Parametrizace řeči

Lukáš Burget ÚPGM FIT VUT Brno, burget@fit.vutbr.cz

FIT VUT Brno

Plán

- Cíle parametrizace
- Mel frekvenční kepstrální koeficienty
- Parametrizace odvozená na datech
- Dynamické koeficienty
- Temporální filtrace

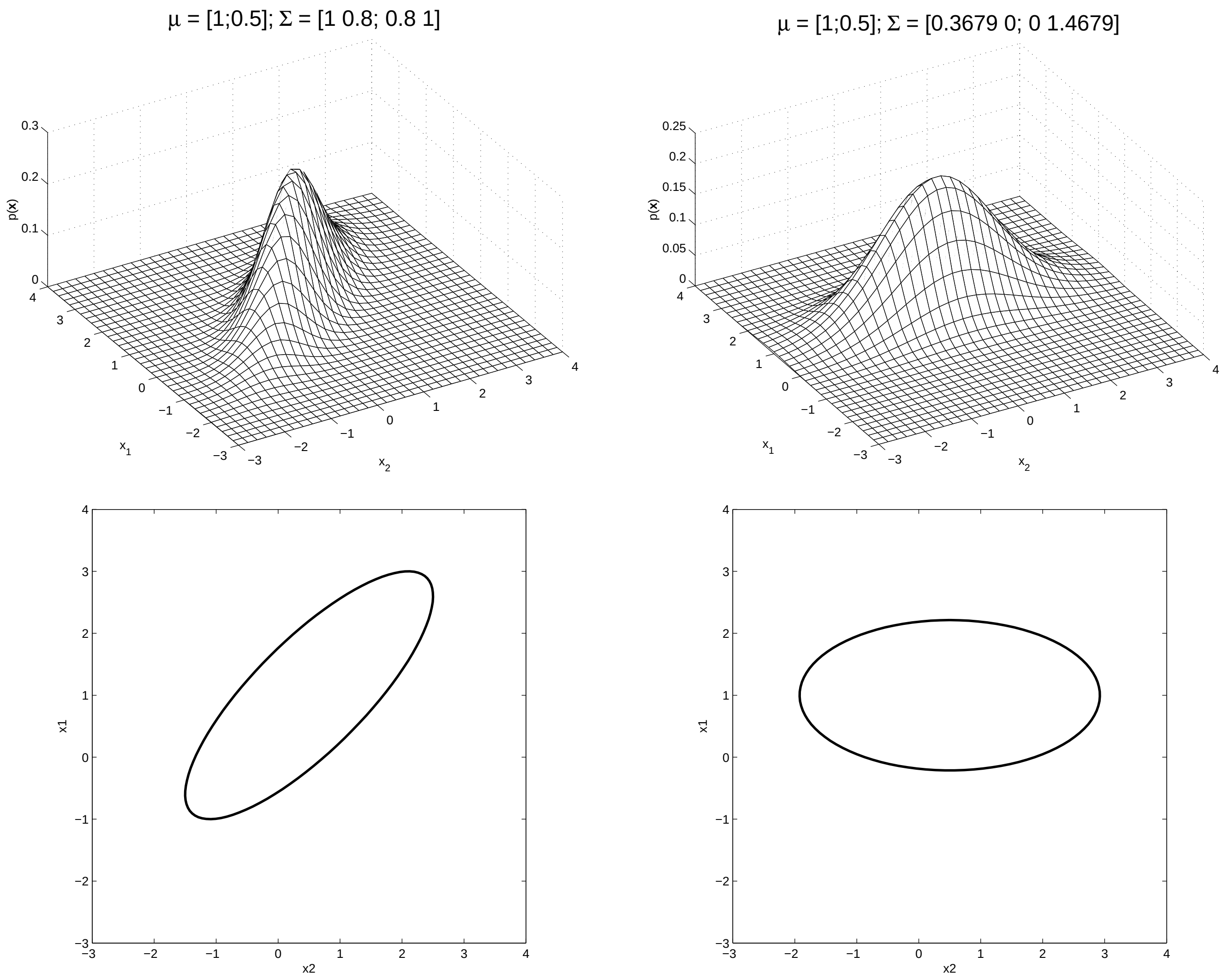
Cíl parametrizace

Přizpůsobení vstupních dat potřebám rozpoznávače

Rozpoznávače mají rády parametry, které jsou

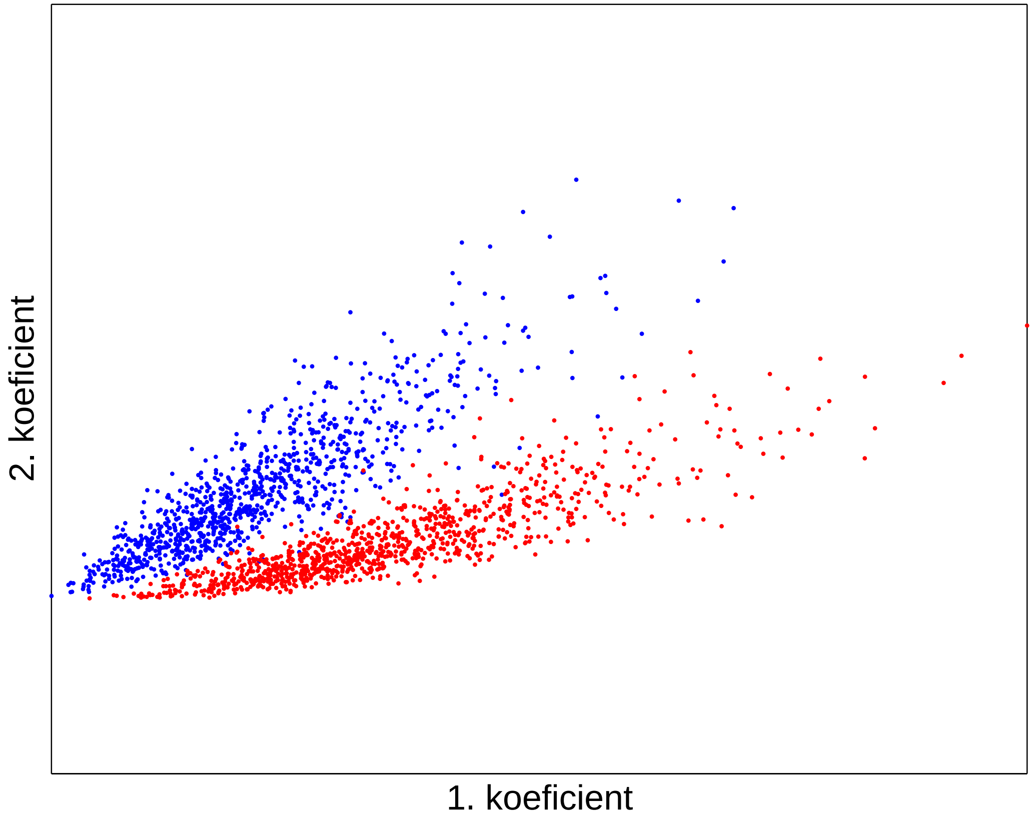
1. gaussovského rozloženi (většinou vícerozměrného),
2. dekorelované,
3. málo dimenzionální.

Dvourozměrné gaussovské rozložení



Příklad parametrizace pro 2D vstupní vektory

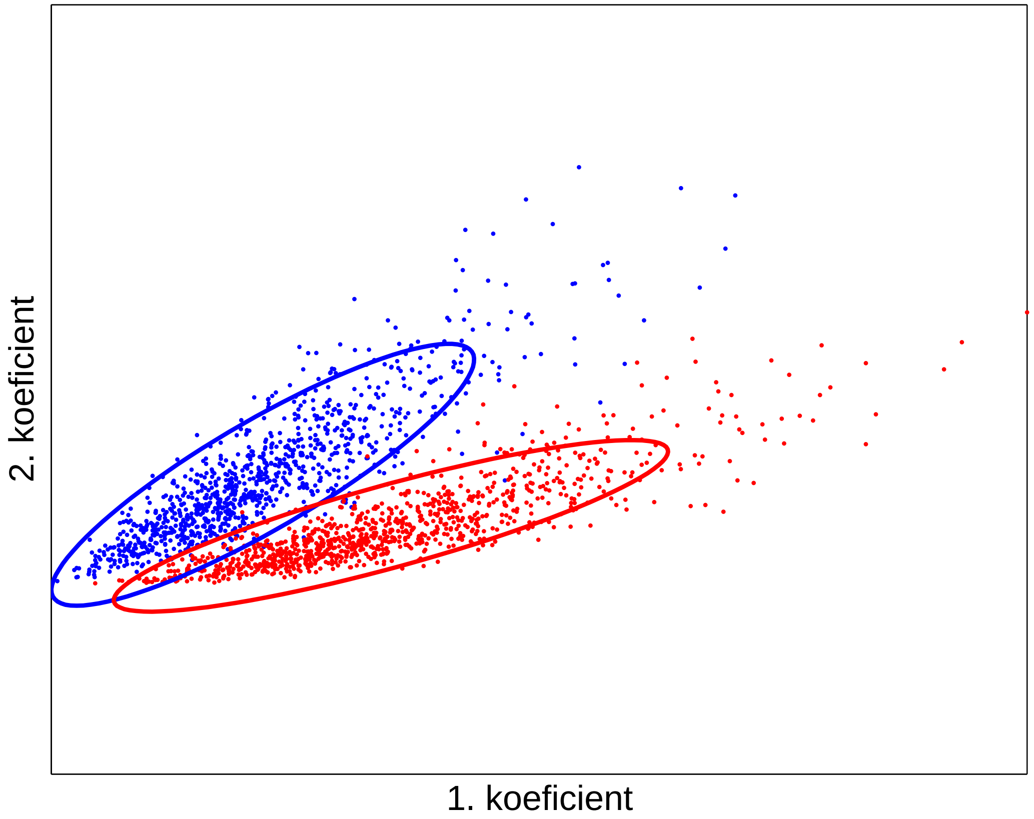
Mnějme vzorky (příklady) 2D rozložení pro dvě třídy.



Příklad parametrizace pro 2D vstupní vektory

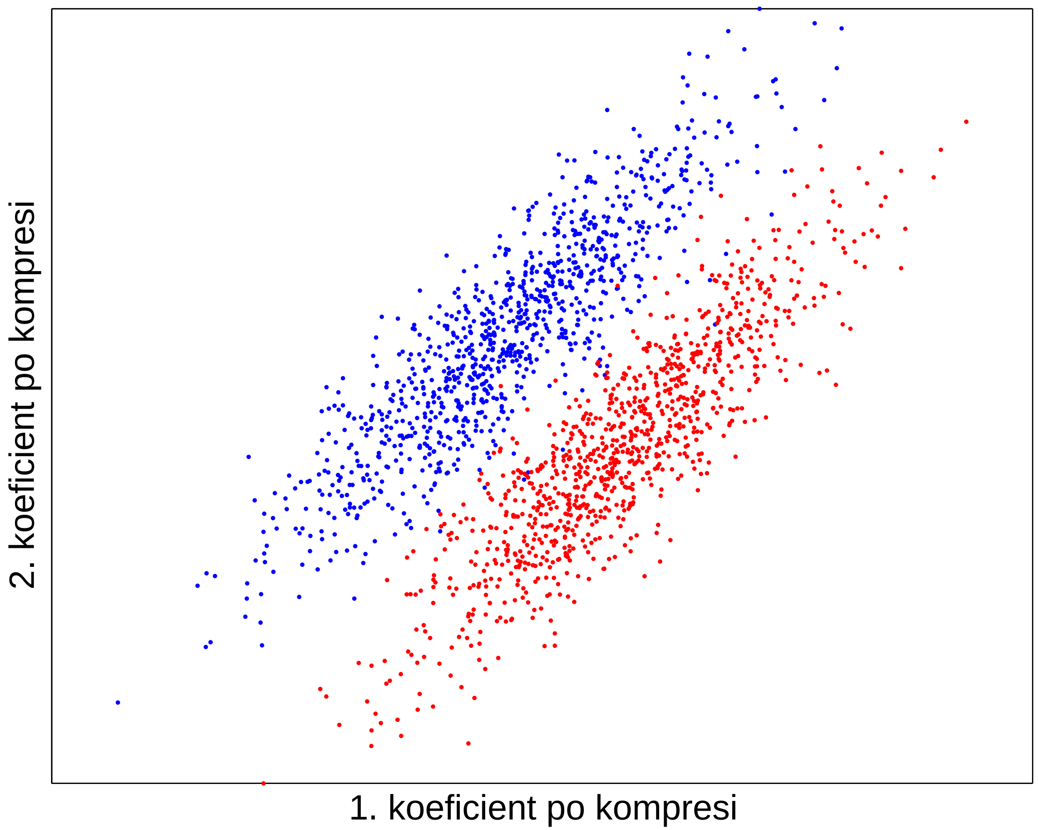
Rozložení není příliš gausovské.

Provedeme třetí odmocninou obou koeficientů.



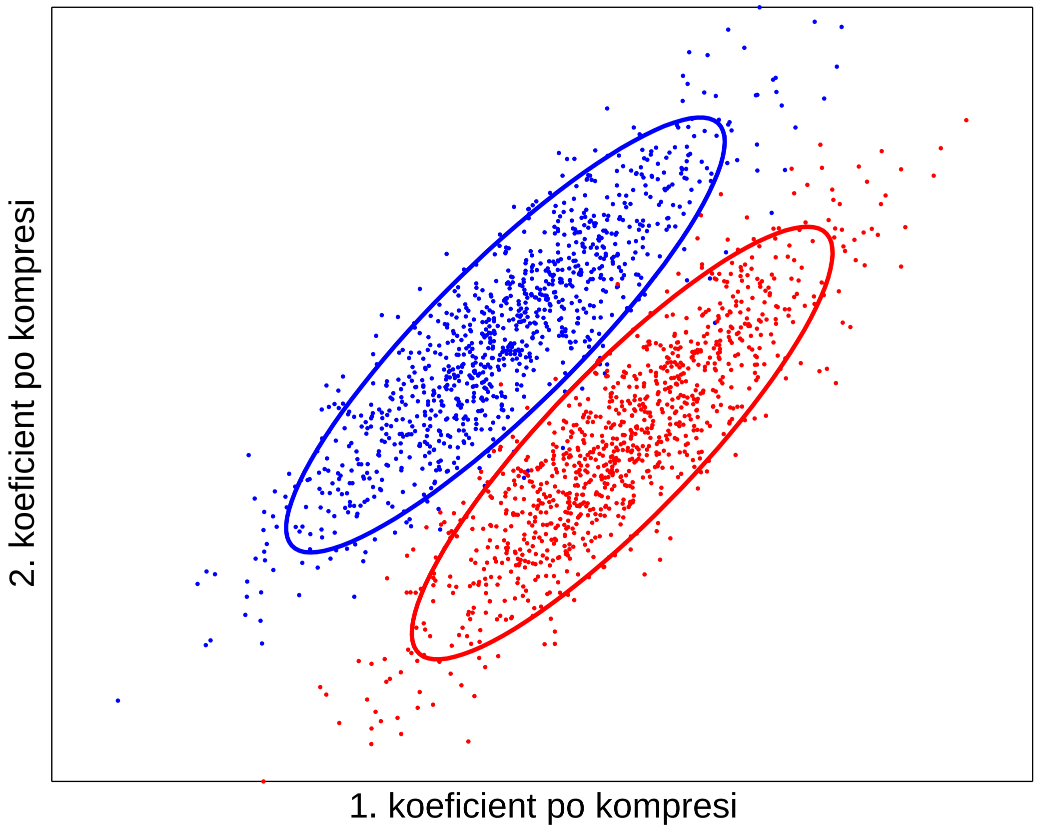
Příklad parametrizace pro 2D vstupní vektory po kompresi

Prostor se komprimuje - nelineárně deformuje...



Příklad parametrizace pro 2D vstupní vektory po kompresi

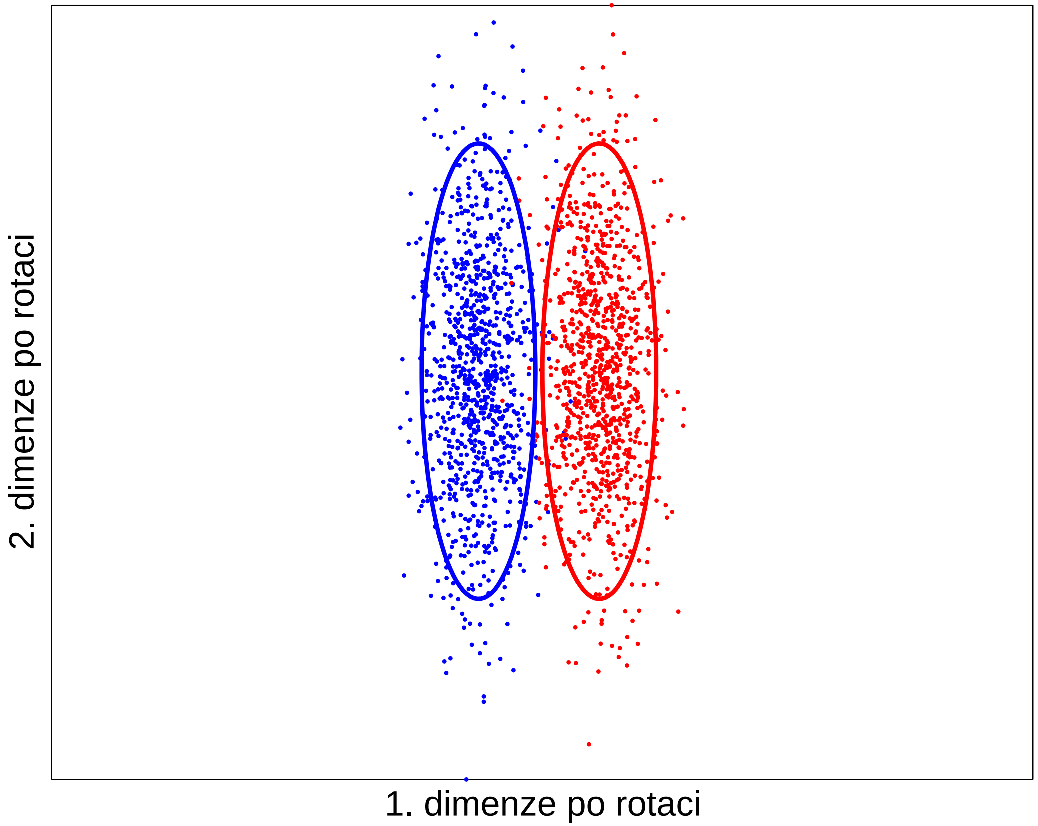
... a rozložení pro každou třídu je nyní gausovské.
Koeficienty jsou ale korelované.
Je vhodné prostor otočit tak aby se koeficienty dekorelovaly.



Příklad parametrizace pro 2D vstupní vektory po dekorelaci

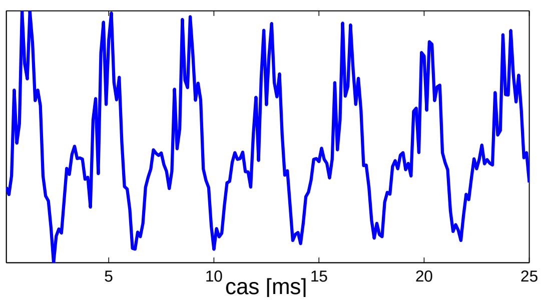
Nyní jsou koeficienty dekorelovány.

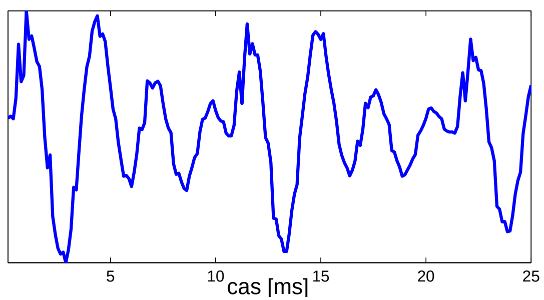
Svislá dimenze je naví zbytečná, protože třídy se v ní zcela překrývají.



MFCC - Segmentace

Řečový signál je nejprve "rozsekán" do segmentů o délce cca 25ms a každy takový segment je pak zpracován nezávisle. Jednotlive kroky MFCC parametrizace si budeme demonstrovat na dvou takových segmentech. Oba segmenty představují foném 'iy' ve slově 'three', jeden je však vysloven ženou a druhý mužem.



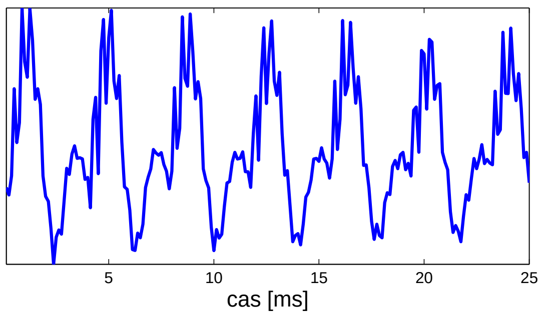


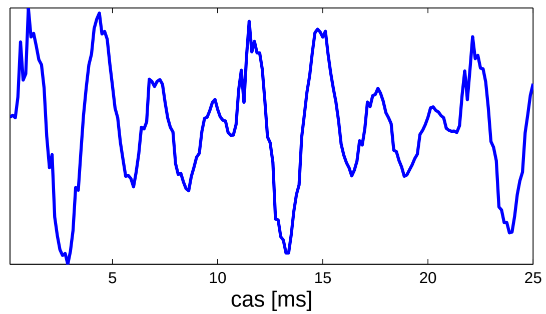
MFCC - Preemfáze

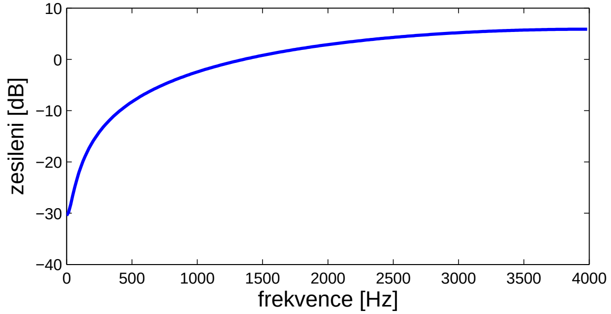
Zvuky o stejné enegii na různých kmitočtech člověk slyší různě hlasitě.

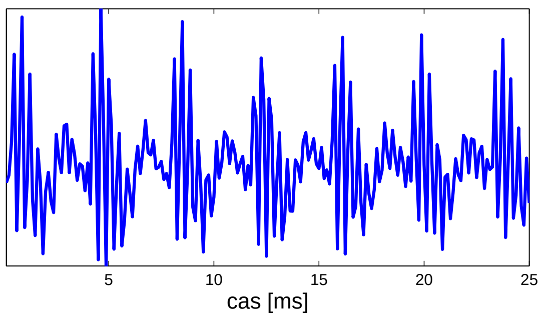
(viz sin_20_200_2000_Hz.wav)

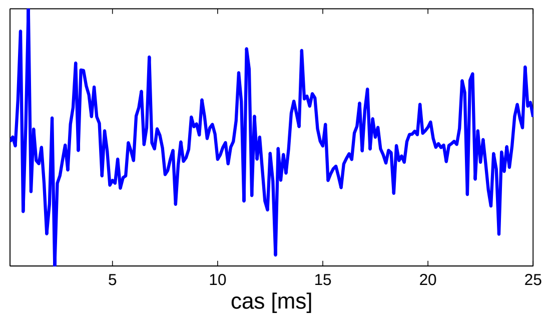
Proto se signál filtruje horni propusti.





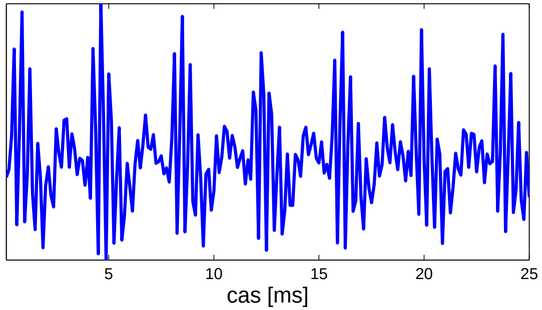


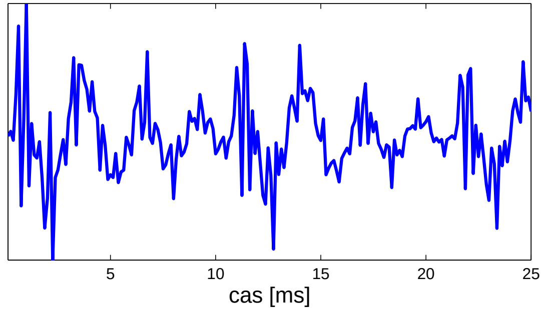


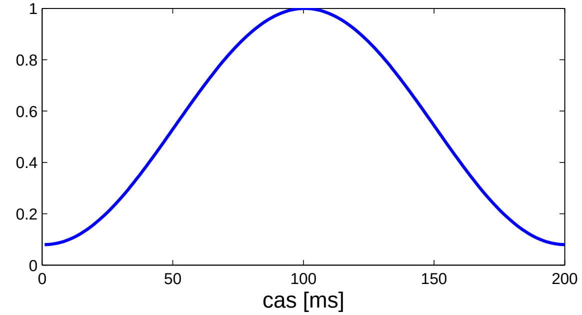


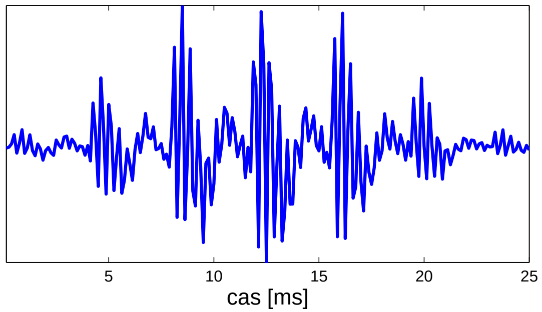
MFCC - Váhování hammingovým oknem

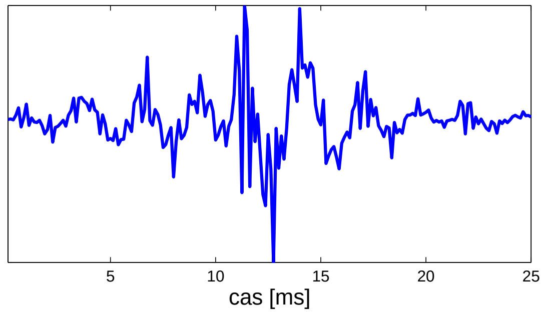
Pro jednotlivé segmenty řeči budeme chtít sočítat jejich spektra.
Diskontinuity na krajích segmentu vedou ke zkresleni (zašumění spekter). To lze kompenzovat váhováním segmentu např. hammingovým oknem.



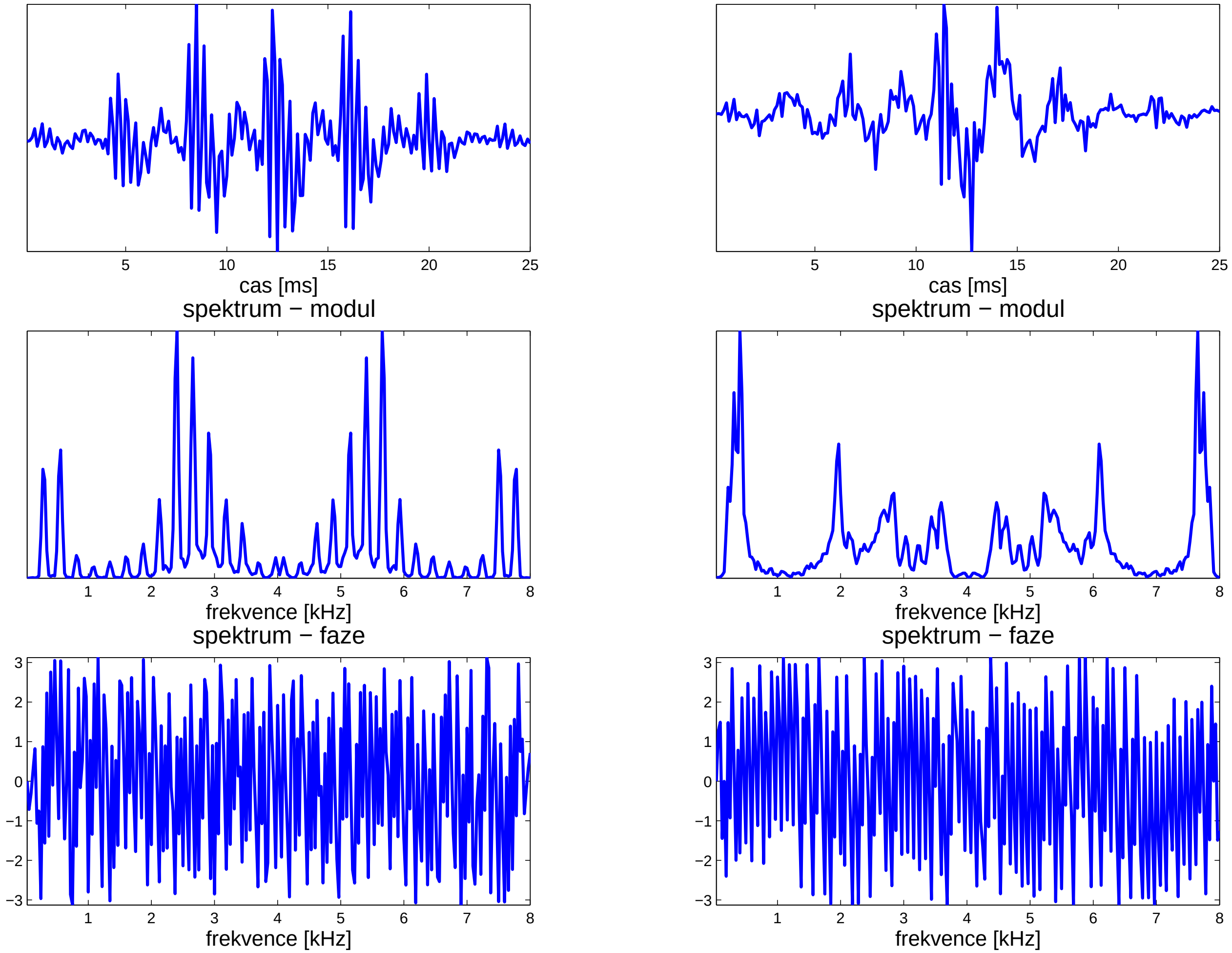






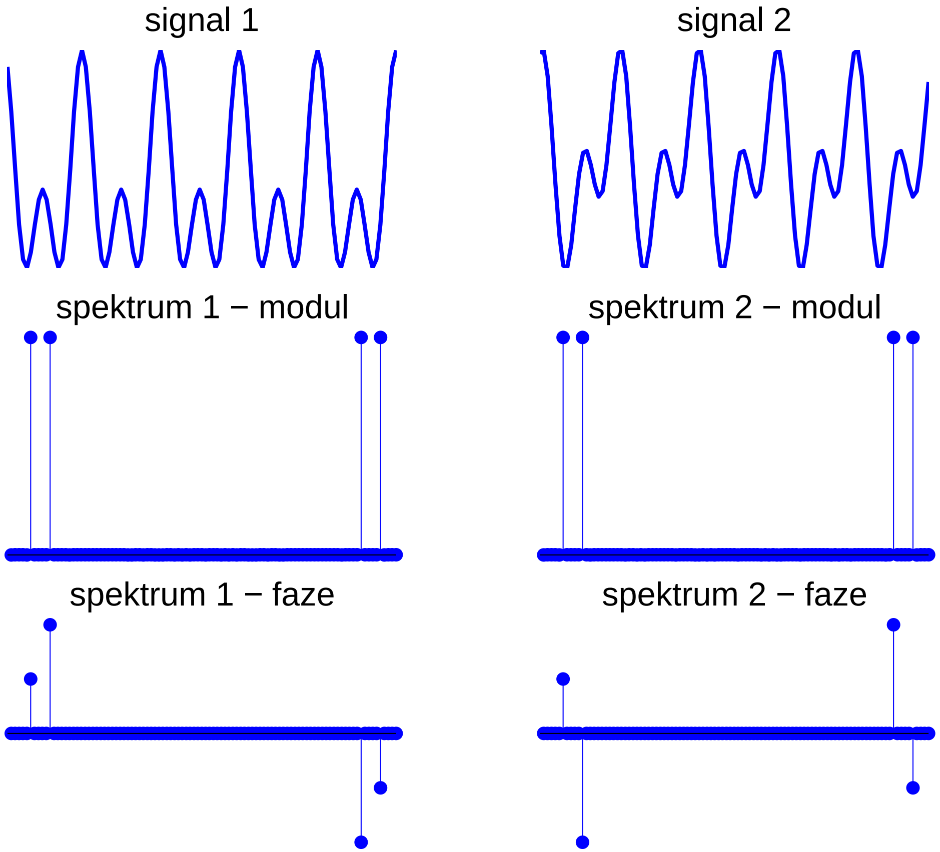


MFCC - Spektrum



Součet dvou sinusovek pro různé fázové posuny

Dva různé součty sinusovek o stejne amplitudě a rekvencích 400 a 800 Hz pro různé vzájemné posunutí obou sinusovek. (viz. different_phase.wav)

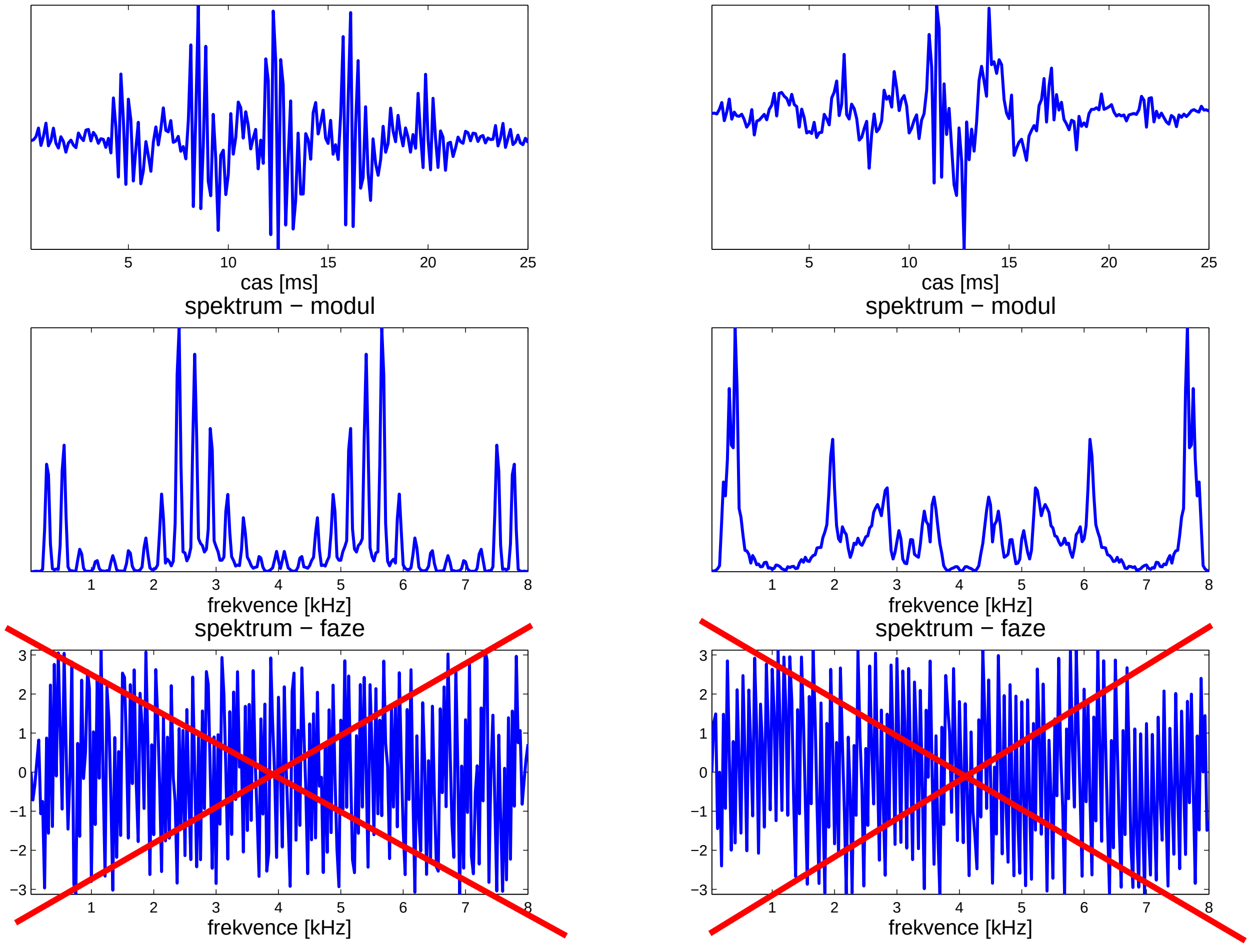


Člověk příliš neslyší rozdily ve fázy.

MFCC - Spektrum

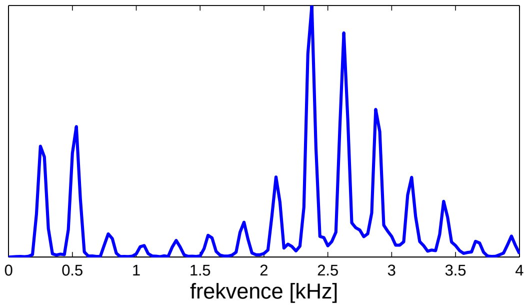
Proto informaci o fázi zahodíme

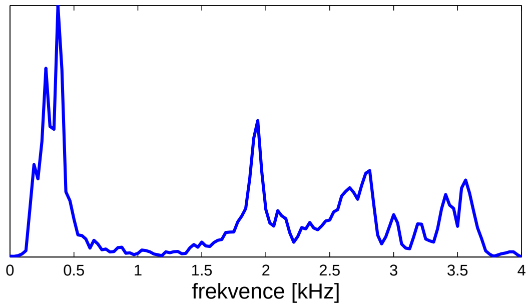
a také budeme dále používat pouze polovinu simetrického spektra

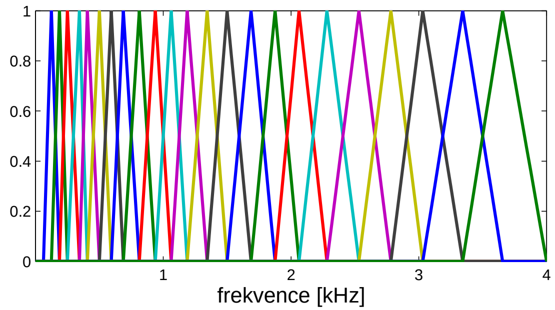


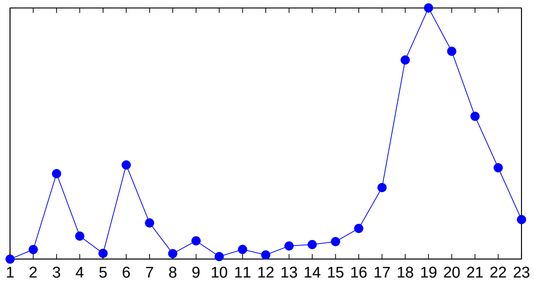
MFCC - Mel banka filtru

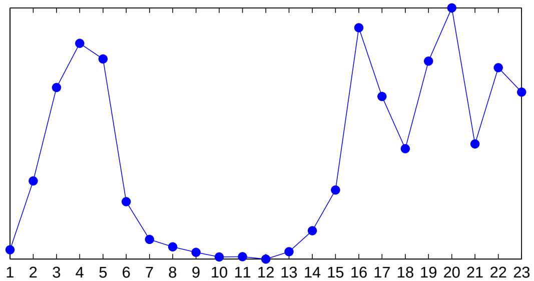
Části spekter váhujeme a integrujeme pomocí trojůhelníkových oken, které se s rostoucí frekvencí rozšiřují. Tím odstraníme jemnou strukturu spektra nesoucí informaci o nedůležitém základním tónu, aplikujeme znalost o různé schopnosti lidí rozlišovat frekvence v různé části spektra (viz sin_100_110_1000_1010_Hz.wav) a redukujeme počet koeficientů (z 129 na 23).





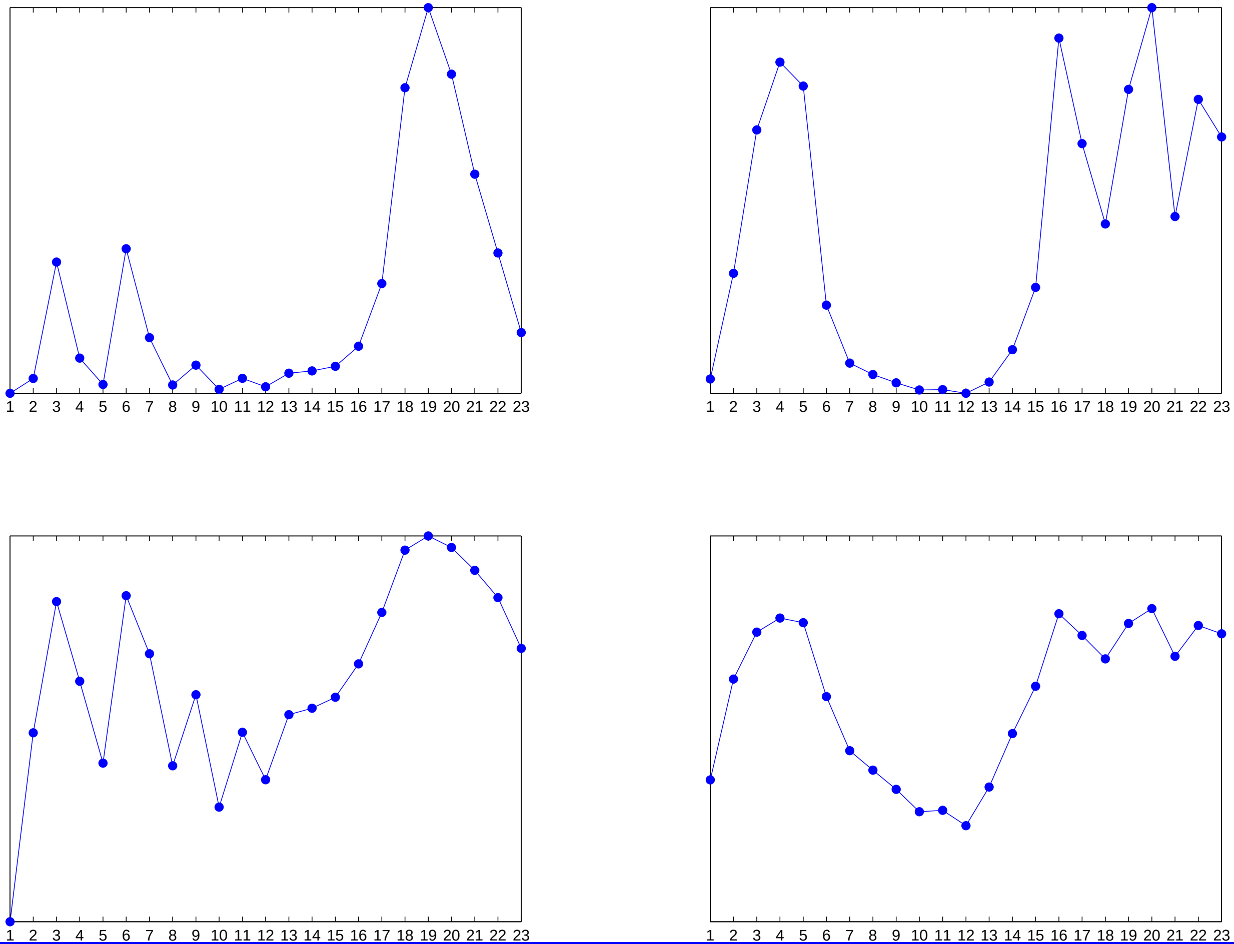






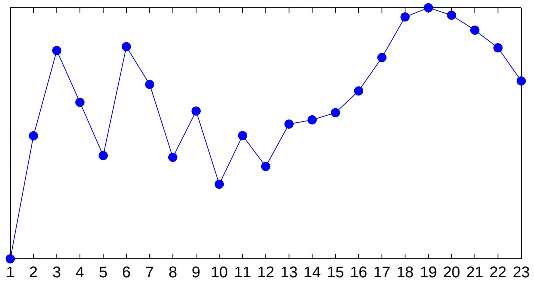
MFCC - Komprese

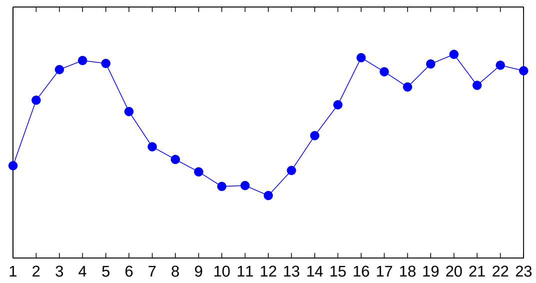
Výstup z banky filtrů je logaritmován.
To odráží logaritmické vnímání hlasitosti lidmi
a vede na “gaussovštější” rozložení koeficientů.

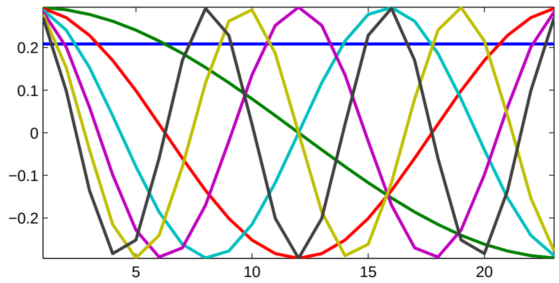


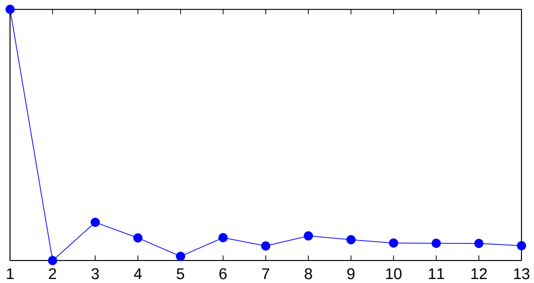
MFCC - Dekorelace

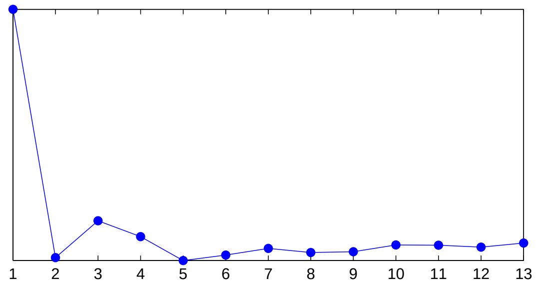
Logaritmický výstup z banky filtrů je promitnut do cosinových bází,
čímž se jednotlivé koeficienty dekorelují.
Proč? Uvidíme za chvíli



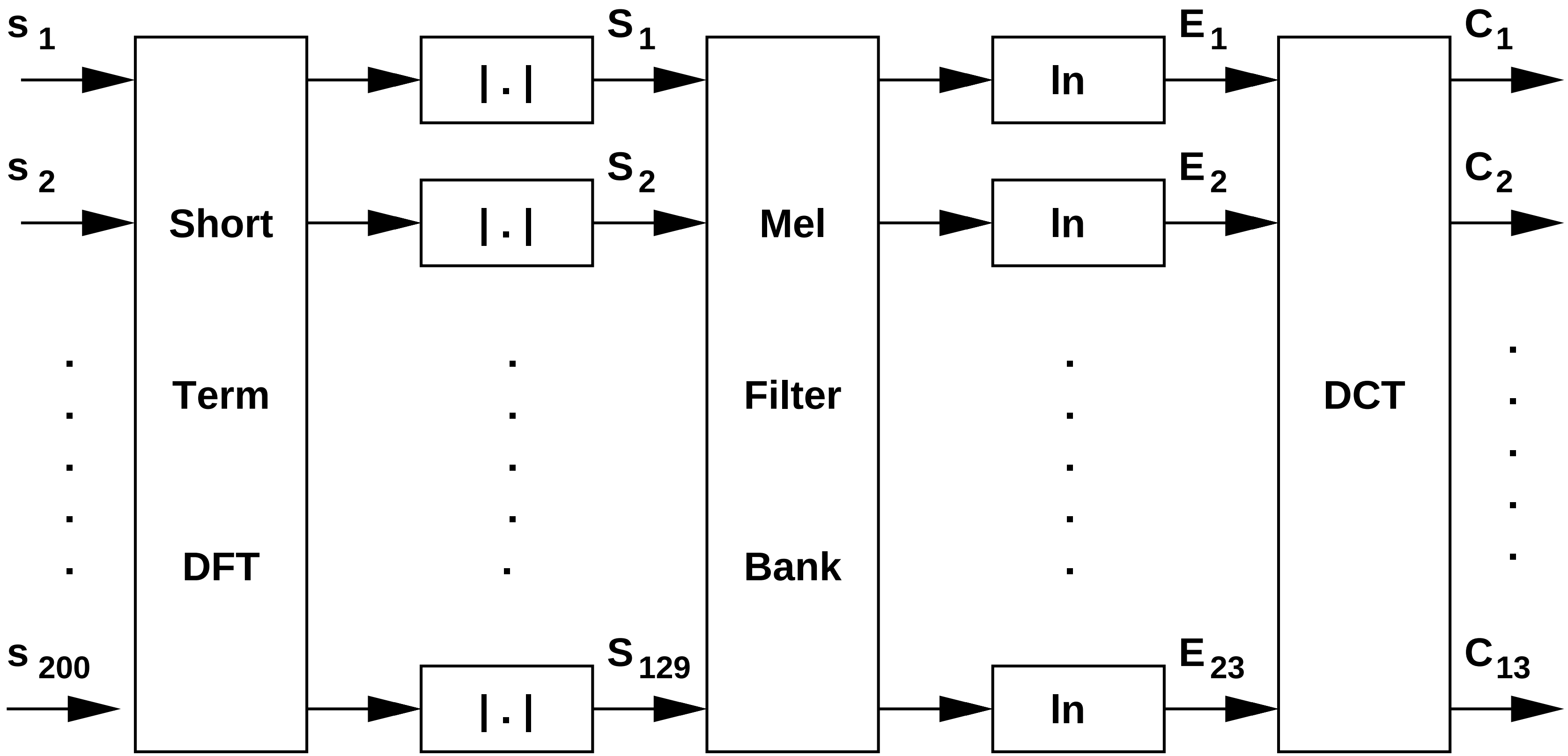




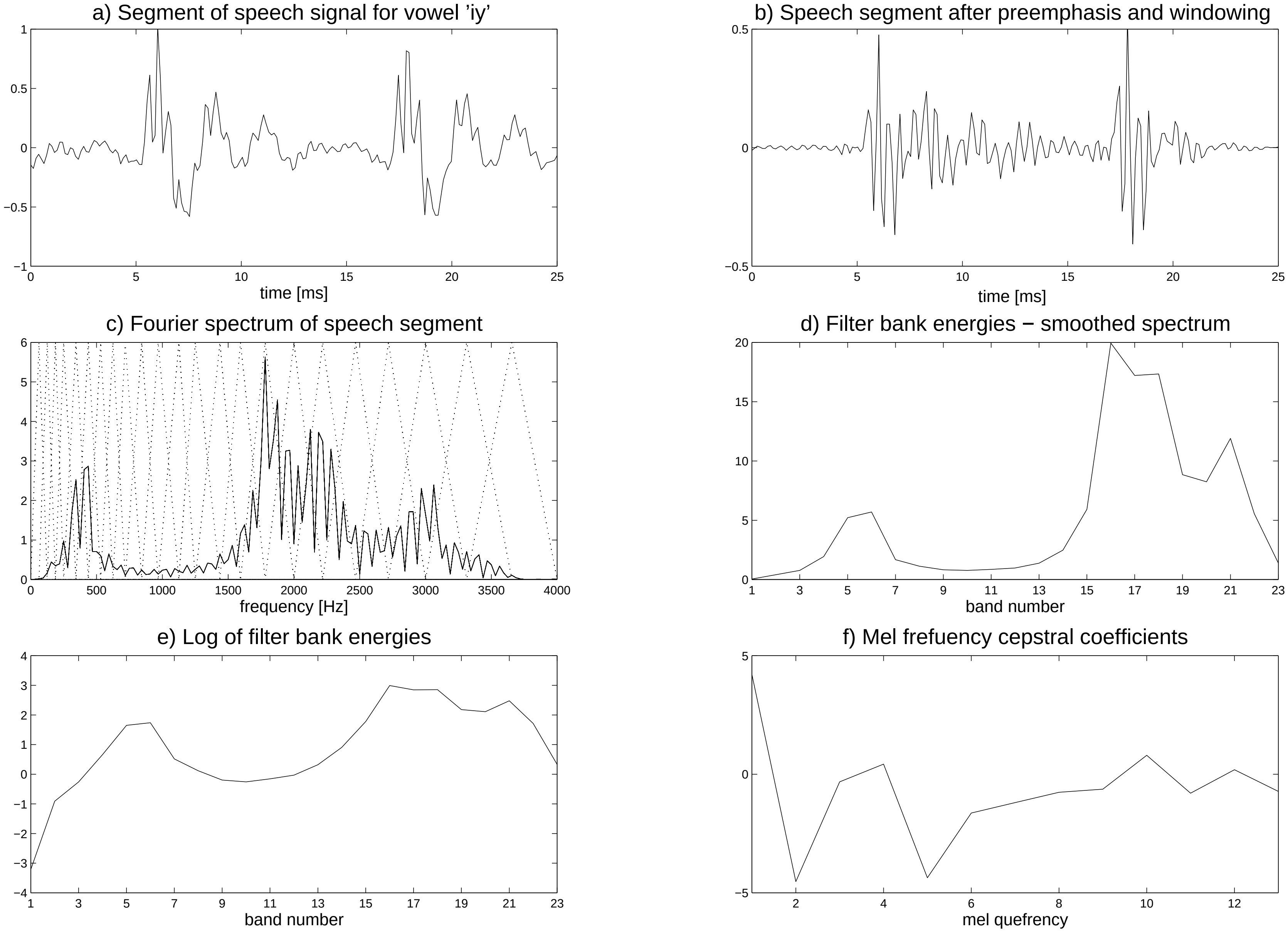




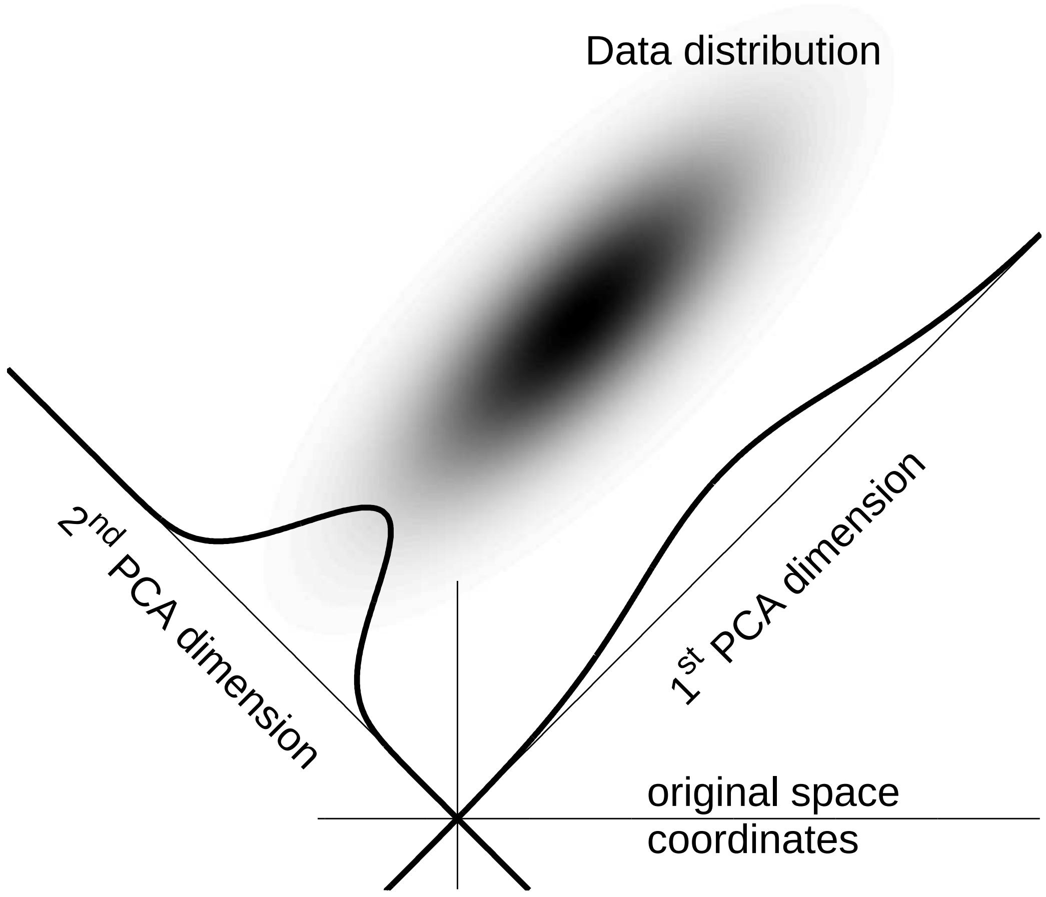
MFCC - Blokové schéma



MFCC - Jednotlivé kroky



Principal Component Analysis (PCA) neboli Karhunen-Loevyho transformace (KLT)



Principal Component Analysis (PCA)

Báze - směry PCA jsou dány vlastními vektory kovarianční matice.
Vlastní čísla udávají variance (důležitosti) v těcho směrech.

$$\mathbf{\Sigma}\mathbf{e}_i = \mathbf{e}_i\lambda$$

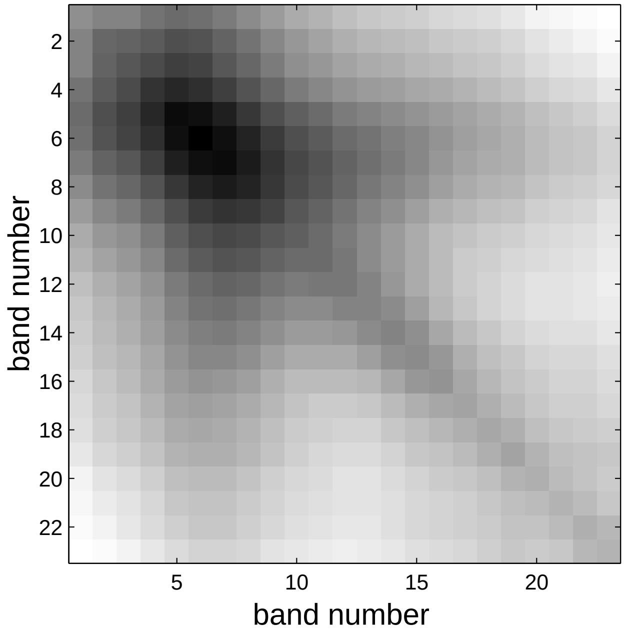
$$\mathbf{\Sigma} = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T]$$

$\mathbf{\Sigma}$	-	kovarianční matice
μ	-	vektor středních hodnot
$\mathbf{e}_i$	-	vlastní vektory
λ	-	vlastní čísla

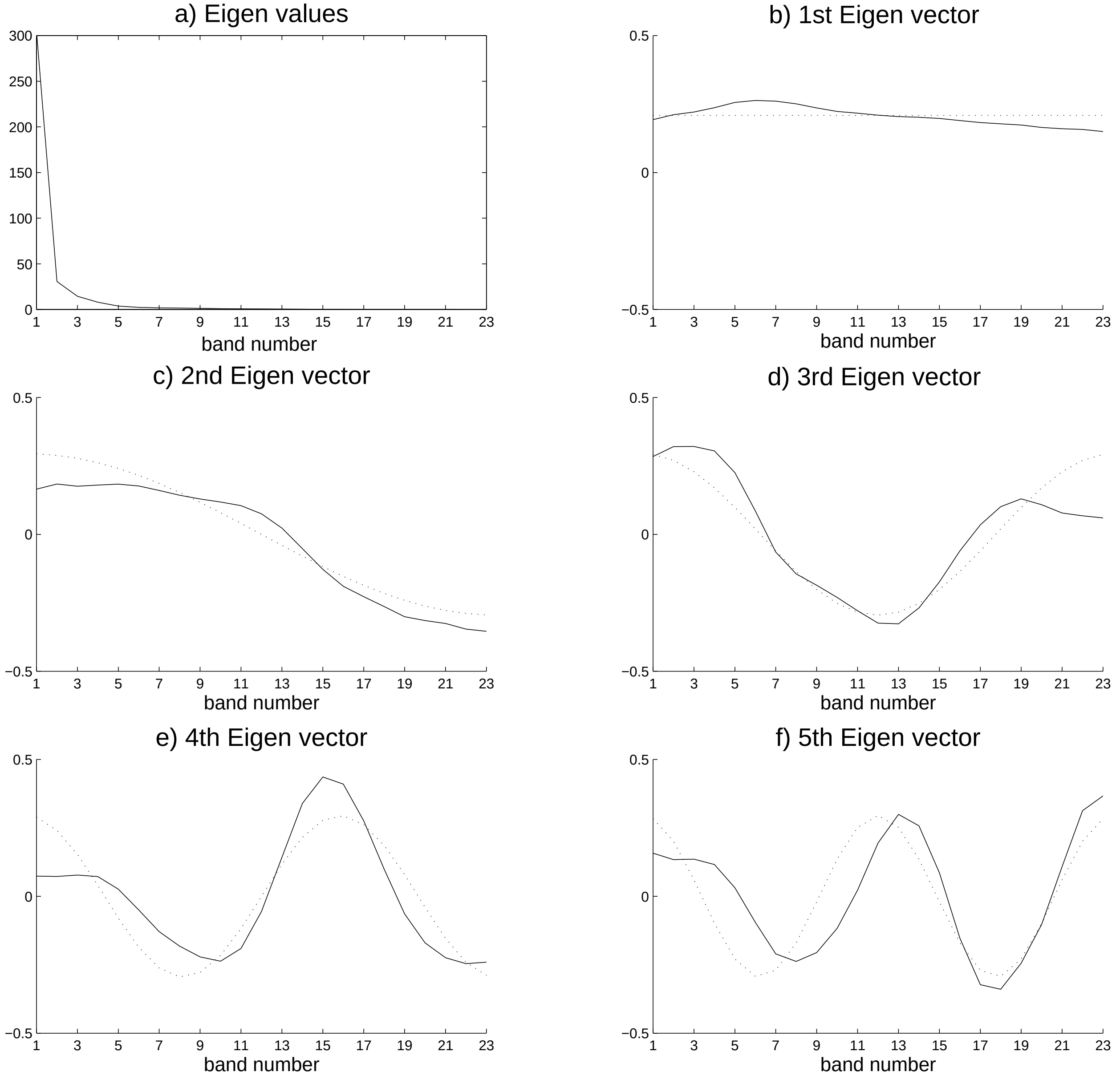
Spektrální báze odvozené pomocí PCA

Abychom u MFCC mohli nahradit DCT za “optimaln” trasformaci odvozenou pomoci PCA, odhadneme kovarianční matici pro vektory představující logaritmický výstup z banky filtrů a pro tuto kovarianční matici spočteme vlastní čísla a vlastní vektory.

Za účelem redukce dimenzí provedeme prijekci pouze do nekolika vlastních vektorů (PCA bází), které odpovídají největším vlastním číslům.

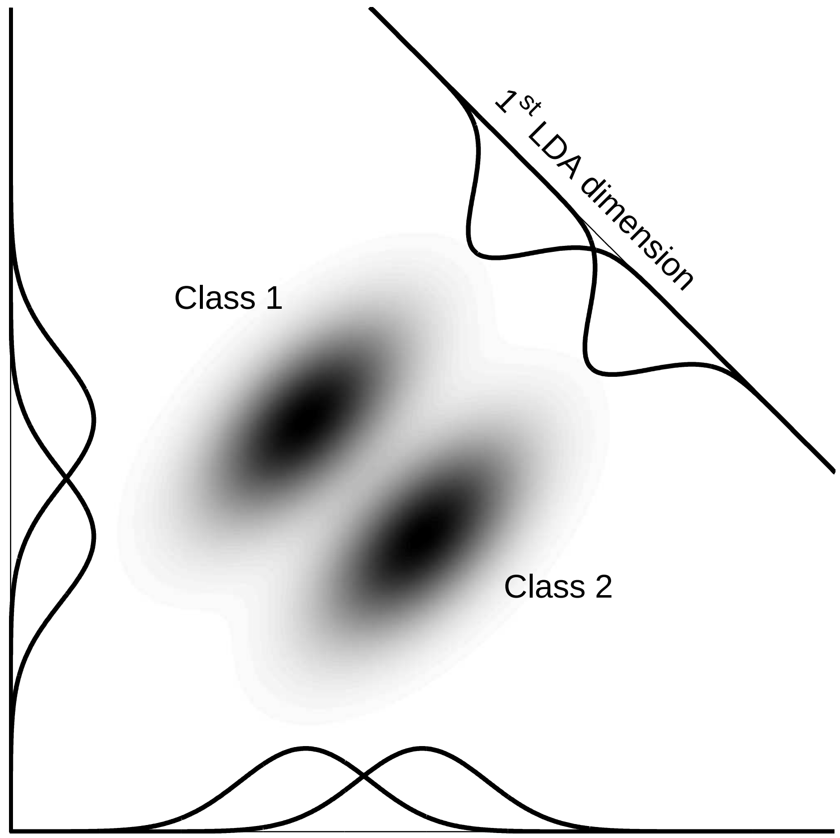


Spektrální báze odvozené pomocí PCA



Linear Discriminant Analysis (LDA)

PCA předpokládá, že důležitost směrů je dána poze variancí a počítá pouze s celkovým rozložením hustoty pravděpodobnosti. Nebere v úvahu třídy. My však potřebujeme především rozlišovat (a dekorelovat) data náležející jednotlivým třídám. Tento problém řeší LDA.



Linear Discriminant Analysis (LDA)

$$[\mathbf{\Sigma}_{wc}^{-1}\mathbf{\Sigma}_{ac}]\mathbf{e}_i = \lambda\mathbf{e}_i$$

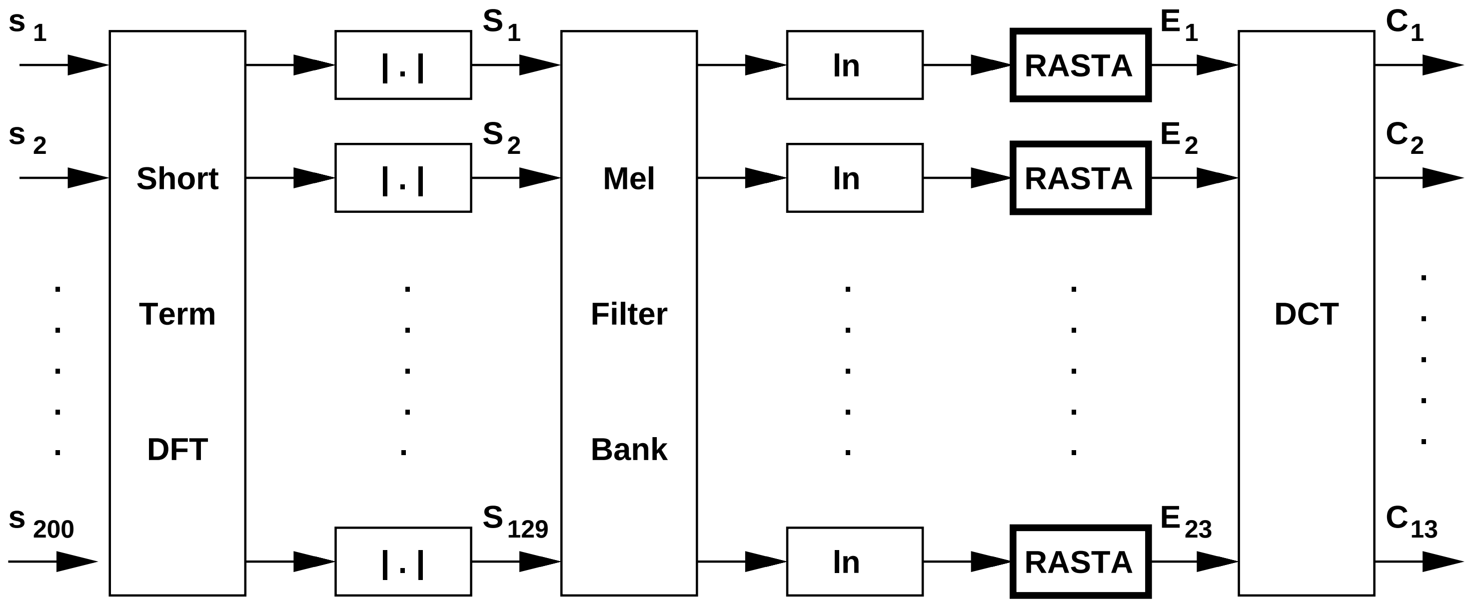
$$\mathbf{\Sigma}_{ac} = E[(\boldsymbol{\mu}_p - \boldsymbol{\mu})(\boldsymbol{\mu}_p - \boldsymbol{\mu})^T]$$

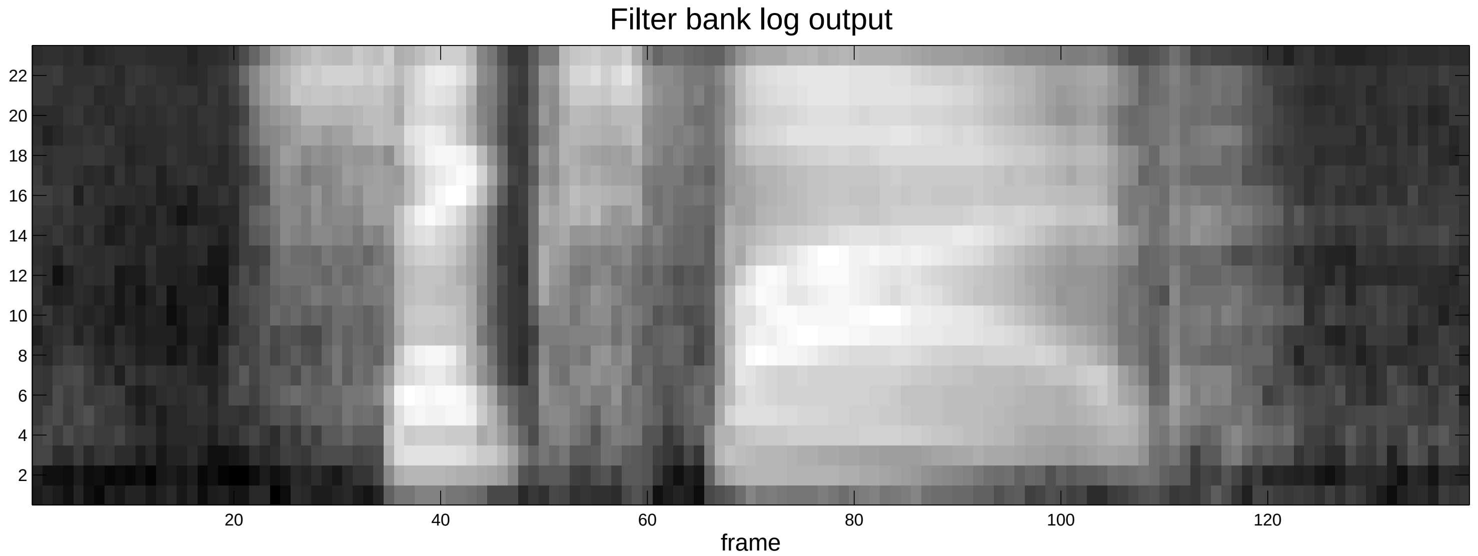
$$\mathbf{\Sigma}_{wc} = E[\mathbf{\Sigma}_p]$$

$\mathbf{\Sigma}_{ac}$, $\mathbf{\Sigma}_{wc}$	-	mezitřídní a vnitrotřídní kovarianční matice
$\mathbf{\Sigma}_p$, μ_p	-	kovarianční matice a střední vektor třídy p
$\mathbf{e}_i$, λ	-	vlastní vektory a vlastní čísla

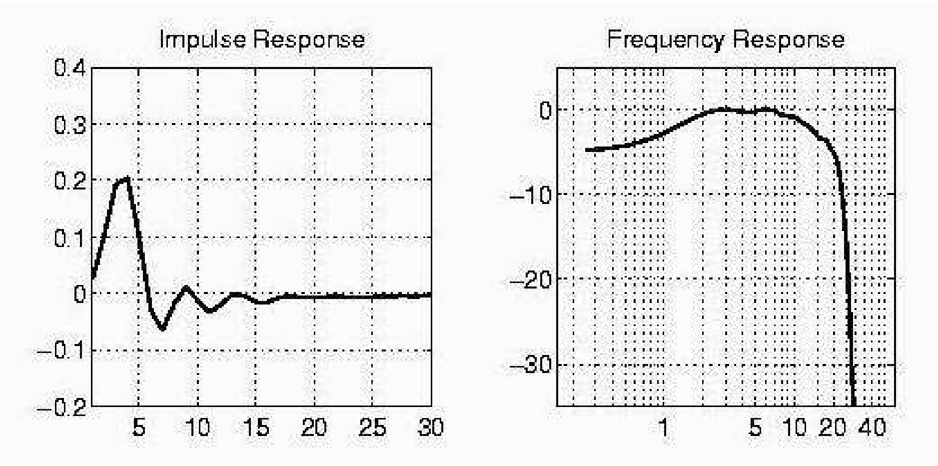
Nedostatek: LDA předpokládá, že všechny třídy mají stejnou kovarianční matici. Tento dedostatek řeší Heteroscedastická linearní diskriminační analýza (HLDA).

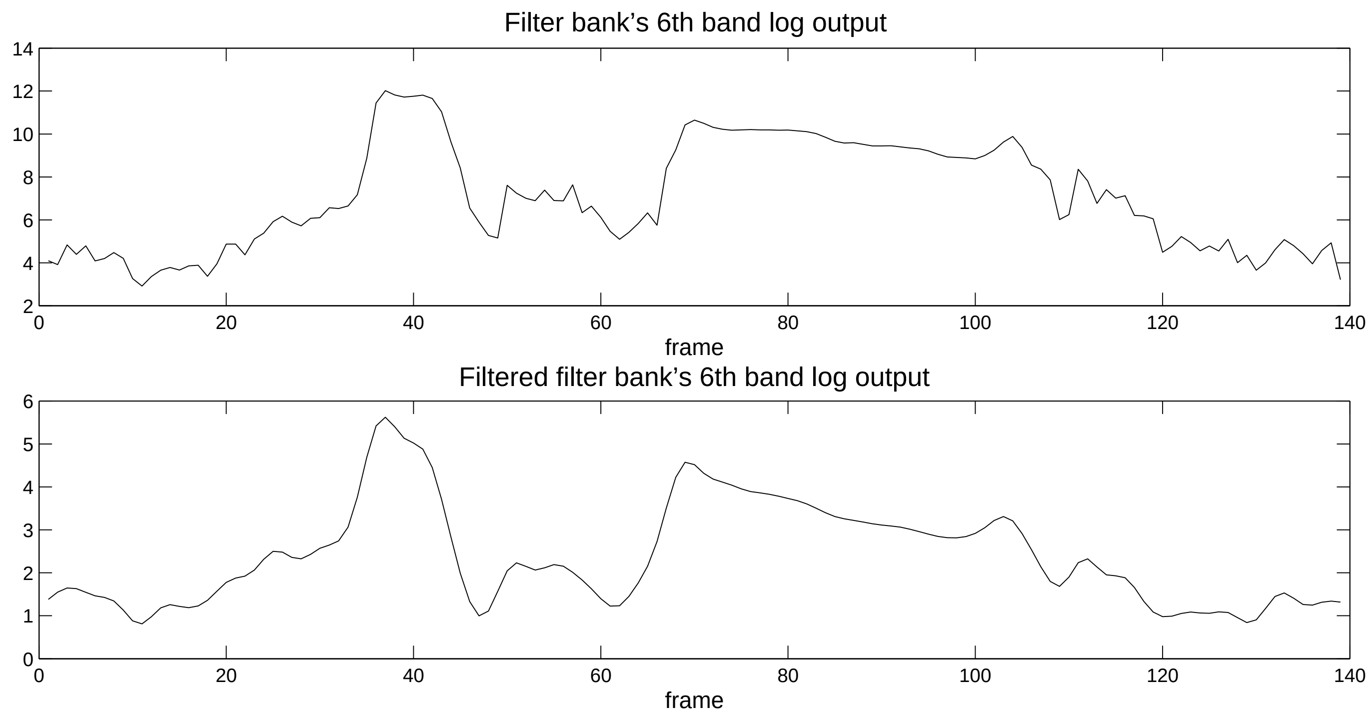
Temporální filtrace



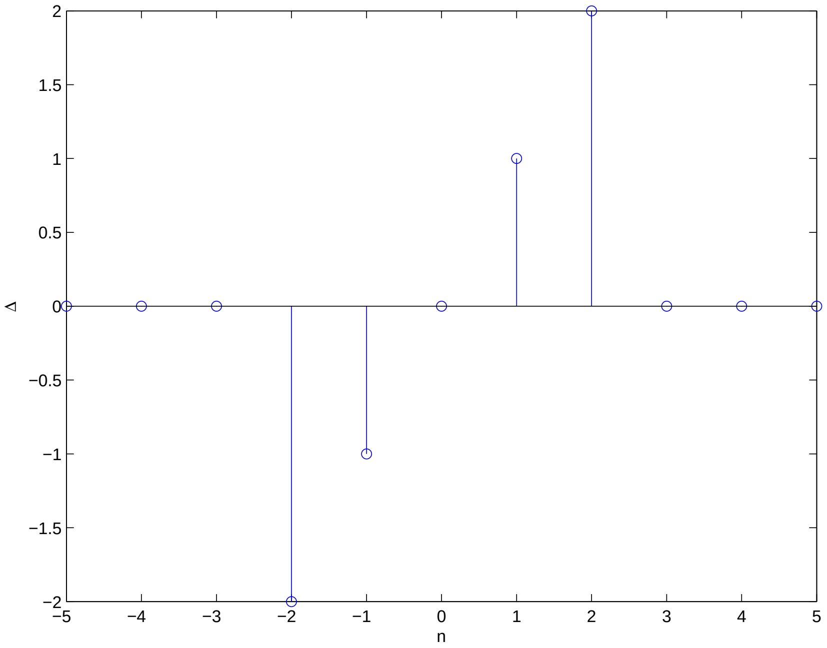


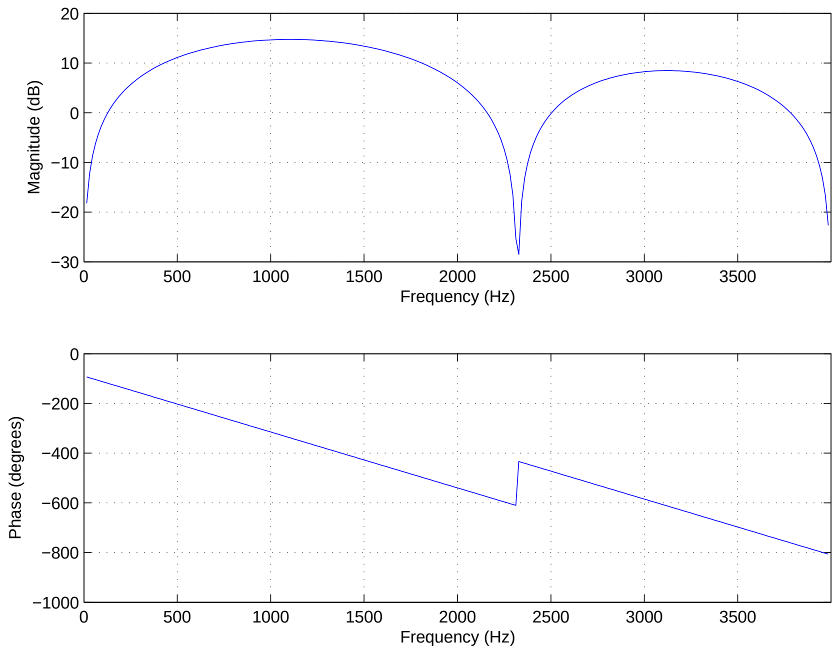
Temporální filtrace





Dynamické koeficienty





Dynamické koeficienty