Clasificador de géneros musicales a partir del procesamiento digital de señales

Salmerón Facundo, Walczak Tomás y Yackel Francisco
Facultad de Ingeniería y Ciencias Hídricas – fyackel@gmail.com

***Resumen*— En este artículo se presenta un enfoque alternativo para la clasificación de los géneros musicales. A raíz de esto, hemos investigado sobre los diferentes métodos que existen para lograrlo. Nuestro proyecto se caracterizó por convertir la señal de audio en un vector numérico y a partir de este obtener distintas características representativas de la señal. El enfoque propuesto tiene un mecanismo de zonificación para llevar a cabo la extracción de estas características localmente, que demostró ser bastante eficaz en relación a un análisis global. Con los resultados obtenidos en las diferentes pruebas realizadas a lo largo del estudio se puede observar que el modelo propuesto sirve para realizar dicha tarea.**

***Palabras clave*—plantilla, género, característica, señal, reconocimiento.**

I. INTRODUCCIÓN

Los géneros musicales son etiquetas creadas por el ser humano para clasificar el estilo de la música. Un género musical se caracteriza por las características comunes compartidas por las distintas piezas musicales. Estas características normalmente están relacionadas con la instrumentación, estructura rítmica y el contenido armónico de la música. El descriptor de géneros es quizás el más utilizado para organizar y gestionar grandes bases de datos de música digital. Anteriormente la clasificación se realizaba manualmente, es por esto que la clasificación automática puede ayudar a sustituir al humano en este proceso, además permite la estructuración y organización de grandes archivos de música y también proporciona una buena manera de comparar y evaluar características que intentan representar el contenido musical.

Los primeros avances con respecto a la clasificación automática se basaban en representar la pieza musical en texturas tímbricas y en relación a los golpes y características relacionadas al pitch, logrando resultados viables, aunque, con el paso del tiempo y la mejoras tecnológicas, se pudieron realizar estudios mas sofisticados, igualmente, esto sigue siendo un problema abierto donde el mismo experimento con seres humanos determinaron que no fueron capaces de clasificar correctamente en gran porcentaje cada género musical, dando una idea de la limitación que llevara la realización de esta implementación. Todos los algoritmos utilizados en el proceso del estudio fueron realizados por los integrantes de este grupo. El objetivo de nuestro trabajo fue realizar un método a partir del cual podamos detectar y diferenciar cuatro géneros musicales a través de los atributos internos de una pista de música ya sea en el tiempo o en el dominio de la frecuencia. A partir de los diferentes temas abordados en el transcurso de la materia y otros en la investigación de esta temática, pudimos realizar una gran aproximación. Los géneros que elegimos para trabajar son: Cumbia, Clásica, Tango y Rock.

II. EXTRACCIÓN DE LA BASE DE DATOS

Para darle solución al problema planteado, creamos una base de datos a partir de un conjunto de canciones que representan a cada uno de los géneros, éste consiste en un grupo de 20 canciones por cada uno de los 4 géneros a estudiar, 80 canciones en total. Este grupo se ha dividido en dos partes, la primera formada por 10 canciones para entrenamiento y la segunda para pruebas y clasificación.
El tratamiento que se ha aplicado a la base de datos ha consistido en dividir a la canción en 3 partes iguales, una representa el inicio, la siguiente la parte del medio y la última el final de la misma. Por cada parte, tomamos 20 segundos y a ellas le aplicamos el análisis y obtención de características. Para evitar quedarnos con introducciones y finales sin sonido o que no aporten datos característicos del género al que pertenecen, a las partes del inicio y del final verificamos que efectivamente eran componentes valiosas de la canción. Se utilizó esta metodología considerando que es mejor un análisis local antes que uno global, ya que cada una de estas partes hace que la diferencia entre las distintas canciones se vea de mejor forma.

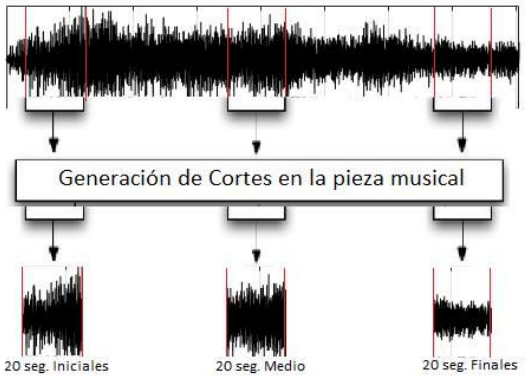

Fig. 1: Señal zonificada

La extensión de cada canción es .mp3, ya que fue la más fácil de conseguir, siendo conscientes de que dicha extensión realiza un procesamiento sobre las canciones quitando información de las mismas, igualmente nuestro trabajo no fue afectado por ello, arrojando buenos resultados. De todas formas elegimos el mejor formato mp3 que existe, que es de 320Kbps.
La elección de cada componente de la base de datos fue realizada teniendo en cuenta las diversas clasificaciones que se pueden encontrar dentro de un mismo género, por

ejemplo, dentro del rock se pueden encontrar varios estilos, como son el punk, el rock nacional, heavy metal, entre mucho otros. Entonces, a la hora de elegir cada componente, estas consideraciones fueron utilizadas. En el rock, se optó por analizar el estilo punk; en la cumbia, elegimos la cumbia santafesina, mientras que en los géneros tango y música clásica se eligieron los rubros más representativos del mismo, ya que los miembros del equipo no estamos muy interiorizados en dichos géneros.

III. ARMADO DE PLANTILLAS.

Para el armado de las distintas plantillas que se corresponderán a cada género elegido, es decir, 4 plantillas distintas, se eligieron 10 canciones por cada género, las cuales van a hacer las veces de base de datos que representarán a los mismos. El tratamiento que se le hizo a cada canción, es el de obtener distintas características, para luego buscar una media de cada una de ellas y así utilizarlas como referencia a la hora de comparar con la canción que se quiere analizar. Debemos recordar que cada canción es dividida en 3 partes, y a cada una de estas se le realiza la obtención de características. Esto se realiza para poder tener una mejor predicción del género buscado, como fue anteriormente dicho. Las características mencionadas fueron:

- Pico máximo de frecuencia
- Energía de corta duración
- Flujo espectral
- Tasa de cruces por cero
- Coeficientes ceptrales en escala de mel.

Pico máximo de frecuencia: La distribución de frecuencias de una señal se puede calcular a través de la transformada de Fourier. Procesos similares ocurren también en el sistema de percepción auditiva de los seres humanos y otros vertebrados, que indica que la composición espectral de una señal es de hecho el portador primario de información. El espectro de frecuencias es una de las características esenciales para reconocimiento de género de la música y se utiliza como la base para obtener muchas otras características.

En base a este espectro calculamos el pico máximo de frecuencia para tener una referencia de la máxima frecuencia presente en cada género.

Para ello calculamos la Transformada de Fourier de la canción, y luego con la función max de Matlab calculamos el valor máximo del módulo del espectro y su posición. El pico máximo de frecuencia es la posición de ese valor máximo encontrado.

$$[\sim, I] = \max(abs(fft(x)))$$

Energía de corta duración: La energía de corta duración de una señal de audio determina la variación de la amplitud en relación al tiempo. Aquí podemos detectar silencios, componentes sonoros y sordos, además de medir, si el sonido es fuerte o no, que es lo que determina la intensidad del sonido.

A cada elemento de la señal le aplicamos el cuadrado y sumamos esos resultados, por último dividimos por el largo de la señal.

$$E = \frac{1}{N}\sum x^2$$

Flujo espectral: Esta característica mide como varía el espectro de forma local. Constituye un buen atributo de percepción, importante en la caracterización del timbre del instrumento musical. En nuestro proyecto, utilizamos dos características estadísticas de esta propiedad, como son la media y la varianza. Para calcularlo dividimos la señal en dos ventanas de hanning de igual tamaño, y a cada porción se le calculó la Transformada de Fourier. Luego se restan las magnitudes de los dos espectros y se le calcula el cuadrado al resultado. A partir de este obtenemos la media y la varianza.

$$F = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2$$

Siendo N cada una de las magnitudes del espectro de las ventanas t.

Tasa de cruces por cero (ZCR): La tasa de cruces por cero puede ser utilizada para determinar el contenido frecuencial de una señal. Para ello determina cada vez que se produce un paso por cero, esto se da cuando dos valores consecutivos de la señal tienen signos distintos.

Si la señal estudiada presenta un ZCR alto, significa que tiene un contenido frecuencial alto, caso contrario, si el ZCR es chico, su contenido frecuencial será bajo. Con esta característica podemos encontrar la frecuencia fundamental de la señal de audio, de suma importancia, ya que podemos determinar el tono.

Para hallar la tasa de cruces por cero, primero normalizamos la señal y luego sumamos la cantidad de veces que ocurre que dos muestras consecutivas difieren de signo.

$$x = x / norm(x)$$

$$ZCR = \sum_{n=1}^{N} |diff(x > 0)|$$

Coeficientes ceptrales en escala de mel: Los coeficientes ceptrales en escala de mel son una característica perceptual que se basa en la transformada de tiempo corto de Fourier. El humano no logra percibir el sonido físico tal cual es, es decir, siguiendo una estructura lineal, sino que hasta aproximadamente 1000Hz puede lograrlo, y luego sigue una distribución logarítmica.

Primero, realizamos un ventaneo de los datos utilizando ventanas de hannning, luego obtuvimos la magnitud de la FFT, a continuación mediante un banco de filtros conseguimos 40 coeficientes cepstrales. Para que los mismos estén en escala de Mel, se le calcula el logaritmo en base diez, y para finalizar buscamos la transformada inversa de Fourier. Ahora, con estos coeficientes en escala de Mel, nos quedamos con los primeros 15, ya que serán los más representativos.

Como conclusión, nuestra plantilla quedará conformada por 20 características por los 3 tramos de cada canción. Estas características son evaluadas en las 10 canciones que utilizamos como base de datos, para luego obtener la media de cada una de las características en los distintos tramos, es decir, cada plantilla devolverá 60 características medias, 20 pertenecerán a las características medias de la primera parte, y así sucesivamente.

IV. PROCESO DE LAS PRUEBAS.

En la realización de las pruebas, lo que hicimos fue introducir una canción distinta a todas las utilizadas para conformar la base de datos, obtuvimos las características de la misma, realizando el mismo procesamiento de división en tres partes, etc.

Para calcular a que genero pertenece, comparamos a la canción ingresada con las plantillas correspondientes a cada género musical, midiendo la distancia euclidea entre ellas. Aquella plantilla que represente la menor distancia con respecto a la canción de prueba será la que se aproxime de mejor manera con las características de la canción, es decir, el género que representa dicha plantilla será el elegido para la canción de prueba.

Destacamos que de las 20 características que obtenemos, 15 de ellas le corresponden a los coeficientes ceptstrales en escala de mel.

En base a diferentes pruebas realizadas donde los coeficientes tenían igual peso que las otras 5 características, vimos que estos en muchas oportunidades determinaban unívocamente el género musical de la canción de prueba, haciendo que el poder de decisión de las demás características sea nulo. Es por ello, que decidimos darle la mitad del peso a dichos coeficientes, encontrándonos con mejores resultados de clasificación.

```
rock_cont =     clasico_cont =

    0.5000          4.0000
    1.5000          0.5000
    2.0000          2.5000


cumbia_cont =   tango_cont =

    0.5000          7.5000
    2.5000          8.0000
    1.0000          7.0000
```

Fig. 2: Resultados de una prueba de tango

V. RESULTADOS

Los datos obtenidos del entrenamiento se han volcado en una tabla para observar claramente los porcentajes de aciertos por cada género, utilizando 10 canciones de prueba por cada uno, como anteriormente fue aclarado.

Género	1er Parte	2da Parte	3ra Parte
Clásica	10	10	10
Cumbia	5	5	7
Rock	7	5	7
Tango	9	10	9

Tabla 1: Resultados de las 40 pruebas.

Como se puede observar, se lograron resultados muy confiables tanto en la música clásica como en el tango, no tan así en los géneros de rock y cumbia. Pensamos que esto puede deberse a la variedad de estilos que existe de cada uno, también antes mencionado.

Género	Aciertos % 1er Parte	Aciertos % 2da Parte	Aciertos % 3ra Parte	Total Aciertos
Clásica	100%	100%	100%	100%
Cumbia	50%	50%	70%	56.67%
Rock	70%	50%	70%	63.33%
Tango	90%	100%	90%	93.33%
Total				78.34%

Tabla 2: Porcentaje de acierto.

VI. CONCLUSIÓN

El método que diseñamos para detectar los géneros musicales tiene un 78.3% de efectividad.

Este estudio o trabajo podría extenderse evaluando que sucede al aumentar canciones en cada una de las bases de datos para los distintos géneros. También podríamos modificar la forma de comparar las canciones de pruebas con las plantillas, por ejemplo aplicando el concepto del producto punto, que mediría el grado de parecido. Además, podríamos evaluar la búsqueda de otra característica para mejorar aún más el detalle de cada género.

REFERENCIAS

[1] Music Genre Recognition Using Spectrograms – Costa, Oliveira, Koerich, Gouyon.
[2] Manipulation, analysis and retrieval systems for audio signals –G. Tzanetakis.
[3] Music Genre Recognition- K. Kosina
[4] Musical Genre Classification of Audio Signals - Tzanetakis