

Aplicación de la transformada de wavelet para el análisis de señales de voz normales y patológicas

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

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

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

Email: cjimenez@uc.edu.ve

Resumen

En este trabajo se utilizó la transformada de wavelet db15, con un nivel de descomposición de 10, para realizar el proceso de filtrado de señales de voz y extraer parámetros que permitan determinar si la señal de voz pertenece a un paciente sano o con alguna patología. Se graficó y tabuló la relación señal a ruido de señales de voces correspondientes a muestras de la pronunciación de una “A” sostenida, clasificadas como normales y patológicas por un experto en el área de procesamiento de voz, observándose en las señales normales un mayor valor de la relación señal a ruido promedio, mayor valor máximo y una menor varianza.

Palabras clave: Wavelet, señales de voz, relación señal a ruido, multiresolución, señal estacionaria.

Applying the wavelet transform for the analysis of normal and pathological voices

Abstract

The wavelet transform db15, with a decomposition level of 10, was used in this research in order to filter voice signals and extract some parameters that would allow us to determine whether the voice signal is a healthy one or if it has some pathology. The signal-to-noise ratio of voice signals corresponding to a sustained “a”, which were classified as normal or pathological by an expert in the area of speech processing, was plotted and tabulated. It was observed that for normal voices the average signal-to-noise-ratio was higher than that for pathological voices, the maximum value of the signal-to-noise-ratio was higher, and the variance of the signal-to-noise-ratio curve was lower.

Keywords: Wavelet, speech, signal-to-noise-ratio, multi-resolution, stationary signal.

1. INTRODUCCIÓN

En el análisis de las señales de voz se utilizan técnicas de procesamiento digital de señales con el objetivo de identificar y caracterizar la calidad de las mismas mediante parámetros. Si se seleccionan estos parámetros convenientemente, se podrían utilizar para comparar voces provenientes de pacientes sanos y de pacientes con patologías de voz, para ayudar al especialista a diagnosticar en forma objetiva y rápida el estado de la señal de voz [1].

La señal de voz presenta un comportamiento no estacionario [2, 3], por lo que se deben aplicar técnicas

de procesamiento de señales que permitan determinar 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 gráfico tiempo-frecuencia. Algunas técnicas utilizadas para esto 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 cualquier mejora de resolución de cualquiera de ellos está limitada por el

mejor resolución en dominio del tiempo, reduce la información en el dominio de la frecuencia, y viceversa.

Debido a que la transformada Wavelet [4-7] 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 extraer parámetros que permitan determinar si corresponde a una voz sana o pertenece a un paciente con alguna patología.

Se utilizó la transformada de wavelet db15, con el nivel de descomposición de 10, y se graficó y tabuló la relación señal a ruido de señales de voces correspondientes a muestras de una “A” sostenida, clasificadas como normales y patológicas por un experto en el área de procesamiento de voz, observándose en las señales normales un mayor valor de la relación señal a ruido promedio, mayor valor máximo y una menor varianza.

2. METODOLOGÍA

2.1 Transformada de wavelet

La transformada de Wavelet permite realizar un análisis de múltiple resolución (MRA), lo cual analiza la señal con resolución diferente a diferentes frecuencias. Se diseña para producir alta resolución en el tiempo y baja resolución en frecuencia para señales de alta frecuencia, y baja resolución en el tiempo y alta resolución en frecuencia para señales de baja frecuencia. En las ecuaciones 1 y 2 se indica la forma como se obtiene la transformada de wavelet continua (CWT) de una función $f(t)$.

$$C(\text{escala}, \text{posición}) = \int_{-\infty}^{\infty} f(t) \cdot \psi(\text{escala}, \text{posición}, t) dt \quad (1)$$

$$CWT(s, \tau) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} f(t) \cdot \psi\left(\frac{t-\tau}{s}\right) dt \quad (2)$$

Donde ψ es la función wavelet (llamada wavelet madre), la cual se debe escalar (s) y desplazar (t) sobre el eje del tiempo. El factor en el denominador (la raíz cuadrada de s) es utilizado como un factor de normalización de la energía.

El parámetro de escalamiento (s) permite comprimir o expandir la función wavelet y en la transformada es utilizada en el denominador ($y(t/s)$). Este parámetro indica el grado de resolución con que se analiza la señal. Un valor alto de este factor ($|s| > 1$) corresponde a una vista global de la señal (expansión de la wavelet) mientras que un factor de escala bajo ($|s| < 1$) corresponde a ver detalles de la señal (se comprime la wavelet). El factor de posición (t) permite desplazar la función wavelet en el eje del tiempo ($y(t-t)$), hasta el intervalo de tiempo que se encuentre definida la función $f(t)$. El resultado que se obtiene cuando se aplica la transformada de wavelet son los coeficientes $C(s, t)$, que son función de la escala y la posición [4-7].

La wavelet utilizada en este proyecto es la Daubechies 15 (db15), la cual no es simétrica y es mostrada en la figura 1 [7,8].

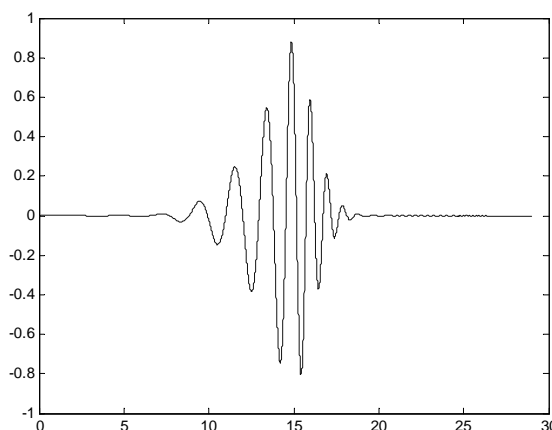


Figura 1. Función Wavelet db15.

2.2 Filtrado de la señal de voz

El análisis de las señales con la transformada de wavelet es equivalente a un proceso de filtrado, donde se realiza una división de los coeficientes, obteniéndose los coeficientes de aproximación (A) y detalle (D). La aproximación son los valores altos de la escala, correspondiente a las componentes de baja frecuencia de la señal, por lo tanto, está asociada a la función de escalamiento que se determina con un filtro pasa bajo. Los detalles son los valores bajos de la escala correspondientes a las componentes de alta frecuencia, y está asociada a la función wavelet que se determina como un filtro pasa alto. En la Figura 2

se muestra un esquema del proceso de filtrado, donde la señal a procesar es pasada por los filtros paso bajo y pasa alto, los cuales son filtros complementarios y se producen dos señales [5, 7, 9, 10].

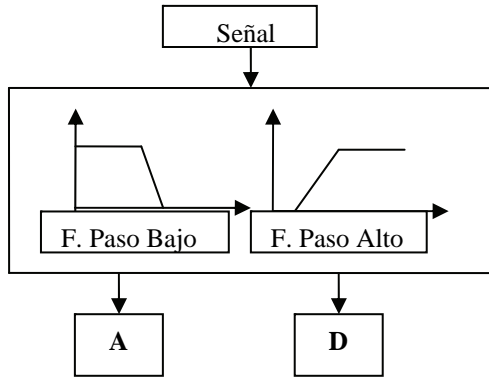


Figura 2. Esquema del proceso de filtrado.

El proceso de filtrado o descomposición se repite, para descomponer la señal en N niveles, cada nivel con una resolución más baja. El número de veces que es filtrada la señal viene determinada por el nivel de descomposición. En la Figura 3 se muestra el árbol de filtrado de wavelet con tres niveles.

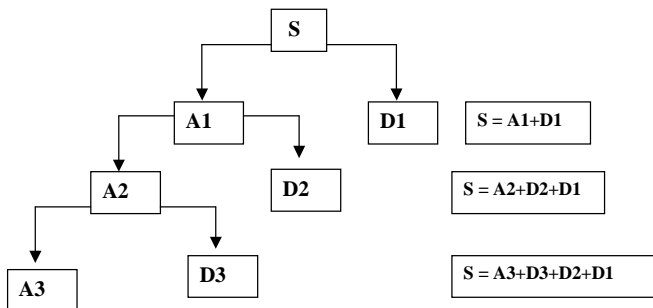


Figura 3. Proceso de descomposición con tres niveles.

Para reducir el ruido de la señal de voz se implementó un programa donde se eliminan las componentes de la señal que están por debajo de un determinado umbral (THR), el cual es calculado según la teoría de Donoho [11-13] mediante la expresión indicada en la ecuación 3.

$$THR = (\sqrt{2 \ln(n)}) \times s \quad (3)$$

Donde:

n = longitud de la señal a analizar

s = estimado del nivel de ruido de la señal, obtenido con la ecuación 4

$$s = \frac{\text{mediana}(\text{modulo}(c))}{0.6745} \quad (4)$$

c = 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 5 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 6 se muestra la expresión utilizada cuando se aplica este tipo de umbral. En la Figura 4 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 5a y 5b se muestran los resultados de las modificaciones de los coeficientes cuando se aplica el umbral duro y el suave [10,12,14].

$$C_m(i, j) = \begin{cases} 0 & \text{si } |C(i, j)| < thr \\ C(i, j) & \text{si } |C(i, j)| \geq thr \end{cases} \quad (5)$$

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

Para determinar la relación señal a ruido (snr) de la señal se utilizó la ecuación 7 [8].

$$snr = 10 * \log_{10} \left[\frac{\sum_{n=0}^{L-1} (x(n))^2}{\sum_{n=0}^{L-1} (\hat{x}(n) - x(n))^2} \right] \quad (7)$$

Donde:

$\hat{x}(n)$ = señal filtrada

$x(n)$ = señal con ruido

snr = relación señal a ruido en decibelios

L = número de muestras de la señal

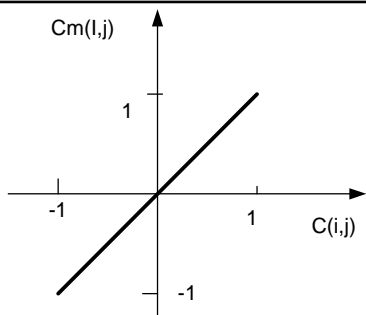


Figura 4. Señal original.

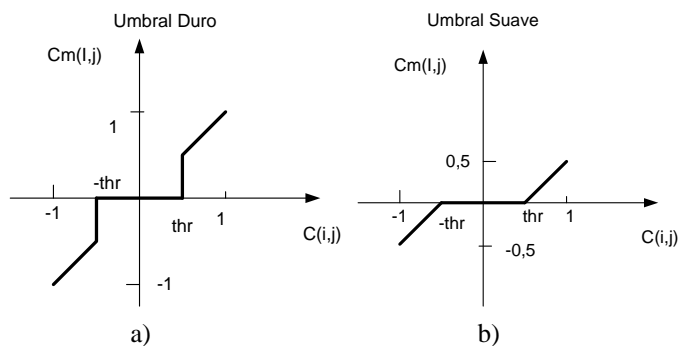


Figura 5. a) Umbral duro. b) Umbral suave.

3. RESULTADOS

3.1 Análisis de señales de voz sanas

En este experimento se analizaron doce señales de voz correspondientes a una “A” sostenida, las cua-

les fueron clasificadas como normales por un experto en el área de procesamiento de voz. Se utilizó para su procesamiento la transformada de Wavelets db15 con el nivel de descomposición de 10. Los resultados de la relación señal a ruido son mostrados en la Tabla 1, en donde se especifican el valor promedio (rango entre 13,767 y 21.7 dB), la varianza, el valor mínimo y el valor máximo (entre 16.047 y 22.7 dB) para cada una de las señales de voz. En la Figura 6 se muestra la relación señal-ruido de las cuatro primeras señales especificadas en la Tabla 1. En la Figura 7 se muestra desde la quinta a la octava señal y en la Figura 8 las últimas cuatro señales de voz sanas de la tabla 1.

3.2 Análisis de señales de voz patológicas

En este experimento se analizaron doce señales de voz correspondientes a una “A” sostenida clasificadas como patológicas por un experto en el área de procesamiento de voz. Se utilizó para su procesamiento la transformada de Wavelet db15 con el nivel de descomposición de 10, con el valor de umbral de 0.1, para todas las señales.

Los resultados de la relación señal a ruido son mostrados en la Tabla 2, en donde se especifican el valor promedio (entre 7,7065 y 16,167 dB), la varianza, el valor mínimo y el valor máximo (entre 13,449 y 20,307 dB) para cada una de la señal de voz. En la Figura 9 se muestra la relación señal-ruido de las cua-

Tabla 1. Relación señal a ruido de señales de voz normales con db15 nivel 10.

Señal	Valor Promedio (dB)	Varianza	Valor Mínimo (dB)	Valor Máximo (dB)
1101-ah	13,767	0,68098	11,653	16,047
1102-6ah	17,289	0,70303	14,948	20,103
1104-5ah	16,368	0,35040	14,781	17,6,73
1106-6ah	17,375	1.45420	14,644	20,317
1108-8ah	16,073	1,47190	13,081	17,,950
1109-9ah	15,669	1,31650	13,220	18,218
2101-6ah	18,063	1,17600	16,166	20,474
2104-7ah	15,609	5,45250	10,293	19,610
2106-5ah	15,105	0,30896	13,778	16,484
2107ah	16,368	0,35040	14,781	17,670
2201-6ah	15,550	0,60234	13,731	17,030
2205-5ah	21,752	0,173720	20,691	22,718

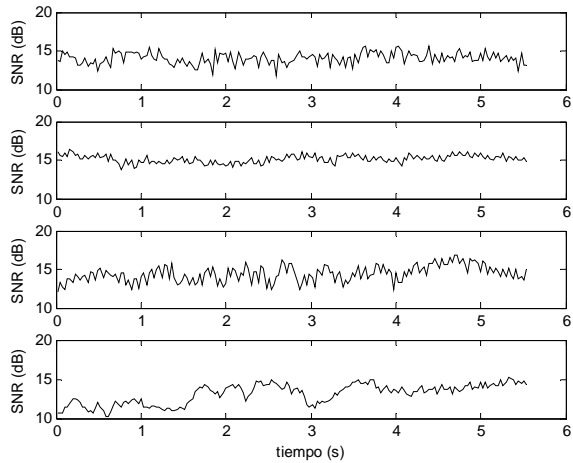


Figura 6. Relación señal a ruido de las cuatro primeras señales de voz normales con db15 nivel 10

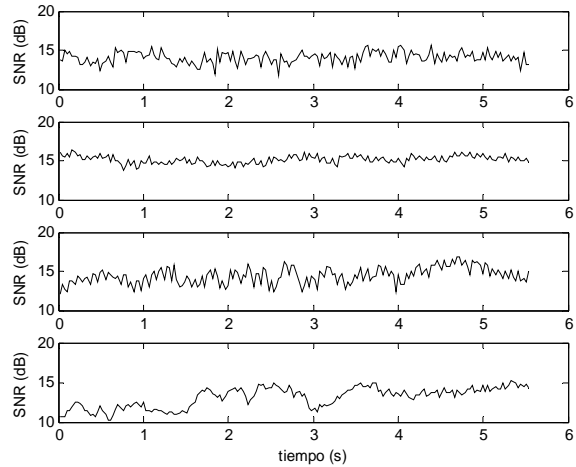


Figura 9. Relación señal a ruido de las cuatro primeras señales de voz patológicas con db15 nivel 10

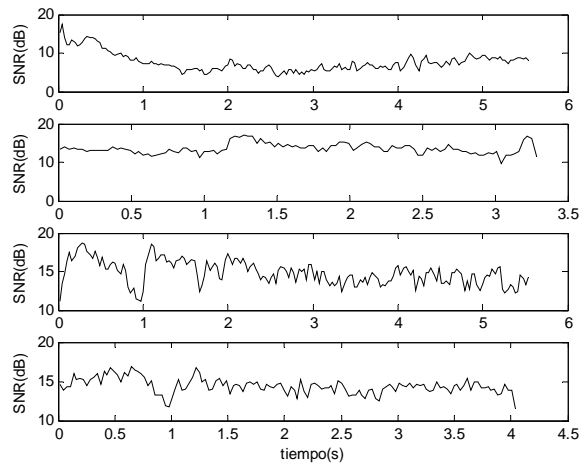


Figura 7. Relación señal a ruido de las señales de voz normales 5,6,7 y 8 con db15 nivel 10.

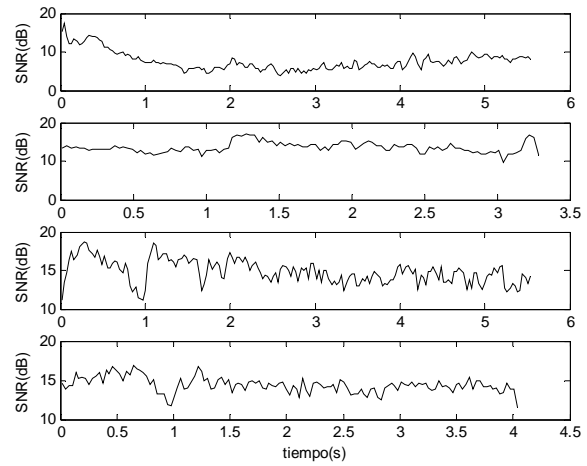


Figura 10. Relación señal a ruido de las señales de voz patológicas 5,6,7 y 8 con db15 nivel 10.

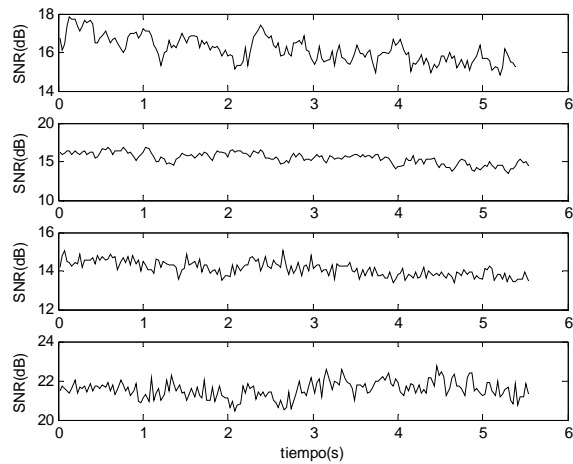


Figura 8. Relación señal a ruido de las señales de voz normales 9,10,11 y 12 con db15 nivel 10.

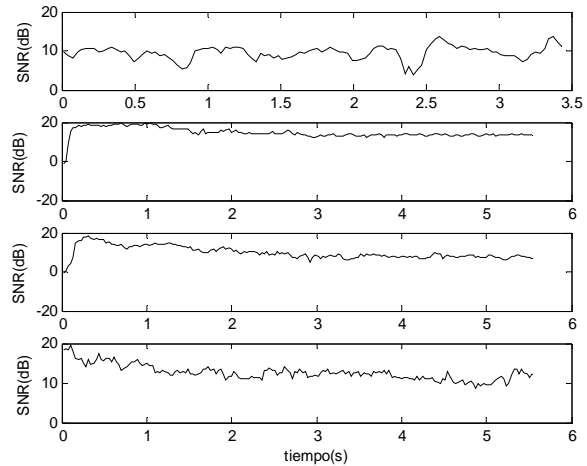


Figura 11. Relación señal a ruido de las señales de voz patológicas 9,10,11 y 12 con db15 nivel 10.

tro primeras señales especificadas en la Tabla 2. En la Figura 10 se muestra la relación señal-ruido desde la quinta a la octava señal y en la Figura 11 las últimas cuatros señales de voz patológicas de la Tabla 2.

Comparando los valores de la relación señal a ruido de las muestras de pacientes sanos con los que presentan alguna patología se observa que las voces correspondientes a los pacientes sanos presentan mayor magnitud de la relación señal-ruido tanto en el promedio como en el máximo, menor valor de la varianza y las formas de ondas presentan un comportamiento menos oscilatorio.

4. CONCLUSIONES

Se utilizaron señales de voz correspondientes a una "A" sostenida, clasificadas como normales y patológicas por un experto en el área de procesamiento de voz. Se aplicó la transformada de Wavelet db15 al conjunto de muestras, se calculó y graficó la relación señal a ruido, se tabularon los valores promedios, la magnitud mínima, la magnitud máxima y su varianza, y se compararon los resultados obtenidos del procesamiento de estas muestras.

Comparando los resultados obtenidos para las señales de voces clasificadas como normales y patológicas, se observó que el valor promedio y el valor máximo de la relación señal a ruido es mayor en las

señales normales que en las patológicas. Así mismo en las muestras patológicas se presentan formas de ondas con más variaciones que en las muestras normales, reflejando una varianza mayor.

4. REFERENCIAS

- [1] Del Pino Paulino, Díaz J. A. y Jiménez C. (2004) "Identificación de los parámetros espectrales que determinan la calidad de la voz". Ingeniería UC. Venezuela.
- [2] Faundez Marcos. (2000) "Tratamiento Digital de Voz e Imagen y Aplicaciones a la Multimedia" Marcombo Boixareu Editores. Barcelona España.
- [3] Obediente Soso Enrique. (2001) "Fonética y Fonología". Universidad de los Andes Mérida. Venezuela.
- [4] Daubechies Ingrid. (1992) "Ten Lectures on Wavelets". Capital City Press Filadelfia USA.
- [5] Kaiser Gerald (1999) "A Friendly Guide to Wavelets" Birkhauser U.S.A.
- [6] Mallat Stephane (1998) "A Wavelet Tour of Signal Processing" Academic. Press USA
- [7] Chui K Charles. (1992) "An Introduction in Wavelet". Academia. Press Inc. Reino Unido.
- [8] Jiménez Carlos, Díaz J. A. y Del Pino P. (2004) "Determinación de la relación señal a ruido de la voz utilizando la transformada de wavelet". Ingeniería UC. Venezuela.

Tabla 2. Relación señal a ruido de señales de voz patológicas con db15 nivel 10.

Señal	Valor promedio (dB)	Varianza	Valor Mínimo (dB)	Valor Máximo (dB)
2202-9ahBAD	14.0358	0.7434	11.7985	15.6239
2204-4ahBAD	15.1460	0.5010	13.7826	16.4072
2207-7ahBAD	14.4219	1.0368	12.1259	16.7991
2208-7ahBAD	13.1120	1.2635	10.1986	15.2145
3101-2ahBAD	7.6192	2.4094	3.9915	17.2424
3102-5ahBAD	13.5506	1.3549	9.7716	17.1036
3103-5ahBAD	14.8406	1.5596	11.1937	18.7190
3105-5ahBAD	14.4081	0.9822	11.5442	16.8624
3106-5ahBAD	9.5183	1.7436	3.9972	13.5593
3201-9ahBAD	14.7580	2.6394	1.0215	19.7743
3203-4ahBAD	9.8652	3.2357	0.5581	18.5604
3207-3ahBAD	12.6375	1.9363	8.6000	19.4750

- [9] García Janer Leonard (1998) “Transformada de Wavelet aplicada a la extracción de información en señales de voz”. Barcelona España. <http://www.eupmt.es/imesd/telematica/veu/thesis.pdf>.
- [10] Cuesta Frau, Kovak D. (2000) “Reducción del ruido en señales electrocardiográficas mediante la transformada de wavelet”. <http://hpk.felk.cvut.cz/~xnovakd1/doc/wavelet.pdf>
- [11] Novak Daniel, Cuesta D. (2000) “Denoising electrocardiogram signal using adaptive Wavelets”. Universidad de Valencia España. http://plutarco.disca.upv.es/~jcperez/Documentos/Comg_sBrnodavid.pdf.
- [12] Donoho D. (1995) “Denoising by soft-thresholding”. IEEE. Trabs. Information theory. Vol 41, N° 3.
- [13] Childers Donald (1997) “Probability and Random Processes”. McGraw Hill U.S.A.
- [14] Papoulis Athanasios. (1984) “Probability Random variables and Stochastic”. Processes. McGraw Hill U.S.A.