



UNIVERSIDAD DE CARABOBO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA ELÉCTRICA



DEPARTAMENTO DE ELECTRÓNICA Y COMUNICACIONES

**EXTRACCIÓN DE ARTEFACTOS EN SEÑALES ELECTRO-FISIOLÓGICAS
UTILIZANDO “ANÁLISIS DE COMPONENTES INDEPENDIENTES ICA”**

FIGUEIRA, CARLOS
GÓMEZ, CARLOS

BÁRBULA, NOVIEMBRE DE 2012



UNIVERSIDAD DE CARABOBO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA ELÉCTRICA



DEPARTAMENTO DE ELECTRÓNICA Y COMUNICACIONES

**TRABAJO ESPECIAL DE GRADO PRESENTADO ANTE LA ILUSTRE
UNIVERSIDAD DE CARABOBO PARA OPTAR AL TÍTULO
DE INGENIERO ELECTRICISTA.**

AUTORES:
FIGUEIRA, CARLOS
GÓMEZ, CARLOS
TUTOR:
PROF. CÉSAR, SEIJAS

BÁRBULA, NOVIEMBRE DE 2012



UNIVERSIDAD DE CARABOBO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA ELÉCTRICA



DEPARTAMENTO DE ELECTRÓNICA Y COMUNICACIONES

CERTIFICADO DE APROBACIÓN

Los abajo firmantes, miembros del jurado asignado para evaluar el trabajo especial de grado titulado **“EXTRACCIÓN DE ARTEFACTOS EN SEÑALES ELECTRO-FISIOLÓGICAS UTILIZANDO “ANÁLISIS DE COMPONENTES INDEPENDIENTES ICA””**, realizado por los bachilleres: Figueira Bolaños Carlos Javier, C.I: 14.607.025, y Gómez Fuentes Carlos Alberto, C.I: 15.088.425, hacemos constar que hemos revisado y aprobado dicho trabajo.

Prof. César Seijas.

TUTOR

Prof. Sergio Villazana.

JURADO

Prof. Carlos Jiménez.

JURADO

BÁRBULA, NOVIEMBRE DE 2012



Dedicatoria

A Dios, por darme la oportunidad de vivir y por estar conmigo en cada paso que doy, por fortalecer mi corazón e iluminar mi mente y por haber puesto en mi camino a aquellas personas que han sido mi soporte y compañía durante todo el periodo de estudio.

A mis padres, por ser el pilar fundamental en todo lo que soy, en toda mi educación, tanto académica, como de la vida, por su incondicional apoyo perfectamente mantenido a través del tiempo.

A Mi hijo, para que veas en mí un ejemplo a seguir.

A mis amigos, Que nos apoyamos mutuamente en nuestra formación profesional y que hasta ahora, seguimos siendo amigos.

Todo este trabajo ha sido posible gracias a ellos.

Carlos Gómez



Dedicatoria

A Mis padres por darme la vida, quererme mucho, creer en mí y porque siempre me han brindado su apoyo. Gracias por darme una carrera para mi futuro, todo esto se los debo a ustedes.

A Mis sobrinos, que de ellos es el futuro. Quiero que sepan que en su tío Carlos tendrán a un amigo y aliado en sus aspiraciones profesionales.

A mis hermanos, tíos, primos, abuelos y amigos. Gracias por haber fomentado en mí el deseo de superación y el anhelo de triunfo en la vida.

Mil palabras no bastarían para agradecerles su apoyo, su comprensión y sus consejos en los momentos difíciles.

A todos, espero no defraudarlos y contar siempre con su valioso apoyo, sincero e incondicional.

Carlos Figueira



Agradecimientos

Gracias a Dios.

A mis padres, que siempre me han dado su apoyo incondicional y a quienes debo este triunfo profesional, por todo su trabajo y dedicación para darme una formación académica y sobre todo humanista y espiritual. De ellos es este triunfo y para ellos es todo mi agradecimiento.

A toda mi familia, por todo su apoyo.

A la UC, la Máxima Casa de Estudios, y a la Escuela de Ingeniería Eléctrica, que bajo la dirección del Prof. Carlos Jiménez y su cercanía para con los estudiantes, ha marcado las pautas para la correcta realización de este trabajo de grado. Al Prof. Sergio Villazana por el apoyo brindado, y sus valiosas críticas y sugerencias. Para mejorar la calidad del trabajo de investigación.

A nuestro tutor, el Prof. César Seijas, por su invaluable apoyo a lo largo de toda la investigación, que fue clave para obtener los resultados esperados.

Carlos Gómez



Agradecimientos

Primero y como más importante, me gustaría agradecer sinceramente a mi tutor de Tesis, Prof. César Seijas, su esfuerzo y dedicación, sus conocimientos, sus orientaciones, su manera de trabajar, su persistencia, su paciencia y su motivación han sido fundamentales para nuestra investigación.

También me gustaría agradecer a los profesores Sergio Villazana y Carlos Jiménez, por sus observaciones y sugerencias, que han tenido relevancia en la correcta realización de este trabajo de grado, y por supuesto les agradezco, sus aportes como excelentes docentes en mi formación académica.

Para ellos,

Muchas gracias por todo.

Carlos Figueira



ÍNDICE GENERAL

DESCRIPCIÓN	PÁG.
PORTADA	I
PÁGINA DE TÍTULO	II
CERTIFICADO DE APROBACIÓN	III
DEDICATORIAS	IV
AGRADECIMIENTOS	VI
ÍNDICE GENERAL	VIII
ÍNDICE DE CONTENIDO	IX
ÍNDICE DE FIGURAS	XII
ÍNDICE DE TABLAS	XXVIII
RESUMEN	XXIX



ÍNDICE DE CONTENIDO

INTRODUCCIÓN..... 1

CAPÍTULO I. PLANTEAMIENTO DEL PROBLEMA..... 3

 1.1 Planteamiento del problema: 4

 1.2 Justificación de la investigación 6

 1.3 OBJETIVOS DE LA INVESTIGACIÓN 7

 1.3.1 Objetivo General: 7

 1.3.2 Objetivos Específicos: 7

 1.4 Alcance: 8

 1.5 Limitaciones: 8

CAPÍTULO II. MARCO TEÓRICO..... 9

 2.1 ANTECEDENTES DE LA INVESTIGACIÓN. 10

 2.2 BASES TEÓRICAS. 12

 2.2.1 Señales electrofisiológicas..... 12

 2.2.1.2 Electromiograma (EMG)..... 12

 2.2.1.3 Electroencefalograma (EEG)..... 12

 2.2.1.4 Polisomnografía..... 14

 2.2.1.5 Electrocardiograma (ECG)..... 15

 2.2.1.6 Bases de datos de señales de electrocardiográficas. 15

 2.2.2 Procesamiento de señales. 20

 2.2.2.1 Introducción..... 20

 2.2.2.2 Señales en tiempo continuo y discreto. 23

 2.2.2.3 Señales en tiempo continuo. 24

 2.2.2.4 Señales en tiempo discreto 26

 2.2.2.5 Representación de las señales Digitales. 28



2.2.3. Señales Electrocardiográficas (ECG).....	29
2.2.3.1 Tipos de ondas que se registran.....	31
2.2.3.2 Principio de medición.....	33
2.2.3.2.1 Derivaciones de miembros.....	34
2.2.3.2.2 Derivaciones precordiales.....	35
2.2.3.3 Análisis de un ECG.....	36
2.2.3.3.2 Ritmo cardíaco.....	37
2.2.4 CLASIFICACIÓN DE SEÑALES.....	38
2.2.5 Proceso determinístico.....	39
2.2.6 Procesos aleatorios - datos aleatorios.....	40
2.2.7 Análisis de datos aleatorios.....	43
2.2.8 Propiedades estadísticas.....	45
2.2.8.1 Propiedades estadísticas de las variables aleatorias.....	45
2.2.8.2 Propiedades estadísticas de los procesos aleatorios.....	48
2.2.9 Estadísticas de alto orden.....	50
2.2.9.1 Momentos de orden n	50
2.2.9.2 Cumulantes.....	53
2.2.9.3 Propiedades de los cumulantes.....	59
2.2.10 Fundamentos básicos del Análisis de Componentes Independientes.....	61
2.2.10.1 Introducción.....	61
2.2.10.2 Representación de datos.....	62
2.2.10.2.1 Consideraciones estadísticas iniciales.....	62
2.2.10.2.2 Reducción de dimensión para la obtención de W	64
2.2.10.2.3 Uso de la independencia para la estimación de W	64
2.2.10.3 Separación Ciega de Fuentes.....	66
2.2.10.3.1 Observaciones de mezclas de señales desconocidas.....	66
2.2.10.3.2 Separación Ciega basada en la independencia.....	68
2.2.10.4 El Análisis de Componentes Principales.....	71
2.2.10.4.1 Introducción.....	71



2.2.10.4.2 Blanqueado	72
2.2.10.4.3 El Análisis de Factores y PCA	77
2.2.10.5 El Análisis de Componentes Independientes.....	78
2.2.10.5.1 Definición.....	78
2.2.10.5.1.1 Definición de ICA basado en un modelo de variables ocultas.	79
2.2.10.5.2 Restricciones en ICA.....	81
2.2.10.5.3 Ambigüedades del análisis ICA	83
2.2.10.5.4 Búsqueda de las componentes independientes.....	84
2.2.10.5.5 Teoría de la Información: una forma alternativa de estimar la transformación ICA	89
2.2.10.5.5.1 Entropía e Información Mutua.	89
2.2.10.5.5.2 La información mutua como medida de la independencia.	90
2.2.10.5.5.3 Entropía de una transformación.....	91
2.2.10.5.5.4 Información mutua y no gaussianidad.....	91
2.2.10.5.5.6 Un ejemplo práctico sobre ICA.....	93
<i>CAPITULO III. MARCO METODOLÓGICO.....</i>	<i>98</i>
3.1. CONSIDERACIONES GENERALES:	99
3.2 DISEÑO Y TIPO DE INVESTIGACIÓN	99
3.3 Técnicas de Recopilación de Datos	101
3.3.1 Análisis Documental:	101
3.3.2 Recolección de Datos:	101
3.4 Técnicas de procesamiento y análisis de datos.....	102
3.4.1 Etapa de adecuación:	102
3.4.2 Etapa de implementación:	102
3.5 Metodología.....	103
<i>CAPÍTULO IV. RESULTADOS.....</i>	<i>108</i>



4.1 Fase I: Estudiar los posibles artefactos que suelen estar presentes en la medición de señales electrofisiológicas.	109
4.2 Fase II: Estudiar y comprender estadísticos de orden superior y medidas de independencia estadística.	110
4.3 Fase III estudiar y comprender el método de Análisis de Componentes Independientes (Independent Components Analysis: “ICA”).....	111
4.4 Fase IV conocer y estudiar las herramientas de software EEGLab y FastICA.	112
<i>CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES.....</i>	<i>198</i>



ÍNDICE DE FIGURAS

Figura 2.2 <i>Ejemplo de señal sinusoidal analógica.</i> [23].....	26
Figura 2.3 <i>Representación gráfica de una señal en tiempo discreto.</i> [23].....	27
Figura 2.4 <i>Ejemplo de una señal sinusoidal en tiempo discreto ($\omega = \pi/6$ y $\theta = \pi/3$).</i> [23].....	28
Figura 2.5 <i>Imagen que muestra los ventrículos, las aurículas, nodos y fibras de Purkinje.</i> [26].....	30
Figura 2.6 <i>Señales de despolarización y repolarización que registra el ECG</i> [26].	30
Figura 2.7 <i>Momento en el que se produce la onda P.</i> [26] 31.....	31
Figura 2.8 <i>Ciclo cardiaco completo.</i> [26]	32
Figura 2.9 <i>Papel de registro del ECG.</i> [27]	33
Figura 2.10 <i>Triangulo de Einthoven.</i> [27]	34
Figura 2.11 <i>Electrodos precordiales.</i> [27]	35
Figura 2.12 <i>Electrodos precordiales.</i> [27]	36
Figura 2.16 <i>Distribuciones Gaussiana y Laplaciana de media 0 y varianza 1.</i> [30].....	65
Figura 2.17 <i>Esquema general que representa los procesos de generación de observaciones y estimación de fuentes en el problema de Separación Ciega de Fuentes.</i> [30]	68
Figura 2.18 <i>Señales originales.</i> [30]	70
Figura 2.19 <i>Señales observadas.</i> [30]	70
Figura 2.20 <i>Señales estimadas.</i> [30]	71
Figura 2.21 <i>Señales originales s_1 (eje horizontal) y s_2 (eje vertical).</i> [30]	75
Figura 2.22 <i>Señales resultantes de la mezcla mediante la matriz A, que denotaremos por x_1 (eje horizontal) y x_2 (eje vertical).</i> [30]	76
Figura 2.23 <i>Señales obtenidas tras blanquear, z_1 (eje horizontal) y z_2 (eje vertical).</i> [30] 77	77
Figura 2.24 <i>Cocktail-Party Effect (el efecto fiesta coctel).</i> [30]	80
Figura 2.25 <i>Comparacion entre distribuciones gaussiana (en azul), laplaciana (en rojo) y uniforme (en verde).</i> [30]	88
Figura 2.26 <i>Densidad de probabilidad conjunta de la distribución (2.96). El eje horizontal representa ‘s_1’ mientras que el vertical representa ‘s_2’.</i> [30]	94
Figura 2.27 <i>Densidad de probabilidad conjunta de la distribución (2.96) tras aplicarle la rotación dada por la matriz A0. El eje horizontal representa ‘x_1’ mientras que el vertical representa ‘x_2’.</i> [30]..	94
Figura 2.29 <i>Rotación de la distribución de la figura 2.28 usando la matriz A. El eje horizontal representa “x_1” mientras que el vertical representa “x_2”.</i> [30]	96
Figura 4.1 <i>Aplicaciones realizadas por EEGLAB.</i> [33].....	112



Figura 4.2 <i>Interfaz gráfica de EEGLAB.</i> [Fuente Propia].....	114
Figura 4.3 <i>Interfaz gráfica de fastICA.</i> [Fuente Propia].....	115
Figura 4.4 <i>Primer paso para el acceso al archivo que contiene la señal a estudiar.</i> [Fuente Propia]...	116
Figura 4.5 <i>Selección del archivo “onda cardiaca sintética”.</i> [Fuente Propia].....	116
Figura 4.6 <i>Forma de onda de la señal ECG sintética, sin ruido alguno.</i> [Fuente Propia].....	117
Figura 4.7 <i>Se selecciona el tipo de arreglo a cargar en este caso un arreglo MATLAB.</i> [Fuente Propia].....	117
Figura 4.8 <i>Colocando el nombre de la señal a estudiar.</i> [Fuente Propia].....	118
Figura 4.9 <i>Descripción de las características de la señal cargada.</i> [Fuente Propia]119	119
Figura 4.10 <i>Ruta para acceder a la opción de graficación de la densidad espectral de potencia de la señal.</i> [Fuente Propia].....	119
Figura 4.11 <i>Opciones para mostrar la grafica de la densidad espectral de potencia de la señal ECG.</i> [Fuente Propia].....	120
Figura 4.12 <i>Densidad espectral de potencia de ECG sintética.</i> [Fuente Propia].....	120
Figura 4.13 <i>Ruta para la obtención de los estudios estadísticos de ECG sintética.</i> [Fuente Propia].....	121
Figura 4.14 <i>Selección del canal a estudiar, en este caso es una sola señal (ECG sintética).</i> [Fuente Propia].....	121
Figura 4.15 <i>Estadísticas de la señal ECG sintética, antes de agregarle ruido.</i> [Fuente Propia].....	122
Figura 4.16 <i>Onda que representa la señal de ruido de línea (60 Hz).</i> [Fuente Propia].....	122
Figura 4.17 <i>Señal ECG con ruido de línea de 60 Hz.</i> [Fuente Propia].....	123
Figura 4.18 <i>Densidad espectral de potencia (PSD) de ECG sintética con ruido.</i> [Fuente Propia].....	124
Figura 4.19 <i>Estadísticas de la señal ECG sintética, después de agregarle ruido.</i> [Fuente Propia].....	124
Figura 4.21 <i>Cargando la matriz “smat” en el fasticag.</i> [Fuente Propia].....	127
Figura 4.22 <i>Matriz “smat” ya cargada.</i> [Fuente Propia]	128
Figura 4.23 <i>Señales sin procesar contenidas en la matriz de observaciones, obtenida de la opción Plot data.</i> [Fuente Propia]	129
Figura 4.24 <i>Reducción de la dimensión de la matriz.</i> [Fuente Propia]	129
Figura 4.25 <i>Señales blanqueadas.</i> [Fuente Propia]	130
Figura 4.26 <i>Componentes independientes obtenidos por fastICA.</i> [Fuente Propia]	131
Figura 4.27 <i>Guardando los resultados en el Work space</i> [Fuente Propia].....	131
Figura 4.28 (a) <i>página principal de physionet donde se muestra la ruta de acceso a la base de datos.</i> (b) <i>ventana donde se seleccionan los datos que se deseen estudiar.</i> [Fuente Propia].....	132



Figura 4.29 (a) Muestra las opciones que se tomaron en este caso para visualizar la onda cuyo registro lleva el nombre de 16539 perteneciente a la base de datos MIT-BIH Normal Sinus Rhythm Database (nsrdb). (b) forma de onda de la señal tratada. [Fuente Propia]..... 134

Figura 4.30 Descarga del archivo con extensión .mat que representa la señal electrocardiográfica, es un arreglo tipo vector, el cual posee las magnitudes leídas continuamente por el electrocardiógrafo al cabo de cierto tiempo. [Fuente Propia]..... 135

Figura 4.31 Forma de onda de la señal ECG 16265 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia]..... 136

Figura 4.32 Densidad espectral de potencia de 16265. [Fuente Propia]..... 136

Figura 4.33 Estadísticas de la señal ECG 16265, antes de agregarle ruido. [Fuente Propia]..... 137

Figura 4.34 Onda que representa la señal de ruido aditivo a la señal ECG 16265. [Fuente Propia]..... 138

Figura 4.35 Onda que representa la señal ECG 16265 contaminada de ruido de línea. [Fuente Propia].. 138

Figura 4.36 Densidad espectral de potencia de 16265 contaminada de ruido de línea. [Fuente Propia].. 139

Figura 4.37 Estadísticas de la señal ECG 16265, después de agregarle ruido. [Fuente Propia]..... 139

Figura 4.38 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia] 139

Figura 4.39 Componentes independientes o fuentes estimadas por fastICA. [Fuente Propia]..... 140

Figura 4.40 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]..... 141

Figura 4.41 Estadísticas de la primera componente. [Fuente Propia]..... 141

Figura 4.42 Estadísticas de la segunda componente. [Fuente Propia]..... 142

Figura 4.43 Estadísticas de la tercera componente. [Fuente Propia]..... 142

Figura 4.44 Forma de onda de la señal ECG 16273 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia]..... 143

Figura 4.46 Estadísticas de la señal ECG 16273, antes de agregarle ruido. [Fuente Propia]..... 144

Figura 4.47 Onda que representa la señal de ruido aditivo a la señal ECG 16273. [Fuente Propia]..... 144

Figura 4.48 Densidad espectral de potencia de 16273 contaminada de ruido de línea. [Fuente Propia]..... 145

Figura 4.49 Estadísticas de la señal ECG 16273, después de agregarle ruido. [Fuente Propia]..... 145

Figura 4.50 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia] 146

Figura 4.51 Componentes independientes o fuentes estimadas por fastICA. [Fuente Propia]..... 146

Figura 4.52 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]..... 147

Figura 4.53 Estadísticas de la primera componente. [Fuente Propia]..... 147

Figura 4.54 Estadísticas de la segunda componente. [Fuente Propia]..... 148

Figura 4.55 Estadísticas de la tercera componente. [Fuente Propia]..... 148



Figura 4.56 Forma de onda de la señal ECG 18184 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia].....	149
Figura 4.57 Densidad espectral de potencia de 18184. [Fuente Propia].....	149
Figura 4.58 estadísticas de la señal ECG 18184, antes de agregarle ruido. [Fuente Propia].....	150
Figura 4.59 Onda que representa la señal ECG 18184 contaminada de ruido de línea. [Fuente Propia]	150
Figura 4.60 Densidad espectral de potencia de 18184 contaminada de ruido de línea. [Fuente Propia].....	151
Figura 4.61 Estadísticas de la señal ECG 18184, después de agregarle ruido. [Fuente Propia].....	151
Figura 4.62 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]...	152
Figura 4.63. Componentes impedientes o fuentes estimadas por fastICA. [Fuente Propia].....	152
Figura 4.64 Densidad espectral de potencia de las componentes independientes. [Fuente Propia].....	153
Figura 4.65 Estadísticas de la primera componente[Fuente Propia].....	153
Figura 4.66 Estadísticas de la segunda componente. [Fuente Propia].....	154
Figura 4.67 Estadísticas de la tercera componente. [Fuente Propia].....	154
Figura 4.68 Forma de onda de la señal ECG 17052 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia].....	155
Figura 4.70 Estadísticas de la señal ECG 17052, antes de agregarle ruido. [Fuente Propia].....	156
Figura 4.71 Onda que representa la señal ECG 17052 contaminada de ruido de línea. [Fuente Propia].....	156
Figura 4.72 Densidad espectral de potencia de 17052 contaminada de ruido de línea. [Fuente Propia]	157
Figura 4.73 Estadísticas de la señal ECG 17052, después de agregarle ruido. [Fuente Propia].....	157
Figura 4.74 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]	158
Figura 4.75 Componentes impedientes o fuentes estimadas por fastICA. [Fuente Propia].....	158
Figura 4.76 Densidad espectral de potencia de las componentes independientes. [Fuente Propia].....	159
Figura 4.77 Estadísticas de la primera componente. [Fuente Propia].....	159
Figura 4.78 Estadísticas de la segunda componente. [Fuente Propia].....	160
Figura 4.79 Estadísticas de la tercera componente. [Fuente Propia].....	160
Figura 4.80 Forma de onda de la señal ECG 16773 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia]	161
Figura 4.81 Densidad espectral de potencia de 16773. [Fuente Propia].....	161
Figura 4.82 estadísticas de la señal ECG 16773, antes de agregarle ruido[Fuente Propia].....	162
Figura 4.83 Onda que representa la señal ECG 16773 contaminada de ruido de línea. [Fuente Propia]...	162
Figura 4.84 Densidad espectral de potencia de 16773 contaminada de ruido de línea. [Fuente propia]...	163
Figura 4.85 Estadísticas de la señal ECG 16773, después de agregarle ruido. [Fuente Propia].....	163



Figura 4.86 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]	164
Figura 4.87 Componentes impedientes o fuentes estimadas por fastICA. [Fuente Propia].....	164
Figura 4.88 Densidad espectral de potencia de las componentes independientes. [Fuente Propia].....	165
Figura 4.89 Estadísticas de la primera componente. [Fuente Propia].....	165
Figura 4.90 Estadísticas de la segunda componente. [Fuente Propia].....	166
Figura 4.91 Estadísticas de la tercera componente. [Fuente Propia].....	166
Figura 4.92 Forma de onda de la señal ECG twa00 de “T-Wave Alternans Challenge Database (twadb)”. [Fuente Propia].....	167
Figura 4.93 Densidad espectral de potencia de twa00. [Fuente Propia].....	167
Figura 4.94 Estadísticas de la señal ECG twa00, antes de agregarle ruido. [Fuente Propia].....	168
Figura 4.95 Onda que representa la señal de ruido aditivo a la señal ECG twa00. [Fuente Propia].....	168
Figura 4.96 Onda que representa la señal ECG twa00, contaminada de ruido aleatorio. [Fuente Propia].....	169
Figura 4.97 Densidad espectral de potencia de twa00, contaminada de ruido aleatorio. [Fuente Propia].....	169
Figura 4.98 Estadísticas de la señal ECG tw00, después de agregarle ruido. [Fuente Propia].....	170
Figura 4.99 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]..	170
Figura 4.100 Componentes impedientes o fuentes estimadas por fastICA. [Fuente Propia].....	171
Figura 4.101 Densidad espectral de potencia de las componentes independientes. [Fuente Propia].....	171
Figura 4.102 Estadísticas de la primera componente. [Fuente Propia].....	172
Figura 4.103 Estadísticas de la segunda componente. [Fuente Propia].....	172
Figura 4.104 Estadísticas de la tercera componente. [Fuente Propia].....	173
Figura 4.105 Forma de onda de la señal ECG twa02 de “T-Wave Alternans Challenge Database (twadb)”. [Fuente Propia].....	173
Figura 4.106 Densidad espectral de potencia de twa02. [Fuente Propia].....	174
Figura 4.107 Estadísticas de la señal ECG twa02, antes de agregarle ruido. [Fuente Propia].....	174
Figura 4.108 Onda que representa la señal ECG twa02 contaminada de ruido aleatorio. [Fuente Propia].....	175
Figura 4.109 Densidad espectral de potencia de twa02, contaminada de ruido aleatorio. [Fuente Propia].....	175
Figura 4.110 Estadísticas de la señal ECG twa02, después de agregarle ruido. [Fuente Propia].....	176
Figura 4.111 Reducción de la matriz antes de obtener los componentes independientes. [Fuente	



Propia].....	176
Figura 4.112 <i>Componentes impeditentes o fuentes estimadas por fastICA.</i> [Fuente Propia].....	177
Figura 4.113 <i>Densidad espectral de potencia de las componentes independientes.</i> [Fuente Propia].....	177
Figura 4.114 <i>Estadísticas de la primera componente.</i> [Fuente Propia].....	178
Figura 4.115 <i>Estadísticas de la segunda componente.</i> [Fuente Propia].....	178
Figura 4.116 <i>Estadísticas de la tercera componente.</i> [Fuente Propia].....	179
Figura 4.117 <i>Forma de onda de la señal ECG twa05 de “T-Wave Alternans Challenge Database (twadb)”.</i> [Fuente Propia].....	179
Figura. 4.118 <i>Densidad espectral de potencia de twa05.</i> [Fuente Propia].....	180
Figura 4.119 <i>Estadísticas de la señal ECG twa05, antes de agregarle ruido.</i> [Fuente Propia].....	180
Figura 4.120 <i>Onda que representa la señal ECG twa05 contaminada de ruido aleatorio.</i>	181
Figura 4.121 <i>Densidad espectral de potencia de twa05 contaminada de aleatorio.</i> [Fuente Propia].....	181
Figura 4.122 <i>Estadísticas de la señal ECG twa05, después de agregarle ruido.</i> [Fuente Propia].....	182
Figura 4.123 <i>Reducción de la matriz antes de obtener los componentes independientes.</i> [Fuente Propia].....	182
Figura 4.124 <i>Componentes impeditentes o fuentes estimadas por fastICA.</i> [Fuente Propia].....	183
Figura 4.125 <i>Densidad espectral de potencia de las componentes independientes.</i> [Fuente Propia].....	183
Figura 4.126 <i>Estadísticas de la primera componente.</i> [Fuente Propia].....	184
Figura 4.127 <i>Estadísticas de la segunda componente.</i> [Fuente Propia].....	184
Figura 4.128 <i>Estadísticas de la tercera componente.</i> [Fuente Propia].....	185
Figura. 4.130 <i>Densidad espectral de potencia de twa08.</i> [Fuente Propia].....	185
Figura 4.131 <i>Estadísticas de la señal ECG twa08, antes de agregarle ruido</i> [Fuente Propia].....	186
Figura 4.132 <i>Onda que representa la señal ECG twa08, contaminada de ruido aleatorio.</i> [Fuente Propia].....	186
Figura 4.133 <i>Densidad espectral de potencia de twa08, contaminada de aleatorio.</i> [Fuente Propia].....	187
Figura 4.134 <i>Estadísticas de la señal ECG twa08, después de agregarle ruido.</i> [Fuente Propia].....	187
Figura 4.135 <i>Reducción de la matriz antes de obtener los componentes independientes.</i> [Fuente Propia].....	188
Figura 4.136 <i>Componentes impeditentes o fuentes estimadas por fastICA.</i> [Fuente Propia].....	188
Figura 4.137 <i>Densidad espectral de potencia de las componentes independientes.</i> [Fuente Propia].....	189



Figura 4.138 <i>Estadísticas de la primera componente.</i> [Fuente Propia].....	189
Figura 4.139 <i>Estadísticas de la segunda componente.</i> [Fuente Propia].....	190
Figura 4.140 <i>Estadísticas de la tercera componente.</i> [Fuente Propia].....	190
Figura 4.141 <i>Forma de onda de la señal ECG twa10 de “T-Wave Alternans Challenge Database (twadb)”.</i> [Fuente Propia].....	191
Figura 4.142 <i>Densidad espectral de potencia de twa10.</i> [Fuente Propia].....	191
Figura 4.143 <i>Estadísticas de la señal ECG twa10, antes de agregarle ruido.</i> [Fuente Propia].....	192
Figura 4.144 <i>Onda que representa la señal ECG twa10, contaminada de ruido aleatorio.</i> [Fuente Propia].....	192
Figura 4.145 <i>Densidad espectral de potencia de twa10, contaminada de aleatorio.</i> [Fuente Propia].....	193
Figura 4.146 <i>Estadísticas de la señal ECG twa10, después de agregarle ruido.</i> [Fuente Propia].....	193
Figura 4.147 <i>Reducción de la matriz antes de obtener los componentes independientes.</i> [Fuente Propia].....	194
Figura 4.148 <i>Componentes impeditivos o fuentes estimadas por fastICA.</i> [Fuente Propia].....	194
Figura 4.149 <i>Densidad espectral de potencia de las componentes independientes.</i> [Fuente Propia].....	195
Figura 4.150 <i>Estadísticas de la primera componente.</i> [Fuente Propia].....	195
Figura4.151 <i>Estadísticas de la segunda componente.</i> [Fuente Propia].....	196
Figura 152 <i>Estadísticas de la tercera componente.</i> [Fuente Propia].....	196



ÍNDICE DE TABLAS

Tabla 1.1 <i>Señales bioeléctricas</i>	5
---	---



Extracción de artefactos en señales electro-fisiológicas utilizando **“Análisis de Componentes Independientes ICA”**

(Realizado por los Brs. Carlos Figueira y Carlos Gomez con el tutor Prof. Cesar Seijas)

RESUMEN

El presente trabajo consistió en implementar una técnica para la extracción de artefactos en señales electrofisiológicas, llamada **“Análisis de Componentes Independientes ICA”** (traducido al inglés se escribe Independent Component Analysis). Para tal fin se tomaron como datos experimentales, registros de señales electrocardiográficas disponibles en la red (www.physionet.org). Se realizó el estudio de todos los fundamentos teóricos concernientes al tema, en los que destacan, el **“Análisis de Componentes Principales PCA”** (traducido al inglés se escribe Principal Component Analysis), **“Separación Ciega Fuentes BSS”** (traducido al inglés se escribe Sources Blind Separation) y estadísticos de orden superior, entre otros. El tratamiento que se le dió a las señales en estudio, para su visualización, adecuación, obtención de características y extracción de artefactos, se realizaron una serie de instrucciones usando el software MATLAB, y bajo esta misma plataforma, se utilizaron los paquetes de software EEGLAB y FastICA.

Palabras claves: ICA, BSS, PCA, EEGLAB, FastICA, ECG, artefactos, señales electrofisiológicas.



INTRODUCCIÓN

Las señales electrofisiológicas son unas de las señales biomédicas más importantes. El hecho que los sistemas biológicos más importantes poseen células excitables hace posible el uso de las señales bioeléctricas para estudiar y monitorear las principales funciones de estos sistemas.

El campo eléctrico se propaga a través del medio biológico, y así el potencial puede adquirirse a distancia desde la superficie del sistema en estudio, eliminándose la necesidad de invadirlo.

Con la llegada del computador, se hizo posible la digitalización de la misma, entendiéndose la digitalización como el proceso de tomar muestras de una señal eléctrica analógica y guardar dichas muestras en formato digital.

Estas señales biológicas demandan aplicaciones de técnicas de análisis no lineal de series de tiempo, dinámica no lineal, teoría de la información y ciencias de la computación, con el fin de generar herramientas que aporten métodos más objetivos de diagnóstico.

Dentro de los problemas en el estudio de estas señales, está la identificación de artefactos, los cuales son contaminaciones de origen externo o interno según sea el caso, que impiden el análisis e interpretación clínica del objeto en estudio.

La presente investigación logró llevar a cabo la implementación del método de separación de fuentes, llamado ICA “análisis de componentes independientes”, y se pudo comprobar su funcionalidad de acuerdo a las propiedades y limitaciones, que condicionan a dicho método. Las señales electrofisiológicas que se utilizaron en este estudio, fueron señales electrocardiográficas **ECG**.

Esta investigación presenta cinco capítulos. El capítulo I describe el problema, plantea los objetivos de la investigación y el alcance del proyecto. El capítulo II contiene el marco teórico, dando a conocer en primer lugar las investigaciones previas que guardan una relación importante con la investigación, luego define los tópicos que comprenden las bases teóricas necesarias para el completo conocimiento de las funciones y el alcance que brinda cada herramienta. El capítulo III presenta la metodología usada, dando a conocer el tipo y



INTRODUCCIÓN



diseño de la investigación, las técnicas de recolección y análisis de datos, y por último presenta las fases que describen las actividades realizadas para el desarrollo de la investigación. El capítulo IV contiene una explicación detallada de los pasos seguidos para el desarrollo de cada una de las fases y los resultados obtenidos en cada una de estas, mostrando finalmente el manejo de la interfaz.

El capítulo 5 presenta las conclusiones generadas de esta investigación, en donde se expone el aprendizaje y el conocimiento nuevo adquirido con la misma, y las recomendaciones las cuales se realizaron con la finalidad de proponer nuevas ideas que ayuden a investigaciones futuras que se relacionen con el presente trabajo.



CAPÍTULO I. PLANTEAMIENTO DEL PROBLEMA.



Capítulo I

EL PROBLEMA

1.1 Planteamiento del problema:

Debido al avance de la tecnología digital, en varios países se han incrementado las aplicaciones del procesamiento digital de señales de origen biológico. El motivo del estudio de las señales electrofisiológicas es que para la obtención de la mayoría de las mismas no es necesario aplicar métodos invasivos al paciente, y estas revelan información importante al ser interpretadas por el médico o experto tratante, a fines de obtener el diagnóstico correspondiente. Otro motivo es la fácil adquisición de dichas señales; la mayor parte de las veces solo precisamos el uso de electrodos que adquieran las señales de la piel u otro lugar del cuerpo humano (según sea lo que se desee medir) y cables que la transmitan a la unidad de procesamiento de la misma. Una ventaja agregada con la llegada del computador es la digitalización de la misma, entendiéndose la digitalización como el proceso de tomar muestras de una señal eléctrica analógica y guardar dichas muestras en formato digital. Luego dicha señal se almacena como datos a efectos de ser estudiada, almacenada, procesada o enviada a algún lugar de interés según sea la necesidad requerida. Existen diferentes fuentes de señales electrofisiológicas que son de sumo interés para el estudio médico, tales como: el corazón, los ojos, el cerebro, los músculos, etc. El más conocido de los estudios es el ECG, en el cual se estudia la actividad cardiaca a través del registro de las señales eléctricas que emite el corazón. Otros son el EEG (electroencefalograma) donde se registra la actividad eléctrica del cerebro; el ERG (electroretinograma) y EOG (electrooculograma) que registran la actividad eléctrica de la retina y el movimiento ocular respectivamente; el EMG (electromiograma) utilizado para la valoración en actividad muscular; EGG asociado a los movimientos peristálticos gastro-intestinales. Los valores que se registran en los diferentes estudios aquí referidos son de valores pequeños, del orden



de los milivoltios (ECG) o en algunos casos del orden de los microvoltios (EEG), según la tabla 1.1 [1]:

Tabla 1.1 *Señales bioeléctricas.*

PARÁMETRO	FRECUENCIA(Hz)	AMPLITUD
EEG	0 – 150	5-300 μ V
ECG	0.01 – 250	0.5-4 mV
EOG	0,1 – 10	1 μ V
EMG	50 – 5000	1mV

[1]

Estas señales biológicas demandan aplicaciones de técnicas de análisis no lineal de series de tiempo, dinámica no lineal, teoría de la información y ciencias de la computación, con el fin de generar herramientas que aporten métodos más objetivos de diagnóstico.

Dentro de los problemas en el estudio de estas señales, está la identificación de artefactos, los cuales son contaminaciones de origen externo o interno según sea el caso, que impiden el análisis e interpretación clínica del objeto en estudio (EEG, ECG, etc.), ocultando información de utilidad en el diagnóstico [2]. En el caso de EEG, también pueden ser confundidos con otros eventos fisiológicos o patrones propios del EEG como ritmo alfa, descargas de punta y onda, delta rítmica frontal intermitente, actividad asimétrica, etc. [3]. Entre los artefactos existentes, se destacan los de tipo fisiológico, que corresponden a potenciales electrofisiológicos de origen extracerebral que nacen en los movimientos involuntarios hechos por el paciente durante el registro del EEG. Estos pueden ser: movimientos oculares, movimientos musculares y el electrocardiograma (EKG) [4]. Estas contaminaciones se adhieren de forma aditiva al EEG. Una gran diversidad de métodos se han planteado en los últimos años para detectar y minimizar artefactos fisiológicos, especialmente los producidos por los movimientos oculares (artefactos oculares) y musculares (artefactos musculares). En la literatura se han reportado métodos basados en redes neuronales [5], análisis en tiempo y frecuencia a través de la transformada Wavelet [6], combinación de los anteriores [7], el uso del análisis de componentes independientes cuyas siglas en inglés son “ICA” (Independent Component Analysis), y la



separación ciega de fuentes cuyas siglas en inglés son “**BSS**” (Blind Source Separation) [8]. En ciertos países, estas nuevas tendencias del estudio de señales de origen biológico (aplicado a EEG y/o ECG), no han sido lo suficientemente difundidas en entidades de salud.

En la presente investigación se plantea utilizar una de las técnicas de separación de fuentes antes mencionadas, conocida como “**Análisis de Componentes Independientes ICA**”, la cual se basa, entre otros conceptos, en la “**Separación Ciega de Fuentes BSS**”, con la finalidad de detectar y extraer en señales electrofisiológicas, artefactos, y de esta manera obtener las señales de interés, separadas de dichos artefactos.

Una vez obtenida la señal depurada, se puede afirmar que los estudios médicos que se le realicen a la misma, tendrán más grado de confiabilidad que en el caso de que la señal en estudio, no se le hayan extraído los artefactos.

1.2 Justificación de la investigación

Los problemas cardiovasculares y neurológicos, demandan la necesidad de desarrollar técnicas no invasivas como la electrocardiografía, y la electroencefalografía respectivamente, con la cual los médicos pueden detectar y diagnosticar lo más acertadamente posible, a sus pacientes. Aunado a esto, es necesario resolver los problemas que se derivan de estas técnicas, que son la presencia de artefactos, que disminuyen la fidelidad de las mediciones obtenidas, alejándolas de los resultados reales, y en consecuencia los diagnósticos son propensos a carecer de precisión, esto a su vez redundaría en tratamientos inadecuados al paciente. Existen métodos de filtrado de artefactos para depurar las señales de los mismos, pero hay que conocer a priori, las frecuencias de estas señales, de modo que no son capaces de filtrar otros artefactos de frecuencia desconocida, en este ámbito ICA es considerablemente útil, ya que se basa en estadísticos de orden superior y otras técnicas matemáticas de separación de fuentes, que la



hacen sumamente útil para la estimación de fuentes originales, y por ende se podrían obtener mayor confianza en los datos obtenidos para el análisis y diagnósticos de pacientes.

1.3 OBJETIVOS DE LA INVESTIGACIÓN

1.3.1 Objetivo General:

Detectar y extraer en señales electro fisiológicas, artefactos usando **“Análisis de Componentes Independientes ICA”**.

1.3.2 Objetivos Específicos:

- Estudiar artefactos en señales electrofisiológicas.
- Estudiar los estadísticos de orden superior y medidas de independencia estadística.
- Estudiar el método de **Análisis de Componentes Independientes (Independent Components Analysis: “ICA”)**,
- Analizar las herramientas de software EEGLAB, y FastICA, necesarias para obtener resultados estadísticos y separación de señales respectivamente.
- Recopilar bases de datos de señales electrofisiológicas, para procesarlas con las herramientas de software EEGLAB, y FastICA.
- Extraer usando **Análisis de Componentes Independientes ICA”**, artefactos y obtener señales depuradas.



1.4 Alcance:

La aplicación de este método (ICA) a las señales usadas, permitirá obtener las señales depuradas (libre de artefactos) para sus posteriores análisis estadísticos, y observación de características. Se usarán como señales experimentales, datos sintéticos generados usando software matemático (“Octave”, software de uso libre) y bases de datos disponibles en la red (“Physionet-MIT”).

1.5 Limitaciones:

Este estudio se basará en análisis por computadora, sobre datos preexistentes, de señales electrocardiográficas (ECG), en el portal web de Physionet-MIT, debido a la complicación que pudiera existir al momento de recolectar datos en centros especializados, y la cantidad de pacientes disponibles para la experimentación.



CAPÍTULO II. MARCO TEÓRICO.



Capítulo II

MARCO TEORICO

2.1 ANTECEDENTES DE LA INVESTIGACIÓN.

Como investigaciones resaltantes en el área, se pueden destacar:

Cesar Seijas 1*, Sergio Villazana 1*, Jorge Guevara 2*, Edilberto Guevara 2* Escuela de Ingeniería Eléctrica, Facultad de Ingeniería, Valencia, Venezuela, ² Escuela de ingeniería Civil, Facultad de Ingeniería, Valencia, Venezuela. PREVISIÓN DE FLUJO DEL RÍO CARONÍ DE VENEZUELA, A TRAVÉS DE MÁQUINAS DE VECTORES DE SOPORTE (SVM).

En este artículo se explora la aplicación de procesos estadísticos y computacionales de elevado desempeño, tales como, Análisis de Componentes Independientes (ICA) como proceso estadístico y algoritmos del cómputo emergente, específicamente, Máquinas de Vectores de Soporte (SVM) como proceso computacional, en el problema de la modelación de series de tiempo. En la esta investigación se presenta un modelo de la serie de tiempo hidrológica del caudal del Río Caroní, conformado por los registros diarios que abarcan un lapso desde 1952 hasta el año 2003. Este modelo está basado en una estructura autorregresional no lineal, implementada con una SVM; de las componentes independientes obtenidas de la aplicación de ICA, a la serie temporal antes mencionada. Las SVM son algoritmos del cómputo emergente desarrollado por Vapnik y sus colaboradores, que han demostrado un excelente desempeño en aplicaciones de regresión de series de tiempo no lineales de señales fisiológicas e hidrológicas por su parte, ICA ha sido utilizado en el problema de predicción de series de tiempo financieras basado en regresión usando redes neuronales artificiales .



Elías Buitrago Bolívar. Estudiante décimo semestre ingeniería biomédica. Universidad Antonio Nariño, Rafael m. Gutiérrez físico, m.sc., ph.d. Director grupo de sistemas complejos, centro de investigaciones universidad Antonio Nariño, Javier E. González Ingeniero electrónico M.sc docente facultad ingeniería electrónica y biomédica universidad Antonio Nariño, Carlos Medina Malo. Neurólogo y neurofisiólogo director fundación liga central contra la epilepsia, Laura victoria guio médico, fundación liga central contra la epilepsia. Residente en neuropediatría, universidad nacional de Colombia. APLICACIÓN CLÍNICA DE SEPARACIÓN CIEGA DE FUENTES Y DIMENSIÓN FRACTAL EN LA IDENTIFICACIÓN DE ARTEFACTOS OCULARES DE EEG.

El Centro de Investigaciones y la Facultad de Ingeniería Electrónica y Biomédica de la Universidad Antonio Nariño (UAN) están desarrollando un proyecto de investigación para introducir los métodos de Separación Ciega de Fuentes y Dimensión Fractal, dirigidos a la minimización de los artefactos fisiológicos en EEGs de instituciones especializadas en neurociencias como la Liga Central Contra la Epilepsia, de Bogotá-Colombia (LICCE).

Damián A. Álvarez Ingeniero electricista. Profesor catedrático Universidad tecnológica de Pereira, Eduardo Giraldo M.sc. ingeniero electricista. Profesor auxiliar Universidad tecnológica de Pereira. ICA APLICADO A LA EXTRACCIÓN DE CARACTERÍSTICAS EN IMÁGENES

En este documento se exponen los fundamentos matemáticos del análisis de componentes independientes (ICA), se muestra su aplicación como técnica de caracterización de imágenes y se valida en la tarea de reconocimiento de rostros sobre las bases de datos CMU y ORL. Se concluyó que la transformación ICA es una técnica apropiada para la extracción de característica en imágenes ya que ofrece una representación robusta en donde las características están lo más decorrelacionadas posible. Permitió el reconocimiento de rostros con porcentajes de acierto promedio del $100\% \pm 0.11$ y $88.56\% \pm 0.24$ para las base de datos CMU y ORL respectivamente.



2.2 BASES TEÓRICAS.

2.2.1 Señales electrofisiológicas

2.2.1.1 Electrooculograma (EOG).

Registra los movimientos oculares mediante electrodos, situados en los ángulos externos de los ojos sobre la piel. Se detectan cambios de voltaje cuando el ojo rota, ya que la polaridad de la retina es negativa respecto al de la cornea. Los electrodos que se usan, al igual que en el registro de EEG, son de oro o de cloruro de plata.

Durante el sueño pueden darse movimientos oculares lentos, oscilatorios como sucede en fases de sueño y movimientos oculares rápidos, los cuales aparecen esporádicamente en salvas y son característicos del sueño REM, aunque también se observan en la fase de adormecimiento. [21]

2.2.1.2 Electromiograma (EMG).

Mide el grado de tono muscular que presentan los músculos esqueléticos. El EMG submentoniano registra los cambios de tensión muscular que se dan en las distintas fases del dormir. Debe hacerse con al menos tres electrodos colocados en la región mentoniana y submentoniana. [21]

2.2.1.3 Electroencefalograma (EEG).

Se comienza hablar de las señales EEG cuando se encuentran evidencias de actividad eléctrica en el cerebro de los animales, el primero en hacerlo fue el inglés Caton en 1874 cuando observó fluctuaciones de la corteza cerebral en conejos y confirmó que esas fluctuaciones estaban relacionadas con la actividad funcional del cerebro. Años más



tarde, en 1912, Kaufmann confirmó la conclusión a la que llegaron algunos y que consistía en afirmar que el potencial eléctrico de la corteza cerebral era modificado por una estimulación sensorial y estaba relacionado con la función nerviosa. Para 1925, Neminski, usó un galvanómetro para sus mediciones trabajando con perros y reportó la actividad eléctrica en la corteza cerebral en muchos de los términos que hoy se usan todavía.

Encontró que el electroencefalograma (llamado así por el mismo) consistía principalmente de fluctuaciones espontáneas en el potencial con una frecuencia de 10 a 15 ciclos por segundo, llamadas ondas de primer orden, y otras fluctuaciones más rápidas con frecuencia de 20 a 32 ciclos por segundo, llamadas ondas de segundo orden.

El psiquiatra suizo Hans Berger (1929) realizó un intento satisfactorio del registro de la actividad eléctrica del cerebro humano. Demostró que el cerebro de un hombre tiene un *golpeteo eléctrico* el cual proviene de las neuronas, no de los vasos sanguíneos o del tejido conjuntivo y que cambia con la edad, con la estimulación sensorial y con varios cambios en el estado fisicoquímico del cuerpo.

Berger llamó *electroencefalograma (EEG)* al registro que logró tomar y es el que se usa hoy en día. Este *EEG* permitió registrar la actividad eléctrica cerebral sobre la superficie del cuero cabelludo y sus diversas variaciones durante este periodo.

Estas variaciones fueron sistematizadas por Loomis, Harvey y Hobart (1937 y 1938). Observando estos autores que la actividad eléctrica cerebral muestra cambios típicos en el transcurso del sueño y que las ondas se hacen más lentas y aumentan de amplitud a medida que el sueño se profundiza. Así establecieron la primera clasificación de las fases del sueño que distinguía cinco patrones distintos de actividad eléctrica cerebral relacionados con diferentes niveles o estados de sueño. Estadio A, Estadio B, Estadio C, Estadio D, Estadio E. Mediante el *EEG* se registran los cambios de voltaje entre dos electrodos colocados sobre el cuero cabelludo.



Estos potenciales se amplifican y pueden enviarse a un osciloscopio para su observación directa o a papel continuo para su análisis visual o a un ordenador el cual puede analizar la amplitud y el espectro de frecuencia de la actividad EEG y cuantificar automatizadamente las señales EEG.

Las señales EEG reflejan en gran parte los cambios de potencial de las membranas de las células nerviosas del encéfalo, en particular de las neuronas corticales. Así se registra la actividad cerebral espontánea que se da durante la vigilia al cual puede clasificarse en cuatro bandas de frecuencia que corresponden con los distintos ritmos cerebrales. Estos se definen por su frecuencia, medidas en hercios (Hz), y por su amplitud, medida en microvoltios (μV).

Los ritmos de alta frecuencia tienen por lo general una amplitud menor que los de baja frecuencia, aunque la amplitud de las señales EEG recogidas sobre el cuero cabelludo están distorsionadas por efecto de los tejidos que rodean el encéfalo. [22]

2.2.1.4 Polisomnografía

El término *polisomnografía* fue propuesto por Holland, Dement y Raynal en 1974, para describir el registro simultáneo de varios parámetros fisiológicos, así como el análisis e interpretación de los mismos. Como técnica, la *polisomnografía* es esencial para la formulación de diagnóstico de los trastornos del dormir. Es un procedimiento complejo que debe ser ejecutado por personal capacitado.

Analizando los datos de la *Polisomnografía* se obtienen una serie de medidas que en conjunto informan de la continuidad y eficiencia del sueño, su organización así como también la composición porcentual de fases con lo que se puede establecer si el patrón de sueño es normal o patológico.



El *EEG*, el *EOG* y el *EMG* se recogen simultáneamente de modo que se pueda observar la relación entre los tres parámetros y asignar a cada unidad de tiempo en que se divide el registro de sueño (época de registro, que conforme a los criterios de cuantificación habituales dura 30 segundos) la fase o estadio correspondiente según las características electroencefalográficas, electrooculográficas y electromiográficas predominantes (mayor al 50%) en dicha etapa.

Estas son las variables que se registran habitualmente en una *Polisomnografía* del sueño humano. Además pueden registrarse el *electrocardiograma* (ECG), la respiración nasal y toraxica, los movimientos de las piernas, y diversos tipos de actividad fisiológica como registros del aparato digestivo. [21]

2.2.1.5 Electrocardiograma (ECG)

Estas señales miden el ritmo cardiaco del corazón. Hay diferentes procedimientos para registrar la actividad ECG durante el registro polisomnográfico. La técnica más simple incluye la colocación de un electrodo sobre la superficie de la piel por debajo de la clavícula derecha.

2.2.1.6 Bases de datos de señales de electrocardiográficas.

Entre las bases de datos de señales electrocardiográficas disponibles en PhysioBank (enlace al cual se accede a través del portal web de [physionet](http://physionet.org)), podemos citar las que siguen:

ANSI/AAMI EC13 Test Waveforms. Base de datos cuyos registros se utilizan para probar y calibrar los diferentes dispositivos utilizados en la monitorización de las señales electrocardiográficas. Está compuesta por 10 registros de corta duración especificados por la American National Standard. Los registros son de dominio público e incluyen tanto



señales sintéticas como reales. Cada uno de ellos contiene una señal electrocardiográfica obtenida mediante muestreo a 720 Hz con una resolución de 12 bits.

European ST-T Database. Utilizada para evaluar todos aquellos algoritmos desarrollados para el análisis de los cambios ocurridos durante el segmento ST y la onda T de una señal electrocardiográfica. Está formada por 90 fragmentos representativos, perfectamente etiquetados (por dos cardiólogos diferentes) y seleccionados a partir de las señales extraídas de los electrocardiogramas registrados sobre 79 pacientes (70 hombres, 8 mujeres y un registro no identificado). En cada uno de los sujetos se ha diagnosticado un posible problema de isquemia de miocardio, además de utilizar otros criterios de selección sobre los fragmentos para obtener una colección de disfunciones cardíacas lo suficientemente representativa (como, por ejemplo, el desplazamiento de la línea de base del segmento ST). La base de datos proporciona un total de 367 episodios de cambio de la línea de base del segmento ST, y 401 episodios de cambio de la onda T, con duraciones que oscilan entre los 30 segundos a varios minutos y con oscilaciones en la amplitud de la señal que van desde los 100 μV hasta más de 1mV. Adicionalmente también se han etiquetado 11 episodios de desplazamiento axial que resultan en un cambio aparente del segmento ST y otros 10 episodios de desplazamiento axial que suponen cambios en la onda T. Cada uno de los registros presentados contiene dos señales de 2 horas de duración, muestreadas a 250 Hz con una resolución de 12 bits y un rango de entrada nominal de 20 mV.

Long-Term ST Database. Se trata de una base de datos concebida para recoger toda la variabilidad que presenta el segmento ST en los electrocardiogramas reales, incluyendo episodios de isquemia, episodios no isquémicos y relacionados con el movimiento axial, episodios de lentitud en el desplazamiento del segmento ST y cualquier variante de la mezcla de estos tres. Está formada por 86 registros de larga duración (de 21 a 24 horas, razón por la cual aparecen fragmentados) conteniendo 2 ó 3 señales electrocardiográficas que han sido etiquetadas latido a latido. Los registros han sido extraídos a partir de 80



pacientes. Precisamente debido a la larga duración de los registros únicamente son accesibles desde Physionet 43 de ellos.

MIT-BIH Arrhythmia Database. En esta base de datos se recogen gran cantidad de episodios de arritmia cardíaca en 48 extractos de media hora de duración correspondientes a los ECG ambulatorios de 47 pacientes, 23 registros seleccionados de forma aleatoria a partir de 4000 ECG ambulatorios de 24 horas de duración, y otros 25 registros elegidos para recoger también aquellas arritmias menos frecuentes pero clínicamente significativas. Los registros (correspondientemente etiquetados) han sido muestreados a 360 Hz con una resolución de 11 bits y con un rango de tensión de 10 mV. La mitad de los registros son accesibles a través de Physionet.

MIT-BIH Noise Stress Test Database. Con 12 registros ECG y 3 registros que contienen el ruido típico que aparece en los ECG ambulatorios [34]. Todos ellos tienen una duración de media hora. Los 3 registros ruidosos han sido obtenidos ensamblando aquellos intervalos de registros que contenían tanto variaciones en la línea de base, como interferencias de carácter muscular o debidas al movimiento de los electrodos.

BIDMC CongestiveHeart Failure Database. Incluye ECGs de larga duración correspondientes a 15 pacientes que presentan fallos cardíacos congestivos de tipo severo. Los registros tienen una duración de 20 horas, conteniendo 2 señales muestreadas a 250 Hz con una resolución de 12 bits y un rango de ± 10 mV.

QT Database. La base de datos QT fue diseñada a partir de la base de datos del MIT con el objetivo de evaluar algoritmos de detección de ondas en el ECG. Esta base de datos consiste en 105 fragmentos de 15 minutos de duración, provenientes de registros Holter de dos canales, seleccionados por varios expertos con el objetivo de incluir una amplia variedad de complejos P, QRS, T y U que aparecen adecuadamente etiquetados e identificados (de 30 a 50 latidos etiquetados de este modo en cada registro). De esta forma,



las técnicas que se apliquen sobre estas señales estarán trabajando dentro de la variabilidad del mundo real. Los límites de cada onda dentro de estos registros han sido determinados manualmente por expertos

Creighton University Ventricular Tachyarrhythmia Database. Incluye 35 registros ECG de aproximadamente 8 minutos de duración, correspondientes a pacientes que han experimentado episodios de taquicardia ventricular, flutter (aleteo) ventricular y fibrilación ventricular. Todos los registros han sido muestreados a 250 Hz con 12 bits de resolución y un rango dinámico de 10 V (10 mV en las señales antes de ser amplificadas). Todos los registros son de acceso libre a través de Physionet.

Intracardiac Atrial Fibrillation Database. Esta base de datos está formada por los registros endocardiales provenientes de la aurícula derecha de 8 pacientes en episodios de fibrilación o palpitación (flutter o aleteo) auricular. Los registros han sido obtenidos a través de un catéter multipolar situado en cuatro regiones separadas del corazón. La frecuencia de muestreo es de 1 KHz. Todas las señales de la base de datos son accesibles a través de la web.

MIT-BIH Atrial Fibrillation Database. Incluye 25 registros ECG de larga duración correspondientes a pacientes que presentan episodios de fibrilación auricular. Los registros de la base de datos tienen una duración aproximada de 10 horas y presentan 2 señales electrocardiográficas muestreadas a 250 Hz con una resolución de 12 bits y un rango de ± 10 mV. Todas las señales están etiquetadas.

MIT-BIH ECG Compression Test Database. Que contiene 168 registros ECG de corta duración (20,48 s cada uno de ellos) seleccionados para proporcionar un conjunto de señales test (prueba) que abarca toda una amplia gama de posibilidades que aparecen en los casos reales y que deben ser resueltas adecuadamente por los diferentes métodos de



compresión de señales electrocardiográficas, en particular aquellas técnicas de compresión con pérdidas.

MIT-BIH Malignant Ventricular Arrhythmia Database. Con 22 registros de media hora de duración pertenecientes a pacientes que han presentado episodios de taquicardia ventricular sostenida, flutter ventricular y fibrilación ventricular.

MIT-BIH Normal Sinus Rhythm Database. Proporciona 18 registros ECG de larga duración perteneciente a 18 pacientes (5 hombres y 13 mujeres) que no presentan ningún tipo de arritmia significativa.

MIT-BIH ST Change Database. En la que se incluyen 28 registros ECG de diferentes duraciones, la mayoría de los cuales han sido obtenidos mientras los pacientes realizaban pruebas de esfuerzo para poder reflejar claramente las variaciones del segmento ST durante los cambios del ritmo cardíaco.

MIT-BIH Supraventricular Arrhythmia Database. En la que se proporcionan 78 registros de media hora de duración que han sido especialmente seleccionados para complementar los ejemplos de arritmia supraventricular proporcionados en la MIT-BIH Arrhythmia Database.

Paroxysmal Atrial Fibrillation Prediction Challenge Database. Esta base de datos fue creada con el objetivo de utilizar sus registros para desarrollar métodos automatizados para la detección de la fibrilación auricular de tipo paroximal. Contiene registros ECG de dos canales muestreados a 128 Hz con una resolución de 16 bits por muestra.



2.2.2 Procesamiento de señales.

2.2.2.1 Introducción.

Las señales generalmente transportan información acerca del estado o del comportamiento de sistemas físicos, y a menudo, estas señales son sintetizadas con el propósito de comunicar información entre personas o entre personas y máquina. Aunque las señales pueden ser representadas en diferentes formas, en todos los casos la información está contenida en un cierto patrón de variaciones. Estas señales son representadas matemáticamente como funciones de una o más variables independientes.

El procesamiento de señales se refiere a la representación, transformación y manipulación de señales y la información que en ella se tiene. Por ejemplo, podríamos separar dos o más señales que se encuentran combinadas o cambiar los componentes o parámetros de alguna de ellas. En el caso de los sistemas de comunicación, generalmente es necesario hacer un procesado previo como modulación, compresión para la transmisión y un procesado para recibirla. [23]

En los años 60, la tecnología del procesamiento de señales fue casi exclusivamente analógica en tiempo continuo, pero la rápida evolución de las computadoras digitales junto con los microprocesadores a la par del desarrollo de importantes avances teóricos tales como el algoritmo de la *Transformada Rápida de Fourier* (TRF) trajo como consecuencia un cambio significativo en la tecnología digital dando como resultado un gran avance en el campo del procesamiento digital de señales.

En el procesamiento digital de señales, estas son representadas como secuencias de números, y su procesamiento se realiza con ayuda de la computación digital.



El valor de una señal, en tiempo continuo o discreto, puede ser continuo o discreto. Si una señal toma todos los valores posibles en un intervalo tanto finito como infinito, se dice que es continua. Por el contrario, si toma valores de un conjunto finito de valores se dice que es discreta. [23]

Normalmente, estos valores son equidistantes en tiempo y por tanto pueden expresarse como un múltiplo de la distancia entre dos valores sucesivos. Una señal en tiempo discreto, que toma valores en un conjunto discreto se denomina *señal digital*.

En la figura 2.1 se muestra una señal digital con tres valores posibles.

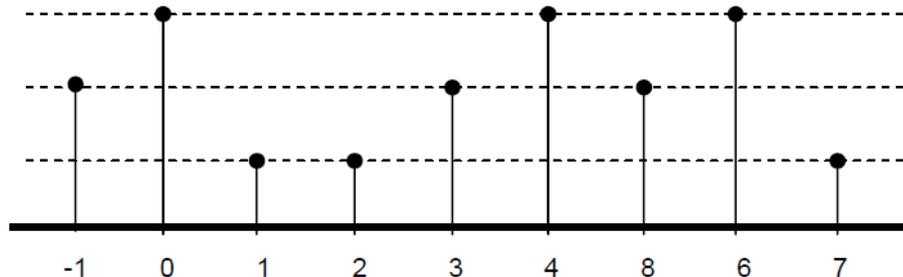


Figura 2.1 *Señal digital con tres valores de amplitud.* [23]

El término “*procesamiento de señales en tiempo discreto*” incluye al procesamiento digital de señales como un caso especial pero no excluye la posibilidad de que estas secuencias sean procesadas con otros métodos tecnológicos para el tiempo discreto.

El procesamiento en tiempo discreto de señales en tiempo real es común en sistemas de comunicación, radares, codificación de voz y video, ingeniería biomédica, por mencionar algunas.



El procesado del reproductor de discos compactos es un ejemplo un poco diferente en cuanto a la forma en cómo se adquiere la información ya que esta se encuentra almacenada en el disco compacto y el procesado final se realiza en tiempo real cuando se requiere de una salida (por ejemplo sonido).

El procesado tradicional de señales, involucra el procesado de una señal para obtener otra señal. Un problema del procesado de señales es su interpretación. En muchos de los casos el objetivo del procesado no es obtener una señal de salida o entenderla, más bien interpretar la entrada o simplemente obtener información de esta. Típicamente, en tales sistemas se aplica un procesado digital (filtrado, estimación de parámetros, etc.) seguido de un sistema de reconocimiento de patrones para producir una representación simbólica.

Esta salida simbólica puede alternadamente ser la entrada de un sistema de procesado simbólico para obtener una interpretación final de la señal.

Para que una señal pueda ser procesada digitalmente ha de ser en tiempo discreto y tomar valores discretos (es decir, ha de ser una señal digital). Si la señal a procesar es analógica, se convierte a digital muestreándola en el tiempo y obteniendo por tanto una señal en tiempo discreto y posteriormente, cuantificando sus valores en un conjunto discreto. El proceso de convertir una señal continua en discreta, denominado cuantificación, es básicamente un proceso de aproximación y se puede lograr por redondeo o truncamiento.

Existen algunas diferencias fundamentales en cuanto a las teorías para el procesamiento de señales en una y varias dimensiones. Muchas aplicaciones de procesamiento de imágenes requieren trasladarse a técnicas de procesamiento en dos dimensiones. Este es el caso para las áreas de codificación de video, imágenes médicas, análisis de fotografías arteriales, análisis satelital, fotografías climáticas y algunas transmisiones de video desde el espacio y la Luna.



El análisis espectral está basado por un lado en el uso de la Transformada de Fourier y por otro en el uso de modelado de señales, este último es particularmente rico e importante en el proceso de señales ya que permite realizar la transformación a la que se ha hecho mención, captar una señal cuya amplitud se define en función del tiempo para transformarla en otra cuya potencia es definida en función de la frecuencia. [24]

El modelado de señales, también juega un importante rol dentro de la compresión y codificación, y las ecuaciones diferenciales fundamentales proveen las bases para el entendimiento de muchas de estas técnicas. Otro asunto de consideración es la adaptación del procesado de señales.

2.2.2.2 Señales en tiempo continuo y discreto.

El concepto de frecuencia es familiar para ingenieros y científicos; es básico en, por ejemplo, el diseño de un receptor de radio, un sistema de alta fidelidad, o un filtro espectral para fotografía en color. Debido a la física sabemos que la frecuencia está íntimamente relacionada con un tipo específico de movimiento periódico llamado oscilación armónica, que se describe mediante funciones sinusoidales. Este concepto de frecuencia está directamente relacionado con el de tiempo. De hecho, sus dimensiones son las inversas del tiempo. Por lo tanto, de acuerdo con esto, se debería esperar que el tiempo (continuo o discreto) afecte a la frecuencia. [25]

Dentro de la ciencia, tecnología, negocios y otros campos estamos constantemente manejando cantidades. Las cantidades son medidas, monitoreadas, grabadas, manipuladas aritméticamente, observadas, o de alguna otra manera utilizadas en la mayoría de los sistemas físicos. Es importante, cuando tratamos con diferentes cantidades, que seamos capaces de representar sus valores con precisión y de manera eficiente.



Existen dos formas de representar el valor numérico de las cantidades: La forma analógica y la forma digital. [25]

Representación Analógica.

En la representación analógica una cantidad se representa por un voltaje, corriente, o el movimiento de un metro que es proporcional al valor de la cantidad. Las cantidades analógicas poseen una característica importante: varían sobre un rango de valores continuos.

Representación Digital.

En la representación digital las cantidades no se representan por cantidades proporcionales sino por símbolos llamados dígitos. Como ejemplo podemos mencionar el reloj digital, el cual representa las horas, los minutos y los segundos en forma de dígitos decimales. Como sabemos el tiempo cambia continuamente, mientras el reloj lo hace digital, por pasos de uno por segundo (o por minuto). En otras palabras, esta representación digital del transcurrir del tiempo cambia en pasos discretos.

La diferencia principal entre las cantidades representadas de forma analógica y las representadas en forma digital es que las primeras son continuas y las segundas son discretas (paso a paso). [25]

2.2.2.3 Señales en tiempo continuo.

La señal continua es una función de una o varias dimensiones. Ejemplos de distintos tipos de señales podemos encontrar en los muy diversos aparatos de medida asociados al estudio de la física, química, biología, medicina, etc. Así por ejemplo, los distintos tipos de electrogramas que son usados en medicina son señales unidimensionales, ya que se



representan por una o varias curvas en función del tiempo. Sin embargo, los distintos tipos de radiografías son señales bidimensionales y los resultados de la tomografía axial computarizada y la resonancia nuclear magnética son señales tridimensionales.

Haciendo uso del lenguaje matemático podemos decir que toda señal es una función matemática que toma un valor en cada punto del espacio en el que está definida. Los resultados matemáticos sobre la aproximación de funciones permiten expresar que cualquier función continua y periódica definida sobre una región finita del espacio puede ser aproximada por una suma infinita de términos, en donde cada término tiene una contribución a la formación de la señal que es independiente y ortogonal a cualquier otro término del desarrollo.

Existen distintas posibilidades a la hora de construir este tipo de aproximaciones, pero por distintos motivos la más usada ha sido aquella en que los términos tienen una significación como señales ondulatorias puras. Es decir, cada término es la contribución de una determinada frecuencia/longitud de onda a la formación de la señal. Así pues toda señal puede ser analizada desde dos puntos de vista, como una función continua sobre un espacio de valores de medida o como una función definida sobre un espacio de frecuencias.

Una simple oscilación armónica se describe matemáticamente mediante la siguiente señal en tiempo continuo:

$$X_a(t) = A \cos(\Omega t + \theta), \quad -\infty < t < \infty \quad (2.1)$$

que se muestra en la figura 2.2. El subíndice a utilizado con $X(t)$ denota una señal analógica. Esta señal está completamente caracterizada por tres parámetros: A es la amplitud de la senoide, Ω es la frecuencia en radianes por segundo (rad/seg), y θ es la fase en radianes. En lugar de Ω a menudo se utiliza la frecuencia F ciclos por segundo o Hertzios (HZ), donde,



$$\Omega = 2\pi F \quad (2.2)$$

La ecuación (2.1) puede escribirse en términos de F como:

$$X_a(t) = A \cos(2\pi Ft + \theta), -\infty < t < \infty \quad (2.3)$$

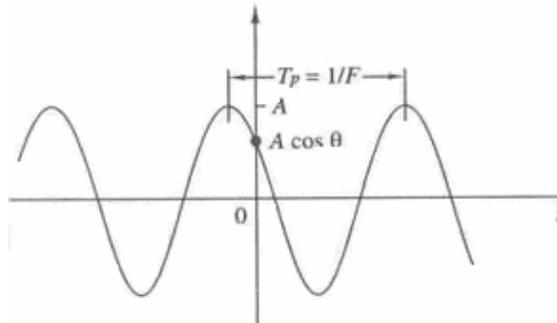


Figura 2.2 *Ejemplo de señal sinusoidal analógica.* [23]

2.2.2.4 Señales en tiempo discreto

Una señal en tiempo discreto $x(n)$ es una función de una variable independiente entera. Gráficamente, se representa como en la figura 2.3. Es importante destacar que una señal en tiempo discreto no está definida para instantes entre dos muestras sucesivas. Igualmente, es incorrecto pensar que $x(n)$ es igual a cero si n no es un entero. Simplemente, la señal $x(n)$ no está definida para valores no enteros de n .

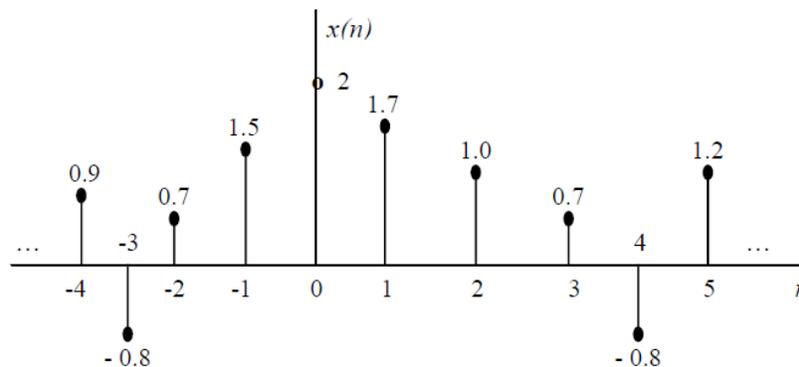


Figura 2.3 *Representación grafica de una señal en tiempo discreto.* [23]

Comúnmente, se refiere a $x(n)$ como la “n-esima muestra” de la señal aun cuando $x(n)$ sea inherentemente en tiempo discreto (es decir, aunque no haya sido obtenida por muestreo de una señal analógica). En el caso en que $x(n)$ haya sido obtenida al muestrear una señal analógica $x_a(t)$ entonces $x(n) = x_a(nT)$, donde T es el periodo de muestreo (tiempo entre muestras sucesivas).

Una señal sinusoidal en tiempo discreto puede expresarse como:

$$x(n) = A \cos(\omega n + \theta), -\infty < n < \infty \quad (2.4)$$

donde, n es una variable entera, denominada numero de muestra, A es la *amplitud* de la senoide, ω es la frecuencia en radianes por muestra, y θ es la *fase* en radianes.

Si, en lugar de ω utilizamos la variable de frecuencia f definida por

$$\omega \equiv 2\pi f \quad (2.5)$$

la relación (2.5) se convierte en:



$$x(n) = A \cos(2\pi f n + \theta), \quad -\infty < n < \infty \quad (2.6)$$

La frecuencia f tiene dimensiones de ciclos por muestra. La figura 2.4 muestra una senoide de frecuencia $\omega = \pi/6$ radianes por muestras ($f = 1/12$ ciclos por muestra) y fase $\theta = \pi/3$.

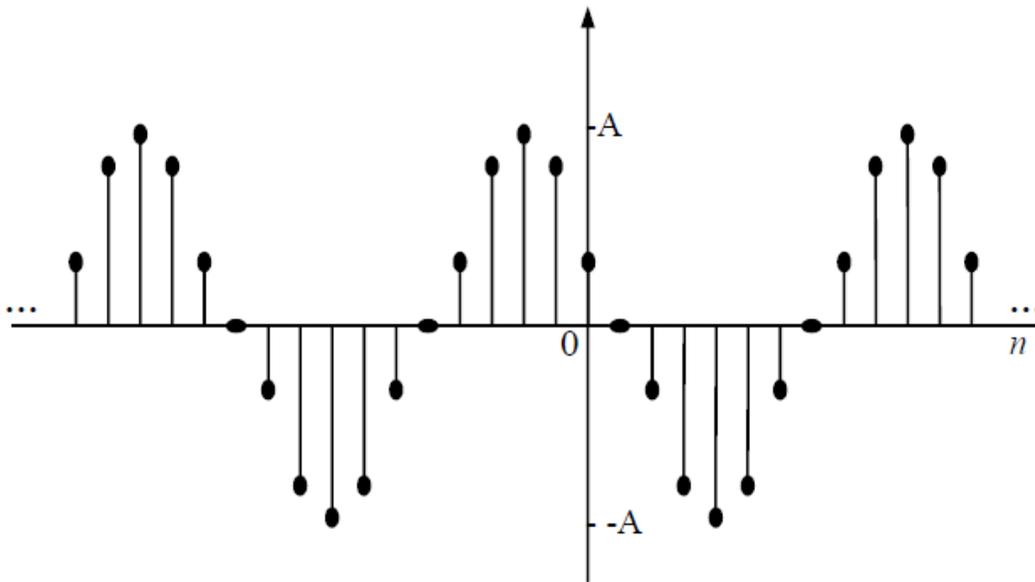


Figura 2.4 *Ejemplo de una señal sinusoidal en tiempo discreto ($\omega = \pi/6$ y $\theta = \pi/3$).* [23]

2.2.2.5 Representación de las señales Digitales.

Las señales que se envían a través del telégrafo son señales digitales; estas son escritas utilizando 26 caracteres del alfabeto, 10 números y símbolos tales como puntos y comas. Estos símbolos pueden ser representados por diferentes niveles de voltaje, por ejemplo, A por 10 V (volts), B por 9.5 V, C por 9 V y así sucesivamente. Este tipo de representación es buena, sin embargo, no es utilizada debido a que es muy susceptible al ruido, cambios de voltaje en la fuente de alimentación y muchos otros disturbios. En la



práctica, esos símbolos son codificados como secuencias de ceros y unos. El símbolo 0 puede ser representado por el voltaje que va desde 0 hasta 0.8 V. El símbolo 1 puede ser representado por un voltaje que va desde 2.0 hasta 5.0 V. [25]

Este tipo de representación de ceros y unos es mucho menos susceptible al ruido además de que es muy fácil su implementación. Por lo tanto, las señales digitales son siempre codificadas como ceros y unos en representaciones físicas. Esto es llamando código binario.

2.2.3. Señales Electrocardiográficas (ECG)

Existen señales que registran la actividad eléctrica cardiaca con lo cual obtenemos información muy útil sobre el funcionamiento del corazón. Este conjunto de señales es mejor conocido como electrocardiograma (ECG). Se muestran como una línea delgada que presenta distintas inflexiones, que corresponden a parámetros de información del estímulo eléctrico del corazón durante las fases de reposo y recuperación. [26]

Existen diferentes tipos de células cardiacas importantes para el campo de la electrocardiografía: [26]

- _Células de actividad automática, eléctricas.
- _Células de actividad contráctil, o de trabajo.
- _Células del tejido conectivo.
- _Células de los vasos sanguíneos.

Las células de actividad automática permiten generar un potencial de tipo diastólico de reposo, al activarse se produce la contracción de células de actividad contráctil (Fase Sistólica). Para ello, será imprescindible el tejido conectivo. Las células de actividad automática tienen mayor despolarización que las de actividad contráctil.

Cada músculo cardiaco se estimula eléctricamente y se contrae. Las células cardiacas en reposo se encuentran cargadas o polarizadas; pero la estimulación eléctrica las “despolariza”, y se contraen. En sentido estricto, una célula polarizada en reposo tiene cargas negativas en su interior y positivas en su exterior. El interior de la célula miocárdica se vuelve positivo cuando la célula recibe un estímulo para contraerse. La figura 2.5 muestra la estructura general del corazón.

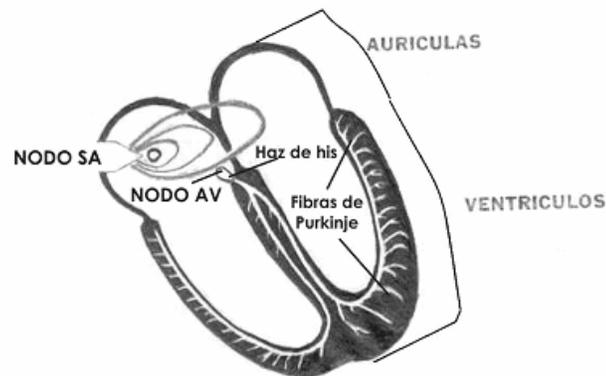


Figura 2.5 *Imagen que muestra los ventrículos, las aurículas, nodos y fibras de Purkinje.* [26]

La despolarización estimula las células del miocardio y hace que se contraigan cuando la carga en cada célula se vuelve positiva. Las ondas de despolarización (el interior de las células se vuelve positivo) y de repolarización (las células recuperan su carga negativa) se registran en el ECG figura 2.6. Este fenómeno es estrictamente eléctrico, y el corazón no presenta ningún movimiento durante esta actividad. Cuando esta actividad eléctrica recorre el corazón, se puede captar desde la piel mediante un equipo de registro sensible como lo son los electrodos externos y se puede registrar como ECG. [26]

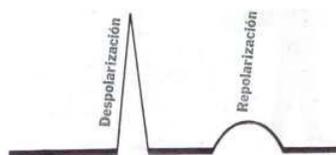


Figura 2.6 *Señales de despolarización y repolarización que registra el ECG* [26].

Cuando la onda positiva de despolarización en las células cardíacas se acerca a un electrodo positivo, el electrocardiograma registra una reflexión positiva, es decir, hacia arriba.

La onda de despolarización nace en el nodo sinoauricular (SA) y se inicia el impulso eléctrico, que se extiende como onda y estimula ambas aurículas con lo cual se produce la llamada *onda P* en el ECG (figura 2.7) dando lugar a la contracción de ambas aurículas.

Este impulso llega al nodo aurículo-ventricular AV, donde ocurre una pausa de 1/10 de segundo, lo que permite que la sangre llegue a los ventrículos. Después de la pausa, el nodo AV es estimulado y se inicia un impulso eléctrico que se dirige hacia abajo por el haz de His y las ramas del mismo.



Figura 2.7 *Momento en el que se produce la onda P.* [26]

2.2.3.1 Tipos de ondas que se registran

Un ciclo cardíaco completo comprende la onda P, el complejo QRS y la onda T y se repite una y otra vez. (Fig. 2.8). La *onda P* representa la contracción de las aurículas; después de esta onda, el impulso llega al nodo AV produciéndose la pausa de 1/10 seg. [26]

El complejo QRS representa el impulso eléctrico que se aleja del nodo AV, y pasa a las fibras de Purkinje y a las células del miocardio. Este complejo en el ECG representa el



inicio de la contracción ventricular. Este complejo se registra con una amplitud $\leq 2,5\text{mV}$ a 5 V y dura de 60 a 100 milisegundos.

La onda *Q* es la primera deflexión hacia abajo del complejo QRS, va seguida de la onda *R* hacia arriba, a menudo esta onda suele no visualizarse en el ECG, cuando existe siempre es la primera en el complejo QRS.

La onda *R* hacia arriba va seguida de una onda *S* hacia abajo. El complejo QRS total representa la actividad eléctrica de la contracción ventricular.

La onda *S* es una deflexión negativa que sigue a la onda *R*, no siempre está presente.

Existe una pausa después del complejo QRS llamado segmento ST que no es sino la parte plana de la línea entre el complejo QRS y la onda *T*.

La onda *T* representa la repolarización de los ventrículos para que se les pueda volver a estimular.

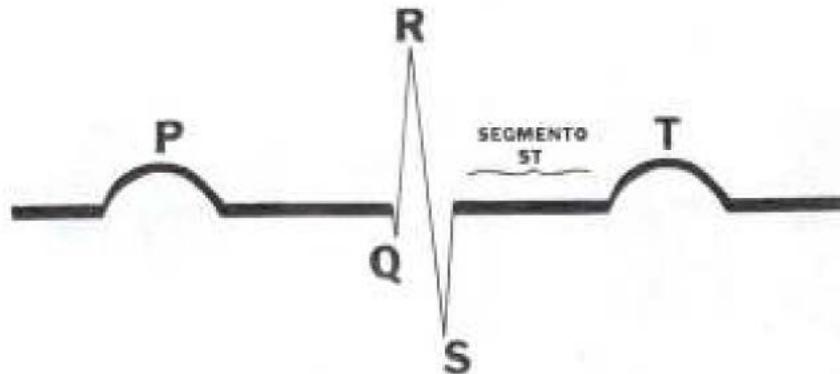


Figura 2.8 *Ciclo cardíaco completo.*[26]



2.2.3.2 Principio de medición.

La señal de salida de nuestro sistema dependerá de la calidad de información de entrada suministrada y a la rapidez con la que los datos sean obtenidos. Por dichos motivos existen ciertos parámetros destacados sobre la adquisición, acondicionamiento y conversión, con el fin de que la información obtenida sea de la mejor calidad posible.

El electrocardiograma se registra generalmente, sobre papel cuadrulado. Las divisiones pequeñas son cuadrados de un milímetro. La altura o profundidad de una onda se miden en milímetros, y representan un voltaje. Las deflexiones hacia arriba se llaman positivas y las deflexiones hacia abajo se llaman negativas. El eje horizontal sobre el cual se traza el ECG representa el tiempo y el eje vertical nos indica el voltaje de la señal así podemos conocer la duración y el voltaje de cualquier parte del ciclo cardiaco (figura 2.9).

El ECG ordinario consta de 12 derivaciones distintas seis de las cuales corresponden a derivaciones de miembros (DI – DII – DIII – AVR – AVL – AVF) y las otras seis a derivaciones son las precordiales (V1 – V2 – V3 – V4 – V5 – V6).

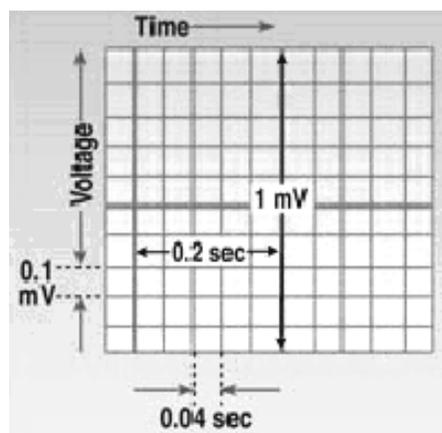


Figura 2.9 *Papel de registro del ECG.* [27]



2.2.3.2.1 Derivaciones de miembros.

Cada derivación constituye un puesto de observación y registro de la actividad eléctrica desde distintos puntos del espacio. Para obtener las derivaciones de miembros, se ponen electrodos en los brazos derecho e izquierdo y en la pierna izquierda, formando un triángulo, en el centro de este se sitúa el corazón, un par de electrodos forma una derivación dando como resultado 3 derivaciones (I – II – III). Si las derivaciones se trasladan al centro del triángulo, forman tres líneas de referencia que se cortan (triángulo de Einthoven) figura 2.10. Como se muestra en el triángulo de Einthoven, cada derivación tiene un principio y un fin partiendo desde un electrodo, el inicio de la flecha indica su polaridad negativa y el final indica su polaridad positiva. [27]

Bipolares:

DI brazo derecho a brazo izquierdo

DII brazo derecho a pie izquierdo

DIII brazo izquierdo a pie izquierdo

Monopolares:

AVR brazo derecho

AVL brazo izquierdo

AVF pie izquierdo

Otra derivación es la derivación AVR. En este caso el brazo derecho es positivo y los demás electrodos de miembros forman una tierra común (negativa). Las dos últimas derivaciones de miembros se llaman AVL y AVF, y se obtienen de manera similar a AVR. En la derivación AVL el brazo izquierdo es positivo y los demás electrodos de miembros se unen a tierra y se consideran negativos. Para la derivación AVF, el electrodo positivo se encuentra en la pierna izquierda.

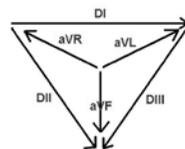


Figura 2.10 *Triangulo de Einthoven.* [27]

Estas tres últimas derivaciones son también derivaciones de miembros. Estas derivaciones se cortan formando ángulos de 60° como las derivaciones I, II y III cuando se trasladan al centro del triángulo. Las derivaciones AVR, AVL y AVF tienen una orientación distinta de las direcciones de las derivaciones I, II y III (son las bisectrices de los ángulos formados por I, II, III). Estas derivaciones de los miembros se encuentran en un mismo plano que puede considerarse el del tórax del paciente (es el plano frontal).

2.2.3.2.2 Derivaciones precordiales

Para obtener las seis derivaciones precordiales, se coloca un electrodo positivo en seis puntos distintos del tórax. En estas derivaciones, el electrodo colocado sobre el pecho se considera positivo. Se enumeran en orden progresivo del 1 al 6 de derecha a izquierda tal como se muestra en la figura 2.11. [27]

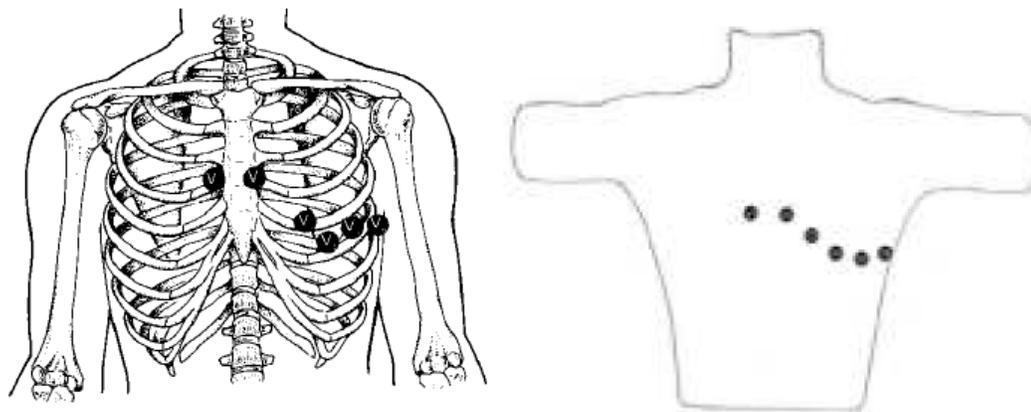


Figura 2.11 *Electrodos precordiales*. [27]

Las derivaciones precordiales se proyectan a través del nódulo AV hacia la espalda de la persona, que constituye el extremo negativo de cada derivación. Debido a esto, este plano divide el cuerpo en mitades superior e inferior, y se llama plano horizontal.

Las derivaciones V1 y V2 se encuentran sobre la parte derecha del corazón, mientras que V5 y V6 están frente al lado izquierdo. V3 y V4 se encuentran sobre la pared común que comparten los ventrículos derecho e izquierdo (tabique interventricular).

2.2.3.3 Análisis de un ECG.

Al reportar el ECG, suelen ponerse en una columna las seis derivaciones de miembros, y en otra las seis derivaciones precordiales (figura 2.12).

Las observaciones de mayor utilidad en la interpretación de un ECG son la frecuencia, el ritmo, el eje, las hipertrofias y los infartos.



Figura 2.12 *Electrodos precordiales.* [27]

2.2.3.3.1 Frecuencia.

La frecuencia es el primer parámetro que debemos medir. Se mide en ciclos/minuto. La frecuencia de los latidos cardiacos depende del nodo SA. Cuando el nodo SA falla otras regiones del corazón son capaces de iniciar el latido del corazón. Se denominan



Marcapasos ectópicos, estos se encuentran en todas las regiones del corazón (aurículas, ventrículos y nodo AV) sin embargo, solo funcionan en caso de emergencia, es decir, cuando deja de funcionar correctamente el nodo SA.

En las aurículas tienen una frecuencia de 75 ciclos/minuto. En situaciones de urgencia, un foco auricular ectópico puede bruscamente empezar a disparar gran frecuencia, de 150-250 ciclos/minuto. Al dejar de funcionar el nodo SA, este marcapasos auricular puede encargarse de la actividad automática. Cuando el estímulo es normal en la aurícula, el nodo AV tiene una frecuencia de 60 ciclos/minuto. Se pueden detectar múltiples enfermedades según la frecuencia cardiaca, sobre todo cuando se encuentra entre los 51-52 ciclos/minuto. [27]

2.2.3.3.2 Ritmo cardiaco

El ECG es la manera más exacta de identificar las arritmias cardiacas (ritmos anormales). El impulso del marcapasos en el nodo SA se extiende por ambas aurículas como una onda de despolarización. El nodo SA (seno auricular) manda impulsos regulares que producen la contracción de la aurícula.

La onda de estimulación producirá la despolarización y se extenderá desde el nodo SA como una onda, y forma la onda P del ECG.

Un ritmo cardiaco normal es aquel en el que la distancia de las ondas suele ser constante, se llama en general ritmo sinusal normal, pues nace en el nodo S A. [27]



2.2.4 CLASIFICACIÓN DE SEÑALES

Un fenómeno físico de interés en ingeniería es normalmente medido en términos de amplitud contra el tiempo, refiriéndose como un registro en el tiempo. A estos registros de n muestras por simplicidad se les conoce como datos [20].

Cualquier dato observado de un fenómeno puede ser clasificado como determinístico y no determinístico, los cuales se definen como [9]:

Determinístico : aquel que puede ser descrito por una relación matemática explícita.

No-Determinístico : no puede ser descrito por relaciones matemáticas explícitas.

De manera que para los datos determinísticos, teniendo su relación matemática, es posible conocer un valor en cualquier tiempo. En cambio con los datos no determinísticos no hay manera de predecir el valor exacto en un instante específico de tiempo.

Los datos no-determinísticos son de carácter aleatorio, conocidos también como datos aleatorios, y pueden ser descritos en términos de postulados, relaciones de probabilidad, y medias estadísticas.

Se dice que la clasificación de datos determinísticos y aleatorios es una descripción única; pero se ha reconocido recientemente otra clasificación donde los datos aparentemente aleatorios son controlados por un proceso determinístico, y se les conoce como caóticos [10]. En la práctica, ningún dato físico es completamente determinístico y ningún dato puede ser totalmente aleatorio [9].

Los datos determinísticos y aleatorios a su vez se dividen en diferentes subgrupos que han sido estudiados ampliamente en [9].



Un panorama general se muestra en la Figura 2.13, algunos de estos subgrupos serán brevemente estudiados en las secciones posteriores.

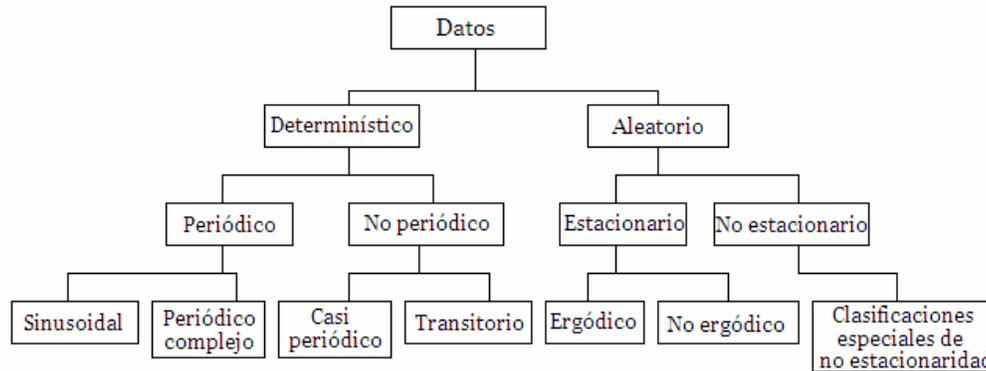


Figura 2.13 *Clasificación de datos* [20].

2.2.5 Proceso determinístico.

Un proceso determinístico es aquel que sus datos tienen un comportamiento determinístico, pueden categorizarse como periódicos o no periódicos.

Los datos periódicos consisten de una forma básica de dimensión finita, que es reproducida infinitamente [20].

Para una señal

$$x(t) = A \text{sen}(2\pi f t + \theta) \tag{2.7}$$

donde:

A = amplitud.

f = frecuencia (ciclos por unidades de tiempo).

θ = desfaseamiento lineal con respecto al origen.

$x(t)$ = valor instantáneo en el tiempo t .



El intervalo requerido para una fluctuación completa o ciclo se le llama período T . El período está ligado a la frecuencia como:

$$T = 1 / f \quad (2.8)$$

Generalmente para una función $x(t)$ que repite sus valores a lo largo del tiempo por una constante T se llama función periódica.

$$x(t) = x(t + nT); n = 1,2,3, \dots \quad (2.9)$$

2.2.6 Procesos aleatorios - datos aleatorios.

Para un fenómeno aleatorio se define su función como una muestra en tiempo finito, y es su representación en un período de tiempo. El conjunto de todas las muestras o colección, es un proceso aleatorio.

El proceso aleatorio hipotéticamente puede ser descrito en cualquier instante de tiempo si la media de la colección de valores lo describe.

El proceso aleatorio puede ser categorizado como estacionario o no. Los procesos aleatorios estacionarios pueden ser ergódicos o no. Los procesos no estacionarios pueden ser categorizados en términos de sus propiedades no estacionarias específicas.

Proceso estacionario.

Sea un proceso aleatorio con valor medio (primer momento) en cualquier instante de tiempo t_i en la colección de datos de la Figura 2.14, dado como [9]:

$$\mu_x(t_1) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x_k(t_1) \quad (2.10)$$



y con una correlación (auto correlación) dada por:

$$R_{xx}(t_1, t_1 + \tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x_k(t_1) x_k(t_1 + \tau) \quad (2.11)$$

Entonces, el proceso aleatorio $\{x(t)\}$ es estacionario bajo las siguientes condiciones.

1. Si $\mu_x(t_1)$ y $R_{xx}(t_1, t_1 + \tau)$ varían, mientras t_1 varía, $\{x(t)\}$ es no estacionario.
2. Si $\mu_x(t_1)$ y $R_{xx}(t_1, t_1 + \tau)$ no varían mientras t_1 varía; el proceso es estacionario en un sentido amplio. Para este proceso el valor medio es una constante, $\mu_x(t_1) = \mu_x$, y la función de auto correlación depende solo del desplazamiento τ , $R_{xx}(t_1, t_1 + \tau) = R_{xx}(\tau)$.
3. Para una colección infinita de momentos de alto orden (ver sección 2.2.9) y de momentos acoplados del proceso aleatorio $\{x(t)\}$; si todos ellos son invariantes en el tiempo entonces $\{x(t)\}$ es fuertemente estacionario o estacionario en un sentido estricto.

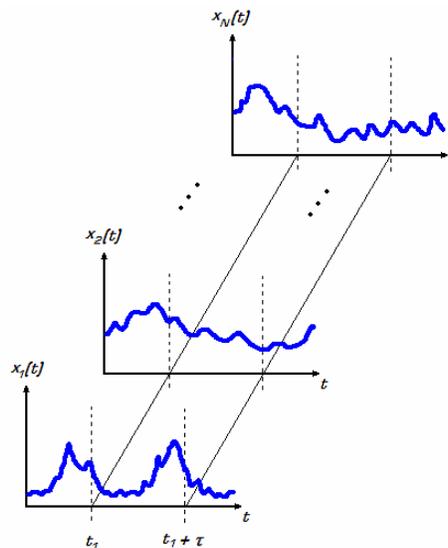


Figura 2.14 Colección de datos aleatorios. [9]



En la figura 2.14 se muestra una colección de datos aleatorios, donde cada par ordenado que se observa en dicha figura, representa como varía la magnitud de cada dato de la colección con respecto al tiempo, que no necesariamente varía de la misma forma que los hacen los demás datos pertenecientes a la colección, es decir, para $t = t_1$, se tiene $x_1(t_1), x_2(t_1), \dots, x_n(t_1)$, con valores determinados no necesariamente relacionados.

Proceso ergódico

Un proceso ergódico se define como sigue: [9].

Para la k -ésima función muestra de $\{x(t)\}$, su valor medio y la correlación están dados por:

$$\mu_x(k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x_k(t) dt \quad (2.12)$$

y con una correlación (auto correlación) dada por:

$$R_{xx}(\tau, k) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x_k(t) x_k(t + \tau) dt \quad (2.13)$$

Si $\{x(t)\}$ es estacionario y $\mu_x(k)$ y $R_{xx}(\tau, k)$ no difieren cuando se calculan para diferentes funciones muestra, el proceso aleatorio se dice que es ergódico, esto es:

$$\mu_x(k) = \mu_x \text{ y } R_{xx}(\tau, k) = R_{xx}(\tau)$$

Solo un proceso estacionario puede ser ergódico, pero un proceso estacionario no necesariamente es ergódico. Para procesos ergódicos todas sus propiedades pueden ser descritas por medias de tiempo de una función muestra.



2.2.7 Análisis de datos aleatorios

Aunque no pueden escribirse ecuaciones matemáticas explícitas para las historias de tiempo producidas por un proceso aleatorio, relaciones de entrada-salida bien definidas existen y son fundamentales para una amplia gama de aplicaciones. Para estos datos o procesos, procedimientos estadísticos deben de ser usados para obtener propiedades y comportamientos estadísticos. [9]

Las propiedades estadísticas básicas de importancia para la descripción de una muestra aleatoria estacionaria definidas por Bendat y Piersol son: [9]

-Media y media cuadrada

Valor medio μ_x : representa la tendencia central de una colección de datos.

Varianza σ_x^2 : es la media de los cuadrados de la desviación del valor medio de una colección de datos medida.

$\mu_x^2 + \sigma_x^2 = \psi_x^2$:: Representa una medida de la combinación de la tendencia central y la dispersión.

-Funciones de densidad de probabilidad

$P(x)$: representa la razón de cambio de probabilidad.

-Funciones de autocorrelación

$R_{xx}(\tau)$: medida de las propiedades de relación de tiempo en los datos, separados por un intervalo.



-Funciones de densidad autoespectral o densidad espectral de potencia, (PSD, siglas en inglés de power spectral density)

$G_{xx}(f)$: son cantidades medidas en la práctica directamente con procedimientos de filtrado que representa la razón de cambio de la media cuadrada con la frecuencia entre $(-\infty, \infty)$.

Para pares de muestras aleatorias de dos procesos estacionarios aleatorios diferentes; las propiedades estadísticas acopladas de importancia son:

- funciones de densidad de probabilidad acoplada
- funciones de correlación cruzada
- funciones de densidad espectral cruzada
- funciones de respuesta a la frecuencia
- funciones de coherencia

Las primeras tres funciones miden las propiedades fundamentales en amplitud, tiempo y frecuencia. Con la función de respuesta a la frecuencia se pueden tratar problemas relacionados con entrada/salida. Con la función de coherencia se mide la exactitud del modelo de entrada/salida.

Las aplicaciones comunes de las funciones de densidad de probabilidad y distribución son:

- Evaluación de la normalidad
- Detección de errores en adquisición de datos
- Indicación de los efectos no lineales
- Análisis de valores extremos

Las aplicaciones principales de las mediciones de correlación son:



- Detección de periodicidades
- Predicción de señales en ruido
- Medición de retrasos en tiempo
- Localización de fuentes de disturbio
- Identificación de rutas de propagación y velocidades

Las aplicaciones típicas de las funciones de densidad espectral son:

- Determinación de las propiedades del sistema a partir de los datos de entrada/salida
- Predicción de los datos de salida a través de los datos de entrada y las propiedades del sistema.
- Identificación de los datos de entrada a través de los datos de salida y las propiedades del sistema.
- Especificaciones de los datos dinámicos para programas de prueba.
- Identificación de fuentes de energía y ruido.
- Filtrado y producción lineal óptimo.

2.2.8 Propiedades estadísticas

En la sección anterior se definieron las clasificaciones de los datos aleatorios, en términos de propiedades estadísticas que serán explicadas a continuación. Estas propiedades son fundamentales para analizar datos y procesos aleatorios, pues mediante ellas éstos pueden ser descritos. Además serán la base para el desarrollo y estudio de las estadísticas de alto orden en la sección 2.2.9. [9]

2.2.8.1 Propiedades estadísticas de las variables aleatorias.

Básicamente son dos propiedades fundamentales; el valor esperado ($E[\]$), y el momento (μ). [27]



Valor esperado

El símbolo E es usado como operador en las variables aleatorias. Es llamado esperanza, valor esperado, o valor promedio. La expresión $E[f(x)]$ es usada para denotar el valor esperado de una función $f(x)$ aplicado a los valores que toma la variable aleatoria x [11] [12]. Se define como [9]:

$$E[x(k)] = \int_{-\infty}^{\infty} x\rho(x)dx \quad (2.14)$$

donde $\rho(x)$ es la función de densidad de probabilidad del proceso aleatorio.

Momentos

El n-ésimo momento de la variable aleatoria x con densidad de probabilidad $\rho(x)$ es definido como:

$$\mu_n(x) = E[x^n] = \int_{-\infty}^{\infty} x^n\rho(x)dx \quad (2.15)$$

El primer momento es llamado media:

$$\mu(x) = E[x] = \int_{-\infty}^{\infty} x\rho(x)dx = \mu_x \quad (2.16)$$

El segundo valor se llama valor cuadrático medio y es:



$$\mu_2(x) = E[x^2] = \int_{-\infty}^{\infty} x^2 \rho(x) dx = \varphi_x^2 \quad (2.17)$$

La varianza es una relación entre la media y el valor cuadrático medio de la forma [13]:

$$\sigma_x^2 = \psi_x^2 - \mu_x^2 \quad (2.18)$$

$$\begin{aligned} \sigma_x^2 &= E[(x - \mu)^2] \\ &= E[x^2 - 2\mu x + \mu^2] \\ &= E[x^2] - E[2\mu x] + E[\mu^2] \\ &= E[x^2] - 2\mu E[x] + \mu^2 \\ &= E[x^2] - 2\mu^2 + \mu^2 \\ &= E[x^2] - \mu^2 \end{aligned} \quad (2.19)$$

Para una secuencia discreta de datos x_i , su momento se puede expresar como:

$$\mu_x = \sum_{i=1}^{N-1} x_i \rho(x_i) \quad (2.20)$$

para los N datos, la probabilidad de cada dato individual es $1/N$, entonces:

$$\mu_x = \frac{1}{N} \sum_{i=1}^{N-1} x_i \quad (2.21)$$

Los momentos de primer orden son simplemente escalares (o vectores) dependiendo de la función. Los momentos de alto orden tienen forma y propiedades matriciales [11].

**Funciones de distribución y densidad de probabilidad.**

Sea $x(k)$ la variable aleatoria de interés. La función de distribución acumulada de probabilidad $P(x)$ se define como la probabilidad que es asignada al conjunto de puntos de k que satisface la desigualdad deseada $x(k) \leq x$ [13].

$$P(x) = \text{Prob} [x(k) \leq x] \quad (2.22)$$

De la ecuación 2.16 se puede ver claramente que:

$$P(a) \leq P(b) \text{ si } a \leq b$$

$$P(-\infty) = 0 \quad P(\infty) = 1$$

Si la variable aleatoria se asume en un rango continuo, entonces la función de densidad de probabilidad $\rho(x)$ puede ser definida como la relación diferencial [13]:

$$\rho(x) = \lim_{\Delta x \rightarrow 0} \left[\frac{\text{prob}[x < x(k) < x + \Delta x]}{\Delta x} \right] \quad (2.23)$$

esto da

$$\rho(x) \geq 0, \quad \int_{-\infty}^{\infty} \rho(x) dx = 1 \quad (2.24)$$

2.2.8.2 Propiedades estadísticas de los procesos aleatorios.

Un proceso aleatorio tiene tres funciones principales usadas para su análisis y caracterización; la media, la correlación y la covarianza.

**Media.**

Sea $x(t)$ un vector de un proceso aleatorio de N datos, su media se define como [9]:

$$E[x(t)] = \int_{-\infty}^{\infty} x(t)\rho(x)dx = \mu_x \quad (2.25)$$

Correlación.

La correlación de un vector de valores $x(t)$ de un proceso aleatorio es definida como: [9]

$$R_{xx}(t_1, t_2) = E[x(t_1)x(t_2)] \quad (2.26)$$

Donde

$$E[x(t_1)x(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t_1)x(t_2)\rho[x(t_1), x(t_2)] dx(t_1)dx(t_2) \quad (2.27)$$

y $\rho[x(t_1), x(t_2)]$ es la función de densidad de probabilidad conjunta de $x(t_1)$ y $x(t_2)$.

Covarianza.

La covarianza de $x(t)$ se define como [9]:

$$\begin{aligned} cov_{xx} &= E[(x(t_1) - E[x(t_1)])(x(t_2) - E[x(t_2)])] \\ &= E[x(t_1)x(t_2) - x(t_1)E[x(t_2)] - E[x(t_1)]x(t_2) + E[x(t_1)]E[x(t_2)]] \\ &= E[x(t_1)x(t_2)] - E[x(t_1)]E[x(t_2)] - E[x(t_1)]E[x(t_2)] + \\ &E[x(t_1)]E[x(t_2)] \\ &= E[x(t_1)x(t_2)] - E[x(t_1)]E[x(t_2)] \end{aligned} \quad (2.28)$$



en el caso que $t_1 = t_2 = t$, $cov_{xx}(t, t)$ representa la varianza de $\{x(t)\}$

$$cov_{xx}(t, t) = E \left[(x(t) - \mu_x(t))^2 \right] = \sigma_x^2(t) \quad (2.29)$$

Para valores arbitrarios de μ_x , la función de covarianza se relaciona con la función de correlación por la función.

$$cov_{xx}(\tau) = R_{xx}(\tau) - \mu_x^2 \quad (2.30)$$

Cuando el proceso $\{x(t)\}$, tiene media cero, su correlación y covarianza son iguales [9] [11]. Los valores de la media y covarianza son los mismos para todos los valores de t , esto quiere decir que son independientes de las transiciones en el tiempo.

Cuando todas las distribuciones de probabilidad son independientes de las relaciones en el tiempo, el proceso aleatorio es estacionario.

2.2.9 Estadísticas de alto orden

Como se planteó anteriormente, los procesos o datos aleatorios a pesar de no tener un modelo exacto que los describan, es posible estudiar dichos procesos conociendo sus propiedades estadísticas. [27]

Las estadísticas hasta ahora descritas pueden describir bien los procesos, sin embargo, puede obtenerse aún más información acerca de ellos con estadísticas de orden superior. A estas estadísticas se les conoce como estadísticas de alto orden (HOS, Higher Order Statistics).

2.2.9.1 Momentos de orden n

Pueden definirse dos funciones que permitan calcular los momentos de una variable aleatoria, estas son la función característica y la función generadora de momentos [12].



Función característica

La función característica de una variable aleatoria x se define como: [14]

$$\Phi_x(\omega) = E[e^{j\omega x}] \quad (2.31)$$

Es una función del número real $-\infty < \omega < \infty$. Si la ecuación 2.31 se escribe en términos de la función de densidad, $\Phi_x(\omega)$ es la transformada de Fourier de $f_x(x)$

$$\Phi_x(\omega) = \int_{-\infty}^{\infty} f_x(x) e^{-j\omega x} dx \quad (2.32)$$

por este hecho, si $\Phi_x(\omega)$ es conocida, $f_x(x)$ puede ser conocida por la transformada inversa de Fourier.

$$f_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Phi_x(\omega) e^{j\omega x} d\omega \quad (2.33)$$

por una diferenciación de $\Phi_x(\omega)$, n veces con respecto a ω y haciendo $\omega = 0$ en la derivada, se puede mostrar que el n -ésimo momento de x está dado por:

$$m_n = (-j)^n \left. \frac{d^n \Phi_x(\omega)}{d\omega^n} \right|_{\omega=0} \quad (2.34)$$

La mayor ventaja de usar $\Phi_x(\omega)$ para encontrar los momentos es que $\Phi_x(\omega)$ siempre existe [12]. Entonces los momentos pueden ser encontrados si $\Phi_x(\omega)$ es conocida. Los momentos de alto orden son una generalización natural de las correlaciones [15].



Función generadora de momentos.

Un promedio estadístico relacionado fuertemente con la función característica es la función generadora de momentos, definida como:

$$M_x(v) = E[e^{vx}] \quad (2.35)$$

donde v es un número real de $-\infty < v < \infty$. Así, $M_x(v)$ está dado por:

$$M_x(v) = \int_{-\infty}^{\infty} f_x(x)e^{vx} dx \quad (2.36)$$

La mayor ventaja de la función generadora de momentos se deriva de su habilidad de dar momentos [46]. Los momentos se relacionan con $M_x(v)$ por la expresión:

$$m_n = \left. \frac{d^n M_x(v)}{dv^n} \right|_{v=0} \quad (2.37)$$

Su desventaja principal, a diferencia de la función característica, es que ésta puede no existir para todas las variables aleatorias. De hecho $M_x(v)$ existe solo si todos los momentos existen. [12]

Una propiedad importante de la función generadora de momentos es que la función generadora de momentos de la suma de variables aleatorias independientes es justamente el producto de las funciones individuales generadoras de momentos. [13]

$$\begin{aligned} M_{x+y}(v) &= E[e^{v(x+y)}] \\ &= E[e^{vx} e^{vy}] \\ &= E[e^{vx}]E[e^{vy}] \end{aligned}$$



$$= M_x(v)M_y(v) \quad (2.38)$$

2.2.9.2 Cumulantes.

Las estadísticas de alto orden, en particular los cumulantes, proporcionan información de alto orden de los datos o señales [16].

Los cumulantes dan la cantidad de correlación de alto orden y la medida de la desviación de la gaussianidad de un proceso [17]. Ya que los momentos son la generalización de las correlaciones, los cumulantes son las combinaciones no lineales específicas de estos momentos [15].

Sean $v = col(v_1, v_2, \dots, v_k)$ y $x = col(x_0, x_1, \dots, x_{k-1})$, colecciones de variables aleatorias (en este caso "col" hace referencia a que v y x están conformados por unas componentes, que a su vez son variables aleatorias). El cumulante de k -ésimo orden de esas variables aleatorias es definido como los coeficientes de (v_1, v_2, \dots, v_k) en la expansión de la serie de Taylor de la función generadora de cumulantes. [18]

$$K_x(v) = \ln E\{e^{jvx}\} \quad (2.39)$$

El cumulante del k -ésimo orden de esta manera se define en términos de los momentos acoplados de orden mayor a k [17]. De tal forma que los cumulantes pueden relacionarse con los momentos, esto es una herramienta útil para encontrarlos.

Relación entre momentos y cumulantes.

Los cumulantes pueden ser encontrados de los momentos y viceversa. "Sea x una colección de variables aleatorias, esto es, $x = col(x_1, x_2, \dots, x_k)$, $I_x = \{1, 2, \dots, k\}$. el conjunto de índices de los componentes de x . Además, si $I \subseteq I_x$ (si I es un subconjunto de I_x), entonces x_I es el vector que consiste de estos componentes de x cuyos índices



pertenecen a I ". Considerando lo anterior, el momento y cumulante del subvector x_1 es $m_x(I) = E[x_1, x_2, \dots, x_k]$ y $c_x(I) = cum[x_1, x_2, \dots, x_k]$ respectivamente, donde "cum" significa cumulante. [16].

La partición de I_x es una colección de conjuntos no interceptados, no vacíos de I , cuya unión forma I_x .

Por ejemplo el conjunto de particiones correspondientes a $k = 3$ consta de las particiones $\{(1,2,3)\}, \{(1), (2,3)\}, \{(2), (1,3)\}, \{(3), (1,2)\}, \{(1), (2), (3)\}$ representando a cada conjunto de particiones como \wp . \wp_q será el conjunto de particiones de I_x de orden q . Así \wp_1 será la partición sencilla de $\{(1,2,3)\}$, \wp_2 constara de tres particiones $P_1: \{(1), (2,3)\}$, $P_2: \{(2), (1,3)\}$, $P_3: \{(3), (1,2)\}$ mientras que \wp_3 consistirá de la partición individual de $\{(1), (2), (3)\}$. Así el cumulante de orden k es obtenido de los momentos de orden menores a k [17], por la fórmula:

$$c_x(I) = cum[x_1, x_2, \dots, x_k] = \sum_{q=1}^k (-1)^{q-1} (q-1)! \sum_{P \in \wp_q} \prod_{I \in P} m_x(I) \quad (2.40)$$

e inversamente los momentos pueden ser obtenidos de los cumulantes por la fórmula:

$$m_x(I) = E[x_1, x_2, \dots, x_k] = \sum_{q=1}^k \sum_{P \in \wp_q} \prod_{I \in P} c_x(I) \quad (2.41)$$

Usando estas ecuaciones se calculan los cumulantes de 1, 2, 3 y 4 orden.

Para el cumulante de primer orden:

Sea $I = \{1\}$ y $k = 1$ $\wp = \{1\}$

$$c_x(I) = \sum (-1)^0 (0)! m_x(I) = E\{x_1\} \quad (2.42)$$

El cumulante de segundo orden es:



$$\begin{aligned} I &= \{1, 2\} \text{ y } k = 2 \text{ } \wp = \{(1,2)\}, \{(1), (2)\} \\ c_{2x}(I) &= \sum_{q=1}^2 (-1)^{q-1} (q-1)! \sum_{P \in \wp_q} \prod_{I \in P} m_x(I) \\ &= m_x(I) + (-1)(1) \prod_{P \in \wp_2} m_x(I) \\ &= E\{x_1 x_2\} - E\{x_1\}E\{x_2\} \end{aligned} \quad (2.43)$$

Para el cumulante de tercer orden:

$$I = \{1, 2, 3\} \text{ y } k = 3$$

$$\wp = \{(1,2,3)\}, \{(1), (2,3)\}, \{(2), (1,3)\}, \{(3), (1,2)\}, \{(1), (2), (3)\}$$

$$\begin{aligned} c_{3x}(I) &= \sum_{q=1}^3 (-1)^{q-1} (q-1)! \sum_{P \in \wp_q} \prod_{I \in P} m_x(I) \\ &= m_x(I) - \sum_{P \in \wp_2} \prod_{I \in P} m_x(I) + (-1)^2 (2)! \prod_{I \in \wp_3} m_x(I) \\ &= E\{x_1 x_2 x_3\} - [E\{x_1\}E\{x_2 x_3\} + E\{x_2\}E\{x_1 x_3\} + E\{x_3\}E\{x_1 x_2\}] \\ &\quad + 2E\{x_1\}E\{x_2\}E\{x_3\} \\ &= E\{x_1, x_2, x_3\} - E\{x_1\}E\{x_2, x_3\} + E\{x_2\}E\{x_1, x_3\} + E\{x_3\}E\{x_1, x_2\} \\ &\quad + 2E\{x_1\}E\{x_2\}E\{x_3\} \end{aligned} \quad (2.44)$$

El cumulante de cuarto orden es:

$$I = \{1, 2, 3, 4\} \text{ y } k = 4$$

$$\begin{aligned} \wp &= \{(1,2,3,4)\}, \{(1), (2,3,4)\}, \{(2), (1,3,4)\}, \{(3), (1,2,4)\}, \{(4), (1,2,3)\} \\ &\quad \{(1,2), (3,4)\}, \{(1,3), (2,4)\}, \{(1,4), (2,3)\}, \{(1), (2,3), (4)\}, \{(2), (1,3), (4)\} \\ &\quad \{(3), (1,2), (4)\}, \{(2), (1,4), (3)\}, \{(1), (2,4), (3)\}, \{(1), (3,4), (2)\}, \{(1), (2), (3), (4)\} \end{aligned}$$

$$c_{4x}(I) = \sum_{q=1}^4 (-1)^{q-1} (q-1)! \sum_{P \in \wp_q} \prod_{I \in P} m_x(I)$$



$$\begin{aligned} &= m_x(I) - \sum_{P \in \emptyset_q} \prod_{I \in P^2} m_x(I) + 2 \sum_{P \in \emptyset_3} \prod_{I \in P^3} m_x(I) + (-1)^3 (3)! \prod_{I \in \emptyset_4} m_x(I) \\ &= E\{x_1, x_2, x_3, x_4\} - [E\{x_1\}E\{x_2, x_3, x_4\} + E\{x_2\}E\{x_1, x_3, x_4\} + E\{x_3\}E\{x_1, x_2, x_4\} + \\ &E\{x_4\}E\{x_1, x_2, x_3\} + E\{x_1, x_2\}E\{x_3, x_4\} + E\{x_1, x_3\}E\{x_2, x_4\} + E\{x_1, x_4\}E\{x_2, x_3\}] + \\ &2E[E\{x_1\}E\{x_2, x_3\}E\{x_4\} + E\{x_2\}E\{x_1, x_3\}E\{x_4\} + E\{x_3\}E\{x_1, x_2\}E\{x_4\} + \\ &E\{x_2\}E\{x_1, x_4\}E\{x_3\} + E\{x_1\}E\{x_3, x_4\}E\{x_2\} + E\{x_1\}E\{x_2, x_4\}E\{x_3\}] - \\ &6[E\{x_1\}E\{x_2\}E\{x_3\}E\{x_4\}] \end{aligned} \quad (2.45)$$

Igualmente calculando momentos a través de los cumulantes.

El momento de primer orden es:

$$\begin{aligned} m_x(I) &= \sum_{q=1}^k \sum_{P \in \emptyset_q} \prod_{I \in P} c_x(I) \\ &= c_x(I) = c(x_1) \end{aligned} \quad (2.46)$$

El momento de segundo orden es:

$$\begin{aligned} m_{2x}(I) &= \sum_{q=1}^2 \sum_{P \in \emptyset_q} \prod_{I \in P} c_x(I) \\ &= c_x(I) + \prod_{I \in \emptyset_2} c_x(I) \\ &= c(x_1 x_2) + c(x_1)c(x_2) \end{aligned} \quad (2.47)$$

Para el momento de tercer orden se tiene:

$$\begin{aligned} m_{3x}(I) &= \sum_{q=1}^3 \sum_{P \in \emptyset_q} \prod_{I \in P} c_x(I) \\ &= c_x(I) + \sum_{P \in \emptyset_2} \prod_{I \in \emptyset_2} c_x(I) + \prod_{I \in \emptyset_3} c_x(I) \\ &= c(x_1, x_2, x_3) + [c(x_1)c(x_2, x_3) + c(x_2)c(x_1, x_3) + \\ &c(x_3)c(x_1, x_2)] \end{aligned}$$



$$+c(x_1)c(x_2)c(x_3) \quad (2.48)$$

El momento de cuarto orden entonces será:

$$\begin{aligned} &= c(x_1, x_2, x_3, x_4) + [c(x_1)c(x_2, x_3, x_4) + c(x_2)c(x_1, x_3, x_4) + c(x_3)c(x_1, x_2, x_4) + \\ &c(x_4)c(x_1, x_2, x_3) + c(x_1, x_2)c(x_3, x_4) + c(x_1, x_3)c(x_2, x_4) + c(x_1, x_4)c(x_2, x_3)] + \\ &[c(x_1)c(x_2, x_3,)c(x_4) + c(x_2)c(x_1, x_3)c(x_4) + c(x_3)c(x_1, x_2)c(x_4) + \\ &c(x_2)c(x_1, x_4)c(x_3) + c(x_1)c(x_3, x_4)c(x_2) + c(x_1)c(x_1, x_3)c(x_3)] + \\ &c(x_1)c(x_2)c(x_3)c(x_4) \end{aligned} \quad (2.49)$$

Una propiedad importante y muy usada en la mayoría de los trabajos de estadísticas de alto orden para variables aleatorias de media cero, gaussianos, es que los cumulantes de segundo y tercer orden son idénticos a los momentos de su orden respectivo

$$c_{2x}(x_1x_2) = m_{2x}(x_1x_1) = E\{x_1x_2\} \quad (2.50)$$

$$c_{2x}(x_1x_2x_3) = m_{3x}(x_1x_1x_3) = E\{x_1x_2x_3\} \quad (2.51)$$

En un caso particular, para un proceso aleatorio estacionario de media cero, las variables aleatorias $x(t), x(t + \tau_1) \dots, x(t + \tau_{k-1})$ dependen o están en función de $\tau_1, \tau_2, \dots, \tau_{k-1}$, entonces se podrá escribir como

$$c_{2x}(\tau) = m_{2x}(\tau) = E\{x(t)x(t + \tau)\} \quad (2.52)$$

$$c_{3x}(\tau_1, \tau_2) = m_{3x}(\tau_1, \tau_2) = E\{x(t)x(t + \tau_1)x(t + \tau_2)\} \quad (2.53)$$

Los cumulantes de alto orden son invariantes al corrimiento de la media [15]. Para estos procesos, el cumulante de primer orden es la media, $c_{1x} = E\{x(t)\}$, y el cumulante de segundo orden es la covarianza.

Los cumulantes con retraso cero tienen nombres especiales: $c_{2x}(0)$ es la varianza



σ_x^2 ; $c_{3x}(0,0)$ y $c_{4x}(0,0,0)$ se conocen como γ_{3x} y γ_{4x} , valores que normalizados son la asimetría (skewness) y la curtosis.

Si el proceso está sistemáticamente distribuido la asimetría es cero, pero no viceversa. Y si el proceso tiene una distribución gaussiana la curtosis es cero pero no viceversa [15] [17], siendo el cumulante de cuarto orden para una muestra de media cero:

$$\begin{aligned} c_{4x}(\tau_1, \tau_2, \tau_3) = & E \left\{ x(t)x(t + \tau_1)x(t + \tau_2)x(t + \tau_3) \right\} - \\ & E \left\{ x(t)x(t + \tau_1) \right\} E \left\{ x(t + \tau_2)x(t + \tau_3) \right\} - \\ & E \left\{ x(t)x(t + \tau_2) \right\} E \left\{ x(t + \tau_1)x(t + \tau_3) \right\} - \\ & E \left\{ x(t)x(t + \tau_3) \right\} E \left\{ x(t + \tau_1)x(t + \tau_2) \right\} \end{aligned} \quad (2.54)$$

Sin embargo, no todos los procesos son gaussianos, de hecho la mayoría de los procesos reales no lo son. Por lo cual definir de esta manera a los cumulantes puede resultar práctico a la hora de su desarrollo computacional pero no se obtendría toda la información generada por los procesos o fenómenos que se desean investigar.

Los cumulantes, entonces, de un proceso aleatorio de segundo y tercer orden serían

$$c_{2x}(\tau) = E \left\{ x(t)x(t + \tau) \right\} - E \left\{ x(t) \right\} E \left\{ x(t + \tau) \right\} \quad (2.55)$$

$$\begin{aligned} c_{3x}(\tau_1, \tau_2) = & E \left\{ x(t)x(t + \tau_1)x(t + \tau_2) \right\} + 2E \left\{ x(t) \right\} E \left\{ x(t + \tau_1) \right\} E \left\{ x(t + \right. \\ & \left. \tau_2) \right\} - \\ & \left[E \left\{ x(t) \right\} E \left\{ x(t + \tau_1)x(t + \tau_2) \right\} + E \left\{ x(t + \tau_1) \right\} E \left\{ x(t)x(t + \right. \\ & \left. \tau_2) \right\} + E \left\{ x(t + \tau_2) \right\} E \left\{ x(t)x(t + \tau_1) \right\} \right] \end{aligned} \quad (2.56)$$

y el de cuarto



$$\begin{aligned}
 c_{4x}(\tau_1, \tau_2, \tau_3) = & E \left\{ x(t)x(t + \tau_1)x(t + \tau_2)x(t + \tau_3) \right\} - E\{x(t)\}E \left\{ x(t + \right. \\
 & \left. \tau_1) x(t + \tau_2)x(t + \tau_3) \right\} - E\{x(t + \tau_1)\}E \left\{ x(t)x(t + \tau_2)x(t + \tau_3) \right\} - E \left\{ x(t + \right. \\
 & \left. \tau_2) \right\}E \left\{ x(t)x(t + \tau_1)x(t + \tau_3) \right\} - E\{x(t + \tau_3)\}E \left\{ x(t)x(t + \tau_1)x(t + \tau_2) \right\} - \\
 & E\{x(t)x(t + \tau_1)\}E \left\{ x(t + \tau_2)x(t + \tau_3) \right\} - E\{x(t)x(t + \tau_2)\}E \left\{ x(t + \tau_1)x(t + \right. \\
 & \left. \tau_3) \right\} - E\{x(t)x(t + \tau_3)\}E \left\{ x(t + \tau_1)x(t + \tau_2) \right\} + 2E\{x(t)\}E\{x(t + \tau_1)\}E \left\{ x(t + \right. \\
 & \left. \tau_2)x(t + \tau_3) \right\} + 2E\{x(t)\}E\{x(t + \tau_2)\}E \left\{ x(t + \tau_1)x(t + \tau_3) \right\} + 2E\{x(t)\}E\{x(t + \right. \\
 & \left. \tau_3)\}E \left\{ x(t + \tau_1)x(t + \tau_2) \right\} + 2E\{x(t + \tau_1)\}E\{x(t + \tau_2)\}E \left\{ x(t)x(t + \tau_3) \right\} + \\
 & 2E\{x(t + \tau_1)\}E\{x(t + \tau_3)\}E \left\{ x(t)x(t + \tau_2) \right\} + \\
 & 2E\{x(t + \tau_2)\}E\{x(t + \tau_3)\}E \left\{ x(t)x(t + \tau_1) \right\} - 6E\{x(t)\}E\{x(t + \tau_1)\}E\{x(t + \right. \\
 & \left. \tau_2)\}E \left\{ x(t + \tau_3) \right\}
 \end{aligned} \tag{2.57}$$

Para aprovechar la propiedad de los cumulantes de ser invariantes a los corrimientos de la media; y no hacer todos los cálculos que computacionalmente representarían un tiempo grande de cálculo, a los procesos de media no cero, se les convierte en procesos de media cero restándoles su media, y así se puede aplicar las definiciones más sencillas de los cumulantes [15].

2.2.9.3 Propiedades de los cumulantes.

Los cumulantes de un proceso estacionario tienen muchas simetrías demostradas por Brillinger y Rosemblat [7], algunas de ellas son para cumulantes de segundo, tercero y cuarto orden:

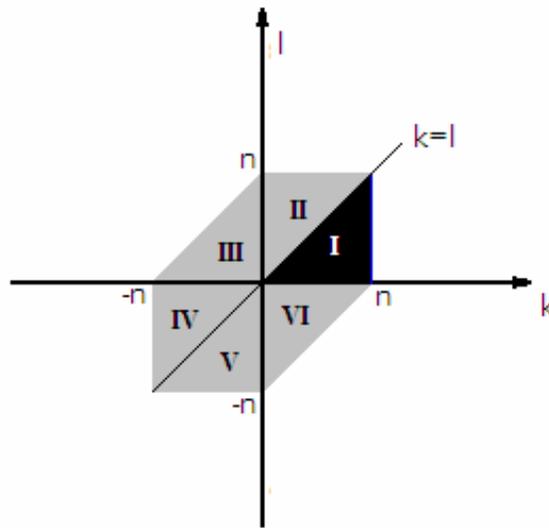
$$c_{2x}(k) = c_{2x}(-k)$$

$$c_{3x}(k, l) = c_{3x}(l, k) = c_{3x}(-k, l - k)$$

$$c_{4x}(k, l, m) = c_{4x}(l, k, m) = c_{4x}(k, m, l) = c_{4x}(-k, l - k, m - k)$$



donde la región fundamental no es el plano $k-l$ entero. Por ejemplo para $c_{2x}(k)$, la región $k \geq 0$ especifica a $c_{2x}(k)$ en todo momento; para $c_{3x}(k, l)$ la región es $0 \leq l \leq k \leq \infty$ (Figura 2.15) dando seis simetrías, y para $c_{4x}(k, l, m)$ es el cono $0 \leq m \leq l \leq k \leq \infty$ [15]. Conociendo el cumulante en una región, puede conocerse el cumulante en todas las regiones. Es importante notar que las simetrías no se conservan para los procesos no estacionarios.



n : número de datos

Figura 2.15 Simetrías del cumulante de tercer orden. [27]

Las propiedades importantes de los cumulantes son [17]:

- Si $\lambda_i, i = 1, \dots, k$ son constantes y $x_i, i = 1, \dots, k$ variables aleatorias

$$cum(\lambda_1 x_1, \dots, \lambda_k x_k) = \left(\prod_{i=1}^k \lambda_i \right) cum(x_1, \dots, x_k)$$

- Los cumulantes son simétricos en sus argumentos

$$cum(x_1, \dots, x_k) = cum(x_{i_1}, \dots, x_{i_k})$$

donde $(x_{i_1}, \dots, x_{i_k})$ es una permutación de $(1, \dots, k)$

- Los cumulantes son aditivos en sus argumentos



$$\text{cum}(x_0 + y_0, z_1, \dots, z_k) = \text{cum}(x_0, z_1, \dots, z_k) + \text{cum}(y_0, z_1, \dots, z_k)$$

- Si α es constante, entonces

$$\text{cum}(\alpha + z_1, z_2, \dots, z_k) = \text{cum}(z_1, \dots, z_k)$$

- Si las variables aleatorias $\{x_i\}$ y $\{y_i\}$ son independientes e $i = 1, \dots, k$ entonces

$$\text{cum}(x_1 + y_1, \dots, x_k + y_k) = \text{cum}(x_1, \dots, x_k) + \text{cum}(y_1, \dots, y_k)$$

- Si un subconjunto de k variables aleatorias $\{x_i\}$ son independientes del resto, entonces $\text{cum}(x_1, \dots, x_k) = 0$. Los cumulantes de un proceso aleatorio independiente idénticamente distribuido son funciones delta, esto es, si $w(t)$ es un proceso independiente idénticamente distribuido entonces $c_{k,w}(\tau_1, \tau_2, \dots, \tau_{k-1}) = \gamma_{k,w}(\tau_1, \tau_2, \dots, \tau_{k-1})$, donde $\gamma_{k,w}$ es el cumulante de k orden de la variable aleatoria w .

2.2.10 Fundamentos básicos del Análisis de Componentes Independientes

2.2.10.1 Introducción

El objetivo fundamental del Análisis de Componentes Independientes (ICA) es el de proporcionar un método que permita encontrar una representación lineal de los datos no gaussianos de forma que las componentes sean estadísticamente independientes o lo más independiente posible. Una representación de este tipo permite obtener la estructura fundamental de los datos en muchas aplicaciones, incluidas la extracción de características y la separación de señales. Seguidamente se va a describir el fundamento teórico en el que se basa dicho método, introduciendo a su vez los conceptos como el Análisis de Componentes Principales (PCA) y la Separación Ciega de Fuentes (BSS). [28]



2.2.10.2 Representación de datos

2.2.10.2.1 Consideraciones estadísticas iniciales

Un problema clásico en estadística y en diversas áreas relacionadas con la misma, es como encontrar una representación conveniente de los datos multidimensionales, entendiendo por ‘representación’ los medios a través de los cuales transformamos dichos datos de forma que su estructura esencial se haga lo más visible o accesible que podamos.

En computación neuronal, este problema pertenece al área del ‘Aprendizaje No Supervisado’, donde dicha representación debe aprenderse a partir del propio conjunto de datos, sin ninguna entrada externa previa, que es lo que constituye la diferencia esencial con el ‘Aprendizaje Supervisado’. En procesamiento de señales, podemos encontrar este mismo problema de la extracción de rasgos, como veremos en el problema de “Separación Ciega de Fuentes” (BSS) más adelante.

Para ilustrar el problema inicial, asumamos que los datos consisten en un número determinado de variables que hemos observado juntas. Denotaremos dicho número de variables por m y el número de observaciones de las mismas por T . Podemos entonces denotar los datos (señales observadas) $x_i(t)$ donde el índice i y t toman valores $i=1, \dots, m$ y $t=1, \dots, T$. Las dimensiones m y T , pueden tomar valores bastante elevados.

Una formulación general del problema se puede realizar considerando que el objetivo es encontrar una función que lleve a cabo la transformación del conjunto original de datos en un espacio m -dimensional a otro espacio n -dimensional, de forma que las variables transformadas en este nuevo espacio nos den tanta información como sea posible sobre las componentes o rasgos que están ocultos en el conjunto de datos originales. En decir, buscamos que las variables transformadas sean las componentes subyacentes que describan la estructura esencial de dichos datos. Es de esperar que dichas componentes



correspondan a alguna causa física en la que estaba inmerso el proceso que generó los datos en primer lugar.

Además, nosotros vamos a considerar únicamente funciones de transformación lineales, ya que la interpretación de la representación de los datos que se lleva a cabo por medio de ellas es más simple, y también su programación. Consecuentemente, cada una de dichas componentes y_i , se puede formular como una combinación lineal de las variables observadas:

$$y_i(t) = \sum_j w_{ij} x_j(t), i = 1, \dots, n, j = 1, \dots, m \quad (2.58)$$

donde los w_{ij} se definen como coeficientes que permiten definir la representación. El problema por tanto se puede reformular como el cálculo de los coeficientes w_{ij} . Mediante el uso del Álgebra lineal es posible expresar la transformación lineal de la ecuación (2.58) como un producto de matrices. Así, agrupando los w_{ij} en una Matriz W , la ecuación anterior se convierte en:

$$\begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} = W \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_m(t) \end{pmatrix} \quad (2.59)$$

que en forma matricial será:

$$y = W \cdot x \quad (2.60)$$

Una aproximación estadística básica consiste en tratar a $x_j(t)$ como un conjunto de T realizaciones de m variables aleatorias. De este modo, cada muestra de la señal $x_j(t)$, $j = 1, \dots, T$ es a su vez una muestra de una variable aleatoria, la cual denotaremos por x_j .



En este contexto, podríamos tratar de determinar la matriz W de acuerdo a ciertas propiedades estadísticas de las componentes transformadas x_j . En las siguientes secciones discutiremos algunas propiedades de esta transformación, y a partir de una de ellas llegaremos al “Análisis de las Componentes Independientes” [29].

2.2.10.2.2 Reducción de dimensión para la obtención de W .

Un principio estadístico clásico que podríamos emplear para la elección de la matriz W es el de limitar el número de las componentes y_i a ser menor, en quizás una o dos unidades que el conjunto de datos originales, determinando entonces W de forma que las y_i contengan tanta información sobre los datos originales como sea. Esto nos lleva a una familia de técnicas conocidas como “Análisis de las Componentes Principales”.

2.2.10.2.3 Uso de la independencia para la estimación de W .

Otro principio estadístico usado para determinar la matriz W es la independencia, es decir, las componentes y_i son estadísticamente independientes unas de otras, lo que significa que el valor que tome cualquiera de ellas no da información alguna sobre el valor que pueda tomar el resto.

Esto resulta sencillo si los datos tiene una distribución Gaussiana, dado que es bastante simple encontrar componentes que sean independientes en este caso, atendiendo a que para datos Gaussianos, las componentes decorrelacionadas son siempre independientes. [29]

Sin embargo, en la realidad los datos no suelen seguir una distribución Gaussiana, y la situación no es tan simple como estos métodos suponen. Muchos conjuntos de datos del mundo real tienen distribuciones supergaussianas, lo que significa que dichas variables aleatorias toman con mayor probabilidad valores que son cercanos al cero o valores muy



grandes, en otras palabras, la función densidad de probabilidad “pdf” (Probability density function) de estos datos es puntiaguda en el cero y tiene las colas densas (debido a los valores grandes que toma lejos del cero), si la comparamos con la pdf de una variable Gaussiana de la misma varianza. Un ejemplo que ilustre esta situación lo encontramos en la figura 2.16.

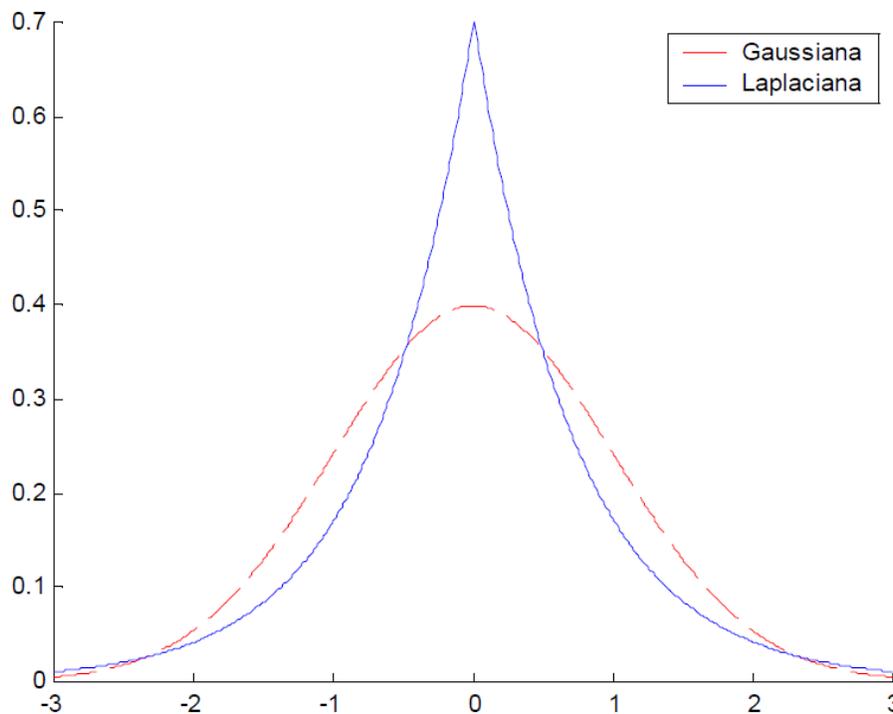


Figura 2.16 *Distribuciones Gaussiana y Laplaciana de media 0 y varianza 1.[30]*

La función densidad de probabilidad de una distribución laplaciana es un ejemplo típico de distribución supergaussiana. En la figura 2.16 se representa una distribución gaussiana (en línea discontinua roja) y una laplaciana (en línea continua azul), estando ambas normalizadas con media cero y varianza unidad.



2.2.10.3 Separación Ciega de Fuentes

El problema de la Separación Ciega de Fuentes (BSS), representa una visión alternativa al de encontrar una representación adecuada de los datos. Además en el podemos observar gran parte del trasfondo de ICA, por lo que su estudio se hace necesario antes de comenzar a analizar el Análisis de Componentes Independientes. [29]

2.2.10.3.1 Observaciones de mezclas de señales desconocidas.

Consideremos un problema donde una fuente emite un cierto número de señales, como por ejemplo, las señales eléctricas correspondientes a diferentes áreas del cerebro o las voces de gente hablando en una misma habitación. Asumamos además que existen una serie de sensores o receptores que van a permitir obtener una grabación de dichas señales mezcladas con unos ciertos coeficientes de ponderación.

Para fijar ideas, supongamos que tenemos tres señales procedentes de la fuente y además tres señales observadas, que han sido recogidas por los sensores. Llamemos $x_1(t)$, $x_2(t)$ y $x_3(t)$ a las señales observadas y $s_1(t)$, $s_2(t)$ y $s_3(t)$ a las señales originales. Se puede comprobar cómo la relación entre ellas vendrá dada por el siguiente sistema de ecuaciones:

$$\begin{aligned}x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t) \\x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t) \\x_3(t) &= a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t)\end{aligned}\tag{2.61}$$

o de forma análoga:

$$x = A \cdot s\tag{2.62}$$

Los coeficientes a_{ij} son constantes que dan los pesos de la mezcla y que agruparemos en una matriz denotada por A. Asumimos que son desconocidos, dado que no



podemos conocer los valores de los coeficientes a_{ij} sin saber todas las propiedades físicas del sistema de mezclado (medio), tarea que puede llegar a ser, generalmente, bastante complicada. Supondremos además que las fuentes originales no son conocidas, de modo que lo único que conocemos son las mezclas.

Lo que nos gustaría hacer es encontrar las señales originales $s_1(t)$, $s_2(t)$ y $s_3(t)$, contando únicamente con las señales mezcla observadas en los sensores $x_1(t)$, $x_2(t)$ y $x_3(t)$. Este es el problema conocido como “Separación Ciega de Fuentes” (BSS). El hecho de incluir el término “Ciega” viene del hecho de que apenas tenemos una mínima información de las señales originales, aunque lo más normal es que no conozcamos nada de las mismas.

Asumiremos que los coeficientes de mezclado a_{ij} toman una serie de valores que permitan que la matriz A sea invertible. De este modo, podemos decir que existe una matriz W con coeficientes w_{ij} tal que nos permitirá separar las señales $s_i(t)$ como:

$$\begin{aligned} s_1(t) &= w_{11}x_1(t) + w_{12}x_2(t) + w_{13}x_3(t) \\ s_2(t) &= w_{21}x_1(t) + w_{22}x_2(t) + w_{23}x_3(t) \\ s_3(t) &= w_{31}x_1(t) + w_{32}x_2(t) + w_{33}x_3(t) \end{aligned} \quad (2.63)$$

o en forma matricial:

$$s = W \cdot x \quad (2.64)$$

Vemos entonces como podremos encontrar la matriz W como la inversa de A , la matriz que formaban los coeficientes a_{ij} , si conocemos dichos coeficientes de mezclado. Ahora observamos que este problema es matemáticamente similar al inicial, donde quisimos encontrar una buena representación de los datos aleatorios $x_i(t)$. Además es posible considerar cada señal $x_i(t)$, $t = 1, \dots, T$ como una muestra de una variable

aleatoria x_i , de forma que el valor de dicha variable viene dado por las amplitudes de esa señal en cada instante de tiempo.

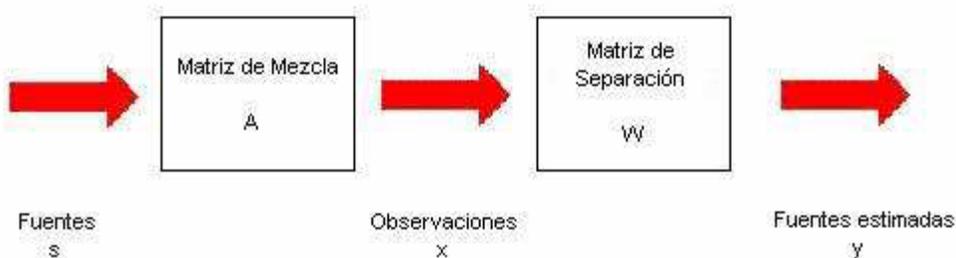


Figura 2.17 *Esquema general que representa los procesos de generación de observaciones y estimación de fuentes en el problema de Separación Ciega de Fuentes.* [30]

La figura 2.17 explica de forma generalizada, que las señales observadas desde distintos puntos de referencia, son el resultado de la mezcla (superposición) de otras señales, consideradas como señales originales (fuentes), dicho resultado se obtiene a través de la matriz de mezcla A , de manera inversa la estimación de las señales “fuentes” son el resultado del producto de la matriz W y las señales observadas.

2.2.10.3.2 Separación Ciega basada en la independencia.

Ahora trataremos de responder a la pregunta de cómo estimar los coeficientes w_{ij} que introducimos en la ecuación (2.63). Para ello, buscaremos un método general que trabaje bajo diferentes circunstancias, y proporcione una respuesta al problema general con el que empezamos: encontrar una buena representación de los datos multidimensionales. Todo lo que tenemos son las señales mezcla, y queremos encontrar una matriz W que nos permita obtener o estimar las señales originales de las fuentes lo mejor posible.

Una solución sorprendentemente simple al problema puede encontrarse suponiendo únicamente que existe independencia estadística entre las señales originales. De hecho, si las señales son no-Gaussianas, esto es suficiente para determinar los coeficientes w_{ij} , de modo que las señales:



$$\begin{aligned}y_1(t) &= w_{11}x_1(t) + w_{12}x_2(t) + w_{13}x_3(t) \\y_2(t) &= w_{21}x_1(t) + w_{22}x_2(t) + w_{23}x_3(t) \\y_3(t) &= w_{31}x_1(t) + w_{32}x_2(t) + w_{33}x_3(t)\end{aligned}\tag{2.65}$$

o en forma matricial

$$y = W \cdot x\tag{2.66}$$

son estadísticamente independientes. Si las señales $y_1(t)$, $y_2(t)$ e $y_3(t)$ son realmente independientes entre sí, entonces serán iguales a las señales originales $s_1(t)$, $s_2(t)$ y $s_3(t)$ (Salvo quizás un factor constante multiplicativo, hecho que no tiene mucha relevancia atendiendo al significado de lo que podemos conseguir).

Usando simplemente esta información sobre la independencia estadística de partida, podemos estimar la matriz W a partir únicamente de las señales mezcla que observamos y conociéndola, podemos estimar las fuentes originales. Veamos a continuación un ejemplo en el que se ilustra este problema que nos servirá para comprender mejor estos conceptos.

Haciendo uso del algoritmo *fastICA*[®] (Gavert, Hurri, Sarela e Hyvarinen), que se ejecutará en el entorno de simulación matemática *MATLAB*[®], trataremos de extraer dos fuentes $s_1(t)$ y $s_2(t)$, a partir de dos observaciones procedentes de la mezcla lineal de ellas y que denotaremos como $x_1(t)$ y $x_2(t)$. Los resultados obtenidos se muestran en las figuras 2.18, 2.19 y 2.20.



SEÑALES ORIGINALES

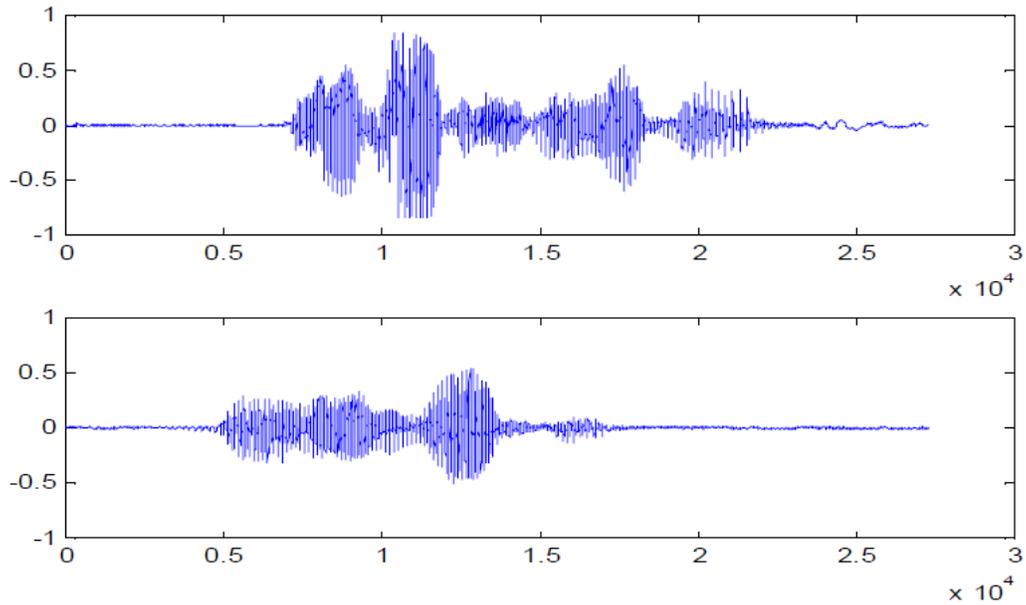


Figura 2.18 Señales originales. [30]

SEÑALES OBSERVADAS

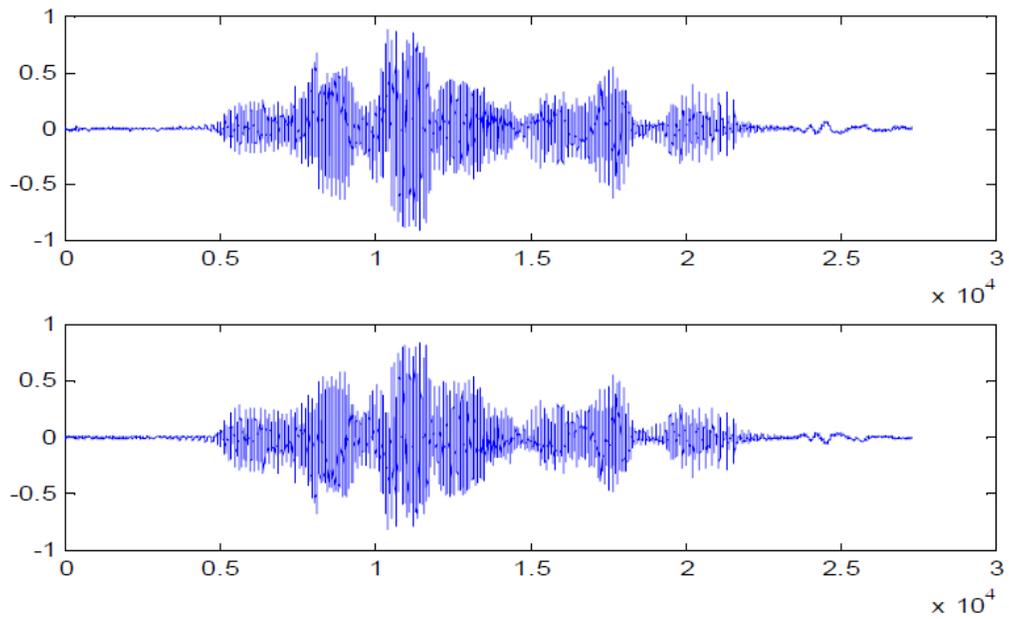


Figura 2.19 Señales observadas. [30]

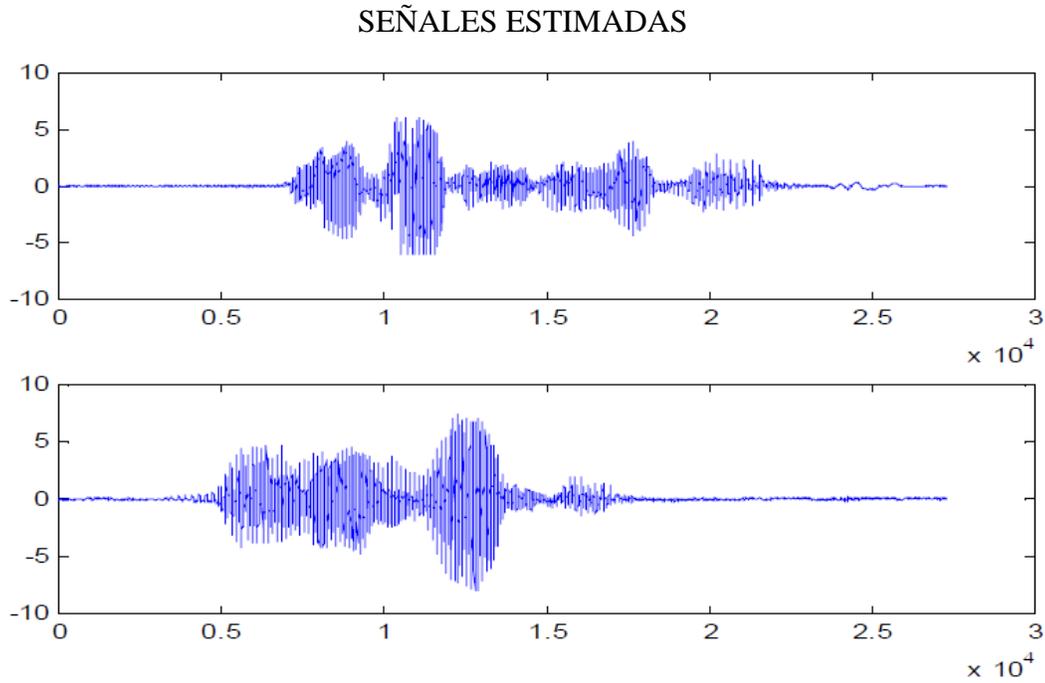


Figura 2.20 *Señales estimadas*. [30]

Podemos observar como las señales estimadas se asemejan a las fuentes originales salvo un factor de escala.

2.2.10.4 El Análisis de Componentes Principales

2.2.10.4.1 Introducción

El Análisis de Componentes Principales (Principal Component Analysis, PCA) es un método clásico de análisis de datos que tiene sus principales aplicaciones en el campo de la extracción de características y compresión de datos. Básicamente equivale a la transformada de Karhunen-Loeve y está estrechamente relacionada con el Análisis de Factores (Factor Analysis, FA). Ambas técnicas están basadas en estadísticos de segundo orden. [31]



El Análisis de Componentes Principales tiene gran utilidad en aplicaciones de reconocimiento facial y de objetos como veremos sobre todo en la parte de este proyecto dedicada a la Factorización No Negativa de Matrices.

2.2.10.4.2 Blanqueado

Un vector de observaciones $\mathbf{z} = (z_1 \dots z_n)^T$ se dice que está blanqueado si los elementos z_i están decorrelacionados y tienen varianzas unidad, es decir:

$$E\{z_i z_j\} = \delta_{ij} \quad (2.67)$$

que en términos de la matriz de covarianzas, implica que, $E\{\mathbf{z}\mathbf{z}^T\} = \mathbf{I}$ siendo \mathbf{I} la matriz identidad. La forma más clara de verlo es a partir del ruido blanco. Se dice que es blanco ya que todas las componentes espectrales son constantes para todo el dominio de la frecuencia. [30]

Además, dado que el blanqueado consiste básicamente en la decorrelación seguida del escalado, es posible usar PCA. El problema del blanqueado será ahora el de obtener una transformación \mathbf{V} tal que aplicada a las observaciones \mathbf{x} de lugar a los datos blanqueados, que denotaremos por \mathbf{z} , esto es

$$\mathbf{z} = \mathbf{V} \cdot \mathbf{x} \quad (2.68)$$

Una forma de verlo, consistiría en hacer uso de la conocida como Descomposición en Valores Singulares (Singular Value Decomposition, SVD).

Llamando $\mathbf{R}_x = E\{\mathbf{z}\mathbf{z}^T\}$, a la matriz de covarianzas de \mathbf{x} , su descomposición en valores singulares dará lugar a:



$$R_x = UDU^T \quad (2.69)$$

donde U es la matriz de autovectores, D la matriz diagonal cuyos elementos de la diagonal se corresponden con los autovalores de R_x (en orden decreciente) y x se supone de media cero [Jenssen00].

Considerando en este caso x como una matriz real y simétrica, el conjunto de autovectores de la matriz de correlación R_x forma un conjunto ortogonal, de forma que se cumple $UU^T = I$, ya que U resulta ser una matriz ortogonal para esas características de x . De este modo, con la siguiente transformación las componentes de z resultan estar decorrelacionadas, dado que pueden ser vistas como una proyección de x en el espacio PCA,

$$y = U^T x \quad (2.70)$$

y por consiguiente podremos calcular la matriz de covarianzas de z de la siguiente manera:

$$\begin{aligned} R_z &= E\{ZZ^T\} = E\{Ux(Ux)^T\} = E\{Uxx^T U^T\} \\ &= UR_x U^T = D^{-\frac{1}{2}} E^T E D E^T E D^{-\frac{1}{2}} = I \end{aligned} \quad (2.71)$$

Por lo que queda demostrado que los nuevos datos están blanqueados ya que su matriz de covarianzas es la matriz identidad. En definitiva, para blanquear los datos existentes, tan solo habrá que realizar la siguiente transformación:



$$z = D^{-\frac{1}{2}}U^T x \quad (2.72)$$

Ya que $z = Vx$, con $V = D^{-\frac{1}{2}}U^T$.

A continuación vamos a estudiar una matriz que cobra especial relevancia y es la que se conoce como Matriz de Blanqueado. Dicha matriz tiene la forma $ED^{-\frac{1}{2}}E^T$ y se puede verificar que es la raíz cuadrada inversa de la matriz de covarianzas de los datos R_x , es decir:

$$ED^{-\frac{1}{2}}E^T = R_x^{-\frac{1}{2}} \quad (2.73)$$

Por último, cabe destacar el hecho de que si las bases PCA vienen dadas por los autovectores de R_x en orden decreciente de sus correspondientes autovalores, podemos descartar los autovectores correspondientes a los autovalores mas pequeños de cara a obtener una dimensión menor, perdiendo la menor información posible en el sentido del error cuadrático medio.

Ejemplo

Para verlo de forma numérica, consideremos el siguiente ejemplo en el que tenemos dos muestras de voz, que denotaremos por x_1 y x_2 y que proceden de la mezcla a su vez de dos señales de voz s_1 y s_2 a partir de la matriz A :

$$A = \begin{pmatrix} 0.8 & -0.3 \\ 0.5 & 0.4 \end{pmatrix} \quad (2.74)$$



La figura 2.21 muestra la grafica “ s_2 vs s_1 ”, siendo estas, las señales originales, se puede observar en este caso, la forma en que están distribuidos los datos.

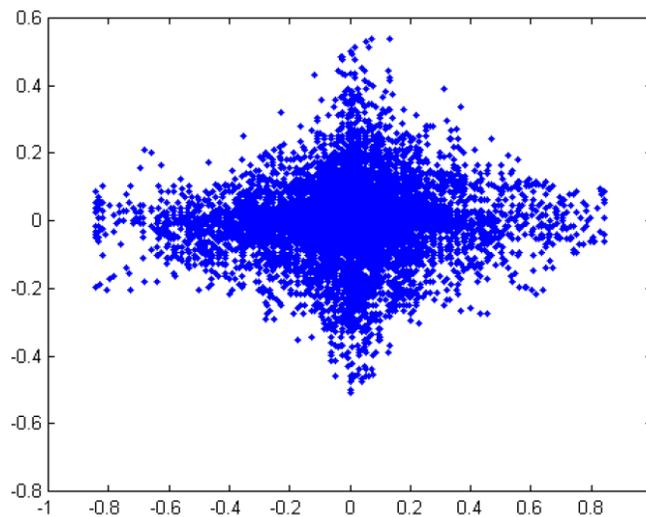


Figura 2.21 Señales originales S_1 (eje horizontal) y S_2 (eje vertical). [30]

La figura 2.22 muestra la grafica “ x_2 vs x_1 ”, siendo estas, el resultado de mezclar, mediante la matriz de mezcla A, las señales originales, se puede observar en este caso, que la forma en que están distribuidos los datos, ha experimentado una rotación, lo cual implica que los datos están mezclados.

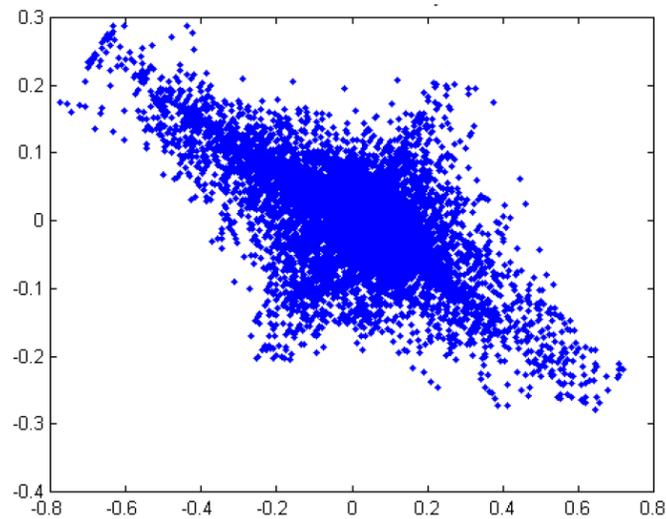


Figura 2.22 Señales resultantes de la mezcla mediante la matriz A , que denotaremos por x_1 (eje horizontal) y x_2 (eje vertical). [30]

Si ahora aplicamos el algoritmo PCA a la matriz formada por los dos vectores x_1 y x_2 , obtendremos los siguientes valores para las matrices U y D :

$$\begin{aligned} U &= \begin{pmatrix} -0.9605 & -0.2782 \\ 0.2782 & -0.9605 \end{pmatrix} \\ D &= \begin{pmatrix} 0.0148 & 0 \\ 0 & 0.0013 \end{pmatrix} \end{aligned} \tag{2.75}$$

por tanto, la matriz W resultará:

$$W = D^{-\frac{1}{2}}U^T = \begin{pmatrix} 0.9611 & -0.2763 \\ 0.2763 & 0.9611 \end{pmatrix} \tag{276}$$

y por último, podremos obtener las nuevas observaciones blanqueadas, que se calcularán como $y = Wx$.



La figura 2.23 muestra la grafica “ z_2 vs z_1 ”, siendo estas, el resultado de aplicarle a las señales observadas, el proceso de blanqueo, mediante la matriz de W.

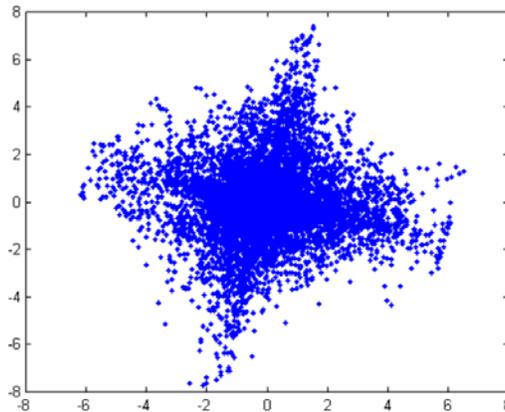


Figura 2.23 Señales obtenidas tras blanquear, z_1 (eje horizontal) y z_2 (eje vertical). [30]

2.2.10.4.3 El Análisis de Factores y PCA

El Análisis de Factores (Factor Analysis, FA), hace referencia a una técnica estadística de análisis de datos muy extendida. Su objetivo consiste en establecer las correlaciones existentes entre un conjunto de variables observadas en términos de un número más pequeño de factores. El Análisis de Factores se desarrolló originariamente en el ámbito de las ciencias sociales y psicología, donde su uso principal radicaba en el desarrollo de técnicas de medida de aspectos como la personalidad o la inteligencia.

Aunque PCA y FA persiguen un mismo objetivo consistente en expresar un conjunto de datos en función de una combinación lineal de un conjunto de datos de dimensión menor (factores en el caso de FA y componentes en el caso de PCA), veremos cómo existe una diferencia esencial entre ambas técnicas. En FA estudia tan solo la varianza compartida entre todos los datos, mientras que PCA analiza todas las varianzas existentes en ese conjunto de datos.



Por este motivo, en la mayoría de las aplicaciones PCA, presenta mejores prestaciones que FA en cuanto a la reducción de dimensiones de los datos estudiados.

2.2.10.5 El Análisis de Componentes Independientes.

2.2.10.5.1 Definición.

Hemos visto como el problema de la ‘separación ciega de fuentes’ se basa en la obtención de una representación lineal en donde las componentes sean estadísticamente independientes. En la práctica, no es siempre posible encontrar una representación donde las componentes sean estadísticamente independientes, sin embargo si será viable la obtención de componentes que al menos sean lo más independientes que se pueda. [30]

Esto nos lleva necesariamente a la definición del Análisis de Componentes Independientes, también conocido como ICA (Independent Component Analysis), que se puede hacer de la siguiente manera:

Dado un conjunto de observaciones de variables aleatorias $\{x_1(t), x_2(t), \dots, x_n(t)\}$, siendo t el tiempo o el índice de las muestras, asumimos que están generadas por una combinación lineal de componentes independientes:

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = A \cdot \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_n(t) \end{pmatrix} \quad (2.77)$$

o en forma matricial:

$$x = A \cdot s \quad (2.78)$$



donde A es una matriz de mezcla desconocida. El Análisis de Componentes Independientes consistirá ahora en estimar tanto la matriz A como las fuentes $s_i(t)$ a partir de las observaciones $x_i(t)$ [29]. Supondremos que el número de observaciones coincide con el de las fuentes originales, si bien esta simplificación no es completamente necesaria para resolver el problema.

De forma alternativa, podríamos definir ICA, como el problema de la obtención de la transformación lineal dada por la matriz W como en (2.60), tal que las variables aleatorias estimadas $y_i(t)$ $i = 1, \dots, n$ sean tan independientes como sea posible. Este planteamiento no difiere en exceso del original (ver sección 2.2.10.2.1), ya que una vez obtenida la matriz A , la matriz W se obtiene invirtiéndola.

Se puede demostrar que el problema está completamente definido, es decir, el modelo planteado en (2.67) puede ser estimado si y solo si las componentes $s_i(t)$ son no gaussianas. Esta condición es esencial y servirá para explicar la diferencia principal que existe entre ICA y PCA, donde la no gaussianidad no se tiene en cuenta.

2.2.10.5.1.1 Definición de ICA basado en un modelo de variables ocultas.

Para definir ICA de una forma rigurosa, es posible usar un modelo de variables ocultas [30]. Se trata de observar n variables aleatorias x_1, \dots, x_n , que se modelan como una combinación lineal de las fuentes s_1, \dots, s_n : [30]

$$x_i = a_{i1} \cdot s_1 + a_{i2} \cdot s_2 + \dots + a_{in} \cdot s_n \quad \forall i = \dots, n \quad (2.79)$$

donde los a_{ij} son coeficientes reales. Por definición los s_i son independientes entre sí.

Este es el modelo básico ICA, que describe como las variables observadas son generadas por un proceso de mezcla de las fuentes s_j . Las componentes s_j son variables

ocultas ya que no se pueden observar de forma directa. Además los a_{ij} pertenecientes a la matriz de mezcla se suponen también desconocidos. Las únicas variables que están “visibles” serán las x_i a partir de las cuales tendremos que estimar las fuentes s_j , y la matriz de mezcla A . Este problema se tendrá que resolver de la manera más general posible.

ICA guarda una estrecha relación con la ‘Separación Ciega de Fuentes’ (BSS). Una “fuente” hace referencia en este contexto a una señal original, por ejemplo una persona hablando en el conocido como “cocktail-party effect” (el efecto fiesta coctel). El hecho de que sea “ciega”, significa que conocemos muy poco (o nada) de la matriz de mezcla, a la vez que se plantean hipótesis muy débiles acerca de las señales “fuente” originales s_j . ICA es quizás el método mas usado en la actualidad para resolver el problema de la separación ciega de fuentes.

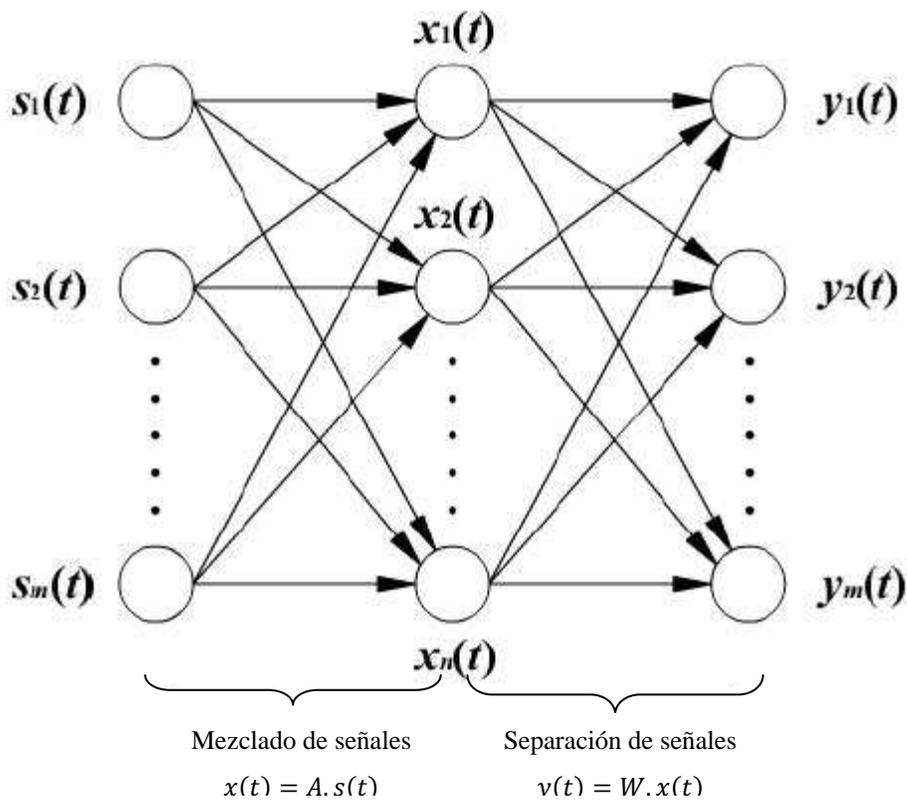


Figura 2.24 Cocktail-Party Effect (el efecto fiesta coctel). [30]



La figura 2.24 hace referencia a Las observaciones $x_i(t)$ con $i = 1, \dots, n$ que proceden de la mezcla de las $s_j(t)$ con $j = 1, \dots, m$ a partir de la matriz A . Las señales estimadas $y_j(t)$ con $j = 1, \dots, m$ se obtendrán a partir de la matriz de separación W . Para que el problema tenga solución, se verificará que el número de observaciones sea mayor o igual que el de señales a estimar, esto es, $n \geq m$. [30]

2.2.10.5.2 Restricciones en ICA

Para poder asegurar que el modelo ICA que estamos planteando tiene solución, es necesario hacer una serie de suposiciones y restricciones que enumeraremos a continuación. [29]

El número de observaciones ha de ser mayor o igual al número de componentes independientes a estimar. En el caso de que dicho número de observaciones sea mayor al de las componentes independientes a estimar, es posible realizar una reducción de la dimensión usando el Análisis de Componentes Principales (PCA), como se indicó anteriormente en el apartado 2.2.10.2.2.

Las componentes independientes se supondrán estadísticamente independientes. En este principio descansa la base de ICA. Sorprendentemente no hace falta mucho más aparte de esta restricción para asegurar que el modelo va a tener solución. Es por esto por lo que ICA es un modelo tan potente y con tantas aplicaciones en diferentes campos.

Básicamente, las variables aleatorias y_1, y_2, \dots, y_n se dice que son independientes si la información en y_i no aporta ninguna información sobre el valor de y_j con $i \neq j$. Técnicamente, la independenciam se puede definir a partir de las densidades de probabilidad. De esta forma, denotemos por $p(y_1, y_2, \dots, y_n)$ a la función densidad de probabilidad conjunta de y_i , y por $p(y_i)$ a la función densidad de probabilidad marginal de y_i . Entonces



consideraremos que las y_i son independientes si y solo si la función densidad de probabilidad conjunta se puede factorizar de la siguiente manera:

$$p(y_1, y_2, \dots, y_n) = p_1(y_1) \cdot p_2(y_2) \cdot \dots \cdot p_n(y_n) \quad (2.80)$$

Las componentes independientes deben presentar distribuciones no gaussianas. De forma intuitiva, se puede pensar que las distribuciones gaussianas son “muy simples”. Los cumulantes de orden superior son cero para distribuciones gaussianas, sin embargo dichos cumulantes son básicos en el análisis ICA. Por tanto, podemos deducir que el análisis ICA es imposible para distribuciones que sean gaussianas.

Por otro lado, cabe recordar que una combinación lineal de componentes gaussianas es a su vez gaussiana y además que para estas distribuciones se cumple que la decorrelación entre dichas componentes va a implicar también independencia (propiedad que no se da en general para cualquier variable aleatoria en este sentido sino en el contrario, es decir, la independencia va a implicar existencia de decorrelación). Esto va a motivar que dada una combinación de componentes independientes gaussianas, sea imposible separarlas por los métodos que estamos describiendo.

Consideraremos que la matriz de mezcla A es cuadrada. Esto implicaría que el número de componentes independientes a estimar es igual al número de observaciones. A partir de esta suposición resulta fácil ver como una vez estimada la matriz de mezcla, al realizarle la inversa, podremos obtener la matriz B , que va a permitir estimar las fuentes originales, esto es:

$$s = Bx \quad (2.81)$$

Lógicamente asumiremos que la matriz de mezcla es invertible (si no lo fuera sería por la existencia de mezclas redundantes y llegaríamos a una situación en la que el número



de componentes independientes y el de las observaciones es distinto y habría que recurrir a técnicas más complejas para llegar a una solución viable).

2.2.10.5.3 Ambigüedades del análisis ICA

En el modelo ICA dado por $x = As$, vemos una serie de ambigüedades que es necesario tener en cuenta antes de seguir desarrollando el modelo. Estas indeterminaciones se especifican en [29] de la siguiente forma:

1. No se puede determinar el orden de las componentes independientes. Esto es debido a que tanto s como A son desconocidas a priori, de forma que si planteamos el modelo ICA como:

$$x = \sum_{i=1}^n a_i s_i \quad (2.82)$$

siendo a_i cada una de las columnas de A y s_i las fuentes, entonces en todo momento sería posible cambiar el orden de los términos en la sumatoria y colocar cualquiera de las componentes independientes en primer lugar. De manera más rigurosa podemos decir que se podría insertar una matriz de permutación P y su inversa de forma que el modelo resultaría así:

$$x = A \cdot P^{-1} \cdot P \cdot s \quad (2.83)$$

De esta manera, la matriz $A \cdot P^{-1}$ resultaría ser la nueva matriz de mezcla a estimar en el análisis ICA.

2. No es posible determinar las energías de las componentes independientes. Esto es debido de nuevo a que al ser A y s desconocidas, cualquier escalar λ_i que multiplicara a una fuente s_i se podría cancelar dividiendo la correspondiente columna a_i de A , esto es:



$$x = \sum_{i=1}^n \frac{1}{\lambda_i} a_i s_i \lambda_i \quad (2.84)$$

La forma más habitual de proceder será suponer que las componentes independientes a estimar van a tener varianza unidad, es decir, $E[s_i^2] = 1$. De esta forma podremos hacer que la matriz A se calcule teniendo en cuenta esta restricción en nuestra solución ICA. También hay que destacar que sigue habiendo problemas con el signo ya que por lo visto anteriormente podríamos multiplicar una fuente por “-1” sin que ello afectara al modelo. Sin embargo este caso es poco relevante en la mayoría de las aplicaciones.

2.2.10.5.4 Búsqueda de las componentes independientes.

A priori puede resultar sorprendente que las componentes independientes puedan ser estimadas a partir de las combinaciones lineales, sin haber impuesto ninguna restricción además de su independencia. Para poder explicar este hecho vamos a ver una serie de aspectos que quizás puedan aclarar esta cuestión. [29]

1. La decorrelación no es suficiente.

Lo primero que tenemos que ver es que la independencia es una propiedad mucho más fuerte que la propia decorrelación. Considerando el problema de la separación ciega de fuentes, vemos que como es posible encontrar multitud de configuraciones de señales decorrelacionadas que por el contrario no serían independientes y por tanto no extraerían las fuentes. Es decir, la decorrelación de por sí, no garantiza una correcta separación de las componentes. Este es el motivo principal por el que PCA no es válido para separar las fuentes ya que exclusivamente permite obtener componentes que son decorrelacionadas, pero poco más.



2. La decorrelación no lineal es la base de ICA

Una forma de ver que la independencia es una condición más fuerte que la decorrelación consiste en establecer que la independencia implica decorrelación no lineal. Si s_1 y s_2 son independientes, entonces las transformaciones no lineales $g(s_1)$ y $h(s_2)$ son decorrelacionadas (su covarianza es cero). Por el contrario, para dos variables aleatorias en general que son decorrelacionadas (pero no independientes), dichas transformaciones no tienen covarianza cero en general. [29]

De esta manera, podríamos intentar abordar nuestro problema ICA usando una forma más potente de decorrelación, es decir, buscando una cierta representación donde las y_i son decorrelacionadas incluso después de aplicarle alguna transformación no lineal. Esto nos llevará a un primer principio de estimación de la matriz de separación W :

- Principio de decorrelación no lineal: Consiste en encontrar la matriz W tal que para todo $i \neq j$ las componentes y_i e y_j están decorrelacionadas, y las correspondientes transformaciones $g(y_i)$ y $h(y_j)$ están decorrelacionadas, siendo g y h funciones no lineales apropiadas.

Si escogemos las no linealidades de una forma apropiada para nuestro fin, esta resulta una aproximación válida para que el método que busquemos halle las componentes independientes.

Sin embargo, llegados a este punto cabría preguntarse la forma de elegir las transformaciones no lineales g y h . Para ello debemos basarnos en principios de teoría de la estimación y teoría de la información, que proporcionarán métodos tales como el de la máxima verosimilitud e información mutua respectivamente.

3. Las componentes independientes son las componentes máximamente gaussianas.



Otro principio que requiere de nuestra atención dentro del estudio del “Análisis de Componentes Independientes”, no es otro que el de la estimación de la máxima no gaussianidad. Este concepto se fundamenta en el conocido como “Teorema Central del Límite”, que establece que la suma de variables no gaussianas dará lugar a una nueva variable que será “mas gaussiana” que las originales.

Esto va a motivar que dada una combinación lineal de las variables observadas $y = \sum_i b_i x_i$, esta será máximamente no gaussiana si es igual a alguna de las componentes independientes que conforman la mezcla. Esto es debido a que si se tratara de una mezcla real de dos o más componentes, estaría más cerca de parecerse a una distribución gaussiana en virtud del Teorema Central del Límite.

Esto nos llevará a formular un segundo principio de la siguiente manera:

- Principio de máxima no gaussianidad: Se trata de encontrar el máximo local de no gaussianidad de la combinación lineal $y = \sum_i b_i x_i$, sujeto a la condición de que la varianza de y ha de ser constante. Cada máximo local implicara que estamos ante una componente independiente.

La kurtosis y su importancia en la medida de la no gaussianidad.

En la práctica, para medir la no gaussianidad haremos uso de la kurtosis. La kurtosis se define a partir de los momentos de tercer y cuarto orden, y para el caso en el que la media sea cero, tendrá la siguiente forma: [29]

$$kurt(y) = E[y^4] - 3 \{E[y^2]\}^2 \quad (2.85)$$

o de forma normalizada:



$$kurt(y) = \frac{E[y^4]}{\{E[y^2]\}^2} - 3 \quad (2.86)$$

A partir de esta definición resulta sencillo establecer medidas de la gaussianidad de una cierta variable aleatoria. Se puede demostrar que si y viene dada por una distribución gaussiana, entonces se verifica que su kurtosis es cero. Por eso, la kurtosis será una buena medida de la no gaussianidad.

De esta forma podremos hacer una clasificación de cualquier distribución, atendiendo al valor que toma su kurtosis y llegaremos a las siguientes conclusiones [29]:

- Una distribución cuya kurtosis vale cero, se denomina gaussiana.

- Si la kurtosis es negativa, dicha variable se denomina subgaussiana. Las distribuciones subgaussianas son en general mas “planas” que las gaussianas.

- Se llaman variables supergaussianas a aquellas cuya kurtosis es positiva. Las distribuciones supergaussianas suelen tener “picos” muy acusados y colas más largas que las gaussianas.

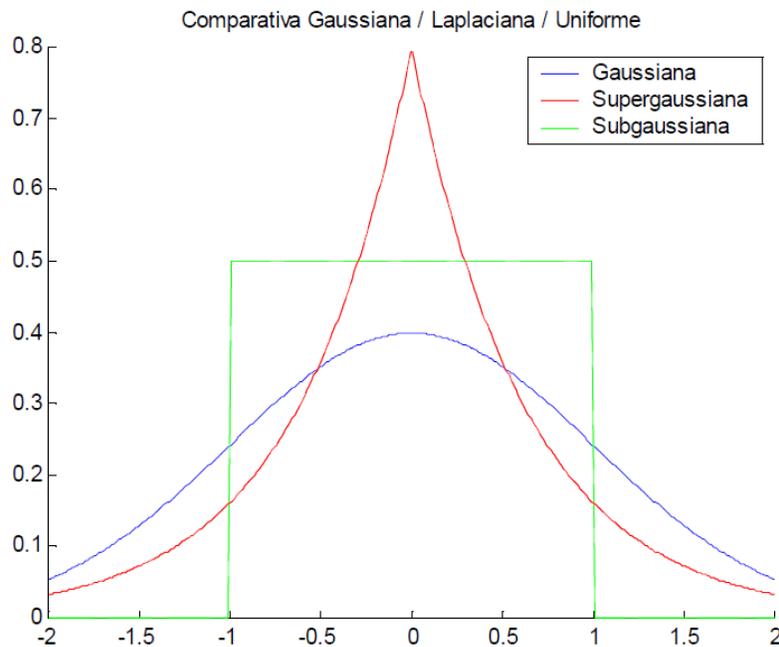


Figura 2.25 Comparación entre distribuciones gaussiana (en azul), laplaciana (en rojo) y uniforme (en verde). [30]

Las distribuciones mostradas en la figura 2.25, representan ejemplos típicos de distribuciones supergaussianas (kurtosis > 0), gaussianas (kurtosis $= 0$) y subgaussianas (kurtosis < 0), en función del valor que tome la kurtosis. [30]

Sin embargo es preciso tomar una serie de precauciones a la hora de tomar la kurtosis como medida de la no gaussianidad. El motivo es que la kurtosis de una variable supergaussiana puede tomar un valor excesivamente elevado, pero en el caso de las variables subgaussianas existe un valor mínimo que puede tomar, que será “-2” (cuando la varianza está normalizada a la unidad). Esto implica que comparar la no gaussianidad de una variable subgaussiana y otra supergaussiana, no sería del todo correcto.



4. La importancia de los estadísticos de orden superior.

En la práctica existen multitud de métodos para realizar la estimación ICA, si bien todos ellos se caracterizan por usar estadísticos que no aparecen en la matriz de covarianzas, son los llamados estadísticos de orden superior. Mediante la matriz de covarianzas podemos decorrelacionar de forma lineal pero no más allá. Es por eso por lo que se hace necesario el uso de dichos estadísticos, como la kurtosis y las correlaciones no lineales, aunque es posible, como se mencionó anteriormente, emplear una amplia variedad de métodos para llevar a cabo el análisis ICA.

5. Los métodos numéricos.

Debido a la enorme carga computacional que puede presentar el problema de estimación ICA, es necesario disponer de herramientas matemáticas potentes y algoritmos que permitan subsanar las necesidades de cálculo existentes. Sin embargo estos serán no lineales, por lo que tendremos que utilizar métodos numéricos.

Algunos algoritmos utilizados son el Método del Gradiente, o los algoritmos FastICA y ThinICA, desarrollados para explotar las características especiales que presenta ICA.

2.2.10.5.5 Teoría de la Información: una forma alternativa de estimar la transformación ICA

2.2.10.5.5.1 Entropía e Información Mutua.

La entropía diferencial H de una variable aleatoria y con densidad $p(y)$ se define como [29]:



$$H(y) = -\int p(y) \log p(y) dy \quad (2.86)$$

La entropía está estrechamente relacionada con la longitud del código de la variable aleatoria. Una versión normalizada de la entropía, viene dada por lo que se conoce como entropía negativa (o negentropy), que se define de la siguiente manera:

$$J(y) = H(y_{gauss}) - H(y) \quad (2.87)$$

siendo y_{gauss} una variable aleatoria gaussiana con la misma matriz de covarianzas que y . La entropía no negativa es siempre positiva y cero en el caso de las variables aleatorias gaussianas.

La información mutua I entre “ m ” variables aleatorias $y_i, i = 1 \dots m$, se define así:

$$I(y_1, y_2, \dots, y_n) = \sum_{i=1}^m H(y_i) - H(y) \quad (2.88)$$

2.2.10.5.2 La información mutua como medida de la independencia.

La información mutua entre variables aleatorias tiene en cuenta la dependencia completa que existe entre dichas variables y no solo la covarianza como ocurre en el caso del Análisis de Componentes Principales (PCA).

Por eso, es posible usar la información mutua de cara a obtener la representación ICA. Para ello definimos el modelo ICA de una variable aleatoria x como una transformación invertible del tipo: [29]



$$y = Bx \quad (2.89)$$

donde la matriz B se calcula de forma que la información mutua de las componentes y_i sea mínima. Además se puede demostrar que la minimización de la información mutua va a llevar a maximizar la independencia de las componentes.

2.2.10.5.5.3 Entropía de una transformación.

Consideremos una transformación invertible de una variable aleatoria x que denotaremos por $y = f(x)$. Llamando $Jf(\zeta)$ a la matriz jacobiana de la función f , la relación entre las funciones densidad de probabilidad de y y de x , que llamaremos p_y y p_x respectivamente se define como: [30]

$$p_y(\eta) = p_x(f^{-1}(\eta)) |det Jf(f^{-1}(\eta))|^{-1} \quad (2.90)$$

Expresando ahora la entropía a partir del valor esperado:

$$H(y) = -E\{\log p_y(y)\} \quad (2.91)$$

e introduciendo el valor de p_y de la ecuación anterior, llegaremos a la ecuación que permite obtener la entropía de una transformación:

$$H(y) = H(x) + E\{\log |det Jf(x)|\} \quad (2.92)$$

2.2.10.5.5.4 Información mutua y no gaussianidad.



A partir de la ecuación (2.92), vamos a construir el razonamiento que nos va a permitir llegar a un resultado para la información mutua. Sea una transformación lineal e invertible $y = Bx$, de forma que su información mutua se podrá escribir como:

$$I(y_1, y_2, \dots, y_n) = \sum_i H(y_i) - H(x) - \log|\det B| \quad (2.93)$$

Ahora veamos lo que ocurre si tomamos las y_i como incorrelacionadas entre sí y con varianza unidad. Esto significa que, $E\{yy^T\} = BE\{xx^T\}B^T = I$, lo cual implica que al tomar determinantes en ambos miembros de la expresión anterior:

$$\det(BE\{xx^T\}B^T) = \det(B)\det(E\{xx^T\})\det(B^T) = \det(I) = 1 \quad (2.94)$$

lo cual lleva a que $\det(B)$ ha de ser constante ya que $\det(E\{xx^T\})$ no depende de B. Además para las, y_i que tienen varianza unidad, la entropía y la entropía negativa difieren solo en una constante y en el signo como vemos en la ecuación (2.87). De esta forma, llegaremos a que:

$$I(y_1, y_2, \dots, y_n) = \text{const.} - \sum_i J(y_i) \quad (2.95)$$

donde el término constante no depende de B.

La ecuación anterior muestra la relación existente entre la información mutua y la entropía no negativa.

Hemos visto en la ecuación (2.95) que encontrar una transformación lineal e invertible que minimice la información mutua es equivalente a encontrar las direcciones en las que entropía negativa se maximiza. Además se sabe que la entropía negativa es una medida de la no gaussianidad por lo que podemos afirmar que la estimación de la



transformación ICA mediante la minimización de la información mutua es equivalente a maximizar la suma de las no gaussianidades de las componentes independientes estimadas.

2.2.10.5.6 Un ejemplo práctico sobre ICA

A continuación vamos a ver un ejemplo numérico sobre el modelo ICA desde el punto de vista estadístico.

Consideremos dos componentes independientes que presentan las siguientes distribuciones uniformes:

$$p(s_i) = \begin{cases} \frac{1}{2\sqrt{3}}, & |s_i| \leq \sqrt{3} \\ 0 & \text{en otro caso} \end{cases} \quad (2.96)$$

donde se verifica que la media va a ser cero y la varianza unidad. La función densidad de probabilidad conjunta de s_1 y s_2 va a ser uniforme en un cuadrado. Esto se debe a que la función densidad de probabilidad conjunta de dos variables independientes es el producto de las marginales como vimos en la ecuación (2.70).

La función densidad de probabilidad conjunta se muestra en la figura 2.26:

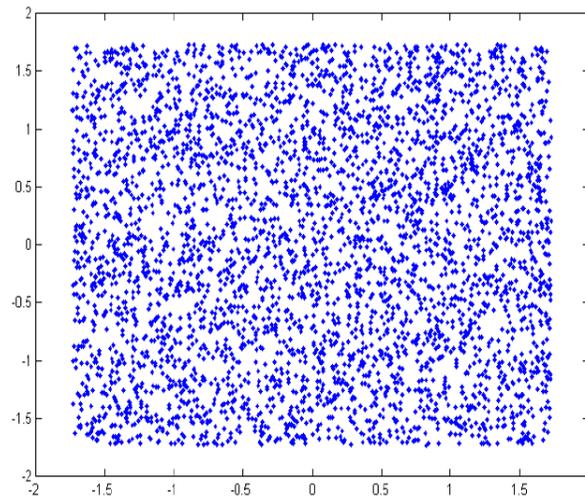


Figura 2.26 Densidad de probabilidad conjunta de la distribución (2.96). El eje horizontal representa ‘ S_1 ’ mientras que el vertical representa ‘ S_2 ’.

A continuación, vamos a proceder a realizar una mezcla de las componentes independientes usando para ello la siguiente matriz de mezcla:

$$A = \begin{pmatrix} 3 & 5 \\ 9 & 6 \end{pmatrix} \tag{2.97}$$

y tras realizar las operaciones pertinentes, mostramos en la figura 2.27 el resultado obtenido:

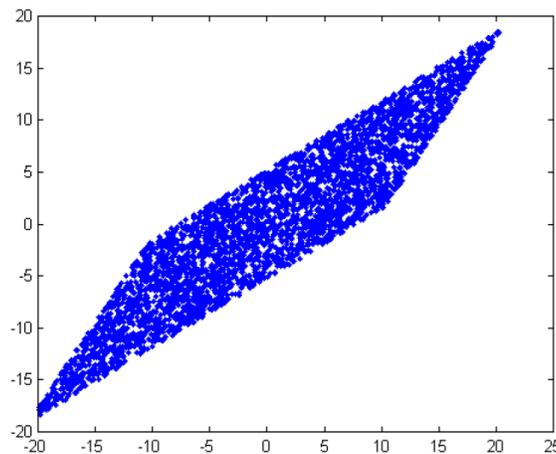


Figura 2.27 Densidad de probabilidad conjunta de la distribución (2.96) tras aplicarle la rotación dada por la matriz A_0 . El eje horizontal representa ‘ x_1 ’ mientras que el vertical representa ‘ x_2 ’. [30]



Observando las figuras anteriores, resulta inmediato ver que las nuevas variables aleatorias generadas x_1 y x_2 ya no son independientes. Una manera sencilla de verlo, consiste en comprobar si es posible determinar el valor de una de ellas a partir de la otra. A partir de la figura 2.27 es posible determinar que si la variable x_1 (representada en el eje horizontal) toma su valor máximo, esto es, en la esquina superior derecha de la distribución, entonces el valor de x_2 quedaría perfectamente determinado. Por el contrario, si se realiza el mismo experimento para la distribución de la figura 2.26, vemos como en este caso, no sería posible determinarlo.

El problema del estimado del modelo ICA va a consistir por tanto en la determinación de la matriz de mezclas \mathbf{A} usando tan solo la información contenida en las variables aleatorias x_1 y x_2 . De forma intuitiva, en el ejemplo que estamos desarrollando se podría estimar esta matriz \mathbf{A} teniendo en cuenta que los ejes del paralelogramo de la figura 2.27 están según las direcciones de las columnas de dicha matriz.

Esto significa que en principio, sería posible estimar el modelo ICA determinando en primer lugar la densidad de probabilidad conjunta de la distribución formada por x_1 y x_2 , y posteriormente localizar los bordes que permitirían obtener la matriz \mathbf{A} .

Por otro lado, podríamos considerar la mezcla de dos componentes independientes que no fueran uniformes (y por tanto subgaussianas), por ejemplo una distribución supergaussiana como es el caso de la laplaciana. Para verlo, consideraremos la distribución laplaciana dada por la siguiente función densidad de probabilidad, donde tomaremos $\lambda = 1.6$

$$p(s_i) = \frac{\lambda}{2} \exp(-\lambda|s_i|) \tag{2.98}$$

cuya representación en el plano formado por s_1 y s_2 se muestra en la figura 2.28:

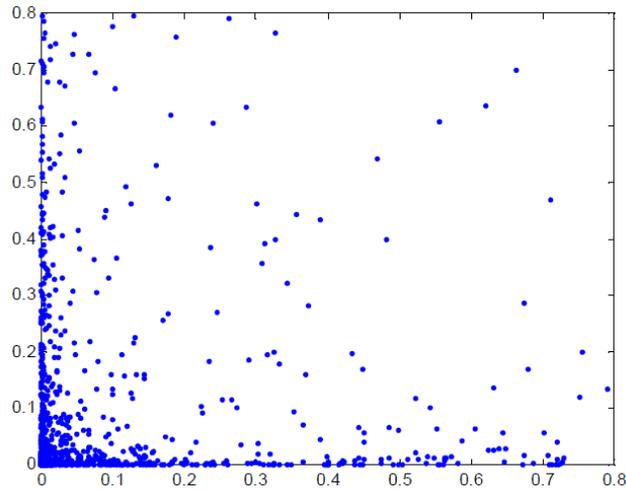


Figura 2.28 Función densidad de probabilidad conjunta de las componentes s_1 y s_2 representadas por distribuciones laplacianas (supergaussianas). En el eje horizontal esta s_1 y en el vertical s_2 . [30]

Si procedemos de igual forma que para la distribución uniforme, realizaremos una rotación usando para ello la misma matriz A dada por la ecuación (2.97) y obtendremos la distribución que se indica en la figura 2.29:

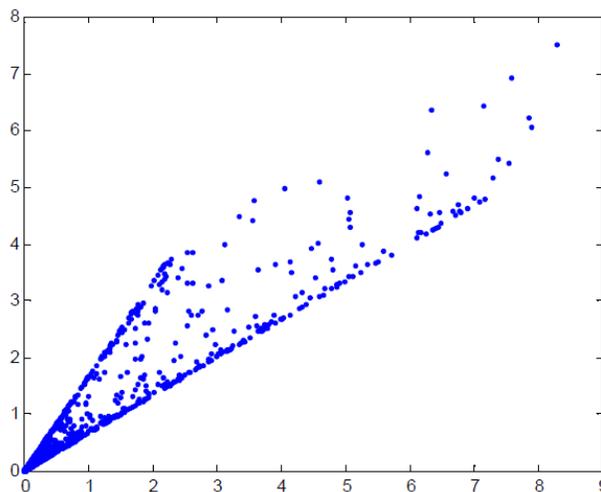


Figura 2.29 Rotación de la distribución de la figura 2.28 usando la matriz A . El eje horizontal representa " x_1 " mientras que el vertical representa " x_2 ". [30]



CAPÍTULO II. MARCO TEÓRICO



En este ejemplo vemos como ya resulta más difícil localizar los bordes que permitirían discernir el valor de los coeficientes que formarían la matriz A . En la práctica veremos cómo usar este método puede resultar computacionalmente complicado e inviable. Lo que realmente necesitaremos será un método que funcione para cualquier distribución de componentes independientes y que lo haga de forma rápida.



CAPITULO III. MARCO METODOLÓGICO



CAPITULO III.

MARCO METODOLÓGICO

3.1. CONSIDERACIONES GENERALES:

Partiendo de las características propias del problema en estudio y de los objetivos planteados al inicio de la investigación, en el presente marco metodológico, se introducirá, de forma sistémica los diversos procedimientos con que se cuenta para recopilar, representar y analizar los datos, con la finalidad de cumplir con el propósito de esta investigación. En este orden de ideas se desarrollarán los aspectos relacionados al diseño y tipo de estudio que se llevará a cabo, que ayudara a delimitar dicho estudio.

3.2 DISEÑO Y TIPO DE INVESTIGACIÓN

El trabajo desarrollado es una investigación de campo, y a su vez cumple con los lineamientos de un proyecto factible ya que los resultados podrán ser medidos con su implementación. Un proyecto factible se define como: “La investigación, elaboración y desarrollo de una propuesta de un modelo operativo viable para solucionar requerimientos o necesidades de organizaciones o grupos sociales; puede referirse a la formulación de políticas, programas, tecnologías, métodos o procesos. El proyecto debe tener el apoyo en una investigación de tipo documental, de campo o un diseño que cumpla con ambas modalidades. [32]



Pasos de una investigación de campo:

- 1) Planteamiento del problema.
- 2) Etapa exploratoria o investigativa. - Lecturas. - Conversaciones con profesores. - Entrevistas a personas que conocen el problema por experiencia personal o debido a sus estudios.
- 3) Delimitaciones operativas del problema.
- 4) Construcción o adquisición de los instrumentos de recolección de datos (computador, conexión a internet).
- 5) Diseño y elaboración de métodos informáticos del proceso que permitan un previo análisis de los datos (pequeñas rutinas elaboradas en matlab).
- 6) Realizar un índice o cronograma de actividades provisorio.
- 7) Discutir el plan de trabajo con el tutor. Luego, hacer un análisis de los datos obtenidos para ver si el plan de análisis va a ser eficaz, gracias a ello se pueden obtener pronósticos de los resultados. La etapa piloto sirve para probar tanto los instrumentos de recolección como de análisis.
- 8) Recolección de los datos.(a través de la pagina physionet).
- 9) Codificación, entrada de datos en computadora, procesamiento y análisis.(software especializado como MATLAB, y toolboxes como fastICA y EEGLab)
- 10) Redacción del borrador. Lectura compartida. Discusión con el tutor.



11) Redacción final. Agregar cuadros, fotografías, gráficos.

3.3 Técnicas de Recopilación de Datos

Para el presente trabajo de investigación la recopilación de datos se realizó haciendo uso de las siguientes técnicas:

3.3.1 Análisis Documental:

Se realizó la recopilación de material bibliográfico que precedía a este trabajo, relacionado con el estudio de “**ANÁLISIS DE COMPONENTE INDEPENDIENTES ICA**” y a fenómenos que son objeto de estudio a través de la técnica antes mencionada, que son los fenómenos aleatorios, ya que es de nuestro conocimiento que muchos de los estudios relacionados al comportamiento de la naturaleza se abordan desde el punto de vista estadístico debido a que no obedecen a leyes determinísticas, es el caso de las señales electrofisiológicas, tales como las provenientes de un electroencefalograma o electrocardiograma, de los cuales también se ha recopilado material teórico disponible en gran cantidad desde la red. También se recolectó material bibliográfico relativo a estadísticos de orden superior, ya que las herramientas informáticas que van a procesar los datos adquiridos se basan en estas técnicas matemáticas.

3.3.2 Recolección de Datos:

En la recolección de datos referidos a la medición del ritmo cardiaco en pacientes sanos y pacientes con distintas patologías, se han tomado una cantidad determinada de archivos, que poseen registros digitalizados de la obtención señales, mediante instrumentos adecuados, provenientes de corazón de distintos pacientes a los cuales se les realizó el estudio electrocardiográfico, para ser analizados con la técnica **ICA**, estos archivos están disponibles en la red (www.physionet.org), que es un portal web que posee bases de datos



de señales electrofisiológicas en las que destaca los ECG, además de herramientas para el análisis de estas señales, tanto software para descargar como software en línea para analizar señales desde el propio navegador de internet, incluyendo rutinas para generar señales sintéticas, como los latidos del corazón, entre otros. Se ha adoptado esta manera de adquirir los datos, dado que se hace considerablemente complicado la recolección de datos directamente en centros especializados, y que posean los instrumentos adecuados de medición y digitalización de las señales de interés, además de la disponibilidad de un número adecuado de pacientes que se requieran para realizar los respectivos estudios.

3.4 Técnicas de procesamiento y análisis de datos

Para el manejo de datos y procesamiento de la información se trabajó una etapa de adecuación y otra de implementación.

3.4.1 Etapa de adecuación:

En esta etapa se construyó en el entorno de trabajo de MATLAB, algunos procedimientos para adecuar y tomar las señales de interés contenidas en los archivos con extensión .mat, y mostrar gráficamente sus características tales como forma de onda, PSD (Power Spectral Density, traducido al español sería: [densidad espectral de potencia](#)), valores estadísticos, entre otros. Además de reproducir copias de dicha señal desplazadas en el tiempo, un valor determinado, esto último con la finalidad de disponer de más de un canal u observación, debido a que el algoritmo fastICA necesita más de un canal para poder realizar el análisis de la componentes independientes.

3.4.2 Etapa de implementación:

Los análisis de los datos se realizaron mediante pruebas usando el paquete fastICA, el cual posee una interfaz gráfica, que aparece en pantalla al escribir la instrucción



“fasticag” en la ventana de comandos de matlab. Una vez cargado el archivo de la señal a estudiar en matlab, con sus acondicionamientos respectivos, se procede a cargar la matriz, que contiene las señales mezcladas, es decir, cada fila de esta matriz, es la señal tomada desde un canal, de esta manera existen otros canales que toman muestra de la misma señal pero desde otro punto físico, y estos van completando la cantidad total de filas con que cuenta la matriz. Al estar cargada la matriz de observación, se realizan ajustes en ella, y en otros parámetros propios del fastICA según sea conveniente, para finalmente obtener las señales “fuente” estimadas por esta toolbox, que serían en consecuencia, las señales deseadas para llevar a cabo sus análisis correspondientes, especialmente a la señal electrocardiográfica, que idealmente estaría libre de artefactos.

3.5 Metodología

Para el desarrollo de este proyecto se cumplieron una serie de fases, como se explica a continuación:

Fase I: Analizar los posibles artefactos que suelen estar presentes en la medición de señales electrofisiológicas.

Para analizar los diferentes artefactos presentes en las señales electrofisiológicas, se recopiló información, relacionada a los ruidos, para estudiarlas siguiendo el orden siguiente:

- Bosquejo sobre las características generales de las señales bioeléctricas.
- Documentación referida a fuentes de ruido e interferencia en un bioamplificador.
- Estudio sobre interferencias capacitivas.
- Estudio sobre interferencias inductivas.



- Recopilación de información referente a potencial de contacto electrodo-
piel.
- Documentación relacionada a interferencias provocadas por otros
potenciales bioeléctricos.
- Análisis sobre interferencias provocadas por otros sistemas fisiológicos.
- Estudio sobre interferencias provocadas por cargas electrostáticas.
- Obtención de información referente a fuentes internas de ruido

Fase II: Analizar y estudiar estadísticos de orden superior y medidas de independencia estadística.

En esta etapa se procedió realizar estudios referidos a variables estadísticas de orden superior, conceptos en los que se basa la técnica de ICA, cubriendo los siguientes temas:

- Estudio del método de blanqueo de características de las señales observadas.
- Información referente al proceso de maximización de la no-gaussianidad.
- Estudio y análisis de la medida de no-gaussianidad: kurtosis.
- Estudio y análisis de la medida de no-gaussianidad: neg-entropía

Fase III Analizar y comprender el método de Análisis de Componentes Independientes (Independent Components Analysis: “ICA”).

Mediante el uso de documentación referida al **Análisis de Componentes Independientes ICA**, se analizaron los siguientes aspectos:

- Fundamentos básicos del análisis de componentes independientes.
- Representación de los datos.
- Consideraciones estadísticas iniciales.



- Reducción de dimensión para la obtención de W .
- Uso de la independencia para la estimación de W .
- Separación ciega de fuentes.
- Observaciones de mezclas de señales desconocidas.
- Separación ciega basada en la independencia.
- El análisis de componentes principales.
- El análisis de componentes independientes.

Fase IV Analizar y estudiar las herramientas de software EEGLAB y FastICA.

Se obtuvo material referido al uso de estas herramientas, a través de tutoriales y asesoramientos de expertos en el uso de los mismos, como el tutor del presente trabajo, para llevar a cabo el análisis en el orden siguiente:

- Inicialización del toolbox fastICA desde la ventana de comandos de MATLAB.
- Exploración de la herramienta de ayuda del toolbox fastICA.
- Observación de las diferentes opciones de utilización del fastICA.
- Carga de datos de demostración (demosig), y realizar el análisis de componentes independientes a dichos datos.
- Inicialización del toolbox EEGLab desde la ventana de comandos de MATLAB.
- Exploración de la herramienta de ayuda del EEGLab.
- Observación de las diferentes opciones de utilización del EEGLab.
- Carga de datos de demostración y realizar estudios estadísticos a las señales de interés mediante esta herramienta.

Fase V Recopilar bases de datos de señales electrofisiológicas.

En esta etapa se obtuvo la base de datos a estudiar de la manera siguiente:



- Se accedió a la página oficial de physionet (www.physionet.org).
- Se obtuvo información referida a la misión y visión de este foro en línea.
- Siguiendo la ruta correspondiente, se acceso a la base de datos disponible.
- Mediante la herramienta de visualización en línea de las señales se fueron seleccionando las señales electrocardiográficas a estudiar.

Fase VI Extraer usando Análisis de Componentes Independientes ICA”, artefactos y obtener señales depuradas.

El procedimiento seguido para obtener las señales separadas mediante fastICA se realizó de la manera siguiente:

- Se cargó el primer archivo con extensión .mat, en el programa MATLAB, el cual contiene la forma de onda del estudio electrocardiográfico correspondiente.
- Se graficó la forma de onda anterior.
- A través de la toolbox EEGLab inicializada desde MATLAB se realizaron estudios estadísticos a la señal correspondiente a esa derivación.
- Mediante procedimientos ejecutados desde la ventana de comandos de MATLAB, se generó una señal de ruido y se añadió a la señal electrocardiográfica.
- Mediante procedimientos ejecutados desde la ventana de comandos de MATLAB, se generó una matriz adecuada a partir del vector que representa la señal electrocardiográfica más ruido.
- A través de la toolbox fastICA inicializada desde MATLAB, se cargó la matriz generada que representa el conjunto de señales electrocardiográficas.
- Obtención de los análisis (extracción de los componentes independientes) realizados por fastICA a la matriz.
- Obtención de los análisis estadísticos de los componentes independientes, realizados por EEGLab.



CAPÍTULO III. MARCO METODOLÓGICO



- Para cada señal electrocardiográfica se realiza el mismo procedimiento descrito anteriormente.
- Redacción de conclusiones y recomendaciones.



CAPÍTULO IV. RESULTADOS



CAPÍTULO IV.

ANÁLISIS DE RESULTADOS

Los resultados de este proyecto se detallarán en este capítulo, según las fases descritas en el capítulo III, correspondiente al marco metodológico.

4.1 Fase I: Estudiar los posibles artefactos que suelen estar presentes en la medición de señales electrofisiológicas.

Se puede destacar que el principal problema de la captación de las señales bioeléctricas proviene de los valores de amplitud y frecuencia de las mismas. Dichas amplitudes son pequeñas y a menudo se encuentran contaminadas de ruido que incluso puede ser superior al valor de la propia señal.

En resumen se muestran las principales fuentes de ruido que actúan sobre el registro de cualquier potencial bioeléctrico. Una posible clasificación de interferencias puede ser:

Interferencias externas al equipo de medida:

Capacitivas:

- _Acoplamiento capacitivo con el paciente.
- _Acoplamiento capacitivo con los conductores y el equipo.

Inductivas:

- _Originadas por la interfaz electrodo-electrolito-piel.
- _Debidas a otros potenciales bioeléctricos.
- _Debidas a otros sistemas fisiológicos.
- _Debidas a cargas electrostáticas.



Interferencias internas al equipo de medida:

- _Provocadas por el transformador de la fuente de alimentación.
- _Debidas al rizado de la fuente de alimentación.
- _Ruido generado por los componentes electrónicos.

Las interferencias externas son todas aquellas que tienen su origen fuera del equipo amplificador. La principal fuente de interferencia externa es, sin duda, la red de distribución de energía eléctrica, que provoca una señal de interferencia de 60 Hz, ya que las interferencias introducidas por otros equipos tienen generalmente frecuencias superiores al ancho de banda de la señal bioeléctrica. Ésta es la responsable de todas las alteraciones debidas por acoplo inductivo y capacitivo. Siendo el acoplo capacitivo de la red eléctrica con el cuerpo del paciente la mayor fuente de ruido presente en el registro de cualquier potencial bioeléctrico.

No obstante, conviene tener presente posibles interferencias o artefactos introducidos por otros biopotenciales como por ejemplo los artefactos introducidos por el sistema respiratorio en el registro del ECG. Estos artefactos pueden minimizarse variando la colocación de los electrodos, de forma que éstos se posicionen en lugares donde su influencia sea mínima.

4.2 Fase II: Estudiar y comprender estadísticos de orden superior y medidas de independencia estadística.

De acuerdo a lo estudiado, se puede determinar fácilmente que las estadísticas de alto orden son más robustas que las estadísticas convencionales, por lo cual se están empleando cada vez más en los estudios de fenómenos o sistemas. Su ventaja principal es dar más información de un proceso usando la misma secuencia de datos. El uso de los cumulantes da ventajas sobre el uso de los momentos por sus características de simetría, inmunidad al ruido y su uso como operador lineal. Este tipo de estadísticas son las que se emplearán en el desarrollo de las herramientas de procesamiento de señales electrofisiológicas.



4.3 Fase III estudiar y comprender el método de Análisis de Componentes Independientes (Independent Components Analysis: “ICA”).

En esta fase de la investigación se dedicó a estudiar los fundamentos del Análisis de Componentes Independientes, y se ha visto la importancia que tiene el problema de la Separación Ciega de Fuentes, necesaria comprender el funcionamiento de ICA.

Luego se comprobó como PCA es una herramienta estadística de vital importancia para la reducción de las dimensiones del problema y a la extracción de características de los datos. En este mismo apartado, se observó el funcionamiento del proceso de blanqueado, que será de gran utilidad como preprocesado antes de realizar el análisis ICA y que permitirá que las observaciones resulten incorrelacionadas y con varianza unidad.

Por último realizó un estudio inicial del Análisis de Componentes Independientes, en el que se definió el modelo básico y se realizaron consideraciones referentes a las restricciones que se han de plantear y a las ambigüedades que presenta dicho modelo. Seguidamente vimos que la decorrelación no es suficiente para poder realizar la separación de las componentes independientes, por lo que se hace necesario recurrir a criterios de búsqueda basados en la no gaussianidad. Mediante el estudio de los estadísticos de orden superior (kurtosis), se llegó a un criterio que permitía establecer una clasificación de las variables aleatorias según su gaussianidad.

Por último se destacó como era posible estimar la matriz ICA mediante la conocida “Teoría de la Información”, cuyo fundamento consiste en minimizar la información mutua entre las salidas de dicha matriz.

Con todo, se puede extraer las siguientes conclusiones:

_ La decorrelación no va a garantizar el éxito en el proceso de estima de las componentes independientes.

_ Las componentes independientes son aquellas que son máximamente gaussianas.

_ Una forma alternativa de calcular la matriz de separación W se puede fundamentar en la utilización de técnicas basadas en la Teoría de la Información.

4.4 Fase IV conocer y estudiar las herramientas de software EEGLAB y FastICA.

El paquete EEGLAB.

EEGLAB es un conjunto de herramientas que trabaja bajo la plataforma de MATLAB, el cual, cuenta con una interfaz gráfica interactiva que permite a los usuarios procesar EEG continuo, EEG relacionados a eventos y otros datos electrofisiológicos usando el Análisis de Componente Independiente y/o el Análisis de Tiempo Frecuencia, así como métodos de promediación como se muestra en la figura 4.1.

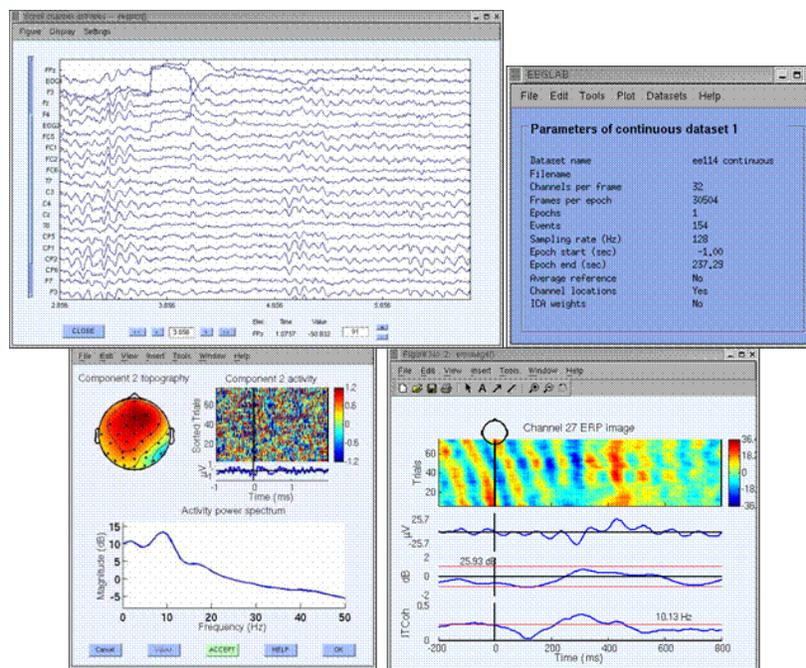


Figura 4.1 Aplicaciones realizadas por EEGLAB. [33]

La colección de funciones disponibles en EEGLAB, permite el procesamiento de la señal, la visualización de los datos de EEG y el mapeado de la imagen de Potenciales Relacionados a Eventos, entre otras cosas. Con el pre-procesamiento de la señal, se pueden filtrar los datos, extraer épocas (una época marca de tiempo para un intervalo de medición o frecuencia de datos, por ejemplo, 10 segundos), eliminar la línea de base y remuestrear los datos. EEGLAB también hace posible que los usuarios puedan eliminar canales de datos, épocas, y/o componentes formadas principalmente por artefactos oculares, musculares y de



la línea de ruido, aceptando o rechazando visualmente las recomendaciones de EEGLAB que son basadas en medidas del procesamiento e información de la señal.

El mapeo de la imagen de PRE (Potenciales Relacionados a Eventos) [33], es una representación de los datos en dos dimensiones, donde se muestran las variaciones de potenciales registrados en un solo canal en todas las épocas a través del tiempo.

Una herramienta esencial de EEGLAB es el Análisis de Componentes Independientes de los datos de EEG, el cual, ha probado ser capaz de aislar o extraer las fuentes generadas por EEG, bajo la suposición de que sus actividades son temporalmente independientes o al menos más temporalmente independientes que cualquier combinación lineal de sus señales. ICA ha sido ampliamente utilizado en la comunidad de investigación de EEG para detectar y quitar los artefactos no deseados.

Una aplicación más de EEGLAB, es el Análisis de Tiempo/Frecuencia, el cual, determina la amplitud espectral relacionada a eventos, las perturbaciones de fase y coherencia en los datos registrados de simples electrodos y/o componentes de ICA.

Para los usuarios experimentados de MATLAB, EEGLAB ofrece un ambiente de programación estructurado para almacenar, tener acceso, medir, manipular y visualizar los datos de EEG relacionados a eventos. Y para los programadores investigadores y desarrolladores creativos de métodos, EEGLAB posee una extensa plataforma de fuente abierta a través de la cual se pueden compartir nuevos métodos con la comunidad de investigación mundial contribuyendo con funciones de EEGLAB que aparezcan automáticamente en el menú de EEGLAB.

Para abrir el EEGLAB, se escribe en la ventana de comandos de MATLAB, la siguiente instrucción:

```
>>eeglab
```


>>fasticag

Luego aparece la siguiente interfaz gráfica mostrada en la figura 4.3.

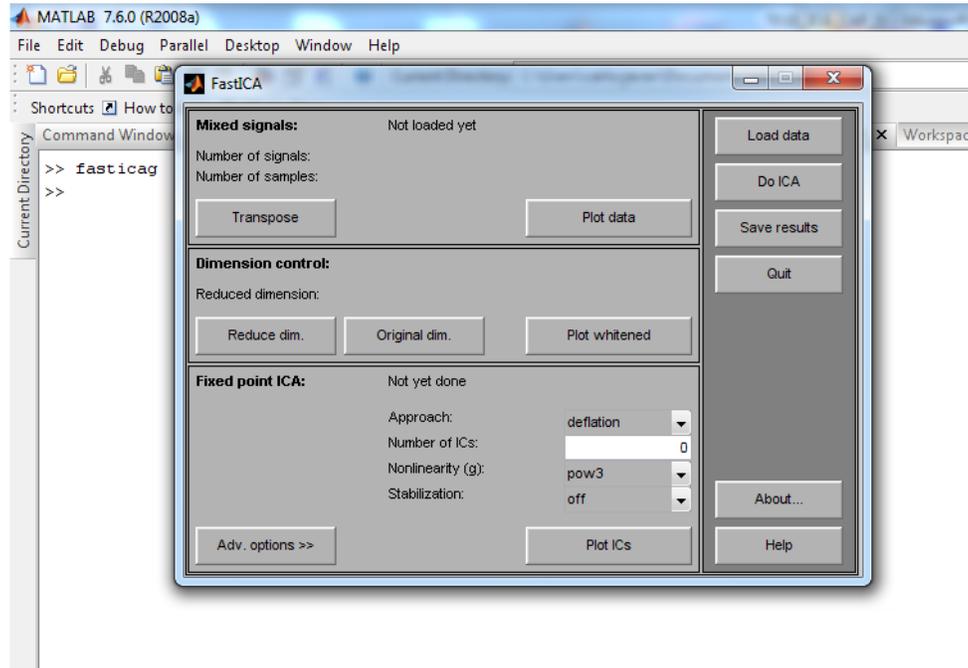


Figura 4.3 Interfaz gráfica de fastICA. [Fuente Propia]

Para ilustrar el uso que se le dió a las herramientas EEGLAB y fastICA, se llevó a cabo el siguiente procedimiento, que es el mismo que se le aplicó a cada una de las señales en estudio, que se trató en la fase VI.

Una vez abierto el programa MATLAB y ambos paquetes se carga el archivo, desde la opción file del menú principal, seguidamente se hace click en Import Data, como se muestra en la figura 4.4.

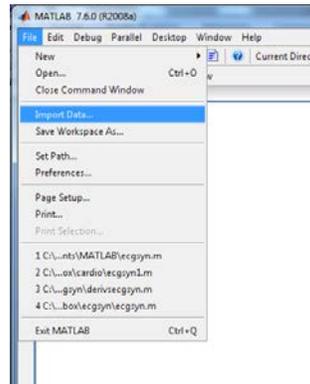


Figura 4.4 Primer paso para el acceso al archivo que contiene la señal a estudiar. [Fuente Propia]

Luego se abre la ventana para ubicar el archivo, que en este caso es “onda cardiaca sintética.mat” (ver figura 4.5).

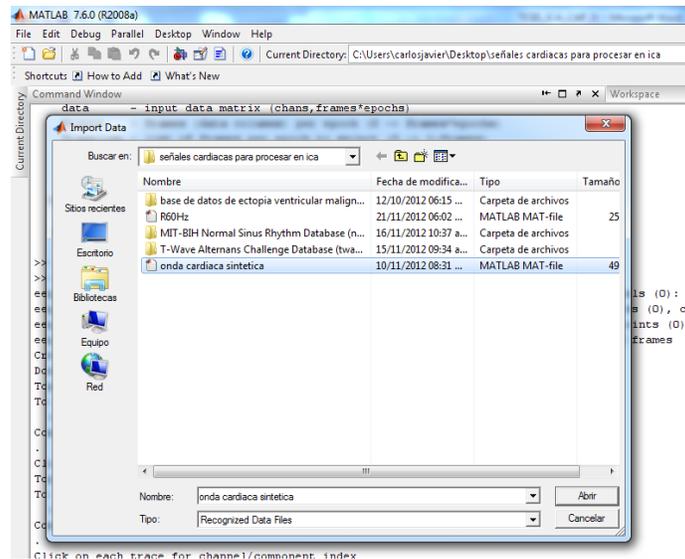


Figura 4.5 Selección del archivo “onda cardiaca sintética”. [Fuente Propia]

Así se le llama a este archivo, ya que pertenece a una onda idealizada, creada con una rutina que funciona en MATLAB, disponible también en la página de physionet.

Para observar la forma de onda de esta señal, se escribe en la ventana de comandos de MATLAB, la siguiente instrucción:



>>plot(s)

Donde “s” es el nombre de la variable (arreglo tipo vector fila), que contiene las lecturas que forman a la señal ECG sintética, con una frecuencia de muestreo de 256 Hz (valor por defecto) y su forma de onda es la que se muestra en la figura 4.6, donde el tiempo leído en segundos, es la lectura en el eje horizontal de la gráfica dividido entre la frecuencia de muestreo (256Hz).

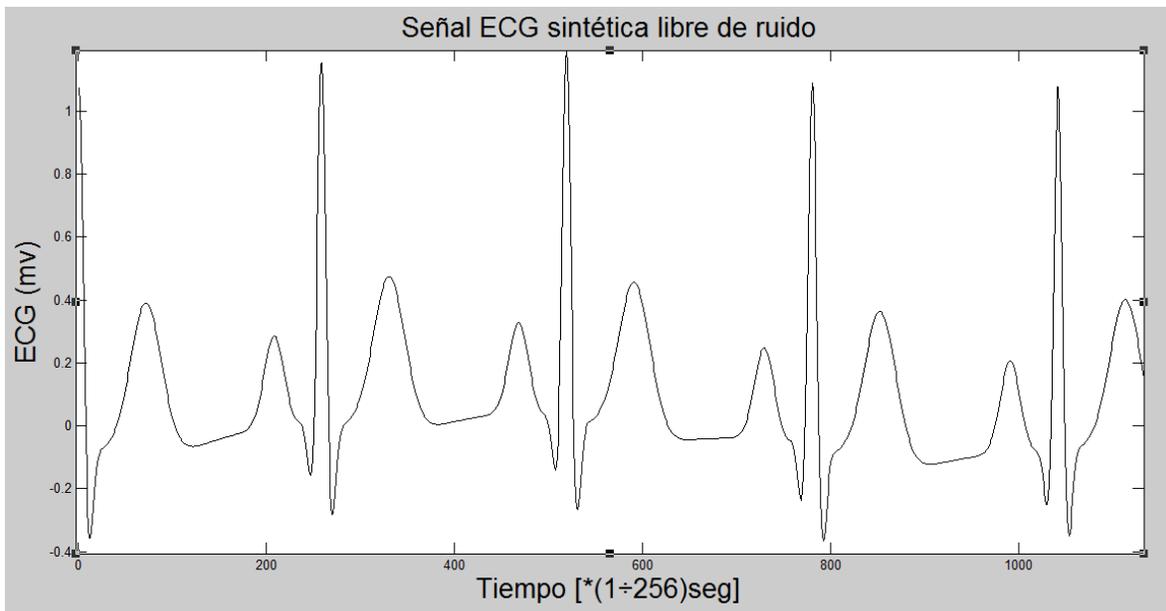


Figura 4.6 Forma de onda de la señal ECG sintética, sin ruido alguno. [Fuente Propia]

A continuación se procedió observar el espectro de potencia de la señal anterior usando EEGLAB de la manera siguiente:

Se escoge la opción que se muestra en la figura 4.7.



CAPÍTULO IV. RESULTADOS

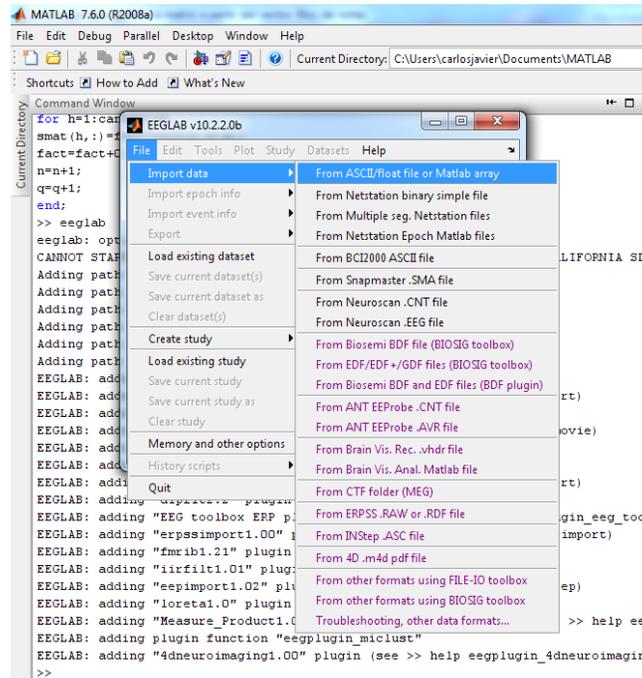


Figura 4.7 Se selecciona el tipo de arreglo a cargar en este caso un arreglo MATLAB. [Fuente Propia]

Se procedió a cargar el vector con la señal ECG sintética en este caso, se llama `s` y se escogió una frecuencia de muestreo de 256 Hz, luego se pulsa “ok”, (ver figura 4.8).

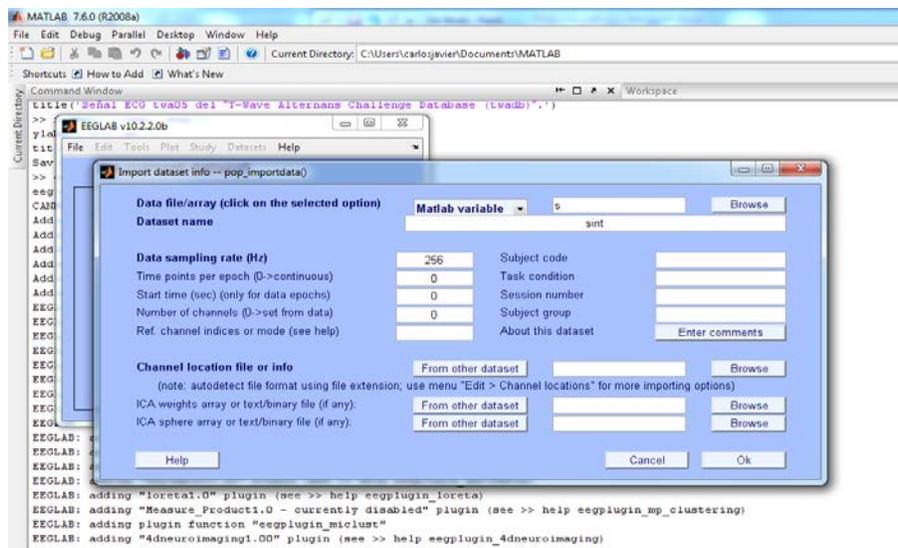


Figura 4.8 Colocando el nombre de la señal a estudiar. [Fuente Propia]



CAPÍTULO IV. RESULTADOS



En la figura 4.9 se observan algunas características de la señal, descritas por EEGLAB.

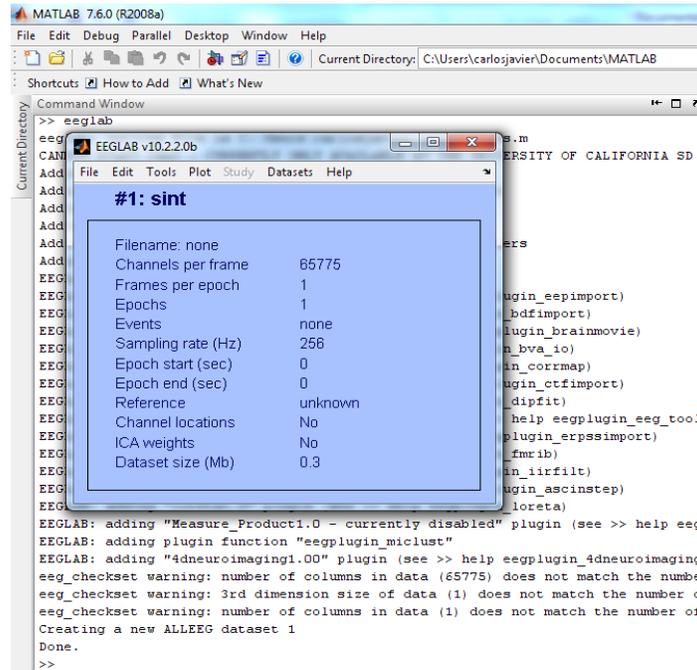


Figura 4.9 Descripción de las características de la señal cargada. [Fuente Propia]

Siguiendo la ruta mostrada en la imagen siguiente se accesa a la ventana que nos permite obtener la gráfica del espectro de potencia de la señal (ver figura 4.10).

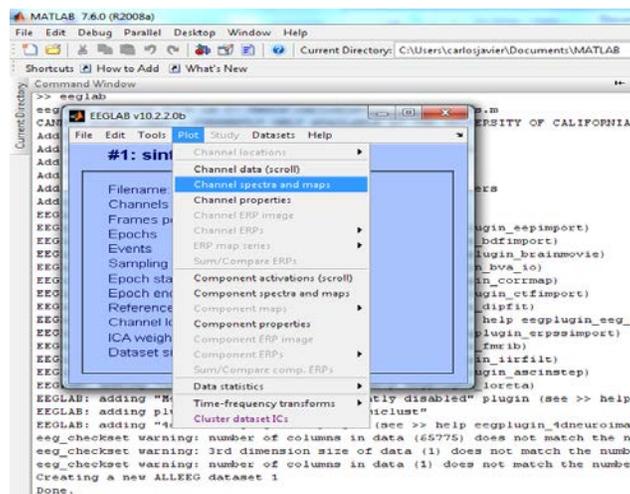


Figura 4.10 Ruta para acceder a la opción de graficación de la densidad espectral de potencia de la señal. [Fuente Propia]

Aquí se pulsa “ok” dejando los parámetros por defecto (ver figura 4.11)

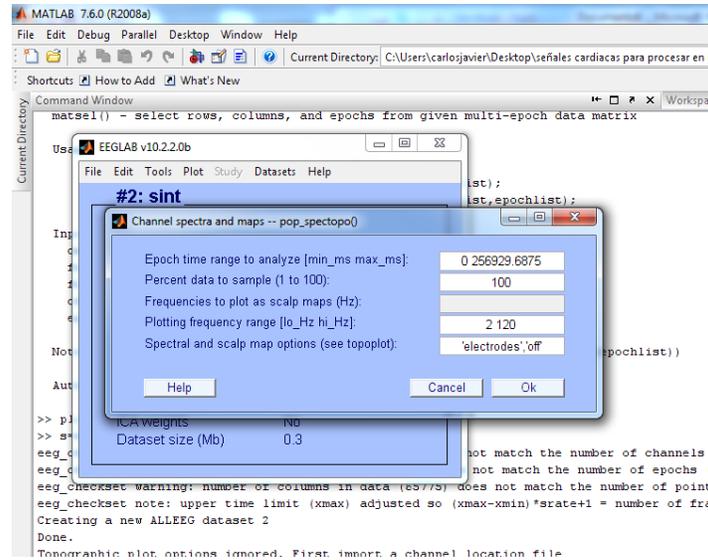


Figura 4.11 Opciones para mostrar la grafica de la densidad espectral de potencia de la señal ECG. [Fuente Propia]

Podemos apreciar la densidad espectral de potencia de la señal ECG sintética en la siguiente figura.

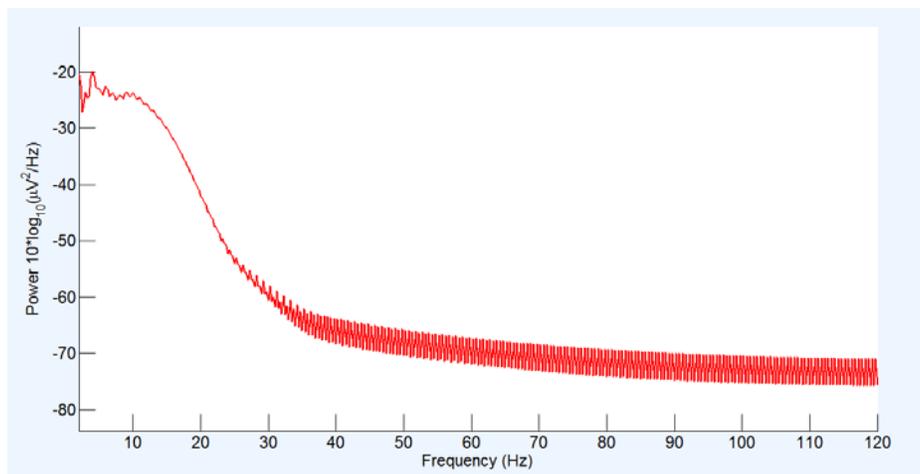


Figura 4.12 Densidad espectral de potencia de ECG sintética. [Fuente Propia]

Seguidamente se procede a observar el estudio estadístico realizado a la señal cardiaca ideal, la imagen mostrada a continuación muestra la ruta de acceso a esta opción de EEGLAB (ver figura 4.13).

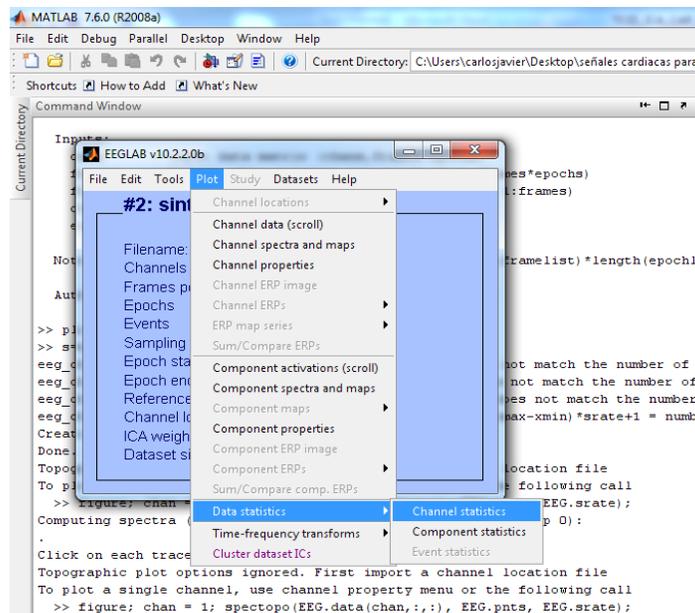


Figura 4.13 Ruta para la obtención de los estudios estadísticos de ECG sintética. [Fuente Propia]

Como en este caso existe un solo canal, solo se pulsa ok, ya que por defecto está el número 1 colocado (ver figura 4.14).

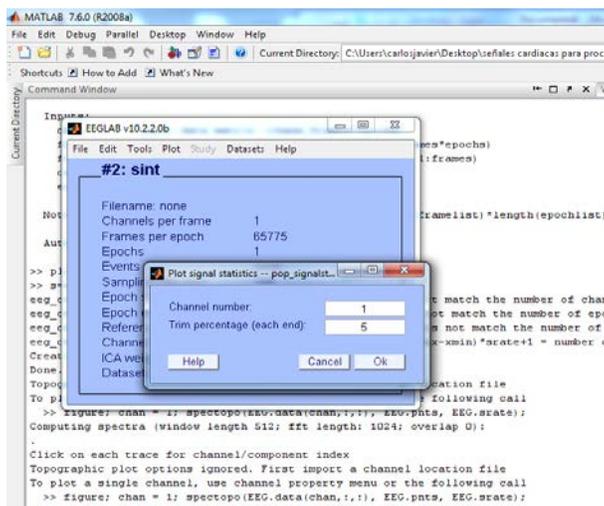


Figura 4.14 Selección del canal a estudiar, en este caso es una sola señal (ECG sintética). [Fuente Propia]

Finalmente se obtiene la imagen de los resultados de los estudios estadísticos de la señal ECG sintética sin presencia de ruido (ver figura 4.15).

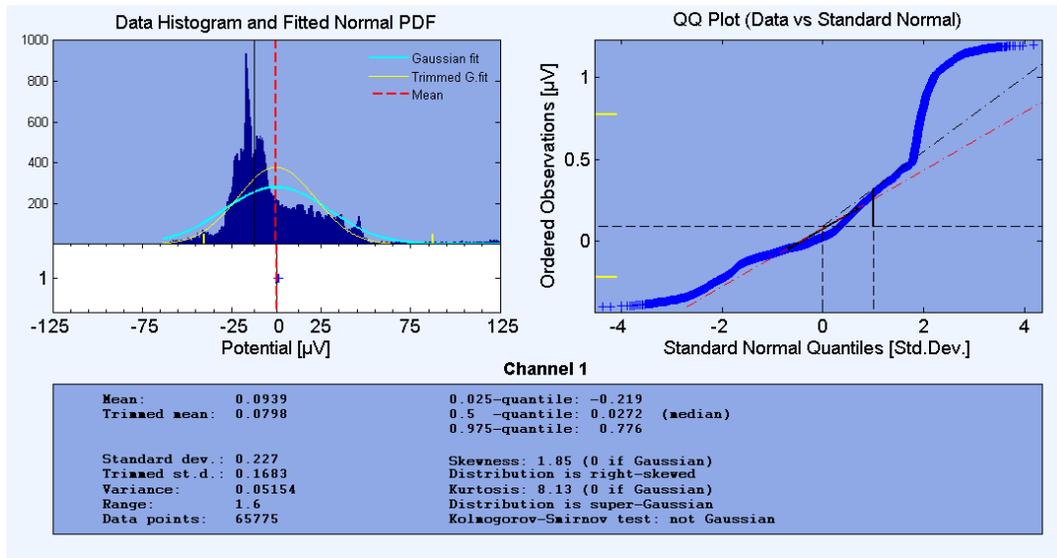


Figura 4.15 Estadísticas de la señal ECG sintética, antes de agregarle ruido. [Fuente Propia]

Aplicando el mismo procedimiento descrito para la onda sintética, se carga el registro que representa la onda de ruido de 60Hz, y se obtiene la forma de onda de la misma, como se muestra en la figura 4.16:

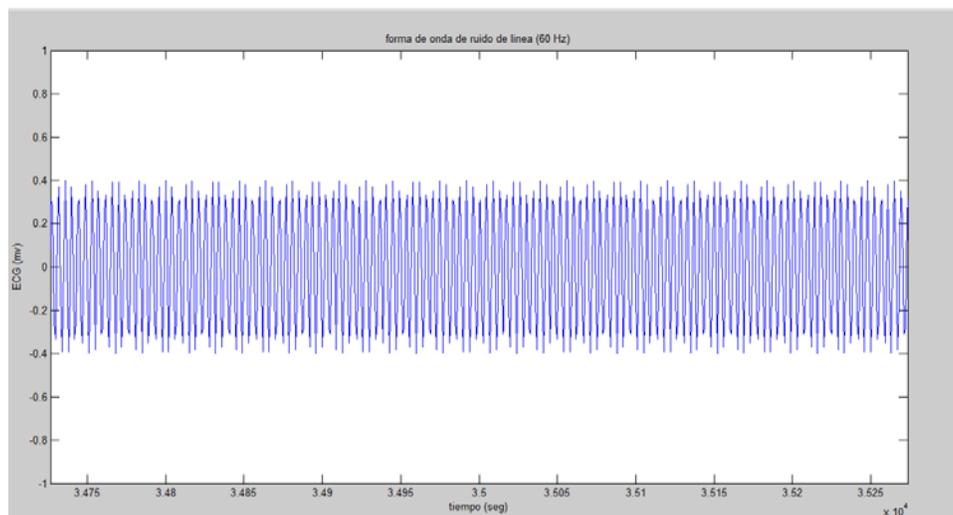


Figura 4.16 Onda que representa la señal de ruido de línea (60 Hz). [Fuente Propia]



El nombre de la variable (arreglo tipo vector fila), que contiene las lecturas que forman a la señal ECG sintética, es “hum”, por lo tanto, si llamamos “s_hum” a la variable que resulta de la suma de las variables “s” y “hum”, dicha variable representa la onda ECG con ruido de línea, entonces procedemos a realizar esta operación en la venta de comandos de MATLAB como sigue:

```
>>s_hum=s+hum;
```

Y para obtener la forma de onda de esta variable resultante, colocamos la siguiente instrucción:

```
>>plot(s_hum)
```

Y aparece la forma de onda mostrada en la figura 4.17.

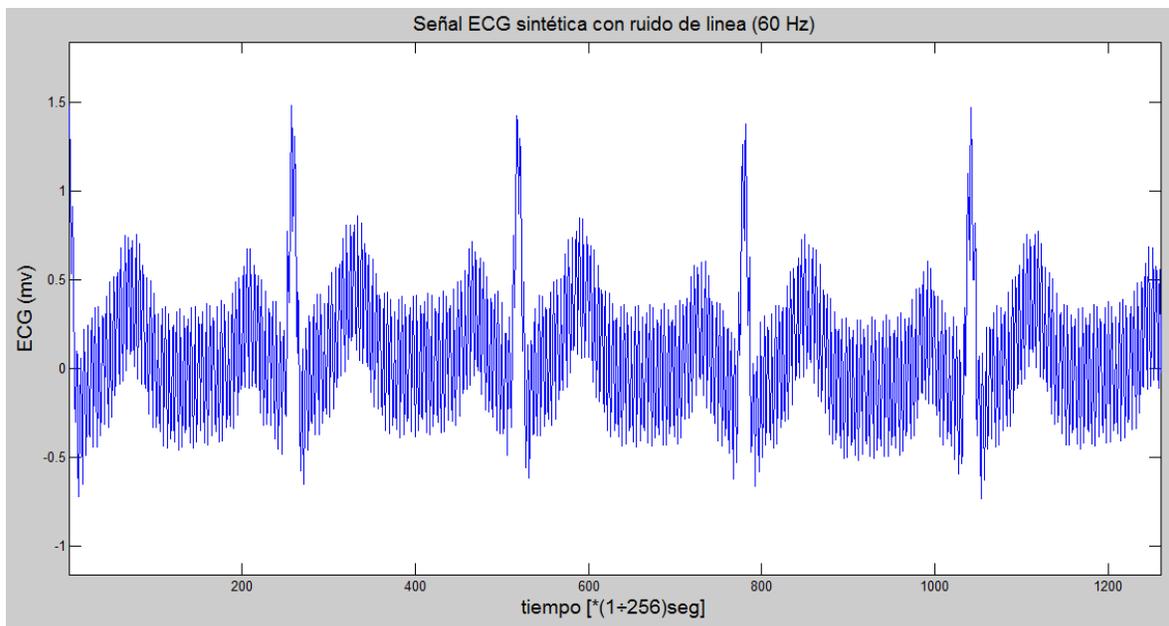


Figura 4.17 Señal ECG con ruido de línea de 60 Hz. [Fuente Propia]

A continuación, se realizaron los estudios a esta última señal, repitiendo los pasos que se aplicaron a la señal ECG idealizada con el fin de observar las diferencias existentes entre ambas señales a nivel estadístico y de su espectro de potencia.

Se obtiene la Gráfica de la densidad espectral de potencia de la señal que ya posee ruido de línea, como se observa en la figura 4.18.

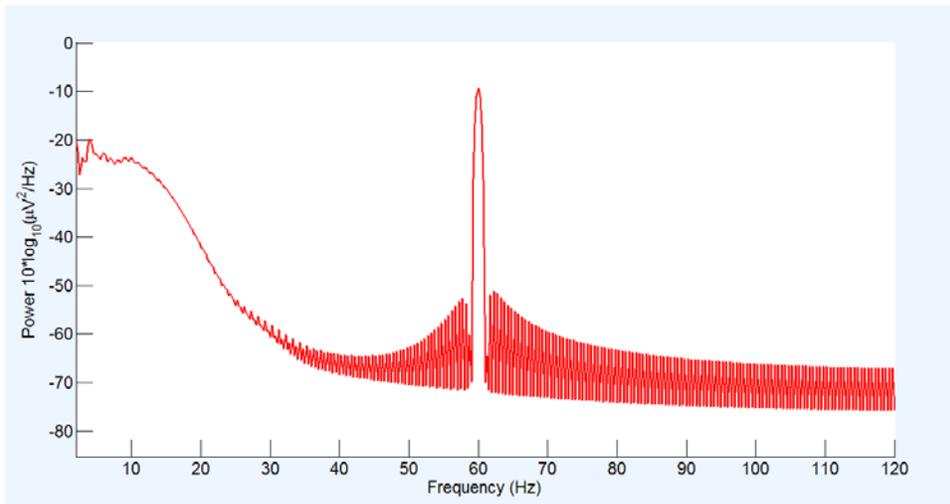


Figura 4.18 Densidad espectral de potencia (PSD) de ECG sintética con ruido. [Fuente Propia]

Obtención de las Estadísticas de la señal ECG ideal con ruido de línea (figura 4.19).

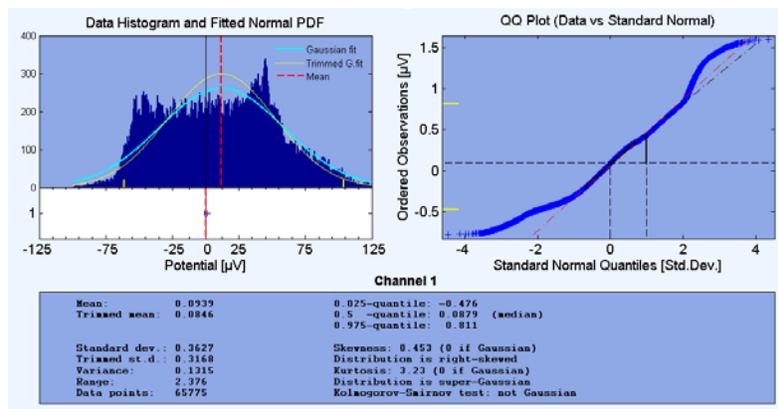


Figura 4.19 Estadísticas de la señal ECG sintética, después de agregarle ruido. [Fuente Propia]



Como ya es sabido, el programa fastICA necesita una matriz, donde cada fila de esa matriz representa un canal, que toma una muestra de la mezcla de las señales desde un punto de referencia particular, esta condición representa un inconveniente para nuestra investigación, ya que solo se cuenta con un canal que es el mostrado anteriormente.

Una manera de solventar este inconveniente fue, formar una matriz donde su primera fila, es el vector de la onda a estudiar, su segunda fila es el mismo vector, pero este empieza desde el segundo valor del vector anterior, y se completa con el primer valor del mismo, la tercera fila empieza desde el tercer valor del vector de la primera fila, y se completa con los dos primeros valores del primer vector, así sucesivamente, hasta completar el número de filas que se fijó previamente. Para ilustrar este artificio, se muestra el siguiente ejemplo:

Sea el vector $v=[1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ 13\ 14\ 15\ 16\ 17\ 18\ 19\ 20\ 21\ 22\ 23\ 24]$.

Y queremos formar la matriz M de 5 filas, de la manera descrita anteriormente, entonces procedemos así:

Primera fila: [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24]
Segunda fila: [2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1]
Tercera fila: [3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1 2]
Cuarta fila: [4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1 2 3]
Quinta fila: [5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1 2 3 4]

El procedimiento en MATLAB para generar esta matriz se muestra a continuación:

```
final_s=size(s);  
if final_s(2)<=1  
s=s';
```



```
end;
final_hum=size(hum);
if final_hum(2)<=1
hum=hum';
end;
final_s=size(s);
s_hum=s+hum;
vector=s_hum;
cantfilas=5;
q=final_s(2);
n=1;
u=1;
for h=1:cantfilas
for j=1:q
smat(h,j)=vector(u);
u=u+1;
if u>q
u=1;
end;
end;
u=h+1;
end;
```

Las primeras instrucciones del procedimiento anterior, se realizan para garantizar que las señales sean vectores fila, “smat” es la matriz generada.

Procedemos a cargar esta matriz en el fastICA, luego de haber iniciado la misma con la instrucción “fasticag”, en la ventana de comandos de MATLAB como se muestra en la figura 4.20.

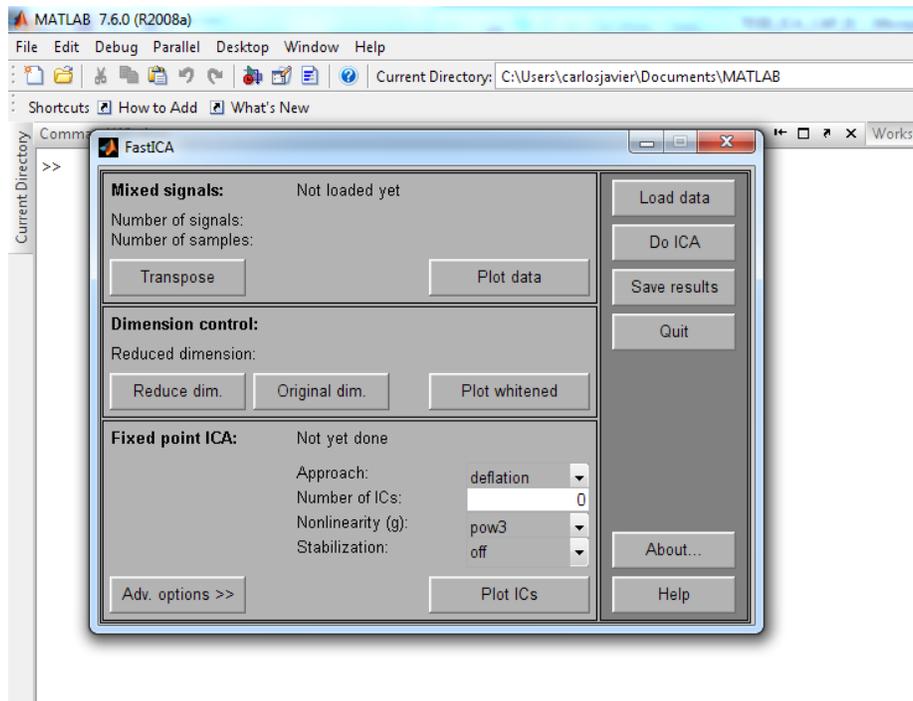


Figura 4.20 Interfaz gráfica de fastICA lista para cargar la matriz de observaciones. [Fuente Propia]

Se pulsa la opción “Load data” y se escribe en el campo disponible en la ventana, el nombre de la matriz de observaciones luego se pulsa “Load” (ver figura 4.21).

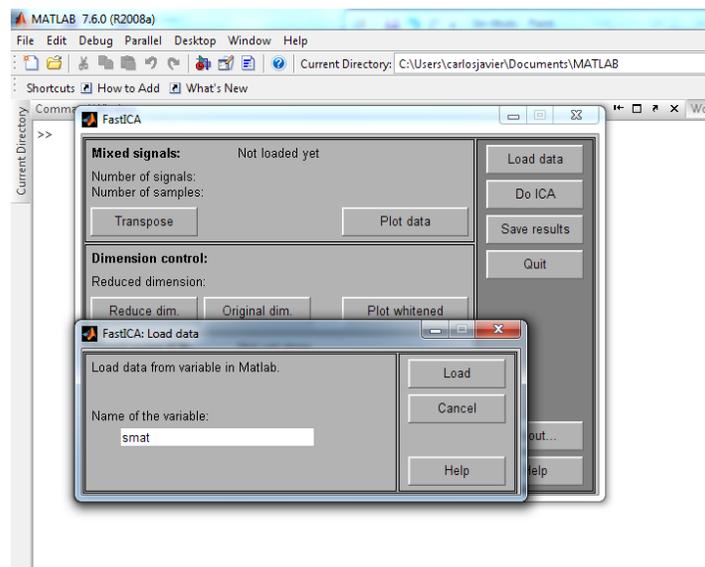


Figura 4.21 Cargando la matriz “smat” en el fasticag. [Fuente Propia]

En la figura 4.22 se muestra a la matriz ya cargada donde se puede observar las diferentes opciones de tratamiento de la matriz antes de realizarle el análisis de componentes independientes.

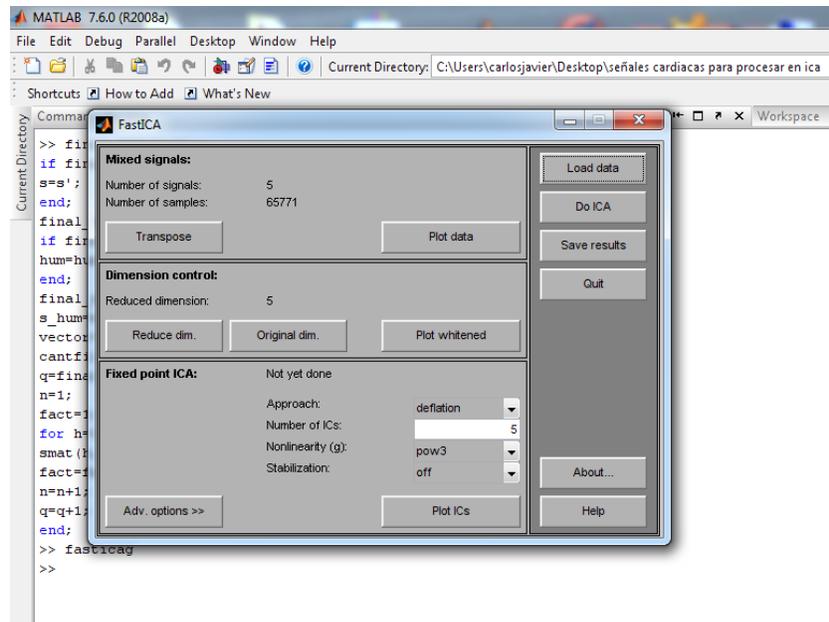


Figura 4.22 Matriz “smat” ya cargada. [Fuente Propia]

En la región denominada "mixed signals" se puede observar cuantas filas y cuantas columnas posee la matriz que se ha cargado, ya que el número de señales (Number of signals), representa el número de filas de la matriz, mientras que el número de muestras (Number of examples), representa el número de columnas de la matriz.

En esa misma región, está una opción llamada “Transpose”, esta se utiliza cuando las señales vienen ordenadas por columnas, de esta manera con solo darle click en ese botón, se transpone la matriz, para poder ser procesada. La opción “Plot data”, sirve para observar las señales provenientes de los canales, sin procesar, en este caso las 5 señales de “smat” (ver figura 4.23).

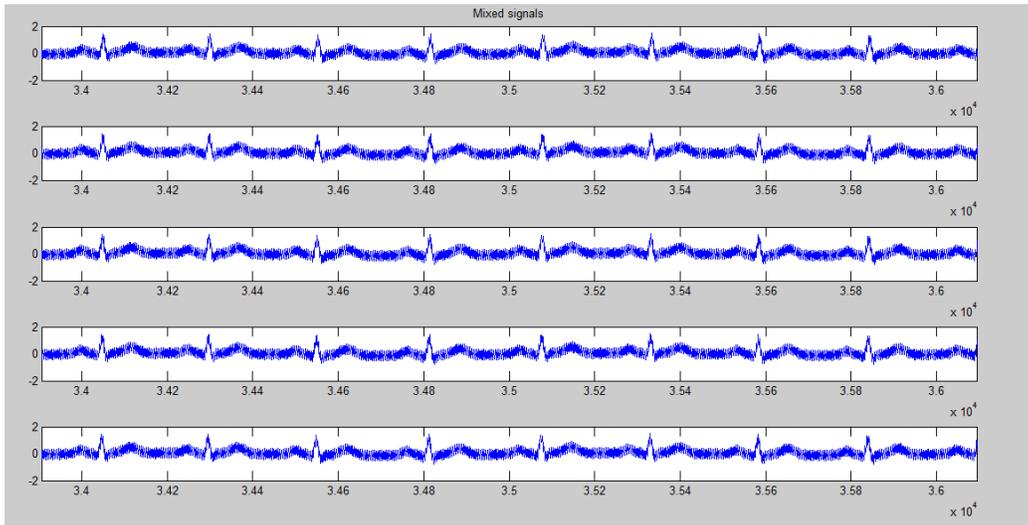


Figura 4.23 Señales sin procesar contenidas en la matriz de observaciones, obtenida de la opción Plot data. [Fuente Propia]

En la región denominada “Dimension control”, se encuentran tres botones, llamados “Reduce dim”, “Original dim.” y “Plot whitened”, respectivamente. En “Reduce dim”, hacemos click, y aparece el recuadro mostrado en la figura 4.24.

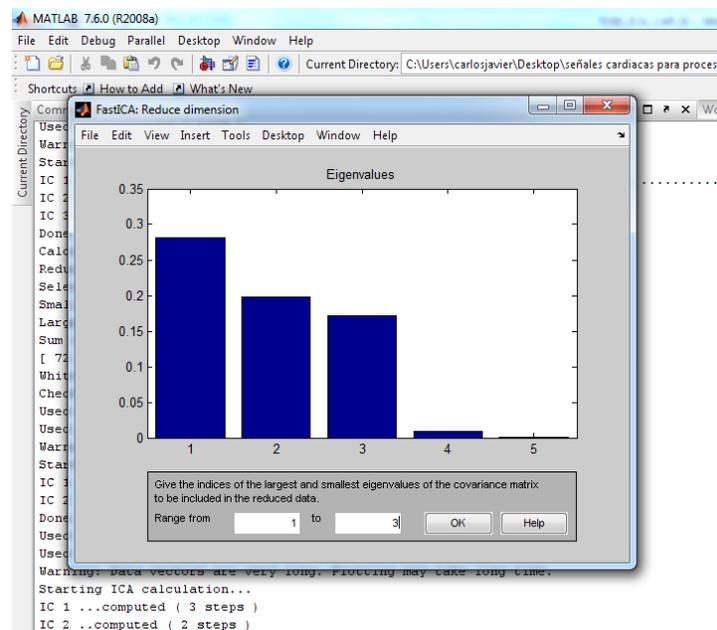


Figura 4.24 Reducción de la dimensión de la matriz. [Fuente Propia]

La selección del subespacio correspondiente a ciertos valores propios de la matriz de covarianza de los datos. Estos valores propios están ordenados de forma decreciente, de manera de desechar los más pequeños a la hora de reducir la dimensión de la matriz, ya que estos no aportarían tanta información, como los de mayor valor. En este caso se eligió reducir a tres dimensiones, ya que aparecen con mayor tamaño. Si se quiere recuperar el número total de dimensiones de la matriz, se da click al botón “Original dim.”, con el botón “Plot whitened”, se obtienen los datos blanqueados, de manera que queden lo más decorrelacionadamente posible, esto es necesario para que el análisis de componentes independientes, sea lo más certero posible (ver figura 4.25).

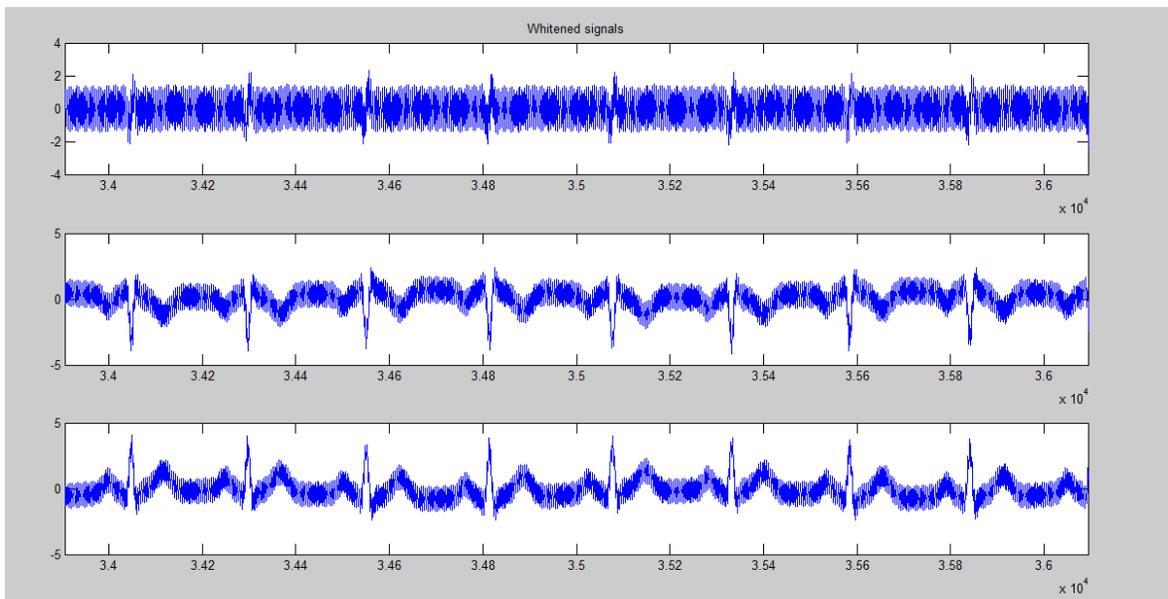


Figura 4.25 Señales blanqueadas. [Fuente Propia]

En la región llamada “fixed point ICA”, aparecen otras opciones de ortogonalización y decorrelación no lineal, que normalmente se dejan sus valores por defecto. Luego al dar click en el botón “Do ICA” o el botón “Plot ICs”, aparecen las componentes independientes (ver figura 4.26). En esta figura se puede apreciar que la gráfica superior corresponde a la señal cardíaca sintética, siendo esta, una de las componentes independientes.

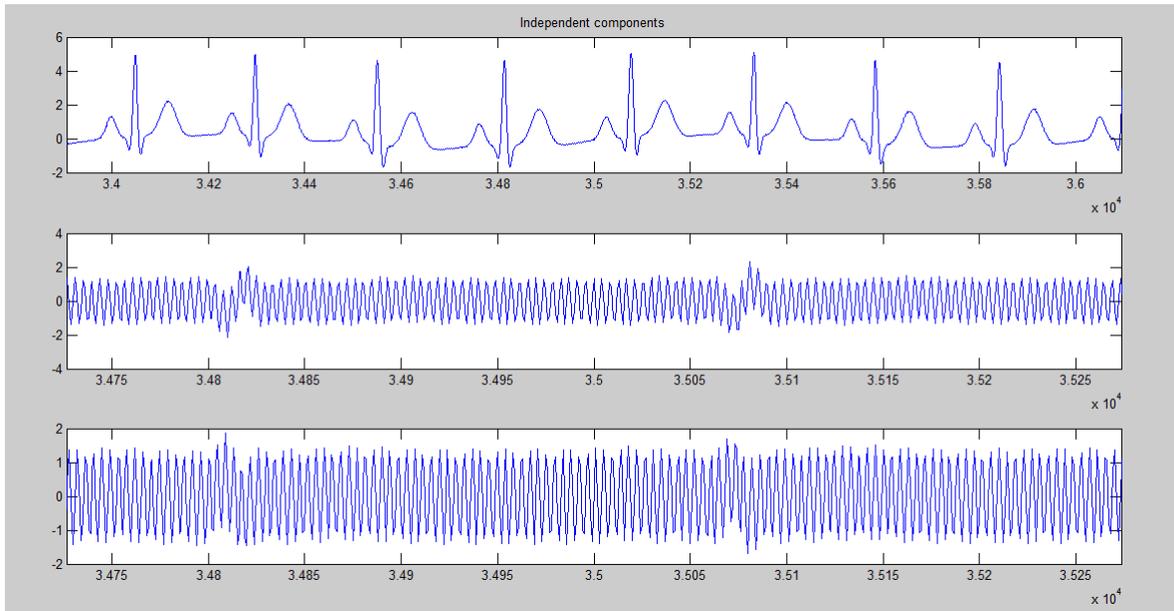


Figura 4.26 Componentes independientes obtenidos por fastICA. [Fuente Propia]

Para guardar los resultados obtenidos se hace click en el botón llamado “Save results” (ver figura 4.27).

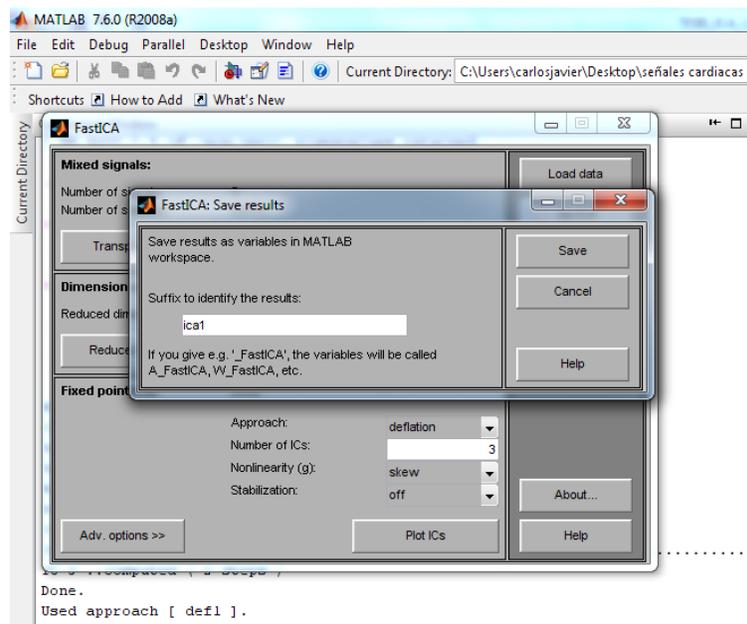
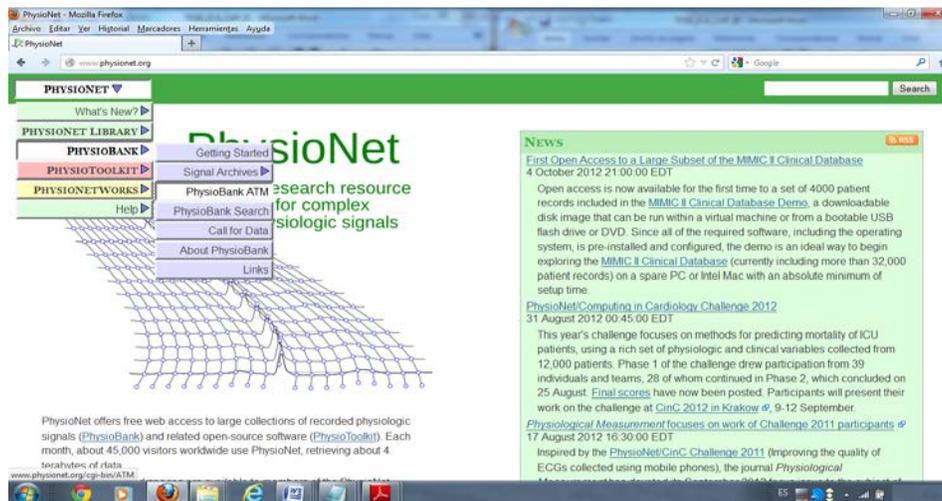


Figura 4.27 Guardando los resultados en el Work space [Fuente Propia]

4.5 Fase V: Recopilar bases de datos de señales electrofisiológicas.

Los datos que representan las señales en estudio, fueron tomados de la base de datos de physionet (<http://www.physionet.org/>), siguiendo la ruta de acceso PHYSIONET > PHYSIOBANK > PhysioBank ATM, tal como que se muestra en la figura 4.28.



(a)



(b)

Figura 4.28 (a) página principal de physionet donde se muestra la ruta de acceso a la base de datos. (b) ventana donde se seleccionan los datos que se deseen estudiar. [Fuente Propia]



En la figura 4.28 (b), se observa la pestaña identificada como “database”, aquí se selecciona la base de datos de la cual se desea obtener las señales de interés.

A continuación, se muestra el procedimiento para seleccionar un registro de una señal electrocardiográfica en particular, en este caso se escogió desde la pestaña “data base”, la base de datos [MIT-BIH Normal Sinus Rhythm Database \(nsrdb\)](#), seguidamente se seleccionó el registro llamado 16539, que se puede visualizar en la pestaña llamada “record”, luego en la siguiente pestaña con el nombre “signal”, se escoge la derivación que se quiere observar, los parámetros anteriormente descritos, pertenecen área de entrada (input), el área de salida (output) posee los siguientes parámetros: Length, Time format, y Data format, que sirven para ajustar convenientemente la visualización de la onda. En el área llamada toolbox se puede escoger la acción a llevar a cabo, bien sea graficar la onda, exportar como archivo con un tipo de extensión determinada (que en nuestro caso es .mat), o mostrar en formato de texto, información relacionada a la señal, la figura 4.29 ilustra la descripción anterior.

Posteriormente, se procede a descargar el archivo que contiene los datos que representan la señal observada con anterioridad, para llevar a cabo este paso, se selecciona en el área toolbox, la opción “Export signals as .mat”, a continuación al lado de la opción “**Download**”, está el nombre del archivo que desea descargar, al cual se le pulsa con el botón primario del mouse para iniciar la descarga (en la figura 4.30 se observan estas opciones), el archivo se descargará en un directorio por defecto desde el navegador a la PC.



CAPÍTULO IV. RESULTADOS

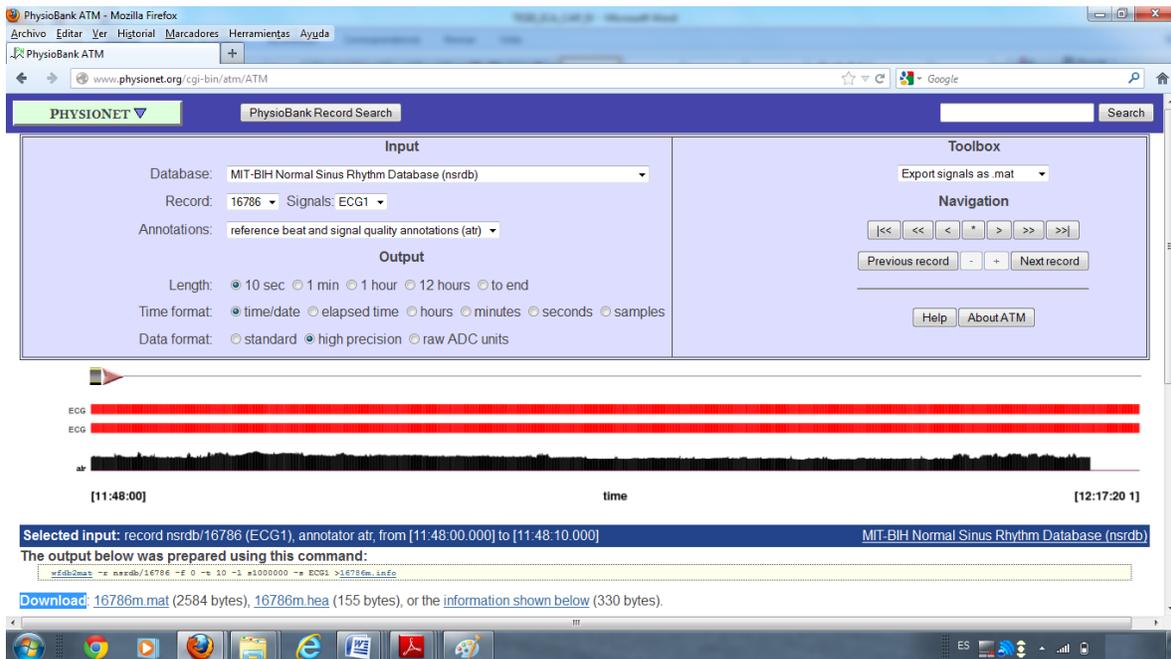


Figura 4.30 Descarga del archivo con extensión .mat que representa la señal electrocardiográfica, es un arreglo tipo vector, el cual posee las magnitudes leídas continuamente por el electrocardiógrafo al cabo de cierto tiempo. [Fuente Propia]

De esta manera se realizaron una serie de descargas de archivos, con el fin de aplicarles el análisis de componentes independientes (ICA), con la herramienta fastICA.

4.6 Fase VI: Extraer usando “Análisis de Componentes Independientes ICA”, artefactos y obtener señales depuradas.

Cabe destacar que las bases de datos disponibles en <http://www.physionet.org/>, ya han sido tratadas en cuanto a supresión de ruido, por lo que no sería factible el uso de las mismas a efectos de aplicarles **ICA**, sin agregarle ruidos típicos a los que normalmente estarían expuestas estas señales, por lo tanto se hace necesario agregarles estos ruidos, que en este caso se usarán uno llamado de línea (de 60Hz) y otro aleatorio, ambos creados



desde MATLAB. Se realizaron varios experimentos con sus respectivos resultados mostrados a continuación:

Experimento 1:

Registro 16265 del “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”.

Para este experimento y los siguientes, se aplicaron los pasos descritos a continuación:

Paso 1: Carga del archivo de la señal electrocardiográfica en MATLAB y graficación de su forma de onda (Ver figura 4.31)

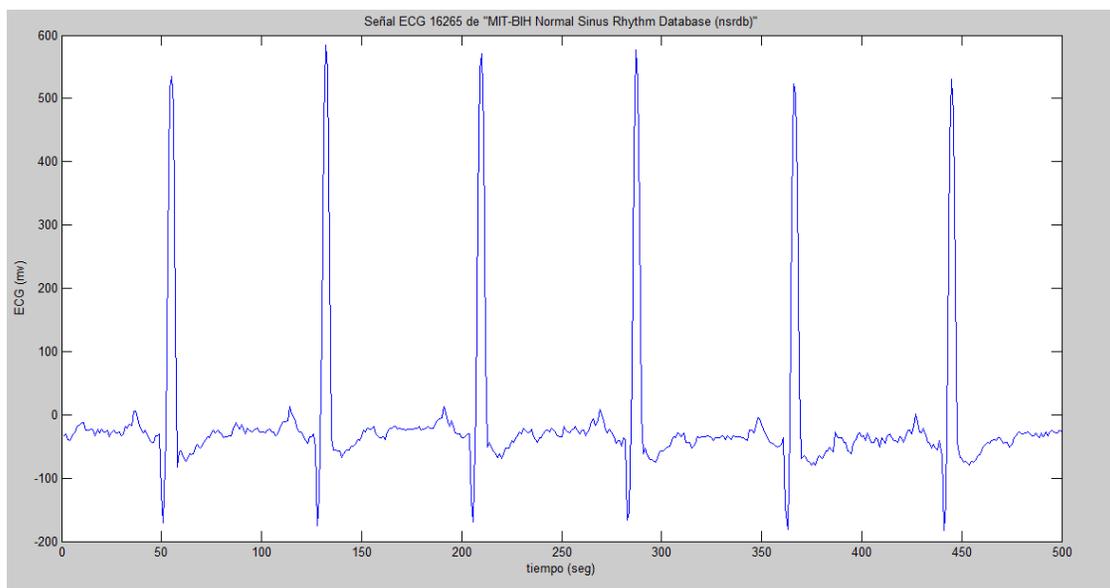


Figura 4.31 Forma de onda de la señal ECG 16265 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia]

Paso 2: Mediante la Utilización de toolbox EEGLab, aplicando los conocimientos obtenidos en la fase 4, para el uso del mismo, se obtiene la siguiente gráfica de la densidad espectral de potencia de la señal anterior (Ver figura 4.32)

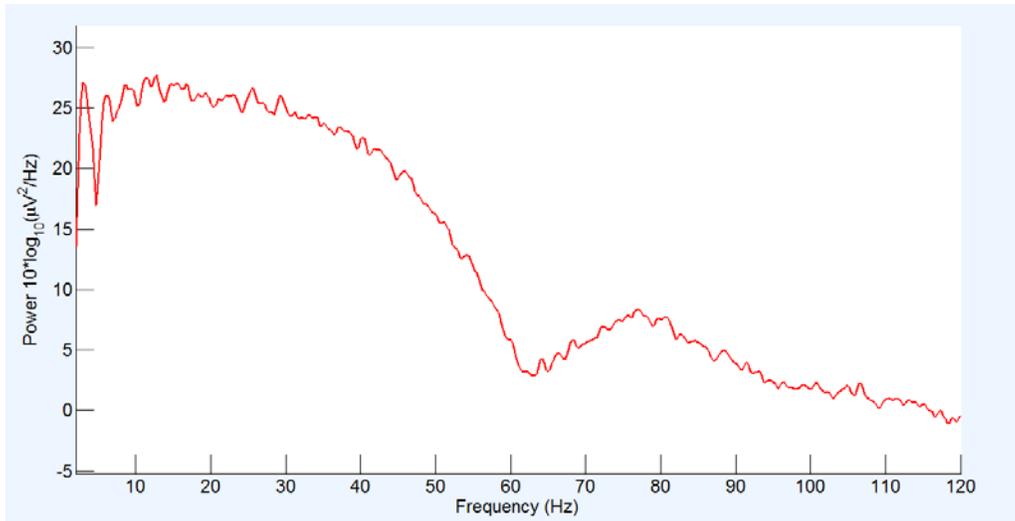


Figura 4.32 Densidad espectral de potencia de 16265. [Fuente Propia]

Paso 3: Obtención de las estadísticas de la señal 16265 con el uso de EEGLab (Ver figura 4.33).

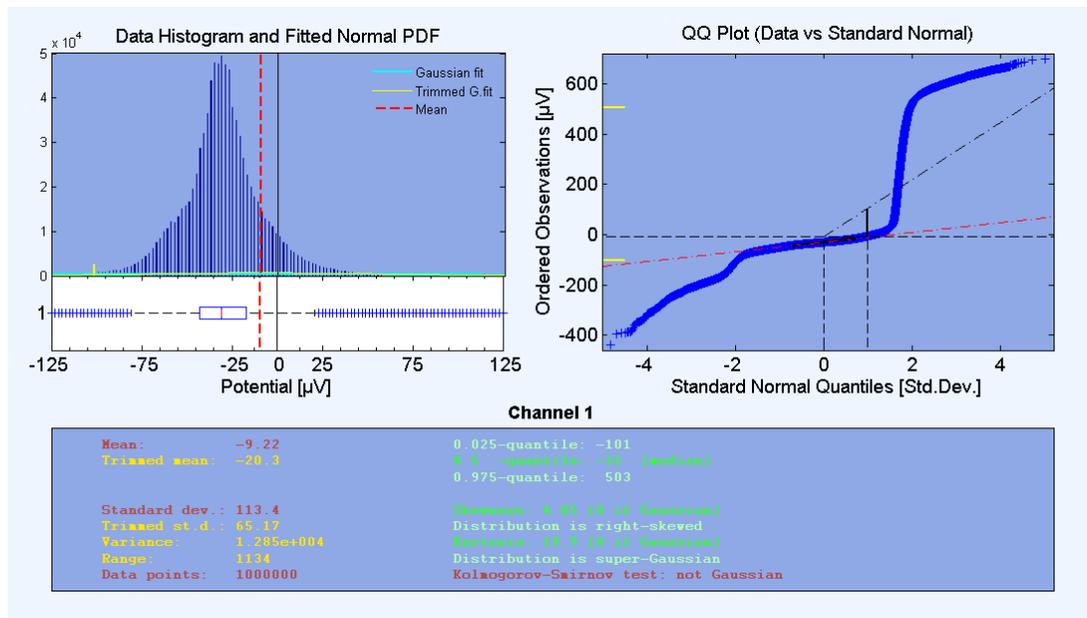


Figura 4.33 Estadísticas de la señal ECG 16265, antes de agregarle ruido. [Fuente Propia]

Generación de la señal de ruido, en este caso es ruido de línea su forma de onda es la mostrada en la figura 4.34:

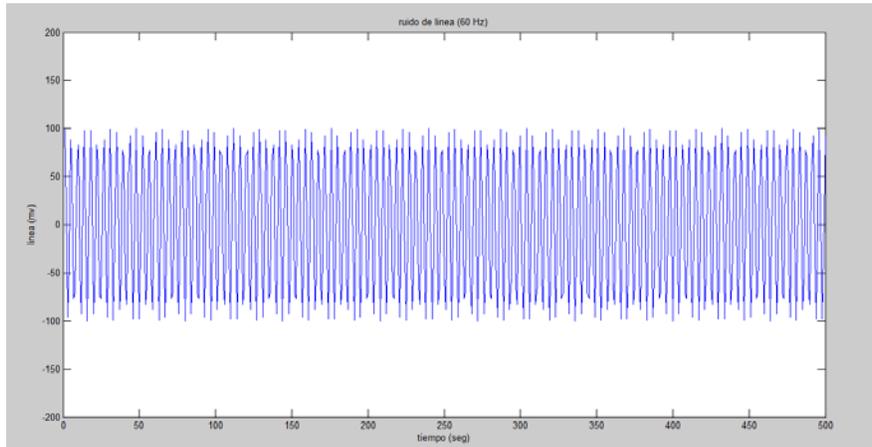


Figura 4.34 Onda que representa la señal de ruido aditivo a la señal ECG 16265. [Fuente Propia]

Esta señal se generó en MATLAB a través de algunos comandos, colocando los parámetros adecuados, los procedimientos para tal fin, están descritos en la fase IV.

Paso 4: Gráfica de la onda electrocardiográfica después de agregarle el ruido de línea (60Hz), (Ver figura 4.35).

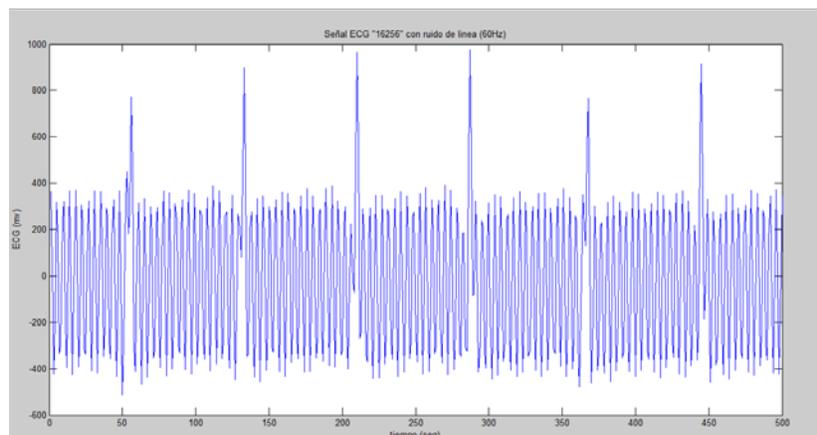


Figura 4.35 Onda que representa la señal ECG 16265 contaminada de ruido de línea. [Fuente Propia]

Paso 5: Procediendo de la misma manera que en el paso 2, se obtiene la Gráfica de la densidad espectral de potencia de la señal que ya posee ruido de línea (ver figura 4.36)

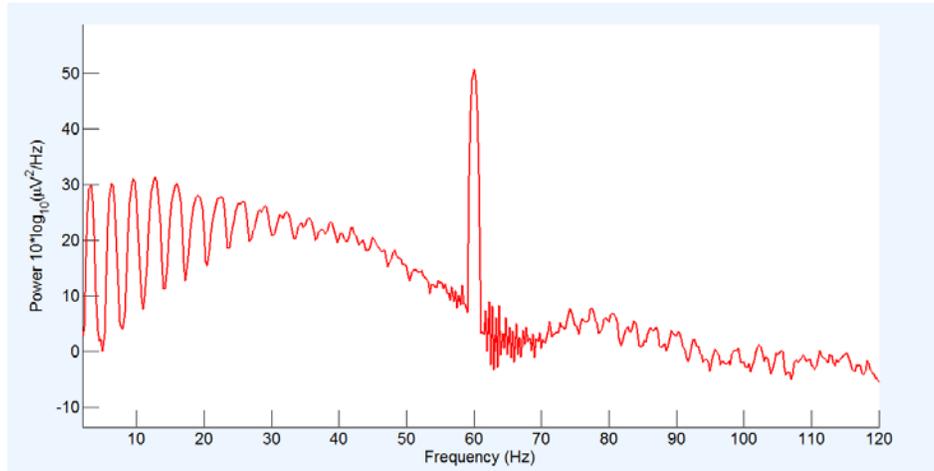


Figura 4.36 Densidad espectral de potencia de 16265 contaminada de ruido de línea. [Fuente Propia]

El pico que se observa en 60Hz (en la figura 4.36), evidencia que existe una señal de esa frecuencia presente de forma aditiva a la señal original.

Paso 6: Obtención de las Estadísticas de la señal 16265 con ruido de línea usando EEGLab, mediante las opciones descritas en la fase IV, para realización del mismo (ver figura 4.37).

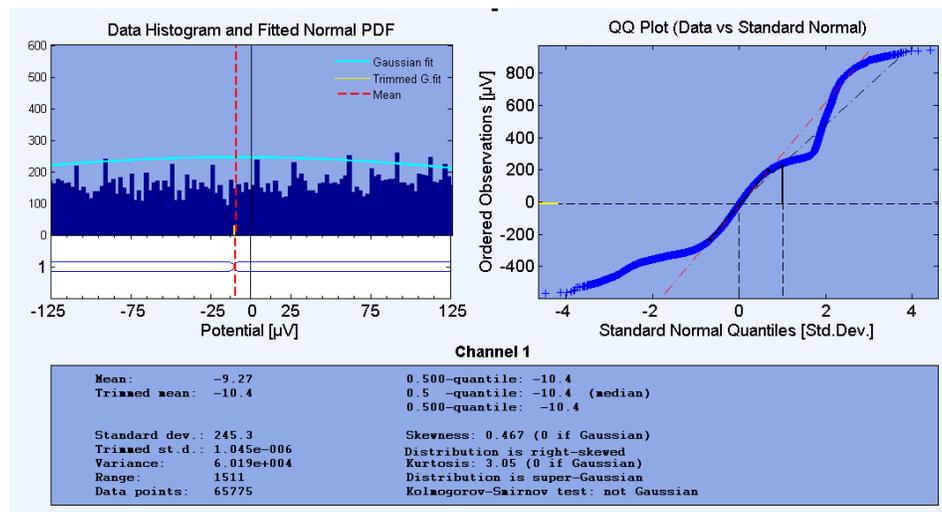


Figura 4.37 Estadísticas de la señal ECG 16265, después de agregarle ruido. [Fuente Propia]

Paso 7: Utilización de fastICA para obtener las fuentes estimadas:

Antes de generar las componentes independientes, se reduce la dimensión de la matriz de mezcla en la opción “Reduce dim”, como se explicó en la fase IV, en la figura 4.38 se observa la reducción de la esta matriz.

Reducción de la dimensión de la matriz de mezcla:

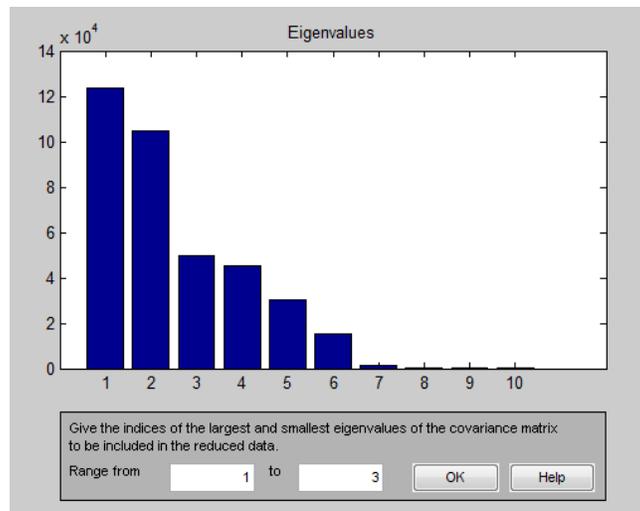


Figura 4.38 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]

Componentes independientes obtenidos después de aplicar el paquete fastICA, (ver figura 3.39).

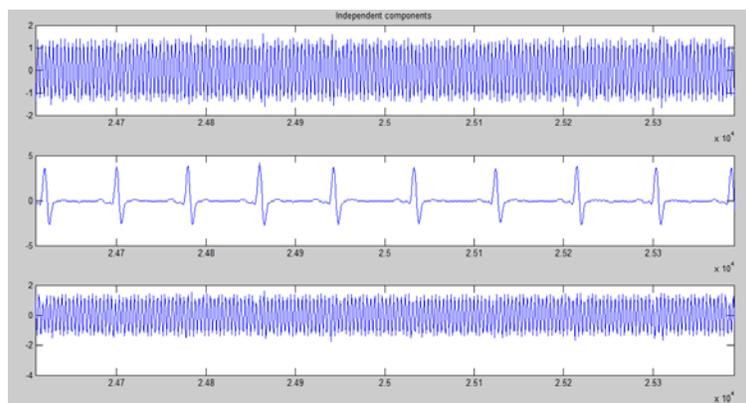


Figura 4.39 Componentes impiedentes o fuentes estimadas por fastICA. [Fuente Propia]

En la fase IV también se detallan los pasos para obtener las componentes independientes usando el fastICA.

Paso 8: Obtención de la Gráfica de la densidad espectral de potencia de las componentes independientes usando EEGLab, (ver figura 4.40).

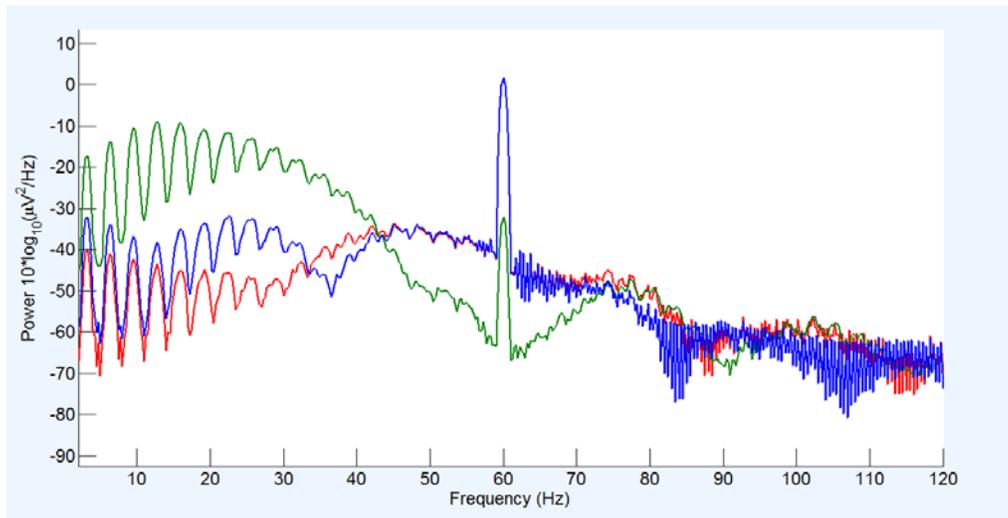


Figura 4.40 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]

Paso 9: Obtención de las estadísticas para cada una de las componentes independientes, (ver figuras 4.41, 4.42 y 4.42).

Componente independiente 1

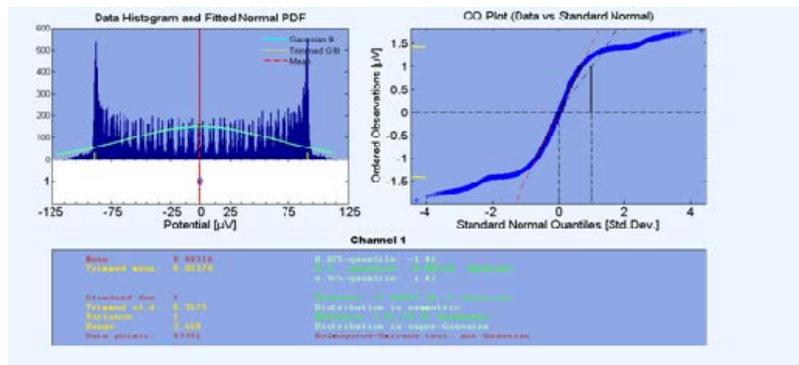


Figura 4.41 Estadísticas de la primera componente. [Fuente Propia]

Componente independiente 2

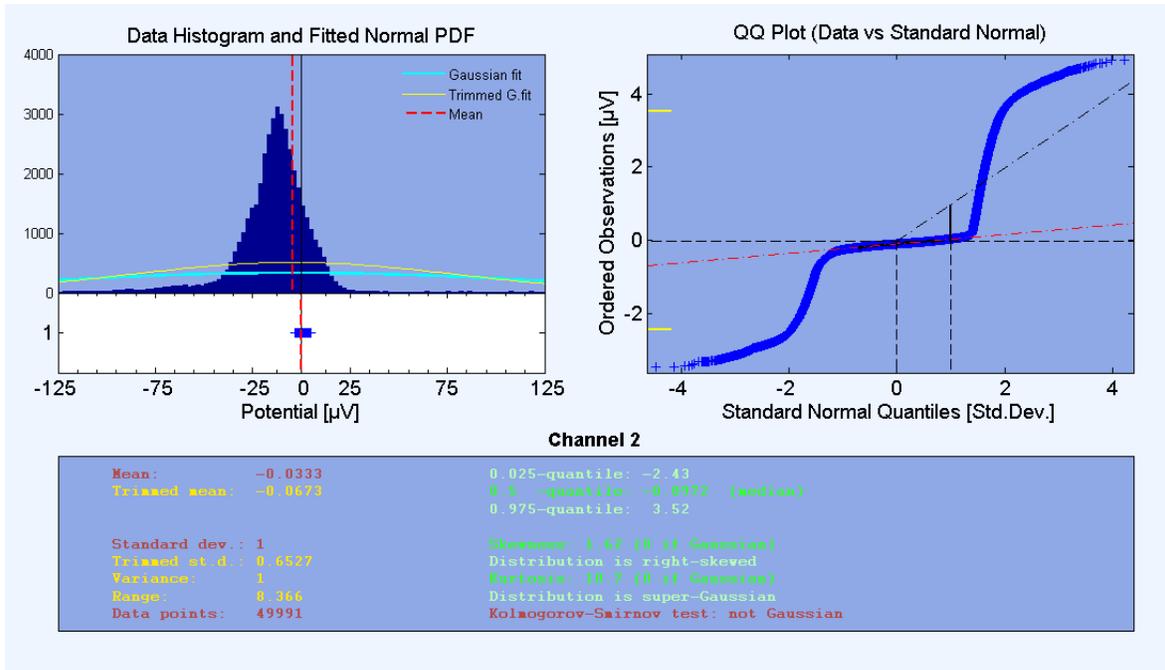


Figura 4.42 Estadísticas de la segunda componente. [Fuente Propia]

Componente independiente 3

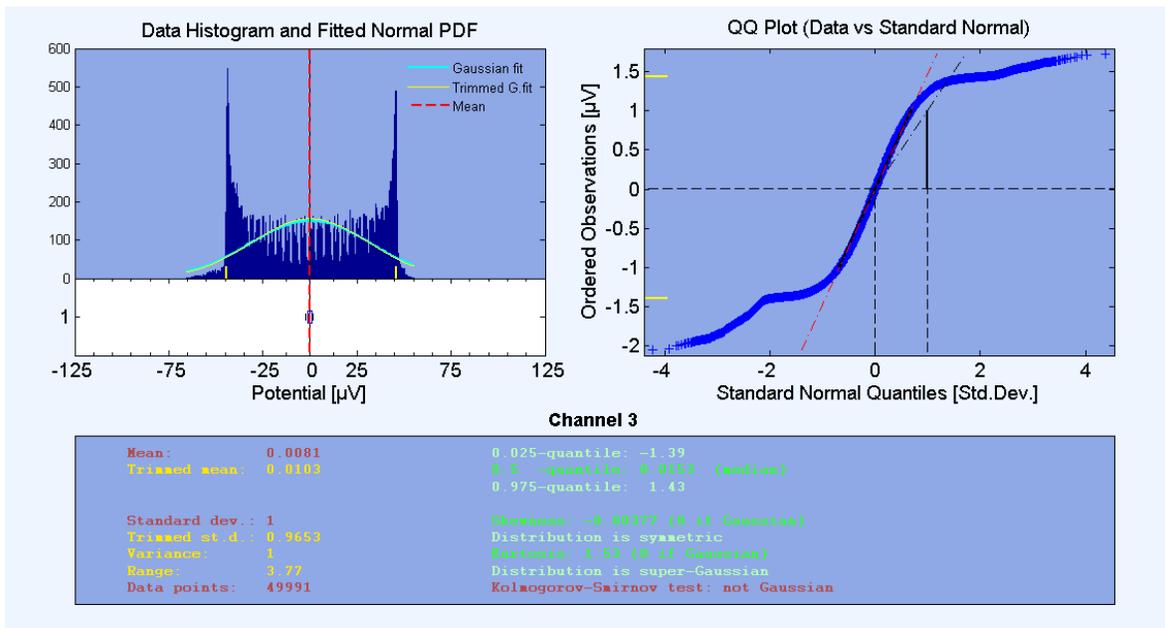


Figura 4.43 Estadísticas de la tercera componente. [Fuente Propia]

Experimento 2:



Registro 16273 del “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”.

Resultados después de aplicar los pasos del 1 al 9:

Paso 1:

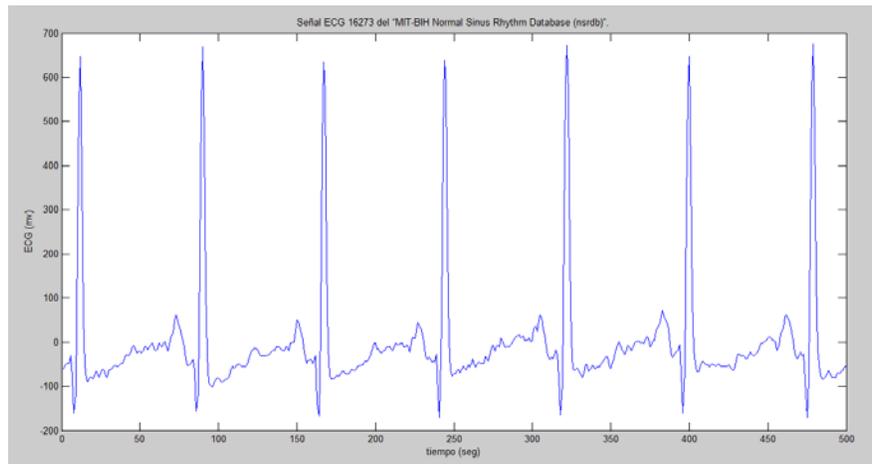


Figura 4.44 Forma de onda de la señal ECG 16273 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia].

Paso 2:

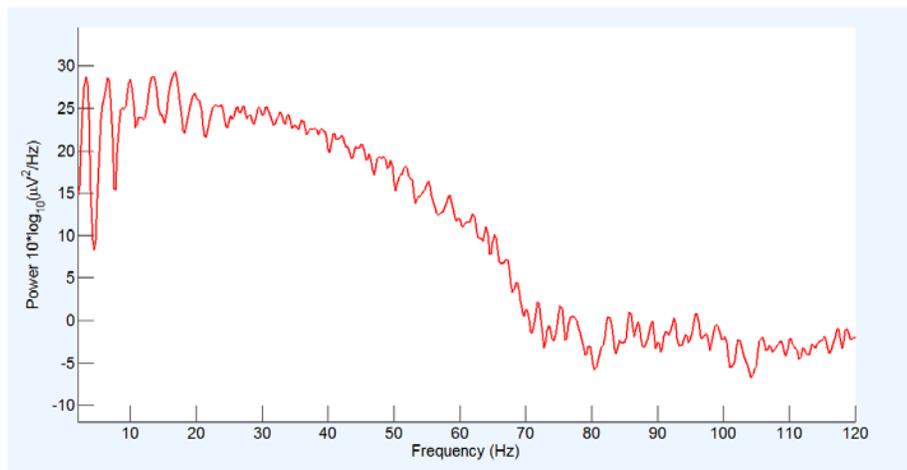


Figura 4.45 Densidad espectral de potencia de 16273. [Fuente Propia]

Paso 3:

CAPÍTULO IV. RESULTADOS

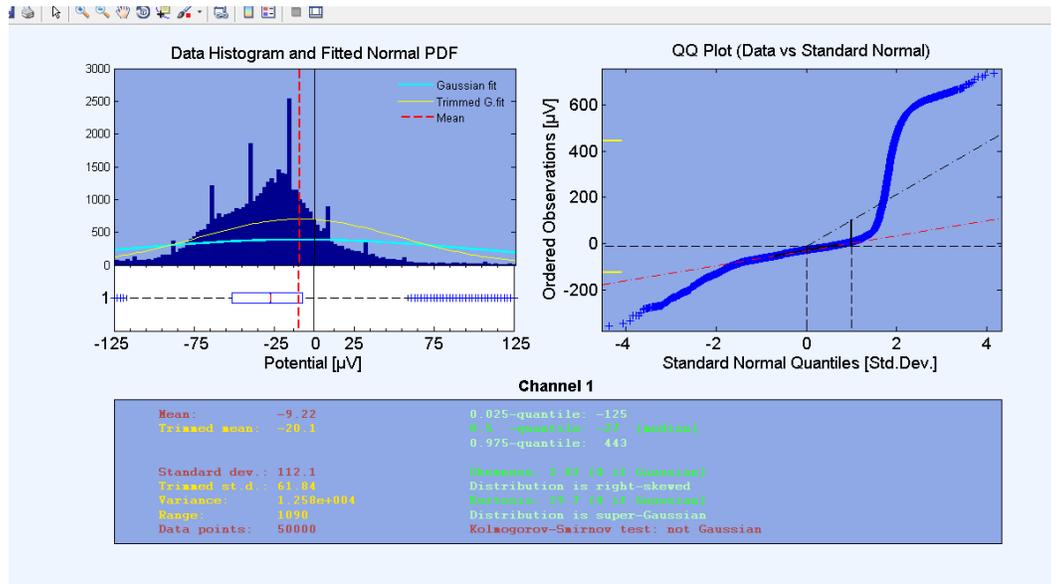


Figura 4.46 Estadísticas de la señal ECG 16273, antes de agregarle ruido. [Fuente Propia]

Para este experimento el ruido aplicado fue de línea (60Hz) al igual que en el experimento 1, en la siguiente figura se observa la onda de la señal ECG 16273 contaminada con dicho ruido.

Paso 4:

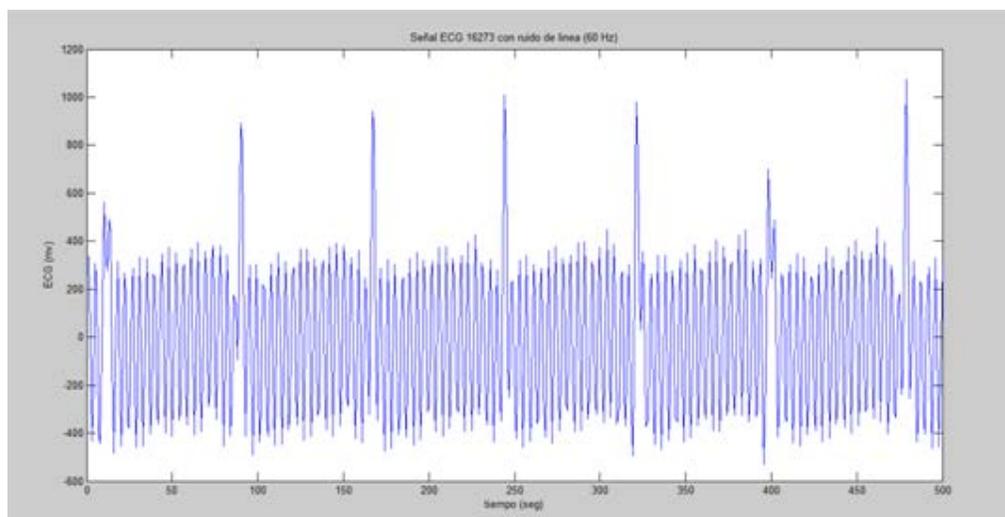


Figura 4.47 Onda que representa la señal de ruido aditivo a la señal ECG 16273. [Fuente Propia]

Paso 5:

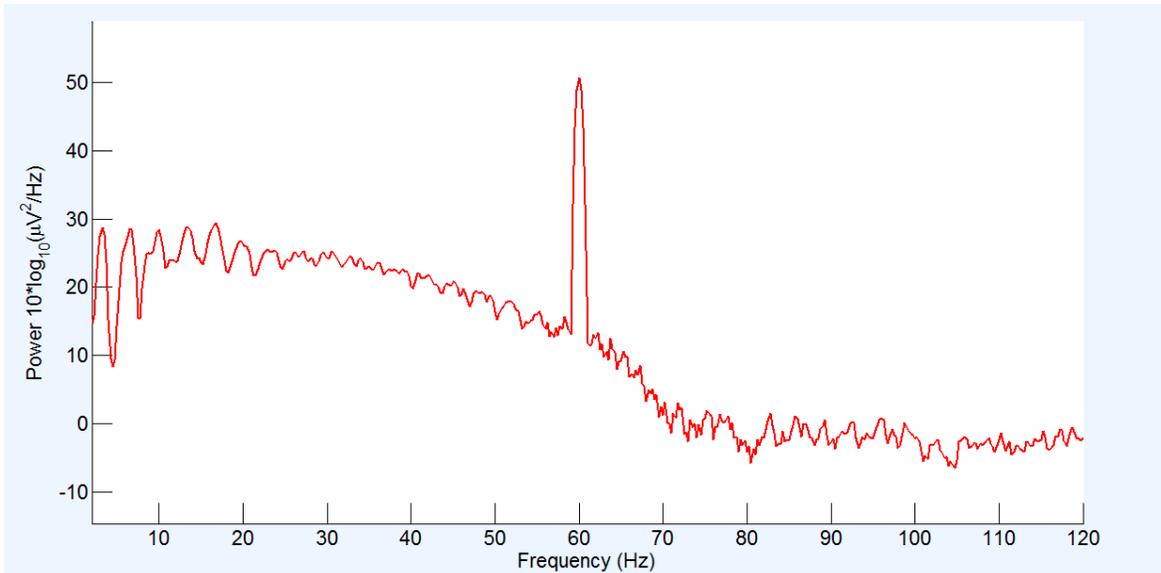


Figura 4.48 Densidad espectral de potencia de 16273 contaminada de ruido de línea. [Fuente Propia].

Paso 6:

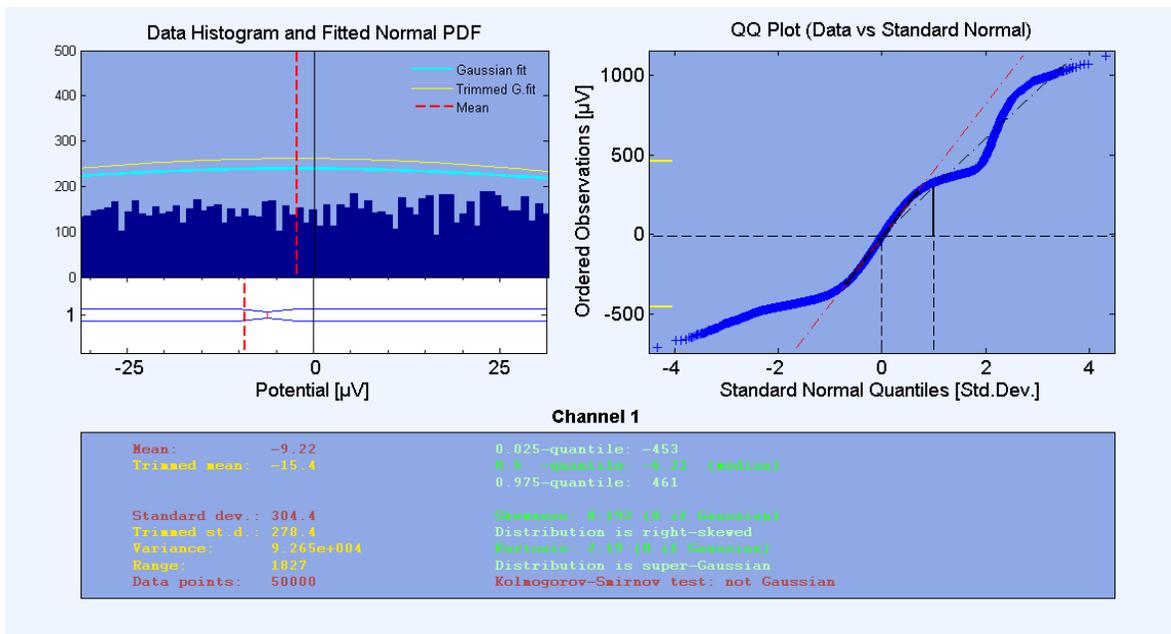


Figura 4.49 Estadísticas de la señal ECG 16273, después de agregarle ruido. [Fuente Propia]

Paso7:

Reducción de la dimensión de la matriz de mezcla (ver figura 4.50).

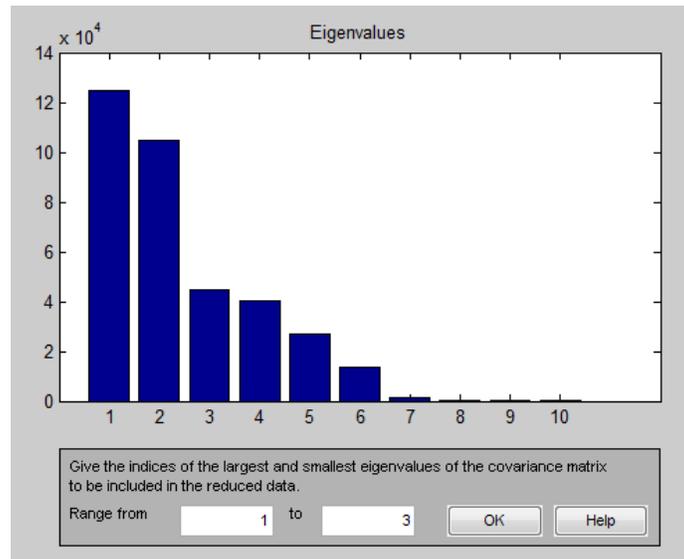


Figura 4.50 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]

Componentes independientes obtenidos después de aplicar el paquete fastICA, (ver figura 4.51).

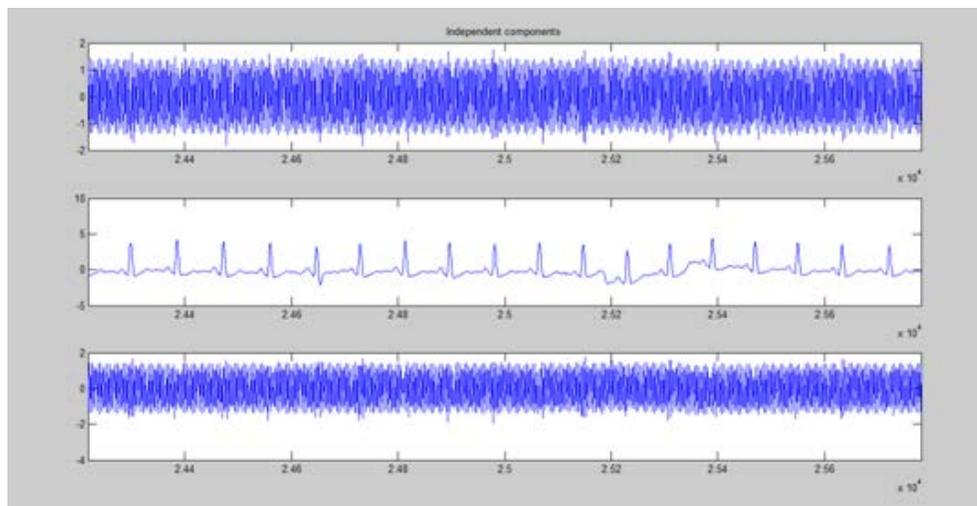


Figura 4.51 Componentes independientes o fuentes estimadas por fastICA. [Fuente Propia]

Paso 8:

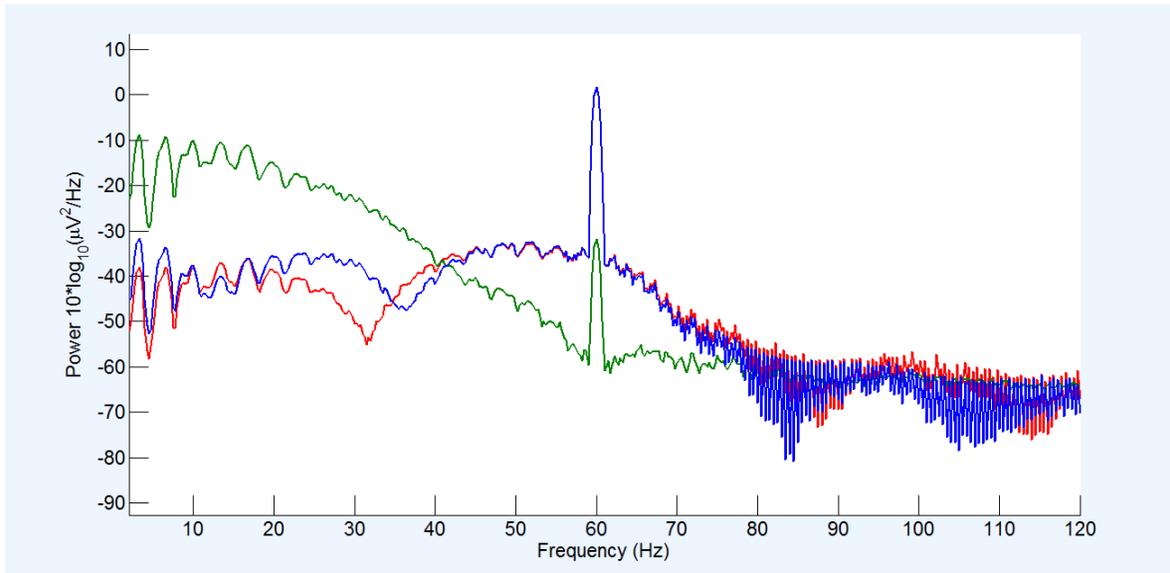


Figura 4.52 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]

Paso 9:

Componente independiente 1

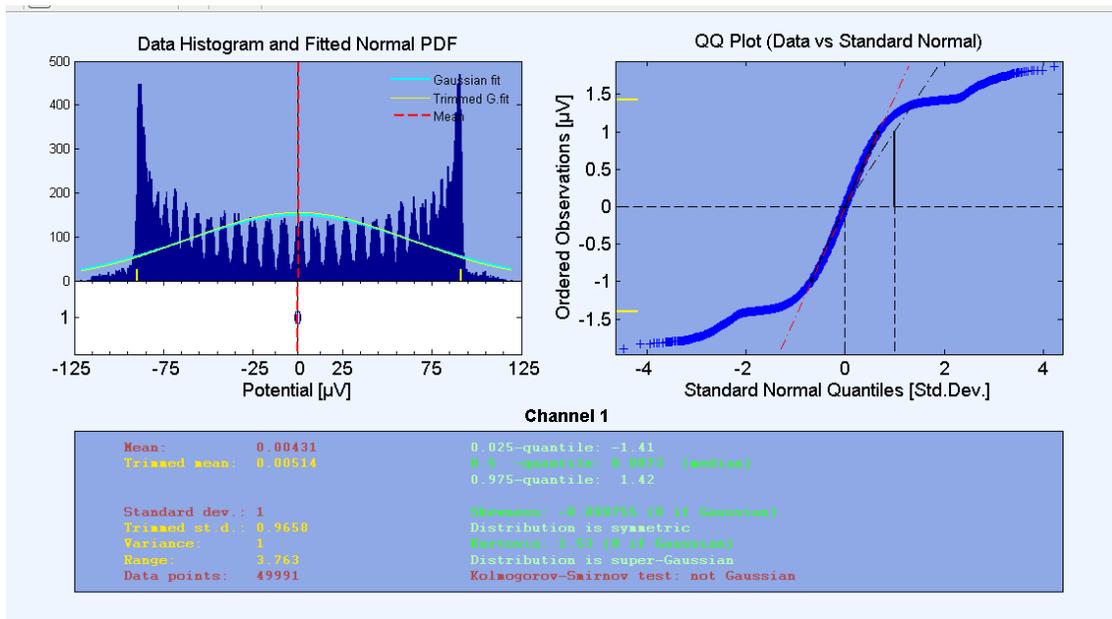


Figura 4.53 Estadísticas de la primera componente. [Fuente Propia]

Componente independiente 2

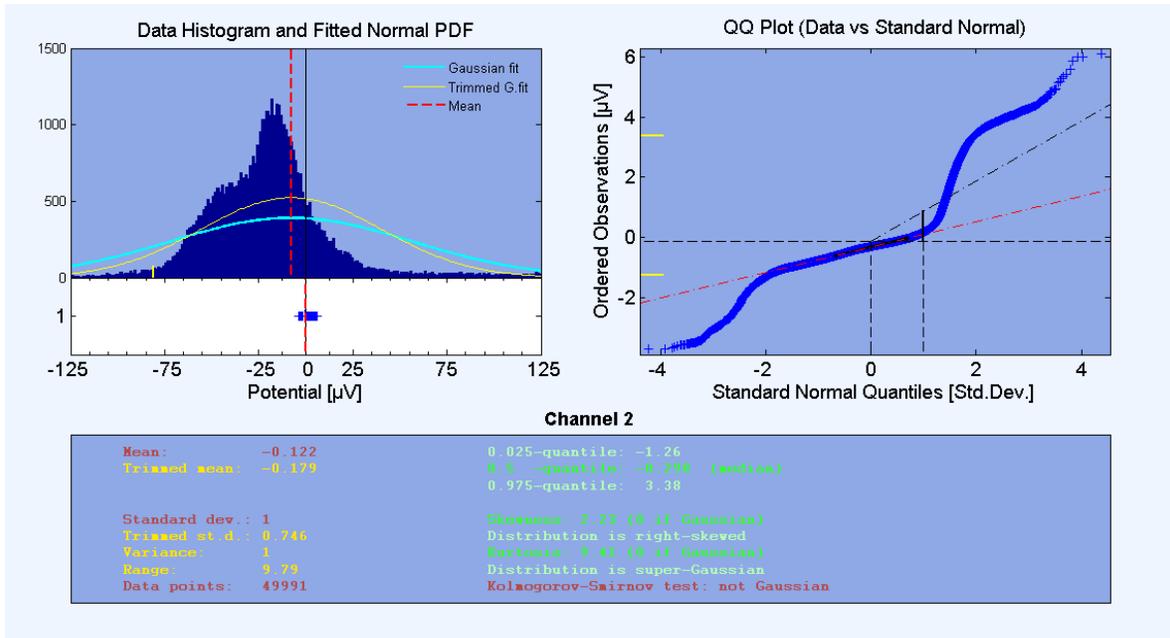


Figura 4.54 Estadísticas de la segunda componente. [Fuente Propia]

Componente independiente 3

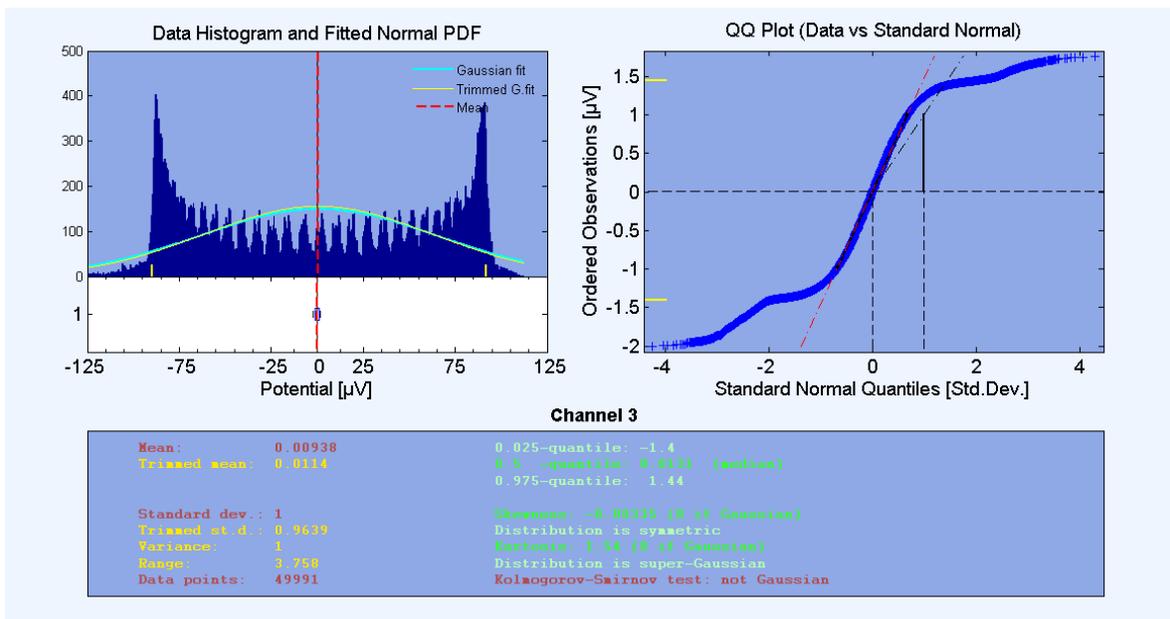


Figura 4.55 Estadísticas de la tercera componente. [Fuente Propia].

Experimento 3



Registro 18184 del “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”.

Resultados después de aplicar los pasos del 1 al 9:

Paso 1:

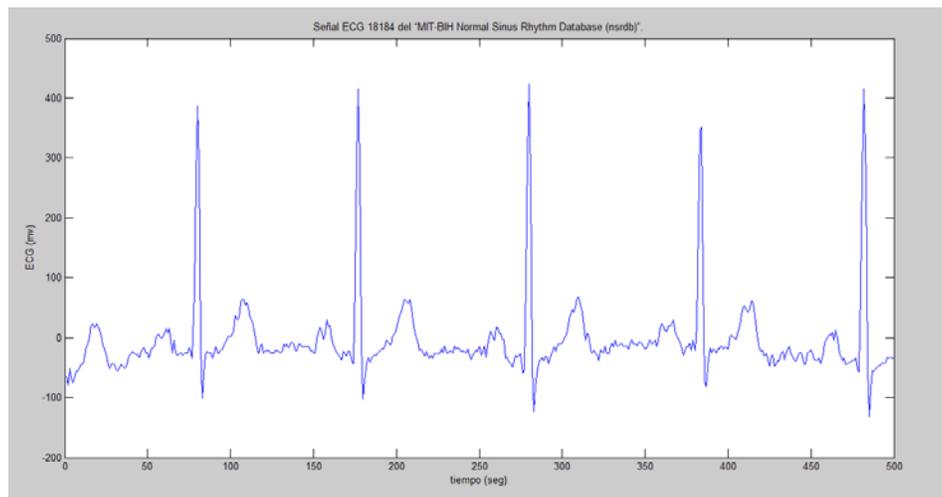


Figura 4.56 Forma de onda de la señal ECG 18184 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia]

Paso 2:

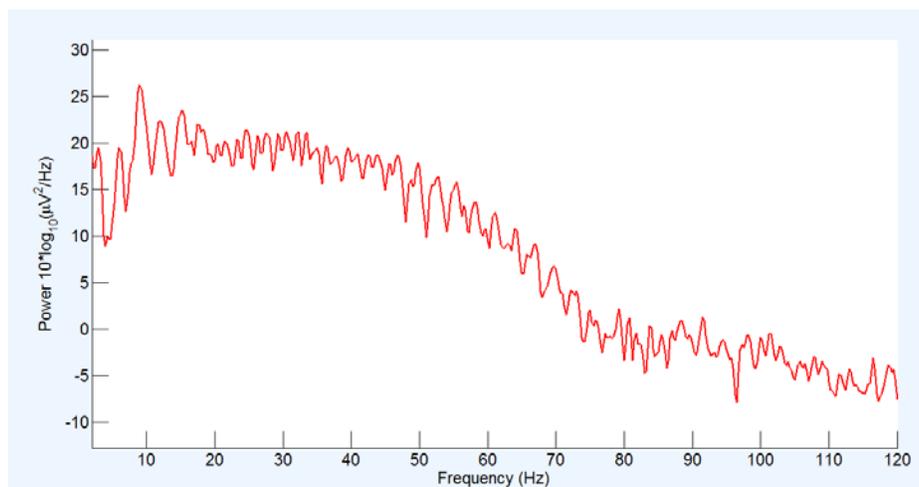


Figura 4.57 Densidad espectral de potencia de 18184. [Fuente Propia]

Paso 3:

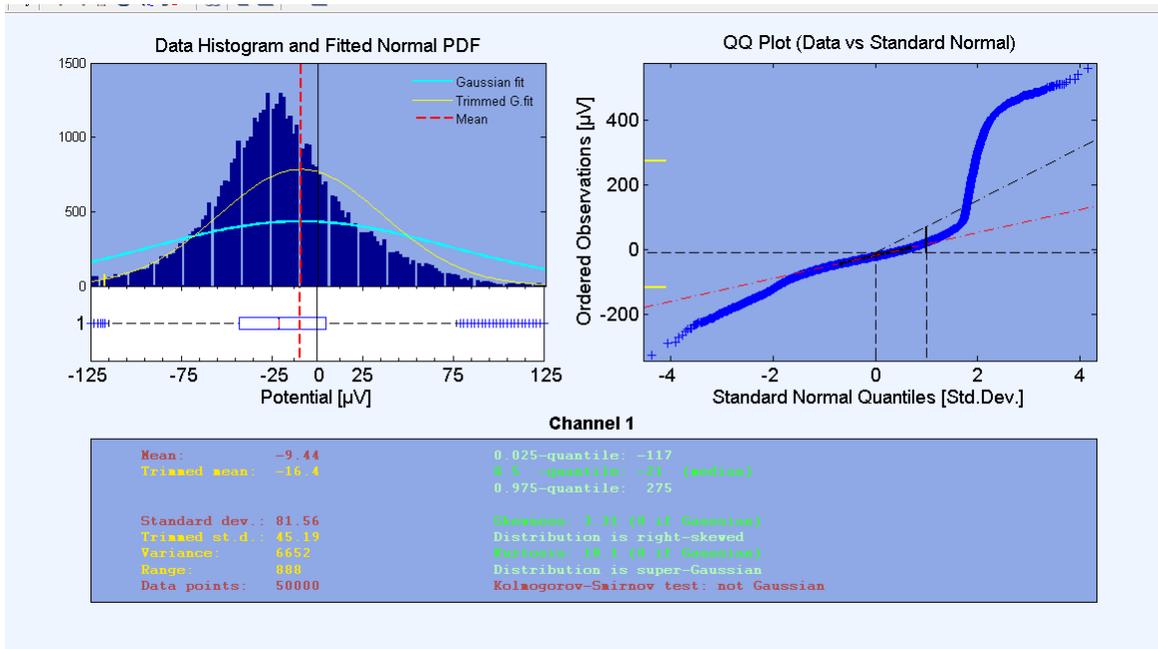


Figura 4.58 estadísticas de la señal ECG 18184, antes de agregarle ruido. [Fuente Propia]

Paso 4:

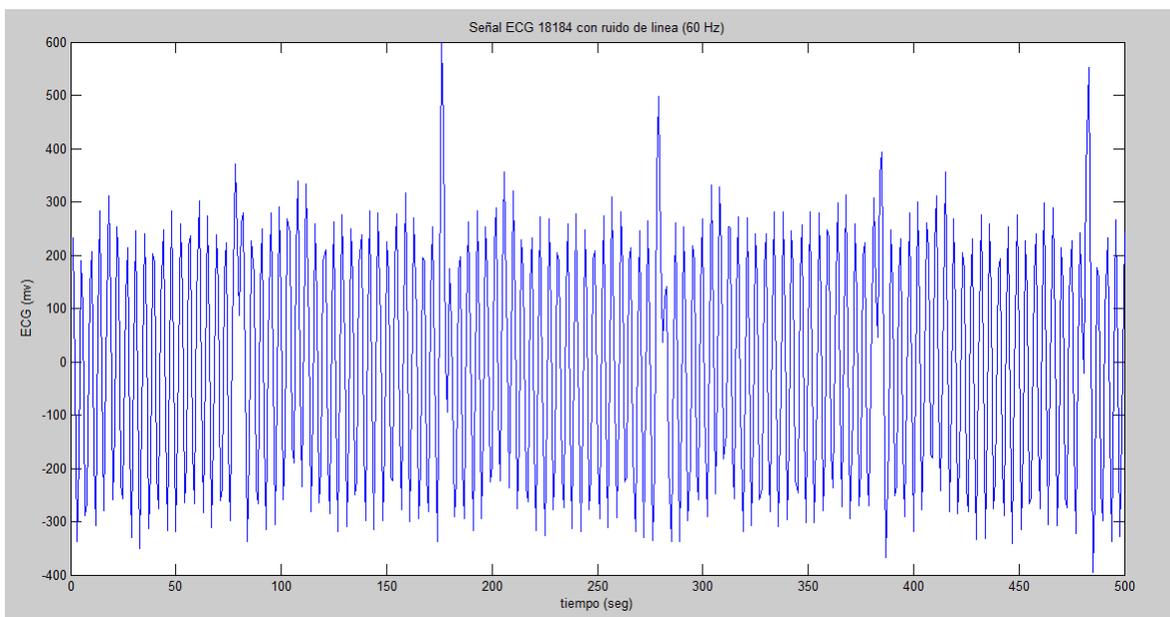


Figura 4.59 Onda que representa la señal ECG 18184 contaminada de ruido de línea. [Fuente Propia]

Paso 5:

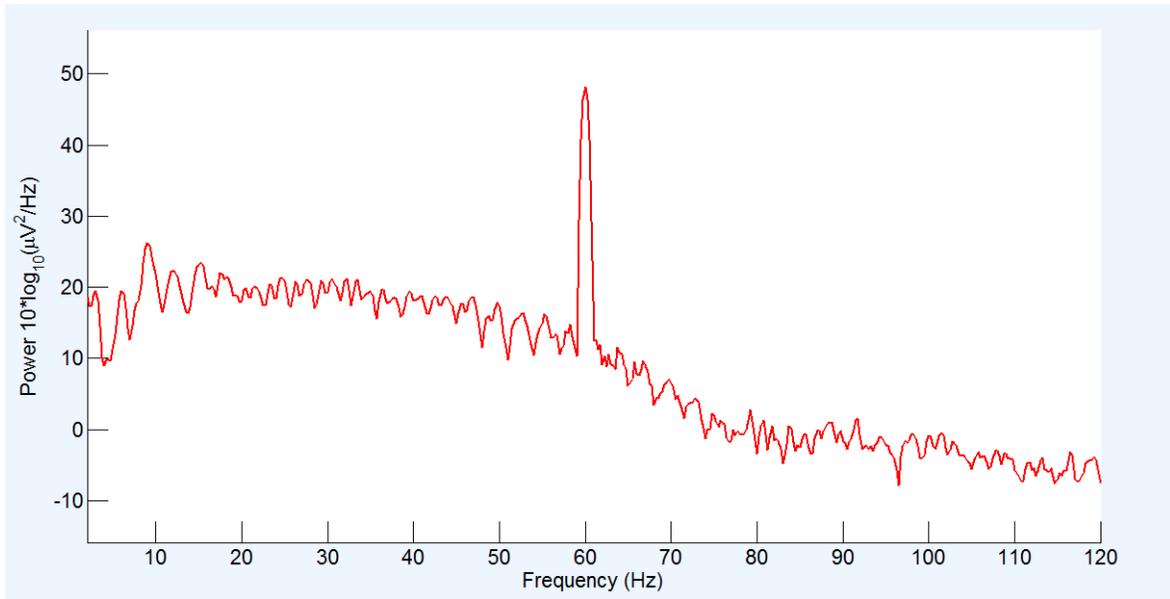


Figura 4.60 Densidad espectral de potencia de 18184 contaminada de ruido de línea. [Fuente Propia].

Paso 6:

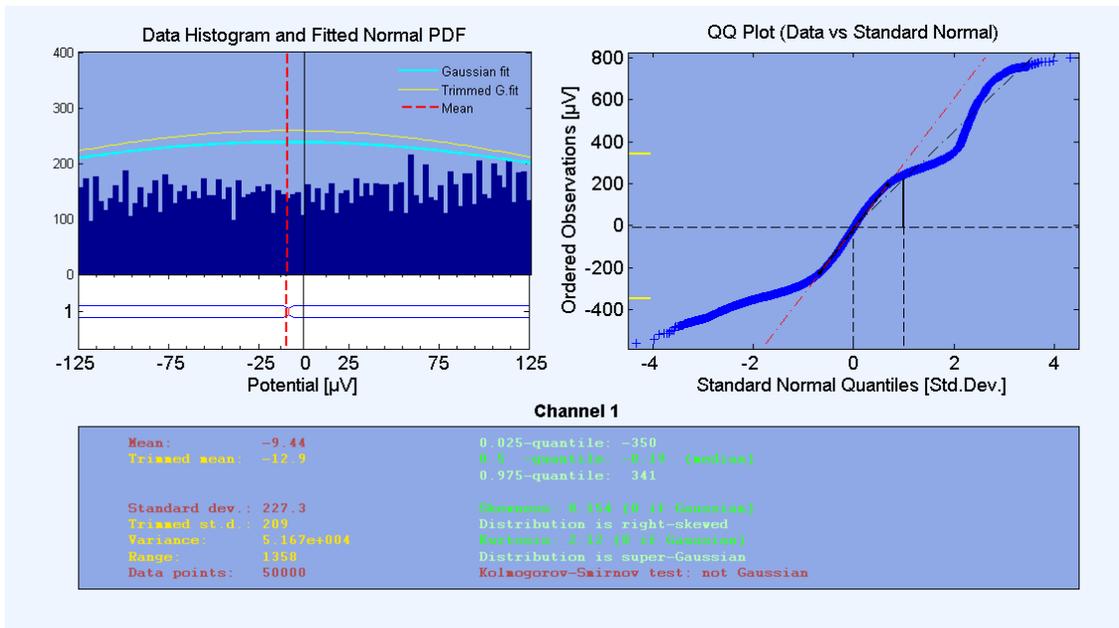


Figura 4.61 Estadísticas de la señal ECG 18184, después de agregarle ruido. [Fuente Propia]

Paso 7:

Reducción de la dimensión de la matriz de mezcla:

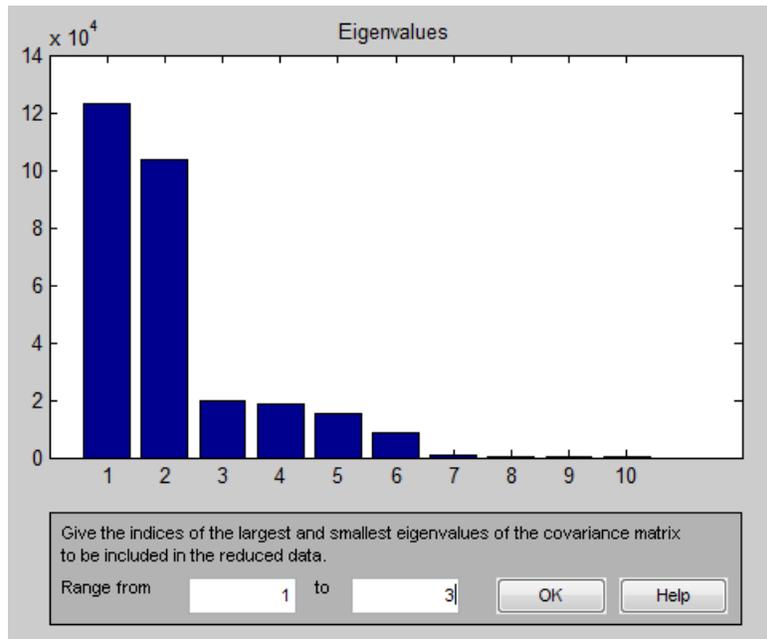


Figura 4.62 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]

Componentes independientes obtenidos después de aplicar el paquete fastICA, (ver figura 4.62).

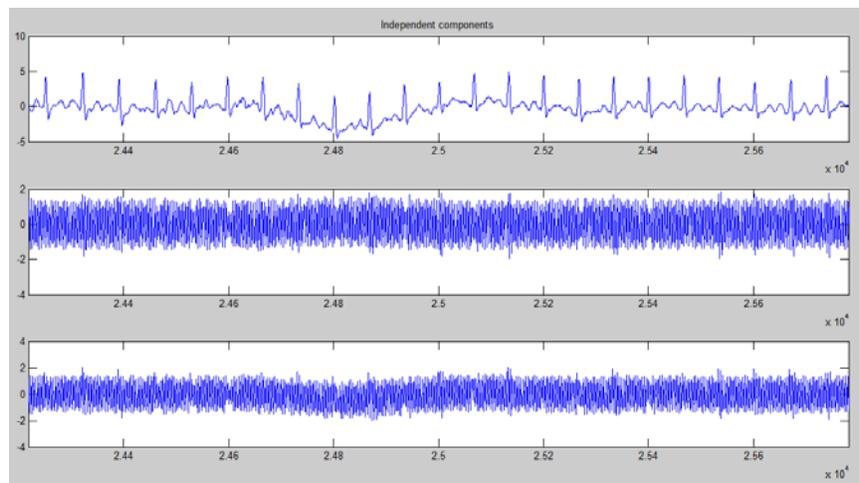


Figura 4.63. Componentes independientes o fuentes estimadas por fastICA. [Fuente Propia].

Paso 8:

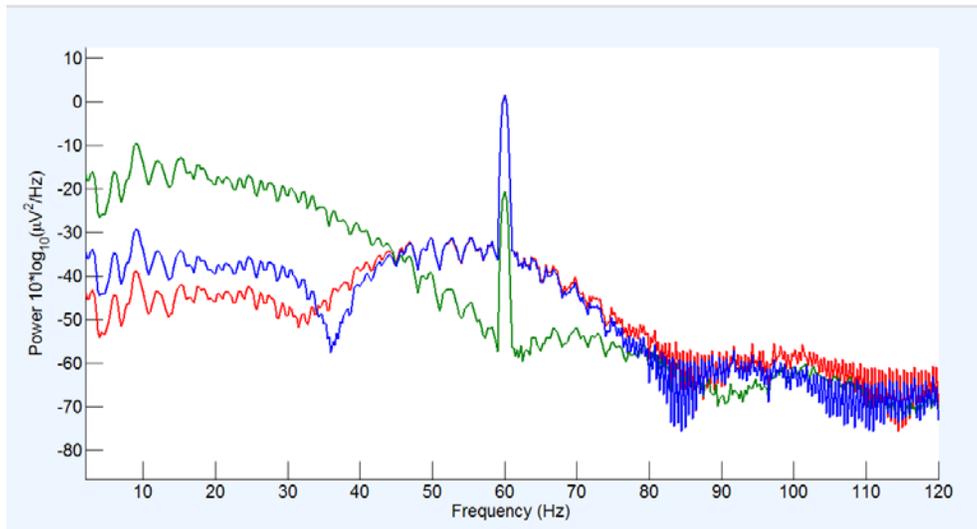


Figura 4.64 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]

Paso 9

Componente independiente 1

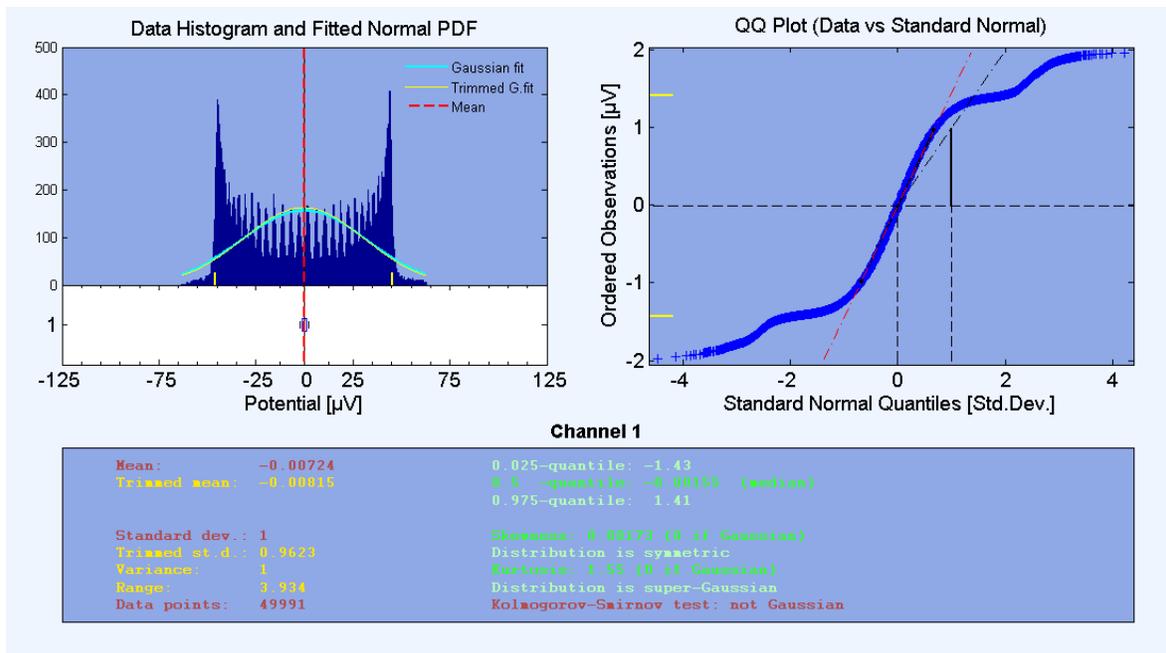


Figura 4.65 Estadísticas de la primera componente [Fuente Propia].

Componente independiente 2

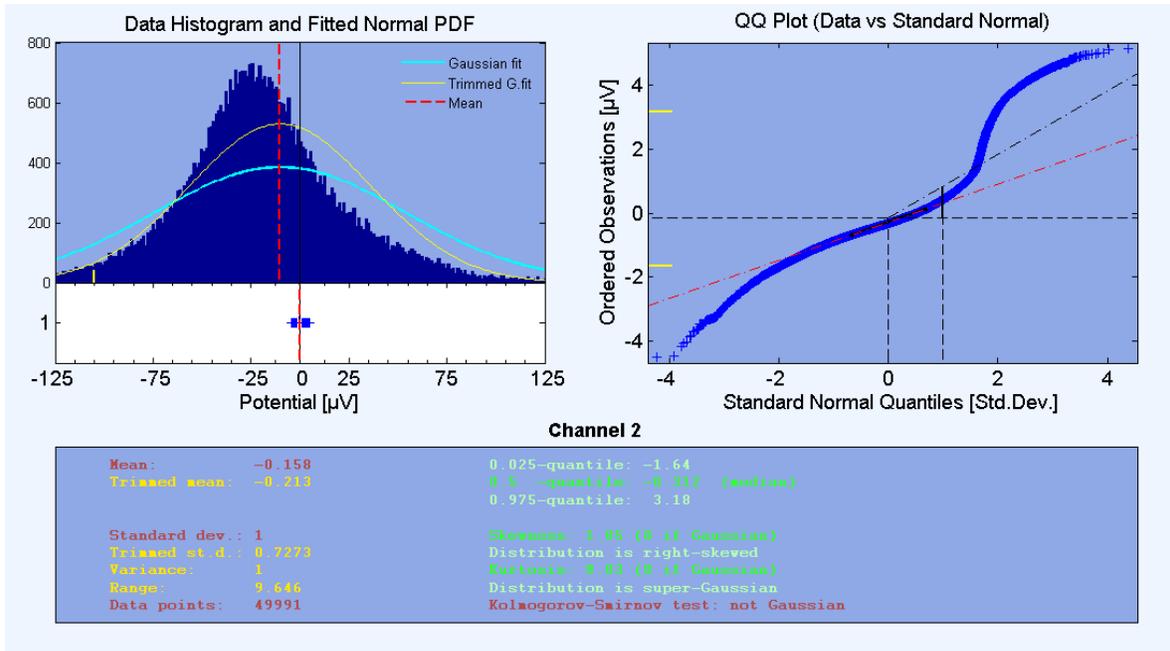


Figura 4.66 Estadísticas de la segunda componente. [Fuente Propia]

Componente independiente 3

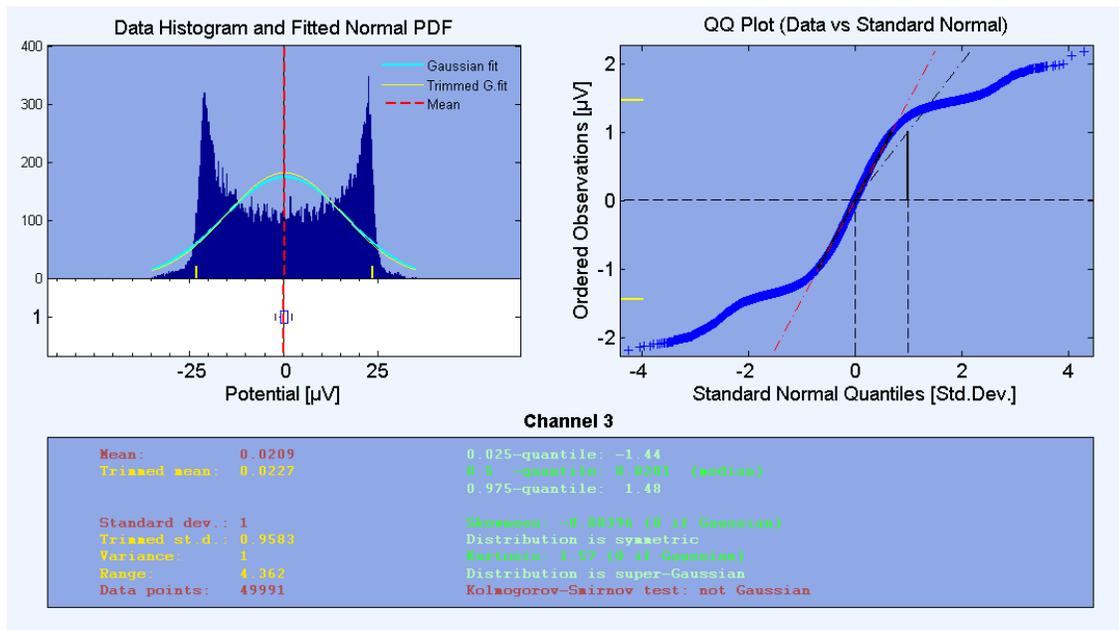


Figura 4.67 Estadísticas de la tercera componente. [Fuente Propia]

Experimento 4:



Registro 17052 del “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”.

Resultados después de aplicar los pasos del 1 al 9:

Paso 1:

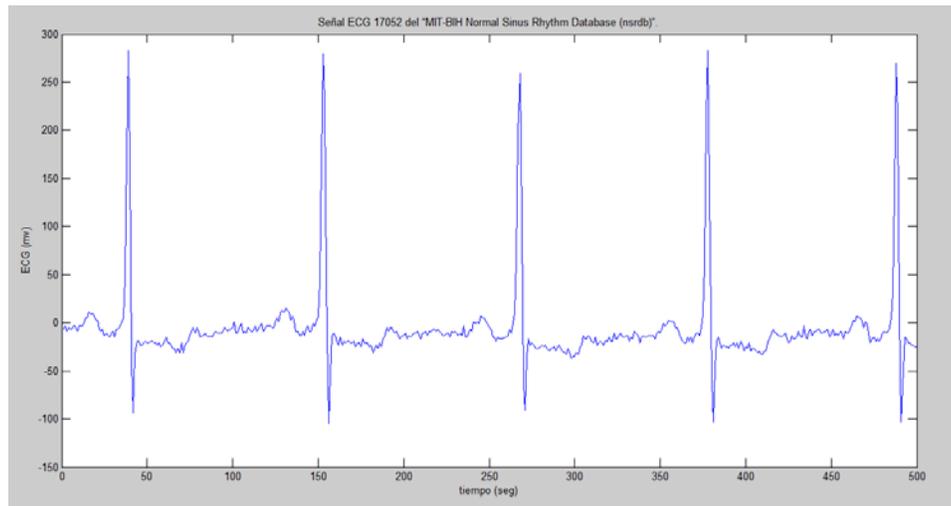


Figura 4.68 Forma de onda de la señal ECG 17052 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia].

Paso 2:

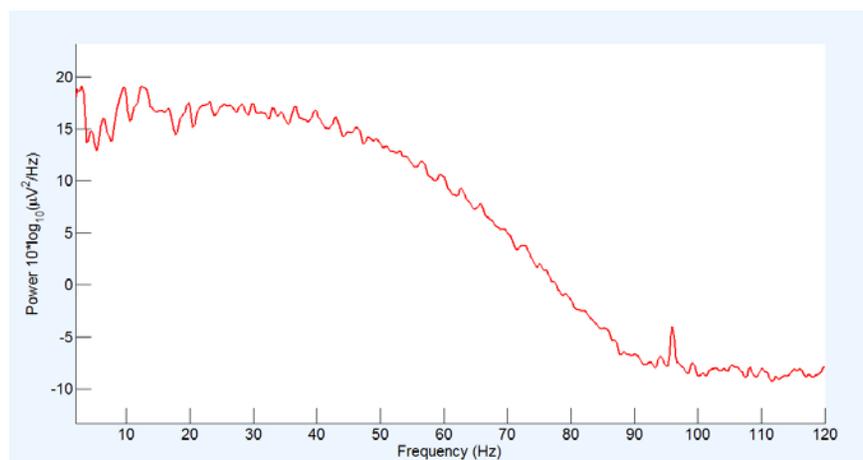


Figura 4.69 Densidad espectral de potencia de 17052.

Paso 3:

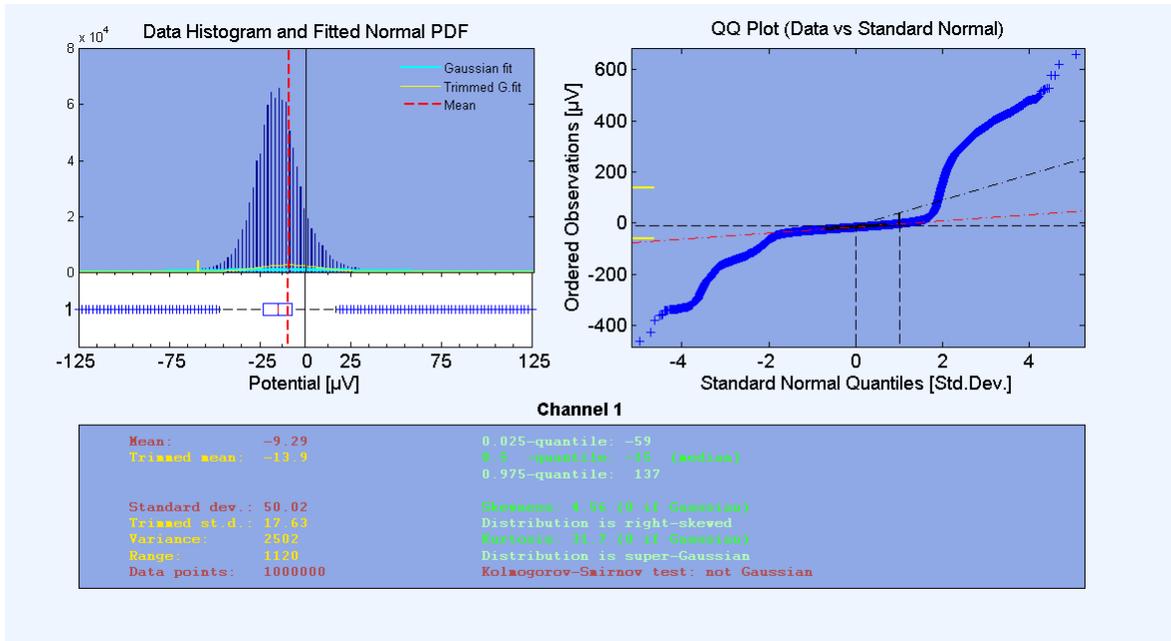


Figura 4.70 Estadísticas de la señal ECG 17052, antes de agregarle ruido. [Fuente Propia]

Paso 4:

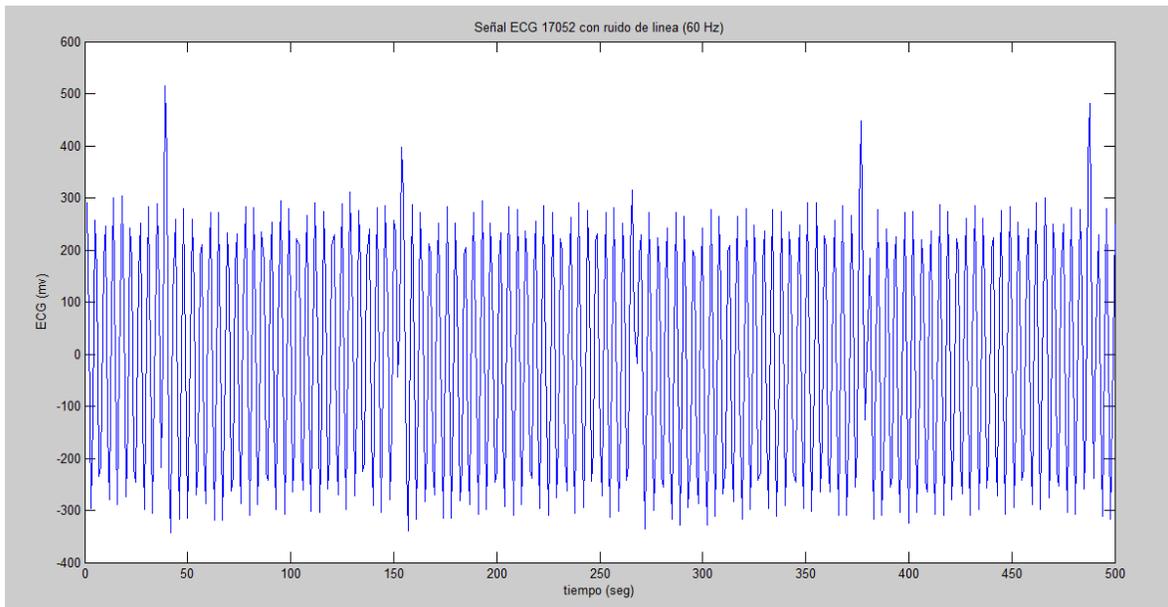


Figura 4.71 Onda que representa la señal ECG 17052 contaminada de ruido de línea. [Fuente Propia].

Paso 5:

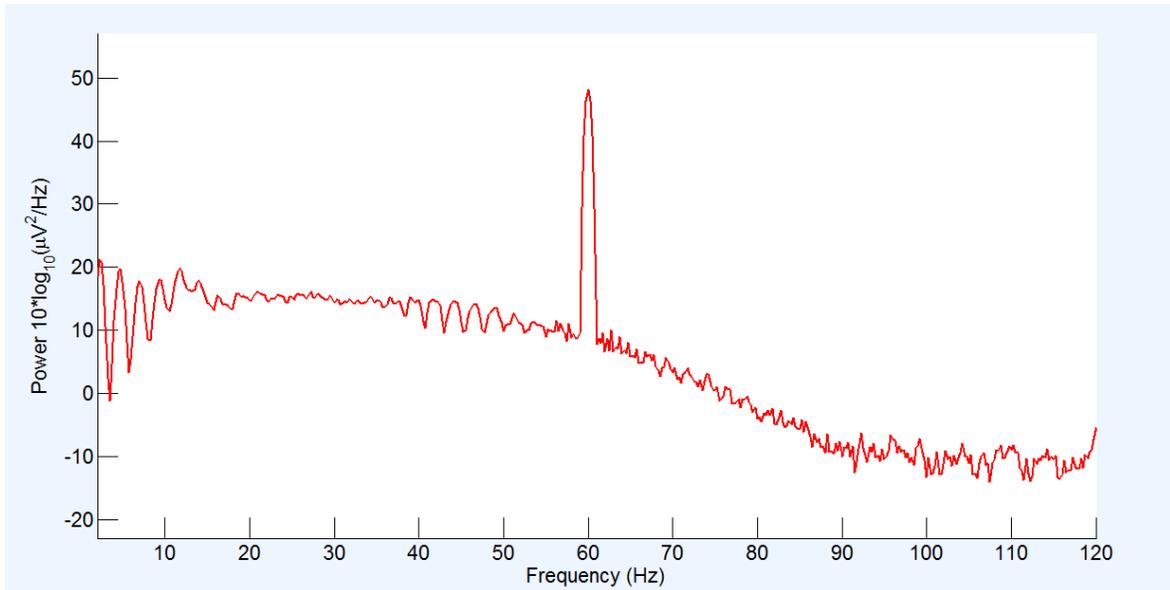


Figura 4.72 Densidad espectral de potencia de 17052 contaminada de ruido de línea. [Fuente Propia]

Paso 6:

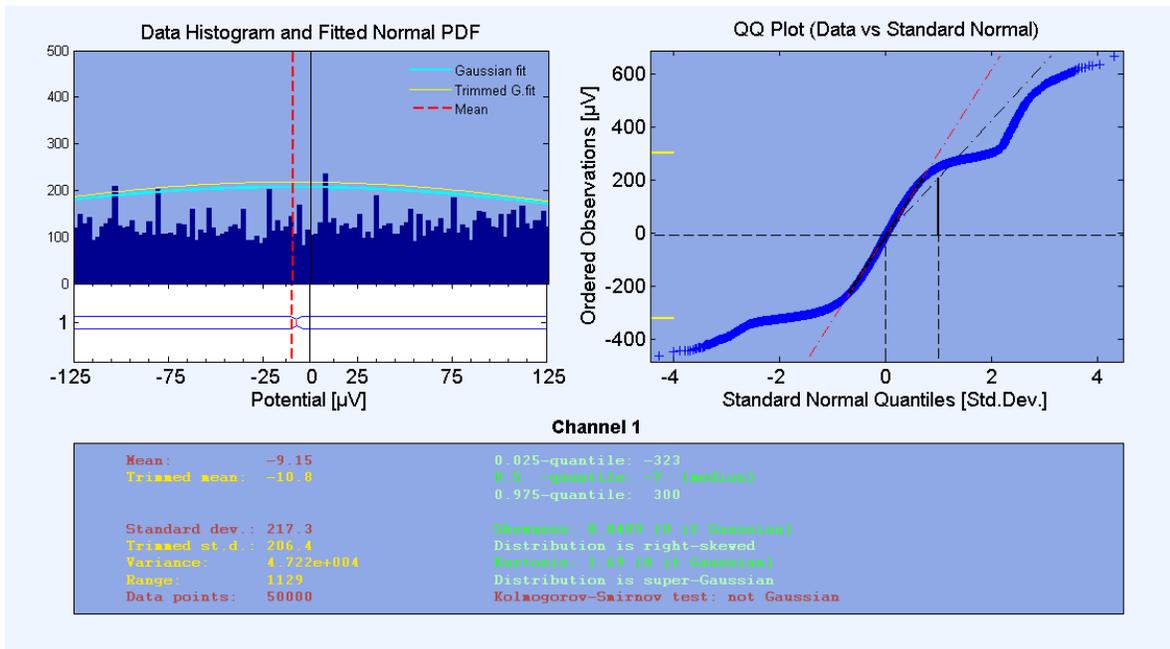


Figura 4.73 Estadísticas de la señal ECG 17052, después de agregarle ruido. [Fuente Propia]

Paso 7:

Reducción de la dimensión de la matriz de mezcla:

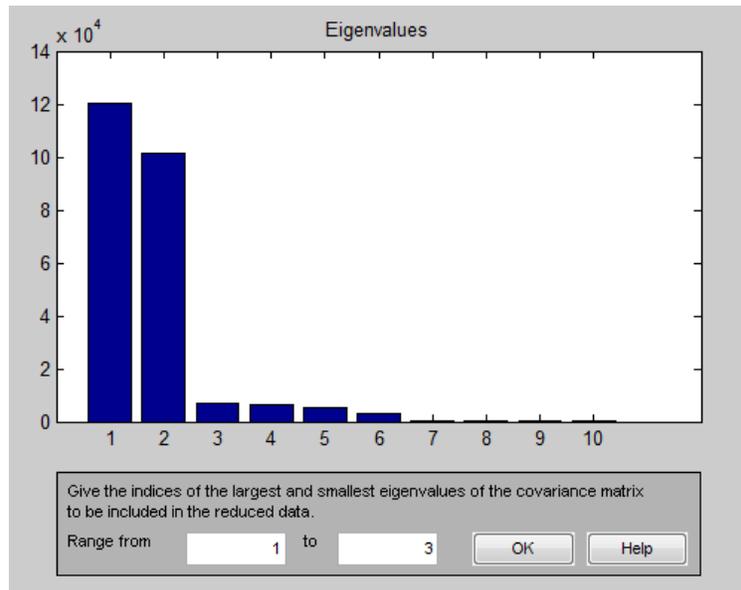


Figura 4.74 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]

Componentes independientes obtenidos después de aplicar el paquete fastICA, (ver figura 4.75).

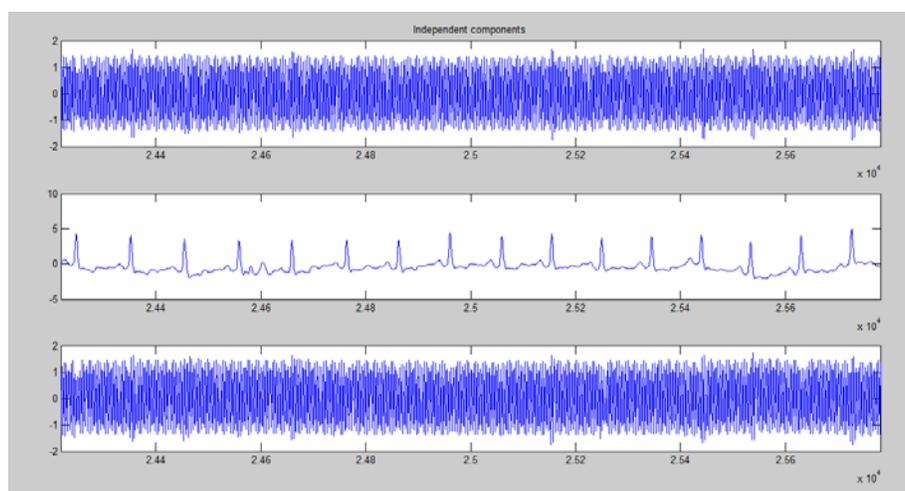


Figura 4.75 Componentes independientes o fuentes estimadas por fastICA. [Fuente Propia]

Paso 8:

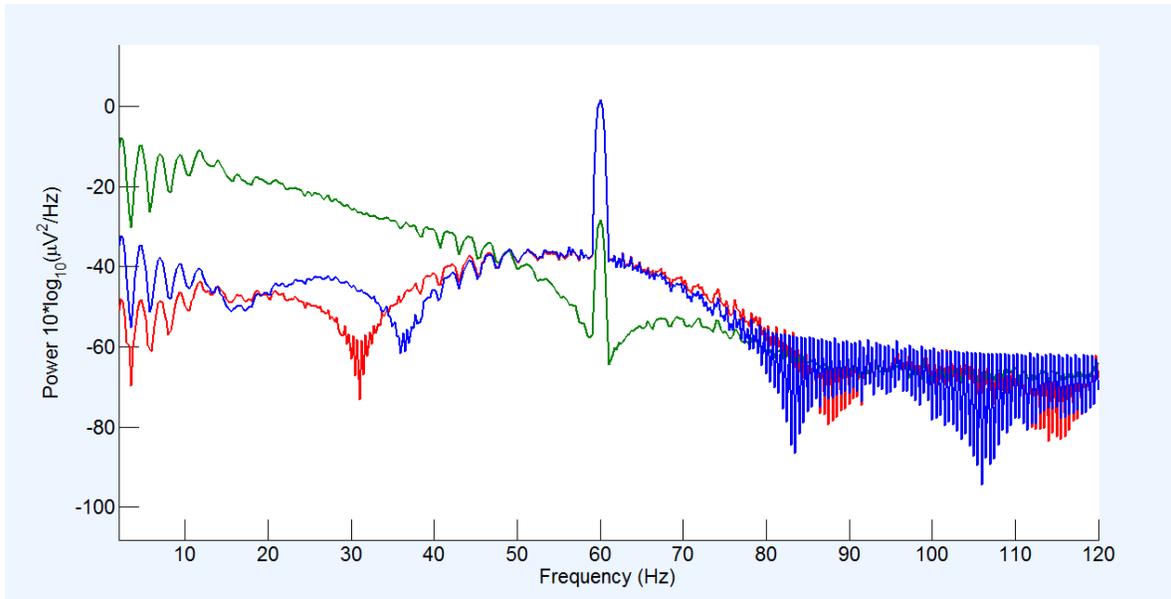


Figura 4.76 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]

Paso 9

Componente independiente 1

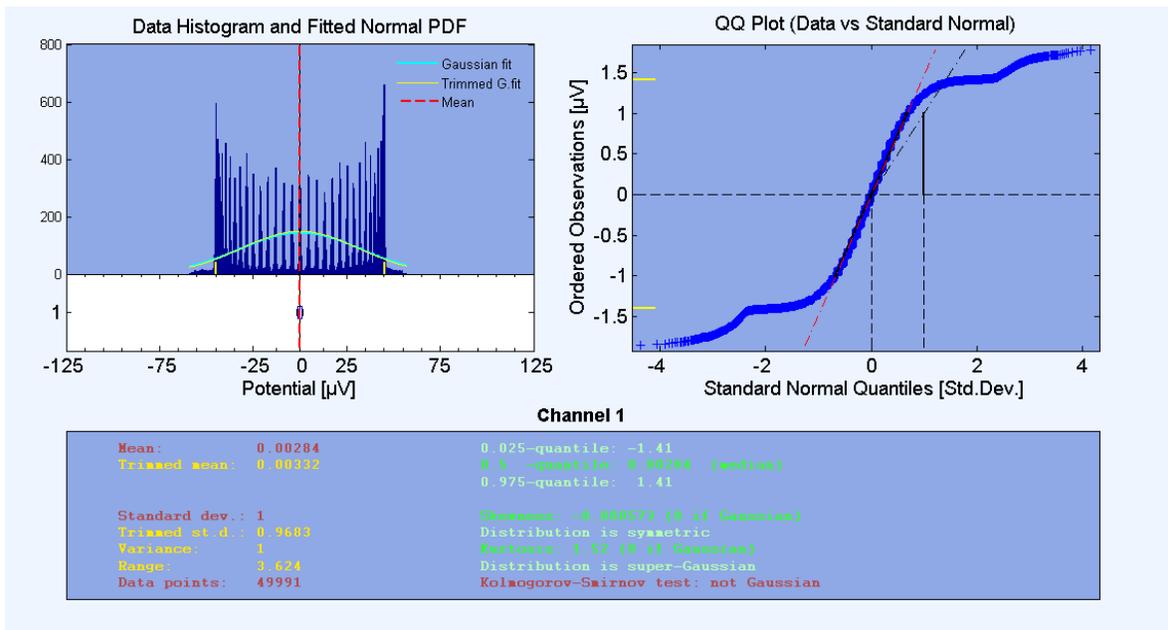


Figura 4.77 Estadísticas de la primera componente. [Fuente Propia].

Componente independiente 2



CAPÍTULO IV. RESULTADOS

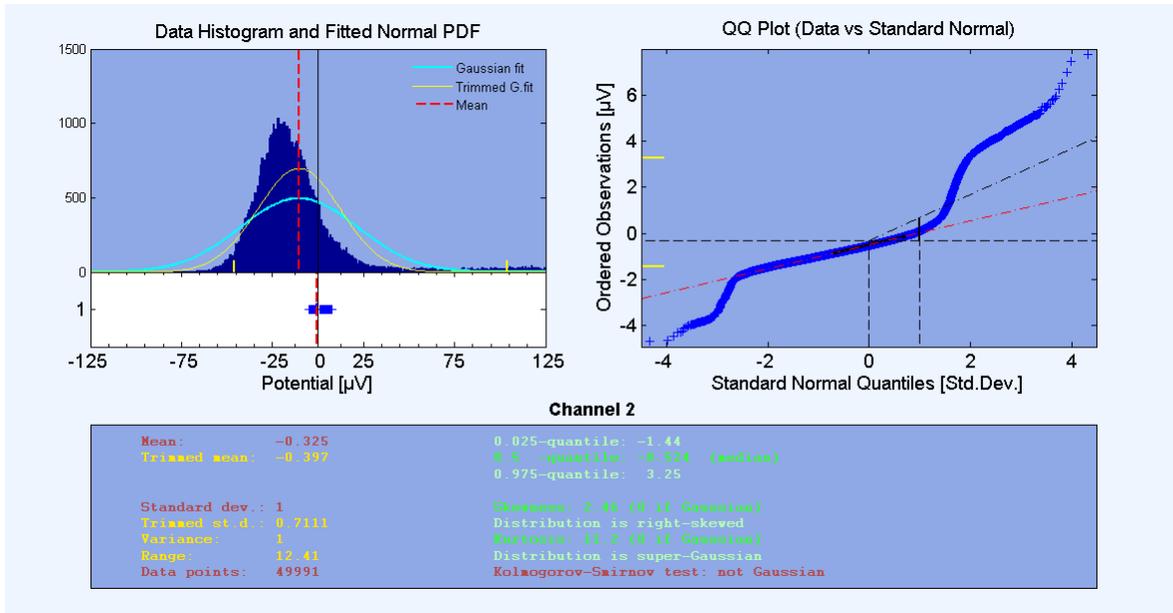


Figura 4.78 Estadísticas de la segunda componente. [Fuente Propia].

Componente independiente 3

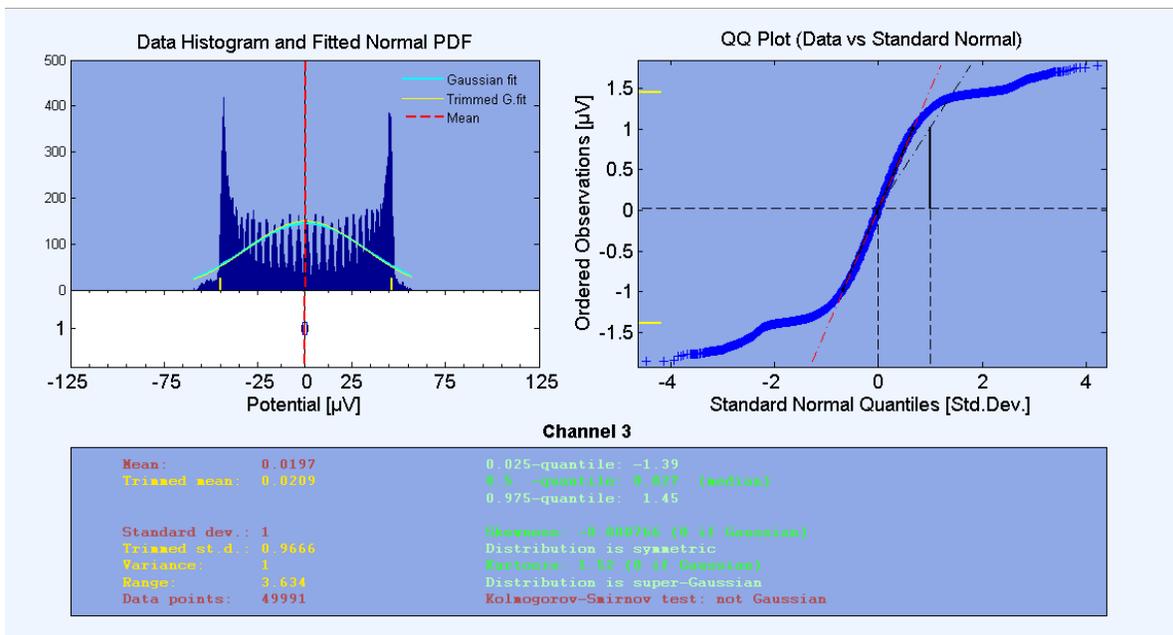


Figura 4.79 Estadísticas de la tercera componente. [Fuente Propia].

Experimento 5:

Registro 16773 del “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”.

Resultados después de aplicar los pasos del 1 al 9:

Paso 1:

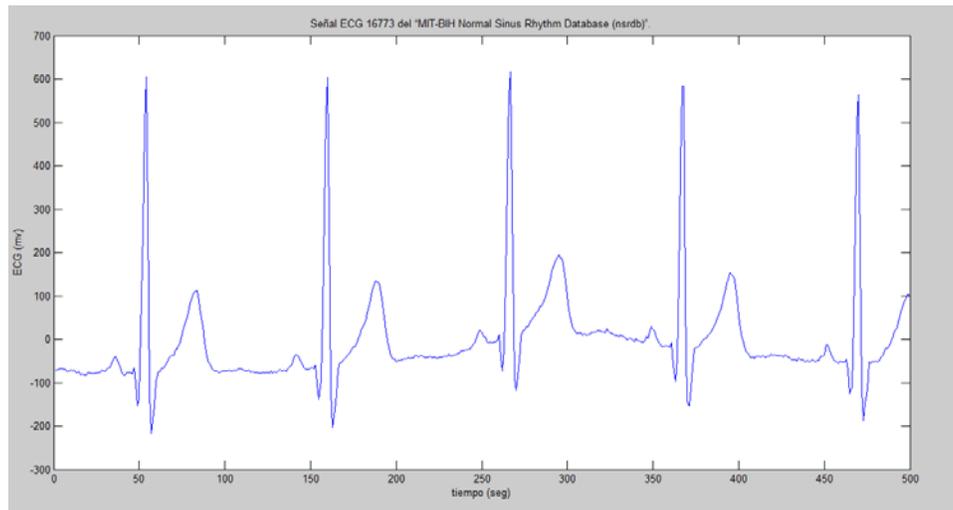


Figura 4.80 Forma de onda de la señal ECG 16773 de “MIT-BIH Normal Sinus Rhythm Database (nsrdb)”. [Fuente Propia]

Paso 2:

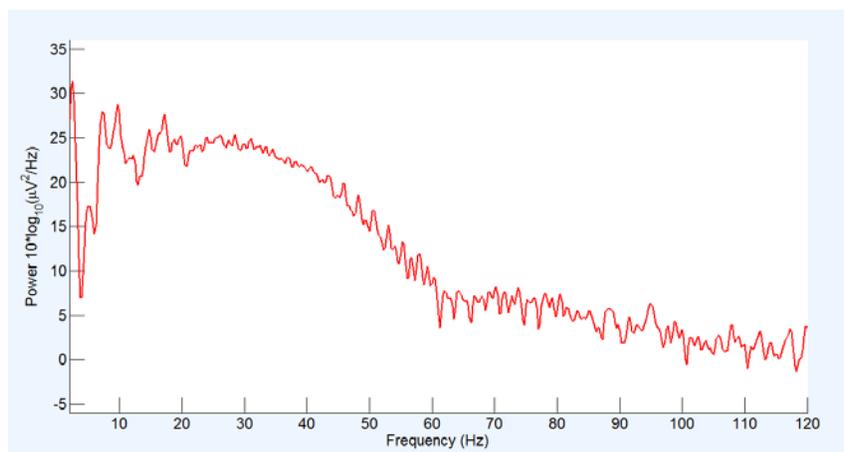


Figura 4.81 Densidad espectral de potencia de 16773. [Fuente Propia]

Paso 3:

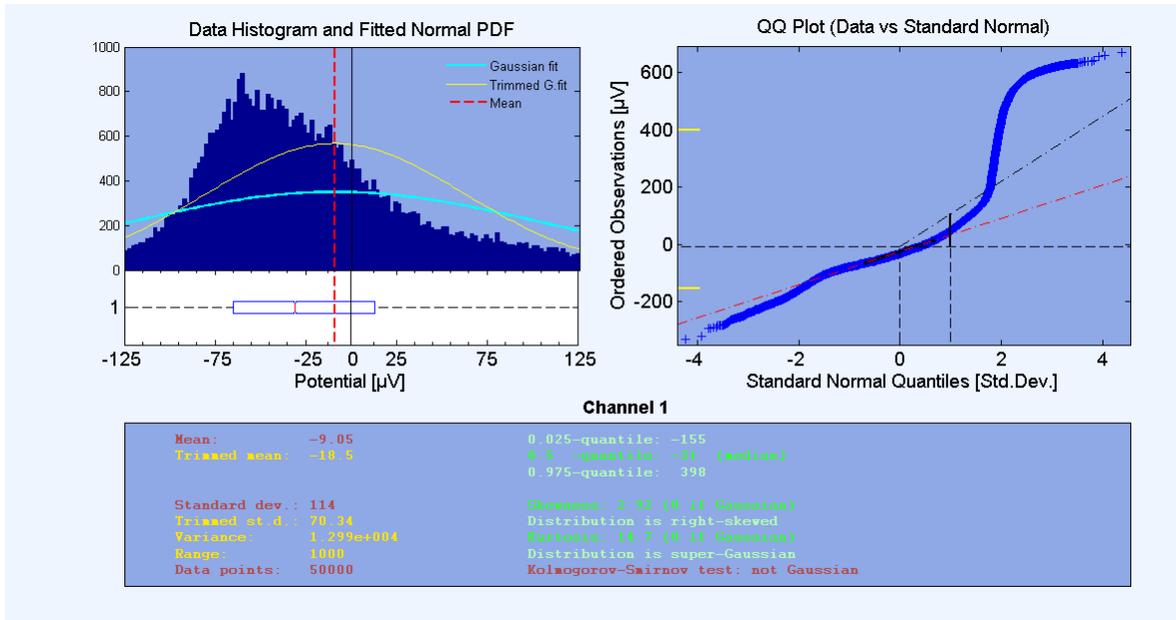


Figura 4.82 estadísticas de la señal ECG 16773, antes de agregarle ruido[Fuente Propia]

Paso 4:

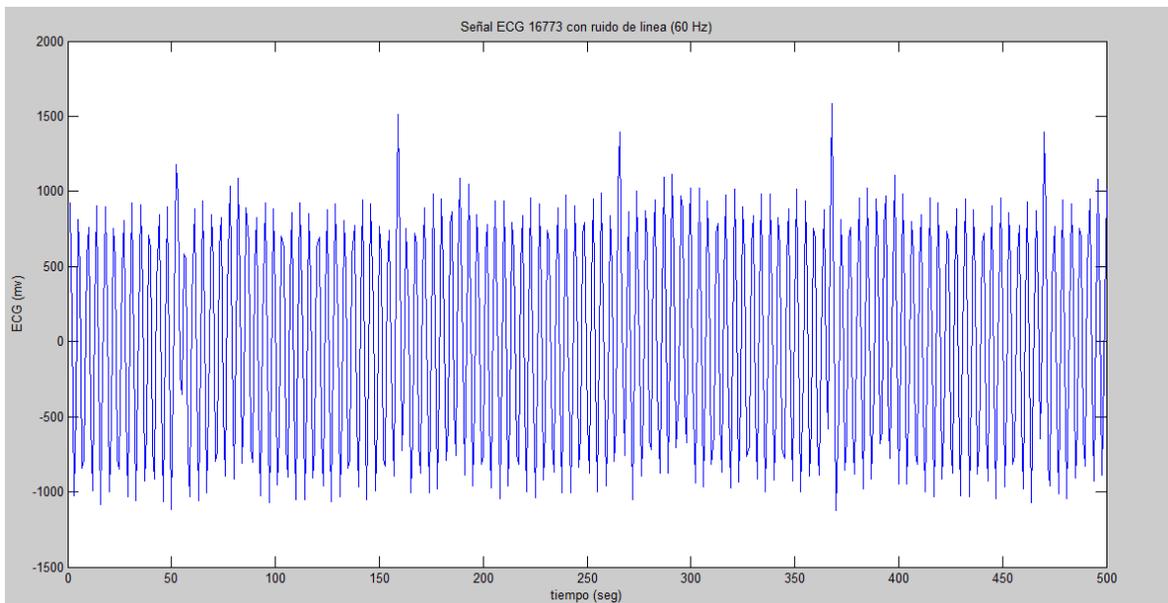


Figura 4.83 Onda que representa la señal ECG 16773 contaminada de ruido de línea. [Fuente Propia]

Paso 5:

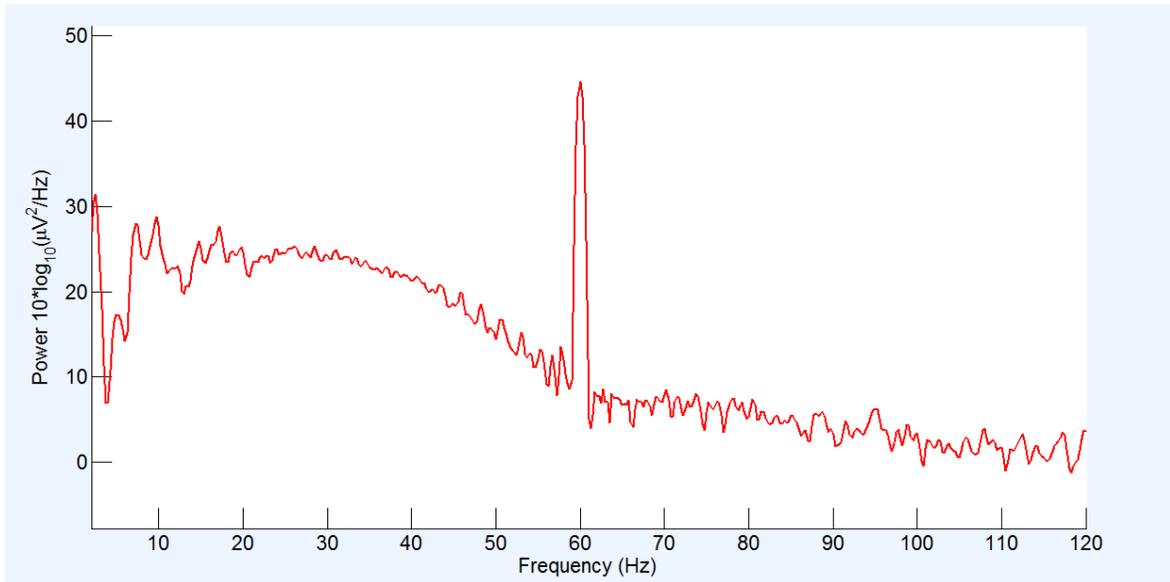


Figura 4.84 Densidad espectral de potencia de 16773 contaminada de ruido de línea. [Fuente Propia].

Paso 6:

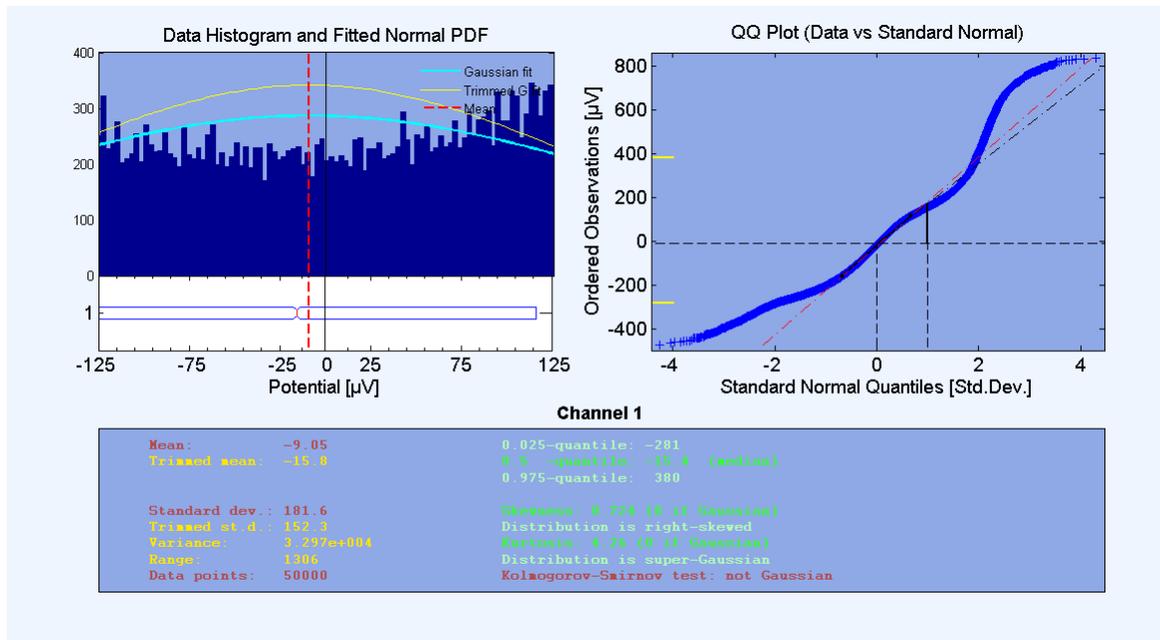


Figura 4.85 Estadísticas de la señal ECG 16773, después de agregarle ruido. [Fuente Propia]

Paso 7:

Reducción de la dimensión de la matriz de mezcla:

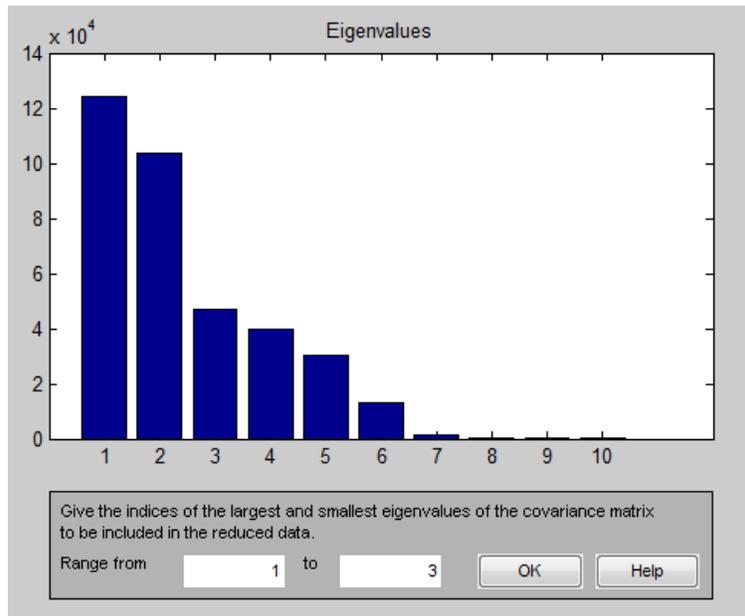


Figura 4.86 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]

Componentes independientes obtenidos después de aplicar el paquete fastICA, (ver figura 4.87).

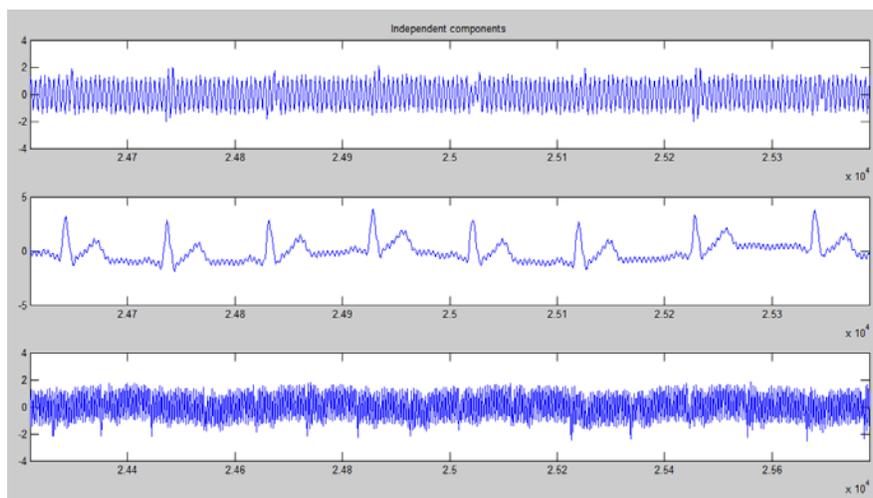


Figura 4.87 Componentes independientes o fuentes estimadas por fastICA. [Fuente Propia]

Paso 8:

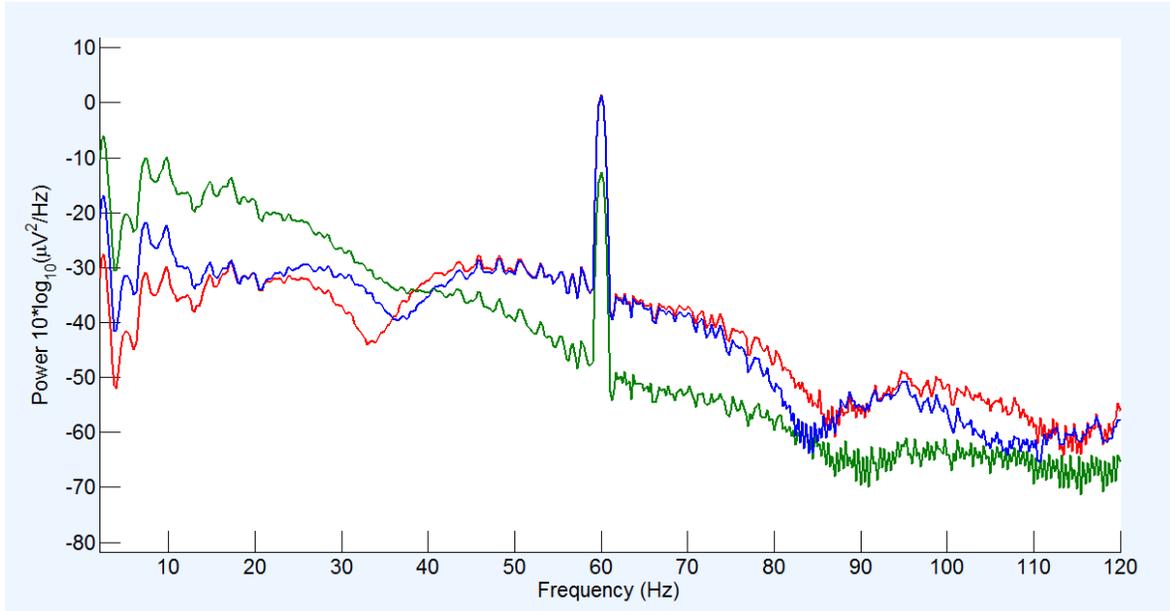


Figura 4.88 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]

Paso 9

Componente independiente 1

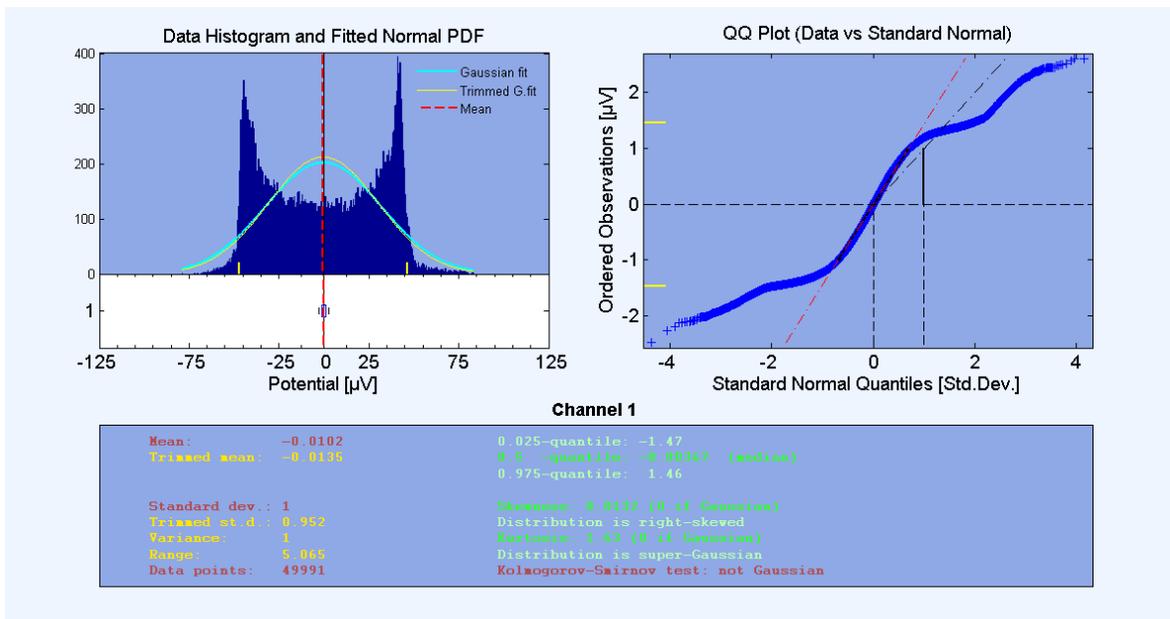


Figura 4.89 Estadísticas de la primera componente. [Fuente Propia]



Componente independiente 2

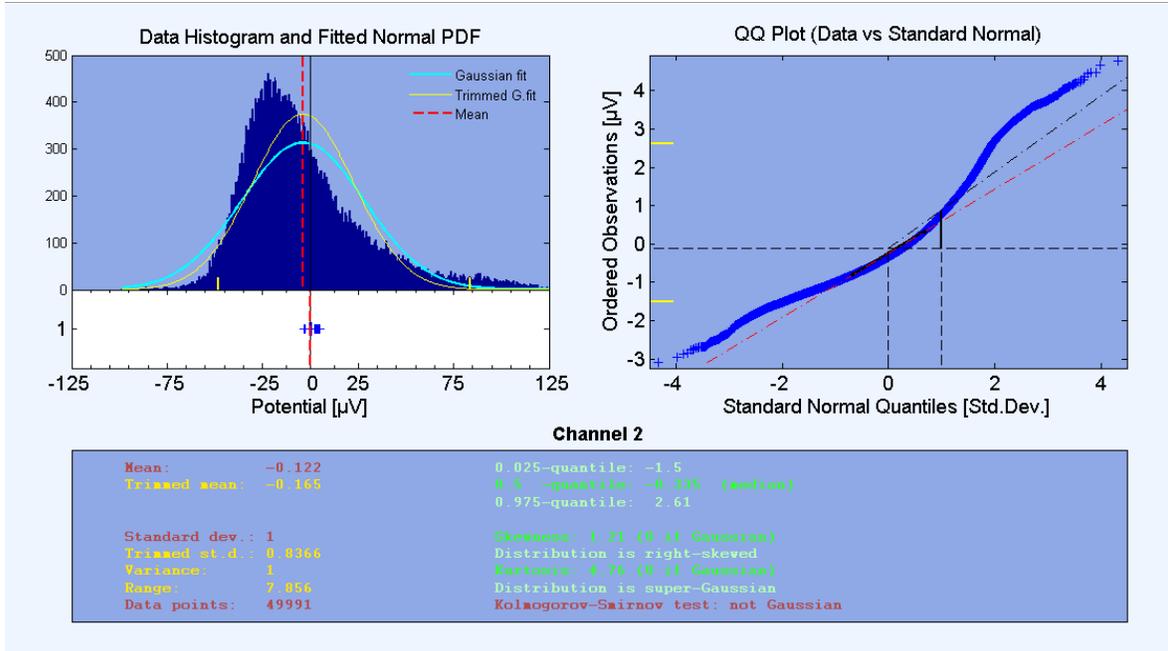


Figura 4.90 Estadísticas de la segunda componente. [Fuente Propia]

Componente independiente 3

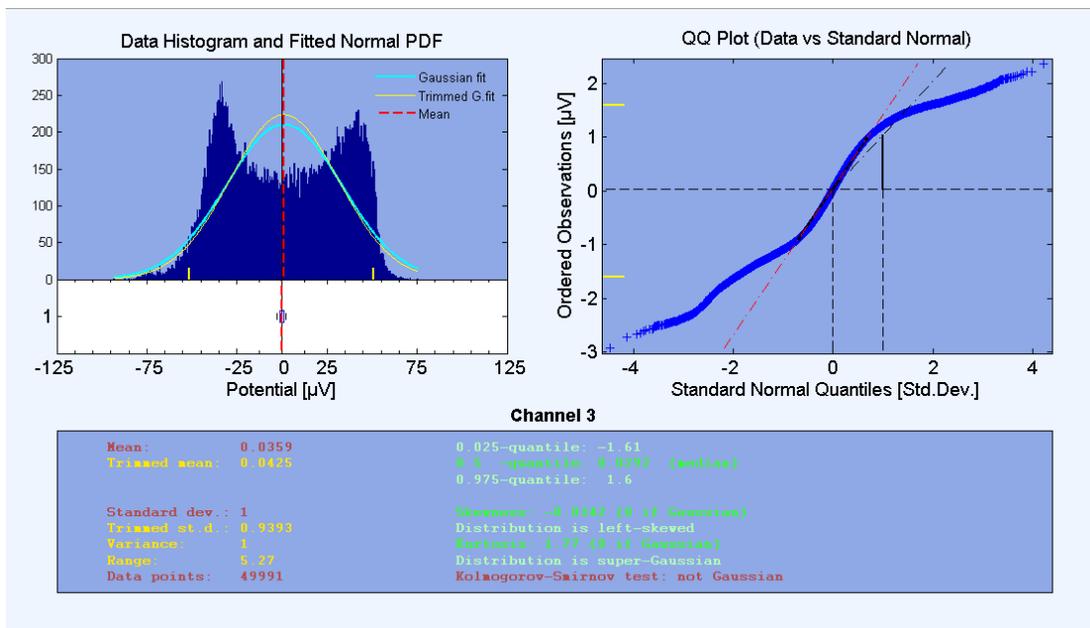


Figura 4.91 Estadísticas de la tercera componente. [Fuente Propia]

Experimento 6

Registro *twa00* de “T-Wave Alternans Challenge Database (*twadb*)”.

A diferencia de los experimentos anteriores, el ruido aplicado a las señales de este y los experimentos restantes será aleatorio.

Resultados después de aplicar los pasos del 1 al 9:

Paso 1:

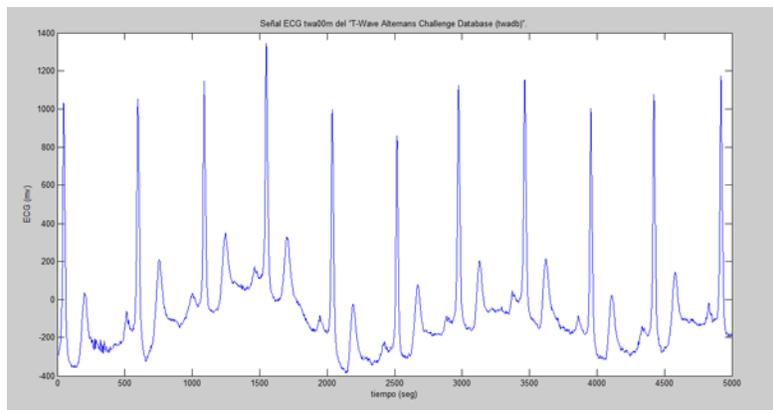


Figura 4.92 Forma de onda de la señal ECG *twa00* de “T-Wave Alternans Challenge Database (*twadb*)”.

[Fuente Propia]

Paso 2:

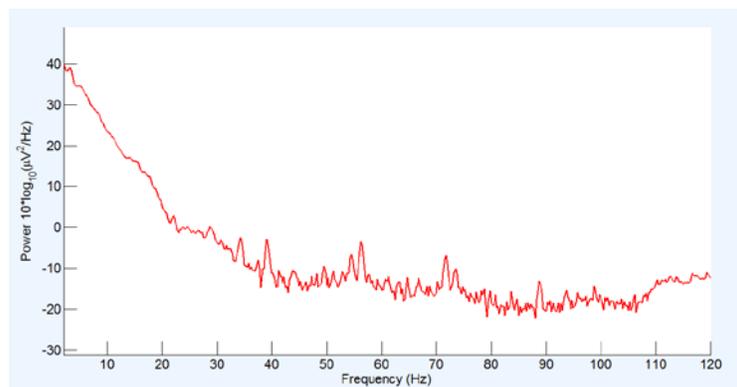


Figura 4.93 Densidad espectral de potencia de *twa00*. [Fuente Propia]

Paso 3:

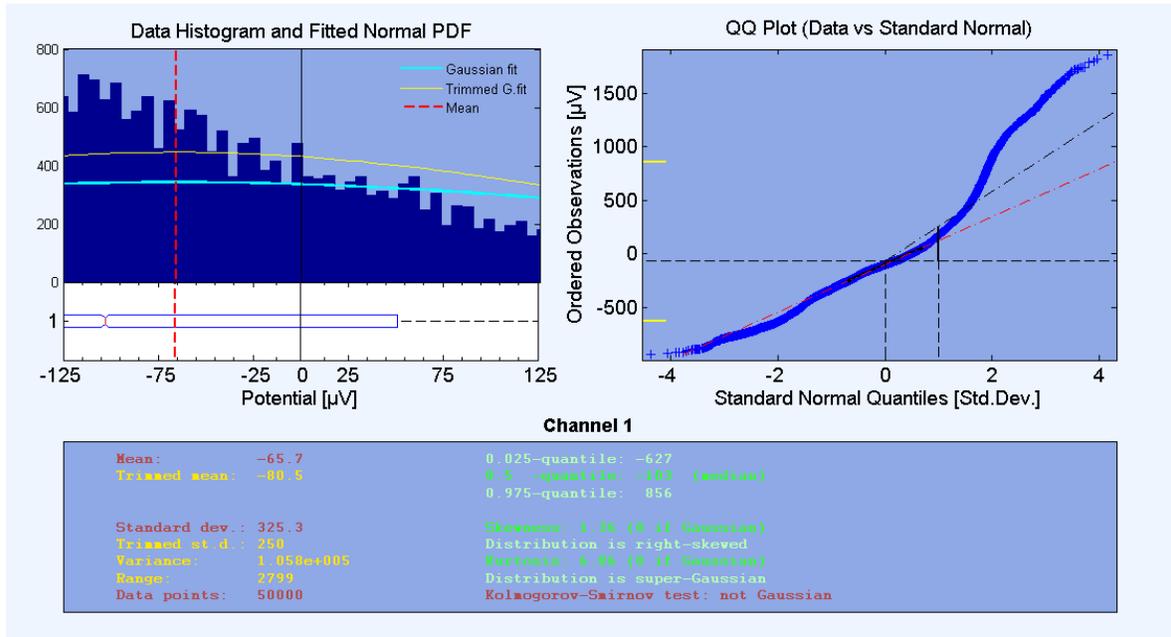


Figura 4.94 Estadísticas de la señal ECG *twa00*, antes de agregarle ruido. [Fuente Propia]

Generación de la señal de ruido, en este caso es ruido aleatorio su forma de onda es la mostrada en la figura 4.95.

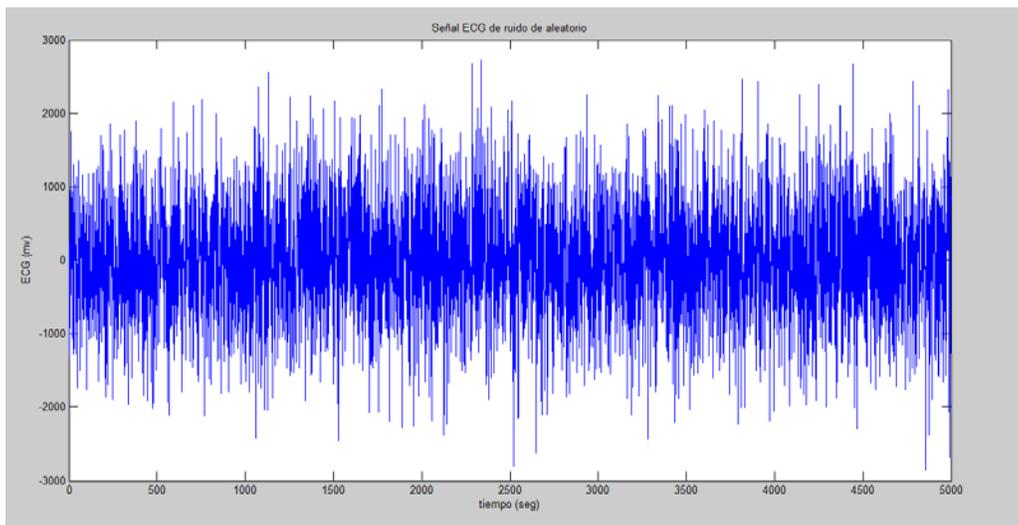


Figura 4.95 Onda que representa la señal de ruido aditivo a la señal ECG *twa00*. [Fuente Propia]

Paso 4:

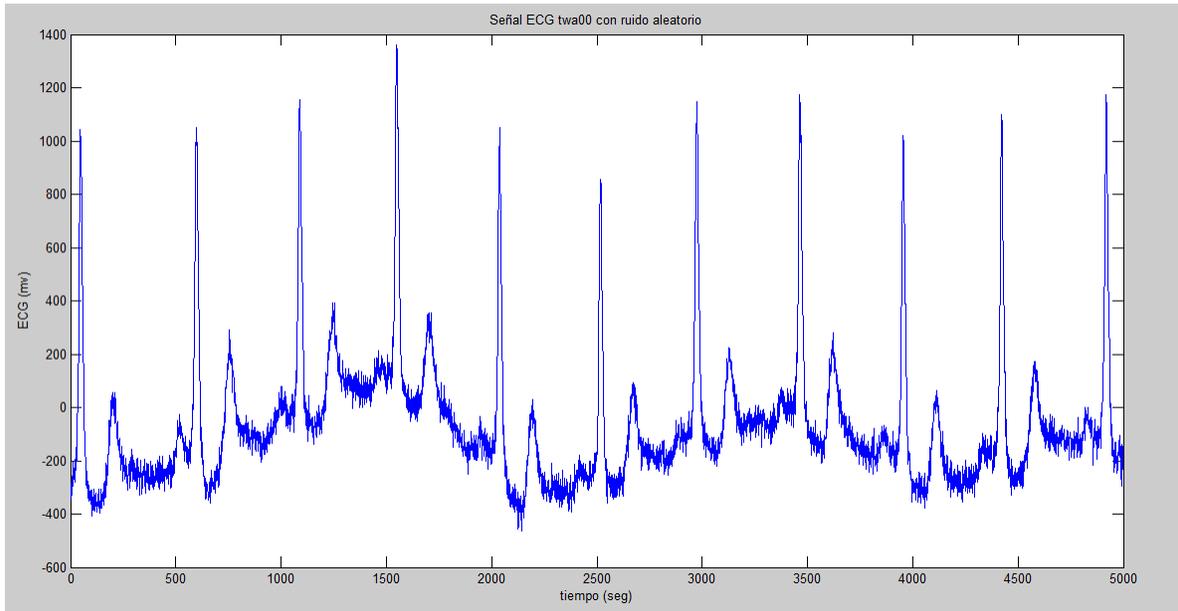


Figura 4.96 Onda que representa la señal ECG *twa00*, contaminada de ruido aleatorio. [Fuente Propia]

Paso 5:

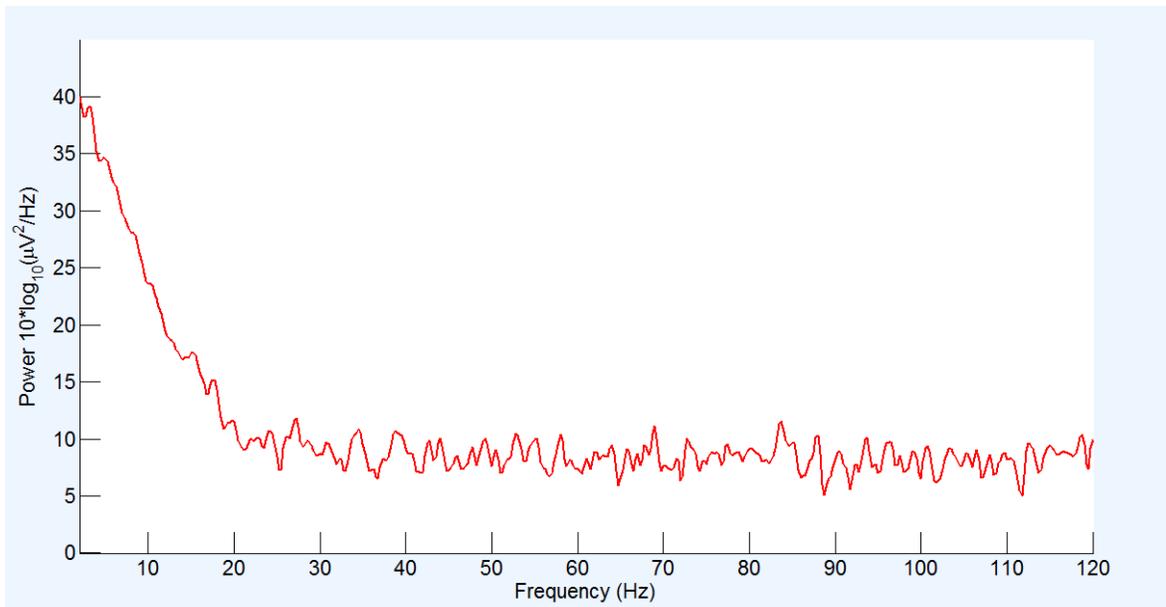


Figura 4.97 Densidad espectral de potencia de *twa00*, contaminada de ruido aleatorio. [Fuente Propia]

Paso 6:

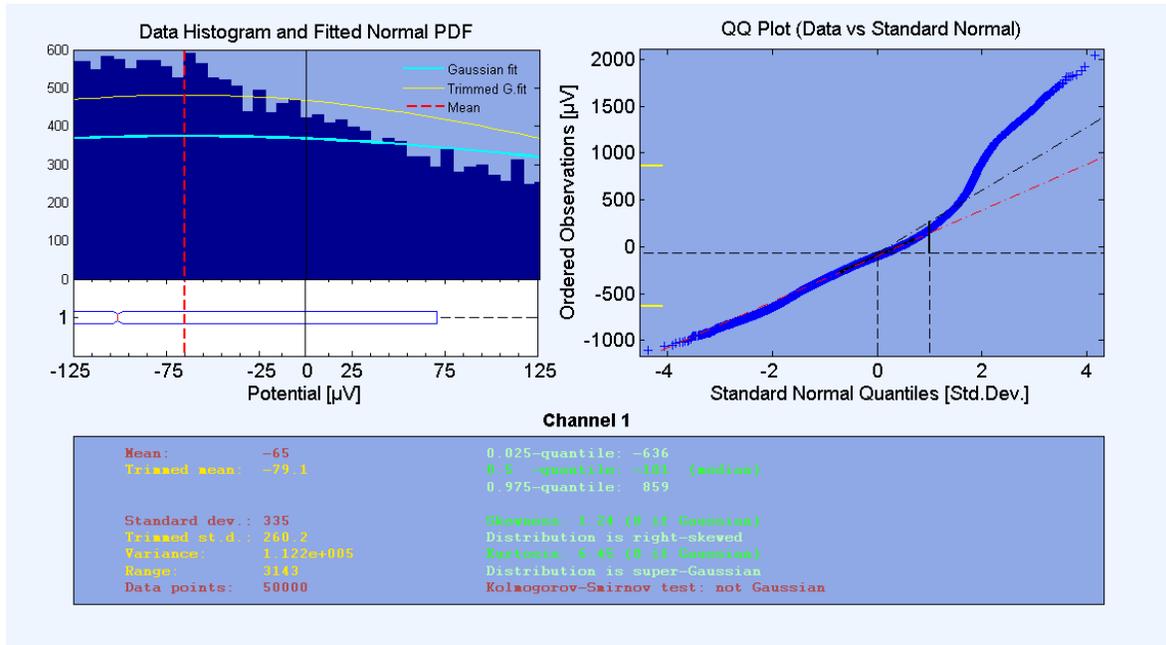


Figura 4.98 Estadísticas de la señal ECG *tw00*, después de agregarle ruido. [Fuente Propia]

Paso 7:

Reducción de la dimensión de la matriz de mezcla:

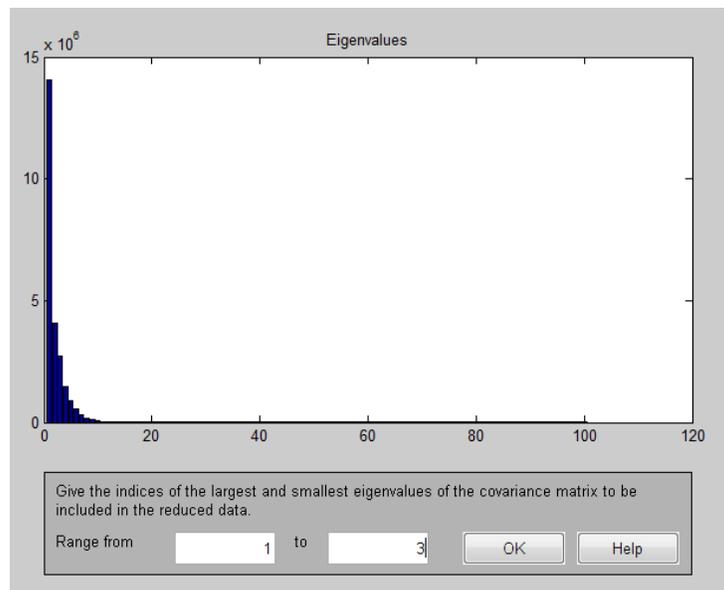


Figura 4.99 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]



Componentes independientes obtenidos después de aplicar el paquete fastICA, (ver figura 4.100).

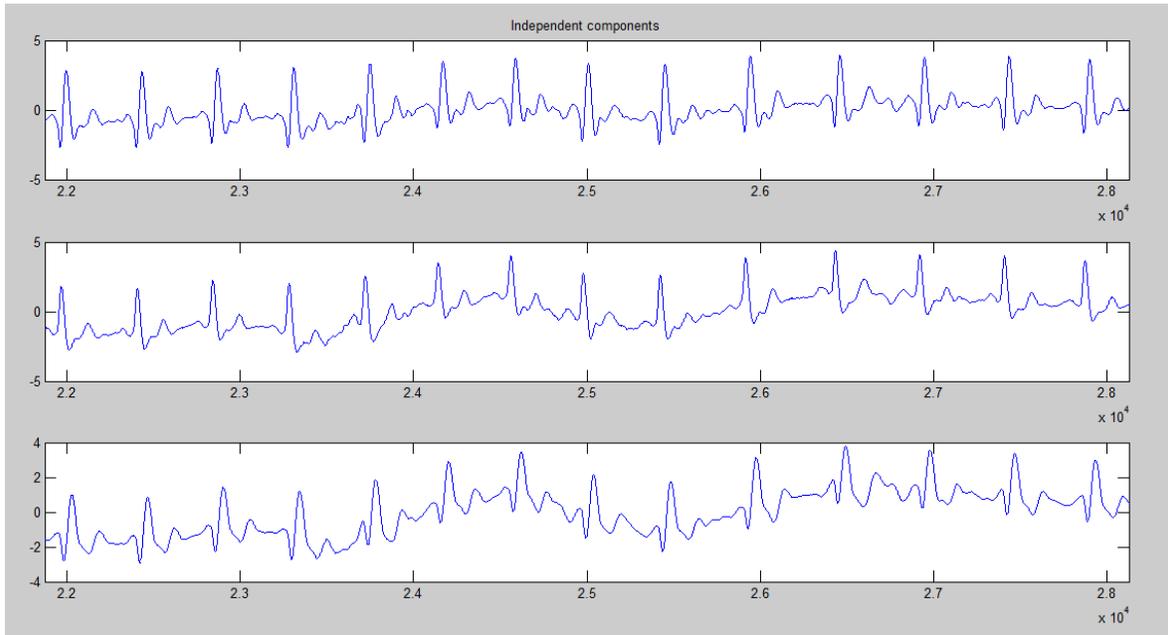


Figura 4.100 Componentes independientes o fuentes estimadas por fastICA. [Fuente Propia]

Paso 8:

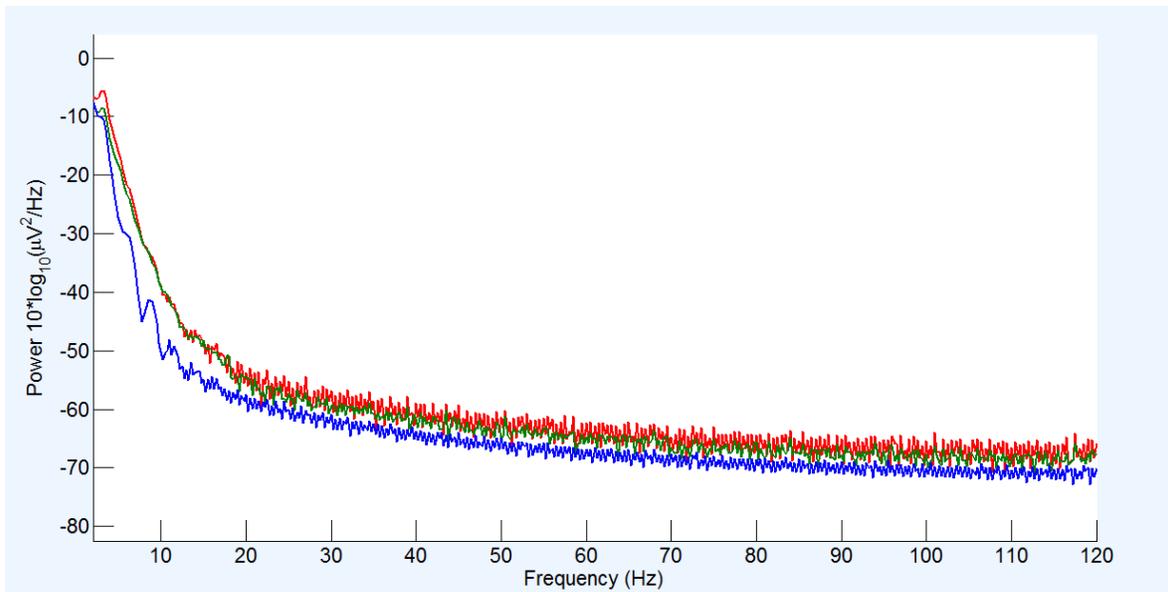


Figura 4.101 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]



Paso 9

Componente independiente 1

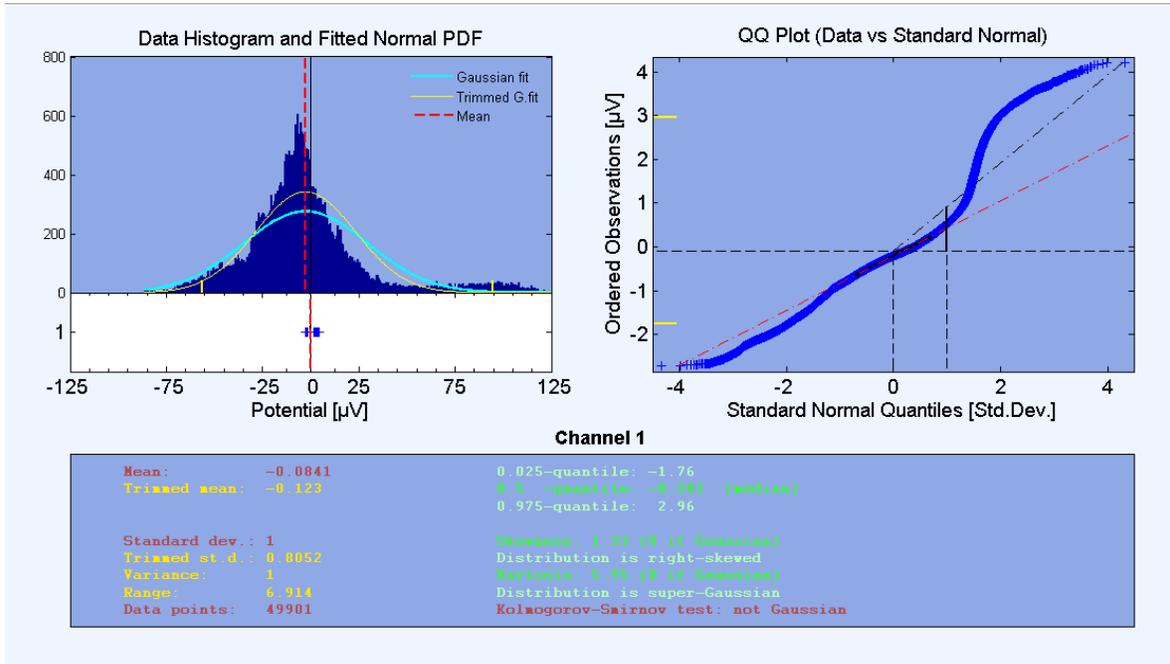


Figura 4.102 Estadísticas de la primera componente. [Fuente Propia]

Componente independiente 2

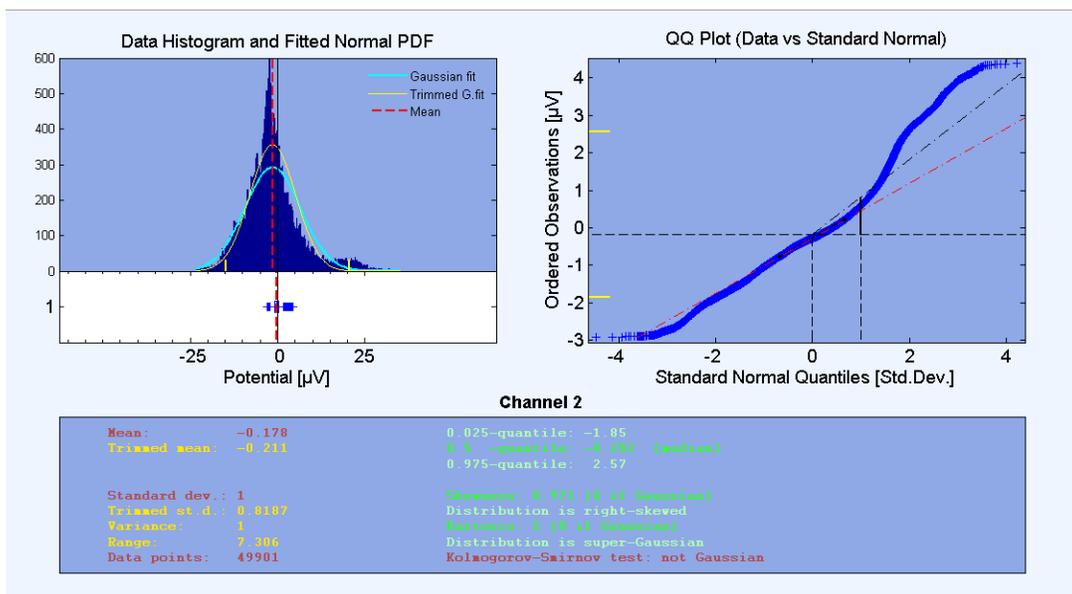




Figura 4.103 Estadísticas de la segunda componente. [Fuente Propia]

Componente independiente 3

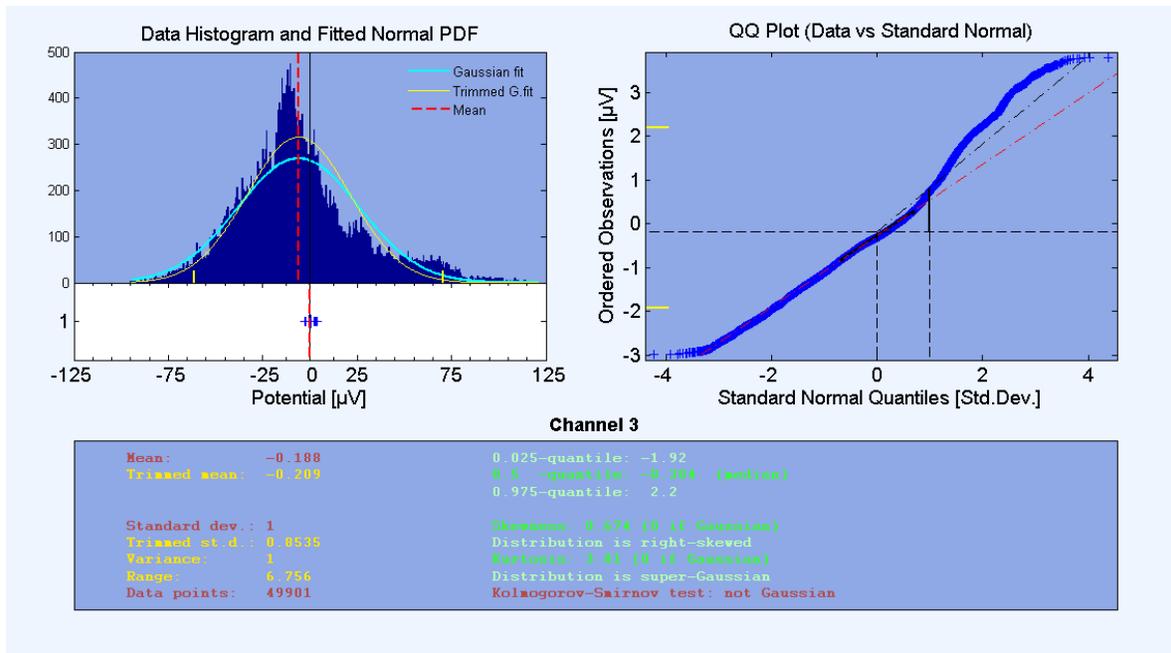


Figura 4.104 Estadísticas de la tercera componente. [Fuente Propia]



Experimento 7:

Registro **twa02** de “T-Wave Alternans Challenge Database (**twadb**)”.

Resultados después de aplicar los pasos del 1 al 9:

Paso 1:

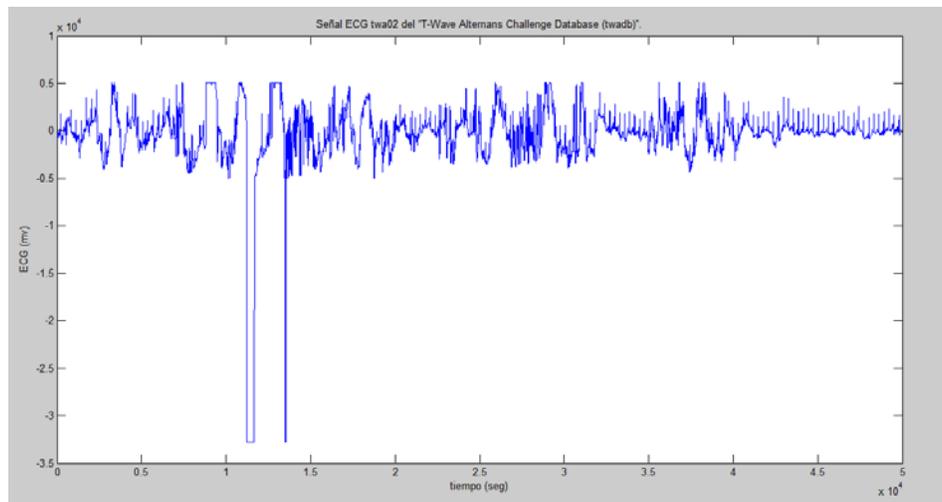


Figura 4.105 Forma de onda de la señal ECG **twa02** de “T-Wave Alternans Challenge Database (**twadb**)”. [Fuente Propia]

Paso 2:

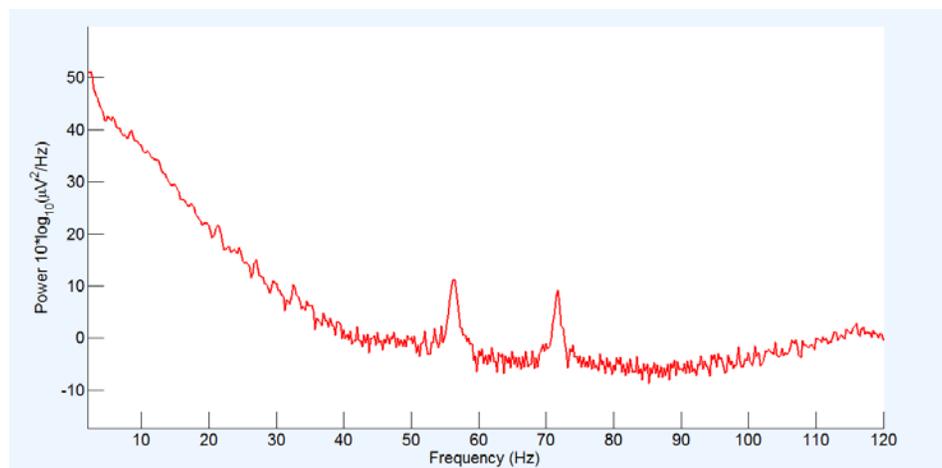


Figura 4.106 Densidad espectral de potencia de **twa02**. [Fuente Propia]

Paso 3:

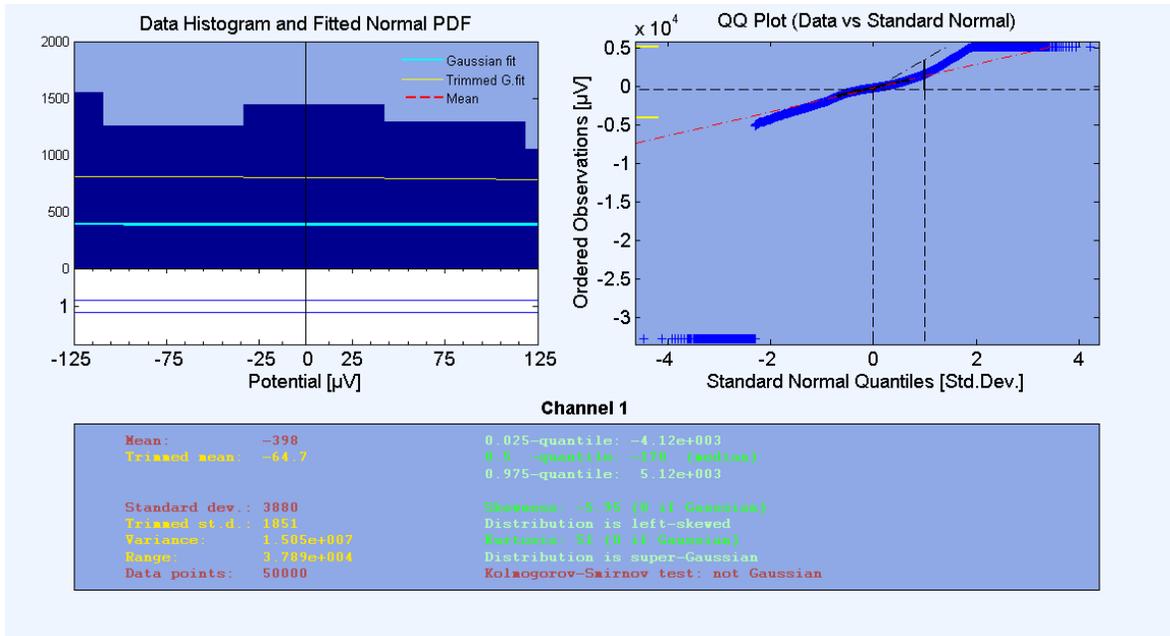


Figura 4.107 Estadísticas de la señal ECG *twa02*, antes de agregarle ruido. [Fuente Propia]

Paso 4:

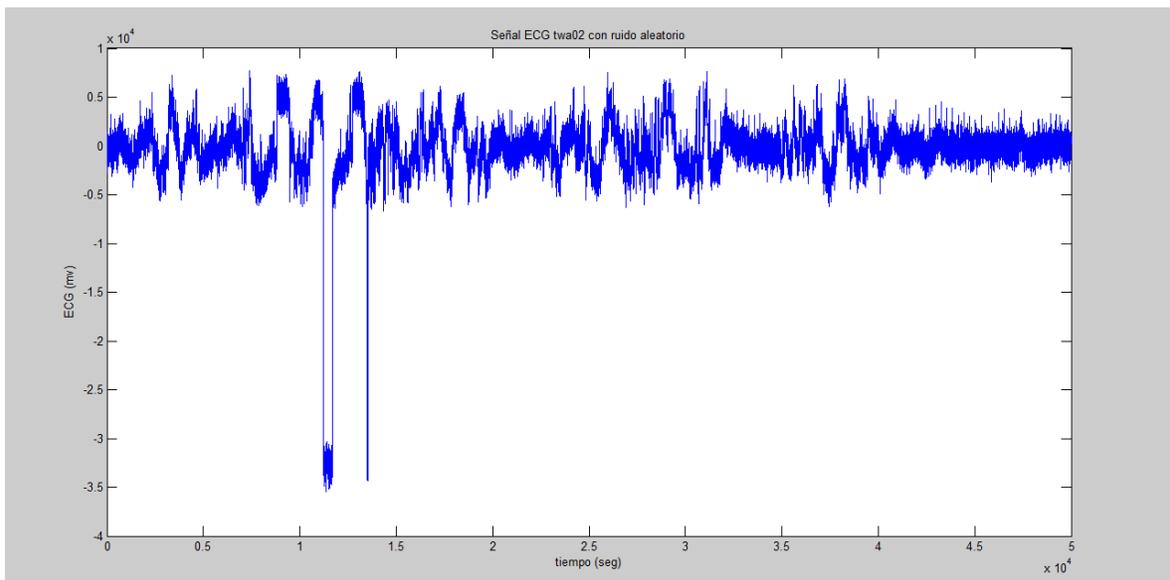


Figura 4.108 Onda que representa la señal ECG *twa02* contaminada de ruido aleatorio. [Fuente Propia]

Paso 5:

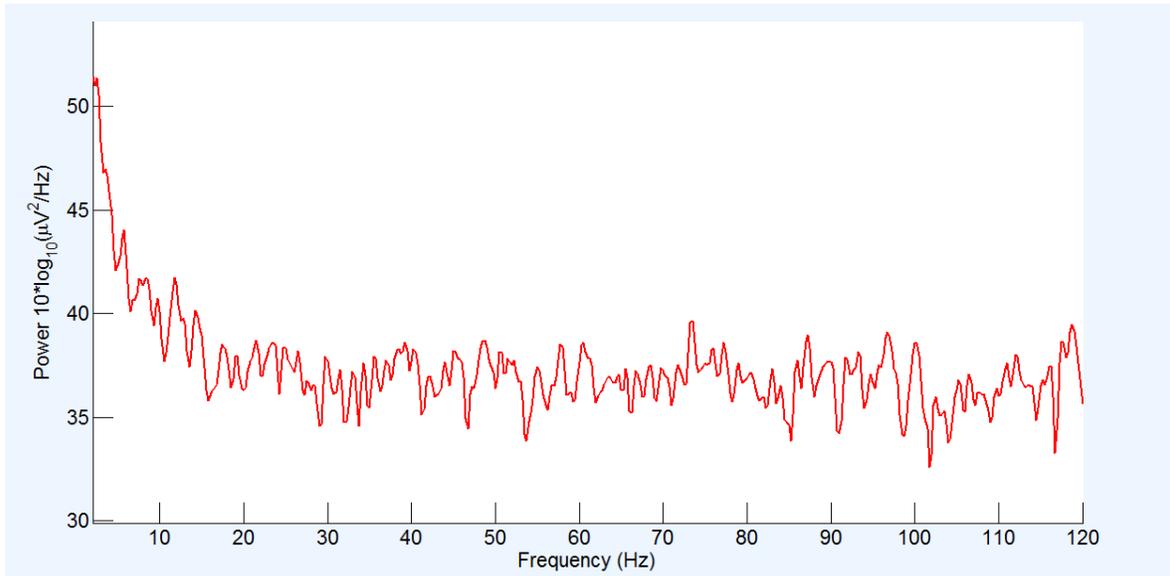


Figura 4.109 Densidad espectral de potencia de *twa02*, contaminada de ruido aleatorio. [Fuente Propia]

Paso 6:

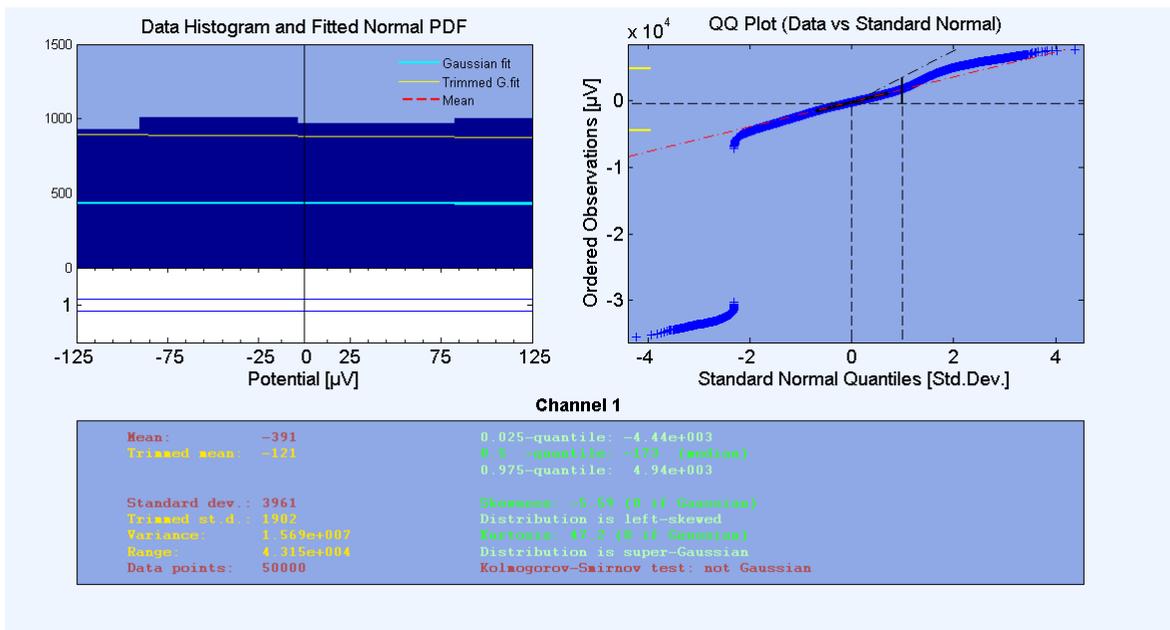


Figura 4.110 Estadísticas de la señal ECG *twa02*, después de agregarle ruido. [Fuente Propia]

Paso 7:

Reducción de la dimensión de la matriz de mezcla:

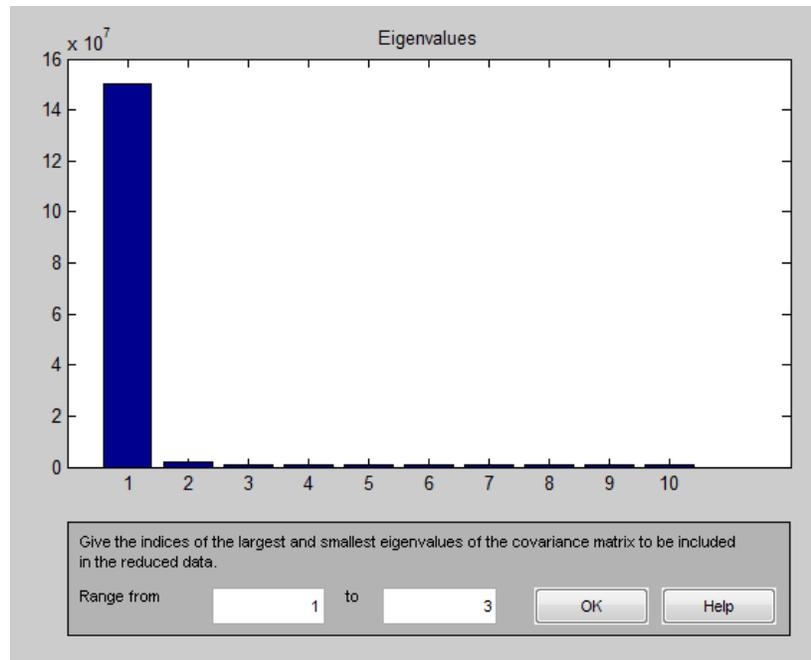


Figura 4.111 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]

Componentes independientes obtenidos después de aplicar el paquete fastICA, (ver figura 4.112).

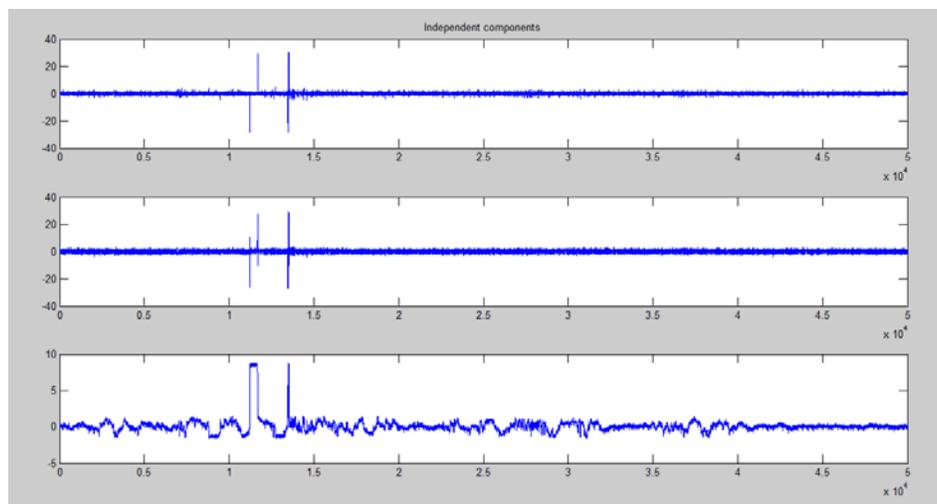


Figura 4.112 Componentes impedientes o fuentes estimadas por fastICA. [Fuente Propia]

Paso 8:

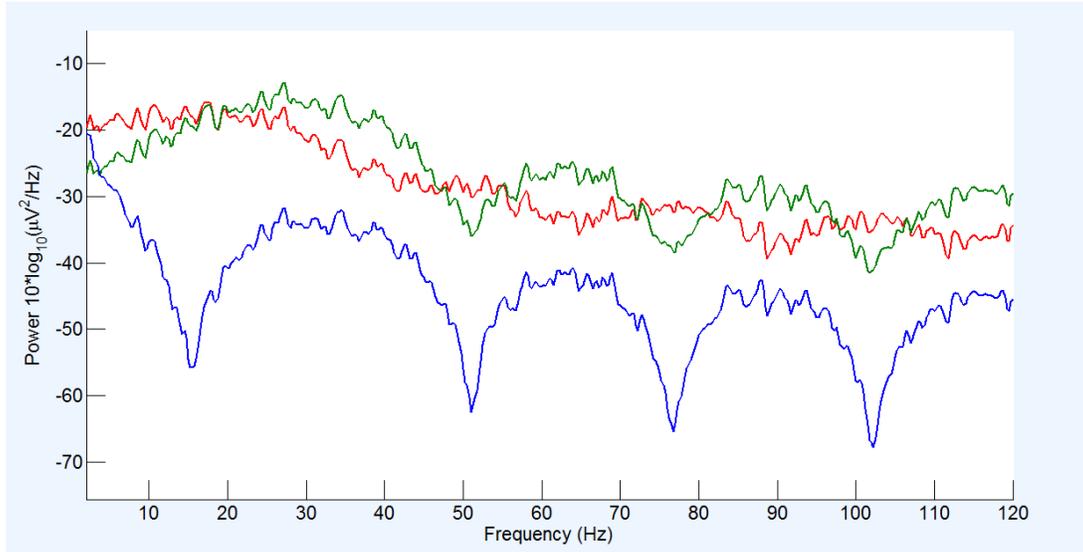


Figura 4.113 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]

Paso 9

Componente independiente 1

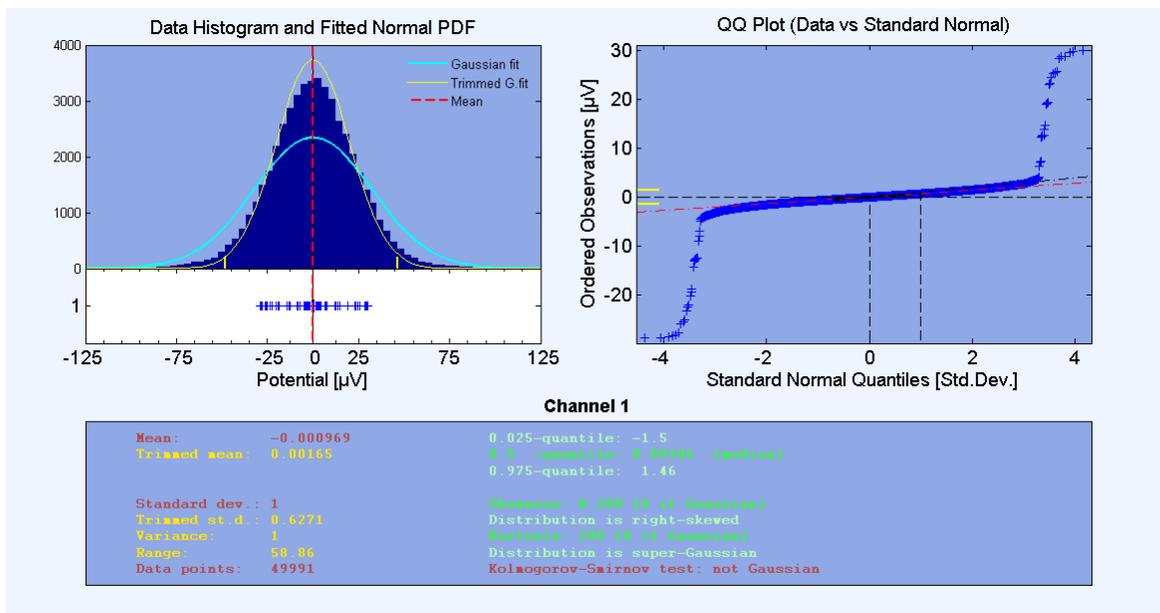


Figura 4.114 Estadísticas de la primera componente. [Fuente Propia]

Componente independiente 2

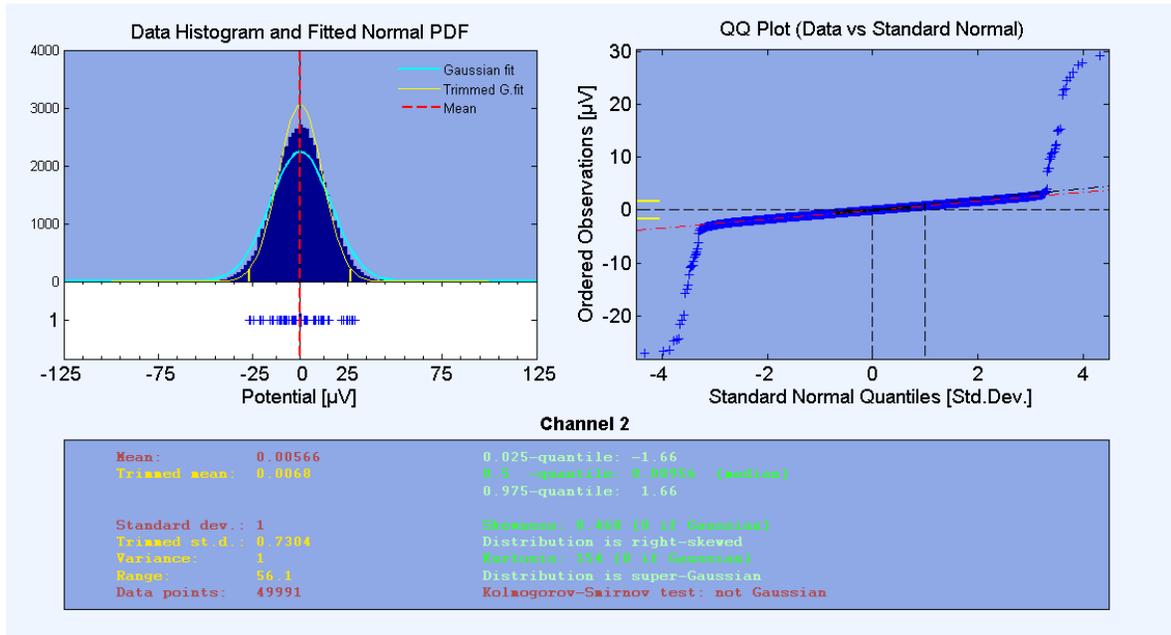


Figura 4.115 Estadísticas de la segunda componente. [Fuente Propia]

Componente independiente 3

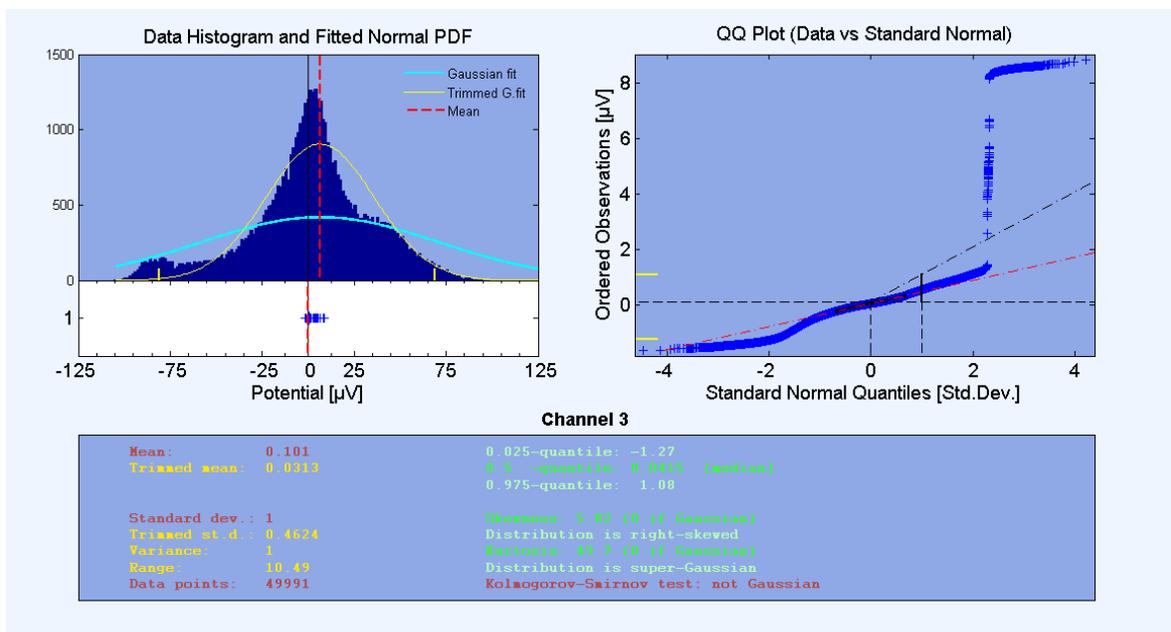


Figura 4.116 Estadísticas de la tercera componente. [Fuente Propia]

Experimento 8:

Registro *twa05* de “T-Wave Alternans Challenge Database (*twadb*)”.

Resultados después de aplicar los pasos del 1 al 9:

Paso 1:

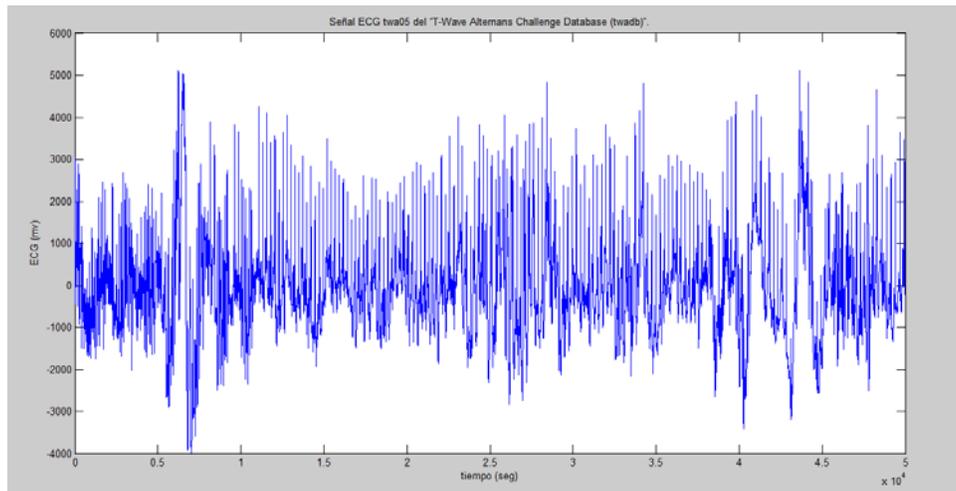


Figura 4.117 Forma de onda de la señal ECG *twa05* de “T-Wave Alternans Challenge Database (*twadb*)”.

[Fuente Propia]

Paso 2:

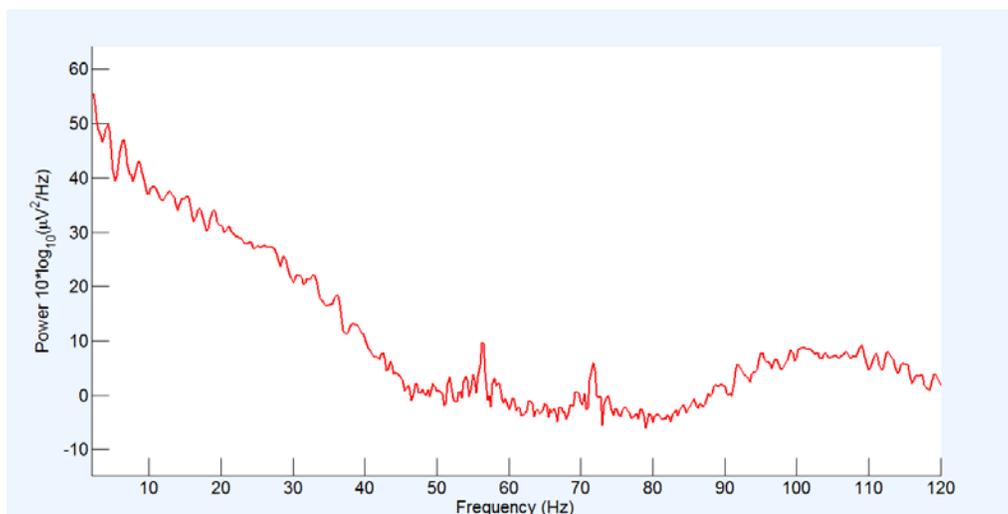


Figura. 4.118 Densidad espectral de potencia de *twa05*. [Fuente Propia]

Paso 3:

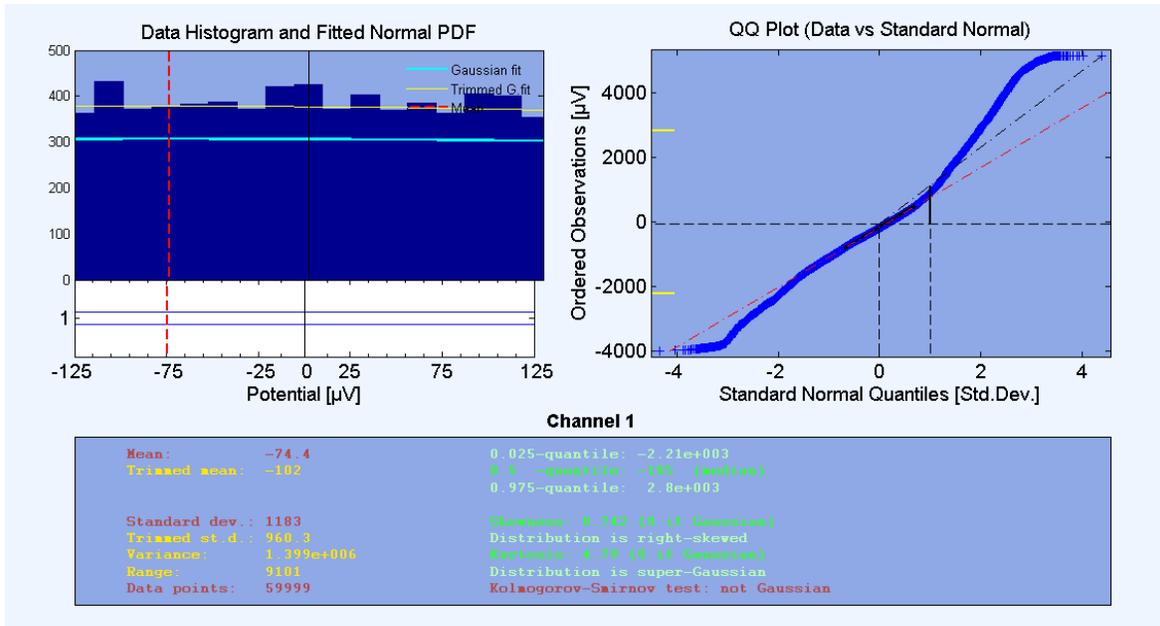


Figura 4.119 Estadísticas de la señal ECG *twa05*, antes de agregarle ruido. [Fuente Propia]

Paso 4:

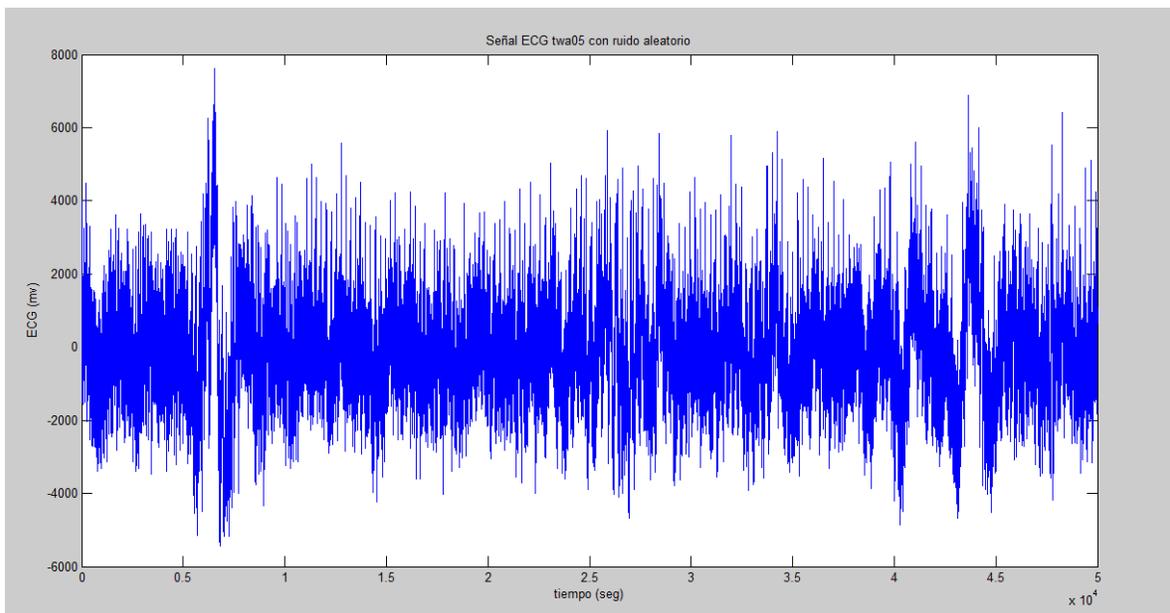


Figura 4.120 Onda que representa la señal ECG *twa05* contaminada de ruido aleatorio.

Paso 5:

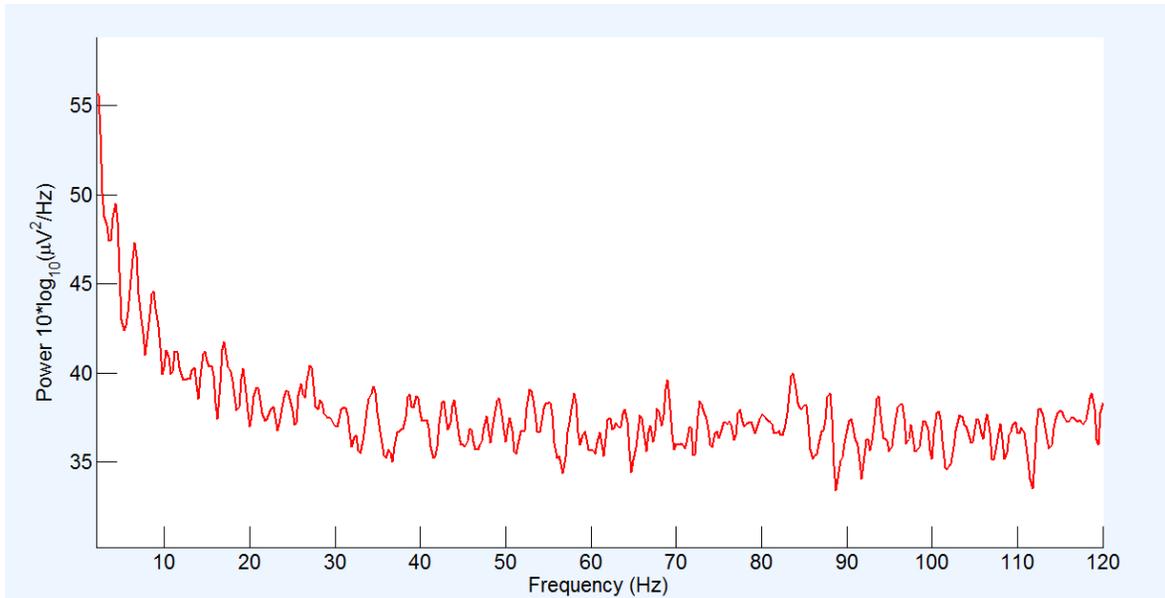


Figura 4.121 Densidad espectral de potencia de *twa05* contaminada de aleatorio. [Fuente Propia]

Paso 6:

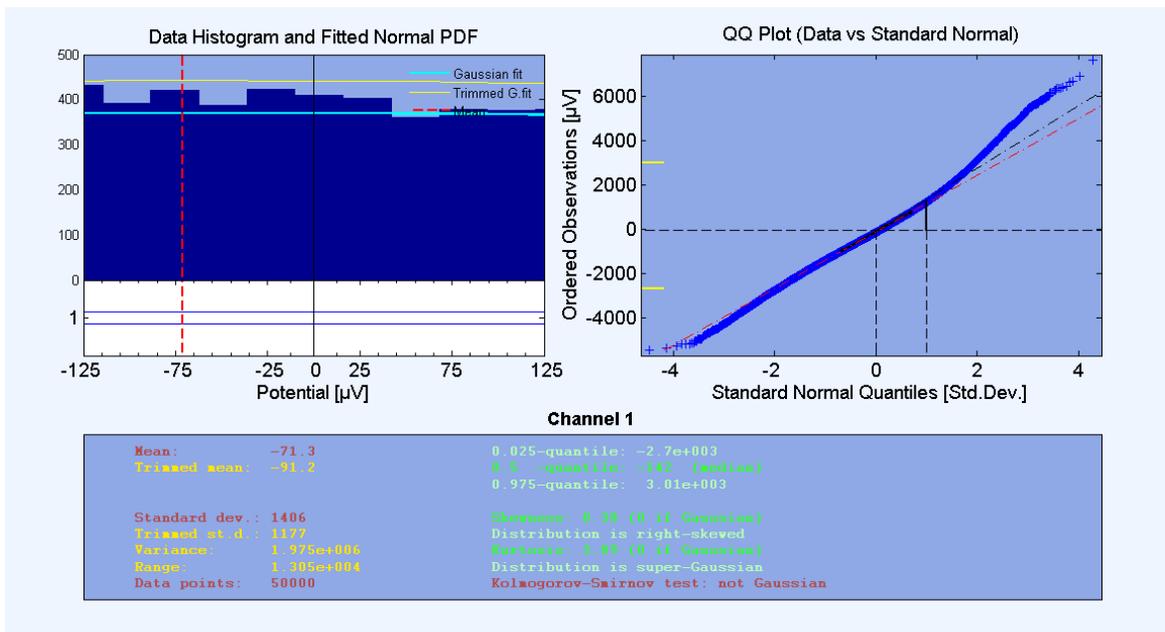


Figura 4.122 Estadísticas de la señal ECG *twa05*, después de agregarle ruido. [Fuente Propia]

Paso 7:

Reducción de la dimensión de la matriz de mezcla:

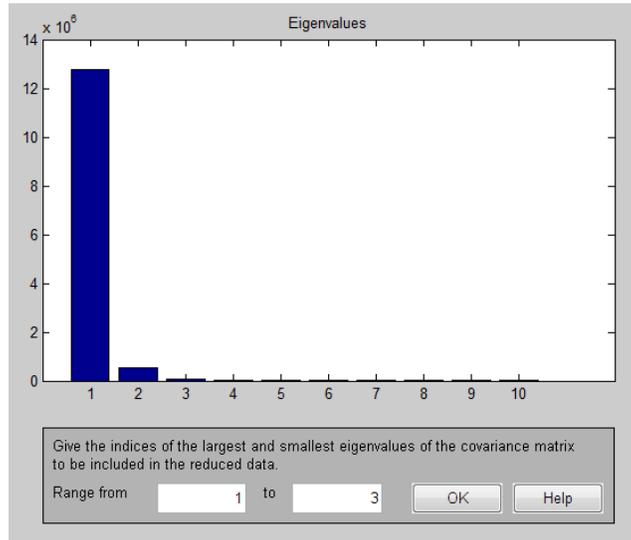


Figura 4.123 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]

Componentes independientes obtenidos después de aplicar el paquete fastICA, (ver figura 4.124)

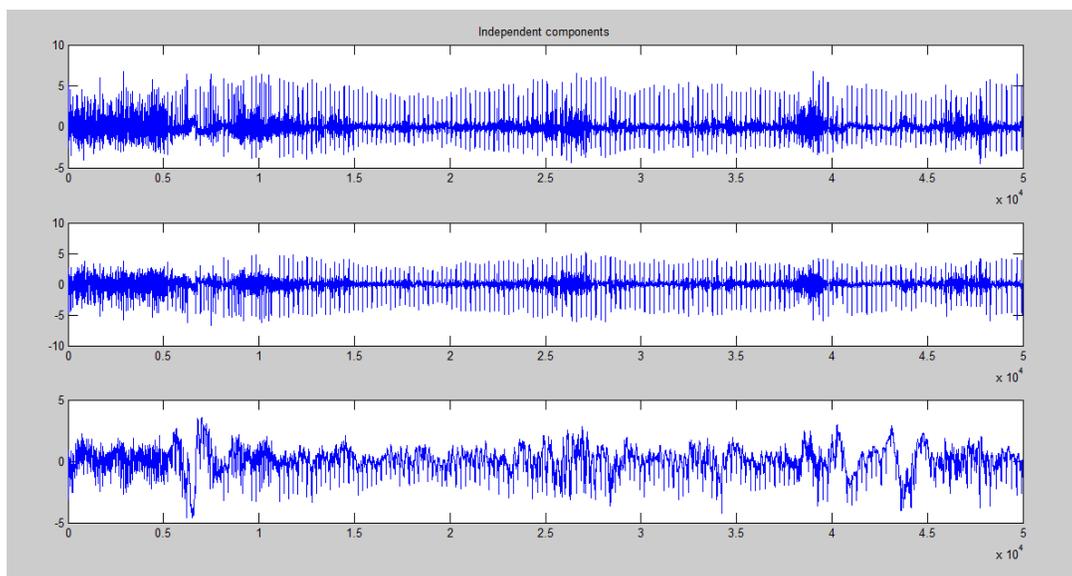


Figura 4.124 Componentes independientes o fuentes estimadas por fastICA. [Fuente Propia]

Paso 8:

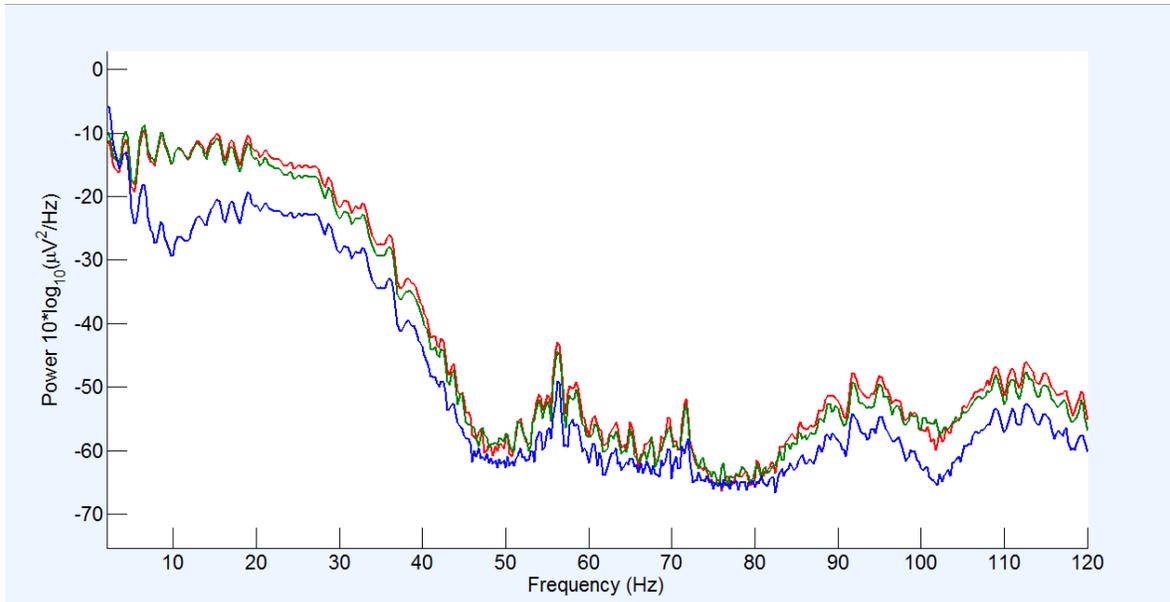


Figura 4.125 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]

Paso 9 Componente independiente 1

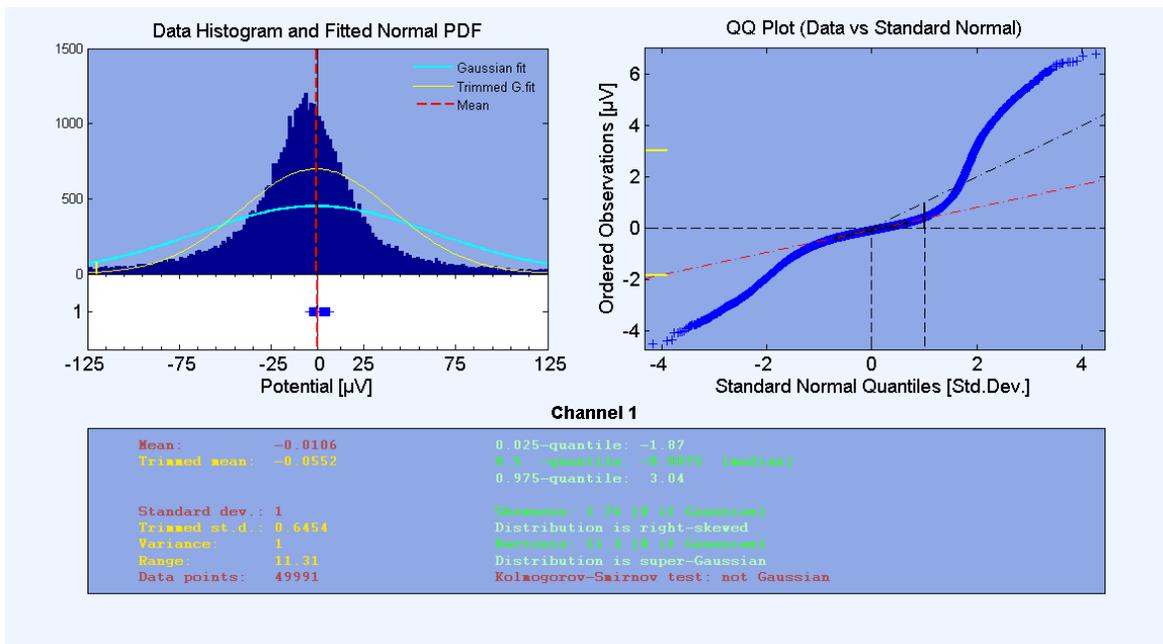


Figura 4.126 Estadísticas de la primera componente. [Fuente Propia]



Componente independiente 2

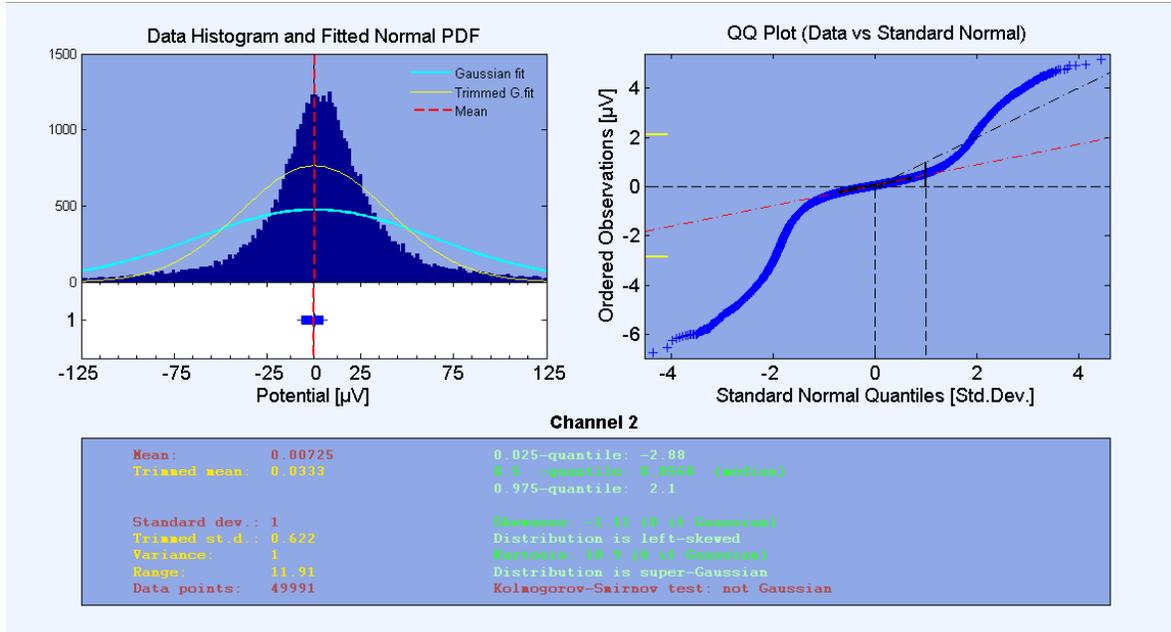


Figura 4.127 Estadísticas de la segunda componente. [Fuente Propia]

Componente independiente 3

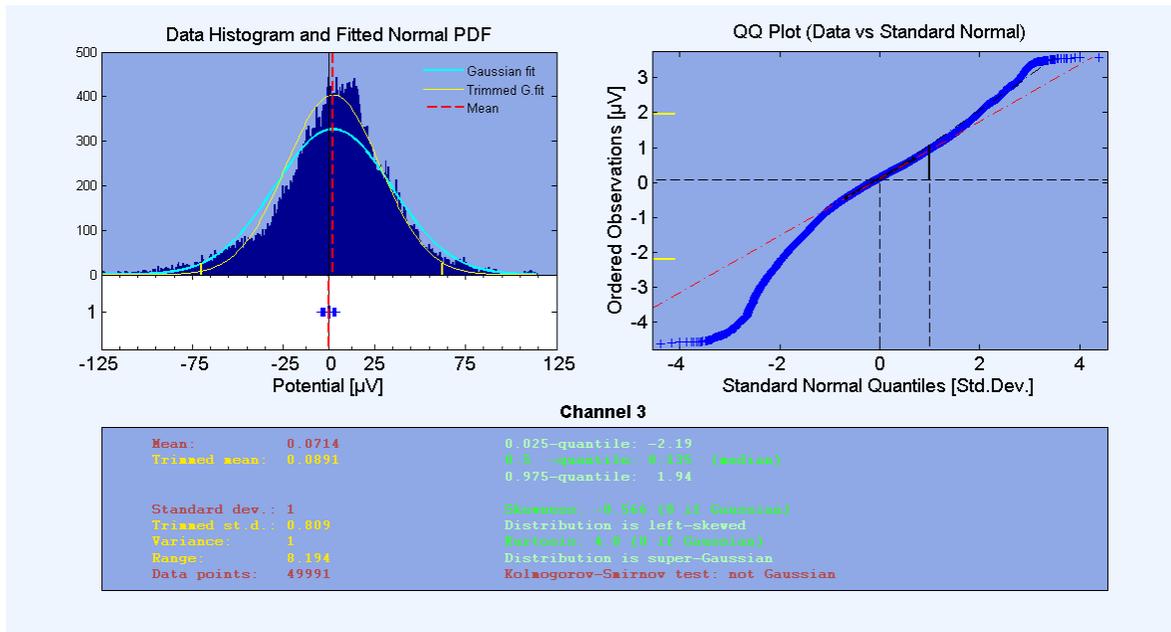


Figura 4.128 Estadísticas de la tercera componente. [Fuente Propia]

Experimento9:

Registro *twa08* de “T-Wave Alternans Challenge Database (*twadb*)”.

Resultados después de aplicar los pasos del 1 al 9:

Paso 1:

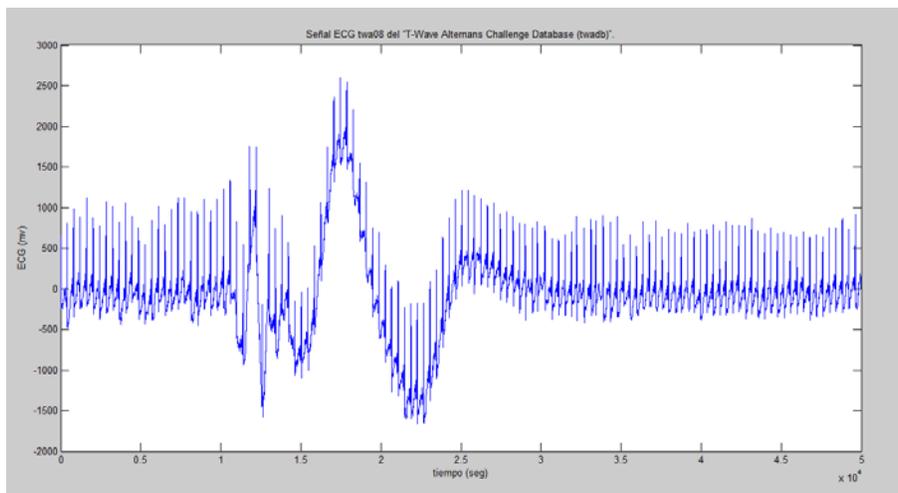


Figura 4.129 Forma de onda de la señal ECG *twa08* de “T-Wave Alternans Challenge Database (*twadb*)”.

[Fuente Propia]

Paso 2:

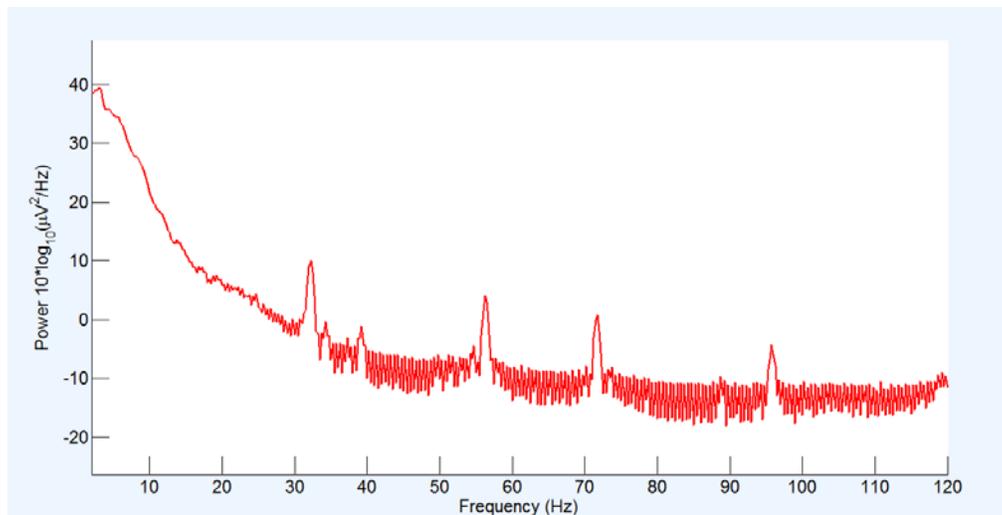


Figura. 4.130 Densidad espectral de potencia de *twa08*. [Fuente Propia]

Paso 3:

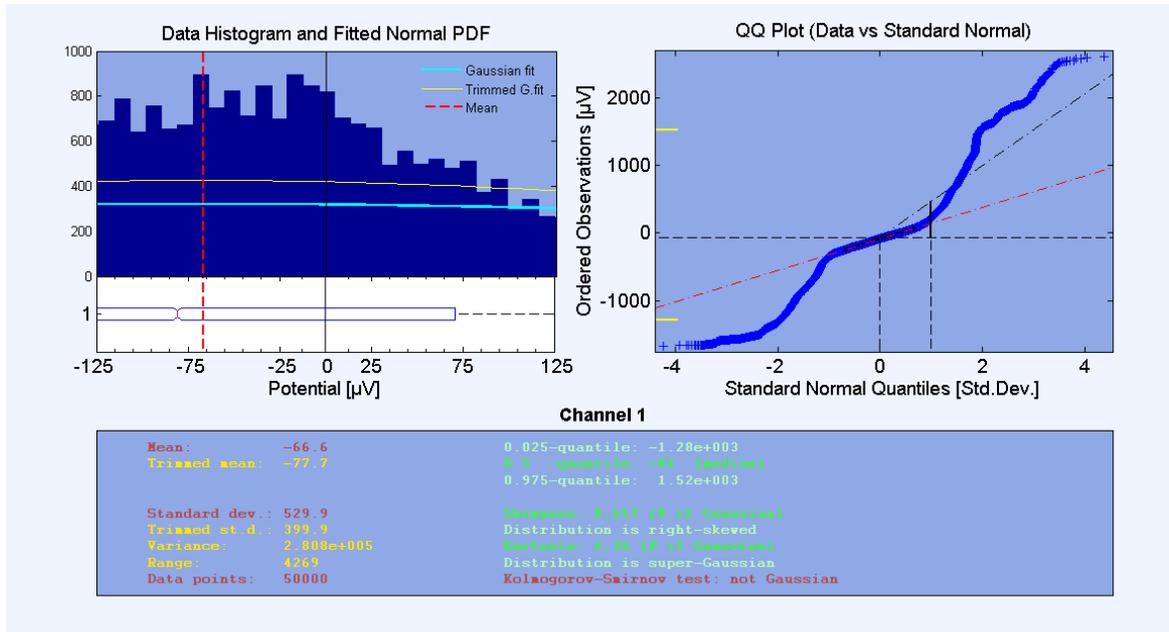


Figura 4.131 Estadísticas de la señal ECG *twa08*, antes de agregarle ruido[Fuente Propia]

Paso 4:

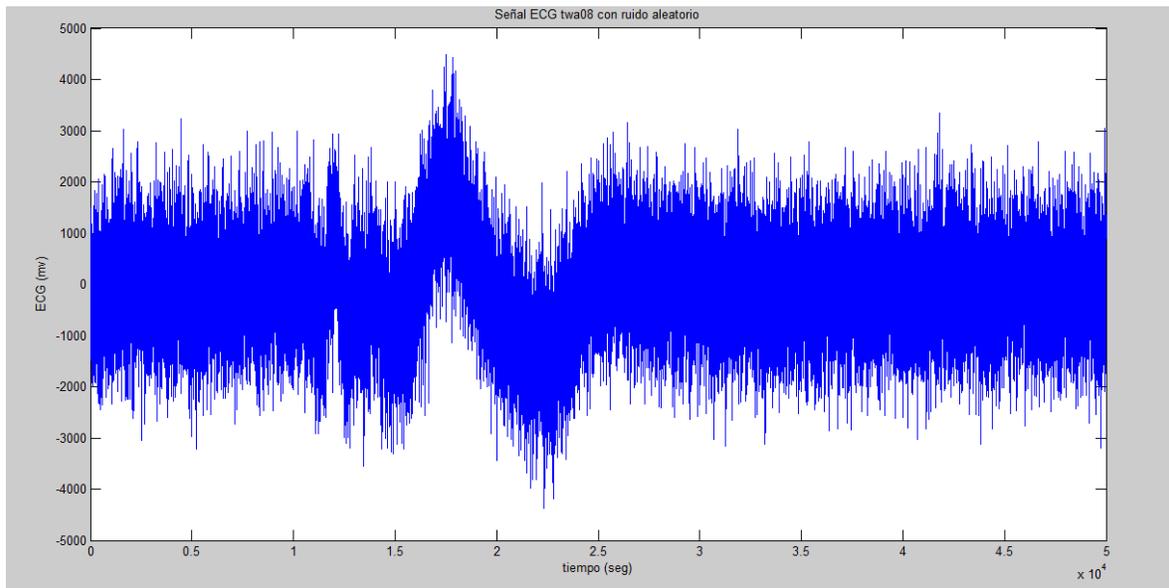


Figura 4.132 Onda que representa la señal ECG *twa08*, contaminada de ruido aleatorio. [Fuente Propia]

Paso 5:

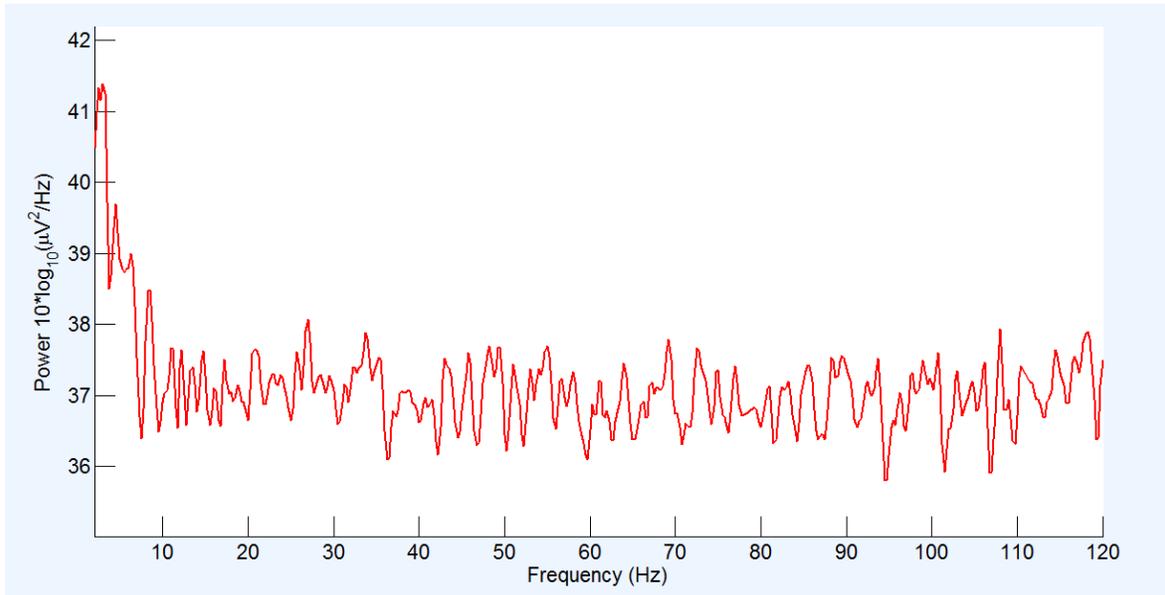


Figura 4.133 Densidad espectral de potencia de *twa08*, contaminada de aleatorio. [Fuente Propia]

Paso 6:

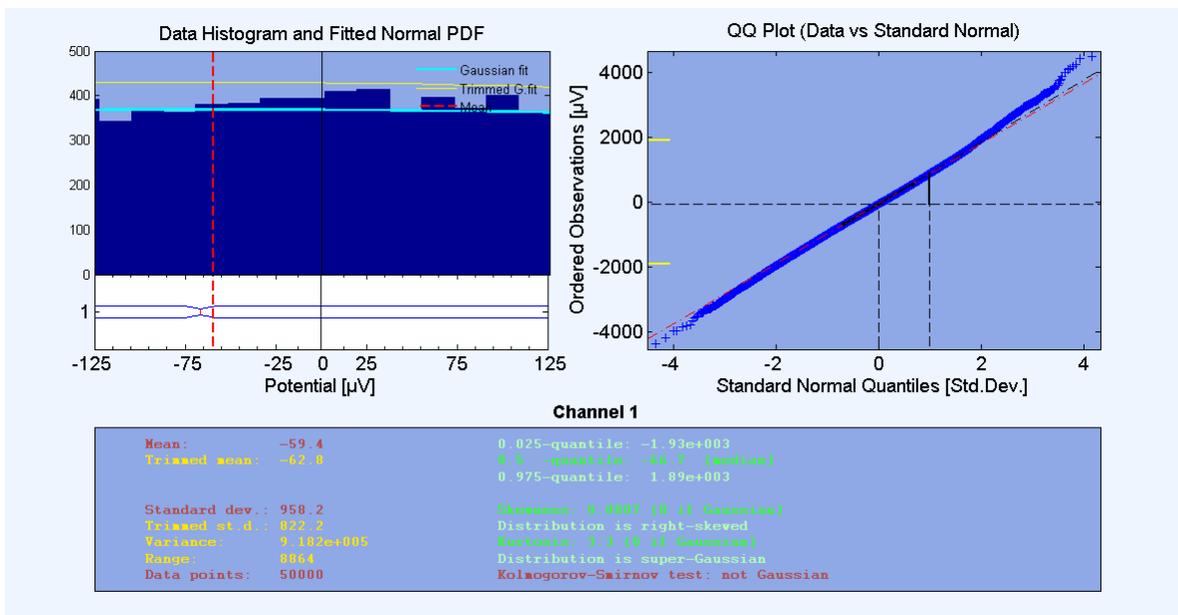


Figura 4.134 Estadísticas de la señal ECG *twa08*, después de agregarle ruido. [Fuente Propia]

Paso 7:

Reducción de la dimensión de la matriz de mezcla:

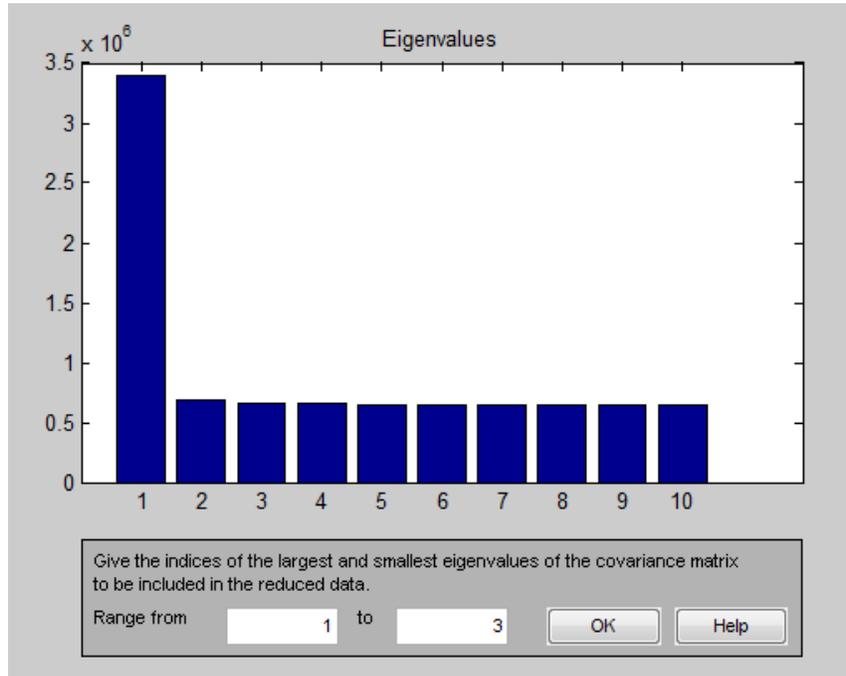


Figura 4.135 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]

Componentes independientes obtenidos después de aplicar el paquete fastICA, (ver figura 4.136).

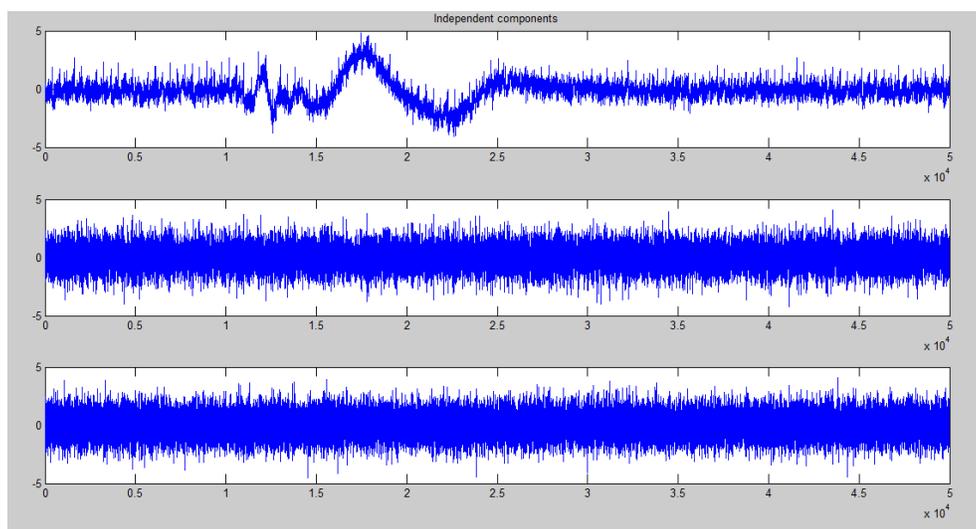


Figura 4.136 Componentes impedientes o fuentes estimadas por fastICA. [Fuente Propia]

Paso 8:

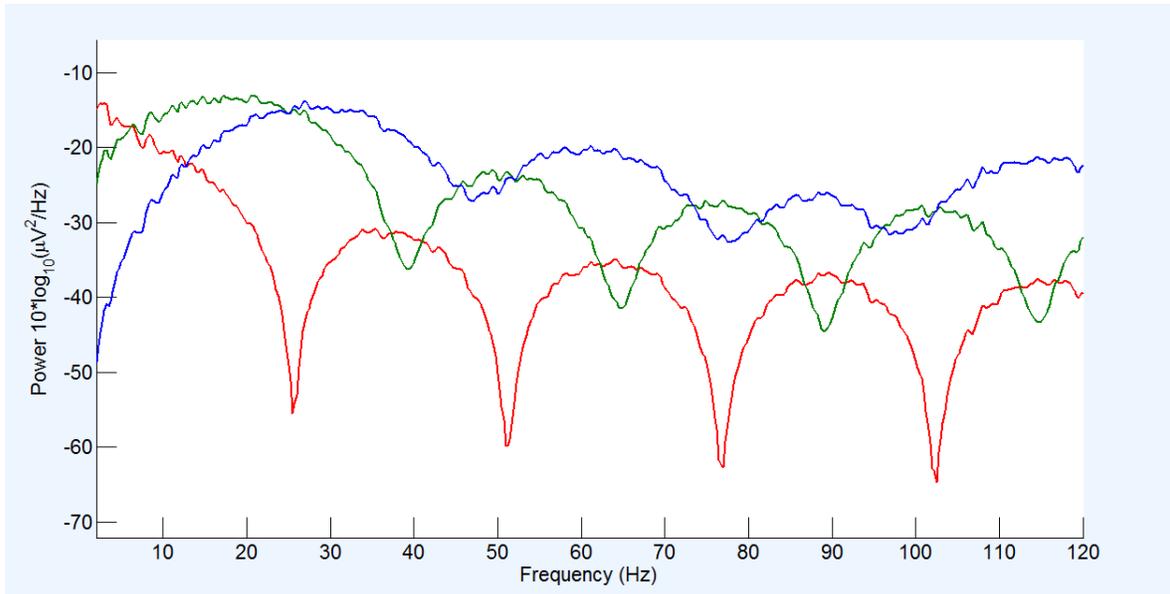


Figura 4.137 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]

Paso 9:

Componente independiente 1

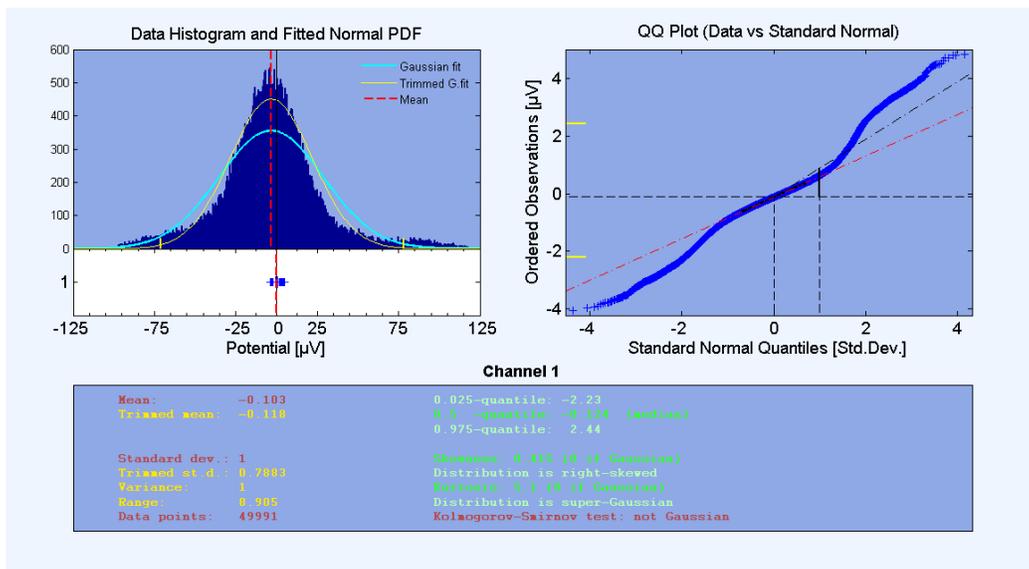


Figura 4.138 Estadísticas de la primera componente. [Fuente Propia]

Componente independiente 2

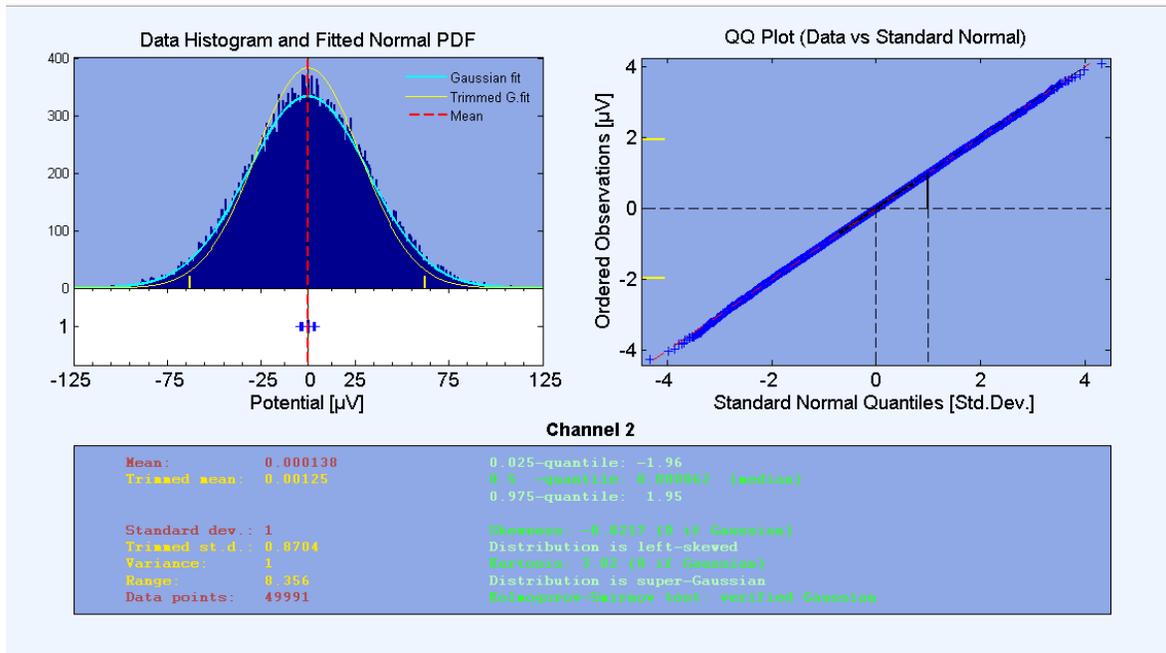


Figura 4.139 Estadísticas de la segunda componente. [Fuente Propia]

Componente independiente 3

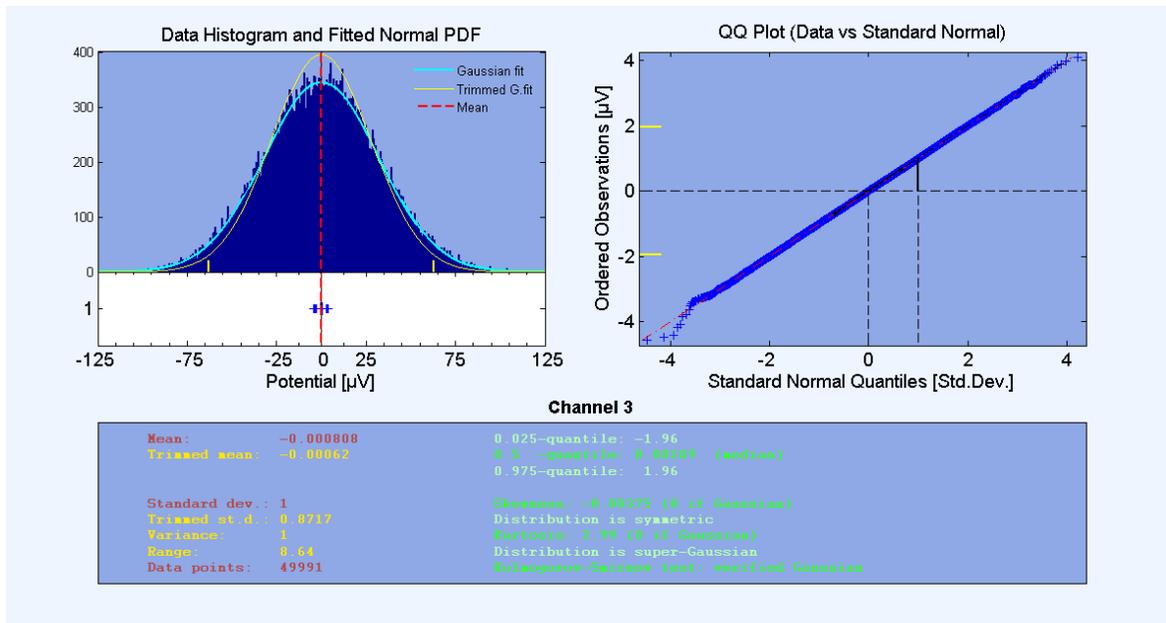


Figura 4.140 Estadísticas de la tercera componente. [Fuente Propia].

Experimento 10:

Registro **twa10** de “T-Wave Alternans Challenge Database (twadb)”.

Resultados después de aplicar los pasos del 1 al 9:

Paso 1:

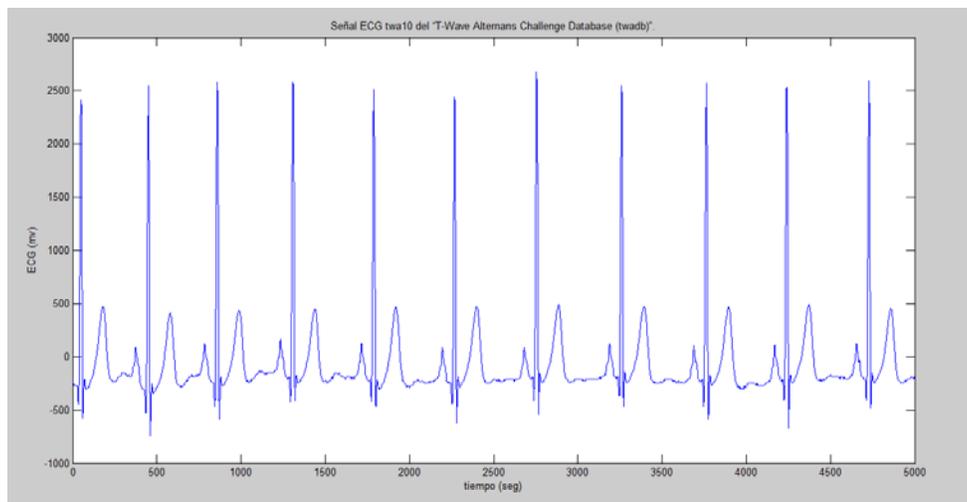


Figura 4.141 Forma de onda de la señal ECG **twa10** de “T-Wave Alternans Challenge Database (twadb)”. [Fuente Propia]

Paso 2:

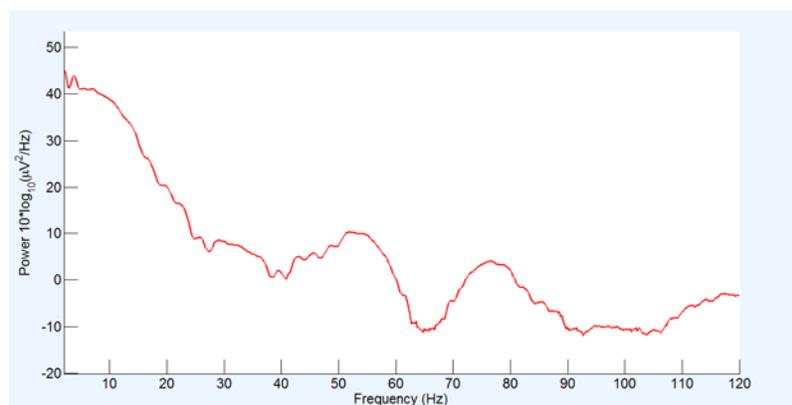


Figura 4.142 Densidad espectral de potencia de *twa10*. [Fuente Propia]

Paso 3:

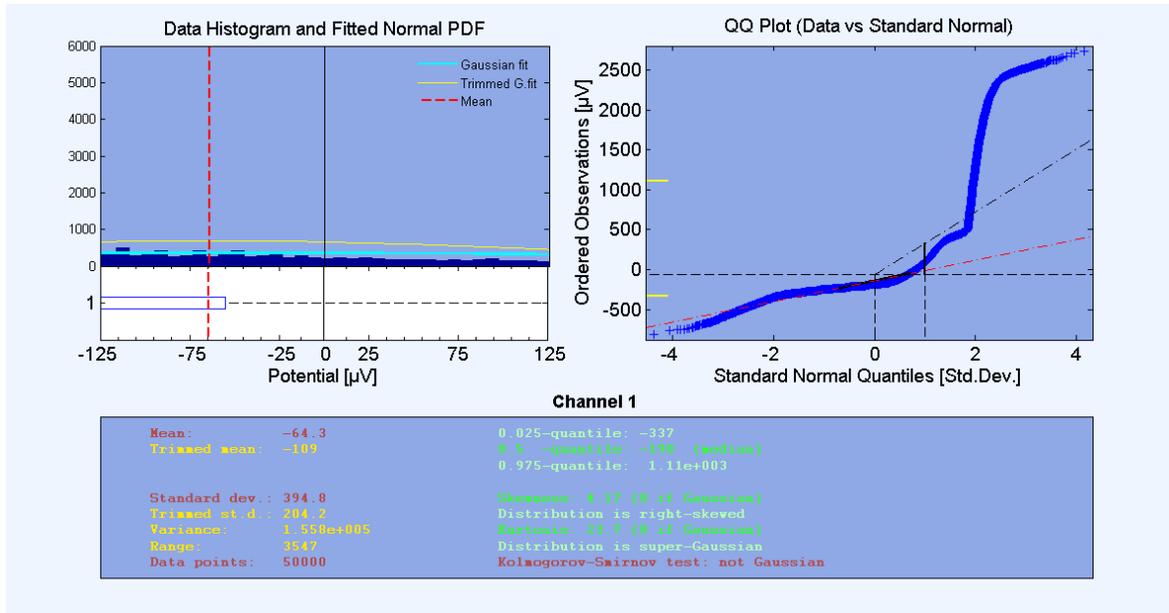


Figura 4.143 Estadísticas de la señal ECG *twa10*, antes de agregarle ruido. [Fuente Propia]

Paso 4:

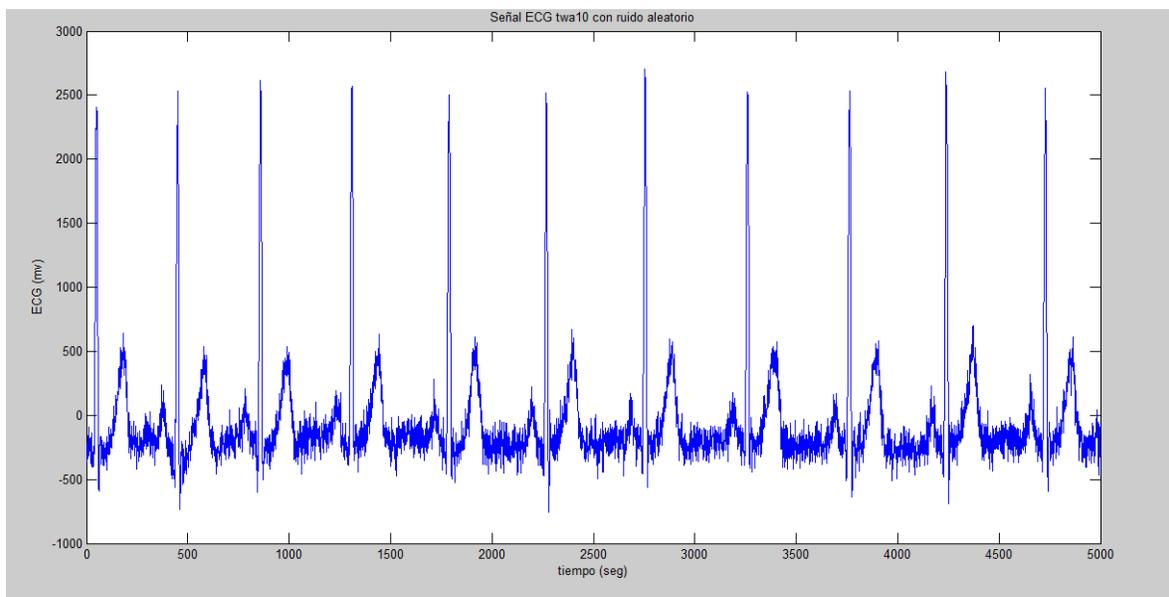


Figura 4.144 Onda que representa la señal ECG *twa10*, contaminada de ruido aleatorio. [Fuente Propia]

Paso 5:

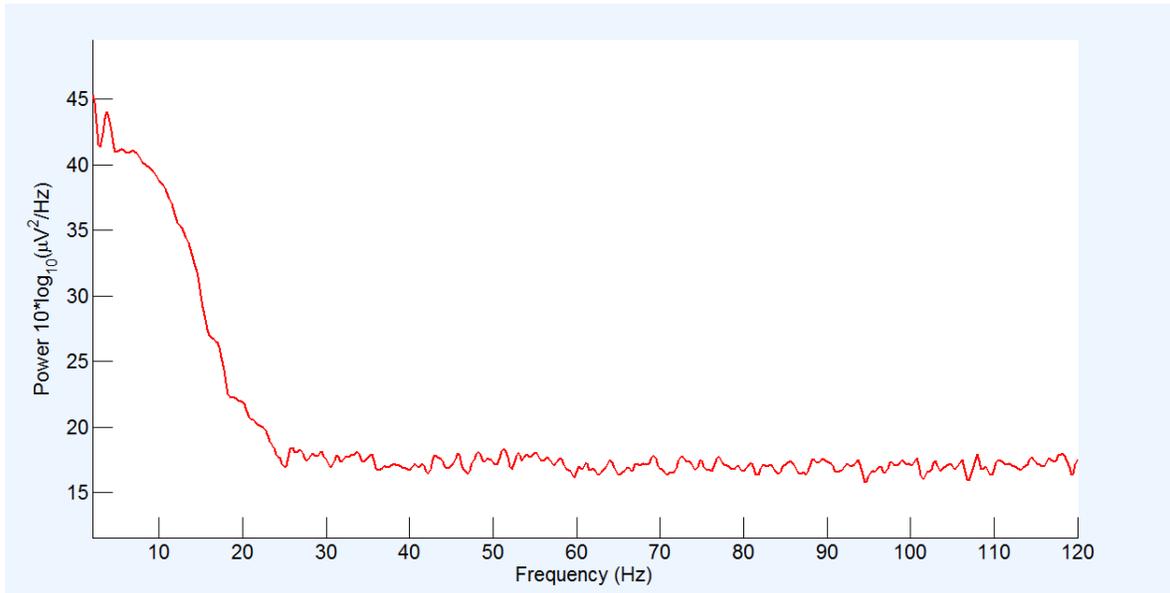


Figura 4.145 Densidad espectral de potencia de *twa10*, contaminada de aleatorio. [Fuente Propia]

Paso 6:

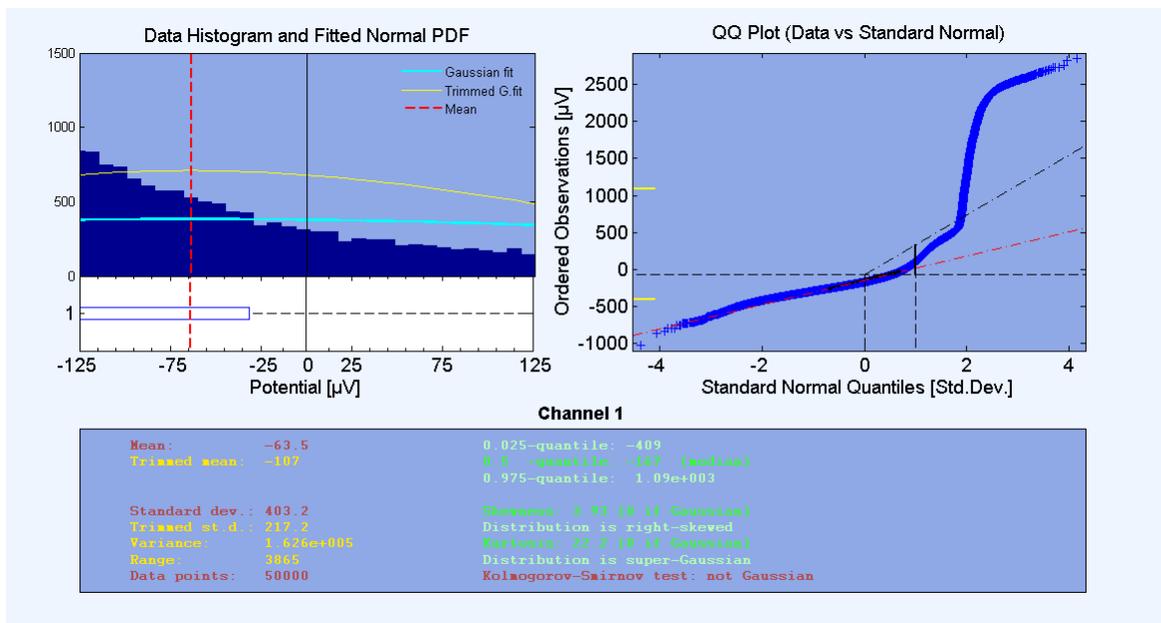


Figura 4.146 Estadísticas de la señal ECG *twa10*, después de agregarle ruido. [Fuente Propia].

Paso 7:

Reducción de la dimensión de la matriz de mezcla:

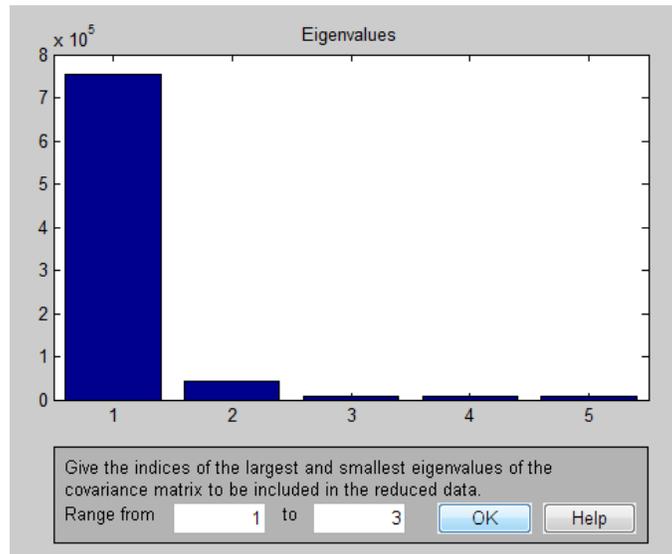


Figura 4.147 Reducción de la matriz antes de obtener los componentes independientes. [Fuente Propia]

Componentes independientes obtenidos después de aplicar el paquete *fastICA*, (ver figura 4.148).

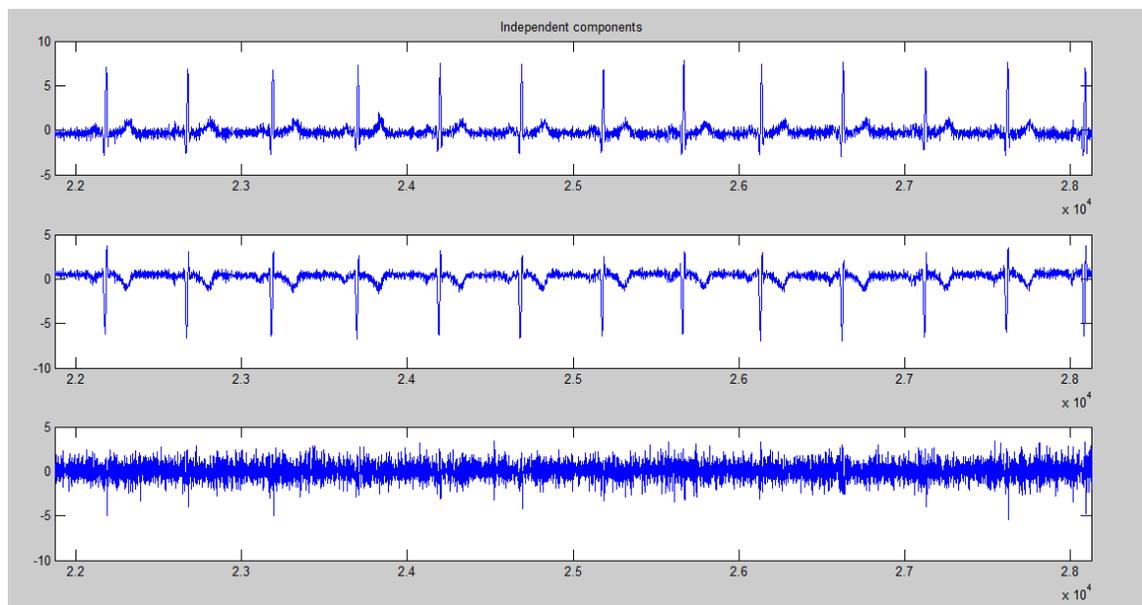


Figura 4.148 Componentes independientes o fuentes estimadas por *fastICA*. [Fuente Propia]

Paso 8:

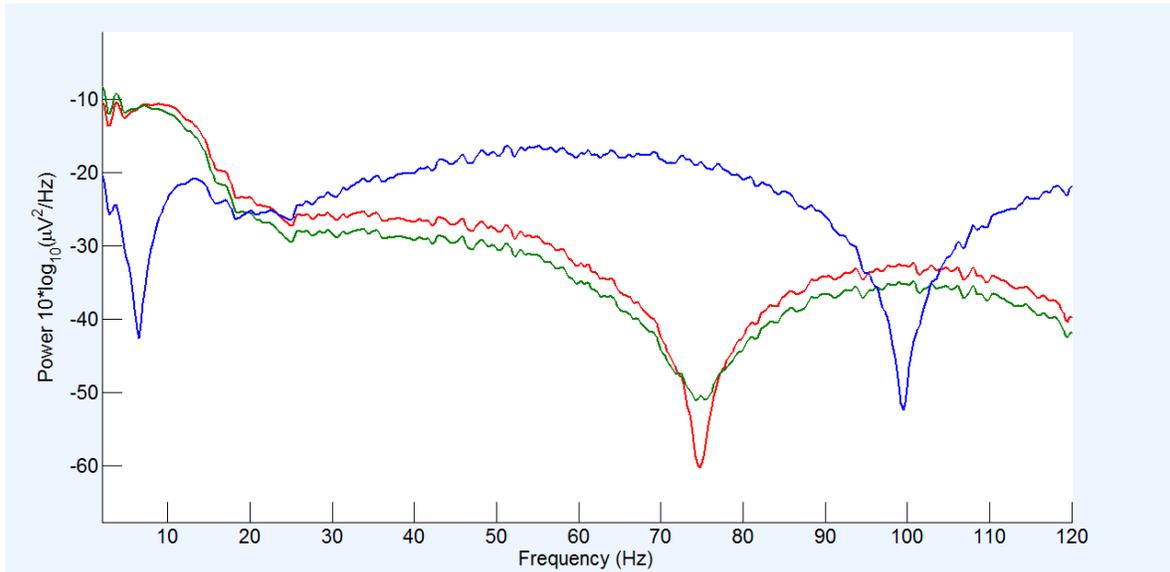


Figura 4.149 Densidad espectral de potencia de las componentes independientes. [Fuente Propia]

Paso 9

Componente independiente 1

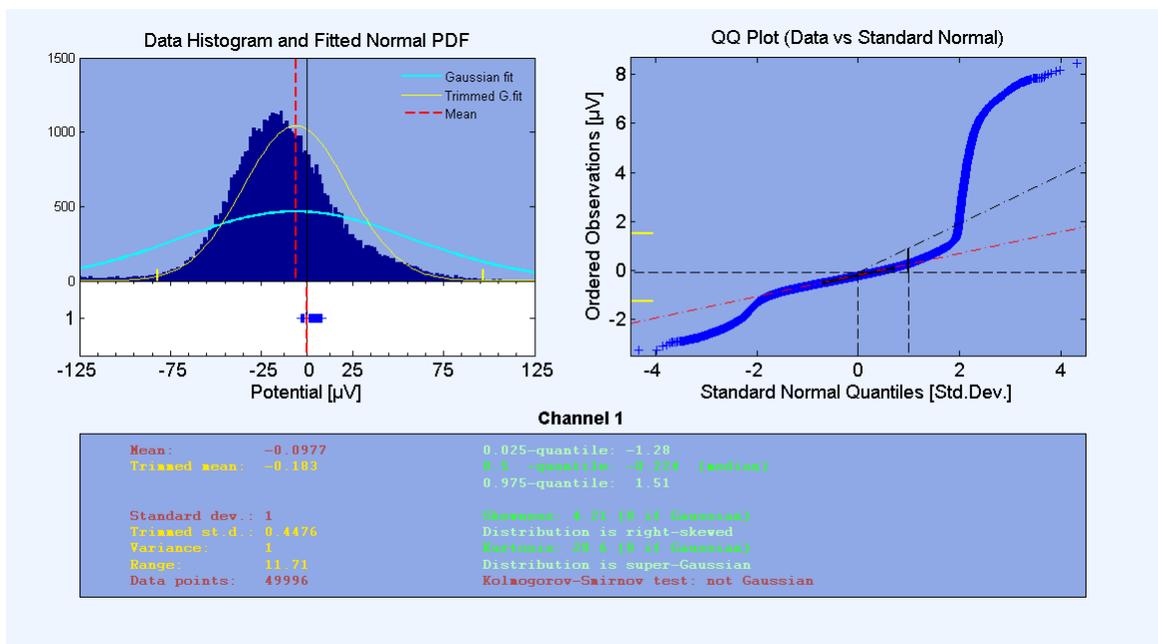


Figura 4.150 Estadísticas de la primera componente. [Fuente Propia]



Componente independiente 2

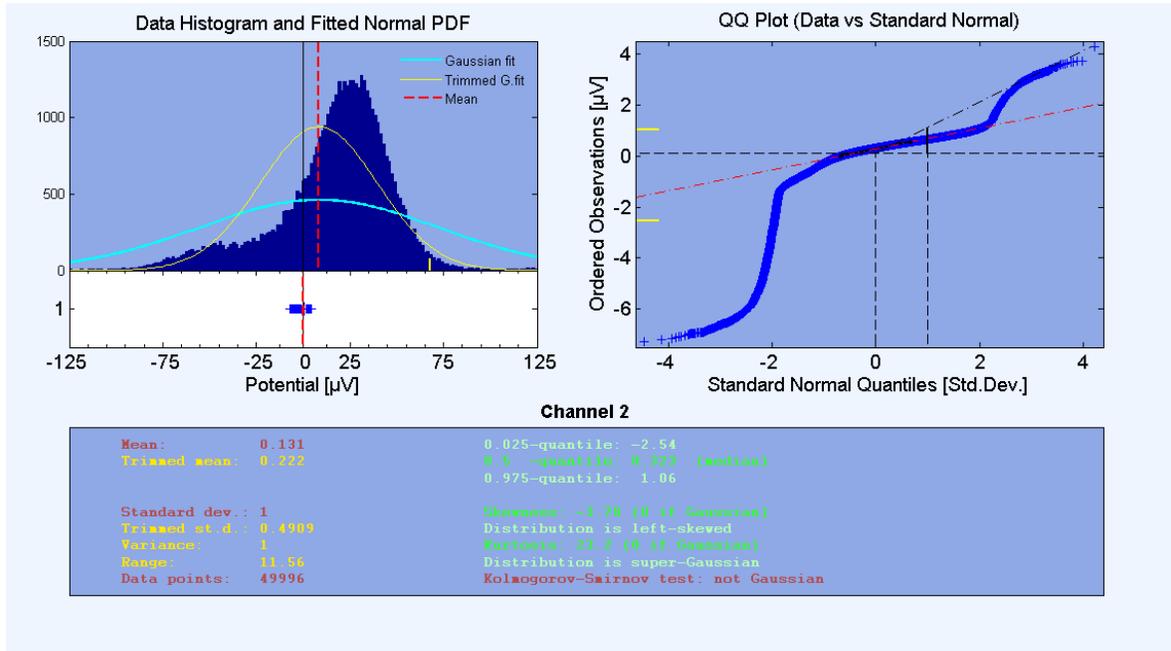


Figura4.151 Estadísticas de la segunda componente. [Fuente Propia]

Componente independiente 3

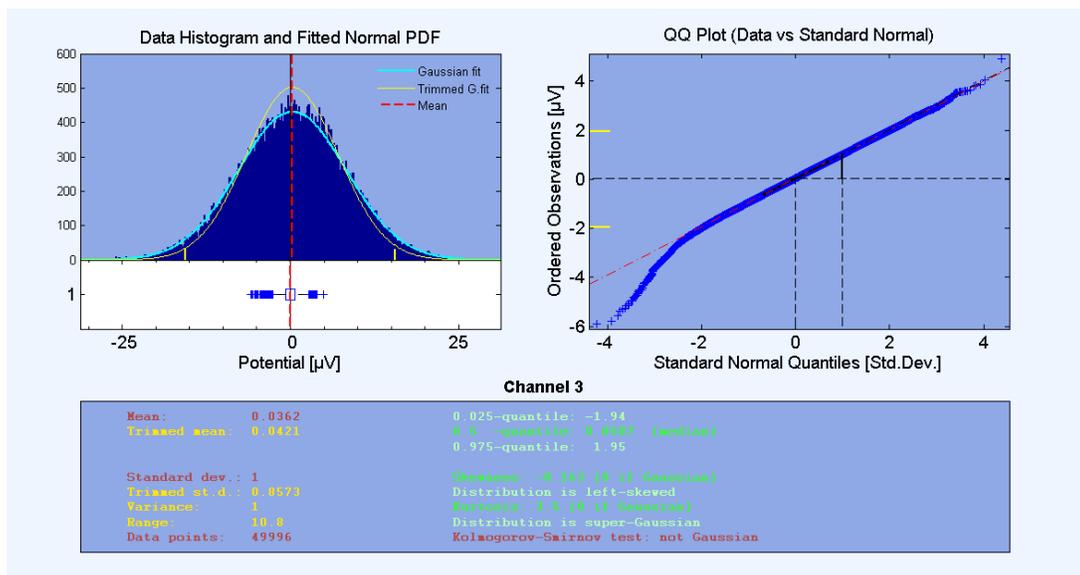


Figura 152 Estadísticas de la tercera componente. [Fuente Propia]



CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES



CONCLUSIONES

La presente investigación logró llevar a cabo la implementación del método de separación de fuentes, llamado ICA “análisis de componentes independientes”, y se pudo comprobar su funcionalidad de acuerdo a las propiedades, restricciones y ambigüedades, que condicionan a dicho método, pudiéndose destacar el caso cuando el ruido añadido a la señal ECG es de tipo aleatorio, cuyo comportamiento es gaussiano donde, en la estimación de las señales fuente, no aparecían bien separadas las señales, la explicación a este resultado, se basa en unas de las restricciones del método ICA, donde especifica que no pueden ser gaussianas las señales originales que intervienen en la mezcla.

Otro aspecto relevante que condicionaba el resultado de los experimentos realizados a varias señales electrocardiográficas, era el hecho de solo contar con un canal para el análisis de componentes independientes, cuando lo necesario es como mínimo dos canales, ya que la matriz de mezcla A debe ser cuadrada e invertible, entre otras condiciones. En este caso se reprodujo la misma señal varias veces, pero desfasadas entre sí, para lograr un efecto de varios canales, para su posterior análisis.

A pesar de estas limitantes, se logró, en el caso de una señal contaminada con ruido de línea, separar de manera satisfactoria las señales originales, obteniéndose nuevamente las mismas, después de aplicarles ICA, y por condiciones del propio método, la media y la varianza de estas señales, resultaron la unidad.

Se destacó la importancia del estudio de los estadísticos de orden superior, que son una de las bases teóricas en la que se sustenta el método de “**Análisis de Componentes Independientes ICA**”, así como las medidas de independencia estadística, que son las que garantizan la efectividad del método de estimación de la fuentes.



RECOMENDACIONES

Dar continuidad y ampliar el campo de aplicaciones donde se puede utilizar este método, como por ejemplo otras señales electrofisiológicas, análisis de vibraciones en elementos mecánicos, señales eléctricas, en fenómenos que respondan a comportamientos estocásticos, fenómenos sociales etc.

Profundizar, en la metodología para obtener resultados más óptimos.

_Elaborar un método más robusto para generar una matriz a partir de un vector que sirva para realizarle el análisis de componentes independientes ICA, en el caso que se cuente con un solo canal se señales mezcladas

Investigar lo referente a las ambigüedades y restricciones de ICA de manera de reducirlas en lo posible y no perder información acerca de la varianza y la media de las fuentes estimadas, entre otros.

Unificar los conceptos básicos en los que se basa el método ICA, con otras investigaciones que persiguen el mismo objetivo, y determinar con certeza, que método responde mejor que otro según sea el caso.



REFERENCIA BIBLIOGRÁFICAS

1. Carlos A. Ramírez R. (cram@unet.edu.ve), Miguel A. Hernández S. (mahs@unet.edu.ve), “Procesamiento en tiempo real de variables Fisiológicas”, Universidad Nacional de Experimental de Táchira, Decanato de Investigación, Grupo de Biomédica. 2001.
2. M. R. Sperling, “EEG reading season”, *J. Clin Neurophysiol.*, 2006;23: pp. 230–237.
3. W. O. Tatum IV, A. M. Husain, S. R. Benbadis and P. W. Kaplan, “Normal adult EEG and patterns of uncertainly significance”, *J Clin Neurophysiol* 2006;23: pp. 194–207.
4. R. Silipo, G. Deco, and H. Bartsch, “Brain tumor classification based on EEG hidden dynamics”, *Intell. Data Anal*, 3, pp. 287–306, 1999.
5. R R. Rao and R. Derakhshani, “A comparison of EEG preprocessing methods using time delay neural networks”, *Proceedings of the 2 International IEEE EMBS Conference on Neural Engineering*, Arlington, Virginia, March 16 - 19, 2005.
6. R. Ksiezzyk, K. Blinowska, P. Durka, “Neural Networks with Wavelet Preprocessing in EEG Artifact Recognition”, Laboratory of Medical Physics, Institute of Experimental Physics, Warsaw University, Hoza 69, 00 681 Warszawa, Poland. Disponible en: http://brain.fuw.edu.pl/~durka/papers/medicon98_arif.pdf.
7. P. Durka, R. Ksiezzyk and K. Blinowska, “Neural networks and wavelets analysis in EEG artifact recognition”, *In Proc. of Second Conference on Neural Networks and Their Applications*, Szczyrk, pp. 167-172, 1996.



8. S. Makeig, A. J. Bell, T. P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data", *Advances in neural information processing systems*, Vol. 8, pp. 145–151, 1996, Cambridge, MA: The MIT Press, 1995.
9. Bendat, Julius; Piersol, Allan. "Random Data Analysis and Measurement Procedures", Third Edition, John Wiley & Sons, USA, 2000.
10. McConnell, Kenneth. G. "Vibration Testing: Theory and Practice", John Wiley & Sons, USA, 1995.
11. Grewal, Mohinder; Andrews, Angus. "Kalman Filtering : Theory and Practice", Prentice Hall, USA, 1993.
12. Peyton, Peebels. "Probability, random variables, and random signal principles", McGraw Hill, Japan, 1980.
13. Ross, Sheldon. "Introduction to Probability and Statistics for engineers and scientists", John Wiley & Sons, USA, 1987.
14. Oliver, Marco. "Parametric Higher-Order Spectral Representation Based on Non-Linear Models", PhD Thesis, University of Sheffield, 1994.
15. Swami, A; Mendel, J; Nikias, C. "Higher-Order Spectral Analysis Toolbox :For Use with MatLab", User's Guide Version 2, MathWorks, 1998.



16. Gatt, George; Kalouptsidis, Nicholas. "Identification of discrete-time state affine state space models using cumulants", Elsevier Science, Automatica No.38, 2002.
17. Mendel, Jerry M. "Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and System Theory: Theoretical Results and Some Applications", Proceedings of the IEEE, Vol. 79, No. 3, 1991.
18. Priestley, M. B. "Spectral Analysis and Time Series", Vol. 2, Multivariate Series, Prediction and Control., Academic Press. Eds., UK, 1981
19. Julio Rodríguez Navarro, Jaime Francisco Aviles Viñas, "Procesamiento digital de señales de vibración con fines de diagnóstico" Centro Nacional De Investigación y Desarrollo Tecnológico. 2003
20. Krishna Garg, "*Digital Signal Processing Algorithms*", CRC Press, 1998, USA.
21. Vergara JM, Brualla J, "Calidad de Sueño y SUBJETIVA Análisis Espectral de Electroencefalograma de Sueño Nocturno ", Revista de Neurología, vol. 28 (N ° 8), 1999, España.
22. Proakis John, Manolakis Dimitris, "*Digital Signal Processing (Principles, algorithms and applications)*", Third ed., Prentice Hall, 1996, USA.
23. Louis Schirm IV. "FFT for Non-FFT People". Redondo Beach, CA. 1979.
24. Bellanger Maurice, "*Digital Processing of Signals (Theory and practice)*", Third Ed., Wiley, 2000, France.
25. Sierra Gilberto, Gomez Maria de Jesus, "*Spectral Analysis of Electrograms During Ventricular Tachycardia in a Canine Model*", Journal of Electrocardiology, Vol. 30 (No. 3), 1997, Canada.
26. Sierra Gilberto, Gomez Maria de Jesus, "*Discrimination Between Monomorphic and Polymorphic Ventricular Tachycardia Using Cycle Length Variability Measured by Wavelet Transform Analysis*", Journal of Electrocardiology, Vol. 31 (No. 3), pags. 245 – 254, 1998, Canada.



27. Rodríguez Julio, Aviles Jaime, “*Procesamiento digital de señales de vibración con fines de diagnostico*”, Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Morelos Noviembre 2003.
28. Gari D. Clifford, “*Blind Source Separation: PCA & ICA*”, Harvart-MIT Health Sciences & Technology, 2009.
29. Hyvärinen, J. Särelä, and R. Vigário. Spikes and bumps: “*Artefacts generated by independent component analysis with insufficient sample size*”. In Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA). Aussois, France, 1999.
30. Aapo Hyvärinen, Juha Karhunen, Erkki Oja, “*Independent Component Analysis*”, Ed. Wiley Interscience, 2001.
31. Mulkern RV, Wong STS, Winalski C, Jolesz. “*Contrast manipulation and artifact assesment of 2D and 3D RARE sequences*”. Magn Reson Imaging FA. 1990.
32. Renie Dubs de Moya, “*El Proyecto Factible: Una modalidad de Investigacion*”, Universidad Pedagógica Experimental Libertador, Caracas Venezuela, 2002.
33. Rosa R. Vázquez A., “*Áreas anatómicas cerebrales activadas en respuesta a estímulos con contenido moral*”, Universidad Autónoma Mepoliitana, Unidad Iztapalapa. División de Ciencias Básicas e Ingeniería, México, D.F., 23 de Abril del 2007.
34. G.B. Moody, W.E. Muldrow, and R.G. Mark. “*A noise stress testfor arrhythmia detectors*”. Computers in Cardiology, 1984.