Text-Independent Speaker Recognition Using Gaussian Mixture Models

Eduardo Martins Barros de Albuquerque Tenório

Centro de Informática
Universidade Federal de Pernambuco
Trabalho de Graduação em Engenharia da Computação

embat@cin.ufpe.br

Recife, 25 de Junho de 2015

Conteúdo

Conteúdo

Reconhecimento de ...

Fala **O que** está sendo dito

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

- Identificar uma pessoa num grupo

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

- Identificar uma pessoa num grupo
- Autenticar um usuário

Reconhecimento de ...

Fala **O que** está sendo dito

- Conteúdo da mensagem
- Estado emocional do locutor
- Sotaque ou dificuldade de articulação

Locutor **Quem** está falando

- Identificar uma pessoa num grupo
- Autenticar um usuário

Este trabalho é focado em reconhecimento de **locutor**

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Verificação Determina se o locutor é quem diz ser

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Verificação Determina se o locutor é quem diz ser

- 1 para 1

Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Verificação Determina se o locutor é quem diz ser

- 1 para 1
- Problema de **conjunto aberto**

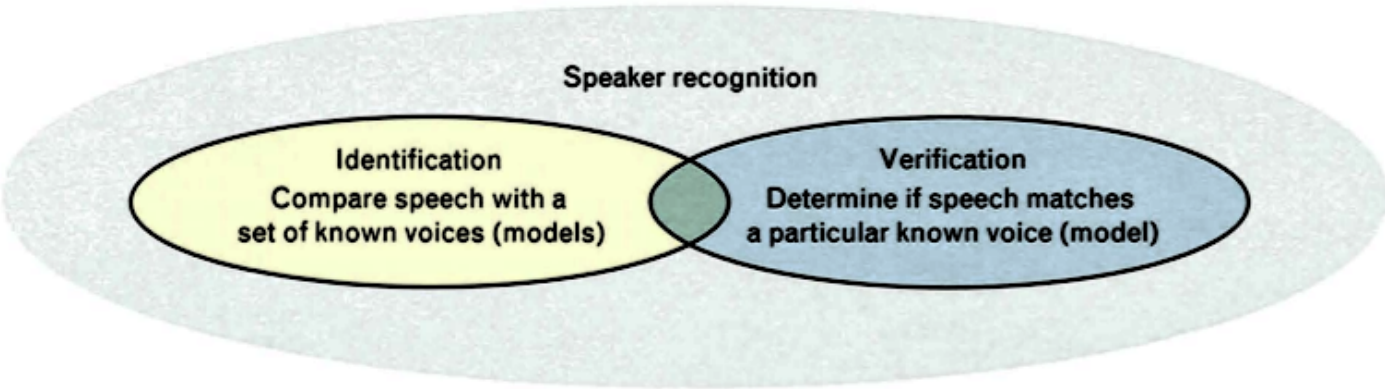
Reconhecimento de Locutor

Identificação Determina a identidade de um locutor dentro de um conjunto não unitário

- 1 para N
- Problema de **conjunto fechado**

Verificação Determina se o locutor é quem diz ser

- 1 para 1
- Problema de **conjunto aberto**



Dependência de texto

Dependente Teste $\in$ Treinamento

Dependência de texto

Dependente Teste $\in$ Treinamento

- Diversos graus de dependência

Dependência de texto

Dependente Teste $\in$ Treinamento

- Diversos graus de dependência
- Teste $\notin$ Treinamento $\Longrightarrow$ Retreinamento

Dependência de texto

Dependente Teste $\in$ Treinamento

- Diversos graus de dependência
- Teste $\notin$ Treinamento $\Longrightarrow$ Retreinamento

Independente Teste $\neq$ Treinamento

Dependência de texto

Dependente Teste $\in$ Treinamento

- Diversos graus de dependência
- Teste $\notin$ Treinamento $\Longrightarrow$ Retreinamento

Independente Teste $\neq$ Treinamento

- Características não textuais

Dependência de texto

Dependente Teste $\in$ Treinamento

- Diversos graus de dependência
- Teste $\notin$ Treinamento $\Longrightarrow$ Retreinamento

Independente Teste $\neq$ Treinamento

- Características não textuais
- Presentes em diferentes sotaques e até *gibberish*

Dependência de texto

Dependente Teste $\in$ Treinamento

- Diversos graus de dependência
- Teste $\notin$ Treinamento $\Longrightarrow$ Retreinamento

Independente Teste $\neq$ Treinamento

- Características não textuais
- Presentes em diferentes sotaques e até *gibberish*

Este trabalho é focado em reconhecimento de locutor **independente de texto**

Modelos de Mistura Gaussiana

GMM **Combinação** de Gaussianas

Modelos de Mistura Gaussiana

GMM **Combinação** de Gaussianas

UBM GMM gerado por diversas **locuções de fundo**

Modelos de Mistura Gaussiana

GMM **Combinação** de Gaussianas

UBM GMM gerado por diversas **locuções de fundo**

AGMM GMM **adaptado** a partir de um UBM

Modelos de Mistura Gaussiana

GMM **Combinação** de Gaussianas

UBM GMM gerado por diversas **locuções de fundo**

AGMM GMM **adaptado** a partir de um UBM

FGMM GMM utilizando **Fractional Covariance Matrix** (FCM)

Objetivos

Implementar sistemas de reconhecimento de locutor e analizar:

Objetivos

Implementar sistemas de reconhecimento de locutor e analizar:

- Taxas de **sucesso** para identificação

Objetivos

Implementar sistemas de reconhecimento de locutor e analizar:

- Taxas de **sucesso** para identificação
  - Diferentes tamanhos de mistura (M)

Objetivos

Implementar sistemas de reconhecimento de locutor e analizar:

- Taxas de **sucesso** para identificação
  - Diferentes tamanhos de mistura (M)
  - Diferentes tamanhos de características ($\mathbf{\Delta}$)

Objetivos

Implementar sistemas de reconhecimento de locutor e analizar:

- Taxas de **sucesso** para identificação
  - Diferentes tamanhos de mistura (M)
  - Diferentes tamanhos de características ($\mathbf{\Delta}$)
- Comparar identificações utilizando GMM e FGMM

Objetivos

Implementar sistemas de reconhecimento de locutor e analizar:

- Taxas de **sucesso** para identificação
  - Diferentes tamanhos de mistura (M)
  - Diferentes tamanhos de características ($\mathbf{\Delta}$)
- Comparar identificações utilizando GMM e FGMM
- Taxas de **falsa detecção** e **falsa rejeição** para verificação

Objetivos

Implementar sistemas de reconhecimento de locutor e analizar:

- Taxas de **sucesso** para identificação
  - Diferentes tamanhos de mistura (M)
  - Diferentes tamanhos de características ($\mathbf{\Delta}$)
- Comparar identificações utilizando GMM e FGMM
- Taxas de **falsa detecção** e **falsa rejeição** para verificação
  - Diferentes tamanhos de mistura (M)

Objetivos

Implementar sistemas de reconhecimento de locutor e analizar:

- Taxas de **sucesso** para identificação
  - Diferentes tamanhos de mistura (M)
  - Diferentes tamanhos de características ($\mathbf{\Delta}$)
- Comparar identificações utilizando GMM e FGMM
- Taxas de **falsa detecção** e **falsa rejeição** para verificação
  - Diferentes tamanhos de mistura (M)
  - Diferentes tamanhos de características ($\mathbf{\Delta}$)

Objetivos

Implementar sistemas de reconhecimento de locutor e analizar:

- Taxas de **sucesso** para identificação
  - Diferentes tamanhos de mistura (M)
  - Diferentes tamanhos de características ($\mathbf{\Delta}$)
- Comparar identificações utilizando GMM e FGMM
- Taxas de **falsa detecção** e **falsa rejeição** para verificação
  - Diferentes tamanhos de mistura (M)
  - Diferentes tamanhos de características ($\mathbf{\Delta}$)
- Comparar verificações utilizando GMM e AGMM

Conteúdo

Identificação

Modelagem Para cada locutor $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

Identificação

Modelagem Para cada locutor $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por $\mathcal{S}_j$

Identificação

Modelagem Para cada locutor $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por $\mathcal{S}_j$
- Treinar um λ_j para cada $\mathcal{S}_j$ através dos $\mathbf{X}_k$

Identificação

Modelagem Para cada locutor $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por $\mathcal{S}_j$
- Treinar um λ_j para cada $\mathcal{S}_j$ através dos $\mathbf{X}_k$

Teste Para um locutor desconhecido $\mathcal{S}$

Identificação

Modelagem Para cada locutor $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por $\mathcal{S}_j$
- Treinar um λ_j para cada $\mathcal{S}_j$ através dos $\mathbf{X}_k$

Teste Para um locutor desconhecido $\mathcal{S}$

- Extrair $\mathbf{X}$ do sinal $\mathbf{Y}$ falado por $\mathcal{S}$

Identificação

Modelagem Para cada locutor $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por $\mathcal{S}_j$
- Treinar um λ_j para cada $\mathcal{S}_j$ através dos $\mathbf{X}_k$

Teste Para um locutor desconhecido $\mathcal{S}$

- Extrair $\mathbf{X}$ do sinal $\mathbf{Y}$ falado por $\mathcal{S}$
- $i = \arg_j \max p(\mathbf{X}|\lambda_j) \implies \mathcal{S} \leftarrow \mathcal{S}_i$

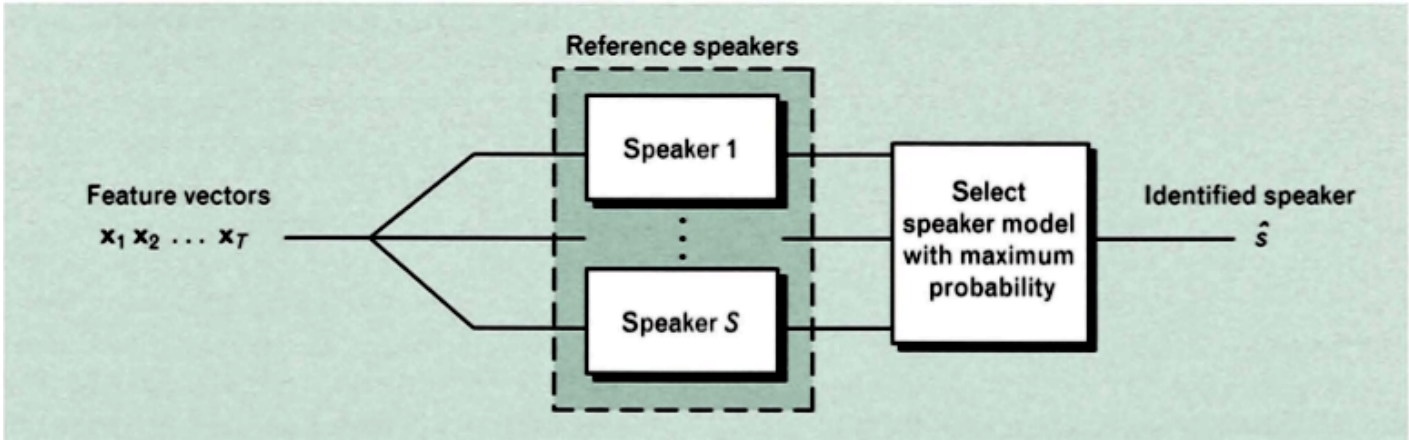
Identificação

Modelagem Para cada locutor $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por $\mathcal{S}_j$
- Treinar um λ_j para cada $\mathcal{S}_j$ através dos $\mathbf{X}_k$

Teste Para um locutor desconhecido $\mathcal{S}$

- Extrair $\mathbf{X}$ do sinal $\mathbf{Y}$ falado por $\mathcal{S}$
- $i = \arg_j \max p(\mathbf{X}|\lambda_j) \implies \mathcal{S} \leftarrow \mathcal{S}_i$



Verificação

Modelagem Para todos os $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

Verificação

Modelagem Para todos os $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por cada $\mathcal{S}_j$

Verificação

Modelagem Para todos os $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por cada $\mathcal{S}_j$
- Treinar um λ_{bkg} através dos $\mathbf{X}_k$ de todos os $\mathcal{S}_j$

Verificação

Modelagem Para todos os $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por cada $\mathcal{S}_j$
- Treinar um λ_{bkg} através dos $\mathbf{X}_k$ de todos os $\mathcal{S}_j$
- Modelar um λ_j para cada $\mathcal{S}_j$

Verificação

Modelagem Para todos os $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por cada $\mathcal{S}_j$
- Treinar um λ_{bkg} através dos $\mathbf{X}_k$ de todos os $\mathcal{S}_j$
- Modelar um λ_j para cada $\mathcal{S}_j$

Teste $\mathcal{S}$ diz ser $\mathcal{S}_C \in \boldsymbol{\mathcal{S}}$

Verificação

Modelagem Para todos os $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por cada $\mathcal{S}_j$
- Treinar um λ_{bkg} através dos $\mathbf{X}_k$ de todos os $\mathcal{S}_j$
- Modelar um λ_j para cada $\mathcal{S}_j$

Teste $\mathcal{S}$ diz ser $\mathcal{S}_C \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}$ do sinal $\mathbf{Y}$ falado por $\mathcal{S}_C$

Verificação

Modelagem Para todos os $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por cada $\mathcal{S}_j$
- Treinar um λ_{bkg} através dos $\mathbf{X}_k$ de todos os $\mathcal{S}_j$
- Modelar um λ_j para cada $\mathcal{S}_j$

Teste $\mathcal{S}$ diz ser $\mathcal{S}_C \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}$ do sinal $\mathbf{Y}$ falado por $\mathcal{S}_C$
- $\Lambda(\mathbf{X}) = \log p(\mathbf{X}|\lambda_C) - \log p(\mathbf{X}|\lambda_{bkg})$

Verificação

Modelagem Para todos os $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por cada $\mathcal{S}_j$
- Treinar um λ_{bkg} através dos $\mathbf{X}_k$ de todos os $\mathcal{S}_j$
- Modelar um λ_j para cada $\mathcal{S}_j$

Teste $\mathcal{S}$ diz ser $\mathcal{S}_C \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}$ do sinal $\mathbf{Y}$ falado por $\mathcal{S}_C$
- $\Lambda(\mathbf{X}) = \log p(\mathbf{X}|\lambda_C) - \log p(\mathbf{X}|\lambda_{bkg})$
- $\Lambda(\mathbf{X}) \geq \theta \implies \textit{aceita}$

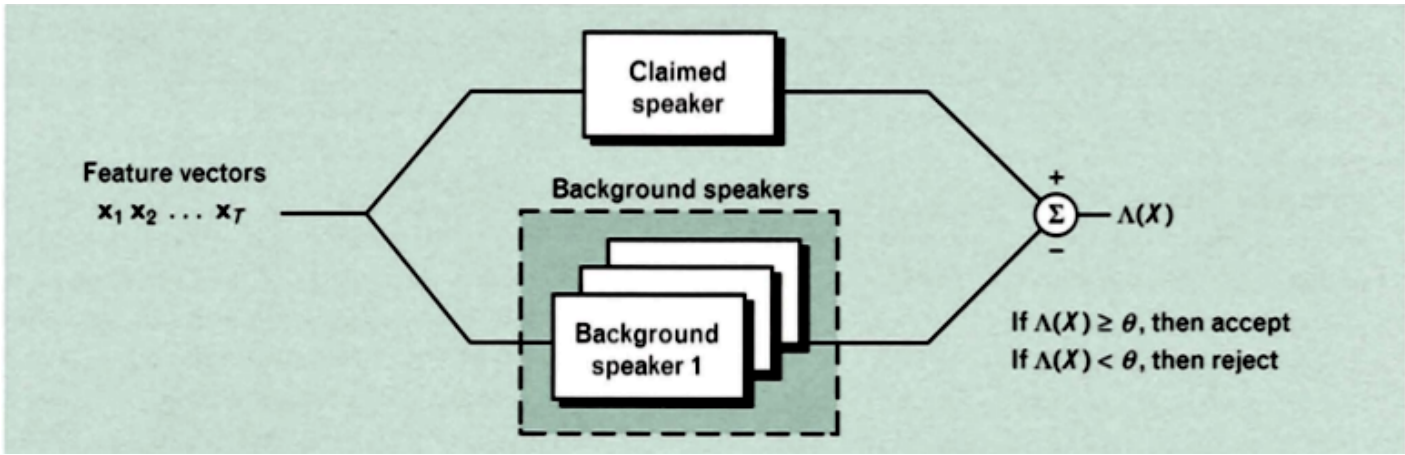
Verificação

Modelagem Para todos os $\mathcal{S}_j \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}_k$ dos sinais $\mathbf{Y}_k$ falados por cada $\mathcal{S}_j$
- Treinar um λ_{bkg} através dos $\mathbf{X}_k$ de todos os $\mathcal{S}_j$
- Modelar um λ_j para cada $\mathcal{S}_j$

Teste $\mathcal{S}$ diz ser $\mathcal{S}_C \in \boldsymbol{\mathcal{S}}$

- Extrair $\mathbf{X}$ do sinal $\mathbf{Y}$ falado por $\mathcal{S}_C$
- $\Lambda(\mathbf{X}) = \log p(\mathbf{X}|\lambda_C) - \log p(\mathbf{X}|\lambda_{bkg})$
- $\Lambda(\mathbf{X}) \geq \theta \implies$ *aceita*



Conteúdo

Características Ideais

- Natural e frequente na fala

Características Ideais

- Natural e frequente na fala
- Facilmente mensurável

Características Ideais

- Natural e frequente na fala
- Facilmente mensurável
- $\uparrow$ variação inter-locutor e $\downarrow$ variação intra-locutor

Características Ideais

- Natural e frequente na fala
- Facilmente mensurável
- $\uparrow$ variação inter-locutor e $\downarrow$ variação intra-locutor
- Constante no tempo e não afetável pela saúde

Características Ideais

- Natural e frequente na fala
- Facilmente mensurável
- $\uparrow$ variação inter-locutor e $\downarrow$ variação intra-locutor
- Constante no tempo e não afetável pela saúde
- Robusta a ruído razoável e a transmissão

Características Ideais

- Natural e frequente na fala
- Facilmente mensurável
- $\uparrow$ variação inter-locutor e $\downarrow$ variação intra-locutor
- Constante no tempo e não afetável pela saúde
- Robusta a ruído razoável e a transmissão
- Difícil de ser produzido artificialmente

Características Ideais

- Natural e frequente na fala
- Facilmente mensurável
- $\uparrow$ variação inter-locutor e $\downarrow$ variação intra-locutor
- Constante no tempo e não afetável pela saúde
- Robusta a ruído razoável e a transmissão
- Difícil de ser produzido artificialmente
- Não ser facilmente modificável pelo locutor

Mel-Frequency Cepstrum Coefficients

Simula a função da **cóclea**

Mel-Frequency Cepstrum Coefficients

	Simula a função da **cóclea**
Escala Mel	Logaritmica

Mel-Frequency Cepstrum Coefficients

Simula a função da **cóclea**

Escala Mel Logaritmica

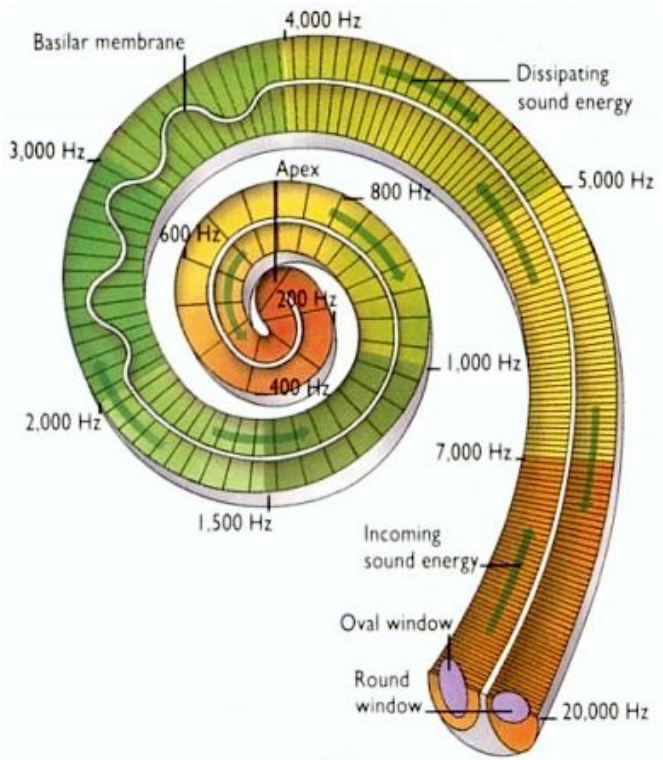
- $f_{mel} = 2595 \log_{10}(1 + \frac{f}{700})$

Mel-Frequency Cepstrum Coefficients

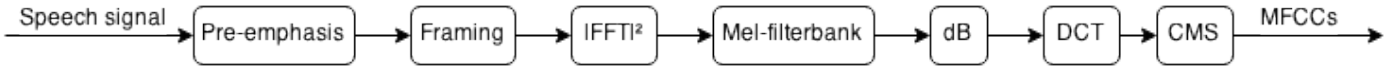
Simula a função da **cóclea**

Escala Mel Logaritmica

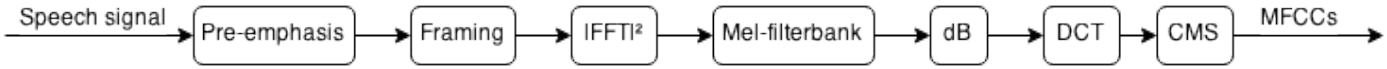
- $f_{mel} = 2595 \log_{10}(1 + \frac{f}{700})$



MFCC - Extração

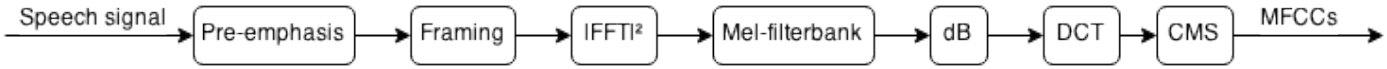


MFCC - Extração



Pré-ênfase **Realça** as altas frequências (opcional)

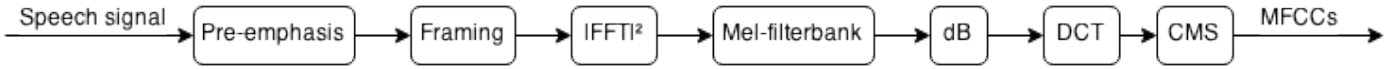
MFCC - Extração



Pré-ênfase **Realça** as altas frequências (opcional)

- $s_{emph}[n] = s[n] - \alpha \cdot s[n-1]$

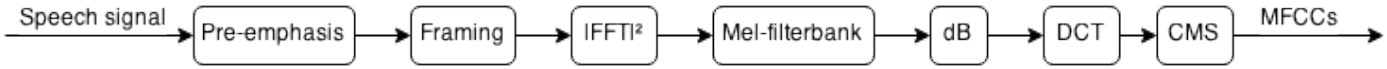
MFCC - Extração



Pré-ênfase **Realça** as altas frequências (opcional)

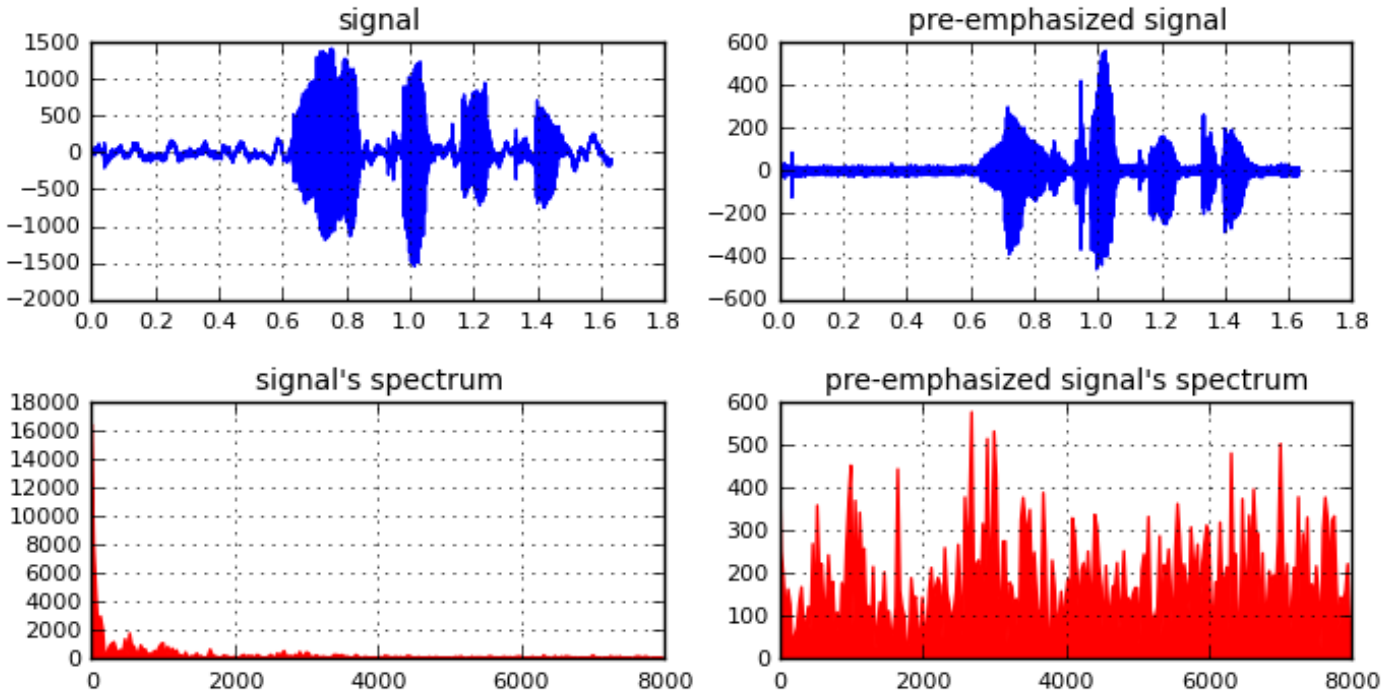
- $s_{emph}[n] = s[n] - \alpha \cdot s[n-1]$
- $\alpha \in [0.95, 0.98]$

MFCC - Extração

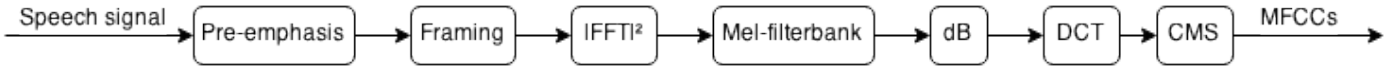


Pré-ênfase **Realça** as altas frequências (opcional)

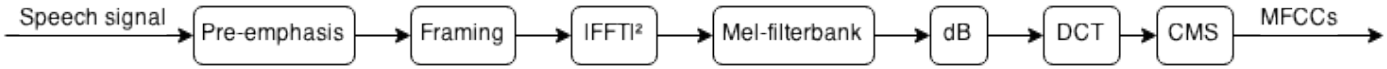
- $s_{emph}[n] = s[n] - \alpha \cdot s[n-1]$
- $\alpha \in [0.95, 0.98]$



MFCC - Extração

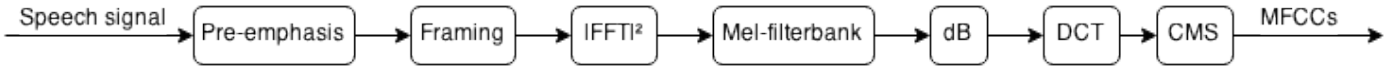


MFCC - Extração



Janelamento Divide o sinal em janelas **superpostas**

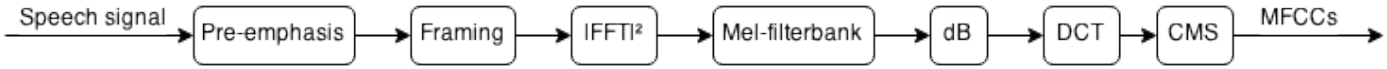
MFCC - Extração



Janelamento Divide o sinal em janelas **superpostas**

- Largura de 20 milissegundos

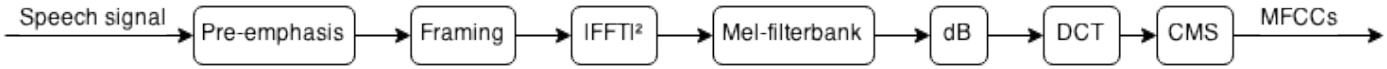
MFCC - Extração



Janelamento Divide o sinal em janelas **superpostas**

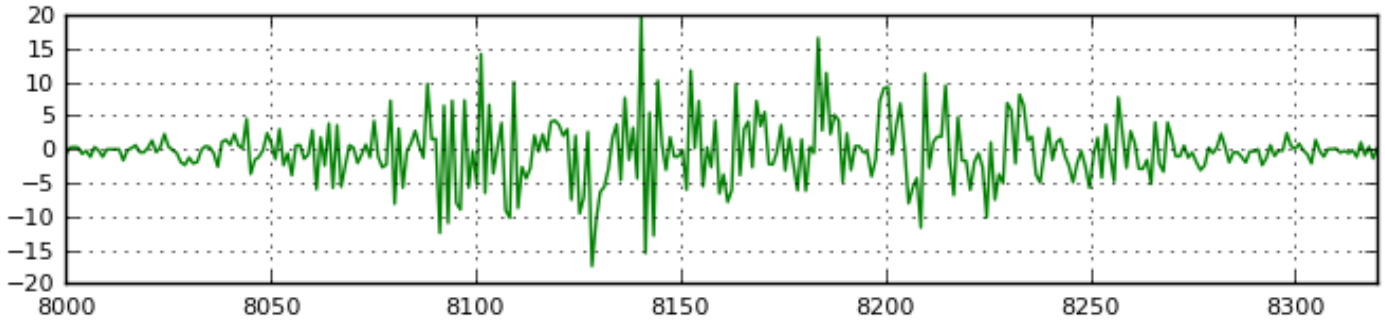
- Largura de 20 milissegundos
- Deslocamento de 10 milissegundos

MFCC - Extração



Janelamento Divide o sinal em janelas **superpostas**

- Largura de 20 milissegundos
- Deslocamento de 10 milissegundos



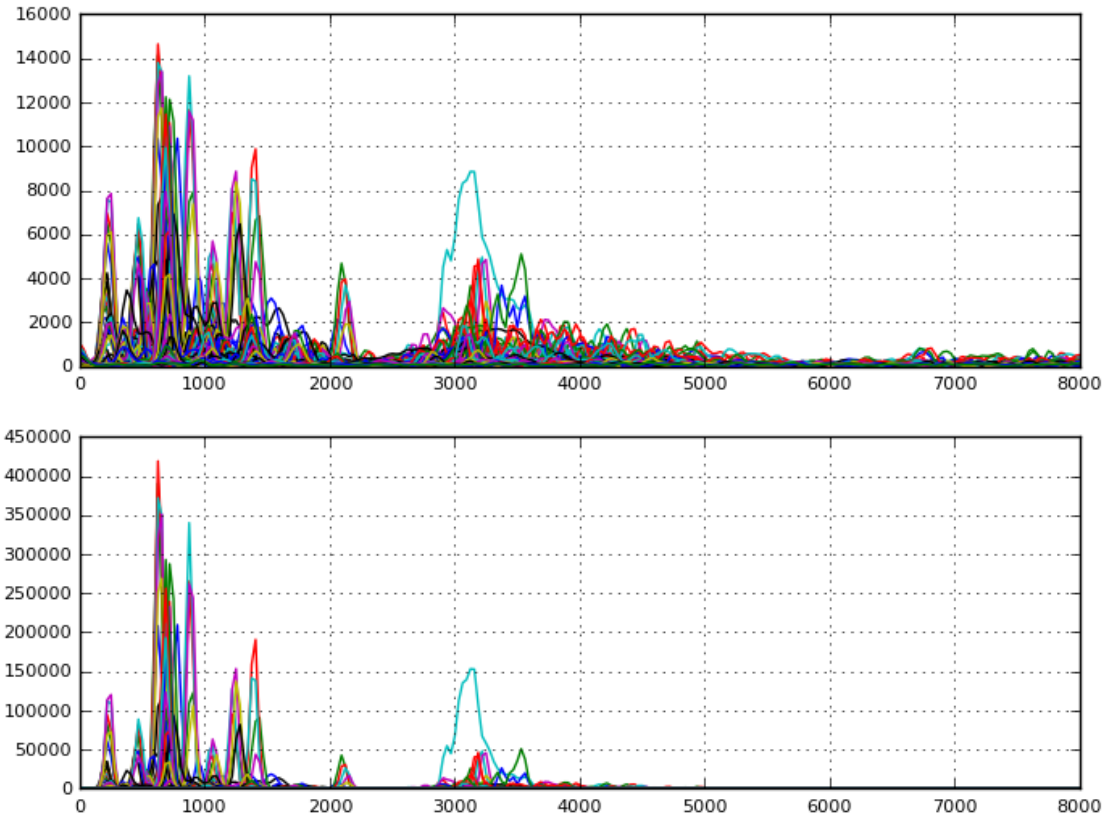
MFCC - Extração

MFCC - Extração

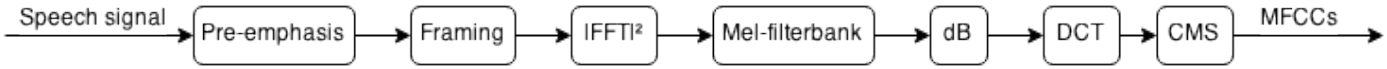
$|FFT|^2$ Calcula o **espectro de potência**

MFCC - Extração

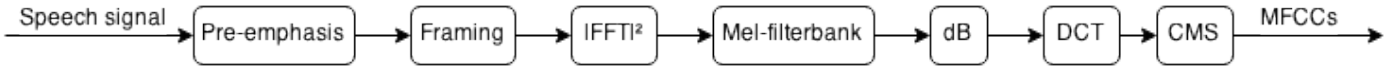
$|FFT|^2$ Calcula o **espectro de potência**



MFCC - Extração

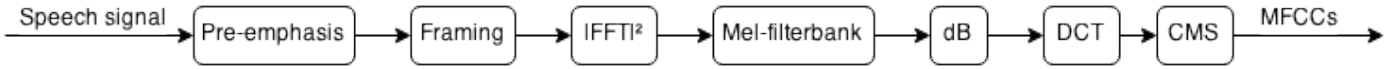


MFCC - Extração

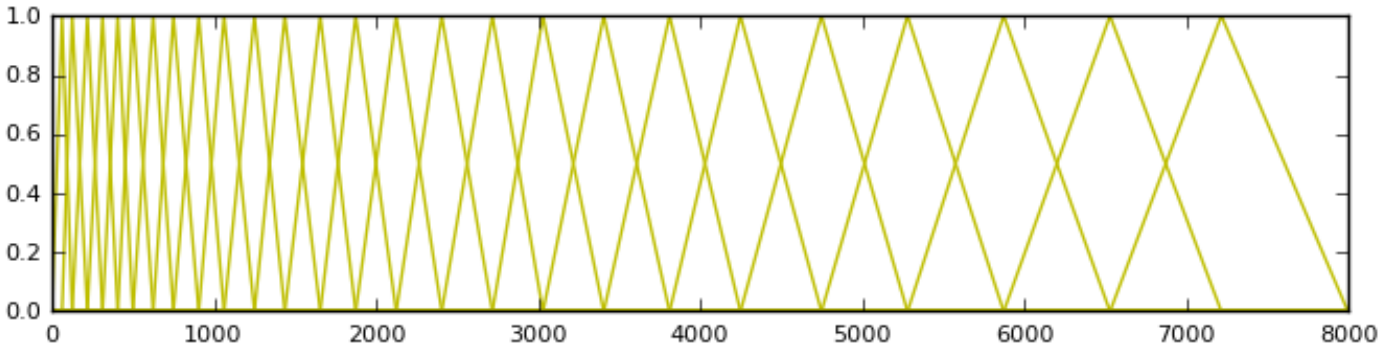


Filtros Espectro em Hz $\Longrightarrow$ espectro em **mels**

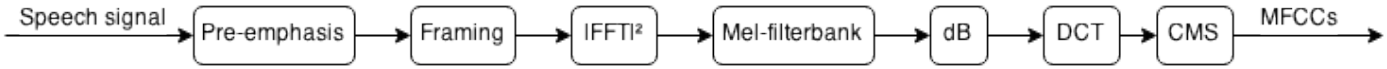
MFCC - Extração



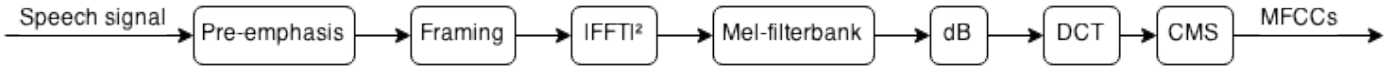
Filtros Espectro em Hz $\Longrightarrow$ espectro em **mels**



MFCC - Extração

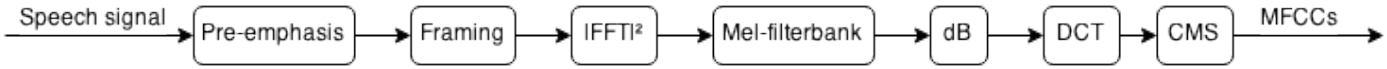


MFCC - Extração

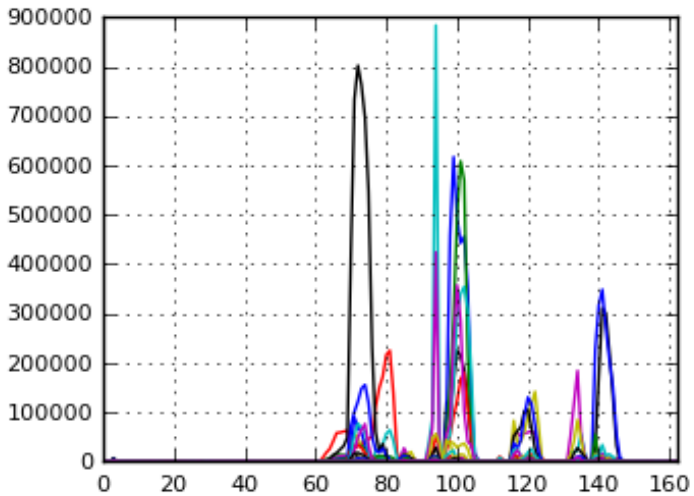


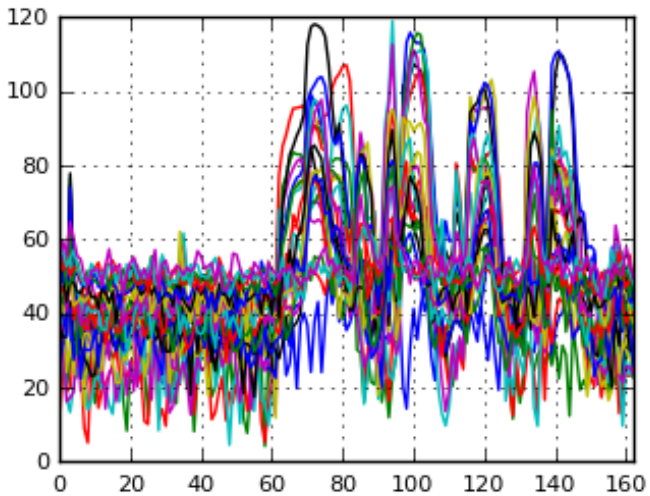
dB Calcula a **sonoridade**

MFCC - Extração



dB Calcula a **sonoridade**





MFCC - Extração

MFCC - Extração

DCT Coeficientes espectrais $\Longrightarrow$ coeficientes **cepstrais**

MFCC - Extração

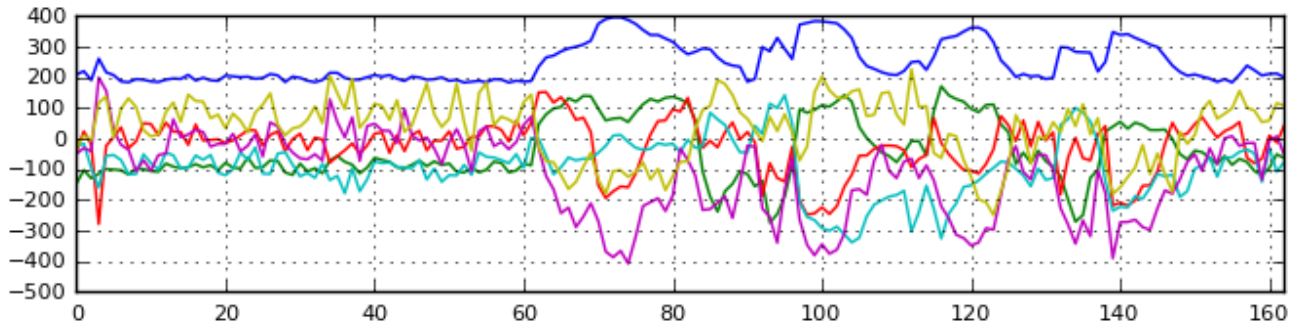
DCT Coeficientes espectrais $\Longrightarrow$ coeficientes **cepstrais**

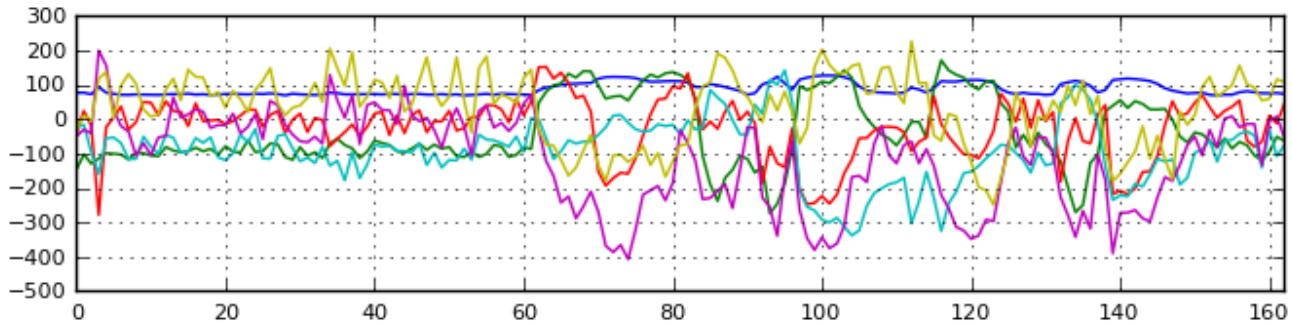
- $c_n = \sum_{k=1}^{K} S_k \cdot \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right], n = 1, 2, ..., L$

MFCC - Extração

DCT Coeficientes espectrais $\Longrightarrow$ coeficientes **cepstrais**

- $c_n = \sum_{k=1}^{K} S_k \cdot \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right], n = 1, 2, ..., L$





MFCC - Extração

MFCC - Extração

CMS **Normaliza** os MFCCs para reduzir perturbações

MFCC - Extração

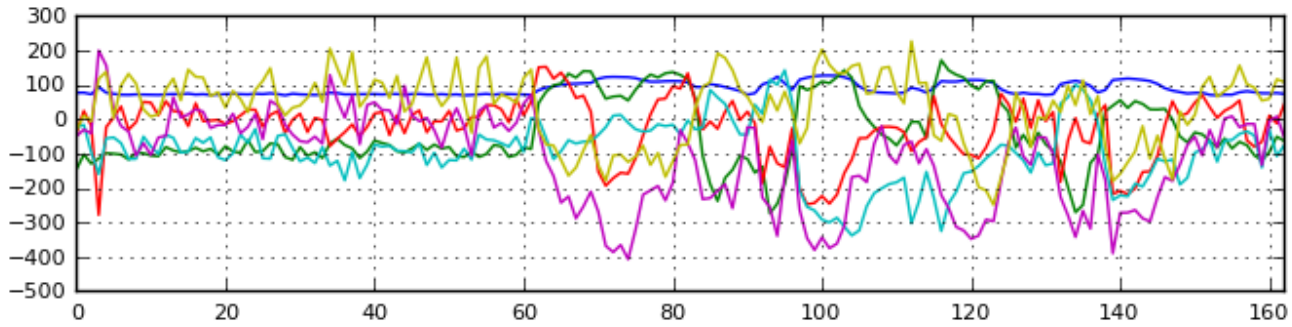
CMS **Normaliza** os MFCCs para reduzir perturbações

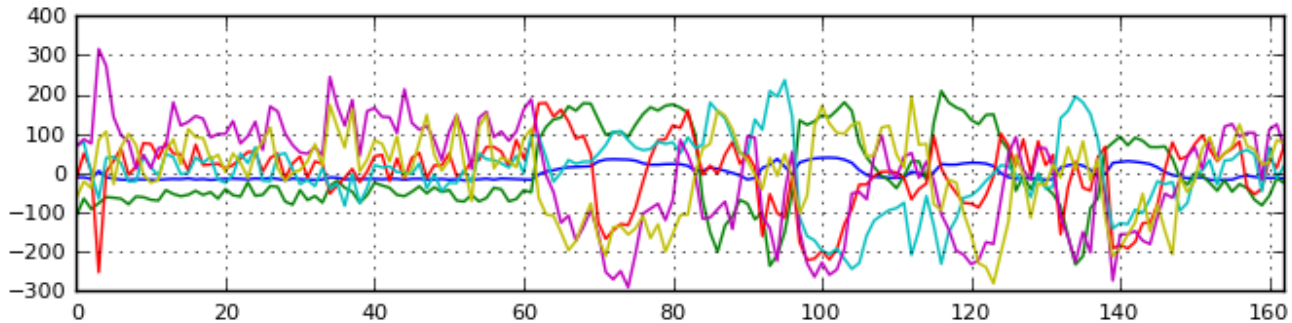
- $c_n = c_n - \frac{1}{T}\sum_{t=1}^{T} c_{n,t}$

MFCC - Extração

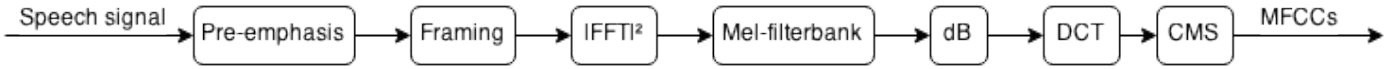
CMS **Normaliza** os MFCCs para reduzir perturbações

- $c_n = c_n - \frac{1}{T}\sum_{t=1}^{T} c_{n,t}$

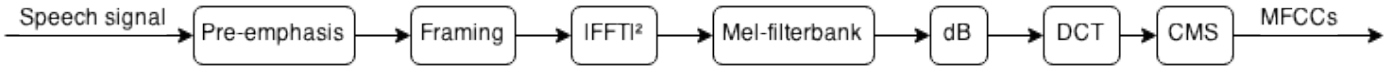




MFCC - Extração

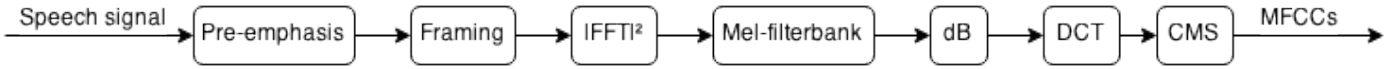


MFCC - Extração



Δs Novos c_n **derivados** dos antigos c_n (opcional)

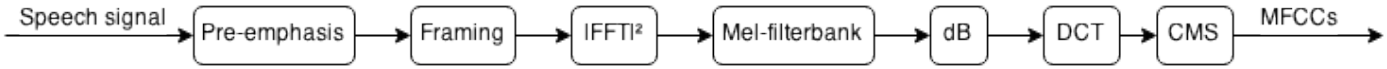
MFCC - Extração



Δs Novos c_n **derivados** dos antigos c_n (opcional)

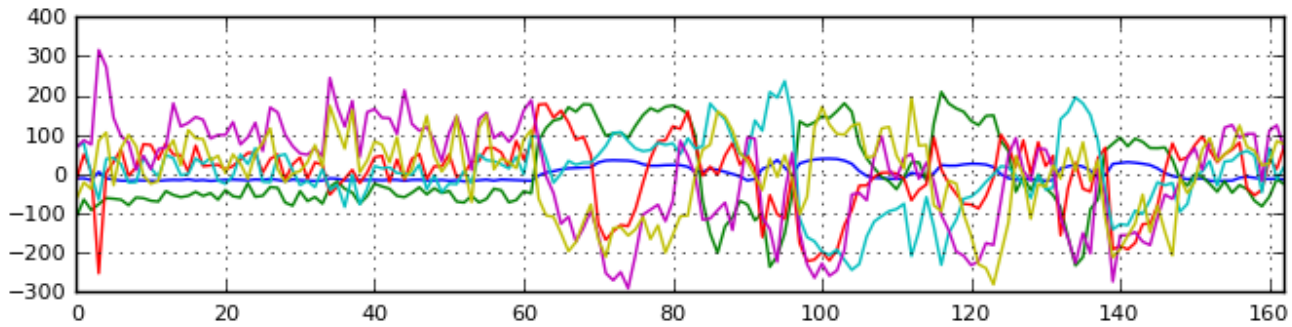
- $\Delta_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2}$

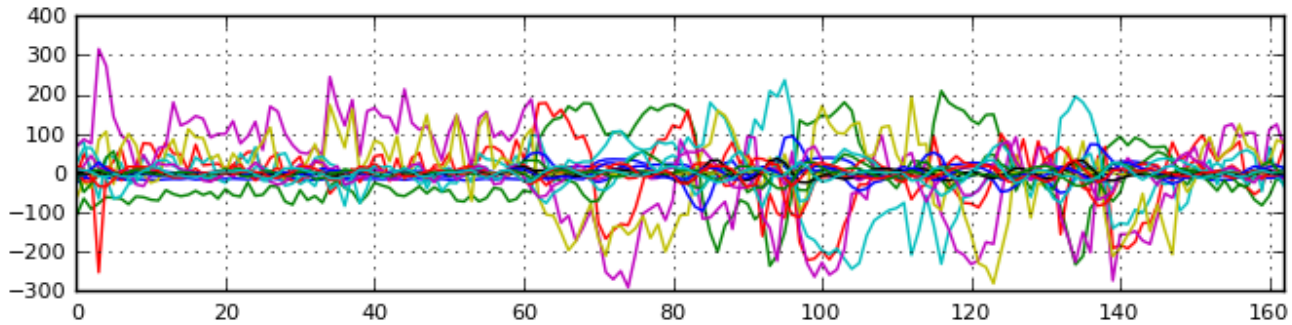
MFCC - Extração



Δs Novos c_n **derivados** dos antigos c_n (opcional)

- $\Delta_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2}$





Conteúdo

Modelos de Misturas Gaussianas

Conteúdo

Experimentos

Conteúdo

Conclusão

Obrigado