

Determinación de la relación señal a ruido de la voz utilizando la transformada de wavelet

Carlos Jiménez⁽¹⁾, José A. Díaz⁽¹⁾, Rahul Shrivastav⁽²⁾, Howard Rothman⁽²⁾, Paulino Del Pino⁽¹⁾

⁽¹⁾Escuela de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de Carabobo, Valencia, Venezuela

⁽²⁾Department of Electrical and Computer Engineering, University of Florida

Email: jadiaz@uc.edu.ve

Resumen

En este trabajo se desarrollaron e implementaron los algoritmos para el cálculo de la relación señal a ruido de señales de voz aplicando la transformada de Wavelet de Daubechies (DB), Symmlet (Sym) y Coiflet(Coif), con los niveles de descomposición de 3, 5, 10 y 15. Se utilizaron señales de prueba con una relación señal a ruido conocida para comprobar el funcionamiento de la metodología aplicada. Igualmente se grabaron señales de voz y se realizó el proceso de filtrado con las diferentes transformadas, comprobándose que con los niveles de descomposición de 10 y 15 se obtiene la mayor eliminación del ruido de la señal en los intervalos de silencio y en las palabras, lográndose escuchar la información filtrada sin deformación y sin el ruido de fondo. Con estos niveles se encontraron los menores valores de la relación señal a ruido en las zonas correspondientes a las pausas de la frase grabada.

Palabras clave: Wavelets, relación señal a ruido, multiresolución, señal estacionaria, procesamiento de señales.

Determination of the signal to noise ratio of the voice using wavelet transform

Abstract

The main purpose of this research was to develop and implement the algorithms to calculate the signal-to-noise ratio of a voice signal using the Wavelet Transform. The Wavelet Transform of Daubechies (DB), Symmlet (Sym) and Coiflet (Coif) with 3,5,10 and 15 levels of decomposition were applied. In order to test the developed algorithms, a signal with a known signal-to-noise ratio was used. Voice signals were also recorded and denoising was made using the mentioned transform. It was verified that the maximum denoising of the original signal was accomplished with 10 and 15 levels of decomposition, resulting in a processed signal without distortion and no background noise. The lowest signal-to-noise ratio values during the unvoiced segments of the recorded sentence were found using these values.

Keywords: Wavelets, signal-to-noise ratio, multiresolution, stationary signal, signal processing.

1. INTRODUCCIÓN

La relación señal a ruido es un indicador de la calidad de una señal y se puede utilizar para caracterizar la señal de voz, para su identificación en presencia de algún tipo de ruido ocasionado durante la transmisión de voz por un medio electrónico, o cuando se graba en presencia de conversaciones simultáneas. Igualmente permite comparar diferentes tipos de voces y caracterizar las patologías.

Como la señal de voz presenta un comportamiento no estacionario, debido a que está compuesta

de partes periódicas que corresponden a las vocales, y partes aleatorias que corresponden a las consonantes, su análisis debe realizarse a través de la Transformada de Wavelet para su filtrado, y luego calcular la relación señal a ruido.

En este proyecto se utilizaron las Transformadas de Daubechies (DB), Symmlet (Sym) y Coiflet (Coif), y los niveles de descomposición 3, 5, 10 y 15, para ser aplicadas a señales de voz y obtener su relación señal a ruido.

2. PLANTEAMIENTO DEL PROBLEMA

El procesamiento digital de la señal de voz se efectúa para caracterizarla según indicadores como punto de inflexión, espectro de frecuencia, relación señal a ruido, etc. La relación señal a ruido permite realizar la comparación de la amplitud de la señal y la varianza del ruido. Se trata de un indicador de la calidad de una señal, y se puede utilizar para caracterizar la señal de voz, para su identificación en presencia de algún tipo de ruido ocasionado durante la transmisión de voz por un medio electrónico, o cuando se graba en presencia de conversaciones simultáneas. Igualmente con la relación señal a ruido se podría identificar y diferenciar las letras vocales y consonantes en una palabra, comparar diferentes tipos de voz y caracterizar diferentes patologías.

Para obtener la relación señal a ruido se deben aplicar algunas de las técnicas utilizadas para realizar el procesamiento digital de la señal de voz, siendo la más común la Transformada de Fourier, con la cual se obtiene el espectro de frecuencia de la señal [1]. Cuando la señal a procesar presenta cambios de frecuencia en un rango de tiempo determinado (señal no estacionaria), la Transformada de Fourier indica el contenido frecuencial, pero no indica dónde ocurren las variaciones de frecuencia.

Como la señal de voz presenta un comportamiento no estacionario, ya que está compuesta de dos partes, una periódica correspondiente a las vocales, y otra aleatoria, correspondiente a las consonantes [2,3], se debe aplicar otra técnica de procesamiento que permita determinar totalmente las características de la señal de voz y su comportamiento en el tiempo. Esto se puede lograr analizando la señal en varios intervalos de tiempo, y definiendo en estos intervalos un espectrograma tiempo-frecuencia. Algunas técnicas utilizadas son la STFT (Short-time Fourier Transform) y la Distribución de Wigner.

Las técnicas anteriores suponen que las señales son estacionarias en un intervalo de tiempo determinado, y el análisis se realiza con una resolución de tiempo-frecuencia fija, por lo que una mejora de resolución de cualquiera de ellos está limitada por el principio de incertidumbre de Heisenberg, es decir, una mejor resolución en dominio del tiempo,

reduce la información en el dominio de la frecuencia, y viceversa.

Debido a que la Transformada Wavelet [4-6] permite analizar la señal en el tiempo y en la frecuencia simultáneamente, y como el segmento de la señal de voz que se analizará es corto, entonces se aplicará esta transformada para analizar la voz y obtener su relación señal a ruido.

3. METODOLOGÍA

El procedimiento metodológico utilizado para el desarrollo de la investigación comprende cinco fases, en las cuales se implementó el programa para procesar las señales de voz y obtener su relación señal a ruido.

3.1. Fase 1: Relación Señal a Ruido (SNR)

En esta fase se determinó la relación señal a ruido (SNR) de una señal de voz. Para lo cual se utilizó la ecuación 1.

$$snr = 10 * \log_{10} \left[\frac{\sum_{n=0}^{L-1} (x(n))^2}{\sum_{n=0}^{L-1} (\hat{x}(n) - x(n))^2} \right] \quad (1)$$

donde:

$\hat{x}(n)$: señal filtrada

$x(n)$: señal con ruido

snr : relación señal a ruido en decibeles

L : número de muestras de la señal

El procedimiento aplicado para obtener la relación consistió en utilizar una señal x a la cual se le añadió ruido blanco con una relación señal a ruido determinada para generar la señal con ruido xn

$$xn = x + wn \quad (2)$$

donde:

xn : señal con ruido gaussiano blanco, que se genera con una relación señal a ruido $snrat$.

wn : Ruido gaussiano blanco con desviación estándar igual a uno y valor medio cero.

x : Señal re-escalada de tal modo que su desviación estándar (srt_d) sea igual al valor de la relación señal a ruido deseada, es decir, $snrat$.

$$x = \frac{x (snrat)}{std(x)} \quad (3)$$

Donde:

snrat: es la raíz cuadrada del valor de la relación señal a ruido deseada

En la Figura 1(a) se muestra una señal senoidal con 1024 muestras, a la cual se le suma un ruido gaussiano con una relación señal a ruido de 10 dB mediante la ecuación 2 (ver Figura 1b). Cuando se le aplicó la ecuación 3 a estas señales, se obtuvo una relación señal a ruido de 10,27 dB, el cual es un valor que se encuentra alrededor del previsto en esta señal (10 dB), por lo que se valida la fórmula utilizada.

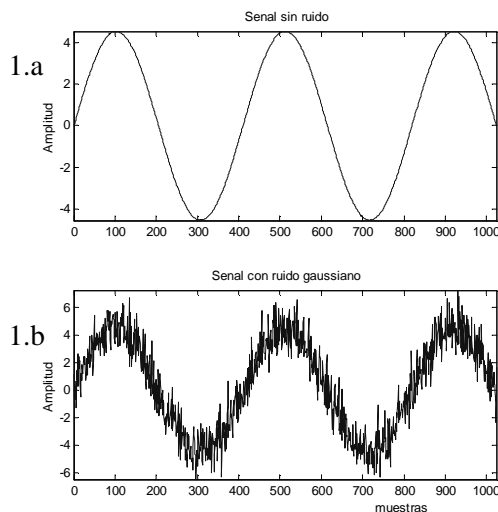


Figura 1. a) Señal sin ruido. b) Señal con ruido.

3.2. Fase 2: Cálculo de la SNR por segmento

En esta fase se implementó el programa para realizar el cálculo de la relación señal a ruido (SNR) segmentando la señal, sin solapamiento y con solapamiento de estos segmentos. Para realizar la segmentación de la señal se deben definir el tamaño del segmento que se va a procesar (N), y el tamaño del segmento que se va a solapar (S). El desplazamiento del segmento se obtiene mediante la diferencia entre N y S. La ubicación (P) del valor correspondiente a la relación señal a ruido calculado en el segmento, se ubica en la mitad del tamaño del segmento procesado.

Un ejemplo del cálculo del segmento que se utiliza para determinar la relación señal a ruido sin

solapamiento es el siguiente:

- Longitud de la señal L = 1024 muestras
- Tamaño del segmento N = 256 muestras, 4 ventanas en total
- Tamaño del solapamiento S = 0

En la Figura 2 se muestra el resultado del procedimiento para obtener el segmento de la señal procesada. Los puntos P1, P2, P3 y P4 representan las posiciones donde se va a colocar el valor de la relación señal a ruido.

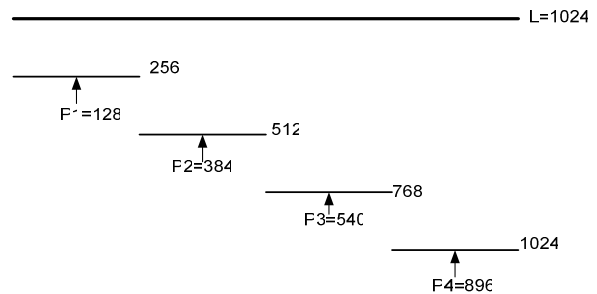


Figura 2. Segmentación de la señal sin solapamiento.

Un ejemplo del cálculo del segmento que se utiliza para determinar la relación señal a ruido con solapamiento es el siguiente:

- Longitud de la señal L = 1024 muestras
- Tamaño de la ventana N = 256 muestras, 4 ventanas en total
- Tamaño del solapamiento S = 128 muestras.
- Desplazamiento de la ventana = N - S = 256 - 128 = 128 muestras

En la Figura 3 se muestra el resultado del procedimiento utilizado para obtener el segmento de la señal que se va a procesar, los puntos P1, P2, P3, P4, P5, P6 y P7 representan las posiciones donde se coloca el valor de la relación señal a ruido.

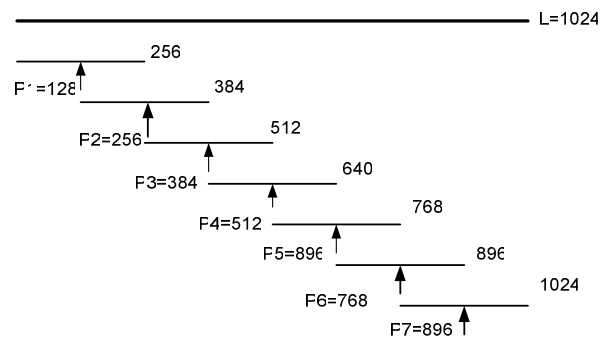


Figura 3. Segmentación de la señal con solapamiento.

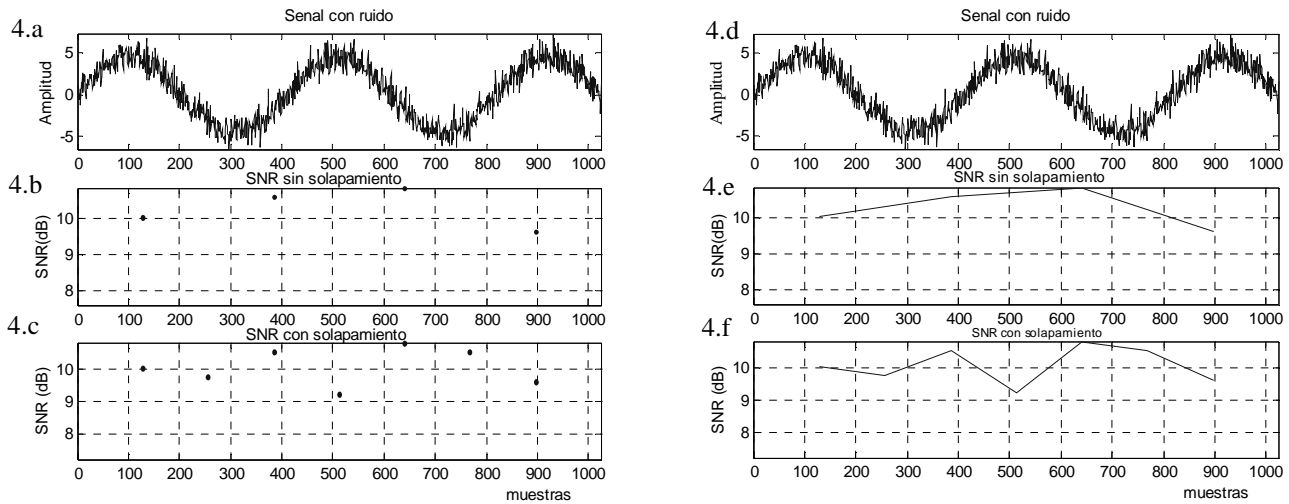


Figura 4. (a) Señal con ruido. (b) SNR sin solapamiento (c) SNR con solapamiento (d) Señal con ruido (e) SNR sin solapamiento con interpolación (f) SNR con solapamiento, con interpolación.

En la Figura 4 se muestra el cálculo de la relación señal a ruido de una señal senoidal (1024 muestras) con ruido blanco gaussiano y una relación SNR=10 dB. El resultado de la relación señal a ruido sin solapamiento se muestra en la Figura 4 (b), y con solapamiento en la Figura 4(c). En las Figuras 4(e) y 4(f) los valores correspondientes a la relación señal a ruido se unieron a través de una línea recta. Se observa que los valores de SNR obtenidos oscilan alrededor de 10 dB en cada segmento donde se realiza su cálculo, con un valor máximo de 11 dB y un valor mínimo de 9,2 dB, con lo cual se demuestra la validación del procedimiento.

3.3. Fase 3: Aplicación de la Transformada de Wavelet a una señal con ruido

En esta fase se aplicó la Transformada de Wavelet a una señal con un determinado nivel de ruido para realizar su filtrado [8-10]. Se implementó un programa para reducir el ruido de una señal con una relación señal a ruido conocida, donde el valor del umbral de ruido de la señal es calculado según la Teoría de Donoho [11-13] mediante la expresión indicada en la ecuación 4

$$THR = (\sqrt{2 \ln(n)}) \times s \quad (4)$$

Donde:

- n: es la longitud de la señal a analizar
- s: es un estimado del nivel de ruido de la señal, obtenido con la ecuación 5

$$s = \frac{\text{mediana}(\text{modulo}(c))}{0.6745} \quad (5)$$

c = Son los coeficientes de los detalles en el nivel uno

El tipo de umbral a aplicar a la señal para reducir el nivel de ruido puede ser: duro (Hard) o suave (Soft). Cuando se aplica el umbral duro, si el valor absoluto del coeficiente es mayor que el umbral calculado, se mantiene el coeficiente, y en el caso contrario se iguala a cero. En la ecuación 6 se muestra la expresión utilizada para aplicar este tipo de umbral, donde $C_m(i, j)$ representan los coeficientes modificados. Cuando se aplica el umbral suave, si el valor absoluto del coeficiente es mayor que el umbral seleccionado se modifica el coeficiente, restando el umbral a su valor absoluto, en caso contrario se iguala el coeficiente a cero. En la ecuación 7 se muestra la expresión utilizada cuando se aplica este tipo de umbral. En la figura 5 se muestra el resultado cuando no se modifican los coeficientes, por lo tanto $C_m(i, j)$ es igual a $C(i, j)$. En las Figuras 6(a) y 6(b) se muestran los resultados de las modificaciones de los coeficientes cuando se aplica el umbral duro y el suave.

$$C_m(i, j) = \begin{cases} 0 & \text{si } |C(i, j)| < thr \\ C(i, j) & \text{si } |C(i, j)| \geq thr \end{cases} \quad (6)$$

$$C_m(i, j) = \begin{cases} 0 & \text{si } |C(i, j)| < thr \\ \text{sgn}(C(i, j))|C(i, j) - thr| & \text{si } |C(i, j)| \geq thr \end{cases} \quad (7)$$

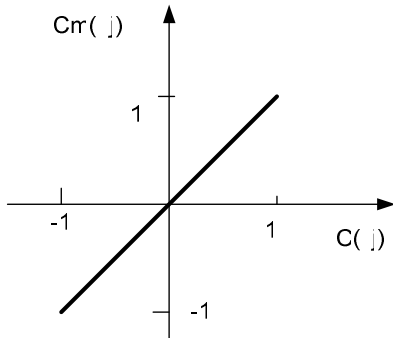


Figura 5. Señal original.

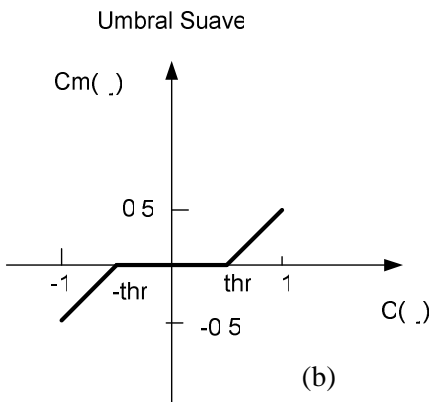
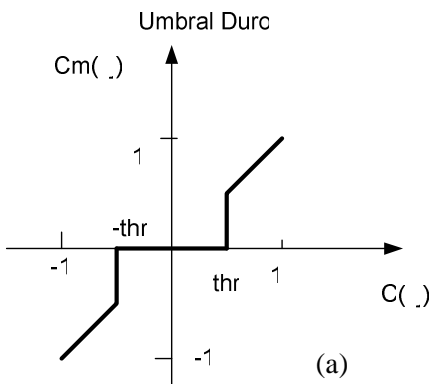


Figura 6. a) Umbral duro. b) Umbral suave.

Se implementó el programa para eliminar el ruido de una señal con una relación señal a ruido conocida. Se utilizó una señal senoidal con un ruido gaussiano con una relación snr igual a 10 dB, y para realizar el proceso de filtrado de la señal se seleccionó

la Wavelet DB5, con un umbral suave igual a 4,0166.

En la Figura 7 se muestra el resultado de este experimento, en donde se observan el ruido extraído (Figura 7b) y la señal filtrada (Figura 7c), la cual conserva su forma senoidal y se le ha eliminado parte del ruido que presentaba la señal original. Se calculó la relación señal a ruido, con la señal original y el ruido extraído y se obtuvo el valor de 10,3362 dB, el cual está alrededor de 10 dB, demostrándose el efecto que tiene la Transformada de Wavelet en el proceso de filtrado de una señal.

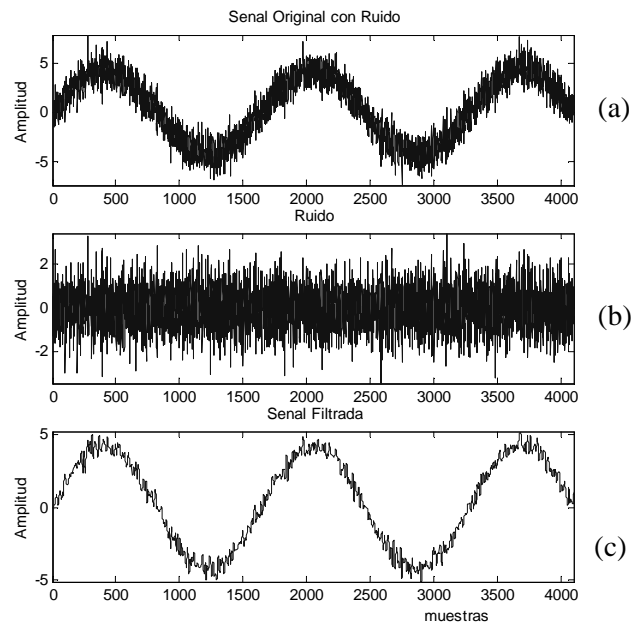


Figura 7. a) Señal con ruido. b) El ruido removido. c) Señal procesada.

3.4. Fase 4: Aplicación de wavelet en la medición de SNR utilizando solapamiento

En esta fase se aplicó la Transformada de Wavelet en la medición de SNR utilizando solapamiento. Se utilizó una señal senoidal con una relación señal a ruido de 10 dB. Para filtrar la señal se utilizó la Transformada de Wavelet DB5, nivel 3, utilizando un umbral suave. En la Figura 8 se muestra el resultado de este experimento. Los valores de la relación señal a ruido obtenidos oscilan entre 10 y 11.5 dB. Se repitió el experimento con una señal senoidal y una relación señal a ruido de 20 dB, obteniéndose los resultados mostrados en la a Figura 8(b), donde los valores de la relación señal a ruido oscilan entre 19 y 22 dB.

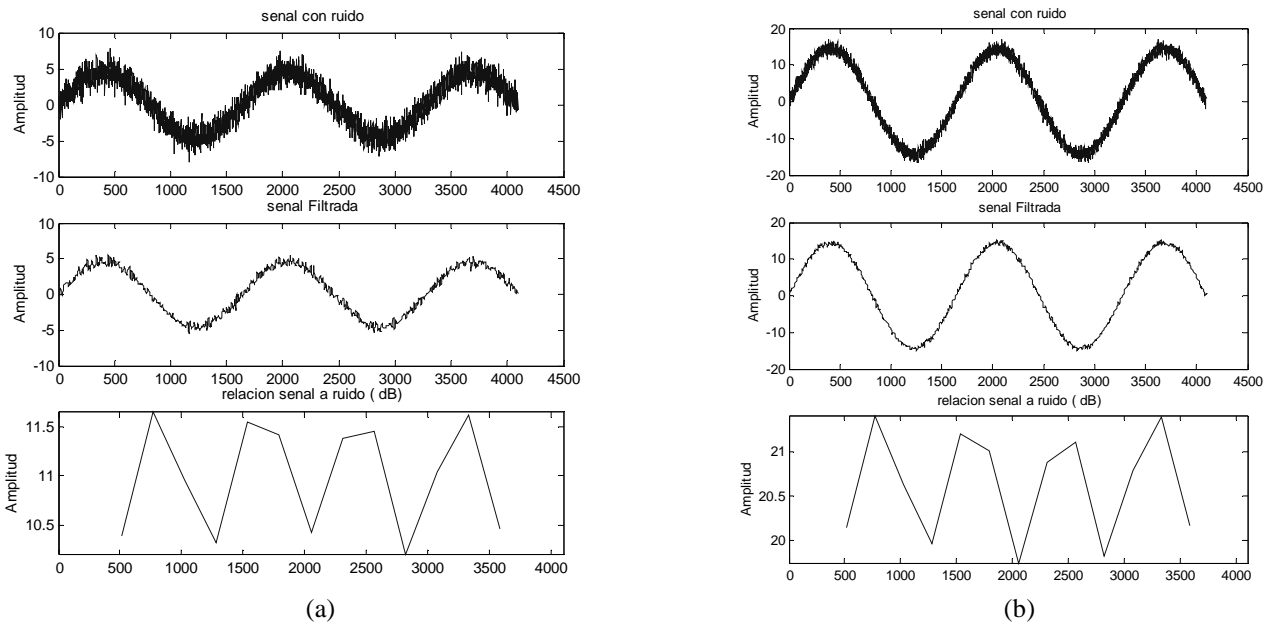


Figura 8. a) Cálculo de snr, con solapamiento, de una señal con SNR de 10 dB. b) Cálculo de snr, con solapamiento, de una señal con SNR de 20 dB.

En estos experimentos se observa cómo las ondas procesadas con la Transformada Wavelet conservan su forma senoidal, sin distorsión, y se le filtra parte del ruido que presentan, obteniéndose con la metodología aplicada una relación señal a ruido en cada segmento procesado alrededor del valor que se le asignó, logrando con esto demostrar la validez del procedimiento aplicado.

3.5. Fase 5: Aplicación de Wavelet para la determinación de la relación señal a ruido de una señal de voz

En esta fase se aplicó la Transformada de Wavelet para eliminar el ruido de una señal de voz y obtener la SNR utilizando solapamiento (ver Figura 9). Se implementó un programa que permite seleccionar el tipo de Transformada Wavelet, el nivel de descomposición, el umbral del ruido, la longitud de la ventana a procesar, y el tamaño del solapamiento.

En la Figura 9(a) se muestra la señal de voz de un niño de 8 años, correspondiente a la frase: “tengo ocho años”. Esta frase se grabó en formato .wav, a una frecuencia de 22050 muestras por segundo y tiene una longitud de 53020 muestras (2,4025 segundos). El procesamiento se realizó con la Transformada de Wavelet DB5 y un nivel de descomposición de 10, el umbral

utilizado es de 0,04362 obtenido a través de la relación de Donoho, la longitud de la ventana es de 30 ms, y se hizo sin solapamiento. En la Figura 9(b), 9(c) se muestra la señal filtrada y el ruido extraído respectivamente. En la Figura 9(d) se observa la relación señal a ruido en función del tiempo.

En la Figura 10 se muestra la magnitud de los espectros de frecuencia de la señal de voz procesada, la señal filtrada y el ruido extraído, los cuales se calcularon con la Transformada de Fourier. Se observa que el mayor contenido frecuencial de la señal de voz a procesar se encuentra en el rango de 0 a 2000 Hz. También se observa la diferencia del contenido de frecuencias de las señales, mostrando lo efectiva que resulta la Transformada de Wavelet para realizar el proceso de filtrado de una señal.

En este experimento se demostró que se puede aplicar la Transformada Wavelet en el proceso de filtrado de las señales de voz, porque cuando se escucha la señal filtrada se comprueba que no se pierde información, y se elimina parte del ruido en las zonas de silencio. Con respecto a la relación señal a ruido, se observa cómo varía su valor en el tiempo, dependiendo del segmento de voz procesado, con valores más bajos entre 5 y 10 dB en las zona de silencio y un valor máximo de 25 dB en un instante de la frase pronunciada.

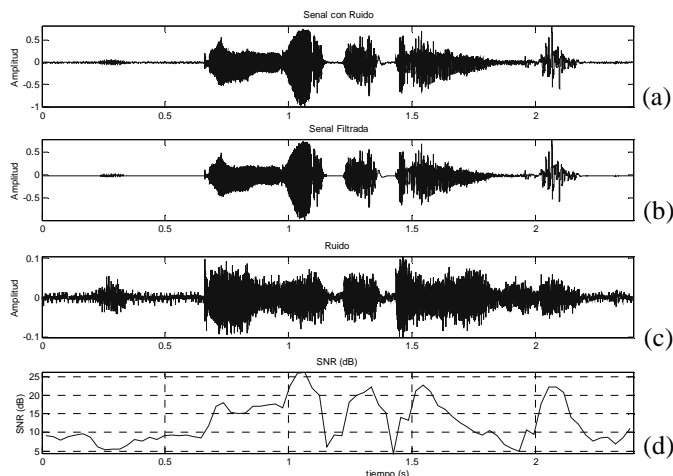


Figura 9. a) Señal de voz a analizar b) Señal filtrada c) Ruido d) Relación señal a ruido.

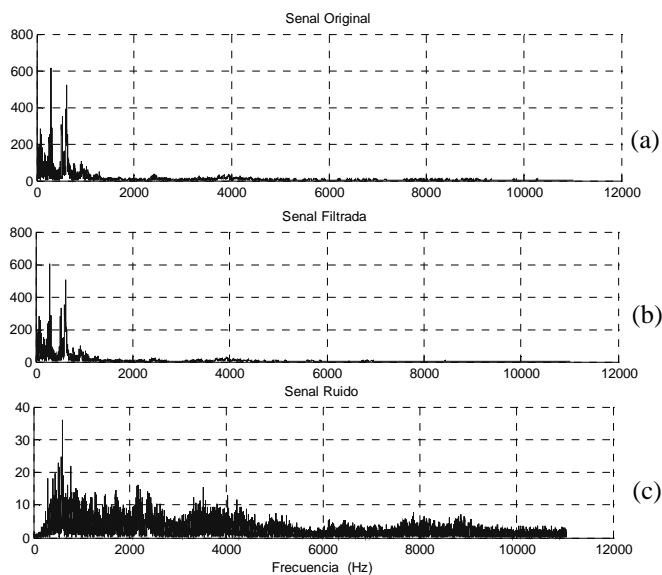


Figura 10. a) Espectro de la señal original de voz b) Espectro de la señal filtrada c) Espectro del ruido.

4. RESULTADOS

En la Tabla 1 se muestran los parámetros utilizados para el procesamiento de la señal de voz y los números de las figuras donde se representan los resultados del procesamiento de la señal de voz: “Hola, Hola, Hola” utilizando las Transformadas de Wavelet DB10, Sym10, Coif 5 y los niveles 3, 5, 10 y 15. El umbral utilizado fue 0,023478, obtenido a través de la relación de Donoho. Se observa en este procesamiento que la mayor disminución del ruido en las pausas o

silencios corresponde a los niveles de descomposición 10 y 15 (Figuras 12a y 12b). En las zonas de silencio se obtienen valores de relación señal a ruido por debajo de 10 dB, los cuales son menores que los procesados con los niveles de descomposición 3 y 5. Además, la señal filtrada se escucha sin variación. Por lo tanto, al utilizar esta Transformada Wavelet se recomienda el nivel de descomposición de 10.

Tabla 1. Análisis de SNR con diferentes Wavelets.

Wavelet	Nivel de Descomposición	Umbral	Figura
DB10	3	0,023478	11a,11b,11c,11d
DB10	5	0,023478	11e,11f,11g,11h
DB10	10	0,023478	12a,12b,12c,12d
DB10	15	0,023478	12e,12f,12g,12h
Sym 10	3	0,023478	13a,13b,13c,13d
Sym 10	5	0,023478	13e,13f,13g,13h
Sym 10	10	0,023478	14a,14b,14c,14d
Sym 10	15	0,023478	14e,14f,14g,14h
Coif 5	3	0,023478	15a,15b,15c,15d
Coif 5	5	0,023478	15e,15f,15g,15h
Coif 5	10	0,023478	16a,16b,16c,16d
Coif 5	15	0,023478	16e,16f,16g,16h

5. CONCLUSIONES

Con el estudio de la Transformada Wavelet se logró determinar que la misma se puede aplicar en el proceso de filtrado de señales de voz para eliminar su ruido y determinar de esta forma su relación señal a ruido.

Se utilizó una señal patrón para determinar si los valores generados por los algoritmos desarrollados eran precisos, y se verificó que los mismos estaban de acuerdo con los valores teóricos, dentro de un margen de error lo suficientemente pequeño.

Se utilizaron las Transformadas Wavelet de Daubechies (DB), Symmlet (Sym) y Coiflet (Coif) con los niveles de descomposición 3, 5, 10 y 15, para realizar el proceso de filtrado de señales de voz

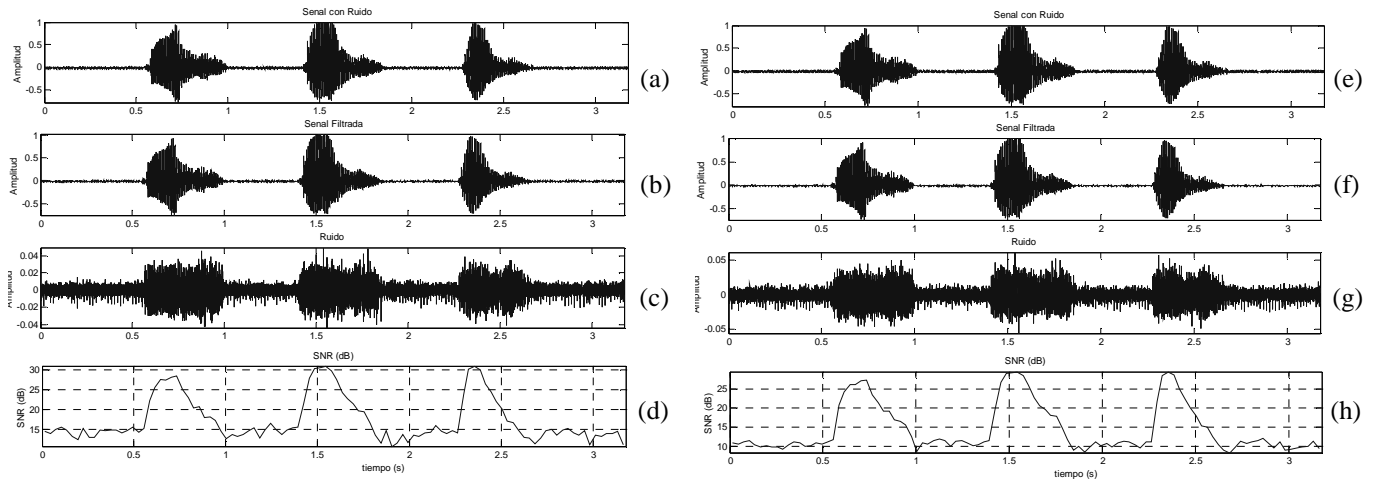


Figura 11. a) Señal de voz a analizar b) Señal filtrada c) Ruido d) Relación señal a ruido con DB10 nivel 3 e) Señal de voz a analizar f) Señal filtrada g) Ruido h) Relación señal a ruido con DB10 nivel 5.

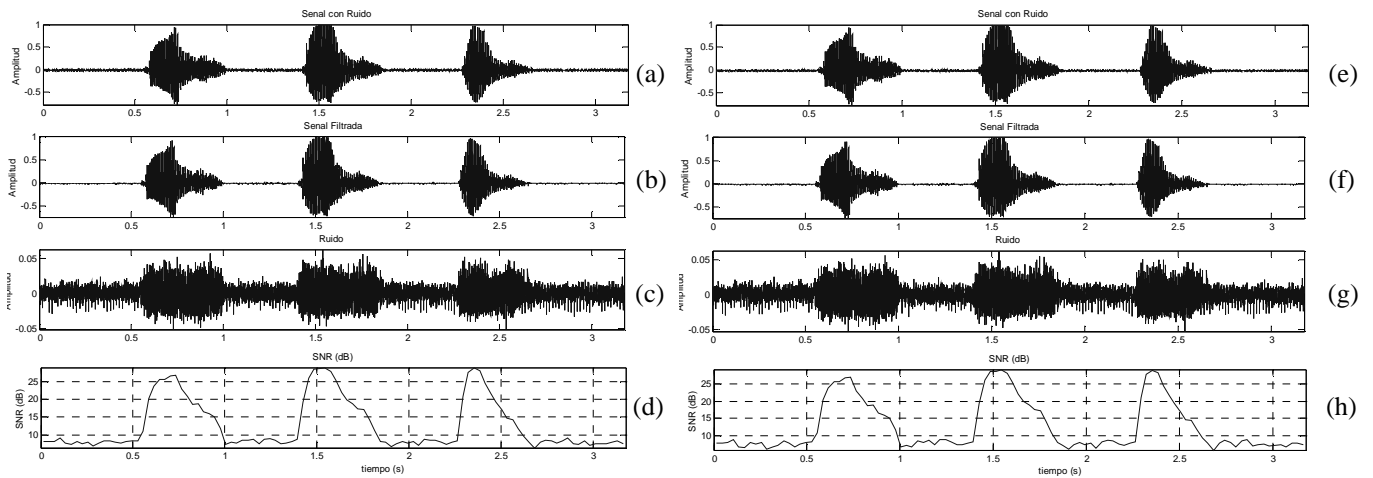


Figura 12. a) Señal de voz a analizar b) Señal filtrada c) Ruido d) Relación señal a ruido con DB10 nivel 10 e) Señal de voz a analizar f) Señal filtrada g) Ruido h) Relación señal a ruido con DB10 nivel 15.

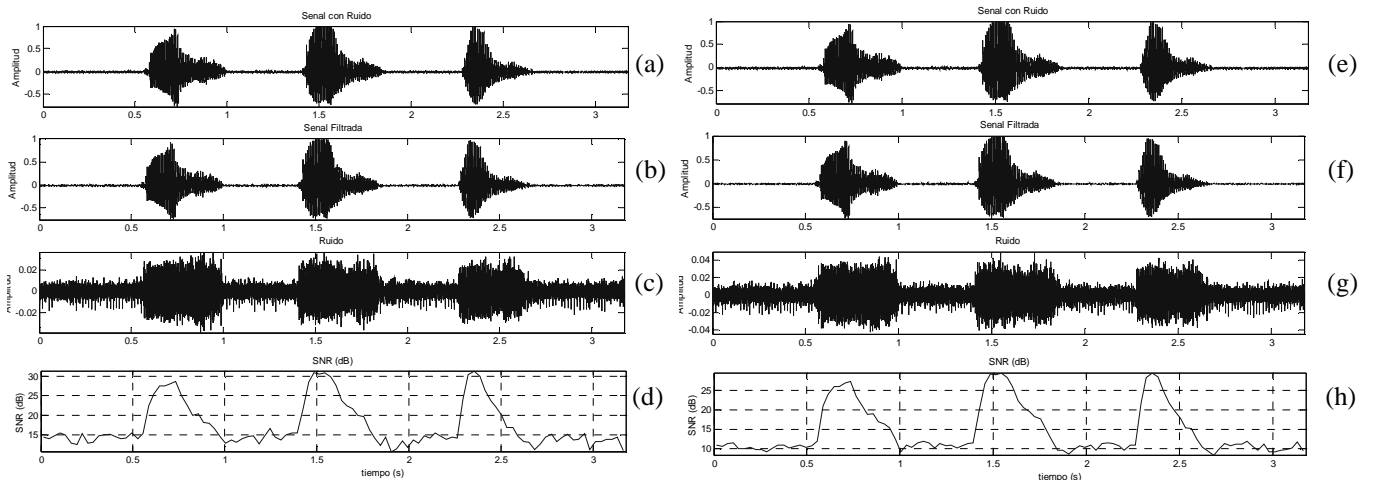


Figura 13. a) Señal de voz a analizar b) Señal filtrada c) Ruido d) Relación señal a ruido con Sym10 nivel 3 e) Señal de voz a analizar f) Señal filtrada g) Ruido h) Relación señal a ruido con Sym10 nivel 5.

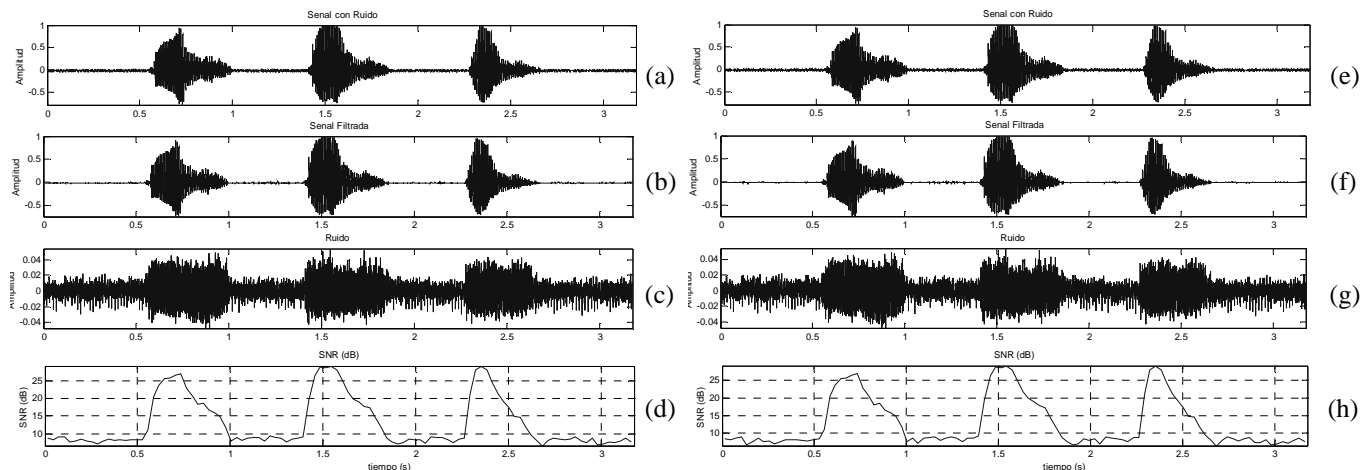


Figura 14. a) Señal de voz a analizar b) Señal filtrada c) Ruido d) Relación señal a ruido con Sym10 nivel 10 e) Señal de voz a analizar f) Señal filtrada g) Ruido h) Relación señal a ruido con Sym10 nivel 15.

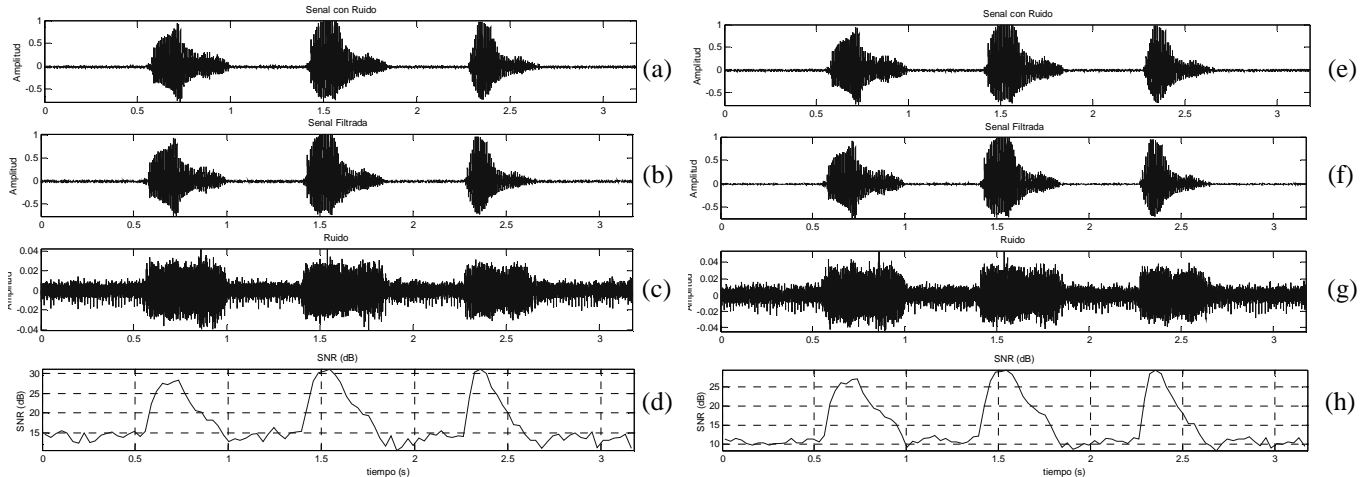


Figura 15. a) Señal de voz a analizar b) Señal filtrada c) Ruido d) Relación señal a ruido con Coif5 nivel 3 e) Señal de voz a analizar f) Señal filtrada g) Ruido h) Relación señal a ruido con Gif nivel 5.

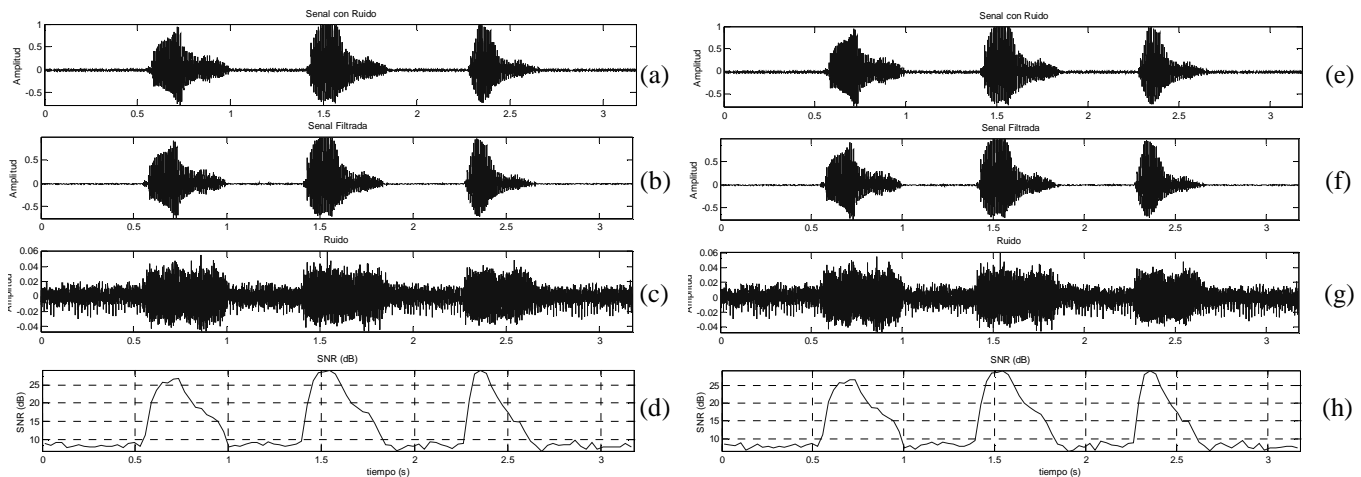


Figura 16. a) Señal de voz a analizar b) Señal filtrada c) Ruido d) Relación señal a ruido con Coif5 nivel 10 e) Señal de voz a analizar f) Señal filtrada g) Ruido h) Relación señal a ruido con Gif5 nivel 15.

formadas por palabras y zonas de silencio, observándose que para los niveles de 10 y 15 se obtiene un mayor filtrado en los intervalos correspondientes al silencio y por lo tanto una menor relación señal a ruido. Con estos niveles de descomposición, las palabras se escuchan sin deformación, eliminándose el ruido de fondo.

Se determinó que la señal de voz filtrada y los valores de la relación señal a ruido son semejantes cuando se aplican las transformadas DB10 y DB15. Igual resultado se obtiene con las Transformadas Sym10 y Sym15.

6. RECOMENDACIONES

Se recomienda utilizar otras Transformadas Wavelet y otros niveles de descomposición en el proceso de filtrado de las señales de voz para comparar los resultados obtenidos con los resultados de esta investigación. También se sugiere realizar estudios sobre la estimación del nivel de ruido en la señal de voz para determinar un umbral apropiado para cada señal a procesar, y utilizar una base de datos de señales de voces patológicas clasificadas según el tipo de enfermedad y aplicar los programas desarrollado.

REFERENCIAS

- [1] Haykin Simon Van Veen B. Señales y Sistemas Editorial Limusa- Mexico. 2001
- [2] Faundez Marcos Tratamiento Digital de Voz e Imagen y Aplicaciones a la Multimedia Marcombo. Boixareu Editores. Barcelona España. 2000
- [3] Obediente Soso Enrique Fonética y Fonología. Universidad de los Andes Mérida. Venezuela. 2001
- [4] Daubechies Ingrid .Ten Lectures on Wavelets. Capital City Press Filadelfia USA.1992
- [5] Kaiser Gerald A Friendly Guide to Wavelets Birkhauser U.S.A. 1999.
- [6] Mallat Stephane A Wavelet Tour of Signal Processing. 1998.
- [7] Chui K Charles. An Introduction in Wavelet .Academia. Press Inc. Reino Unido. 1992
- [8] García Janer Leonard Transformada de Wavelet aplicada a la extracción de información en señales de voz. Barcelona España. <http://www.eupmt.es/imesd/telematica/veu/thesis.pdf>. 1998

- [9] Cuesta Frau, Kovak D. Reducción del ruido en señales electrocardiográficas mediante la transformada de wavelets. <http://hpk.felk.cvut.cz/~xnovakd1/doc/wavelet.pdf> . 2000
- [10] Novak Daniel , Cuesta D. Denoising electrocardiogram signal using adaptive Wavelets. Universidad de Valencia España. http://plutarco.disca.upv.es/~jcperez/Documentos/Comg_sBrnodavid.pdf 2000
- [11] Donoho D. Denoising by soft – thresholding. IEEE. Trabs. Information theory. Vol 41, num3 95
- [12] Childers Donald Probability and Random Processes. McGrawHill U.S.A.1997.
- [13] Papoulis Athanasios Probability Random variables and Stochastic. Processes. Mc. Graw Hill U.S.A.1984.