Filtrado de señales de voz a través de “wavelet shrinkage”[1]

Diego R. Llanos Ferraris, Valentín Cardeñoso Payo

Departamento de Informática, Universidad de Valladolid
Edificio de Tecnología de la Información, Campus Miguel Delibes
47071 Valladolid - España
Tfno: (+34) 983 423670 - Fax: (+34) 983 423671
e-mail: diego@infor.uva.es valen@infor.uva.es

Resumen

El filtrado de señales de voz con el objeto de eliminar en lo posible los ruidos producidos por el sistema de grabación utilizado constituye una etapa previa en todo sistema de análisis del habla. En este documento se describen las experiencias que se están realizando en el filtrado de señales de voz utilizando transformadas wavelet ([5]). Se han desarrollado un conjunto de módulos implementados en Matlab [11] que permiten filtrar una señal de voz, apoyándose en las funciones de cálculo en el dominio wavelet (UviWave) realizadas por el grupo de Teoría de Señal de la Universidad de Vigo (España) [8]. Se han implementado diferentes criterios de filtrado, evaluando sus ventajas e inconvenientes, y propuesto otros que están siendo desarrollados en la actualidad. Se propone además una función de medida para evaluar la calidad del filtrado realizado. Finalmente, se evalúan veinticinco filtros wavelet con los criterios indicados, estableciéndose una comparativa utilizando la función de calidad antedicha.

1 Introducción

La técnica conocida como “wavelet shrinkage” ([5], [3], [4]) hace referencia a la reconstrucción de una señal afectada por ruido a través del siguiente procedimiento:

1. Obtener la transformada wavelet de la señal.
2. Minimizar los coeficientes cercanos a cero de la transformada.
3. Obtener la transformada inversa de la señal.

La ventaja de este procedimiento respecto de un filtrado por bandas de frecuencia reside en que se obtiene una señal casi libre de ruido, sin modificar apenas las características de la señal (presencia de picos de alta frecuencia, etcétera). Este resultado es muy distinto al que se obtiene mediante los métodos tradicionales de suavizado, que sólo consiguen eliminar el ruido a costa de suavizar también los componentes de la señal ([7]). Entre estos métodos podemos citar la utilización de splines con elección adaptativa del parámetro de tensión [6]. Otros métodos, como la estimación en Series de Fourier, respetan las características de la señal, pero no eliminan realmente el ruido [4].

[1]Publicado en las Actas IV Jornadas de Informática, organizadas por el Grupo de Arquitectura y Concurrencia (GAC), Departamento de Ingeniería Telemática, Universidad de Las Palmas de Gran Canaria. Depósito Legal GC-582-1998, ISBN 84-87526-61-6. Julio 1998.

Supongamos que estamos interesados en filtrar una función $f(t)$ compuesta de $n = 2^{j+1}$ muestras de la forma $y_i = f(t_i) + \sigma z_i$, con $i = 1, ...n$, con los valores de t_i equidistantes y con z_i ruido blanco. Donoho y Johnstone [5] han propuesto un método que consta de tres partes para recuperar $f(t)$:

1. Aplicar el filtrado wavelet piramidal propuesto por Cohen, Daubechies, Jawerth y Vial [1] al conjunto de datos, con lo que se obtiene un conjunto de coeficientes wavelet $w_{j,k}$, con $j = j_0, ..., J$ las diferentes escalas de resolución, y $k = 0, ..., 2^j - 1$ el número de muestras en cada escala.
2. Aplicar una contracción no lineal a los coeficientes wavelet obtenidos

$$\nu(w) = sign(w)(|w| - t)_+ \tag{1}$$

   utilizando para ello un umbral u igual a

$$u = \frac{\sqrt{2log(n)}\sigma}{\sqrt{n}} \tag{2}$$

   lo que lleva a unos nuevos coeficientes $\hat{w}_{j,k}$.
3. Poner a cero todos los coeficientes $\hat{w}_{j,k}$ para $j > J$ (lo que equivale a eliminar el residuo después de calcular la transformada wavelet para los J escalas superiores), y realizar la transformación inversa, lo que permite obtener un conjunto de muestras $\hat{f}(t)$.

Este método contrae los coeficientes wavelet cercanos a cero de forma empírica. Lo que se consigue con la contracción no lineal descrita en el paso 2 es lo siguiente:

- Si el valor absoluto del coeficiente w es mayor que el umbral u, se resta ese umbral del coeficiente.
- Si el valor absoluto del coeficiente es menor o igual al umbral, ese coeficiente se pone a cero.

Esta contracción es denominada por Donoho y Johnstone como "contracción suave" (*soft thresholding*) Alternativamente, también se propone una "contracción firme" (*hard thresholding*) consistente en poner a cero todos los coeficientes por debajo del umbral, dejando el resto inalterados.

De la lectura del artículo de Donoho y Johnstone se deduce que la elección del umbral u debe optimizarse para cada señal en particular, ya que como puede verse en el segundo paso del algoritmo de filtrado, dicho umbral depende de una constante σ relacionada con el nivel de ruido presente en la señal, valor desconocido en la práctica.

Hemos realizado una implementación del algoritmo en Matlab [11], utilizando un *toolbox* desarrollado en la Universidad de Vigo y denominado UviWave [8]. Dicha implementación tiene por objeto la evaluación del comportamiento de la técnica descrita en el filtrado de señales de voz, al objeto de utilizarla como etapa previa de los algoritmos de extracción de características de la señal que están siendo desarrollados en el Departamento de Informática de la Universidad de Valladolid ([6], [9]).

2 Implementación del algoritmo propuesto

La implementación del algoritmo de filtrado de señal de voz ha sido desarrollado utilizando como estructura básica el diagrama de bloques que aparece en la figura 1.

Como puede verse en dicha figura, la señal obtenida de la grabación (señal en bruto) pasa por una etapa previa de adecuación (1). En esta etapa sólo se efectúa un cambio de formato para su tratamiento con Matlab, no realizándose ninguna modificación en sus componentes frecuenciales ni temporales. A continuación la señal de trabajo se introduce en el módulo de filtrado (2). El filtrado se realiza sobre la transformada wavelet de la señal, obteniéndose luego, mediante la transformada inversa, una señal libre de ruido. Esta señal se compara con la original en un módulo de evaluación del filtrado (3). Veremos en detalle cada uno de estas tres etapas, analizando en las secciones siguientes los problemas asociados con la elección del criterio de filtrado y del filtro de espejo en cuadratura (QMF) utilizados.

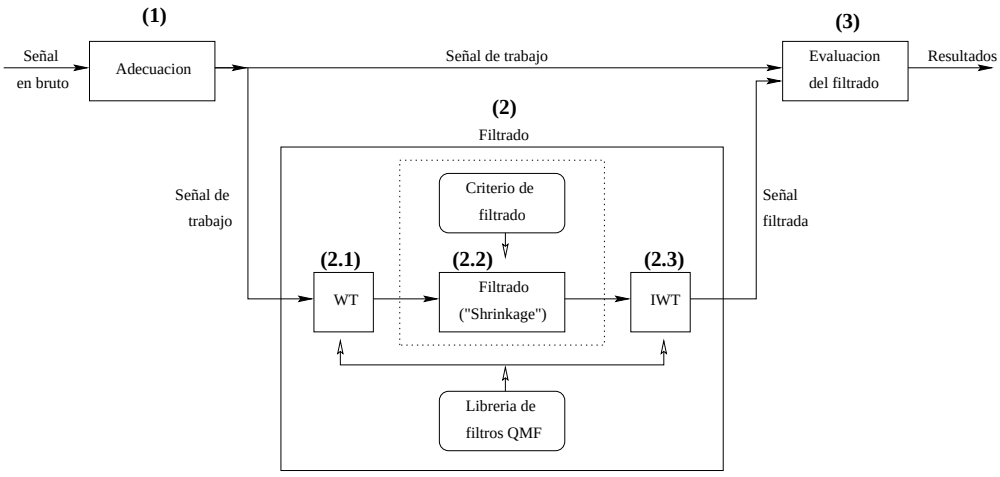

Figura 1: Diagrama de bloques del método de filtrado en el dominio wavelet.

2.1 Adecuación de un conjunto de muestras de voz

Se tomaron archivos con muestras de voz del corpus existente en nuestro Departamento, de unos siete segundos de duración y muestreadas a 8 KHz. Esta frecuencia de muestreo suele ser suficiente para capturar toda la información relevante de la señal de voz en condiciones de habla normales.

2.2 Desarrollo del módulo de filtrado wavelet

Al realizarse el filtrado en el dominio wavelet, se hace necesario disponer de un conjunto de funciones que calculen la transformada wavelet directa e inversa. Para ello hemos utilizado las funciones desarrolladas por el Departamento de Teoría de Señal de la Universidad de Vigo, y presentes en un toolbox de dominio público denominado UviWave ([8]).

Sea una señal v con 2^J componentes. Para calcular la transformada wavelet de dicha señal se hace necesario disponer de un par de filtros QMF de análisis wavelet h (paso bajo) y g (paso alto) ([7]). Con dicho par de filtros puede obtenerse un número k de escalas ($k < J$) de la transformada wavelet de la señal. El vector obtenido tras calcular dicha transformada wavelet -haciendo uso del bloque (2.1) de la figura 1- consta de las siguientes partes:

- Los 2^{J-k} primeros componentes del vector son el residuo obtenido tras aplicar el filtro wavelet iterativamente k veces.
- A continuación aparecen los 2^{J-k-1} componentes de la k-ésima escala, los de la escala $k-1$, $k-2$, etcétera, hasta llegar a los 2^{J-1} componentes de la escala 1 (los obtenidos en la primera iteración del algoritmo).

Como puede verse, el módulo WT no separa el resultado del cálculo de cada escala, sino que devuelve un vector con todas las escalas juntas y el residuo correspondiente. Esto ha hecho necesaria la creación de funciones que permitan el tratamiento de la señal en el dominio wavelet.

La figura 2 es una ampliación de la zona enmarcada en línea de puntos de la figura 1. El filtrado se realiza a través de un módulo que implementa el criterio *hard thresholding* propuesto por Donoho [5]. Dicho criterio consiste en tratar cada escala por separado, colocando a cero los coeficientes de dicha escala menores que un umbral determinado y dejando el resto inalterados.

Como puede verse en la figura, se hace necesario disponer de un vector con los umbrales de filtrado para cada escala. En la sección 3 se discuten los posibles criterios de elección de dicho umbral. Una

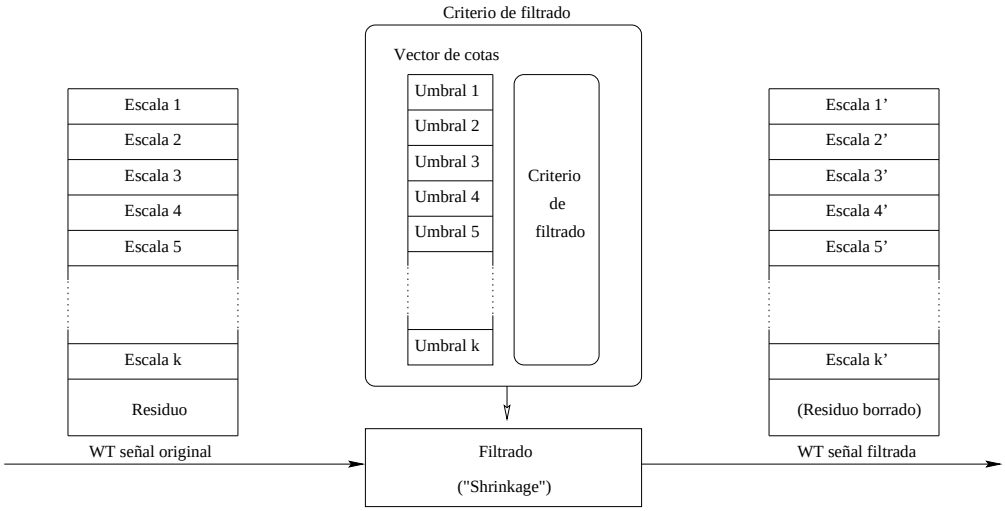

Figura 2: Detalle del módulo de filtrado.

vez hecho esto, el módulo (2.3) de la figura 1 calcula la transformada inversa de la señal, obteniéndose nuevamente una representación de la señal de voz en el dominio del tiempo.

2.3 Evaluación de la calidad del filtrado

Es evidente que para comparar los resultados obtenidos con diferentes criterios de filtrado es necesario disponer de un método de evaluación que permita determinar la calidad del filtrado realizado. Al igual que se hace en la bibliografía consultada, se utilizó en un principio un criterio estrictamente visual. Para ello se desarrollaron un conjunto de funciones que muestran superpuestas ambas señales, así como su diferencia en valor absoluto y sus correspondientes espectrogramas.

Otro criterio que se ha utilizado en la evaluación del filtrado ha sido comparar la energía de la señal original con la de la señal filtrada. Dado que el filtrado elimina componentes wavelet antes de realizar la transformación inversa, la energía de la señal reconstruida será menor o igual que la de la señal original. Para la eliminación del ruido en grabaciones lo que se pretende es que la energía de la señal debida al ruido sea cero, mientras que la energía de la señal de voz deberá ser modificada lo menos posible. Por lo tanto, una forma de verificar el filtrado de la señal es calcular la energía de la señal reconstruida, y su diferencia porcentual respecto de la energía de la señal original. Esta diferencia deberá ser despreciable en los segmentos de voz de la señal (ya que la energía de la señal de voz es mucho mayor que la debida al ruido) y máxima en los períodos de silencio de la misma. Este criterio fue el utilizado en el módulo (3) de evaluación del filtrado (figura 1).

3 Elección del umbral de *thresholding*

Como hemos dicho en la sección anterior, el funcionamiento del proceso de filtrado depende de la elección del umbral de filtrado. Dicho umbral es diferente para cada escala, y en su elección debe tenerse en cuenta las características de las componentes de la señal en esa escala.

Se han ensayado diferentes criterios para la elección del umbral de filtrado. Dichos criterios se expondrán a continuación, examinando el resultado obtenido al aplicar cada uno de ellos. Asimismo, se proponen criterios adicionales, en los que se está trabajando en la actualidad.

3.1 Elección del umbral a partir de la energía

El primer criterio utilizado para la elección de los valores de umbral de filtrado en cada escala es el siguiente: elegir los valores sucesivamente para cada escala a partir de la primera de forma que se maximizara el efecto de reducción del ruido en las zonas de silencio de la señal, y se minimizara en las zonas de la señal correspondientes a las vocales (en donde su energía es máxima). Para ello, utilizamos las funciones desarrolladas para la evaluación de la calidad del filtrado, descritas en la sección 2.3.

La señal elegida para el ensayo es una señal muestreada a 8 kHz, de unos siete segundos de duración, en la cual una voz de hombre pronuncia la frase en inglés "*seven two six eight zero nine*".

El procedimiento de elección de umbrales fue el siguiente:

1. Para la primera banda se ensayó con diferentes valores de umbral. El valor elegido finalmente fue el valor más pequeño que permitía una reducción máxima del ruido de fondo en las zonas de silencio. Para valores más altos que él, el ruido en las zonas de silencio no disminuía, y en cambio comenzaba a verse afectada la energía de la señal de voz en más de un cinco por ciento.
2. Una vez establecido el umbral de filtrado para la primera banda, se calculó la transformada de la señal para las dos primeras bandas y, filtrando la señal en la primera con el umbral antedicho, se buscó el umbral de filtrado más adecuado para la segunda, siguendo el mismo criterio.
3. El proceso se repitió hasta la décima escala. A partir de la décima escala los resultados mejoraban eliminando el residuo en lugar de seguir calculando escalas por debajo (en este caso, las escalas 11 y siguiente se corresponde con la banda de frecuencias por debajo de los 8 Hz).

En la figura 3 puede observarse el fragmento de la señal correspondiente a la palabra "*six*". Una vez realizado el proceso descrito sobre esta señal, el resultado obtenido fue el siguiente:

- Reducción de la energía de la señal en las zonas de silencio: 99,864 por ciento.
- Reducción de la energía de la señal en las zonas correspondiente a consonantes fricativas: 4,404 por ciento.
- Reducción de la energía de la señal en las zonas de vocales: 0,089 por ciento.

En la figura 4 aparece la señal una vez realizado el filtrado. Aunque la mejora visual es evidente, los espectrogramas de ambas señales revelan mejor el comportamiento del filtro (figuras 5 y 6).

3.2 Elección del umbral adaptado a las componentes de la escala

Tras examinar los valores utilizados como umbrales de filtrado de la señal a las diferentes escalas, hemos podido comprobar que en todos los casos el umbral filtraba aproximadamente el 50 por ciento de todos los coeficientes de la escala. Utilizando este criterio de forma estricta -es decir, eligiendo los umbrales de modo que el 50 por ciento de los coeficientes fueran eliminados- los resultados descritos en el apartado anterior mejoraron ligeramente.

El porcentaje de coeficientes a desechar guarda relación con las características de la grabación, ya que si dicha grabación presenta una diferente proporción de silencios respecto de la duración total, el porcentaje de coeficientes a desechar cambia. Además, dicha proporción también cambia si la grabación presenta un alto nivel de ruido, ya que la eliminación del mismo porcentaje de coeficientes lleva a una pérdida de matices apreciable en la señal.

Por lo tanto, este criterio de eliminar exactamente la mitad de los coeficientes no puede extrapolarse directamente a grabaciones realizadas en otras circunstancias. Sin embargo, a la vista de los resultados parece razonable suponer que para obtener un filtrado adecuado debe eliminarse el mismo porcentaje de coeficientes a todas las escalas analizadas. Estamos trabajando en la cuestión de cómo seleccionar dicho porcentaje.

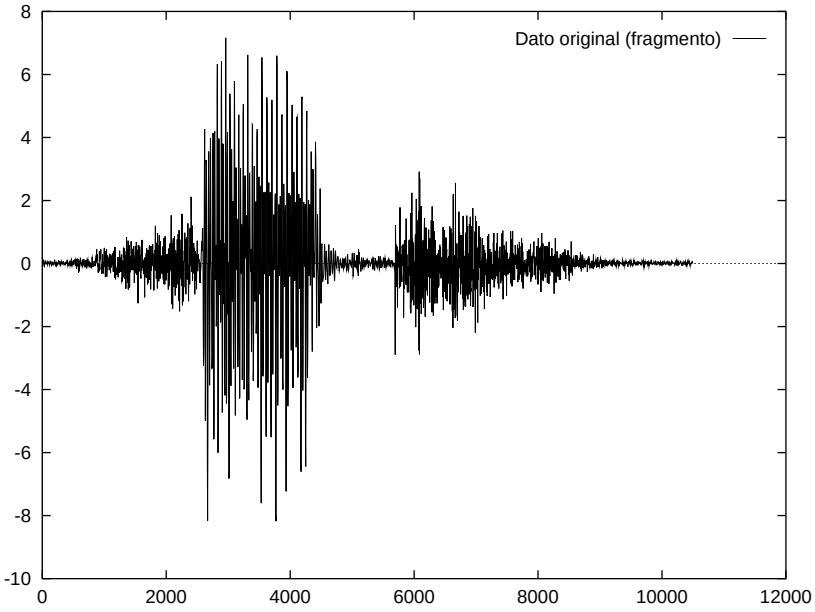

Figura 3: Fragmento de la señal correspondiente a "s*ix*".

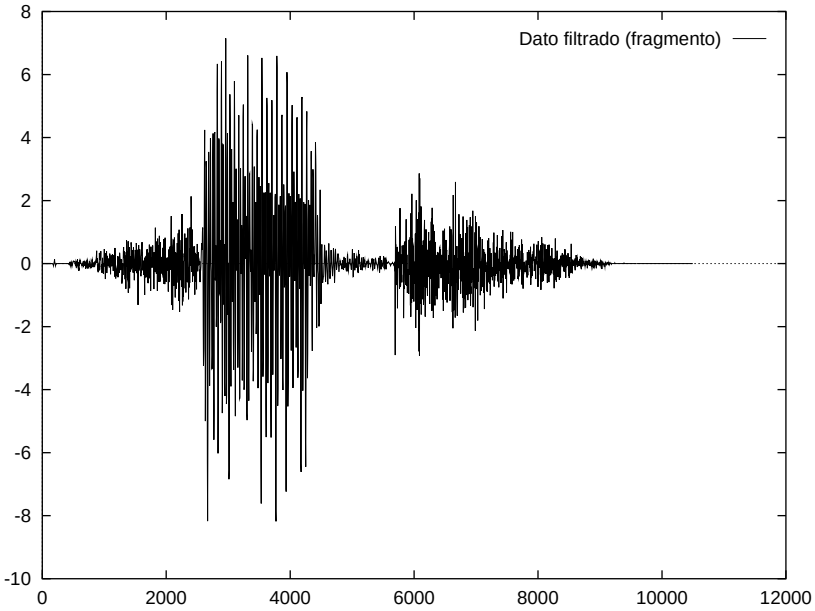

Figura 4: Señal de la figura 3 una vez filtrada.

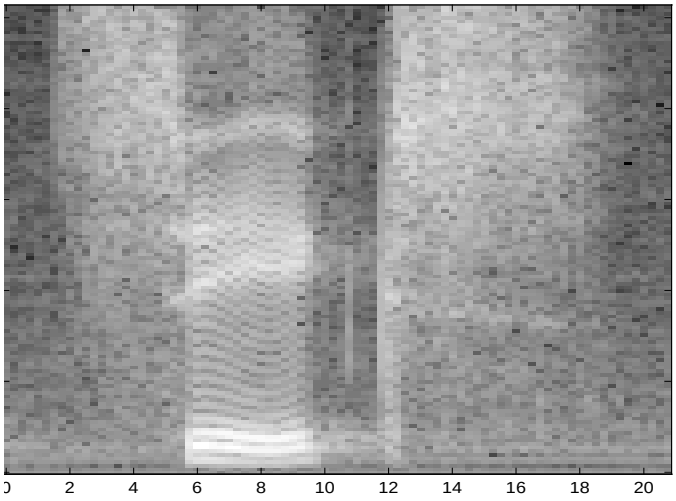

Figura 5: Espectrograma de la palabra “s*ix*” antes del filtrado.

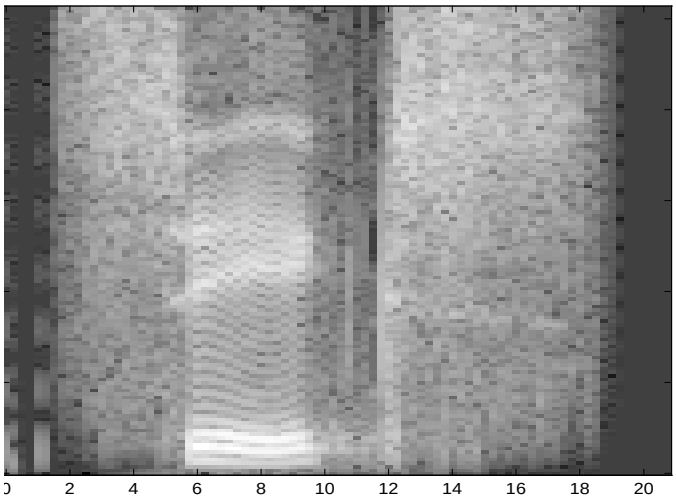

Figura 6: Espectrograma de la palabra “s*ix*” después del filtrado.

Nombre	Dif. vocales	Dif. fricativas	Dif. silencio	Puntuación
Daubechies 16	0.100124	3.900061	100.000000	20
Daubechies 4	0.092005	4.345299	100.000000	19
Filt3	0.089798	4.404413	99.864239	17
Daubechies 18	0.094582	4.534459	100.000000	13
Symmlet 5	0.094685	3.955151	96.668563	10
Symmlet 4	0.094761	4.453808	100.000000	10
Daubechies 14	0.097866	4.348625	100.000000	10
Coiflet 1	0.095324	4.562462	100.000000	10
Symmlet 7	0.095380	4.379250	99.872295	9
Symmlet 6	0.096476	4.479102	99.870660	8

Tabla 1: Comparativa de resultados de filtrado para los diez mejores filtros de entre los 25 analizados. Las diferencias descritas son porcentuales (véase texto).

3.3 Utilización del criterio *soft-thresholding*

Se ha implementado también el criterio *soft-thresholding* (propuesto por Donoho [2]). El vector de umbrales utilizado fue el propuesto en la sección anterior. Sin embargo, hemos podido apreciar el problema siguiente: todas las componentes de la señal sufren una atenuación muy fuerte pese a pertenecer a fragmentos de voz. Sin duda, este resultado puede mejorarse utilizando umbrales menores, pero de todos modos la atenuación de todas las componentes persistirá, por la misma naturaleza del criterio de filtrado. Por lo tanto, este criterio parece menos útil en este caso que el de *hard-thresholding*, utilizado en el apartado anterior.

3.4 Otros criterios de elección de umbral de filtrado

Los criterios propuestos por Donoho, ya comentados en la sección 1, presentan el inconveniente de que requieren conocer previamente el nivel de ruido de la señal. Dicho valor es desconocido en la práctica, por lo que sólo es posible una estimación utilizando técnicas heurísticas. De obtenerse dicha estimación, podría construirse un filtro adaptativo, que filtrara las componentes de la señal en función de las características de ésta. Este filtro sería de gran utilidad como etapa previa en el procesamiento de señal de voz, por lo que estamos trabajando actualmente en el desarrollo de dicha heurística.

4 Elección del filtro QMF a utilizar

Para comparar los efectos del uso de diferentes filtros QMF sobre de la señal, se efectuaron pruebas de filtrado con diferentes filtros propuestos en la bibliografía.

Se examinaron un total de venticinco filtros: el filtro Haar, el filtro Beylkin, la familia de 5 filtros Coiflet, la familia de 9 filtros Daubechies, la familia de 7 filtros Symmlet y el filtro Vaidyanathan ([10]), además de un filtro biortogonal (“Filt3” de Uviwave) ([8]).

Como criterio de evaluación se ha utilizado el descrito más arriba, esto es, una reducción máxima del ruido y una modificación mínima de la señal hablada. Para este segundo caso se distinguió entre la modificación de la señal correspondiente a vocales y la modificación de la señal correspondiente a sonidos fricativos. La comparativa se realizó como sigue:

- Se realizó el filtrado de una señal utilizando cada uno de los filtros mencionados anteriormente.

- Se utilizó el módulo de evaluación de filtrado para obtener los porcentajes de modificación de la señal filtrada respecto de la señal original, en las tres categorías indicadas (silencio, fricativas y vocales).
- Se ordenaron los filtros según la reducción de la energía en las zonas de silencios en orden decreciente, otorgándose diez puntos al que presentaba una reducción mayor, nueve al siguiente, etcétera.
- Se ordenaron nuevamente los filtros según la reducción de energía en las fricativas, en orden creciente. Se otorgó diez puntos al que menor reducción de la señal presentaba, nueve al siguiente, etcétera.
- Se ordenaron los filtros en función de la reducción de la energía en las vocales, en orden creciente, volviéndose a asignar diez puntos al de menor reducción, y así sucesivamente.
- Finalmente, se sumaron las puntuaciones obtenidas para cada filtro.

Una vez realizado los cálculos con el vector de cotas propuesto en 3.2, el resultado final es el que aparece en la tabla 1. En ella puede verse que dos filtros de la familia de Daubechies son los que permiten unos mejores resultados. En concreto, consideramos que el filtro "Daubechies 16" es el mejor de los filtros analizados, al eliminar al completo el ruido modificando la energía de la señal para los sonidos fricativos en menos de un cuatro por ciento.

5 Conclusiones

La aplicación de las técnicas vistas al filtrado de señales de voz presenta notables ventajas respecto al filtrado paso bajo convencional, ya que permite eliminar componentes de alta frecuencia de la señal debidos al ruido sin suavizar el resto de detalles de la señal.

Un aspecto importante es el de la elección de los umbrales de filtrado. Como se ha visto, existen diferentes criterios, que podemos clasificar en dos tipos: los que utilizan técnicas heurísticas (por ejemplo, que elimine la energía de los silencios de la grabación), y otros que para su aplicación necesitan conocer el nivel de ruido de la señal, que es precisamente un dato desconocido. Estamos trabajando en la cuestión de la elección automática del umbral de filtrado para cada escala en función de las características de la señal, lo que posibilitaría la obtención de un filtro adaptativo, de gran utilidad como etapa de preproceso de la señal de voz.

La calidad del filtrado realizado depende, como hemos visto, de una correcta elección del filtro QMF que realiza el paso al dominio wavelet de la señal a filtrar. Hemos establecido una comparativa utilizando criterios fácilmente reproducibles, lo que nos ha permitido seleccionar el filtro mejor adaptado a los criterios establecidos de calidad de filtrado.

Entre las posibles aplicaciones de esta técnica figuran su uso como etapa previa a la aplicación sobre la señal de algoritmos de detección de pitch ([6]).

Referencias

[1] COHEN, A., DAUBECHIES, I., JAWERTH, B., AND VIAL, P. Multiresolution analysis, wavelets, and fast algorithms on an interval. Tech. rep., Comptes Rendus Acad. Sci. París, 1992.

[2] DONOHO, D. L. De-noising via soft-thresholding. *Tech. Report, Statistics, Stanford* (1992).

[3] DONOHO, D. L. Wavelet shrinkage and wavelet-vaguelette decomposition: A 10-minute tour. *International Conference on Wavelets and Applications* (June 1992).

[4] DONOHO, D. L. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noise data. *Proceedings of Symposium in Applied Mathematics 00* (1993).

[5] DONOHO, D. L., AND JOHNSTONE, I. M. Ideal space adaptation via wavelet shrinkage. *Tech. Report, Statistics, Stanford* (1992).

[6] ESCUDERO, D., AND CARDEÑOSO, V. Procesamiento de curvas de frecuencia fundamental. Tech. rep., Departamento de Informática, Universidad de Valladolid, 1998.

[7] RIOUL, O., AND VETTERLI, M. Wavelets and signal processing. *IEEE SP Magazine* (Octubre 1991), 14–38.

[8] S. GONZÁLEZ SÁNCHEZ, N. GONZÁLEZ PRELCIC, S. J. G. G. Uviwave version 3.0 (wavelet toolbox for matlab). *Grupo de Teoría de la Señal, Universidad de Vigo* (1995).

[9] SILVA-VARELA, H., AND CARDEÑOSO, V. Phonetic classification in spanish using especialized neural nets hierarchically organized. Tech. rep., Departamento de Informática, Universidad de Valladolid, 1998.

[10] WICKERHAUSER, M. V. *Adapted Wavelet Analysis - From Theory to Software*, 1 ed. IEEE Press, 1994.

[11] WORKS, T. M. *Matlab versión 4 - Guía del Usuario*, 1 ed. Prentice-Hall, 1996.