

# Segmentación de interlocutores a partir de señales de audio utilizando cadenas escondidas de Markov y técnicas de selección automática de modelos

Rafael de Jesús Robledo Juárez

`rrobledo@cimat.mx`

Asesor: Dr. Salvador Ruíz Correa



**CIMAT**

Centro de Investigación en Matemáticas, Guanajuato  
Departamento de Ciencias de la Computación

xx de noviembre del 2013

# Contenido

**1** Introducción

**2** Speaker Diarization

**3** Modelo

**4** Pruebas

**5** Trabajo futuro

# Contenido

- 1 **Introducción**
  - Problema
  - Motivación
  - Principales enfoques
  - Trabajo previo
- 2 Speaker Diarization
- 3 Modelo
- 4 Pruebas
- 5 Trabajo futuro

# Problema

- ▶ Se considera que se tiene una señal de audio con información de nuestro interés, y se requiere segmentar de acuerdo a las personas que participan en la grabación.
- ▶ *Speaker Diarization*: el problema consiste en identificar el número de interlocutores que participan en una grabación de audio, y además encontrar en qué segmentos de la grabación habla cada persona.
- ▶ Dos tareas principales:
  1. Encontrar el número total de personas que hablan en la conversación.
  2. Identificar los momentos en los que habla cada participante.

# Motivación

- ▶ La tarea de speaker diarization es importante en diferentes procesos que se realizan con las grabaciones de audio, tales como la identificación y navegación por segmentos en específico,
- ▶ También resulta útil para la búsqueda y recuperación de información en grandes volúmenes de secuencias de audio.
- ▶ Es una etapa importante en el procesamiento de voz. Tanto para reconocimiento como transcripción de voz.

# Principales enfoques

You can create overlays. . .

- ▶ using the `pause` command:
- ▶ First item.
- ▶ Second item.

# Trabajo previo

You can create overlays. . .

- ▶ using the `pause` command:
  - ▶ First item.
  - ▶ Second item.
- ▶ using overlay specifications:
  - ▶ First item.
  - ▶ Second item.
- ▶ using the general `uncover` command:
  - ▶ First item.
  - ▶ Second item.

# Contenido

## 1 Introducción

## 2 Speaker Diarization

- Componentes del sistema
- Procesamiento acústico

## 3 Modelo

## 4 Pruebas

## 5 Trabajo futuro

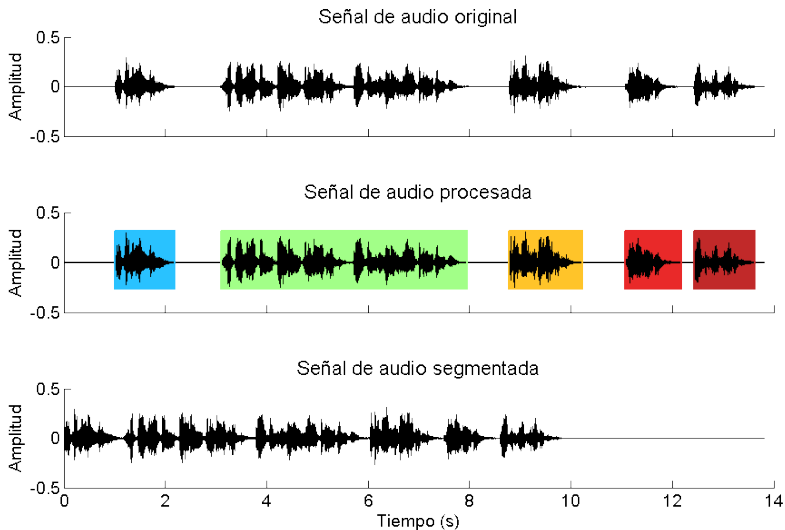


# Speaker Diarization

# Formulación matemática

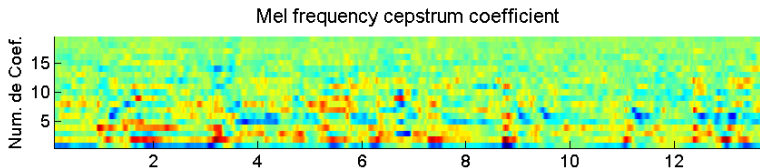
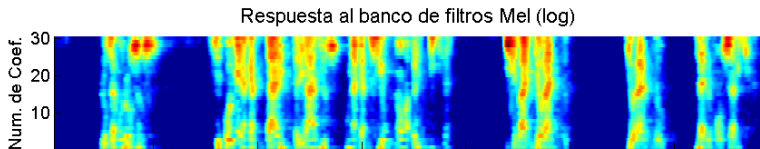
# Componentes del sistema

# Eliminación de ruido / Detección de silencios



# Mel Frequency Cepstrum Coefficient

- FFT (ventana) -> Banco de filtros triangular (Mel Scale) -> Log -> DCT -> MFCC



# Contenido

1 Introducción

2 Speaker Diarization

3 Modelo

- Hidden Markov Model
- Resolver HMM con EM
- Bootstrap

4 Pruebas

5 Trabajo futuro

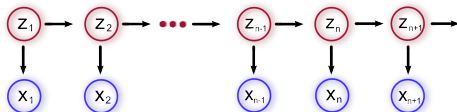
# Hidden Markov Model

- Cadena de Markov de primer orden



$$p(x_1, \dots, x_N) = \prod_{n=1}^N p(x_n \mid x_1, \dots, x_{n-1}) = p(x_1) \prod_{n=2}^N p(x_n \mid x_{n-1}) \quad (1)$$

- Modelo estocástico de Markov en el que el estado de la cadena es parcialmente observado.



- Modelar proceso bivariado en el tiempo. Una variable observada y una variable latente asociada.

# Hidden Markov Model

- Agregar una variable latente  $z_n$  (discreta), que corresponda a cada observación  $x_n$ .

$$z_{n+1} \perp z_{n-1} \mid z_n \quad (2)$$

$$p(x_1, \dots, x_N, z_1, \dots, z_N) = p(z_1) \left[ \prod_{n=2}^N p(z_n \mid z_{n-1}) \right] \prod_{n=1}^N p(x_n \mid z_n). \quad (3)$$

- Mezcla de distribuciones en la que la densidad tiene una distribución dada por  $p(x|z)$



# Parámetros del HMM

- Probabilidad de cambio entre estados dada una **matriz de transición** **A**

$$A_{jk} \equiv p(z_{nk} = 1 \mid z_{n-1,j} = 1) \quad (4)$$

$$p(z_n \mid z_{n-1}, \mathbf{A}) = \prod_{k=1}^K \prod_{j=1}^K A_{jk}^{z_{n-1,j} \cdot z_{n,k}} \quad (5)$$

- **Vector de distribución inicial**  $\pi$  para variable latente.

$$\pi_k \equiv p(z_{1k}) \quad (6)$$

$$p(z_1 \mid \pi) = \prod_{k=1}^K \pi_k^{z_{1k}} \quad (7)$$

- **Probabilidad de emisión** de una variable observada  $x_n$  dada una variable latente  $z_n$ .

$$p(x_n \mid z_n, \phi) = \prod_{k=1}^K p(x_n \mid \phi_k)^{z_{nk}} \quad (8)$$

# HMM con EM

- Probabilidad conjunta del modelo

$$p(\mathbf{X}, \mathbf{Z} \mid \theta) = p(z_1 \mid \pi) \left[ \prod_{n=2}^N p(z_n \mid z_{n-1}, \mathbf{A}) \right] \prod_{n=1}^N p(x_n \mid z_n, \phi) \quad (9)$$

- Parámetros del modelo  $\theta = \{\pi, \mathbf{A}, \phi\}$
- Función de verosimilitud completa

$$\mathcal{Q}(\theta, \theta^{old}) = \sum_{\mathbf{Z}} p(\mathbf{Z} \mid \mathbf{X}, \theta^{old}) \log p(\mathbf{X}, \mathbf{Z} \mid \theta) \quad (10)$$

- Prob. marginal de una variable latente, prob. conjunta de dos variables latentes consecutivas

$$\gamma(z_n) = p(z_n \mid \mathbf{X}, \theta^{old}) \quad (11)$$

$$\xi(z_{n-1}, z_n) = p(z_{n-1}, z_n \mid \mathbf{X}, \theta^{old}) \quad (12)$$

# HMM con EM

- Prob. marginal de  $z_{nk} = 1$ , prob. conjunta de  $z_{n-1,j}, z_{nk}$

$$\gamma(z_{nk}) = \mathbb{E}[z_{nk}] = \sum_{\mathbf{z}} \gamma(\mathbf{z}) z_{nk} \quad (13)$$

$$\xi(z_{n-1,j}, z_{nk}) = \mathbb{E}[z_{n-1,j} \cdot z_{nk}] = \sum_{\mathbf{z}} \gamma(\mathbf{z}) z_{n-1,j} \cdot z_{nk} \quad (14)$$

- Función de verosimilitud completa (reescrita con (13), (14))

$$\begin{aligned} Q(\theta, \theta^{old}) = & \sum_{k=1}^K \gamma(z_{1k}) \log \pi_k + \sum_{n=2}^N \sum_{j=1}^K \sum_{k=1}^K \xi(z_{n-1,j}, z_{nk}) \log A_{jk} + \\ & \sum_{n=1}^N \sum_{k=1}^K \gamma(z_{nk}) \log p(x_n | \phi_k) \end{aligned} \quad (15)$$

- Parámetros estimados por EM:

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{j=1}^K \gamma(z_{1j})}, \quad A_{jk} = \sum_{n=2}^N \frac{\xi(z_{n-1,j}, z_{nk})}{\sum_{l=1}^K \xi(z_{n-1,j}, z_{nl})} \quad (16)$$

# Algoritmo backward-forward

$$\gamma(z_n) = p(z_n | X) = \frac{p(X | z_n)p(z_n)}{p(X)} \quad (17)$$

$$\gamma(z_n) = \frac{p(x_1, \dots, x_n, z_n)p(x_{n+1}, \dots, x_N | z_n)}{p(X)} \quad (18)$$

$$\gamma(z_n) = \frac{\alpha(z_n)\beta(z_n)}{p(X)} \quad (19)$$

$$(20)$$

donde

$$\alpha(z_n) \equiv p(x_1, \dots, x_n, z_n) \quad (21)$$

$$\beta(z_n) \equiv p(x_{n+1}, \dots, x_N | z_n) \quad (22)$$

$$\alpha(z_n) = p(x_n | z_n) \sum_{z_{n-1}} \alpha(z_n | z_{n-1}) \quad (23)$$

$$\alpha(z_1) = p(z_1)p(x_1 | z_1) = \prod_{k=1}^K \{\pi_k p(x_1 | \phi_k)\}^{z_{1k}} \quad (24)$$

# Algoritmo backward-forward

$$\beta(z_n) = \sum_{z_{n+1}} \beta(z_{n+1}) p(x_{n+1} \mid z_{n+1}) p(z_{n+1} \mid z_n) \quad (25)$$

# Bootstrap

# Contenido

- 1 Introducción
- 2 Speaker Diarization
- 3 Modelo
- 4 Pruebas**
  - Pruebas con datos sintéticos
- 5 Trabajo futuro

# Numero fijo de speakers



# Numero variable de speakers

# Resumen

- ▶ The **first main message** of your talk in one or two lines.
  - ▶ The **second main message** of your talk in one or two lines.
  - ▶ Perhaps a **third message**, but not more than that.
- 
- ▶ Outlook
    - ▶ Something you haven't solved.
    - ▶ Something else you haven't solved.

# Contenido

1 Introducción

2 Speaker Diarization



3 Modelo

4 Pruebas

**5 Trabajo futuro**

# Trabajo futuro

# Referencias I

-  A. Author.  
*Handbook of Everything.*  
Some Press, 1990.
-  S. Someone.  
On this and that.  
*Journal of This and That*, 2(1):50–100, 2000.