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

Conteúdo

1 Introdução

2 Sistemas de Reconhecimento de Locutor

3 Extração de Características

4 Modelos de Mistura Gaussianas

5 Experimentos

6 Conclusão

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
- 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
- Problema de **conjunto aberto**

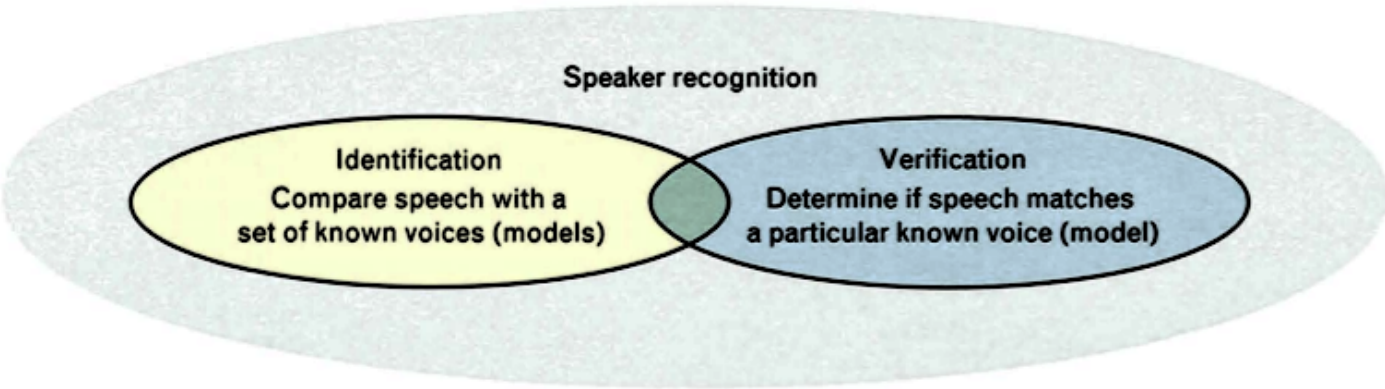
Reconhecimento de Locutor

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

Com Teste $\in$ Treinamento

Dependência de texto

Com Teste $\in$ Treinamento

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

Dependência de texto

Com Teste $\in$ Treinamento

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

Sem Teste $\neq$ Treinamento

Dependência de texto

Com Teste $\in$ Treinamento

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

Sem Teste $\neq$ Treinamento

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

Dependência de texto

Com Teste $\in$ Treinamento

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

Sem 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

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

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

FGMM GMM **fracionário** utilizando 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)
  - 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)
  - 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

1 Introdução

2 Sistemas de Reconhecimento de Locutor

3 Extração de Características

4 Modelos de Mistura Gaussianas

5 Experimentos

6 Conclusão

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 $\mathcal{S}_j$
- 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 $\mathcal{S}$
- $i = \arg_j \max p(\mathbf{X}|\lambda_j) \implies \mathcal{S} \leftarrow \mathcal{S}_i$

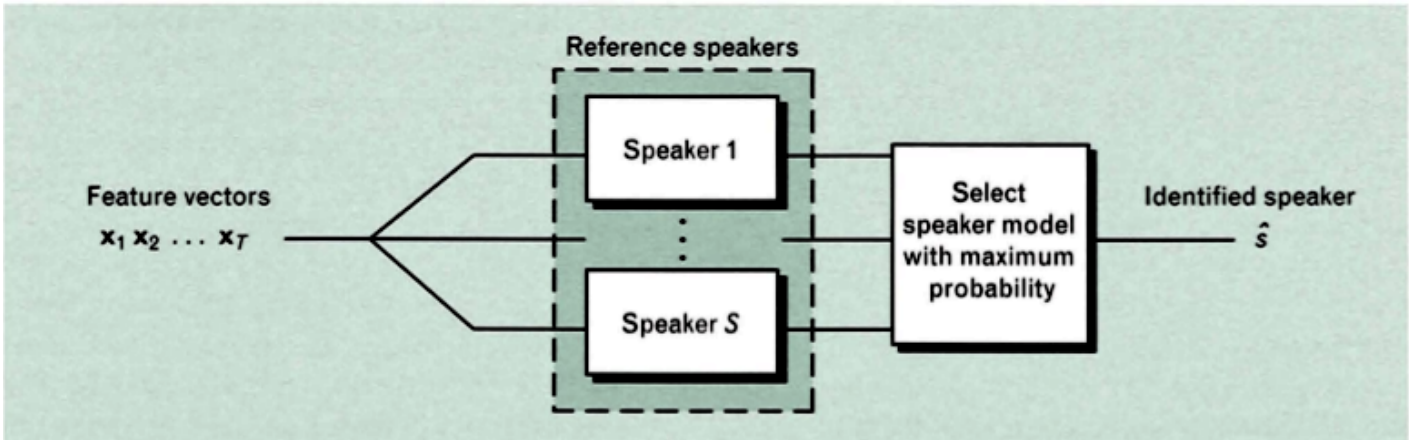
Identificação

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 $\mathcal{S}_j$
- 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 $\mathcal{S}_C$
- $\Lambda(\mathbf{X}) = \log p(\mathbf{X}|\lambda_C) - \log p(\mathbf{X}|\lambda_{bkg})$
- $\Lambda(\mathbf{X}) \geq \theta \implies \textit{aceita}$

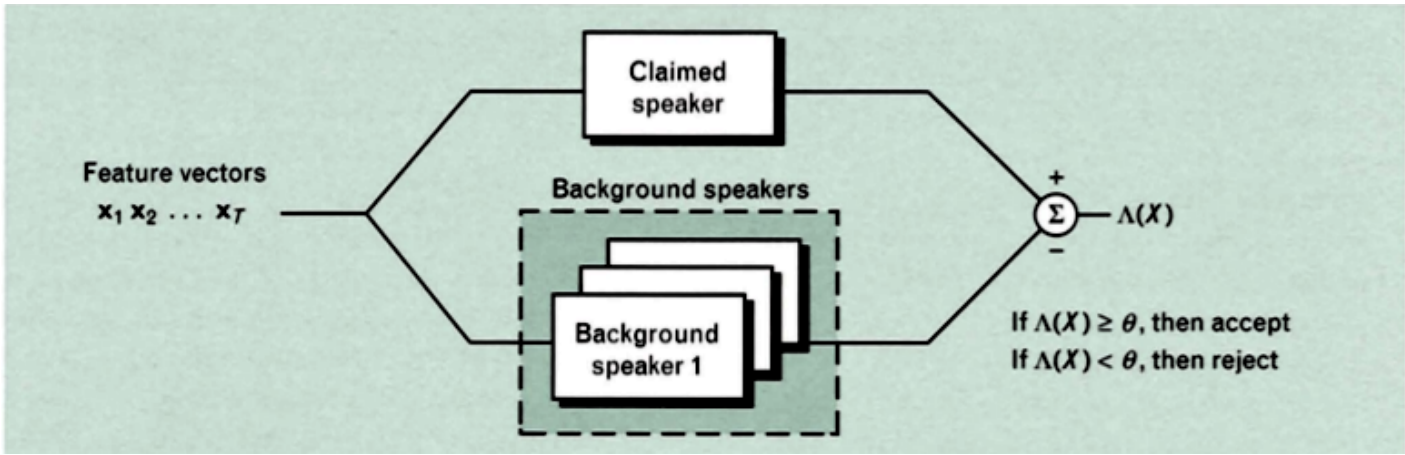
Verificação

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

1 Introdução

2 Sistemas de Reconhecimento de Locutor

3 Extração de Características

4 Modelos de Mistura Gaussianas

5 Experimentos

6 Conclusão

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

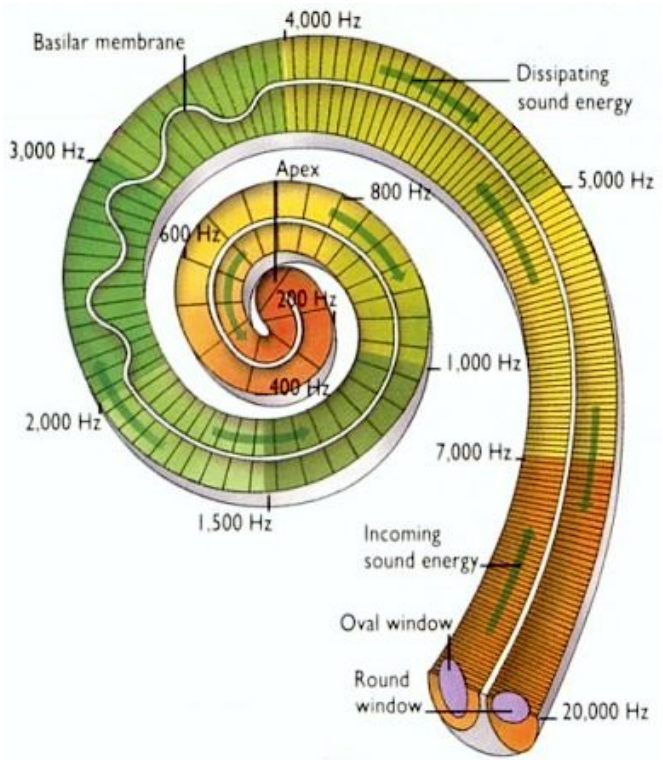
- $f_{mel} = 2595 \log_{10}(1 + \frac{f}{700})$

Mel-Frequency Cepstrum Coefficients

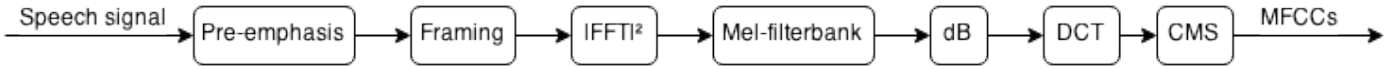
Simula a função da **cóclea**

Escala Mel Logaritmica

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

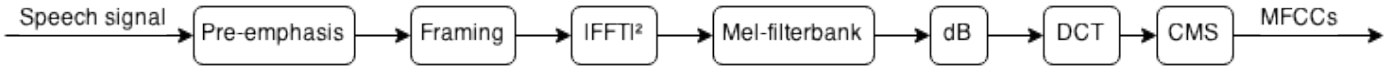


MFCC - Extração



Pre-emphasis **Realça** as frequências altas (opcional)

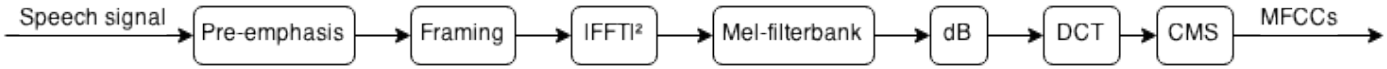
MFCC - Extração



Pre-emphasis **Realça** as frequências altas (opcional)

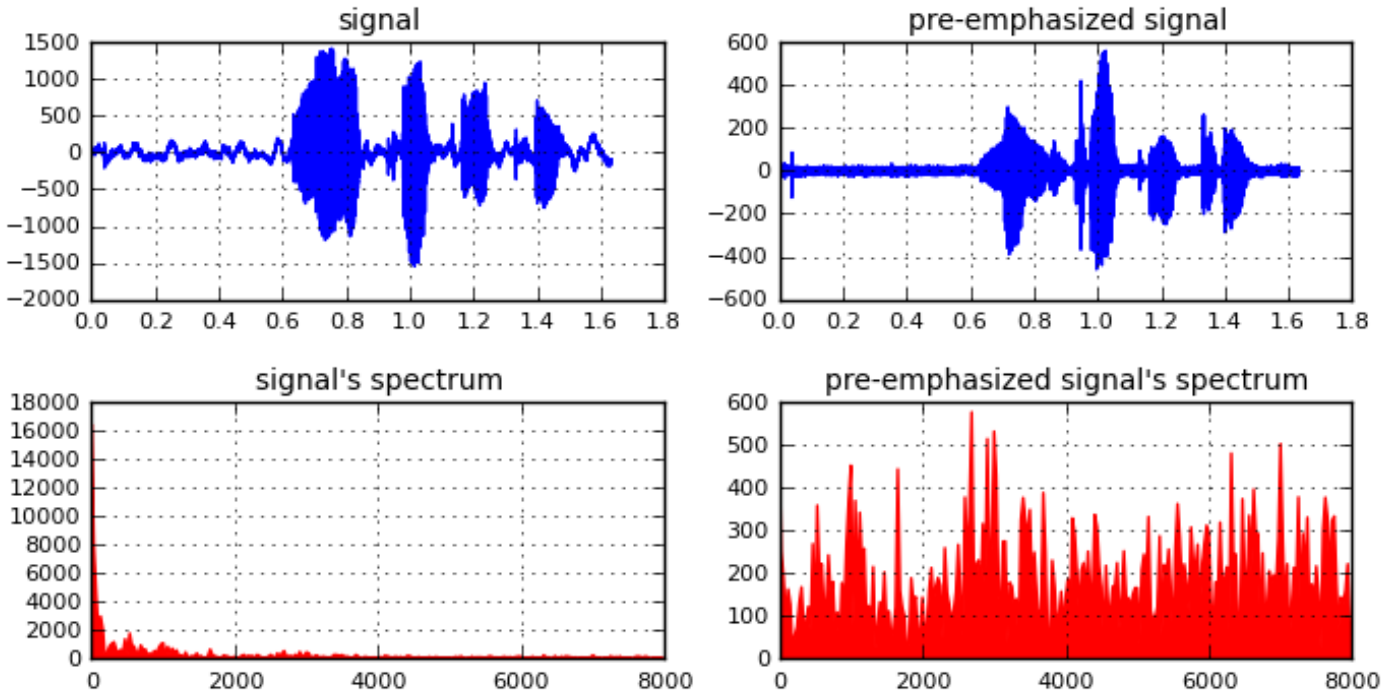
- $s_{emph}[n] = s[n] - \alpha \cdot s[n-1]$
- $\alpha \in [0.95, 0.98]$

MFCC - Extração

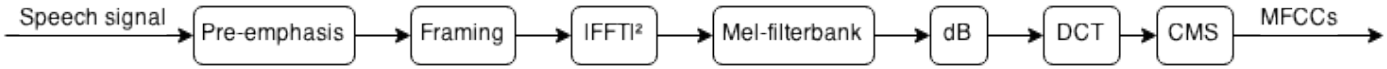


Pre-emphasis **Realça** as frequências altas (opcional)

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

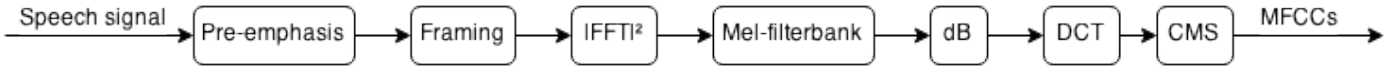


MFCC - Extração



Framing Divide o sinal em janelas **superpostas**

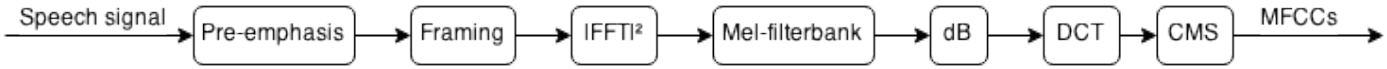
MFCC - Extração



Framing Divide o sinal em janelas **superpostas**

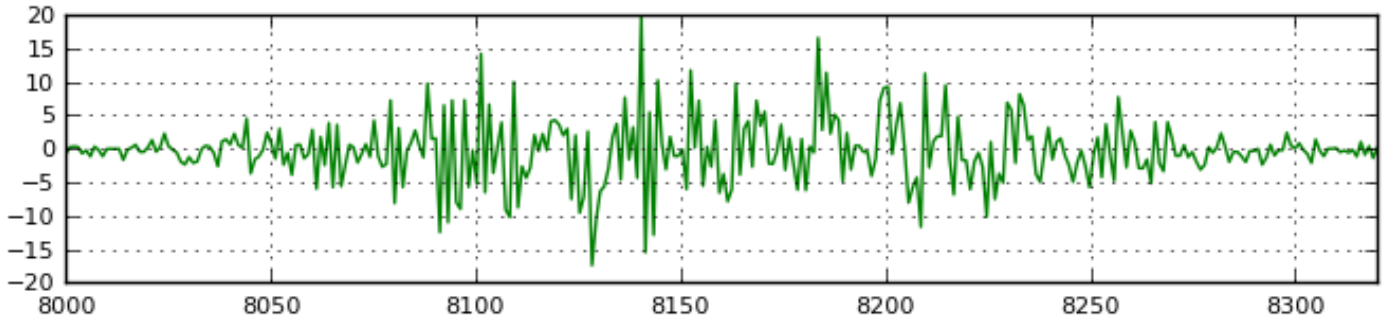
- Janela de Hamming
- Largura de 20 milissegundos
- Deslocamento de 10 milissegundos

MFCC - Extração



Framing Divide o sinal em janelas **superpostas**

- Janela de Hamming
- Largura de 20 milissegundos
- Deslocamento de 10 milissegundos

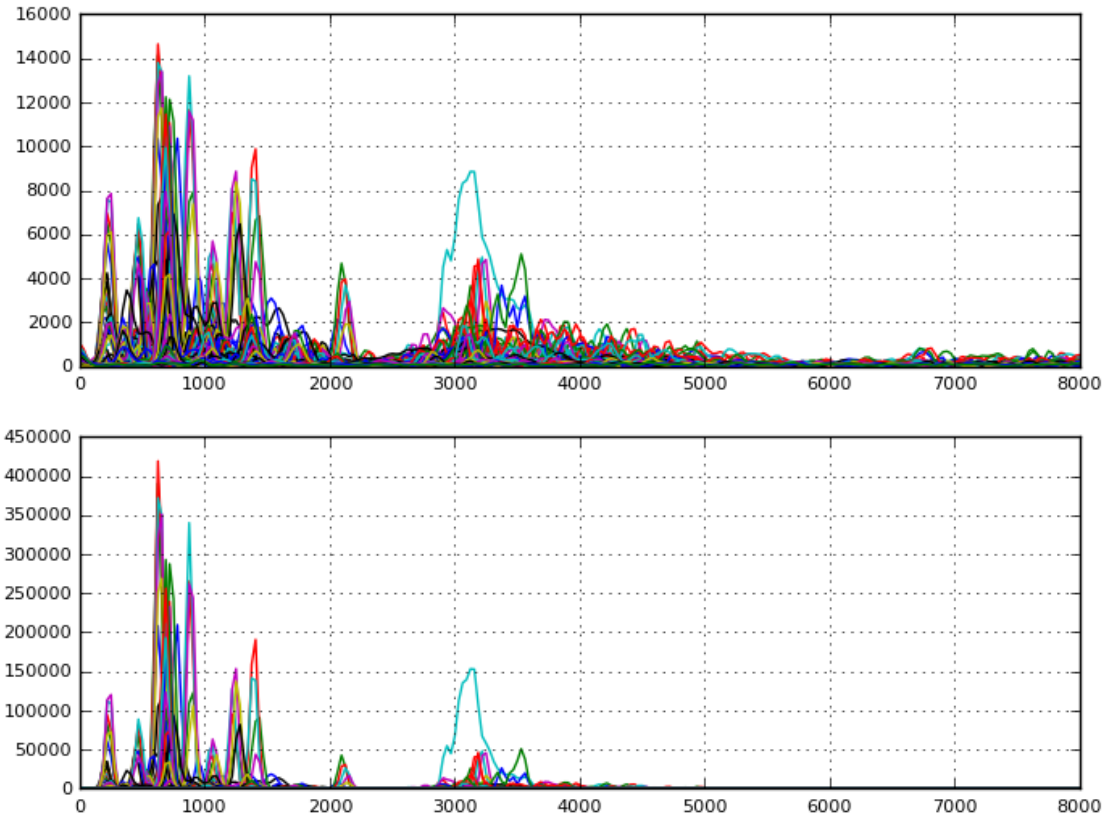


MFCC - Extração

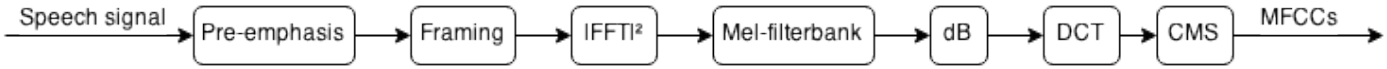
$|FFT|^2$ Calcula o **espectro de potência**

MFCC - Extração

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

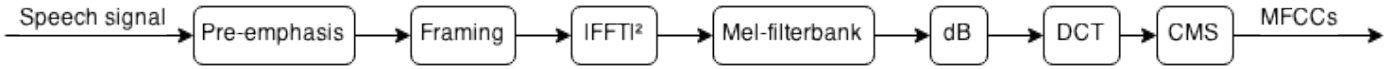


MFCC - Extração

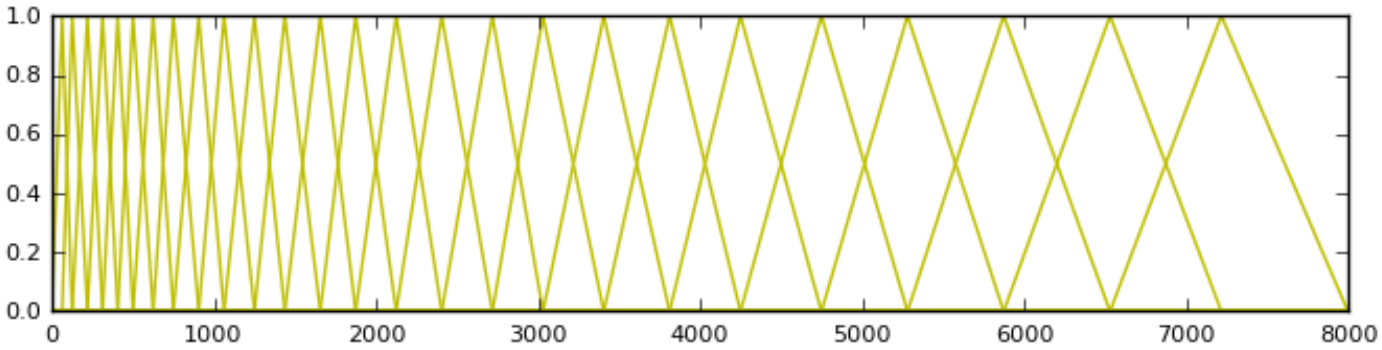


Mel-filterbank Espectro em Hz $\Longrightarrow$ espectro em **mels**

MFCC - Extração

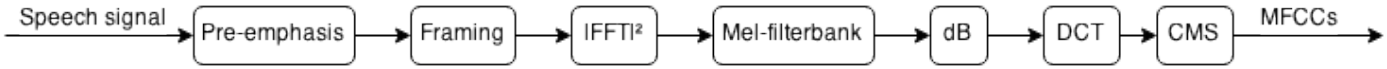


Mel-filterbank Espectro em Hz $\Longrightarrow$ espectro em **mels**



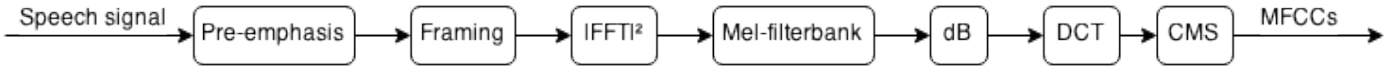
Na escala mel, as larguras são iguais

MFCC - Extração

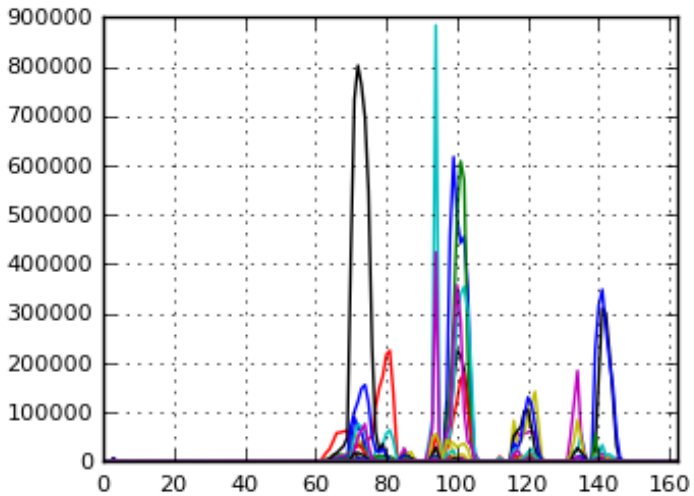


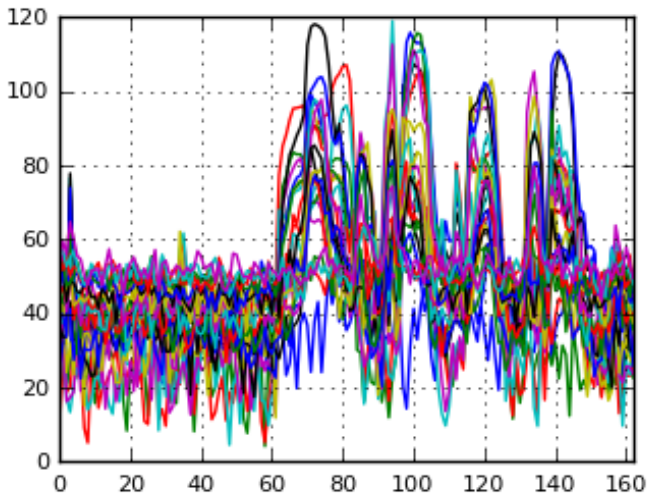
dB Calcula a **sonoridade**

MFCC - Extração

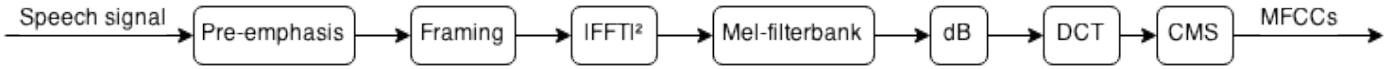


dB Calcula a **sonoridade**



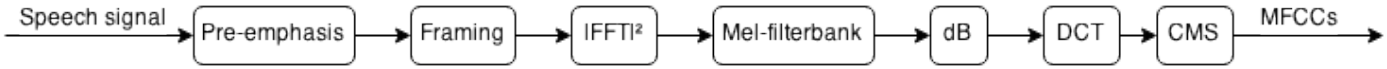


MFCC - Extração



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

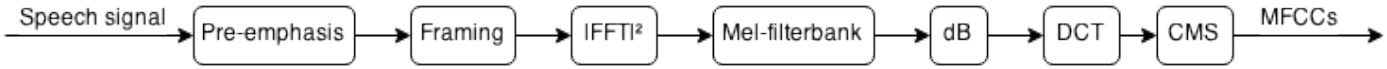
MFCC - Extração



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

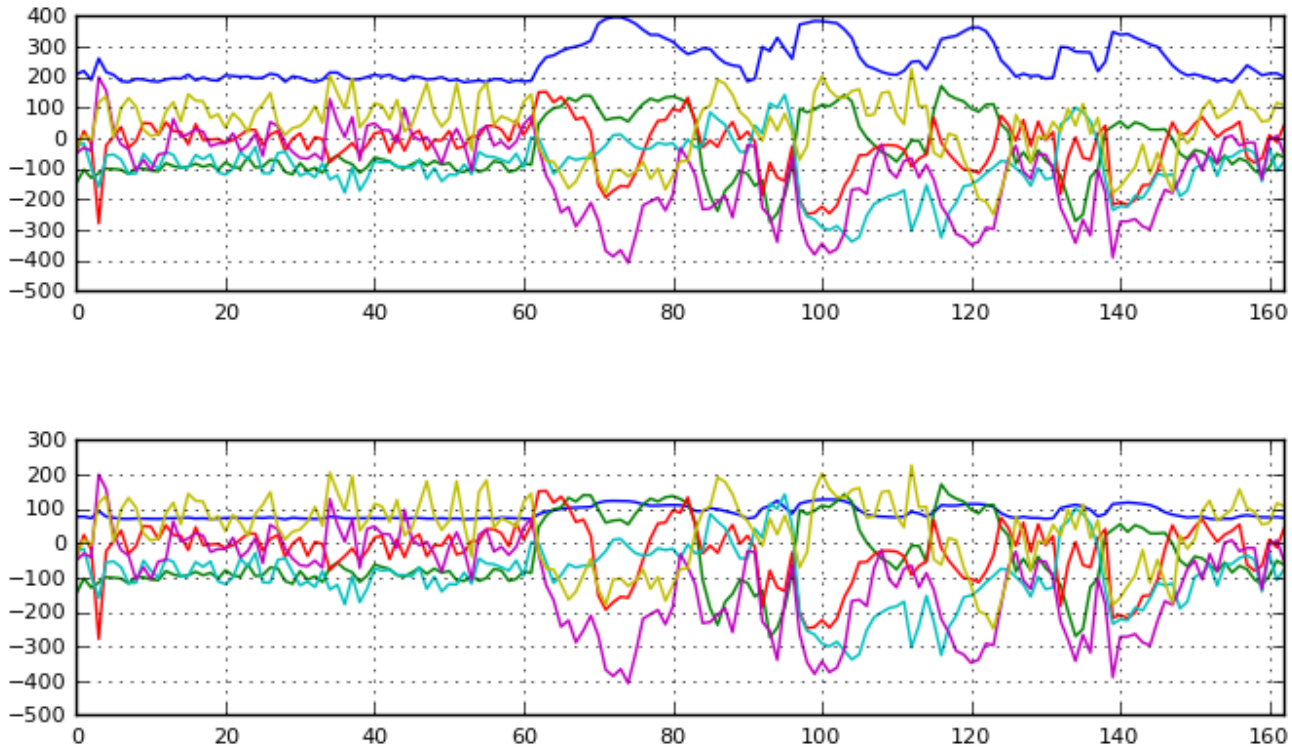
- $c_n = \sum_{k=1}^{K} S_k \cdot \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K}\right], n = 1, 2, ..., L$

MFCC - Extração

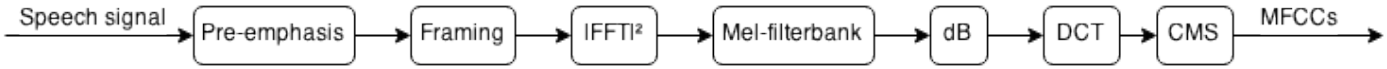


DCT Coeficientes espectrais $\Longrightarrow$ coeficientes **cepstrais**

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

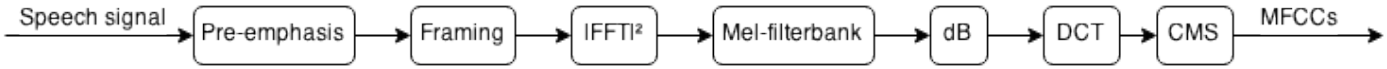


MFCC - Extração



CMS **Normaliza** os MFCCs para reduzir perturbações

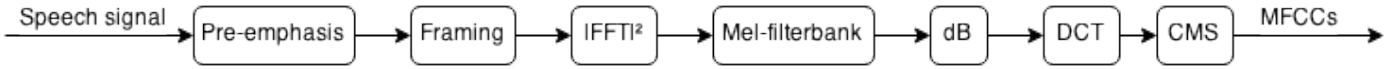
MFCC - Extração



CMS **Normaliza** os MFCCs para reduzir perturbações

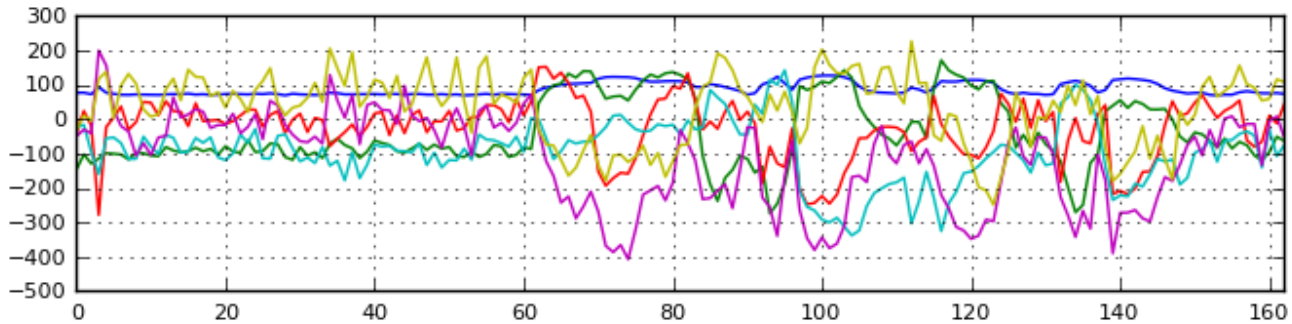
- $c_n = c_n - \frac{1}{T}\sum_{t=1}^{T} c_{n,t}$

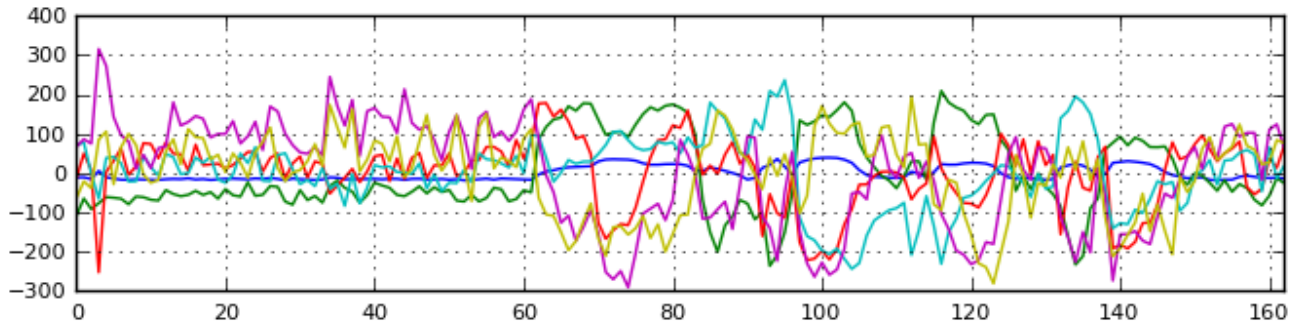
MFCC - Extração



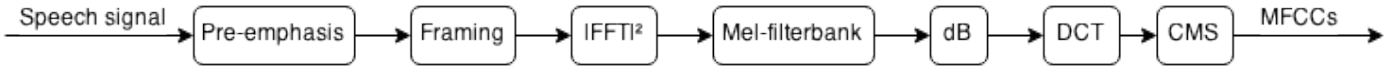
CMS **Normaliza** os MFCCs para reduzir perturbações

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



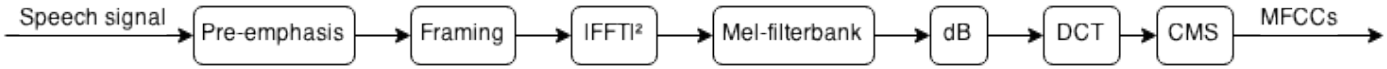


MFCC - Extração



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

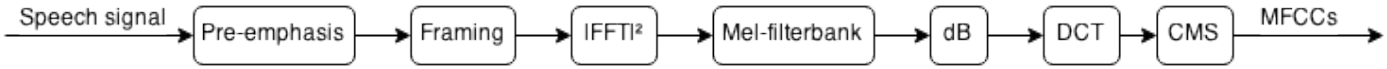
MFCC - Extração



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

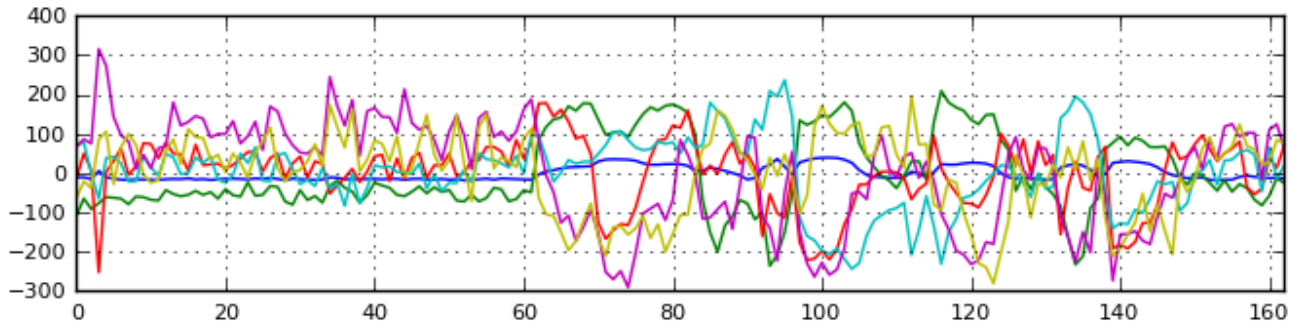
- $\Delta_t = \frac{\sum_{n=1}^{N} n(c_{t+n} - c_{t-n})}{2\sum_{n=1}^{N} n^2}$

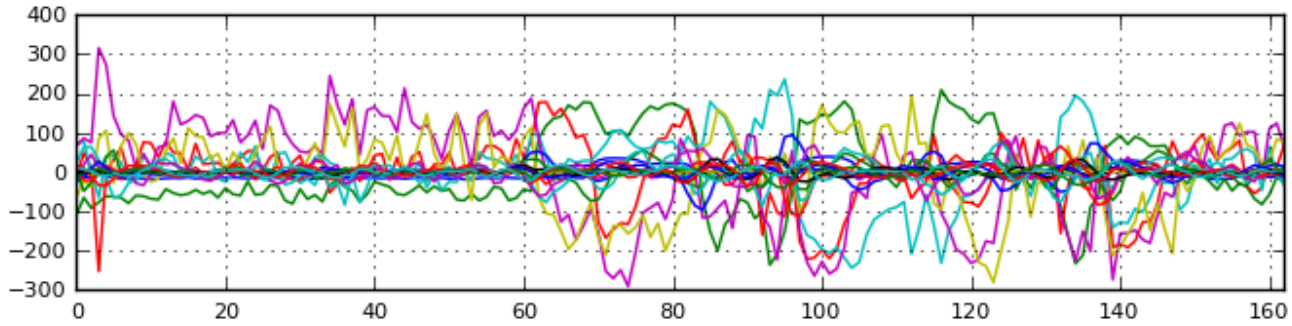
MFCC - Extração



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

1 Introdução

2 Sistemas de Reconhecimento de Locutor

3 Extração de Características

4 Modelos de Mistura Gaussianas

5 Experimentos

6 Conclusão

Definição

GMM $p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i p_i(\mathbf{x})$

Gaussiana $p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$

Definição

GMM $p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i p_i(\mathbf{x})$

Gaussiana $p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$

$\lambda = \{(w_i, \boldsymbol{\mu}_i, \mathbf{\Sigma}_i)\}, i = 1, ..., M$

$\mathbf{\Sigma}$ diagonal $\implies \boldsymbol{\sigma}^2$

Definição

GMM $p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i p_i(\mathbf{x})$

Gaussiana $p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$

$\lambda = \{(w_i, \boldsymbol{\mu}_i, \mathbf{\Sigma}_i)\}, i = 1, ..., M$

$\mathbf{\Sigma}$ diagonal $\implies \boldsymbol{\sigma}^2$

Dada uma sequência $\mathbf{X}$

Definição

GMM $p(\mathbf{x}|\lambda) = \sum_{i=1}^{M} w_i p_i(\mathbf{x})$

Gaussiana $p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$

$\lambda = \{(w_i, \boldsymbol{\mu}_i, \mathbf{\Sigma}_i)\}, i = 1, ..., M$

$\mathbf{\Sigma}$ diagonal $\implies \boldsymbol{\sigma}^2$

Dada uma sequência $\mathbf{X}$

- $p(\mathbf{X}|\lambda) = \prod_{t=1}^{T} p(\mathbf{x}_t|\lambda)$.
- Função não linear de λ
- Estimar com o Expectation-Maximization (EM)

Expectation-Maximization

Estimar $\lambda^{(k+1)}$ a partir de λ^k

Expectation-Maximization

Estimar $\lambda^{(k+1)}$ a partir de λ^k

Obedecer $p(\mathbf{X}|\lambda^{(k+1)}) \geq p(\mathbf{X}|\lambda^{(k)})$

Expectation-Maximization

Estimar $\lambda^{(k+1)}$ a partir de λ^k

Obedecer $p(\mathbf{X}|\lambda^{(k+1)}) \geq p(\mathbf{X}|\lambda^{(k)})$

Calcular *E-Step* e *M-Step* para cada k até convergir

Expectation-Maximization

Estimar $\lambda^{(k+1)}$ a partir de λ^k

Obedecer $p(\mathbf{X}|\lambda^{(k+1)}) \geq p(\mathbf{X}|\lambda^{(k)})$

Calcular *E-Step* e *M-Step* para cada k até convergir

E-Step $P(i|\mathbf{x}_t) = \frac{w_i p_i(\mathbf{x}_t)}{\sum_{k=1}^{M} w_k p_k(\mathbf{x}_t)}$

Expectation-Maximization

Estimar $\lambda^{(k+1)}$ a partir de λ^k

Obedecer $p(\mathbf{X}|\lambda^{(k+1)}) \geq p(\mathbf{X}|\lambda^{(k)})$

Calcular *E-Step* e *M-Step* para cada k até convergir

E-Step $P(i|\mathbf{x}_t) = \frac{w_i p_i(\mathbf{x}_t)}{\sum_{k=1}^{M} w_k p_k(\mathbf{x}_t)}$

M-Step Adaptar os parâmetros

Expectation-Maximization

Estimar $\lambda^{(k+1)}$ a partir de λ^k

Obedecer $p(\mathbf{X}|\lambda^{(k+1)}) \geq p(\mathbf{X}|\lambda^{(k)})$

Calcular *E-Step* e *M-Step* para cada k até convergir

E-Step $P(i|\mathbf{x}_t) = \frac{w_i p_i(\mathbf{x}_t)}{\sum_{k=1}^{M} w_k p_k(\mathbf{x}_t)}$

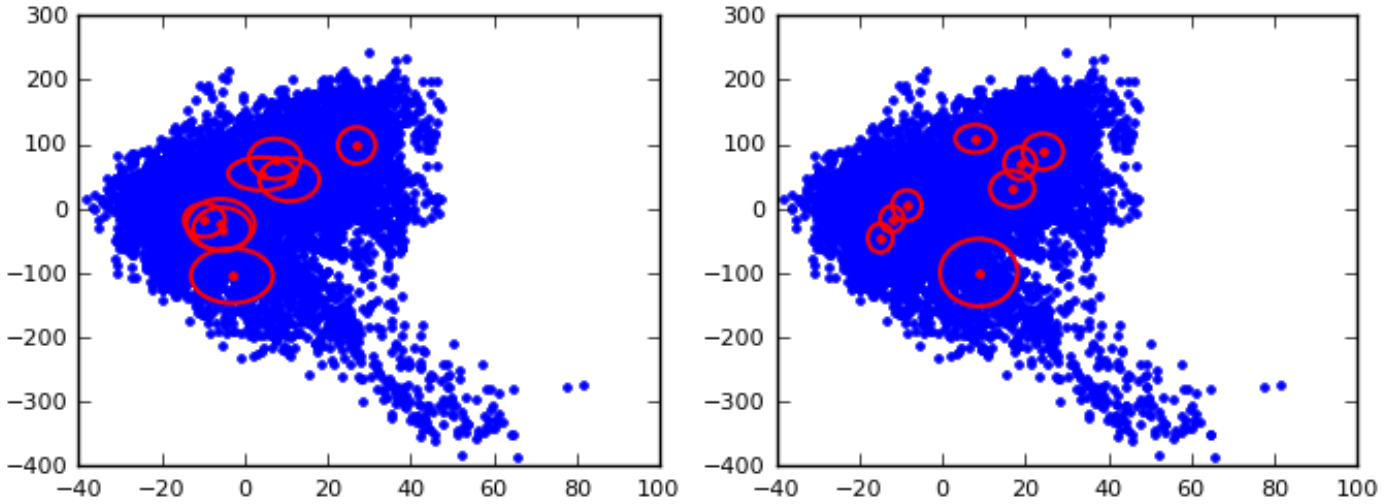
M-Step Adaptar os parâmetros

Pesos $\overline{w}_i = \frac{1}{T} \sum_{t=1}^{T} P(i|\mathbf{x}_t, \lambda)$

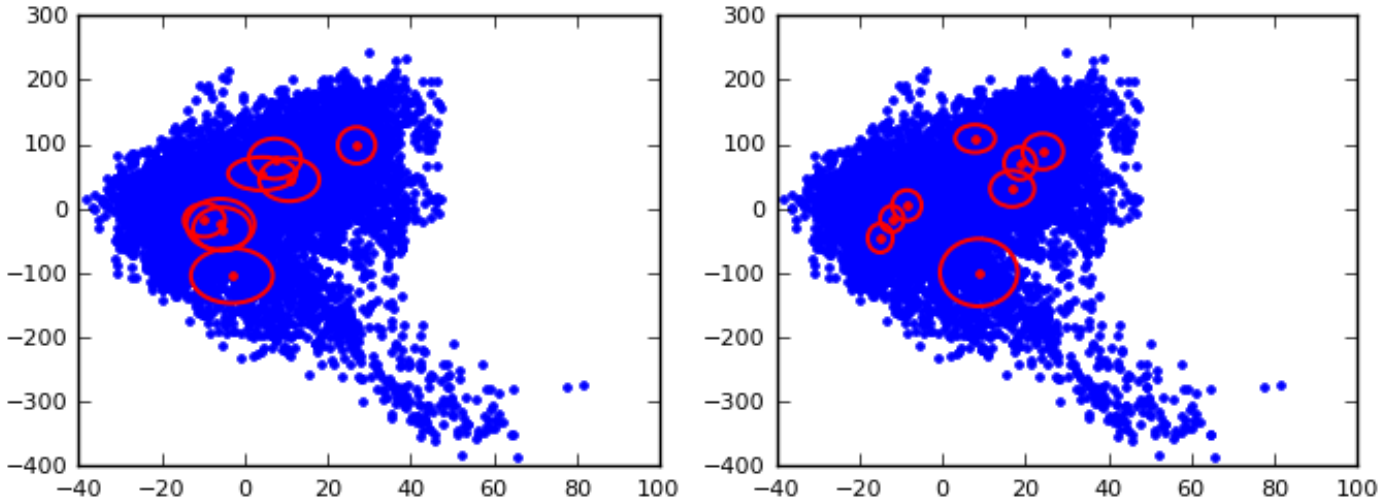
Médias $\overline{\boldsymbol{\mu}}_i = \frac{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)\mathbf{x}_t}{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)}$

Variâncias $\overline{\boldsymbol{\sigma}}_i^2 = \frac{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)\mathbf{x}_t^2}{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)} - \overline{\boldsymbol{\mu}}_i^2$

Expectation-Maximization

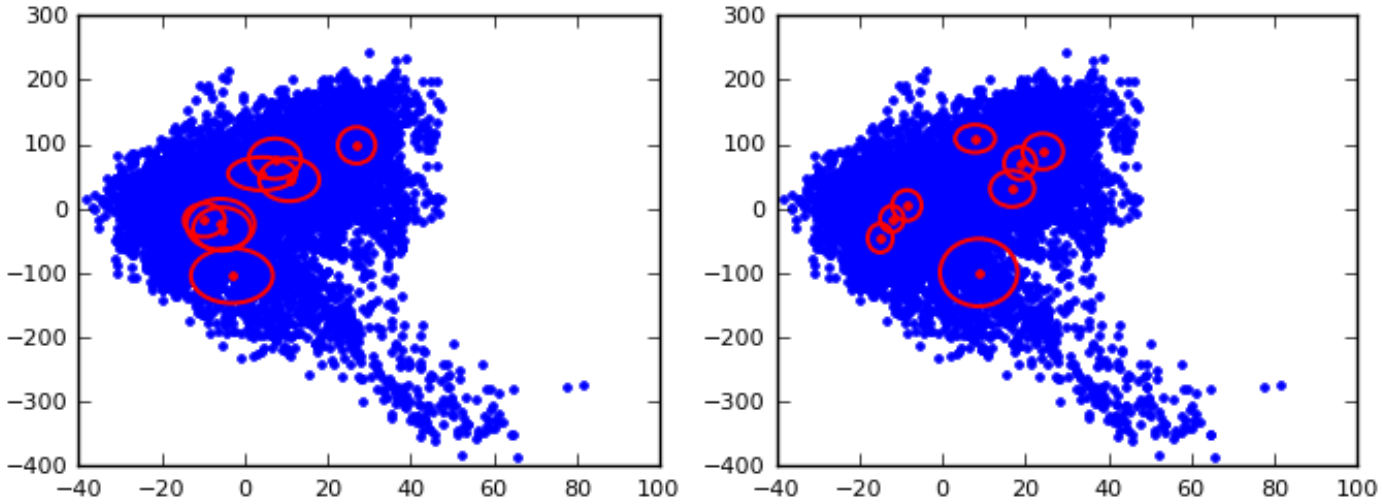


Expectation-Maximization



$M = 8$ e $\Delta = 0$

Expectation-Maximization



$M = 8$ e $\Delta = 0$

Inicialização *k-means* com 1 iteração

Limiar 10^{-3}

Universal Background Model

Utiliza locuções de todos os locutores registrados

Universal Background Model

Utiliza locuções de todos os locutores registrados

Realça características comuns

Universal Background Model

Utiliza locuções de todos os locutores registrados

Realça características comuns

$\mathbf{X}$ específico a $\mathcal{S} \implies \uparrow \Lambda(\mathbf{X})$

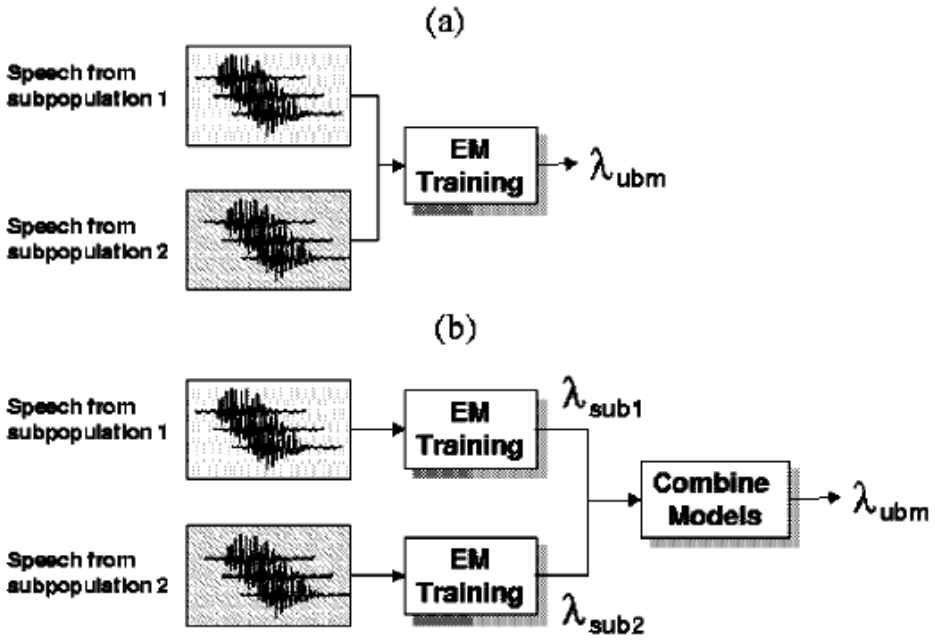
Universal Background Model

Utiliza locuções de todos os locutores registrados

Realça características comuns

$\mathbf{X}$ específico a $\mathcal{S} \implies \uparrow \Lambda(\mathbf{X})$

Escolhido o tipo (b)



Adapted Gaussian Mixture Model

Adaptação λ_{bkg} treinado $\Longrightarrow$ λ_j para cada $\mathcal{S}_j$

Adapted Gaussian Mixture Model

Adaptação λ_{bkg} treinado $\Longrightarrow$ λ_j para cada $\mathcal{S}_j$

Modelagem mais rápida que EM

Adapted Gaussian Mixture Model

Adaptação λ_{bkg} treinado $\Longrightarrow$ λ_j para cada $\mathcal{S}_j$

Modelagem mais rápida que EM

Composto de *E-Step* e *MAP-Step*

Adapted Gaussian Mixture Model

Adaptação λ_{bkg} treinado $\Longrightarrow$ λ_j para cada $\mathcal{S}_j$

Modelagem mais rápida que EM

Composto de *E-Step* e *MAP-Step*

E-Step Semelhante ao *E-Step* do EM

Adapted Gaussian Mixture Model

Adaptação λ_{bkg} treinado $\Longrightarrow$ λ_j para cada $\mathcal{S}_j$

Modelagem mais rápida que EM

Composto de *E-Step* e *MAP-Step*

E-Step Semelhante ao *E-Step* do EM

- $n_i = \sum_{t=1}^{T} P(i|\mathbf{x}_t)$
- $E_i(\mathbf{x}) = \frac{1}{n_i} \sum_{t=1}^{T} P(i|\mathbf{x}_t)\mathbf{x}_t$
- $E_i(\mathbf{x}^2) = \frac{1}{n_i} \sum_{t=1}^{T} P(i|\mathbf{x}_t)\mathbf{x}_t^2$

Adapted Gaussian Mixture Model

Adaptação λ_{bkg} treinado $\Longrightarrow$ λ_j para cada $\mathcal{S}_j$

Modelagem mais rápida que EM

Composto de *E-Step* e *MAP-Step*

E-Step Semelhante ao *E-Step* do EM

- $n_i = \sum_{t=1}^{T} P(i|\mathbf{x}_t)$
- $E_i(\mathbf{x}) = \frac{1}{n_i} \sum_{t=1}^{T} P(i|\mathbf{x}_t)\mathbf{x}_t$
- $E_i(\mathbf{x}^2) = \frac{1}{n_i} \sum_{t=1}^{T} P(i|\mathbf{x}_t)\mathbf{x}_t^2$

MAP-Step Adapta os parâmetros

Adapted Gaussian Mixture Model

Adaptação λ_{bkg} treinado $\Longrightarrow$ λ_j para cada $\mathcal{S}_j$

Modelagem mais rápida que EM

Composto de *E-Step* e *MAP-Step*

E-Step Semelhante ao *E-Step* do EM

- $n_i = \sum_{t=1}^{T} P(i|\mathbf{x}_t)$
- $E_i(\mathbf{x}) = \frac{1}{n_i} \sum_{t=1}^{T} P(i|\mathbf{x}_t)\mathbf{x}_t$
- $E_i(\mathbf{x}^2) = \frac{1}{n_i} \sum_{t=1}^{T} P(i|\mathbf{x}_t)\mathbf{x}_t^2$

MAP-Step Adapta os parâmetros

Pesos $\hat{w}_i = [\alpha_i n_i / T + (1 - \alpha_i) w_i]\gamma$

Médias $\hat{\boldsymbol{\mu}}_i = \alpha_i E_i(\mathbf{x}) + (1 - \alpha_i)\boldsymbol{\mu}_i$

Variâncias $\hat{\boldsymbol{\sigma}}_i^2 = \alpha_i E_i(\mathbf{x}^2) + (1 - \alpha_i)(\boldsymbol{\sigma}_i^2 + \boldsymbol{\mu}_i^2) - \hat{\boldsymbol{\mu}}_i^2$

Adapted Gaussian Mixture Model

γ normaliza os pesos

Adapted Gaussian Mixture Model

γ normaliza os pesos

Coeficiente $\alpha_i = \frac{n_i}{n_i + r}$

Adapted Gaussian Mixture Model

γ normaliza os pesos

Coeficiente $\alpha_i = \frac{n_i}{n_i + r}$

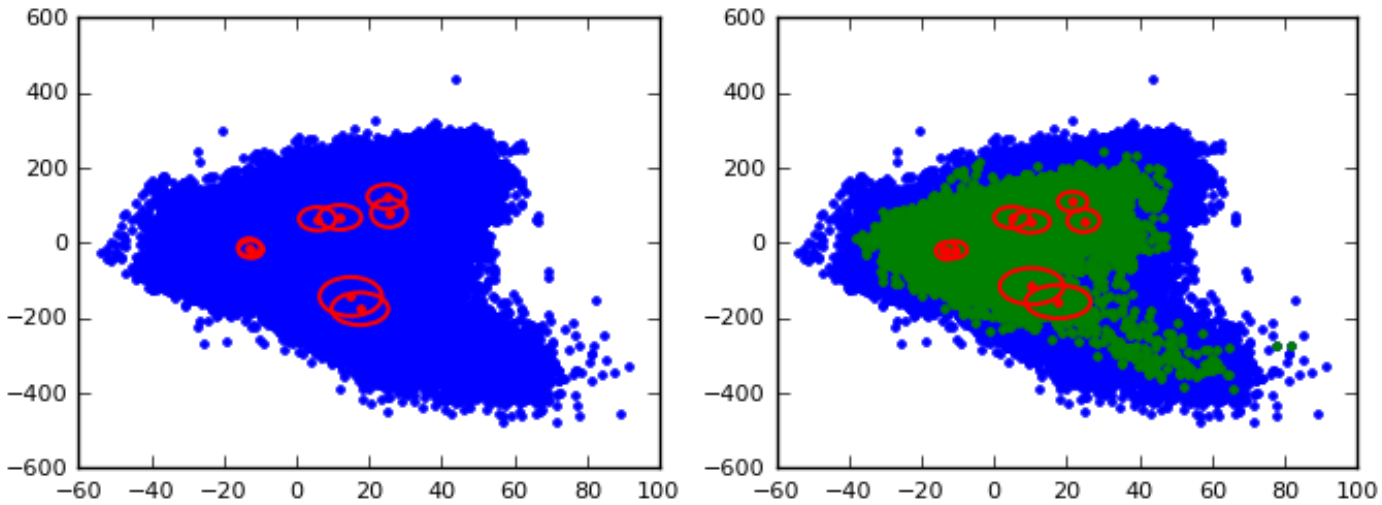
- $\alpha_i \to 0 \implies$ manter os antigos parâmetros
- $\alpha_i \to 1 \implies$ adaptar para os novos parâmetros

Adapted Gaussian Mixture Model

γ normaliza os pesos

Coeficiente $\alpha_i = \frac{n_i}{n_i + r}$

- $\alpha_i \to 0 \implies$ manter os antigos parâmetros
- $\alpha_i \to 1 \implies$ adaptar para os novos parâmetros



Pesos, médias e variâncias adaptados

Fractional Gaussian Mixture Model

GMM com $\boldsymbol{\Sigma}$ fracionário

Fractional Gaussian Mixture Model

GMM com $\mathbf{\Sigma}$ fracionário

- $\sigma^2 = E[(X^r - \mu^r)^2]$
- $\overline{\boldsymbol{\sigma}}_i^2 = \frac{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)(\mathbf{x}_t^r - \overline{\boldsymbol{\mu}}_i^r)^2}{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)}$

Fractional Gaussian Mixture Model

GMM com $\mathbf{\Sigma}$ fracionário

- $\sigma^2 = E[(X^r - \mu^r)^2]$
- $\overline{\boldsymbol{\sigma}}_i^2 = \frac{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)(\mathbf{x}_t^r - \overline{\boldsymbol{\mu}}_i^r)^2}{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)}$

Problema $\mathbb{C}$

Fractional Gaussian Mixture Model

GMM com $\mathbf{\Sigma}$ fracionário

- $\sigma^2 = E[(X^r - \mu^r)^2]$
- $\overline{\boldsymbol{\sigma}}_i^2 = \frac{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)(\mathbf{x}_t^r - \overline{\boldsymbol{\mu}}_i^r)^2}{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)}$

Problema $\mathbb{C}$

Antes $c_n = c_n + (1 - \min_t c_{n,t})$

Depois $\mu_n = \mu_n - (1 - \min_t c_{n,t})$

Fractional Gaussian Mixture Model

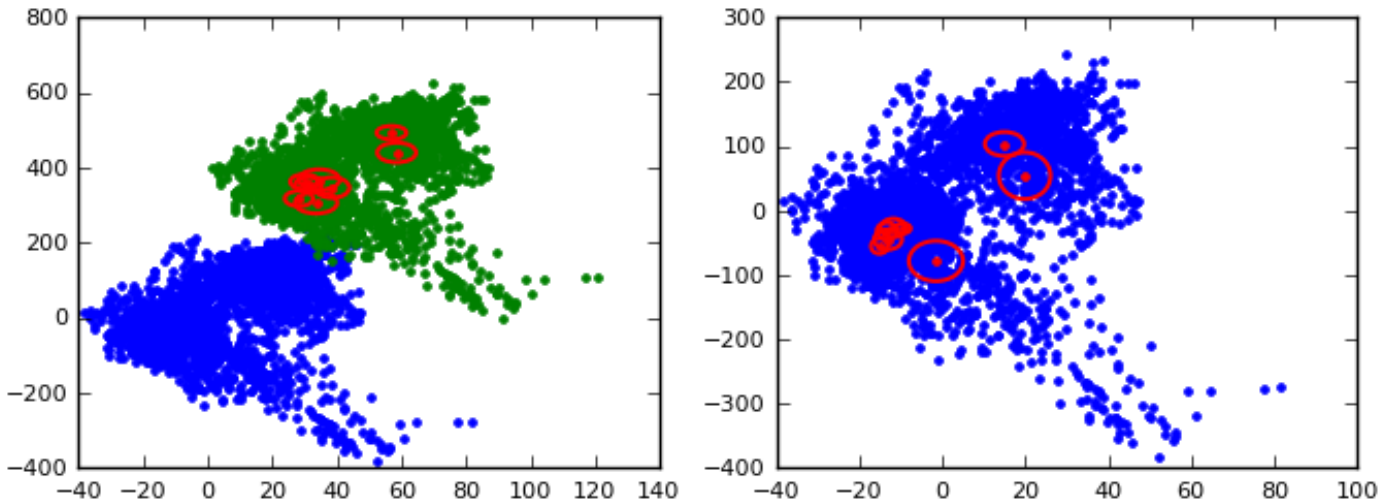
GMM com $\mathbf{\Sigma}$ fracionário

- $\sigma^2 = E[(X^r - \mu^r)^2]$
- $\overline{\boldsymbol{\sigma}}_i^2 = \frac{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)(\mathbf{x}_t^r - \overline{\boldsymbol{\mu}}_i^r)^2}{\sum_{t=1}^{T} P(i|\mathbf{x}_t,\lambda)}$

Problema $\mathbb{C}$

Antes $c_n = c_n + (1 - \min_t c_{n,t})$

Depois $\mu_n = \mu_n - (1 - \min_t c_{n,t})$



Conteúdo

1 Introdução

2 Sistemas de Reconhecimento de Locutor

3 Extração de Características

4 Modelos de Mistura Gaussianas

5 Experimentos

6 Conclusão

Corpus

Base MIT Mobile Device Speaker Verification Corpus

Corpus

Base MIT Mobile Device Speaker Verification Corpus

54 locuções/locutor em 3 níveis de ruído

Corpus

Base MIT Mobile Device Speaker Verification Corpus

54 locuções/locutor em 3 níveis de ruído

- **Baixo** Escritório calmo
- **Médio** Saguão de edifício
- **Alto** Cruzamento movimentado

Corpus

Base MIT Mobile Device Speaker Verification Corpus

54 locuções/locutor em 3 níveis de ruído

- **Baixo** Escritório calmo
- **Médio** Saguão de edifício
- **Alto** Cruzamento movimentado

3 sessões distintas

- **Enroll 1** Treinamento dos modelos
- **Enroll 2** Teste de detecção (e identificação)
- **Imposter** Teste de rejeição

Corpus

Base MIT Mobile Device Speaker Verification Corpus

54 locuções/locutor em 3 níveis de ruído

- **Baixo** Escritório calmo
- **Médio** Saguão de edifício
- **Alto** Cruzamento movimentado

3 sessões distintas

- **Enroll 1** Treinamento dos modelos
- **Enroll 2** Teste de detecção (e identificação)
- **Imposter** Teste de rejeição

Session	Training	Test	#female	#male
Enroll 1	**X**		22	26
Enroll 2		**X**	22	26
Imposter		**X**	17	23

Codificação

Linguagem Python 3.4.3

Codificação

Linguagem Python 3.4.3

Frameworks NumPy 1.8.1, SciPy 0.14.0 e Matplotlib 1.4

Codificação

Linguagem Python 3.4.3

Frameworks NumPy 1.8.1, SciPy 0.14.0 e Matplotlib 1.4

Parâmetros A implementação utilizou:

Codificação

Linguagem Python 3.4.3

Frameworks NumPy 1.8.1, SciPy 0.14.0 e Matplotlib 1.4

Parâmetros A implementação utilizou:

- # filtros $= 26$
- # coefficientes $= 19$
- Δs de ordem 0, 1 e 2, com $K = 2$
- Energy appending e CMS
- $r = 16$ para AGMM
- *threshold* $= 10^{-3}$ no EM
- $M = 8, 16, 32, 64, 128$

Percalços

Inicialização Em 2 passos

Percalços

Inicialização Em 2 passos

- Escolha de médias aleatórias
- *k-means* $\Longrightarrow$ Novas médias + pesos e variâncias

Percalços

Inicialização Em 2 passos

- Escolha de médias aleatórias
- *k-means* $\Longrightarrow$ Novas médias + pesos e variâncias

Variâncias Podem reduzir significativamente

Percalços

Inicialização Em 2 passos

- Escolha de médias aleatórias
- *k-means* $\Longrightarrow$ Novas médias + pesos e variâncias

Variâncias Podem reduzir significativamente

- $\sigma^2_{min} = 0.01$
- $\sigma^2 < \sigma^2_{min} \Longrightarrow \sigma^2 \leftarrow \sigma^2_{min}$

Percalços

Inicialização Em 2 passos

- Escolha de médias aleatórias
- *k-means* $\Longrightarrow$ Novas médias + pesos e variâncias

Variâncias Podem reduzir significativamente

- $\sigma^2_{min} = 0.01$
- $\sigma^2 < \sigma^2_{min} \Longrightarrow \sigma^2 \leftarrow \sigma^2_{min}$

Monotonic FGMM viola $\log p(\mathbf{X}|\lambda^{(k+1)}) \geq \log p(\mathbf{X}|\lambda^k)$

Percalços

Inicialização Em 2 passos

- Escolha de médias aleatórias
- *k-means* $\Longrightarrow$ Novas médias + pesos e variâncias

Variâncias Podem reduzir significativamente

- $\sigma^2_{min} = 0.01$
- $\sigma^2 < \sigma^2_{min} \Longrightarrow \sigma^2 \leftarrow \sigma^2_{min}$

Monotonic FGMM viola $\log p(\mathbf{X}|\lambda^{(k+1)}) \geq \log p(\mathbf{X}|\lambda^k)$

- $|1 - r| \Longrightarrow \downarrow$ estimação

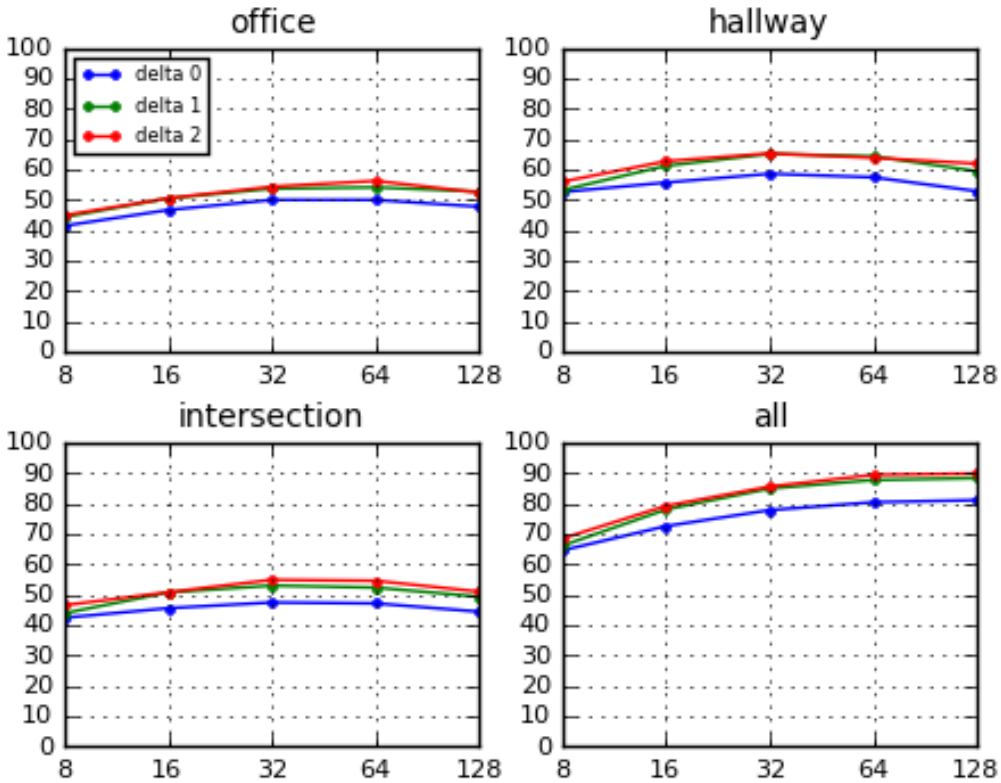
Identificação

SSGMM Single Speaker GMM

Δ	M	Office	Hallway	Intersection	All
0	8	41.55	52.66	42.48	64.66
	16	46.76	55.79	45.64	72.65
	32	50.08	58.68	47.53	77.93
	64	50.08	57.52	47.22	80.52
	128	47.84	52.93	44.48	81.21
1	8	44.41	53.28	43.98	66.20
	16	50.58	61.30	50.81	78.12
	32	53.78	65.20	53.09	85.03
	64	54.21	64.43	52.43	87.85
	128	52.82	59.53	49.42	88.46
2	8	45.02	56.06	46.60	68.56
	16	50.62	62.81	50.89	79.32
	32	54.44	65.39	54.98	85.69
	64	56.33	63.93	54.67	89.54
	128	52.47	62.00	51.08	89.97

Identificação

SSGMM Single Speaker GMM



Identificação

SSFGMM Single Speaker FGMM

Identificação

SSFGMM Single Speaker FGMM

$$r = r_0 + (-1)^u \delta$$

Identificação

SSFGMM Single Speaker FGMM

$$r = r_0 + (-1)^u \delta$$

- $r_0 = 1,\ u \in \{0, 1\}$
- $\delta \in \{0.01, 0.05\}$

Identificação

SSFGMM Single Speaker FGMM

$$r = r_0 + (-1)^u \delta$$

- $r_0 = 1,\ u \in \{0, 1\}$
- $\delta \in \{0.01, 0.05\}$

$\uparrow |1 - r| \implies \downarrow$ representação

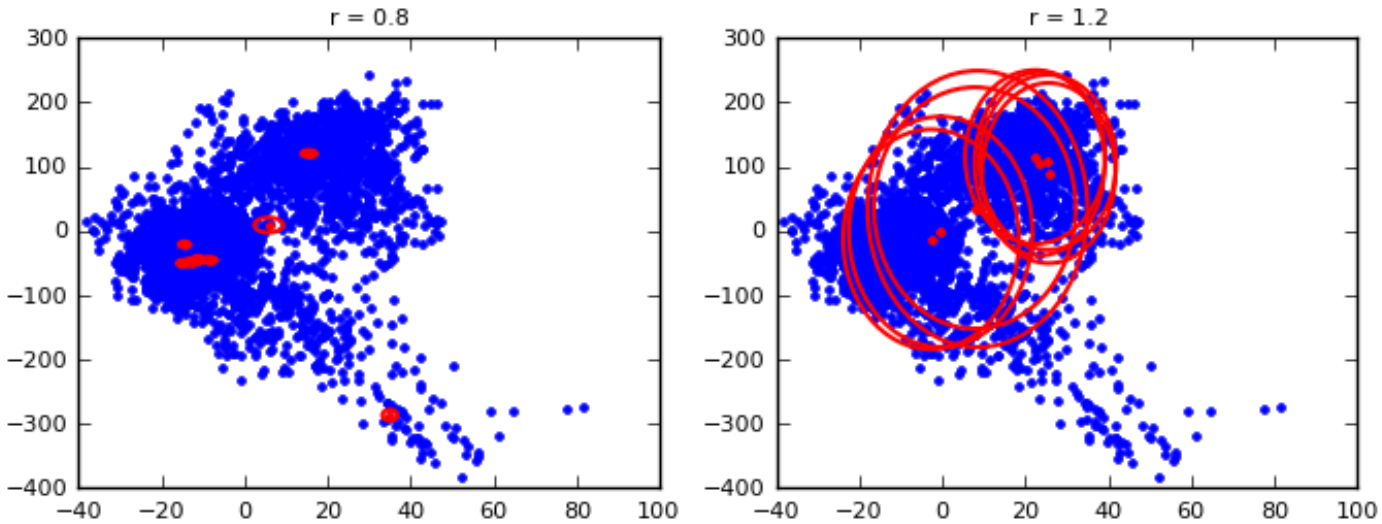
Identificação

SSFGMM Single Speaker FGMM

$$r = r_0 + (-1)^u \delta$$

- $r_0 = 1,\ u \in \{0, 1\}$
- $\delta \in \{0.01, 0.05\}$

$\uparrow |1 - r| \implies \downarrow$ representação



Identificação

SSFGMM Single Speaker FGMM

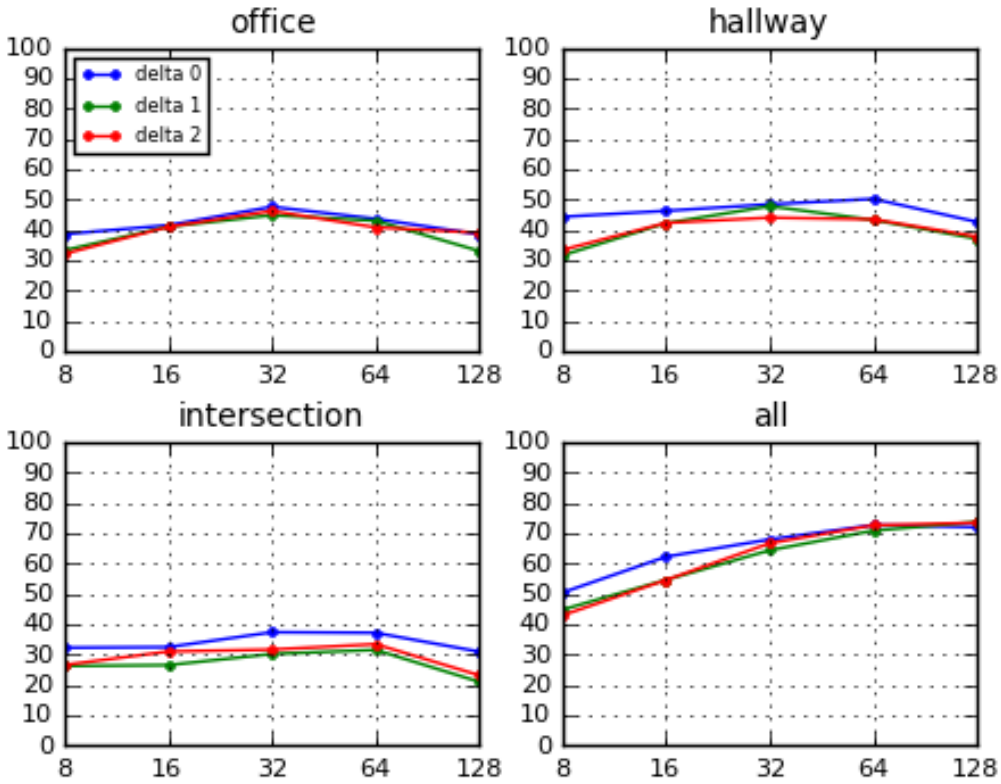
$r = 0.95$

Δ	M	Office	Hallway	Intersection	All
0	8	38.70	44.41	32.37	50.50
	16	41.63	46.37	32.56	62.35
	32	47.72	48.53	37.46	68.06
	64	43.75	50.31	37.27	72.80
	128	38.62	42.75	31.06	72.15
1	8	33.37	31.67	26.35	44.87
	16	41.13	42.32	26.62	54.71
	32	44.95	47.92	30.29	64.47
	64	43.13	43.36	31.64	70.95
	128	33.14	37.15	21.10	73.84
2	8	32.21	33.49	26.66	43.02
	16	41.09	42.40	31.10	54.67
	32	46.33	44.14	31.75	66.78
	64	40.93	43.60	33.53	72.72
	128	39.16	37.89	23.26	73.53

Identificação

SSFGMM Single Speaker FGMM

$r = 0.95$



Identificação

SSFGMM Single Speaker FGMM

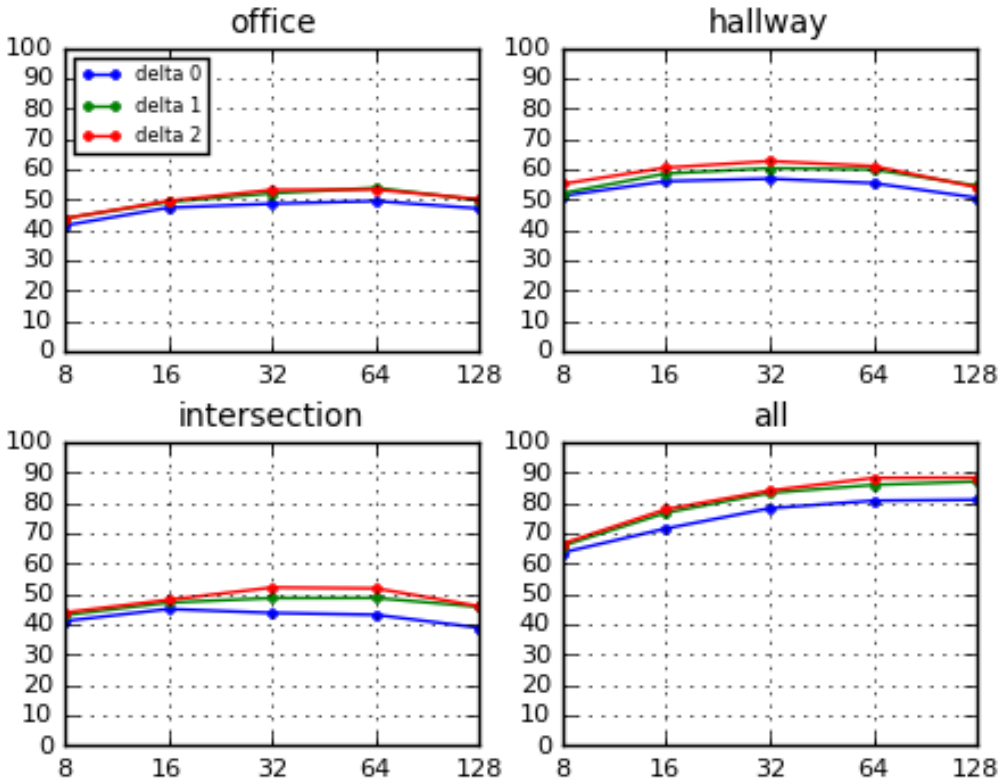
$r = 0.99$

Δ	M	Office	Hallway	Intersection	All
0	8	41.55	51.31	41.13	63.70
	16	47.42	56.13	45.10	71.64
	32	48.73	56.98	43.83	78.32
	64	49.61	55.52	43.21	80.83
	128	47.15	50.69	38.93	81.13
1	8	43.90	52.16	43.09	65.90
	16	49.31	58.68	47.22	76.85
	32	52.16	60.42	48.73	83.37
	64	53.94	60.03	48.77	86.03
	128	49.88	54.63	45.83	87.15
2	8	43.87	55.25	43.94	66.63
	16	49.65	60.61	48.11	77.97
	32	53.28	62.77	52.20	84.14
	64	53.40	61.11	51.93	88.31
	128	50.23	54.17	46.03	88.43

Identificação

SSFGMM Single Speaker FGMM

$r = 0.99$



Identificação

SSFGMM Single Speaker FGMM

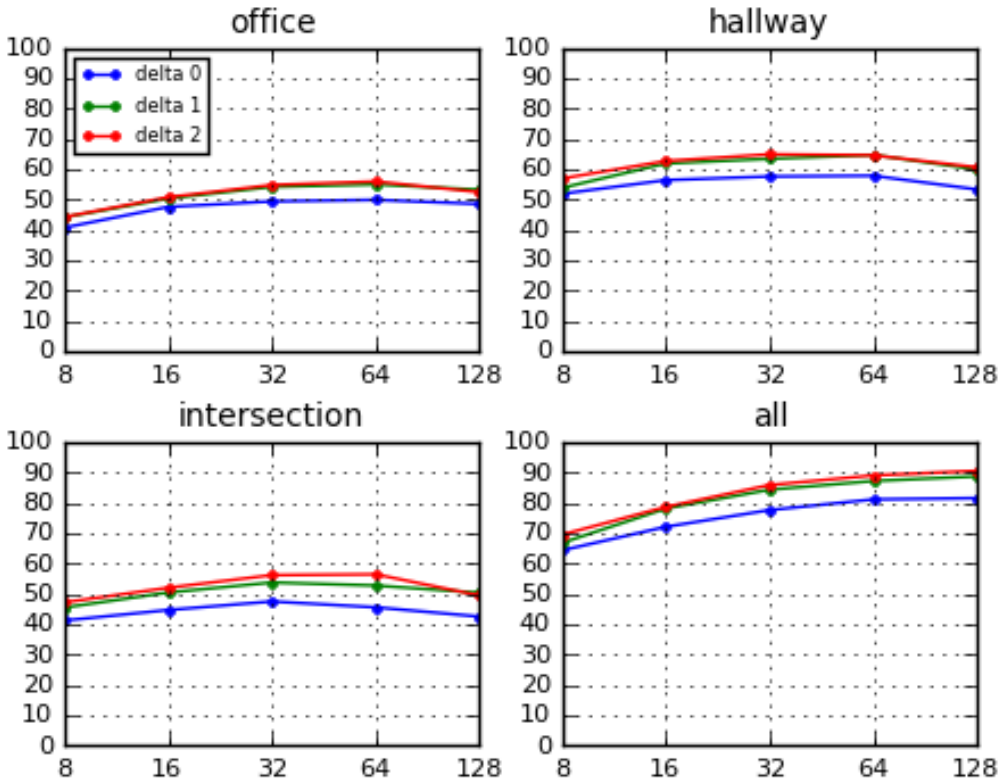
$$r = 1$$

Δ	M	Office	Hallway	Intersection	All
0	8	40.86	52.01	41.32	64.47
	16	47.69	56.52	44.79	72.22
	32	49.50	57.72	47.61	77.74
	64	50.00	57.95	45.68	81.25
	128	48.65	53.43	42.63	81.67
1	8	44.25	53.97	45.60	66.94
	16	50.42	62.00	50.54	78.24
	32	54.28	63.54	53.86	84.45
	64	55.09	64.81	52.85	87.31
	128	53.32	59.99	50.46	88.85
2	8	44.37	57.06	47.30	69.64
	16	50.89	62.81	52.12	78.78
	32	54.90	65.01	56.29	86.00
	64	56.06	64.70	56.56	89.16
	128	52.55	60.73	49.58	90.66

Identificação

SSFGMM Single Speaker FGMM

$r = 1$



Identificação

SSFGMM Single Speaker FGMM

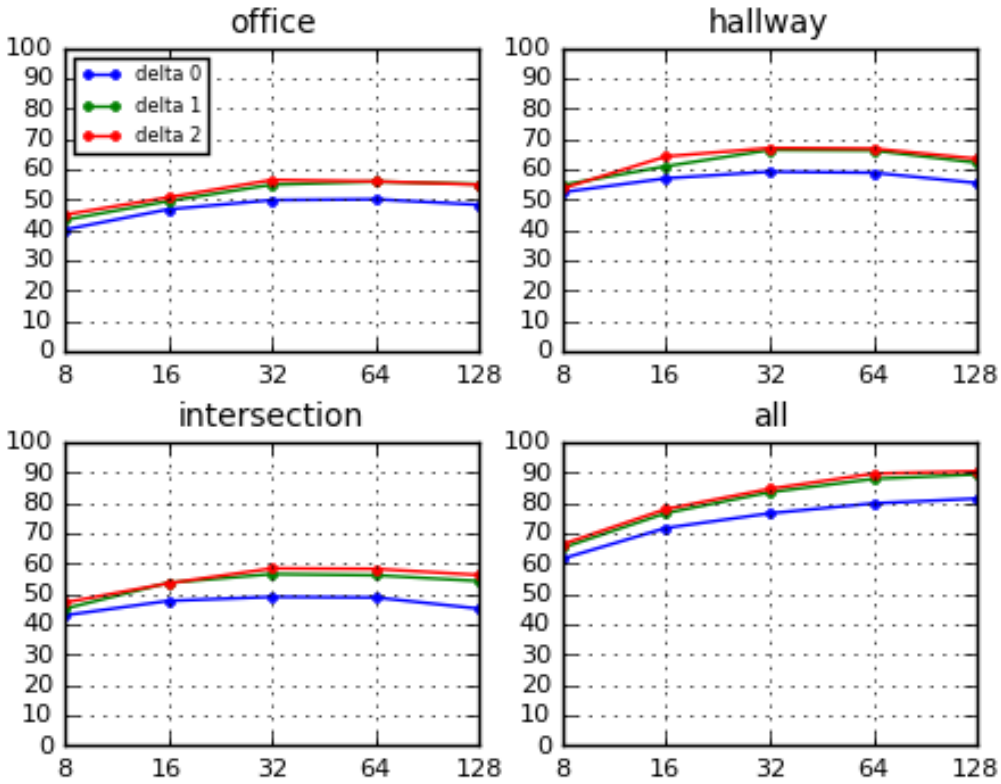
$r = 1.01$

Δ	M	Office	Hallway	Intersection	All
0	8	40.16	52.51	43.02	61.69
	16	46.88	57.10	47.80	71.84
	32	49.92	59.30	49.11	76.66
	64	50.19	58.95	48.92	79.94
	128	48.38	55.56	45.22	81.52
1	8	43.36	54.90	45.18	65.28
	16	49.58	61.07	53.74	76.74
	32	55.02	66.44	56.64	83.60
	64	56.02	66.28	56.25	88.00
	128	55.17	62.23	54.32	89.51
2	8	45.10	53.74	47.22	66.44
	16	50.81	64.31	53.59	78.05
	32	56.56	67.09	58.49	84.72
	64	56.10	66.90	58.33	89.74
	128	55.02	63.54	56.33	90.55

Identificação

SSFGMM Single Speaker FGMM

$r = 1.01$



Identificação

SSFGMM Single Speaker FGMM

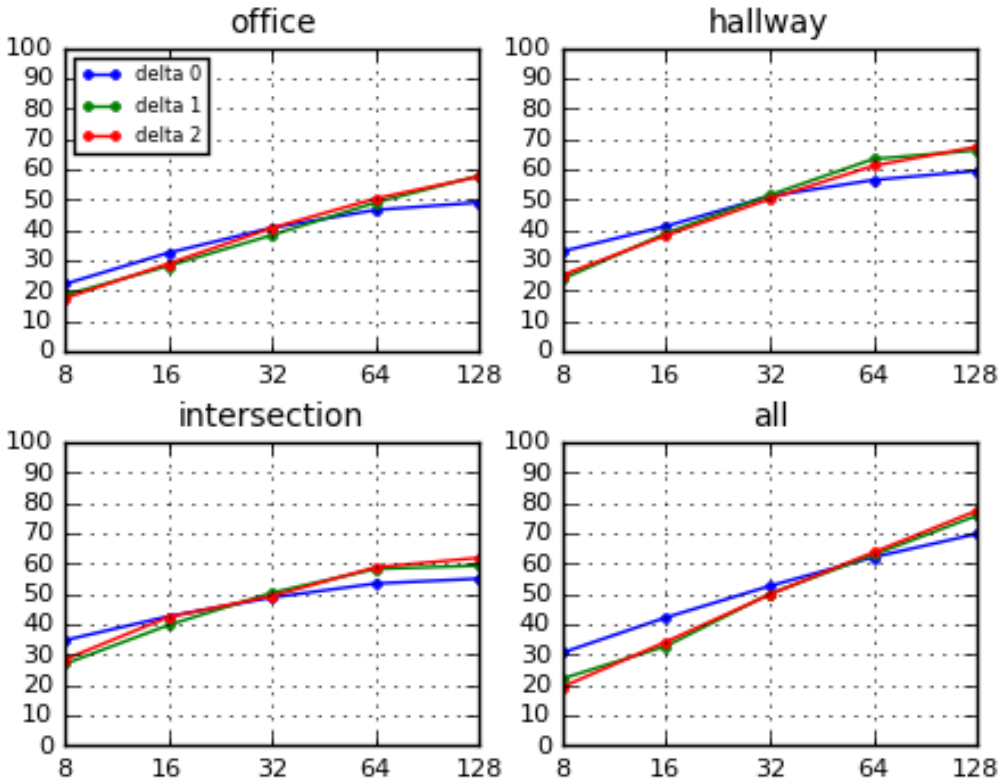
$r = 1.05$

Δ	M	Office	Hallway	Intersection	All
0	8	22.22	33.02	34.80	30.71
	16	32.52	41.32	42.67	42.32
	32	40.78	51.20	48.92	52.70
	64	46.68	56.56	53.51	62.19
	128	49.15	59.57	55.13	69.91
1	8	18.56	23.88	26.97	22.15
	16	28.20	39.00	39.78	32.87
	32	38.39	51.58	50.46	50.08
	64	49.11	63.46	58.33	62.96
	128	57.99	66.32	59.41	75.96
2	8	17.52	25.15	28.43	19.41
	16	28.94	38.27	42.44	34.30
	32	40.74	50.31	49.38	49.88
	64	50.42	61.23	58.87	63.77
	128	57.68	67.52	62.00	77.55

Identificação

SSFGMM Single Speaker FGMM

$r = 1.05$



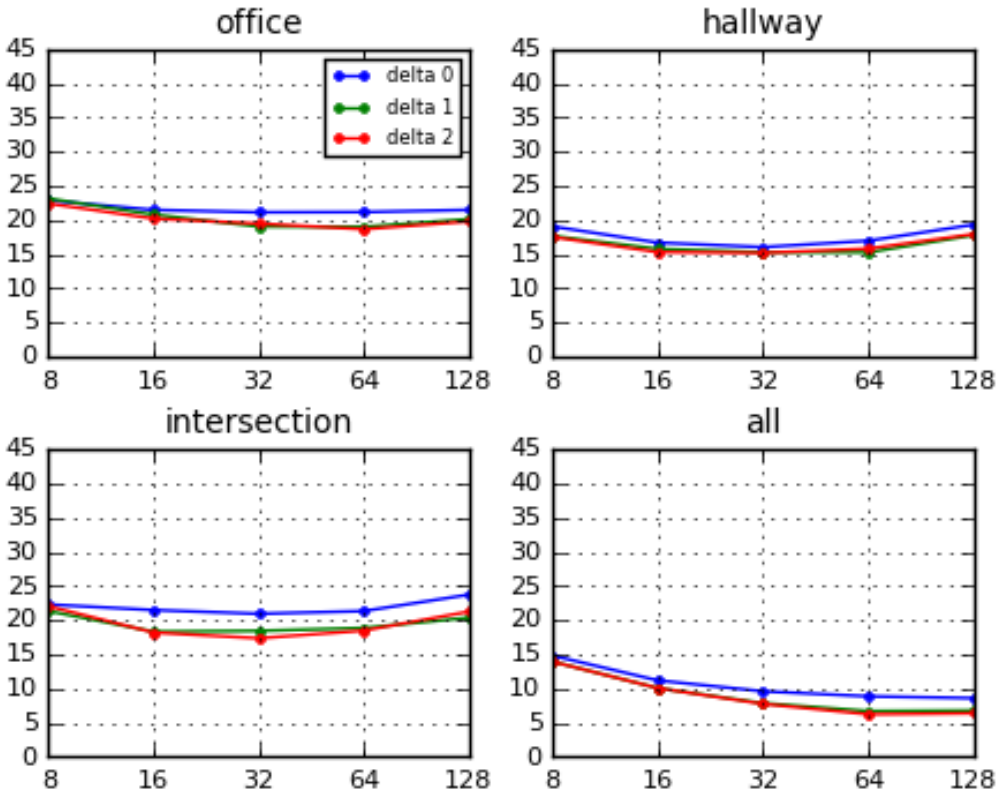
Verificação

SSGMM Single Speaker GMM

Δ	M	Office	Hallway	Intersection	All
0	8	22.88	19.06	22.30	14.81
	16	21.49	16.71	21.49	11.19
	32	21.14	16.05	20.94	9.61
	64	21.18	16.98	21.34	8.87
	128	21.49	19.33	23.74	8.60
1	8	23.15	17.67	21.34	13.93
	16	20.80	15.78	18.33	10.07
	32	19.06	15.31	18.45	7.87
	64	19.02	15.28	18.87	6.72
	128	20.14	17.79	20.37	6.79
2	8	22.42	17.52	22.03	13.92
	16	20.22	15.32	18.20	10.06
	32	19.48	15.20	17.36	7.75
	64	18.67	15.82	18.48	6.25
	128	19.80	17.94	21.26	6.40

Verificação

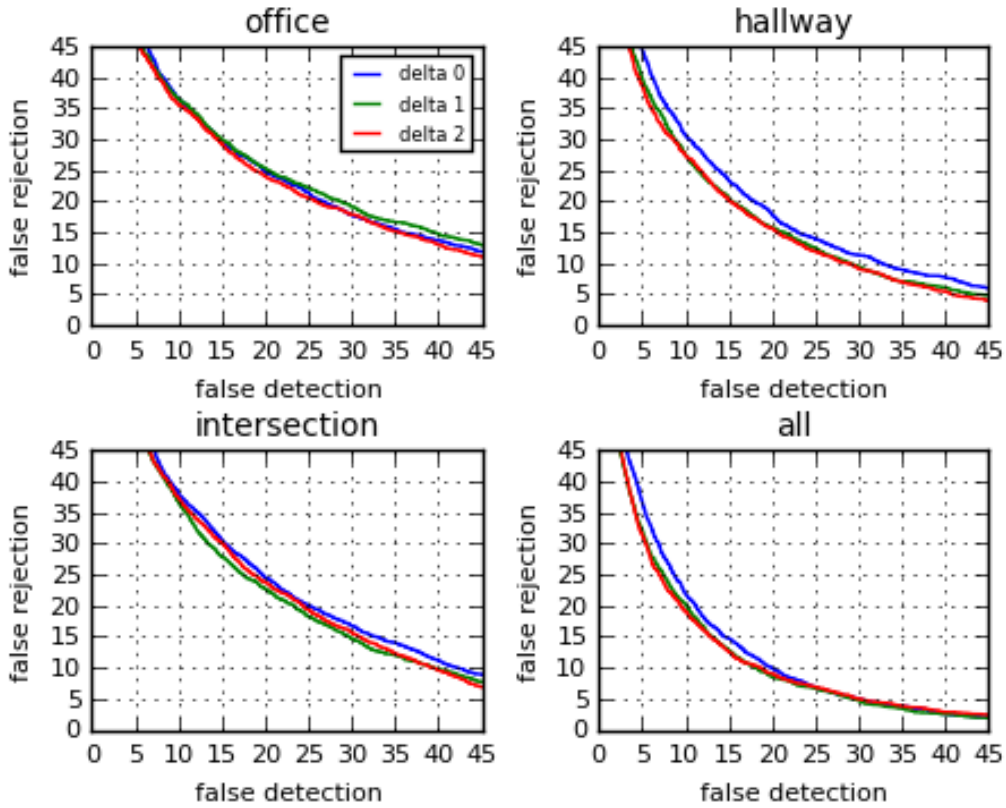
SSGMM Single Speaker GMM



Verificação

SSGMM Single Speaker GMM

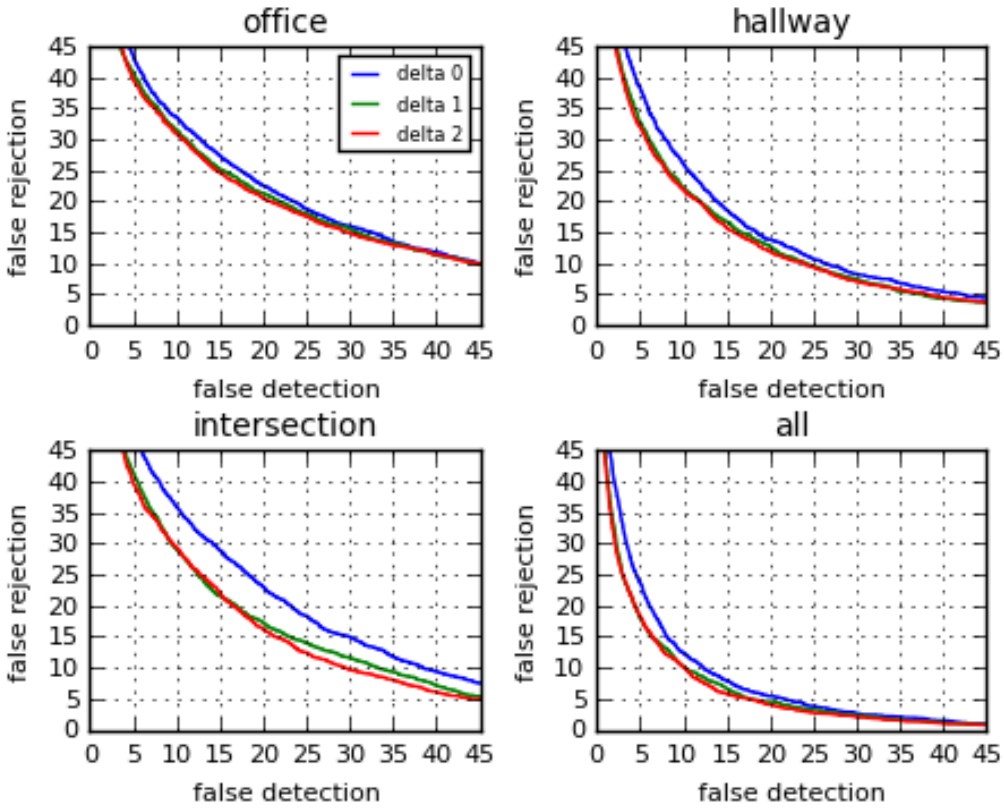
$M = 8$



Verificação

SSGMM Single Speaker GMM

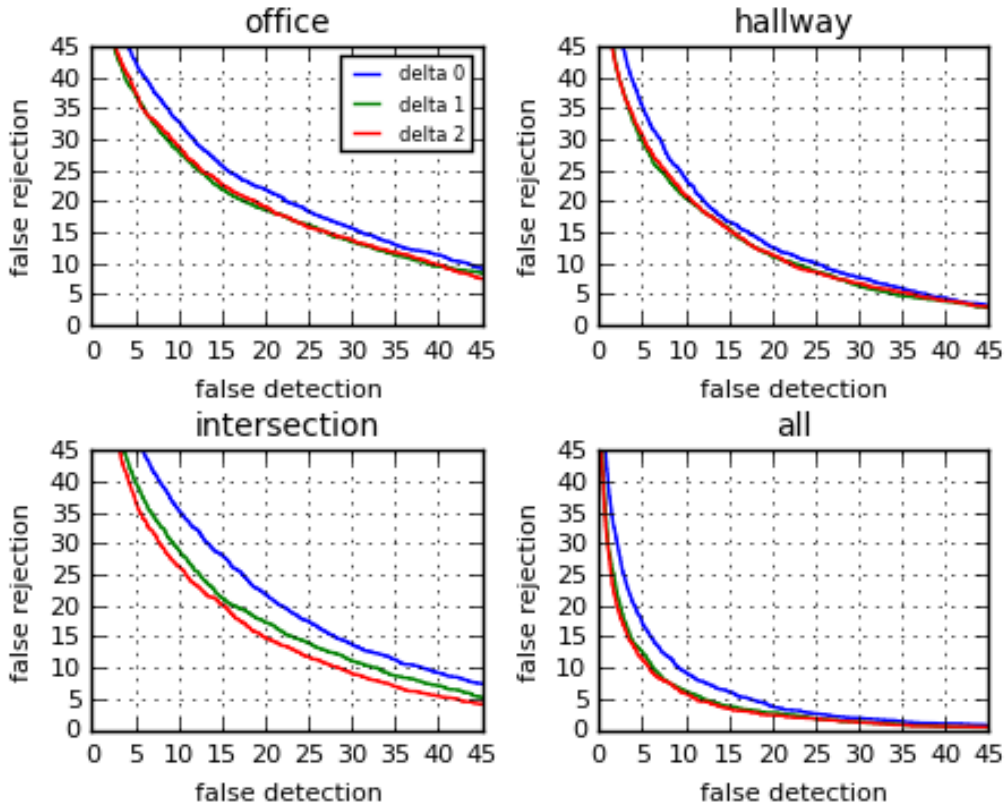
$M = 16$



Verificação

SSGMM Single Speaker GMM

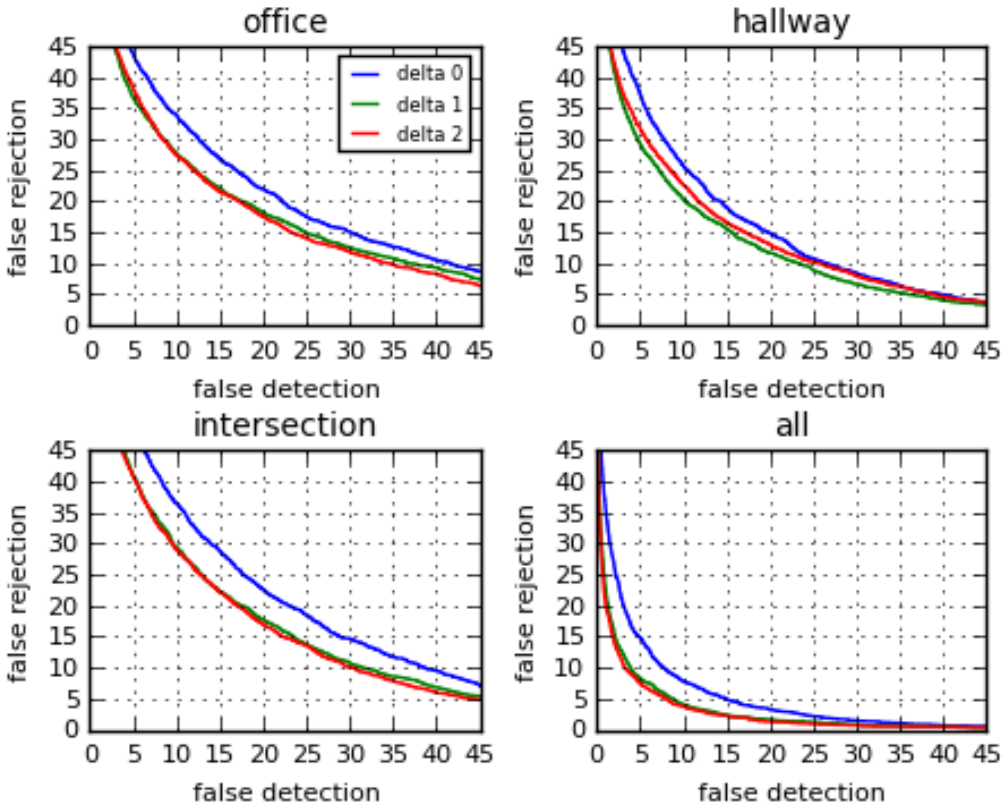
$M = 32$



Verificação

SSGMM Single Speaker GMM

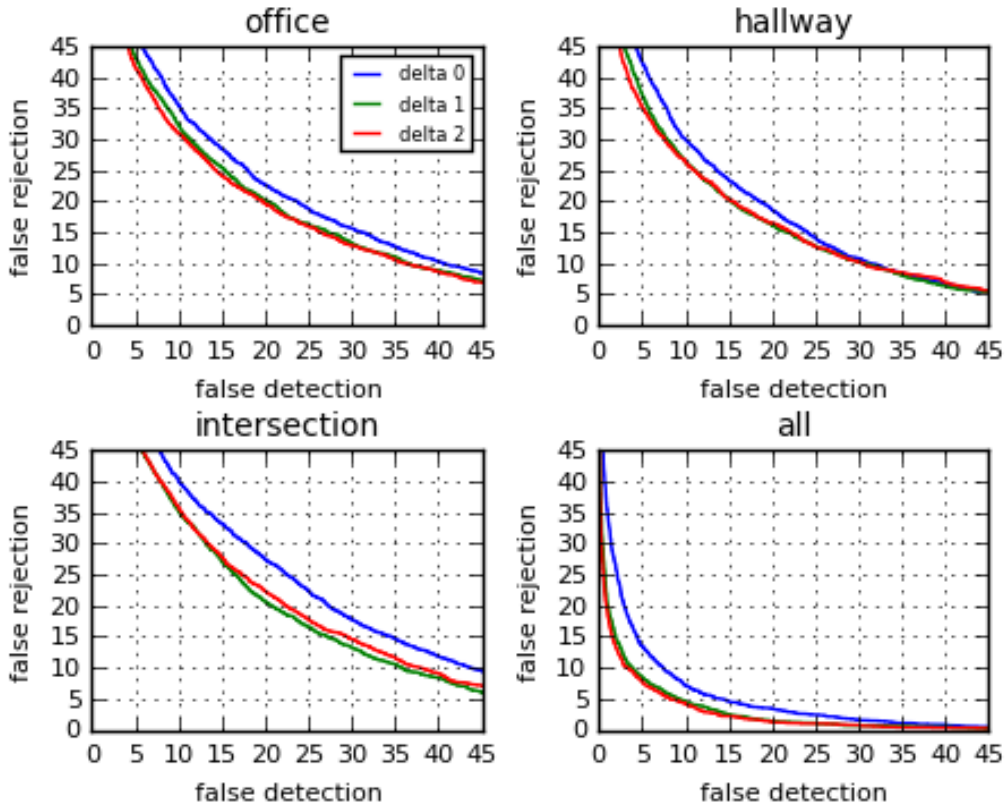
$M = 64$



Verificação

SSGMM Single Speaker GMM

$M = 128$



Verificação

SSAGMM Single Speaker AGMM

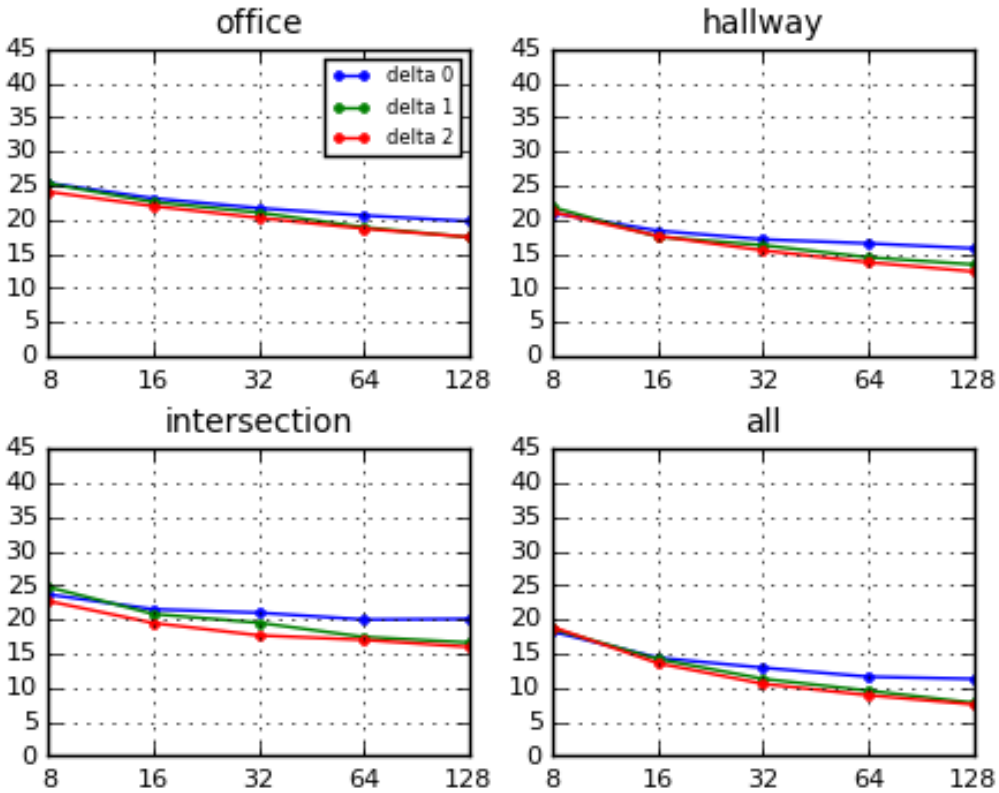
Adaptação Médias

Δ	M	Office	Hallway	Intersection	All
0	8	25.38	21.00	23.66	18.21
	16	23.14	18.40	21.49	14.39
	32	21.71	17.13	20.99	12.93
	64	20.64	16.55	19.98	11.61
	128	19.79	15.82	20.07	11.29
1	8	25.31	21.83	24.73	18.59
	16	22.61	17.52	20.80	14.20
	32	21.07	16.28	19.52	11.30
	64	18.90	14.51	17.44	9.58
	128	17.44	13.46	16.62	7.80
2	8	24.11	21.13	22.68	18.87
	16	21.99	17.63	19.47	13.59
	32	20.29	15.51	17.67	10.57
	64	18.71	13.77	17.01	8.91
	128	17.48	12.43	15.97	7.56

Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias

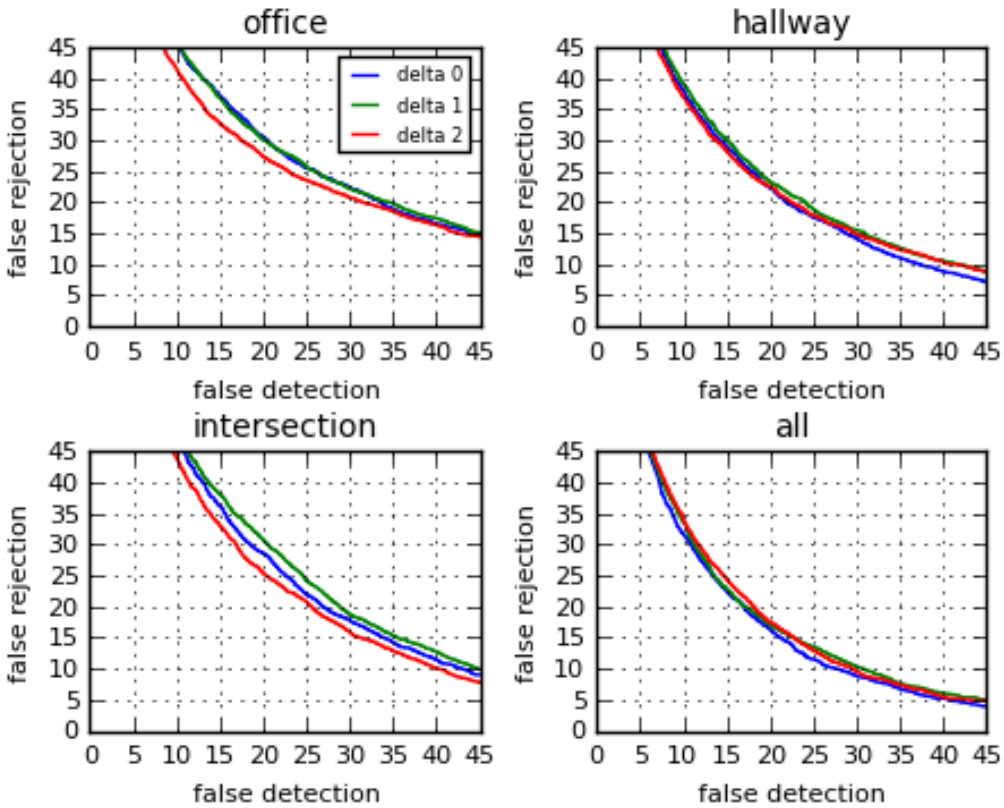


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias

$M = 8$

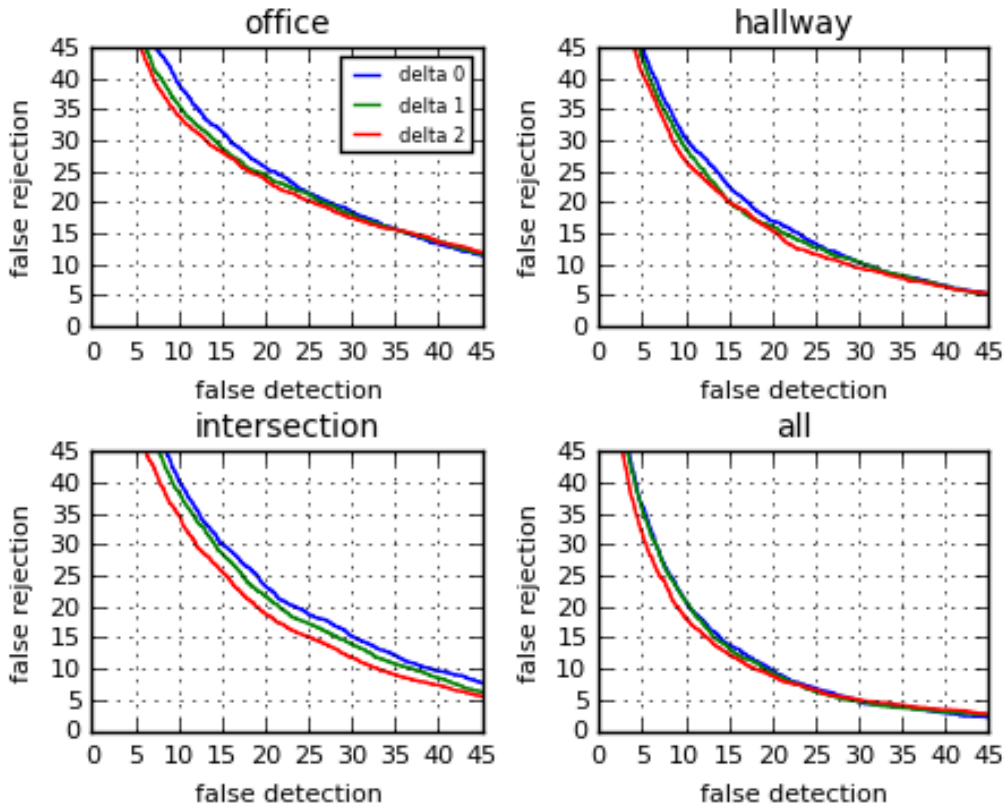


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias

$M = 16$

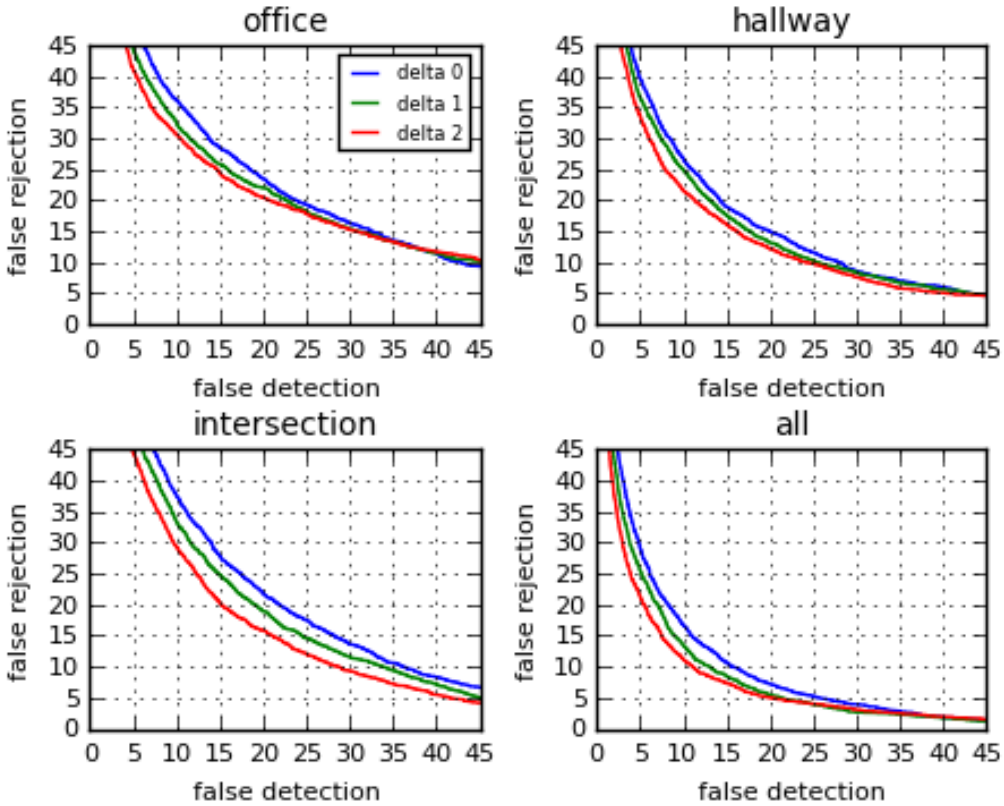


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias

$M = 32$

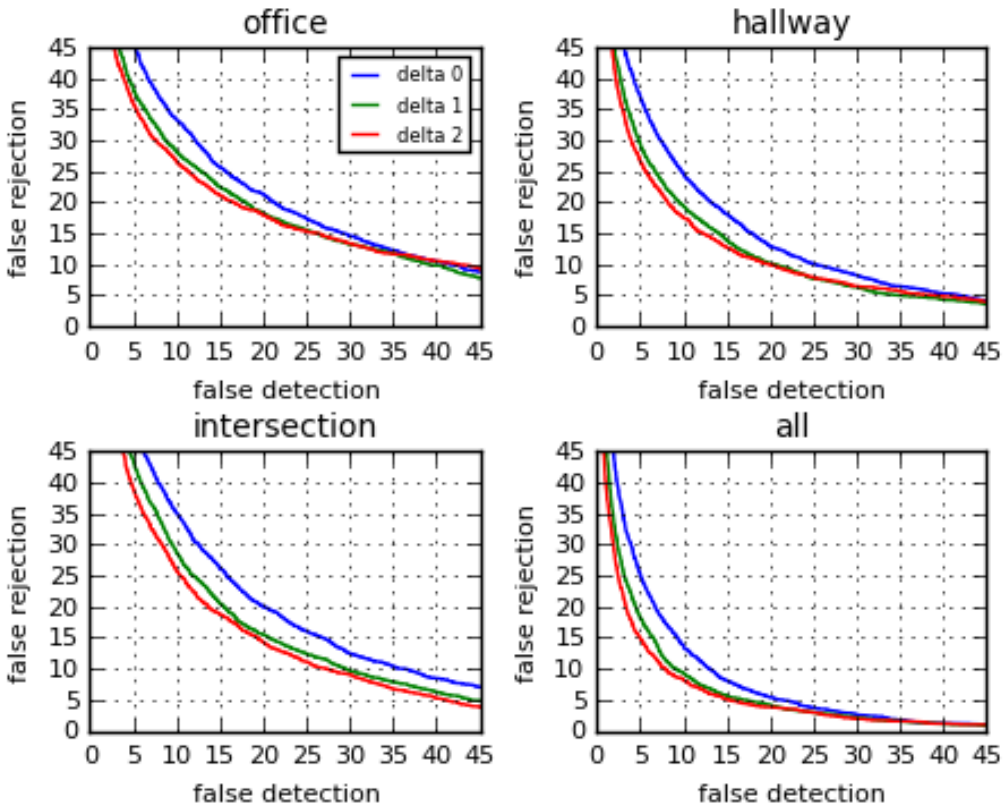


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias

$M = 64$

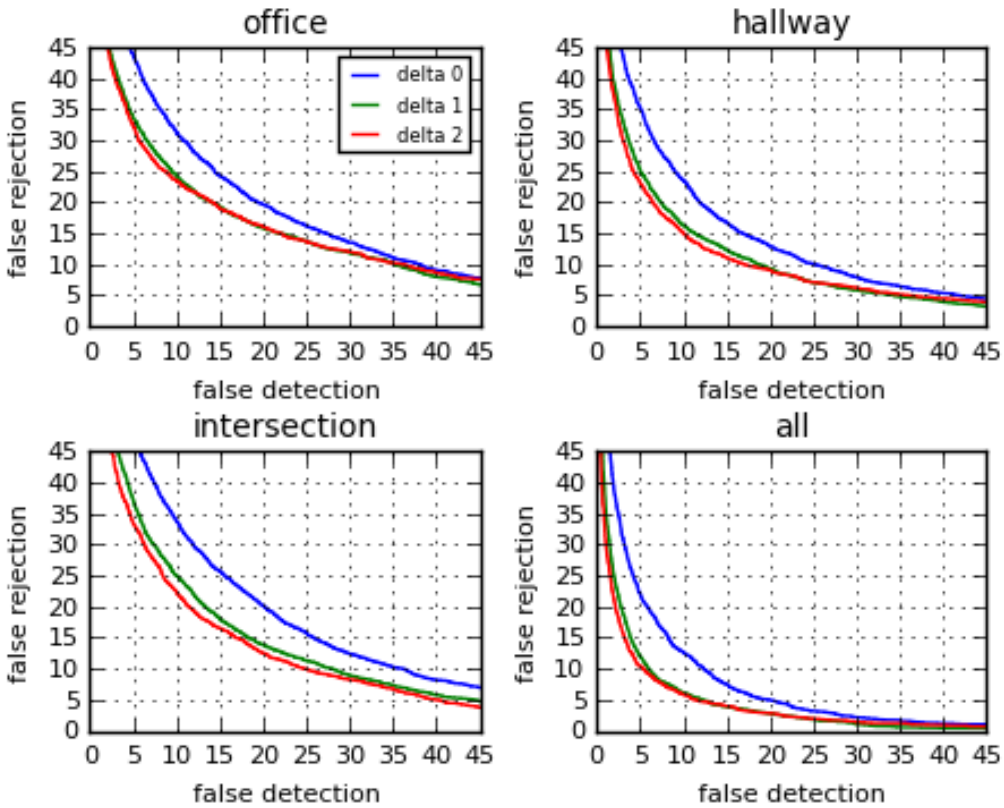


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias

$M = 128$



Verificação

SSAGMM Single Speaker AGMM

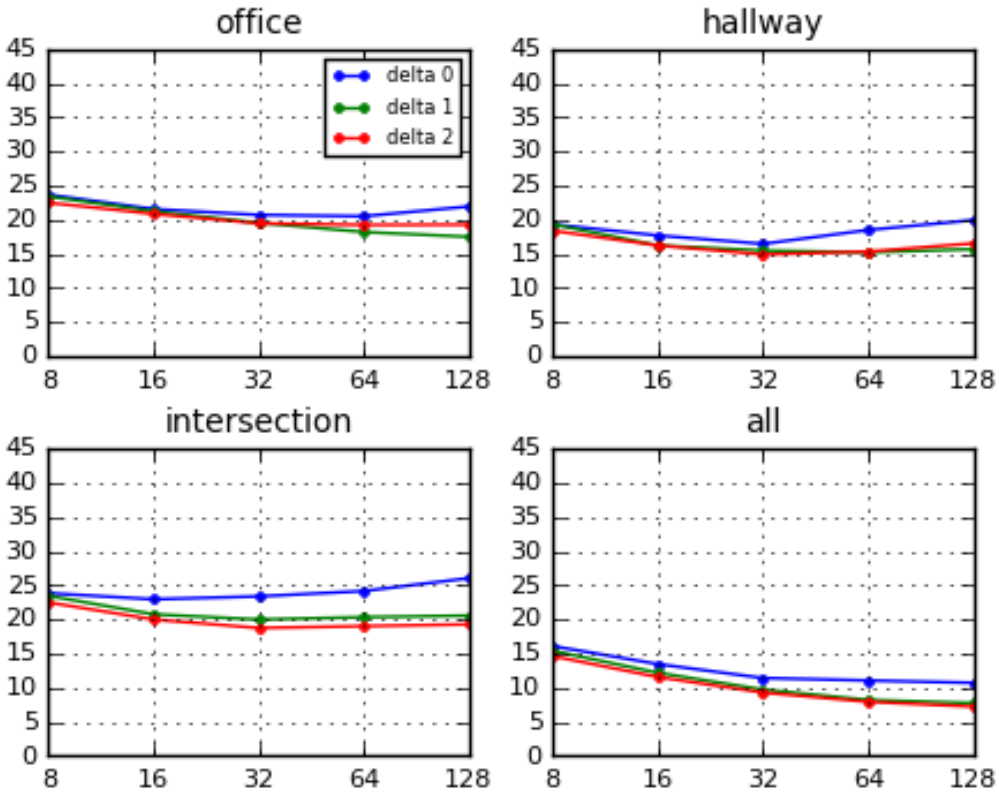
Adaptação Médias e variâncias

Δ	M	Office	Hallway	Intersection	All
0	8	23.68	19.32	23.84	16.09
	16	21.57	17.71	22.96	13.46
	32	20.72	16.48	23.38	11.42
	64	20.52	18.51	24.16	11.07
	128	21.95	19.96	26.04	10.72
1	8	23.42	19.33	23.45	15.39
	16	21.26	16.24	20.76	12.19
	32	19.56	15.50	19.98	9.72
	64	18.22	15.24	20.33	8.22
	128	17.52	15.69	20.56	7.75
2	8	22.49	18.40	22.49	14.62
	16	20.87	16.24	19.99	11.56
	32	19.48	14.93	18.75	9.30
	64	19.25	15.36	19.02	7.94
	128	19.29	16.55	19.28	7.25

Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias e variâncias

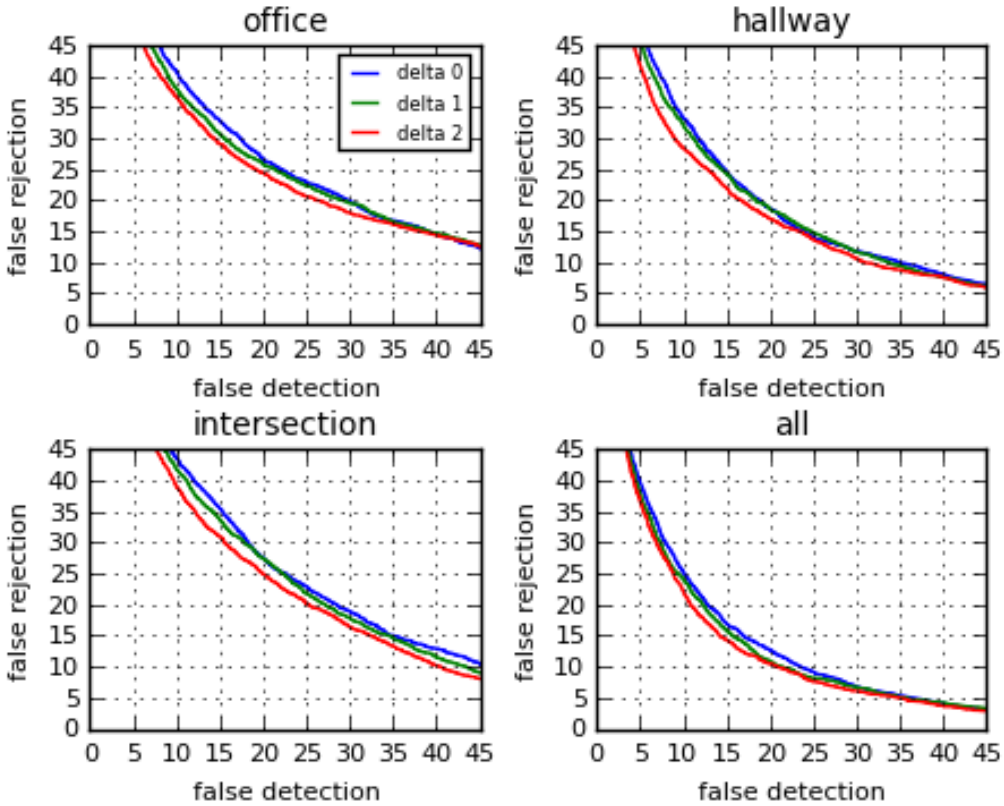


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias e variâncias

$M = 8$

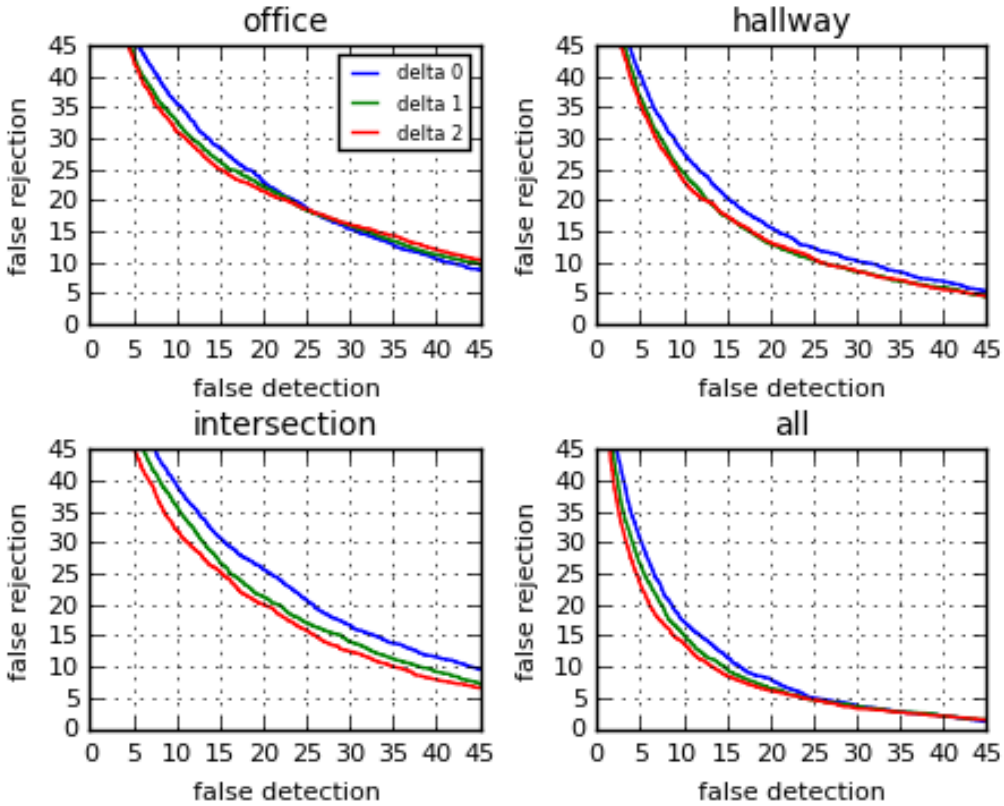


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias e variâncias

$M = 16$

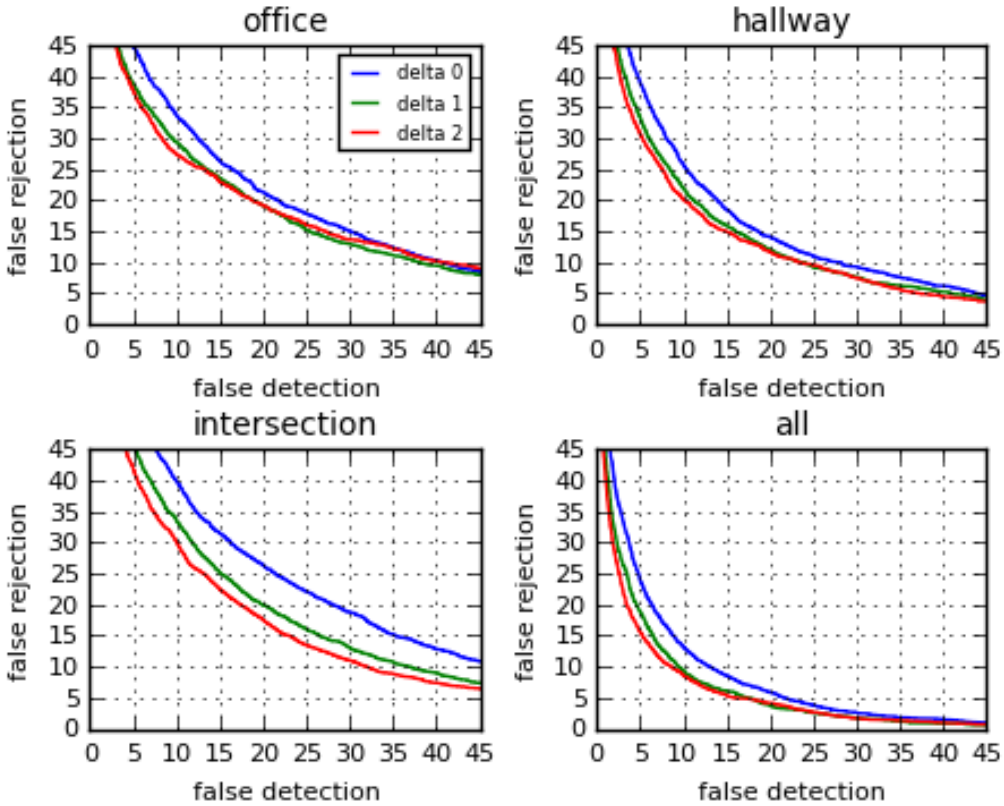


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias e variâncias

$M = 32$

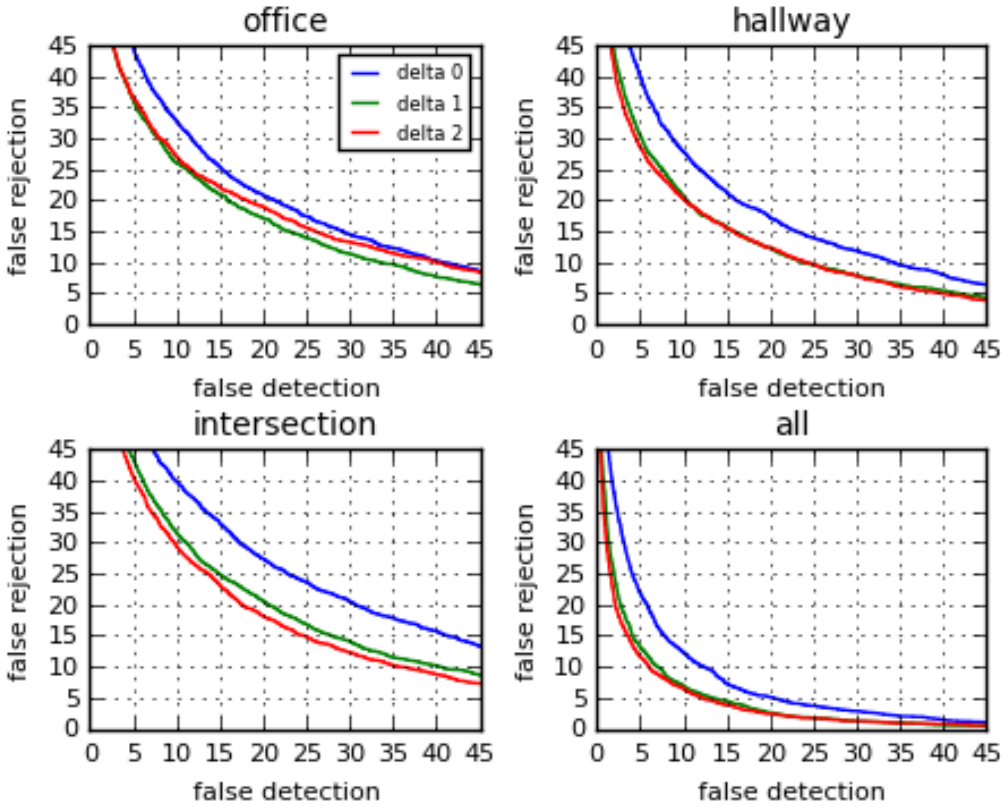


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias e variâncias

$M = 64$

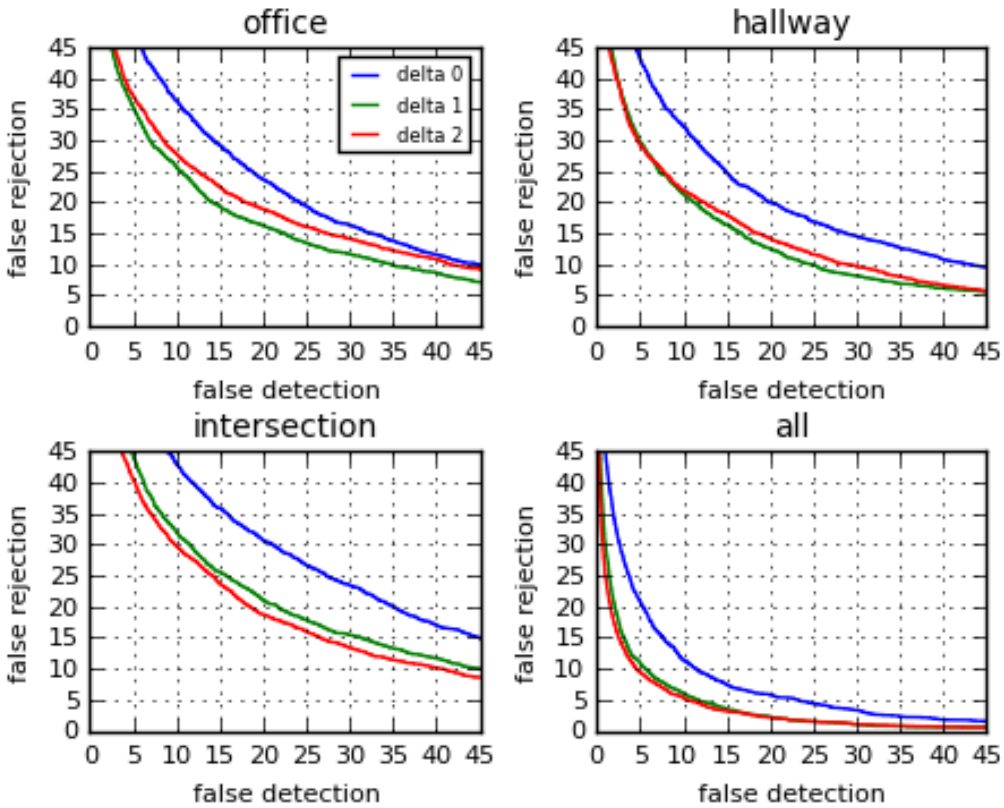


Verificação

SSAGMM Single Speaker AGMM

Adaptação Médias e variâncias

$M = 128$



Verificação

SSAGMM Single Speaker AGMM

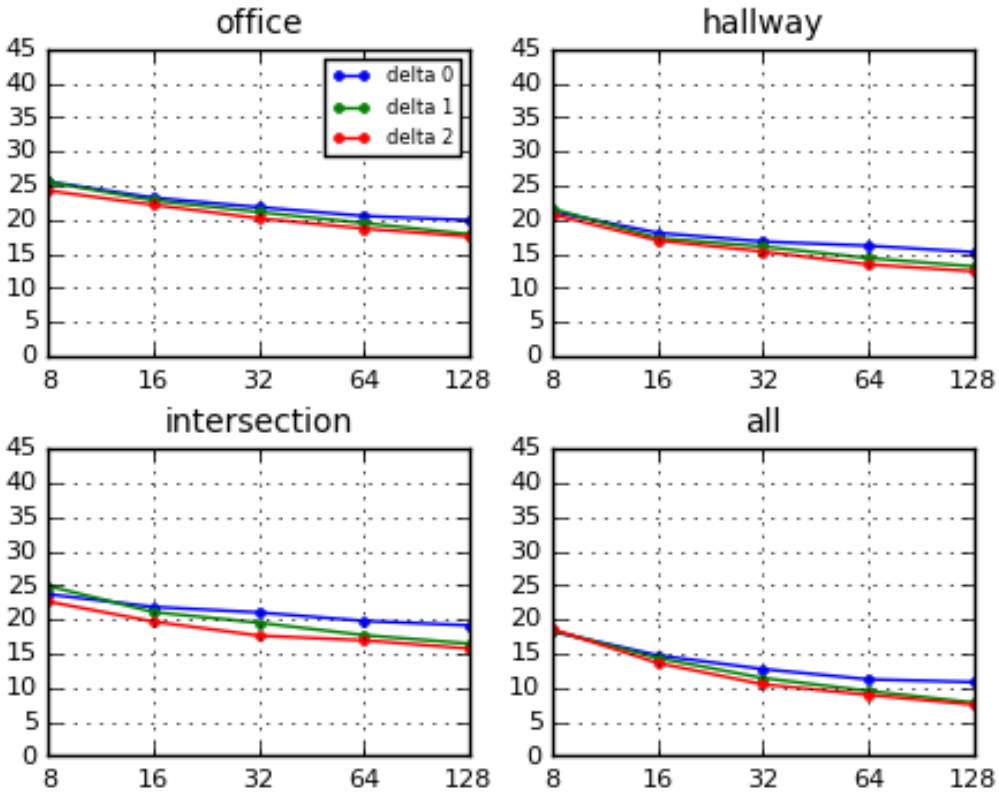
Adaptação Pesos e médias

Δ	M	Office	Hallway	Intersection	All
0	8	25.58	21.17	23.68	18.21
	16	23.23	18.09	21.83	14.74
	32	21.84	16.82	21.03	12.73
	64	20.56	16.20	19.78	11.23
	128	19.95	15.28	19.14	10.84
1	8	25.54	21.60	24.88	18.36
	16	22.84	17.32	21.07	14.35
	32	21.14	16.09	19.52	11.42
	64	19.52	14.40	17.71	9.53
	128	17.90	13.19	16.47	7.84
2	8	24.27	20.76	22.61	18.56
	16	22.18	16.98	19.68	13.62
	32	20.22	15.36	17.64	10.49
	64	18.72	13.47	16.95	8.96
	128	17.55	12.46	15.74	7.52

Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos e médias

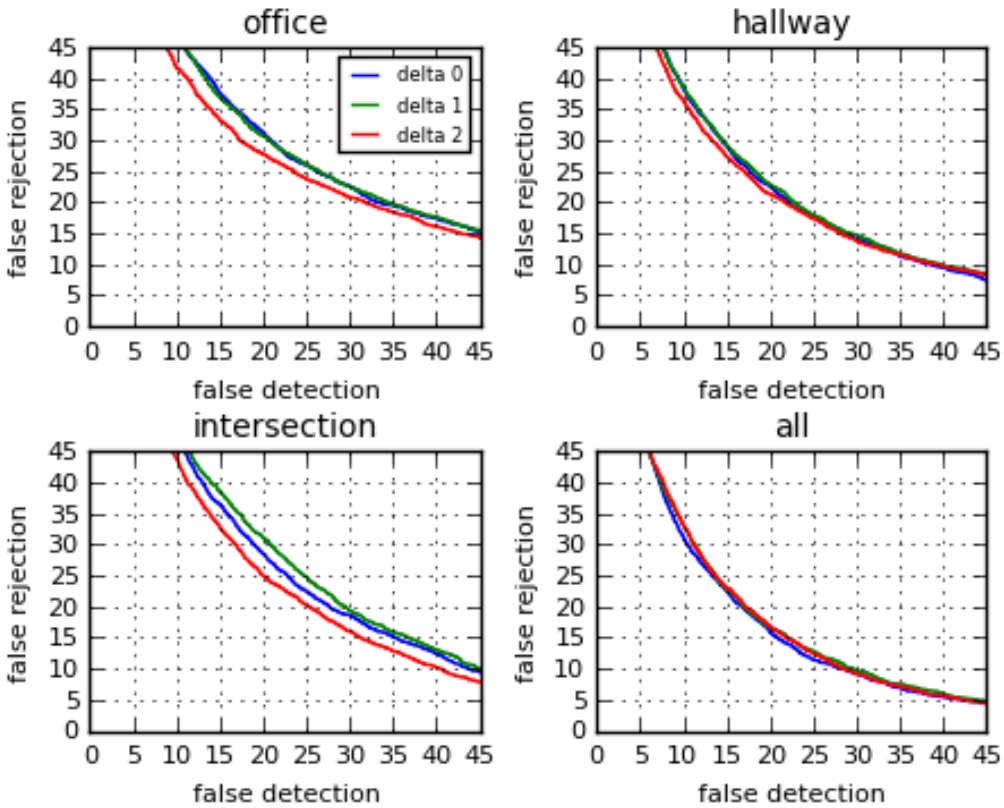


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos e médias

$M = 8$

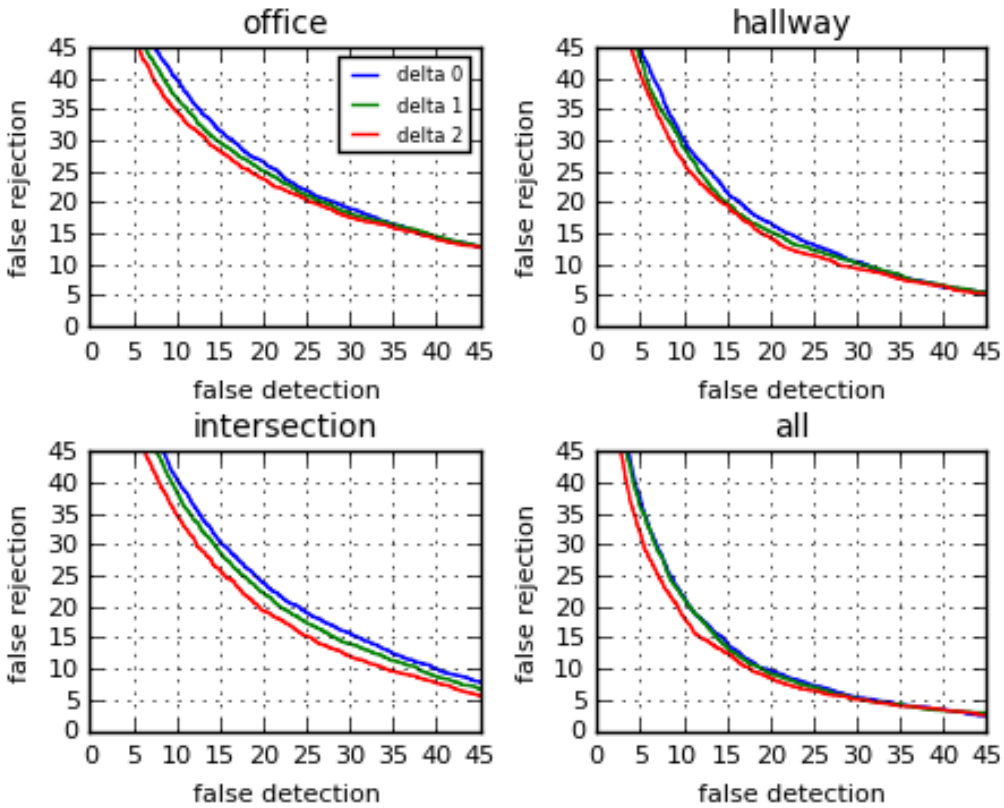


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos e médias

$M = 16$

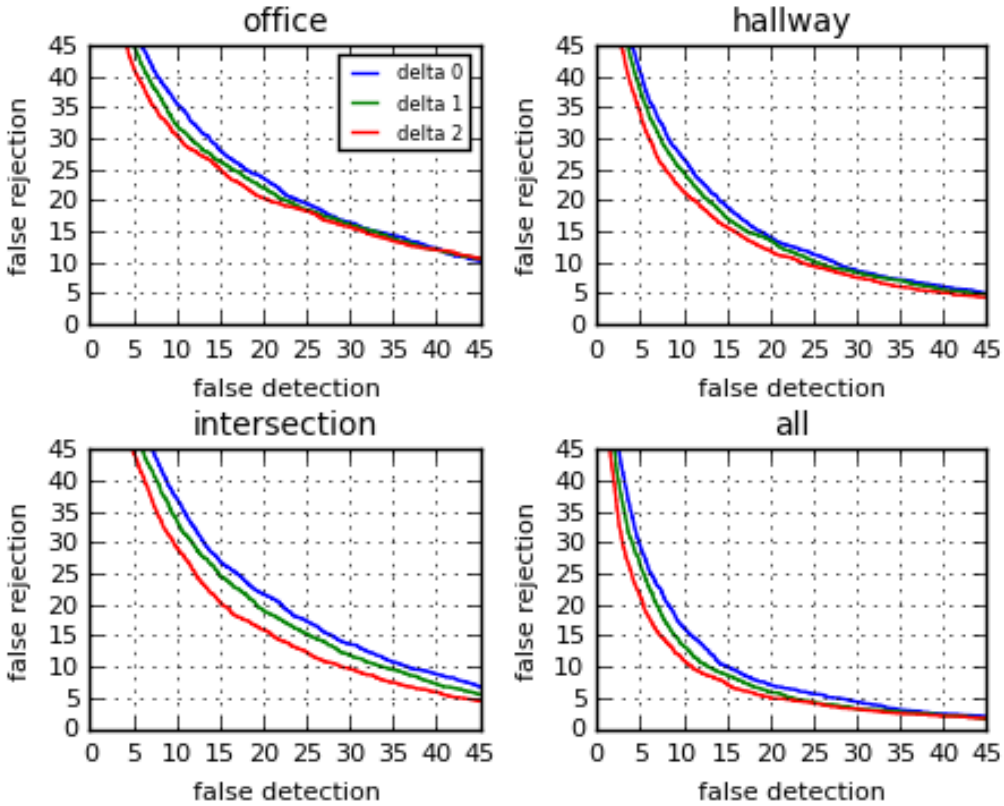


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos e médias

$M = 32$

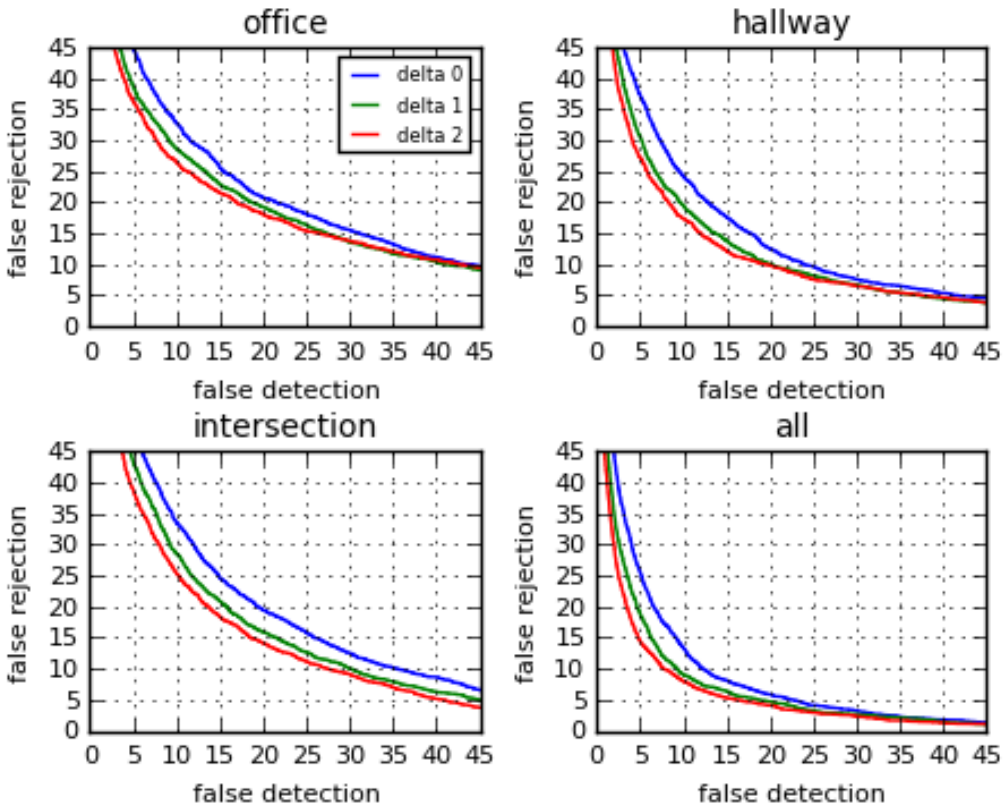


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos e médias

$M = 64$

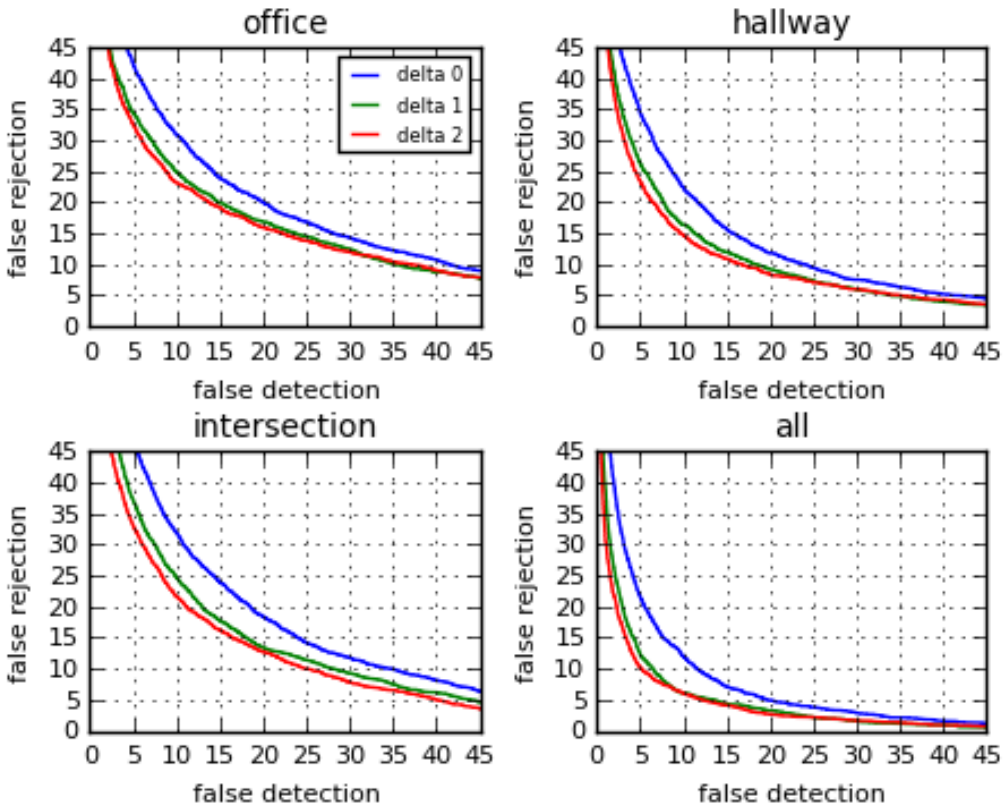


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos e médias

$M = 128$



Verificação

SSAGMM Single Speaker AGMM

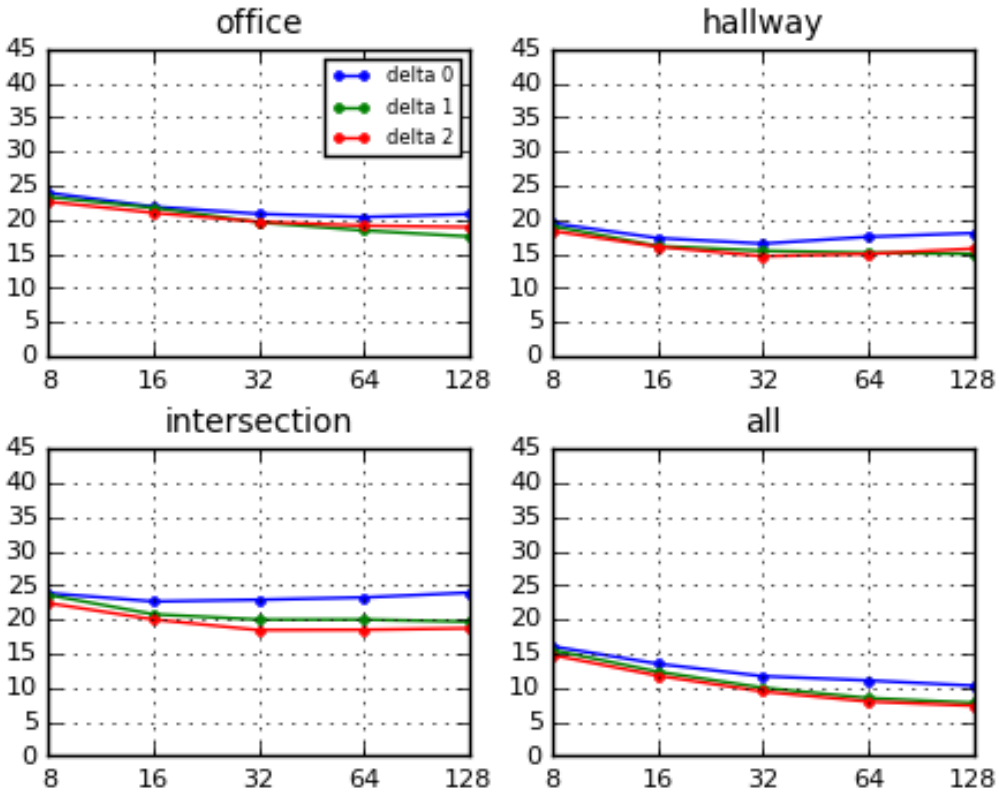
Adaptação Pesos, médias e variâncias

Δ	M	Office	Hallway	Intersection	All
0	8	23.96	19.49	23.84	16.04
	16	21.92	17.33	22.64	13.50
	32	20.87	16.51	22.88	11.69
	64	20.41	17.51	23.23	11.07
	128	20.84	18.06	23.92	10.30
1	8	23.35	19.06	23.58	15.51
	16	21.76	16.16	20.76	12.35
	32	19.64	15.47	19.98	9.99
	64	18.44	15.16	19.98	8.49
	128	17.55	15.01	19.68	7.80
2	8	22.65	18.36	22.38	14.78
	16	21.03	16.06	19.99	11.77
	32	19.71	14.67	18.44	9.44
	64	19.14	15.01	18.49	7.99
	128	18.94	15.78	18.71	7.37

Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos, médias e variâncias

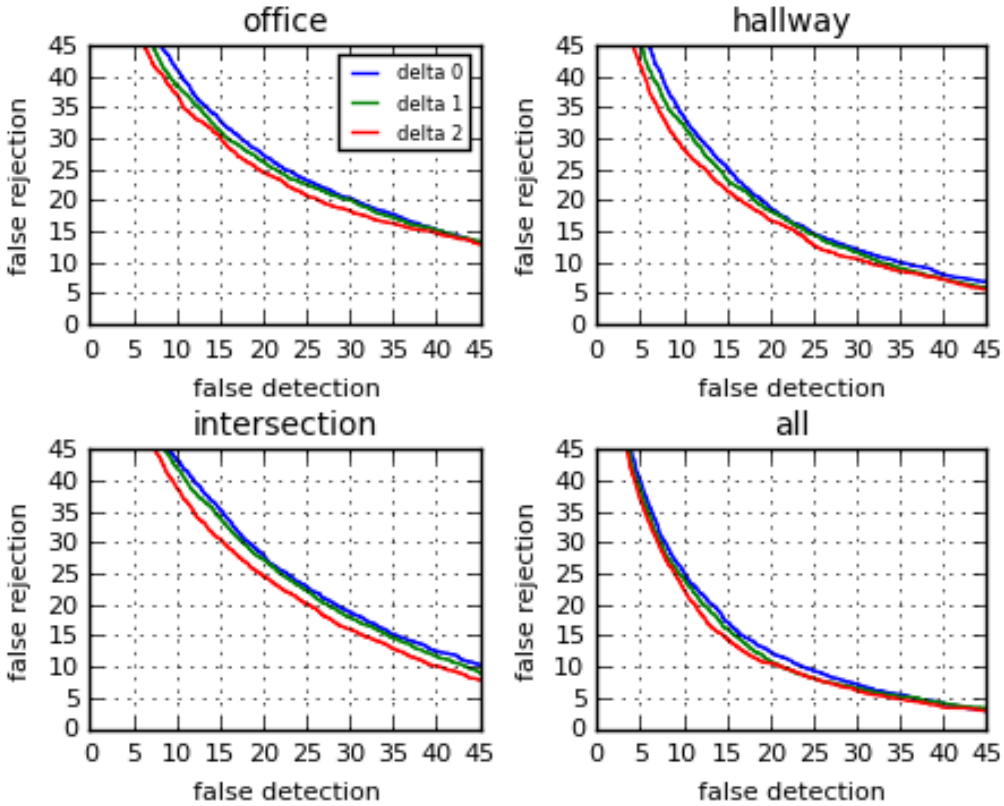


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos, médias e variâncias

$M = 8$

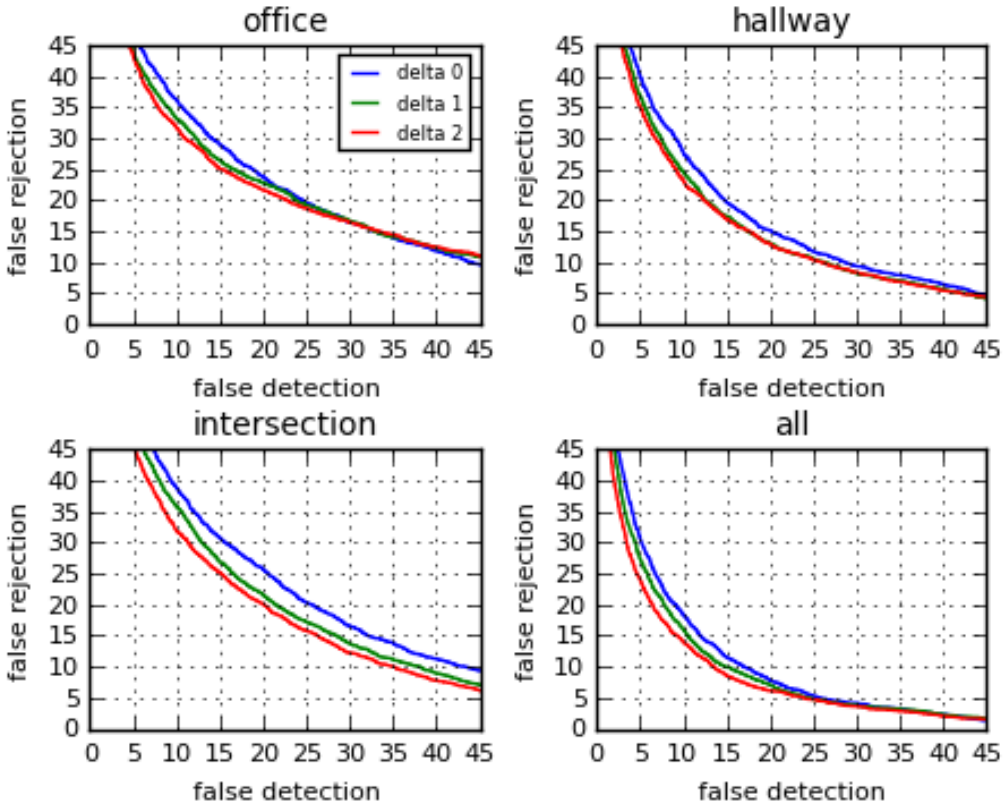


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos, médias e variâncias

$M = 16$

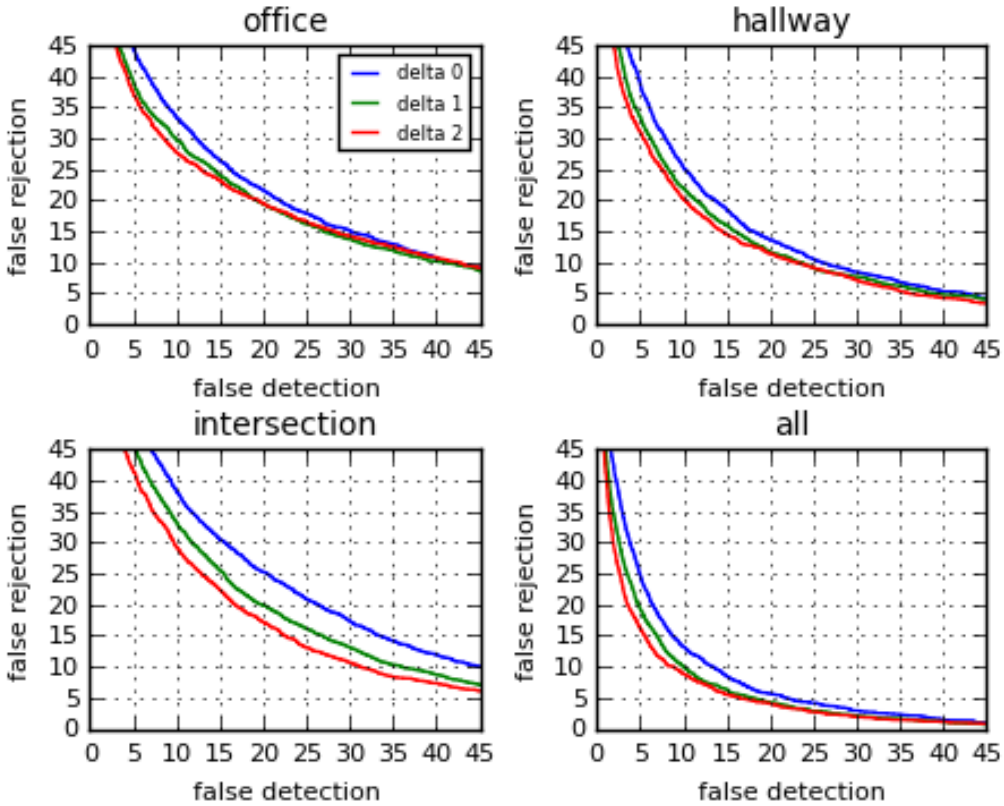


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos, médias e variâncias

$M = 32$

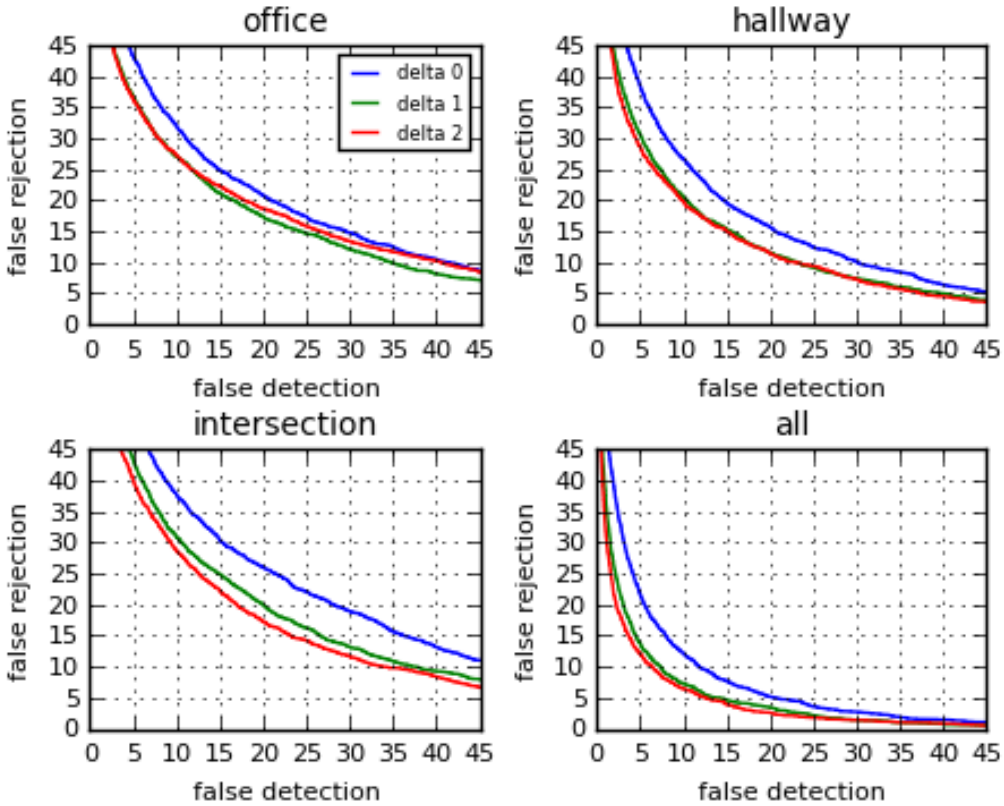


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos, médias e variâncias

$M = 64$

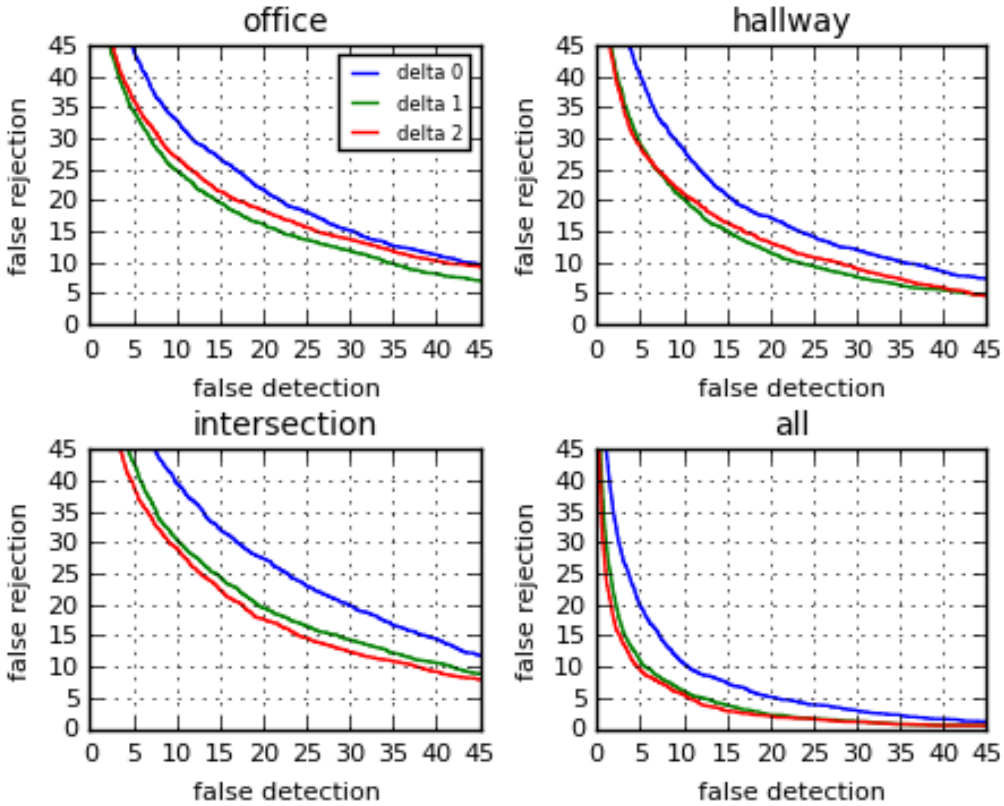


Verificação

SSAGMM Single Speaker AGMM

Adaptação Pesos, médias e variâncias

$M = 128$



Conteúdo

1 Introdução

2 Sistemas de Reconhecimento de Locutor

3 Extração de Características

4 Modelos de Mistura Gaussianas

5 Experimentos

6 Conclusão

Conclusão

GMM é uma ótima modelagem para reconhecimento de locutor independente de texto

Conclusão

GMM é uma ótima modelagem para reconhecimento de locutor independente de texto

Identificação com FGMM apresentou resultados inferiores ao esperado

Conclusão

GMM é uma ótima modelagem para reconhecimento de locutor independente de texto

Identificação com FGMM apresentou resultados inferiores ao esperado

- Investigar melhor a teoria
- Problema de calibragem do r?

Conclusão

GMM é uma ótima modelagem para reconhecimento de locutor independente de texto

Identificação com FGMM apresentou resultados inferiores ao esperado

- Investigar melhor a teoria
- Problema de calibragem do r?

Verificação com GMM apresenta bons resultados

Conclusão

GMM é uma ótima modelagem para reconhecimento de locutor independente de texto

Identificação com FGMM apresentou resultados inferiores ao esperado

- Investigar melhor a teoria
- Problema de calibragem do r?

Verificação com GMM apresenta bons resultados

- Testar com valores maiores de M
- Utilizar outras bases

Conclusão

GMM é uma ótima modelagem para reconhecimento de locutor independente de texto

Identificação com FGMM apresentou resultados inferiores ao esperado

- Investigar melhor a teoria
- Problema de calibragem do r?

Verificação com GMM apresenta bons resultados

- Testar com valores maiores de M
- Utilizar outras bases

Verificação com AGMM é uma boa alternativa

Conclusão

GMM é uma ótima modelagem para reconhecimento de locutor independente de texto

Identificação com FGMM apresentou resultados inferiores ao esperado

- Investigar melhor a teoria
- Problema de calibragem do r?

Verificação com GMM apresenta bons resultados

- Testar com valores maiores de M
- Utilizar outras bases

Verificação com AGMM é uma boa alternativa

- Boas modelagens sempre adaptam as médias
- Testar com diferentes valores de r

Dúvidas?