

Separación ciega de fuentes no-determinada aplicada a mezclas de voz con base en la transformada wavelet discreta

Undetermined blind source separation of speech mixtures based on discrete wavelet transform

Camilo Andrés Lemus¹ Dora María Ballesteros¹

Recibido 14 de abril de 2011, aceptado 25 de septiembre de 2012

Received: April 14, 2011 Accepted: September 25, 2012

RESUMEN

La separación ciega de fuentes, conocida como BSS por sus siglas en inglés (Blind Source Separation), es una técnica de procesamiento de señales que consiste en estimar fuentes en señales mezcladas linealmente, utilizando métodos como el ICA, para señales fuentes estadísticamente independientes. Uno de los algoritmos BSS más conocidos es el algoritmo JADE, el cual exige que el número de señales independientes coincida con el número de señales observadas (sensores). En situaciones reales, el número de sensores es menor al número de señales fuentes (BSS no-determinado) y el problema no tiene solución. En este proyecto se propone una solución para BSS no-determinado, adicionando una etapa de preprocesamiento y una etapa de descomposición basada en la transformada wavelet discreta. Nuestro modelo, el cual hemos denominado DWT+BSS, crea una señal virtual observada a partir de una señal real observada y utiliza los coeficientes wavelet de las señales observadas como entradas al algoritmo clásico JADE. El modelo se valida con señales de voz y audio, obteniendo índices de similitud entre las señales fuentes y las estimadas por encima de 0.7.

Palabras clave: Separación ciega de fuentes, transformada wavelet, señal de voz, índice de similitud, análisis de componentes independientes.

ABSTRACT

Blind Source Separation, BSS, is a signal processing technique which estimates sources from linearly mixed signals and it uses methods such as ICA for sources that are statistically independent. Among the best known BSS algorithms is the JADE method, which requires that the number of independent signals match the number of observed signals (sensors). In the real world, the number of sensors is lower than the number of sources (undetermined BSS) and therefore the problem has no solution. This work proposes a solution for undetermined BSS by pre-processing and decomposition stages based on the Discrete Wavelet Transform (DWT). Our proposal, which it is known as DWT+BSS, creates a virtual observed signal from a real observed signal and it uses the wavelet coefficients of the observed signals as the inputs of the classical JADE algorithm. We validated our model with voice and audio signals obtaining indexes of similarity over 0.7 between the original and the estimated sources.

Keywords: Blind source separation, wavelet transform, speech signal, index of similarity, ICA.

INTRODUCCIÓN

La separación ciega de fuentes (BSS) es una técnica de procesamiento de señales que consiste en estimar

señales fuentes a partir de sus mezclas lineales observadas [1-2]. Tradicionalmente, las señales fuentes deben ser estadísticamente independientes y debe existir el mismo número de señales observadas como

¹ Programa Ingeniería en Telecomunicaciones. Grupo de investigación TIGUM. Universidad Militar Nueva Granada. Kra. 11 No. 101-80. Bogotá, Colombia. E-mail: camilo.lemus@unimilitar.edu.co; lemus.camilo@gmail.com; dora.ballesteros@unimilitar.edu.co

de señales fuentes. Uno de los métodos matemáticos que más se ha utilizado en el problema BSS es el método de análisis de componentes independientes (ICA), el cual ha sido el punto de partida de algoritmos como el JADE [3-5].

En los últimos años la transformada wavelet se ha utilizado en modelos BSS, con el fin de mejorar los resultados obtenidos a través de los algoritmos basados en ICA. En [6], los autores utilizan la transformada wavelet como etapa de preprocesamiento de las señales observadas y las señales obtenidas son más similares a las señales fuentes originales que con el solo algoritmo BSS tradicional. En [7], la entrada al algoritmo BSS corresponde a todos los coeficientes wavelet de las señales observadas; mientras en [8] se seleccionan los coeficientes wavelet más representativos a partir de un criterio de análisis estadístico.

Por otro lado, en ambientes reales en los que el número de sensores o señales observadas, n , es menor al número de fuentes, m , el método ICA no se puede utilizar para separar las señales fuentes, ya que éste necesita que n sea igual a m . A este tipo de problemas se le conoce como BSS no-determinado. Algunos autores han propuesto soluciones a BSS no-determinado basadas en la transformada wavelet, en las que se estiman señales observadas a partir de señales reales observadas y de esta forma el número total de señales observadas se hace igual al número de señales fuentes a estimar. Por ejemplo, en [9-10] se estima una señal observada a partir del radio de dos señales observadas, basado en una hipótesis de dispersión. La desventaja consiste en que la calidad de las señales fuentes estimadas depende notablemente del tamaño de la ventana de tiempo utilizada y no existe un criterio global para la selección de la ventana de tiempo.

En este trabajo se combinan las dos alternativas de uso de la transformada wavelet en el problema de BSS no-determinado, y se comparan los resultados en ambientes ruidosos y sin ruido de señales de audio, voz y voz-audio. Se presenta un modelo, denominado DWT+BSS, el cual a partir del filtrado wavelet de una señal observada real, estima una señal observada virtual. De esta forma se iguala el total de señales observadas al total de señales fuentes a estimar. Las señales observadas se transforman al dominio wavelet y sus coeficientes ingresan

al algoritmo JADE. A la salida del algoritmo se obtienen los coeficientes wavelet de las señales fuentes estimadas, y al aplicarles la transformada wavelet inversa se transforman al dominio temporal.

Dentro de las condiciones de diseño del modelo se tienen los parámetros de filtrado wavelet que incluyen el número de niveles de descomposición, la base wavelet, el tipo de umbral y el valor del umbral [11]. Con el propósito de presentar una condición de diseño que sea global, el criterio de selección de los parámetros se basa en el índice de similitud entre la señal virtual observada y la señal observada real. Se aplican pruebas en señales de voz capturadas en ambientes reales (ruidosos) y señales de audio sintéticas. Adicionalmente, se utiliza de nuevo el índice de similitud como parámetro de verificación de la calidad de las señales estimadas. Si el índice se acerca uno, implica que la similitud de las señales es alta; si se acerca a cero, la similitud de las señales se hace nula. Se comparan los resultados con dos casos de señales virtuales tanto para voz como para audio, con el propósito de establecer el rango del criterio de selección de los parámetros de diseño en el bloque de filtrado.

El documento está organizado como sigue. Primero, presentamos un análisis teórico del modelo clásico BSS y del modelo propuesto para BSS no-determinado. Segundo, presentamos el diseño experimental y el índice de similitud. Tercero, presentamos algunos de los resultados obtenidos en el proyecto. Cuarto, incluimos la sección de conclusiones y, por último, la sección de agradecimientos.

ANÁLISIS TEÓRICO

En esta sección presentamos la formulación matemática del modelo clásico BSS y nuestra propuesta para el caso de BSS no-determinado.

Modelo clásico BSS

La Separación Ciega de Fuentes es una técnica capaz de estimar información contenida en mezclas de señales, asumiendo que cada una tiene una distribución no gaussiana y es estadísticamente independiente de las demás. En el caso BSS de audio se disponen de m micrófonos que reciben señales de n fuentes diferentes, generando un sistema de ecuaciones lineales, de acuerdo con:

$$\begin{aligned}
 x_1(t) &= a_{11}s_1 + a_{12}s_2 + \dots + a_{1n}s_n \\
 x_2(t) &= a_{21}s_1 + a_{22}s_2 + \dots + a_{2n}s_n \\
 &\dots \\
 x_m(t) &= a_{m1}s_1 + a_{m2}s_2 + \dots + a_{mn}s_n
 \end{aligned}
 \tag{1}$$

Donde $x_1(t), x_2(t), \dots, x_m(t)$ corresponden a las señales captadas por los micrófonos resultado de la combinación lineal de las n señales fuente y las constantes a_{mn} se relacionan con la distancia de las fuentes a los micrófonos. En forma matricial, el sistema de ecuaciones se representa como:

$$X = A.S \tag{2}$$

Donde X es la matriz de las señales observadas, A la matriz cuadrada de $n \times n$ que contiene los coeficientes de las mezclas aleatorias, y S la matriz que contiene las señales fuente (Figura 1). Este sistema tiene solución siempre y cuando $m \geq n$, es decir, se disponga de al menos el mismo número de sensores que de señales fuentes.

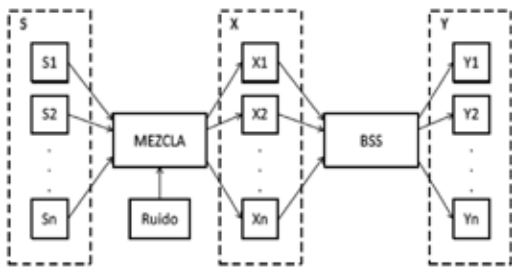


Figura 1. Modelo BSS clásico. Basado en [1].

Cuando $n > m$ (mayor número de fuentes que de sensores) el modelo se transforma en BSS no-determinado [12]. En este caso, tal y como está planteado el algoritmo JADE, no se tiene solución al problema, es decir, no se pueden estimar las n señales fuentes.

Modelo propuesto BSS no-determinado

La solución que se plantea en nuestro modelo DWT+BSS consiste en crear señales virtuales a partir de las señales reales observadas, utilizando una etapa de filtrado que crea una señal virtual a partir de una señal observada real utilizando umbralización en el dominio wavelet. En nuestro modelo, las entradas al algoritmo JADE corresponden a los coeficientes en el dominio wavelet de las señales observadas, denominados W_{s1} y a la salida

del algoritmo se obtienen los coeficientes de las señales estimadas, W_{xi} ; de tal forma, que las señales estimadas en el dominio del tiempo Y_i , corresponden a la reconstrucción wavelet de los coeficientes W_{xi} . En la Figura 2 se presenta el modelo DWT+BSS para dos señales fuente S_1 y S_2 y una sola señal observada real, X_1 .

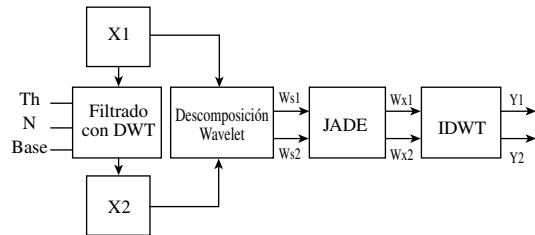


Figura 2. Modelo propuesto. Ejemplo $n=2, m=1$.

En forma general, para un número m de señales observadas y un número n de señales fuentes, en una situación no-determinada, el modelo se comporta de acuerdo a las ecuaciones (3) a (7).

Las m primeras señales observadas X cumplen con la ecuación (3) como una superposición lineal de sus señales fuentes S :

$$\begin{bmatrix} X_1 \\ \dots \\ X_m \end{bmatrix} = \begin{bmatrix} A_1 \\ \dots \\ A_m \end{bmatrix} \cdot \begin{bmatrix} S_1 \\ \dots \\ S_m \end{bmatrix} \tag{3}$$

La n señal observada (sensor virtual) se reconstruye a partir del filtrado de cualquiera de las señales observadas reales, a partir de (4):

$$X_n = \hat{X}_i \quad i = \forall [1 \dots m] \tag{4}$$

Donde \hat{X}_i es la señal filtrada de la observación i y X_n es la señal virtual.

Al conjunto de señales reales y virtuales se les aplica la transformada wavelet discreta para obtener los coeficientes w ; utilizando $W(\cdot)$ como el operador de la transformada:

$$w_s = W(X) \tag{5}$$

Una vez obtenidos los n vectores de coeficientes wavelet, se aplica el algoritmo JADE tradicional, de la forma:

$$w_s \xrightarrow{JADE} w_x \quad (6)$$

A los coeficientes estimados se les aplica la transformada wavelet inversa, denotada por el operador $W(\cdot)^{-1}$. La salida corresponde a las señales estimadas Y , en el dominio del tiempo.

$$Y = W(w_x)^{-1} \quad (7)$$

En este modelo se tienen en cuenta algunos parámetros relacionados con el filtrado por medio de la transformada wavelet y a la descomposición wavelet de las señales. Para el filtrado, la señal real se descompone utilizando N niveles de descomposición, con *base* wavelet y umbral *th*; de tal forma que N , *base* y *umbral* son parámetros de entrada al bloque de filtrado. Para la descomposición, las entradas al bloque son N y *base*. Debido a que no se conoce a priori el comportamiento en tiempo de las señales fuentes, no se pueden fijar los parámetros de filtrado y descomposición, porque estos pueden variar de señal a señal.

El ajuste de las variables de diseño se basa en el índice de similitud entre la señal observada real y la señal observada virtual. Si el índice de similitud es alto, las señales observadas (real y virtual) son estadísticamente similares y no se pueden estimar las señales fuentes; por el contrario, si el índice de similitud es bajo, el sensor virtual no se comporta como un combinatorio lineal de todas las señales fuentes, lo que imposibilita la estimación de al menos una de las señales fuente.

DISEÑO EXPERIMENTAL

En esta sección presentamos el diseño del experimento para la validación del modelo propuesto. Adicionalmente, seleccionamos el parámetro a utilizar para medir la similitud entre las señales reales y estimadas, y entre las señales fuentes originales y las señales fuentes estimadas.

Creación de los datos de prueba

Las pruebas incluyen dos escenarios. El primero consiste en señales de voz capturadas en ambientes reales, y el segundo en señales de audio del banco de señales de Matlab ©.

Para el primer escenario, se capturan las señales fuentes por separado y se mezclan de acuerdo a la ecuación (3). Los valores del vector A son obtenidos aleatoriamente en un rango comprendido entre 0 y 1.

Para el segundo escenario se utilizan los registros *train* y *chirp* de Matlab ©. De nuevo se mezclan de acuerdo a la ecuación (3) con ponderación aleatoria.

Aunque en una situación real no se dispone de las señales fuentes originales, la mezcla se realiza de forma artificial con el propósito de validar nuestro modelo, a partir de la similitud entre las señales fuentes obtenidas y las señales fuentes originales.

Selección del parámetro de similitud

En la literatura se encuentran varios parámetros de comparación entre señales de voz/audio, como la relación-sígnal-a-ruido (SNR), el radio de verosimilitud (LLR), la evaluación de la percepción de la calidad de voz (PESQ) y el promedio del puntaje de opinión (MOS). Recientemente, ha sido propuesto el cuadrado del coeficiente de correlación (ρ^2) como un índice de distorsión de las señales de voz [13-14]. Hemos seleccionado el índice de distorsión como nuestro parámetro de similitud de las señales de voz, por varias razones: el resultado es independiente del rango dinámico de las señales a comparar, es normalizado, lo que permite establecer el valor óptimo y el valor mínimo [15], y por último es un parámetro que se calcula con bajo costo computacional.

El índice de similitud se calcula a partir de la siguiente ecuación:

$$\rho^2(S, \hat{S}) = \left[\frac{\sum_{i=1}^m (S(i) - S_{mean}) \cdot (\hat{S}(i) - \hat{S}_{mean})}{\sqrt{\sum_{i=1}^m (S(i) - S_{mean})^2} \cdot \sqrt{\sum_{i=1}^m (\hat{S}(i) - \hat{S}_{mean})^2}} \right]^2 \quad (8)$$

Donde S, \hat{S}, S_{mean} and \hat{S}_{mean} son la señal original, la señal estimada, el promedio de la señal original y el promedio de la señal estimada, respectivamente. Su valor está comprendido entre 0 y 1. Si es cero, significa que las señales son completamente diferentes; pero si es uno, significa que las señales son perceptualmente iguales. Ese índice representa la correlación lineal entre las dos señales de voz.

RESULTADOS

A continuación presentamos algunos resultados del modelo propuesto para el caso BSS no-determinado, para dos señales fuentes y un solo micrófono. Primero se analiza el impacto del bloque de filtrado dentro del modelo propuesto y seguidamente el impacto del bloque de descomposición wavelet. Se valida con señales de voz y señales de audio.

Bloque de filtrado

En este apartado analizamos cómo cambia la calidad de las señales fuentes estimadas si se cambian los valores de los parámetros de diseño del bloque de filtrado (*N, base, th*). Dado que el criterio de selección de los parámetros está dado por el índice de similitud entre la señal virtual y la señal observada real, nos centramos en definir qué rango de valores es más recomendable para estimar con mayor calidad las señales fuentes. Primero presentamos un ejemplo con señales de voz capturadas en entorno real y seguidamente con señales de audio.

La Figura 3 presenta un ejemplo con señales de voz. La señal fuente 1 corresponde al mensaje “hola”, mientras que la señal fuente 2 al mensaje “señales”. Finalmente se grafica la señal capturada por el micrófono.

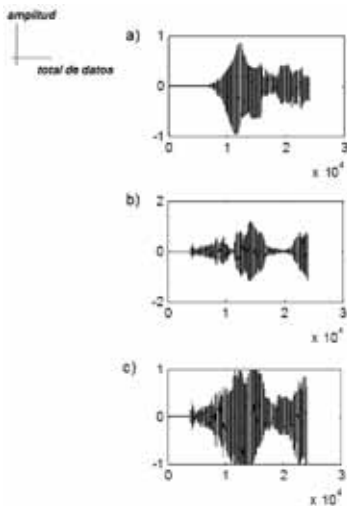


Figura 3. Señales de voz: a) fuente 1, b) fuente 2, c) señal observada real.

Con el propósito de establecer qué tan similar debe ser la señal virtual en relación a la señal observada real, presentamos dos casos. En el primero, la señal

virtual es altamente similar a la señal observada real ($\rho^2=0,98$), mientras en el segundo caso, la señal virtual tiene un nivel medio de similitud con la señal observada real ($\rho^2=0,56$). En la Figura 4 se presentan la señal virtual para el caso 1 y sus señales fuentes estimadas; mientras en la Figura 5 la señal virtual caso 2 y sus señales fuentes estimadas.

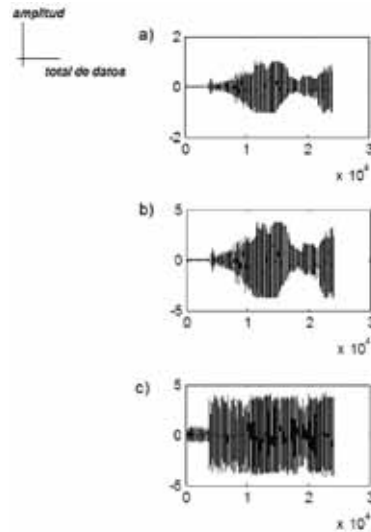


Figura 4. Caso 1, señales de voz: a) señal virtual, b) señal fuente estimada 1, c) señal fuente estimada 2.

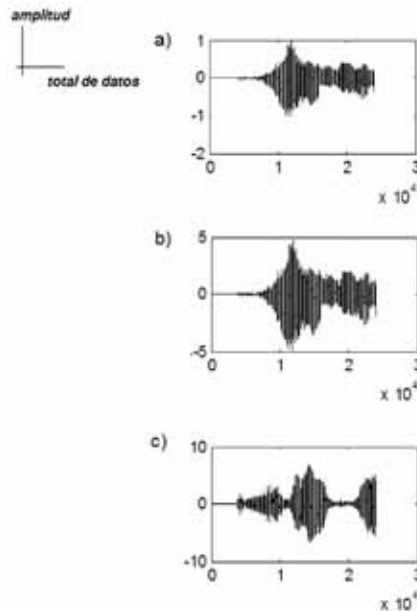


Figura 5. Caso 2, señales de voz: a) señal virtual, b) señal fuente estimada 1, c) señal fuente estimada 2.

De acuerdo a las gráficas, las señales estimadas en la Figura 5 son más similares a las señales fuentes originales, que las encontradas en la Figura 4. En el primer caso, sólo se logró estimar una señal con éxito.

Matemáticamente, el índice de similitud entre las señales fuentes estimadas y las señales fuentes originales se presenta en la Tabla 1. Se corrobora que el caso 2 brinda mejores resultados que el caso 1.

Tabla 1. Similitud entre señales fuentes originales y señales fuentes estimadas: señales de voz.

Caso 1		Caso 2	
$\rho^2(S_1, Y_1)$	$\rho^2(S_2, Y_2)$	$\rho^2(S_1, Y_1)$	$\rho^2(S_2, Y_2)$
0.51	< 0.1	0.71	0.76

Para la segunda parte de la validación se trabaja con señales de audio. En la Figura 6 se grafican las señales *train* y *chirp* y la señal observada con el sensor real. De nuevo se analizan dos señales virtuales, una con un alto índice de similitud con la señal observada real (0,97) y otra con un nivel medio de similitud (0,5).

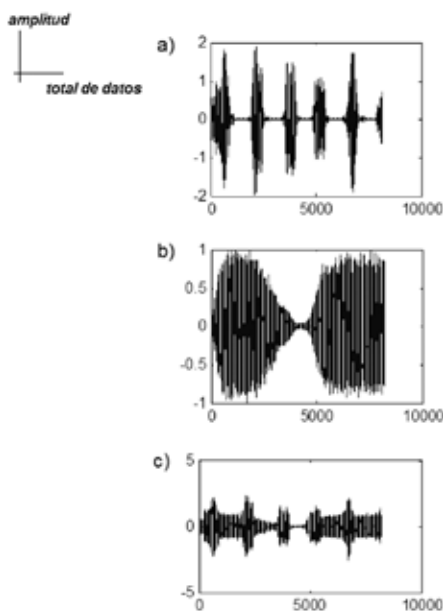


Figura 6. Señales de audio: a) fuente 1, b) fuente 2, c) señal observada real.

El primer caso se grafica en la Figura 7 y el segundo caso en la Figura 8.

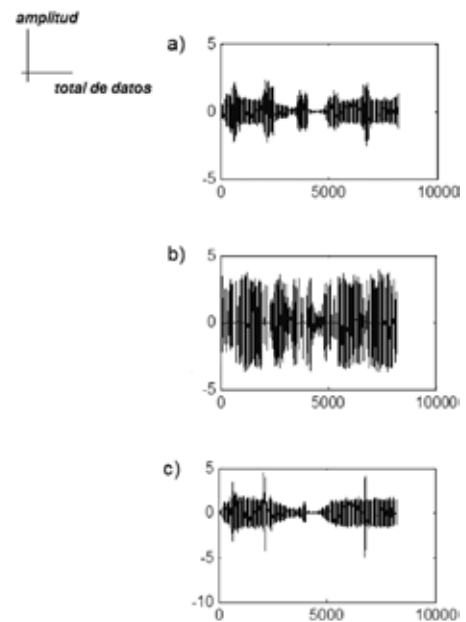


Figura 7. Caso 1, señales de audio: a) señal virtual, b) señal fuente estimada 1, c) señal fuente estimada 2.

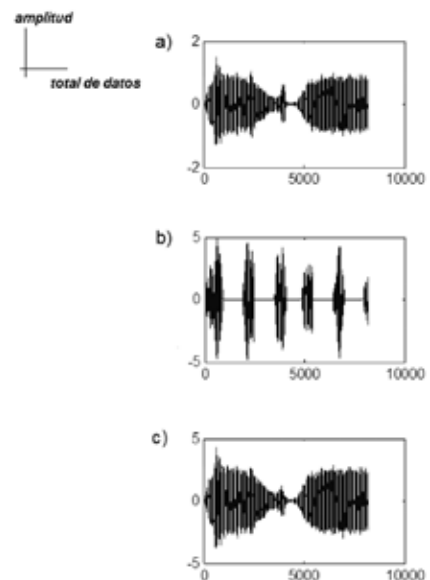


Figura 8. Caso 2, señales de audio: a) señal virtual, b) señal fuente estimada 1, c) señal fuente estimada 2.

De acuerdo a los resultados, el caso 2 estima de mejor forma las señales fuentes en comparación con el caso 1. Los resultados matemáticos de similitud se presentan en la Tabla 2.

Tabla 2. Similitud entre señales fuentes originales y señales fuentes estimadas: señales de audio.

Caso 1		Caso 2	
$\rho^2(S_1, Y_1)$	$\rho^2(S_2, Y_2)$	$\rho^2(S_1, Y_1)$	$\rho^2(S_2, Y_2)$
0	0,6	0,94	0,90

De nuevo, un alto nivel de similitud entre la señal observada real y la virtual no arroja los mejores resultados a la hora de estimar las señales fuente.

Bloque de filtrado

En este segundo apartado evaluamos la importancia del bloque de descomposición wavelet dentro del modelo. Se estiman las señales fuentes a partir de las señales observadas y sus versiones virtuales, utilizando las condiciones del caso 2 y eliminando el bloque de descomposición dentro del modelo. Es decir, al bloque JADE entran las señales observadas (real y virtual) en el dominio del tiempo.

Tabla 3. Similitud entre señales fuentes originales y señales fuentes estimadas: señales de voz y audio, sin bloque de descomposición.

Señales de voz		Señales de audio	
$\rho^2(S_1, Y_1)$	$\rho^2(S_2, Y_2)$	$\rho^2(S_1, Y_1)$	$\rho^2(S_2, Y_2)$
0,69	0,75	0,91	0,88

De acuerdo con los resultados de la Tabla 1 y la Tabla 2 (segundo caso) y los obtenidos en la Tabla 3, se puede establecer que el bloque de descomposición mejora ligeramente la calidad de las señales fuentes estimadas.

CONCLUSIONES

Se ha propuesto un modelo para el caso BSS no-determinado, el cual permite estimar todas las señales fuentes a partir de un número menor de señales observadas reales. El núcleo del modelo se basa en la creación de señales virtuales a partir de las señales observadas reales, a partir de un proceso de filtrado en el dominio wavelet. A diferencia de otros modelos, que estiman la señal virtual a partir del radio de señales observadas reales utilizando un modelo matemático complejo, nuestra propuesta se basa en el filtrado wavelet con un ajuste de

parámetros a partir del índice de similitud entre la señal virtual y la señal observada real. La gran ventaja de nuestro modelo es que es global, para lo cual se puede ajustar a todo tipo de señales de voz (con diferente duración, escala de tiempo, frecuencia de muestreo, amplitud, género, idioma, entre otras) teniendo en cuenta solamente el índice de similitud. Experimentalmente se estableció que si este índice entre la señal virtual y la señal observada real está en el rango de [0,5 0,7], la calidad de las señales estimadas tendrán un índice de similitud con las señales fuentes originales superior al 0,7. Este valor fue contrastado con un conjunto de pruebas mayor al presentado en este documento.

La inclusión de la etapa de descomposición dentro del modelo propuesto jugó un papel menos importante que el bloque de filtrado, pero en todo caso permitió mejorar ligeramente la calidad de las señales estimadas. Es decir, si la entrada al bloque JADE corresponde a los coeficientes wavelet en vez de las señales en el dominio temporal, la calidad de las señales estimadas puede ser mejor.

AGRADECIMIENTOS

El proyecto ha sido financiado por el Fondo de Investigaciones de la Vicerrectoría de Investigaciones de la Universidad Militar Nueva Granada, a través del proyecto ING641.

Los autores agradecemos a los revisores quienes con sus aportes contribuyeron a enriquecer el documento.

REFERENCIAS

[1] J.F. Cardoso. “Blind signal separation: statistical principles”. Proceedings of the IEEE, Vol. 86, pp. 2009-2025. October, 1998.
 [2] A.W. Bronkhorst. “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions”. Acta Acustica united with Acustica, Vol. 86, Issue 1, pp. 117-128. January, 2000.
 [3] S. Houcke and A. Chevreuril. “Blind source separation of a mixture of communication sources emitting at various baud-rates”. IEEE International conference on Acoustics, Speech, and Signal Processing. Vol. 3, pp. 3053-3056. May, 2002.

- [4] J.F. Cardoso and B.H. Laheld. "Equivariant adaptive source separation". IEEE Transactions on Signal Processing. Vol. 14, Issue 12, pp. 3017-3030. 1996.
- [5] J.F. Cardoso. "Estimating equations for source separation". IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 5, pp. 3449-3452. 1997.
- [6] X. Liu, X. Tan, S Wei and A. Anghuwo. "Blind Source Separation Algorithm Based on wavelet denoising". 1st International Conference on Pervasive Computing, Signal Processing and Applications. 2010.
- [7] Z. Zhao, T. Xie, Z. Zhao and X. Kong. "A Blind Source Separation Method Based on the Time Delayed Correlations and the Wavelet Transform". 8th International Conference on Signal Processing. 2006.
- [8] M.R. Mirarab, M.A Sobhani and A.A. Nasiri. "A New Wavelet Based Blind Audio Source Separation Using Kurtosis". 3rd International Conference on Advanced Computer Theory and Engineering. 2010.
- [9] T. Hamada, K. Nakano and A. Ichijo. "Wavelet-based Underdetermined Blind Source Separation of Speech Mixtures". International Conference on Control, Automation and Systems 2007. October, 2007.
- [10] O. Yilmaz and S. Rickard. "Blind separation of speech mixtures via time-frequency masking". IEEE Trans. Signal Process. Vol. 52, Issue 7, pp. 1830-1847. 2004.
- [11] C.A. Lemus y D.M. Ballesteros. "Algoritmo de separación ciega de fuentes en el dominio wavelet". Certificado de Registro de Soporte Lógico-Software. Libro: 13, Tomo: 27, Partida: 94. Dirección Nacional de Derechos de Autor, Ministerio del Interior y de Justicia de Colombia. Fecha de registro: 11 de Octubre 2010.
- [12] S. Makino. "Blind Speech Separation". Signals and Communication Technology. Springer, p. 432. Dordrecht, Holanda. 2007.
- [13] J. Benesty, C. Jingdong and H. Yiteng. "On the Importance of the Pearson Correlation Coefficient in Noise Reduction". IEEE Transactions on Audio, Speech, and Language Processing. Vol. 16, Issue 4, pp. 757-765. 2008.
- [14] J. Benesty, C. Jingdong, H. Yiteng and I. Cohen. "Chapter: Pearson Correlation Coefficient". Book: Noise Reduction in Speech Processing, Springer Topics in Signal Processing, 2009. Vol. 2, pp. 1-4. DOI: 10.1007/978-3-642-00296-0_5.
- [15] D.M. Ballesteros L. and J.M. Moreno A. "On the ability of adaptation of speech signals and data hiding". Expert Systems with Applications, Vol. 39, pp. 12574-12579. 2012. DOI: 10.1016/j.eswa.2012.05.027.