

# El audio por ordenador

**Un breve estudio y una pequeña prueba de concepto aplicado sobre todo a la música procesada por computador.**

## Introducción

Este trabajo en un principio, iba a ser un estudio (no demasiado profundo) de los procedimientos de extracción de características de audio procesado por ordenador, o lo que es lo mismo, Music Information Retrieval (MIR) para luego, poder clasificar canciones y partes de ellas según diferentes criterios y, finalmente, dar a luz una pequeña prueba de concepto que recomendara canciones, fragmentos y otros según criterios y utilidad para el usuario y, dependiendo de cómo y dónde encajarían esas canciones o fragmentos en un nuevo proyecto del artista.

No obstante, al darme cuenta de la falta de parte del conocimiento de base necesario para afrontar la tarea, opté por otra idea.

Finalmente, ha sido el estudio de lo básico del audio procesado por ordenador, las características propias del audio contenido en cada fichero que hacen posible el entendimiento, clasificación y recomendación de canciones, piezas, tramas de audio, etc. además de los primeros procesos necesarios para diseñar procesos que extraigan estas características y el diseño de los primeros modelos que puedan reconocer los parámetros y estos ejemplos de audio.

## Las bases del audio por ordenador

### Fichero de audio WAV

El sonido en directo se diferencia bastante del sonido grabado y almacenado, concretamente en un fichero WAV.

Este tipo de formato, a diferencia de otros como MP3, no comprime las muestras de audio y por eso es más fácil que el audio mantenga mayor calidad. Además, es más fácil de estudiar.

#### Muestreo (Sampling)

El muestreo consiste en intentar replicar una señal continua (como el sonido en directo) transformándola en un conjunto de valores discretos, que es lo que un ordenador puede comprender.

#### Muestra (Sample)

Cada muestra recoge la amplitud de la onda en un intervalo temporal determinado.

Cada muestra posee un parámetro importante llamado profundidad de bits (bit depth) que mide el tamaño de los datos que puede almacenar cada muestra.

Esta variable es la que determina la calidad del audio grabado. A mayor profundidad de bits, mayor calidad y mayor espacio ocupa la muestra.

Gracias a esta técnica, en el campo de la música, se puede reproducir bastante fielmente en diferido, canciones o piezas que han sido grabadas en directo y guardadas en ficheros.

## Características instantáneas del audio (Features)

Las llamadas características instantáneas de un fichero de audio componen un conjunto de elementos que describen cómo es la señal física que produce ese sonido representado en el fichero.

Estos parámetros, posteriormente compondrás las características básicas musicales del sonido que reproduce ese fichero tales como, la altura de una nota

### Características basadas en frecuencia

#### 1. Spectral Centroid - Centro de gravedad espectral

Indica el punto del espectro de frecuencias que tiene más peso por unidad de tiempo. Se podría traducir en la media de las frecuencias que suenan por unidad de tiempo.

#### 2. Spectral Rolloff

Representa la frecuencia a la que las altas frecuencias se hacen cero. ???

#### 3. Spectral Bandwidth - Ancho de banda espectral

El ancho de banda medio de la señal de audio.

#### 4. Mel-Frequency Cepstral Coefficients(MFCCs)

Este indicador es un subconjunto de entre 10 y 20 parámetros que describe la forma del espectro de frecuencias. Se usa para modelar las características de la voz humana.

#### 5. Chroma Feature (Pitch Chroma) - Tono

Agrupar las frecuencias en las 12 notas musicales de la escala.

#### 6. Spectral Crest - Cresta espectral

Cataloga la intensidad de los picos de la onda.

#### 7. Spectral Decrease - Disminución espectral

#### 8. Spectral Kurtosis - Apuntamiento espectral

Mide el apuntamiento de la onda de frecuencia. Ésto describe la impulsibilidad de la señal.

#### 9. Spectral Skewness - Asimetría espectral

Estudia la forma de la onda de frecuencia en relación con la asimetría.

#### 10. Spectral Slope - Pendiente espectral

Es la velocidad con la que el espectro de la señal se va apagando en las altas frecuencias.

### **11. Spectral Spread - Propagación espectral**

Desviación promedio de la señal entorno al centro de gravedad (centroid). Las señales multitónicas tienen una propagación alta, al contrario que las monotonales.

### **12. Tonal Power Ratio - Ratio de potencia tonal**

## **Características basadas en potencia**

### **1. Spectral Flatness - Planitud espectral**

Es el ratio entre la media aritmética y geométrica del espectro de potencia. Cuantas más bandas existan en la señal mayor será el flatness (no mayor que 1 ni menor que 0) y viceversa.

El reducido número de bandas se asocia a una señal tonal pura y la presencia de numerosas bandas se asocia a ruido (el ruido es una señal multitonal gracias a que distribuye la potencia entre muchas bandas).

### **2. Spectral Flux - Flujo espectral**

Es una medida que indica cómo de rápido varía el flujo de potencia de la señal.

## **Características basadas en tiempo**

### **1. Autocorrelation Coefficients - Coeficientes de autocorrelación**

La autocorrelación mide cuánto se asemeja la muestra (sample) actual con una tomada con un desfase determinado de tiempo. Si son iguales, este coeficiente es 0.

Según avanza la señal en el tiempo, la correlación decrecerá hasta cierto punto, entonces comenzará a ascender hasta alcanzar un máximo local. Ésto se puede utilizar para reconocer el tempo de la canción.

### **2. Autocorrelation maximum - Valores máximo de autocorrelación**

Mide los máximos de la función creada por la autocorrelación.

### **3. Peak Envelope - Ámbito del pico**

### **4. Root Mean Square**

### **5. Standard Deviation**

### **6. Zero Crossing Rate - Ratio de cruce por punto 0**

Es una medida de cuánto la amplitud de la señal de audio cruza el punto cero por unidad de tiempo.

## Datasets

### FMA (Dataset For Music Analysis)

Este dataset trae 8000 pistas de audio (en el caso del fma\_small) que pertenecen a numerosos géneros.

Está pensado para realizar la extracción de características de audio (features) y clasificar las canciones en función del género.

Las pistas se deben almacenar en los subdirectorios nombrados con números del interior de la ruta /datasets/fma/data.

Además, aglutina características básicas extraídas de los ficheros de audio, que son almacenadas en ficheros csv en el interior del directorio /datasets/fma/data/fma\_metadata.

Se pueden descargar fácilmente de la página de Github del proyecto:  
<https://github.com/mdeff/fma>.

Trae ciertos cuadernillos de Jupyter y scripts de Python para analizar y preprocesar los datos.

analysis.ipynb - Explora los datos y aporta una primera vista de los mismos.  
baselines.ipynb - Es una comparación de los clasificadores de SciKit Learn y Keras aplicados al problema de clasificación de géneros.  
creation.ipynb - Actualiza los ficheros csv en caso de que se actualizen los ficheros de audio.  
usage.ipynb - Realiza un pequeño análisis estadístico de las características de audio y de los metadatos presentes en los ficheros csv y da una guía para el desarrollo de modelos.

### GiantSteps Key

Este conjunto de datos se compone de 604 ficheros de audio que deben ser guardados en el directorio /datasets/giantsteps-key-dataset/audio.

Se pueden descargar en formato mp3 utilizando el script /datasets/giantsteps-key-dataset/audio\_dl.sh.

Además, es necesario convertirlos a wav, para lo que se provee de otro script: /datasets/giantsteps-key-dataset/convert\_audio.sh.

También, en la ruta /datasets/giantsteps-key-dataset/annotations se encuentran la asignación de género y tonalidad de cada canción para el entrenamiento de los modelos.

Fue creado para realizar sobre él la extracción de parámetros de audio y clasificar las canciones por géneros y por tonalidad.

El dataset se encuentra en el siguiente repositorio:  
<https://github.com/GiantSteps/giantsteps-key-dataset>

## Proyecto Audio Content Analysis

El mecanismo de extracción de las características ha sido, en parte obtenido y en parte inspirado en un proyecto que proporciona una introducción a los métodos de Music Information Retrieval (MIR) y su implementación.

El MIR no es otra cosa que la extracción de las características del audio procesado por ordenador que antes he mencionado además de algunas otras.

En algunos aspectos es un poco avanzado porque presupone cierto conocimiento de procesamiento de señales por ordenador y de ondas sonoras.

Los scripts que componen este mecanismo MIR se encuentran en los siguientes directorios de la ruta raíz del trabajo:

- /features
- /features2
- /helper\_functions
- /pitch\_track
- /rhythm

Además de los scripts:

- basic\_plot.py
- feature\_computation.py
- fundamental\_frequency\_computation.py
- key\_detection.py.

Realizan respectivamente:

Dibujado de gráficos básicos para echar un primer vistazo a un fichero de audio.

Ejecuta las funciones de extracción de features del directorio /features

Ejecuta el análisis de las características relacionadas con la altura de musical del audio.

Determina la tonalidad del fichero de audio que se está analizando.

## Conclusión

En el repositorio se encuentra todo lo necesario (parte cedida y parte de creación propia) para probar los funcionamientos de los distintos mecanismos, trastear con ellos pudiéndolos editar y analizarlos medianamente a fondo.

Para ello, en su construcción se ha evitado en la medida de lo posible el uso de librerías de funciones predefinidas optando por desarrollar funciones propias que describen mejor los procesos.

Me he dado cuenta de que por medio de este estudio he podido establecer cierta relación entre lo que son las características musicales de una canción u obra y los parámetros físicos que es lo que, al fin y al cabo, maneja el ordenador (o una simulación de ellos).

Ambos ámbitos casi siempre suelen estar separados y no se hace por complementar el conocimiento de un área con el de la otra, cosa que creo es muy beneficiosa para aprender sobre ellas.

Así que, aunque no he cumplido alguno de mis objetivos principales que declaré al principio, he sacado conocimiento sobre un área poco común y que me fascina.