INSTITUTO POLITÉCNICO NACIONAL

Centro de Investigación en Computación
Secretaría de Investigación y Posgrado

LABORATORIO DE TIEMPO REAL Y AUTOMATIZACIÓN

CLASIFICADOR DIFUSO DE SEÑALES ACÚSTICAS AMBIENTALES BASADO EN ANÁLISIS DE COMPONENTES INDEPENDIENTES

T E S I S

QUE PARA OBTENER EL GRADO DE

MAESTRO EN CIENCIAS EN INGENIERÍA DE CÓMPUTO CON OPCIÓN EN SISTEMAS DIGITALES

P R E S E N T A

ING. MARÍA GUADALUPE LÓPEZ PACHECO

DIRECTORES DE TESIS: DR. LUIS PASTOR SÁNCHEZ FERNÁNDEZ
DR. HERÓN MOLINA LOZANO

MÉXICO, D.F. **2011**

SIP-14 bis

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 13:30 horas del día 25 del mes de noviembre de 2011 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

"Clasificador difuso de señales acústicas ambientales basado en análisis de componentes independientes"

Presentada por la alumna:

LÓPEZ	**PACHECO**	**MARÍA GUADALUPE**
Apellido paterno	Apellido materno	Nombre(s)

Con registro:

B	0	9	1	6	7	9

aspirante de: **MAESTRÍA EN CIENCIAS EN INGENIERÍA DE CÓMPUTO CON OPCIÓN EN SISTEMAS DIGITALES**

Después de intercambiar opiniones los miembros de la Comisión manifestaron ***APROBAR LA TESIS***, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de Tesis

Dr. Luis Pastor Sánchez Fernández

Dr. Herón Molina Lozano

Dr. Sergio Suárez Guerra

Dr. Oleksiy Pogrebnyak

Dr. Luz Noé Oliva Moreno

Dr. Víctor Hugo Ponce Ponce

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Luis Alfonso Villa Vargas

INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de *México, Distrito Federal* el día *25* del mes de *Noviembre* del año *2011*, la que suscribe *María Guadalupe López Pacheco* alumna del Programa de *Maestría en Ciencias en Ingeniería de Cómputo con opción en sistemas digitales* con número de registro *B091679*, adscrito al *Centro de Investigación en Computación*, manifiesta que es autora intelectual del presente trabajo de Tesis bajo la dirección de *Dr. Luis Pastor Sánchez Fernández* y *Dr. Herón Molina Lozano* y cede los derechos del trabajo titulado *Clasificador difuso de señales acústicas ambientales basado en análisis de componentes independientes*, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección *lpachecob09@sagitario.cic.ipn.mx*. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

María Guadalupe López Pacheco

Nombre y firma

RESUMEN

La evaluación del ruido se hace generalmente considerando el impacto de una fuente de ruido específica. Sin embargo, en prácticamente todos los entornos, un gran número de fuentes distintas contribuyen al ruido ambiental en una determinada zona. Las principales causas de la contaminación acústica son aquellas relacionadas con las actividades humanas como el transporte terrestre, el comercio, actividades recreativas, entre otras.

Debido a esto surge la necesidad de realizar la separación de fuentes que se presentan mezcladas en el ambiente, así como de conocer el tipo de fuente. Esta información es útil para tomar las medidas necesarias que ayuden a reducir los niveles de ruido.

Se presenta un método que permite la separación ciega de fuentes en campo, en un ambiente controlado. Se divide principalmente en dos etapas, la primera es una de pre-procesamiento compuesta por filtrado, wavelets y deconvolución; en la segunda etapa se utiliza Análisis de Componentes Independientes (*ICA, Independent Component Analysis*). Los experimentos se realizaron utilizando cinco tipos de fuentes, comunes en zonas urbanas: motores de vehículos, claxon, silbato, sirenas y voz. Se obtuvo la separación de 2, 3 y 4 fuentes mezcladas en ambientes controlados, grabadas mediante micrófonos de tipo industrial con comportamiento direccional.

Además, se diseñó un clasificador basado en técnicas de agrupamiento difuso, capaz de identificar los cinco tipos de fuentes considerados, con un buen desempeño. La extracción de rasgos característicos de las fuentes es una parte fundamental en el comportamiento el clasificador, se muestran los resultados obtenidos al utilizar distintos rasgos.

Se presentan resultados satisfactorios en la separación de mezclas en campo y en su clasificación, los mismos son una aproximación para ayudar a resolver el problema de contaminación acústica en ambientes reales.

ABSTRACT

The noise evaluation is usually done by considering the impact of a specific noise source. However, in almost all environments, a large number of different sources contribute to noise in a particular area. The main causes of noise pollution are those related to human activities such as transport, trade, leisure activities, among others.

Because of this, there is the need to obtain the source separation of sources that appear mixed in the atmosphere, as well as knowing the type of source. This information is useful to take the necessary actions to help reduce noise levels.

This work shows a method for blind source separation of sources in field, in a controlled environment. It is mainly divided into two stages, the first is a pre-processing consists of filtering, wavelets and deconvolution; the second stage uses Independent Component Analysis (ICA). The experiments were performed using five types of sources common in urban areas like: motor vehicles, horns, whistles, sirens and voice. Separation was obtained with 2, 3 and 4 sources mixed in controlled environments, recorded by industrial-type microphone with directional behavior.

Besides, a classifier was designed based on fuzzy clustering techniques, able to identify five types of sources considered, with a good performance. The extraction characteristics of the sources is a fundamental part of the classifier performance, this work shows the results obtained using different features.

The results presented are satisfactory in the separation of mixtures in field and in their classification. They are an approach to help solve the problem of noise in real environments.

ÍNDICE

Índice de Figuras

Índice de Tablas

CAPÍTULO 1. INTRODUCCIÓN

1.1 Antecedentes

El concepto de ecología ha tomado gran importancia durante las últimas tres décadas, esta rama de la biología estudia las relaciones entre el individuo y su medio. Uno de los factores que la ecología urbana desea reducir es la contaminación acústica que se define como el exceso de sonido que altera las condiciones normales del ambiente en una determinada zona, ya que esta puede causar grandes daños tanto la salud auditiva, física y mental de las personas como en la calidad de vida de la población, si no se controla adecuadamente [1,2].

El ruido se puede clasificar en: ruido continuo, el cual se produce por maquinaria que opera sin interrupción como ventiladores y bombas; ruido intermitente, producido por maquinaria que opera en ciclos, o cuando pasan vehículos aislados o aviones donde el nivel de ruido aumenta y disminuye rápidamente; ruido impulsivo, ruido de impactos o explosiones, es breve y abrupto, su efecto sorprendente causa gran molestia [3].

Para tomar las medidas necesarias que ayuden a reducir la contaminación acústica, se requiere identificar la fuente generadora del ruido. En este sentido se han hecho investigaciones sobre separación ciega de fuentes, la cual consiste en determinar las fuentes originales que se encuentran mezcladas en señales recuperadas a través de un arreglo de transductores, sin información alguna de las señales originales ni de las ponderaciones de la mezcla. La técnica más ampliamente usada en separación ciega de fuentes es Análisis de Componentes Independientes (*ICA,* por sus siglas en inglés).

En el Centro de Investigación en Computación del Instituto Politécnico Nacional se realizó un proyecto de investigación aprobado y apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACYT, México) denominado “Sistema Avanzado de Monitoreo Ambiental de Sonidos y Vibraciones”, clave: 51283-Y, [4].

En este mismo centro, se presentaron las siguientes tesis de la Maestría en Ciencias en Ingeniería de Cómputo, con los siguientes títulos: “Sistema distribuido para análisis de ruidos ambientales” y “Sistema de monitoreo de ruidos ambientales producidos por aviones en el AICM (Aeropuerto Internacional de la Ciudad de México)”, con la finalidad de analizar el ruido generado en zonas urbanas.

En la Universidad de Vigo, España, se realizaron los siguientes trabajos: “Análisis de componentes independientes en separación de fuentes de ruido de tráfico en vías interurbanas” y “Sustracción espectral de ruido en separación ciega de fuentes de ruido de tráfico” , con resultados satisfactorios obtenidos en separación de fuentes agrupadas en clases aplicando la técnica de ICA, dando lugar a una separación ciega con suficiente grado de independencia entre fuentes, caracterizada por una marcada diferencia en su distribución espectral de energía [5,6].

Los estudios realizados sobre separación de fuentes relacionadas con el ruido ambiental son escasos, sin embargo es de suma importancia en nuestro país, ya que México cuenta con un sistema de monitoreo permanente en la centro histórico del D.F., donde se desea reducir los niveles de contaminación acústica. Por lo que se hace necesario, separar las distintas fuentes generadoras de ruido y analizarlas por separado.

1.2 Planteamiento del problema

La evaluación del ruido se hace generalmente considerando el impacto de una fuente de ruido específica. Sin embargo, un gran número de fuentes distintas contribuyen al ruido ambiental en una determinada zona. Por lo que surgen principalmente los siguientes problemas:

1. Como las fuentes que producen la contaminación acústica se presentan mezcladas en el ambiente, no es posible cuantificar el aporte individual de cada una de ellas y por tanto facilitar el control y la reducción de su impacto ambiental.

2. No existe un modelo computacional capaz de identificar las fuentes generadoras de ruido típicas en zonas urbanas.

1.3 Justificación

Un informe de la Organización Mundial de la Salud (OMS), considera 50 dB como el límite superior deseable. Por encima de este nivel, el sonido resulta molesto para el descanso y la comunicación. En la Cd. De México se han reportado niveles de hasta 80 dB por el día y 52 dB durante la noche.

Las variaciones de presión sonora audible varían en un margen que puede ir desde los 0 dB hasta 130 dB, 0 dB corresponde al umbral auditivo medio de una persona, y una presión sonora de aproximadamente 130 dB es tan alta que causa dolor y es llamado umbral del dolor [3].

Inicialmente, los daños producidos por una exposición prolongada no son permanentes, sobre los 10 días desaparecen. Sin embargo, si la exposición a la fuente de ruido no cesa, las lesiones son definitivas. La sordera irá creciendo hasta que se pierda totalmente la audición. El déficit auditivo provocado por el ruido ambiental se llama *socioacusia* y se presenta como un silbido en el oído.

No sólo el ruido prolongado es perjudicial, un sonido repentino de 160 dB, como el de una explosión o un disparo, pueden llegar a perforar el tímpano o causar otras lesiones irreversibles. La contaminación acústica, además de afectar al oído puede provocar efectos psicológicos negativos y otros efectos fisiopatológicos. Dentro de los efectos fisiopatológicos se tiene:

a) A más de 60 dB: Dilatación de las pupilas y parpadeo acelerado; agitación respiratoria, aceleración del pulso y taquicardias; aumento de la presión arterial y dolor de cabeza; menor irrigación sanguínea y mayor actividad muscular, los músculos se ponen tensos y dolorosos, sobre todo los del cuello y espalda.
b) A más de 85 dB: Disminución de la secreción gástrica, gastritis o colitis; aumento del colesterol y de los triglicéridos, con el consiguiente riesgo cardiovascular, en enfermos con problemas cardiovasculares, arteriosclerosis o problemas coronarios, los ruidos fuertes y súbitos pueden llegar a causar hasta un infarto; aumenta la glucosa en sangre, en los enfermos de diabetes, la elevación de la glucemia de manera continuada puede ocasionar complicaciones médicas a largo plazo.

Efectos psicológicos: Insomnio y dificultad para conciliar el sueño, fatiga, estrés (por el aumento de las hormonas relacionadas con el estrés como la adrenalina), depresión y ansiedad, irritabilidad y agresividad, histeria y neurosis, aislamiento social, entre otras [7,8,9].

Conocer el nivel de ruido específico que cada fuente aporta, es sumamente importante para: planear nuevos desarrollos de zonas residenciales, zonas industriales, autopistas, aeropuertos; solucionar las quejas de los ciudadanos durante el proceso de planificación o después, evaluar la conformidad o no conformidad de las fuentes de ruido (plantas industriales, parques de atracciones, aeropuertos, autopistas, ferrocarriles, entre otros) según la normativa y la legislación.

Solo las grandes ciudades de países desarrollados o con economías emergentes presentan una estimación del impacto del ruido ambiental. Por ejemplo en *El Libro Verde de la Unión Europea* sobre la *Futura Política de Ruido* (1996); se estima que, en términos del número de personas afectadas por el ruido, el 20% de la población (unos 80 millones de personas) sufre niveles de ruido inaceptables que causan alteraciones en el sueño, molestias y efectos adversos sobre la salud. Otros 170 millones de ciudadanos viven en Europa, en áreas donde los niveles de ruido causan una seria molestia durante el día [1].

1.4 Hipótesis

Es posible lograr una separación ciega de señales acústicas que afectan principalmente zonas urbanas, capturadas en campo en un ambiente controlado y tomando en cuenta las fuentes producidas por actividades comerciales, de transporte terrestre y recreativas, e identificarlas mediante un clasificador difuso basado en técnicas de agrupamiento.

1.5 Objetivos

1.5.1 Objetivo general

Desarrollar un método para obtener la separación ciega de señales acústicas ambientales y diseñar un clasificador basado en técnicas de agrupamiento difuso tipo 2.

1.5.2 Objetivos particulares

1. Obtener la separación ciega de fuentes de ruido ambiental en un ambiente controlado, utilizando análisis de componentes independientes.
2. Obtener rasgos característicos de las fuentes estimadas en la separación, útiles para el diseño del clasificador difuso.
3. Diseñar y modelar el clasificador difuso utilizando técnicas de agrupamiento difuso.
4. Realizar pruebas para validar el desempeño del método propuesto.

1.6 Alcances del trabajo

La separación de las señales acústicas ambientales, se realizó utilizando fuentes producidas por actividades sociales, comerciales y de transporte. Se realizaron pruebas de laboratorio y en campo, que consisten en generar señales de ruido ambiental en un medio controlado para ser capturadas por sensores de tipo industrial. Utilizando esta mezclas se aplicó la etapa de pre-procesamiento y posteriormente el método de *ICA*, para separarlas y estimar las fuentes originales. En los experimentos se consideraron 2, 3 y 4 fuentes fijas, utilizando arreglos predefinidos de micrófonos industriales.

Se caracterizaron los tipos de fuentes consideradas obteniendo indicadores o rasgos característicos utilizando análisis espectral, análisis de octavas, componentes impulsivas, componentes tonales, entre otros.

Con los rasgos característicos de cada fuente se generaron grupos (*clusters*) útiles para el entrenamiento del sistema difuso tipo-2 mediante técnicas de agrupamiento difuso como *c-medias difuso*, *c-medias posibilístico* y *Gustafson-Kessel*, principalmente.

1.7 Contribuciones

- Un método que permite separar y clasificar las distintas fuentes principales de ruido ambiental que contribuyen a la contaminación acústica en un ambiente controlado. para su posterior análisis individual. Con el objetivo de realizar posteriormente el análisis acústico de las fuentes individuales y ayudar a tomar las medidas necesarias que reduzcan los niveles de ruido ambiental.

- Se desarrolló un clasificador difuso basado en técnicas de agrupamiento difuso tipo-2 para identificar señales acústicas ambientales, utilizando rasgos característicos de cada señal fuente definida.

1.8 Método de investigación y desarrollo utilizado

El método utilizado para realizar la investigación de este trabajo se presenta a continuación:

- Establecer los problemas a resolver.
- Definir el objetivo principal y los objetivos particulares.
- Investigar el estado del arte sobre separación ciega de fuentes y clasificación.
- Realizar pruebas para separación ciega de fuentes de mezclas lineales, para caracterizar el algoritmo de ICA.
- Capturar mezclas de sonidos en campo en un ambiente controlado, donde se consideraron 2, 3, y 4 fuentes fijas.
- Realizar experimentos de laboratorio, para capturar mezclas en un lugar cerrado considerando las fuentes fijas y un arreglo de micrófonos lineal.
- Recopilar señales acústicas ambientales para conformar una base de datos, con sonidos disponibles en internet, considerando sonidos de voz, sirenas, claxon, silbato y autos.
- Analizar y evaluar distintos rasgos característicos de los cinco tipos de fuentes considerados, que resultan útiles para diseñar un clasificador difuso.
- Diseñar un clasificador difuso tipo-2 capaz de identificar entre cinco fuentes que predominan en la ruido presente en zonas urbanas.
- Reportar resultados del método propuesta para separación y clasificación.

1.9 Organización del trabajo

Capítulo 2. Se presenta una breve descripción de los diferentes métodos de análisis de componentes independientes, las técnicas más utilizadas para diseñar clasificadores difusos tipo-2 y sus aplicaciones.

Capítulo 3. Base teórica de la metodología utilizada para la separación ciega de fuentes para mezclas lineales y convolutivas. Descripción de las principales técnicas de agrupamiento difuso para clasificación.

Capítulo 4. Descripción del método utilizado para separación ciega de fuentes de mezclas lineales y convolutivas, extracción de rasgos y clasificación.

Capítulo 5. Experimentos y resultados utilizando el método para separación de señales acústicas y el clasificador difuso diseñado.

Conclusiones. Se muestran las conclusiones sobre los resultados obtenidos, así como las prestaciones y limitaciones del sistema propuesto.

Anexos A. Conceptos básicos de procesos aleatorios.

Anexos B. Características del equipo de medición.

Referencias Web de señales acústicas. Se citan las referencias donde se encuentran disponibles los sonidos de cada señal fuente utilizados para entrenar el clasificador.

CAPÍTULO 2. ESTADO DEL ARTE

2.1 Ruido ambiental

El ruido ambiental es un problema mundial, últimas noticias relacionadas con los problemas de ruido ambiental abundan, gran esfuerzo y grandes sumas de dinero son invertidas en los conflictos producidos por el ruido ambiental. Sin embargo, la forma en que el problema es tratado difiere enormemente de un país a otro y depende mucho en la cultura, la economía y la política. Pero el problema persiste incluso en las zonas donde gran cantidad de recursos se han utilizado para la regulación, evaluación y amortiguación de fuentes de ruido o para la creación de barreras acústicas. Por ejemplo, grandes esfuerzos se han realizado para reducir el ruido del tráfico desde la fuente, de hecho, los coches de hoy en día son más silenciosos que los fabricados hace diez años, pero el volumen de tráfico ha aumentado por lo que el nivel de molestia va en aumento. La fabricación de vehículos más silenciosos podría haber aliviado el problema durante un período pero ciertamente no lo ha quitado. No hay estimaciones a nivel mundial del impacto y el coste del ruido ambiental, sin embargo, un ejemplo destacado que cubre la mayor parte de Europa es el Libro Verde de la Unión sobre la política de ruido Futuro (1996).

El sonido puede ser definido como cualquier variación de la presión que el oído humano puede detectar, un movimiento de onda se hace cuando un elemento es colocado lo más cercano a una partícula de aire en movimiento. Este movimiento se extiende progresivamente a las partículas de aire adyacentes más allá de la fuente, como las fichas de dominó. Dependiendo del medio, el sonido se propaga en diferentes velocidades. En el aire, el sonido se propaga a una velocidad de aproximadamente 340 m/s. En líquidos y sólidos, la velocidad de propagación es mayor - 1500 m/s en el agua y 5000 m/s en el acero. En la Tabla 2.1 se listan los procedimientos más comunes para el control de ruidos que son utilizados para el diseño arquitectónico y para controlar la contaminación acústica en zonas urbanas.

Tabla 2.1 Procedimientos más comunes para control de ruidos

Objetivos de control de ruidos	**PROCEDIMIENTOS DE CONTROL**					
	Silenciar la fuente	Barreras	Aislamiento de la vibración	Absorción	Enmasca-ramiento	Protección personal
Reducir el nivel general del ruido para:						
Mejorar la comunicación	**X**		**X**	**X**		
Incrementar la comodidad	**X**		**X**	**X**		
Reducir el riesgo de dañar el oído	**X**		**X**	**X**		
Reducir los ruidos extraños para:						
Mejorar la intimidad		**X**			**X**	
Incrementar la comodidad		**X**	**X**			
Mejorar la comunicación		**X**	**X**			

Tabla 2.1 ***Continúa…***

Objetivos de control de ruidos	PROCEDIMIENTOS DE CONTROL					
	Silenciar la fuente	Barreras	Aislamiento de la vibración	Absorción	Enmasca-ramiento	Protección personal
Proteger a varias personas contra fuentes localizadas de ruidos molestos	X	X	X	X		X
Proteger a varias personas contra muchas fuentes distribuidas de ruidos molestos	X	X	X	X		X
Proteger a una persona contra una fuente localizada de ruidos molestos	X	X	X			X
Proteger a varias personas contra muchas fuentes distribuidas de ruidos molestos		X				X
Eliminar ecos y murmullos				X		
Reducir la reverberación				X		
Eliminar vibraciones molestas			X			

En comparación con la presión estática del aire (10^5 Pa), las variaciones de presión de sonido audible son muy pequeños que van desde los 20 µPa ($20x10^{-6}$Pa) a 100 Pa. En los niveles típicos de ruido, 20 µPa corresponde al promedio umbral de audición de la persona. Es por ello que se llama umbral de audición. Una presión de sonido de aproximadamente 100 Pa es tan fuerte que causa dolor y es llamada umbral del dolor. La relación entre estos dos extremos es más de un millón a uno.

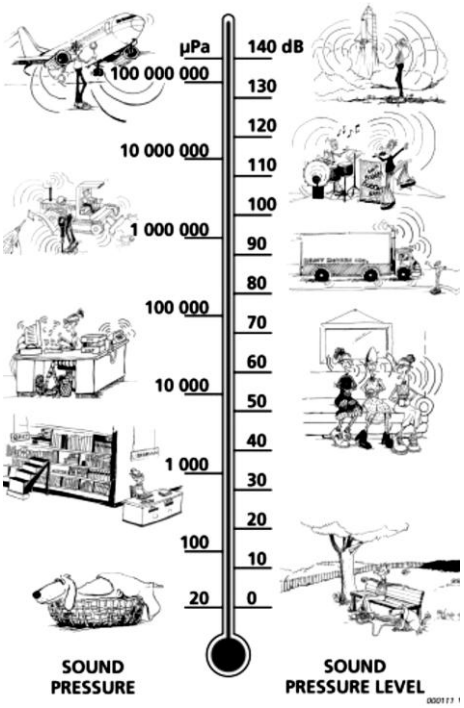


Fig. 2.1 Escala en µPa comparada con la escala en dB.

Una aplicación directa de las escalas lineales en Pa, para medición de la presión del sonido conduce a números muy grandes. Y, así como el oído reacciona logarítmicamente en vez de forma lineal a los estímulos, es más práctico expresar los parámetros acústicos como una relación logarítmica del valor medido a un valor de referencia. Esta relación logarítmica se llama decibelios o dB, la escala en dB hace los números manejables. La ventaja de usar dB puede verse claramente en la Fig. 2.1, aquí la escala lineal con sus largos números es convertida a una escala manejable de 0 dB en el umbral de audición (20 µPa) a 130 dB en el umbral del dolor (~ 100 Pa).

El número de variaciones de presión por segundo se denomina frecuencia del sonido y se mide en hertz (Hz). La audición normal para un adulto sano va del rango de aprox. 20 Hz a 20000 Hz. En términos de niveles de presión del sonido, el rango del sonido audible va desde el umbral de audición 0 dB hasta el umbral de dolor a 130 dB y más. Sin embargo un aumento de 3 dB representa duplicar la presión del sonido (una confusión común es el hecho de que al doblar el voltaje da como resultado un incremento de 6dB, mientras que si se dobla la potencia sólo resulta un incremento de 3dB). Así, el menor cambio perceptible es de aprox. 1 dB, Fig. 2.2.

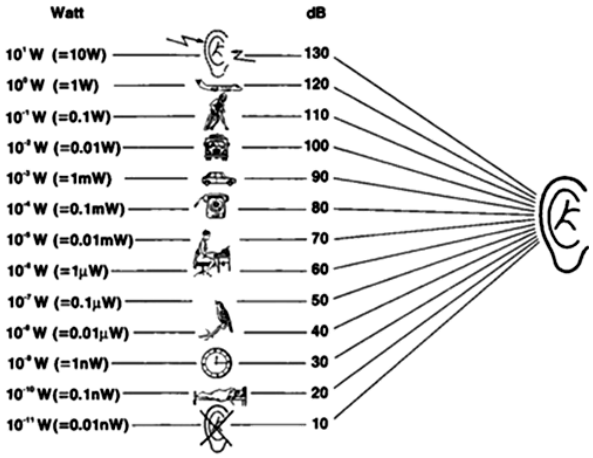


Fig. 2.2 Umbral acústico

Cuando se mide el ruido, es necesario saber de qué tipo de ruido se trata para elegir los parámetros de medición adecuados, el equipo a usar y la duración de la medición, a continuación se describen los tipos de ruidos más comunes sobre los que se realizan mediciones para su análisis.

Ruido Continuo. Es producido por la maquinaria que funciona sin interrupción de la misma manera, por ejemplo, los sopladores, bombas y equipos de proceso. Midiendo por tan sólo unos minutos con el equipo de mano es suficiente para determinar el nivel de ruido. Si los tonos o frecuencias bajas son escuchados, el espectro de frecuencia puede ser medido para su posterior análisis.

Ruido Intermitente. Cuando la maquinaria opera en ciclos, o cuando los vehículos individuales o aviones pasan cerca, el nivel de ruido aumenta y disminuye rápidamente. Para cada ciclo de una fuente de ruido, el nivel de ruido puede ser medido solo para ruido continuo. Sin embargo, la duración del ciclo debe ser anotado. Un solo vehículo o aeronave que pasa es llamado evento. Para medir el ruido de un evento, el nivel de sonido expuesto es medido, la combinación del nivel y la duración. El máximo nivel de presión de sonido también puede ser utilizado. Una serie de eventos similares se pueden ser medidos para establecer un promedio confiable.

Ruido impulsivo. El ruido de impactos o explosiones, por ejemplo, de un martillo, troqueladora o arma de fuego, es llamado ruido impulsivo. Es breve y abrupto, y su efecto escandaloso causa mayor molestia de la que podría esperarse en una simple medición de sonido. Para cuantificar la impulsividad del ruido, la diferencia entre una respuesta rápida y un parámetro de respuesta lenta puede ser usada [1].

Las variables de la señal acústica fijadas por las normas internacionales son: nivel de presión acústica medida en decibeles (dB), campo de frecuencia del sonido emitido en Hertz (Hz), la distancia entre la fuente y el receptor, y la presencia de otras fuentes de ruido, en la Tabla 2.2 se encuentra la equivalencia del nivel de presión del sonido en dB, con respecto a la distancia medida en metros.

Tabla 2.2 Equivalencia del nivel de presión del sonido con base en la distancia

m	dB (A)[1]																					
1	65	70	75	80	85	90	92	94	96	98	100	102	104	106	108	110	112	114	116	118	120	
2	59	64	69	74	79	84	86	88	90	92	94	96	98	100	102	104	106	108	110	112	114	
3	55	60	65	70	75	80	82	84	86	88	90	92	94	96	98	100	102	104	106	108	110	
5	51	56	61	66	71	76	78	80	82	84	86	88	90	92	94	96	98	100	102	104	106	
10	45	50	55	60	65	70	72	74	76	78	80	82	84	86	88	90	92	94	96	98	100	
20	39	44	49	54	59	64	66	68	70	72	74	76	78	80	82	84	86	88	90	92	94	
30	35	40	45	50	55	60	62	64	66	68	70	72	74	76	78	80	82	84	86	88	90	
50	–	36	41	46	51	56	58	60	62	64	66	68	70	72	74	76	78	80	82	84	86	
100		–	–	40	45	50	52	54	56	58	60	62	64	66	68	70	72	74	76	78	80	
200				–	39	44	46	48	50	52	54	56	58	60	62	64	66	68	70	72	74	
300					–	40	42	44	46	48	50	52	54	56	58	60	62	64	66	68	70	
500						–	38	40	42	44	46	48	50	52	54	56	58	60	62	64	66	
1000							–	–	–	38	40	42	44	46	48	50	52	54	56	58	60	
2000										–	–	–	38	40	42	44	46	48	50	52	54	
3000													–	–	38	40	42	44	46	48	50	
5000															–	–	38	40	42	44	46	

[1] ***Medido en la escala de un medidor estándar de intensidad sonora***

2.2 Separación ciega de fuentes

El problema de la separación es antiguo en la ingeniería eléctrica y se ha estudiado bien, existen muchos algoritmos para la separación de señales mixtas. El problema de la separación ciega es más difícil, ya que sin el conocimiento de las señales que se han mezclado, no es posible el diseño de pre-procesamiento apropiado para separarlas. La única hipótesis formulada por Herault y Jutten era la independencia, pero las limitaciones adicionales son necesarias en la distribución de probabilidad de las fuentes. Si asumimos que las señales fuente tienen distribución Gaussiana, entonces el problema no tiene solución general. Las investigaciones posteriores han demostrado que el mejor rendimiento obtenido por la red de Herault-Jutten se presenta cuando las señales fuente tienen distribución sub-Gaussiana (Cohen et al, 1992), es decir, para las señales cuya curtosis es menor que el de una distribución Gaussiana.

En el campo de redes neuronales, este modelo de red se vio ensombrecido en su momento por la red de Hopfield, que pronto perdió popularidad ante el algoritmo de propagación hacia atrás del perceptron multicapa. En 1994, el campo de redes neuronales se había movido de algoritmos de aprendizaje supervisado hacia el aprendizaje no-supervisado. En 1995, se obtuvieron los primeros resultados del método de ICA, conocido como Infomax por Bell y Sejnowski, que depende de invertir una matriz. Amari en 1997 y Cardoso en 1996, se dieron cuenta de forma independiente, que el algoritmo de Infomax se puede mejorar mediante el uso de la pendiente natural, acelerando la convergencia mediante la eliminación de la inversión de la matriz. Este descubrimiento permitió aplicar este algoritmo en una variedad de problemas del mundo real.

Análisis de componentes independientes (***ICA**, por sus siglas en inglés*), es una nueva técnica para separar datos o señales que fue desarrollada a partir de 1986, cuando Jeanny Herault y Chirstian Jutten presentaron un modelo de red neuronal recurrente y un algoritmo de aprendizaje basado en la regla de Hebb que es capaz de separar ciegamente mezclas de señales independientes.

Varios enfoques diferentes se han adoptado para la separación ciega de fuentes, como son: la máxima verosimilitud, los métodos de Bussgang basados en cumulantes, la búsqueda de proyección y los métodos basados en la negentropía. Todos estos están estrechamente relacionados con el marco de Infomax (Lee, Girolami, Bell y Sejnowski, 1998). Por lo tanto, un gran número de investigadores que han atacado a ICA a partir de una variedad de direcciones diferentes, convergen en un conjunto común de principios. Herault y Jutten mencionan en su documento de 1986, "*No podemos probar la convergencia de este algoritmo a causa de la no linealidad de la ley la adaptación y la no estacionariedad de las señales*". Todavía no se cuenta con una explicación adecuada de por qué ICA converge para tantos problemas, casi siempre con las

mismas soluciones, incluso cuando las señales originales no son obtenidas de fuentes independientes.

A pesar de la separación ciega de mezclas de señales generadas por simulación es un punto de referencia útil, un problema más difícil es aplicar ICA a las grabaciones de las señales del mundo real donde las fuentes se desconocen. Un ejemplo importante es la aplicación de Infomax sobre señales electroencefalográficas (EEG) de los potenciales de cuero cabelludo de los humanos. Las señales eléctricas procedentes del cerebro son muy débiles en el cuero cabelludo, en el rango de microvolts, además hay más componentes derivadas de los movimientos del ojo y los músculos. Ha sido un reto difícil eliminar estas componentes sin alterar las señales del cerebro. ICA es ideal para esta tarea, ya que las señales obtenidas por los sensores corresponden a diferentes mezclas lineales de las señales del cerebro y los movimientos. El algoritmo Infomax extendido ha demostrado ser el mejor método para separar estas señales, es decir, separar fuentes sub-Gaussianas como ruido, de las señales del cerebro, que generalmente son super-gaussianas (Jung et al. 1998).

ICA se puede aplicar a muchos problemas en que las mezclas no son ortogonales y las señales de origen no son Gaussianas. La mayoría de señales portadoras de información tienen estas características. Hay muchos problemas de interés teórico en ICA que aún no se han resuelto y hay muchas aplicaciones nuevas, como en minería de datos, que aún no se han explorado [10].

Existen diversas variables que deben ser tomadas en cuenta para determinar el desempeño de ICA. El primer paso es decidir si es necesario un pre-procesamiento de los datos. Algunos pasos de pre-procesamiento pueden ser el blanqueado, centrado y reducción de dimensión utilizando PCA (*Principal Component Analysis*). El segundo paso es decidir la forma de estimación de la matriz de mezcla $\boldsymbol{W}$.

Para estimar la matriz de mezcla, se pueden considerar los algoritmos que incluyen la maximización de la *no-gaussianidad*, la estimación de máxima verosimilitud, la reducción al mínimo de información mutua, los métodos tensoriales, *decorrelación* no lineal, entre otros. Una vez que se ha tomado esta decisión, los componentes independientes se pueden encontrar utilizando (2.1).

$$\boldsymbol{s}(t) = \boldsymbol{W}\boldsymbol{x}(t) \qquad (2.1)$$

ICA tiene muchas aplicaciones en diferentes áreas como separación ciega de fuentes, reconocimiento de patrones, mejoras en las comunicaciones, entre otras. Una aplicación importante de ICA es limpiar las señales, eliminando el ruido y otros componentes no deseados; ya que el ruido puede ser separado de las señales de audio, diversos dispositivos se verían beneficiados con esta tecnología como: audífonos, altavoces, teléfonos móviles y teleconferencias [11].

Otras aplicaciones en las que se aplica el método de ICA se listan a continuación:

- Procesamiento de señales médicas, por ejemplo: electroencefalograma (EEG) y electrocardiograma (ECG), donde se pueden separar las señales del cerebro o del corazón, respectivamente.
- Separación de señales de voz [12].
- Problemas de comunicación incluyendo la propagación de varias rutas en los sistemas móviles [13,14].
- Procesamiento de imágenes – Extracción de características [15].
- En física, se puede utilizar para mejorar el análisis de los espectros de las reacciones nucleares.

2.2.1 Mezclas convolutivas

En los modelos simples, la mezcla consiste en una suma de señales fuente con diferente ponderación. Sin embargo, en muchas aplicaciones del mundo real, como por ejemplo en acústica, el proceso de mezclado es más complejo. En este caso las mezclas son ponderadas y retardadas, y cada fuente contribuye a la suma de múltiples retardos correspondientes a los trayectos múltiples por la cual una señal acústica se propaga hacia un micrófono. Esta suma de diferentes fuentes filtradas es llamada mezcla convolutiva. En muchas situaciones se desea recuperar todas las fuentes de mezclas grabadas o al menos separar algunas de ellas. Sin embargo, puede ser útil identificar el proceso de mezclado en sí mismo, ya que puede revelar información sobre la naturaleza del sistema de mezclado físico.

Existen múltiples aplicaciones de separación ciega en mezclas convolutivas. En acústica diferentes fuentes de sonido se graban simultáneamente con varios micrófonos. Estas fuentes pueden ser voz o música, también las señales registradas bajo el agua con un sonar pasivo [16]. En las comunicaciones de radio, matrices de antenas reciben mezclas de comunicación compuestas por diferentes señales [17,18]. La separación de fuentes también se ha aplicado a datos astronómicos o imágenes de satélite [19]. Por último, los modelos convolutivos se han utilizado para interpretar datos funcionales de imágenes del cerebro y los voltajes de bio-señales [20,21,22,23].

Nos enfocaremos en el problema de separación de mezclas convolutivas acústicas, en primer lugar se introduce el modelo básico de mezclas convolutivas. En el tiempo t, una mezcla de N señales fuentes $\mathbf{s}(\mathrm{t}) = (\mathrm{s}_1(\mathrm{t}), \dots, \mathrm{s}_\mathrm{n}(\mathrm{t}))$ se reciben en una matriz de sensores. Las señales recibidas se denotan por $\mathbf{x}(\mathrm{t}) = (\mathrm{x}_1(\mathrm{t}), \dots, \mathrm{x}_\mathrm{m}(\mathrm{t}))$. El modelo convolutivo presenta la siguiente relación entre la

m señal mixta, las señales de las fuentes originales y algunos sensores de ruido aditivo $\mathbf{v}_{\mathrm{m}}(\mathrm{t})$, en forma matricial.

$$\boldsymbol{x}(t) = \sum_{k=0}^{K-1} \boldsymbol{A}_k \boldsymbol{s}(t-k) + \boldsymbol{v}(t) \qquad (2.2)$$

La señal recibida $\mathbf{x}(\mathrm{t})$ es una combinación lineal de las versiones filtradas de cada una de las señales fuente, la matriz $\mathbf{A}_{\mathrm{k}}$ de MxN representa los coeficientes del filtro de mezcla. En la práctica estos coeficientes pueden cambiar en el tiempo, pero para simplificar el modelo se supone que son estacionarios. En teoría, los filtros pueden ser de longitud infinita (como los sistemas IIR), sin embargo, en este caso es suficiente asumir que $\mathrm{k} < \infty$. En el dominio Z, el modelo convolutivo, se escribe en la ecu. (2.3), donde $\mathbf{A}(\mathrm{z})$ es una matriz de filtros polinomiales tipo FIR (*Finite Impulse Response*).

$$\boldsymbol{X}(z) = \boldsymbol{A}(z)\boldsymbol{S}(z) + \boldsymbol{V}(z) \qquad (2.3)$$

Existen algunos casos especiales de mezclas convolutivas que simplifican la ecuación (2.2):

Modelo de mezclado instantáneo: Se asume que todas las señales de llegada a los sensores ocurren al mismo tiempo sin ser filtradas, el modelo convolutivo se simplifica (2.4), donde $\mathbf{A} = \mathbf{A}_0$ es una matriz de MxN que contiene los coeficientes de mezclado. Se han desarrollado muchos algoritmos para resolver este modelo, como se describió anteriormente.

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{v}(t) \qquad (2.4)$$

Fuentes con retardos: Asumiendo un espacio libre de reverberación con retardos de propagación el modelo de mezclado se simplifica de la siguiente manera

$$\boldsymbol{x}_m(t) = \sum_{n=1}^{N} \boldsymbol{a}_{mn} \boldsymbol{s}_n(t - k_{mn}) + \boldsymbol{v}_m(t) \qquad (2.5)$$

Donde k_{mn} es el retardo de propagación entre la fuente n y el sensor m.

Modelo libre de ruido: En la mayoría de los algoritmos se supone que el modelo convolutivo es libre de ruido, es decir:

$$\boldsymbol{x}(t) = \sum_{k=0}^{K-1} \boldsymbol{A}_k \boldsymbol{s}(t-k) \qquad (2.6)$$

Fuentes sobre y sub-determinadas: Usualmente se asume que el número de sensores es igual o mayor al número de fuentes, en el cual los métodos lineales son suficientes para invertir el mezclado lineal. Sin embargo, si el número de fuentes es mayor al número de sensores el problema es sub-determinado y aun conociendo el sistema de mezclado lineal no hay métodos desarrollados para recuperar las fuentes.

La Fig. 2.3 muestra la taxonomía de los métodos más importantes en la separación ciega de fuentes cuando las mezclas son convolutivas.

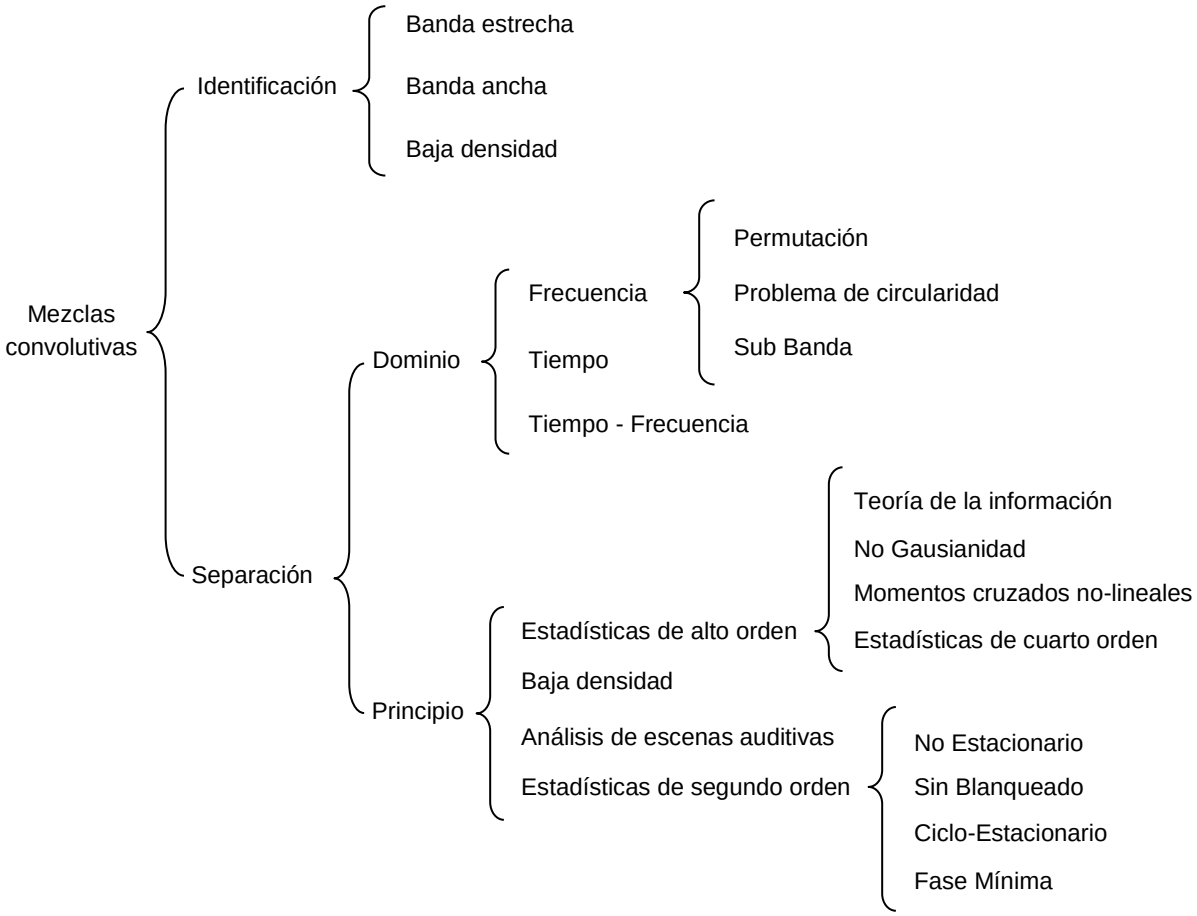


Fig. 2.3. Taxonomía de los métodos más relevantes para separación ciega de fuentes convolutiva.

Los algoritmos de separación ciega de fuentes se basan en consideraciones previas sobre las fuentes y el modelo de mezclado del sistema. En general, se considera que las fuentes son independientes o por lo menos no correlacionadas. Los criterios de separación se pueden dividir en métodos basados en estadística de alto orden (HOS, *High Order Statistics*), y métodos basados en estadística de segundo orden (SOS, *Second Order Statistics*).

En la Tabla 2.3 se muestra un resumen del modelo de las señales capturadas y el modelo de separación para cada dominio utilizado frecuentemente en mezclas convolutivas.

Tabla 2.3. Modelo de mezcla convolutiva y ecuación de separación correspondiente en los diferentes ámbitos descritos para la separación ciega de fuentes.

	PROCESO DE MEZCLADO	MODELO DE SEPARACION
Tiempo	$x_m(t) = \sum_{n=1}^{N} \sum_{k=0}^{K-1} a_{mn} s_n(t-k) + v_m(t)$ $\boldsymbol{x}(t) = \sum_{k=0}^{K-1} \boldsymbol{A_k s}(t-k) + \boldsymbol{v}(t)$	$y_n(t) = \sum_{m=1}^{M} \sum_{l=0}^{L-1} w_{nml} x_m(t-l)$ $\boldsymbol{y}(t) = \sum_{l=0}^{L-1} \boldsymbol{W_l x}(t-l)$
Dominio – Z	$\boldsymbol{X}(z) = \boldsymbol{A}(z)\boldsymbol{S}(z) + \boldsymbol{V}(z)$	$\boldsymbol{Y}(z) = \boldsymbol{W}(z)\boldsymbol{X}(z)$
Dominio de la Frecuencia	$\boldsymbol{X}(w) = \boldsymbol{A}(w)\boldsymbol{S}(w) + \boldsymbol{V}(w)$	$\boldsymbol{Y}(w) = \boldsymbol{W}(w)\boldsymbol{X}(w)$
Forma de Bloques Toeplitz	$\hat{\boldsymbol{x}}(t) = \hat{\boldsymbol{A}}\hat{\boldsymbol{s}}(t)$	$\hat{\boldsymbol{y}}(t) = \widehat{\boldsymbol{W}}\hat{\boldsymbol{x}}(t)$

En la separación convolutiva también asume que los sensores reciben N versiones de las fuentes linealmente independientes, esto significa que las fuentes deben proceder de diferentes lugares en el espacio (o al menos que las señales se emiten en diferentes orientaciones) y que hay al menos tantas fuentes como sensores, es decir, $M \geq N$. Algunos algoritmos hacen suposiciones más estrictas, como que los datos no se solapen en el dominio tiempo-frecuencia, o por ejemplo en las comunicaciones por radio, una suposición razonable de las fuentes es la ciclo-estacionariedad o el hecho de que señales de la fuente sólo toman valores discretos. Mediante el uso de tales consideraciones sobre las fuentes estadísticas a veces es posible cambiar la condición del número de sensores, por ejemplo, que $M < N$. Los diferentes métodos utilizados según los criterios sobre el número de fuentes y sensores para la separación de fuentes se resumen en la Tabla 2.4.

En Nishikawa et al. (2003) [24], se muestra un enfoque comparando las ventajas y desventajas de la separación en el dominio del tiempo y en el dominio de la frecuencia. Esto se resume en la Tabla 2.5. La gran mayoría de los algoritmos para separación convolutiva han sido evaluados en

simulación. Desafortunadamente, en el proceso real, se han utilizado una gran variedad de salas para evaluarlos y no están claros los resultados.

Tabla 2.4. Métodos aplicados considerando el número de fuentes y sensores para la separación

M > N	M = N	M < N
• Métodos de subespacios • Reducción del problema a mezcla instantánea	• Fuentes asimétricas por cumulantes de 2° y 3° orden • Criterio de separación basado en SOS y HOS para sistemas de 2 x 2 • Fuentes no-correlacionadas con espectros de potencia distinta • Cumulantes de orden superior, principio ML • Filtros cruzados conocidos • Funciones impares no-lineales	• Cumulantes cruzados • Baja densidad de datos en tiempo y frecuencia

Tabla 2.5. Ventajas y desventajas de separación de mezclas convolutivas en el dominio del tiempo y en el dominio de la frecuencia.

Dominio del tiempo		Dominio de la frecuencia	
Ventajas	*Desventajas*	*Ventajas*	*Desventajas*
La consideración de independencia se mantiene para señales de banda completa. Alta convergencia cerca del punto óptimo.	Degradación de convergencia en ambientes con gran reverberación. Se requiere ajustar muchos parámetros en cada iteración.	La mezcla convolutiva puede ser transformada en un problema de mezcla instantánea para cada segmento de frecuencia. Debido a la FFT los cálculos, se guardan en comparación a una aplicación en el dominio del tiempo.	Para cada banda de frecuencia, existe ambigüedad en la permutación y el escalamiento, que necesita ser resuelta. Problemas con pocas muestras en cada banda de frecuencia puede causar que la consideración de independencia falle. La convolución circular deteriora el desempeño de separación. La inversión de la matriz W no se puede garantizar.

Las principales preocupaciones son la sensibilidad del micrófono al ruido (a menudo no es mayor que -25dB), la no linealidad de los sensores y posiblemente la fuerte reverberación. Se hace notar que solo un pequeño subconjunto de las investigaciones reportan la evaluación de los algoritmos con grabaciones reales.

La Tabla 2.6 muestra una lista de resultados obtenidos utilizando la relación señal a interferencia (SIR, *Signal Interference Ratio*), que se define como un promedio de múltiples canales de salida. Los resultados SIR no son directamente comparables, ya que dependen del instrumento de medición, la sala de grabación y el *SIR* en las mezclas grabadas. Los resultados reportados tienen un SIR de 10 a 20 dB para dos fuentes estacionarias, por lo general de 2 a 10 segundos es suficiente para generar estos resultados. Sin embargo, nada se puede decir de este estudio sobre fuentes móviles [25]. Los artículos reportan que la separación para más de dos fuentes sigue siendo un reto. Todavía hay mucho trabajo que queda por hacer.

Tabla 2.6. Una visión general de los algoritmos aplicados en habitaciones, donde mejoras en el SIR han sido reportadas.

Tamaño aprox. de la sala [m]	T_{60} [ms]	N Fuentes	M Sensores	SIR [dB]
6 × 3 × 3	300	2	2	13
6 × 3 × 3	300	2	2	8–10
6 × 3 × 3	300	2	2	12
6 × 3 × 3	300	2	2	5.7
6 × 3 × 3	300	2	2	18–20
	50	2	2	10
	250	2	2	16
6 × 6 × 3	200	2	2	< 16
6 × 6 × 3	150	2	2	< 15
6 × 6 × 3	150	2	2	< 20
	500	2	2	6
4 × 4 × 3	130	3	2	4–12
4 × 4 × 3	130	3	2	14.3
4 × 4 × 3	130	3	2	< 12
4 × 4 × 3	130	2	2	7–15
4 × 4 × 3	130	2	2	4–15
4 × 4 × 3	130	2	2	12
4 × 4 × 3	130	6	8	18
4 × 4 × 3	130	4	4	12
	130	3	2	10
Oficina		2	2	5.5–7.6

Tabla 2.6. Continúa…

Tamaño aprox. de la sala [m]	T_{60} [ms]	N Fuentes	M Sensores	SIR [dB]
6 × 3 × 3	300	2	2	13
6 × 3 × 3	300	2	2	8–10
6 × 3 × 3	300	2	2	12
6 × 3 × 3	300	2	2	5.7
6 × 3 × 3	300	2	2	18–20
	50	2	2	10
	250	2	2	16
6 × 6 × 3	200	2	2	< 16
6 × 6 × 3	150	2	2	< 15
6 × 6 × 3	150	2	2	< 20
	500	2	2	6
4 × 4 × 3	130	3	2	4–12
4 × 4 × 3	130	3	2	14.3
4 × 4 × 3	130	3	2	< 12
4 × 4 × 3	130	2	2	7–15
4 × 4 × 3	130	2	2	4–15
4 × 4 × 3	130	2	2	12
4 × 4 × 3	130	6	8	18
4 × 4 × 3	130	4	4	12
	130	3	2	10
Oficina		2	2	5.5–7.6
6 × 5	130	2	8	1.6–7.0
8 × 7	300	2	2	4.2–6.0
15 × 10	300	2	2	5–8.0
		2	2	< 10
Oficina		2	2	6
Varias salas		2	2	3.1–27.4
Sala pequeña		2	2	4.7–9.5
4 × 3 × 2		2	2	< 10
4 × 3 × 2		2	2	14.4
4 × 3 × 2		2	2	4.6
		2	2	< 15
6 × 7	580	2	3	< 73
	810	2	2	< 10
Sala de conf.		4	4	14

2.3 Clasificador difuso

El agrupamiento de datos ha sido ampliamente aceptado en una gama de actividades de exploración de datos, análisis y desarrollo de modelos en ciencia, ingeniería, economía, así como las disciplinas biológicas y médicas. Áreas como la minería de datos, análisis de imágenes, reconocimiento de patrones, modelado y bioinformática son sólo ejemplos tangibles de numerosas actividades que explotan los conceptos y algoritmos de agrupamiento de datos tratados como herramientas esenciales para la formulación de problemas y el desarrollo de soluciones específicas o un vehículo para facilitar los mecanismos de interpretación.

El papel de la agrupación difusa se vuelve muy importante dentro del marco general de la agrupación y clasificación de datos. La agrupación ayuda a obtener una visión de la estructura de datos, facilita el análisis de datos y ayuda a formar bloques esenciales para las actividades de modelado. Los fundamentos conceptuales de los conjuntos difusos son atractivos, debido a su capacidad para cuantificar el nivel de la composición de elementos en los grupos detectados y la pertenencia parcial al grupo, [26]. Una rápida exploración, revela la importancia que tiene la investigación sobre agrupamiento difuso. Por ejemplo, en *IEEE Xplore* se pueden encontrar alrededor de 800 artículos, más de la mitad de estos se han publicado después del 2000.

El análisis de grupos (clusters) es una de las principales técnicas de reconocimiento de patrones, también es un acercamiento al aprendizaje no supervisado. Los métodos convencionales de agrupamiento (duros) restringen que cada punto del conjunto de datos pertenece exactamente a un cluster. La teoría de conjuntos difusos propuesta por Zadeh [27] en 1965 dio una idea de la incertidumbre de pertenencia que fue descrita por una función de pertenencia. El uso de los conjuntos difusos proporciona información imprecisa de la pertenencia a una clase [28].

Aplicaciones de la teoría de los conjuntos difusos y para el análisis de conglomerados de datos se propusieron en los trabajos de Bellman, Kalaba y Zadeh [29] y Ruspini [30]. Estos trabajos abren la puerta a la investigación en la agrupación difusa.

El agrupamiento es una tarea de aprendizaje no supervisado que tiene como objetivo descomponer un conjunto dado de objetos en subgrupos o grupos basados en la similitud. El objetivo es dividir el conjunto de datos de tal manera que los objetos, perteneciente al mismo grupo sean lo más similares posibles, mientras que los objetos pertenecientes a diferentes grupos sean tan diferentes como sea posible. El análisis de clusters es principalmente una herramienta para el descubrimiento de la estructura que antes estaba oculta en un conjunto de objetos desordenados. En este caso se supone que es la agrupación "verdadera" o natural que existe en los datos. Sin embargo, la asignación de objetos a las clases y la descripción de estas clases son desconocidas.

Los métodos de agrupación también se pueden utilizar para fines de reducción de datos. Basados en criterios matemáticos, se puede decidir sobre la composición de los grupos y así lograr la clasificación de conjuntos de datos de forma automática. Los métodos de agrupación utilizan distintas funciones de distancia que equivalen a una medición de similitud.

Un concepto común en todos los enfoques de agrupación es que están basados en los centros de cada grupo, es decir, el cluster está representado por un grupo prototipos $C_i,\ i = 1, \dots, c$. Los prototipos se utilizan para representar la estructura o distribución de los datos de cada clúster. Con esta representación del clúster, se denota formalmente el conjunto de prototipos como $\boldsymbol{C} = \{C_1, \dots, C_c\}$. Cada prototipo C_i es una n-tupla de parámetros que consiste en un centro del grupo c_i (parámetro de localización) y algunos parámetros adicionales sobre el tamaño y la forma del cluster.

Los algoritmos de C-medias difuso y C-posibilístico difuso se derivan del algoritmo de C-medias duro, también conocido como k-medias. En estos, cada prototipo único consiste en los vectores de centros, $\mathrm{C_i} = (\mathrm{c_i})$, de manera que los datos asignados a un grupo están representados por un punto prototipo en el espacio. Se considera una medida de distancia d, por ejemplo, la distancia Euclidiana.

Los algoritmos basados en funciones objetivos J, toman criterios matemáticos que cuantifican los modelos de cluster que integran los prototipos y la partición de datos. Las funciones objetivo tienen que reducirse al mínimo para obtener soluciones optimas, así una vez definido un criterio de optimización, la tarea de agrupamiento puede ser formulado como un problema de optimización de funciones. Es decir, los algoritmos deben determinar la mejor descomposición de un conjunto de datos en un número predefinido de clusters reduciendo al mínimo su función objetivo.

Las formas básicas de los algoritmos de partición dura, difusa, y C-medias posibilística buscan un número predefinido de los clusters c en un conjunto de datos, donde cada uno de los grupos está representado por el vector centro. Sin embargo, estos algoritmos difieren en la forma en que asignan los datos a los clusters. En el análisis clásico (duro) cada dato se asigna a exactamente a un solo clúster. El enfoque del agrupamiento difuso modifica el requisito de que los datos tienen que ser asignados a uno y sólo un cluster, aquí los puntos de datos puede pertenecer a más de un grupo e incluso con diferentes grados de pertenencia a los diferentes clusters [26].

CAPÍTULO 3. MARCO TEÓRICO

3.1 Pre-procesamiento de señales

El teorema de Nyquist o de muestreo, indica que para poder replicar con exactitud la forma de una onda es necesario que la frecuencia de muestreo sea superior al doble de la máxima frecuencia a muestrear. El aliasing se produce cuando la frecuencia de muestreo es inferior a la frecuencia Nyquist y por lo tanto insuficiente para hacer el muestreo correctamente con lo cual inventa frecuencias fantasmas que no tiene nada que ver con la original. Afecta más a las frecuencias altas, que se pierden antes, por lo tanto los tonos agudos se verán más afectados por el aliasing.

Los sonidos se caracterizan por el tono o frecuencia, intensidad o fuerza, y distribución espectral de energía o calidad. Una persona promedio puede escuchar de 20 a 20000 cps (ciclos o vibraciones por segundo). Los sonidos de alta frecuencia o de tono alto molestan más a la mayoría de las personas, que los sonidos de tono bajo de la misma intensidad. Sin embargo, los sonidos de tono alto se atenúan más rápidamente en el aire que los de tono bajo.

La intensidad es una evaluación subjetiva de la presión del sonido o su nivel. Debido a que la respuesta humana a la fuerza del sonido varía con la frecuencia, cualquier medida de fuerza debe incluir la frecuencia así como la presión o la intensidad para que pueda ser significativa. En acústica, la relación 10:1 se llama bel. En la práctica, la unidad que se utiliza con mayor frecuencia es el decibel (dB), que es igual a 0.1 bel.

El sonido es la vibración de un medio elástico, bien sea gaseoso, líquido o sólido. Cuando nos referimos al sonido audible por el oído humano, se está hablando de la sensación detectada por nuestro oído, que producen las rápidas variaciones de presión en el aire por encima y por debajo de un valor estático, dado por la presión atmosférica (alrededor de 100.000 pascales).

Cuando las variaciones de presión se centran entre 20 y 20.000 veces por segundo (igual a una frecuencia de 20 Hz a 20 kHz) el sonido es potencialmente audible. Los sonidos muy fuertes son causados por grandes variaciones de presión, por ejemplo una variación de 1 pascal se oiría como un sonido muy fuerte, siempre y cuando la mayoría de la energía de dicho sonido estuviera contenida en las frecuencias medias (1kHz - 4 kHz) que es donde el oído humano es más sensitivo.

La frecuencia de una onda sonora se define como el número de pulsaciones (ciclos) que tiene por unidad de tiempo (segundo). La unidad correspondiente a un ciclo por segundo es el *Hertz* (Hz). Las frecuencias más bajas corresponden a los sonidos "graves", son sonidos de vibraciones

lentas. Las frecuencias más altas corresponden a lo que llamamos sonidos "agudos" y son vibraciones muy rápidas. El espectro de frecuencias audible varía según cada persona, edad, etc. Sin embrago, se acepta como intervalo de 20 Hz a 20 kHz.

En acústica, el decibel se utiliza para comparar la presión sonora en el aire, con una presión de referencia. Este nivel de referencia es una aproximación al nivel de presión mínimo que hace que nuestro oído sea capaz de percibirlo. El nivel de referencia varía lógicamente según el tipo de medida que estemos realizando. No es el mismo nivel de referencia para la presión acústica, que para la intensidad acústica o para la potencia acústica. A continuación se dan los valores de referencia.

- Nivel de Referencia para la Presión Sonora (en el aire) = 0.00002 = 2E-5 Pa (rms)
- Nivel de Referencia para la Intensidad Sonora (en el aire) = 0.000000000001 = 1E-12 w/m^2
- Nivel de Referencia para la Potencia Sonora (en el aire) = 0.00000000001 = 1E-12 w

El nivel de presión sonora se calcula tomando que la intensidad acústica en el campo lejano es proporcional al cuadrado de la presión acústica y se define como:

$$Lp = 10\,log\left(\frac{p^2}{pr}\right) = 20\ log\left(\frac{p}{pr}\right) \qquad (3.1)$$

donde *Lp* = Nivel de Presión sonora; *p* la presión medida (Pa); *pr* la presión de referencia (2E-5 Pa), así el nivel de referencia siempre se corresponde con el nivel de 0 dB.

El uso del decibel (dB) permite la compresión o expansión de una escala, para la simplificación de cálculos en donde se involucran cantidades muy grandes. Además, hay que tener en cuenta que el comportamiento del oído humano está más cerca de una función logarítmica que de una lineal, ya que no percibe la misma variación en las diferentes escalas de nivel, ni en las diferentes bandas de frecuencias.

El ruido se define como un sonido indeseable, por ejemplo un sonido que para una persona no es demasiado fuerte, para otra puede ser molesto; lo que es confortable en una fábrica puede ser indeseable en una escuela; la música que disfruta un aficionado puede considerarse como ruido para un vecino que está tratando de dormir.

3.1.1 Wavelets

En el ámbito de las ciencias aplicadas usualmente se representa una señal física mediante una función del *tiempo* $\boldsymbol{s}(t)$ o en el dominio de la *frecuencia* por su Transformada de Fourier $\boldsymbol{s}(w)$. Ambas representaciones naturales, resultantes de la habitual modalidad de enfocar el universo real. Las mismas contienen exactamente la misma información sobre la señal, respondiendo a enfoques distintos y complementarios. Asumiendo que la señal es aperiódica y de energía finita, estas representaciones se relacionan mediante el par de fórmulas de *Fourier*:

$$\boldsymbol{s}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\boldsymbol{s}}(\omega) e^{l\,\omega t}\, d\omega \qquad (3.2)$$

$$\hat{\boldsymbol{s}}(\omega) = \int_{-\infty}^{\infty} \boldsymbol{s}(t) e^{-l\,\omega t}\, dt \qquad (3.3)$$

donde t representa el tiempo y ω la frecuencia angular. Por lo tanto, la información en uno de los dominios puede recuperarse a partir de la información desplegada en el otro. Esto plantea el problema de las representaciones en *tiempo-frecuencia*.

La transformada *wavelet* pertenece a una serie de técnicas de análisis de señal denominadas comúnmente análisis multi-resolución. Lo que significa que es capaz de variar la resolución de los parámetros que analiza (escala, concepto relacionado con la frecuencia y tiempo) a lo largo del análisis.

$$WT\boldsymbol{x}(\tau, a) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} \boldsymbol{x}(t) h * \left(\frac{t-\tau}{a}\right) dt \qquad (3.4)$$

La principal característica de este método es que permite conocer qué frecuencias componen una señal en cada instante con las siguientes resoluciones:

- Para las altas frecuencias consigue una buena resolución en el tiempo que permite su exacta localización temporal, aún a cambio de perder resolución en frecuencia.
- Para las componentes de bajas frecuencias lo más relevante es conocer su frecuencia aún a costa de perder resolución temporal.

Las wavelets son familias de funciones que se encuentran en el espacio y se emplean como funciones de análisis, examinan a la señal de interés para obtener sus características de espacio, tamaño y dirección; la familia está definida por:

$$h_{a,b} = \frac{h\left(\frac{x-b}{a}\right)}{\sqrt{|a|}} \qquad donde\ a, b\ \epsilon\ \mathbb{R}, \qquad a \neq 0 \tag{3.5}$$

Son generadas a partir de funciones madre **h**(x). A esa función madre se le agregan un par de variables que son la escala (a) que permite hacer dilataciones y contracciones de la señal y la variable de traslación (b), que nos permite mover la señal en el tiempo.

Existen diferentes wavelets que tienen definiciones establecidas, sin embargo la elección de un tipo de wavelet depende de la aplicación específica que se le vaya a dar. En este trabajo se hace uso de la transformada wavelet con filtro Daubechies, ya que es la que presenta mejores resultados para esta aplicación [31].

La transformada wavelet con filtro Daubechies puede tener orden N, dependiendo del número de momentos de desvanecimiento que se deseen, N es un entero positivo y denota el número de coeficientes del filtro que tiene esa wavelet. En la Fig. 3.1 se observa la wavelet con filtro Daubechies de orden 4. Esta wavelet cuenta con las características de ortogonalidad y biortogonalidad. La ecuación que describe la función wavelet con filtro Daubechies, se muestra a continuación:

$$m_0(\varepsilon) = \left(\frac{1+e^{-2\pi i\varepsilon}}{2}\right)^2 \left(\frac{1+\sqrt{3}}{2} + \frac{1-\sqrt{3}}{2}\right) e^{-2\pi i\varepsilon} \tag{3.6}$$

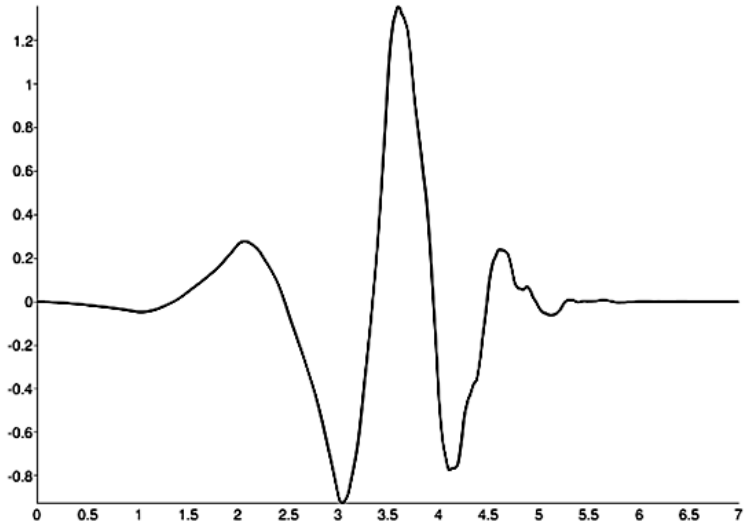


Fig. 3.1. Transformada Wavelet con filtro Daubechies de orden 4.

3.1.2 Blanqueado

En la mayoría de los trabajos desarrollados para el Análisis de Componentes Independientes, se utiliza una primera etapa conocida como *blanqueado espacial (whitening),* en la cual se realiza una transformación $\boldsymbol{V}$ de las señales de entrada $\boldsymbol{e}(t)$, de forma que resultan en la señal $\boldsymbol{z}(t)$, con media cero, varianza unitaria y decorrelacionadas, como se indica a continuación:

$$\boldsymbol{z}(t) = \boldsymbol{V}\boldsymbol{e}(t) \tag{3.7}$$

Este tratamiento de la señal simplifica el problema de ICA. Con esta transformación las señales se decorrelacionan, dicha condición es necesaria pero no suficiente para la independencia estadística. El primer paso a realizar en el blanqueado, es hacer que los vectores de entrada $\boldsymbol{e}(t)$tengan media cero, con el propósito de obtener el centrado de las variables. Esto se logra restando su media, es decir:

$$\boldsymbol{e}(t + \Delta t) \leftarrow \boldsymbol{e}(t) - E\{\boldsymbol{e}(t)\} \tag{3.8}$$

Las componentes de los vectores blanqueados $\boldsymbol{z}(t)$ resultantes tienen la característica de estar decorrelacionadas y con su varianza normalizadas. Por lo tanto se tiene que:

$$E\{\boldsymbol{z} \cdot \boldsymbol{z}^T\} = \boldsymbol{I} \tag{3.9}$$

La medida de la independencia lineal de los vectores se denominada *correlación*. Por lo tanto la decorrelación o correlación cero, indica que se tiene una independencia lineal entre vectores. Esto es, dos variables y_1 y y_2 se dice que están decorrelacionadas si su covarianza es cero.

$$cov(\boldsymbol{y}_1, \boldsymbol{y}_2) = E\{\boldsymbol{y}_1\boldsymbol{y}_2^T\} - E\{\boldsymbol{y}_1\}E\{\boldsymbol{y}_2^T\} = 0 \tag{3.10}$$

Asumiendo que las variables tienen media de cero, su covarianza es igual a su correlación:

$$corr(\boldsymbol{y}_1, \boldsymbol{y}_2) = E\{\boldsymbol{y}_1\boldsymbol{y}_2\} \tag{3.11}$$

La decorrelación de los datos de entrada $\boldsymbol{e}(t)$está dada por la transformación de la matriz $\boldsymbol{V}$. Para encontrar esta matriz, cuyos vectores son linealmente independientes, se parte de la matriz de covarianza $\boldsymbol{C}_e = E\{\boldsymbol{e} \cdot \boldsymbol{e}^T\}$ de los vectores de entrada $\boldsymbol{e}(t)$. Así, la matriz $\boldsymbol{V}$ es estimada a partir de los valores y los vectores propios (normalizados) de la matriz de covarianza ($\boldsymbol{C}_e$). Los vectores propios ($\boldsymbol{E}$) proporcionan la dirección de las componentes, mientras que los valores propios ($\boldsymbol{D}$) proporcionan la amplitud de las distribuciones (varianza). Por lo tanto la transformación $\boldsymbol{V}$ está dada por:

$$\boldsymbol{V} = \boldsymbol{D}^{-1/2}\boldsymbol{E}^T \tag{3.12}$$

Esta matriz existe siempre y cuando los valores propios sean positivos. Escribiendo la correlación en términos de sus vectores ($\boldsymbol{E}$) y valores propios ($\boldsymbol{D}$), se denota como:

$$\boldsymbol{C}_e = \boldsymbol{EDE}^T \tag{3.13}$$

Donde $\boldsymbol{E}$ representa los vectores propios, cuya característica es la de ortogonalidad, y sus vectores satisfacen la siguiente condición:

$$\boldsymbol{E}^T\boldsymbol{E} = \boldsymbol{EE}^T = \boldsymbol{I} \tag{3.14}$$

Esto demuestra que:

$$E\{\boldsymbol{z}\cdot\boldsymbol{z}^T\} = \boldsymbol{V}E\{\boldsymbol{e}\cdot\boldsymbol{e}^T\}\boldsymbol{V}^T = \boldsymbol{D}^{-1/2}\boldsymbol{E}^T\boldsymbol{EDE}^T\boldsymbol{ED}^{-1/2} = \boldsymbol{I} \tag{3.15}$$

Donde $\boldsymbol{D}$ es una matriz de los valores propios $\boldsymbol{D} = diag(d_1, \dots, d_n)$, que representan la varianza. Considerando la normalización de las varianzas, se demuestra que la covarianza de z es una matriz unitaria, tomando el término de *blanqueado*. De acuerdo con las expresiones anteriores se puede deducir que:

a) Los vectores propios de la matriz de correlación $\boldsymbol{R}$ correspondientes a los vectores de entrada $\boldsymbol{e}(t)$, de media cero, definen vectores unitarios representando las direcciones principales a lo largo de las cuales las varianzas toman sus valores máximos.
b) Los valores propios asociados, definen los valores máximos de esas varianzas.
c) Proporciona datos decorrelacionados.

Otra manera para efectuar el blanqueado es utilizando la siguiente matriz de blanqueado propuesta por Bell y Sejnowski, en su procedimiento para la separación de fuentes basado en la maximización de la entropía:

$$\boldsymbol{V} = 2\sqrt{E^{-1}\{\boldsymbol{e}\cdot\boldsymbol{e}^T\}} \tag{3.16}$$

Los trabajos con enfoque estadístico dividen el problema de separación en dos etapas, ver Fig. 3.2. En la primera se transforman las señales de entrada, $\boldsymbol{e}(t)$, para obtener las señales $\boldsymbol{x}(t)$, con media cero, decorrelacionadas y con varianza unitaria. Es decir, la transformación $\boldsymbol{V}$ de (5.3), debe ser tal que:

$$\langle\boldsymbol{x}(t)\rangle = 0 \quad y \quad \langle\boldsymbol{x}(t)\boldsymbol{x}^T(t)\rangle = \boldsymbol{I} \tag{3.17}$$

con ello, las covarianzas son nulas: $< \boldsymbol{x}_i \boldsymbol{x}_j >= 0$, *si* $i \neq j$, y las varianzas de cada señal son la unidad $< \boldsymbol{x}_i^2 >= 1, \forall i$. La decorrelación es un prerrequisito necesario (*pero no suficiente*) para obtener la independencia estadística.

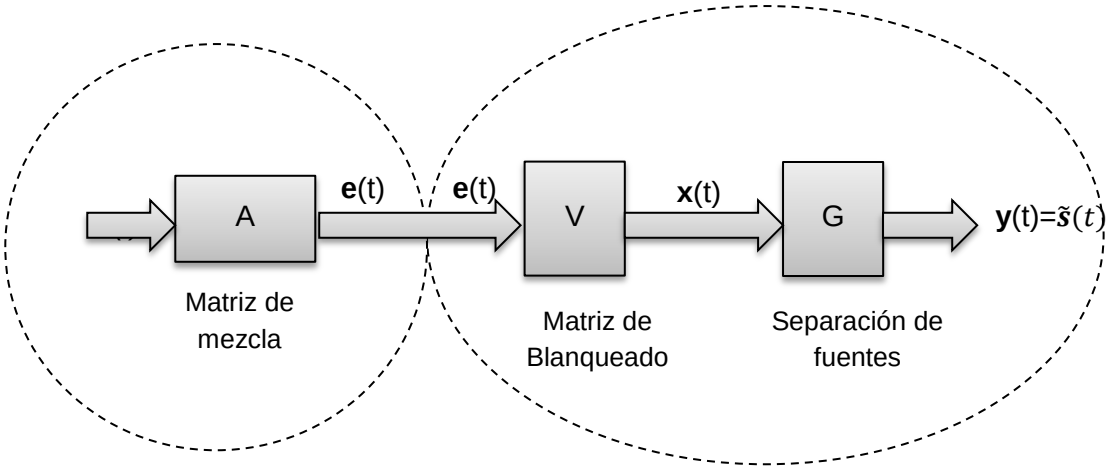
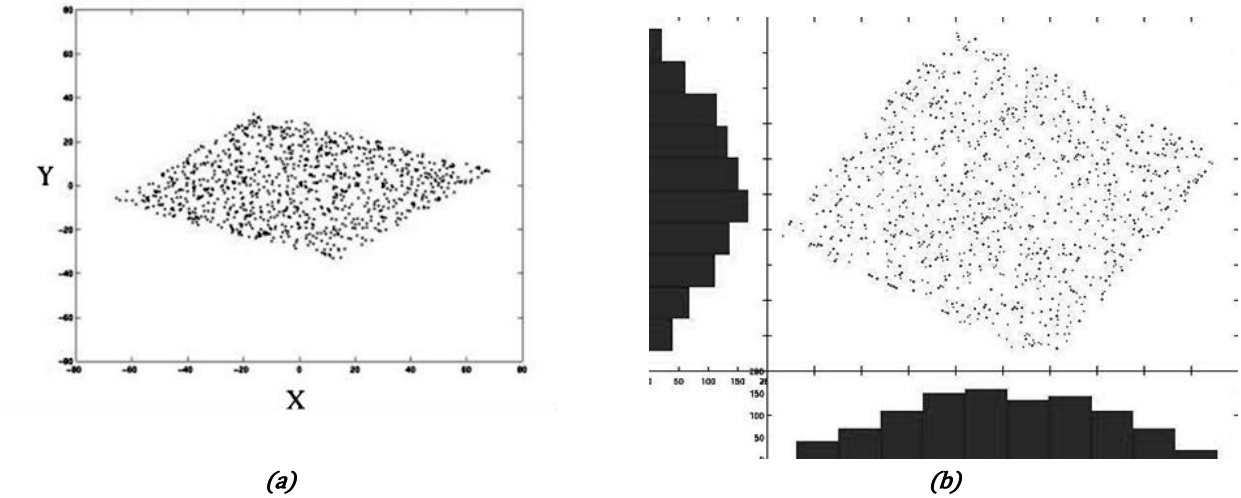


Fig. 3.2 Diagrama de bloques de la separación ciega de fuentes en dos etapas

El blanqueado espacial consiste básicamente de tres procesos, que son:

- Centrado de las variables,
- Normalización de la varianza, y
- Decorrelación.

En la Fig. 3.3 se muestran los planos bidimensionales a) de dos señales mezcladas y en b) las señales decorrelacionadas, cuyos vectores principales se encuentran en la dirección de máxima varianza y muestran ortogonalidad.



(a) *(b)*

Fig. 3.3. Distribución de dos señales blanqueadas. a) Señales mezcladas, b) Señales blanqueadas.

Estos procesos son necesarios para los algoritmos de ICA, basados en la maximización de la *no-gaussianidad*, como INFOMAX. Debido a que la decorrelación es una condición necesaria pero no suficiente para la independencia estadística, al encontrar una de las direcciones de las componentes independientes (*señal blanqueada*), se efectúa una rotación con el propósito de que cada una de las componentes se localice en la dirección de los ejes coordenados, cuya característica es la de independencia. A continuación, se presentan las etapas para la implementación del algoritmo del blanqueado.

1. Centrado de las variables, el cual se logra restando su media aritmética:

$$\boldsymbol{e}(t) \leftarrow \boldsymbol{e}(t) - E\{\boldsymbol{e}(t)\} \tag{3.18}$$

2. Estimación de la matriz $\boldsymbol{V}$, que es estimada a partir de la siguiente expresión (*decorrelación*):

$$\boldsymbol{V} = \boldsymbol{D}^{-1/2}\boldsymbol{E}^{T} \tag{3.19}$$

donde $\boldsymbol{D}$ es una matriz de los valores propios $\boldsymbol{D} = diag(d_1, \dots, d_n)$ y $\boldsymbol{E}$ representa los vectores propios [32].

3. Transformación lineal $\boldsymbol{V}$, donde $E\{\boldsymbol{z} \cdot \boldsymbol{z}^T\} = \boldsymbol{I}$ (*normalización*):

$$\boldsymbol{z}(t) = \boldsymbol{V}\boldsymbol{e}(t) \tag{3.20}$$

3.2 Análisis de componentes independientes

La separación ciega de fuentes (*Blind Source Separation, BSS*) consiste en la recuperación de señales originales a partir de mezclas observadas. El éxito de los algoritmos consiste en separar cada una de estas señales originales y escucharlas una a una sin influencia de las demás, es un concepto más amplio que el de *eliminación de ruido*. Al realizar la separación, las fuentes individuales se denominan *componentes independientes*, de ahí el término análisis en componentes independientes (*Independent Component Analysis, ICA*) íntimamente ligado al de separación ciega de fuentes.

Para el estudio de la separación de p fuentes desconocidas a partir de q señales mezcladas o sensores, se establecen las siguientes consideraciones:

1. Las fuentes son estadísticamente independientes entre sí.
2. Las fuentes tienen distribución no-Gaussiana, como máximo sólo una de ellas puede ser de distribución Gaussiana.
3. El modelo de mezcla se conoce (lineal, convolutivo o no-lineal).

Además de estas consideraciones, se hace la indicación del número de señales y de sensores, y los procedimientos dependerán del valor $p < q, p = q, p > q$ y combinaciones. El modelo general de mezcla es el siguiente:

$$\boldsymbol{Y}(t) = \boldsymbol{H}\big(\boldsymbol{s}(t), \dots, \boldsymbol{s}(t-m)\big) + \boldsymbol{B}(t) \qquad (3.21)$$

donde $\boldsymbol{H}$ es una función desconocida que depende sólo del canal y de los sensores, $\boldsymbol{B}(t)$ es un vector de ruido Gaussiano aditivo independiente de las fuentes. A partir de (3.21) se pueden definir tres tipos de mezclas:

A) El *modelo lineal instantáneo* de mezcla, donde la función $\mathbf{H}$ en (3.21) es una matriz real al suponerse que el canal no tiene memoria, verificándose que:

$$\boldsymbol{Y}(n) = \mathbf{H}\boldsymbol{S}(n) + \boldsymbol{B}(n) \qquad (3.22)$$

siendo n el tiempo discretizado e $\boldsymbol{Y}(n)$, $\boldsymbol{S}(n)$ y $\boldsymbol{B}(n)$ vectores que contienen las señales fuente y ruido.

B) El *modelo convolutivo*, una generalización lineal de (3.21) en el que se supone que el canal tiene memoria, verificándose que:

$$\boldsymbol{Y}(n) = [\boldsymbol{H}(z)]\boldsymbol{S}(n) + \boldsymbol{B}(n) = \sum_{l} \mathbf{H}(l)\boldsymbol{S}(n-l) + \boldsymbol{B}(n) \qquad (3.23)$$

donde $\mathbf{H}(l)$ es una matriz $q \times p$ que representa el efecto del canal y $\boldsymbol{H}(z)$ es una matriz de filtros lineales indicando el efecto de la fuente i sobre la observación j , es decir:

$$[\boldsymbol{H}(z)] = \big(h_{ji}(z)\big) = \sum_{l} \mathbf{H}(l)z^{-1} \qquad (3.24)$$

C) *El modelo no-lineal*, en el que la función $\boldsymbol{H}$ en (3.21) es cualquier función no-lineal. Una aproximación que hacen algunos autores consiste en suponer un modelo post-nolineal, es decir, la aplicación de una función no-lineal a una mezcla lineal, quedando así la ecuación (3.21) como:

$$\boldsymbol{Y}(t) = \boldsymbol{H}_1[\mathbf{H}_2\boldsymbol{S}(n)] + \boldsymbol{B}(t) \qquad (3.25)$$

donde $\boldsymbol{H}_1$ una función no-lineal tipo sigmoide, por ejemplo la tangente hiperbólica y $\boldsymbol{H}_2$ una matriz real.

La primera consideración indica que las fuentes deben ser estadísticamente independientes entre sí, es fundamental. Por definición, dos variables aleatorias u_i y u_j son independientes si su función de densidad de probabilidad conjunta es:

$$p(u_i, u_j) = p(u_i)p(u_j) \tag{3.26}$$

Para verificar la hipótesis de independencia, se utilizan una serie de conceptos entre los que destacan los siguientes:

1. Divergencia de Kullback-Leibler. Se define como:

$$\delta(pu_i\,, pu_j) = \int p_{u_i}(v) log \frac{p_{u_i}(v)}{p_{u_j}(v)} dv \geq 0 \tag{3.27}$$

de manera que

$$\delta\left(p(u_i)\,, p(u_j)\right) = 0 \qquad si \quad p(u_i) = p(u_j) \tag{3.28}$$

2. Una distancia similar, la Información Mutua, se define como:

$$i(p_u) = \int p_u(\boldsymbol{V}) log \frac{p_u(\boldsymbol{V})}{\prod_{i=1}^{N} p_{u_i}(v_i)} d\boldsymbol{V} \geq 0 \tag{3.29}$$

donde $\boldsymbol{V}$ es un vector aleatorio y v_i son sus componentes. La información mutua vale cero si y sólo si estas componentes son independientes entre sí.

3. Momentos y Cumulantes. Ambos se definen a partir de funciones características. La primera función característica es:

$$\Phi_U(\boldsymbol{V}) = \int \exp(j\boldsymbol{V}^T\boldsymbol{U})\ dF(\boldsymbol{U}) \tag{3.30}$$

donde $F(\boldsymbol{U})$ es la función de distribución de $\boldsymbol{U}$. La segunda función característica es:

$$\Psi_U(\boldsymbol{V}) = \ln\{\Phi(\boldsymbol{V})\} \tag{3.31}$$

El momento de orden q del vector $\boldsymbol{U}$ se define como:

$$Mom_q(u_1, \dots, u_q) \leq u_1, \dots, u_q \geq (-j)^q \frac{\partial^q \Phi_U(\boldsymbol{V})}{\partial v_1 \dots \partial v_{q\|V=0}} \tag{3.32}$$

Y el cumulante de orden q de $\boldsymbol{U}$ se define como:

$$Cum_q(u_1, \dots, u_q) = (-j)^q \frac{\partial^q \Psi_U(\boldsymbol{V})}{\partial v_1 \dots \partial v_q}|_{V=0} \qquad (3.33)$$

La independencia estadística de las señales implica que los cumulantes cruzados de todos los órdenes deben ser cero, aunque en la práctica sólo se consideran cumulantes de orden cuatro como máximo. La estadística de orden dos no es suficiente para llevar a cabo la separación, si las fuentes tienen matrices de covarianza similares. La estadística de orden tres tampoco es válida para señales con función de densidad de probabilidad simétrica [33].

3.2.1 Mezclas lineales

El análisis de componentes independiente (*ICA*) tiene un enfoque estadístico y parte de la hipótesis de que las señales originales o fuentes son estadísticamente independientes y los procedimientos que lo siguen, se basan en propiedades de las fuentes y de las mezclas realizadas, caracterizadas por sus distribuciones de probabilidad. Se pretende que si las señales originales son estadísticamente independientes, las señales recuperadas también deben de serlo. Por lo tanto, los métodos relacionados con ICA consideran como condición primordial la independencia estadística, buscando dentro de espacios multidimensionales las componentes principales de las variables a determinar. El modelo para mezclas lineales o instantáneas, tiene la siguiente forma general:

$$\boldsymbol{e}(t) = \boldsymbol{A}\boldsymbol{s}(t) \qquad (3.34)$$

Donde $\boldsymbol{A}$, es la matriz de mezcla, $\boldsymbol{s}(t)$ son las n señales fuente, y $\boldsymbol{e}(t)$ son las señales mezcladas. En este caso se considera que tanto el medio como los sensores se comportan de forma lineal, Fig. 3.4.

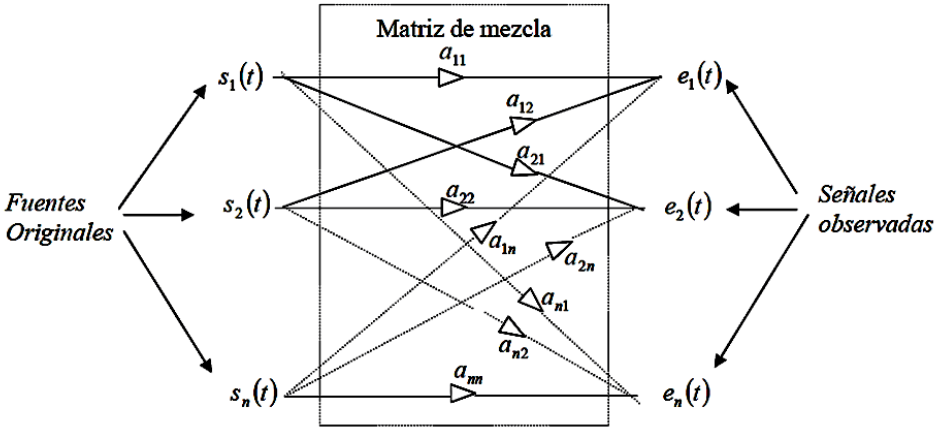


Fig. 3.4 Modelo de mezcla lineal

3.2.1.1 Maximización de la información

El principio de maximización de la información también llamado INFOMAX, se encuentra muy relacionado al de máxima verosimilitud, este método se basa en la medida de la entropía relativa de dos o más distribuciones de probabilidad.

En la Fig. 3.5 se muestra la arquitectura de la red neuronal, en la que se efectúa la maximización de la información de transferencia y la minimización de la información mutua entre sus salidas. Cada entrada $e_i(t)$ está formada por la mezcla de las señales y cada salida $u_i(t)$ es la señal estimada de la fuente independiente.

El objetivo es encontrar una matriz ***W*** inversa a la matriz de mezcla ***A***, que recupere las señales de las fuentes independientes. La regla de aprendizaje, propuesta por Bell y Sejnowski [34] para obtener la matriz ***W*** se deriva de la maximización de la entropía $H(\hat{\boldsymbol{S}})$ de las salidas de la red.

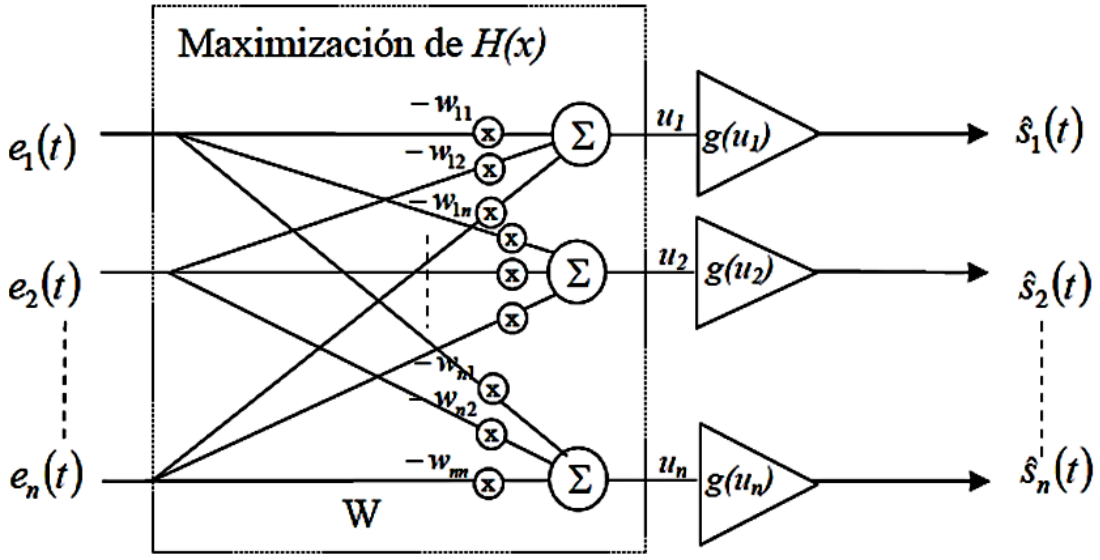


Fig. 3.5. Arquitectura de red neuronal para algoritmo INFOMAX

INFOMAX está basado en conocer la razón de información, por lo tanto es necesario definir conceptos como la entropía, que se encuentra definida como:

$$H(\boldsymbol{X}) = \sum_{i=1}^{n} P(x_i) log \frac{1}{P(x_i)} \qquad (3.35)$$

Donde $\boldsymbol{X}$ es el conjunto de variables aleatorias de x_i. La entropía conjunta de la salida de la red, está dada por:

$$H(\hat{\boldsymbol{s}}_1, \hat{\boldsymbol{s}}_2, \dots, \hat{\boldsymbol{s}}_n) = H(\hat{\boldsymbol{s}}_1) + H(\hat{\boldsymbol{s}}_2) + \cdots + H(\hat{\boldsymbol{s}}_n) - I(\hat{\boldsymbol{s}}_1, \hat{\boldsymbol{s}}_2, \dots, \hat{\boldsymbol{s}}_n) \tag{3.36}$$

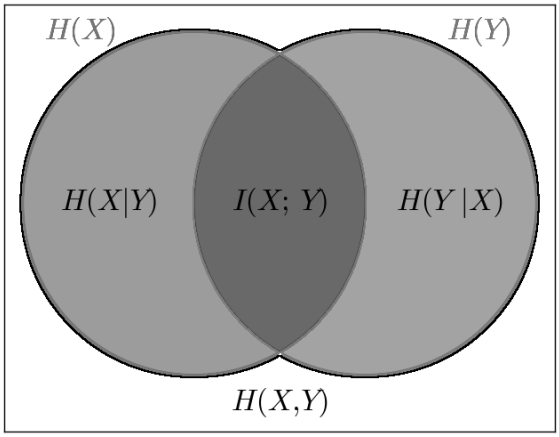


Fig. 3.6 Entropía conjunta

Donde $H(\hat{\boldsymbol{s}}_i)$ son las entropías marginales de la *i-esima* salida, e $I(\hat{\boldsymbol{s}}_1, \hat{\boldsymbol{s}}_2, \dots, \hat{\boldsymbol{s}}_n)$ es la información mutua entre las salidas. Maximizar la entropía conjunta $H(\hat{\boldsymbol{s}}_1, \hat{\boldsymbol{s}}_2, \dots, \hat{\boldsymbol{s}}_n)$ de la salida de la red, equivale a maximizar las entropías marginales $H(\hat{\boldsymbol{s}}_i)$ y minimizar la información mutua $I(\hat{\boldsymbol{s}}_i)$. Por lo tanto, el valor máximo que se puede alcanzar, es cuando no existe información mutua entre sus salidas. En consecuencia, el término de $I(\hat{\boldsymbol{s}}_1, \hat{\boldsymbol{s}}_2, \dots, \hat{\boldsymbol{s}}_n)$ será igual a cero y su entropía marginal será igual a su entropía conjunta (entropía máxima), [35].

$$H(\hat{\boldsymbol{s}}_1, \hat{\boldsymbol{s}}_2, \dots, \hat{\boldsymbol{s}}_n) = H(\hat{\boldsymbol{s}}_1) + H(\hat{\boldsymbol{s}}_2) + \cdots + H(\hat{\boldsymbol{s}}_n) \tag{3.37}$$

La función de transferencia entre $\hat{s}_i$ y u_i se denota como:

$$\frac{p(u_i)}{p(\hat{s}_i)} = \left|\frac{\partial g_i(u_i)}{\partial u_i}\right| \tag{3.38}$$

La relación entre $\hat{\boldsymbol{s}}_{\boldsymbol{i}}$, u_i y la función de transferencia no lineal es:

$$p(\hat{\boldsymbol{s}}_i) = \frac{p(u_i)}{\left|\frac{\partial g_i(u_i)}{\partial u_i}\right|} \tag{3.39}$$

Para una distribución uniforme de $\hat{\boldsymbol{s}}_i$, se tiene que:

$$p(u_i) = \left|\frac{\partial g_i(u_i)}{\partial u_i}\right| \tag{3.40}$$

La función no lineal $g_i(u_i)$ es empleada principalmente para minimizar la información mutua y así obtener salidas independientes, además permite la minimización de la correlación de orden superior (por su expansión en series de Taylor). Como se observa en la ecuación (3.36), la independencia estadística se obtiene cuando la información mutua de la red es cero y la entropía conjunta es igual a la suma de sus entropías marginales, donde la expresión de la entropía marginal se representa como:

$$H(\hat{\boldsymbol{s}}_i) = -E\{\log p(\hat{\boldsymbol{s}}_i)\} \tag{3.41}$$

La no linealidad realiza una transformación de la variable u_i a $\widehat{\boldsymbol{s}_i}$. Por lo tanto la ecuación (3.39) se puede expresar como:

$$p(\hat{\boldsymbol{s}}_i) = \frac{p(u_i)}{\left|\frac{\partial \hat{\boldsymbol{s}}_i}{\partial u_i}\right|} \tag{3.42}$$

Sustituyendo en la ecuación (3.41), tenemos:

$$H(\hat{\boldsymbol{s}}_i) = -E\left\{\log\left(\frac{p(u_i)}{\left|\frac{\partial \hat{\boldsymbol{s}}_i}{\partial u_i}\right|}\right)\right\} \tag{3.43}$$

Ahora reescribiendo la ecuación (3.36), se obtiene de forma compacta:

$$H(\hat{\boldsymbol{s}}) = -\sum_{i=1}^{n}\left\{E\left(\log\left(\frac{p(u_i)}{\frac{\partial \hat{\boldsymbol{s}}_i}{\partial u_i}}\right)\right)\right\} - I(\hat{\boldsymbol{s}}) \tag{3.44}$$

Realizando la derivada de la ecuación anterior respecto a la matriz $\boldsymbol{W}$ se obtiene:

$$\frac{\partial H(\hat{\boldsymbol{s}})}{\partial \boldsymbol{W}} = \frac{\partial}{\partial \boldsymbol{W}}(-I\,(\hat{\boldsymbol{s}})) - \frac{\partial}{\partial \boldsymbol{W}}\sum_{i=1}^{n}\left\{E\left(\log\left(\frac{p(u_i)}{\frac{\partial \hat{\boldsymbol{s}}_i}{\partial u_i}}\right)\right)\right\} \tag{3.45}$$

En esta ecuación se observa la relación de la entropía conjunta con la información mutua. La minimización directa de la información mutua genera la maximización de la entropía conjunta. En el caso en el que la fuente estimada y la función no lineal resulten en un valor $I(\hat{s})$ diferente de cero, existirá un error. Por lo tanto, en la ecuación (3.45) se plantea que el proceso de minimización de $\boldsymbol{I}(\hat{s})$ depende de las fuentes estimadas y de la función no lineal empleada. Sin embargo, el término de error para las aplicaciones propuestas empleando una función logística y señales con distribución super-gaussiana, se puede considerar como despreciable. En este caso, el término del error desaparece y el máximo de la entropía conjunta se obtiene derivando la entropía $H(\hat{s})$ con respecto a $\boldsymbol{W}$. Esto es calculando el gradiente de $H(\hat{s})$.

$$\frac{\partial H(\hat{s})}{\partial \boldsymbol{W}} = \frac{\partial}{\partial \boldsymbol{W}}(-E\{\log|\boldsymbol{J}|\}) = \frac{\partial}{\partial \boldsymbol{W}}\log|\boldsymbol{J}| \tag{3.46}$$

El término $\boldsymbol{J}$ es el valor absoluto del Jacobiano. Donde la relación entre la densidad de salida $p(\hat{\boldsymbol{s}})$ y la densidad de entrada $p(\boldsymbol{e})$ puede ser definida como $p(\hat{\boldsymbol{s}}) = p(\boldsymbol{e})\,\boldsymbol{J}(\boldsymbol{e})$ [35]. Por lo tanto $\boldsymbol{J}$ puede ser definido como una transformación de $p(\boldsymbol{e})$ a $p(\hat{\boldsymbol{s}})$:

$$\boldsymbol{J}(\boldsymbol{e}) = det\begin{bmatrix} w_{11}\dfrac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{1}}}{\partial u_1} & \cdots & w_{1n}\dfrac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{1}}}{\partial u_n} \\ \vdots & \ddots & \vdots \\ w_{n1}\dfrac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{n}}}{\partial u_1} & \cdots & w_{nn}\dfrac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{n}}}{\partial u_n} \end{bmatrix} \tag{3.47}$$

Como no existe una conexión entre las salidas después de las neuronas, las derivadas parciales de $\partial \hat{\boldsymbol{s}}_{\boldsymbol{i}}/\partial u_j$, tomarán un valor diferente de cero únicamente en los casos donde $i = j$. Por lo tanto, la ecuación anterior se puede definir como:

$$\boldsymbol{J}(e) = det(\boldsymbol{W})\prod_{i=1}^{n}\left|\frac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{i}}}{\partial u_i}\right| \tag{3.48}$$

Substituyendo en la ecuación (3.46), tenemos:

$$\frac{\partial H(\hat{\boldsymbol{s}})}{\partial \boldsymbol{W}} = \frac{\partial}{\partial \boldsymbol{W}}\log\left(|\det(\boldsymbol{W})|\prod_{i=1}^{n}\left|\frac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{i}}}{\partial u_i}\right|\right) = \frac{\partial}{\partial \boldsymbol{W}}\log|\det(\boldsymbol{W})| + \sum_{i=1}^{n}\frac{\partial}{\partial \boldsymbol{W}}\log\left|\frac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{i}}}{\partial u_i}\right| \tag{3.49}$$

El primer término de la ecuación anterior se puede expresar como:

$$\frac{\partial}{\partial \boldsymbol{W}} \log|\det(\boldsymbol{W})| = \frac{(adj\ \boldsymbol{W})^T}{\det \boldsymbol{W}} = (\boldsymbol{W}^T)^{-1} \tag{3.50}$$

El segundo término de la ecuación (3.49) puede ser determinado como:

$$\frac{\partial}{\partial w_{ij}} \sum_{i=1}^{n} \log \left| \frac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{i}}}{\partial u_i} \right| = \frac{1}{\frac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{i}}}{\partial u_i}} \frac{\partial^2 \hat{\boldsymbol{s}}_{\boldsymbol{i}}}{\partial {u_i}^2} e_j^T \tag{3.51}$$

Se define la derivada de la no linealidad $\hat{\boldsymbol{s}}_{\boldsymbol{i}}$ con respecto a u_i como una aproximación de la densidad de la fuente $p(u_i)$, de la siguiente manera:

$$p(u_i) = \frac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{i}}}{\partial u_i} \tag{3.52}$$

Por lo tanto, se obtiene de la ecuación (3.51)

$$\frac{\partial}{\partial w_{ij}} \sum_{i=1}^{n} \log \left| \frac{\partial \hat{\boldsymbol{s}}_{\boldsymbol{i}}}{\partial u_i} \right| = \frac{\frac{\partial p(u)}{\partial u}}{p(u)} e^T \tag{3.53}$$

Sustituyendo en ambos términos en la ecuación (3.49), se obtiene:

$$\frac{\partial H(\hat{s})}{\partial \boldsymbol{W}} = \left(\boldsymbol{W}^{\boldsymbol{T}}\right)^{-1} + \left(\frac{\frac{\partial p(u)}{\partial u}}{p(u)} \right) e^T \tag{3.54}$$

Esta ecuación es resultado del gradiente de la función de la entropía. Una forma de maximizar la entropía es por medio del gradiente *natural*, el cual se logra multiplicando por $\boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{W}$. Por lo tanto, la ecuación anterior se expresa como:

$$\Delta \boldsymbol{W} \propto \frac{\partial H(\hat{\boldsymbol{s}})}{\partial \boldsymbol{W}} \boldsymbol{W}^{\boldsymbol{T}} \boldsymbol{W} = \left[\boldsymbol{I} + \left(\frac{\frac{\partial p(u)}{\partial u}}{p(u)} \right) \boldsymbol{u}^{\boldsymbol{T}} \right] \boldsymbol{W} \tag{3.55}$$

Donde **I** es la matriz identidad y el segundo factor denota la no linealidad. Definiendo el siguiente termino $\varphi(\boldsymbol{u})$ como una función no lineal:

$$\varphi(\boldsymbol{u}) = -\frac{\frac{\partial p(\boldsymbol{u})}{\partial \boldsymbol{u}}}{p(\boldsymbol{u})} \tag{3.56}$$

Sustituyendo en la ecuación (3.55) se obtiene la siguiente ecuación, que se conoce como *regla de aprendizaje,* y se emplea en el modelo neuronal recursivo para el ajuste de los pesos sinápticos en donde en cada iteración se realiza la actualización hasta encontrar los valores óptimos.

$$\Delta \boldsymbol{W} \propto [\boldsymbol{I} - \varphi(u)u^T]\boldsymbol{W} \tag{3.57}$$

La función no lineal propuesta por Bell y Sejnoski fue una función logística $g(u) = tanh(u)$, así la regla de aprendizaje se rescribe como:

$$\Delta \boldsymbol{W} \propto [\boldsymbol{I} - tanh(u)u^T]\boldsymbol{W} \tag{3.58}$$

A este algoritmo de aprendizaje se le conoce como INFOMAX original, ya que solo puede separar señales con distribuciones súper gaussianas. La ecuación muestra el caso general para distribuciones sub y super gaussianas, conocido como INFOMAX extendido, desarrollado por Girolami.

$$\Delta \boldsymbol{W} \propto [\boldsymbol{I} - \boldsymbol{K}\, tanh(u)u^T - uu^T]\boldsymbol{W} \tag{3.59}$$

Donde $\boldsymbol{K}$ representa una matriz con elementos sólo en su diagonal principal, los cuales efectúan una conmutación de signo para cada una de las distribuciones. Girolami empleó el signo de la curtosis como criterio para definir la conmutación.

k_i = 1 : distribución súper-gaussiana.
k_i = -1 : distribución sub-gaussiana.

3.2.1.2 Algoritmo de punto fijo

El algoritmo FastICA o de punto fijo es una versión mejorada del gradiente, por lo que comenzamos con este método. El método del gradiente de la curtosis busca maximizar el valor absoluto de la curtosis, el cual consiste en derivar a la curtosis respecto a **w**.

$$\frac{\partial|kur(w^T z)|}{\partial w} = 4\, sign(kurt(w^T z))[E\{z(w^T z)^3\} - 3w\|w\|^2] \quad (3.60)$$

Obteniendo la dirección del vector w en donde el valor absoluto de la curtosis de $y = w^T z$ es máximo, la proyección de w se efectuará dentro de un círculo unitario en donde en cada iteración efectuará una normalización de w. Por lo tanto, el algoritmo del gradiente se expresa como:

$$\Delta\, w \propto sign\big(kurt(w^T z)\big)E\{z(w^T z)^3\} \quad (3.61)$$

$$w \leftarrow \frac{w}{\|w\|} \quad (3.62)$$

Una versión en línea de este algoritmo se obtiene omitiendo el valor esperado de la ecuación (3.61) resultando en:

$$\Delta\, w \propto sign\big(kurt(w^T z)\big)z(w^T z)^3 \quad (3.63)$$

También la curtosis puede ser estimada en línea utilizando la siguiente expresión iterativa:

$$\Delta\, \gamma \propto ((w^T z)^4 - 3) - \gamma \quad (3.64)$$

FastICA es un algoritmo de punto fijo, el cual consiste en realizar iteraciones en la dirección del gradiente, multiplicado por un escalar para encontrar el valor óptimo de w, resultando en una versión más eficiente, alcanzando una convergencia más rápida y más estable, como lo indica la siguiente expresión:

$$w \propto [E\{z(w^T z)^3\} - 3\|w\|^2 w] \quad (3.65)$$

El algoritmo de punto fijo calcula el nuevo valor de w en cada iteración asignado un nuevo valor, de la siguiente manera:

$$w \longleftarrow E\{z(w^T z)^3\} - 3w \quad (3.66)$$

En cada iteración y actualización de w se efectuará la división entre su norma conservando su magnitud. El vector resultante w, da la dirección de una de las componentes independientes de la señal blanqueada. Esto es, cuando la red converge, el vector actual y el nuevo apuntarán en la misma dirección (w ó $-w$) y por lo tanto la actualización será la misma. El algoritmo FastICA tiene la ventaja sobre el del gradiente, en que su comportamiento es cúbico, lo cual hace su convergencia más rápida [32].

Los pasos para implementar el algoritmo FASTICA [36], son los siguientes:

1. Inicializar aleatoriamente el vector $w(0)$ de norma 1. Tomar k=1
2. Sea $\mathrm{w(k)} = \mathrm{E}\left\{\mathrm{x}\left(\mathrm{w(k-1)^T x}\right)^3\right\} - 3\mathrm{w(k-1)}$. La esperanza matemática puede ser estimada usando un vector x con varias muestras (por ejemplo, 1,000 puntos).
3. Dividir w(k) por su norma.
4. Si $\left|\mathrm{w(k)^T w(k-1)}\right|$ no es cercana a 1, tomar $k = k + 1$ e ir al paso 2. De otra forma, tomar como salida el vector $w(k)$.

3.2.2 Mezclas convolutivas

El problema de la separación ciega de fuentes (*BSS, Blind Source Separation*) consiste en la recuperación de un conjunto de N señales que no pueden ser observadas directamente, llamadas fuentes, a partir de otro conjunto de M señales que sí que pueden observarse. Los algoritmos basados en BSS buscan encontrar un sistema de separación que invierta la mezcla realizada por el sistema de mezcla, devolviendo una estimación de las señales originales, siendo la condición principal requerida para la aplicación de estos algoritmos que las señales fuentes sean mutuamente independientes. Si bien la forma exacta en que se mezclan las fuentes independientes no es conocida, a la hora de abordar un problema de BSS se hace necesario suponer un modelo de mezcla que se ajuste lo más posible a la realidad para poder desarrollar algoritmos matemáticos que lo resuelvan. La suposición más frecuente es considerar que las mezclas son lineales. Por otro lado, los algoritmos BSS convolutivos (*CBSS*) consideran que pueden existir contribuciones retardadas, lo cual supone sustituir los productos escalares por convoluciones,

$$x_i[n] = \sum_{j=1}^{N} h_{ij}[n] * s_j[n], \qquad i = 1,2,\dots,M \tag{3.67}$$

donde los términos h_{ij} representan filtros lineales de longitud finita (*FIR*). Si utilizamos notación matricial tenemos:

$$x[n] = \boldsymbol{A} * s[n] \tag{3.68}$$

En cualquier caso, tanto en el modelo de mezcla lineal instantánea como en el de mezcla convolutiva, la solución del problema BSS consiste en encontrar una matriz $\boldsymbol{W}$ que invierta la mezcla hecha por la matriz $\boldsymbol{A}$ de manera que se obtenga una estimación de las señales independientes originales, a las que nos referiremos por $\hat{s}_i$.

Los métodos más importantes para resolver este problema son los siguientes:

- Estadísticos de orden dos. Estos procedimientos utilizan filtros FIR estrictamente causales o bien métodos de subespacio, siempre que el número de fuentes sea inferior al número de sensores. En general, se trata de reducir el problema a una mezcla instantánea con la estadística de orden dos para, posteriormente, utilizar estadística de orden cuatro y obtener la separación.
- Estadísticos de alto orden. El primer algoritmo fue propuesto por Jutten para dos señales, considerando el canal como un filtro de respuesta impulsional finita (*FIR*), generalizando para este tipo de mezclas lo que había desarrollado para mezclas lineales, es decir, separa las fuentes imponiendo la anulación de momentos cruzados.

El modelo convolutivo considera que el medio de propagación actúa como un filtro con *p* entradas (las fuentes) y *q* salidas (las mezclas detectadas por los sensores). Usualmente se admite que este filtrado, introducido por el medio y los sensores, es lineal y estacionario. Para este modelo la *mezcla espectral* corresponde a la mezcla convolutiva en el dominio de la frecuencia.

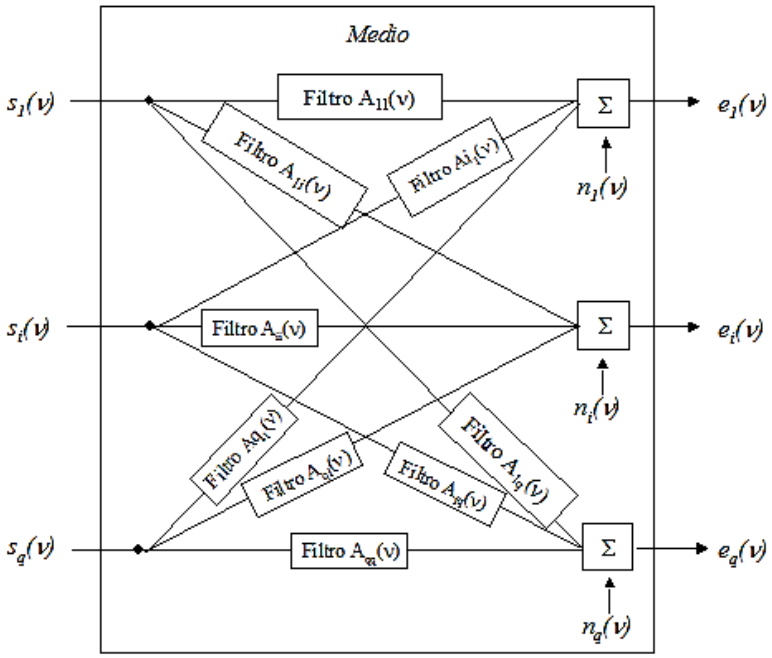


Fig. 3.7 Modelo de mezcla convolutiva

Este modelo se puede descomponer en varios problemas de separación de mezclas instantáneas. El sistema de separación $\boldsymbol{W}$ para cada frecuencia puede ser elegido independientemente utilizando los algoritmos establecidos para mezclas instantáneas. Una vez recuperadas las fuentes de cada frecuencia, se calcula la trasformada inversa de Fourier para obtener las fuentes en el dominio temporal. En la práctica, los sistemas de mezcla y separación suelen modelarse como filtros tipo FIR, ver Fig. 3.7.

3.3 Clasificador difuso

La clasificación es una tarea de aprendizaje no supervisado que propone una descomposición de un conjunto dado de objetos en subgrupos o grupos basados en su semejanza. El objetivo es dividir el conjunto de datos de tal manera que los objetos pertenecientes al mismo grupo sean lo más similares posibles, mientras que los objetos pertenecientes a diferentes grupos sean tan diferentes como sea posible.

El análisis de grupos o *clusters* es una herramienta útil para el descubrimiento de estructuras ocultas en un conjunto de objetos desordenados. Sin embargo, la asignación de los objetos a las clases y las descripciones de estas clases son desconocidas. Mediante la organización de objetos similares en grupos, se intenta reconstruir la estructura desconocida de cada grupo, de forma tal que represente las distintas categorías de los objetos.

Todos los métodos de clasificación que se consideran en este trabajo son algoritmos de partición (agrupación difusa), es decir, dado un entero positivo C, se propone encontrar la mejor partición de los datos en c_i grupos, basada en sus rasgos característicos. Los métodos de clasificación o agrupamiento difuso son diferentes de las técnicas jerárquicas.

Los algoritmos de agrupamiento difuso se basan en funciones objetivo J, las cuales son criterios matemáticos que cuantifican los modelos de agrupamiento, abarcando los prototipos y la partición de datos. Las funciones objetivo tienen que reducirse al mínimo para obtener soluciones óptimas. Una vez definido un criterio de optimización, la tarea de agrupamiento puede ser formulada como un problema de optimización de funciones.

En su forma básica el algoritmo FCM (*Fuzzy C-Means*) busca un número predefinido de C grupos, en un determinado conjunto de datos, donde cada uno de los grupos está representado por su vector centro. Sin embargo, el algoritmo FPCM (*Fuzzy Posibilistic C-Means*) difiere en la forma de asignar los datos a los grupos o clusters, es decir, qué tipo de particiones de los datos se formaran. En el análisis duro (*hard*) de agrupamiento, cada dato es asignado a exactamente un solo grupo, por lo que en este tipo de particiones se pueden generar conjuntos vacíos y resulta

inadecuado para datos que se encuentren casi a la misma distancia entre dos o más clusters. Tales puntos de datos pueden representar un tipo de grupo híbrido o una mezcla de objetos, es decir, que son similares a dos o más tipos. Una partición dura restringe por completo la asignación de esos puntos de datos a solo uno de los clusters, aunque deberían pertenecer a dos o más grupos.

3.3.1 Extracción de rasgos característicos

El conjunto de datos que se proporciona como entrada al clasificador se caracteriza por ser un conjunto fijo con valores predeterminados de rasgos característicos o atributos. El uso de un conjunto fijo de rasgos característicos impone una restricción, ya que en aplicaciones diferentes se tienen características diferentes. Por ejemplo, si las instancias a clasificar fueran vehículos de transporte, entonces el número de ruedas es una característica que se aplica a muchos vehículos, pero no a los barcos, mientras que el número de mástiles podría ser una característica que se aplica a los barcos pero no para vehículos terrestres.

Existen principalmente dos tipos de atributos, los numéricos y los nominales. Atributos numéricos, a veces llamados atributos continuos, están representados por números reales o en valores enteros. Los atributos nominales toman valores de un conjunto predefinido y finito de posibilidades y son a veces llamados categóricos. Las cantidades nominales tienen valores que son símbolos distintos, que sirven sólo como etiquetas o nombres, por ejemplo, el atributo de clima podría tener los valores: soleado, nublado y lluvioso. Existes otros tipos que introducen niveles de medición, como de intervalo y relación.

3.3.2 Algoritmos de clasificación difusa

Los modelos basados en ambigüedades de datos son utilizados en varios campos. A veces, estos modelos se utilizan como alternativas más simples a los modelos probabilísticos. Otras veces son usados para estudiar los datos que, por su naturaleza intrínseca, no puede ser conocidos o cuantificados con exactitud y, por tanto son considerados difusos. Un ejemplo típico de datos difusos es un juicio humano o un término lingüístico. El concepto de número difuso puede ser utilizado eficazmente para describir el concepto de ambigüedad asociado con una evaluación subjetiva. Cada vez que se pide cuantificar nuestras sensaciones o percepciones, parece que la cuantificación tiene un grado de arbitrariedad. Sin embargo, cuando la información es analizada a través de técnicas no-difusas, es decir certeras, se considera como si fuera exacta.

El objetivo de las técnicas difusas (*fuzzy*, en inglés) es incorporar toda la ambigüedad de los datos originales. Por lo tanto, los modelos basados en datos difusos usan más información que los modelos donde la incertidumbre original de los datos es ignorada o arbitrariamente cancelada.

Además, los modelos basados en datos difusos son más generales debido a que un número certero puede ser considerado como un número difuso especial que no tiene un significado difuso asociado [37].

En el proceso de análisis de datos, con el fin de tener en cuenta los problemas de heterogeneidad, puede ser necesario realizar un pre-procesamiento como el centrado, la normalización y la estandarización de los datos [38].

- Centrado, teniendo en cuenta el promedio de los centros.
- Normalización de los centros, dividiendo los centros, por ejemplo c_{ij} entre un factor de normalización $\bar{c}_j$.
- Estandarización de los centros, utilizando por ejemplo $c_{ij}{}^{*} = \bar{c}_{ij} / \left(\frac{1}{l} \sqrt{\sum_{i=1}^{l} \bar{c}_{ij}{}^{2}} \right)$.

La normalización de los centros trata los problemas de heterogeneidad de las unidades de medida y/o del tamaño de las variables.

3.3.2.1 Fuzzy C-Means

La mayoría de las técnicas de agrupamiento difuso se basan en optimizar una función objetivo como *Fuzzy C-Means* (*FCM*) o alguna modificación de ésta. La función objetivo base de una gran familia de algoritmos de agrupamiento difuso es la siguiente:

$$J(Z; U, C) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^{m} \|z_k - c_i\|_B^2 \qquad (3.69)$$

Donde $\mathrm{Z} = \{\mathrm{z}_1, \mathrm{z}_2, \ldots, \mathrm{z}_\mathrm{N}\}$ son los datos que deben ser clasificados, el valor de la función de objetivo es una medida ponderada del error cuadrático que se comete al representar los c clusters por los prototipos c_i. La matriz de partición difusa U esta definida como:

$$U = [\mu_{ik}] \in M_{fc} \qquad (3.70)$$

El vector de centros C, que se busca determinar está compuesto por:

$$C = [c_1, c_2, \ldots, c_c] \qquad (3.71)$$

Este algoritmo está basado en la siguiente norma:

$$D_{ik\,B}^2 = \|z_k - c_i\|_B^2 = (z_k - c_i)^T B (z_k - c_i) \qquad (3.72)$$

El parámetro m es un exponente que determina que tan difusos son los grupos resultantes, y se define como:

$$m \in [1, \infty) \tag{3.73}$$

Minimizar la función objetivo (3.69) es un problema de optimización no lineal que puede ser resuelto de muchas formas, pero la más utilizada es el algoritmo Fuzzy C-Means. Los puntos estacionarios de la función objetivo, se encuentran añadiendo la condición de que la suma de las pertenencias de un punto a todos los clusters debe ser igual a 1, mediante:

$$J(Z; U, C, \lambda) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m D_{ik}^2 + \sum_{k=1}^{N} \lambda_k \left(\sum_{i=1}^{c} \mu_{ik} - 1 \right) \tag{3.74}$$

Igualando a cero las derivadas parciales de J con respecto a U, C y λ, las condiciones necesarias para que (3.69) alcance su mínimo son:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} \left(D_{ik} / D_{jk} \right)^{2/(m-1)}}, \quad 1 \le i \le c \quad 1 \le k \le N \tag{3.75}$$

$$c_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m Z_k}{\sum_{k=1}^{N} (\mu_{ik})^m}, \quad 1 \le i \le c \tag{3.76}$$

La ecuación (3.81) proporciona un valor para c_i como la media ponderada de los datos que pertenecen a un cluster, donde los pesos son las funciones de pertenencia. Existen muchas extensiones y modificaciones al algoritmo básico c-means que se ha descrito. Estos nuevos métodos pueden ser clasificados en tres grandes grupos:

- Algoritmos que utilizan una medida de la distancia adaptativa (una norma diferente para cada cluster). Esto posibilita la detección de clusters de datos con estructuras (tamaños y formas) diferentes.
- Algoritmos basados en prototipos lineales (norma constante y prototipos variables).
- Algoritmos basados en prototipos no lineales.

De entre los algoritmos con distancia adaptativa, cabe destacar el de Gustafson-Kessel (GK), algoritmo que extiende el algoritmo básico fuzzy c-means eligiendo una norma diferente B_i para cada cluster. Este algoritmo, es ampliamente usado dado que los clusters hiperlipsoidales que busca, detectan de forma correcta los comportamientos quasi-lineales que pueden existir en un conjunto de datos.

3.3.2.2 Fuzzy C-Means II

El algoritmo difuso C-Medias (Fuzzy C-Means, FCM) tipo-2 es una extensión del algoritmo convencional Fuzzy C-Means, en este los valores de membresía para cada patrón son extendidos al tipo 2 por la asignación de *dos* grados de membresía tipo-1.

De este modo, los centros de cada grupo que son estimados por los grados de membresía tipo-2 pueden converger a un lugar más deseable que los centros obtenidos con el método FCM tipo-1, cuando se tiene presencia de ruido. En el método convencional FCM tipo-1, los centros de grupos son obtenidos minimizando la siguiente función objetivo de error al cuadrado:

$$J_m(U,V:X) = \sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ik})^m \, \|x_k - v_i\| \qquad (3.77)$$

En FCM-2 se maneja una incertidumbre asociada al parámetro difuso "m", el cual controla el grado de pertenencia de la C-partición final en el algoritmo Fuzzy C-Means tipo-1 (FCM). Para diseñar y manejar la incertidumbre del parámetro m, se extiende el conjunto de patrones a un conjunto difuso tipo 2 de intervalos, usando dos parámetros, $m1$ y $m2$.

Las agrupaciones para un conjunto de datos son obtenidas por la distancia medida entre el dato a evaluar y el centro del grupo. Las agrupaciones determinadas para cada cluster denotan el grado en que cada dato de entrada pertenece a un grupo. Además, cuando la ubicación del centro de un grupo es actualizado, el grado de pertenencia indica la cantidad con la que contribuye cada centro [6]. Por otro lado, cuando un patrón de entrada es considerado como un punto de ruido (localizado lejos de todos los grupos), los datos se asigna a todos los grupos con un grado de pertenencia de l/n, donde n indica el número de grupos, por lo tanto, puede ocurrir una agrupación indeseable. En el método FCM tipo-2, si el valor de la agrupación para un patrón es más grande, se considera que tienen menos incertidumbre frente a una agrupación más pequeña [2].

El uso de agrupaciones difusas tipo 2 mejora el método convencional FCM ya que las agrupaciones pueden ser mejor representadas, tomando en cuenta la densidad de cada cluster. Los centros de grupos estimados tienden a converger a un lugar más deseable que en los grupos obtenidos por el método de tipo-1, incluso en la presencia de ruido.

3.3.2.3 Gustafson Kessel

El algoritmo Gustafson-Kessel sustituye a la distancia euclídea por la distancia de Mahalanobis, a fin de adaptarse a diversas formas y tamaños de los grupos [39]. Para un grupo i, la distancia Mahalanobis se define como:

$$d^2(\boldsymbol{x}_j, \boldsymbol{C}_i) = (\boldsymbol{x}_j - \boldsymbol{c}_i)^T \sum_i^{-1} (\boldsymbol{x}_j - \boldsymbol{c}_i) \qquad (3.78)$$

Donde Σ_i es la matriz de covarianza de cada grupo. Usando la distancia Euclidiana, como en el algoritmo FCM, se puede asumir que $\forall i, \Sigma_i = \boldsymbol{I}$, es decir, todos los grupos tienen la misma covarianza que equivale a la matriz identidad. De esta manera solo es posible detectar grupos esféricos, no se pueden identificar grupos que tienen diferentes formas o tamaños. El algoritmo Gustafson-Kessel modela cada grupo $\boldsymbol{\Gamma}_i$ por su centro $\boldsymbol{c}_i$ y por su matriz de covarianza $\Sigma_i, i = 1, \ldots, c$. Así los centros prototipo de cada grupo están representados por $\boldsymbol{C}_i = (\boldsymbol{c}_i, \Sigma_i)$, donde $\boldsymbol{c}_i$ y Σ_i deben ser calculados.

La estructura de la matriz Σ_i definida positiva $p \times p$ representa la forma del grupo *i*. El tamaño de los grupos, si es un dato conocido, puede ser controlado usando la contante $\varphi > 0$, lo que implica que el $\det(\Sigma_i) = \varphi_i$, usualmente se asume que los grupos tienen un tamaño unitario, es decir $\det(\Sigma_i) = 1$.

La función objetivo (*utilizando la distancia* Mahalanobis *ecu (*3.78)), las ecuaciones para encontrar los centros y las ecuaciones para calcular los grados de pertenencia son las mismas que las utilizadas en FCM, y se definen respectivamente de la siguiente manera:

$$J_f(X, U_f, C) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 \qquad (3.79)$$

$$\boldsymbol{c}_i = \frac{\sum_{j=1}^{n} u_{ij}^m \boldsymbol{x}_j}{\sum_{j=1}^{n} u_{ij}^m} \qquad (3.80)$$

$$\boldsymbol{u}_{ij} = \frac{1}{\sum_{l=1}^{c} \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{\frac{1}{m-1}}} = \frac{d_{ij}^{-\frac{2}{m-1}}}{\sum_{l=1}^{c} d_{lj}^{-\frac{2}{m-1}}} \qquad (3.81)$$

Para calcular las matrices de covarianza, se utilizan las siguientes ecuaciones:

$$\Sigma_i = \frac{\Sigma_i^*}{\sqrt[p]{det(\Sigma_i^*)}} \qquad \textbf{\textit{donde}} \quad \Sigma_i^* = \frac{\sum_{j=1}^{n} u_{ij}(\boldsymbol{x}_j - \boldsymbol{c}_i)(\boldsymbol{x}_j - \boldsymbol{c}_i)^T}{\sum_{j=1}^{n} u_{ij}} \tag{3.82}$$

El algoritmo Gustafson-Kessel extrae más información de los datos, que los algoritmos basados en la distancia euclidiana. Sin embargo es más sensible a la inicialización, por lo tanto se recomienda inicializarlo realizando algunas iteraciones con FCM o PCM dependiendo del tipo de partición considerada. En comparación con el FCM o PCM, el algoritmo Gustafson-Kessel presenta mayor demanda computacional debido al cálculo de las matrices inversas.

Un aspecto importante relacionado con la agrupación es la valoración de la calidad de los grupos obtenidos: el agrupamiento (*clustering*) es una tarea de aprendizaje sin supervisión, lo que significa que existen datos que no son asociados con los grupos que indican la salida deseada, por lo tanto, no se proporciona una referencia con la cual los resultados obtenidos puedan ser comparados. Algunos criterios están específicamente dedicados al agrupamiento difuso, por ejemplo el criterio de entropía de la partición, calcula la entropía de los grados de pertenencia obtenidos.

$$PE = -\sum_{i,j} u_{ij} \log u_{ij} \tag{3.83}$$

Este criterio se puede utilizar para evaluar cuantitativamente la calidad de la agrupación y para comparar diferentes algoritmos. También se puede aplicar para comparar los resultados obtenidos con un único algoritmo, cuando los parámetros cambian. En particular, se puede utilizar para seleccionar el número óptimo de grupos o clúster, aplicando el algoritmo para varios valores de c, se obtiene el valor de c_i óptimo [40].

El algoritmo GK es bastante adecuado para el propósito de clasificación, ya que tiene las siguientes propiedades:

- La dimensión de los clusters viene limitada por la medida de la distancia y por la definición del prototipo o centro de los clusters como un punto.
- En comparación con otros algoritmos, GK es relativamente insensible a la inicialización de la matriz de partición.
- Como el algoritmo está basado en una norma adaptativa, no es sensible al escalado de los datos, con lo que se hace innecesaria la normalización previa de los mismos.
- El algoritmo GK puede detectar clusters de diferentes formas (clusters hiperelipsoidales), no solo subespacios lineales.

Sin embargo, presenta las siguientes desventajas:

- La carga computacional es bastante elevada, sobre todo en el caso de grandes cantidades de datos.
- Cuando el número de datos disponibles es pequeño, o cuando los datos son linealmente dependientes, pueden aparecer problemas numéricos ya que la matriz de covarianzas se hace casi singular. El algoritmo GK no podrá ser aplicado a problemas puramente lineales, en el caso ideal de no existir ruido.
- Si no hay información al respecto, los volúmenes de los clusters se inicializan con valores iguales. De esta forma, no se podrán detectar clusters con grandes diferencias en tamaño.

CAPÍTULO 4. SISTEMA DE SEPARACIÓN CIEGA Y CLASIFICACIÓN DIFUSA

4.1 Introducción

El sistema de separación ciega y clasificación de señales de audio ambientales tiene como objetivo separar señales de audio mezcladas de forma natural en el ambiente, como se mencionó anteriormente, esta separación se conoce como "*ciega*", ya que no se tiene conocimiento previo de las señales que se encuentran presentes en la mezcla, el número (*cantidad*) de las distintas señales fuentes y la ponderación de cada una de las fuentes. Una vez separadas las señales fuente, el sistema es capaz de identificar el tipo de señal del que se trata.

Este sistema está constituido por varias etapas que se listan a continuación, tomando en cuenta que se inicia con al menos con 2 mezclas de señales. Esto debido a la condición que presenta ICA, donde el número de mezclas de entrada debe ser mayor o igual al número de señales fuentes que se deseen separar. El sistema está compuesto por las siguientes etapas:

1. Pre-procesamiento
2. Separación ciega mediante ICA
3. Extracción de rasgos característicos
4. Clasificación difusa

En la Fig. 4.1 se muestra el diagrama de bloques del sistema propuesto, se puede observar que el primer paso es aplicar una etapa de pre-procesamiento donde las mezclas son filtradas, y se obtienen su descomposición wavelet, a cada sub-banda obtenida al aplicar la transformada wavelet se aplica la técnica de ICA y se reconstruyen las señales separadas en el dominio del tiempo. Hasta aquí se tiene el primer objetivo de separar las señales fuentes presentes en las mezclas iniciales.

El siguiente paso es obtener rasgos característicos que representan el grupo o clase al que pertenece dicha fuente, estos rasgos son la entrada al clasificador difuso. El clasificador previamente entrenado, evalúa los rasgos para identificar el tipo de señal fuente. De esta manera se logra separar *n-fuentes* presentes en *m*-mezclas donde $m > n$ e identificar (con previo conocimiento de las posibles señales fuentes presentes en las mezclas) el tipo de señales fuentes que predominan en la muestra.

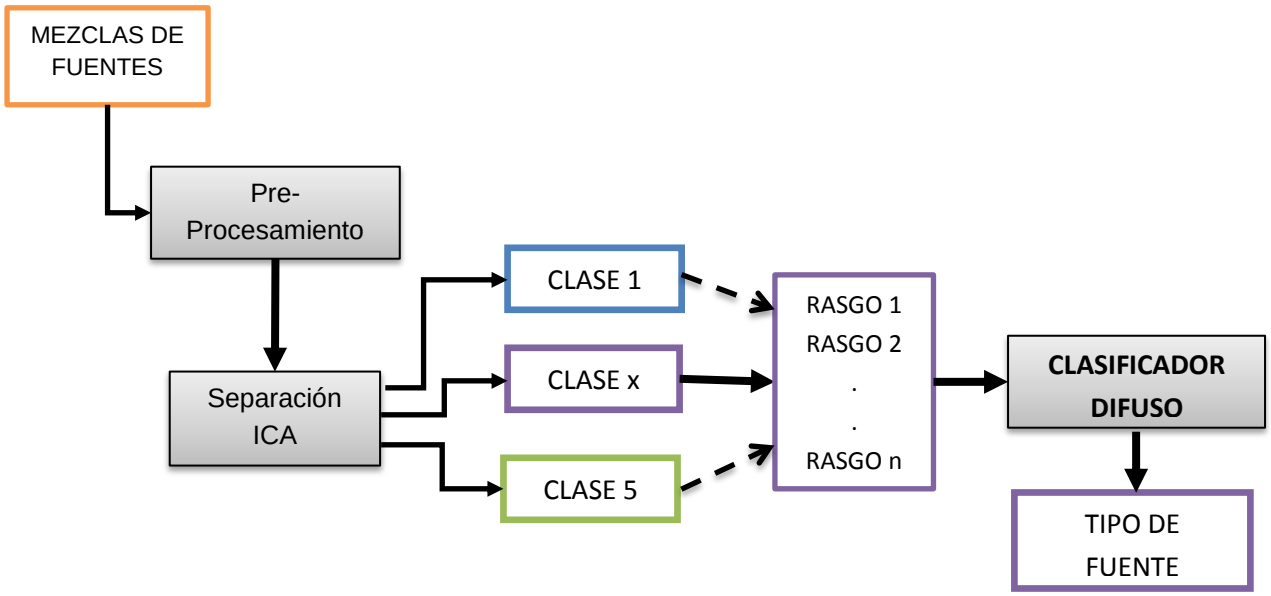


Fig. 4.1. Diagrama de bloques del sistema de separación y clasificación

Debido a que el número de señales fuentes que se pueden separar está sujeto al número de mezclas, en ambientes reales es muy difícil encontrar una separación fiel de las señales mediante ICA. Lo que se obtiene son señales donde predomina una señal fuente, para el caso del monitoreo de ruido ambiental estos resultados ayudan a identificar las fuentes que generan la mayor cantidad de ruido. Una ventaja que presenta el clasificador difuso es que además de identificar el tipo de fuente que predomina en la mezcla, se puede obtener una proporción de otras fuentes que estén presentes en la mezcla.

A continuación se detalla el funcionamiento de cada una de las etapas del sistema, la descripción de cada etapa se explica en un contexto general, es decir, sin tomar en cuenta restricciones del número de mezclas de entrada, señales fuente presentes en dichas mezclas, posibles tipos de fuentes, y rasgos característicos específicos para identificar ciertos tipos de señales de audio.

4.2 Pre-procesamiento

La etapa de pre-procesamiento consta de dos fases, la primera es aplicar un filtro blanqueador que eliminar frecuencias menores a 80 Hz. La segunda fase es descomponer cada una de las mezclas utilizando la transformada wavelet. La transformada wavelet pertenece a una serie de técnicas de análisis de señal denominadas comúnmente análisis *multiresolución*.

Su principal característica es que permite conocer qué frecuencias componen una señal en cada instante con las siguientes resoluciones:

- Para las altas frecuencias consigue una buena resolución en el tiempo que permita su exacta localización temporal, aún a cambio de perder resolución en frecuencia.
- Para las componentes de bajas frecuencias lo más relevante es conocer su frecuencia aún a costa de perder resolución en el dominio del tiempo.

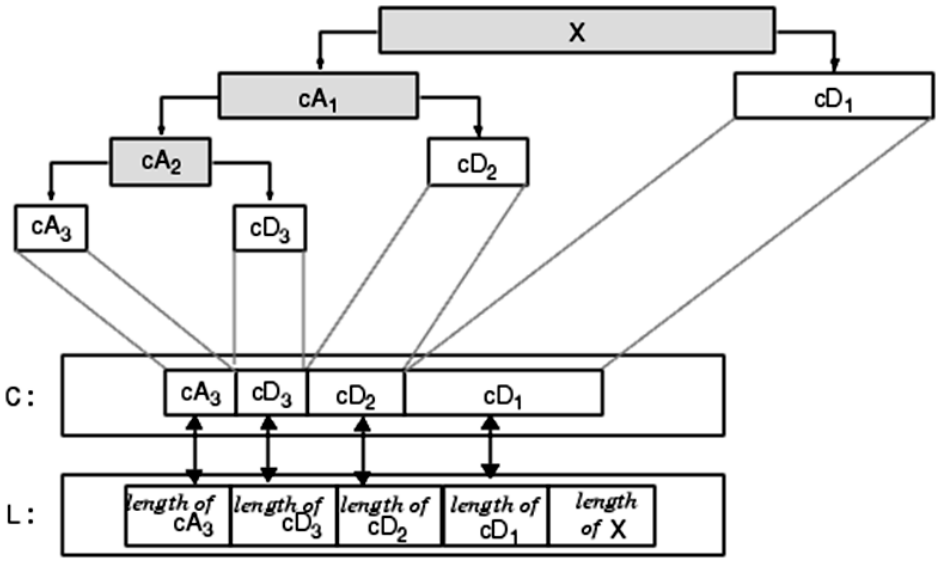


Fig. 4.2 Descomposición de una señal en bandas wavelet

Por definición, los algoritmos de separación ciega de fuentes, como ICA, sólo son capaces de separar señales estadísticamente independientes. El problema que se presenta es que la mayor parte de señales que nos encontraremos en el mundo real no son totalmente independientes.

Cuando se intenta conseguir la estimación a partir de mezclas de señales tomadas en el mundo real, los resultados que se obtienen no son tan buenos como se podría esperar. Pero este contratiempo se puede solventar en la mayoría de los casos aplicando un pre-procesamiento a las mezclas de entrada sobre las que se aplica el algoritmo BSS.

Este análisis previo consistiría básicamente en dividir las señales de entrada en una serie de sub-señales, siendo ahora estas sobre las que se aplicará el algoritmo ICA. Para conseguir la división indicada se seguirá un procesamiento como el mostrado en la Fig. 4.3.

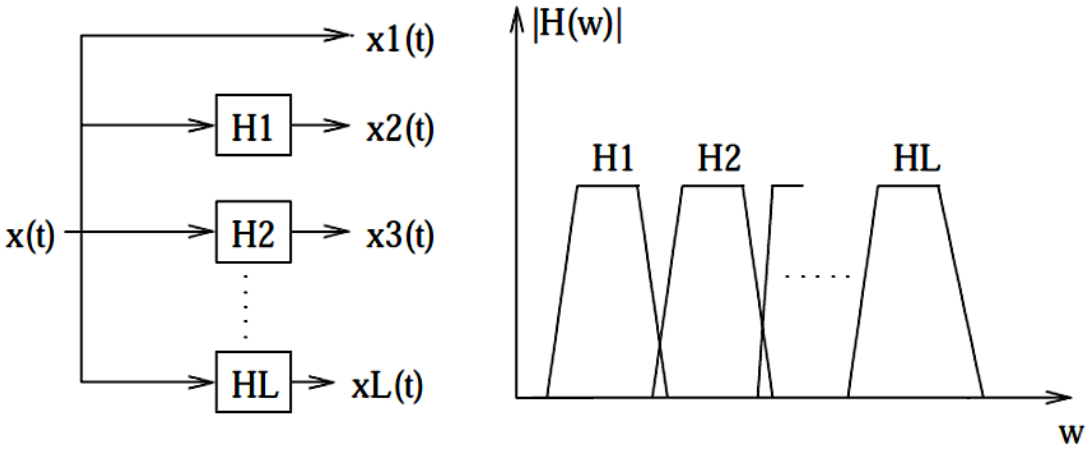


Fig. 4.3 Filtrado en sub-bandas de la mezcla x(t)

Una alternativa para llevar a cabo este procedimiento es aplicar la transformada Wavelet. Esta transformada consigue una división tanto en el dominio del tiempo como en el dominio de la frecuencia, esta división es adecuada para el análisis de señales de audio puesto que nos permite conseguir una mayor resolución espectral en las frecuencias donde se encuentra concentrada la mayor parte de los sonidos producidos en el centro histórico de la Ciudad de México.

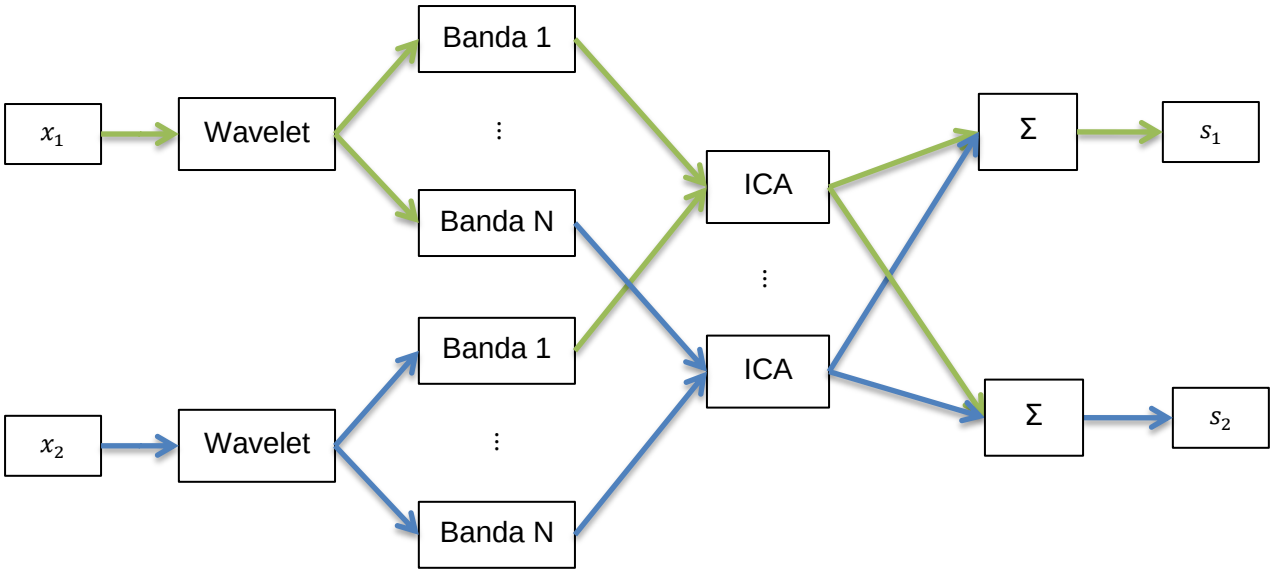


Fig. 4.4 Separación en sub-bandas utilizando wavelets

El procedimiento consiste en dividir cada señal de entrada en varias sub-señales, mediante la transformada Wavelet, y aplicar el algoritmo de separación a las bandas de igual frecuencia de

cada señal, por lo tanto, el algoritmo se aplicará tantas veces como bandas obtengamos de una señal. Posteriormente se compondrán las dos señales de salida simplemente a partir de la suma de las bandas estimadas, este método se muestra en la Fig. 4.4.

Como se mencionó anteriormente, la transformada wavelet permite aumentar la resolución de las frecuencias bajas en el espectro y mantener la buena resolución de las frecuencias altas en el tiempo, sin embargo no existe un criterio definido para evaluar la calidad de las diferentes familias de wavelets debido a que dependen de la aplicación y características requeridas como simetría, momentos de desvanecimiento, regularidad, etc.

Utilizando este método wavelet-ICA se comprobó que la trasformada wavelet con filtro Daubechies presenta mejores resultados que otras familias de wavelets, además se realizaron varios experimentos dividendo en bandas de 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40 y 50, (ver, Fig. 4.5); obteniendo un mejor desempeño cuando las señales de entrada se descomponen en 4 sub-bandas para la mayoría de las señales [31].

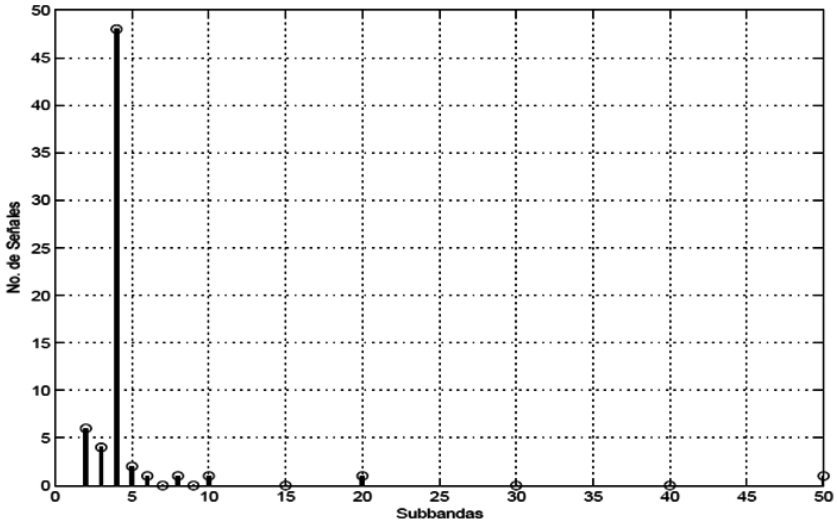


Fig. 4.5 Histograma número de señales con mejores resultados por sub-banda

La característica de suavizado de la wavelet con filtro Daubechies de orden 4 (Fig. 4.6), la hace más apropiada para detectar los cambios en la señal. Es efectiva para la eliminación de ruido ya que producen curvas más suaves a diferencia de otras wavelets.

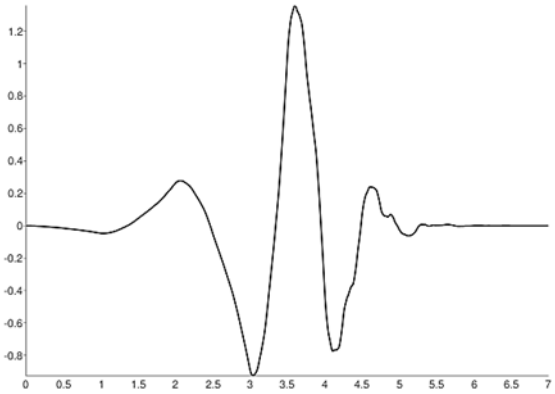


Fig. 4.6 Función wavelet con filtro Daubechies de orden 4

Es importante considerar que a la salida del algoritmo de separación ICA, deben sumarse las bandas correspondientes a cada señal fuente, por lo tanto deben ordenarse, ya que las señales estimadas pueden no estar en el mismo orden que las señales de entrada. Una forma de ordenar las señales es utilizando la máxima correlación entre la estimación y la señal original.

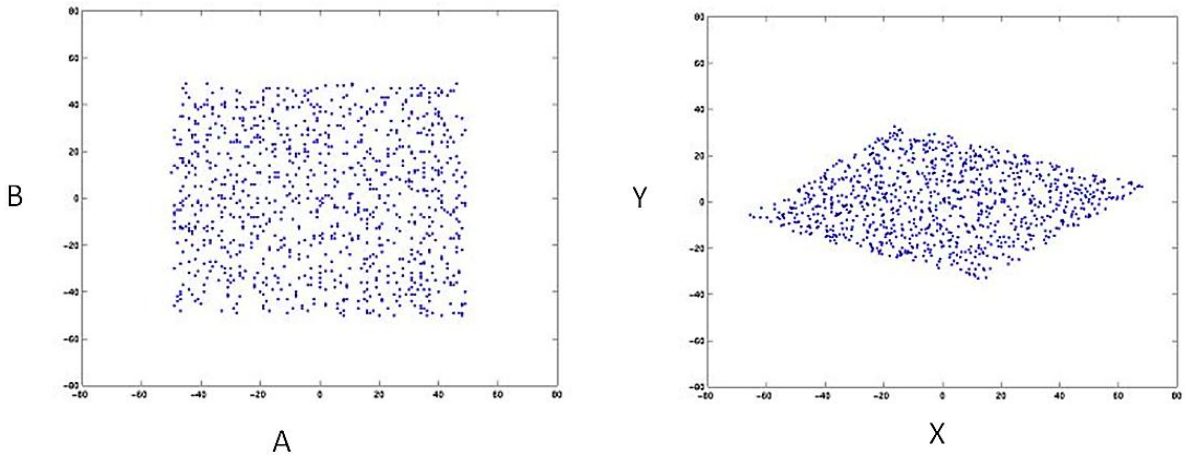
4.3 Metodología para separación ciega de fuentes

El análisis de componentes independientes, en adelante ICA, fue presentado en 1986 por Jeanny Herault y Christian Jutten en Utah como una red neuronal basada en la ley de aprendizaje de Hebb capaz de realizar una separación ciega de señales. En concreto, este algoritmo trata de separar un número determinado de señales estadísticamente independientes a partir un número idéntico de señales de entrada que son suma lineal de las primeras.

La primera aplicación inmediata de ICA, es la eliminación de artefactos. Se trata de separar estas últimas señales no deseadas pudiendo realizar la clasificación únicamente sobre las señales originales resultado de la actividad neuronal. La restricción en el uso de esta técnica para un caso general, son:

- Las fuentes, es decir, las señales originales que se mezclan y que ICA deberá recuperar posteriormente, deben ser linealmente independientes.
- El retardo de propagación a través del medio en el que se mezclan las señales tiene que ser despreciable.
- Las señales originales deben ser analógicas y su función de distribución de probabilidad no puede ser gaussiana.
- El número de componentes independientes es el mismo que el de señales originales.

La separación ciega de señales, se realiza sobre dos o más mezclas de las señales fuentes, donde el número de mezclas debe ser mayor o igual al número de señales fuente que se deseen separar. En la Fig. 4.7, se observa la distribución de datos de dos señales independientes A y B. Estas señales son mezcladas linealmente para obtener las mezclas Y y X.

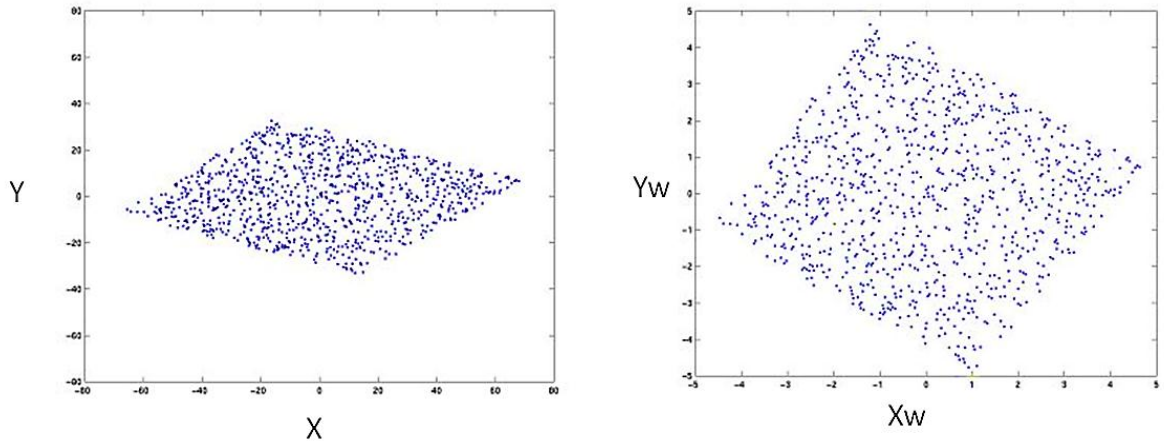


(a) **(b)**

Fig. 4.7 Distribución de datos, a) señales independientes y b) señales mezcladas

Paso1. Antes de aplicar el método de ICA, se deben blanquear los datos. Lo que significa remover cualquier correlación entre los datos, estos es cumplir con la ecuación (4.1). El proceso de blanqueado es un cambio lineal de coordenadas de los datos mezclados, ver Fig. 4.8.

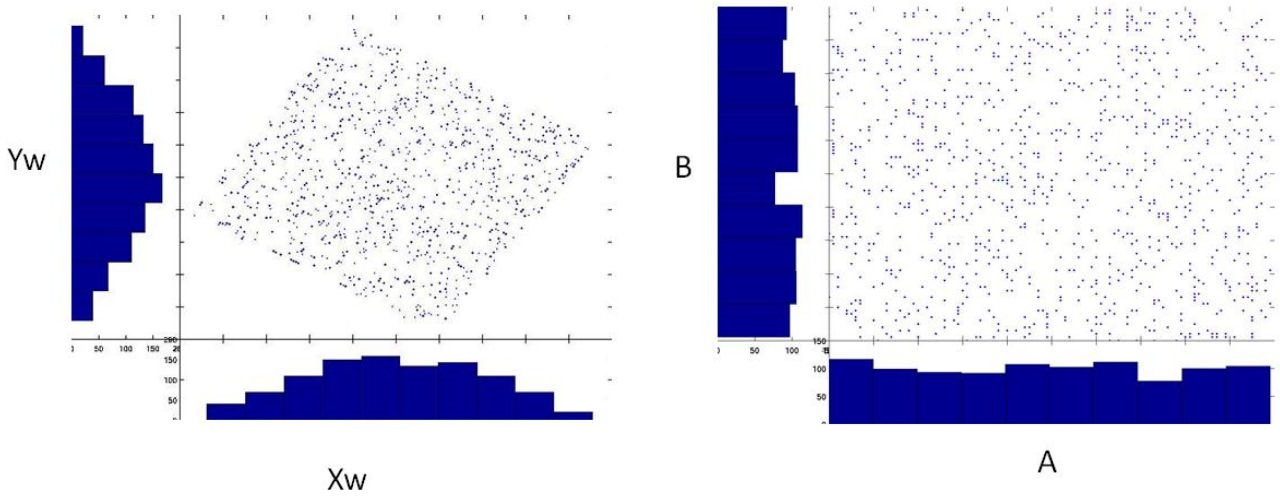
$$C = E\{X_w {Y_w}^T\} = E\{X_w\}E\{{Y_w}^T\} = 0 \qquad (4.1)$$



(a) **(b)**

Fig. 4.8 Distribución de datos, a) señales mezcladas y b) señales blanqueadas

Paso 2. El método de ICA rota la matriz de blanqueado a sus fuentes originales (A,B). Esta rotación se realiza minimizando la *gaussianidad* de los datos proyectados en ambos ejes, Fig. 4.9.



(a) **(b)**

Fig. 4.9 Histograma de la distribución de datos, a) señales blanqueadas y b) señales separadas por ICA

Características principales de ICA

- ICA permite obtener fuentes originales que sean estadísticamente independientes.
- Debido a que ICA separa las fuentes, maximizando la no-gaussianidad, las fuentes gaussianas no pueden ser separadas.
- Aun cuando las fuentes no son totalmente independientes, ICA puede encontrar el espacio donde son mayormente independientes.
- El método de ICA no puede obtener la amplitud original de las fuentes mezcladas.

Indeterminaciones de ICA

- El método de ICA no puede obtener la amplitud original de las fuentes mezcladas, esta puede ser atenuada o amplificada.
- Se puede permutar el orden de las fuentes originales, sin variar el orden observaciones e(t).

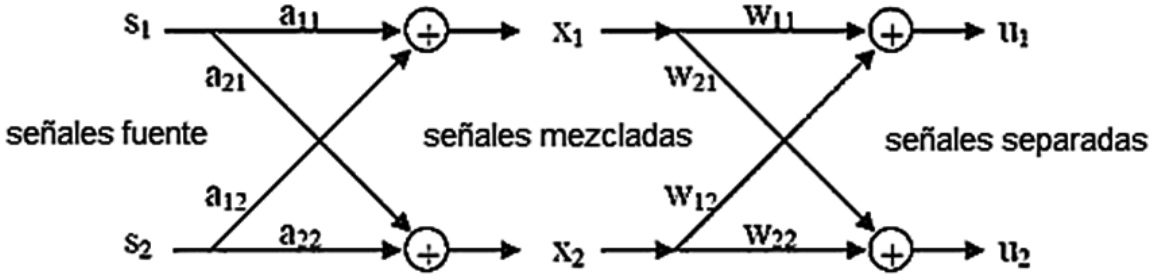


Fig. 4.10 Diagrama de bloques del modelo lineal

Modelo convolutivo

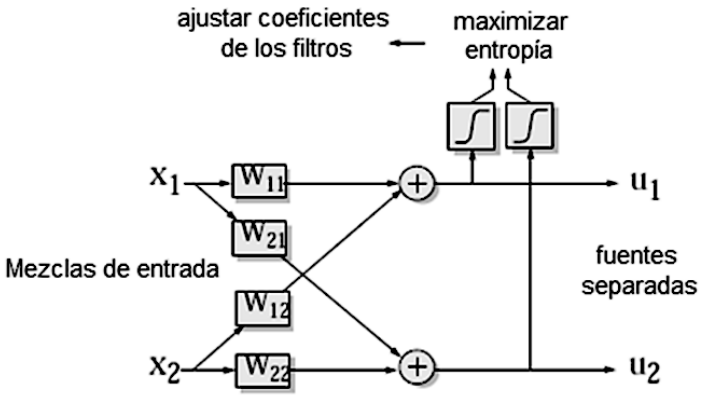


Fig. 4.11 Diagrama de bloques del modelo convolutivo

Ecuaciones Modelo convolutivo:

$$\boldsymbol{u}(t) = (\boldsymbol{I} + \boldsymbol{W}_0)^{-1}(\boldsymbol{x}(t) - \sum_{k=1}^{M-1} \boldsymbol{W}_k \boldsymbol{u}(t-k)) \tag{4.2}$$

$$\Delta \boldsymbol{W}_0 \propto -(\boldsymbol{I} + \boldsymbol{W}_0)(\boldsymbol{I} + \hat{\boldsymbol{y}}\boldsymbol{u}^T) \tag{4.3}$$

$$\Delta \boldsymbol{W}_k \propto -(\boldsymbol{I} + \boldsymbol{W}_k)\hat{\boldsymbol{y}}u_{t-k}^T \tag{4.4}$$

4.4 Sistema de clasificación

Para obtener los rasgos característicos se generaron las siguientes clases, con base en las posibles fuentes a identificar en el Centro de la Cuidad de México.

Clase 1	Clase 2	Clase 3	Clase 4	Clase 5
Silbato para control de tráfico	Vehículos: Carros Motos Camiones	Voz Voz Alta Aplausos Risas	Sirenas: Ambulancias Patrullas Bomberos	Claxon de distintos vehículos

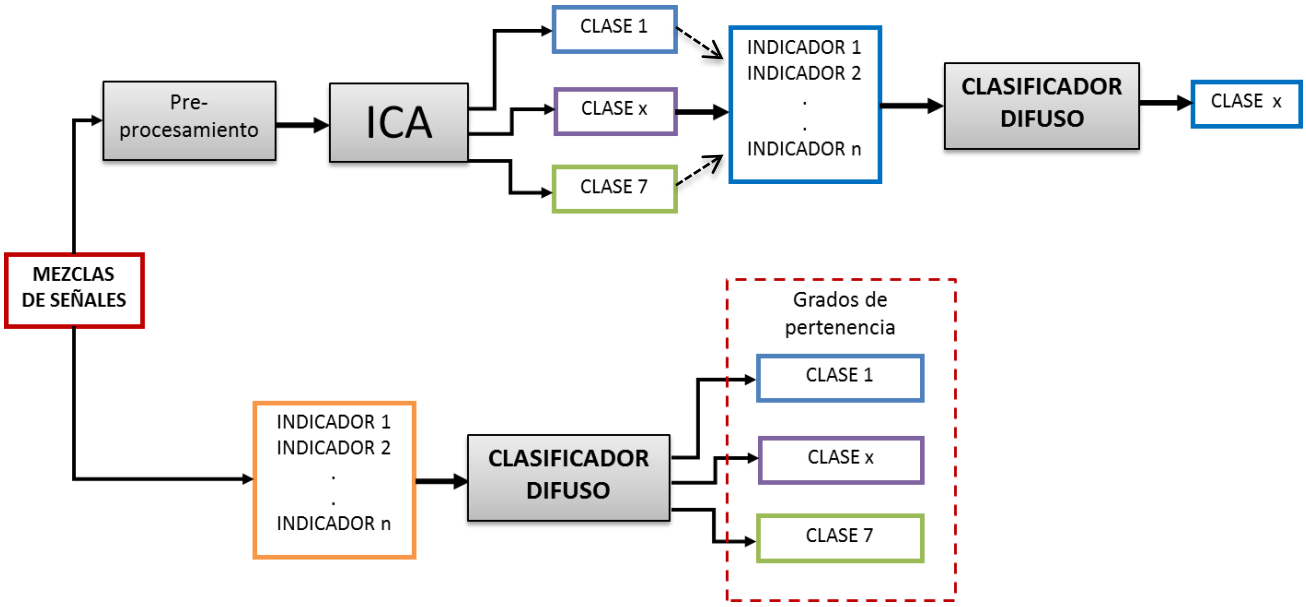


Fig. 4.12 Diagrama de bloques del clasificador difuso

Extracción de Rasgos Característicos. Para lograr diferenciar estas clases, se eligieron los siguientes descriptores de sonido para caracterizar las clases obtenidas del algoritmo de ICA y clasificarlas mediante técnicas de agrupamiento difuso. Se tomaron en cuenta los siguientes índices que establecen las normas para evaluar los sonidos:

CTE: Componentes Tonales emergentes
CI: Componentes Impulsivas
CBF: Componentes de Bajas Frecuencias
NSCE: Nivel Sonoro Continuo Equivalente
VP: Valores Pico

Por otro lado se consideraron los siguientes índices basados en los siguientes cálculos para representar las señales fuente:

E: Energía

$$E = \sum_{n=0}^{N} x[n]^2 \qquad (4.5)$$

Z: Cruces por cero

$$SC = \frac{\sum f_i a_i}{\sum a_i} \qquad (4.6)$$

SC: Centroide espectral

$$ZCR = \frac{1}{2}\sum_{n=1}^{N}|sign(x[n]) - sign(x[n-1])| \qquad (4.7)$$

Se propone aplicar un pre-procesamiento a la mezcla de entrada para lograr una descomposición utilizando wavelets, se hace pasar por el método de ICA para obtener una separación en clases. A cada clase obtenida se extraen los rasgos característicos, estos rasgos son los datos de entrada al clasificador difuso, el cual nos dará el grado de pertenencia a cada una de las clases definidas con las cuales fue entrenado. Por otro lado, se obtienen los rasgos a la mezcla inicial para utilizarlos en el clasificador y obtener de qué clases se encuentra compuesta la mezcla.

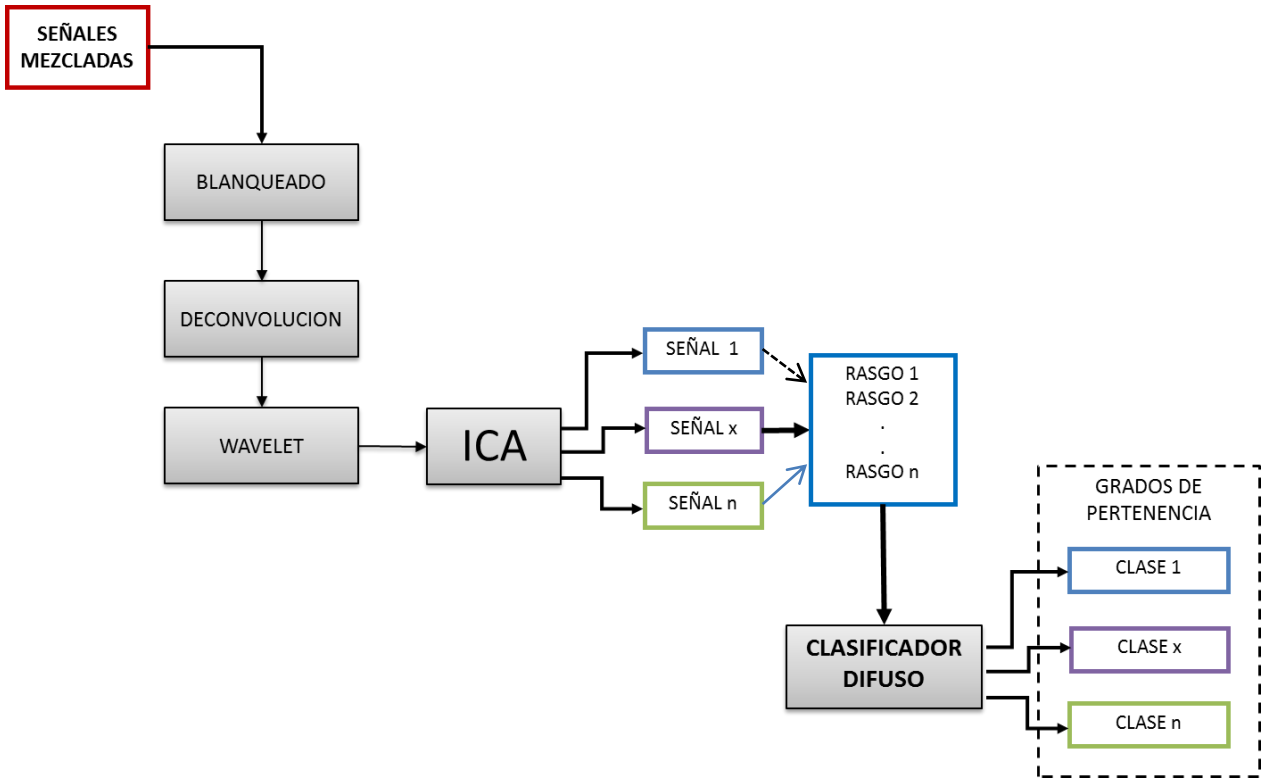


Fig. 4.13 Sistema de separación y clasificación para mezclas reales

CAPÍTULO 5. SEPARACIÓN Y CLASIFICACIÓN DE SEÑALES ACÚSTICAS

5.1 Experimentos y resultados sobre mezclas lineales

Para la separación de mezclas lineales o instantáneas se utilizaron señales de audio en formato *wav* (*WAVeform audio file format)*, estas señales representan cinco de los sonidos más comunes y que contribuyen en mayor medida a la contaminación acústica presente en zonas urbanas.

El formato *WAV (o WAVE)* es un formato de audio digital, normalmente sin compresión de datos que se utiliza para almacenar sonidos, admite archivos mono y estéreo a diversas resoluciones y velocidades de muestreo, utiliza principalmente el formato PCM (no comprimido) y, al no tener pérdida de calidad, es adecuado para uso profesional. Para tener calidad CD de audio se necesita que el sonido se grabe a 44100 Hz y a 16 bits [41].

Fig. 5.1 Señales acústicas ambientales en zonas urbanas

Las señales a separar corresponden a sonidos típicos que predominan en el ruido ambiental generado en las zonas urbanas y que se encuentran relacionadas a las actividades económicas, de transporte y recreativas, Fig. 5.1**.** Las señales que se consideran en este trabajo son las siguientes:

1. Sonidos generados por el ser humano, debido a las actividades económicas presentes en la zona (Centro Histórico de la Ciudad de México), así como actividades recreativas. Dentro de esta categoría se considera la voz producida por multitudes (murmullos), aplausos, risas, gritos (voz alta).
2. Sonidos producidos por los motores de automóviles, camiones, motos, etc. En la extracción de rasgos y entrenamiento del clasificador se considera el sonido constante, es decir en las señales analizadas no se encuentra presente el efecto *Doppler*, tampoco se considera el sonido de los autos al acelerar o frenar.
3. Sonidos producidos por sirenas en general, es decir, se incluyen los sonidos producidos por patrullas, ambulancias y camiones de bomberos. Las normas internacionales establecen que el valor de las sirenas en dB tiene que estar 15 dB por encima del nivel del ruido ambiental y en cualquier caso deben ser superior a los 65 dB, mientras la frecuencia central de la señal debe estar entre los 300 y 3000 Hz y diferenciarse en lo posible de la frecuencia central en la cual el ruido ambiental es más fuerte.
4. Sonidos generados por el claxon de los vehículos, ya que este sonido es molesto y puede contribuir en gran medida al ruido ambiental de la zona. Sobre todo en las zonas más transitadas donde existen conflictos viales.
5. Otra fuente de ruido a considerar es el producido por los silbatos que utilizan los policías para controlar el tránsito. Este ruido puede resultar molesto para las personas que se encuentran cerca de ellos y por tiempos prolongados, de ahí el interés de considerar este sonido para evaluar su contribución al ruido ambiental.

Estos sonidos no son los únicos que contribuyen al ruido ambiental en las zonas urbanas de mayor concentración de población, sin embargo se considera que son los que más prevalecen y pueden ser molestos para el confort de los habitantes.

Tabla 5.1 Señales de ruido ambiental consideradas para separación y clasificación

Señales de ruido ambiental	Descripción
Sirena	Ambulancias, patrullas, camiones de bomberos.
Claxon	Carros, camiones, motos, entre otros.
Silbato	Control de tránsito
Voz	Generada por multitudes, risas, voz alta, aplausos
Vehículos	Motor de carros, camiones, motos, entre otros.

Otros sonidos presentes, son la música en volumen alto (por ejemplo, utilizado por comercios para publicidad), las fábricas, los aviones (sobre todo en áreas cercanas a los aeropuertos), la maquinaria de construcción (utilizada en mantenimiento o construcción de edificios, puentes, etc.),

así como también actividades recreativas temporales como: conciertos, ferias, exposiciones, entre otros, sin embargo estas fuentes, no son consideradas en esta investigación, ya que son ocasionales y/o su aporte al ruido ambiental está limitado a zonas muy específicas. Otro motivo importante para delimitar las categorías de sonidos y clasificarlas, es para ser congruentes con las restricción que establece la teoría de ICA, ya que en las mediciones realizadas se utilizaron hasta 4 sensores (micrófonos, *ver Anexo B*), por lo que se pueden separar como máximo 4 fuentes generadoras, sin considerar en el modelo que considera ruido adicional.

Las señales que se utilizaron en los siguientes experimentos tienen frecuencias de muestreo (*fs*) de 8000, 11025, 16000, 22050, 32000, 44100 y 48000 Hz; el número de bits de 8 y 16 bits. Algunas de las señales descargadas cuentan con 2 canales (*estéreo*). En este trabajo se utilizó un solo canal (*monocanal*) en todas las señales. La Tabla 5.2 describe brevemente los sonidos contenidos y el número de muestras de cada tipo de fuente con las que cuenta la base de datos utilizada en los siguientes experimentos. En el caso del tipo de fuente vehículos y silbatos se tiene menor cantidad de muestras ya que solo son considerados los sonidos constantes, por lo que no existen muchas variaciones entre muestras, a diferencia de los sonidos de sirenas, voz y claxon.

Tabla 5.2 Base de datos de las fuentes de ruido consideradas

Tipo de Fuente	Descripción	No. de muestras
Vehículos	*Sonido del motor de distintos vehículos como carros, camiones, motos, etc.*	76
Silbato	*Sonido de silbato para control de tránsito.*	95
Sirena	*Sonidos de sirenas de patrullas, ambulancias y bomberos.*	139
Claxon	*Claxon de vehículos terrestres comunes en zonas urbanas.*	144
Voz	*Murmullos de personas hablando al mismo tiempo, risas, voz de mujer, voz de hombre, aplausos.*	178

Como se mencionó anteriormente, una etapa previa incluida en gran parte de los algoritmos de separación ciega es efectuar un blanqueado de las señales captadas, con el fin de decorrelacionarlas y facilitar notablemente el proceso de separación, como se muestra a continuación. El blanqueado consiste en someter a los vectores ***e(t)*** a una transformación lineal **V** de forma que se obtengan unos nuevos vectores ***x(t)*** dados por:

$$\boldsymbol{x}(t) = \boldsymbol{V} \cdot \boldsymbol{e}(t) \qquad (5.1)$$

de forma que:

$$E\{\boldsymbol{x}(t)\boldsymbol{x}^{\boldsymbol{T}}(t)\} = \boldsymbol{I} \qquad (5.2)$$

La matriz **V** puede ser estimada a partir de una muestra de los vectores ***e(t)*** calculando de alguna forma su matriz de covarianza y después normalizando, utilizando los valores propios. Antes de nada se hace que los vectores de entrada ***e(t)*** tengan media cero. Esto se logra sin más que restarles su media, es decir:

$$\boldsymbol{e}(t) \leftarrow \boldsymbol{e}(t) - E\{\boldsymbol{e}(t)\} \tag{5.3}$$

De esta forma los datos se normalizan con respecto a la estadística de primer orden. Las componentes de los vectores blanqueados ***x(t)*** resultan estar decorrelacionadas y normalizadas de forma que su varianza sea la unidad.

Experimento 1. Se realizaron dos mezclas lineales utilizando dos sonidos como señales fuente. La mezcla contiene el sonido de una sirena y el de un silbato. La figura muestra las señales originales y las dos mezclas en el dominio del tiempo, obtenidas linealmente mediante la matriz de mezcla:

$$A = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \tag{5.4}$$

En la Fig. 5.2 se muestran las señales originales en el dominio del tiempo, y las dos mezclas obtenidas a partir de estas.

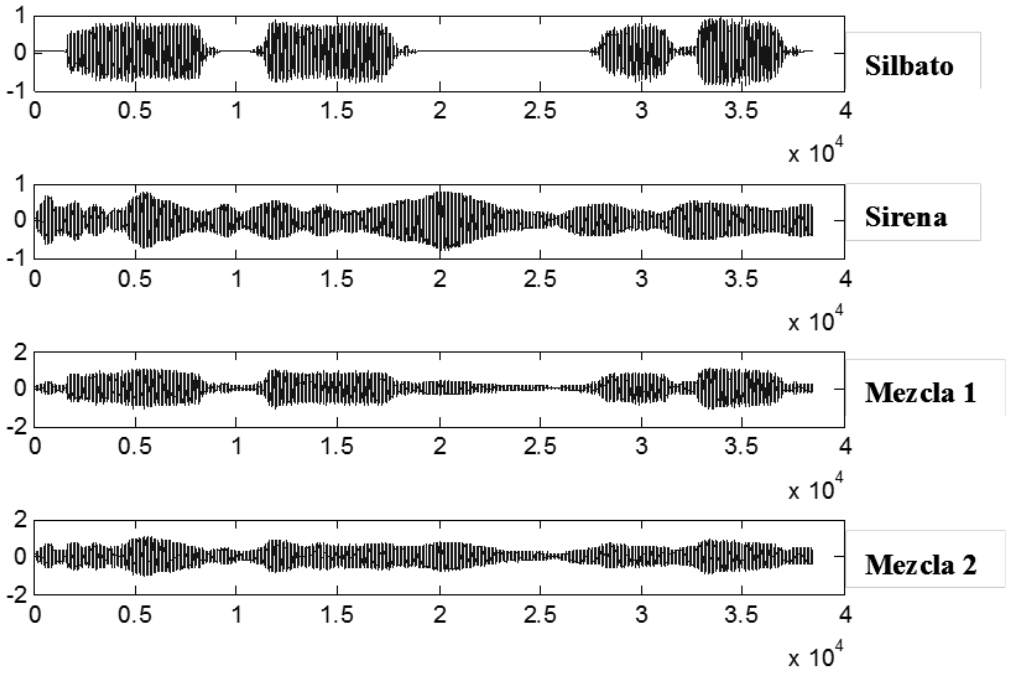
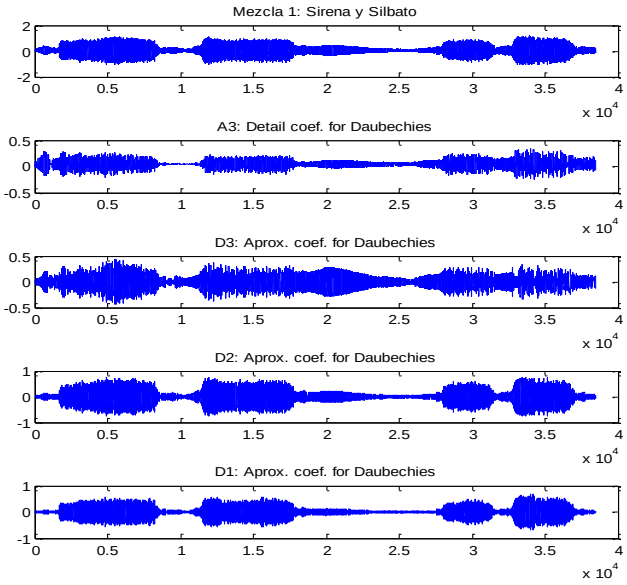
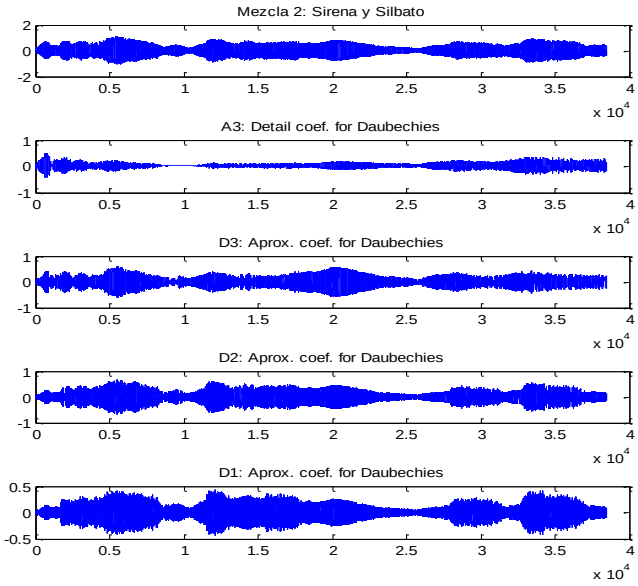


Fig. 5.2 Señales originales y mezclas lineales

Las gráficas (Fig. 5.3) corresponden a los 4 niveles obtenidas de las mezclas entre el silbato y la sirena en el dominio del tiempo, utilizando la transformada wavelet con filtro Daubechies. Resultado después de aplicar a cada una de las bandas el método FASTICA, como se puede observar se obtiene una separación exitosa manteniendo la descomposición en bandas hecha por la aplicación de la transformada wavelet, Fig. 5.4.



(a)



(b)

Fig. 5.3 Descomposición Wavelet de las mezclas lineales en 4 niveles

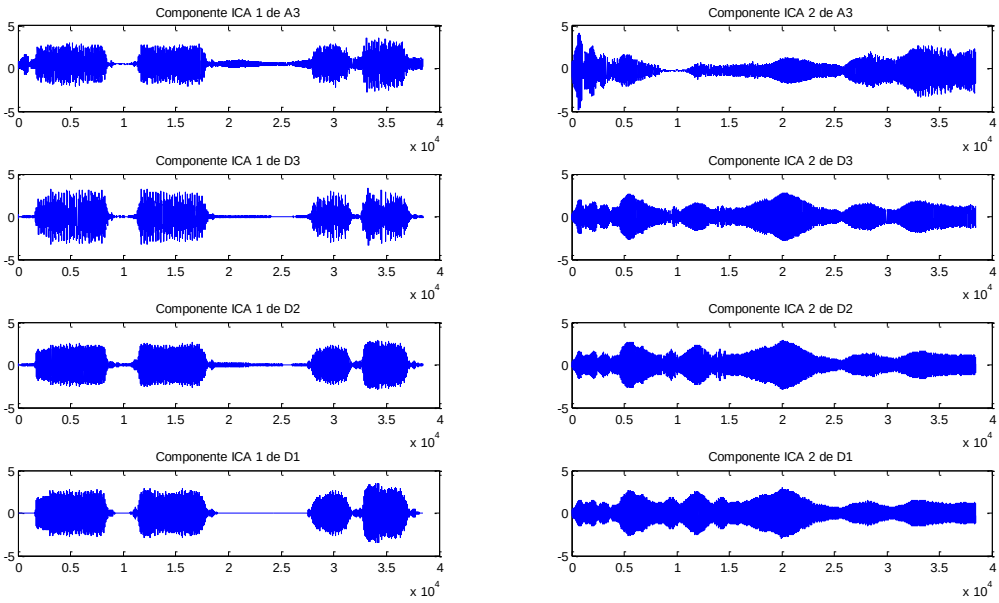
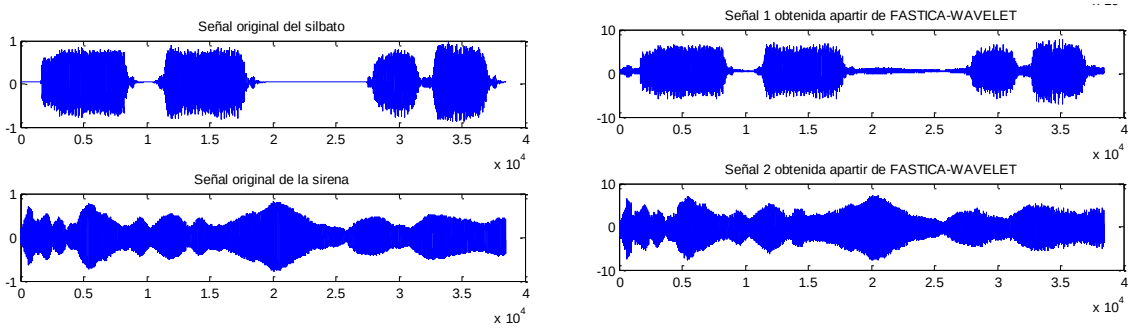


Fig. 5.4 Separación con ICA de la descomposición en bandas Wavelets

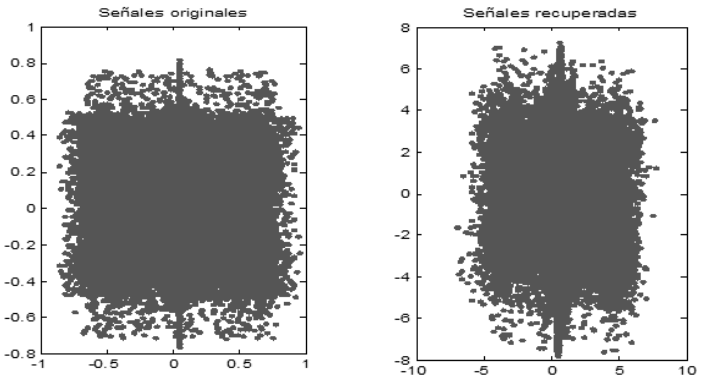
Después de sumar las bandas que componen cada señal obtenida a través de ICA, se comparan con las señales originales en la Fig. 5.5, como se puede observar, se obtiene una buena separación, sin embargo las señales estimas no son totalmente independientes, en ambas señales obtenidas se tiene de todavía parte de la mezcla original.



(a) *(b)*

Fig. 5.5 Comparación de los sonidos sirena y silbato entre: a) señales originales y b) señales estimadas

En Fig. 5.6, se muestra la distribución de los datos de la señal del silbato vs sirena, de las señales originales a la izquierda y de las señales obtenidas mediante Wavelet-FASTICA a la derecha, donde se observa que las señales estimadas aún presentan dependencia estadística.



(a) *(b)*

Fig. 5.6 Distribución de los datos de sirena vs silbato, a) señales originales y b) señales estimadas

Experimento 2. Con el propósito de caracterizan la respuesta del modelo FastICA, en la separación ciega de mezclas lineales, se realizaron mezclas de dos tonos $\boldsymbol{s}_1$ y $\boldsymbol{s}_2$ variando las ponderaciones en la matriz de mezcla $\boldsymbol{A}$ y reportando la relación señal a ruido (*SNR, Signal Noise Ratio*) en dB de las componentes encontradas, es decir las señales estimadas $\boldsymbol{y}_1$ y $\boldsymbol{y}_2$, la frecuencia de muestreo (fs) de los tonos generados es de 44,100 Hz, ver Tabla 5.3. Como se puede observar, si la ponderación de las señales presentes en la mezcla es muy próxima la SNR disminuye, por otro lado la SNR aumenta cuando en las frecuencias altas que sean predominantes den la mezcla.

Tabla 5.3 Caracterización de FastICA variando matriz de mezcla

Variación	Frecuencia [Hz]		Matriz de Mezcla	SNR [dB]		OBSERVACIONES
	s1	s2	A	y1	y2	
MEZCLA	800	2K	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	48.4	42.1	• Si la ponderación de las señales en las mezclas son muy próximas la SNR disminuye. • Si predomina en la mezcla la frecuencia más alta, la SNR aumenta.
	800	2K	$\begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}$	22.65	21.15	
	800	2K	$\begin{bmatrix} 1 & 0.9 \\ 0.1 & 1 \end{bmatrix}$	34.6	46.64	
	800	2K	$\begin{bmatrix} 1 & 0.1 \\ 0.9 & 1 \end{bmatrix}$	32.22	35.6	
	800	2K	$\begin{bmatrix} 1 & 0.7 \\ 0.4 & 1 \end{bmatrix}$	42.17	44	
	800	2K	$\begin{bmatrix} 1 & 0.4 \\ 0.7 & 1 \end{bmatrix}$	36.7	35	
	800	2K	$\begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix}$	40.13	44.7	
	800	2K	$\begin{bmatrix} 1 & 0.01 \\ 0.01 & 1 \end{bmatrix}$	40.2	44.5	

La Tabla 5.4, muestra los resultados obtenidos al variar la amplitud de las señales fuente manteniendo fija la relación de mezcla en 0.5, al igual que el caso anterior la $fs = 44{,}100$. Se

observa que si la amplitud de las señales originales disminuye también disminuye la SNR en las componentes independientes, por lo tanto la señale con mayor amplitud presente en la mezcla presenta una mejor separación indicando una valor de SNR mayor. También mejora el desempeño en las señales que predominan en amplitud y además tienen frecuencias más altas.

Tabla 5.4 Caracterización de FastICA variando la amplitud de las fuentes

Variación	Frecuencia [Hz]		Matriz de Mezcla	SNR [dB]		OBSERVACIONES
	s1	s2	A	y1	y2	
AMPLITUD	800	2K	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.8 \end{bmatrix}$	41.3	35.26	• Conforme disminuye la amplitud de las señales mezcladas disminuye la SNR. • Se obtiene una mejor SNR en la señal que predomina en la mezcla. • Se obtiene una mejor separación cuando la señal que predomina es de mayor frecuencia.
	800	2K	$\begin{bmatrix} 0.8 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	40.65	42	
	800	2K	$\begin{bmatrix} 0.6 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	34.7	42.8	
	800	2K	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.6 \end{bmatrix}$	34.5	29.7	
	800	2K	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 0.1 \end{bmatrix}$	37.3	39.65	
	800	2K	$\begin{bmatrix} 0.1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	32.5	37	

El ultimo parámetro a variar es la frecuencia de los tonos de entrada (señales fuente), manteniendo fija la amplitud en la unidad y mezcladas con una ponderación de 0.5. En este caso se aprecia que conforme las frecuencias de las señales fuente son más próximas, la SNR obtenida de las señales estimadas al realizar la separación disminuye, sin embargo el desempeño sigue siendo muy bueno. Por lo tanto, se concluye que el rango de frecuencias de las señales fuente no afecta en el proceso de separación ciega, utilizando FastICA.

Tabla 5.5 Caracterización de Infomax variando la frecuencia de las fuentes

Variación	Frecuencia [Hz]		Matriz de Mezcla	SNR [dB]		OBSERVACIONES
	s1	s2	A	y1	y2	
FRECUENCIAS	500h	1k	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	52.7	53.5	• Conforme las frecuencias de las señales a mezclar estén más próximas, la SNR disminuye, aunque la disminución está en el orden de 40 dB.
	800h	1k	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	53.95	53.9	
	900h	1k	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	48.5	43.3	
	950h	1k	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	45	74.3	
	1k	10k	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	74.8	79.5	
	10k	11k	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	69.7	68	

5.2 Experimentos y resultados sobre mezclas reales

En las señales reales (*grabadas directamente con sensores, ver Anexo B)* el primer paso es aplicar un filtro pasa altas a 80 Hz, ya que en términos de frecuencia, la voz humana está normalmente entre el rango de 80 Hz y 1100 Hz considerando todo el rango de voces masculinas y femeninas. Lo único que se graba en esas frecuencias es ruido del micrófono, ruido residual de los cables de alimentación, etc.

De esta manera se elimina ruido provocado por situaciones que no son tomadas en cuentas en esta investigación, ya que la voz, presenta las frecuencias más bajas de las señales que se desean separar, descritas anteriormente, en la Tabla 5.6 se hace una comparación del nivel de prisión del sonido con algunos de los sonidos más comunes.

Tabla 5.6 Comparación entre el nivel de presión del sonido y sonidos comunes

dB (A)[1]	Intensidad
140	Avión de chorro o fuego de artillería
130	Umbral del dolor
120	Umbral de percepción táctil
100	Interior de avión de hélice
90	Orquesta sinfónica o banda
80	Interior de automóvil a alta velocidad
70	Conversación frente a frente
50	Interior de una oficina publica
40	Interior de una oficina privada
30	Interior de una recamara
20	Interior de un teatro vacío
0	Umbral de audición

***1** Medido en la escala de un medidor estándar de intensidad sonora*

Experimento 3. Se captaron mezclas reales en campo abierto, utilizando tres micrófonos direccionales y tres señales distintas (voz, piano y sonido de un perro ladrando). Los micrófonos se colocaron en dirección a las fuentes, es decir cada micrófono estaba alineado a cada fuente. Las fuentes se encontraban separadas a una distancia de 15 metros de los micrófonos y a una distancia angular de 60° entre ellos, como se muestra en la Fig. 5.7. La frecuencia de muestreo utilizada es de $fs = 25{,}000\ muestras/s$ y el tamaño es de 725,000 muestras (*29 segundos*).

Una vez captadas las señales, se aplica el filtro pasa altas de 80 Hz y se realiza el proceso de blanqueado a las mezclas grabadas, se obtiene 4 bandas de cada mezcla utilizando la transformada wavelet de Deubechies, en la Fig. 5.8 se observa la descomposición wavelet de una mezcla.

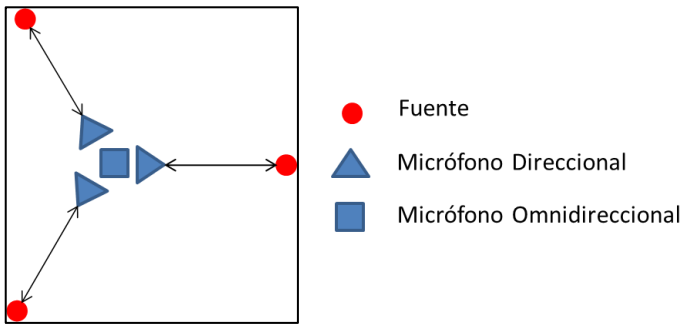


Fig. 5.7 Arreglo de 3 micrófonos direccionales y 1 micrófono omnidireccional

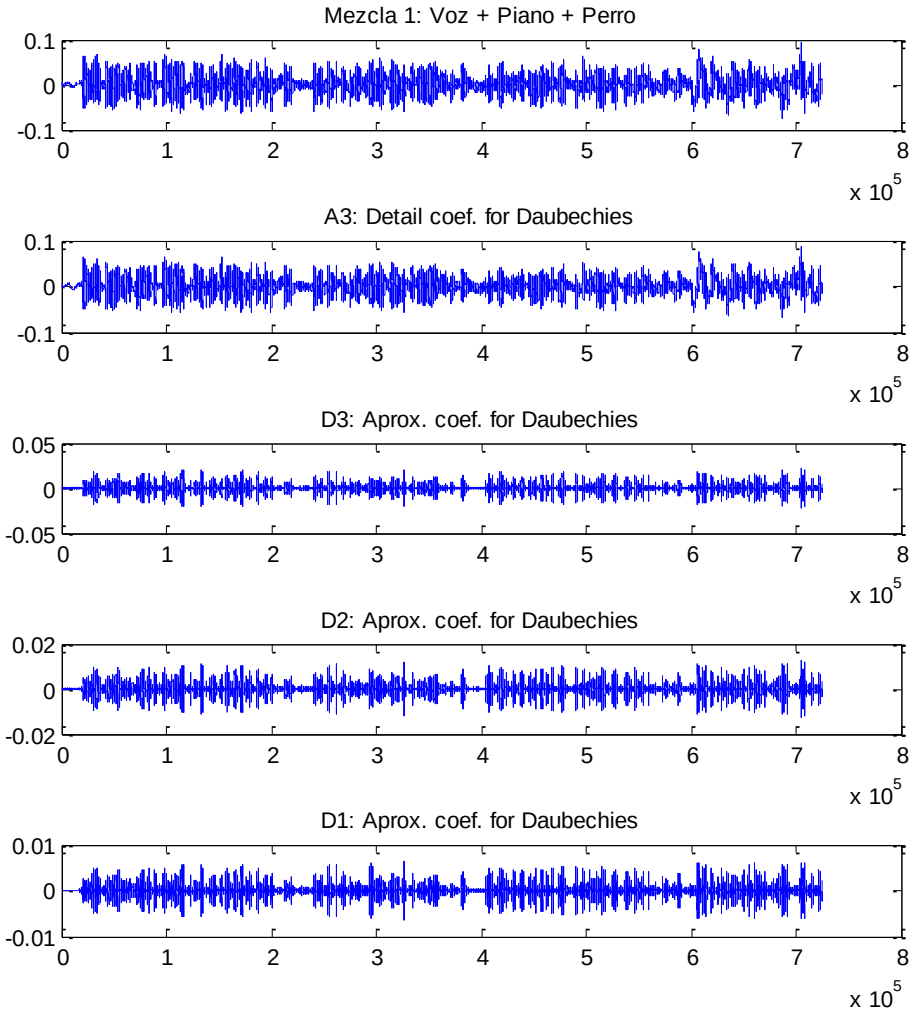
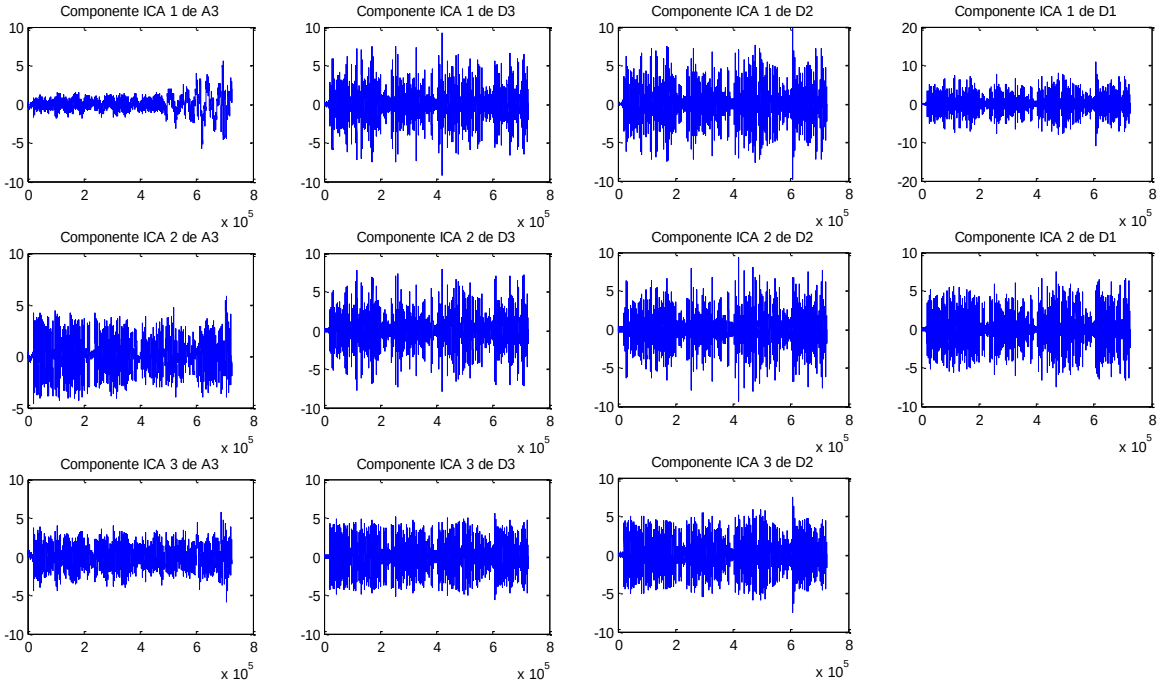


Fig. 5.8 Descomposición en bandas de una mezcla real con tres señales fuente

Es siguiente paso es aplicar el algoritmo FastICA a los grupos por bandas de cada una de las 3 mezclas, es decir, los vectores de entrada para FastICA se expresan de la siguiente manera:

$$\begin{aligned} \boldsymbol{x}_1(t) &= [a31 \quad a32 \quad a33]^T \\ \boldsymbol{x}_2(t) &= [d31 \quad d32 \quad d33]^T \\ \boldsymbol{x}_2(t) &= [d21 \quad d22 \quad d23]^T \\ \boldsymbol{x}_4(t) &= [d11 \quad d12 \quad d13]^T \end{aligned} \tag{5.5}$$

La siguiente gráfica muestra las componentes obtenidas con FastICA a cada nivel de banda, para A3 y D3, se detectaron tres componentes independientes. Para la banda D2 se tienen 3 componentes independientes y para la banda D1, se obtuvieron solo dos componentes al aplicar el algoritmo de FASTICA.



(a) **(b)**

Fig. 5.9 Componentes independientes obtenidas a partir de las bandas a) A3 y D3, b) D2 y D1

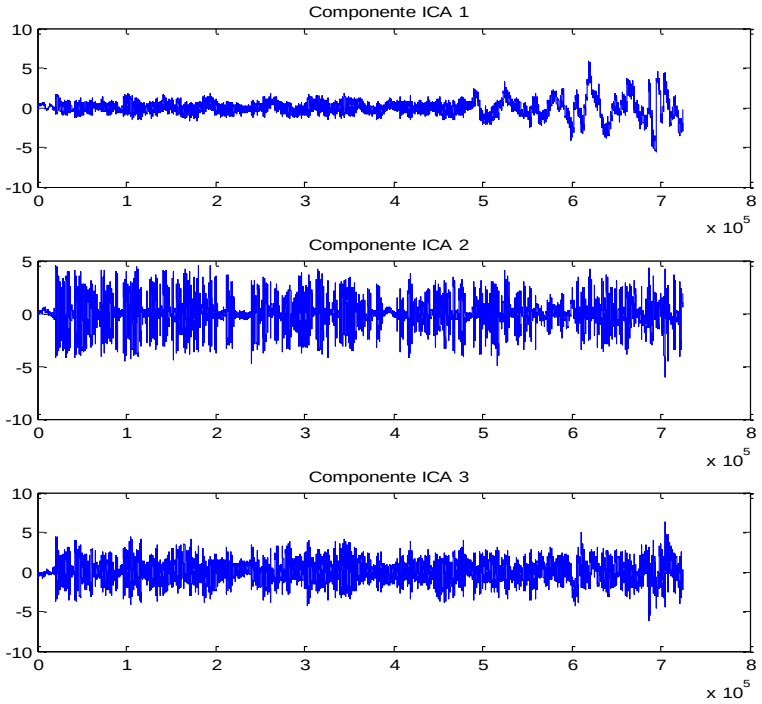
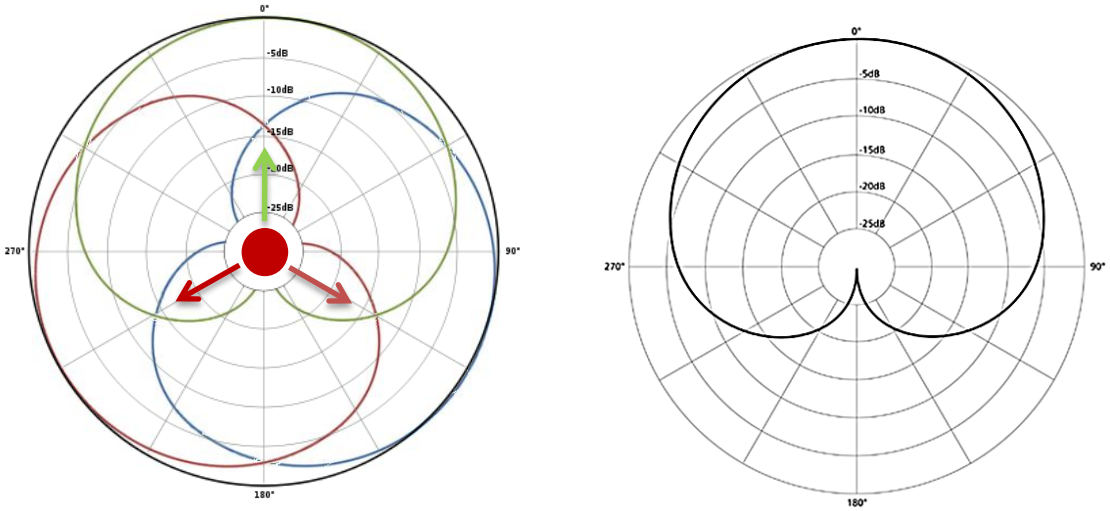


Fig. 5.10 Señales reales separadas utilizando Wavelet-FastICA

Experimento 4. La Fig. 5.11a muestra la distribución de los sensores con una separación angular de 60° entre ellos, son 3 micrófonos direccionales apuntando a cada una de las tres fuentes situadas a 15 m del centro, donde se encuentran los 3 micrófonos. El diagrama polar de la Fig. 5.11b, muestra el comportamiento de cada micrófono direccional.



(a) **(b)**

Fig. 5.11 a) Distribución física de los micrófonos, b) Diagrama polar del comportamiento de un micrófono direccional

Usando la distribución anterior se grabaron señales de tonos generados a las frecuencias de 500 Hz y 1 kHz, la Fig. 5.12 muestra las tres mezclas obtenidas por los micrófonos, después de ser filtradas. Como se observa en las mezclas se encuentran presentes los dos tonos con distintas ponderaciones.

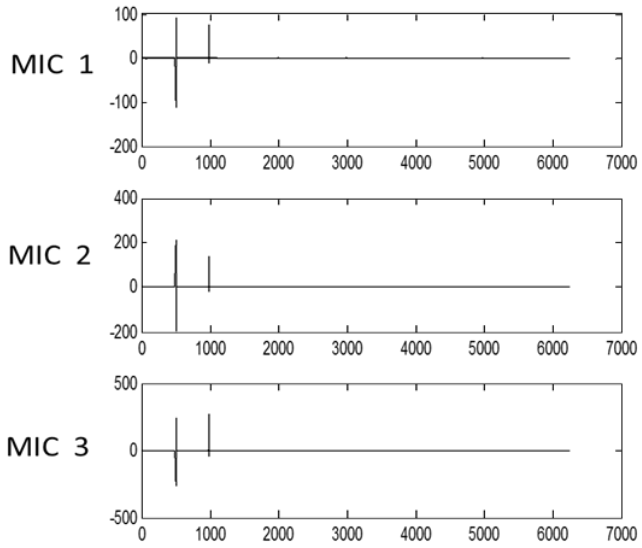


Fig. 5.12 Espectro de frecuencia de mezclas reales, compuestas por tonos de 500 Hz y 1 kHz

Aplicando el algoritmos de separación ciega de fuentes FastICA, se obtuvieron dos componentes Fig. 5.13, la primer componente corresponde al tono de 1kHz y la segunda al tono de 500 Hz. En este caso se obtiene una separación satisfactoria ya que los tonos no están muy próximos entre sí, además se utilizaron las tres mezclas para separar solo dos componentes, es decir, con mayor información.

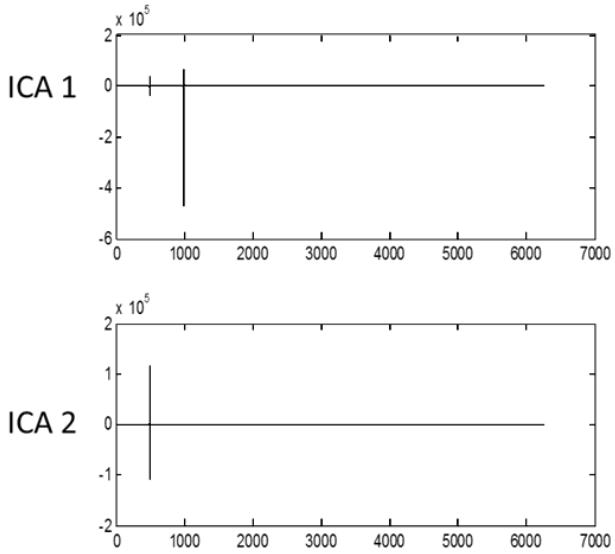


Fig. 5.13 Señales de tonos obtenidas al aplicar FastICA

Experimento 5. Utilizando la distribución de sensores de la Fig. 5.11a, se grabaron señales de tonos con frecuencias de 500 Hz, 1 kHz y 2 kHz. Las siguientes graficas muestran el espectro de frecuencia de las componentes independientes obtenidas al aplicar el algoritmo FastICA, Fig. 5.14. Se observa que las señales estimadas no son tonos puros, pero si se obtiene mayor presencia en cada frecuencia de las señales originales, esto debido a que las mezclas en ambientes reales nos son lineales.

Para comparar los resultados de FastICA y aplicar Wavelet-FastICA se calculó la relación señal a ruido (*SNR, Signal Noise Ratio*) de las señales estimadas utilizando cada metodo. Como se observa en la Tabla 5.7, se obtienen mejores resultados utilizando Wavelet-FastICA, donde la frecuencia de 500 Hz si logra ser recuperada.

Tabla 5.7 Relación señal a ruido de tonos separados con FastICA y Wavelet-FastICA

Frequencies	*FastICA [dB]*	*Wavelet-FastICA [dB]*
500 Hz	+6.02	-10.36
1 kHz	-7.95	-5.53
2 kHz	-6.02	-9.95

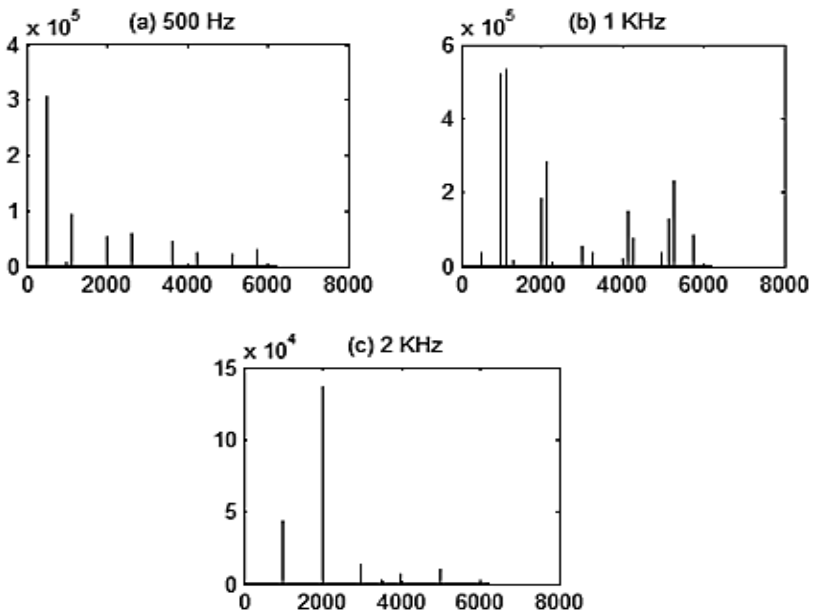


Fig. 5.14 Tonos separados mediante Wavelet-FastICA de mezclas reales

Experimento 6. Para este ejercicio se utilizaron las mezclas que se describieron en el experimento anterior, es decir, bajo las mismas condiciones. Las señales corresponden a tonos de 500 Hz, 1kHz y 2kHZ, sin embargo ahora se considera el modelo convolutivo. La Fig. 5.15 muestra las componentes encontradas al aplicar los filtros deconvolutivos y después FastICA.

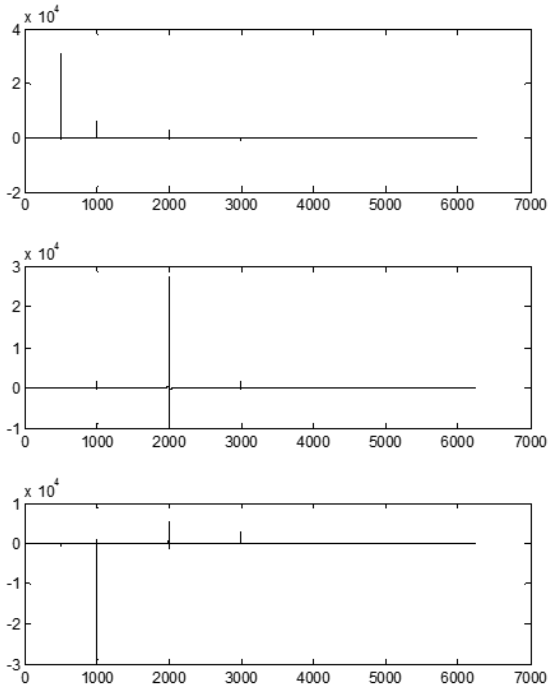


Fig. 5.15 Señales estimadas a partir de mezclas reales utilizando el modelo deconvolutivo-ICA

Como se aprecia en la Fig. 5.15, se obtiene una mejor separación al considerar el modelo convolutivo y aplicar un filtro de deconvolución antes de obtener las componentes independientes mediante ICA. Para comparar el desempeño del método deconvolutivo, con el modelo lineal antes utilizado, se calculó el índice SNR. Es claro que los resultados en la separación mejoran si se utiliza el modelo convolutivo, Tabla 5.8.

Tabla 5.8 Relación señal a ruido de tonos separados con FastICA y modelo Convolutivo

Frecuencias [Hz]	Deconvolución SNR [dB]	ICA SNR [dB]
500	-13.47	-10.36
1000	-15.06	-5.53
2000	-24.06	-9.95

Experimento 6. En este experimento se utilizó un arreglo de micrófonos lineal, separados entre sí 40.5 cm, y con 1.5m lineales de separación entre los sensores y las fuentes. Las fuentes fueron simuladas por bocinas convencionales de computadora. Como se observa en la Fig. 5.16, se colocaron 2, 3 y 4 fuentes, en un cuarto cerrado.

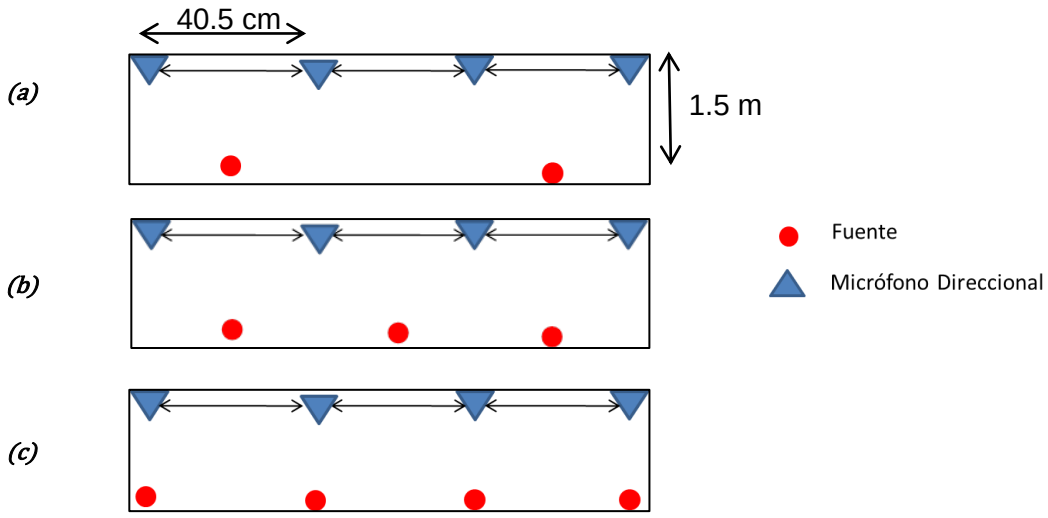


Fig. 5.16 Arreglo de micrófonos direccionales con a) 2 fuentes, b) 3 fuentes y c) 4 fuentes

En (a) las fuentes tienen una separación de 80 cm entre ellas, en el inciso (b) y (c) la separación entre las fuentes es de 40.5 cm colocadas como se observa en la figura anterior. En las Fig. 5.17 y Fig. 5.18, se tienen fotografías de la distribución física de los cuatro micrófonos utilizados y las fuentes simuladas por bocinas convencionales de computadora. La diferencia de alturas entre las fuentes y los sensores es de 60 cm.

Fig. 5.17 Bocinas utilizadas para simular 2, 3 y 4 fuentes en un cuarto cerrado

Fig. 5.18 Arreglo lineal de 4 micrófonos para mediciones en interior

Las señales que se generaron fueron tonos a 800Hz y 1kHz, y los siguientes sonidos: camión, voz, silbato y sirena. Se exponen los resultados obtenidos al separar los tonos de 800 Hz y 1kHz, la razón de utilizar tonos es que facilita observar gráficamente los resultados de la separación. El espectro de frecuencia de las mezclas grabadas por el arreglo lineal de micrófonos se muestra en la Fig. 5.19. Pruebas y resultados del experimento anterior:

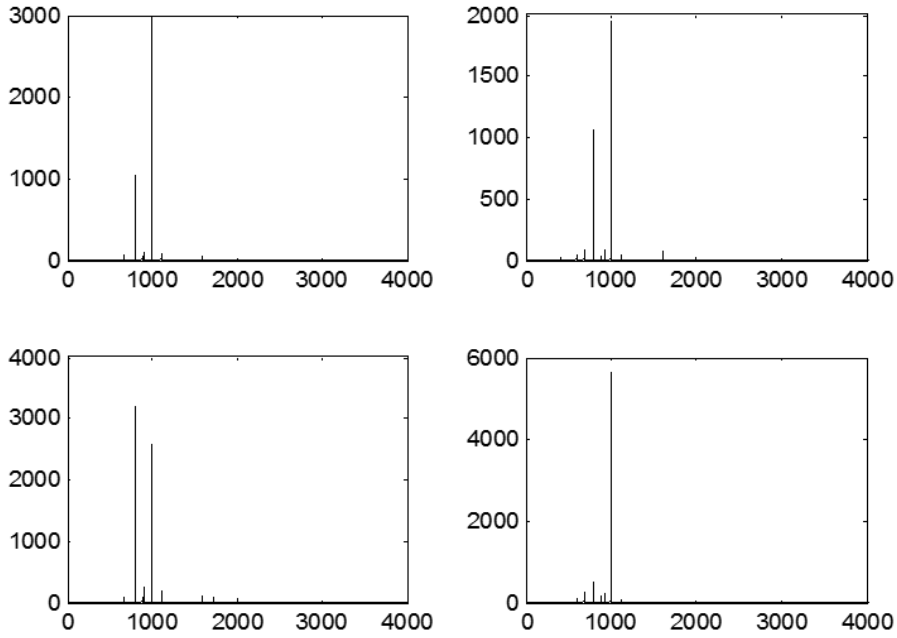
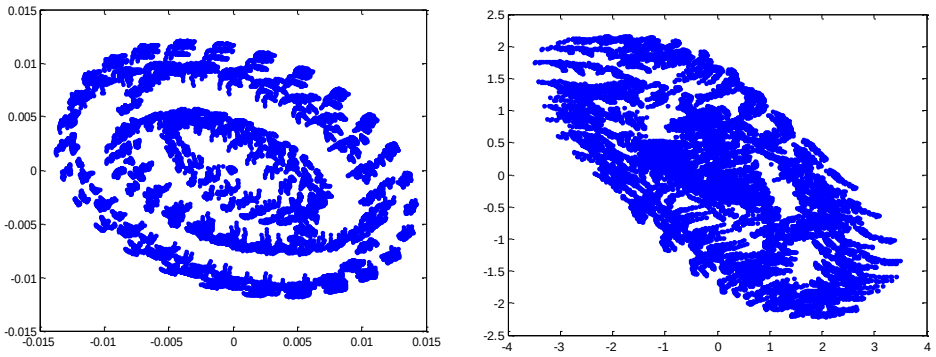


Fig. 5.19 Espectro de frecuencia de mezclas captadas por arreglo lineal de micrófonos

La Fig. 5.20a, muestra la distribución de los datos de 2 de las mezclas obtenidas y la Fig. 5.20b representa la distribución de las misma mezclas después de ser blanqueadas. En las señales blanqueadas se elimina la correlación que existe entre las mezclas, sin cumplir aun con la condición de ser independientes.



(a) *(b)*

Fig. 5.20 Distribución de datos a) señales mezcladas, b) señales blanqueadas

Una vez filtradas y blanqueadas las mezclas, se realiza el proceso de deconvolución ciega, para esto a partir de las mezclas blanqueadas se ajustan los coeficientes de los filtros, la Fig. 5.21 muestra los filtros obtenidos para eliminar la convolución existente entre las mezclas de los dos tonos.

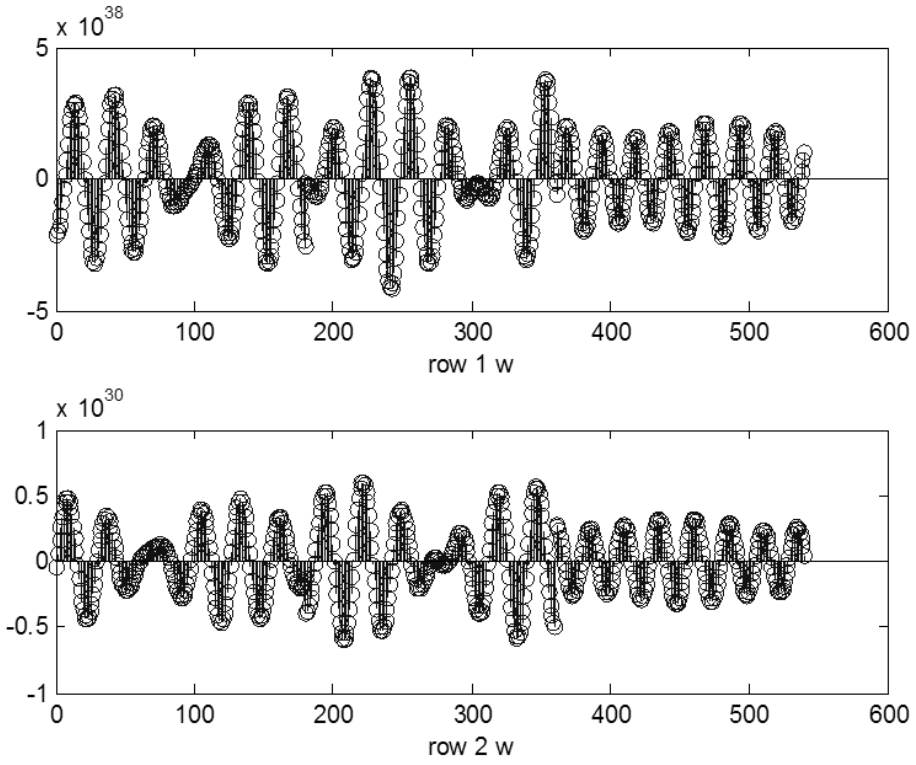
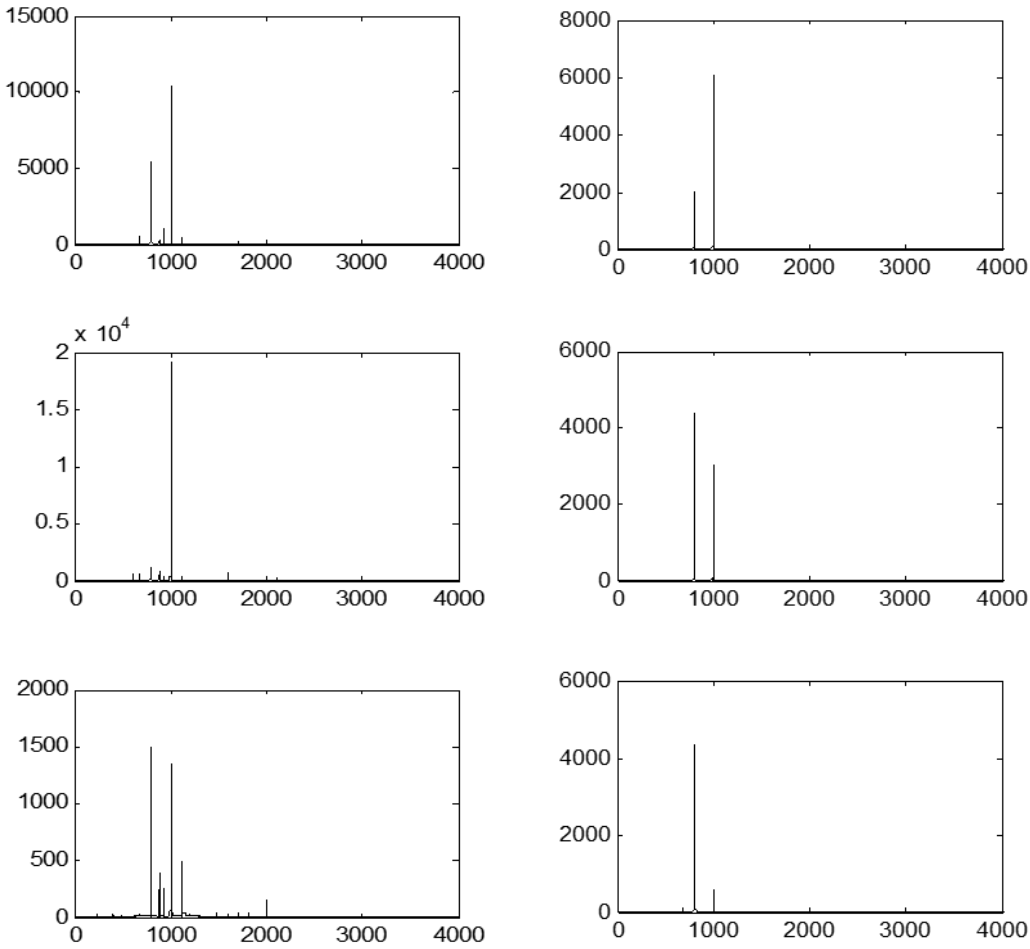


Fig. 5.21 Filtros para deconvolución de tonos a 800 Hz y 1kHz

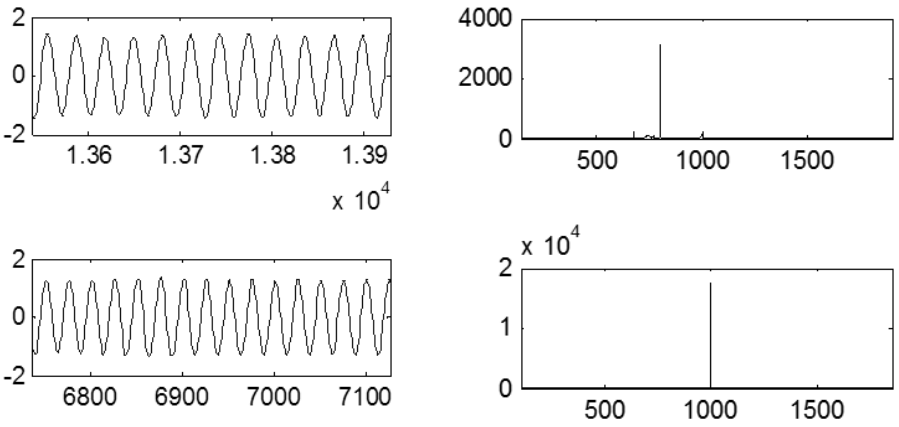
El resultado al deconvoluciar las señales blanqueadas se observa en la Fig. 5.22, a la izquierda se tienen el espectro de frecuencia de las señales blanqueadas y a la derecha el espectro de las señales filtradas con los filtros de deconvolución. Como se puede apreciar se eliminan frecuencias que no pertenecen a las señales originales y que se producen debido a la convolución que ocurre al captarlas en un ambiente real.

Las señales de la Fig. 5.22 pueden ahora ser separadas utilizando el algoritmo de FastICA para mezclas lineales. Se obtiene una buena separación al aplicar FastICA, ver Fig. 5.23, como se puede observar las señales estimadas contienen los tonos de las señales originales, es decir, de 800 Hz y 1kHz.



(a) *(b)*

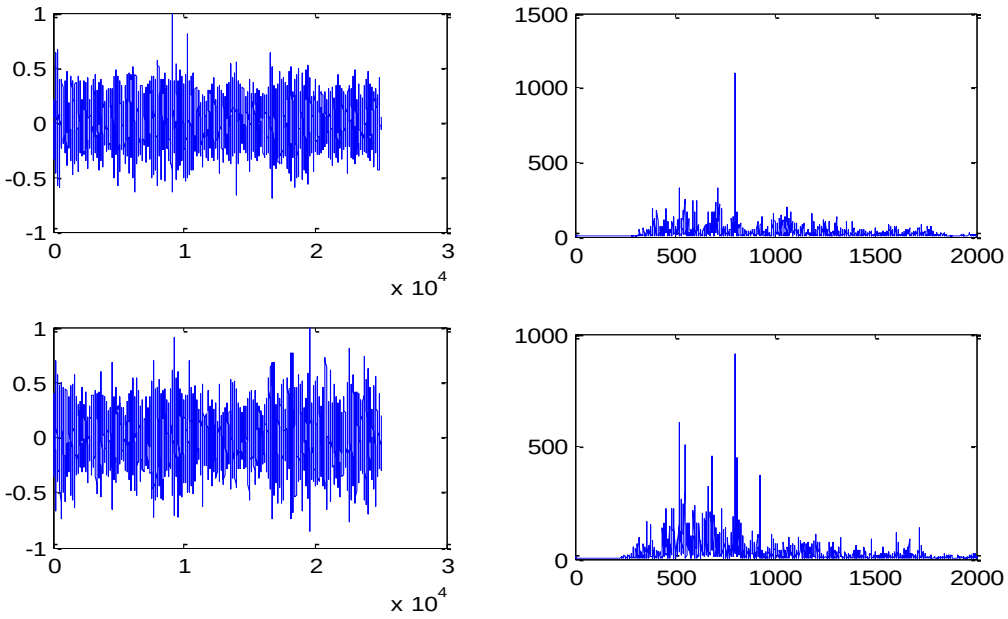
Fig. 5.22 Respuesta en frecuencia de a) señales blanqueadas, b) señales deconvolucionadas



(a) *(b)*

Fig. 5.23 Señales separadas con FastICA, a) dominio del tiempo, b) espectro de frecuencia

Experimento 7. En la siguiente prueba se mezcló el sonido del motor de un carro y un tono a 800 Hz. Las mezclas reales de los 4 sensores se muestran en la Fig. 5.24a en el dominio del tiempo y Fig. 5.24b el espectro de frecuencia.



(a) *(b)*

Fig. 5.24 Mezcla real de tono 800 Hz y carro a) dominio del tiempo, b) espectro de frecuencia

De forma similar al experimento anterior, se filtran y blanquean las 4 mezclas grabadas, posteriormente se obtienen los filtros de deconvolución Fig. 5.25, para eliminar los efectos de la convolución presentes en las mezclas originales.

Las señales deconvolucionadas se muestran en la Fig. 5.26, tanto en el dominio del tiempo como el espectro de frecuencia. Como se puede observar, en la primera mezcla prácticamente desaparece el sonido del carro y en la segunda se encuentran presentes las dos señales originales, esto facilita la tarea de ICA por la gran diferencia de proporción presente en las mezclas.

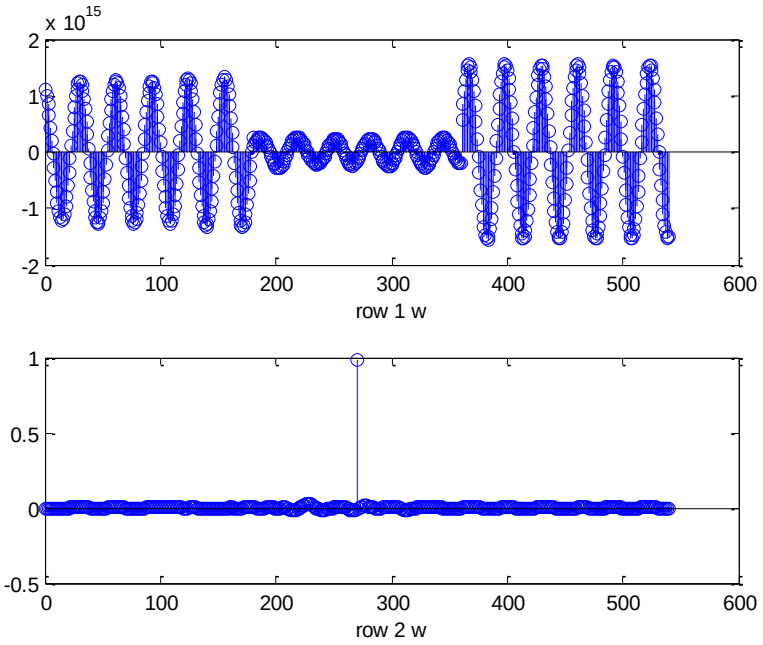


Fig. 5.25 Filtros de deconvolución para sonido de carro y tono de 800 Hz.

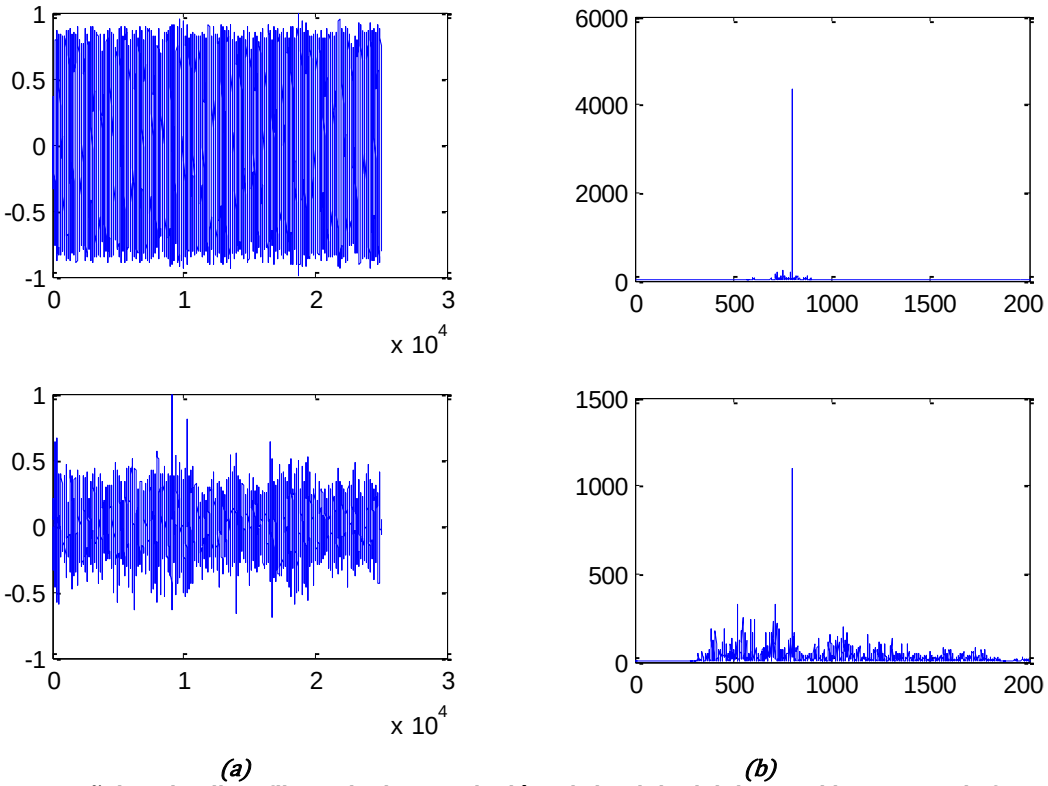


Fig. 5.26 Señales al aplicar filtros de deconvolución, a) dominio del tiempo, b) espectro de frecuencia

El paso final es obtener las componentes independientes aplicando FastICA, la separación se muestra en la Fig. 5.27, donde se logra separar el tono de 800 Hz del sonido producido por un carro satisfactoriamente.

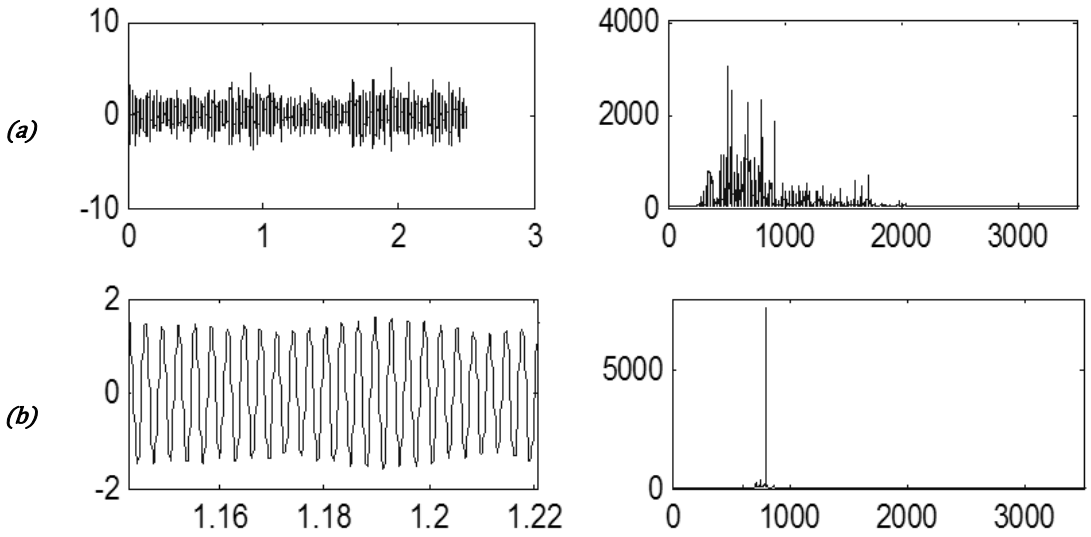
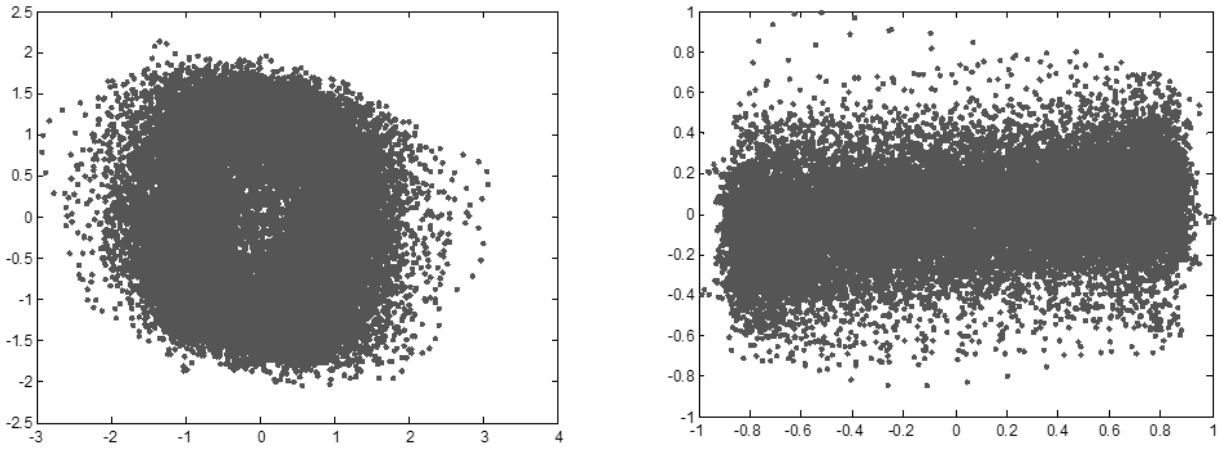


Fig. 5.27 Componentes separadas por ICA en tiempo y frecuencia de a) señal de carro, b) tono 800 Hz

La Fig. 5.28, muestra la distribución de los datos de las señales blanqueadas, es decir las señales decorrelacionadas y las señales después de aplicar los filtros de deconvolución. Al aplicar los filtros, se logran definir las componentes presentes en la mezcla por lo que mejora el desempeño del algoritmo de separación.



(a) *(b)*

Fig. 5.28 Distribución de datos a) señales blanqueadas, b) señales deconvolucionadas

Experimento 8. En este caso se mezclaron dos sonidos, el motor de un carro y el sonido de una silbato. En las gráficas de la Fig. 5.29 se muestras dos de las mezclas reales en el dominio del tiempo y el espectro de frecuencia.

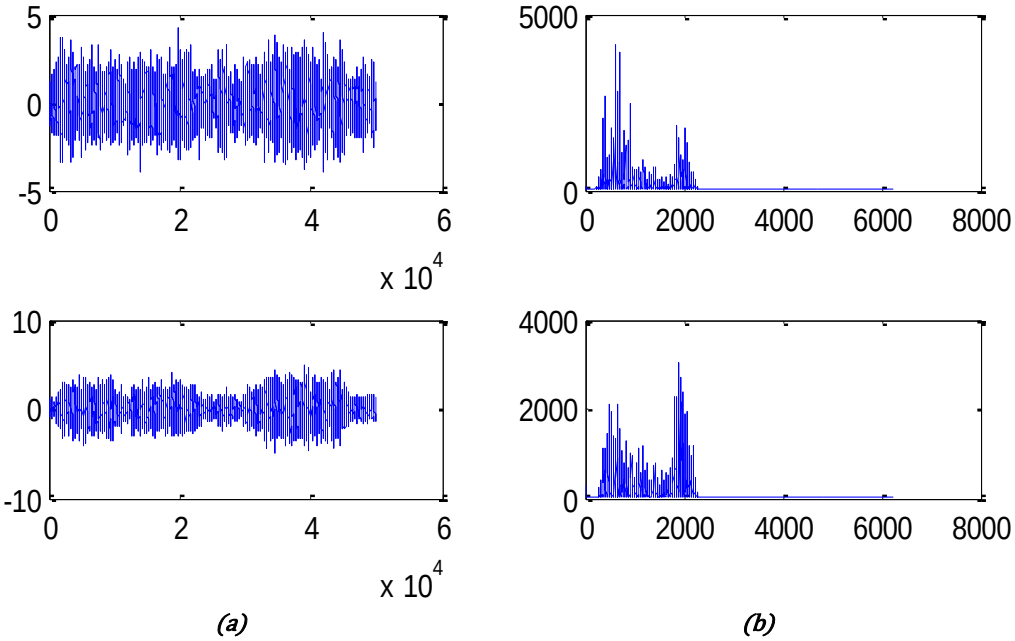


Fig. 5.29 Mezclas reales de carro y silbato a) dominio del tiempo, b) espectro de frecuencia

Estas mezclas reales fueron filtradas para eliminar las frecuencias por debajo de 80 Hz y blanqueadas para decorrelacionar las señales. Antes de realizar la separación ciega es necesario deconvolucionar las mezclas, para esto se utilizan los filtros de la Fig. 5.31, que fueron calculados con el método deconvolución ciega descrito en el capítulo 4.

En este experimento, solo se obtuvo la señal correspondiente al sonido del silbato de forma independiente, en la otra componente obtenida están mezclados los dos sonidos fuente donde predomina el sonido del silbato. Esto debido a la diferencia de amplitud entre las dos señales fuente.

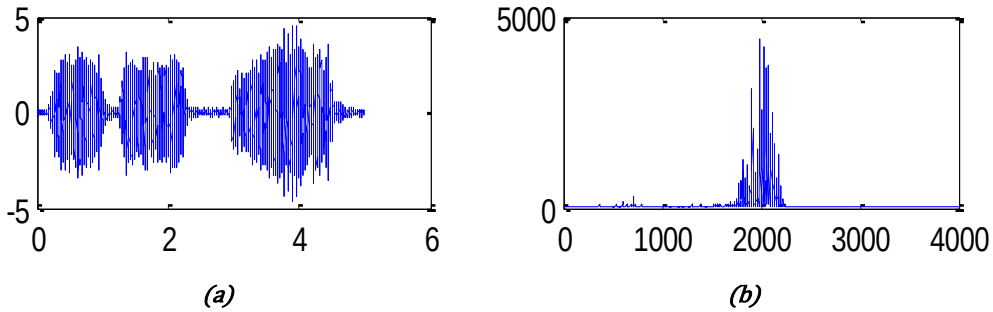


Fig. 5.30 Señal de silbato obtenida mediante Deconvolución-FastICA de mezclas reales

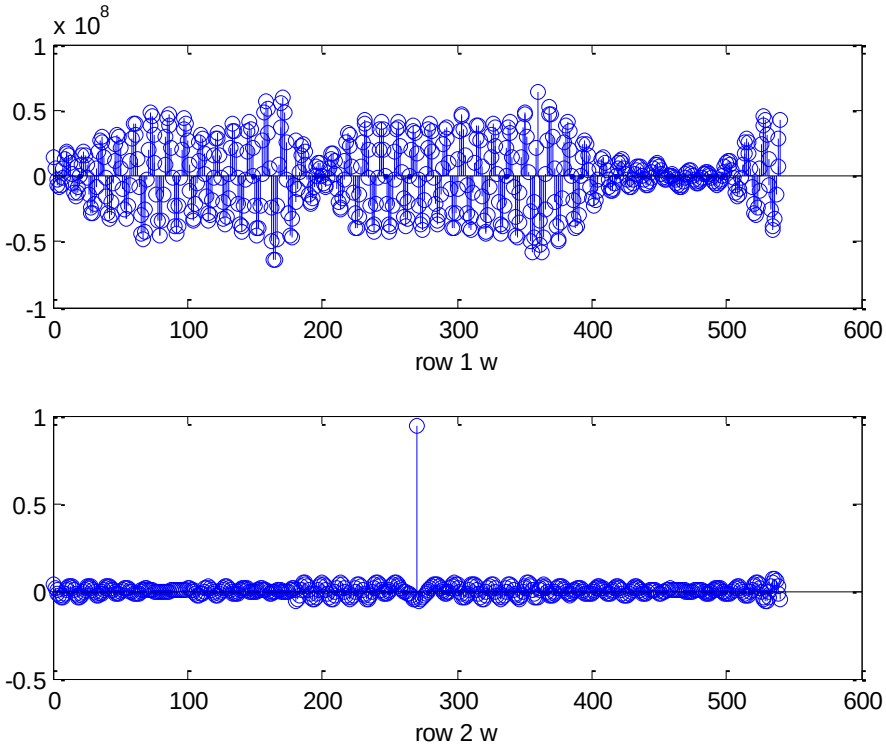


Fig. 5.31 Filtros de deconvolución para las señales reales de silbato y carro

5.3 Clasificación de señales acústicas ambientales

Para lograr diferenciar los distintos tipos de fuentes: voz, sirenas, silbatos, claxon y vehículos. Se eligieron los siguientes descriptores de sonido para caracterizar cada tipo de fuente, para el entrenamiento del clasificador se utilizan las señales que conforman la Base de datos con señales descargadas de internet (ver Tabla 5.2), una vez entrenado, el clasificador debe ser capaz de identificar el tipo de sonido que predomina en las señales estimadas obtenidas al realizar la separación ciega de fuentes mediante el proceso antes descrito (filtrado-blanqueado-deconvolución-FastICA).

Al extraer los siguientes índices: Componentes Tonales Emergentes (*CTE*), Componentes Impulsivas (*CI*), Componentes de Bajas Frecuencias (*CBF*), Nivel Sonoro Continuo Equivalente (*NSCE*) y Valores Pico (*VP*), no se logra identifica los diferentes tipos de fuentes, ya que en la mayoría de las señales se presenta una corrección en estos índices con valores de 0, 3, 6 ó 9, según lo indica las normas ambientales [9,7]. Las Fig. 5.32 - Fig. 5.36 muestran la distribución de los índices extraídos para separar las fuentes: autos, claxon, perro, sirenas y voz.

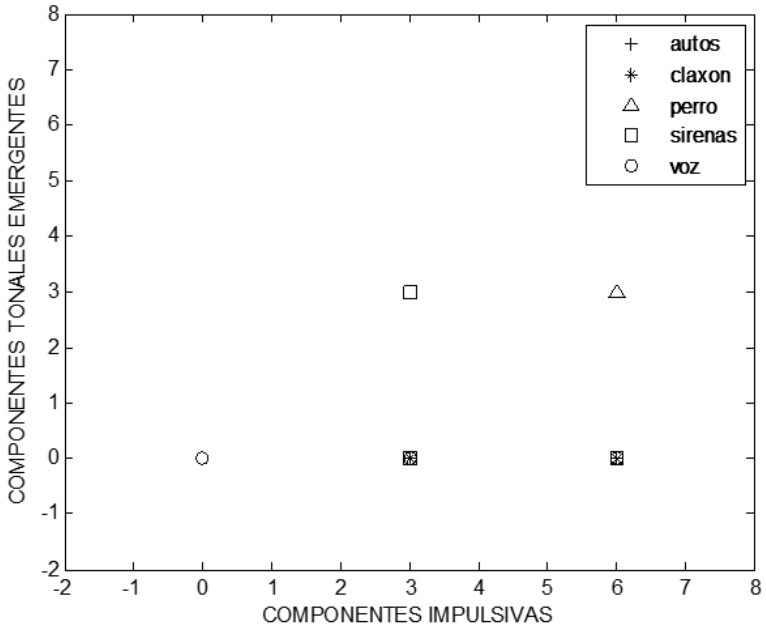


Fig. 5.32 Componentes Impulsivas vs Componentes Tonales Emergentes

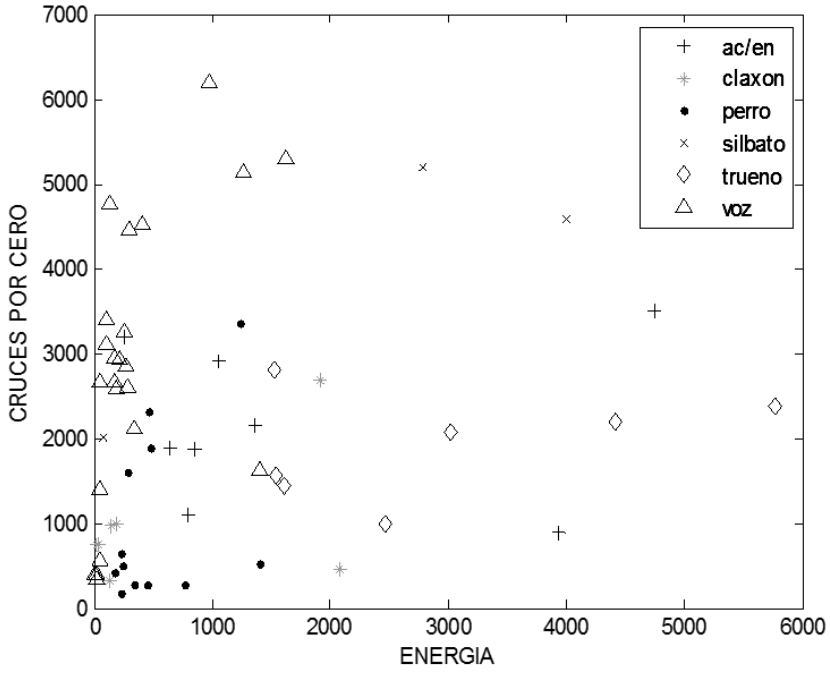


Fig. 5.33 Cruces por cero vs Energía

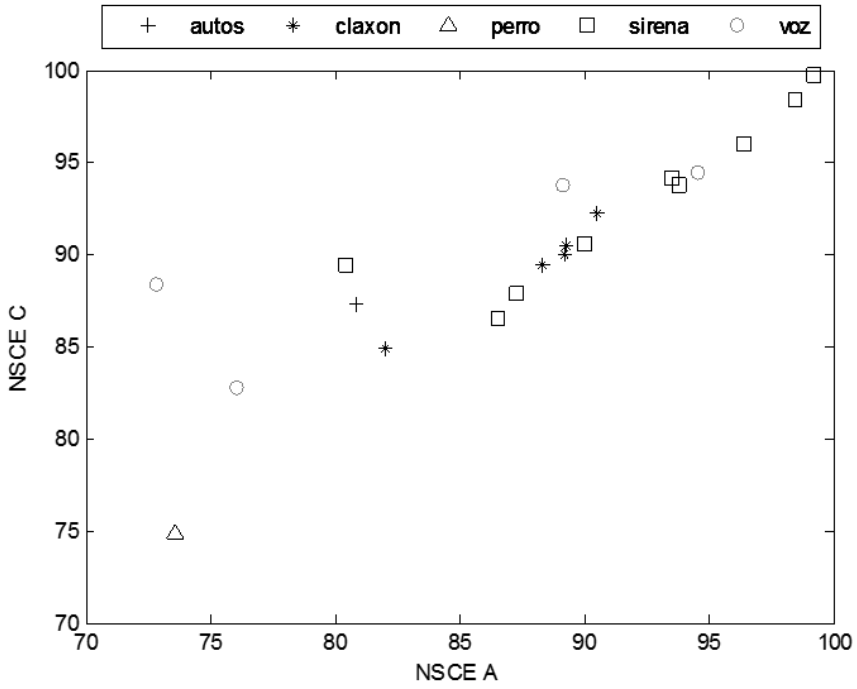


Fig. 5.34 NSCE C vs NSCE A

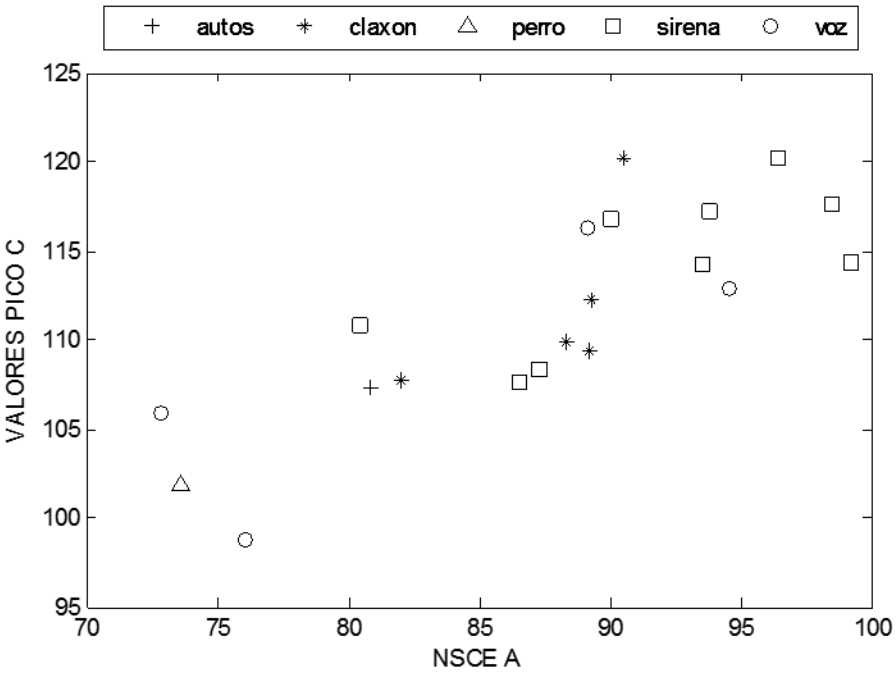


Fig. 5.35 Valores Pico C vs NSCE A

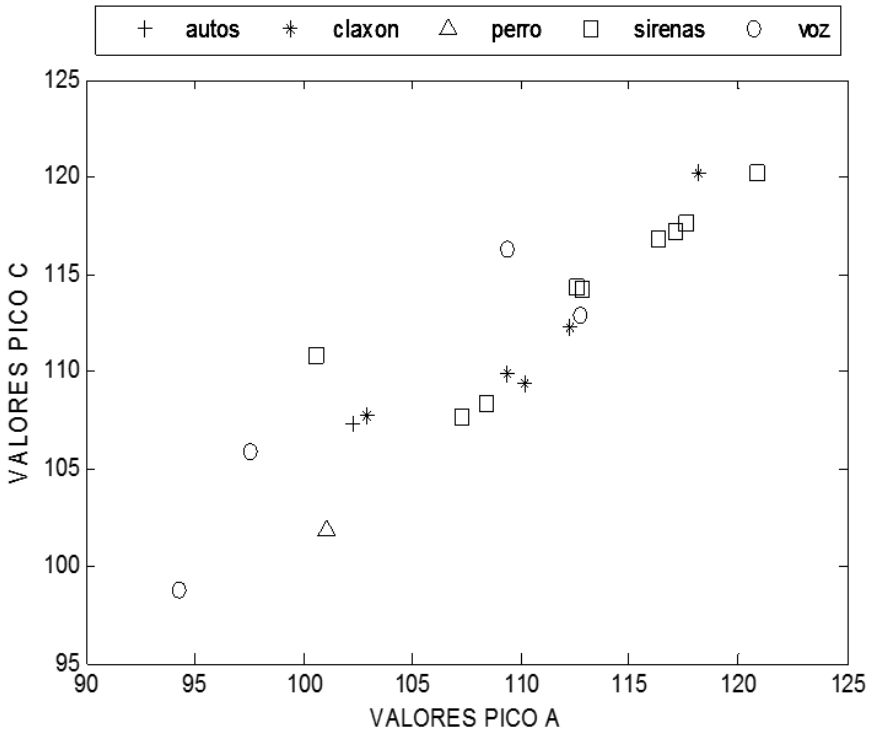


Fig. 5.36 Valores Pico C vs Valores Pico A

Como se puede observar en las gráficas anteriores, es posible identificar grupos de distintas clases utilizando los índices acústicos, sin embargo no es posible separar todas las clases propuestas. En la Tabla 5.9, se muestran los pares de índices que presentan una diferencia clara para identificar y clasificar las fuentes entre sí, sin embargo no todas las clases pueden ser clasificadas utilizando estos identificadores, por ejemplo, identificar la clase de silbatos de sirenas o entre autos y sirenas.

Tabla 5.9 Clasificación entre señales fuente, utilizando índices acústicos como rasgos característicos.

Grupos	Truenos	Silbatos	Perro	Claxon	Ac/En	Autos	Sirenas
Voz	E-Z	E-Z	E-Z	E-Z	E-Z	E-SC	PA-NA
Truenos		Z-SC	Z-SC	E-SC	Z-SC	--	--
Silbatos			E-Z	E-Z	E-Z	E-SC	--
Perro				NC-PC	E-SC	NC-PC	NA-PC
Claxon					E-SC	NA-PA	NA-PC
Acelerar/Enfrenar						--	--
Autos (Tráfico)							E-Z

Fuzzy C-Means Tipo 2. El propósito es representar y gestionar la incertidumbre que se produce en los grupos con diferente volumen y/o densidad para un conjunto determinado. En consecuencia, la gestión de la incertidumbre ayuda a los centros de los grupos para converger a una posición más razonable.

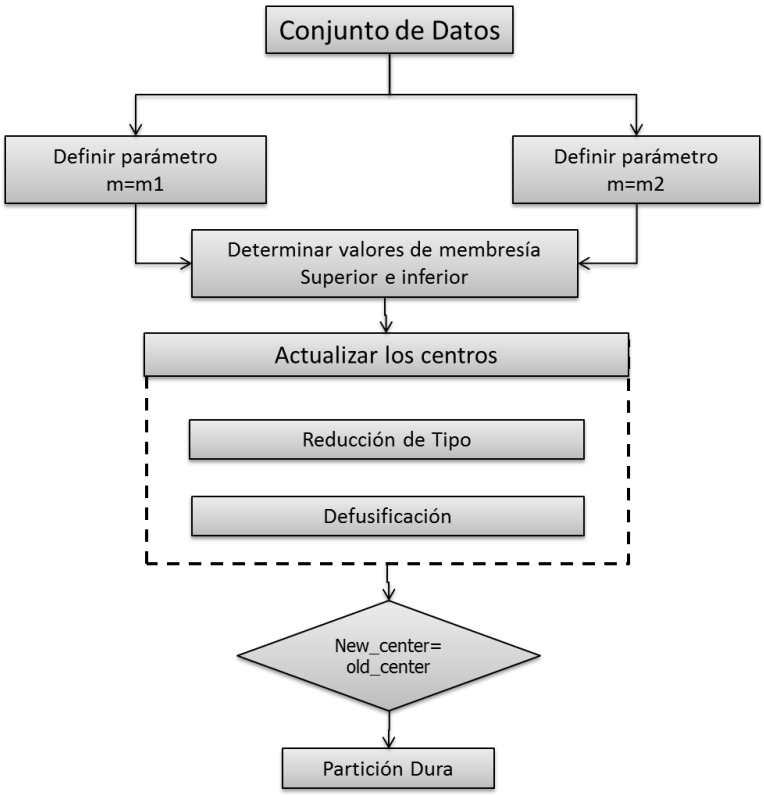


Fig. 5.37 Diagrama de flujo del algoritmo FCM-2

Se realizaron pruebas utilizando técnicas básicas de minería de datos, y extrayendo como rasgos característicos los coeficientes LPC en valores numéricos y nominales, también se evaluaron los índices acústicos descritos anteriormente. Las técnicas utilizadas de minería de datos son: 1 Regla (*One Rule*), *divide y conquistarás*, C4.5 y K-Medias. Se realizaron pruebas para identificar 4, 5, 6 y 8 señales fuentes originales. La Tabla 5.10, muestra los resultados en porcentaje de aciertos y error, obtenidos al evaluar estas técnicas.

Tabla 5.10 Evaluación de técnicas de minería de datos

Técnica	BD	No. Instancias	No. Grupos	Cross Validation	% Aciertos	% Error
1 Regla	LPC Numérico Pos.	9	4	3	11.11%	88.89%
C4.5	LPC Numérico Pos.	9	4	3	22.22%	77.78%
K-Means	LPC Numérico Pos.	9	4	3	33.33%	66.67%
1 Regla	LPC Numérico Real	9	4	3	11.11%	88.89%

Evaluación de técnicas de minería de datos (continua…)

C4.5	LPC Numérico Real	9	4	3	22.22%	77.78%
K-Means	LPC Numérico Real	9	4	3	33.33%	66.67%
Divide - Conquistar	LPC Nominal	9	4	s/n	100%	0%
1 Regla	Rasgos Numérico Norm	100	8	33	36%	64%
C4.5	Rasgos Numérico Norm	100	8	33	63%	37%
K-Means	Rasgos Numérico Norm	100	8	33	35%	65%
1 Regla	Rasgos Numérico Real	102	8	33	41.17%	58.82%
C4.5	Rasgos Numérico Real	102	8	33	61.76%	38.23%
K-Means	Rasgos Numérico Real	102	8	33	35.29%	64.71%
1 Regla	Rasgos Numérico Norm	90	6	30	38%	61%
C4.5	Rasgos Numérico Norm	90	6	30	66%	33%
K-Means	Rasgos Numérico Norm	90	6	30	38.89%	61.11%
1 Regla	Rasgos Numérico Norm	71	4	23	50.7%	49.3%
C4.5	Rasgos Numérico Norm	71	4	23	78.87%	21.12%
K-Means	Rasgos Numérico Norm	71	4	23	50.7%	49.3%
1 Regla	Rasgos Nominales	100	8	33	31%	69%
Divide - Conquistar	Rasgos Nominales	100	8	33	44%	49%
C4.5	Rasgos Nominales	100	8	33	47%	53%
Divide - Conquistar	Rasgos Nominales	87	6	s/n	77%	22%
Divide - Conquistar	Rasgos Nominales	80	5	s/n	81.25%	18.75%

La Tabla 5.11, se resumen las técnicas con los mejores porcentajes de aciertos, como se observa, se obtienen mejores resultados al utilizar rasgos nominales (etiquetas) que valores numéricos. También se obtienen mejores resultados utilizando como rasgos característicos los coeficientes LPC, se calcularon 12 coeficientes. Al incrementar el número de clases a clasificar, el desempeño disminuye, generando mejores resultados el método "*Divide y conquistarás*", que el método C4.5.

Tabla 5.11 Evaluación de rasgos característicos con algoritmos de minería de datos

Method	Database	Instances	Classes	% Hits	% Error
Divide and Conquer	LPC-Label	9	4	100%	0%
C.4.5	LPC-Numeric	9	3	95%	5%
Divide and Conquer	Features-Label	80	5	81.25%	18.75%
Divide and Conquer	Features-Label	87	6	77%	22%
C4.5	Features-Numeric	71	4	78.87%	21.12%
C.4.5	LPC-Numeric	9	4	77.78%	22.22%
C4.5	Features-Numeric	100	8	63%	37%

La Fig. 5.38, muestra la distribución de los datos por clases utilizados para evaluar técnicas básicas de minería de datos. Las clases consideradas en este experimento fueron sonidos de voz producidos por multitudes (etiquetada como gente), sonidos de camiones, claxon, ladrido de perros, silbatos para control de tránsito, sirenas, truenos y voz.

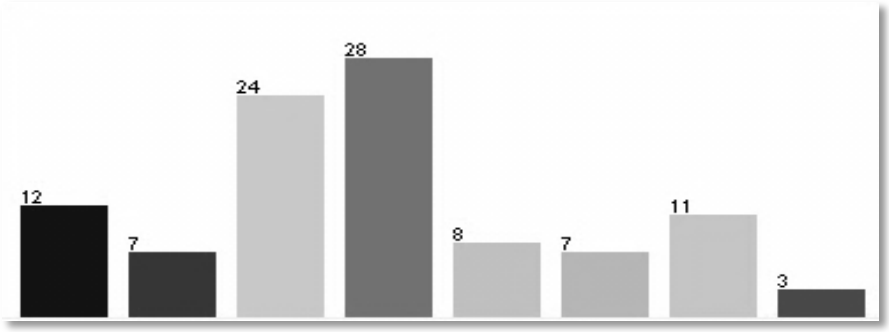


Fig. 5.38 Distribución de los datos por categoría: gente, camión, claxon, perro, silbato, sirena, trueno, voz.

La Fig. 5.39, muestra en árbol de decisión generado con el método C4.5, utilizando como rasgos característicos solo el coeficiente LPC No. 10, este árbol clasifica entre tres clases diferentes: camión, gente y sirena.

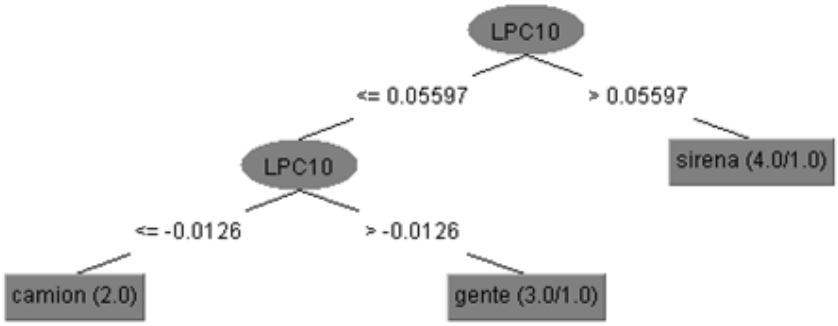


Fig. 5.39 Clasificación con C4.5 árbol de decisión utilizando LPC

Utilizando como rasgos característicos la energía de la señal a clasificar y la energía de señales prototipo representativas de cada clase, se obtuvieron grupos o clusters, y mediante los algoritmos FCM-2, se obtuvieron sus centros. La Fig. 5.40, muestra los grupos generados en 3D para identificar la clase de automóviles. Se obtiene 6 grupos diferentes que representan las 79 señales utilizadas para identificar esta clase. De manera similar, en la Fig. 5.41, se muestran los grupos obtenidos para representar la clase de sirenas, cuenta con 6 grupos diferentes para representar las 139 señales de la BD generada. Se obtuvieron de manera similar, los grupos representativos de las demás clases: claxon, voz y silbato.

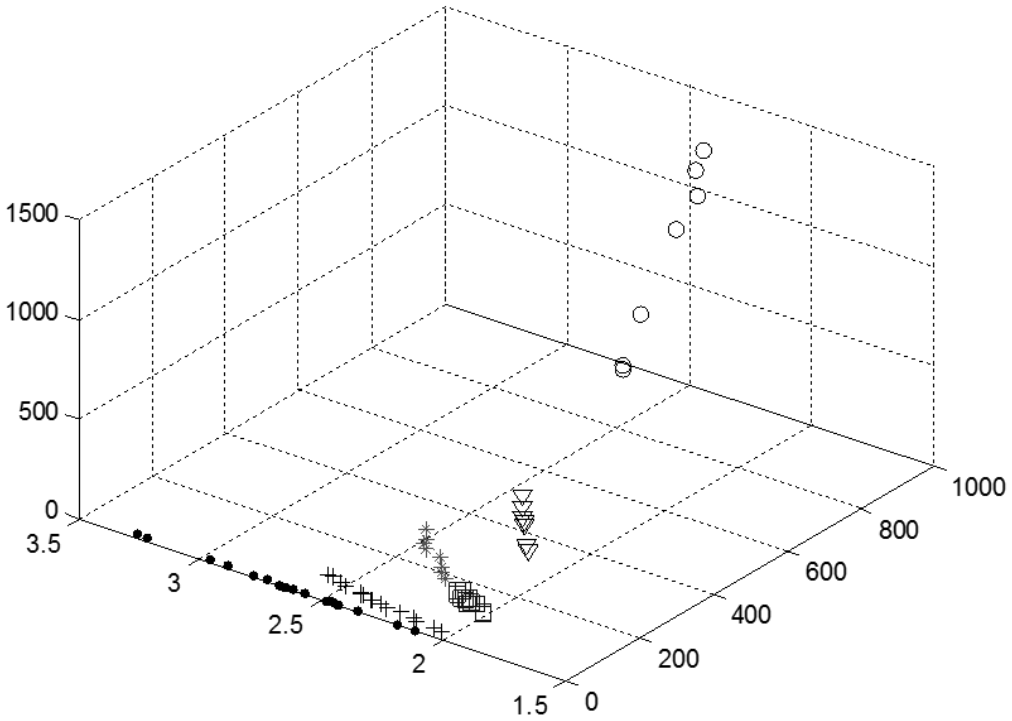


Fig. 5.40 Clasificador para identificar la clase de automóviles

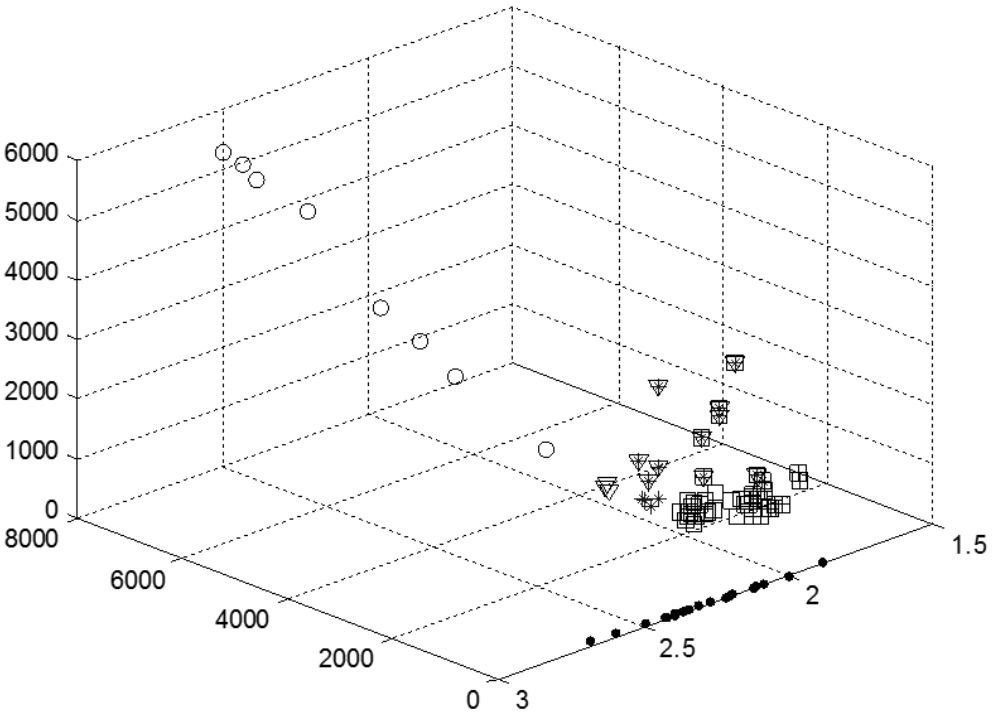


Fig. 5.41 Clasificador para identificar la clase de sirenas

La Fig. 5.42 muestra los grupos que representan la clase de *claxon, y los centros de grupos obtenidos al aplicar el algoritmo FCM-2, los datos están representados por puntos y los centros representados con cuadros. Los asteriscos representan los rasgos extraídos a una señal con el sonido de un claxon, como se puede observar, esta señal pertenece con mayor grado de pertenencia un grupo de este clasificador.*

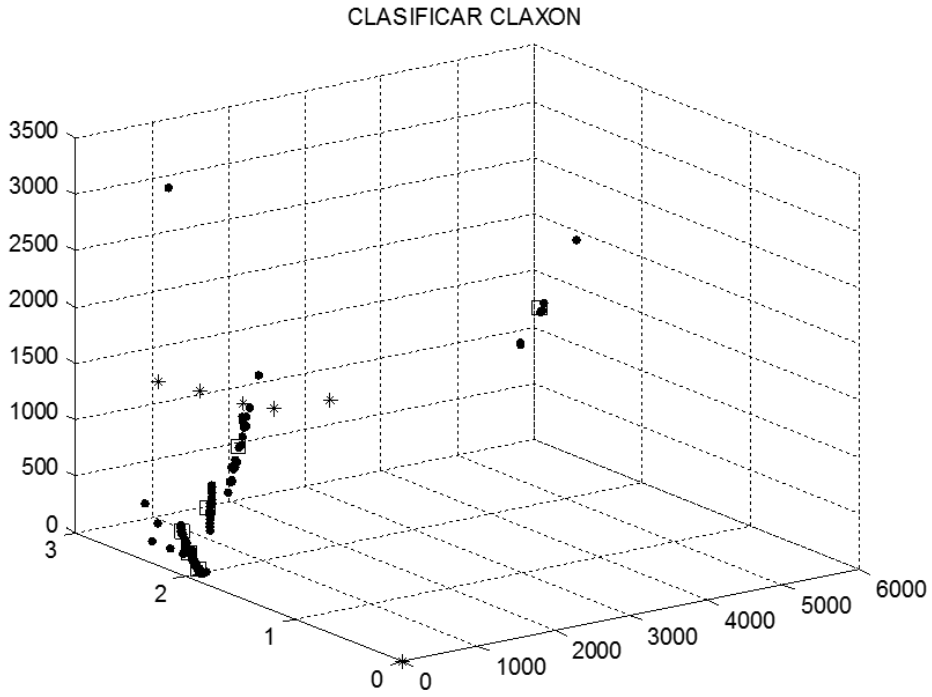


Fig. 5.42 Clasificador para identificar la clase de claxon y sus centros de cada cluster.

Así, se extraen los rasgos de la señal que se desea evaluar y se obtiene sus grados de pertenencia a cada cluster de cada clase de señales definidas: sirena, voz, autos, silbatos y claxon. Para diferenciar los resultados generados por cada clasificador se aplican las siguientes reglas de conocimiento:

1. Las señales solo pueden pertenecer a un solo grupo o cluster en cada clasificador, evaluando el grado de pertenencia $u(i,i)$ de la matriz de grados de pertenencia. Si la señal a clasificar tiene 2 o más grados de pertenencia mayores a 0.8, la señal es descartada para esta clase.
2. Solo se consideran los grados de pertenencia mayores a 0.8, es decir, si en un clasificador los grados de pertenencia $u(i,i)$ son menores, entonces la señal no pertenece a esta clase.
3. Si existen dos o más señales en distintos clasificadores con valores de $u(i,i) > 0.8$, se toma como clase el que tenga el mayor valor para $u(i,i)$.

4. Si existen dos o más señales en distintos clasificadores con valores de $u(i,i) > 0.8$, pero iguales entre sí, entonces se toma como clase, la que presente la menor distancia al dicho centro.

De esta manera se logran clasificar las 5 clases definidas entre sí, la siguiente tabla resume los porcentajes de acierto y error obtenidos.

Tabla 5.12 Resultados de clasificador FCM-2 para cada clase.

Clasificador	Aciertos	Errores	%Aciertos	%Error
Autos	67	9	88.15	11.84
Sirena	118	21	84.89	15.10
Claxon	132	12	91.66	8.33
Silbato	84	11	88.42	11.57
Voz	143	35	80.33	19.66

CONCLUSIONES

1. Al realizar la etapa de pre-procesamiento que implica aplicar un filtro blanqueador, realizar el blanqueado espacial las mezclas y descomponer en bandas wavelet, para posteriormente aplicar el método de análisis de componentes independientes se logra satisfactoriamente la separación ciega de fuentes en un ambiente controlado, considerando las fuentes fijas, con un SNR de atenuación de entre 10 dB y 20 dB. Se observó que los resultados varían dependiendo de la amplitud de las señales fuente presentes en las mezclas y las ponderaciones con que se encuentren mezcladas.

2. El desempeño del clasificador depende de los rasgos característicos extraídos, se evaluaron distintos rasgos característicos como índices para el control de ruido, energía de la señal, potencia, cruces por cero, valores pico, coeficientes LPC. Los mejores resultados de clasificación se obtienen al utilizar la energía de las señales y diseñando un clasificador difuso FCM-2 para cada clase.

3. Con base en los rasgos característicos se clasificaron los cinco tipos de sonidos que son voz, claxon, sirenas, claxon, y silbatos. Mediante la técnica de agrupamiento difuso de FCM tipo 2, se logró una exitosa separación. Además se realizaron pruebas con técnicas de minería de datos para obtener reglas de conocimiento que permitieran la clasificación, obteniendo un porcentaje de aciertos entre el 77% y 95%, dependiendo del tipo de fuente del que se trate y el número de fuentes que se encuentre mezcladas simultáneamente.

4. Utilizando el método propuesta se logra el objetivo de separar fuentes mezcladas en campo en un ambiente controlado, tomando en cuenta mezclas de 2, 3 y 4 fuentes fijas con un SNR de atenuación mayor al reportado en otros trabajos de separación ciega de fuentes tomando en cuenta hasta 4 fuentes simultaneas, además se obtiene una correcta clasificación con errores entre el 80% y 90% para los tipos de fuente considerados.

TRABAJO FUTURO

Además de los utilizados en esta investigación: Infomax y FastICA, existen diversos algoritmos de ICA que pueden ser empleados para resolver los problemas aquí expuestos y evaluar su desempeño.

Debido a que la extracción de rasgos característicos es una etapa fundamental para la clasificación, es necesario investigar y proponer diferentes métodos para extracción de rasgos que caractericen los tipos de sonidos presentes en zonas urbanas, que permitan mejorar el desempeño de los clasificadores.

En este trabajo solo se consideraron cinco tipos de sonidos que comúnmente contribuyen en la contaminación acústica de zonas urbanas, sin embargo es conveniente considerar otros tipos de fuentes, no solo de presentes en zonas urbanas sino también aquellos sonidos que perturban el bienestar y confort de la vida diaria del ser humano, como los presentes en fábricas, escuelas, oficinas, aeropuertos, entre otros.

Para la etapa de clasificación se utilizó una técnica de agrupamiento difuso tipo-2, así como también algunas técnicas de minería de datos, sin embargo existen muchas técnicas de clasificación que se pueden evaluar para identificar este tipo de señales.

El arreglo de micrófonos que se utilice para captar las mezclas influye en gran medida a como se formen las mezclas, para lograr una buena separación ciega es muy interesante contar con investigaciones sobre como colocar y distribuir los micrófonos (sensores) de medición tomando en cuenta las fuentes, reflexiones, retardos, el ambiente, etc.

Los experimentos se realizaron considerando las fuentes fijas, sin embargo en los ambientes reales las fuentes están en constante movimiento, por lo que se requiere una metodología donde se consideren este factor y sus implicaciones.

El método de ICA se pueden aplicar a una gran variedad de problemas donde se desee realizar una separación ciega de fuentes como en reconocimiento de voz, seguidor de locutor, estudios de vibraciones, cancelación de ruido en canales de comunicación, también es muy utilizado para cancelación de ruido en señales débiles (principalmente en señales biológicas), donde los métodos actuales para eliminación de ruido no son aplicables.

PRODUCTOS DE LA INVESTIGACIÓN

Referencias de los artículos publicados a partir de esta investigación:

1. *López, M.G., Molina, H., Sánchez, L.P. & Oliva, L.N., "Blind Source Separation of Audio Signals Using Independent Component Analysis and Wavelets",* CONIELECOMP 2011 *(21st International Conference On Electronics Communications And Computers)*, IEEE Catalog Number: CFP11363-CDR, ISBN: 978-1-4244-9557-3, Puebla, México, Febrero 2011.

2. *López, M.G., Molina, H., Sánchez, L.P. & Oliva, L.N., "Separation and Identification of Environmental Noise Signals using Independent Component Analysis and Data Mining Techniques",* CERMA 2011, Cuernavaca, México, Noviembre 2011.

REFERENCIAS BIBLIOGRÁFICAS

[1] Brüel&Kjaer Sound & Vibration Measurement A/S, "Ruido Ambiental," 2000.

[2] Ecología urbana. Wikipedia. [Online]. http://es.wikipedia.org/wiki/Ecolog%C3%ADa_urbana

[3] Leo L. Beranek, "Noise Reduction," Peninsula Publishing, 1991.

[4] Luis P. Sánchez Fernández, "Sistema Avanzado de Monitoreo Ambiental de Sonidos y Vibraciones," Proyecto CONACYT, Clave: 51283-Y, Vigencia: 2007-2010.

[5] Fernando J. Mato Méndez and Manuel Sobreira Seoane, "Análisis de componentes independientes en separación de fuentes de ruido de tráfico en vías interurbanas," in *E.T.S.I. Telecomunicación*, Universidad de Vigo, España, 2008.

[6] Fernando J. Mato Méndez and Manuel Sobreira Seoane, "Sustracción espectral de ruido en separación ciega de fuentes de ruido de tráfico," in *E.T.S.I. Telecomunicación*, Universidad de Vigo, España, 2008.

[7] Norma Oficial Mexicana, "Evaluación del Ruido Producido por Fuentes Fijas," NOM-081-ECOL-1994, 1994.

[8] Norma Oficial Mexicana, "Evaluación de la exposición del cuerpo humano a las vibraciones," NOM-025-STPS-1995, 1995.

[9] Norma Oficial Mexicana, "Determinación de los Niveles de ruido Ambiental," NOM-AA-62-1979, 1979.

[10] Te-Won Lee, *Independet Component Analysis. Theory and Applications*, Tercera edición ed.: Kluwer Academic Publishers, 2001.

[11] Keshia N. Leach, "A Survey Paper on Independent Component Analysis," in *Proceedings of the Thirty-Fourth Southeastern Symposium on System Theory*, 2002, pp. 239-242.

[12] J. Larsen, T. Kolenda L. Hansen, "On Independent Component Analysis for Multimedia Signals," in *Multimedia Image and Video Processing*.: CRC Press, 2000, ch. 7, pp. 175-200.

[13] A. Hyvärinen, R. Vigario, J. Hurri, E. Oja J. Karhunen, "Applications of Neural Blind Source Separation to Signal and Image Processing," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 131-134, 1997.

[14] J. Schiller, *Mobile Communications*. NY: Adison Wesley, 2000.

[15] A. Hyvärinen, J. Karhunen, E. Oja J.Hurri, "Image Feature Extraction Using Independent Component Analysis," *Proceedings of the IEEE Nordic Signal Processing Symposium (NORSIG)*, no. 475-478, 1996.

[16] Martin Kermit, *Independent Component Analysis: Theory and Applications*., April 2000.

[17] E. Oja, L. Wang, R. Vigario, J. Joutsensalo J. Karhunen, "A Class of Neural Networks for

Independet Component Analysis," *IEEE Trans. Neural Net.*, vol. 8, no. 3, pp. 486-504, Mayo 1997.

[18] Rüdiger Brause Björn Arlt, "The Principal Independent Components of Images," Internal Report FB Infromatik.

[19] A. Eide, K. Agehed, V. Beeanovia, T. Lindblad M. Kermit, "Independet Component Analysis as an Oder Identification Tool," *Elsevier*, October 2000.

[20] K. Itoh T. Yamaguchi, "An Algebraic Solution to Independent Component Analysis," *Optics Communications*, vol. 178, no. 59-64, Octubre 2000.

[21] Erikki Oja, Aapo Hyvärinen, and Juha Karhunen, *Independet Component Analysis*.: John Wiley & Sons, Inc., 2001.

[22] S. Unadkat, H. Patel, T. Schulte, L. Medsker S. Nayeri, "Independent Component Analysis in Physics," *International Joint Conference on Neural Networks*, pp. 1926-1929, 2001.

[23] K. Torkkola, "Blind Separation for Audio Signals - Are We There Yet?," *Proc. Workshop on Independet Component Analysis and Blind Source Separation*, pp. 1-6, 1999.

[24] H. Saruwatari, K. Shikano T. Nishikawa, "Blind source separation of acoustic signals based on multistage ICA combining frequency-domain ICA and time-domain ICA," *IEICE Trans. Fundamentals*, vol. E86-A, no. 4, pp. 846-858, Sep 2003.

[25] Jan Larsen, Ulrik Kjems, Lucas C. Parra Michael S. Pedersen, "A survey of Convolutive Blind Source Separation Methods," *Multichannel Speech Processing Handbook*, vol. 51, pp. 1065-1084, 2007.

[26] W. Pedrycz J. Valente de Oliveira, *Advances in Fuzzy Clustering and its Applications*.: John Wiley & Sons, Ltd., 2007.

[27] L.A. Zadeh, "Fuzzy sets," *Inform. and Control*, no. 8, pp. 338-353, 1965.

[28] M.S. Yang, "A Survey of Fuzzy Clustering," *Mathl. Comput. Modeling*, vol. 18, no. 11, pp. 1-16, 1993.

[29] R. Kalaba, L.A. Zadeh R. Bellman, "Abstraction and pattern classification," *Journal Mathematics Analysis and Applications*, no. 2, pp. 581-586, 1966.

[30] E.H. Ruspini, "A new approach to clustering," *Inform. and Control*, no. 15, pp. 22-32, 1969.

[31] C. Sanchez, J. Rieta, M. Fernandez, J. Ballesteros R. Alcaraz, "Separacion Ciega de Fuentes en el Dominio Wavelet. Aplicacion a señales de audio," in *Universidad de Castilla, La Mancha*.

[32] Luz N. Oliva Moreno, "Sistemas de Procesamiento de Señales para el Análisis de Información Multidimensional," CINVESTAV, 2008.

[33] Carlos G. Puntonet, "Procedimientos y Aplicaciones en Separación de Señales (BSS-ICA)," Departamento de Arquitectura y Tecnología. Universidad de Granada, 2000.

[34] A.J. Bell and T.J. Sejnowski, "An Information-Maximization Approach to Blind Separation and

Blind Deconvolution," *Neural Computation*, vol. 7, pp. 1129-1159, 1995.

[35] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, Tercera ed.: Mc Graw Hill, 1991.

[36] E Oja and A. Hyvärinen, "A Fast Fixed-Point for Independet Component Analysis," *Neuronal Computation*, 1997.

[37] P. D'Urso and P. Giordani, "A weighted fuzzy c-means clustering model for fuzzy data," in *Computational Statistics & Data Analysis*, 2006, pp. 1496-1523.

[38] P. D'Urso and P. Giordani, "Fitting of fuzzy linear regression models with multivariate response," *International Mathematical Journal*, vol. 3, no. 6, pp. 655-664, 2003.

[39] E.E., Kessel, W.C. Gustafson, "Fuzzy clustering with a fuzzy covariance matrix," in *Proc. of the IEEE Conference on Decision and Control*, San Diego, California, 1979, pp. 761-766.

[40] W. Pedrycz J. Valente de Oliveira, *Advances in Fuzzy Clustering and its Applications*.: Ed. John Wiley & Sons, Ltd., 2007.

[41] Wikipedia. (2011, Abril) [Online]. http://es.wikipedia.org/wiki/Waveform_Audio_Format

[42] Juan Ramón Castro Rodríguez, "Tutorial Type-2 Fuzzy Logic: Theory and Applications," in *3rd International Seminar on Computational Intelligence*, 2006.

[43] Erikki Oja and Hyvärinen Aapo, "Independent Component Analysis: Alogorithms and Applications," , 2000.

[44] Frank Chung-Hoon Rhee, "Uncertain Fuzzy Clustering: Insights and Recommendations," *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, pp. 44-56, 2007.

[45] Cheul Hwang and Frank Chung-Hoon Rhee, "Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach to C-Means," *IEEE Transaction on Fuzzy Systems*, vol. 15, no. 1, pp. 107-120, 2007.

[46] Jerry M. Mendel, "Type-2 Fuzzy Sets and Systems: An Overview," *IEEE Computational Intelligence Magazine*, vol. 2, no. 1, pp. 20-29, 2007.

[47] Manuel Recuero López, "Acústica arquitectónica. Soluciones prácticas," Editorial Paraninfo, 1992.

[48] Manuel Recuero López, "Sistemas para aislamiento acústico," Brüel & Kjaer, 1988.

[49] Aapo Hyvärinen, "Survey on Independent Component Analysis," *Neural Computing Surveys*, vol. 2, pp. 94-128, 1999.

[50] Shoji Makino, Audrey Blin, Ryo Mukai, Hiroshi Sawada Shoko Araki, "Underdetermined Blind Separation for Speech in Real Enviroments with Sparseness and ICA," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, Kyoto, Japan, 2004, pp. 881-884.

[51] C. Jutten and J. Herault, "Independet Component Analysis versus Principal Component Analysis," in *Signal processing IV*, Grenoble, France, 1988, pp. 643-646.

[52] A. Mansour and M. Kawamoto, "ICA papers classified according to their applications and performances," in *IEICE Trans. Fundamentals*, vol. E86-A, N.3, March 2003.

[53] A. Mansour, A. K. Barros, and Ohnishi N., "Blind Separation of Sources: Methods, Assumptions ans Applications," in *IEICE Trans. Fundamentals*, vol. E83-A, N.8, August 2000.

[54] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing. Learning Algorithms and Applications.*: Wiley, 2002.

[55] Arun D. Kulkarni, *Computer Vision an Fuzzy-Neural Systems*.: Prentice Hall, 2001.

REFERENCIAS WEB DE SEÑALES ACÚSTICAS

A continuación se listan las principales referencias web, de donde se descargaron las señales de audio utilizadas en este trabajo para la separación de mezclas lineales y en la extracción de rasgos característicos. Los rasgos utilizados en las técnicas de minería de datos y el entrenamiento del clasificador difuso, se basan completamente en las señales obtenidas de estos sitios web.

1. http://www.freesound.org
2. http://www.findsounds.com
3. http://www.1motormart.com/sounds
4. http://www.galls.com/wav
5. http://www.wolo-mfg.com
6. http://cd.textfiles.com/maxsounds/WAV
7. http://www.horseless.com
8. http://mgcf.free.fr/sounds
9. http://www.abrivosports.com
10. http://www.ilovewavs.com/Effects
11. http://www.weblust.com/sounds
12. http://www.arbitrosdefutbol.com.ar/silbatos
13. http://www.hecticfilms.com/downloads/sound%20effects/Misc
14. http://www.karlgraf.ch
15. http://www.mediacollege.com/downloads/sound-effects
16. http://911porsche.free.fr
17. http://d21c.com/LooneyRon/sounds
18. http://homepage.powerup.com.au
19. http://sep800.mine.nu/files/sounds
20. http://www.autotech.com/media
21. http://www.corshamref.org.uk/whistle
22. http://www.ecst.csuchico.edu
23. http://www.ee.columbia.edu/~dpwe/sounds
24. http://www.sfstoledo.org/students/classes/GameDesign/September/Game_Resources/WaveSounds
25. http://www.smhc.info/geluiden
26. http://www.sucaicool.com/mid/renqun
27. http://www.tnlc.com/eep/sfx
28. http://www.worldtune.com/audio
29. http://accad.osu.edu
30. http://billor.chsh.chc.edu.tw/sound
31. http://caad.arch.ethz.ch

32. http://mustangworld.com/ourpics/sounds
33. http://new.wavlist.com/soundfx
34. http://podcasts.nytimes.com/podcasts
35. http://psychic3d.free.fr/extra_fichiers
36. http://stuweb1.gulfcoast.edu/stu34
37. http://tdwhs.nwasco.k12.or.us/staff/lewing/resources/sound
38. http://themusichouse.com/musich
39. http://webdocs.cs.ualberta.ca/~bmw/Noises
40. http://www.4uall.net/free-sound-effects/SF-free-sound-effects
41. http://www.able2products.com/siren%20sounds
42. http://www.acoustica.com/sounds
43. http://www.dro.ca
44. http://www.earthstation1.com/SFXs/SFX_Wavs
45. http://www.federalsignaldistribution.com/sounds
46. http://www.foln.org/internallinks/folmusic
47. http://www.freespecialeffects.co.uk/soundfx
48. http://www.microsoft.com/games/motocross/audio
49. http://www.payer.de
50. http://www.republic-of-loafdom.com/Media/Soundclips
51. http://www.ringelkater.de/Sounds
52. http://www.thepocket.com/wavs
53. http://www.uvm.edu/~naguiar/wavs
54. http://www.websites-graphics1.de/songs/wavs
55. http://apeyman.free.fr/sfx/humains
56. http://ch.catholic.or.kr/toykye/legio/wave
57. http://compsci.cis.uncw.edu
58. http://condor.wesleyan.edu/courses/2000f/phys192/01/sounds
59. http://cordo.free.fr
60. http://darren-john-dwyer.com/ChatRooms/Sounds
61. http://diamondridge.com/software/mailalert/sounds
62. http://elektron-bbs.dyndns.org/files/multimed/wav
63. http://ftp.iinet.net.au/pub/games/serverdownload/hlds/cstrike/sound
64. http://ftp.tux.org/pub/X-Windows/games/freeciv/incoming/sounds
65. http://gibsonexhaust.com/sounds/mp3
66. http://home.hccnet.nl
67. http://humlog.homestead.com/~site/sounds/sound_effects
68. http://jsf3.homestead.com/files
69. http://k.domaindlx.com/piracykings/tones
70. http://korea119.pe.kr/midi

ANEXO A. CONCEPTOS BÁSICOS DE PROCESOS ALEATORIOS

Teorema del límite central

El Teorema del Límite Central o Teorema Central del Límite indica que, bajo condiciones muy generales, la distribución de la suma de variables aleatorias tiende a una distribución gaussiana cuando la cantidad de variables es muy grande. Existen diferentes versiones del teorema, en función de las condiciones utilizadas para asegurar la convergencia. Una de las más simples establece que es suficiente que las variables que se suman sean independientes, idénticamente distribuidas, con valor esperado y varianza finitas.

Una muestra no tiene que ser muy grande para que la distribución de muestreo de la media se acerque a la normal. Los estadísticos utilizan la distribución normal como una aproximación a la distribución de muestreo siempre que el tamaño de la muestra sea al menos de 30, pero la distribución de muestreo de la media puede ser casi normal con muestras incluso de la mitad de ese tamaño. La importancia del teorema del límite central es que nos permite usar estadísticas de muestra para hacer inferencias con respecto a los parámetros de población sin saber nada sobre la forma de la distribución de frecuencias de esa población más que lo que podamos obtener de la muestra.

Medidas de no-gaussianidad

Curtosis

La curtosis se define como el momento de cuarto orden de datos aleatorios. Dados algunos datos aleatorios y, la curtosis denotada como $kurt(y)$, se expresa de la siguiente manera:

$$Kurt(y) = E\{y^4\} - 3(E\{y^2\})^2$$

Si la variable aleatoria y tiene promedio cero y varianza unitaria $E\{y^2\} = 1$, entonces la ecuación anterior se convierte en:

$$Kurt(y) = E\{y^4\} - 3$$

La curtosis de una variable gaussiana es 0, mientras la curtosis de una variable no-gaussiana es diferente de cero. Debido a esto, la curtosis puede ser usada como una medida de no-gausianidad, dando como resultado tanto valores positivos como negativos. Variables con valores positivos de curtosis son llamadas supergausianas, y las variables con valores negativos se conocen como subgausianas. La función de densidad de probabilidad (f.d.p) para datos

supergausianos y subgausianos tiene diferentes características. La f.d.p para variables supergausianas es alargada, y la f.d.p para datos subgausianos tiene una forma plana. La matriz de mezcla W se elige de forma tal que $Kurt(W^T x)$ entregue un extremo local o componente independiente. [11]

Negentropía

Otro método para medir la no-gausianidad es la negentropía. La negentropía está basada en la entropía diferencial, que es una medida de la cantidad de información. Mientras más aleatorios e impredecibles son los datos, la entropía de estos es mayor. La entropía H de una variable aleatoria y con densidad $p(\eta)$ es

$$H(y) = -\int p_y(\eta) log\, p_y(\eta) dp_y(\eta)$$

Una variable gaussiana tiene un valor de entropía grande para todas las variables aleatoria con varianza unitaria. La negentropia J, es:

$$J(y) = H(y_{gauss}) - H(y)$$

y_{gauss} es una variable aleatoria gaussiana con la misma correlación y covarianza de y. Como la negentropía esta normalizadas, esta siempre es no-negativa y es cero si la distribución es gaussiana. Este método utiliza momentos de alto orden. La negentropía de y se convierte en:

$$J(y) = \frac{1}{12} E\{y^3\}^2 - \frac{1}{48} kurt(y)^2$$

Otra expresión para calcula la negentropia es basada en la expectación matemática, una función no cuadrática G remplaza los polinomios y^3 y y^4. En este caso la *negentropia* es:

$$J(y) \propto [G\{y\} - G\{v\}]^2$$

Donde v es una variable gaussiana con varianza unitaria y promedio cero. La clave es escoger la función G correctamente, las elecciones frecuentes para G son:

$$G_1(y) = 1/a_1 \log\cosh a_1 y \text{ donde } 1 \leq a_1 \leq 2$$

$$G_2(y) = -\exp(-y^2/2)$$

$$G_3(y) = 1/4\, y^4$$

Ejemplo de este método es el algoritmo FastICA.

ANEXO B. CARACTERÍSTICAS DE MICROFONOS INDUSTRIALES

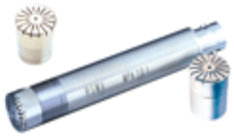


Fig. B.1 Micrófono MP201

Diameter	1/2"
Standards (IEC61672)	Class I
Microphone	MP201
Optimized	Free Field
Preamplifier	MA231(TEDS optional)
Frequency Response (Hz)	20 ~ 20k
Open-circuit Sensitivity (mV/Pa) (±2dB)	45
Output Impedance (Ω)	< 50
Dynamic Range (dBA)	16 ~ 134
Inherent Noise (dBA)	< 16
Operating Temperature (°C)	-30 ~ 80
Operating Humidity (RH)	0 ~ 95%
Temperature Coefficient (dB/°C)	0.005
Humidity Coefficient (dB/%RH)	0.003
Pressure Coefficient (250 Hz) (dB/kPa)	-0.004
Length (mm)	91
Input Connector	BNC
Corresponding Model with TEDS	MPA261