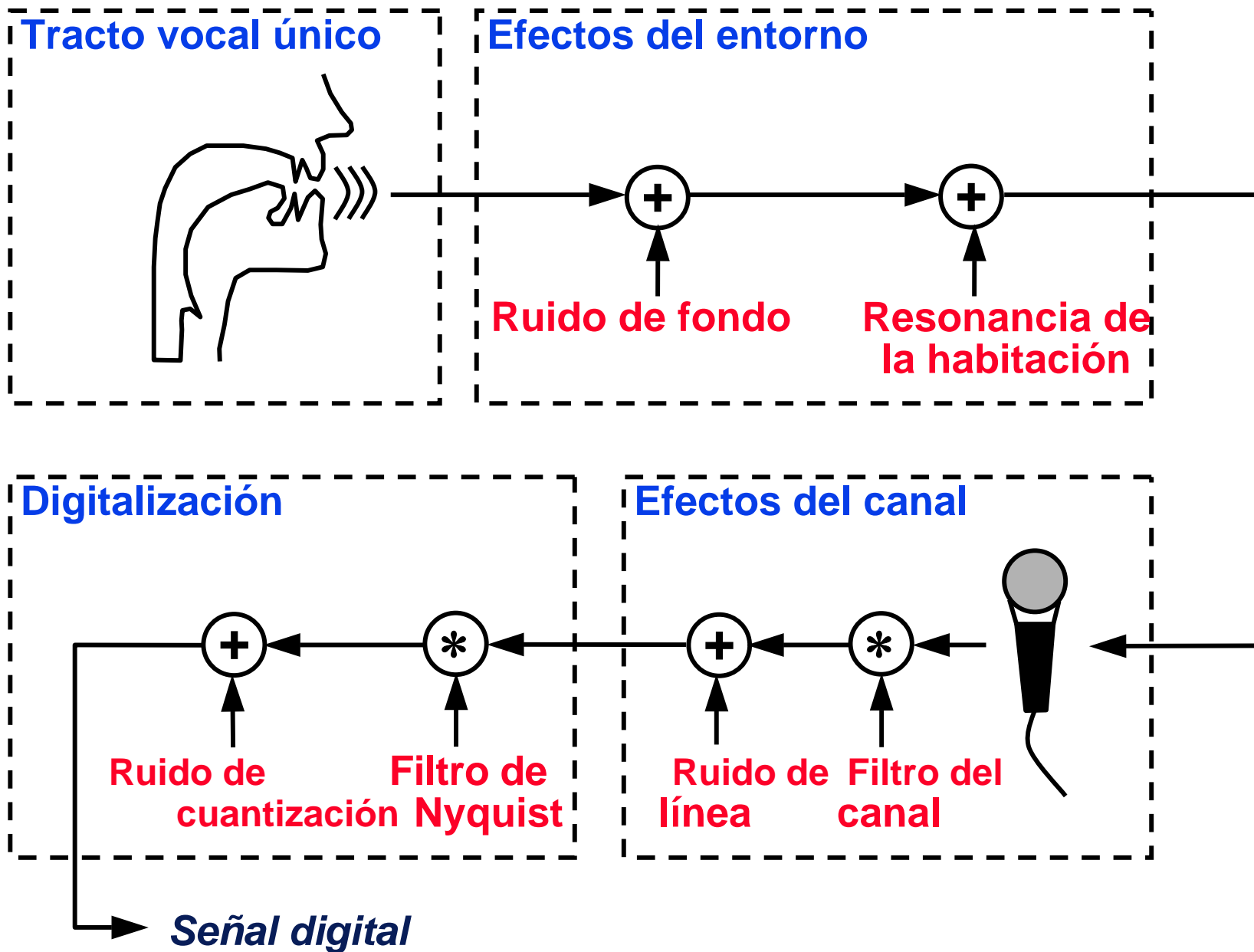


Adaptación del hablante

Profesor: T. J. Hazen

- **Visión general**
- **Métodos de adaptación**
 - Normalización de la longitud del tracto vocal
 - Adaptación bayesiana
 - Adaptación transformacional
 - Coeficiente de ponderación del hablante de referencia
 - *Eigenvoices* para la adaptación del hablante
 - Adaptación estructural
 - Agrupamiento jerárquico de hablantes
 - Coeficiente de ponderación de grupos de hablantes
- **Resumen**

Grabación típica digital del discurso



MIT Estimaciones de variabilidad

- **Los reconocedores deben estimar la variabilidad de los hablantes**
- **Enfoque estándar: Entrenamiento independiente del hablante (SI)**
 - Los datos de entrenamiento agruparon a muchos hablantes distintos
- **Problemas con enfoques de modelado primario:**
 - Los modelos son heterogéneos de varianza alta
 - Se necesitan muchos parámetros para construir modelos exactos
 - Los modelos no proporcionan ninguna restricción de hablante
 - Puede que los nuevos datos aún no se parezcan a los datos de entrenamiento

MIT Facilitando la restricción

- **Sería aconsejable que los reconocedores también proporcionaran la restricción**
 - Las fuentes de variación permanecen normalmente fijas durante el enunciado
 - Mismo hablante, micrófono, canal, entorno
- **Posibles soluciones:**
 - Normalizar los datos de entrada para igualar modelos (ej., normalización)
 - Adaptar modelos para igualar datos de entrada (ej., adaptación)
- **Ideas clave:**
 - Las fuentes de variabilidad son normalmente sistemáticas y consistentes
 - Unos cuantos parámetros pueden describir una gran variación sistemática
 - Las correlaciones internas al hablante existen entre sonidos distintos

- El modelo acústico predice la probabilidad de observaciones acústicas dadas las unidades fonéticas:

$$P(A | U) = P(\vec{a}_1, \vec{a}_2, \dots, \vec{a}_N | u_1, u_2, \dots, u_n)$$

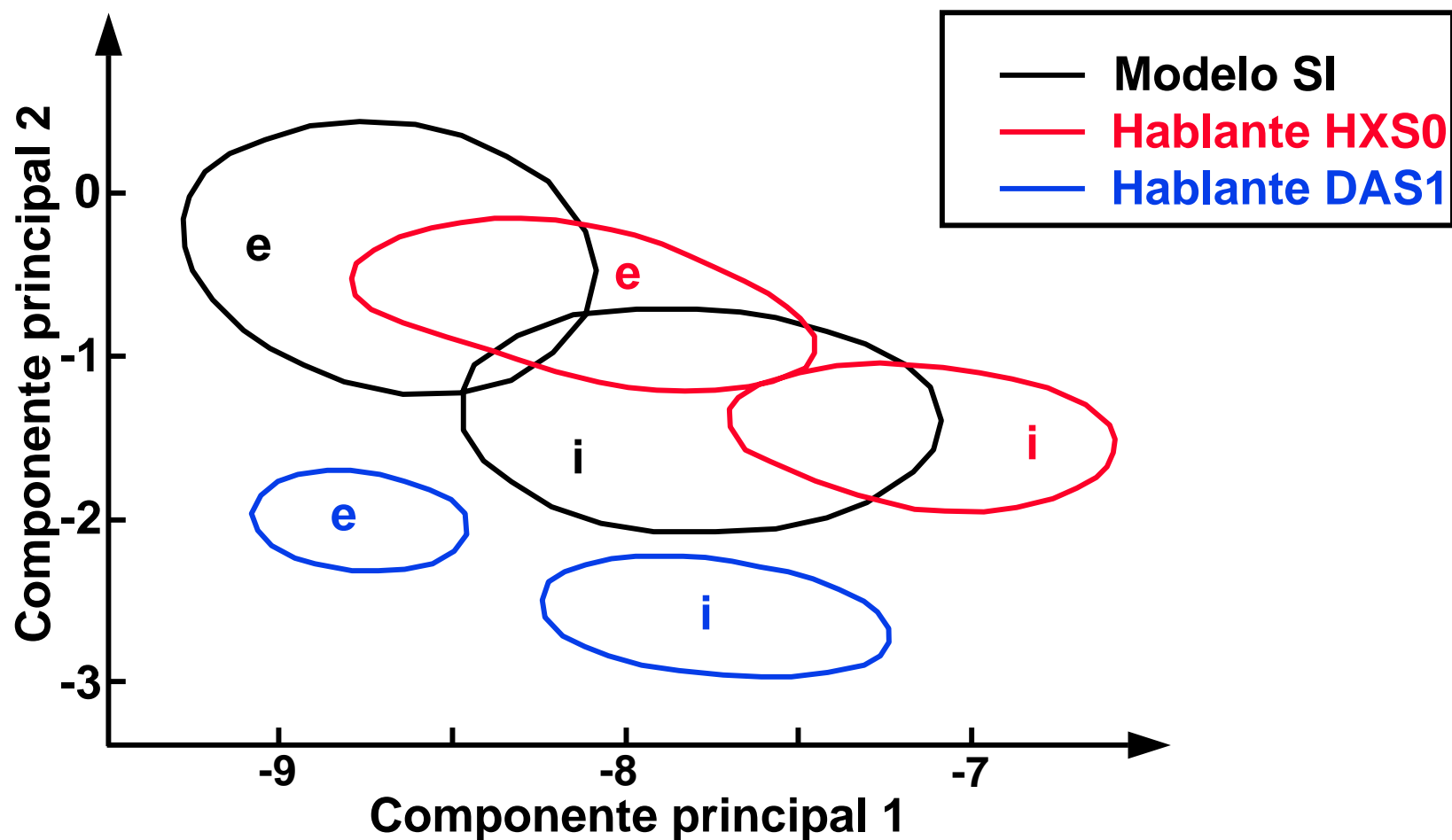
- Se requiere normalmente una suposición de independencia para hacer posible el modelado:

$$P(A | U) = \prod_{i=1}^N P(\vec{a}_i | U)$$

- Esta suposición de independencia puede ser perjudicial
 - Las correlaciones acústicas entre eventos fonéticos se ignoran
 - No se facilita ninguna restricción de las observaciones previas

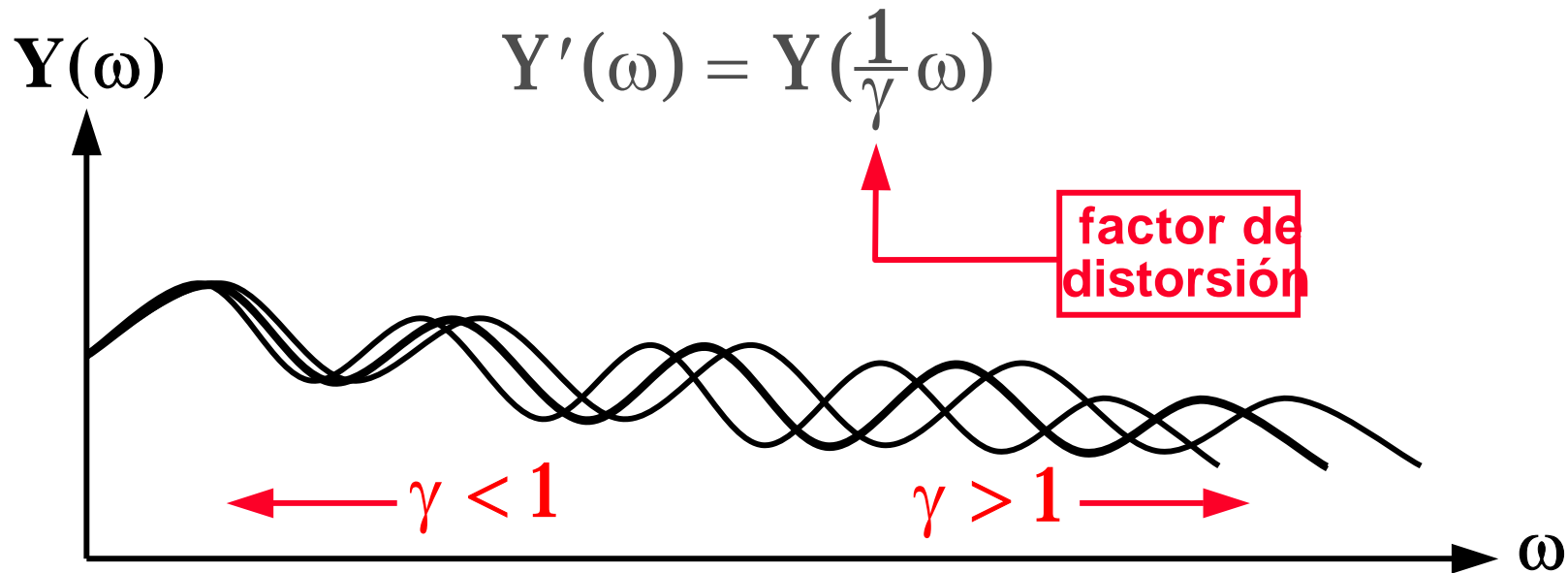
Variabilidad y correlación

- Línea de contornos de probabilidad isométricos para los fonos [i] y [e]
- Un modelo SI y dos modelos dependientes del hablante (SD)
- Los contornos SD son más estrictos que los SI y van correlacionados con cada uno



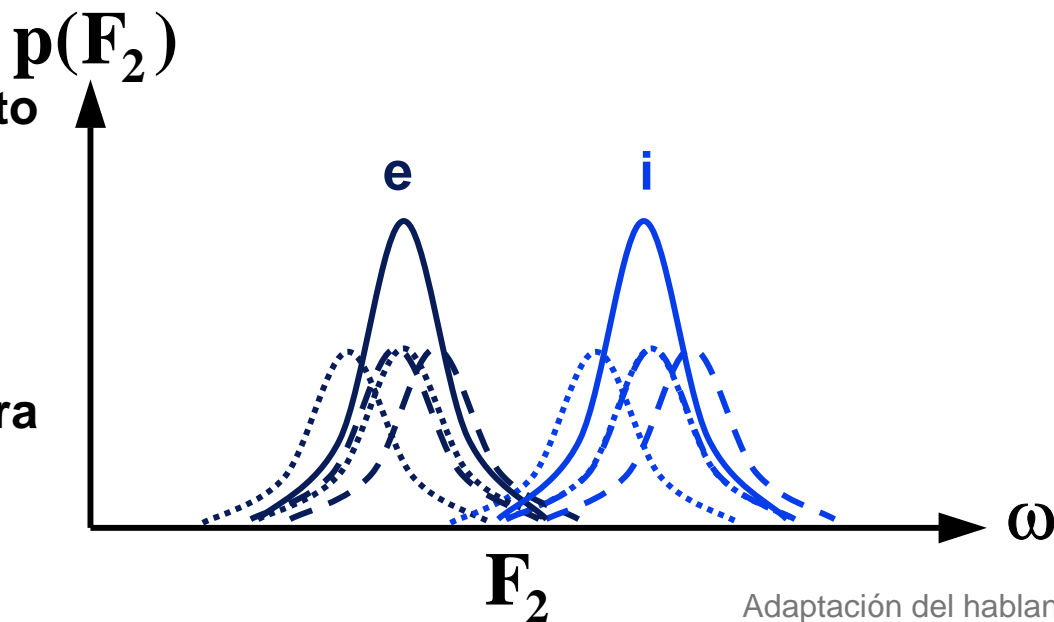
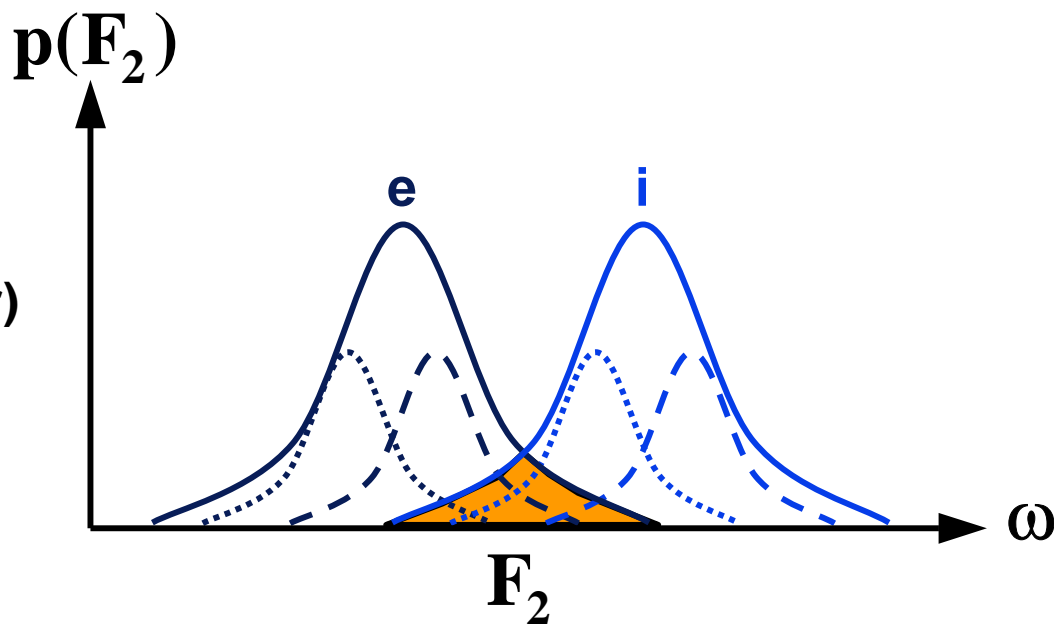
Normalización de la longitud del tracto vocal

- La longitud del tracto vocal afecta a las frecuencias del formante:
 - tractos vocales más cortos \Rightarrow frecuencias del formante más altas
 - tractos vocales más largos \Rightarrow frecuencias del formante más bajas
- La normalización de la longitud del tracto vocal (VTLN) trata de ajustar el discurso de entrada para presentar una longitud del tracto vocal “media”
- Método: Distorsionar la escala de frecuencia



Normalización de la longitud del tracto vocal (cont.)

- Ilustración: segundo formante para [e] y [i]
- Los modelos SI presentan un gran solapamiento (región de error)
- Los modelos SD poseen menores varianzas y regiones de error
- Distorsión de los espectros de todos los hablantes del entrenamiento para ajustarse mejor al modelo SI
- Entrenar el modelo SI de VTLN
- Entrenar a hablantes de prueba para ajustarse al modelo SI de VTLN



Normalización de la longitud del tracto vocal

- Durante las pruebas, se utiliza el enfoque de ML para hallar el factor de distorsión:

$$\gamma = \arg \max_{\gamma} p(\mathbf{X}^{\gamma} \mid \Theta_{\text{VTLN}})$$

- El factor de distorsión se halla mediante la búsqueda de fuerza bruta
 - Conjunto discreto de factores de distorsión probados en posibles rangos
- Referencias:
 - Andreou, Kamm y Cohen, 1994
 - Lee y Rose, 1998

MIT Reconocimiento dependiente del hablante

- **Condiciones del experimento:**
 - Tarea de gestión de recursos DARPA (vocabulario de 1000 palabras)
 - Reconocedor SUMMIT basado en segmentos mediante gramática de pares de palabras
 - Modelos de mezclas de gaussianas para 60 unidades independientes del contexto:
 - **Grupo de entrenamiento dependiente del hablante:**
 - * 12 hablantes con 600 enunciados de entrenamiento y 100 enunciados de prueba por hablante
 - * ~80.000 parámetros en cada grupo del modelo acústico SD (dependiente del hablante)
 - **Grupo de entrenamiento independiente del hablante:**
 - * 149 hablantes con 40 enunciados de entrenamiento por hablante (5960 enunciados totales)
 - * ~400.000 parámetros en el grupo del modelo acústico SI (independiente del hablante)
- **Resultados de la tasa de error por palabra (WER) con el grupo de pruebas SD**
 - El reconocedor SI presentaba un 7'4% de WER
 - La media del reconocedor SD presentaba un 3'4% de WER
 - El reconocedor SD presentaba un 50% menos de errores utilizando un 80% menos de parámetros

Definiciones de adaptación

- **Los modelos dependientes del hablante no existen para nuevos usuarios**
- **El sistema debe aprender características de los nuevos usuarios**
- **Tipos de adaptación:**
 - **Inscrita frente a instantánea**
 - * ¿Se utiliza un grupo pregrabado de datos de adaptación o se emplean los datos de prueba como datos de adaptación?
 - **Supervisada frente a no supervisada**
 - * ¿Es la ortografía de los datos de adaptación conocida o desconocida?
 - **Por lotes frente a en línea**
 - * ¿Se presentan los datos de adaptación de repente o uno por uno?

Definiciones de adaptación (cont.)

- **Objetivo:** Ajustar parámetros del modelo para hacerlos corresponder con los datos de entrada
- **Definiciones:**
 - X es un grupo de datos de adaptación
 - Λ es un grupo de parámetros de adaptación, tal como:
 - * Género e velocidad de hablante
 - * Vector de medias de unidades fonéticas
 - * Matriz de transformación global
 - Θ es un conjunto de parámetros del modelo acústico utilizados por el reconocedor
- **Método:**
 - Λ se estima desde X
 - Θ se ajusta según Λ

Definiciones de adaptación (cont)

- **La obtención de Λ es un problema de estimación:**
 - Pocos puntos de datos de adaptación \Rightarrow n° pequeño de parámetros en Λ
 - Muchos puntos de datos de adaptación \Rightarrow n° más grande de parámetros en Λ
- **Ejemplo:**
 - Supongamos que Λ contiene sólo un único parámetro λ
 - Supongamos que λ representa la probabilidad de que el hablante sea masculino
 - λ se estima a partir de los datos de adaptación X
 - El modelo adaptado al hablante podría representarse como:

$$P(\vec{a} | \Theta_{sa}) = \lambda P(\vec{a} | \Theta_{\text{masculino}}) + (1 - \lambda) P(\vec{a} | \Theta_{\text{femenino}})$$

Adaptación bayesiana

- Un método para la adaptación directa de parámetros modelos
- Más práctico con grandes cantidades de datos de adaptación
- También conocida como adaptación de máxima probabilidad *a posteriori* (MAP)
- Expresión general para la adaptación MAP del vector de medias de una función de densidad gaussiana:

$$\vec{\mu} = \arg \max_{\vec{\mu}} p(\vec{\mu} | X) = \arg \max_{\vec{\mu}} p(\vec{\mu} | \vec{x}_1, \dots, \vec{x}_N)$$

- Aplicar la regla de Bayes:

$$\vec{\mu} = \arg \max_{\vec{\mu}} p(X | \vec{\mu}) p(\vec{\mu})$$

The diagram illustrates the decomposition of the MAP equation. The term $p(X | \vec{\mu})$ is labeled as "probabilidad observación" (observation probability), and the term $p(\vec{\mu})$ is labeled as "modelo a priori" (prior model). Red arrows point from these labels to their respective terms in the equation.

Adaptación bayesiana (cont.)

- Suponga que las observaciones son independientes:

$$p(\mathbf{X} | \vec{\mu}) = p(\vec{\mathbf{x}}_1, \dots, \vec{\mathbf{x}}_N | \vec{\mu}) = \prod_{n=1}^N p(\vec{\mathbf{x}}_n | \vec{\mu})$$

- Funciones de probabilidad modeladas con gaussianas:

$$p(\vec{\mathbf{x}} | \vec{\mu}) = \mathbf{N}(\vec{\mu}; \mathbf{S}) \quad p(\vec{\mu}) = \mathbf{N}(\vec{\mu}_{\text{ap}}; \mathbf{S}_{\text{ap}})$$

- Parámetros de adaptación hallados a partir de \mathbf{X} :

$$\Lambda = \{ \vec{\mu}_{\text{ml}}, \mathbf{N} \} \quad \vec{\mu}_{\text{ml}} = \frac{1}{N} \sum_{n=1}^N \vec{\mathbf{x}}_n$$

estimación de máx.
probabilidad (ML)

Adaptación bayesiana (cont.)

- La estimación MAP para un vector de medias parece ser:

$$\vec{\mu}_{\text{map}} = \mathbf{S}(\mathbf{N}\mathbf{S}_{\text{ap}} + \mathbf{S})^{-1} \vec{\mu}_{\text{ap}} + \mathbf{N}\mathbf{S}_{\text{ap}} (\mathbf{N}\mathbf{S}_{\text{ap}} + \mathbf{S})^{-1} \vec{\mu}_{\text{ml}}$$

- La estimación MAP es una interpolación de la media de estimaciones de ML y de la media *a priori*:

- Si N es pequeña: $\vec{\mu}_{\text{map}} \approx \vec{\mu}_{\text{ap}}$

- Si N es grande: $\vec{\mu}_{\text{map}} \approx \vec{\mu}_{\text{ml}}$

- La adaptación MAP puede ampliarse para el control de todos los parámetros de mezcla de gaussianas
 - Referencia: Gauvain y Lee, 1994

Adaptación bayesiana (cont.)

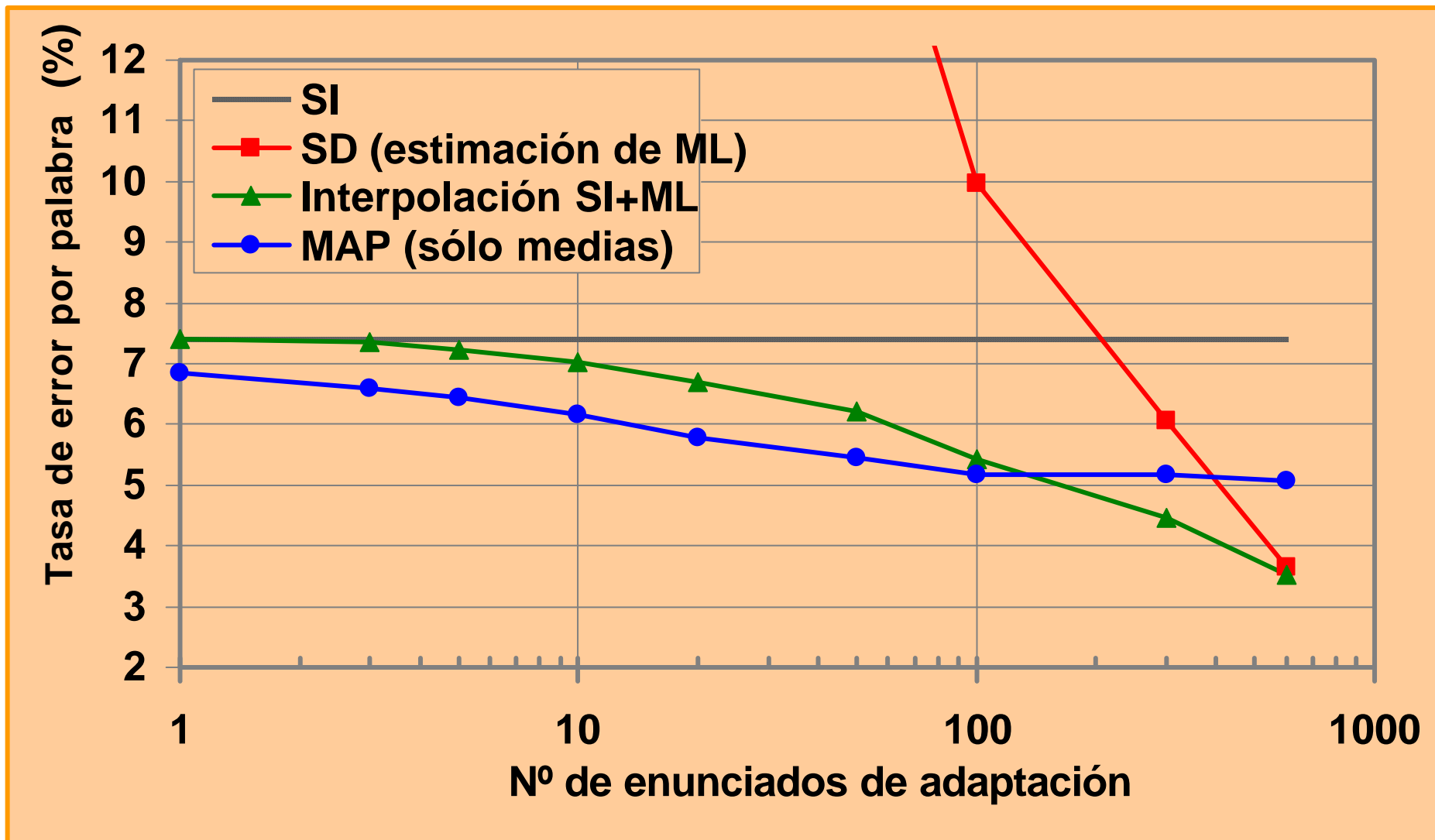
- **Ventajas de MAP:**
 - Basado en un marco matemático sólido
 - Converge con el modelo dependiente del hablante en el límite
- **Desventajas de MAP:**
 - La adaptación es muy lenta debido a la suposición de independencia
 - Es sensible a los errores durante la adaptación no supervisada
- **La adaptación de interpolación del modelo se aproxima a MAP**
 - No precisa de modelo a priori
 - También converge con el modelo dependiente del hablante en el límite
 - Expresado como:

$$p_{sa}(\vec{x}_n | \mathbf{u}) = \frac{N}{N+K} p_{ml}(\vec{x}_n | \mathbf{u}) + \frac{K}{N+K} p_{si}(\vec{x}_n | \mathbf{u})$$

K empíricamente determinada

Adaptación bayesiana (cont.)

- Grupo de pruebas SD de gestión de recursos de adaptación supervisada:



Adaptación transformacional

- Las técnicas de transformación son la forma más común de adaptación utilizada hoy en día
- Idea: Ajustar los parámetros del modelo mediante una transformación compartida globalmente o entre unidades distintas dentro de una clase
- Traslación global del vector de medias:

$$\forall \mathbf{p} \quad \bar{\mu}_p^{sa} = \bar{\mu}_p^{si} + \vec{\mathbf{v}}$$

Diagram illustrating the global translation of the mean vector. The equation shows that the source mean vector $\bar{\mu}_p^{sa}$ is equal to the target mean vector $\bar{\mu}_p^{si}$ plus a shared translation vector $\vec{\mathbf{v}}$. The universal quantifier $\forall \mathbf{p}$ and the translation vector $\vec{\mathbf{v}}$ are highlighted with red boxes. Below the equation, two red boxes provide context: "adaptar vectores de medias de todos los modelos fonéticos" points to the universal quantifier, and "vector de traslación compartido" points to the translation vector.

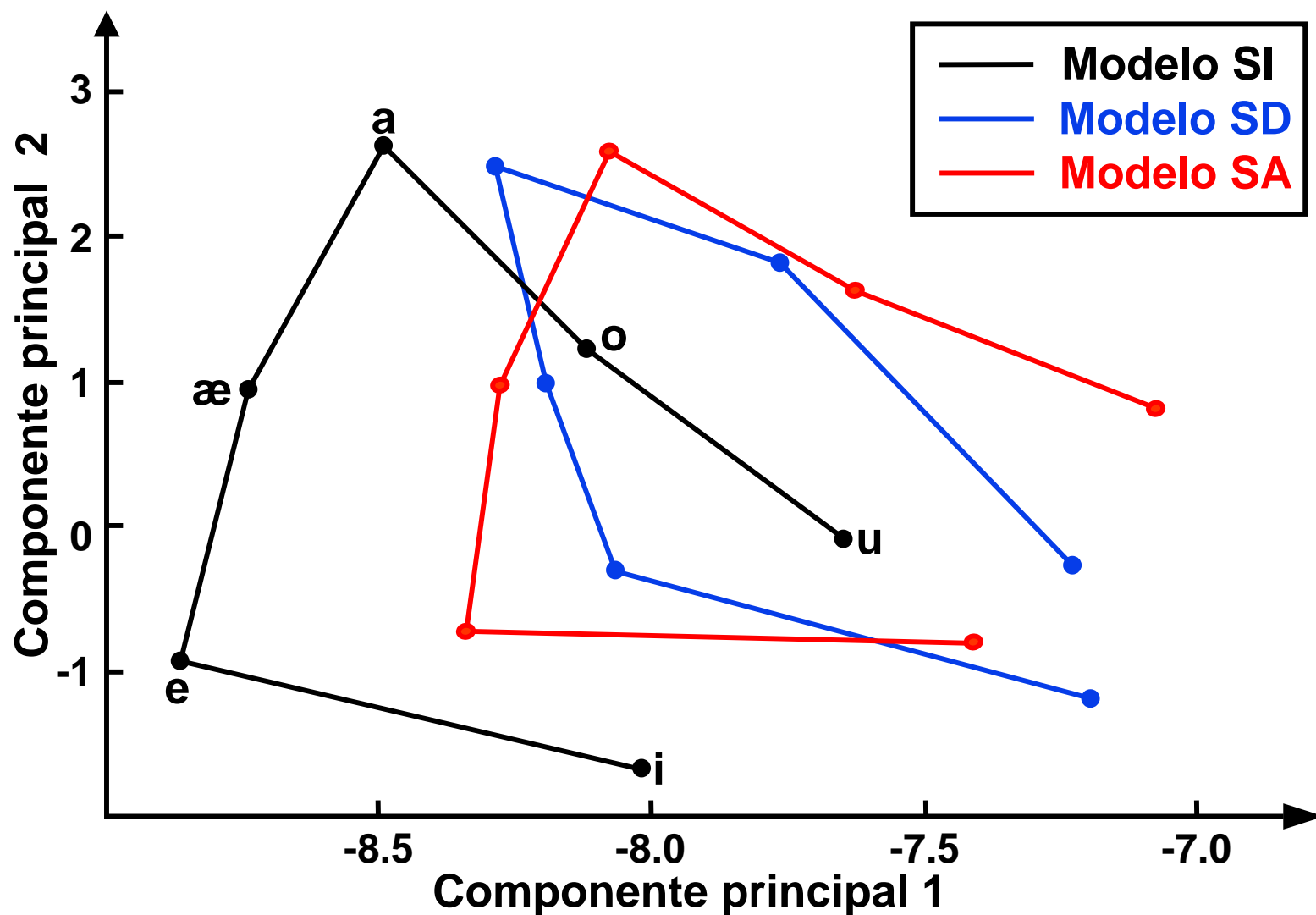
- Traslación, rotación y escalada global del vector de medias:

$$\forall \mathbf{p} \quad \bar{\mu}_p^{sa} = \mathbf{R} \bar{\mu}_p^{si} + \vec{\mathbf{v}}$$

Diagram illustrating global translation, rotation, and scaling of the mean vector. The equation shows that the source mean vector $\bar{\mu}_p^{sa}$ is equal to the target mean vector $\bar{\mu}_p^{si}$ multiplied by a shared rotation and scaling matrix \mathbf{R} , plus a shared translation vector $\vec{\mathbf{v}}$. The universal quantifier $\forall \mathbf{p}$ and the matrix \mathbf{R} are highlighted with red boxes. Below the equation, a red box labeled "escalada compartida y matriz de rotación" points to the matrix \mathbf{R} .

Adaptación transformacional (cont.)

- Modelo SI rotado, escalado y trasladado para ajustarse al modelo SD:



Adaptación transformacional (cont.)

- Parámetros de transformación hallados mediante la estimación de ML:

$$[\mathbf{R}, \vec{\mathbf{v}}] = \arg \max_{\mathbf{R}, \vec{\mathbf{v}}} p(\mathbf{X} | \mathbf{R}, \vec{\mathbf{v}})$$

- Ventajas:

- Los modelos de unidades sin datos de adaptación se adaptan según las observaciones de otras unidades
- No precisa modelo *a priori* (Puede que esto sea un punto débil)

- Desventajas:

- Funciona pobremente (peor que MAP) con pequeñas cantidades de datos
- Supone que todas las unidades podrían adaptarse de la misma forma

- La técnica se conoce comúnmente como regresión lineal de probabilidad máxima (MLLR)

- Referencia: Leggetter & Woodland, 1995

Coefficiente de ponderación del hablante de referencia

- La interpolación de modelos a partir de "hablantes de referencia"
 - Se aprovecha de relaciones fonéticas internas al hablante
- Ejemplo que utiliza vectores de medias a partir de hablantes de entrenamiento:
 - Los datos de entrenamiento contienen hablantes de referencia R
 - El reconocedor contiene modelos fonéticos P
 - Una media se entrena para cada modelo p y cada hablante r: $\vec{\mu}_{p,r}$
 - Se crea una matriz de *vectores de hablantes* a partir de medias entrenadas:

$$\vec{m}_r = \begin{bmatrix} \vec{\mu}_{1,r} \\ \vdots \\ \vec{\mu}_{P,r} \end{bmatrix}$$

vector del hablante

$$M = \begin{bmatrix} \vec{\mu}_{1,1} & \cdots & \vec{\mu}_{1,R} \\ \vdots & \ddots & \vdots \\ \vec{\mu}_{P,1} & \cdots & \vec{\mu}_{P,R} \end{bmatrix}$$

matriz del hablante

cada columna es un
vector del hablante

Coeficiente de ponderación del hablante de referencia (cont.)

- El objetivo consiste en hallar el vector de hablante más probable para el nuevo hablante
- Hallar la combinación de pesos de los vectores del hablante de referencia:

$$\vec{m}_{sa} = M\vec{w}$$

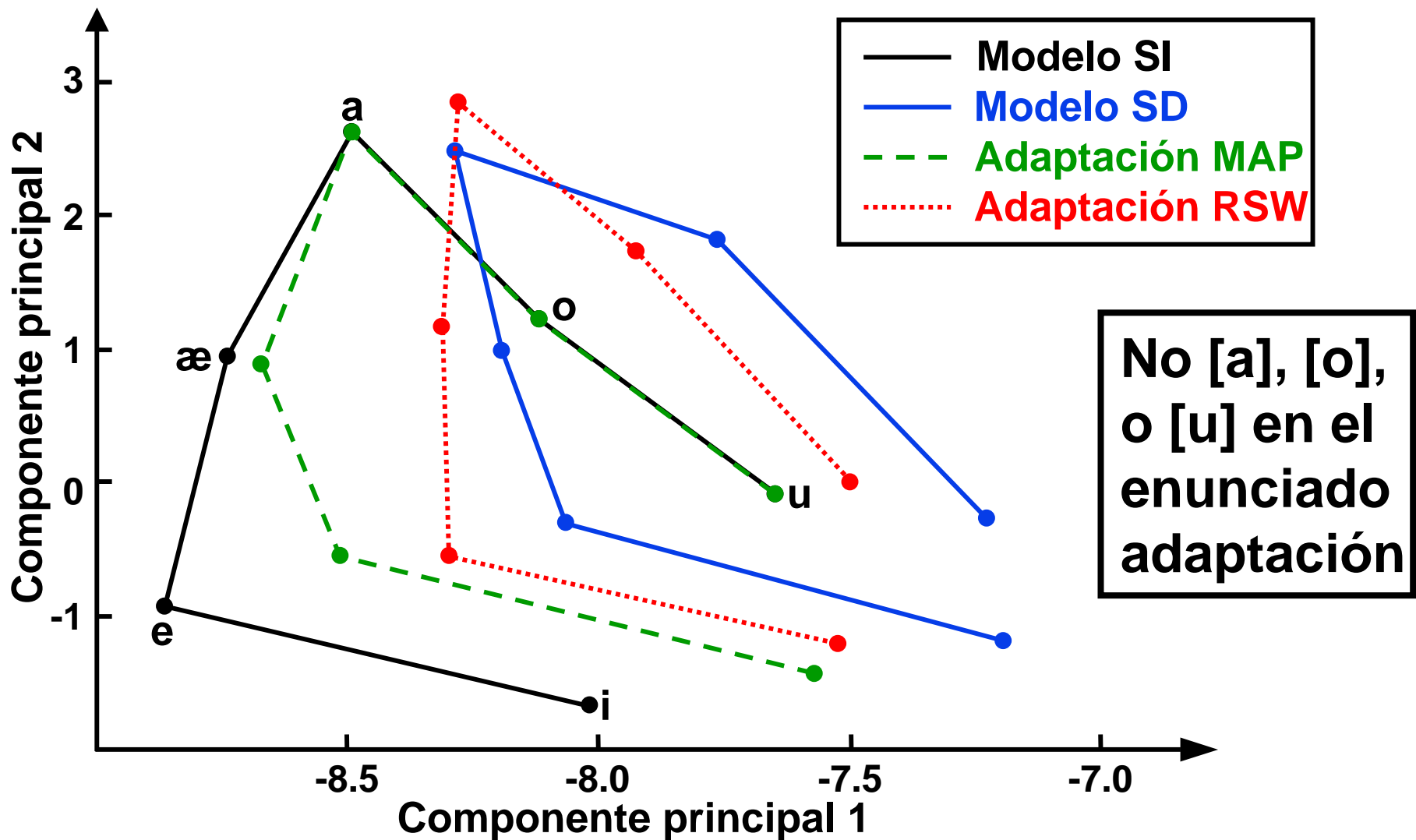
- Estimación de probabilidad máxima del vector de ponderación:

$$\vec{w} = \underset{\vec{w}}{\operatorname{argmax}} p(X | M, \vec{w})$$

- El vector de ponderación global es robusto frente a los errores introducidos durante la adaptación no supervisada
- Los métodos iterativos pueden utilizarse para hallar el vector de ponderación
 - Referencia: Hazen, 1998

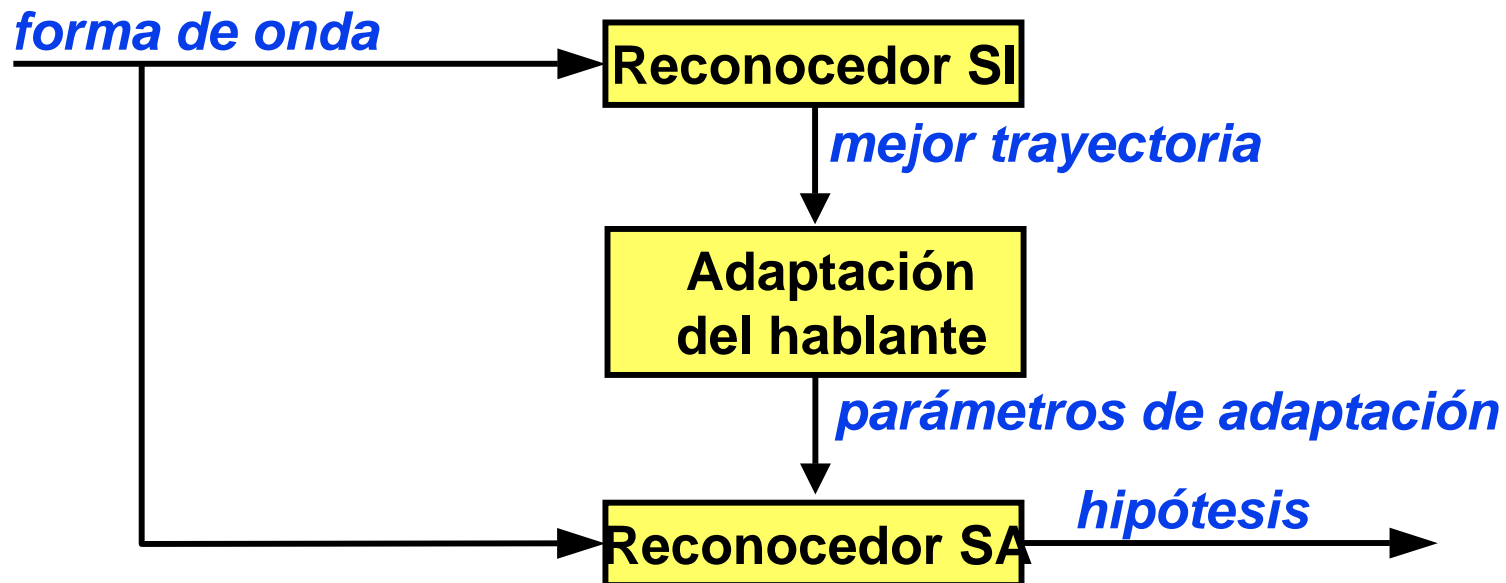
Coeficiente de ponderación del hablante de referencia (cont.)

- La adaptación del vector de medias con un enunciado de adaptación:



Arquitectura de adaptación no supervisada

- Arquitectura del sistema de adaptación no supervisada:



- En modo fuera de línea, los modelos adaptados se utilizan para volver a reconocer la forma de onda original
 - Conocido en ocasiones como adaptación instantánea
- En modo en línea, los modelos SA se utilizan en la siguiente forma de onda

Experimento de adaptación no supervisada

- **Adaptación no supervisada, instantánea**
 - Adaptar y probar en el mismo enunciado
 - No supervisada \Rightarrow los errores de reconocimiento afectan a la adaptación
 - Instantáneo \Rightarrow los errores de reconocimiento se reafirman

Método de adaptación	WER	Reducción
SI	8.6%	---
Adaptación MAP	8.5%	0.8%
Adaptación RSW	8.0%	6.5%

- **RSW es más robusto frente a errores que MAP**
 - La estimación RSW es “global” \Rightarrow emplea el enunciado completo
 - La estimación MAP es “local” \Rightarrow emplea sólo una clase fonética

MIT *Eigenvoices*

- Las *eigenvoices* amplían ideas de la ponderación del hablante de referencia
 - Referencia: Kuhn, 2000
- El objetivo consiste en aprender rasgos no correlacionados del espacio del hablante
- Comenzar creando una matriz del hablante:

$$\mathbf{M} = \begin{bmatrix} \vec{\mu}_{1,1} & \cdots & \vec{\mu}_{1,R} \\ \vdots & \ddots & \vdots \\ \vec{\mu}_{P,1} & \cdots & \vec{\mu}_{P,R} \end{bmatrix}$$

- Realizar análisis Eigen (componentes principales) de M
 - Cada eigenvector representa una dimensión independiente (ortogonal) en el espacio del hablante
 - Las dimensiones de ejemplo que este método típicamente aprende son género, volumen, monotonidad, etc.

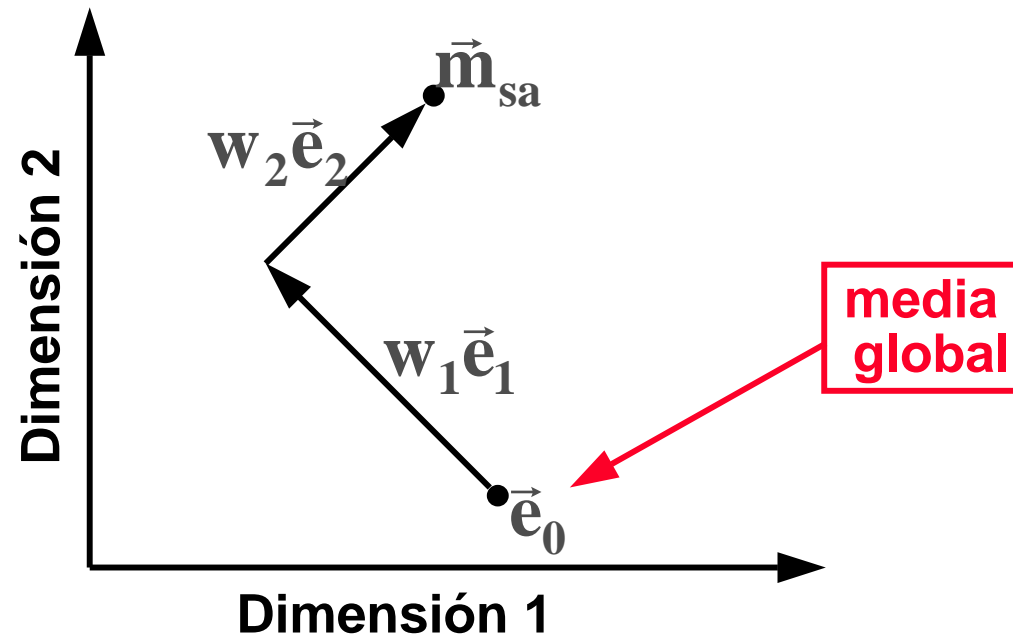
Eigenvoces (cont.)

- Hallar eigenvectores R:

$$\mathbf{E} = \{ \vec{e}_0; \vec{e}_1; \dots; \vec{e}_R \}$$

- El vector del nuevo hablante es una combinación de vectores superiores N:

$$\vec{m}_{sa} = \vec{e}_0 + w_1 \vec{e}_1 + \dots + w_N \vec{e}_N$$



Eigenvoces (cont.)

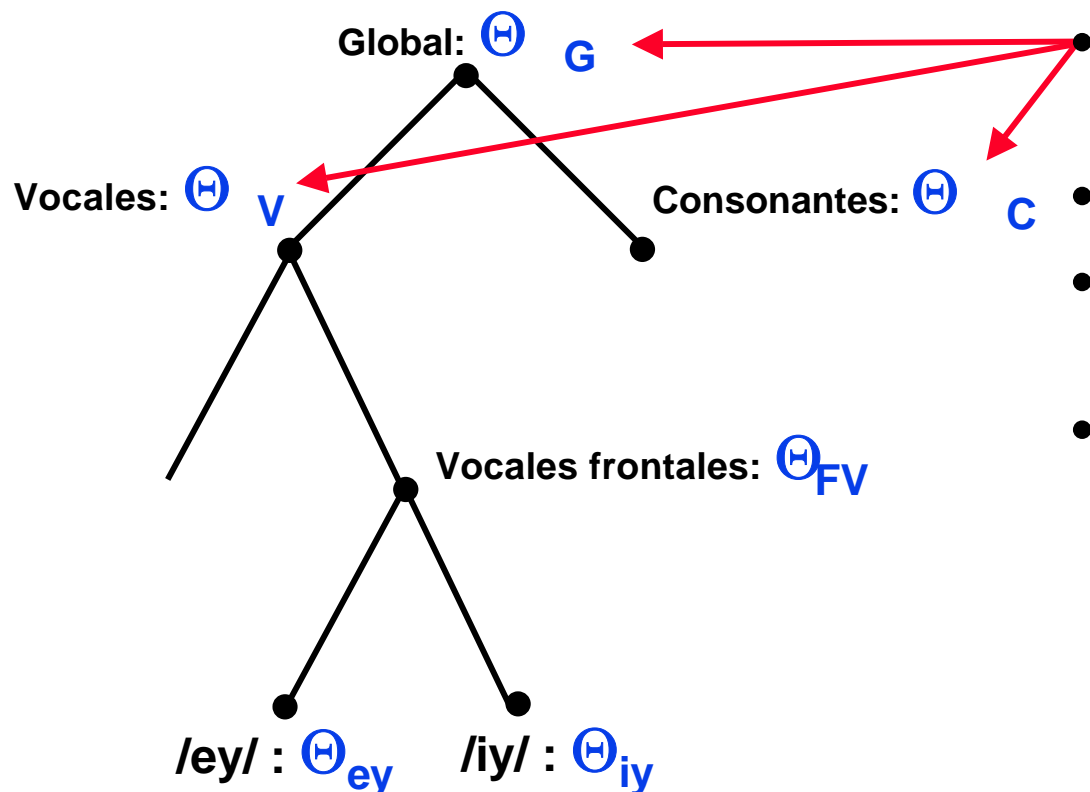
- El procedimiento de adaptación es muy similar a RSW:

$$\vec{w} = \arg \max_{\vec{w}} p(X | E, \vec{w})$$

- La adaptación de eigenvoces puede ser muy rápida
 - Unos cuantos vectores pueden generalizarse a muchos tipos de hablantes
 - Sólo se necesitan un pequeño número de adaptaciones fonéticas para obtener beneficios considerables

MIT Adaptación estructural

- **Parámetros de adaptación organizados en estructura arbórea**
 - El nodo raíz representa la adaptación global
 - Los nodos ramas realizan la adaptación en clases compartidas de modelos
 - Los nodos hojas realizan la adaptación específica al modelo



Parámetros de adaptación aprendidos para cada nodo del árbol

- Cada nodo tiene un peso: W_{nodo}
- Pesos basados en la disponibilidad de los datos de adaptación
- Cada camino desde la raíz a la hoja sigue esta restricción:

$$\sum_{\forall \text{node} \in \text{path}} W_{\text{nodo}} = 1$$

Adaptación estructural

- **Adaptación estructural basada en la combinación ponderada de la adaptación realizada en cada nodo del árbol:**

$$p_{sa}(\vec{x} | u, \text{tree}) = \sum_{\forall \text{nodes} \in \text{path}(u)} w_{\text{node}} p(\vec{x} | u, \Theta_{\text{node}})$$

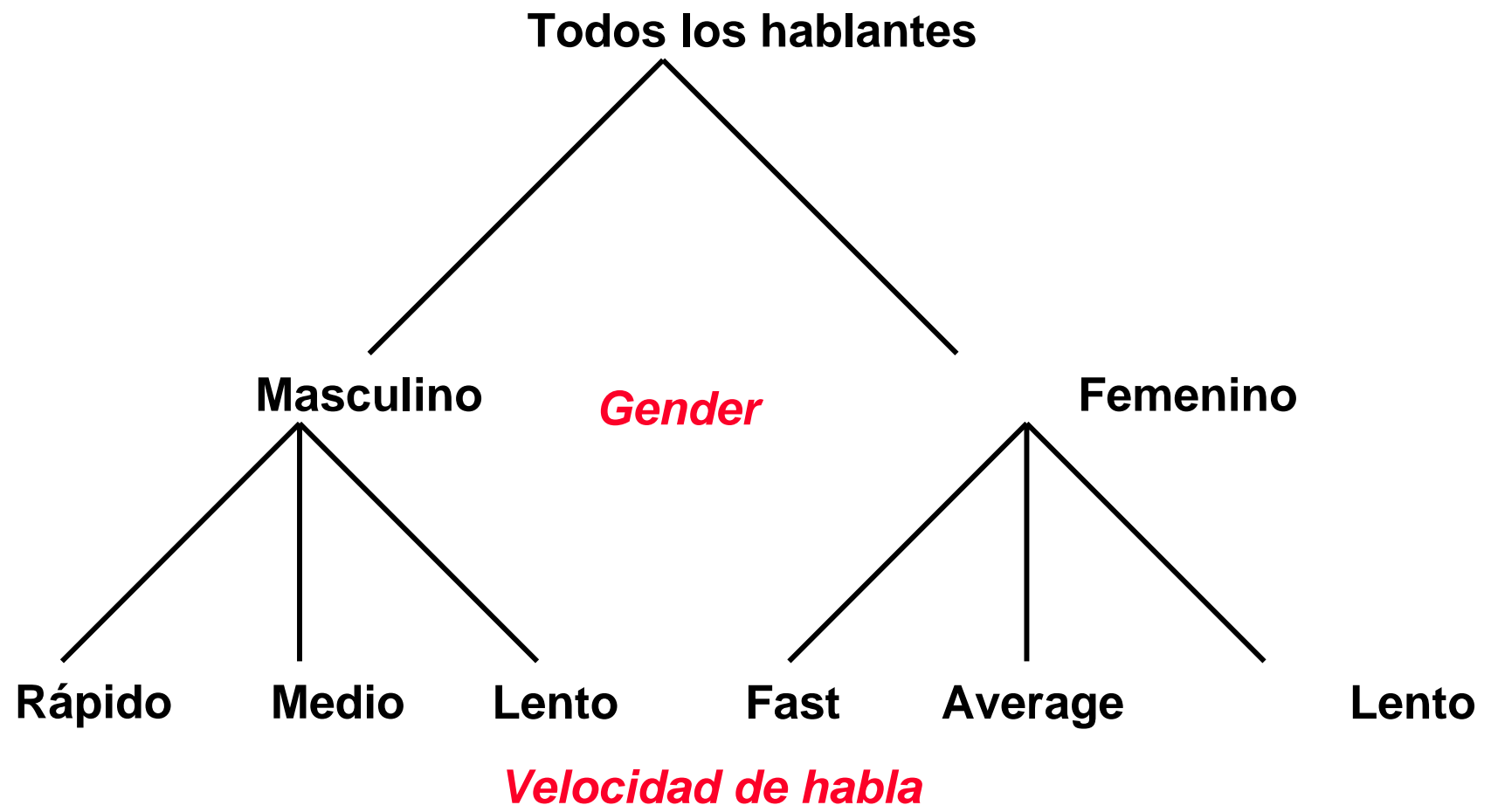
- **La adaptación estructural se ha aplicado a una variedad de técnicas de adaptación del hablante:**
 - **MAP** (Referencia: Shinoda & Lee, 1998)
 - **RSW** (Referencia: Hazen, 1998)
 - **Voces eigen** (Referencia: Zhou & Hanson, 2001)
 - **MLLR** (Referencia: Siohan, Myrvoll & Lee, 2002)

Agrupamiento jerárquico de hablantes

- **Idea:** Usar el modelo entrenado a partir de un grupo de hablantes más parecido al hablante actual
- **Enfoque:**
 - Se crea un árbol jerárquico usando hablantes del grupo de entrenamiento
 - El árbol separa a los hablantes en clases parecidas
 - Modelos distintos construidos para cada nodo del árbol
 - Un hablante de prueba es comparado con todos los nodos del árbol
 - El modelo del mejor nodo asociado se utiliza durante el reconocimiento
- **Los hablantes pueden agruparse...**
 - ...manualmente basándose en propiedades predefinidas del hablante
 - ...automáticamente basándose en la similitud acústica
- **Referencias:**
 - Furui, 1989
 - Kosaka y Sagayama, 1994

Agrupamiento jerárquico de hablantes

- Ejemplo de una jerarquía de hablantes creada manualmente:



Agrupamiento jerárquico de hablantes (cont.)

- **Problema: Modelo más específico \Rightarrow menos datos de entrenamiento**
- **Equilibrio entre robustez y precisión**
- **Una solución: interpolar modelos generales y específicos**
- **Ejemplo de combinación del modelo de ML entrenado dependiente del género con el modelo IS, para obtener el modelo interpolado dependiente del género:**

$$p_{igd}(\vec{x}_n | \mathbf{u} = \mathbf{p}) = \lambda p_{mlgd}(\vec{x}_n | \mathbf{u} = \mathbf{p}) + (1 - \lambda) p_{si}(\vec{x}_n | \mathbf{u} = \mathbf{p})$$

- **λ valores hallados usando la interpolación eliminada**
 - Referencia: X.D. Huang, *et al*, 1996

Coeficiente de ponderación del grupo de hablantes

- El agrupamiento jerárquico de hablantes selecciona un modelo
- La ponderación del grupo de hablantes combina modelos:

$$p_{sa}(\vec{x}_n | u = p) = \sum_{m=1}^M w_m p_m(\vec{x}_n | u = p)$$

- Pesos determinados utilizando el algoritmo de EM
- Los pesos pueden ser globales y basados en clases
- Ventaja: Decisiones *fáciles* menos rígidas que las *difíciles*
 - Referencia: Hazen, 2000
- Desventaja:
 - El tamaño del modelo podría agrandarse demasiado con muchos grupos
 - Hacen falta métodos de aproximación para el tiempo real
 - Referencia: Huo, 2000

Experimento de agrupamiento de hablantes

- Experimento de adaptación instantánea no supervisada
 - Grupo de pruebas IS de gestión de recursos
- Los modelos de grupos de hablantes usados para la adaptación:
 - 1 modelo SI
 - 2 modelos dependientes del género
 - 6 modelos dependientes del género y de la velocidad del habla

Modelos	WER	Reducción
SI	8.6%	---
Dependiente del género	7.7%	10.5%
Dependiente del género y de la velocidad	7.2%	16.4%
Interpolación del grupo de hablantes	6.9%	18.9%

- **La adaptación mejora el reconocimiento al restringir los modelos a características del hablante actual**
- **Buenas propiedades de los algoritmos de adaptación:**
 - **explican un conocimiento a priori sobre hablantes**
 - **son capaces de adaptar modelos de unidades que no son observados**
 - **ajustan el número de parámetros de adaptación a la cantidad de datos**
 - **son robustos frente a errores durante la adaptación no supervisada**
- **La adaptación es importante para aplicaciones del "mundo real"**

- A. Andreou, T. Kamm y J. Cohen, “Experiments in vocal tract normalization,” *CAIP Workshop: Frontiers in Speech Recognition II*, 1994.
- S. Furui, “Unsupervised speaker adaptation method based on hierarchical spectral clustering,” ICASSP, 1989.
- J. Gauvain y C. Lee, “Maximum *a posteriori* estimation for multivariate Gaussian mixture observation of Markov chains,” *IEEE Trans. On Speech and Audio Processing*, April 1994.
- T. Hazen, *The use of speaker correlation information for automatic speech recognition*, PhD Thesis, MIT, January 1998.
- T. Hazen, “A comparison of novel techniques for rapid speaker adaptation,” *Speech Communication*, May 2000.
- X.D. Huang, *et al*, “Deleted interpolation and density sharing for continuous hidden Markov models,” ICASSP 1996.
- Q. Huo y B. Ma, “Robust speech recognition based on off-line elicitation of multiple priors and on-line adaptive prior fusion,” ICSLP, 2000.

References

- T. Kosaka y S. Sagayama, “Tree structured speaker clustering for speaker-independent continuous speech recognition,” *ICASSP*, 1994.
- R. Kuhn, *et al*, “Rapid speaker adaptation in Eigenvoice Space,” *IEEE Trans. on Speech and Audio Processing*, November 2000.
- L. Lee y R. Rose, “A frequency warping approach to speaker normalization,” *IEEE Trans. On Speech and Audio Proc.*, January 1998.
- C. Leggetter y P. Woodland, “Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models,” *Computer Speech and Language*, April 1995.
- K. Shinoda y C. Lee, “Unsupervised adaptation using a structural Bayes approach,” *ICASSP*, 1998.
- O. Siohan, T. Myrvoll y C. Lee, “Structural maximum a posteriori linear regression for fast HMM adaptation,” *Computer Speech and Language*, January 2002.
- B. Zhou y J. Hanson, “A novel algorithm for rapid speaker adaptation based on structural maximum likelihood Eigenspace mapping,” *Eurospeech*, 2001.