

# 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

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências

# Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências

# 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**

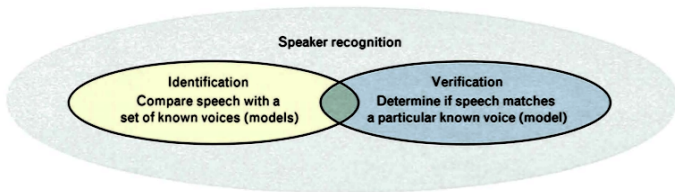
# Reconhecimento de Locutor

**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  $\implies$  Retreinamento

# Dependência de texto

Dependente Teste  $\in$  Treinamento

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

Independente Teste  $\neq$  Treinamento

# Dependência de texto

Dependente Teste  $\in$  Treinamento

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

Independente Teste  $\neq$  Treinamento

- Características não textuais

# Dependência de texto

Dependente Teste  $\in$  Treinamento

- Diversos graus de dependência
- Teste  $\notin$  Treinamento  $\implies$  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  $\implies$  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 analisar:

# Objetivos

Implementar sistemas de reconhecimento de locutor e analisar:

- Taxas de **sucesso** para identificação

# Objetivos

Implementar sistemas de reconhecimento de locutor e analisar:

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



# Objetivos

Implementar sistemas de reconhecimento de locutor e analisar:

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

# Objetivos

Implementar sistemas de reconhecimento de locutor e analisar:

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

# Objetivos

Implementar sistemas de reconhecimento de locutor e analisar:

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

# Objetivos

Implementar sistemas de reconhecimento de locutor e analisar:

- Taxas de **sucesso** para identificação
  - Diferentes tamanhos de mistura ( $M$ )
  - Diferentes tamanhos de características
- 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 analisar:

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

# Objetivos

Implementar sistemas de reconhecimento de locutor e analisar:

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

# Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor**
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências

# Identificação

Modelagem Para cada locutor  $S_j \in \mathcal{S}$



# Identificação

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

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

# Identificação

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

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

# Identificação

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

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

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

# Identificação

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

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

- Extrair  $\mathbf{X}_k$  dos sinais  $\mathbf{Y}_k$  falados por  $\mathcal{S}_j$
- Treinar um  $\lambda_j$  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  $\mathcal{S}$
- $i = \arg_j \max p(\mathbf{X}|\lambda_j) \implies \mathcal{S} \leftarrow \mathcal{S}_i$

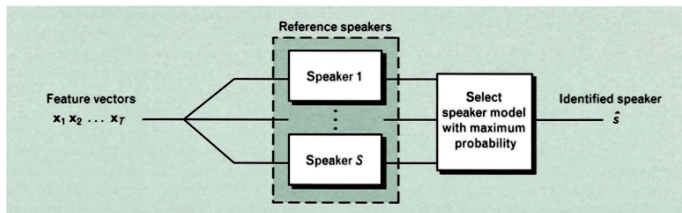
# Identificação

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

- Extrair  $\mathbf{X}_k$  dos sinais  $\mathbf{Y}_k$  falados por  $\mathcal{S}_j$
- Treinar um  $\lambda_j$  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  $\mathcal{S}$
- $i = \arg_j \max p(\mathbf{X}|\lambda_j) \Rightarrow \mathcal{S} \leftarrow \mathcal{S}_i$



# Verificação

Modelagem Para todos os  $S_j \in \mathcal{S}$

# Verificação

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

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



# Verificação

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

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

# Verificação

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

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

# Verificação

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

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

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

# Verificação

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

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

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

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

# Verificação

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

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

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

- 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  $\mathcal{S}_j \in \mathcal{S}$

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

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

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

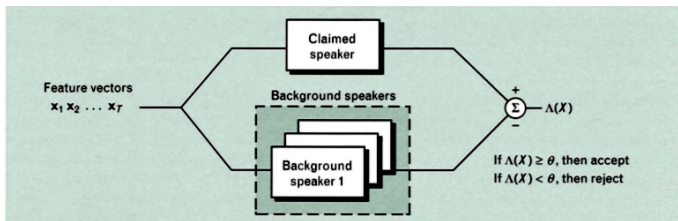
# Verificação

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

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

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

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



# Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características**
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências



# 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
- Robusto 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
- Robusto 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
- Robusto 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

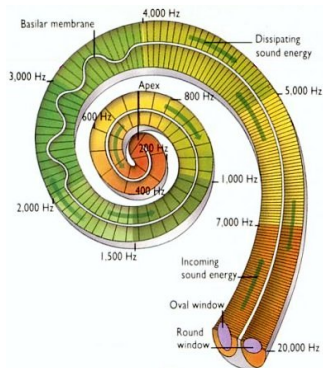
- $f_{mel} = 2595 \log_{10}(1 + \frac{f}{700})$

# Mel-Frequency Cepstrum Coefficients

Simula a função da **cóclea**

Escala Mel Logaritmica

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



# MFCC - Extração

Pré-ênfase Realça as altas frequências (opcional)

# MFCC - Extração

Pré-ênfase Realça as altas frequências (opcional)

Janelamento Divide o sinal em janelas superpostas

# MFCC - Extração

Pré-ênfase Realça as altas frequências (opcional)

Janelamento Divide o sinal em janelas superpostas

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

# MFCC - Extração

- Pré-ênfase Realça as altas frequências (opcional)
- Janelamento Divide o sinal em janelas superpostas
- $|FFT|^2$  Calcula o espectro de potência
- Filtros Espectro em Hz  $\implies$  espectro em mels

# MFCC - Extração

- Pré-ênfase Realça as altas frequências (opcional)
- Janelamento Divide o sinal em janelas superpostas
- $|FFT|^2$  Calcula o espectro de potência
- Filtros Espectro em Hz  $\implies$  espectro em mels
- dB Calcula a sonoridade



# MFCC - Extração

- Pré-ênfase Realça as altas frequências (opcional)
- Janelamento Divide o sinal em janelas superpostas
- $|FFT|^2$  Calcula o espectro de potência
- Filtros Espectro em Hz  $\implies$  espectro em mels
- dB Calcula a sonoridade
- DCT Coeficientes espectrais  $\implies$  coeficientes cepstrais

# MFCC - Extração

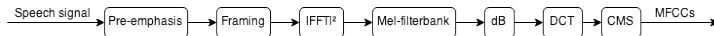
- Pré-ênfase Realça as altas frequências (opcional)
- Janelamento Divide o sinal em janelas superpostas
- $|FFT|^2$  Calcula o espectro de potência
- Filtros Espectro em Hz  $\implies$  espectro em mels
- dB Calcula a sonoridade
- DCT Coeficientes espectrais  $\implies$  coeficientes cepstrais
- CMS Normaliza os MFCCs para reduzir perturbações

# MFCC - Extração

- Pré-ênfase Realça as altas frequências (opcional)
- Janelamento Divide o sinal em janelas superpostas
- $|FFT|^2$  Calcula o espectro de potência
- Filtros Espectro em Hz  $\implies$  espectro em mels
  - dB Calcula a sonoridade
- DCT Coeficientes espectrais  $\implies$  coeficientes cepstrais
- CMS Normaliza os MFCCs para reduzir perturbações
  - $\Delta s$  Novos coeficientes a partir dos antigos (opcional)

# MFCC - Extração

- Pré-ênfase Realça as altas frequências (opcional)
- Janelamento Divide o sinal em janelas superpostas
- $|FFT|^2$  Calcula o espectro de potência
- Filtros Espectro em Hz  $\implies$  espectro em mels
  - dB Calcula a sonoridade
- DCT Coeficientes espectrais  $\implies$  coeficientes cepstrais
- CMS Normaliza os MFCCs para reduzir perturbações
  - $\Delta s$  Novos coeficientes a partir dos antigos (opcional)



# Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas**
- 5 Experimentos
- 6 Conclusão
- 7 Referências

# Modelos de Misturas Gaussianas

# Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos**
- 6 Conclusão
- 7 Referências

# Experimentos



# Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão**
- 7 Referências

# Conclusão

# Conteúdo

- 1 Introdução
- 2 Sistemas de Reconhecimento de Locutor
- 3 Extração de Características
- 4 Modelos de Mistura Gaussianas
- 5 Experimentos
- 6 Conclusão
- 7 Referências**

# Referências

# Obrigado