

Universidad de las Ciencias Informáticas Facultad 9



Título:

Separación Automática de Fuentes de Sonido.

Trabajo de Diploma para Optar por el Título de Ingeniero en Ciencias Informática

Autor:

Yanet García García

Tutor:

Ing. Solangel Rodríguez Vázquez

Co-Tutor:

MSc. Yaneisis Pérez Heredia

Ciudad de la Habana

Junio 2010

DECLARACIÓN DE AUTORÍA

Declaro que soy la única autora de este trabajo y autorizo a la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmo la presente a los ____ días del mes de _____ del año _____.

Yanet García García

Autora

Ing. Solangel Rodríguez Vázquez

Tutora

MSc. Yaneisis Pérez Heredia

Tutora

DEDICATORIA

Le dedico este logro ante todo a la persona más importante de mi vida, porque lo hicimos entre las dos, por todo el amor que me ha dado. Siempre serás mi Ángel mami.

A mi papá que siempre me ha amado y enseñado a caminar por el camino correcto, por depositar toda su confianza en que yo iba a ser alguien en la vida.

A mi hermana, que en todo momento ha estado apoyándome, por confiar siempre en mí y asentarme como ejemplo.

A una persona muy especial que me ha dado su cariño, su apoyo, su ayuda en los buenos y en los malos momentos; y sobre todas las cosas nunca ha dejado de confiar en mí, mi novio Yordan.

Resumen

La separación de fuentes de sonido se refiere a la tarea de poder obtener las señales producidas por cada una de las fuentes de sonido de una mezcla compleja acústica. Por tal motivo este ha sido un tema de intensa investigación en los últimos años debido a las múltiples aplicaciones que tiene dentro del tratamiento de señales, teniendo sus orígenes en el procesamiento de voz como respuesta al efecto “Cocktail Party”. Efecto que consiste en la habilidad que tiene el ser humano de poder centrar su atención y escuchar a una sola persona, aun habiendo más personas hablando a la misma vez o existiendo otras fuentes sonoras.

El procesamiento de señales digitales es una de las líneas de trabajo que se desarrolla dentro de la Universidad de las Ciencias Informáticas (UCI), donde precisamente la separación de fuentes de sonido permitiría un mejor despliegue dentro de los productos que aquí se crean.

En el presente trabajo se lleva a cabo un estudio bibliográfico sobre las técnicas que permiten la separación automática de fuentes de sonido, donde las fuentes que se desean separar no se conocen de antemano. Para ello la separación se realiza mediante la utilización de las propiedades comunes de las fuentes de sonido del mundo real y sus principales características. Los planteamientos que aquí se describen para la separación están dados por el uso del aprendizaje no supervisado y del modelo sinusoidal. El resultado obtenido está dado por la propuesta de la técnica no supervisada utilizando PCA y NMF y la técnica basada en el modelo sinusoidal.

Palabras claves: Separación de Fuentes de Sonido, Técnica, Sonido

Índice

| | |
|--|----|
| Introducción | 1 |
| Capítulo 1: Fundamentación Teórica | 5 |
| 1.1 Introducción..... | 5 |
| 1.2 Conceptos asociados al dominio del problema..... | 5 |
| 1.2.1 Técnica..... | 5 |
| 1.2.2 Sonido..... | 5 |
| 1.2.3 Sonido Analógico..... | 6 |
| 1.2.4 Sonido Digital..... | 6 |
| 1.2.5 Fuente de Sonido..... | 7 |
| 1.3 Precedentes Históricos..... | 7 |
| 1.4.1 Separación Ciega de Fuentes (BSS)..... | 9 |
| 1.4.2 Modelos de la separación ciega de fuentes..... | 10 |
| 1.4.2.1 Modelo lineal:..... | 10 |
| 1.4.2.2 Modelo no lineal:..... | 13 |
| 1.5 Conclusiones..... | 14 |
| Capítulo 2: Solución Propuesta | 15 |
| 3.1 Introducción..... | 15 |
| 3.2 Técnicas basadas en el aprendizaje no supervisado para la separación de fuentes..... | 15 |
| 2.2.1 Análisis de Componentes Principales (PCA)..... | 17 |
| 3.2.1.1 Cálculos de los componentes principales..... | 17 |
| 3.2.1.2 Proceso de extracción de factores..... | 18 |
| 2.2.2 Análisis de Componentes Independientes (ICA)..... | 23 |
| 2.2.2.1 Hipótesis para el Análisis de Componentes Independientes..... | 23 |
| 2.2.2.2 Definición y propiedades fundamentales de Independencia..... | 24 |
| 2.2.2.3 Variables no correlacionadas son sólo parcialmente independientes..... | 25 |
| 2.2.2.4 ¿Por qué están prohibidas las variables de Gauss?..... | 26 |
| 2.2.2.5 Reducción de la información mutua..... | 26 |
| 2.2.2.6 Indeterminaciones del ICA..... | 27 |

| | | |
|---------|---|----|
| 2.2.3 | Factorización no Negativa de Matrices (NMF)..... | 28 |
| 2.2.3.1 | Definición de la función de coste. | 29 |
| 2.2.3.2 | Reglas de actualización multiplicativas. | 29 |
| 2.2.3.3 | Pruebas de convergencia. | 31 |
| 2.2.4 | Comparación entre las técnicas PCA, ICA y NMF..... | 33 |
| 2.2.5 | Separación no supervisada utilizando PCA y NMF..... | 35 |
| 2.2.5.1 | Extracción de las bases auditivas para PCA..... | 36 |
| 2.2.5.2 | Extracción de bases auditivas mediante NMF. | 37 |
| 3.3 | Técnica de separación basada en el modelo sinusoidal..... | 38 |
| 3.3.1 | Modelo de la señal..... | 38 |
| 3.3.2 | Resolución de superposición de sobre-tonos..... | 39 |
| 2.3.3 | Método basado en el Modelado Sinusoidal. | 40 |
| 2.3.4 | Formulación en el dominio de frecuencia. | 41 |
| 2.3.5 | Fase de estimación..... | 42 |
| 2.3.6 | Estimación de la amplitud. | 44 |
| 2.3.6.1 | Solución de los componentes que se superponen por mínimos cuadrados. | 44 |
| 2.3.7 | Estimación de frecuencias. | 46 |
| 2.3.8 | Combinar las frecuencias fundamentales en notas musicales. | 48 |
| 2.3.9 | Tratamiento no lineal por mínimos cuadrados..... | 49 |
| 2.3.9.1 | Algoritmos..... | 49 |
| 2.4 | Principios matemáticos..... | 50 |
| 2.4.1 | Definición de variables aleatorias. | 50 |
| 2.4.2 | Función de distribución acumulada. | 50 |
| 2.4.3 | Función de densidad de probabilidad..... | 51 |
| 2.4.4 | Vectores aleatorios. | 52 |
| 2.4.5 | Funciones de Distribución y densidad de vectores Aleatorios. | 52 |
| 2.4.6 | Función de Distribución y Densidad conjunta y marginales..... | 53 |
| 2.4.7 | Esperanzas..... | 53 |
| 2.4.8 | Distribuciones de Probabilidad Importantes..... | 55 |
| 2.4.8.1 | Distribución Uniforme. | 55 |

| | |
|---|-----------|
| 2.4.8.2 Distribución normal..... | 56 |
| 2.4.8.3 Distribución de Laplace. | 57 |
| 2.4.9 Correlación y Covarianza Cruzadas..... | 57 |
| 2.4.10 Estimación de Funciones Estadísticas..... | 58 |
| 2.4.11 Decorrelación y Blanqueado. | 59 |
| Capítulo 3: Validación de la Propuesta..... | 61 |
| 3.1 Introducción..... | 61 |
| 3.2 Método para la validación de la propuesta. | 61 |
| 3.3 Indicaciones para la implementación. | 66 |
| 3.4 Conclusiones..... | 67 |
| Conclusiones..... | 68 |
| Recomendaciones..... | 69 |
| Referencias Bibliográficas..... | 70 |

Introducción

Los seres humanos presentan la capacidad auditiva para identificar y separar señales en ambientes donde múltiples fuentes de sonido se mezclan con mayor o menor nivel. Esto se puede ver en una fiesta donde intervienen distintas fuentes de sonido, cada uno de los hablantes, la música propia y el sentir de los pasos de baile. Sin embargo cuando se sostiene una conversación, por ejemplo se nota como al concentrarse lo suficiente en este sonido que proviene de la persona que interesa, se puede conseguir en poco tiempo que su sonido se destaque sobre el resto del círculo sonoro en el que se encuentra. La capacidad del ser humano de seleccionar una determinada señal o sonido en un ambiente ruidoso es conocida como el Problema de las Fiestas Ruidosas (*Cocktail Party Effect*) en referencia a la facultad del cerebro del hombre de centrarse en una única voz e ignorar a otras voces de interlocutores y/o sonidos que se producen simultáneamente con similar amplitud en un entorno ruidoso como ocurre en las fiestas[1]. Aunque los seres humanos presentan la capacidad de extraer perceptualmente un sonido de una mezcla, los sistemas automáticos para realizar una escucha discriminada constituye un gran reto en la disciplina del aprendizaje automático. La separación ciega de fuentes (*Blind Source Separation, BSS*)[2] es una potente técnica capaz de solucionar esto en el mundo del procesamiento de señales y ha sido durante los últimos años y en la actualidad una de las líneas que más interés despierta en este ámbito.

Se entiende por técnica de Separación Ciega de Fuentes (*Blind Separation of Sources, BSS*) a aquellas cuyo objetivo es estimar las señales producidas por diferentes fuentes a partir de una o varias señales de mezcla de todas ellas. Esta técnica consiste en la recuperación de fuentes no observadas a partir de diferentes ‘mezclas’ de señales observadas. Suponiendo que se conversa frente a un par de micrófonos. Salvo que se hable por turnos, los micrófonos recogerán la superposición¹ de las voces. Tomar las grabaciones y tratar de separar las voces es, a grandes rasgos lo que se conoce como problema de “Separación Ciega de Fuentes” donde el término (ciego) se debe a que no se conoce la manera en que son mezcladas las fuentes. En muchas situaciones es necesario extraer una o más señales originales (fuentes) teniendo como único dato de referencia la mezcla recogida por uno o más sensores [4].

¹ Superposición de ondas armónicas: Cuando dos ondas se encuentran en un punto o una región del espacio, trae como resultado una nueva onda armónica cuya amplitud depende de la diferencia de fase entre las dos ondas originales. 3. Autores, C.d. *Superposición de ondas*. 2003 [citado 13/02/2010]; Disponible en : <http://www.ehu.es/acustica/index.html>.

Este problema inicialmente fue planteado en el año 1985 por los profesores Jutten¹ y Héroult² [7], como una aplicación a la neurofisiología. Empleando un modelo, simplificándose un movimiento en la contracción de un músculo, se cuantifica dicha contracción mediante dos sensores (mezcla).

Actualmente, muchos proyectos de investigación se han enfocado en buscar soluciones a este problema por las importantes aplicaciones que tiene hoy en día. La separación de fuentes de ruido de tráfico interurbano tras la aplicación de sustracción espectral de ruido [8]. También en el campo de la biomédica al realizar un electrocardiograma a una mujer embarazada, es necesario separar las señales generadas por la madre de las generadas por el feto. En Telecomunicaciones la separación de señales es de gran utilidad en los procesos de eliminación de ruido e interferencias, tanto en imágenes como en sonido. La separación de fuentes de sonidos submarinos podría distinguir las señales recogidas por un sonar provenientes de un banco de peces de las que provienen del fondo del mar. En definitiva, cualquier situación en la que no se tenga acceso a la información original, sino a una mezcla de varias señales y ruido es susceptible de ser resuelta mediante técnicas de Separación Ciega de Señales.[1]

En el Departamento de Procesamiento de Imágenes y Señales Digitales de la facultad 9 perteneciente a la Universidad de las Ciencias Informáticas (UCI) se trabaja en el desarrollo de software en el ámbito de procesamiento de audio, y con el objetivo de ampliar sus funcionalidades se piensa insertar técnicas de separación de fuentes de sonido por las múltiples aplicaciones que esta tiene; tales como en la búsqueda por contenido en base de datos multimedia. También para la extracción de voz de una mezcla ruidosa o la voz de un cantante de una mezcla musical podría ser interesante no sólo para propósito de remezcla, sino para la transcripción automática de letras de canciones o melodías, reconocimiento de artistas y otras tareas de recuperación musical.

¹ Christian Jutten. Recibió el grado de doctorado en 1981 y el grado de Doctor en Ciencias en 1987 en el Instituto Nacional Politécnico de Grenoble (Francia). Ha sido Editor Asociado de la IEEE. Durante 20 años, sus Intereses de investigación son la separación de fuentes y análisis de componentes independientes y el aprendizaje en redes neuronales. 5. Autores, C.d., *Christian Jutten*. Hindawi Publishing Corporation, 2010.

² Jeanny Héroult es profesor emérito de la Universidad Joseph Fourier, Grenoble. En el campo de investigación se ha preocupado con el modelado de Recursos Naturales y Redes Neuronales Artificiales, a distintos niveles: desde la membrana celular a las grandes redes de adaptación. Es revisor de revistas internacionales en redes neuronales, Procesamiento de Señales e Imágenes 6.IEEE, T.I., *The IEEE Workshop on Neural Networks for Signal Processing*. September 2-4 2009: Grenoble, France..

Esta situación genera la necesidad de llevar a cabo una investigación para darle respuesta al siguiente **problema científico**: ¿Qué técnica es la adecuada para gestionar la separación de fuentes de sonido para el procesamiento de medias en el Departamento de Señales Digitales de la facultad 9?

Para la solución de este problema se plantea el siguiente **objetivo**: Proponer una técnica que permita la separación automática de fuentes de sonido para el procesamiento de medias en el Departamento de Señales Digitales de la facultad 9.

Objeto de estudio: Separación automática de fuentes de sonido.

Campo de acción: Técnicas que permitan la separación automática de fuentes de sonido para el procesamiento de medias en el Departamento de Señales Digitales de la facultad 9.

La idea a defender en esta investigación: La propuesta de una técnica de separación automática de fuentes de sonido para el procesamiento de medias permitirá la futura implementación de dicha técnica ampliando las funcionalidades que desarrollan en el Departamento de Señales Digitales de la facultad 9.

Para dar solución a nuestro problema se proponen un grupo **de Tareas de investigación** a desarrollar como son:

- Describir el estado del arte de las posibles técnicas a utilizar para la separación automática de fuentes de sonido.
- Identificar las posibles técnicas para la separación automática de fuentes de sonido.
- Analizar las técnicas de separación automática de fuentes de sonido.
- Proponer la técnica de separación automática de fuentes de sonido.
- Valorar la factibilidad y aportes de las mejoras propuestas.

Para realizar las tareas enunciadas se emplearon los siguientes Métodos de Investigación Científica.

Métodos Teóricos:

- **Histórico Lógico**: Permite estudiar de forma analítica la trayectoria histórica real de los fenómenos, su evolución y desarrollo. El método permitió realizar la primera parte de la investigación, al hacer un análisis bibliográfico del tema, para su posterior estudio y determinar a través de la evaluación

de la bibliografía conceptos de la temática, que permiten conocer el estado de la evolución actual del fenómeno.

- **Analítico Sintético:** Permite la división del fenómeno en sus múltiples relaciones y componente lo que facilita su estudio. El método arrojó como resultado elementos importantes que forman parte del objeto de estudio de la investigación, después de analizado todos los documentos, teorías y fuentes consultadas.
- **Sistémico:** Permitió estudiar y determinar cada una de las partes que integran la investigación, unir los datos, definir cada elemento que está estrechamente relacionado con otro y darle solución al problema.

Métodos Empíricos

- **Entrevista:** Permitió ampliar más la información para la investigación, constituyó un medio de conocimiento para poner en práctica una información clara y detallada del objeto de estudio planteado.

En sentido general con este trabajo se espera obtener como resultado la técnica que permita el sonido por sus fuentes.

Capítulo 1: Fundamentación Teórica

1.1 Introducción.

En el presente capítulo se realiza un estudio del estado del arte asociado a las técnicas existentes en el ámbito de la separación de sonido. En una primera sección se tratan los conceptos asociados al dominio del tema en cuestión. Posteriormente se realiza un acercamiento histórico sobre los inicios de los estudios alrededor de la separación ciega de señales donde se destacan las principales figuras en el desarrollo investigativo en esta área. Acerca de la separación ciega de señales se describen sus características, se describen los modelos de separación, es decir, el modelo de mezcla lineal y el modelo de mezcla no lineal que constituyen un punto de partida para lograr un adecuado análisis de las técnicas, así como una muestra de las posibles aplicaciones que tiene esta temática en diversos campos donde se incluye el referente al procesamiento de sonido.

1.2 Conceptos asociados al dominio del problema.

1.2.1 Técnica.

La palabra técnica proviene del griego *téchne*, que se ha traducido como “arte” o “ciencia”. Una técnica es un procedimiento que tiene como objetivo la obtención de un resultado determinado, ya sea en la ciencia, en la tecnología, en el arte o en cualquier otro campo. En otras palabras, una técnica es un conjunto de reglas, normas o protocolos que se utiliza como medio para llegar a un cierto fin. Usualmente, la técnica requiere del uso de herramientas y conocimientos muy variados, que pueden ser tanto físicos como intelectuales.[9]

Otros autores definen a la técnica como un procedimiento o un conjunto de procedimientos prácticos, en vistas al logro de un resultado, o a varios resultados concretos, valiéndose de herramientas o instrumentos, y utilizando el método inductivo y/o analógico, en cualquier campo del saber o del accionar humano. Técnica es el modo y los medios empleados para llegar al fin propuesto.[10]

1.2.2 Sonido.

La Real Academia Española define el sonido como:

- Sensación producida en el órgano del oído por el movimiento vibratorio de los cuerpos, transmitido

por un medio elástico, como el aire.

- Significación y valor literal que tienen en sí las palabras.

El sonido es la interpretación que hace el cerebro de las variaciones de presión que genera un objeto vibrante en un medio. Para que esta vibración pueda ser escuchada por los humanos el objeto debe oscilar aproximadamente entre 20 y 20 000 veces por segundo debido a que las frecuencias de ondas que somos capaces de escuchar se encuentran entre 20 Hertz (20 Hz) a 20 KHz.[11]

1.2.3 Sonido Analógico.

El sonido analógico es aquél que se almacena, procesa y reproduce gracias a circuitos electrónicos y otros dispositivos de carácter analógico como son la cinta magnética (en casete o en bobina) o el disco de vinilo.

La tecnología aplicada al procesamiento del sonido analógico funciona de la siguiente manera: la onda sonora produce una vibración en el aire que es captada por un micrófono. Este convierte la vibración en una señal eléctrica que viaja por un cable hasta el aparato grabador. La grabación se produce de forma magnética (en cinta de casete o bobina), mediante unas cabezas grabadoras/reproductoras del sonido que hacen contacto físico con la superficie de la cinta. Esto conlleva un desgaste del material con el uso.

1.2.4 Sonido Digital.

Pero en computadoras, en el mundo informático, no se trabaja con el sonido de forma analógica sino con datos digitales. El sonido de forma digital, es más complejo que el analógico, pues no se almacena mediante oscilaciones de ondas sino convertido en ceros y unos, lenguaje digital. El sonido digital se obtiene de dos procesos: el muestreo y la cuantificación digital.

La frecuencia de muestreo, quiere decir el espacio de tiempo que se va a estar tomando la muestra de la señal. La tasa de muestreo se va a mostrar en Hercios, son las muestras ó pulsos por segundos. Para el espectro audible (20 a 20 000 Hz) la tasa de muestreo debe ser algo más de 40 000 Hz. Por ejemplo los CDs de audio se reproducen hasta 44,1kHz. Esto es planteado por la Ley de Nyquist¹ (Si la tasa de

¹ Harry Nyquist Ingeniero estadounidense de origen sueco. Especializado en cibernética. Durante sus 37 años de servicio con el Sistema Bell recibió 138 patentes y publicó 12 artículos técnicos. Entre sus principales trabajos está la descripción del (ruido de Johnson-Nyquist) y la Ley de Nyquist. 13. Autores, C.d., *Harry Nyquist*.

muestreo es dos veces de la frecuencia más alta que contiene el sonido, se puede muestrear el audio sin pérdida de calidad).

La cuantificación o también conocido por Bit de profundidad representan el número de niveles que habrá en cada espacio de tiempo. La conversión del nivel de muestras fijadas en el proceso de muestreo, en un valor entero de rango finito y predeterminado. [12]

1.2.5 Fuente de Sonido.

Se entiende por "fuentes de sonido" como aquellos aparatos a cuya salida se obtiene una señal eléctrica, de amplitud directamente proporcional a la amplitud de la señal de audio. Esta señal eléctrica sólo ha de ser amplificada convenientemente y conectada a unos altavoces para poder ser escuchada.[14]

El sonido es un fenómeno vibratorio transmitido en forma de ondas. Para que se genere un sonido es necesario que vibre alguna fuente.[15]

Teniendo en cuenta lo planteado anteriormente se concluye que una fuente de sonido es aquel objeto material vibrante que transmite sonido en forma de ondas, a través del aire.

1.3 Precedentes Históricos.

El Problema de las Fiestas Ruidosas (Cocktail Party Problem) es uno de los problemas que se presentan con más frecuencia en el mundo de del procesamiento de sonido. La separación ciega de fuentes (Blind Source Separation, BSS)[2], es una potente técnica capaz de solucionar este problema en el mundo de las comunicaciones.

El problema de la separación ciega de señales consiste en la recuperación de las señales originales a partir de las mezclas detectadas por sensores (un conjunto de N fuentes), conociendo tan sólo estas últimas. El término de Build (ciego) proviene de desconocer tanto la fuente como el mecanismo que las mezcla. Estas mezclas de señales tienen lugar en el medio en que se propagan y en los sensores, y como característica de este método es que no se disponen de datos a priori, no se cuenta con ninguna información de las señales originales ni de la forma en que fueron mezcladas.

El primer artículo de la separación ciega de fuentes fue publicado a mediados de la década de los 80 (1985-1986) por Christian Jutten y Jeanny Héroult profesores de la Universidad de Grenoble, en las actas del congreso sobre "Redes Neuronales y Computación" celebrado en Snowbird(EE.UU). En él se presentaba que habían conseguido separar fuentes independientes mediante el uso de redes neuronales

de realimentación que usaban como algoritmo de aprendizaje el algoritmo de Hebb [7], como una aplicación a la neurofisiología. Empleando un modelo simplificándose un movimiento en la contracción de un músculo, se cuantifica dicha contracción mediante dos sensores (mezcla). Es a partir de estas señales obtenidas que se plantean obtener las fuentes, determinadas como la posición angular y la velocidad de movimiento de una partícula. En palabras del propio Jutten, la motivación para su trabajo habría nacido tras una discusión informal, de cafetería sobre mecanismos que tiene el cerebro para extraer información útil de entre la amalgama de estímulos que recibe. De esta forma resurgió un período de intensa investigación en la separación ciega de fuentes. En 1994 Comon¹ propuso funciones de energía relacionadas con la minimización de la información mutua con lo que dio a conocer de forma general el análisis de componentes independientes, término que fue usado por primera vez por este autor.

En el campo de las comunicaciones, la separación de fuentes viene como consecuencia del uso del mismo medio físico por parte de distintas señales siendo la situación crítica cuando el canal presenta ruido de amplitud parecida a las señales de interés siendo las técnicas de separación de señales básicas en el pre-procesado del conjunto, un ejemplo se encuentra en trabajos de Cardoso² y Comon en el 1996 [18]. Además en la separación de hablantes, interesaría reproducir el problema de las fiestas ruidosas.

Una de las aplicaciones de más importancia es la eliminación de redundancia en conjuntos grandes de datos, esta se resuelve mediante una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables), fenómeno que es capaz de reproducir también y que ha sido estudiado por diversos autores hasta la fecha mediante el análisis de componentes principales.

¹ Pierre Comon alcanzó su Doctorado en 1985, en la Universidad de Grenoble, Francia. Más tarde recibió la habilitación para conducir Investigaciones "en 1995, de la Universidad de Niza Sophia-Antipolis (UNS). Su interés investigativo ha sido principalmente en el área de procesamiento de señales. Actualmente es director de la escuela doctoral Information and Communication Sciences and Technologies (STIC), y también preside el departamento de Signal Images Systèmes (SIS), desde enero 2008.16. P.Comon (2009) *Pierre COMON*. Universié Nice

² Jean-François Cardoso es investigador del Departamento de Procesamiento de Imágenes y Señales de Telecom, París. Director de Investigación de CNRS, pionero de los métodos de procesamiento de señales estadística conocida como "separación de fuentes" o "análisis de componentes independientes". 17. Autores, C.d., *Jean-François Cardoso* Hindawi Publishing Corporation, 2010.

1.4 Fundamentos de la Separación de Fuentes de Sonido.

1.4.1 Separación Ciega de Fuentes (BSS).

El problema de la separación ciega de fuentes se plantea en los siguientes términos: un determinado proceso combina, superpone o genéricamente “mezcla” señales a las que se les conoce como las “fuentes”. Las señales mezcladas reciben el nombre de “observaciones”. El adjetivo “ciega” enfatiza que no se sabe como estas señales se han mezclado[22].

Para una mejor explicación se parte de la suposición de que se está en una sala de concierto escuchando a una orquesta de música sinfónica. Al concentrarse lo suficiente en el sonido que proviene de un determinado instrumento, se puede conseguir en poco tiempo que su sonido se destaque sobre el resto de la orquesta independientemente de todos los instrumentos que estén sonando armoniosamente. De alguna forma el cerebro está aplicando una técnica de separación en la que es posible concebir cada instrumento como una fuente independiente[23].

Suponiendo ahora que se colocan una cantidad M de micrófonos distribuidos por toda la sala de conciertos. Cada uno de ellos permitirá obtener una grabación de la orquesta al completo, que se notarán $\mathbf{X}_i(\mathbf{t})$ y que serán las observaciones, donde cada \mathbf{X}_i representa una amplitud y \mathbf{t} representa el tiempo. Por otro lado, suponiendo que la orquesta está compuesta por N instrumentos, cada uno de ellos se considerarán como una fuente y llamadas como $\mathbf{S}_j(\mathbf{t})$.

De esta forma las señales de cada micrófono (*observaciones*) son una suma ponderada del sonido emitido por cada instrumento de la orquesta (*fuentes*) por lo que es posible expresar las señales registradas por los micrófonos como:

$$\begin{aligned}x_1(t) &= a_{11}s_1 + a_{12}s_2 + \dots + a_{1N}s_N \\x_2(t) &= a_{21}s_1 + a_{22}s_2 + \dots + a_{2N}s_N \\&\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\x_M(t) &= a_{M1}s_1 + a_{M2}s_2 + \dots + a_{MN}s_N\end{aligned}\tag{1}$$

Donde $\mathbf{a}_{11}, \dots, \mathbf{a}_{MN}$ son unos parámetros que pueden considerarse dependientes de la distancia de cada micrófono a los instrumentos de manera que si se puede estimar las fuentes originales $\mathbf{S}_1(\mathbf{t}), \mathbf{S}_2(\mathbf{t}),$

... $S_n(t)$, e incluso la forma en que han sido mezcladas usando solamente las señales registradas $X_1(t)$, $X_2(t)$, ... $X_m(t)$. Hablando en términos de acústica esto se denomina el problema de las fiestas ruidosas (*Cocktail Party Effect*), que se llamará también el problema de la separación ciega de fuentes. Como por ejemplo, considerando las formas de onda procedentes de cuatro instrumentos musicales diferentes.

Aplicando el modelo más simplificado, se podrían haber obtenido cuatro grabaciones de cuatro micrófonos, mediante un proceso de mezcla lineal e instantánea de fuentes. El problema de la separación ciega de fuentes consiste en recuperar la información de los cuatro instrumentos musicales usando solamente la información de las observaciones. En realidad de conocer los parámetros a_{ij} se podría resolver la ecuación lineal (1) empleando métodos clásicos de álgebra lineal. Sin embargo, el problema es mucho más complejo, pues se desconocen a_{ij} . Un primer caso en el camino de resolver este problema podría pasar por ejemplo de algún tipo de información sobre las propiedades estadísticas de las fuentes $S_j(t)$, para estimar los parámetros a_{ij} . Realmente, puede resultar suficiente con realizar la suposición de que las fuentes $S_1(t_1)$, $S_2(t_2)$,... $S_n(t_n)$ sean *estadísticamente independientes* para cada instante de tiempo t y que el sistema de mezcla sea lineal, invariante en el tiempo e invertible.

De manera que las cuatro señales estimadas mediante un método de BSS poseen un parecido enorme con las fuentes originales y, aunque su signo esté invertido o su amplitud no sea la original, esto no suele ser un problema excesivamente importante.

1.4.2 Modelos de la separación ciega de fuentes.

Un modelo de separación ciega de fuentes, está condicionado por las consideraciones que se le realicen a la mezcla de las fuentes. Atendiendo al modo en que son mezcladas las fuentes aparecen dos modelos de separación ciega de fuentes.

1.4.2.1 Modelo lineal:

Modelo de mezcla lineal instantáneo

El modelo de mezcla instantánea y lineal (o llamado también lineal sin memoria) es el más sencillo. Este modelo supone que el valor de las observaciones $X_i(t)$ en un instante dado es una función lineal de los valores de fuentes originales $S_i(t)$ en ese mismo instante de tiempo y no influyen en la mezcla en ningún caso de los valores de las fuentes en instantes de tiempo anteriores. Una función lineal sobre un conjunto

de variables se reduce a un conjunto de coeficientes reales que multiplican a los datos originales. De esta forma, se puede describir este modelo matemáticamente mediante la siguiente expresión:

$$x_i(t) = \sum_{j=1}^n a_{ij} s_j(t), \quad i = 1, \dots, n \quad (2)$$

o de forma resumida, utilizando vectores de variables:

$$\mathbf{x}(t) = \underline{\mathbf{A}}\mathbf{s}(t), \quad \underline{\mathbf{A}} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{1n} & \dots & a_{nn} \end{bmatrix}. \quad (2.1)$$

La matriz $\underline{\mathbf{A}}$ se denomina matriz de mezcla, y es claramente la encargada de realizar la combinación lineal de las fuentes \mathbf{S} . Al asumir que es el número de observaciones o de mezclas es igual al de fuente, la matriz de mezcla $\underline{\mathbf{A}}$ es cuadrada tal y como se muestra en la expresión (2.1).

Por tanto el problema de la separación ciega de fuentes lineal e instantáneamente modelado, se podría resolver si se tuviera una matriz $\underline{\mathbf{W}}$ llamada matriz de separación, cuya inversa es similar a la matriz de mezcla $\underline{\mathbf{A}}$ y sus mismas dimensiones, con lo que, multiplicándola por las señales mezcladas se podría obtener una reconstrucción o estimación $\mathbf{y}(t)$ de las fuentes originales. Naturalmente, es necesario que la matriz de separación $\underline{\mathbf{W}}$ sea no singular. De esta forma, la ecuación del modelo para la separación es:

$$\mathbf{y}(t) = \underline{\mathbf{W}}\mathbf{x}(t) \quad (2.2)$$

Modelo de mezcla lineal convolutivo.

En el modelo de mezcla convolutiva, llamada también mezcla con memoria o con retardos, se supone que el valor de las observaciones ($\text{vec } \mathbf{X}_i(\mathbf{t})$) en un instante dado es un vector de las funciones y de valores de las fuentes originales $\mathbf{S}_i(\mathbf{t})$ en ese mismo instante de tiempo anteriores. Matemáticamente, se podría representar este modelo en el caso general (lineal y no lineal) mediante una serie de funciones como sigue:

$$\mathbf{x}(t) = F(\mathbf{s}(t), \mathbf{s}(t-1), \mathbf{s}(t-2), \dots, \mathbf{s}(t-n)) \quad (2.3)$$

Esta ecuación trata de modelar fenómenos reales como ecos en las comunicaciones y reverberaciones de las señales. El modelo de mezcla convolutiva considera que el medio de propagación actúa como un filtro con \mathbf{p} entradas (las fuentes) y \mathbf{q} salidas (las señales detectadas por los sensores). Usualmente se admite que este filtrado introducido por el medio y los sensores, es lineal y estacionario. De esta manera las señales captadas pueden expresarse de la forma:

$$x_i(t) = \sum_{j=1}^p \int a_{ij}(t-\tau) s_j(\tau) d\tau + n_i(t) \quad \forall i = 1, \dots, q \quad (2.4)$$

Donde $\mathbf{a}_{ij}(\mathbf{t})$ representa la respuesta a un impulso de un filtro que modela la propagación entre la fuente \mathbf{i} y el sensor \mathbf{j} . El término $\mathbf{n}_i(\mathbf{t})$ corresponde a un posible ruido aditivo, que no siempre tiene por qué estar presente.

y el sensor \mathbf{j} . El término $\mathbf{n}_i(\mathbf{t})$ corresponde a un posible ruido aditivo, que no siempre tiene por qué estar presente.

En el dominio de la frecuencia, ν , la ecuación (2.3) puede escribirse como:

$$x_i(\nu) = \sum_{j=1}^p A_{ij} s_j(\nu) + n_i(\nu) \quad \forall i = 1, \dots, q \quad (2.5)$$

donde la matriz $\mathbf{A}(\nu)$, de dimensión $\mathbf{p}^* \mathbf{q}$, que contiene todas las funciones de transferencia entre las \mathbf{p} fuentes y los \mathbf{q} sensores. De esta manera el problema pasaría a ser lineal pero con el dominio de la frecuencia.

Las técnicas de separación ciega de fuentes que se basan en el modelo de mezcla lineal convolutiva tratan de explotar principalmente la llamada diversidad espacial que es el fenómeno que se produce por el hecho de la separación física de los distintos sensores, lo que hace que cada uno de ellos capte, en cada instante de tiempo, distintas mezclas de las fuentes es esencialmente espacial y busca la estructura a través de los sensores, no a lo largo del tiempo.[24, 25]

Presencia de ruido

El ruido suele considerarse, dentro de la separación de fuentes, bajo tres perspectivas distintas:

1. El ruido incide en el medio como una señal más. En este caso los algoritmos que se proponen consideran al ruido como una de las p señales y lo representan correctamente como si se tratara de una fuente cualquiera, realizando la separación, si es posible.
2. Las señales originales están afectadas por ruido. En esta situación los algoritmos de separación actúan transparentemente con respecto al ruido, sin incrementarlo ni reducirlo, recuperándose las señales con la misma relación señal-ruido original.
3. El ruido corrompe las señales ya mezcladas, en el interior del medio o en los sensores. Este es un caso en el que el ruido debe ser tenido en cuenta por las técnicas de separación para reducirlo y, si es posible, eliminarlo. Para modelar este caso, se suele considerar que el ruido es aditivo y acotado.[24]

1.4.2.2 Modelo no lineal:

Un modelo no lineal se considera un conjunto de N señales fuente que tras un proceso de transformación $F_i(\Sigma)$ son detectadas por unos sensores dando lugar a M observaciones.

$$\mathbf{x}(t) = \mathbf{F}(\mathbf{s}(t)) \quad (2.6)$$

Donde $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n)^T$ es una familia de funciones no lineales de mezcla, que transforma las señales originales desconocidas en las mezclas observadas. En general el proceso transformador solo dependerá del canal donde se propagan las señales y de los parámetros de los sensores.

El problema de la separación ciega de señales no lineal, consiste en encontrar una familia de funciones $\mathbf{G} = (\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_n)^T$ tales que, aplicadas sobre las mezclas observadas, devuelvan un conjunto de señales y que sean equivalentes en su forma de onda a las señales originales \mathbf{s} .

Matemáticamente el modelo de separación en el caso no lineal corresponde a obtener:

$$\mathbf{y}(t) = \mathbf{G}(\mathbf{x}(t)) \quad (2.7)$$

Como es lógico, en el mejor de los casos se debe encontrar un conjunto de funciones G que sean las funciones inversas de F , de forma que $G(\mathbf{X}(t)) = s(t)$ y se pueden recuperar las funciones originales:

$$G(x) = F^{-1} \quad \rightarrow \quad g_i(x) = f_i^{-1}(x) \quad \forall i = 1, \dots, n. \quad (2.8)$$

1.5 Conclusiones

En este capítulo se abordaron los principales conceptos que son la base que dan soporte a esta investigación, y que seguirán tratándose en los capítulos siguientes. Además se hizo una reseña de los precedentes históricos del tema de la investigación, teniéndose una visión de cómo se encuentra el tema a nivel mundial. Se analizó los fundamentos teóricos de la Separación Ciega de Fuentes (BSS) y sus modelos que ayudan a entender la base teórica de la investigación, agregado a esto también se describen las representaciones de las señales como elemento indispensable para entender términos principales la Separación de Fuentes de Sonido.

Con el estudio realizado del tema se tienen todos los aspectos para seguir desarrollando las siguientes tareas de la investigación donde se caracterizarán las distintas técnicas de Separación de Fuentes de Sonido.

Capítulo 2: Solución Propuesta.

3.1 Introducción.

En el presente capítulo se aborda los elementos fundamentales que aparecen en el trabajo de acuerdo con la situación polémica planteada en el capítulo 1. A su vez, quedan identificadas las técnicas que logran obtener como resultado la separación de las diferentes fuentes de sonidos; agrupadas estas en dos categorías de acuerdo a las líneas en las que pueden ser aplicadas. También se incluyen un grupo de principios matemáticos, para una mejor comprensión de los procesos que son tratados en el desarrollo del trabajo.

3.2 Técnicas basadas en el aprendizaje no supervisado para la separación de fuentes.

El aprendizaje, así como la inteligencia, estudia un amplio espectro de procesos que son difíciles de precisar. Revisando un diccionario se puede encontrar algo como “adquirir conocimiento, ó, el entendimiento de este, ó, la estructuración del mismo, mediante estudio, instrucción o la experiencia. El aprendizaje automático o aprendizaje maquinal (*machine learning*)[30], comprende diferentes mecanismos, reglas, enfoques y tecnologías mediante los cuales un computador puede aprender a desarrollar tareas que los seres humanos hacen de forma natural y rápida, como por ejemplo; reconocer imágenes, entender el lenguaje natural y tomar decisiones. Usualmente el aprendizaje maquinal se refiere a cambios en un sistema que realiza tareas usando Inteligencia Artificial (IA). Dichas tareas involucran reconocimiento, diagnóstico, planeación, control, etc.

El aprendizaje automático es un concepto primordial, y diferenciador de las técnicas estadísticas más clásicas que fue concebido hace aproximadamente cuatro décadas con el objetivo de desarrollar métodos computacionales que implementarían varias formas de aprendizaje, en particular, mecanismos capaces de inducir conocimiento a partir de datos. Los principales métodos para el aprendizaje están tipificados en diferentes ramas de la inteligencia computacional, pero la gran mayoría se encuentran bajo modelos probabilísticos, estadísticos y algebraicos.

En términos más generales se puede afirmar que el conocimiento maquinal se basa en la idea de utilizar las percepciones no sólo para actuar, sino también para mejorar la habilidad de un agente para actuar en

el futuro, basado en las experiencias del pasado.

Es posible realizar una clasificación de las maneras de aprendizaje según como sea el proceso supervisado, y no supervisado.

El *aprendizaje supervisado* produce una función que establece una correspondencia entre las entradas y las salidas deseadas del sistema; es aquel en el cual se utilizan ejemplos de entrenamiento para "Supervisar" la manera como se adquiere el conocimiento, de tal forma que el sistema se debe ajustar perfectamente al (los) patrón(es) que se utilizaron como entrenadores[31].

En el *aprendizaje no supervisado* todo el proceso de modelado se lleva a cabo sobre un conjunto de ejemplos formado tan sólo por entradas al sistema, está relacionado con el agrupamiento según patrones de similitud entre los datos. Los datos no están clasificados (no etiquetados) y el sistema se encarga de organizarlos dividiéndolos en grupos[31].

El aprendizaje automático está relacionado con el reconocimiento de patrones (*pattern recognition* o *matching*), concepto anterior en el tiempo al de aprendizaje automático y con un enorme potencial práctico. Este campo estudia el desarrollo y aplicación de sistemas complejos basados en técnicas "blandas" (redes neuronales, lógica borrosa, algoritmos evolutivos, etc.) para la tarea de clasificación adaptable de patrones, en una doble vertiente de reconocimiento y de focalización (conocimiento contextual). Este tipo de sistemas tiene su aplicación en problemas de los que no se dispone de un modelo matemático, o el modelo es demasiado complejo, donde las propiedades estadísticas de los datos son muy variables. Tal es el caso, por ejemplo, de patrones visuales basados en imágenes aéreas o satelitales, de problemas de clasificación y diagnóstico en algunos campos (como la medicina o la balística)[32].

Esta sección ofrece una visión general de las actuales técnicas de aprendizaje sin supervisión que han demostrado producir resultados aplicables de separación de señales por sus fuentes. El empleo de estas técnicas se debe a que en el aprendizaje no supervisado estas intentan aprender de las características de las fuentes a partir de las observaciones por el hecho que no hay conocimiento a priori. El aprendizaje no supervisado trata los objetos de entrada como un conjunto de variables aleatorias, siendo construido un modelo de densidad de probabilidad para el conjunto de datos donde la reducción de la redundancia y la maximización de la información mutua son parámetros claves[33]. El Análisis de Componentes Principales (PCA), el Análisis de Componentes Independientes (ICA) y la Factorización no Negativa de

Matrices (NMF) son técnicas que suelen utilizar estos principios basados en teoría de la información y análisis estadístico.

2.2.1 Análisis de Componentes Principales (PCA).

El Análisis de Componentes Principales (*Principal Component Analysis, PCA*)[34] es una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables) planteada para resolver el problema de BSS. Es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible. Los nuevos componentes principales o factores serán una combinación lineal de las variables originales, y además serán independientes entre sí.[35]

3.2.1.1 Cálculos de los componentes principales.

El PCA parte una serie de variables a las que se llamarán fuentes s_1, s_2, \dots, s_n sobre un grupo de objetos y se trata de calcular, a partir de ellas, un nuevo conjunto de variables x_1, x_2, \dots, x_n (observaciones), incorreladas entre sí, cuyas varianzas vayan decreciendo progresivamente.[36]

Cada x_j (donde $j = 1, \dots, n$) es una combinación lineal de las s_1, s_2, \dots, s_n originales, es decir:

$$\begin{aligned} x_j &= a_{j1}s_1 + a_{j2}s_2 + \dots + a_{jn}s_n \\ &= \mathbf{a}'_j \mathbf{s} \end{aligned} \quad (2.21)$$

Donde $x_j = \mathbf{a}'_j \mathbf{s}$ representa una manera reducida de x_j siendo $\mathbf{a}'_j = (a_{1j}, a_{2j}, \dots, a_{nj})$ un vector de constantes, y

$$\mathbf{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_n \end{bmatrix} \quad (2.22)$$

Obviamente, si lo que se quiere es maximizar la varianza, una forma simple podría ser aumentar los coeficientes \mathbf{a}_{ij} . Por ello, para mantener la ortogonalidad de la transformación se impone que el módulo del vector $\mathbf{a}'_j = (a_{1j}, a_{2j}, \dots, a_{nj})$ sea

$$\mathbf{a}'_j \mathbf{a}_j = \sum_{k=1}^n a_{kj}^2 = 1 \quad (2.23)$$

El primer componente se calcula eligiendo \mathbf{a}_1 de modo que \mathbf{x}_1 tenga la mayor varianza posible, sujeta a la restricción de que $\mathbf{a}'_1\mathbf{a}_1 = 1$. El segundo componente principal se calcula obteniendo \mathbf{a}_2 de modo que la variable obtenida, \mathbf{x}_2 esté incorrelada con \mathbf{x}_1 .

Del mismo modo se eligen $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, incorrelados entre sí, de manera que las variables aleatorias obtenidas vayan teniendo cada vez menor varianza.

2.2.1.2 Proceso de extracción de factores.

Para obtener los componentes principales se realiza el proceso de extracción de factores donde se debe elegir \mathbf{a}_1 de modo que se maximice la varianza de \mathbf{x}_1 sujeta a la restricción de que $\mathbf{a}'_1\mathbf{a}_1 = 1$ donde se requiere encontrar un vector \mathbf{a}_1 que satisfaga esta condición.

$$\text{Var}(s_1) = \text{Var}(a'_1 s) = a'_1 \Sigma a_1 \quad (2.24)$$

El método habitual para maximizar una función de varias variables sujeta a restricciones es el método de los multiplicadores de Lagrange.

Puede verificarse que las condiciones para que el vector \mathbf{a}_1 maximice la función $\mathbf{a}'_1 \Sigma \mathbf{a}_1$ sujeta a la restricción $\mathbf{a}'_1 \mathbf{a}_1 = k$ es precisamente las mismas para que el vector \mathbf{a}_1 maximice $a'_1 \Sigma a_1 - \lambda_1 (a'_1 \Sigma a_1 - k)$, λ_1 es la constante conocida como multiplicador de Lagrange

Se puede observar que la incógnita es precisamente \mathbf{a}_1 (el vector desconocido que da la combinación lineal óptima).

Así, se construye la función L :

$$L(\mathbf{a}_1) = \mathbf{a}'_1 \Sigma \mathbf{a}_1 - \lambda (\mathbf{a}'_1 \mathbf{a}_1 - 1) \quad (2.25)$$

y se busca el máximo, derivando e igualando a 0:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{a}_1} &= 2\Sigma \mathbf{a}_1 - 2\lambda \mathbf{a}_1 = 0 \implies \\ (\Sigma - \lambda I) \mathbf{a}_1 &= 0. \end{aligned} \quad (2.26)$$

Esto es, en realidad, un sistema lineal de ecuaciones. Pero para que el sistema tenga una solución distinta de 0 la matriz $(\Sigma - \lambda \mathbf{I})$ tiene que ser singular. Esto implica que el determinante debe ser igual a cero:

$$|\Sigma - \lambda \mathbf{I}| = 0 \quad (2.27)$$

y de este modo, λ es un autovalor de Σ . La matriz de covarianzas Σ es de orden n y si además es definida positiva, tendrá n autovalores distintos, $\lambda_1, \lambda_2, \dots, \lambda_n$ tales que, por ejemplo, $\lambda_1 > \lambda_2 > \dots > \lambda_n$. [37]

Se tiene que, desarrollando la expresión anterior,

$$\begin{aligned} (\Sigma - \lambda \mathbf{I}) \mathbf{a}_1 &= 0 \\ \Sigma \mathbf{a}_1 - \lambda \mathbf{I} \mathbf{a}_1 &= 0 \\ \Sigma \mathbf{a}_1 &= \lambda \mathbf{I} \mathbf{a}_1 \end{aligned} \quad (2.28)$$

entonces,

$$\begin{aligned} \text{Var}(x_1) &= \text{Var}(\mathbf{a}'_1 \mathbf{x}) = \mathbf{a}'_1 \Sigma \mathbf{a}_1 = \lambda \mathbf{a}'_1 \mathbf{a}_1 = \\ &= \lambda \mathbf{a}_1 \mathbf{a}_1 = \lambda \cdot 1 = \lambda \end{aligned} \quad (2.29)$$

Luego, para maximizar la varianza de \mathbf{x}_1 se tiene que tomar el mayor autovalor, dígame λ_1 , y el correspondiente autovector \mathbf{a}_1 .

En realidad, \mathbf{a}_1 es un vector que da la combinación de las variables originales que tiene mayor varianza, esto es, si $\mathbf{a}'_1 = (\mathbf{a}_{11}, \mathbf{a}_{12}, \dots, \mathbf{a}_{1n})$, entonces

$$x_1 = \mathbf{a}'_1 \mathbf{s} = a_{11}s_1 + a_{12}s_2 + \dots + a_{1n}s_n \quad (2.30)$$

El segundo componente principal, dígame $\mathbf{x}_2 = \mathbf{a}'_2 \mathbf{s}$, se obtiene mediante un argumento parecido. Además, se quiere que \mathbf{x}_2 esté incorrelado con el componente anterior \mathbf{x}_1 , es decir, $\text{Cov}(\mathbf{x}_2, \mathbf{x}_1) = 0$. Por lo tanto:

$$\begin{aligned} \text{Cov}(x_2, x_1) &= \text{Cov}(\mathbf{a}'_2 \mathbf{s}, \mathbf{a}'_1 \mathbf{s}) = \\ &= \mathbf{a}'_2 \cdot E[(s - \mu)(s - \mu')] \cdot \mathbf{a}_1 = \\ &= \mathbf{a}'_2 \Sigma \mathbf{a}_1, \end{aligned} \quad (2.31)$$

es decir, se requiere que $\mathbf{a}'_2 \Sigma \mathbf{a}_1 = 0$.

Como se tenía que $\Sigma \mathbf{a}_1 = \lambda \mathbf{a}_1$, lo anterior es equivalente a

$$\mathbf{a}'_2 \Sigma \mathbf{a}_1 = \mathbf{a}'_2 \lambda \mathbf{a}_1 = \lambda \mathbf{a}'_2 \mathbf{a}_1 = 0, \quad (2.32)$$

esto equivale a que $\mathbf{a}'_2 \mathbf{a}_1 = 0$, es decir, que los vectores sean ortogonales.

De este modo, se tiene que maximizar la varianza de \mathbf{x}_2 , es decir, $\mathbf{a}_2' \Sigma \mathbf{a}_2$, sujeta a las siguientes restricciones

$$\begin{aligned} \mathbf{a}'_2 \mathbf{a}_2 &= 1, \\ \mathbf{a}'_2 \mathbf{a}_1 &= 0. \end{aligned} \quad (2.33)$$

Se toma la función:

$$L(\mathbf{a}_2) = \mathbf{a}'_2 \Sigma \mathbf{a}_2 - \lambda(\mathbf{a}'_2 \mathbf{a}_2 - 1) - \delta \mathbf{a}'_2 \mathbf{a}_1 \quad (2.34)$$

y se deriva:

$$\frac{\partial L(\mathbf{a}_2)}{\partial \mathbf{a}_2} = 2\Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}_2 - \delta \mathbf{a}_1 = 0 \quad (2.35)$$

si se multiplica por \mathbf{a}'_1 , entonces

$$2\mathbf{a}'_1 \Sigma \mathbf{a}_2 - \delta = 0 \quad (2.36)$$

porque

$$\begin{aligned} \mathbf{a}'_1 \mathbf{a}_2 &= \mathbf{a}'_2 \mathbf{a}_1 = 0 \\ \mathbf{a}'_1 \mathbf{a}_1 &= 1. \end{aligned} \quad (2.37)$$

Luego

$$\delta = 2\mathbf{a}'_1 \Sigma \mathbf{a}_2 = 2\mathbf{a}'_2 \Sigma \mathbf{a}_1 = 0, \quad (2.38)$$

ya que $Cov(\mathbf{x}_2, \mathbf{x}_1) = 0$.

De este modo, $\frac{\partial L(\mathbf{a}_2)}{\partial \mathbf{a}_2}$ queda finalmente como:

$$\begin{aligned} \frac{\partial L(\mathbf{a}_2)}{\partial \mathbf{a}_2} &= 2\Sigma\mathbf{a}_2 - 2\lambda\mathbf{a}_2 - \delta\mathbf{a}_1 = 2\Sigma\mathbf{a}_2 - 2\lambda\mathbf{a}_2 = \\ (\Sigma - \lambda I)\mathbf{a}_2 &= 0 \end{aligned} \tag{2.39}$$

Usando los mismos razonamientos que antes, se elige λ como el segundo mayor autovalor de la matriz Σ con su autovector asociado \mathbf{a}_2 .

Los razonamientos anteriores se pueden extender, de modo que en el caso del j -ésimo componente, se busca una combinación lineal $\mathbf{X}_j = \mathbf{a}_j' \mathbf{S}$ de las variables originales que sea ortogonal a todas las combinaciones previas. La maximización de la varianza sujeta a estas restricciones lleva a una maximización de una expresión que involucra a j multiplicadores de Lagrange.

Entonces todos los componentes \mathbf{x} (en total \mathbf{n}) se pueden expresar como el producto de una matriz formada por los autovectores, multiplicada por el vector \mathbf{s} que contiene las variables originales $\mathbf{x}_1, \dots, \mathbf{x}_n$. [38]

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{2.40}$$

donde

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \quad \mathbf{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix} \tag{2.41}$$

Como

$$\begin{aligned} \text{Var}(x_1) &= \lambda_1 \\ \text{Var}(x_2) &= \lambda_2 \\ &\dots \\ \text{Var}(x_n) &= \lambda_n \end{aligned} \tag{2.42}$$

la matriz de covarianzas de \mathbf{x} será

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_p \end{pmatrix} \quad (2.43)$$

porque $\mathbf{x}_1, \dots, \mathbf{x}_n$ se han construido como variables incorreladas.

Se tiene que

$$\Lambda = \text{Var}(X) = A' \text{Var}(S) A = A' \Sigma A \quad (2.44)$$

o bien

$$\Sigma = A \Lambda A' \quad (2.45)$$

ya que \underline{A} es una matriz ortogonal (porque $\mathbf{a}'_i \mathbf{a}_i = 1$ para todas sus columnas) por lo que

$$A A' = I \quad (2.46)$$

Para el cálculo de los componentes principales a partir de la matriz de correlaciones habitualmente, se calculan los componentes sobre variables originales estandarizadas, es decir, variables con media 0 y varianza 1. Esto equivale a tomar los componentes principales, no de la matriz de covarianzas sino de la matriz de correlaciones (en las variables estandarizadas coinciden las covarianzas y las correlaciones). Así, los componentes son autovectores de la matriz de correlaciones y son distintos de los de la matriz de covarianzas. Si se actúa así, se da igual importancia a todas las variables originales.

En la matriz de correlaciones todos los elementos de la diagonal son iguales a 1. Si las variables originales están tipificadas, esto implica que su matriz de covarianzas es igual a la de correlaciones, con lo que la variabilidad total (la traza) es igual al número total de variables que hay en la muestra. La suma total de todos los autovalores es \mathbf{n} y la proporción de varianza recogida por el autovector j -ésimo (componente) es

$$\frac{\lambda_j}{\mathbf{n}} \quad (2.47)$$

Uno de los objetivos del cálculo de componentes principales es la identificación de los mismos, es decir, averiguar qué información de la muestra resumen. Sin embargo este es un problema difícil que a menudo resulta subjetivo. Habitualmente, se conservan sólo aquellos componentes que recogen la mayor parte de la variabilidad, hecho que permite representar los datos según dos o tres dimensiones si se conservan dos o tres ejes factoriales, pudiéndose identificar entonces grupos naturales entre las observaciones.

2.2.2 Análisis de Componentes Independientes (ICA).

La independencia estadística como principio básico para la resolución del problema de la separación ciega de señales relaciona estrechamente a dicho problema con el uso de la técnica de Análisis de Componentes Independientes (*Independent Component Analysis, ICA*); suponiendo que las señales lineales originales se mezclan linealmente y que es posible recoger estas mezclas con los sensores correctos.

Esta técnica se puede definir como un vector aleatorio \mathbf{X} donde en su forma más genérica sería la búsqueda de su transformación lineal $\mathbf{S} = \mathbf{W}\mathbf{X}$, tal que las componentes S_i son lo más independientes posibles, en el sentido de maximizar una función $F_s(S_1, S_2, S_3, \dots, S_n)$ que cuantifica la independencia.

Tomando el modelo de mezcla lineal instantáneo donde el valor de las observaciones \mathbf{X}_i es una función lineal de los valores de fuentes originales S_i se representa matemáticamente sin tener en cuenta el tiempo como:

$$\mathbf{X} = \mathbf{A}\mathbf{s} \quad (3.3)$$

Este es un modelo genérico, lo que significa que describe cómo los datos observados son generados por un proceso de mezcla de los componentes S_i . Los componentes independientes son variables latentes, lo que significa que no puede observarse directamente. También se supone que la matriz de mezcla es desconocida.

2.2.2.1 Hipótesis para el Análisis de Componentes Independientes.

Para el análisis de componentes independientes, se deben establecer ciertas hipótesis sobre el conjunto de datos y sobre el modo de mezcla de forma que puedan estimar de una manera adecuada las componentes independientes y que éstas coincidan con las fuentes originales desconocidas[40].

1. Los componentes independientes van a ser asumidos estadísticamente independientes. Hipótesis fundamental sobre la que se rige la técnica de análisis de componentes independientes.
2. Las componentes independientes no tienen distribuciones gaussianas. Cuando los componentes independientes siguen distribuciones gaussianas, las mezclas generadas se pueden separar simplemente mediante métodos de correlación. Sin embargo, el análisis de componentes independientes emplea estadísticos de un gran orden para los cuales una variable de distribución normal toma valores nulos, y así no podría aplicarse la técnica a este tipo de distribuciones.
3. Se asume que la matriz de mezcla es cuadrada (el número de observaciones o mezclas es igual al de componentes a estimar). Eliminándose así la variante de un análisis de componentes sobre-completo (*overcomplete ICA*), es decir, el número de observaciones es menor que el de componentes independientes.
4. La matriz de mezcla es invertible. Ya que de no ser el caso alguna mezcla podría ser redundante y por tanto eliminada, volviendo a la partida de que hay números diferentes de mezclas y fuentes y ya fue eliminado en el planteamiento anterior.

2.2.2.2 Definición y propiedades fundamentales de Independencia.

Para definir el concepto de independencia, considere a dos valores escalares de variables aleatorias \mathbf{y}_1 , \mathbf{y}_2 . Básicamente, las variables aleatorias \mathbf{y}_1 , \mathbf{y}_2 se dice que son independientes si la información obtenida por \mathbf{y}_1 no parte ninguna de la obtenida por \mathbf{y}_2 , y viceversa. Este caso se ha observado anteriormente con las variables \mathbf{s}_1 , \mathbf{s}_2 pero no con las variables de mezcla \mathbf{X}_1 , \mathbf{X}_2 .

Técnicamente, la independencia puede ser definida por la densidad de probabilidad (fdp), denotando a $p(\mathbf{y}_1, \mathbf{y}_2)$ como la función de densidad de probabilidad de \mathbf{y}_1 , \mathbf{y}_2 . También se va a designar por $p_1(\mathbf{y}_1)$ la fdp marginal de \mathbf{y}_1 como independiente cuando la fdp de \mathbf{y}_1 se considera:

$$p_1(\mathbf{y}_1) = \int p(\mathbf{y}_1, \mathbf{y}_2) d\mathbf{y}_2, \quad (3.31)$$

y lo mismo para \mathbf{y}_2 . Entonces se determina que \mathbf{y}_1 y \mathbf{y}_2 son independientes si y sólo si el conjunto es fdp factorizable en la siguiente manera:

$$p(y_1, y_2) = p_1(y_1)p_2(y_2). \quad (3.32)$$

Esta definición se extiende, naturalmente, para cualquier número n de variables aleatorias, en cuyo caso la densidad conjunta debe ser un producto de n términos.

La definición puede utilizarse para determinar la propiedad más importante de variables aleatorias independientes. Dados dos funciones, h_1 y h_2 , se tiene:

$$E\{h_1(y_1)h_2(y_2)\} = E\{h_1(y_1)\}E\{h_2(y_2)\}. \quad (3.33)$$

Esto se puede comprobar como sigue:

$$\begin{aligned} E\{h_1(y_1)h_2(y_2)\} &= \int \int h_1(y_1)h_2(y_2)p(y_1, y_2)dy_1dy_2 \\ &= \int \int h_1(y_1)p_1(y_1)h_2(y_2)p_2(y_2)dy_1dy_2 = \int h_1(y_1)p_1(y_1)dy_1 \int h_2(y_2)p_2(y_2)dy_2 \\ &= E\{h_1(y_1)\}E\{h_2(y_2)\}. \end{aligned} \quad (3.34)$$

2.2.2.3 Variables no correlacionadas son sólo parcialmente independientes.

Una forma débil de la independencia es la no correlación. Dos variables aleatorias y_1 y y_2 se dice que están correlacionadas, si su covarianza es cero:

$$E\{y_1y_2\} - E\{y_1\}E\{y_2\} = 0 \quad (3.35)$$

Estas variables son independientes, donde no están correlacionadas, como se muestra en la ecuación (3.33), teniendo a $h_1(y_1) = y_1$ y $h_2(y_2) = y_2$.

Por otra parte, la no correlación no implica la independencia. Esto se ve por ejemplo si se supone que (y_1, y_2) son valores discretos y siguen una distribución tal que los dos son de probabilidad $1/4$ con cualquiera de los siguientes valores: $(0, 1)$, $(0, -1)$, $(1, 0)$, $(-1, 0)$. Entonces, y_1 y y_2 no están correlacionadas, de modo que pueden ser simplemente calculados.

$$E\{y_1^2y_2^2\} = 0 \neq \frac{1}{4} = E\{y_1^2\}E\{y_2^2\}. \quad (3.36)$$

Por lo que la condición en la ecuación. (3.33) es violada, y las variables no pueden ser independientes.

Al implicar la independencia una no correlación se limita el procedimiento de estimación de modo que siempre da estimaciones no correlacionadas de los componentes independientes. Esto reduce el número de parámetros independientes, y simplifica el problema.

2.2.2.4 ¿Por qué están prohibidas las variables de Gauss?

La restricción fundamental en el ICA es que los componentes independientes no deben ser valores gaussianos. Para ver por qué variables gaussianas hacen imposible al ICA, se asume que la matriz de mezcla es ortogonal y es gaussiana. Luego \mathbf{x}_1 y \mathbf{x}_2 son gaussianas, no correlacionadas y de varianza única. La densidad conjunta está dada por:

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right) \quad (3.37)$$

Esta distribución es mostrada en la Figura 10. La figura muestra que la densidad es totalmente simétrica. Por lo tanto, no contiene ninguna información sobre las direcciones de las columnas de la matriz de mezcla $\underline{\mathbf{A}}$. Esta es la razón por lo que $\underline{\mathbf{A}}$ no puede ser estimada.

Con más rigor, se puede probar que cualquier distribución ortogonal gaussiana $(\mathbf{x}_1, \mathbf{x}_2)$ tiene exactamente la misma distribución que $(\mathbf{x}_1, \mathbf{x}_2)$, y que \mathbf{x}_1 y \mathbf{x}_2 son independientes. Así, en el caso de variables gaussianas, sólo se puede estimar el ICA hasta una transformación ortogonal. En otras palabras, la matriz $\underline{\mathbf{A}}$ no es identificable para componentes independientes gaussianos.

2.2.2.5 Reducción de la información mutua.

Otro enfoque para la estimación del ICA, está inspirado en la teoría de la información, referente a la reducción de la información mutua. Utilizando el concepto de entropía diferencial, se define la información mutua entre las que \mathbf{m} (escalar) representa a las variables aleatorias, \mathbf{y}_i , $i = 1 \dots \mathbf{m}$ de la siguiente manera:

$$I(y_1, y_2, \dots, y_m) = \sum_{i=1}^m H(y_i) - H(\mathbf{y}). \quad (3.38)$$

La información mutua es una medida natural de la dependencia entre variables aleatorias. De hecho, es equivalente a la conocida divergencia entre la densidad conjunta $f(\mathbf{y})$ y el producto de sus

densidades marginales. Donde siempre es no negativa, y cero si y sólo si las variables son estadísticamente independientes.

Le información mutua se puede interpretar mediante la interpretación de la entropía como la longitud del código. El término $H(\mathbf{y}_i)$ tiene la longitud de los códigos de los \mathbf{y}_i cuando estos se codifican por separado, y $H(\mathbf{y})$ da la longitud del código cuando \mathbf{y} es codificado como un vector aleatorio independiente, para todos los componentes que están codificados en el mismo código. La información mutua pone de manifiesto cómo la reducción de la longitud del código se obtiene mediante la codificación del vector de enteros en lugar de los componentes por separado. Sin embargo, si las \mathbf{y}_i son independientes, estas no ofrecen ninguna información entre sí, y no se podría obtener correctamente el código de las variables por separado, sin aumentar la longitud de código.

Dado que la información mutua es la información medida natural de la teoría de la independencia de variables aleatorias, se puede utilizar como criterio para la búsqueda de la transformada del ICA. En este caso siendo una alternativa al modelo del método de estimación, se define al ICA como un vector aleatorio \mathbf{X} en la búsqueda de una transformación invertible como $\mathbf{S} = \mathbf{W}\mathbf{X}$, donde la matriz \mathbf{W} está determinada de manera que la información mutua de los componentes transformados \mathbf{S}_i se reducen al mínimo. De modo que encontrar una transformación invertible \mathbf{W} que minimiza la información mutua es aproximadamente equivalente a la búsqueda de direcciones en las que se maximiza la negentropía. Rigurosamente, hablando, esto muestra que la estimación de ICA por la minimización de la información mutua es equivalente a maximizar la suma de de las estimaciones no gaussianas, cuando las estimaciones se ven obligadas a estar correlacionadas.

La restricción de no correlación, no es de hecho necesaria, pero simplifica considerablemente en cuanto a los cálculos.

2.2.2.6 Indeterminaciones del ICA.

Mediante este conjunto de restricciones se puede decir que el modelo ICA es identificable, esto es, que se puede estimar la matriz de mezcla y las componentes independientes salvo algunas indeterminaciones de escalado y permutaciones. Estas indeterminaciones son:

1. Permutación: El orden de la componente independiente no se puede restablecer. Obviamente, se puede permutar el orden de las componentes independientes estimadas y éstas seguirán

manteniendo la propiedad de independencia. Como el análisis de componentes independientes tan solo parte del conocimiento del conjunto de observaciones, resulta imposible obtener la estimación de las fuentes en el mismo orden que las originales.

2. Escalado: No es posible determinar la amplitud de las señales independientes. Efectivamente, si un conjunto de variables $Y_1, Y_2, Y_3, \dots, Y_n$ es independiente, esta se mantiene si son multiplicadas dichas variables aisladamente por un coeficiente real. Nótese que incluso los escalados pueden ser invertibles, por lo que se pueden dar por válidas soluciones en las que no sólo cambia la magnitud de las señales estimadas respecto a las originales, sino también su signo. No obstante, esta indeterminación no suele ser un problema mayor en el campo de la separación ciega de señales, puesto que el objetivo real de la separación es recuperar la forma de las señales originales, sin tener demasiada importancia sus respectivas magnitudes. En cualquier caso, se suelen fijar las magnitudes de las estimaciones, usualmente normalizando las señales para que posean varianza unidad: $E \{Y^2\}=1$.

Debido a lo anteriormente mencionado, con frecuencia las señales recuperadas aparezcan en el orden distinto a las señales originales y afectadas por un factor de escala que las aumenta o disminuye. Esto es un hecho aceptado por todos los autores que trabajan en separación ciega de fuentes, y se considera válida la separación aún cuando concurren estas dos indeterminaciones.

2.2.3 Factorización no Negativa de Matrices (NMF).

Un problema bastante extendido en diferentes técnicas de análisis de datos consiste en encontrar una representación adecuada de los datos. Un tipo de representación de gran utilidad será aquella donde se permita reducir las dimensiones de los datos a la vez que muestre ciertas características del conjunto de éstos que permanezcan ocultas a priori. La Factorización no Negativa de Matrices (*Non-negative Matrix Factorization, NMF*) es una técnica de representación lineal y no negativa de conjunto de datos de reciente creación capaz de resolver el problema BSS con los mejores resultados; su principal utilidad consiste en encontrar una representación lineal de los datos, que han de ser no negativos.

Una de las propiedades principales de la Factorización No Negativa de Matrices radica en que generalmente suele proporcionar una representación dispersa de los datos que permite codificarlos de forma que el resultado sea bastante sencillo y fácil de interpretar.

Partiendo que $\underline{\mathbf{X}}$ es una matriz no negativa generada a partir de $\underline{\mathbf{A}}$ y $\underline{\mathbf{S}}$, se debe encontrar los factores de las matrices $\underline{\mathbf{A}}$ y $\underline{\mathbf{S}}$ [41] [42].

$$\underline{\mathbf{X}} \approx \underline{\mathbf{A}}\underline{\mathbf{S}}$$

Esta técnica está basada en actualizaciones iterativas de las matrices $\underline{\mathbf{A}}$ y $\underline{\mathbf{S}}$. Las dimensiones de la matriz $\underline{\mathbf{X}}$ está representada por $\mathbf{n} \times \mathbf{m}$ donde \mathbf{m} es el número de observaciones de las señales \mathbf{n} . Esta matriz se obtiene a partir de la factorización de la matriz $\underline{\mathbf{A}}$ de dimensión $\mathbf{n} \times \mathbf{r}$ y la matriz $\underline{\mathbf{S}}$ de dimensión $\mathbf{r} \times \mathbf{m}$. Normalmente \mathbf{r} es más pequeño que \mathbf{n} ó \mathbf{m} de modo que $\underline{\mathbf{A}}$ y $\underline{\mathbf{S}}$ son más pequeñas que la matriz original $\underline{\mathbf{X}}$, esta elección de \mathbf{r} conllevará un considerable ahorro de información a almacenar.

2.2.3.1 Definición de la función de coste.

Para encontrar una representación del tipo $\underline{\mathbf{X}} \approx \underline{\mathbf{A}}\underline{\mathbf{S}}$ es necesario previamente definir una función de coste que permite cuantificarla calidad de la aproximación. Esta función de coste se puede construir a partir de una medida típica de la distancia entre dos matrices no negativas genéricas, que se denotan por $\underline{\mathbf{C}}$ y $\underline{\mathbf{B}}$. La función de distancia más conocida es la distancia euclídea, que se va a describir seguidamente[42]:

$$\|\underline{\mathbf{C}} - \underline{\mathbf{B}}\|^2 = \sum_i (C_{ij} - B_{ij})^2 \quad (3.40)$$

siendo C_{ij} y B_{ij} los elementos de las matrices $\underline{\mathbf{C}}$ y $\underline{\mathbf{B}}$. Esta función está acotada inferiormente por cero, valor que toma cuando $\underline{\mathbf{C}} = \underline{\mathbf{B}}$.

Basado en medir la calidad del problema de optimización NMF se puede formular lo siguiente:

$$\text{Minimizar } \|\underline{\mathbf{X}} - \underline{\mathbf{A}}\underline{\mathbf{S}}\|^2 \text{ para } \underline{\mathbf{A}} \text{ y } \underline{\mathbf{S}}, \text{ tal que } \underline{\mathbf{A}}, \underline{\mathbf{S}} \geq 0. \quad (3.41)$$

Hay que tener en cuenta que aunque la función $\|\underline{\mathbf{X}} - \underline{\mathbf{A}}\underline{\mathbf{S}}\|^2$ es convexa en sólo $\underline{\mathbf{A}}$ ó sólo $\underline{\mathbf{S}}$, no es convexa en ambas variables. Por lo tanto, cualquier método de optimización basado en el descenso de gradiente pueden encontrar sólo los mínimos locales de la función objetivo en (3.41). [43]

2.2.3.2 Reglas de actualización multiplicativas.

Para resolver lo planteado anteriormente se van a tomar los siguientes teoremas.

Teorema 1. La distancia euclídea $\|\underline{\mathbf{X}} - \underline{\mathbf{A}}\underline{\mathbf{S}}\|^2$ es estrictamente no decreciente bajo la siguiente regla de actualización de las matrices $\underline{\mathbf{A}}$ y $\underline{\mathbf{S}}$:

$$\begin{aligned}
 A_{ij} &\leftarrow A_{ij} \frac{(X S^T)}{\underline{A} \underline{S} S^T} \\
 S_{ij} &\leftarrow S_{ij} \frac{(A^T X)}{\underline{A}^T \underline{A} S}
 \end{aligned}
 \tag{3.42}$$

La distancia euclídea es invariante bajo estas reglas de actualización si y sólo si \underline{A} y \underline{S} se encuentran en un punto estacionario de la distancia.

Teorema 2. La divergencia $D(\underline{C} - \underline{B})$ es estrictamente no decreciente bajo la siguiente regla de actualización de las matrices \underline{A} y \underline{S} :

$$\begin{aligned}
 S_{\bar{y}} &\leftarrow S_{\bar{y}} \frac{\sum_i \underline{A}_{\bar{y}} X_{\bar{y}} / (\underline{A} \underline{S})_{\bar{y}}}{\sum_j \underline{A}_{\bar{y}}} \\
 A_{\bar{y}} &\leftarrow A_{\bar{y}} \frac{\sum_i S_{\bar{y}} X_{\bar{y}} / (\underline{A} \underline{S})_{\bar{y}}}{\sum_i \underline{A}_{\bar{y}}}
 \end{aligned}
 \tag{3.43}$$

La divergencia es invariante bajo estas reglas de actualización si y sólo si \underline{A} y \underline{S} se encuentran en un punto estacionario de la divergencia[44].

Hay que destacar que cada actualización consiste en una multiplicación por un factor. En particular, se puede comprobar cómo este factor multiplicativo es la unidad cuando $\underline{X} = \underline{A} \underline{S}$ de forma que esta reconstrucción se corresponde necesariamente con un punto fijo en las reglas de actualización previamente descritas.

Las reglas de actualización multiplicativa descritas en (4.42) se pueden derivar de las reglas de actualización aditiva fundamentada en un progresivo descenso del gradiente. Para ilustrar esto considera una regla aditiva para actualizar la matriz \underline{S} que reduce la distancia cuadrática de (4.41) para una determinada \underline{A} que se puede escribir de la siguiente forma:

$$S_{ij} \leftarrow S_{ij} + \eta_{ij} [(\underline{A}^T \underline{X})_{ij} - (\underline{A}^T \underline{A} \underline{S})_{ij}].
 \tag{3.44}$$

Si los η_{ij} se toman todos iguales a una constante positiva arbitrariamente pequeña, la expresión anterior se corresponde con el algoritmo clásico del gradiente. Mientras menor sea esta constante, mayor será la reducción que se produzca en $\|\underline{X} - \underline{A} \underline{S}\|$.

Con el fin de obtener la regla de actualización de la multiplicativa de $\underline{\mathbf{S}}$, como se muestra en el *Teorema 1* se debe mostrar de la siguiente manera:

$$\eta_{ij} = \frac{S_{ij}}{(\underline{\mathbf{A}}^T \underline{\mathbf{A}} \underline{\mathbf{S}})_{ij}}. \quad (3.45)$$

Incluso si η_{ij} no pequeños se puede demostrar con el descenso del gradiente que siempre se conduce a una función de costos decrecientes. Esto será probado en la próxima sección.[42, 45]

2.2.3.3 Pruebas de convergencia.

Para demostración de los *Teoremas 1* y *2* de la última sección se necesita denominar las funciones auxiliares $\mathbf{G}(\mathbf{s}, \mathbf{s}')$. Dichas funciones se definen de la siguiente manera:

Definición 1. $\mathbf{G}(\mathbf{s}, \mathbf{s}')$ es una función auxiliar de $\mathbf{F}(\mathbf{s})$ si es continua y

$$G(\mathbf{s}, \mathbf{s}') \geq F(\mathbf{s}), \quad G(\mathbf{s}, \mathbf{s}) = F(\mathbf{s}) \quad (3.46)$$

lo satisface.

La función auxiliar es un concepto utilizado por el siguiente lema. [46]

Lema 1. Si \mathbf{G} es una función auxiliar, entonces \mathbf{F} es no creciente bajo la actualización de

$$\mathbf{s}^{t+1} = \arg \min_{\mathbf{s}} G(\mathbf{s}, \mathbf{s}^t) \quad (3.47)$$

Prueba:
$$F(\mathbf{s}^{t+1}) \leq G(\mathbf{s}^{t+1}, \mathbf{s}^t) \leq G(\mathbf{s}^t, \mathbf{s}^t) = F(\mathbf{s}^t)$$

Obviamente, si \mathbf{s}^t es un mínimo local de $\mathbf{G}(\mathbf{s}, \mathbf{s}^t)$ entonces $\mathbf{F}(\mathbf{s}^{t+1}) = \mathbf{F}(\mathbf{s}^t)$. Por otra parte, si $\mathbf{F}(\mathbf{s})$ es diferenciable y si sus derivadas son continuas en un pequeño intervalo de \mathbf{s}^t , esto también implica que las derivadas $\nabla \mathbf{F}(\mathbf{s}^t) = 0$. Así, iterando la actualización de la ecuación (3.47) se obtiene una secuencia de de las estimaciones que convergen en un mínimo local \mathbf{s}_{min} de la función objetivo $\mathbf{F}(\mathbf{s})$:

$$F(\mathbf{s}_{min}) \leq \dots \leq F(\mathbf{s}^{t+1}) \leq F(\mathbf{s}^t) \leq \dots \leq F(\mathbf{s}^2) \leq F(\mathbf{s}^1) \leq F(\mathbf{s}^0). \quad (3.48)$$

El siguiente lema demuestra como la convergencia de la función objetivo $\mathbf{F} = \|\underline{\mathbf{X}} - \underline{\mathbf{A}} \underline{\mathbf{S}}\|$ se puede probar si se determina una adecuada función auxiliar $\mathbf{G}(\mathbf{s}, \mathbf{s}^t)$ de las normas de actualización de los *Teoremas 1* y *2* tratadas además en la ecuación (3.47).[44]

Lema 2. Si $\underline{\mathbf{K}}(\mathbf{s}^t)$ es la matriz diagonal

$$K_{ij}(\mathbf{s}^t) = \delta_{ij}(\underline{\mathbf{A}}^T \underline{\mathbf{A}} \mathbf{s}^t)_i / s_i^t \quad (3.49)$$

entonces

$$G(\mathbf{s}, \mathbf{s}^t) = F(\mathbf{s}^t) + (\mathbf{s} - \mathbf{s}^t) \nabla F(\mathbf{s}^t) + \frac{1}{2} (\mathbf{s} - \mathbf{s}^t) \underline{\mathbf{K}}(\mathbf{s}^t) (\mathbf{s} - \mathbf{s}^t) \quad (3.50)$$

Es una función auxiliar para

$$F(\mathbf{s}) = \frac{1}{2} \sum_i (\mathbf{x}_i - \sum_j \underline{\mathbf{A}}_{ij} s_j)^2. \quad (3.51)$$

(Aquí la notación " \mathbf{x}_i " indica la i -ésimas columnas de la matriz $\underline{\mathbf{X}}$)

Demostración: Como se puede ver directamente en (3.50) $\mathbf{G}(\mathbf{s}, \mathbf{s}^t) = F(\mathbf{s})$ de tal manera que $\mathbf{G}(\mathbf{s}, \mathbf{s}^t) \geq F(\mathbf{s})$ sólo necesita ser demostrado de forma explícita. Para ello compara la expansión de Taylor de $F(\mathbf{s})$.

$$F(\mathbf{s}) = F(\mathbf{s}^t) + (\mathbf{s} - \mathbf{s}^t) \nabla F(\mathbf{s}^t) + \frac{1}{2} (\mathbf{s} - \mathbf{s}^t) (\underline{\mathbf{A}}^T \underline{\mathbf{A}}) (\mathbf{s} - \mathbf{s}^t) \quad (3.52)$$

con la definición de $\mathbf{G}(\mathbf{s}, \mathbf{s}^t)$ en (3.50). Obviamente, $\mathbf{G}(\mathbf{s}, \mathbf{s}^t) \geq F(\mathbf{s})$ sólo si se mantiene

$$0 \leq (\mathbf{s} - \mathbf{s}^t) [\underline{\mathbf{K}}(\mathbf{s}^t) - \underline{\mathbf{A}}^T \underline{\mathbf{A}}] (\mathbf{s} - \mathbf{s}^t). \quad (3.53)$$

Para de ver si se puede cumplir la desigualdad de la matriz

$$M_{ij}(\mathbf{s}^t) = s_i^t (\underline{\mathbf{K}}(\mathbf{s}^t) - \underline{\mathbf{A}}^T \underline{\mathbf{A}})_{ij} s_j^t, \quad (3.54)$$

que es sólo un reajuste de $\underline{\mathbf{K}} - \underline{\mathbf{A}}^T \underline{\mathbf{A}}$, se comprueba la semidefinitud positiva.

$$\begin{aligned}
 \mathbf{v}^T \underline{\mathbf{M}} \mathbf{v} &= \sum_{ij} v_i M_{ij} v_j \\
 &= \sum_{ij} s_i^t (\underline{\mathbf{A}}^T \underline{\mathbf{A}})_{ij} s_j^t v_i^2 - v_i s_i^t (\underline{\mathbf{A}}^T \underline{\mathbf{A}})_{ij} h_j^t v_j \\
 &= \sum_{ij} (\underline{\mathbf{A}}^T \underline{\mathbf{A}})_{ij} s_i^t s_j^t \left[\frac{1}{2} v_i^2 + \frac{1}{2} v_j^2 - v_i v_j \right] \\
 &= \frac{1}{2} \sum_{ij} (\underline{\mathbf{A}}^T \underline{\mathbf{A}})_{ij} s_i^t s_j^t (v_a - v_b)^2 \\
 &\geq 0.
 \end{aligned} \tag{3.55}$$

Ahora se puede demostrar la convergencia del *Teorema 1*:

Demostración del *Teorema 1*. Sustituyendo $\mathbf{G}(\mathbf{s}, \mathbf{s}^t)$ en (3.17) para (3.20) el resultado en la regla de actualización es

$$\mathbf{s}^{t+1} = \mathbf{s}^t - K(\mathbf{s}^t)^{-1} \nabla F(\mathbf{s}^t). \tag{3.56}$$

Dado que la ecuación (3.50) es una función auxiliar, \mathbf{F} es no creciente para esta regla de actualización, de acuerdo con el *Lema 1*. Por último, si los componentes de (3.56) se escriben de forma explícita para la regla de actualización del *Teorema 1* se obtiene:[42]

$$s_i^{t+1} = s_i^t \frac{(\underline{\mathbf{A}}^T \mathbf{x})_i}{(\underline{\mathbf{A}}^T \underline{\mathbf{A}} \mathbf{s}^t)_i}. \tag{3.57}$$

NMF es un tipo de representación lineal y no negativa de un conjunto de datos. NMF impone que todos los valores de ambas matrices sean no negativos, lo que significa que los datos son descritos utilizando sólo componentes aditivas. Esta restricción impuesta por NMF está motivada principalmente por la consideración de que en la mayoría de los sistemas físicos reales las cantidades implicadas no pueden ser negativas. Por otra parte la técnica NMF permite optimizar los algoritmos que se ejecuten utilizando los datos de salida de este porque reducen la dimensionalidad de los datos sin alterar el resultado para realizar una clasificación.

2.2.4 Comparación entre las técnicas PCA, ICA y NMF.

El Análisis de Componentes Principales (PCA), el Análisis de Componentes Independientes (ICA) y la Factorización no Negativa de Matrices (NMF) pueden considerarse como técnicas basadas

factorizaciones de matrices con diferentes elecciones en la función objetivo o en las restricciones que imponen en la representación.

En PCA, se trata de buscar las componentes principales de las observaciones que consisten en los autovalores de la matriz de covarianza de estas. En PCA la clave es maximizar la varianza de las componentes sujetas a condiciones de ortogonalidad entre ellas. Por otro lado el objetivo de ICA es buscar la máxima independencia posible entre las componentes que forman las observaciones. De este modo se trata de representar los datos de una forma más eficiente y que preserve su información. PCA, es una técnica de reducción de la dimensión. PCA surge como respuesta a: encontrar direcciones incorreladas del espacio que tengan máxima varianza y encontrar el sistema de referencia ortogonal del subespacio que proporcione la proyección óptima de las observaciones en el sentido de mínimos cuadrados. Un primer enfoque trataría de encontrar unas nuevas variables (llamadas factores) no correlacionadas entre sí y que sean función lineal de las originales y que expliquen la mayor parte de la varianza total de los datos. El segundo enfoque pretende encontrar un subespacio de modo que la diferencia entre las observaciones y sus proyecciones tengan el menor error cuadrático medio. El punto de partida de ICA es la asunción de que las componentes buscadas son estadísticamente independientes.[47]

Mientras que PCA e ICA no imponen ninguna restricción en cuanto al signo que han de tener las entradas de las matrices de las bases y de codificación, las técnicas basadas en la Factorización No Negativa de Matrices han de presentar la peculiaridad de que dichas matrices han de contener elementos que han de ser *no negativos*. Esto conlleva que la descripción de los datos se realiza tan sólo a partir de sumas. Esta restricción se debe principalmente a dos motivos. Por un lado, que el conjunto de datos que se maneja nunca puede ser negativo. En estos casos, la interpretación de resultados mediante técnicas basadas en PCA o ICA sería más dificultosa. En segundo lugar, la no negatividad se usa como argumento para ilustrar la idea que las partes se combinan siempre de forma aditiva (nunca a partir de restas) para dar lugar a un todo.

Determinando las bases a partir de técnicas ICA. En general, se puede afirmar que mediante ICA se pretende minimizar la dependencia estadística entre los elementos de la base resultante. Sin embargo hay algo más que diferencia ICA y NMF. Hasta ahora, en el estudio de la Factorización No Negativa de Matrices, no se ha realizado ninguna consideración acerca de la dependencia estadística de las variables aleatorias a estimar. Sin embargo en el Análisis de Componentes Independientes se exigen una serie de restricciones que han de cumplir las variables para que puedan ser estimadas correctamente: han de ser

no gaussianas y estadísticamente independientes. Esto plantea un pequeño problema en la aplicación del ICA cuya solución pasa por PCA. La suposición de independencia no es especialmente recomendable ya que es bastante probable que algunas de estas partes en la mezcla aparezcan de forma simultánea en los datos.

La mayor diferencia entre NMF y otras técnicas de factorización que han sido aplicadas a la separación de fuentes, tales como Análisis de Componentes Principales (PCA), y Análisis de Componentes Independientes (ICA), radica en la restricción impuesta a los vectores base ($\underline{\mathbf{A}}$) y a los codificantes ($\underline{\mathbf{S}}$) de no utilizar valores negativos. Gracias a esta restricción, se obtiene una representación de los datos basada en partes o secciones, ya que sólo se permiten combinaciones aditivas, nunca sustractivas.

De esta manera, los factores pueden ser interpretados como partes de los datos o, dicho de otro modo, como elementos que tienden a aparecer juntos. Por el contrario, otras técnicas de factorización como las mencionadas anteriormente, permiten que los valores de $\underline{\mathbf{A}}$ y $\underline{\mathbf{S}}$ tengan cualquier signo, lo que conlleva a cancelaciones complejas de elementos positivos y negativos durante la reconstrucción del conjunto original de datos. En otras palabras, NMF tiende a producir factores que permiten una interpretación contextual relativamente sencilla, mientras que aquellos obtenidos mediante las otras técnicas no contienen un “significado” intuitivo.[48]

2.2.5 Separación no supervisada utilizando PCA y NMF.

Las técnicas fundamentadas en el aprendizaje no supervisado donde la representación de datos está basada en restricciones de independencia estadística han sido aplicadas con éxito para la separación de señales por sus fuentes.

Un conjunto multidimensional de datos (denotado genéricamente en el trabajo mediante una matriz \mathbf{X}) puede ser representado mediante diversas técnicas de transformación de atributos. No obstante, la utilidad de dichos métodos de representación dependen tanto de su práctica respecto a la naturaleza de los datos como de la utilización que se requiera hacer de los mismos. Por ejemplo, El Análisis de Componente Principales (PCA) es útil para describir los datos sobre un espacio ortogonal [34]. Por su parte el Análisis de Componentes Independientes (ICA) da lugar a representaciones significativas cuando el conjunto de datos analizado responde a un modelo generativo de variables estadísticamente independientes[40]. Por otro lado, la Factorización de Matrices no Negativas (NMF) descompone linealmente conjuntos de datos bajo la restricción de la no negatividad de los mismos [46].

Pese a sus distintos enfoques, todas estas técnicas de representación tienen en común su capacidad para representar los datos dentro de una mezcla acústica, lo que resulta ventajoso desde el punto de vista de su almacenamiento y posterior procesamiento.

Cuando se considera la aplicación de este tipo de técnicas al análisis de sonido digital, es preciso definir el espacio de datos sobre el que se opera. Así la matriz \mathbf{X} como ha sido definida anteriormente contiene la información auditiva a analizar. En términos generales se tiene en cuenta que la matriz \mathbf{X} está dada por la relación que existe entre el número de fuentes \mathbf{N} con el número de micrófonos \mathbf{M} , es decir la matriz \mathbf{X} (de dimensiones $\mathbf{M} \times \mathbf{N}$) representa toda la mezcla acústica.

2.2.5.1 Extracción de las bases auditivas para PCA.

Como se describe en (2.2.1) el Análisis de Componentes Principales ofrece las direcciones principales de variación de los datos, así como la información relativa a la importancia de los mismos, siendo la descomposición de valores singulares (*Singular Value Decomposition*, SVD) [49] una vía para encontrarlas. La SVD es una técnica que factoriza cualquier matriz \mathbf{X} de \mathbf{M} filas, \mathbf{N} columnas y un rango \mathbf{K} en el siguiente producto:

$$X_{M \times N} = U_{M \times K} \cdot \Sigma_{K \times K} \cdot (V_{N \times K})^T \cdot m_{M \times 1} \quad (3.58)$$

Donde \mathbf{U} y \mathbf{V} son dos matrices ortogonales, Σ es una matriz diagonal con los números reales. La matriz \mathbf{U} contiene, en sus columnas, las componentes principales de \mathbf{X} , en la diagonal de Σ , sus valores singulares y en las columnas \mathbf{V} , las componentes principales de \mathbf{X}^T .

$$X_{M \times N} \approx U_{M \times R} \cdot \Sigma_{R \times R} \cdot (V_{N \times R})^T \cdot m_{M \times 1} = \Lambda \quad (3.59)$$

Con SVD se puede reducir la dimensionalidad de los datos al proporcionar una mejor aproximación de rango R (siendo $R < K$) de una matriz \mathbf{X} cualquiera, en el sentido de los mínimos cuadrados.

Para representar un conjunto de datos de audio \mathbf{X} en función de sus componentes principales, se realiza SVD de rango R sobre (3.59). Como resultado se obtienen los R vectores base (columnas de \mathbf{U}) y las proyecciones de cada columna de \mathbf{X} sobre esta base en $\mathbf{C} = \Sigma \mathbf{V}^T$, que representa (por filas) los niveles de cada columna \mathbf{U} . Hay que destacar que las bases de \mathbf{U} están formadas por la información que corresponde a la información auditiva. Para reducir el consumo de la aplicación de PCA mediante SVD, se realiza el cálculo incremental de la SVD con actualización de la media. A partir de una inicialización

que utiliza las primeras S columnas de \mathbf{X} en (3.60), se obtienen las nuevas matrices \mathbf{U}_{t+1} , $\mathbf{\Sigma}_{t+1}$ y \mathbf{V}_{t+1} y la nueva columna media \mathbf{m}_{t+1} a partir de \mathbf{U}_t , $\mathbf{\Sigma}_t$, \mathbf{V}_t , \mathbf{m}_t en (3.61) y S nuevas columnas de \mathbf{X} , es decir, $\mathbf{x}_{tS+1}, \dots, \mathbf{x}_{tS+S}$ (3.62).

$$X_0 = [x_1 \cdots x_2] = U_0 \cdot \Sigma_0 \cdot (V_{t+1})^T + m_0 \quad (3.60)$$

$$X_t = U_t \cdot \Sigma_t \cdot (V_t)^T + m_0 \quad (3.61)$$

$$X_{t+10} = [X_t \quad x_{tS+1} \cdots x_{tS+S}] = U_{t+1} \cdot \Sigma_{t+1} \cdot (V_{t+1})^T + m_{t+1} \quad (3.62)$$

Si el número de valores singulares se limita en cada paso de la iteración y se hace menor que el rango de \mathbf{X} , las ecuaciones 3.60, 3.61 y 3.62 pasan a ser aproximaciones.

2.2.5.2 Extracción de bases auditivas mediante NMF.

La Factorización en Matrices No Negativas se introduce ahora como técnica para aproximar una matriz no-negativa \mathbf{X} (de dimensiones $M \times N$) mediante el producto de dos matrices no negativas: una matriz \mathbf{W} ($M \times R$) y una matriz \mathbf{H} ($R \times N$).

$$X \approx W \cdot H = \Delta$$

Al igual que ICA, NMF asume la existencia de un modelo generativo. Sin embargo, NMF considera que el conjunto de datos multidimensional observado es el resultado de la suma ponderada de R fuentes (o vectores base) no negativas. De acuerdo a este modelo generativo, los R vectores base no negativos están contenidos en las columnas de \mathbf{W} y los pesos asociados a cada vector base se encuentran en las filas de \mathbf{H} . La matriz $\mathbf{\Lambda}$ denota la aproximación de la secuencia \mathbf{X} .

Desde el punto de vista algorítmico, la aproximación de la (ecuación 6) puede realizarse mediante la optimización de una función de coste (3.63) [46]. Ésta se optimiza iterativamente mediante la aplicación de las reglas de multiplicativas de actualización propuestas en (sección 2.2.3.2), donde se denota la división elemento a elemento y \mathbf{T} denota la transposición matricial (3.64).

$$FC = \|\mathbf{X} - \mathbf{\Lambda}\|^2 \quad (3.63)$$

$$H \leftarrow H \frac{W^T X}{W^T \Lambda}, \quad W \leftarrow W \frac{X H^T}{\Lambda \cdot H^T} \quad (3.64)$$

Teniendo en cuenta la descomposición lineal de PCA y NMF la combinación de estas dos técnicas permite llevar a cabo no solo la separación de las fuentes de sonido, sino también que lo lograría a un coste computacional razonable.

3.3 Técnica de separación basada en el modelo sinusoidal.

En esta sección se hace referencia a la separación basada en el modelo sinusoidal; debido a que en aplicaciones musicales, el modelo más utilizado es precisamente el modelo sinusoidal. Este modelo tiene principalmente en cuenta la armoniosidad de las fuentes sonoras, lo cual lo hace muy útil en la separación de instrumentos musicales mediante seguimiento de pitch.

Una descomposición eficiente de los sonidos producidos por instrumentos musicales está dado por el modelo sinusoidal más un ruido, que representa la señal como una suma determinista y de partes aleatorias, o, como una suma de un conjunto de sinusoides además de un ruido residual[50]. Los componentes sinusoidales son producidos por un sistema en vibración que es armónico.

3.3.1 Modelo de la señal.

El modelo sinusoidal de un frame de $x(n)$, $n = 0, \dots, N - 1$ donde la señal puede ser escrita como

$$x(n) = \sum_{h=1}^H a_h \cos(f_h n / f_s + \theta_h) + r(n), \quad n = 0, \dots, N - 1, \quad (3.65)$$

Donde n es el índice del tiempo, N es la longitud del frame, a_h , f_h , y θ_h son la amplitud, la frecuencia, y la fase inicial de la senoide h^{th} , respectivamente y $r(n)$ es el residual. En la mayoría de los casos se supone que los parámetros dentro de cada frame son fijos, a pesar de que los parámetros varían con el tiempo. La estimación de los parámetros se realiza en el dominio de la frecuencia. Cada senoide corresponde a un pico en el espectro de la amplitud, y sus frecuencias se pueden calcular al ponderarse por los picos más destacados.

El modelo sinusoidal es una poderosa herramienta para el análisis de señales musicales, ya que puede conservar las frecuencias exactas de los parciales armónicos. Esto permite la asociación de las sinusoides en las fuentes de sonido y la estimación en un alto nivel de dicha información para las notas tocadas por cada fuente. Las amplitudes de las sinusoides se han utilizado, por ejemplo, para estimar los parámetros de sonido de instrumentos de cuerda.

Para fuentes de sonido armónicos, las frecuencias de las sinusoides son enteros simples de aproximadamente la multiplicación de la frecuencia fundamental. Tomando esto en cuenta en el modelo

de la señal, mediante la estimación de estructuras armónicas en lugar de los sinusoides individuales. Para la suma de las M fuentes el modelo de señal se muestra como

$$x(n) = \sum_{m=1}^M \sum_{h=1}^{H_m} a_{m,h} \cos(2\pi f_{m,h}n + \theta_{m,h}) + r(n), \quad n = 0, \dots, N - 1 \quad (3.66)$$

Donde H_m es el número de tonos en las fuentes de m , así como $a_{m,h}$, $f_{m,h}$, y $\theta_{m,h}$, son la amplitud, frecuencia y la fase del sobre-tono h^{th} . Suponiendo que las fuentes son armónicas, de modo que $f_{m,h} \approx hf_{m,1}$, donde $f_{m,1}$ es la frecuencia fundamental de la fuente m .

3.3.2 Resolución de superposición de sobre-tonos.

Lograr la separación de señales superpuestas en el dominio del tiempo y la frecuencia resulta un problema difícil de resolver con cualquier técnica de separación que se utilice. En el caso de modelación sinusoidal armónica las relaciones de frecuencia fundamentales son complejas. Cuando dos o más fuentes se encuentran relacionadas armónicamente, algunos tonos tienen aproximadamente la misma frecuencia de manera que estos chocan, o, se superponen. Este fenómeno es común en las señales musicales, ya que en la música las frecuencias fundamentales están normalmente relacionadas de manera armónica.

La forma aproximada de la amplitud del espectro de los sonidos naturales es por lo general lentamente variable en función del tiempo y la frecuencia. También se ha observado que el principio de suavidad espectral es una señal importante en la fusión espectral de los componentes en las fuentes [51]. Debido a esta mencionada propiedad acústica y de la percepción del sonido humano, las amplitudes de la superposición de parciales se pueden aproximar bastante bien mediante la interpolación (superposición, combinación) de frames adyacentes o parciales. Para algunos instrumentos musicales son los armónicos de enganche de fase [52], de modo que la interpolación también podría utilizarse para resolver las fases de acumulación parciales. Sin embargo, el bloqueo de fase no se puede suponer, en general, ya que hay también instrumentos musicales cuando esto no se cumple [52]. Dado que las fases son perceptivamente menos importantes, no es necesaria una cuidadosa estimación de las fases. O bien la fase de la señal mezcla se puede asignar para todos los armónicos superpuestos, o sus fases se pueden generar para producir transiciones suaves entre frame. El método más comúnmente usado para aproximar la amplitud de $a_{m,h}$ de la h^{th} , parcial de m fuente es por interpolación lineal de dos parciales adyacentes $h - 1$ y $h + 1$, de modo que

$$\hat{a}_{m,h} = \frac{1}{2}(a_{m,h-1} + a_{m,h+1}) \quad (3.67)$$

Aunque podrían utilizarse otros parciales, las adyacentes suelen producir los mejores resultados, ya que sus amplitudes se correlacionan más con la amplitud de la superposición parcial. En relación con un grupo de choque como sinusoides individuales, se puede medir la amplitud mezcla a_{mz} , la frecuencia y la fase. Las amplitudes interpoladas de cada fuente pueden ser procesadas de forma que las sinusoides procesadas son el resultado de una mezcla sinusoidal con una amplitud a_{mz} . Esto puede ser obtenido por

$$\tilde{a}_i = a_{mz} \frac{\hat{a}_i}{\sum_{j=1}^J \hat{a}_j} \quad (3.68)$$

Donde $\hat{a}_j, j = 1, \dots, J$ son las amplitudes de las sinusoides interpoladas.

La exactitud de la interpolación depende, en particular en la veracidad de la estimación de la frecuencia fundamental, ya que se utiliza para determinar qué elementos parciales son interpolados. Debido a la múltiple estimación de frecuencias fundamentales resulta una tarea difícil la detección o la estimación de las frecuencias fundamentales de un armónico donde el sonido puede fallar. Los errores en la estimación de la frecuencia fundamental pueden causar grandes errores en la estimación de los parciales, por ejemplo, en la interpolación de amplitudes parciales que se superponen. Como el número de las fuentes concurrentes aumenta, el número de la superposición de parciales aumenta, y puede haber situaciones en las que la mayoría de ellos se superponen. Por lo tanto, no debería haber un método que utilice también las amplitudes medias para el componente de los grupos en superposición.

2.3.3 Método basado en el Modelado Sinusoidal.

Para lograr la eficiencia computacional y fácil aplicación de una técnica de separación basada en el modelo sinusoidal, se propone el trabajo en de dos etapas, en primer lugar el número de sonidos y sus frecuencias fundamentales se estiman, para después estimar los parámetros variables en el tiempo y realizar la separación.

Utilizando la misma notación para las variables como en la expresión (3.66), el modelo de $x(n)$ para las fuentes de M se escribe como

$$\hat{x}(n) = \sum_{m=1}^M \sum_{h=1}^{H_m} a_{m,h} \cos(2\pi f_{m,h} n + \theta_{m,h}), \quad n = 0, \dots, N - 1 \quad (3.69)$$

Cuando múltiples fuentes ($M > 1$) están presentes, cada una de ellas se modela como una parte separada del modelo de la señal. La minimización del error de reconstrucción entre la señal que se observa $x(n)$ y

el modelo $\hat{x}(n)$ en general no garantizan que la señal de cada fuente esté representada por separado como parte del modelo. Por ejemplo, si dos tonos tienen la misma frecuencia, de ellos se puede representar la energía de la señal en esa frecuencia. Por lo tanto, se imponen restricciones para los parámetros del modelo para que sea imposible que una fuente se represente con una parte de otra fuente. En la práctica, las amplitudes de la superposición de parciales se restringen de manera que no se permitan grandes cambios entre los sobretonos adyacentes de un sonido, resultando los sonidos separados en el espectro continuo.

Si se supone en (3.69) que el residual es una distribución normal de ruido blanco, la máxima estimación de mínimos cuadrados es el estimador de probabilidad de las sinusoides individuales.

2.3.4 Formulación en el dominio de frecuencia.

El método propuesto se puede formular en el tiempo y dominio de la frecuencia. Se presenta la formulación de dominio de la frecuencia, ya que permite una vista computacional de estimación eficiente de las fases por el procesamiento de lo real y lo imaginario a partir del espectro por separado. Además, cada una de las sinusoides se localiza en el dominio de la frecuencia, y por lo tanto una banda de frecuencia estrecha es suficiente para la estimación de sus parámetros. Esto lleva a una aproximación que reduce la complejidad computacional de manera significativa.

En esta sección se presenta los parámetros de (3.69) donde se estima el modelo sinusoidal para pequeños frames dada la frecuencia de estimación aproximada $\hat{f}_{m,h}$. Para activar la fase de estimación, el modelo se reescribe como

$$x(n) = \sum_{m=1}^M \sum_{h=1}^{H_m} \alpha_{m,h} \cos(2\pi f_{m,h}n) + \beta_{m,h} \text{sen}(2\pi f_{m,h}n) \quad (3.70)$$

Donde $\alpha_{m,h} = a_{m,h} \cos(\theta_{m,h})$ y $\beta_{m,h} = -a_{m,h} \text{sen}(\theta_{m,h})$

La frecuencia de transformación del modelo se desarrolla de manera que el espectro de los términos del coseno es real y el espectro de las condiciones seno es imaginario. En la práctica, esto requiere un cambio de fase de la base de la transformada de Fourier. Los principales beneficios de la operación son un menor consumo de memoria (los términos son ya real o imaginario), y menor complejidad computacional (los parámetros se pueden resolver por separado para las partes real e imaginaria).

La transformada de Fourier de una señal de valor real, que es simétrica con respecto al origen también es real. Según lo sugerido por Harris [53], la frecuencia se transforma de manera que al calcularse el índice

de tiempo $n = N / 2$ (T siendo la longitud del frame en las muestras) es considerado como el origen. Esta técnica es llamada a menudo como "ventanas de fase cero", y en la práctica puede ser aplicada utilizando la normal transformada rápida de Fourier (*Fast Fourier Transform, FFT*) y restando un término de fase lineal $\pi N \kappa / K$ del espectro resultante de la fase. Por lo tanto, la DFT de fase cero es igual a la otra DFT normal, pero la fase de la base se cambia. Después de este proceso, la DFT de cada coseno multiplicado por la función ventana es real, y la DFT de un término sinusoidal (función impar) multiplicado por $\varpi(n)$ que es imaginario. En particular, según el teorema de la modulación, la transformada de Fourier $\varpi(n)\cos(f_0 n)$ es igual a $\frac{1}{2}[W(f - f_0) + W(f + f_0)]$, y la transformada de Fourier $\varpi(n)\sin(f_0 n)$ es igual a $\frac{1}{2}[W(f - f_0) - W(f + f_0)]$, donde i es la unidad imaginaria y $W(f)$ es la transformada de Fourier $\varpi(n)$.

Aplicando el resultado anterior en (3.70), se obtiene el modelo de dominio de la frecuencia para $X(f)$, la transformada de Fourier de frame de ventana $x(n)$. Lo real y las partes imaginarias, $\Re\{X(f)\}$ y $\Im\{X(f)\}$, se puede escribir por separado, como

$$\begin{aligned}\Re\{X(f)\} &= \sum_{m=1}^M \sum_{h=1}^H \alpha_{m,h} H_{m,h}^{\Re}(f) \\ \Im\{X(f)\} &= \sum_{m=1}^M \sum_{h=1}^H \beta_{m,h} H_{m,h}^{\Im}(f)\end{aligned}\quad (3.71)$$

Donde los escalares $\alpha_{m,n}$ y $\beta_{m,h}$ son reales

$$H_{m,h}^{\Re} = \frac{1}{2} [W(f - f_{m,h}) + W(f + f_{m,h})] \quad (3.72)$$

Es real, y

$$H_{m,h}^{\Im} = \frac{i}{2} [W(f - f_{m,h}) - W(f + f_{m,h})] \quad (3.73)$$

Es imaginario, para todos los $m = 1, \dots, M$, y $h = 1, \dots, H_m$.

2.3.5 Fase de estimación.

En la primera iteración de las fases se calcula utilizando la frecuencia de las estimaciones $f_{m,h}$ dada por la estimación de la frecuencia fundamental múltiple (*Multiple Fundamental Frequency Estimation, MFFE*).

Para la estimación de las fases de la superposición de los componentes. En primer lugar, los componentes que se superponen son detectados al encontrar grupos de sinusoides donde las frecuencias

están dentro de un límite preestablecido. En la estimación de fase, cada grupo se considera como una senoide única, cuya frecuencia es la media de la superposición de frecuencias. Denotando el número total de senoideas por J .

Tomando las funciones de base en (3.72) y (3.73) por dos K -por- J matrices H_{\Re} y H_{\Im} , que contiene la parte real y la parte imaginaria de los espectros senoideas, respectivamente. Cada columna de las matrices es la DFT de una senoide individual evaluada en las frecuencias discretas $k = 1, \dots, K$. Implementando la evaluación de la DFT mediante el cálculo $W(k)$ con una alta frecuencia de resolución, y luego eligiendo los índices correspondientes a cada una de las senoideas.

El modelo para el vector del espectro $x = [X(1), \dots, X(K)]^T$ se puede escribir por separado, por su parte real x_{\Re} y parte imaginaria x_{\Im} como

$$\begin{aligned} x_{\Re} &= H_{\Re} \alpha \\ x_{\Im} &= H_{\Im} \beta \end{aligned} \quad (3.74)$$

Donde α y β son vectores, el elemento j^{th} está relacionado con la amplitud y la fase de la senoide j^{th} como $\alpha_j = a_j \cos(\theta_j)$ y $-\sin \beta_j = -a_j (\theta_j)$.

Debido a que las partes reales e imaginarias son ortogonales, la solución por mínimos cuadrados para sus parámetros se puede obtener por separado. La solución de mínimos cuadrados de (3.74) es

$$\begin{aligned} \hat{\alpha} &= (H_{\Re}^T H_{\Re})^{-1} H_{\Re}^T x_{\Re} \\ \hat{\beta} &= (H_{\Im}^T H_{\Im})^{-1} H_{\Im}^T x_{\Im} \end{aligned} \quad (3.75)$$

Las filas de H_{\Re} y H_{\Im} son linealmente independientes si dos o más senoideas no tienen frecuencias iguales. Dado que estos casos fueron detectados y agrupados con anterioridad, las matrices tienen rango completo, y en las inversas (3.75) existe.

De la solución anterior, la fase θ_j de la senoide j^{th} se obtiene como la fase de la variable compleja $(\alpha_j + i\beta_j)$. Si los componentes no se superponen, también la solución de mínimos cuadrados para las amplitudes se pueden obtener como $\hat{a}_j = \sqrt{\alpha_j^2 + \beta_j^2}$.

La solución anterior de K -por- J $H_{\mathfrak{R}}$ y $H_{\mathfrak{I}}$ requiere $(K + J/3)J^2$ operaciones flotantes de puntos (flops), tanto para $\hat{\alpha}$ y $\hat{\beta}$, lo que hace $2(K + J/3)J^2$ flops en total.

En la estimación sinusoidal de parámetros, los métodos anteriores (por ejemplo [54, 55]) resuelve las partes real e imaginaria simultáneamente, lo que requiere $4(K + 2J/3)J^2$ flops. Por lo tanto, la solución de las partes reales e imaginarias por separado reducen la complejidad de cómputo por el factor 2...4, dependiendo de la relación de K y J . Por otra parte, las matrices son reales o imaginarias, que requieren un menor número de sumas y multiplicaciones que cuando se trabaja variables complejas [54]. La complejidad computacional de las soluciones de mínimos cuadrados (3.75) se puede reducir considerablemente por una aproximación donde las funciones de base son agrupadas en subconjuntos de acuerdo con sus frecuencias, y los parámetros son resueltos por separado para cada subconjunto.

2.3.6 Estimación de la amplitud.

Esta sección presenta dos métodos alternativos para la estimación de la amplitud de las componentes que se superponen, las cuales se basan en la continuidad de la posición del espectro. El primer, que utiliza modelos lineales de las series y él un segundo que utiliza un regulador no lineal.

El modelo lineal en las series armónicas incluye la suavidad espectral como principio básico de modelo de señal, y garantiza que la suma de las amplitudes de la superposición de sinusoides es igual a la amplitud media para el grupo de sinusoides.

Las amplitudes de un sonido armónico se modelan como una suma ponderada de las funciones de base fija. Denotando las amplitudes de los armónicos m de una fuente por un vector $a_m = [a_{m,1}, \dots, a_m, H_m]^T$ para m fuentes, la forma lineal se escribe con una matriz de G_m de H -por- L , donde $H \geq L$. Cada columna de la matriz es una función base, l^{th} es la entrada de vector u_m siendo el peso de la función base l^{th} , por lo que el modelo de las amplitudes se escribe como

$$a_m = G_m u_m \quad (3.76)$$

En primer lugar se presenta cómo las amplitudes pueden resolver un determinado G_m , y a continuación son analizados los diferentes tipos de modelos.

2.3.6.1 Solución de los componentes que se superponen por mínimos cuadrados.

Con base en el teorema de modulación y la fórmula de Euler, la transformada de Fourier del coseno de

ventana $w(n) \cos(2\pi f_{m,h}n + \theta_{m,h})$ en el (3.69) puede ser escrita por

$$H_{m,h}(f) = \frac{1}{2} [e^{i\theta_{m,h}} W(f - f_{m,h}) + e^{-i\theta_{m,h}} W(f + f_{m,h})] \quad (3.77)$$

Usando la frecuencia actual y las estimaciones de fase, éstas se escriben en K -por- H_m de la matriz H_m para cada fuente m . La columna h^{th} de H_m es el espectro h^{th} de la sinusoide de la fuente m evaluada en las frecuencias discretas $k = 1, \dots, K$. El modelo para el espectro de m fuente es $H_m a_m$, y cuando se aplica el modelo lineal de (3.76), el modelo puede ser escrito como $H_m G_m u_m$. Ahora se denota $Y_m = H_m G_m$, de modo que el modelo es igual a $Y_m u_m$. El espectro m de la fuente se modela como una suma ponderada de las columnas de Y_m .

Para las fuentes de las matrices M se combinan como $Y = [Y_1, \dots, Y_M]$. Denotando las fuentes utilizando los vectores $u^T = [u_1^T, \dots, u_M^T]$, el modelo del espectro de x es igual a $x = Yu$. Dependiendo del modelo lineal, los resultados se limitan a cualquiera de los valores reales o no negativos. Las restricciones pueden ser aplicadas al reescribir el modelo como $x_{\Re \Im} = Y_{\Re \Im} u$, donde $x_{\Re \Im} = \begin{bmatrix} x_{\Re} \\ x_{\Im}/i \end{bmatrix}$, y $Y_{\Re \Im} = \begin{bmatrix} Y_{\Re} \\ Y_{\Im}/i \end{bmatrix}$, Y_{\Re} y Y_{\Im} son las partes reales e imaginarias de Y , respectivamente. Ahora la solución de mínimos cuadrados para los aumentos de valor real se obtiene como

$$\hat{u} = (Y_{\Re \Im}^T Y_{\Re \Im})^{-1} Y_{\Re \Im}^T x_{\Re \Im} \quad (3.78)$$

Cuando los beneficios se limitan a valores no negativos, por mínimos cuadrados la solución $x_{\Re \Im} = Y_{\Re \Im} u$ se obtiene mediante mínimos cuadrados no negativos [56].

Con todos los modelos, las amplitudes se resuelven de la resultante por \hat{u}_m está dado por $\hat{a}_m = G_m u_m$. Para reducir la complejidad computacional, algunos modelos (por ejemplo, el modelo defrecuencia de banda) permiten una aproximación donde las ganancias se resuelven para un subconjunto de funciones de base, como se hizo en la estimación de fase en el final de la sección. 2.3.2.

Modelo polinomial fijo: Un ejemplo sencillo de un modelo lineal es el Vandermon de matriz $[G_m]_{h,l} = f_{m,h}^{l-1}$, para las amplitudes de de los sobretonos de orden polinomial $(L - 1)^{\text{th}}$. Los resultados obtenidos con este modelo no son muy buenos. La principal razón de esto está dada porque las bases de las funciones polinómicas tienen en su mayoría altas frecuencias, mientras que los espectros de audio suelen

tener una estructura de paso bajo. Un inconveniente del modelo polinomial es que cada función de base cubre la gama de frecuencias, por lo que sería conveniente utilizar sólo los armónicos adyacentes en la interpolación.

Las características de paso bajo tienen un mejor uso en el modelo $[G_m]_{h,l} = f_{m,h}^{l+1}$, donde las amplitudes están dadas por un polinomio con potencias negativas. En el modelo polinomial el buen resultado está dado al contar también con valores negativos.

Suavización no lineal: El efecto de la superposición de parciales, o, interferencia armónica pueden ser disminuidos por el post-procesamiento de las amplitudes estimadas utilizando el método de suavizado espectral propuesto por Klapuri en [57]. En este se calcula una versión del espectro suavizado para la amplitud de cada fuente, donde si la amplitud resultante del sobretono es inferior a la amplitud original, la amplitud es sustituida por el nuevo. A diferencia de la interpolación en otros métodos, que no requiere conocer los sobretonos que se superponen. Se ocupa de todos los tonos para obtener una mejor solidez mediante la reducción de la exactitud de toda la amplitud estimada.

2.3.7 Estimación de frecuencias.

En cada iteración del procedimiento iterativo de estimación de las frecuencias, usando las estimaciones de la amplitud de corriente y la fase para linealizar la dependencia de la frecuencia en las proximidades de las estimaciones actuales, y reducir al mínimo la reconstrucción del error del modelo lineal.

La relación de cada sobretono de frecuencia $f_{m,h}$, que tiene como frecuencia fundamental $f_{m,1}$ se considera como base, de modo que la estructura armónica se mantiene. Por lo tanto, no se actualizan las frecuencias de los parciales por separado, pero sí la estructura armónica. En general, los radios de frecuencias de sobretonos en diferentes frecuencias fundamentales no son los mismos, pero para las pequeñas desviaciones de frecuencia dentro de un marco MFFE la aproximación es bastante exacta.

Para cada parcial, se escribe el error entre las frecuencias $f_{m,h}$, y la estimación actual $\hat{f}_{m,h}$ como $\Delta_{m,h} = \hat{f}_{m,h} - f_{m,h}$. Dado que los radios de frecuencia son fijos, el error de cada sobretono es linealmente dependiente del error de la frecuencia fundamental: $\Delta_{m,h} = \Delta_{m,1} f_{m,h} / f_{m,1}$

El modelo para el espectro es

$$X(f) = \sum_{m=1}^M \sum_{h=1}^H a_{m,h} H_{m,h}(f) \quad (3.80)$$

Donde

$$H_{m,h}(f) = \frac{1}{2} [e^{i\theta_{m,h}} W(f - f_{m,h}) + e^{-i\theta_{m,h}} W(f + f_{m,h})] \quad (3.81)$$

La serie de Taylor de W en cada $f \pm f_{m,h}$ es

$$\begin{aligned} W(\hat{f} \pm f_{m,h}) &= W(f \pm f_{m,h} \pm \Delta_{m,h}) \\ &= W(f \pm f_{m,h}) \pm W'_{m,h}(f \pm f_{m,h}) \Delta_{m,h} + o(\Delta_{m,h}^2) \end{aligned} \quad (3.82)$$

Donde W' denota la derivada de W , $o(\Delta_{m,h}^2)$ incluye los términos del segundo y de orden superior, los cuales son pequeños en las proximidades de la frecuencia. Hay que descartar que de $o(\Delta_{m,h}^2)$, se obtiene la aproximación lineal

$$W(f \pm f_{m,h}) \approx W(f \pm \hat{f}_{m,h}) \mp W'_{m,h}(f \pm \hat{f}_{m,h}) \Delta_{m,h} \quad (3.83)$$

Sustituyendo (3.83) a (3.81), se obtiene

$$H_{m,h}(f) \approx \hat{H}_{m,h}(f) + \hat{H}'_{m,h}(f) \Delta_{m,h} \quad (3.84)$$

Donde

$$\hat{H}_{m,h}(f) = \frac{1}{2} [e^{i\theta_{m,h}} W(f - \hat{f}_{m,h}) + e^{-i\theta_{m,h}} W(f + \hat{f}_{m,h})] \quad (3.85)$$

Y

$$\hat{H}'_{m,h}(f) = \frac{1}{2} [e^{i\theta_{m,h}} W'(f - \hat{f}_{m,h}) - e^{-i\theta_{m,h}} W'(f + \hat{f}_{m,h})] \quad (3.86)$$

Denotando la estimación del espectro actual

$$\hat{X}(f) = \sum_{m=1}^M \sum_{h=1}^H a_{m,h} \hat{H}_{m,h}(f) \quad (3.87)$$

Sustituyendo (3.84) y (3.87) a (3.80), la aproximación para el espectro se convierte en

$$X(f) \approx \hat{X}(f) + \sum_{m=1}^M \sum_{h=1}^H a_{m,h} \hat{H}'_{m,h}(f) \Delta_{m,h} \quad (3.88)$$

Ahora se puede escribir el error de la estimación del espectro en función del error de la frecuencia fundamental

$$X(f) - \hat{X}(f) \approx \sum_{m=1}^M \sum_{h=1}^H a_{m,h} \hat{H}'_{m,h}(f) \frac{\hat{f}_{m,h}}{\hat{f}_{m,1}} \Delta_{m,1} \quad (3.89)$$

El error de modelado de los espectros discretos pueden ser escritos como

$$X - \hat{X} \approx \Omega \Delta \quad (3.90)$$

Donde x y \hat{x} son el espectro observado y modelado, respectivamente, y la m^{th} columna de la matriz Ω es $\sum_{h=1}^{H_m} a_{m,h} \frac{\hat{f}_{m,h}}{\hat{f}_{m,1}} \hat{H}'_{m,h}(f)$ evaluada en frecuencias discretas $k = 1, \dots, K$, y el elemento Δ de m^{th} contiene los errores correspondientes $\Delta_{m,1}$ de la frecuencia fundamental, que es conocido.

Los derivados de $W'(f)$ en la definición (3.86) de los términos $H'_{m,h}(f)$ se calcularán del siguiente modo: la transformada de Fourier de una señal de $nw(n)$ es igual a $iW'(f)$ [58], así que W' queda como la DFT de la señal $-inw(n)$. La solución de mínimos cuadrados para valores reales del vector de error de la frecuencia fundamental Δ se obtiene como

$$\Delta = (\Omega_{\Re\Im} \Omega_{\Re\Im})^{-1} \Omega_{\Re\Im}^T (x_{\Re\Im} - \hat{x}_{\Re\Im}) \quad (3.91)$$

Donde $\Omega_{\Re\Im} = \begin{bmatrix} \Omega_{\Re} \\ \Omega_{\Im}/i \end{bmatrix}$, Ω_{\Re} y Ω_{\Im} son la parte real e imaginaria de Ω , respectivamente, y $\hat{x}_{\Re\Im} = \begin{bmatrix} \hat{x}_{\Re} \\ \hat{x}_{\Im}/i \end{bmatrix}$, \hat{x}_{\Re} y \hat{x}_{\Im} son la parte real e imaginaria del modelo del espectro, respectivamente. Las nuevas estimaciones de frecuencia para cada parcial se ofrecen como $\tilde{f}_{m,h} = \tilde{f}_{m,h} + \Delta_{m,1} \hat{f}_{m,1} / \hat{f}_{m,1}$.

En general, un número adecuado de iteraciones depende de la estimación de frecuencia la señal. Aumentando el número de iteraciones puede aumentar la calidad monótonamente cuando sólo una de las fuentes está presente. Sin embargo, cuando múltiples sonidos están presentes, los armónicos de otras fuentes cercanas pueden perturbar la estimación de la frecuencia fundamental de una fuente. En este caso, un gran número de iteraciones de estimación de frecuencia pueden causar que la frecuencia fundamental de una fuente se ajuste a una fuente equivocada.

2.3.8 Combinar las frecuencias fundamentales en notas musicales.

En los apartados anteriores se explica la estimación de parámetros en un solo frame MFFE. En los instrumentos puede ocurrir que durante una nota musical, esta se puede dividir en varios frames adyacentes MFFE. El objetivo es combinar las señales separadas en cada frame MFFE para obtener todas las notas musicales.

Dado que la variación de la frecuencia fundamental de una nota musical es a menudo relativamente pequeña. Utilizando un método para estimar las notas utilizando modelos ocultos al modelar las

características MFFE. Se calcula un conjunto de notas, cada una de ellas tiene una frecuencia fundamental, tiempo de latencia. En lugar de las estimaciones de frecuencia fundamental primas, se utilizan las frecuencias fundamentales de los acontecimientos teniendo en cuenta el método de mínimos cuadrados no lineales. Dado que en el método de estimación de notas no se producen estimaciones de las frecuencias de los sobretonos, se asume que están en armonía.

Las notas separadas se pueden clasificar en las fuentes de sonido. Dado que el rendimiento de separación de los algoritmos existentes es muy limitado, no hay estudios donde la separación y el agrupamiento se realicen de forma totalmente automática con solo señales acústicas. Every[59] utiliza el reconocimiento de patrones y la agrupación de enfoque en un material guiando la separación por una frecuencia MIDI.

2.3.9 Tratamiento no lineal por mínimos cuadrados.

Como el objetivo está en la estimación de parámetros de las fuentes de acumulación, la simulación se concentra sólo en la propuesta del algoritmo no lineal por mínimos cuadrados dentro de un frame MFFE.

2.3.9.1 Algoritmos.

Los algoritmos de separación discutidos permiten un gran número de variaciones y en las distintas aplicaciones. Las principales diferencias entre los algoritmos están en la estimación de los parciales de superposición.

Los algoritmos evaluados son los siguientes:

- **Filtrado peine.** En este método la amplitud de una senoide se determina directamente desde el espectro de amplitud de la mezcla a la frecuencia parcial. Así, el método no incluye un mecanismo para resolver armónicos superpuestos. La aplicación del método se basa en el modelo sinusoidal.
- **Interpolación lineal y normalización.** Interpolación lineal de acuerdo con (2.3.2) es el método más utilizado, y aquí está seguido por la normalización, ya que proporciona una mejora.
- **Suavizado no lineal.** Aquí se aplica el filtro no lineal de Klapuri[57] en el espectro de amplitud de cada sonido para reducir el efecto de la superposición de componentes.

En el filtrado de peine se marca la diferencia entre la relación señal/distorsión (RSD) de segmentación y la RSD normal con respecto a los restantes algoritmos. A pesar de que el filtrado de peine no tiene un mecanismo de solución de superposición de armónicos, los otros métodos no producen un mejor

promedio de RSD de segmentación. Esto está dado porque en parte del procedimiento de evaluación: el filtro de peine ajusta la amplitud de una senoide de manera que la amplitud de la mezcla es igual a la frecuencia de la senoide. Así en el filtrado de peine pues se muestra el menor error entre la RSD de segmentación y la RSD normal.

2.4 Principios matemáticos.

En esta sección se tratan los principios matemáticos necesarios para entender los planteamientos hechos. De este modo, serán introducidas algunas de las propiedades básicas de las variables aleatorias y otros términos empleados.

2.4.1 Definición de variables aleatorias.

Una variable aleatoria es una función que asigna un valor numérico a cada suceso elemental del espacio muestral. Es decir, una variable aleatoria es una variable cuyo valor numérico está determinado por el resultado del experimento aleatorio. Matemáticamente es una aplicación

$$X : \Omega \rightarrow \mathbb{R} \quad (4.0)$$

Que da un valor numérico, del conjunto de los reales \mathbb{R} , a cada suceso en el espacio Ω de los resultados posibles del experimento. Las variables aleatorias son usualmente designadas por letras mayúsculas (ej. X, Y, \dots), y un valor particular de una variable aleatoria es denominado por una letra minúscula (ej. x, y, \dots). [60, 61]

2.4.2 Función de distribución acumulada.

La función de distribución acumulada $F_X(x_0)$ corresponde a la probabilidad de una variable aleatoria real X tome un valor numérico menor o iguala x_0 , ó representa la suma de las probabilidades hasta alcanzar el valor de interés. Simbólicamente, lo anterior se expresa como:

$$F_X(x_0) = P(X \leq x_0) \quad \text{para} \quad x_0 \in \mathbb{R} \quad (4.1)$$

La función de probabilidad acumulada $F_X(x)$ cumple con las siguientes propiedades:

1. $F_X(x_0)$ es no negativo:

$$F_X(x_0) \geq 0 \quad \forall x_0 \in \mathbb{R}.$$

2. F_X es no decreciente:

$$F_X(x_1) \leq F_X(x_2) \quad \forall x_1 \leq x_2, \quad x_1, x_2 \in \mathbb{R}.$$

3. El recorrido de la función de probabilidad acumulada es:

$$0 \leq F_X(x_0) \leq 1 \quad \forall x_0 \in \mathbb{R}.$$

4. El valor de la función de probabilidad acumulada cuando el valor de la variable es demasiado grande se acerca a uno:

$$F_X(\infty) = 1.$$

5. El valor de la función de probabilidad acumulada cuando el valor de la variable es demasiado pequeño se acerca a cero:

$$F_X(-\infty) = 0.$$

Su representación gráfica tiene la forma escalonada, siendo coincidentes los saltos con las probabilidades p_i correspondientes a los valores x_i de la variable X . [62]

2.4.3 Función de densidad de probabilidad.

La función de densidad de probabilidad (fdp) p_X de una variable continua X es una función que se integra para obtener la probabilidad que la variable aleatoria toma un valor en un intervalo predefinido: [63]

$$\int_{x_1}^{x_2} p_X(t) dt = P(x_1 \leq X \leq x_2). \quad (4.2)$$

La fdp está relacionada con la función de distribución de probabilidad como sigue:

$$p_X(x_0) = \left. \frac{dF_X(x)}{dx} \right|_{x=x_0}. \quad (4.3)$$

De manera análoga, se puede relacionar de forma inversa ambas funciones:

$$F_X(x_0) = \int_{-\infty}^{x_0} p_X(t) dt. \quad (4.4)$$

La fdp cumple con las siguientes propiedades:

1. p_x es no negativo

$$p_X(x_0) \geq 0 \quad \forall x_0 \in \mathbb{R}.$$

2. El recorrido de la fdp es

$$0 \leq p_X(x_0) \leq 1 \quad \forall x_0 \in \mathbb{R}.$$

3. La fdp está normalizada

$$\int_{-\infty}^{\infty} p_X(t) dt = 1.$$

2.4.4 Vectores aleatorios.

En la práctica, la naturaleza aleatoria de un proceso puede necesitar ser descrito por más de una variable aleatoria. Por eso, se define un vector aleatorio como sigue: X_1, \dots, X_n variables aleatorias definidos sobre el mismo espacio probabilístico. El vector ordenado $[X_1, \dots, X_N]$ se llama variable aleatoria N-dimensional o vector aleatorio de dimensión N .

Más formalmente, un vector aleatorio es una función con dominio del espacio probabilístico Ω y de destino \mathbb{R}^N :

$$[X_1, \dots, X_N]: \Omega \rightarrow \mathbb{R}^N \quad (4.5)$$

2.4.5 Funciones de Distribución y densidad de vectores Aleatorios.

En el caso multivariable se define la función de distribución acumulada y la función de densidad de probabilidad en analogía con el caso univariable:

Sea $X = [X_1, \dots, X_N]$ un vector aleatorio real y sean X_1, \dots, X_N variables reales aleatorias de un mismo

experimento estadístico. La función de distribución acumulada X se define como

$$F_X(x_1, \dots, x_N) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_N \leq x_N). \quad (4.6)$$

De acuerdo con esta definición la función de probabilidad de un vector aleatorio puede ser determinada en la ecuación (4.3) si se sustituyen las derivadas totales por derivadas parciales respecto a cada una de las variables[64]:

$$p_{\mathbf{X}}(x_{1_0}, x_{2_0}, \dots, x_{N_0}) = \frac{\partial F_{\mathbf{X}}}{\partial x_1} \Big|_{x_1=x_{1_0}} \frac{\partial F_{\mathbf{X}}}{\partial x_2} \Big|_{x_2=x_{2_0}} \cdots \frac{\partial F_{\mathbf{X}}}{\partial x_N} \Big|_{x_N=x_{N_0}} \quad (4.7)$$

2.4.6 Función de Distribución y Densidad conjunta y marginales.

Los conceptos anteriores pueden ser extendidos también a problemas con dos vectores aleatorios diferentes \mathbf{X} y \mathbf{Y} donde la dimensión m del vector \mathbf{Y} puede ser diferente que la dimensión n del vector \mathbf{X} . Si se concatenan los dos vectores para obtener un “supersector” $\mathbf{Z} = [\mathbf{X}, \mathbf{Y}]$ las fórmulas precedentes pueden ser utilizadas directamente. En tal escenario se define la función de distribución conjunta

$$F_{\mathbf{X},\mathbf{Y}}(x_0, y_0) = P(\mathbf{X} \leq x_0, \mathbf{Y} \leq y_0) \quad (4.8)$$

y la función de densidad conjunta

$$p_{\mathbf{X},\mathbf{Y}}(x_0, y_0) = \frac{\partial F_{\mathbf{X},\mathbf{Y}}(x, y)}{\partial x} \Big|_{x=x_0} \frac{\partial F_{\mathbf{X},\mathbf{Y}}(x, y)}{\partial y} \Big|_{y=y_0} \quad (4.9)$$

donde x_0 e y_0 son un vector constante de valores concretos del conjunto de variables que forman \mathbf{X} e \mathbf{Y} , respectivamente. Asimismo, la función de distribución se deriva de la función de densidad conjunta

$$F_{\mathbf{X},\mathbf{Y}}(x_0, y_0) = \int_{-\infty}^{x_0} \int_{-\infty}^{y_0} p_{\mathbf{X},\mathbf{Y}}(\mu\nu) d\mu d\nu. \quad (4.10)$$

Dado la función de densidad conjunta las dos funciones de densidad de probabilidad individuales $p_{\mathbf{X}}(x)$ de \mathbf{X} y $p_{\mathbf{Y}}(y)$ de \mathbf{Y} , llamadas funciones de densidad marginales, pueden ser también obtenidas[65]. Por eso, se integra la función de densidad conjunta sobre uno de los dos vectores aleatorios como sigue

$$p_{\mathbf{X}}(x) = \int_{-\infty}^{\infty} p_{\mathbf{X},\mathbf{Y}}(x, \mu) d\mu \quad (4.11)$$

$$p_{\mathbf{Y}}(y) = \int_{-\infty}^{\infty} p_{\mathbf{X},\mathbf{Y}}(\nu, y) d\nu \quad (4.12)$$

2.4.7 Esperanzas.

En esta sección serán expuestas las esperanzas de variables y vectores aleatorios. Estas esperanzas se utilizan frecuentemente en la práctica ya que pueden ser estimadas de observaciones de variables aleatorias e incluso si la fdp de dichas variables no se conoce de manera exacta.

Generalmente, la esperanza de una magnitud $\mathbf{g}(\mathbf{x})$ derivada del vector aleatorio \mathbf{X} se define como:

$$E\{\mathbf{g}(\mathbf{x})\} = \int_{-\infty}^{\infty} \mathbf{g}(\mathbf{x})p_{\mathbf{X}}(\mathbf{x})d\mathbf{x}. \quad (4.13)$$

Aquí la magnitud $\mathbf{g}(\mathbf{x})$ puede ser una cantidad escalar, un vector o una matriz.

Un caso especial de las esperanzas son los momentos de un vector aleatorio. Estos momentos se obtienen si $\mathbf{g}(\mathbf{x})$ es el resultado del producto de componentes del vector aleatorio \mathbf{X} . Generalmente, el momento de orden n de un vector aleatorio definido como

$$\mathbf{m}_{\mathbf{X}}^{\{n\}} = \int_{-\infty}^{\infty} \mathbf{x}^n p_{\mathbf{X}}(\mathbf{x})d\mathbf{x}. \quad (4.14)$$

En detalle, el primer momento $\mu_{\mathbf{X}} = \mathbf{m}_{\mathbf{X}}^{\{1\}}$ corresponde al vector media del vector \mathbf{X} mientras que el segundo momento define la correlación $\mathbf{R}_{\mathbf{X}}$ entre pares de componentes del vector aleatorio $\mathbf{X} = [X_1, \dots, X_N]^T$:

$$\mathbf{R}_{\mathbf{X}} = (r_{ij})_{N \times N} = \mathbf{m}_{\mathbf{X}}^{\{2\}} = E\{\mathbf{X}\mathbf{X}^T\} \quad (4.15)$$

donde

$$r_{ij} = E\{X_i X_j\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j p_{X_i, X_j}(x_i, x_j) dx_j dx_i. \quad (4.16)$$

Esta matriz $\mathbf{R}_{\mathbf{X}}$ tiene algunas propiedades importantes:

1. $\mathbf{R}_{\mathbf{X}}$ es simétrica.
2. $\mathbf{R}_{\mathbf{X}}$ es semidefinida positiva.
3. Todos los valores propios de $\mathbf{R}_{\mathbf{X}}$ son reales y no negativos.
4. Todos los valores propios de $\mathbf{R}_{\mathbf{X}}$ son reales y no ortogonales.

A menudo, la media está restando a los vectores aleatorios antes que los momentos están determinados. Estos momentos se nombran momentos centrados y son solamente significativos si sus ordenes son más grandes que 1. Correspondiendo al momento de segundo orden, correlación, se denomina al momento centrado de segundo orden covarianza. La correspondiente matriz de covarianza $\mathbf{C}_{\mathbf{X}}$ se calcula como sigue:

$$\underline{\mathbf{C}}_X = (c_{ij})_{N \times N} = E\{(X - \mu_X)(X - \mu_X)^T\} \quad (4.17)$$

donde los elementos

$$c_{ij} = E\{(X_i - \mu_i)(X_j - \mu_j)\} \quad (4.18)$$

Se llaman covarianzas. Esta matriz $\underline{\mathbf{C}}_X$ tiene las mismas propiedades que la matriz $\underline{\mathbf{R}}_X$.

Para el caso especial de una sola variable X el vector media se reduce al valor medio $\mu_X = E\{X\}$, la matriz de correlación se reduce al momento segundo $E\{X^2\}$ y la matriz de covarianza se reduce a la variación de X .

$$\sigma_X^2 = E\{(X - \mu_X)^2\}. \quad (4.19)$$

Además, la raíz σ_X de la variación se denomina desviación estándar[66].

2.4.8 Distribuciones de Probabilidad Importantes.

En este apartado se exponen algunas de las funciones de densidad de probabilidad más importantes. Estas funciones se utilizan con mucha frecuencia en el procesamiento de señales.

2.4.8.1 Distribución Uniforme.

Las distribuciones uniformes corresponden al experimento de elegir dos puntos al azar dos fijos m y n . Como la probabilidad de elegir cualquier punto es la misma, la función de densidad tendrá la misma altura en todos los puntos entre m y n , es decir se trata de una función constante desde m a n de la altura $1/(n - m)$ [63]

$$p_X(x) = \begin{cases} \frac{1}{n-m} & \text{si } x \in [m, n] \\ 0 & \text{resto} \end{cases} \quad (4.20)$$

Aquí, la misma coincide con el punto medio del segmento $[m, n]$

$$\mu_X = \frac{n - m}{2},$$

y la varianza se determina como

$$\sigma_X^2 = \frac{(n - m)^2}{12}.$$

2.4.8.2 Distribución normal.

La distribución normal multivariable, también llamada de Gauss o distribución gaussiana, tiene algunas propiedades específicas que la hacen única entre todas las funciones de densidad de probabilidad.

Sea \mathbf{X} un vector aleatorio de dimensión n . \mathbf{X} se llama vector aleatorio gaussiano si su función de probabilidad de densidad está determinada por la siguiente expresión:

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det \underline{\mathbf{C}}_{\mathbf{X}})^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^T \underline{\mathbf{C}}_{\mathbf{X}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})\right) \quad (4.21)$$

Donde $\det \underline{\mathbf{C}}_{\mathbf{X}}$ es el determinante de $\underline{\mathbf{C}}_{\mathbf{X}}$. La matriz de covarianza $\underline{\mathbf{C}}_{\mathbf{X}}$ se supone definida positiva implicando que su inversa existe. Por un vector gaussiano \mathbf{X} puede ser demostrado que

$$E\{\mathbf{X}\} = \boldsymbol{\mu}_{\mathbf{X}}, \quad E\{(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^T\} = \underline{\mathbf{C}}_{\mathbf{X}} \quad (4.22)$$

Por eso, $\boldsymbol{\mu}_{\mathbf{X}}$ y $\underline{\mathbf{C}}_{\mathbf{X}}$ se llaman vector media y matriz de covarianza respectivamente, de la distribución normal multivariada. Además, el conocimiento de $\boldsymbol{\mu}_{\mathbf{X}}$ y $\underline{\mathbf{C}}_{\mathbf{X}}$ ya es suficiente para definir completamente la fdp gaussiana, es decir, que momentos de orden superior no proporcionan ninguna información sobre dicha fdp [63, 67].

La razón más importante para la popularidad de la fdp gaussiana proviene del teorema del límite central que indica que la distribución de la suma de las variables aleatorias tiende a una distribución gaussiana. Más formalmente, sea X_k la suma de la secuencia de variables aleatorias idénticamente distribuidas $\{Z_i\}$:

$$X_k = \sum_{i=1}^k Z_i. \quad (4.23)$$

Como la media μ_{X_k} y la covarianza σ_{X_k} de X_k puede crecer sin límite en el caso de $K \rightarrow \infty$, se considera la variable normalizada Y_k en lugar de X_k :

$$Y_k = \frac{X_k - \mu_{X_k}}{\sigma_{X_k}}. \quad (4.24)$$

En este caso el teorema del límite central afirma que la distribución de Y_k converge a una distribución normal con media cero y varianza unidad cuando $K \rightarrow \infty$.

Este teorema se extiende fácilmente a vectores aleatorios idénticamente distribuidos Z_k con media μ_x y C_x matriz común. En este caso, la distribución del vector aleatorio

$$Y_k = \frac{1}{\sqrt{k}} \sum_{i=1}^k (Z_i - \mu_Z) \quad (4.25)$$

Converge a una distribución gaussiana con media cero y matriz de covarianza C_z .

Este teorema es la razón de que se asuma que muchas de las variables observadas en el mundo real sigan una distribución gaussiana. Al ser la suma de un conjunto suficiente de otras variables desconocidas, sin importar sus respectivas funciones de densidad de probabilidad, se puede asegurar que la variable que corresponde a la suma sigue una distribución similar a la normal.

Las consecuencias del teorema del límite central en BSS son profundas, puesto que se puede asumir que, con un número suficiente de señales participando en la mezcla, ésta tenderá a ser gaussiana. Por eso, un primer método auditivo para recuperar las señales que dieron origen a la mezcla es tratar de alejar las distribuciones de las estimaciones lo más posible de la distribución gaussiana, es decir, aumentar la no-gaussianidad.

2.4.8.3 Distribución de Laplace.

Otra distribución importante en el procesamiento de datos es la distribución de laplace:

$$p_X(x) = \frac{\lambda}{2} \exp(-\lambda|x|) \quad (4.26)$$

Esta distribución con su grande concentración de valores en el centro y sus alrededores (suponiendo que la media de la señal en cuestión es precisamente cero) se usa a menudo para describir las señales de voz[68].

2.4.9 Correlación y Covarianza Cruzadas.

Basándose en la función de densidad conjunta (ver ecuación 3.64) las operaciones de esperanzas pueden también ser extendidos a funciones $g(\mathbf{x}, \mathbf{y})$ de dos vectores aleatorios \mathbf{X} y \mathbf{Y} como sigue:

$$E\{g(\mathbf{x}, \mathbf{y})\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\mathbf{x}, \mathbf{y}) p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \quad (4.27)$$

Análogamente, se definen la matriz de correlaciones cruzadas las covarianzas cruzadas respectivamente como:

$$\begin{aligned} \mathbf{R}_{\mathbf{X}, \mathbf{Y}} &= E\{\mathbf{X}\mathbf{Y}^T\}, \\ \mathbf{C}_{\mathbf{X}, \mathbf{Y}} &= E\{(\mathbf{X} - \boldsymbol{\mu}_X)(\mathbf{Y} - \boldsymbol{\mu}_Y)^T\} \end{aligned} \quad (4.28)$$

La medida más habitualmente utilizada para cuantificar la correlación cruzada entre dos variables aleatorias \mathbf{X} y \mathbf{Y} es el coeficiente de correlación lineal de Pearson $r_{Pearson}$. El coeficiente de Pearson mide el grado de asociación lineal entre dos variables cualesquiera, y puede calcularse dividiendo la covarianza de ambas entre el producto de las desviaciones típicas de las dos variables:

$$r_{Pearson}(X, Y) = \frac{C_{X,Y}}{\sigma_X \sigma_Y}. \quad (4.29)$$

$r_{Pearson}(X, Y)$ tiene las propiedades siguientes:

1. El recorrido de $r_{Pearson}(\mathbf{X}, \mathbf{Y})$ es

$$-1 \leq r_{Pearson}(X, Y) \leq 1 \quad \forall X, Y \in \mathbb{R}$$

$r_{Pearson}(\mathbf{X}, \mathbf{Y})$ es positivo si existe una relación directa entre ambas variables y es negativo si la relación es inversa.

2. Un valor de $r_{Pearson}(\mathbf{X}, \mathbf{Y})$ de +1 ó -1 indica una relación lineal perfecta entre \mathbf{X} y \mathbf{Y} , mientras que un valor 0 indica que no existe una relación lineal.

Finalmente, se puede ver que un valor de cero de $r_{Pearson}(\mathbf{X}, \mathbf{Y})$ no indica necesariamente que no exista correlación entre \mathbf{X} y \mathbf{Y} , ya que las variables pueden presentar una relación no lineal[68].

2.4.10 Estimación de Funciones Estadísticas.

Las correlaciones y covarianzas se emplean como herramientas estadísticas para determinar las posibles dependencias lineales entre las distintas variables aleatorias. Hasta ahora, todas las fórmulas de

Las correlaciones y covarianzas se emplean como herramientas estadísticas para determinar las posibles dependencias lineales entre las distintas variables aleatorias. Hasta ahora, todas las fórmulas de momento se han determinado basándose en el conocimiento de la función de densidad de la probabilidad. Como ya se comentaba al principio del (2.4.7), normalmente no se conoce esta función. No obstante, es posible realizar una estimación de estos valores si se conoce un conjunto x_1, x_2, \dots, x_k de valores que toma el conjunto de variables que forman el vector aleatorio \mathbf{X} [68].

En este caso, tienen que ser distinguidos dos estimadores diferentes dependiendo de que se trate de un momento normal o con un momento centrado. En primer caso, el de momentos normales de orden n , se usa el estimador

$$\frac{1}{K} \sum_{k=1}^K x_i^n \quad (4.30)$$

Mientras que el primer caso de momentos centrados de orden n se aplicaría la función

$$\frac{1}{K-1} \sum_{k=1}^K (x_i - \mu_x)^n \quad (4.31)$$

Por ejemplo, el estimador de $\mu_x = E\{X\}$ es

$$\hat{\mu}_X = \frac{1}{K} \sum_{k=1}^K x_k \quad (4.32)$$

y la covarianza cruzada entre dos variables aleatorias \mathbf{X} y \mathbf{Y} se determina como

$$\hat{c} = \frac{1}{K-1} \sum_{k=1}^K (x_k - \mu_X)(y_k - \mu_Y). \quad (4.33)$$

2.4.11 Decorrelación y Blanqueado.

Se dice que dos vectores \mathbf{X} e \mathbf{Y} están incorrelados si su matriz de covarianzas cruzadas, $y(\mathbf{x}, \mathbf{y})$, es la matriz nula. De forma equivalente, podemos decir que dichos vectores \mathbf{X} e \mathbf{Y} están incorrelados si su matriz de correlaciones cruzadas es $\mathbf{R}_{\mathbf{x}, \mathbf{y}} = \mu_x \mu_y$.

Refiriéndose al concepto de decorrelación entre los componentes de un mismo vector aleatorio \mathbf{X} , se refiere a la matriz de covarianzas definida en la ecuación (4.16). Obviamente, cada uno de los

componentes X_i de \mathbf{X} está perfectamente correlado consigo mismo, por lo que la condición de decorrelación relativa a la matriz de covarianza es

$$\mathbf{X} \text{ está decorrelado si } \mathbf{C}_{\mathbf{X}}(X_1, \dots, X_N)^T = \begin{bmatrix} \sigma_{X_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{X_2}^2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sigma_{X_n}^2 \end{bmatrix} \quad (4.34)$$

donde $\sigma_{X_i} = E \{X_i - \mu_{X_i}\}$ corresponde al momento central de segundo orden, o sea, a la varianza.

Como caso particular, se dice que un vector aleatorio está blanqueado si su media es cero para todos sus componentes y su matriz de covarianzas es la matriz identidad \mathbf{I} . Al imponer que la media sea cero, lógicamente la matriz de correlación coincide con la de covarianza.[68]

$$\mathbf{X} \text{ está blanqueado si } \mu_{\mathbf{X}} = 0, \mathbf{C}_{\mathbf{X}} = \mathbf{R}_{\mathbf{X}} = \mathbf{I}. \quad (4.35)$$

El blanqueo de las señales suele ser un paso habitual de pre-procesamiento de los datos en el análisis de componentes independiente, en muchos casos necesario para realizar estimaciones correctas de independencia.

2.5 Conclusiones

En este capítulo se analizaron las diferentes técnicas que permiten obtener como resultado la separación de las fuentes de sonido en una mezcla acústica con el menor error posible. Sobre las técnicas basadas en el aprendizaje no supervisado, *la separación no supervisada utilizando PCA y NMF* es la propuesta obtenida luego del análisis de cada una de las técnicas dentro de esta categoría, debido a que la combinación de estas dos técnicas no sólo arroja como consecuencia la separación de fuentes de sonido sino que a la vez permite llevar a cabo un coste computacional razonable. Por otra parte también es propuesta una técnica de separación basada en el modelo sinusoidal de las señales, basándose precisamente en que el modelo sinusoidal es el más utilizado en aplicaciones musicales ya que este tiene en cuenta la armoniosidad de las fuentes sonoras, esto figura en su utilidad para la separación de instrumentos musicales.

Capítulo 3: Validación de la Propuesta.

3.1 Introducción.

Las técnicas de separación de fuentes de sonido tienen como objetivo brindar la posibilidad de obtener a partir de una media cada uno de los sonidos por sus fuentes. En el presente capítulo se realizará una evaluación de las propuestas descritas en el apartado anterior teniendo en cuenta las aplicaciones directas dentro de los productos informáticos que se desarrollan en el Departamento de Señales Digitales. Con esta evaluación se pretende proporcionar una representación del avance del trabajo así como de la eficiencia de la propuesta realizada.

3.2 Método para la validación de la propuesta.

Para realizar la evaluación técnica de las propuestas se utilizó el método de Experto, que es el que permite, según las decisiones tomadas por expertos, tomar o rechazar determinadas propuestas. Este grupo de expertos, seleccionados por su amplio conocimiento en el ámbito que se desarrolló la investigación, serán los encargados de analizar las convergencias de opiniones en torno al problema que aborda la investigación.

Paso1: Se elabora los criterios de evaluación de acuerdo a las características de la propuesta y organizarlos por grupos.

◆ **Grupo No. 1:** Criterio de mérito científico.

- Valor científico de la investigación.
- Calidad de la investigación.
- Novedad científica.

◆ **Grupo No. 2:** Valoración de la propuesta.

Validación de la Propuesta

- La propuesta de la combinación de las técnicas PCA y NMF es correcta.
- La propuesta de técnica de separación basada en el modelo sinusoidal es correcta.

◆ **Grupo No. 3:** Criterios de implantación.

- Necesidad real de implantación de la propuesta.
- Posibilidad de ser aplicada la propuesta.
- Integración de la propuesta con el proceso de desarrollo de software (PDSW).

◆ **Grupo No. 4:** Criterios de impacto.

- Repercusión en los proyectos productivos.
- Posibilidad de aplicación.
- Impacto en el centro para el cual está destinado.

Paso 2: Se le asigna un peso relativo a cada criterio de acuerdo al porcentaje que representa y los intereses a evaluar, sumando un valor total de 100.

- Grupo No. 1.....30.
- Grupo No. 2.....20.
- Grupo No. 3.....25.
- Grupo No. 4.....25

Paso 3: Se organiza un comité de expertos con una cantidad de 4 expertos teniendo en cuenta su especialidad, grado científico y currículum de cada uno.

Paso 4: Se entrega a cada experto la propuesta para que sea estudiada, y evaluada a través de dos modelos, en el primero se valorará el peso relativo de cada criterio y en el segundo se realizará una evaluación cuantitativa de dichos criterios en una escala del 1- 5. Seguidamente se registra una apreciación cualitativa de la propuesta de excelente, bueno, aceptable, cuestionable y malo; así como consideraciones a dicha propuesta.

Validación de la Propuesta

Paso 5: Calcular por cada criterio el peso promedio, partiendo de los pesos dados por los expertos. El peso promedio por cada criterio se muestra en la siguiente tabla:

- Sea C el número de criterios que van a evaluarse.
- E el número de expertos que realizan la evaluación.
- G: es el peso del grupo al que pertenecen los criterios.

| G | C/E | E ₁ | E ₂ | E ₃ | E ₄ | ΣE | E _p |
|----|-----------------|----------------|----------------|----------------|----------------|-----|----------------|
| 30 | C ₁ | 10 | 8 | 8 | 9 | 35 | 8.75 |
| | C ₂ | 11 | 15 | 12 | 15 | 53 | 13.2 |
| | C ₃ | 9 | 7 | 10 | 6 | 32 | 8 |
| 20 | C ₄ | 11 | 12 | 11 | 10 | 44 | 11 |
| | C ₅ | 9 | 8 | 9 | 10 | 36 | 9 |
| 25 | C ₆ | 7 | 8 | 8 | 7 | 30 | 7.5 |
| | C ₇ | 10 | 8 | 8 | 10 | 36 | 9 |
| | C ₈ | 8 | 9 | 9 | 8 | 34 | 8.5 |
| 25 | C ₉ | 7 | 8 | 7 | 8 | 30 | 7.5 |
| | C ₁₀ | 8 | 10 | 9 | 10 | 37 | 9.25 |
| | C ₁₁ | 10 | 7 | 9 | 7 | 33 | 8.25 |
| T | | 100 | 100 | 100 | 100 | 400 | 100 |

E₁: Ing. Reinier Pupo Ruiz.

E₂: Ing. Jean Michael Suarez Perez.

E₃: Msc. Yurdik Cervantes Mendoza.

E₄: Ing. Edmis Deivis Semanat Aldana.

Paso 6: Se verifica la consistencia en el trabajo de los expertos, para lo que se utiliza el coeficiente de concordancia de Kendall y el estadígrafo Chi cuadrado (χ^2). Se sigue el procedimiento siguiente:

◆ Para cada criterio se determina:

- ΣE: Sumatoria del peso dado por cada experto.
- E_p: Puntuación promedio del peso dado por cada experto.
- MΣE: Media de los ΣE ($M\Sigma E = \Sigma E/C_n$).
- ΔC: Diferencia entre ΣE y MΣE.

Validación de la Propuesta

- ◆ Se determina la desviación de la media, que posteriormente se eleva al cuadrado para obtener la dispersión (S) por la expresión: $S = \Sigma (\Sigma E - \Sigma \Sigma E / C)^2$.
- ◆ Conociendo la dispersión se puede calcular el coeficiente de concordancia de Kendall (W). Dado por la siguiente expresión: $W = (S / E^2) / (C^3 - C) / 11$.
- ◆ Calcular el Chi cuadrado real conociendo el valor del coeficiente de Kendall teniendo en cuenta la siguiente expresión: $X^2 = E (C - 1) W$.

A continuación se muestran los datos obtenidos luego de realizar los pasos anteriores

| C/E | E ₁ | E ₂ | E ₃ | E ₄ | ΣE | E _p | ΔC | ΔC ² |
|-----------------|----------------|----------------|----------------|----------------|-----|----------------|-------|-----------------|
| C ₁ | 10 | 8 | 8 | 9 | 35 | 8.75 | 1.36 | 1.8496 |
| C ₂ | 11 | 15 | 12 | 15 | 53 | 13.25 | 16.14 | 260.4996 |
| C ₃ | 9 | 7 | 10 | 6 | 32 | 8 | 4.36 | 19.0096 |
| C ₄ | 11 | 12 | 11 | 10 | 44 | 11 | 7.64 | 58.3696 |
| C ₅ | 9 | 8 | 9 | 10 | 36 | 9 | 0.36 | 0.1296 |
| C ₆ | 7 | 8 | 8 | 7 | 30 | 7.5 | 6.36 | 40.4496 |
| C ₇ | 10 | 8 | 8 | 10 | 36 | 9 | 0.36 | 0.1296 |
| C ₈ | 8 | 9 | 9 | 8 | 34 | 8.5 | 2.36 | 5.5696 |
| C ₉ | 7 | 8 | 7 | 8 | 30 | 7.5 | 6.36 | 40.4496 |
| C ₁₀ | 8 | 10 | 9 | 10 | 37 | 9.25 | 1.36 | 1.8496 |
| C ₁₁ | 10 | 7 | 9 | 7 | 33 | 8.25 | 3.36 | 11.2896 |
| T | 100 | 100 | 100 | 100 | 400 | 100 | 50.5 | 434.026 |
| MΣE | 36.36 | | | | | | | |
| W | 0.226 | | | | | | | |
| X ² | 9.04 | | | | | | | |

Posteriormente, se compara el x^2 real, con el valor del dato estadístico, siendo $\alpha = 0.01$ y $C = 10$ y debe cumplirse que $x^2_{\text{real}} < x^2_{(\alpha, c-1)}$ para que el trabajo realizado por los expertos sea valorado de consistente. $x^2_{(\alpha, c-1)} = x^2_{(0.01, 10)} = \mathbf{23.2093}$.

Por tanto $9.04 < 23.2093$, quedando demostrada la consistencia en el trabajo realizado por los expertos.

Validación de la Propuesta

Paso No. 7: Una vez comprobada la consistencia del trabajo de expertos se puede determinar el nivel de aceptación de la propuesta entre los expertos $P \times c$ donde “P” es el peso relativo de cada criterio y c es la calificación promedio dada por los expertos.

P: peso promedio de cada criterio.

$$P = (E_p)/100$$

c: calificación promedio de cada criterio concebida por los expertos.

| Criterios | Calificación(c) | | | | | P | P*c |
|-----------------|-----------------|---|---|---|---|--------|--------|
| | 1 | 2 | 3 | 4 | 5 | | |
| C ₁ | | | | X | | 0.0875 | 0.35 |
| C ₂ | | | | | X | 0.1325 | 0.6625 |
| C ₃ | | | | X | | 0.08 | 0.32 |
| C ₄ | | | | X | | 0.11 | 0.44 |
| C ₅ | | | | X | | 0.09 | 0.36 |
| C ₆ | | | | X | | 0.075 | 0.3 |
| C ₇ | | | | X | | 0.09 | 0.36 |
| C ₈ | | | | X | | 0.085 | 0.34 |
| C ₉ | | | | X | | 0.075 | 0.3 |
| C ₁₀ | | | | X | | 0.0925 | 0.37 |
| C ₁₁ | | | | X | | 0.0825 | 0.33 |

Paso 8: Se calcula el Índice de Aceptación (IA) de la propuesta.

Se calcula a partir de la siguiente fórmula teniendo en cuenta que $\sum(P * c) = 4.1325$

$$IA = (P * c) / 5$$

Si $(P * c) = 4.1325$ entonces $IA = 4.1325 / 5$ y se obtiene que $IA = 0.8265$.

Paso 9: determinar la probabilidad de éxito de la propuesta a partir de los rangos predefinidos del índice de aceptación.

$IA > 0,7$Existe alta probabilidad de éxito.

$0,7 > IA > 0,5$Existe probabilidad media de éxito.

$0,5 > IA > 0,3$Existe baja probabilidad de éxito.

$0,3 > IA$No existe probabilidad de ningún éxito.

Dado el resultado de IA es igual a **0.8265** entonces se puede concluir que la probabilidad de éxito es: ALTA; y se establece por los expertos una valoración final de: BUENO.

3.3 Indicaciones para la implementación.

Las técnicas propuestas son capaces de ofrecer datos de gran relevancia y con una alta capacidad de producir resultados fácilmente interpretables lo que trae consigo que las mismas presentan una complejidad computacional elevada y además requiere de una representación e interactividad adecuada para que la interpretación de los resultados que produce sea efectiva. Por ello la mayoría de las implementaciones solo están disponibles en MATLAB (de "matrix laboratory" (laboratorio matricial)) que es a su vez un entorno interactivo y un lenguaje de programación de alto nivel[74].

Como lenguaje permite construir herramientas reusables que se pueden agrupar en Toolbox, que son conjuntos de archivos que amplían el entorno de Matlab y que sirven para trabajar en problemas particulares como pueden ser defensa (reconocimiento de patrones, procesamiento de señales, compresión de datos), electrónica (modelado no lineal, síntesis de voz, visión por ordenador), economía (análisis financiero, análisis predictivo), industria (control de procesos, identificación en tiempo real), medicina, robótica (control de trayectorias, sistemas de visión) ,reconocimiento y síntesis del habla, telecomunicaciones (control de datos e imágenes, servicios de información automatizada, traducción del lenguaje hablado en tiempo real, diagnóstico, sistemas de enrutamiento), etc.

Por su funcionalidad, Matlab permite realizar tareas de cálculo complejas de forma más rápida que con los lenguajes de programación tradicionales, y además está disponible para las plataformas Unix, Windows y Mac OS X[74].

Como las técnicas propuestas presentan un alto orden estadísticos se puede apreciar como las Statistics Toolbox le da un rango ancho de herramientas para la realización de estos cálculos estadísticos donde se puede trabajar directamente con funciones de probabilidad en el tema, tales como las distribuciones de probabilidad y de estadística descriptiva que son principios matemáticos fundamentales discutidos. Además para el cálculo matricial y vectorial se puede hacer un buen uso de Matlab con funciones especiales y elementales que representa, dígame matriz identidad y otras matrices elementales.[75]

Como uno de los objetivos fundamentales de separar las diferentes fuentes de sonido está en poder obtener las voces. El trabajo en Matlab permite el análisis de voz propiamente donde el estudio se refiere

a la acústica en si como un modo de generar modelos que van a estar acoplados en a las necesidades de la investigación en cuestión. Por otra parte en la representación y trabajo de señales para lo que se emplea la transformada de Fourier que es una herramienta matemática que permite pasar de una representación en el dominio temporal a una representación en el dominio frecuencial de modo que la Transformada de Fourier es un procedimiento matemático que mapea cualquier señal analógica estacionaria a una serie infinita de sinusoides, cada una de ellas con una determinada amplitud y fase[76].

De forma general Matlab permite realizar:

- Cálculos intensivos desde un punto de vista numérico.
- Gráficos y visualización avanzada.
- Lenguaje de alto nivel basado en vectores, arreglos y matrices.
- Colección muy útil de funciones de aplicación.

3.4 Conclusiones.

Basándose en las respuestas emitidas por los expertos se puede concluir en que la propuesta realizada cumple con todas las metas trazadas en un porcentaje alto por lo que el resultado de la validación por este método es positivo.

- La propuesta de la combinación de las técnicas de Análisis de Componentes Principales (PCA) y Factorización no Negativa de Matrices (NMF) fue valorada por los expertos como más recomendable para su implementación.
- Al aplicar el método y analizar los resultados se obtuvo una probabilidad de éxito alta indicando que la aplicación de la propuesta proporcionará resultados favorables y que lo planteado hasta el momento brinda un aporte significativo capaz de resolver los problemas existentes.

Una vez utilizado el Método de Expertos y por ende, validadas las actividades y tareas, se llega a la conclusión de que las propuestas desarrolladas en el trabajo investigativo pueden ser aplicadas para ampliar las funcionalidades que brindan los diferentes productos desarrollados en el Departamento de Señales Digitales. Demostrando así la factibilidad y utilidad de las propuestas para lograr una mayor calidad en el PDSW de los diferentes productos que se realizan en el ámbito del procesamiento de sonido digital.

Conclusiones

La presente investigación cumple satisfactoriamente con su objetivo general que ha sido determinar la técnica adecuada para gestionar la separación automática de fuentes de sonido para el procesamiento de media en el Departamento de Señales Digitales de la facultad 9.

Los principales resultados obtenidos en este trabajo se relacionan a continuación:

- Se realizó un detallado estudio del proceso de Separación Ciega de Fuentes de Sonido, se analizaron los diferentes modelos en los que se basa BSS con el objetivo de comprender como funcionan estas técnicas.
- Se analizó cada una de las técnicas que logran separar las señales por sus fuentes, teniendo en cuenta su estructura y características.
- Se propuso la separación no supervisada utilizando PCA y NMF como técnica basadas en el aprendizaje no supervisado. También es propuesta una técnica de separación basada en el modelo sinusoidal de las señales, basándose precisamente en que el modelo sinusoidal es el más utilizado en aplicaciones musicales ya que este tiene en cuenta la armoniosidad de las fuentes sonoras.
- Se llevó a cabo la evaluación de los criterios reconocidos por el Método de Expertos, evidenciando la aplicabilidad de la propuesta mediante un índice de aceptación con una probabilidad de éxito alta donde los expertos denotan una mayor aceptación a la combinación de las técnicas PCA y NMF debido al menor coste computacional que ofrece su implementación.

Recomendaciones

Una vez cumplido los objetivos trazados en la investigación se recomienda:

- Desarrollar un sistema capaz de gestionar la separación automática de fuentes de sonido con la práctica de las técnicas propuestas.
- Que la propuesta desarrollada sea utilizada dentro del Departamento de Señales Digitales para validarla en la práctica.
- Hacer un estudio más profundo de las diferentes técnicas que permiten la separación automática de fuentes de sonido.

Referencias Bibliográficas

1. F. Rojas, C.G.P., *Separación Ciega de Señales: un nuevo enfoque mediante Algoritmos Genéticos*. febrero de 2002.
2. COMON, P., *Séparation de mélanges de signaux* in *I-Revues*. 12-16 junio de 1989: Gretsi Conferencia (Francia)
3. Autores, C.d. *Superposición de ondas*.2003 [citado 13/02/2010]; Disponible: <http://www.ehu.es/acustica/index.html>.
4. Raúl Alcaraz, C.S., José Joaquín Rieta, Marcos D. Fernández, José A. Ballesteros, *Separación Ciega de Fuentes en el Dominio Wavelet. Aplicación a señales de audio.*, U.d.C.L.M. E.U.P. Cuenca, Editor: Valencia.
5. Autores, C.d., *Christian Jutten*. Hindawi Publishing Corporation, 2010.
6. IEEE, T.I., *The IEEE Workshop on Neural Networks for Signal Processing*. September 2-4 2009: Grenoble, France.
7. Hyvärinen, A., *Survey on Independent Component Analysis*, in *Neural Computing Surveys*. 1999: Finland.
8. Fernando J. Mato Méndez, M.S.S. (20 - 22 de Octubre de 2008) *SUSTRACCIÓN ESPECTRAL DE RUIDO EN SEPARACIÓN CIEGA DE FUENTES DE RUIDO DE TRÁFICO*.
9. Oliveri, P.F.J. *Reflexiones sobre la técnica en Aristóteles*. Grupo de Investigación Máthesis de la Universidad Nacional de Mar del Plata 2000 [citado 23 de febrero de 2010; Disponible: url:<http://www.favanet.com.ar/mathesis/>].
10. Hilda, *Técnica concepto*, in *De Conceptos*. 17 Octubre, 2008.
11. Saco, A. (2003 -2004) *Apuntes de sonido digital*.
12. Autores, C.d., *El sonido*, in *DigitalFotored*. 2005.
13. Autores, C.d., *Harry Nyquist*.
14. SonidoZero. *Fuentes de sonido 2003-2005* [citado 14/01/2010]; Disponible: <http://www.sonido-zero.com>.
15. Autores, C.d. *Fuente de sonido*. 2001 [citado 14 de enero de 2010]; Disponible: <http://www.ceidis.ula.ve>.
16. P.Comon (2009) *Pierre COMON*. Universié Nice,
17. Autores, C.d., *Jean-François Cardoso* Hindawi Publishing Corporation, 2010.

18. J Cardoso, P.C., *Independent Component Analysis, a survey of some algebraic methods*. 1996: Paris.
19. Autores, C.d., *Fred Attneave HAHP*(The Archives de History of American Psychology), septiembre 2009.
20. Elizabeth Mazzola & Vince Vales (May 2007) *Archives of the History of American Psychology* Archives of the History of American Psychology
21. David L. G. Noakes, J.R.B., *THE FAMILY THAT GEORGE BUILT*. ANIMAL BEHAVIOR SOCIETY, 2005 -2010.
22. Clemente, R.M., *Ecuaciones lineales para el problema de la separación ciega de fuentes.*, in *Departamento de Ingeniería Electrónica*. 30 de junio de 2000, Universidad de Sevilla: Sevilla.
23. Pedro A. Carrión Pérez, P.A.C.P., José Joaquín Rieta Ibáñez, Juan Ródenas García, *Procesado de señales biomédicas*, U.d.C.L. Mancha, Editor. 2007.
24. Stadlthanner, K., ed. E.d.I.U.d. Granada. 2006, Granada.
25. Carlos Vayá, J.J.R., David Moratal-Pérez, César Sánchez. *Rendimiento de los Algoritmos BSS Convolutivos en el Estudio de la Fibrilación Auricular*. 2008 [citado 2010; Disponible: <http://w3.iec.csic.es/>].
26. Virtanen, T., . *Sound Source Separation in Monaural Music Signals*. 2006, Tampere University of Technology: Tampere.
27. E. Urrestarazu, J.I., *El análisis de componentes independientes (ICA) en el estudio de señales electroencefalográficas*, in *Departamento de Neurología*. 30 - 05 - 2005, Universidad de Navarra: Navarra.
28. Juliana Muñoz, J.R., Edison Duque. *Principal and independent component analysis applied to noise reduction in electrocardiographic signals*. in *Scientia et Technica Año XIV*. septiembre 2008. Pereira: Universidad Tecnológica de Pereira.
29. Puntonet, C., *Sistema para separación ciega de señales*. 1999, Universidad de Granada: Santa Lucía.
30. Alpaydin, E., *Introduction to machine learning. Adaptive computation and machine learning*. ilustrada ed. 2004: MIT Press, 2004. 415 páginas.
31. Ryszard S. Michalski, R.S.M., Gheorghe Tecuci, *Machine learning: a multistrategy approach*. revisada ed. Vol. Volumen 4 de Machine Learning, a Multistrategy Approach, Jerzy W. Bala. 1994: Morgan Kaufmann, 1994. 782 páginas.
32. F. Herrera, J.C.R., J.S. Aguilar (2004) *Red Española de Minería de Datos y Aprendizaje*. Universidad de Sevilla .

33. Geoffrey E. Hinton, T.J.S., *Unsupervised learning: foundations of neural computation*. ilustrada ed. 1999, Estados Unidos: Bradford Books. 398 páginas.
34. Jolliffe, I.T., ed. *Principal Component Analysis*. Segunda ed., ed. S.V. INC. 2002,1986, 2, ilustrada: New York. 487.
35. Gurrea, M.T., *Análisis de componentes principales*, Universitat Oberta de Catalunya (UOC): Catalunya.
36. Galeano, P., in *Estadística III*. 2009-2010, Universidad Carlos III de Madrid: Madrid.
37. Krzanowski, W.J., *Principles of multivariate analysis: a user's perspective*. 2, revisada, ilustrada ed. 1988,2000, New York: Oxford University Press, 2000. 586.
38. Dunteman, G.H., *Principal Component Analysis*, ed. 07-069. 1989: Sange Publications,Inc. 96.
39. A. Rodríguez Ruíz, C.G.L., *CARACTERÍSTICAS MORFO-FUNCIONALES DE Chirostoma Consocium, Ch.Chapalae Y Ch. Ocotlanae DEL LAGO DE CHAPALA, MÉXICO*, in *Departamento de Ecología*. 5 de abril de 1988, Universidad de Sevilla: Sevilla, España.
40. Oja, A.H.a.E., *Independent Component Analysis: Algorithms and Applications*, in *Neural Networks Research Centre*. 2000, Helsinki University of Technology: Helsinki,Finland.
41. Bravo Alcobendasn, B., *CLASIFICACIÓN NO SUPERVISADA DE DOCUMENTOS*, in *División de Ingeniería Informática y Telecomunicación*. septiembre 2008, Universidad San Pablo- CEU. p. 135.
42. Daniel D. Lee, H.S.S., *Algorithms for Non-negative Matrix Factorization*, in *Nature*. 1999.
43. Virtanen, T., *Non-Negative Matrix Factorization And Its Application to Audio*. 2008, Virtanen, Tuomas: Tampere University of Technology.
44. Sajda, P.D., Shuyan; Parra, Lucas, *Recovery of constituent spectra using non-negative matrix factorization*, in *Department of Biomedical Engineering, Adaptive Image and Signal Processing*,, Columbia University, Sarno® Corporation: New York, Princeton.
45. Zdunek, R.K., Raul, Hori Gen, He Zhaohui, *Extended SMART Algorithms for Non-Negative Matrix Factorization*, in *BSI, Laboratory for Advanced Brain Signal Processing; Amari Research Unit for Mathematical Neuroscience*: Wako-shi.
46. D. Lee, D.H.S., *Sebastian Algorithms for Non-negative Matrix Factorization*, in *Nature*. 1999.
47. Mariam de la Iglesia Vayá, *Análisis de series temporales mediante ICA*, Universidad Plitécnica de Valencia: Valencia.
48. José L. García Sarasa, G.P.d.I.V., *Desarrollo de un sistema web para el análisis de datos en bioinformática utilizando técnicas de factorización ositiva de matrices.*, in *SISTEMAS INFORMÁTICOS*. 2006/2007, UNIVERSIDAD COMPLUTENSE DE MADRID: Madrid. p. 57.

49. Zaballa, I., *Valores singulares. ¿Qué son?.¿Paraqué sirven?*, in *Departamento de Matemática Aplicada y EIO*. 2006, Universidad del País Vasco: País Vasco.
50. Serra, X., *Musical Sound Modeling with Sinusoids plus Noise*. 1997, Audiovisual Institute, Pompeu Fabra University: Barcelona, España.
51. S, B.A., *Auditory scene analysis: the perceptual organization of sound*. 2, ilustrada, reimpresión ed. WV 272 B833a. 1994: MIT Press, 1994. 773.
52. Horner Fletcher, N.R., Thomas D., *The physics of musical instruments*. 2, ilustrada ed. 1998, New York: Springer. 756.
53. Hartmann, W.M., *Signals, sound, and sensation*. ilustrada ed. BF251.H35. 1997, New York: Springer, 1997. 647.
54. Press, W.H.V., William T.; Teukolsky, Saul A.; Flannery, Brian P., *Numerical recipes in C: the art of scientific computing*. 2 ed, ed. N. 10011-4211. Vol. Volumen 2. 1992, New York: Cambridge University Press. 994.
55. Depalle, P.H., T. , *Extraction of spectral peak parameters using a short-time Fourier transform modeling and no sidelobe windows*. 1997, New Paltz-EUA: In Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,
56. Lawson, C.L.H., Richard J., *Solving least squares problems*. 3, ilustrada ed. Vol. Classics in applied mathematics. 1995, New Jersey SIAM. 337.
57. Klapuri, A.P., *Multiple fundamental frequency estimation based on harmonicity and spectral smoothness*. 2003: IEEE Transactions on Speech and Audio Processing.
58. Pavlovich Tolstov, G.S., Richard A., *Fourier series*. ilustrada ed. 1976, New Jersey: Courier Dover Publications. 336.
59. Every, M.R., *Separation of musical sources and structure from single-channel polyphonic recordings*, in *Department of Electronics*. febrero, 2006, University of York: New York. p. 215.
60. Autores, C.d. *Definición de variable aleatoria*. [citado 2010 8 de febrero]; Disponible: <http://estadistica.bio.ucm.es/>.
61. David Ruiz Muñoz, A.M.S.S., *Apuntes de Estadística*, J.C.M. Coll, Editor. 2006.
62. Autores, C.d. *FUNCION DE PROBABILIDAD ACUMULADA*. octubre 2004 [citado 2010 6 de febrero]; Disponible: <http://www.virtual.unal.edu.co>.
63. Autores, C.d. (2001) *Cálculo aplicado a probabilidad y estadística*.
64. Navarro, J.d.I.H., ed. *Estadística aplicada*. 3, ilustrada ed., ed. Ediciones Díaz de Santos. 2003, Ediciones Díaz de Santos, 2003 358 páginas.
65. Martínez Barbeito, J., ed. *Introducción al cálculo estocástico*. 2003, Netbiblo, 2003: España. 315.
66. Vicente Quesada Paloma, A.G.P., ed. *Lecciones de cálculo de probabilidades*. ilustrada ed. 1988, Ediciones Díaz de Santos, 1988: Madrid. 475 páginas.