

128-5



Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Ciudad de México

División de Ingeniería y Arquitectura

Ingeniería en Electrónica y Comunicaciones

Departamento de Ingeniería Eléctrica y Electrónica

Estimación de glucosa en agua mediante espectroscopia de infrarrojo cercano

Autores:	Arturo Trueba	711683
	José Luis Álvarez	953600
	Manuel Juárez	955082

Asesores:	Dr. Jorge Eduardo Brieva
	Dr. Pedro Ponce

México D.F. a 1 de Diciembre del 2004.

EGIA. PROYE

rn b10768081

QD95

T78

2004

42

1. La diabetes y el monitoreo de glucosa.

1.1 La diabetes.

La diabetes es un desorden del metabolismo, el proceso que convierte el alimento que ingerimos en energía. La insulina, que es una hormona segregada por el páncreas, es el factor más importante en este proceso ya que durante la digestión se descomponen los alimentos para crear la glucosa que pasa a la sangre, donde la insulina le permite entrar en las células.

En las personas que sufren diabetes, una de dos componentes del metabolismo falla y dependiendo de la parte que falla en este proceso, podemos clasificar a la diabetes en dos tipos:

El tipo I o insulino-dependientes. Es llamado también diabetes juvenil porque normalmente comienza en la infancia aunque también puede ocurrir en el adulto. Ésta se caracteriza porque el páncreas no produce insulina, o produce muy poca. Debido a que el cuerpo no es capaz de producir insulina, las células no procesan la glucosa en la sangre, con lo cual, el paciente debe inyectarse insulina para poder vivir.

El tipo II, surge en adultos y en ésta, el páncreas sí produce insulina pero las células no pueden aprovechar la que se produce.

En ambos casos, este comportamiento provoca un aumento en la permanencia de la glucosa en la sangre por arriba de los niveles normales; lo que altera en su conjunto el metabolismo de los hidratos de carbono, los lípidos y las proteínas.

Dado que no existe cura para la diabetes, la mejor forma para que las personas afectadas por esta enfermedad cuiden su salud es controlar dicho desorden manteniendo los niveles de glucosa lo más cercano posible a los niveles normales (76-110mg/dL). Un buen control de los niveles de azúcar es posible mediante una dieta planificada, actividad física, toma correcta de medicamentos y chequeos frecuentes del nivel de azúcar en la sangre.

La revisión de los niveles de azúcar en la sangre se puede realizar mediante glucómetros portátiles o bien en laboratorios de análisis médico. Los métodos conceptuales para medir los niveles de azúcar en la sangre se explican brevemente a continuación.

1.2 Métodos de medición de glucosa.

A los distintos métodos de medición de glucosa existentes los podemos clasificar en dos grupos, los métodos de medición invasivos (como la punción o los implantes) y métodos de medición no invasivos, siendo estos últimos los de nuestro principal interés.

1.2.1 Métodos de medición no invasivos

1.2.1.1 Tecnología transdérmica

La tecnología transdérmica consiste en un sistema análogo a los parches que distribuyen nicotina, a diferencia de que la glucosa se mueva hacia fuera y no hacia dentro del cuerpo. Normalmente la piel es una excelente barrera por lo que los métodos transdérmicos requieren de algún tipo de transportador los cuales incluyen el uso de corriente eléctrica (iontoforesis), energía acústica (sonoforesis) y químicos que cambian la permeabilidad de la glucosa a través de la piel (difusión pasiva) o la permeabilidad mecánica (microporación). Estos métodos pueden causar cierta irritación en la piel de los pacientes.

Incluso con la ayuda de los transportadores, la muestra de fluido obtenida puede ser varios órdenes de magnitud menor que la concentración en la sangre. También será muy pequeña en volumen del orden de microlitros. Ambos factores hacen necesaria la utilización de sensores innovadores y sistemas novedosos para manejar fluidos.

La compañía Cygnus desarrolló un dispositivo que extrae glucosa a través de la piel por iontoforesis y mide la muestra extraída electroquímicamente usando la reacción de la glucosa oxidasa. Está pensado para utilizarse junto con sistemas de medición tradicional para asistir en el manejo de la diabetes. Se usa como un reloj y el paciente puede conocer sus mediciones de glucosa durante 12 horas. El sensor se encuentra en la parte inferior del dispositivo y se adhiere a la piel. Cada sensor contiene electrodos para aplicar la corriente de extracción, dos discos de hidrogel para recolectar la glucosa y electrodos para sensar la glucosa.

Cada 20 minutos se pasa por un ciclo de extracción y sensado. Primero se aplica una corriente de 0.3mA a la piel. Los iones que hacen la corriente, mueven junto con ellos la glucosa hacia los discos de hidrogel. Durante el periodo de sensado que sigue al de extracción, la glucosa pasa por la reacción enzimática electroquímica ya mencionada.

El sudor que puede interferir con las mediciones debido a que también contiene glucosa y se detecta midiendo la conductividad de la piel a través de detectores de sudoración colocados en el dispositivo. Un termistor, también integrado, mide los cambios en la temperatura que pudieran interferir con la señal del sensor. Es necesario un algoritmo que considere estos factores para convertir la medición obtenida en una concentración de glucosa útil para el paciente. Alrededor del 25% de las mediciones son erróneas.

Otras técnicas incluyen el uso de frecuencias bajas ultrasónicas para ayudar al transporte de la glucosa a través de la piel. El ultrasonido provoca la formación y el colapso de bolsas de gases disueltos en un gel aplicado en la piel. Esta cavitación crea chorros microscópicos de fluido que llegan a la parte externa de la piel, y sirven como caminos para que el fluido intersticial que contiene glucosa salga.

Por otro lado se tiene la difusión pasiva la cual consiste en extraer durante 5 minutos glucosa a un parche transdérmico como el de la nicotina. El parche también incluye agentes que cambian de color con la glucosa. Un medidor óptico lee la magnitud del cambio en el color.

Por último se tiene el acceso a través de microporos. Esto se logra mediante arreglos de lasers a la medida que crean estos micro poros en la piel, cada hoyo es del tamaño de un cabello humano. Una vez que se ha perforado la piel, por así decirlo, el usuario utiliza un sistema con el cual mediante succión al vacío se obtiene la muestra de líquido intersticial la cual después se pasa por un sensor de glucosa electroquímico.

1.2.1.2 Métodos ópticos

La medición óptica del nivel de glucosa se realiza enfocando un rayo de luz dentro del cuerpo, la luz es modificada por el tejido después de su transmisión a través del área objetivo. Una firma óptica, o una huella digital del contenido del tejido, son producidas por la luz difusa que escapa del tejido que ha penetrado. La absorción de la luz por la piel se debe a sus componentes químicos (agua, hemoglobina, melanina, grasa y glucosa). La transmisión de la luz en cada longitud de onda es una función del grosor, color y estructura de la piel, hueso, sangre y cualquier otro material por el cual la luz pase. [11]

La concentración de glucosa se puede determinar analizando los cambios en la señal óptica en cuanto a la longitud de onda, la polarización o la intensidad de la luz. El volumen muestreado por estos métodos depende de la zona de medición. La correlación con la glucosa en sangre se basa en el porcentaje de la muestra de fluido que es sangre intersticial, intracelular o capilar. Los doctores Roe y Smoller han descrito el siguiente ejemplo. El fluido visto a través de la extremidad es 63% intracelular y 37% extracelular, del cual 27% es intersticial y 10% es plasma. Un valor de glucosa en sangre de 100mg/dL es equivalente a un promedio de glucosa en la muestra de tejido de 38mg/dL promedio del cual 26% se debe a la sangre, 58% al fluido intersticial y 16% al fluido intracelular. Lo que la medida de glucosa en la muestra de tejido significa clínicamente con respecto a la terapia todavía está en investigación.

El método óptico no sólo depende de los cambios en la concentración en todos los componentes del cuerpo medidos, además cambia la razón de estos fluidos en el tejido (por el nivel de actividad, la dieta o fluctuaciones de hormonas) y esto a su vez, afecta la medida de glucosa. También ocurren problemas debido a cambios en el tejido después de la calibración original y a la falta de transferencia de calibración de una parte del cuerpo a otra. Los cambios en el tejido incluyen: cambio de la fuente de sangre para el fluido corporal medido, medicamentos que afectan la razón de los fluidos en el tejido, cambios diarios en los vasos sanguíneos, el proceso de envejecimiento, enfermedades y la actividad metabólica de la persona.

Espectroscopia Raman

La espectroscopia Raman mide la luz dispersa que se ha influenciado por la oscilación y rotación del dispersor. Se han intentado varias técnicas Raman en la sangre, agua, suero y soluciones de plasma así como en el ojo, pero todavía existen varios problemas antes de que se puedan hacer estudios en humanos. Los problemas analíticos incluyen la inestabilidad en la longitud de onda e intensidad del láser, errores debido a otros químicos en la muestra de tejido y largos tiempos de adquisición espectral.

Espectroscopia foto-acústica

La espectroscopia fotoacústica utiliza un rayo óptico para calentar rápidamente la muestra de tejido y generar una onda de presión acústica que se puede medir por un micrófono. En cuanto a la determinación de glucosa en sangre, este método puede proveer más estabilidad que la espectroscopia convencional cuando algunos parámetros físicos son favorables. Se ha mostrado una excelente correlación entre la señal fotoacústica y el nivel de glucosa en sangre en dedos índices de pacientes sanos y diabéticos. Actualmente la instrumentación se hace a la medida, es costosa y sensible a los parámetros ambientales. La técnica también está sujeta a interferencia química de moléculas biológicas y a interferencia física de cambios en la presión y la temperatura.

Cambios en la dispersión

La medición de la dispersión consiste en revisar constantemente los cambios en el coeficiente de dispersión del tejido y se utiliza para la determinación de la glucosa en tejidos y humanos.

El aumento de la glucosa en el tejido es proporcional al aumento en el índice refractivo del mismo. Mediciones en el abdomen de pacientes diabéticos muestran una correlación excelente entre la señal dispersa y los niveles de glucosa en sangre. Varios parámetros contribuyen a un corrimiento fisiológico natural del parámetro de dispersión. Se necesita de métodos que compensen el corrimiento de la señal.

Cambios en la polarización

Aunque el cambio en la señal óptica por la glucosa es pequeño, la glucosa es un buen rotor óptico. Esta característica se ha utilizado para llevar a cabo pruebas in-vitro. La intensidad de la luz corresponde a la cantidad de luz presente. La piel no es un lugar factible para las mediciones debido a sus altas propiedades de dispersión de la luz. El humor acuoso del ojo se ha sugerido y se han publicado mediciones en solución. El uso de varias longitudes de onda minimiza la pobre especificidad de esta técnica. Otras sustancias óptimamente activas interfieren con el análisis así como variaciones en la temperatura y en el pH del tejido muestreado.

Espectroscopia infrarroja

Los sistemas de medición de glucosa infrarrojos en la superficie epidérmica permiten la investigación de profundidades titulares en el rango de los 10 a los 50 micrómetros usando una banda de longitud de onda de la región infrarroja entre los 700 y los 1000 nm. Estos sistemas no miden la glucosa en los tejidos que contienen sangre. Una técnica de reflexión total atenuada se ha usado para la mucosa oral, sin embargo, las desventajas incluyen la contaminación de la glucosa del sitio de medición por comida y las variaciones altas en la saliva. Si se utiliza la sangre completa como muestra existe interferencia de la albúmina, células rojas, la gama globulina y los cambios en la temperatura y el pH. Además, la glucosa en la saliva cambia considerablemente y no refleja los niveles de glucosa en sangre.

Espectroscopia del infrarrojo cercano

La glucosa produce una de las señales de absorción de NIR (Near Infrared – infrarrojo cercano) más débiles por unidad de concentración de los componentes principales del cuerpo. La medición de glucosa por espectroscopia NIR permite la exploración del tejido en profundidades del rango de 1 a 100 milímetros con una disminución general en la profundidad de penetración conforme se aumenta el valor de la longitud de onda. La transmisión NIR a través del lóbulo auricular, la membrana de los dedos, la cutícula del dedo o reflejada de la piel del antebrazo y la mucosa del labio se ha intentado en la región NIR entre 1000nm y 2500nm. Mediciones de reflexión difusa NIR se han hecho en el dedo y la cutícula y han mostrado una buena correlación con el nivel de glucosa en sangre, pero el 10% de las predicciones no son clínicamente aceptables.

Estudios de reflexión difusa en el labio interno también han mostrado una buena correlación con el nivel de glucosa en sangre y han indicado un retraso en el tiempo de 10 minutos entre la glucosa en sangre y la señal medida. Los niveles de glucosa en la saliva (un componente de la medición en el labio) no reflejaron los niveles de glucosa en sangre. Parámetros físicos y químicos como la variación en la presión, la temperatura, los triglicéridos y la albúmina interfieren con la medición de glucosa. También pueden ocurrir errores debido a variaciones del medio ambiente como cambios en la temperatura, la humedad, la hidratación de la piel, el dióxido de carbono y la presión atmosférica. Se requiere una validación extensiva de los datos y pruebas de la ecuación de predicción de glucosa para determinar si la correlación de glucosa es consistente en todas las condiciones clínicas importantes en todo tipo de paciente.

1.3 Definición del problema y propuesta de solución.

1.3.1 Problemática

La diabetes es una enfermedad que padecen un gran número de personas en todo el mundo afectando su calidad de vida en mayor o menor medida según el correcto tratamiento de esta. En dicha enfermedad, el páncreas no produce la suficiente insulina, o bien, las células del organismo no son capaces de utilizarla en forma adecuada lo que conlleva a que la glucosa (azúcar) que consumimos y producimos diariamente no se transforme en la fuente principal de energía para el cuerpo y como consecuencia la elevación de los niveles de azúcares en la sangre y el mal aprovechamiento de los alimentos.

Los tratamientos actuales para la enfermedad consisten en la aplicación de dosis de insulina mediante inyecciones y bombas de infusión para los casos de diabetes tipo I y el consumo de pastillas para el tipo II. En ambos casos es necesaria la adquisición de los niveles de glucosa en la sangre.

A pesar de todos los avances en el tratamiento de la diabetes, la educación del paciente sobre su propia enfermedad sigue siendo la herramienta fundamental para el control de la misma. La gente que sufre de diabetes, a diferencia de aquellos con muchos otros problemas médicos, no puede simplemente tomarse unas pastillas o insulina por la mañana, y olvidarse de su condición el resto del día. Cualquier diferencia en la dieta, el ejercicio, el nivel de estrés, u otros factores pueden afectar el nivel de azúcar en la sangre, por lo tanto, cuanto mejor conozcan los pacientes los efectos de estos factores, mejor será el control que puedan ganar sobre su condición.

Actualmente el monitoreo de glucosa se realiza con mayor efectividad mediante la pinchadura de un dedo con una lanceta y la posterior colocación de la sangre en una tira reactiva para su medición en un glucómetro.

1.3.2 Propuesta de solución

Sobre todo para el caso de la diabetes del tipo insulino-dependiente, una adecuada supervisión de los niveles de glucosa implica una serie de monitoreos con mayor regularidad o frecuencia. De presentarse o realizarse con los métodos convencionales de tipo invasivo, resulta ser una verdadera molestia. El caso ideal sería el de obtener dicha información con cierta regularidad y que además no requiera de ningún agente o dispositivo tanto de punción como transdérmico.

Es por ello que se desea implementar algoritmos para la detección de glucosa en soluciones acuosas, para su utilización como base o referencia en el posterior diseño de un glucómetro no invasivo. Este debe facilitar la adquisición de datos para el paciente a la vez que se reducen las molestias por la medición y tener así una obtención de datos en menores periodos de tiempo para un mejor control de los tratamientos a aplicar.

El propósito del desarrollo de estas técnicas es el de ayudar a mejorar la calidad de vida de las personas que padecen esta enfermedad.

1.3.3 Método a utilizar

En base a la experiencia del trabajo anteriormente realizado con espectroscopia del infrarrojo cercano y el uso de regresión PLS para la detección, se decidió continuar utilizando la espectroscopia como método de adquisición de datos pero cambiando los métodos de decisión por métodos basados en sistemas inteligentes como redes neuronales y lógica difusa [1].

La pequeña región de longitud de onda del infrarrojo cercano (1100 a 2500 nm) tiene algunas propiedades que son ideales para diagnósticos in vivo no invasivos. En este rango existe una ventana óptica entre la región en la cual la luz visible se absorbe por la sangre y los pigmentos de la piel y las longitudes de onda mayores para las cuales la absorción del agua predomina.

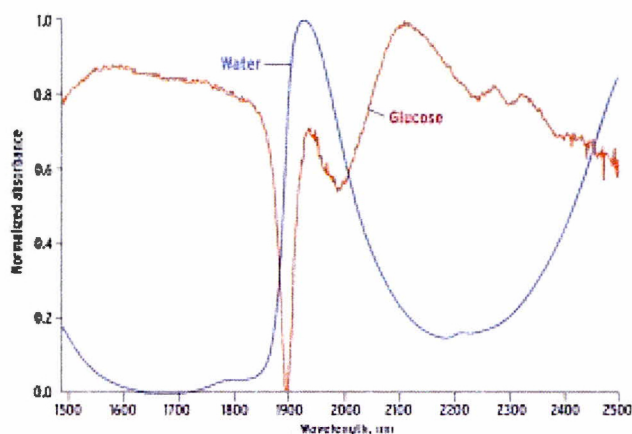


Figura 1.1 Absorción del agua y la glucosa para diferentes longitudes de onda [2].

El espectro NIR consiste de sobretonos y combinaciones de las vibraciones fundamentales principalmente de los enlaces de carbón, nitrógeno y oxígeno con hidrógeno. El espectro NIR de un sistema acuoso muestra bandas débiles, anchas y que se traslapan. La posición y la intensidad de las señales varían de acuerdo a la vecindad química (efectos de los lazos con hidrógeno). La influencia de sales disueltas y la temperatura en el espectro NIR de un sistema acuoso es conocida. Ya que la proporción normal de glucosa en sangre y en tejido es solo el 0.1% del contenido de agua, las variaciones espectrales debido a la concentración de glucosa son extremadamente pequeñas. La evaluación del espectro medido es complicada por varias influencias: el agua es un fuerte absorbente y también el mayor componente del tejido vivo. Además ocurren procesos biológicos que dependen del tiempo como el pulso y la respiración. Otras fuentes de variación son las mediciones espectrales erróneas producidas por una presión irregular de la cabeza medidora en el dedo. Como consecuencia, es necesario utilizar métodos de calibración multivariable para evaluar el espectro.

Anteriormente se evaluaron los obstáculos que presentaba la espectroscopia como método de medición no invasivo de glucosa y se mencionaron maneras de reducir cada uno de estos problemas con el fin de demostrar la factibilidad del método de medición propuesto, NIR. Ahora que se ha comprobado que es posible llegar a resultados para soluciones dadas, se intenta disminuir al máximo los problemas existentes para la detección de glucosa.

Los principales obstáculos mencionados son los siguientes [2]:

1. Variaciones en la línea de base de los datos espectroscópicos debido a la instrumentación o condiciones ambientales
2. Alta absorción de fondo intrínseca debido al agua.
3. Ruido de alta frecuencia debido al detector.
4. Gran número de traslapes de los espectros de absorción debido a otros elementos.
5. Las propiedades ópticas de la piel son complejas.
6. Degradación de la señal de interés (interferencia óptica de otros componentes).

Es entonces la solución de los primeros tres problemas presentes, un primer paso para la realización de un glucómetro no invasivo, por lo que se realiza en un inicio únicamente la detección de glucosa en agua.

Para dicha detección se seleccionó el método de espectroscopía del infrarrojo cercano, la cual nos ofrece ciertas ventajas y desventajas, con respecto a la espectroscopía infrarroja, para lograr los objetivos que se desean cumplir.

1.3.3.1 Ventajas método NIR

- Es menos afectado por la interferencia de agua que MIR lo que nos permite trabajar sin un prisma de reflectancia total atenuada (ATR).
- Longitud de onda menor permite mayor penetración; lo cual es necesario para medir glucosa en sangre en los capilares y glucosa en los fluidos intersticiales.

Además, los prismas ATR tienen dos inconvenientes:

- Se fabrican normalmente de seleniuro de zinc (ZnSe) el cual no debe estar en contacto con la piel humana pues el seleniuro es venenoso.
- Necesitan alineación de su precisión periódicamente.

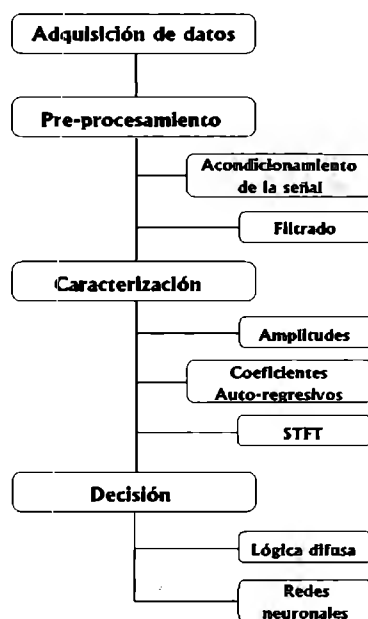
Como se puede ver el problema del detector se resuelve con la simple elección del método infrarrojo cercano sobre el infrarrojo.

1.3.3.2 Desventajas método NIR

- Espectro más plano que el MIR
- Bandas combinacionales y sobretonos en la región NIR para varias sustancias biológicas.

Las dos desventajas que se mencionan del método NIR son las que se buscan resolver mediante la implementación de algoritmos de obtención de la concentración de glucosa a partir de los datos espectrales.

1.3.4 Sistema propuesto



2. Metodología

El primer paso del procedimiento consiste en la captura de la información necesaria mediante un espectroscopio. Esta captura, como sabemos de antemano, contiene variación de línea de base y mucho ruido debido a la instrumentación y a las condiciones ambientales, es por eso necesaria una etapa de pre-procesamiento que quite estas perturbaciones. Debido a que desconocemos el tipo y la forma de estas perturbaciones implementamos diversos tipos de filtrado: FIR, IIR y wavelets. Las señales obtenidas del filtrado son del orden de 600 a 1000 elementos, lo cual dificulta su análisis pues estas son muy grandes para ser procesadas por un equipo de cómputo. Para superar este obstáculo fue necesario extraer de las señales la parte más importante; la que contiene la información relevante, y para esto empleamos dos métodos de caracterización: coeficientes autorregresivos y STFT (Transformada del Fourier del tiempo corto, por sus siglas en inglés). Con los coeficientes autorregresivos podemos obtener la información que caracterice a la señal con muchos menos puntos, mientras que con STFT podemos observar gráficamente la parte donde se concentra la mayor cantidad de energía y así tomar esa parte de la señal. Una vez con la señal limpia tenemos que identificar la concentración de glucosa en ésta, para eso utilizamos diversos métodos: redes neuronales, lógica difusa y ANFIS, estos son métodos de decisión inteligentes con capacidad de adaptación, en este caso adaptación al comportamiento de la glucosa.

2.1 Etapa de Pre-Procesamiento

2.1.1 Diseño de filtros digitales

En el proceso de diseño de un filtro digital se determinan los coeficientes de un filtro FIR o IIR causal que aproxima de forma precisa las especificaciones de respuesta en frecuencia. La elección de uno de los dos tipos de filtro, FIR o IIR, dependerá de la naturaleza del problema y de las especificaciones de la respuesta en frecuencia deseada. En la práctica, los filtros FIR se emplean en problemas de filtrado donde hay un requisito de fase lineal dentro de la banda de paso del filtro. Si no existe este requisito es posible escoger uno de tipo IIR o FIR. Como regla general los filtros IIR tienen lóbulos laterales menores en la banda de rechazo que un filtro FIR con el mismo tipo de diseño. Por ésta razón, si se puede tolerar alguna distorsión en la fase debido a su no-linealidad, o simplemente no es importante, se prefiere un filtro IIR. Esto sucede debido a que su

implementación involucra menos parámetros, requiere menos memoria y tiene menor complejidad computacional.

Diseñar un filtro es encontrar su función de transferencia (realizable y estable). El siguiente paso es su realización mediante una estructura adecuada. En la mayoría de las aplicaciones el objetivo es desarrollar una aproximación a una especificación de la respuesta en magnitud.

Las especificaciones en las bandas de paso y de rechazo se dan con ciertas tolerancias. La banda de transición permite que la magnitud disminuya del valor de la banda de paso al valor de la banda de rechazo. Por ejemplo, la magnitud $G(e^{j\omega})$ de un filtro pasa bajas se muestra en la figura 2.1

En la banda de paso, definida por $0 \leq \omega \leq \omega_p$, se requiere que la magnitud se aproxime a la unidad con un error de $\pm \delta_p$.

$$1 - \delta_p \leq G(e^{j\omega}) \leq 1 + \delta_p \quad \omega \leq \omega_p \quad (2.1)$$

En la banda de rechazo, definida por $\omega_s \leq \omega \leq \pi$, se requiere que la magnitud se aproxime a cero con un error máximo de δ_s .

$$G(e^{j\omega}) \leq \delta_s, \quad \omega_s \leq \omega \leq \pi \quad (2.2)$$

Donde:

ω_p = frecuencia de banda de paso.

ω_s = frecuencia de banda de rechazo.

δ_p, δ_s son los valores pico de rizo.

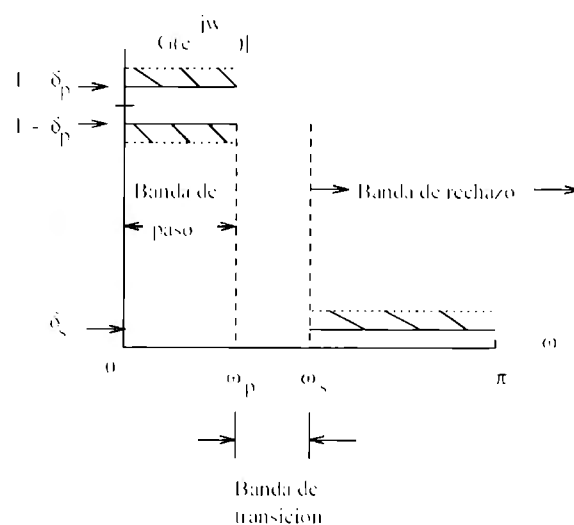


Figura 2.1 Especificaciones de un filtro pasa bajas [6]

Puesto que $G(e^{j\omega})$ es periódica y la $|G(e^{j\omega})|$ de un filtro de coeficientes reales es una función par, las especificaciones se requieren solo para $0 \leq \omega \leq \pi$.

Con una frecuencia las especificaciones se dan en función de la función de pérdida o atenuación $A = -20 \log_{10}(|G(e^{j\omega})|)$ en dB. El rizo pico de la banda de paso α_p , y la atenuación mínima de la banda de rechazo α_s , se dan en dB. Las especificaciones de un filtro digital están dadas por:

$$\alpha_p = -20 \log_{10}(1 - \delta_p) \text{ dB}$$

$$\alpha_s = -20 \log_{10}(\delta_s) \text{ dB}$$

Alternativamente, las especificaciones de un filtro digital de pueden dar en función de la respuesta en magnitud, como se muestra en la figura 2.2

En éste caso, se supone que el valor máximo es igual a 1; y la máxima desviación en la banda de paso $\frac{1}{\sqrt{1+\epsilon^2}}$ está dada por el valor mínimo de la magnitud de la banda de paso. El valor máximo de la magnitud en la banda de rechazo se denota por $1/A$.

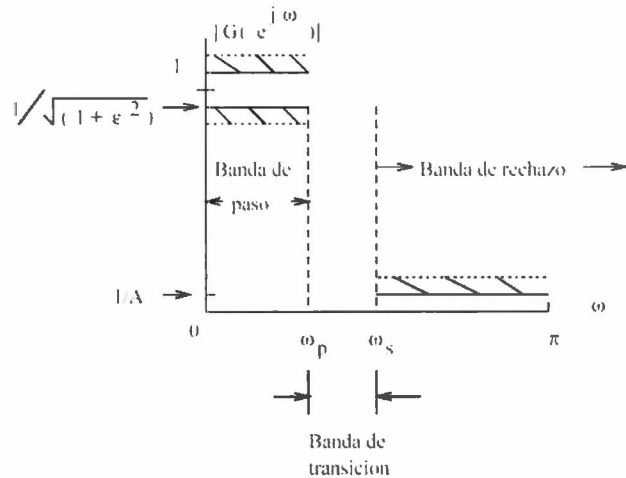


Figura 2.2 Especificaciones alternas de un filtro pasa bajas [6]

Para las especificación normalizada, el valor máximo de la ganancia es de 0 dB; y la cantidad α_{\max} esta dada por :

$$\alpha_{\max} = -20 \log_{10} \left(\frac{1}{\sqrt{1+\epsilon^2}} \right) \text{ dB} \quad (2.3)$$

$$\alpha_{\max} = 20 \log_{10} (1 + \epsilon^2) \quad \text{dB} \quad (2.4)$$

En la mayoría de las aplicaciones las frecuencias de paso y de rechazo se especifican en Hz junto con la frecuencia de muestreo del filtro digital. Puesto que las técnicas de diseño de filtros usan frecuencias angulares de tiempo discreto ω_p y ω_s , es necesario normalizar las frecuencias críticas antes de aplicar los algoritmos de diseño. Para una frecuencia de muestreo F_T en Hz y para frecuencias de paso y de rechazo f_p y f_s en Hz, respectivamente, las frecuencias angulares normalizadas en radianes están dadas por

$$\omega_p = \frac{\Omega_p}{F_T} = \frac{2\pi f_p}{F_T} = 2\pi f_p T \quad (2.5)$$

$$\omega_s = \frac{\Omega_s}{F_T} = \frac{2\pi f_s}{F_T} = 2\pi f_s T \quad (2.6)$$

2.1.1.1 Filtro IIR (respuesta infinita al impulso)

Los filtros de Respuesta Infinita al Impulso, o filtros IIR, de la misma manera que los filtros FIR, son sistemas LTI (lineales e invariables en el tiempo) que recrean un amplio rango de diferentes respuestas en frecuencia.

Consideremos un sistema discreto caracterizado por la ecuación de diferencias:

$$y(n) = -\sum_{k=1}^N a_k y(n-k) + \sum_{k=0}^M b_k x(n-k) \quad (2.7)$$

Mediante la transformada Z se puede obtener la función de transferencia caracterizada por (2.7):

$$H(z) = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=1}^N a_k z^{-k}} \quad (2.8)$$

De esta caracterización se obtienen polos y ceros, los cuales dependen de la elección de los parámetros del sistema $\{b_k\}$ y $\{a_k\}$, y determinan las características de la respuesta en frecuencia del sistema.

Diseño de filtros IIR

Para filtros IIR, el método más usado se basa en filtros analógicos. Las especificaciones del filtro digital se convierten a especificaciones de un filtro analógico prototipo, se determina la función de transferencia del filtro analógico que satisface las especificaciones y finalmente se transforma a la función de transferencia del filtro digital.

Este método se basa en el mapeo entre los dominios s y z , de tal manera que se mantengan las propiedades fundamentales de la respuesta en frecuencia. La función de mapeo debe ser tal que:

1. El eje imaginario ($j\Omega$) en el plano s se mapea al círculo unitario del plano z
2. Una función de transferencia analógica estable, se transforma en una función de transferencia digital estable.

Las transformaciones más comunes son la invarianza al impulso y la transformación bilineal.

El orden del filtro depende del orden del filtro digital y se obtiene a partir de la formula 2.9

$$N \geq \frac{\log_{10} \left(\frac{10^{0.1As-1}}{10^{0.1Ap-1}} \right)}{2 \log_{10} \left(\frac{\omega_s}{\omega_p} \right)} \quad (2.9)$$

Por otra parte existen funciones en Matlab para determinar el orden mínimo de la función de transferencia de un filtro IIR digital que satisface las especificaciones. Dichas funciones son:

$$[N, Wn]=buttord(Wp,Ws, Rp, Rs) \quad (2.10)$$

$$[N, Wn]=cheb1ord(Wp,Ws, Rp, Rs) \quad (2.11)$$

$$[N, Wn]=cheb2ord(Wp,Ws, Rp, Rs) \quad (2.12)$$

$$[N, Wn]=ellipord(Wp,Ws, Rp, Rs) \quad (2.13)$$

2.1.1.2 Filtro FIR (respuesta finita al impulso)

Los filtros FIR tienen valores de respuesta al impulso diferentes de cero, solamente durante un periodo finito de tiempo. Este tipo de filtros tienen la ventaja de poseer respuestas de fase lineales, estabilidad inherente, menos susceptibilidad a la cuantificación de los coeficientes y una rápida implementación.

La ecuación de diferencias de un filtro FIR esta dada por:

$$y(n) = h(0)x(n) + h(1)x(n-1) + \dots + h(N)x(n-N) = \sum_{m=0}^N h(m)x(n-m) \quad (2.14)$$

La respuesta en frecuencia se obtiene de la siguiente manera:

$$H(e^{j\omega}) = \sum_{n=0}^N h_d(n)e^{-j\omega n} \quad (2.15)$$

Diseño de filtros FIR

El diseño de filtros FIR se basa en una aproximación directa de la respuesta en magnitud especificada, frecuentemente con el requisito de que la respuesta sea lineal. Entre los

métodos de diseño se cuentan el de ventanas y el de muestreo en frecuencia. Otros métodos se basan en técnicas iterativas de optimización para minimizar el error entre la respuesta en frecuencia deseada y la del filtro generado por computadora.

El método más sencillo de diseño de filtros FIR es el método de ventanas. Este método comienza con una respuesta en frecuencia ideal deseada que puede representarse por:

$$H_d(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h_d[n] e^{-j\omega n} \quad (2.16)$$

La respuesta en frecuencia del filtro discreto paso bajas ideal es una forma de onda periódica rectangular en el dominio de la frecuencia. La respuesta al impulso correspondiente con una frecuencia de corte ω_c esta dada por la transformada inversa de Fourier en tiempo discreto como:

$$h_d[n] = \frac{\text{sen}(\omega_c n)}{\pi n}, -\infty < n < \infty. \quad (2.17)$$

Como la respuesta al impulso es de duración infinita, debemos de truncar una parte de ésta para hacerla finita y dado que la respuesta al impulso de un sistema físico realizable, es cero para valores menores que cero, desplazamos la respuesta al impulso truncada. Por lo tanto la respuesta al impulso truncada de un filtro realizable esta dada por:

$$h[n] = \frac{\text{sen}\left(\omega_c \left(\frac{N}{2} - n\right)\right)}{\pi \left(\frac{N}{2} - n\right)}, n = 0, 1, \dots, n, \quad (2.18)$$

donde N es el orden del filtro.

Como un filtro ideal es prácticamente irrealizable, el diseño de un filtro real es la aproximación de la respuesta en frecuencia ideal dada. Recordando que un filtro ideal posee una atenuación infinita en la banda de rechazo y la banda de transición con ancho cero, para diseñar un filtro práctico es necesario multiplicar la respuesta al impulso ideal $h(n)$ truncada por una ventana $w(n)$ seleccionada y usar el resultado $h(n)w(n)$ como la respuesta al impulso.

Dichas ventanas son utilizadas para modificar la $h(n)$ truncada de tal forma que se logre reducir el tamaño de las ondulaciones en frecuencia de la respuesta del filtro. La respuesta en frecuencia de un filtro depende de las características de la ventana usada, mientras más suave sea la ventana más se ensancha la banda de transición y la atenuación se incrementa.

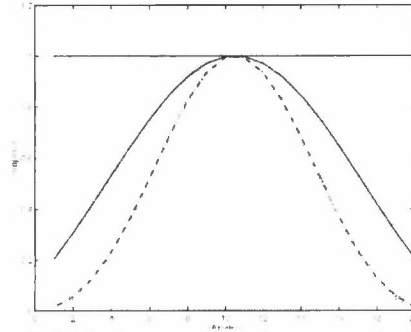


Figura 2.3 Ventanas de Kaiser para $b=0,3$ y 6 [6]

Dadas las frecuencias de frontera de la banda de paso y la banda de rechazo, ω_p y ω_s , y la correspondientes magnitudes de atenuación, A_c y A_r , para un filtro paso bajas, se puede determinar el orden del filtro y el parámetro b como se muestra a continuación.

$$\delta_c = \frac{10^{0.05A_p} - 1}{10^{0.05A_p} + 1} \quad (2.24)$$

$$\delta_r = 10^{-0.05A_r} \quad (2.25)$$

Luego, δ es igual al mínimo de δ_c y δ_r . La atenuación A es calculada mediante:

$$A = -20 \log_{10}(\delta) \quad (2.26)$$

El parámetro β esta definido por:

$$\beta = \begin{cases} 0 & A < 21 \\ 0.5842(A-21)^{0.4} + 0.07886(A-21) & 21 \leq A \leq 50 \\ 0.1102(A-8.7) & A > 50 \end{cases} \quad (2.27)$$

El orden del filtro esta dado por :

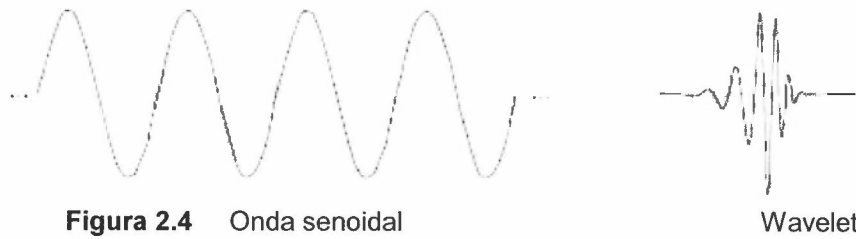
$$N \geq \begin{cases} 0.9222 \frac{2\pi}{\omega_s - \omega_p} & A < 21 \\ \frac{(A-7.95)}{14.36} \frac{2\pi}{\omega_s - \omega_p} & \text{otro} \end{cases} \quad (2.28)$$

2.1.1.3 Wavelets

Un wavelet es una forma de onda de duración limitada que tiene un valor promedio de cero.

Si comparamos wavelets con ondas senoidales, las cuales son la base del análisis de Fourier, las senoidales no tienen duración limitada – se extienden de menos infinito a más

infinito. Donde las senoidales son suaves y predecibles, los wavelets tienden a ser irregulares y asimétricas.



El análisis de Fourier consiste en descomponer una señal en ondas senoidales de varias frecuencias. Similarmente, al análisis wavelet es la descomposición de una señal en versiones con corrimiento y escaladas del wavelet original (wavelet madre). El análisis wavelet analiza una señal unidimensional en tiempo y frecuencia.

Con solo ver la forma de los wavelets y de las ondas senoidales, se puede observar intuitivamente que las señales con cambios bruscos pueden ser mejor analizadas con un wavelet irregular que con una suave senoidal.

También tiene sentido que las características locales pueden ser mejor descritas con wavelets, las cuales tiene alcance local.

Relación entre filtros y las formas wavelets

El objetivo de una transformada wavelet es descomponer una señal en base a otra función base. Proponemos el análisis wavelet porque permite el análisis en tiempo y frecuencia de una señal unidimensional.

El análisis es desarrollado por una convolución con un filtro monodimensional cuyos coeficientes se definen por la siguiente función compleja:

$$\Phi_{k_x}(x) = (1 + \cos(2\pi F_0 x)) \times e^{-j2\pi k_x F_0 x} \quad (2.29)$$

$k_x F_0$ – frecuencia central

donde $|x| < \frac{1}{2F_0}$, $k_x \neq -1, 0, 1$ y k_x es entero.

Esta función es muestreada y ventaneada de tal manera que, $2N + 1$ es el número total de coeficientes, $N = INT(1/2F_0)$ donde N es el orden del filtro wavelet y sólo puede ser entero.

Los parámetros del filtro son F_0 y k_x . De esta manera, la función Φ_{k_x} define un filtro pasabandas con una frecuencia central $k_x F_0$ y con un ancho de banda proporcional a F_0 . El valor de k_x , da la selectividad del filtro.

Por ejemplo un filtro wavelet con selectividad $k_x=4$ forma la siguiente wavelet.

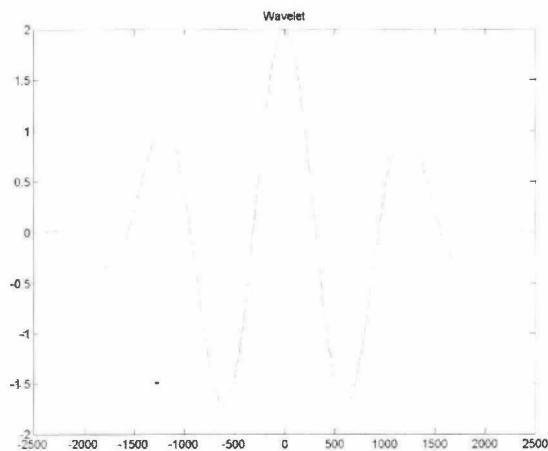


Figura 2.5 Wavelet madre $\Phi(x)$

Por otro lado, el corrimiento de un wavelet se refiere a retrasar o adelantar la forma de onda del wavelet. Matemáticamente, atrasar por k una función $f(x)$ se representa como $f(x-k)$.

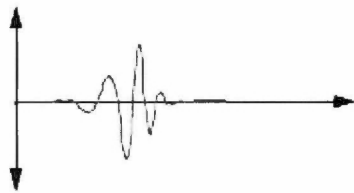
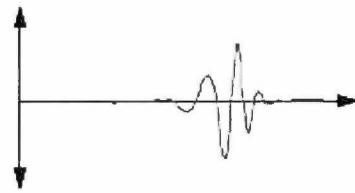


Figura 2.6 Función wavelet $\Phi(x)$



Función wavelet desplazada $\Phi(x-k)$

2.2 Etapa de caracterización

2.2.1 Coeficientes autorregresivos.

Posterior a la etapa de pre-procesamiento, continua la etapa de caracterización de la señal.

La idea principal de esta etapa es identificar y obtener algunos parámetros o características que nos permitan definir a nuestra señal de manera sencilla.

Nuestra etapa de decisión, posterior a la etapa de caracterización, involucra el procesamiento de algoritmos de lógica difusa y redes neuronales. Por resultados anteriores, sabemos que las pruebas realizadas con las muestras en redes neuronales no fueron del todo satisfactorias.

Consideramos que un número menor de parámetros puede ser adecuado para facilitar el entrenamiento en la etapa de Decisión, así como para la obtención de mejores resultados.

Nuestro sistema propuesto involucra el uso de un método para el modelado de señales llamado Autorregresivo (AR). Este método es utilizado muchas veces para el modelado de filtros o para la obtención de espectros de potencia estimados, es decir aproximados.

El método AR es uno muy adecuado para el modelado de sistemas estocásticos. Es decir, sistemas de los cuales no se sabe bien a bien cuál será su comportamiento. El resultado de éste método es una función que predice el comportamiento de un sistema como el original basándose en muestras anteriores y en las probabilidades de que cada muestra se comporte de cierta manera.

En el caso de la estimación del espectro de potencia, el método requiere del cálculo de ciertos coeficientes. Estos coeficientes son lo que se utilizan para reconstruir el espectro estimado de potencia de una señal dada. A estos se les conoce como los coeficientes autorregresivos.

Una idea importante del uso del modelado por AR es el hecho de que se intenta representar a la señal a través de pocos parámetros (número de coeficientes) pero guardando la máxima información posible.

Un tipo de caracterización ampliamente empleado es el espectro de frecuencia de una señal, el cual se obtiene después de aplicar la transformada de Fourier. Nuestra etapa de decisión pudiera estar basada en esta caracterización semejante para obtener la concentración de glucosa. Sin embargo, la longitud del vector de nuestras muestras es grande (alrededor de 500 muestras).

Además de lo mencionado anteriormente, el tamaño del vector se incrementa aún más, debido a que, en la etapa de pre-procesamiento donde se implementaron filtros a la señal para el mejor acondicionamiento de la misma, el resultado fue un vector mayor y en ocasiones hasta del doble en el número de muestras debido a las convoluciones hechas en el procedimiento.

Implementar vectores tan grandes a una red neuronal dificulta y demora su procesamiento, además de que puede volverla menos estable y con posibles errores en los resultados.

La idea principal sobre el uso del modelado autorregresivo es el poder reconstruir el espectro de potencia de nuestras señales con este método a partir de pocos coeficientes.

Creemos que diferenciar los espectros de las señales a través de los coeficientes autorregresivos puede funcionar como parámetros que se introducirán en el sistema de decisión para obtener la concentración de glucosa en solución.

Cabe mencionar que, hasta ahora, no contamos con información que pueda validar el uso de este método y que ello se dilucidará posteriormente al realizar la etapa de Decisión.

Estimación espectral autorregresiva.

El espectro de potencia estimado para un proceso autorregresivo de orden N se define con la siguiente expresión:

$$\hat{P}_x(e^{j\omega}) = \frac{\hat{b}(0)^2}{1 + \sum_{k=1}^N \hat{a}_k(k) e^{-jk\omega}} \quad (2.30)$$

Esta expresión usa los coeficientes obtenidos apartir de una autocorrelación normal.

El arreglo de estas ecuaciones de autocorrelación, queda definido por la ecuación de Yule-Walker:

$$\begin{bmatrix} r_x(0) & r_x(1) & r_x(N-1) \\ r_x(1) & r_x(0) & r_x(N-2) \\ r_x(2) & & \\ r_x(N-1) & & r_x(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_N \end{bmatrix} = - \begin{bmatrix} r_x(1) \\ r_x(2) \\ r_x(N) \end{bmatrix} \quad (2.31)$$

La obtención de las autocorrelaciones normales se obtiene de:

$$\hat{r}_x(k) = \frac{1}{L} \sum_{n=0}^{L-k-1} x(n)x(n+k) \quad (2.32)$$

Posterior a eso se debe calcular la varianza existente de la señal y está definida así:

$$\hat{b}(0)^2 = r_x(0) + \sum_{k=1}^N \hat{a}_k r_x(k) \quad (2.33)$$

Una vez obtenidos los coeficientes, así como la varianza, estos valores pueden incorporarse a la ecuación (2.30) y obtener el espectro estimado de potencia.

Selección del orden del modelo

La selección del orden del modelo es una cuestión importante. Si el orden del modelo es pequeño, entonces el espectro resultante es suave y de poca resolución. En el caso contrario, si el orden del modelo es muy alto, el espectro estimado agregará picos a nuestro espectro, entre otros aspectos debido a la partición de la línea espectral.

Existen diferentes criterios que ayudan a escoger el orden del modelo. La mayoría se basan en minimizar el error del modelo y escoger el orden del proceso que produce dicho error.

En principio, el error es una función decreciente debido al cálculo de la autocorrelación de los datos de la señal. Sin embargo, algunos criterios incluyen en sus expresiones un término de penalización lineal directamente hacia el orden del modelo.

Uno de estos criterios es el de Información de Akaike (AIC), el cual se obtiene de minimizar la función, además de que agrega un término de penalización, $2(N+1)$, para los coeficientes que no reducen el error en la predicción.

$$AIC = n(\log \hat{\sigma}_y^2 + 1) + 2(N + 1) \quad (2.34)$$

Como se mencionó anteriormente, diferenciar los espectros de potencia a través de los coeficientes autorregresivos tal vez pueda funcionar como procedimiento para evaluar en la red neuronal. Por tal motivo, es muy importante que tanto los valores de los coeficientes como el número de ellos, sean lo más ajustados para resultar con un espectro estimado de potencia muy preciso del original.

Para observar el funcionamiento de nuestro algoritmo implementado, es prudente cotejar nuestros resultados con algunos ejemplos en la literatura.

Encontrados en la literatura [9], tenemos dos ejemplos de dos diferentes partes del método del modelado autorregresivo de señales.

En el primero, la parte que corresponde al algoritmo es la del cálculo de los coeficientes. A este se proporcionan tanto las autocorrelaciones estimadas, así como la varianza de valor unitario de un sistema de orden tres:

$$\hat{R}_x(0) = 2.0307, \hat{R}_x(1) = -1.3893, \hat{R}_x(2) = 0.6778, \hat{R}_x(3) = -0.3126$$

Así como los coeficientes autorregresivos estimados de este sistema son:

$$\hat{\alpha}_1 = 0.8932, \hat{\alpha}_2 = 0.3222, \hat{\alpha}_3 = 0.0815$$

Los resultados obtenidos con nuestros algoritmos en Matlab fueron:

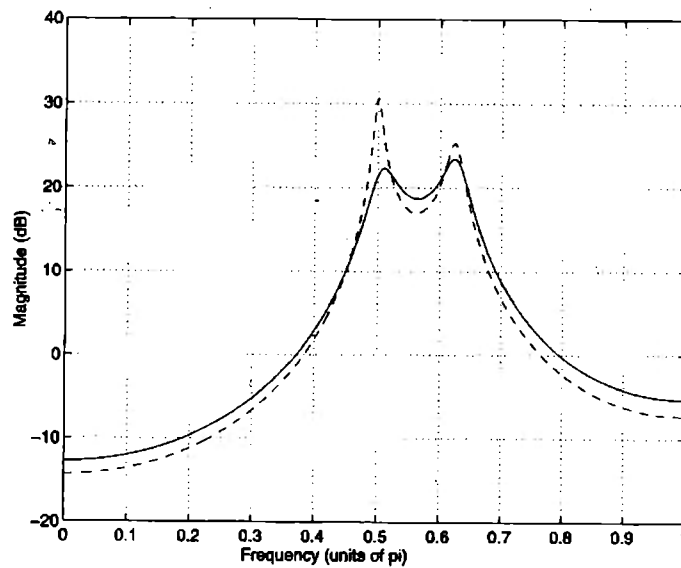
$$a1 = 0.8775, a2 = 0.3224, a3 = 0.0816$$

En el segundo ejemplo [8], la parte implementada es la que corresponde a la reconstrucción y comparación del espectro de potencia de un proceso AR. A este se proporcionan la varianza y los coeficientes correspondientes a un proceso de orden cuatro:

$$\hat{b}(0) = 1.4844$$

$$\hat{a}_1 = 0.7100, \hat{a}_2 = 1.6198, \hat{a}_3 = 0.6193, \hat{a}_4 = 0.6908$$

Gráficamente, el resultado obtenido con nuestro algoritmo en Matlab, corresponde con el ejemplo mostrado de la literatura. El modelado en la literatura también fue resuelto a través del método de Yule-Walker.



(b)

Figura 2.7 Espectro de potencia estimado, método de Yule-Walker

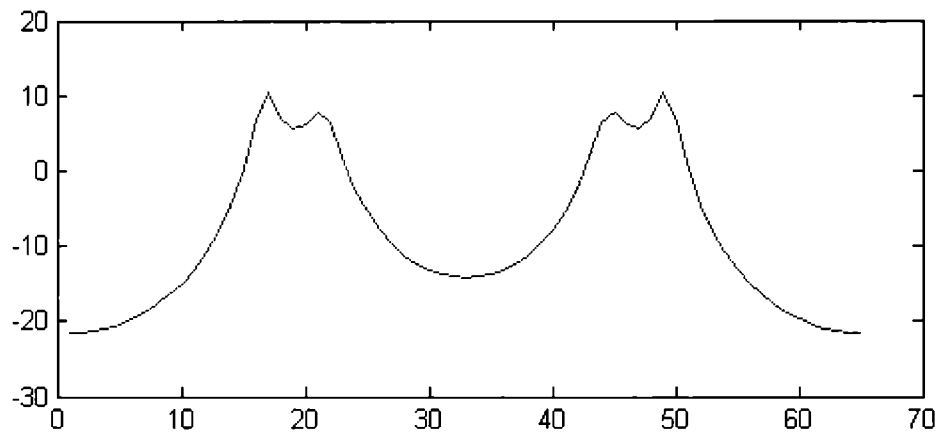


Figura 2.8 Espectro de potencia estimado, comprobación

A continuación se muestra el procedimiento de manera completa para validar nuestro algoritmo. La evaluación se realizará para una señal de audio.

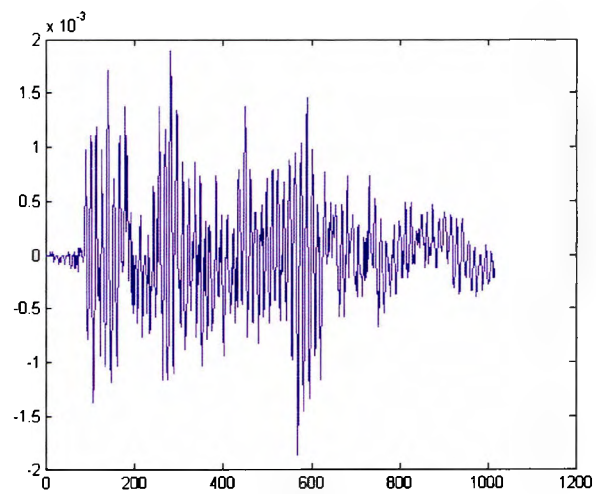


Figura 2.9 Señal de audio

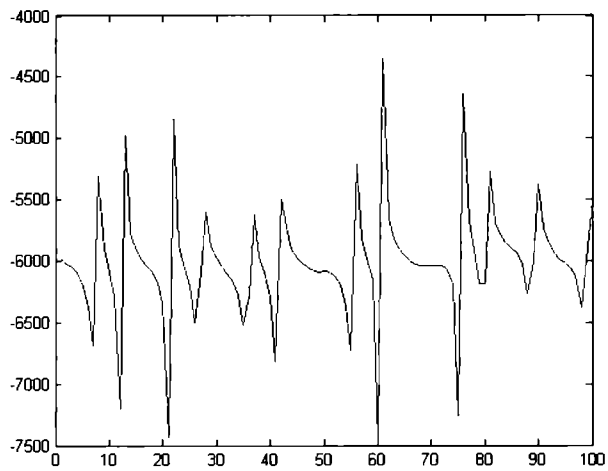


Figura 2.10 Comportamiento de AIC, N=100

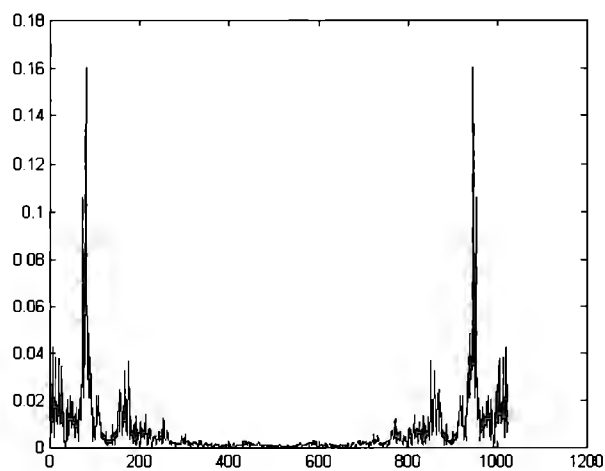


Figura 2.11 FFT, 1024 puntos

Como se podrá observar de la gráfica del comportamiento de AIC, podemos suponer que con aproximadamente 10, 20, 60 y 76 coeficientes se podría tener un buen desempeño de modelado autorregresivo. Se hicieron las pruebas con estos valores para el orden del modelo y los mejores resultados fueron con 10 y 20. A continuación se muestran dichos resultados. Las gráficas, aunque son hechas a base de trazos suaves son las que mejor representan el espectro de la señal original.

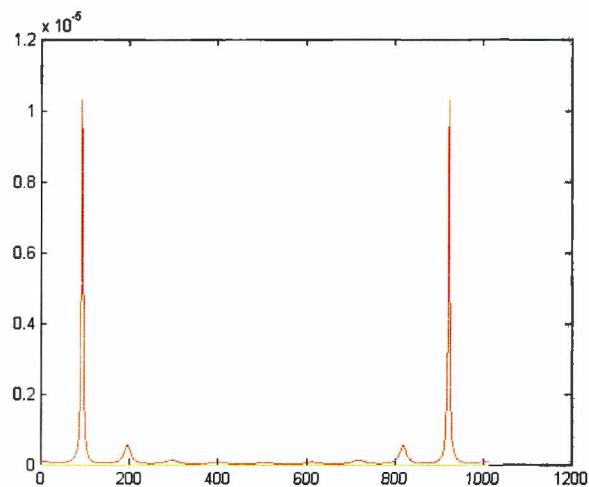


Figura 2.12 Espectro estimado de potencia, modelado AR(10)

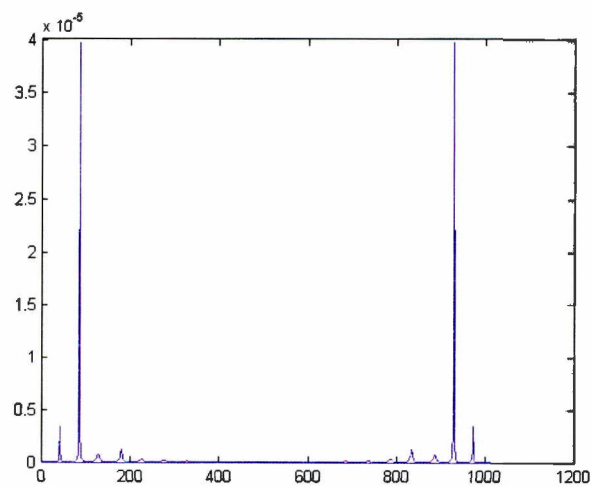


Figura 2.13 Espectro estimado de potencia, modelado AR(10)

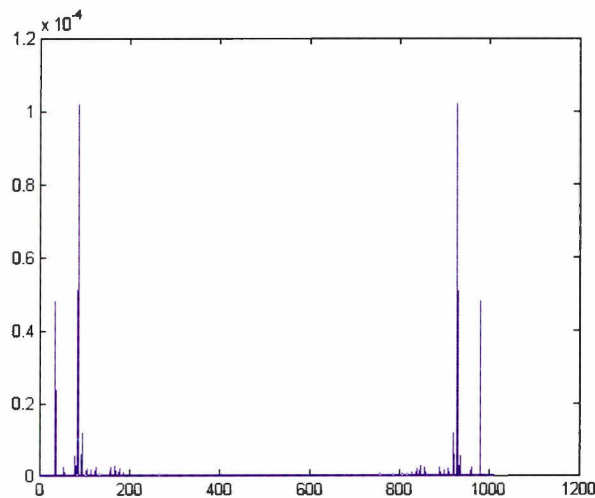


Figura 2.14 Espectro estimado de potencia, modelado AR(100)

2.2.2 Diagramas tiempo - frecuencia

El Análisis de Fourier informa acerca de la presencia o ausencia de determinadas frecuencias en la señal que debemos analizar. Estas frecuencias, poseen cierto carácter universal: el dominio de interés puede estar relacionado con ondas físicas (óptica, vibraciones, sonido,...) o se basa en ciertas periodicidades de acontecimientos (economía, biología, astronomía,...). El Análisis de Fourier es una técnica de gran utilidad para el análisis de las funciones (señales) periódicas o suficientemente regulares, sobre todo desde el descubrimiento de la transformada rápida de Fourier. Pero para señales no periódicas o irregulares presenta eventualmente más inconvenientes que ventajas.

Es importante observar que, con la transformación de Fourier, desaparecen los aspectos temporales de la señal, lo que resulta decisivo en circunstancias como:

- Señal finita: el espectro de Fourier no revela el instante de comienzo de la emisión de la señal ni su término.
- Señal con singularidades: dicho espectro no contiene información sobre la ubicación en el tiempo de las mismas, únicamente puede constatar su presencia.

En particular, si se trata de un análisis en tiempo real, la señal es transformada a medida que los datos son adquiridos. Pero el espectro resultante es ficticio, porque, en cada instante, se desconoce el futuro de la señal, que puede contener frecuencias y comportamientos dignos de tener en cuenta.

La raíz del problema se encuentra en la estructura de las funciones analizadoras. Son funciones eternas y, por lo tanto, no aptas para describir acontecimientos transitorios.

En un esfuerzo para corregir esta deficiencia. Dennis Gabor (1946) adaptó la transformada de Fourier para analizar sólo una pequeña sección de la señal en el tiempo

– una técnica de ventaneo de la señal. La adaptación de Gabor, llamada la Transformada de Fourier del Tiempo Corto (STFT – por sus siglas en inglés), grafica la señal como una función bidimensional del tiempo y la frecuencia.

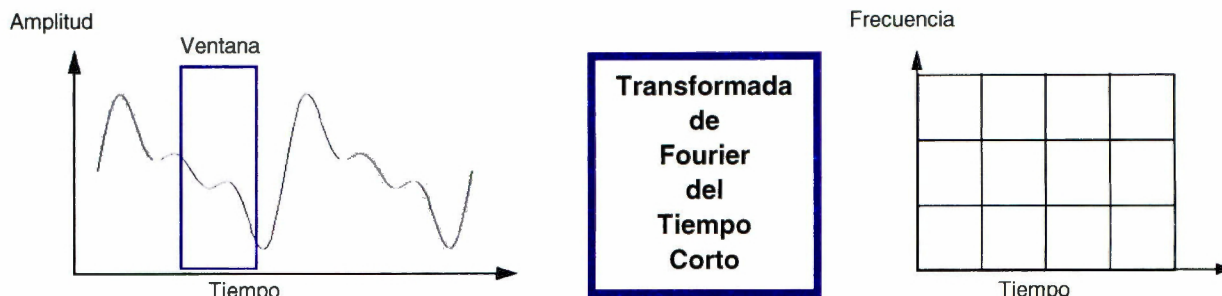


Figura 2.15

La STFT representa un tipo de compromiso entre el análisis en tiempo y frecuencia de una señal. Este provee información de cuándo y a qué frecuencias ocurre un evento en la señal. Sin embargo, solo se puede obtener esta información con precisión limitada, y esa precisión es determinada por el tamaño de la ventana.

Mientras que el compromiso de la STFT entre la información de tiempo y frecuencia pueda ser útil, la desventaja es que una vez elegido un tamaño particular para la ventana de tiempo, la ventana es la misma para todas las frecuencias. Muchas señales requieren una aproximación más flexible - una donde podamos variar el tamaño de la ventana para determinar de manera más exacta el tiempo o la frecuencia.

2.3 Etapa de decisión

Los sistemas de inteligencia artificial nos ayudan a realizar tareas propias de la deducción humana, que son complejas para los mismos humanos o que simplemente requerirían de la colaboración de expertos en la materia. Así los sistemas inteligentes realizan toma de decisiones de manera autónoma con el fin de obtener resultados que de otra forma tendrían que requerir la presencia de personas calificadas. Entre los métodos con los que contamos para auxiliarnos en el desarrollo de sistemas “inteligentes” se encuentran la lógica difusa y las redes neuronales.

2.3.1 Lógica difusa

La lógica difusa fue creada en la década de los 60's por L.A. Zadeh y ha comprobado su utilidad en una gran variedad de aplicaciones. Este tipo de lógica pretende obtener resultados exactos a partir de datos imprecisos por lo que se ha convertido en una herramienta de gran utilidad en aplicaciones electrónicas o computacionales.

En 1965 Zadeh concluyó que las teorías de lógica clásica ponían mucho énfasis y precisión además de que no podían describir sistemas complejos.

En cierta forma, la lógica difusa puede ser vista como un método que permite trasladar oraciones del lenguaje común a un lenguaje matemático y extiende la capacidad de la

lógica clásica. Por ejemplo al calificar que el cielo es azul, uno está tentado a graduar que tan azul es, “muy azul”, “poco azul”, etc., sin tener que dar un valor exacto de que tan azul es en realidad. De esta forma, la lógica difusa procura crear aproximaciones matemáticas en la resolución de ciertos tipos de problemas. En cierto sentido, opera de manera similar al razonamiento humano.

Conjuntos difusos

Para comenzar la realización de sistemas difusos debemos primeramente ampliar la teoría de conjuntos y la lógica booleana; de manera que un individuo pueda pertenecer parcialmente a un conjunto y que las operaciones lógicas además de unos y ceros, puedan ser 0.01 o 0.75, por ejemplo.

De manera común se tiene el concepto de conjunto como una colección bien definida de elementos, en la que es posible determinar para un objeto cualquiera, en un universo dado, si acaso éste pertenece o no al conjunto. La decisión, naturalmente, es “sí pertenece” o bien “no pertenece”.

Con los conjuntos difusos podemos definir subconjuntos, de una manera tal, que cualquier elemento pueda pertenecer a ellos en diferentes grados. Dichos conjuntos difusos pueden ser caracterizados por una función de membresía μ_F , la cual puede tomar valores entre cero y uno e indica el grado con que pertenece a un valor lingüístico, tal como el cielo es azul.

En otras palabras, un conjunto difuso es una función cuyo dominio es el universo y cuyo rango es el intervalo $[0, 1]$. En tanto el grado de pertenencia sea más cercano a 1 tanto más pertenecerá el elemento a un conjunto; y mientras el grado de pertenencia sea más cercano a 0 menos pertenecerá el elemento al conjunto.

Los conjuntos difusos pueden ser representados por un elemento genérico u y su función de membresía de tal manera que, para un universo U continuo el conjunto difuso puede ser representado por $F = \int \mu_F(u)/u$, y para un U discreto el conjunto difuso se

representa mediante $F = \sum_{i=1}^n \mu_F(u_i)/u_i$.

Como habíamos comentado previamente, las funciones de membresía o de pertenencia definen la forma que tendrán los conjuntos difusos. Cada función de membresía tiene ventajas o desventajas sobre las demás. Algunas funciones de membresía comúnmente usadas y ya definidas son las siguientes:

1. Triangular con parámetros a, b, c :

$$\mu_A(x) = \begin{cases} 0, & x \leq a \\ x - a, & a < x \leq b, \\ b - a \\ c - x, & b < x \leq c \\ c - b \\ 0, & c < x \end{cases} \quad (2.35)$$

2. Trapezoidal con parámetros a, b, c, d:

$$\mu_i(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x \leq b \\ 1, & b < x \leq c \\ \frac{d-x}{d-c}, & c < x \leq d \\ 0, & d < x \end{cases} \quad (2.36)$$

3. Gaussiana con m y s

$$\mu_i(x) = e^{-\frac{(x-m)^2}{2s^2}} \quad (2.37)$$

4. "Fuzzy Singleton" con parámetro x_0

$$\mu_i(x) = \begin{cases} 1, & x = x_0; \\ 0, & x \neq x_0. \end{cases} \quad (2.38)$$

Existen además los conceptos de **Support, Crossover Point y Fuzzy Singleton**.

Support de un conjunto difuso F se refiere al conjunto clásico (en el que los elementos pueden solo pertenecer o no) de todos los puntos u en U tal que la función de pertenencia debe ser mayor que cero.

Un *crossover point* hace referencia al elemento u en U que posee una función de pertenencia de 0.5.

Un *singleton* es el conjunto difuso cuyo *support* es un solo punto en U con función de pertenencia de 1.

Operaciones básicas con conjuntos difusos

Se definen las operaciones básicas de lógica difusa mediante los conjuntos difusos A y B en U y sus respectivas funciones de pertenencia.

Podemos definir la unión, intersección y complemento de la siguiente forma:

$$\begin{aligned} (A \cap_i B)(x) &= T(A(x), B(x)) && \text{para toda } x \in X \\ (A \cup_i B)(x) &= S(A(x), B(x)) && \text{para toda } x \in X \end{aligned}$$

Donde T y S son cualquier norma T o norma S respectivamente, algunas veces denotadas como \star_T o \star_S .

$$\overline{A}(x) = n(A(x)) \quad \text{para toda } x \in X$$

Donde n denota negación.

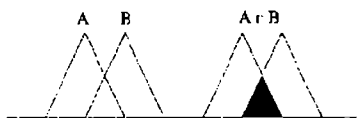
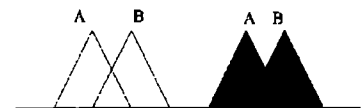
A continuación se presenta las tablas de las distintas normas T y S.

Nombre	Norma T	Norma S
Zadeh	$M(x, y) = \min(x, y)$	$M'(x, y) = \max(x, y)$
Algebraic	$\prod(x, y) = x \cdot y$	$\prod'(x, y) = x + y - x \cdot y$
Bounded	$W(x, y) = \max(x + y - 1, 0)$	$W'(x, y) = \min(x + y - 1, 0)$
Fodor	$\min_u(x, y) = \begin{cases} \min(x, y) - si - (x + y) > 1 \\ 0 - -si - -(x + y) \leq 1 \end{cases}$	$\max_l(x, y) = \begin{cases} \max(x, y) - si - (x + y) > 1 \\ 1 - -si - -(x + y) \leq 1 \end{cases}$
Drastic	$Z(x, y) = \begin{cases} \min(x, y) - si - \max(x, y) = 1 \\ 0 - -si - -\max(x, y) \neq 1 \end{cases}$	$Z'(x, y) = \begin{cases} \max(x, y) - si - \min(x, y) = 0 \\ 1 - -si - -\min(x, y) \neq 0 \end{cases}$
Einstein	$E(x, y) = \frac{xy}{2 - (x + y - xy)}$	$E'(x, y) = \frac{x + y}{1 + xy}$

Tabla 1. Normas T y Norma S'

Para el caso específico de Zadeh, las operaciones básicas se realizan como se describe a continuación:

Unión: Esta dada por el valor máximo de los grados de pertenencia para toda u, de los conjuntos difusos A y B.
 $M(x, y) = \max(x, y)$



Intersección: El valor mínimo de los grados de pertenencia de los conjuntos A y B. $M'(x, y) = \min(x, y)$

Complemento: Esta dado por la diferencia entre el valor máximo que puede obtener un conjunto difuso, 1, y los valores de la función de pertenencia para toda u del conjunto al que se le desea sacar el complemento.

Productos cartesianos, relación y composición

Un operador conjunción determina un operador de *unión* de conjuntos difusos. Para dos conjuntos usuales A y B, su *producto cartesiano* consta de todas las parejas ordenadas de la forma (a, b) donde $a \in A$ y $b \in B$. Así pues, si \star_T es un operador conjunción donde A y B son conjuntos difusos en universos X y Y, su *producto cartesiano* es el conjunto difuso $A \times B : X \times Y \rightarrow [0, 1], (x, y) \rightarrow A(x) \star_T B(y)$.

Una relación, en el sentido usual, entre dos conjuntos es un subconjunto de su producto cartesiano. Por tanto, se puede considerar a una *relación difusa* entre dos universos como un conjunto difuso en su producto cartesiano.

Una composición crea una relación difusa entre dos conjuntos difusos que no se encuentran directamente relacionados sin pasar por un conjunto difuso intermedio y se encuentra denotada por R o S = $[(x, z), \text{Max}(\mu_R(x, y) \star_T \mu_S(y, z))]$, donde R es una relación difusa en $X \times Y$ y S es una relación difusa en $Y \times Z$. (\star puede ser un operador de la

clase *triangular*, *norms*, *namely*, *minimum*, *algebraic product*, *bounded product* o *drastic product*, en este caso se estableció como el operador de norma T , \min).

Etiquetas lingüísticas

Una *etiqueta lingüística* es el nombre de un conjunto difuso. Es decir, es una terceta donde por ejemplo se tiene (Nombre, A, X), donde *Nombre* es el nombre asociado al conjunto difuso *A* en el universo *X*.

Por ejemplo, si $X = [0,200]$ éste es el conjunto de velocidades posibles de automóviles, medidas en km/hr. La etiqueta lingüística *Lento* puede corresponder a un conjunto difuso.

Lógica difusa y razonamiento aproximado:

Existen dos reglas de inferencia o implicación llamadas *Generalizad Modus Ponens* (GMP) y *Generalizad Modus Tollens* (GMT)

GMT esta basado en relaciones de negación y GMP esta basado en condiciones afirmativas.

El modo de implicación más comúnmente utilizado es GMP y se realiza mediante instrucciones del tipo **if-then**.

Usamos implicación para conocer, en base a los antecedentes, la salida. Una ventaja de usar la implicación es que se pueden usar las operaciones AND y OR dado que no afectan las reglas que son válidas.

Existen dos tipos de sistemas difusos muy empleados:

- **Tipo Mamdani**
- **Tipo Takagi-Sugeno**

El primero utiliza un método de composición difusa llamado Min-Max. De manera sencilla y complementando lo descrito anteriormente, los métodos de composición definen distintas posibilidades o caminos en cuanto a las relaciones entre conjuntos difusos. Para el caso Min-Max, éste se puede definir como tomar el mínimo de los antecedentes con el máximo de las consecuencias. Este método de composición es uno de los más empleados debido a que ha sido uno con los que mejores resultados se han logrado.

El segundo es muy parecido al primero y también es ampliamente usado. La diferencia radica al final del sistema difuso, donde la función de membresía ya no tiene ninguna de las formas convencionales descritas anteriormente. En cambio, se utiliza un polinomio para evaluar la regla. Este sistema difuso es el que se utiliza para los sistemas ANFIS que describiremos posteriormente.

Para estos métodos existen varias formas de calcular el valor real de salida. Para los sistemas tipo Mamdani el más comúnmente usado es el de Centroide, el cual se expresa de manera siguiente:

$$Salidareal = \frac{\sum_a^b \mu(x) \cdot x}{\sum_a^b \mu(x)} \quad (2.39)$$

Para el tipo Takagi-Sugeno, la forma de calcular es la misma, solo que ahora no se hace una sumatoria de los valores proyectados por la función de membresía sino a través de evaluar en el polinomio.

La estructura de lógica difusa queda entonces como tal:

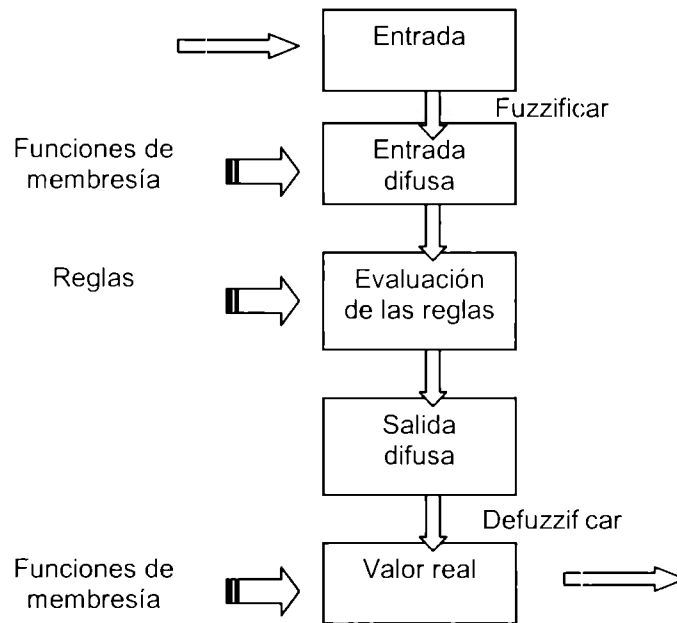


Fig 2.16 Diagrama de flujo de un sistema difuso

2.3.2 Redes neuronales artificiales

Las Redes Neuronales Artificiales (ANN's de Artificial Neural Networks) son una simulación abstracta de los sistemas nerviosos biológicos, formados de elementos simples de procesamiento altamente interconectados llamados "neuronas" o "nodos". Cada neurona consiste de una unión, la cual suma las entradas de las otras neuronas por medio de conexiones con pesos asociados, y de una función de activación que se activa si la fuerza combinada de la señal de entrada es superior a un cierto nivel, en el caso general el valor de activación de la neurona viene dado por una función de activación $f(\cdot)$, que genera la salida de la neurona. La salida de una neurona se distribuye para utilizarla como entrada de otras neuronas. Las neuronas transmiten señales a otras neuronas por medio de sus conexiones, que atenúan o amplifican la señal transmitida dependiendo del peso que tengan asociado. Este proceso es a menudo modelado como una regla de propagación representada por la función de red $u(\cdot)$.

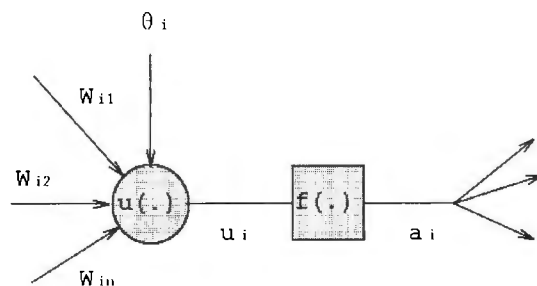


Fig 2.17 Estructura básica de una neurona.

Las ANN's se pueden clasificar en tres diferentes tipos dependiendo de su estructura de retroalimentación: recurrentes, localmente recurrentes y no recurrentes. Un tipo especial de ANN no recurrente es la **ANN de conexiones de atrás hacia adelante** (FNN de Feedforward Neural Network), que consiste de capas de neuronas con conexiones asociadas a pesos que conectan las salidas de las neuronas en una capa a las entradas de las neuronas en la siguiente capa.

Taxonomía de las redes neuronales artificiales

Existen dos fases en toda aplicación de las redes neuronales: la fase de aprendizaje o entrenamiento y la fase de prueba. En la fase de entrenamiento, se usa un conjunto de datos o patrones de entrenamiento para determinar los pesos (parámetros de diseño) que definen el modelo neuronal. Una vez entrenado este modelo, se usará en la llamada fase de prueba o funcionamiento directo, en la que se procesan los patrones de prueba que constituyen la entrada habitual de la red, analizándose de esta manera las prestaciones definitivas de la red.

- Fase de Aprendizaje: una característica de las redes neuronales es su capacidad de aprender. Aprenden por la actualización o cambio de los pesos sinápticos que caracterizan a las conexiones. Los pesos son adaptados de acuerdo a la información extraída de los patrones de entrenamiento nuevos que se van presentando. Normalmente, los pesos óptimos se obtienen optimizando (minimizando o maximizando) alguna "función de energía".
- Fase de Prueba: los parámetros de diseño de la red neuronal se han obtenido a partir de unos patrones representativos de las entradas que se denominan patrones de entrenamiento. Los resultados pueden ser tanto calculados de una vez como adaptados iterativamente, según el tipo de red neuronal, y en función de las ecuaciones dinámicas de prueba. Una vez calculados los pesos de la red, los valores de las neuronas de la última capa, se comparan con la salida deseada para determinar la validez del diseño.

Redes neuronales supervisadas y no supervisadas

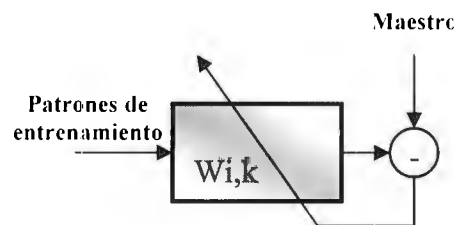
Las redes neuronales se clasifican comúnmente en términos de sus correspondientes algoritmos o métodos de entrenamiento: redes no supervisadas, y redes de entrenamiento supervisado. Para las redes de pesos fijos no existe ningún tipo de entrenamiento.

Entrenamiento supervisado

Los datos para el entrenamiento están constituidos por varios pares de patrones de entrenamiento de entrada y de salida. El hecho de conocer la salida implica que el entrenamiento se beneficia de la supervisión de un maestro. Dado un nuevo patrón de entrenamiento, por ejemplo, (k+1)-ésimo, los pesos serán adaptados de la siguiente forma:

$$W_{i,k+1} = W_{i,k} + ? W_{i,k}$$

en la siguiente figura se puede ver un sistema de entrenamiento supervisado:



Entrenamiento no supervisado

Para los modelos de entrenamiento No Supervisado, el conjunto de datos de entrenamiento consiste sólo en los patrones de entrada. Por lo tanto, la red es entrenada sin el beneficio de un maestro. La red aprende a adaptarse basada en las experiencias recogidas de los patrones de entrenamiento anteriores.

Algoritmos de adaptación de pesos

Para un ADALINE

La regla delta Widrow-Hoff

La regla delta Widrow-Hoff, puede ser usada para adaptar el vector de pesos de un adaline (ADAPtative LINear Element, Elemento lineal adaptable). La regla delta minimiza el error cuadrático medio entre la salida deseada y la salida actual en todas las k y puede ser escrito como sigue

$$w_{k+1} = \begin{cases} w_k + \frac{\alpha e_k x}{x^T x} & \text{si } x^T x \neq 0 \\ w_k & \text{si } x^T x = 0 \end{cases} \quad (2.40)$$

donde k = índice de tiempo o iteración,

$w_k = [w_{1,k}, w_{2,k}, \dots, w_{n,k}]^T \in \mathfrak{R}^n$ es el vector de pesos en el tiempo k ,

$x = [x_1, x_2, \dots, x_n]^T \in \mathfrak{R}^n$ es el vector de entrada,

$e_k = y_d - y_k \in \mathfrak{R}$ es el error en el tiempo k , y

$\alpha \in \mathfrak{R}$ es el factor de reducción.

Asumimos $x \neq 0$ que implica que $x^T x \neq 0$. La dinámica de error es

$$e_{k+1} = (1 - \alpha)e_k \quad (2.41)$$

El error e_k converge de manera asintótica a 0 a una razón de $1 - \alpha$ si y solo si $0 < \alpha < 2$.

También asumimos que el mismo vector de entrada se presenta en iteraciones sucesivas. Si el vector de entrada es constante, adaptamos el vector de pesos hasta que el error es suficientemente pequeño, y si el vector de entrada cambia respecto al tiempo, continuamos adaptando el vector de pesos. Esta continua adaptación de pesos es necesaria para las aplicaciones, donde el vector de entrada puede ser no constante.

Generalización de la regla delta.

Si $\theta: \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ es un operador. El vector w_k se adapta de acuerdo a la regla

$$w_{k+1} = \begin{cases} w_k + \frac{\alpha e_k}{x^T \theta(x)} \theta(x) & \text{si } \theta(x)^T \neq 0 \\ w_k & \text{si } \theta(x)^T = 0 \end{cases} \quad (2.42)$$

Si θ es el operador identidad $\theta(x) = x$, el algoritmo de adaptación se vuelve la regla delta. Si por otro lado

$$\theta(x) = \begin{bmatrix} \text{sgn}(x_1) \\ \vdots \\ M \\ \vdots \\ \text{sgn}(x_n) \end{bmatrix} \quad \text{donde} \quad \text{sgn}(x) = \begin{cases} +1 & \text{si } x > 0 \\ -1 & \text{si } x < 0 \end{cases} \quad (2.43)$$

asumimos que $x^T \theta(x) \neq 0$ para toda $x \neq 0$. Sin embargo, si $x^T \theta(x) = 0$ para alguna $x \neq 0$, establecemos $w_{k+1} = w_k$.

Para una FNN de dos capas

Notación de operadores

Si $\Psi: \mathfrak{R}^p \rightarrow \mathfrak{R}^p$ es un operador con la siguiente propiedad:

$$\Psi(-x) = -\Psi(x) \quad (2.44)$$

Solo consideramos el Ψ que tiene la siguiente forma diagonal

$$\Psi(x) = \begin{bmatrix} \Psi_1(x_1) \\ M \\ \Psi_p(x_p) \end{bmatrix} \quad \text{donde} \quad \psi: \mathfrak{R} \rightarrow \mathfrak{R} \quad (2.45)$$

Usamos las siguientes funciones de \mathfrak{R} a \mathfrak{R} , comunes en las redes neuronales.

$$\text{idn}(x) = x, \quad (2.46)$$

$$\text{sat}(x) = \begin{cases} +1 & \text{si } x > +1 \\ x & \text{si } x \in [-1, +1], \\ -1 & \text{si } x < -1 \end{cases} \quad (2.47)$$

$$\text{sgn}(x) = \begin{cases} +1 & \text{si } x > 0 \\ -1 & \text{si } x < 0 \end{cases} \quad (2.48)$$

$$\text{sig}(x) = \frac{1 - e^{-x}}{1 + e^{-x}}, \quad (2.49)$$

$$\text{tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (2.50)$$

Tenemos que $x^T \Psi(x) \neq 0$ si y solo si $x \neq 0$. Ahora podemos describir la estructura multicapa de las FNN's usando vectores, matrices, aritmética matriz / vector, y operadores Ψ .

El algoritmo de actualización de la matriz de pesos esta dado por

$$W_{i,k+1} = W_{i,k} + \eta W_{i,k} \quad i = 1, 2 \quad (2.51)$$

donde

$$\Delta W_{1,k} = \begin{cases} -2z_{1,k} \eta_1^T(x) & \text{si } \eta_1^T(x) x \neq 0 \\ \eta_1^T(x) x & \\ 0 & \text{si } \eta_1^T(x) x = 0 \end{cases} \quad (2.52)$$

y

$$\eta W_{2,k} = \begin{cases} -2W_{2,k} - A e_k \eta_2^T(y_{1,k}) & \text{si } \eta_2^T(y_{1,k}) y_{1,k} \neq 0 \\ \eta_2^T(y_{1,k}) y_{1,k} & \\ 0 & \text{si } \eta_2^T(y_{1,k}) y_{1,k} = 0 \end{cases} \quad (2.53)$$

Aquí θ_1 y θ_2 son operadores.

La dinámica de error de la FNN es

$$e_{k+1} - e_k = -Ae_k \quad \text{ó} \quad e_{k+1} = (I_{n_2} - A)e_k \quad (2.54)$$

Para hacer que el error converja de manera asintótica a 0, debemos escoger A de tal manera que los eigenvalores de $I_{n_2} - A$ estén contenidos en la unidad de disco abierto del plano complejo.

Para una FNN de tres capas

El algoritmo de actualización de la matriz de pesos esta dado por

$$W_{i,k+1} = W_{i,k} + \eta W_{i,k} \quad i = 1,2,3$$

donde

$$\Delta W_{1,k} = \begin{cases} -2z_{1,k} \theta_1^1(x) & \text{si } \theta_1^1(x)x \neq 0 \\ \theta_1^1(x)x & \\ 0 & \text{si } \theta_1^1(x)x = 0 \end{cases} \quad (2.55)$$

$$\eta W_{2,k} = \begin{cases} -2z_{2,k} \theta_2^1(y_{1,k}) & \text{si } \theta_2^1(y_{1,k})y_{1,k} \neq 0 \\ \theta_2^1(y_{1,k})y_{1,k} & \\ 0 & \text{si } \theta_2^1(y_{1,k})y_{1,k} = 0 \end{cases} \quad (2.56)$$

$$\eta W_{3,k} = \begin{cases} Ae_k \theta_3^1(y_{2,k}) & \text{si } \theta_3^1(y_{2,k})y_{2,k} \neq 0 \\ \theta_3^1(y_{2,k})y_{2,k} & \\ 0 & \text{si } \theta_3^1(y_{2,k})y_{2,k} = 0 \end{cases} \quad (2.57)$$

Aquí θ_1 , θ_2 y θ_3 son operadores.

La dinámica de error de la FNN es

$$e_{k+1} = (I_{n_3} - A)e_k \quad (2.58)$$

Para hacer que el error converja a 0, debemos escoger A de tal manera que los eigenvalores de $I_{n_3} - A$ estén contenidos en la unidad de disco abierto del plano complejo.

Estructura de conexión de atrás hacia delante

Los aspectos más característicos de las redes neuronales artificiales son la estructura de conexión y el tamaño de la red.

Una red neuronal se determina por la neurona y la matriz de pesos. El comportamiento de la red depende en gran medida del comportamiento de la matriz de pesos. Hay tres tipos de capas de neuronas: la de entrada, las ocultas y la de salida. Entre dos capas de neuronas existe una red de pesos de conexión, que puede ser hacia adelante.

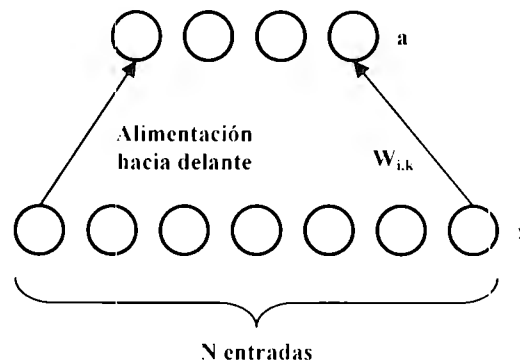


Fig. 2.19 Alineación hacia adelante

Conexiones hacia delante: para todos los modelos neuronales, los datos de las neuronas de una capa inferior son propagados hacia las neuronas de la capa superior por medio de las redes de conexiones hacia adelante.

Las conexiones sinápticas pueden ser total o parcialmente interconectadas. También es posible que las redes sean de una capa con el modelo de pesos hacia atrás o bien el modelo multicapa hacia adelante. Es posible así mismo, el conectar varias redes de una sola capa para dar lugar a redes más grandes.

Tamaño de las redes neuronales

En una red multicapa de propagación hacia adelante, puede haber una o más capas ocultas entre las capas de entrada y salida. El tamaño de las redes depende del número de capas y del número de neuronas ocultas por capa.

- *Número de capas:* en una red multicapa, hay una o más capas de neuronas *ocultas* entre la entrada y la salida. El número de capas se cuenta a menudo a partir del número de capas de pesos (en vez de las capas de neuronas).
- *Número de neuronas ocultas:* El número de unidades ocultas está directamente relacionado con las capacidades de la red. Para que el comportamiento de la red sea correcto (esto es, generalización), se tiene que determinar apropiadamente el número de neuronas por cada capa oculta.

Generalización

La generalización se puede extender a un problema de interpolación. El sistema es entrenado por un gran conjunto de muestras de entrenamiento basados en un procedimiento de aprendizaje supervisado. Una red se considera que esta entrenada con éxito si puede aproximar los valores de los patrones de entrenamiento y puede dar interpolaciones "suaves" para el espacio de datos no entrenado. El objetivo de la Generalización es dar una respuesta correcta a la salida para un estímulo de entrada que no ha sido entrenado con anterioridad. El sistema debe inducir la característica saliente del estímulo a la entrada y detectar la regularidad. Tal habilidad para el descubrimiento de esa regularidad es crítica en muchas aplicaciones. Esto hace que el sistema funcione eficazmente en todo el espacio, incluso si ha sido entrenado solo por un conjunto limitado de ejemplos.

2.3.3 ANFIS

La lógica difusa es un método ideal para expresar conocimiento incierto e incompleto, sin embargo, el principal problema que presenta esta técnica es la concepción de las reglas del tipo *if-then*. Dicha generación automática de estas reglas ha sido investigada intensivamente y se ha llegado a su realización valiéndose de la ayuda de la capacidad de aprendizaje de las redes neuronales artificiales.

Uno de los diferentes sistemas neuro-difusos existentes es el llamado ANFIS (Adaptive Neuro Fuzzy Inference System), significa sistema de inferencia difuso basado en redes neuronales adaptables. Como su nombre lo dice es un modelo híbrido de sistemas difusos con redes neuronales.

ANFIS es un método para sintonizar una base de reglas existentes mediante un algoritmo de aprendizaje basado en la recolección de datos de entrenamiento, lo que permite a la base de reglas adaptarse a su entorno.

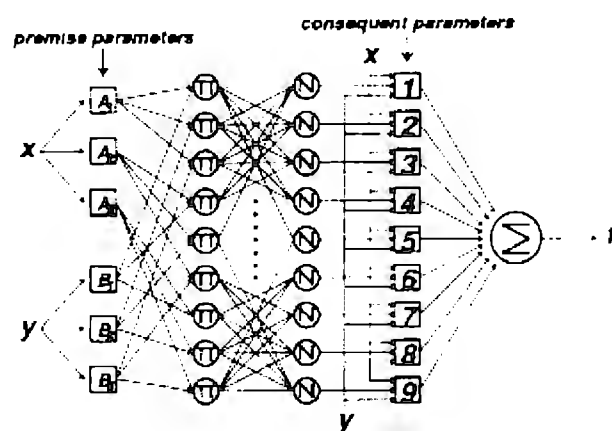


Fig. 2.20 Estructura general de un sistema ANFIS [13]

Una diferencia principal de ANFIS con las redes neuronales es que la capa de entrada de estos sistemas contiene funciones de membresía que proyectarán las entradas hacia conjuntos difusos. Cabe señalar que esta capa de entrada requiere de algunos parámetros o premisas que ayudarán a modificar la forma de la función de membresía hacia alguna de las mencionadas en el apartado de lógica difusa. Generalmente la función empleada es de tipo campana. A este tipo de capas donde se pueden utilizar otros parámetros para modificar su característica se conocen como nodos adaptables. Estos nodos generalmente se encuentran en la capa de entrada y en la penúltima.

En la penúltima capa también se encuentra un nodo adaptable y es ahí donde se realiza la evaluación de las reglas tipo Takagi-Sugeno.

Existen además otros nodos que se les conoce como fijos. Generalmente se encuentran en las capas intermedias u ocultas.

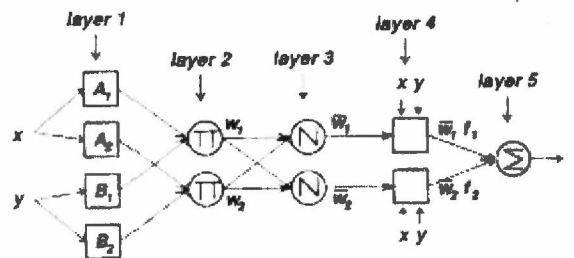


Fig. 2.21 Función de cada capa de un sistema ANFIS [13]

Como se puede apreciar en este diagrama de manera general, las capas intermedias también realizan una labor importante en el proceso de entrenamiento.

Supongamos un sistema difuso con dos entradas, x e y, junto con una salida z y que tiene dos reglas de tipo TSK.

Regla 1: Si x es A1 e y es B1 entonces $f_1 = p_1x + q_1y + r_1$;

Regla 2: Si x es A2 e y es B2 entonces $f_2 = p_2x + q_2y + r_2$;

Capa 1: Fuzzifica cada entrada, y cada nodo de la capa 1 se define mediante una función

$$O_i^1 = \mu_{A_i}(x) = \frac{1}{1 + \left[\frac{x - c_i}{\alpha_i} \right]^{2b_i}} \quad (2.59)$$

Capa 2: Las capas intermedias realizan funciones distintas. La segunda capa realiza la multiplicación de las entradas para obtener la fuerza de operación de las entradas y su saber así su mayor o menor contribución a los pasos siguientes.

$$O_i^2 = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), i = 1, 2 \quad (2.60)$$

Capa 3: Calcula la razón de disparo de cada regla. Ésta realiza una normalización de las salidas anteriores. En base a estos resultados es que se lleva a cabo la evaluación de las reglas Takagi-Sugeno.

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}, i = 1, 2 \quad (2.61)$$

Capa 4: Se realiza la evaluación de las reglas Takagi-Sugeno

$$O_i^4 = \bar{w}_i f_i = w_i (p_i x + q_i y + r_i). \quad (2.62)$$

Capa 5: Posteriormente, la capa de salida realiza un promedio ponderado de las entradas respecto de los polinomios activados y entrega el resultado.

$$O_i^5 = \frac{\sum \bar{w}_i f_i}{\sum \bar{w}_i} \quad (2.63)$$

3. Diseño del experimento.

3.1 Adquisición de datos

Se trabajó para la adquisición de datos en el Instituto Nacional de Astrofísica Óptica y Electrónica, INAOE. Ahí se cuenta con dos espectrómetros. El primero se encuentra en sus laboratorios de óptica, está calibrado y otorga medidas con certificación internacional. Éste trabaja en el rango de longitudes de onda de 300 a 900nm. El segundo se encuentra en sus laboratorios de electrónica, también está calibrado pero no con los requisitos necesarios para la certificación internacional, este último trabaja en el rango de longitudes de onda entre 1200 y 2500 nm.

3.1.1 Equipo utilizado

Las muestras se tomaron en el laboratorio de microelectrónica del Instituto Nacional de Astrofísica Óptica y Electrónica que se encuentra en Tonantzintla, Puebla. El equipo que se utilizó fue un espectrómetro Bruker Vector 22 el cual puede trabajar en el rango infrarrojo e infrarrojo cercano. Para este proyecto se utilizó la fuente NIR que es un foco de halógeno de tungsteno y trabajando en el rango espectral de $10,000\text{ cm}^{-1}$ a $4,000\text{ cm}^{-1}$. La temperatura y la humedad en la celda de medición estaban controladas con nitrógeno. Esto es importante pues no se puede utilizar este espectrómetro para hacer experimentos sobre los efectos de la temperatura y condiciones ambientales en las mediciones los cuales sabemos que existen [4].

3.1.2 Datos disponibles

Se tiene un grupo de datos que está formado por 35 muestras, cada una con 5 barridos y que van de 500mg/dl hasta 50mg/dl, en este caso la resolución es de 5 cm^{-1} .

El agua absorbe fuertemente energía infrarroja, por ello, los espectros están dominados por este efecto y los espectros de soluciones se parecen mucho a los espectros de agua.

Para obtener el espectro de la glucosa, al espectro de la solución se le resta el espectro del agua de acuerdo a la ley de Lambert – Beer que se describe como sigue:

$$A = \log_{10}\left(\frac{I_0}{I}\right) = \log_{10}\left(\frac{1}{T}\right) = \epsilon cl \quad (3.1)$$

Donde:

A = absorción

I_0 = intensidad de la luz incidente

I = intensidad de luz absorbida

T = intensidad de luz transmitida

ϵ = coeficiente de absorción molar

c = la concentración de la sustancia analizada

l = longitud de la muestra

Esta ecuación se puede leer como: la cantidad de absorción de una sustancia en particular es directamente proporcional a la concentración de la sustancia. Eso es lo que nos importa, pues es esta propiedad la que justifica la resta del espectro del agua al espectro de la concentración.

Una vez obtenidos los espectros se les aplica el tratamiento de filtrado, se caracterizan y posteriormente se pasarán a la red neuronal para obtener una medida de la concentración de glucosa correspondiente al espectro.

3.2 Pre-procesamiento de las señales.

3.2.1. Filtrado

Para lograr un eficiente reconocimiento de las concentraciones de glucosa, por medio de los algoritmos de decisión que se van a implementar, es necesario eliminar la parte del espectro que no nos provee información útil y que por el contrario podría dificultar el entrenamiento o aprendizaje del sistema.

Es por ello que se realizaron distintos tipos de filtrado para identificar el más óptimo a implementar.

3.3 Sistema de decisión

Se implementan métodos de decisión basados en sistemas inteligentes como lo son redes neuronales, lógica difusa y una combinación de estos dos que son los sistemas ANFIS. Se decidió utilizar estos métodos ya que en trabajos anteriores [1] se concluyó que métodos como la regresión PLS eran una opción viable para la generalización de los resultados. Se espera que estos métodos logren una buena aproximación a los valores reales de las concentraciones de glucosa en las soluciones preparadas.

4. Resultados

4.1 Resultados de la adquisición.

A continuación se muestra el resultado de la concentración de 50 mg/dl que se obtuvo anteriormente en cuanto a la adquisición de espectros de infrarrojo cercano de soluciones de glucosa con agua, así como el espectro de agua únicamente.

El grupo de datos disponibles se adquirió con la intención de tener espectros para concentraciones que estuvieran en un rango limitado. En este grupo la concentración mayor es de 500mg/dl.

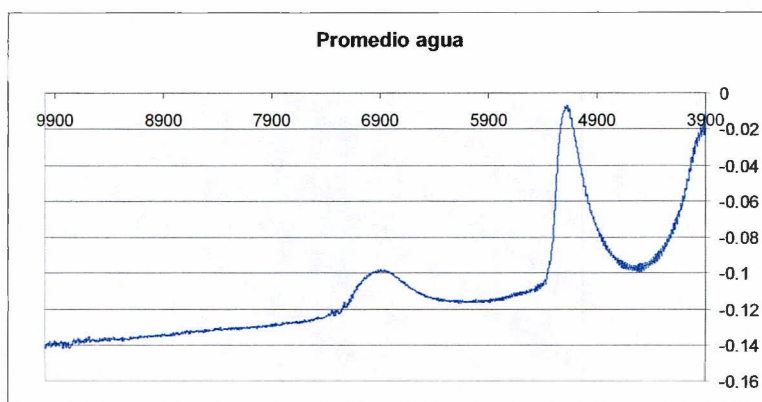


Figura 4.1 Espectro NIR del promedio de 5 muestras de agua

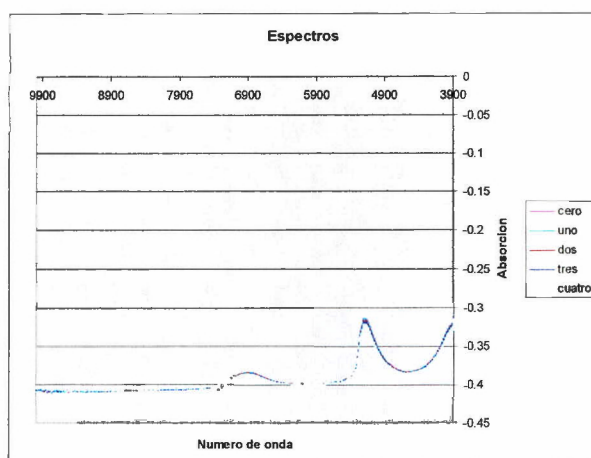


Figura 4.2 Espectro NIR de 5 muestras de solución de glucosa con concentración de 50 mg/dl

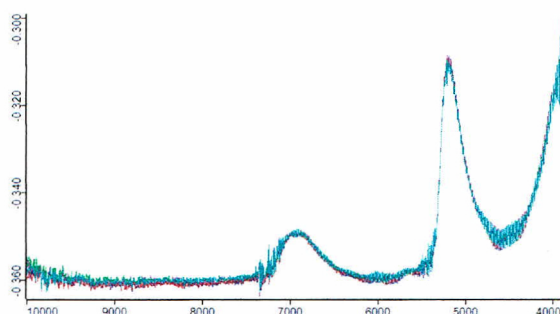


Figura 4.3 Espectro NIR de 5 muestras de solución de glucosa concentración de 238 mg/dl [1]

Todos los espectros comparten la forma del espectro de agua porque como ya se mencionó, el efecto que predomina es la absorción NIR del agua. Debido a lo mencionado anteriormente a los espectros de soluciones se les resta el promedio de los espectros de agua.

El resultado de la resta mencionada se muestra en las siguientes figuras de algunas de las concentraciones utilizadas.

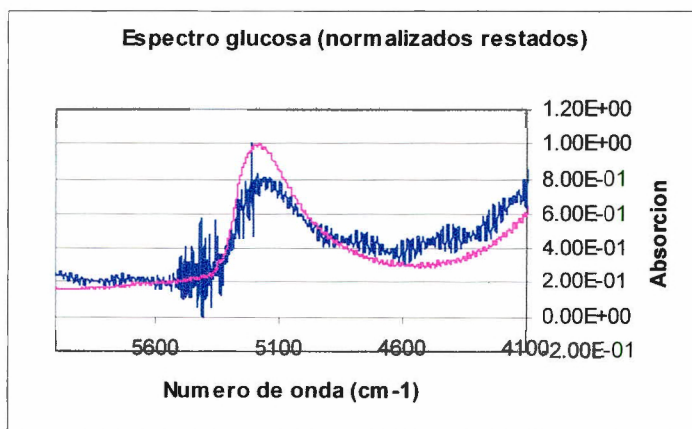


Figura 4.4 Espectro de glucosa para una concentración de 50 mg/dl

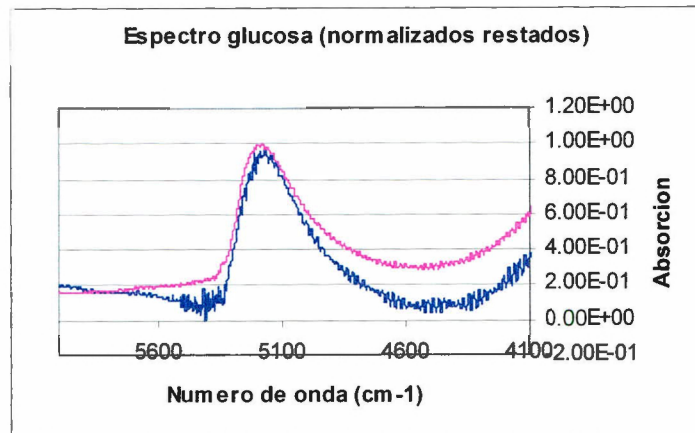


Figura 4.5 Espectro de glucosa para una concentración de 90 mg/dl

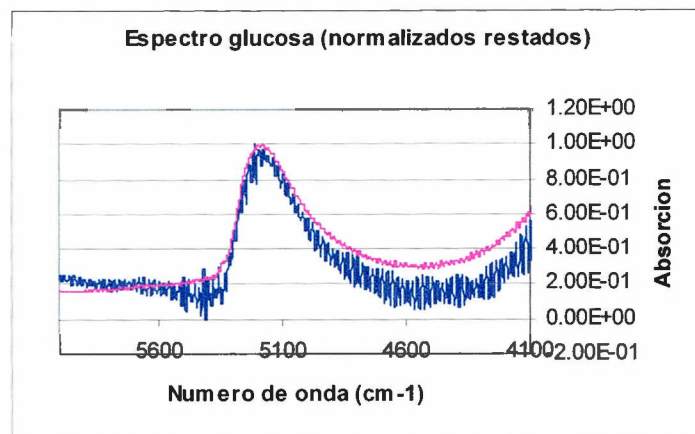


Figura 4.6 Espectro de glucosa para una concentración de 166 mg/dl

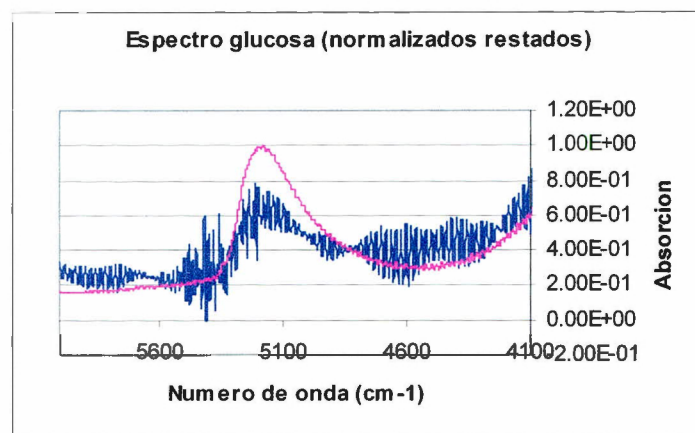


Figura 4.7 Espectro de glucosa para una concentración de 238 mg/dl

Se hace el mismo tratamiento para todos los espectros que se obtuvieron. En las gráficas, ambos espectros están normalizados con el fin de que se observen

correctamente pero en realidad, el orden de absorción NIR de la glucosa es aproximadamente 3 veces menor que el del agua.

Como se puede ver en la mayoría de los casos el espectro de la resta queda bastante ruidoso. Sin embargo se sabe que el efecto de la absorción del agua es mínimo en el rango que va desde 4900 cm^{-1} hasta 4100 cm^{-1} ; esto se puede ver en las gráficas presentadas anteriormente [2].

Además se sabe que en esa región existen 3 zonas de absorción que son característicos para la glucosa. Es por ello que se escoge la región anteriormente mencionada y donde la información que nos da el espectro sobre la concentración de glucosa es más importante.

4.2 Etapa de pre-procesamiento

En la figura 4.8 se muestran las señales de glucosa de diversos niveles de concentraciones y en las cuales se observan las variaciones de la línea de base así como el ruido de alta frecuencia que las caracteriza.

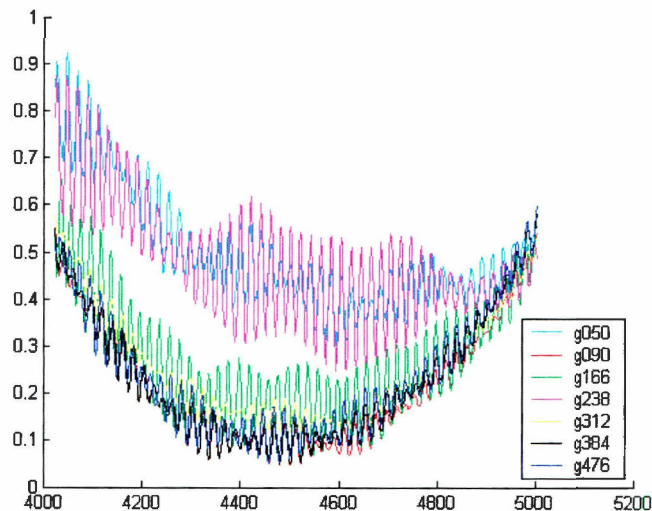


Figura 4.8 Señales de glucosa sin filtrar

4.2.1 Filtros FIR

Ventana Hamming

Se sacaron las respuestas en frecuencia de la señal g050 y se tomaron como referencia para sacar las aproximaciones de las frecuencias de corte. Estas respuestas en frecuencias se calcularon con las funciones *freqz* y *fft* de Matlab.

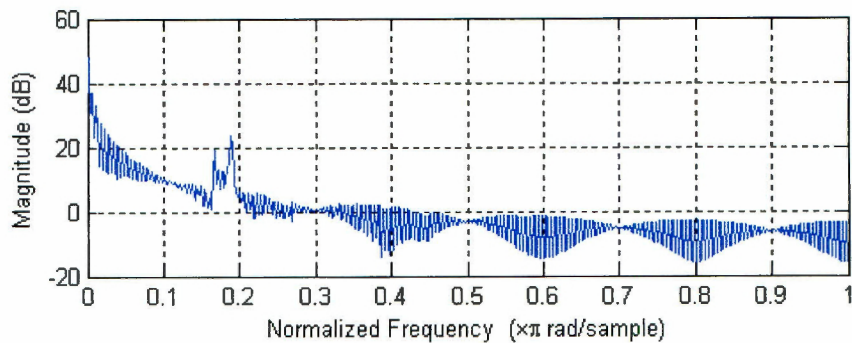


Figura 4.9 Respuesta en frecuencia de la señal original g050 (Decibeles)

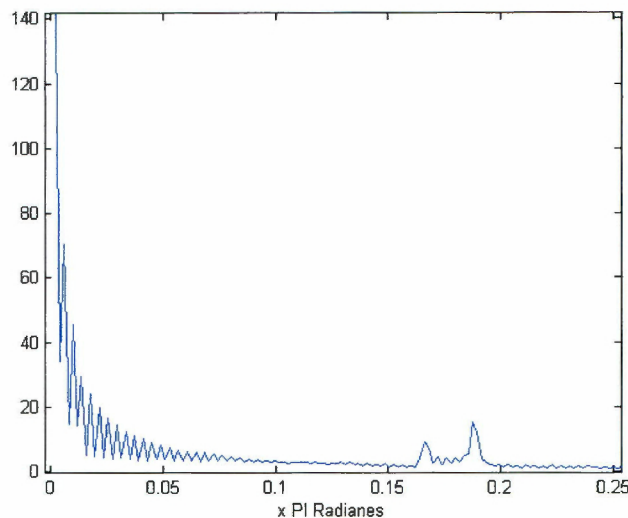


Figura 4.10 Respuesta en frecuencia

Se puede observar en las figuras 4.9 y 4.10 que una frecuencia de corte aproximada que nos permita eliminar las altas frecuencias se puede escoger en 0.1π rad.

La frecuencia de corte baja se escoge de forma tal que elimine las bajas frecuencias que componen las variaciones de la línea de base.

Al no tener una referencia que nos sirviera como base para establecer nuestras frecuencias, estas se establecieron en base al análisis de Fourier tomando aproximaciones y variando los valores de las frecuencias.

Se utilizaron las formulas 2.20 para determinar el orden mínimo del filtro y las frecuencias de corte apropiadas dados los límites de las bandas de paso y las bandas de rechazo. Dichos límites son para la banda de paso y banda de rechazo bajas 0.02π y 0.05π respectivamente y los límites altos se escogieron en 0.12π y 0.15π . El orden resultante fue de 330 y las frecuencias de corte se calculan como: $\omega_{ci} = 0.05\pi - 0.015\pi = 0.035\pi$ y $\omega_{cs} = 0.12\pi + 0.015\pi = 0.135\pi$.

Las frecuencias mencionadas anteriormente resultaron en un filtrado relativamente bueno sin embargo seguía presentando componentes de alta frecuencia; una vez identificadas

las componentes se tomó como una nueva referencia de corte alta la frecuencia 0.04π (límites paso banda y rechaza banda de 0.03π y 0.05π respectivamente) y corte baja 0.01π , pero se mantuvo constante el orden del filtro.

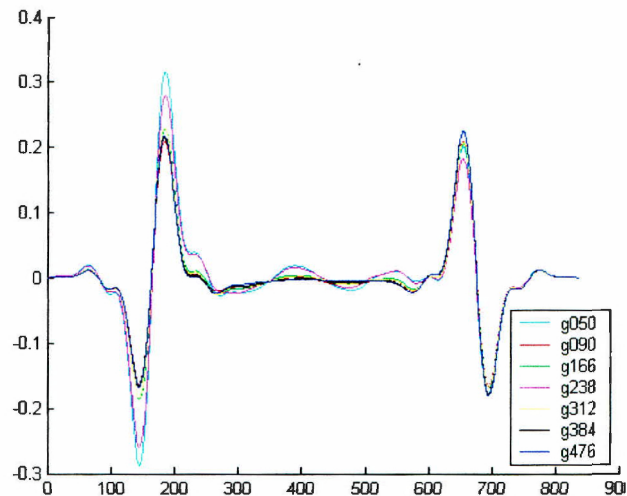


Figura 4.11 Señales filtradas

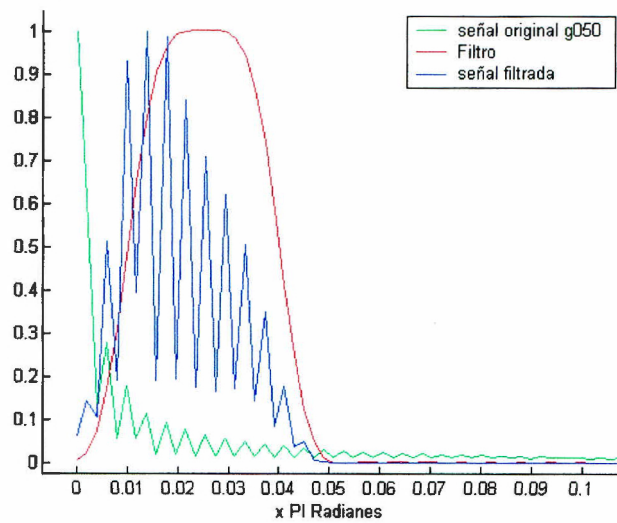


Figura 4.12 Respuesta en frecuencia normalizada

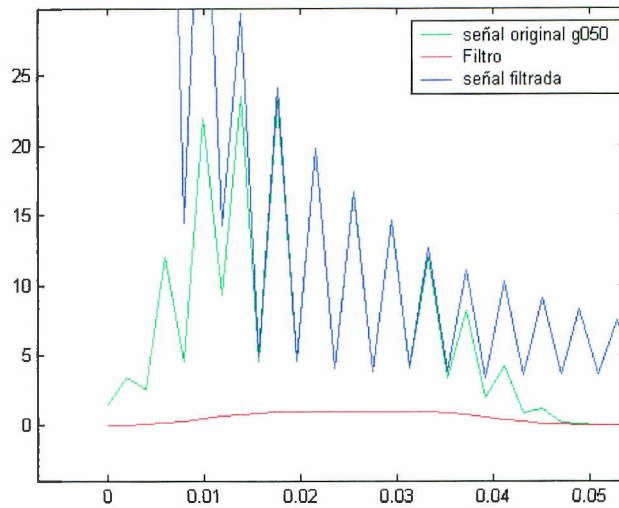


Figura 4.13 Respuesta en frecuencia, sin normalizar.

Se volvió a realizar el filtrado con una nueva frecuencia de corte de 0.03π (límites paso banda y rechaza banda de 0.02π y 0.04π respectivamente) y los resultados fueron muy similares tanto en respuesta en frecuencia como en “tiempo” y se muestran en la imagen 4.14.

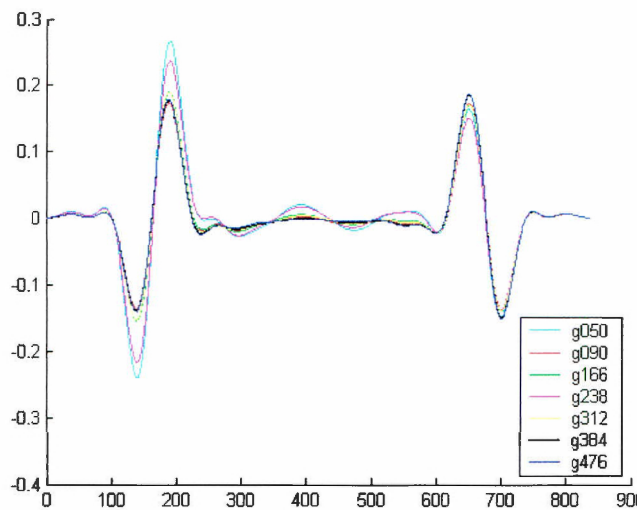


Figura 4.14 Señales filtradas

Ventana Kaiser:

Como se sabe, estos filtros tienen la ventaja de que pueden variar su parámetro beta que modifica la atenuación mínima en la banda de rechazo y la forma de la ventana. Se escogió un valor de rizo arbitrario $\delta_r = 0.03$ y se convirtió a una atenuación en decibeles A, $A = -20 \log_{10}(0.03) = 30.4575$. Sustituimos la atenuación en decibeles en la fórmula 2.22 y se obtiene el valor de $\beta = 2.18089$. El orden del filtro se determinó con la fórmula

2.23 usando los valores límites de paso banda alto y rechaza banda alto de 0.03π y 0.05π respectivamente, $M = 156.73 \approx 158$. Los resultados fueron los siguientes.

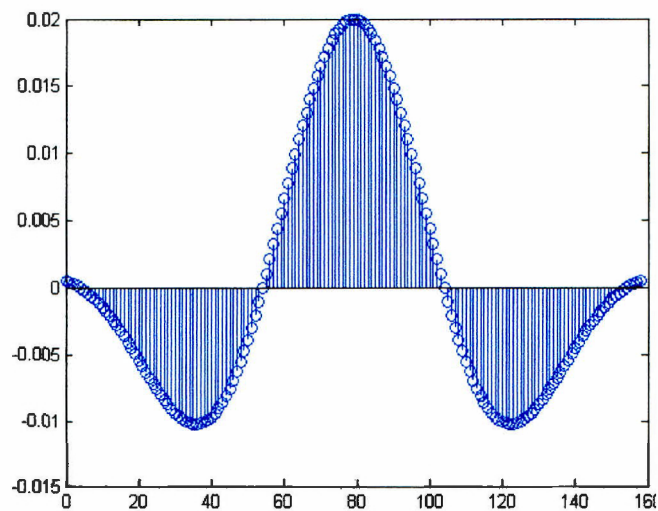


Figura 4.15 Respuesta al impulso del filtro, Beta = 2.1809

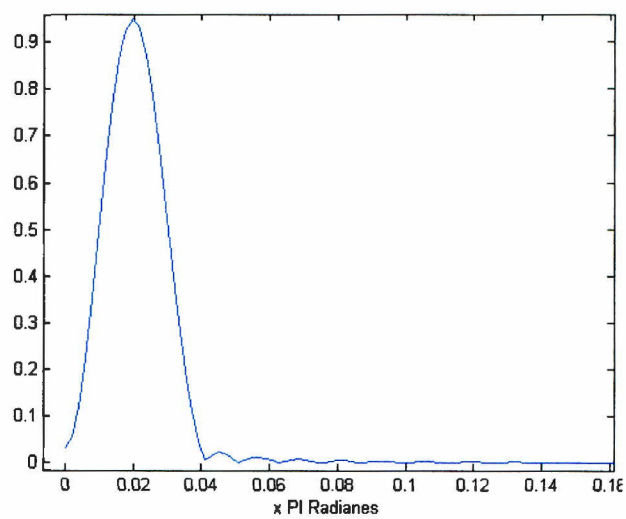


Figura 4.16 Respuesta en frecuencia del filtro

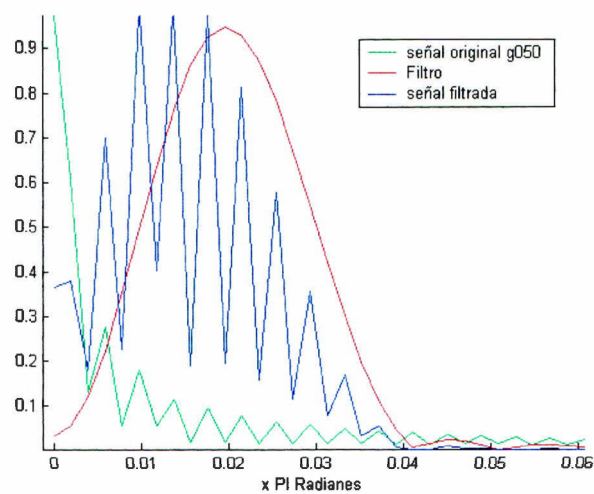


Figura 4.17 Respuesta en frecuencia normalizada

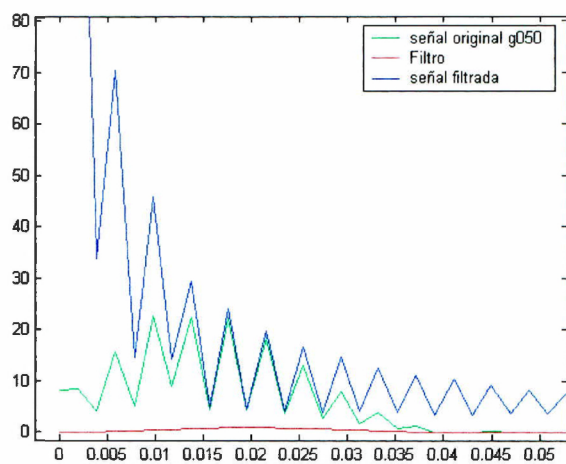


Figura 4.18 Respuesta en frecuencia sin normalizar

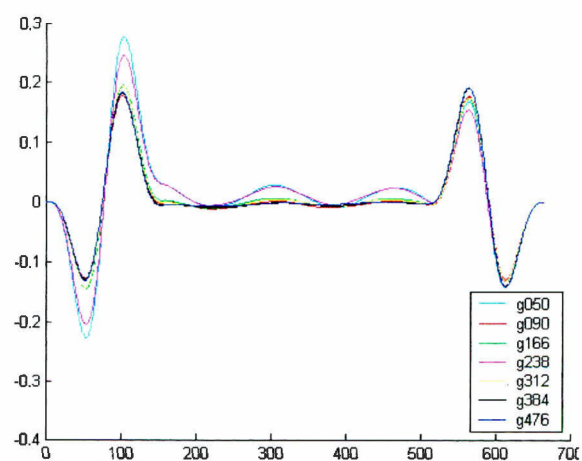


Figura 4.19 Señales filtradas

Se observa en la figura 4.19 que existen aun componentes de baja frecuencia que están afectando a las señales g050 y g166 por lo que se muestran con un offset, para solucionarlo se recorrió la frecuencia de corte baja a 0.011π , el resultado se presenta en las imágenes 4.20 y 4.21:

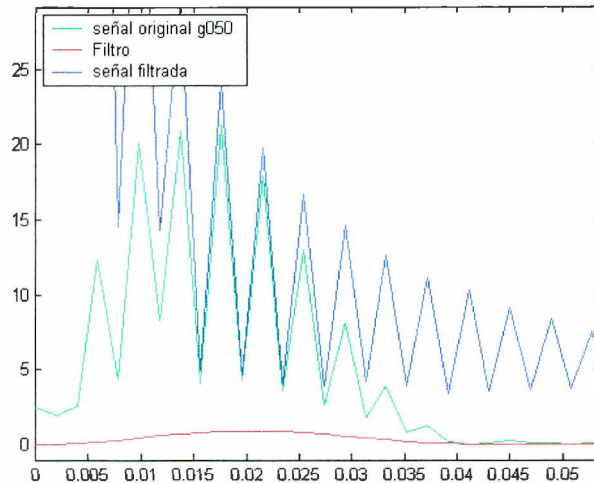


Figura 4.20 Respuesta en frecuencia

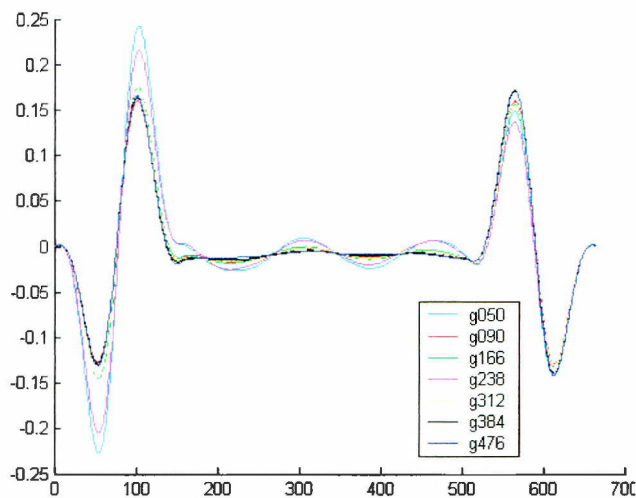


Figura 4.21 Señales filtradas

Como se ve en las figuras anteriores, los lóbulos laterales del filtro ayudaron a acentuar curvas de las señales filtradas que no se presentaban cuando se eliminaban totalmente las componentes de frecuencias existentes después de la frecuencia de corte.

4.2.2 Filtros IIR

Dados los resultados anteriores se prosiguió a realizar el filtrado mediante filtros IIR. En este tipo de filtro contábamos con la referencia de un trabajo ya realizado sobre la determinación de glucosa en matrices de agua [2]. En dicho documento se establecía una frecuencia central y un ancho de banda óptimos del filtro para determinados números de

factores PLS sin embargo al implementar los valores mencionados, filtro butterworth orden 4, frecuencia central 0.0625 y mitad del ancho de banda 0.0609, se obtuvieron resultados poco satisfactorios que se muestran a continuación.

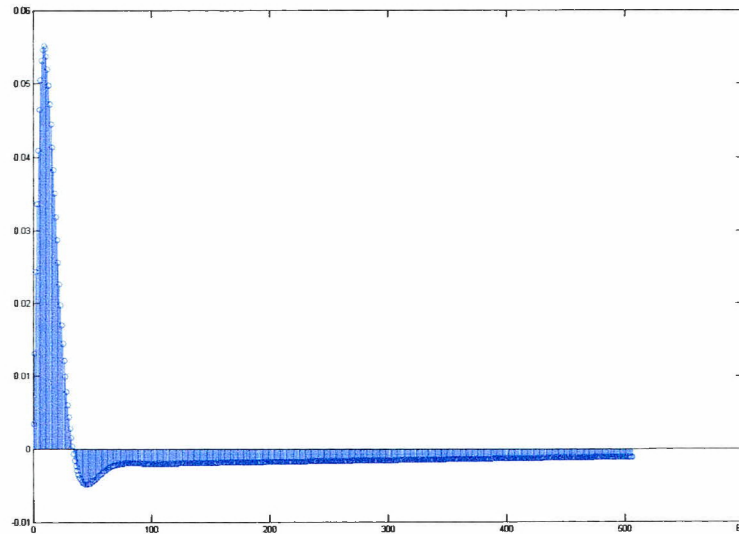


Figura 4.22 Respuesta al impulso del filtro referenciado

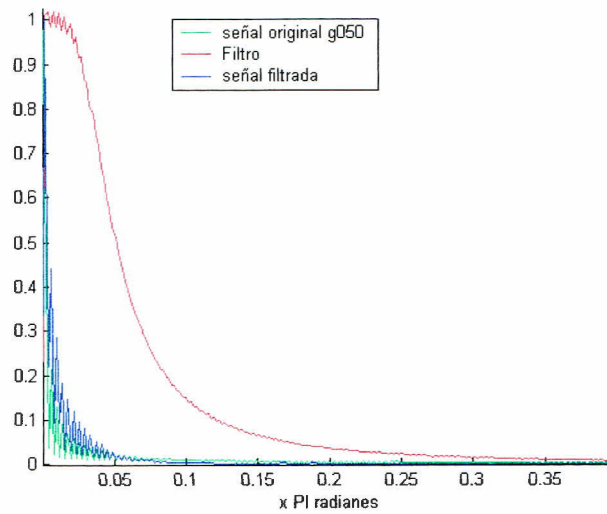


Figura 4.23 Respuestas en frecuencia, g050 original, filtro, señal filtrada

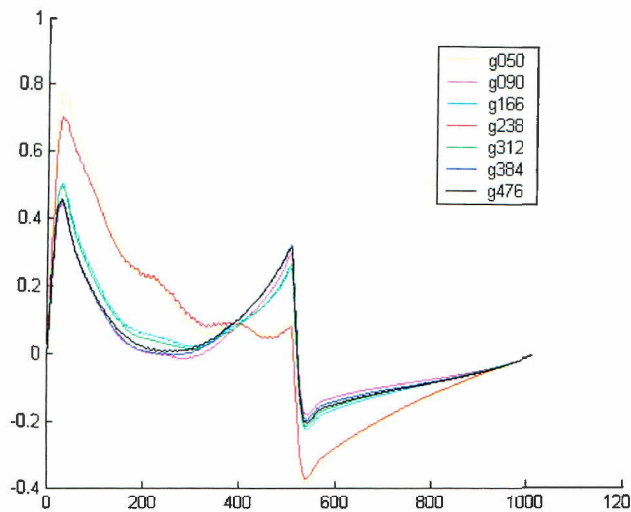


Figura 4.24 Señales filtradas

Es notorio mediante las imágenes anteriores que el filtro establecido en el documento mencionado no cumple con los requerimientos que debemos cubrir como son la eliminación de las componentes de DC y de alta frecuencia. Dados estos resultados se prosiguió a buscar unas frecuencias de corte más óptimas para nuestras necesidades.

Dado que ya se tenía la referencia de las frecuencias establecidas en los filtros FIR, se tomaron estas y se definió el filtro IIR de manera experimental. Se estableció un orden del filtro 6 y unas frecuencias de corte baja de 0.01π , frecuencia de corte alta de 0.025π y se obtuvieron los siguientes resultados:

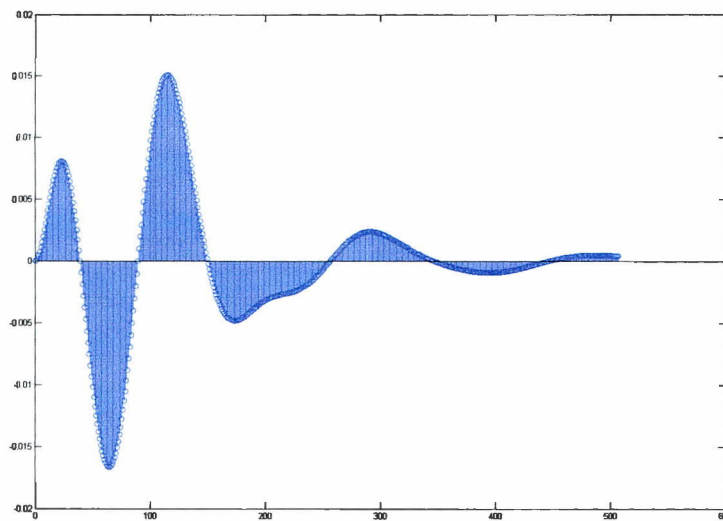


Figura 4.25 Respuesta al impulso

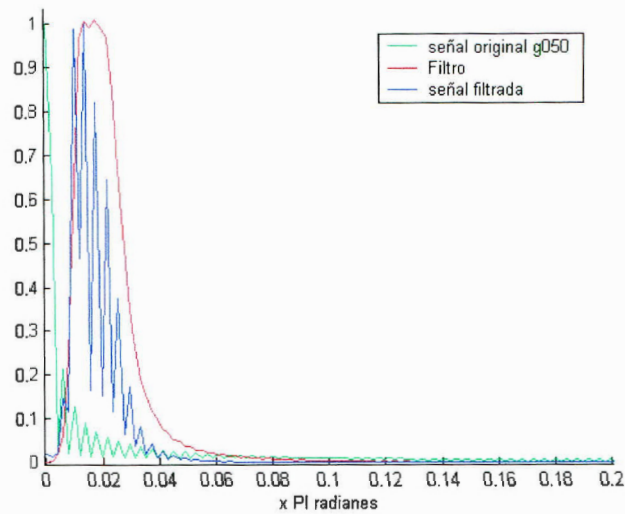


Figura 4.26 Respuesta en frecuencia normalizadas

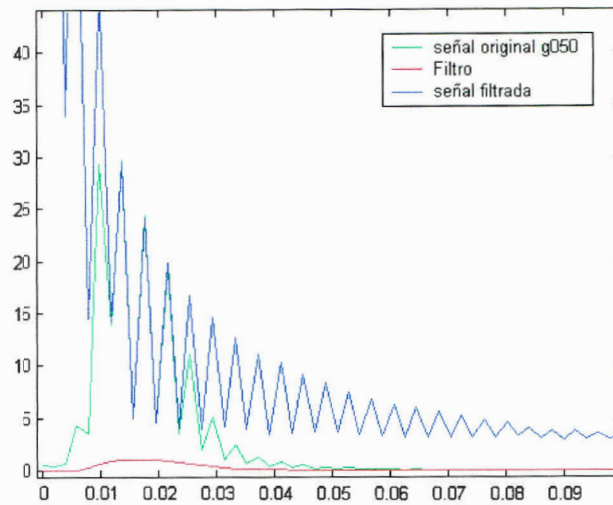


Figura 4.27 Respuesta en frecuencia normalizadas

Como se observa en la figura 4.27, este tipo de filtrado nos ayuda a eliminar la componente de más baja frecuencia totalmente mientras toma en su totalidad las componentes que nos interesan y atenúa las que nos pueden ocasionar ruido de alta frecuencia. Los resultados de este filtrado se observan en la figura 4.28.

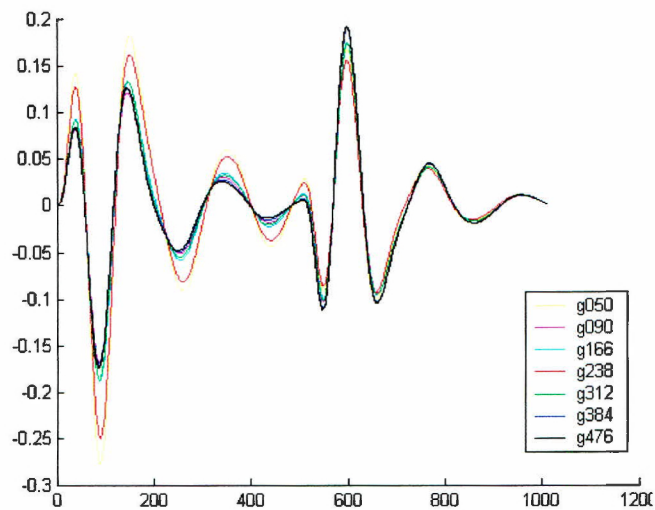


Figura 4.28 Señales filtrada

Dado que las componentes pequeñas que se están filtrando podrían contener también información útil se estableció la frecuencia de corte baja en 0.01π y la frecuencia de corte alta se recorrió a la derecha hasta establecer la frecuencia de corte 0.0465π entregándonos al parecer mejores resultados.

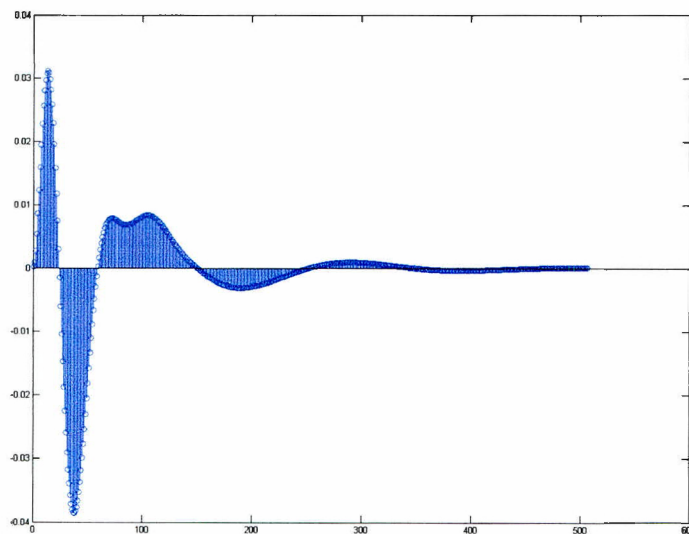


Figura 4.29 Respuesta al impulso del filtro

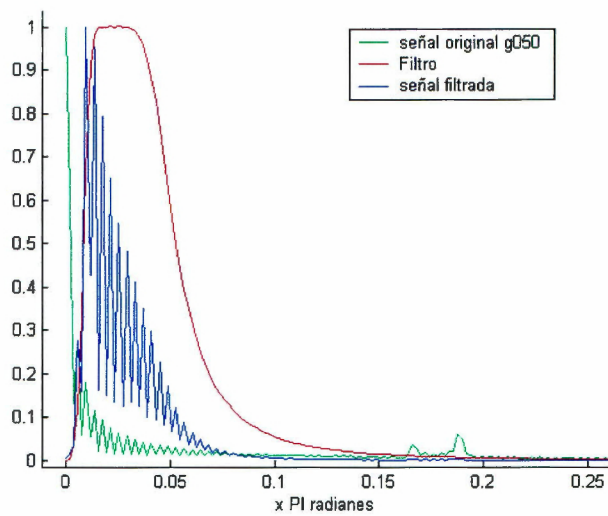


Figure 4.30 Respuesta en frecuencia normalizada

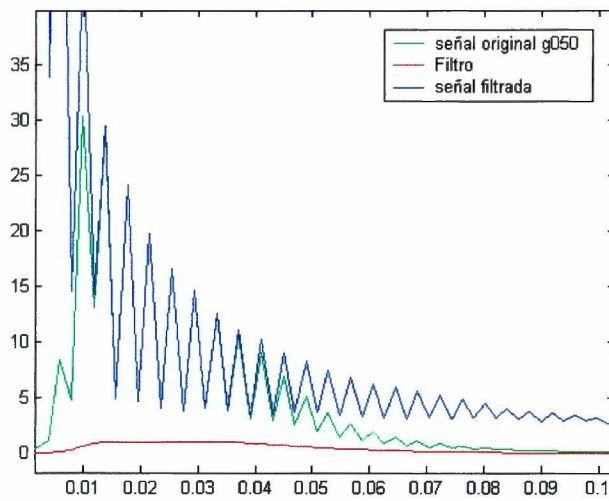


Figura 4.31 Respuesta en frecuencia sin normalizar

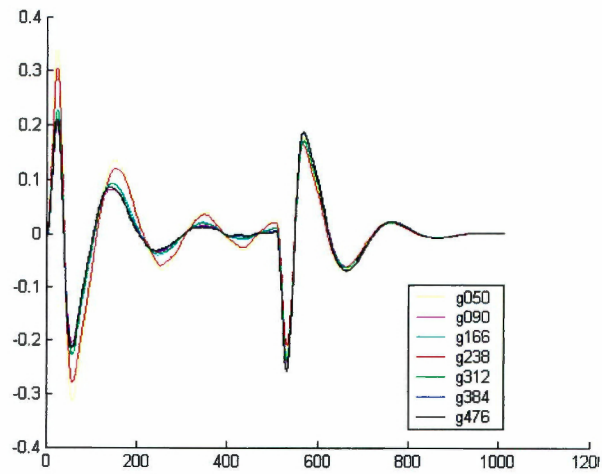


Figura 4.32 Señales filtradas.

4.2.3 Wavelets

Con el fin de evaluar el comportamiento de la etapa de decisión y facilitar el aprendizaje de ésta, se generaron diversas alternativas de filtros wavelet. Se utilizaron 4 valores para k_x , que da la selectividad del filtro ($k_x = 2, 3, 4$ y 5). La elección de F_0 se realiza en función del rango espectral (entre $10,000 \text{ cm}^{-1}$ y $4,000 \text{ cm}^{-1}$), que para efectos de simulación la variación de esta no afecta pues se ajusta a la separación entre muestras a $x = 1/2F_0$. Dada $k_x = 2, 3, 4$ y 5 , ajustamos el orden del filtro y elegimos las siguientes combinaciones de parámetros:

K	N
2	80
3	150
4	240
5	340

Tabla 4.1 Combinación de parámetros usada.

La Figura 4.33 muestra el wavelet obtenido para $k_x = 2$, $N = 40$.

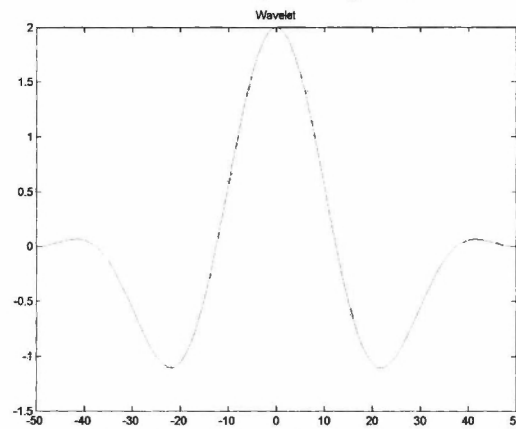


Figura 4.33 Wavelet; $k_x=2$.

Para validar la homogeneidad de los resultados con los distintos wavelets, se hicieron pruebas con siete concentraciones distribuidas entre las treinta y cinco de nuestro grupo de datos (50 mg/dl, 90 mg/dl, 166 mg/dl, 238 mg/dl, 312 mg/dl, 384 mg/dl y 476 mg/dl Figura 4.35), aunque los resultados expuestos en este trabajo se basan principalmente en la concentración de 50 mg/dl.

Para $k_x=2$, $N=40$ se obtuvieron los siguientes resultados.

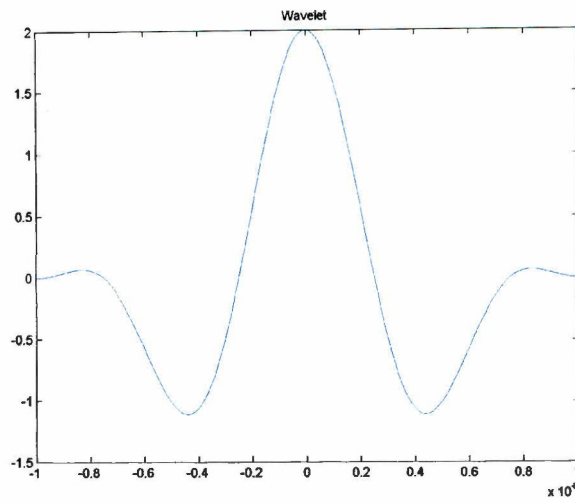


Figura 4.35 Wavelet; $k_x=2$, $N=80$.

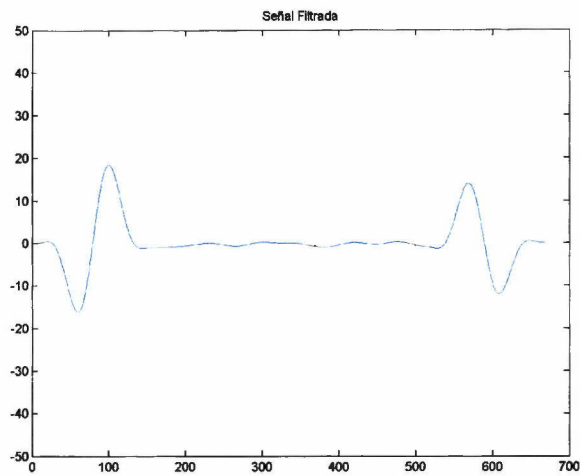


Figura 4.36 Señal filtrada; $k_x=2$, $N=80$.

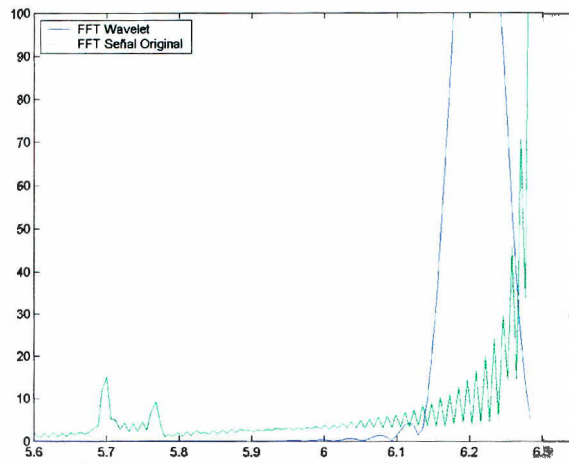


Figura 4.37 Respuesta en frecuencia; $k_x=2$, $N=80$.

Para $k_x=3$, $N=150$.

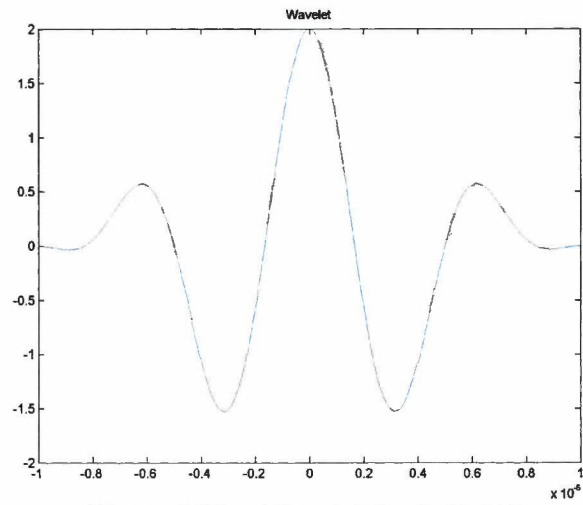


Figura 4.38 Wavelet; $k_x=3$, $N=150$.

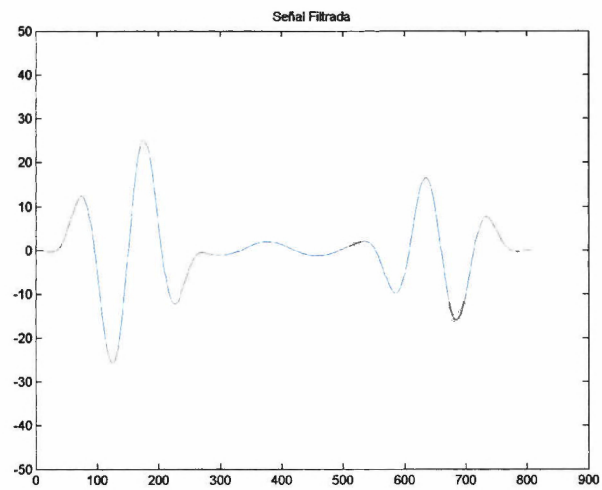


Figura 4.39 Señal filtrada; $k_x=3$, $N=150$.

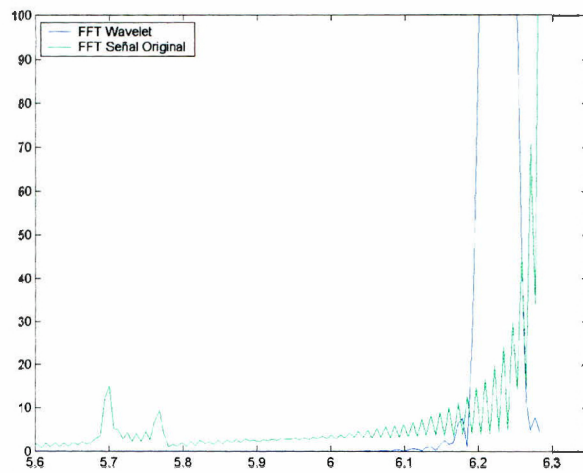


Figura 4.40 Respuesta en frecuencia; $k_x=3$, $N=150$.

Para $k_x = 4$, $N = 240$.

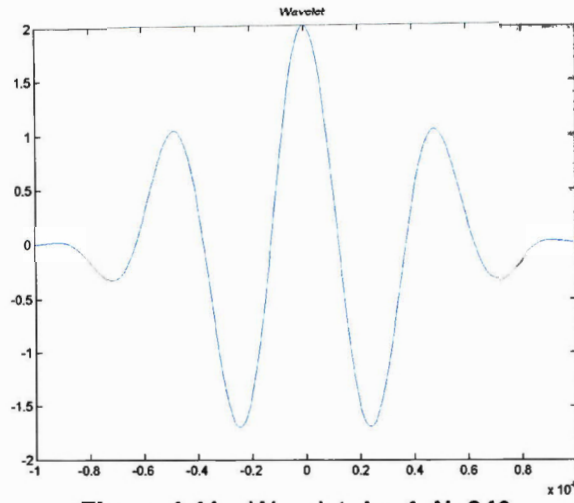


Figura 4.41 Wavelet; $k_x=4$, $N=240$.

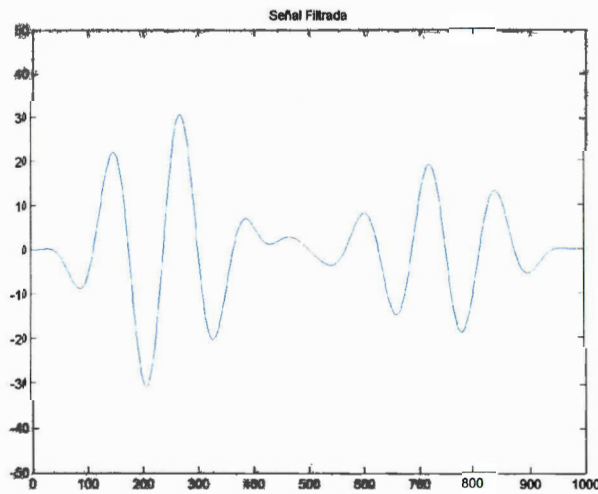


Figura 4.42 Señal filtrada; $k_x=4$, $N=240$.

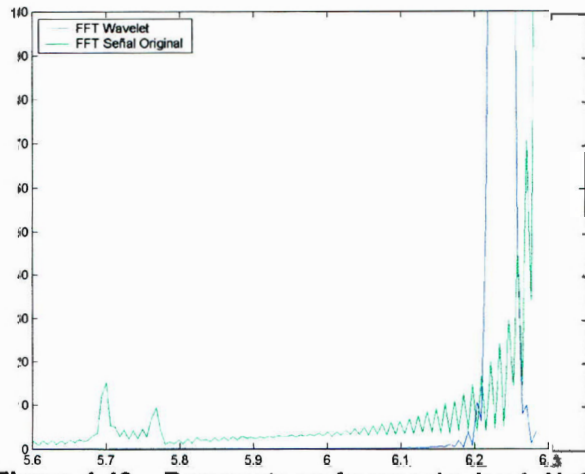


Figura 4.43 Respuesta en frecuencia; $k_x=4$, $N=240$.

Para $k_x=5$, $N=340$.

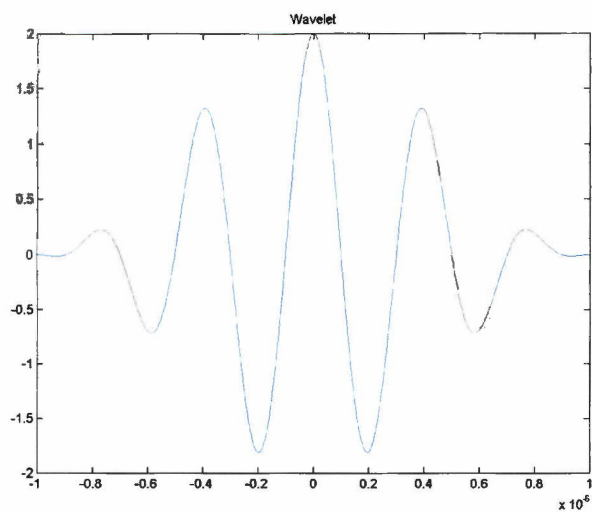


Figura 4.44 Wavelet; $k_x=5$, $N=340$.

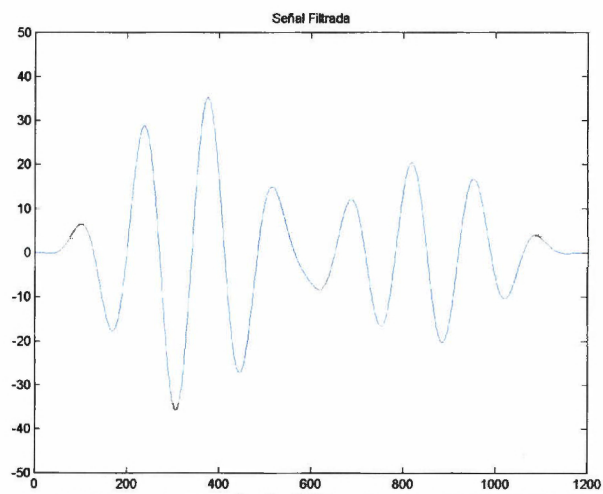


Figura 4.45 Señal filtrada; $k_x=5$, $N=340$.

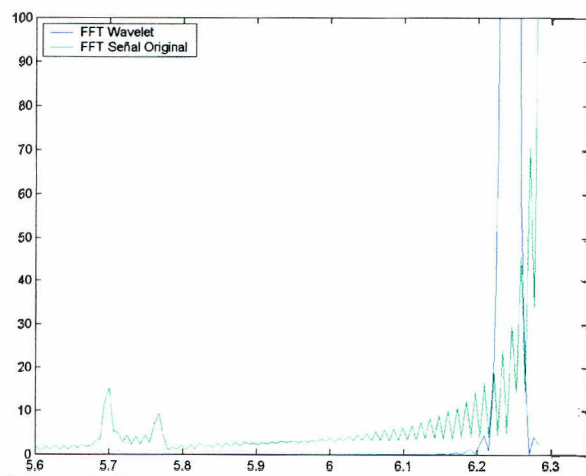


Figura 4.46 Respuesta en frecuencia; $k_x=5$, $N=340$.

De la ecuación (2.29) y de las figuras anteriores observamos que entre mayor es el índice de selectividad k_x , se reduce el ancho de banda analizado, y se recorre la frecuencia central, corrimiento que debe ser compensado con el incremento en N. De manera similar entre mayor sea el orden del filtro, se incrementa el ancho de banda y esto debe ser compensado con un k_x más grande. También vemos que incrementar la selectividad del filtro k_x aumenta la dinámica del wavelet, efecto que se refleja en la señal filtrada.

4.3 Etapa de caracterización

4.3.1 Coeficientes autorregresivos

Realizando pruebas con las señales obtenidas del filtrado realizamos el modelado de señales por AR para obtener el espectro estimado de potencia de estas señales.

Los resultados siguientes se obtuvieron de la señal g476 filtrada a través de wavelets que se muestra en la figura 4.47

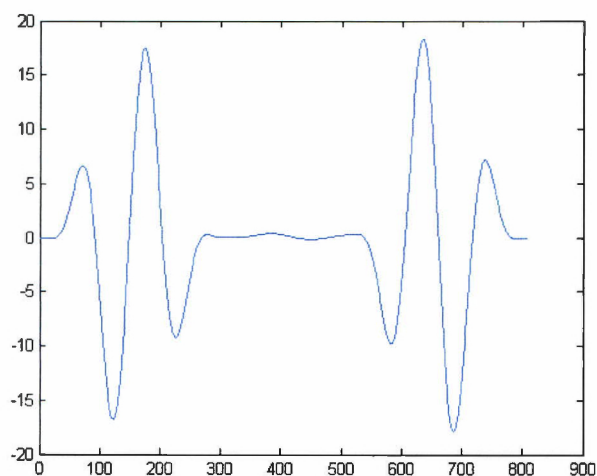


Figura 4.47 g476 filtrada mediante wavelets, $K_x=4$, $N=240$

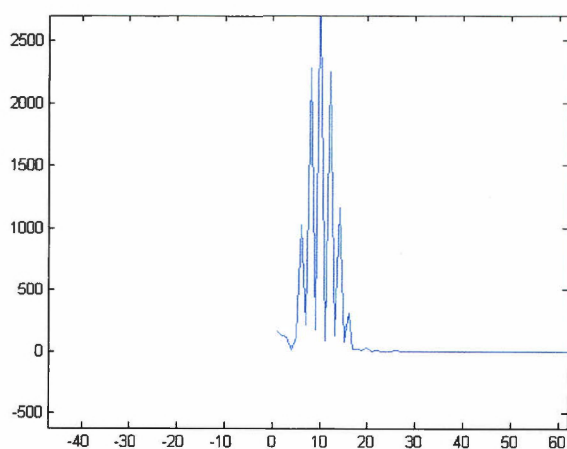


Figura 4.48 FFT de 1024 puntos de la señal filtrada

Se muestra en la figura 4.49 el resultado de nuestra aproximación al espectro de potencia con FFT a través de modelado Autorregresivo por el método de Yule-Walker.

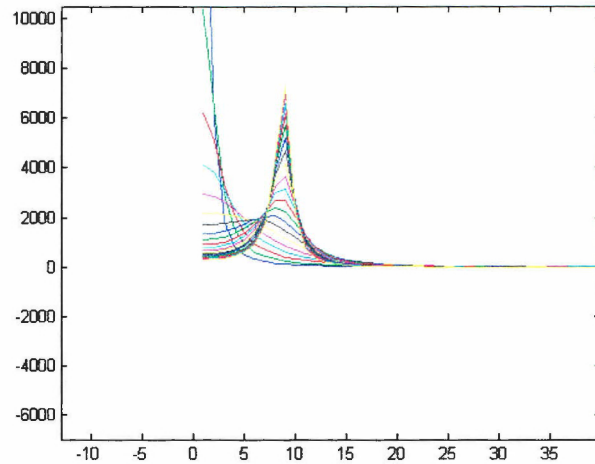


Figura 4.49 Aproximación del espectro de potencia

Cómo se puede apreciar en la gráfica, se notan diferentes trazos. Cada uno de estos trazos corresponde al espectro estimado de potencia de la señal con un número de coeficientes distinto. Para el modelado de esta señal se utilizaron desde uno hasta veinte coeficientes para realizarlo. Conforme se añaden coeficientes al procedimiento, el trazo de la señal se modifica de una señal suave al principio hacia una señal mucho más parecida al espectro de potencia original. Puede observarse que el espacio entre los trazos de los últimos coeficientes se encuentra mucho más cerrado que en los primeros. Es notoria la semejanza entre la FFT de la señal filtrada y la composición por AR. La siguiente gráfica muestra el comportamiento del AIC o Criterio de Información de Akaike.

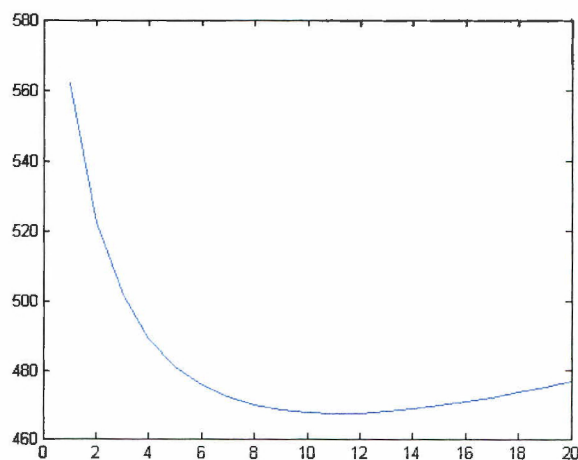


Figura 4.50 Criterio de Akaike

Al aplicar el Criterio de Información de Akaike pudimos observar que, efectivamente, el comportamiento de este criterio proporciona un valor mínimo al variar el número de coeficientes a utilizar. Para efectos prácticos de la implementación de estos parámetros en la etapa de Decisión, hemos decidido fijar el orden del sistema a sólo 20 coeficientes. Valor aceptado dentro del intervalo de valores mínimos. Esto aclara el porqué del menor espaciado entre los trazos con coeficientes mayores.

Para efectos prácticos del uso de estos parámetros dentro de la red neuronal, se desprende que el número de coeficientes para modelar el espectro de potencia estimado de las señal, no puede ser igual o mayor que el número de muestras tomadas para crearlo.

Entonces, queda restringida la aplicación del algoritmo empleado al cálculo de solo igual número de coeficientes que el tamaño de la señal a modelar.

Al observar el comportamiento de éste método al variar el número de coeficientes, notamos que para un número reducido de coeficientes, la señal se comporta suavemente. Al hacer uso de varios coeficientes la señal tiende a corregirse o parecerse más a la original. Sin embargo, un número de coeficientes muy grande puede desfigurarse el espectro de la señal.

4.3.2 Diagramas tiempo-frecuencia.

Las muestras originales antes del filtrado despliegan los siguientes diagramas tiempo – frecuencia (STFT); se muestran los resultado para 2 de la señales únicamente.

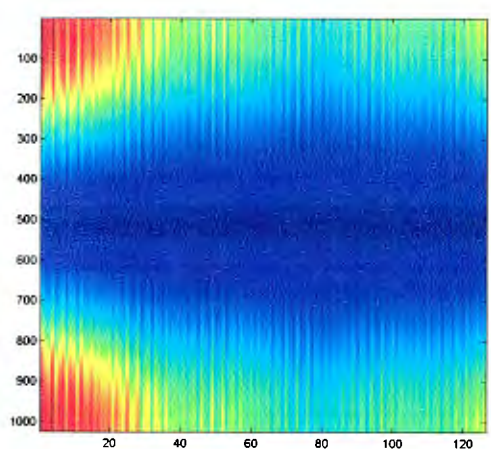


Figura 4.51 Diagrama tiempo–frecuencia g050

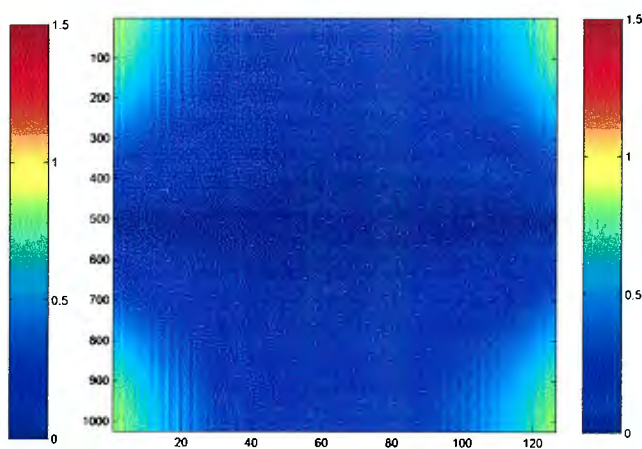


Figura 4.52 Diagrama tiempo–frecuencia g090

Los diagramas tiempo – frecuencia (STFT) para las señales filtradas se muestran a continuación.

Filtrado IIR

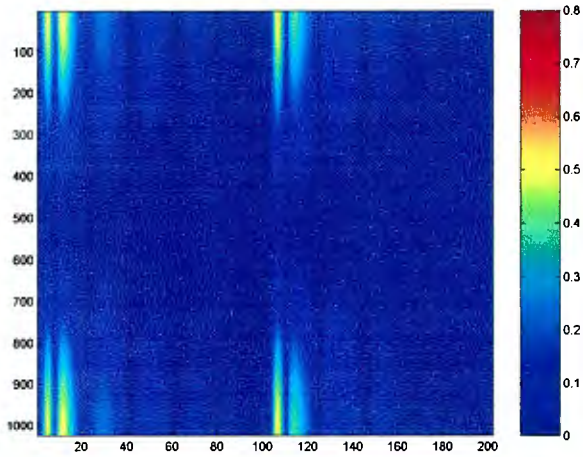


Figura 4.53 Diagrama tiempo - frecuencia g050 filtrada, IIR orden 6

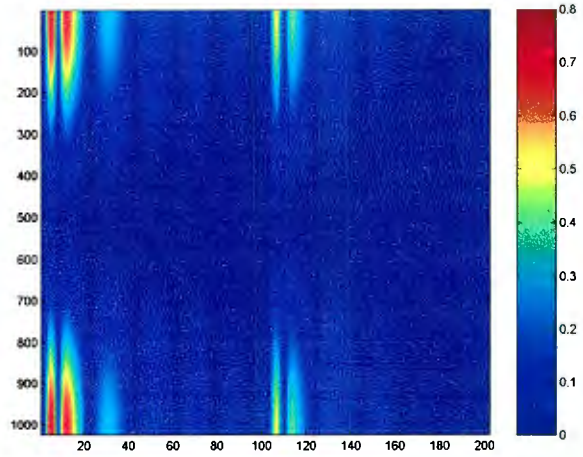


Figura 4.54 Diagrama tiempo - frecuencia g090 filtrada, IIR orden 6

Filtrado FIR, ventana Hamming

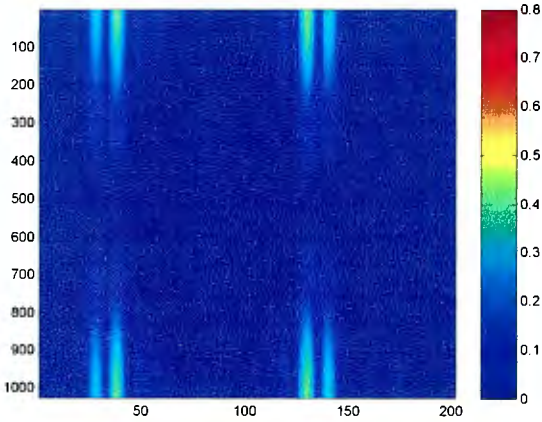


Figura 4.55 Diagrama tiempo - frecuencia g050 filtrada

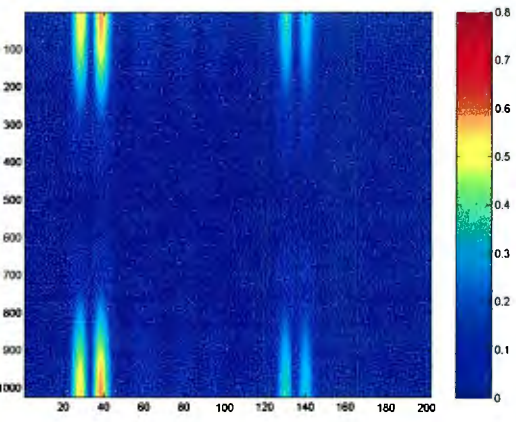


Figura 4.56 Diagrama tiempo - frecuencia g090 filtrada

Filtros FIR, ventana Kaiser

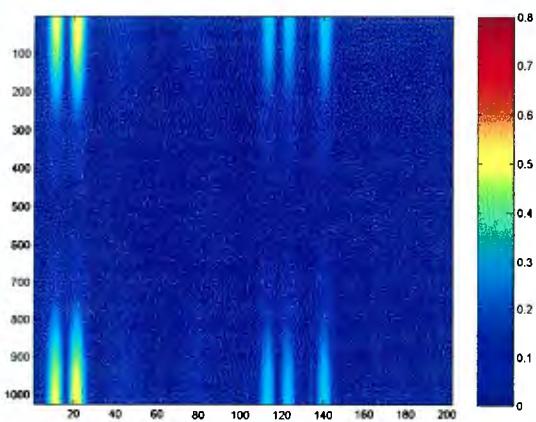


Figura 4.57 Diagrama tiempo - frecuencia g050 filtrada

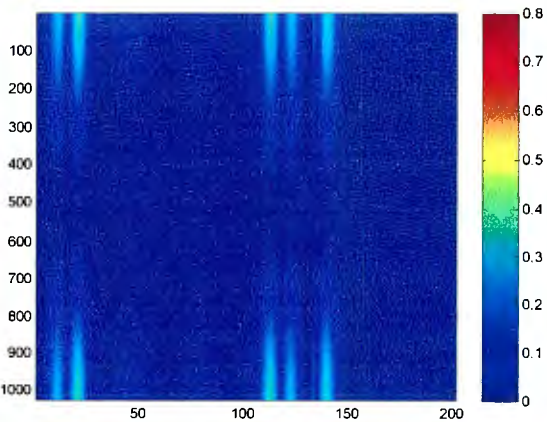


Figura 4.58 Diagrama tiempo - frecuencia g090 filtrada

Wavelets

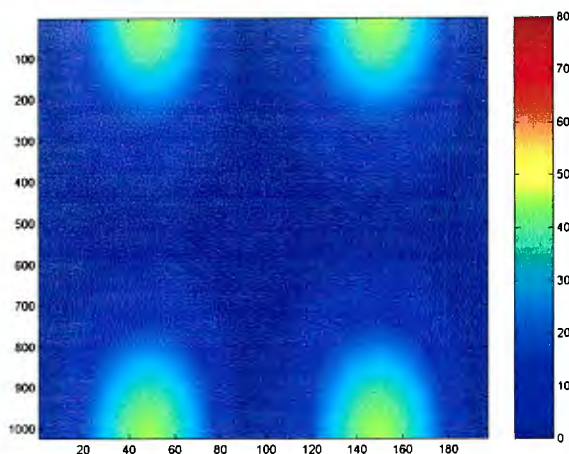


Figura 4.59 Diagrama tiempo - frecuencia g090 filtrada; $Kx=4$, $N=240$.

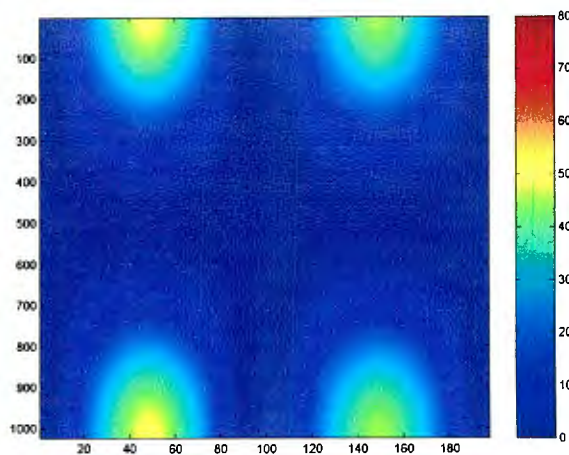


Figura 4.60 Diagrama tiempo - frecuencia g166 filtrada; $Kx=4$, $N=240$.

4.4 Etapa de decisión

4.4.1 Lógica difusa y coeficientes autorregresivos

Como se había mencionado en la etapa de caracterización por coeficientes autorregresivos; éstos nos servirán para encontrar parámetros y reducir la cantidad de elementos que formarán parte de nuestra etapa de decisión.

Para cada concentración entonces, se obtuvieron veinte coeficientes. Estos coeficientes son característicos de cada concentración dado que con ellos mismos se puede reconstruir el espectro de potencia estimado.

La idea que se tuvo de utilizar lógica difusa con los coeficientes autorregresivos es la siguiente:

Primeramente se define un dominio donde se encuentran todos los coeficientes (el primero de los veinte, por ejemplo) de todas y cada una de las concentraciones. En este dominio se encuentran varias reglas de membresía, una por cada concentración. En este caso todas las funciones de membresía son de la misma forma. Cada una de estas funciones de membresía tiene como valor central al valor del coeficiente; de esta manera el valor máximo de la función de membresía coincide con el valor del coeficiente y por lo mismo ahí se da la máxima pertenencia.

Así existen diferentes dominios como elementos hay por concentración. Cada dominio con igual número de funciones de membresía correspondientes a todas las concentraciones. Lo anterior corresponde a los segmentos de Entrada y Entrada difusa a través de Funciones de membresía.

Cada función de membresía tiene asociada una variable lingüística, en este caso el nombre de la variable es el nombre de la concentración.

Correspondiente al segmento de evaluación de las reglas es la construcción de las mismas. Cada concentración queda definida como una combinación puntual de todos sus coeficientes. El operador AND es el que define las reglas para la obtención de cada concentración.

En los segmentos de Salida difusa y Valor real se tiene también un dominio en el que se tienen los valores de las concentraciones y sus respectivas funciones de membresía. Para estas funciones, sus valores centrales son los correspondientes a cada concentración (por ejemplo 50, 55, 125 etc.) y que son los resultados esperados.

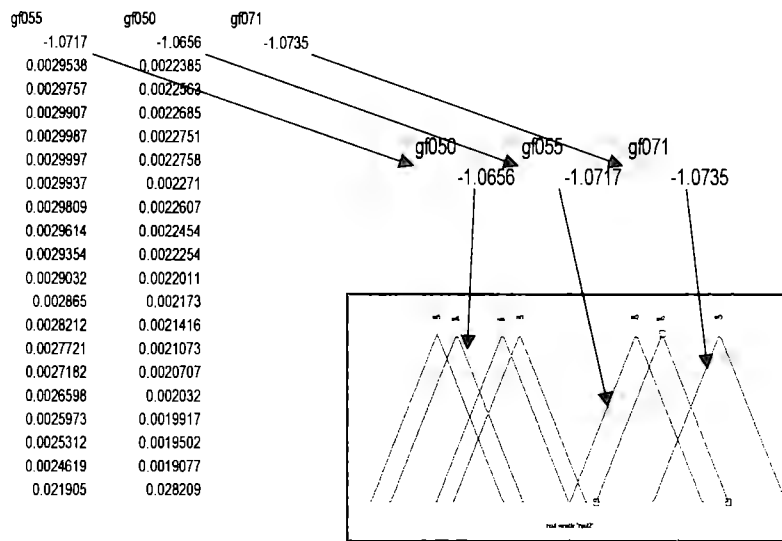


Fig.4.61 Asignación de coeficientes por cada función de membresía.

Una de las ventajas esperadas por el uso de lógica difusa es el hecho de poder reconocer concentraciones cuyos coeficientes no coinciden con los valores centrales de las funciones de membresía pero que tienen un cierto grado de pertenencia a éstas.

Por la naturaleza de las señales filtradas, los coeficientes autorregresivos tienen el mismo comportamiento o difieren muy poco entre sí. Esto dificultó el poder reconocer las concentraciones adecuadamente debido a que hubo muchas funciones de membresía que se activaban simultáneamente. Observamos que conforme se aumentaban elementos por concentración o las mismas concentraciones, hacer la sintonía de las funciones para reducir el número de funciones activadas se volvía más complicado y no entregaba buenos resultados.

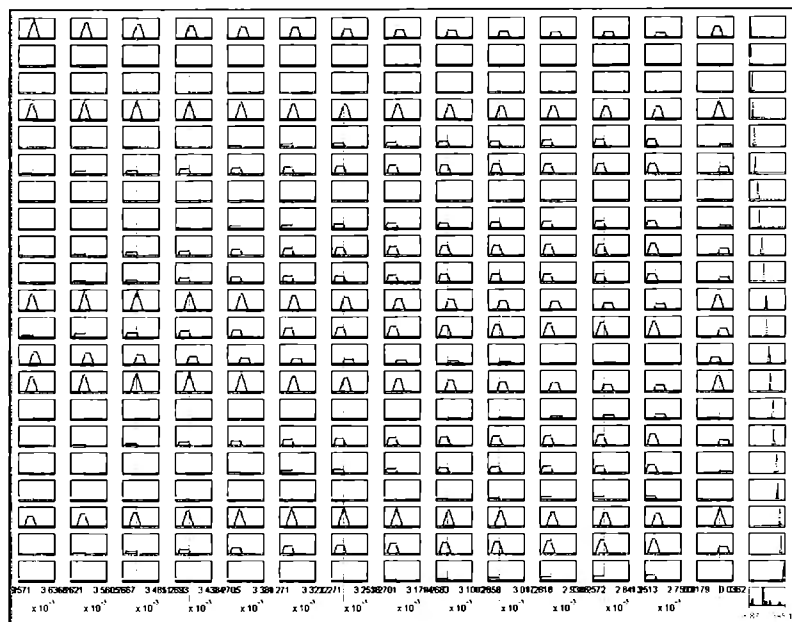


Fig. 4.62 Activación de las funciones de membresía

No es extraño ver que haya más de una función de membresía activada, por el contrario, es una de las fuerzas de la lógica difusa porque ello permite generalizar otros resultados; sin embargo, fueron muchas las que se activaban, cada una de estas activaciones corresponde a valores de concentraciones en ocasiones muy distintos por lo que los resultados se alejaban mucho del valor real.

4.4.2 ANFIS y coeficientes autorregresivos

Para el entrenamiento con sistema ANFIS se implementaron veinte coeficientes por concentración. Lamentablemente y debido a la falta de capacidad de nuestras computadoras, este sistema no pudo realizarse adecuadamente.

Decidimos realizar algunas pruebas con solo dos concentraciones de cinco elementos cada una. La convergencia del error de entrenamiento se alcanza muy rápidamente; sus resultados son bastante buenos. Las concentraciones entrenadas se reconocen con exactitud.

Hay que mencionar que las concentraciones no entrenadas no pudieron ser reconocidas. Sabemos que el entrenamiento con ANFIS es especialmente bueno para hacer generalizaciones, sin embargo, consideramos que estos resultados no fueron buenos debido a que la cantidad de elementos por concentración fue de solo cinco y que son realmente pocos para tener mejores resultados.

La disminución de elementos se debe a que el tiempo de procesamiento es realmente grande. Observamos la estructura generada y es realmente elaborada tan solo para cinco elementos por concentración.

4.4.3 Redes neuronales artificiales

Se propusieron 3 estructuras diferentes de redes neuronales con los siguientes números de neuronas: 50 neuronas capa 1, 10 neuronas capa 2, una neurona capa 3; 50 neuronas capa 1, 25 neuronas capa 2, 10 neuronas capa 3 y una neurona capa 4.

Los vectores de entrada para esta red y para todas las demás fueron treinta concentraciones de las treinta y cinco existentes. Cada una de estas concentraciones contiene 50 elementos seleccionados mediante STFT.

La idea de tomar solo treinta concentraciones se debe a que; con las cinco restantes, se harán pruebas sobre reconocimiento y generalización de concentraciones no entrenadas. Las concentraciones que fueron suprimidas para este entrenamiento fueron 83, 166, 250, 333, 454.

Caso 1

Las concentraciones suministradas en el primer caso fueron las correspondientes al filtrado tipo IIR de orden 4, realizadas en el proyecto anterior. En el presente documento se muestran únicamente los mejores resultados de las estructuras mencionadas.

Para el caso de las señales filtradas con los filtros IIR de orden 4 los mejores resultados se obtuvieron de las estructuras de 3 y 4 capas, para el caso de la red de la gran mayoría de la eficiencia de los resultados se encuentran por encima del 90%. Sólo dos de ellos muestran resultados menores con 84.8% y 89.8 % de efectividad; para el caso de 4 capas la eficiencia en los datos entrenados fue del 96.2% en promedio, sin embargo para todos los casos en las concentraciones entrenadas no se obtuvieron buenos resultados.

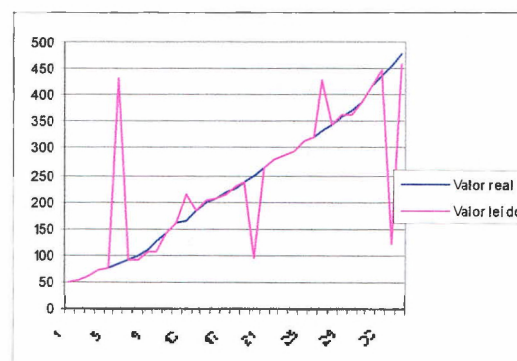


Figura 4.63. Estimaciones con red de 4 capas

Caso 2

Las redes neuronales entrenadas con las señales filtradas con filtros FIR fueron igualmente con 4 estructuras diferentes y 50 puntos de la señal en total y que se seleccionaron mediante el uso de STFT, sin embargo los resultados no fueron del todo

satisfactorios, sobretodo para el caso de los filtros FIR con ventana hamming los resultados presentan 85.56 % de eficiencia en promedio en el mejor de los casos para las concentraciones entrenadas, pero para los casos no entrenados este error aumenta considerablemente. De igual forma las señales filtradas con filtros FIR y ventana Kaiser tienen un comportamiento con una eficiencia de 87.62% para las señales entrenadas.

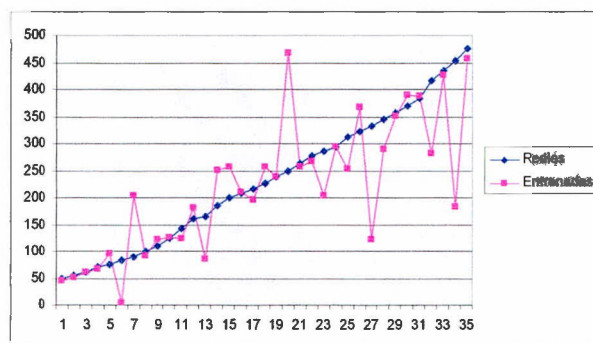


Figura 4.64. Estimaciones con filtrado FIR ventana Kaiser, red de 4 capas

Caso 3

En el caso de los filtros tipo IIR se aprecian mejores resultados, siguiendo el mismo procedimiento de múltiples entrenamientos y el mismo tipo de estructuras. En la grafica siguiente se observa el comportamiento de la detección de concentraciones

En este caso se tiene un porcentaje de error promedio de tan solo 0.005% en las concentraciones entrenadas pero este error promedio aumenta a 12.317% cuando se toma en cuenta la estimación de las concentraciones no entrenadas para verificar su capacidad de generalización, sin embargo si tomamos los casos más críticos en cuanto al porcentaje de error que son las concentraciones de 166 y 454 mg/dl, estos presentan errores de 259.6 % y 100.708%.

La figura 4.65 muestra la aproximación que tuvo la red mencionada a los valores reales de las concentraciones.

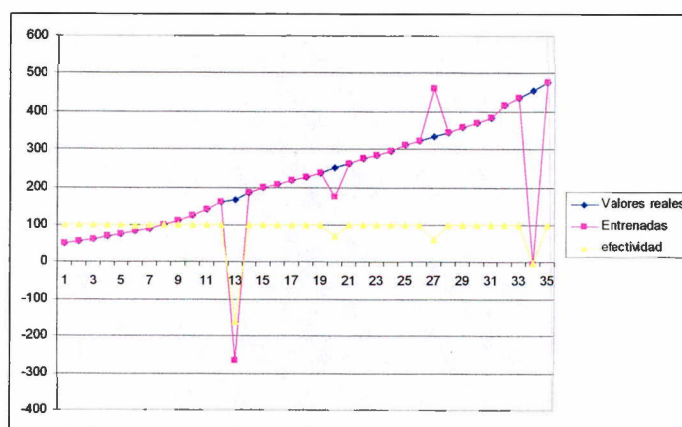


Figura 4.65. Estimación con filtrado IIR, red de 4 capas

Caso 4

Para este caso se entrenaron las redes con las mismas características de estructura mencionadas pero utilizando las concentraciones correspondientes al filtrado con wavelets $N = 80$, $K_x = 2$. Los mejores resultados se obtienen con la red neuronal de 3 capas. El caso de 3 capas se muestra en la grafica 4.66 donde se observa que la efectividad de esta red es muy buena para las concentraciones entrenadas y para las no entrenadas se genera error. En este caso la efectividad de tres de las concentraciones no entrenadas es aceptable, pero en el caso de la concentración 7 y 21 la aproximación es muy baja.

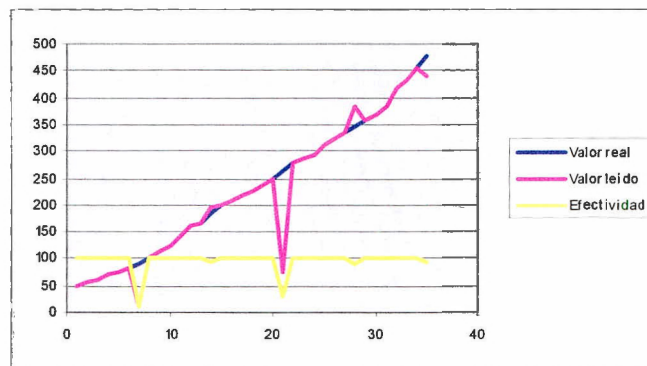


Figura 4.66. Estimaciones con red de 3 capas

Posterior a estos resultados se repitieron al igual que en los otros casos varias veces el entrenamiento y se llegó a los resultados mostrados en la figura 4.67 en donde todas arrojan una efectividad mayor del 90%. Es necesario mencionar que una de las concentraciones con menor efectividad en la simulación (94.1%) fue la concentración 416, que si fue entrenada. La efectividad global, es decir, la efectividad tomando en cuenta las concentraciones tanto entrenadas como las no entrenadas de la red es del 99.06%.

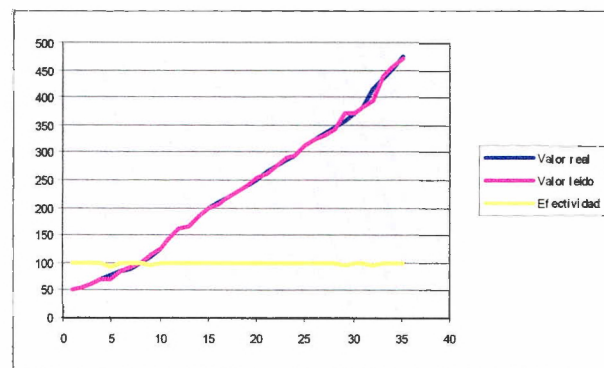


Figura 4.67. Estimaciones con red de 3 capas

Queda pendiente verificar que la red anterior realmente ha generalizado y pueda detectar las concentraciones con otros barridos de espectroscopia diferentes a los promediados en el presente trabajo, ya que al parecer estos últimos resultados son los ideales para implementarse en un dispositivo medico.

5. Conclusiones y perspectivas:

Se diseñaron distintos tipos de filtros que se utilizaron para la eliminación de las altas frecuencias y las variaciones en la línea de base que presentaban las señales obtenidas mediante la espectroscopia. Dado que en la primera etapa del proyecto no se tenía aun un sistema de decisión; no era posible verificar la validez de los resultados obtenidos de la etapa de pre-procesamiento y caracterización. Sin embargo, una vez definidos los métodos de decisión se logró comprobar la eficacia de los filtros realizados mostrando un mejor desempeño en las señales tratadas mediante wavelets y filtros tipo IIR.

Uno de los problemas que presentó la realización del presente proyecto fue la limitada capacidad de cómputo con la que se contaba, lo que no permitió el análisis o entrenamiento de las redes neuronales con las señales completas o incluso la exploración de métodos como los son los sistemas ANFIS, los cuales mejorarían notablemente los resultados obtenidos al combinar dos de los métodos más efectivos de inteligencia artificial para el aprendizaje y generalización. Este contratiempo nos orilló aun más a considerar métodos de caracterización que permitieran reducir el número de parámetros. Para lograr esto nos enfocamos en el uso de los coeficientes autorregresivos, así como el uso de los diagramas tiempo-frecuencia o STFT para solucionar dicho problema. En el caso de los coeficientes autorregresivos, a pesar de que se comprobó su eficacia con la estimación del espectro de una señal de audio, no nos arrojó resultados adecuados cuando se utilizaron junto con la lógica difusa o las redes neuronales.

En el caso de los diagramas tiempo-frecuencia nos permitieron seleccionar cincuenta puntos de las partes más importantes de las señales filtradas para su inserción en las redes neuronales. Como ya se mencionó, gracias al problema que se presentó en cuanto a la falta de capacidad de cómputo, fue estrictamente necesario el uso de las técnicas anteriormente descritas y aunque utilizamos solamente cincuenta puntos de la señal, obtuvimos resultados que se podrían considerar muy buenos tomando en cuenta que el número total de puntos fluctuaban entre los seiscientos y los mil. Esto último nos hace creer que el sistema de medición mediante espectroscopia del infrarrojo cercano podría funcionar en un futuro pero es necesario lograr la precisión que se requiere para los dispositivos médicos, lo que parece viable cuando se considera lo logrado con las limitantes mencionadas.

Bibliografía

- [1] **Munir Dabaghi Richerand, José Rodríguez Gutiérrez, Pablo Orta Chaia**, “*Diseño de un glucómetro no invasivo*”, en *Proyectos de ingeniería*, Diciembre, 2003.
- [2] **F. M. Ham, I. N. Kostanic, G. M. Cohen, and B. R. Gooch**, “*Determination of Glucose Concentrations in an Aqueous Matrix from NIR Spectra Using Optimal Time-Domain Filtering and Partial Least-Squares Regression*”, in *IEEE Trans. on Biomedical Eng.* Vol. 44, No. 6 June, 1997.
- [3] **F. M. Ham, G. M. Cohen, K. Patel, and B. R. Gooch**, “*Multivariate determination of glucose using NIR spectra of human blood serum*,” in *Proc. IEEE Eng. Med. Biol. Soc.* 16th Annu. Int. Conf., pt. II, 1994.
- [4] **K. H. Hazen, M. A. Arnold, and G. W. Small**, “*Temperature-sensitive near-infrared spectroscopic measurement of glucose in aqueous solutions*,” *Appl. Spect.*, vol. 48, no. 4, pp. 477–483, 1994.
- [5] **F. M. Ham, G. M. Cohen and Byoungcho Cho**, “*Glucose sensing using infrared absorption spectroscopy and a hybrid artificial neural network*”, Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vol. 13, no. 4, 1991.
- [6] **Sundarajan, D.** “*Digital Signal Processing: Theory and practice*”, World Scientific Publishing. USA.2003.
- [7] **Strang, G., Nguyen, T.** “*Wavelets and Filter Banks*”. Wellesley – Cambridge Press. USA.1994.
- [8] **Hayes, Monson H.**, “*Statistical digital signal processing and modeling*”, Prentice Hall, E.E.U.U. 1996
- [9] **M.D. Srinath, P.K. Rajasekaran, R. Viswanathan**, “*Introduction to statistical signal processing with applications*”, Englewood Cliffs, N.J. : Prentice Hall, c1996
- [10] **Tamada, Janet A., Lesho, Matthew and Michael J. Tierney.** *Keeping a Watch on Glucose*, Medical Electronics, IEEE Spectrum, USA: 2002
- [11] **K. Danzer, Ch. Fischbacher, K.U. Jagemann, K.J. Reichelt.** *Espectroscopia por reflexión difusa de luz cuasi-infrarroja para el monitoreo no invasivo de la glucosa en sangre.* Institut für Anorganische und Analytische Chemie, Friedrich-Schiller-Universität, Institute for Physical High-Technology Jena. Jena, Alemania.F
- [12] **Czogala, Ernest; Leski, Jacek**, *Fuzzy and Neuro-Fuzzy Intelligent Systems.* ; Editorial: Physcia-Verlag Heidelberg New York, Ed. 2000
- [13] **Jhy-Shing, Roger Jang**, *ANFIS: Adaptive-Network-Based Fuzzy Inference System*, IEEE Transaction on systems, man and cybernetics, vol. 23, No. 3 May/June 1993.
- [14] **Leondes, Cornelius**, *Fuzzy theory systems : techniques and applications*, San Diego, Calif. ; London : Academic, c1999.

Índice.

<u>Tema</u>	<u>Hoja</u>
1. <u>La diabetes y el monitoreo de glucosa.</u>	1
1.1 <u>La diabetes.</u>	1
1.2 <u>Métodos de medición de glucosa.</u>	2
1.2.1 Método de medición no invasivo.	3
1.2.1.1 Tecnología transdérmica.	3
1.2.1.2 Métodos ópticos.	4
1.3 <u>Definición del problema y propuesta de solución.</u>	6
1.3.1 <u>Problemática.</u>	6
1.3.2 <u>Propuesta de solución.</u>	7
1.3.3 <u>Método a utilizar.</u>	7
1.3.3.1 <u>Ventajas método NIR.</u>	8
1.3.3.2 <u>Desventajas método NIR.</u>	9
1.3.4 <u>Sistema propuesto.</u>	9
2. <u>Metodología.</u>	10
2.1 <u>Etapa de Pre-procesamiento</u>	
2.1.1. <u>Diseño de filtros digitales</u>	
2.1.1.1 <u>Diseño filtros IIR</u>	
2.1.1.2 <u>Diseño filtros FIR</u>	
2.1.1.3 <u>Diseño wavelets</u>	10
2.2 <u>Etapa de Caracterización</u>	11
2.2.1 <u>Coeficientes autorregresivos</u>	19
2.2.2 <u>Diagramas tiempo-frecuencia</u>	26
3. <u>Diseño de Experimento.</u>	28
3.1 <u>Adquisición de datos.</u>	28
3.1.1 <u>Equipo utilizado.</u>	28
3.1.2 <u>Datos disponibles.</u>	28
4. <u>Resultados.</u>	30
4.1 <u>Resultados de la adquisición.</u>	30
4.2 <u>Etapa de Pre-procesamiento</u>	34
4.2.1 <u>Filtros FIR.</u>	34
4.2.2 <u>Filtros IIR.</u>	44
4.2.3 <u>Wavelets.</u>	49
4.3 <u>Etapa de Caracterización.</u>	55
4.3.1 <u>Coeficientes autorregresivos.</u>	55
4.3.2 <u>Diagramas tiempo-frecuencia.</u>	58
4.4 <u>Etapa de Decisión</u>	
4.4.1 <u>Lógica difusa y coeficientes autorregresivos</u>	
4.4.2 <u>ANFIS y coeficientes autorregresivos</u>	
4.4.3 <u>Redes neuronales artificiales</u>	
5. <u>Conclusiones y perspectivas.</u>	63
6. <u>Bibliografía.</u>	64
