

03063



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

**“ANÁLISIS DE SEÑALES ULTRASÓNICAS DE
ENSAYOS NO DESTRUCTIVOS UTILIZANDO REDES
NEURONALES”**

T E S I S

QUE PARA OBTENER EL GRADO DE

**MAESTRO EN CIENCIAS
(COMPUTACIÓN)**

P R E S E N T A

FERNANDO FLORES MANGAS

DIRECTORA DE TESIS: DRA. LUCIA MEDINA GÓMEZ

México, D. F.

2005.

11349093



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Análisis de señales ultrasónicas de ensayos no
destructivos utilizando redes neuronales

Fernando Flores Mangas

fflores@uxmcc2.iimas.unam.mx

Posgrado en Ciencia e Ingeniería de la Computación,
Universidad Nacional Autónoma de México,

Asesora: Dra. Lucía Medina Gómez

3 de octubre de 2005

Autorizo a la Dirección General de Bibliotecas de la
UNAM a difundir en formato electrónico e impreso el
contenido de mi trabajo recepcional.
NOMBRE: FERNANDO FLORES
MANGAS
FECHA: 6 - OCTUBRE - 2005
FIRMA: 

Índice

Agradecimientos	V
1. Introducción	1
1.1. Trabajo relacionado	2
1.2. Organización de los capítulos	4
2. Ultrasonido	5
2.1. Física del fenómeno	6
2.1.1. Propagación	6
2.1.2. Atenuación	8
2.1.3. Reflexión	8
2.1.4. Refracción	10
2.1.5. Difracción	11
2.2. Descripción de un sistema ultrasónico	13
2.3. Transductor Ultrasónico	14
2.3.1. Tipos de transductores	15
2.3.2. Parámetros del transductor	15
2.3.3. Modos de excitación del transductor	16
2.4. Amplificador ultrasónico	17
2.5. Convertidor Analógico-Digital	18
2.6. Imágenes Acústicas	20
3. Redes Neuronales	23
3.1. Descripción general	23
3.1.1. Estructura Neuronal	24
3.1.2. Topología de la red	25
3.1.3. Aprendizaje	27
3.2. Redes multicapas de perceptrones	29
3.2.1. Retropropagación	30
3.2.2. Neurona de sesgo	33
3.2.3. Modificaciones a Retropropagación	33

4. Metodología	37
4.1. Preprocesamiento	37
4.1.1. Envolverte	38
4.1.2. Normalización	39
4.1.3. Corrección por atenuación	40
4.1.4. Homogeneidad	40
4.1.5. Interpolación	44
4.2. Red Neuronal: Construcción	46
4.2.1. Datos reales e ideales	46
4.2.2. Conjuntos de entrenamiento	47
4.2.3. Entrenamiento	48
4.3. Red Neuronal: Uso	53
5. Experimentación y resultados	55
5.1. Preprocesamiento	56
5.1.1. Envolverte	57
5.1.2. Corrección por atenuación	57
5.1.3. Normalización	58
5.2. Conjuntos de entrenamiento	62
5.3. Entrenamiento de la red	65
5.3.1. Tamaño de la capa de entrada	66
5.3.2. Tamaño de la capa intermedia	68
5.3.3. Tasas de aprendizaje y constante de momento	69
5.4. Uso de la red	73
6. Conclusiones	77
6.1. Preprocesamiento y construcción de muestras	78
6.2. Entrenamiento de la red	79
6.3. Uso de la red	80
Bibliografía	83

Índice de figuras

2.1. Medio elástico.	6
2.2. Propagación longitudinal.	7
2.3. Propagación transversal.	7
2.4. Onda Superficial.	8
2.5. Refracción.	10
2.6. Efecto de las diferencias en las velocidades.	10
2.7. Ley de Snell.	11
2.8. Difracción.	12
2.9. Sombra producida por la ausencia de refracción.	13
2.10. Ruido temporal.	20
2.11. Imagen Acústica.	21
3.1. Estructura Neuronal.	24
3.2. Funciones de Activación.	25
3.3. Red neuronal acíclica de una capa.	26
3.4. Red neuronal acíclica de dos capas.	26
3.5. Red neuronal recurrente.	27
3.6. Cálculo de Δw_{41}	32
4.1. Arco de visión de un transductor ultrasónico.	41
4.2. Arcos muestrales a diferente profundidad con α constante.	42
4.3. Resolución en profundidad (Δs).	43
4.4. Resolución Horizontal (Δt).	44
4.5. La región muestral (arriba), el spline (en medio) y la muestra (abajo).	46
4.6. Conjuntos de datos.	47
4.7. Diagrama de flujo del entrenamiento de la red.	49
4.8. Interfaz de las propiedades de la red neuronal.	51
4.9. Evolución de la medida de error del conjunto de entrenamiento (línea punteada) y del conjunto de prueba (línea continua).	52
4.10. Interfaz de análisis visual de resultados.	54
5.1. Interfaz de herramientas de preprocesamiento.	56
5.2. Envoltente de una traza.	57

5.3. Envolvente de una imagen acústica 3D.	58
5.4. Corrección de atenuación.	59
5.5. Uso del rango dinámico.	60
5.6. Interfaz de características ultrasónicas.	61
5.7. Región muestral situada en la posición central de las fallas. . . .	61
5.8. Región muestral inconsistente con los parámetros del estudio. . .	62
5.9. Interfaz de imagen ideal.	63
5.10. Interfaz de construcción de conjuntos de entrenamiento.	64
5.11. Muestras de una región muestral con diferente tamaño.	65
5.12. Análisis de una señal real utilizando redes con capa de entrada de tamaño 3, 5, 10 y 20.	67
5.13. Análisis de una señal real utilizando redes con capa intermedia de tamaño 1, 2, 5, 10 y 20.	68
5.14. Error medio cuadrático de 1000 épocas de entrenamiento.	71
5.15. Efecto del momento en la aproximación a un mínimo.	72
5.16. Efecto del momento en el error medio cuadrático.	72
5.17. Uso de una red para analizar diferentes estudios.	74
5.18. Resultados de tres diferentes redes.	75

Agradecimientos

A mi asesora, la Dra. Lucia Medina Gómez, gracias por todo el apoyo académico y personal. Con su esfuerzo ha contribuido enormemente en mi formación hacia la investigación científica formal. Gracias por ese trato tan especial y particularmente por la infinita confianza que siempre ha tenido en mí.

A todas las personas que participaron en el proceso de revisión, gracias por ser tan generosos con su tiempo, pero más importante, por enriquecer este trabajo con sus valiosos comentarios.

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) y a la Dirección General de Estudios de Posgrado (DGEP) de la UNAM, gracias por el apoyo económico recibido durante la duración del programa.

~ o ~

A mis papas, de quienes ciertamente nunca dejaré de aprender, gracias por enseñarme a siempre seguir creciendo. Mil gracias por su inagotable cariño.

A mi hermanita, a Alan y a Alex, por querernos tanto y por llenar de alegría nuestras vidas.

A mi amada Moni, gracias por todo tu amor, por permitirme invertir nuestro tiempo en este trabajo, por todos estos años en los que hemos crecido juntos y ¡por todos los que nos faltan!

A todos mis amigos del posgrado y a todos los que alguna vez estuvieron en la UNAM, mil gracias por llenar mi vida de experiencias memorables.

Gracias a Dios por rodearme de todas estas personas maravillosas y por permitirme lograr todas mis metas.

Capítulo 1

Introducción

El propósito de este trabajo de investigación es evaluar la capacidad de una red neuronal de perceptrones para analizar un conjunto de señales ultrasónicas de ensayos no destructivos (estudio), intentando producir una descripción geométrica de los elementos reflectivos.

En este documento se presenta el contexto en el que se desarrolló la investigación y las aportaciones realizadas, describiendo en un principio los aspectos esenciales del problema, como la naturaleza de las señales y las técnicas utilizadas para su obtención, algunos elementos fundamentales de los métodos de análisis, la metodología propuesta y los resultados obtenidos.

La manera de abordar el problema consiste en encontrar las características físicas que impiden obtener una descripción clara de los elementos reflectivos presentes en estudios ultrasónicos y utilizar estas características junto con una red neuronal para revertir las perturbaciones en la señal.

La motivación para realizar este trabajo consiste en facilitar (o posiblemente habilitar) el uso de tecnologías ultrasónicas de ensayos no destructivos al utilizar un sistema de análisis automatizado de las señales.

Todas las descripciones de conceptos relacionados, ya sea con señales ultrasónicas o con redes neuronales, son descritos con la profundidad necesaria para comprender la totalidad del documento y no se esperan conocimientos previos del lector mas que de matemáticas y cálculo básico.

1.1. Trabajo relacionado

Existe una gran variedad de aplicaciones relacionadas al análisis de señales ultrasónicas, basadas en principios físicos, matemáticos o mecanismos de clasificación. No obstante esta diversidad de perspectivas (y resultados), es posible realizar cierta agrupación.

Una idea común es la de efectuar una etapa de preprocesamiento para producir un vector que agrupa un conjunto de características que alimenta a la red neuronal. Uno de estos casos utiliza una red probabilística para realizar clasificaciones usando como entrada estos vectores de características [1]. Otro utiliza filtros morfológicos y procesamiento de imágenes como la forma de construir el vector que alimenta al clasificador [2].

También existen propuestas donde la señal ultrasónica completa (y no un vector de características) entra a la red después de pasar por alguna etapa de preprocesamiento. En [3] se explota la idea de que las señales ultrasónicas tienen mas información sobre las características de las fallas en el dominio de la frecuencia, en cambio son mas claras con respecto a la posición de las fallas en el dominio del tiempo. Así pues sugieren obtener la transformada de Fourier de la señal completa para alimentar con esta nueva señal a una red neuronal y decidir sobre la naturaleza de las fallas.

Un enfoque diferente consiste en hacer que la red neuronal forme parte de las etapas previas a la clasificación. Por ejemplo en [4] mencionan la dificultad de interpretar señales ultrasónicas dada la escasa relación entre los patrones que se observan y los defectos reales, presentando el problema de reconocimiento de formas. La solución involucra una etapa de extracción de características usando una red neuronal, seguida de una etapa de clasificación de objetos. Otro ejemplo en este sentido consiste en reconstruir la señal de ruido basándose en las características del sistema ultrasónico y del material, utilizando una red neuronal. El resultado se resta de la señal original [5] en un proceso de restauración.

Existen enfoques en los que las redes son utilizadas tanto en la etapa de preprocesamiento, para realzar las características útiles de la señal, como en la etapa de clasificación. En [6] una primera red incrementa la relación señal a ruido y una segunda realiza la localización de las fallas. Por el contrario en [7] utilizan la red como un aproximador de funciones para relacionar la distancia entre el transductor y la falla con el voltaje de un mecanismo de detección electromagnético, como una parte del proceso de análisis general.

Otro enfoque sugiere el uso de múltiples redes para distinguir entre diferentes

tipos de fallas. Luego se utiliza un sistema de votación para decidir sobre la clase ganadora [8]. Por otro lado existen propuestas que no consideran el tipo de fallas y cuyo objetivo es únicamente distinguir la existencia o ausencia de estas. En este caso, existe una propuesta que sugiere alimentar a la red con la señal ultrasónica completa [9].

La propuesta de [10] es un algoritmo en el que se utilizan principios de propagación para definir un límite, que al ser superado, asume la presencia de una falla, produciendo una decisión preliminar. Después se utiliza una red neuronal de base radial para tomar la decisión final.

Una de las propuestas sugiere construir una señal utilizando una red neuronal que asume un tipo de falla para luego comparar esta señal sintética con la señal real. El proceso continua hasta encontrar el tipo de falla que hace que la red produzca la señal mas cercana a la real [11].

Otras alternativas precinden de algoritmos basados en redes neuronales. Por ejemplo existe un modelo basado en diagramas de Voronoi con el que es posible simular la estructura granular de algunos materiales y así distinguir entre el ruido de fondo y el reflejo producido por una grieta [12]. Se encontró también la idea de buscar contornos en imágenes ultrasónicas y utilizar análisis tiempo-frecuencia para realizar clasificaciones [13]. Otro ejemplo se basa en la idea de que una señal ultrasónica que registra una falla difiere suficientemente de una distribución conocida de señales sin fallas [14].

Por último se presentan algunos de los trabajos que orientaron o fortalecieron la decisión de utilizar algún procedimiento en la metodología que se propone.

En primer lugar, la idea de [15], donde se utiliza una red del tipo ART2, en la que se aplica una etapa de preprocesamiento con el fin de obtener la envolvente de la señal y aparte se utilizan wavelets para resaltar algunas características. Esta señal preprocesada ingresa a la red que produce el resultado final. La idea del cálculo de la envolvente contribuye en la metodología que se presenta en el Capítulo 4.

En segundo lugar, en [16] se introduce la idea de que las fallas ocurren a diferentes resoluciones dentro de una misma señal ultrasónica y consecuentemente utilizar esquemas que realizan un mapeo de estos defectos a una resolución única no resulta en una solución necesariamente óptima. Esta idea fortalece la motivación del algoritmo de construcción de regiones propuesto en la Sección 4.1.5.

Finalmente, existe la propuesta de usar un árbol de redes neuronales, donde

la primera red realiza una clasificación burda que determina cual de las redes de la siguiente etapa debe ser utilizada para producir el resultado final. Se utiliza un vector de características tomadas de la señal para alimentar a las redes, aunque no todas las características entran en todas las redes [17]. Esta idea reafirma la necesidad de seleccionar cuidadosamente la información que ingresa a la red, ya que no necesariamente al incluir mas información se producen mejores resultados.

1.2. Organización de los capítulos

El capítulo siguiente a la introducción consiste en la descripción de los detalles relacionados con las técnicas de ultrasonido, incluyendo algunas descripciones de la física del fenómeno utilizadas para el desarrollo de la solución y mencionando los parámetros que intervienen al momento de obtener y analizar una señal. Luego sigue un capítulo dedicado a redes neuronales, mencionando detalles como la arquitectura tanto de las neuronas como de las redes. Se habla también de los algoritmos de aprendizaje y se profundiza en las características de las redes de perceptrones, justificando su selección.

El cuarto capítulo describe la metodología utilizada para construir un algoritmo capaz de lograr el objetivo propuesto. En este sentido se habla de las etapas clave en la construcción del mecanismo de análisis, como el preprocesamiento de las señales y el entrenamiento de la red neuronal. Este capítulo termina describiendo como se utiliza el clasificador construido.

El quinto capítulo presenta, los resultados que cada etapa produce, intentando evidenciar la contribución de cada una para conseguir el objetivo final.

El capítulo final resume las conclusiones producidas a lo largo del desarrollo de esta investigación, haciendo comentarios tanto de los componentes como del sistema en general, mencionando posibilidades de trabajo futuro. El capítulo (y el documento) termina con algunos comentarios sobre la utilidad de sistemas de análisis como éste, enmarcado en los retos que presenta la industria en nuestros días.

Capítulo 2

Ultrasonido

El ultrasonido es un fenómeno acústico (mecánico) que se propaga por cualquier medio gracias al movimiento de las moléculas. El prefijo *ultra* a la palabra *sonido* sugiere frecuencias mayores a las audibles por el oído humano, esto es mayores a 20 *KHz*.

El número de aplicaciones que utilizan ultrasonido crece continuamente en disciplinas médicas o, como en este caso, industriales, en donde se utiliza para describir las características al interior de una pieza.

La forma de obtener información de una pieza en un *Ensayo No Destructivo* (END) utilizando señales ultrasónicas es interpretando las perturbaciones que la onda ultrasónica sufre dentro del material. Este principio se utiliza desde hace muchos años en aplicaciones como el SONAR y mas recientemente en la bien conocida ecografía médica. En la industria se utiliza para conocer las características físicas como la geometría o la constitución de la pieza y además para producir descripciones sobre la existencia de fallas como burbujas, grietas o impurezas.

Este capítulo describe en primer lugar los detalles de la física que permiten llevar a cabo ensayos no destructivos basados en ultrasonido, incluyendo algunos fenómenos como el de propagación, el de atenuación, el de reflexión, etc. Luego se mencionan los componentes requeridos en un sistema de ultrasonido, seguidos de algunas descripciones mas detalladas de aquellos que resultan mas importantes, como el transductor y el convertidor analógico-digital. Por último se mencionan algunos detalles de la construcción de imágenes acústicas como una herramienta de análisis.

2.1. Física del fenómeno

La base de los END que utilizan ultrasonido son las vibraciones y deformaciones temporales producidas por la energía acústica. A nivel atómico, las moléculas se mueven en un gran número de patrones, aunque la mayoría de estos movimientos resultan irrelevantes para el análisis ultrasónico. Sin embargo, el movimiento al unísono de grandes conjuntos de átomos produce una onda de energía mecánica que origina zonas de tensión y compresión, que junto con la inercia, produce oscilaciones acústicas en un medio elástico.

Una descripción sintética de este medio es la que se muestra en la Figura 2.1, donde las moléculas son pequeñas masas unidas por resortes.

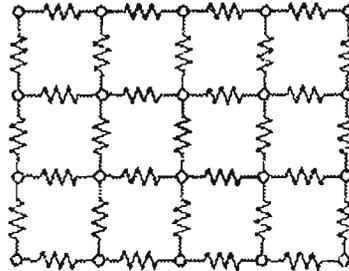


Figura 2.1: Medio elástico.

La rigidez de estos resortes y el efecto de la inercia definen parámetros importantes del fenómeno de propagación, como la velocidad, por ejemplo.

La Ecuación 2.1 relaciona la *Impedancia Acústica* o resistencia del material a la propagación del sonido (Z), la velocidad de propagación (v) y la densidad del material (ρ).

$$Z = \rho v \quad (2.1)$$

2.1.1. Propagación

La forma en que se propagan las oscilaciones dentro de un material sólido define diferentes modos. Los dos más importantes son el longitudinal y el transversal.

1. Longitudinal: son oscilaciones que ocurren en la dirección de propagación de la energía. Producen fuerzas de compresión y dilatación modificando

temporalmente la densidad del material conforme a su paso. El desplazamiento de una onda longitudinal se muestra en la Figura 2.2, los puntos representan la posición de moléculas o conjuntos de ellas.



Figura 2.2: Propagación longitudinal.

2. Transversal: las partículas oscilan en direcciones ortogonales a la de propagación. Para que este modo de propagación ocurra se necesita que el medio sea sólido. Las ondas transversales son relativamente débiles cuando su magnitud se compara con la de las longitudinales, pues de hecho en algunos casos se generan utilizando la energía de las últimas. La Figura 2.3 muestra el desplazamiento de una onda transversal. Una vez más los puntos muestran la posición de moléculas o grupos de ellas.

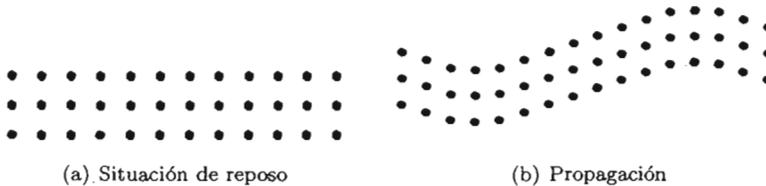


Figura 2.3: Propagación transversal.

Existen algunas otras formas de propagación que ocurren en las superficies de los materiales o en placas muy delgadas. Su caracterización es complicada y la energía que se propaga en estas formas es mucho menor a la que lo hace en ondas transversales o longitudinales. La Figura 2.4 muestra, por ejemplo, la forma en que se propaga una onda superficial.

Existen también algunos otros fenómenos relacionados con la propagación de la energía acústica, como la atenuación, la dispersión, la reflexión, la refracción y la difracción. La caracterización de la mayoría de estos es importante al analizar resultados de END basados en ultrasonido.



Figura 2.4: Onda Superficial.

2.1.2. Atenuación

La intensidad de una onda acústica que viaja por cualquier medio disminuye con respecto a la distancia, a este fenómeno se le denomina *atenuación*. Las fuentes de atenuación son dos: la *dispersión* y la *absorción*. La dispersión se debe a reflexiones producidas por pequeñas variaciones en la densidad del material que desvían la energía de la dirección de propagación original. La absorción consiste en la transformación de la energía acústica en alguna otra forma de energía, como por ejemplo calor. En un material idealmente conductor, la atenuación se debe únicamente a la dispersión y en ese caso la energía que entra a la pieza sale de ella conservando su magnitud. Sin embargo ese tipo de material no existe en la naturaleza.

La Ecuación 2.2 describe la forma en que la energía decrece al viajar una distancia z desde un punto de referencia donde la magnitud es A_0 .

$$A = A_0 e^{-\alpha z} \quad (2.2)$$

La amplitud A es la amplitud reducida y la constante α es el coeficiente de atenuación de la onda viajando en la dirección de z .

Por lo general, la atenuación es proporcional al cuadrado de la frecuencia del sonido, aunque existe un amplio número de factores que afectan a este fenómeno, como la pureza del material e incluso la forma en que la pieza fue fabricada. Por lo general se recomienda utilizar coeficientes de atenuación basados en mediciones del material que se analiza.

2.1.3. Reflexión

Las ondas ultrasónicas se reflejan al encontrar fronteras de materiales con diferentes impedancias acústicas (ver Sección 2.1). La fracción de energía incidente que es reflejada se puede calcular debido a que la velocidad y la presión

local de un lado al otro de la frontera deben ser continuas. El coeficiente que indica la proporción de energía reflejada o *coeficiente de reflexión* (r), con respecto a la energía incidente se calcula utilizando la Ecuación 2.3 que involucra la impedancia acústica de los materiales a ambos lados de la frontera; es decir, del material de origen (Z_1), como del de destino (Z_2) y consecuentemente sus densidades y velocidades de propagación.

$$r = \left(\frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2 \quad (2.3)$$

La Tabla 2.1 muestra las impedancias y velocidades acústicas de algunos materiales, así como sus densidades.

Material	Impedancia (Z) [Adimensional]	Velocidad (v) 1×10^4 [m/s]	Densidad (ρ) Kg/dm ³
Agua	0.1480	1.48	1.00
Carne	0.1674	1.55	1.08
Concreto	0.8395	3.65	2.30
Aluminio	1.7696	6.32	2.80
Plomo	2.4624	2.16	11.40
Titanio	2.7315	6.07	4.50
Latón	3.6637	4.28	8.56
Fierro	3.7440	4.80	7.80
Cobre	4.1614	4.66	8.93
Acero Inoxidable	4.5162	5.79	7.80

Tabla 2.1: Impedancias, velocidades y densidades de algunos materiales.

La tabla está ordenada con base en las impedancias, sin embargo la intensidad de una reflexión es directamente proporcional a la *diferencia* de impedancias e inversamente proporcional a la suma de sus magnitudes.

La otra parte de la energía incidida (aquella que no fue reflejada) es transmitida al interior de la pieza. El coeficiente que indica la proporción de energía transmitida con respecto a la incidida o *coeficiente de transmisión* (t) es el complemento de r , es decir:

$$t = 1 - \left(\frac{Z_2 - Z_1}{Z_2 + Z_1} \right)^2 \quad (2.4)$$

Esta energía que logra cruzar la frontera para internarse en la pieza sufre un fenómeno conocido como refracción.

2.1.4. Refracción

La *refracción* consiste en un cambio en la trayectoria de la onda ultrasónica al paso por una diferencia de impedancias. La razón consiste en la diferencia de velocidades acústicas de las ondas en ambos materiales.



Figura 2.5: Refracción.

Un ejemplo de los efectos de la refracción se muestra en la Figura 2.6, donde se hace incidir una onda ultrasónica plana que viaja en un material con velocidad v_1 y cruza la frontera para ingresar en un material con velocidad v_2 . En el primer caso, la onda se dobla como en la Figura 2.6a, pues la velocidad en el material origen es menor que en el de destino. El caso opuesto se muestra en la Figura 2.6b.

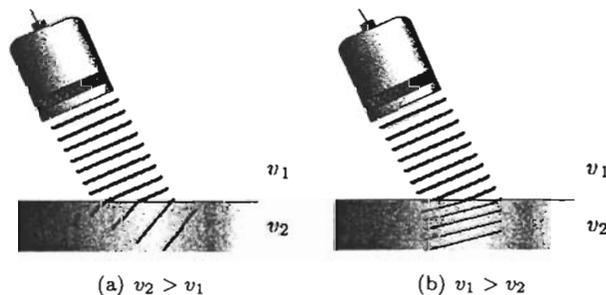


Figura 2.6: Efecto de las diferencias en las velocidades.

Este fenómeno está descrito por la *Ley de Snell*, expresada mediante la Ecuación

ción 2.5, que describe la relación entre el ángulo de incidencia (Θ_1) y el de transmisión (Θ_2) con respecto a las velocidades de propagación en los materiales (v_1 y v_2).

$$\frac{\sin(\Theta_1)}{v_1} = \frac{\sin(\Theta_2)}{v_2} \quad (2.5)$$

El diagrama de la Figura 2.7 muestra la relación entre los cuatro factores.

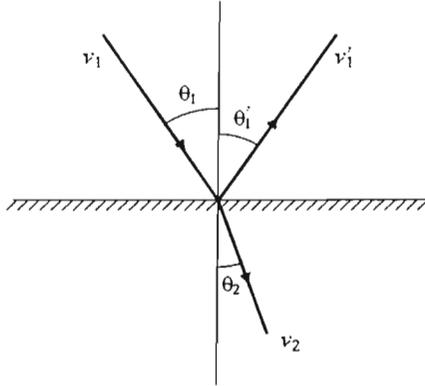


Figura 2.7: Ley de Snell.

Note también como el ángulo de reflexión (Θ_1') es idéntico al ángulo de incidencia ya que la onda se desplaza por el mismo material, conservando su velocidad.

2.1.5. Difracción

La *difracción* es un fenómeno de las ondas en general, y en particular de las acústicas, que son capaces de rodear obstáculos o de pasar a través de espacios pequeños comparados con la longitud de onda.

La Figura 2.8 muestra un ejemplo de cada una de estas situaciones. El punto oscuro representa la fuente de una onda de energía acústica, el punto claro representa un obstáculo y la línea es una pared. No obstante la existencia de estos elementos, la onda acústica alcanza las regiones marcadas gracias al fenómeno de la difracción.

La difracción es la razón principal para utilizar frecuencias de ultrasonido, pues la longitud de onda es inversamente proporcional a la frecuencia. Utilizando

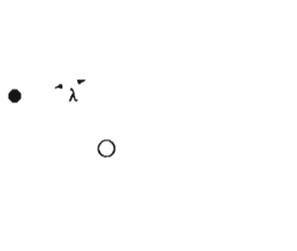


Figura 2.8: Difracción.

ultrasonido se producen longitudes de onda pequeñas y consecuentemente se evita que la onda de energía pase de largo a los obstáculos, haciendo que la energía sea reflejada para construir la descripción del material que se analiza. No obstante, la longitud de onda sigue siendo uno de los factores críticos que definen la resolución de un estudio, pues por lo general el tamaño del elemento visible mas pequeño es mayor que la mitad del tamaño de la longitud de onda.

Por otro lado, uno de los efectos negativos producidos por el hecho de utilizar frecuencias altas (que consecuentemente evitan la difracción) consiste en la formación de *sombras* producidas por la ausencia de energía que alcance a los elementos reflectivos mas profundos alineados en la dirección de propagación, haciendo imposible su descripción. La imagen de la Figura 2.9 muestra de manera sintética un ejemplo de este efecto. Los óvalos simulan elementos reflectivos. La sombra se proyecta en la dirección de propagación de la onda cubriendo algunos elementos a mayor profundidad.

De igual forma, la Figura 2.11 muestra una imagen real en la que también se observan sombras, particularmente al fondo de la pieza.

Una vez descritos los fenómenos relacionados con la propagación de la energía acústica, lo siguiente es mencionar los detalles de los ensayos no destructivos utilizando este tipo de energía.

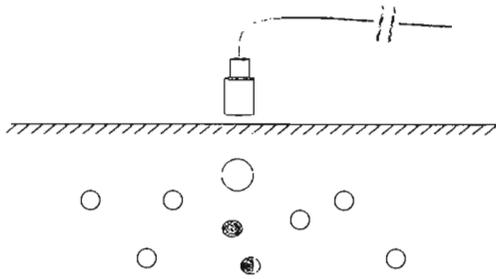


Figura 2.9: Sombra producida por la ausencia de refracción.

2.2. Descripción de un sistema ultrasónico

Todo sistema ultrasónico requiere de al menos tres componentes básicos: Un sistema de emisión, el medio a analizar y un sistema de recepción.

1. Sistema de Emisión: Es el encargado de producir el pulso ultrasónico que genera las perturbaciones dentro del medio. Generalmente tiene tres componentes: un generador de un pulso eléctrico, un transductor que transforma la energía eléctrica del pulso en energía acústica y un mecanismo de acoplamiento que se encarga de facilitar el ingreso de la energía acústica al interior de la pieza, intentando disminuir las pérdidas en la(s) interfase(s).
2. Medio a analizar: Que recibe el pulso ultrasónico y se perturba de acuerdo a sus características.
3. Sistema de Recepción: Tiene al menos dos componentes, un sensor (o un arreglo de éstos) que percibe las perturbaciones del medio y las convierte de energía mecánica en energía eléctrica y un sistema de análisis de señales que utiliza la señal eléctrica para construir alguna descripción útil. En la actualidad este sistema de análisis comienza con un convertidor analógico digital que toma como entrada las señales del transductor y produce como salida un archivo de valores reales con los que es posible, al menos idealmente, reconstruir la señal acústica.

Dado el número de factores que intervienen en un END, cada aplicación requiere un diseño a la medida, en donde los componentes de todo el sistema se adecúan para satisfacer características específicas. Algunos de los factores importantes se muestran en la Tabla 2.2.

Característica	Rango de Valores
Tipo de propagación	Longitudinal, transversal, etc.
Velocidad de propagación	$0.3 \rightarrow 6500 \text{ m/s}$
Rango de frecuencias	$0.03 \rightarrow 20 \text{ MHz}$
Atenuación	$1 \rightarrow 5000 \text{ dB/m}$
Resolución Axial	$1 \times 10^{-8} \rightarrow 1 \times 10^{-1} \text{ m}$
Resolución Transversal	$0 \rightarrow \infty \text{ m}$
Rango dinámico	100 dB
Tipo de Acoplamiento	Sólido, Líquido, Gaseoso
Frecuencia de Repetición de pulso	$0 \rightarrow 20000 \text{ Pulsos/s}$
Tipo de inspección	Pulso-eco, Transmisión, otros
Tamaño de la pieza	$0 \rightarrow \infty \text{ m}$

Tabla 2.2: Características en aplicaciones de ultrasonido y sus rangos.

Un componente extra que provee de gran funcionalidad es el sistema que brinda la posibilidad de presentar los datos de un END mediante una *imagen acústica*, que se obtiene al situar las trazas de varios pulsos ultrasónicos tomados mediante un barrido espacial una a lado de la otra. Así es posible generar imágenes de dos e incluso de tres dimensiones que facilitan el proceso de evaluación. Para hacer el barrido se utilizan dos técnicas, una que funciona mediante arreglos de transductores que registran las perturbaciones en diferentes regiones del material y otra que lo hace mediante un solo transductor que se mueve para adquirir trazas contiguas.

A continuación se presentan algunas descripciones más detalladas de los componentes de un END que influyen en el análisis que se realiza en esta investigación, como son el transductor ultrasónico y el convertidor analógico digital.

2.3. Transductor Ultrasónico

El transductor ultrasónico se encarga de transformar la energía acústica en energía eléctrica y viceversa. Para lograr esta transformación se han explotado diversos fenómenos naturales, como el electromagnético o la electrostricción. Algunos de los tipos de transductores se describen a continuación.

2.3.1. Tipos de transductores

El tipo de transductor ultrasónico mas común en END es el llamado *piezo-eléctrico*, aunque existen también los *capacitivos*, los *acústico electromagnéticos* y los *láser*, entre otros.

1. Transductores Piezoeléctricos: Su funcionamiento se basa en el fenómeno de la electrostricción que consiste en la capacidad de algunos materiales para producir un potencial eléctrico al experimentar un cambio de presión y viceversa. Gracias a su gran rendimiento, que se traduce en una excelente relación señal a ruido y a su alta eficiencia, son los más utilizados en aplicaciones de END. El rango de frecuencias va desde 0.02 MHz hasta 100 MHz. Su principal desventaja es la necesidad de un mecanismo de acoplamiento, ya sea líquido o sólido para transmitir la energía acústica a la pieza de análisis.
2. Transductores Capacitivos: Funcionan muy bien para aplicaciones en el aire y principalmente como herramientas para medir distancias. El funcionamiento se basa en la oscilación de una lámina flexible de algún material conductor al producir un potencial eléctrico con respecto a otra lámina rígida. Tienen una frecuencia máxima de oscilación de 2 MHz y utilizan pulsos muy amplios, limitando la resolución.
3. Transductores acústico electromagnéticos: Las vibraciones del material analizado se registran utilizando una bobina que las transforma en corriente eléctrica. Su uso se limita al análisis superficial (menores a 1 mm) y funcionan únicamente con materiales conductores de electricidad.
4. Transductores Láser: El láser genera un pulso de alta energía y corta duración que al incidir sobre la superficie de la pieza produce ondas mecánicas de varios MHz de frecuencia en el interior. Otro láser ilumina la zona de interés y las oscilaciones de la superficie se registran utilizando interferometría. Esta técnica presenta ventajas como evitar la necesidad de entrar en contacto físico con la pieza y de no requerir ningún medio de acoplamiento, aunque la solución es sumamente costosa.

2.3.2. Parámetros del transductor

Todos los transductores se pueden caracterizar funcionalmente utilizando algunos parámetros físicos como la penetración, la sensibilidad, la resolución y

la direccionalidad del haz que producen. Algunos datos importantes de cada una de estas se presentan a continuación:

1. **Penetración:** un pulso ultrasónico sufre de atenuación al desplazarse al interior de un material. Además la frecuencia del pulso es un factor importante, pues a mayor frecuencia, mayor atenuación.
2. **Sensibilidad:** Se encuentra relacionada con la longitud de onda (y por ende con la frecuencia) y con el material utilizado para construir el transductor. La sensibilidad es la capacidad de detectar características que reflejan poca energía y se ve favorecida con longitudes de onda menores.
3. **Resolución Axial:** Es la capacidad de distinguir elementos reflectivos en la dirección de propagación del pulso. La duración del pulso inducido se encuentra inversamente relacionada con la resolución axial. Sin embargo, para producir un pulso muy corto se necesita un transductor con alto ancho de banda.
4. **Resolución Transversal:** Define la capacidad de distinguir elementos reflectivos cercanos entre si, situados en direcciones ortogonales a la dirección de propagación del pulso. La direccionalidad del haz está directamente relacionada con la frecuencia del pulso inducido e inversamente con el diámetro del transductor.

2.3.3. Modos de excitación del transductor

Existen tres modos comúnmente usados para excitar un transductor que son: continua, monopulsada y multipulsada. Su uso depende de la aplicación.

1. **Excitación continua:** El transductor se somete a una corriente eléctrica oscilante a su frecuencia de resonancia. Su uso es poco frecuente y dado que el transmisor que emite la energía ultrasónica siempre está en funcionamiento pues es necesario utilizar un segundo transductor para hacer las lecturas. Además es importante considerar que existe la posibilidad de encontrar ondas estacionarias al interior de la pieza
2. **Excitación monopulsada:** Es la forma de excitación mas utilizada. Consiste en descargar un intenso pulso de energía de muy corta duración, intentando que al menos uno de los flancos sea tan vertical (rápido) como sea posible, explotando así toda la gama de frecuencias del transductor y maximizando la resolución axial.

3. Excitación multipulsada: En un intento por mejorar la transferencia de energía del transductor a la pieza, un transductor puede ser excitado utilizando trenes de pulsos de número, frecuencia y amplitud variables. Sin embargo, la duración del pulso resultante es mayor que con la excitación monopulsada, lo cual afecta negativamente la resolución axial.

En cualquiera de los dos modos de excitación pulsados existe una limitación física motivada por el principio de propagación de la energía a una cierta velocidad, que determina el tiempo mínimo entre un par de impulsos contiguos. Para calcular este tiempo se suma la duración del recorrido del transductor a la pieza mas el tiempo de recorrido dentro de la pieza (momento de adquisición de datos) mas el tiempo de cancelación del ruido residual. No esperar este tiempo entre un par de mediciones produciría resultados ruidosos.

Una vez excitado el transductor y acoplada la energía acústica en la pieza de análisis, es necesario cambiar el modo del transductor de transmisor a receptor (o en su caso activar el transductor destinado a recibir) para registrar las perturbaciones que provienen del interior. Esta señal pasa del transductor al convertidor analógico-digital pero no sin antes pasar por una etapa de amplificación.

2.4. Amplificador ultrasónico

Para que en esta etapa se conserven las características útiles de la señal, el amplificador debe cumplir ciertas características, como son: ganancia constante en todo el rango dinámico, suficiente ancho de banda y bajos niveles de ruido.

1. Ganancia constante sobre todo el rango dinámico: Se refiere a la capacidad de aumentar la potencia de la señal de entrada con el mismo factor, independientemente de la amplitud de la señal original. En algunos casos es útil considerar el uso del logaritmo de la señal como entrada al amplificador y luego obtener la exponencial de la señal amplificada, particularmente cuando el rango dinámico de la señal original es demasiado grande.
2. Ancho de banda: Esta es una de las características mas críticas de un amplificador en ultrasonido, pues típicamente es necesario al menos duplicar el ancho de banda del transductor para conservar las características de frecuencia mas alta de la señal. Además, es importante que la ganancia

del amplificador sea constante sobre todo el ancho de banda (o al menos en la banda de operación).

3. Nivel de ruido: Es necesario que el nivel de la señal mas débil supere al nivel de ruido del amplificador en al menos en 6 dB. De lo contrario, la señal ultrasónica se vuelve indistinguible ante la presencia del ruido.

Además de estas tres características indispensables, algunos amplificadores proveen funcionalidad adicional, por ejemplo utilizando un conjunto de filtros es posible eliminar o acentuar bandas de frecuencias o minimizar componentes de ruido. También es posible calcular la envolvente, facilitando la detección de algunas características en etapas posteriores. Incluso hay herramientas que compensan la atenuación por distancia, proveyendo a la etapa conversión analógico-digital de una señal homogénea en este sentido.

Luego de amplificar y acondicionar la señal ultrasónica, ésta pasa a la etapa de análisis, que usualmente comienza con un convertidor analógico-digital.

2.5. Convertidor Analógico-Digital

El objetivo del *Convertidor Analógico-Digital* (CAD) es producir una descripción de la señal original utilizando muestras de la amplitud. Estas muestras deben ser tomadas a una frecuencia constante de modo que sea posible reproducirla con un nivel de detalle funcional. La salida del CAD se conecta usualmente a una computadora que almacena los datos en un disco para su análisis posterior. Los factores mas importantes en un CAD son la frecuencia de muestreo, el tamaño de la palabra usado para representar la muestra y el tamaño de la memoria temporal que almacena las trazas. Todos estos factores asociados con el convertidor, mas algunos otros, asociados al fenómeno de propagación de la energía acústica, se incorporan en la Ecuación 2.6 para obtener una relación entre ellos. La ecuación es:

$$M = \frac{2Lf_sT}{c}, \quad (2.6)$$

donde M es la cantidad (en bits) de memoria necesarios para almacenar el conjunto de muestras de una traza, L es la profundidad (en metros) que se pretende estudiar de la pieza, f_s es la frecuencia de muestreo (en segundos⁻¹), T es el tamaño (en bits) de cada muestra y c es la velocidad de propagación del sonido dentro de la pieza (en *metros/segundo*).

Algunos valores típicos de estas variables son los siguientes: para la frecuencia de muestreo $f_s = 10 \rightarrow 100 \text{ MHz}$ y para el tamaño de palabra $T = 8, 10, 12$ ó 16 bits . Así, por ejemplo, una pieza de aluminio ($c_{Al} = 6320 \text{ m/s}$) de 5 cm de profundidad ($L = 0.05 \text{ m}$) muestreada a $f_s = 50 \text{ MHz}$ y con un tamaño de palabra de $T = 8 \text{ bits}$ requiere un total de 5406 bits (676 bytes) para almacenar una traza completa.

Además de los parámetros de muestreo, es necesario considerar algunos factores al momento de digitalizar un estudio ultrasónico, principalmente relacionados con el momento de iniciar la adquisición de cada traza. Estos son: el retardo, el modo y el ruido temporal.

1. Retardo: Una vez disparado el pulso de energía acústica hacia la pieza, este realiza un viaje por el material acoplante. Durante ese tiempo no es necesario adquirir muestras pues su información es por lo general irrelevante. En consecuencia, es posible definir el inicio del momento de adquisición algunos instantes después de disparar el pulso.
2. Modo: Se refiere al modo de inicio de adquisición de datos y típicamente hay dos opciones, uno llamado *modo automático*, en el que la adquisición comienza justo después del retardo y el otro llamado *modo sincronizado* en el que la adquisición inicia al recibir el primer eco después del retardo. Esta última opción es particularmente útil pues de manera intrínseca define una referencia de distancia (basado en el tiempo o en el número de muestras) entre cualquier eco producido en el interior y el borde de la pieza asociado al primer eco.
3. Ruido Temporal: Se produce cuando existe incertidumbre en el tiempo que transcurre entre la emisión del pulso y el momento de inicio de adquisición de datos. Si la señal digital se convierte después en una imagen, el ruido temporal impedirá una clara visualización de trazas contiguas. Este ruido también introduce errores en los análisis. La Figura 2.10 muestra el efecto del ruido temporal.

El cuidado en estos tres factores permite optimizar el uso de memoria y favorece la adquisición de señales más descriptivas y de mejor calidad. Sin embargo, es necesario considerar también el ruido asociado a cualquier proceso de digitalización, principalmente por el hecho de la longitud de palabra finita y por posibles inconsistencias del convertidor.

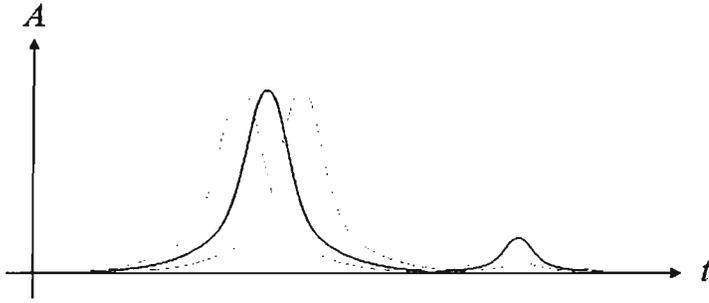


Figura 2.10: Ruido temporal.

2.6. Imágenes Acústicas

Una imagen acústica se obtiene al representar los valores de las trazas contiguas utilizando algún código de colores relacionado con la intensidad de la señal, como niveles de gris. La Figura 2.11 muestra una imagen acústica.

Las imágenes acústicas permiten realizar descripciones cualitativas de las características de una pieza, aunque su análisis también permite hacer descripciones cuantitativas, sobre todo si se conocen las características más importantes como la resolución axial, la resolución lateral, etc. Las características más relevantes en una imagen acústica se describen a continuación, algunas de ellas son definidas durante el proceso de adquisición.

1. **Resolución axial:** El concepto es el mismo de la resolución axial del transductor, aunque en este caso no se utiliza la señal analógica que este produce, sino con la señal digital que se obtiene del CAD. La menor de las resoluciones axiales (del transductor o del CAD) es la resolución axial de la imagen acústica.
2. **Resolución lateral:** La resolución lateral es la capacidad de separar reflectores que se encuentran alineados con el eje de desplazamiento del transductor. Cuando se analizan señales en las que los reflectores se encuentran muy cerca del transductor, la resolución lateral es aproximadamente igual al diámetro del transductor. No así para inspecciones de reflectores lejanos al transductor, en donde el ancho del pulso crece linealmente con relación a la profundidad. Este es uno de los factores que producen mayores dificultades para interpretar señales ultrasónicas.

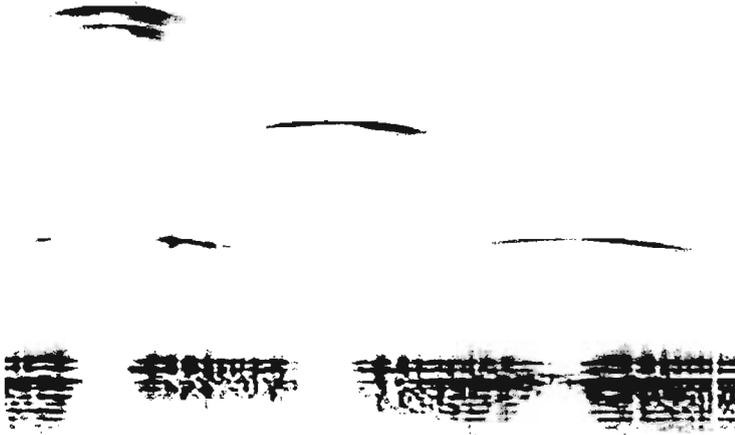


Figura 2.11: Imagen Acústica.

3. Rango dinámico: En un contexto de imágenes, el rango dinámico es el rango de diferencia tonal entre la parte más clara y la más oscura de una imagen. Con respecto a imágenes acústicas, se refiere a la diferencia entre la amplitud máxima y la mínima representadas en la imagen. Cabe señalar que tener un rango dinámico muy amplio no es necesariamente bueno, ya que el número de niveles diferentes de amplitud es finito.¹ En ese caso, un rango dinámico amplio puede asociar un buen porcentaje de los niveles útiles a amplitudes poco frecuentes o quizá no importantes (como la amplitud del reflejo que se produce en la interfaz entre el material acoplante y la pieza, que típicamente es muy grande). Idealmente, el rango dinámico se debe ajustar exactamente al rango de amplitudes que contienen la información importante de la señal.
4. Estabilidad de la inspección: Se refiere a la capacidad de repetir la medición obteniendo los mismos resultados. Si esto se logra se puede concluir que los parámetros se mantienen constantes a lo largo de todo el estudio.
5. Frecuencia de adquisición: Años atrás, la parte del proceso que definía el límite de la frecuencia con que se adquieren trazas era la electrónica,

¹El número de niveles cuantizables depende del tamaño de palabra del convertidor analógico-digital: Si $T = n \text{ bits}$ existen 2^n niveles, por ejemplo si $T = 8$ existen 256 niveles o si $T = 16$ existen 65536 niveles.

últimamente ese límite se define más por cuestiones de disipación de la energía acústica al interior de la pieza. Este parámetro es importante para imágenes acústicas sólo cuando se actualiza la imagen a tiempo real de acuerdo con el proceso de adquisición.

La metodología descrita en el capítulo 4 utiliza las propiedades de la física del fenómeno de propagación ultrasónica como una fuente significativa de información para realizar el análisis de las señales. Consecuentemente, el contenido de este capítulo resulta fundamental.

Capítulo 3

Redes Neuronales

La idea de Red Neuronal Artificial, comúnmente referida como *Red Neuronal* (RN) consiste en la definición de un modelo que intenta reproducir las funciones de aprendizaje y generalización del cerebro humano utilizando herramientas matemáticas. Algunas características biológicas fueron particularmente importantes para la definición de este modelo. Por ejemplo, el hecho de que el conocimiento está distribuido en múltiples componentes, implica que ningún elemento es completamente responsable de ninguna decisión, permitiendo construcción de modelos mas robustos. También la idea de mantener comunicación local entre las neuronas, creando la posibilidad de incluir múltiples elementos al producir las respuestas a diferentes estímulos. Por último, la capacidad de adaptación del cerebro, permitiendo abstraer las características apropiadas para cada fenómeno en un proceso de aprendizaje.

3.1. Descripción general

Las siguientes secciones especifican los tres componentes que definen a una RN: la estructura de cada neurona, la topología de la red y el mecanismo de aprendizaje. La cuarta sección describe el funcionamiento de la “Neurona de Sesgo” y la quinta presenta algunas modificaciones al algoritmo original de retropropagación que incrementan su desempeño.

3.1.1. Estructura Neuronal

La neurona es la unidad mínima de procesamiento en una RN y típicamente se utilizan tres componentes para describirla: Una función sináptica o de entrada, una función somática o de activación y una función de salida. La Figura 3.1 muestra estas tres componentes y el flujo interno de la información.

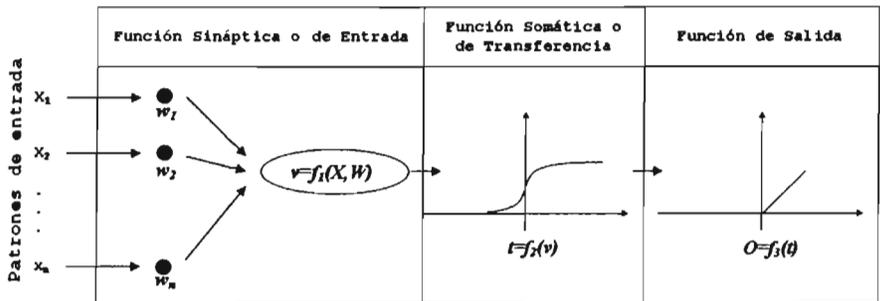


Figura 3.1: Estructura Neuronal.

La neurona recibe los estímulos de todas las entradas de manera independiente y cada uno de ellas es ponderada por un peso para luego pasar a la función sináptica. Cada peso (w_{ji}) representa la fuerza de cada conexión sináptica de la neurona anterior i a la siguiente j . Es precisamente en estos pesos donde se almacena la información que describe al fenómeno. Los pesos sinápticos se calculan durante el proceso de aprendizaje (descrito en la Sección 3.2.1) utilizando los patrones de un conjunto de muestras.

Una vez ponderadas las entradas por los pesos, la función sináptica incorpora todos los valores en un escalar único. Un ejemplo de una función comúnmente usada es un sumador del tipo:

$$v_j = \sum_i x_i w_{ji}. \quad (3.1)$$

El valor v_j se conoce como el *campo local inducido* de la neurona j . Este nivel de activación llega a la función *Somática* o de *Activación* cuya salida representa el nivel de activación de la neurona. Algunos ejemplos de funciones típicas utilizadas para este fin son las que se muestran en la Figura 3.2.

Al final, la salida de la función de activación pasa al elemento conectivo de salida de la neurona que típicamente se modela utilizando una función lineal.

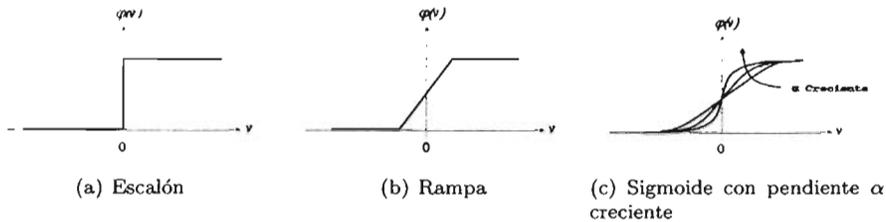


Figura 3.2: Funciones de Activación.

3.1.2. Topología de la red

La escasa capacidad de un elemento de procesamiento como el de una neurona no es suficiente para realizar procesos de clasificación o aproximación complejos, sin embargo, un conjunto de estos elementos interconectados si lo es. La forma en que se conectan más de uno de estos elementos define la *topología de la red* y existen diferentes clasificaciones de acuerdo a diversos factores como la dirección en la que fluye la información en la red, la conectividad y el número de capas.

Una de las clasificaciones mas generales considera tres tipos de arquitecturas:

1. Acíclicas de una capa: Esta es la instancia mas simple de una red neuronal con una arquitectura de capas. Consiste en un grupo de perceptrones que corresponden uno a uno con las salidas de la red y cuyas entradas provienen de los estímulos y no de las salidas de alguna neurona, constituyendo una red estrictamente “*hacia adelante*” o *acíclica*. Un ejemplo se muestra en la Figura 3.3, que en particular tiene 3 nodos de entrada y 3 perceptrones de salida. En estas redes sólo existe una capa de perceptrones.
2. Acíclicas multicapas: En esta clase, la característica principal es la existencia de una o más capas *ocultas*, cuya función es la de intervenir entre los estímulos externos de la red y las salidas de la misma. Esta(s) capa(s) adicional(es) permiten a la red extraer características mas complejas gracias a los conjuntos adicionales de pesos (y de conexiones sinápticas).

En este tipo de red, las salidas de los nodos de la capa de entrada son tomadas como las entradas de las neuronas de la segunda capa. Las salidas de las neuronas de la segunda capa se utilizan como entradas de la tercera y así sucesivamente hasta llegar a la capa de salida que constituye la respuesta final de la red al estímulo presentado en la capa de entrada.

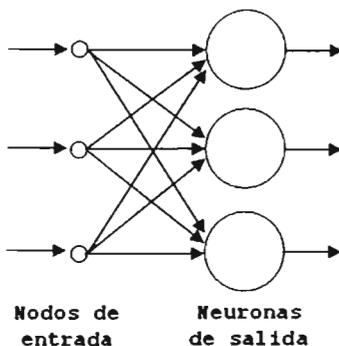


Figura 3.3: Red neuronal acíclica de una capa.

La Figura 3.4 muestra un ejemplo de una red acíclica de dos capas con 5 nodos de entrada, 4 neuronas en la capa oculta y dos en la de salida.

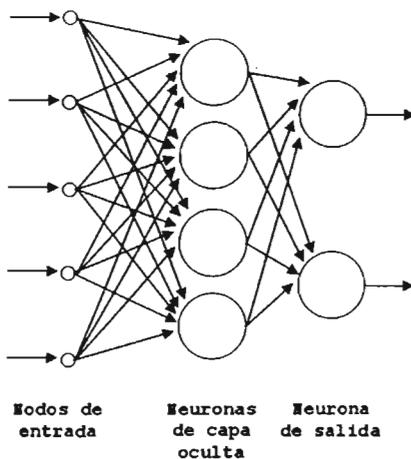


Figura 3.4: Red neuronal acíclica de dos capas.

De dicho ejemplo se puede decir también que la arquitectura de esa red es *fuertemente conectada*, pues todos los nodos de una capa anterior se encuentran conectados con los de la capa siguiente. Si se presenta el caso en que no todos los elementos de una capa anterior se encuentran conectados con la capa siguiente entonces se habla de una red *parcialmente conectada*.

3. Redes recurrentes: Este tipo de arquitecturas se distinguen de las acíclicas en que existe al menos un ciclo de retroalimentación. Al igual que las redes acíclicas, existe la posibilidad de construir redes recurrentes de una o mas capas. Los ciclos se pueden presentar entre nodos de capas diferentes, de la misma capa e incluso de un mismo nodo.

La presencia de ciclos de retroalimentación en una RN modifica profundamente las capacidades de aprendizaje y consecuentemente el desempeño. Su uso involucra conexiones que requieren elementos de retraso para posibilitar el cálculo de las salidas, y aunque las neuronas fueran elementos con comportamiento lineal, usando retroalimentación, el comportamiento general de la red se convierte en uno no lineal. La Figura 3.5 muestra una RN recurrente de dos capas.

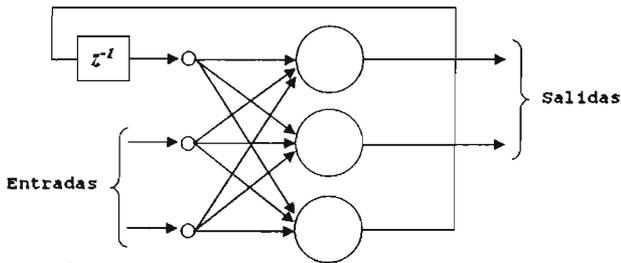


Figura 3.5: Red neuronal recurrente.

Es importante mencionar que la topología de la red se encuentra íntimamente relacionada con el algoritmo de aprendizaje utilizado para entrenarla.

3.1.3. Aprendizaje

Una vez seleccionada la arquitectura y el tipo de neuronas, es necesario entrenarla antes de usarla como un clasificador. Esto se logra presentando un conjunto de muestras una por una, en una etapa de entrenamiento. Al final de este proceso, el conocimiento adquirido por la red queda representado en los pesos de las conexiones que existen entre los nodos que la conforman (y no en ninguna otra aunque quizá fuera deseable).

El proceso de entrenamiento de una RN comienza comúnmente definiendo aleatoriamente el conjunto de pesos de todas las conexiones, seguido de la presentación uno a uno de los patrones que describen el fenómeno de interés,

ajustando los pesos hasta cumplir con algún criterio, como el alcance de una medida de error menor a un límite. Por último se calcula una segunda medida de error utilizando un conjunto de datos desconocido (diferente al de entrenamiento) para evaluar el éxito del aprendizaje y generalización del fenómeno en cuestión.

La selección de los patrones con los que se entrena a la RN es un aspecto particularmente importante pues estos deben describir por completo al fenómeno, sabiendo que la red sólo funcionará bien para el problema que le fue enseñado. Esto significa que deberá haber patrones de todas las clases e incluso de las fronteras entre las clases.

Existen diversos procesos de aprendizaje y en general se pueden clasificar en tres tipos:

1. **Aprendizaje supervisado:** En este tipo de aprendizaje se presentan a la red patrones reales, tanto de entrada como de salida. La red utiliza el patrón de entrada para calcular uno de salida que compara con el patrón real. Esto constituye una medida de error que se utiliza para calcular modificaciones a los pesos buscando reducirla.
2. **Aprendizaje no-supervisado:** Este tipo de aprendizaje requiere que la red encuentre automáticamente los criterios para realizar la clasificación, utilizando alguna medida de similitud entre los patrones de entrada y adaptando los pesos de acuerdo a esta medida.
3. **Aprendizaje estocástico:** Los métodos de este tipo utilizan procesos aleatorios y distribuciones de probabilidad para minimizar una función de energía definida con relación al desempeño de la red.

El conjunto de tipos de redes concebibles es sumamente grande dado que las opciones en su construcción son muy variadas: desde el tipo de neurona y la arquitectura, hasta el mecanismo utilizado para su entrenamiento. Más aún, las capacidades de cada tipo de red pueden estar profundamente relacionadas con alguna aplicación en particular. Sin embargo, existen algunos tipos de redes que son sumamente populares, como los mapas auto organizados de Kohonen [18], las redes de “teoría de resonancia adaptativa” de Carpenter y Grossberg [19] y por supuesto las redes de perceptrones, basadas en un concepto inicial de una sola capa presentado hace más de sesenta años por McCulloch y Pitts [20], que tardó más de 40 años en evolucionar a multicapas con Ackley et al. [21] y

todavía un poco más en encontrar un algoritmo de aprendizaje verdaderamente eficaz, hasta que el grupo de Rumelhart et al. [22] propusieron el uso de “retropropagación” (aunque el algoritmo fuera descrito más de 10 años antes en la tesis doctoral de Werbos).

Desde entonces las redes de multicapas de perceptrones han sido profundamente estudiadas y su comportamiento se ha descrito analíticamente. Entre sus principales ventajas se encuentran la gran simplicidad en su funcionamiento y su buen desempeño, particularmente en aplicaciones de clasificación con grandes conjuntos de datos, donde su comportamiento es tan bueno como el clasificador óptimo de Bayes [23, 24, 25, 26].

Por estas razones, las redes multicapas de perceptrones fueron seleccionadas para formar parte de la solución como la herramienta de clasificación, aunque la decisión también resultó influenciada por experiencias previas que arrojaron resultados positivos. La experimentación descrita en el Capítulo 5 (y en general, la de todo el trabajo de investigación) fue realizada utilizando únicamente este tipo de red y existe la posibilidad de que alguna otra produzca mejores resultados.

La siguiente sección describe su arquitectura, algoritmo de aprendizaje y funcionamiento.

3.2. Redes multicapas de perceptrones

Las *Redes Multicapas de Perceptrones* (RMP) son redes de arquitectura acíclica multicapas, con perceptrones cuya función acumulativa es lineal, la función de activación es alguna sigmoide no lineal suave (derivable) y la función de salida es lineal. El hecho de que existan una o más capas ocultas de perceptrones permite a la red aprender tareas complejas, ya que permite extraer características menos evidentes de los patrones de entrada. El que la función de transferencia sea no-lineal es importante pues de lo contrario la funcionalidad quedaría reducida a la de una red de una sola capa.

Estas características, combinadas con la posibilidad de aprender mediante algún algoritmo de entrenamiento, habilitan el poder de cómputo de una RMP. Uno de los algoritmos de aprendizaje más populares y funcionales hasta ahora es el de retropropagación.

3.2.1. Retropropagación

Desde su propuesta como algoritmo de aprendizaje, el algoritmo de retropropagación ha sido uno de los principales responsables del exitoso surgimiento de las redes neuronales. Su nombre proviene de la forma como se propaga una medida de error calculada en las capas de salida hacia las capas de atrás de la red, aunque en realidad el proceso de propagar hacia atrás se refiere a una parte del cálculo de un gradiente que es, de hecho, una operación común para muchos algoritmos.

Retropropagación es un algoritmo basado en la técnica del “descenso” en la dirección opuesta del gradiente, utilizando una medida de error de la red como valor a minimizar. La base del algoritmo consiste en calcular el gradiente de la función de error y dar un paso en la dirección opuesta (para minimizar en vez de maximizar), repitiendo cuantas veces sea necesario. Debido a que al menos localmente, el algoritmo se dirige hacia un valor cada vez mas pequeño, se espera llegar eventualmente a un mínimo.

Como en la mayoría de los algoritmos de aprendizaje, existen parámetros sensibles que afectan la probabilidad de éxito en el aprendizaje y junto con ellos existen también cientos de heurísticos que orientan su selección. Además, en la búsqueda por acelerar estos procesos y de hacerlos mas robustos, se han propuesto modificaciones, algunas muy funcionales y otras intrascendentes. El funcionamiento del algoritmo se detalla a continuación.

Una vez definida la arquitectura, inicializados los pesos y seleccionado el conjunto de patrones de entrenamiento es posible comenzar el algoritmo de retropropagación. En primer lugar se presenta uno de estos patrones a la red, aplicando los valores a los nodos de entrada. La información se propaga hacia adelante por toda la red y se produce un vector de salida O . Este vector se compara con los valores que la red debería producir y la diferencia provee una medida de error para cada una de las neuronas de salida. El objetivo de todo el proceso es minimizar la energía del error ε de toda la red:

$$\varepsilon = \frac{1}{2} \sum_{v_i} (z_i - o_i)^2, \quad (3.2)$$

donde z_i son los valores deseados. El algoritmo lo logra modificando los pesos en la dirección de máximo descenso, calculado con base en el gradiente de la

Ecuación 3.2 con respecto a cada uno de los pesos de la red, es decir:

$$\Delta w_{ij} = -\eta \frac{\partial \varepsilon}{\partial w_{ij}}, \quad (3.3)$$

donde η es el coeficiente de aprendizaje que define el tamaño del paso para cada corrección.

Para calcular esta derivada parcial de la energía del error (ε) con respecto al peso (w_{ij}), se utiliza la regla de la cadena, ya que la parcial de ε esta en función del error (e) de la neurona j , que se obtiene utilizando:

$$e_j = z_j - o_j. \quad (3.4)$$

Como se puede apreciar, e_j esta en función del valor observado (o) en la neurona j y éste a su vez en función de la suma de los estímulos ponderados por los pesos v_j (ecuación 3.1), que ciertamente incluye a w_{ji} , es decir:

$$\frac{\partial \varepsilon}{\partial w_{ij}} = \frac{\partial \varepsilon}{\partial e_j} \frac{\partial e_j}{\partial o_j} \frac{\partial o_j}{\partial v_j} \frac{\partial v_j}{\partial w_{ij}}. \quad (3.5)$$

Calculando las derivadas y sustituyendo en 3.3 queda:

$$\Delta w_{ij} = -\eta \delta_j x_j, \quad (3.6)$$

donde δ_j representa los errores de las capas siguientes y se obtiene utilizando la ecuación:

$$\delta_j = \phi'(v_j) \sum_k \delta_k w_{jk}, \quad (3.7)$$

en la cual ϕ' es la derivada de la función de activación y x_j es el estímulo de entrada. Se puede observar entonces que para calcular el error en una neurona de la capa k , se deben conocer todos los errores de la capa $k + 1$, es decir, primero se calcula el error en la capa de salida y luego se propaga hacia la capa anterior, repitiendo el proceso hasta la capa de entrada. Para el momento en que se alcanza la primera capa, todos los ajustes han sido calculados y es hasta entonces cuando son modificados todos los pesos utilizando las Δw_{ij} obtenidas durante el proceso.

Ejemplo

En esta sección se presenta un ejemplo de cómo calcular el ajuste en una neurona. La topología de la red consiste en la capa de entrada, una capa oculta y una de salida, cada capa con tres nodos. El ajuste que se calcula corresponde al peso que existe entre el primer nodo de entrada y el primer nodo oculto. La Figura 3.6 muestra los detalles del problema. Las líneas claras son pesos que existen en la arquitectura pero que resultan intrascendentes para el cálculo del ajuste al peso en cuestión. Las líneas oscuras muestran pesos que influyen para el cálculo. La línea mas gruesa muestra la conexión cuyo peso se pretende ajustar. Todos los nodos fueron numerados consecutivamente para permitir las referencias durante la explicación. Las capas también fueron identificadas con literales con el mismo fin. Utilizando la nomenclatura, el peso cuyo ajuste se pretende calcular es w_{41} y va de la capa i a la j .

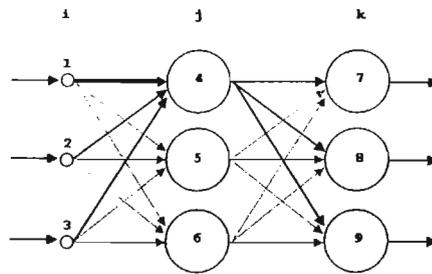


Figura 3.6: Cálculo de Δw_{41} .

Sustituyendo nuestros valores sobre la Ecuación 3.6, queda:

$$\Delta w_{41} = -\eta \delta_4 x_1, \quad (3.8)$$

asumiendo que el valor de la tasa de aprendizaje es η y que la entrada que llega por el nodo 1 es x_1 . Ahora, es necesario calcular todos los errores de las capas posteriores, en este caso los de la capa k , para obtener δ_4 .

Estos quedan:

$$\begin{aligned} \delta_7 &= \phi'(v_7) \cdot (z_7 - o_7) \\ \delta_8 &= \phi'(v_8) \cdot (z_8 - o_8) \\ \delta_9 &= \phi'(v_9) \cdot (z_9 - o_9) \end{aligned} \quad (3.9)$$

Luego, los errores δ_7 , δ_8 y δ_9 se propagan a la capa j para calcular δ_4 :

$$\delta_4 = \phi'(v_4) \cdot (\delta_7 w_{74} + \delta_8 w_{84} + \delta_9 w_{94}) \quad (3.10)$$

En donde sólo falta calcular v_4 :

$$v_4 = x_1 w_{41} + x_2 w_{42} + x_3 w_{43} \quad (3.11)$$

Así se puede obtener Δw_{41} utilizando la Ecuación 3.8.

3.2.2. Neurona de sesgo

En una RMP existe un límite que define si una neurona debe responder activamente a los estímulos o no, justo como en las neuronas del cerebro humano que se excitan sólo si la suma de los estímulos supera cierto valor. En el pasado se pensaba que cada uno de estos límites debería ser fijado a mano para cada neurona pero el problema encontró una solución mucho más práctica. Esta solución consiste en incluir una neurona artificial de entrada cuyo valor es siempre uno. Esta *neurona de sesgo* se conecta a todas las demás neuronas (de todas las capas) y el peso asociado a cada conexión actúa como el límite para excitar a esa neurona. La gran ventaja consiste en que se puede utilizar el algoritmo de retropropagación para ajustar también estos pesos como si fueran conexiones normales de la red, evitando la necesidad de ajustarlos a mano.

3.2.3. Modificaciones a Retropropagación

Las siguientes son dos modificaciones que incrementan el desempeño del algoritmo de retropropagación. Una reduce el número de iteraciones necesarias para completar el entrenamiento, la otra incrementa la capacidad de generalización de la red.

Inercia o Momento

Como en casi cualquier algoritmo basado en el gradiente, el problema de definir adecuadamente las tasas de aprendizaje es complicado. La función de error que se intenta minimizar generalmente guarda una relación totalmente no-lineal con los pesos de la red. Sin embargo, el hecho de que se utilice una función lineal de una derivada parcial para modificarlos implica que se asume una función de error al menos localmente lineal. Consecuentemente, la tasa de aprendizaje debe

de representar el vecindario en donde se presume esta linealidad, motivando la selección de tasas de aprendizaje tan pequeñas como sea posible.

Las consecuencias de utilizar tasas de aprendizaje muy pequeñas son un elevado tiempo de cómputo y la posibilidad de alcanzar un mínimo local y permanecer ahí indefinidamente. Existe una solución a este problema y consiste en la inclusión de un término de inercia o momento, modificando la ecuación para calcular los ajustes de la siguiente forma:

$$\Delta w_{ij}(n) = -\eta \delta_j x_j + \mu \cdot \Delta w_{ij}(n-1). \quad (3.12)$$

En esta ecuación, μ es el coeficiente de inercia o momento, su valor esta entre cero y uno, generalmente cercano a uno. La modificación permite realizar ajustes considerando la magnitud y dirección de las $(n-1)$ modificaciones anteriores al momento de calcular la modificación actual (n) , eliminando los efectos oscilatorios y fortaleciendo las tendencias de largo plazo. El efecto final es un aprendizaje mucho mas rápido incluso con tasas de aprendizaje muy pequeñas.

Disminución de pesos

Una red neuronal cuyos pesos tienen magnitudes pequeñas es capaz de hacer mejores generalizaciones [27]. Para mantener los pesos en magnitudes pequeñas se utiliza un algoritmo de disminución de pesos que consiste en dos modificaciones al algoritmo original. La primera es la incorporación de la suma de los pesos a la medida del error general de la red, así en el proceso de minimizar esta medida, los pesos también se hacen mas pequeños. La segunda consiste en multiplicar los pesos por una constante ligeramente menor que uno en cada etapa de aprendizaje, disminuyendo la magnitud de cada peso en una forma similar al de una exponencial. La Ecuación 3.13 muestra como queda el cálculo de Δw_{ij} utilizando disminución de pesos (y momento).

$$\Delta w_{ij}(n) = d \cdot (-\eta \delta_j x_j + \mu \cdot \Delta w_{ij}(n-1)). \quad (3.13)$$

En esta ecuación, los valores típicos de d se encuentran entre $1.0 - 1.0 \times 10^{-6} = 0.999999$ y $1.0 - 1.0 \times 10^{-8} = 0.99999999$. El efecto de disminución de pesos decrece cuando $d \rightarrow 1$.

Estas dos son las modificaciones mas comunes al algoritmo de retropropagación y se incluyen en la explicación porque fueron implementadas en el algoritmo

de aprendizaje utilizado en el proyecto. En la sección de resultados se presentan algunos ejemplos del desempeño del algoritmo aplicando las modificaciones y algunas comparaciones.

Capítulo 4

Metodología

En este capítulo se presenta la metodología de análisis de señales ultrasónicas para detección de fallas. En el proceso se utiliza una red neuronal que explota la información de las señales para realizar el análisis, pero en las etapas de preprocesamiento se utilizan también algunas definiciones de la física asociada al fenómeno de propagación de la energía acústica en un medio físico. El capítulo incluye todas las etapas del proceso de análisis, partiendo del conjunto de datos tomados directamente del convertidor analógico-digital conectado al transductor ultrasónico hasta llegar a las imágenes resultantes que describen características de las fallas.

Dentro del proceso, la red neuronal es la herramienta principal para realizar el análisis. Sin embargo, los datos que la alimentan deben mantener ciertas características cuya observación define la calidad de los resultados pues condicionan la capacidad de aprendizaje de la red. La primera sección de este capítulo presenta la descripción de estas etapas de preprocesamiento. La segunda muestra como se utilizan los datos preprocesados para lograr el entrenamiento de la red y la tercera como utilizar la red entrenada para analizar conjuntos de datos.

4.1. Preprocesamiento

El objetivo de la etapa de preprocesamiento en un problema de clasificación consiste en identificar la información relevante en los datos con el fin de facilitar la distinción entre clases. Se pone especial atención en factores como el ruido, la redundancia o la escasa capacidad descriptiva de la fuente y se apli-

can algoritmos conocidos orientados a corregir tales defectos. En esta etapa se realizan también las transformaciones necesarias para adecuar el conjunto de datos pensando en su ingreso a una red neuronal, utilizando operaciones como escalamiento o normalización e incluso se calculan transformaciones que enfatizan las características del fenómeno de interés, como la envolvente (usando la transformada de Hilbert) o la distribución de las componentes frecuenciales (utilizando análisis de Fourier).

Para el desarrollo de esta investigación se encontró una amplia gama de factores que pueden ser corregidos o mejorados en la etapa de preprocesamiento. De todos estos se evaluó de manera cualitativa su utilidad y practicidad para decidir su inclusión. En algunos casos se encontró que las mejoras que producen dichos algoritmos no contribuyen de manera substancial al proceso de análisis y al final fueron excluidos del proceso [28] buscando además mantener la eficiencia.

Las operaciones que sí fueron incluidas son: *a)* enfatizar las características importantes de la señal mediante el cálculo de la envolvente, *b)* reducir los efectos de la atenuación, *c)* estandarizar las entradas de la red neuronal llevando los valores del conjunto de datos al intervalo $[0, 1]$, *d)* establecer condiciones de homogeneidad en la información que entra a la red y *e)* construir las muestras de tamaño específico. Los detalles de cada operación se muestran a continuación.

4.1.1. Envolvente

En ensayos no destructivos, la *envolvente* de una señal ultrasónica es comúnmente utilizada como un mecanismo para encontrar la posición relativa de una falla al interior de una pieza [29] pues, en esta señal, es posible observar un pico o máximo local situado en la vecindad de la falla. La razón consiste en que la envolvente representa la magnitud de la energía cinética [30, 31], debida al movimiento de las partículas al interior del material [32], donde la correlación entre las diferencias de amplitud y las diferencias en densidad del material es mucho mas alta, comparada con la señal original que representa la energía potencial debida a la deformación del medio físico (elástico).

Para calcular la envolvente de la señal en \mathbb{R} se utiliza la transformada de Hilbert:

$$\mathbf{H}[y(t)] = \frac{1}{\pi} \int \frac{y(\tau)}{t - \tau} d\tau, \quad (4.1)$$

pero también existe un algoritmo para obtener $\mathbf{H}[y(n)]$ de una función en \mathbb{N} basado en la transformada discreta de Fourier. Para lograrlo es necesario seguir

una serie de pasos y el primero consiste en obtener la FFT de $y(n)$ usando:

$$Y(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} y(n) e^{-j2\pi kn/N}. \quad (4.2)$$

Luego es necesario pasar el resultado $Y(k)$ por la siguiente función:

$$T(k) = \begin{cases} 0 & k < N/2 \\ Y(k) & k = N/2 \\ 2Y(k) & k > N/2. \end{cases} \quad (4.3)$$

Por último, se obtiene $H[y(n)]$ calculando la transformada inversa de Fourier de $T(k)$:

$$\mathbf{H}(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} T(k) e^{+j2\pi nk/N}. \quad (4.4)$$

Cabe mencionar que este proceso únicamente enfatiza las características útiles para el análisis que se realiza con la red neuronal, manipulando el conjunto de datos mediante un proceso matemático. Es decir, el cálculo de la transformada de Hilbert no agrega ninguna información al conjunto de datos.

La energía de la señal que resulta de este proceso se encuentra linealmente relacionada con la energía de la señal original y la magnitud de las componentes se encuentra en un rango cuyos límites no se conocen. El siguiente paso será estandarizar la magnitud de estas componentes previo a su ingreso a la red neuronal.

4.1.2. Normalización

Las redes neuronales de perceptrones tienen en su mayoría funciones de activación (Ver Sección 3.1.1) que poseen un rango sensible limitado, situado en un intervalo centrado en el origen.

Con el fin de explotar al máximo las capacidades de clasificación de la red al permitir que los pesos se ajusten en los rangos de mayor sensibilidad, las entradas deben encontrarse en un rango limitado y cercano al origen.

En este caso se decidió realizar un escalamiento al intervalo $[0, 1]$ que aunque tiene un sesgo hacia \mathbb{R}^+ , contribuye en el cálculo del campo local inducido sin sesgo, pues las entradas se multiplican por pesos cuyo valor inicial es aleatorio en el intervalo $[-1, 1]$. Este proceso estandariza también a los conjuntos de datos por posibles diferencias en parámetros asociados con el proceso de adquisición,

como la magnitud del pulso inducido o la ganancia del transductor.

Como en el caso anterior, es igualmente necesario señalar que el escalamiento no agrega ninguna información al conjunto de datos.

4.1.3. Corrección por atenuación

La atenuación es uno de los fenómenos presentes en la propagación de la energía acústica (ver Sección 2.1.2). Es posible caracterizarla utilizando la ecuación:

$$A_z = A_0 e^{-\alpha z}. \quad (4.5)$$

Donde A_0 es la amplitud de la señal ultrasónica en algún punto y A_z es la amplitud reducida después de que la señal ha viajado una distancia z del punto original.

La interfaz permite seleccionar dos puntos donde se observan elementos reflectivos similares a profundidades diferentes para obtener α , que luego se utiliza en una función inversa para calcular una nueva señal en la cual las amplitudes son comparables sin importar la distancia al transductor.

4.1.4. Homogeneidad

El problema de generar muestras homogéneas a partir de conjuntos de datos que no lo son, produjo uno de los retos más importantes de este trabajo de investigación y se ve en la propuesta de solución una de las principales contribuciones. Los principios en los que se basa tal propuesta se encuentran totalmente relacionados con la física del fenómeno de propagación de una onda acústica en un medio físico. Sin embargo, la información agregada de manera implícita al conjunto de datos resulta de suma importancia para el análisis de señales ultrasónicas mediante redes neuronales.

Las fuentes de heterogeneidad en señales ultrasónicas son muchas y muy diversas, así como el efecto que producen. Diferencias en el ángulo de visión del transductor, en el material que se analiza, en la frecuencia de muestreo o en el tamaño del paso (ver Sección 2.3.2), son algunos de los factores que producen conjuntos de datos completamente heterogéneos. Incluso en un mismo conjunto de datos, donde todos estos parámetros se mantienen constantes, un par de muestras pueden resultar heterogéneas debido a que la distancia de cada una de estas con respecto al transductor modifican su especificidad.

Esta sección describe todos los factores (contemplados) que introducen heterogeneidad en los conjuntos de datos y la forma en que esas diferencias fueron eliminadas. La solución que se propone consiste en calcular para cada punto (p), una *región muestral* (h_p) que incluya la mayoría de los puntos que de alguna forma contienen información relacionada con p , para luego construir una muestra (X_p) que describa a p de forma homogénea (utilizando el mecanismo de interpolación de la siguiente sección).

La región muestral h_p

Dado que un transductor ultrasónico registra información de una región y no sólo de un punto, se produce una situación de ambigüedad, donde una única muestra no describe por completo una región y tampoco una muestra describe una región única.

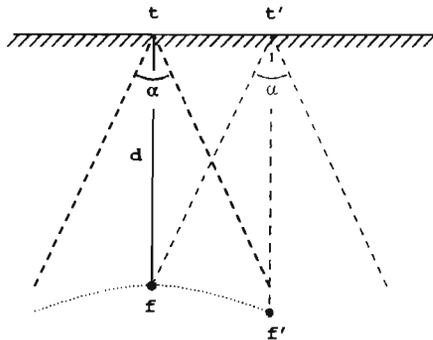


Figura 4.1: Arco de visión de un transductor ultrasónico.

La Figura 4.1 muestra las fuentes de ambigüedad mediante un ejemplo. El punto f representa una falla puntual localizada a una distancia d de la superficie. Un transductor con un ángulo de visión α se sitúa en la posición horizontal t y registra la falla sin distorsión. Otro transductor de características idénticas se localiza en la posición t' y debido a que la falla situada en f produce un reflejo de energía dentro del rango de visión del transductor situado en t' , éste también la registra, aunque con una distorsión en su posición (f'). Este es un ejemplo de como el punto f puede estar descrito por muestras en diferentes regiones.

La región en donde existe información relacionada al punto f se muestra como el arco punteado de la Figura 4.1 y, de manera general, dicha región se describe idealmente con una hipérbola, aunque para hacerlo en función de la

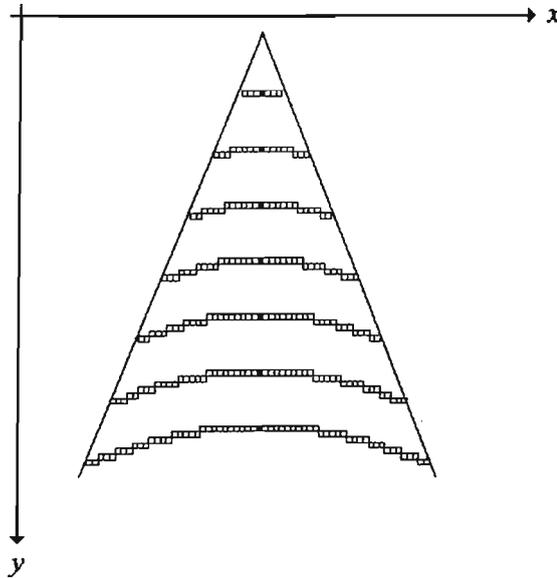


Figura 4.2: Arcos muestrales a diferente profundidad con α constante.

profundidad (d), la posición del transductor (τ) y el ángulo de visión (α) que son parámetros conocidos en análisis ultrasónicos, es posible utilizar las ecuaciones 4.6 y 4.7.

$$y(d, \theta) = d\sqrt{1 + [\sin(\theta/2)]^2} \quad (4.6)$$

$$x(d, \theta, \tau) = \tau + d \sin(\theta/2) \quad (4.7)$$

Con $-\alpha/2 \leq \theta \leq \alpha/2$ y donde x y y son posiciones absolutas (horizontal y vertical respectivamente) en la pieza de análisis.

La Figura 4.2 muestra un conjunto de ejemplos de los arcos muestrales que se generan a diferentes distancias de un transductor con una α definida. Note que los ejemplos consideran el hecho de que las señales sean discretas. Idealmente, el punto correspondiente al cuadrado obscuro de cada arco es completamente descrito por la región de los cuadrados claros.

Resolución

Otra fuente importante de heterogeneidad en los conjuntos de datos es la resolución que, además de definir factores típicos como el tamaño del detalle

mas pequeño que se aprecia en una señal, en este caso define también la forma de los arcos muestrales. La resolución en una señal digital está definida por la frecuencia de muestreo y, en el caso de señales ultrasónicas de ensayos no destructivos existen al menos dos frecuencias de muestreo importantes. La primera asociada al tiempo que existe entre cada muestra tomada de la señal producida por el transductor (f_s) y la segunda asociada al desplazamiento lineal del transductor (f_t) para obtener trazas adyacentes. Utilizando f_s es posible calcular la

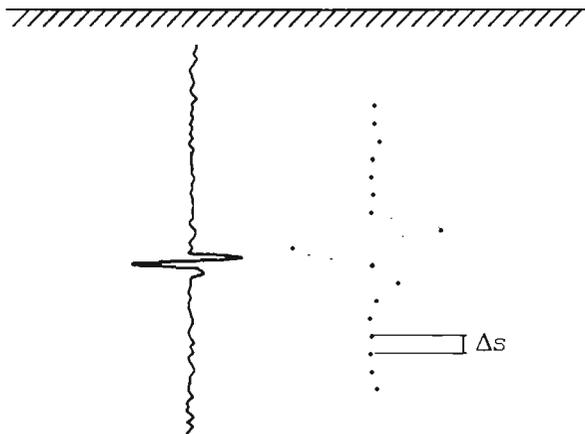


Figura 4.3: Resolución en profundidad (Δs).

distancia Δs que existe entre cada muestra de una misma traza (en la dirección a la que apunta el transductor), aunque además es necesario conocer la velocidad de propagación del sonido (c) dentro del medio de estudio. Para hacerlo se debe considerar que la información importante en este tipo de análisis es la que se refleja, de alguna forma, desde el interior de la pieza, es decir, la energía inducida hace un viaje desde el transductor hasta la causa del eco y otro de regreso. Así, la ecuación para calcular la resolución Δs queda:

$$\Delta s = \frac{c}{2f_s} \quad (4.8)$$

La Figura 4.3 muestra la interpretación física que se puede dar a Δs : cada muestra tiene una profundidad asociada a la pieza o al medio de análisis.

Con respecto a la frecuencia de muestreo f_t y a la distancia Δt , solo es necesario conocer el tamaño del paso que se utilizó durante el proceso de adquisición pues es una variable controlada en ese momento. La Figura 4.4 muestra

la interpretación física de Δt .

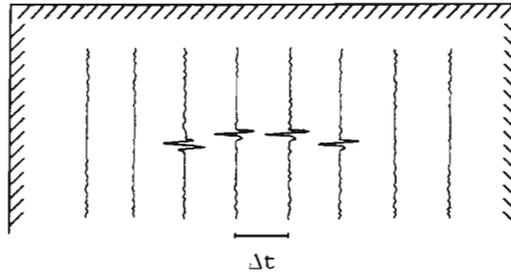


Figura 4.4: Resolución Horizontal (Δt).

Conociendo Δt y Δs , es posible hacer un mapeo entre la posición de una muestra y la información de ésta dentro de la pieza física, permitiendo que las descripciones tengan una relación espacial congruente con la realidad.

4.1.5. Interpolación

Hasta ahora se ha definido una región homogénea que de alguna manera elimina las diferencias en los parámetros de adquisición de una señal ultrasónica y al mismo tiempo, incluye la información más relevante para cada punto. Sin embargo, esta región no puede ser utilizada eficazmente como el conjunto de datos de entrada en una red neuronal de perceptrones pues su tamaño cambia de un punto a otro. Para solucionar este problema se propone utilizar un mecanismo de interpolación de donde sea posible tomar un número constante de muestras.

La tarea de seleccionar el mejor mecanismo de interpolación no es trivial, pues la literatura reporta una amplia variedad de métodos con características muy diversas. En este caso se decidió utilizar un interpolador basado en *Splines Naturales* (SN) ya que sus cualidades se alinean con las necesidades de la aplicación, aunque posiblemente existan otros mecanismos de interpolación que superen su desempeño. Las características más importantes de los SN son: estabilidad gracias a un polinomio de grado tres, eficiencia en el cálculo y curvatura mínima.

1. Polinomios de grado 3: Comparados con un polinomio de aproximación en donde el grado crece hasta $n - 1$ para interpolar n datos, el grado de un SN es siempre constante, pero aún más importante es el hecho de que los

valores interpolados no sean inestables, es decir, que se mantengan cerca de los valores reales.

2. Eficiencia de cálculo: Utilizando la Ecuación 4.9 es posible calcular los coeficientes del spline en un algoritmo de complejidad $O(n)$.

$$\begin{aligned}
 S_i(x) = & \frac{s_i}{6h_i}(x_{i+1} - x)^3 \\
 & + \frac{s_{i+1}}{6h_i}(x - x_i)^3 \\
 & + \left(\frac{f_{i+1}}{h_i} - \frac{s_{i+1}h_i}{6} \right) (x - x_i) \\
 & + \left(\frac{f_i}{h_i} - \frac{s_i h_i}{6} \right) (x_{i+1} - x)
 \end{aligned} \tag{4.9}$$

Teniendo los coeficientes, el tiempo necesario para realizar k interpolaciones es del orden de $O(k \log n)$.

3. Curvatura mínima: Existen varios tipos de splines cuyas diferencias se basan en las condiciones iniciales utilizadas para su construcción. En el caso de los splines naturales estas condiciones son $S''(x_0) = S''(x_n) = 0$ y esta característica permite definir funciones con curvatura mínima.

Una vez calculados los coeficientes del spline con la información de la región muestral (h_p), la muestra asociada a un punto p se construye tomando un número constante n de interpolaciones equidistantes. La Figura 4.5 muestra dos ejemplos de este proceso, uno en el que la región muestral tiene cardinalidad ($|h_p|$) mayor al número de elementos (n) de la muestra ($|h_p| \approx 60$, $n = 10$) y otro en el que pasa lo contrario, $|h_p| < n$ ($|h_p| = 7$, $n = 10$). En ambos casos los elementos de las muestras se toman del spline, mientras que los datos reales sólo se utilizan para su construcción.

Por último, cabe mencionar que la correcta selección de n resulta relativamente importante pues no sólo determina el tamaño de cada muestra sino también el número de neuronas de la primera capa de la red. Una n grande introduce redundancia innecesaria, al tomar muestras con componentes muy cercanas entre si, aparte de que incrementa el número de operaciones necesarias para entrenar y usar la red. Una n pequeña no sería capaz de describir apropiadamente la información del fenómeno. El capítulo de experimentación muestra que (afortunadamente) el algoritmo no resulta extremadamente sensible a cambios en este parámetro.

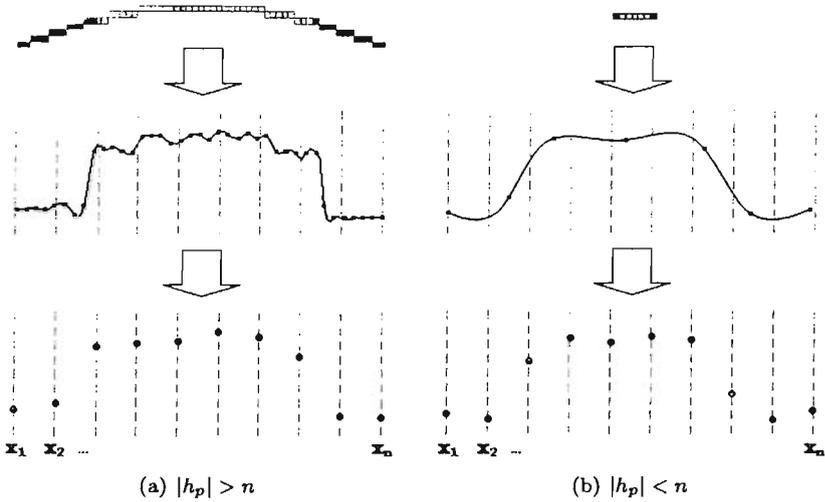


Figura 4.5: La región muestral (arriba), el spline (en medio) y la muestra (abajo).

4.2. Red Neuronal: Construcción

El entrenamiento es el proceso en el que una red neuronal aprende y generaliza el fenómeno de interés. Para ello es necesario definir un número importante de muestras con las que se describe el fenómeno de la mejor manera posible. Esta sección muestra como construir este conjunto con base en *datos reales* y la forma en que se utiliza información de las fallas conocidas en una pieza para generar *datos ideales*, necesarios ambos durante el entrenamiento. Se describe también la interfaz que se propone con el fin de definir estos conjuntos y por último se describe el algoritmo de entrenamiento de la red y los detalles de su implantación.

4.2.1. Datos reales e ideales

Los *datos reales* corresponden a aquellos obtenidos directamente de la señal ultrasónica, aunque en nuestro caso se utilizan aquellos que fueron previamente procesados con los procedimientos anteriormente descritos. En cambio los *datos ideales* corresponden a un conjunto de datos con características idénticas a los reales (número de muestras, resoluciones, etc.) pero cuya información es sintéticamente construida a partir de la descripción que existe de las fallas conocidas de la pieza, como su posición, tamaño y forma. La Figura 4.6 muestra un con-

junto de datos reales y el conjunto de datos ideales asociado. En este caso, la pieza tiene tres fallas creadas de manera artificial, que consisten en perforaciones de 4.2 mm de diámetro.

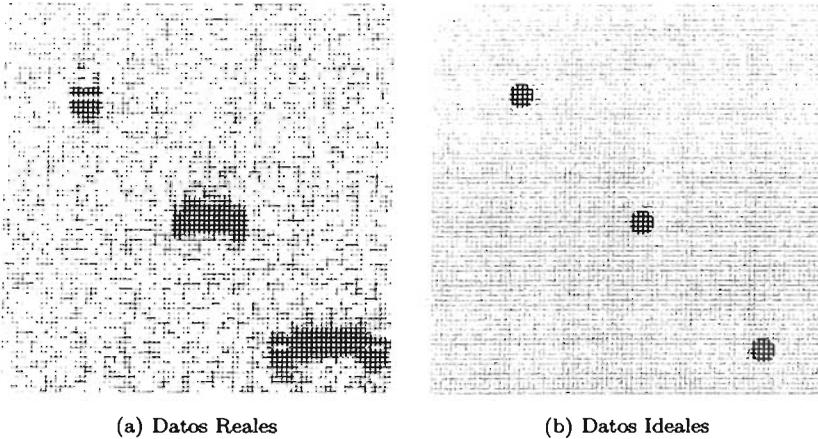


Figura 4.6: Conjuntos de datos.

Para crear los conjuntos de datos artificiales se aprovecha el mapeo que existe entre la posición física en la pieza y el índice de cada muestra en el conjunto de datos reales, permitiendo la representación exacta de la falla en un conjunto de datos sintético. Este proceso requiere de la interacción de un usuario que, al observar la pieza, ingresa al algoritmo la información de las fallas; pero una vez terminado el entrenamiento de la red, usando Δt , Δs y α , el análisis se realiza de manera automática.

4.2.2. Conjuntos de entrenamiento

El conjunto de entrenamiento es quizá el componente mas importante de todo el proceso de creación de una red neuronal, pues constituye el único medio por el que ésta obtendrá información del fenómeno analizado. Siendo así, resulta igualmente importante cuidar que este conjunto sea una descripción correcta del fenómeno, que sea completo dentro de lo posible y que no contenga información irrelevante.

Dentro del contexto de análisis de señales ultrasónicas, un conjunto de entrenamiento descriptivo se puede pensar en función de las combinaciones que existen entre los datos reales y lo que se espera de la red (datos ideales), para

incluir al conjunto muestras de todas las combinaciones. Es lógico pensar que deberá haber muestras de regiones en las que los datos reales sugieren la presencia de una falla donde de hecho la hay, pero también es necesario incluir muestras de regiones en donde no parece haberla y que en realidad no la hay. Así, la red aprende a distinguir correctamente “verdaderos positivos” y “verdaderos negativos”. Asimismo es importante incluir muestras de regiones en las que aparentemente existe una falla y en realidad no la hay y viceversa, enseñando a la red a distinguir “falsos positivos” y “falsos negativos”. Ciertamente, es indispensable que toda esta información sea correcta.

Con el objetivo de que el conjunto sea completo, es deseable incluir muestras de situaciones eventuales, ya que contribuirán a construir un clasificador más robusto, pero también es necesario evitar información redundante, ya sea eliminando componentes altamente correlacionadas dentro de todas las muestras o descartando muestras de situaciones irrelevantes.

4.2.3. Entrenamiento

El entrenamiento de una red neuronal es el proceso en el que se abstrae el fenómeno a partir de un conjunto de muestras ajustando los pesos de las conexiones entre las neuronas. El algoritmo utilizado es *retropropagación*, aunque su aplicación es sólo una parte del proceso. La secuencia de pasos completa está descrita en el diagrama de flujo de la Figura 4.7. Mas adelante se describe cada paso.

Secuencia aleatoria

El conjunto de datos de entrenamiento se construye definiendo regiones que incluyen características importantes del fenómeno y por esta razón, frecuentemente existen secuencias ordenadas de datos dentro de este conjunto, es decir, se encuentra primero un subconjunto donde sólo se aprecia una característica y luego otro donde se aprecia otra. Pensando en que el siguiente paso consiste en definir dos subconjuntos en los que deben existir muestras que caractericen el fenómeno completo, se construye una secuencia aleatoria de muestras intentando eliminar cualquier orden dentro del conjunto. El algoritmo que se utiliza intercambia la posición de todos los registros, seleccionando el destino de cada uno de manera aleatoria y cuidando, con miras a reducir el tiempo de ejecución, que los registros ya cambiados (por ser el destino de alguno anterior), no sean cambiados por segunda vez.

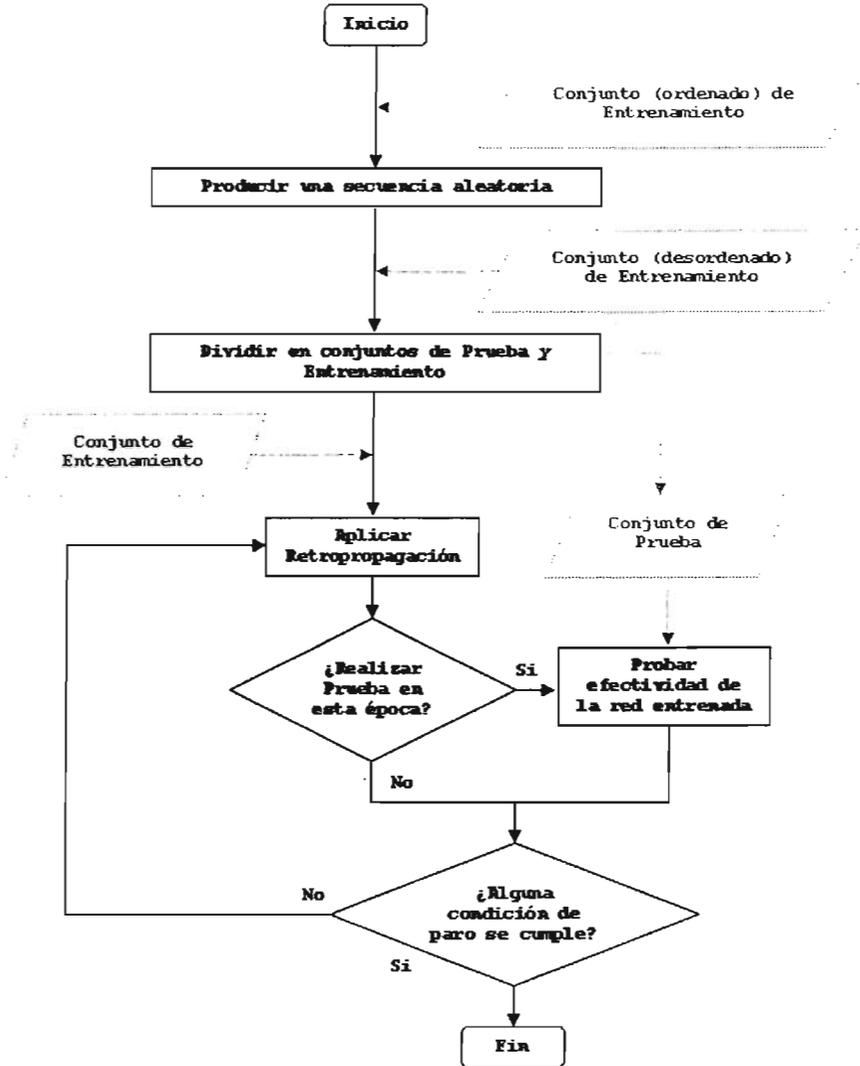


Figura 4.7: Diagrama de flujo del entrenamiento de la red.

Subconjuntos de entrenamiento y prueba

Es necesario evaluar periódicamente el desempeño de la red conforme avanza el entrenamiento. Para hacerlo existen dos opciones comúnmente usadas: una en la que se utilizan las mismas muestras con las que se efectúa el entrenamiento y otra donde se utilizan muestras distintas. La primera resulta atractiva pues se puede utilizar el conjunto de datos original completo y así lograr posiblemente una mejor descripción del fenómeno, el problema es que al hacer la evaluación se obtiene un indicador de que tan bien aprende la red a predecir ese (y posiblemente sólo ese) conjunto de datos. En cambio, si el proceso de evaluación se realiza con un conjunto diferente al de entrenamiento, la evaluación está mucho más relacionada con el nivel de aprendizaje de la red para describir el fenómeno (y no sólo un conjunto de datos). Así, resulta interesante la idea de sacrificar algún porcentaje del conjunto de datos original y dejarlo como el *conjunto de prueba* y el resto efectivamente utilizarlo como el *conjunto de entrenamiento*. Algorímicamente esto se logra utilizando un índice que indica donde terminan los datos de entrenamiento y comienzan los de prueba.

Retropropagación

Esta sección presenta los detalles de implantación de esta fase del entrenamiento (la descripción del algoritmo se encuentra en la Sección 3.2.1). Todas las funciones utilizadas fueron desarrolladas con la intención de permitir modificaciones a casi todos los valores de la aplicación para poder realizar una experimentación exhaustiva. La interfaz, mostrada en la Figura 4.8, permite definir el número de neuronas (en la capa intermedia e incluso en la de entrada), el tipo de función de activación de cada capa, las tasas de aprendizaje, las proporciones de los conjuntos de entrenamiento y prueba, las condiciones de paro y la frecuencia con que se debe probar el desempeño alcanzado por la red. Cabe señalar que la interfaz muestra valores por defecto, evitando la necesidad de ingresarlos cada vez. La descripción de como fueron implantados todos los pasos se hará siguiendo la secuencia de ocurrencia en la aplicación.

Comenzando por la definición de la arquitectura de la red, se utiliza una red de tres capas y de éstas el único parámetro fijo es el tamaño de la capa de salida, que consiste en una única neurona que indica la existencia o no de una falla en el punto asociado a la muestra de entrada. Los parámetros libres son: el tamaño de la capa de entrada, definido al momento de indicar cuantas muestras se toman del spline para construir el conjunto de datos de entrenamiento, y

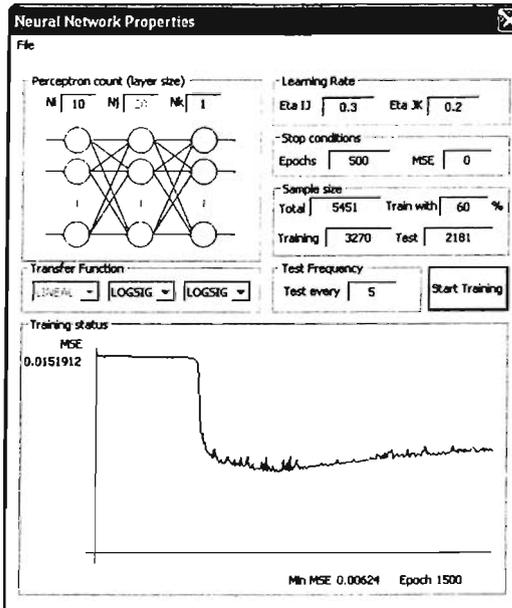


Figura 4.8: Interfaz de las propiedades de la red neuronal.

el tamaño de la capa intermedia, que debe ser definida en cualquier momento antes de empezar el entrenamiento. Estos son dos parámetros importantes que condicionan la capacidad de aprendizaje de la red. La sensibilidad de la red al tamaño de la capa de entrada (y entonces de la muestra de entrada) se describe en la Sección 4.1.5 y está relacionada con la redundancia o la falta de información. La sensibilidad de la red al tamaño de la capa intermedia, esta relacionada con la complejidad del fenómeno que se pretende representar con la red, pero hasta ahora no existen formas analíticas para encontrarlo y su selección se basa en heurísticos combinados con experiencia.

El siguiente paso es la inicialización y se ejecuta antes de aplicar retropropagación por primera vez. En este paso se definen los pesos iniciales de las conexiones neuronales conforme a la función de activación que se pretende utilizar en cada capa. Para la función Logarítmica Sigmoide, se utilizan pesos aleatorios en el rango $[0.1, 0.9]$, para la Tangente Sigmoide el rango es $[-0.1, 0.1]$ y para la Lineal $[0, 1]$. Todo esto con el fin de mantener los valores cercanos a la región sensible de la función de transferencia.

Una vez definida la arquitectura e inicializados los pesos, es posible comenzar

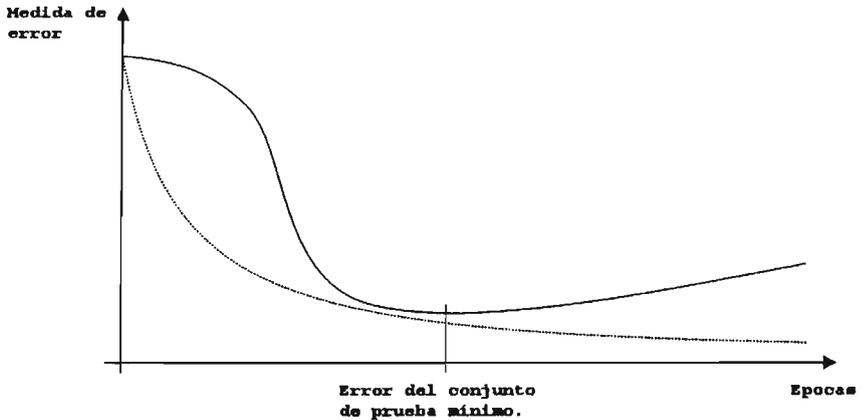


Figura 4.9: Evolución de la medida de error del conjunto de entrenamiento (línea punteada) y del conjunto de prueba (línea continua).

el entrenamiento de la red. Para ello existen dos estrategias de aprendizaje, una en la que todas las muestras son presentadas a la red para calcular una medida de error y luego corregir los pesos, denominada *acumulativa*. La otra en la que después de presentar cada muestra a la red se calcula el error y se corrigen los pesos, denominada *Delta o paso-a-paso*.

En el algoritmo se utiliza corrección *Delta o paso-a-paso* ya que los conjuntos de datos de señales ultrasónicas son grandes (mas de 10,000 muestras en promedio) y en estos casos el entrenamiento acumulativo produce una convergencia muy lenta.

Al paso de todas las muestras por la red para corregir los pesos, independientemente de la estrategia de aprendizaje, se le denomina *época*.

Como parte del entrenamiento de una red, es necesario evaluar periódicamente su desempeño, pues sólo así se obtiene información que sugiere la continuación o detención del proceso. Para este fin, el entrenamiento se interrumpe y se evalúa utilizando el conjunto de prueba, considerando el valor de una medida de error calculada con base en la diferencia entre los resultados observados y los deseados. En la aplicación, el usuario es responsable de la frecuencia de estas pruebas. El valor por defecto es probar cada 5 épocas y el programa genera una gráfica a tiempo real de la medida de error.

La decisión sobre el momento para detener el proceso de entrenamiento esta dictada por la capacidad de la red, no sólo de reproducir los datos de entrena-

miento, sino de clasificar correctamente conjuntos de datos diferentes. Además, durante el entrenamiento de una red neuronal existe un fenómeno conocido como sobreajuste, que consiste en hacer que la red aprenda a reproducir exactamente el conjunto de datos de entrenamiento, pero perdiendo la generalidad del fenómeno y consecuentemente, produciendo resultados pobres con cualquier otro conjunto de datos. Este fenómeno se observa generalmente en etapas avanzadas del entrenamiento, es decir, tras haber aplicado retropropagación un gran número de veces. La Figura 4.9 muestra este fenómeno. La gráfica del error calculado al conjunto de datos de entrenamiento tiene una forma asintótica, mientras que el error al conjunto de prueba crece tras haber alcanzado un mínimo. La época en la que se produce este mínimo es el momento óptimo para detener el proceso de entrenamiento.

Una vez que se ha entrenado la red, los pesos se guardan en un archivo para su uso futuro.

4.3. Red Neuronal: Uso

Explotar una red neuronal es en general un proceso sencillo: se especifican las entradas, la red hace los cálculos y se obtiene una salida. Este caso no es la excepción, aunque además es necesario aplicar la secuencia de pasos de preprocesamiento que fueron descritos en la Sección 4.1 con el fin de preparar cualquier conjunto de datos. Hecho esto, es posible construir las muestras asociadas a cada punto para ingresarlas a la red y así obtener un resultado también para cada punto. Las características que describen al conjunto de datos de entrada no cambian para el conjunto de datos resultante, es decir, este último mantiene las mismas resoluciones (Δs y Δt) permitiendo una interpretación inmediata de los resultados.

La aplicación es capaz de evaluar cualquier conjunto de datos utilizando cualquier red neuronal ya creada y tiene una interfaz, mostrada en la Figura 4.10, que permite realizar una detallada comparación visual entre el conjunto de datos de entrada (a la izquierda en la interfaz) con el resultado del análisis (a la derecha) utilizando acercamientos, y desplazamientos.

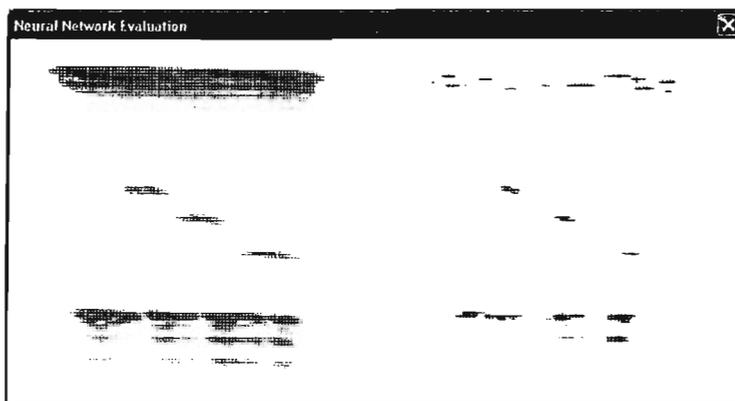


Figura 4.10: Interfaz de análisis visual de resultados.

Capítulo 5

Experimentación y resultados

Este capítulo describe el proceso de experimentación, incluyendo detalles como la selección de los parámetros importantes, su justificación y algunos resultados.

Como introducción al capítulo se presenta una descripción de los detalles técnicos del ambiente de experimentación. Mas adelante el orden de presentación es el mismo del Capítulo 4: primero el preprocesamiento, seguido de la construcción de los conjuntos de datos, luego el entrenamiento de la red y por último el uso de ésta.

Detalles técnicos

La fase de experimentación fue realizada utilizando una aplicación desarrollada *ex professo*. El código fue escrito en el lenguaje Visual C++ utilizando el producto “Microsoft Visual Development Environment 2003 Version 7.1.3088” sobre el sistema operativo “Microsoft Windows XP 2002”. El equipo de cómputo es una PC con procesador Pentium IV, 512 MB de memoria RAM, 40 GB de disco duro y una tarjeta de video nVidia GeForce 4 que favorece el despliegue de gráficos complejos.

Las partes del programa que utilizan cualquier tipo de despliegue gráfico, ya sea para inspección visual de las señales de entrada o para análisis de los resultados, fueron resueltas utilizando la librería OpenGL, que tras muchos años sigue siendo una de las herramientas mas versátiles y cada vez mejor documentadas

para este fin, pero sobre todo muy eficiente. Esta última característica cobra especial importancia si se pretende manipular una superficie tridimensional de mas de un millón de polígonos a tiempo real.

Dentro de la aplicación, cada parte del análisis de señales se desarrolló en una ventana independiente, mostrando únicamente los detalles de la información indispensable, en un intento por simplificar la utilización. Así pues, existen siete ventanas o interfaces encargadas de: la inspección de las señales de entrada, el preprocesamiento, la definición e inspección de características ultrasónicas, la construcción de imagen ideal, la construcción de conjuntos de datos, el entrenamiento de la red neuronal y la inspección de resultados.

El sistema es capaz de ejecutarse en cualquier computadora con sistema operativo Windows 98 o superior. La aplicación final que contiene todas las herramientas de visualización, las de procesamiento digital de señales y los algoritmos de aprendizaje y uso de la red, mide menos de 0.5 Megabytes y su uso de memoria RAM depende del tamaño del conjunto de datos de entrada.

5.1. Preprocesamiento

La secuencia de preprocesamiento consiste en el cálculo de la envolvente, la corrección por atenuación y la normalización. Esto establece las condiciones que permiten la construcción de muestras homogéneas al considerar las variaciones intrínsecas de la física del fenómeno de propagación.



Figura 5.1: Interfaz de herramientas de preprocesamiento.

La interfaz de las herramientas de preprocesamiento se muestra en la Figura

5.1.

5.1.1. Envoltente

El primer paso en la etapa de procesamiento es obtener la envoltente de la señal de entrada original. Para ello no hay más que aplicar el algoritmo descrito en la Sección 4.1.1. Todos los parámetros del algoritmo son características intrínsecas de cada señal, como el tamaño de las trazas y el número de éstas. La Figura 5.2a muestra una traza original, y en 5.2b se observa la envoltente. Este proceso se repite para cada traza.

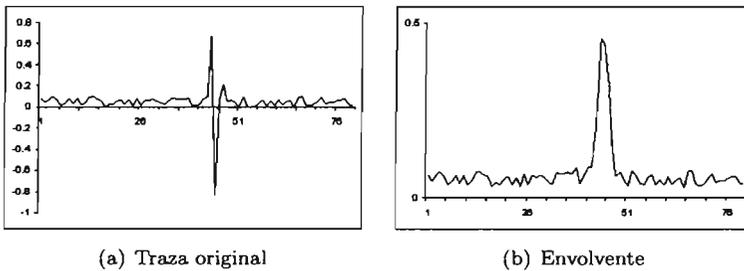


Figura 5.2: Envoltente de una traza.

La Figura 5.3 muestra el resultado de calcular la envoltente de una señal real. El resultado es particularmente claro en la vista ortogonal al plano zy (gráfica inferior). Todos los lóbulos con valores menores que cero han desaparecido.

5.1.2. Corrección por atenuación

Una vez calculada la envoltente, lo siguiente es hacer la corrección por atenuación. Para lograrlo es necesario calcular la constante α de la Ecuación 2.2. Ya sea que este valor se obtenga utilizando las constantes asociadas a las características físicas del material o, de manera empírica, utilizando la aplicación que permite seleccionar dos puntos donde se asumen amplitudes similares a profundidades diferentes. Una vez teniendo α , el programa calcula el coeficiente de atenuación para cada profundidad y multiplica los valores de las trazas por estos coeficientes. El resultado es una señal donde las amplitudes son comparables sin importar la profundidad.

Las líneas sobre la imagen de la interfaz de herramientas de preprocesamiento mostrada en la Figura 5.1 tienen la función de indicar los puntos que contienen

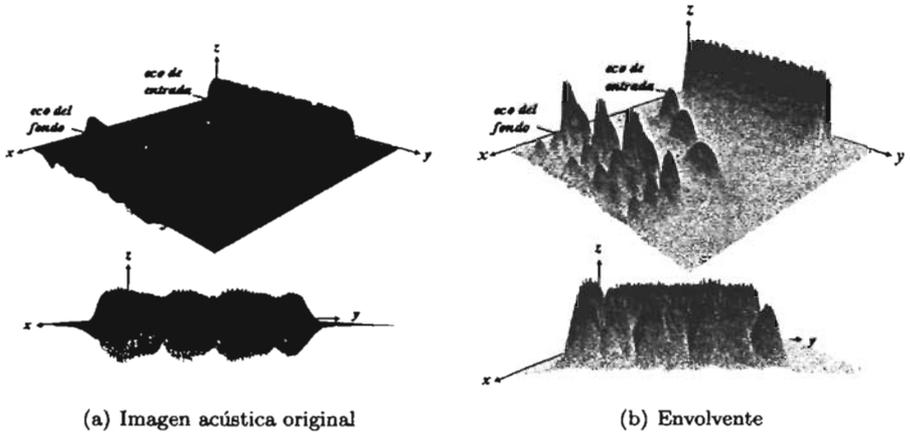


Figura 5.3: Envolvente de una imagen acústica 3D.

la información relevante para calcular la constante α .

La Figura 5.4 muestra el efecto de esta corrección. Note como en la envolvente de la señal, los arcos de la señal tienen amplitudes que decrecen con la profundidad (eje x), en cambio en la señal corregida estos tienen amplitudes constantes. Como un efecto colateral, se puede observar un importante incremento en la amplitud del eco del fondo de la pieza y un decremento del eco de entrada.

5.1.3. Normalización

La normalización debe ser la última etapa de preprocesamiento antes de construir las muestras, pues sólo así es posible garantizar que el intervalo de los datos es apropiado para su uso en una red neuronal. En este paso es importante cuidar que el rango dinámico (descrito en la Sección 2.6) sea aprovechado de la mejor forma, principalmente porque las amplitudes más altas en señales ultrasónicas son por lo general de información irrelevante, por ejemplo los ecos de entrada o de fondo, en los que se registran intensidades de mucho mayor magnitud que las registradas al percibir el reflejo de alguna falla en el interior. Sin embargo, existe información sobre estas fronteras y se puede utilizar para eliminar datos de la imagen acústica, permitiendo que el rango dinámico sea aprovechado de la mejor forma por la información relevante. La Figura 5.5a muestra la señal con la corrección de atenuación, incluyendo las muestras de información irrelevante.

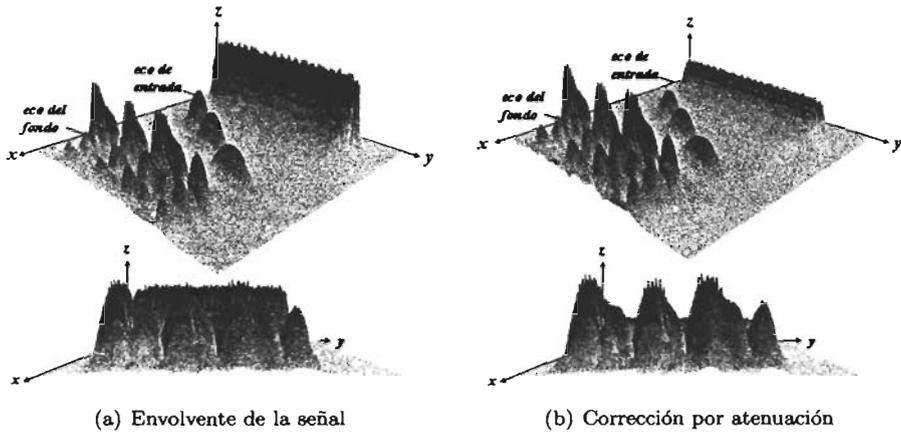


Figura 5.4: Corrección de atenuación.

Note como el eco del fondo se acentúa debido a la corrección por atenuación, limitando el rango dinámico para la información de las fallas.

En cambio, la imagen 5.5b, muestra la señal normalizada sin las secciones de los ecos de entrada y de fondo. El rango dinámico esta completamente utilizado por las información de las fallas.

Muestras Homogéneas

Una vez normalizada, las características básicas de la señal están listas, ya sea para formar parte de un conjunto de entrenamiento (y con este entrenar una nueva RN) o para su análisis con una RN ya entrenada. Sin importar el uso que se vaya a dar, es necesario obtener los arcos descritos en la Sección 4.1.4 que conforman la región muestral y así construir los patrones que alimentan a la red.

Para construir las regiones muestrales es necesario conocer las características del estudio ultrasónico (ángulo de visión del transductor, resoluciones horizontal y en profundidad, etc.) e ingresarlas en el programa, ya sea utilizando la interfaz o leyendo estos valores de un encabezado del mismo archivo donde se almacena la señal.

La Figura 5.6 muestra la interfaz de características ultrasónicas que permite leer y editar estos valores. La parte izquierda muestra la imagen acústica producida por un conjunto de señales y, sobre ésta, la región visible por el transductor

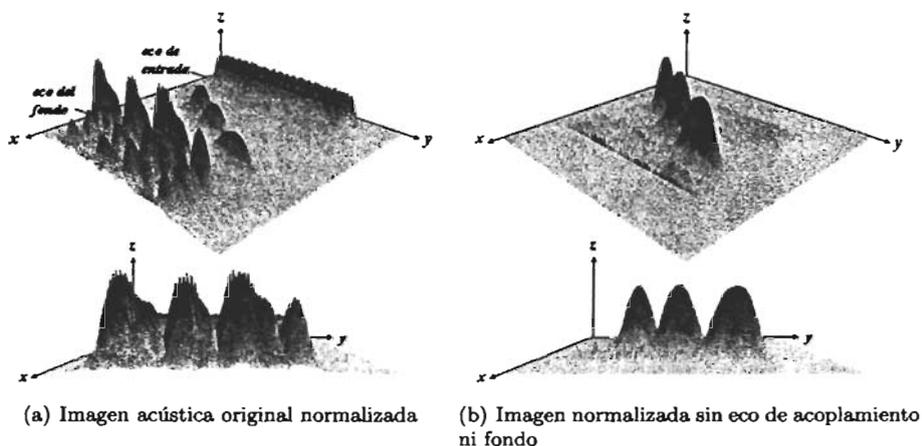


Figura 5.5: Uso del rango dinámico.

en la posición actual y la región muestral. Del lado derecho hay tres secciones, la primera relacionada con los parámetros de profundidad de la pieza como son el número de muestras, la frecuencia de muestreo del convertidor analógico digital (en MHz), la velocidad de propagación (en mm/s), la distancia del transductor a la primera muestra de la señal (en muestras), la máxima profundidad analizable (en cm) y la resolución (en mm). Estos dos últimos valores son de sólo lectura y se calculan en función de las otras características del estudio. La segunda sección consiste en las características horizontales del estudio, como son el número de trazas obtenidas, la resolución o distancia entre éstas (expresado en milímetros), el ángulo de visión (en grados) y el ancho muestreado total (en cm). La tercera y última sección describe la posición del arco muestral que se presenta en la gráfica.

Cada vez que el usuario modifica alguno de estos valores, el programa calcula las regiones muestrales para todas las profundidades y el cambio se refleja inmediatamente en la región muestral desplegada sobre imagen acústica, brindando retroalimentación a tiempo real sobre los efectos que las modificaciones producen. Este arco se puede desplazar a cualquier punto sobre la imagen permitiendo comprobar visualmente la congruencia entre las características conocidas del estudio y las regiones muestrales que estas características producen, permitiendo hacer ajustes para corregir inconsistencias. La Figura 5.7 muestra ejemplos de tres regiones muestrales. En este caso, estas efectivamente coinciden con los

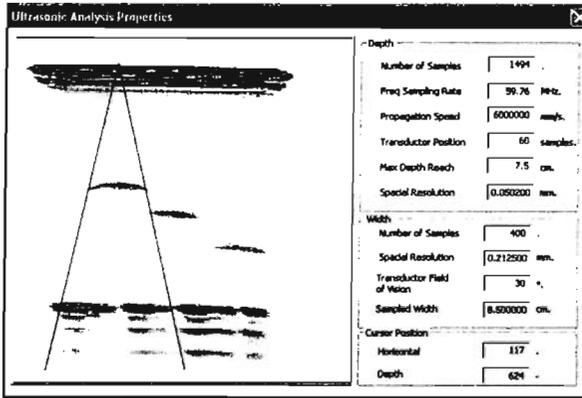


Figura 5.6: Interfaz de características ultrasónicas.

arcos producidos por las fallas en la imagen acústica.

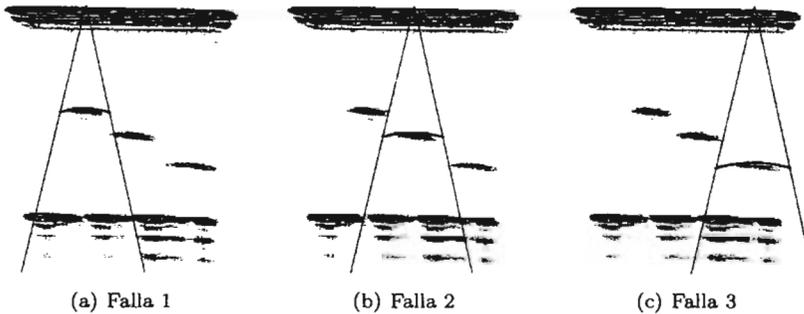


Figura 5.7: Región muestral situada en la posición central de las fallas.

Un ejemplo de regiones muestrales inconsistentes con las características del estudio se muestra en la Figura 5.8. En particular, el error consiste en la especificación de una frecuencia de muestreo en profundidad menor a la real. El resultado son arcos notoriamente más grandes que los producidos por las fallas.

Habiendo especificado las características ultrasónicas de la señal, el único parámetro adicional necesario para construir las muestras que ingresarán a la red es el *tamaño* o número de elementos de estas. La fuente que especifica el tamaño depende del uso que se pretende dar a las muestras. En el caso de que fueran a ser utilizadas para entrenar una nueva red, el tamaño debe ser consistente

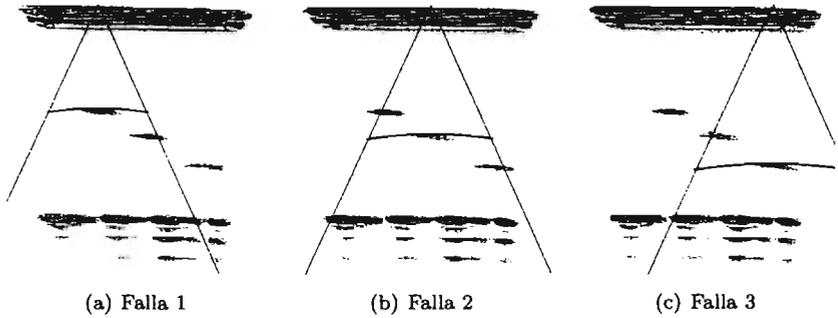


Figura 5.8: Región muestral inconsistente con los parámetros del estudio.

con el número de nodos de la capa de entrada de la RN que será construida y el usuario lo especifica. Para el caso en que las muestras sean la entrada de una red ya existente (y supuestamente entrenada), el tamaño deberá ser igual al número de nodos de la capa de entrada de aquella red que realizará el análisis y está especificado intrínsecamente por la red.

Conociendo este valor, el programa utiliza el algoritmo automático de interpolación basado en splines (descrito en la Sección 4.1.5) que no requiere ningún parámetro.

Utilizando esta secuencia de etapas de preprocesamiento se obtienen muestras homogéneas para cada punto.

5.2. Conjuntos de entrenamiento

El conjunto de entrenamiento es una colección de muestras homogéneas asociadas individualmente con la respuesta ideal que se espera de la red. Luego entonces, es necesario conocer estas respuestas ideales para construirlo.

El programa tiene una interfaz que permite definir regiones de la pieza cuya impedancia acústica es diferente. La Figura 5.9 muestra esta interfaz. Del lado izquierdo se observa la imagen acústica producida por el conjunto de datos, y sobrepuestas en negro, las regiones cuya impedancia acústica es diferente, que en el caso de esta imagen corresponden a 3 perforaciones de 4.2 mm. Del lado derecho en la parte superior se muestran algunas características de la imagen que facilitan la tarea de situar las regiones de impedancia diferente. Del mismo lado, pero en la parte inferior se encuentran los controles que facilitan la definición

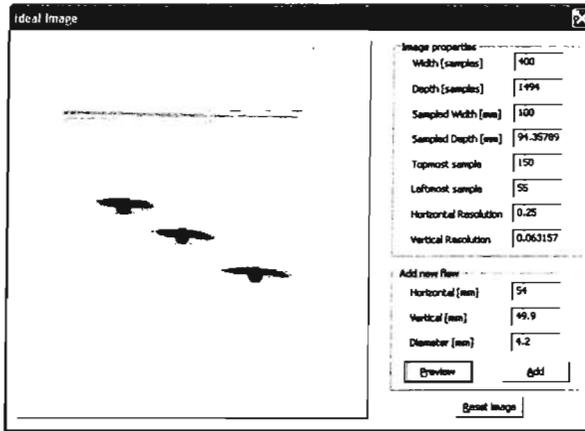


Figura 5.9: Interfaz de imagen ideal.

de estas regiones.

Una vez definida la imagen ideal, se cuenta con toda la información necesaria para construir un conjunto de datos que contiene todas las muestras que se pueden generar utilizando todos los datos del estudio y toda la información de la imagen ideal. Sin embargo este conjunto es, por lo general, poco apropiado para entrenar una RN debido a que existe un enorme sesgo en el número de muestras que existen para cada una de las clases. En particular, la clase “muestra-sin-falla” tendrá un porcentaje mucho mayor de participación en las muestras comparado con el que tendrá la clase “muestra-con-falla” .

Para evitar esta situación, se construyó otra interfaz que agrupa las funciones relacionadas con la construcción de los conjuntos de datos en la cual es posible seleccionar regiones de la imagen acústica para componer un conjunto de entrenamiento balanceado. La Figura 5.10 muestra esta interfaz. Las líneas sobre la imagen indican la frontera de la región que se agrega al conjunto de datos.

Esta parte del programa es también la encargada de construir un archivo donde se almacena toda la información relacionada con este conjunto, desde el tamaño de cada muestra, hasta los valores de cada una de ellas, incluyendo por supuesto las respuestas ideales asociadas.

El funcionamiento es como sigue: se selecciona la opción de crear un nuevo archivo de conjunto de entrenamiento y se asigna un nombre. En ese archivo el programa guarda las características de la arquitectura de la red que se pretende

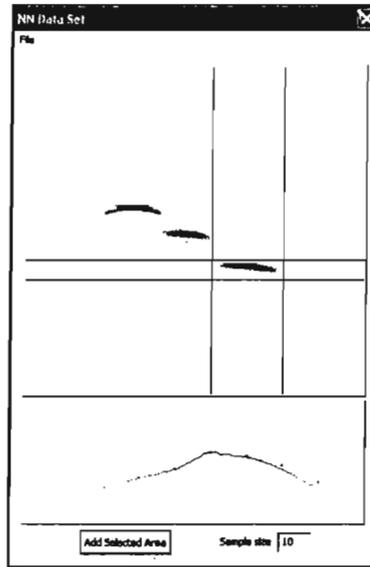


Figura 5.10: Interfaz de construcción de conjuntos de entrenamiento.

entrenar, es decir el tamaño de la capa de entrada. El usuario especifica este valor utilizando el control que se encuentra en la parte inferior de la interfaz. Por último se definen una a una las regiones a incluir dentro del conjunto de datos y, para cada una de ellas, el programa toma la posición de los puntos para construir su región muestral, toma los valores de la imagen real en las posiciones asociadas a esta región y construye las muestras homogéneas incluyendo el valor de salida, tomado de la imagen ideal.

Parte de la experimentación consistió en variar el tamaño de las muestras, aunque los resultados de análisis completos se muestran hasta la sección relacionada con el entrenamiento de la red. No obstante, en la última parte de esta sección se incluye la Figura 5.11 que contiene una secuencia de imágenes de regiones muestrales idénticas pero de las que se construyen muestras de diferentes tamaños. Estas imágenes fueron tomadas de la gráfica inferior de la interfaz de construcción de conjuntos de entrenamiento (Figura 5.10) y ayudan a describir mejor el proceso de construcción de las muestras.

En 5.11a se muestra la imagen acústica de todo el estudio y, sobrepuesta la región muestral. En las figuras 5.11b-f, la línea de la gráfica representa la interpolación lineal de los valores de la imagen acústica en las posiciones de la

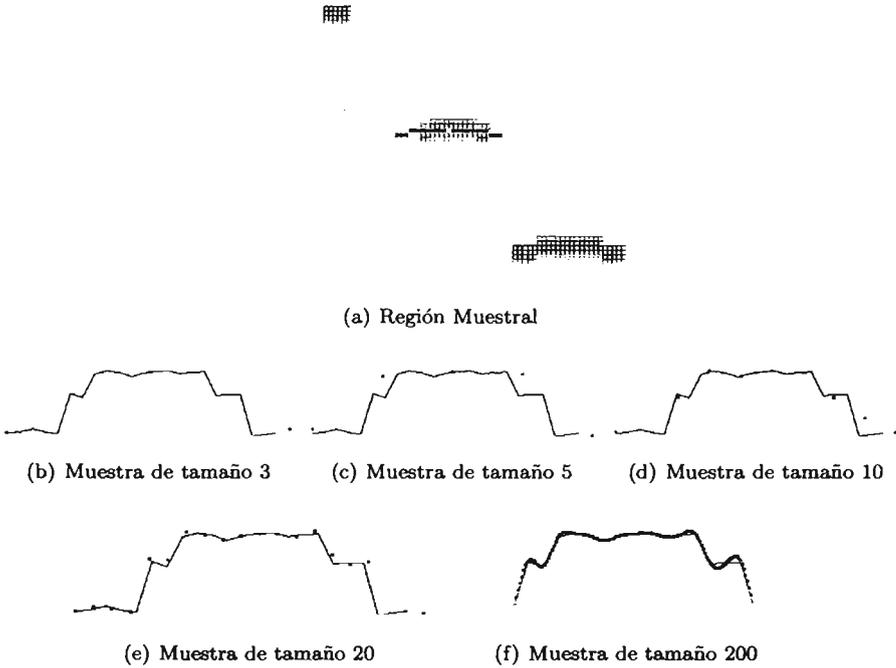


Figura 5.11: Muestras de una región muestral con diferente tamaño.

región muestral y los puntos representan los valores que tomaría una muestra del tamaño especificado para la posición que produce dicha región muestral. La forma que produce el interpolador de splines resulta particularmente visible en 5.11f.

5.3. Entrenamiento de la red

Esta sección describe las combinaciones de los parámetros que fueron seleccionados al optimizar los resultados obtenidos al entrenar y utilizar la red neuronal. Entre ellos se encuentran el tamaño de la capa de entrada (y consecuentemente el tamaño de las muestras homogéneas) y el de la capa intermedia, las tasas de aprendizaje de cada capa (η_{ji} y η_{kj}), la constante de momento (μ), la constante de disminución de pesos (d), la función de transferencia para cada capa e incluso el número de épocas con que se entrena a la red.

Con el objeto de mantener la objetividad a lo largo de toda la experimentación, cada parámetro fue modificado manteniendo todos los demás constantes, incluso la semilla del generador de números aleatorios con que se obtienen los pesos iniciales de la red.

Los resultados mostrados en esta sección fueron construidos utilizando parámetros alrededor de aquellos que, tras cierta experimentación inicial, se sabe le permiten a la red generar resultados aceptables, intentando hacer que el efecto producido por las modificaciones a la variable que se analiza fuera un poco mas claro.

5.3.1. Tamaño de la capa de entrada

El tamaño de la capa de entrada se puede interpretar como el nivel de detalle que contiene cada muestra, teniendo en mente que ésta será toda la información que se proveerá a la red para realizar la clasificación de cada punto. Un tamaño muy pequeño limita la cantidad de información que el clasificador recibe y consecuentemente reduce sus capacidades. Un tamaño muy grande introduce redundancia (la diferencia entre muestras contiguas se vuelve despreciable), incrementando linealmente el consumo de recursos (memoria y tiempo de procesador) pero marginalmente el desempeño. Resulta entonces indispensable encontrar el tamaño que permite a la red obtener clasificaciones adecuadas utilizando la menor cantidad de recursos.

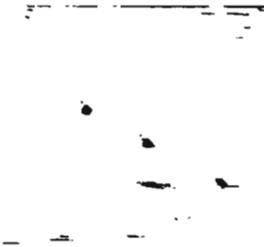
La experimentación con este parámetro se hizo tomando un valor inicial y abriendo el rango en ambos sentidos. El rango total explorado alcanzó valores de tres a veinte nodos de entrada.

Los resultados de las clasificaciones obtenidas utilizando algunos de los valores dentro de este rango se muestran en la Figura 5.12. Estas imágenes son el producto de analizar una señal real utilizando una red entrenada con conjuntos de datos de regiones idénticas, pero con muestras construidas del tamaño especificado. El entrenamiento incluyó únicamente las regiones de las fallas 1 y 3 (señaladas en 5.12a), permitiendo utilizar la falla 2 como un criterio de evaluación.

Como se puede observar, la calidad de las clasificaciones crece de acuerdo con el número de muestras que ingresan a la red, aunque la diferencia resulta casi imperceptible entre el resultado producido por la red de 10 nodos de entrada comparado con el producido por la red de 20, sin embargo el tiempo requerido para entrenar a la red (conservando el número de iteraciones) se duplica, al igual



(a) Imagen acústica original



(b) Imagen resultante. Capa de entrada de tamaño 3



(c) Imagen resultante. Capa de entrada de tamaño 5



(d) Imagen resultante. Capa de entrada de tamaño 10



(e) Imagen resultante. Capa de entrada de tamaño 20

Figura 5.12: Análisis de una señal real utilizando redes con capa de entrada de tamaño 3, 5, 10 y 20.

que el tiempo de análisis y el consumo de memoria.

5.3.2. Tamaño de la capa intermedia

El tamaño de la capa intermedia influye profundamente en las capacidades de aprendizaje de la red, ya que se encuentra directamente relacionado con el número de pesos que ésta tiene, sabiendo que los pesos son los únicos elementos que almacenan las características del fenómeno que se analiza. Sin embargo, no existe ningún algoritmo reportado que indique el número óptimo de neuronas de las capas intermedias y en general este valor se obtiene de manera empírica.

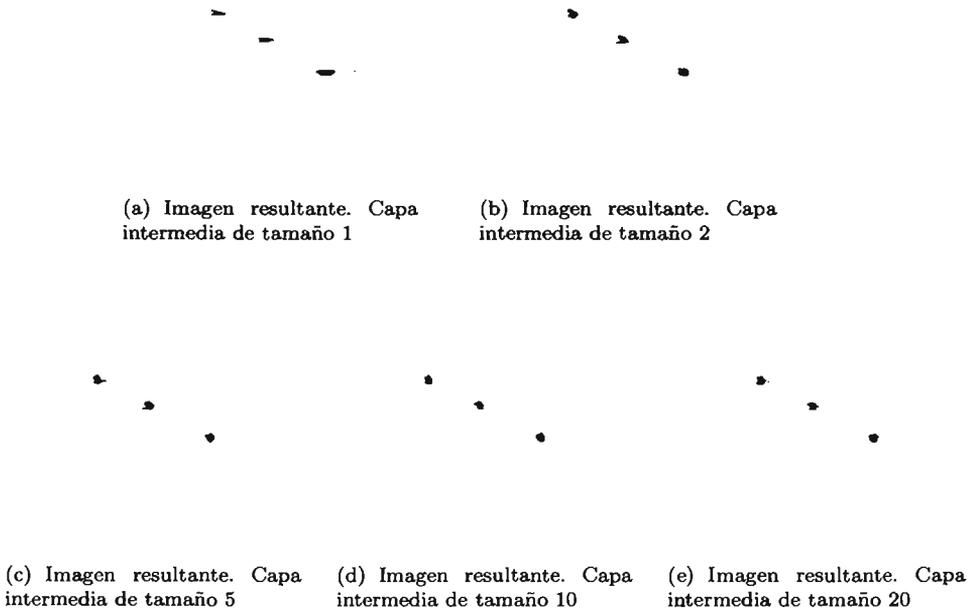


Figura 5.13: Análisis de una señal real utilizando redes con capa intermedia de tamaño 1, 2, 5, 10 y 20.

Las consecuencias de errar en la selección de este parámetro son similares a las de errar en la selección del tamaño de la capa de entrada, es decir, un número muy pequeño limita las capacidades de aprendizaje de la red, particularmente los detalles y un número muy grande entorpece el entrenamiento además de

propiciar el fenómeno del sobreajuste.

La experimentación consistió en construir redes con arquitecturas cuya capa intermedia tuviera entre uno y veinte neuronas. Los resultados de algunas de estas se muestran en la Figura 5.13. Para realizar las pruebas, el tamaño de la capa de entrada se fijó en 10 nodos. La imagen acústica original es la misma que la presentada en la Figura 5.12a.

Como se puede observar, las redes cuya arquitectura tiene apenas uno o dos nodos en la capa intermedia tienen algunos problemas para definir adecuadamente los contornos de las fallas. No así las redes de 10 y 20 neuronas.

5.3.3. Tasas de aprendizaje y constante de momento

Las tasas de aprendizaje y la constante de momento no afectan directamente en el desempeño de la RN en su labor de clasificación, pero si lo hacen durante el proceso de entrenamiento. Recordemos que, en la búsqueda del mínimo en la función de error, el valor η_{ji} limita la variación máxima de los pesos de la capa oculta, mientras que η_{kj} lo hace para la capa de salida. Por otro lado, la constante de momento μ especifica la cantidad de “inercia” que tienen esas variaciones. Así pues, dada la correlación en los efectos de los tres parámetros, resulta difícil considerarlos de manera independiente. De hecho, durante el proceso de experimentación se encontraron valores para las tasas de aprendizaje que, al momento de agregar el efecto del momento, tuvieron que ser corregidos.

El espacio de búsqueda donde se pretende encontrar la combinación de η_{ji} , η_{kj} y μ que produce resultados aceptables es amplio. Sin embargo, existen modelos heurísticos que permiten realizar la búsqueda experimental de una forma mas o menos orientada. En este sentido se presume por ejemplo, que la capa interna de la red realiza una clasificación mas burda comparada con aquella que realiza la capa de salida en donde el resultado de la clasificación es tan fino como las condiciones asociadas al aprendizaje lo permitan. En consecuencia, se sugiere utilizar un valor mas grande para la tasa de aprendizaje relacionada con los pesos que llegan a la capa oculta y uno mas pequeño para la tasa de aprendizaje de los pesos que llegan a la capa de salida, en un intento por obtener mejores resultados al alinear la selección de los parámetros con el funcionamiento del modelo.

La primera parte de la experimentación se enfocó en encontrar las tasas de aprendizaje que, sin utilizar momento, produjeran resultados aceptables a lo largo de un número de épocas constante. La Tabla 5.1 muestra algunos de

los resultados obtenidos durante esta parte de la experimentación. En las dos primeras columnas se especifican las respectivas tasas de aprendizaje. La tercera muestra el mínimo del error medio cuadrático del conjunto de datos de prueba alcanzado tras mil épocas de entrenamiento.

η_{ji}	η_{kj}	EMC 1000 épocas
0.1	0.05	0.02499
0.2	0.1	0.01674
0.3	0.2	0.01524
0.5	0.3	0.01498
0.5	0.5	0.01457
0.7	0.5	0.01416
0.9	0.9	0.01517

Tabla 5.1: Error medio cuadrático para diferentes tasas de aprendizaje.

De la tabla se observan algunos resultados esperados pero otros no tanto. En la categoría de los predecibles se encuentra por ejemplo el pobre desempeño obtenido al utilizar tasas de aprendizaje sumamente pequeñas, como $\eta_{ji} = 0.1$, $\eta_{kj} = 0.05$. La evolución del entrenamiento utilizando estas tasas se presenta en la gráfica de la Figura 5.14a. Ahí se observa cómo el progreso se detiene durante las primeras 200 épocas de entrenamiento, posiblemente por causa de un mínimo local cuya frontera superó a la máxima variación de pesos permitida por las tasas de aprendizaje.

Otro resultado previsto era la disminución gradual de la función de error durante las primeras etapas del entrenamiento al incrementar ligeramente las tasas de aprendizaje. Este ajuste permite explorar más rápidamente el espacio de solución del problema durante el proceso de entrenamiento, como consecuencia del incremento en la variación máxima de los pesos. Las gráficas 5.14a, 5.14b y 5.14c muestran esta situación.

El resultado sorprendente fue el buen desempeño obtenido con tasas de aprendizaje cercanas a uno. Por ejemplo en el entrenamiento cuya evolución se muestra en la Figura 5.14d, donde resulta evidente que las variaciones de los pesos son exageradamente reactivas, no obstante se alcanzó (casualmente) una combinación de pesos extraordinariamente funcional, dadas las circunstancias.

Con respecto a la constante de momento, se presume que el efecto acumulado de múltiples correcciones es mucho más objetivo que el de una sola. En consecuencia se utiliza una constante de inercia cercana a uno (ya que si

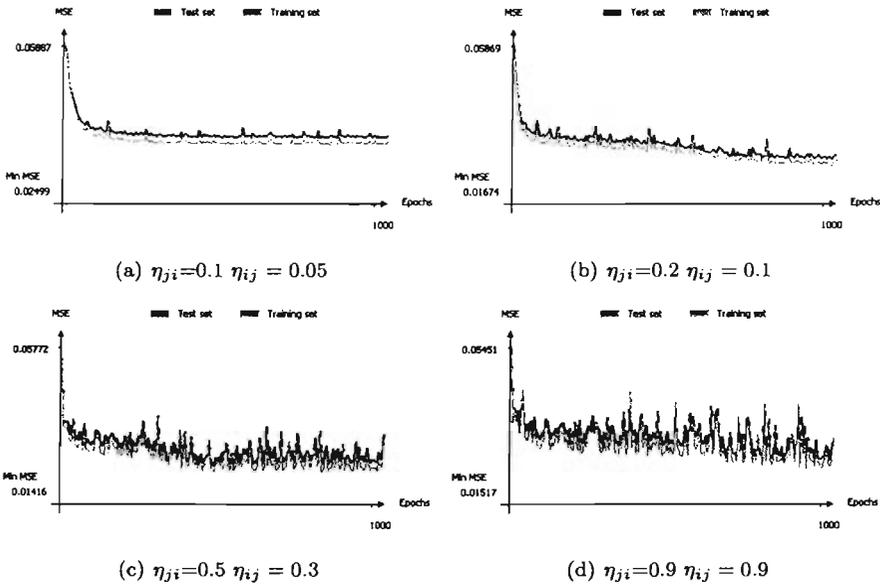


Figura 5.14: Error medio cuadrático de 1000 épocas de entrenamiento.

$\mu \rightarrow 1$ entonces $\Delta w_{ji}(t) \rightarrow \Delta w_{ji}(t - 1)$), reduciendo el efecto de correcciones individuales y posiblemente ruidosas. Además, el uso de una constante de momento cercana a uno produce otros dos efectos favorables al entrenamiento de la red. El primero es el esperado efecto acelerador al dirigirse hacia el mínimo (presuntamente global) de la función de error, permitiendo modificaciones a los pesos incluso mayores a las definidas por las tasas de aprendizaje y facilitando el escape de mínimos locales. El segundo, quizá menos evidente, es el efecto desacelerador (similar al que se produciría al reducir las tasas de aprendizaje), que ocurre durante la aproximación final al mínimo de la función de error, provocado principalmente por los cambios de dirección. La Figura 5.15 muestra estos efectos utilizando un ejemplo sintético de aproximación al mínimo de una función.

La imagen 5.15a presenta el escenario general. En ésta se muestra cómo gracias al efecto de inercia del momento es posible salir del mínimo local.

Por otro lado, en 5.15b se observa el acercamiento de la región del recuadro de 5.15a. En esta imagen se observa cómo gracias a los cambios de dirección, los efectos de las correcciones se hacen mas pequeños durante la aproximación final. La imagen es recursiva pues se puede considerar como la amplificación del

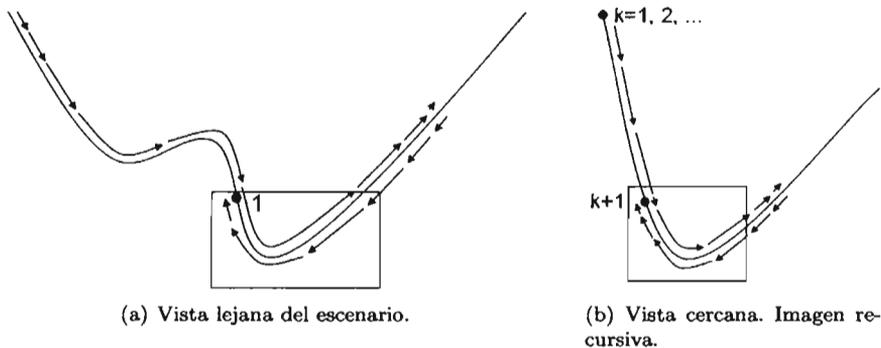


Figura 5.15: Efecto del momento en la aproximación a un mínimo.

recuadro marcado en si misma en siguientes iteraciones del proceso de entrenamiento.

Es de esperarse entonces que, tras un número específico de épocas, el error producido por una red entrenada utilizando el momento sea menor que el producido por otra red entrenada sin éste. La Figura 5.16 muestra algunos resultados en este sentido utilizando el mismo formato que en la Figura 5.14.

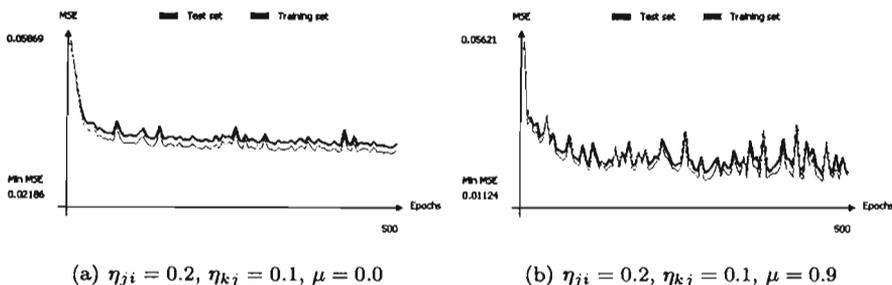


Figura 5.16: Efecto del momento en el error medio cuadrático.

Aunque la evolución del entrenamiento en 5.16b es ruidosa, el error mínimo alcanzado es 20% menor que el mas pequeño de la Tabla 5.1 utilizando el mismo número de épocas gracias al efecto de desaceleración que permite alcanzar mínimos más profundos.

5.4. Uso de la red

Hasta ahora, la medida utilizada para evaluar el éxito en el entrenamiento de la red ha sido el error medio cuadrático, que con el fin de mantener la objetividad, se calculó utilizando una parte del conjunto de datos de entrenamiento que no fuera utilizada como entrada para el algoritmo de retropropagación. Sin embargo, este conjunto de prueba está formado por muestras tomadas de un mismo estudio ultrasónico, con lo cual no es posible decidir si la red aprendió la generalidad del fenómeno o si únicamente encontró características del estudio con el que fue entrenada.

Con el objeto de obtener conclusiones más generales, se definieron dos nuevos escenarios de experimentación. El primero consiste simplemente en utilizar la red entrenada con un estudio y validarla con uno distinto. El segundo consiste en construir un conjunto de entrenamiento utilizando muestras de diferentes estudios ultrasónicos para analizar tanto estudios que contribuyen a formar este conjunto de datos mixto, como otros que no. Ambos escenarios permiten evaluar de mejor manera la capacidad de generalización durante el aprendizaje de la red.

Con el primer escenario se obtuvieron resultados exitosos, aunque sólo parcialmente, pues tras haber confirmado que la red era capaz de clasificar correctamente al estudio con el que fue entrenada, se encontró que también era capaz de hacerlo con estudios distintos, aunque no con todos. La Figura 5.17 muestra esta situación. En la parte superior se observan algunas imágenes acústicas y en la inferior, los correspondientes resultados del análisis realizado por una red. Esta red fue entrenada utilizando un conjunto de muestras construido con base en uno de los estudios. Resultados muy similares se obtienen al entrenar a la red utilizando conjuntos de entrada construidos con los estudios 5.17b ó 5.17c, pero no con 5.17d, por ejemplo.

Este escenario de experimentación permite concluir con respecto a la capacidad de la red para aprender algunas de las características del fenómeno, pero también evidencia la falla al intentar hacer una generalización universal, pues de alguna forma el clasificador se adecúa a un tipo de estudios ultrasónicos.

El segundo escenario define un espacio de experimentación sumamente amplio, ya que existe un gran número de combinaciones posibles para construir conjuntos de datos mixtos. Recordemos que después de producir cada conjunto de datos es necesario entrenar una red y luego evaluarla con una variedad de estudios diferentes para producir conclusiones. Luego entonces se intentó acotar

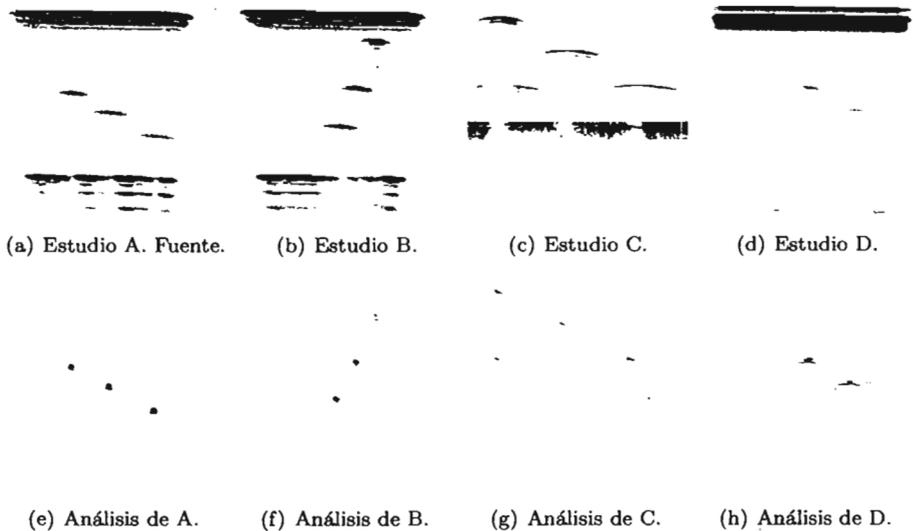


Figura 5.17: Uso de una red para analizar diferentes estudios.

este espacio intentando predecir grupos de estudios ultrasónicos que resultaran mas difíciles de aprender para la red, basándose en las dificultades para clasificar correctamente distintos estudios dentro del escenario de experimentación anterior. Así por ejemplo, un conjunto de datos apropiado para este escenario sería aquel que se forma al unir las muestras producidas por la imagen acústica 5.17a con las de 5.17d. Una muestra de los resultados obtenidos se presenta en la Figura 5.18, así como los medios para realizar comparaciones.

Las imágenes de la izquierda (5.18a y 5.18d) muestran los estudios ultrasónicos originales. Las centrales (5.18b y 5.18e) muestran la salida de dos redes entrenadas de manera individual, cada una con conjuntos de datos basados en su respectiva imagen fuente. Las de la derecha (5.18c y 5.18f) muestran la salida de una red entrenada con un conjunto de datos mixto en el que participan ambas fuentes.

Como es posible observar, los resultados pierden resolución cuando se enseña a la red a clasificar estudios de múltiples fuentes, aunque definitivamente obtiene mejores capacidades de generalización. Note por ejemplo como el resultado del análisis de la Figura 5.18f, producido por una red entrenada con un conjunto mixto, es mucho mejor que aquel de la Figura 5.17h, en el que la misma señal se analizaba utilizando una red entrenada para un estudio diferente. El mismo

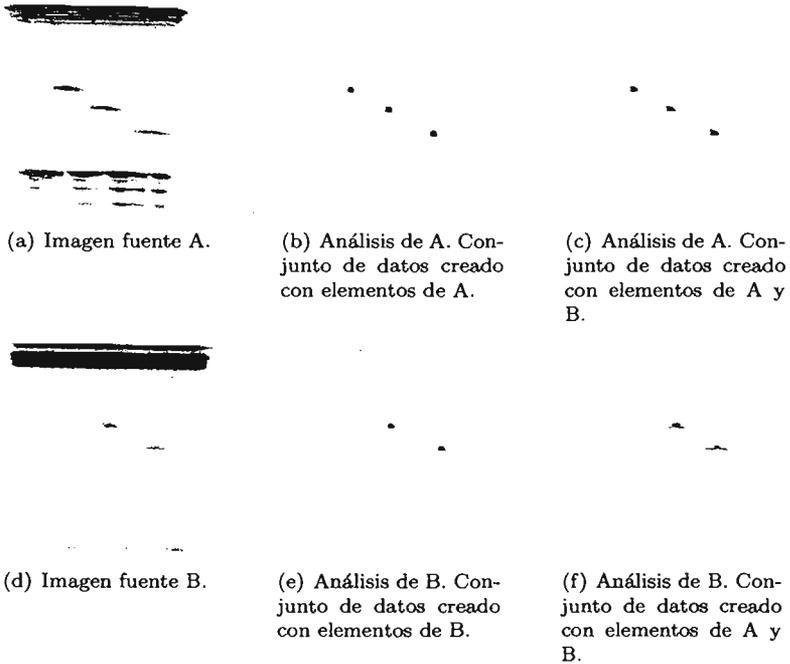


Figura 5.18: Resultados de tres diferentes redes.

fenómeno ocurrió en mayor o menor medida con diferentes combinaciones de estudios ultrasónicos como fuentes para construir conjuntos de entrenamiento mixtos. A partir de lo anterior se puede concluir que existe un compromiso entre la capacidad de generalización de la red y la generalidad de los datos que ésta es capaz de analizar.

Capítulo 6

Conclusiones

El objetivo del proyecto consistió en evaluar la capacidad de una red neuronal de perceptrones para analizar señales ultrasónicas de ensayos no destructivos. En este sentido se pueden generar diversas conclusiones, algunas categóricas y otras que requieren de construir nuevos escenarios de experimentación para eliminar la ambigüedad. Este capítulo resume las conclusiones y contribuciones más importantes y se presentan también algunas situaciones que, ante la necesidad de definir una frontera final, no fueron incorporadas dentro del marco de esta investigación, convirtiéndose en posibilidades de trabajo futuro.

Con el objetivo de lograr que el sistema fuera capaz de realizar el análisis propuesto de señales ultrasónicas, se diseñó una metodología que consiste en una etapa de preprocesamiento y construcción de muestras homogéneas que permite eliminar algunos de los fenómenos intrínsecos en los sistemas basados en ultrasonido, y una vez construidas es posible entrenar o usar a la red con ellas. Para entrenarla, se construye un subconjunto de muestras y se aplica el algoritmo de retropropagación que ajusta los pesos de las conexiones de la red para aprender el fenómeno descrito por las muestras. Una vez entrenada, la red es capaz de analizar otros conjuntos de datos.

El orden de presentación en este capítulo es el mismo utilizado en capítulos anteriores: preprocesamiento, entrenamiento de la red y uso de la misma.

6.1. Preprocesamiento y construcción de muestras

La secuencia de pasos que conforman la etapa de preprocesamiento constituye una de las principales contribuciones de esta investigación. Utilizando únicamente las propiedades físicas, tanto del sistema de adquisición, como del fenómeno de propagación del sonido, fue posible construir la representación de cada punto mediante regiones que contienen la mayor parte de la información relevante de dicho punto. Posteriormente se realiza un muestreo uniforme de la información ultrasónica de dicha región utilizando un mecanismo basado en splines, permitiendo construir muestras de tamaño uniforme capaces de alimentar a una red neuronal.

Así fue posible extraer información de una señal heterogénea (perturbada por fenómenos como atenuación, amplificación, pérdidas por acoplamiento y apertura del ángulo de visión del transductor) para construir muestras homogéneas, que constituyen los elementos que proveen de información a la red.

Sin embargo, se encontraron dos situaciones relacionadas con esta etapa que eventualmente condicionan el funcionamiento del sistema. La primera es la exactitud de los parámetros que describen al sistema de adquisición y la segunda es la información que se obtiene en una situación de oclusión (ver Sección 2.1.5).

Con respecto a los parámetros del sistema de adquisición, el problema consiste en que las muestras que alimentan a la red dependen profundamente de las regiones que se construyen utilizando los parámetros y resulta indispensable que la información relacionada con el punto de análisis se encuentre dentro de la región. De lo contrario, la red recibe datos que no describen el punto de análisis y consecuentemente sus resultados son imprevisibles.

Para evitar esta situación se proponen algunas opciones. La más evidente es asegurarse de obtener todos los parámetros relacionados con el estudio ultrasónico (como el ángulo de visión del transductor, la velocidad de propagación del material, y las resoluciones y la distancia entre el transductor y la pieza). Si todos estos datos son correctos, la región que se construye corresponde exactamente con el arco que cualquier elemento reflectivo produce en una secuencia de trazas ultrasónicas y la red tendrá toda la información para describirlo.

Otra opción factible cuando se desconoce uno de los parámetros, consiste en aproximarlos utilizando la retroalimentación visual que produce la interfaz de propiedades ultrasónicas de la aplicación. Esto se logra sobreponiendo el ar-

co sintético sobre los arcos de los elementos reflectivos presentes en el estudio, modificando el parámetro hasta encontrar el valor que produce superposiciones completas del arco sintético sobre todos los arcos de la imagen acústica. Esta ayuda resulta útil, por ejemplo, cuando se desconoce la distancia entre el transductor y la pieza.

Una solución alternativa, propuesta como trabajo futuro, consiste en modificar la forma de la región con la que se construyen las muestras (que actualmente es un arco delgado) por una que abarque todas las regiones posibles que las variaciones en los parámetros puedan tener, es decir, una región invariante a los parámetros (como puede ser un arco grueso o incluso un cuadrado). La ventaja de esta solución consiste en que el sistema de clasificación prescinde de información adicional al estudio ultrasónico, a costa de un crecimiento en la región de análisis (lo que se traduce en mas nodos de entrada en la RN) y posiblemente en una pérdida de generalidad de la red entrenada, pues el desempeño se vincularía aún mas con los estudios que contribuyeron al conjunto de entrenamiento.

Con respecto a la situación de la información ante una oclusión, se observa que la región muestral podría no resultar completamente descriptiva, particularmente cuando existen sombras producidas por elementos reflectivos de menor profundidad. Por lo tanto, quizá sea importante incluir una característica adicional que indique a la red la proporción de energía absorbida por los elementos reflectivos, tanto a mayor como a menor profundidad, ya que esta información podría ayudar a la red ya sea a describir de una mejor forma la región de análisis o posiblemente a concluir que no existe información suficiente para producir una respuesta.

6.2. Entrenamiento de la red

La conclusión mas importante a este respecto sugiere la existencia de un compromiso entre la capacidad que tiene la red para describir los detalles de un tipo de estudios y la generalidad con que aprende el fenómeno. Esto se observó de manera reiterada al entrenar redes utilizando conjuntos de datos formados por fuentes mixtas, cuyo desempeño era bueno para indicar la posición de los elementos reflectivos pero no describía con detalle su contorno, comparado con las redes entrenadas utilizando un solo tipo de datos. Sin embargo, lo que ocurre en una aplicación real (que opera dentro de algún sistema de producción o de control de calidad) es que el equipo de estudios ultrasónicos es siempre

el mismo y la red que produce los resultados puede ser entrenada de manera individual, de modo que no sea necesario llevar el aprendizaje de la red a ningún nivel de generalización, conservando así toda la resolución.

Otra conclusión con respecto al entrenamiento de la red indica que el desempeño no se encuentra sensiblemente relacionado con su arquitectura. Es importante resaltar este hecho porque contrasta con muchas otras aplicaciones de RN en donde existen por lo general arquitecturas que producen resultados muy superiores. En esta aplicación, el rango de valores para la cantidad de nodos de capa de entrada, como de neuronas en la capa oculta es amplio sin modificar profundamente la funcionalidad. Cabe señalar, sin embargo, que la opción de utilizar una sola neurona en la capa de salida es posiblemente desafortunada, principalmente porque la información relacionada con la región de análisis, aunque se encuentra enfocada hacia el resultado de un punto, también se encuentra relacionada con el resultado de toda una región. Es probable que la idea de construir una red con múltiples salidas y asociarlas a regiones en la vecindad del punto de análisis produzca mejores resultados, simplemente por agregar criterios de evaluación. Esta es la primera modificación que el autor sugiere para continuar la investigación.

Una conclusión mas, respecto al entrenamiento de la red apunta a que no resulta indispensable construir conjuntos de entrenamiento balanceados. Incluso se puede afirmar que los mejores resultados durante la fase de entrenamiento fueron obtenidos utilizando conjuntos de entrenamiento construido utilizando casi la totalidad de la imagen (con lo cual, el porcentaje de participación de la clase “muestra-sin-falla” es muy superior al de la clase “muestra-con-falla”).

Por último se puede afirmar que para el entrenamiento de la red, las modificaciones al algoritmo original de retropropagación son clave, tanto para acelerar el proceso como para incrementar la eficacia. Gracias a esto fue posible duplicar el número de experimentos realizados, con respecto a lo planeado.

6.3. Uso de la red

Los resultados obtenidos por la red satisfacen la mayoría de las expectativas generadas sobre ésta. De hecho, la única expectativa insatisfecha consiste en la incapacidad de producir una generalización total del fenómeno ultrasónico (es decir, que cualquier red fuera capaz de analizar cualquier señal), aunque en el camino se comprendió la dificultad. Ciertamente es necesario considerar, por

ejemplo, que la distribución de la energía absorbida por cada transductor es diferente a lo largo del ángulo de visión, (posiblemente mas intensa al centro y menos en las orillas, por ejemplo). Consecuentemente se puede pensar que, en el intento por disminuir la medida de error durante el entrenamiento, la misma red ajusta los pesos que relaciona a cada entrada compensando las variaciones en intensidad producidas por el cambio de ángulo, vinculándose con un transductor en particular, a la consecuente pérdida de generalidad.

Dentro de la industria, la utilidad de un sistema de análisis de señales ultrasónicas de ensayos no destructivos utilizando redes neuronales responde perfectamente con la creciente competencia global que ha llevado a cambios dramáticos en los procesos productivos. Esta alineación se produce al crear la posibilidad de automatizar un proceso de control de calidad que se traduce en un beneficio inmediato para el consumidor con una mínima inversión en tecnología.

Por otra parte, nos encontramos en un momento histórico en el que los activos de industrias como la aeronáutica o de la construcción encuentran fallas estructurales con cada vez mayor frecuencia (producidas por el proceso natural de envejecimiento), que eventualmente comprometen el funcionamiento de sus equipos o el valor de sus inversiones. Esto incrementa la necesidad de mecanismos de monitoreo que, junto a un sistema como el propuesto, facilita (y habilita) su uso a personas sin entrenamiento especializado.

En general, el enfoque de la industria ha cambiado hacia la idea de producir con una calidad controlada, intentando construir productos sin fallas y evitando la necesidad de revisar si éstos las tienen, sin embargo las fallas y la necesidad de descubrirlas siguen en muchos procesos productivos y en casi todos los procesos de mantenimiento, por lo que la investigación y el desarrollo de herramientas como la propuesta forman parte de un área de investigación sumamente vigente.

Referencias

- [1] S. Liu, I. C. Ume y A. Achari. “Defects Pattern Recognition for Flip-Chip Solder Joint Quality Inspection with Laser Ultrasound and Interferometer.” *IEEE Transactions on Electronics Packaging Manufacturing*, 27(1):59–66, January 2004.
- [2] J. Chang, G. Han, J. M. Velazco, N. C. Griswold, J. F. Duque-Carrillo y E. Sánchez-Sinencio. “Cork Quality Classification System Using a Unified Image Processing and Fuzzy-Neural Network Methodology.” *IEEE Transactions on Neural Networks*, 8(4):964–974, July 1997.
- [3] A.J. McGrail, D.W. Auckland y B.R. Varlow. “Detecting and Classifying Flaws within Insulating Materials using Ultrasound.” *IEEE International Symposium on Electrical Insulation*, June 1994.
- [4] I. Dunlop y A. McNab. “Shape Classification of Flaw Indications in Three-dimensional Ultrasonic Images.” *IEE Proceedings - Science, Measurement and Technology*, 142(4):307–312, July 1995.
- [5] T. Chady, M. Enokizono y R. Sikora. “Signal Restoration Using Dynamic Neural Network Model for Eddy Current Nondestructive Testing.” *IEEE Transactions on Magnetics*, 37(5):3737–3740, September 2001.
- [6] M.S. Unluturk y J. Saniie. “Deconvolution Neural Networks for Ultrasonic Testing.” En IEEE, editor, “IEEE Ultrasonics Symposium,” páginas 715–719. 1995.
- [7] R. Sikora, J. Sikora, E. Cardelli y T. Chady. “Artificial Neural Network Application for Material Evaluation by Electromagnetic Methods.” Informe Técnico, Technical University of Szczecin, Warsaw University of Technology, University of Perugia, Oita University, 1999.

- [17] V. Maurizio, B. Daniela y D. C. Daniele. "Application of Hierarchical Neural Networks to Pattern Recognition for Quality Control Analysis in Steel-Industry Plants." En IEEE, editor, "International Workshop on Neural Networks for Identification, Control, Robotics, and Signal/Image Processing," 1996.
- [18] T. Kohonen. "Self-organized formation of topologically correct feature maps." *Biological Cybernetics*, 43:59-69, 1982.
- [19] G. A. Carpenter y S. Grossberg. "The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network." *IEEE Computer*, 21(3):77-88, 1988. ISSN 0018-9162.
- [20] W. McCulloch y W. Pitts. "A logical calculus of the ideas immanent in nervous activity." *Bulletin of Mathematical Biophysics*, 7:115-133, 1943.
- [21] D. H. Ackley, G. E. Hinton y T. J. Sejnowski. "A Learning Algorithm for Boltzmann Machines." *Cognitive Science*, 1985.
- [22] D. E. Rumelhart, G. E. Hinton y R. J. Williams. "Learning internal representation by error propagation." En D. E. Rumelhart y J. L. McClelland, editores, "Parallel Distributed Processing: Explorations in the Microstructure of Cognition," volumen 1, páginas 318-362. Cambridge, MA: MIT Press, 1986.
- [23] W. D. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley y B. W. Suter. "The multilayer perceptron as an approximation to a Bayes optimal discriminant function." *IEEE Transactions on Neural Networks*, 1(4):296-298, December 1991.
- [24] J. B. Hampshire y B. A. Pearlmutter. "Equivalence proofs for multi-layer perceptron classifiers and the Bayesian discriminant function." En Touretzky, Elman, Sejnowski y Hinton, editores, "Connectionist models," 1990.
- [25] F. Kanaya y S. Miyake. "Bayes statistical behavior and valid generalization of pattern classifying neural networks." *IEEE Transactions on Neural Networks*, 2(4):471-475, July 1991.
- [26] S. Miyake y F. Kanaya. "A neural network approach to a Bayesian statistical decision problem." *IEEE Transactions on Neural Networks*, 2(5):538-540, September 1991.