
Metodología para la caracterización de bioseñales
empleando análisis de complejidad sobre espacio de
embebimiento no uniforme

Jorge Andrés Gómez García

Tesis de Maestría



Universidad Nacional de Colombia
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería Eléctrica,
Electrónica y Computación
Manizales, 2010

Metodología para la caracterización de
bioseñales empleando análisis de
complejidad sobre espacio de embebimiento
no uniforme

Jorge Andrés Gómez García

Tesis para optar por el grado de
Magíster en Ingeniería - Automatización Industrial

Director
Prof. Germán Castellanos Domínguez

Universidad Nacional de Colombia - Manizales
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería eléctrica, electrónica y computación
Manizales
2010

Biosignal Characterization Methodology by means of complexity analysis on non-uniform embedding space

Jorge Andrés Gómez García

Thesis for the degree of
Master in Engineering - Industrial Automation

Supervisor
Prof. Germán Castellanos Domínguez

Universidad Nacional de Colombia - Manizales
Faculty of Engineering and Architecture
Department of Electrical, Electronic and Computing
Engineering
Manizales
2010

*”El único verdadero viaje de descubrimiento
consiste no en buscar nuevos paisajes,
sino en mirar con nuevos ojos”*
Marcel Proust

Agradecimientos

Agradezco a Dios por el regalo de la vida.

Agradezco a mis padres, Adiela y Rubiel, a mi hermana Viviana, y a toda mi familia por su paciencia y apoyo incondicional.

Agradezco al profesor Germán Castellanos, por su colaboración y dirección a lo largo del desarrollo de esta tesis. También a todas las personas del *Grupo de Control y Procesamiento Digital de Señales*, especialmente a Sebastián García López, Hermes Fandiño, Diego Hernán Peluffo, y Eliana García por el regalo de su amistad. Agradezco también a Carolina Ospina, Paola Castro y Jorge Alberto Jaramillo por todo el apoyo prestado.

Extiendo un agradecimiento especial a Julián David Arias, jurado de esta tesis, por su constante orientación y acompañamiento.

Agradezco también a todas las personas que de una forma u otra hicieron posible la realización de este trabajo, a todos mil Gracias.

Información del trabajo de tesis

Título:

Metodología para la caracterización de bioseñales empleando análisis de complejidad sobre espacio de embebimiento no uniforme.

Desarrollado por:

Jorge Andrés Gómez García.

Director:

Profesor César Germán Castellanos Domínguez.
Universidad Nacional de Colombia, sede Manizales.

Jurados:

Julián David Arias Londoño. Universidad Nacional de Colombia, sede Manizales.
Juan Ignacio Godino Llorente. Universidad Politécnica de Madrid, España.

Este trabajo fue apoyado por el programa Jóvenes Investigadores e Innovadores p-2008-0781 del Instituto Colombiano para el Desarrollo de la Ciencia y la Tecnología Colciencias.

Índice general

Agradecimientos	ix
Información del trabajo de tesis	xi
Índice general	xv
Lista de figuras	xvii
Lista de tablas	xx
Lista de algoritmos	xxi
Resumen	xxiii
Abstract	xxv
Notación	xxvii
Abreviaturas	xxix
I Preliminares	1
1 Introducción	3
1.1 Planteamiento del problema	3
1.2 Organización del documento	5
1.3 Objetivos	6
1.3.1 Objetivo general	6
1.3.2 Objetivos específicos	6
1.4 Estado del Arte	7
II Marco Teórico	11
2 Representación de sistemas dinámicos en el espacio de estados	13
2.1 Embebimiento Uniforme	13
2.1.1 Obtención de los parámetros óptimos para el embebimiento uniforme	15

2.2	Embebimiento no uniforme	17
2.2.1	Embebimiento no uniforme, método de Judd	18
2.2.2	Embebimiento no uniforme, método de Ragulskis	22
3	Caracterización	27
3.1	Introducción	27
3.2	Dimensión de Correlación (d_2)	27
3.3	Exponente de Hurst	29
3.4	Máximo Exponente de Lyapunov (LLE)	30
3.5	Entropías	31
3.5.1	Entropía de Kolmogorov	33
3.5.2	Entropía Aproximada	35
3.5.3	Entropía Muestreada	37
3.5.4	Entropía Aproximada de Kernel Gaussiano	38
3.5.5	Entropía Fuzzy	38
III	Marco Experimental	43
4	Resultados	45
4.1	Metodología	45
4.2	Prueba sobre Voz	47
4.2.1	Base de Datos	47
4.2.2	Resultados método de Judd y Mees	47
4.2.3	Resultados método de Ragulskis	51
4.3	Prueba sobre base de datos de FCG	54
4.3.1	Base de Datos	54
4.3.2	Resultados método de Judd y Mees	55
4.3.3	Resultados método de Ragulskis	58
4.4	Prueba sobre HRV	61
4.4.1	Base de Datos	61
4.4.2	Resultados método de Judd y Mees	63
4.4.3	Resultados método de Ragulskis	66
IV	Discusiones y Conclusiones	69
5	Discusiones	71
6	Conclusiones y trabajo futuro	75
6.1	Conclusiones	75
6.2	Trabajo Futuro	77
V	Apéndices	79
A	Clasificación	81

A.1	Modelos de Mezcla de Gaussianas (Gaussian Mixture Models – GMM)	81
A.2	Algoritmo de Maximización de la Esperanza (Maximization Expectation – EM)	82
B	Curvas ROC y DET	83
C	Publicaciones	87
	Bibliografía	89

Lista de figuras

2.1	Gráfica ilustrativa del embebimiento uniforme y no uniforme.	18
2.2	Reconstrucción de un atractor para diferentes embebimientos	26
3.1	Ejemplo ilustrativo de los diferentes estimadores de regularidad	41
4.1	Metodología de un sistema de detección de patologías basado en dinámica no lineal y análisis de complejidad	45
4.2	Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Judd en señales de voz	48
4.3	Curvas ROC y DET para señales de Voz. Método de Judd	50
4.4	Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Ragulskis en señales de voz	52
4.5	Curvas ROC y DET para señales de Voz. Método de Ragulskis	53
4.6	Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Judd en señales FCG	56
4.7	Curvas ROC y DET para señales FCG. Método de Judd	57
4.8	Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Ragulskis en señales FCG	60
4.9	Curvas ROC y DET para señales FCG. Método de Ragulskis	60
4.10	Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Judd en señales HRV	64
4.11	Curvas ROC y DET para señales HRV. Método de Judd	65
4.12	Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Ragulskis en señales HRV	67
4.13	Curvas ROC y DET para señales HRV. Método de Ragulskis	68
B.1	Distribución de las diferentes cantidades derivadas de la matriz de confusión a partir de las densidades por clase	83
B.2	Curva ROC típica	84
B.3	Curva DET típica	85

Lista de tablas

4.1	Máximo, mínimo, media y desviación estándar sobre cada uno de las características usando la técnica de embebimiento no uniforme de Judd en señales de voz	48
4.2	Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Judd en señales de voz	49
4.3	Sensibilidad y Especificidad para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Judd en señales de voz . . .	50
4.4	Mínimo, máximo, media y desviación estándar sobre cada una de las características, usando la técnica de embebimiento no uniforme de Ragulskis en señales de voz	51
4.5	Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Ragulskis en señales de voz	52
4.6	Sensibilidad, Especificidad y Acierto para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Ragulskis en señales de voz	53
4.7	Tiempo de cómputo por ventana usando embebimiento uniforme y no uniforme para señales normales y patológicas.	54
4.8	Mínimo, máximo, media y desviación estándar sobre cada uno de las características, usando la técnica de embebimiento no uniforme de Judd y Mees en señales FCG	55
4.9	Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Judd y Mees en señales FCG	57
4.10	Sensibilidad y Especificidad para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Judd en señales FCG . . .	58
4.11	Mínimo, máximo, media y desviación estándar sobre cada uno de las características, usando la técnica de embebimiento no uniforme de Ragulskis en señales FCG	59
4.12	Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Ragulskis en señales FCG	59
4.13	Sensibilidad, Especificidad y Acierto para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Ragulskis en señales FCG	61

4.14	Mínimo, máximo, media y desviación estándar sobre cada uno de las características, usando la técnica de embebimiento no uniforme de Judd en señales HRV	63
4.15	Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Judd en señales HRV	64
4.16	Sensibilidad y Especificidad para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Judd en señales HRV	65
4.17	Mínimo, máximo, media y desviación estándar sobre cada uno de las características, usando la técnica de embebimiento no uniforme de Ragulskis en señales HRV	66
4.18	Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Ragulskis en señales HRV	67
4.19	Sensibilidad, Especificidad y Acierto para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Ragulskis en señales HRV	68

Lista de algoritmos

1	Algoritmo de Judd y Mees para hallar el embebimiento uniforme	22
2	Algoritmo R/S para la estimación del exponente de Hurst	30
3	Algoritmo de Estimación de ApEn	37
4	Algoritmo de Estimación de SampEn	39
5	Algoritmo EM	82

Resumen

En el presente estudio se expone una metodología basada en dinámica no lineal y medidas de complejidad para la detección automática de patologías. El análisis con dinámica no lineal requiere una etapa de reconstrucción en el espacio de estados de la señal estudiada, proceso que es logrado mediante 2 técnicas no convencionales de embebimiento no uniforme. El primer enfoque basa la reconstrucción en un procedimiento que fusiona modelado y embebimiento bajo un criterio de información llamado Mínima Longitud de Descripción. El segundo enfoque basa la reconstrucción en consideraciones geométricas, buscando el máximo esparcimiento del atractor reconstruido en todas las proyecciones planares. Tras la reconstrucción con estas dos técnicas se procede a la caracterización utilizando medidas típicas de complejidad, y estadísticas de regularidad de la serie de tiempo. Para comprobar la influencia de la reconstrucción mediante las técnicas de embebimiento, y la potencia de las características escogidas se usan tres bases de datos de bioseñales diferentes: señales fonocardiográficas, de variabilidad del ritmo cardíaco y de voz. Los resultados muestran que la metodología puede ser utilizada en labores de detección de patologías.

Abstract

The present work presents a nonlinear dynamics and complexity based methodology for automatic detection of pathologies. The nonlinear dynamics analysis requires a reconstruction stage in order to reconstruct the signal on the state space. This is achieved by 2 non conventional embedding techniques. A first approach uses a procedure that fuses modelling and embedding the time series, and uses an Information criterion called Minimum Description Length. A second approach bases reconstruction on geometrical consideration looking for the largest spread of the reconstructed attractor on all the planar projections. After the reconstruction with these two techniques, characterization is made by using complexity measures, and regularity statistics of the time series. To test the influence of the embedding techniques, and the power of the chosen features, three biosignal were used: Phonocardiographics, Heart Rate Variability, and Voice. Results show that the proposed methodology could perform well on pathologic detection tasks.

Notación

$\mathbf{s}(t)$	Serie de tiempo
$C(\cdot)$	Integral de Correlación
m	Dimensión de embebimiento
d	Dimensión real del sistema. Número de variables dinámicas del sistema
τ	Tiempo de retardo
d_w	Ventana de embebimiento
\mathbf{l}	Vector de retardos
$\mathbf{u}(\cdot)$	Vector de estado d-dimensional
$\mathbf{x}(\cdot)$	Vector de estado reconstruidos m-dimensional
$\mathbf{M}(\cdot)$	Matriz de Embebimiento
d_2	Dimensión de Correlación (Correlation Dimension)
H	Entropía de Shannon
H_{KS}	Entropía de Kolmogorov

Abreviaturas

MDL	Minimum Description Length
ApEn	Entropía Aproximada (Approximate Entropy)
SampEn	Entropía Muestreada (Sample Entropy)
GapEn	Entropía Aproximada de Kernel Gaussiano (Gaussian Kernel Approximate Entropy)
FuzzyEn	Entropía Fuzzy (Fuzzy Entropy)
FNN	Falsos vecinos (False Nearest Neighbors)
LLE	Máximo Exponente de Lyapunov (Lyapunov Largest Exponent)
ECG	Electrocardiográficas
FCG	Fonocardiográficas
PSG	Polisomnografía
HRV	Variabilidad del ritmo cardíaco (Heart rate Variability)
OSA	Apnea obstructiva del sueño (Obstructive sleep apnea)
GMM	Modelos de Mezclas de Gaussianas (Gaussian Mixture Models)
ROC	Receiver-Operation Curve
DET	Detection-Error Tradeoff

Parte I

Preliminares

1. Introducción

1.1 Planteamiento del problema

El análisis de sistemas dinámicos con técnicas lineales ha sido el paradigma de trabajo usado durante mucho tiempo. Sin embargo, en muchos tipos de señales se presenta comportamiento irregular que no ha sido posible de explicar por medio de estas técnicas y que para fines prácticos es considerado el producto de fuentes externas aleatorias, ajenas a la señal estudiada.

Por otra parte, el análisis dinámico no lineal puede a partir de ecuaciones determinísticas generar respuestas irregulares, mostrándose como una herramienta poderosa de análisis capaz de darle sentido a los comportamientos extraños percibidos en las señales.

El análisis basado en dinámica no lineal requiere del conocimiento de las variables dinámicas del sistema y la evolución en el tiempo de las mismas. Como usualmente solo se tiene acceso a series de tiempo, y debido a que reconstruir a partir de ellas la dinámica de un sistema resulta complejo, se han asegurado bajo ciertas consideraciones generales, reconstrucciones equivalente capaces de mantener las mismas propiedades que la dinámica original del sistema. Dicho proceso recibe el nombre de embebimiento, y requiere de la escogencia de ciertos parámetros necesarios para una correcta reconstrucción, para el mejoramiento de la observabilidad de las trayectorias con que evolucionan los estados del sistema, o incluso entre otras, para la buena convergencia de algoritmos de estimación de características [1].

En sistemas con una sola frecuencia fundamental los embebimientos convencionales funcionarán bien, siendo posible que se presentasen problemas en sistemas con múltiples frecuencias que pudieran llevar a representaciones incorrectas. Con esto en mente, se han desarrollado procedimientos capaces de trabajar con las múltiples periodicidades en las señales. Dichos procedimientos están dentro de la categoría de embebimientos no uniformes, capaces de variar la reconstrucción dependiendo de la dinámica de la señal. Es de resaltar sin embargo, que dichas técnicas han sido solamente estudiadas

para labores predictivas, obviando su estudio en tareas de reconocimiento de patrones.

Ahora bien, posterior a la etapa de embebimiento, es necesaria una caracterización de las bioseñales de tal manera que se puedan encontrar medidas que puedan ayudar a la discriminancia entre clases en labores de detección de patologías. Típicamente en el análisis dinámico no lineal de señales, se usan características de complejidad basadas entre otras, en las dimensiones del objeto reconstruido en el espacio de estados, las tasas de divergencia promedio de las trayectorias trazadas por los estados a medida que pasa el tiempo, la autosimilaridad de la serie de tiempo embebida, o la memoria a largo plazo de las series de tiempo. Estas características requieren sin embargo que la dinámica sea determinística, suposición que no es del todo válida debido a los componentes estocásticos presentados por influencias indirectas como el ruido.

Como resultado, medidas basadas en entropías, y que no necesitan suponer determinismo u estocasticidad para su cálculo, resultan importantes a la hora de llevar a cabo labores de detección de patologías.

El problema queda entonces presentado: como pueden explotarse las técnicas basadas en dinámica no lineal para tratar con señales de múltiples frecuencias, y como pueden determinarse características capaces de ayudar a la discriminación de patologías y que no exijan que la naturaleza de las señales sea puramente determinística.

1.2 Organización del documento

Este documento está compuesto por 4 secciones: Preliminares, Marco teórico, Marco experimental, Discusiones y Conclusiones.

Los contenidos serán los siguientes: En el capítulo 2 se presentan conceptos relacionados con dinámica no lineal y algunas medidas típicas utilizadas en caracterización. Se hace un recorrido por las técnicas clásica de reconstrucción de series de tiempo en el espacio de estados, junto con los parámetros típicos para lograr dichos fines. También se explican las técnicas no convencionales de reconstrucción no uniforme utilizadas a lo largo del trabajo, al igual que las características clásicas de la dinámica no lineal, y algunos estimadores de entropía de la serie de tiempo.

El capítulo 4 presenta la metodología de trabajo utilizada, así como los conjuntos de pruebas realizados y los resultados obtenidos. Este capítulo contiene 3 secciones correspondientes a los 3 tipos de pruebas realizados: sobre señales de voz, Fonocardiográficas y de variabilidad cardíaca.

Los resultados de los experimentos son discutidos en el capítulo 5, presentando a su vez, las conclusiones de la tesis y el trabajo futuro.

1.3 Objetivos

1.3.1 Objetivo general

Desarrollar una metodología para la detección automática de patologías basada en dinámica no lineal, usando técnicas de embebimiento no uniforme y medidas de complejidad.

1.3.2 Objetivos específicos

- Comparar la influencia del embebimiento no uniforme en labores de reconocimiento de patrones, mediante 2 enfoques diferentes de estimación de parámetros para el embebimiento no uniforme.
- Evaluar la capacidad discriminante que las medidas de regularidad basadas en entropía, pueden proveer en labores de detección de patologías.
- Validar los resultados obtenidos mediante curvas ROC y DET.

1.4 Estado del Arte

Mucho de lo que se conoce de sistemas fisiológicos se ha aprendido usando teoría de sistemas lineales, implicando una concepción en la que la dinámica intrínseca del proceso analizado es gobernada por el paradigma de que pequeñas causas llevan a pequeños efectos [1]. Debido a que modelando sistemas con ecuaciones lineales, se pueden generar solamente soluciones periódicamente oscilantes o de crecimiento exponencial; todos los comportamientos irregulares o aperiódicos, son atribuidos tradicionalmente a influencias externas aleatorias (como el ruido), o a interacciones entre un gran número de componentes constituyentes. Sin embargo, estas no resultan ser las únicas fuentes de comportamiento irregular. Los sistemas dinámicos no lineales pueden producir comportamientos muy irregulares con ecuaciones puramente determinísticas, debido a pequeños cambios en parámetros de control y en variaciones en las condiciones iniciales, que pudieran explicar el comportamiento irregular observado en señales fisiológicas. En la práctica, se ha demostrado comportamiento no lineal en bioseñales, proveyendo un desafío a los enfoques tradicionales de estudio [2–6].

Entre muchas otros fenómenos fisiológicos no lineales, se puede considerar por ejemplo el relativo a la producción de voz; cuyo comportamiento no lineal, proviene de fenómenos como la relación presión-flujo en la glotis, las curvas de esfuerzo-estiramiento en los tejidos de las cuerdas vocales y efectos asociados con la colisión de las cuerdas vocales [7, 8].

Otro ejemplo es visto en los intervalos entre latidos de un corazón humano sano, que muestran una fluctuación en una forma compleja e irregular. Presumiblemente, dichas fluctuaciones vienen de la existencia de comportamiento no lineal originado por el sistema nervioso, específicamente relacionado con las entradas neuroautónomas. La estimulación del sistema parasimpático decrementa la actividad en el nodo seno del corazón, mientras que el sistema simpático la incrementa. El efecto de ambos estímulos genera una fluctuación en la tasa cardíaca que no puede ser explicada por el análisis lineal [9, 10].

El cerebro ha mostrado también comportamiento no lineal introducido a nivel celular, debido a que sus componentes constituyentes, las neuronas, están gobernadas por fenómenos no lineales como saturaciones y umbrales. Luego este comportamiento es visto a nivel macroscópico en una red neuronal como el cerebro. [11–13].

Se pueden también nombrar los mecanismos de producción de sonidos respiratorios. Los sonidos pulmonares son producidos probablemente por turbulencia en los bronquios

lobulares o segmentarios, mientras que los sonidos traqueales son producidos por ruido de cavitacion, ambos conocidos por presentar comportamiento no lineal [14].

En el contexto de detección de patologías, el análisis con dinámica no lineal ha generado un creciente interés debido a que se ha mostrado que cambios en medidas basadas en dinámica no lineal pueden indicar estados de disfunción patofisiológica [7].

Ahora bien, para poder hacer uso de análisis con dinámica no lineal sobre series de tiempo, es necesaria una etapa de reconstrucción de la dinámica intrínseca del sistema analizado. Dicho procedimiento, llamado embebimiento, llevará la serie de tiempo a un espacio m dimensional llamado espacio de fases, donde se podrá analizar la evolución de los estados del sistema. El teorema de Takens [15] provee la reconstrucción típica con una técnica llamada método de los retardos, que requiere del cómputo de dos parámetros: la dimensión de embebimiento m y el tiempo de retardo τ . Este último parámetro se escogerá de tal manera que optimice el esparcimiento de la serie de tiempo embebida. Dado que esta reconstrucción utiliza solamente un tiempo de retardo τ llevará también el nombre de embebimiento uniforme.

A pesar de que el embebimiento uniforme ha sido exitoso en la reconstrucción de la dinámica de series de tiempo fisiológicas, se pueden presentar problemas cuando las señales analizadas presentan múltiples periodicidades fuertes con escalas de tiempo muy diferentes. Esto porque un tiempo de retardo pequeño será óptimo para los componentes de alta frecuencia, mientras que un tiempo de retardo grande, lo será para las bajas frecuencias y la modulación, siendo un compromiso inadecuado [16]. Para solucionar dicho problema el embebimiento no uniforme fue propuesto [16], reemplazando el único vector de retardo τ por un vector de retardo \mathbf{l} , capaz de hacer frente a las diferentes frecuencias presentes en la señal.

Varios métodos han sido propuestos para resolver el problema de la estimación adecuada del vector de retardos. La primera proposición encontrada en la literatura fue dada por Judd [16], que basa la escogencia del vector de retardos en un principio venido de teoría de la información llamado Mínima Longitud de Descripción, y que combina el embebimiento y modelado de una serie de tiempo en un solo procedimiento, de tal manera que lo que se escoja para evaluar la calidad del embebimiento se relacione con lo que se escoja para evaluar la calidad del modelo. Otra proposición es la dada por [17], que propone el uso de redes neuronales de autopropagación entrenadas por aprendizaje estructural con regularizador. En [18] se propuso un método que usa consideraciones geométricas para seleccionar el vector de retardo óptimo. Para tales fines se define una función objetivo del esparcimiento de la serie de tiempo embebida en todas las

proyecciones planares, que debe ser maximizada para encontrar los parámetros óptimos del vector de retardo.

El test Gamma, un procedimiento de modelado no lineal, que permite examinar la naturaleza de una hipotética relación entrada-salida en los datos analizados [19], ha sido también utilizado en la consecución de un vector de retardos óptimo. Para ello, se escogen m valores pasados (la dimensión de embebimiento) y se hace una optimización de los 2^{m+1} combinaciones de elementos posibles de vectores de retardos. Cada combinación será entrada a un modelo que buscará la minimización del error de predicción [20].

En general, las técnicas de embebimiento no uniforme han tenido aceptación en el contexto de predicción, siendo su labor en tareas de reconocimiento de patrones un campo inexplorado.

Ahora bien, en la caracterización con medidas de dinámica no lineal, se usan convencionalmente medidas como el Máximo Exponente de Lyapunov (LLE) que cuantifica la sensibilidad a condiciones iniciales, la dimensión de correlación (d_2) que mide la autosimilaridad de la serie de tiempo embebida, y el Exponente de Hurst que calcula la correlación existente entre diferentes puntos en la serie de tiempo [1, 21]. Sin embargo el cálculo de LLE y d_2 requieren que la dinámica de las señales sea puramente determinística, suposición que es inadecuada debido a los componentes estocásticos que pueden presentarse en series de tiempo fisiológicas.

Existen una serie de medidas caracterizantes de la complejidad de una señal, que no hacen suposiciones acerca de su naturaleza (determinística o estocástica)[22]. Algunas de las medidas usadas en este contexto son: la Entropía Aproximada (Approximate Entropy – ApEn) [23], la Entropía Muestreada (Sample Entropy – SampEn) [24], la Entropía Aproximada de Kernel Gaussiano (Gaussian Approximate Entropy – GapEn)[25], y la Entropía Fuzzy (Fuzzy Entropy – FuzzyEn) [26].

ApEn es un estimador de regularidad, basado en la entropía de Kolmogorov, que mide la probabilidad de que sucesiones de patrones que están cerca, permanezcan cerca en las próximas comparaciones incrementales [23]. A mayor probabilidad de permanecer cerca, se obtiene un valor menor de ApEn [27], que clínicamente está asociados con patologías, es decir con señales más regulares, y con menor aleatoriedad aparente [27].

ApEn ha sido exitosa en el estudio y discriminación de enfermedades múltiples como: episodios de hipertensión intracranial en pacientes pediátricos con lesiones cerebrales traumáticas [28] [29], análisis sobre series de tiempo generadas por pacientes

esquizofrénicos [30], estudios de la variabilidad del ritmo cardíaco por enfermedades y debido al envejecimiento [31], entre otras.

Sin embargo dicha medida posee problemas, debido a que en la búsqueda de épocas similares en la serie de tiempo, hace comparaciones consigo misma generando un sesgo en la estimación. Para resolver dicho inconveniente la Entropía Muestreada fue propuesta en [24].

Esta estadística ha sido exitosa en el estudio de diferentes señales fisiológicas como: la variabilidad cardíaca de neonatos [32], la detección de Estenosis Aórtica usando señales fonocardiográficas en perros [33], el análisis de electroencefalogramas en neonatos [34], entre otras.

Debido a que ApEn y SampEn tiene problemas por estar formuladas sobre la función de Heaviside [26], se han propuesto otros estimadores de regularidad: GapEn [22], que reemplaza la función de Heaviside de ApEn por una de Kernel Gaussiano, y FuzzyEn [26] que lo hace por una función de pertenencia Fuzzy.

Parte II

Marco Teórico

2. Representación de sistemas dinámicos en el espacio de estados

En el presente capítulo se hace una breve introducción a los sistemas dinámicos, y al estudio de los mismos con técnicas no lineales. Se hace un breve recorrido por las técnicas convencionales de representación de señales en el espacio de estados, explicando también dos enfoques no convencionales de reconstrucción de series de tiempo.

2.1 Embebimiento Uniforme

Un sistema dinámico es aquel cuyo comportamiento evoluciona en el tiempo. Si este comportamiento cambia continuamente en el tiempo (o a intervalos discretos de tiempo para el caso discreto) será matemáticamente descrito por un conjunto de ecuaciones diferenciales (o en diferencias) autónomas ordinarias de primer orden. Para el caso discreto éstas tendrán la forma [3]:

$$\mathbf{u}(t + 1) = F(\mathbf{u}(t)) \quad (2.1)$$

Donde $\mathbf{u}(t)$ son vectores de estado d -dimensionales, cuyos elementos son las variables dinámicas (internas) del sistema; y $F(\cdot)$ son las reglas dinámicas y determinísticas que gobiernan el comportamiento de estas variables. Para los fines de esta tesis $F(\cdot)$ será considerada no lineal.

La representación de sistemas dinámicos se da en un espacio \mathbb{R}^d , llamado espacio de estados (o de fase), cuyos ejes coordenados son las variables dinámicas del proceso, y sobre los que es posible representar cualquier estado que el sistema pudiera tener. Si adicionalmente los sistemas son disipativos, habrá una atracción a un subconjunto del espacio de fases, que recibirá el nombre de atractor del sistema [1].

Ahora bien, como usualmente no se tiene acceso a los estados de un sistema, se suele utilizar una función de observación $h(\cdot)$ sobre el proceso, que mapee los estados d -dimensionales no observables a valores escalares. La sucesión de estas observaciones a medida que el tiempo transcurre, da lugar a series de tiempo $\mathbf{s}(t)$.

$$\mathbf{s}(t) = h(\mathbf{u}(t)) \quad (2.2)$$

Si n mediciones son hechas, entonces $\mathbf{s}(t) = (x_1, x_2, \dots, x_n)$, donde cada observación escalar será una proyección de las variables internas del sistema.

Si bien es deseable obtener a partir de la serie de tiempo la dinámica intrínseca del sistema, el proceso de proyección puede ser no lineal y puede mezclar diferentes variables internas, por lo que una reconstrucción del espacio de estados original puede resultar ser compleja, aún teniendo alto conocimiento del proceso de observación que da origen a la serie de tiempo [1]. Por esta razón se busca una reconstrucción equivalente, que será suficiente y a veces más productiva para los fines buscados, que la reconstrucción exacta del atractor original. Por equivalente se habla de reconstrucciones homeomórficas (u equivalentes topológicamente) capaces de conservar las propiedades topológicas del sistema, es decir, aquellas que se guardan bajo transformaciones como estiramiento, traslación, o rotación del objeto sin que implique romper o conectar puntos o regiones que no sean adyacentes [2].

El teorema de Takens provee una solución en reconstrucción para una serie de tiempo, tal que las propiedades del atractor original sean topológicamente equivalentes a la reconstrucción formada por vectores de fase m -dimensionales, con $m \geq 2d + 1$, en un proceso llamado embebimiento y con una reconstrucción de la forma:

$$\mathbf{x}(t + 1) = f(\mathbf{x}(t)) \quad (2.3)$$

Donde f y F serán homeomórficas entre sí, y el vector de estado reconstruido $\mathbf{x}(t) \in \mathbb{R}^m$ será:

$$\mathbf{x}(t) = [x(t), x(t - \tau), x(t - 2\tau), \dots, x(t - (m - 1)\tau)] \quad (2.4)$$

Con m la dimensión de embebimiento o el mínimo número de coordenadas para representar la señal sin que se presenten traslapes en el espacio de estados, y τ , el tiempo de retardo que influenciará la forma y la dispersión del atractor reconstruido.

La colección de todos los estados del sistema forma la matriz de embebimiento \mathbf{M} :

$$\mathbf{M} = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(n - (m - 1)\tau)]^T \quad (2.5)$$

Donde cada $\mathbf{x}(\cdot)$ representa un estado.

Usualmente a (2.4) se le llama embebimiento por retardo de tiempo o embebimiento uniforme, esto debido a que la reconstrucción se hace con un parámetro τ constante que crea vectores de estado, con elementos mapeados a intervalos iguales.

2.1.1 Obtención de los parámetros óptimos para el embebimiento uniforme

La correcta reconstrucción de una serie de tiempo depende del cálculo correcto de los parámetros τ y m . Aunque algunos autores afirman que lo más importante a la hora de embeber es la consecución de una cantidad igual al producto de $m \cdot \tau$, en el presente capítulo se tratarán por aparte los dos parámetros y las formas usuales de obtención de los mismos.

Obtención del tiempo de retardo τ

A pesar de que cualquier valor τ es válido en sistemas libres de ruido, una correcta estimación podría facilitar el análisis. Así, si un tiempo de retardo pequeño es escogido comparado con la escala de tiempo del sistema, vectores de estado fuertemente correlacionados serán producidos. Por otra parte, un valor grande generará vectores de estado casi independientes, dispersos sobre todo el espacio de fases [1].

La idea clave en el escogencia del tiempo de retardo es que los elementos que reconstruyen el vector de estado deben estar lo suficientemente cerca en el tiempo tal que estén dinámicamente relacionados, y sin embargo lo suficientemente lejos para que no resulten redundantes [2].

Una de las reglas más simples para estimar el valor de τ , es examinar en la serie de tiempo la función de correlación entre pares de datos como una función de su separación en el tiempo τ_0 , y definir una función de correlación como:

$$R(\tau_0) = \frac{\langle x(t) \cdot x(t + \tau_0) \rangle}{\langle x(t)^2 \rangle} \quad (2.6)$$

Con $\langle . \rangle$ denotando un promedio sobre todos los puntos en la serie de tiempo. Sobre esta función se determina el tiempo t_c en el que se presenta el primer cruce por

cero de $R(\tau_0)$, llamado tiempo de correlación. Debido a que se busca un τ que genere alta correlación pero que permita también que la dinámica del sistema sea exhibida, un pequeño múltiplo (2 o 3) del t_c es escogido como valor óptimo de τ [2].

Un enfoque más sofisticado usa la información mutua, que provee una medida de la probabilidad de que puntos vecinos, $x(t)$ e $x(t + \tau)$ no sean estadísticamente independientes. Si $P(x(t))$ y $P(x(t + \tau))$ son distribuciones de probabilidad de los puntos vecinos, y $P(x(t), x(t + \tau))$ es la distribución de probabilidad conjunta, entonces la información mutua se define como:

$$I(\tau) = \sum_t P(x(t), x(t + \tau)) \log \frac{P(x(t), x(t + \tau))}{P(x(t)).P(x(t + \tau))} \quad (2.7)$$

Para cierto τ lo suficientemente grande la probabilidad conjunta será igual al producto de las distribuciones separadas (marginales) y la información mutua será cero. Así un tiempo de retardo τ se escoge tal que la información mutua alcance su primer mínimo [35].

Obtención de la dimensión de embebimiento m

En cuanto a la estimación de m , se debe asegurar que dicho valor sea lo suficientemente grande para que el atractor reconstruido sea propiamente embebido en el sentido topológico [2]. A pesar de que cualquier valor mayor al m óptimo es aceptable, no es del todo recomendable dado que agregará coordenadas redundantes en el espacio de estados, entorpeciendo su posterior análisis.

El método típico de estimación de la dimensión de embebimiento es mediante el criterio de los falsos vecinos (False Nearest Neighbors – FNN), cuyo principio se basa en la búsqueda de puntos que sean vecinos en el espacio de fases, pero que no deberían serlo dado que su evolución temporal es diferente [1].

Para vectores de embebimiento $\mathbf{x}(t)$ como en (2.4). Los correspondientes vectores reconstruidos, correspondientes a los vecinos más cercanos serán:

$$\mathbf{x}^{NN}(t) = [x(t'), x(t' - \tau), \dots, x(t' - (m - 1)\tau)] \quad (2.8)$$

Si el vector de estado $\mathbf{x}^{NN}(t)$ es un verdadero vecino de $\mathbf{x}(t)$, entonces vino al vecindario de $\mathbf{x}(t)$ por orígenes dinámicos. Si es un falso vecino provino de una proyección de una dimensión mayor, debido a que la presente dimensión m no representa adecuadamente el atractor en el espacio de estados. Aumentando la dimensión a $m + 1$, el falso

vecino se remueve de la vecindad de $\mathbf{x}(t)$; así al revisar cada punto $\mathbf{x}(t)$ y buscar la dimensión en que se remueven todos los falsos vecinos, se logrará estimar la dimensión de embebimiento adecuada [36].

El criterio FNN se explica a continuación [37]:

Defínase $\hat{\mathbf{x}}(t)$ como el vector $\mathbf{x}(t)$ de dimensión $m + 1$, y $\hat{\mathbf{x}}(t)^{NN}$ el vector $\mathbf{x}(t)^{NN}$ de dimensión $m + 1$ tal como:

$$\hat{\mathbf{x}}(t) = [x(t), x(t - \tau), x(t - 2\tau), \dots, x(t - m\tau)] \quad (2.9)$$

$$\hat{\mathbf{x}}(t)^{NN} = [x(t'), x(t' - \tau), x(t' - 2\tau), \dots, x(t' - m\tau)] \quad (2.10)$$

Se tendrá que el incremento en la distancia entre estos dos puntos es dado solamente por la diferencia entre los últimos componentes. Así:

$$\|\hat{\mathbf{x}}(t) - \hat{\mathbf{x}}(t)^{NN}\|^2 - \|\mathbf{x}(t) - \mathbf{x}(t)^{NN}\|^2 = [x(t - m\tau) - x(t' - m\tau)]^2 \quad (2.11)$$

Normalizando la distancia entre ambos puntos se tendrá:

$$\frac{|x(t - m\tau) - x(t' - m\tau)|^2}{\|\mathbf{x}(t) - \mathbf{x}(t)^{NN}\|^2} \geq R_T \quad (2.12)$$

Donde R_T es un valor de umbral. Un R_T pequeño hará que verdaderos vecinos se cuenten como falsos, por el contrario uno grande hará que algunos falsos vecinos no sean incluidos. Valores típicos están entre 10 y 30, siendo 15 un buen punto de partida.

2.2 Embebimiento no uniforme

En señales con múltiples periodicidades fuertes con escalas de tiempo muy diferentes, un tiempo de retardo τ pequeño será óptimo para los componentes de alta frecuencia, mientras que para las de baja frecuencia un τ grande es apropiado. Un compromiso en valores es inadecuado para ambas escalas de tiempo [16].

Con esto en mente, el embebimiento no uniforme fue propuesto para tratar con diferentes escalas de tiempo dentro de una misma señal. Contrario al embebimiento convencional que usa un único valor de retardo τ , el embebimiento no uniforme usa un vector de retardos $\mathbf{l} = (l_1, l_2, \dots, l_n)$, y define la reconstrucción en (2.4) como:

$$\mathbf{x}(t) = [x(t - l_1), x(t - l_2), \dots, x(t - l_n)] \quad (2.13)$$

La selección de los elementos óptimos en \mathbf{l} presenta sin embargo un problema por la explosión combinatoria de embebimientos posibles a medida que la dimensión m aumenta [16] (m en el embebimiento no uniforme será el número de elementos en el vector de retardos).

En la gráfica 2.1 se ilustra el principio del embebimiento uniforme y no uniforme para una serie de tiempo.

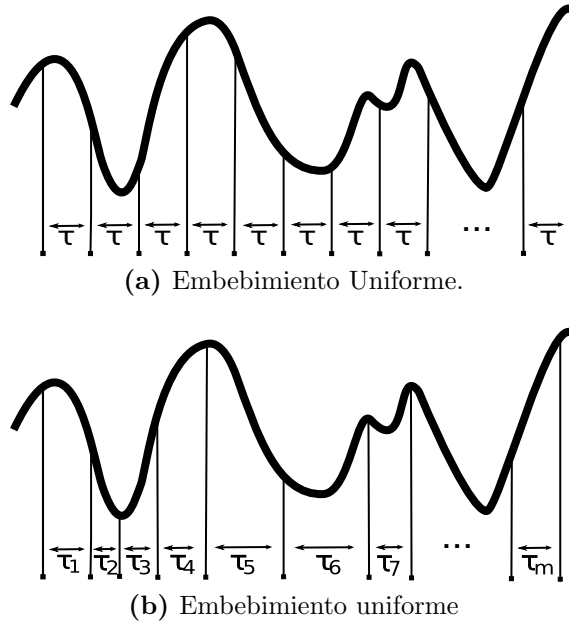


Figura 2.1: La Gráfica 2.1a ilustra el embebimiento uniforme con un único valor de retardo τ que implica intervalos muestreados a espaciamientos iguales. La gráfica 2.1b ilustra el embebimiento no uniforme m con diferentes valores de retardo muestreando la señal a espaciamientos diferentes.

A pesar de que pocas metodologías se pueden encontrar en la literatura para resolver el problema de la estimación óptima de \mathbf{l} , dos métodos serán presentados: el propuesto por Judd [16], y el propuesto por Ragulskis [18].

2.2.1 Embebimiento no uniforme, método de Judd

La reconstrucción óptima utilizando embebimiento no uniforme depende fundamentalmente de encontrar dos parámetros: el vector de retardos \mathbf{l} y la ventana de embebimiento d_w . Este último será el mayor elemento de \mathbf{l} , que puede pensarse como la anchura de una ventana que se desliza a través de la serie de tiempo, y será típicamente

definido por $l_n = d_w = \tau m$, con m la dimensión de embebimiento, y τ el tiempo de retardo como en el embebimiento uniforme.

Para hallar \mathbf{l} , Judd y Mees proponen un procedimiento en que conjuntamente se realiza embebimiento y modelado; así, el criterio correcto para optimizar el embebimiento es lo que se usa para medir la calidad del modelo [16].

El método propone primeramente modelar una serie de tiempo, tal que:

$$\mathbf{x}(t+1) = \mathbf{y}(t) = f(\mathbf{x}(t)) + \epsilon_t \quad (2.14)$$

Con f una función desconocida capaz de capturar la dinámica del sistema, ϵ_t los errores de modelado (asumidos Gaussiano por simplicidad) y $\mathbf{x}(\cdot)$ los vectores de estado reconstruidos como en (2.13).

Un problema que se presenta es la correcta selección de las funciones f capaces de modelar efectivamente el proceso. Para solucionar esto, Judd y Mees proponen usar funciones pseudo-lineales debido a las propiedades deseables que presentan: existencia de métodos eficientes para el ajuste de sus parámetros y su capacidad de modelar una gran variedad de no linealidades [38].

Éstas son de la forma:

$$f(\mathbf{x}(t)) = \sum_{i=1}^{mb} \lambda_i \phi_i(\mathbf{x}(t)) \quad (2.15)$$

Donde ϕ_i serán funciones bases del modelo y λ_i serán constantes de peso enteras. Ahora, reescribiendo (2.14) con (2.15), la ecuación se transforma en:

$$\mathbf{y}(t) = \sum_{i=1}^{mb} \lambda_i \phi_i(\mathbf{x}(t)) + \epsilon_t \quad (2.16)$$

Debido a que las funciones bases ϕ_i , serán en general, funciones no lineales que dependen de varios parámetros que deben ser optimizados, implicando a su vez procedimientos complicados, lentos, y tendientes a caer en mínimos locales; es propuesto generar un gran número de funciones base fijas, no solo al comienzo del modelamiento sino a medida que el proceso de construcción de modelos progrese, en vez de optimizar parámetros de unas cuantas funciones base.

Adicionalmente como estas funciones ϕ_i pueden pertenecer a muchos tipos diferentes, se suelen restringir a 3 familias específicas: función constante, funciones lineales

de los componentes coordenados de $\mathbf{x}(t)$, y funciones de base radial. Estas últimas son funciones que dependen solamente de la distancia radial de un punto llamado centro, y que serán $\phi_r(\frac{\mathbf{x}(\cdot)-c_j}{r_j})$ para centros c_j y radios r_j , y $\phi_r(\cdot)$ del tipo gaussiano, tophat gaussiano, sigmoide, wavelet, por nombrar algunas.

Ahora bien, para hallar el mejor modelo se debe encontrar también la cantidad óptima de funciones bases, es decir el valor del límite mb en (2.16), basando la escogencia en un criterio de información llamado Mínima Longitud de Descripción (Minimum Description Length – MDL).

A mayor cantidad de funciones base es posible obtener un mejor modelo, sin embargo demasiadas lo sobreajustarán demasiado al ruido. Por el contrario, un número insuficiente dejará de capturar las no linealidades propias de la serie de tiempo. Un compromiso es adecuado, y el MDL será un criterio que se encargará de decidir cuando se ha alcanzado el punto óptimo.

El criterio del MDL está basado en el principio de la navaja de Occam: lo más simple es lo mejor. Aplicado al contexto del modelado de series de tiempo, se debe entonces buscar el modelo más sencillo, que genere los menores errores de predicción y produzca buenas propiedades topológicas al reconstruir el atractor.

Para un modelo particular el MDL es proporcional al número de bytes de información requeridos para reconstruir la serie de tiempo original. Una analogía deja ver mejor el concepto: si se fueran a transmitir los datos, el MDL sería la compresión ganada al transmitir solo los parámetros de un modelo, el error de predicción y el modelo a utilizar, en vez de transmitir la serie de tiempo completa [16], es decir, el MDL de una serie de tiempo será la compresión de los datos de precisión finita, ganada por el modelado de los mismos.

Así, si un modelo es pobre será más económico simplemente enviar la serie de tiempo, si por el contrario el modelo esta sobreajustado la descripción del modelo aumenta (el modelo se ajusta demasiado al ruido), mientras que si el modelo es bueno, la descripción y los errores de predicción del modelo serán compactos [37].

La descripción de longitud de una serie de tiempo particular estará dada una expresión de la forma [39]:

$$DL \approx (\text{Número de datos}) \times \log(\text{suma de cuadrados de errores de predicción}) \\ + (\text{Penalización por precisión numérica de parámetros})$$

Habiendo definido el criterio de optimalidad usado por Judd, es posible ahora ex-

plicar el proceso de consecución de un modelo óptimo para un cierto vector de retardo $\mathbf{x}(\cdot)$:

Defínase \mathbf{V} :

$$\mathbf{V} = \begin{pmatrix} \phi_1(\mathbf{x}(1)) & \phi_2(\mathbf{x}(1)) & \cdots & \phi_{mb}(\mathbf{x}(1)) \\ \phi_1(\mathbf{x}(2)) & \phi_2(\mathbf{x}(2)) & \cdots & \phi_{mb}(\mathbf{x}(2)) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}(n)) & \phi_2(\mathbf{x}(n)) & \cdots & \phi_{mb}(\mathbf{x}(n)) \end{pmatrix}_{n \times mb} \quad (2.17)$$

Y λ un vector:

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{mb})_{mb \times 1}^T \quad (2.18)$$

El criterio de optimalidad está dado por la minimización sobre λ del error cuadrático medio de predicción $\mathbf{e} = \mathbf{y} - \mathbf{V}\lambda$, dado por:

$$\min_{\lambda} \{\|\mathbf{y} - \mathbf{V}\lambda\|\} \quad (2.19)$$

El problema de selección consiste ahora en encontrar la menor combinación lineal de funciones base ϕ_i , capaces de modelar efectivamente la serie de tiempo. En otras palabras se deben decidir cuales λ_i deberán ser diferentes de cero. Este será el llamado problema restringido de selección, que puede ser resuelto al minimizar para crecientes valores de k :

$$\begin{aligned} & \min_{\lambda} \{\|\mathbf{y} - \mathbf{V}\lambda\|\} \\ & \text{s.a. : } N(\lambda) = k, \quad k = 0, 1, 2, \dots \end{aligned} \quad (2.20)$$

Con $N(\lambda) = k$ el número de componentes no cero de λ .

Se tendrá entonces que hallar el MDL para cada k , aumentado su valor hasta que no haya un incremento significativo del MDL del modelo. En el caso particular, de modelos pseudo-lineales con error de ajuste distribuidos normalmente, el MDL será de la forma [38]:

$$S(k) = \left(\frac{n}{2} - 1\right) \ln \left(\frac{\mathbf{e}^T \mathbf{e}}{n}\right) + (k + 1) \left(\frac{1}{2} + \ln \gamma\right) - \sum_{j=1}^k \ln \delta_j \quad (2.21)$$

Donde γ está relacionado con la escala de los datos, $Q = n \hat{\mathbf{V}}^T \hat{\mathbf{V}} / \mathbf{e}^T \mathbf{e}$, δ resuelve $[Q\delta]_j = 1/\delta_j$, \mathbf{V} como en (2.17), y $\hat{\mathbf{V}}$ una matriz formada por columnas de \mathbf{V} que corresponden a elementos no cero de λ .

El resultado final de esta optimización será un modelo óptimo para un vector de retardos específico, y deberá ser repetido para todos los vectores de retardos posibles. Debido al alto costo computacional que eso implica, un vector de retardos inicial es analizado. Luego, es evaluada si la inclusión de un nuevo elemento provee un mejor modelo medido por vías del criterio del MDL. Este proceso es repetido hasta que se incluya la ventana de embebimiento d_w .

El Algoritmo 1 resume el procedimiento para un único vector de retardos $\mathbf{x}(\cdot)$ [40] [16].

Algoritmo 1 Algoritmo de Judd y Mees para hallar el embebimiento uniforme

1. Normalizar las columnas de \mathbf{V} (2.17).
 2. Sea $\mathbf{e}_B = \mathbf{y}$, $B = 0$. Defínase también $S_0 = \left(\frac{N}{2} - 1\right) \ln \left(\frac{\mathbf{y}^T \mathbf{y}}{N}\right) + 0.5 + \ln \gamma$.
 3. Sea $\mu = \mathbf{V}^T \mathbf{e}_B$ y j el índice del componente de μ con máximo valor absoluto. Los componentes del vector μ medirán que tan cerca cada una de las funciones base que no estén en uso actualmente alcanzarán el error del modelo actual. Se extiende el actual modelo con las funciones base que mejor encuentre el error del modelo: $B' = BU\{j\}$
 4. Se calcula (con la nueva base B') el $\lambda_{B'}$ tal que $\mathbf{y} - \mathbf{V}_{B'} \lambda_{B'}$ sea minimizado. Sea $u' = \mathbf{V}^T \mathbf{e}_{B'}$. Sea o el índice en B' correspondiente al componente de u' con el menor valor absoluto. o será el índice de la función base que hace la menor contribución al modelo actual extendido.
 5. Si $o \neq j$ entonces $B = B' \setminus o$ (se quita la función base o que menos aporte si esta no es la última que se agregado). Se calcula el λ_B tal que $\mathbf{y} - \mathbf{V}_B \lambda_B$ sea minimizado, junto a $\mathbf{e}_B = \mathbf{y} - \mathbf{V}_B \lambda_B$. Después se vuelve al paso 3.
 6. Si $o = j$ se encontró la mejor base. Defínase entonces $B_k = B$ con $k = |B|$. Encuentre δ tal que $(\mathbf{V}_B^T \mathbf{V}_B \delta)_j = \frac{1}{\delta_j}$ para cada $j = \{1, 2, \dots, k\}$ y calcúlese $S_k = \left(\frac{N}{2} - 1\right) \ln \frac{\mathbf{e}_B^T \mathbf{e}_B}{N} + (k + 1)(0.5 + \ln \gamma) - \sum_{j=1}^k \ln \delta_j$
 7. Si $S_k < S_{k-1}$ vaya a 3 y continúe hasta que el MDL deje de bajar.
 8. La base B_k tal que S_k es mínima es el modelo óptimo.
-

2.2.2 Embebimiento no uniforme, método de Ragulskis

Ragulskis et al. [18] proponen el uso de un método determinístico para la selección de un conjunto óptimo de vectores de retardo. Aunque esta selección es usada como argumento de entrada a sistemas de inferencia Fuzzy para evaluar su capacidad predictiva;

la presente tesis se enfoca en el estudio de la discriminancia que dicha metodología de embebimiento pudiera proveer.

El objetivo del método de Ragulskis es la consecución del máximo esparcimiento del atractor en el espacio de fases, para lo cual usa consideraciones geométricas buscando la mayor circularidad en todas las proyecciones planares del atractor embebido.

Con esto en mente, es generada entonces una función de calidad de que tan circular es el conjunto de proyecciones 2-dimensionales del atractor; luego, una función objetivo es propuesta para hallar los parámetros del vector de retardo \mathbf{l} que maximizan la circularidad de la serie embebida.

La reconstrucción para un embebimiento de este tipo, reescribe la ecuación (2.4) como:

$$\begin{aligned} \mathbf{x}(t) &= (x(t), x(t + \tau_1), \dots, x(t + \tau_1 + \dots + \tau_{(m-1)})) \\ t &= 1, 2, \dots, N - \sum_{i=1}^{m-1} \tau_i \end{aligned} \quad (2.22)$$

Y define la matriz de embebimiento \mathbf{M} como en (2.5), la distancia de tiempo entre dos puntos adyacentes en la serie de tiempo escalar como δ , y la ventana de embebimiento como $\delta \sum_{i=1}^{m-1} \tau_i$.

El análisis comienza con una reconstrucción 2-dimensional de una serie de tiempo armónica de la forma:

$$\begin{aligned} \mathbf{s}(t) &= A.\sin(\omega\delta(t - 1) + \Phi) \\ t &= 1, 2, \dots, N \end{aligned} \quad (2.23)$$

Donde A es la Amplitud, ω la frecuencia cíclica, y Φ la fase de la función armónica. El mapeo al espacio de estados 2-dimensional da origen a una elipse en un plano de fases con X_1 e X_2 como ejes coordenados, y cuya ecuación estará dada por:

$$X_2 = X_1.\cos(\omega\tau\delta) + \sqrt{A^2 - X_1^2}\sin(\omega\tau\delta) \quad (2.24)$$

Y los radios de las elipses mapeadas como:

$$\begin{aligned} r_1 &= A\sin(\omega\tau\delta)/\sqrt{(1 - \cos(\omega\tau\delta))} \\ r_2 &= A\sin(\omega\tau\delta)/\sqrt{(1 + \cos(\omega\tau\delta))} \end{aligned} \quad (2.25)$$

Se define ahora una función de calidad de la circularidad de la serie de tiempo como:

$$Q_1 = \frac{E}{\pi A^2} \quad (2.26)$$

Donde E es el área de la elipse mapeada, y el denominador será el área de un círculo de radio A en el espacio de estados. Un factor $Q_1 = 1$ será el caso ideal con un mapeo totalmente circular, y $Q_1 = 0$ será el peor caso con un mapeo hecho sobre una línea recta.

El factor de calidad (2.26) se puede reescribir utilizando (2.25) como:

$$Q_1 = \frac{E}{\pi A^2} = \frac{\pi r_1 r_2}{\pi A^2} = |\sin(\omega\tau\delta)| \quad (2.27)$$

Ahora, para el caso m -dimensional habrán $m(m-1)/2$ proyecciones planares del atractor en ejes coordenados $X_i - X_j$; $1 \leq i \leq m$, $1 \leq j \leq m$, $i \neq j$. Si el embebimiento es uniforme los τ_i serán iguales entre ellos, y se deberá definir un factor de calidad para todas las posibles combinaciones de j e i como:

$$Q_{ij} = |\sin(\omega(j-i)\tau\delta)| \quad (2.28)$$

Como Q_{ij} depende de la diferencia entre j e i , una nueva notación es presentada:

$$\begin{aligned} Q_1 &= Q_{ij|j-i|=1} \\ Q_2 &= Q_{ij|j-i|=2} \\ &\vdots \\ Q_{m-1} &= Q_{ij|j-i|=m-1} \end{aligned}$$

Se define finalmente una función de calidad de embebimiento como el promedio de los parámetros de calidad de embebimiento en todas las posibles proyecciones planares:

$$Q(\tau, \omega) = \frac{2}{m(m-1)} \sum_{k=1}^{m-1} (m-k) Q_k \quad (2.29)$$

$$= \frac{2}{m(m-1)} \sum_{k=1}^{m-1} (m-k) \sin(k\omega\tau\delta) \quad (2.30)$$

La anterior expresión se puede ampliar para el caso en el que se tenga un embe-

bimienta no uniforme formado por τ_i diferentes:

$$Q(\tau_1, \dots, \tau_{m-1}, \omega) = \frac{2}{m(m-1)} \left(\sum_{i=1}^{m-1} |\sin(\omega\delta\tau_i)| \right. \\ \left. + \sum_{i=1}^{m-2} |\sin(\omega\delta(\tau_i + \tau_{i+1}))| + \dots + |\sin(\omega\delta \sum_{j=1}^{m-1} \tau_j)| \right) \quad (2.31)$$

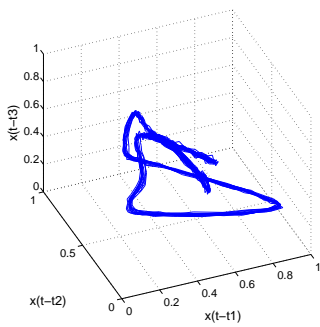
Debido a que cada componente armónico de la serie de tiempo original será afectado por la función de calidad cuando la serie de tiempo se embebe en el espacio de estados m -dimensional, se define una función objetivo $F(\cdot)$ que caracteriza la magnitud del esparcimiento del atractor en el espacio de estados, dada por:

$$F(\tau_1, \dots, \tau_{m-1}) = \frac{\pi}{2} \frac{\int_0^\infty A(\omega)Q(\tau_1, \dots, \tau_{m-1})d\omega}{\int_0^\infty A(\omega)d\omega} \quad (2.32)$$

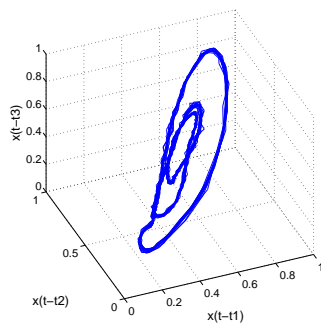
Con $A(\omega)$ la amplitud del espectro de Fourier de la serie de tiempo, y los límites de la integral involucrando todas las frecuencias posibles de la señal.

Ahora, una maximización de los parámetros de la función objetivo dará origen a la mejor combinación de vectores de retardo. En la presente tesis se utilizó para tales fines un algoritmo genético.

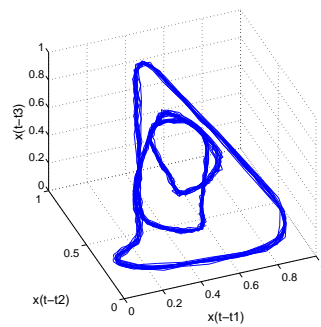
Finalmente para ilustrar el efecto del embebimiento en una señal fisiológica, se muestra en la figura 2.2 una gráfica comparativa de los tres métodos de embebimiento sobre un frame de una señal de voz.



(a) Embebimiento Uniforme. $\tau = 9$



(b) Embebimiento no uniforme de Judd. $\mathbf{l} = (1, 2, 5)$



(c) Embebimiento no uniforme de Ragulskis. $\mathbf{l} = (5, 10, 15)$

Figura 2.2: Reconstrucción de un atractor para un frame de una señal normal embebida en \mathbb{R}^3 . (a) muestra la reconstrucción para un embebimiento uniforme, (b) usando el método de embebimiento no uniforme de Judd, y (c) con el método de embebimiento no uniforme de Ragulskis

3. Caracterización

3.1 Introducción

Dado que el Teorema de Takens provee una reconstrucción topológicamente equivalente del espacio de fases original de un sistema, utilizando solamente una serie de observaciones; es deseable el cálculo de aquellas cantidades que permanecen invariantes ante los suaves cambios de coordenadas que la reconstrucción implica. Adicionalmente se busca que dichas cantidades no dependan de las condiciones iniciales que dieron origen al atractor en el espacio de fases, así como cierta robustez a los parámetros necesarios en la reconstrucción, de tal manera que puedan ser usadas para una válida caracterización del sistema dinámico.

Con esto en mente, se pueden encontrar dos tipos de medidas invariantes: Invariantes estáticas que dependen principalmente de la medida invariante (y su soporte), y que aparecen como dimensiones del atractor y exponentes de masa que tienen que ver con varias funciones de correlación estáticas del atractor. E invariantes dinámicas que dependen de las propiedades de las trayectorias, e incluyen varias entropías, exponentes de Lyapunov, y los momentos de las fluctuaciones en los exponentes de Lyapunov [41].

Entre la infinidad de medidas invariantes que pueden encontrarse, tanto estáticas como dinámicas, la caracterización utilizando dinámica no lineal típicamente suele utilizar al Máximo Exponente de Lyapunov, y la dimensión de Correlación. En la presente tesis sin embargo, se estudiará el poder caracterizante de las anteriores, junto con el Exponente de Hurst, y 4 medidas de la regularidad de series de tiempo, todas aproximaciones de la Entropía de Kolmogorov.

3.2 Dimensión de Correlación (d_2)

La dimensión, más que el número de variables dinámicas que se necesitan para representar un sistema dinámico, resultan ser un indicador del grado de complejidad de un sistema. Debido a esto, el rastreo de cambios en la dimensión debido a patologías,

puede llegar a convertirse en un criterio de diagnóstico útil en labores de detección de enfermedades [2].

Para estimar la dimensión de un objeto se han propuesto diferentes métodos; todos sin embargo, compartiendo la idea de medir el objeto a una escala r , de tal manera que se ignoren las irregularidades de tamaño menor a r y se estudie como las medidas varían conforme r tiende a 0.

El método más simple para estimar la dimensión de un objeto, consiste en cubrir el atractor con una grilla regular de cajas (hiper-volúmenes) de longitud r , tal que sea $M(r)$ el número de cajas que contienen al menos un punto:

$$M(r) \propto r^{d_B}, \quad r \rightarrow 0$$

Que reescrito quedará:

$$d_B = \lim_{r \rightarrow 0} \frac{\log(M(r))}{\log(r)} \quad (3.1)$$

d_B recibe el nombre de dimensión de capacidad ó de cajas (box counting dimension). En la práctica este estimado de dimensión no suele ser utilizado debido a su alto costo computacional, y a la necesidad de tener series de tiempo con alta resolución para obtener buenos estimados de dimensión [1].

En contraparte, Grassberger y Procaccia [42] propusieron un método para estimar la dimensión de una manera más eficiente. Sea $\mathbf{x}(\cdot)$ un vector de estado reconstruido como en (2.4), sea también $N_x(r)$ el número de puntos en el atractor dentro de una hiper-esfera de radio r centrada en $\mathbf{x}(\cdot)$. La mayoría de los puntos encerrados en la hiper-esfera vendrán por orígenes dinámicos, más que por cercanía con otros puntos correlacionados temporalmente. Así $N_x(r)$ será una medida de que tan frecuentemente una trayectoria visita un r -vecindario de $\mathbf{x}(\cdot)$. Si r es variado, se encontrará que a medida que el número de puntos en la esfera crece:

$$N_x(\varepsilon) \propto \varepsilon^{d_p} \quad (3.2)$$

Donde d_p será la llamada dimensión punto a punto (pointwise dimension) en $\mathbf{x}(\cdot)$. Debido a que d_p puede depender significativamente del $\mathbf{x}(\cdot)$ que se escoge, se suele obtener un promedio de $N_x(r)$ sobre muchos $\mathbf{x}(\cdot)$, tal que:

$$C(r) \propto r^{d_2} \quad (3.3)$$

Siendo d_2 la llamada dimensión de correlación (correlation dimension) y $C(r)$ la integral de correlación.

La dimensión de correlación será un identificador de la autosimilaridad de un atractor. Esta propiedad caracteriza estructuras geométricas que tienen comportamiento replicante bajo magnificaciones apropiadas, es decir, cuando una parte del objeto es observada bajo una magnificación adecuada, la parte aparentará ser igual a la totalidad, no habiendo manera de saber a que escala es visto el objeto con tan solo considerar una sección.

Un método efectivo para estimar la dimensión de correlación es utilizando el estimador de Takens, que para una serie de tiempo de longitud n , define la integral de correlación como:

$$C(r) = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i,j=1}^n \Theta(r - \|\mathbf{x}(i) - \mathbf{x}(j)\|) \quad (3.4)$$

Donde Θ es la función de Heaviside y r la medida de tolerancia.

Ampliando la definición en (3.3), se espera que a medida en que $r \rightarrow 0$ entonces $C(r) = \phi.r^v$; donde ϕ es una constante que refleja la lacunaridad, una medida de como los datos llenan el espacio [43]; y v es el exponente de correlación equivalente a d_2 .

Para una cierta cantidad N_p , de distancias r_p escogidas aleatoriamente entre $\mathbf{x}(\cdot)$, independientes y distribuidas aleatoriamente de acuerdo a la probabilidad:

$$P(r_p < r) = C(r) = \phi.r^v \quad (3.5)$$

El estimador de Takens, $T_2(r)$, dará el valor v que maximiza (3.5) como:

$$T_2(r) = \left[\frac{-1}{N_p - 1} \sum_{p=1}^{N_p} \log \left(\frac{r_p}{r} \right) \right]^{-1} \quad (3.6)$$

En el límite $r \rightarrow 0$ y $N_p \rightarrow \infty$; $T_2(r)$ será la estimación de d_2 [44].

3.3 Exponente de Hurst

El exponente de Hurst es una medida de la memoria a largo plazo del sistema, es decir, cuantifica la influencia de valores pasados en el valor actual de la serie de tiempo. Puede verse también como una medida de la persistencia del proceso a seguir tendencias o de

la correlación existente entre diferentes puntos en la serie de tiempo [21].

El exponente de Hurst varía en valor entre 0 y 1, donde un valor menor a 0.5 mostrará correlación negativa, implicando un proceso antipersistente en el que habrá tendencia a regresar al punto de donde se comenzó; así, si la tendencia de la serie es positiva, es posible que haya un cambio a tendencia negativa en la próxima iteración. Un valor cercano a 0.5 indicará que no habrá correlación entre los valores actuales y pasados, un indicador de movimiento Browniano. Un valor mayor a 0.5 indicará correlación positiva, y será un indicador de un proceso persistente en el que la dirección de la trayectoria tenderá a seguir igual.

El exponente de Hurst se puede calcular con el método del rango reescalado explicado en el Algoritmo 2:

Algoritmo 2 Algoritmo R/S para la estimación del exponente de Hurst

Supóngase una serie de tiempo $\mathbf{s}(t)$ de n elementos, dada en un periodo de observación $1 \leq t \leq T$. Se analiza la serie sobre un intervalo arbitrario $s + 1 \leq t \leq s + \tau$, cuya diferencia de tiempo es τ . Sobre un intervalo $[s + 1, s + \tau]$:

1. Calcúlese el valor medio: $\mu = \frac{1}{n} \sum_{i=1}^n x(i)$
 2. Calcúlese la media ajustada a los datos Y : $Y_t = x(t) - \mu$, $t = 1, 2, \dots, n$
 3. Calcúlese la serie acumulativa desviada Z : $Z_t = \sum_{i=1}^t Y_i$, $t = 1, 2, \dots, n$
 4. Calcúlese las series de rango R : $R_t = \max(Z_1, Z_2, \dots, Z_t) - \min(Z_1, Z_2, \dots, Z_t)$, $t = 1, 2, \dots, n$
 5. Calcúlese las series de desviación estándar S : $S_t = \sqrt{\frac{1}{t} \sum_{i=1}^t (x(i) - me)^2}$, con me la media desde $x(1)$ a $x(t)$.
 6. Calcúlese las series de rango Reescalado (R/S): $(R/S)_t = R_t/S_t$, $t = 1, 2, \dots, N$
 7. Se grafica (R/S) contra el tiempo en coordenadas logarítmicas. La pendiente de la línea de regresión de dicha gráfica se aproxima al exponente de Hurst.
-

3.4 Máximo Exponente de Lyapunov (LLE)

Un sistema dinámico se dice determinístico si el presente (estado inicial) completamente determina el futuro. A su vez, sistemas dinámicos usualmente aparentan ser caóticos cuando tienen dependencia sensitiva a condiciones iniciales, haciendo que dos trayecto-

rias que comienzan muy cerca juntas, rápidamente divergan una de otra logrando por tanto futuros diferentes [45].

El exponente propiamente promediado de este incremento es característico para el sistema, y cuantifica la fuerza de la no linealidad presente; este será el llamado Exponente de Lyapunov. En un sistema dinámico existen tantos exponentes de Lyapunov como dimensiones en el espacio de fases, sin embargo el más importante, es el Máximo Exponente de Lyapunov (Lyapunov Largest Exponent - LLE)[1].

Sean s_{n1} y s_{n2} dos puntos en el espacio con distancia $\|s_{n1} - s_{n2}\| = \delta_0 \ll 1$ y sea $\delta_{\Delta n}$ la distancia un tiempo después Δn , entre dos trayectorias emergiendo de esos puntos, $\delta_{\Delta n} = \|s_{n1+\Delta n} - s_{n2+\Delta n}\|$. Entonces LLE es determinado por:

$$\delta_{\Delta n} \simeq \delta_0 \exp(\lambda \Delta n), \delta_{\Delta n} \ll 1, \Delta n \gg 1$$

$$\lambda(\delta_0) = \lim_{\Delta n \rightarrow \infty} \lim_{\|\delta_0 \rightarrow 0\|} \frac{1}{\Delta n} \log \frac{\|\delta_{\Delta n}\|}{\|\delta_0\|}$$

El valor del LLE será un indicador del comportamiento del sistema, así: Valores positivos señalan caos¹ en el atractor, valores infinitos sugieren señales compuestas puramente por ruido, valores negativos mostrarán atractores con puntos fijos estables y valores cero mostrarán ciclos límites estables[1].

3.5 Entropías

La entropía cuantifica la incertidumbre de una variable aleatoria; esto es, la confusión. Así, si las probabilidades de un conjunto de futuros eventos tienden a ser iguales, menor será la información que se obtendrá sobre que lo pasará en el futuro, lo que a su vez significará un mayor estado de entropía [46].

Dada una variable aleatoria discreta X , con un rango de valores posibles $\Theta = \{x_1, x_2, \dots, x_n\}$, función de probabilidad masa $p(x_i) = P_r\{X = x_i\}$ y esperanza E , se definirá la Entropía de Shannon, denotada por $H(X)$, como [47, 48]:

$$H(X) = - \sum_{x_i \in \Theta} p(x_i) \log p(x_i) = -E[\log p(x_i)] \quad (3.7)$$

Para una serie de tiempo, es decir, la secuencia de n variables aleatorias $X_i =$

¹El término caos se usa por consideraciones históricas sobre las cuales el LLE fue desarrollado. En el análisis de series de tiempo se prefiere el término dinámica no lineal

X_1, \dots, X_n con un rango de valores $\Theta_1, \dots, \Theta_n$, se define una entropía conjunta:

$$H_n = - \sum_{x_1 \in \Theta_1} \dots \sum_{x_n \in \Theta_n} p(x_1, \dots, x_n) \log p(x_1, \dots, x_n) \quad (3.8)$$

$$H_n = H(X_1, X_2, \dots, X_n) \quad (3.9)$$

Con $p(x_1, \dots, x_n) = P_r\{X_1 = x_1, \dots, X_n = x_n\}$ la probabilidad conjunta para las n variables X_1, \dots, X_n .

Dado que:

$$H(X, Y) = H(X) + H(Y|X)$$

Se puede reescribir (3.9) como:

$$\begin{aligned} H(X_1, X_2) &= H(X_1) + H(X_2|X_1) \\ H(X_1, X_2, X_3) &= H(X_1) + H(X_2, X_3|X_1) \\ &\vdots \\ H(X_1, X_2, \dots, X_n) &= H(X_1) + H(X_2|X_1) + \dots + H(X_n|X_{n-1}, \dots, X_1) \\ H_n = H(X_1, X_2, \dots, X_n) &= \sum_{i=1}^n H(X_i|X_{i-1}, \dots, X_1) \end{aligned} \quad (3.10)$$

Esta última relación es llamada la regla de la cadena, y deja ver una propiedad fundamental de la entropía: El estado de un sistema en un instante X_n se determina por su historia X_1, X_2, \dots, X_{n-1} . Gracias a esto a medida que crece la serie de tiempo, la entropía conjunta también lo hará.

Para evitar esta dependencia entre la cantidad de variables aleatorias (longitud de la serie de tiempo) y el valor calculado de entropía conjunta se define entonces una tasa de entropía:

$$h = \lim_{n \rightarrow \infty} \frac{H_n}{n} \quad (3.11)$$

Para procesos estacionarios se puede probar que [48]:

$$\lim_{n \rightarrow \infty} \frac{H_n}{n} = \lim_{n \rightarrow \infty} H(X_n|X_{n-1}, \dots, X_1) \quad (3.12)$$

3.5.1 Entropía de Kolmogorov

Los anteriores conceptos se pueden ampliar para series de tiempo embebidas en el espacio de estados, al cubrir con cajas el atractor embebido y calcular la probabilidad de que un punto aleatorio esté sobre cada caja.

Considérese un sistema dinámico en \mathbb{R}^m , con m las variables dinámicas del sistema. Supóngase que el espacio de fase se particiona en N hipercubos de radio ϵ y de contenido ϵ^m , haciendo n mediciones del sistema espaciadas un τ determinado. Sea $p(k_1, k_2, \dots, k_N)$ la probabilidad conjunta de que el estado del sistema se encuentre en el hipercubo k_1 en el tiempo $t = \tau$, en el hipercubo k_2 en el tiempo $t = 2\tau$ y en el hipercubo k_N en tiempo $t = N\tau$. Sea también $N_b(i)$ el número de puntos dentro de la partición i y N_p el número total de puntos [47], tal que se puede escribir una relación para la probabilidad $P_b(\cdot)$ en cada partición, como la razón entre el número de puntos en la partición sobre el número total de puntos en el espacio de estados.

La Entropía de cada partición será expresada por la fórmula de Shannon como en la ecuación (3.7) [49], que permite ser reescrita en términos de la suma de correlación, para encontrar el estimado de entropía sin tener que recurrir a un algoritmo de conteo de cajas[50].

Para ello se define primero una función auxiliar $C_i(r)$, que contará el número de puntos que se encuentran a una distancia menor a r de un punto de referencia $\mathbf{x}(\cdot)$, escalado sobre el total de puntos de la serie de tiempo:

$$C_i(r) = \frac{1}{N_p} (\text{Número de puntos } \mathbf{x}(j) \text{ con } |\mathbf{x}(j) - \mathbf{x}(i)| < r) \quad (3.13)$$

Nótese que el numerador de $C_i(r)$ cuenta el número de puntos dentro del hipercubo de radio r (usando la norma Takens ó infinito); lo cual resulta ser una aproximación a $N_b(\cdot)$. Relacionando entonces con la probabilidad queda:

$$C_i(r) \approx \frac{N_b(i)}{N_p} = P_b(i) \quad (3.14)$$

La suma promediada sobre $C_i(r)$ será la suma de correlación $C(r)$:

$$C(r) = \frac{1}{N_p} \sum_{i=1}^{N_p} C_i(r) \quad (3.15)$$

$$C(r) = \frac{1}{N_p^2} (\text{Número de pares de puntos } (X_i, X_j) \text{ tal que } |X_i - X_j| < r) \quad (3.16)$$

$$C(r) = \frac{1}{N_p^2} \Theta(r - \|X(i) - X(j)\|) \quad (3.17)$$

Que también se puede relacionar con la aproximación de probabilidad que se hizo anteriormente:

$$C(r) = \frac{1}{N_p} \sum_{i=1}^{N_p} C_i(r) \approx \frac{1}{N_p} \sum_{i=1}^{N_p} P_b(i) \quad (3.18)$$

Reescribiendo la entropía de Shannon (3.7) se obtendrá finalmente:

$$\begin{aligned} H &= - \sum_b P_b \log P_b = - \sum_b \frac{N_b}{N_p} \log P_b = - \frac{1}{N_p} \sum_b \sum_{i \in b} \log P_b \\ &= - \frac{1}{N_p} \sum_i \log P_b(i) = - \frac{1}{N_p} \sum_i \log C_i(r) \end{aligned} \quad (3.19)$$

Esta entropía H se puede interpretar como la ganancia de información promedio al hacer una medición del sistema con una precisión r , o como la cantidad promedio de información requerida para especificar el estado de un sistema a una precisión r [49].

La entropía de Shannon ahora se amplía para todas las particiones, hallando una tasa de producción de información global sobre el atractor. Ésta será la llamada entropía de Kolmogorov (H_{KS}), que es obtenida usando la relación (3.11) y llevando el límite de tiempo a infinito para promediar variaciones, y los límites del tiempo entre mediciones y el del radio del hipercubo tendientes a cero [49].

Así:

$$H_{KS} = - \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n\tau} \sum_{k_1, \dots, k_n} p(k_1, \dots, k_n) \log p(k_1, \dots, k_n) \quad (3.20)$$

Si el sistema es estacionario, se puede reescribir como:

$$H_{KS} = - \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n\tau} H_n \quad (3.21)$$

Reemplazando (3.12) en (3.21), se tendrá:

$$H_{KS} = - \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{\tau} H(X_n | X_{n-1}, \dots, X_1) \quad (3.22)$$

Finalmente:

$$H_{KS} = \lim_{\tau \rightarrow 0} \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} (H_{n+1} - H_n) \quad (3.23)$$

Como las series de tiempo son de orden finito, a medida que H_n se vuelve más grande con respecto a la longitud de la serie de tiempo la tasa de entropía tiende a subestimarse [47], por lo que Grassberger y Procaccia [51] proponen el hallazgo de la entropía K_2 , un límite inferior sobre la entropía de Kolmogorov, que resulta ser fácilmente calculable de una señal experimental y que para casos típicos es numéricamente cercano a H_{KS} .

La ecuación propuesta es de la forma [52]:

$$K_2 = \lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \frac{1}{m} \lim_{n \rightarrow \infty} (-\log C^m(r)) \quad (3.24)$$

$$= \lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} \left(\log \left(\frac{C^m(r)}{C^{m+1}(r)} \right) \right) \quad (3.25)$$

Esta suma de correlación usa norma Euclídea, en contraposición a la fórmula de Kolmogorov que usa norma Takens.

Eckman y Ruelle hacen un refinamiento y proponen encontrar la entropía mediante [23]:

$$H_{ER} = \lim_{r \rightarrow 0} \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} [\phi^m(r) - \phi^{m+1}(r)] \quad (3.26)$$

Con:

$$\phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \log C_i^m(r) \quad (3.27)$$

3.5.2 Entropía Aproximada

Para cuantificar la entropía de una serie de tiempo se usa normalmente la entropía de Kolmogorov (o algunas de sus variantes); sin embargo su cálculo presenta problemas debido al ruido de baja amplitud presente en los datos que afecta seriamente el estimado de la misma. Ésto, sumado a la exigencia de series de tiempo largas (requeridas para alcanzar la convergencia de la entropía) y un costo computacional relativamente alto, hace evidente la necesidad de una medida basada en la entropía de Kolmogorov, capaz de cuantificar la regularidad de una serie de tiempo y de evitar sus inconvenientes [53].

Pincus [23] [53] propone entonces la Entropía Aproximada (Approximate Entropy – ApEn); una medida que se basa en la búsqueda de patrones (épocas) similares dentro de una serie de tiempo, y que ha resultado exitosa en el estudio y discriminación de enfermedades múltiples como: episodios de hipertensión intracraneal en pacientes pediátricos

con lesiones cerebrales traumáticas [28, 29], análisis sobre series de tiempo generadas por pacientes esquizofrénicos [30], estudios de la variabilidad del ritmo cardíaco por enfermedades y debido al envejecimiento [31], entre otras.

ApEn medirá el promedio negativo de los logaritmos naturales, de la probabilidad condicional de que dos secuencias similares de tamaño m permanecerán similares al aumentar el tamaño del patrón de comparación a $m + 1$ [24].

Dada una serie de tiempo de N puntos embebida en \mathbb{R}^m , con m la dimensión de embebimiento y la longitud del patrón a comparar, y matriz de embebimiento \mathbf{M} como en 2.5; se deberá llevar a cabo una comparación uno a uno de todos los vectores de estados en \mathbf{M} . El criterio de similaridad lo dará una cantidad r , que dirá similar un vector a otro si la diferencia entre sus componentes es menor a r . Luego es incrementado el tamaño del patrón a $m + 1$, es decir la serie es embebida en \mathbb{R}^{m+1} , calculándose la probabilidad condicional de que los vectores que se hayan escogidos como similares en la comparación del patrón m , continuarán siendo similares al aumentar la longitud del patrón a $m + 1$.

Nótese que el cómputo de la matriz de embebimiento requiere del cálculo del tiempo de retardo τ , para poder embeber la señal en \mathbb{R}^m y en \mathbb{R}^{m+1} . En el caso del embebimiento no uniforme de Judd, el tiempo de retardo no será un valor único escalar, sino un vector de escalares cuya cantidad determinará la dimensión m (ó $m + 1$) de embebimiento; lo que supone primeramente hallar el vector de embebimiento óptimo, para definir la dimensión m como la cantidad de elementos de este vector. La dimensión $m + 1$ exigirá entonces la consecución de un vector subóptimo de tamaño $m + 1$, para embeber en \mathbb{R}^{m+1} .

En Ragulskis el procedimiento es ligeramente diferente dado que las dimensiones m y $m + 1$ ya está determinadas de antemano, faltando solamente calcular vectores de retardo de dichas dimensiones para poder lograr los embebimientos.

Después de hacer las aclaraciones correspondientes se puede proceder a definir ApEn:

$$\text{ApEn}(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \quad (3.28)$$

Donde ϕ será:

$$\phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \log C_i^m(r) \quad (3.29)$$

Y $C_i^m(r)$ como en (3.13).

Para no depender de la amplitud absoluta de la señal el valor de r se suele escoger

como [24]:

$$r = \alpha \cdot std(\mathbf{s}(t)) \quad (3.30)$$

Con α una constante, y std la desviación estándar de la serie de tiempo.

Se describe en el Algoritmo 3 el procedimiento que debe seguirse para el cálculo de la medida [30].

Algoritmo 3 Algoritmo de Estimación de ApEn

1. Formar la matriz de embebimiento como en (2.5), y definir la distancia entre dos vectores reconstruidos $\mathbf{x}(i)$ y $\mathbf{x}(j)$ como:

$$d[\mathbf{x}(i), \mathbf{x}(j)] = \max_{k=1, \dots, m} (|\mathbf{x}(i+k-1) - \mathbf{x}(j+k-1)|) \quad (3.31)$$

2. Para un $\mathbf{x}(i)$ dado se cuenta el numero de j , tal que $d[\mathbf{x}(i), \mathbf{x}(j)] \leq r$, denotado como n_{im} .
3. Se define $C_i^m(r)$, que mide dentro de una tolerancia r , la frecuencia de patrones similares al dado de longitud de ventana m .

$$C_i^m(r) = \frac{n_{im}}{N - m + 1} \quad (3.32)$$

4. Se calcula el logaritmo de cada $C_r^m(i)$ y se promedia sobre i , definiendo a su vez $\phi^m(r)$:

$$\phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_i^m(r) \quad (3.33)$$

5. Se incrementa m , y se hallan $C_i^{m+1}(r)$ y $\phi^{m+1}(r)$.
6. Finalmente se define ApEn por

$$\text{ApEn}(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \quad (3.34)$$

3.5.3 Entropía Muestreada

ApEn presenta un problema que puede afectar la estimación de la medida de regularidad de la serie de tiempo estudiada. Para evitar la ocurrencia de $\log(0)$ en su cálculo, cada vector plantilla se cuenta a sí mismo en las comparaciones; situación que genera dos problemas principales: convierte a ApEn en una medida fuertemente dependiente

de la longitud de la serie de tiempo, haciendo que series de tiempo cortas tengan un estimado menor a lo esperado; y afecta la consistencia de la medida, esto es, si un conjunto de datos es de mayor complejidad que otro, deberá seguirlo siendo para todas las pruebas realizadas [24].

Richman et al. [24] proponen la Entropía Muestreada (Sample Entropy – SampEn) como una medida capaz de hacer frente a los problemas propios de ApEn. SampEn se define como:

$$\text{SampEn}(m, r, N) = -\log \left(\frac{A^m(r)}{A^{m+1}(r)} \right) \quad (3.35)$$

Donde m , r y N estarán definidos como en ApEn, y A será como (3.29), pero sin hacer auto-comparaciones para de esta manera evitar el sesgo.

El procedimiento que se debe seguir para la estimación de la estadística [30] se describe en el Algoritmo 4.

3.5.4 Entropía Aproximada de Kernel Gaussiano

Según [22], existen problemas en la validez y exactitud del estimado de regularidad a pesar de las mejoras que ofrece SampEn sobre ApEn, debido principalmente a la discontinuidad de la función de Heaviside sobre la que ApEn y SampEn están formuladas.

Para resolver dicho problema Xu et.al [22] propusieron la Entropía Aproximada de Kernel Gaussiano (Gaussian Approximate Entropy – GapEn), que reemplaza la función de Heaviside en la suma de correlación por una función de Kernel Gaussiano, lo que implica que (3.32) de ApEn se convierta en:

$$C_i^m(r) = \frac{\sum_{j=1, j \neq i}^{N-m+1} \exp \left(-\frac{(\|\mathbf{x}(i), \mathbf{x}(j)\|_1)^2}{10r^2} \right)}{N - m} \quad (3.41)$$

El Algoritmo para su cálculo será igual al Algoritmo 3, pero reemplazando la función en (3.32) por (3.41).

3.5.5 Entropía Fuzzy

La Entropía Fuzzy (Fuzzy Entropy – FuzzyEn) [26] es otra familia de estadísticas desarrollada para contrarrestar los efectos que la función de Heaviside produce sobre la estimación de regularidad. Contrario a ApEn y SampEn, FuzzyEn reemplaza la

Algoritmo 4 Algoritmo de Estimación de SampEn

1. Fórmese la matriz de embebimiento como en 2.5 y defínase la distancia entre $\mathbf{x}_m(i)$ y $\mathbf{x}_m(j)$ como en (3.31).
2. Para un $\mathbf{x}(i)$ dado, se cuenta el número de j tal que $d[\mathbf{x}_m(i), \mathbf{x}_m(j)] \leq r$, denotado como B_i .

3. Para todo i , se tiene:

$$B_i^m(r) = \frac{1}{N - m - 1} B_i \quad (3.36)$$

4. Se define $B^m(r)$ como:

$$B^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} B_i^m(r) \quad (3.37)$$

5. Se incrementa la dimensión a $m + 1$ y se calcula A_i como el número de ocurrencias en que $d[\mathbf{x}_{m+1}(i), \mathbf{x}_{m+1}(j)] \leq r$. Después se define $A_i^m(r)$ como:

$$A_i^m(r) = \frac{1}{N - m - 1} A_i \quad (3.38)$$

6. Luego $A^m(r)$ será:

$$A^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} A_i^m(r) \quad (3.39)$$

Así $B^m(r)$ es la probabilidad de que dos secuencias corresponderán para m puntos, mientras $A^m(r)$ es la probabilidad de que dos secuencias corresponderán para $m + 1$ puntos.

7. Se estima SampEn por:

$$\text{SampEn}(m, r, N) = -\ln \left[\frac{A^m(r)}{B^m(r)} \right] \quad (3.40)$$

función de Heaviside por una función de pertenencia fuzzy.

La Entropía Fuzzy se define como:

$$\text{FuzzyEn}(m, n, r, N) = \ln \phi^m(n, r) - \ln \phi^{m+1}(n, r) \quad (3.42)$$

Con:

$$\phi^m(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^m \right)$$

$$D_{ij}^m = \exp(-(d_{ij}^m)^n / r)$$

D_{ij}^m se ha definido para ser una función de pertenencia fuzzy sigmoidal con r e n parámetros que determinan la anchura, y el gradiente de la frontera de la función [54] y d_{ij}^m una función de distancia definida por:

$$d_{ij}^m = d[\mathbf{x}_i^m, \mathbf{x}_j^m] = \max_{k \in (0, m-1)} |\mathbf{x}(i+k) - u0(i) - (\mathbf{x}(j+k) - u0(j))| \quad (3.43)$$

Donde $u0(i) = m^{-1} \sum_{j=0}^{m-1} \mathbf{x}(i+j)$ removerá la línea base de los vectores de embebimiento.

El parámetro n actúa como el peso de la similaridad de los vectores de estado. Un valor $n > 1$ da peso al grado de similaridad de vectores cercanos y quita peso a los lejanos, mientras que un valor de $n < 1$ funciona al contrario [54]. Para los fines de la tesis se utiliza el valor típico $n = 2$.

Finalmente para ilustrar la idea general de los diferentes estimadores de regularidad se considera el ejemplo mostrado en la figura 3.1.

Los puntos similares a $x(1)$, es decir aquellos dentro de la franja $x(1) \pm r$ se pintarán de color azul, los similares a $x(2)$ con un punto rojo, y los similares a $x(3)$ con un punto verde. Ahora bien, considérese el vector plantilla de tamaño $m = 2$ y $\tau = 1$, dado por $x(1) - x(2)$ (azul, rojo). A todas las secuencias que presenten dicho patrón en la serie de tiempo, se les medirá la distancia y a través de alguna de las funciones ilustradas en la parte izquierda de la gráfica se les medirá el grado de similaridad del vector plantilla y del vector comparado. Luego se repite el procedimiento para la secuencia de tamaño $m+1 = 3$ dada por $x(1) - x(2) - x(3)$ (azul, rojo, verde), contando todas las incidencias del caso. Dicho procedimiento debe repetirse para todas las secuencias de tamaño m y $m+1$, computando finalmente el estimado de regularidad con alguno de los métodos explicados anteriormente.

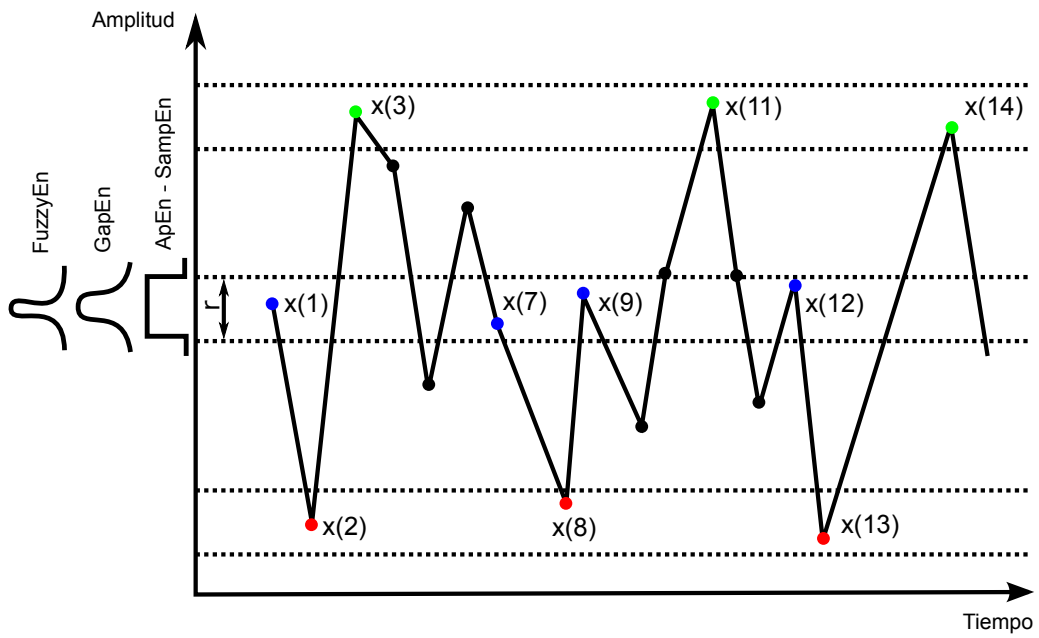


Figura 3.1: Ejemplo ilustrativo de los diferentes estimadores de regularidad

Parte III

Marco Experimental

4. Resultados

4.1 Metodología

Un sistema de detección de patologías basado en dinámica no lineal y análisis de complejidad, se muestra en la figura 4.1.

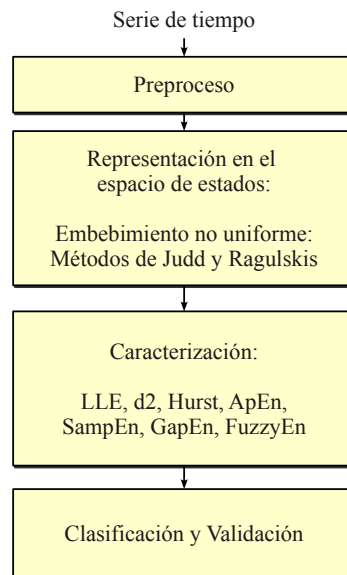


Figura 4.1: Metodología de un sistema de detección de patologías basado en dinámica no lineal y análisis de complejidad

La primera etapa en la metodología es el preproceso de las señales de entrada, que incluye procedimientos como la normalización y ventaneo de las señales a una longitud específica. La normalización empleada con el método de Judd es una cero-uno, que para una serie de entrada $s(\cdot)$, definirá la señal normalizada como:

$$s(\cdot)_{normalizada} = \frac{s(\cdot) - \min(s(\cdot))}{\max(s(\cdot)) - \min(s(\cdot))} \quad (4.1)$$

Con $\min(\cdot)$ y $\max(\cdot)$, el mínimo y máximo de la señal respectivamente.

Por otra parte, el método de Ragulskis mostró los mejores resultados al utilizar una normalización entre $[-1,1]$, por lo que dicho procedimiento fue empleado en todas las pruebas.

La segunda etapa implica el mapeo de la serie de tiempo a un espacio de estados en \mathbb{R}^m para su posterior análisis con dinámica no lineal. Típicamente, y como se explicó en la sección 2.1, el embebimiento uniforme es usado para estos fines; sin embargo, en el presente estudio se evaluará la influencia del embebimiento no uniforme en la detección de patologías, mediante dos enfoques diferentes: El presentado por Judd (sección 2.2.1), y el presentado por Ragulskis (sección 2.2.2). Este último procedimiento requiere una etapa de optimización, que para los fines de esta tesis será hecha utilizando algoritmos genéticos.

La tercera etapa en la metodología es la caracterización de la serie de tiempo embebida, para lo cual se incluyen 7 características divididas en dos grupos. El primer grupo está conformado por 3 características clásicas de la dinámica no lineal: El Máximo Exponente de Lyapunov, la Dimensión de Correlación y el Exponente de Hurst. La última medida a pesar de ser independiente del embebimiento utilizado, se incluye por haber contribuido a una mejora en las tasas de acierto en clasificación. El segundo grupo llamado de entropías incluye 4 características de complejidad basadas en la entropía de Kolmogorov que resultan ser estimadores de regularidad. Estos son: ApEn, SampEn, GapEn, y FuzzyEn.

La cuarta etapa es la de clasificación y validación, para la cual se emplea cross-validación 11-fold, con un 80% de muestras para entrenar clasificadores, y el 20% restante para evaluar desempeños. El clasificador escogido es un Modelo de mezcla de Gaussianas (Gaussian Mixture Models – GMM) al que se le varía el número de Gaussianas de 2 a 6. Los resultados se presentan en forma de matriz de confusión, a partir de los cuales se hallan la especificidad y sensibilidad definidas como:

$$Sensibilidad = \frac{VP}{VP + FN} \quad (4.2)$$

$$Especificidad = \frac{VN}{VN + FP} \quad (4.3)$$

donde VP son Verdaderos positivos, VN Verdaderos negativos, FP Falsos positivos, y FN Falsos negativos.

Adicionalmente el desempeño se medirá con curvas ROC y DET.

La metodología será probada en tres tipos de bioseñales: Voz, Electrocardiográficas (ECG) y Fonocardiográficas (FCG), usando los métodos de embebimiento no uniforme de Judd y Ragulskis. En total, se tendrán entonces 6 grupos de pruebas, que determinarán el desempeño en clasificación de cada característica por separado, así como agrupadas por conjuntos: clásicas, entropías, y todas las características.

4.2 Prueba sobre Voz

4.2.1 Base de Datos

Los sistemas para detección automática de voces patológicas han recibido gran atención en los últimos años, debido a sus ventajas sobre métodos tradicionales de detección de trastornos del sistema vocal; tales como su naturaleza no invasiva, la reducción en tiempo y costo de análisis, y la posibilidad de no solo identificar sino también monitorear las patologías en el sistema vocal [7].

En el presente grupo de pruebas se evaluará la metodología de la figura 4.1, en un esquema de detección automática de patologías vocales. Se usará como base de datos de prueba la desarrollada por The Massachusetts Eye and Ear Infirmary Voice Laboratory (MEEIVL) [55], cuyos registros contienen la fonación sostenida de la vocal $|ah|$ de pacientes con una variedad de patologías vocales: orgánicas, neurológicas, y desórdenes traumáticos; tomadas en un cuarto a prueba de sonido, a una frecuencia de muestreo de 44.1kHz y 16 bits de resolución. Un subconjunto de 173 registros de hablantes patológicos y 53 normales han sido seleccionados de acuerdo a lo enumerado por [56], para tener un diagnóstico y una distribución similar de edades entre ambos grupos.

Todas las grabaciones fueron remuestreadas a 25Khz, para posteriormente utilizar análisis de tiempo corto, con ventanas cuadradas de 55ms de duración y 50% de traslape. Adicionalmente para el cálculo de las entropías se utiliza un r como en (3.30) y un α de valor 0.35, como se describe en [57].

4.2.2 Resultados método de Judd y Mees

La tabla 4.1 muestra el mínimo, máximo, media y desviación estándar entre todas las características medidas usando el método de Judd y Mees. Los diagramas de cajas de

Característica	Clase	Máximo	Mínimo	Media	Std.
FuzzyEn	Normal	0,75	-0,07	0,11	0,08
	Patológico	0,96	-0,07	0,09	0,10
GapEn	Normal	17,26	6,46	10,32	1,31
	Patológico	37,88	6,45	14,31	4,13
SampEn	Normal	1,18	0,02	0,12	0,10
	Patológico	2,55	0,02	0,32	0,28
ApEn	Normal	1,18	0,01	0,10	0,09
	Patológico	1,53	0,00	0,31	0,26
LLE	Normal	0,07	-0,10	0,00	0,01
	Patológico	0,37	-0,22	0,00	0,04
d ₂	Normal	6,89	1,17	2,16	0,52
	Patológico	104,78	0,00	3,06	2,30
Hurst	Normal	0,76	-0,30	0,26	0,16
	Patológico	0,97	-0,08	0,68	0,20

Tabla 4.1: Máximo, mínimo, media y desviación estándar sobre cada uno de las características usando la técnica de embebimiento no uniforme de Judd en señales de voz

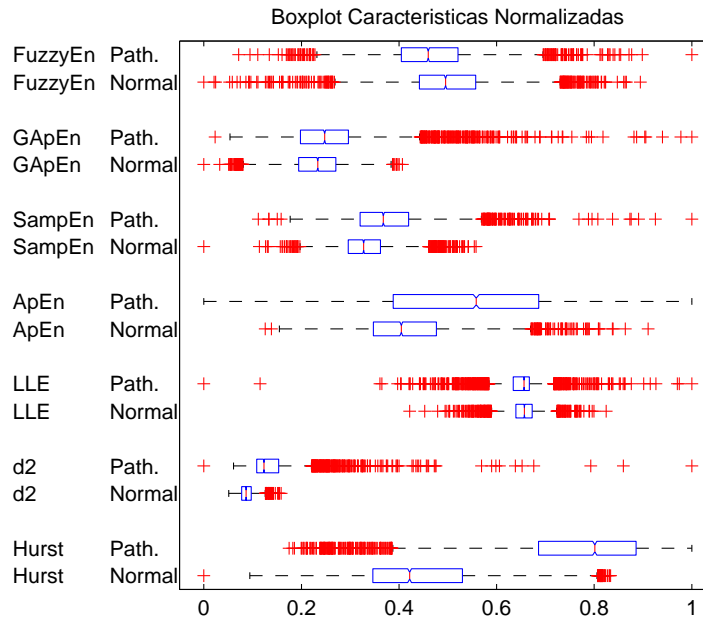


Figura 4.2: Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Judd en señales de voz

las clases patológicas y normal se muestran en la figura 4.2.

Del diagrama de cajas es posible observar el gran solapamiento entre clases de casi todas las medidas de entropía, especialmente FuzzyEn y GapEn. Por otra parte LLE a pesar de tener una distribución de datos muy concentrada (vista por el tamaño de

la caja en el diagrama) presenta una gran cantidad de atípicos tanto por debajo del cuartil 25 como por encima del 75, siendo más apreciable este fenómeno en la clase patológica.

d_2 muestra las cajas más pequeñas entre todas las medidas, lo que implica una gran concentración de valores sobre un rango pequeño. Es de resaltar también su gran cantidad de atípicos principalmente en la clase patológica, mostrándose valores tan altos como el dado por el máximo que en mucho supera el valor de la media. La gráfica también deja ver la aparente separabilidad entre clases que ofrecería la característica.

Por otra parte el Exponente de Hurst muestra poco traslape entre cajas normales y patológicas. Dicho fenómeno es visto en la tabla a través de medias diferentes entre clases con una pequeña desviación estándar (normal $0,26 \pm 0,16$ y patológica $0,68 \pm 0,20$).

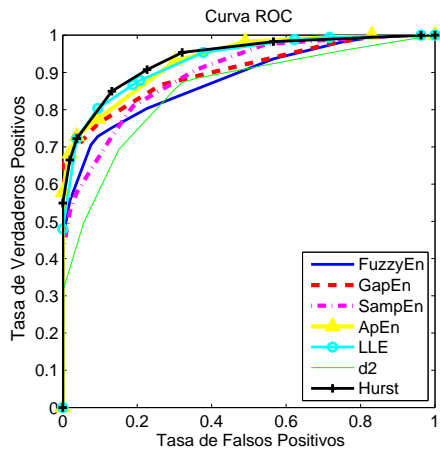
Resultados sobre características individuales

La tabla 4.2 muestra los resultados de clasificación para las características individuales. Las curvas ROC y DET para dichas pruebas se muestran en las figuras 4.3a y 4.3b respectivamente.

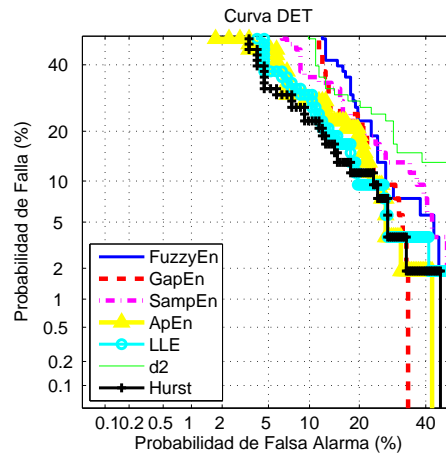
Características	Sensibilidad	Especificidad	Eficiencia
FuzzyEn	0.90	0.73	$77.43 \pm 5.45\%$
GapEn	0.83	0.76	$77.87 \pm 5.41\%$
SampEn	0.81	0.75	$76.54 \pm 5.52\%$
ApEn	0.92	0.78	$81.85 \pm 5.02\%$
LLE	0.81	0.87	$85.84 \pm 4.54\%$
d_2	0.73	0.79	$77.87 \pm 5.41\%$
Hurst	0.87	0.87	$87.61 \pm 4.29\%$

Tabla 4.2: Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Judd en señales de voz

Los resultados de la tabla muestran el poder caracterizante del Exponente de Hurst, por sobre las otras medidas con un 87.61% de acierto, seguida de LLE con un 85.84%. ApEn, mostró el mejor desempeño, con 81.85%, entre todas las características basadas en entropías, por sobre SampEn, GapEn, y FuzzyEn a pesar de que en teoría propiciaban la mejora en la estimación de complejidad. Las curvas ROC dejan ver el buen desempeño en términos generales de todas las características, mientras que la linealidad de las curvas DET hace pensar en la gaussividad de las características aplicadas



(a) Curva ROC para las características individuales usando la técnica de embebimiento de Judd



(b) Curva DET para las características individuales usando la técnica de embebimiento de Judd

Figura 4.3: Curvas de desempeño para el clasificador GMM, usando el método de embebimiento de Judd en señales de voz

al proceso, exceptuando quizás GapEn y d_2 .

Resultados por grupos de características

La tabla 4.3 muestra los resultados al combinar diferentes conjuntos de características.

Conjunto de Características	Sensibilidad	Especificidad	Eficiencia
Clásicas	0.94	0.94	94.24 ± 3.03%
Entropías	0.90	0.79	85.84 ± 4.54%
Todas	0.94	0.97	96.46 ± 2.40%

Tabla 4.3: Sensibilidad y Especificidad para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Judd en señales de voz

Los aciertos usando las características clásicas resultan ser mejores que los logrados utilizando solamente las basadas en entropía. Sin embargo, el uso en conjunto de ambos grupos de características, propició las mejores tasas de clasificación con un 96.46% de acierto.

4.2.3 Resultados método de Ragulskis

La tabla 4.4 muestra el mínimo, máximo, media y desviación estándar entre todas las características medidas usando el método de Ragulskis. Los diagramas de cajas de las clases patológicas y normal se muestran en la figura 4.4.

Característica	Clase	Máximo	Mínimo	Media	Std.
FuzzyEn	<i>Normal</i>	1,24	-0,42	0,45	0,20
	<i>Patológico</i>	1,13	-0,60	0,38	0,17
GapEn	<i>Normal</i>	16,47	7,21	11,18	1,28
	<i>Patológico</i>	31,66	7,15	13,06	3,25
SampEn	<i>Normal</i>	1,28	-0,86	0,41	0,21
	<i>Patológico</i>	2,98	-0,43	0,59	0,31
ApEn	<i>Normal</i>	0,89	-0,30	0,34	0,15
	<i>Patológico</i>	1,47	-0,23	0,46	0,21
LLE	<i>Normal</i>	0,04	-0,06	0,00	0,01
	<i>Patológico</i>	0,08	-0,15	0,00	0,01
d₂	<i>Normal</i>	3,50	1,12	1,94	0,31
	<i>Patológico</i>	33,09	0,00	3,03	1,26
Hurst	<i>Normal</i>	0,76	0,30	0,26	0,16
	<i>Patológico</i>	0,97	-0,08	0,68	0,20

Tabla 4.4: Mínimo, máximo, media y desviación estándar sobre cada una de las características, usando la técnica de embebimiento no uniforme de Ragulskis en señales de voz

En general los diagramas de cajas dejan ver una gran cantidad de atípicos presentes en todas las medidas.

También se muestra de nuevo a FuzzyEn (tal como con Judd) como la característica con mayor traslape, al igual que LLE que muestra una media y desviación estándar similar entre ambas clases, y un boxplot fuertemente traslapado con una gran cantidad de atípicos que hacen pensar en pobres desempeños en clasificación.

Por otra parte, d_2 muestra poco traslape entre cajas, y una buena concentración de valores en un rango muy pequeño, constatable en valores de medias diferentes entre clases con desviaciones estándar aceptables (normal $1,94 \pm 0,31$ y patológico $3,03 \pm 1,26$). Las otras medidas basadas en entropía muestran igualmente buenos resultados en términos de separabilidad aparente entre cajas.

Resultados sobre características individuales

La tabla 4.5 muestra los resultados de clasificación para las características individuales. Las curvas ROC y DET son también mostradas en las figuras 4.5a y 4.5b respectiva-

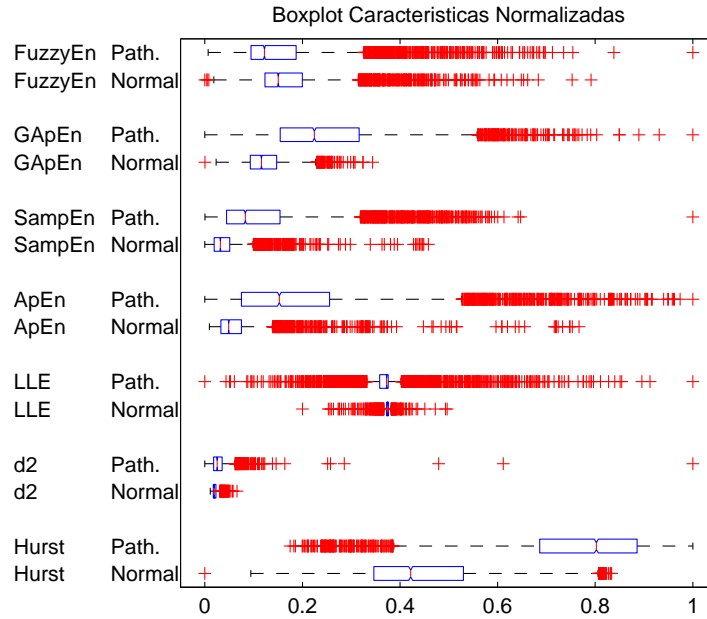


Figura 4.4: Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Ragulskis en señales de voz

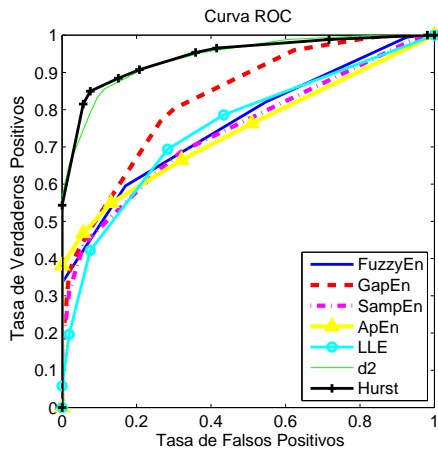
mente.

Característica	Sensibilidad	Especificidad	Eficiencia
FuzzyEn	0.73	0.65	67.25 ± 6.11%
GapEn	0.71	0.78	76.99 ± 5.48%
SampEn	0.79	0.54	60.17 ± 6.38%
ApEn	0.88	0.51	60.17 ± 6.38%
LLE	0.75	0.69	71.23 ± 5.90%
d₂	0.88	0.80	82.74 ± 4.92%
Hurst	0.87	0.88	87.61 ± 4.30%

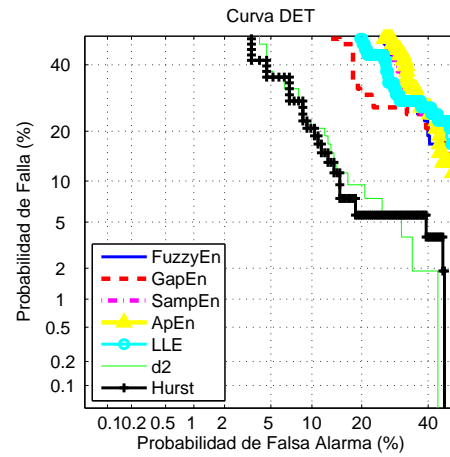
Tabla 4.5: Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Ragulskis en señales de voz

Los mejores resultados, al igual que con Judd, fueron obtenidos al usar el Exponente de Hurst con un 87.61% de acierto, seguido de d_2 con una tasa de 82.74%. Ambos resultados se aproximan mucho en las curvas ROC y DET, aunque manteniéndose siempre por encima Hurst.

Entre todas las medidas de regularidad, la que mostró las mejores desempeños fue GapEn, mientras que la peor tasa de acierto, en contraparte a lo del método de Judd,



(a) Curva ROC para las características individuales usando la técnica de embebimiento de Ragulskis



(b) Curva DET para las características individuales usando la técnica de embebimiento de Ragulskis

Figura 4.5: Curvas de desempeño para el clasificador GMM, usando el método de embebimiento de Ragulskis sobre señales de voz.

corrió por cuenta de ApEn con que anteriormente se habían mostrado los mayores aciertos en clasificación.

A partir de la linealidad de las curva DET es posible inferir la gaussividad de la gran mayoría de las característica con el método de Ragulskis sobre las señales de voz. Este fenómeno no es visto en Hurst y GapEn que muestran ciertas tendencias extrañas afectando la linealidad de la curva.

Resultados por grupos de características

La tabla 4.6 muestra los resultados al combinar diferentes conjuntos de características.

Tipo de características	Sensibilidad	Especificidad	Eficiencia
Clásicas	0.94	0.94	93.80 ± 3.14%
Entropías	0.71	0.83	80.53 ± 5.16%
Todas	0.94	0.94	93.80 ± 3.14%

Tabla 4.6: Sensibilidad, Especificidad y Acierto para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Ragulskis en señales de voz

Los mejores resultados resultan ser los dados por características clásicas con un 93.80%, en contraparte al 80.53% de las característica basadas en entropía. Aparente-

mente la inclusión de las características de entropía no mejoró el desempeño que mostraron las características clásicas por si solas.

Finalmente es de señalar el gran costo computacional del embebimiento no uniforme en las señales de voz, frente al esquema clásico de embebimiento. La tabla 4.7 muestra el tiempo de cómputo gastado en caracterizar una ventana para una señal normal y patológica, usando embebimiento uniforme y no uniforme. Las pruebas fueron hechas en Matlab 7.6, bajo Ubuntu 9.10, en un computador con 2gb de Ram y procesador Intel Core 2 Quad Q6700 de 2.67 GHz.

	Tiempo de cómputo (seg.)		
	Embebimiento no uniforme		Embebimiento Uniforme
Señales	Judd	Ragulskis	
Normal	785.78	752.86	2.14
Patológico	360.43	331.51	2.52

Tabla 4.7: Tiempo de cómputo por ventana usando embebimiento uniforme y no uniforme para señales normales y patológicas.

4.3 Prueba sobre base de datos de FCG

4.3.1 Base de Datos

Típicamente, los sonidos cardíacos son producidos por pares, en patrones repetidos regularmente entre latido y latido. El primer sonido conocido como $S1$, es producido por el cierre de las válvulas tricúspide y mitral, y el segundo, $S2$, es producido por el cerramiento de las válvulas aórtica y pulmonar. El intervalo de tiempo entre $S1$ y $S2$ se llama sístole, mientras que el intervalo entre $S2$ y $S1$ recibe el nombre de diástole. En condiciones normales, el flujo de sangre dentro del corazón es laminar y por tanto silencioso; pero cuando este se vuelve turbulento, causa vibración en los tejidos circundantes volviéndose perceptible acústicamente. Dicho fenómeno recibe el nombre de soplo, que de acuerdo en el instante en que se presente recibirá la denominación de sistólico o diastólico [58].

El uso de fonocardiogramas (FCG) permite representar gráficamente las señales acústicas del corazón, facilitando el análisis de cardiopatías como los soplos cardíacos.

En este trabajo, se utilizó una base de datos de señales FCG, tomada de 148 sujetos adultos mediante un estetoscopio electrónico, y un estándar de 3 derivaciones. Cada

señal fue digitalizada a 44.1 Khz con 16 bits por muestra. 50 pacientes fueron etiquetados como normales y 98 patológicos con soplos cardíacos, posterior diagnóstico del especialista. Cada señal FCG fue después dividida por latidos, escogiendo los 360 mejores, 180 de cada clase, después de una inspección visual y sonora dada por el cardiólogo para evitar la presencia de fenómenos como ruido u artefactos, que pudieren perjudicar el rendimiento de los algoritmos [58]. De los 180 segmentos patológicos solamente se trabajó con los que presentaban soplos diastólicos, un total de 66. Posteriormente cada señal FCG fue remuestreada a 4 Khz según lo indicado por [59], y segmentada para separar los eventos $S1$, $S2$, sístole y diástole. A pesar de que en [60] recomiendan un $\alpha=0.10$ para el cálculo de las entropías, las mayores discriminancias se lograron con un $\alpha=0.35$.

4.3.2 Resultados método de Judd y Mees

La tabla 4.8 muestra el mínimo, máximo, media y desviación estándar entre todas las características medidas usando el método de Judd y Mees. Los diagramas de cajas de las clases patológicas y normal se muestran en la figura 4.6.

Característica	Clase	Máximo	Mínimo	Media	Std
FuzzyEn	<i>Normal</i>	0,12	0	0,03	0,04
	<i>Patológico</i>	0,12	0,05	0,09	0,02
GapEn	<i>Normal</i>	0,9	0,41	0,47	0,1
	<i>Patológico</i>	0,49	0,44	0,45	0,01
SampEn	<i>Normal</i>	1,78	0,5	0,79	0,21
	<i>Patológico</i>	0,96	0,65	0,79	0,07
ApEn	<i>Normal</i>	1,91	0,47	0,86	0,23
	<i>Patológico</i>	1,03	0,85	0,94	0,04
LLE	<i>Normal</i>	0,28	-0,03	0,08	0,09
	<i>Patológico</i>	0,32	-0,06	0,11	0,09
d₂	<i>Normal</i>	2,02	1,36	1,62	0,16
	<i>Patológico</i>	1,83	1,3	1,51	0,11
Hurst	<i>Normal</i>	0,85	0,52	0,72	0,07
	<i>Patológico</i>	0,72	0,44	0,56	0,08

Tabla 4.8: Mínimo, máximo, media y desviación estándar sobre cada uno de las características, usando la técnica de embebimiento no uniforme de Judd y Mees en señales FCG

Los resultados muestran una aparente separabilidad entre clases de las características FuzzyEn (media normal 0.03 ± 0.04 , media patológica 0.09 ± 0.02), y Hurst (media normal 0.72 ± 0.08 , media patológica 0.56 ± 0.07), junto con un rango de valores entre

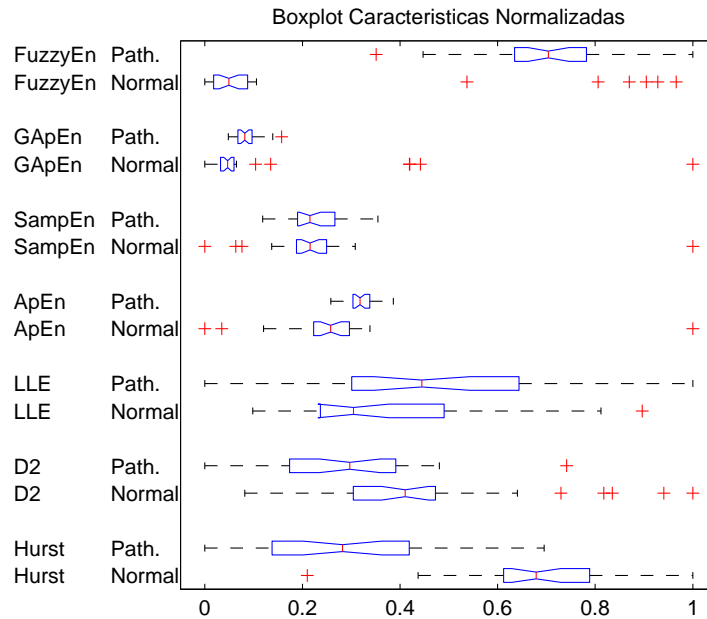


Figura 4.6: Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Judd en señales FCG

máximo y mínimo pequeño, que hace pensar en las buenas propiedades discriminantes de ambas medidas. Dicho resultado se puede comprobar en el poco traslape entre clases en el boxplot de la figura 4.6.

Por otra parte, SampEn muestra medias similares entre clases, aunque con una grande desviación estándar comparada con el rango de la señal. Dicho fenómeno es también constatable del gráfico y el aparente traslape entre cajas.

Visualmente también se pueden comprobar no muy grandes traslapes entre cajas en LLE y d_2 , y una mayor separabilidad en GapEn y ApEn.

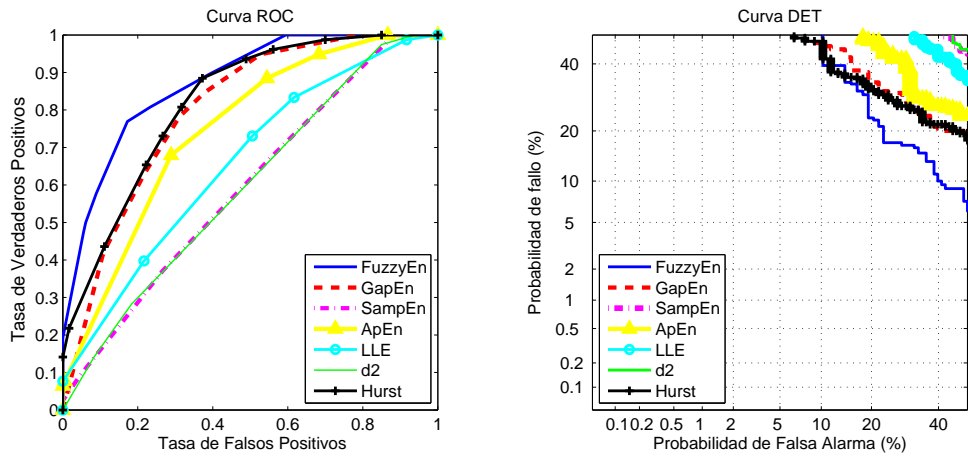
Resultados sobre características individuales

La tabla 4.9 muestra los resultados de clasificación para las características individuales. Las curvas ROC y DET para dichas pruebas se muestran en las figuras 4.7a y 4.7b respectivamente.

Los resultados muestran el poder en clasificación de FuzzyEn por sobre las otras características, con una eficiencia del $82.56 \pm 4.63\%$, seguida muy cerca por el Exponente de Hurst con un $79.07 \pm 4.96\%$. Las curvas ROC muestran la tendencia superior

Característica	Eficiencia	Sensibilidad	Especificidad
FuzzyEn	$82.56 \pm 4.63 \%$	0.78	0.94
GapEn	$70.16 \pm 5.58 \%$	0.66	0.80
SampEn	$55.04 \pm 6.07 \%$	0.54	0.56
ApEn	$66.28 \pm 5.77 \%$	0.62	0.77
LLE	$60.08 \pm 5.98 \%$	0.59	0.62
d₂	$54.65 \pm 6.07 \%$	0.52	0.60
Hurst	$79.07 \pm 4.96 \%$	0.82	0.72

Tabla 4.9: Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Judd y Mees en señales FCG



(a) Curva ROC para las características individuales usando la técnica de embebimiento de Judd en señales FCG

(b) Curva DET para las características individuales usando la técnica de embebimiento de Judd en señales FCG

Figura 4.7: Curvas de desempeño para el clasificador GMM, usando el método de embebimiento de Judd sobre señales FCG

de FuzzyEn por sobre el exponente de Hurst, constatando a su vez, la aparente separabilidad que mostraban los diagramas de cajas de ambas características.

Por otra parte d_2 junto con SampEn, mostraron los peores resultados en acierto. Dichos comportamientos son bien vistos en la curva ROC donde se muestran curvas prácticamente sobre la bisectriz de la gráfica.

La aparente linealidad de las curvas DET hace pensar en la gaussividad de las características sobre las señales FCG, a excepción sin embargo, de FuzzyEn donde se observa una ligera desviación en la gráfica.

Resultados por grupos de características

La tabla 4.10 muestra los resultados al combinar diferentes conjuntos de características.

Conjunto de Características	Eficiencia	Sensibilidad	Especificidad
Clásicas	77.91 ± 5.06 %	0.75	0.85
Entropías	97.29 ± 1.98 %	0.99	0.93
Todas	96.90 ± 2.12 %	0.98	0.94

Tabla 4.10: Sensibilidad y Especificidad para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Judd en señales FCG

Los mejores resultados se lograron al utilizar las características basadas en entropía, con un desempeño superior al 97%. Las características clásicas alcanzaron apenas un 77.91%. Mientras que la combinación de características clásicas y basadas en entropías lograron un 96.90% de acierto, significando una disminución en desempeño a lo mostrado por las características clásicas por si solas.

4.3.3 Resultados método de Ragulskis

Resultados sobre características individuales

La tabla 4.11 muestra el mínimo, máximo, media y desviación estándar entre todas las características medidas usando el método de Ragulskis. Los diagramas de cajas de las clases patológicas y normal se muestran en la figura 4.8.

La tabla muestra a FuzzyEn con la misma media y desviación estándar entre clases, lo que se traduce en un gran traslape entre clases en los diagramas de cajas. LLE presenta un fenómeno parecido pero seguido de desviaciones estándares un poco diferentes entre clases.

Por otra parte, Hurst muestra medias diferentes con desviaciones estándares pequeñas y un rango pequeño de valores mínimo-máximo, que hacen pensar en buena discriminancia.

Visualmente, el boxplot deja ver también el gran traslape en GapEn, aunque con un rango muy concentrado de valores para la clase patológica y uno muy amplio con la clase normal (std. normal 0.32, y std. patológico 0.03).

Característica	Clase	Máximo	Mínimo	Media	Std
FuzzyEn	<i>Normal</i>	0,15	0,01	0,06	0,04
	<i>Patológico</i>	0,20	0,00	0,06	0,04
GapEn	<i>Normal</i>	0,84	0,04	0,49	0,32
	<i>Patológico</i>	0,48	0,32	0,41	0,03
SampEn	<i>Normal</i>	1,42	0,94	1,13	0,10
	<i>Patológico</i>	1,91	0,82	1,39	0,29
ApEn	<i>Normal</i>	1,52	1,02	1,21	0,12
	<i>Patológico</i>	1,77	0,54	1,27	0,35
LLE	<i>Normal</i>	0,79	0,10	0,51	0,14
	<i>Patológico</i>	0,98	-0,05	0,56	0,20
D2	<i>Normal</i>	1,80	0,99	1,46	0,33
	<i>Patológico</i>	1,74	1,22	1,47	0,12
Hurst	<i>Normal</i>	0,88	0,52	0,71	0,06
	<i>Patológico</i>	0,85	0,44	0,59	0,08

Tabla 4.11: Mínimo, máximo, media y desviación estándar sobre cada uno de las características, usando la técnica de embebimiento no uniforme de Ragulskis en señales FCG

La tabla 4.12 muestra los resultados de clasificación para las características individuales. Las curvas ROC y DET son también mostradas en las figuras 4.9a y 4.9b respectivamente.

Característica	Eficiencia	Sensibilidad	Especificidad
FuzzyEn	54.48 ± 7.81 %	0.51	0.57
GapEn	92.30 ± 4.18 %	0.92	0.92
SampEn	80.77 ± 6.18 %	0.91	0.70
ApEn	80.12 ± 6.26 %	0.88	0.71
LLE	56.41 ± 7.78 %	0.66	0.46
d₂	88.46 ± 5.01 %	0.88	0.88
Hurst	75.64 ± 6.73 %	0.76	0.74

Tabla 4.12: Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Ragulskis en señales FCG

Los resultados muestran el desempeño superior de GapEn por sobre las otras características con un 92.30%, contra un 88.46% de d₂.

En estas pruebas se presentaron dos resultados totalmente chocantes a comparación a lo mostrado por el método de Judd. Allí FuzzyEn presentaba los mejores resultados, mientras que el presente método lo muestra con los peores aciertos. Por otra parte d₂ que con Judd mostraba las peores tasas de clasificación, presenta ahora los segundos mejores aciertos.

Las curvas ROC muestran el superior desempeño de GapEn por sobre las otras

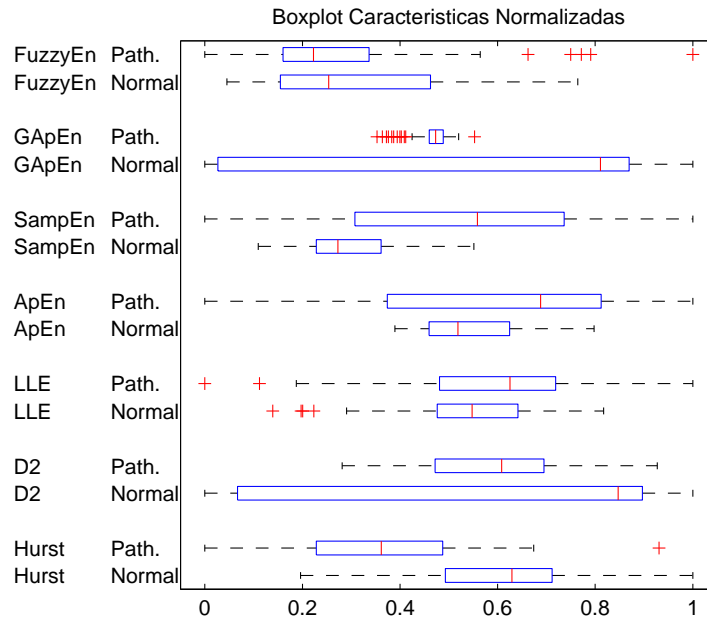
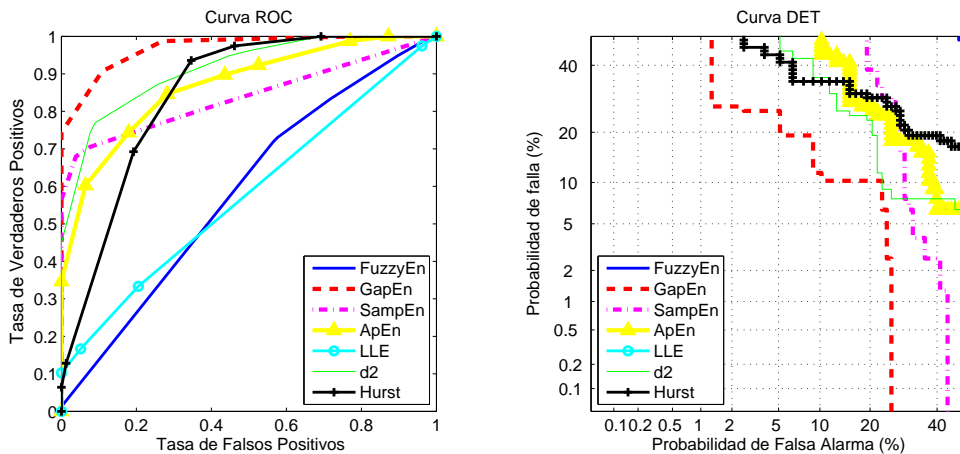


Figura 4.8: Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Ragulskis en señales FCG



(a) Curva ROC para las características individuales usando la técnica de embebimiento de Ragulskis

(b) Curva DET para las características individuales usando la técnica de embebimiento de Ragulskis

Figura 4.9: Curvas de desempeño para el clasificador GMM, usando el método de embebimiento de Ragulskis sobre señales FCG.

características y el pobre desempeño de LLE y d_2 . Las curvas DET no muestran ninguna linealidad por lo que no se podría inferir gaussividad de las características

sobre el proceso.

Resultados por grupos de características

La tabla 4.13 muestra los resultados al combinar diferentes conjuntos de características.

Conjunto de Características	Eficiencia	Sensibilidad	Especificidad
Clásicas	$91.02 \pm 4.48 \%$	0.89	0.92
Entropías	$94.23 \pm 3.65 \%$	0.88	1
Todas	$94.23 \pm 3.65 \%$	0.88	1

Tabla 4.13: Sensibilidad, Especificidad y Acierto para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Ragulskis en señales FCG

Las pruebas muestran el buen desempeño de las medidas basadas en entropía (94.23%) por sobre las características clásicas (91.02%). La inclusión de las características clásicas en las pruebas basadas en entropía no contribuyó de ninguna manera a mejorar el desempeño en clasificación.

4.4 Prueba sobre HRV

4.4.1 Base de Datos

La apnea obstructiva del sueño (Obstructive sleep apnea – OSA) es un desorden médico caracterizado por un colapso anormal de la vía aérea faríngea durante el sueño, que produce despertares repetitivos. Las presentaciones clínicas más comunes de la apnea incluyen ronquidos, pausas en la respiración, y excesiva somnolencia durante el día. La apnea también está asociada con una gran variedad de consecuencias relacionadas con la salud que puede interferir con las actividades normales, incluyendo un riesgo incrementado de choques automovilísticos, anormalidades cognitivas como la depresión, y enfermedades cardiovasculares [61].

La polisomnografía (PSG) durante la noche entera es el método más confiable de diagnóstico de la enfermedad, que implica la medición de varios parámetros fisiológicos que incluyen: EEG, ECG, Electromiografía (EMG) de rodillas y piernas, flujo nasal, Electrooculograma (EOG), exámenes del flujo nasal, movimientos abdominales y torácicos, y saturación de oxígeno en la sangre; presenta sin embargo, problemas debido al alto costo del sistema y la incomodidad de los electrodos conectados al cuerpo.

Dado que se ha mostrado relación entre la apnea y variaciones en el ritmo cardíaco (Heart Rate Variability – HRV), un análisis de este tipo de señales puede resultar productivo en tareas de diagnóstico relacionadas con la detección de la apnea [62].

La base de datos utilizada en el estudio es una de Apnea-ECG de 70 registros grabados en horas de sueño, con señales ECG y respiratorias de buena calidad. Las grabaciones fueron seleccionadas a partir de dos estudios anteriores en los cuales el ECG no fue el foco de análisis. En el primer estudio, llevado a cabo entre 1993 y 1995, se investigó el efecto de la apnea obstructiva del sueño en la presión arterial en sujetos con apnea del sueño moderada y severa. Todos los sujetos fueron monitorizados por dos noches consecutivas, y luego de un intervalo de 4 semanas, se repitieron las grabaciones por dos noches más, consecutivas. Los ECGs fueron digitalizados a 100Hz. Se seleccionaron 27 registros de 9 sujetos para la base de datos. El número de registros por sujeto varía entre 1 y 4, dependiendo de la calidad de la señal. El segundo estudio fue llevado a cabo entre 1998 y 1999, con el fin de crear un conjunto estándar de registros de sueño. En este estudio, el foco de la investigación fue el de la grabación de un electroencefalograma de varios canales. Este estudio se llevó a cabo con voluntarios saludables y pacientes seleccionados con apnea del sueño. Todos los sujetos se grabaron durante dos noches consecutivas. Las señales ECG se digitalizaron a 200Hz, y luego fueron remuestreadas a 100Hz para tener consistencia con los datos del primer estudio. Se seleccionaron 43 registros de 23 sujetos de este segundo estudio para ser incluidos en la base de datos; por cada sujeto no se seleccionaron más de dos registros.

Las señales fueron etiquetadas por expertos en los intervalos en donde se presentaban desórdenes en la respiración (apneas o hipoapneas), marcando el inicio y el final de cada episodio. Posteriormente las etiquetas se mapearon con una resolución de 1 minuto. La duración de los registros ECG varía entre 401 y 578 minutos. Con base en el número de minutos con desórdenes respiratorios, se definen tres grupos de registros: los registros que tienen menos de 5 minutos de desórdenes respiratorios se colocaron en el grupo normales o de control (20 registros); el grupo de apnea se define como registros con 100 o más minutos con desórdenes respiratorios (40 registros); finalmente, se define un grupo intermedio llamado apnea límite con algunas apneas de importancia incierta (10 registros) [63]. La base de datos estará compuesta finalmente por 1800 segmentos de 3 minutos para clase, muestreados a 100Hz en donde la etiqueta de cada segmento corresponde a la etiqueta del minuto central. El parámetro de entropías usado es un $\alpha=0.2$, tal como se indica en [64].

4.4.2 Resultados método de Judd y Mees

La tabla 4.14 muestra el mínimo, máximo, media y desviación estándar entre todas las características medidas usando el método de Judd. Los diagramas de cajas de las clases patológicas y normal se muestran en la figura 4.10.

Características	Clase	Máximo	Mínimo	Media	Std
FuzzyEn	<i>Normal</i>	1,12	-0,04	0,03	0,06
	<i>Patológico</i>	0,26	-0,03	0,01	0,03
GapEn	<i>Normal</i>	406,06	16,31	62,28	50,45
	<i>Patológico</i>	388,16	14,62	44,75	35,65
SampEn	<i>Normal</i>	1,92	0,04	0,33	0,28
	<i>Patológico</i>	1,75	0,04	0,23	0,21
ApEn	<i>Normal</i>	1,94	0,05	0,39	0,29
	<i>Patológico</i>	1,84	0,04	0,28	0,24
LLE	<i>Normal</i>	3,38	-0,10	0,44	0,21
	<i>Patológico</i>	1,52	-0,29	0,40	0,19
d₂	<i>Normal</i>	2,54	0,93	1,26	0,17
	<i>Patológico</i>	2,54	0,81	1,20	0,17
Hurst	<i>Normal</i>	1,00	0,05	0,85	0,12
	<i>Patológico</i>	1,00	0,72	0,95	0,03

Tabla 4.14: Mínimo, máximo, media y desviación estándar sobre cada uno de las características, usando la técnica de embebimiento no uniforme de Judd en señales HRV

Se observa una gran cantidad de atípicos junto a gran solapamiento entre clases en los diagramas de cajas de casi todas las medidas de entropía, fenómeno especialmente visto en FuzzyEn y constatable por un gran rango entre mínimo y máximo de ambas clases, una media casi igual (0.03 para normal y 0.01 para patológico) y una gran desviación estándar (0.06 normal y 0.03 patológico).

GapEn por su parte muestra un rango máximo-mínimo grande, con una desviación estándar relativamente alta, y una gran cantidad de atípicos.

Tanto ApEn como SampEn dejan ver en el diagrama una relativa separabilidad entre cajas. Mientras que Hurst, a pesar de tener el menor solapamiento entre cajas normales y patológicas (media normal 0.85 ± 0.12 y media patológica 0.95 ± 0.03), presenta una gran cantidad de atípicos especialmente en la clase normal.

Resultados sobre características individuales

La tabla 4.15 muestra los resultados de clasificación para las características individuales. Las curvas ROC y DET para dichas pruebas se muestran en las figuras 4.11a y

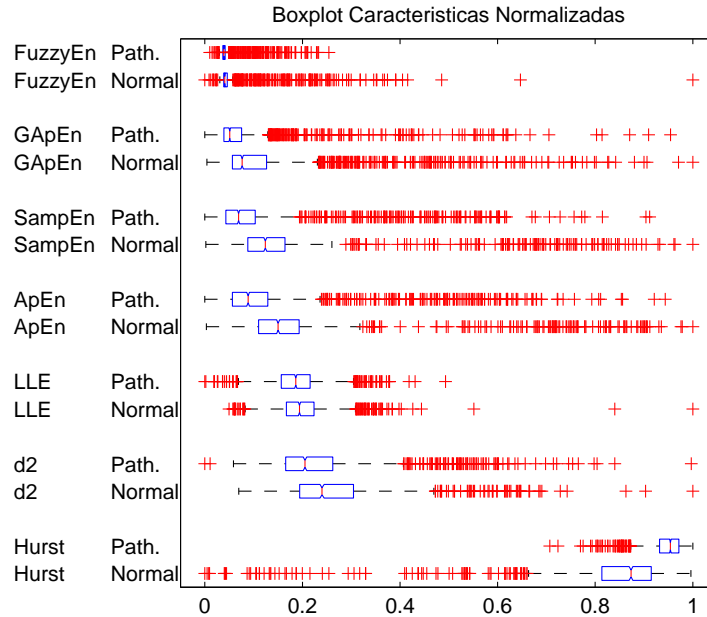


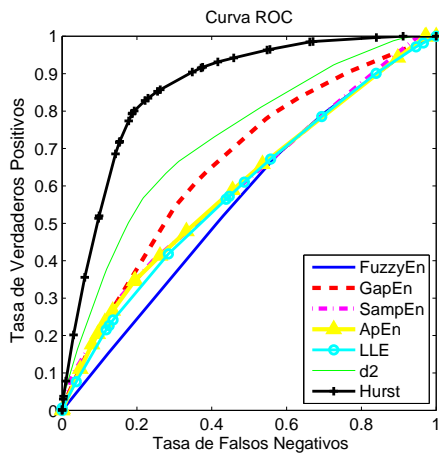
Figura 4.10: Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Judd en señales HRV

4.11b respectivamente.

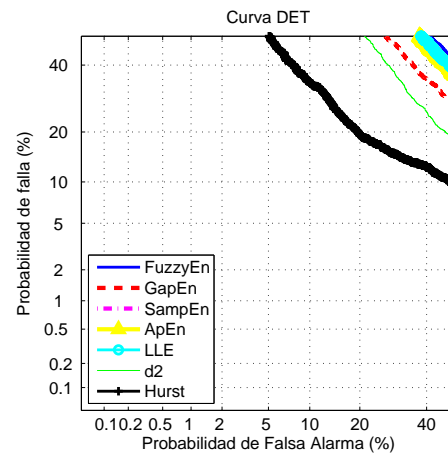
Característica	Sensibilidad	Especificidad	Eficiencia
FuzzyEn	0.79	0.46	62.64 ± 1.58 %
GapEn	0.84	0.52	67.78 ± 1.52 %
SampEn	0.64	0.80	72.36 ± 1.46 %
ApEn	0.66	0.78	72.02 ± 1.47 %
LLE	0.68	0.38	53.16 ± 1.63 %
d₂	0.63	0.56	59.47 ± 1.60 %
Hurst	0.78	0.83	80.66 ± 1.29 %

Tabla 4.15: Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Judd en señales HRV

Los mejores resultado fueron mostrados por Hurst con un acierto de 80.66 ± 1.29 , seguido de ApEn y SampEn, con un acierto de alrededor del 72%. Las otras medidas no superaron el 70%, inclusive con resultado de clasificación tan bajos como un 53.16 ± 1.63 dado por LLE. Las curvas ROC y DET constatan los pobres resultados obtenidos por casi todas las características, mostrando a la gran mayoría sobre la bisectriz de la gráfica, y a un exponente de Hurst que supera ampliamente los desempeños de las



(a) Curva ROC para las características individuales usando la técnica de embebimiento de Judd



(b) Curva DET para las características individuales usando la técnica de embebimiento de Judd

Figura 4.11: Curvas de desempeño para el clasificador GMM, usando el método de embebimiento de Judd sobre señales HRV

demás características.

La linealidad de las curvas DET deja ver también la gaussividad del proceso con todas las características utilizadas.

Resultados por grupos de características

La tabla 4.16 muestra los resultados al combinar diferentes conjuntos de características.

Conjunto de Características	Sensibilidad	Especificidad	Eficiencia
Clásicas	0.78	0.84	$81.11 \pm 1.28\%$
Entropías	0.66	0.79	$72.88 \pm 1.45\%$
Todas	0.72	0.83	$77.72 \pm 1.36\%$

Tabla 4.16: Sensibilidad y Especificidad para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Judd en señales HRV

Los resultados muestran que el desempeño de las características clásicas es el mejor por sobre los otros grupos de pruebas con un 81.11%. El desempeño con las características basadas en entropías resulta ser bajo (72.88%), y su combinación con las características clásicas reduce el buen desempeño que estas presentan por sí solas.

4.4.3 Resultados método de Ragulskis

La tabla 4.17 muestra el mínimo, máximo, media y desviación estándar entre todas las características medidas usando el método de Ragulskis. Los diagramas de cajas de las clases patológicas y normal se muestran en la figura 4.12.

Característica	Clase	Máximo	Mínimo	Media	Std
FuzzyEn	<i>Normal</i>	0,97	0,06	0,34	0,12
	<i>Patológico</i>	0,80	0,07	0,34	0,10
GapEn	<i>Normal</i>	255,57	9,29	31,43	14,22
	<i>Patológico</i>	235,96	11,22	29,14	19,38
SampEn	<i>Normal</i>	2,25	0,27	0,86	0,23
	<i>Patológico</i>	1,39	0,16	0,81	0,20
ApEn	<i>Normal</i>	1,79	0,34	0,69	0,24
	<i>Patológico</i>	1,32	0,13	0,70	0,22
LLE	<i>Normal</i>	0,62	0,14	0,19	0,11
	<i>Patológico</i>	0,92	0,27	0,15	0,12
d₂	<i>Normal</i>	3,05	1,39	2,03	0,27
	<i>Patológico</i>	2,63	1,09	1,77	0,24
Hurst	<i>Normal</i>	1,00	0,05	0,85	0,12
	<i>Patológico</i>	1,00	0,72	0,95	0,03

Tabla 4.17: Mínimo, máximo, media y desviación estándar sobre cada uno de las características, usando la técnica de embebimiento no uniforme de Ragulskis en señales HRV

Se muestra un gran traslape entre clases en casi todas las características, especialmente en las basadas en entropía, hecho comprobable, de medias parecidas entre clases: FuzzyEn (normal $0,34 \pm 0,12$ y patológico $0,34 \pm 0,12$), GapEn (normal $31,43 \pm 14,22$ patológico $29,14 \pm 19,38$), SampEn (normal $0,86 \pm 0,23$, patológico $0,81 \pm 0,20$), ApEn ($0,69 \pm 0,24$, patológico $0,70 \pm 0,22$).

d₂ muestra poco traslape entre cajas, mientras que de nuevo Hurst presenta el menor traslape entre cajas de todas las características analizadas.

Resultados sobre características individuales

La tabla 4.18 muestra los resultados de clasificación para las características individuales. Las curvas ROC y DET son también mostradas en las figuras 4.13a y 4.13b respectivamente.

De nuevo el resultado más fuerte es el dado por Hurst, con un acierto del $80.67 \pm 1.29\%$, y unas curvas ROC y DET que claramente superan al resto de las características.

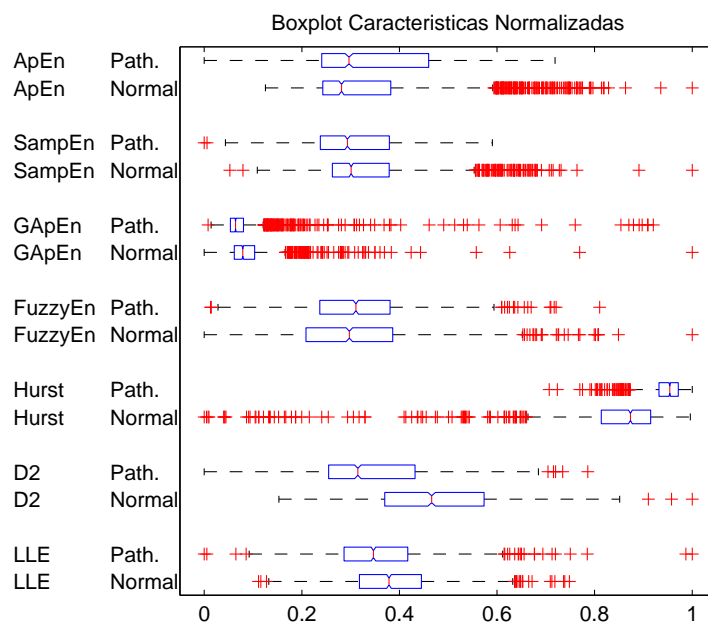


Figura 4.12: Diagrama de cajas normalizado usando la técnica de embebimiento no uniforme de Ragulskis en señales HRV

Característica	Sensibilidad	Especificidad	Eficiencia
FuzzyEn	0.37	0.74	55.55 ± 1.62 %
GapEn	0.61	0.69	64.83 ± 1.56 %
SampEn	0.66	0.56	60.91 ± 1.59 %
ApEn	0.77	0.44	60.25 ± 1.59 %
LLE	0.60	0.52	56.27 ± 1.62 %
d₂	0.80	0.60	70.39 ± 1.49 %
Hurst	0.78	0.83	80.67 ± 1.29 %

Tabla 4.18: Sensibilidad, Especificidad y Eficiencia de clasificación para las características individuales usando la técnica de embebimiento no uniforme de Ragulskis en señales HRV

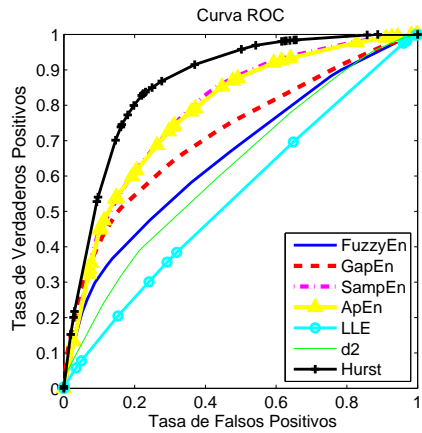
FuzzyEn y LLE dieron los peores resultados de acierto con valores inferiores al 60%.

La linealidad de las curvas DET deja ver la gaussividad de las características sobre el proceso.

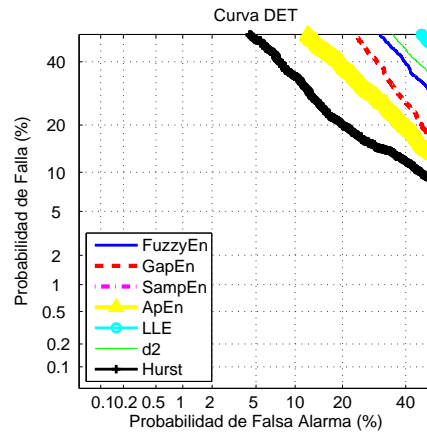
Resultados por grupos de características

La tabla 4.19 muestra los resultados al combinar diferentes conjuntos de características.

Las características clásicas son las que mejor acierto proveen con un 80.22%. Las



(a) Curva ROC para las características individuales usando la técnica de embebimiento de Ragulskis



(b) Curva DET para las características individuales usando la técnica de embebimiento de Ragulskis

Figura 4.13: Curvas de desempeño para el clasificador GMM, usando el método de embebimiento de Ragulskis sobre señales HRV.

Tipo de características	Sensibilidad	Especificidad	Eficiencia
Clásicas	0.78	0.82	80.22 ± 1.30%
Entropías	0.70	0.71	70.66 ± 1.49%
Todas	0.76	0.82	79.19 ± 1.32%

Tabla 4.19: Sensibilidad, Especificidad y Acierto para diferentes conjuntos de características usando la técnica de embebimiento no uniforme de Ragulskis en señales HRV

basadas en entropía dan un acierto mucho menor al conseguido con las clásicas, y afectan el desempeño en las pruebas de todas las características en conjunto, desmejorando el acierto que las características clásicas tienen por si solas.

Parte IV

Discusiones y Conclusiones

5. Discusiones

Tres tipos de bioseñales fueron analizadas: Voz, FCG, y señales HRV, todas de morfología y comportamientos muy diferentes. Las señales de voz, por el hecho de ser producidas por el sostenimiento en la pronunciación de una vocal, presentan una frecuencia fundamental definida así como diferentes armónicos. Esto la hace ideal para el estudio con embebimientos no uniformes capaces de trabajar con las diferentes periodicidades que presentan. Las señales FCG son básicamente señales acústicas, que a pesar de tener eventos repetitivos y sincronizados muy marcados, presentan dinámicas rápidas seguidas de silencios o eventos patológicos. Las señales HRV por otra parte no presentan ni morfología, ni eventos definidos.

Las pruebas sobre señales de voz mostraron los buenos resultados del esquema metodológico usando las técnicas basadas en embebimiento no uniforme, especialmente la propuesta por Judd. Con esta se presentaron los mejores resultados, al lograr tasas de acierto del $96.46 \pm 2.40\%$, a expensas de un costo computacional muy alto. La técnica de Ragulskis alcanzó tasas de del $93.80 \pm 3.14\%$ a un costo computacional alto, pero menor a Judd.

Los resultados dejan ver también que los conjuntos de características basados en medidas de regularidad mostraron tasas de desempeño menores a las de las características clásicas de la dinámica no lineal.

Por otra parte, las medidas basadas en entropía mostraron una fuerte dependencia al embebimiento, que hicieron que las tasas de acierto entre método y método variara considerablemente. Tanto así, que ApEn en Judd mostró los mejores resultados, mientras que en Ragulskis generaba los peores desempeños. Es de resaltar también el pobre desempeño en general de FuzzyEn que a pesar de ser pensada para mejorar las estimaciones de regularidad de ApEn y SampEn, no demostró un gran poder caracterizante.

Comparando los resultados obtenidos con los dados por [57], en que el acierto máximo fue del 92.7% utilizando un esquema de embebimiento uniforme con medidas de complejidad, se demuestra el desempeño superior del esquema metodológico presentado por sobre lo dado por la referencia.

Las pruebas con señales FCG mostraron buenos resultados. Con la técnica de Judd, se logró un desempeño máximo del 97.29% al utilizar las características basadas en entropía, mientras que con la técnica de Ragulskis se logró un máximo en acierto del 94.23% con las características clásicas. Cabe destacar, como con estas bioseñales la metodología que logró los mejores aciertos, no dependió de Hurst (y su implícita independencia al embebimiento) para elevar las tasas de clasificación.

En [58], con un esquema de pruebas similar usando solamente características clásicas con embebimiento uniforme, se lograron tasas de acierto del 97.73%. Los resultados muestran que Judd puede lograr desempeños comparables a lo mostrado por el estado del arte, aunque con conjuntos de características diferentes, mientras que Ragulskis de nuevo muestra un desempeño inferior a lo encontrado con el embebimiento uniforme.

Nuevamente se mostraron resultados contradictorios que no permiten afirmar que una característica es mejor que otra. En el caso de Ragulskis, d_2 mostró buenos desempeños, contrario a lo mostrado con Judd donde esta misma característica mostró los peores aciertos. Esto también es visto en como los mejores aciertos con Judd, se lograron con las características basadas en entropía, mientras que con Ragulskis fueron las clásicas las que obtuvieron los mejores desempeños.

Las pruebas sobre las señales HRV mostraron resultados comparables con el estado del arte. Con el método de Judd se lograron aciertos máximos de $81.11 \pm 1.28\%$ con sensibilidad del 78% y especificidad del 84% usando solamente las características clásicas, al igual que con Ragulskis, en que estas mismas características lograron aciertos de $80.22 \pm 1.30\%$, sensibilidad del 78% y especificidad del 82%.

Hurst demostró ser la características que proveía los mejores desempeños, en contraparte a LLE que mostró los peores aciertos con ambos métodos de embebimiento no uniforme. Debe destacarse la no contribución del embebimiento no uniforme a las tasas de acierto; esto es comprobable de la fuerte dependencia en Hurst para elevar desempeños. En la literatura se puede encontrar procedimientos para la detección de OSA basados en análisis con dinámica no lineal que han mostrado resultados similares a los encontrados en el presente trabajo. En [64] utilizaron ApEn como indicador de la presencia de OSA en señales HRV, logrando sensibilidades máximas del 75% y especificidad del 61.1%. Adicionalmente, mediante el uso de otra característica basada en la dinámica no lineal llamada titración numérica (Numerical Titration – NL) lograron sensibilidad del 81.3% y especificidad del 72.2%. En el presente trabajo, utilizando solamente ApEn se lograron resultados de sensibilidad de 66% y de especificidad de 78% con el método de Judd, y sensibilidad del 77% y especificidad del 44% con el

método de Ragulskis. A pesar de que los resultados con ApEn muestran ser menores, es de destacar como NL mostró resultados aproximadamente similares a los dados por el conjunto de características clásicas usando ambos métodos de embebimiento no uniforme.

En otro estudio para la detección de OSA usando señales HRV [65], se lograron aciertos del 73% usando ApEn, y del 57% usando otra medida de las correlaciones a largo término de la serie de tiempo llamada *Detrended fluctuations analysis (DFA)*. El resultado de ApEn de la referencia es más consistente con lo hallado en el presente trabajo con el método de Judd, donde se tienen aciertos de $72.02 \pm 1.47\%$.

Por otra parte, en [66] se lograron aciertos del 85%, sensibilidad del 73.36% y especificidad de 89.33%, al usar 3 grupos de características: características medidas en el tiempo, características basadas en la correlación, y otro grupo usado normalmente en labores de interfaz cerebro-computador. En este último grupo se incluye el Máximo Eigenvalor Normalizado de la matriz de embebimiento y una medida basada en entropía, con los que se lograron por si solos sensibilidades de 73.36%, y especificidades del 87.38%. Este último resultado es muy aproximado también a lo logrado en con Judd principalmente, aunque el resultado de las 3 características en conjunto muestran resultados superiores a los encontrados con el esquema metodológico.

Los bajos resultados en clasificación pueden ser debidos al fuerte preproceso que ha tenido la base de datos después de su obtención, debido a que los filtrados pueden "blanquear" la dinámica no lineal del sistema y afectar fuertemente el análisis [67].

6. Conclusiones y trabajo futuro

6.1 Conclusiones

En términos generales el esquema metodológico presentado, utilizando el embebimiento no uniforme de Judd, mostró desempeños superiores o equivalentes a lo presentado al estado del arte. En el único caso en que no se lograron resultados cuanto menos comparables fueron en las pruebas con señales HRV; esto probablemente producido por el fuerte preproceso que llevaba consigo la base de datos que podría haber disminuido los aciertos en clasificación.

Es de destacar también, que a pesar de que Judd puede ayudar a mejorar las tasas de acierto, el costo computacional se eleva ostensiblemente. Tal como lo mostró la tabla 4.7, el tiempo de cálculo con el método de Judd podía ser hasta 300 veces superior al tiempo de cálculo del embebimiento uniforme. Resulta por tanto importante considerar si es más importante el tiempo de cómputo o la precisión en alguna aplicación particular, para definir si el esquema metodológico resultará ser productivo para los fines buscados, o si más bien uno basado en embebimiento uniforme es mejor aplicable.

Por otra parte, el método de Ragulskis mostró desempeños inferiores a los del método de Judd en todos los casos de prueba, y en muchas ocasiones a los del embebimiento uniforme; por lo que se puede llegar a pensar, a partir de los resultados obtenidos, que el método de Ragulskis no resulta recomendable en labores de reconocimiento de patrones, siendo a veces preferible el enfoque clásico.

Sobre las características se puede decir que las basadas en entropía mostraron una alta dependencia al embebimiento, presentándose una falta de consistencia en los resultados entre método y método. Por otra parte, las características clásicas mostraron, en buena parte de las pruebas, un desempeño superior al mostrado por las entropías. Sin embargo la combinación de ambos tipos de características ayudó en ciertos casos a aumentar el acierto de clasificación, por lo que no debería descartarse el uso de las características basadas en entropía en labores de reconocimiento de patrones.

Es de resaltar el buen desempeño mostrado por el exponente de Hurst que en varias

pruebas mostró ser la característica que generaba los mayores desempeños, y que en muchos casos resultó favorecida del "blindaje" a los embebimientos. Cabe destacar, que en general tanto ApEn como SampEn (a excepción de unos cuantos casos como en FCG) mostraron mejores aciertos que GapEn y FuzzyEn, a pesar de que en teoría deberían mejorar la estimación de regularidad.

En resumen se podría afirmar a la luz de los resultados, que el esquema metodológico presentado, utilizando el embebimiento no uniforme de Judd, ayuda a mejorar los desempeños en tareas clasificatorias en comparación con esquemas similares basados en embebimiento uniforme; esto a expensas de un costo computacional relativamente mayor. El método de Ragulskis presenta en casi todos los casos desempeños menores a las técnicas clásicas de embebimiento, que hacen pensar que es inadecuado en labores de reconocimiento de patrones.

En cuanto a las medidas basadas en entropías, se puede señalar que entre prueba y prueba los desempeños variaron mucho, por lo que no es deducible la existencia de una medida de regularidad superior a las demás.

6.2 Trabajo Futuro

Como trabajo futuro se propone el estudio y desarrollo de otras técnicas basadas en embebimientos no uniformes, cuyo objetivo principal sea la caracterización más que la predicción de series de tiempo.

Se pretende hacer una extensión al método de Ragulskis, usando igualmente consideraciones geométricas, pero buscando criterios de esparcimiento mas que de circularidad en las proyecciones planares, de tal manera que el criterio de selección del vector de retardo resulte más fuerte.

Se desean también estudiar otras medidas basadas en entropías que resulten igualmente discriminantes en labores de detección de patologías.

Parte V
Apéndices

A. Clasificación

A.1 Modelos de Mezcla de Gaussianas (Gaussian Mixture Models – GMM)

La idea central tras los GMM es la estimación de una función de densidad de un conjunto de datos, por medio de una superposición lineal de Gaussianas. Para un vector de características D -dimensional, \mathbf{f} , se tendrá una función de densidad como:

$$p(\mathbf{f}|\zeta) = \sum_{k=1}^K \pi_k p_k(\mathbf{f}) \quad (\text{A.1})$$

Donde K será el número total de Gaussianas; π_k coeficientes de mezcla que cumplen $0 \leq \pi_k \leq 1$ y $\sum_{k=1}^K \pi_k = 1$; $p_k(\cdot)$ funciones de densidad Gaussianas unimodales con vector de medias μ_k de $D \times 1$, y matriz de covarianza Σ_k de $D \times D$, como:

$$p_k(\mathbf{f}) = N(\mathbf{f}|\mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{f} - \mu_k)^T (\Sigma_k)^{-1} (\mathbf{f} - \mu_k) \right\} \quad (\text{A.2})$$

Y $\zeta = [\pi_k, \mu_k, \Sigma_k]$ los parámetros del modelo.

Introduciendo una variable latente binaria K -dimensional, \mathbf{z} , tal que cuando uno de sus elementos $z_k = 1$, todos los otros elementos serán 0; se podrá reescribir (A.1) como:

$$p(\mathbf{f}|\zeta) = \sum_{\mathbf{z}} p(\mathbf{z}) p(\mathbf{f}|\mathbf{z}) \quad (\text{A.3})$$

Ahora bien, para modelar un conjunto de N observaciones $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_N]$ usando GMM, se tendrá una función:

$$\log[p(\mathbf{F}|\zeta)] = \sum_{n=1}^N \log p(\mathbf{f}_n|\zeta) \quad (\text{A.4})$$

Cuyos parámetros ζ , deben ser maximizados, típicamente utilizando el algoritmo

de Maximización de la esperanza.

Para cada clase entonces se estiman diferentes GMM. Luego a cada observación \mathbf{f} que deba ser clasificada, se le calcula la probabilidad a posteriori:

$$p(\zeta|\mathbf{f}) = \frac{p(\mathbf{f}|\zeta)}{\sum p(\mathbf{f}|\zeta)p(\zeta)} \quad (\text{A.5})$$

La observación se asignara a la clase de mayor probabilidad a posteriori [68, 69].

A.2 Algoritmo de Maximización de la Esperanza (Maximization Expectation – EM)

Es usado para maximizar la función de logaritmo-verosimilitud con respecto a los parámetros de la ecuación (A.4). El algoritmo se explica a continuación [68, 70]:

Algoritmo 5 Algoritmo EM

1. Inicializar lo parámetros $\zeta = [\pi_i, \mu_i, \Sigma_i]$ y evaluar el valor inicial de la función de logaritmo-verosimilitud (A.4)
2. **Paso E.** Evaluar la probabilidad a posteriori:

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{f}_n | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{f}_n | \mu_j, \Sigma_j)} \quad (\text{A.6})$$

3. **Paso M.** Reestimar los parámetros con:

$$N_k = \sum_{n=1}^N \gamma(z_{nk})$$

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) \mathbf{f}_n \quad (\text{A.7})$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (\mathbf{f}_n - \mu_k^{new})(\mathbf{f}_n - \mu_k^{new})^T \quad (\text{A.8})$$

$$\pi_k^{new} = \frac{N_k}{N} \quad (\text{A.9})$$

4. Evaluar la logaritmo-verosimilitud (A.4) y buscar convergencia de los parámetros. Si un criterio de convergencia no es satisfecho, vuelva al paso 2.
-

B. Curvas ROC y DET

La manera tradicional de cuantificar el acierto diagnóstico de un test es mediante sensibilidad y especificidad, tal como se definió en (4.3). Estos parámetros describen la fracción de pacientes (enfermos o sanos) que son clasificados correctamente. La sensibilidad o tasa de verdaderos positivos describe la fracción de pacientes enfermos que realmente tienen un resultado de test positivo. La especificidad o tasa de falsos negativos describe la probabilidad de resultado de test negativo en individuos no enfermos.

La figura B.1 muestra las funciones de densidad por clase para pacientes enfermos y sanos, y la ubicación gráfica de los diferentes valores encontrados en la matriz de confusión.

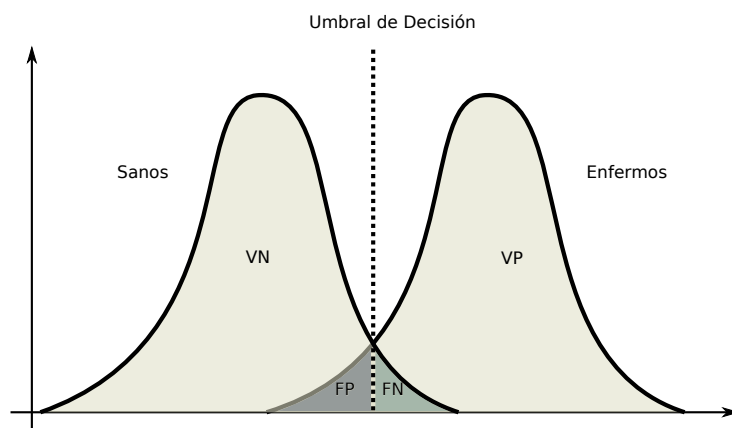


Figura B.1: Distribución de las diferentes cantidades derivadas de la matriz de confusión a partir de las densidades por clase

Mediante la modificación del umbral de decisión es posible afectar la sensibilidad y la especificidad obtenida, por lo que es deseable una prueba que compare resultados por medios que sean independientes del valor escogido de umbral. Para dichas tareas la *Receiver-Operation Curve* (ROC) fue propuesta [71].

La curva ROC es una gráfica de todos los pares sensibilidad-especificidad resultantes de variar continuamente el umbral de decisión sobre el rango entero de resultados

observados. En cada caso la curva ROC grafica el traslape entre las distribuciones al pintar la sensibilidad vs 1-especificidad [72].

Una ejemplo de Curva ROC típica se muestra en la figura B.2.

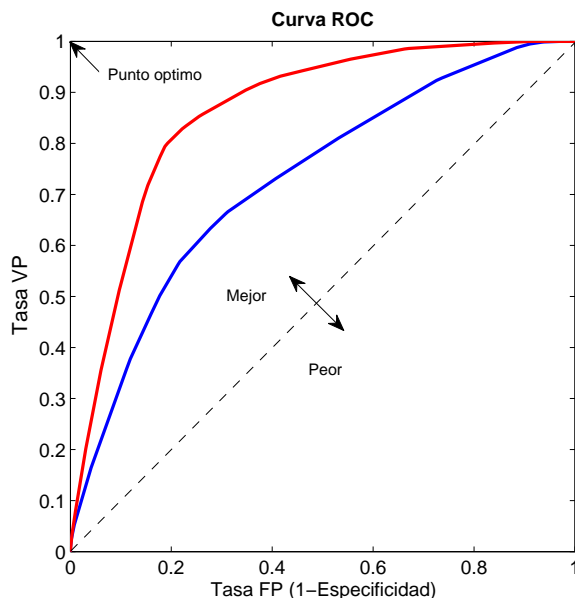


Figura B.2: Curva ROC típica

Los mejores resultados se lograran a medida que el la gráfica ROC alcance un codo más cercano a la esquina superior izquierda. En el caso mostrado en la figura B.2, la curva en rojo, muestra mejor desempeño que la curva en azul, lo que implica que las densidades de clase de la curva roja tienen menos traslape que el que se presenta para la curva azul.

Las *Detection Error Trade-off Curves* (DET), son otra forma de representación de la información obtenida. Las diferencias con las curvas ROC serán más del tipo estético, donde cada una de las gráficas enfatizará más un aspecto u otro del rendimiento del sistema [73].

A diferencia de las curvas ROC, las curvas DET grafican en el eje y la tasa de falsas aceptaciones (FN), además que utilizan en sus ejes escalas de desviaciones normales. En el caso de que las distribuciones sean cercanas a la normalidad, las gráficas usando escalas de desviaciones normales serán aproximadamente lineales [73, 74].

La gráfica B.3 muestra una curva DET típica, donde se ilustran dos curvas. La curva azul tendrá mejor rendimiento que la curva roja. Las curvas que se acerquen

más a la esquina inferior izquierda de la gráfica serán las que presentarán los mejores resultados.

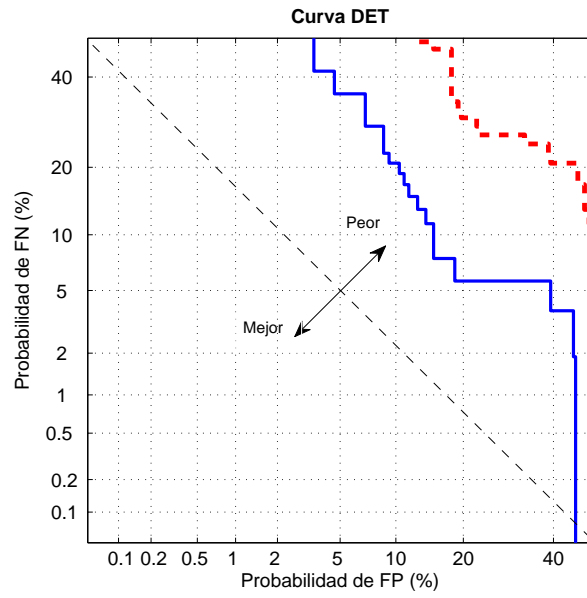


Figura B.3: Curva DET típica

C. Publicaciones

- Detección automática de voces patológicas empleando medidas de complejidad estocástica sobre embebimiento no uniforme (J. A. Gómez-García, J.D Arias-Londoño, J.I Godino-Llorente, C.G Castellanos-Domínguez), En XIV Simposio de tratamiento de Señales, Imágenes y Visión artificial, STSIVA2009, 2009.
- Análisis de la tolerancia al ruido de características basadas en dinámica no lineal sobre señales FCG. (C. Ospina-Aguirre, J. A. Gómez-García, C.G Castellanos-Domínguez) Revista TECNO LÓGICAS. *accepted*
- Evaluation of Feature Extraction Techniques on Evoked-related potentials for Detection of Attention-Deficit/Hyperactivity Disorder (P. Castro-Cabrera, J. Gomez-Garcia, F. Restrepo, O. Moscoso, G. Castellanos-Dominguez). 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society *submitted*
- Non-linear characterization of pathological sustained vowel speech signal by using non-uniform embedding. *to be submitted*

Bibliografía

- [1] H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis*. 2003. 1.1, 1.4, 2.1, 2.1, 2.1.1, 2.1.1, 3.2, 3.4
- [2] Mark Shelhamer. *Nonlinear Dynamics in Physiology: A State-space Approach*. World Scientific Publishing Company, December 2006. 1.4, 2.1, 2.1.1, 2.1.1, 2.1.1, 3.2
- [3] B. Henry, N. Lovell, and F. Camacho. Nonlinear biomedical signal processing : Dynamic analysis and modelling (ieee press series in biomedical engineering.). *ISBN: 9780780360129*, pp. 1-39, 2001. 2.1
- [4] Ary L. Goldberger, Luis A. N. Amaral, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, C.-K. Peng, and H. Eugene Stanley. Fractal dynamics in physiology: Alterations with disease and aging. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 1):2466–2472, 2002.
- [5] Velickov Slavco. *Nonlinear Dynamics and Chaos with Applications to Hydrodynamics and Hydrological Modelling*. Taylor & Francis, 2004.
- [6] Thomas Schreiber. Interdisciplinary application of nonlinear time series methods. *Phys. Rep*, 308:1–64, 1998. 1.4
- [7] J-J. Jiang, Y. Zhang, and C. McGilligan. Chaos in voice, from modeling to measurement. *Journal of Voice*, vol. 20, no. 1, pp. 2-17, 2006. 1.4, 4.2.1
- [8] Y. Zhang and J-J. Jiang. Nonlinear dynamic analysis in signals typing of pathological human voices. *Electronics Letters*, vol. 39, no. 13, pp. 1021-1023, 2003. 1.4
- [9] C.-K. Peng, S. Havlin, J.M. Hausdorff, J.E. Mietus, H.E. Stanley, and A.L. Goldberger. Fractal mechanisms and heart rate dynamics: Long-range correlations and their breakdown with disease. *Journal of Electrocardiology*, 28(Supplement 1):59 – 65, 1995. Research and Technology Transfer in Computerized Electrocardiology. 1.4
- [10] Akihiko Kikuchi, Toshiyuki Shimizu, Akiko Hayashi, Tsuguhiro Horikoshi, Nobuya Unno, Shiro Kozuma, and Yuji Taketani. Nonlinear analyses of heart rate variability in normal and growth-restricted fetuses. *Early Human Development*, 82(4):217 – 226, 2006. 1.4
- [11] N. Kannathal, U. Rajendra Acharya, C.M. Lim, and P.K. Sadasivan. Characterization of eeg—a comparative study. *Computer Methods and Programs in Biomedicine*, 80(1):17 – 23, 2005. 1.4

- [12] C. J. Stam. Nonlinear dynamical analysis of eeg and meg: Review of an emerging field. *Clinical Neurophysiology*, 116(10):2266–2301, October 2005.
- [13] Henri Korn and Philippe Faure. Is there chaos in the brain? ii. experimental evidence and related models. *Comptes Rendus Biologies*, 326(9):787 – 840, 2003. 1.4
- [14] C. Ahlstrom, A. Johansson, P. Hult, and P. Ask. Chaotic dynamics of respiratory sounds. *Chaos, Solitons & Fractals*, 29(5):1054 – 1062, 2006. 1.4
- [15] Floris Takens. Detecting strange attractors in turbulence. In *Nonlinear optimization*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381. Springer, 1981. 1.4
- [16] K. Judd and A. Mees. Embedding as a modelling problem. *Physica D 120 (1998) 273-286*, 1998. 1.4, 2.2, 2.2, 2.2, 2.2.1, 2.2.1, 2.2.1
- [17] Yusuke Manabe and Basabi Chakraborty. A novel approach for estimation of optimal embedding parameters of nonlinear time series by structural learning of neural network. *Neurocomputing*, 70(7-9):1360–1371, 2007. Advances in Computational Intelligence and Learning - 14th European Symposium on Artificial Neural Networks 2006, 14th European Symposium on Artificial Neural Networks 2006. 1.4
- [18] M. Ragulskis and K. Lukoseviciute. Non-uniform attractor embedding for time series forecasting by fuzzy inference systems. *Neurocomput.*, 72(10-12):2618–2626, 2009. 1.4, 2.2, 2.2.2
- [19] Antonia Jones. New tools in non-linear modelling and prediction. *Computational Management Science*, 1(2):109–149, 07 2004. 1.4
- [20] A. P. M. Tsui, A. J. Jones, and A. Guedes de Oliveira. The construction of smooth models using irregular embeddings determined by a gamma test analysis. *Neural Computing & Applications*, 10:318–329, 2002. 1.4
- [21] Julien Clinton Sprott. *Chaos and Time-Series Analysis*. Oxford University Press, 2003. 1.4, 3.3
- [22] L.S. Xu, K.Q. Wang, and L. Wang. Gaussian kernel approximate entropy algorithm for analyzing irregularity of time-series. *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21, 2005*. 1.4, 3.5.4
- [23] S. M. Pincus. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA Vol. 88, pp. 2297-2301*, 1991. 1.4, 3.5.1, 3.5.2
- [24] J-S. Richman and J-R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol 278: H2039-H2049*, 2000. 1.4, 3.5.2, 3.5.2, 3.5.3
- [25] L.S. Xu, K.Q. Wang, and L. Wang. Gaussian kernel approximate entropy algorithm for analyzing irregularity of time-series. *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21, 2005*. 1.4

- [26] W. Chen, Z. Wang, H. Xie, and W Yu. Characterization of surface emg signal based on fuzzy entropy. *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 15, No. 2, 2007. 1.4, 3.5.5
- [27] S. Pincus and Goldberg A. Physiological time-series analysis: what does regularity quantify? *The American Physiological Society*, 1994. 1.4
- [28] R. Hornero, M. Aboy, D. Abasolo, J. McNames, W. Wakeland, and B. Goldstein. Complex analysis of intracranial hypertension using approximate entropy. *Crit Care Med*, vol. 34, no. 1, pp. 87–95, 2006. 1.4, 3.5.2
- [29] R. Hornero, M. Aboy, D. Abásolo, J. McNames, and B. Goldstein. Interpretation of approximate entropy: analysis of intracranial pressure approximate entropy during acute intracranial hypertension. *IEEE Trans Biomed Eng*, vol. 52, no. 10, pp. 1671–1680, 2005. 1.4, 3.5.2
- [30] R. Hornero, D. Abásolo, N. Jimeno, C-I. Sánchez, J. Poza, and M. Aboy. Variability, regularity, and complexity of time series generated by schizophrenic patients and control subjects. *IEEE Trans Biomed Eng*, vol. 53, no. 2, pp. 210–218, 2006. 1.4, 3.5.2, 3.5.2, 3.5.3
- [31] M-M. Platisa and V. Gal. Dependence of heart rate variability on heart period in disease and aging. *Physiol Meas*, vol. 27, no. 10, pp.989-998. [Online]. Available: <http://dx.doi.org/10.1088/0967-3334/27/10/005>, 2006. 1.4, 3.5.2
- [32] Douglas Lake, Richman J-S, Griffin M-P, and Moorman. Sample entropy analysis of neonatal heart rate variability. *Am J Physiol Regul Integr Comp Physiol* 283, 2000. 1.4
- [33] C. Ahlstrom, K. Hoglund, P. Hult, J. Haggstrom, C. Kwart, and P. Ask. Assessing aortic stenosis using sample entropy of the phonocardiographic signal in dogs. *Biomedical Engineering, IEEE Transactions on*, 55(8):2107 –2109, aug. 2008. 1.4
- [34] Dandan Zhang, Haiyan Ding, Yunfeng Liu, Congle Zhou, Haishu Ding, and Datian Ye. Neurodevelopment in newborns: a sample entropy analysis of electroencephalogram. *Physiological Measurement*, 30(5):491, 2009. 1.4
- [35] Gregory L. Baker and J. P. Gollub. *Chaotic dynamics : an introduction*. Cambridge University Press, Cambridge [England] ; New York :, 1990. 2.1.1
- [36] H. Abarbanel. *Analysis of observed chaotic data*. Springer-Verlag, 1996. 2.1.1
- [37] M. Small. *Applied Nonlinear Time Series Analysis, Applications In Physics, physiology and finance*. 2002. 2.1.1, 2.2.1
- [38] K Judd and A. Mees. On selecting models for nonlinear time series. *Physica D* 82 426-444, 1995. 2.2.1, 2.2.1
- [39] T. Nakamura, K. Judd, and A. Mees. Refinements to model selection for nonlinear time series. *Internation Journal of Bifurcation and Chaos*. Vol 13, Number 5, 2003. 2.2.1

- [40] M. Small and C.K. Tse. Optimal embedding parameters: A modeling paradigm. 2.2.1
- [41] Itamar Procaccia. The static and dynamic invariants that characterize chaos and the relations between them in theory and experiments. *Physica Scripta*, 1985(T9):40, 1985. 3.1
- [42] Peter Grassberger and Itamar Procaccia. Characterization of strange attractors. *Phys. Rev. Lett.*, 50(5):346–349, Jan 1983. 3.2
- [43] Charles R. Tolle, Timothy R. McJunkin, David T. Rohrbaugh, and Randall A. LaViolette. Lacunarity definition for ramified data sets based on optimal cover. *Physica D: Nonlinear Phenomena*, 179(3-4):129 – 152, 2003. 3.2
- [44] A. Guerrero and L. A. Smith. Towards coherent estimation of correlation dimension. *Physics Letters A*, 318(4-5):373 – 379, 2003. 3.2
- [45] S.H Strogatz. *Nonlinear Dynamics and Chaos With Applications to Physics, Biology, Chemistry, and Engineering*. 1994. 3.4
- [46] J. Olivella, A. Corominas, and R. Pastor. An entropy-based measurement of working time flexibility. *European Journal of Operational Research*, 2008. 3.5
- [47] M. Costa, A-L. Goldberg, and C-K. Peng. Multiscale entropy analysis of biological signals. *Phys Rev E* 2005;71:021906, 2005. 3.5, 3.5.1, 3.5.1
- [48] T-M. Cover and J-A. Thomas. *Elements of information theory*. John Wiley and Sons, 2006. 3.5, 3.5
- [49] J. Crutchfield and K. Kaneko. *Directions in Chaos*. World Scientific, Singapore, 1987, Vol. 1, p. 272.pag 1872-1882, 1987. 3.5.1, 3.5.1
- [50] I. Cohen and I. Procaccia. Computing the kolmogorov entropy from time signals of dissipative and conservative dynamical systems. *Physical review A. Volume 31, number 3*, 1984. 3.5.1
- [51] P. Grassberger. Generalized dimensions of strange attractors. *Physical letters Volume 97A, number 6. September 5. 1983*, 1983. 3.5.1
- [52] J-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Reviews of modern physics, vol 57, 3*, 1985. 3.5.1
- [53] S. M. Pincus, I. Gladstone, and R. Ehrenkranz. A regularity statistic for medical data analysis. *J Clin Monit* 1991;7:335-345, 1991. 3.5.2
- [54] W. Chen, J. Zhuang, W Yu, and Z. Wang. Measuring complexity using fuzzyen, apen, and sampen. *Medical Engineering & Physics* 31 (2009) 61?68, 2009. 3.5.5, 3.5.5
- [55] *Massachusetts Eye and Ear Infirmary. Voice disorders database. version 1.03. [CD-ROM]. 1994. Lincoln Park, NJ: Kay Elemetrics Corp.* 4.2.1

- [56] V. Parsa and Jamieson D. Identification of pathological voices using glottal noise measures. *Journal of Speech, Language and Hearing Research*, vol. 43, no. 2, pp. 469-485, 2000. 4.2.1
- [57] J.D. Arias-Londoño, J.I. Godino-Llorente, and G. Castellanos-Domínguez. Short time analysis of pathological voices using complexity measures. *Proceedings of 3rd Advanced Voice Function Assessment International Workshop*, 2009. 4.2.1, 5
- [58] Edilson Delgado-Trejos, Andrés Felipe Quiceno-Manrique, Juan Ignacio Godino-Llorente, and César Germán Castellanos-Domínguez. Digital auscultation analysis for heart murmur detection. *Annals of Biomedical Engineering*, 37(2). 4.3.1, 5
- [59] A. F. Quiceno-Manrique, J. I. Godino-Llorente, M. Blanco-Velasco, and G. Castellanos-Domínguez. Selection of dynamic features based on time-frequency representations for heart murmur detection from phonocardiographic signals. *Annals of Biomedical Engineering*, Volume 38:118 – 137, 2009. 4.3.1
- [60] M. Akay, Y.M. Akay, D. Gauthier, R.G. Paden, W. Pavlicek, F.D. Fortuin, J.P. Sweeney, and R.W. Lee. Dynamics of diastolic sounds caused by partially occluded coronary arteries. *Biomedical Engineering, IEEE Transactions on*, 56(2):513 –517, feb. 2009. 4.3.1
- [61] James M. Parish and Virend K. Somers. Obstructive Sleep Apnea and Cardiovascular Disease. *Mayo Clinic Proceedings*, 79(8):1036–1046, 2004. 4.4.1
- [62] Wenlong Xu and Zhijie Pan. Definition of sleep apnea event by one minute hrv spectrum analysis. In *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*, pages 2292 –2294, 16-18 2008. 4.4.1
- [63] A. F. Quiceno-Manrique. Análisis tiempo-frecuencia por métodos no paramétricos orientado a la detección de patologías en bioseñales. Master’s thesis, Universidad Nacional de Colombia, Sede Manizales. Colombia, 2009. 4.4.1
- [64] Zhi-De Deng, Chi-Sang Poon, N.M. Arzeno, and E.S. Katz. Heart rate variability in pediatric obstructive sleep apnea. In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, pages 3565 –3568, aug. 2006. 4.4.1, 5
- [65] H. Nazeran, R. Krishnam, S. Chatlapalli, Y. Pamula, E. Haltiwanger, and S. Cabrera. Nonlinear dynamics analysis of heart rate variability signals to detect sleep disordered breathing in children. In *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, pages 3873 –3878, aug. 2006. 5
- [66] H. Dickhaus, C. Maier, and M. Bauch. Heart rate variability analysis for patients with obstructive sleep apnea. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 1, pages 507 – 510 vol.1, 2001. 5
- [67] James Theiler and Stephen Eubank. Don’t bleach chaotic data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 3(4):771–782, 1993. 5

- [68] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. A.1, A.2
- [69] Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn. Speaker verification using adapted gaussian mixture models. In *Digital Signal Processing*, page 2000, 2000. A.1
- [70] Jeff Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Technical Report TR-97-021, ICSI, 1997. A.2
- [71] Arian R. van Erkel and Peter M. Th. Pattynama. Receiver operating characteristic (roc) analysis: Basic principles and applications in radiology. *European Journal of Radiology*, 27(2):88 – 94, 1998. B
- [72] Charles E. Metz. Basic principles of roc analysis. *Seminars in Nuclear Medicine*, 8(4):283 – 298, 1978. B
- [73] *Biometric System and Data Analysis: Design, Evaluation, and Data Mining*. Springer, 2009. B
- [74] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki. The det curve in assessment of detection task performance. pages 1895–1898, 1997. B