

Separación de señales usando análisis de componentes principales y muestreo compresivo con mediciones mínimas

Eduardo Rivera, Rodolfo Moreno, Héctor Pérez, y Mariko Nakano

Escuela Superior de Ingeniería Mecánica y Eléctrica, Unidad Culhuacan, Instituto Politécnico Nacional, Av. Santa Ana 1000, Coyoacan, Ciudad de México, 04440, México. (correo-e: lalomxp@hotmail.com; moar09@hotmail.com; hmperezm@ipn.mx; mnakano@ipn.mx)

Recibido Jun. 26, 2019; Aceptado Ago. 26, 2019; Versión final Sep. 15, 2019, Publicado Feb. 2020

Resumen

Con el fin de aumentar la eficiencia del Análisis de Componentes Independientes (ICA) y reducir la complejidad computacional del sistema, este artículo propone una metodología basada en la técnica de muestreo compresivo, usando mediciones mínimas. Esta permite comprimir y modificar las características Gaussianas de las señales de audio. ICA es uno de los esquemas más utilizados para la separación a ciegas de fuentes (BSS), a partir únicamente de las mezclas recibidas en un conjunto de sensores. Sin embargo, el ICA requiere que las señales involucradas no sean de tipo Gaussiano, o solamente una de ellas lo sea, características que no satisfacen las señales de audio, las cuales son en general de tipo Gaussiano. La metodología propuesta permite obtener la separación de las señales en forma más eficiente, con una menor complejidad computacional, aun cuando todas ellas sean de tipo Gaussiano.

Palabras clave: separación de fuentes; muestreo compresivo; ICA; algoritmo de gradiente ascendente

Signal separation using principal component analysis and compressive sampling with minimum measurements

Abstract

To increase the efficiency of the independent component analysis (ICA) and reduce the computational complexity of the system, this paper proposes a methodology based on the compressive sampling with minimum measurements. This allows compressing and modifying the Gaussian characteristics of audio signals. ICA is one of the most widely used schemes for separating the sources involved in an audio mixture received in a set of sensors. However, for proper operation it is required that the signals involved in the mixture do not have Gaussian characteristics or at most only one of them be Gaussian, characteristics that are not satisfied by the audio signals, which have in general Gaussian characteristics. The proposed methodology allows obtaining a better separation with lower computational complexity, and with a more efficient separation of the mixed signals, even if the signals involved in the mixture are of Gaussian toe.

Keywords: signal separation; compressive sampling; ICA; ascendant gradient algorithm

INTRODUCCIÓN

En la actualidad es posible encontrar un gran número de situaciones en las cuales un conjunto de señales se mezcla e interfiere entre sí, haciendo necesaria su separación para su posterior procesamiento. Lo anterior, ha incentivado el desarrollo de diversos esquemas que permitan recuperar las señales originales, a partir únicamente del conocimiento de las mezclas recibidas en un grupo de sensores, lo que se conoce como separación a ciegas de fuentes ó BSS, por sus siglas en inglés. Una de las primeras propuestas para resolver el problema de BSS fue desarrollada por Herault et al. (1985) y por Jutten y Herault, (1988) quienes intentaban resolver ciertos problemas que se presentan en neurofisiología. Los resultados obtenidos proporcionaron una solución matemática adaptativa; en la cual, a partir de la información obtenida, los autores intentaban obtener la posición angular y la velocidad del movimiento. Estas investigaciones estimularon la aplicación del BSS en la solución de problemas tales como: la supresión de ruido; la codificación eficiente de música, aplicaciones médicas, mezclado de audio en estudios de grabación, robótica, mejora de la calidad de voz, cancelación de interferencias en comunicaciones satelitales (Cano et al., 2016; Stone, 2004) entre otras aplicaciones; ya que una de las múltiples ventajas que ofrecen las técnicas basadas en separación ciega de fuentes, es que éstas no necesitan de un conjunto de datos de entrenamiento y no requieren de ningún tipo de conocimiento a priori.

El proceso BSS se puede clasificar como: mezclas instantáneas, mezcla convolutivas, mezclas lineales instantáneas o mezclas no lineales (Pertila, 2013). Debido a que la naturaleza de las mezclas es diferente en cada caso, se requieren esquemas de separación diferentes, pese a que en todos los casos se considera que las señales originales son independientes entre sí y que solo se dispone de las señales mezclas arribando a los micrófonos. Este artículo aborda, únicamente el problema de separación de mezclas lineales instantáneas con ganancias desconocidas, el cual se conoce como el problema de fiesta de coctel (Tharwat et al., 2018), donde existen n fuentes sonoras captadas por un arreglo de n micrófonos, a partir de los cuales se desea obtener cada una de las fuentes (Kitamura y col., 2018; Laufer-Goldshtein y Ronen, 2018), como se muestra en la Figura 1. Con esta finalidad se han propuesto varios esquemas, siendo algunos de los más exitosos aquellos basados en el Análisis de Componentes Independientes (ICA) (Jutten et al., 2010); en el cual la mezcla se expresa como como:

$$X = A^T S, \tag{1}$$

donde

$$S = [s_1(t), s_2(t), \dots, s_n(t)]^T, \tag{2}$$

es el conjunto de señales generadas por n fuentes con muestras cada una; con $t = 1, 2, \dots, N$, A de dimensión $n \times n$ es la matriz de mezclas y

$$X = [x_1(t), x_2(t), x_3(t), \dots, x_n(t)]^T \tag{3}$$

es el conjunto de n mezclas lineales instantáneas. El objetivo de estos esquemas es obtener el inverso de la matriz A , denotado como, W , en forma tal que la estimación de las señales originales está dada como:

$$\hat{S} = W^T X. \tag{4}$$

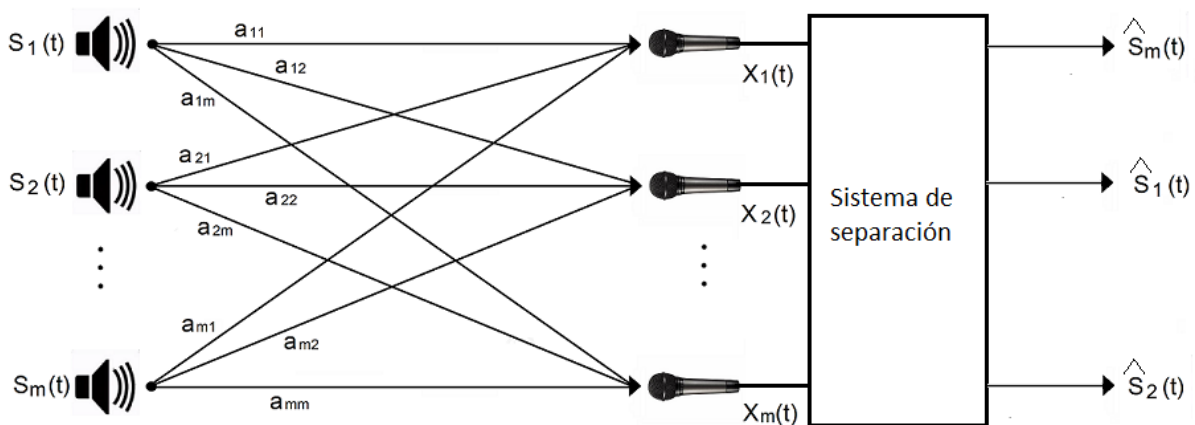


Fig. 1: Proceso de separación a ciegas de mezclas instantáneas de fuentes.

Este procedimiento permite separar cada una de estas señales y procesarlas una a una sin influencia de las demás, como si sólo una de las señales estuviera presente. Sin embargo, uno de los requerimientos principales para el adecuado funcionamiento del ICA es el hecho de que las señales mezcladas no sean de tipo Gaussiano, o que solo una de ellas lo sea. Esto limita su empleo en la separación de señales de audio ya que éstas, son en general de naturaleza Gaussiana.

El método ICA se puede aplicar a diversos problemas en donde las mezclas no son ortogonales y las señales originales no tienen características Gaussianas; mismas que poseen una gran cantidad de señales útiles. Atendiendo a las características de las señales mezcladas, existen diversas variables que pueden ser tomadas en cuenta para mejorar el desempeño del ICA. Así, es importante decidir si es necesario realizar un preprocesamiento de los datos de entrada como lo son: la reducción de la correlación, el centrado y reducción de dimensión de los datos utilizando análisis de componentes principales (PCA) Mosquera et al. (2018), Fontalvo-Herrera et al. (2018), etc. Una vez hecho esto se debe decidir la forma en la cual se estimarán las señales originales, en la mayoría de los casos por medio del inverso de la matriz de mezcla W . Para lo cual se pueden considerar algoritmos que incluyan la eliminación de las características Gaussianas, la estimación de la máxima verosimilitud, la reducción al mínimo de la información mutua, los métodos tensoriales, la decorrelación no lineal, entre otros. Una que se ha hecho lo anterior, las componentes independientes se pueden encontrar empleando algunos de los métodos reportados en la literatura (Udin et al. 2018; Stone, 2004).

Durante los últimos años, el interés en el muestreo compresivo se ha incrementado de manera importante (Eldar y Kutyniok, 2012), pues esta rama emergente del procesamiento de señales permite muestrear y comprimir simultáneamente, entre otras, las señales de audio (Moreno et al., 2014). El empleo del muestreo compresivo permite reemplazar el modelo tradicional limitado en banda planteado por el teorema de Nyquist por un modelo disperso, donde se asume el hecho de que un número pequeño de proyecciones lineales no adaptativas sobre una señal compresible, contienen suficiente información para reconstruirla y procesarla (Candes y Wakin, 2008; Donoho, 2006), lo que da como resultado una gran reducción de las tasas de muestreo y por lo mismo en complejidad computacional requerida para la estimación de las componentes principales de las señales mezcladas.

Tomando en cuenta la reducción de las tasas de muestreo y el hecho de que el muestreo compresivo permite modificar las características Gaussianas de las señales de audio, varios esquemas para llevar a cabo la separación ciega de fuentes se han desarrollado empleando el muestreo compresivo (Xu et al., 2014, Karahanoglu y Erdogan, 2013; Xu y Wang, 2010, Xu, 2009, Eldar y Kutyniok, 2012; Bao, 2013). Entre ellos tenemos el algoritmo desarrollado por Karahanoglu y Erdogan (Karahanoglu y Erdogan, 2014), en el cual inicialmente a las señales mezcladas se les aplica el muestreo compresivo para modificar las características Gaussianas de las mismas. Seguidamente las señales muestreadas se descomprimen y posteriormente se aplica el ICA para estimar la matriz de mezcla W , resolviendo con ello, el problema BSS. Si bien este esquema modifica significativamente la naturaleza Gaussiana de las mezclas, la complejidad del sistema es en general alta dado que implica la descompresión de las mezclas y la solución del ICA, empleando señales de dimensión N . Otro esquema de BSS basado en muestreo compresivo e ICA fue propuesto por Xu et al. (2014), en el cual se aplica el muestreo compresivo distribuido (Baron et al., 2009) a las mezclas de entrada. Posteriormente usando las mezclas comprimidas se calcula el inverso de la matriz de mezcla directamente, sin reconstruir las mezclas originales, usando tramas de 512 puntos comprimidas a 300 puntos. Sin embargo, aunque no se reconstruyen las mezclas originales, usando el DCS las señales de mezcla se reducen únicamente de 512 a 300 puntos, al llevar a cabo la estimación de la matriz de mezclas mediante ICA.

Para resolver los problemas mencionados, este artículo propone un algoritmo basado en muestreo compresivo e ICA el cual opera con tramas de 512 y 1024 puntos, empleando para la estimación del inverso de la matriz de mezclas, un número de muestras que va desde 15 hasta un máximo de 50 puntos lo cual reduce significativamente el tiempo y costo computacional. Los resultados obtenidos muestran que el algoritmo propuesto proporciona una separación de fuentes precisa pese al bajo número de muestras empleado.

El artículo está organizado de la siguiente manera. A continuación, se presenta el desarrollo del sistema propuesto. Seguidamente se proporciona la evaluación experimental del esquema propuesto y finalmente se proporcionan las conclusiones de esta investigación.

ALGORITMO PROPUESTO

El esquema propuesto se muestra en la Figura 2, en el cual inicialmente se reciben n señales mezcladas a las cuales se les aplica la DCT (Khayam, 2003; Ramos et al., 2016) para obtener la representación dispersa de cada mezcla. A continuación, estas representaciones dispersas se comprimen empleando técnicas de

muestreo compresivo. Posteriormente se lleva a cabo la estimación del inverso de la matriz de mezclado, mediante el ICA usando las versiones comprimidas de las mezclas. Esto permite reducir el espacio en la memoria y el tiempo de procesamiento requerido. Finalmente, las señales originales se recuperan realizando la multiplicación del vector de mezclas por la matriz de mezcla inversa.

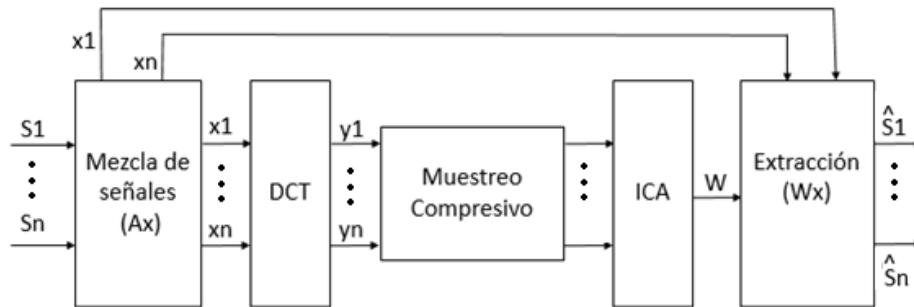


Fig. 2: Diagrama a bloques del algoritmo propuesto

Muestreo Compresivo

La teoría del muestreo compresivo subyace sobre dos principios: el de dispersión y el de incoherencia y con estos. Establece que un vector disperso $X^0 \in R^N$ puede ser recuperado a partir de un conjunto pequeño de mediciones no lineales y aleatorias (incoherentes) de la forma:

$$b = \Phi X^0, \quad K \ll N, \tag{5}$$

donde K es el número de muestras diferentes de cero, en una señal dispersa o compresible. Aunque las señales de audio no son dispersas de origen, estas si son compresibles, lo que significa que se pueden trasladar a otro dominio y con esto dispersar, mediante el empleo de la transformada Coseno Discreta (DCT por sus siglas en inglés). Así, la DCT de la j -ésima mezcla está dada por (Khayam, 2003):

$$f_j(u) = \alpha(u) \sum_{t=0}^{N-1} x_j(t) \cos\left[\frac{\pi(2t+1)u}{2N}\right], \quad u = 0,1,2,\dots,N-1, \tag{6}$$

donde

$$\alpha(u) = \begin{cases} \sqrt{1/n}, & u = 0, \\ \sqrt{2/n}, & u \neq 0. \end{cases} \tag{7}$$

De esta forma, es posible obtener una señal compresible en el dominio de la frecuencia (Khayam, 2003). Una vez obtenidas las señales dispersas, $f_j(u), j=1,2,\dots,n$ se genera un vector F de dimensión $N \times 1$ es cual es una representación dispersa de las señales mezcladas

$$X = [x_1(t), x_2(t), x_3(t), \dots, x_n(t)]^T. \tag{8}$$

En términos de notación matricial se generan

$$y = \Phi F \tag{9}$$

mediciones no lineales y aleatorias (incoherentes), donde F es la versión dispersa de X , en la cual solo $K \ll N$ entradas son distintas de cero y Φ corresponde a la matriz de sensado. La teoría del muestreo compresivo, como se muestra en la Figura 3, establece que una señal dispersa con K elementos distintos de cero, puede ser reconstruida a partir de un vector comprimido con solo $M \geq \log N$ mediciones lineales de la forma:

$$y = \Phi F, \tag{10}$$

donde y es un vector de dimensión $M \times 1$ mientras que Φ representa una matriz de dimensión de $M \times N$, la cual es incoherente en base a las funciones de la DCT (Candes y Wakin, 2008; Donoho y Huo, 2001). Finalmente, con esta información, en caso de ser necesario, se procede a recuperar la señal por medio de la minimización

de la norma l_1 . Así, si F es suficientemente dispersa, la recuperación por medio de la minimización de la norma l_1 es probablemente exacta, debido a que intrínsecamente la norma favorece la dispersión de señales, por lo que para una matriz Φ de dimensión $M \times N$ y un vector $y \in R^M$ la solución dada por:

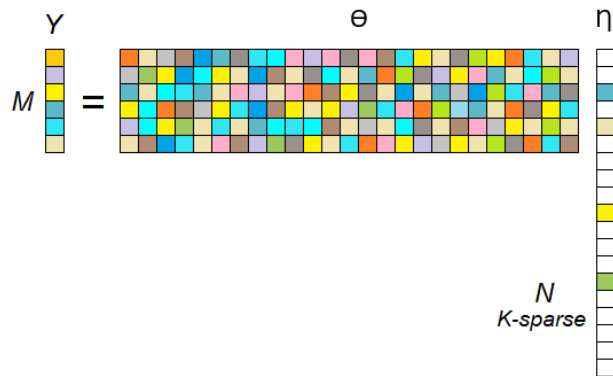


Fig. 3: Mediciones en muestreo compresivo

$$\min_{x \in R^p} \|y - \Phi F\|_{l_1}, \tag{11}$$

De una manera concreta el algoritmo selecciona un conjunto de n vectores linealmente independientes, pertenecientes a una base de funciones ϕ , para efectuar la reconstrucción.

Estimación de W

Suponga que W es una matriz de $M \times M$ la cual permite recuperar las señales originales a partir de sus mezclas;

$$\hat{S} = [\hat{s}_1(t), \hat{s}_2(t), \hat{s}_3(t), \dots, \hat{s}_n(t)]. \tag{12}$$

Es el conjunto de señales recuperadas y

$$Y = [y_1(t), y_2(t), y_3(t), \dots, y_n(t)]^T, \tag{13}$$

es el conjunto de mezclas obtenidas a partir de la aplicación del muestreo compresivo a las señales mezcladas. Existen diversos algoritmos que permite obtener W a partir del conjunto de mezclas Y . Uno de los más usados es el algoritmo del gradiente ascendente (Udin et al., 2018; J. Stone, 2004) el cual estima la matriz W óptima maximizando una a una la entropía de las componentes de \hat{S} , donde la entropía de la i -ésima componente de la j -ésima muestra recuperada está dada por

$$H(S) = H(F) + E \left[\sum_{i=1}^n \ln (P_s(\hat{S}_i)) \right] + \ln(|W|), \tag{14}$$

donde $P_s(S)$ es la función de densidad de probabilidad (PDF) de S ; siendo la ij -ésima componente de W en la iteración $t+1$ esta dada por

$$W_{ij}(t + 1) = W_{ij}(t) + \mu \nabla(i, j), \quad i = 1, 2, \dots, n, j = 1, 2, \dots, n. \tag{15}$$

Dado que la entropía de F , $H(F)$, no se ve afectada por W su contribución al gradiente puede ser ignorada así

$$\nabla(i, j) = \frac{\partial h(\hat{S}_i)}{\partial W_{i,j}}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, n. \tag{16}$$

Dado que

$$h(S_i) = E \left[\sum_{i=1}^n \ln (P_s(\hat{S}_i)) \right] + \ln(|W|), \tag{17}$$

y tomando en cuenta que la PDF es la derivada de la función de probabilidad acumulativa de \mathbf{S} , esto es

$$P_s(\hat{S}_i) = g'(\hat{S}_i), \tag{18}$$

de (16) y (17) se obtiene

$$W_{ij}(t+1) = W_{ij}(t) + E \left[\sum_{i=1}^n \frac{g''(\hat{S}_i)}{g'(\hat{S}_i)} y_j \right] + (W^T)_{ij}^{-1}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, n. \tag{19}$$

Definiendo $K_j(t)$ como

$$K_j(t) = \sum_{i=1}^n \frac{g''(\hat{S}_i(t))}{g'(\hat{S}_i(t))} y_j(t) \tag{20}$$

y reemplazando el valor esperado por su promedio temporal se obtiene

$$W_{ij}(t+1) = W_{ij}(t) + (W^T)_{ij}^{-1} + \frac{1}{T} \sum_{t=1}^T K_j(t), \quad i = 1, 2, \dots, n, j = 1, 2, \dots, n \tag{21}$$

donde T es el tamaño de la ventana. Finalmente (21) se puede reescribir como

$$W_{ij}(t+1) = W_{ij}(t) + (W^T)_{ij}^{-1} + G_{i,j}(t), \quad i = 1, 2, \dots, n, j = 1, 2, \dots, n. \tag{22}$$

Donde $G_{i,j}(t)$ se puede estimar en forma recursiva como

$$G_{i,j}(t) = \left(1 - \frac{1}{T}\right) G_{i,j}(t-1) + \frac{1}{T} K_j(t). \tag{23}$$

Tomando en cuenta que la PDF de una señal súper-Gaussiana puede ser representada como

$$g(\hat{S}_i) = \tanh(\hat{S}_i), \tag{24}$$

y que la PDF de una señal sub-Gaussiana se puede representar como

$$g(\hat{S}_i) = \hat{S}_i^3, \tag{25}$$

después de algunas operaciones, de (14) se obtiene que; cuando las señales mezcladas tienen una PDF de tipo súper-Gaussiana $K_j(t)$ esta dada por

$$K_j(t) = -2 \tanh(\hat{S}_i(t)). \tag{26}$$

Por otro lado, cuando las señales de entrada tienen PDFs de Tipo sub-Gaussiano, $K_j(t)$ estará dado por

$$K_j(t) = 3 \hat{S}_i(t). \tag{27}$$

Una vez estimado el inverso de la matriz mezclas W , las señales recuperadas se obtienen haciendo

$$\hat{S}(t) = W^T X(t). \tag{28}$$

Complejidad computacional

Considere que se tienen n señales mezcladas, cada una con N muestras de entrada por trama, las cuales se reducen a M muestras por trama usando técnicas de muestreo compresivo. Con este fin inicialmente se calcula la DCT de la trama de N muestras. lo cual tiene una complejidad del orden de $O(N \log_2 N)$. Seguidamente el vector obtenido se multiplica por una matriz de sensado de $M \times N$, lo cual tiene una complejidad del orden de

$O(MN)$. Tomando en cuenta que se tienen n mezclas, el cálculo del muestreo compresivo de las n mezclas tiene una complejidad de

$$O(nN \log_2 N + nMN). \quad (29)$$

Para estimar la matriz W , por medio del ICA se requiere un número de operaciones del orden de

$$O(M(2n^3 + n)). \quad (30)$$

Así para estimar la matriz W se requiere un número de operaciones del orden de

$$O(2Mn^3 + Mn + nN \log_2 N + nMN). \quad (31)$$

Tomando en cuenta que el número de muestras de entrada es mucho mayor que el número de mezclas se tiene que el orden del número de operaciones requeridas por muestra está aproximadamente dado por

$$O(Mn + n \log_2 N). \quad (32)$$

La Tabla 1 muestra una comparación entre la complejidad computacional del algoritmo propuesto y aquellos propuestos por (Xu et al., 2014; D. Baron et al., 2009).

Tabla 1. Complejidad computacional por muestra del sistema propuesto compara con los algoritmos de Xu y col. y Baron y col.

| Propuesto | | | | Xu y Col. | | Baron y Col. | |
|-----------|-----|------|-----|-----------|-----|--------------|-------|
| M=15 | | M=50 | | | | | |
| n=2 | n=3 | n=2 | n=3 | n=2 | n=3 | n=2 | n=3 |
| 50 | 75 | 120 | 180 | 224 | 236 | 20808 | 30000 |

RESULTADOS Y DISCUSIÓN

Para evaluar el algoritmo propuesto, se generó una base de datos, que consiste en 100 señales de audio con 35,000 muestras cada una. Posteriormente, se propusieron dos matrices de mezcla A para efectuar las mezclas con 2 y 3 señales cuyas dimensiones son 2×2 y 3×3 respectivamente. A continuación, se dividió cada señal de mezcla en tramas de 512 muestras para un primer experimento y posteriormente 1024 para un segundo experimento, para después aplicar el algoritmo propuesto a cada una de las tramas, evaluando en esta etapa la naturaleza Gaussiana de las señales de audio por medio de la medición de la kurtosis normalizada dada por (25). Finalmente se evalúa la calidad de las señales recuperadas.

Cabe mencionar que se realizaron 10 experimentos en los cuales, cada gráfica muestra el resultado de kurtosis y correlación obtenido para la señal mencionada al encontrarse en una mezcla aleatoria de 2 señales ó 3 señales según sea el caso, sin posibilidad de repetir alguna señal de la base de datos.

Evaluación de las características Gaussianas de las señales de entrada.

Siendo requisito para poder usar ICA en la separación ciega de fuentes, que las señales mezcladas sean no gaussianas o cuando mucho solo una de ellas lo sea, se evalúa la kurtosis de cada mezcla, después de aplicar el muestreo compresivo, como método para conocer la naturaleza Gaussiana de las señales. Por lo que, para llevar a cabo la estimación de la kurtosis, se emplea la siguiente relación:

$$kurt(s) = \frac{E\{s^4\}}{(E\{s^2\})^2} - 3. \quad (33)$$

Donde $E\{s^2\}$ y $E\{s^4\}$ son el segundo y el cuarto momento de las mezclas de entrada. Cuando el resultado es igual a 0 implica que la señal bajo análisis, s , es Gaussiana. Por su parte cuando el resultado es mayor a 0, s sigue una distribución súper-Gaussiana, y finalmente, si el resultado es menor a 0, s sigue una distribución sub-Gaussiana. Para este trabajo, utilizamos señales de voz y música de naturaleza gaussiana, de tal forma que nos permiten evaluar el desempeño del muestreo compresivo en la reducción de las características Gaussianas de este tipo de señales. En las Figuras 4 y 5 se muestran los valores de la kurtosis obtenidos usando las señales de audio empleadas, cada una con 35,000 muestras, segmentadas en tramas de 512 y 1024 muestras, respectivamente. En estas figuras se muestra la kurtosis obtenida cuando las tramas de las

señales de entrada de dimensión 512 o 1024, se comprimen en tramas de 15 hasta 50 muestras, usando el muestreo compresivo, la kurtosis aplicada a la trama completa usando muestreo compresivo con factor de compresión igual a 1, también se muestra para comparación para tramas de 512 y 1024 muestras. Los resultados obtenidos muestran que, con el empleo del CS, se modifica la naturaleza Gaussiana de las señales utilizadas, principalmente en las señales de voz, haciéndolas sub-Gaussianas o super-Gaussianas según sea el caso de manera que la herramienta ICA puede ser empleada.

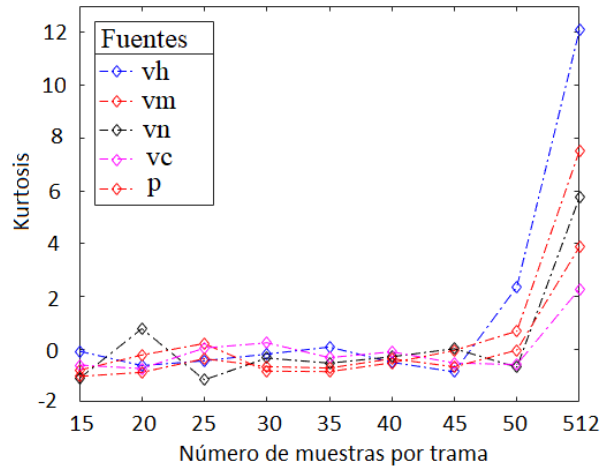


Fig. 4(a): kurtosis normalizada después de aplicar CS a las mezclas para tramas de 512 puntos, donde vh denota voz de hombre, vm voz de mujer, vn vons de niño, vc voz cantada y p percusiones

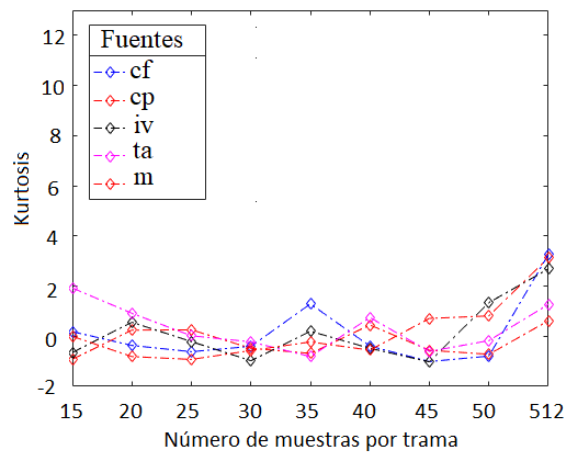


Fig. 4(b): kurtosis normalizada después de aplica CS a las mezclas para tramas de 512 puntos, donde cf denota cuerdas flotantes, cp cuerdas percutidas, iv instrumentos de viento, ta tonos agudos y m melodías.

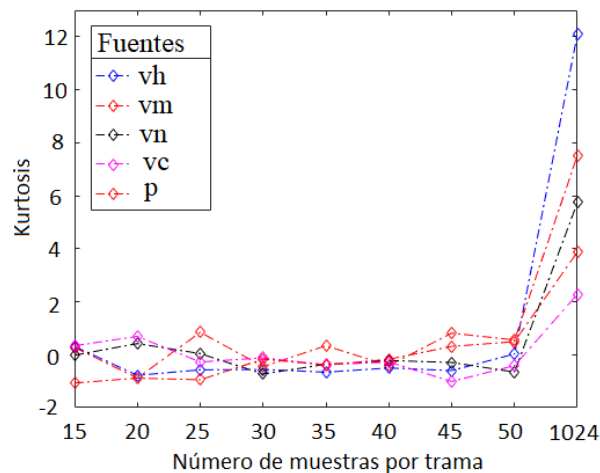


Fig. 5(a): kurtosis normalizada después de aplica CS a las mezclas para tramas de 1024 puntos, donde vh denota voz de hombre, vm voz de mujer, vn vons de niño, vc voz cantada y p percusiones.

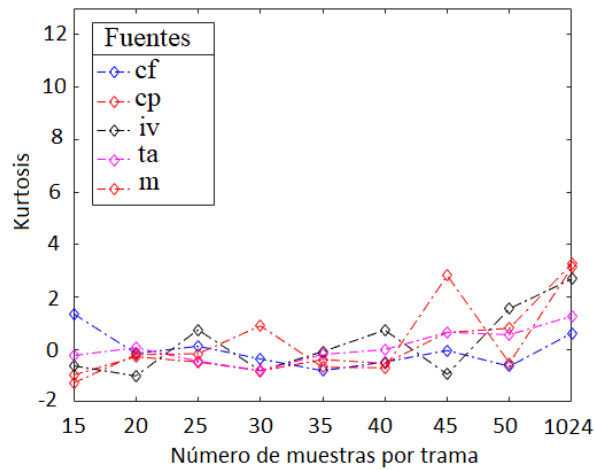


Fig. 5(b): kurtosis normalizada después de aplica CS a las mezclas para tramas de 1024 puntos, donde cf denota cuerdas flotantes, cp cuerdas percutidas, iv instrumentos de viento, ta tonos agudos y m melodías.

Similitud entre las señales originales y las señales recuperadas.

Para cumplir con el objetivo planteado, el cual consiste en encontrar el inverso de la matriz A, y con el fin de realizar una comparación con el artículo publicado por Xu et al. (2014), y evaluar la viabilidad del algoritmo propuesto, se emplearán las siguientes matrices de mezcla A₁ y A₂, para 2 y 3 señales, respectivamente:

$$A_1 = \begin{bmatrix} 0.9222 & -0.3867 \\ 0.4509 & -0.8926 \end{bmatrix} \tag{34}$$

$$A_2 = \begin{bmatrix} -0.1748 & 0.8197 & 0.1552 \\ -0.9706 & 0.0765 & 0.5500 \\ -0.1657 & 0.5677 & -0.8206 \end{bmatrix} \tag{35}$$

A continuación, para evaluar la calidad de las señales recuperadas por el algoritmo propuesto, se empleó la correlación cruzada (Mosquera et al., 2018) entre la señal original y la señal recuperada, la cual está dada por:

$$R_{s\hat{s}} = \sum_{t=1}^N \frac{(S_1(t) - \bar{S}_1) (S_2(t) - \bar{S}_2)}{\| (S_1(t) - \bar{S}_1) \| \| (S_2(t) - \bar{S}_2) \|} \tag{36}$$

Donde S₁ denota el valor promedio de S₁(t).

La Figura 6 muestra la correlación entre las señales recuperadas y las señales originales separadas de la mezcla de 2 fuentes, cuando el inverso de la matriz de mezclas se obtiene usando muestreo compresivo con un número de muestras que oscila entre 15 y 50 muestras para tramas de 512 puntos. Por su parte la Figura 7 muestra la correlación entre las señales recuperadas y originales cuando el inverso de la matriz de mezclas de dimensión de 2x2 se obtiene usando muestreo compresivo con un número de muestras que oscila entre 15 y 50 muestras para tramas de 1024 muestras. Las Figuras 8 y 9 muestran la capacidad del sistema propuesto cuando se requiere recuperar tres señales mezcladas entre sí. Esto es, se requiere que el sistema estime la matriz inversa de dimensión 3x3, que permita la recuperación de las señales mezcladas, de manera que la correlación cruzada entre las señales recuperadas y las originales, sea superior de 0.9, cuando se emplean tramas ya sea de 512 o 1024 muestras. En ambos casos las señales se representaron por medio de 15 hasta 50 muestras obtenidas usando muestreo compresivo. Analizando los valores de correlación cruzada obtenidos cuando el sistema propuesto se emplea para separar dos y tres señales mezcladas, podemos observar que en ambos casos el valor de la correlación se mantiene alto, asegurando de esta forma que las señales separadas son claramente audibles. Así bien, el algoritmo arroja buenos resultados para tramas de 512 y 1014 puntos cuando se requiere separar ya sean dos o tres señales mezcladas entre sí.

Como se puede analizar en los datos mostrados en las Figuras 6-9, en algunos casos la correlación cruzada entre las señales originales y las recuperadas es ligeramente superior a 0.7, cuando las señales son de baja frecuencia y la tasa de compresión es alta, (512/15) o (1024/15) cuando se emplean 512 o 1024 muestras por trama respectivamente. Sin embargo, aunque las señales recuperadas son inteligibles; su calidad mejora notablemente cuando la tasa de compresión disminuye a (512/50) ó (1024/50), respectivamente.

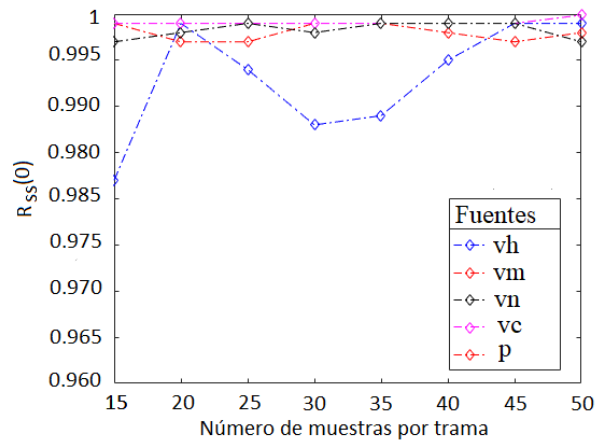


Fig. 6(a): Correlación para tramas de 512 puntos en mezcla de 2 señales, donde vh denota voz de hombre, vm voz de mujer, vn voz de niño, vc voz cantada y p percusiones.

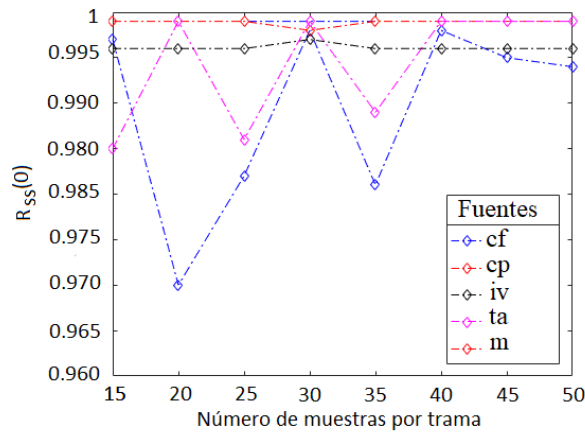


Fig. 6(b): Correlación para tramas de 512 puntos en mezcla de 2 señales, donde cf denota cuerdas flotantes, cp cuerdas percutidas, iv instrumentos de viento, ta tonos agudos y m melodías.

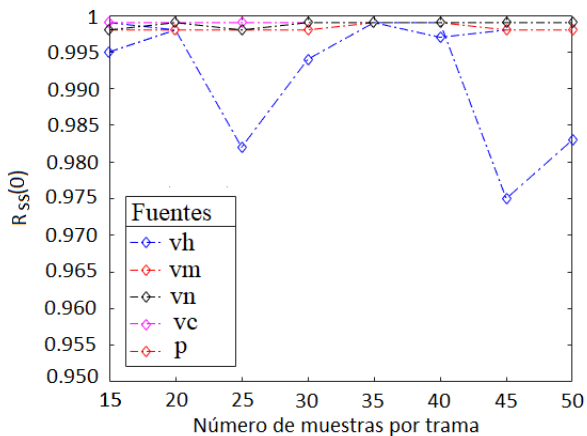


Fig. 7(a): Correlación para tramas de 1024 puntos en mezcla de 2 señales, donde vh denota voz de hombre, vm voz de mujer, vn voz de niño, vc voz cantada y p percusiones.

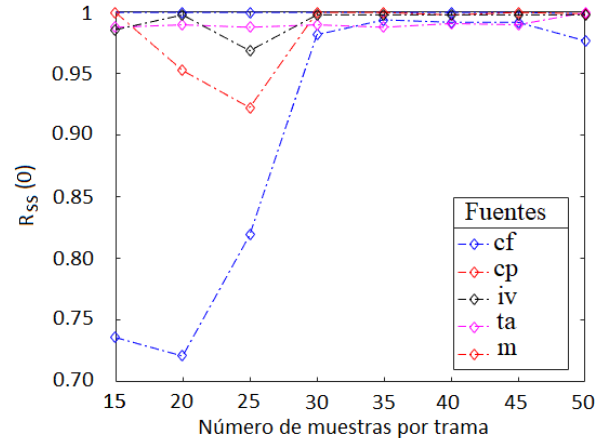


Fig. 7(b): Correlación para tramas de 1024 puntos en mezcla de 2 señales, donde cf denota cuerdas flotantes, cp cuerdas percutidas, iv instrumentos de viento, ta tonos agudos y m melodías.

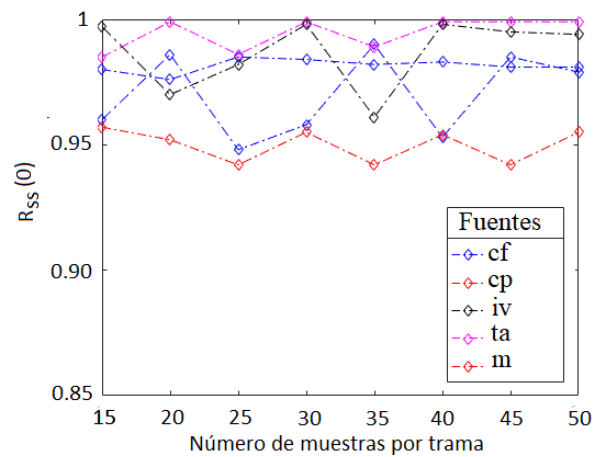
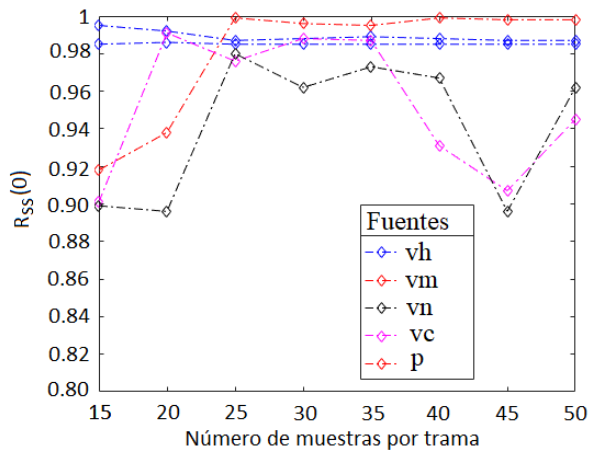


Fig. 8(a): Correlación para tramas de 512 puntos en mezcla de 3 señales, donde vh denota voz de hombre, vm voz de mujer, vn voz de niño, vc voz cantada y p percusiones.

Fig. 8(b): Correlación para tramas de 512 puntos en mezcla de 3 señales, donde cf denota cuerdas flotantes, cp cuerdas percutidas, iv instrumentos de viento, ta tonos agudos y m melodías.

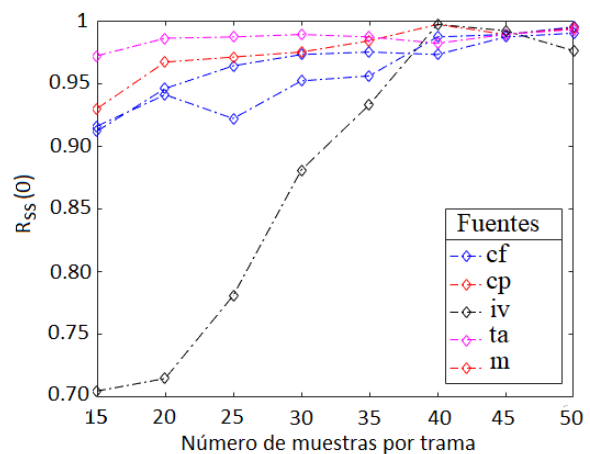
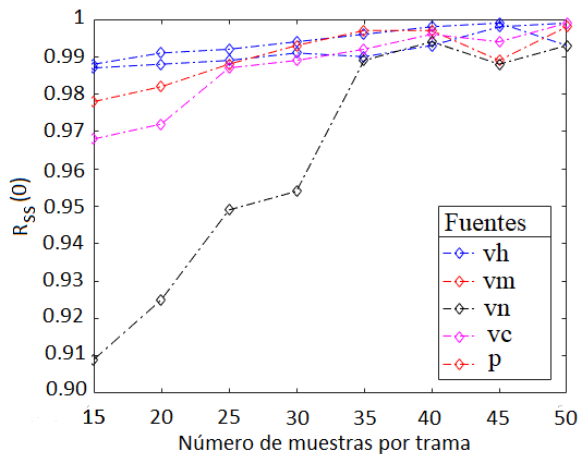


Fig. 9(a): Correlación para tramas de 1024 puntos en mezcla de 3 señales, donde vh denota voz de hombre, vm voz de mujer, vn voz de niño, vc voz cantada y p percusiones.

Fig. 9(b): Correlación para tramas de 1024 puntos en mezcla de 3 señales, donde cf denota cuerdas flotantes, cp cuerdas percutidas, iv instrumentos de viento, ta tonos agudos y m melodías.

Así si se incrementa la complejidad computacional del sistema es posible mejorar la calidad de las señales recuperadas. Esto implica que, para cierto tipo de señales, es necesario un mayor número de muestras para poder calcular el inverso de la matriz de mezcla, de forma tal, que el valor de la correlación entre las señales original y recuperada este por encima de 0.5, a pesar de que éstas están contaminadas por señales inter-clase, es decir, señales que corresponden a otras fuentes involucradas en la mezcla. El índice de correlación cruzada obtenido es similar al reportado por otros esquemas con una complejidad computacional mucho mayor (Calderón-Piedras et al., 2014). De esta forma la señal de entrada del algoritmo puede ser cualquier señal de audio como voz, instrumentos musicales, etc. En el siguiente ejemplo, Figura 10, se muestran los resultados obtenidos de la separación de dos señales de audio mezcladas, las cuales corresponden a una voz de mujer y a una voz de hombre. Por su parte, en la Figura 11 se muestran los resultados de la separación de tres señales de audio mezcladas, las cuales corresponden a dos voces (una voz de mujer y una voz de hombre) y a un instrumento musical.

Como se puede analizar en las imágenes anteriores (Figuras 10 y 11), las cuales representan una trama de la separación de 2 y 3 señales mezcladas, se puede visualizar que las señales recuperadas son muy parecidas a las señales originales, sin embargo, cabe mencionar que una característica de ICA es que esta herramienta no puede determinar la varianza de las componentes independientes que conforman la mezcla, es decir, que se presenta una diferencia en la amplitud original de las fuentes sonoras. Otro punto que mencionar es que cuando una de las señales mezcladas contiene silencios, en la etapa de separación, estos silencios presentarán sonidos inter-clase, sin embargo, la señal predominante es claramente inteligible.

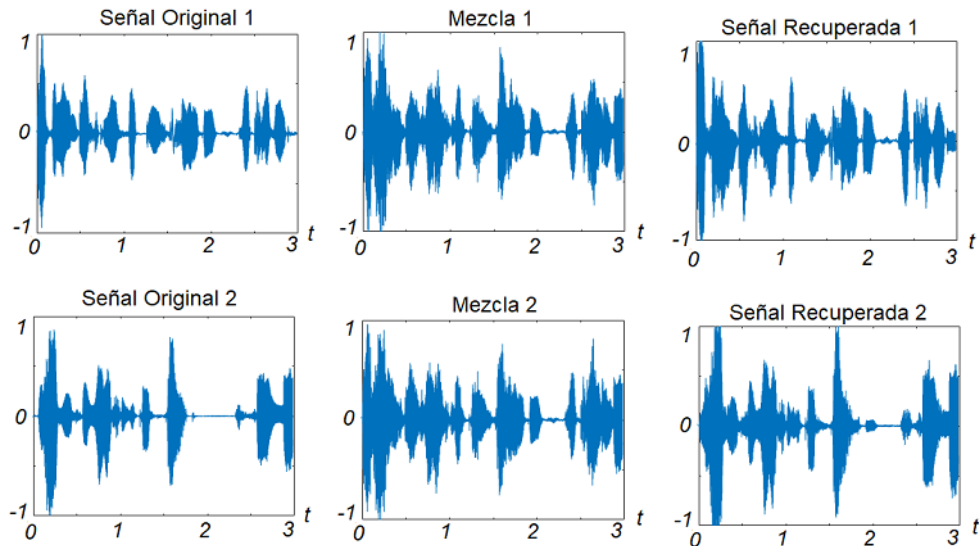


Fig. 10: Separación de voces de mujer y hombre usando ICA y muestreo compresivo

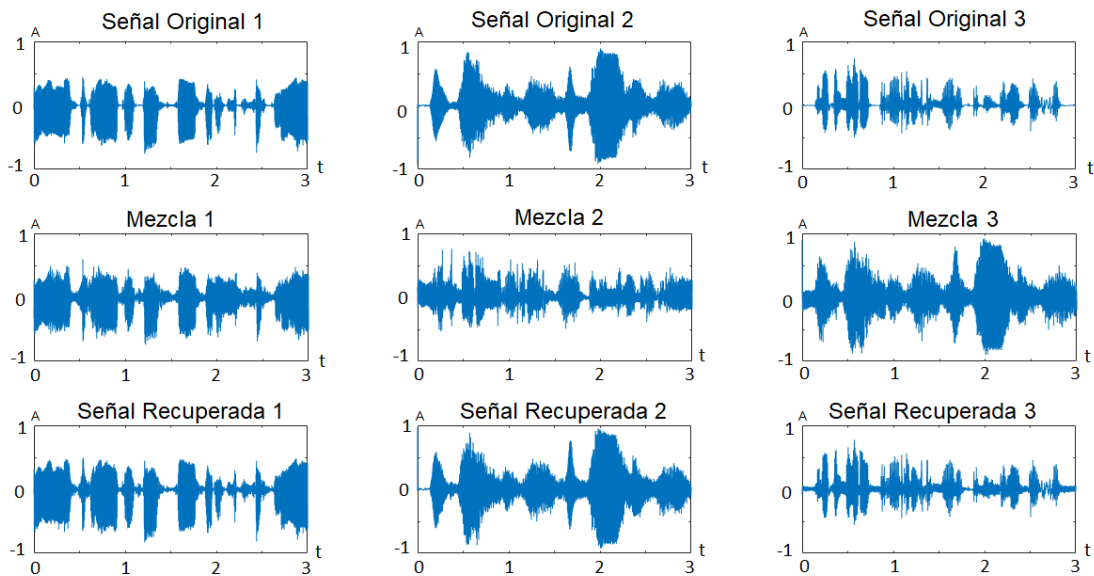


Fig. 11: Separación de dos voces y un instrumento musical usando ICA y muestreo compresivo

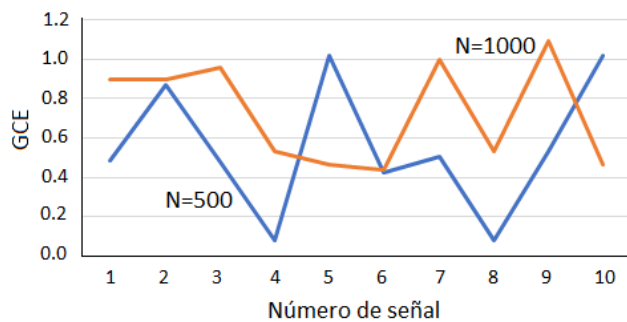


Fig. 12: GCE obtenido cuando la matriz de mezclas se obtiene usando las señales mencionadas anteriormente.

Otro parámetro que puede ser empleado para evaluar la precisión del algoritmo propuesto, es el GCE (error de interferencia generalizada) la cual está dada por (Xu et al., 2014)

$$GCE = \min_{B \in \Pi} \|A - A^*\| \tag{29}$$

Donde Π es el conjunto de todas las matrices invertibles de dimensión $n \times n$. Aquí si A es equivalente a A^* , GCE tiende a cero, por lo cual GCE se puede emplear para comparar la estimación obtenida por dos

esquemas diferentes. Así entre más pequeño sea el valor del GCE, mayor parecido tienen la matriz original a la matriz estimada. La Figura 12 muestra el GCE obtenido por el sistema propuesto cuando se emplea para separar tres señales mezcladas, con tramas de 512 y 1024 muestras respectivamente. En este caso se muestra en el eje de las abscisas el número de la señal, siendo el mismo orden que en las figuras anteriores donde 1 es asignado a la voz de hombre y 10 es asignado al bajo. Estas señales se combinaron con las otras 9 aleatoriamente y sin repetición para evaluar el sistema propuesto. Por lo que podemos observar que el menor valor obtenido en nuestro esquema se encuentra entre 0.0781 y 1.09, mientras que el mínimo valor reportado por Xu et al. (2014) es de 1.81. Por lo cual el sistema propuesto con mediciones mínimas proporciona una mejor estimación que el propuesto por Xu et al. (2014).

La calidad de las señales separadas usando el sistema propuesto puede ser evaluada también mediante el uso de las métricas siguientes: la Relación Señal a Distorsión (SDR), la Relación Señal a Interferencia (SIR) y la Relación Señal a Artefacto (SAR) (Vincent E. et al. 2008; 2010; 2014). La Tabla 3 muestra los resultados obtenidos empleando el sistema propuesto, así como los resultados obtenidos empleando el esquema propuesto por Mirzaei et al. (2015) para comparación. Los resultados obtenidos muestran, que salvo en el caso del SIR, el sistema propuesto proporciona mejores resultados que otros esquemas previamente reportados.

Tabla 2. Métricas usadas para evaluar el sistema propuesto y su comparación y el esquema reportado en Mirzaei (2015)

| Esquema | SDR | SIR | SAR |
|--------------------------|-------|------|------|
| Propuesto | 13.69 | 13.7 | 36.6 |
| Complex NMF | 12.8 | 18.1 | 14.3 |
| Binary Masking | 8.9 | 17.7 | 10.1 |
| l_0 -norm minimization | 6.1 | 10.7 | 9.5 |

CONCLUSIONES

En el presente trabajo se desarrolló un método de separación ciega de fuentes acústicas, a partir de grabaciones de mezclas instantáneas basado en Análisis de Componentes Independientes y reduciendo la complejidad computacional del muestreo compresivo a partir de realizar las mínimas mediciones requeridas para modificar la naturaleza gaussiana de la señal. A través de este método, se logró obtener la separación de varias fuentes mezcladas simultáneamente, dando como resultado señales estimadas con una fuente predominante.

Uno de los aportes más importantes de la presente investigación es que a partir de las muestras comprimidas del algoritmo, dadas por las mediciones mínimas, se estimó la matriz de mezcla directamente de la compresión, sin necesidad de reconstruir la mezcla, con lo cual se ahorra tiempo de procesamiento y se disminuye el costo computacional sin afectar la calidad de la separación de las señales. A diferencia de otros algoritmos que emplean únicamente ICA como herramienta, este método contribuye a lograr con éxito la separación de señales gaussianas y sub-Gaussianas, por medio del empleo del muestreo compresivo. A su vez, se logra el análisis y el procesamiento independiente de cada una de las fuentes involucradas. Algunas aplicaciones de este algoritmo en otros campos del procesamiento de audio pueden ser: el reconocimiento de voz, la ecualización y la acústica forense. Se propone como trabajo a futuro, debido al número pequeño de muestras, trabajar en tiempo real, así como emplear las señales recuperadas para desarrollar esquemas de clasificación de señales que se encuentran mezcladas entre sí.

AGRADECIMIENTOS

Los autores agracen al Instituto Politécnico Nacional (IPN) de México y al Consejo Nacional de Ciencia y Tecnología (CONACyT) de México por su apoyo económico durante la realización de esta investigación.

REFERENCIAS

- Bao G., Z. Ye, X. Xu, y Y. Zhou, *A Compressed Sensing Approach to Blind Separation of Speech Mixture Based on a Two-layer Sparsity Model*, doi: 10.1109/TASL.2012.2234110, IEEE Trans. Audio, Speech, Lang. Processing, 21(5), 899–906 (2013)
- Baron D., M. Duarte, M. Wakin, S. Sarvotham y otro autor, *Distributed Compressive Sensing*. arXiv:0901.3403[cs.IT] (2009)

- Calderon-Piedras J., A. Orijuela-Cañon y D. Sanabria-Quiroga, *Blind Source Separation from Single Channel Audio Recording using ICA Algorithms*. Int. Symp. on Image, Signal Processing and Artificial Vision, Armenia, Colombia, 1-5 (2014).
- Candes E. y M. Wakin, *An Introduction to Compressive Sampling*. IEEE Signal Processing Magazine, 25(2), 47-21, (2008)
- Cano E., D. Fitzgerald y K. Brandenburg, *Evaluation of Quality of Sound Source Separation Algorithms: Human Perception vs Quantitative Metrics*, 24th European Signal Processing Conference, Budapest, Hungría, 1758–1762, (2016)
- Donoho D. L., *Compressed Sensing*, IEEE Trans. on Information Theory 52(4), 1289–1306, (2006)
- Eldar Y. y G. Kutyniok, *Compressive Sensing: Theory and Applications*, Cambridge University Press, New York, USA (2012)
- Foncalvo-Herrera T., E. Delahoz y A. Mendoza-Mendoza, *Aplicación de Minería de Datos para la Clasificación de Programas Universitarios de Ingeniería Industrial Acreditados en Alta Calidad en Colombia*, doi: 10.4067/S0718.07642018000300089, Información Tecnológica 29(3), 89-96 (2018)
- Herault J. y C. Jutten, *Detection de Grandeurs Primitives Dans un Message Composite par une Architecture de Calcul Neuromimetique en Apprentissage Non Supervise*, X Colloque GRETSI, Nice, Francia, 1017-1022, (1985)
- Jutten C. y J. Herault, *Independent Component Analysis Versus Principal Component Analysis*, European Signal Processing Conference, Roma, Italia, 643-646 (1988)
- Karahanoglu N. y H. Erdogan, *Compressed Sensing Signal Recovery via Forward-backward Pursuit*, doi: 10.1016/j.dsp.2013.05.007, Digital Signal Processing, 23(5), 1539–1548 (2013)
- Khayam A., *The Discrete Cosine Transform (DCT): Theory and Application*, Department of Electrical and Computer Engineering Michigan State University (2003)
- Kitamura D., S. Mogami, Y. Mitsui, N. Takamune y otros cuatro autores, *Generalized Independent Low Rank Matrix Analysis Using Heavy-Tailed Distributions for Blind Source Separation*. doi: 10.1186/s13634-018-0549-5, EURASIP Journal on Advances in Signal Processing, 28(1), 1–25, (2018)
- Laufer-Goldshtein B. y T. Ronen, *Source Counting and Separation Based on Simplex Analysis*. doi: 10.1109/TSP.2018.2876349, IEEE Transactions on Signal Processing, 66(24), 6458-6473 (2018)
- Mirzaei S., H. Van Hamme y Y. Norouzi, *Blind Audio Source Counting and Separation of Anechoic Mixtures Using Multichannel Complex NMF Framework*, doi:10.1016/j.sigpro.2015.03.006, Signal Processing, 115, 27-37 (2015)
- Mosquera R., O. Castrillón y L. Parra, *Predicción de Riesgos Psicosociales en Docentes de Colegios Públicos Colombianos Utilizando Técnicas de Inteligencia Artificial*, doi: 10.4667/S0718-07642018000400267, Información Tecnológica 29(4), 267-280 (2018)
- Moreno R., M. Martínez, M. Nakano y H. Pérez, *Desarrollo de Algoritmos para Muestreo Compresivo Aplicado a Señales de Audio*, doi: 10.4067/S0718-07642014000500019, Información Tecnológica, 25(5), 137-146 (2014)
- Pertila P., *Online Blind Speech Separation Using Multiple Acoustic Speaker Tracking and Time Frequency Masking*, doi:10.1016/j.csl.2012.08.003, Journal of Computer Speech and Language 27(3), 683-702 (2013)
- Ramos O., D. Rojas y J. Saby, *Reconocimiento de Patrones Vocálicos Mediante la Implementación de una Red Neuronal Artificial Utilizando Sistemas Embebidos*, doi:10.4067/S0718-07642016000500015, Información Tecnológica, 27(5), 133-142 (2016)
- Stone B., *Independent Component Analysis: A Tutorial Introduction*, MIT Press, Cambridge Mas. (2004)
- Tharwat A., T. Gaber, A. Ibrahim y A. Hassaniien, *Independent Component Analysis: A Detailed Tutorial*, doi: 10.3233/AIC-170729, AI Communications, 30,169-190 (2018)
- Udin Z., A. Ahmad., M. Iqbal y Z. Kalem, *Adaptive Step Size Gradient Ascent ICA Algorithm for Wireless MIMO Systems*, doi: 10.1155/2018/7038531, Mobile Information Systems, Hindawi, Article ID 7038531 (2018)
- Vincent E., S. Araki y P. Bofill, *Signal Separation Evaluation Campaign: AC community-Based Approach to Large-Scale Evaluation*, Int. Conf. on Independent Component Analysis and Signal Separation (ICA), Paraty, Brasil, 734–741 (2008)
- Vincent E., M. Jafari, S. Abdallah, M. Plumbley y otro autor, *Probabilistic Modeling Paradigms for Audio Source Separation*, Machine Audition: Principles, Algorithms and Systems. IGI Global Press, 162–185, (2010)
- Vincent E., N. Bertin, R. Gribonval y F. Bimbot, *From Blind to Guided Audio Source Separation: How Models and Side Information Can Improve the Separation of Sound*, doi: 10.1109/MSP.2013.2297440, IEEE Signal Proc. Magazine 31(3), 107–115 (2014)
- Xu T. y W. Wang, *A Compressed Sensing Approach for Underdetermined Blind Audio Source Separation with Sparse Representation*, IEEE Inter. Conference on Statist. Signal Process., Taipei, Taiwan, 493–496 (2009)
- Xu T. y W. Wang, *A Block-Based Compressed Sensing Method for Undetermined Blind Speech Separation Incorporating Binary Mask*, IEEE Int. Conf. Acoustic, Speech and Signal Process., Dallas, Texas, 2022–2025 (2010)
- Xu H., N. Fu, L. Qiao, W. Yu y otro autor, *Compressive Blind Mixing Matrix Estimation of Audio Signals*, doi: 10.1109/TIM.2013.2292359, IEEE Transaction Instrumentation and Measurements, 63(5), 1253-1261 (2014)