

CLASIFICACION AUTOMÁTICA

“TRABAJO DE APLICACIONES DE AUDIO”

CARLOS ANDRES ORTIZ QUINTERO
NATALIA CARDONA BUSTAMANTE

DOCENTE:

RICARDO ZAPATA

13 DE AGOSTO

UNIVERSIDAD PONTIFICIA BOLIVARIANA

MEDELLIN

2008



1. CLASIFICACIÓN AUTOMÁTICA

La clasificación automática de archivos nace como una necesidad ante la abundancia de información en el medio en el que vivimos y nos desenvolvemos en nuestra vida diaria. Este trabajo hace un enfoque en la clasificación automática de archivos de audio.

La música a través de la historia ha estado presente en todas las culturas, dando como resultado la enorme variedad de música existente. A través de la historia el hombre ha hecho una clasificación musical creando etiquetas puramente subjetivas, claramente condicionada por factores culturales, sociales e históricos. Ante la existencia de tantos géneros musicales y después de haber conformado una biblioteca musical tan amplia se hace necesario una clasificación y organización.

Para realizar la clasificación de la música, en principio lo mas claro es hacer una clasificación por género. La gran cantidad de programas de software que hacen este tipo de clasificación, agrupan los archivos de audio en determinadas cantidades de géneros de todo tipo de música, pero la gran mayoría errados ya que solo existen 6 géneros en lo que podemos clasificar la música y no mas de 15 como lo sugieren dicho programas, esto según estudios de la Universidad de Alcalá, mediante el “Diseño y evaluación de herramientas para la caracterización automática de señales sonoras”.

Es entonces de gran interés estudiar las soluciones para lograr una buena técnica de clasificación, la clave radica en el reconocimiento de patrones de la señal acústica. La solución que generalmente contienen las implementaciones de un sistema de reconocimiento se basan en tres módulos principales:

- Procesamiento
- Extracción de características.
- Reconocimiento

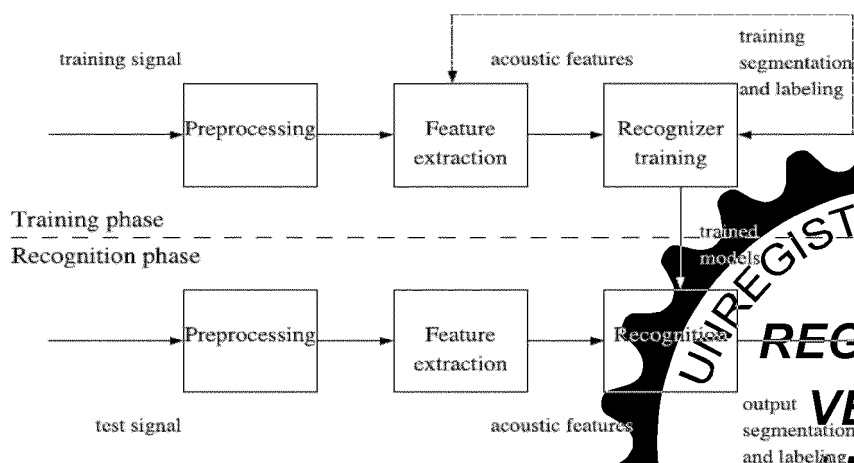


Figura 1. Sistema de reconocimiento de audio

La figura 1 muestra el diagrama general de un sistema de reconocimiento de audio. La figura dos muestra el diagrama de flujo del procesamiento de la señal de audio y la extracción de características



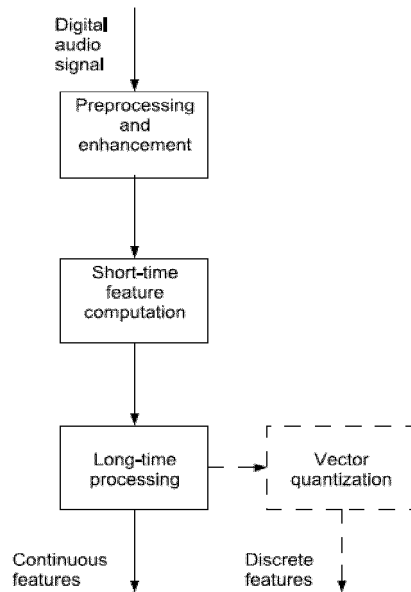
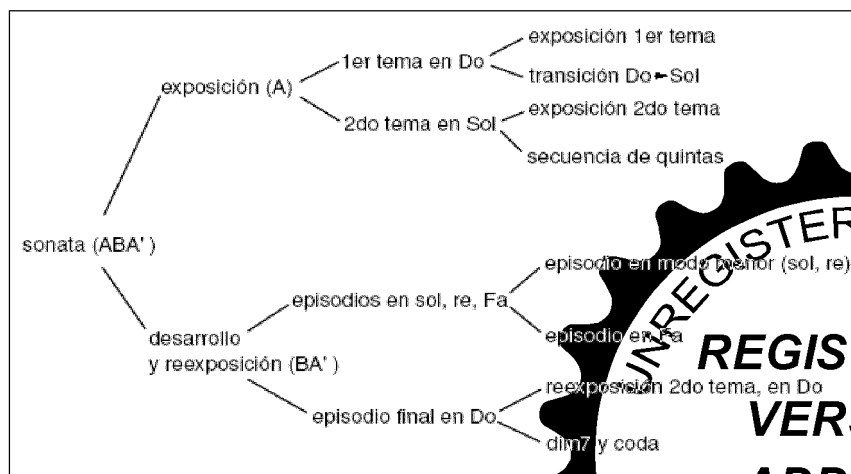


Figura 2. Flujo de datos de un extractor de características acústicas

En ocasiones para el estudio de las características de la música se hacen comparaciones con la literatura. Un dato interesante es que según la teoría de G. K. Zipf se dice que: “el número de repeticiones de la n -ésima palabra más frecuente es, muy aproximadamente, inversamente proporcional a n . Si, por ejemplo, la décima palabra más usada en un texto aparece 300 veces, la ley de Zipf predice que la centésima aparecerá unas 30 veces.” Esto también podría funcionar en la música respecto a su altura y duración o también por tríos de notas, acordes y sucesiones de intervalos. Se puede relacionar música con literatura ya que existe en una la gramática, morfológica y semántica, y en la otra existe la melodía, armónica y dinámica en la música. Para todo este análisis lo más importante es el estudio de las señales segmentadas. Un ejemplo claro de segmentación es el primer movimiento de la sonata para piano en do mayor (K. 545) de Mozart.



Resultado de los tres primeros pasos de segmentación del primer movimiento de la sonata K. 545 de Mozart partiendo de la estructura ternaria tradicional (forma sonata ABA') de Alperga.

Figura 3.



2. TECNICAS DE PROCESAMIENTO DE SEÑAL Y EXTRACCIÓN DE CARACTERÍSTICAS

- 2.1 Aproximaciones por procesamiento de tramas cortas de tiempo: se basa en dividir la señal en varios cuadros, segmentándola en cortas distancias igual espaciadas en el tiempo. Sobre estos segmentos de audio se aplican métodos de procesamiento, como la transformada discreta de Fourier (DFT) y el método de autocorrelación de predicción lineal.
- 2.2 Mejoras y Pre-procesamiento de la señal: Puede incluir filtrado con ajuste de pre-énfasis (en aplicaciones de voz), generalmente se utiliza un filtro FIR de primer orden, pasa altos, para enfatizar las altas frecuencias la voz. El pre-procesamiento también puede incluir operaciones de mejoramiento para remover ciertas distorsiones de la señal, La distorsión puede ser acústica (ambiental) ruido o canal ruidoso. Dos métodos conocidos para mejorar la señal son: Spectral subtraction y Wiener filtering.
- 2.3 Representación espectral de cada trama de tiempo: las principales herramientas de modelamiento espectral que son utilizadas para el reconocimiento voz/audio son, la transformada discreta de Fourier (DFT), códigos de predicción lineal (LP) y la representación Cepstral del espectro de magnitud.
- 2.4 Características perceptuales: El término "características perceptuales " se utiliza en el contexto actual para referirse a las características que tengan en cuenta los aspectos más humanos de la percepción auditiva. Algunas implementaciones que se basan en los parámetros relacionados con las características perceptuales son: Auditory Filterbanks, Mel Frequency Cepstral Coefficients (MFCC) y Perceptual linear prediction (PLP).
- 2.5 Características en tramas de tiempo especializadas: La mayor parte de parámetros utilizados en las implementaciones encontradas en la literatura se pueden dividir en tres tipos, dependiendo de si están relacionados con el timbre, el ritmo o el pitch de la señal. Parámetros mas comunes:
- Short-time energy STE: La STE se define como la energía media de la señal en cada trama de análisis.
 - Logarithmic short-time energy LGSTE. Es el logaritmo (generalmente en base 10) de STE.
 - Loudness (sonoridad) o LOUD. Parámetro basado en el modelo perceptual del sistema auditivo humano. Puede ser usado como una característica.
 - Función de autocorrelación.
 - ZCR (Zero Crossing Rate): EL ZCR o tasa de cruces por cero. Este parámetro nos da una idea de lo ruidosa que es la señal. Aunque de manera estricta no es un parámetro relacionado con el timbre, sino



temporal, sí que se suele incluir dentro de esta familia de parámetros, y tiene un gran interés para el problema de la discriminación voz/música.

- **Spectral Centroid:** Está definido como el “centro de gravedad” del espectro de magnitud, es decir, la frecuencia que divide el espectro en dos porciones aproximadamente iguales. Este parámetro se puede relacionar con la brillantez del sonido, por lo tanto es un importante constituyente de la percepción del timbre.
 - **HZCRR (High Zero-Crossing Rate Ratio):** Este parámetro, propuesto en [14], se calcula a partir de la tasa de cruces por cero definida anteriormente, y se define como el número de tramas cuya tasa de cruces por cero está 1.5 veces por encima de la tasa de cruces por cero media en una ventana que contiene M tramas.
 - **LSTER (Low Short-Time Energy Ratio):** De manera similar al HZCRR, el LSTER se obtiene a partir del STE, tomando en lugar del valor exacto, una medida de la variación del parámetro trama a trama. El LSTER se define como el número de tramas cuya STE sea menor que la mitad de la media del STE en una ventana que contenga M tramas.
 - **Spectral rolloff:** Se define como la frecuencia por debajo de la cual se concentra el PR% de la magnitud del espectro. Suele considerarse un valor típico de $PR = 0.85$. Se trata de un parámetro que nos da una idea de la forma del espectro.
 - **Spectral flux:** El spectral flux se define como el cuadrado de la diferencia entre las magnitudes de los espectros de dos tramas consecutivas. Se trata de una medida de la cantidad de cambios locales.
- 2.6 **Procesamiento de términos largos:** En este caso no se trabaja con segmentos de audio sino que se procesa el audio completo. La ventaja de este procesamiento es que no se altera el espectro como se hace con el procesamiento de tramas cortas, donde se modula la señal al aplicársele alguna ventana.
- 2.7 **Vector Quantization (VQ):** Después del procesamiento de tramas cortas y tramas largas de tiempo, se genera un vector que contiene una secuencia de características que representan la señal de audio, este vector es de la forma $\{X_n\}$, la dimensión d espacial del vector depende del número de características usado. Este vector juega un papel muy importante en los métodos de reconocimiento de patrones para clasificación de audio.



3. MÉTODOS DE RECONOCIMIENTO DE PATRONES Y CLASIFICACIÓN

Existen diversos métodos para reconocimiento de patrones y posteriormente clasificación del audio, algunos de ellos son:

- 3.1 Discriminantes lineales de Fisher, método sugerido por la Universidad de Alcalá, el cual estudia del timbre, el ritmo y el pitch de la señal.
- 3.2 Algoritmo K-NN (K- Nearest Neighbor), ilustrado en la figura 3. Es un método de clasificación muy simple aunque también muy potente. Para explicar su funcionamiento, supongamos que disponemos de un conjunto de entrenamiento con L vectores de características agrupados en C clases distintas. Para obtener la clase a la que pertenece un nuevo vector X basta con buscar los K vecinos más próximos al vector de test X, y ponderar las clases a las que pertenecen, generalmente utilizando un criterio de mayoría. Aunque es posible utilizar distintas medidas de distancia, en la mayor parte de los casos se opta por emplear la distancia euclidiana.

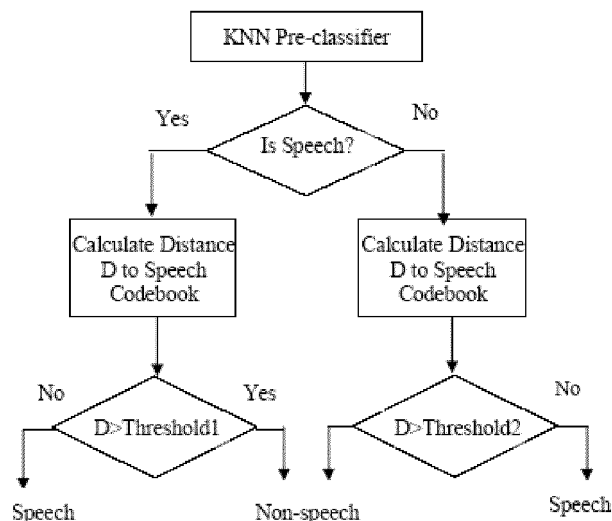


Figura 4. Algoritmo KNN

- 3.3 Métodos Bayesianos de clasificación: Usa métodos probabilísticos para clasificación y una función discriminante que determina a qué clase pertenece.
- 3.4 SVM (Suport Vector Machines): Permite clasificar las señales de audio además de la recuperación de datos. Este método no es solo usado en audio si no también en video y en ocasiones funciona mucho mejor que los algoritmos mencionado como el K-NN. Mas específicamente hablando es un procesamiento que se le hace ya sea a la voz, la música, otros para obtener los llamados vectores de características los cuales sintetizan el contenido frecuencias de los archivos musicales para clasificarlos. En algunas ocasiones las maquinas de soporte vectorial se emplean solo con voz, para comparar voces entre si y determinar a quien corresponde. Se usan entonces bases de registros de voz en donde están almacenadas las características vectoriales de los usuarios que serán comparados entre si.



para determinar a quien corresponde la voz. Hecha ya la extracción de todos los parámetros vectoriales se procede a la clasificación de las voces. De aquí se hacen las comparaciones y el análisis en tiempo real. El esquema que explica un poco más es el siguiente.

Lo mas importante de del trabajo con maquinas de soporte de vectores es que con un procedimiento mas arduo puede ayudar a clasificar canciones en cada uno de sus géneros musicales, pues se le puede extraer las características vectoriales a las canciones de ciertos géneros y con registros musicales se realizan comparaciones y obtener una clasificación completamente automática de canciones. Aunque de igual manera hay cabida para los porcentajes de error, pero si es una muy buena opción para desarrollar usando maquinas de soporte vectorial.

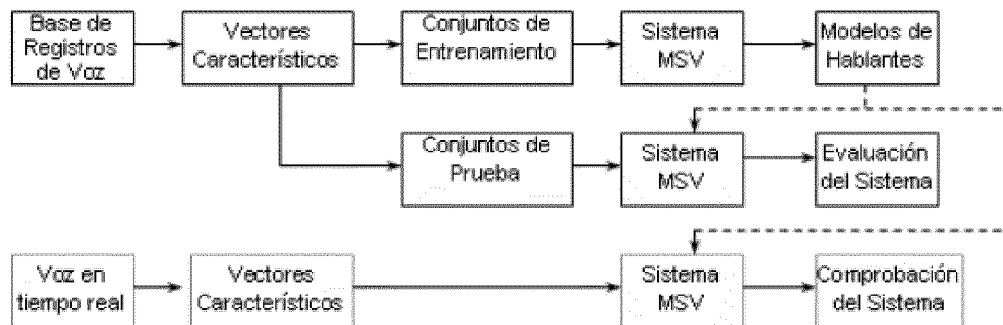


Figura 5. Método SVM

Tomado de http://mcyti.izt.uam.mx/Tesis_egresados/TesisGabrielPedroza.pdf

3.5 Otros métodos de gran interés son Gaussian mixture model (GMM), Hidden Markov model (HMM), Artificial Neural Networks y LSP.

La figura 6 ilustra un sistema completo de clasificación automática de señales de audio



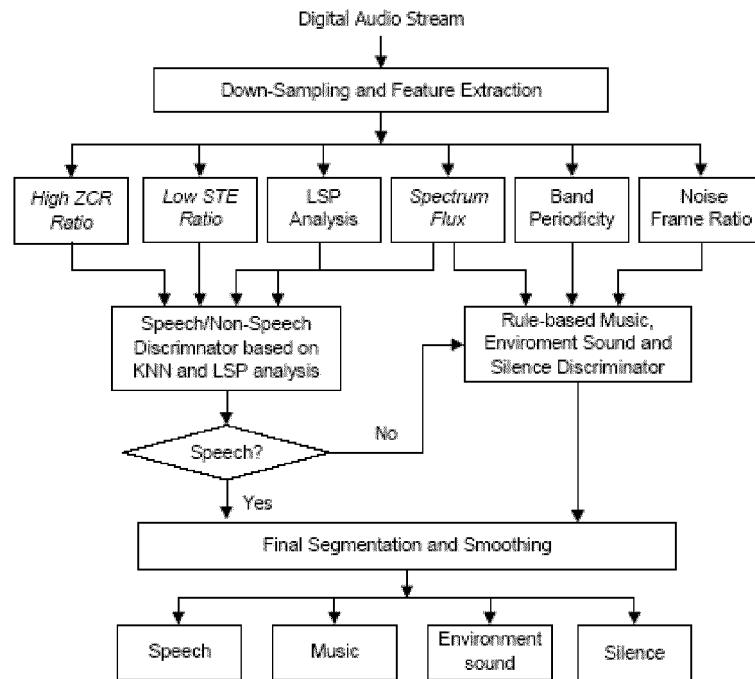


Figura 6. Diagrama de bloques de un sistema de clasificación y segmentación de audio

Después de clasificar las señales de audio en estos cuatro dominios, se siguen aplicando los métodos de reconocimiento de patrones para continuar con una jerarquización.

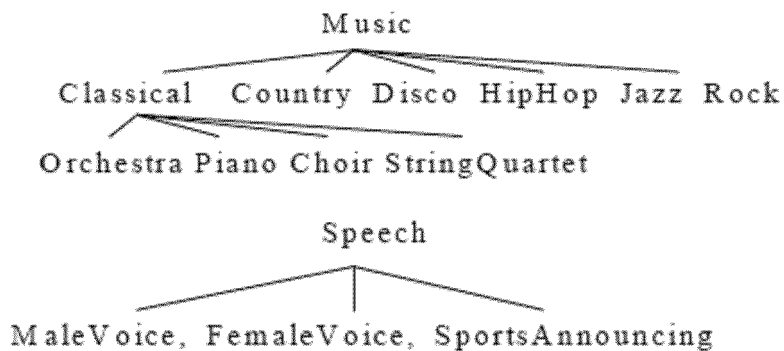


Figura 7. Jerarquía de clasificación por género

4. APLICACIONES DE LA CLASIFICACION AUTOMÁTICA DE AUDIO

Las aplicaciones prácticas pueden darse en tres dominios: voz, música, y ambiente de sonido. Voz y música, son las segmentaciones más importantes.

4.1 **Aplicaciones en voz**, la aplicación más importante envuelve el reconocimiento automático de voz (ASR). Aplicaciones prácticas en las que ASR es central, incluyen dictación, transcripción de voz grabada, aplicaciones en telefonía, aplicaciones de control e interfaces de voz.



- 4.2 **Aplicaciones en Música.** Transcripción de música, *Query by humming*, identificación de género musical, *audio thumbnailing* por identificación de segmentos recurrentes y Análisis estructural de música. La transcripción de música se refiere a un proceso completo de escritura bajo notación musical (por ejemplo: transcripción de acordes y melodía). *Query by humming* (Recuperación de un sonido específico basándose en la melodía), en esta aplicación el análisis del Pitch es la característica central de extracción. El análisis estructural de música puede ser muy útil en aplicaciones de búsqueda y recuperación. La Identificación automática de regiones musicales como coro, verso y secciones de diferentes instrumentos, puede ser de gran ayuda en la búsqueda de archivos de audio y *audio thumbnailing*.

5. GRUPOS DE INVESTIGACIÓN QUE TRABAJAN EN LA CLASIFICACIÓN AUTOMÁTICA DE AUDIO

- China Speech group: este grupo ha desarrollado aplicaciones interesantes basadas en reconocimiento de patrones y en la voz y clasificación de audio, algunas aplicaciones recientes interesantes son:

A Trainable Text-to-Speech Synthesis
Finding Music: Query-By-Humming and Music Steering
A Voice User Interface with Intelligent Correction
Discriminative Training of HMM Models
Enhancing human-human communication: Speech Indexing

<http://research.microsoft.com/Speech/>

- Departamento de Teoría de la Señal y Comunicaciones de la Universidad de Alcalá

<http://www2.uah.es/teose/inicio.htm>

- MIR Group: Vienna University of Technology Institute of Software Technology and Interactive Systems Information & Software Engineering Group. Proyectos en los que han participado:

DELOS
MUSCLE
IMPACT

<http://www.ifs.tuwien.ac.at/mir/index.html>

- Music, Mind and Machine Group

<http://sound.media.mit.edu/index.html>



6. PRODUCTOS COMERCIALES QUE EXISTEN EN EL MEDIO

Los productos más comerciales para clasificar música automáticamente son programas de software que permiten varios tipos de opciones según sea la necesidad del usuario, ya que no todos buscan lo mismo y en lo mismo, es decir, hay algunos quienes precisan de solo clasificación por artista y por nombre de la canción y otros necesitan de opciones mas avanzadas como categorías, géneros, duración de canciones, entre otros.

Mas allá hay que tener en cuenta el software para ciertas plataformas, es por ello que en lo referente al sistema operativo de Linux se tiene como opción el software que ofrece iPod mediante el iTunes (también usado en sistemas Windows).

- **iTunes:** se ha constituido como un reproductor de muy buena acogida entre quienes gustan de la clasificación ordenada de música, además que recopila toda la música existente en el disco duro y lo hace bastante atractivo ya que no todos los usuarios saben que tanta música existe en sus discos, problema al cual denominamos “caos musical”. En otras palabras pone en orden toda la colección musical y permite organizarla en un estilo de biblioteca musical clasificándola por género, artista, tipo de música, entre otras opciones más.
- **AmaroK:** para usuarios de Linux, tiene características muy similares a las del iTunes, pues de igual manera permite organizar los archivos por tipo de música, artistas, duración, etc. AmaroK permite también editar datos de las canciones y si se desconoce el nombre de la canción busca la información en la Web. Otra característica de AmaroK es que permite leer las letras de las canciones que están sonando en el momento, además de poder adquirir las carátulas y los álbumes enteros por la Web.
- **SongBird:** Mas conocido también como el pajarito de FireFox. Para sistemas Linux y Windows. Promete ser una excelente alternativa para clasificación de música desde Internet. Este software no permite solo reproducir las canciones del computador si no que también automáticamente genera otras listas de otras melodías halladas en la Web, para ello se apoya en Pitchfork Media donde se consiguen listas de las canciones en MP3. Desde el mismo programa puedes acceder a bajar música, a comprarla, tenes acceso a la Radio y a podcast. El pajarito de FireFox es bastante nuevo en el medio, su apariencia puede cambiar un poco, tiene muchas ventajas y muchas alternativas por ensayar, es por ello que esta en prueba y hay que experimentarlo para no tener problemas y estar familiarizado con el.
- **MusicIP Mixer:** Disponible para todas las plataformas, es un programa que desde el principio, cuando se dispone a organizar los archivos del computador, te ofrece la alternativa de organizarlos como se desee y como se te haga mas cómodo para escoger una canción a tu preferencia. En pocas palabras es un software que le ayuda al usuario a escuchar lo que se quiere oír.



Existen entonces gran cantidad de productos los cuales permiten la clasificación de la música de manera automática, unos mas nuevos, otros mas viejos, hay muchísimos de donde escoger de acuerdo a los gustos y preferencias. Otros programas desde software son:

- **1st MP3 Tag Editor**: Para hacer un buen uso de esta herramienta es necesario una conexión a Internet para obtener toda la información de tus canciones automáticamente desde la Web, usando FreeDB puedes rellenar información de la canción tal como el nombre del tema, el título del álbum, duración, etc. Te ofrece la opción además de buscar temas de música e inmediatamente obtener información de todo un álbum. También podrás insertar, cambiar o borrar tags a tus archivos.
- **Activa MP3 Tray**: Herramienta de fácil uso que permite la administración de todos los archivos Mp3 existentes en el computador para organizarlos automáticamente y clasificarlos según generes con posibilidad de submenús y grupos de canciones.
- **CDAID (CD Audio Information Downloader)**: automáticamente desde Internet permite ingresar a una base de datos y obtener nombres, artistas, títulos del disco, duraciones, etc.
- **CD-Tree**: permite organizar los Cds y las canciones en diversas categorías, se pueden realizar búsquedas rápidas y eficaces, se pueden crear copias de seguridad y se puede clasificar por tamaño, según estadísticas, entre otros.
- **MP3 Filename Formatter**: es una buena solución para cuando se hacen grabaciones utilizando un extractor o convertidor que la canción pierde su nombre y queda indicada solo con un "Track #", así pues podemos imitar las grabaciones con propiedad y tener los nombres correctos de la canción y el artista.
- **MP3 Organizer**: Es un organizador de directorios para hacer mas fácil la búsqueda de archivos MP3, normalmente por las letras iniciales de cada directorio creado.
- **Music Label**: permite almacenar el nombre de la canción, el autor, tipo de disco, tipo de música y otros mas avanzados tales como empresa discográfica, precio y comentarios. Permite la creación de paginas Web, creando listas, con audio, video y datos. Opciones avanzadas: recordatorios de compra de discos, estadísticas de compra, discos sin usar en un periodo y hasta la gestión de seguros para nuestros discos. Como se puede notar en esta descripción es un muy buen software que va más allá de una simple clasificación.
- **R-porter**: bastante útil para Cds de audio. Permite búsquedas avanzadas y precisas desde una base de datos de almacenamiento de registros que tiene preconfigurados una serie de filtros para realizar búsquedas por autor.



tamaño, tipo, género, categoría, entre otros. Tiene un menú de ayuda bastante útil para el manejo del software.

- **Tag Rename:** hay otros programas como este, que presta otros servicios tales como corrección de errores de los archivos de música, además de organizar, archivar y otorgar información de los archivos desde la Web para una mejor clasificación de los álbumes

Estos son unos de los pocos software que permiten una clasificación de música automáticamente y algunos de ellos con otros servicios más, para más información acerca de ellos, o para las descargas gratuitas se pueden acudir a la siguiente página Web: http://descargas.orange.es/descargas/Hogar_Familia_y_Ocio/Musica/Organizadores_de_Colecciones_de_Musica/

REFERENCIAS:

- Clasificación automática de voz/música utilizando discriminantes lineales de Fisher. Enrique Alexandre y Manuel Rosa. Universidad de Alcalá.
- Música, Complejidad, información. Damián Horacio Zanette. Abril de 2008
- A Robust Audio Classification and Segmentation Method. Lie Lu, Hao Jiang and HongJiang Zhang. Microsoft Research.
- http://mcyti.izt.uam.mx/Tesis_egresados/TesisGabrielPedroza.pdf
- <http://www.mouse.cl/2006/productos/06/27/02.asp>
- http://descargas.orange.es/descargas/Hogar_Familia_y_Ocio/Musica/Organizadores_de_Colecciones_de_Musica/
- Content-Based Audio Classification and Retrieval Using SVM Learning, Stan Z. Li, Microsoft Research China, 5/F Beijing Sigma Center.
- Methods of Automatic Audio Content Classification, Thesis, Helsinki University of Technology.
- Automatic Musical Genre Classification Of Audio Signals, T. George, E. George, C. Perry, ismir
- ..

