5. Separación Ciega de Fuentes basada en la estructura temporal de las señales de voz

5.1 Introducción

La *Separación Ciega de Fuentes (BSS)*, dentro del marco del *Análisis de Componentes Independientes (ICA)* es una herramienta de gran utilidad en el campo de la ingeniería. Existen multitud de aplicaciones tales como la eliminación de ruido aditivo en señales o imágenes, eliminación de efectos de diafonía en las telecomunicaciones, entre otras.

Como se ha estudiado en secciones previas la Separación Ciega de Fuentes se basa en el empleo de estadísticos de orden superior, sin tener en cuenta la estructura temporal de las señales originales. Estos algoritmos necesitan de una gran carga computacional ya que en la mayoría de los casos emplean optimización no lineal.

Seguidamente desarrollaremos una aplicación basada en [Ikeda] e [Ikeda2] que permitirá obtener un método para implementar la Separación Ciega de Fuentes en el caso de señales estructuradas temporalmente, como es el caso de las señales de voz.

Un problema fundamental que nos encontramos en la separación de señales de voz son los *retardos* existentes y las reflexiones que se producen, dado que nos encontramos en un entorno real. Esto implicará que las mezclas ya no sean instantáneas, sino convolutivas. Para separar las señales haremos habrá que trasladar las señales al dominio tiempo-frecuencia y una vez allí aplicar un algoritmo de decorrelación a cada componente frecuencia. Una dificultad añadida que veremos en el desarrollo del problema será el de encontrar el orden exacto de las componentes frecuenciales. Para solventar esta situación es donde precisamente hacemos uso de la estructura temporal de las señales de voz.

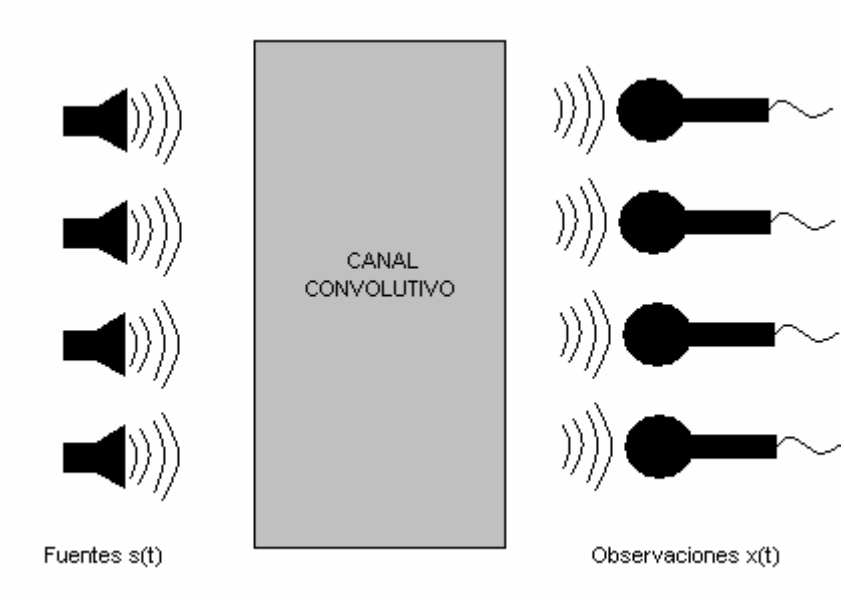

Figura 5.1 *Esquema físico del proceso de generación de las observaciones.*

5.2 Estudio teórico del problema

5.2.1 Consideraciones iniciales sobre las señales de voz

En primer lugar vamos a establecer una serie de hipótesis que serán de gran utilidad de cara a desarrollar posteriormente el algoritmo que implementa la separación de las señales de voz. Dichas fuentes originales se caracterizan por las siguientes características:

1. Las señales se suponen estacionarias en un pequeño intervalo de tiempo. Se suele considerar que la voz humana es estacionaria en un periodo no superior a los quince milisegundos.
2. Las señales son intrínsecamente no estacionarias debido a la amplitud de modulación.

El objetivo del método es el de representar las señales en el dominio tiempo-frecuencia, también llamado *espectrograma* y proceder posteriormente a su separación para obtener las señales estimadas.

Las observaciones se corresponden con mezclas de tipo convolutivas (e independientes del tiempo) de las señales originales, esto es:

$$\mathbf{x}(t) = \mathbf{A} * \mathbf{s}(t) \tag{5.1}$$

donde cada elemento de **A**(t) es una función de transferencia desconocida.

5.2.2 La transformada localizada de Fourier

La propiedad anteriormente estudiada referente a la estacionalidad de las señales de voz nos permite el uso de la conocida como *transformada localizada de Fourier* en un período de tiempo lo suficientemente corto como para no incumplir las restricciones, lo que condicionará el tamaño de la ventana a elegir. La segunda de las propiedades enunciadas en el apartado anterior nos permite discernir una forma de combinar las componentes frecuenciales que se extraigan en el algoritmo de separación.

La Separación Ciega de Fuentes no tiene en cuenta las posibles permutaciones que se puedan dar en las señales separadas sin embargo, por el hecho de aplicarle dicha técnica a cada frecuencia por separado, a fin de cuentas tendremos que saber en que orden han de colocarse. Supondremos que los espectros en frecuencia serán continuos, por lo que en la práctica podremos combinar cada una de las frecuencias e ir construyendo los espectrogramas por separado.

La aplicación de la transformada localizada de Fourier a las observaciones ***x(t)***, da lugar a la siguiente representación en el dominio *tiempo-frecuencia*:

$$\hat{\mathbf{x}}(\omega, t_s) = \sum_t e^{-j\omega t}\mathbf{x}(t)w(t-t_s)$$
$$\omega = 0, \frac{2\pi}{\mathrm{N}}, \ldots, \frac{2\pi(N-1)}{\mathrm{N}} \tag{5.2}$$
$$t_s = 0, \Delta T, 2\Delta T$$

donde ω hace referencia a la frecuencia, N denota el número de puntos en la transformada discreta de Fourier, w es una ventana del tipo Hamming, Hanning o Kaiser, t_s la posición de la ventana y ΔT es el desplazamiento de la misma. Se ha de cumplir que la longitud de la ventana sea menor que el periodo de tiempo en el que se verifica que la señal es estacionaria.

De forma análoga se puede definir la transformada localizada *inversa* de Fourier como:

$$\mathbf{x}(t) = \frac{1}{2\pi} \cdot \frac{1}{W(t)} \sum_{t_s} \sum_{\omega} e^{j\omega(t-t_s)} \hat{\mathbf{x}}(\omega, t_s) \tag{5.3}$$

donde

$$W(t) = \sum_{t_s} w(t-t_s) \tag{5.4}$$

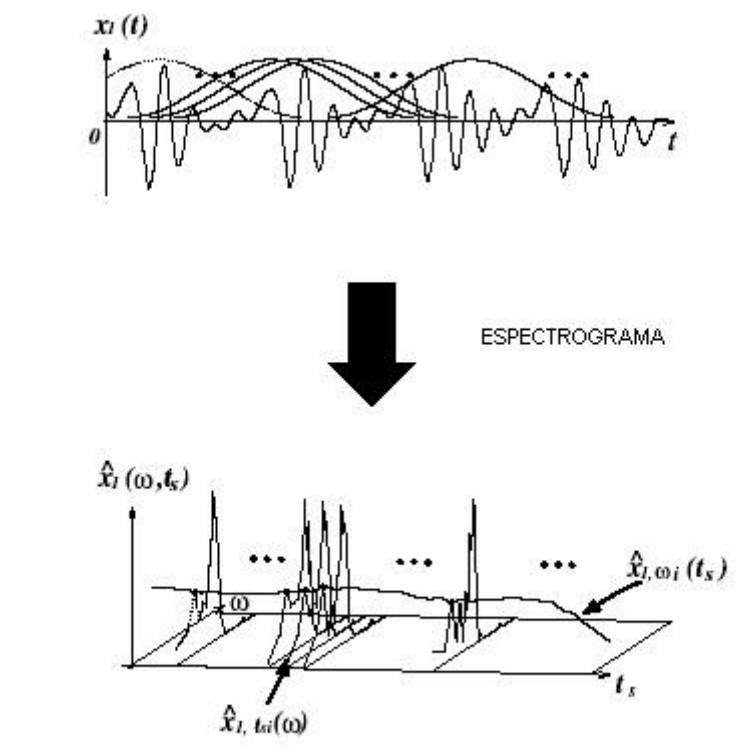

Figura 5.2 *Espectrograma. Dada una observación* $x_1(t)$, *es posible establecer una representación tiempo-frecuencia* $\hat{x}(\omega, t_s)$ *en la que para cada instante de tiempo* t_s *se puede observar el valor que toma cada una de las componentes frecuenciales que componen la señal.*

Por otro lado, la relación entre las observaciones y las fuentes originales se puede expresar de la siguiente forma:

$$\hat{\mathbf{x}}(\omega, t_s) = \hat{\mathbf{A}}(\omega)\hat{\mathbf{s}}(\omega, t_s) \tag{5.5}$$

donde $\hat{\mathbf{A}}(\omega)$ es la transformada de Fourier de **A**(t) y $\hat{\mathbf{s}}(\omega, t_s)$ la correspondiente de **s**(t).

5.2.3 Descripción del método de separación

En primer lugar, vamos a ver un algoritmo de estimación de la matriz de separación, que servirá de base para el método descrito en esta aplicación. Este algoritmo, llamado de Molgedey-Schuster [Molgedey], se fundamenta en la eliminación, en la medida de lo posible, de la correlación cruzada en las señales estimadas.

5.2.3.1 El algoritmo de Molgedey-Schuster

Consideremos que las señales fuente son débilmente estacionarias y las observaciones se corresponden con mezclas instantáneas (es decir, la matriz **A**) es constante. En este caso, la matriz de correlación de las observaciones se puede formular de la siguiente forma:

$$\left\langle \mathbf{x}(t)\mathbf{x}(t+\tau)^T \right\rangle = \mathbf{A}\left\langle \mathbf{s}(t)\mathbf{s}(t+\tau)^T \right\rangle \mathbf{A}^T = \mathbf{A}\begin{pmatrix} \mathbf{R}_{s1}(\tau) & \cdots & 0 \\ 0 & \ddots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \mathbf{R}_{sn}(\tau) \end{pmatrix} \tag{5.6}$$

donde $\mathbf{R}_{si}(\tau)$ es la matriz de autocorrelación de la señal de entrada $s_i(t)$.

Sea **W** una matriz de separación tal que se verifique que $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$, es decir que permita que las componentes estimadas sean mutuamente independientes sin conocer a priori la matriz de mezcla **A** ni las distribuciones de probabilidad de las señales originales $\mathbf{s}(t)$. Idealmente sería deseable que **W** fuera la inversa de la matriz **A** si bien ya que carecemos de información sobre las amplitudes de las señales fuente no será del todo posible. Tendremos que la relación entre ambas matrices será de la siguiente forma:

$$\mathbf{WA} = \mathbf{PerD} \tag{5.7}$$

donde **Per** es una matriz de permutación (es decir, una matriz en la que todos los elementos de cada fila y columna son cero salvo uno de ellos que vale ‘1’) y **D** es una matriz diagonal.

A partir de la ecuación anterior es posible determinar la matriz de correlaciones de las señales reconstruidas:

$$\left\langle \mathbf{y}(t)\mathbf{y}(t+\tau)^T \right\rangle = \left\langle \left(\mathbf{PerDs}(t)\right)\left(\mathbf{PerDs}(t+\tau)\right)^T \right\rangle = \begin{pmatrix} \lambda_1^2,\mathbf{R}_{s1'}(\tau) & \cdots & 0 \\ 0 & \ddots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_n^2,\mathbf{R}_{sn'}(\tau) \end{pmatrix} \quad (5.8)$$

denotando los índices *1', 2', ..., n'* las permutaciones de los índices *1, 2, ...,n* por los correspondientes que vienen dados por la matriz **Per**. λ_i hace referencia al i-ésimo elemento de la diagonal de **D**. El cálculo de la matriz **W** se hará de forma que se diagonalice la matriz de correlación para diferentes valores de τ.

El algoritmo propuesto permite simplificar el concepto de diagonalización simultánea de la matriz de correlación de las observaciones para cada valor de τ. Lo que se pretende es encontrar **W** tal que:

$$\mathbf{W}\left\langle \mathbf{x}(t)\mathbf{x}(t+\tau)^T \right\rangle \mathbf{W}^T = \mathbf{\Lambda}_i, \qquad i = 1, ..., r \quad (5.9)$$

siendo $\mathbf{\Lambda}_i$ matrices diagonales.

Existen diferentes métodos que permiten obtener una solución para el cálculo de **W**. Por ejemplo podemos usar un método basado en el gradiente que permite minimizar el error:

$$L(\mathbf{W}) = \sum_{i=1}^{r}\sum_{j\neq k}\left|\mathbf{W}\left\langle \mathbf{x}(t)x(t+\tau_i)^T \right\rangle \mathbf{W}^T\right|_{jk}^2 \quad (5.10)$$

El algoritmo de resolución consta de dos pasos, *blanqueo* y *rotación*. En el primero de ellos se realizará la operación de ortogonalización de las señales fuentes en las coordenadas de observación. Definamos la matriz de covarianzas de las observaciones:

$$\mathbf{C}_\mathbf{x} = \left\langle \mathbf{x}(t)\mathbf{x}(t)^T \right\rangle \quad (5.11)$$

y su raíz cuadrada inversa:

$$\sqrt{\mathbf{C}_\mathbf{x}^{-1}} = \sqrt{\mathbf{\Lambda}^{-1}}\mathbf{S}^T \quad (5.12)$$

donde **S** y **Λ** son una matriz ortogonal y diagonal respectivamente, que satisfacen:

$$\mathbf{C}_\mathbf{x} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^T \quad (5.13)$$

y $\sqrt{\mathbf{\Lambda}^{-1}}$ es una matriz diagonal que verifica que sus elementos son la raíz cuadrada de los de $\mathbf{\Lambda}^{-1}$.

Debido a la que las señales son débilmente estacionarias, la matriz de correlación es independiente del tiempo. De esta forma, podremos transforma el vector de observaciones como:

$$\mathbf{x}'(t) = \sqrt{\mathbf{C}_{\mathbf{x}}^{-1}}\mathbf{x}(t) \tag{5.14}$$

y conseguiremos que la matriz de correlaciones del nuevo vector de observaciones $\mathbf{x}'(t)$ esté ortogonalizada, es decir:

$$\left\langle \mathbf{x}'(t)\mathbf{x}'(t)^T \right\rangle = \sqrt{\mathbf{C}_{\mathbf{x}}^{-1}}\mathbf{C}_{\mathbf{x}}\sqrt{\mathbf{C}_{\mathbf{x}}^{-1}}^T = \mathbf{I} \tag{5.15}$$

siendo **I** la matriz identidad. De forma intuitiva podemos ver como las observaciones se pueden entender como las proyecciones de las fuentes sobre ciertas direcciones del espacio. Como ya se ha estudiado en secciones previas, aunque las direcciones de las observaciones originales son en general no ortogonales, el blanqueo permite que sí lo sean.

Tras el blanqueo de las señales originales, es preciso realizar una rotación que vendrá determinada por la eliminación de los elementos que estén fuera de la diagonal principal de la matriz de correlación para diferentes valores del retraso τ. Una forma de realizar dicha rotación es a partir de una cierta matriz, que llamaremos **Q** que va a minimizar el siguiente doble sumatorio:

$$\sum_{k=1}^{r}\sum_{i\neq j}\left|(QM_kQ^T)_{ij}\right|^2 \tag{5.16}$$

donde $(QM_kQ^T)_{ij}$ representa el ij-ésimo elemento de la matriz QM_kQ^T y:

$$\mathbf{M}_k = \left\langle \mathbf{x}'(t)\mathbf{x}'(t+\tau_k)^T \right\rangle, \qquad k = 1, ..., r \tag{5.17}$$

Para resolver este problema de diagonalización simultánea podemos usar el método de Jacobi propuesto en [Cardoso]. De esta forma llegaremos a que la matriz de separación **W** será:

$$\mathbf{W} = \mathbf{Q}\sqrt{\mathbf{C}_x^{-1}} \tag{5.18}$$

Una ventaja de este algoritmo en relación a otros es que exclusivamente se usan estadísticos de segundo orden. Además incluso se podría emplear cuando las observaciones estén inmersas en ruido blanco, que se puede eliminar usando tan sólo las correlaciones cruzadas para $\tau \neq 0$.

5.2.3.2 Aplicación del algoritmo de Molgedey-Schuster

En primer lugar, fijaremos la frecuencia ω para los espectrogramas, esto es, $\hat{\mathbf{x}}_{\omega}(t_s) = \hat{\mathbf{x}}(\omega, t_s)$, que será por tanto función de t_s. A continuación aplicaremos el algoritmo de *Molgedey-Schuster* para estimar la transformada de la matriz de separación $\hat{\mathbf{A}}(\omega)$.

Hemos de tener en cuenta que $\hat{\mathbf{x}}_{\omega}(t_s)$ es en general un valor complejo por lo que el algoritmo descrito anteriormente se ha de extender al plano complejo para que sea realizable. Esto se puede hacer de forma sencilla tan sólo sustituyendo la matriz hermítica y unitaria por la matriz simétrica y ortogonal respectivamente.

Una vez aplicado el algoritmo, obtendremos una serie de señales estimadas, que serán función de t_s además de mutuamente independientes en ω:

$$\hat{\mathbf{u}}_{\omega}(t_s) = \mathbf{W}(\omega)\hat{\mathbf{x}}_{\omega}(t_s) \tag{5.19}$$

Debido a que la separación ciega de fuentes no puede resolver el problema de las permutaciones y cambios de escala en las señales estimadas, para cada una de las frecuencias existen cambios de orden y amplitud. Esto significa que aunque coloquemos $\hat{\mathbf{u}}_{\omega}(t_s)$ en su orden correspondiente en ω, esos espectrogramas están en realidad están combinados y sus amplitudes no son las que corresponden.

5.2.3.3 Resolución del problema de la permutación y escalado

El fundamento que permite solventar esta situación se basa en desensamblar los espectrogramas haciendo uso de la propiedad de independencia que se da en cada una de las frecuencias para las que se han calculado las señales estimadas. Para ello, definamos los espectrogramas procedentes de este proceso de división como:

$$\hat{\mathbf{v}}_{\omega}(t_s; i) = \mathbf{W}(\omega)^{-1} \begin{pmatrix} 0 \\ \vdots \\ \hat{u}_{i,\omega}(t_s) \\ \vdots \\ 0 \end{pmatrix} \tag{5.20}$$

donde el índice i denota la dependencia de los espectrogramas con ω en la i-ésima componente independiente de $\hat{\mathbf{u}}_{\omega}(t_s)$. Se puede observar como implícitamente, i es función de la frecuencia ω, es decir $i = i(\omega)$. Para solventar el problema de los escalados, al multiplicar por $\mathbf{W}$ y $\mathbf{W}^{-1}$, se elimina dicha ambigüedad.

Sin embargo, aún permanece el problema de las permutaciones. Basándonos en la no estacionariadad de las señales fuentes encontraremos una forma de resolverlo, de forma que haciendo uso del segundo principio enunciado en el *apartado 5.2.1*, si los espectrogramas de las señales paso de banda, $\hat{\mathbf{v}}_{\omega}(t_s; i)$, se originan a partir de las

señales fuentes, parece lógico pensar que van a estar bajo la influencia de un mismo proceso de modulación en amplitud. Para verlo, consideremos un operador ε que definiremos como:

$$\varepsilon\hat{\mathbf{v}}_{\omega}(t_s;i)=\frac{1}{2\mathrm{M}}\sum_{t_s'=t_s-M}^{t_s+M}\sum_{j=1}^{n}\left|v_{j,\omega}(t_s';i)\right| \tag{5.21}$$

siendo M una constante positiva mientras que $v_{j,\omega}(t_s';i)$ denota el j-ésimo elemento de $\hat{\mathbf{v}}_{\omega}(t_s;i)$. Por otro lado podemos definir su producto escalar como:

$$\varepsilon\hat{\mathbf{v}}_{\omega}(i)\cdot\varepsilon\hat{\mathbf{v}}_{\omega'}(j)=\sum_{t_s}\varepsilon\hat{\mathbf{v}}_{\omega}(t_s;i)\varepsilon\hat{\mathbf{v}}_{\omega'}(t_s;j) \tag{5.22}$$

y su norma:

$$\left\|\varepsilon\hat{\mathbf{v}}_{\omega}(i)\right\|=\sqrt{\varepsilon\hat{\mathbf{v}}_{\omega}(i)\cdot\varepsilon\hat{\mathbf{v}}_{\omega}(i)} \tag{5.23}$$

El problema de las permutaciones se resuelve tan sólo ordenando los diferentes valores de los resultados de aplicar el operador ε. Como podemos comprobar, la ordenación se va a realizar a partir de las correlaciones entre las envolventes de las señales paso de banda.

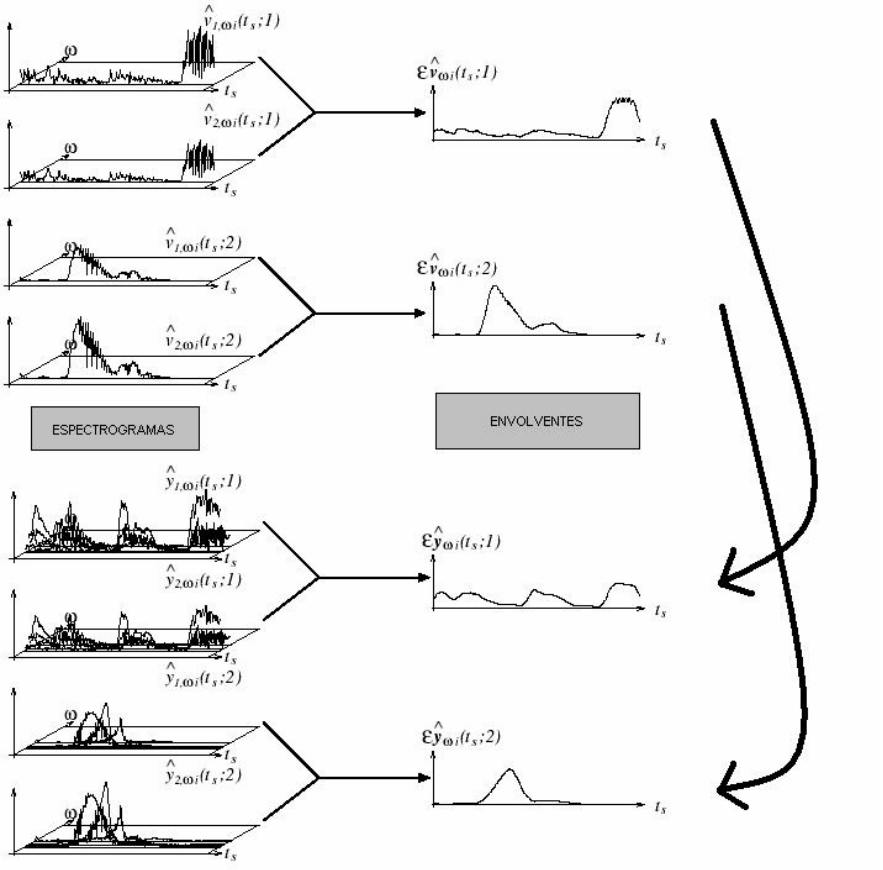

Figura 5.3 *Resolución del problema de las permutaciones*

De forma esquemática, podemos establecer los siguientes pasos que permitirán obtener los espectrogramas ya ordenados de forma adecuada:

1. Ordenar ω en el orden de menor a mayor correlación entre las componentes independientes en ω, esto es:

$$sim(\omega) = \sum_{i \neq j} \frac{\varepsilon\hat{\mathbf{v}}_{\omega}(i) \cdot \varepsilon\hat{\mathbf{v}}_{\omega}(j)}{\left\|\varepsilon\hat{\mathbf{v}}_{\omega}(i)\right\| \left\|\varepsilon\hat{\mathbf{v}}_{\omega}(j)\right\|} \qquad (5.24)$$
$$\text{si } sim(\omega_1) \leq \text{si } sim(\omega_1) \leq ... \leq \text{si } sim(\omega_N)$$

2. Para ω_1, asignar $\hat{\mathbf{v}}_{\omega_1}(t_s;i)$ a $\hat{\mathbf{y}}_{\omega_1}(t_s;i)$ de la siguiente forma:

$$\hat{\mathbf{y}}_{\omega_1}(t_s;i) = \hat{\mathbf{v}}_{\omega_1}(t_s;i), \quad i = 1, ..., n \qquad (5.25)$$

3. Para ω_k, encontrar la permutación *σ(i)* que maximiza la correlación entre las envolventes de ω_k y las envolventes acumuladas desde ω_1 hasta ω_{k-1}. Esto se consigue a partir de la maximización de:

$$\sum_{i=1}^{n} \varepsilon\hat{\mathbf{v}}_{\omega_k}(\sigma(i)) \cdot \left(\sum_{j=1}^{k-1} \varepsilon\hat{\mathbf{y}}_{\omega_j}(i) \right) \qquad (5.26)$$

teniendo en cuenta todas las permutaciones posibles de σ desde $i = 1, ..., n$.

4. Asignar la permutación adecuada a $\hat{\mathbf{y}}_{\omega_k}(t_s;i)$:

$$\hat{\mathbf{y}}_{\omega_k}(t_s;i) = \hat{\mathbf{v}}_{\omega_k}(t_s;\sigma(i)) \qquad (5.27)$$

Una vez realizadas estas operaciones, podremos resolver el problema de las permutaciones y obtener los correspondientes espectrogramas:

$$\hat{\mathbf{y}}_{\omega}(\omega, t_s;i) = \hat{\mathbf{y}}_{\omega}(t_s;i) \qquad (5.28)$$

En último lugar, aplicando la transformada inversa tal y como se definió en la ecuación (5.3) podemos obtener las componentes independientes estimadas:

$$\mathbf{y}(t;i) = \frac{1}{2\pi} \cdot \frac{1}{W(t)} \sum_{t_s} \sum_{\omega} e^{j\omega(t-t_s)} \hat{\mathbf{y}}(\omega, t_s) \qquad (5.29)$$

donde *i* denota el índice de la componente independiente. Hay que tener en cuenta que se han extraído las señales $\mathbf{y}(t;i)$, que tienen la misma dimensión que $\mathbf{x}(t)$, siendo *i* el número de fuentes. Es decir, si tenemos dos entradas, obtendremos 4 salidas según este algoritmo. Además hay que reseñar que de la definición se extrae que $\sum_i \mathbf{y}(t;i) = \mathbf{x}(t)$.

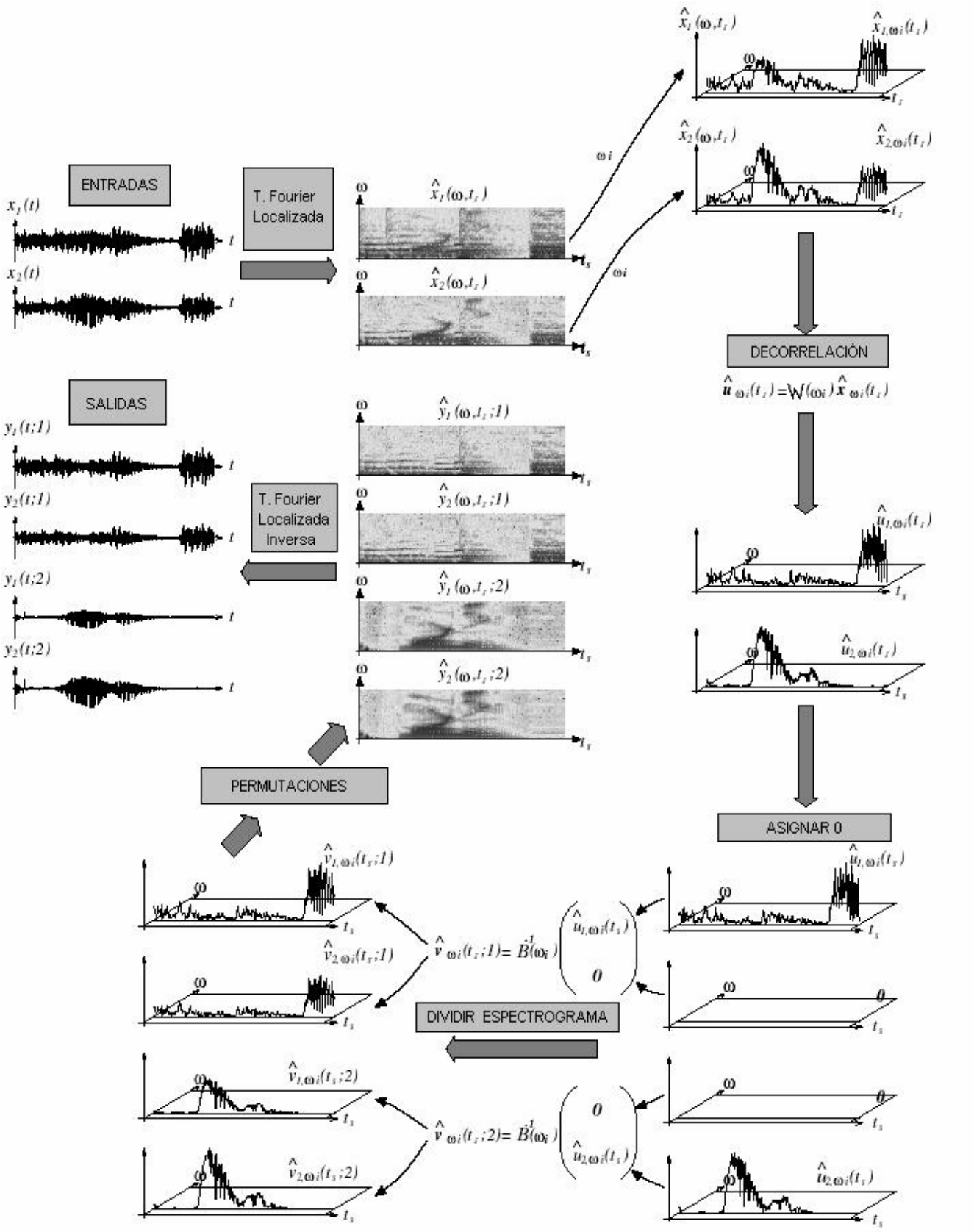

Figura 5.4 *Esquema general del algoritmo propuesto.*

5.3 Simulaciones y resultados

En este apartado se van a presentar los experimentos realizados con el algoritmo expuesto previamente y sus correspondientes resultados, que se analizarán de cara a observar la influencia de ciertos parámetros que permitan mejorarlos.

5.3.1 Generación de datos

Para las simulaciones se han tomado dos fuentes que han sido mezcladas artificialmente de forma convolutiva. Las señales originales s_1 y s_2 proceden de la grabación en un ordenador personal a una frecuencia de 8 KHz y 8 bits por muestra. En primer lugar, s_1 representa el vocablo japonés 'konbanwa' (que significa 'buenas tardes'), mientras que la señal s_2 contiene las palabras 'good morning'. Dichas señales se muestran seguidamente:

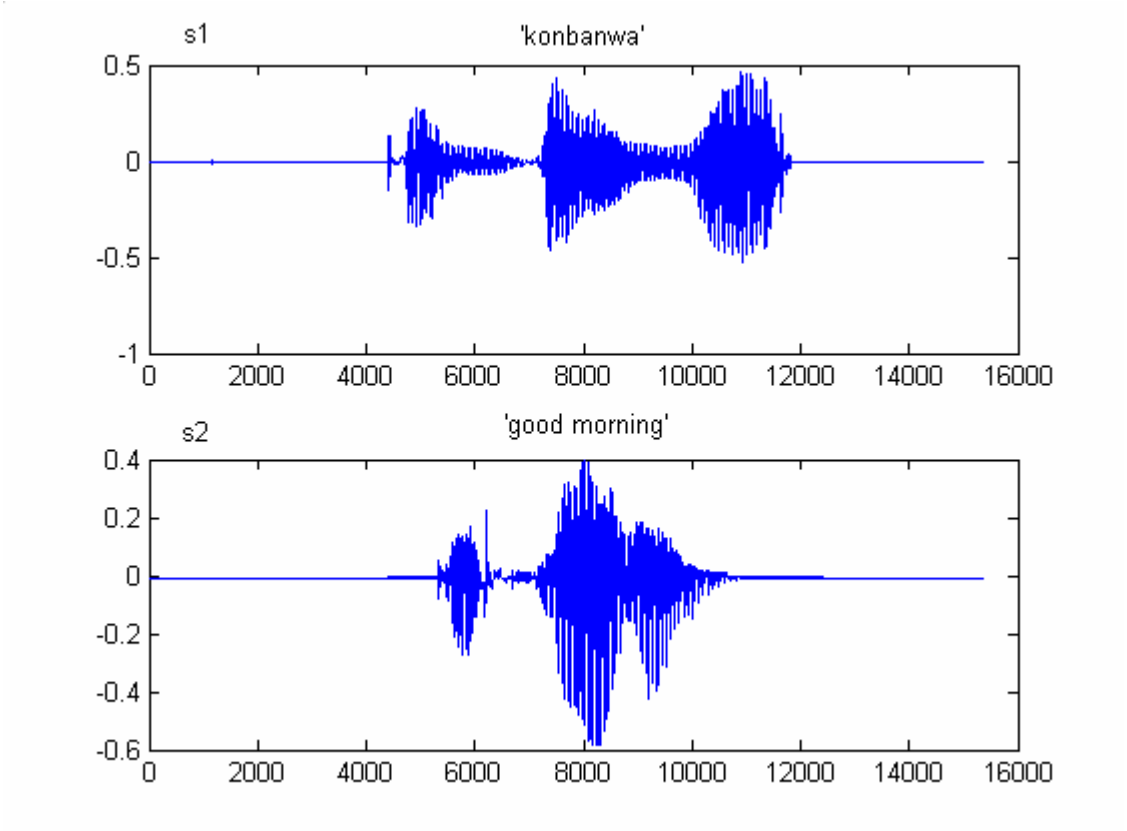

Figura 5.5 *Señales originales s_1 y s_2*

Una vez obtenidas las grabaciones procedemos a mezclarlas de forma artificial para obtener las observaciones, que llamaremos x_1 y x_2. Para generar la mezcla recurrimos a la siguiente expresión:

$$\mathbf{x}(t) = \mathbf{A} * \mathbf{s}(t) = \begin{pmatrix} \sum_j a_{1j} * s_j(t) \\ \sum_j a_{2j} * s_j(t) \end{pmatrix} \tag{5.30}$$

donde la convolución entre se define de la siguiente forma:

$$a_{ij} * s_j(t) = \sum_{\tau=0}^{\infty} a_{ij}(\tau) s_j(t-\tau) \tag{5.31}$$

siendo los a_{ij} los coeficientes de la matriz de mezcla, que se tomarán de forma completamente aleatoria ya que no se tiene información de la configuración geométrica del recinto en el que se han realizado las grabaciones.

5.3.2 La Relación Señal a Ruido como medida de calidad

La calidad de las señales recuperadas, se medirá a partir de la *Relación Señal a Ruido (Signal to Noise Ratio, SNR)*. Para ello definiremos las siguientes expresiones:

$$sg_i(t;i) = a_{ij} s_j(t) \tag{5.32}$$

$$error_i(t;i) = y_i(t;i) - sg_i(t;j) \tag{5.33}$$

$$SNR_{ij} = 10\log_{10} \frac{\sum_t sg_i(t;j)^2}{\sum_t error_i(t;j)^2} \tag{5.34}$$

5.3.3 Ajuste de los parámetros

Antes de iniciar la simulación en Matlab, habrá que ajustar una serie de parámetros que permitan que la calidad de las señales recuperadas sea máxima. Estos parámetros serán:

1. *Desplazamiento de la ventana*, que denotaremos por ΔT.
2. *Número de matrices a diagonalizar de forma simultánea*, parámetro que llamamos r en el estudio teórico.
3. *Longitud de la ventana* empleada en la transformada localizada de Fourier, N.
4. *Tipo de ventana* utilizada.

Durante los diferentes experimentos que se vayan realizando iremos fijando alguno de los parámetros en función de la calidad de los resultados que proporcionen.

5.3.4 Estimación de las componentes independientes

5.3.4.1 Experimento 1. Efectos de *r* y ΔT

En primer lugar realizaremos cuatro medidas fijando el parámetro *r* (número de matrices a diagonalizar de forma simultánea) a los valores 20, 40 y 60, de forma que para cada uno de ellos se variará el solapamiento entre las 10 y 200 muestras con incrementos de 10 muestras (dado que la frecuencia de muestreo es de 8 KHz, estos valores oscilarán entre los 1.25 y 25.0 mseg). El tamaño de la ventana Hanning que usaremos es de 256 muestras (32 mseg), si bien en el apartado siguiente estudiaremos el efecto que tiene la variación de este parámetro sobre los resultados.

Seguidamente vamos a mostrar la SNR obtenida para las señales estimadas correspondientes a s_1 y s_2.

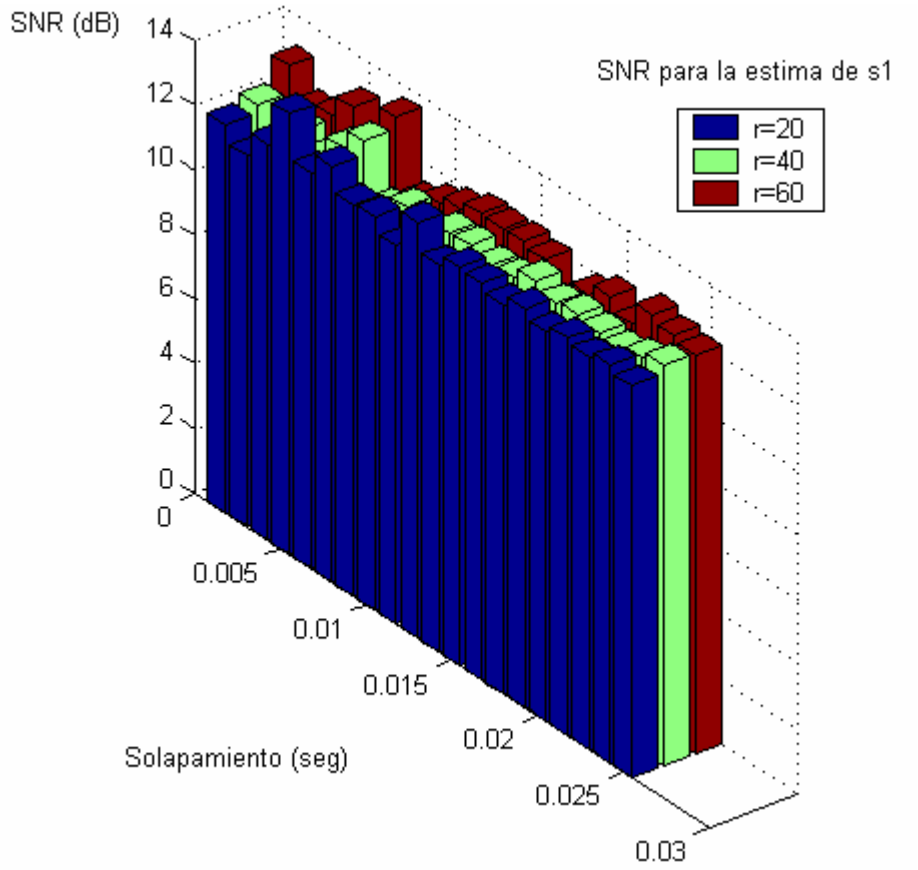

Figura 5.6 *Relación Señal a Ruido en dB pertenecientes a la estima de la señal s_1. Se muestra la evolución para tres valores concretos de 'r', variando el solapamiento desde 1.25 a 25.0 mseg. El tamaño de la ventana es de 32 mseg.*

En el caso de la señal estimada correspondiente a la primera fuente, vemos como la zona en la que la SNR es mayor se corresponde con un solapamiento en torno a los 5 milisegundos independientemente de cual sea el número de matrices simultáneas a

diagonalizar. Esto valores estarán en torno a los 13.5 decibelios, lo que implica que la relación de diafonía resultará aproximadamente 1/22.

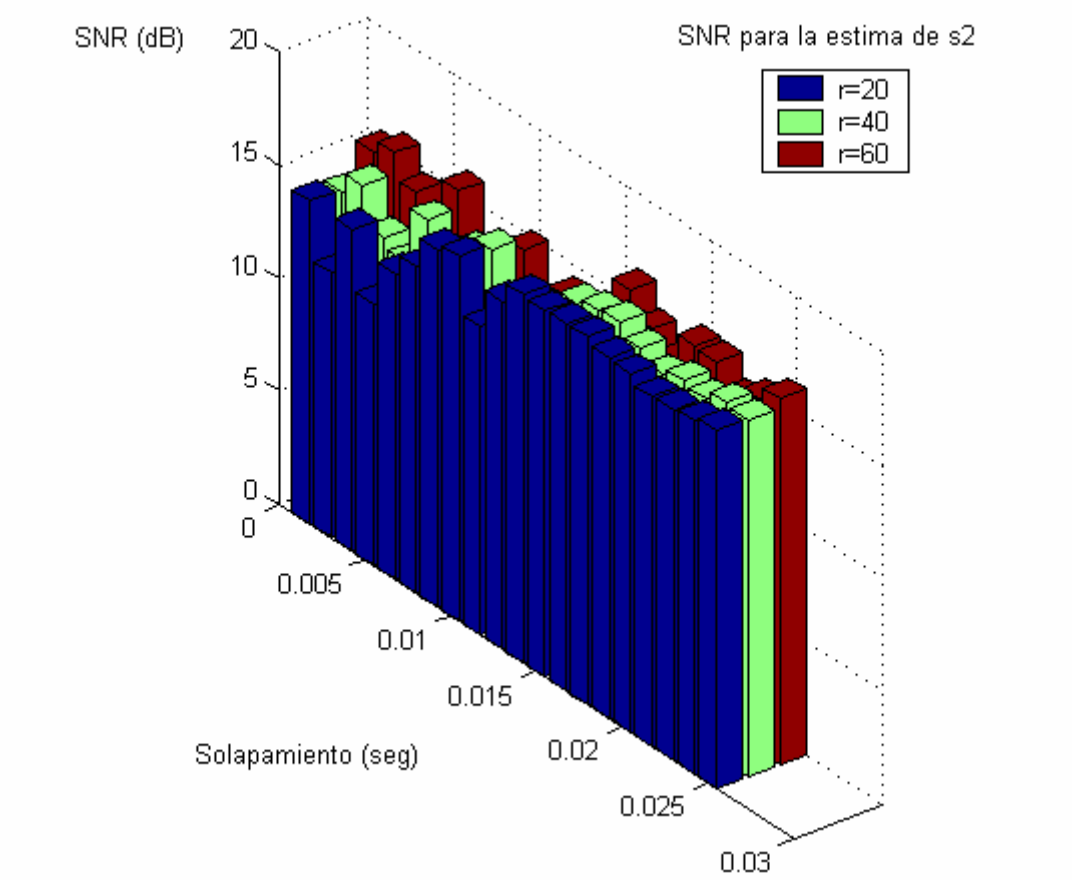

Figura 5.7 *Relación Señal a Ruido en dB pertenecientes a la estima de la señal s_2. Se muestra la evolución para tres valores concretos de 'r', variando el solapamiento desde 1.25 a 25.0 mseg. El tamaño de la ventana es de 32 mseg.*

En el caso de la segunda señal, podemos comprobar como la Relación Señal a Ruido resultante toma unos valores ligeramente superiores a los obtenidos en el caso de la señal s_1. Los valores máximos de la SNR se sitúan en torno a los 16.5 decibelios, lo cual implica una relación de diafonía de aproximadamente 1/45.

En la siguiente tabla comparativa se muestran los valores máximos de la Relación Señal a Ruido (y el consiguiente valor del solapamiento en el que se dan) para cada una de las señales s_1 y s_2 asociados a cada uno de los tres valores de 'r' empleados en la simulación.

Señal s_1			
r	*20*	*40*	*60*
SNR máxima (dB)	13.6471	13.2814	13.6517
Solapamiento (mseg)	5	7.5	7.5

Tabla 5.1 *Valores máximos de la Relación Señal a Ruido asociada a la señal s_1 y su correspondiente valor de solapamiento.*

Señal s_2			
r	*20*	*40*	*60*
SNR máxima (dB)	16.3109	16.4029	16.6920
Solapamiento (mseg)	17.5	17.5	16.25

Tabla 5.2 *Valores máximos de la Relación Señal a Ruido asociada a la señal s_2 y su correspondiente valor de solapamiento.*

A la vista de los resultados anteriores, podemos verificar como para el caso de ambas señales, la mejor SNR se da para un valor de $r = 60$, si bien los valores correspondientes al solapamiento difieren en aproximadamente 10 milisegundos.

De todas formas, dado que para cada simulación el valor de 'r' ha de ser el mismo para las dos señales, observando las *figuras 5.7* y *5.8* tendremos que tomar un valor de compromiso en el que se consiga que ninguna de las señales recuperadas se vea afectada en exceso. El valor elegido para 'r' será de 60 ya que aquí se localiza el máximo de SNR para ambas señales. El principal problema radica en la elección del solapamiento, ya que a partir de las figuras anteriores podemos comprobar como por regla general un máximo de SNR en una de señales viene acompañado por un valor relativamente bajo en la otra. Por tanto se ha pensado en estimar el valor del solapamiento como aquel que hace que la Relación Señal a Ruido de ambas señales se sitúe por encima de la media de forma similar en mayor medida. Observando los resultados obtenidos, podemos comprobar como el valor del solapamiento que lo verifica es precisamente el primero de ellos, es decir, 1.25 milisegundos.

Las señales recuperadas en este caso se muestran en la siguiente figura:

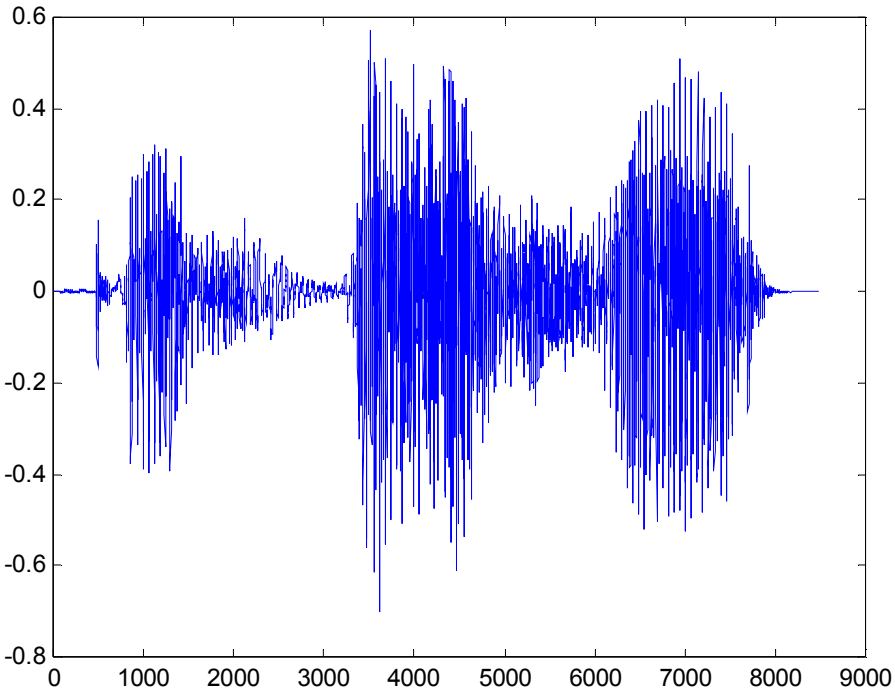

Figura 5.8 *Señal s_1 para los valores elegidos de los parámetros (r = 60 y solapamiento de 1.25 milisegundos).*

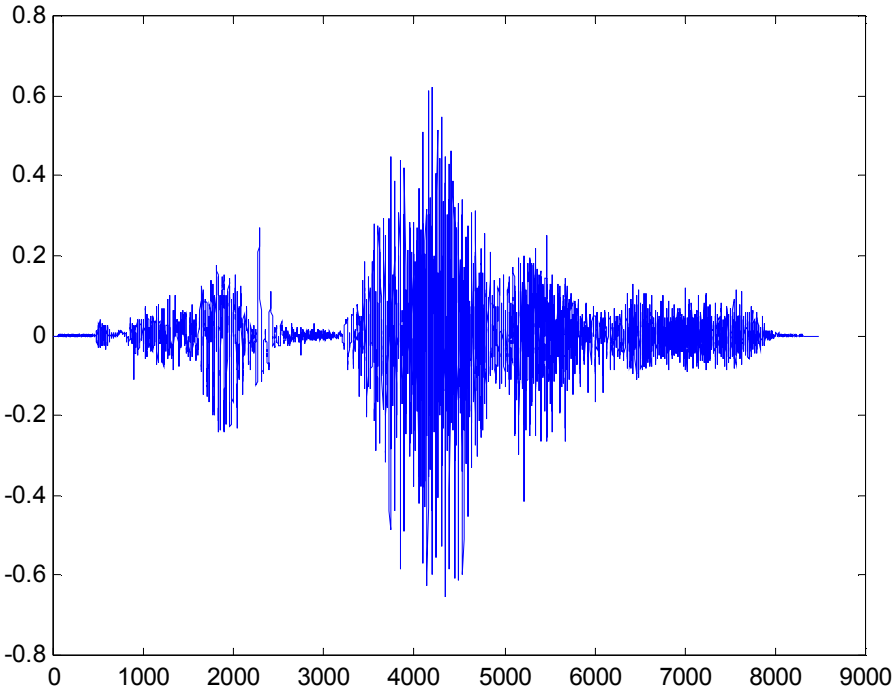

Figura 5.9 *Señal s_2 para los valores elegidos de los parámetros (r = 60 y solapamiento de 1.25 milisegundos).*

5.3.4.2 Experimento 2. Efecto del tamaño de la ventana

Seguidamente pasaremos a ver el efecto que produce en los resultados el cambio del tamaño de la ventana, considerando valores más pequeños para así poder tener una mayor resolución en el tiempo y promediar parámetros que se suponen que no varían en el intervalo en el que la voz se supone estacionaria [Huang].

A raíz de los resultados obtenidos en el apartado anterior y una vez fijado el número de matrices a diagonalizar de forma simultánea a una valor de 60 y habiendo situado el solapamiento en torno a los 1.25 milisegundos, procederemos a repetir el proceso tomando $N = 64, N = 128, N = 256$ y un rango de valores del solapamiento entre los 1.25 y 7.5 milisegundos.

Los resultados obtenidos se muestran en las siguientes gráficas:

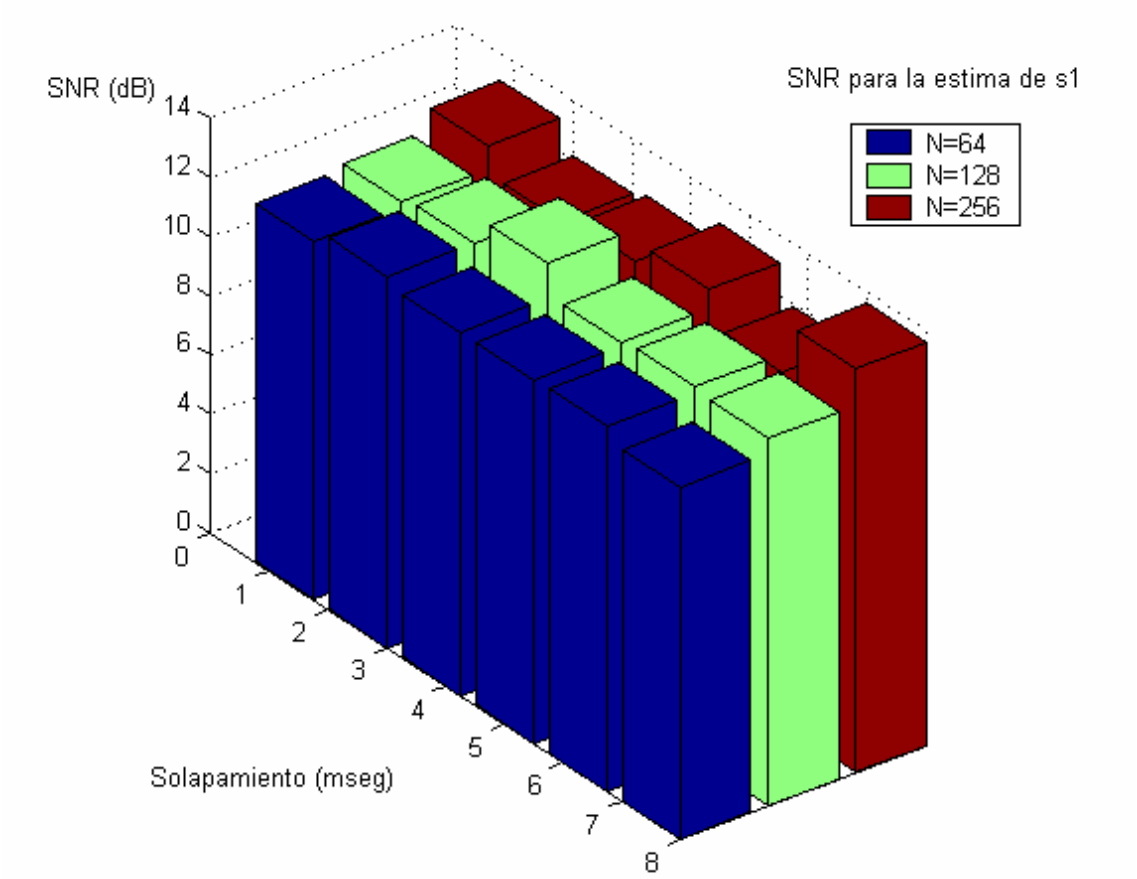

Figura 5.10 *Relación Señal a Ruido en dB pertenecientes a la estima de la señal s_1. Se muestra la evolución para tres valores concretos del tamaño de la ventana, variando el solapamiento desde 1.25 a 7.5 mseg. El número de matrices simultáneas a diagonalizar es 60.*

Para el caso de la primera señal podemos comprobar como los valores obtenidos resultan bastante parejos, localizando los máximos en los valores que se presentan en la siguiente tabla:

Señal s_1			
N	*64*	*128*	*256*
SNR máxima (dB)	12.5576	13.5020	13.6517
Solapamiento (mseg)	5	3.75	7.5

Tabla 5.3 *Valores máximos de la Relación Señal a Ruido asociada a la señal s_1 y su correspondiente valor de solapamiento.*

A la vista de los resultados de la *tabla 5.3* verificamos que el máximo valor de la Relación Señal a Ruido se vuelve a localizar para una tamaño de la ventana de 256 muestras, esto es 32 milisegundos (si bien el valor máximo para $N = 128$ es bastante similar). Por tanto, la relación de diafonía coincide con la obtenida en el apartado anterior y será 1/22.

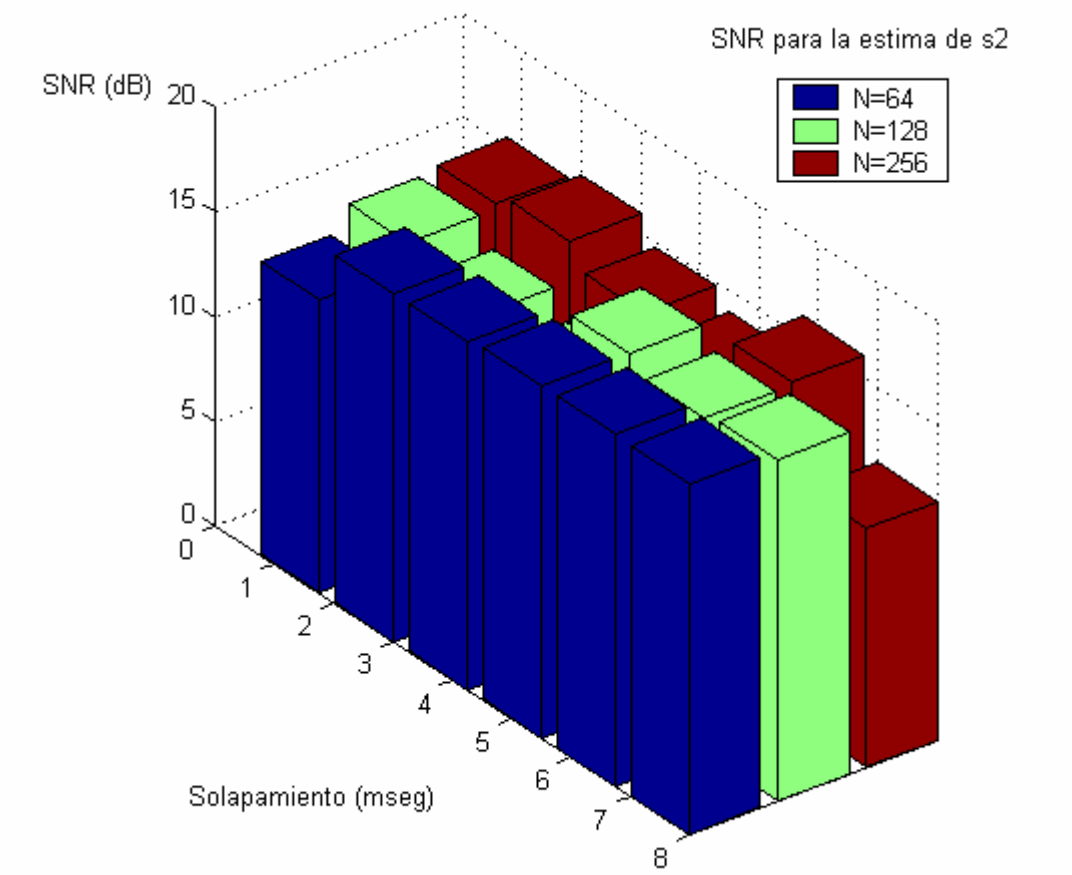

Figura 5.11 *Relación Señal a Ruido en dB pertenecientes a la estima de la señal s_2. Se muestra la evolución para tres valores concretos del tamaño de la ventana, variando el solapamiento desde 1.25 a 7.5 mseg. El número de matrices simultáneas a diagonalizar es 60.*

En el caso de la señal s_2 se van a observar ligeras variaciones con respecto a los resultados del apartado anterior, como podemos comprobar en la siguiente tabla:

Señal s_2			
N	*64*	*128*	*256*
SNR máxima (dB)	16.8749	16.8076	16.1088
Solapamiento (mseg)	5	5	6.25

Tabla 5.4 *Valores máximos de la Relación Señal a Ruido asociada a la señal s_2 y su correspondiente valor de solapamiento.*

A la vista de los resultados de la *tabla 5.4* verificamos que el máximo valor de la Relación Señal a Ruido se verifica cuando el tamaño de la ventana es de 64 muestras, esto es 8 milisegundos (si bien el valor máximo para $N = 128$ es bastante similar). En este caso, la relación de diafonía se ve incrementada con respecto al apartado anterior y toma un valor aproximado de 1/49.

Llegados a este punto, el principal problema con el que nos encontramos consiste en decidir qué valor de *N* tomar como más adecuado para continuar con el último de los experimentos. Al igual que en el apartado anterior, será necesario establecer un criterio de decisión, que basaremos en la Relación Señal a Ruido.

Observando las *tablas 5.3* y *5.4* podemos comprobar como para el caso $N = 256$ y un solapamiento de 1.25 milisegundos se consigue que las SNR resultantes para las dos señales se localicen conjuntamente por encima de la media en mayor medida. En este caso, las Relaciones Señal a Ruido resultantes serán de 13.0834 dB para la primera señal, y 15.3905 dB para la segunda, lo cual implica diafonías de 1/20 y 1/35 respectivamente. Las señales recuperadas coincidirán con las mostradas en las *figuras 5.8* y *5.9*.

5.3.4.3 Experimento 3. Efectos del tipo de ventana empleada.

Una vez hemos observado la influencia del tamaño de la ventana, el solapamiento y el número de matrices simultáneas a diagonalizar, la última variante a realizar consistirá en ver el efecto que tiene en los resultados obtenidos la utilización de un tipo de ventana u otra. Dado que hasta ahora, se ha venido empleando una ventana del tipo *Hanning*, desarrollaremos un nuevo experimento en el que se establecerán comparaciones con ventanas de *Hamming*, *Blackman* y *Rectangular*.

Para el caso de la ventana de *Hanning*, los resultados obtenidos en los apartados anteriores se pueden resumir en la siguiente tabla:

	Señal s_1			***Señal s_2***		
r	*20*	*40*	*60*	*20*	*40*	*60*
SNR máxima (dB)	13.6471	13.2814	13.6517	16.3109	16.4029	16.6920
Solapamiento (mseg)	5	7.5	7.5	17.5	17.5	16.25
SNR media (dB)	12.2787	12.1397	12.3271	15.0153	14.9742	14.9431
	12.2485			14.9775		

Tabla 5.5 *Ventana de Hanning. Valores medios máximos de la Relación Señal a Ruido asociada a las señales s_1 y s_2, y su correspondiente valor de solapamiento*.

En las siguientes simulaciones se fijará un tamaño de ventana de 256 muestras e iremos variando el valor de 'r' y del solapamiento como se hizo en los experimentos anteriores.

a. Ventana de Hamming

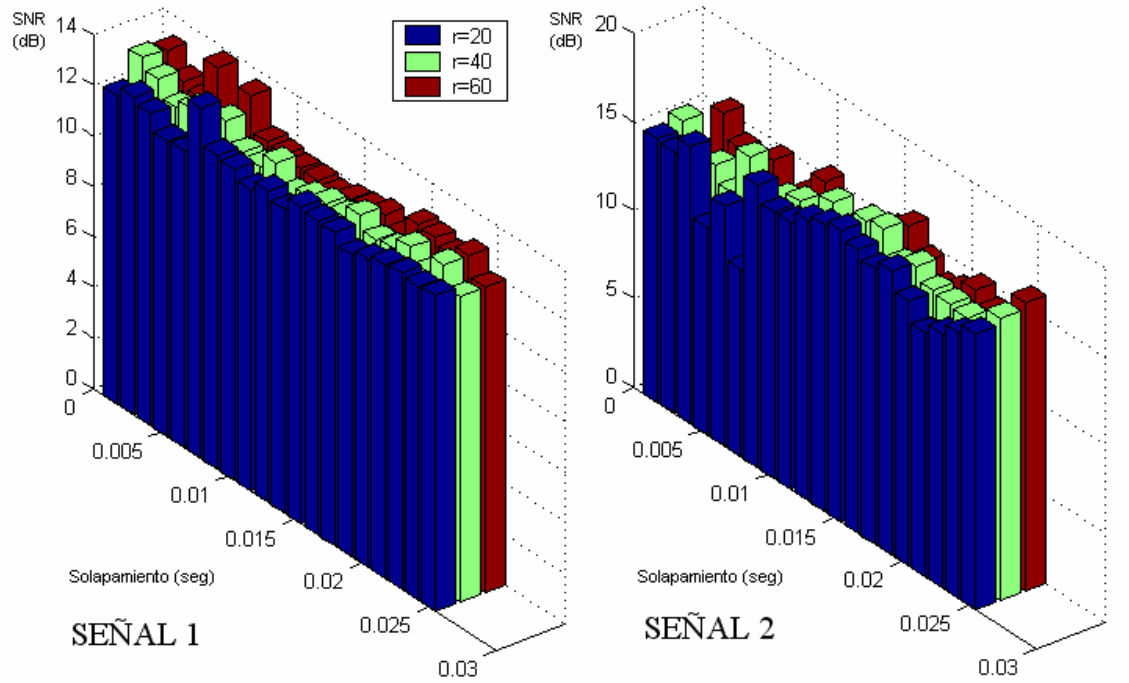

Figura 5.12 *Relación Señal a Ruido en dB pertenecientes a la estima de la señales s_1 y s_2 al emplear una ventana de Hamming de 256 puntos. Se muestra la evolución para tres valores concretos de 'r', variando el solapamiento desde 1.25 a 25.0 mseg.*

En la siguiente tabla se muestran los valores máximos de Relación Señal a Ruido para cada una de las señales, así como su valor medio:

	Señal s_1			***Señal s_2***		
r	*20*	*40*	*60*	*20*	*40*	*60*
SNR máxima (dB)	13.8708	13.3371	13.8767	16.5084	16.6782	16.3050
Solapamiento (mseg)	7.5	1.25	5	15	16.25	25
SNR media (dB)	12.3445	12.4318	12.4527	14.8778	14.7654	14.3697
	12.4097			14.6710		

Tabla 5.6 *Ventana Hamming. Valores medios máximos de la Relación Señal a Ruido asociada a las señales s_1 y s_2, y su correspondiente valor de solapamiento.*

b. Ventana Blackman

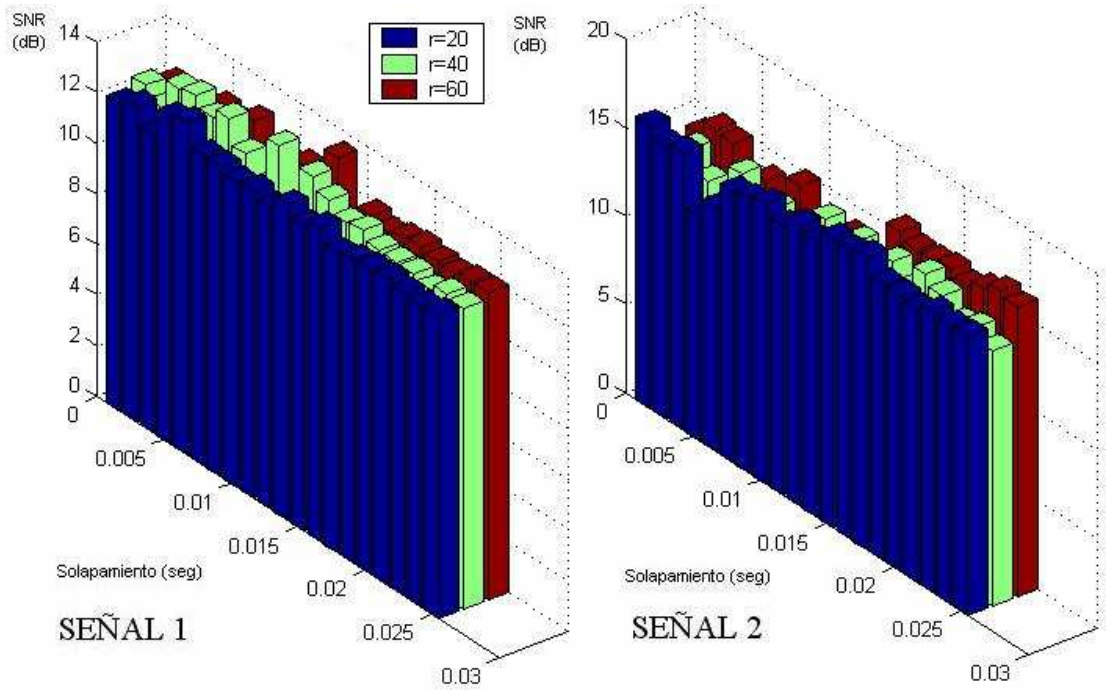

Figura 5.13 *Relación Señal a Ruido en dB pertenecientes a la estima de la señales s_1 y s_2 al emplear una ventana de Blackman de 256 puntos. Se muestra la evolución para tres valores concretos de 'r', variando el solapamiento desde 1.25 a 25.0 mseg.*

En la siguiente tabla se muestran los valores máximos de Relación Señal a Ruido para cada una de las señales, así como su valor medio:

	Señal s_1			***Señal s_2***		
r	*20*	*40*	*60*	*20*	*40*	*60*
SNR máxima (dB)	13.1505	13.5890	13.6003	16.5970	16.2826	16.6116
Solapamiento (mseg)	6.25	11.25	13.75	17.5	20	23.75
SNR media (dB)	12.2241	12.4580	12.1794	15.5103	14.2899	14.5547
	12.2872			14.7850		

Tabla 5.7 *Ventana Blackman. Valores medios máximos de la Relación Señal a Ruido asociada a las señales s_1 y s_2, y su correspondiente valor de solapamiento.*

c. Ventana Rectangular

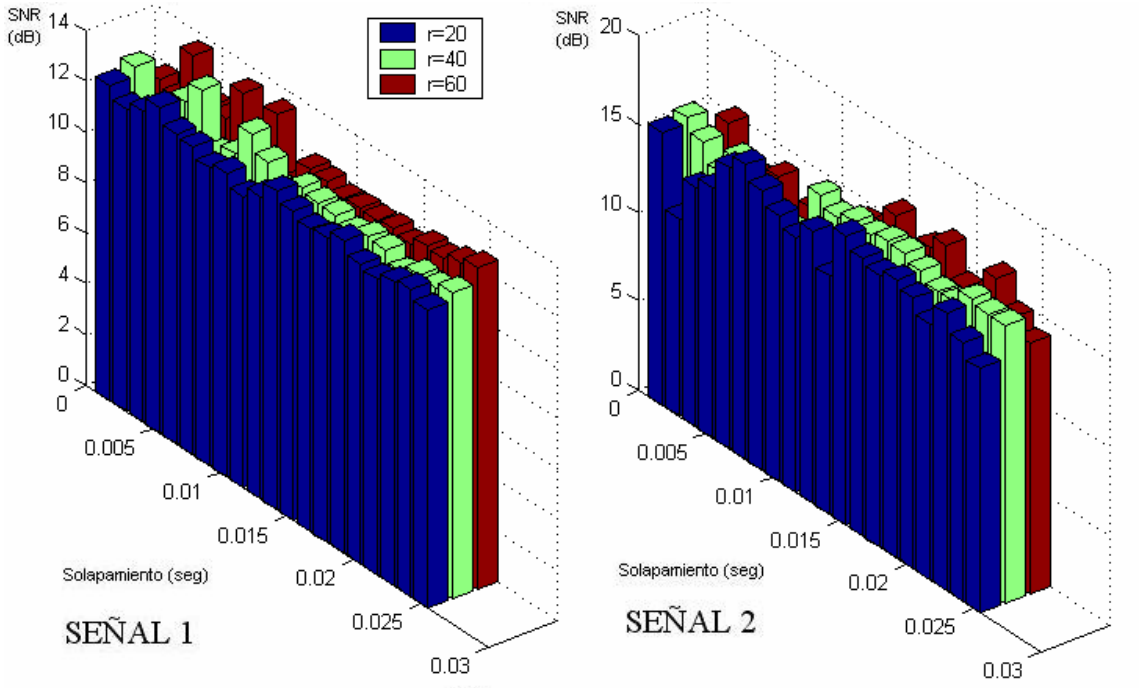

Figura 5.14 *Relación Señal a Ruido en dB pertenecientes a la estima de la señales s_1 y s_2 al emplear una ventana de Rectangular de 256 puntos. Se muestra la evolución para tres valores concretos de 'r', variando el solapamiento desde 1.25 a 25.0 mseg.*

En la siguiente tabla se muestran los valores máximos de Relación Señal a Ruido para cada una de las señales, así como su valor medio:

	Señal s_1			***Señal s_2***		
r	*20*	*40*	*60*	*20*	*40*	*60*
SNR máxima (dB)	12.8929	13.5956	13.7234	16.6599	16.2325	16.5918
Solapamiento (mseg)	5	6.25	3.75	7.5	11.25	18.75
SNR media (dB)	12.0999	12.1878	12.2848	14.8856	14.9389	14.2588
	12.1908			14.6944		

Tabla 5.8 *Ventana Rectangular. Valores medios máximos de la Relación Señal a Ruido asociada a las señales s_1 y s_2, y su correspondiente valor de solapamiento.*

Comparativa entre las diferentes ventanas

A la vista de los resultados obtenidos podemos comprobar como ninguna ventana en particular consigue mejorar en exceso la Relación Señal a Ruido de las señales estimadas. Con todo, la ventana de *Hanning* permite recuperar ambas señales con una SNR por encima de la media en mayor medida, y sin que ninguna de ellas se vea afectada negativamente.

5.3.4.4 Resultados definitivos

Con la información obtenida a partir de los tres experimentos, estamos ya en condiciones de establecer los valores de los cuatro parámetros que modelan el problema:

Parámetro	Valor
Tipo de ventana	Hanning
Tamaño de la ventana	256 muestras (32 mseg)
Solapamiento entre ventanas	10 muestras (1.25 mseg)
Nº matrices a diagonalizar	60 matrices

Tabla 5.9 *Valores designados para el conjunto de los parámetros de la simulación.*

Las señales definidamente recuperadas en estas condiciones se muestran en las siguientes figuras:

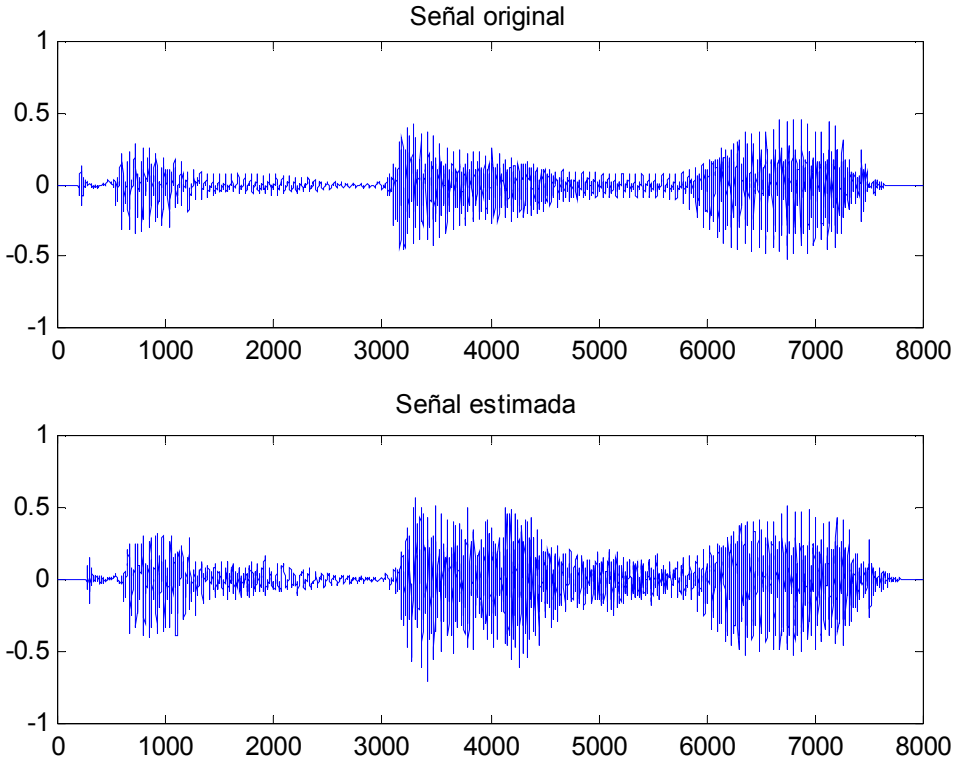

Figura 5.15 *Señal s_1 original y estimada para los valores de los parámetros definitivos.*

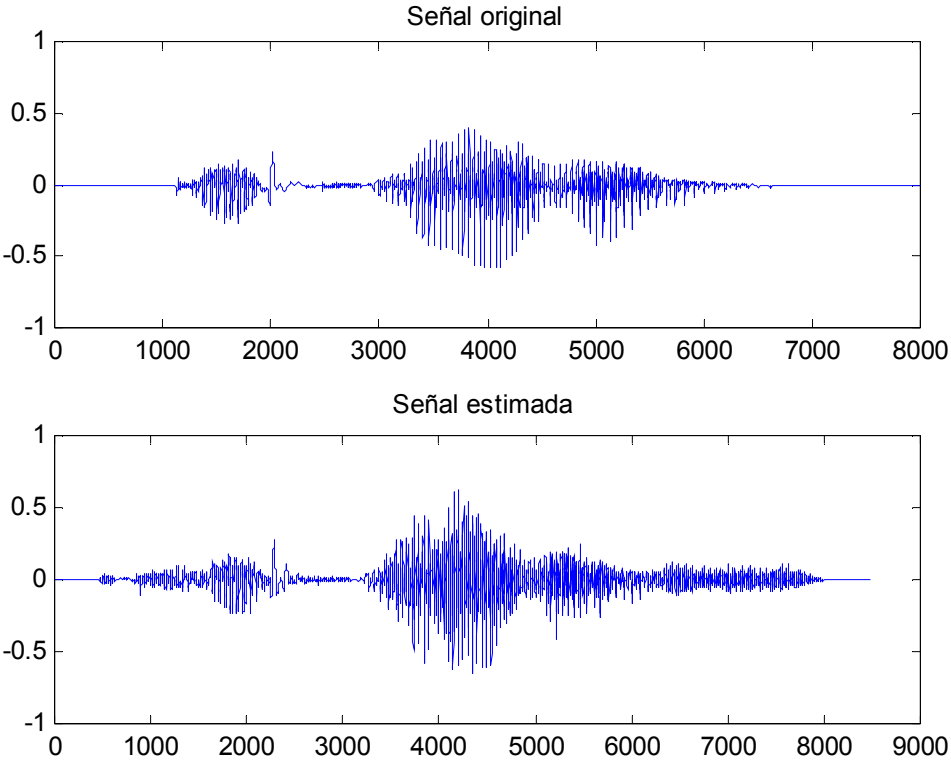

Figura 5.15 *Señal s_1 original y estimada para los valores de los parámetros definitivos.*

5.4 Conclusiones

El problema de la Separación Ciega de Fuentes planteado en los capítulos teóricos se corresponde con una aproximación, por lo que es necesario establecer un modelo que tenga en cuenta situaciones más próximas a la realidad. Las observaciones no se obtienen como combinaciones lineales de las señales de las fuentes, sino que provienen de un proceso de convolución que permite así simular los posibles efectos de retardos y reflexiones de un entorno real. Además, supondremos que las señales son de voz, con lo cual se podrán aplicar ciertas hipótesis de estacionaridad que permitirán resolver el problema siguiendo una determinada sistemática.

Planteamos un método de separación basado en el tratamiento del espectrograma, lo cual conlleva dos problemas asociados. Debido a que la separación ciega de fuentes no puede resolver el problema de las permutaciones y cambios de escala en las señales estimadas, para cada una de las componentes frecuenciales de las mismas se dan cambios de *orden* y *amplitud*. El problema de escalado no presenta excesivas complicaciones, mientras que para solventar el referente a las permutaciones, hay que hacer uso de la propiedad de no estacionaridad intrínseca de las señales de voz.

Una vez diseñado el algoritmo de separación, se realizaron diferentes simulaciones variando cada uno de los cuatro parámetros característicos que modelan el problema (tamaño y tipo de ventana, solapamiento entre ventanas y número de matrices simultáneas a diagonalizar), con el objeto de verificar qué combinación de los mismos permitía recuperar las señales con una Relación Señal a Ruido mejor para ambas.