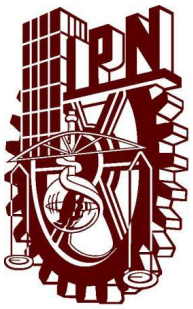


INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

Laboratorio de Robótica y Mecatrónica

METODOLOGÍA PARA LA RECUPERACIÓN DE MELODÍAS MEDIANTE REDES NEURONALES DE RETARDO TEMPORAL

TESIS

QUE PARA OBTENER POR EL GRADO DE:

DOCTOR EN CIENCIAS DE LA COMPUTACIÓN

P R E S E N T A:

M.C. Laura Elena Gómez Sánchez

Directores de tesis:
Dr. Juan Humberto Sossa Azuela
Dr. Ricardo Barrón Fernández

México, D.F. Mayo 2013

SIP-14 BIS

INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 11:00 a.m. del día 13 del mes de Diciembre de 2012 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

"METODOLOGÍA PARA LA RECUPERACIÓN DE MELODÍAS MEDIANTE REDES NEURONALES DE RETARDO TEMPORAL"

Presentada por la alumna:

GÓMEZ	**SÁNCHEZ**	**LAURA ELENA**
Apellido paterno	Apellido materno	Nombre(s)

Con registro:

B	0	8	1	5	0	5

aspirante de: **DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron ***APROBAR LA TESIS***, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de tesis

Dr. Juan Humberto Sossa Azuela

Dr. Ricardo Barrón Fernández

Dr. Sergio Suárez Guerra

Dr. Oleksiy Pogrebnyak

Dr. José Luis Oropeza Rodríguez

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Luis Alfonso Villa Vargas

INSTITUTO POLITECNICO NACIONAL
CENTRO DE INVESTIGACION
EN COMPUTACION
DIRECCION

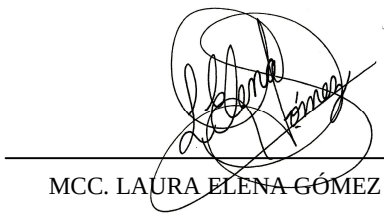
INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de **México, DF** el día **8** del mes **ENERO** del año **2013**, el (la) que suscribe **LAURA ELENA GÓMEZ SÁNCHEZ** alumno (a) del Programa de **DOCTORADO EN CIENCIAS DE LA COMPUTACIÓN** con número de registro **B081505**, adscrito a **CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN**, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de **DR. JUAN HUMBERTO SOSSA AZUELA Y DR. RICARDO BARRÓN FERNÁNDEZ** y cede los derechos del trabajo intitulado **METODOLOGÍA PARA LA RECUPERACIÓN DE MELODÍAS MEDIANTE REDES NEURONALES DE RETARDO TEMPORAL**, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección **lenis45@hotmail.com**. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.



MCC. LAURA ELENA GÓMEZ SÁNCHEZ

A DIOS

Por darme la oportunidad de vivir, por su infinita bondad y amor.

A mi esposo

Julio Fernando Jiménez Vielma

Por compartir tu vida junto a la mía, hombro con hombro, eres el gran amor de mi vida.

A mí amada hija

Lakshmi Michelle Jiménez Gómez

Por ser lo mejor que Dios y la vida me ha dado, mi mayor tesoro.

A mis padres

E. Natividad Sánchez Morales
Miguel Raúl Gómez Pérez

Por ser el pilar fundamental en todo lo que soy y por su apoyo incondicional.

A mi hermano

Miguel Raúl Gómez Sánchez

Por ser una de las personas más especiales en mi vida y apoyarme siempre.

La música es el verdadero lenguaje universal.
Carl Maria von Weber

Sería posible describir todo científicamente, pero no tendría ningún sentido; carecería de significado el que usted describiera a la sinfonía de Beethoven como una variación de la presión de la onda auditiva.
Albert Einstein

La música es el arte más directo, entra por el oído y va al corazón.
Ástor Piazzolla

AGRADECIMIENTOS

Desde estas líneas pretendo expresar mi más sincero agradecimiento a todas aquellas personas que durante estos años de trabajo han estado a mi lado, que de una u otra forma han contribuido a que esta tesis haya llegado a su fin.

Durante esta etapa tuve el privilegio de conocer a excelentes investigadores y personas muy valiosas; a todos ellos gracias por sus enseñanzas, consejos y apoyo que me brindaron para culminar exitosamente esta etapa como estudiante de doctorado.

Antes que nada quiero agradecer a Dios por darme la oportunidad de vivir y por estar conmigo en cada paso que doy, por fortalecer mi corazón e iluminar mi mente y por haber puesto en mi camino a aquellas personas que han sido mi soporte y compañía durante mi vida.

Al Instituto Politécnico Nacional y al Centro de Investigación en Computación por abrirme las puertas en esta etapa valiosa de mi vida. A CONACYT por el apoyo otorgado para la realización de un sueño más. Al Centro de Investigaciones en Óptica, por brindarme la oportunidad de tener acceso a sus instalaciones con las estancias de investigación.

Este trabajo de investigación se realizó en el marco de los siguientes proyectos de investigación: SIP 20121311, SIP 20131182 y CONACYT 155014.

A todas las personas que han estado conmigo, que me han apoyado, que me han aconsejado, que me han dado su amistad, que me han tendido su mano, quiero dedicarles este trabajo. Primero que todo quiero agradecerle primero a Dios por regalarme todas y cada una de las oportunidades a lo largo de este recorrido que apenas inicia, oportunidades que han sido la base para forjar mi carácter y definir el rumbo de mi vida. De manera muy especial quiero mostrar mi gratitud y mi más sincero cariño a mis directores de tesis al Dr. Juan Humberto Sossa Azuela y al Dr. Ricardo Barrón Fernández, por haber confiado en mi persona, por la paciencia y dirección de esta tesis doctoral. Fueron el pilar principal para la realización de esta investigación, gracias por los conocimientos transmitidos, sus consejos y el tiempo que me brindaron para darme un jalón de orejas como para escucharme y orientarme, me han permitido crecer como persona. Para mí no solo han sido excelentes directores de tesis que he podido tener, sino que me han demostrado en todo momento que son amigos a los que siempre podré recurrir, mil gracias.

Al Dr. Oleksiy Pogrebnyak por todo el apoyo que me ha dado, el tiempo que me brindó para atender mis dudas o peticiones. Al Dr. Sergio Suárez Guerra y al Dr. José Luis Oropeza gracias por darme la oportunidad y el tiempo que han dedicado para leer este trabajo.

Al Dr. Francisco Javier Cuevas de la Rosa, por brindarme unos minutos de su tiempo para darme consejos y comentarios que beneficiaron la realización de esta investigación. Gracias por su incondicional apoyo y amistad.

Gracias al personal de la DTE: Lic. Lourdes Olvera, Silvia Arteaga, María del Carmen Morales, María Nieves Galicia, Elda Baranda por su paciencia y orientación que me brindaron estos años, brindándome su ayuda incondicional

en todo momento.

Gracias a mis papás que son una bendición de Dios, por estar siempre a mi lado en todo momento, por creer siempre en mí y apoyarme en todas las decisiones que he tomado a lo largo de mi vida, hayan sido buenas o malas, por enseñarme a luchar por lo que quiero y a terminar lo que he empezado, los amo. Gracias por las enseñanzas y el amor que me muestran día a día.

A mi hermano, que con su amor me ha enseñado a salir adelante, gracias por existir, te amo, pero sobre todo gracias por estar siempre en los momentos más importantes de mi vida.

A mi princesita hermosa Lakshmi, no tengo palabras para expresar todo lo que siento por ti, eres el motor de mi vida, que me impulsa a ser mejor cada día. Por soportar estar separados, siempre mostrando madurez a pesar de estar tan pequeña, sin reproches solo esperando con ansias el término de esta etapa; platicando nuestros sueños y anhelos de estar juntos y disfrutarnos día a día. Tan solo con tu mirada, tu sonrisa o una caricia reflejas todo el amor que tienes para brindarnos, espero en Dios me dé la sabiduría para ser una excelente madre, amiga y confidente que mereces, él ya me dio el mejor de los regalos y tesoros que es el tenerte en mi vida, te amo hija.

A ti, Fer mi gran compañero de vida, mi gratitud y todo mi amor, por estar siempre hombro con hombro en las buenas y en las malas. Porque gracias a tu apoyo incondicional, a tus desvelos y al amor que sin duda me has entregado, me he convertido en una mejor persona. Nuestro camino lleva poco recorrido, aún muchas cosas por compartir, sueños y logros vendrán poco a poco, pero lo más importante es que estamos juntos para vivirlos. Agradezco a Dios infinitamente por permitirme tener a dos grandes tesoros en mi vida, llenando de dicha y amor cada día de mi vida, los amo.

Índice general

Índice de figuras

Índice de tablas

RESUMEN

El uso de las redes neuronales artificiales generalmente se ha catalogado en identificación y control de sistemas, clasificación, pronóstico o predicción debido al gran potencial que tienen para realizar dichas tareas. Dado un conjunto de datos una red neuronal artificial es capaz de extraer relaciones entre los objetos para después predecir valores futuros; en la mayoría de los casos son muy buenas en el reconocimiento de patrones complejos.

Una red neuronal de retardo temporal (TDNN) permite asociar patrones de salida con patrones de entrada. En esta investigación se propone utilizar dicha arquitectura para la codificación de melodías, resultando en un nuevo paradigma reconocimiento de melodías, tomando un nuevo paradigma al no recurrir a la extracción de características utilizadas por diversos investigadores en la recuperación de información musical.

Las redes son entrenadas con las señales originales de cada melodía, y no con descriptores tradicionales. La aportación más importante de esta investigación es que los parámetros internos de la red neuronal funcionan como el descriptor de la melodía. De igual forma esta red se usa como predictor, permitiendo modelar cada melodía. El desempeño de la propuesta es probado usando varias bases de datos.

Descriptores: Recuperación de melodías, Melodías, Redes neuronales

ABSTRACT

The use of artificial neural networks has generally been catalogued into system control and identification, classification, forecast or prediction due to the great potential that such networks have to perform those tasks. Given a set of data, an artificial neural network is capable of extracting relationships among the objects in order to later predict future values; in most cases, these networks are very good at recognizing complex patterns.

A time delayed neural network (TDNN) allows the association of output patterns with input patterns. In the present work, the use of such architecture is suggested for encoding melodies, which turns out to be a new paradigm in melody recognition by not resorting to the extraction of the characteristics used by different researchers in the recovery of musical information.

The networks are trained with the original signals of each melody, and not with traditional descriptors. The most important contribution of this research is that the internal parameters of the neural network serve as the melody descriptor. In the same fashion, this network is used as a predictor, which allows the modeling of each melody. The performance of the proposal is tested using various data bases.

Descriptors: Melody recovery, Melodies, Neural networks.

Términos

En éste apartado se listan los términos usados en éste texto. Se incluye una descripción breve.

Algoritmo genético: Sistema adaptativo, el cual se basa en la evolución y selección natural.

Algoritmo *K*-medias: Es uno de los algoritmos más simples y conocidos de agrupamiento, sigue una forma fácil y simple para dividir una base de datos dada en *k* grupos (fijados a priori).

Algoritmo de correspondencia: Basicamente buscan patrones exactos dentro de una secuencia.

Análisis espectral: Se refiere a la acción de descomponer algo complejo en partes simples o identificar en ese algo complejo las partes más simples que lo forman. Es un proceso que cuantifica las diversas intensidades de cada frecuencia.

Armónico: Es el resultado de una serie de variaciones adecuadamente acomodadas en un rango o frecuencia de emisión, denominado paquete de información.

Autocorrelación: Método que se utiliza para encontrar patrones repetitivos dentro de una señal.

Contorno melódico: Se obtiene al unir las notas de una melodía con una línea continua.

Descriptor: Palabra clave que defina el contenido de un documento.

Distancia de Mahalanobis: Es una medida de distancia, su utilidad radica en que es una forma de determinar la similitud entre dos variables aleatorias multidimensionales.

Distancia Earth Mover's: Es una medida de la distancia entre dos distribuciones de probabilidad sobre una región.

Escala: Es una sucesión ordenada, en forma consecutiva, de todas las notas de un entorno sonoro particular de manera simple y esquemática.

Espectograma: Es el resultado de calcular el espectro de tramas enventanadas de una señal. Resulta una gráfica

tridimensional que representa la energía del contenido frecuencial de la señal según va variando ésta a lo largo del tiempo.

Espectro de frecuencias: Caracteriza qué distribución de amplitudes presenta para cada frecuencia un fenómeno ondulatorio (sonoro, luminoso o electromagnético) que sea superposición de ondas de varias frecuencias.

Firma digital: Es un esquema matemático cuya utilidad es la de demostrar la autenticidad de un documento electrónico o de un mensaje.

Formato HDF5: Es un modelo de datos, biblioteca y formato de archivo para almacenar y gestionar datos. Soporta una ilimitada variedad de tipos de datos, y está diseñado para entrada/salida flexible y eficiente y para datos complejos de gran volumen. Es portable y extensible, permitiendo que las aplicaciones evolucionen en el uso de esta biblioteca.

Frecuencia: Es una medida para indicar el número de repeticiones de cualquier fenómeno o suceso periódico en la unidad de tiempo. Para calcular la frecuencia de un evento, se contabilizan un número de ocurrencias de este teniendo en cuenta un intervalo temporal, luego estas repeticiones se dividen por el tiempo transcurrido.

Frecuencia fundamental: Es la frecuencia más baja del espectro de frecuencias, tal que las frecuencias dominantes pueden expresarse como múltiplos de esta frecuencia fundamental.

Hertz: (Heinrich Rudolf Hertz) Suceso o fenómeno repetido una vez por segundo, casi siempre hay una relación en el número de Hertz con las ocurrencias. Originalmente se conoció como ciclo por segundo. Las pulsaciones del corazón o el tempo musical se miden como golpes por minuto.

Huella digital: Es un mecanismo para defender los derechos de autor y combatir la copia no autorizada de contenidos, que consiste en introducir una serie de bits imperceptibles sobre un producto de soporte técnico de forma que se puedan detectar las copias ilegales.

Indexación: Su propósito es ejecutar la elaboración de un índice que contenga de forma ordenada la información, esto con la finalidad de obtener resultados de forma sustancialmente más rápida y relevante al momento de realizar una búsqueda.

Longitud de onda: Es la distancia comprendida entre dos crestas o dos valles.

MIDI: Protocolo ideado para comunicar instrumentos musicales entre sí, y también con ordenadores; es como la “partitura” de una pieza musical.

Onda: Es una propagación de alguna propiedad de un medio, por ejemplo, densidad, presión, campo eléctrico o campo magnético, que se propaga a través del espacio transportando energía. El medio perturbado puede ser de naturaleza diversa como aire, agua, un trozo de metal, el espacio o el vacío.

Periodo: Es el tiempo transcurrido entre dos puntos equivalentes de la oscilación. Es el lapso mínimo que separa dos instantes en los Sue el sistema se encuentra exactamente en el mismo estado. El periodo de oscilación de una onda

es el tiempo empleado por la misma en completar una longitud de onda.

Sonido: Radica en ondas sonoras consistentes en oscilaciones de la presión del aire, que son convertidas en ondas mecánicas en el oído humano y percibidas por el cerebro.

Tono: Es la cualidad que permite distinguir un sonido grave de otro agudo.

Tono de contorno: Son tonos que cambian durante su realización, esto es, puede ser ascendente, descendente o cualquier elevación y descenso.

Transformada de Fourier: En procesamiento de señales, suele considerarse como la descomposición de una señal en componentes de frecuencias diferentes. Es básicamente el espectro de frecuencias de una función.

Ventana de Hanning: Forza las extremidades hacia cero, pero también agrega distorsión a la forma de onda que se está analizando, bajo la forma de modulación de amplitud, eso es la variación en amplitud de la señal sobre la grabación de tiempo.

Capítulo 1

Introducción

Desde la antigüedad se ha tenido la necesidad de controlar el almacenamiento y la recuperación de información, por lo mismo se ha llegado a la creación de métodos y técnicas que permitan dichas actividades, así como la conservación e identificación de la misma. A pesar de ser tareas simples, el acceso rápido a ella se está haciendo cada vez más difícil. La idea de utilizar computadoras para la búsqueda de fragmentos relevantes de información se dio en [13].

Con la invención en 1946 de las tecnologías computacionales fue de progresiva e inmediata aplicación en la naciente esfera de la información, especialmente para solucionar las preocupaciones dominantes en el lapso de la explosión documental, sobre como localizar y buscar información puntualmente.

En 1950, el norteamericano Calvin Mooers, propuso la creación de un área que afrontara “los aspectos intelectuales de la descripción de la información y sus especificaciones para la búsqueda, además de cualquier sistema, técnica o instrumento que se utilice en la operación: la recuperación de información” [76].

Pese a que en 1953, en Gran Bretaña y Estados Unidos se realizaron pruebas para evaluar el desempeño del entonces controvertido sistema “Uniterm” de Mortimer Taube, primer sistema enfocado a la temática de indexación y recuperación de información; no obtuvieron el peso suficiente para establecerse como el inicio emblemático de este campo. No fue sino hasta 1957 en el Cranfield Institute of Technology y otras instituciones asociadas, cuando comenzaron a realizar una serie de pruebas que representaron el comienzo real de la investigación sobre recuperación de la información como disciplina empírica.

Una ilustración básica de cualquier sistema de recuperación de información (SRI) se muestra en la Fig. 1.1.

Un SRI consta de tres componentes fundamentales de cualquier sistema: entrada, salida y proceso. El problema principal es el obtener una representación y el documento de consulta adecuado para que la computadora lo procese. La mayoría de los sistemas de recuperación sólo almacenan una representación del documento, lo que significa que dicho documento se pierde una vez que se haya procesado.

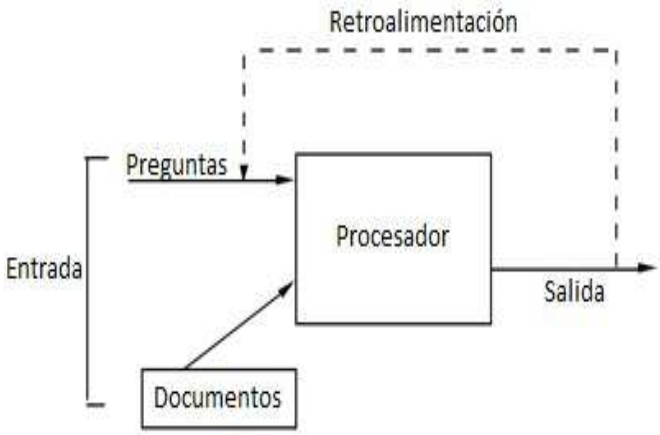


Figura 1.1: Sistema de recuperación de información (SRI)

El proceso de recuperación de información consiste en extraer de una colección de documentos, aquellos que se ajustan a las especificaciones de una determinada petición. En concreto, se trata de una comparación sistemática entre los documentos y la petición de información. La operación resulta relativamente sencilla, sin embargo el ignorar detalles conduciría a una duplicación de la información en los resultados.

Cuando el sistema de recuperación de información se encuentra en línea, es posible para el usuario cambiar su petición durante la búsqueda, para obtener una mejora en la recuperación, mediante estrategias de búsqueda en respuesta a una consulta dada. A este procedimiento se le conoce comúnmente como retroalimentación. Por último, la salida es normalmente un conjunto de citas o números de documentos [86].

Para tratar de evitar ese problema, se tiene que hacer uso de los descriptores de búsqueda, teniendo en cuenta las posibles formas de expresión de un concepto (sinónimos, conceptos más generales y más específicos, etc.) ya que de otro modo podría llegarse a una pérdida importante de información. Por lo tanto, es importante preparar adecuadamente el perfil de búsqueda, operación que resulta la más importante en el proceso de recuperación de información. Finalmente habrá que comparar si los documentos obtenidos satisfacen los requisitos del solicitante. A menudo, la información que en verdad se necesita no coincide exactamente con lo que se pide, esto se debe, normalmente a que el usuario no conoce con precisión sus necesidades.

La recuperación de información es un campo interdisciplinario; cubre tantas disciplinas que eso genera normalmente un conocimiento parcial desde tan solo una u otra perspectiva. En los últimos años, gran cantidad de material musical es accesible a usuarios domésticos a través de redes y almacenamiento masivo, por lo que el campo musical, junto a las ciencias de la computación, planteó una disciplina derivada de la recuperación de información, conocida como recuperación de información musical (MIR por sus siglas en inglés), la cual comenzó a madurar a finales de 1990.

Kassler en el trabajo [35] dio a conocer el primer sistema programado en lenguaje ensamblador, utilizando partituras para encontrar posiciones que cumplieran ciertos criterios; poco después, las limitantes comenzaron a surgir, una de ellas era el reconocer si una melodía era copia de otra. Su investigación tomo un giro enfocándose en el problema de reconocimiento musical óptico, que consiste en transcribir las partituras impresas a un formato de almacenamiento

denominado MIDI.

El querer automatizar el acceso a la información de la música a través del uso de las computadoras digitales ha intrigado a musicólogos, informáticos, bibliotecarios e incluso a los amantes de la música por igual; cada uno tiene su propósito en mente, por lo tanto pueden apreciarse muchos enfoques. De acuerdo a ediciones pasadas de la Computing in Musicology (Hewlett and Selfridge-Field, eds.) que se han dado a la tarea de informar sobre esta disciplina. Se puede observar que algunos investigadores han diseñado complejas herramientas informáticas [33] para analizar todas las facetas que ha vivido la música.

Hasta hace dos décadas es que los investigadores, se han enfocado realmente a la recuperación de información musical automática. Inicialmente la música se clasificaba por título, compositor o género, comúnmente conocido como metadato; información que requería conocer el usuario para una búsqueda satisfactoria, dado a que no todo usuario era experto en el tema, los resultados eran limitados [34]. En consecuencia, la investigación actual se enfoca a características propias de la música, basándose en su contenido musical; mediante consultas por tarareo. Actualmente se enfrenta a muchos obstáculos teóricos y técnicos.

Un punto vital que une a todos los enfoques hasta ahora analizados es que tienen algún tipo de deficiencia.

1.1. Hipótesis

Mientras se sigan utilizando descriptores tradicionales para la recuperación o recomendación de información musical, estos sistemas seguirán teniendo sus inconvenientes y limitantes, por lo anterior es necesario probar si es posible la particion y codificacion de una melodia mediante los pesos de una red neuronal de retardo temporal para su posterior recuperación.

1.2. Objetivo

Poner en operación una nueva metodología que utilice redes de retardo temporal para la codificación y recuperación de melodías.

1.2.1. Objetivos particulares

A continuación se describen los objetivos particulares que se persiguen con esta investigación:

1. Se realizó una nueva técnica para codificar melodías mediante los pesos sinápticos de una red neuronal de retardo temporal.

2. Tomando como base los resultados obtenidos en el punto anterior, se realizó la recuperación de melodías mediante redes de retardo temporal.
3. Se realizó la recuperación ó recomendación de melodías mediante el uso de redes neuronales de retardo temporal, utilizando su estructura como descriptor propio.
4. Para la recuperación ó recomendación no se realizó ningún pre procesamiento a las melodías.
5. Se utilizaron melodías con diferentes frecuencias de muestreo, para obtener un menor costo computacional en el entrenamiento de las melodías.

1.3. Aportaciones científicas

A continuación se describen las aportaciones que surgen a partir de esta investigación:

1. Se aplicó un modelo de red neuronal de retardo temporal (TDNN) para la recuperación ó recomendación de información musical.
2. El descriptor utilizado no proviene de ninguno tradicional, sino del mismo reconocedor.

1.4. Organización de la tesis

A continuación se describe el contenido de cada capítulo que compone esta tesis doctoral:

1. En el capítulo 1, se da una introducción al tema de recuperación de información en general, posteriormente mostrando el enfoque de recuperación de información musical.
2. En el capítulo 2, se presenta un breve estado del arte donde se describen las aportaciones relevantes en los últimos años.
3. En el capítulo 3, se describen los conceptos necesarios utilizados en los siguientes capítulos.
4. En el capítulo 4, se describe la metodología propuesta para esta investigación mediante el uso de redes neuronales de retardo temporal para la recuperación o recomendación de melodías.
5. En el capítulo 5, se presentan los experimentos y resultados realizados con redes neuronales de retardo temporal para la recuperación o recomendación de melodías.
6. En el capítulo 6, se presenta una serie de conclusiones obtenidas hasta el momento, así como trabajo a futuro.

Capítulo 2

Estado del arte

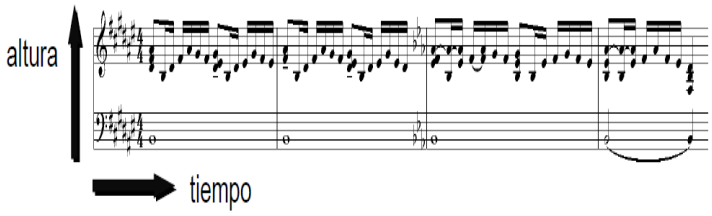
El ser humano tiene capacidades y habilidades que ha desarrollado con el paso del tiempo, destacando la imaginación, memoria, pensamiento, percepción, emociones y sentimientos. Debido a ello puede reconocer infinidad de objetos, escenas o melodías que se encuentren en su entorno, sin embargo en algunas ocasiones no cuenta con la información suficiente para realizarlo o simplemente no recuerda la letra de una melodía que sea de su agrado. Gracias a los sistemas de recuperación que existen en diversas áreas de investigación en la actualidad, se puede realizar una recuperación satisfactoria. Enfocándonos en la recuperación de información musical, se puede obtener desde el título, autor, intérprete, así como la reproducción de la melodía que se estaba buscando.

2.1. Recuperación de información musical

Ha sido definida por Stephen Downie como “la investigación multidisciplinaria que se esfuerza por desarrollar sistemas innovadores de búsqueda basados en el contenido, interfaces novedosas y mecanismos para que el extenso mundo de la música esté al alcance de todos” [18]. Esta área se organiza de acuerdo a los casos según cada tipo de consulta y de acuerdo a la forma de comparar la entrada con la salida. La consulta y la salida pueden ser mediante información textual (metadatos), fragmentos de música, grabaciones, partituras o características de la música. Esta disciplina se divide en tres áreas principales de estudio, en las figuras 2.1(a), 2.1(b) y 2.1(c) se puede observar la representación de cada una de ellas:

2.1.1. Análisis simbólico

A mediados de los años 90’s muchos investigadores estudiaron sobre el problema de similitud en la música mediante el análisis de representaciones simbólicas, tales como datos de música MIDI, partituras musicales, seguimiento de tono, así como las técnicas de concordancia utilizadas para comparar las transcripciones de cada

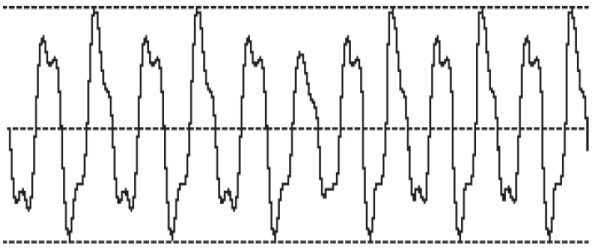


(a)

TEST.MID

```
/home/jose/Project/Library/Final/test.mid;819;1;120;6;
Copyright © 2005 by J M Inesta;
4/4;C major;1;1;1;
1;Tenor Sax       ;1; 69;1;mono;0.000;1.000;1; 1;0;0;0;
2;Electric guitar ;3; 42;3;mono;0.002;0.025;2; 1;0;0;0;
3;Fingered bass   ;5; 58;5;mono;0.010;0.825;2;18;0;0;0;
4;Drums           ;7;176;7;poly;0.500;0.500;2; 4;0;0;0;
5;Honky Tonk piano;9; 74;9;mono;0.000;0.000;2; 0;21;33;0;
```

(b)



(c)

Figura 2.1: a) Análisis simbólico (Recuperación a través de partituras), b) Metadatos (Recuperación usando metadatos), c) de señales acústicas (Recuperación mediante señales sonoras musicales).

canción ([22], [52], [84], [83], [59], [67]).

Los archivos MIDI poseen información importante que ha sido utilizada para el reconocimiento de melodías. En la Fig. 2.2 se puede observar la representación de las notas con archivos MIDI. Con el uso de la frecuencia fundamental y la notación U, D y R se describe el estado de las notas obtenidas por el contorno melódico, si la nota es mayor, menor o igual a la nota anterior respectivamente. Al aplicar autocorrelación se realiza la identificación de melodías [22].

En McNab [53], se menciona que su investigación se centra en la recuperación de partituras musicales, transcribiendo automáticamente una melodía; las notas se dividen en segmentos, se identifica la frecuencia de cada nota y se realiza un etiquetado. Para comparar las melodías se determina la exactitud del contorno melódico y el intervalo de las secuencias de las notas, como se observa en la Fig. 2.3 donde se muestra una gráfica generada por esta propuesta. A diferencia de los métodos basados en cadenas, las melodías son vistas como un conjunto de eventos, las notas son definidas por su aparición en el tiempo, el tono y su duración. Este método considera a las partituras y las consultas como conjuntos de notas, pero en lugar de encontrar superconjuntos, utiliza la distancia Earth Mover's para comparar conjuntos.

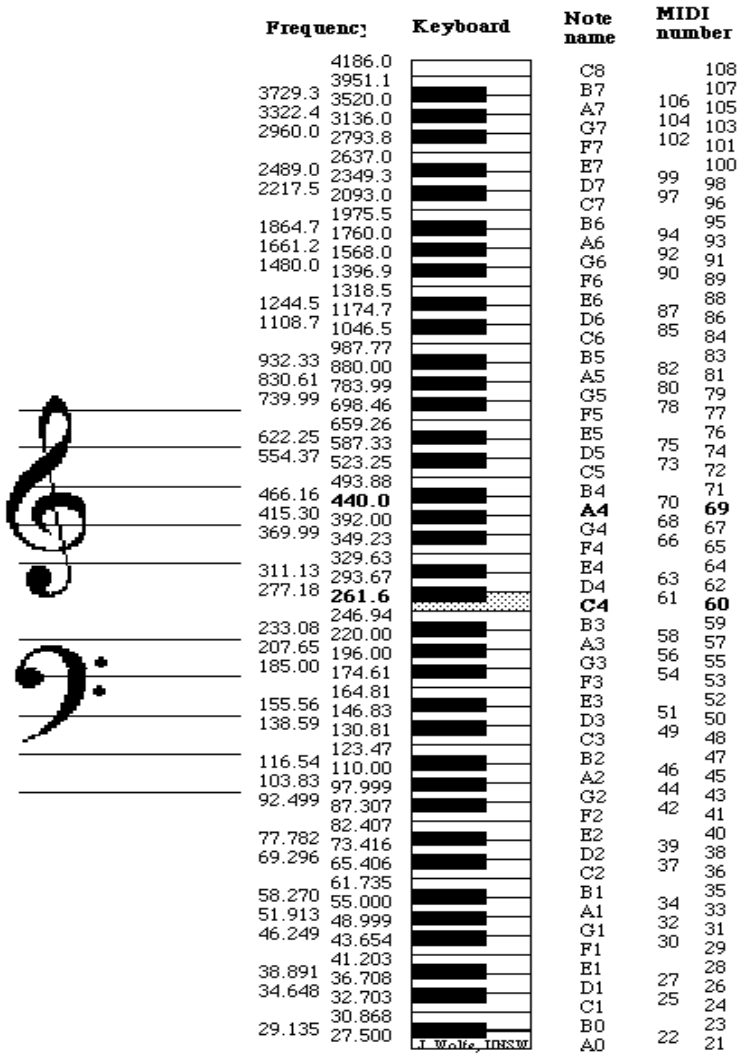


Figura 2.2: Representación de notas musicales en MIDI

Las melodías pueden ser representadas por una cadena de caracteres, donde cada caracter describe una nota o un par de notas consecutivas; de las cuales se pueden obtener secuencias de intervalos y de tonos, contornos brutos (describen la dirección de los intervalos), etc.; los algoritmos de correspondencia, la búsqueda de la subsecuencia común más larga o de las ocurrencias de una cadena en otra, son utilizados para el cálculo de distancias y de esta manera se puede encontrar la similitud entre melodías. Con frecuencia, la variación de una melodía que se llega a percibir como melódicamente similar, puede contener muchas notas más; por ello el cálculo de una medida de similitud para las cadenas sin producir falsos positivos no es fácil. Los algoritmos de búsqueda Knuth-Morris-Pratt y Boyer-Moore [36], [17] se han hecho presentes en ese tipo de reconocimiento.

El objetivo de utilizar métodos probabilísticos, es para determinar las propiedades probabilísticas de las melodías candidatas y compararlas con las obtenidas de las consultas. Un claro ejemplo, es el sistema GUIDO [30], en la Fig. 2.4 se puede observar un ejemplo de la notación musical que utiliza dicho sistema, donde se utilizan las cadenas de Markov para modelar los contornos melódicos y rítmicos de las melodías, designando una cadena para cada melodía.

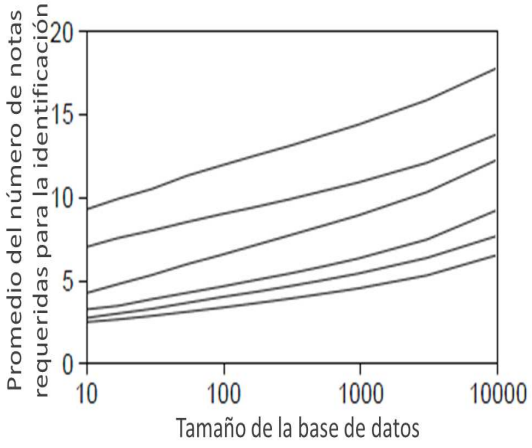
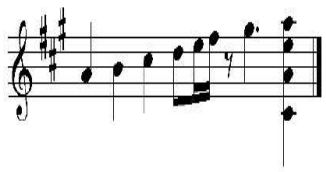


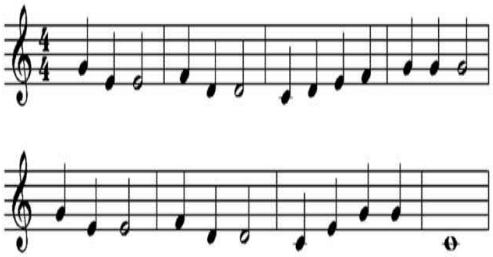
Figura 2.3: Visualización de una consulta con el sistema propuesto por McNab [53]

En cada cadena de Markov, los estados de transición pueden corresponder a cierto intervalo de tiempo o duración de la nota, y las probabilidades de transición revelan la cantidad de ocurrencias de los diferentes estados posteriores. Las matrices de transición están organizadas en forma de árbol con el objetivo de descartar datos con probabilidades de transición cero en una etapa temprana de la búsqueda.

```
[ \key<"A"> a1/4 h c#2 d/8 e/16 f#16 _/8
  g#*1/4. {a1/4,c#,e2,a2} ]
```



```
[ \key<"C"> \meter<"4/4"> g1/4 e e/2
f/4 d d/2 c/4 d e f g g g/2 g/4 e e/2
f/4 d d/2 c/4 e g g c/1 ]
```



```
{ [ \tempo<"Vivace">
  \meter<"5/8"> \intens<"p"> \sl(\bm(g1*1/8 a b)
  \bm(b& c2) \bm(c# b1 a b& a&)) ],
[ \meter<"5/8"> \sl(g1*3/8 d/4 c#*3/8 d/4) ] }
```

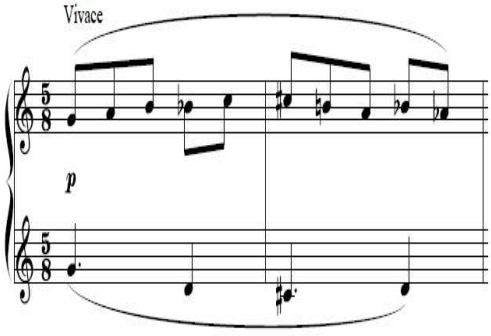


Figura 2.4: Notación musical utilizada por GUIDO

Su enfoque consta de tres capas, cada una encuentra las ocurrencias de las invariantes de transposición en un patrón de consulta dado dentro de una base de datos. Un patrón que se repite varias veces es utilizado eficientemente para respuestas rápidas al usuario, mientras que un patrón cuando es detectado en raras ocasiones toma más tiempo en una búsqueda futura [45].

Musipedia en [61], [80], [81] y [82], solamente trata de calcular la distancia existente entre la cadena consulta y las que se encuentran en su base de datos, sin importar las longitudes de éstas, por lo que sería necesario elegir subcadenas adecuadas para calcular dicha distancia. Se propone un buscador rápido de melodías, que puede recuperar melodías de una base de datos basándose en la frecuencia de consulta de las canciones. Los temas son recogidos de las consultas realizadas por los usuarios, el índice se incrementa y se actualiza.

La música se puede representar, por medio de partituras musicales, tales como el formato MIDI, la otra forma se basa en las señales acústicas que son muestreadas en una determinada frecuencia. La recuperación de música basada en su contenido suele apoyarse en un conjunto de características extraídas tales como: tono, intervalo, duración y escala. Un enfoque común es representar a la música como una cadena utilizando dichas características. Para representar el contorno de tiempo se pueden utilizar tres caracteres: U (up), D (down) y R (repeat) con el fin de encontrar cadenas similares de la melodía fuente.

Cuando se realiza una consulta tarareando una melodía o mediante la notación basada en el contenido, el sistema la interpreta como una señal o una sucesión de notas respectivamente, de los cuales extrae características como el tono y el contorno del tiempo en [66] y [67]. La notación LSR permite describir el contorno de tiempo calculando la duración de las notas, se clasifican de tres maneras: R (por una repetición en un tiempo previo), L (por un tiempo largo en un tiempo previo) y S (por un tiempo corto en un tiempo previo).

En la Fig. 2.5 se puede ver el diagrama general del sistema Musemble.

El sistema de búsqueda rápida de melodías crea un índice dinámico de consulta para melodías frecuentes. Transformando a cadenas la información extraída de las melodías, mediante las notaciones UDR y LSR. Se busca primero en la base de datos, si se encuentra una coincidencia con el índice, se ajusta la entrada de las variables y se incluye en el conjunto de resultados [68].

Se propone un nuevo régimen para la reformulación de consultas para mejorar el rendimiento de recuperación de información musical. Las melodías son analizadas y representadas con las notaciones antes mencionadas, tomando como principales características la altura o la duración de cada nota. Existen diversas técnicas para analizar y extraer el tono de contorno, intervalo y duración de las consultas vocales. En general, los métodos para la detección de tono y la duración de la música puede dividirse aproximadamente en dos categorías: dominio del tiempo y el dominio de frecuencia. En el dominio del tiempo las técnicas utilizadas son el cruce por cero y la autocorrelación. Para el dominio de frecuencia, la transformada rápida de Fourier (FFT) que se basa en la propiedad de que cada forma de onda se puede dividir en simples ondas sinusoidales.

Para mejorar la eficiencia en la recuperación de información (IR) se realiza una reformulación de la consulta, mediante una retroalimentación con técnicas evolutivas aplicando un algoritmo genético para dicha recuperación, sin embargo aún no ha sido ampliamente adoptado en el ámbito de recuperación musical ([24] y [63]).

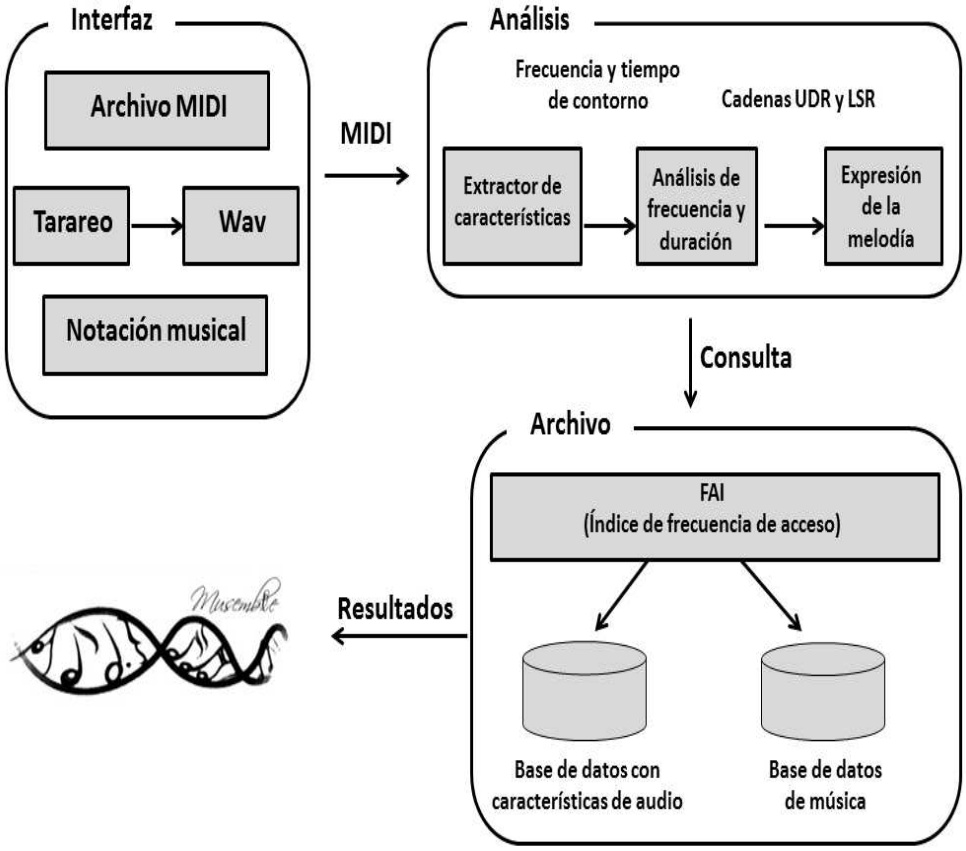


Figura 2.5: Sistema Musemble

Se proponen dos métodos para el análisis de energía y se incorpora una red neuronal feedforward (FFNN) para reducir el espacio de búsqueda, ésta se compone de las tres capas: entrada, oculta y salida. Su aprendizaje es supervisado y todas las neuronas se activan por medio de la función sigmoide [64]. En la Fig. 2.6 se muestra el modelo de la red neuronal.

La base de datos almacena todos los segmentos de las melodías y son clasificados mediante el algoritmo *K*-medias como preprocesamiento. El patrón de entrada se obtiene de la población que fue generada por el algoritmo genético mediante selección de torneo; dicho patrón se forma por el tono y la duración. El paso siguiente es comparar el patrón de entrada con el de salida en cada categoría para encontrar la similitud entre estos. Una vez que se han procesado las características se muestra un resultado al usuario, debiendo retroalimentar al sistema si es que el resultado no ha sido el esperado [65]. La Fig. 2.7 muestra un diagrama general de este sistema con las últimas actualizaciones realizadas.

2.1.2. Metadatos

Por medio de los metadatos universales de las melodías y de extraer una serie de características que representan el espectro, el ritmo y los cambios de acorde, se reúnen en un único vector para determinar la similitud de las melodías. A partir de las agrupaciones que realizó en [49] se determina la similitud al comparar los modelos obtenidos.

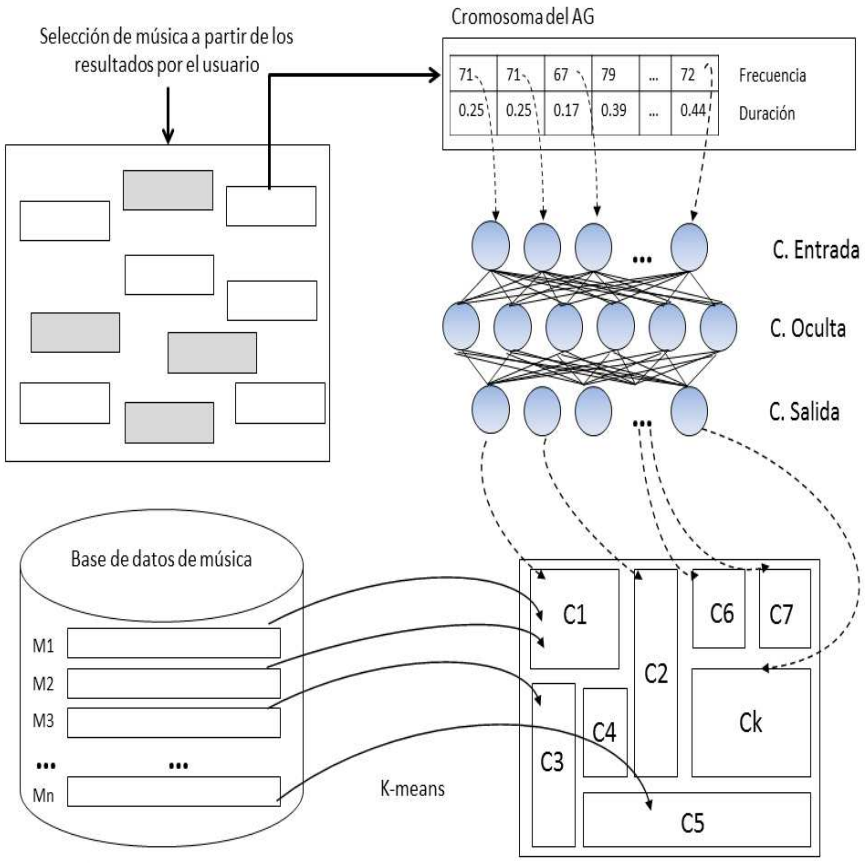


Figura 2.6: Estructura del modelo de FFNN en Musemble

Los servidores MusicBrainz, Moodlogic o AMG proporcionan dos tipos de metadatos a los sistemas de Distribución de Música Electrónica (EMD por sus siglas en inglés) y a los editores de metadatos. El primer tipo se enfoca a la información general como nombre del artista, canciones, álbum, etc., el siguiente es más detallado en cuanto al género, instrumentos, estilo, entre otros. Las búsquedas pueden ser por medio de un servidor local o por internet, como se muestra en la Fig. 2.8, el usuario debe proporcionar los datos requeridos para que el sistema muestre una lista de posibles melodías que consultó [40].

Pachet en el 2005 [58] también tiene catalogados los metadatos, la información general es un factor importante en este tipo de almacenamiento, una clasificación o asociación de los diferentes géneros que tenga un artista, así como un pequeño análisis acústico sobre el ritmo de las melodías.

Se evalúa una nueva técnica híbrida para recuperación de información musical basado en metadatos y consultas por zumbido (HQMS) presentando dos filtros en serie: el de metadatos y por consulta tarareada, como se muestra en la Fig. 2.9. En el primer filtro, el usuario debe proporcionar información que esté ligado con el metadato de la canción que desea buscar, logrando así la reducción a un conjunto de resultados posibles que van a un almacén temporal. Las melodías que se encuentran en la base de datos previamente fueron transformadas a series de tiempo contando con un indexado. En el segundo filtro, la consulta también es transformada, de esta manera se realizan los gráficos correspondientes, para ver la similitud que existe [2].

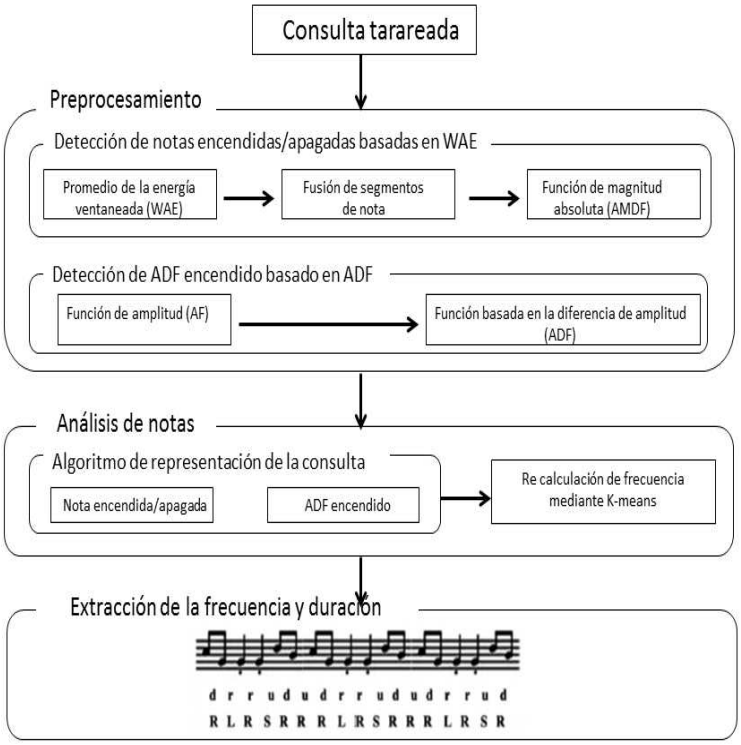


Figura 2.7: Preprocesamiento de sistema Musemble

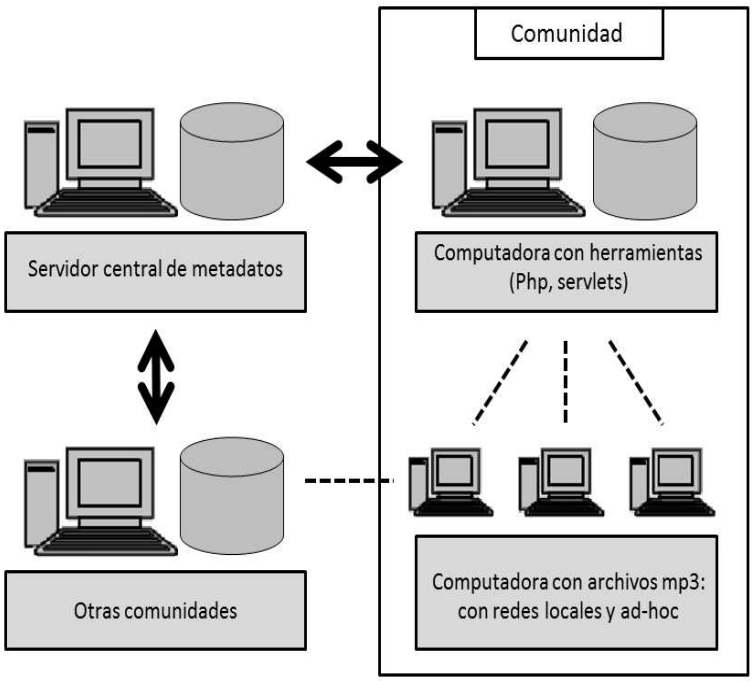


Figura 2.8: Interacción entre servidores

Diseñar la base de datos con un millón de melodías, proporciona metadatos y análisis de audio, capturando la mayor parte de la información, los datos se almacenan utilizando el formato HDF5 para el manejo eficiente de los

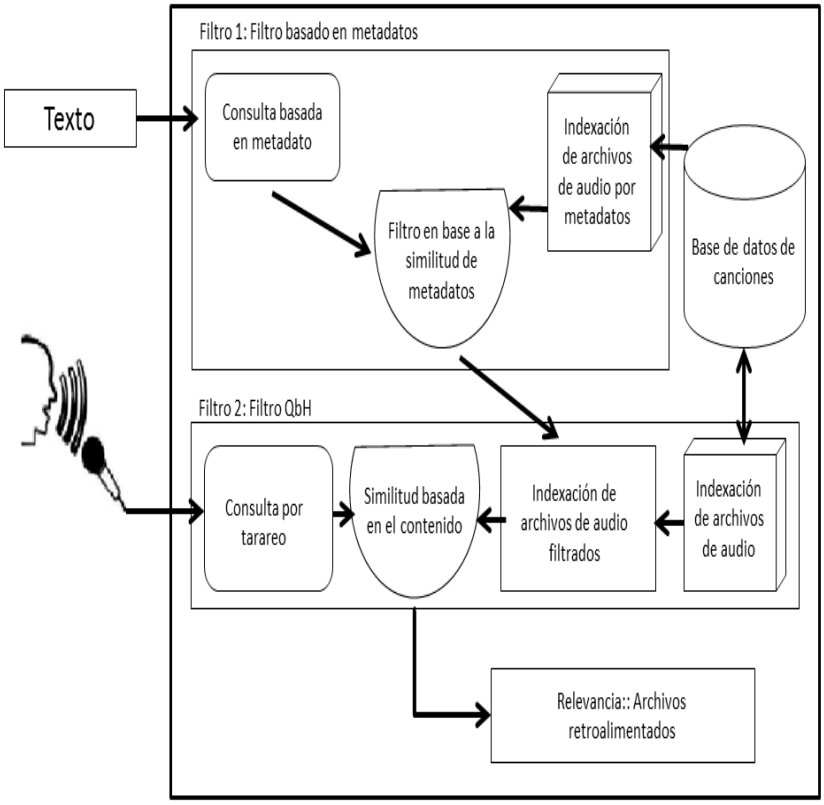


Figura 2.9: Sistema HQMS

datos heterogéneos de información, tales como nombres del artista, género, autor, entre otros, de igual forma las características acústicas como el tono, timbre y volumen [5].

Se propone un enfoque basado en el contenido mediante el análisis de timbre, tiempo y tono además del uso de metadatos. Siendo de gran ayuda el uso de este último, para mejorar los resultados obtenidos, no basta la recuperación musical basándose en el contenido de las melodías [9].

2.1.3. Análisis de señales acústicas

Existen tres modelos que se han utilizado para la generación del sonido musical: modelos físicos o de instrumentos musicales (describen el sonido a partir de parámetros mecánicos y acústicos), modelos del espectro (basados en el espectro STFT, se extraen características de un sonido real, se usan los componentes sinusoidales así como la reproducción del sonido original y transformaciones), y los modelos abstractos (a partir de una fórmula abstracta).

La Síntesis del Modelo Espectral (SMS por sus siglas en inglés) es una técnica de análisis que extrae las características perceptivas de una amplia variedad de sonidos, la representación que resulta del análisis es intuitiva y fácil de asignar a parámetros musicales útiles. La parte central del sistema es el análisis, se trata de un algoritmo complejo que requiere la configuración manual de unos cuantos parámetros de control. El trabajo adicional puede

automatizar el proceso de análisis, en particular si hay una especialización para un grupo de sonidos.

La síntesis de la representación estocástica determinista es simple y se puede realizar en tiempo real, dicha representación se almacena una vez que ha sido calculada y la transformación del sonido se realiza de forma interactiva [78].

Una forma de comparar grabaciones de audio de forma significativa, es extraer una descripción abstracta de la señal de audio. En la tabla 2.1 se observar la lista de algunas características que analizó en segmentos con una duración de 25 y 40 milisegundos [93].

Tabla 2.1: Características consideradas como importantes para Wold

Característica	Descripción
Volumen	Por medio de la energía de la señal, determina si la música que se percibe es intensa o fuerte
Frecuencia fundamental	La transformada de Fourier proporciona la descomposición de una señal en componentes de frecuencias diferentes
Croma	Vector que contiene la energía espectral de cada una de las 12 clases de tonos tradicionales de la escala
Tono (brillo y ancho de banda)	El brillo es la medida de la frecuencia más alta de una señal. El ancho de banda se puede calcular como el promedio de la magnitud ponderada entre los componentes espectrales y la Transformada Corta de Fourier
Coeficientes Cepstrales de las Frecuencias de Mel	Son una representación de una señal ventaneada en el tiempo que ha sido derivada de aplicar la Transformada Rápida de Fourier, pero en una escala de frecuencias no lineal, las cuales se aproximan al comportamiento del sistema auditivo humano

Presenta un sistema para recuperar documentos de audio, mediante una indexación de melodías basado en histogramas de MFCC (Coeficientes Cepstrales de las Frecuencias de Mel). Investiga la aplicación de dos algoritmos de correspondencia aproximada para recuperar música [21].

Se da paso al primer sistema diseñado para recuperar melodías de una base de datos llamado MELDEX. El usuario debe proporcionar pocas notas cantadas de alguna melodía por medio de un micrófono, las cuales son transformadas en notaciones musicales; posteriormente se busca en la base de datos el patrón o patrones similares a la consulta. Se hace uso del contorno melódico, intervalos musicales y el ritmo, son analizados a través de un algoritmo de programación dinámica para comparar secuencias musicales [52].

En Blackburn [7] presenta un sistema que emplea el uso de los contornos melódicos basado en el contenido incluyendo archivos de audio digital y formato MIDI. Se centra en trabajar directamente con las señales de audio, analizando el timbre y aspectos temporales del sonido en lugar de las descripciones teóricas de la música y sin realizar alguna transcripción de la música. Se propone la extracción de características por medio del análisis de la Transformada Rápida de Fourier (FFT), Codificación de Predicción Lineal (LPC), Coeficientes Cepstrales de las Frecuencias de Mel

(MFCC) almacenando la información en vectores, que son utilizados para el reconocimiento de patrones mediante técnicas estadísticas, esto se observa en la Fig. 2.10 [16].

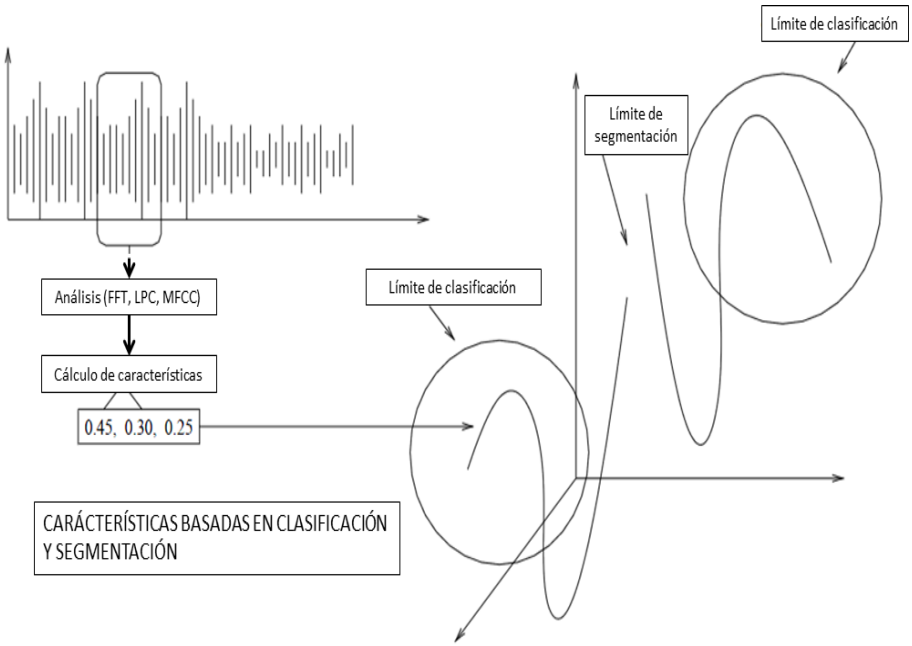


Figura 2.10: Clasificación y segmentación de segmentos

Se presenta un método para comparar melodías basándose únicamente en el contenido del audio. Analizando la firma espectral de cada melodía y clasificándolas mediante el algoritmo K-means, en [49].

El sistema Shazam sirve para la identificación de música, utilizando huellas digitales. El extractor de características se utiliza para describir segmentos cortos de una grabación, ignorando las distorsiones típicas causadas por altavoces, micrófonos sencillos y conexiones a teléfonos móviles, así como el ruido de fondo que pueda existir. Por lo general, sólo unos bytes por cada segmento de la grabación, se almacenan en un índice de la base de datos, junto con los apuntadores donde se producen las grabaciones [14].

Basándose en las preferencias musicales del usuario y de su género, propone un método de recuperación musical, permitiendo a dicho usuario descubrir nuevas canciones que podría desear posteriormente, mediante métodos de retroalimentación por relevancia para mejorar el rendimiento de la propuesta y así reducir la carga de usuarios en los datos de entrada al sistema de aprendizaje. Mediante el método TreeQ se entrena un cuantificador vectorial Fig. 2.11, en lugar de una modelización de los datos de sonido, por medio de una representación espectral, realizando el cálculo de los coeficientes cepstrales de frecuencias de Mel transformándolos en vectores de características de 13 dimensiones (12 coeficientes MFCC más energía). [29].

Utiliza un extractor de características para la conversión de señales PCM (Modulación por Pulsos Codificados) en grupos que pueden ser tratados de la misma manera como conjuntos de notas [46], [44] y [85].

Se presenta un nuevo sistema de consulta por tarareo, su diagrama se puede observar en la Fig. 2.12, buscando en una base de datos los temas relacionados a la consulta mostrando al usuario dichas opciones. El sistema captura

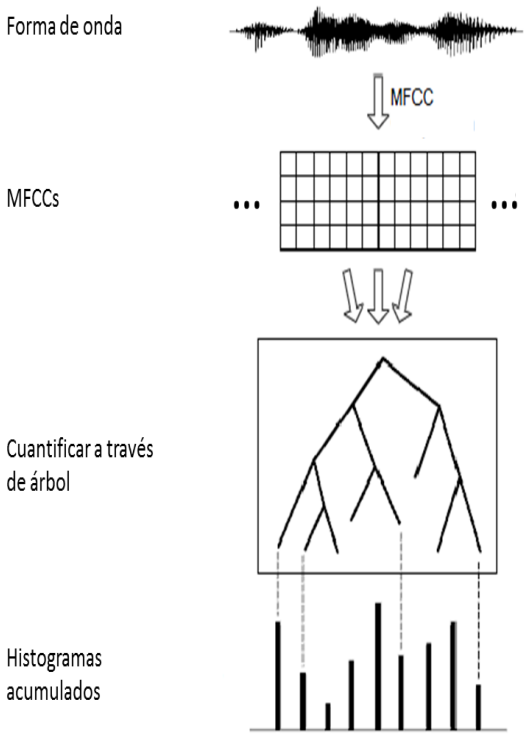


Figura 2.11: Esquema de cuantificación vectorial y el método TreeQ

ventanas de 10 milisegundos de audio, analizando el tono, la amplitud y la frecuencia fundamental que son almacenados en vectores para utilizarlos posteriormente. Una vez que se tienen dichos vectores, por medio de la distancia Mahalanobis se determina la similitud que haya entre el vector capturado y los que se encuentran en la base de datos [48].

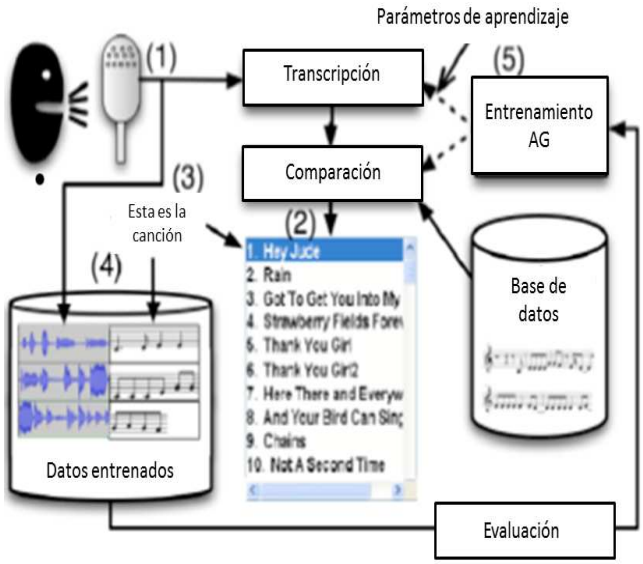


Figura 2.12: Esquema de cuantificación vectorial y el método TreeQ

Se enfoca en la recuperación de información en varios idiomas, de información multimedia y de información basada en la semántica. Con la explosión de los medios de comunicación digital que está disponible en Internet y mediante técnicas de búsqueda permite presentar a los usuarios de forma rápida y precisa, resultados posibles. La recuperación en el área de multimedia no es por medio de texto, sino por la música e imágenes. La recuperación de información basada en la semántica va más allá de la recuperación de información clásica sino mediante el uso de los conceptos representados en documentos y consultas para mejorar el rendimiento de la recuperación [62].

Se describe un enfoque actual para la extracción automática de melodías, la metodología utilizada es mediante el análisis espectral y la estimación de la frecuencia fundamental de la señal de audio, para obtener el espectrograma, terminando esta fase se calcula la frecuencia instantánea que sirve para detectar los picos espectrales que permitirán encontrar los inicios de los tonos más altos. Una vez que se tienen los tonos se procesan para construir voces, se dice que una voz se define por su magnitud, su frecuencia y el rango de su frecuencia.

Desde 1940 se han realizado muchos intentos para diseñar sistemas automáticos para la clasificación de melodías. En la actualidad, el patrimonio cultural de algunas instituciones ha dado alta prioridad a la digitalización y clasificación de colecciones de melodías, intentando apoyar la investigación Folk Song (FSR) para la clasificación e identificación de las posibles variantes que tienen las canciones populares existentes en la base de datos del Instituto Meertens basándose en el contenido melódico utilizando diferentes modelos para la recuperación de información musical que pueden ser explotados en los motores de búsqueda por medio de la musicología cognitiva y la musicología en general [39].

Se presenta la descripción de un algoritmo de extracción de melodías, la recogida de datos e indicadores utilizados para la evaluación es por medio del análisis espectral aplicando un filtro de sonoridad a la señal tratando de mejorar la percepción humana. Posteriormente se aplica la Transformada de Fourier de Tiempo Reducido utilizando ventanas de Hanning de 46 milisegundos. Se obtienen los picos del espectro y el espectro de fase para calcular la frecuencia instantánea como se propone Dreesler en [19], que proporciona una estimación refinada de la frecuencia de pico.

La generación de la función de relevancia se calcula por medio de los picos cepstrales restantes; una vez que se ha creado, cada segmento se refina volviendo a calcular la suma de sus picos armónicos. Para determinar si la melodía está presente o no, se maneja mediante un filtrado de sonorización.

Después de varias iteraciones se produce una media de la cual se retiran ciertos errores, enmarcando el pico perteneciente al segmento más sobresaliente, obteniendo la selección de la melodía según Salomón en [73].

Se presenta una versión actualizada con dos variantes en este sistema, optimizando la extracción de características de una melodía. En el primer bloque del sistema se analiza la señal de audio y extraer picos espectrales (sinusoides) que se usarán para construir la función de relevancia en el bloque siguiente. Este proceso se compone de tres pasos principales: pre-filtrado, transformada y la frecuencia/corrección de la amplitud [74].

El nuevo método se basa en la creación y caracterización de frecuencias de contorno, para distinguir entre lo melódico y no melódico. Esto conduce al desarrollo de una nueva detección de sonoridad, minimizar errores y una nueva técnica para selección de melodías [75].

2.2. Resumen

En este capítulo se revisaron los artículos más importantes en el campo de la recuperación de información musical. A través de este capítulo se puede apreciar el desarrollo del área de la recuperación de información musical y sus avances hasta nuestros días. También, se hace una importante recopilación de los artículos pioneros en el área escritos por Serra, McNab, Logan, Hoashi y Salomón, entre otros. Todos los artículos presentados en este capítulo están organizados por la clasificación que tiene la recuperación de información musical.

Capítulo 3

Marco teórico

La música es una forma de expresión abstracta, ya que mediante símbolos se identifican sonidos que corresponden y tienen un significado en específico; por otro lado, ésta puede ser interpretada de diferentes formas, ya que cada persona tiene un criterio propio. Comúnmente se dice que la música es como el arte de los sonidos con un propósito expresivo. Según esta definición, la música está caracterizada por tres sentidos básicos: arte, ciencia y lenguaje.

Puede verse como un lenguaje, ya que transmite mensajes en base a un código propio y particular que se conoce como Lenguaje Musical.

3.1. Melodía

La melodía puede definirse como una sucesión entre diferentes sonidos de altura, intensidad y duración variable, formando frases con sentido expresivo. Es uno de los elementos que forman el lenguaje musical.

Algunas definiciones académicas referentes a la melodía son:

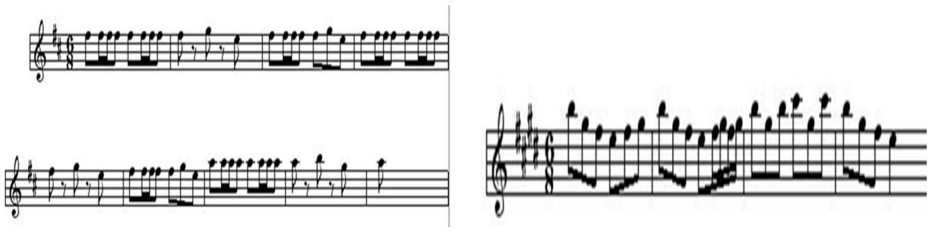
- “La melodía es una sucesión de sonidos de diferente altura y duración” [79].
- “Sucesión de notas o intervalos que comunican o expresan una idea estética o musical”.
- “Sucesión coherente de notas” diccionario de música de Harvard.

3.1.1. Características de la melodía

Después de conocer lo que es una melodía, es conveniente saber los conceptos musicales que se relacionan con ella [10].

Intervalo Es la distancia que existe entre los sonidos o diferencia de altura entre dos notas musicales. Dependiendo de sus características, pueden ser: ascendentes o descendentes, armónico o melódico, según el número y especie, conjunto o disjunto, y simples o compuestos.

Línea melódica Es la línea o trayectoria sonora que constituye el tema de la obra musical, la cual se puede encontrar repetida en otra frase de la melodía o con alguna pequeña variación. Se tienen en cuenta las altitudes y el movimiento lineal, pero no su ritmo. Puede ser: recta u ondulada. En las Fig. 3.1(a) y 3.1(b) se puede observar la representación de ellas.



(a) (b)

Figura 3.1: a) Línea melódica recta y (b)Línea melódica ondulada

Frase melódica Es un fragmento, que dentro de una melodía tiene un sentido determinado porque dan origen a la formación de secciones y subsecciones cada vez de menor categoría. Está formada por varias secciones y partes, tales como:

- **Semifrases**: son las divisiones más importantes en que se puede dividir el fragmento.
- **Periodos**: son las divisiones más importantes en que se puede dividir la semifrase.
- **Subperiodos**: son las divisiones más importantes del periodo.

Las frases son normales suelen tener ocho compases aunque dependerá del tipo de música del que se trate y del compás que se emplee. No obstante, la frase de ocho compases es la frase más perfecta por excelencia y la más usada y didáctica.

Tonalidad Conjunto de sonidos ordenados mediante relaciones mutuas, estando estas determinadas por un sonido básico llamado tónica (sonido básico y organización de los sonidos de una escala, siendo también la nota que da nombre a dicha escala. Esa nota es el eje de la construcción armónica y melódica. Es la nota de reposo de la escala).

Modalidad Cuando se habla de modalidad se dice que cada tonalidad puede presentarse de dos formas diferentes: modo mayor y modo menor. Es la ordenación de tonos y semitonos. Una tonalidad es mayor cuando la tríada de su tónica es mayor y es menor cuando dicha tríada es menor.

Transporte de canciones El transporte en música se refiere al cambio de una melodía de un tono a otro. Antes de realizarlo, se debe comprobar la tonalidad de origen y pensar qué transporte es el que se desea hacer. Tiene como finalidad el ofrecer la posibilidad de que una pieza musical pueda ser interpretada por voces o instrumentos para los que sería imposible o muy difícil. Éste puede ser leído o escrito.

3.2. Sonido digital

Se dice que el sonido es un fenómeno físico, generado por oscilaciones o vibraciones de las partículas que componen un medio y que son generalmente producidas por la presión acústica ejercida por una fuente sonora, este fenómeno se puede dar en diferentes medios; como lo son los gases, los sólidos y los líquidos. El modo más conocido de propagación del sonido, sin dudas, lo es el de los gases, ya que el aire es el medio gaseoso donde los humanos se encuentran inmersos desde su nacimiento.

Sin embargo, el audio es la tensión eléctrica o magnética proporcional a un sonido, se genera por medio de elementos transductores como lo son los micrófonos y suele estar acompañado de varios pasos y procesos para su tratamiento, almacenamiento y reproducción.

El audio digital, se refiere a la conversión de un sonido a datos digitales (valores de 0 y 1), los cuales pueden ser tratados de forma parecida a la que se trata el audio real, pero siempre se debe tener en cuenta que no se trata de audio en sí, sino datos, iguales a los datos almacenados en documentos de texto común. Resumiendo, el sonido es lo que llega al oído humano y el audio es cuando un sonido ha sido convertido en tensiones eléctricas para ser procesado o almacenado [55].

3.2.1. Audio digital

Es la representación de señales sonoras mediante un conjunto de datos binarios, entonces se puede decir que el sonido digital es una adaptación del sonido real que se realiza a través de un medio electrónico: la tarjeta de sonido. La calidad de sonido está determinada por las características técnicas que tenga dicha tarjeta, como:

El muestreo es el proceso mediante el cual se mide la frecuencia del sonido tomando muestras en intervalos de tiempos regulares. Es el proceso básico en la transformación del sonido analógico en sonido digital.

La frecuencia de muestreo; es la cantidad de muestras tomadas de una onda, como se visualiza en la Fig. 3.2. A mayor cantidad de frecuencia de muestreo el sonido digitalizado será más parecido al original. Cuanta más alta sea ésta

la captura del sonido será más precisa y, en consecuencia, el sonido digital será de mayor calidad.

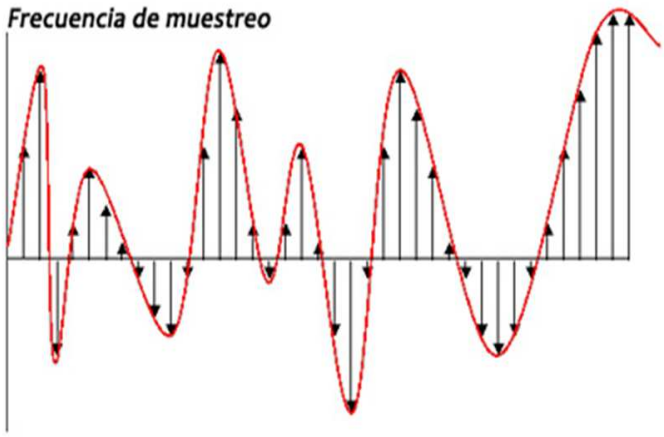


Figura 3.2: Frecuencia de muestreo

El teorema del Muestreo de Nyquist-Shannon dice que se puede reproducir de manera exacta una onda si la frecuencia de muestreo es, como mínimo, el doble de la más alta que se pueda escuchar. Para el oído humano esta frecuencia corresponde a 20,000 Hz, por lo tanto la frecuencia de muestreo más adecuada será de 40,000 Hz. Algunos estudios aumentan esta cifra hasta los 44,100 Hz, que es la que se suele usar.

La resolución del sonido está directamente relacionada con la frecuencia de muestreo. Se refiere al número de dígitos binarios, 1 y 0, que componen cada muestra, la Fig. 3.3 muestra un ejemplo de este tipo. Su unidad de medida es el bit y hace referencia al tamaño de cada una de esas muestras. Lo habitual es trabajar con 16 bits aunque se puede hacer también con 8 o con 32. Si la resolución de un audio es de 8 bits significa que hemos tomado 256 valores para la muestra. A más resolución y más frecuencia, mayor será la calidad del sonido.

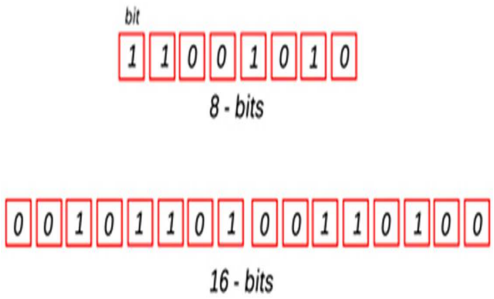


Figura 3.3: Resolución del sonido

El principal problema del audio en una biblioteca digital es el tamaño de los archivos, a mayor calidad mayor espacio, e incluso con características pobres de calidad los archivos son grandes de cualquier manera. Asimismo, los recursos internos del equipo que reproduce el sonido se ven sometidos a un pesado procesamiento. Por ejemplo, el tamaño de una grabación de un minuto de 8-bit monofónico, a diferentes frecuencias de muestreo, se muestra en la tabla 3.1, y el tamaño aproximado de un minuto de grabación a 16-bit estereofónico, se muestra en la tabla 3.2.

Tabla 3.1: Grabación de audio a 8 bits

Frecuencias de muestreo (KHz)	Tamaño del archivo (Mb)
22,050	2.52
44,100	5.04
48,000	5.49

Tabla 3.2: Grabación de audio a 16 bits

Frecuencias de muestreo (KHz)	Tamaño del archivo (Mb)
22,050	4.04
44,100	10.08
48,000	10.98

3.2.2. Formatos de sonido

Para manejar el sonido digital con facilidad se desarrollaron algunas formas para almacenar archivos de manera que fuesen lo más pequeños posible, sin perder demasiada calidad. Estas formas reciben el nombre de formatos, cada formato tiene asociada una extensión que sirve para nombrarlo e identificarlo.

Existen muchos formatos de sonido, cada uno desarrollado por un fabricante, y, desde luego, no todos los programas son capaces de "leer" todos los formatos, de aquí la utilidad de los editores de sonido que, además de grabar y reproducir sonido, pueden servir para cambiar un tipo de formato en otro. Dentro de estos formatos se puede establecer una clasificación general:

- **Formatos sin compresión**: Son los que almacenan el sonido tal cual se graba sin realizar ningún tipo de modificación. Desde el punto de vista de calidad de sonido son los mejores, pero tienen un gran inconveniente producen archivos de enorme tamaño entre 2,6 y 10,4 Mb (mega bites) por minuto.
- **Formatos con compresión**: Son los que almacenan el sonido de forma comprimida, realizando una transformación que hace que el archivo sea de menor tamaño. Todos los formatos comprimidos producen una pérdida de calidad con respecto al sonido original, pérdida que será mayor cuanto mayor sea el porcentaje de compresión que utilicemos. En la Tabla 3.3 se muestra una comparativa de los principales formatos de sonido.

Formatos WAV

El formato WAV, (Waveform Audio File) es un formato de archivo originario de Microsoft Windows 3.1; su extensión es WAV. Es el formato para almacenar sonidos más utilizado por los usuarios de Windows, lo flexible de

Tabla 3.3: Grabación de audio a 16 bits

Formato	Desarrollado	Calidad	Tamaño/minuto
MIDI	Dave Smith	Sonido digital puro	21 Mb
WAV	Microsoft	Muy buena	5.3 Mb
MP3	Moving Picture Expert Group	Buena (depende del archivo wav original)	440 Kb
CDA		Excelente	5.3 Mb

este formato lo hace muy usado para el tratamiento del sonido pues puede ser comprimido y grabado en distintas calidades y tamaños de muestreo.

Aunque los archivos WAV pueden tener un excelente sonido comparable al del CD (16 bytes y 44,100 Hz estéreo) el tamaño necesario para esa calidad es demasiado grande (una canción convertida a WAV puede ocupar fácilmente entre 20 y 30 Mb).

Los archivos WAV pueden guardarse con varios tipos de codificación. Pese a que el archivo de sonido sigue siendo muy grande, si lo comparamos con otros tipos de compresión, la realidad es que nos proporciona la suficiente reducción de tamaño para que resulten más manejables, sobre todo cuando se trata de sonidos de corta duración. Si se desea grabar sonidos de larga duración es irrelevante aplicar estas compresiones puesto que los archivos seguirán siendo muy grandes; emplear en su lugar, si las características de la aplicación lo permiten, formatos con compresión.

El principal problema que se puede encontrar con los archivos WAV grabados con condiciones mínimas es la baja calidad del sonido, los ruidos e incluso cortes en el sonido, por esta razón casi siempre hay que usar muestras de sonido.

3.3. Redes neuronales

Si se tuviera que definir la principal característica que separa al ser humano del resto de los animales seguramente, la gran mayoría de los humanos, responderían que es la capacidad de raciocinio. Esta capacidad ha permitido desarrollar una tecnología propia de tal manera que, en estos momentos, esa tecnología se orienta a descubrir su origen. ¿Cómo funciona el cerebro?, ¿se pueden construir modelos artificiales que lo emulen?, ¿se pueden desarrollar máquinas inteligentes? Todas estas preguntas han conducido a un desarrollo exponencial de un campo multidisciplinario del conocimiento conocido como Inteligencia Artificial (IA). Este campo se podría dividir en dos clases que se pueden definir como “macroscópico” y “microscópico”.

En el primero de ellos se intenta modelizar el funcionamiento del cerebro basándose en reglas del tipo “si ocurre esto entonces…”, el nombre de macroscópico se debe a que no se toma en cuenta en ningún momento la estructura interna del cerebro sino que modeliza su comportamiento en base a un funcionamiento que podríamos definir como global.

En la segunda aproximación se parte de la estructura que presenta el cerebro de tal forma que se construyen modelos que tienen en cuenta dicha estructura. De esta forma aparecen “neuronas artificiales” que se combinan entre sí para formar “estructuras multicapas” que, a su vez, pueden combinarse para formar “comités de expertos”, etc. Esta forma de combinación recuerda la estructura en niveles del cerebro. Esta aproximación de la IA conocida como redes neuronales ha sufrido, en los últimos años, un incremento espectacular en publicaciones, aplicaciones comerciales, número de congresos celebrados, etc.

3.3.1. ¿Qué son las redes neuronales?

No existe una definición general de red neuronal artificial, existiendo diferentes según el texto o artículo consultado. Aquí algunas definiciones encontradas:

- Sistema caracterizado por una red adaptativa combinada con técnicas de procesado paralelo de la información [36].
- Sistema de procesado de la información que tiene características de funcionamiento comunes con las redes neuronales biológicas [20].
- Una red neuronal es un modelo computacional, paralelo, compuesto de unidades procesadoras adaptativas con una alta interconexión entre ellas [25].
- Desde la perspectiva del reconocimiento de patrones las redes neuronales son una extensión de métodos clásicos estadísticos [6].
- Sistemas de procesado de la información que hacen uso de algunos de los principios que organizan la estructura del cerebro humano [47].
- Modelos matemáticos desarrollados para emular el cerebro humano [15]].

Las definiciones expuestas son una muestra, ya que cada autor las define a su manera. Parece ser que en todas ellas aparece el componente de simulación del comportamiento biológico. Lo que sí tienen en común estos elementos con el cerebro humano es la distribución de las operaciones a realizar en una serie de elementos básicos que, por analogía con los sistemas biológicos, se conocen como neuronas. Estos elementos están interconectados entre sí mediante una serie de conexiones que, siguiendo con la analogía biológica, se conocen como pesos sinápticos. Estos pesos varían con el tiempo mediante un proceso que se conoce como aprendizaje. Así se puede definir el aprendizaje de una red como el proceso por el cual modifica las conexiones entre neuronas, pesos sinápticos, para realizar la tarea deseada.

3.3.2. Revisión histórica

Cuando se narra la corta pero intensa historia de las redes neuronales también conocidas como modelos conexionistas se suele fijar el origen en los trabajos de McCulloch y Pitts. Sin embargo, existen trabajos anteriores

que abrieron el camino a estos investigadores. Entre esos trabajos puede destacar el realizado por Karl Lashley en los años 20. En su trabajo de 1950 se resumen su investigación de 30 años; destaca que el proceso de aprendizaje es un proceso distribuido y no local a una determinada área del cerebro [43]. Un estudiante de Laksley, D. Hebb estudia lo realizado por su maestro, y determina una de las reglas de aprendizaje más usadas en la regla del conexionismo y que, lógicamente, se conoce con el nombre de aprendizaje hebbiano. Las contribuciones de este investigador aparecen en [28], en el capítulo 4 se da, por primera vez, una regla para la modificación de las sinapsis, es decir, una regla de aprendizaje fisiológica. Además propone que la conectividad del cerebro cambia continuamente conforme un organismo aprende cosas nuevas, creándose asociaciones neuronales con esos cambios. En su postulado de aprendizaje, Hebb sigue lo sugerido por Ramón y Cajal al afirmar que la efectividad de una sinapsis variable entre dos neuronas se incrementa por una repetida activación de una neurona sobre otra a través de esta sinapsis. Desde un punto de vista neurofisiológico la regla planteada por Hebb sería una regla variante-temporal, con un alto mecanismo interactivo que incrementa la eficacia sináptica como una función de la actividad pre y post sináptica. Desde un punto de vista conexionista la regla de Hebb es un tipo de aprendizaje no supervisado (no se necesita ningún “maestro”) en el que las conexiones entre dos neuronas se incrementan si ambas se activan al mismo tiempo.

La siguiente gran contribución a considerar es el trabajo de McCulloch y Pitts. Este tipo de neurona es un dispositivo binario (salida 0 ó 1), tiene un umbral de funcionamiento por debajo del cual está inactiva y puede recibir entradas excitadoras o inhibitorias cuya acción es absoluta: si existe alguna de estas entradas la neurona permanece inactiva. El modo de trabajo es simple, si no existe ninguna entrada inhibitoria se determina la resultante de las entradas excitadoras y si ésta es mayor que el umbral, la salida es 1 y si no, la salida es 0 [51].

Se puede comprobar que este modelo puede sintetizar algunas de las funciones lógicas. En la Fig. 3.4se puede observar un ejemplo de este modelo.

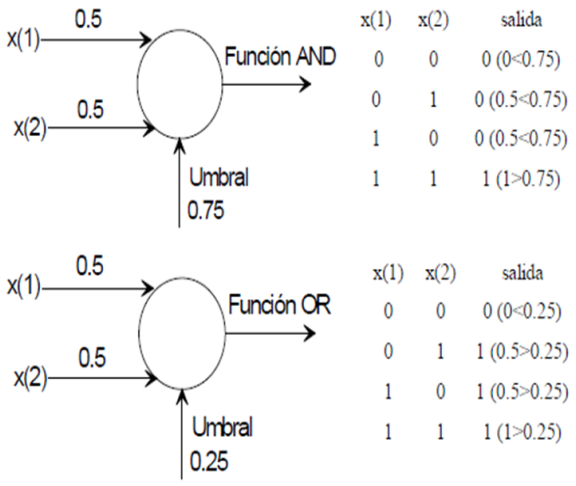


Figura 3.4: Ejemplo con McCulloch Pitts

Se puede observar que, con un elemento tan simple como el que se acaba de definir, se pueden implementar un gran número de funciones lógicas mediante su combinación con elementos similares. Además, dado el estado de la neurofisiología en 1943, el modelo McCulloch-Pitts se acercaba a lo conocido por esa época acerca de la actividad sináptica neuronal. Esta capacidad de modelizar funciones lógicas desató la euforia por estos elementos individuales: si se pueden modelizar funciones lógicas, ¿por qué no implementar un sistema de conocimiento mediante el uso de

estas neuronas?.

En 1956, Rochester, Holland, Haibit y Duda presentan un trabajo en el que, por primera vez, se verifica mediante simulaciones una teoría neuronal basada en el postulado de Hebb. Para realizar este trabajo especialmente práctico, se tuvieron que hacer varias suposiciones que, inicialmente no estaban en el trabajo de Hebb. Por ejemplo se señaló el valor de las sinapsis que, en principio, podía crecer sin límite.

Otro gran genio matemático, John Von Neumann, se planteó ideas conexionistas, su trabajo fue reconocido después de su muerte. En el sugiere la posibilidad de mejorar las computadoras para el estudio del sistema nervioso central. Debido a esto John Von Neumann se puede considerar como uno de los padres de la computación.

En 1958 se producen las aportaciones de Selfridge y Rosenblatt. Éstas plantean implementaciones físicas de sistemas conexionistas. Selfridge propone el sistema conocido como Pandemonium [77]. Este sistema consta de una serie de capas compuestas por lo que se conocen como “demonios”. Cada una de las diferentes capas de este sistema se reparten las diferentes tareas a realizar.

Por su parte, Rosenblatt, quince años después del estudio de McCulloch-Pitts, presenta una nueva aproximación al problema de reconocimiento de patrones mediante la introducción del perceptrón. Él planteó un dispositivo que realizara tareas que le interesaran a los psicólogos (como él). El hecho que fuera una máquina capaz de aprender la hacía irresistiblemente atractiva para los ingenieros [70].

En 1960 Widrow y Holff presentan su ADALINE. Este sistema estaba regido por un algoritmo de aprendizaje muy sencillo denominado LMS (Least Men Square). Con este trabajo se propone un sistema adaptativo que puede aprender de forma más precisa y rápida que los perceptrones existentes [92]. En su trabajo posibilitó el desarrollo de un área del procesado digital de señales (control de sistemas) que se conoce con el nombre de procesado (control) adaptativo [27].

Block presenta en 1962 un trabajo que estudia los perceptrones más concretamente, presenta resultados sobre el perceptrón “MARK I” con 400 dispositivos receptores fotosensitivos dispuestos en una matriz 20 por 20 con un conjunto de 8 unidades de salida [8]. Para el año de 1970 el profesor Marvin Minsky junto con el matemático Seymour Papert, formulan el trabajo titulado Perceptrons que paraliza durante 10 años el avance de este campo de la inteligencia artificial. Este trabajo, puso de manifiesto las limitaciones de los perceptrones. Estas limitaciones hacían referencia a la clase de problemas que se podían resolver usando estos elementos. Demostraron que un perceptrón solo podía resolver problemas linealmente separables que son los menos ocurrentes. Además los autores expusieron sus opiniones sobre las extensiones de los perceptrones (a sistemas multicapa); ellos plantearon su absoluta inutilidad práctica [54].

Kohonen y Anderson proponen el mismo modelo de memoria asociativa de forma simultánea. A modo de demostración de los diferentes campos de conocimiento que engloban los sistemas conexionistas estos autores tienen una formación diferente (Kohonen es ingeniero eléctrico y Anderson es neurofisiólogo). En el modelo artificial planteado, la neurona es un sistema lineal que se usa como regla de aprendizaje la regla de Hebb modificada. Esta premisa los coloca frente a un asociador lineal ([3], [37] y [38]).

En 1980, Stephen Grossberg, uno de los autores más fructíferos en el campo de las redes neuronales, establece un nuevo principio de auto-organización desarrollando las redes neuronales conocidas como ART (Adaptive Resonance

Theory). Grossberg ha planteado diferentes modelos neuronales que han presentado una gran utilidad (principalmente en el campo del reconocimiento de patrones) [23]. En 1982 J. Hopfield publica un trabajo clave para el resurgimiento de las redes neuronales, gran parte del impacto de este trabajo se debió a la fama de Hopfield como distinguido físico teórico. En él, desarrolla la idea del uso de una función de energía para comprender la dinámica de una red neuronal recurrente con uniones sinápticas simétricas [31]. En este primer trabajo, Hopfield sólo permite salidas bipolares (0 ó 1, ± 1). En un trabajo posterior amplía la función de energí¬a planteada para estos sistemas permitiendo la salida continua de las neuronas [32]. El principal uso de estas redes ha sido para memorias y como instrumento para resolver problemas de optimización [1].

En el mismo año Kohonen publica un importante artículo sobre mapas auto-organizativos que se ordenan de acuerdo a unas simples reglas. El aprendizaje que se da en el modelo planteado no necesita de un “maestro”; esto es un tipo de aprendizaje no supervisado [Kohonen 1982]. Al año siguiente, en el número especial sobre modelos neuronales de la revista especializada IEEE Transactions on Systems, Man and Cybernetics, aparecen dos trabajos de gran importancia en el desarrollo de las redes neuronales. Fukushima, Miyake e Ito presentan una red neuronal, el Neocognitron, de tal forma que combinando ideas del campo de la fisiología, ingeniería y de la teoría neuronal crean un dispositivo que es capaz de ser aplicado con éxito en problemas de reconocimiento de patrones. Este trabajo supone un perfeccionamiento de un modelo anterior presentado por los mismos autores y conocido como Cognitron.

El segundo trabajo, presentado por Barto, Sutton y Anderson estudia el aprendizaje reforzado y su aplicación en control. En este trabajo se plantea este nuevo tipo de aprendizaje en el que, a diferencia de trabajos anteriores sobre modelos supervisados, no es necesario un conocimiento total del error cometido por la red; lo único que se necesita es conocer el signo del error [4].

En 1986 aparece un trabajo que, junto al de Hopfield, resucitará el interés por las redes neuronales. En este trabajo Rumelhart, Hinton y Williams, desarrollan el algoritmo de aprendizaje de retropropagación (back-propagation) para redes neuronales multicapa dando una serie de ejemplos en los que se muestra la potencia del método desarrollado [72]. A partir de ese año, el número de trabajos sobre redes neuronales ha aumentado exponencialmente apareciendo un gran número de aportes, tanto a los métodos de aprendizaje como a las arquitecturas y aplicaciones de las redes neuronales. Se podrí¬a destacar de entre todos estos aportes el trabajo de Broomhead y Lowe y el de Poggio y Girosi sobre el diseño de redes neuronales en capas usando RBF (Radial Basis Functions)([12] y [60]), el trabajo intensivo desarrollado sobre las máquinas de vectores de soporte [87], el desarrollo de la unión entre elementos neuronales y difusos y, por último, los trabajos sobre neuronas de pulsos (spike neurons).

Finalmente se encuentran uno de los motores en el desarrollo de las redes neuronales: la predicción en series temporales. Una generalización de las redes TDNN (orientadas especialmente para ser usadas con series temporales) la realizó Eric Wan [90]. En su trabajo los pesos sinápticos, conexiones sinápticas, eran filtros digitales.

3.3.3. Modelos neuronales

En todo modelo artificial de neurona se tienen cuatro elementos básicos:

1. Un conjunto de conexiones, pesos o sinapsis que determinan el comportamiento de la neurona. Estas conexiones pueden ser excitadoras (presentan un signo positivo), o inhibidoras (conexiones negativas).

2. Un sumador que se encarga de sumar todas las entradas multiplicadas por las respectivas sinapsis.

3. Una función de activación no lineal para limitar la amplitud de la salida de la neurona.

4. Un umbral exterior que determina el umbral por encima del cual la neurona se activa.

Esquemáticamente, una neurona artificial quedaría representada por la Fig. 3.5.

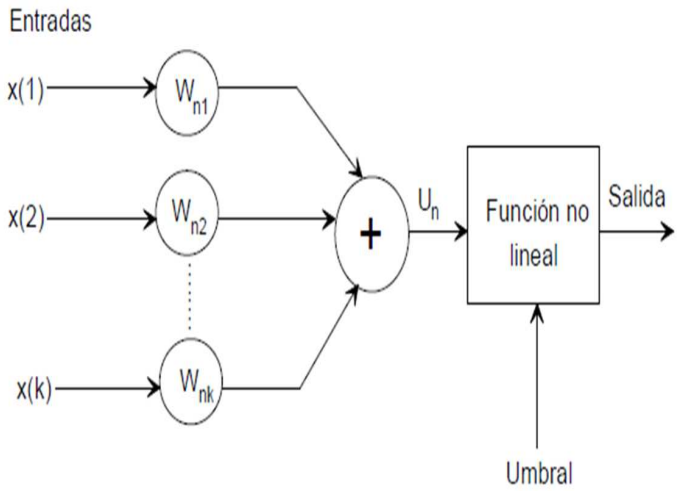


Figura 3.5: Esquema de un modelo neuronal

Matemáticamente las operaciones a realizar serían:

$$U_n = \sum_{j=1}^{k} W_{nj} \cdot x\,(j) \tag{3.1}$$

y

$$\text{salida} = \rho\,[U_n - umbral] \tag{3.2}$$

Donde ρ es una función no lineal conocida como función de activación. Normalmente se asocia el umbral a la salida U_n mediante una entrada (que vale -1) y un peso adicional asociado. Es decir:

$$umbral = -W_{n0} \Rightarrow U_n = \sum_{\substack{j=0 \\ x(0)=1}}^{k} W_{nj} \cdot x(j) \Rightarrow \text{salida} = \rho(U_n) \tag{3.3}$$

El modelo descrito es el más usual, sin embargo, aparecen otros modelos que no realizan un promedio de las entradas directamente sino que, antes de multiplicar por los pesos se plantea una transformación de dichas entradas. Así, se tiene:

- Transformación cuadrática:

$$U_n = \sum_{j=1}^{k} W_{nj} \cdot x^2(j) \tag{3.4}$$

- Transformación polinómica:

$$U_n = \sum_{j=1}^{k} \sum_{s=1}^{k} W_{njs} \cdot x(j) \cdot x(s) \tag{3.5}$$

- Transformación esférica:

$$U_n = \frac{1}{\rho^2} \sum_{j=1}^{k} (X(j) - W_{nj})^2 \tag{3.6}$$

El modelo planteado es el más común pero hay que destacar que es estático; uno más general consideraría salidas anteriores: tendríamos un modelo dinámico [91]. La neurona definida de esta forma tendría memoria. Matemáticamente este hecho se expresaría como:

$$\text{salida}_n = F(\text{salidas}_{n-k}, \text{entradas}) \quad k = 1, \ldots, n-1 \tag{3.7}$$

Es decir, la salida en el instante n depende no sólo de las entradas como en el caso anterior sino que ahora aparece una dependencia con las salidas anteriores.

En cuanto a las funciones de activación existe un gran número inspiradas, todas ellas, en modelos biológicos. Algunas de estas funciones son:

- Función signo o umbral:

$$\text{salida} = \begin{cases} 1 & U_n \geq 0 \\ 0 & U_n < 0 \end{cases} \tag{3.8}$$

  Cuando una neurona usa esa función de activación se habla del modelo de McCulloch-Pitts.

- Sigmoide:

$$\text{salida} = \frac{1}{1 + e^{(-a \cdot U_n)}} \tag{3.9}$$

  donde a fija la pendiente de la función en el origen. Aumentando esta constante, la sigmoide se asemeja a la función signo. Las funciones definidas varían entre 0 y 1; se pueden definir a partir de ellas otras funciones que varían entre -1 y 1 simplemente escalando las salidas entre éstos límites.

- Función lineal a tramos:

$$\text{salida} = \begin{cases} 1 & U_n \geq \frac{1}{2} \\ U_n + \frac{1}{2} & -\frac{1}{2} > U_n > \frac{1}{2} \\ 0 & U_n = -\frac{1}{2} \end{cases} \tag{3.10}$$

- Función Gaussiana:

$$\text{salida} = K_i \cdot e^{\left(\frac{U_n - K_2}{K_3}\right)^2} \tag{3.11}$$

  siendo K_i constante

3.3.4. Arquitecturas neuronales

Los elementos básicos comentados anteriormente se pueden conectar entre sí para dar lugar a las estructuras neuronales o modelos conexionistas que se pueden clasificar de diferentes formas según el criterio usado.

Según el número de capas

- **Red neuronal monocapa**: Corresponde a la red neuronal más sencilla ya que se tiene una capa de neuronas que proyectan las entradas a una capa de neuronas de salida donde se realizan diferentes cálculos, como se muestra en la Fig.3.6. La capa de entrada, por no realizar ningún cálculo, no se cuenta de ahí el nombre de redes neuronales con una sola capa.

- **Red neuronal multicapa**: Es una generalización de la anterior existiendo un conjunto de capas intermedias entra la entrada y la salida (capas ocultas), en la Fig. 3.7 se observa su esquema. Este tipo de red puede ser total o parcialmente conectada.

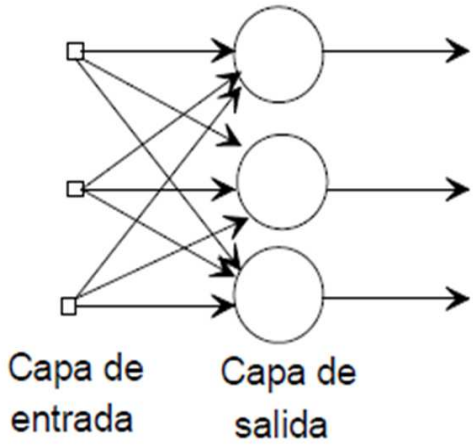


Figura 3.6: Red neuronal monocapa

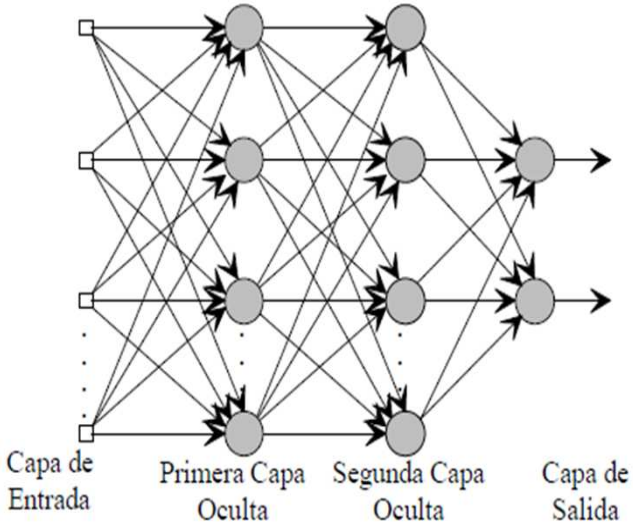


Figura 3.7: Red neuronal multicapa

Según el tipo de conexiones

- **Red neuronal no recurrente**: En esta red la propagación de las señales se produce en un sentido solamente, no existiendo la posibilidad de realimentaciones. Lógicamente estas estructuras no tienen memoria.
- **Red neuronal recurrente**: Esta red está caracterizada por la existencia de lazos de retroalimentación. Estos lazos pueden ser entre neuronas de diferentes capas, neuronas de la misma capa o, más sencillamente, entre una misma neurona. Esta estructura recurrente la hace especialmente adecuada para estudiar la dinámica de sistemas no lineales. En la Fig. 3.8 se muestra el esquema de una red recurrente.

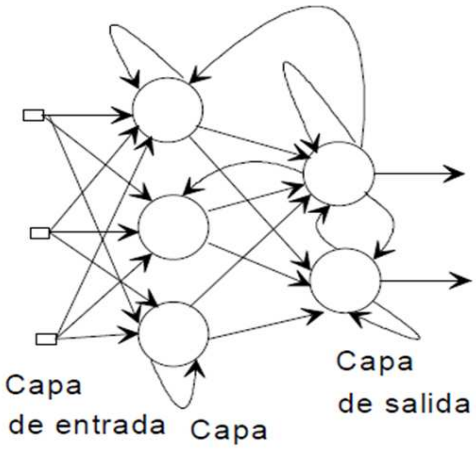


Figura 3.8: Red neuronal recurrente

Según el grado de conexión

- **Red neuronal totalmente conectada**: En este caso las neuronas de una capa se encuentran conectadas con las de la capa siguiente (redes no recurrentes) o con las de la anterior (redes recurrentes).

- **Red parcialmente conectada**: En este caso no se da la conexión total entre neuronas de diferentes capas.

3.3.5. Métodos de aprendizaje

En una red neuronal es necesario definir un procedimiento por el cual las conexiones del dispositivo varían para proporcionar la salida deseada (algoritmo de aprendizaje). Los métodos de aprendizaje se pueden dividir en las siguientes categorías [69], como se puede observar en la Fig. 3.9.

La primera división en los métodos de aprendizaje es entre algoritmos supervisados y no supervisados. En los algoritmos no supervisados no se conoce la señal que debe dar la red neuronal (señal deseada). La red en este caso se organiza ella misma agrupando, según sus características, las diferentes señales de entrada. Estos sistemas proporcionan un método de clasificación de las diferentes entradas mediante técnicas de agrupamiento o clustering.

El aprendizaje supervisado presenta a la red las salidas que debe proporcionar ante las señales que se le presentan. Se observa la salida de la red y se determina la diferencia entre ésta y la señal deseada. Posteriormente, los pesos de la red son modificados de acuerdo con el error cometido. Este aprendizaje admite dos variantes: aprendizaje por refuerzo o por corrección. En el aprendizaje por refuerzo sólo conocemos si la salida de la red se corresponde o no con la señal deseada, es decir, nuestra información es de tipo booleana (verdadero o falso). En el aprendizaje por corrección se conoce la magnitud del error y ésta determina la magnitud en el cambio de los pesos.

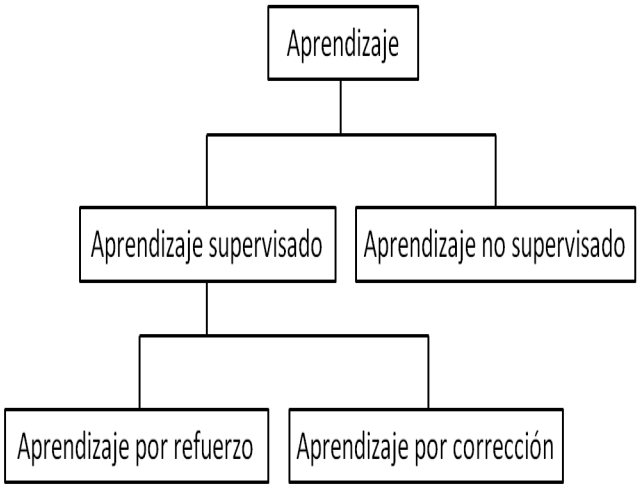


Figura 3.9: Métodos de aprendizaje

3.3.6. Estructuras neuronales

Las redes neuronales se pueden usar en una serie de estructuras según la aplicación a la que está destinado el sistema.

Estructura directa

En la Fig. 3.10 se muestra el esquema de bloques de esta estructura. Se puede apreciar el sistema, al principio desconocido y la red neuronal tienen las mismas entradas por lo que se conseguirá el mínimo error (objetivo de la red neuronal) cuando la salida de una red neuronal y la señal deseada sean iguales, o lo que es lo mismo, cuando la función de transferencia de la red neuronal sea igual a la del sistema desconocido. Así pues esta estructura tiene como finalidad la modelización de funciones de transferencia de sistemas de los que, en principio, no se conoce nada pero se tiene la posibilidad de excitarlos con una determinada entrada y así conocer su salida.

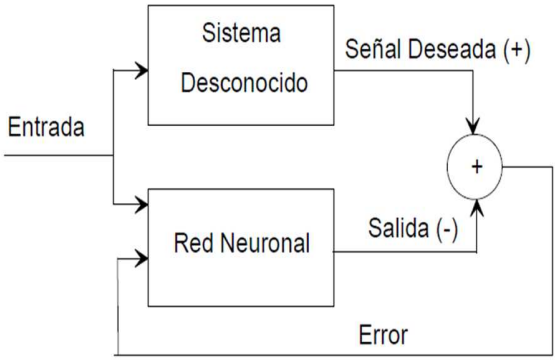


Figura 3.10: Esquema de bloques de la estructura directa

Estructura inversa

En la Fig. 3.11 se muestra el esquema de bloques de esta estructura. El mínimo error en esta estructura se obtendrá cuando la salida de la red neuronal sea la entrada al sistema desconocido lo que conlleva que la función de transferencia de la red neuronal sea la inversa del sistema desconocido. Hay que destacar que el perfecto funcionamiento de esta estructura depende de la estabilidad de la inversa de la función de transferencia del sistema desconocido.

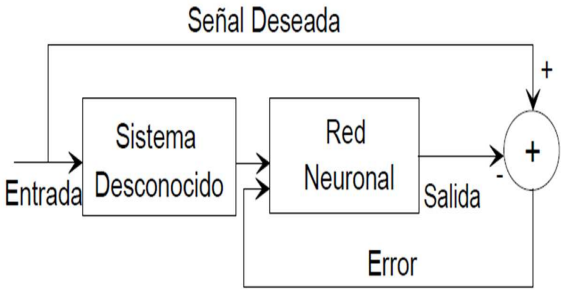


Figura 3.11: Esquema de bloques de la estructura inversa

Estructura con retardo

En la Fig. 3.12 se muestra el esquema de bloques de esta estructura. Según la definición de red neuronal, esta estructura tiende a minimizar la diferencia entre la señal deseada (señal de entrada en el instante) y la salida de la red neuronal que será un determinado valor obtenido con valores de la señal. Se intenta, pues, modelizar la señal actual a partir de los valores anteriores de ésta. Este sistema se puede usar, pues, en problemas de predicción (a partir de las muestras pasadas se puede estimar la siguiente) y de control (si se conoce la evolución del sistema se puede alterar los parámetros de dicho sistema para cambiar dicha evolución).

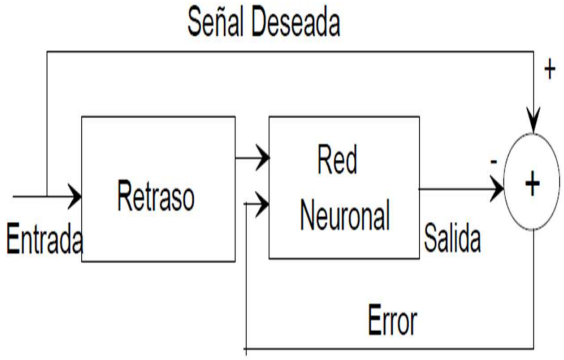


Figura 3.12: Esquema de bloques de la estructura con retardo

3.4. Redes de retardo temporal (TDNN)

Cuando se usan redes multicapa para el tratamiento de secuencias se suele aplicar una idea muy simple que consiste en que la entrada de la red se componga de no sólo el valor de la secuencia en un determinado instante, sino en instantes anteriores. Es como alimentar la red con una ventana temporal.

Esta idea de introducir el estado de una variable en diversos instantes en la red no sólo se puede aplicar a la entrada sino también a las activaciones de las neuronas. Una red donde las activaciones de algunas neuronas son simplemente una copia de las activaciones de otras en instantes anteriores se la denomina Red Neuronal de Retardo Temporal o Time-Delay Neural Network (TDNN) ([41], [42], [88] y [89]).

Las neuronas con las que se trabajan en redes multicapa con retrasos temporales responden a la ecuación:

$$x_i = f_i\left(\sum_{j=1} w_{ij} \cdot x_j\right) \tag{3.12}$$

Como se observa no existe una dependencia temporal, y la propagación o cálculo de las activaciones se realiza desde la capa superior a la inferior como en cualquier red multicapa. En estas redes un paso de tiempo hay que entenderlo como iteración. La conexión entre la neurona j y la i, introduciendo retrasos temporales, se realizará como:

$$\begin{aligned} x_i &= f_i\left(\textstyle\sum_1 w_{i1} \cdot x_1\right) \\ x_1 &= x_j\left(t - \tau_1\right) \end{aligned} \tag{3.13}$$

donde t significa interacción y τ_1 es el retraso temporal. Las neuronas x_1 son simplemente copias de la activación de x_j en instantes o iteraciones anteriores. Se puede dar otra interpretación que consiste en asignar a los pesos distintas velocidades de conexión, siendo unos más lentos que otros, con lo cual en vez de tener una capa con varias neuronas conteniendo copias de las activaciones de la j tendríamos sólo la neurona j pero conectada a la i con varios pesos de distinta velocidad. La ecuación 3.14 se transformaría en:

$$x_i = f_i\left(\sum_j \sum_k w_{ijk} \cdot x_j\right) \tag{3.14}$$

donde w_{ijk} correspondería al peso que conecta la neurona j con la con retraso o velocidad k. En la figura 3.13 se visualiza la estructura de una red neuronal de ratardo temporal. Esta interpretación tiene una gran importancia ya que es bien sabido que existen retrasos temporales significativos en los axones y sinapsis de las redes de neuronas biológicas

[11].

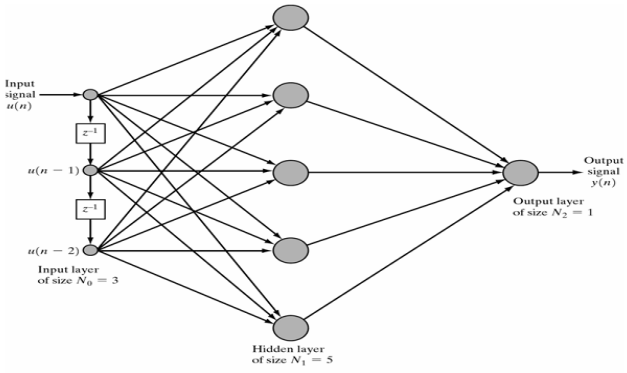


Figura 3.13: Estructura de una red neuronal de retardo temporal

El algoritmo de aprendizaje a utilizar puede ser perfectamente el Backpropagation [71] aunque también es posible utilizar otro tipo de algoritmos que son variaciones del anterior teniendo en cuenta la existencia de retrasos temporales ([56] y [57]).

Método Levenberg Marquardt

Fue diseñado para encontrar las raíces de funciones formadas por la suma de los cuadrados de funciones no lineales, siendo el aprendizaje de redes neuronales, una aplicación especial de este algoritmo. Levenberg Marquardt es una variación del método de iterativo de Newton para encontrar las raíces de una función.

Puede aplicarse en cualquier problema donde se necesite encontrar los valores de las raíces de una función; en el caso de las redes neuronales artificiales, la función es el error cuadrático medio de las salidas de la red y las raíces de esta función son los valores correctos de los pesos sinápticos.

En la siguiente ecuación, se presenta como se localiza un valor mínimo (x_{min}) de una función de una variable $f(x)$, utilizando la primera y segunda derivada de acuerdo al método de Newton.

$$x_{min}(t+1) = x_{min} - \frac{f'(x_{min}(t))}{f''(x_{min}(t))} \tag{3.15}$$

Con base en esta ecuación se puede inferir la ecuación 3.16, donde se minimice el error global E_p en el espacio de los pesos sinápticos representado por la matriz W.

$$W(t+1) = W(t) - \frac{E'_p}{E''_p} \tag{3.16}$$

La segunda derivada del error global $(\left(E_p''\right)$ corresponde a la matriz Hessiana H y la primera derivada $(\left(E_p'\right)$ la conocemos como el vector gradiente G. El vector gradiente y la matriz Hessiana de la función de error los podemos calcular utilizando la regla de la cadena. Así, el vector gradiente se compone por las derivadas parciales del error con respecto a cada uno de los pesos w_i de la red, el elemento (i, j) de la matriz Hessiana se calcula con las segundas derivadas parciales del error con respecto a los pesos w_i y w_j.

Debido a la carga computacional que implica calcular de manera exacta la matriz H, se hace una estimación de la misma [50]. Debido a esto, en la fórmula 3.16 se introduce un mecanismo de control para evitar los problemas que se puedan tener en la actualización de pesos de la red, dando origen a la ecuación 3.17.

$$W\left(t+1\right) = W\left(t\right) - \left(H + \lambda I\right)^{-1} G \tag{3.17}$$

El mecanismo de control para garantizar la convergencia del algoritmo consiste en introducir un factor λI. En primer lugar se prueba la ecuación del método de Newton. Si al evaluarla, el algoritmo no converge (el valor del error comienza a crecer), se elimina este valor y se incrementa el valor de λ en la ecuación 3.17, con el fin de minimizar el efecto de la matriz en la actualización de pesos. Si λ es muy grande, el efecto de la matriz H prácticamente desaparece y la actualización de pesos se hace esencialmente con el algoritmo de gradiente descendente. Si el algoritmo tiene una clara tendencia hacia la convergencia se disminuye el valor de λ con el fin de aumentar el efecto de la matriz H. De esta manera se garantiza que el algoritmo se comporta con un predominio del método de Newton.

El método Levenberg Marquardt mezcla sutilmente el método de Newton y el método Gradiente Descendente en una única ecuación para estimar la actualización de los pesos de la red neuronal.

El método Levenberg Marquardt converge a la solución en menos iteraciones que la regla delta generaliza, pero cada iteración requiere más tiempo, ya que se realizan más operaciones. Cuando se entrena a la red utilizando una gran cantidad de ejemplos, es mejor utilizar Levenberg Marquardt, ya que permite obtener la solución en un menor tiempo.

3.5. Clasificación MIR

Como se mencionó en el capítulo 2, la recuperación de información musical se conforma de una clasificación que depende de las características utilizadas.

1. Análisis simbólico
2. Metadatos
3. Análisis de señales acústicas

Cada clasificación de la recuperación de información musical, se basa en las siguientes fuentes de información:

- **Análisis simbólico**: Las fuentes que contienen representaciones simbólicas y estructuradas de música, pueden ser las partituras digitales (secuencias) en formatos privados (editores de partituras y secuenciadores) o en formatos públicos como MIDI o MusicXML, entre otros. Los datos de mayor realce en este caso es la información melódica, la armónica y la rítmica. También se utiliza la sonoridad y el timbre, aunque es menos frecuente el uso de éstos.
- **Metadatos**: Es la información que acompaña a la música y que no es la propia ella; son los datos técnicos, administrativos, descriptivos, estructurales, etc. La información que usa puede variar de nivel: administrativos (autor, textos, datos de catalogación,$ldots$), descriptivos (género, instrumentación,$ldots$), técnicos (tipo de fichero, URL,$ldots$), de uso (derechos de autor,$ldots$), etc.
- **Análisis de señales acústicas**: En este caso particular, la fuente son las señales de audio que supuestamente contienen música (formatos de ficheros WAV, MP3, etc.). La información utilizada es altura, sonoridad, timbre, duración, textura, etc. En estos sistemas no se ha utilizado la señal directamente porque se piensa que contiene demasiada información irrelevante para la mayoría de las tareas en MIR.

3.5.1. Análisis simbólico

Como se mencionó anteriormente la información proviene de datos estructurados y representaciones simbólicas, como lo son:

- Standard MIDI Files:
  - Pro: Enorme volumen de datos existentes
  - Contra: Severas limitaciones para la representación de la música (para control)
- MusicXML:
  - Pro: Compatibiliza representación y control
  - Contra: Ficheros muy grandes y no está tan extendido
- Otros:
  - Formatos propietarios: Finale, Encore, Sibelius, Band in a Box, etc.
  - Formatos abiertos: ABC, Humdrum, Essen, Lilypond, etc.

Su representación tiene dos dimensiones vertical (altura) y horizontal (duración), las magnitudes en estas dimensiones pueden representarse mediante símbolos, cuando los símbolos de altura y duración están ligados, se dice que la representación es acoplada y si son independientes, es desacoplada.

La representación de las melodías se puede llevar a cabo mediante:

- Cadenas
  - Pro: Algoritmos muy rápidos para construcción y análisis
  - Contra: Resultados muy sensibles a los códigos seleccionados
- Árboles
  - Pro: La duración (ritmo) queda implícitamente codificado en la estructura → resultados menos sensibles a la codificación
  - Contra: Su construcción y análisis son más lentos
- N-Cadenas
  - Palabras musicales
  - Cadenas de Markov (n-gramas)
  - Herramientas disponibles del procesamiento del lenguaje natural

Los problemas presentados en este tipo de recuperación de información musical, son los siguientes:

- Segmentación (motivos, frases y partes estructurales)
- Reconocimiento de:
  - Melodías
  - Géneros (estilos musicales)
  - Tonalidades
  - Acordes
  - Modos (alegre, triste, melancólico, agresivo, etc.)
  - Autores
  - Intérpretes
  - etc.
- Detección y seguimiento de métrica y tempo

Los descriptores que se suelen utilizar son descriptores estadísticos:

- Contadores
  - Notas y silencios (cortos y largos)
- Melódicos:
  - Vertical:

        - Alturas: rango, media relativa y desviación
        - Intervalos: rango, media relativa y desviación
    - Horizontal:
        - Duraciones de notas: rango, media relativa y desviación
        - IOIs: rango, media relativa y desviación
        - Duraciones de silencios: rango, media relativa y desviación
- Armónicos:
    - Notas no diatónicas: cuenta, grado media y desviación de grados
- Rítmicos:
    - Cuenta de notas sincopadas
- Normalidad de las distribuciones de
    - Alturas, duraciones de notas y silencios, IOIs, intervalos y grados no diatónicos

En su mayoría utilizan archivos MIDI, así como partituras; como datos estructurados para la transcripción.

3.5.2. Metadatos

El metadato es una información estructurada que describe, explica, localiza o bien la hace más fácil de recuperar, usar o administrar un recurso. Los tipos de metadato son:

- Alto nivel:
    - Descriptivos:
        - Título, fecha, lugar
        - Compositor
        - Autor de la interpretación o de la secuenciación
        - Género, instrumentación, etc.
    - Estructurales:
        - Tamaños
        - Movimientos, índices de contenidos
        - Métricas, tempos, tonalidades, etc.
    - Administrativos:
        - Cómo, cuándo, por quién, ... ha llegado hasta aquí
    - DE uso:

    - Gestión de derechos

- Bajo nivel:
    - Técnicos:
        - Temporización, resolución, sincronismos, etc.
        - Descriptores melódicos, armónicos y rítmicos.

3.5.3. Análisis de señales acústicas

Dentro de esta categoría existen 4 criterios de categorización, los cuales dependen del nivel de complejidad.

- Temporalidad:
    - Estáticas: tomadas en un instante dado (muestras, decenas de ms)
    - Dinámicas (p.ej., medias o desviaciones de las estáticas a lo largo del tiempo)
- Extensión temporal:
    - Globales: descripción de toda la señal (p.ej., sonoridad)
    - Locales: sólo de una parte (p.ej., tiempo de ataque)
- Nivel de abstracción:
    - Según lo intuitivas que sean
- Proceso de extracción:
    - o Directamente de la forma de onda (p.ej., cruces por cero)
    - o De una transformación de la onda (Fourier, wavelets,...)
    - o Relacionadas con algún modelo de la señal (fuente, filtro, auditivo,$ldots$)

Algunos de los problemas que aún se tienen en recuperación de información musical basada en audio son:

- Identificación de la frecuencia fundamental
- Detección de inicios de notas
- Transcripción automática
- Clasificación de géneros musicales
- Organización de bases de datos musicales
- Segmentación de audio
- Identificación y separación de instrumentos
- Organización de efectos sonoros

3.6. Resumen

En este capítulo se describió a detalle lo que es una melodía así como los formatos que existen y el que es utilizado en esta investigación además de una reseña histórica de las redes neuronales artificiales. También se revisó, para una mejor comprensión de esta investigación, el concepto de una red neuronal de retardo temporal además del método Levenberg Marquardt. Al igual que el enfoque general que se tiene en cada una de las ramas que contempla la recuperación de información musical.

Capítulo 4

Metodología para la recuperación de melodías

4.1. Metodología aplicada

Hasta hace algunos años, la recuperación de información musical se había centrado en técnicas que se enfocaban en el dominio de la frecuencia por los resultados prometedores que ofrecía esta área, analizando los armónicos de las melodías para obtener a partir de ellos, descriptores tradicionales o frecuencias fundamentales; sin embargo los mejores resultados se obtuvieron mediante el análisis espectral de la señal.

Recordando un poco lo mencionado en el capítulo anterior el audio digital es la representación de señales sonoras (melodías) mediante un conjunto de datos binarios. Rho, Hoashi, Little por mencionar algunos de los investigadores que han utilizado este tipo de archivos de audio se han enfocado a extraer diferentes descriptores tradicionales como frecuencia fundamental, contorno melódico, coeficientes cepstrales de frecuencias de Mel entre otros; explotando así, nuevamente el área del dominio de la frecuencia.

El dominio del tiempo había sido olvidado, porque se argumentaba que se contaba con demasiada información irrelevante, pese a que se cuenta con buenas técnicas. Para comprobar lo contrario en este trabajo de investigación, se utilizó la melodía original para llevar a cabo la recuperación o recomendación de información musical.

En este capítulo se propone utilizar las melodías en su estado original, sin realizar ningún pre procesamiento o tratar de adaptarlas a un modelo tradicional. Las señales se introducen directamente a una red de retardo temporal, obteniendo así nuestro propio descriptor, para probar si este tipo de estructura es capaz de detectar información musical, el aplicar cambios a los archivos podría desvirtuar la información contenida en él. A continuación se explica a detalle cada paso de esta propuesta.

4.2. Red neuronal de retardo temporal

Este tipo de redes en sus inicios se usó en el reconocimiento y clasificación de fonemas [88] obteniendo un 98.3 % de efectividad, mejor que otros métodos, mostrando ser robusto ante variaciones contextuales y temporales. Este tipo de redes son capaces de modelar sistemas donde la salida $y\,[t]$ tiene una dependencia no lineal de un intervalo de tiempo limitado de la entrada $u\,[t]$:

$$y\,(t) = F\,[u\,(t-m)\,,\ldots,u\,(t-1)\,,u\,(t)\,,u\,(t+1)\,,\ldots,u\,(t+n)] \tag{4.1}$$

Con este tipo de redes se pueden procesar datos de series temporales como una colección de patrones estáticos de entrada/salida relacionados en función del tamaño de la ventana de entrada. Debido a la ausencia de realimentación, su arquitectura corresponde con la de un perceptron multicapa y se puede entrenar usando un algoritmo estándar de retro propagación del error (backpropagation [71]).

Se construyen pares de entrada/salida para la red, formados por el archivo musical para muestras temporales distintas alrededor de un instante dado t_i. La entrada es $S\,(f, t_{i+j})$ para $j \in [-m, +n]$, siendo m y n el número de ventanas consideradas antes y después del instante central t_i. La salida consiste en una codificación del archivo musical.

Para efectos de esta investigación, la TDNN se usa como predictor. El predecir no es una tarea fácil en ningún campo científico. Los físicos o los matemáticos nos hablan de que el crecimiento de errores, o caos, impide una predicción con certeza de un sistema dinámico. En el campo de la meteorología y economía se asumen los riesgos que conllevan sus predicciones. En el caso de medicina, la predicción se limita a establecer diagnósticos previos y, a partir de ahí, realizar el tratamiento adecuado para la salud del paciente. En nuestro caso, predecir el dato siguiente por medio de un error de recuperación, nos ayuda a detectar la melodía que se esté consultando.

La predicción debe entenderse como un intento permanente de anticipación de un futuro incierto y sobre el que, además, podemos incidir en algunos casos. La predicción no es un fin en sí misma, sino que forma parte de un proceso complejo de toma de decisiones, es por ello aconsejable exponer las técnicas de predicción en el contexto de las situaciones reales en que se aplican, es decir, hay que considerar el entorno. La predicción como se mencionó, es una tarea compleja que exige, en ocasiones, la utilización de estadísticas muy complicadas, uno de los aspectos fundamentales en la construcción del modelo de predicción es el de la elección del número de retardos.

Si clasificamos las técnicas de predicción según el tipo de información que utilizan, ésta puede ser:

- Información subjetiva
- Información histórica
- Información relacional o causal

En la primera, la predicción toma como base la propia opinión del experto sobre el futuro de la cuestión en estudio, la segunda utiliza la propia evolución del fenómeno objeto del estudio en periodos anteriores, siendo la característica clave de este enfoque el estudio de un fenómeno en sí mismo, a través de su evolución temporal (series temporales) y la tercera toma como base, las reglas internas de funcionamiento del tema cuyo comportamiento se trata de predecir.

Nosotros vamos a utilizar la predicción mediante redes de neuronales de retardo temporal en base al análisis aislado de series, ya que por las características intrínsecas de las redes, es el enfoque más idóneo

4.3. Señales utilizadas

Los archivos de audio que se utilizan se encuentran en formato WAV, se ha hablado de este formato en el capítulo anterior. Existen parámetros básicos que se deben tomar en cuenta:

- El número de canales: 1 para mono, 2 para estéreo, 4 para cuadrafónico, etc.
- Frecuencia de muestreo: El número de muestras tomadas por segundo en cada canal.
- Número de bits por muestra: Habitualmente 8 o 16 bits.

Como regla general, las muestras de audio multicanal suelen organizarse en tramas. Una trama es una secuencia de tantas muestras como canales, correspondiendo cada una a un canal. En este sentido el número de muestras por segundo coincide con el número de tramas por segundo. En estéreo, el canal izquierdo suele ser el primero.

La calidad del audio digital depende fuertemente de los parámetros con los que la señal de sonido ha sido adquirida, pero no son los únicos parámetros importantes para determinar la calidad.

En un principio se optó por archivos con frecuencia de muestreo a 44,100 Hz con resolución de 16 bits por muestra, recordando que esa calidad corresponde a la de un CD de música). Sin embargo a lo largo de las pruebas, sólo en algunas de ellas se determinó utilizar diferentes frecuencias de muestreo para examinar que tan robusta es para la recuperación de información musical, en la Tabla 4.1 se pueden observar dichas frecuencias de muestreo, siendo este factor el único modificado. Al momento de comenzar el entrenamiento de las melodías con las TDNN solo se utiliza un canal, para maximizar el desempeño de la propuesta y evitar problemas de cómputo.

Tabla 4.1: Frecuencias de muestreo utilizadas

Frecuencia de muestreo (Hz)
22,050
24,000
32,000
44,100

Como se mencionó en el capítulo 3; la frecuencia de muestreo, es la cantidad de muestras tomadas de una señal, entre más grande sea el tamaño de la muestra más parecido será al sonido original.

4.4. Descripción de la red TDNN

De acuerdo a la literatura el análisis espectral de la señal en el dominio de la frecuencia ha brindado mejores resultados que las técnicas enfocadas al análisis de la misma en el dominio del tiempo. Algunas de las características como el tono, duración, ritmo, correlación cruzada, FFT entre otros están ampliamente ligados a la firma digital. Sin embargo, utilizar directamente la información contenida en la melodía sin hacer uso de firmas digitales o de funciones tradicionales, se evita el pre-procesamiento de la información que estas requieren.

En este trabajo se utiliza la melodía original para realizar la recuperación de información musical. No se ha realizado ningún cambio o tratamiento previo a las melodías para tratar de adaptarlas a un modelo tradicional, se introducen directamente a la TDNN, dicho procesamiento se muestra en la Fig. 4.1. Cada melodía es entrenada por una TDNN independiente.

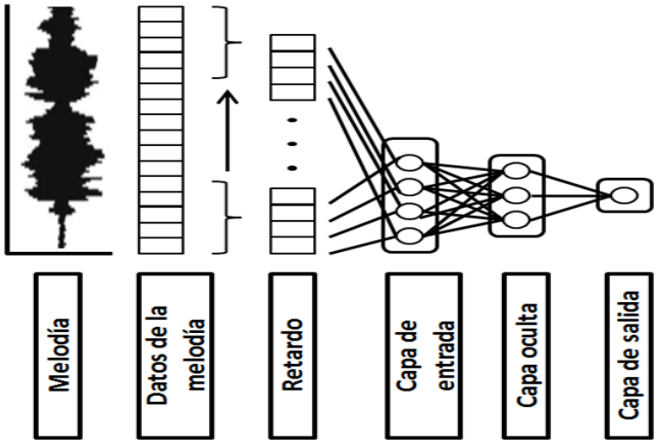


Figura 4.1: Propuesta de la red TDNN

4.4.1. Esquema de la propuesta con TDNN

Para esta investigación se tienen dos bases de datos, posteriormente se darán las características cada una de ellas. Cada melodía es entrenada en una red neuronal de retardo temporal ($TDNN\,(\,)$), donde i^{th} es la melodía, al terminar el entrenamiento se obtiene una matriz de pesos $WNN\_(i)$.

De cada melodía que se almacena en la base de datos, se obtiene un vector de datos que puede ser de longitud variable. El número de retardos es igual al número de neuronas en la capa de entrada para el primer bloque de datos, en otras palabras se hace un ventaneo del vector. Cada una de las ventanas es la siguiente entrada de la red. Dicho proceso se aplica en toda la melodía.

La matriz obtenida es nuestro descriptor musical, descartando por completo cualquier descriptor tradicional, las cuales son almacenadas en una nueva base de datos llamada descriptores. Esto se puede observar en la Fig. 4.2.

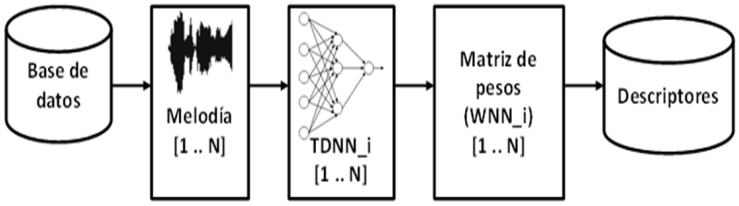


Figura 4.2: Estructura de entrenamiento de las melodías con TDNN

Para la recuperación de una melodía se introduce un segmento consulta, dicho segmento entra a una red neuronal de retardo temporal con los pesos sinápticos previamente entrenados, obteniendo los errores de recuperación por cada red ($Re\_(i)$). Este error se genera a partir de la comparación del segmento consulta en relación con la señal estimada o la predicción de la señal obtenida desde la red. Estos errores son almacenados en un vector, finalmente se aplica $argmin\,(\,)$ retornando un índice (n^*), indicando que red neuronal tuvo el menor error. Este procedimiento se ilustra en la Fig. 4.3.

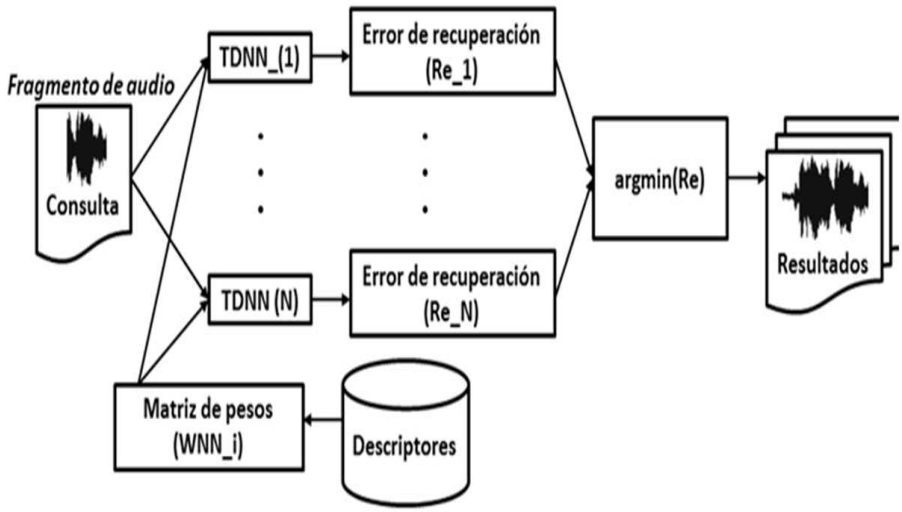


Figura 4.3: Procedimiento de recuperación de una melodía usando el modelo propuesto

El error de recuperación está dado por:

$$Re\_i = \frac{\sum_{j=1}^{N} (x_j - y_j)^2}{w} \tag{4.2}$$

donde x_i son las matrices de pesos previamente entrenados, y_i es el segmento a reconocer, y w es el número de ventanas en las que el segmento fue dividido.

Aunado a la idea principal de obtener el $argmin\,(\,)$, cuando se realiza una consulta de melodías; también se puede hacer uso de funciones de ordenamiento con las que cuenta Matlab, para este caso se ha utilizado $sort\,(x)$, el cual permite ordenar los elementos de un vector en orden ascendente. Por medio de la orden $find$ se encuentra la posición en el vector resultante de los errores obtenidos.

En la literatura se menciona que no existe una solución única para la arquitectura de red neuronal que se debe usar en determinado problema, el diseño de redes neuronales es más un arte que una ciencia, especialmente al escoger el número de capas ocultas, el número de neuronas en cada capa oculta, y las funciones de activación a usarse. Estos factores dependen enteramente del diseñador, aunque hay reglas que surgen más que nada de los resultados obtenidos en otros problemas.

Está la base teórica del Teorema Universal de la Aproximación como sustento de que cualquier función continua puede aproximarse mediante un perceptron multicapa con una capa oculta, pero este teorema no dice nada sobre el tiempo requerido de entrenamiento ni si la implementación es fácil y óptima. Con base a este teorema han sugerido métodos empíricos para determinar el número de neuronas ocultas que debe tener un perceptron multicapa según el problema a resolver, pero no siempre se tienen buenos resultados [26].

Entre más capas ocultas o neuronas se tengan en dichas capas, mejor será la aproximación de la función tratada, pero se debe tomar en cuenta que si se exagera en el número de neuronas ocultas, se puede obtener una red inestable o de un tiempo de entrenamiento muy lento, especialmente con algoritmos de entrenamiento cuyos cálculos son de alta complejidad como el método de Levenberg-Marquardt.

Después de una serie de pruebas se llego a la configuración más eficiente de la red, permitiendo obtener resultados favorables, por lo cual para efectos de esta investigación sólo se utiliza una capa oculta.

Una demostración de la propuesta puede servir para entenderla un poco más. En la Fig. 4.4 se pueden observar los datos que se obtienen de una melodía, los cuales son asignados a un vector que servirá para el entrenamiento de ésta.

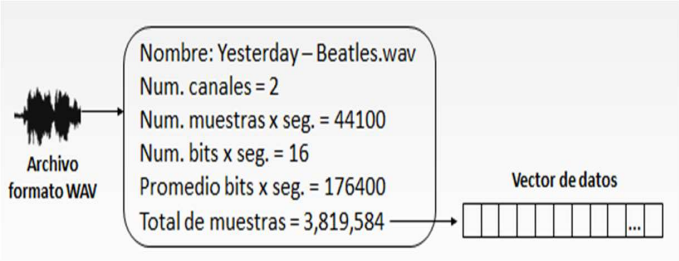


Figura 4.4: Vector de datos obtenido al leer un archivo WAV

Una vez que se tiene el vector de datos se fijan los parámetros que se utilizarán para configurar la TDNN, como se ve en la Tabla 4.2. En la siguiente Fig. 4.5 se muestra el estado inicial de la red.

De acuerdo a los parámetros utilizados, los primeros cinco datos son los que entran a la red, después del primer cálculo, se obtiene la predicción del siguiente dato del vector de datos de la melodía que se está entrenando como se puede ver en las Fig. 4.6 y 4.7, este proceso se realiza hasta terminar la melodía. De esta forma se almacenan los

Tabla 4.2: Parámetros de entrenamiento para la TDNN

Num. retardos = 5
Neuronas capa oculta = 4
Num. iteraciones = 50

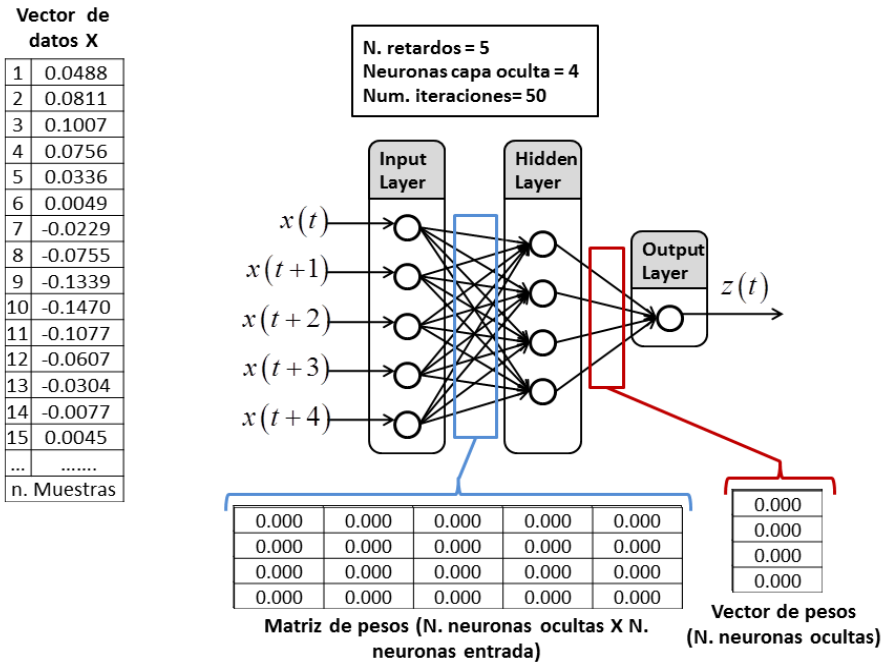


Figura 4.5: Estado incial de la TDNN

parámetros internos de la red en una nueva base de datos, los cuales son utilizados cuando se realiza alguna consulta. Hay que resaltar que esos parámetros son nuestro descriptor propio, que hasta la fecha ningún investigador los ha llegado a utilizar.

De esta forma es como se realiza el entrenamiento de cada melodía. La recuperación o recomendación se lleva a cabo tal como se describió en la Fig. 4.3.

En el siguiente capítulo en el área de pruebas se vuelve a tomar el tema del uso de las diferentes frecuencias de muestreo. En una de las pruebas de esta investigación se habla sobre el ruido Gaussiano, el cual fue agregado al segmento consulta. Es el ruido cuya densidad de probabilidad responde a una distribución normal (o distribución de Gauss). La distribución de Gauss o distribución normal es una distribución de probabilidad muy importante, este ruido es el único que presenta una distribución de Gauss, independientemente de que exista una correlación del ruido en el tiempo o no. Este tipo de ruido se caracteriza porque su energía o densidad es constante sobre todas las frecuencias de la señal.

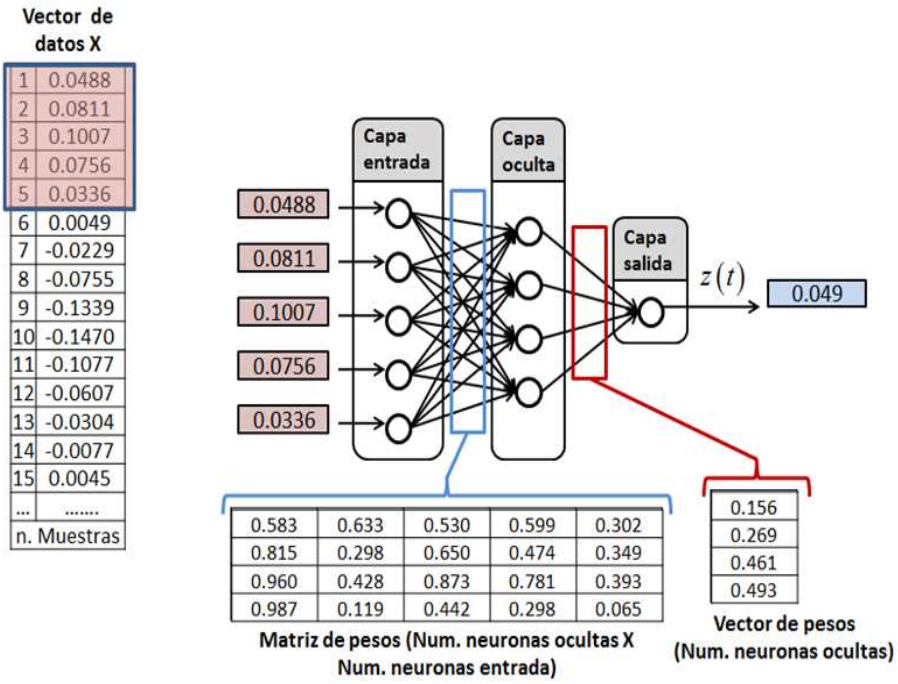


Figura 4.6: Primer cálculo de la TDNN

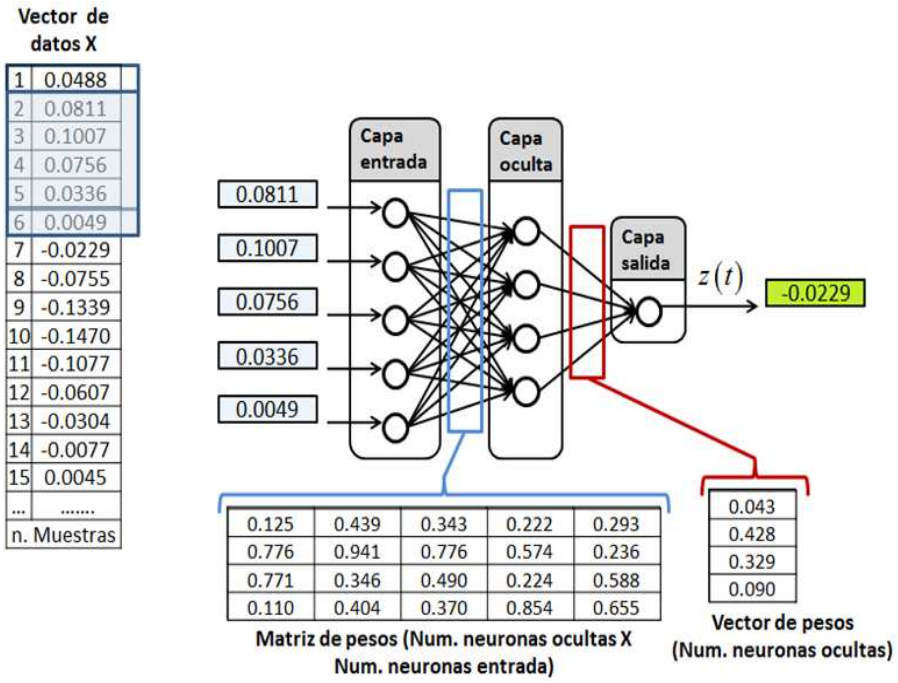


Figura 4.7: Segundo cálculo de la TDNN, que terminará con todo el vector de datos

4.5. Resumen

En este capítulo se describieron algunas características principales para obtener una recuperación musical, desde las señales utilizadas, como la descripción del entrenamiento de las redes neuronales de retardo temporal (TDNN) así como la consulta de una melodía. Se propuso una forma simple de analizar el entrenamiento por medio de un ejemplo. También una pequeña explicación del ruido gaussiano que se utilizó en algunas pruebas.

Capítulo 5

Pruebas y resultados

La metodología propuesta en esta investigación ha sido programada en Matlab, sin olvidar que los datos de entrada no han sido normalizados o que hayan sufrido algún tipo de pre procesamiento. La topología de la TDNN seleccionada consiste de tres capas:

- Entrada (el número de neuronas es igual al número de retardos que tiene la red)
- Oculta)
- Salida (corresponde a la predicción, ya que es una neurona)

El algoritmo utilizado para el entrenamiento es el backpropagation; este procedimiento ajusta iterativamente todos los pesos de la red con el fin de disminuir el error obtenido en la unidad de salida, utilizando el método Levengerg-Marquardt como función de activación. Los pesos de conexión están inicializados aleatoriamente a [0:3]. Por razones de la velocidad de convergencia de todas las muestras entrenadas que se presentan una vez que los pesos se actualizan.

Los archivos de audio que se utilizan son en formato WAV, tipo estéreo polifónico y con una resolución de 16 bits por muestra. Por el momento la longitud de las canciones son de un minuto; dado el procesamiento de las redes en Matlab nos limita a esta cantidad, para no tener problemas de memoria.

Las pruebas realizadas se dividen en dos grupos, porque se trabajó con dos bases de datos diferentes.

- **Base de datos 1**
  - Melodías de Disney
  - Frecuencia de muestreo: 44,100 Hz
  - Total de melodías 800

- **Base de datos 2**
  - Melodías de Beatles y Elvis Presley
  - Frecuencia de muestreo: 22,050 Hz; 24,000 Hz; 32,000 Hz y 44,100 Hz
  - Total de melodías 1,000

5.1. Pruebas iniciales

El objetivo de estas pruebas fue para probar la metodología propuesta, variar los parámetros y de ahí partir con la mejor configuración para pruebas futuras.

Básicamente consistió en realizar el entrenamiento correspondiente de cada melodía con las que se cuentan en la base de datos 1, cabe aclarar que en esta etapa inicial el mismo conjunto de entrenamiento sirvió para realizar las consultas de las melodías; posteriormente se agregó ruido gaussiano.

5.1.1. Prueba 1

La configuración utilizada fue:

- Iteraciones: 10, 25, 50 y 75
- Neuronas capa entrada: 8
- Neuronas capa oculta: 5, 6, 7 y 8
- Segmento consulta: 6,900 datos (aproximadamente 1 segundo)

Los resultados se muestran en las Tablas 5.1 y 5.2, mientras que en las figuras se muestran los errores de entrenamiento y recuperación obtenidos.

Tabla 5.1: Tabla de errores de entrenamiento y recuperación con diferentes números de neuronas

	Entrenamiento			Recuperación		
Número neuronas	Minimo	Promedio	Máximo	Minimo	Promedio	Máximo
5	2.99E-04	2.48E-03	6.12E-03	8.45E-03	**1.41E-02**	**2.21E-02**
6	**2.93E-04**	**2.48E-03**	**5.61E-03**	**5.19E-03**	1.88E-02	5.05E-02
7	5.59E-04	3.67E-03	6.46E-03	6.09E-03	1.95E-02	4.83E-02
8	2.94E-04	3.52E-03	6.43E-03	1.01E-02	1.69E-02	3.00E-02

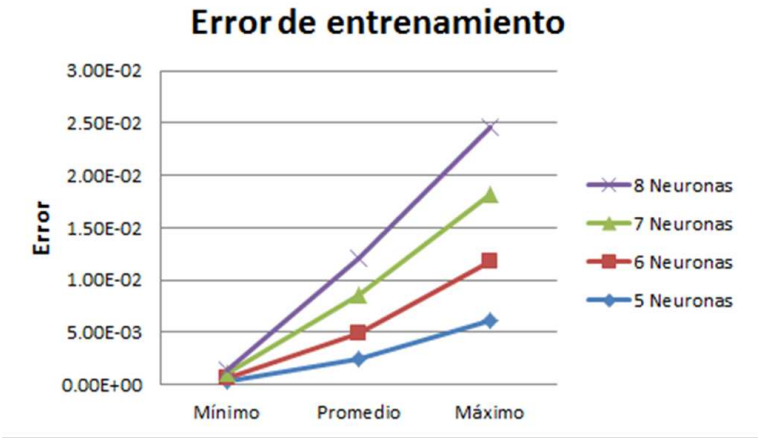


Figura 5.1: Gráfica del error de entrenamiento con diferente número de neuronas

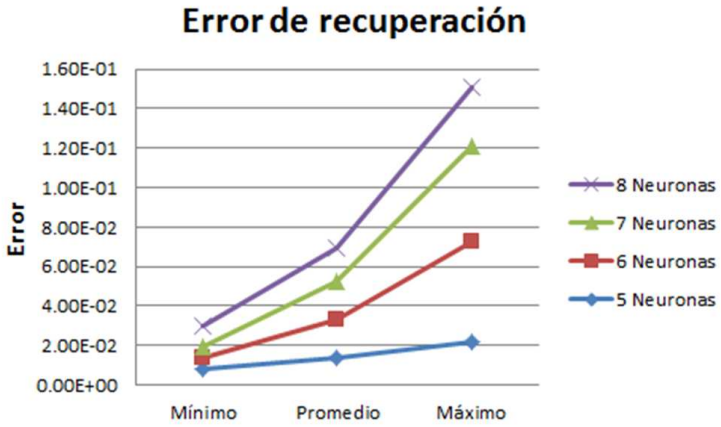


Figura 5.2: Gráfica del error de recuperación con diferente número de neuronas

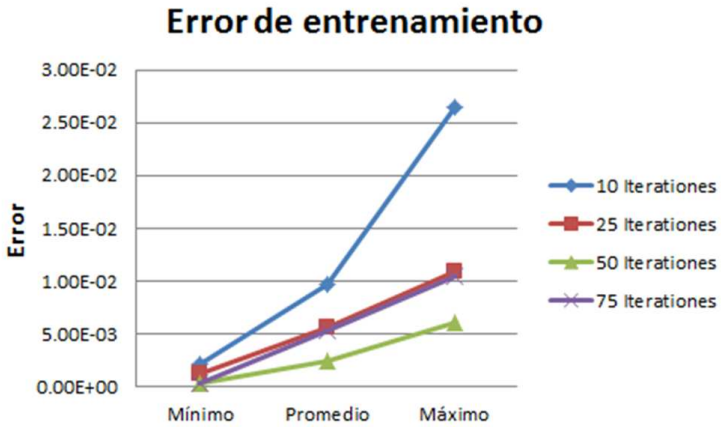


Figura 5.3: Gráfica del error de entrenamiento con diferente número de iteraciones

Tabla 5.2: Tabla de errores de entrenamiento y recuperación con diferentes números de iteraciones

	Entrenamiento			Recuperación		
Número neuronas	Minimo	Promedio	Máximo	Minimo	Promedio	Máximo
10	2.07E-03	9.67E-03	2.64E-02	9.96E-03	3.00E-02	6.37E-02
25	1.15E-03	5.65E-03	1.09E-02	5.49E-03	2.14E-02	5.68E-02
50	2.99E-04	2.48E-03	6.12E-03	8.45E-03	1.41E-02	2.21E-02
75	3.55E-04	5.29E-03	1.05E-02	7.94E-03	2.11E-02	5.08E-02

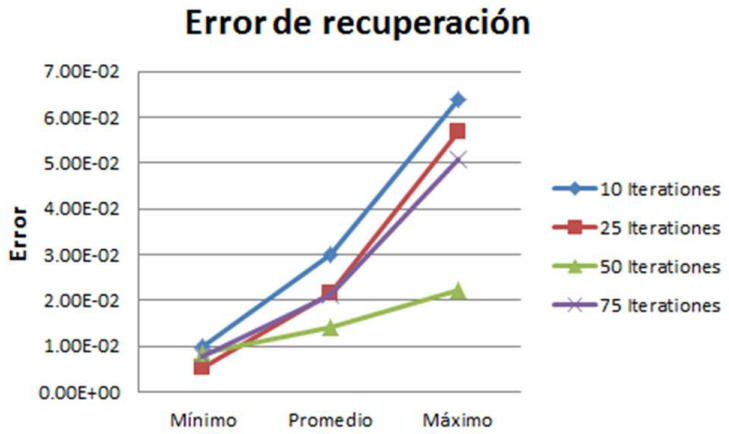


Figura 5.4: Gráfica del error de recuperación con diferente número de iteraciones

El error mínimo obtenido durante el entrenamiento fue con 50 iteraciones, mientras que el de recuperación fue con 25 iteraciones. Sin embargo con 6 neuronas en la capa oculta los errores obtenidos fueron favorables tanto en entrenamiento como en recuperación de las melodías. Estas pruebas sirvieron para detectar los mejores parámetros de configuración para la recuperación de melodías mediante redes neuronales de retardo temporal.

5.1.2. Prueba 2

La configuración utilizada fue:

- Iteraciones: 10, 30 y 50
- Neuronas capa entrada: 8
- Neuronas capa oculta: [5..50] con saltos de 5
- Segmento consulta: 7,100 datos (aproximadamente 1 segundo)

De las ejecuciones realizadas se obtuvieron los errores promedio de entrenamiento y recuperación, éstos se muestran en las Tablas 5.3 y 5.4, así como sus respectivas gráficas en las Fig. 5.5 y 5.6.

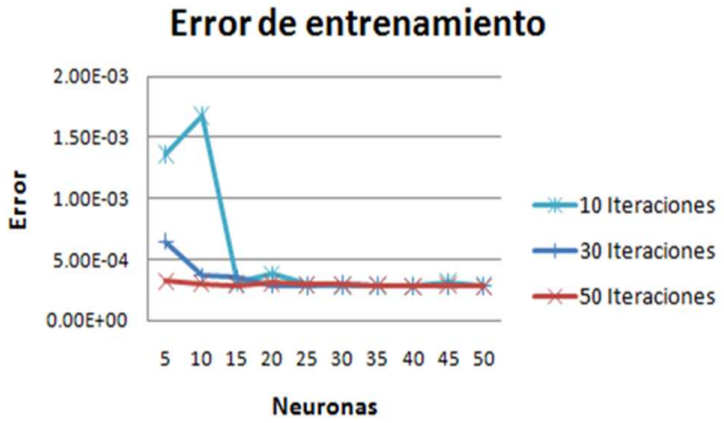


Figura 5.5: Gráfica de errores de entrenamiento de la red TDNN

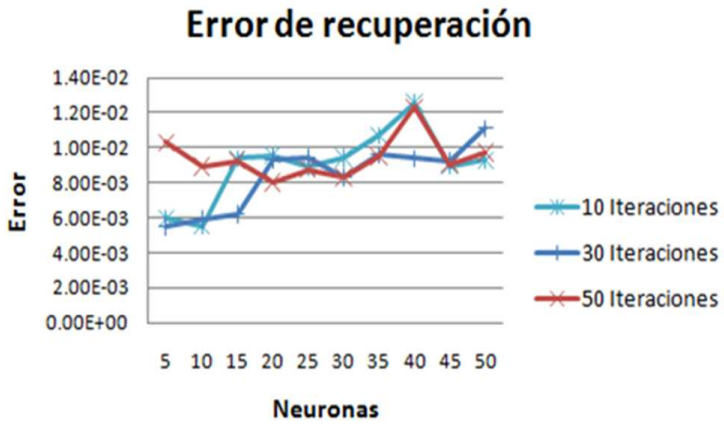


Figura 5.6: Gráfica de errores de recuperación de la red TDNN

El error mínimo obtenido durante el entrenamiento fue con 10 iteraciones y 40 neuronas en la capa oculta, sin embargo para la recuperación se reflejaron los errores mínimos con 10 iteraciones y 10 neuronas. Con tan solo una ventana de 7,100 muestras de consulta es posible realizar el reconocimiento de la melodía; un minuto de melodía equivale a 2,646,000 muestras, eso quiere decir que con menos del 1 % se realiza satisfactoriamente la búsqueda.

5.1.3. Prueba 3

La configuración utilizada fue la misma que en el subtema 5.1.2, salvo que en esta prueba se agregó ruido gaussiano al segmento consulta. El error de entrenamiento más bajo fue de 2.07e-04 estabilizándose después de 40 iteraciones. Se obtuvo un 96 % de recuperación perfecta de las melodías, debido a que cada melodía es codificada en su propia red neuronal el error de recuperación es independiente de las demás melodías. Para este caso el tamaño del segmento consulta creció a 10,500 muestras, debido al ruido agregado.

Tabla 5.3: Errores de entrenamiento de la red TDNN aplicada a melodías

	10 Iteraciones			30 Iteraciones			50 Iteraciones		
Neuronas	Minimo	Promedio	Máximo	Minimo	Promedio	Máximo	Minimo	Promedio	Máximo
5	1.36E-03	6.50E-03	1.10E-02	6.45E-04	5.55E-03	1.05E-02	3.19E-04	5.37E-03	1.07E-02
10	1.68E-03	5.71E-03	1.06E-02	3.63E-04	5.43E-03	1.05E-02	3.00E-04	5.17E-03	1.04E-02
15	3.05E-04	5.26E-03	1.05E-02	3.52E-04	5.17E-03	1.06E-02	2.87E-04	5.23E-03	1.06E-02
20	3.80E-04	5.26E-03	1.04E-02	2.87E-04	5.19E-03	1.04E-02	3.05E-04	5.00E-03	1.04E-02
25	2.90E-04	5.26E-03	1.05E-02	2.85E-04	5.14E-03	1.04E-02	2.94E-04	4.95E-03	1.02E-02
30	2.85E-04	5.22E-03	1.04E-02	2.99E-04	4.98E-03	1.01E-02	2.92E-04	5.01E-03	1.04E-02
35	2.82E-04	5.32E-03	1.04E-02	2.85E-04	**4.92E-03**	1.02E-02	2.88E-04	4.97E-03	1.02E-02
40	**2.79E-04**	5.62E-03	**1.03E-02**	2.82E-04	4.97E-03	1.02E-02	**2.76E-04**	4.96E-03	1.01E-02
45	3.17E-04	5.04E-03	1.03E-02	2.84E-04	5.08E-03	1.03E-02	2.88E-04	4.94E-03	1.01E-02
50	2.84E-04	**5.03E-03**	1.03E-02	**2.80E-04**	4.96E-03	**1.01E-02**	2.80E-04	**4.86E-03**	**1.01E-02**

Tabla 5.4: Errores de recuperación de la red TDNN aplicada a melodías

	10 Iteraciones			30 Iteraciones			50 Iteraciones		
Neuronas	Minimo	Promedio	Máximo	Minimo	Promedio	Máximo	Minimo	Promedio	Máximo
5	5.96E-03	2.32E-02	5.72E-02	**5.49E-03**	1.51E-02	2.30E-02	1.03E-02	**1.47E-02**	**2.21E-02**
10	**5.54E-03**	1.68E-02	2.59E-02	5.92E-03	2.15E-02	4.97E-02	8.93E-03	1.73E-02	2.99E-02
15	9.41E-03	1.69E-02	3.14E-02	6.21E-03	1.55E-02	2.50E-02	9.27E-03	2.01E-02	4.41E-02
20	9.56E-03	1.90E-02	4.52E-02	9.33E-03	2.18E-02	5.81E-02	**8.02E-03**	2.12E-02	5.54E-02
25	8.93E-03	1.70E-02	3.98E-02	9.45E-03	**1.41E-02**	**2.23E-02**	8.75E-03	1.53E-02	2.37E-02
30	9.44E-03	1.97E-02	4.85E-02	8.31E-03	2.01E-02	3.78E-02	8.32E-03	2.04E-02	5.32E-02
35	1.07E-02	2.89E-02	7.97E-02	9.59E-03	1.49E-02	2.44E-02	9.50E-03	2.26E-02	6.09E-02
40	1.26E-02	1.94E-02	3.58E-02	9.38E-03	2.07E-02	5.26E-02	1.23E-02	1.87E-02	3.75E-02
45	8.95E-03	1.76E-02	3.79E-02	9.26E-03	1.84E-02	4.36E-02	9.05E-03	1.73E-02	3.57E-02
50	9.31E-03	**1.44E-02**	**2.08E-02**	1.11E-02	1.64E-02	2.74E-02	9.72E-03	1.81E-02	3.40E-02

La Fig. 5.7 muestra el error de recuperación frente al porcentaje del segmento de música utilizado para la recuperación. Se puede apreciar que cuanto menos sea el segmento consulta es más grande el error de recuperación.

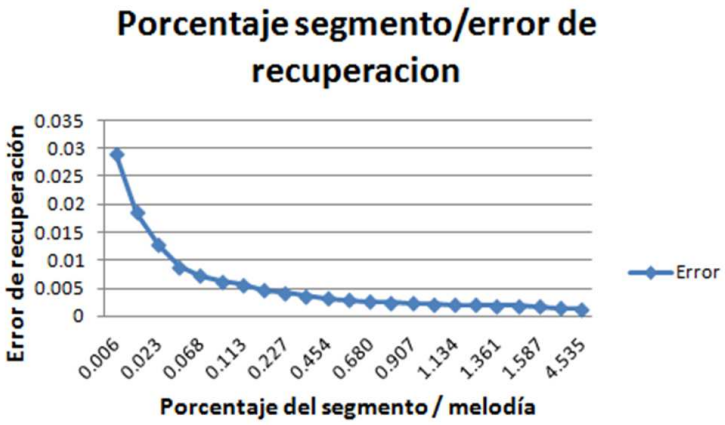


Figura 5.7: Porcentaje de segmento consulta contra el error de recuperación

Finalmente en la Fig. 5.8 se muestran los errores de recuperación con diferentes porcentajes de ruido. Se puede apreciar que entre la gama de ruido de 0 a 50 %, el comportamiento es el mismo, por ende la recuperación de la melodía es satisfactoria. Del 50 % al 75 % aún es aceptable la recuperación, sin embargo al rebasar el porcentaje de ruido en el segmento consulta, la recuperación decae. En ocasiones recupera y en otras no.

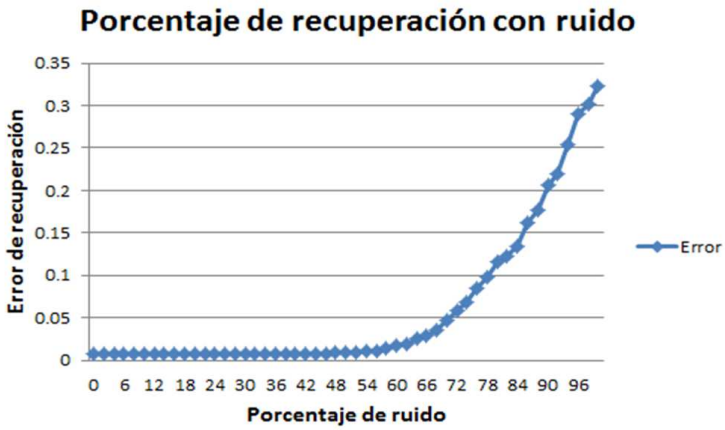


Figura 5.8: Porcentaje de recuperación con ruido

Partiendo de los resultados obtenidos en las tres pruebas anteriores, se analizó la recomendación de melodías utilizando diferentes frecuencias de muestreo, el objetivo fue disminuir el tiempo de entrenamiento, sin afectar la recuperación o recomendación de melodías.

Además de probar la base de datos 2, puesto que para esta etapa se contó con un conjunto de entrenamiento y otro de prueba, la variante para esta base de datos es el tener varias versiones de alguna melodía, que al realizar la consulta permitiera analizar la capacidad de reconocimiento de información musical propuesta en esta investigación.

5.1.4. Prueba 4

La configuración utilizada fue:

- Frecuencia de muestreo: 22:050 Hz; 24,000 Hz; 32,000 y 44,100 Hz
- Iteraciones: 50
- Neuronas capa entrada: 10
- Neuronas capa oculta: 5

La Tabla 5.5 resume las principales características de los datos utilizados.

Tabla 5.5: Características del audio digital (WAV)

Frecuencia de muestreo (Hz)	Tamaño de la ventana consulta (Muestras)	Tiempo (seg)
22,050	15,000 – 58,000	0.7 - 2.6
24,000	20,000 – 63,000	0.8 - 2.6
32,000	30,000 – 70,000	0.9 - 2.1
44,100	30,000 – 70,000	0.7 - 1.5

La Tabla 5.6, muestra el porcentaje de recuperación en función de la frecuencia de muestreo y el tamaño de la ventana de consulta. En cada caso, llega a una recuperación perfecta sin el uso de la recomendación de las melodías.

En la Tabla 5.7, a través de la recomendación de canciones, se puede ver el aumento de la tasa de recuperación mantiene en la mayoría de los casos en un 93 %. Recordando que gestiona un umbral de 5 melodías, hay un caso en la frecuencia de 32,000 Hz, se obtiene una recuperación del 100 % a través de la recomendación de melodías pero con umbral utilizado de 8.

Por ejemplo, una melodía se compone de 7.826.688 muestras, se observó que con sólo 10,000 muestras de la melodía se recupera correctamente a través de la recomendación de las melodías. Del mismo modo, las frecuencias que dan los mejores resultados son los de 32,000 y 44,100 Hz, ya que tiene un número mayor de muestras para realizar una recuperación exitosa. Es importante destacar que incluso con diferentes frecuencias de muestreo el porcentaje de recuperación es aceptable.

5.1.5. Prueba 5

La configuración utilizada fue:

Tabla 5.6: Porcentaje de recuperación con frecuencia de muestreo diferente y tamaño de ventana de consulta

Tam. ventana consulta	% recuperacion	Tam. ventana consulta	% recuperacion	Tam. ventana consulta	% recuperacion	Tam. ventana consulta	% recuperacion
22,050		24,000		32,000		44,100	
20.000	73.33	20,000	86.67	30,000	80	30,000	93.33
22,000	73.33	25,000	86.67	45,000	80	35,000	93.33
30,000	80	35,000	86.67	55,000	80	39,000	93.33
35,000	80	50,000	86.67	60,000	93.33	40,000	100
55,000	93.33	58,000	86.67	70,000	100		
55,000	100	63,000	100				

Tabla 5.7: Porcentaje de recomendación de melodías

Tam. ventana consulta	% recomendación	Tam. ventana consulta	% recomendación	Tam. ventana consulta	% recomendación	Tam. ventana consulta	% recomendación
22,050		24,000		32,000		44,100	
20,000	93.33	20,000	93.33	30,000	100	30,000	100
22,000	93.33	25,000	93.33	45,000	100	35,000	100
22,000	93.33	35,000	93.33	55,000	100	39,000	100
35,000	93.33	50,000	93.33	60,000	100	40,000	100
55,000	100	58,000	93.33	70,000	100		
55,000	100	63,000	100				

- Frecuencia de muestreo: 44,100 Hz; 32,000 Hz; 24,000 Hz y 22,050 Hz
- Iteraciones: 15, 25 y 35
- Neuronas capa entrada: 10
- Neuronas capa oculta: 5 y 10

Es importante recordar que al terminar el entrenamiento de cada melodía se obtiene su matriz de pesos, las cuales se utilizarán como descriptores. La Tabla 5.8 resume las características principales de los datos utilizados.

Tabla 5.8: Características del audio digital (WAV)

Frecuencia de muestreo (Hz)	Tamaño de la ventana consulta (Muestras)	Tiempo (seg)
22,050	25,000 – 60,000	1.1 - 2.7
24,000	20,000 – 54,000	0.8 - 2.3
32,000	15,000 – 45,000	0.4 - 1.4
44,100	10,000 – 40,000	0.3 - 0.9

Como se puede observar la ventana de consulta tiene diferentes valores, ya que a menor frecuencia se cuenta con menos información para realizar un reconocimiento con un segmento pequeño de consulta. Las Tablas 5.9 y 5.10 muestran el rendimiento obtenido en cada configuración utilizada. La Tabla 5.9 muestra el tamaño necesario de la ventana consulta para una recuperación perfecta.

Tabla 5.9: Recuperación perfecta

Tasa de muestreo (Hz):		22,050	24,000	32,000	44,100
Neuronas en la capa oculta	Iteraciones	Rango de la ventana de consulta (Muestras)			
5	15	58,000	54,000	32,000	32,000
	25	41,000	40,000	35,000	29,000
	35	48,000	43,000	38,000	18,000
10	15	59,000	41,000	35,000	40,000
	25	43,000	57,000	39,000	31,000
	35	57,000	46,000	45,000	25,000

La Tabla 5.10 muestra el tamaño mínimo para obtener una recomendación de 10 melodías para el usuario.

El porcentaje de recuperación por recomendación de melodías puede observarse en la Tabla 5.11.

Las llamadas redes neuronales con retrasos pueden ser usadas para describir el contenido musical de melodías para su posterior recuperación, basándose en partes de dichas melodías. Mediante las TDNN se logró resolver el problema planteado sin necesidad de realizar un pre procesamiento, evitando así el utilizar algún descriptor tradicional o firma digital de la melodía.

Tabla 5.10: Recomendación de 10 melodías

Tasa de muestreo (Hz):		22,050	24,000	32,000	44,100
Neuronas en la capa oculta	Iteraciones	Rango de la ventana de consulta (Muestras)			
5	15	42,000	38,000	27,000	25,000
	25	31,000	29,000	24,000	21,000
	35	35,000	31,000	31,000	15,000
10	15	44,000	36,000	25,000	32,000
	25	36,000	52,000	28,000	24,000
	35	51,000	42,000	37,000	19,000

Tabla 5.11: Porcentaje de recuperación por recomendación

Tasa de muestreo (Hz):		22,050	24,000	32,000	44,100
Neuronas en la capa oculta	Iteraciones	Porcentaje %			
5	15	72	74	92	89
	25	79	79	85	92
	35	75	75	83	94
10	15	80	80	84	82
	25	77	77	87	86
	35	74	74	81	87

A diferencia de otras técnicas de MIR, la melodía original se puede considerar como una serie temporal que se introduce directamente en la TDNN, la salida de la red codifica una descripción de la melodía en la matriz de pesos.

Con los resultados obtenidos en esta experimentación se ha observado que el sistema funciona muy bien incluso al trabajar con diferentes frecuencias de muestreo, los mejores porcentajes se obtuvieron con frecuencias de 32,000 y 44,100 Hz, debido a que se tiene una mejor calidad de audio, sin embargo las frecuencias restantes logran realizar una buena recomendación musical. En el presente trabajo se analizó el desempeño del método propuesto al usar versiones diferentes de las melodías aprendidas para la recuperación.

5.1.6. Prueba 6

La configuración utilizada fue:

- Frecuencia de muestreo: 44,100 Hz
- Iteraciones: 35
- Neuronas capa entrada: 10
- Neuronas capa oculta: 5

De la segunda base de datos se tomaron aleatoriamente un conjunto de melodías, sin importar si pertenecen al conjunto de entrenamiento o al de consulta. Esta acción fue para grabarlas con ruido ambiental, y así someterlas tanto al sistema Midomi como a la propuesta manejada en esta investigación. Dichas pruebas sirvieron para realizar la comparativa de dichos sistemas.

La configuración que se utiliza fue tomada de la prueba 5.1.5, con la que se obtuvieron los mejores resultados. En la Tabla 5.12 se muestran los rangos de tiempo utilizados para melodías sin ruido y con ruido.

Tabla 5.12: Rango de tiempos

	Tiempo (seg)	
	Sin ruido	Con ruido
Midomi	5 – 11	1 - 18
TDNN	1.5 – 3	3.5 - 6

Como se puede observar en ambos casos el segmento consulta se incrementa debido al ruido ambiental con el que cuenta la consulta.

Para mostrar un panorama general de esta prueba, en la Tabla 5.13 y 5.14 se muestran algunos de los resultados obtenidos.

En ambos se obtuvieron resultados favorables, sin embargo en el tiempo de consulta de acuerdo a nuestra metodología supera al sistema Midomi, como se puede observar en las tablas anteriores.

5.2. Resumen

En este capítulo se presentó el comportamiento de las redes neuronales de retardo temporal propuesto para la recuperación o recomendación de melodías ante diferentes casos de investigación. Se pudo observar el comportamiento de las redes y su desempeño es bastante aceptable.

Tabla 5.13: Melodías sin ruido

			Midomi			TDNN	
No.	Nombre de la Medolía	Tiempo	Tipo de recuperación	Recomendación	Tiempo	Tipo de recuperación	Recomendación
1	A hard days night	5	Perfecta	0	1.5	Perfecta	0
2	Across the universe	11	Recomendación	3	2	Perfecta	0
3	Back in the USSR	11	Recomendación	8	2	Recomendación	4
4	Blue jay way	11	Recomendación	4	3	Recomendación	4
5	Dont pass me by	11	Recomendación	14	2.5	Recomendación	8
6	Eleanor rigby	11	Recomendación	4	2.5	Recomendación	3
7	For you blue	11	Recomendación	11	3	Recomendación	3
8	Get back	12	Recomendación	12	2	Perfecta	0
9	Glass onion	12	Recomendación	3	3	Recomendación	3
10	Happiness is a warm gun	12	Recomendación	11	3	Recomendación	4
11	Help	5	Perfecta	0	1.5	Perfecta	0
12	Here comes the sun	11	Recomendación	18	3	Recomendación	10
13	In my life	6	Recomendación	11	2	Recomendación	7
14	Let it be	11	Recomendación	6	1.5	Perfecta	0
15	Michelle	5	Recomendación	7	2	Recomendación	5
16	Old brown shoe	11	Recomendación	40	2	Recomendación	10
17	Revolution	9	Recomendación	6	2	Recomendación	5
18	Something	11	Recomendación	2	2.5	Recomendación	2

Tabla 5.14: Melodías con ruido

		Midomi			TDNN	
No.	Nombre de la Medolía	Tiempo	Tipo de recuperación	Recomendación	Tiempo	Tipo de recuperación	Recomendación
1	A hard days night	11	Perfecta	0	3.5	Perfecta	0
2	Across the universe	12	Recomendación	3	4	Perfecta	0
3	Back in the USSR	12	Recomendación	10	4	Recomendación	7
4	Blue jay way	12	Recomendación	4	5	Recomendación	5
5	Dont pass me by	18	Recomendación	14	4.5	Recomendación	8
6	Eleanor rigby	12	Recomendación	7	6	Recomendación	8
7	For you blue	15	Recomendación	11	4.5	Recomendación	6
8	Get back	15	Recomendación	12	4	Perfecta	8
9	Glass onion	12	Recomendación	5	4	Recomendación	5
10	Happiness is a warm gun	12	Recomendación	11	6	Recomendación	7
11	Help	11	Recomendación	8	3.5	Recomendación	8
12	Here comes the sun	15	Recomendación	18	4.5	Recomendación	12
13	In my life	11	Recomendación	11	4	Recomendación	12
14	Let it be	11	Recomendación	10	3.5	Recomendación	9
15	Michelle	11	Recomendación	7	4	Recomendación	5
16	Old brown shoe	18	Recomendación	50	4	Recomendación	16
17	Revolution	13	Recomendación	6	4	Recomendación	7
18	Something	15	Recomendación	7	4.5	Recomendación	7

Capítulo 6

Conclusiones y futuras líneas de investigación

En este capítulo se dan las conclusiones a las que se llegó a lo largo de esta investigación. También se enlistan algunas directivas para investigaciones futuras con miras a dar respuesta a algunas de las preguntas que quedaron sin responder.

Con las publicaciones realizadas en esta investigación, se pudo llegar a una serie de conclusiones parciales que permitieron observar y dar seguimiento al desarrollo de esta investigación.

6.1. Conclusiones

La recuperación de información musical realmente es un campo nuevo de investigación, a lo largo de los años se llevó a cabo su clasificación en análisis simbólico, metadatos y análisis de señales acústicas. Anteriormente diferentes investigaciones se han enfocado a explotar el dominio de la frecuencia y realizar extracción de características; sin embargo como se muestra en este trabajo en el dominio del tiempo es prometedor para la recuperación o recomendación musical.

Se puso en operación una nueva metodología que utilice redes de retardo temporal para la codificación y recuperación de melodías. Para ello, se puso en operación una nueva técnica para codificar melodías mediante los pesos sinápticos de una red neuronal de retardo temporal. Se puso en operación una nueva manera para recuperar melodías mediante redes de retardo temporal sobre la base de los resultados obtenidos. Se logró recuperar o recomendar melodías mediante el uso de redes neuronales de retardo temporal, utilizando su estructura como descriptor propio y sin realizar ningún preprocesamiento a las melodías. La metodología propuesta permite operar con las melodías de diferentes frecuencias de muestreo, que permite obtener un menor costo computacional en el entrenamiento de las melodías.

Se propuso la recuperación de información musical con una estructura que había dado resultados en clasificación y reconocimiento de fonemas. Las redes neuronales de retardo temporal permitieron la recuperación de melodías, trayendo consigo una gran aportación notable sobre las técnicas usadas anteriormente. Se han obtenido grandes logros y buenos resultados, a lo largo de esta investigación; el implementar redes neuronales de retardo temporal (TDNN) utilizando las melodías originales, sin realizarles ningún pre procesamiento y convirtiendo los parámetros internos de la red como nuestro propio descriptor nos ha permitido realizar la recuperación o recomendación de melodías.

El uso de diferentes frecuencias de muestreo en los archivos de audio también permitió obtener resultados favorables, de esta forma se redujo el tiempo de entrenamiento de las melodías. Algo importante a destacar es que con esta propuesta, se obtuvieron los primeros resultados que presentan un gran desempeño a partir de señales de audio digital en formato WAV, lo cual nos permite abrir paso a la recuperación de información musical en el dominio del tiempo.

En la mayoría de los experimentos realizados, se utilizó menos del 1 % de una melodía y los resultados obtenidos fueron muy aceptables. De esta forma los resultados obtenidos permitieron cumplir con los aportes que se mencionaron en el capítulo 1.

En general, la conclusión más importante que se puede hacer sobre esta investigación es que, a pesar de los resultados favorables que se obtuvieron, sólo marcan el comienzo de un campo a ser estudiado más a fondo y obtener resultados prometedores.

6.2. Trabajo futuro

Actualmente existe mucho trabajo que se puede realizar en el campo de la recuperación de información musical, si se sigue trabajando el área que había sido apartada en el dominio del tiempo. Convirtiéndose en la parte central de esta investigación.

A partir de esta investigación se pueden desprender varios trabajos que están dentro del área de la inteligencia artificial. Entre los trabajos más importantes se pueden citar:

1. Recuperación o recomendación de melodías con una variante de las redes neuronales de retardo temporal, las redes neuronales de retardo distribuido

2. Recuperación o recomendación de melodías por medio de memorias asociativas morfológicas.

3. Programación paralela con GPU’s, esto permitiría procesar la información en tiempos mucho más reducidos que los actuales.

6.2.1. Redes neuronales de retardo distribuido

Esta red se deriva de las TDNN, por eso se considera probarlas para la recuperación de información musical. A diferencia de las TDNN solo se realiza el retardo en la capa de entrada, la red neuronal de retardo distribuido (DistDelayNet) puede distribuir las líneas de retardo en toda la red. Existiendo una línea de retardo en cada entrada y los pesos de cada capa. Esto permite que la red tenga una respuesta finita dinámica a los datos de entrada en serie.

6.2.2. Memorias asociativas

El objetivo básico de una memoria asociativa es recuperar correctamente patrones completos a partir de patrones de entrada, los cuales pueden aparecer alterados con ruido. De acuerdo con esto y con respecto al problema que se intenta resolver, una memoria asociativa M puede ser vista como un sistema de entrada y salida, donde las entradas van a ser los rasgos descriptivos del objeto a clasificar y la salida la clase a la que pertenece el objeto.

Teniendo una base de datos de archivos de audio musical, cada uno de estos archivos sin modificaciones de dominio o filtros especiales se entrena en una memoria asociativa (MAM_i), la salida de esta red se obtiene una matriz de pesos (M_i) para cada archivo, esta matriz se almacena, es decir como propio descriptor musical, desechando por completo cualquier descriptor tradicional.

Finalmente para poder recuperar un melodía se introduce un segmento de una melodía, este segmento entra a la memoria asociativa con la matriz de pesos (M_i) previamente entrenada, a partir de aquí se obtiene un error de recuperación (e_i) para cada memoria asociativa, este error se genera de la comparación del segmento consultado con respecto a la señal recuperada de la memoria asociativa, estos errores son guardados en un vector, finalmente a este vector se le aplica el $argmin_{()}$, lo que regresa un índice (n^*), el cual indica en que memorias asociativa se obtuvo el menor error, de esa manera se verifica con que melodía se entrenó esa memoria.

6.2.3. Programación paralela con GPU's

La programación paralela ha cobrado una gran importancia en los últimos años debido a la difusión y uso extendido de los procesadores multinúcleo, la computación en red y los procesadores gráficos programables.

Una de las áreas importantes a tratar con la programación paralela es el procesamiento de señales. La filosofía de diseño de las arquitecturas many-cores como la GPU y sus avances están regidos por la industria del videojuego y su constate demanda de mejores prestaciones. La idea subyacente es optimizar el ancho de banda de muchos hilos al ser ejecutados en paralelo, de forma que si alguno de ellos está esperando la finalización de una operación, se le asigna trabajo para que no permanezca ocioso. Las memorias caché son pequeñas, su función es ayudar a mantener el ancho de banda definido para todos los hilos paralelos. Estas características determinan por ello, que la mayor parte de la arquitectura está dedicada a cómputo y no a técnicas para disminuir la latencia.

Otro punto de discrepancia entre ambos tipos de arquitecturas es el ancho de banda de la memoria, las GPU mantuvieron siempre una brecha en el ancho de banda diez veces superior al de las CPU contemporáneas. Esto obedece a que las arquitecturas de propósito general deben optimizar el ancho de banda para atender a todas las aplicaciones, operaciones de entrada/salida y funciones del sistema operativo coexistentes en el sistema. Por el contrario en la GPU con su modelo de memoria más simple y menor número de limitaciones, es más fácil lograr un mayor ancho de banda de memoria.

El paralelismo es una forma de computación en la cual varios cálculos pueden realizarse simultáneamente, si nos basamos en el principio de dividir los problemas grandes para obtener varios problemas pequeños, que son posteriormente solucionados en paralelo. Puede brindar un menor costo computacional para el entrenamiento de melodías, tan solo hablar de redes neuronales de retardo temporal, lleva implícito que el entrenamiento de las melodías es un tanto lento, por esa razón daría un aporte más a la recuperación de información musical.

6.3. Publicaciones surgidas a partir de esta investigación

A partir de las investigaciones llevadas a cabo durante el desarrollo de esta tesis se han obtenido por el momento las siguientes publicaciones:

6.3.1. En revistas arbitradas

- Gómez, L.E., Sossa, J.H., Barron, R. and Jiménez J.F. (2012). Redes neuronales dinámicas aplicadas a la recomendación musical, *Polibits*, Vol 48. (por aparecer).

- Laura E. Gómez, Humberto Sossa, Ricardo Barrón, Julio F. Jiménez (2012). A new methodology for music retrieval based on dynamic neural networks, International Journal of Hybrid Intelligent Systems *IJHIS*, Vol 9 (1). pag. 1-11.

6.3.2. Foros indexados por ISI proceedings

- Laura E. Gómez, Humberto Sossa, Ricardo Barrón, Julio F. Jiménez (2010). Dynamic Neural Networks Applied to Melody Retrieval. In Grigori Sidorov, Arturo Hernández Aguirre, Carlos A. García, eds., Mexican International Conference on Artificial Intelligence (*MICAI 2010*), vol. 6438, of *Lecture Notes in Computer Science*, 269-279, Springer-Verlag Berlin Heidelberg.

6.3.3. En memorias en conferencias

- Gómez, L.E., Sossa, J.H., Barrón, R. and Jiménez, J.F. A New Approach to Music Information Retrieval using Dynamic Neuronal Networks. In M.A. Martínez and A. Alarcón, eds., *Proceedings of CORE 2010 on Advances in Computer Science and Engineering. CORE 2010*. México, D.F., May 26-28 2010, vol. 45. pag. 41-51.
- Gómez, L.E., Jiménez, J.F., Sossa, J.H., Cuevas, F.J., Pogrebnyak, O. and Barrón, R. Implementation of a swarm intelligence algorithm to a mobile device. In M.A. Martínez and A. Alarcón, eds., *Proceedings of CORE 2010 on Advances in Computer Science and Engineering. CORE 2010*. México, D.F., May 26-28 2010, vol. 45. pag. 317-326.

Referencias

[1] A. Adenso and G. Fred. *Optimización heurística y redes neuronales*. Paraninfo, 1996.

[2] N. Ali and M. Mshtaq. Hybrid query by humming and metadata search system (hqms) analysis over diverse features. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2(9):58, 2011.

[3] J. Anderson. *Neurocomputing: Foundations of Research*, chapter A simple neural network generating an interactive memory, pages 181–192. MIT Press, 1972.

[4] A. Barto, R. Sutton, and C. Anderson. *Neurocomputing: Foundations of Research*, chapter Neuronlike adaptive elements that can solve difficult learning control problems, pages 535–549. MIT Press, 1989.

[5] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere. The million song dataset. In A. Klapuri and C. Leider, editors, *Proceedings 12th International Society for Music Information Retrieval (ISMIR).*, ISMIR, 2011.

[6] C. M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, 1996.

[7] S. Blackburn and D. DeRoure. A tool for content based navigation of music. In *Proceedings of the sixth ACM international conference on Multimedia*, pages 361–268, New York, NY, USA, 1998. ACM.

[8] H. Block. *Neurocomputing: Foundations of Research*, chapter The perceptron a model for Brain Functioning, pages 139–150. MIT Press, 1989.

[9] D. Bogdanov and P. Herrera. How much metadata do we need in music recommendation? a subjective evaluation using preference sets. In A. Klapuri and C. Leider, editors, *Proceedings 12th International Conference on Music Information Retrieval (ISMIR).*, volume DBLP:conf/ismir/2011 of *ISMIR*, Miami, Florida, USA, October 2011. University of Miami, University of Miami.

[10] F. Borrero. Los elementos de la música. In *Innovación y experiencias educativas*, number 13. Dep. legal: GR2922/2007, 2007.

[11] V. Braitenberg. *On the Texture of Brains: An Introduction to Neuroanatomy for the Cybernetically Minded*. Springer-Verlag, 1977.

[12] D. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex systems*, 2:321–355, 1988.

[13] V. Bush. As we may think. The Atlantic Monthly, July 1945. http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

[14] P. Cano, E. Batlle, T. Kalker, and J. Haitsma. A review of algorithms for audio fingerprinting. In *IEEE Workshop on Multimedia Signal Processing*, pages 169–173, May 2002.

[15] G. Chen and X. Dong. *From Chaos to Order: Methodologies, Perspectives, and Applications*. World scientific series on non-linear science: Monographs and treatises. World Scientific, 1998.

[16] P. R. Cook and G. Tzanetakis. Audio information retrieval (air) tools. In *Proceedings 1st International Symposium on Music Information Retrieval (ISMIR).*, ISMIR, Plymouth, Massachusetts, USA, October 2000. ISMIR.

[17] T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to algorithms*. MIT Press and McGraw-Hill, 2001.

[18] J. S. Downie. The scientific evaluation of music information retrieval systems: Foundations and future. *Computer Music Journal*, 28(2):12–23, June 2004.

[19] K. Dressler. Sinusoidal extraction using an efficient implementation of a multi-resolution fft. In *Proc. of the Int. Conf. on Digital Audio Effects (DAFx)*, pages 247–252, Montreal, Quebec, Canada, September 2006.

[20] L. V. Fausett. *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Prentice Hall international editions. Prentice-Hall, 1994.

[21] J. Foote. Content-based retrieval of music and audio. In *Multimedia Storage and Archiving System II, Proceedings of International Society for Optics and Photonics (SPIE)*, volume 3229, pages 138–147. SPIE, 1997.

[22] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith. Query by humming: musical information retrieval in an audio database. In *Proceedings of ACM Multimedia 95*, pages 231–236, 1995.

[23] S. Grossberg. *Neurocomputing: Foundations of research*, chapter How does a brain build a cognitive code, pages 349–399. MIT Press, 1889.

[24] B.-j. Han, S. Rho, and E. Hwang. An efficient voice transcription scheme for music retrieval. In *International Conference on Multimedia and Ubiquitous Engineering (MUE)*, pages 366–371, Seoul, Korea, April 2007. IEEE Computer Society.

[25] M. H. Hassoun. *Fundamentals of Artificial Neural Networks*. A Bradford Book. MIT Press, 1995.

[26] S. Haykin. *Neural Networks, A Comprehesive Foundation*. Macmillan, 1994.

[27] S. S. Haykin. *Adaptive filter theory*. Prentice-Hall information and system sciences series. Prentice Hall, University of California, 4 edition, 2009.

[28] D. Hebb. *Neurocomputing: Foundations of Research*, chapter Introduction and chapter 4, the firts stage of perception: growth and assembly, pages 45–56. MIT Press, 1989.

[29] K. Hoashi, K. Matsumoto, and N. Inoue. Personalization of user profiles for content-based music retrieval based on relevance feedback. In *Proceeding of the eleventh ACM international conference on Multimedia*, pages 110–119, New York, USA, 2003.

[30] H. Hoos, K. Renz, and M. Görg. Guido/mir - an experimental musical information retrieval system based on guido music notation. In *Proceedings 2nd International International Symposium on Music Information Retrieval (ISMIR).*, pages 41–50, 2001.

[31] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 79, pages 2554–2558, April 1982.

[32] J. J. Hopfield. Neurons with graded response have collective computational propierties like those of two-state neurons. In *Proceedings of the National Academy of Sciences*, volume 81, pages 82–86, 1984.

[33] D. Huron. Humdrum: Music tools for unix systems. *Computing in Musicology*, 7:66–67, 1991.

[34] D. Huron and B. Aarden. Cognitive issues and approaches in music information retrieval. *Unpliblished writing. Retrieved*, 9(9):0, January 2002. from http://www.music-cog.ohio-state.edu/Huron/Publications/huron.aarden.MIR.html.

[35] M. Kassler. *Perspectives of New Music*, volume 4, chapter Toward Musical Information Retrieval, pages 59–67. Perspectives of New Music, 1966.

[36] D. E. Knuth, M. J.H., and V. R. Pratt. Fast pattern matching in strings. *SIAM Journal on Computing*, 6(2):323–350, 1977.

[37] T. Kohonen. *Correlation matrix memories*, chapter Correlation matrix memories, pages 174–180. MIT Press, 1972.

[38] T. Kohonen. Correlation matrix memories. *IEEE Transactions on Computer*, C-21(4):10, 1972.

[39] P. V. Kranenburg, J. Garbers, A. Volk, F. Wiering, L. P. Grijp, and R. C. Veltkamp. Collaboration perspectives for folk song research and music information retrieval: The indispensable role of computational musicology. *Journal of interdisciplinary music studies*, 4(1):17–43, 2010.

[40] A. La Burthe, F. Pachet, and J.-J. Aucouturier. Editorial metadata in the cuidado music browser: between universalism and autism. In *Proceedings of the WedelMusic Conference*, Liverpool, UK, September 2003.

[41] K. Lang and G. Hinton. A time-delay neural network architecture for speech recognition. Technical report, Camegie Mellon University, 1988. Tech. Rept. CMU-CS-88-152.

[42] K. J. Lang, A. H. Waibel, and G. E. Hinton. A time-delay neural network architecture for isolated word recognition. *Neural Networks*, 3(1):23–43, January 1990. Publisher Elsevier Science Ltd.

[43] K. S. Lashley. In search of the engram. In *In Society for Experimental Biology Symposium No. 4: Physiological Mechanisms in Animal Behavior (1950)*, pages 478–505. Ambridge University Press, 1950.

[44] K. Lemström and J. Tarhio. Transposition invariant pattern matching for multi-track strings. *Nordic Journal of Computing*, 10(3):185–205, September 2003. Publisher Association Nordic Journal of Computing.

[45] K. Lemström, G. Wiggins, and D. Meredith. A threelayer approach for music retrieval in large databases. In *Second International Symposium on Music Information Retrieval*, pages 13–14, Bloomington, USA, 2001.

[46] K. Lemström, G. Wiggins, and D. Meredith. The c-brahms project. In ISMIR, editor, *Proceedings International Conference on Music Information Retrieval (ISMIR).*, page March, 2003.

[47] C.-T. Lin, C. T. Lin, and C. S. G. Lee. *Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems*. Prentice Hall PTR, 1996.

[48] D. Little, D. Raffensperger, and B. Pardo. A query by humming system that learns from experience. In *Proceedings 8th International Conference on Music Information Retrieval (ISMIR).*, ISMIR, pages 335–338, Vienna, Austria, September 2007. Austrian Computer Society.

[49] B. Logan and A. Salomon. A content-based music similarity function. Technical Report Series CRL 2001/02, Cambridge Research Laboratory, June 2001.

[50] T. Masters. *Advanced Algorithms for Neural Network: A C++ Sourcebook*. John Wiley, 1995.

[51] W. S. McCulloch. *Neurocomputing: Foundations of Research*, chapter A logical calculus of the ideas immanent in nervous system, pages 18–27. MIT Press, 1943.

[52] R. J. McNab, L. A. Smith, D. Bainbridge, and W. I. H. The new zealand digital library melody index. *D-Lib Magazine*, 3(5):0, May 1997.

[53] R. J. McNab, L. A. Smith, I. H. Witten, C. L. Henderson, and S. J. Cunningham. Towards the digital music library: tune retrieval from acoustic input. In *Proceedings of the first ACM international conference on Digital libraries*, pages 11–18, Bethesda, Maryland, United States, 1996. ACM, New York, NY, USA.

[54] M. Minsky and S. Papert. *Neurocomputing: Foundations of research*, chapter Perceptrons, pages 161–173. MIT Press, 1969.

[55] J. Nápoles. Audio y sonido profesional. Academia Centroamericana, Agosto 2008.

[56] K. Narenda and K. Parthasarathy. Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, 1(1):4–27, 1990.

[57] K. Narenda and K. Parthasarathy. Gradient methods for the optimization of dynamical systems containing neural networks. *IEEE Transactions on Neural Networks*, 2(1):252–262, 1991.

[58] F. Pachet. *Encyclopedia of Knowledge Management*, chapter Knowledge management and musical metadata. Idea Group, 2005.

[59] D. Park and E. Hwang. Popularity-adaptive index scheme for fast music retrieval. In *Proceedings of the 2002 IEEE International Conference on Multimedia and Expo, ICME 2002, Lausanne, Switzerland. August 26-29, 2002.*, volume 1, pages 121–124, Lausanne, Switzerland, August 2002. IEEE.

[60] T. Poggio and F. Girosi. Networks for approximation and learning. In *Proceedings IEEE*, volume 78, pages 1481–1497, 1990.

[61] L. Prechelt and R. Typke. An interface for melody input. *ACM Transactions on Computer-Human Interaction*, 8(2):133–149, June 2001. publisher ACM.

[62] F. Ren and D. Bracewell. Advanced information retrieval. *Journal Electronic Notes in Theoretical Computer Science (ENTCS)*, 225:303–317, 2009.

[63] S. Rho, B.-j. Han, E. Hwang, and M. Kim. An adaptation framework for qbh-based music retrieval. In *nowledge-Based Intelligent Information and Engineering Systems, 11th International Conference, (KES). XVII Italian Workshop on Neural Networks, Proceedings, Part I*, volume 4692 of *Lecture Notes in Computer Science*, pages 596–603, Vietri sul Mare, Italy, September 2007. Springer.

[64] S. Rho, B.-j. Han, E. Hwang, and M. Kim. Musemble: A music retrieval system based on learning environment. In *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, (ICME)*, pages 1463–1466, Beijing, Beijing, China, July 2007. IEEE.

[65] S. Rho, B.-j. Han, E. Hwang, and M. Kim. Musemble: A novel music retrieval system with automatic voice query transcription and reformulation. *Journal of Systems and Software*, 81(7):1065–1080, 2008. Publisher Elsevier).

[66] S. Rho and E. Hwang. Fast melody finding based on memorable tunes. In *1st. International Symposium on Computer Music Modeling and Retrieval*, pages 227–239, Montpellier, France, 2003.

[67] S. Rho and E. Hwang. Fmf (fast melody finder): a web-based music retrieval system. In *Lecture Notes in Computer Science 277*, volume 277, pages 179–192. Springer-Verlag, 2004.

[68] S. Rho and E. Hwang. Fmf: Query adaptative melody retrieval system. *Journal of systems and software (JSS)*, 79(1):43–56, 2006.

[69] R. Rojas. *Neural networks: A systematic intriduction*. Springer, 1996.

[70] F. Rosenblatt. *Neurocomputing: Foundations of research*, chapter The perceptron: A probabilistic model for information storage and organization in the brain, pages 92–113. MIT Press, 1958.

[71] D. Rumelhart, H. G.E., and R. Williams. *Neurocomputing: Foundations of Research*, chapter Learning internal representations by error propagation, pages 675–695. MIT Pres, 1986.

[72] D. Rumelhart, G. Hinton, R. Williams, and S. D. I. f. C. S. University of California. *Learning Internal Representations by Error Propagation*. ICS report. Institute for Cognitive Science, University of California, San Diego, 1985.

[73] J. Salamon and E. Gómez. Melody extraction from polyphonic music audio. In *Proceedings Music Information Retrieval Evaluation eXchange (MIREX)*, Utrecht, The Netherlands, 2010.

[74] J. Salamon and E. Gómez. Melody extraction from polyphonic music: Mirex 2011. In *Proceedings Music Information Retrieval Evaluation eXchange (MIREX)*, Miami, USA, 2011.

[75] J. Salamon and E. Gómez. Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech & Language Processing*, 20(6):1759–1770, 2012.

[76] T. Saracevic. Information science. *Journal of the American Society for Information Science (JASIST)*, 50 (2):1051–1063, 1999.

[77] O. G. Selfridge. Pandemonium: a paradigm for learning", mechanisation of thought processes. In *Proceedings of a symposium Held at the National Physical Laboratory*, November 1958.

[78] I. J. Serra, X. Smith. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14:12–24, 1990.

[79] E. Toch and R. Gerhard. *La melodía*. Nacional, 1947.

[80] R. Typke, P. Giannopoulos, R. C. Veltkamp, F. Wiering, and R. Van Oostrum. Using transportation distances for measuring melodic similarity. In *Proceedings 4th International Conference on Music Information Retrieval (ISMIR).*, ISMIR, Baltimore, Maryland, USA, October 2003.

[81] R. Typke, P. Giannopoulos, R. C. Veltkamp, F. Wiering, and R. Van Oostrum. Using transportation distances for measuring melodic similarity. Technical report UU-CS-2003-024, Universiteit Utrecht, 2003.

[82] R. Typke, R. C. Veltkamp, and F. Wiering. A search method for notated polyphonic music with pitch and tempo fluctuations. In *Proceedings 5th International Conference on Music Information Retrieval (ISMIR).*, ISMIR, pages 281–288, Barcelona, Spain, October 2004.

[83] A. Uitdenbogerd and J. Zobel. Melodic matching techniques for large music databases. In *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, MULTIMEDIA '99, pages 57–66, New York, NY, USA, 1999. ACM.

[84] A. L. Uitdenbogerd and J. Zobel. Manipulation of music for melody matching. In *Proceedings of the sixth ACM international conference on Multimedia*, MULTIMEDIA '98, pages 235–240, New York, NY, USA, 1998. ACM.

[85] E. Ukkonen, K. Lemström, and V. Mökinen. Geometric algorithms for transposition invariant content-based music retrieval. In *Proceedings 4th International Conference on Music Information Retrieval (ISMIR).*, ISMIR, pages 193–199, 2003.

[86] C. Van Rijsbergen. *Information retrieval*. PhD thesis, Department of computing science. University of Glasgow, 1979.

[87] V. Vapnik. *The Nature of Statistical Learning Theory*. Statistics for Engineering and Information Science. Springer, 1999.

[88] A. Waibel. Modular construction of time-delay neural networks for speech recognition. *Neural Computation*, 1(1):39–46, March 1989. Publisher MIT Press, Cambridge, MA, USA.

[89] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang. Phoneme recognition using time-delay neural networks. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 37(1):328–339, 1989.

[90] E. Wan. *Finite impulse response neural networks with applications in time series prediction*. PhD thesis, Universidad de Stanford, 1993.

[91] A. S. Weigend and G. N. A. Time series prediction: Forecasting the future and understanding the past. In *Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis*, Santa Fe Institute studies in the sciences of complexity: Proceedings volumes. University of California, Addison-Wesley Pub. Co., May 2009.

[92] B. Widrow, J. Glover, J. McCool, J. Kaunitz, C. Williams, R. Hearn, J. Zeidler, J. Eugene Dong, and R. Goodlin. Adaptive noise cancelling: principles and applications. In *Proceedings IEEE*, volume 63, pages 1692–1716, Decembre 1975.

[93] E. Wold, T. Blum, D. Keislar, and J. Wheaton. Content-based classification, search, and retrieval of audio. *IEEE MultiMedia*, 3(3):27–36, September 1996. Publisher IEEE Computer Society Press.