

# Reconocimiento de música

Jose Javier Calvo Moratilla

Reconocimiento automático del habla  
Curso 2021/2022

## 1. Introducción

Los reconocedores de música llevan con nosotros muchos años en nuestro bolsillo, instalados en un dispositivo móvil, una herramienta mágica que nos ha ayudado a obtener el nombre de la canción que escuchamos en directo en una discoteca o en la cafetería donde almorcamos todos los días.

Tener una herramienta que te permite identificar la canción que suena, demuestra el poder real de las nuevas tecnologías dentro del campo de la inteligencia artificial, concretamente en el reconocimiento de sonido.

El reconocimiento de canciones no se centra en identificar una canción con las mismas características, hay aplicaciones que profundizan y utilizan el ingenio para poder detectar canciones mediante otras fuentes de sonido, como los silbidos o el tarareo.

La competencia entre las aplicaciones es bastante dura, pero hay herramientas que gozan de mayor prestigio por la aparición en películas y series que han viralizado su uso, gracias a las redes sociales.

Cómo se observará en los siguientes puntos las aplicaciones de reconocimiento de canciones no se limitan a la detección exclusivamente, tienen un alto componente social para poder conectar con el mayor número de redes sociales posible, para compartir toda la experiencia de uso de las aplicaciones.

En primer lugar en el presente trabajo se hace una aproximación con el marco teórico del sonido, la adquisición, discretización y tratamiento para poder ser utilizada en la tarea de reconocimiento en las dos aplicaciones más importantes del mercado, Shazam y Soundhound.

En último lugar se dará a conocer una implementación demo del funcionamiento de un reconocedor de canciones en lenguaje python para realizar una aproximación del funcionamiento de dichas aplicaciones.

## 2. Shazam

Es una aplicación creada en el año 2002 de la empresa Shazam Entertainment Limited, fundada por Chris Barton, Philip Inghelbrecht, Avery Wang, y Dhiraj Mukherjee en el año 1999. La empresa fue adquirida en 2018 por Apple Inc por 400 millones de dólares.



Figura 1: Logotipo Shazam

### 2.1. Funcionalidades

La aplicación te permite encontrar el nombre de una canción en segundos, permite escuchar y añadirlas a listas de reproducción de Apple Music, puedes obtener la letra de la canción, ver los videos de las canciones que han identificado en Apple Music o Youtube o recibir recomendaciones relacionadas.

La aplicación es multiplataforma y está ampliamente conectada con todas las redes sociales como facebook, WhatsApp, Instagram, twitter, etc.



Figura 2: Interfaz Shazam

Te permite añadir la app como widget en tu dispositivo móvil para poder utilizar la aplicación con mayor rapidez y es completamente gratuita, no te muestra ningún tipo de publicidad molesta.

La fortaleza de negocio de shazam es que si un usuario compra una canción habiendo detectado una canción desde la aplicación, se lleva una comisión de la venta.

Uno de los puntos negativos es que no puede detectar una canción por el silbido o tarareo como otras aplicaciones como SoundHound, que sí que tiene dicha funcionalidad, todo gracias al *Query By Humming (QbH)*.

Actualmente la aplicación está utilizando funcionalidades de reconocimiento de imágenes de códigos QR para colaborar en campañas publicitarias mediante el uso de su app y por último están poniendo en práctica la realidad aumentada.

## 2.2. Marco teórico

El sonido es una vibración mecánica de presión que se propaga a través de un medio como el aire. En el cuerpo humano el tímpano es el encargado de transformar dichas vibraciones en señales eléctricas que se transmiten al cerebro a través del nervio auditivo.

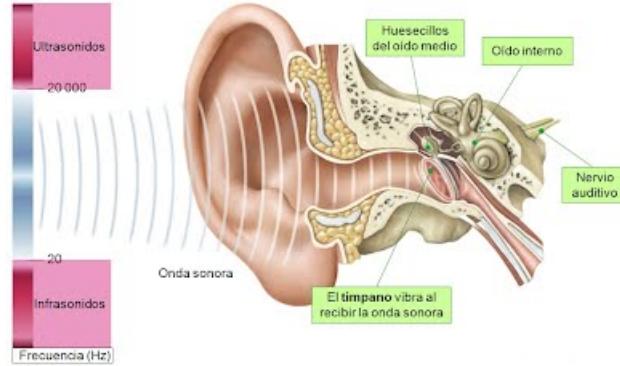


Figura 3: El oído humano, sjcalasannciencias2

El ser humano observa el comportamiento de la naturaleza para intentar emularlo, por ello los dispositivos de grabación son capaces de transformar el sonido a señales análogicas.

Para poder utilizar la señal de manera digital y ser almacenada, ésta tiene que preprocesar, tiene que discretizar.

La discretización consiste en representar la señal analógica con unos valores digitales que representan la amplitud de la señal, añadiendo una pequeña cantidad de error.

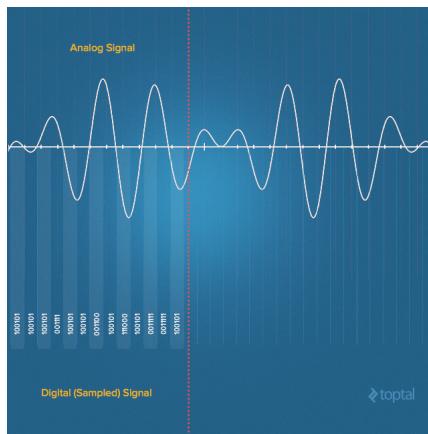


Figura 4: Proceso de discretización, toptal.com

El teorema de Nyquist-Shannon nos indica la tasa de muestreo necesaria para poder capturar el máximo de frecuencias posibles, perfectamente oíbles por el ser humano, para poder ser reconstruida la señal de nuevo, que en el caso del oído humano tiene que ser el rango de los humanos multiplicado por dos. Las frecuencias del oído humano van desde los 20 Hz a los 20000 Hz, por ello la tasa de muestreo utilizada en los medios digitales es de 44100 Hz.

Gracias a la investigación de Jean-Baptiste Joseph Fourier, la demostración de la serie de Fourier permite representar una señal periódica sólo con el valor de las frecuencias, amplitud y fase de cada sinusode, convirtiendo la señal en una suma de senos y cosenos que permiten realizar una aproximación de la señal real.

Los parámetros a y b de la siguiente ecuación corresponde con los coeficientes de Fourier:

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos \frac{2n\pi}{T} + b_n \sin \frac{2n\pi}{T} t$$

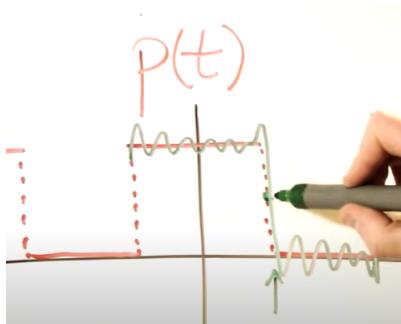


Figura 5: Aproximación serie de fourier de una señal cuadrada, El traductor de ingeniería

Gracias a la serie de Fourier se obtiene la representación discreta o la frecuencia de dominio como huella digital característica para poder identificar un sonido concreto.

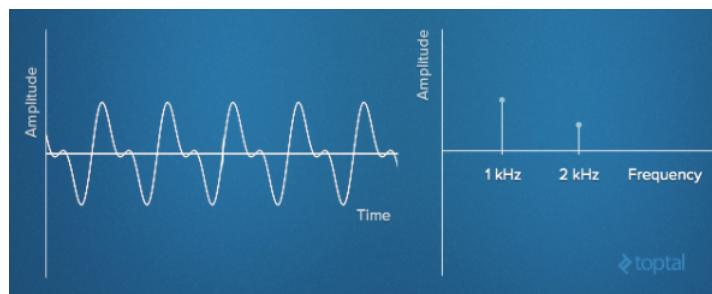


Figura 6: Transformación de la señal al dominio de frecuencias, toptal.com

Para el proceso en un primer se ha utilizado el algoritmo *Fast Fourier Transform o FFT* con un coste

$$O(n^2)$$

que con la variación del algoritmo llamada *Cooley-Tukey o DFT* se consigue un coste computacional de

$$O(n * \log n)$$

En resumen, se obtienen todas las frecuencias y las magnitudes correspondientes de una señal en un intervalo temporal.

El objetivo para el diseño de la huella digital es encontrar las frecuencias más importantes que identifican a una canción, para ello se eligen diferentes intervalos de frecuencia para poder identificar los rasgos más importantes de una canción, frecuencias de los instrumentos, voces de las canciones, etc.

Mediante el uso de dichos intervalos se obtiene de cada intervalo la frecuencia máxima observada, generando una huella digital de la misma. Dichas características se almacenan en una base de datos con el nombre de "Hashtag", junto a la duración del sonido y la etiqueta de la canción correspondiente.

Hash Tag	Time in Seconds	Song
30 51 99 121 195	53.52	Song A by artist A
33 56 92 151 185	12.32	Song B by artist B
39 26 89 141 251	15.34	Song C by artist C

Figura 7: Características que identifican a un sonido, [toptal.com](http://toptal.com)

La información que hace referencia a las frecuencias observadas en los intervalos de importancia prefijados en la construcción de la huella digital puede no ser del todo relevante en la diferenciación de canciones, ya que dos canciones pueden tener la misma base, por ello se utiliza un algoritmo de distribución en el tiempo.

Es importante ya que una canción larga puede contener en el tiempo las frecuencias obtenidas con la grabación realizada, por ello se comprueba las frecuencias en diferentes intervalos de la duración completa de la canción para ver si son coincidentes con el extracto de sonido grabado.

Una vez ejecutados los algoritmos se obtiene una solución con las canciones más probables, siendo la canción más probable la elegida como resultado final.



Figura 8: Búsqueda Shazam

### 3. SoundHound

Soundhound es una aplicación de reconocimiento de música creada en el año 2005 por su fundador Keyvan Mohajer. Su campo de acción es el reconocimiento del habla, el entendimiento de lenguaje natural, reconocimiento de música y tecnologías de búsqueda.



Figura 9: Logotipo SoundHound

### 3.1. Funcionalidades

La aplicación permite utilizar los asistentes de voz para ejecutar multitud de comandos para interactuar con la aplicación, como por ejemplo pedir la letra de una canción concreta.

Una de las principales desventajas de Soundhound es la publicidad que puedes eliminar si compras una versión pro de la misma.

El programa permite identificar tanto la música como el silbido o tarareo de la misma, a diferencia de Shazam que solo es capaz de identificar las canciones. Otra de las características que la diferencia es que muestra todas las posibles canciones y no solo una.

La aplicación muestra imágenes del álbum al que pertenece, escuchar la canción con limitaciones, letra de la canción e información relacionada como la bibliografía del cantante o la gira de conciertos.

El tiempo de ejecución es más largo que la búsqueda que realiza Shazam y otra de las funcionalidades destacables es que permite compartir la información rápidamente por redes sociales y comprar las canciones en Amazon.



Figura 10: Interfaz SoundHound

Otra de las herramientas de SoundHound es la recomendación de canciones similares a las canciones identificadas en el historial. Dependiendo de la versión gratuita o no te permite tener más funcionalidades disponibles.

### 3.2. Marco teórico

Soundhound a diferencia de Shazam utiliza el paradigma *Query By Humming (QbH)* (6) para la tarea *retrieval* de canciones, que si nos remitimos la traducción significa encontrar una canción en una base de datos tarareando.

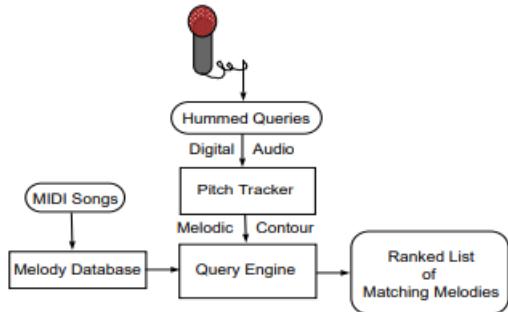


Figura 11: Arquitectura QbH, Cornell University

Para discriminar entre canciones se observa el contorno melódico, definido como la secuencia de diferencias relativas de tono entre notas sucesivas. Según Stephen Handel (7) el contorno melódico es el elemento más importantes que los seres humanos utilizamos para discriminar entre dos canciones. Al observar notas consecutivas se puede ver la evolución de éstas en el tiempo y ser utilizada dicha evolución para recuperar las canciones que hacen *matching* con la búsqueda.

Las notas se codifican con las siglas (U, D, S). Si la nota que sigue es igual a la anterior la transición se codifica con la letra S, si es más alta que la nota anterior con la letra U y si es más baja con la letra D, por ejemplo el inicio de la 5<sup>a</sup> Sinfonía de Beethoven se codifica como (- S S D U S D).

El siguiente paso es identificar el tono para poder hacer la codificación por ello se rastrean las ráfagas de aire salientes de la glotis. Para poder identificar el tono se pueden utilizar tres aproximaciones diferentes, la Autocorrelación, la Máxima Verosimilitud o el Análisis Ceps-trum.

Una vez identificados los tonos y realizada la codificación *UDS* se realiza la búsqueda en la base de datos. Se puede realizar una búsqueda difusa para meter posible ruido a la señal ya que no todas las personas tararean de la misma manera.

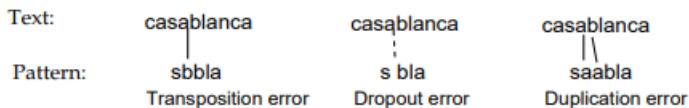


Figura 12: Posibles errores en QbH, Cornell University

Se utiliza el algoritmo propuesto por Baesa-Yates y Perleberg (8) en el que se trata de detectar cadenas con K desajustes, obteniendo a la salida las canciones que a pesar de dichos desajustes hacen más *matching* con alguna canción de la base de datos.

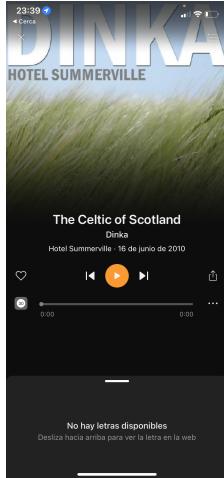


Figura 13: Búsqueda SoundHound

## 4. Demo Shazam en python

Se ha utilizado la herramienta shazam-demo de Peacecwz (4) para probar el funcionamiento de shazam en python. El código permite en primer lugar codificar una canción en la base de datos, para poder ser utilizada como referencia cuando la aplicación escucha una canción de fondo.

Se recomienda utilizar el entorno de anaconda en una distribución Ubuntu para poder ejecutar el código.

### 4.1. Librerías

Para poder utilizar la herramienta se precisa de las siguientes librerías en un entorno con python 2.7:

Librerías:
pylint
numpy
termcolor
pyaudio
wave
pydub
matplotlib
scipy

### 4.2. Resetear Base de datos

Mediante la ejecución del fichero reset.py la base de datos que almacena los *hashtags* que identifican a cada canción almacenada disponibles para ser reconocibles por el sistema.

El código utiliza la librería SQLiteDatabase para conectarse a la base de datos, creando un fichero dónde se almacena la base de datos en la carpeta `./db`.

### 4.3. Almacenar archivos mp3

El usuario debe de almacenar las canciones que se quieren detectar en la carpeta `./mp3` y una vez almacenadas en la carpeta se ejecuta el el fichero analyze.py para detectar las canciones de la carpeta y crear sus respectivas entradas en la base de datos para ser reconocibles.

#### 4.4. Detectar canciones

Una vez almacenadas todas las canciones se ejecuta el fichero listen.py con el parámetro -s 5 para que el micrófono del ordenador detecte la canción, si hay una coincidencia el fichero nos devuelve la información que hace referencia a la canción almacenada en la base de datos.

### 5. Conclusiones

En general las dos aplicaciones logran su cometido pero con diferente tecnología y estrategias. A parte muestran una conexión con redes sociales y la venta de canciones on-line.

Observando a Shazam la tecnología es más rápida que la de Soundhound, es completamente gratuita y se centra solo en mostrarte la canción que necesitas, en el caso se Soundhound te muestra la lista de las canciones más probables, cubriendo la posibilidad de no acertar con la primera canción.

Shazam es completamente gratuita pero se está aliando con marcas para poner en práctica campañas mediante el uso de nuevas tecnologías y sacar así beneficio económico, otra de las fuentes es la recomendación de canciones a tiendas de música on-line.

Soundhound muestra una política diferente, diferencia dos servicios, uno gratuito y otro de pago dónde se tienen mayores funcionalidades. A parte de la aplicación disponen de otros servicios que van orientados a ámbito profesional.

Después de probar personalmente las aplicaciones, las dos pueden detectar canciones, pero no he conseguido que Soundhound encuentre una canción con un tarareo, demostrando que la tecnología no es del todo efectiva.

Son las aplicaciones de reconocimiento de canciones más utilizadas a día de hoy por los usuarios en la red, pero poco a poco las empresas de smartphones están dotando a sus asistentes de voz la posibilidad de reconocer canciones, incluso con el tarareo.

## Referencias

- [1] El mejor identificador de canciones Consultado en <https://tecnologia-facil.com/como-hacer/identificador-canciones/>.
- [2] How does Shazam work? Consultado en <https://www.toptal.com/algorithms/shazam-reconocimiento-de-algoritmos-de-musica-huellas-dactilares-y-procesamiento>.
- [3] SoundHound Consultado en <https://www.todotech.com/android/apps/n112/soundhound-android-app-review.html#:~:text=SoundHound%20es%20una%20aplicaci%C3%B3n%20que,con%20la%20que%20disfrutar%C3%A1s%20mucho..>
- [4] Shazam-demo Consultado en <https://github.com/peacecwz/shazam-demo>.
- [5] Avery Li-Chun Wang An Industrial-Strength Audio Search Algorithm. *Shazam Entertainment, Ltd.*, 2003.
- [6] Asif Ghias, Jonathan Logan, David Chamberlin, Brian C. Smith Query By Humming, Musical Information Retrieval in an Audio Database *ACM Multimedia, Cornell University*, 1995.
- [7] Stephen Handel An Introduction to the Perception of Auditory Events. *The MIT Press*, 1989.
- [8] Ricardo A. Baesa-Yates and Chris H. Perleberg. Fast and practical approximate string matching. *Combinatorial Pattern Matching, Third Annual Symposium*, pages 185-192, 1992.