



# FUNDAMENTOS DEL RECONOCIMIENTO AUTOMÁTICO DE LA VOZ



## *“Métodos de encaje de patrones”*

Agustín Álvarez Marquina



### Introducción. Encaje de patrones



#### Objetivos:

- Asociar secuencias temporales de vectores de rasgos agrupados de la fase anterior con las unidades de reconocimiento.
- Proporcionar una medida del grado de encaje de dichas secuencias supuesta la hipótesis de su pertenencia a una determinada unidad de reconocimiento.



- **Distorsión dinámica temporal.**
- **Redes neuronales.**
- **Modelos ocultos de Markov.**



- **El método de distorsión dinámica temporal o *DTW* (*Dynamic Time Warping*) [SAK78], [SAK92] es uno de los algoritmos aplicados en reconocimiento de voz [RAB81] más antiguos e importantes.**
  - En la actualidad ha cedido paso a otros procedimientos como son los modelos ocultos de Markov.
    - Si bien esta técnica aún se continúa utilizando, tiene un número de limitaciones que restringen su uso a sistemas con vocabularios pequeños.
    - En sistemas de mayor tamaño, el número de plantillas a generar y el coste computacional de las búsquedas es intratable.





## Distorsión dinámica temporal (II)



- La manera más fácil de reconocer una palabra aislada pasa por compararla con un conjunto de plantillas previamente almacenadas y determinar cual es la que proporciona un mejor encaje.
  
- Sin embargo, este objetivo se complica por dos factores:
  - La duración de la palabra no tiene que ser la misma que la de las plantillas.
  - El ritmo con el que pronuncia esa palabra no tiene por que ser constante.



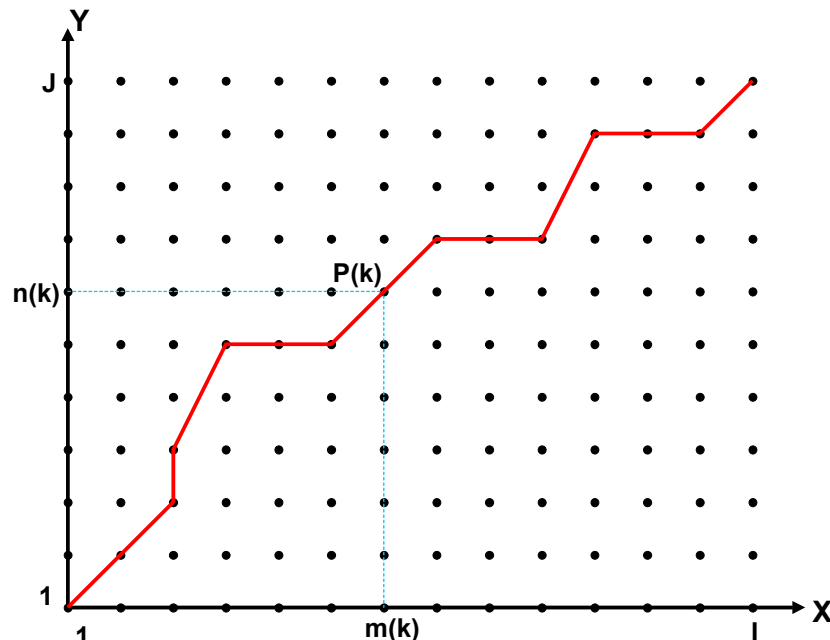
## Distorsión dinámica temporal (III)



- Resumiendo, el alineamiento óptimo, entre plantillas almacenadas y las plantillas producidas en un determinado momento, puede ser no lineal.



- Sean  $X = (x_1, x_2, \dots, x_i)$  e  $Y = (y_1, y_2, \dots, y_j)$  dos patrones de voz aislados (ej. palabras). La disparidad promedio entre  $X$  e  $Y$ ,  $D(X, Y)$ , se basa en alguna medida de distancia entre los vectores  $x_i$  e  $y_j$  que denotaremos como  $d(i, j)$ .
- La distorsión temporal no lineal de los patrones  $X$  e  $Y$  puede representarse por un camino  $\{ P(k) = (m(k), n(k)), k = 1, K \}$  en el plano  $(i, j)$  definida por las dos secuencias de vectores  $X$  e  $Y$ .



**Figura 1.** Ejemplo de camino de distorsión temporal para los patrones de voz  $X$  e  $Y$ .

- La disparidad entre las tramas  $X$  e  $Y$  a lo largo de un camino  $P$  viene dada por:

$$D_P(X, Y) = \sum_{k=1}^K d(P(k)) \cdot \frac{w(k)}{N(w)}$$

- $w(k)$  es un peso y  $N(w)$  es el factor de normalización. Existe un número de posibles caminos  $P(k)$ , que corresponden a diferentes funciones de distorsión para los patrones de voz.

- El objetivo será encontrar el camino que minimice  $D_P(X, Y)$ , siendo una elección natural el tomar el mínimo sobre todos los posibles caminos:

$$D(X, Y) = \min_P (D_P(X, Y))$$

- Para resolver este problema se emplean técnicas inspiradas en los algoritmos de programación dinámica.



## Distorsión dinámica temporal (VIII)



- Con objeto de tener en cuenta ciertos aspectos físicos del problema y limitar el número de caminos a considerar, se imponen algunas restricciones a las funciones de distorsión:

- Puntos de inicio y final de tramas.

$$P(1) = (1, 1)$$

$$P(K) = (I, J)$$



## Distorsión dinámica temporal (IX)



- El camino no podrá tener una pendiente negativa.
- Continuidad Local. Para minimizar la pérdida de información se restringen los movimientos locales. Un ejemplo de restricciones recogidas en [RAB93] es:

$$g(i, j) = \min \begin{cases} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i, j-1) + d(i, j) \end{cases}$$





## Distorsión dinámica temporal (X)



- Restricciones de pendiente. Se expresan como funciones de costo para el cálculo del peso  $w(k)$ .
  - Algunos ejemplos que aparecen en [SAK78] son:

$$w(k) = n(k) - n(k-1) + m(k) - m(k-1)$$

$$w(k) = n(k) - n(k-1)$$

$$w(k) = m(k) - m(k-1)$$

$$w(k) = \{n(k) - n(k-1) + m(k) - m(k-1)\}$$



## Distorsión dinámica temporal (XI)



- **Por su parte, el factor de normalización suele ser:**

$$N(w) = I + J$$

- **Finalmente, para el caso descrito anteriormente, el algoritmo de programación dinámica puede expresarse de la siguiente manera:**



### 1 Inicio:

$$g(1,1) = d(1,1) \cdot w(1)$$

### 2 Recursión:

Para  $1 \leq i \leq I$

Para  $1 \leq j \leq J$

$$g(i, j) = \min \begin{cases} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i, j-1) + d(i, j) \end{cases}$$

### 3 Terminación:

$$D(X, Y) = g(I, J) / (I + J)$$

○ Las técnicas de distorsión temporal se han aplicado con éxito a los problemas de reconocimiento de palabras aisladas y conectadas (series de palabras concatenadas sin la presencia de pausas artificiales entre ellas).

- Sin embargo para este último caso, la potencia computacional requerida puede ser excesiva.





## Distorsión dinámica temporal (XIV)



○ Con el objetivo de aliviar este problema se han propuesto diferentes técnicas:

- Algoritmo de programación dinámica en dos niveles (*two-level dynamic programming algorithm*).
  - La idea es dividir el proceso de encaje de patrones en dos fases. Durante la primera se trabaja en el nivel de palabra y durante la segunda se trata el nivel de habla conectada.



## Distorsión dinámica temporal (XV)



- Algoritmo de construcción de nivel (*level building algorithm*).
  - Es una implementación más eficaz que la anterior, conseguida a través de la realización de decisiones parciales sobre las palabras durante la programación dinámica.
- Algoritmo de pasada única (*one-pass algorithm*).
  - Resulta muy útil para aplicaciones en tiempo real ya que su computación se realiza trama a trama de forma síncrona.

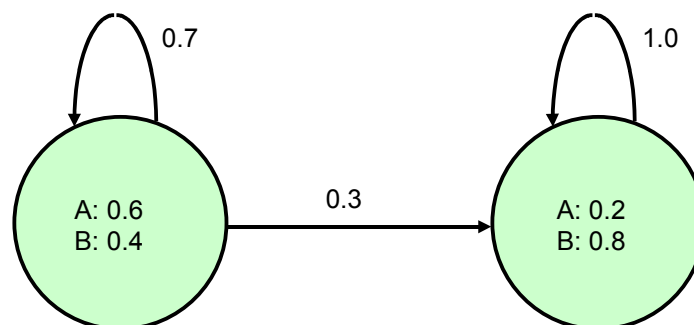




- Las redes neuronales permiten integrar de manera sencilla las fases de clasificación y de encaje de patrones.
- La estructura de las redes empleadas en reconocimiento de voz deben contemplar los aspectos dinámicos del proceso (redes recurrentes o con retardo temporal).
- Su empleo en esta fase es muy poco usual por los problemas para poder definir una organización jerárquica útil.



- Un modelo oculto de Markov o *HMM (Hidden Markov Model)* [BAK75], [BAH83], [RAB86], [POR88], [RAB89], [PIC90], es una colección de estados conectado por transiciones.



**Figura 2.** Modelo oculto de Markov con 2 estados y 2 símbolos de salida A y B.





## Modelos ocultos de Markov (II)



- El proceso comienza en un estado/s diseñado/s para tal efecto.
- En cada paso de ejecución se toma una transición a un nuevo estado y se genera un símbolo de salida asociado a ese estado.
- La elección de la transición y del símbolo es aleatoria y está gobernada por distribuciones de probabilidad.



## Modelos ocultos de Markov (III)



- Un modelo oculto de Markov puede considerarse como un caja negra donde la secuencia de símbolos de salida generados a lo largo del tiempo es visible, pero la secuencia de estados por los que se ha pasado para generar la anterior serie se desconoce.
- Esto es por lo que se llaman modelos ocultos.





## Modelos ocultos de Markov (IV)



### ○ Cuando se aplican los modelos ocultos de Markov al reconocimiento de la voz:

- Los estados se interpretan como modelos acústicos, indicando las ocurrencias de sonidos que son más probables durante los correspondientes segmentos de habla.
- Las transiciones incorporan restricciones de tipo temporal acerca de cómo son las secuencias de aparición de esos sonidos.



## Modelos ocultos de Markov (V)



### ○ Estados y transiciones pueden ser utilizados para modelar distintas jerarquías del proceso del habla: desde fonemas hasta oraciones pasando por palabras.

### ○ La utilización de este tipo de estructuras debe resolver 3 problemas básicos:

- ① Problema de reconocimiento.
- ② Problema de decodificación.
- ③ Problema de aprendizaje o entrenamiento.



### 1 Problema de reconocimiento.

- Consiste en escoger aquel modelo de entre un grupo de éstos, que mejor represente al conjunto de etiquetas obtenidas a partir de la cuantificación de las correspondientes plantillas espectrales.
- Formalmente, conocida una secuencia de observaciones  $O = \{o_1, o_2, \dots, o_T\}$  y el correspondiente modelo de Markov  $\lambda = (A, B, \pi)$ , siendo  $A = \{a_{ij}\}$  la distribución de probabilidad para la transición entre los estados  $i$  y  $j$ ,  $B = \{b_j(k)\}$  la distribución de probabilidad para la observación  $k$  en el estado  $j$  y  $\pi$  la distribución inicial de estados, se pretende calcular  $Prob(O | \lambda)$ .

- Una solución eficiente a este problema es el algoritmo hacia adelante y atrás (*forward-backward algorithm*) [BAU70], [RAB86]:

$$Prob(O/\lambda) = \sum_{i=1}^N \alpha_T(i)$$

$$\alpha_1(i) = \chi_i b_i(o_1), \quad 1 \leq i \leq N$$

$$\alpha_{t+1}(j) = \left[ \sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}), \quad 1 \leq t \leq T-1$$
$$1 \leq j \leq N$$



## Modelos ocultos de Markov (VIII)



### 2 Problema de decodificación.

- Descubrir la secuencia oculta de estados, es decir, conocida una secuencia de observaciones  $O = \{o_1, o_2, \dots, o_T\}$  calcular la secuencia de estados  $I = \{i_1, i_2, \dots, i_T\}$ , que sea óptima siguiendo algún criterio.
- Este segundo problema puede resolverse a partir del algoritmo de Viterbi [VIT67].
  - Algoritmo similar al anterior con la excepción de que en vez de tomar la suma de los valores de probabilidad en los anteriores estados se toma el máximo.



## Modelos ocultos de Markov (IX)



- Para ello se define una función de máxima probabilidad  $\delta_t(i)$ , siendo  $i$  el estado y  $t$  el instante de tiempo:

$$\delta_1(i) = \chi_i b_i(o_1)$$

$$\delta_t(j) = \left[ \max_{1 \leq i \leq N} (\delta_{t-1}(i) \cdot a_{ij}) \right] \cdot b_j(o_t) \quad \begin{matrix} 2 \leq t \leq T \\ 1 \leq j \leq N \end{matrix}$$



### ③ Problema de aprendizaje o entrenamiento.

- Consiste en construir un modelo de manera que recoja el conocimiento con el que se ha entrenado de forma óptima, es decir, construir el modelo  $\lambda = (A, B, \pi)$  para que  $Prob(O | \lambda)$  sea máxima.
- El caso se resuelve a través de un algoritmo iterativo de tipo gradiente que se conoce como algoritmo Baum-Welch [BAU70], [BAH83]. Este procedimiento proporciona un método para reestimar unos nuevos valores de  $a'_{ij}$  y  $b'_j(k)$  de forma que se cumpla que  $Prob(O | \lambda') \geq Prob(O | \lambda)$ :

$$a'_{ij} = \frac{\sum_{t=1}^T \gamma_t(i, j)}{\sum_{t=1}^T \sum_{j=1}^N \gamma_t(i, j)} \quad b'_j(k) = \frac{\sum_{t=1}^T \sum_{j=1}^N \gamma_t(i, j)}{\sum_{t=1}^T \sum_{j=1}^N \gamma_t(i, j)}$$

$$\gamma_t(i, j) = \frac{\alpha_t(i) \cdot a_{ij} b_j(o_{t+1}) \cdot \beta_{t+1}(j)}{P(O | \lambda)}$$

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \cdot \beta_{t+1}(j)$$



## Modelos ocultos de Markov (XII)



- El esquema de modelos ocultos de Markov que hemos considerado hasta ahora corresponde a una situación en la cual el espacio acústico se ha dividido en un número moderado de regiones como resultado del proceso de cuantificación vectorial.
- Así la distribución de probabilidad a lo largo del espacio acústico se representa únicamente por un simple histograma para cada una de las entradas del libro de códigos empleado.



## Modelos ocultos de Markov (XIII)



- Este enfoque, que se conoce como de modelos de densidad discreta, es computacionalmente muy efectivo pero presenta el problema de decidir el tamaño de los libros de códigos.
  - Un tamaño pequeño introducirá mucho ruido de cuantificación y un tamaño grande puede hacer que muchos centroides no estén lo suficientemente representados en los datos de entrenamiento y producir con ello una degradación en las tasas de reconocimiento.







## Modelos ocultos de Markov (XIV)



- Una solución a este problema puede ser el empleo de modelos de densidad continua [LIP82].
- En este caso la distribución de probabilidad a través del espacio acústico se modela directamente asumiendo que presenta una determinada forma paramétrica, generalmente en la forma de una mezcla de K distribuciones Gaussianas [RAB89].



## Modelos ocultos de Markov (XV)



$$b_j(y) = \sum_{k=1}^K c_{jk} G(y, \mu_{jk}, C_{jk})$$

- En este caso los parámetros  $b_j(y)$  dependen de un conjunto de pesos  $c_{jk}$ , que ponderan a cada Gaussiana G de media  $\mu_{jk}$  y covarianza  $C_{jk}$ . La suma de todos los pesos de ponderación de las Gaussianas debe ser igual a 1.





## Modelos ocultos de Markov (XVI)



- Este tipo de modelos proporciona para problemas con un gran número de vocabulario el mejor resultado, aunque con un coste computacional de 2 órdenes de magnitud mayor.
- El principal inconveniente que presenta esta aproximación es que los parámetros no están compartidos por todos los estados.
  - Si el valor de  $K$  es grande podemos tener una falta de representatividad para muchos de los estados y un valor pequeño de  $K$  haría errónea la asunción de que la distribución de probabilidad puede modelarse por medio de una mezcla de Gaussianas.



## Modelos ocultos de Markov (XVII)



- Por último existe un caso intermedio que se conoce como modelos de densidad semicontinua [HUA90].
- En este caso existe un libro de códigos común para todos los estados, que describe el conjunto de clases acústicas.
  - Sin embargo aquí las clases no se representan por centroides discretos sino por medio de funciones de densidad continua (típicamente Gaussianas) a través del espacio de una clase.





## Modelos ocultos de Markov (XVIII)



- De esta forma se evitan los errores de cuantificación que se producen cuando un vector se transforma automáticamente en el centroide de la clase a la que pertenece.
- Al mismo tiempo este enfoque proporciona una solución mejor en cuanto a tiempos de ejecución.

$$b_j(y) = \sum_{k=1}^L c_{jk} G(y, \mu_k, C_k)$$



## Modelos ocultos de Markov (XIX)



- El número de elementos del libro de códigos aparece representado por  $L$  y se corresponde también con el número de Gaussianas.
- Estas funciones como en el caso anterior se recalculan durante el entrenamiento al mismo tiempo que los parámetros de los modelos ocultos.
  - Esta es otra diferencia con los modelos discretos donde el libro de códigos permanece fijo durante todo el proceso.





## Modelos ocultos de Markov (XX)



○ Para los diferentes tipos de estructuras de los HMM estudiadas, encontramos dos variantes que merece la pena comentar:

- La primera variante consiste en asociar las observaciones con los arcos mas que con los estados de un HMM [BAH83].
  - En este tipo de modelo, podemos definir transiciones nulas que no producen ninguna salida. Estas transiciones proporcionan una manera eficaz de describir fenómenos de elisión fonética.



## Modelos ocultos de Markov (XXI)



- La segunda variante se corresponde con el concepto de **unión de parámetros** (*parameter tying*) [BAH83].
- La idea es reducir el número de parámetros independientes por medio del establecimiento de relaciones de equivalencia entre parámetros pertenecientes a diferentes estados.
  - Este método es particularmente interesante cuando la cantidad de datos de entrenamiento para generar los modelos es insuficiente.
  - Es importante mencionar que los aspectos matemáticos del algoritmo de entrenamiento no se ven afectados por la ligadura de parámetros.





## Modelos ocultos de Markov (XXII)



- Otro aspecto importante en los trabajos de investigación relacionados con los *HMM* es el punto relativo al modelado del tiempo de permanencia en un estado.
  - Lo usual es obtener las distribuciones de probabilidad que modelan la duración por cada estado a partir de los datos de entrenamiento.
  - Estas probabilidades se emplean durante una fase de postproceso con objeto de matizar las puntuaciones obtenidas en el algoritmo de Viterbi.
    - Esto puede ser especialmente útil en habla continua [ANA95], [BUR96].



## Modelos ocultos de Markov (XXIII)



- Aunque el marco teórico para incorporar informaciones referentes al tiempo dentro de los *HMM* está bien desarrollado [RAB89], su coste computacional es usualmente demasiado alto.
- El entrenamiento de los modelos ocultos de Markov, tal y como lo hemos desarrollado, se realiza siguiendo un criterio de máxima semejanza o ML (Maximum Likelihood).
  - Se intenta maximizar la probabilidad de una secuencia(s) de observaciones siendo conocido el modelo asociado.





## Modelos ocultos de Markov (XXIV)



- Aunque este criterio posee buenas propiedades asintóticas, requiere por lo general de grandes conjuntos de entrenamiento para alcanzar una estimación fiable.
  - Para resolver posibles problemas asociados con la escasez de datos, encontramos técnicas de suavizado tales como la interpolación borrada (deleted interpolation) [JEL80].



## Modelos ocultos de Markov (XXV)



- Sin embargo, encontramos en la literatura otros criterios alternativos:
  - ① Criterio de máxima información mutua o *MMI* (*Maximum Mutual Information*) [BAH86].
    - Se basa en la idea de diseñar todos los HMM al mismo tiempo, de forma que las capacidades discriminativas de cada modelo puedan ser aumentadas.
    - Suponiendo que las unidades de habla representadas por medio de los modelos HMM son equiprobables, este criterio estima los parámetros HMM haciendo máximo:





## Modelos ocultos de Markov (XXVI)



$$I_i = \max_{\lambda} \left( \log P(O^i | \lambda_i) - \log \sum_{j=1}^M P(O^i | \lambda_j) \right)$$

siendo  $M$  el número de modelos,  $\lambda_i$  el modelo correcto y  $\lambda_j$  el resto de modelos. Cuando se aplica a todas las secuencias de entrenamiento, la expresión queda de la siguiente forma:

$$I = \max_{\lambda} \left( \sum_{i=1}^M \left( \log P(O^i | \lambda_i) - \log \sum_{j=1}^M P(O^i | \lambda_j) \right) \right)$$



## Modelos ocultos de Markov (XXVII)



- El algoritmo de entrenamiento Baum-Welch puede también extenderse al caso *MMI* [GOP89], [NOR91]. Así mismo, este tipo de enfoque puede utilizarse en sistemas híbridos red neuronal/*HMM* [RIG94].

### ② Criterio de mínima información discriminante o MDI (Maximum Discrimination Information) [EPH89].

- Mejora en el proceso de entrenamiento de los *HMM* realizado mediante la selección de los parámetros que minimizan la información de discriminación entre la distribución de probabilidad de la fuente (conjunto de observaciones) y la del modelo de oculto de Markov.





## Modelos ocultos de Markov (XXVIII)



- La información de discriminación entre dos distribuciones de probabilidad  $P$  y  $Q$ , con funciones de densidad de probabilidad  $p$  y  $q$ , puede expresarse por medio de la siguiente ecuación: [EPH87]:

$$D(Q \parallel P) = \int q(y) \ln \left( \frac{q(y)}{p(y)} \right) dy$$

- La idea detrás del criterio  $MDI$  es compensar los desajustes entre las medidas y el modelo.
- El procedimiento comienza con la estimación de los  $HMM$  siguiendo el criterio  $ML$ .



## Modelos ocultos de Markov (XXIX)



- Entonces para un HMM dado se estima la distribución de probabilidad de la fuente, minimizando la información de discriminación sobre todas las distribuciones de probabilidad de la fuente que son coherentes con las medidas.
- Finalmente, dada una distribución de probabilidad de la fuente, se estima el HMM que minimiza la información de discriminación sobre el conjunto de modelos ocultos de Markov.
- La reestimación se realiza usando una variante modificada del algoritmo Baum-Welch [EPH87].







## Modelos ocultos de Markov (XXX)



### ③ Criterio del mínimo error en la clasificación o *MEC* (*Minimum Error Classification*) [BAH88a], [JUA92], [REI95], [RAH97].

- ❑ Método conocido también como entrenamiento correctivo (corrective training) o entrenamiento discriminativo (discriminative training).
- ❑ Constituye una fase de postproceso que busca aumentar el poder de discriminación de los modelos por medio de la reestimación de los parámetros de éstos.



## Modelos ocultos de Markov (XXXI)



- ❑ El procedimiento se centra en las porciones de los HMM, que más importancia tienen para la discriminación entre unidades similares que deben ser reconocidas. Consiste en un paso adicional comparado con el ML, puesto que el procedimiento de entrenamiento no necesita ser reformulado.
- ❑ Este modelo también puede aplicarse a sistemas que empleen redes de neuronas [LEE95].





## Modelos ocultos de Markov (XXXII)



- El modelado de los modelos ocultos de Markov de primer orden, asume que la probabilidad de transición entre estados en el momento  $t+1$  depende solamente del estado de la cadena de Markov en el tiempo  $t$ .
- En los modelos de segundo orden *HMM2* [MAR94], [MAR97] la probabilidad de transición de estado en el instante  $t+1$  depende de los estados de la cadena en los momentos  $t$  y  $t-1$ .



## Modelos ocultos de Markov (XXXIII)



- Los experimentos llevados a cabo en tareas de reconocimiento de dígitos conectados muestran una mejora en las prestaciones en el caso de los *HMM2*.
- Sin embargo si al caso *HMM* (*HMM1*) se le añade una etapa de postproceso que tenga en cuenta la duración de los segmentos [GON94], los resultados son similares.





## Modelos ocultos de Markov (XXXIV)



- **Los modelos ocultos de Markov, tanto de primer como de segundo orden, sufren de una limitación conocida como la asunción de la independencia de las observaciones.**
  - Con ello se asume que la probabilidad de un vector de características acústicas, asociado a un estado particular, no depende de los otros vectores de ese mismo estado.
  - En la mayoría de los casos esta asunción de independencia no es válida.



## Modelos ocultos de Markov (XXXV)



- **El uso de propiedades dinámicas a corto plazo en los vectores de rasgos ayuda a solventar este problema.**
  - La dependencia estadística entre el vector de observación actual y el último observado puede modelarse explícitamente por medio estimación gaussiana [WEL87].
  - Otra posibilidad es incluir un método que condicione la probabilidad del vector de observación actual en el estado actual con los vectores de observación anteriores [PAL93].





## Modelos ocultos de Markov (XXXVI)



- Las restricciones secuenciales también se pueden capturar por medio de una red neuronal recurrente [ROB94].
- **Una traza de voz se representa como un punto que se mueve en el espacio de parámetros según se producen cambios articulatorios.**
  - Esto constituye la trayectoria del habla.



## Modelos ocultos de Markov (XXXVII)



- Teniendo en cuenta que un punto puede pertenecer a diferentes trayectorias, los modelos para reconocimiento de voz deben basarse en las trayectorias y no en las posiciones geométricas individuales del espacio de parámetros.
  - La independencia de observaciones presente en los HMM no preserva la información de trayectorias.
  - Diferentes trayectorias comparten la misma función de distribución de probabilidad. Las clases con las trayectorias no pueden representarse bien, porque la información sobre la continuidad de cada trayectoria individual se pierde.
  - Las trayectorias se agrupan y esto conlleva a una menor discriminación de contextos fonéticos complejos.





## Modelos ocultos de Markov (XXXVIII)



- Una forma de tener en cuenta este aspecto es modelar la función de densidad de probabilidad de un gran vector, compuesto por la concatenación de las tramas de la secuencia de los vectores de observación, a través de una distribución gaussiana multivariable.
  - En otras ocasiones, las mezclas se introducen en el nivel de segmentos.
  - De esta forma, una distribución gaussiana representa una clase simple de trayectorias, mientras que un modelo de mezcla puede representar múltiples clases de agrupación de trayectorias.



## Modelos ocultos de Markov (XXXIX)



- Una nueva aproximación consiste en considerar a las trayectorias como observaciones de una variable aleatoria y emplear un modelo de trayectorias con mezcla estocástica o *STM (Stochastic Mixture Trajectory)* [GON94], [OST96].
  - De esta forma, las unidades del habla basadas en fonemas, se modelan como agrupaciones de trayectorias en su espacio de parámetros.
  - Las trayectorias se modelan por medio de mezclas de secuencias de estados de las funciones de densidad Gaussianas multivariables y se optimizan a nivel de secuencias de estados.
  - La duración de las trayectorias se estima por medio de distribuciones  $\Gamma$ .





## Bibliografía (I)



- [ANA95] A. Anastasakos, R. Schwartz and H. Shu, "Duration Modeling in Large Vocabulary Speech Recognition", *Proc. of ICASSP'95*, Detroit, Estados Unidos, 9-12 mayo 1995, Detroit, Estados Unidos, 9-12 mayo 1995, pp. 628-631.
- [BAH75] L. R. Bahl, "Decoding for Channels with Insertions, Deletions, and Substitutions with Applications to Speech Recognition", *IEEE Transactions on Information Theory*, Vol. IT-21, N° 4, julio 1975, pp. 404- 411.
- [BAH83] L. R. Bahl, F. Jelinek and R. L. Mercer, "A Maximum Likelihood Approach to Continuous Speech Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, N° 2, marzo 1983, pp. 179-190.
- [BAH86] L. R. Bahl et al., "Maximum Mutual Information Estimation of Hidden Markov Model Parameters for Speech Recognition", *Proc. of ICASSP'86*, pp. 49-52.
- [BAH88] L. R. Bahl et al., "Acoustic Markov Models Used in the Tangora Speech Recognition System", *Proc. of ICASSP'88*, Nueva York, Estados Unidos, 11-14 abril 1988, pp. 497-500.
- [BAU70] L. E. Baum et al., "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains", *The Annals of Mathematical Statistics*, Vol. 41, N° 1, 1970, pp. 164-171.
- [BUR96] D. Burshtein, "Robust Parametric Modeling of Durations in Hidden Markov Models", *IEEE Transactions on Speech and Audio Processing*, Vol. 4, N° 3, mayo 1996, pp. 240-242.



## Bibliografía (II)



- [EPH87] Y. Ephraim, J. G. Wilpon and L. R. Rabiner, "A Linear Predictive Front-End Processor for Speech Recognition in Noisy Environments", *Proc. of ICASSP'87*, Dallas, Estados Unidos, 6-9 abril 1987, pp.1324-1327.
- [EPH89] Y. Ephraim, A. Dembo and L. R. Rabiner, "A Minimum Discrimination Information Approach for Hidden Markov Modeling", *IEEE Transactions on Information Theory*, Vol. 35, N° 5, septiembre 1989, pp. 1001-1013.
- [GON94] Y. Gong and J. P. Haton, "Stochastic Trajectory Modeling for Speech Recognition", *Proc. of ICASSP'94*, Adelaida, Australia, 19-22 abril 1994, Vol. I, pp. 57-60.
- [GOP89] P. S. Gopalakrisnan et al., "A Generalization of the Baum Algorithm to Rational Objective Functions", *Proc. of ICASSP'89*, Glasgow, Reino Unido, 23-26 mayo 1989, pp. 631-634.
- [HUA90] X. D. Huang, Y. Ariki and M. A. Jack, *Hidden Markov Models for Speech Recognition*, Edinburgh University Press, 1990.
- [JEL80] F. Jelinek and R. L. Mercer, "Interpolated Estimation of Markov Source Parameters from Sparse Data", *Pattern Recognition in Practice*, E. S. Gelsema and L. N. Kanal editores, North-Holland Publishing Company 1980.
- [JUA92] B. H. Juang and S. Kataragi, "Discriminative Learning for Minimum Error Classification", *IEEE Transactions on Signal Processing*, Vol. 40, N° 12, diciembre 1992, pp. 3043-3054.





## Bibliografía (III)



- [LEE95] T. Lee, P. C. Ching and L. W. Chan, "An RNN Based Speech Recognition System with Discriminative Training", *Proc. of EUROSPEECH'95, Madrid, septiembre 1995*, pp. 1667-1670.
- [LIP82] L. A. Liporace, "Maximum Likelihood Estimation for Multivariate Observations of Markov Sources", *IEEE Transactions on Information Theory*, Vol. IT-28, N° 5, septiembre 1982, pp. 729-734.
- [MAR94] J. F. Mari and J. P. Haton, "Automatic Word Recognition based on Second-Order Hidden Markov Models", *ICSLP'94, Yokohama, Japón, septiembre 1994*, pp. 247-250.
- [MAR97] J. F. Mari and J. P. Haton, "Automatic Word Recognition Based on Second-Order Hidden Markov Models", *IEEE Transactions on Speech and Audio Processing*, Vol. 5, N° 1, enero 1997, pp. 22-25.
- [NOR91] Y. Normandin and S. D. Morgera, "An Improved MMIE Training Algorithm for Speaker-Independent, Small Vocabulary, Continuous Speech Recognition", *Proc. of ICASSP'91, Toronto, Canadá, 14-17 mayo 1991*, pp. 537-540.
- [OST96] M. Ostendorf, V. V. Digalakis and O. A. Kimball, "From HMM's to Segments Models: A Unified View of Stochastic Modeling for Speech Recognition", *IEEE Transactions on Speech and Audio Processing*, Vol. 4, N° 5, septiembre 1996, pp. 360-378.
- [PAL93] K. K. Paliwal, "Use of Temporal Correlation Between Successive Frames in a Hidden Markov Model Based Speech Recognizer", *Proc. of ICASSP'93, Minneapolis, Estados Unidos, 27-30 abril 1993*, Vol. II, pp.215-218.



## Bibliografía (IV)



- [PIC90] J. Picone, "Continuous Speech Recognition Using Markov Models", *IEEE ASSP Magazine*, Vol. 7, N° 3, julio 1990, pp. 26-41.
- [POR88] A. B. Poritz, "Hidden Markov Models: A Guided Tour", *Proc. of ICASSP'88, Nueva York, Estados Unidos, 11-14 abril 1988*, pp. 7-13.
- [RAB81] L. R. Rabiner and S. E. Levinson, "Isolated and Connected Word Recognition- Theory and Selected Applications", *IEEE Transactions on Communications*, Vol. COM-29, N° 5, mayo 1981, pp. 621-659.
- [RAB86] L. R. Rabiner and B. H. Juang, "An Introduction to Hidden Markov Models", *IEEE ASSP Magazine*, Vol. 3, N° 1, enero 1996, pp. 4-16
- [RAB89] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proc. of the IEEE*, Vol. 77, N° 2, febrero 1989, pp. 257-286
- [RAB93] L. R. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, N. J., 1993.
- [RAH97] M. Rahim, Y. Bengio and Y. LeCun, "Discriminative Features and Model Design for Automatic Speech Recognition", *Proc. of EUROSPEECH'97, Rodas, Grecia, septiembre 1997*, pp. 75-78.





## Bibliografía (V)



- [REI95] W. Reich and G. Ruske, "Discriminative Training for continuous speech recognition", *Proc. of EUROSPEECH'95*, Madrid, septiembre 1995, pp. 537-540.
- [RIG94] G. Rigoll, "Maximum Mutual Information Neural Networks for Hybrid Connectionist-HMM Speech Recognition Systems", *IEEE Transactions on Speech and Audio Processing*, Vol. 2, N° 1, Parte II, enero 1994, pp. 175-184.
- [ROB94] A. Robinson, "An Application of Recurrent Nets to Phone Probability Estimation", *IEEE Transactions on Neural Networks*, Vol. 5, N° 2, marzo 1994, pp. 298-305.
- [SAK78] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", *IEEE Transactions on Acoustic, Speech, and Signal Processing*, Vol. ASSP-26, N° 1, febrero 1978, pp. 43-49.
- [SAK92] H. Sakoe, "Dynamic Programming-Based Speech Recognition Algorithms", *Advances in Speech Signal Processing*, S. Furui and M. M. Sondhi editores, Marcel Dekker Inc., 1992, pp. 487-507.
- [VIT67] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm", *IEEE Transactions on Information Theory*, Vol. IT-13, N° 2, abril 1967, pp. 260-309.
- [WEL87] C. J. Wellekens, "Explicit Correlation in Hidden Markov Models for Speech Recognition", *Proc. of ICASSP'87*, Dallas, Estados Unidos, 6-9 abril 1987, pp. 384-386.

