Departamento de Computación

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

TESIS DE LICENCIATURA

Organización Tímbrica de Instrumentos Musicales utilizando Redes Neuronales

Alumnos: **Néstor Spedalieri** **L.U. 745/90**
Jorge Xifra **L.U. 274/89**

Director: **Dr. Enrique Segura**

Pabellón 1 - Planta Baja - Ciudad Universitaria
(1428) Buenos Aires - Argentina
http://www.dc.uba.ar

INDICE

Resumen 4
Abstract 5
Introducción 6
Capítulo 1 - Antecedentes 9
Grey 10
Toivianen 11
Prandoni 11
DePoli y Prandoni 12
Fujinaga 13
Fraser y Fujinaga 14
Martin 14
Martin y Kim 15
Capítulo 2 - Conceptos necesarios 16
2.1. Breve reseña de los instrumentos musicales 16
Familias de Instrumentos 16
Instrumentos de viento de bronce 16
Instrumentos de viento de madera 16
Instrumentos de cuerda tocados con arco 16
Instrumentos de cuerda pulsados 17
Instrumentos percusivos 17
Instrumentos diversos 17
Características de generación de los sonidos de los instrumentos 17
Instrumentos de cuerda 17
Instrumentos de viento cilíndricos con sección abierta 17
Instrumentos de viento cilíndricos con sección cerrada 18
Instrumentos de viento cónicos 18
Instrumentos de percusión 18
Resumen de Características de los Instrumentos 18
Terminología 19
2.2. Kohonen 20
Descripción Detallada 20
2.3. Wavelets 23
Transformada Discreta de Wavelets 25
Capítulo 3 - Descripción del Modelo 28
3.1. Breve descripción del modelo 28
Esquema 28
3.2. Sonidos 30
Sonidos elegidos para el entrenamiento (por qué los MUMs) 30
3.3. Procesamiento del sonido 30
¿Sonido Mono, Estéreo, Envolvente o Cuadrafónico? 31
Cantidad de sonido a Procesar 31
¡Bajen el Volumen! 32
Un Kilo de Sonido bien trozado, por favor 32
Es hora de achicar el sonido 33

3.4. Mapas Inferiores de Kohonen **35**

3.5. Mapa Superior de Kohonen **37**

Capítulo 4 - Resultados Experimentales ***42***

4.1. Información adicional para la organización **42**

Variables del Modelo (Kohonen) 42

Factor de aprendizaje de Kohonen 44

4.2. Resultado y análisis de la aplicación del modelo **44**

Conclusiones y Trabajos futuros ***58***

Anexos ***60***

Anexo A. Ejemplo de pre-procesamiento de algunos sonidos **60**

1. Violín 60
2. Piano 1 63
3. Guitarra Eléctrica 66
4. Oboe 69
5. Trompeta 72

Anexo B. MUMs **75**

Material de referencia ***76***

Resumen

El presente trabajo muestra un modelo para la generación de mapas tímbricos en dos dimensiones de sonidos de instrumentos musicales. Esto se logra utilizando una arquitectura basada principalmente en la descomposición de los sonidos en bandas de frecuencias, con el objetivo de emular a grandes rasgos al oído humano, el cual analiza la información recibida por octavas. Se utiliza la transformación de Wavelets para la división en frecuencias, y el aprendizaje posterior se realiza con una estructura multinivel de Redes Neuronales Auto-Organizativas de Kohonen. En el análisis de los sonidos se toma en cuenta tanto el ataque como la fase estable de los mismos. Todos los sonidos utilizados tienen la misma frecuencia fundamental y provienen de las grabaciones de la McGill University (McGill University Master Samples). El modelo propuesto puede servir como base para una posterior implementación de clasificadores de sonidos y/o reconocedores de instrumentos.

Palabras clave: timbre, organización, espacio tímbrico, instrumentos musicales, sonido, kohonen, wavelets

Abstract

This study presents a model for the generation of two dimensional timbre maps of musical instruments sounds. This is achieved using an architecture based mainly on the decomposition of sound in frequency bands, with the intention of modelling the behavior of the human ear, which analyzes the received information by octaves. The wavelet transformation is used for the partition of the sound in frequency bands and the subsequent learning stage is performed by a multi-level structure of Kohonen Self Organizing Maps. In analysing the sounds, both the attack phase and the steady phase are taken into account. All sounds have the same fundamental frequency and were taken from the McGill University Master Samples. The proposed model could serve as the basis for the implementation of sound classification and/or instrument recognition.

Keywords: timbre, organization, timbre space, musical instruments, sound, kohonen, wavelets

Introducción

El timbre de un instrumento musical (o en forma extensiva, de cualquier sonido, incluso la voz humana) es la característica del sonido producido, que nos permite identificar *cuál* es la fuente que produce ese sonido, o nos permite decir que dos sonidos que tienen la misma frecuencia, intensidad y duración, provienen de instrumentos o fuentes diferentes. En pocas palabras, el timbre es aquello que nos permite diferenciar, por ejemplo, un piano de un violín. Esta es una definición informal, para explicar en forma simple el concepto de timbre.

Científicamente, el concepto de timbre es aún hoy un punto difícil de definir. Existen varias definiciones que intentan aproximarse, pero al no haberse establecido aún exactamente qué características de un sonido forman el timbre, no llegan a ser completas. Esta indefinición, acerca de lo que se quiere involucrar cuando se habla del timbre, hace que no exista una lista *ya definida* de características únicas de un sonido, en la cual se puedan basar estudios de generación de mapas tímbricos o reconocedores de instrumentos. Existen, por supuesto, un gran número de características que **sí** se sabe influyen en el timbre, pero aún no se han logrado sistemas computacionales que puedan demostrar una perfomance similar a la que puede lograr un oyente humano al reconocer sonidos. El uso de redes neuronales es un aporte que permite darle un nuevo enfoque al estudio de lo que hace que dos sonidos tengan timbres similares, ya que no se basa en características elegidas de antemano para el estudio, sino que se aprovecha la capacidad de estas redes de "aprender" o de "auto-organizarse" basándose en características que se encuentren durante el entrenamiento de las mismas.

Para comprender mejor las dificultades que se encuentran al intentar definir *timbre*, veamos algunas de las definiciones más comunes encontradas en la literatura sobre el tema:

> "...hasta que cuidadosos trabajos científicos hayan sido realizados en el tema, difícilmente pueda decirse algo más acerca del timbre, que es una dimensión ´multidimensional´". [LIC/1951].

> "Timbre es aquel atributo de la sensación auditiva en términos del cual un oyente puede juzgar que dos sonidos presentados en forma similar y teniendo la misma intensidad (loudness) y tono (pitch) son disímiles. El timbre depende primariamente del espectro del estímulo, pero también depende de la forma de onda, la presión del sonido, la localización en frecuencia del espectro, y las características temporales del estímulo" [AME/1960]

> "[El concepto de timbre] puede referirse a las particularidades del tono que sirve para identificar que un sonido musical se origina en algún instrumento o familia de instrumentos, por ejemplo, que es un oboe, o quizás algún tipo de instrumento con doble lengüeta, o quizás algún instrumento de viento. También puede ser usado por un músico para denotar la calidad tonal de la perfomance en alguna fuente instrumental dada, tal como un tono de oboe `oscuro`, `apagado`, `brillante`, `estridente`" [GRE/1975].

> "En esencia, entonces, el timbre puede comprender cualquier subconjunto de fenómenos acústicos que no sean el tono (pitch) e intensidad (loudness) (y, podríamos agregar, duración y ubicación espacial)" [GRE/1975].

> "Ha habido un largo debate acerca de si es esencial, cuando se trata el tema del timbre, tener en cuenta las rápidas evoluciones en el tiempo de la señal, tales cómo la fase de ataque. Parece, sin embargo, que estas pistas de tiempo son especialmente necesarias cuando se trata de *reconocer la fuente del sonido*, al mismo tiempo que son mucho menos importantes en *evaluar la calidad* de un tono" [PRA/1994].

> "Al contrario de otras características de los sonidos musicales, tales como el tono (pitch) o intensidad (loudness), el timbre no puede ser relacionado directamente a una dimensión física; su percepción es el resultado de la presencia y de la ausencia de muchas propiedades diferentes del sonido, el peso de percepción de las cuales es aún muy poco claro". [DEP/1997]

> "Abundan los argumentos acerca de la importancia relativa de varias características del timbre, pero aún no se ha construido ningún sistema que pueda reconocer instrumentos con alguna generalidad significativa"[MAR/1998a].

Como se puede apreciar, en los últimos 40 años ha habido muchos intentos de categorizar los sonidos en cuanto a su timbre, aunque todavía no se ha logrado una definición precisa o un método exacto (matemático) para representar todas las características del timbre.

Una organización automática de los instrumentos musicales de acuerdo a su timbre serviría para varios fines, entre ellos la identificación de la fuente que produce el sonido, categorizar nuevos sonidos sobre la base de su similitud o diferencia con sonidos ya existentes, crear nuevos sonidos que se sitúen entre dos ya existentes, utilizando para eso las características obtenidas en el análisis de los instrumentos.

Una extensión interesante, que seguramente se alcanzará algún día, es la facilidad de contar con un sistema que pueda transcribir automáticamente y en tiempo real la música proveniente de un conjunto de instrumentos (por ejemplo una orquesta), separando la escritura de acuerdo a qué instrumento musical ejecuta cada parte.

El enfoque brindado en este trabajo intenta llegar a una organización o mapeo de los instrumentos musicales de acuerdo a su timbre, dependiendo en la menor medida posible de las subjetividades de oyentes humanos y basándose no en representaciones sintetizadas de los instrumentos musicales, sino en sonidos reales provenientes de los instrumentos originales. El estudio se realiza basándose en sonidos aislados, es decir que se analizan los sonidos de los instrumentos en forma independiente, y no insertos en una frase musical o acompañados de otros instrumentos.

El resto del trabajo está dividido en cuatro secciones o capítulos, más el agregado de algunos anexos relevantes:

Capítulo 1	**Antecedentes** Este capítulo muestra una recorrida por los distintos enfoques históricos en el tratamiento del tema y brinda el *state of the art* de las investigaciones en el campo.
Capítulo 2	**Conceptos Necesarios** En este capítulo se brinda todos los conceptos relevantes, tanto de la terminología del área que se estudia, como de los métodos y técnicas utilizados en el presente trabajo.
Capítulo 3	**Descripción del Modelo** En este capítulo se encuentra la descripción del modelo que proponemos para generar un mapa tímbrico, utilizando la técnica de Wavelets para procesar los sonidos, los que a su vez alimentan a una arquitectura de Redes Neuronales Auto-Organizativas de Kohonen.
Capítulo 4	**Resultados Experimentales** En este capítulo se detallan varias pruebas experimentales realizadas con el modelo propuesto y se muestran los resultados obtenidos.
Conclusiones	Aquí se rescatan las conclusiones sobre el trabajo realizado y se proponen alternativas futuras de estudio.
Anexos	En estos anexos se brinda mayor información sobre algunos ejemplos de procesamiento de sonidos y grabaciones de la Universidad McGill.

Capítulo 1 - Antecedentes

El enfoque más antiguo sobre como caracterizar el timbre se basó en hacerlo a través del espectro de la onda periódica que constituye un sonido, más precisamente del espectro del steady-state de esa onda, es decir de la parte de la señal que es estable, y no tiene modificaciones a lo largo del tiempo. Este enfoque deja de lado otras fases del sonido, como el ataque (el intervalo entre el comienzo del sonido y el momento en que se estabiliza) y el decay (el intervalo entre el comienzo de la disminución del ataque de un sonido hasta su estabilización).

Posteriormente se empezó a considerar que quizás era necesario incluir en el análisis estas otras fases, como así también otros factores que también influían, en especial todas las variaciones que se producen en el sonido a lo largo de la duración del mismo, como por ejemplo las diferentes amplitudes sucesivas de los diferentes armónicos que componen el sonido.

En la actualidad, la investigación continúa por varias ramas, con autores trabajando sólo con el steady-state, y otros con las variaciones temporales del sonido, en este último caso estudiando sólo el ataque de un sonido o considerando además del ataque todas las variaciones del sonido en su duración.

En cuanto a los métodos que se han utilizado para intentar llegar a un mapa de sonidos o espacio tímbrico, ha habido diferentes enfoques a lo largo del tiempo. Estos enfoques dependen, en gran medida, de cuál sea el objetivo final del estudio, el cual puede estar dirigido a reconocer instrumentos, a obtener mapas o espacios tímbricos, o bien a obtener información para sintetizar nuevos sonidos sobre la base de las características de otros (musicalmente, sintetizar es componer un sonido partiendo de elementos básicos).

Podríamos dividir a los investigadores en este tema en tres grandes grupos:

el primer grupo está compuesto por investigadores cuya meta es lograr extraer las características básicas que produce un cierto instrumento, para después automáticamente poder sintetizar todo el rango de sonidos de ese instrumento, de forma tal que posteriormente se puedan utilizar estos sonidos fácilmente sin tener que contar con el instrumento deseado. Un segundo grupo está formado por otros investigadores que buscan encontrar un “espacio tímbrico” cuya organización responde a ciertas características, de forma de poder pasar de un sonido a otro (o en otras palabras, de un instrumento a otro) a través de modificaciones en ciertas características del sonido original (parametrización del espacio tímbrico), y un tercer grupo contiene a quienes buscan obtener una organización de los instrumentos en base a sus timbres de forma de poder reconocer sonidos o decir a cual sonido conocido es similar en timbre uno nuevo que se agrega, organizarlos por taxonomías, y tratar de entender el proceso que utiliza el ser humano para diferenciar entre timbres diferentes. Estos grupos a su vez pueden estar insertos dentro de áreas de investigación más amplias, como puede ser inteligencia artificial, sistemas multimedia, transcripción musical automática, CASA (Computational Auditory Scene Analysis), etc.

Si bien existen muchos estudios históricos relacionados al análisis de los sonidos, como por ejemplo los de Helmholtz [HEL/1954], hemos decidido centrarnos en las investigaciones más relacionadas a los espacios tímbricos y a la generación automática de los mismos.

Se han realizado estudios basándose en el juicio de oyentes humanos, que caracterizan a los sonidos por diversos atributos y luego sobre la base de estos atributos se obtiene el espacio tímbrico

[GRE/1975, GRE/1977, GRE/1978], y otros estudios que se basan en un análisis más automático sin precisar estos oyentes [PRA/1994, MAR/1998b].

En cuanto al tipo de sonidos o señales de entrada para estos estudios, algunos investigadores [PRA/1994, MAR/1998b]han utilizado sonidos de instrumentos reales, es decir, grabaciones de los mismos, mientras que otros investigadores [GRE/1975], en cambio, han analizado los sonidos reales, extraído la mayor información posible, y basándose en ésta han sintetizado sonidos para efectuar los experimentos, de forma de poder modularlos y generar sonidos para el mismo instrumento con diferentes tonos (pitch) y uniformar características tales como la intensidad, duración, y frecuencia entre los diferentes sonidos que forman parte del experimento.

La mayoría de los estudios y experimentos que han sido realizados se han basado en el análisis exclusivo de sonidos aislados, es decir sonidos individuales que no están precedidos ni sucedidos por otros del mismo instrumento, ni son simultáneos con otros sonidos. Existen autores [MAR/1998a], sin embargo, que sugieren que el "entorno" de un sonido es también importante para el reconocimiento de la fuente que lo produce. La idea es que el oyente también considera la sucesión de notas, extrayendo de ella información adicional que no está presente en el análisis de tonos aislados, y que le permite identificar con mayor seguridad un instrumento o familia de estos.

En los puntos siguientes se detallan una serie de trabajos e investigaciones, como ejemplos más representativos de los estudios actuales en este campo.

Grey

Este es el estudio clásico sobre los espacios tímbricos [GRE/1975] [GRE/1977] [GRE/1978], y referencia obligada de todos los estudios posteriores. Si bien no se trata de un sistema computacional *per sé*, sirvió como base para comparaciones de los estudios posteriores, tanto de los realizados con oyentes humanos como con los realizados con procesamiento computacional de los sonidos.

La idea básica fue analizar las respuestas a pares de sonidos, por parte de oyentes humanos. Estos oyentes, que eran personas entrenadas musicalmente, daban un "grado de similitud" a cada par posible de sonidos, tomados entre varios sonidos de instrumentos musicales (16 en total).

El estudio se realizó en varios pasos: en primer lugar se trabajó con sonidos provenientes de instrumentos reales, se los analizó, se extrajo sus características principales y en base a estas se los sintetizó, obteniendo de esta forma sonidos "sintéticos" pero que tenían en común la misma intensidad (loudness), duración y tono (pitch), tratando de eliminar de esta forma cualquier diferencia en estos factores que pudiera alterar el estudio tímbrico propiamente dicho. Estos sonidos fueron presentados de a pares a los oyentes, y estos los clasificaron de acuerdo a su similitud o no. Basándose en esta clasificación se utilizó un multidimensional scaling algorithm (MDS), para tratar de establecer distancias métricas entre los sonidos en un espacio euclidiano, y un algoritmo de clustering para agrupar los sonidos similares sin importar su organización espacial.

Basándose en esto se obtuvo un "mapa" tridimensional, de forma tal que hubiera consistencia entre ambos análisis.

Se puede encontrar una relación entre el primer eje del mapa y la distribución espectral de la energía (brillo o brightness) en los sonidos, el segundo eje pareciera mostrar una relación con la forma de los patterns de onset-offset de los sonidos, principalmente en la *sincronía* en los ataques y decaimientos colectivos de los armónicos superiores. Esta segunda dimensión también pareciera indicar un agrupamiento de acuerdo a familias musicales, con algunas excepciones. La tercera dimensión también estaría relacionada a patterns temporales, tales como la baja amplitud y energía en alta frecuencia en el ataque del sonido. También puede interpretarse a la segunda y tercera

dimensión combinadas, pareciendo de esta forma que representan la dispersión de clusters relacionados con familias de instrumentos, en forma cilíndrica alrededor de la primera dimensión (brillo).

Las conclusiones del trabajo indican que previamente (en estudios anteriores) siempre se encontraba la dimensión que representa el brillo (brightness), pero que era complicado encontrar explicaciones para la segunda dimensión, aunque se la relacionaba con las características temporales que ayudan a la identificación por familias, pero sin saber cuáles podían ser estas características. Al considerarse en este estudio una dimensión más, se puede llegar a una interpretación de cuales pueden ser estas características para encontrar un agrupamiento por familias. Existen excepciones a estos agrupamientos, y se intentó explicarlas con la posibilidad de que ciertos factores físicos puedan ser más importantes que la tendencia de agrupación por familia, y se sugiere principalmente los componentes de articulación del sonido durante la fase de ataque del mismo.

Existe un cierto grado de subjetividad debido al uso de oyentes humanos para establecer las relaciones entre los sonidos, por lo cual las conclusiones e hipótesis sobre los resultados arrastran esta subjetividad.

Se concede también que el estudio se ha realizado con tonos aislados, y que sería muy útil realizar estudios similares con sonidos dentro de contextos de frases musicales, donde pueda obtenerse más información sobre el instrumento.

Toivianen

La idea del trabajo [TOI/1992] fue lograr una organización (o mapa) de timbres mediante la utilización de mapas auto-organizables de Kohonen [KOH/1988].

Básicamente el modelo empleado puede dividirse en dos partes. La primera consta de un procesamiento de las señales (o sonidos) mediante análisis de Fourier (FFT), que descompone la señal en una sucesión de “ventanas”. Cada sonido queda entonces representado por una sucesión de componentes, y utilizando todos los componentes de todos los sonidos se entrena una red de Kohonen, de forma que se obtenga un mapa u organización que represente todo el espacio de “fotos instantáneas” de los sonidos. Todos los sonidos que se presentan para el entrenamiento están “tocados” en 440 Hz (nota LA). Luego se toma cada sonido, es decir su representación en sucesión de ventanas o componentes, y se obtiene un vector que contiene las distintas posiciones dentro del mapa de cada una de las ventanas. Una vez que se tienen estos “trace vectors” de cada uno de los sonidos, existe una segunda etapa, en la que se entrena una nueva red de Kohonen, utilizando estos vectores. La salida final es un mapa de dos dimensiones con una organización de los sonidos en base a su timbre. Los resultados del trabajo son prometedores, pero con varias limitaciones que afectan la organización final de los sonidos. Entre las limitaciones de este trabajo se puede incluir la baja resolución de frecuencias en el procesamiento de los sonidos, y el hecho de que se utilizaron, para los entrenamientos, sonidos artificiales que buscaban imitar a los de los instrumentos reales. De esta forma, todos los sonidos poseen una cierta similitud y se pierden elementos que pueden ser útiles para el estudio del timbre.

Prandoni

Este trabajo [PRA/1994] intenta emular el estudio realizado anteriormente por Grey [GRE/1975], pero reemplazando los oyentes humanos por un procesamiento de los sonidos en forma computacional, para tratar de evitar las subjetividades.

Analiza la parte estacionaria (steady-state) de la señal, basándose en valores que se obtuvieron a partir de procesar los sonidos con el método de Mel-Frequency Cepstral Coefficients.

Los sonidos se relacionan de acuerdo a la distancia que tiene en el espacio de coeficientes. Teniendo en cuenta estas relaciones se realiza un análisis cualitativo-cuantitativo siguiendo las líneas del experimento de Grey. Se obtiene un espacio tímbrico bidimensional con buenas propiedades de clasificación generales y cuyos ejes están en estrecha relación con parámetros que son fácilmente derivables del espectro de las señales.

Los sonidos son pre-procesados mediante MFCC, obteniendo una sucesión de coeficientes que representan, cada uno, una ventana de 32 ms sobre el sonido original, y estas ventanas están distanciadas por 4ms. De cada sonido se toman 800 ms. De esta forma se busca seguir la evolución temporal del sonido. Una ventaja interesante de este método sobre otros, es que permite reducir casi en un 95% el espacio de datos, ya que el MFCC actúa como un método de compresión, conservando las características principales del sonido original.

Una vez obtenidos los coeficientes que representan a los sonidos, se los procesa mediante un multidimensional algorithm y también de un algoritmo de clustering.

Una decisión que se toma es la dejar de lado los detalles temporales y concentrarse lo más posible en las características de las frecuencias de los sonidos. Se parte de la idea que las características temporales son muy útiles para *identificar la fuente de un sonido*, pero son mucho menos importantes para *evaluar la calidad* de un sonido.

Todos los sonidos utilizados para el estudio están la frecuencia de 262 Hz aproximadamente (C4, nota do), y fueron obtenidos de las grabaciones de la Universidad McGill [OPO/1987].

Nuevamente el eje principal o con relación más directa es el que representa el brillo (brightness) de los sonidos. La segunda dimensión es más difícil de caracterizar, aunque pareciera estar relacionada con el concepto de *presencia* utilizado en la ecualización de audio, y que representa la medida de energía contenida en una banda particular del espectro, ubicada entre los 700Hz y 900 Hz.

Básicamente, el estudio pone en evidencia que el uso de la técnica MFCC tiene propiedades interesantes para ser utilizada en el estudio de sonidos musicales (antes se la usaba principalmente para el reconocimiento de voz humana).

Tomando como base el estudio del steady-state de la señal, se obtiene una organización en base a dos características: el brillo y la presencia. Se dejan de lado los factores temporales, los cuales, se afirma, están ligados a la forma física del instrumento, y a los que se asocia más con un reconocedor de instrumentos musicales, que con una organización o mapa tímbrico de los sonidos.

DePoli y Prandoni

Este trabajo [DEP/1997] se basa en gran parte en el mencionado en el punto anterior [PRA/1994], y conserva muchas de sus características, tales como nota en la que están los sonidos, fuente de los mismos, técnica MFCC, etc.

La diferencia reside en que se utilizaron posteriormente mapas de Kohonen (SOM) [KOH/1988]. Se utilizaron para entrenar la red 40 instrumentos. Con MFCC obtienen la sucesión de ventanas (tales como se explica en el punto anterior), y cada una de las 27 ventanas de cada sonido esta formada por 6 coeficientes.

El mapa de Kohonen utilizado es rectangular, de 15*30 neuronas (450). Se entrena la red con los 40 * 27 vectores, y luego se obtiene el “trace vector” de cada sonido. Luego “se dibuja” ese “trace vector” sobre la red de 15*30, es decir que el sonido de un instrumento determinado se visualiza por las líneas que unen a las neuronas que disparan sus “trace vectors”.

Se observa una fuerte relación entre el eje principal del mapa y el concepto de distribución de energía espectral (brightness), lo cual tiene una analogía a los espacios tímbricos obtenidos en los estudios realizados por Grey [GRE/1975]. El segundo eje no tiene una relación directa con un orden determinado, aunque pareciera indicar los agrupamientos locales en los cuales se acercan espacialmente las familias musicales.

Se utilizó también un análisis de componentes principales (PCA) que es una proyección lineal (en diferencia a Kohonen), y se obtuvo un mapa de tres dimensiones (en base a un conjunto de sonidos diferente al experimento anterior, obtenidos de un sintetizador, y esta vez en la frecuencia 311.1 Hz - nota mi bemol 4 -). El procesamiento en MFCC es similar al del experimento anterior. Se realizó un análisis PCA y se llegó a que el 80% de la varianza está concentrada en los primeros tres componentes. De aquí se puede obtener un mapa en tres dimensiones. Se pueden realizar varias interpretaciones de este mapa, y después de diferentes análisis se observa que el eje principal está asociado al brillo (brightness, spectral energy distribution), por lo tanto instrumentos tales como el oboe, basson, y el horn (brillantes) están a máxima distancia geométrica de los sonidos más oscuros, como el vibráfono, la guitarra y el piano. El segundo eje pareciera estar asociado a los valores de energía de la banda que va de 0.6 a 6 Khz. La tercera dimensión parece estar asociada a aspectos más sutiles de la característica espectral, ya que está relacionado a los contenidos de energía de una región angosta del espectro, situado alrededor de los 700Hz. (el mismo concepto de presencia del punto anterior). La conclusión es que la tercera dimensión es un factor de diferenciación en la calidad del timbre musical que actúa de una forma independiente de la calidad llamada brillo (brightness).

"En otras palabras, los detalles temporales tales como la etapa de ataque, la cual dejamos afuera, son mucho más sensibles a modificaciones o ajustes, ligados como están a lineamientos de la estructura instrumental; no es sorprendente entonces que las pistas temporales sean la clave para reconocer un instrumento, ya que parecen permanecer constantes entre los diferentes matices del color tonal. Teniendo esto en cuenta, y del lado de la percepción musical, los resultados mostrados aquí parecen proveer buenos argumentos para el debate general sobre el rol de los detalles temporales en la percepción del timbre. Mientras Grey consideraba la fase de ataque como un factor preponderante para determinar la calidad del timbre, otros investigadores como Sundberg [SUN/1991] mantienen la preponderancia de las características de la fase de steady-state cuando *evalúan* la calidad del timbre, y mueven la importancia del ataque al acto de *reconocer* sonidos. Los espacios tímbricos físicos que obtuvimos algorítmicamente parecen apoyar este segundo punto de vista"

Fujinaga

Este estudio [FUJ/1998] busca lograr un reconocedor de instrumentos musicales, y se basa para ello en el análisis del steady-state de la señal de 39 instrumentos musicales, tocados en tonos (pitch) diferentes, 1338 espectros en total.

La técnica utilizada es una combinación de un clasificador de vecino k-cercano (k-nearest neighbor) y un algoritmo genético. Este último se usa para encontrar el conjunto óptimo de pesos de las características para mejorar la clasificación.

La ventaja del método es que no requiere entrenamiento, su aprendizaje es incremental y muy rápido. El mayor problema es que exige mucha memoria, ya que se deben guardar en ella todos los ejemplos.

La idea es que los sonidos se categorizan por su similitud con uno o más sonidos almacenados.

Las características de los sonidos utilizadas en este experimento fueron: la masa o la integral de la forma (momento de cero orden), el centroide (momento de primer orden), la desviación standard (momento de segundo orden), la skewness (momento de tercer orden), curtosis (momento de cuarto orden), momentos centrales de orden alto (hasta el décimo), la frecuencia fundamental, y las amplitudes de los parciales armónicos, que resultaron en cientos de características.

Los sonidos provienen de los samples de la Universidad McGill [OPO/1987], y basándose en el steady-state se calculo las amplitudes y fases para todos los armónicos de las notas, hasta los 10 Khz.

Según los análisis iniciales, el centroide fue la mejor característica de reconocimiento (en el orden del 20%). Después de utilizar algoritmos genéticos, los mejores resultados se obtuvieron utilizando 7 características: la fundamental, la integral del espectro, el centroide, la desviación standard, la skewness, y los dos primeros armónicos.

Entre las conclusiones del estudio, surge que el grado de reconocimiento varía mucho de acuerdo al instrumento que se intente reconocer: trompeta asordinada y corno francés fueron reconocidos un 90% de las veces y el violín pizzicato y el violín martelé solo lo fueron en un 8% y 15% respectivamente. El promedio en general fue del 50%.

Fraser y Fujinaga

Se basa en el anterior trabajo [FUJ/1998], pero esta vez [FRA/1998] se analiza la parte del ataque de varios sonidos. Se vuelve a utilizar un clasificador del vecino k-cercano y algoritmos genéticos.

Se tomaron ventanas de análisis de 2048 puntos cada una, y en cada una de ellas se calcularon varias características: la masa o integral de la curva (momento de cero orden), el centroide (momento de primer orden), la desviación standard (momento de segundo orden), la skewness (momento de tercer orden), y el tono estimado. Luego, y para considerar la evolución dinámica del espectro, se calculo la distancia total recorrida por cada característica sobre las diferentes ventanas de análisis, además de la media y la desviación standard de estas distancias. Lo que se almacena por cada tono son la velocidad, media y desviación standard para cada una de las 5 características, más las cinco características medidas en la última ventana (20 características en total).

Los algoritmos genéticos consumen mucho tiempo en encontrar los pesos, pero luego el clasificador utiliza un tiempo insignificante.

Se observa una mejora de entre el 10% y el 20% en el grado de reconocimiento, con relación al estudio anterior. La tasa de reconocimiento creció de 50% a 64% con relación al mismo estudio mencionado.

Martin

Este trabajo [MAR/1998c] está inserto en la temática de CASA (Computational Auditory Scene Analysis), que es el enfoque de crear sistemas de computación que aprendan a reconocer fuentes de sonido en ambientes de audición complejos.

En él se detallan una serie de características acústicas, relacionadas con propiedades físicas de los objetos que producen los sonidos, y se presenta el correlograma log-lag como una representación de la señal que codifica muchas de las características propuestas.

Se consideran instrumentos musicales de las familias de cuerdas, metales y maderas. La representación correlograma consiste en tres pasos: 1) la señal acústica “cruda” se pasa por un banco de filtros “gammatone”, que intenta modelar la resolución de frecuencia de la cóclea (caracol óseo del oído interno); 2) las salidas de los filtros son rectificadas a “media onda” y suavizadas ligeramente como si fuera un modelo “tosco” de la transducción de las “inner hair cells” (en el oído interno). En los canales de frecuencias altas, estas operaciones remueven estructuras temporales pequeñas mientras preservan las envolventes (curvas) de las señales. 3) la salida de cada canal es sujeta a una autocorrelación, implementada por una simple arquitectura delay/multiply/smooth. La salida de la autocorrelación es computada como una función de tiempo para lags espaciados uniformemente en una escala logarítmica. Además la autocorrelación zero-lag da una medida de la energía short-time en cada canal. Aunque realizar estos cálculos es costoso, son fácilmente adaptables a arquitecturas de procesamiento paralelo.

En la representación de correlograma, que es tridimensional, la primera dimensión (posición de la cóclea) muestra una resolución de frecuencia, que es capaz de resolver los primeros 5 o 6 armónicos de una señal periódica. La segunda dimensión (lag de autocorrelación) es una representación logarítmica de la periodicidad, que se corresponde con la resolución de pitch casi logarítmica que exhiben los humanos. La tercera dimensión es el tiempo.

Las características que se proponen para el análisis son: Pitch, Modulación de frecuencia, Envolvente espectral (curva espectral), Centroide espectral (se relaciona directamente con cualidades subjetivas como el brillo (brightness)), intensidad, envolvente de amplitud, modulación de amplitud, asincronía de onset e inharmonía o inharmonicidad, entre otras.

Dentro del área de la que se habla en este estudio, se habla de organizar los instrumentos musicales en una jerarquía, basada en propiedades acústicas: en el nivel más alto los tonos se dividen en transient (sonidos breves, como los percusivos) o sostenidos. Dentro de los sostenidos se dividen en soplados o tocados con arco. Los soplados pueden dividirse en maderas o metales.

Cada una de las categorías mencionadas tiene propiedades acústicas características. Por ejemplo los sonidos “transients” tienen onsets rápidos y decaimiento logarítmico. Dentro de la clase de los sostenidos, las cuerdas con arco tienen onsets muy largos (los armónicos parciales tardan un largo tiempo - más de 250ms - en llegar a su steady-state). Dentro de la clase de los vientos, los bronces tienden a tener estructuras simples de formantes, así como “blips” de amplitud y modulaciones de pitch características en el onset.

En este estudio también se hace una crítica a tratar de identificar sonidos considerando sonidos aislados. De hecho, una parte corta de una frase musical lleva a un mejor reconocimiento que los tonos aislados.

Todo esto es una propuesta que no esta implementado aún.

Martin y Kim

La idea final es conseguir un reconocedor de instrumentos musicales [MAR/1998b], y para ello se aplicó una técnica estadística de reconocimiento de patrones, dentro de una taxonomía jerárquica. Se utilizaron características acústicas salientes (representadas por 31 coeficientes) de cada sonido. Se utilizaron 15 instrumentos orquestales, con su rango completo de sonidos. La idea de procesamiento de los sonidos es la que se menciona en el trabajo anterior.

Utilizando divisiones de 70%/30% entre los conjuntos de entrenamiento y testeo, se construyeron clasificadores basados en modelos gausianos, a los que se llegó a través del análisis de multidiscriminante de Fisher. Estos clasificadores distinguieron entre tonos transient y continuos con un 99% de perfomance correcta. Las familias de instrumentos fueron identificadas con un 90% aproximadamente de exactitud y el reconocimiento de instrumentos individuales se logró con un 70% de exactitud.

La idea básica detrás de esta investigación es que la identificación de instrumentos se realiza primero por la familia y luego el instrumento. Para reducir los requerimientos de entrenamiento, se emplearon análisis de discriminante múltiples de Fisher en cada punto de decisión de la taxonomía. La técnica de Fisher proyecta de un espacio de características multidimensional a un espacio de menores dimensiones. Además de la técnica de Fisher, se probaron dos variedades de clasificadores Knn (k nearest neighbor). Estos trabajan memorizando todos los ejemplos de entrenamiento. Cuando aparece un nuevo ejemplo para clasificar, el sistema encuentra los k ejemplos más cercanos en el espacio de características y el nuevo ejemplo es clasificado por regla de mayoría.

Actualmente estos investigadores están trabajando en poder ampliar este análisis al estudio de frases musicales, no solo de tonos aislados, de forma de poder generalizar aún más los resultados obtenidos, en un contexto más realista.

Capítulo 2 - Conceptos necesarios

2.1. Breve reseña de los instrumentos musicales

Familias de Instrumentos

Tradicionalmente los instrumentos se dividen en familias, basándose principalmente en sus características básicas de construcción o forma de obtener el sonido. Las grandes divisiones que pueden hacerse son las siguientes: instrumentos de viento de bronce, instrumentos de vientos de madera, instrumentos de cuerda tocados con arco, instrumentos de cuerda pulsados e instrumentos percusivos

Veamos una breve descripción de cada una de estas familias:

Instrumentos de viento de bronce

Dentro de esta familia podemos encontrar a instrumentos tales como la tuba, la trompeta, el corno francés (o trompa), la corneta y el trombón. Aparte del material con el que están hechos, estos instrumentos comparten ciertas características como método de emitir el sonido, como por ejemplo el cambio de longitud del recorrido interno del aire dentro del instrumento (esto se consigue mediante válvulas que desvían el aire por diferentes tubos produciendo mayor o menor recorrido) por el cual se logran diferentes entonaciones, la importancia de los labios del ejecutante (que son los que producen el sonido al vibrar en la boquilla) y el tipo de boquilla (en forma cónica o de taza) que usa el ejecutante para ingresar el aire en el instrumento.

Instrumentos de viento de madera

Dentro de esta familia encontramos instrumentos tales como el fagot, el oboe, el clarinete y el corno inglés. Si bien la familia suele denominarse “maderas”, existen instrumentos, como la flauta traversa, que pueden estar hechos de metal. La característica básica de esta familia es que su sonido se consigue introduciendo el aire en el instrumento a través de un agujero de soplido o de una boquilla con lengüeta simple o doble y que el recorrido interno del aire no tiene diferentes longitudes (como en el caso de los bronces) sino que las diferentes notas se consiguen al dejar abiertos o cerrar agujeros hechos a lo largo del tubo hueco que es el instrumento. Existen instrumentos que de alguna forma son híbridos, como por ejemplo el saxofón (alto, tenor y barítono), que, si bien están hechos de materiales similares a los de la familia de bronces, logran su sonido al introducir el aire mediante una boquilla con lengüeta simple y cerrar o abrir agujeros a lo largo del instrumento (sin cambiar la longitud del recorrido del aire)

Instrumentos de cuerda tocados con arco

Aquí se agrupan instrumentos tales como el violín, la viola, el violoncello y el contrabajo. La característica básica de esta familia es que el sonido se produce al hacer pasar un arco por las cuerdas tensas, que al vibrar transmiten esa vibración a una caja de resonancia (madera), que la amplifica. Las diferentes notas se consiguen al “acortar” la longitud de la cuerda al presionar con los dedos las cuerdas contra el diapasón (mango).

Instrumentos de cuerda pulsados

Aquí se puede agrupar a instrumentos como el arpa, la guitarra, el laúd, e incluso a instrumentos de arco que sean ejecutados pulsando las cuerdas y no a través de un arco.

Instrumentos percusivos

Aquí encontramos instrumentos como los timbales, el xilofón, el vibráfono, las campanas tubulares y la marimba. Su sonido se produce al golpear sobre el instrumento con una baqueta, que puede tener una cabeza suave (recubierta con fieltro), o una cabeza dura (hecha de alguna madera especialmente dura), o bien al golpear con un martillo o maza, como puede ser el caso de las campanas tubulares.

Instrumentos diversos

También tenemos instrumentos como el piano y el órgano que comparten características con más de una familia. El piano posee cuerdas de alambre estiradas que producen el sonido al ser golpeadas por pequeños martillos que se mueven en respuesta a la presión sobre las teclas del instrumento. El órgano, a su vez, produce su sonido cuando el aire pasa a través de tubos de diferentes diámetros y longitudes, que se abren y se cierran al presionar o liberar las teclas.

Características de generación de los sonidos de los instrumentos

Instrumentos de cuerda

El sonido de un instrumento de cuerda no es simplemente el sonido producido por la señal sinusoidal derivada directamente de la longitud de la cuerda que se pulsa, sino que esta formado por la adición de otras componentes, llamados armónicos, que se producen al dividir la cuerda en partes iguales. Es decir, al tocar una cuerda "al aire", también están resonando las frecuencias de una cuerda de la mitad de longitud de la original, de un tercio de la longitud de la original, de un cuarto, etc. Esta sucesión de armónicos es infinita teóricamente, y la presencia e importancia de cada uno de estos armónicos durante el tiempo que dura el sonido (pueden ir creciendo o decayendo) es lo que le da a cada instrumento su color particular (o timbre). Entonces, el sonido de una cuerda es la suma de todos estos armónicos. Esta vibración producida en la cuerda necesita algo que la haga resonar, para poder ser escuchada. Si una cuerda no tuviera una caja de resonancia, su sonido sería casi imperceptible. De aquí que los instrumentos de cuerda necesiten una caja, llena de aire, donde la vibración provocada en la cuerda haga mover al aire de la caja, resonando contra la madera. Debido a esto es que instrumentos de cuerda de distintas maderas suenan en forma distinta, a pesar de estar produciendo la misma nota y con cuerdas de los mismos materiales. El hecho es que hay maderas que resuenan mas que otras, produciendo diferentes sonidos o colores.

Instrumentos de viento cilíndricos con sección abierta

En los instrumentos de viento con sección abierta, el sonido se produce por el desplazamiento y vibración del aire a través de un cilindro abierto en sus dos extremos. La presión ejercida en una punta del instrumento (al soplar) llegará hasta el final de cilindro. Una parte de esta presión (o energía) se perderá, pero otra parte actuará como un reflejo negativo volviendo hacia el tubo. El sonido se produce cuando esta presión vuelve a la punta del cilindro donde se introdujo el aire. El sonido más grave que puede obtenerse tiene una longitud de onda del doble de la longitud del cilindro.

También se producen armónicos que son submúltiplos de la onda fundamental, formando así el sonido característico de estos instrumentos. Al cerrar los diferentes orificios del cilindro, se va modificando la presión interna, produciendo de esta forma el rango completo de sonidos del instrumento.

Entre los instrumentos de esta familia se encuentran la flauta traversa, el píccolo, la flauta dulce y todas sus variantes.

Instrumentos de viento cilíndricos con sección cerrada

Existen otros instrumentos de viento que poseen un extremo cerrado, pues el aire se introduce por una válvula que solo lo deja pasar en una dirección. La idea básica de como se produce el sonido en estos instrumentos es la misma que en los de sección abierta. La diferencia fundamental es que en estos instrumentos, al tener su extremo cerrado, provocan que la presión o energía "rebote" en el extremo cerrado, provocando un nuevo viaje hacia el otro extremo del cilindro, y otro regreso a lo largo de éste. De esta forma, el instrumento produce un sonido cuya longitud de onda es 4 veces la longitud del cilindro, logrando así un sonido más grave con la misma longitud de instrumento que uno de sección abierta. Debido a este cambio de longitud de la onda producida, la secuencia de armónicos generada por estos instrumentos difieren de los de sección abierta, a pesar de tener la misma longitud física, dándoles su sonido característico, ya que solo forman su sonido con los armónicos impares de la serie, no generando los armónicos pares (es decir los que vibran a 2, 4, 6 veces de la longitud de onda del instrumento).

Entre los instrumentos de este grupo se encuentra, por ejemplo, el clarinete.

Instrumentos de viento cónicos

Otros instrumentos están formados por un tubo cónico, uno de cuyos extremos se encuentra cerrado por una válvula, ya sea del instrumento o los labios del ejecutante. A través de este cambio de forma y del uso de una doble válvula se consigue un instrumento que tiene las características de los instrumentos de sección cerrada, en cuanto a producir sonidos con mayor longitud de onda que el tamaño del instrumento, pero que, a diferencia de estos, forma sus sonidos con toda la secuencia de armónicos sin "filtrar" los armónicos pares.

Entre estos instrumentos se encuentran el oboe, el fagot, la trompeta, el trombón y el corno.

Instrumentos de percusión

Los instrumentos percusivos se diferencian de los anteriores en que algunos de los componentes que forman su sonido pueden ser inarmónicos, es decir componentes que no son múltiplos enteros de la frecuencia fundamental, y además pueden incluir componentes de ruido, no presentes en otros tipos de instrumentos. Estos componentes dependen del material y tamaño de los materiales utilizados para construir los instrumentos.

Entre estos instrumentos podemos encontrar la marimba, el tambor y el bombo.

Resumen de Características de los Instrumentos

En la tabla que se encuentra a continuación podemos observar las diferentes características de los instrumentos musicales, basándose en su construcción distintiva y a su mecanismo de generación del sonido.

Instrumento	Tipo	SubTipo	Pieza de soplido	Caña	Llaves	Sordina	Baqueta	Cuerpo
Trompeta	Viento	Metal	Taza	No	Si	No		
Trompeta Muted	Viento	Metal	Taza	No	Si	Si		
Trombón	Viento	Metal	Taza	No	No	No		
Trombón Muted	Viento	Metal	Taza	No	No	Si		
Corneta	Viento	Metal	Taza	No	Si	No		
Corneta Muted	Viento	Metal	Taza	No	Si	Si		
Corno Francés	Viento	Metal	Taza	No	Si	No		
Corno Francés Muted	Viento	Metal	Taza	No	Si	Si		
Tuba	Viento	Metal	Taza	No	Si			
basson (Fagot)	Viento	Madera	Boquilla	Doble	Si			
Corno Ingles	Viento	Madera	Boquilla	Doble	Si			
Oboe	Viento	Madera	Boquilla	Doble	Si			
Flauta Traversa	Viento	Metal (Madera)	Agujero	No	Si			
Clarinete	Viento	Madera	Boquilla	Simple	Si			
Saxo Tenor	Viento	Metal	Boquilla	Simple	Si			
Violín	Cuerda	de arco						Hueco
Viola	Cuerda	de arco						Hueco
Cello	Cuerda	de arco						Hueco
Contrabajo	Cuerda	de arco						Hueco
Arpa	Cuerda	punteado						Hueco
Guitarra Eléctrica	Cuerda	punteado						Sólido
Bajo Eléctrico	Cuerda	punteado						Sólido
Piano	Cuerda	golpe en cuerdas metálicas						
Vibráfono HM	Percusivo	Metal					Dura	
Vibráfono SM	Percusivo	Metal					Suave	
Xilofón	Percusivo	Madera					Dura	
Celesta	Percusivo	Metal					Dura	
Tubular Bells	Percusivo	Metal					Mazo	
Marimba	Percusivo	Madera					Dura	

Tabla 2-1 - Características de los Instrumentos

Terminología

Generalmente se utilizan algunos conceptos básicos para definir dimensiones de los sonidos. Podemos nombrar algunas, por ejemplo: pitch (frecuencia fundamental, altura o tono), loudness (amplitud de la onda, intensidad), color tonal (espectro de la onda) y articulación (evolución de la onda en el tiempo).

El *tono* de un sonido es la frecuencia de su onda fundamental, es decir la frecuencia de la onda sinusoidal pura sin incluir los armónicos superiores.

La *intensidad* de un sonido es la amplitud de la onda y da una idea de la potencia de la señal.

El *espectro* de un sonido es su representación de la amplitud en función de la frecuencia, y permite analizar un sonido sobre la base de sus distintos componentes de frecuencia y sus respectivas intensidades (por ejemplo Análisis de Fourier).

La articulación es la evolución de la amplitud de la onda en función del tiempo, y nos permite ver las diferentes fases de un sonido. Es a través de esta representación donde se puede llegar a evaluar un sonido en attack, decay, sustain y release (ataque, decaimiento, sostén y liberación). El ataque es la fase inicial de un sonido, quizás la más compleja y variable, luego viene una fase llamada decay donde la intensidad y la complejidad disminuyen rápidamente, hasta alcanzar un estado relativamente estable, llamado sustain (o también steady-state). Luego al interrumpirse el estímulo que produce al sonido, viene la fase de release, donde el sonido desaparece, lenta o inmediatamente, de acuerdo a la forma de ejecución.

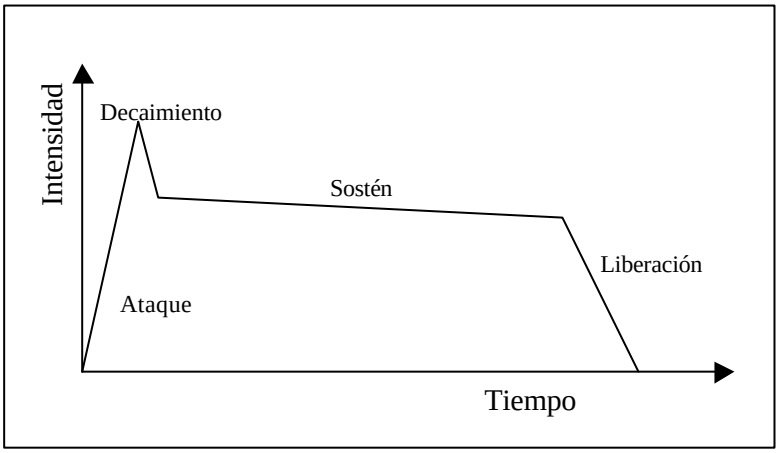


Figura 2-1 – Evolución en intensidad de un sonido

Entre las características de los sonidos también son importantes los *formantes*. Un formante es una región de frecuencias que resuenan con diferente intensidad a lo largo de la duración del sonido, y dependiendo de cada instrumento. La estructura y movimiento de estos formantes se conservan prácticamente a lo largo de toda la gama de sonidos de un instrumento, ayudando a darle su color o timbre particular.

2.2. Kohonen

Los mapas de Kohonen [KOH/1988] son redes neuronales no supervisadas auto-organizativas (Self-Organizing Maps - SOM), en el cual el algoritmo subyacente busca “clusters” en los datos.

Esta categoría de redes no supervisadas, en las cuales están Kohonen y Análisis de Componente Principal, le permite al investigador agrupar objetos, sobre la base de la cercanía percibida en un espacio n-dimensional, donde n es el número de variables u observaciones realizadas en cada objeto.

Descripción Detallada

Es una red neuronal no supervisada con aprendizaje competitivo entre las neuronas y sin capas ocultas.

La **entrada** a Kohonen usualmente es un conjunto de vectores de longitud N (posiciones ξ_1 a ξ_N) de valor continuo (reales), que definen un punto en un espacio n-dimensional

Las unidades **de salida** están estructuradas en un array (generalmente de una o dos dimensiones), y están totalmente conectadas con todas las entradas (ξ_1 a ξ_N) vía la matriz de pesos w_{ij} (Ver Figura 2-2 – Red de Kohonen)

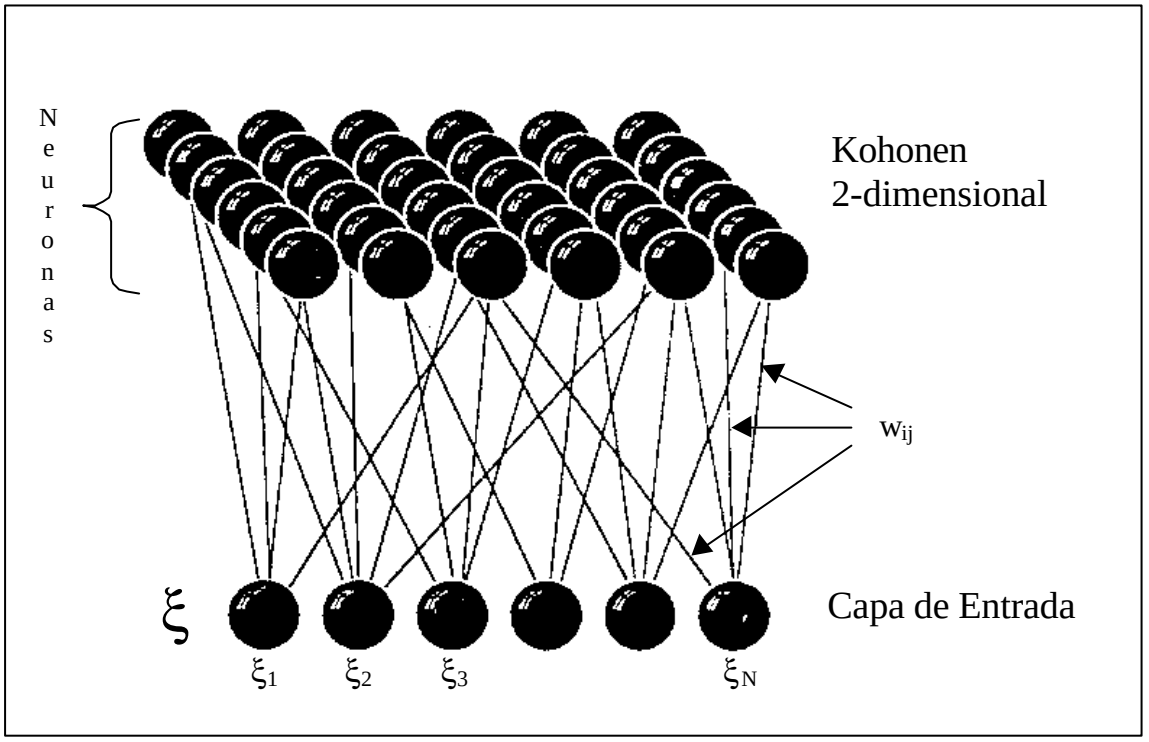


Figura 2-2 – Red de Kohonen

Ejemplificación:

Si las unidades de salidas (neuronas) están agrupadas en un esquema de dos dimensiones, digamos 10x10, y la dimensión de los vectores de entrada es 6, entonces cada una de las 100 neuronas de salida está conectada con las 6 posiciones de los vectores de entrada. Se puede ver como que cada neurona tiene 6 conexiones de entrada. Matemáticamente hablando, todos los pesos se pueden representar como una matriz tridimensional de 10x10x6.

La matriz de pesos w en la posición [0,0] estaría referenciando a una neurona particular, y en esa posición de la matriz tendríamos un vector de 6 posiciones.

Durante la fase de entrenamiento de la red una regla competitiva de aprendizaje es utilizada, eligiendo la neurona ganadora *i**, como la neurona cuyo vector de pesos esté más cercano al vector de entrada.

$$|w_{i*} - \xi| \leq |w_i - \xi| \quad \text{(para toda neurona i)} \qquad \text{(Ecuación 2-1)}$$

Notar que el subíndice i se refiere a una neurona de la Red de Kohonen y que su forma varía según la estructura topológica de interconexión de las neuronas. Ej. : i puede ser de la forma [x] ó [x,y] ó incluso [x,y,z] para una estructura tridimensional.

Una vez obtenida la neurona ganadora, se aplica la regla de aprendizaje de Kohonen que es ([KOH/1982], [KOH/1988])

$$\Delta w_{ij} = \eta \Lambda(i,i^*) (\xi_j - w_{ij}) \qquad \text{(Ecuación 2-2)}$$

para todos los i y los j (neuronas y posiciones del vector de entrada respectivamente). La **función de vecindad** Λ es 1 para i = i* y decrece conforme i se aleja de la neurona ganadora i* en el arreglo de salida. De esta manera, las neuronas cercanas a la ganadora, y también la ganadora, tienen sus pesos cambiados en forma apreciable, mientras que aquellas más alejadas, donde $\Lambda(i,i^*)$ es pequeño, experimentan poco o ningún cambio. Aquí es donde la información topológica es provista, pues neuronas cercanas reciben actualizaciones parecidas y terminan respondiendo a entradas parecidas.

La regla (Ecuación 2-2) arrastra el vector de peso de la neurona ganadora w_{i^*} hacia la entrada ξ. Pero también arrastra los w_i cercanos. Por lo tanto, podemos pensarlo como una suerte de red elástica en el espacio de entrada que quiere asemejarse lo más posible a las entradas. La red elástica tiene la topología del array de salida (en nuestro ejemplo 10x10) y cada punto de la red (en nuestro caso 100) tiene como coordenada, en el espacio de entrada, el peso de la neurona correspondiente. Es bueno tener esta representación en mente al intentar graficar las redes de Kohonen.

Durante el periodo de aprendizaje de la Red de Kohonen, existen dos fases bien diferenciadas. La primera es la fase de Ordenamiento, en donde se busca lograr una organización general de los vectores de entrada, de forma que se agrupen por su similitud. La segunda es la fase de Convergencia, y en esta fase se busca que las posiciones de la matriz que se parecen más a un vector de entrada, se parezcan aún más.

La diferencia fundamental es que en la fase de Ordenamiento el aprendizaje afecta siempre a un conjunto de neuronas (dependiendo de la vecindad), mientras que en la fase de Convergencia las neuronas aprenden individualmente. En forma figurada, podríamos decir que la fase de Ordenamiento realiza el trabajo “bruto”, mientras que la fase de Convergencia realiza el “fino”.

Se pueden probar mapas de dos dimensiones a una dimensión, a pesar de la imposibilidad de preservar toda la topología. Mapas de una dimensión a una dimensión también son posibles, y mucho del trabajo de análisis teórico ha sido para este caso.

Aplicaciones para “feature mapping” han sido hechas en muchas áreas, incluyendo “sensory mapping”, control de motores, reconocimiento de voz, “vector quantization” y optimización combinatoria.

Para profundizar en la teoría de Redes Neuronales se puede leer "Introduction to the theory of neural computation" de Hertz, Krogh y Palmer [HER/1991b].

Para profundizar en Redes Neuronales No Supervisadas Auto-Organizativas se puede leer “Self Organization and Associative Memory” de T. Kohonen [KOH/1988].

2.3. Wavelets

La Transformada de Wavelets es una técnica de análisis de señales, como la Transformada de Fourier, pero a diferencia de ésta, tiene una ventana de análisis variable.

En esencia la transformada de Wavelet opera una transformación de la señal en un nuevo espacio, en donde ciertas características (ej. : la evolución de las distintas frecuencias) son más evidentes.

En particular la Transformada de Wavelet es de interés en el análisis de señales no estacionarias (como el ataque de un instrumento musical), en donde la Short-Time Fourier Transform (STFT) es menos óptima.

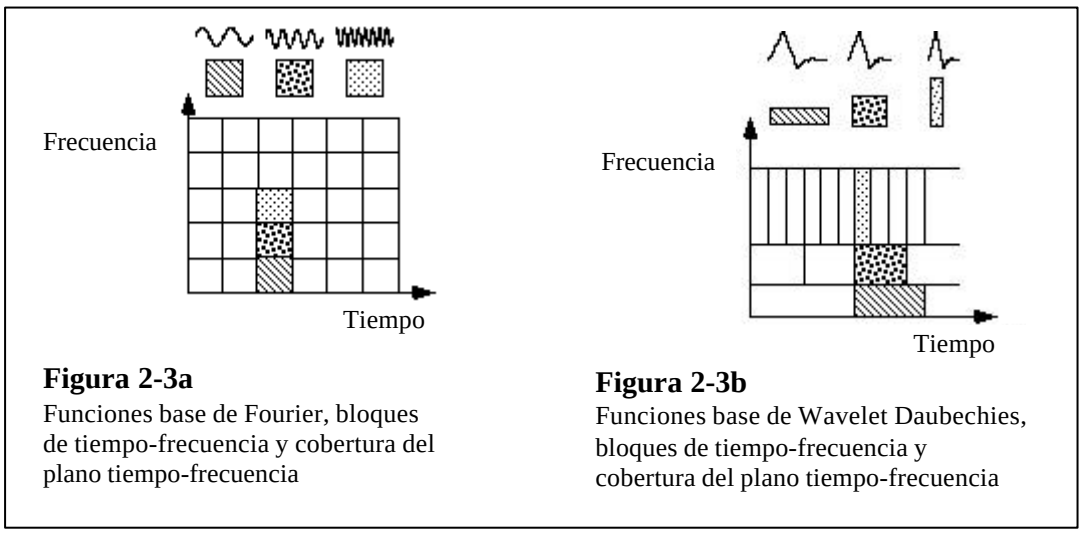


Figura 2-3a
Funciones base de Fourier, bloques de tiempo-frecuencia y cobertura del plano tiempo-frecuencia

Figura 2-3b
Funciones base de Wavelet Daubechies, bloques de tiempo-frecuencia y cobertura del plano tiempo-frecuencia

Figura 2-3 - Wavelets

La principal diferencia con STFT es que ésta usa un único tamaño de ventana de análisis (que puede ser visto como filtros de banda), mientras que TW usa ventanas más cortas en tiempo, a altas frecuencias (análisis más detallado) y ventanas más grandes a bajas frecuencias.

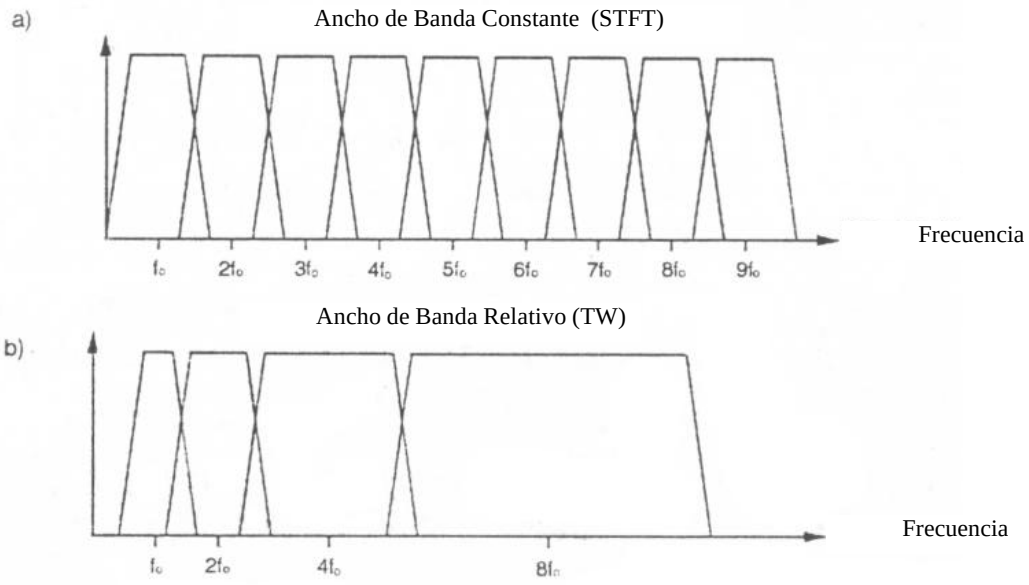
En la Figura 2-3a se puede apreciar las ventanas de análisis (bloques) que utiliza Fourier. Notar que todas tienen el mismo tamaño (en tiempo y frecuencias analizados).

En la Figura 2-3b podemos apreciar como las ventanas de la TW son más cortas en tiempo a medida que crece la frecuencia (análisis más detallado) y que a frecuencias más altas se cubren más frecuencias en la misma ventana. Cada bloque representa (tanto en a como en b) la concentración esencial en el plano tiempo-frecuencia de una función base determinada.

La Transformación de Wavelet se puede ver como una descomposición de la señal sobre un conjunto de funciones básicas. Todas estas funciones básicas se obtienen a partir de una única wavelet mediante dilataciones y contracciones (escalamientos), como también de corrimientos (shifts). La wavelet base puede ser vista como un filtro pasabanda en donde solo pasan ciertas frecuencias. (En la Figura 2-5 se puede ver un ejemplo de cómo se ve una Wavelet particular)

Al mirar las wavelets como bancos de filtros, se los puede ver como una serie de filtros en donde se mantiene constante la siguiente relación (al contrario que STFT)

$$A = \frac{\text{Rango de Frecuencia del filtro}}{\text{Frecuencia Central a Analizar}}$$



Tomando STFT y TW como filtros, se puede ver en a) que Fourier analiza con un ancho constante, mientras que Wavelet en b), toma mayores ventanas a mayores frecuencias.

Figura 2-4 – Ventana de Análisis de STFT y TW

Este tipo de filtros se usa para modelar la respuesta en frecuencia de la cóclea situada en el oído interno y está, por lo tanto, adaptada a la percepción auditiva, ej. : sonidos de instrumentos musicales. Los filtros que la satisfacen están naturalmente distribuidos en octavas.

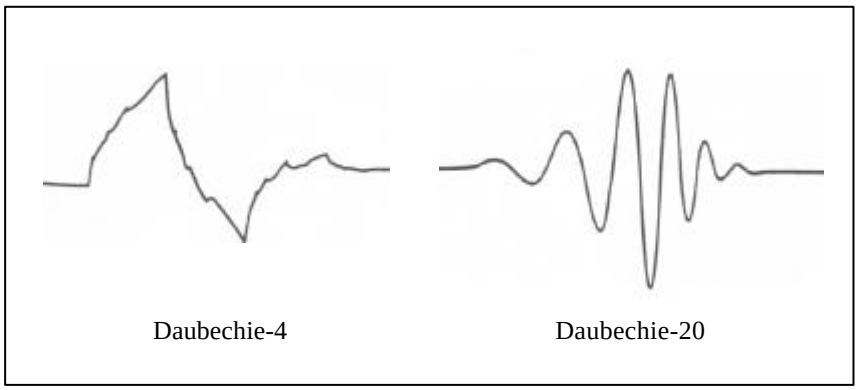
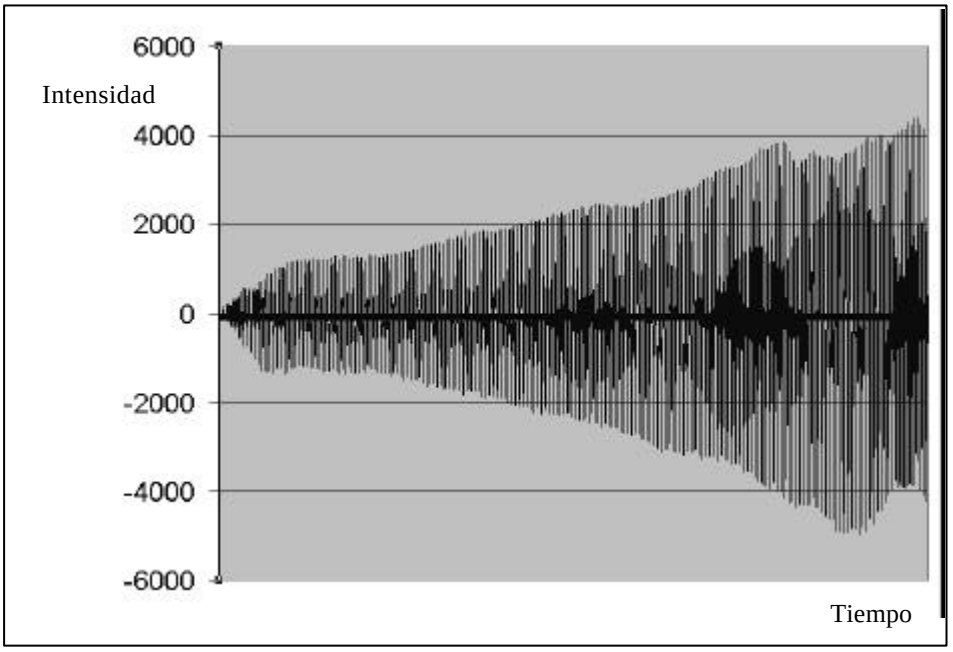


Figura 2-5 - Distintas wavelets base para realizar las transformaciones

El resultado del análisis de wavelet es un conjunto de coeficientes de wavelets que indican que tan cerca esta la señal a la función wavelet base.

Figura 2-6 - 32768 valores iniciales del sonido C4 del Violín (743 ms)



Aunque la teoría de Wavelet tiene un sostén matemático complejo, su Transformada es fácil de instrumentar y alcanza velocidades de ejecución comparables a las obtenidas con otras transformadas, con mayor ahorro de memoria.

Transformada Discreta de Wavelets

La Transformada Discreta de Wavelets (DWT) es una operación lineal rápida que opera en un vector de datos cuya longitud es un entero potencia de dos, transformándolo en un vector numéricamente diferente de la misma longitud.

La transformada de wavelets utiliza unas funciones básicas que tienen el nombre de “mother functions” o “wavelets” (ver Figura 2-3), a diferencia de la STFT, que utiliza senos y cosenos.

Lo que hace interesante a las funciones básicas de wavelets es que, al contrario de senos y cosenos, las funciones individuales de wavelets están bastante localizadas en el espacio; simultáneamente, al igual que senos y cosenos, las funciones individuales de wavelets están bastante localizadas en frecuencia.

Al contrario que senos y cosenos, que definen una única transformada de Fourier, no existe un único conjunto de wavelets, de hecho hay infinitos conjuntos. Básicamente, los diferentes conjuntos hacen diferentes compromisos entre que tan compactamente están localizados en el espacio y cuan suaves son.

En Figura 2-6, Figura 2-7 y Figura 2-8 se puede ver sucesivamente, la señal de un instrumento musical (Violín), su transformada Wavelet y otra visualización de la misma transformada.

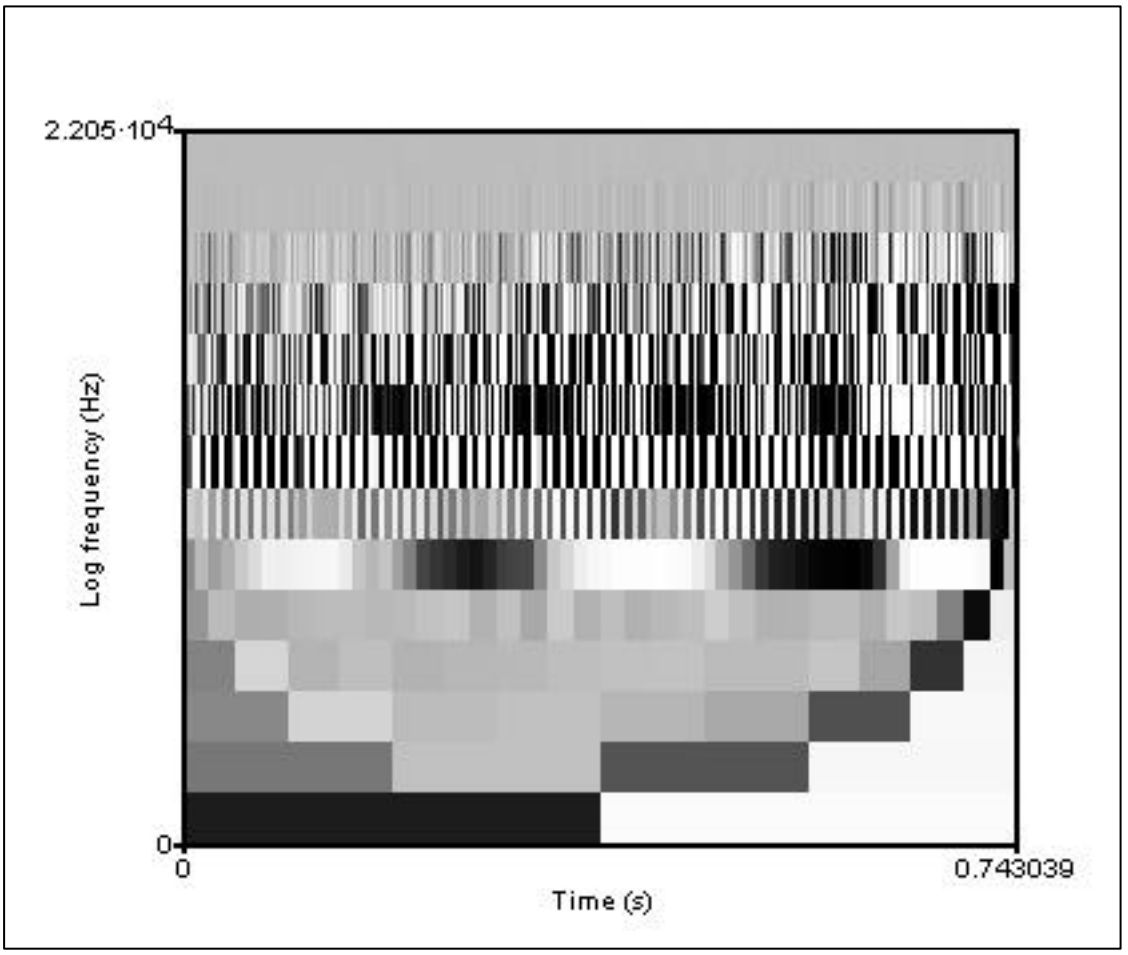


Figura 2-7 - Representación tradicional del Wavelet del Violín de la Figura 2-6

Si la señal de entrada es de tamaño 1024 y la frecuencia máxima es f, entonces la transformada de Wavelet (discreta) genera los siguientes vectores como resultado (el contenido de los vectores son coeficientes de wavelets)

	Cantidad de valores	Rango de frecuencia analizado
Vector 10	512	f/2 hasta f
Vector 9	256	f/4 hasta f/2
Vector 3	4	f/512 hasta f/256
Vector 2	2	0 hasta f/512
Vector 1	2	coeficientes madre

Tabla 2-2

Cada uno de ellos representa la evolución temporal de la señal original en un rango de frecuencia determinado

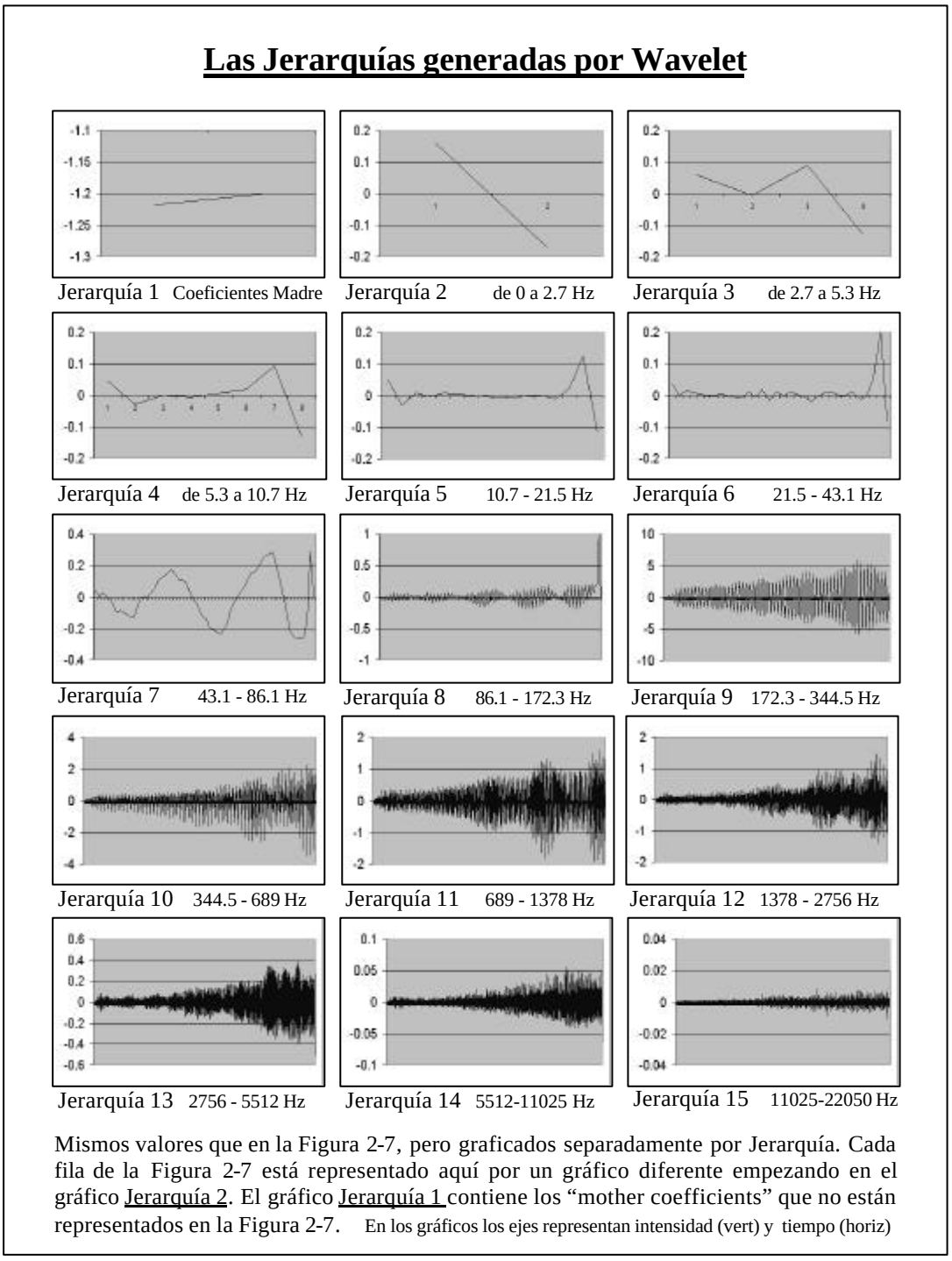


Figura 2-8 - Representación alternativa del Wavelet del Violín de la Figura 2-6

Para leer un poco más de Wavelets se puede leer "Wavelets and signal processing", de Olivier Rioul y Martin Vetterli [RIO/1991], o también se puede consultar el sitio http://www.amara.com/current/wavebiblio.html, donde hay una gran cantidad de Bibliografía para los interesados en conseguir material (libros y links).

Capítulo 3 - Descripción del Modelo

En este capítulo detallamos el modelo que proponemos y hemos implementado, y también todo lo relacionado a los sonidos utilizados y su pre-procesamiento. Comenzaremos dando una visión global del modelo.

3.1. Breve descripción del modelo

El modelo desarrollado está basado en la descomposición de los sonidos musicales en octavas. A cada una de las partes de esta descomposición la llamamos *jerarquía*, y cada conjunto de jerarquías equivalentes, de todos los sonidos, es utilizado para entrenar un Mapa Auto-organizativo de Kohonen [KOH/1988]. Es decir, supongamos que tenemos una jerarquía que contiene frecuencias entre 172 y 345 Hz, entonces de cada sonido tomaremos esta jerarquía y con este nuevo conjunto de elementos entrenaremos el mapa correspondiente. Queda claro que tendremos un mapa por cada jerarquía en la que se hayan dividido los sonidos. Basándose en donde se ubica cada sonido en los distintos mapas, luego se entrenará otro mapa o red, que es el que nos brindará la información que buscamos.

Para poder generar el mapa topológico de distribución de instrumentos (resultado final), el modelo pasa por una etapa de aprendizaje, después de la cual ya es posible observar la distribución u organización de los sonidos o, si se quiere, consultarlo con relación a nuevos instrumentos (típicamente no usados en el periodo de entrenamiento) para determinar el parecido con los instrumentos existentes.

Esquema

La siguiente es una descripción esquemática del modelo que hemos desarrollado e implementado, separado en sus tres bloques: entrada (alimentación del sistema), procesamiento y salida (organización obtenida).

Entrada:

Sonidos aislados, es decir notas individuales, provenientes de instrumentos musicales y producidos en la nota C4 (262 Hz, DO de la cuarta octava).

Procesamiento:

Cada sonido (tomado como un vector de muestras) es pre-procesado antes de ingresar al entrenamiento neuronal. Este pre-procesamiento (Paso 1 de **Figura 3-1**) tiene como resultado final N vectores (Paso 2), en donde cada uno representa la evolución temporal del sonido en una determinada banda de frecuencias (en realidad son N-1 bandas, pero este concepto será explicado en detalle en el item 0 dentro de Procesamiento del sonido en la página 32).

Cada uno de estos N vectores alimenta N redes neuronales diferentes (Paso 3 de **Figura 3-1**).

Después que las redes inferiores se entrenaron, cada vector obtiene una posición final (X,Y) para cada red Neuronal (Paso 4).

Las N posiciones (X,Y) combinadas "representan" al sonido original. Una vez realizado esto con todos los sonidos del conjunto de entrenamiento, estas N posiciones combinadas alimentan otro Mapa de Kohonen (red de Kohonen Superior – Paso 5 de **Figura 3-1**), para obtener como resultado final una distribución espacial en 2D de la relación entre todos los sonidos.

Salida / Resultado Final:

Mapa topológico en 2 dimensiones, que contiene la distribución de todos los instrumentos de entrada, agrupados por similitud o características comunes.

Permite también analizar instrumentos nuevos y determinar su parecido con los instrumentos utilizados en el aprendizaje

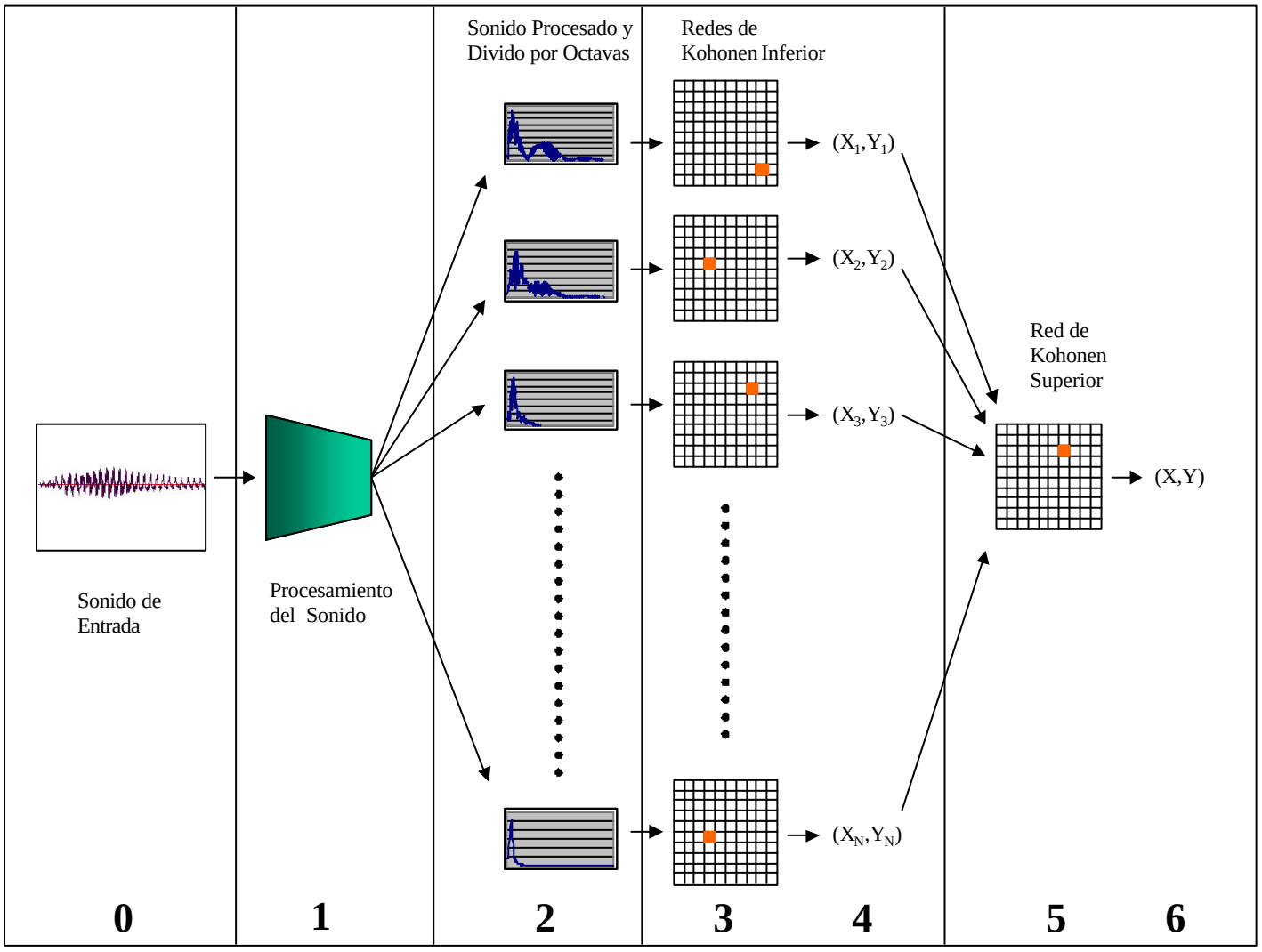


Figura 3-1 – Arquitectura del Modelo

3.2. Sonidos

Sonidos elegidos para el entrenamiento (por qué los MUMs)

Existen varios motivos por los cuales se eligió el conjunto de sonidos de la McGill University (MUMs, McGill University Master Samples) [OPO/1987] para estas investigaciones, entre los más importantes podemos destacar:

a) Los MUMs tienen prácticamente todo el espectro de notas para cada instrumento, lo cual permitiría expandir el presente modelo sin cambiar de fuente;

b) Son grabaciones de instrumentos reales (y no sintetizados), lo cual asegura una mayor fidelidad de las entradas utilizadas; y

c) Se han grabado en forma profesional y con todas las condiciones técnicas necesarias para registrar fielmente los instrumentos.

También es importante el hecho que varios trabajos de investigación lo usan como fuente de sonidos para sus análisis ([MAR/1998a], [MAR/1998b], [PRA/1994], [DEP/1997], [FRA/1998] y otros), con lo cual estamos trabajando con la misma base utilizada por otros investigadores (y facilitaría, para un futuro próximo, una comparación más directa de los resultados obtenidos).

Para tener un mayor conocimiento de los instrumentos involucrados (material de construcción, forma de uso, familias, etc.), sugerimos leer el Capítulo 2 – Conceptos Básicos.

Para nuestros experimentos utilizamos un conjunto de 22 instrumentos, básicamente instrumentos típicos de una orquesta, además de algunos más modernos como el bajo o la guitarra eléctrica.

3.3. Procesamiento del sonido

En los mapas de Kohonen es sumamente importante tener una buena representación de los elementos de entrada, y es por eso que una buena (o mala) representación, puede ser la diferencia entre una red de Kohonen que se organice y otra que no [KOH/1982].

Debido a la naturaleza de este tipo de redes, y de que le cuesta “aprender” si las entradas tienen muchas dimensiones, es que procesamos los sonidos de entrada con el objetivo de reducir la complejidad del problema.

Solo a modo de ejemplo, para sonidos grabados con un muestreo de 44Khz en modo estéreo (como son los MUMs), la cantidad de muestras varía entre 80.000 y 1.200.000 muestras por sonido (que son entre 1 y 14 segundos aproximadamente)

El procesamiento de los sonidos incluye los siguientes 5 pasos:

¿Sonido Mono, Estéreo, Envolvente o Cuadrafónico?

Los sonidos incluidos en los CDs MUMs, están grabados en modo estéreo, por lo que la primera determinación que se tomó fue pasar esta información a mono, transformación que no afecta a nuestros experimentos, ya que un instrumento musical es básicamente monoaural, y además nos permite reducir la cantidad de muestras a procesar a la mitad.

En cada uno de los siguientes pasos mostraremos los resultados en forma gráfica con el violín, para poder tener una mejor visión de los diferentes procesos de transformación involucrados.

Cantidad de sonido a Procesar.

El segundo tema a analizar es:

¿Cuánto se va a procesar del sonido?
¿Tomamos el sonido completo o solo una parte?
¿Si es solo una parte, que parte?
¿El ataque (comienzo), el decaimiento, la fase estable (el medio), o la liberación (final del sonido)?

Luego de analizar cuidadosamente estos puntos se decidió trabajar tomando aproximadamente 743 milisegundos desde el inicio del sonido, fundamentándonos principalmente en las siguientes razones:

a) En muchos de los trabajos consultados, los autores coinciden en que el ataque es uno de los momentos más importantes para poder identificar un sonido;

b) Todos los sonidos utilizados en este trabajo tienen un ataque menor a 500 milisegundos. [MAR/1998b]. Por ejemplo, uno de los sonidos que más demora en alcanzar su fase estable es el violín, casi 500 ms, debido al estilo de ejecución utilizado en esta grabación. Podría haberse utilizado algún estilo de ejecución diferente cuya fase de ataque fuera mucho más corta;

c) Queremos no solo tomar el ataque, sino también incluir la fase estable de sonido (steady state). Varios autores inclusive creen que la fase estable puede ser suficiente para clasificar el sonido ([GRE/1975] y [FUJ/1998]). El ataque y la fase estable son estructuralmente diferentes, por lo cual ambos proveen información importante sobre el sonido, además de que la relación de duración entre el ataque y la fase estable también permite identificar a los sonidos;

d) Con 743 milisegundos obtenemos 32768 muestras y para el procesamiento en el item 0 (Wavelets) precisamos 2^N muestras, donde N tiene el mismo significado que veníamos usando hasta ahora, o sea, la cantidad de partes u octavas en que se divide el sonido. El utilizar 32768 muestras nos permite considerar luego 15 jerarquías (ya que 2^{15} = 32768). De estas 15 jerarquías, la primera estará conformada por coeficientes madre de wavelets, y las 14 restantes contendrán información representativa del contenido y evolución espectral de los sonidos en cada banda de frecuencias.

Como resultado de esta elección nuestro modelo está analizando el ataque y parte de la fase estable de los instrumentos, ver Figura 3-2.

Estrictamente hablando, y dependiendo del sonido, puede que inclusive en algunos sonidos estemos incluyendo parte o toda la fase de liberación (en general, los sonidos percusivos son mas bien cortos).

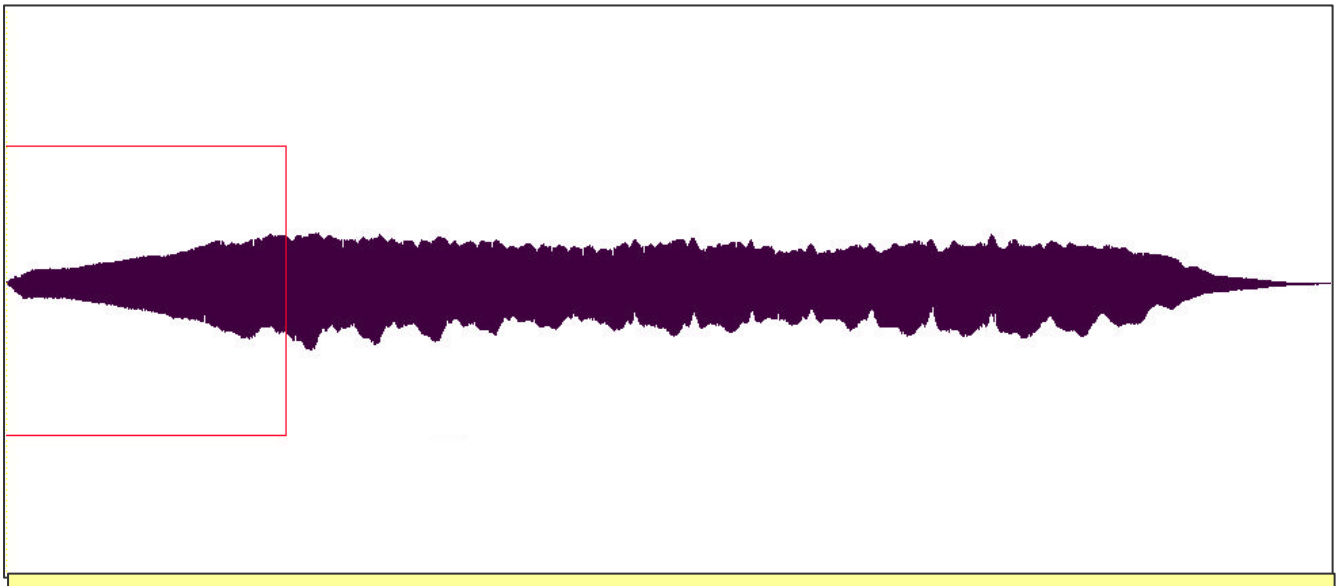


Figura 3-2 – Gráfico del Violín

¡Bajen el Volumen!

En este paso `normalizamos` la señal, para que la intensidad (volumen) de todos los sonidos sea la misma.

Para ello el valor más alto dentro de cada sonido es llevado a 1 y el resto modificado de manera proporcional. De esta forma, la intensidad máxima de todos los sonidos es la misma.

Un Kilo de Sonido bien trozado, por favor.

En este paso aplicamos la transformación de Wavelets a cada sonido (si no esta familiarizado con el procesamiento con Wavelets sugerimos leer el Capítulo 2 – Conceptos Básicos, antes de seguir), lo que da como resultado la descomposición del sonido en 15 partes. La cantidad de partes en que es descompuesto el sonido viene determinado por el tamaño de la muestra ($2^{15} = 32768$).

Usando la terminología de Wavelets, a cada una de estas 15 partes la llamaremos Jerarquía.

En la Jerarquía Nro. 1 están los coeficientes madre de Wavelets (mother-function coefficients) y el resto de las Jerarquías representan la evolución temporal del sonido en las diferentes octavas.

Estrictamente hablando, las Jerarquías no representan octavas musicales sino rangos de frecuencias determinados (Ver Tabla 3-1 – Página 36) que tienen la misma estructura de rango de frecuencias que las octavas musicales, pues cada jerarquía es el doble de la anterior; por eso es que las mencionamos como octavas. Cada Jerarquía corresponde a la octava que va aproximadamente de la nota FA al MI siguiente.

Nota: Dentro de las posibles funciones de Wavelet usamos las Wavelets de Daubechies de 12 coeficientes, pues son de las más utilizadas y tienen un compromiso medio entre el nivel de calidad y la complejidad para generar las Wavelets.

La Figura 3-3 muestra esta descomposición.

Es hora de achicar el sonido

Debido a la naturaleza de Wavelets, los vectores para las octavas más elevadas tienen una gran cantidad de valores (en nuestro caso, por ejemplo, hay vectores de 512, 1024, 2048, 4096, 8192 y 16384 valores), por eso es que se decidió comprimirlos, para que puedan ser manejados mejor por Kohonen (con redes de Kohonen que tienen entradas de grandes dimensiones no se obtiene un buen grado de convergencia)

Después de analizar varios métodos de compresión, optamos por usar el promedio en valor absoluto de los valores. Este método toma el valor absoluto de todos los valores del vector y luego promedia de a M valores, donde M es de la forma 2^m (ver ejemplo más adelante).

También se analizó, entre otras posibilidades, calcular solamente el promedio, pero fue descartado por análisis de los resultados obtenidos, pues para las jerarquías superiores (mayores a 10), los datos generados por Wavelet en base a las señales musicales eran generalmente una secuencia de números positivos y negativos alternados, haciendo que el promedio anulara estos valores.

A modo de ejemplo del método elegido, si estamos reduciendo todas las jerarquías a 256 valores, entonces la jerarquía 10 (512 valores) es reducida en relación 2 a 1, la jerarquía 11 (1024 valores) 4 a 1, y así siguiendo hasta la jerarquía 15 (16484 valores), que es reducida en relación de 64 a 1. En nuestro modelo, cada uno de los 256 valores finales de cada jerarquía representa siempre 2.9 ms del sonido original (recordemos que hemos tomado aproximadamente los 743 ms iniciales de cada sonido).

Como se puede observar en la Figura 3-4, la intensidad del sonido se mantiene y la principal diferencia que se aprecia es debido al comportamiento natural de la función promedio que tiende a nivelar los extremos (picos y valles)

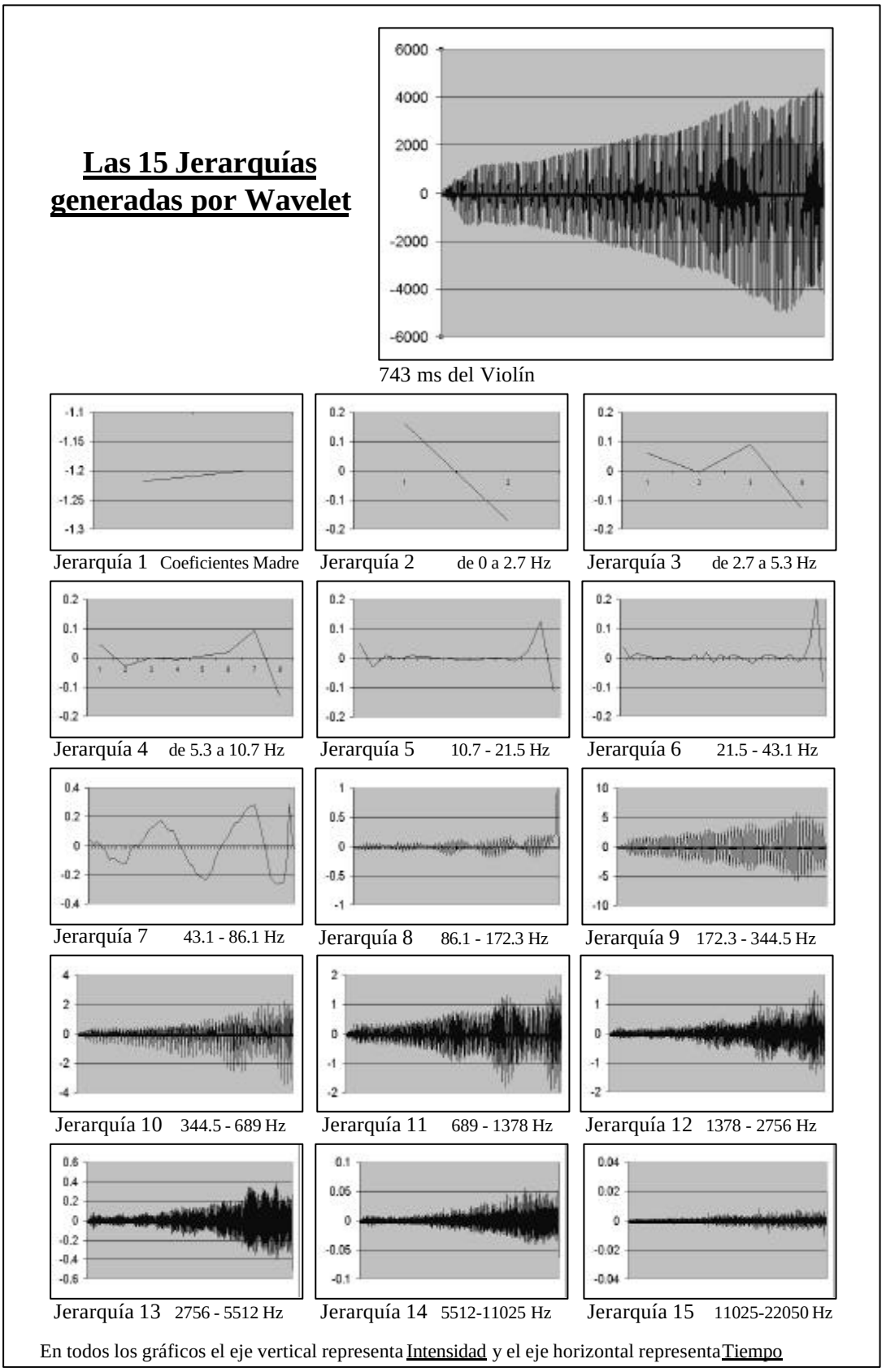


Figura 3-3 – Las 15 Jerarquías generadas por Wavelet

Gráfico 3

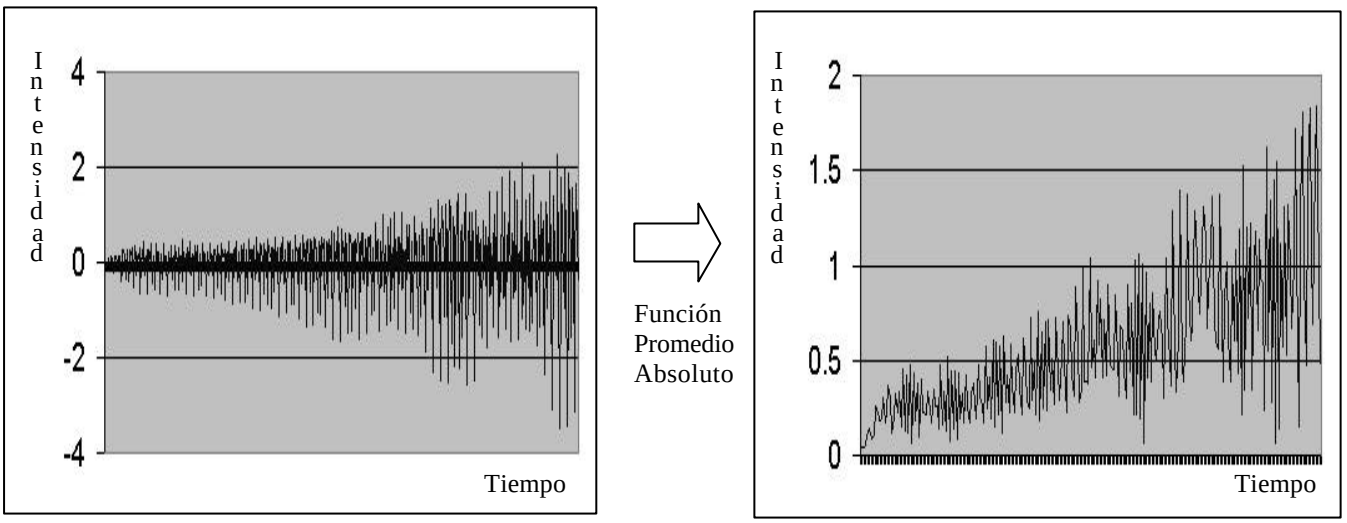


Figura 3-4 – Función de Compresión

3.4. Mapas Inferiores de Kohonen

Una vez que los sonidos han sido procesados, los resultados son utilizados para entrenar a las redes de Kohonen (Paso 3 y 4 de Figura 3-1 en Página 29).

Como soslayamos en la introducción al modelo, contamos con N redes de Kohonen y cada una de ellas será entrenada con los elementos provenientes de cada Jerarquía, permitiéndole analizar la similitud de los diferentes sonidos según el rango de frecuencia que representen los vectores de entrada (ver Tabla 3-1 - Más adelante).

De esta manera, por ejemplo, el mapa de Kohonen que analice la Jerarquía 10, estará analizando el comportamiento evolutivo de todos los sonidos de entrada entre las frecuencias 344.5 Hz y 689 Hz.

Nota I: usamos indistintamente los términos “Mapas de Kohonen” y “Redes de Kohonen”, pues por un lado son Redes Neuronales y por otro, la bibliografía se refiere a ellos como Mapas Auto-organizativos de Kohonen (SOM - Self-Organizing Maps).

Nota II: Para familiarizarse con la teoría de los Mapas Auto-organizativos de Kohonen, sugerimos leer el Capítulo 2 – Conceptos Básicos.

Con respecto al tamaño de los mapas de Kohonen, tomamos como primera aproximación una red de 10x10, para tener al menos una relación de 1 a 5 entre la cantidad de sonidos de entrenamiento y el espacio de salida de la red neuronal, tomando en cuenta un conjunto de entrenamiento de 22 sonidos.

El objetivo de utilizar los Mapas Auto-organizativos de Kohonen, es para que la propia red neuronal agrupe todos los elementos de entrada sin intervención externa (para evitar el factor subjetivo a la hora de agrupar los instrumentos, es decir, se utiliza aprendizaje no supervisado).

Las Redes de Kohonen tienen la particularidad de agrupar los componentes “más parecidos”, o características comunes; en nuestro modelo este “parecido”, está dado por la función Distancia Euclidiana (norma).

Jerarquía	Cantidad de Elementos generados por Wavelet	Frecuencias Que comprende (en Hz)	Frecuencia central (en Hz)
1	2	Coeficientes Madre	-
2	2	0 a 2.7	1.3
3	4	2.7 a 5.3	4
4	8	5.3 a 10.7	8
5	16	10.7 a 21.5	16.1
6	32	21.5 a 43.1	32.3
7	64	43.1 a 86.1	64.6
8	128	86.1 a 172.3	129.2
9	256	172.3 a 344.5	258.4
10	512	344.5 a 689	516.7
11	1024	689 a 1378.1	1033.5
12	2048	1378.1 a 2756.2	2067.5
13	4096	2756.2 a 5512.5	4134.3
14	8192	5512.5 a 11025	8268.7
15	16384	11025 a 22050	16537.5

Nota: el DO C4 tiene una frecuencia central de 262 Hz

Tabla 3-1 – Rangos de Frecuencia de Wavelet

Una vez que todas las redes Inferiores de Kohonen han sido entrenadas, se hace pasar todos los elementos del conjunto de entrenamiento por las mismas. Esto significa que, por cada jerarquía, se da como entrada cada sonido al mapa de Kohonen y la neurona que es más parecida a ese sonido es activada.

Como resultado de esto, cada jerarquía de cada sonido, resulta en una posición (X,Y) para cada red Inferior de Kohonen.

A modo de ejemplo he aquí una corrida real:

En la jerarquía 1 de la flauta “cae” en la posición (4,0).
En la jerarquía 2 de la flauta “cae” en la posición (5,2)
En la jerarquía 3 de la flauta “cae” en la posición (1,7)
…
…
…
En la jerarquía 15 de la flauta “cae” en la posición (7,3)

Esta línea se interpreta como :
Cuando se da como entrada al sonido “flauta” en el mapa de Kohonen de la Jerarquía 1, se activa la neurona de la posición (4,0)

Entonces, a partir de este momento la flauta estará representada por el vector compuesto por los 15 pares (X,Y), resultantes de su posición en cada una de las 15 Redes Inferiores:

J 1	J 2	J 3	J 4	J 5	J 6	J 7	J 8	J 9	J 10	J 11	J 12	J 13	J 14	J 15
(4,0)	(5,2)	(1,7)	(6,0)	(3,0)	(6,0)	(0,6)	(0,9)	(9,9)	(7,8)	(0,9)	(5,3)	(6,5)	(4,6)	(7,3)

Computacionalmente estos valores quedan representados como un vector de 30 números enteros.

Esta es una posible representación final de todos los sonidos, después de "pasarlos" a través de todas las jerarquías inferiores ya entrenadas; y se resalta la flauta, que es el instrumento que usamos como ejemplo.

Instrumento	**Jerarquía**														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Piano 1	(1,6)	(9,9)	(2,0)	(3,6)	(6,7)	(0,9)	(0,0)	(9,0)	(3,2)	(0,6)	(9,0)	(4,0)	(5,7)	(2,2)	(3,1)
Piano 2	(8,2)	(5,7)	(5,3)	(0,4)	(6,3)	(4,4)	(3,9)	(6,0)	(5,6)	(0,3)	(3,0)	(2,0)	(0,7)	(5,2)	(7,0)
Tubular Bells	(3,4)	(8,2)	(0,2)	(7,5)	(4,5)	(2,3)	(2,5)	(3,8)	(0,9)	(5,0)	(8,3)	(0,3)	(1,5)	(7,0)	(9,2)
Vibráfono HM	(0,9)	(9,4)	(4,1)	(4,4)	(7,5)	(1,5)	(6,0)	(7,9)	(4,4)	(4,1)	(5,0)	(0,0)	(1,9)	(4,0)	(5,0)
Vibráfono SM	(9,8)	(0,5)	(6,8)	(2,0)	(5,0)	(9,3)	(9,3)	(0,3)	(1,0)	(2,2)	(1,0)	(8,2)	(4,9)	(0,0)	(0,1)
Marimba, Sinfónica	(7,4)	(3,2)	(4,5)	(7,3)	(2,5)	(2,1)	(9,6)	(3,2)	(6,0)	(0,0)	(2,1)	(9,0)	(6,9)	(1,1)	(5,2)
Guitarra Eléctrica	(8,9)	(0,9)	(9,7)	(0,0)	(3,7)	(6,5)	(7,2)	(3,5)	(0,5)	(2,5)	(9,5)	(3,2)	(3,6)	(9,2)	(9,5)
Bajo Eléctrico (Deep)	(9,9)	(0,7)	(7,6)	(0,2)	(7,1)	(9,6)	(9,0)	(1,0)	(1,3)	(2,0)	(0,1)	(9,1)	(3,8)	(2,0)	(1,0)
Violín	(6,8)	(0,1)	(3,9)	(4,0)	(2,1)	(7,3)	(5,2)	(1,7)	(7,7)	(8,1)	(0,7)	(9,9)	(9,3)	(0,6)	(0,6)
Viola	(8,7)	(2,8)	(9,4)	(1,7)	(8,7)	(5,7)	(0,3)	(6,6)	(9,6)	(9,0)	(1,5)	(8,7)	(7,2)	(0,9)	(2,7)
Cello	(1,2)	(8,7)	(2,2)	(9,3)	(4,9)	(6,9)	(5,8)	(9,6)	(9,3)	(9,3)	(6,2)	(0,6)	(0,0)	(9,9)	(9,9)
Contrabajo	(2,0)	(5,5)	(6,0)	(6,7)	(0,0)	(0,0)	(0,9)	(9,9)	(6,9)	(9,9)	(2,8)	(6,9)	(9,0)	(9,5)	(6,9)
Trompeta	(3,2)	(4,0)	(0,9)	(4,9)	(0,6)	(4,0)	(4,6)	(2,9)	(0,8)	(7,0)	(3,3)	(6,6)	(3,0)	(6,5)	(4,4)
Trombón	(3,7)	(6,9)	(7,4)	(2,3)	(9,3)	(1,7)	(2,7)	(4,7)	(2,9)	(4,6)	(4,9)	(9,5)	(8,5)	(1,3)	(2,3)
Tuba	(0,0)	(9,0)	(0,0)	(9,9)	(0,9)	(0,3)	(9,9)	(9,3)	(9,0)	(9,6)	(0,3)	(7,4)	(8,9)	(0,4)	(0,3)
Corno Francés	(4,9)	(1,3)	(4,7)	(4,2)	(4,2)	(4,2)	(3,3)	(0,7)	(2,5)	(5,9)	(5,7)	(7,0)	(9,7)	(0,2)	(2,0)
Saxo Tenor	(6,1)	(6,0)	(0,4)	(9,0)	(0,3)	(9,0)	(6,6)	(0,5)	(4,9)	(2,8)	(5,4)	(3,5)	(4,4)	(6,9)	(3,9)
Oboe	(9,5)	(4,9)	(8,2)	(0,6)	(9,5)	(3,6)	(2,1)	(5,9)	(3,7)	(7,2)	(9,8)	(0,9)	(6,0)	(2,7)	(0,9)
Corno Ingles	(8,0)	(2,0)	(2,5)	(6,2)	(2,3)	(5,1)	(6,4)	(2,7)	(1,7)	(6,5)	(7,9)	(4,7)	(5,2)	(3,9)	(4,7)
Basson	(5,3)	(7,5)	(9,0)	(9,6)	(9,0)	(9,9)	(7,8)	(6,3)	(7,4)	(0,9)	(3,6)	(5,1)	(7,7)	(3,4)	(2,5)
Clarinete	(9,3)	(2,6)	(8,9)	(1,9)	(9,9)	(3,9)	(4,0)	(6,8)	(7,2)	(5,3)	(7,6)	(3,9)	(2,2)	(7,7)	(6,5)
Flauta Traversa	**(4,0)**	**(5,2)**	**(1,7)**	**(6,0)**	**(3,0)**	**(6,0)**	**(0,6)**	**(0,9)**	**(9,9)**	**(7,8)**	**(0,9)**	**(5,3)**	**(6,5)**	**(4,6)**	**(7,3)**

Tabla 3-2 – Resultados Intermedios del Modelo

Los distintos vectores que representan a cada sonido, son la entrada de la red Superior de Kohonen.

3.5. Mapa Superior de Kohonen

Llegados a este punto del modelo cada sonido ha sido pre-procesado, dividido en partes y cada parte analizada en una Mapa de Kohonen Inferior.

El mapa de Kohonen de esta sección (Paso 5 y 6 de Figura 3-1) tiene la función de agrupar las salidas de todos los Mapas Inferiores de Kohonen.

El Mapa Superior de Kohonen es entrenado con todos los sonidos del conjunto de entrenamiento, representados a esta altura como vectores de longitud 30 (son las filas de Tabla 3-2).

Antes de introducir la siguiente modificación del modelo, es necesario mencionar un poco de teoría de los armónicos de los instrumentos musicales.

Como hemos visto en los gráficos de los instrumentos, el sonido producido por un instrumento musical es muy complejo. Esto se debe a la presencia de otras frecuencias además de la frecuencia *fundamental* (la que da el tono). Estas otras frecuencias son llamadas armónicos y son múltiplos enteros de la frecuencia fundamental. Estos armónicos son una secuencia infinita de ondas de frecuencia cada vez más elevada.

A modo de ejemplo, si tocamos la Nota E2 (82.4 Hz – Mi de la segunda octava) las frecuencias de los 3 siguientes armónicos son:

82.4 * 2 = 164.8 Hz (segundo armónico)
82.4 * 3 = 247.2 Hz (tercer armónico)
82.4 * 4 = 329.6 Hz (cuarto armónico)

La siguiente tabla muestra un ejemplo de la distribución de los armónicos de un sonido en las diferentes Jerarquías de nuestro modelo, basándonos en la misma nota mi del ejemplo anterior:

Armónico Nro.	Frecuencia	Jerarquía
Fundamental	82.4 Hz (E2)	7
Segundo	164.8 Hz (E3)	8
Tercero	247.2 Hz	9
Cuarto	329.6 Hz (E4)	9
Quinto	412.0 Hz	10
Sexto	494.4 Hz	10
Séptimo	576.8 Hz	10
Octavo	659.2 Hz (E5)	10
Noveno	741.6 Hz	11
Décimo	824.0 Hz	11
11	906.4 Hz	11
12	988.8 Hz	11
13	1071.2 Hz	11
14	1153.6 Hz	11
15	1236.0 Hz	11
16	1318.4 Hz (E6)	11
17	1400.8 Hz	12
18	1483.2 Hz	12
19	1565.6 Hz	12
20	1648.0 Hz	12
21	1730.4 Hz	12
22	1812.8 Hz	...

Tabla 3-3 – Relación Armónicos - Jerarquías

Como regla aproximada podemos decir que cada jerarquía por encima de la siguiente a la fundamental contiene el doble de armónicos que la jerarquía anterior.

Es decir:

La jerarquía X	contiene	el armónico fundamental
La jerarquía X+1	contiene	el armónico siguiente
La jerarquía X+2	contiene	los 2 armónicos siguientes
La jerarquía X+3	contiene	los 4 armónicos siguientes
La jerarquía X+4	contiene	los 8 armónicos siguientes

Lo importante de lo que acabamos de ver, es que los sonidos están compuestos, además de la frecuencia fundamental, por otras ondas de frecuencia mayor a la fundamental.

Los elementos percusivos son aun más complejos en cuanto a los parciales que lo componen, ya que generalmente sus parciales no son armónicos, es decir sus frecuencias no son un múltiplo entero de la fundamental.

Por ejemplo en el tambor, los parciales tienen aproximadamente la relación: 1, 1.59, 2.14, 2.65, 2.92, 3.16, 3.50, 3.60, 3.65, etc., todo esto con relación a su fundamental. Esta secuencia de parciales recibe el nombre de patrón de Chaldni (Chaldni patterns) y es producida por la vibración del parche. Esta secuencia también se ve afectada por el aire dentro de la caja de resonancia del instrumento.

Entonces, como explicamos, un sonido tiene una frecuencia fundamental (armónico principal) y una serie de armónicos superiores que de acuerdo a como evolucionan en el tiempo le dan a cada sonido su color particular ([DEP/1997], [PAR/1994] y [MAR/1998b]). Debido a las características de construcción de muchos instrumentos también pueden influir en el sonido frecuencias más bajas que la fundamental, que resuenan al ejecutar el instrumento.

En nuestro modelo hemos optado por incluir en el análisis una jerarquía más, por debajo de la jerarquía que contiene a la frecuencia fundamental, como una forma de tener en cuenta la posible presencia de información en estas frecuencias. El resto de las jerarquías, que contienen información de las frecuencias entre 0 y 86 Hz aproximadamente, no es tenido en cuenta para el análisis debido a la poca información que contiene.

Esto se puede apreciar gráficamente en la Figura 3-3 de la página 34, (y figuras similares del Anexo A. Ejemplo de pre-procesamiento de algunos sonidos). Como se puede ver, el armónico fundamental (usualmente el de mayor intensidad) se encuentra en la Jerarquía 9 (pues es la que contiene la frecuencia del DO de la 4ta octava). También puede verse que las jerarquías que contienen frecuencias menores no poseen información significativa, salvo la jerarquía 8.

A pesar de que las Jerarquías Inferiores no son utilizadas en el análisis final, igualmente son entrenadas, porque una de las posibles variaciones del modelo sería, no descartar completamente los valores de estas jerarquías, sino utilizarlos en forma ponderada en el análisis final.

Una vez entrenado el Mapa Superior de Kohonen, y por ende el modelo completo, se obtienen tres tipos de aplicaciones/funcionalidades:

1. Un mapa topológico de los instrumentos, según el parecido que nuestro modelo encuentra entre los diferentes instrumentos (Ver un ejemplo en la Figura 3-5 – Resultado Final del Modelo). En este mapa se puede apreciar agrupaciones por familias de sonidos (con algunas

excepciones). Para facilitar la visualización de las familias, las hemos coloreado en forma manual.

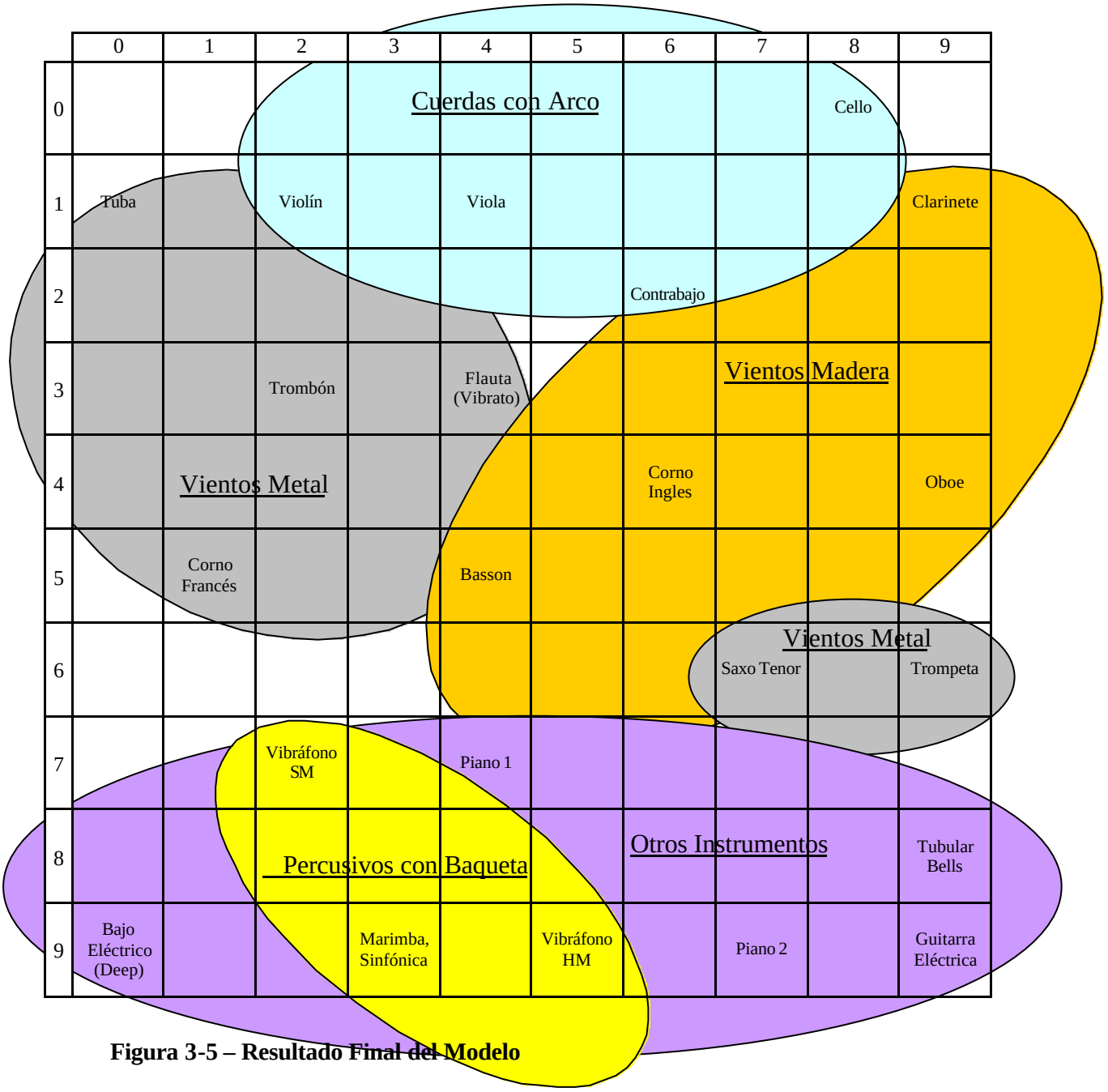


Figura 3-5 – Resultado Final del Modelo

2. La capacidad de identificar instrumentos conocidos, sobre la base de los sonidos usados en el conjunto de entrenamiento. Con esta funcionalidad sería posible, con una herramienta completa, que el sistema identifique un instrumento al escuchar una sola nota del mismo. Decimos que sería posible (y no que es posible) pues recordamos que el sistema actual está modelado para el caso de la nota DO de la cuarta octava de los instrumentos, y que haría falta generalizarlo para lograr una independencia de la frecuencia.

3. La capacidad de clasificar instrumentos nuevos, sobre la base de los sonidos usados en el conjunto de entrenamiento. Esta funcionalidad permite clasificar instrumentos nuevos, según la familia de sonidos en donde "caigan" en el Mapa de Kohonen, o inclusive de cuales sonidos queden más cerca. Por ejemplo, si en la Figura 3-5 un instrumento nuevo y desconocido activa la neurona de la posición [1,3] (entre el violín y la viola), es altamente probable que sea un instrumento de cuerda, tocado con arco y con caja de resonancia.

Solo a modo de ejemplo, vemos en la Figura 3-5 el resultado final de un mapa topológico generado por nuestro modelo.

Notar que en este caso el modelo no pudo agrupar la Trompeta y al Saxo Tenor con el resto de los vientos de metal.

Nota: De este gráfico NO debe deducirse que la viola es igual de parecida al violín que a la flauta. Ver en Capítulo 4 - Resultados Experimentales, más detalles sobre las distancias entre celdas. Recordemos que Kohonen no se expande uniformemente, sino que intenta cubrir todo el espacio disponible (mediante una transformación no lineal, lo que es causa de algunas anomalías), por lo que un análisis más detallado revelaría que la viola se encuentra a distancia 10 de la flauta y a distancia 4 del violín.

Capítulo 4 - Resultados Experimentales

4.1. Información adicional para la organización

En esta sección se comentan las elecciones y decisiones tomadas durante la implementación del modelo.

Toda la implementación y varias herramientas de visualización fueron programadas en lenguaje C, utilizando el compilador Borland. Las únicas librerias externas utlizadas fueron las funciones de procesamiento de la Trasnformada Discreta de Waveletes, tomadas de "Numerical Recipes in C" [PRE/1992]

Como herramienta de manipulación de sonidos se utilizó el software CoolEdit 96 creado por David Johnston. La figura de la página 26 y algunas verificaciones de la transformación de wavelet se realizaron mediante el software Praat [BOE/1999]. Los gráficos de las páginas 55 y 56 fueron generados utilizando el software MatLab.

Variables del Modelo (Kohonen)

Las redes de Kohonen [KOH/1988] pueden tener distintas formas y dimensiones. En nuestro caso hemos optado por utilizar redes cuadradas, de diez por diez neuronas cada una. Esta elección se basó en considerar la cantidad de sonidos que se iban a utilizar para el entrenamiento (22) y en tener el suficiente "espacio" en la red para lograr una buena distribución. Es decir, tenemos una cantidad de neuronas cinco veces superior a la cantidad de vectores de entrada, aproximadamente, lo que es similar a otras experiencias realizadas con redes de Kohonen.

Como se vio en el Capitulo 2, el entrenamiento de una Red de Kohonen se divide en 2 etapas o fases: fase de Ordenamiento y fase de Convergencia. Estas dos fases son bastante similares en su funcionamiento, a excepción de algunas diferencias conceptuales, causadas por el objetivo que persigue cada una.

En la fase de Ordenamiento se busca lograr una organización general de los vectores de entrada, de forma que se agrupen por su similitud. Para esto, se alimenta a la red con todos los vectores de entrada, en forma sucesiva y volviendo a procesar el mismo conjunto de vectores, una y otra vez, en forma cíclica. Cada una de estas "vueltas" donde es ingresado y procesado todo el conjunto de entrada se llama *época* o *ciclo*. En nuestra implementación práctica del modelo propuesto, consideramos dos casos diferentes, las redes inferiores y la red superior o final. Para las redes inferiores utilizamos una cantidad de épocas igual a 256.000, la cual fue apropiada para lograr una estabilidad en la organización final. Para la red superior utilizamos 150.000 épocas.

Al tomar un vector de entrada y encontrar la neurona que activa en la red, se actualiza ésta neurona de acuerdo a un valor de aprendizaje, que va decreciendo conforme vayan pasando las épocas. También se actualiza una cierta vecindad de la neurona ganadora, y esta vecindad también va decreciendo en función de las épocas transcurridas. En nuestro caso, y tanto en las redes inferiores como en la superior, utilizamos 0,32 como valor inicial de este factor de aprendizaje (eta) y 0 como valor final. El rango de vecindad también fue el mismo para los dos casos, variando entre 10 y 1, es decir, en las últimas épocas de esta fase solo se actualiza la neurona ganadora, y en una proporción muy pequeña.

En la fase de Convergencia, lo que se busca es que las posiciones de la matriz que se parecen más a un vector de entrada, se parezcan aún más. Esto se logra actualizando únicamente la neurona ganadora, mediante un factor de aprendizaje (eta) que también va disminuyendo con el tiempo. Tanto en las redes inferiores como en la superior partimos de un valor inicial de aprendizaje de 0,11 disminuyendo hasta ser nulo. En las redes inferiores se utilizó una cantidad de épocas igual a 25.600 (un diez por ciento de la cantidad de épocas de la fase de ordenamiento) y en la red superior se utilizaron 15.000 épocas. Las diferencias entre el número de épocas para las distintas redes se basa en el hecho que la dimensionalidad de las entradas de la red superior es notablemente menor a la de las redes inferiores, llegando a una estabilización de la red con mucho menos trabajo.

El programa que implementa el modelo utiliza archivos de configuración para definir los parámetros que hemos mencionado y que modifican su comportamiento.

Existe un archivo general de configuración (kohonen.cfg), un archivo por cada red inferior de Kohonen (Jerar_XX.cfg) y un archivo para el comportamiento de la red superior de Kohonen (Superior.cfg). A continuación se muestran los valores de las variables más importantes que definen el comportamiento del modelo durante el aprendizaje de la corrida que mostraremos en este capítulo. Hay variables o parámetros que son necesarios definir para que el programa funcione pero que no reproducimos aquí por ser irrelevantes en este momento.

Jerar_XX.cfg

Variable	Descripción
OutputFil=10 OutputCol=10	Tamaño de la Matriz de salida de la kohonen
IteracionesOrdenamiento=256000 IteracionesConvergencia=25600	Cantidad de iteraciones del ciclo de Ordenamiento y del ciclo de Convergencia de Kohonen
OrdEtaInicio=0.32 OrdEtaFin=0 OrdVecindadInicio=10 OrdVecindadFin=1	Eta (Ver Ecuación 2-2) y vecindad (Ver Ecuación 2-2) para la fase de ordenamiento
CnvEtaInicio=0.11 CnvEtaFin=0 CnvVecindadInicio=1 CnvVecindadFin=1	Eta y vecindad para fase de convergencia

- Estos valores son todos iguales para las jerarquías 8 a 15 (principalmente porque la estructura de los datos de entrada es la misma para todas estas jerarquías).
- Las variables Eta y Vecindad decrecen linealmente sobre la cantidad de épocas.

Superior.cfg

Variable	Descripción
FirstJerarq=8 LastJerarq=15	Estos dos parámetros indican el rango de Jerarquías Inferiores que serán utilizadas para entrenar la red superior
OutputFilSuperior=10 OutputColSuperior=10	Tamaño de la Matriz de salida de la Red de Kohonen Superior
IteracionesOrdenamientoSuperior=150000 IteracionesConvergenciaSuperior=15000	Cantidad de iteraciones del ciclo de Ordenamiento y del ciclo de Convergencia de Kohonen Superior
OrdEtaInicioSuperior=0.32 OrdEtaFinSuperior=0 OrdVecindadInicioSuperior=10 OrdVecindadFinSuperior=1	Eta y vecindad para fase de ordenamiento de Kohonen Superior

CnvEtaInicioSuperior=0.11 CnvEtaFinSuperior=0 CnvVecindadInicioSuperior=1 CnvVecindadFinSuperior=1	Eta y vecindad para fase de convergencia de Kohonen Superior

Factor de aprendizaje de Kohonen

Sobre la Ecuación 2-2 que mencionamos en el capítulo 2,

$$\Delta w_{ij} = \eta \Lambda(i,i^*) (\xi_j - w_{ij})$$

solo queda por definir la forma de la función Λ (vecindad), pues η (eta) toma su valor directamente de los archivos de configuración, decreciendo en forma lineal.

La función de vecindad que utilizamos es

$$\Lambda(i,i^*) = e^{\left(\frac{-\text{distancia}(i,i^*)^2}{2 \cdot \text{vecindad}^2} \right)}$$

sugerida por Hertz [HER/1991] como una función típica para $\Lambda(i,i^*)$.
La variable vecindad decrece linealmente durante al aprendizaje y sus límites superior e inferior también están definidos en los archivos de configuración.

Nota: No confundir variable vecindad con la función de vecindad. La variable vecindad define cuales neuronas hay que modificar, es decir el radio del vecindario a partir de la neurona ganadora, mientras que la función de vecindad influye sobre la proporción en que se las modificará.

4.2. Resultado y análisis de la aplicación del modelo

Los instrumentos siguientes fueron los utilizados en la generación de los mapas tímbricos.

Piano 1	Guitarra Eléctrica	Trompeta	Corno Ingles
Piano 2	Bajo Eléctrico (Deep)	Trombón	Basson
Tubular Bells	Violín	Tuba	Clarinete
Vibráfono Hard Mallet	Viola	Corno Francés	Flauta Traversa
Vibráfono Soft Mallet	Cello	Saxo Tenor	
Marimba, Sinfónica	Contrabajo	Oboe	

Se presentan a continuación las redes finales para cada una de las jerarquías o bandas de frecuencia propuestas (la organización que aquí se muestra se obtuvo de una corrida típica, y es representativa, con pequeñas variaciones, de todos los resultados obtenidos en las diferentes corridas). Si bien el resultado final sobre el ordenamiento o mapeo de los sonidos está representado únicamente en el mapa final, ayuda a tener una idea sobre como trabaja el modelo mirando los pasos intermedios, de los cuales también pueden extraerse algunas conclusiones interesantes.

Jerarquía 8										
	0	1	2	3	4	5	6	7	8	9
0				Vibráfono SM		Saxo Tenor		Corno Francés		Flauta (Vibrato)
1	Bajo Eléctrico (Deep)							Violín		
2								Corno Ingles		Trompeta
3			Marimba, Sinfónica			Guitarra Eléctrica			Tubular Bells	
4								Trombón		
5										Oboe
6	Piano 2			Basson			Viola		Clarinete	
7										Vibráfono HM
8										
9	Piano 1			Tuba			Cello			Contrabajo

Esta jerarquía representa a los armónicos de los sonidos que están comprendidos entre 86 y 172 Hertz, aproximadamente, y si bien estos armónicos están por debajo de la frecuencia fundamental de la nota que se está considerando (262 Hz aprox.), se incluyó esta jerarquía debido a que presentaba información que se consideró útil, sobre todo para los sonidos de instrumentos naturalmente graves.

Como punto interesante puede destacarse la agrupación en la esquina superior derecha de varios instrumentos de viento de sonido más brillante, asi mismo junto a instrumentos como el violín. Todos estos instrumentos se destacan por tener poco contenido espectral en esta banda de frecuencias. En todo el borde opuesto, costado izquierdo e inferior, encontramos instrumentos de sonido característicamente más grave u oscuro, como son los pianos, el contrabajo, el bajo eléctrico, la tuba, etc. Es decir sonidos que naturalmente suenan más grave.

Jerarquía 9										
	0	1	2	3	4	5	6	7	8	9
0						Guitarra Eléctrica			Trompeta	Tubular Bells
1	Vibráfono SM			Bajo Eléctrico (Deep)				Corno Ingles		
2						Corno Francés				Trombón
3			Piano 1					Oboe		
4					Vibráfono HM					Saxo Tenor
5							Piano 2			
6	Marimba, Sinfónica									Contrabajo
7			Clarinete		Basson			Violín		
8										
9	Tuba			Cello			Viola			Flauta (Vibrato)

Esta jerarquía representa a los armónicos de los sonidos que están comprendidos entre 172 y 344 Hertz, aproximadamente, y es la banda de frecuencia que contiene a la frecuencia fundamental de los sonidos evaluados. Es quizás en esta banda de frecuencias donde menos diferencia tendría que encontrarse entre los diferentes sonidos, debido precisamente a la gran importancia en todos ellos de la frecuencia del DO, es decir, en 262 Hz. Sin embargo es interesante notar como en la esquina inferior derecha se han agrupado todos los instrumentos de cuerda tocados con arco, junto con la flauta, algo que también ocurre en el estudio realizado por Grey [GRE/1975]. También se empiezan a observar grupos de sonidos que se consolidarán en la estructura final, como por ejemplo la agrupación de la marimba, los vibráfonos, el piano1, el bajo eléctrico y la guitarra eléctrica, que ocupa todo el sector superior izquierdo del mapa. Los vientos de metal han quedado casi en su totalidad en el extremo superior derecho, a excepción de la tuba que está en el extremo opuesto, y que se podría explicar por medio de un pequeño twist durante la fase de entrenamiento de las redes.

Jerarquía 10										
	0	1	2	3	4	5	6	7	8	9
0	Marimba, Sinfónica			Piano 2			Piano 1			Basson
1										
2	Bajo Eléctrico (Deep)		Vibráfono SM			Guitarra Eléctrica			Saxo Tenor	
3										
4		Vibráfono HM					Trombón			
5	Tubular Bells			Clarinete						Corno Francés
6						Corno Ingles				
7	Trompeta		Oboe						Flauta (Vibrato)	
8		Violín								
9	Viola			Cello			Tuba			Contrabajo

Esta jerarquía representa a los armónicos de los sonidos que están comprendidos entre 344 y 689 Hertz, aproximadamente, y es la banda de frecuencia donde se empiezan a considerar los armónicos superiores, con frecuencia mayor igual al doble de la frecuencia fundamental.

Aquí también se pueden ver agrupaciones interesantes, como por ejemplo que continua la agrupación de los pianos, los vibráfonos, la marimba, el bajo y la guitarra eléctrica. Esto guarda también relación con lo que ocurre en el trabajo de Prandoni [PRA/1994] donde los pianos se agrupan con las cuerdas punteadas (bajo y guitarra eléctrica, en este caso). Se observa en el sector inferior izquierdo la agrupación de las cuerdas tocadas con arco menos graves, es decir, el violín, la viola y el violoncello. Esta situación se condice con la situación real, por la cual cuando estos instrumentos se ejecutan en este rango de frecuencias muchas veces se hace difícil distinguir uno de otro, a menos que se escuche una frase musical más larga, donde queden en evidencia las diferencias de espectro entre ellos. El contrabajo se encuentra un poco más alejado de este grupo.

Es importante destacar que en el modelo presentado no solo se tiene en cuenta la importancia del contenido en las bandas de frecuencia para comparar dos sonidos, sino que también entra en juego, y en gran forma, como se “mueve” ese contenido a lo largo de la duración del sonido. Por lo cual, aparte de la frecuencia, se esta evaluando también la forma en que se articula el sonido, razón que ayuda a que las cuerdas, por ejemplo, se aproximen más entre ellas que si solo se hiciera una cuantificación del contenido espectral en un momento determinado.

Jerarquía 11										
	0	1	2	3	4	5	6	7	8	9
0		Bajo Eléctrico (Deep)		Tuba				Violín		Flauta (Vibrato)
1	Vibráfono SM					Viola				
2		Marimba, Sinfónica							Contrabajo	
3	Piano 2			Trompeta			Basson			
4										Trombón
5	Vibráfono HM				Saxo Tenor			Corno Francés		
6			Cello							
7							Clarinete			Corno Ingles
8				Tubular Bells						
9	Piano 1					Guitarra Eléctrica			Oboe	

Esta jerarquía representa a los armónicos de los sonidos que están comprendidos entre 689 y 1378 Hertz, aproximadamente, y es la banda de frecuencia que incluye las frecuencias involucradas en lo que en la literatura de audio, y en algunos trabajos tímbricos [PRA/1994] [DEP/1997], se denomina *presencia*, es decir la banda comprendida entre 700 y 900 Hz aproximadamente.

Aquí seguimos observando la similitud entre los vibráfonos, la marimba, los pianos. Es interesante notar que en este rango de frecuencias se han agrupado juntos tres instrumentos como el violín, la viola y el contrabajo (junto con la flauta, que siempre aparece relacionada a las cuerdas), pero sin embargo el violoncello se encuentra separado de este grupo por una "barrera" de instrumentos de viento. Los instrumentos de viento de madera, como el oboe, el clarinete y el corno inglés tienden a situarse lejos de instrumentos de sonido más "oscuro", como pueden ser el vibráfono, el piano, la marimba y el bajo eléctrico. Lejos de los vientos de madera se ubica el fagot (basson), más cercano a los instrumentos de bronce, como la trompeta, corno francés y trombón, que es también una de las "anomalías" presentes en el mapa tímbrico de Grey [GRE/1975].

La distribución de los vientos de metal podría verse como un arco formado por estos, separando a las cuerdas de los demás instrumentos. Este arco puede verse como formado por la tuba, la trompeta, el saxo tenor, el corno francés y el trombón.

Jerarquía 12										
	0	1	2	3	4	5	6	7	8	9
0	Vibráfono HM			Tubular Bells			Cello			Oboe
1										
2	Piano 2									
3			Guitarra Eléctrica			Saxo Tenor				Clarinete
4	Piano 1							Corno Ingles		
5		Basson		Flauta (Vibrato)						
6							Trompeta			Contrabajo
7	Corno Francés				Tuba					
8			Vibráfono SM					Viola		
9	Marimba, Sinfónica	Bajo Eléctrico (Deep)				Trombón				Violín

Esta jerarquía representa a los armónicos de los sonidos que están comprendidos entre 1378 y 2756 Hertz, aproximadamente, y es la banda de frecuencia que incluye los armónicos de orden quinto a décimo de la frecuencia original.

En este mapeo encontramos similitudes a banda anterior, con la agrupación de las cuerdas, salvo el violoncello. El grupo tradicional de percusivos, cuerdas punteadas y pianos, queda dividido por la aparición del corno francés (bronce) y el basson (fagot, madera), mostrando que en este rango de frecuencias los sonidos de estos instrumentos de viento se hacen cada vez más complejos en su composición armónica. Uno de los vibráfonos, los pianos, la guitarra eléctrica y las campanas tubulares permanecen agrupados. También se ve una agrupación entre los metales tuba, trombón y trompeta. El cello en este caso se aleja nuevamente de las demás cuerdas, acercándose a los vientos de madera, de sonido más brillante.

Jerarquía 13										
	0	1	2	3	4	5	6	7	8	9
0	Cello							Piano 2		
1						Tubular Bells				Vibráfono HM
2			Clarinete							
3	Trompeta						Guitarra Eléctrica		Bajo Eléctrico (Deep)	
4					Saxo Tenor					Vibráfono SM
5			Corno Ingles					Piano 1		
6	Oboe					Flauta (Vibrato)				Marimba, Sinfónica
7			Viola					Basson		
8						Trombón				Tuba
9	Contrabajo			Violín				Corno Francés		

Esta jerarquía representa a los armónicos de los sonidos que están comprendidos entre 2756 y 5512 Hertz, aproximadamente, y es la banda de frecuencia que incluye los armónicos desde al límite anterior hasta el orden vigésimo.

Aquí vemos algunos metales ocupando el extremo inferior derecho (tuba, corno francés, trombón) junto al basson (fagot) que sigue alejándose del resto de los vientos de madera. El grupo clásico de instrumentos percusivos, pianos y cuerdas punteadas ocupa el extremo superior derecho.

Jerarquía 14										
	0	1	2	3	4	5	6	7	8	9
0	Vibráfono SM		Corno Francés		Tuba		Violín			Viola
1		Marimba, Sinfónica		Trombón						
2	Bajo Eléctrico (Deep)		Piano 1					Oboe		
3					Basson					Corno Ingles
4	Vibráfono HM						Flauta (Vibrato)			
5			Piano 2							
6						Trompeta				Saxo Tenor
7	Tubular Bells							Clarinete		
8										
9			Guitarra Eléctrica			Contrabajo				Cello

Esta jerarquía representa a los armónicos de los sonidos que están comprendidos entre 5512 y 11025 Hertz, aproximadamente, y es la banda de frecuencia que incluye los armónicos desde al límite anterior hasta el orden cuadragésimo.

Aquí nos encontramos en un rango de frecuencias muy alto, donde empiezan a pesar mucho la cantidad y complejidad de todos los armónicos considerados, y esto permite ver algunas agrupaciones interesantes. El violín y la viola (naturalmente más agudos) se han separado del contrabajo y el cello (más graves), y se encuentran cerca de algunos vientos de madera como el oboe y el corno inglés, de sonido más brillante. Algunos instrumentos de metal (corno francés, tuba y trombón) se han agrupado en el lado superior, separando a las cuerdas del grupo percusivo.

Jerarquía 15										
	0	1	2	3	4	5	6	7	8	9
0		Vibráfono SM		Tuba			Violín			Oboe
1	Bajo Eléctrico (Deep)									
2	Corno Francés			Trombón		Basson		Viola		
3		Piano 1								Saxo Tenor
4					Trompeta			Corno Ingles		
5	Vibráfono HM		Marimba, Sinfónica							
6						Clarinete				Contrabajo
7	Piano 2			Flauta (Vibrato)						
8										
9			Tubular Bells			Guitarra Eléctrica				Cello

Esta última jerarquía representa a los armónicos de los sonidos que están comprendidos entre 11025 y 22050 Hertz, aproximadamente, y es la banda de frecuencia que llega hasta los límites de audición del ser humano (alrededor de los 20 Khz). Aquí las componentes a estudiar son mucho más sutiles y, quizás, más insignificantes que en las jerarquías anteriores, pero se incluyeron, de forma de tener en cuenta para el estudio todo el espectro de audición normal de un oyente humano. Es en cierta forma muy similar a la jerarquía anterior y pueden observarse aproximadamente las mismas agrupaciones, con algunas mínimas diferencias.

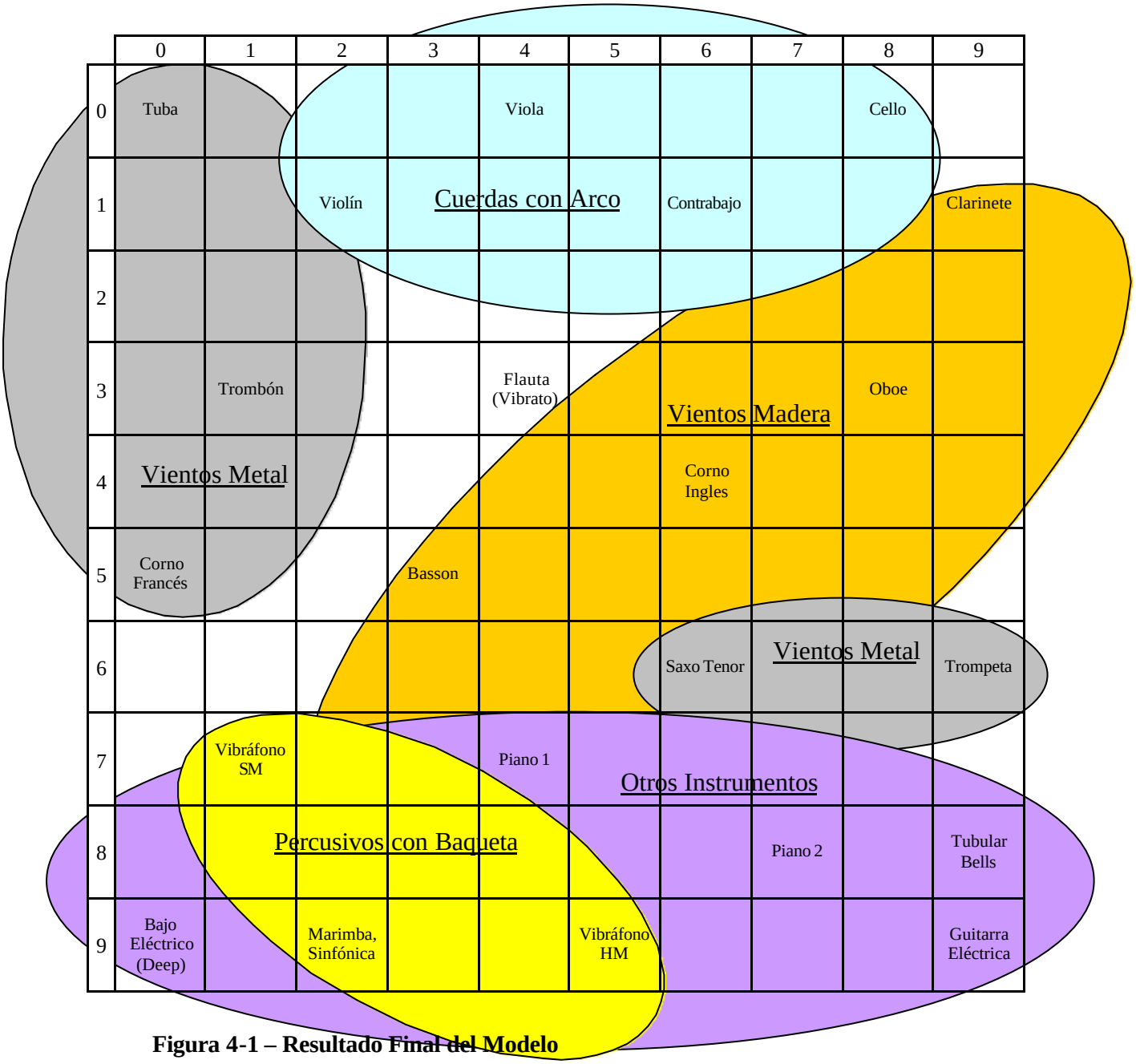


Figura 4-1 – Resultado Final del Modelo

En el mapeo u organización final se tienen en cuenta todos los mapas que hemos visto, desde la jerarquía 8 hasta la 15, y basándose en el comportamiento de un sonido a través de todas esas bandas de frecuencias se organizan los diferentes sonidos de los instrumentos.

Se puede observar una interesante agrupación de todas las cuerdas tocadas con arco en la parte superior del mapa, con el violín, la viola, el violoncello y el contrabajo muy próximos entre si.

Los vientos de madera se encuentran en la zona central y superior derecha del mapa, con la excepción del basson (fagot) que se encuentra en una posición algo alejada de estos, y más cercano

a los vientos de metal, como suele encontrarse en los estudios que se han hecho anteriormente [GRE/1975].

Los vientos de metal se han agrupado en dos bloques como puede verse, por un lado tuba, trombón y corno francés, y por el otro el saxo tenor y la trompeta, aunque vistos como un arco que separa a las cuerdas de los instrumentos percusivos puede verse que es el basson (fagot) el vínculo entre ellos, siendo este, como hemos visto, un instrumento que tiende a agruparse con los metales.

En todo el sector inferior del mapa vemos a los instrumentos percusivos, pianos, y cuerdas punteadas, mostrando en forma definitiva lo que se fue viendo en cada una de las diferentes jerarquías. Dentro de ellos, se ve una sub agrupación de instrumentos muy similares, como son el vibráfono (tocado con baquetas suaves y duras) y la marimba. Los pianos quedan a la derecha de este grupo separándolo de las campanas tubulares y la guitarra eléctrica. Estos instrumentos, tanto la guitarra eléctrica, el piano, los vibráfonos, etc., a diferencia de los vientos, tienen una fase de ataque muy importante, que los diferencia del resto de los instrumentos.

Para realizar un análisis detallado del mapa es preciso conocer el valor final de cada neurona (vectores de 16 valores). Un ejemplo claro del tipo de conocimiento que obtenemos al realizar un análisis detallado es el siguiente:

> En la Figura 4-1 se aprecia que la viola y el trombón están a la misma distancia del violín sobre el Mapa de Kohonen (a distancia raíz$(1+2^2) \approx 2.23$), lo que NO podemos deducir es que la viola y el trombón tienen el mismo grado de similitud con el violín. Teniendo en cuenta que Kohonen no se expande uniformemente, sino que intenta cubrir todo el espacio disponible [KOH/1988], un análisis más detallado nos revela que la distancia entre la neurona que representa al violín y la neurona que representa al trombón es de 12 (comparación en dimensión 16), mientras que la distancia entre las neuronas del violín y la viola es de 7.8, lo cual hace al violín más parecido a la viola que al trombón.

Por lo tanto es muy útil para los análisis el tener conocimiento de la disposición espacial de las neuronas. Evidentemente esto es un problema, pues las neuronas de la red superior de Kohonen tienen una dimensión de 16, lo que dificulta 'levemente' (;-)) el obtener una representación gráfica de las mismas.

Uno de los métodos que hemos utilizado para representar parcialmente esta información, consiste en ver las distancias de todas las neuronas con respecto a un sonido elegido. Esto nos permite visualizar en forma rápida la distancia entre un sonido y el resto de los sonidos (o neuronas).

Tomando el caso del violín (que 'cae' en la neurona [1,2] de la Red Superior), comparamos su valor 16-dimensional contra todas las neuronas de la red. El resultado de esta comparación euclidiana se puede apreciar en la Figura 4-2 – Distancias desde el Violín.

De esta manera se puede observar gráficamente las distancias que mencionábamos unos párrafos más arriba (en realidad no estamos comparando contra la viola o el trombón sino contra las neuronas que los representan).

Con este método obtenemos una comprensión de la relación entre los sonidos desde el punto de vista del violín, lo cual nos da una idea muy clara de cuales sonidos se parecen al violín y cuales no.

Para tener una idea más acabada de la estructura final de la red es preciso aplicar este método por separado para cada sonido, pues los resultados de un sonido no sirven para analizar a ningún otro par de sonidos; es decir, el conocer las distancias recientemente mencionadas entre viola, trombón y el violín nos ofrece poca o ninguna información de la distancia (16-dimensional) entre la viola y el trombón.

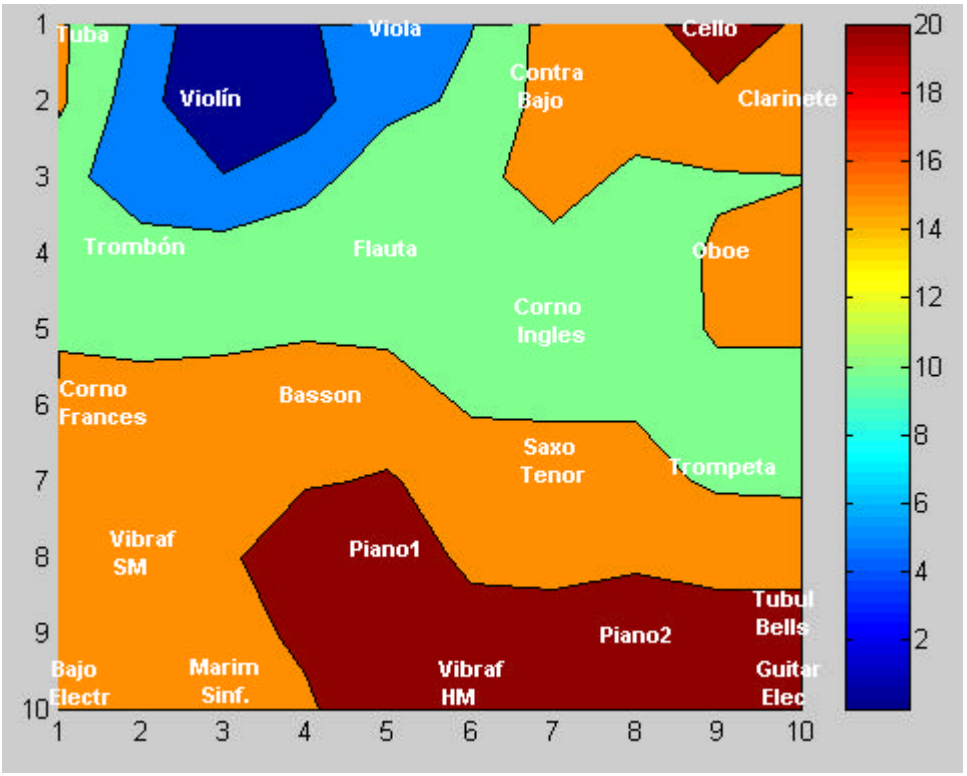


Figura 4-2 – Distancias desde el Violín

Otras observaciones interesantes sobre esta misma figura:

- A pesar de que el Violín y Tuba están cerca en la representación gráfica en dos dimensiones de Kohonen, la distancia entre sus neuronas es mediana-alta, lo cual concuerda con la diferencia real entre los sonidos de estos instrumentos.

- Si un sonido de origen desconocido es pasado a través de todo el modelo y 'cae' en el color azul (es decir, activa una de las neuronas de esa área) es altamente probable que sea un violín, si 'cae' en el color celeste es probable que sea de la misma familia.

Otro de los métodos utilizados para analizar la disposición espacial de las neuronas, consiste en generar todas las distancias entre neuronas vecinas, esto es, generar las diferencias verticales, horizontales y diagonales entre todos los pares de neuronas adyacentes. Este gráfico (Figura 4-3) permite visualizar las variaciones locales entre neuronas.

Un punto muy importante al analizar esta figura, es que hay que tener presente que la transitividad no se cumple, con esto queremos decir que:

> si las neuronas *a*,*b* son adyacentes y las neuronas *b*,*c* son adyacentes, entonces
> si *a* esta lejos de *b* y *b* lejos de *c*, no se puede deducir nada de la distancia en que se encuentran *a* y *c*; básicamente porque la figura no representa información sobre el ángulo n-dimensional que las separa.

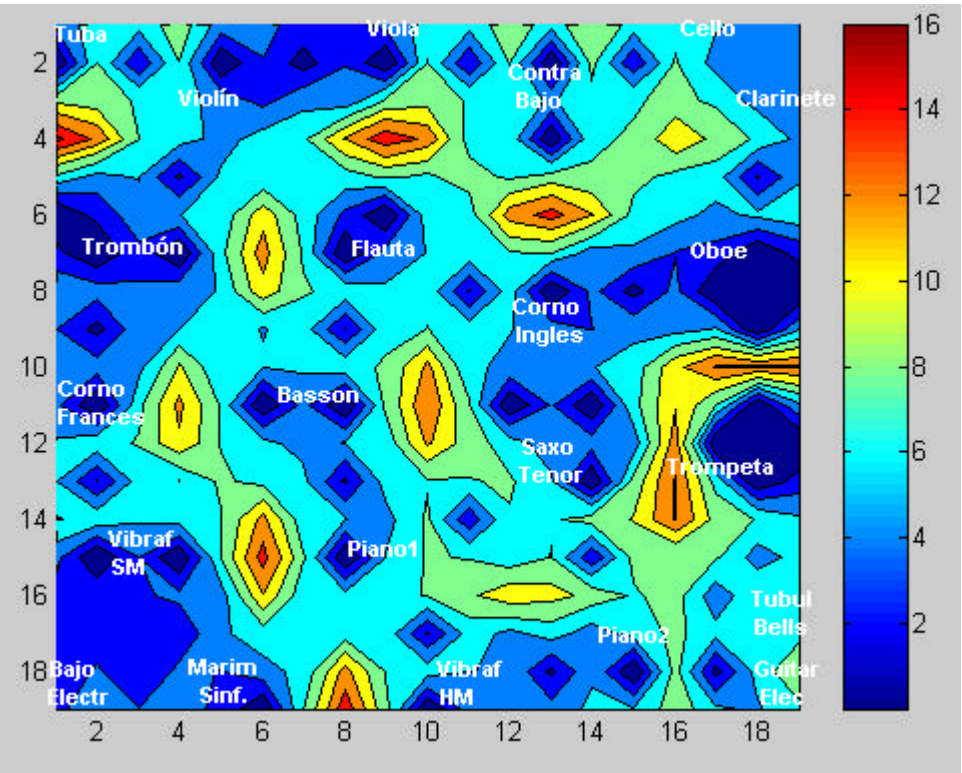


Figura 4-3 - Distancias entre neuronas

Una forma práctica de interpretar la figura es pensarla como representando un plano con curvas de nivel, de esta forma podemos considerar a los valores bajos (azules y celestes) como valles y los valores altos (verdes, amarillos y rojos) como montañas que separan los diferentes valles. Interpretándolo en este sentido, se considera que los pares de instrumentos adyacentes que están conectados por superficies azules o celestes guardan similitud entre si, y si están separados por áreas verdes, amarillas o rojas presentan mayores diferencias.

Las siguientes observaciones se desprenden del análisis de la Figura 4-3 - Distancias entre neuronas:

Se observa una anomalía alrededor de la trompeta, pues está muy lejos (en 16-d) de los sonidos que lo rodean (dicho gráficamente, la trompeta esta rodeada de montañas). A ningún otro sonido del gráfico le sucede esto. Es como si no se pareciera a los sonidos que están alrededor de ella, como si no perteneciera a ese lugar. Es posible que durante el periodo de aprendizaje hubiese quedado separada de los instrumentos de su familia, y esto podría deberse a un ‘twist’ [HER/1991b] de la red de Kohonen durante el periodo de aprendizaje (en estudios anteriores, como por ejemplo [GRE/1975], la trompeta se encuentra cerca de instrumentos como el basson (fagot) y el corno francés, y en [DEP/1997] las trompetas se encuentran cerca del oboe y el clarinete en mi bemol).

La única forma de tener una idea cabal de esta anomalía es aplicando el método de la Figura 4-2 comparando contra la trompeta. Verificando esto, notamos que la trompeta está más cerca del trombón que de la flauta, el basson o el saxo tenor (sonidos todos estos que se encuentran entre la

trompeta y el trombón). Esta situación podría también deberse a una condición irresoluble para la transformación que efectúa Kohonen a un espacio de 2 dimensiones.

Se puede notar también que existe una marcada separación entre los instrumentos de cuerda y los vientos de madera (observar el área debajo de viola, contrabajo y cello).

Conclusiones y Trabajos futuros

Como resultado de estos experimentos e investigaciones se pueden obtener varias conclusiones. El uso de wavelets para extraer información de los sonidos por octavas brinda una excelente representación de la evolución de la señal analizada en cada banda de frecuencias a lo largo del tiempo. Su comportamiento es superior al análisis de Fourier por ventanas para simular el desarrollo del sonido en el tiempo, ya que se amolda automáticamente al rango de frecuencias que se está analizando, mediante el escalamiento de sus funciones base. Quizás la mayor desventaja que presenta este método como representación de la señal, es que no brinda una compresión de las dimensiones de entrada que facilite su posterior procesamiento, ya sea con redes de Kohonen u otros métodos (la conocida compresión por wavelets, muy utilizada actualmente en video y audio, no es de aplicación práctica en nuestro caso, ya que está orientada principalmente a la compresión de espacio ocupado, y para que pueda utilizarse nuevamente la señal comprimida hay que recuperarla mediante el proceso inverso, lo que nos dejaría con el mismo problema de tamaño en nuestro caso). Las redes de Kohonen no soportan bien entradas con una dimensión muy grande si se las quiere "mapear" a una cantidad mucho menor de dimensiones (dos en nuestro caso). Por este motivo es que nos vimos obligados a realizar una compresión de datos posterior, que indudablemente puede producir una pérdida de detalle de las señales. Aunque incluso en estas condiciones, creemos que los resultados obtenidos han sido interesantes, y en algunos casos muestran gran similitud con estudios e investigaciones anteriores. En la actualidad parece haber, entre los investigadores del tema, una cierta preferencia por otros métodos de representación de señales como el método de Mel Frequency Coefficients. Éste ha sido muy utilizado en el campo de investigación fonética, y solo recientemente ha empezado a utilizarse para otro tipo de sonidos. Su ventaja reside principalmente en la detección de coeficientes significativos para breves segmentos de sonido (por lo que debe utilizarse con un método de ventanas como la Transformada Rápida de Fourier) y en la importante compresión de datos que brinda, posibilitando un análisis mucho más preciso posteriormente con técnicas como las redes de Kohonen, por su reducida dimensionalidad.

Pueden plantearse diferentes alternativas como continuación de este estudio, de forma de extenderlo para obtener mejores resultados, o quizás resultados más parecidos a un espacio tímbrico intuitivo o conceptual. Entre estas posibilidades se pueden sugerir las siguientes:

- Utilizar en todos los niveles inferiores y/o en el nivel superior de Kohonen, redes de tres dimensiones. En la investigación fundamental de Grey [GRE/1975], el resultado del experimento con oyentes humanos se muestra en una representación tridimensional, ya que Grey consideró que sus datos no podían representarse en forma adecuada con un mapa bidimensional. Para este caso sería muy útil obtener o desarrollar herramientas de visualización en tres dimensiones que faciliten la comprensión de los espacios obtenidos.

- Sería muy interesante tratar de extender este modelo de nota única (es decir, todos los instrumentos deben tocarse en la misma nota) a un modelo que pueda generalizar la información de cada instrumento y acepte entradas que puedan ser notas de diferentes frecuencias fundamentales. Quizás una idea para este modelo sería entrenar a las redes no solo con una nota por instrumento, sino utilizando un rango más amplio por cada instrumento. De esta forma podría estudiarse si las notas de un mismo instrumento tienden a agruparse naturalmente, o si hay otras relaciones que la red descubre y que pueden ser útiles para el objetivo propuesto. Otra alternativa podría ser generar varios mapas tímbricos, en el que cada uno tiene como entrada sonidos de una frecuencia particular. De esta forma habría varios mapas

tímbricos similares al de la Figura 4-1 – Resultado Final del Modelo de la página 53, cada uno generado a partir de diferentes notas bases (una por octava). El uso de los MUMs es apropiado para este caso, ya que se cuenta con todo el rango de sonidos de cada instrumento.

- Si bien pensamos que es necesario tener en cuenta tanto el ataque como la parte estable de los sonidos, sería interesante probar este modelo o alguna variación del mismo para que solo considere el ataque o la parte estable. Esto podría brindar un buen estudio comparativo entre las distintas alternativas y estudiar la que pueda ser más óptima para lo que se busca, ya sean reconocedores o clasificadores de sonidos.

- También puede ser útil investigar diferentes ponderaciones o importancias relativas de las jerarquías inferiores, de forma de encontrar la combinación de ponderaciones que logren una organización más parecida a la que se pretendería de forma intuitiva. Esta forma de ponderar las diferentes jerarquías podría ser una técnica para resaltar bandas determinadas de frecuencia, de modo que el análisis tenga una mayor similitud con el sistema auditivo humano, del cual se sabe que posee mayor sensibilidad en ciertos rangos de frecuencia. Tal cual está planteado nuestro modelo, esto sería relativamente fácil de implementar, ya que solo consiste en modificar valores en la parte algorítmica de la programación hecha, más precisamente en la función que calcula la distancia entre dos vectores.

Como reflexión final, puede decirse que si bien aún hoy no es posible crear reconocedores o clasificadores de instrumentos que cuenten con la misma capacidad que un oyente humano, indudablemente, con el desarrollo de nuevas técnicas (tanto de procesamiento de señales como de aprendizaje automático) será posible contar en un futuro con sistemas computacionales que efectúen tareas que hoy solo están dentro de la imaginación, como pueden ser sistemas de transcripción automática en tiempo real de presentaciones en vivo de conjuntos musicales, sistemas de arreglos y acompañamiento automático en tiempo real que reaccionen al instrumento que se está ejecutando en un momento determinado, y sistemas multimedia de mayor complejidad a los que se puede encontrar actualmente.

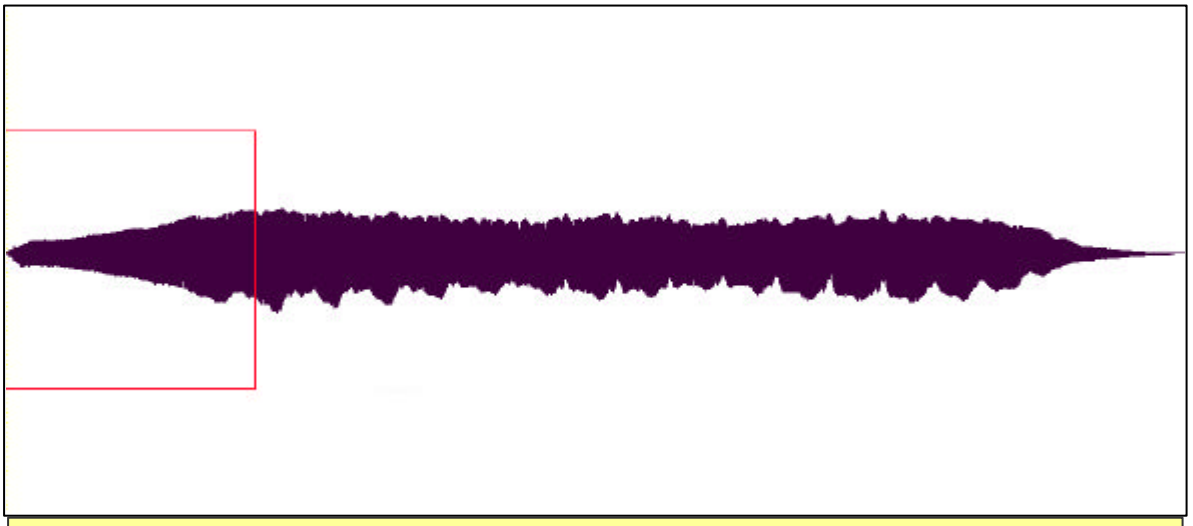
Anexos

Anexo A. Ejemplo de pre-procesamiento de algunos sonidos

Este anexo presenta algunos instrumentos, de familias diferentes, y el procesamiento por el que pasan antes de ingresar a las redes de Kohonen.

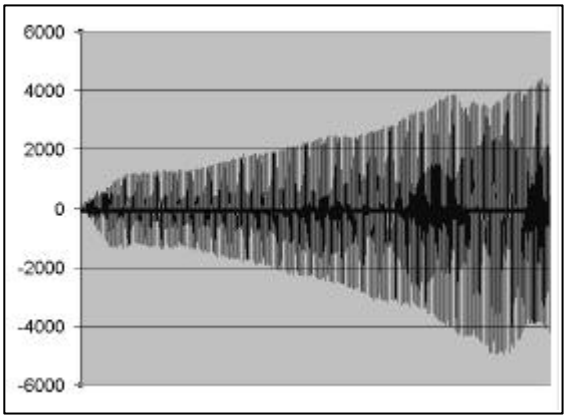
Nota: En todos los gráficos de este anexo el eje vertical representa la intensidad y el eje horizontal representa el tiempo.

1. Violín

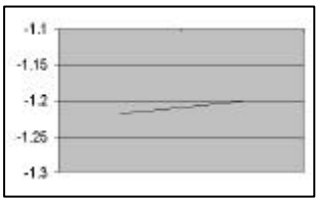


Violín	Longitud del sonido original: 3.56 seg.	Zona recuadrada: 743 ms
	Amplitud máxima del Sonido: 4973 (sobre 32767 que es el máximo)	

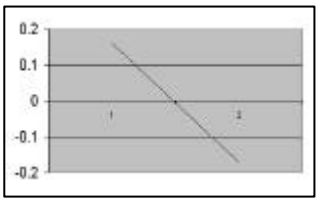
Las 15 Jerarquías generadas por Wavelet



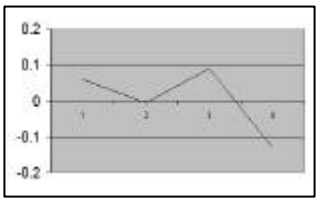
743 ms del Violín



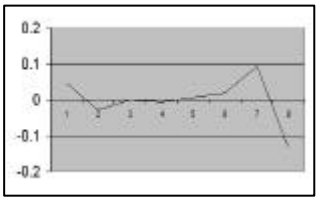
Jerarquía 1 Coeficientes Madre



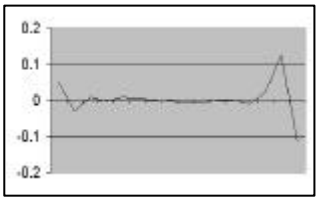
Jerarquía 2 de 0 a 2.7 Hz



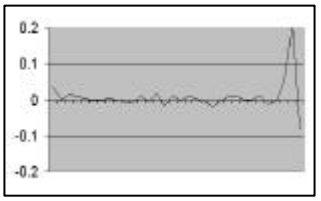
Jerarquía 3 de 2.7 a 5.3 Hz



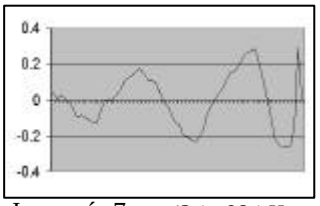
Jerarquía 4 de 5.3 a 10.7 Hz



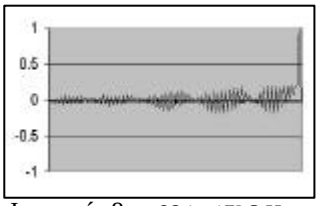
Jerarquía 5 10.7 - 21.5 Hz



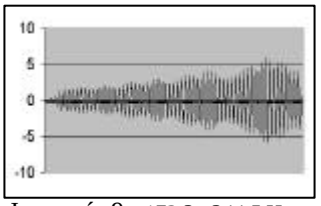
Jerarquía 6 21.5 - 43.1 Hz



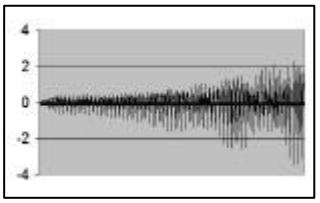
Jerarquía 7 43.1 - 86.1 Hz



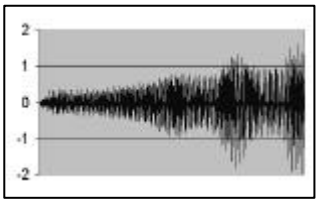
Jerarquía 8 86.1 - 172.3 Hz



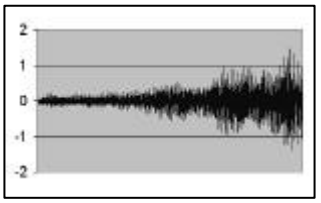
Jerarquía 9 172.3 - 344.5 Hz



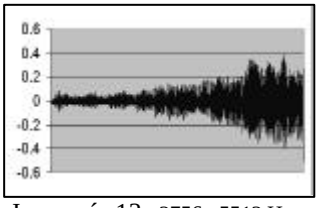
Jerarquía 10 344.5 - 689 Hz



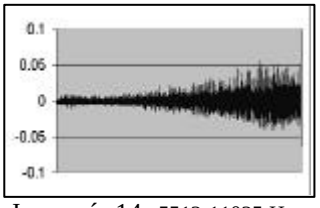
Jerarquía 11 689 - 1378 Hz



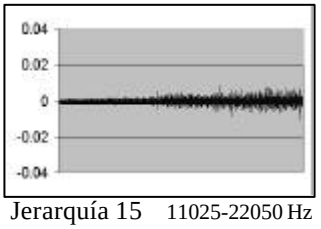
Jerarquía 12 1378 - 2756 Hz



Jerarquía 13 2756 - 5512 Hz

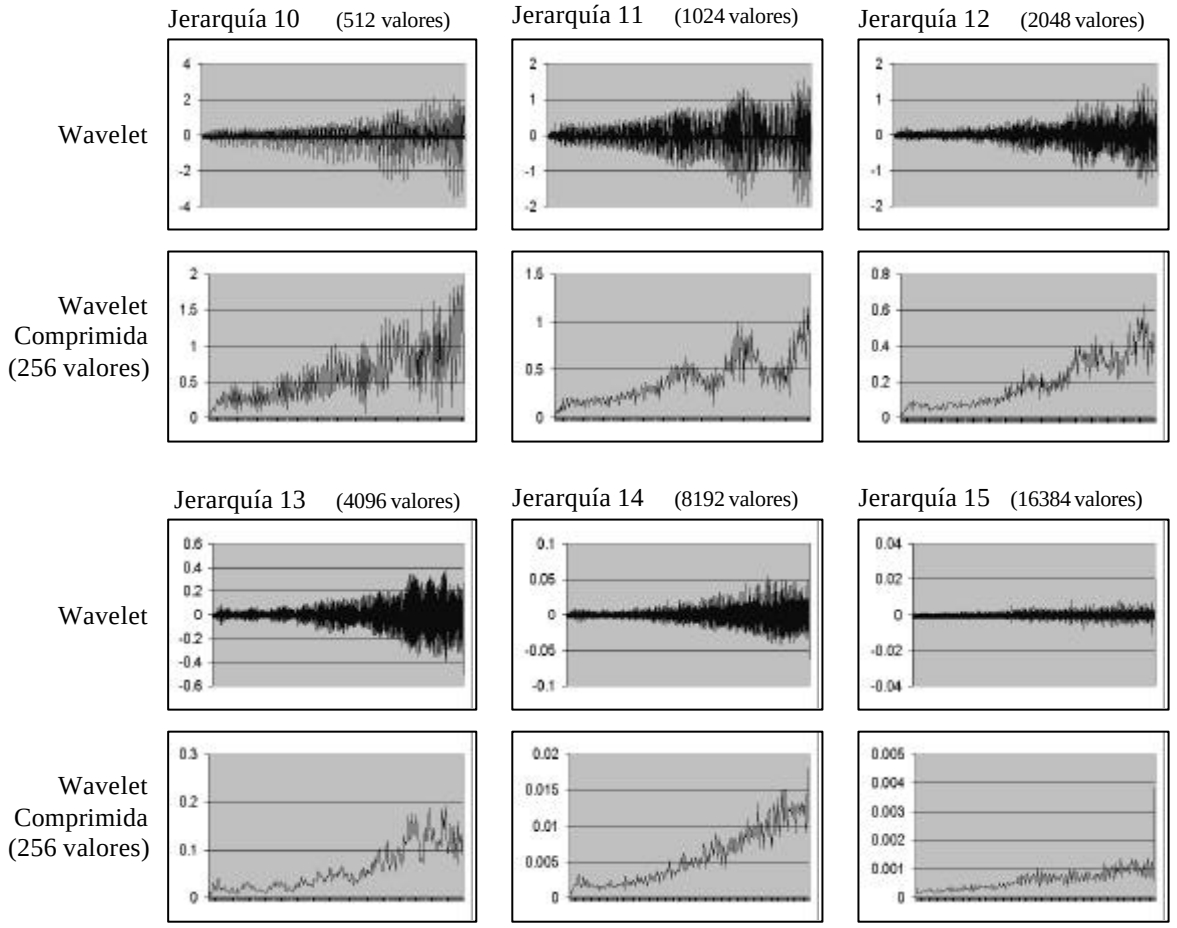


Jerarquía 14 5512-11025 Hz

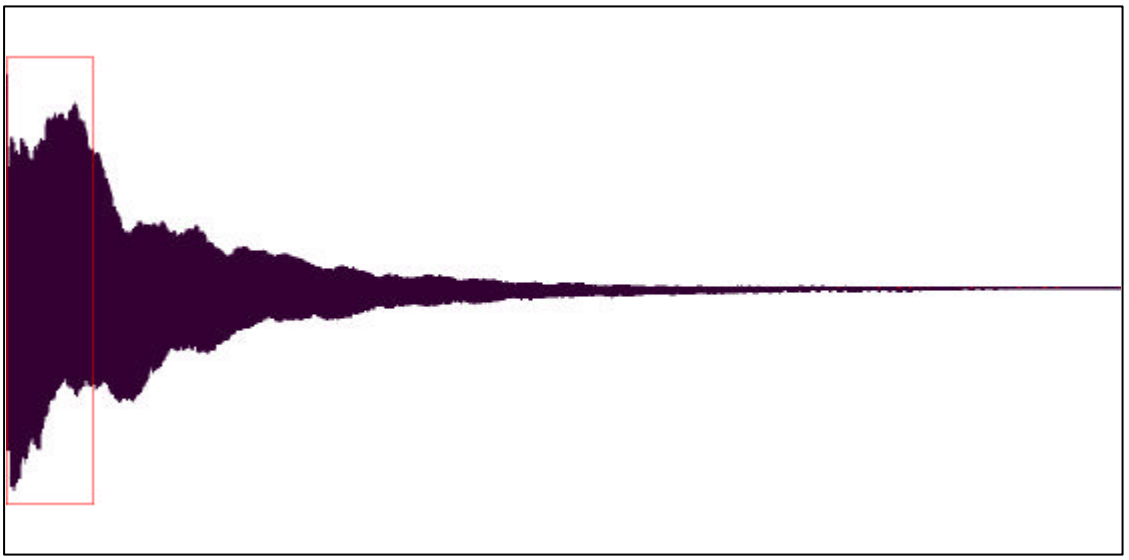


Jerarquía 15 11025-22050 Hz

Violín - Compresión de las Jerarquías Superiores



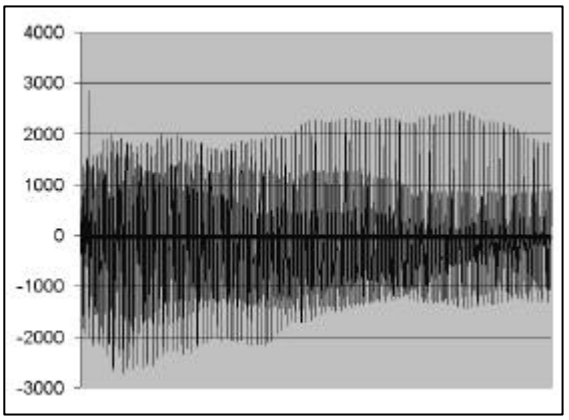
2. Piano 1



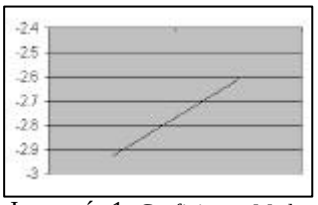
Piano 1

Longitud del sonido original: 9.5 seg. (tiene 15 segundos, pero en el gráfico solo están dibujados los primeroes 9.5 seg.)
Zona recuadrada: 743 ms
Amplitud máxima del Sonido: 2830 (sobre 32767 que es el máximo)

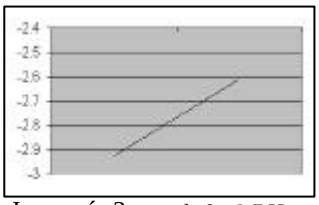
Las 15 Jerarquías generadas por Wavelet



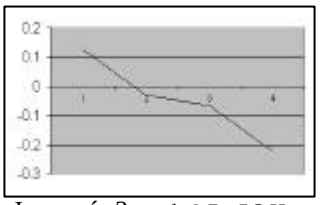
743 ms del Piano 1



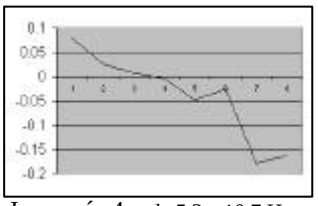
Jerarquía 1 Coeficientes Madre



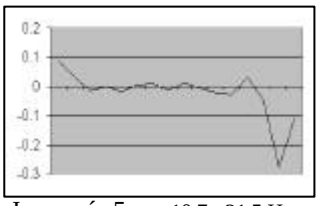
Jerarquía 2 de 0 a 2.7 Hz



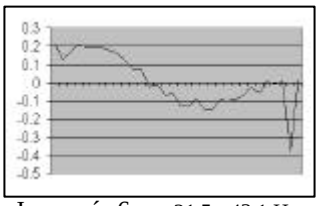
Jerarquía 3 de 2.7 a 5.3 Hz



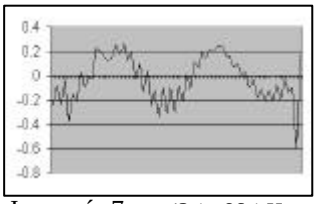
Jerarquía 4 de 5.3 a 10.7 Hz



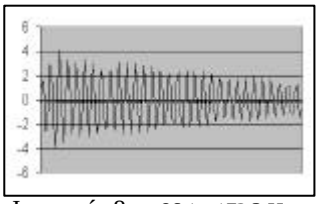
Jerarquía 5 10.7 - 21.5 Hz



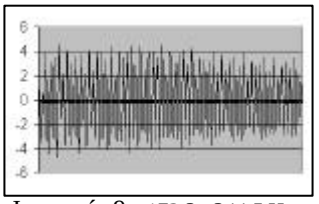
Jerarquía 6 21.5 - 43.1 Hz



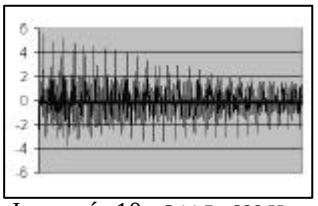
Jerarquía 7 43.1 - 86.1 Hz



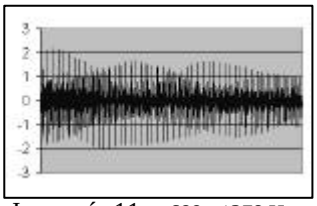
Jerarquía 8 86.1 - 172.3 Hz



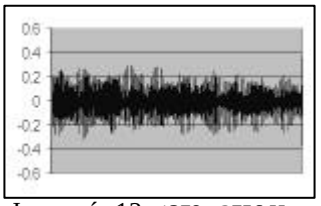
Jerarquía 9 172.3 - 344.5 Hz



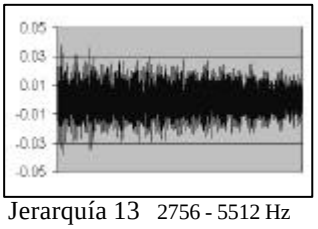
Jerarquía 10 344.5 - 689 Hz



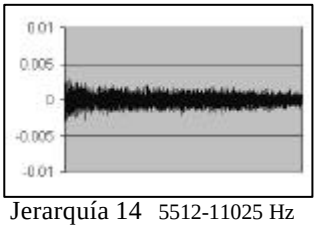
Jerarquía 11 689 - 1378 Hz



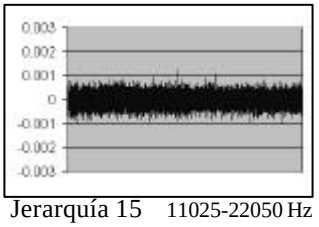
Jerarquía 12 1378 - 2756 Hz



Jerarquía 13 2756 - 5512 Hz

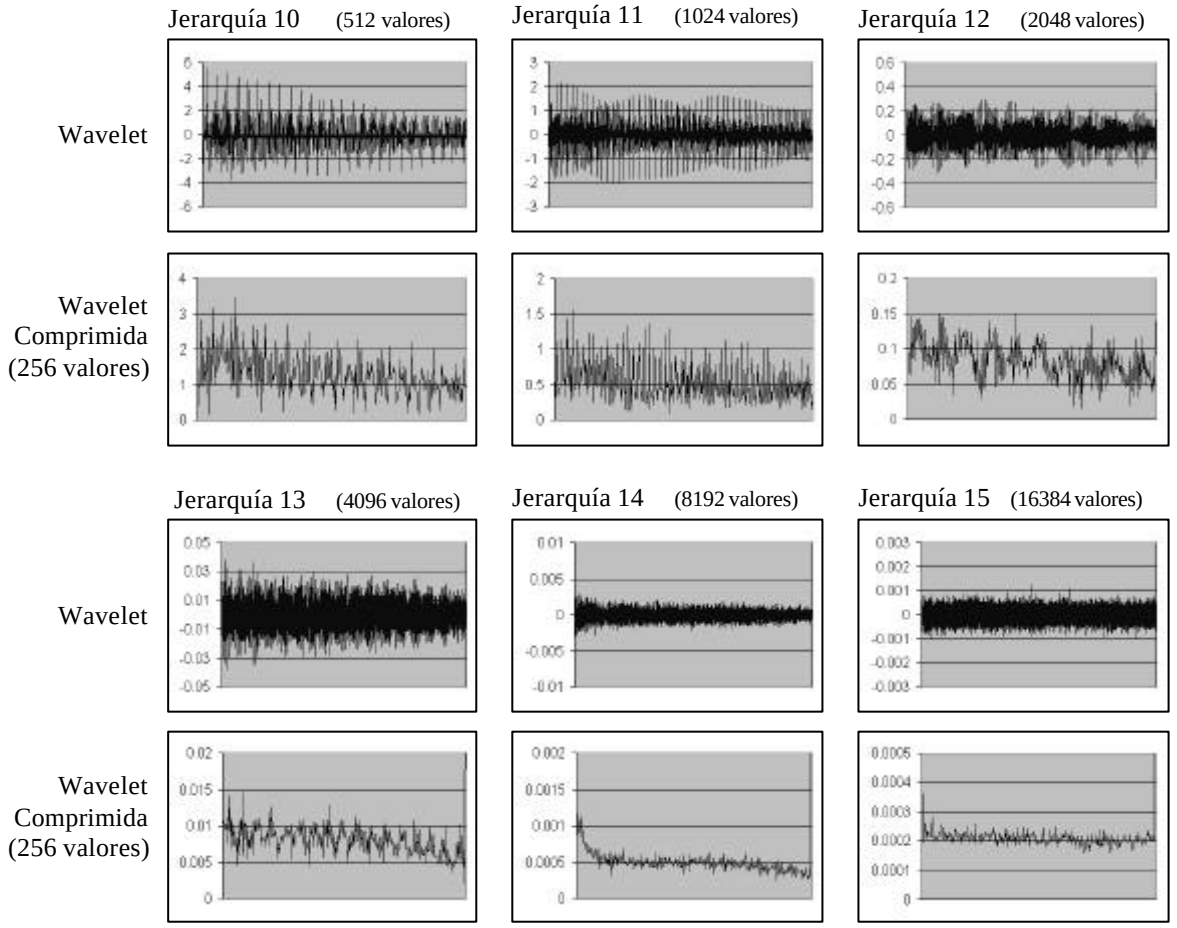


Jerarquía 14 5512-11025 Hz

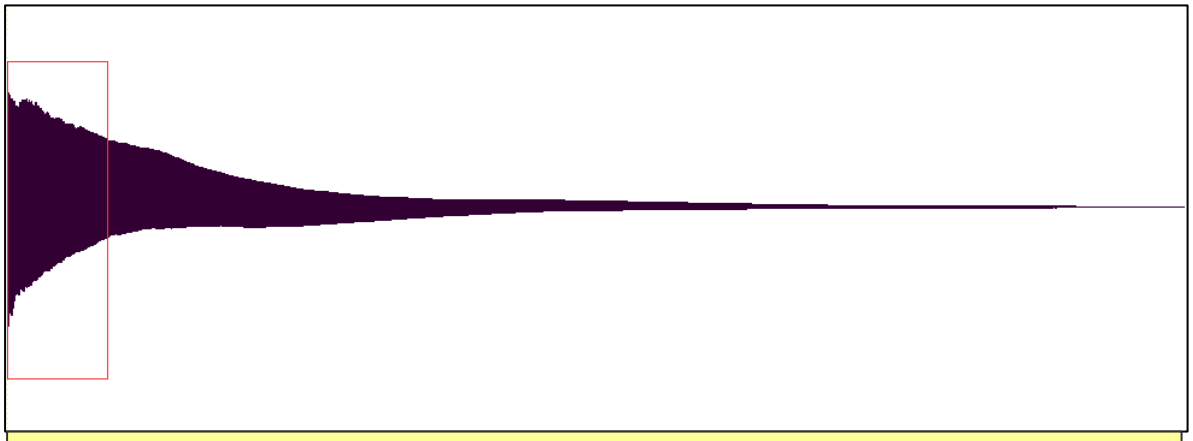


Jerarquía 15 11025-22050 Hz

Piano 1 - Compresión de las Jerarquías Superiores



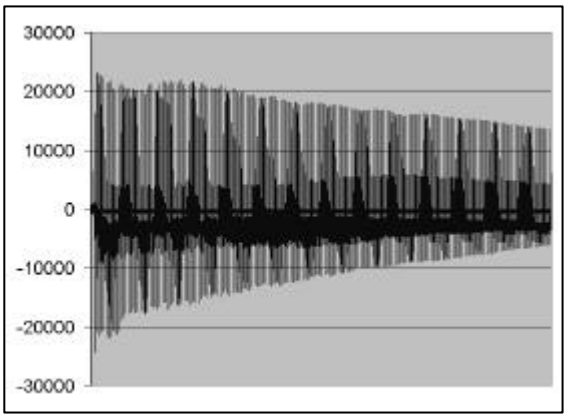
3. Guitarra Eléctrica



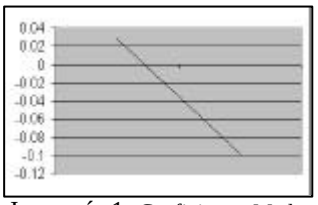
Piano 1

Longitud del sonido original: 8 seg. (tiene 12 segundos, pero en el gráfico solo están dibujados los primeroes 8 seg.)
Zona recuadrada: 743 ms
Amplitud máxima del Sonido: 24270 (sobre 32767 que es el máximo)

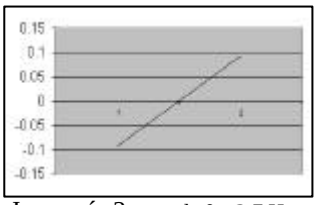
Las 15 Jerarquías generadas por Wavelet



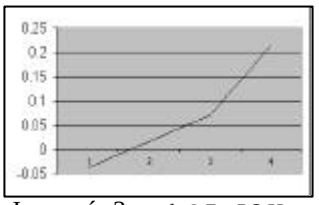
743 ms de la Guitarra Eléctrica



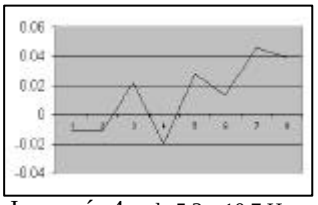
Jerarquía 1 Coeficientes Madre



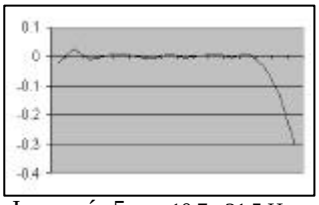
Jerarquía 2 de 0 a 2.7 Hz



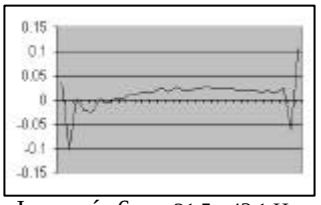
Jerarquía 3 de 2.7 a 5.3 Hz



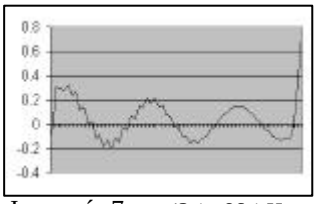
Jerarquía 4 de 5.3 a 10.7 Hz



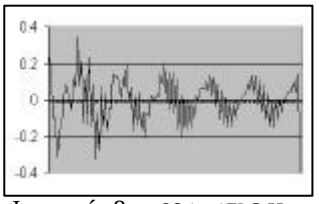
Jerarquía 5 10.7 - 21.5 Hz



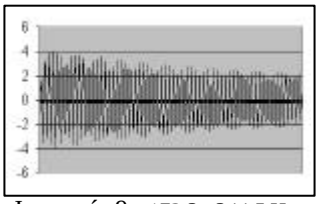
Jerarquía 6 21.5 - 43.1 Hz



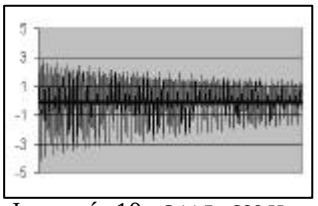
Jerarquía 7 43.1 - 86.1 Hz



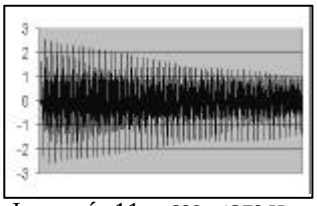
Jerarquía 8 86.1 - 172.3 Hz



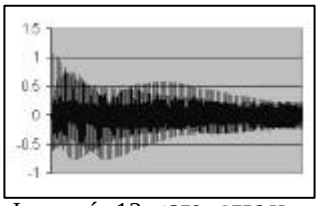
Jerarquía 9 172.3 - 344.5 Hz



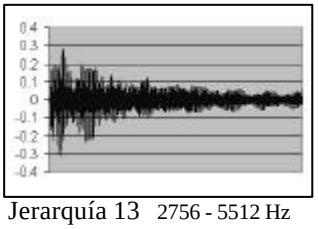
Jerarquía 10 344.5 - 689 Hz



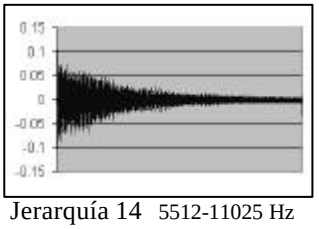
Jerarquía 11 689 - 1378 Hz



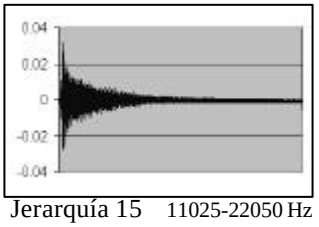
Jerarquía 12 1378 - 2756 Hz



Jerarquía 13 2756 - 5512 Hz

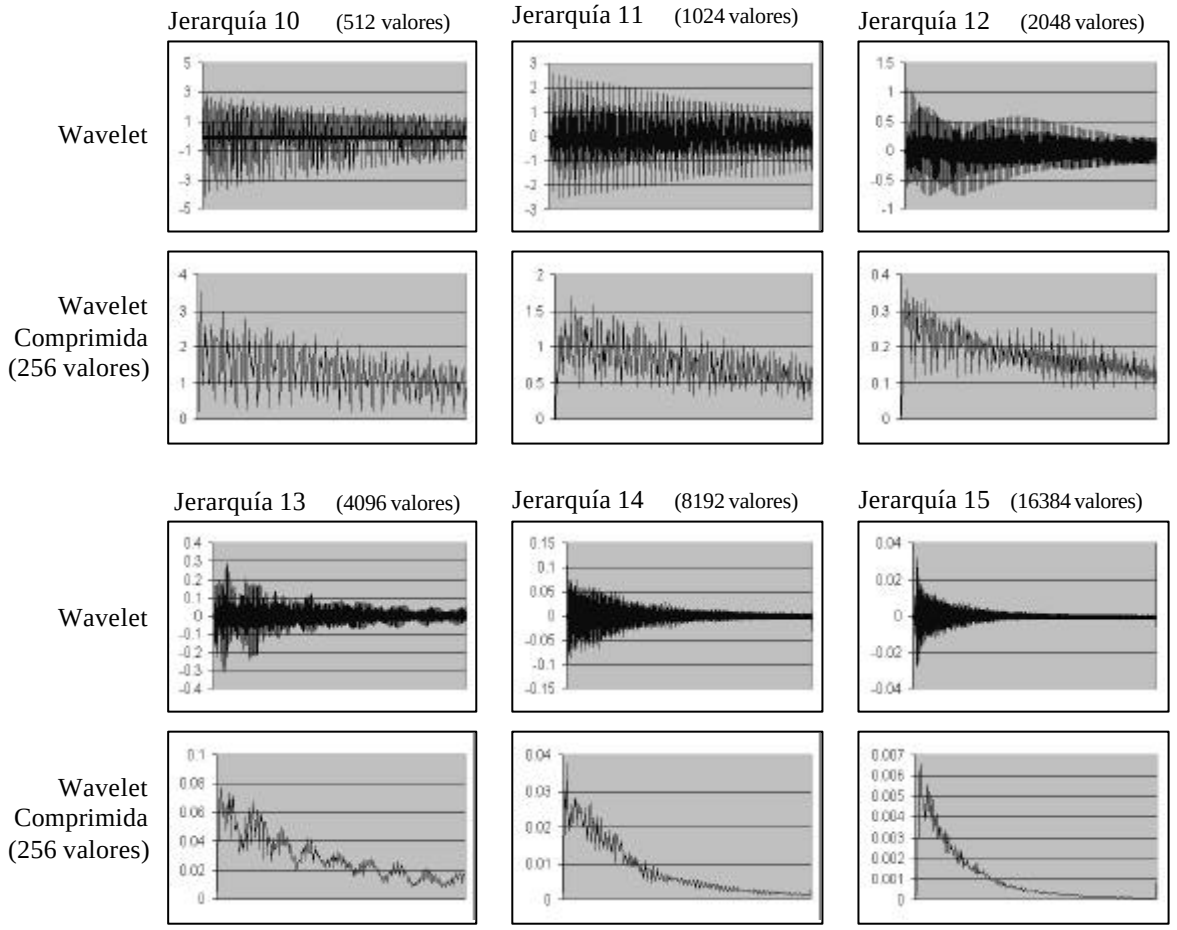


Jerarquía 14 5512-11025 Hz

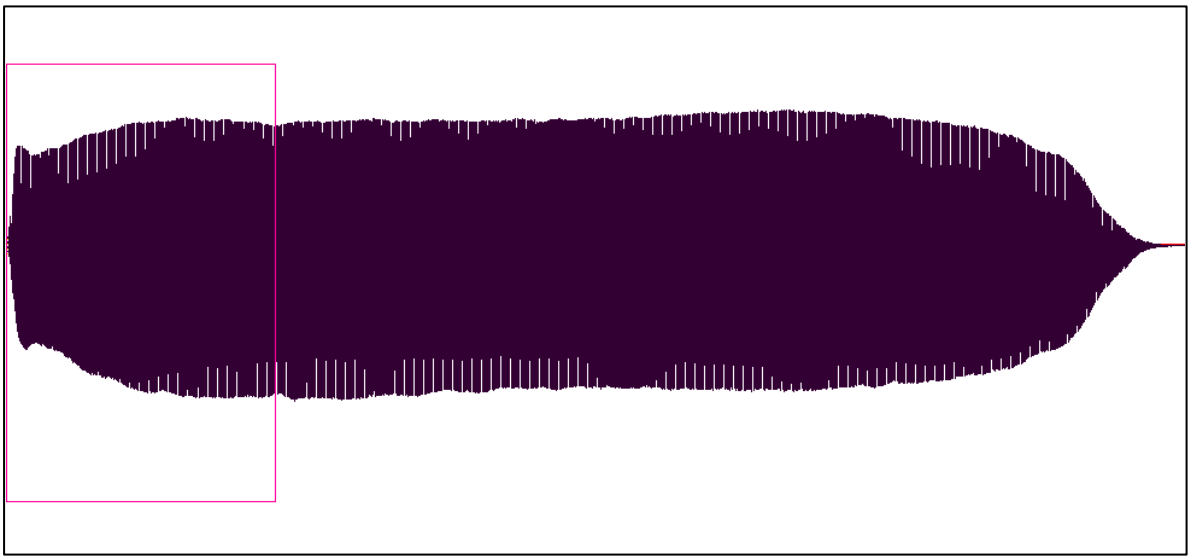


Jerarquía 15 11025-22050 Hz

Guitarra Eléctrica - Compresión de las Jerarquías Superiores

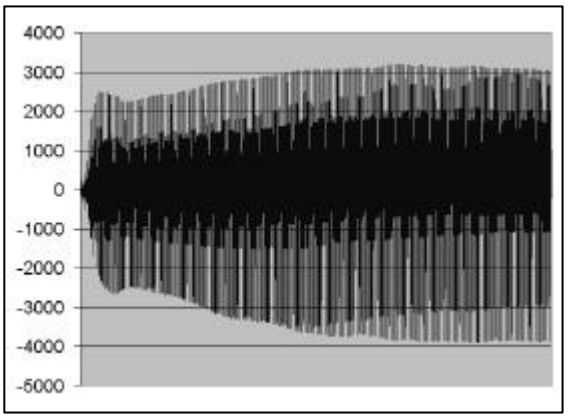


4. Oboe

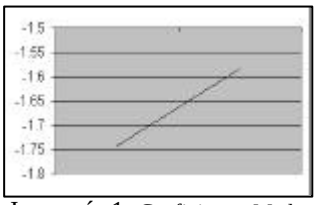


Oboe Longitud del sonido original: 3.5 seg. Zona recuadrada: 743 ms
Amplitud máxima del Sonido: 3960 (sobre 32767 que es el máximo)

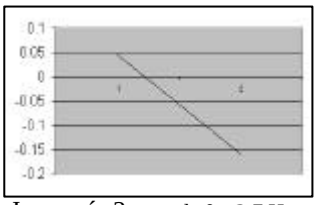
Las 15 Jerarquías generadas por Wavelet



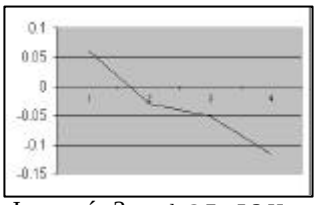
743 ms del Oboe



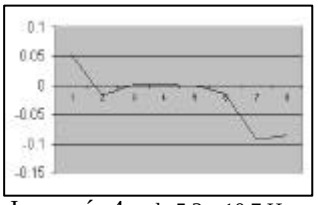
Jerarquía 1 Coeficientes Madre



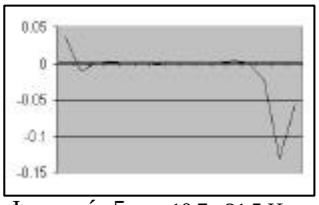
Jerarquía 2 de 0 a 2.7 Hz



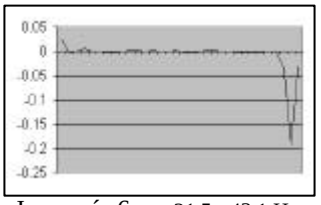
Jerarquía 3 de 2.7 a 5.3 Hz



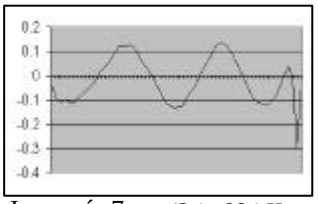
Jerarquía 4 de 5.3 a 10.7 Hz



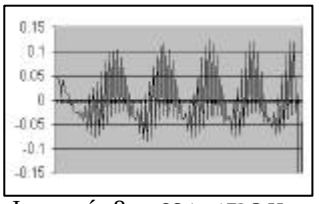
Jerarquía 5 10.7 - 21.5 Hz



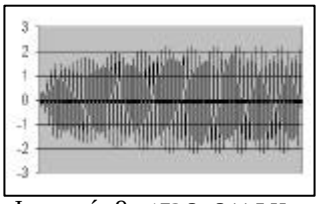
Jerarquía 6 21.5 - 43.1 Hz



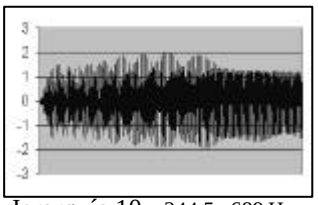
Jerarquía 7 43.1 - 86.1 Hz



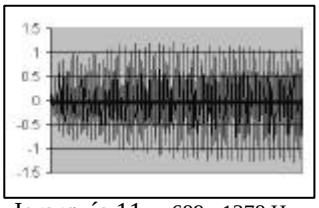
Jerarquía 8 86.1 - 172.3 Hz



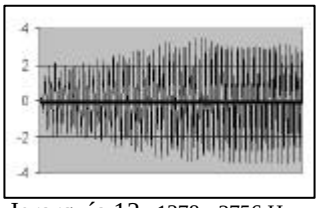
Jerarquía 9 172.3 - 344.5 Hz



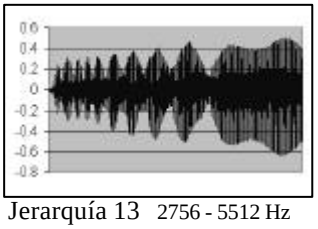
Jerarquía 10 344.5 - 689 Hz



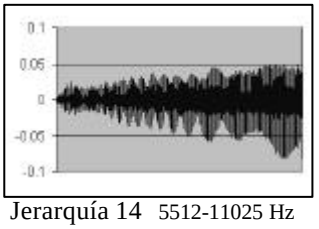
Jerarquía 11 689 - 1378 Hz



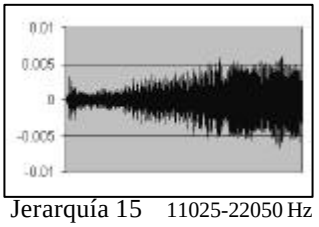
Jerarquía 12 1378 - 2756 Hz



Jerarquía 13 2756 - 5512 Hz

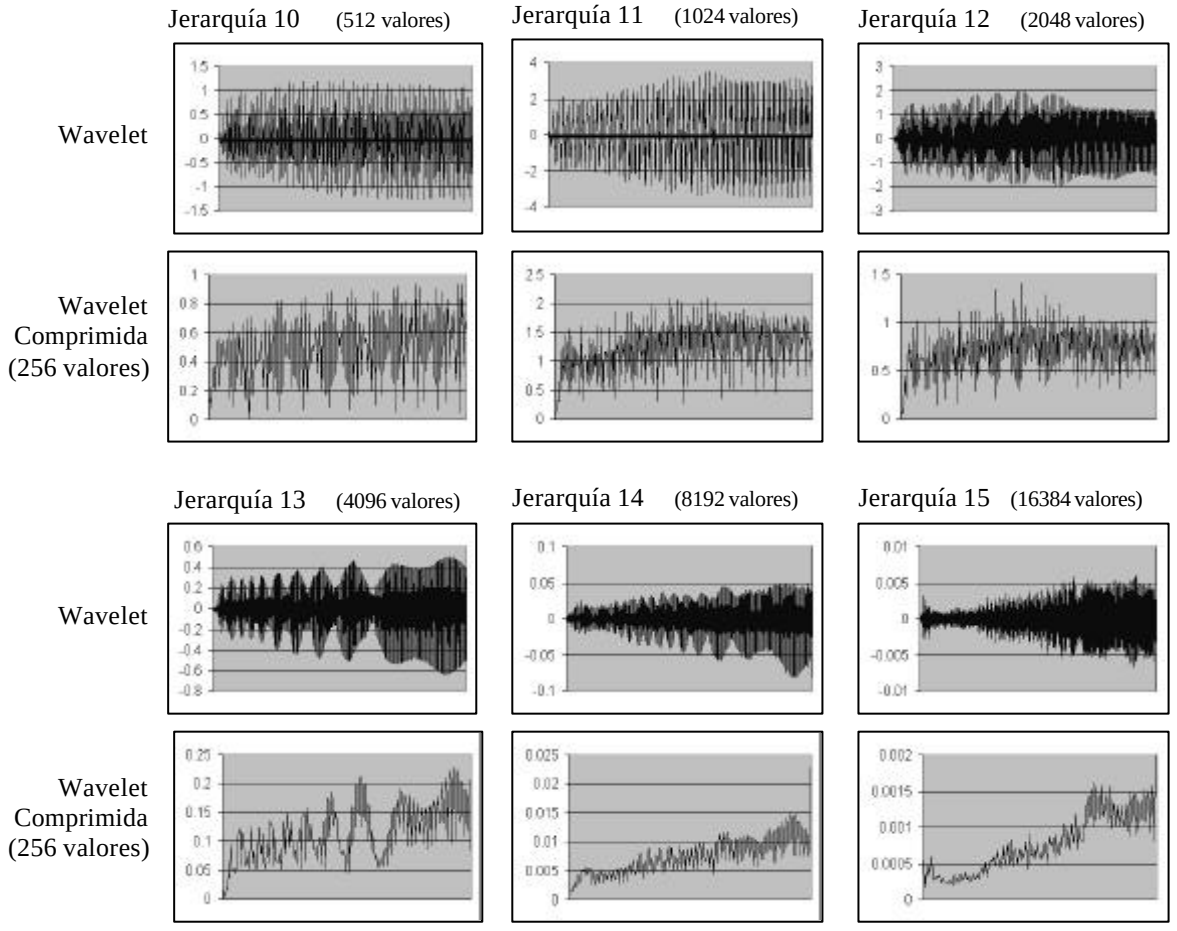


Jerarquía 14 5512-11025 Hz



Jerarquía 15 11025-22050 Hz

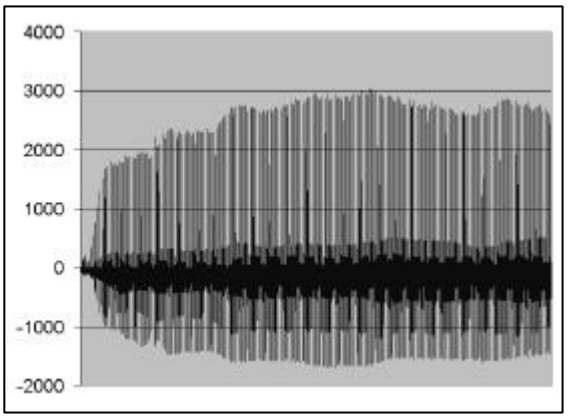
Oboe - Compresión de las Jerarquías Superiores



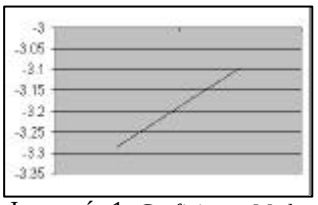
5. Trompeta

Trompeta Longitud del sonido original: 7.4 seg. Zona recuadrada: 743 ms
Amplitud máxima del Sonido: 3040 (sobre 32767 que es el máximo)

Las 15 Jerarquías generadas por Wavelet

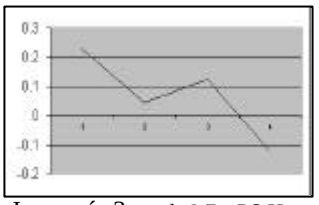


743 ms de la Trompeta

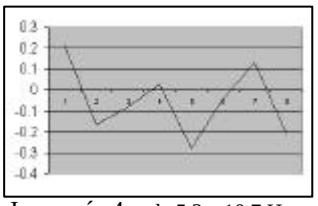


Jerarquía 1 Coeficientes Madre

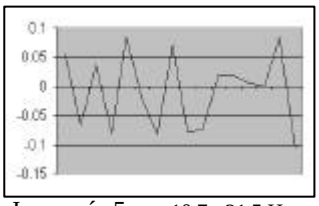
Jerarquía 2 de 0 a 2.7 Hz



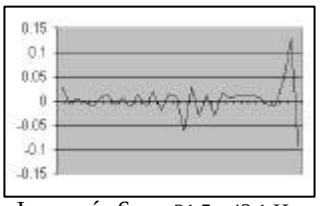
Jerarquía 3 de 2.7 a 5.3 Hz



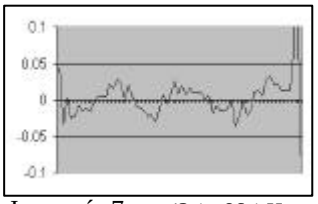
Jerarquía 4 de 5.3 a 10.7 Hz



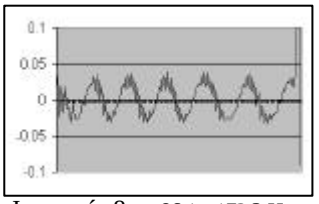
Jerarquía 5 10.7 - 21.5 Hz



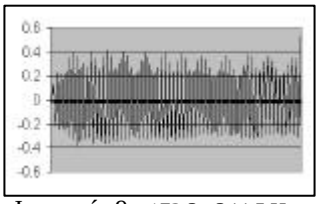
Jerarquía 6 21.5 - 43.1 Hz



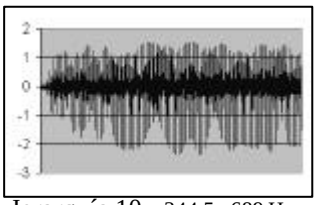
Jerarquía 7 43.1 - 86.1 Hz



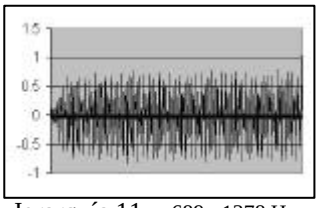
Jerarquía 8 86.1 - 172.3 Hz



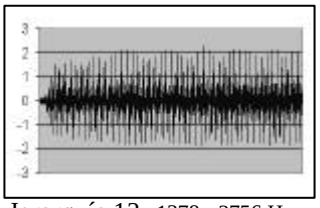
Jerarquía 9 172.3 - 344.5 Hz



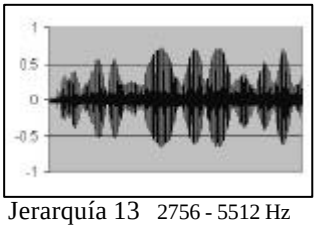
Jerarquía 10 344.5 - 689 Hz



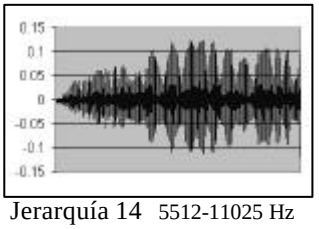
Jerarquía 11 689 - 1378 Hz



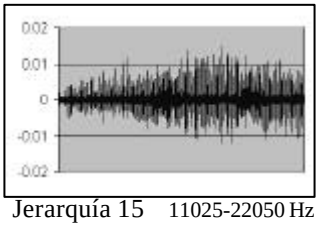
Jerarquía 12 1378 - 2756 Hz



Jerarquía 13 2756 - 5512 Hz

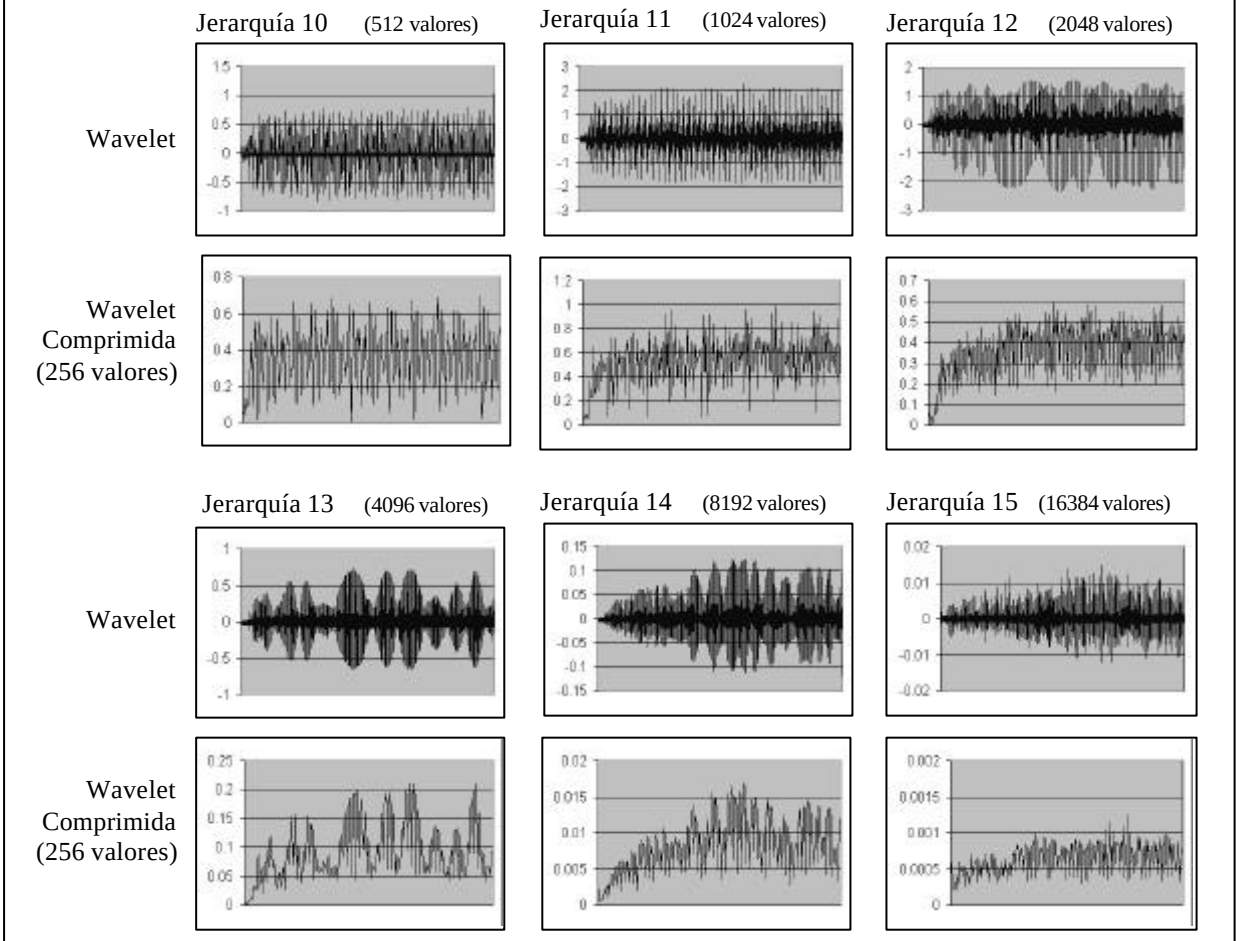


Jerarquía 14 5512-11025 Hz



Jerarquía 15 11025-22050 Hz

Trompeta - Compresión de las Jerarquías Superiores



Anexo B. MUMs

Las grabaciones de la McGill University (MUMs, McGill University Master Samples) fueron realizadas por Frank Opolko and Joel Wapnick en 1987 [OPO/1987], con la idea de registrar todos los instrumentos posibles, en toda su extensión cromática, con todas sus variantes tímbricas y utilizando instrumentos musicales de excelente nivel. Todas las grabaciones ser realizaron en equipos de última tecnología para representar el sonido de los instrumentos lo mas fielmente posible.

Todas las grabaciones están incluidas en 11 CDs, y han sido utilizadas para muchos estudios y análisis sobre sonidos de instrumentos musicales.

Los sonidos del violín, la viola, el violoncello, el contrabajo, la flauta, los oboes, los cuernos, los clarinetes, fagotes, trompetas, trombones, tubas y el saxo tenor, fueron grabados a un grabador "Sony" PCM 3202 DASH (A/D rate: 44.1 Khz, sin énfasis). Los sonidos de los pianos, las marimbas, el vibráfono y las campanas tubulares fueron grabados con el mismo equipamiento, pero agregándole un efecto de "ambiente" en el canal derecho.

Se utilizaron micrófonos de condensación de alta calidad marca "B & K" (modelo 4003), junto con preamplificadores del mismo fabricante. En algunos casos se adosaron un par de micrófonos de condensación "Sennheiser" MKH40. La salida de estos micrófonos se pasó luego a través de una consola de mezcla automática “Sony” MCI 3630.

Los sonidos del bajo eléctrico, la guitarra eléctrica, bajo acústico, la trompeta en si bemol asordinada y el vibráfono con baqueta suave fueron grabados directamente a una maquina "Sony" R-DAT (modelo 2500). Se usaron los mismos micrófonos y preamplificadores "B & K" que en los sonidos anteriores.

Para grabar el órgano de tubos se agregaron micrófonos omnidireccionales "B & K" (modelos 4006 y 4007).

Durante la fase de grabación se trató de eliminar en la mayor medida posible el uso de procesadores de nivel de sonido (como compresores y limitadores de señales) para preservar al máximo la personalidad y ataque de cada instrumento. Las grabaciones se realizaron en diferentes tipos de ambientes acústicos, de forma de lograr el mejor entorno para cada tipo de instrumento.

Material de referencia

[ARF/1991] Arfib D. "Analysis, Transformation, and Resynthesis of Musical Sounds with the help of a Time-Frequency representation", en Representations of Musical Signals (editores: De Poli G., Piccialli A. Y Roads C.), The MIT PRESS, Cambridge, Massachusets, - pag. 87-118. 1991

[AME/1960] "American Standard Acoustical Terminology", American Standards Association, Inc. , New York. 1960

[BOE/1999] Boersma P, Weenink D "Praat, a system for doing phonetics by computer", Software para procesamiento de Señales. 1999.

[COS/1998] Cosi P. "Auditory modeling and neural networks", unpublished paper. 1998.

[DEP/1997] De Poli G., Prandoni P. "Sonological model for timbre characterization", Journal of New Music Research, Vol.26, n.2. 1997.

[EVA/1991] Evangelista G. "Wavelet transforms that we can play", en Representations of Musical Signals (editores: De Poli G., Piccialli A. y Roads C.), The MIT PRESS, Cambridge, Massachusets, - pag. 119-136. 1991

[FRA/1998] Fraser A., Fujinaga I. "Toward real-time recognition of acoustic musical instruments", unpublished paper. 1998.

[FUJ/1998] Fujinaga I. "Machine recognition of timbre using steady-state tone of acoustical musical instruments", Proceedings of the International Computer Music Conference. 1998.

[GRE/1975] Grey J. "An exploration of musical timbre" Ph.D. Dissertation, Dept. of Music, Report No. STAN-M-2, Stanford University, Stanford, CA. 1975.

[GRE/1977] Grey J. "Multidimensional perceptual scaling of musical timbres", Journal of the Acoustical Society of America, Vol.61 n.5 – pag. 1270-1277.1977.

[GRE/1978] Grey J. "Timbre discrimination in musical patterns", Journal of the Acoustical Society of America, Vol.64 n.2 – pag. 467-472.1978.

[HEL/1954] Helmholtz H. L. F. "On the sensations of tone as a physiological basis for the theory of music", Ellis A. J. traductor, Dover, New York. 1954.

[HER/1991a] Hertz J., Krogh A., Palmer R. "Introduction to the theory of neural computation", Addison-Wesley Publishing Company. Chapter 8, Unsupervised Hebbian Learning. 1991

[HER/1991b] Hertz J., Krogh A., Palmer R. "Introduction to the theory of neural computation", Addison-Wesley Publishing Company. Chapter 9, Unsupervised Competitive Learning. 1991

[HLA/1992] Hlawatsch F., Boudreaux-Bartels G.F. "Linear and quadratic time-frequency signal representations", IEEE Signal Processing Magazine, April – pag.21-67. 1992.

[KOH/1982] Kohonen T., “Clustering, taxonomy, and topological maps of patterns”, Proceedings of the 6th international conference on pattern recognition. (Piscataway, NJ: IEEE) - pag. 114-128. 1982

[KOH/1988] Kohonen T., “Self Organization and Associative Memory”, Springer Verlag, Berlin, Heidelberg, New York, 2nd edition. 1988

[KRO/1991] Kronland-Martinet R., Grossmann A. "Application of Time-Frequency and Time-Scale Methods (Wavelet Transforms) to the Analysis, Synthesis, and Transformation of Natural Sounds", en Representations of Musical Signals (editores: De Poli G., Piccialli A. Y Roads C.), The MIT PRESS, Cambridge, Massachusets, - pag. 45-85. 1991

[LAD/1989] Laden B., Keefe D. "The representation of pitch in a neural net model of chord Classification", Computer Music Journal, Vol.13, no. 4, Winter 1989, reimpreso en Music and Connectionism (editores: Todd P., Loy D.), The MIT PRESS, Cambridge, Massachusets, - pag. 64-83. 1991

[LEM/1991] Leman M. "The ontogenesis of tonal semantics: results of a computer study", Music and Connectionism (editores: Todd P., Loy D.), The MIT PRESS, Cambridge, Massachusets, - pag. 100-127. 1991

[LIC/1951] Licklider,J.C.R. "Basic Correlates of the Auditory Stimulus ", Handbook of Experimental Psychology, S. S. Stevens, ed. Wiley, New York . 1951, tal como se cita en [GRE/1975].

[MAR/1998a] Martin K., Scheirer E., Vercoe B. “Music content analysis through models of auditions”, presented at ACM Multimedia´98, Workshop on Content Processing of Music for Multimedia Applications, Bristol UK, 12 September. 1998.

[MAR/1998b] Martin K., Youngmoo E. “Musical instrument identification: A pattern-recognition approach”, Session presented at the136th meeting of the Acoustical Society of America, October 13th. 1998

[MAR/1998c] Martin K. “Toward automatic sound source recognition: Identifying musical instruments”, Presented at the NATO Computational Hearing Advanced Study Institute, Il Ciocco, Italy, July 1-12. 1998.

[OPO/1987] Opolko F., Wapnick J. “McGill University Master Samples [Compact Disc]” , McGill University, Quebec Montreal, Vol.1, 2, 3, 5 & 8.1987.

[PRA/1994] Prandoni P. “An analysis-based timbre space”, unpublished paper. 1994.

[PRE/1992] Press William H., Teukolsky Saul A., Vetterling William T., Flannery Brian P. "Numerical Recipes in C, The Art of Scientific Computing, Second Edition", Cambridge University Press, Cambridge - Capítulo 13, Sección 10. 1992 (con correcciones en 1993, 1994,1995 y 1997)

[RIO/1991] Rioul O., Vetterli M. “Wavelets and signal processing”, IEEE Signal Processing Magazine, October – pag.14-38. 1991.

[RIS/1991] Risset J "Timbre analysis by synthesis: Representations, Imitations, and Variants for Musical Composition", en Representations of Musical Signals (editores: De Poli G., Piccialli A. Y Roads C.), The MIT PRESS, Cambridge, Massachusets, - pag. 7-43. 1991

[SAN/1989] Sano H., Jenkins B. "A neural network model for pitch perception", Computer Music Journal, Vol.13, no. 3, Fall 1989, reimpreso en Music and Connectionism (editores: Todd P., Loy D.), The MIT PRESS, Cambridge, Massachusets, - pag. 42-53. 1991

[SUN/1991] Sundberg J., "The science of musical sounds", Academic Press, San Diego. 1991

[TAM/2000] Támara Patiño V. "Compresión de Señales empleando wavelets", Departamento de Matemáticas, Facultad de Ciencias, Universidad de los Andes, Colombia. Paper sin publicar. 2000.

[TOI/1992] Toivianen P. "The organization of timbres: a two-stage neural network model", Workshop Notes of the ECAI 92 Workshop on Artificial Intelligence and Music (G. Widmer, Editor) Vienna: ECCAI. 1992.

[XIF/1994] Xifra J., Spedalieri N. "Organización del timbre de los instrumentos musicales" Trabajo Práctico de la Materia Laboratorio de Redes Neuronales, Carrera de Lic. en Cs. de la Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina. 1994.