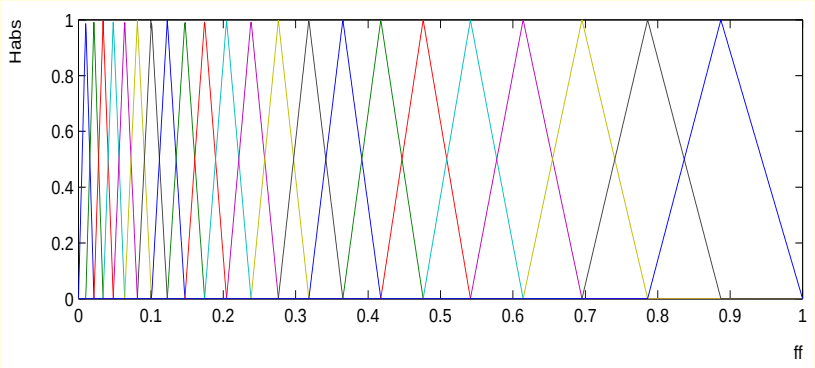
MFCC - Melovské kepstrální koeficienty

Blokové schéma výpočtu mel-kepstrálních koeficientů:



Výpočet energie v jednom pásmu

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



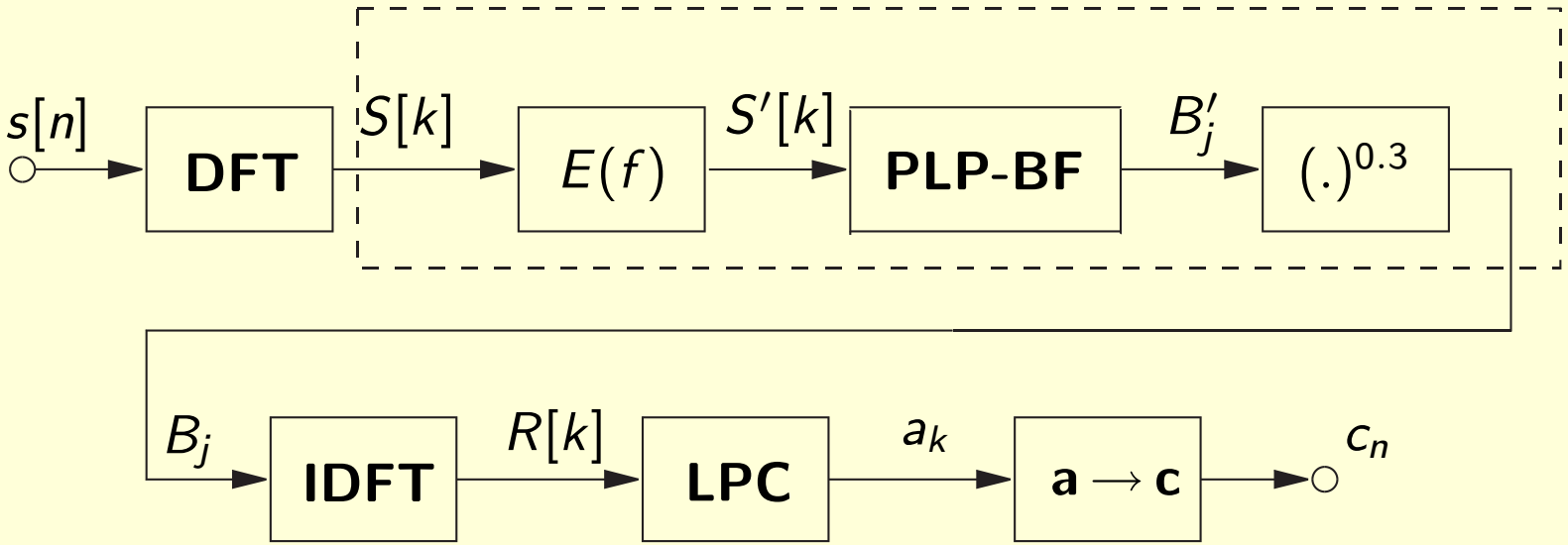
Výpočet kepstra pomocí DCT

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

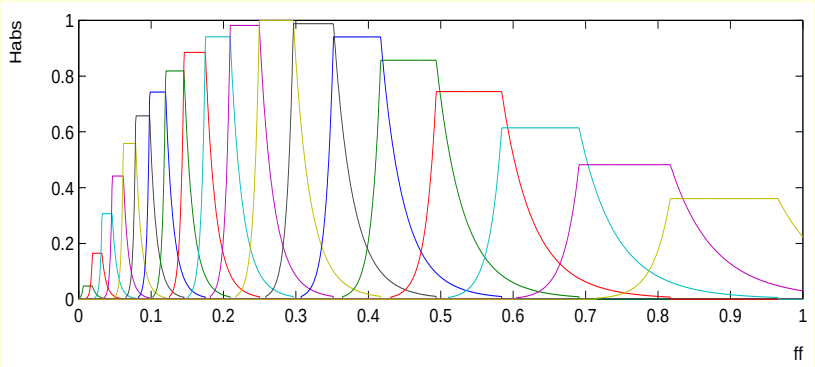
- **nejrozšířenější příznaky používané v ASR**
- často používané i pro **identifikaci a verifikaci řečníka**

PLP kepstrální koeficienty

Blokové schéma výpočtu PLP kepstrálních koeficientů:

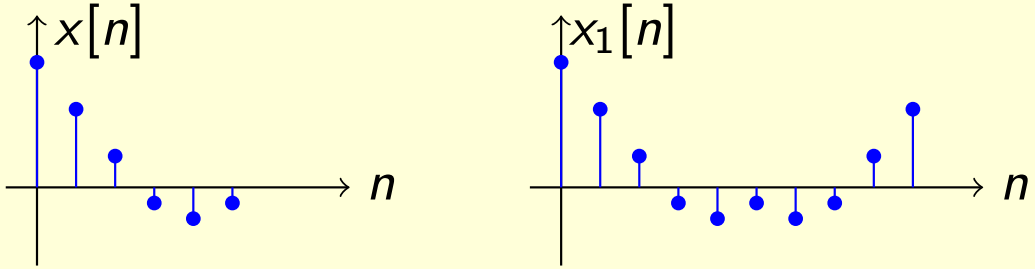


- Barkova stupnice (kritická pásma)
- křivky stejné hlasitosti
- zákon slyšení
- kepstrum se počítá na bázi lineární predikce



DCT a souvislost s DFT: DCT-1, $2N-2$ extenze

Definice DCT-1:



$$X^{c1}[k] = 2\sum_{n=0}^{N-1} \alpha[n]x[n]\cos\frac{\pi kn}{N-1}$$

$$x[n] = \frac{1}{N-1}\sum_{k=0}^{N-1} \alpha[k]X^{c1}[k]\cos\frac{\pi kn}{N-1}$$

$$\alpha[n] = \begin{cases} \frac{1}{2} & \text{pro } n = 0, n = N-1, \\ 1 & 1 \le n \le N-2 \end{cases}$$

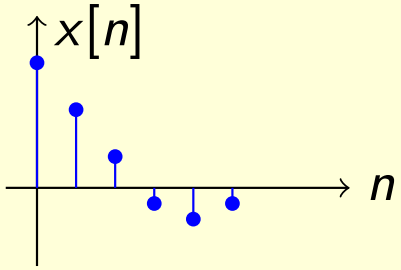
Souvislost s DFT: (N je délka $x[n]$, $2N-2$ extenze)

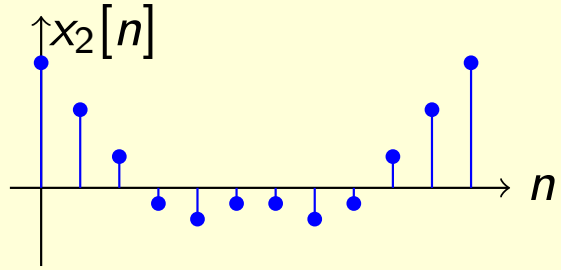
$$X_1[k] = \mathcal{DFT}\{x_1[n]\} \quad \Rightarrow \quad X^{c1}[k] = X_1[k] \text{ pro } k = 0,..,N-1$$

POUŽITÍ - **výpočet DFT reálného kepstra (nahrazuje IFFT)**
(DFT spektrum - 2N-2 symetrie, ss složka)

DCT a souvislost s DFT: DCT-2, $2N$ extenze

Definice DCT-2:





$$X^{c2}[k] = 2\sum_{n=0}^{N-1} x[n]\cos\frac{\pi k(2n+1)}{2N}$$

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} \beta[k]X^{c2}[k]\cos\frac{\pi k(2n+1)}{2N}$$

$$\beta[n] = \begin{cases} \frac{1}{2} & \text{pro } n = 0, \\ 1 & 1 \le n \le N-1 \end{cases}$$

Souvislost s DFT: (N je délka $x[n]$, $2N-2$ extenze)

$$X_2[k] = \mathcal{DFT}\{x_2[n]\} \Rightarrow X^{c2}[k] = X_2[k]\cdot e^{-j\frac{\pi k}{2N}} \text{ pro } k = 0,..,N-1$$

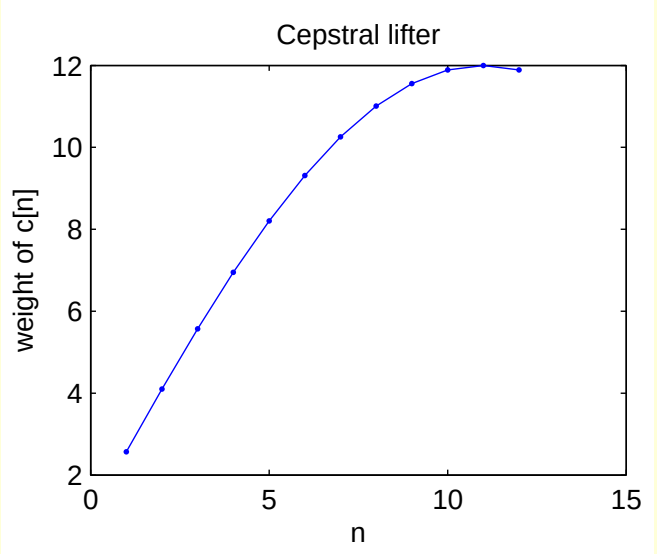
POUŽITÍ - **výpočet MFCC (nahrazuje IFFT)**

(Spektrum na výstupu BF - 2N symetrie, není ss složka)

Liftrace kepstrálních koeficientů

Liftrace - modifikace (váhování) hodnot kepstra

$$c'[n] = \left(1 + \frac{L}{2} \sin \frac{\pi n}{L}\right) \cdot c[n]$$

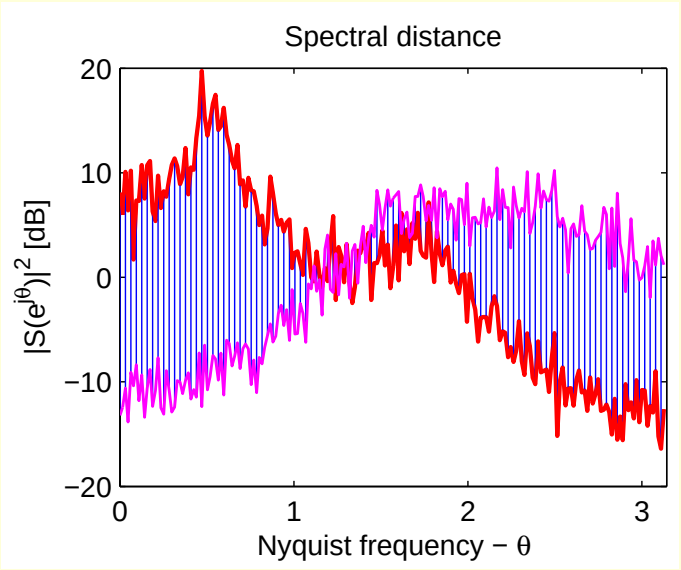


- zvýraznění vyšších kepstrálních koeficientů
  (vyšší kepstrální koeficienty - výrazně menší numerická hodnota)
- zlepšení modelovacích vlastností (výpočet parametrů GMM)

Spektrální vzdálenost

Spektrální vzdálenost (L_2-norma)

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



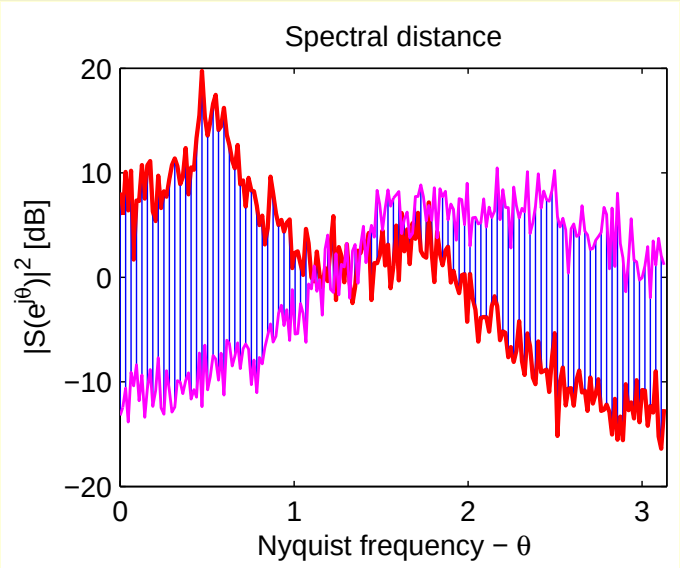
Spektrální vzdálenost na bázi L_2**-normy**

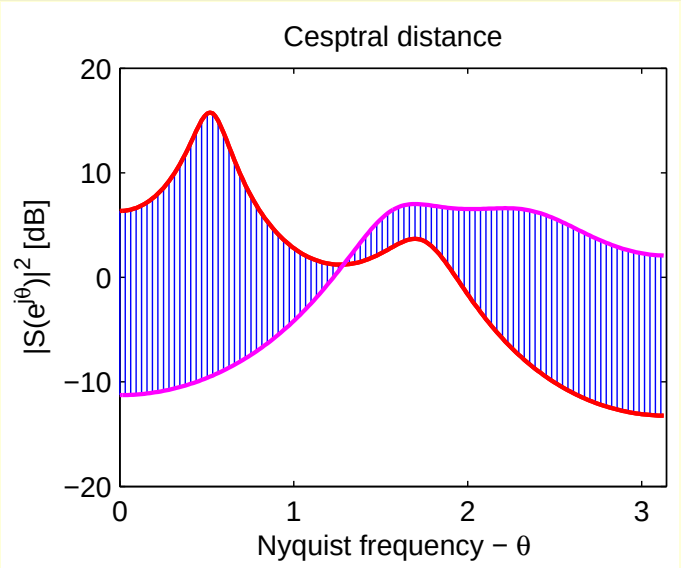
$\longrightarrow$ kvantifikuje plochu ohraničnou dvěma spektry (křivkami)

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

SHRNUTÍ - Používané příznaky:

- základní frekvence (charakteristika hlasu)
- formantové kmitočty (souvislost s délkou vokálního traktu)

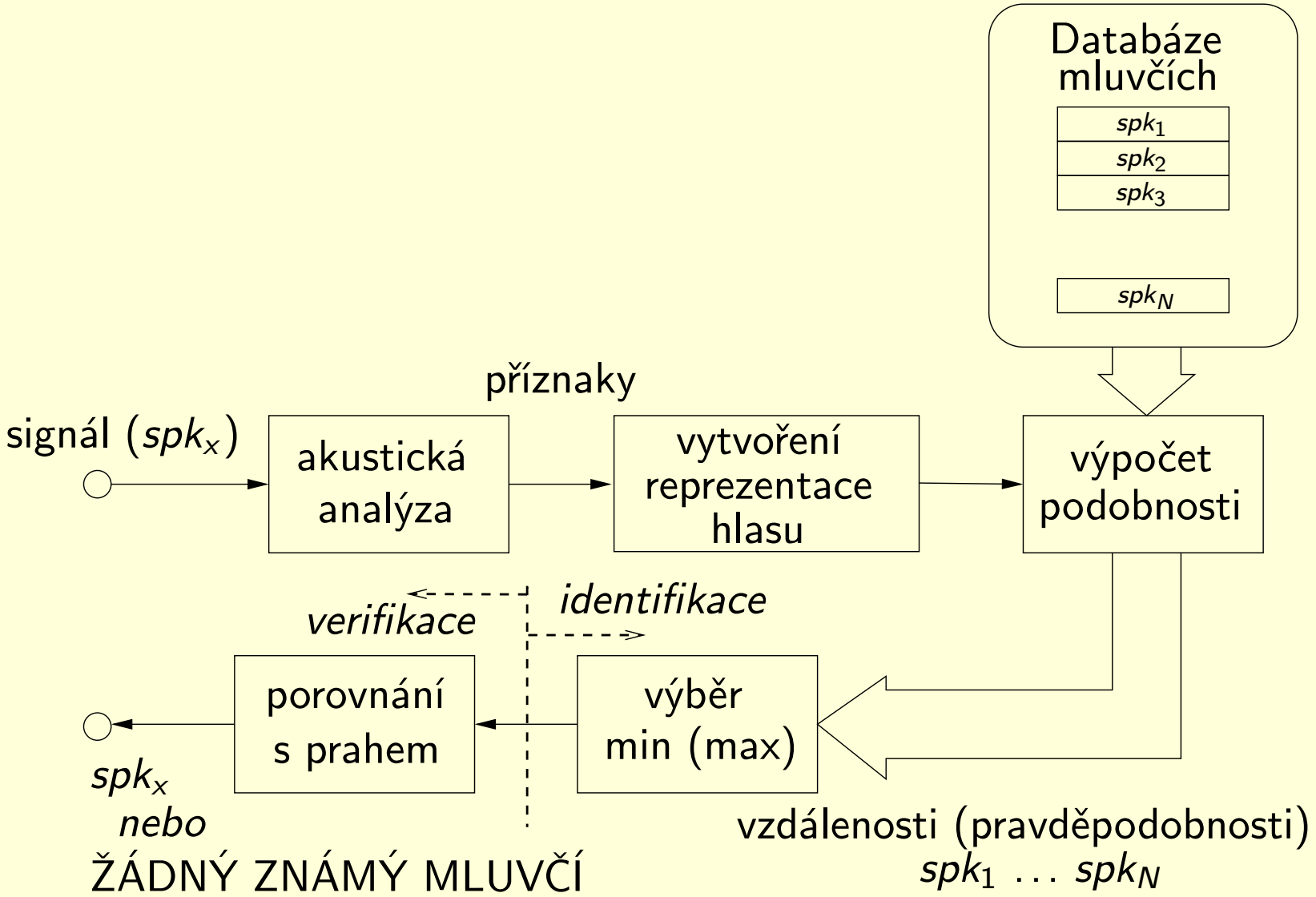
- **MFCC**, PLPC - obecně používané
  (možnost vyhlazení variability mezi mluvčími)
- LPC kepstrální příznaky
  (variabilita mezi mluvčími, přímá souvislost formanty
  malá robustnost vůči šumu)
- parametry AR modelu
  (menší robustnost, Itakurova vzdálenost)

- **kombinované vektory příznaků** pro komplexnější rozhodování
  (f_o, formanty, často textově závislé výskyty)

II. část

Rozpoznávání řečníka

Identifikace mluvčího (v otevřené množině)



- rozpoznání neznámého mluvčího (největší podobnost hlasu)
- **VÝSLEDEK** = **ID mluvčího / skupiny** nebo **ZAMÍTNUTÍ**

Reprezentace hlasu řečníka

Reprezentace mluvčího na bázi vzorů

- Vzorová promluva
  (jako míra podobnosti se počítá vzdálenost mezi vzorovým průběhem a verifikovanou promluvou - princip DTW)

- Kódová kniha používaných parametrů
  (mírou podobnosti je měří kumulované vzdálenosti aktuálních příznakových vektorů od uložených typických reprezentantů)

Reprezentace mluvčího na bázi pravděpodobnostních modelů

- Statistické modely na bázi GMM
  (Gaussian Mixture Models - směsi Gaussovských funkcí modelujících typickou reprezentaci příznaků pro daného řečníka - mírou podobnosti je emitovaná pravděpodobnost)

Rozpoznávání na základě časových funkcí

- vzorová promluva
- průběh energie - rytmizace promluvy
- průběh f_o - intonace v promluvě

Vzdálenost mezi promluvami :

- normalizace délky promluvy - prosté prodloužení, zkrácení (problém pro různou rychlost v rámci promluvy)
- kumulovaná vzdálenost na bázi DTW

$$dist_s = \mathrm{dk}\ (\mathbf{O}, \mathbf{O}^s)$$

Vzdálenost mezi všemi segmenty:

$$d_{i,j} = \sum_{k=1}^{p} (c_k[i] - \overline{c}_k[j])^2 \quad \text{pro } i, j = 1, ..., N$$

Kumulovaná vzdálenost algoritmem dynamického programování:

$$dk_{i,j} = \min\left(dk_{i-1,j} + d_{i,j}, dk_{i,j-1} + d_{i,j}, dk_{i-1,j-1} + d_{i,j}\right)$$

$$\text{pro } i, j = 1, ..., N$$

Klasifikační metody při rozpoznávání řečníka - VQ

klasifikace na bázi VQ

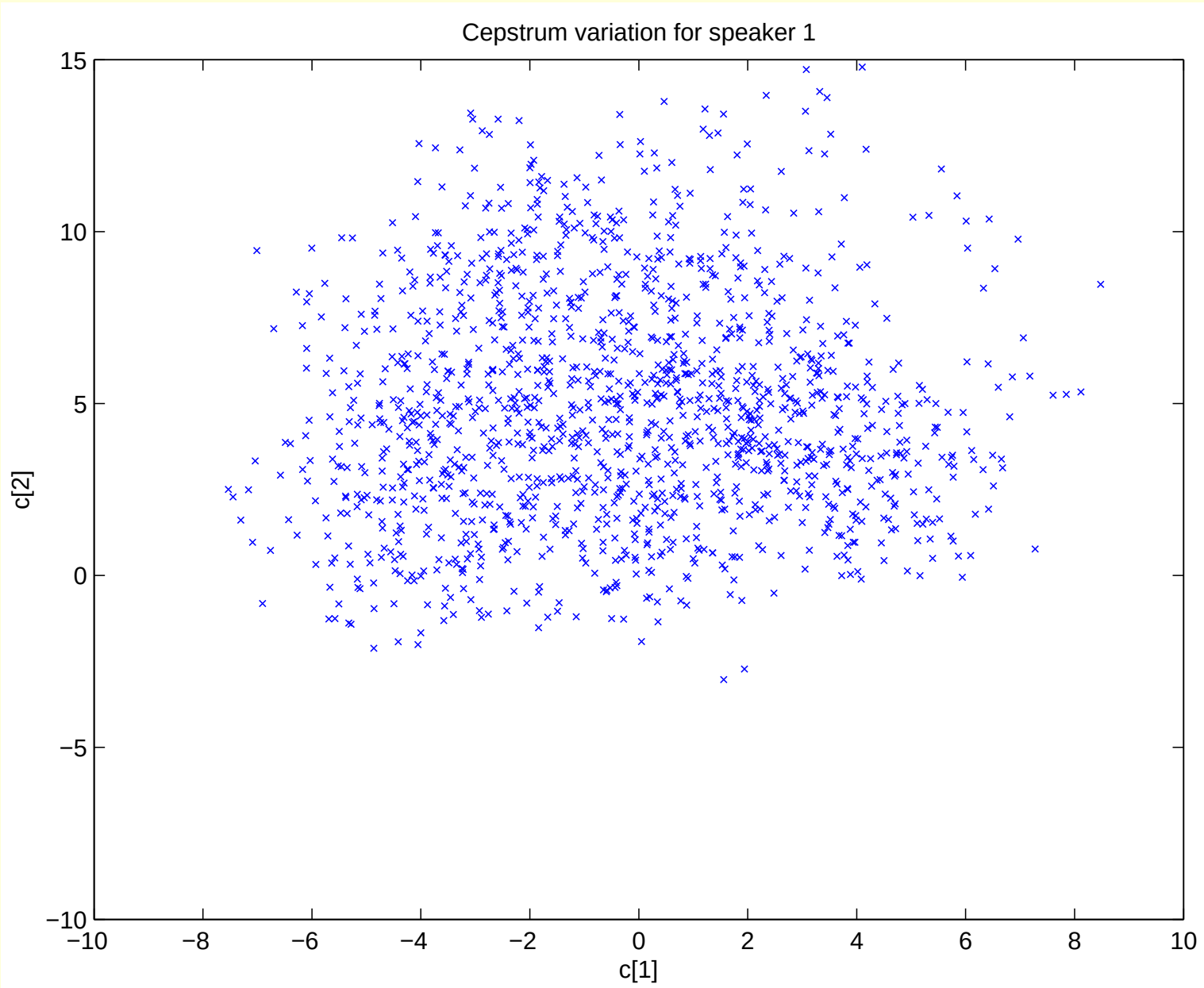
(měření průměrné vzdálenosti aktuálních příznakových vektorů od uložených typických reprezentantů)

---

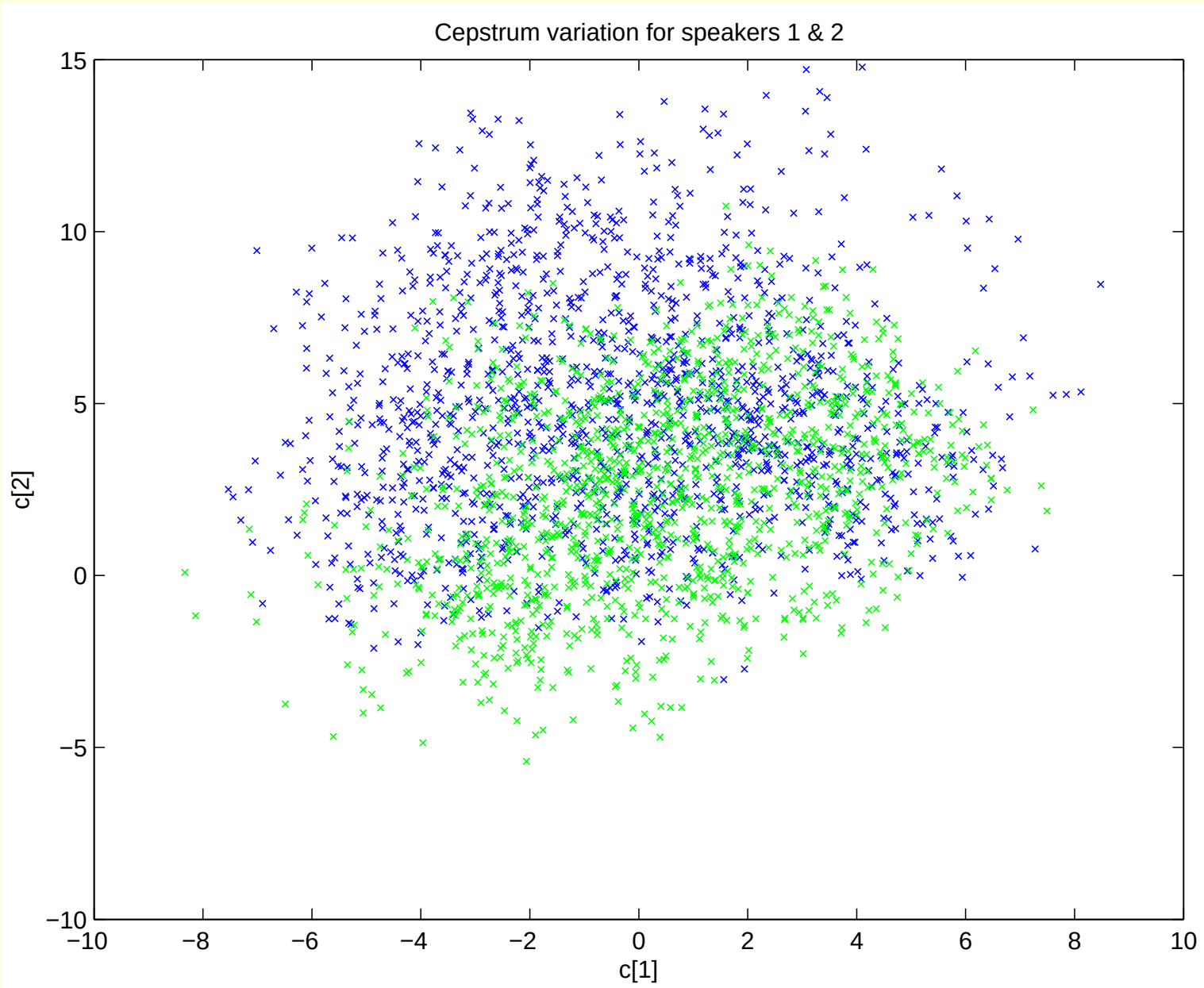
c_{VQ}^{s} ... typická reprezentace - kódová kniha
(výpočet na bázi K-means algoritmu)

$$dist_s = \text{mean}\left(\text{cd}\left(c_i, \underset{c_{VQ,s}}{\text{argmin}}\, \text{cd}\left(c_i, c_{VQ}^{s}\right)\right)\right)$$

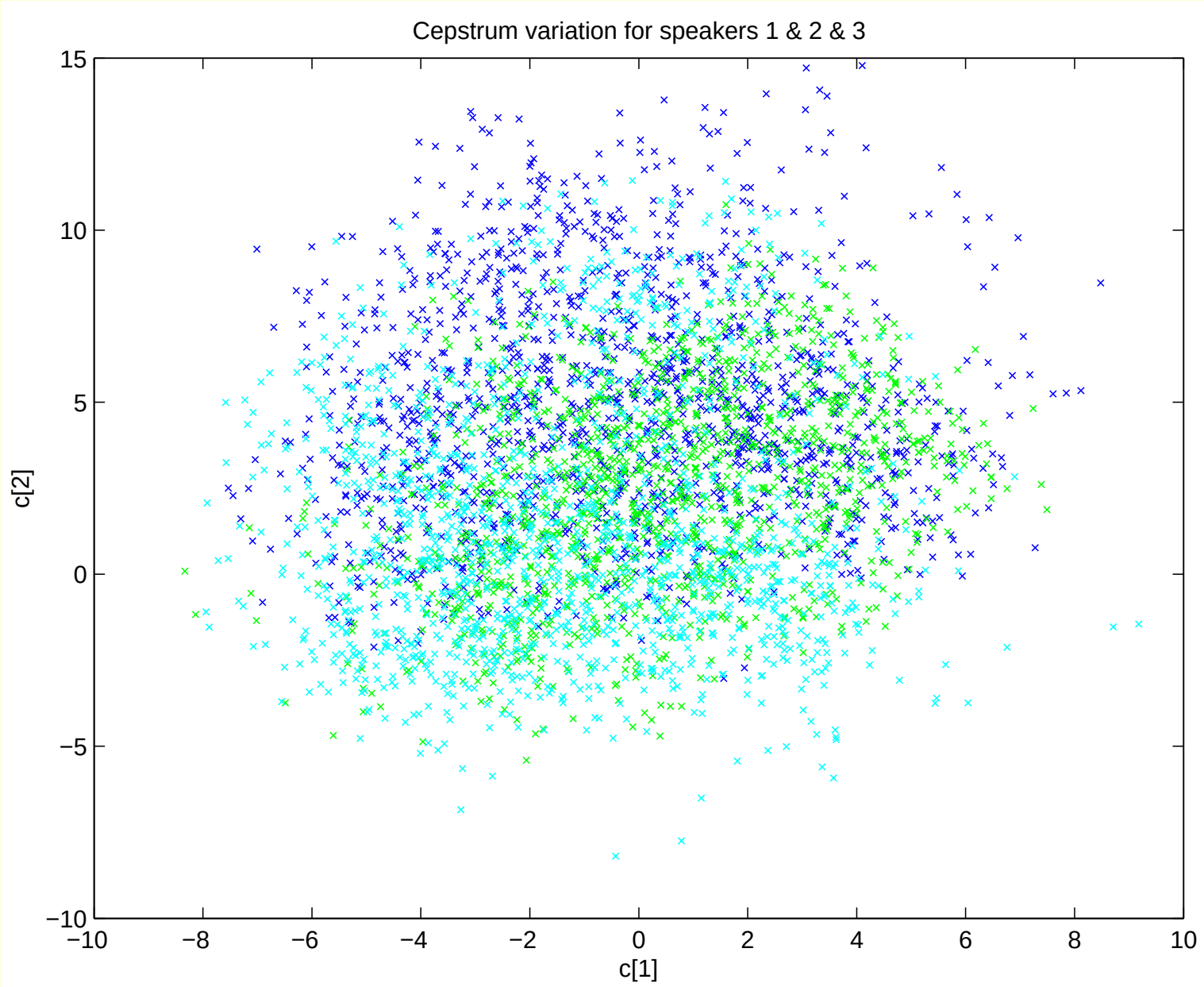
Rozložení kepstra u řečníka - 1



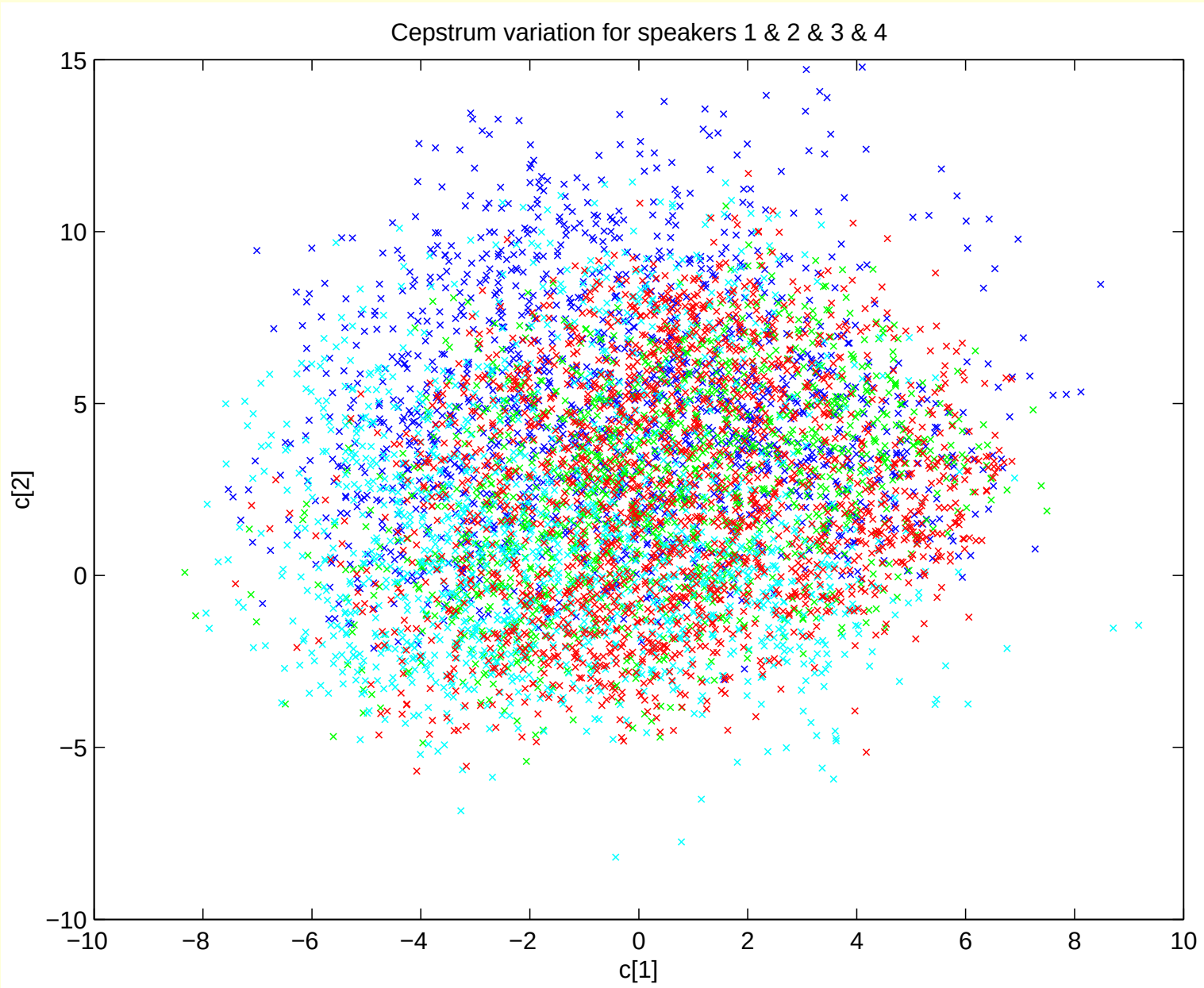
Rozložení kepstra u řečníka - 1 & 2



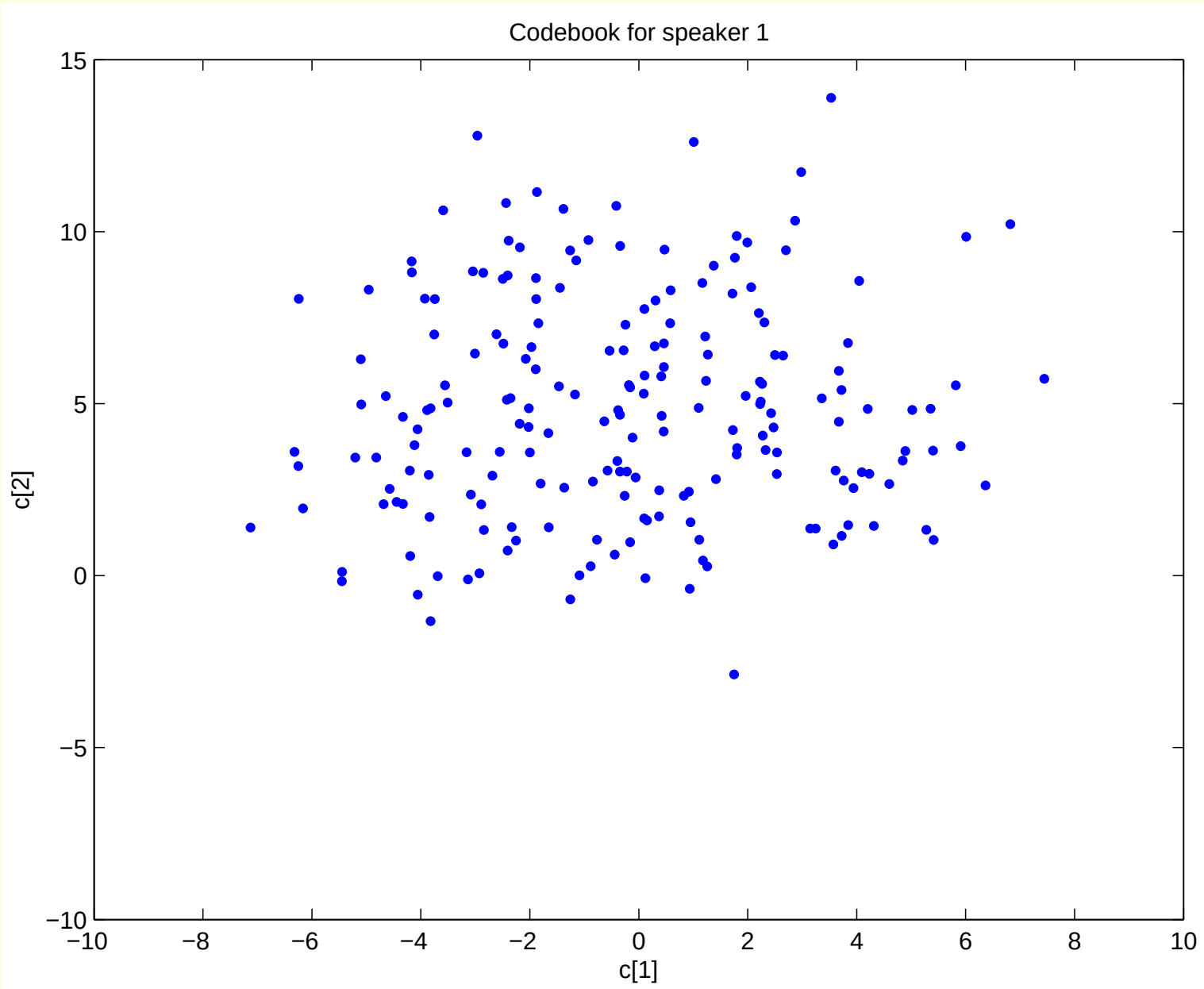
Rozložení kepstra u řečníka - 1 & 2 & 3



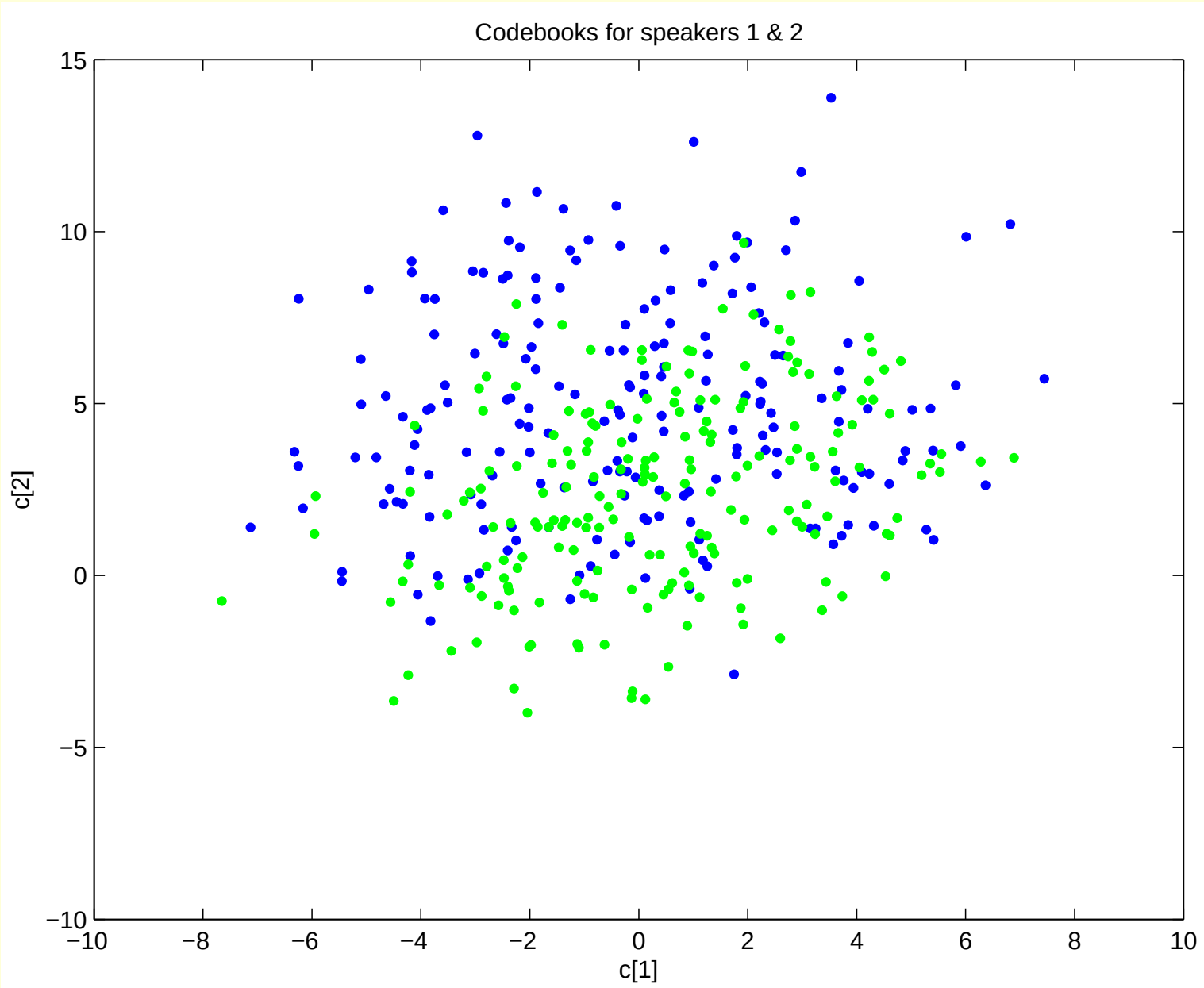
Rozložení kepstra u řečníka - 1 & 2 & 3 & 4



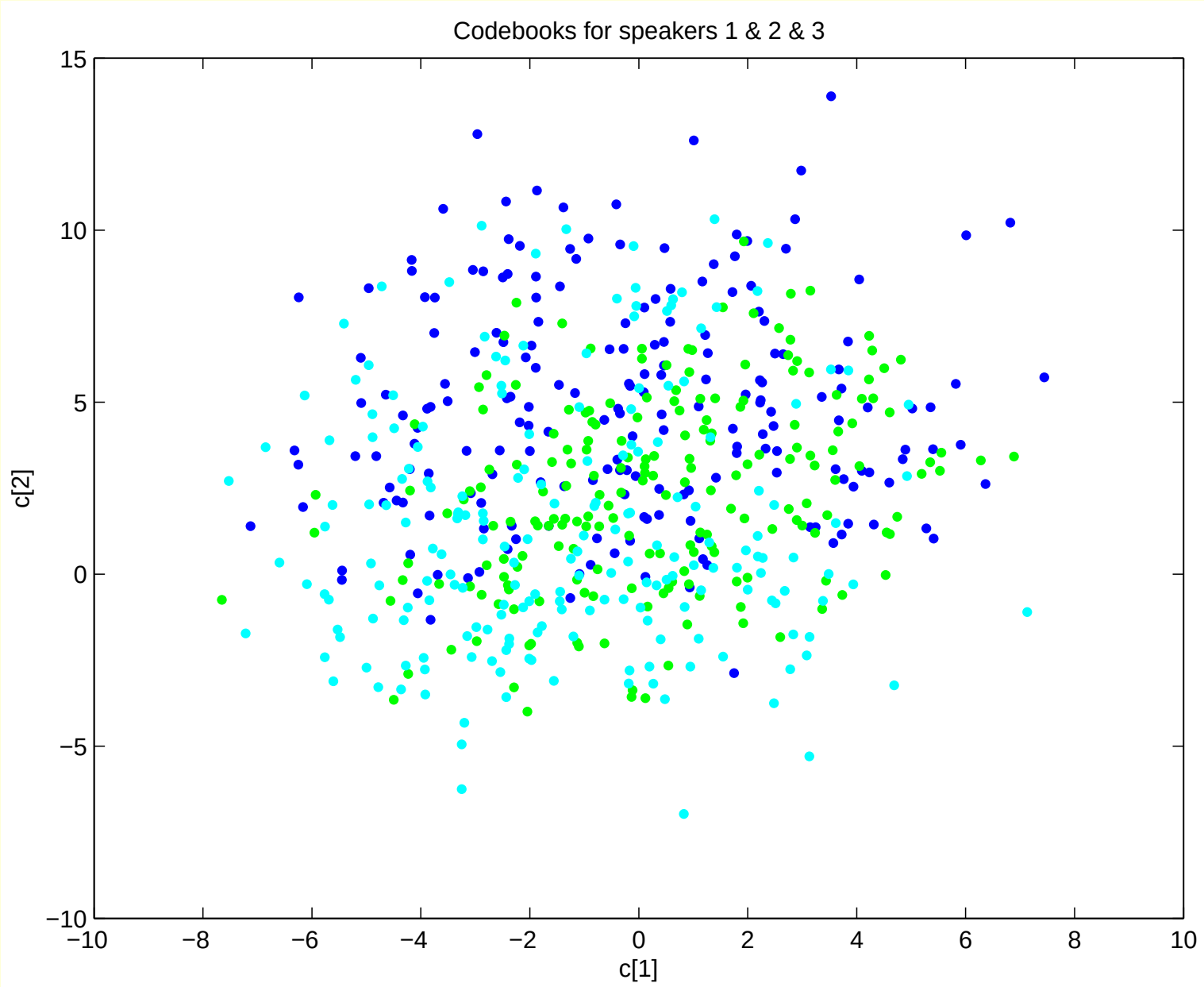
Kódová kniha řečníka - 1



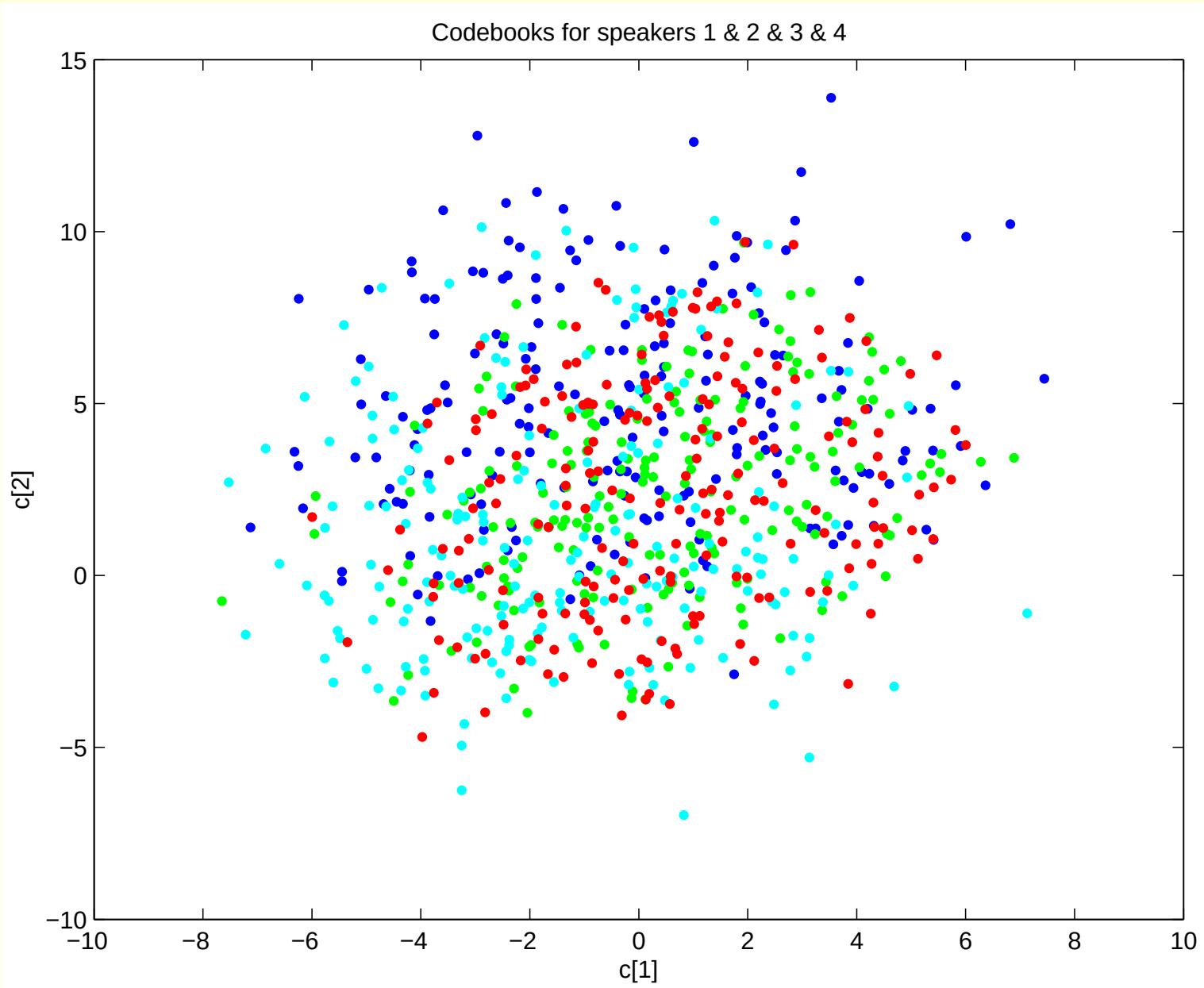
Kódová kniha řečníka - 1 & 2



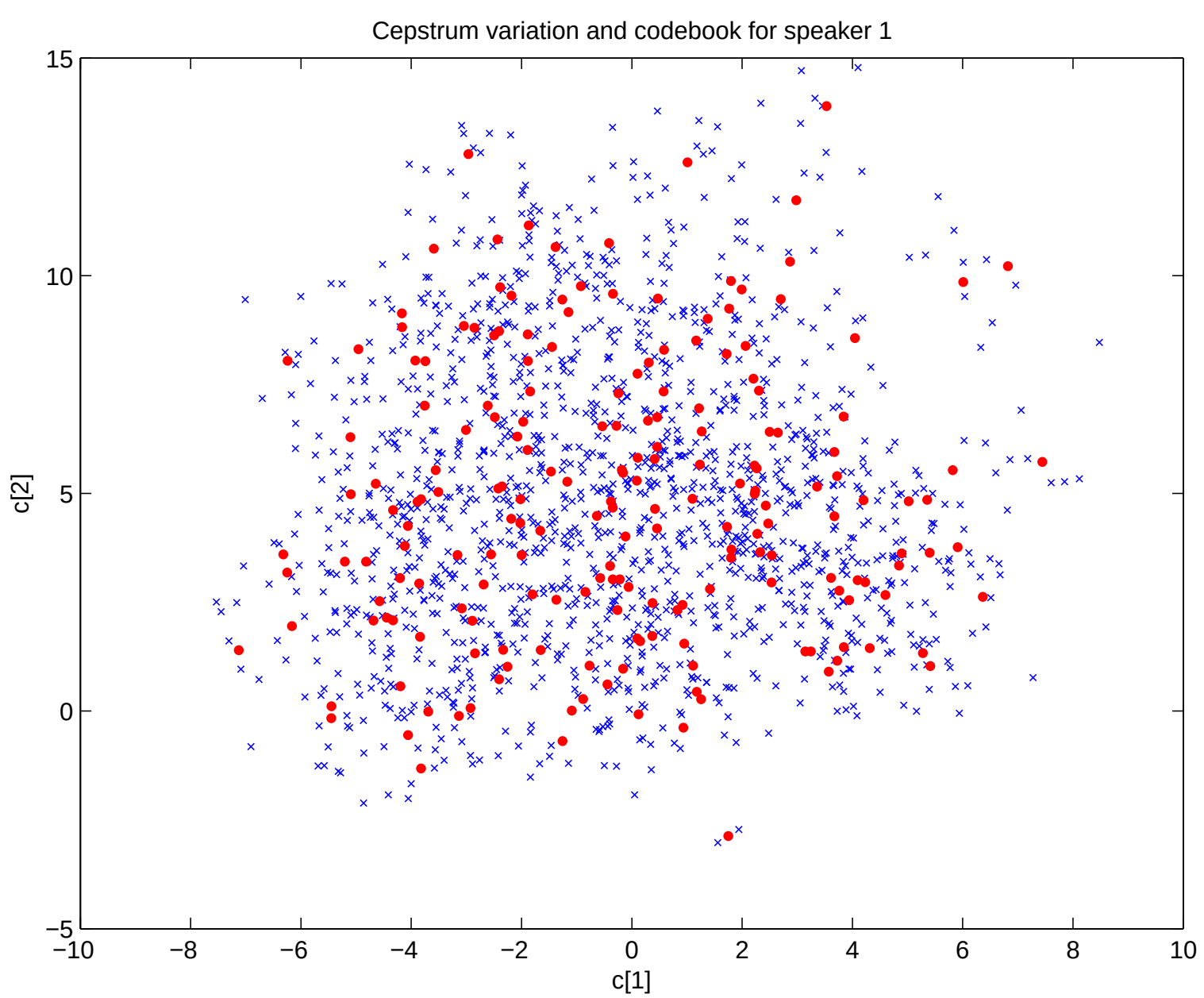
Kódová kniha řečníka - 1 & 2 & 3



Kódová kniha řečníka - 1 & 2 & 3 & 4



Variace kepstra a kódová kniha řečníka - 1



Statistické modelování na bázi GMM

$$p(\mathbf{o}|\lambda^s) = \sum_{i=1}^{M_s} c_i^s \cdot \mathcal{N}(\mathbf{o}, \boldsymbol{\mu}_i^s, \mathbf{C}_i^s)$$

- více směsí modeluje lépe variabilitu příznaků pro daného řečníka
- počty směsí: 8-256 (model řečníka), 512-2048 (model okolí)
  (počty směsí zavisí na množství trénovacích dat)

$$\mathbf{O} = (\mathbf{o_1}, \mathbf{o_2}, ..., \mathbf{o_n})$$

$$P(\mathbf{O}|\lambda^s) = \prod_{j=1}^{N} p(\mathbf{o_j}|\lambda^s)$$

- hodnota pravděpodobnosti se počítá z celé promluvy
  (detektor řeči pro vyřazení neřečových úseků)

Rozhodování v úlohách rozpoznávání řečníka

- **identifikace**

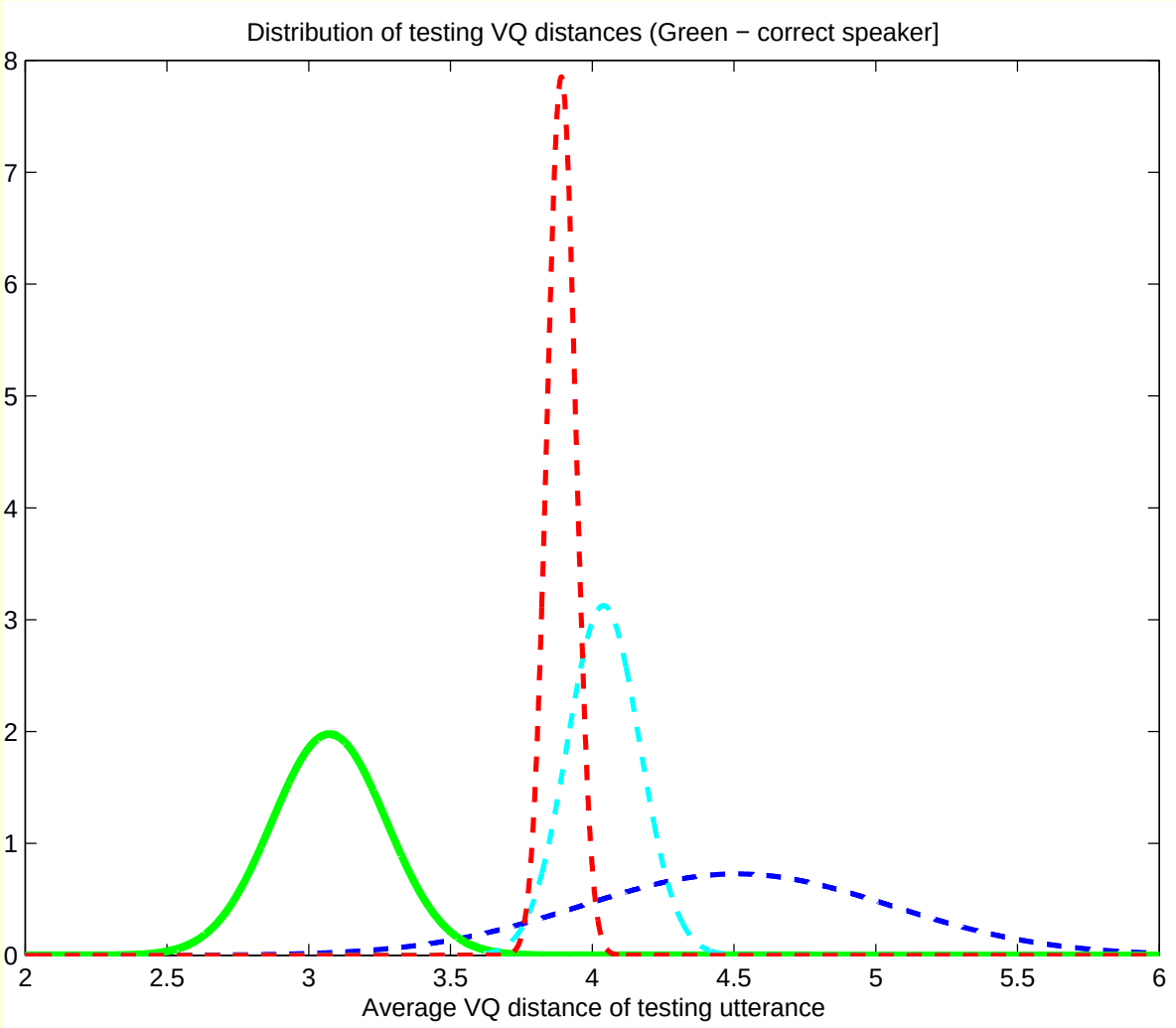
$$spk = \underset{s}{\text{argmin}}\, dist_s \quad \text{pro} \quad s = 1, 2, ..., L$$
$$spk = \underset{s}{\text{argmax}}\, P(\mathbf{O}|\lambda^s) \quad \text{pro} \quad s = 1, 2, ..., L$$

- **verifikace**

$dist_s < dist_{thr}$ .... předpokládaná identita PŘIJATA
$dist_s > dist_{thr}$ .... předpokládaná identita ZAMÍTNUTA

$P(\mathbf{O}|\lambda^s) > P_{thr}$ .... předpokládaná identita PŘIJATA
$P(\mathbf{O}|\lambda^s) < P_{thr}$ .... předpokládaná identita ZAMÍTNUTA

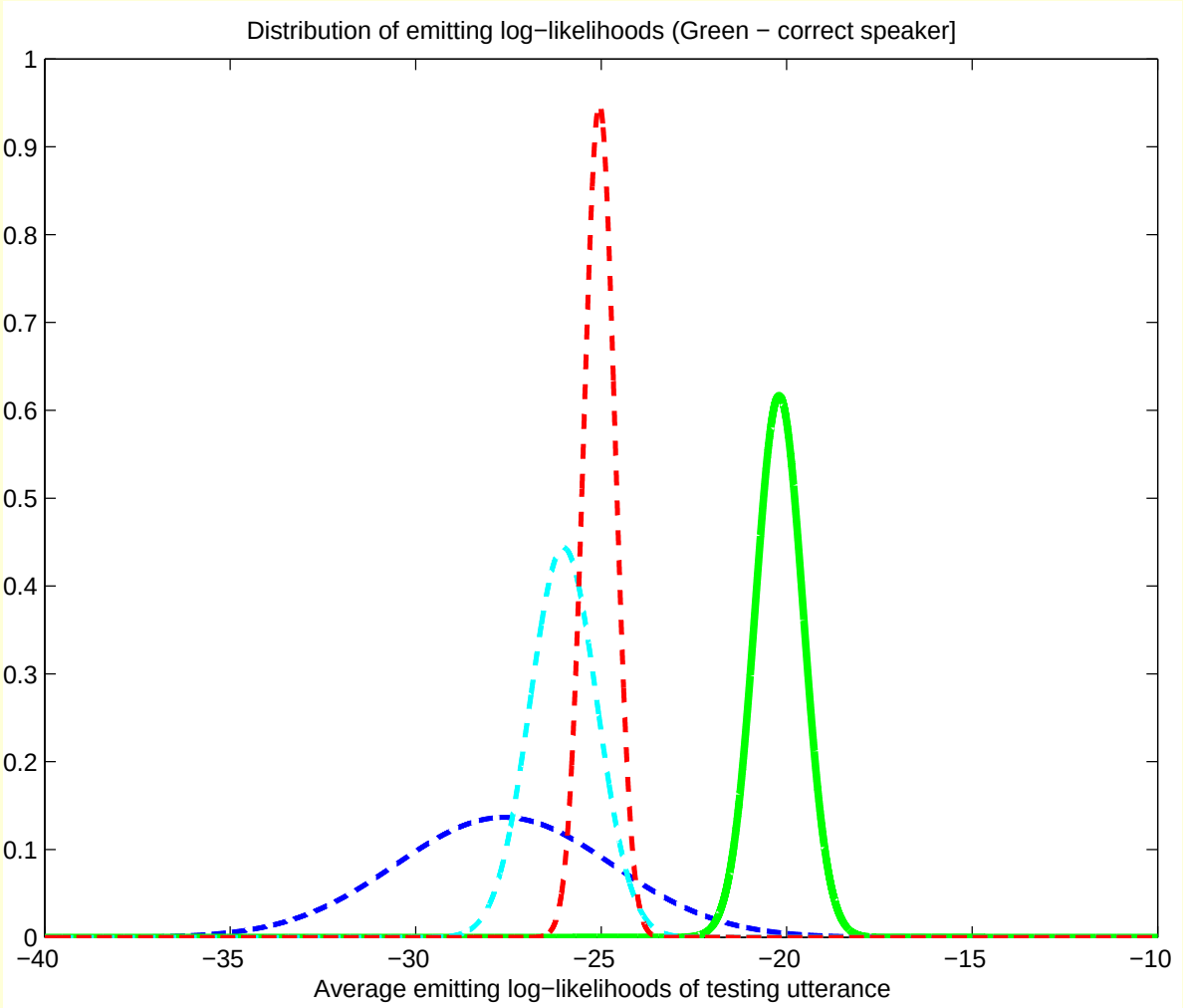
Statistiky výsledků pro 4 řečníky a 1 kódovou knihu



Kódová kniha - zdroj: 12 promluv (12 x 5s), cca 2000 segmentů
- velikost kódové knihy: 200

Testování - 6 promluv (6 x 5s), cca 1200 segmentů

Statistiky výsledků pro 4 řečníky a 1 GMM model



GMM model - zdroj: 12 promluv (12 x 5s), cca 2000 segmentů

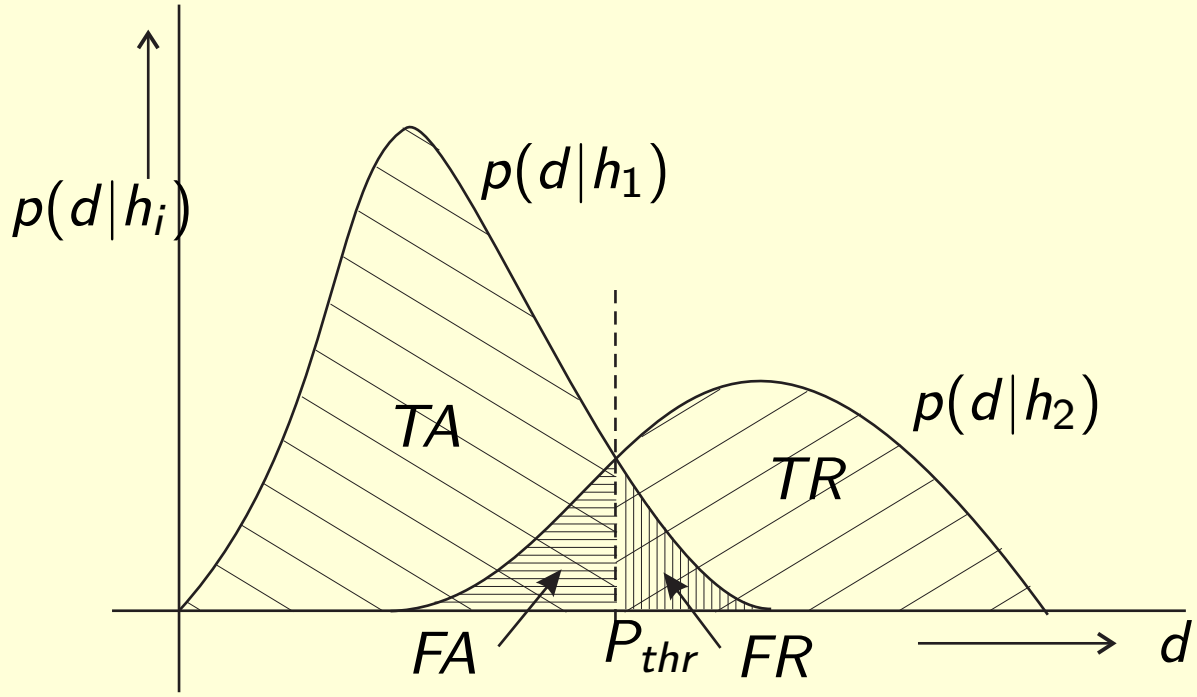
- počet vážených směsí v GMM: 6

Testování - 6 promluv (6 x 5s), cca 1200 segmentů

III. část

Příklady systémů rozpoznávání řečníka

Hodnotící kritéria při verifikaci mluvčího



Vyhodnocování klasifikace:

TA - True acceptance

FA - False acceptance: $R_{FA} = \frac{N_{FA}}{N_{podv}}$

TR - True rejection

FR - False rejection $R_{FR} = \frac{N_{FR}}{N_{spref}}$

EER - Equal Error Rate :

$$R_{EER} = R_{FR}(P_{thr,EER}) = R_{FA}(P_{thr,EER})$$

Příklady systémů

J. R. Campbell: Speaker Recognition. Department of Defense Fort Meade, MD, at http://scgwww.epfl.ch

- přehled různých systémů verifikace

Autor	Příznaky	Metoda	Text	Vstup	Error
Atal'74	kepstr.	Pattern	depend.	LAB	2% (1s)
Fururi'81	nor. kepstr.	Pattern	depend.	TEL	0,2% (3s)
Doddington'85	FB	DTW	depend.	LAB	0,8% (6s)
Tishby'91	kepstr.	HMM	10 digits	TEL	2,8% (1,5s) 0,8% (3,5s)
Reynolds'96	MFCC+Δ	GMM	indep.	TEL match.	11% (3s) 6% (10s) 3% (30s)
Reynolds'96	MFCC+Δ	GMM	indep.	TEL mism.	16% (3s) 8% (10s) 5% (30s)

Příklady systémů

NIST 2010 - Speaker verification evaluations.

- výsledky verifikace pro rozdílné evaluační podmínky
- GMM-UBM systémy (UBM - Universal Background Model)
- EER - Equal Error Rate

	mic-mic	mic-mic2	mic-tel	tel-tel
Systém 1 - muži	8,39	17,29	16,24	15,68
Systém 1 - ženy	13,5	23,47	18,42	17,18
Systém 1 - AVG	10,94	20,38	17,54	16,52
Systém 2	6,00	8,64	5,32	5,11

Systém 1 - 8kHz, 25/10 ms, preemfáze, 16 MFCC ($+\Delta$, $+\Delta\Delta$), log energie, energetický VAD, normalizace příznakových vektorů, 512 směsí

Systém 2 - 8kHz, 25/10 ms, 19 MFCC & $c[0]$ ($+\Delta$), detektor řeči na bázi automatického přepisu (rozpoznávání), normalizace příznakových vektorů, adaptace akustických modelů, 512 směsí

Příklady systémů

Vlasta Radová: Rozpoznávání řečníka. ZČU Plzeň. Prosinec 2004.

Komplexní systém využívající kombinované příznaky a víceúrovňové rozhodování

- textově závislá verifikace
  - EER 0.5 % - vstup z mikrofonu (16 kHz)
  - EER 2 % - vstup z telefonní linky (8 kHz)

- textově nezávislá verifikace
  - EER 2 % - vstup z mikrofonu (16 kHz)
  - EER 10 % - vstup z telefonní linky (8 kHz)

Děkuji vám za pozornost !