

# CONFORMADO DE HAZ PARA VOZ CON WAVELETS

Antonio Satué Villar

Dep. Telecom. y Arquít. De Computadores  
Escuela Universitaria Politécnica de Mataró  
satue@eupmt.es

Juan Fernández Rubio

Dep. Teoría de la Señal y Comunicaciones  
Esc. Téc. Sup. Ingenieros Telecom. Barcelona  
juan@gps.tsc.upc.es

## RESUMEN

En este artículo se estudian las ventajas de realizar una transformación de los datos en un conformador de banda ancha utilizando transformada wavelet. Con este objetivo, se simulan unas señales directas e interferentes que inciden en una agrupación de micrófonos en un entorno ruidoso. Las señales directa (deseada) e interferente se obtienen de una base de datos de voz. Las simulaciones realizadas muestran que la transformada wavelet reduce el tiempo de convergencia. En el artículo también se comprueba que la transformada wavelet no afecta a la calidad de la señal de salida (evaluada con un método de reconocimiento de locutor).

## 1. INTRODUCCIÓN

Si se desean adquirir señales de voz con un micrófono en un entorno ruidoso o con interferencias, el diagrama de recepción de nuestro micrófono debe tener un máximo en la dirección de la señal deseada y debe minimizar las interferencias. Para resolver este problema, se pueden agrupar varios micrófonos y combinar sus diagramas mediante unas ponderaciones (pesos) para obtener las características deseadas.

Si la señal de entrada fuera estacionaria, los pesos podrían ser fijos. Como en la práctica esto no es así, los pesos deben adaptarse basándose en las variaciones de la señal de entrada. Es decir, algoritmos adaptativos. Estos algoritmos son una aproximación de la solución óptima de Wiener; cada solución depende de la dispersión de los autovalores de la entrada. La transformada wavelet define bandas no constantes (en la FFT las bandas son de la misma amplitud) y esto es más apropiado para trabajar con señales de voz (se trabaja con octavas, que son bandas de ancho no constante). Así, convertimos un problema complejo de conformado de haz en problemas menos complejos con señales de banda estrecha. Además, esta descomposición es adecuada para diferentes coeficientes de adaptación ( $\mu$ ).

Este artículo está estructurado de la siguiente forma. En la siguiente sección se revisa la estructura GSC y la transformada wavelet. En la sección 3 se presentan algunos experimentos basados en simulaciones. Finalizamos el artículo con algunas conclusiones acerca del resultado de los experimentos.

## 2. ESTRUCTURA WAVELET-GSC

En el artículo partimos de la estructura GSC, propuesta por Griffiths y Jim [1]. Tiene dos ramas:

- Una rama superior, llamada quiescente, con un vector de pesos  $w_q$  que hace uso de las restricciones.
- Una rama inferior, que consta de una matriz de bloqueo (B) que no permite el paso de la señal deseada, seguida por un conformador sin restricciones (pesos  $w_a$ ).

La rama inferior solo procesa las señales interferentes, las cuales se restan de la rama superior. De esta forma, la señal de salida queda libre de interferencias.

Chen y Fang [2] aplicaron el algoritmo LMS en el dominio frecuencial al conformador de Griffiths-Jim, y obtuvieron una mayor velocidad de convergencia. En este artículo partimos de una estructura que realiza la transformada wavelet (DWT) de los datos (dominio wavelet). De esta forma, se inserta un bloque DWT entre la matriz B y los pesos  $w_a$ , tal como se muestra en la figura 1.

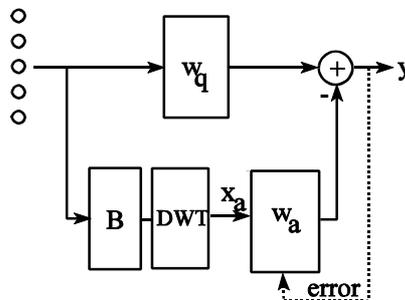


Figura 1. Conformador Wavelet-GSC

Hay distintas familias de transformada wavelet, cada una con sus propiedades. En este artículo utilizaremos la Daubechies 2 (db2). Una wavelet está caracterizada por 2 filtros (un paso-alto y un paso-bajo). Los coeficientes son las salidas pares de estos filtros. Podemos definir transformadas de nivel 2 transformando por separado los coeficientes paso-alto y paso-bajo, etc. Para más detalles, consultar las referencias [3] y [4]. En la figura 2 podemos ver los coeficientes que resultan de una transformación de nivel 2 que solo itera los coeficientes paso-bajo.

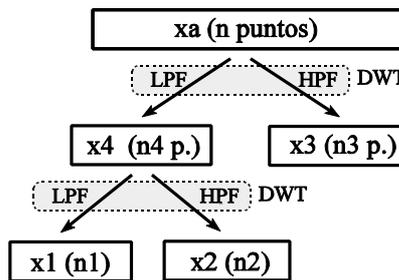


Figura 2. DWT de 2 niveles

Con la descomposición anterior y una estructura como la de la figura 3, podemos adaptar cada tipo de pesos ( $w_1, w_2, w_3$ ) con un algoritmo LMS con un valor distinto de  $\mu$ , acelerando así la convergencia.

$$w_i(n+1) = w_i(n) + \mu_i \cdot y^*(n) \cdot x_i(n) \quad i=1:3 \quad (1)$$

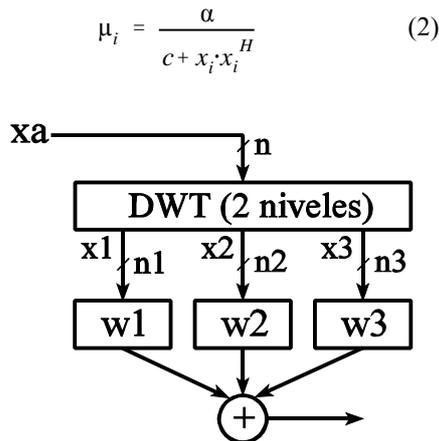


Figura 3. Ponderación de los coeficientes DWT

### 3. EXPERIENCIAS

En [5] se presentó una estructura mono-sensor que adaptaba pesos con transformada wavelet, pero con señal de entrada sintética. Vamos a incorporar esta estructura a un conformador GSC que trate con señales de voz, como vemos en la figura 4.

El escenario en las simulaciones es:

- ne=15 micrófonos equiespaciados media longitud de onda (frecuencia central = 2 KHz).
- nt=19 taps por filtro
- ruido gaussiano de media 0 y varianza 0.1
- señales incidentes de voz de 45000 muestras ( $f_{\text{muestreo}} = 8 \text{ KHz}$ ), incidiendo a  $0^\circ$  (deseada) y  $-30^\circ$  (interferente) con potencia media igual a 10000 (40 db).

Nosotros queremos ver que utilizando DWT no se degrada apreciablemente la calidad de la señal de salida. Para ello, usaremos un método simple de reconocimiento de locutor. Se evalúa la distancia de esfericidad aritmético-armónica [6]

$$\mu(C_j, C_{\text{test}}) = \log [tr(C_{\text{test}} \cdot C_j^{-1}) \cdot tr(C_j \cdot C_{\text{test}}^{-1})] - 2 \cdot \log(m) \quad (3)$$

donde  $C_{\text{test}}$  es la matriz de covarianza de la salida del array,  $C_j$  es la matriz de covarianza del usuario a testear y  $m$  la dimensión de las matrices (20 en todos los experimentos).

Se define un factor de mérito como

$$\text{Factor de Merito} = (\mu_{IS} - \mu_{DS}) - (\mu_{IE} - \mu_{DE}) \quad (4)$$

donde

- $\mu_{DE} = \mu(C_j \text{ deseada, entrada del array})$
- $\mu_{IE} = \mu(C_j \text{ interferente, entrada del array})$
- $\mu_{DS} = \mu(C_j \text{ deseada, salida del array})$
- $\mu_{IS} = \mu(C_j \text{ interferente, salida del array})$

Un Factor de Mérito igual a 0 significa que entrada y salida son iguales (el array no funciona correctamente).

En todas las simulaciones realizadas, este factor de mérito ha sido el mismo usando DWT que sin usarla.

### 4. CONCLUSIONES

Después de realizar diversas simulaciones, se observa que al introducir la DWT, la convergencia del algoritmo es más

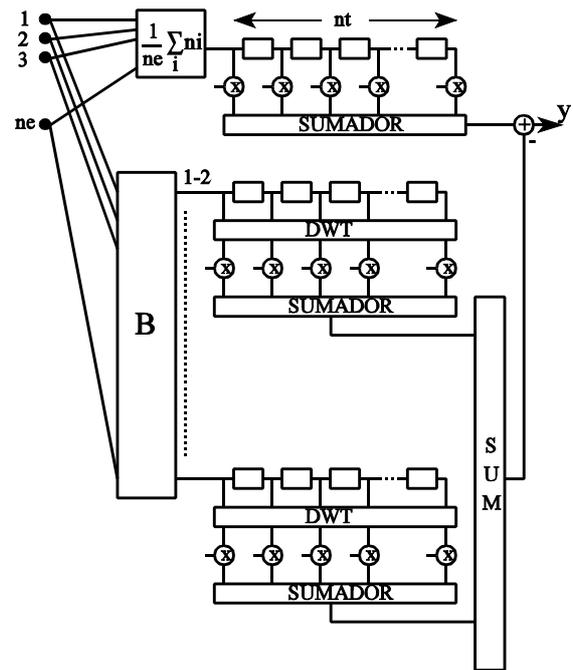


Figura 4. GSC con DWT

rápida y que la calidad subjetiva de la señal de salida evaluada por un sistema de reconocimiento de locutor es casi la misma, independientemente del número de niveles wavelet de descomposición (se ha probado con 1, 2 y 3 niveles).

También se ha visto que eliminando alguno de los coeficientes wavelet, la degradación es pequeña (más pequeña que cuando eliminamos taps sin aplicar la DWT).

### AGRADECIMIENTOS

Este trabajo ha sido realizado con financiación del proyecto CICYT TIC2000-1669-C04-02.

### 5. REFERENCIAS

- [1] Griffiths, Jim, "An alternative approach to linearly constrained adaptive beamforming", *IEEE Trans. Antenna Prop.* vol AP.30 núm. 1 pp 27-34, aug. 1982
- [2] Chen, Fang, "Frequency-domain implementation of Griffiths-Jim adaptive beamformer", *Journal of Acoust. Soc. America* vol.91 n.6 pp.3354-3366 jun 92
- [3] Daubechies, *Ten lectures on wavelets*, CBMS-NSF Society for Industrial and Applied Mathe, 1992
- [4] Rioul, Vetterli, "Wavelets and signal processing", *IEEE Signal Processing Magazine*, pp.14-38, oct 91
- [5] Petraglia, Mariane, Torres, Julio C. B., "Performance Analysis of an Adaptive Filter Structure Employing Wavelets and Sparse Subfilters", *X European Signal Processing Conference (Eusipco 2000)*
- [6] Bimbot F., Mathan L., "Text-free speaker recognition using an arithmetic-harmonic sphericity measure" pp.169-172, *Eurospeech 1991*
- [7] Strobel, Norbert, Rabenstein, Rudolf, "Robust Speaker Localization Using a Microphone Array", *X European Signal Processing Conference (Eusipco 2000)*