ALINEACIÓN TEMPORAL PARA EL RECONOCIMIENTO DE PALABRAS

Federico Miyara

1. Introducción

El objetivo del reconocimiento de la voz orientado a las palabras es comparar dos emisiones vocales aisladas a fin de determinar si corresponden o no a la misma palabra. Cuando una de las palabras pertenece a una base de datos o corpus previamente almacenado en el sistema, será posible utilizar esta técnica para identificar la palabra emitida.

Una primera idea sería intentar comparar, instante a instante, las formas de onda de ambas emisiones. Esta estrategia no resulta, sin embargo, eficaz, debido a que la conexión entre la forma de onda temporal y el sonido percibido es demasiado débil, si no virtualmente inexistente.

El oído está, más bien, orientado a la percepción de la información espectral del sonido que a su forma de onda. Podríamos intentar comparar, entonces, los espectros. Ahora bien, el espectro de una señal es una propiedad global, es decir, de *toda* la señal, y no representa de una manera significativa las variaciones de distribución de la energía en las diferentes frecuencias a lo largo del tiempo. Por este motivo es necesario primero subdividir la emisión en pequeños intervalos denominados *cuadros* (en inglés, *frames*) de entre 10 ms y 30 ms y analizar el espectro de cada cuadro separadamente (análisis de espectro a corto plazo). Si bien este enfoque tendría mejores posibilidades que la comparación directa de las formas de onda, en la práctica tampoco funciona, dado que algunas modificaciones simples del espectro, como por ejemplo un cambio en la frecuencia fundamental, no alteran la identidad de la palabra emitida pero afectan drásticamente la comparabilidad.

Un examen de los fenómenos acústicos y fonéticos involucrados en la emisión vocal permite concluir que lo verdaderamente distintivo de la identidad fonética de los sonidos emitidos es la forma en que se distribuyen las resonancias del tracto vocal, es decir, sus formantes, y si el tipo de excitación es tonal (cuasiperiódica) o no tonal (ruido de banda ancha).

Las resonancias pueden obtenerse a partir de la envolvente del espectro. Los *coeficientes de predicción lineal* (LPC) permiten una aceptable detección de estas resonancias. El *cepstro* resulta todavía más robusto. Podemos concebir, por lo tanto, la idea de reemplazar la información temporal de cada cuadro por la información cepstral, dada por medio de un vector constituido por los *coeficientes cepstrales* de menor orden (en general son suficientes los primeros 10 a 20 coeficientes). Este vector se denomina *vector cepstral*.

La comparación de los vectores cepstrales de dos cuadros (frames) simultáneos correspondientes a sonidos perceptivamente iguales (por ejemplo dos instancias del fonema /a/) puede realizarse a través de la *distancia cepstral*, es decir, la distancia euclídea entre los vectores cepstrales. Los resultados se corresponden bastante bien con la apreciación subjetiva, en el sentido de que dos sonidos similares exhiben una distancia cepstral pequeña, y en cambio dos sonidos perceptivamente diferentes están separados por una distancia cepstral considerable.

El enfoque anterior tropieza con una importante dificultad, consistente en que las duraciones de los fonemas en dos emisiones diferentes de una misma palabra no son iguales (figura 1). Inclusive la duración total de las emisiones puede diferir bastante, dependiendo de cuestiones como la personalidad del hablante, el estado de ánimo, la tranquilidad o el nerviosismo, etc. Ello implicará una considerable probabilidad de que si la comparación se efectúa entre cuadros *simultáneos* sus vectores cepstrales no coincidan por corresponder a fonemas diferentes. Esta situación podría darse a pesar de percibirse las palabras como idénticas.

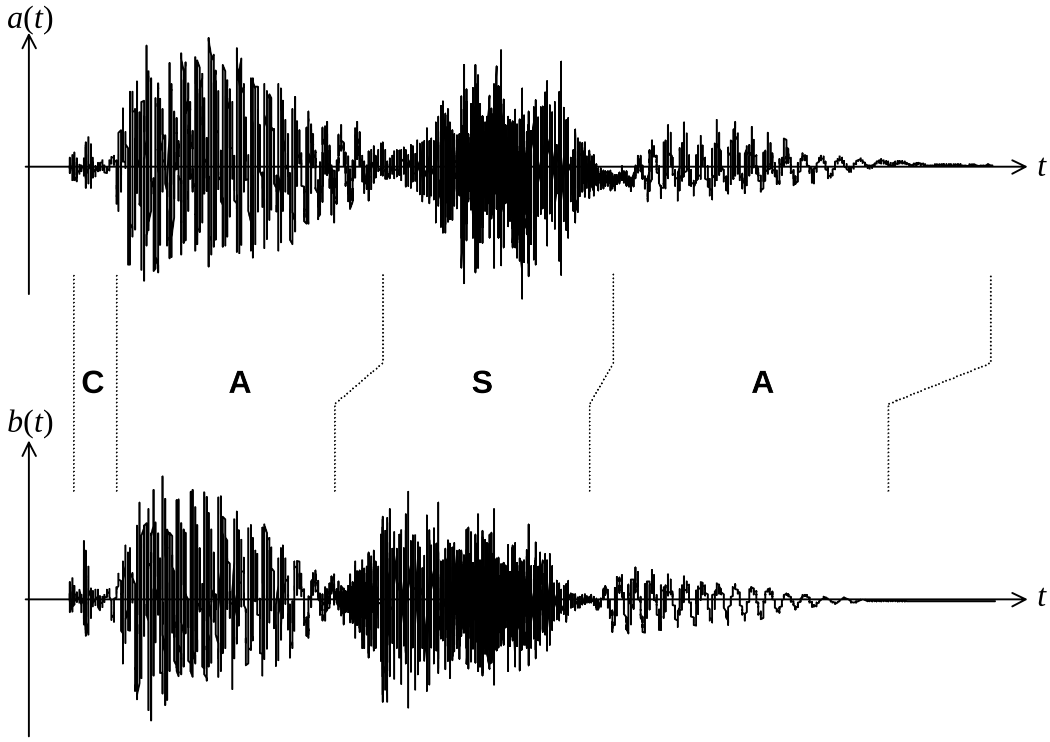


Figura 1. Dos diferentes emisiones de la palabra “casa”. La segmentación entre los diferentes fonemas se ha marcado en línea de trazos, poniendo en evidencia el problema de desalineación temporal.

La solución a este problema consiste en llevar a cabo un proceso de *alineación temporal* de los sucesivos cuadros (frames) de ambas emisiones, de tal manera de comparar cuadros (frames) *fonéticamente homólogos* (es decir, correspondientes a igual posición fonética en las dos palabras emitidas).

2. Distorsión temporal

La mencionada alineación de los cuadros de una y otra emisión puede efectuarse por medio de una *distorsión temporal*. Antes de introducir formalmente este concepto, tengamos en cuenta que las emisiones, que denominaremos $a(t)$ y $b(t)$, han sido subdivididas en n_a y n_b cuadros respectivamente, y cada cuadro ha sido reemplazado por un vector cepstral (dicho vector podría tener una estructura mixta; por ejemplo, las primeras 15 componentes podrían ser los primeros coeficientes cepstrales, la siguiente un valor binario 0 ó 1 según que la señal correspondiente sea de carácter tonal o no, y la última, el valor de la frecuencia fundamental estimada en el caso de sonidos tonales). Llamando A_i al vector correspondiente al i-ésimo cuadro de la señal $a(t)$ y B_j al vector correspondiente al j-ésimo cuadro de la señal $b(t)$, las señales totales quedan sustituidas por

$$A = \left[A_1, \ldots, A_i, \ldots, A_{n_a}\right], \tag{1}$$

$$B = \left[B_1, \ldots, B_j, \ldots, B_{n_b}\right]. \tag{2}$$

Una primera idea para la distorsión temporal sería aplicar una función ϕ al índice i de la señal A para obtener el índice j de la señal B, es decir:

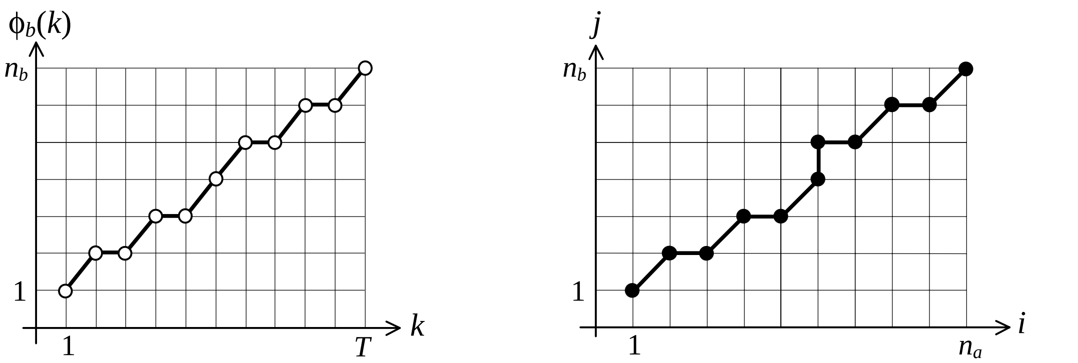
$$j = \phi(i).$$

Sin embargo, a veces puede ser conveniente que un mismo vector cepstral de A quede alineado con dos o más vectores de B (por ejemplo si un fonema de B es mucho más corto que el correspondiente de A). En tal caso ϕ dejaría de ser una función. Para evitar esto conviene adoptar un índice $k = 1, \ldots, T$ correspondiente a un *tiempo normalizado* y hacer que i y j sean funciones de k (figura 2):

$$i = \phi_a(k) \qquad k = 1, \ldots, T \tag{3}$$

$$j = \phi_b(k) \qquad k = 1, \ldots, T. \tag{4}$$

Esto permite lograr una alineación aún en casos como el indicado en la figura 1, en los que algunos fonemas de A son más largos que los homólogos de B (las dos /*a*/) en tanto para otros fonemas la relación se invierte (la /*s*/).



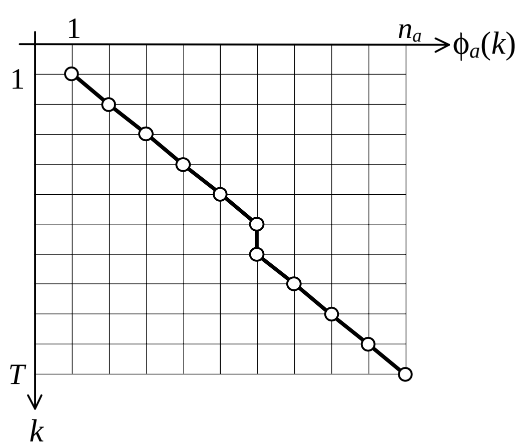


Figura 2. Composición de dos funciones de distorsión temporal, $\phi_a(k)$ y $\phi_b(k)$, dependientes de un único tiempo normalizado k.

Las ecuaciones (3) y (4) pueden interpretarse como las ecuaciones paramétricas de un camino discreto en el plano *i-j* como el indicado en la figura 2. Cada punto de dicho camino corresponde a la alineación del cuadro *i*-ésimo de la emisión *A* con el cuadro *j*-ésimo de la emisión *B*, como se indica en la figura 3.

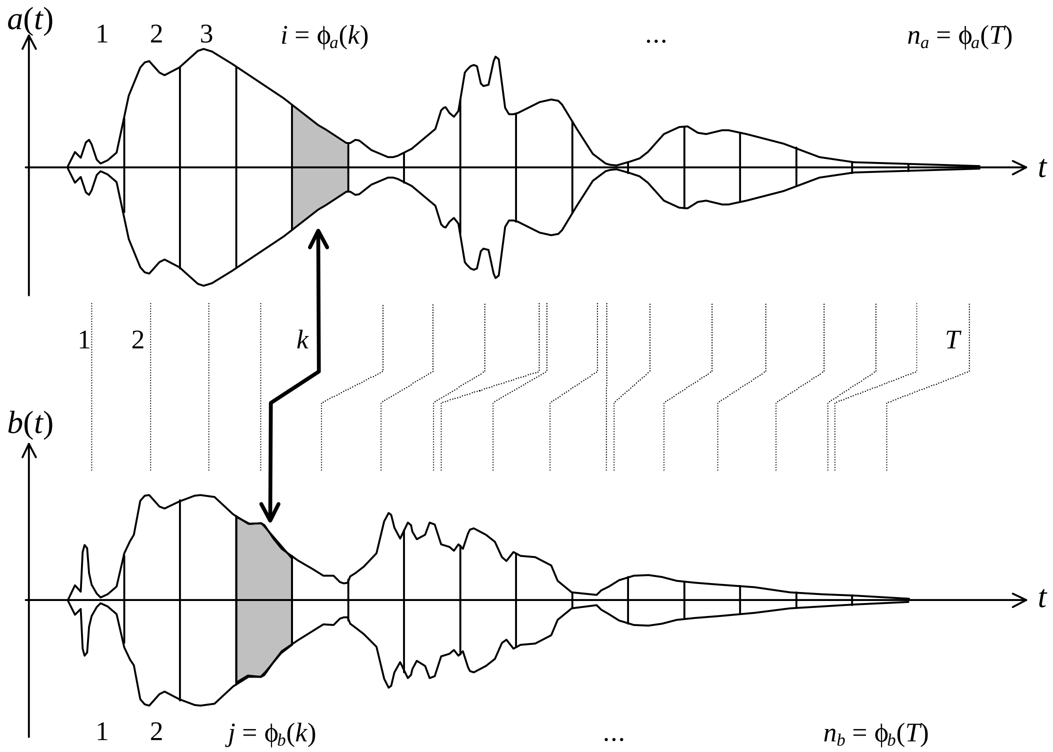


Figura 3. Las formas de onda correspondientes a la figura 1 se han subdividido en cuadros (frames) y alineado por medio de una función $\phi = (\phi_a, \phi_b)$. Se han sombreado los cuadros alineados correspondientes al tiempo normalizado *k*. *Nota: Por claridad las formas de onda se han representado mediante sus respectivas envolventes.*

Es conveniente en algunos casos aligerar la notación considerando una función $\phi(k)$ a valores en el plano *i-j* dada por

$$\phi(k) \quad = \quad \big(\phi_a(k), \phi_a(k)\big) \qquad k = 1, ..., T. \tag{5}$$

Recordemos que nuestro objetivo es comparar las dos emisiones desde el punto de vista del cepstro de sus cuadros (frames) fonéticamente homólogos, es decir aquéllos que corresponden a fonemas iguales. En este contexto, cada elección de ϕ traerá aparejada una decisión (acertada o no) sobre qué frames han de considerarse homólogos. Podemos definir a partir de dicha alineación, una *distancia* entre ambas emisiones igual a la suma de las distancias entre los vectores cepstrales que representan a cuadros homólogos. Es decir,

$$d_\phi(A,B) \quad = \quad \sum_{k=1}^{T} d\Big(A_{\phi_a(k)},\, B_{\phi_b(k)}\Big), \tag{6}$$

donde

$$d\left(A_{\phi_a(k)}, B_{\phi_b(k)}\right)^2 \;=\; \sum_{h=1}^{L}\left(c_{A_{\phi_a(k)},h} \;-\; c_{B_{\phi_b(k)},h}\right)^2 . \tag{7}$$

En esta expresión $c_{A_{\Phi_a(k)},h}$ es el h-ésimo coeficiente cepstral del cuadro ϕ_a(k) de la señal A, y análogamente para B. Observemos que la distancia definida en (6) depende de la elección que se haya efectuado de la función de distorsión temporal ϕ.

En la práctica puede ser conveniente aplicar una ponderación $m(k)$ a los términos de la ecuación (6), obteniéndose

$$d_\phi(A,B) \;=\; \sum_{k=1}^{T} d\left(A_{\phi_a(k)}, B_{\phi_b(k)}\right) m(k) . \tag{8}$$

Esto permite asignar un mayor peso a determinados cuadros según su importancia relativa para la comparación de las dos emisiones.

3. Selección de la función de distorsión temporal

Hasta ahora hemos introducido conceptualmente las funciones de distorsión temporal $\phi = (\phi_a, \phi_b)$ y la distancia entre A y B correspondiente a cada una de ellas. Es claro que potencialmente existe un número muy grande de tales funciones (igual a la cantidad de posibles caminos entre (1,1) y (n_a, n_b)),[1] por lo cual será necesario dar algún criterio para seleccionar la que permite la "mejor" alineación temporal. Intuitivamente, la mejor alineación sería aquélla que permitiera comparar entre sí cuadros fonéticamente parecidos, es decir, aquéllos cuya distancia cepstral sea inherentemente pequeña. Dado que esta condición debe cumplirse para cada par de cuadros de A y B a comparar, equivale a requerir que la distancia total $d_\phi(A, B)$ sea mínima. En otras palabras, la distorsión temporal *óptima* será aquélla que hace mínima la distancia entre A y B:

$$\phi_{opt} \;=\; \arg\left(\min_{\phi}\left\{d_\phi(A,B)\right\}\right) , \tag{9}$$

Definimos, entonces, la *distancia* entre A y B como dicha distancia mínima:

$$d(A,B) \;=\; \min_{\phi}\left\{d_\phi(A,B)\right\} \;=\; d_{\phi_{opt}}(A,B). \tag{10}$$

La determinación de dicha distancia mínima requeriría, en principio, la inspección de todos los posibles caminos y el cálculo de la distancia correspondiente a cada uno de ellos. Esta técnica sería computacionalmente ineficiente, y en lugar de ello se utilizan algoritmos de *programación dinámica*, que analizaremos más adelante.

[1] Sin imponer ninguna restricción, la cantidad exacta es $(n_a \cdot n_b)^T$. La cantidad se reduce notablemente imponiendo algunas condiciones adicionales, por ejemplo la monotonía de ϕ_a y ϕ_b.

3.1. Restricciones sobre las funciones de distorsión temporal

Hemos visto que la cantidad total de posibles funciones de distorsión es muy grande. Sin embargo, como veremos, no todas tienen sentido en el contexto del reconocimiento de la palabra, lo cual hace necesario reducir el conjunto aplicando algunas restricciones. Son ellas las condiciones de contorno, de monotonía, de continuidad local y globales.

3.1.1. Condiciones de contorno

Las condiciones de contorno establecen que el primer cuadro de A debe compararse con el primer cuadro de B y, análogamente, el último cuadro de A debe compararse con el último de B, es decir:

$$\phi_a(1) = \phi_b(1) = 1 \tag{11}$$

$$\phi_a(T) = n_a, \qquad \phi_b(T) = n_b. \tag{12}$$

Esto garantiza que se compare la totalidad de la emisión A con la totalidad de la emisión B, evitando que la comparación de las emisiones de dos palabras tales como "fra*casa*" y "*casa*ca" arroje resultados positivos debido a que contienen dos fragmentos similares.

3.1.2. Condiciones de monotonía

El orden de sucesión de los fonemas es esencial para la identidad de una palabra. Así, no es lo mismo "casa" que "saca", a pesar de que las dos poseen los mismos fonemas. Si sólo estuviéramos interesados en la alineación que minimiza la distancia (en el sentido de la ecuación (8)), es evidente que se lograría un mejor ajuste alineando los cuadros iniciales de "saca" con los cuadros centrales de "casa" (ya que ambos grupos de cuadros contienen el fonema /s/). El resultado de la comparación sería positivo, cuando en realidad son palabras diferentes.

Las condiciones de monotonía para las funciones ϕ_a y ϕ_b impiden que esto suceda. Dichas condiciones pueden expresarse como:

$$\phi_a(k+1) \geq \phi_a(k) \qquad k = 1, ..., T-1 \tag{13}$$

$$\phi_b(k+1) \geq \phi_b(k) \qquad k = 1, ..., T-1. \tag{14}$$

El significado de la monotonía puede apreciarse en la figura 2. La monotonía implica que no pueden reexaminarse cuadros que ya han sido superados.

3.1.3. Condiciones de continuidad local

En la emisión normal de la palabra hablada no cualquier transición es posible. Ello se debe a que la respuesta transitoria de los órganos articulatorios (la glotis, el velo, la lengua y los labios) es relativamente lenta. La transición entre el fonema nasal /n/ y el fonema abierto /a/, que se muestra en la figura 4, es un ejemplo ilustrativo. Dicha transición dura más de 40 ms, que puede representar hasta 6 cuadros (dependiendo de la duración y el solapamiento elegidos para los cuadros). Si se permitiera cualquier alineación que cumpla con las condiciones de contorno y de monotonía, podríamos estar aceptando saltos que pasen por alto estas transiciones.

Otra situación se da en el caso de los fonemas que, a la inversa del caso anterior, son de corta duración, como en el caso de las consonantes oclusivas o explosivas. Un ejemplo aparece en el fonema inicial de la palabra “casa”, mostrado en las figuras 1 y 3. Si se permitiera saltear varios cuadros, podría correrse el riesgo de omitir información importante sobre estos fonemas, que a veces resultan esenciales para discriminar una palabra.

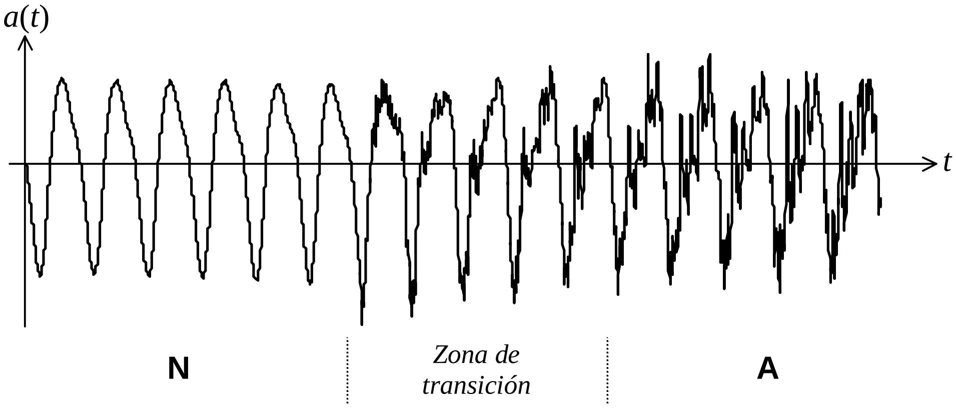


Figura 4. Las condiciones de continuidad local se originan en las transiciones graduales entre fonemas, como la ilustrada en la figura. Allí se muestra la forma de onda del segmento central de la sílaba “na”. La zona de transición corresponde al momento en que la lengua se separa del alvéolo permitiendo la salida de aire por los labios.

El tipo de dificultades señaladas puede resolverse imponiendo las denominadas *condiciones de continuidad local*. Estas condiciones se presentan en la forma de pequeños repertorios de transiciones elementales aceptables junto con la prescripción de que todos los caminos que unen (1, 1) con (n_a, n_b) deberán estar formados por una sucesión de tales caminos elementales. En la figura 5 se muestran dos juegos simples de caminos elementales.

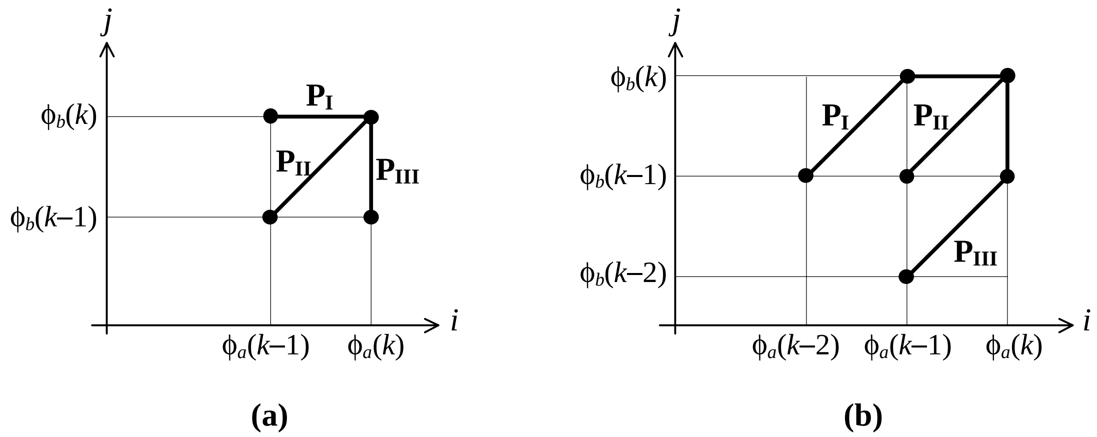


Figura 5. Dos repertorios o juegos de caminos elementales que permiten garantizar condiciones de continuidad local. Elegido un repertorio (por ejemplo el **(a)**), cualquier camino que conecte (1, 1) con (n_a, n_b) deberá estar formado por tramos pertenecientes al repertorio.

En la figura 6 se muestra la descomposición de un camino que une el punto (1, 1) con el (n_a, n_b) utilizando los caminos o transiciones elementales del tipo (b) de la figura 5. Recordemos que cada punto (i, j) de un camino implica la alineación del i-ésimo cuadro de la emisión A con el j-ésimo cuadro de la emisión B. Llamando **P** al camino completo, podemos representar dicho camíno en la forma

$$\mathbf{P} \;=\; \mathbf{P_I}\,\mathbf{P_I}\,\mathbf{P_{II}}\,\mathbf{P_{II}}\,\mathbf{P_I}\,\mathbf{P_{II}} \tag{15}$$

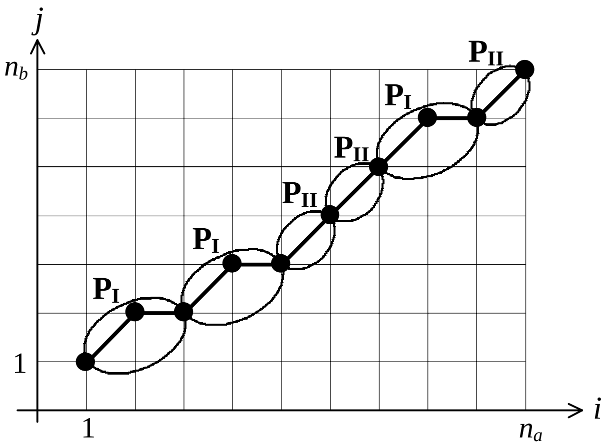


Figura 6. Descomposición de un camino que une los puntos extremos (1, 1) y (n_a, n_b) en una sucesión de caminos elementales pertenecientes al repertorio (b) de la figura 5.

3.1.4. Condiciones globales

Por último, existen condiciones globales que afectan a la totalidad de un camino. La primera aparece como mera consecuencia de las condiciones de continuidad local. Supongamos que adoptamos el repertorio (b). A diferencia del repertorio (a), que admite transiciones elementales horizontales y verticales, el repertorio (b) no las admite.[2] Ello implica la existencia de cotas máxima y mínima para la "pendiente" media de cada transición elemental. Llamando Q_r a la pendiente del camino elemental $\mathbf{P_r}$, tenemos:

$$\begin{aligned} Q_\mathrm{I} &= 1/2 \\ Q_\mathrm{II} &= 1 \\ Q_\mathrm{III} &= 2 \end{aligned} \tag{16}$$

es decir

$$\begin{aligned} Q_\text{mín} &= 1/2 \\ Q_\text{máx} &= 2. \end{aligned} \tag{17}$$

Estos valores definen unas "cuñas" de salida del punto inicial y de llegada al punto final cuya intersección constituye la región admisible para el paso del camino.[3] Esto se ejemplifica en la figura 7.

[2] Si bien los caminos elementales contienen transiciones verticales, ellas son inseparables de las transiciones diagonales.

[3] Que el camino seleccionado esté incluido en esta región es condición necesaria pero en general no suficiente, particularmente si el número de puntos del camino es grande.

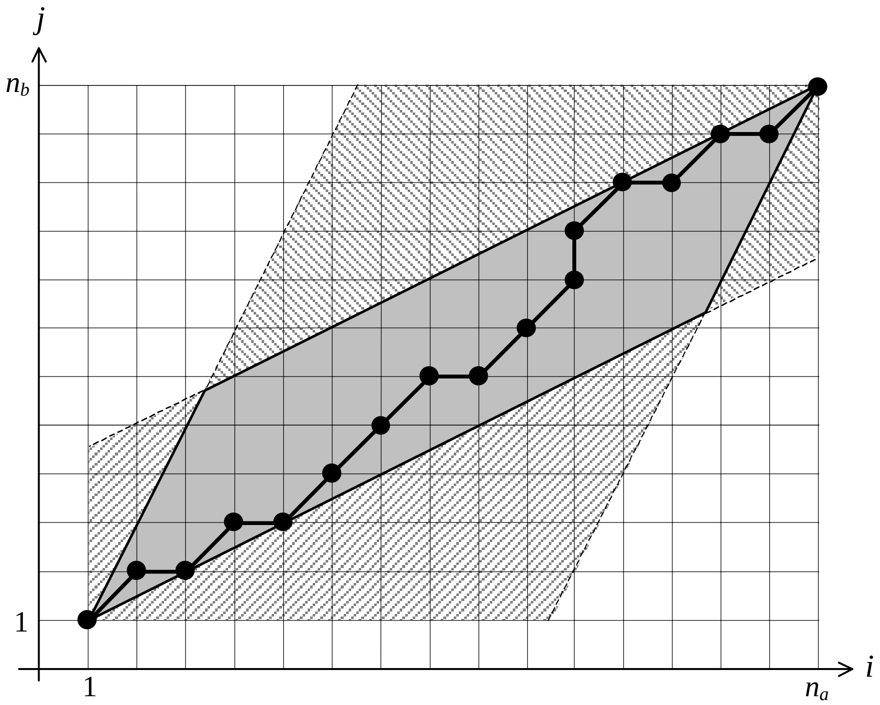


Figura 7. Condición global derivada de las cotas sobre la pendiente de los caminos elementales. La cuña con vértice en (1, 1) representa todos los puntos alcanzables por medio de una sucesión de caminos elementales válidos partiendo de dicho punto inicial. La cuña con vértice en (n_a, n_b) representa todos los puntos desde los que puede alcanzarse dicho punto final por medio de una sucesión de caminos elementales válidos. La intersección de ambas cuñas proporciona la región cuadrangular donde deben yacer todos los caminos válidos.

Una segunda condición global, que puede coexistir con la anterior produciendo una región aceptable todavía más estrecha, consiste en imponer una cota máxima a la distorsión temporal relativa entre ambas emisiones. Esa restricción se enuncia en la forma

$$\left|\phi_a(k) - \phi_b(k)\right| \leq N, \tag{18}$$

o bien

$$\phi_a(k) - N \leq \phi_b(k) \leq \phi_a(k) + N. \tag{19}$$

Así planteada, la condición implica que los caminos aceptables se encuentran en una banda de altura $\pm N$ alrededor de la identidad ($j = i$). Intersectando esa banda con el paralelogramo resultante de las limitaciones de pendiente, se obtiene la figura 8.

Esta restricción es aplicable especialmente cuando no hay diferencias demasiado notables en las duraciones de las emisiones a comparar. Si existen diferencias marcadas, la distorsión acumulada hacia el final de las emisiones puede llevar a que no exista ningún camino aceptable, por lo cual no podrá calcularse la distancia entre las emisiones.

Una excepción interesante es cuando antes de la distorsión no lineal del tiempo se aplica una distorsión lineal con el propósito de igualar la duración de ambas emisiones (medida en cantidad de cuadros). En este caso la cota global indicaría el máximo desplazamiento relativo entre fonemas homólogos bajo la hipótesis de igual duración. Esta diferencia, presente en los casos de *acentos agógicos* (énfasis logrado aumentando la duración de un determinado fonema), pertenece generalmente al *campo expresivo*, más que al *léxico* o *semántico*.

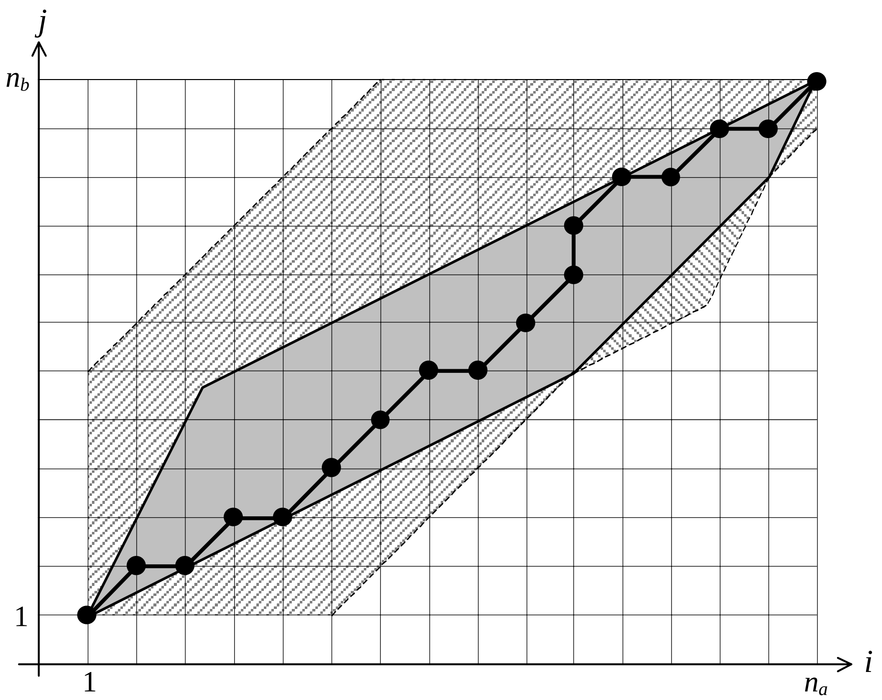


Figura 8. Condición global de acotación de la máxima distorsión temporal. La intersección con el paralelogramo de la figura 7 da la zona admisible sombreada.

3.2. Optimización de los caminos mediante programación dinámica

La ecuación (9) brinda un criterio para hallar la distorsión temporal óptima, es decir, la que ofrece la mejor alineación entre dos emisiones *A* y *B*. La aplicación de dicho criterio en forma directa, aun con la considerable reducción de caminos aceptables dada por las condiciones restrictivas de la sección anterior, es computacionalmente ineficiente. Debemos, por consiguiente, introducir un algoritmo que no requiera la exploración de todas las posibilidades para alcanzar el óptimo.

Un algoritmo más eficiente para resolver problemas de optimización de este tipo es la *programación dinámica*, que permite reducir la cantidad de casos explorados apoyándose en el denominado *principio de optimalidad*. Este principio establece que si un camino que une dos puntos X e Y pasando por una serie de puntos intermedios es óptimo según determinado criterio, entonces al subdividirlo en dos tramos XY e YZ, cada uno de ellos también será el camino óptimo entre sus respectivos extremos. La programación dinámica aplica este principio recursivamente aumentando en una unidad la cantidad de puntos en cada iteración.

Podemos plantear el criterio de optimización como el de minimizar una función "costo" adecuadamente elegida. En nuestro caso convendrá elegir el "costo" de un camino dado (o de la distorsión temporal correspondiente) como la distancia cepstral ponderada entre *A* y *B* definida por la ecuación (8), aquí reproducida:

$$d_{\phi}(A,B) \;=\; \sum_{k=1}^{T} d\left(A_{\phi_a(k)},\, B_{\phi_b(k)}\right) m(k)\,. \tag{20}$$

A efectos de utilizar técnicas recursivas, definiremos a su vez una distancia parcial que involucre sólo h puntos ($h < T$):

$$d_{\phi,h}(A,B) \;=\; \sum_{k=1}^{h} d\left(A_{\phi_a(k)},\, B_{\phi_b(k)}\right) m(k)\,. \tag{21}$$

El "costo" parcial para conectar el punto (1, 1) con el (i, j) será la mínima distancia parcial posible, abarcando todos los caminos posibles, es decir,

$$\psi_{(i,\,j)}(A,B) \;=\; \min_{\phi,h}\left\{d_{\phi,h}(A,B) \;/\; \left(\phi_a(h), \phi_b(h)\right) = (i, j)\right\}. \tag{22}$$

Para aplicar este algoritmo es necesario considerar los posibles caminos descompuestos en sucesiones de caminos elementales del repertorio elegido. Supongamos un camino cualquiera correspondiente a una distorsión temporal ϕ = (ϕ_a, ϕ_b), y sea k un valor del tiempo normalizado tal que

$$(i, j) \;=\; (\phi_a(k),\, \phi_b(k)). \tag{23}$$

Supongamos además que el camino elemental que conduce a este punto[4] comienza en (l, m). Si la cantidad de tramos que lo forman es Q, entonces

$$(l, m) \;=\; (\phi_a(k-Q),\, \phi_b(k-Q)). \tag{24}$$

Podemos asignar un "costo" $\zeta((l, m), (i, j))$ al camino elemental que une los puntos (l, m) e (i, j), dado por

$$\zeta\left((l,m),(i,j)\right) \;=\; \sum_{q=0}^{Q-1} d\left(A_{\phi_a(k-q)},\, B_{\phi_b(k-q)}\right) m(k-q)\,. \tag{25}$$

Por ejemplo, si se tratara del camino $\mathbf{P_I}$ del repertorio (b) de la figura 5, entonces

$$(l, m) \;=\; (\phi_a(k-2),\, \phi_b(k-2)). \tag{26}$$

Este costo corresponde a la suma de las distancias (ponderadas) entre los vectores cepstrales alineados según lo indican los puntos del camino elemental excepto los correspondientes al primer punto. En el ejemplo, tendremos:

$$\zeta\left((l,m),(i,j)\right) \;=\; d\left(A_{\phi_a(k)},\, B_{\phi_b(k)}\right) m(k) + d\left(A_{\phi_a(k-1)},\, B_{\phi_b(k-1)}\right) m(k-1)\,.$$

[4] Obsérvese que si el repertorio ha sido adecuadamente elegido, existe a lo sumo un camino elemental que une dos puntos dados. (Si los puntos están muy alejados o fuera del área admisible, no habrá ninguno.)

El algoritmo indica que

$$\psi_{(i,\,j)}(A,B) \;=\; \min_{(l,m)}\left\{\psi_{(l,m)}(A,B)+\zeta\big((l,m),(i,j)\big)\right\}. \tag{27}$$

Este algoritmo se aplicará iterativamente hasta que $(i, j) = (n_a,n_b)$.

3.2.1. Eficiencia computacional de la programación dinámica

A efectos de determinar cuán eficiente es la programación dinámica para resolver problemas de optimización de caminos, consideremos un total de N puntos y supongamos que queremos obtener el camino óptimo para pasar del punto n al punto m en T tramos. Supondremos, por simplicidad, que la única restricción será que no se puede pasar de un punto a sí mismo. El proceso se ilustra en el ejemplo de la figura 9, en la que se muestran $N = 4$ puntos, siendo la cantidad de tramos $T = 4$. Si bien no se indica el "costo" de cada camino, se supone que los caminos indicados con trazo grueso son los que han resultado óptimos entre sus respectivos puntos extremos cuando se consideran la cantidad de tramos correspondiente a esa iteración.

La cantidad de caminos que es necesario explorar a fin de evaluar sus costos es, utilizando programación dinámica,

$$C_{PD}(n, m) \;=\; (N-1) \;+\; (N-1)^2 \;+\; (T-3)\,N(N-1) \;+\; (N-1).$$

En efecto, a partir de la tercera iteración, la cantidad de caminos a explorar se estabiliza en N (N –1) excepto en la última iteración, ya que el punto de destino (el m) es uno solo.

La ecuación anterior se puede simplificar, resultando

$$C_{PD} \;=\; ((T-2)\,N \;+\; 1)\,(N-1), \tag{28}$$

que es del orden de $T\,N^2$. Si se hubieran explorado todas las posibilidades, la cantidad de caminos investigados habría sido

$$C \;=\; (N-1)^T\,, \tag{29}$$

que crece mucho más rápidamente. Por ejemplo, si $N = 10$ puntos y se utilizan $T = 7$ tramos, la exploración exhaustiva implica investigar $9^7 = 4782969$ caminos, en tanto que la programación dinámica permite realizar sólo ((7 – 2)10 + 1) 9 = 459 pruebas. Las diferencias se hacen más notables a medida que la cantidad de puntos y la cantidad de tramos crece.

Por otra parte, el análisis realizado corresponde al caso en que la única restricción es no poder pasar de un punto a sí mismo, por lo que desde cada punto se puede pasar a los restantes $N - 1$. Si agregamos las restricciones comentadas en 3.1 la carga computacional baja enormemente, ya que en cada tramo sólo pueden ser alcanzados unos pocos puntos, y no los $N - 1$ aquí supuestos.

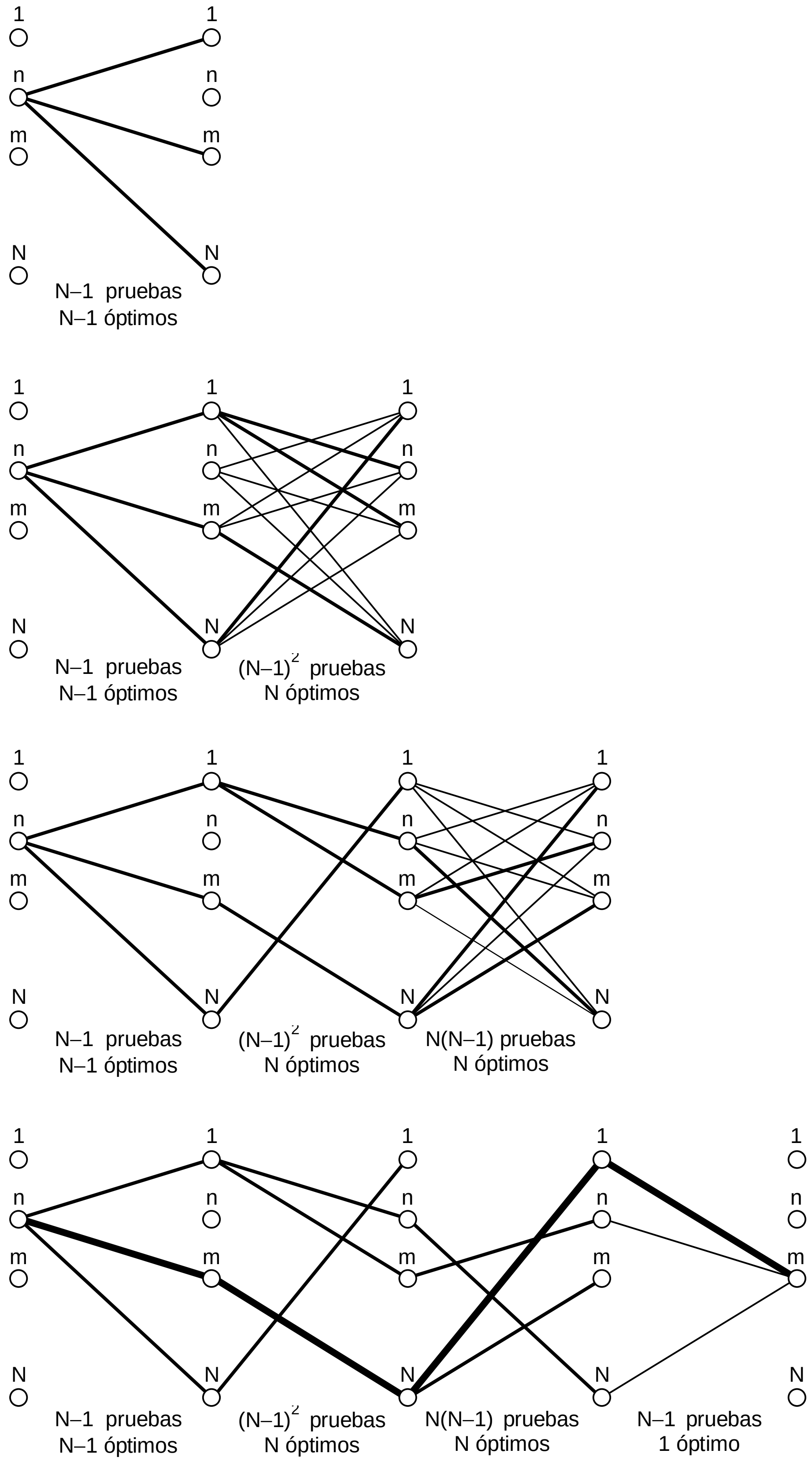


Figura . Proceso iterativo para obtención del camino óptimo entre el punto *n* y el *m* de un conjunto de *N* puntos. Las líneas gruesas son las correspondientes al mínimo costo entre los respectivos puntos extremos y la cantidad de tramos hasta esa iteración.

REFERENCIAS

Mary Jo Creaney-Stockton "Isolated Word Recognition Using Reduced Connectivity Neural Networks With Non-Linear Time Alignment Methods". PhD dissertation - Department of Electrical and Electronic Engineering -University of Newcastle-Upon-Tyne. August, 1996. Chapter 3. Disponible en http://www.moonstar.com/~morticia/thesis/chapter3.html

Rabiner, Lawrence; Juang, Biing-Hwang. "Fundamentals of Speech Recognition". Prentice Hall International.. Englewood Cliffs, New Jersey, USA, 1993.