

Clasificador de géneros musicales a partir del procesamiento digital de señales

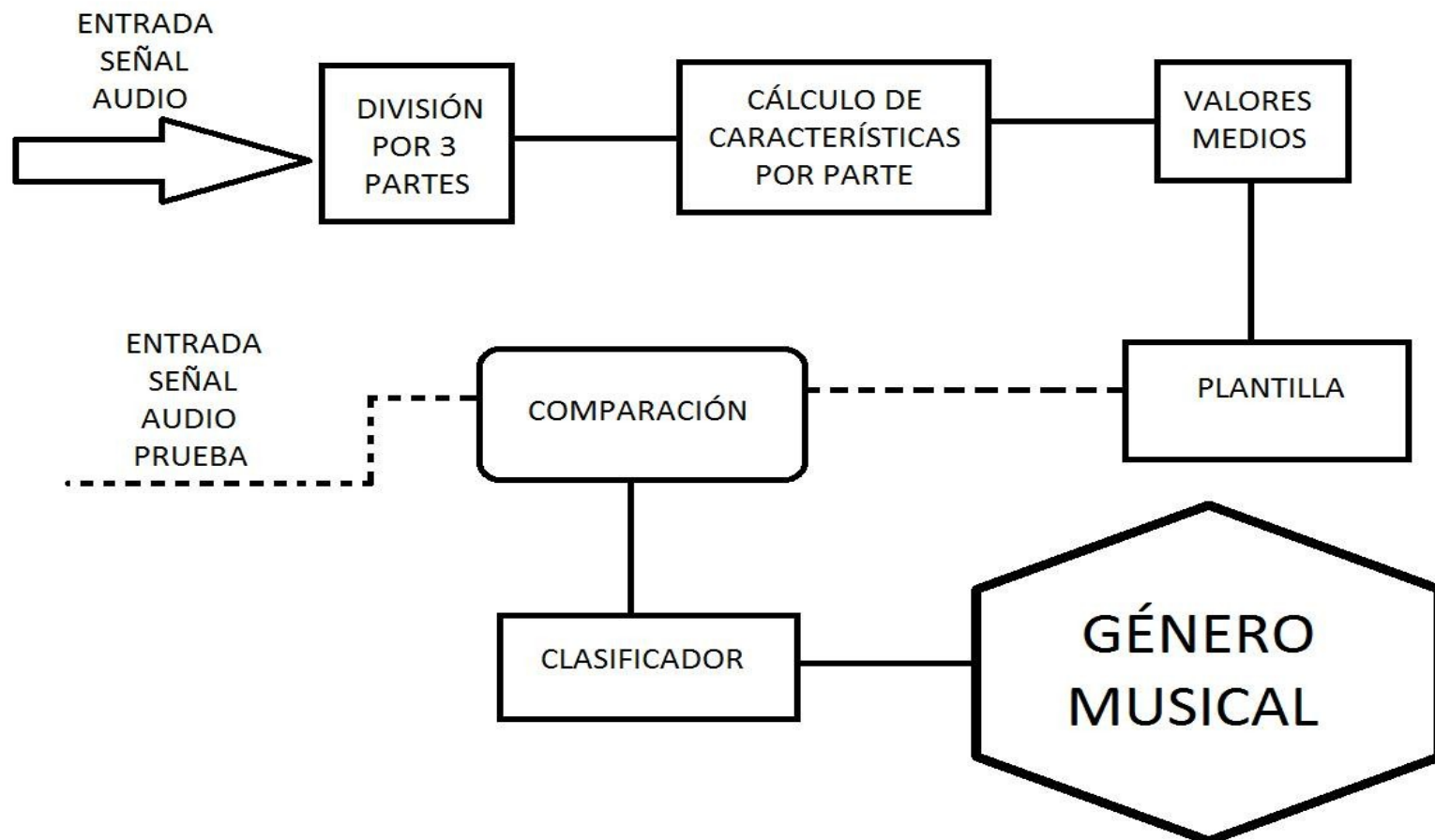
- *Salmerón Facundo*
- *Walczak Tomás*
- *Yackel Francisco*

Motivación y Objetivos

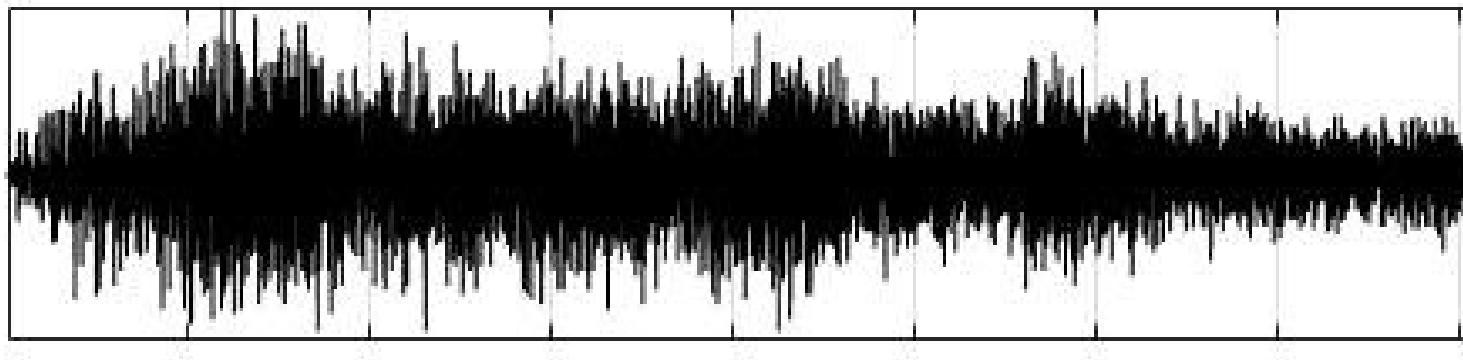
- Lograr una clasificación automática de géneros musicales.
- Afianzar y utilizar herramientas aprendidas en el curso.
- Estudiar las señales de audio de manera local, zonificando a la misma en tres partes y obteniendo características representativas de las mismas, para luego poder clasificarlas de manera automática.

Pipeline del Proyecto

Pasos para la realización del proyecto

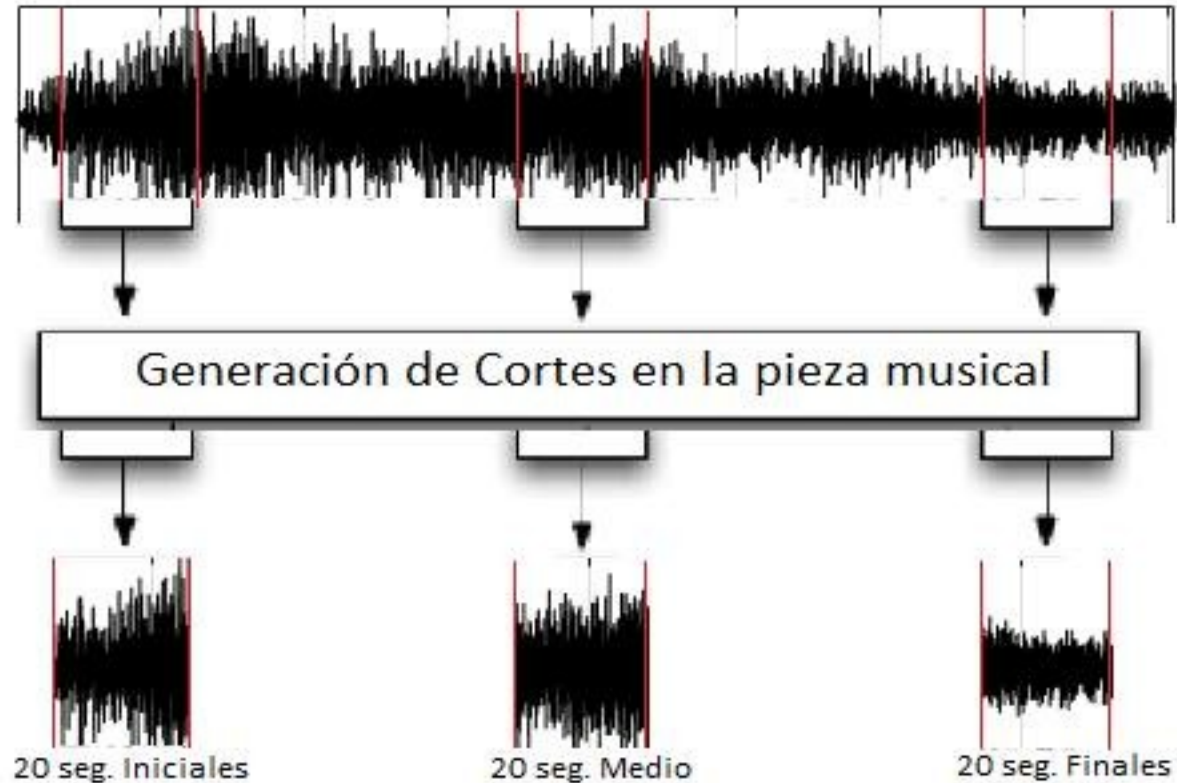


Señal de entrada



Señal de audio

Zonificación



Dividimos a la señal en tres partes iguales

Características

- Pico máximo de frecuencia
- Energía de corta duración
- Flujo espectral \rightarrow Media y Varianza
- Tasa de cruces por cero (ZCR)
- Coeficientes cepstrales en escala de Mel

Pico máximo de frecuencia

- El pico máximo de frecuencia lo utilizamos para tener una referencia de la máxima frecuencia presente en cada género.

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

Energía de corta duración

La energía de corta duración de una señal de audio determina la variación de la amplitud en relación al tiempo

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

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.
- Utilizamos la media y la varianza.

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

Tasa de cruces por cero (ZCR)

- La tasa de cruces por cero puede ser utilizada para determinar el contenido frecuencial de una señal. Con esta característica podemos encontrar la frecuencia fundamental de la señal de audio, de suma importancia, ya que podemos determinar el tono

$$x = x / \text{norm}(x)$$

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

Coeficientes cepstrales (Mel)

- Los coeficientes cepstrales 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.
- Calculamos 40 coeficientes, nos quedamos con los primeros 15, que serán los más representativos.

Armado de plantilla

- Plantilla conformada por 20 características por los 3 tramos de cada canción, es decir, 60 características por canción.
- Calculamos dichas características a las 10 canciones que conformarán nuestra base de datos por cada género.
- Utilizamos la media por cada característica.
- Obtendremos como salida 60 características medias por cada plantilla, 20 representarán las primeras partes, 20 las partes del medio y 20 las partes del final de cada género.

Comparación

- Utilizamos una señal de audio de prueba, a la cual le calculamos las 60 características.
- Buscamos la distancia mínima, característica a característica, con las 4 plantillas, guardando en un contador la cantidad de ocurrencias por cada característica.
- Como salida tendremos 4 contadores con 3 valores cada uno, expresando la cantidad de ocurrencias de cada parte de la canción de prueba con todas las características medias por parte.

Comparación

- Finalmente, podemos visualizar a que genero se vincula cada una de las partes.
- Coeficientes les dimos la mitad del peso.

```
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|
```

Canción de tango como
prueba

Resultados

- 10 canciones de pruebas por cada género

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

Resultados

- Porcentaje de acierto.

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%

Conclusiones

- El método que diseñamos para detectar los géneros musicales tiene un 78.3% de efectividad.
- Estudio podría extenderse evaluando que sucede al aumentar canciones en cada una de las bases de datos para los distintos géneros.
- 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.
- Evaluar la búsqueda de otra característica para mejorar aún más el detalle de cada género.



¿Preguntas?