



EMPLEO DE LA TRANSFORMACIÓN DE FOURIER FRACCIONAL
PARA EXTRAER CARACTERÍSTICAS DISCRIMINANTES EN EL
RECONOCIMIENTO DE LOCUTORES

IDRISS TYLER SANDOVAL VILLAMIZAR
Ingeniera Electrónica

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
MAESTRÍA EN INGENIERÍA ELECTRÓNICA
BUCARAMANGA
2012



EMPLEO DE LA TRANSFORMACIÓN DE FOURIER FRACCIONAL
PARA EXTRAER CARACTERÍSTICAS DISCRIMINANTES EN EL
RECONOCIMIENTO DE LOCUTORES

IDRISS TYLER SANDOVAL VILLAMIZAR
Ingeniera Electrónica

Trabajo de Investigación presentado como requisito para optar al título de
Magíster en Ingeniería Electrónica.

Director
Dr. Yezid Torres Moreno
Profesor Titular

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE
TELECOMUNICACIONES
MAESTRÍA EN INGENIERÍA ELECTRÓNICA
BUCARAMANGA
2012

A mis padres y hermanos
+ *A Doña Ceci, y A Don Antonio*

AGRADECIMIENTOS

A todos y a cada una de las personas que en el transcurso de mi proceso de formación estuvieron aportando conocimientos relacionados con el tema de investigación.

A mi familia, quienes han demostrado en todos estos años su apoyo incondicional.

A mi director Yezid Torres Moreno por su orientación, dedicación y aportes que hicieron posible este trabajo.

A Jaime Meneses por permitirme hacer parte del grupo GOTS, y brindarme la oportunidad de dar a conocer el proyecto a cada uno de los miembros.

A los que trabajaron y trabajan con la Fraccional. A los integrantes del grupo de Óptica y Tratamiento de Señales GOTS por todas sus contribuciones.

RESUMEN

TÍTULO: EMPLEO DE LA TRANSFORMADA DE FOURIER FRACCIONAL PARA EXTRAER CARACTERÍSTICAS DISCRIMINANTES EN EL RECONOCIMIENTO DE LOCUTORES*

AUTOR: IDRIS TYLER SANDOVAL VILLAMIZAR**

PALABRAS CLAVES: FRFT, reconocimiento de hablantes, reconocimiento de voz, procesamiento de señal

CONTENIDO: En este trabajo de investigación se presenta un estudio de las características discriminantes en el reconocimiento de locutores (o hablantes), las cuales están representadas en dominios mixtos tiempo-frecuencia donde se requiere identificar las particularidades que permiten reconocer a los hablantes empleando una nueva alternativa de representación de la señal de voz en el espacio de Fourier de orden fraccional.

En la parte inicial se aborda lo concerniente al reconocimiento de locutores y sus generalidades. Posteriormente se introduce la forma integral de la transformada de Fourier fraccional (FRFT) y sus propiedades más relevantes, y se presentan los resultados experimentales del cálculo de la FRFT de un centenar de órdenes simultáneos en cada una de las señales de voz. Esta representación se realiza en seis (6) locutores, tres (3) hombres y tres (3) mujeres, quienes realizan la fonación de diferentes palabras de habla inglesa suministradas de una base de datos específica y con una frecuencia de muestreo de 16 kHz. A partir de una representación tiempo-orden en coordenadas rectangulares se obtiene una imagen de correlación en donde se extraen los parámetros relación señal a ruido SNR, relación altura de pico lóbulo lateral PSR, y la relación energía valor pico de correlación PCE, los cuales permiten evaluar las características métricas de la discriminación en los locutores.

* Trabajo de Investigación.

** Facultad de Ingeniería Físico-Mecánicas. Maestría en Ingeniería Electrónica, Director: Dr. Yezid Torres M.

ABSTRACT

TITLE: USING THE FRACTIONAL FOURIER TRANSFORM TO EXTRACT DISCRIMINATING FEATURES IN SPEAKER RECOGNITION*

AUTHOR: IDRIS TYLER SANDOVAL VILLAMIZAR**

KEYWORDS: Fractional Fourier Transform, Speaker recognition, voice recognition, signal processing

CONTENT: This research is aimed at studying the discriminating characteristics on the recognition, which are represented in mixed time-frequency domain where it is required to identify the characteristics that enable recognition of the speakers. Here we propose a new alternative representation of the voice signal in the fractional Fourier space, trying to identify a set of new features in advance cannot be evaluated with the use of standard Fourier Transform.

The first part, deals with a fast review to recognition of speakers and general aspects. Subsequently we introduce the integral form of the fractional order Fourier transform (FRFT) and its most important properties, and we present experimental results of the FRFT calculus of a hundred orders simultaneous each of the voice signals. This representation is done in six (6) speakers, three (3) men and three (3) women who undertake the phonation of English of different words supplied from a specific database and a sampling frequency of 16 kHz. From time-order representation in rectangular coordinates we obtain a set of correlation image parameters. Are extracted signal to noise ratio SNR, Peak sidelobe ratio PSR and Peak to Correlation energy PCE parameters, this set is the metric to assess the discrimination characteristics for the speakers.

* MsC. Research Work

** Physics Mechanical Engineering Faculty. Electrical, Electronic and Telecommunication Engineering School.
Director: Dr. Yezid Torres M.

ÍNDICE GENERAL

INTRODUCCIÓN	14
CAPITULO 1	17
RECONOCIMIENTO DEL LOCUTOR.....	17
1.1 Aplicaciones del reconocimiento del locutor.....	19
1.2 Extracción de características	19
1.3 Coeficientes LPC.....	20
1.4 Coeficientes cepstrales de frecuencia mel	21
1.5 Métodos de reconocimiento de locutores	21
1.5.1 Reconocimiento de locutores por cuantificación vectorial VQ	22
1.5.2 Reconocimiento de locutores por NN (del inglés, “nearest neighbour”), o vecino más cercano	22
1.5.3 Reconocimiento de locutor mediante Redes Neuronales	23
1.5.4 Modelos ocultos de Markov (“HMM, Hidden Markov Models”)	23
CAPITULO 2	25
TRANSFORMADA DE FOURIER FRACCIONAL (FRFT).....	25
2.1 Generalidades	25
2.2 Definición	26
2.3 Propiedades	29
2.4 Representación tiempo-orden para la FrFT	37
2.5 Algoritmo de cálculo	38
CAPITULO 3	41
PRUEBAS EXPERIMENTALES	41
3.1 Análisis bajo la transformada de Fourier fraccional	46
3.2 Correlación y autocorrelación	49
3.3 Correlación de un mismo locutor	50
CAPITULO 4	52
PARÁMETROS DE CORRELACIÓN.....	52
4.1 Parámetro relación señal a ruido (del inglés SNR”signal to noise ratio”)...	52
4.2 Relación altura de pico lóbulo lateral PSR (“Peak sidelobe ratio”).....	54
4.3 Relación pico a energía de correlación PCE(“Peak to Correlation energy”)55	
CONCLUSIONES	60
REFERENCIAS	62

LISTA DE ANEXOS

Anexo A

A1. Representación de los MFCC.....	71
-------------------------------------	----

Anexo B

B1. Sistemas lineales especiales con sus núcleos de transformación.....	77
B2. Ejemplo de correlación.....	81
B3. Tablas resultados de correlación.....	84

LISTA DE FIGURAS

Fig. 1	Forma de cálculo de los LPC.....	20
Fig. 2	Forma de cálculo de los coeficientes cepstrales	21
Fig. 3	Reconocimiento del locutor por cuantificación vectorial.....	22
Fig. 4	Reconocimiento de Locutores por NN.....	23
Fig. 5	Forma de cálculo de los HMM Aplicación de HMMs a un sistema de verificación de locutor	24
Fig. 6	Funciones rectángulo, Módulo y Fase.....	30
Fig. 7	Módulo de la FRFT normalizada al valor máximo de la función rectángulo de las funciones senoidales de la figura precedente.	31
Fig. 8	Señal seno con variación en las frecuencias izquierda 1,1 Hz. y derecha 0,31 Hz. .	31
Fig. 9	Módulo de la FRFT normalizando al valor máximo de las señales sinusoidales de la figura precedente.	32
Fig. 10	Fases de la transformada de Fourier Fraccional de 100 órdenes de forma simultánea en funciones seno de la figura anterior.	32
Fig. 11	Señal rectángulo, abajo escalada la de arriba por un factor $M=3$	33
Fig. 12	Módulos de la FRFT de la función rectángulo con orden 0,75. La de abajo es la de arriba escalada por un factor $M=3$	33
Fig. 13	Fases de la FRFT de la función rectángulo con orden 0,75, la de abajo es la de arriba escalada por un factor $M=3$	34
Fig. 14	Función Rectángulo con diferentes valores de M , definidos en las ecuaciones precedentes en el texto.....	35
Fig. 15	FrFT de las funciones rectángulo escaladas en el eje del tiempo. Arriba señal original y a la derecha el caso $M=3$, abajo izquierda $M=1/3$ y a la der. $m=-3$	35
Fig. 16	Funciones seno con diferentes valores de M , para distintos escalamientos en el tiempo.....	36
Fig. 17	Módulos de la FrFT de las funciones seno, arriba izq. frecuencia: 0,017 Hz. Der. frecuencia 5,52 mHz. Abajo izq. frecuencia: 0,027 Hz. Der. frecuencia 5,58 mHz	37
Fig. 18	Etapas del sistema general utilizado.....	38
Fig. 19	Cálculo de la FrFT del centenar de órdenes	38
Fig. 20	Seudocódigo implementado para la FrFT.....	40
Fig. 21	Señal de voz de una de las frases para seis locutores	41
Fig. 22	Señal de voz, una fonación particular de los seis locutores, las señales de la derecha pertenecen a los hombres y las de la izquierda a las mujeres.....	43
Fig. 23	Histograma de la longitud de las señales de voz seleccionadas de la base EUSTACE.....	44
Fig. 24	Ejemplo de señal de voz interpolada y diezmada a un valor de $3/2$	45
Fig. 25	Ejemplo del error en una señal, en rojo señal diezmada a partir de la señal interpolada y en azul señal original.....	45
Fig. 26	Calculo de la FrFT para 101 órdenes de una señal tomada de la base de datos seleccionada.	46
Fig. 27	Representación Tiempo–Orden del módulo de la FrFT de las señales de voz mostradas en la Fig. 22.....	47
Fig. 28	Módulo de la función de autocorrelación de una señal de voz femenina	48
Fig. 29	Módulo de la correlación cruzada de una señal femenina y otra masculina.....	48
Fig. 30	Autocorrelación arriba izquierda y correlaciones cruzadas con las señales de los otros cinco locutores.....	50

Fig. 31	Fonación de una misma palabra pronunciada por una misma persona en diferentes momentos.....	50
Fig. 32	Correlación de la misma palabra en un locutor para diferentes realizaciones en momentos distintos.....	51
Fig. 33	Correlación de la misma palabra en un locutor, se correlaciona la primera realización con las demás.	51
Fig. 34.	(a.) Representación Tiempo–Orden del módulo de la FrFT con 10 órdenes. (b) Correlación de la primera realización con las demás. Arriba a la izquierda la autocorrelación de la primera realización.....	57
Fig. 35.	(a.) Representación Tiempo–Orden del módulo de la FrFT con 25 órdenes. (b) Correlación de la primera realización con las demás. Arriba a la izquierda la autocorrelación de la primera realización.....	58
Fig. 36.	(a.) Representación Tiempo–Orden del módulo de la FrFT con 50 órdenes. (b) Correlación de la primera realización con las demás. Arriba a la izquierda la autocorrelación de la primera realización.....	59

LISTA DE TABLAS

Tabla 1 SNR de la palabra design.....	54
Tabla 2 SNR de la palabra Jhon.....	54
Tabla 3 PSR de la palabra design.....	54
Tabla 4 PSR de la palabra Jhon.....	56
Tabla 5 PCE de la palabra design.....	56
Tabla 6 PCE de la palabra Jhon.....	56

INTRODUCCIÓN

La transformada de Fourier de orden fraccional FrFT, es una representación en el dominio mixto tiempo-frecuencia, que permite un análisis más detallado de la señal, y donde la transformada de Fourier estándar TF, es sólo un caso particular. Esta herramienta matemática es empleada en diversas áreas de la ciencia, para las señales de voz representadas en este espacio se puede afirmar que es un tema de continuo desarrollo.

Actualmente la investigación del reconocimiento del habla usando la transformada de Fourier fraccional está en sus inicios, por lo que son modelos en los que se experimenta todavía y que han dado cierta motivación con las investigaciones en curso.

La Universidad Industrial de Santander y su Grupo de Investigación en Óptica y Tratamiento de Señales (GOTS), ha venido realizando trabajos referentes a la FrFT, en los que se citan: “Holografía Numérica por Transformación de Fourier Fraccional” [39], “Encriptación Óptica de Señales usando la FrFT” [40], y “Óptica Fraccional de Fourier” [9], entre otros [16-17]; pero no se había empleado en las señales de voz.

A nivel internacional se encuentran algunos artículos que emplean la FrFT en el reconocimiento de voz; en la referencia [1] se plantea un método de representación de la señal de voz, en donde se utilizan los coeficientes cepstrales en la frecuencia Mel (MFCC) como método de extracción de características. Un estudio de estos coeficientes combinados con el tono para los sistemas de reconocimiento de locutores se proponen en las referencias [23,27]; en la primera se desarrolla un modelo de funciones de probabilidad conjunta relacionadas con el tono y los vectores de características. De forma similar en la referencia [34] se propone un estudio de las características acústicas del habla basadas en

los MFCC. Otra investigación que tiene que ver con estos temas pero aplicado no en el tono sino en la identificación del lenguaje se referencia en [36].

Al igual se propone en la referencia [25] un método novedoso, el cual representa la señal de voz, basándose en la FrFT; se argumenta que las características individuales derivadas de la FrFT proporcionan un eficiente reconocimiento, en donde el cálculo eficaz del proceso de extracción de características lleva a niveles aceptables en comparación con los parámetros MFCC. Este artículo inicia con algunos fundamentos de la FrFT, presenta dos tipos de métodos de definición de características denominados: “Mean Energy Within Critical Bands (MECB)”, y “Differential Mean Energy within Critical Bands (DMECB)”. Estas características emplean diferentes órdenes de la FrFT. En la parte experimental se emplea independientemente del texto un conjunto cerrado; la base de datos de la voz está en el lenguaje Mandarin, y cuenta con secciones grabadas de 47 locutores hombres. Otro método, en este caso, de frecuencia instantánea aplicada al reconocimiento de voz aparece en la referencia [22], este método se utiliza para producir un nuevo tipo de representación tiempo-frecuencia.

En la referencia [20] se encuentra una monografía que muestra ciertas representaciones bidimensionales de la señal de voz; tales como la representación tipo Gabor, representación Cepstral, representación de la clase de Cohen, entre otras. Asimismo, en la referencia [21] se abordan representaciones tiempo- frecuencia con aplicación a la voz.

Con relación a la extracción de características según los métodos convencionales, en la referencia [24] se discute una técnica que compensa las diferencias, en las características de los locutores.

En áreas similares a la voz, se encuentran artículos relacionados con acústica entre ellos se pueden citar en la referencia [31] un estudio de señales de sonar en donde es empleada la distribución Choi Williams arrojando resultados eficientes. De los mismos autores en la referencia [32] se aborda un estudio previo a los resultados obtenidos con señales de sonar,

donde en esta oportunidad se emplea la distribución de Wigner. Dentro de esta línea de investigación se encuentra en la referencia [38] una caracterización de canales bajo el agua, usando la transformada de Fourier fraccional se expresa la señal recibida como una suma de sinusoides cuyas frecuencias se relacionan directamente con los tiempos de llegada.

Una de las aplicaciones encaminadas hacia la separación del ruido y la señal de voz se propone en la referencia [30], y en la referencia [33] se estudia una nueva técnica de filtrado adaptativo para suprimir ruido no estacionario. El filtrado en un dominio fraccional se referencia en [7] y diversos tópicos que involucran la FrFT en las referencias [10-15, 28,29].

La investigación en temas relacionados con la transformada de Fourier fraccional se ha incrementado en los últimos años; son diversas las áreas de aplicación que se han propuesto por citar un ejemplo en protección de derechos de autor [35] o en imágenes [37] y en ocasiones se ha finalizado obteniendo muy buenos resultados.

Aunque en el reconocimiento de locutores se revele una gran expectativa en las aplicaciones, existen ciertos problemas que esperan ser solucionados.

CAPITULO 1

RECONOCIMIENTO DEL LOCUTOR

Sin lugar a dudas, la voz es la forma más eficaz, natural, en que los seres humanos se comunican. Existen situaciones en donde la comunicación se realiza hombre-máquina, resultando económico, sencillo y práctico para las personas.

Dependiendo de la aplicación, la máquina (o computador), deberá confirmar previamente que el usuario que realiza la comunicación esté autorizado para llevarla a cabo. Es aquí, donde se aborda el problema de *reconocimiento del locutor*, presente en aplicaciones de tipo identificación y/o verificación.

El reconocimiento del locutor (o hablantes) es un proceso de identificación y/o verificación de una persona que realiza la fonación, se fundamenta en las características individuales presentes en la señal del habla y son diferentes a las del resto de los locutores.

Una de las dificultades en el reconocimiento de hablantes con relación al reconocimiento de voz estriba en la inexistencia de una característica acústica especial que permita una identificación directa.

Los métodos de reconocimiento de personas mediante la voz son una forma de identificación biométrica que posee ciertas ventajas como: la imposibilidad de sufrir pérdidas y olvidos, es decir, no se memoriza ninguna clave, existe una elevada complejidad de duplicación la cual brinda una mayor seguridad a los sistemas biométricos.

No obstante las desventajas del uso de la voz consiste en que las cualidades físicas son variables y de fácil modificación por el método de transmisión. Si se da el caso en que se admita un sistema con amplia variación de voz se podría aceptar un locutor erróneo. Por

esta razón resulta esencial que las cualidades físicas de la voz, sean estables, y que no se puedan imitar por terceros o se afecten por características de transmisión.

Algunos de los factores de las fuentes de variación de la señal de voz entre locutores se deben a la fisiología del hablante, al modo de hablar y al vocabulario utilizado.

Las características fisiológicas se manifiestan en la envolvente espectral, en las propiedades de las fuentes de la voz y en su comportamiento dinámico. La calidad y el tono de la voz se relacionan con las características estáticas y temporales de la envolvente espectral y con el “pitch” (o tono).

En la clasificación de los sistemas de reconocimiento del locutor debido a la dependencia o no del texto pronunciado se encuentra:

Sistema dependiente del texto: requiere que el hablante pronuncie las mismas palabras o frases usadas en el entrenamiento del sistema, los resultados de este tipo de sistemas varían en función del texto empleado.

Sistema independiente del texto: proceso de identificación o verificación que funciona correctamente para cualquier texto usado en los procesos de entrenamiento o test [26], posee una complejidad o tasa de error superior a la dependiente del texto.

Según como se determine el conjunto de locutores existen dos grupos:

- Conjunto cerrado: identifica a un locutor entre un conjunto de locutores registrados; cuanto mayor sea el número de locutores posibles, mayor será la probabilidad de error.
- Conjunto abierto: al sistema puede acceder cualquier persona registrada o no, se decide si un locutor determinado pertenece a un conjunto de locutores conocidos.

En la verificación del locutor los resultados se expresan en función de los tipos de errores:

Tasa de falsa aceptación: es la probabilidad de dar como válido a un impostor, es decir, verificar afirmativamente un locutor diferente de quien dice ser.

Tasa de falso rechazo: probabilidad de dar como no válido a un usuario correcto, es decir, verificar negativamente un locutor que realmente es quien dice ser.

1.1 Aplicaciones del reconocimiento del locutor

En sistemas de seguridad tales como: el control de acceso a recintos, la adquisición de informaciones restringidas, entre otros, hacen parte de la aplicabilidad que posee la identificación y verificación de locutores.

Otras de las aplicaciones a mencionar están, *el reconocimiento a sospechosos*, en este caso mediante grabaciones (pueden ser policiales) se identifica a qué persona pertenece la voz en la grabación; y *la transcripción automática de reuniones*, que consiste en un conjunto de conferencistas de los cuales se transcriben frases habladas y se indaga las identidades de las personas que hicieron la oratoria.

Una ventaja que presenta el uso de la voz en aplicaciones de seguridad consiste en que por ser la voz una característica biométrica, personal e intransferible es más fácil de enviar por ejemplo por medio telefónico comparado con sistemas como huellas dactilares, análisis de retina, etc.

1.2 Extracción de características

Se puede definir como un proceso que divide la señal de voz en tramas del orden de 10 a 30 ms de forma de la onda y las transforma en un vector N-dimensional de características,

obteniéndose una secuencia de vectores que se comparan con los diferentes modelos, para así determinar a qué locutor pertenecen [3].

El proceso de reducir la cantidad de datos se denomina *parametrización*, el cual disminuye la complejidad computacional del proceso de reconocimiento y transforma la señal de voz a un nuevo espacio de características, donde es más sencillo distinguir los hablantes. Los vectores de características producidos por cada individuo se consideran muestras con una función densidad de probabilidad. Las distribuciones de los diferentes locutores se solaparán en ciertas zonas del espacio N-dimensional, pero será posible distinguir los locutores entre sí, si las características poseen un alto grado de capacidad discriminativa.

Los coeficientes LPC y Cepstrum, cada uno con sus respectivas familias derivadas, son las características más usadas en el reconocimiento [3].

1.3 Coeficientes LPC

Se obtienen realizando un análisis lineal predictivo (de ahí su nombre en inglés, Linear Predictive Coding, LPC), describen la función de transferencia del tracto vocal y permiten construir una envolvente espectral que resalta los grupos de armónicos reforzados por la función de transferencia, la forma de cálculo se representa en la Fig.1. Los LPC tienen su aplicación en la medida espectral de tramos de la señal de voz de corta duración. La desventaja es que se ven afectados por las frecuencias características de los sistemas de grabación y de transmisión.

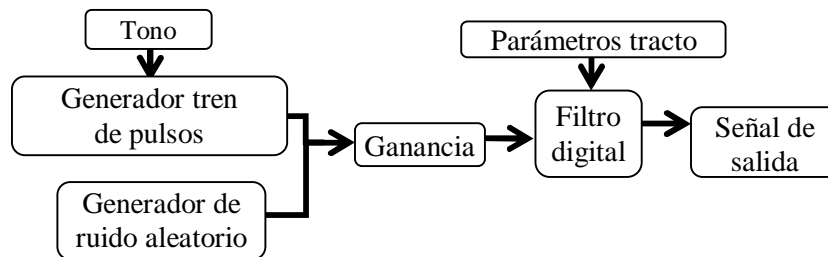


Fig. 1 Forma de cálculo de los LPC

1.4 Coeficientes cepstrales de frecuencia mel

“Mel-frequency cepstral coefficients” (MFCC) son los más conocidos. Se derivan de la TF (transformación de Fourier) o Transformada DCT (del inglés “Discrete Cosine Transform”) y son ampliamente usados en el reconocimiento del habla. El cálculo de los MFCC (Fig. 2) está basado en el análisis de espacio corto (análisis espectral). La diferencia básica entre TF o la DCT y los MFCC es que en los MFCC las bandas de frecuencia están situadas logarítmicamente, (según la escala Mel), que modela la respuesta auditiva humana más apropiadamente que las bandas espaciadas linealmente de TF o DCT.

Pueden obtenerse mediante la transformada inversa de Fourier del logaritmo del espectro, la ventaja es que se pueden obtener a través de los coeficientes LPC. La envolvente espectral reconstruida, basándose en los coeficientes cepstrales es más suave que la reconstruida basándose en los coeficientes LPC y proporciona una representación más estable entre las pronunciaciones repetidas de un mismo locutor [4-5]. Algunos resultados de la simulación de los MFCC de orden fraccional se enseñan en los anexos. (Anexo A1)

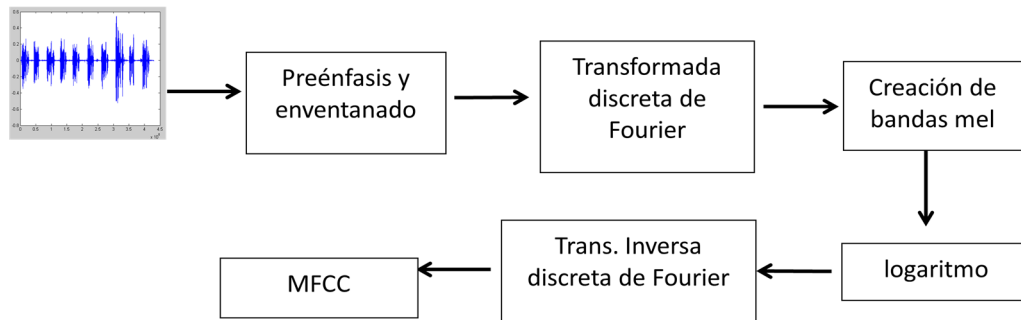


Fig. 2 Forma de cálculo de los coeficientes cepstrales

1.5 Métodos de reconocimiento de locutores

Cuando se aborda el estudio de los métodos de reconocimiento de locutores se encuentran en la literatura semejanzas y diferencias entre ellos, que los hacen específicos en

determinadas aplicaciones; se debe tener presente las características como la cantidad de datos a manipular, si el sistema es dependiente o independiente del locutor, y la forma como se realiza la toma de decisión.

1.5.1 Reconocimiento de locutores por cuantificación vectorial VQ

La cuantificación vectorial (del inglés Vectorial Quantification) es una generalización de la cuantificación escalar, es un proceso irreversible, que consiste en la manipulación conjunta de grupos de N muestras, siendo N la dimensión de los vectores como aparece en la Fig. 3. En éste tipo de cuantificación se usa un libro de códigos (“codebook”), que es aquella tabla que contiene los vectores posibles a la salida del cuantificador.

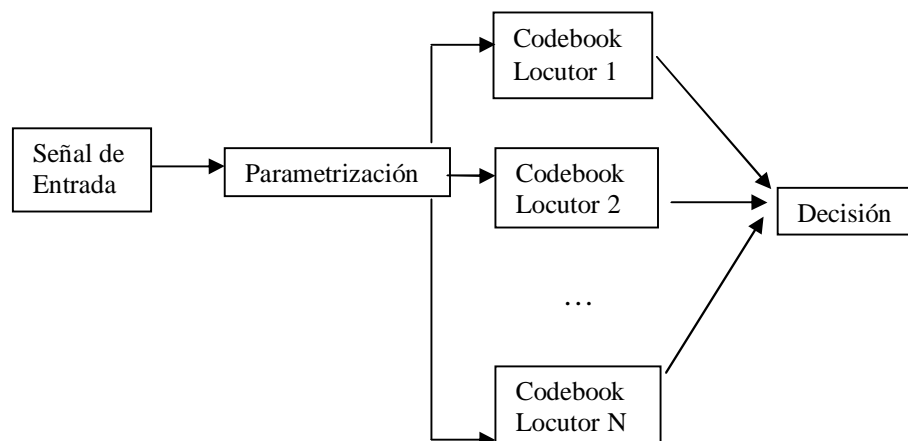


Fig. 3 Reconocimiento del locutor por cuantificación vectorial

Es decir, la cuantificación vectorial radica en dividir el espacio vectorial en sectores, cada uno de los cuales será representado por un sólo vector.

1.5.2 Reconocimiento de locutores por NN (del inglés, “nearest neighbour”), o vecino más cercano

No agrupa los vectores de la secuencia de entrenamiento para formar un codebook por locutor, como lo hace el método anterior. El reconocimiento por vecino más cercano

consiste en calcular la distancia entre las tramas de test de entrada y las de entrenamiento almacenadas previamente como se muestra en la Fig. 4; obteniendo así una matriz de distancias para cada uno de los locutores, en donde se promedia la mínima distancia entre tramas, y se compara con el locutor que proporcione la menor distancia.

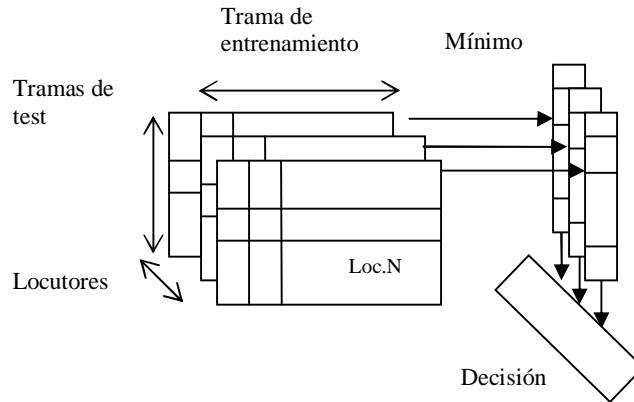


Fig. 4 Reconocimiento de Locutores por NN

1.5.3 Reconocimiento de locutor mediante Redes Neuronales

La aproximación al problema del reconocimiento de locutor, se lleva a cabo con una parametrización de la señal de entrada, que da una serie de vectores de entrenamiento para cada locutor. A partir de estos vectores se entrena una red neuronal, que actuará como clasificador; para la identificación del locutor; éste entrenamiento discriminativo pretende que la red sea capaz de discriminar los vectores de un locutor frente a los de los demás. El entrenamiento discriminativo tiene el inconveniente de que cuando se desee añadir un nuevo locutor a la base de datos, será necesario volver a entrenar toda la red neuronal.

1.5.4 Modelos ocultos de Markov (del inglés, “HMM, Hidden Markov Models”)

Son autómatas de estados finitos que pretenden modelar el mecanismo de producción del habla. Su capacidad de discriminación se debe a que están constituidos por un conjunto de estados, parecidos a los estados que atraviesa el tracto vocal cuando hablamos [3]. Dependiendo de cómo evolucionen los estados, los modelos HMM se clasifican en:

- HMM ergódicos

- HMM izquierda a derecha

En el reconocimiento de locutores (RL) no se usa HMM izquierda a derecha, puesto que la identidad de un locutor no constituye de forma directa una secuencia fonética ordenada en el tiempo. Los modelos ergódicos son los empleados en el RL, éstos no presentan un orden secuencial en las transiciones entre los distintos estados del modelo, es decir permiten cualquier transición entre estados. Una aplicación de HMMs a un sistema de verificación de locutores es la mostrada en la Fig. 5.

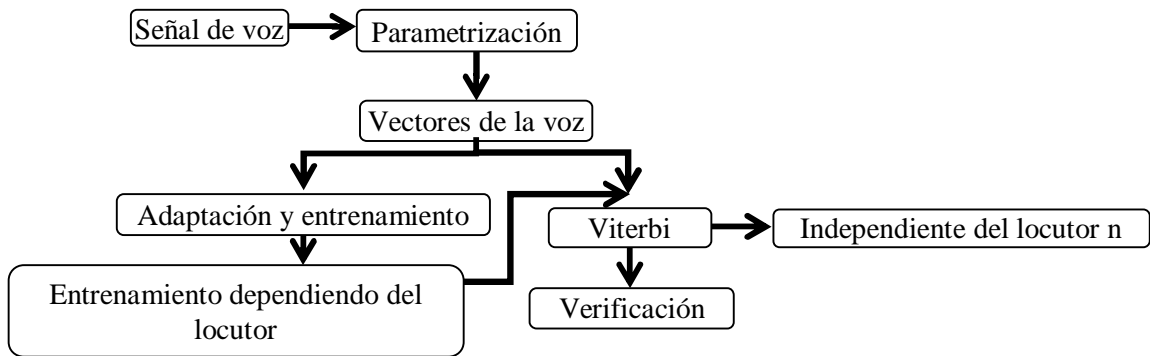


Fig. 5 Forma de cálculo de los HMM Aplicación de HMMs a un sistema de verificación de locutor

CAPITULO 2

TRANSFORMADA DE FOURIER FRACCIONAL (FRFT)

2.1 Generalidades

La transformada de Fourier fraccional pertenece al conjunto de las transformaciones canónicas lineales; introducida por Víctor Namias [42] y posteriormente en investigaciones de Ozaktas[2,6,14] Almeida [43] y Pellat-Finet [41] en el área de la óptica y el procesamiento de señales respectivamente.

La primera noción de la FrFT aparece en un trabajo de Wigner (1929) y consecutivamente autores como Condon (1937) y Patterson (1959) hacen alusión a esta transformada. En el año de 1980 el matemático Namias emplea la FrFT como un método novedoso para resolver la ecuación de Schrödinger [42].

La FrFT es conocida asimismo como una generalización de la transformada de Fourier estándar, con un parámetro de orden fraccional α . Este orden paramétrico puede ser un número real o imaginario [19]. La representación de la señal en los órdenes intermedios es un paso muy gradual del espacio directo al espacio recíproco que se llama espacio tiempo-frecuencia en este caso. En síntesis la transformada fraccional tiene un potencial para la generalización y el mejoramiento, en áreas en donde la transformada de Fourier y el concepto del dominio frecuencial¹ se emplea.

¹ Es un caso especial del dominio continuo de Fourier fraccional, el cual está íntimamente relacionado con representaciones tiempo-frecuencia (o espacio-frecuencia)

También, la FrFT ha encontrado un importante rol en el estudio de sistemas ópticos, los cuales permiten una generalización en cuanto a la noción del dominio frecuencial, incrementando el conocimiento en el espacio directo-frecuencia.

La transformada de Fourier Fraccional puede escribirse como:

- La integral lineal de la transformación.
- La potencia fraccional de la transformada de Fourier.
- La rotación de la representación de Wigner en el plano tiempo – frecuencia.
- La transformación de coordenadas, operadores multiplicación y diferenciación.
- Y la solución a una ecuación diferencial.

2.2 Definición

La forma integral de la transformada de Fourier fraccional en una dimensión es:

$$F^a[f(x)](x') = \int_{-\infty}^{\infty} K_a(x, x') f(x') dx' \quad (1)$$

con núcleo² $K_a(x, x') = K_\alpha e^{i\pi(\cot\alpha x^2 - 2\csc\alpha x x' + \cot\alpha x'^2)}$ (2)

donde $\alpha \equiv \frac{a\pi}{2}$ y $K_\alpha = \sqrt{1 - i \cot \alpha}$. (3)

Para $a = 0$ y $a = \pm 2$ el núcleo se define como

$$K_0(x, x') \equiv \delta(x - x') \quad \text{y} \quad K_{\pm 2}(x, x') \equiv \delta(x + x') \quad (4)$$

Para $a = 1$, se encuentra que $K_\alpha = 1$ y

$$F^1[f(x)](x') = \int_{-\infty}^{\infty} e^{-i2\pi x x'} f(x') dx' \quad (5)$$

² Algunos sistemas lineales especiales con sus núcleos de transformación son demostrados en los anexos. (Apéndice B1)

Ésta última expresión, corresponde a la transformada de Fourier estándar de $f(x)$; de forma semejante $F^{-1}[f(x)](x')$ es la transformada de Fourier estándar inversa de $f(x)$.

Puesto que $\alpha \equiv \frac{a\pi}{2}$ aparece en la ecuación (1), sólo en el argumento de las funciones trigonométricas, la definición es periódica en a (o en α) con periodo 4 (o 2π) respectivamente, de ahí surge el operador identidad si la FrFT tiene orden cero ($a=0$) u orden cuatro ($a=4$), y la transformada de Fourier estándar si el orden es uno ($a=1$) y la inversa si el orden es menos uno ($a=-1$).

2.2.1 La potencia fraccional de la transformada de Fourier.

Al querer abordar otra forma de enunciar la FrFT se define el a -ésimo orden de la transformada de Fourier fraccional como la a -ésima potencia fraccional de la transformada estándar de Fourier. Denotando la función de Gauss_Hermite como ψ_l que es la función propia de la transformada de Fourier estándar con su respectivo valor propio λ_l se define la FrFT como:

$$F^a \psi_l = \lambda_l^a \psi_l = \left(e^{-\frac{il\pi}{2}} \right)^a \psi_l = e^{-\frac{ial\pi}{2}} \psi_l \quad (6)$$

Para encontrar la transformada fraccional de una función $f(x)$, primero se expande ésta como una superposición lineal de las funciones propias de la FrFT:

$$f(x) = \sum_{l=0}^{\infty} C_l \psi_l(x) \quad \text{siendo} \quad C_l = \int \psi_l(x') f(x') dx' \quad (7)$$

$$\text{Luego} \quad F^a f(x) = \sum_{l=0}^{\infty} e^{-\frac{ial\pi}{2}} C_l \psi_l(x) = \int \sum_{l=0}^{\infty} e^{-\frac{ial\pi}{2}} \psi_l(x) \psi_l(x') f(x') dx' \quad (8)$$

Con núcleo de transformación

$$K_a(x, x') = \sum_{l=0}^{\infty} e^{-\frac{ial\pi}{2}} \psi_l(x) \psi_l(x') \quad (9)$$

2.2.2 La rotación de la representación de Wigner en el plano tiempo – frecuencia.

El a-ésimo orden de la FrFT es una transformación canónica lineal definida por la matriz de transformación M y corresponde a una rotación de la distribución de Wigner de la señal en el plano tiempo-frecuencia.

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \cos \alpha & \operatorname{sen} \alpha \\ -\operatorname{sen} \alpha & \cos \alpha \end{bmatrix} \quad \text{donde } \alpha \equiv \frac{a\pi}{2} \quad (10)$$

$$W_{f_a(x, x')} \equiv W_f(x \cos \alpha - x' \operatorname{sen} \alpha, x \operatorname{sen} \alpha + x' \cos \alpha) \quad (11)$$

2.2.3 La transformación de coordenadas, operadores multiplicación y diferenciación.

Especificando el operador de multiplicación U y el operador de diferenciación D en los dominios de tiempo (x) y frecuencia (x') se definen las siguientes propiedades

$$\begin{aligned} Uf(x) &= xf(x) \\ Df(x) &= \frac{1}{i2\pi} \frac{d}{dx} f(x) \\ -UF(x') &= \frac{1}{i2\pi} \frac{d}{dx'} F(x') \\ DF(x') &= x'F(x') \end{aligned} \quad (12)$$

Los operadores U_a y D_a se definen como una rotación de los anteriores:

$$\begin{bmatrix} U_a \\ D_a \end{bmatrix} \equiv \begin{bmatrix} \cos \alpha & \operatorname{sen} \alpha \\ -\operatorname{sen} \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} U \\ D \end{bmatrix} \quad (13)$$

Tal que $U_0 = U$ $U_1 = D$ $D_0 = D$ $D_1 = -U$

Luego la representación en el dominio de la FrFT $f_a(x_a)$ de una función f que satisfaga las propiedades se escribe como cualquiera de estas dos posibilidades:

$$\begin{aligned} U_a f_a(x_a) &= x_a f_a(x_a) \\ D_a f_a(x_a) &= \frac{1}{i2\pi} \frac{d}{dx_a} f_a(x_a) \end{aligned} \quad (14)$$

Que no es más que una generalización de la Ecuación 12. Siendo x_a la variable asociada al dominio de la a-ésima FrFT.

2.2.4 Una ecuación diferencial.

Esta definición se introduce de tal manera que la FrFT $f_a(x)$ de una función $f_0(x)$ corresponda a la solución de una ecuación diferencial con condiciones iniciales $f_0(x)$.

Esta ecuación es conocida como la ecuación diferencial del oscilador armónico en mecánica cuántica [2].

$$\left[-\frac{1}{4\pi} \frac{\partial^2}{\partial x^2} + \pi x^2 - \frac{1}{2} \right] f_a(x) = i \frac{2}{\pi} \frac{\partial f_a(x)}{\partial a} \quad (15)$$

La solución $f_a(x)$ de esta ecuación es la FrFT de orden a-ésimo de la función $f_0(x)$.

2.3 Propiedades

Las propiedades básicas de la transformada de Fourier estándar son análogas a la de la transformada de Fourier fraccional, citando algunas de ellas se tiene que:

Linealidad: Sí se consideran dos funciones $f(x)$ y $g(x)$ la transformada de Fourier fraccional cumple que

$$F^a[cf(x)+bg(x)](x') = cF^a[f(x)](x') + bF^a[g(x)](x') \quad (16)$$

Siendo b y c constantes complejas.

Desplazamiento: También conocido como el teorema de corrimiento

$$F^a[f(x-x_0)](x') = F^a[f(x'-x_0 \cos \alpha)]e^{i\pi \sin \alpha (x_0^2 \cos \alpha - 2x'x_0)} \quad (17)$$

Adición del índice

$$F^{a1}[f(x)](x')F^{a2}[f(x)](x') = F^{a1+a2}[f(x)](x') \quad (18)$$

Teorema de Parseval

$$\int f(x)g^*(x)dx = \int F^a[f(x)](x')G^{*a}[g(x)](x')dx' \quad (19)$$

Presentando un ejemplo de simulación de una de las propiedades si se tienen dos funciones rectángulo, definidas para un mismo número de muestras con diferente ancho de pulso como se observa en la fig. 6, se calcula el módulo y la fase de la señal donde se evidencia los cambios de un orden a otro, en el orden 0 del módulo de la señal se tiene la función rectángulo, y en el orden 1 una función sinc. Los órdenes varían de cero a uno en pasos de 0,01.

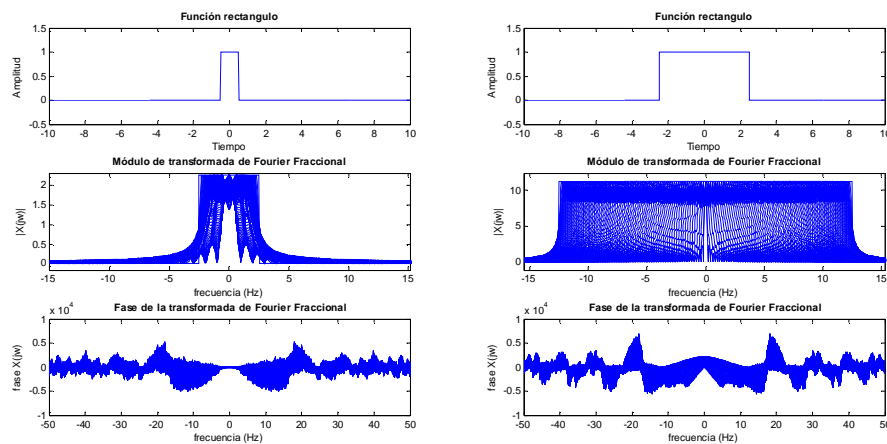


Fig. 6 Funciones rectángulo, Módulo y Fase

La FRFT normalizada al valor máximo de la función rectángulo para un centenar de órdenes, permite visualizar una menor energía para el rectángulo más estrecho como era de esperarse. (Ver fig. 7)

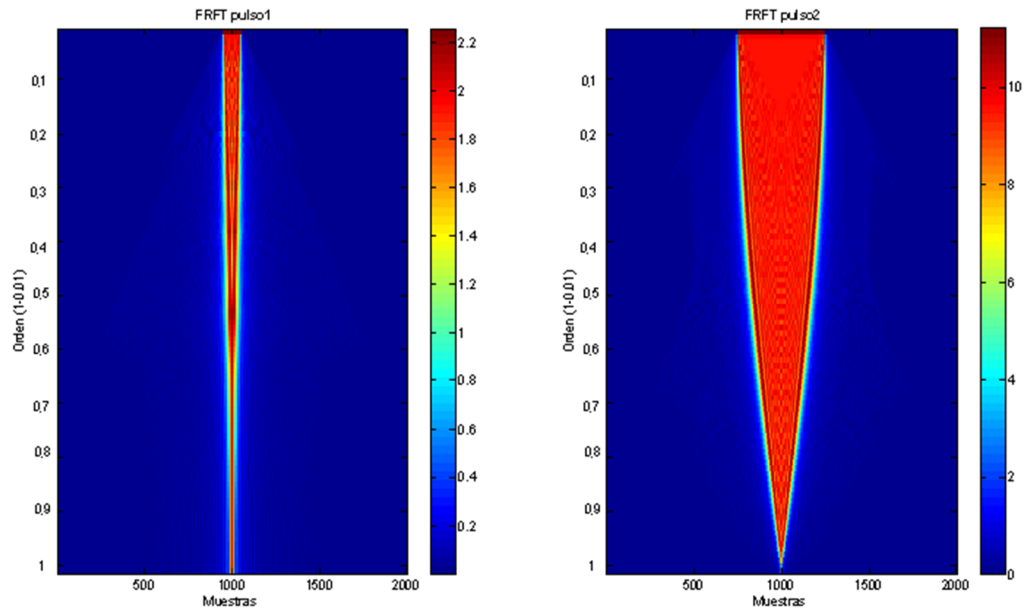


Fig. 7 Módulo de la FRFT normalizada al valor máximo de la función rectángulo de las funciones senoidales de la figura precedente.

De forma similar se realiza la simulación de una función sinusoidal con diferente frecuencia (Fig. 8, 9 y 10).

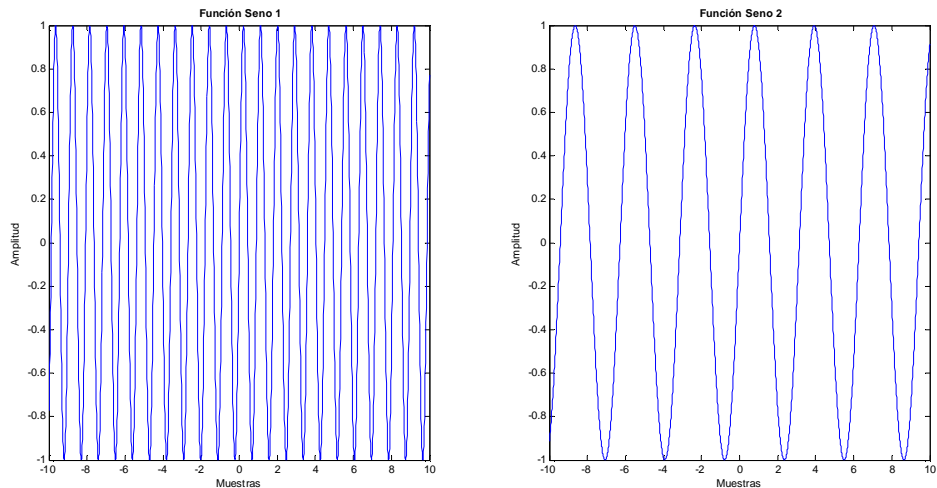


Fig. 8 Señal seno con variación en las frecuencias izquierda 1,1 Hz. y derecha 0,31 Hz.

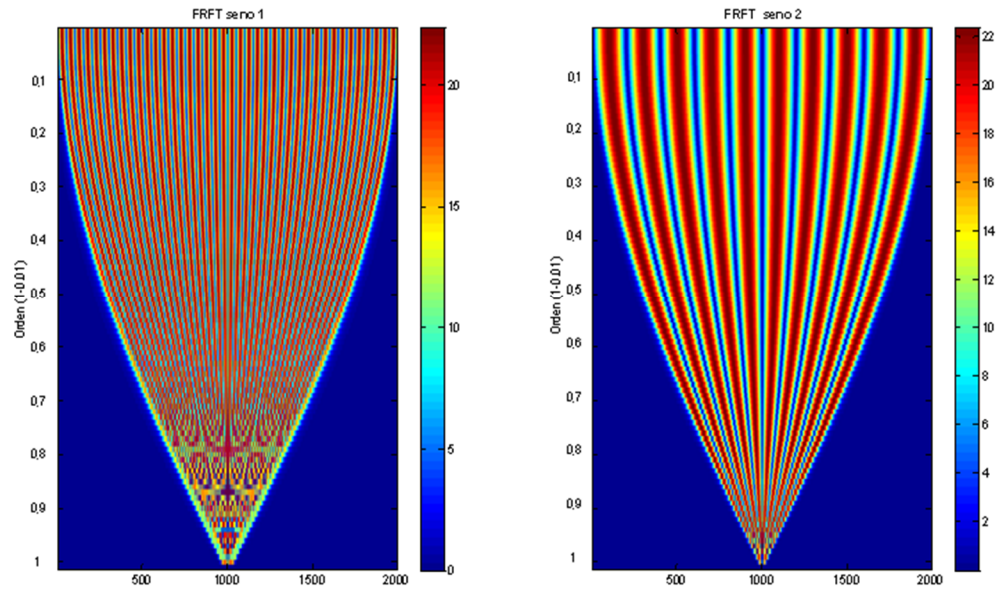


Fig. 9 Módulo de la FRFT normalizando al valor máximo de las señales sinusoidales de la figura precedente.

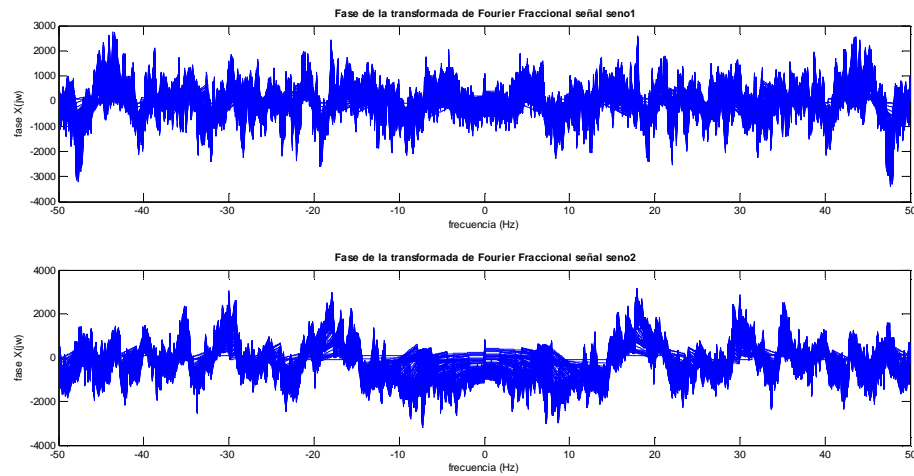


Fig. 10 Fases de la transformada de Fourier Fraccional de 100 órdenes de forma simultánea en funciones seno de la figura anterior.

Otro ejemplo es la aplicando la propiedad de escalamiento en este caso a una función rectángulo, ver Fig. 11

$$rect(u) \text{ y } \frac{1}{M} rect\left(\frac{U}{M}\right)$$

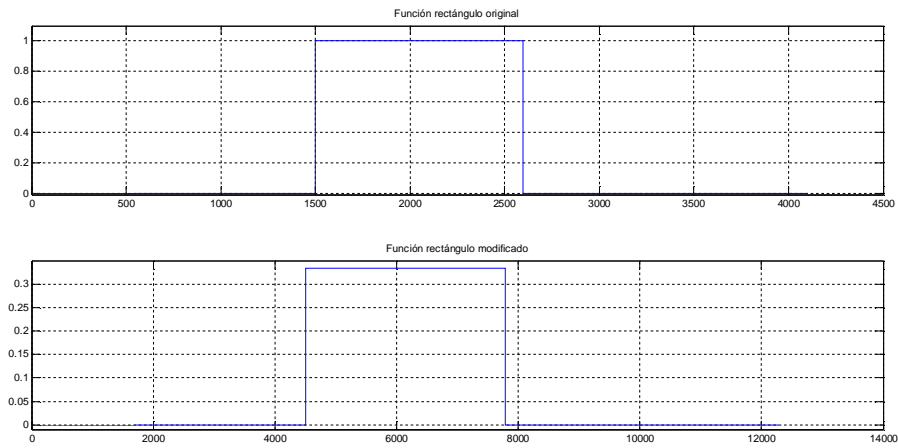


Fig. 11 Señal rectángulo, abajo escalada la de arriba por un factor $M=3$.

El módulo y la fase de la FrFT de la función rectángulo para un orden fraccional específico (0,75) son mostrados en las figuras 12 y 13.

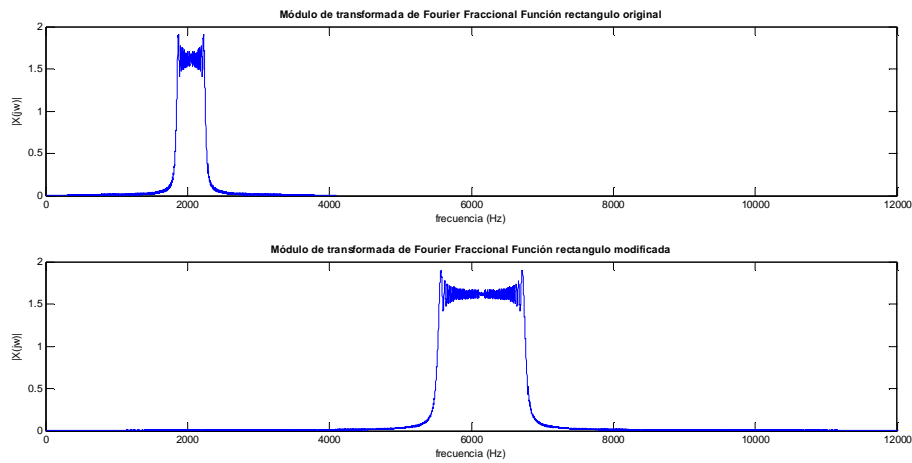


Fig. 12 Módulos de la FRFT de la función rectángulo con orden 0,75. La de abajo es la de arriba escalada por un factor $M=3$.

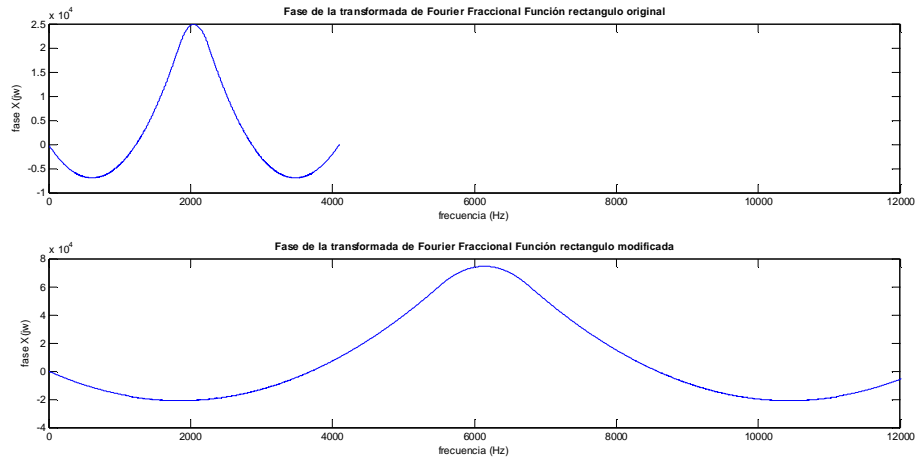


Fig. 13 Fases de la FRFT de la función rectángulo con orden 0.75, la de abajo es la de arriba escalada por un factor M=3.

Otra simulación realizada de la propiedad de escalamiento consistió en definir una función rectángulo original como

$$rect(u) = \begin{cases} 1 & 1500 < u < 2596 \\ 0 & \text{el resto} \end{cases} \quad \begin{array}{l} \text{En la simulación el cero para el resto fue} \\ \text{tomado desde } 1-1500 \text{ y } 2596-4096 \end{array}$$

- Para M=3 $rect\left(\frac{u}{3}\right) = \begin{cases} 1 & 4500 < u < 7788 \\ 0 & \text{el resto} \end{cases}$

- Para M=1/2 $rect(2u) = \begin{cases} 1 & 750 < u < 1298 \\ 0 & \text{el resto} \end{cases}$

- Para M=-3 $rect\left(\frac{u}{-3}\right) = \begin{cases} 1 & -4500 < u < -7788 \\ 0 & \text{el resto} \end{cases}$

En la Fig. 14 se encuentra la función rectángulo original con sus respectivos escalamientos para M=3, 1/2 y -3.

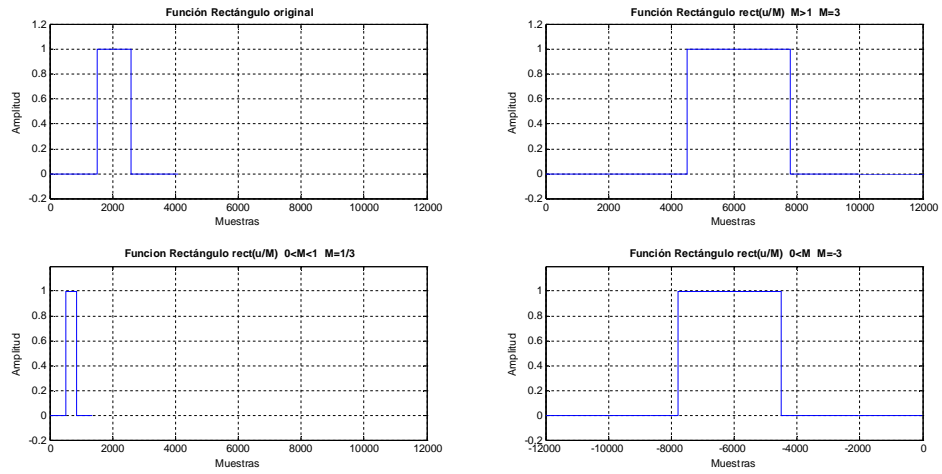


Fig. 14 Función Rectángulo con diferentes valores de M, definidos en las ecuaciones precedentes en el texto.

Al calcular el módulo de la FrFT y usando la función `imagesc` de matlab para su visualización los resultados se muestran en la Fig. 15.

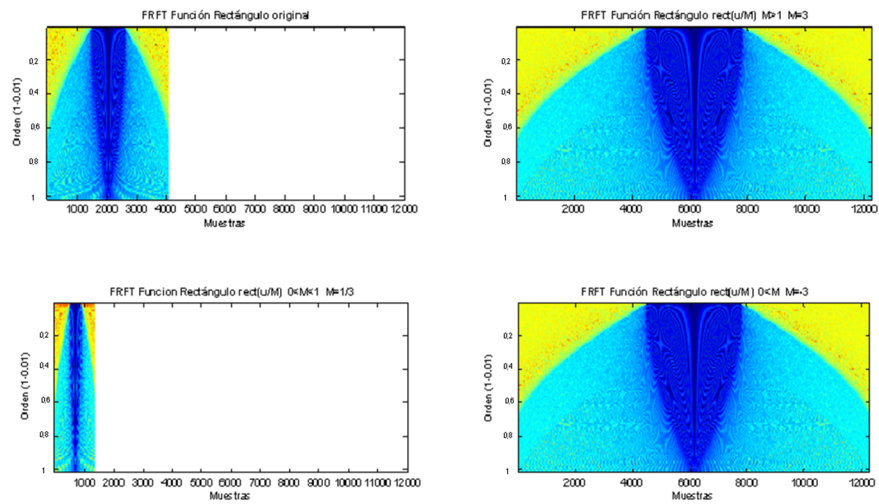


Fig. 15 FrFT de las funciones rectángulo escaladas en el eje del tiempo. Arriba señal original y a la derecha el caso $M=3$, abajo izquierda $M=1/3$ y a la derecha $M= -3$

De igual manera en la figura 16 aparece una señal sinusoidal definida como:

$$\text{sen}(t) \quad -50 < t < 50 \quad \text{En pasos de } 0,1$$

Y sus respectivos escalamientos:

- Para $M=3$

$$\text{sen}\left(\frac{t}{m}\right) = \text{sen}\left(\frac{t}{3}\right)$$

- Para $M= 1/2$

$$\text{sen}\left(\frac{t}{m}\right) = \text{sen}(2t)$$

- Para $M= -3$

$$\text{sen}\left(\frac{t}{m}\right) = \text{sen}\left(\frac{t}{-3}\right)$$

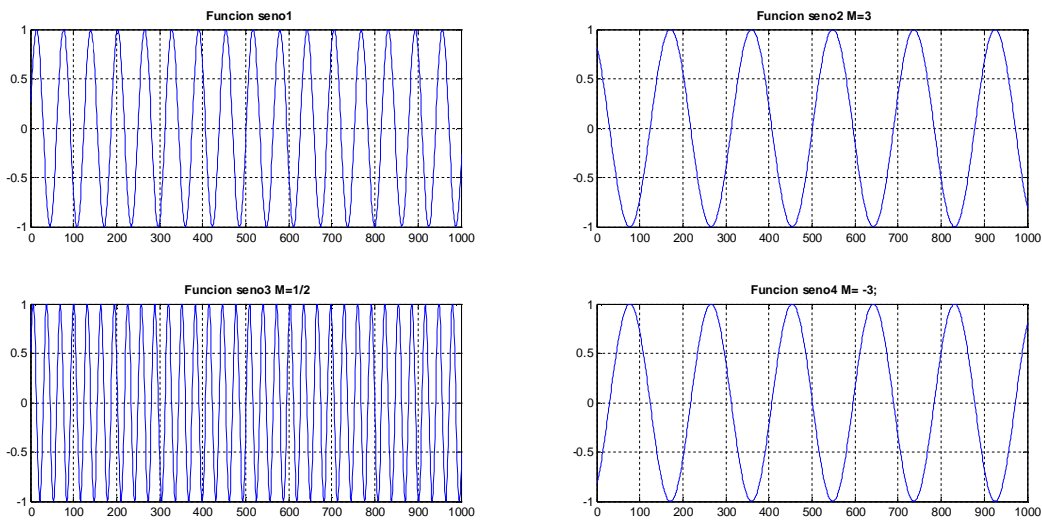


Fig. 16 Funciones seno con diferentes valores de M, para distintos escalamientos en el tiempo

En la figura 17 se muestra el módulo de la FrFT de cada una de las señales sinusoidales precedidas.

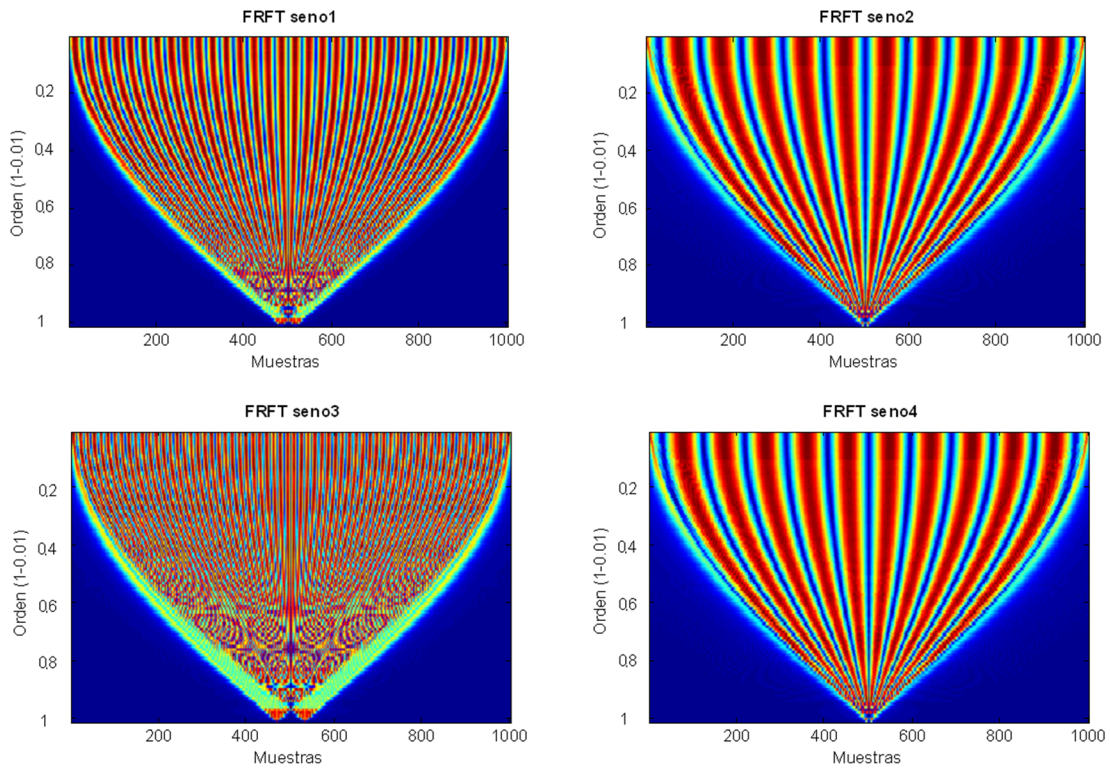


Fig. 17 Módulos de la FrFT de las funciones seno, arriba izq. frecuencia: 0,017 Hz. Der. frecuencia 5,52 mHz. Abajo izq. frecuencia: 0,027 Hz. Der. frecuencia 5,58 mHz

Con estas simulaciones se indaga la forma de respuesta de la transformada fraccional y se observa sus similitudes y diferencias con la transformada estándar.

2.4 Representación tiempo-orden para la FrFT

Existen diferentes formas de representar la información en una señal, como principal alternativa se propone la representación tiempo–orden (o espacio–orden) por considerarla apropiada para el caso de señales de voz y suficientemente redundante en contenido de información. Con el fin de evaluar tal representación se plantea la metodología descrita a continuación: Se inicia con una captura de las señales de voz; una vez pre-procesada la señal se realiza la transformada de Fourier de orden fraccional. El orden de la FrFT es variado desde el orden cero al orden uno en pasos de 0,01. El número de pasos fue un criterio seleccionado de forma arbitraria, y se decidió que fuera pequeño ya que se apreciaría un número considerable de transformadas de orden fraccional; al implementar

pasos de 0,01 se obtiene la representación de la señal de voz en un centenar de órdenes (exactamente 101 órdenes). Existen dos posibilidades para la representación Tiempo-Orden. Puesto que se trata de una señal unidimensional, puede ser en un espacio de coordenadas polares o bien de coordenadas rectangulares [2]. En términos generales en la investigación se planteó lo mostrado en la Fig. 18 donde antes de ser aplicada la FRFT se realizó un previo tratamiento de la señal de voz.

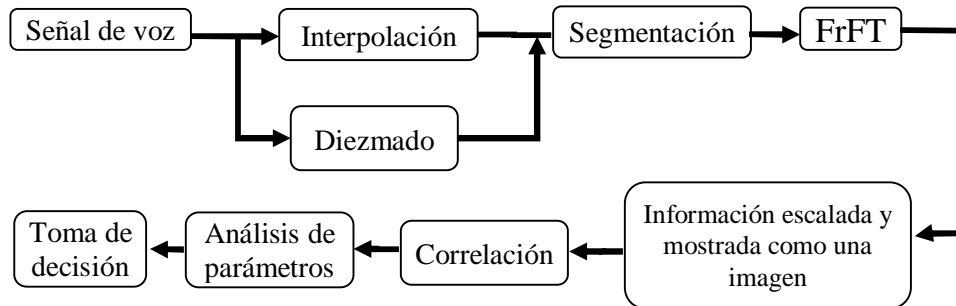


Fig. 18 Etapas del sistema general utilizado

2.5 Algoritmo de cálculo

Para el cálculo de la FrFT del centenar de órdenes de forma simultánea en las señales de voz se implementó el algoritmo mostrado en el diagrama de la fig. 19, donde el valor de la normalización depende de la energía en la señal.

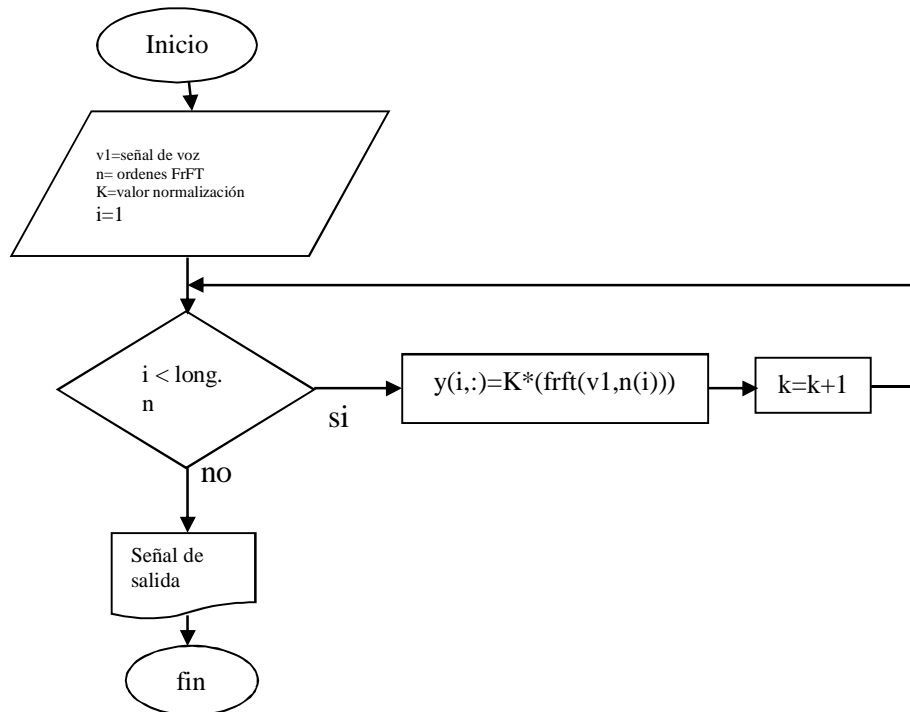
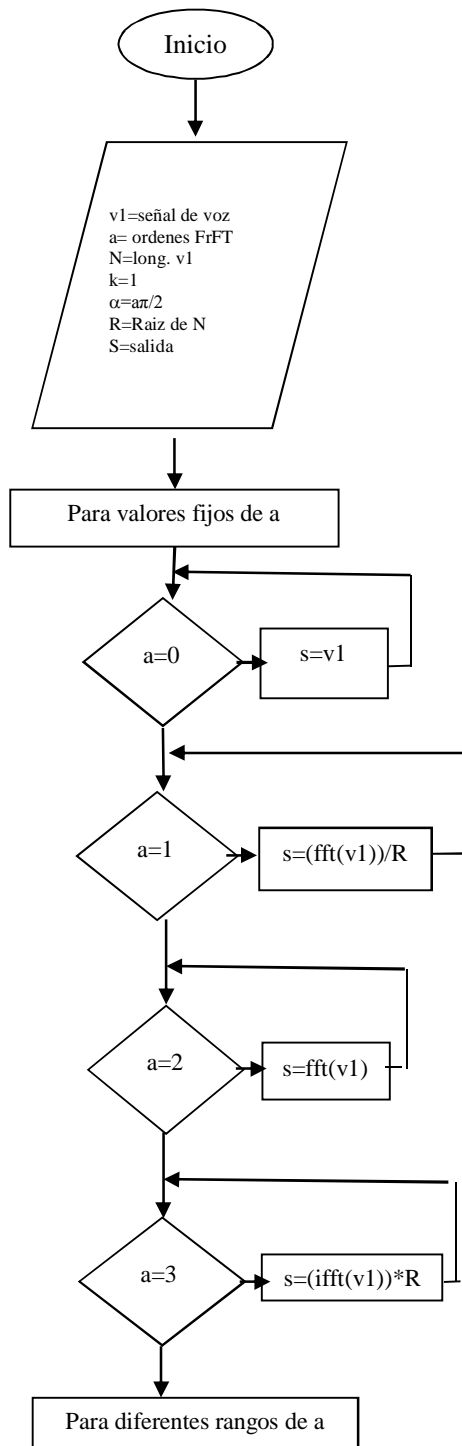


Fig. 19 Cálculo de la FrFT del centenar de órdenes



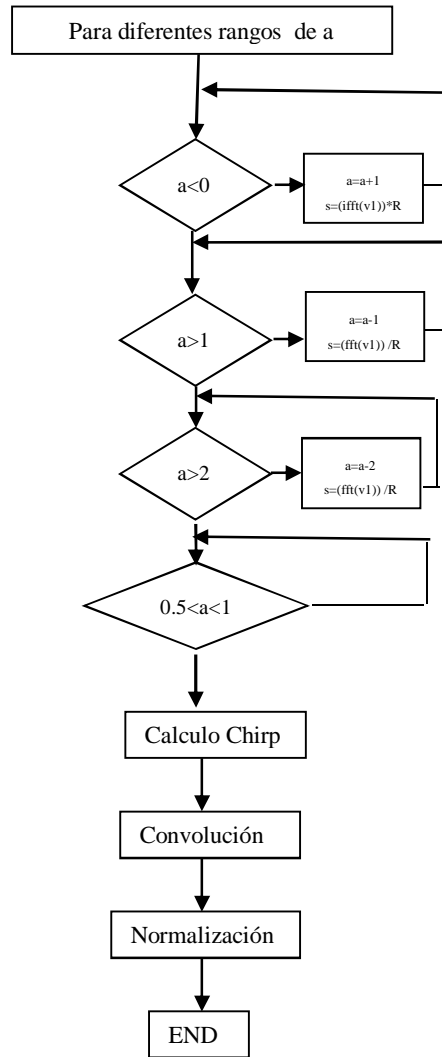


Fig. 20 Seudocódigo implementado para la FrFT

Una vez leídas las señales de voz, se introducen los parámetros de entrada como el centenario de órdenes y el valor de normalización con relación a la energía; posteriormente de forma secuencial se realiza el cálculo de la FRFT. En la Fig. 20 se presente el pseudocódigo del algoritmo implementado.

CAPITULO 3

PRUEBAS EXPERIMENTALES

Como una primera etapa posterior a la adquisición, se realiza un pre-procesamiento de la señal de voz, utilizando palabras extraídas de la base de datos modelo de locutores EUSTACE³, que comprende 4608 frases registradas por seis locutores, tres hombres y tres mujeres, con una frecuencia de muestreo de 16 kHz. Una representación de una de las frases pertenecientes a la base de datos se muestra en la Fig. 21.

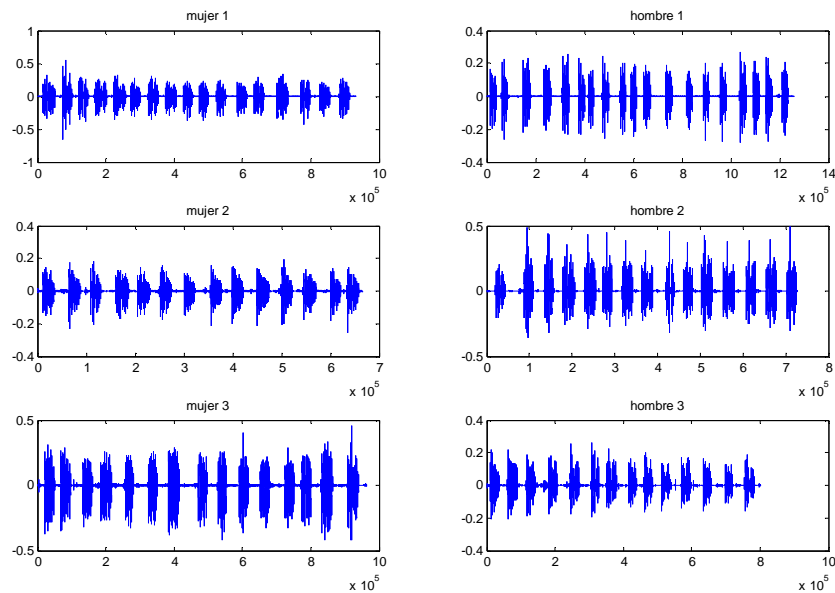


Fig. 21 Señal de voz de una de las frases para seis locutores

Tomando las frases se seleccionan las palabras que hacen parte del sistema a implementar. Para determinar la muestra de palabras a usar, se calcula de acuerdo al muestreo simple aleatorio por proporción definido como:

³ Centro de investigación tecnológica de la voz, Universidad de Edimburgo, Inglaterra, De acceso libre, de uso no comercial. <http://www.cstr.ed.ac.uk/>

$$n = \frac{N\hat{p}\hat{q}}{(N-1)\frac{B^2}{Z^2} + \hat{p}\hat{q}} \quad (20)$$

Donde $\hat{p} = 0,5$ (proporción a favor) y $\hat{q} = 0,5$ (proporción en contra) con un error de 0,1 y un nivel de confiabilidad de 95% de una población general de 3360 palabras pronunciadas por los 6 locutores, se obtiene una muestra de 93,40 palabras. Como son 6 locutores, se aproxima a 96 palabras [18].

Las palabras seleccionadas de forma arbitraria pronunciadas por los 6 locutores son:

1. Jhon
2. designed
3. the cap
4. Kate
5. explained
6. censorship
7. hope
8. find
9. partnership
10. now
11. continue
12. treatment
13. recommend
14. again
15. please
16. presuppose

Debido a que las señales de voz son de diferentes longitudes, se requiere para efectos de comparación que todas tengan la misma cantidad de muestras, como se muestra en la Fig. 22.

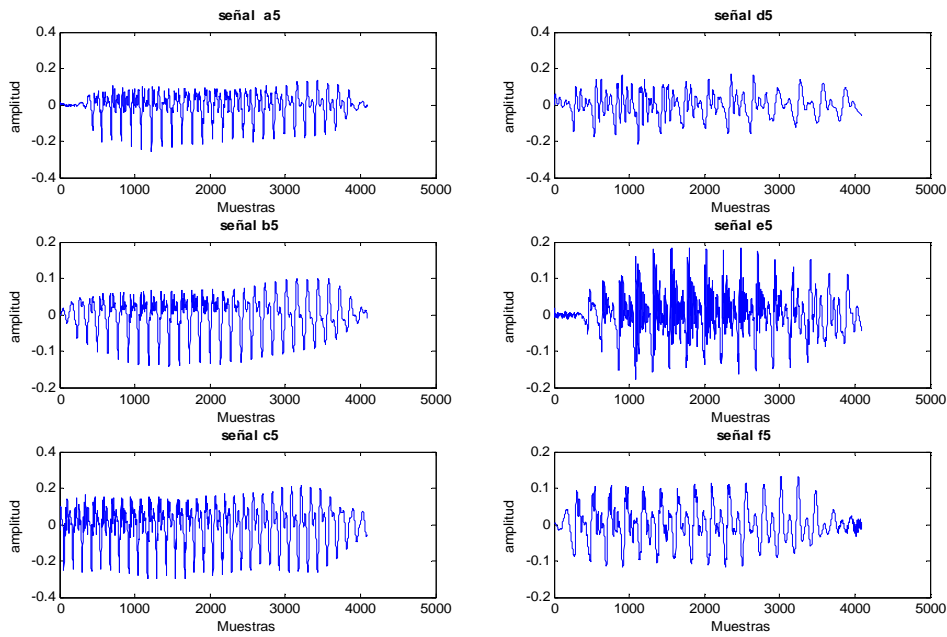


Fig. 22 Señal de voz, una fonación particular de los seis locutores, las señales de la derecha pertenecen a los hombres y las de la izquierda a las mujeres.

Para determinar el número de muestras a emplear a la hora de segmentar la señal de voz se realiza un histograma, éste es utilizado para encontrar el rango más frecuente de las tramas de voz que se han obtenido de las palabras de la base de datos; en ocasiones una misma palabra tiene diferentes muestras dependiendo del locutor. Como se evidencia en el histograma de la Fig. 23, el rango más predominante se encuentra entre el intervalo 3651 y 5851 muestras.

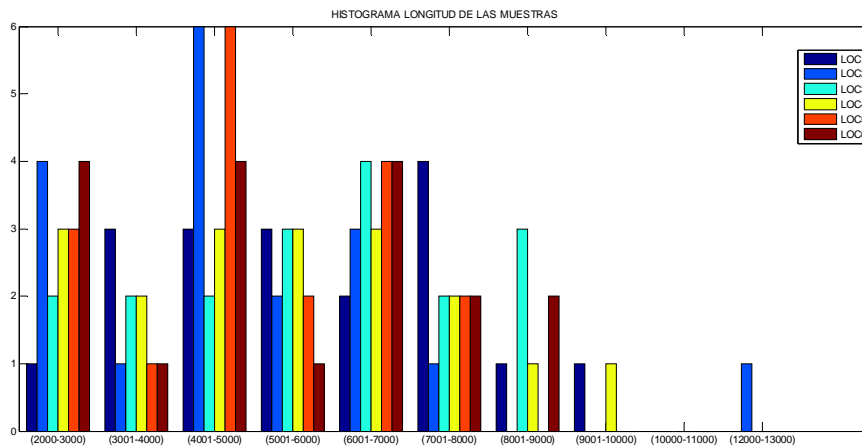


Fig. 23 Histograma de la longitud de las señales de voz seleccionadas de la base EUSTACE.

Otro factor que resulta ventajoso es tomar el número de muestras como un número potencia de dos, el más cercano es 4096 (2^{12}), con ello se consigue una mayor velocidad a la hora de implementar los algoritmos propios de la transformada de Fourier fraccional. Este valor no crea dependencia en el sistema, así como se tomó este valor se puede escoger otro valor arbitrario dentro del rango predominante del histograma.

Después de tener en cuenta estos parámetros se diezma o se interpola la señal como se muestra en la Fig. 24, el encargado de determinar si se interpola o se diezma la señal de voz es el número de muestras por encima o por debajo de 4096, respectivamente.

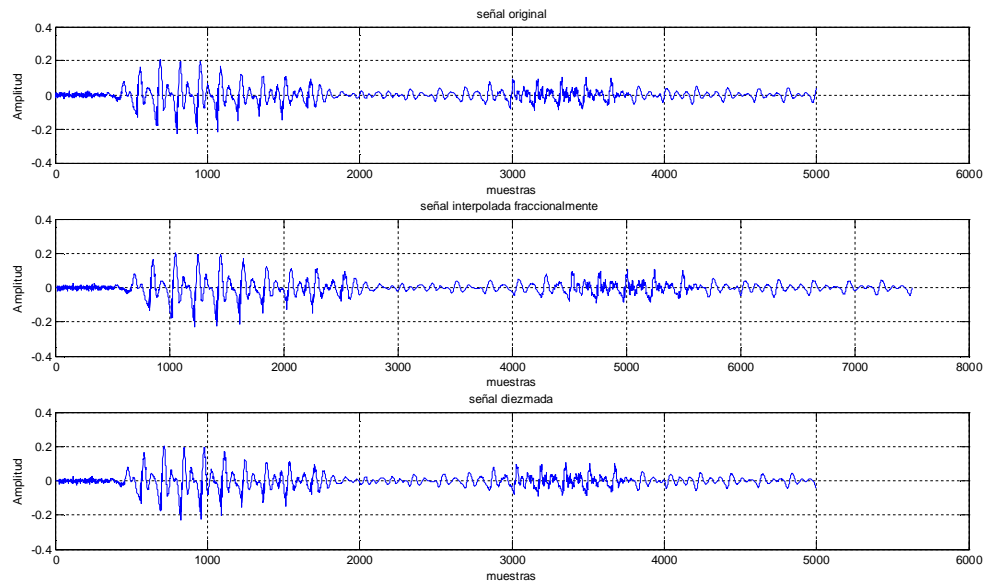


Fig. 24 Ejemplo de señal de voz interpolada y diezmada a un valor de 3/2

La señal de voz original y la señal interpolada o diezmada según sea el caso, son analizadas calculando el error cuadrático medio introducido por el algoritmo de cálculo implementado, garantizando que no exista pérdida de información en la señal truncada o interpolada. Para este caso concreto el error cuadrático medio fue $1.4879e-006$. La Fig. 25 ilustra las ligeras modificaciones del diezmado con relación a la señal original.

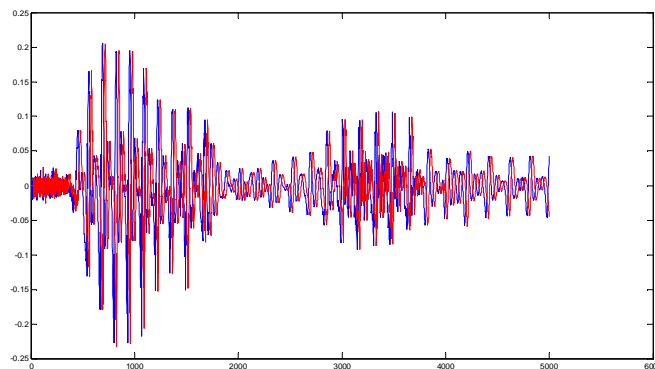


Fig. 25 Ejemplo del error en una señal, en rojo señal diezmada a partir de la señal interpolada y en azul señal original

3.1 Análisis bajo la transformada de Fourier fraccional

La transformada de Fourier fraccional se determina para un centenar de órdenes en las señales de voz mediante un algoritmo digital (Ver Fig. 26). Al realizar las pruebas iniciales se observó que no existe un criterio para elegir el orden⁴ al calcular la FrFT en las simulaciones, ya que cada orden representa una transformación particular en la señal y son diversas las características que tiene la señal al ser transformada, por ello se consideró un número de órdenes calculados de forma simultánea aumentando significativamente la redundancia.

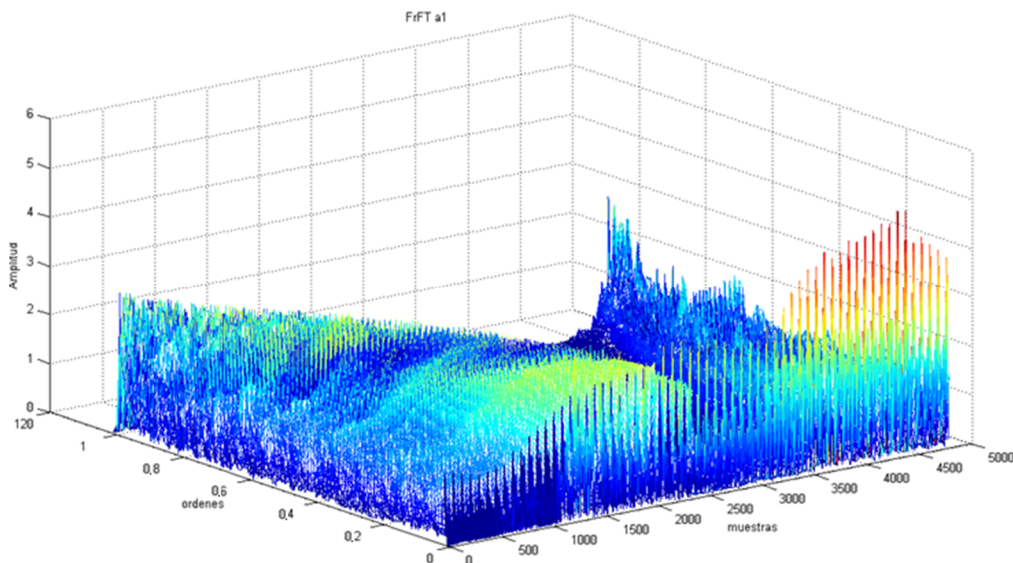


Fig. 26 Cálculo de la FrFT para 101 órdenes de una señal tomada de la base de datos seleccionada.

Finalmente, se realiza la representación Tiempo-Orden tal y como se muestra en la Fig. 27, en donde se observan las diferencias entre hombres y mujeres, los valores más altos del espectro están en las mujeres (izquierda), y en los hombres se ve una distribución casi uniforme con valores típicos por debajo de 1,2. Para este caso particular.

⁴ Se hace la salvedad de que existen órdenes que brindan una mejor representación de la señal para su aplicación al reconocimiento de locutores.

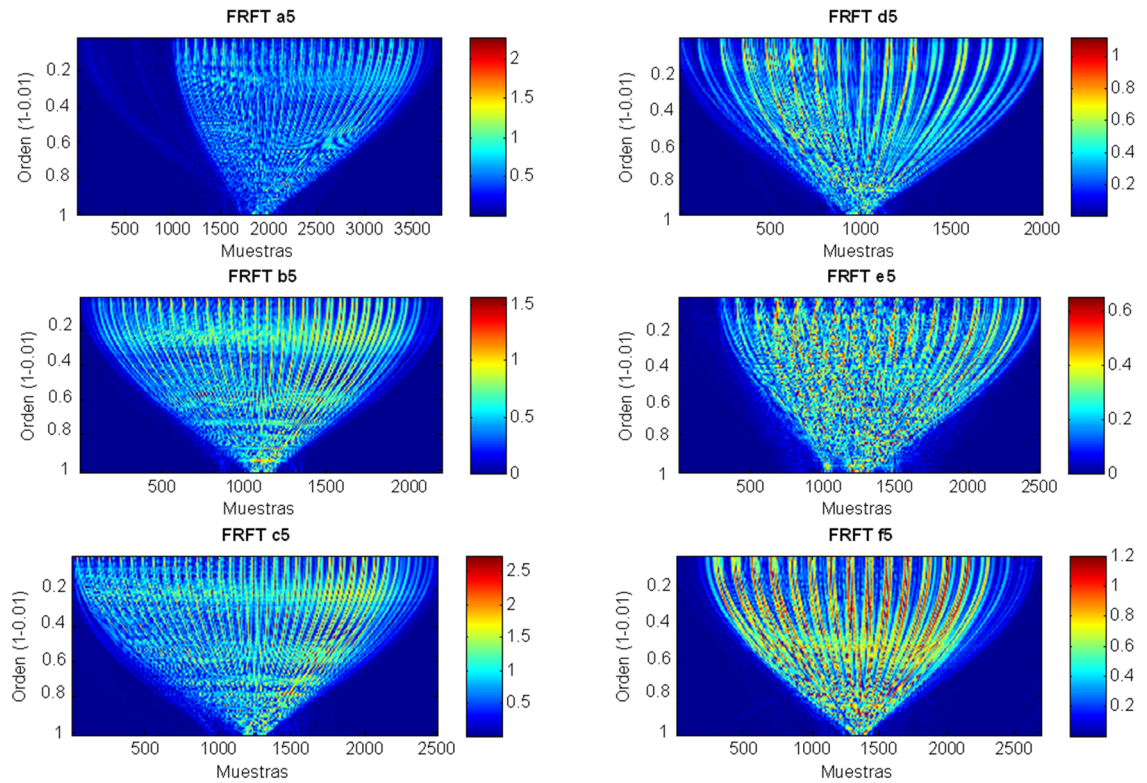


Fig. 27 Representación Tiempo–Orden del módulo de la FrFT de las señales de voz mostradas en la Fig. 22

En la misma figura, la representación de la señal de voz en el espacio directo (tiempo) se encuentra en la primera línea superior de cada diagrama tiempo-orden; y en la primera línea inferior se encuentra la señal de voz en el espacio recíproco (frecuencia pura o espectral). Esta representación tiempo-orden se replica para cada una de las palabras analizadas. Con el fin de verificar la calidad y poder de discriminación de la representación propuesta se procedió a realizar autocorrelaciones y correlaciones cruzadas entre las diferentes señales en Tiempo-Orden, como se muestra en las figuras 28 y 29.

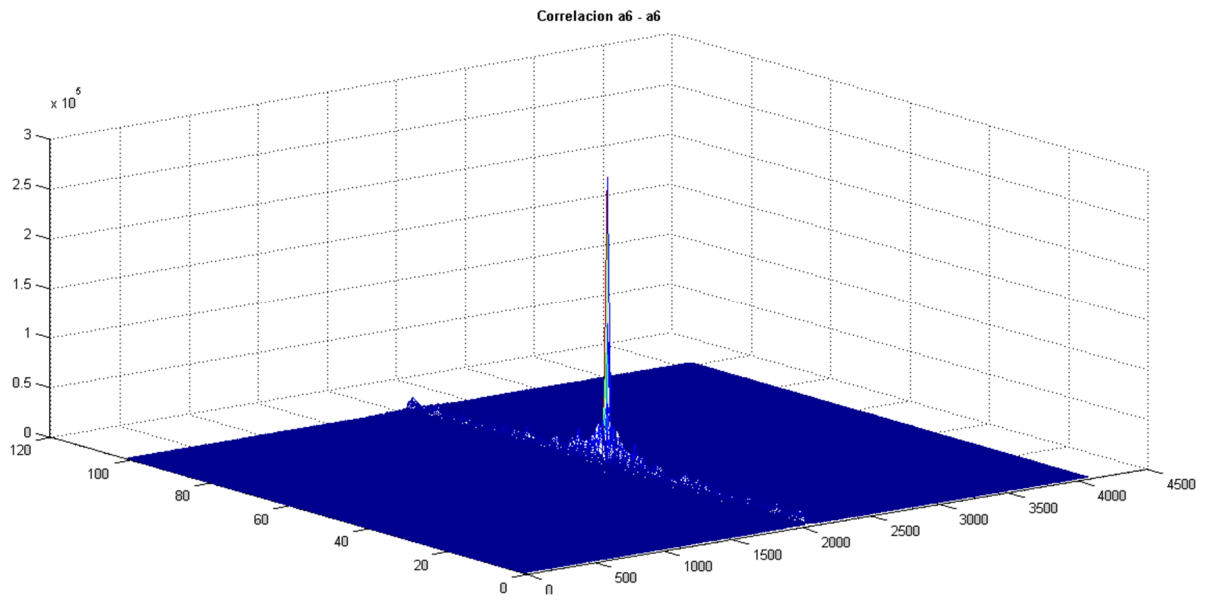


Fig. 28 Módulo de la función de autocorrelación de una señal de voz femenina

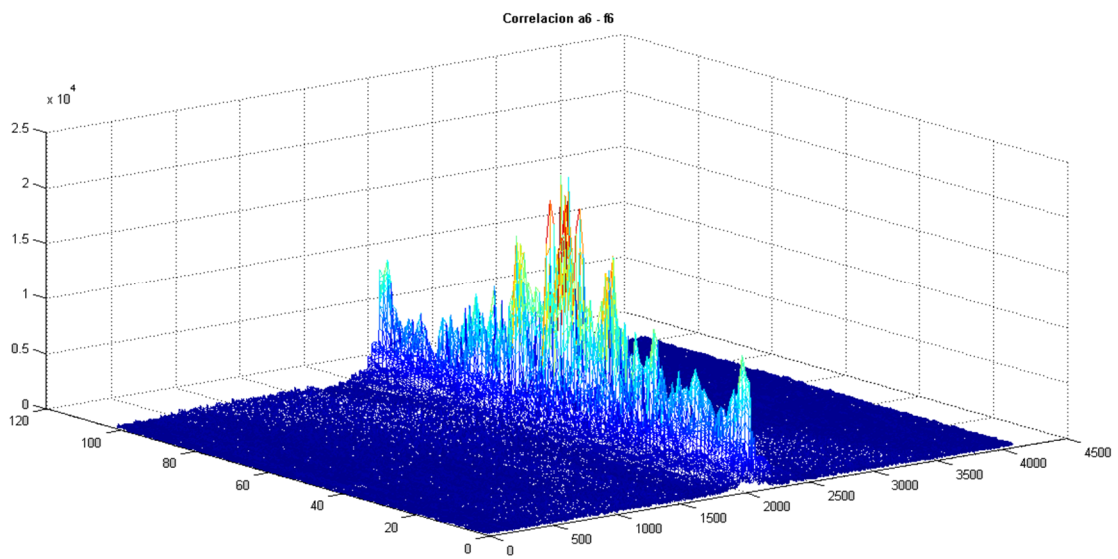


Fig. 29 Módulo de la correlación cruzada de una señal femenina y otra masculina

De acuerdo a los resultados obtenidos en la correlación, el pico máximo es presentado en la autocorrelación como era de esperarse, y en la correlación cruzada se presenta una señal con claras diferencias entre ellas, obsérvese que las escalas verticales difieren en más de un órdenes de magnitud. En el apéndice B2 se encuentra un ejemplo de correlación.

3.2 Correlación y autocorrelación

Para cuantificar el grado de similitud entre dos señales una vez calculadas la FrFT se procede a emplear la correlación, y obtener una medida objetiva que permita una clasificación en los locutores. Al realizar el cálculo de la transformada de Fourier fraccional (FrFT) se obtiene una señal con parte real e imaginaria, tomando la secuencia $x(n)$ como la transformada calculada de la fonación de una mujer, y $y(n)$ como la transformada de otra persona, sea hombre o mujer se define la correlación cruzada como

$$C_{xy}(k) = \sum_{-\infty}^{\infty} x(n)y^*(n-k) \quad (21)$$

Siendo k el parámetro de corrimiento definido en todo el plano real, la secuencia no ira desde $-\infty$ a $+\infty$ ya que depende del número de muestras que tiene la señal a correlacionar. La secuencia $y(n-k)$ con su respectivo complejo conjugado (denotada como $y^*(n-k)$). Los subíndices xy indican cuales son las señales correlacionadas. Si en lugar de calcular la C_{xy} se calcula C_{yx} se obtendrá la señal reflejada con respecto a $k=0$, sí las señales no son reales la operación en general no es conmutativa. Si $y(n)=x(n)$ se tiene la autocorrelación definida como

$$C_{xx}(k) = \sum_{-\infty}^{\infty} x(n)x^*(n-k) \quad (22)$$

Las señales correlacionadas son las transformadas de Fourier de orden fraccional tomadas de las señales de voz. Esta señal está representada de forma bidimensional

(tiempo, orden). La FrFT se ordena en una matriz de 101x4096, una vez correlacionadas los resultados son como los mostrados en la Fig. 30.

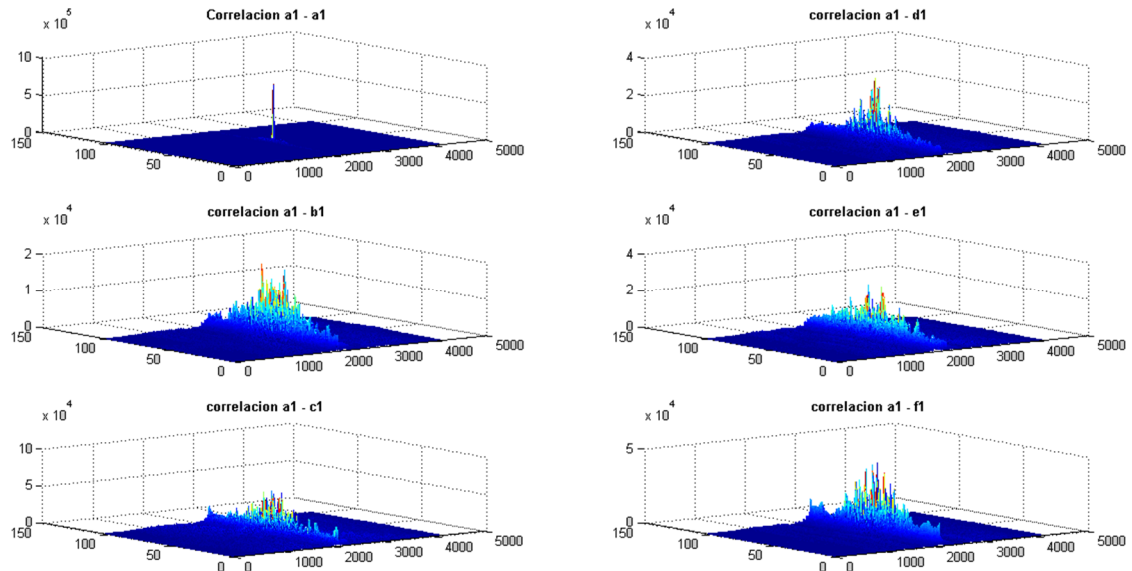


Fig. 30 Autocorrelación arriba izquierda y correlaciones cruzadas con las señales de los otros cinco locutores

3.3 Correlación de un mismo locutor

Una de las pruebas llevadas a cabo en la etapa de correlación fue tomar una palabra pronunciada por la misma persona en tiempos diferentes y observar cual era el comportamiento que se suscitaba. En la Fig. 31 se muestran estas señales obtenidas y posteriormente, en las figuras 32 y 33 la correlación entre ellas.

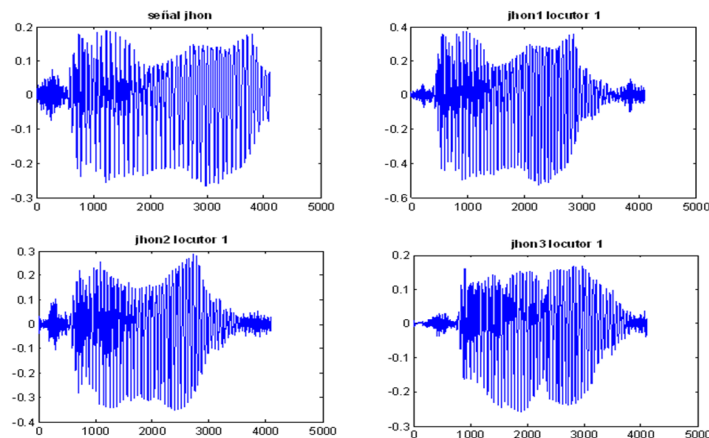


Fig. 31 Fonación de una misma palabra pronunciada por una misma persona en diferentes momentos

El algoritmo responde de forma muy diferente que cuando se emplea la misma señal con diferentes personas. En la Fig. 33 se detecta como la señal converge a un punto central y como en el plano generado se desplaza en ciertas regiones de amplitud.

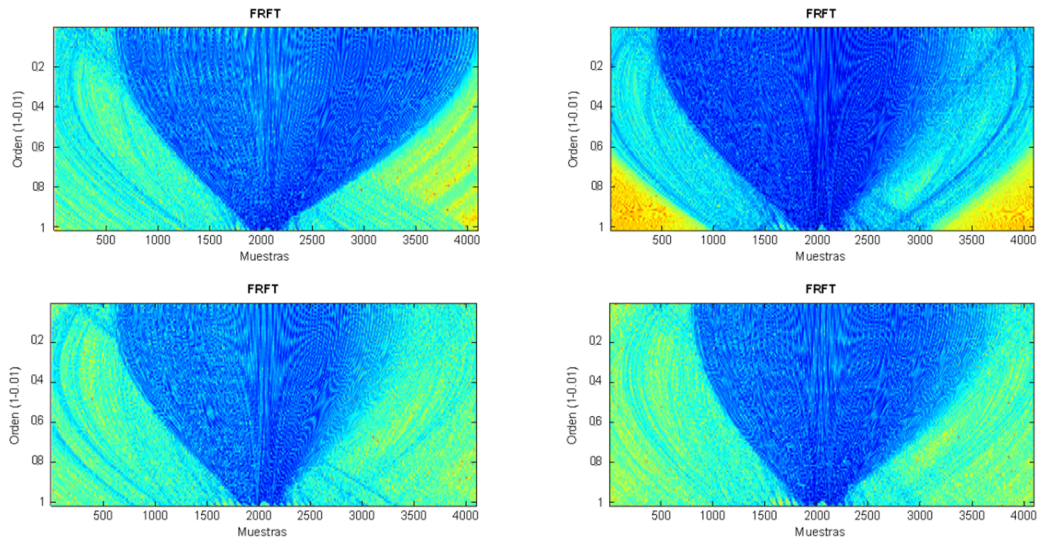


Fig. 32 Correlación de la misma palabra en un locutor para diferentes realizaciones en momentos distintos

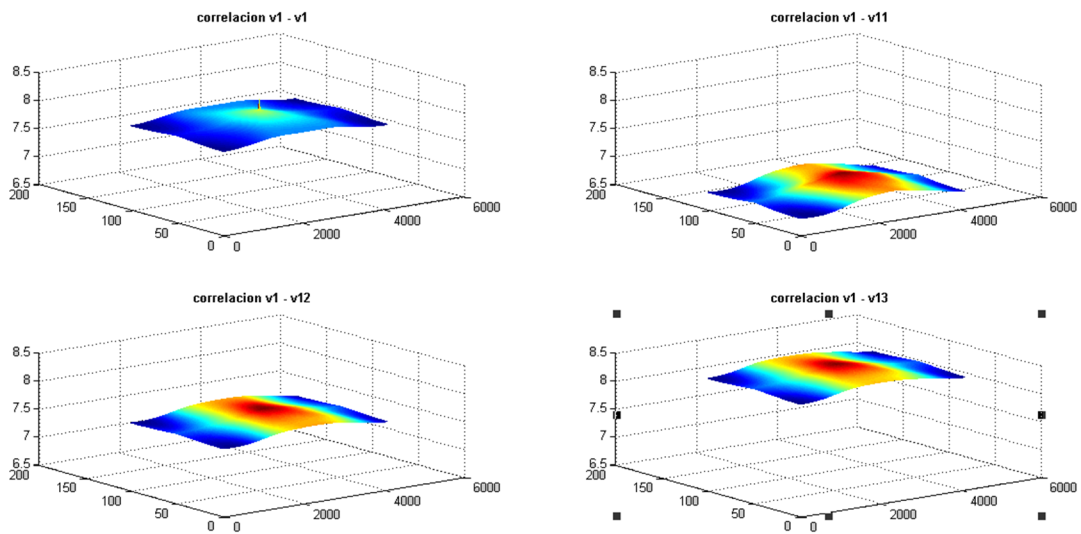


Fig. 33 Correlación de la misma palabra en un locutor, se correlaciona la primera realización con las demás. Arriba a la izquierda la autocorrelación de la primera realización.

CAPITULO 4

PARÁMETROS DE CORRELACIÓN

Para validar la confiabilidad de los resultados obtenidos se procede a calcular las métricas de la correlación que se enuncian a continuación.

4.1 Parámetro relación señal a ruido (del inglés SNR "signal to noise ratio")

La relación señal a ruido es definida como la relación entre la intensidad del pico de correlación y la desviación estándar del ruido. Se define matemáticamente como la media (μ) sobre la desviación estándar (σ);

$$SNR = \frac{P_c}{\sqrt{\frac{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} n(x_i, y_j) - \overline{n(x_i, y_j)}}{N_1 N_2}}} \quad (23)$$

Donde P_c es la media del pico de correlación de la FrFT. Y el denominador como la variación de las medidas con respecto a los picos de correlación en presencia de ruido.

Para la realización de los cálculos se emplea la desviación estándar de una región de 40x40 píxeles alrededor del pico de correlación y una región excluyente que contiene el pico de correlación de 30x30 píxeles.

Cuando se calcula el parámetro SNR y éste tiene un valor menor que la unidad, muestra que la señal está por debajo del nivel de ruido, si por el contrario el SNR posee un valor mayor, el ruido tiene menos efecto sobre la calidad de la correlación.

En la Tabla 1, se muestra el cálculo del parámetro relación señal a ruido de una señal de voz (pronunciación de design), la cual se le ha calculado la correlación cruzada y la autocorrelación de la FrFT.

Locutores	SNR (db)
L1-L1	24,8495
L1-L2	17,5527
L1-L3	16,4334
L1-L4	16,9148
L1-L5	15,4448
L1-L6	15,5195

Tabla 1. SNR de la palabra design para diferentes locutores

Se evidencia que el parámetro SNR de 24,8495 dB. es cuando se realiza la autocorrelación del locutor 1, indicando que es el valor más alto, por tanto es cuando la señal presenta menos ruido con relación a los demás locutores.

La comparación con los demás locutores sean hombres o mujeres es mucho menor que cuando se realiza la autocorrelación. El SNR para un mismo locutor en diferentes momentos pronunciando la misma palabra (Jhon) arroja los resultados mostrados en la Tabla 2. El valor más significativo corresponde a la autocorrelación.

Locutores	SNR(db)
Señal original L13-L1	82,7010
Señal original L13-L11	97,3773
Señal original L13-L12	110,9208
Señal original L13-L13	151,1662

Tabla 2. SNR de la palabra Jhon para diferentes realizaciones de un mismo locutor.

4.2 Relación altura de pico lóbulo lateral PSR (del inglés “Peak sidelobe ratio”)

La relación altura de pico lóbulo lateral se usa para cuantificar la agudeza del pico de correlación, se define como

$$PSR = \frac{P_c - \mu}{\sigma} \quad (24)$$

Donde μ y σ representan la media y la desviación estándar del área donde está el pico de correlación. Si se calcula el parámetro PSR y su valor es más significativo con respecto a los demás indica que el pico de correlación es agudo.

Analizando los resultados de la Tabla 3, cuando se toma una señal de voz y se le calcula el PSR en la correlación de la FrFT se aprecia que el pico de correlación más agudo se presenta en el caso de la autocorrelación, y la diferencia con respecto a los demás locutores es amplia.

Locutores	PSR
L1-L1	81,6052
L1-L2	7,8899
L1-L3	6,2296
L1-L4	7,0103
L1-L5	7,0397
L1-L6	4,0516

Tabla 3. PSR de la palabra design para diferentes locutores.

Este parámetro también fue implementado para un mismo locutor en diferentes momentos pronunciando la misma palabra. La Tabla 4 presenta los resultados obtenidos. Se evidencia que se obtiene un margen amplio cuando se realiza la autocorrelación.

Locutores	PSR
Señal original L13-L1	1,0202
Señal original L13-L11	1,6024
Señal original L13-L12	2,2363
Señal original L13-L13	25,999

Tabla 4. PSR de la palabra Jhon para diferentes realizaciones del mismo locutor.

4.3 Relación pico a energía de correlación PCE (“Peak to Correlation energy”)

Es la medida de la concentración de la energía en el pico de correlación. Se define

como

$$PCE = \frac{P_c}{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |C(x_i, y_j)|^2} \quad (25)$$

Donde el denominador representa la energía de la función de correlación.

Analizando los resultados de la Tabla 5 cuando se toma una señal de voz, y se le calcula el PCE en la correlación de la FrFT se obtiene que el valor más significativo es cuando se realiza la autocorrelación, para la correlación cruzada se observa que los valores mantienen una diferencia grande.

Locutores	PCE
L1-L1	458,439
L1-L2	24,9138
L1-L3	46,2539
L1-L4	44,0927
L1-L5	35,2547
L1-L6	37,0613

Tabla 5. PCE de la palabra design para diferentes locutores

De forma general en las tablas 7 y 8 del anexo B (literal B3), se presentan los resultados de los parámetros de correlación para cada una de las dieciséis palabras utilizadas.

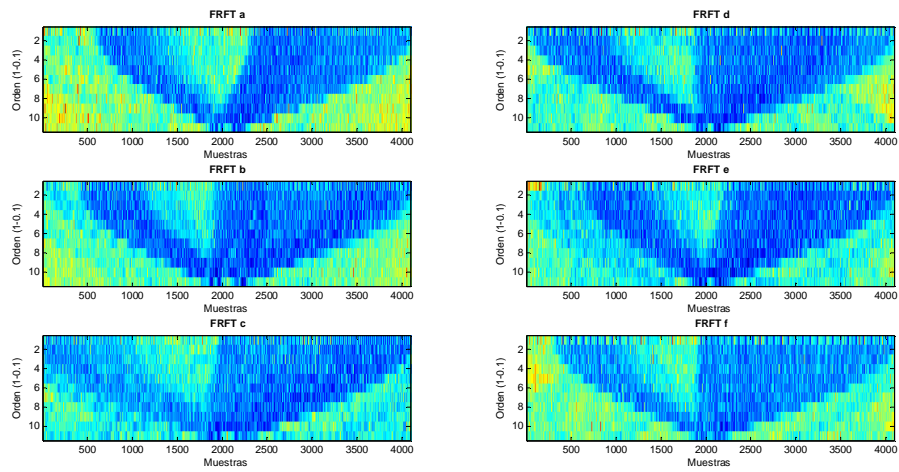
El PCE de un mismo locutor pronunciando las mismas palabras en diferentes momentos corresponde a los siguientes datos.

Locutores	PCE
Señal original L13-L1	9,7015
Señal original L13-L11	3,2098
Señal original L13-L12	7,2372
Señal original L13-L13	19,3875

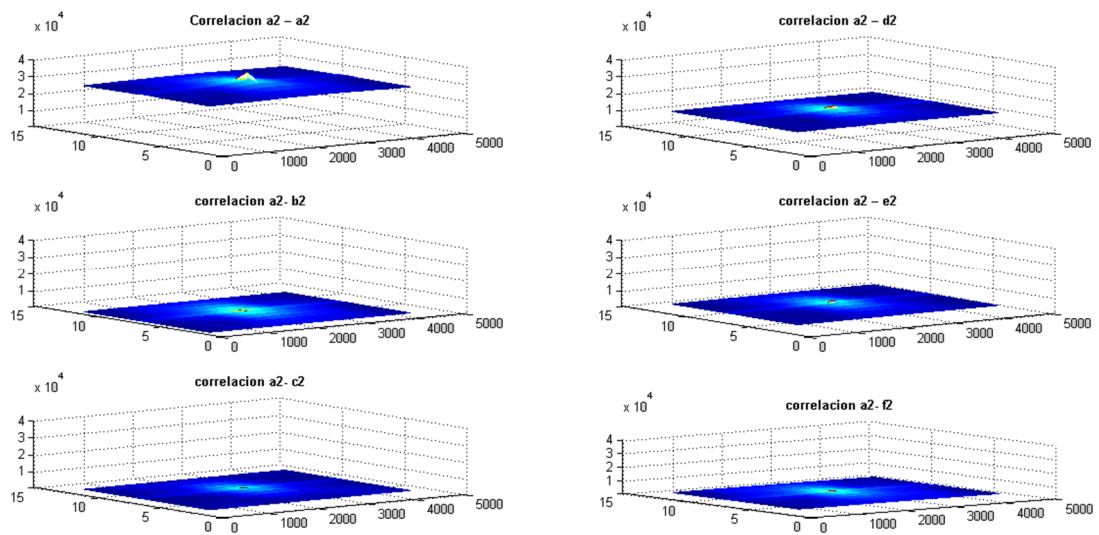
Tabla 6. PCE de la palabra Jhon para diferentes realizaciones del mismo locutor.

En términos generales, cuando se realiza las simulaciones de un centenar de ordenes en la transformada de Fourier fraccional se puede encontrar problemas de redundancia de información. La ventaja de tener esta cantidad de datos variando suavemente los órdenes permite tener una representación más detallada de la señal, sin embargo se puede limitar el estudio a órdenes específicos o a un número más reducido de órdenes, lo que conlleva a aumentar notoriamente la velocidad de respuesta al implementar los algoritmos. Realizando diversas simulaciones y analizando los parámetros de correlación se puede llegar a establecer un margen adecuado de los órdenes a emplear.

Una de las simulaciones llevadas a cabo consistió en tomar un número más reducido de órdenes, se analizan grupos de órdenes de 10, 25 y 50 simultáneamente. La representación Tiempo- orden del módulo de la FrFT y su respectiva correlación se muestra en las figuras 34 a 36.

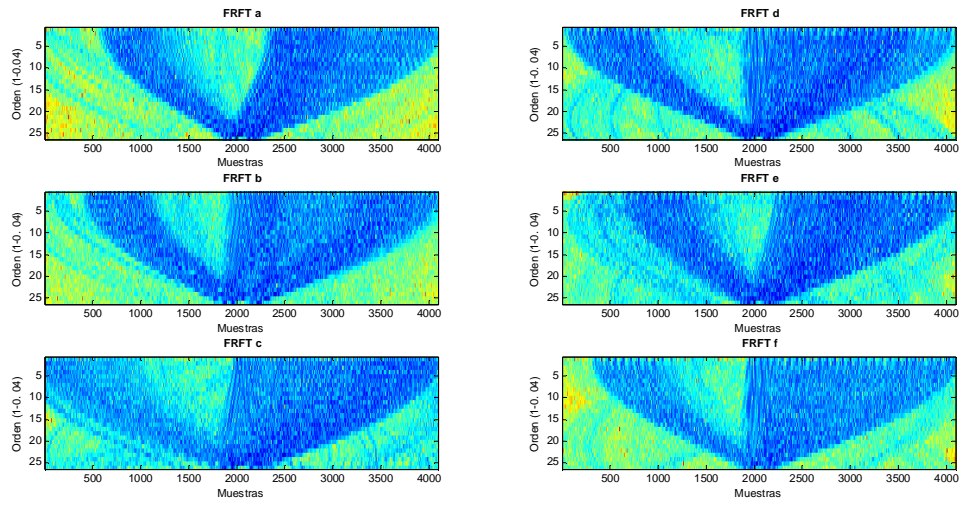


(a)

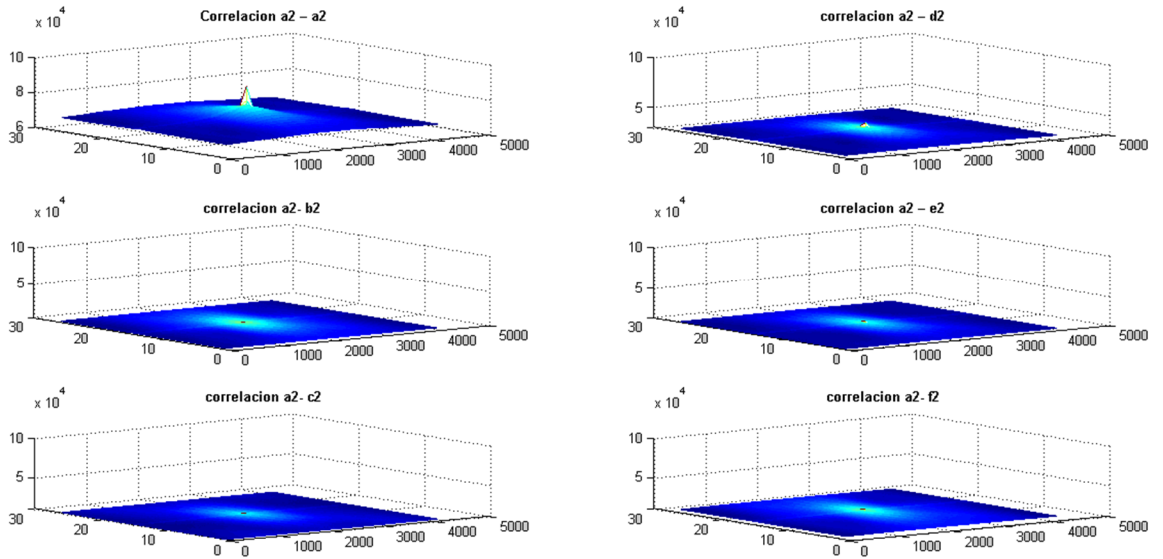


(b)

Fig. 34. (a) Representación Tiempo–Orden del módulo de la FrFT con 10 órdenes. (b) Correlación de la primera realización con las demás. Arriba a la izquierda la autocorrelación de la primera realización.

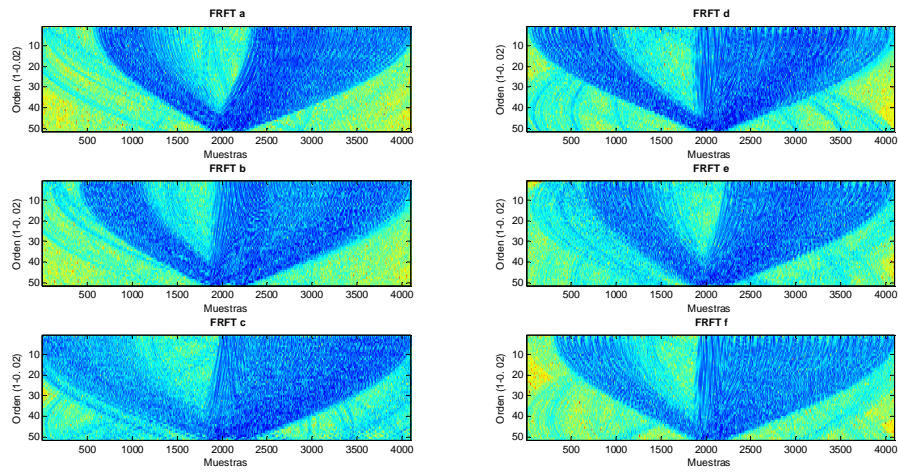


(a)

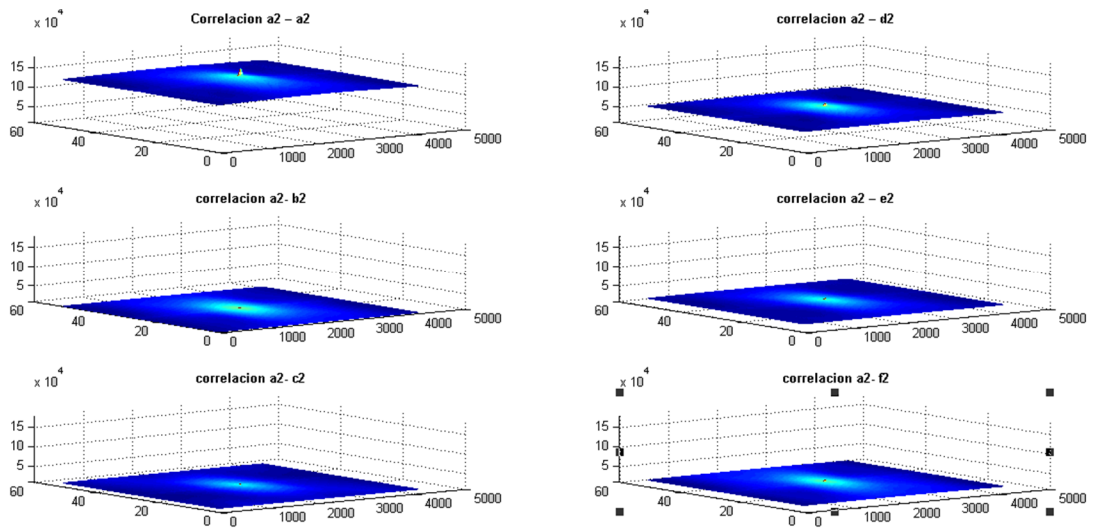


(b)

Fig. 35. (a.) Representación Tiempo–Orden del módulo de la FrFT con 25 órdenes. (b) Correlación de la primera realización con las demás. Arriba a la izquierda la autocorrelación de la primera realización.



(a)



(b)

Fig. 36. (a.) Representación Tiempo–Orden del módulo de la FrFT con 50 órdenes. (b) Correlación de la primera realización con las demás. Arriba a la izquierda la autocorrelación de la primera realización.

CONCLUSIONES

Emplear la representación tiempo-orden en la FrFT en el tratamiento de señales, permite indagar sobre nuevas posibilidades, buscando una discriminación de características inherentes en las señales de voz aprovechando la redundancia de información de la representación misma, en contravía de las representaciones clásicas que buscan desde el comienzo una reducción de la información.

El tratamiento simultáneo de un número considerable de órdenes en la transformada de Fourier de orden fraccional, resulta ser ventajoso en la representación Tiempo-Orden, ya que permite ver las variaciones al pasar de un orden a otro gradualmente. Es incluyente porque considera sus extremos, la representación en el espacio directo, en este caso el tiempo, y la representación en el espacio recíproco frecuencial o de Fourier estándar.

El algoritmo implementado por simulación digital muestra buenos resultados, aquellos que permiten ser evaluados cuantitativamente frente a otras técnicas bien conocidas buscando en lo posible una discriminación de hablantes. Los parámetros de correlación sobre la representación redundante tiempo-orden, evaluados condujeron a extraer características propias en cada locutor.

Cualquier tecnología biométrica empleada de forma apropiada para una aplicación específica lleva resultados fiables, sencillos y muy cómodos de usar en múltiples entornos. La implementación de la FRFT es una herramienta importante para el procesamiento de señales que conlleva a investigaciones en áreas que requieren explorar hacia la búsqueda de mejores resultados.

Los experimentos numéricos conducen a la necesidad de adelantar un estudio detallado para la selección apropiada del número de órdenes fraccionales que deben ser tenidos en cuenta al momento de realizar la representación redundante espacio directo-orden.

Es necesario continuar con el trabajo, tal vez depurando la representación espacio directo-orden, de manera que se pueda pasar a las etapas de verificación y reconocimiento de locutores, y porque no del reconocimiento de voz. Para ello habría que buscar un espacio de representación apropiado de los resultados de la correlación de las señales bidimensionales redundantes espacio directo-orden, que parecen ser lo suficientemente discriminantes.

REFERENCIAS

- [1] Sarikaya, R., Gao Y., Saon G. "Fractional Fourier Transform Features For Speech Recognition", Acoustics, Speech, and Signal Processing, I-529, 2004. Proceedings. (ICASSP '04). IEEE International Conference on.
- [2] Ozaktas, Kutay M. and Mendlovic, "The Fractional Fourier Transform: with Applications in Optics and Signal Processing", John Wiley & Sons, New York, 2001.
- [3] Zanuy Faundez M., "Tratamiento digital de voz e imagen", Primera edición, Editores Alfaomega, España, 90pp., 2001.
- [4] Furui, Sadaoki, "Digital Speech Processing, Synthesis and Recognition", Marcel Dekker, Inc, USA, 291pp., 1994.
- [5] Furui, Sadaoki, "An overview of speaker recognition technology", Proceeding of the ESCA. Workshop on Automatic Speaker/Recognition technology, vol. 4, pp.4072-4075, 2002.
- [6] Mendlovic D. Z, Kutay A., Ozaktas H, et. al., "Improved acoustic signals discrimination using fractional Fourier transform based phase-space representations", Optics Communications, vol. 190, Issue 1-6, pp.95-101, 2001.
- [7] Gonon G., Fellah Z., Depollier C., "Filtering in Fractional Fourier Domains: Application to the Biot's Waves", Laboratoire d'Informatique de l'Université du Maine, Francia, 5pp., 1998.

- [8] A. V. Oppenheim, R. Schafer, "Tratamiento de Señales en tiempo discreto". Editorial Prentice Hall, 904pp., 2000.
- [9] Pierre P. F. "Lecciones de Óptica de Fourier Fraccional", Ediciones UIS: Bucaramanga, ISBN: 958-8187-32-X, 252pp. 2004.
- [10] Champeney, "Transformada de Fourier y sus aplicaciones físicas", Academic Press, New York, 256pp. 1973.
- [11] Papoulis, A. "The Fourier Transforms and Its Applications", McGraw-Hill, New York, 328pp. 1962.
- [12] Bracewell, R., "The Fourier Transforms and It's Applications" McGraw-Hill, New York, 381pp. 1965.
- [13] Cariolaro, G. et al., "A unified framework for the fractional Fourier transform", IEEE Trans. Signal Processing, Italia, vol. 46, pp 3206 – 3219. 1998.
- [14] Ozaktas, H.M. and Mendlovic, D., "Fractional Fourier Optics", Journal of Optical Society America, vol.12, pp. 742–751, 1995.
- [15] Chang Soo, Yen M, Tseng C., "Discrete Fractional Fourier Transform Base on Orthogonal Projection", IEEE Trans. Signal Processing, vol. 47, No 5, pp. 1335 – 1348. 1999.
- [16] Cornejo J, Torres Y. "Implementación de un Algoritmo para Calcular la Transformada de Fourier Fraccional usando un Procesador Digital de Señales". Universidad Industrial de Santander, Escuela de Física. Tesis de Pregrado, 2003.

- [17] Cornejo J, Torres Y. "Correlador Híbrido de Transformada Conjunta empleando un DSP", Universidad Industrial de Santander, Escuela de Física. Tesis de Maestría, 2005.
- [18] Martínez C. "Estadística y Muestreo", ECOE ediciones, Bogotá, edición 12, 393pp., 2005.
- [19] Wolf. K.B."Integral Transforms in Science and Engineering". Plenum Press, New York., 255pp, 1979.
- [20] Demars C., "Représentations bidimensionnelles du signal de parole; éléments de monographie", <http://christian.jl.demars.free.fr/rapport2004.pdf.gz>, 407pp. 2005.
- [21] Demars C., "Représentation du signal analytique, représentation temps-fréquence instantanée", <http://christian.jl.demars.free.fr/safi/safi.pdf.gz>, 108pp., 2006.
- [22] Demars C., Gauvain J., "Application de la distribution λ^2 a la reconnaissance de la parole", Dans journal .14 ème JEP, pp. 283-287, Juin 1985, Paris
- [23] Ezzaidi H, Rouai J, et Al., "Combining pitch and MFCC for speaker recognition systems", ERMETIS, Université du Québec à Chicoutimi, Canada, pp.2825-2828, 2001.
- [24] Deshmukh, Om; Espy-W, Carol Y.; Juneja, A; University of Maryland, Department of Electrical & Computer Engineering, A. V. Williams Bldg., College Park, 20752, USA. Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on ISSN: 1520-6149 , ISBN: 0-7803-7402-9 . 10.1109/ICASSP.2002.5743787, pp. 593-596, FL, USA. 2002.

- [25] Jinfang W., Jinbao W., "Automatic Identification Advanced Technologies", 2005. Fourth IEEE Workshop on, Changchun, pp. 16-18, China.
- [26] Zanuy Faundez, Satué Villar A., "Identificación De Locutor Sobre Base De Datos Bilingüe", Escola Universitària Politècnica de Mataró (Adscrita UPC), pp. 632-633, 2002.
- [27] Ezzaidi H, et Al., "Pitch and MFCC dependent GMM models for speaker", citeseer.ist.psu.edu/646091.html. pp. 43-46, 2004.
- [28] Sejdić, et Al., "Time–frequency feature representation using energy concentration: An overview of recent advances", Digital Signal Processing, Elsevier. 19, pp. 153–183, 2009.
- [29] Wen, Zhang, Zhao et Al., "The Analysis of the Discrete Fractional Fourier Transform Algorithms", pp. 979 -982,2009. 1-4244-3508, 2009 IEEE.
- [30] Zhenli, Zhang, et Al., "On the application of fractional Fourier transform for enhancing noisy speech", IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications Proceeding. pp. 289 – 292, Vol. 1, 10.1109/MAPE.2005.1617906, 2005.
- [31] Barbu, M.; Kaminsky, E.; Trahan, R.E.; , "Acoustic Seabed Classification using Fractional Fourier Transform and Time-Frequency Transform Techniques", OCEANS 2006. pp.18-21, Sept. 2006
 [10.1109/OCEANS.2006.306889].URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4099044&isnumber=4098825>.

- [32] Barbu, M.; Kaminsky, E.J.; Trahan, R.E.; , "Fractional Fourier transform for sonar signal processing", OCEANS, 2005. Proceedings of MTS/IEEE, pp.1630-1635 Vol. 2, 17-23 Sept. 2005. [10.1109/OCEANS.2005.1639989].URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1639989&isnumber=34367>.
- [33] Aldirmaz, S.; Durak, L.; Serbes, A. "Fractional Fourier domain LMS-based adaptive filtering algorithms in active noise control", Signal Processing, Communication and Applications Conference, 2008. SIU 2008. IEEE 16th, pp.20-22, April,2008.
URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4632655&isnumber=4632536>.
- [34] Hui Yin; Nadeu, C.; Hohmann, V.; Xiang Xie; Jingming Kuang; , "Order Adaptation of the Fractional Fourier Transform Using the Intraframe Pitch Change Rate for Speech Recognition," Chinese Spoken Language Processing, 2008. ISCSLP '08. 6th International Symposium on. pp. 16-19 Dec. 2008.doi:10.1109/CHINSL.2008.ECP.60.
URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4730314&isnumber=4730254>.
- [35] Ji Xi; Li Zhao; Chong Wang; Qing Yun Wang; , "A Scheme of Robust Audio Watermarking Based on FRFT for Copyright Protection", Management and Service Science, 2009. MASS '09. International Conference on, 4pp. 20-22 Sept. 2009. [10.1109/ICMSS.2009.5305753]
URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5305753&isnumber=5300803>.

- [36] Wei-Qiang Zhang; Liang He; Tao Hou; Jia Liu; , "Fractional Fourier transform based auditory feature for language identification", Circuits and Systems, 2008. APCCAS 2008. IEEE Asia Pacific Conference on, pp. 209-212, Nov. 30 2008-Dec. 3 2008. [10.1109/APCCAS.2008.4745997]
URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4745997&isnumber=4745943>.
- [37] Yongjin Wang; Ling Guan; , "Recognizing Human Emotional State From Audiovisual Signals", Multimedia, IEEE Transactions on , vol.10, pp. 659-668, June 2008.
URL:<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4523966&isnumber=4523952>.
- [38] Gervaise, Cedric; Stephan, Yann; Simard, Yvan; , "On the use of time frequency methods in a passive acoustic monitoring system", Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, pp.5646-5649, 14-19 March 2010
URL:<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5495262&isnumber=5494886>.
- [39] Torres R. et. Al. "Holografía numérica por transformación de Fourier fraccional", Memorias XX Congreso Nacional de Física. Armenia: Sociedad Colombiana de Física, pp.140 – 142, v.1, 2003.
- [40] Lizarazo Z, et. Al. "Transformada de Fourier fraccional aplicada a la Encriptación de Señales", Memorias VII Simposio de Tratamiento de Señales, Imágenes y Visión Artificial, pp.128 – 130, v.1, 2002.

- [41] Pellat-Finet P. *Optique de Fourier: Théorie métaxiale et fractionnaire*, primera edición, springer, Francia, 548pp., 2009.
- [42] Namias V., J. “The fractional order Fourier transform and its application in quantum mechanics”, *J. Inst. Math. Appl.* Vol. 25, pp. 241-265, 1980.
- [43] Almeida L.B., “The fractional Fourier transform and time-frequency representation”, *IEEE Trans. Sig. Proc.* Vol- 42, pp.3084- 3091, 1994.

ANEXOS

ANEXOS

Anexo A

A1. Representación de los MFCC

Apéndice B

B1. Sistemas lineales especiales con sus núcleos de transformación

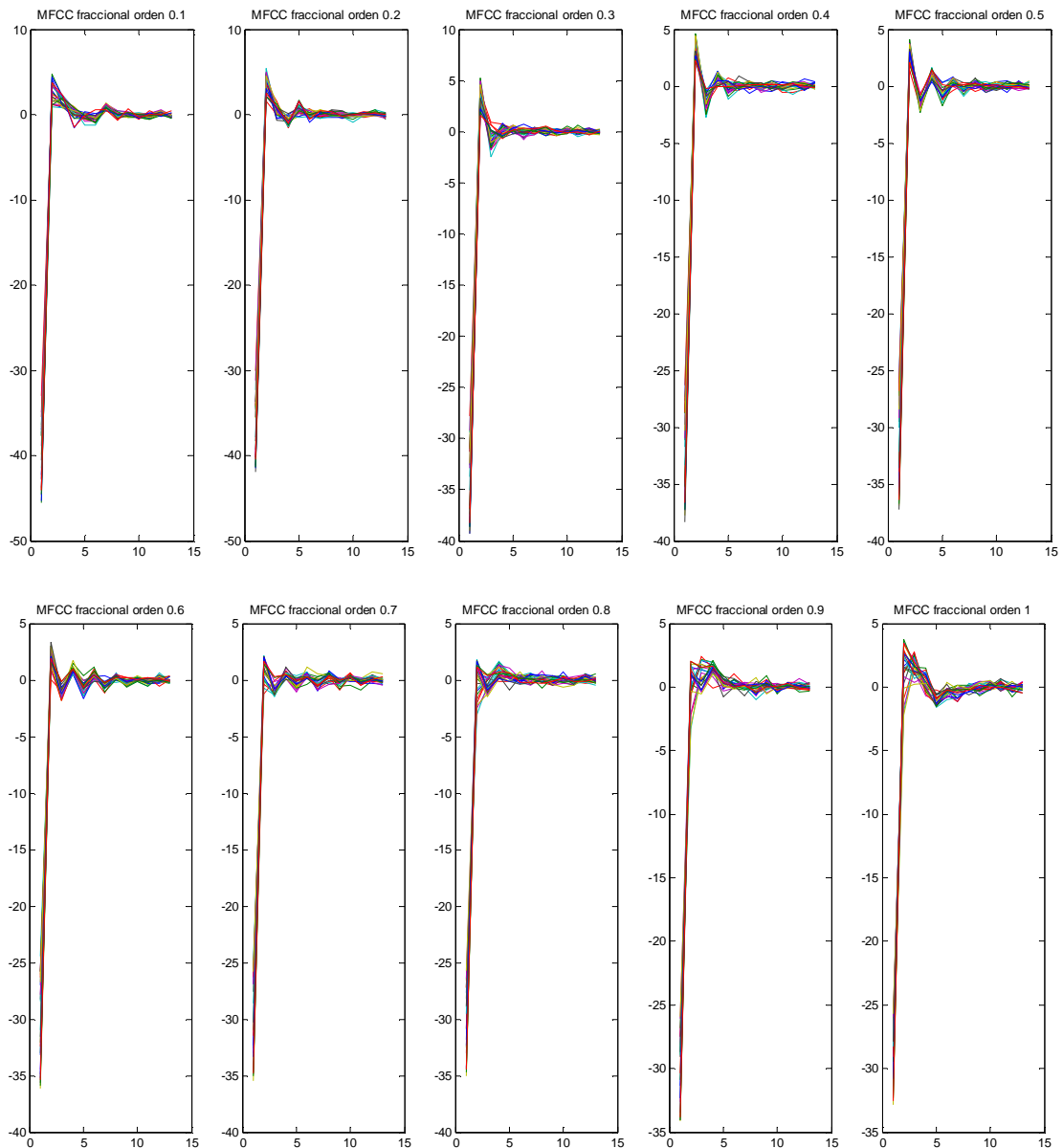
B2. Ejemplo de correlación

B3. Tablas Resultados de correlación

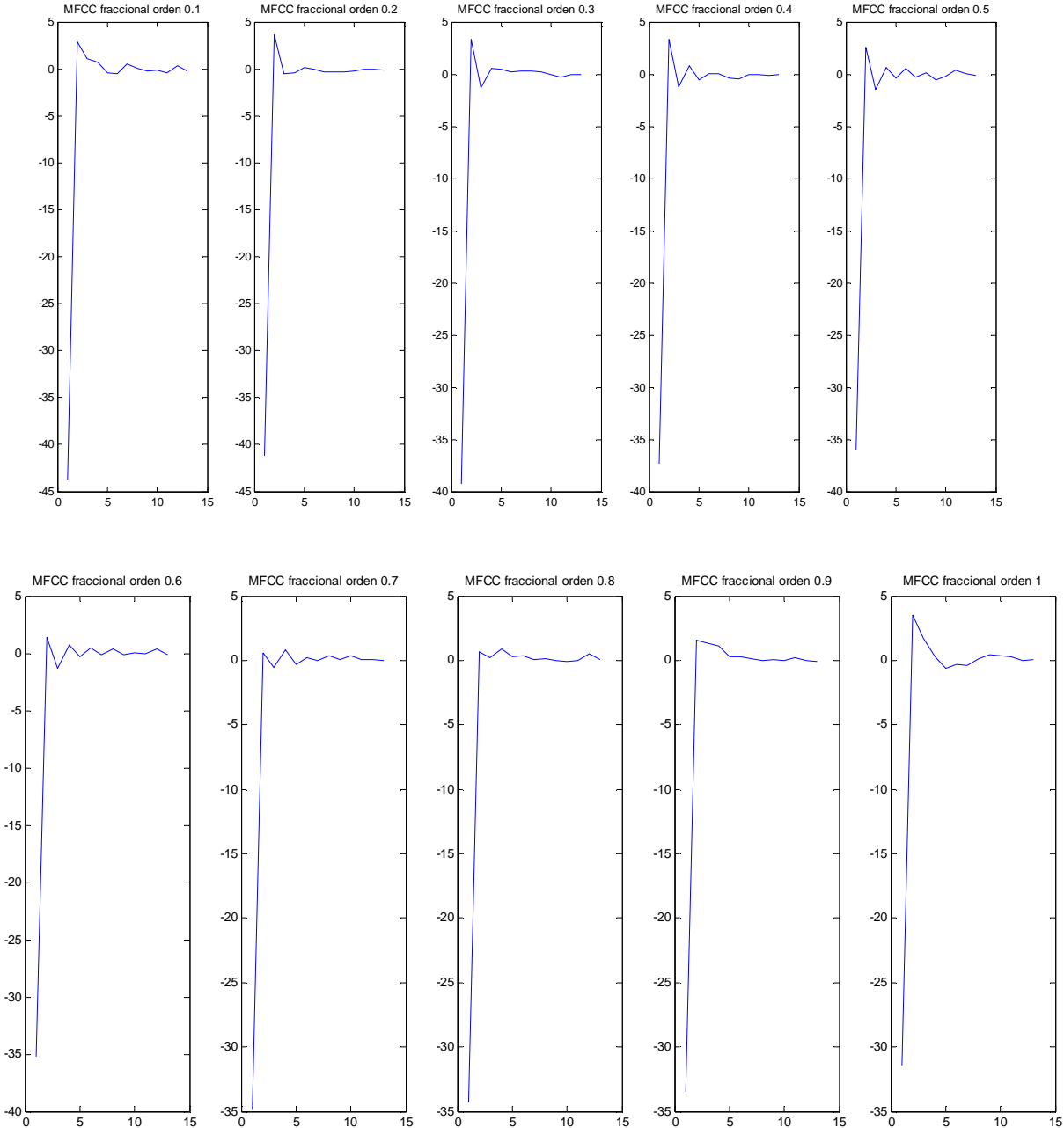
Anexo A

A1. REPRESENTACION DE LOS MFCC

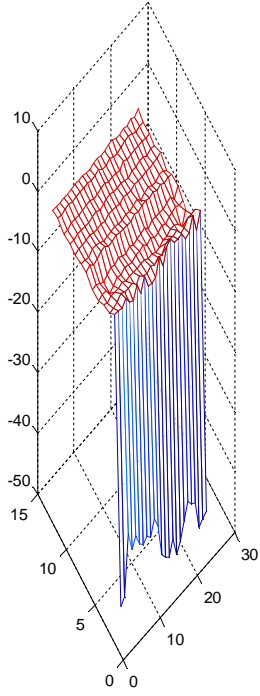
Si en lugar de usar la transformada de Fourier estándar en los cálculos de los MFCC de una señal de voz empleamos la transformada de Fourier fraccional para un número específico de ordenes simultáneos obtendríamos los siguientes resultados



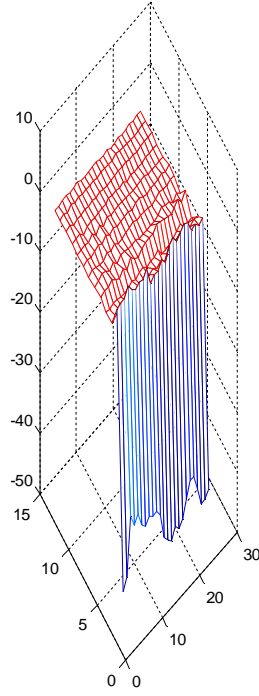
Las imágenes a continuación corresponden a los cálculos de MFCC de una señal de voz usando la FrFT donde se varia el orden del espacio temporal al espacio frecuencial con un paso de 0,1 que corresponden a tener diez órdenes; más adelante se emplea otra forma de representar los mismos datos, y por último se realiza la misma señal de voz en un locutor distinto.



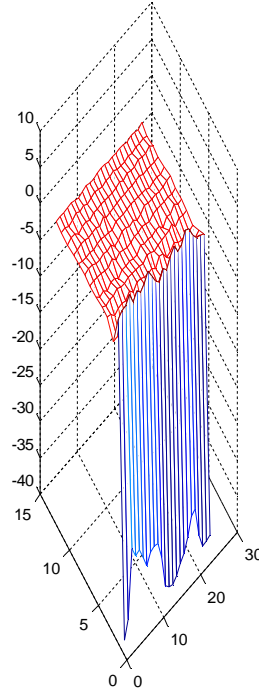
MFCC fraccional orden 0.1



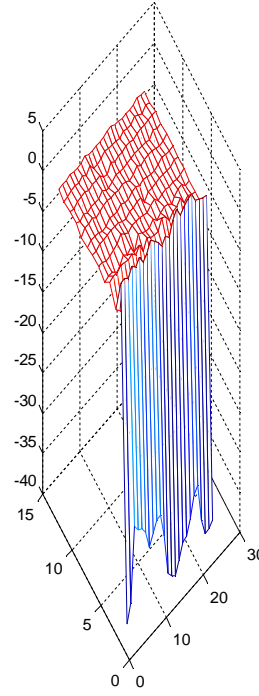
MFCC fraccional orden 0.2



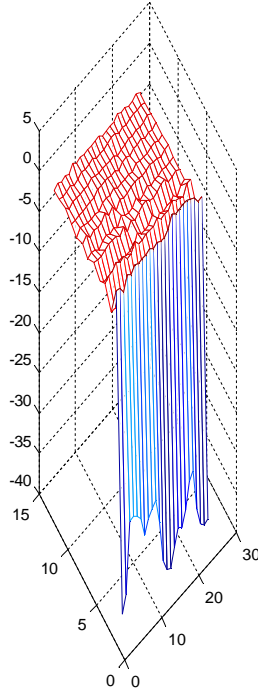
MFCC fraccional orden 0.3



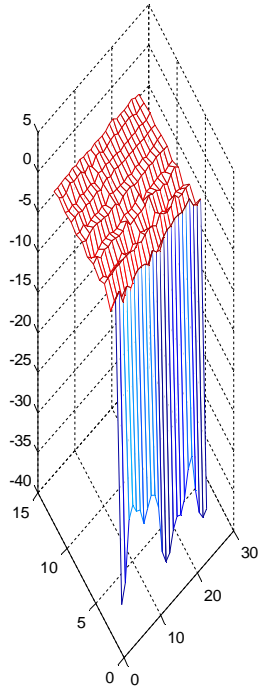
MFCC fraccional orden 0.4



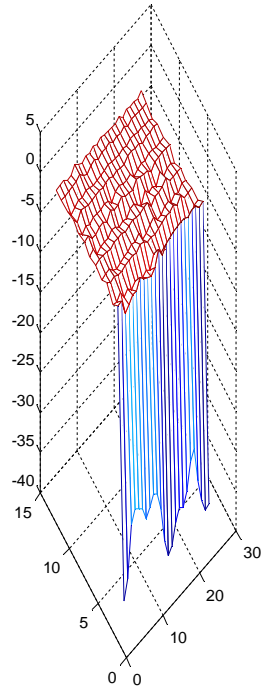
MFCC fraccional orden 0.5



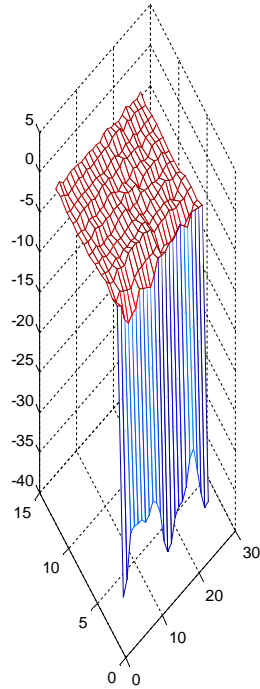
MFCC fraccional orden 0.6



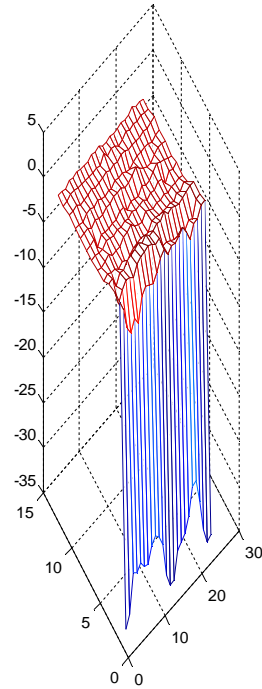
MFCC fraccional orden 0.7



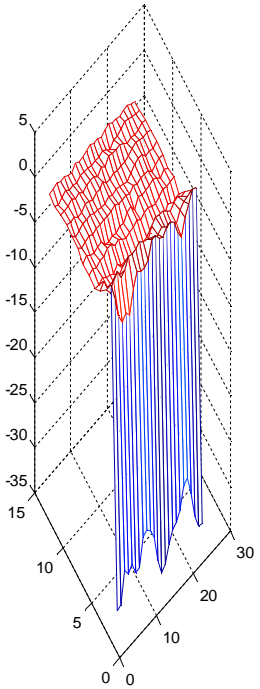
MFCC fraccional orden 0.8



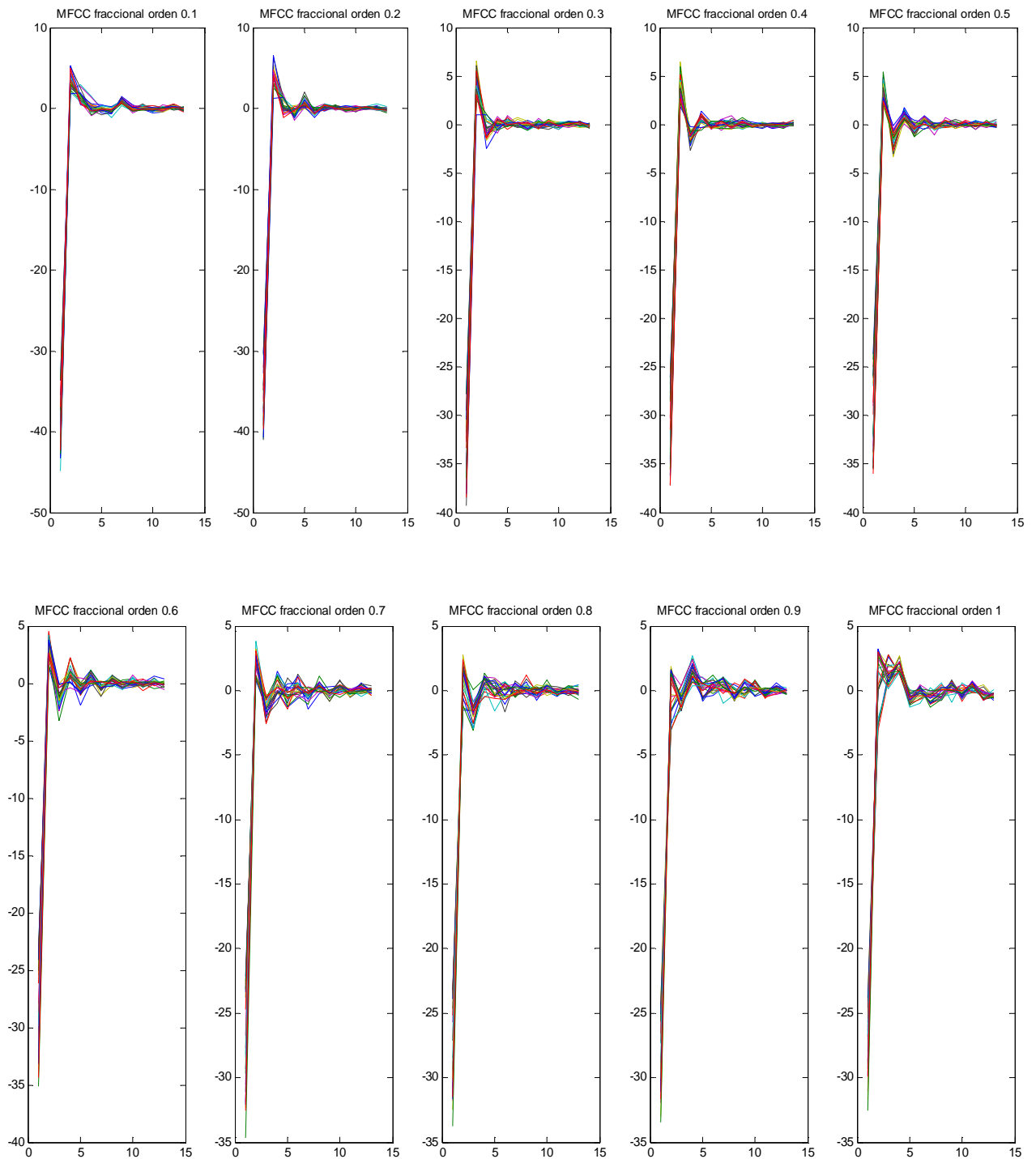
MFCC fraccional orden 0.9



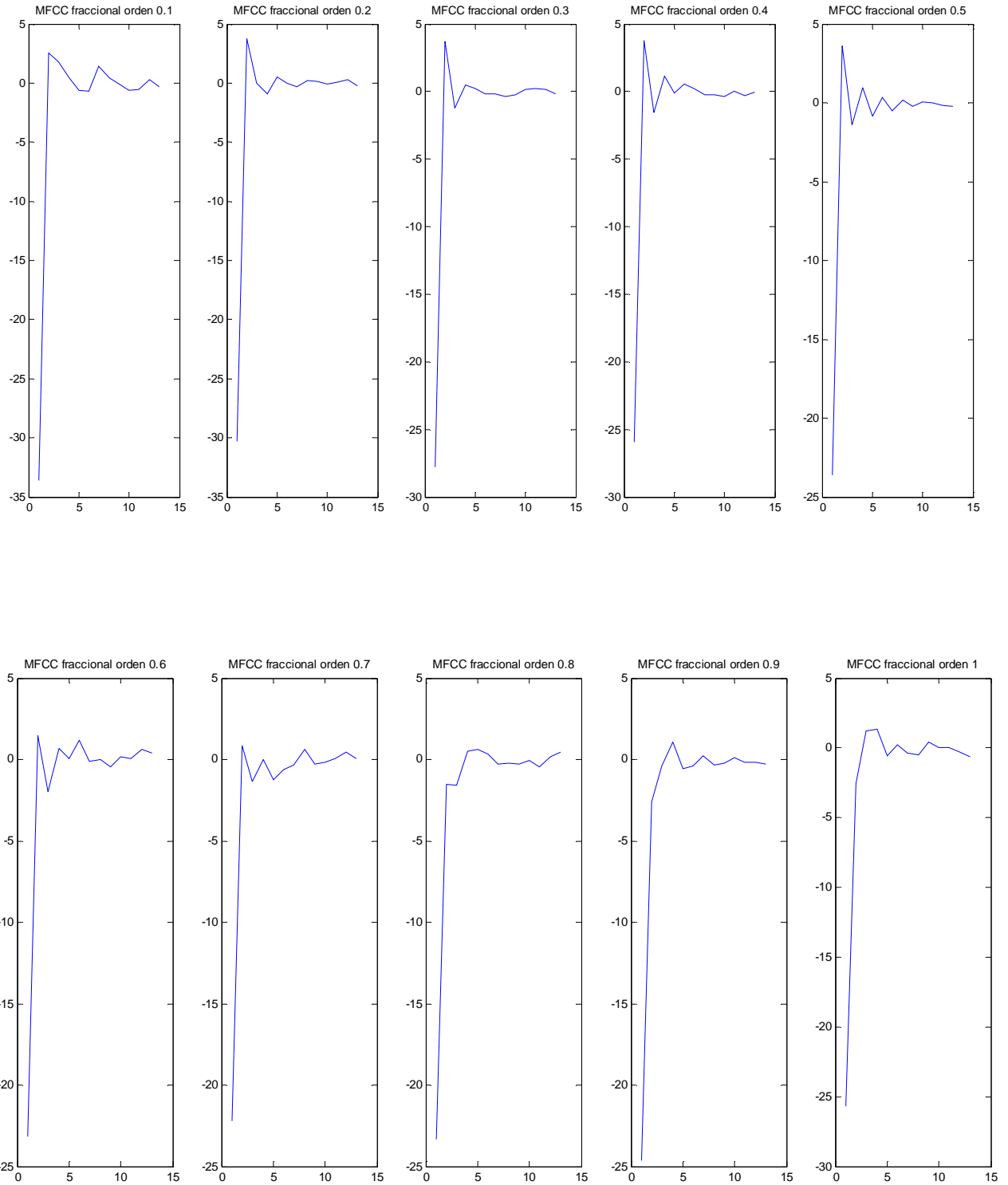
MFCC fraccional orden 1



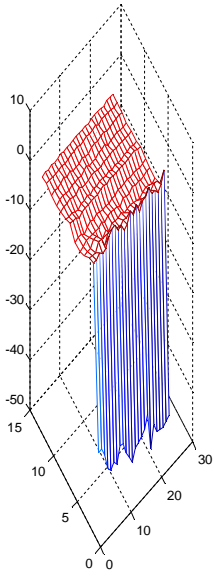
REPRESENTACION DE LOS MFCC SEÑAL 2



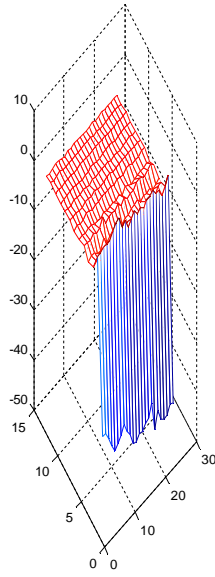
REPRESENTACION DE LOS MFCC SEÑAL 2



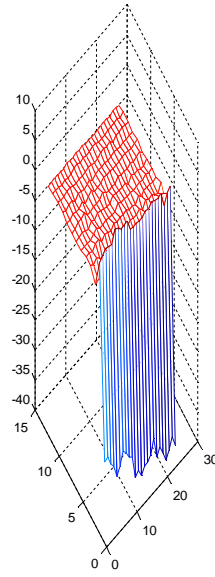
MFCC fraccional orden 0.1



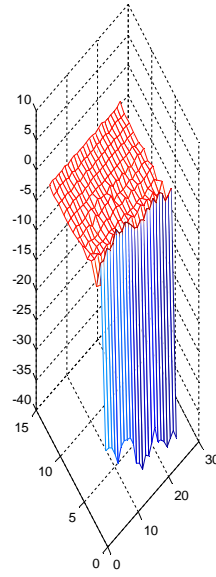
MFCC fraccional orden 0.2



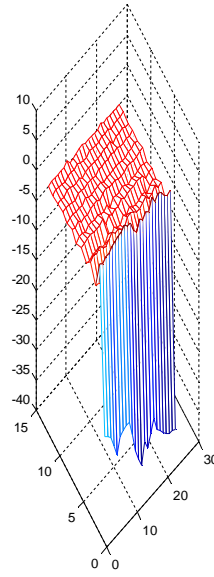
MFCC fraccional orden 0.3



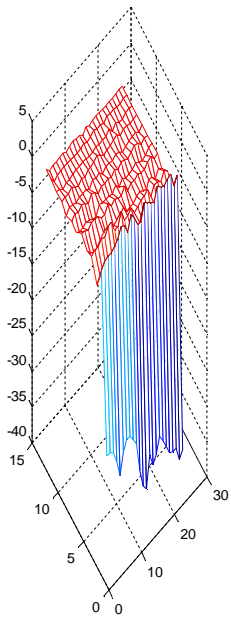
MFCC fraccional orden 0.4



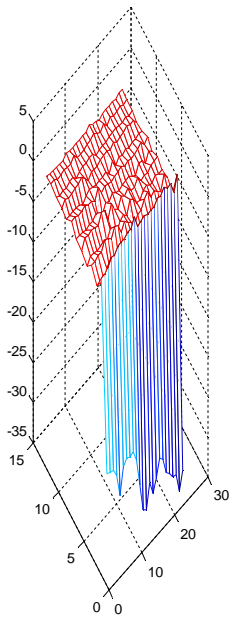
MFCC fraccional orden 0.5



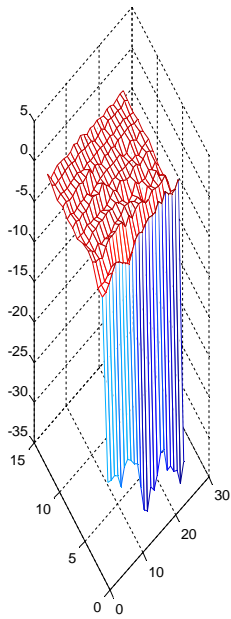
MFCC fraccional orden 0.6



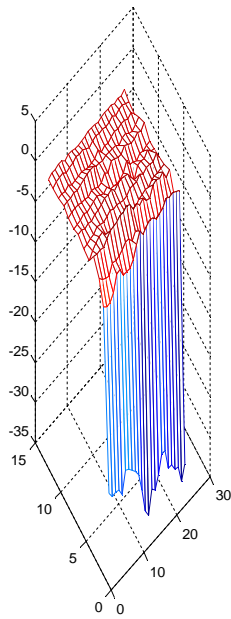
MFCC fraccional orden 0.7



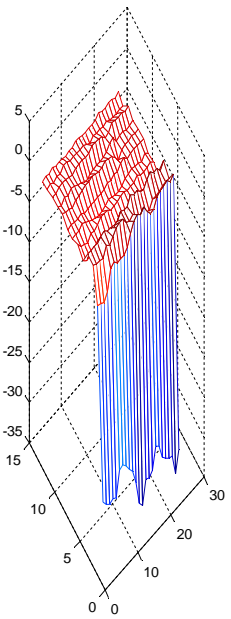
MFCC fraccional orden 0.8



MFCC fraccional orden 0.9



MFCC fraccional orden 1



Anexo B

B1. SISTEMAS LINEALES ESPECIALES CON SUS NÚCLEOS DE TRANSFORMACIÓN

	Kernel	Kernel Inverso
a. Identidad (I)	$\delta(u - u')$	$\delta(u - u')$

$$f(u) = \int f(u')\delta(u - u')du' = f(u) \text{ por la propiedad de tamizado}$$

	Kernel	Kernel Inverso
b. Paridad (P)	$\delta(u + u')$	$\delta(u + u')$

$$\begin{aligned} \int_{-\infty}^{\infty} \delta(u + u')f(u')du' &= f(-u) \\ \int_{-\infty}^{\infty} \delta(-(-u - u'))f(u')du' &= f(-u) \\ f(u) = \int g(u')\delta(u + u')du' &= g(-u) = f(u) \end{aligned}$$

	Kernel	Kernel Inverso
c. Escalamiento ($M_m \text{ m}\xi R$)		

$$\sqrt{|M|}\delta(u - Mu') \qquad \frac{1}{\sqrt{|M|}}\delta\left(u - \frac{u'}{M}\right)$$

Conociendo que $\delta(Mu) = \frac{\delta(u)}{|M|} \Rightarrow \delta\left(M\left(-u' + \frac{u}{M}\right)\right) = \frac{\delta\left(\frac{u}{M} - u'\right)}{|M|}$

$$\begin{aligned} g(u) &= \int \sqrt{|M|}\delta(u - Mu')f(u')du' = \\ &= \sqrt{|M|}\int \delta(u - Mu')f(u')du' = f(u) \\ &= \sqrt{|M|}\int \frac{\delta\left(\frac{u}{M} - u'\right)}{|M|}f(u')du' = \frac{1}{\sqrt{|M|}}f\left(\frac{u}{M}\right) \end{aligned}$$

$$f(u) = \frac{1}{\sqrt{|M|}} \int \delta\left(u - \frac{u'}{M}\right) g(u') du' = \frac{1}{M} \int f\left(\frac{u'}{M}\right) \delta\left(u - \frac{u'}{M}\right) du' = \frac{M}{M} \int \delta(v) f\left(\frac{M(u-v)}{M}\right) du' = f(u)$$

	Kernel	Kernel Inverso
d. Traslación (Shift) (SH _ξ ξR)	$\delta(u - u' + \xi)$	$\delta(u - u' - \xi)$

$$\int_{-\infty}^{\infty} \delta(u - u' + \xi) f(u') du' = g(u)$$

$$g(u - \xi) = \int_{-\infty}^{\infty} \delta(-u - \xi - u' + \xi) f(u') du' = \int_{-\infty}^{\infty} \delta(u - u') f(u') du' = f(u)$$

$$g(u) = f(u - \xi)$$

$$f(u) = \int g(u') \delta(u - u' - \xi) du' = g(u - \xi) = f(u - \xi + \xi) = f(u)$$

	Kernel	Kernel Inverso
e. Corrimiento de fase (PH _ξ ξR)	$\ell^{i2\pi\xi u'} \delta(u - u')$	$\ell^{-i2\pi\xi u'} \delta(u - u')$

$$g(u) = \int_{-\infty}^{\infty} e^{i2\pi\xi u'} \delta(u - u') f(u') du' = e^{i2\pi\xi u} \int_{-\infty}^{\infty} \delta(u - u') f(u') du' = e^{i2\pi\xi u} f(u)$$

$$f(u) = \int g(u') \ell^{-i2\pi\xi u'} \delta(u - u') du' = g(u) \ell^{-i2\pi\xi u} = f(u) \ell^{-i2\pi\xi u} \ell^{i2\pi\xi u} = f(u)$$

	Kernel	Kernel Inverso
f. Filtro Multiplicativo (Δh)	$h(u') \delta(u - u')$	$\frac{1}{h(u')} \delta(u - u')$

$$g(u) = \int f(u') h(u) \delta(u - u') du = f(u) h(u) \quad y$$

$$f(u) = \int g(u') \frac{1}{h(u')} \delta(u - u') du = \frac{g(u)}{h(u)} = \frac{f(u) h(u)}{h(u)} = f(u)$$

	Kernel	Kernel Inverso
g. Multiplicación chirp (\mathbb{Q}_q $q\xi\mathbb{R}$)	$\ell^{-i\pi q u^2} \delta(u - u')$	$\ell^{i\pi q u^2} \delta(u - u')$

$$g(u) = \int f(u') \delta(u - u') \ell^{-i\pi q u'^2} du' = f(u) \ell^{-i\pi q u^2}$$

$$f(u) = \int g(u') \delta(u - u') \ell^{i\pi q u'^2} du' = g(u) \ell^{i\pi q u^2} = f(u) \ell^{i\pi q u^2} \ell^{-i\pi q u^2} = f(u)$$

	Kernel	Kernel Inverso
h. Filtro de convolución (ΔH)	$h(u - u')$	$h^{-1}(u - u')$

$$g(u) = \int f(u') h(u - u') du' = f(u) * h(u) \quad y$$

$$f(u) = \int g(u'') h^{-1}(u - u'') du'' = \int \int f(u') h(u'' - u') h^{-1}(u - u'') du' du'' =$$

$$\int \int f(u') h^{-1}(u - u'') h(u'' - u') du'' du' = \int f(u') \delta(u - u') du' = f(u)$$

	Kernel Inverso	Kernel
i. Convolución chirp (\mathbb{R}_r $r\xi\mathbb{R}$)	$\ell^{-\frac{i\pi}{4}} \sqrt{\frac{1}{r}} \ell^{\frac{i\pi q(u-u')^2}{r}}$	

$$\ell^{\frac{i\pi}{4}} \sqrt{\frac{1}{r}} \ell^{\frac{i\pi q(u-u')^2}{r}}$$

Transformada de Fresnel

$$g(u) = \int \ell^{-\frac{i\pi}{4}} \sqrt{\frac{1}{r}} \ell^{\frac{i\pi q(u-u')^2}{r}} f(u') du' = \int \sqrt{\frac{1}{r}} \text{chirp}\left(\frac{u-u'}{\sqrt{r}}\right) f(u') du' = \sqrt{\frac{1}{r}} \text{chirp}\left(\frac{u}{\sqrt{r}}\right) * f(u)$$

$$f(u) = \int \ell^{\frac{i\pi}{4}} \sqrt{\frac{1}{r}} \ell^{\frac{-i\pi q(u-u')^2}{r}} g(u') du' = \int \sqrt{\frac{1}{r}} \text{chirp}\left(\frac{u'-u}{\sqrt{r}}\right) g(u') du' = \sqrt{\frac{1}{r}} \text{chirp}\left(\frac{-u}{\sqrt{r}}\right) * g(u) =$$

$$\sqrt{\frac{1}{r}} \text{chirp}\left(\frac{-u}{\sqrt{r}}\right) * \sqrt{\frac{1}{r}} \text{chirp}\left(\frac{u}{\sqrt{r}}\right) * f(u) = \frac{1}{r} \text{chirp}\left(\frac{-u}{\sqrt{r}}\right) * \text{chirp}\left(\frac{u}{\sqrt{r}}\right) * f(u) = \frac{1}{r} \delta(u) * f(u)$$

$$= f(u)$$

		Kernel	Kernel
		Inverso	
j.	Multiplicación por Coordenadas (U)	$u' \delta(u - u')$	$\frac{1}{u'} \delta(u - u')$

$$g(u) = \int f(u') u' \delta(u - u') du' = u f(u)$$

$$f(u) = \int \frac{g(u')}{u'} \delta(u - u') du' = \frac{1}{u} g(u) = \frac{1}{u} u f(u) = f(u)$$

		Kernel	Kernel Inverso
k.	Derivación (D)	$(i2\pi)^{-1} \delta'(u - u')$	$(i2\pi) \text{step}(u - u')$

$$g(u) = \int_{-\infty}^{\infty} (i2\pi)^{-1} \delta'(u - u') f(u') du'$$

$$= \frac{1}{i2\pi} \int_{-\infty}^{\infty} \frac{d\delta(u - u')}{du} f(u') du' =$$

$$= \frac{1}{2i\pi} \int \lim_{c \rightarrow 0} \left(\frac{\delta(u - u' + c) - \delta(u - u')}{c} \right) f(u') du'$$

$$= \frac{1}{2i\pi} \int \lim_{c \rightarrow 0} \frac{\delta(u - u' + c)}{c} f(u') du' - \frac{\delta(u - u') f(u') du'}{c}$$

$$= \frac{1}{2i\pi} \lim_{c \rightarrow 0} \frac{f(u + c)}{c} - \lim_{c \rightarrow 0} \frac{f(u)}{c} = \frac{1}{2i\pi} \lim_{c \rightarrow 0} \frac{f(u + c) - f(u)}{c} = \frac{1}{2i\pi} \frac{df(u)}{du}$$

$$f(u) = \int g(u') (i2\pi) \text{step}(u - u') du' = (i2\pi) \int g(u') \text{step}(u - u') du' = (i2\pi) \int_0^{\infty} g(u') du' =$$

$$(i2\pi) \int_0^{\infty} \frac{1}{i2\pi} \frac{d}{du} f(u') du' = f(u)$$

		Kernel	Kernel Inverso
l.	Transformada de Fourier (F)	$\ell^{-i2\pi u u'}$	$\ell^{i2\pi u u'}$

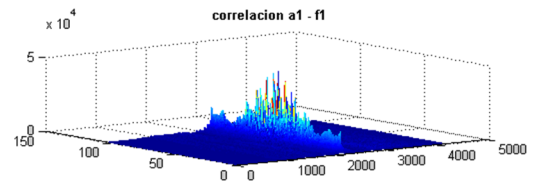
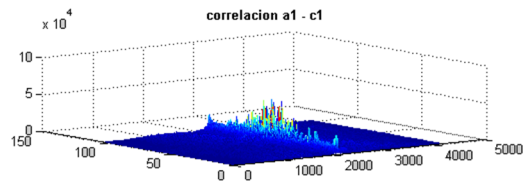
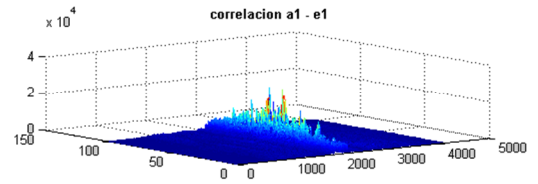
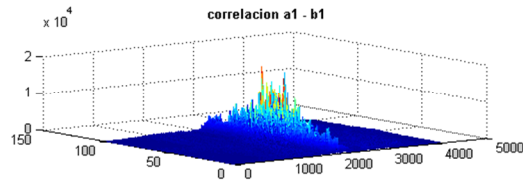
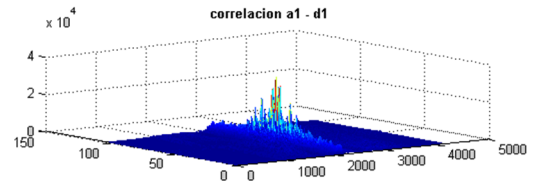
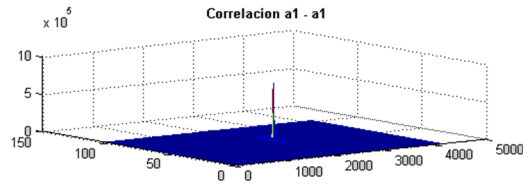
$$g(u) = \int f(u') \ell^{-i2\pi u u'} du'$$

$$f(u) = \int g(u') \ell^{i2\pi u u'} du' = \int \int f(u'') \ell^{-i2\pi u' u''} du'' \ell^{i2\pi u u'} du' = \int \int f(u'') \ell^{-i2\pi u' u''} \ell^{i2\pi u u'} du' du'' =$$

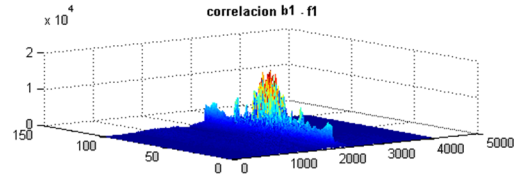
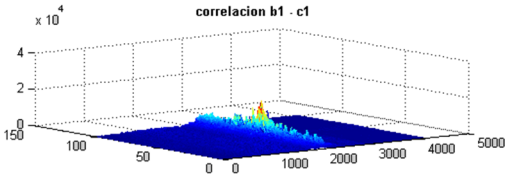
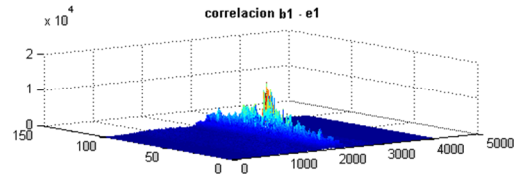
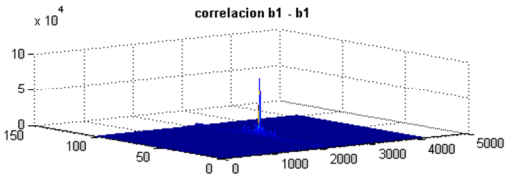
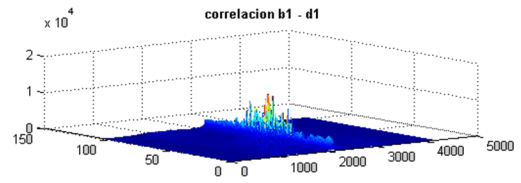
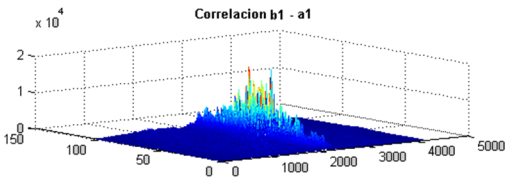
$$\int f(u'') \delta(u - u'') du'' = f(u)$$

B2. EJEMPLO DE CORRELACIÓN

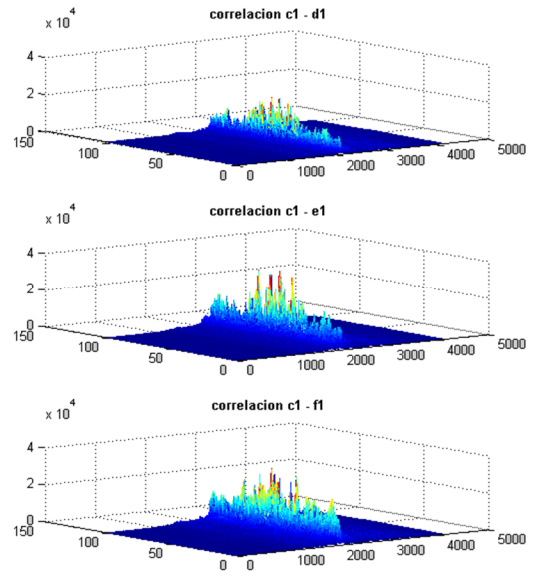
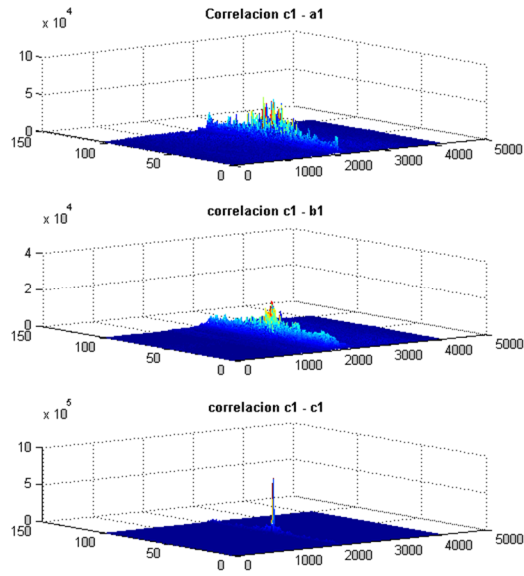
A



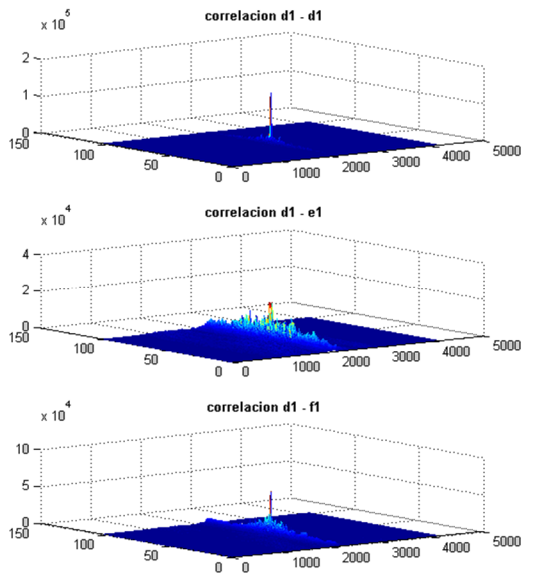
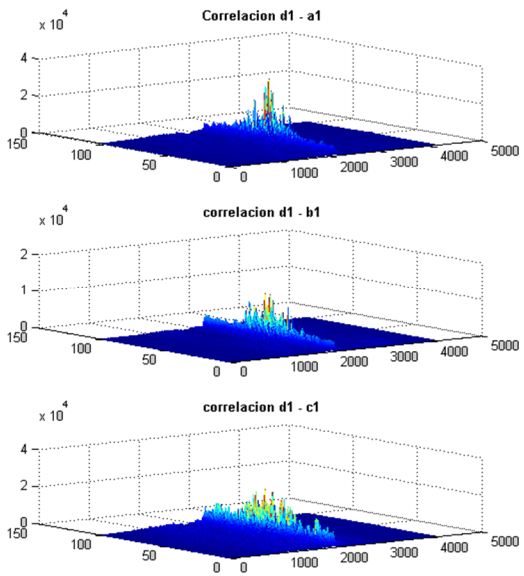
B



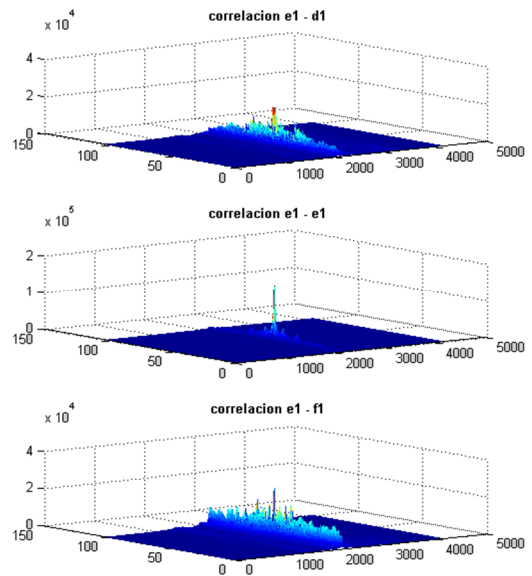
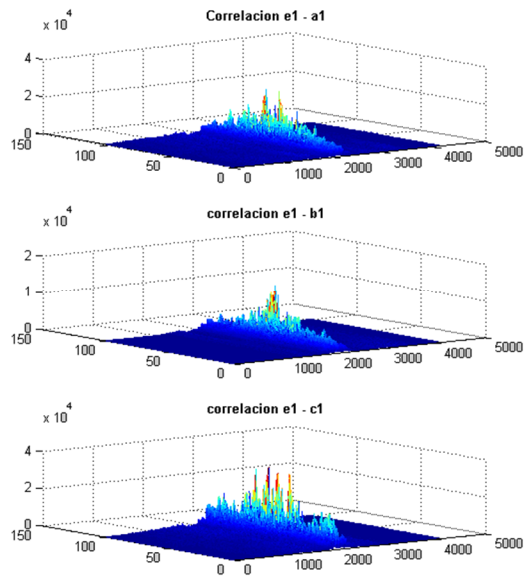
C



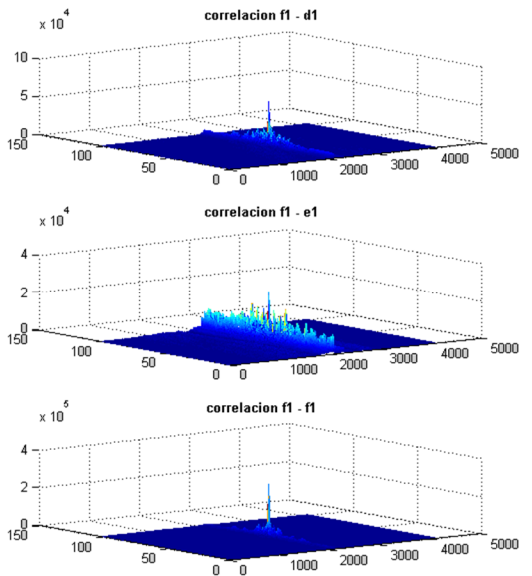
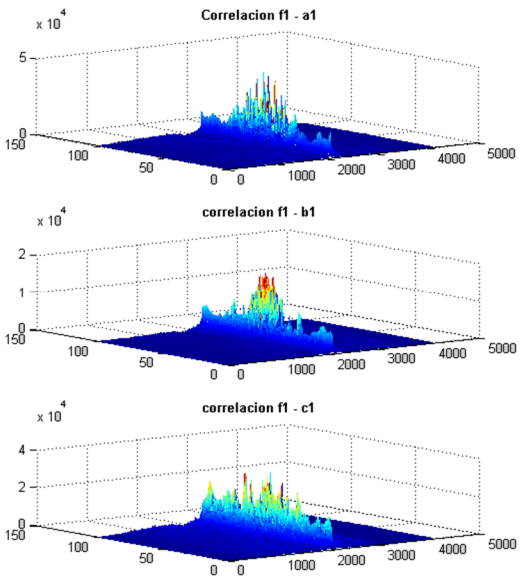
D



E



F



B3. Tablas Resultados de correlación

PARAMETROS DE CORRELACION

Palabra Parámetro	1	2	3	4	5	6	7	8	9	LOCUTOR
PSR	88,0334	81,6052	30,5960	76,1001	29,9107	46,0840	61,4268	63,1116	61,4415	L1-L1
	5,2727	7,8899	4,1345	8,5484	4,9514	4,9726	16,8756	4,8714	8,5893	L1-L2
	5,0886	6,2296	6,2968	7,3251	3,3536	6,6148	8,0019	4,4419	6,0348	L1-L3
	8,1815	7,0103	6,2968	11,5858	3,4997	4,7899	9,3846	5,2168	13,6282	L1-L4
	5,6448	7,0397	4,5173	7,3355	4,5023	4,3918	4,7482	6,0327	8,0401	L1-L5
	5,7504	4,0516	3,3723	4,7001	4,4076	4,4811	10,3910	4,0963	15,2474	L1-L6
PCE	12499	458,8439	602,7940	297,6250	635,4870	378,5313	441,8911	207,3317	179,185	L1-L1
	31,2445	24,9138	40,7811	8,7397	43,4728	31,0470	21,6431	27,6070	47,1581	L1-L2
	92,8262	46,2539	101,2591	27,9453	74,5111	44,4578	33,7763	30,9445	25,1710	L1-L3
	55,4265	44,0927	101,2591	24,8553	61,3500	28,0399	20,4395	20,7778	35,8571	L1-L4
	45,8710	35,2547	62,8313	13,4473	49,3212	26,5565	16,4117	19,3418	32,6049	L1-L5
	76,3872	37,0613	46,0905	8,0220	44,0481	24,9661	28,4675	19,4765	54,1577	L1-L6
SNR (db)	26,5275	24,8495	22,2186	25,5329	21,7630	22,9818	26,7998	24,6336	26,1582	L1-L1
	15,6634	17,5527	14,5574	16,7934	14,9091	14,8657	20,1460	15,3277	17,2200	L1-L2
	15,7581	16,4334	16,0376	16,4741	14,0737	16,3867	17,6469	15,0510	16,6089	L1-L3
	17,3352	16,9148	16,0376	17,9509	14,9117	15,0174	16,9613	14,6699	19,5126	L1-L4
	15,0447	15,4448	14,2789	15,7078	14,5330	14,2177	16,1406	15,0186	17,9457	L1-L5
	16,0830	15,5195	14,5239	14,4714	13,7342	14,2078	18,6246	14,3814	19,5592	L1-L6

Tabla 1 PSR, PCE Y SNR de las palabras 1-9 del locutor 1 Vs los seis locutores

PARAMETROS DE CORRELACION

Palabra Parámetro	10	11	12	13	14	15	16	LOCUTOR
PSR	55,4350	105,2034	79,6052	90,6052	54,6052	62,6052	85,6052	L1-L1
	5,1830	5,7208	11,8899	7,8899	11,8899	9,8899	9,8899	L1-L2
	12,0344	4,4814	10,2296	6,2296	10,2296	8,2296	8,2296	L1-L3
	7,7378	6,7701	11,0103	7,0103	11,0103	9,0103	9,0103	L1-L4
	9,7290	44,2932	11,0397	7,0397	11,0397	9,0397	9,0397	L1-L5
	7,7071	4,9137	8,0516	4,0516	8,0516	6,0516	6,0516	L1-L6
PCE	76,1767	583,5238	477,8439	277,8439	281,8439	279,8439	377,8439	L1-L1
	24,4454	17,9366	31,9138	21,9138	25,9138	23,9138	26,9138	L1-L2
	39,0141	31,1567	53,2539	43,2539	47,2539	45,2539	48,2539	L1-L3
	13,4426	24,3502	51,0927	41,0927	45,0927	43,0927	46,0927	L1-L4
	9,4525	86,8499	42,2547	32,2547	36,2547	34,2547	37,2547	L1-L5
	15,2912	40,4314	44,0613	34,0613	38,0613	36,0613	39,0613	L1-L6
SNR (db)	25,0879	26,1759	31,8495	25,8495	29,8495	27,8495	28,8495	L1-L1
	16,1727	17,1429	20,5527	16,5527	18,5527	17,5527	18,5527	L1-L2
	18,7473	19,4476	19,4334	15,4334	17,4334	16,4334	17,4334	L1-L3
	17,9671	18,6761	19,9148	15,9148	17,9148	16,9148	17,9148	L1-L4
	17,3965	17,3568	18,4448	14,4448	16,4448	15,4448	16,4448	L1-L5
	17,5335	17,3750	18,5195	14,5195	16,5195	15,5195	16,5195	L1-L6

Tabla 2 **PSR, PCE Y SNR** de las palabras 10-16 del locutor 1 Vs los seis locutores