

UNIVERSIDAD DE ALCALÁ  
Escuela Politécnica Superior

**GRADO EN INGENIERÍA EN TECNOLOGÍAS DE  
TELECOMUNICACIÓN**

Trabajo Fin de Grado

**CLASIFICADOR DE GÉNEROS MUSICALES  
BASADO EN ALGORITMOS DE SOFT COMPUTING**

**Autor:** Víctor Alonso Sánchez  
Director/es: Dr. Enrique Alexandre Cortizo

**TRIBUNAL:**

**Presidente:** Dr. Sancho Salcedo Sanz

**Vocal 1º:** Dr. Lucas Cuadra Rodríguez

**Vocal 2º:** Dr. Enrique Alexandre Cortizo

**CALIFICACIÓN:** .....

**FECHA:** .....



# AGRADECIMIENTOS

Como no podía ser menos, quiero aprovechar este documento para dar las gracias a todas esas personas que han sido muy importantes para mí en mi día a día durante esta etapa.

En primer lugar, agradecer al Dr. Enrique Alexandre su dedicación y esfuerzo a la hora de explicarme los conceptos necesarios para el desarrollo de este trabajo, por ayudarme cuando no sabía cómo continuar y por resolverme todas las dudas que me han ido surgiendo durante estos meses.

En segundo lugar, quiero dar las gracias a mi familia, que siempre han estado pendientes de cómo me iba yendo y que siempre se han alegrado por todos los logros que he ido consiguiendo. En especial, dar las gracias a mis padres, Daniel y Ángela y a mi hermana, Raquel, por estar ahí, al pie del cañón, desde el día en que comencé esta carrera hasta el final. Ellos han sido los que han aguantado carros y carretas, cabreos, frustraciones y malos momentos, pero también han estado ahí compartiendo todas las alegrías y buenas noticias. Por eso, muchísimas gracias.

Las siguientes líneas se las quiero dedicar a todos mis compañeros de la UAH. Han sido muchos años juntos, tantos que os habéis convertido en una segunda familia para mí. Sin vuestro apoyo y vuestra ayuda durante todos estos años, esto no habría sido posible. Quiero hacer una mención especial a Laura Cornejo. Durante los meses en los que he estado trabajando en este proyecto, ella ha sido un gran apoyo para mí, ayudándome en todo momento y resolviéndome un sinfín de dudas.

También quiero dar las gracias a Jesica Martín, por estar apoyándome desde el minuto uno, haciéndome *feedback* de todo lo que iba escribiendo y dándome consejos para mejorarlo.

Y no podía faltar una mención especial para Mario Blanca, mi amigo, mi hermano. Todas las cosas que nos han pasado durante tantos años nos han hecho más fuertes y por eso quiero compartir este trabajo contigo. Espero que sigamos juntos por mucho tiempo y que estemos presentes en los acontecimientos importantes que nos vayan sucediendo.

Por todo esto, muchísimas gracias

A todos.



# ÍNDICE GENERAL

ÍNDICE GENERAL.....	4
ÍNDICE DE FIGURAS .....	6
ÍNDICE DE TABLAS.....	8
RESUMEN.....	10
ABSTRACT .....	12
RESUMEN EXTENDIDO.....	14
1. INTRODUCCIÓN .....	16
2. BASE DE DATOS.....	18
2.1 INTRODUCCIÓN.....	18
2.2 BÚSQUEDA DE LA BASE DE DATOS.....	18
2.3 ADECUACIÓN DE LA BASE DE DATOS.....	18
3. PROCESADO DE SEÑAL Y EXTRACCIÓN DE CARACTERÍSTICAS .....	22
3.1 PROCESADO DE SEÑAL .....	22
3.2 EXTRACCIÓN DE CARACTERÍSTICAS.....	23
3.3 IMPLEMENTACIÓN EN MATLAB .....	28
4. CLASIFICACIÓN DE GÉNEROS MUSICALES .....	30
4.1 INTRODUCCIÓN.....	30
4.2 REDES NEURONALES ARTIFICIALES .....	30
4.3 SVM (SUPPORT VECTOR MACHINE): .....	32
4.4 IMPLEMENTACIÓN DE SVM EN MATLAB .....	35

5.	IMPLEMENTACIÓN DEL CLASIFICADOR Y PRIMEROS RESULTADOS..	38
5.1	CONSTRUCCIÓN DEL CLASIFICADOR .....	38
5.2	PRIMEROS RESULTADOS .....	39
6.	MEJORAS EN EL CLASIFICADOR .....	42
6.1	NORMALIZACIÓN DE LA MATRIZ DE CARACTERÍSTICAS.....	42
6.2	IGUALACIÓN DEL NÚMERO DE ELEMENTOS EN LA BASE DE DATOS.....	44
6.3	AGRUPACIÓN DE GÉNEROS MUSICALES.....	46
6.4	CLASIFICACIÓN DE GÉNEROS SIMILARES. ....	49
7.	CONCLUSIONES .....	54
8.	LÍNEAS FUTURAS .....	56
	BIBLIOGRAFÍA .....	58

# ÍNDICE DE FIGURAS

Figura 1. Ventana de Hanning en el dominio temporal y espectral .....	24
Figura 2. Proceso para obtener los MFCCs.....	26
Figura 3. Escalado y suavizado de los coeficientes de Mel .....	27
Figura 4. Esqueleto de una red neuronal artificial.....	31
Figura 5. En esta figura se muestra el hiperplano óptimo para la clasificación de dos clases ....	32
Figura 6. Función kernel polinomial .....	33
Figura 7. SVM con kernel perceptrón (Redes Neuronales) .....	34
Figura 8. Efecto de la transformación con kernel Gaussiano en función de sigma.....	34
Figura 9. Validación cruzada para $K = 5$ . .....	38





# ÍNDICE DE TABLAS

Tabla 1. Comparativa entre los géneros de la base de datos original y la base de datos modificada utilizada en el desarrollo del trabajo.....	19
Tabla 2. Estructura de la matriz de características. Los datos son reales .....	28
Tabla 3. Géneros musicales con sus respectivas etiquetas .....	29
Tabla 4. Prototipo de la función de entrenamiento de la SVM .....	36
Tabla 5. Prototipo de la función de predicción de la SVM .....	36
Tabla 6. Porcentaje de acierto para cada género sin normalizar la matriz de características .....	39
Tabla 7. Probabilidad de acierto media, coef de correlación medio y MSE media del clasificador al introducir la matriz de características sin normalizar.....	40
Tabla 8. Matriz de confusión de los géneros musicales sin normalizar la matriz de características .....	40
Tabla 9. Porcentaje de acierto para cada género normalizando la matriz de características .....	42
Tabla 10. Probabilidad de acierto media, coef de correlación medio y MSE media del clasificador al introducir la matriz de características normalizada. ....	43
Tabla 11. Matriz de confusión de los géneros musicales normalizando la matriz de características .....	43
Tabla 12. Porcentaje de acierto para cada género normalizando la matriz de características y equiparando el número de elementos de cada género. ....	44
Tabla 13. Probabilidad de acierto media, coef de correlación medio y MSE media del clasificador al introducir la matriz de características normalizada y la base de datos compensada. ....	45
Tabla 14. Matriz de confusión de los géneros musicales normalizando la matriz de características e igualando el número de canciones de cada género. ....	45
Tabla 15. Combinación de géneros musicales similares.....	46

Tabla 16. Nuevos géneros musicales con sus respectivas etiquetas.....	47
Tabla 17. Porcentaje de acierto para cada género normalizando la matriz de características y equiparando el número de elementos de cada género después de combinar géneros. ....	47
Tabla 18. Probabilidad de acierto media, coef de correlación medio y MSE media del clasificador al introducir la matriz de características normalizada y la base de datos compensada después de combinar géneros musicales. ....	48
Tabla 19. Matriz de confusión de los géneros musicales normalizando la matriz de características e igualando el número de canciones de cada género. ....	48
Tabla 20. Porcentaje de acierto para los subgéneros del género CLASICAL.....	49
Tabla 21. Probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género CLASICAL. ....	50
Tabla 22. Matriz de confusión para los subgéneros del género CLASICAL. ....	50
Tabla 23. Porcentaje de acierto para los subgéneros del género JAZZ/BLUES. ....	50
Tabla 24. Probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género JAZZ/BLUES.....	51
Tabla 25. Matriz de confusión para los subgéneros del género JAZZ/BLUES. ....	51
Tabla 26. Porcentaje de acierto para los subgéneros del género METAL. ....	51
Tabla 27. Probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género METAL. ....	52
Tabla 28. Matriz de confusión para los subgéneros del género METAL.....	52
Tabla 29. Porcentaje de acierto para los subgéneros del género ROCK.....	52
Tabla 30. Probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género ROCK.....	53
Tabla 31. Matriz de confusión para los subgéneros del género ROCK. ....	53

# RESUMEN

En este Trabajo de Fin de Grado se pretende diseñar un clasificador de géneros musicales basado en algoritmos de soft computing que comparan las características tímbricas de distintas canciones pertenecientes a varios géneros.

En primer lugar se creará una base de datos de todas las canciones que se van a utilizar. Posteriormente se extraerán las características pertinentes de cada canción para después implementar un clasificador para distinguir entre los distintos géneros musicales. Finalmente se buscarán mejoras en función de los resultados obtenidos.

**Palabras clave:** género musical, base de datos, características, clasificador.



# ABSTRACT

In this Final Grade Work will be sought to design a classifier based on musical genres soft computing algorithms that compare the timbral characteristics of different songs belonging to different genres.

First, a database of every song will be built. The next step is to extract the relevant characteristics of each song and then to implement a classifier to distinguish between the different musical genres. Finally, improvements will be sought on the basis of the results obtained.

**Key words:** musical genre, database, characteristics, classifier.



# RESUMEN EXTENDIDO

Los motivos principales por los cuales se ha decidido desarrollar este trabajo fin de grado son la gran variedad de géneros musicales que existen en la actualidad y la complejidad que existe a la hora de decidir a qué tipo de género musical pertenece una canción. En este capítulo se describirá, mediante un amplio resumen, cómo se ha llevado a cabo el desarrollo de este proyecto.

El primer capítulo de este trabajo es una breve introducción en la que se sitúan las bases del trabajo y el desarrollo del mismo.

El segundo capítulo trata de la obtención de la base de datos de canciones y de cómo ha sido etiquetada. También se verá como se ha ido modificando la base de datos original para obtener mayor variedad de géneros y distintos resultados.

El tercer capítulo habla sobre las distintas características que se han extraído de cada canción. Este capítulo tendrá una connotación más teórica en la que se explicará el procesado de señal que se realiza en cada archivo de audio. Se verá cuantas características se han obtenido y cómo se obtienen de forma teórica. Finalmente se expone cómo se consiguen dichas características en el entorno de trabajo utilizado, que se trata de Matlab en este caso.

En el cuarto capítulo se habla de los métodos que se utilizan para la clasificación. Primero se hará una introducción que exponga lo que son las redes neuronales artificiales, que serán la base de este trabajo, ya que intentan emular el aprendizaje humano en una máquina. Después de esta visión general de las redes neuronales, se hará hincapié en una máquina de aprendizaje en concreto (SVM) que es la utilizada en el desarrollo de esta práctica y se mostrarán las diferencias que existen entre esta máquina y las redes neuronales artificiales. Primero se hará una descripción teórica del algoritmo y después se mostrará la manera de implementarlo en Matlab para su correcto funcionamiento.

En el quinto apartado de este trabajo se recoge como se ha implementado el clasificador musical en sí. Este clasificador funciona introduciendo un grupo de las características obtenidas previamente y comparará con otro grupo distinto y se obtendrán los resultados. Como estos resultados no serán muy buenos, se verá cómo se puede mejorar dicho clasificador, cambiando parámetros, normalizando las características, etc.

En el sexto capítulo se habla de todas las mejoras mencionadas anteriormente. Primero se aplican normalizaciones a las características obtenidas y después se igualará el número de archivos de audio para cada género, para que no haya grandes diferencias entre ellos. Para mejorar el acierto del clasificador, se va a proponer combinar géneros parecidos, realizar la clasificación, y después

ver como clasificaría las canciones dentro de esa combinación. Por último se mostrarán los resultados obtenidos con estas mejoras.

Por último se mostrarán las conclusiones sobre la realización del proyecto y las líneas futuras en las que se puede aplicar este proyecto.



# 1. INTRODUCCIÓN

La música ha estado presente en la sociedad desde hace miles de años. Ya en la prehistoria, los hombres utilizaban ritmos musicales para diversas funciones, ya fuesen rituales de caza, fiestas, rituales religiosos e incluso para la comunicación entre ellos. Estos ritmos han ido evolucionando con el paso de los años, adquiriendo mayor complejidad y diversidad. Así llegamos a la actualidad, en la que la inmensa variedad musical es abrumadora, por lo que la clasificación de ésta se torna harto complicada.

Debido a la gran cantidad de música existente surge la necesidad de clasificarla para facilitar la búsqueda de alguna canción en concreto. Hay diferentes métodos de clasificación, como puede ser por artista, año o incluso condición anímica. Pero la más común es la clasificación por género musical.

Este tipo de clasificación es la más complicada de realizar, ya que existe una gran imprecisión a la hora de distinguir dos canciones de géneros parecidos, con las mismas características rítmicas o instrumentales. A este problema hay que añadir el factor humano, ya que dos personas distintas pueden interpretar de forma diferente una misma canción y clasificarla en géneros distintos.

Debido al gran coste que supone la clasificación de géneros musicales, tanto temporal como económico, la industria musical se ha esforzado por buscar formas eficientes para realizar este trabajo. Para solucionar este problema se ha desarrollado este trabajo fin de grado, el cual se encarga de optimizar el tiempo de ejecución y los costes que conlleva la clasificación de un repertorio musical amplio, manteniendo un porcentaje de acierto aceptable.

Para conseguirlo se seguirán unos sencillos pasos. El primero será la construcción de una base de datos lo suficientemente extensa y con gran variedad de géneros musicales y etiquetarla de forma correcta. A partir de esta base de datos se procederá a obtener las características musicales de cada canción, las cuales servirán como patrón para que el clasificador las diferencie correctamente y pueda clasificarlas de forma óptima. Por último, una vez implementado el clasificador y comprobando que funciona correctamente, se plantearán ciertas mejoras para obtener un buen porcentaje de acierto. Los resultados finales que obtendremos serán los que nos sirvan para este trabajo fin de grado.



## **2. BASE DE DATOS**

### **2.1 INTRODUCCIÓN**

Para la correcta realización de este proyecto es necesaria la obtención de una base de datos extensa y perfectamente etiquetada. Gracias a esta base de datos se podrán extraer las características de todas las canciones pertenecientes a los distintos géneros musicales. En los siguientes apartados de este capítulo, se mostrará cómo se ha obtenido la base de datos y como se ha ido adecuando a las necesidades que presenta este trabajo.

### **2.2 BÚSQUEDA DE LA BASE DE DATOS**

Como bien es sabido, internet es una fuente de información casi sin límite, de la cual se puede obtener cualquier tipo de información si se busca con el criterio adecuado. Es por ello que la base de datos utilizada será extraída de la red.

Concretamente, la base de datos sobre la que se ha trabajado ha sido extraída de la página del MIREX. MIREX (Musical Information Retrieval Evaluation eXchange) son unas conferencias que se vienen celebrando desde el año 2005 que se centran en la investigación de la información musical.

En el apartado de clasificación por género musical dentro de la web del MIREX, hay unos enlaces de descarga gratuita, de los cuales se ha obtenido la base de datos principal para este trabajo, llamada Music Technology Group (2004) [5]. Esta base de datos viene dividida en dos partes, una para entrenamiento y otra para validación y así comprobar que el porcentaje de acierto que se obtiene después de entrenar es válido.

La base de datos está dividido en 8 géneros distintos: clasical, electronic, jazz, metal, punk, rock, pop y world. En total, la base de datos completa cuenta con unos 1000 archivos de audio.

### **2.3 ADECUACIÓN DE LA BASE DE DATOS**

Una vez descargada la base de datos, hay que adecuarla para poder trabajar con ella en Matlab. Pero primero se han hecho varios cambios con respecto a la base de datos original descargada. Debido a que dentro de varios géneros había canciones que presentaban muchas diferencias al oído, se ha decidido aumentar el número de géneros musicales. Además, se ha decidido añadir

algún género más, para dotar de más diversidad a la base de datos. En la Tabla 1 se muestran las diferencias existentes entre la base de datos original descargada de la página web del MIREX y la que finalmente se ha utilizado en el desarrollo del trabajo.

<i>Base de datos original</i>	<i>Base de datos final</i>
<b>CLASICAL</b>	CHOIR
	LYRIC
	ORCHESTRA
<b>JAZZ</b>	JAZZ
	BLUES
<b>ELECTRONIC</b>	ELECTRONIC
<b>POP</b>	POP
<b>METAL</b>	METAL
	METALCORE
	DEATH METAL
	POWER METAL
	GOTHIC METAL
<b>ROCK</b>	ROCK
	PUNK ROCK
	FOLK ROCK
	ROCK & ROLL
<b>PUNK</b>	HIP HOP
<b>WORLD</b>	WORLD
	FLAMENCO
	LATIN
	FOLK

*Tabla 1. Comparativa entre los géneros de la base de datos original y la base de datos modificada utilizada en el desarrollo del trabajo.*

Como se puede apreciar, la base de datos original contaba con 8 géneros distintos y la nueva aumenta este número a 21. A priori, puede ser una complicación a la hora de clasificar correctamente cada canción dentro de su género, pero se verá posteriormente que se obtendrá un porcentaje de acierto aceptable.

También se cambiaron algunas canciones de las que traía originalmente la base de datos, eliminando alguna y añadiendo otras ya que con las que traía, el número era insuficiente.

Después de esto, se cambió el formato de cada archivo de audio, ya que el formato original era *.mp3* y no es compatible con Matlab.

Utilizando un software gratuito, se cambió el formato de cada canción a *.wav*, que es un formato compatible con Matlab. Para trabajar con este formato se utilizó la librería incluida en Matlab “wavread”. La función que utilizamos para obtener información de cada canción es la que da nombre a la librería.

Finalmente, la base de datos que se utilizó constaba de 1234 archivos de audio. En algunos géneros, el número de canciones es ligeramente superior al de otros, lo cual dificultará su correcta clasificación, pero eso se solucionará más adelante.



### **3. PROCESADO DE SEÑAL Y EXTRACCIÓN DE CARACTERÍSTICAS**

En este capítulo se va a contar cómo se han extraído las distintas características de los archivos de audio de la base de datos. Con la aplicación de diversas técnicas de procesado de señal se van a obtener unas 10 características distintas de cada canción.

Primero se va a estudiar las distintas técnicas que se aplican para obtener las características. Después se van a obtener las distintas características tímbricas de cada archivo de audio. Por último, se mostrará cómo se han implementado estas técnicas en el entorno de trabajo utilizado para la obtención de todas las características.

Toda la información de este apartado se ha extraído del paper de George Tzanetakis y Perry Cook “*Musical Genre Classification of Audio Signals*” publicado el año 2002.

#### **3.1 PROCESADO DE SEÑAL**

Antes de extraer cualquier característica hay que hacer un procesado de la señal para obtener mejores resultados. Para ello se decide tomar una ventana de entramado de tamaño 512 muestras. Se elige este tamaño debido a que las señales acústicas poseen una naturaleza aleatoria, lo cual supone un gran inconveniente a la hora de extraer información de ellas. Para solucionar este problema se toman valores de ventana muy pequeños. Con estos valores de trama y sabiendo que la frecuencia de muestreo es de 11 kHz obtenemos una duración de trama de 46.55 ms aproximadamente. Dada la corta duración de la trama, se consigue que la señal encapsulada sea prácticamente estacionaria, lo que facilitará la extracción de características de ella. Otra ventaja del entramado de la señal es que se puede conseguir mejor información, ya que se realizará la media y la desviación típica de todas las tramas de cada archivo de audio.

De esta forma se obtienen dos medidas de una sola característica, en lugar de si se calculase dicha característica en cuestión de la señal de audio sin entamar.

## 3.2 EXTRACCIÓN DE CARACTERÍSTICAS

En este apartado se va a explicar de forma teórica las técnicas que se van a manejar para la extracción de características de los distintos archivos de audio. Como se dijo al inicio de este capítulo, toda esta información ha sido extraída de [6].

Este trabajo se va a centrar en la extracción de características musicales, relacionadas con el timbre. Dentro de este grupo de características se pueden distinguir dos tipos: características temporales y características espectrales.

- **Características temporales:**

Dentro de este grupo, se han calculado la energía de la señal en alta frecuencia y el número de cruces por cero.

**Energía de la señal en alta frecuencia (HFE):** se calcula en cada trama obteniendo el módulo al cuadrado de la trama, como se muestra en la ecuación (1).

$$E_k = \sum_{n=0}^N |x[n]|^2 \quad (1)$$

Después de calcular la energía de cada trama, se calculan la media y la desviación típica de la energía de la señal. Como mejora, se han eliminado las zonas de la señal que no tienen sonido. Esto se ha hecho estableciendo un umbral y eliminando las partes de la señal que no superen ese umbral.

**Número de cruces por cero (zerocrossing):** con este parámetro se calcula la cantidad de ruido que tiene una señal viendo cuantas veces cruza por cero. Cuantas más veces pase la señal por el cero, más ruidosa es. Esta característica se utiliza para distinguir la voz de la parte instrumental. Se ha hecho uso de la ecuación (2) para obtener el zerocrossing.

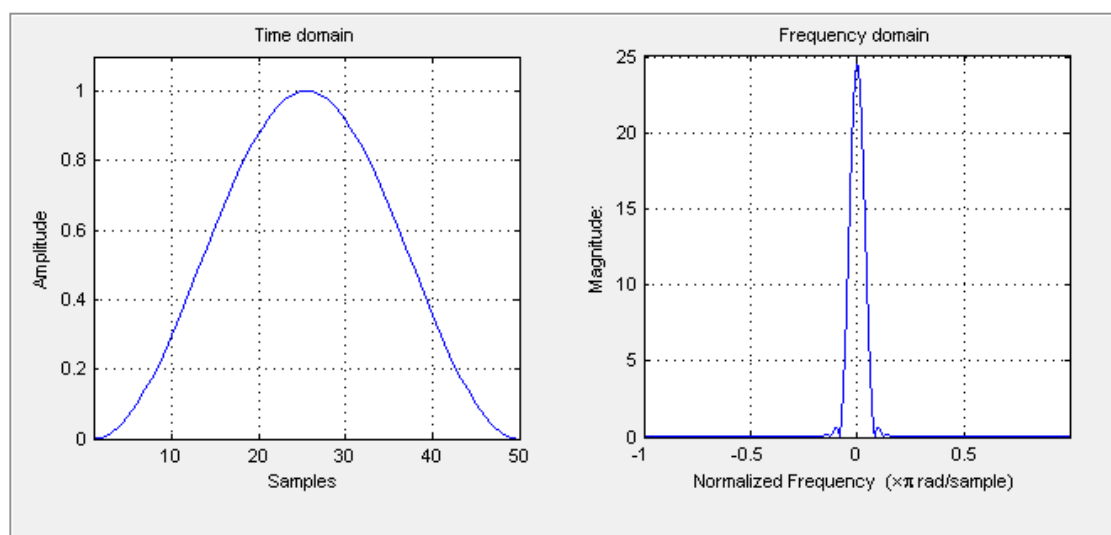
$$Z_k = \frac{1}{2} \cdot \sum_{n=0}^{N-1} |\text{sign}(x[n]) - \text{sign}(x[n+1])| \quad (2)$$

Como en el caso anterior, esto se calcula para cada trama y posteriormente se obtiene la media y la desviación típica.



- **Características espectrales:**

Para obtener este tipo de características se ha decidido utilizar una ventana de Hanning para cada trama. La ventana de Hanning proporciona un suavizado en los bordes de la trama, lo cual evita cambios bruscos entre tramas. Este es uno de los motivos por los que se ha elegido este tipo de ventana para esta parte. Otro de los motivos es que, en el dominio espectral, una multiplicación en el dominio temporal se transforma en una convolución. Por eso, para calcular las características espectrales es necesario utilizar una función que al pasarla al dominio de la frecuencia, se aproxime más a una función delta, ya que esta función no modifica el espectro de la señal original. En la siguiente figura se muestra como es una ventana de Hanning en el dominio temporal y en el dominio espectral.



*Figura 1. Ventana de Hanning en el dominio temporal y espectral*

En la imagen se puede ver cómo se aproxima mucho a una delta en el dominio de la frecuencia, por lo que es la mejor opción para este proyecto y presenta una mejor solución que una ventana rectangular, por ejemplo, ya que su transformada en el dominio de la frecuencia es la de una función *sinc*.

Las características espectrales que se han extraído de los archivos de audio de la base de datos son el centroide espectral, el rolloff, el flujo espectral y los MFCC.

**Centroide espectral:** se define como el centro de gravedad del espectro de la señal. Es la frecuencia que divide al espectro en dos partes iguales. Este parámetro se relaciona con

el “brillo” del sonido, por lo que a los sonidos más brillantes les corresponde un valor del centroide más alto. Para el cálculo del centroide en cada trama se utiliza la ecuación (3), que se muestra a continuación.

$$C_k = \frac{\sum_{n=0}^N (M_k[n] \cdot n)}{\sum_{n=0}^N M_k[n]} \quad (3)$$

$M_k[n]$  es la amplitud de la transformada de Fourier de la trama  $k$  a la frecuencia  $n$ .

Una vez calculado el centroide de todas las tramas, como en los casos anteriores se calculan la media y la desviación típica, ya que son los valores que utilizaremos para nuestro comparador.

**Factor de Rolloff:** el factor de Rolloff es el valor de la frecuencia  $R$  por debajo de la cual se encuentra el 85% del módulo del espectro de la señal. De esta manera se puede tener una idea de la forma que puede adoptar el espectro de la señal. Se calcula haciendo uso de la ecuación (4).

$$\sum_{n=0}^R M_k[n] = 0.85 \cdot \sum_{n=0}^N M_k[n] \quad (4)$$

Como en todos los casos, una vez calculado el factor de Rolloff de todas las tramas se calculan la media y la desviación típica.

**Flujo espectral:** esta característica muestra la rapidez con la que cambia la energía del espectro de la señal. Se calcula, como se muestra en la ecuación (5), como la diferencia al cuadrado del módulo de dos tramas consecutivas del espectro.

$$F_k = \sum_{n=0}^N (M_k[n] - M_{k-1}[n])^2 \quad (5)$$

Una vez calculados todos los flujos de todas las tramas, se calculan la media y la desviación típica, como en los casos anteriores.

**Coefficientes cepstrales de Mel (MFCC):** la obtención de los coeficientes cepstrales de Mel es una de las técnicas más utilizadas y más importantes en sistemas de reconocimiento de voz. Estos coeficientes vienen derivados de la aplicación del Cepstrum. El Cepstrum es un operador que transforma una convolución en el dominio

temporal en una suma en el dominio de la frecuencia. De esta forma se consigue diferenciar en una señal de audio la voz del resto de sonidos. La definición del operador Cepstrum es la transformada inversa de Fourier del logaritmo del espectro de la señal de voz, como se muestra en la ecuación (6).

$$\text{Cepstrum}(s[n]) = \hat{s}[n] = F^{-1}[\log(|F(s[n])|)] \quad (6)$$

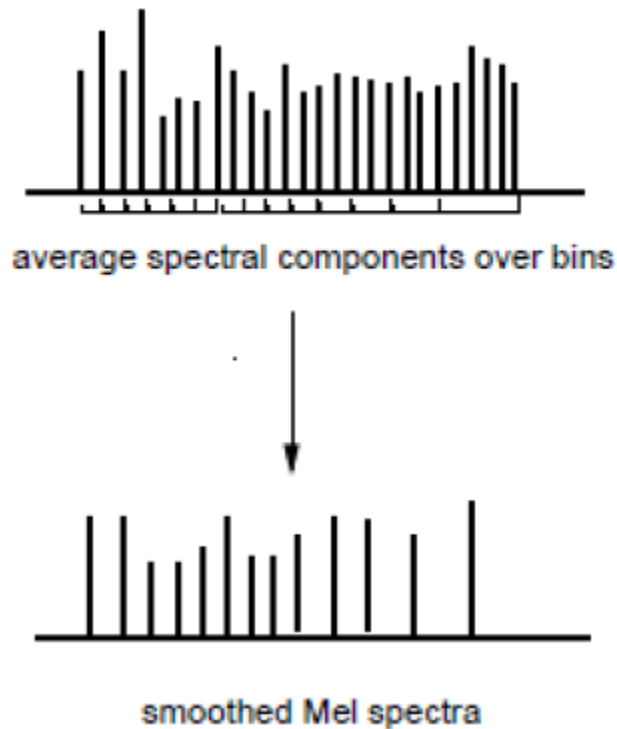
Los pasos que hay que seguir para la obtención de los MFCCs se muestran en la siguiente figura:



*Figura 2. Proceso para obtener los MFCCs*

Como ya se ha mencionado durante todo este capítulo, se trabaja con tramas de la señal original. Una vez tenemos entramada la señal de audio se aplica la DFT a cada trama para

trabajar en el dominio espectral. Posteriormente se aplica un logaritmo al módulo del espectro de la señal. Esto se hace porque se ha averiguado que el volumen de la señal de audio se puede aproximar de forma logarítmica. El penúltimo paso es suavizar el espectro quedándose con las frecuencias más significativas, como se muestra en la siguiente figura.



*Figura 3. Escalado y suavizado de los coeficientes de Mel*

Para el habla, las frecuencias más bajas son más significativas, por eso los puntos del espectro no están equiespaciados por norma general, sino que sigue la escala frecuencial de Mel. Esta escala está basada en la relación que existe entre la frecuencia real y el tono percibido, ya que el oído humano no funciona de forma lineal. Para aproximar esta escala se ha utilizado un banco de filtros.

Por último, como los coeficientes de Mel están muy correlados, hay que aplicar a cada trama una transformada para descorrelarlos. Esto se consigue aplicando la transformada discreta del coseno (DCT).

Finalmente, como en el resto de casos, se calcula la media y la desviación típica de cada uno de los coeficientes calculados.

### 3.3 IMPLEMENTACIÓN EN MATLAB

Para obtener las características mencionadas en el apartado anterior, ha sido necesario crear varios scripts en Matlab, uno para cada característica, siguiendo las ecuaciones descritas anteriormente. Después se ha creado otro que organiza todas las características en una matriz de la siguiente forma: de la primera columna a la columna 39, corresponden a la media y a la desviación típica de los coeficientes de Mel (se han calculado 20 coeficientes). Después, de la columna 40 a la 49 corresponden, en este orden, a la media y la desviación típica de la energía de la señal, el centroide, el Rolloff, el flujo espectral y el zero crossing. La última columna corresponde a la etiqueta del género musical al que pertenece el archivo. En cuanto al número de filas, éste es igual al número de archivos de audio de la base de datos. En la Tabla 2 se muestra un ejemplo del formato de la matriz de características.

MFCC 1	...	MFCC 40	M(HFE)	Std(HFE)	M(centroide)	Std(centroide)	M(Rolloff)	Std(Rolloff)	M(flux)	Std(flux)	M(ZCE)	Std(ZCE)	Etiqueta
0,026		-0,344	0,023	0,104	0,100	0,016	0,074	-0,009	-0,256	-0,237	0,084	-0,092	1
0,016		0,195	0,007	0,015	0,415	0,082	0,373	0,064	-0,311	-0,227	0,355	0,138	2
0,287		-0,117	0,214	0,115	0,367	0,051	0,348	0,006	0,484	0,709	0,337	0,211	3
0,153		0,065	-0,075	-0,082	-0,020	-0,033	-0,027	-0,084	0,676	0,717	-0,148	-0,075	4
-0,523		0,077	-0,113	-0,096	-0,338	-0,130	-0,390	-0,159	0,311	0,508	-0,264	-0,217	5
-0,527		-0,065	-0,113	-0,095	-0,286	-0,050	-0,350	-0,111	-0,144	0,136	-0,205	-0,124	6

Tabla 2. Estructura de la matriz de características. Los datos son reales

La asignación del número de la etiqueta que corresponde a cada género musical se muestra en la Tabla 3.

<i>Género musical</i>	<i>Etiqueta</i>
<i>ORCHESTRA</i>	1
<i>CHOIR</i>	2
<i>LYRIC</i>	3
<i>ELECTRONIC</i>	4
<i>FOLK</i>	5
<i>HIP HOP</i>	6
<i>JAZZ</i>	7
<i>BLUES</i>	8
<i>LATIN</i>	9
<i>POP</i>	10
<i>WORLD</i>	11
<i>DEATH METAL</i>	12
<i>GOTHIC METAL</i>	13
<i>METAL</i>	14
<i>METALCORE</i>	15
<i>POWER METAL</i>	16
<i>FOLK ROCK</i>	17
<i>PUNK ROCK</i>	18
<i>ROCK</i>	19
<i>ROCK &amp; ROLL</i>	20
<i>FLAMENCO</i>	21

*Tabla 3. Géneros musicales con sus respectivas etiquetas*

Una vez construida la matriz de características, se guarda en formato *.mat*, el cual es compatible con Matlab, para cargarla posteriormente e introducirla al comparador.

## **4. CLASIFICACIÓN DE GÉNEROS MUSICALES**

### **4.1 INTRODUCCIÓN**

Para la clasificación de los distintos géneros musicales se va a implementar una red neuronal artificial. Este tipo de redes emulan el funcionamiento del sistema nervioso humano a la hora de procesar información. Por eso, las redes neuronales artificiales presentan características propias del cerebro humano, como el aprendizaje, la generalización o la abstracción.

El aprendizaje consiste en adquirir el conocimiento de las cosas mediante el estudio o la experiencia. Esto hace que las redes neuronales artificiales posean la capacidad de cambiar dinámicamente según las condiciones del medio.

Gracias a la generalización se pueden extraer conclusiones globales de casos particulares. De esta forma, las redes neuronales pueden obtener resultados correctos a entradas distorsionadas (dentro de un margen).

Por último, la abstracción separa las cualidades de un objeto para considerarlas individualmente. Esto permite diferenciar un conjunto de datos que no muestren características comunes.

Todo esto hace que una red neuronal artificial tenga un comportamiento no lineal, lo que permite procesar información de procesos no lineales. Dado que la mayoría de los fenómenos que se dan en la naturaleza son no lineales, este hecho es muy útil para estudiarlos.

A continuación se va a explicar con más detalle el funcionamiento de las redes neuronales artificiales para entender el funcionamiento de la máquina empleada en este trabajo.

### **4.2 REDES NEURONALES ARTIFICIALES**

Las redes neuronales artificiales constan, típicamente, de tres capas. Cada capa tiene una serie de neuronas interconectadas que permiten aproximar cualquier función continua si se definen los hiper parámetros de la red de forma correcta.

En la siguiente figura se muestra la estructura de una red neuronal típica donde se pueden apreciar los 3 estratos que la componen y que serán descritos a continuación.

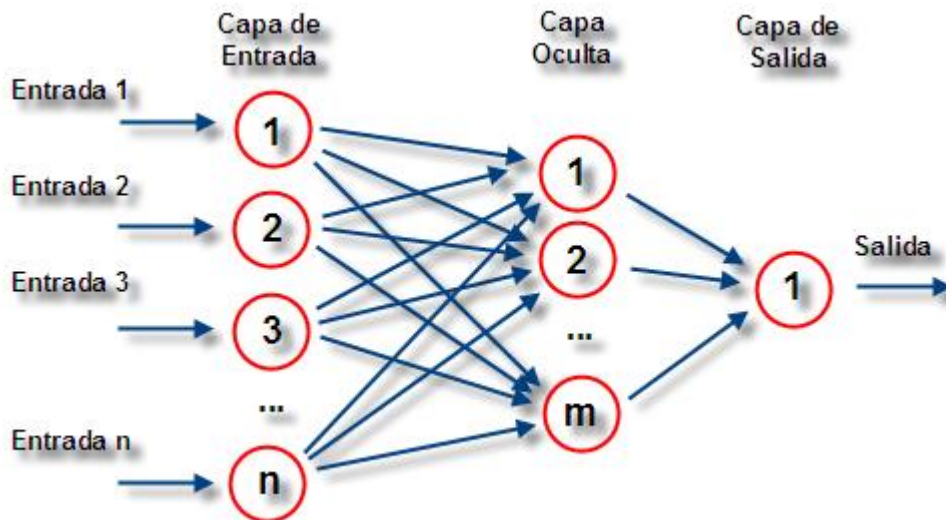


Figura 4. Esqueleto de una red neuronal artificial (Wikipedia) [14]

**Capa de entrada:** en esta capa se sitúan las neuronas que contienen los patrones de entrada a la red.

**Capa oculta:** en esta capa, la entrada depende de la salida de la capa anterior y su salida pasa a la capa posterior

**Capa de salida:** es la última capa de la red neuronal y es donde se encuentran las neuronas que contienen las salidas de la red neuronal.

A la hora de implementar estas redes, lo más habitual es fijar el número de neuronas  $M$  e inicializar los pesos de forma aleatoria. Posteriormente se entrena la red con un conjunto de entrenamiento mediante el algoritmo de “back propagation” junto con métodos de optimización basados en el gradiente. Finalmente, mediante el proceso de validación cruzada se obtiene el número óptimo de neuronas. Este proceso es muy lento y tiene un coste computacional muy alto debido al elevado número de pasos que se requieren para entrenar la red.

Por esta razón, han ido surgiendo diferentes algoritmos que optimizan el entrenamiento y el aprendizaje. Uno de ellos es el SVM (Support Vector Machine). Dicho algoritmo mejora el tiempo de ejecución sin disminuir la precisión, por lo que se ha decidido utilizar en el presente trabajo. En el siguiente apartado se explicará cómo funciona este tipo de máquina clasificadora.



### 4.3 SVM (SUPPORT VECTOR MACHINE)

Las máquinas de soporte vectorial o máquinas de vectores de soporte (Support Vector Machines, SVMs) son un conjunto de algoritmos de aprendizaje supervisado.

Estos métodos están propiamente relacionados con problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento (de muestras) podemos etiquetar las clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra. Intuitivamente, una SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases por un espacio lo más amplio posible. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de su proximidad pueden ser clasificadas a una u otra clase.

Más formalmente, una SVM construye un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad muy alta (o incluso infinita) que puede ser utilizado en problemas de clasificación o regresión. Una buena separación entre las clases permitirá una clasificación correcta. La siguiente figura ha sido extraída del documento [11].

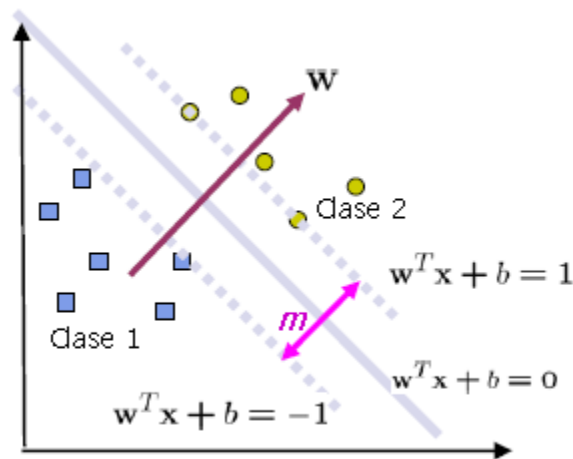


Figura 5. En esta figura se muestra el hiperplano óptimo para la clasificación de dos clases [11]

Como en la mayoría de los métodos de clasificación, los datos de entrada se ven como un vector  $n$ -dimensional (una lista de  $n$  números). La SVM busca un hiperplano que separe de forma óptima los puntos que pertenecen a una clase, de los que pertenecen a la otra. En ese concepto de "separación óptima" es donde reside la característica fundamental de las SVM: este tipo de algoritmos buscan el hiperplano que tenga la máxima distancia (margen) con los puntos que estén más cerca del mismo.

Por eso también a veces se conoce a las SVM como "clasificadores de margen máximo". De esta forma, los puntos del vector que son etiquetados con una categoría estarán a un lado del hiperplano y los casos que se encuentren en la otra categoría estarán al otro lado.

Los modelos basados en SVMs están estrechamente relacionados con las redes neuronales. Usando una función kernel, resultan un método de entrenamiento alternativo para clasificadores polinomiales, funciones de base radial y perceptrón multicapa.

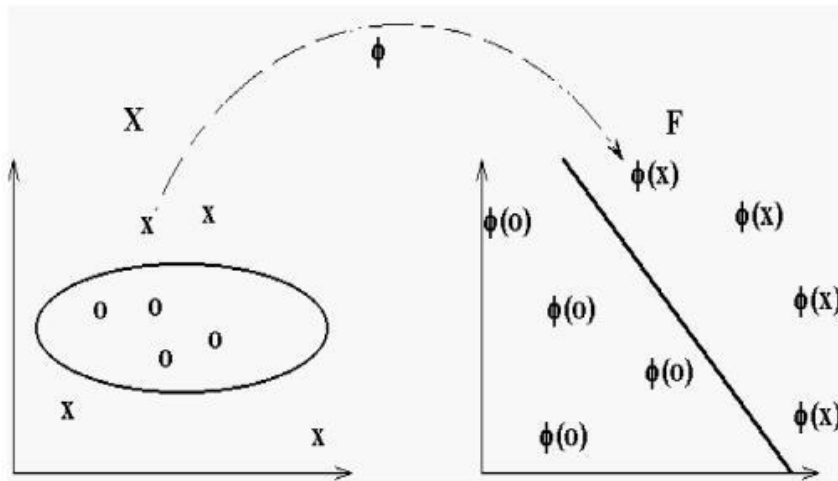
La manera más sencilla de realizar la separación de dos clases, es como se ha visto en la *Figura 5*, mediante una línea recta, un plano recto o un hiperplano n-dimensiona. Desgraciadamente, la mayoría de parámetros de la naturaleza no son lineales y tampoco suelen ser en dos dimensiones, por lo que un algoritmo SVM se suele enfrentar a curvas no lineales de separación, casos donde los conjuntos de datos no pueden ser completamente separados o clasificadores con más de dos categorías, como es el caso de este proyecto.

Debido a este problema, es necesario implementar una función kernel, proyectando la información a un espacio de características de mayor dimensión el cual aumenta la capacidad computacional de la máquina de aprendizaje lineal. Es decir, con una SVM mapeamos los puntos de entrada a un espacio de características de una dimensión mayor.

Los distintos tipos de funciones kernel y sus ecuaciones se muestran a continuación.

- **Polinomial-homogénea:**

$$K(x_i, x_j) = (x_i \cdot x_j)^n \tag{7}$$



*Figura 6. Función kernel polinomial (Wikipedia [8])*

- **Perceptrón:**

$$K(x_i, x_j) = \|x_i - x_j\| \quad (8)$$

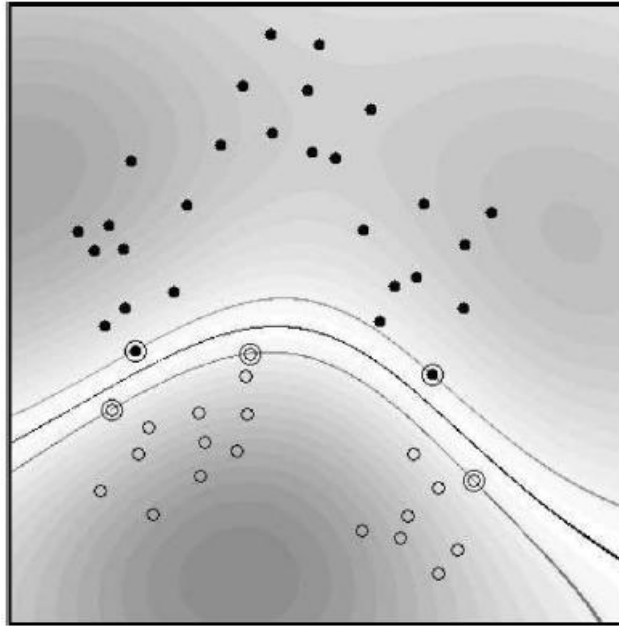


Figura 7. SVM con kernel perceptrón (Redes Neuronales) (Wikipedia [8])

- **Función de base radial Gaussiana:**

$$K(x_i, x_j) = e^{\frac{-(x_i - x_j)^2}{2 \cdot \sigma^2}} \quad (9)$$

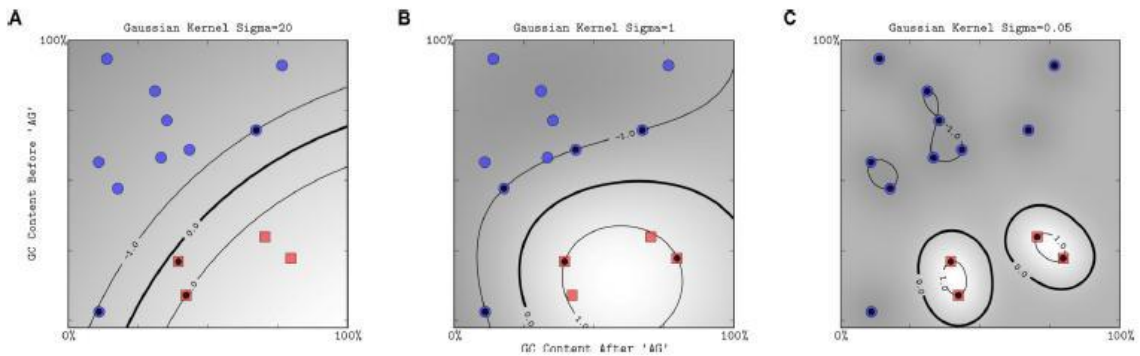


Figura 8. Efecto de la transformación con kernel Gaussiano en función de sigma (Wikipedia [8])

- **Sigmoidal:**

$$K(x_i, x_j) = \tanh(x_i \cdot x_j - \theta) \quad (10)$$

A pesar de todas las ventajas que presenta la SVM frente a las redes neuronales, también se pueden dar problemas. Uno de los más comunes es el error de *overfitting* o sobreentrenamiento. Este fenómeno se da cuando se intenta abarcar todos los datos del universo que se pretende estudiar, ya que si se hace esto, el modelo no puede ser generalizado para otros datos de entrada.

Con el fin de permitir cierta flexibilidad, las SVM manejan un parámetro  $C$  que controla la compensación entre los errores de entrenamiento y los márgenes rígidos, creando así un margen blando (soft margin) que permita varios errores pero al mismo tiempo los penalice.

Como conclusión se puede decir que SVM es más rápido y menos costoso de implementar que una red neuronal artificial ya que el entrenamiento es muy eficiente y es extremadamente robusto para generalización.

Por todos estos motivos se ha utilizado este tipo de algoritmo para la clasificación de géneros musicales en este trabajo.

En el siguiente apartado se hablará de cómo ha sido implementada la máquina en Matlab.

#### **4.4 IMPLEMENTACIÓN DE SVM EN MATLAB**

A la hora de usar SVM en Matlab se optó por descargar una librería con las funciones de la SVM ya programadas. Dentro de esta librería se encuentran las dos funciones que se han utilizado para el entrenamiento y el test de la red.

Las funciones que se han utilizado son *svmtrain* y *svmpredict* las cuales se van a explicar con más detalle a continuación.

- *svmtrain* es la función que permite entrenar la red neuronal. En la siguiente tabla se muestra el prototipo de la función y explica los argumentos de entrada y salida de dicha función.

<pre>SVMStruct = svmtrain(Training,Group) SVMStruct = svmtrain(Training,Group,Name,Value)</pre>	
<b>Argumentos de entrada</b>	
Training	Matriz de datos de entrenamiento, donde cada fila corresponde a una observación o replicar, y cada columna corresponde a una característica o variable.
Group	Cada elemento de group especifica el grupo de la fila correspondiente del entrenamiento.
<b>Argumentos de salida</b>	
SVMStruct	Estructura que contiene información acerca del entrenamiento de la SVM.

Tabla 4. Prototipo de la función de entrenamiento de la SVM

- *svmpredict* es la función que da la solución a unas características de entrada una vez que la red neuronal está entrenada. El prototipo de esta función se muestra en la tabla 5.

<pre>[predicted_label, accuracy, decision_values/prob_estimates] = svmpredict(testing_label_vector, testing_instance_matrix, model, ['libsvm_options']);</pre>	
<b>Argumentos de entrada</b>	
testing_label_vector	Vector de las etiquetas de test
testing_instance_matrix	Matriz que contiene las entradas de los datos con los que se va a realizar el test
model	Es la salida de la función <i>svmtrain</i>
<b>Argumentos de salida</b>	
predicted_label	Matriz que guarda los datos ya clasificados
accuracy	Probabilidad de acierto
decision_values/prob_estimates	Es una matriz que tiene los datos con los que se ha hecho la clasificación.

Tabla 5. Prototipo de la función de predicción de la SVM

En el siguiente capítulo se va a hablar de cómo se ha implementado la SVM en el clasificador y de los resultados que se han obtenido en primera instancia.



## 5. IMPLEMENTACIÓN DEL CLASIFICADOR Y PRIMEROS RESULTADOS.

En este nuevo capítulo se va a presentar como se ha implementado el clasificador y se va a explicar su funcionamiento de forma detallada. Después se mostrarán los primeros resultados que se han obtenido.

Como se ha explicado en el apartado anterior, el clasificador implementado es una SVM funcionando en modo clasificación.

### 5.1 CONSTRUCCIÓN DEL CLASIFICADOR

Lo primero que había que hacer es dividir la base de datos en “datos de entrenamiento” y “datos de test”. Para esto se ha elegido utilizar un 80% de las canciones de la base de datos para entrenamiento y el 20% restante, como test.

Además de esto, para aumentar el número de archivos con los que trabajar, se ha decidido hacer una validación cruzada. Esto se ha hecho utilizando la función de Matlab *crossvalind* [10]. Esta función devuelve un vector de índices en función del parámetro K. Como la idea sigue siendo utilizar un 80% para entrenamiento y un 20% para test, se ha decidido que el valor de K sea 5.

El esquema de cómo funciona un *K-fold crossvalidation* con  $K = 5$  se muestra en la figura 9, extraída del artículo de Zhang y Wu, mencionado en la bibliografía.

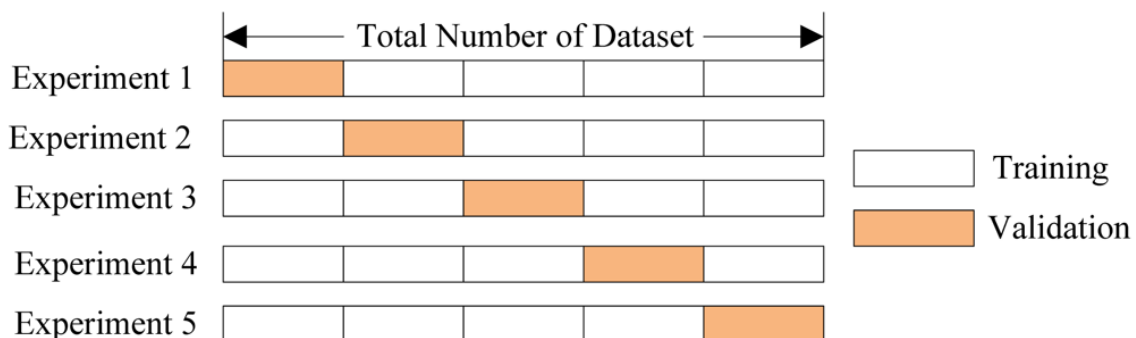


Figura 9. Validación cruzada para  $K = 5$ . Figura extraída de [7].

Una vez generados los índices, a cada canción de la base de datos se le asigna uno de forma aleatoria. Finalmente, con las funciones explicadas en el apartado anterior (*svmtrain* y *svmpredict*) se realiza el entrenamiento y la comprobación del funcionamiento de la máquina.

A continuación se van a presentar los primeros resultados que se han obtenido en la clasificación.

## 5.2 PRIMEROS RESULTADOS

Lo primero que introducimos al clasificador es la matriz de características sin normalizar. Los datos que se van a mostrar a continuación son el porcentaje de acierto para cada género musical, la probabilidad de acierto que tiene el clasificador en media contando todos los géneros musicales y la matriz de confusión completa. La tabla 6 recoge el porcentaje de acierto para cada género musical si no se normaliza la matriz de características.

<i>Género musical</i>	<i>% acierto</i>
<i>ORCHESTRA</i>	78.72%
<i>CHOIR</i>	0%
<i>LYRIC</i>	0%
<i>ELECTRONIC</i>	3.9%
<i>FOLK</i>	0%
<i>HIP HOP</i>	0%
<i>JAZZ</i>	0%
<i>BLUES</i>	2.4%
<i>LATIN</i>	0%
<i>POP</i>	5.6%
<i>WORLD</i>	20.96%
<i>DEATH METAL</i>	0%
<i>GOTHIC METAL</i>	0%
<i>METAL</i>	0%
<i>METALCORE</i>	0%
<i>POWER METAL</i>	0%
<i>FOLK ROCK</i>	0%
<i>PUNK ROCK</i>	0%
<i>ROCK</i>	0%
<i>ROCK &amp; ROLL</i>	0%
<i>FLAMENCO</i>	0%

Tabla 6. Porcentaje de acierto para cada género sin normalizar la matriz de características



En la tabla 7 se puede ver cuál es la probabilidad de acierto media, el MSE media y el coeficiente de correlación medio si no se normaliza la matriz de características.

Probabilidad de acierto media	11,91 %
MSE media	95,44125
Coficiente de Correlación Medio	0,009292

Tabla 7. Probabilidad de acierto media, coef de correlación medio y MSE media del clasificador al introducir la matriz de características sin normalizar.

En la tabla 8 se muestra la matriz de confusión entre todos los géneros de la base de datos si no se normaliza la matriz de características.

	ORCHESTRA	CHOIR	LYRIC	ELECTRONIC	FOLK	HIP HOP	JAZZ	BLUES	LATIN	POP	WORLD	DEATH METAL	GOTHIC METAL	METAL	METAL CORE	POWER METAL	FOLK ROCK	PUNK ROCK	ROCK	ROCK & ROLL	FLAMENCO	
ORCHESTRA	111	0	0	4	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0
CHOIR	27	0	0	1	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0
LYRIC	39	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0
ELECTRONIC	77	0	0	4	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0
FOLK	44	0	0	2	0	0	0	0	0	2	15	0	0	0	0	0	0	0	0	0	0	0
HIP HOP	41	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0
JAZZ	35	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0
BLUES	22	0	0	2	0	0	0	0	1	0	6	0	0	0	0	0	0	0	0	0	0	0
LATIN	35	0	0	2	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
POP	64	0	0	2	0	0	0	0	0	5	19	0	0	0	0	0	0	0	0	0	0	0
WORLD	97	0	0	1	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0
DEATH METAL	35	0	0	1	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
GOTHIC METAL	35	0	0	1	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0
METAL	41	0	0	2	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0
METALCORE	39	0	0	1	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0
POWER METAL	41	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0
FOLK ROCK	21	0	0	1	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0
PUNK ROCK	33	0	0	0	0	0	0	0	0	1	12	0	0	0	0	0	0	0	0	0	0	0
ROCK	29	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0
ROCK & ROLL	35	0	0	2	0	0	0	1	0	0	7	0	0	0	0	0	0	0	0	0	0	0
FLAMENCO	36	0	0	1	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0

Tabla 8. Matriz de confusión de los géneros musicales sin normalizar la matriz de características

Como se puede observar, los resultados son bastante malos. Esto es debido a que la SVM requiere que se le introduzcan datos normalizados entre -1 y 1. Además, se puede comprobar que la mayoría de la clasificación se centra en los géneros de ORCHESTRA, ELECTRONIC, POP y WORLD. Esto se debe a que el número de archivos es superior al número de archivos que hay en el resto de las clases. En los futuros experimentos que se van a realizar, se solventarán esos problemas para mejorar los resultados.

## 6. MEJORAS EN EL CLASIFICADOR

En este capítulo se van a presentar todas las mejoras que se han ido realizando progresivamente para mejorar el porcentaje de acierto del clasificador para todos los géneros musicales.

Todas estas mejoras se irán comentando de forma individual y a medida que se van implementando nuevas mejoras, éstas se irán acumulando a las anteriores.

### 6.1 NORMALIZACIÓN DE LA MATRIZ DE CARACTERÍSTICAS.

La primera mejora que se ha propuesto es la de normalizar la matriz de las características obtenidas anteriormente de la base de datos de canciones. La normalización que se ha aplicado es la de dividir por el valor máximo, ya que para trabajar con una SVM es conveniente que los datos estén acotados entre un mínimo de -1 y un máximo de 1.

Una vez realizado este cambio, en la tabla 9 se muestra el porcentaje de acierto para cada género musical.

<i>Género musical</i>	<i>% acierto</i>
<i>ORCHESTRA</i>	75,89%
<i>CHOIR</i>	0,00%
<i>LYRIC</i>	0,00%
<i>ELECTRONIC</i>	8,91%
<i>FOLK</i>	0,00%
<i>HIP HOP</i>	0,00%
<i>JAZZ</i>	2,13%
<i>BLUES</i>	0,00%
<i>LATIN</i>	2,38%
<i>POP</i>	16,67%
<i>WORLD</i>	37,90%
<i>DEATH METAL</i>	2,44%
<i>GOTHIC METAL</i>	1,96%
<i>METAL</i>	8,62%
<i>METALCORE</i>	0,00%
<i>POWER METAL</i>	0,00%
<i>FOLK ROCK</i>	3,57%
<i>PUNK ROCK</i>	0,00%
<i>ROCK</i>	0,00%
<i>ROCK &amp; ROLL</i>	2,22%
<i>FLAMENCO</i>	0,00%

Tabla 9. Porcentaje de acierto para cada género normalizando la matriz de características

A continuación, en la tabla 10, se muestra es la probabilidad de acierto media, el MSE media y el coeficiente de correlación medio si se normaliza la matriz de características.

Probabilidad de Acierto Media	14,91 %
MSE media	88,40079
Coficiente de Correlación Medio	0,01984

*Tabla 10. Probabilidad de acierto media, coef de correlación medio y MSE media del clasificador al introducir la matriz de características normalizada.*

Finalmente, en la tabla 11, se muestra la matriz de confusión entre todos los géneros de la base de datos si se normaliza la matriz de características.

	ORCHESTRA	CHOIR	LYRIC	ELECTRONIC	FOLK	HIP HOP	JAZZ	BLUES	LATIN	POP	WORLD	DEATH METAL	GOTHIC METAL	METAL	METAL CORE	POWER METAL	FOLK ROCK	PUNK ROCK	ROCK	ROCK & ROLL	FLAMENCO	
ORCHESTRA	107	0	0	2	0	0	0	0	0	0	32	0	0	0	0	0	0	0	0	0	0	0
CHOIR	20	0	0	1	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0
LYRIC	32	0	0	3	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0
ELECTRONIC	64	0	0	9	0	0	0	0	0	3	25	0	0	0	0	0	0	0	0	0	0	0
FOLK	43	0	0	1	0	0	0	0	0	1	18	0	0	0	0	0	0	0	0	0	0	0
HIP HOP	30	0	0	5	0	0	0	0	0	0	14	0	0	1	0	0	0	0	0	0	0	0
JAZZ	27	0	0	2	0	0	1	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0
BLUES	24	0	0	1	0	0	0	0	1	1	4	0	0	1	0	0	0	0	0	0	0	0
LATIN	23	0	0	0	0	0	0	0	1	1	17	0	0	0	0	0	0	0	0	0	0	0
POP	51	0	0	3	0	0	0	0	0	15	21	0	0	0	0	0	0	0	0	0	0	0
WORLD	75	0	0	1	0	0	0	0	0	1	47	0	0	0	0	0	0	0	0	0	0	0
DEATH METAL	20	0	0	3	0	0	0	0	0	1	16	1	0	0	0	0	0	0	0	0	0	0
GOTHIC METAL	24	0	0	6	0	0	0	0	0	2	18	0	1	0	0	0	0	0	0	0	0	0
METAL	31	0	0	4	0	0	0	0	0	1	17	0	0	5	0	0	0	0	0	0	0	0
METALCORE	35	0	0	2	0	0	0	0	0	1	19	0	0	0	0	0	0	0	0	0	0	0
POWER METAL	31	0	0	2	0	0	0	0	0	1	18	0	0	0	0	0	0	0	0	0	0	0
FOLK ROCK	16	0	0	1	0	0	0	0	0	2	8	0	0	0	0	0	1	0	0	0	0	0
PUNK ROCK	26	0	0	1	0	0	0	0	0	2	17	0	0	0	0	0	0	0	0	0	0	0
ROCK	21	0	0	2	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0
ROCK & ROLL	23	0	0	1	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	1	0
FLAMENCO	28	0	0	1	0	0	0	0	0	2	16	0	0	0	0	0	0	0	0	0	0	0

*Tabla 11. Matriz de confusión de los géneros musicales normalizando la matriz de características*

Se ha comprobado que normalizando la matriz de características, los resultados mejoran ligeramente, aunque siguen sin ser buenos. Esto se debe a que el número de elementos de cada género musical es distinto. Como se observa en la matriz de confusión, la clasificación de los elementos se sigue acumulando en los géneros que más canciones tienen en la base de datos.

Por ello, la siguiente mejora que se propone, es la de equiparar el número de canciones que debe haber en todos los géneros, sin que haya una diferencia de más de 20 elementos entre cada uno de ellos.

## 6.2 IGUALACIÓN DEL NÚMERO DE ELEMENTOS EN LA BASE DE DATOS.

Como se ha comentado en el apartado anterior, se va a intentar equiparar el número de elementos de los distintos géneros musicales. Esto se ha hecho reduciendo el número de canciones de los géneros que tenían más de 70 temas, a 50 solamente. Así se consigue una base de datos más uniforme y se puede augurar que el porcentaje de acierto va a aumentar.

A continuación se muestran los resultados obtenidos después de ejecutar el clasificador, añadiendo estas mejoras a la matriz de características.

En la tabla 12 se muestra el porcentaje de acierto para cada género musical con la matriz de características normalizada y el número de archivos de cada género compensado.

<i>Género musical</i>	<i>% acierto</i>	<i>Género musical</i>	<i>% acierto</i>
<i>ORCHESTRA</i>	12,00%	<i>METAL</i>	20,69%
<i>CHOIR</i>	34,38%	<i>METALCORE</i>	8,77%
<i>LYRIC</i>	40,00%	<i>POWER METAL</i>	19,23%
<i>ELECTRONIC</i>	18,00%	<i>FOLK ROCK</i>	17,86%
<i>FOLK</i>	15,87%	<i>PUNK ROCK</i>	26,09%
<i>HIP HOP</i>	12,00%	<i>ROCK</i>	39,47%
<i>JAZZ</i>	25,53%	<i>ROCK &amp; ROLL</i>	22,22%
<i>BLUES</i>	35,48%	<i>FLAMENCO</i>	10,64%
<i>LATIN</i>	23,81%		
<i>POP</i>	10,00%		
<i>WORLD</i>	8,00%		
<i>DEATH METAL</i>	31,71%		
<i>GOTHIC METAL</i>	39,22%		

Tabla 12. Porcentaje de acierto para cada género normalizando la matriz de características y equiparando el número de elementos de cada género.

A continuación, en la tabla 13, se muestra es la probabilidad de acierto media, el MSE media y el coeficiente de correlación medio si se normaliza la matriz de características y se compensan los distintos géneros musicales.

Probabilidad de Acierto Media	21,57 %
MSE media	58,29726
Coeficiente de Correlación Medio	0,040193

*Tabla 13. Probabilidad de acierto media, coef de correlación medio y MSE media del clasificador al introducir la matriz de características normalizada y la base de datos compensada.*

Por último, en la tabla 14, se muestra la matriz de confusión entre todos los géneros de la base de datos si se normaliza la matriz de características y se equipara el número de canciones de cada género musical.

	ORCHESTRA	CHOIR	LYRIC	ELECTRONIC	FOLK	HIP HOP	JAZZ	BLUES	LATIN	POP	WORLD	DEATH METAL	GOTHIC METAL	METAL	METAL CORE	POWER METAL	FOLK ROCK	PUNK ROCK	ROCK	ROCK & ROLL	FLAMENCO
ORCHESTRA	6	0	6	7	2	0	3	0	2	5	1	3	2	2	0	2	0	2	3	4	0
CHOIR	0	11	2	3	2	0	2	2	1	1	0	1	0	0	0	1	0	1	3	1	1
LYRIC	0	0	20	4	2	0	1	1	1	3	1	2	1	4	0	0	0	2	4	4	0
ELECTRONIC	2	0	0	9	2	2	3	0	3	8	2	5	0	3	3	1	0	1	2	3	1
FOLK	1	0	3	7	10	2	3	2	0	7	5	1	2	4	5	2	2	2	3	2	0
HIP HOP	1	0	2	6	4	6	3	0	4	6	0	1	1	2	3	0	0	4	4	0	3
JAZZ	0	0	2	5	4	0	12	0	3	2	1	2	2	2	3	1	0	2	4	2	0
BLUES	2	1	0	1	1	0	0	11	1	3	1	0	0	0	0	1	0	5	0	1	3
LATIN	1	0	2	3	2	2	4	1	10	5	3	1	0	1	1	1	0	0	2	1	2
POP	2	1	2	4	3	0	3	3	1	5	1	1	5	2	0	2	0	6	3	3	3
WORLD	0	1	4	4	4	0	2	0	1	6	4	1	3	2	2	1	1	9	2	1	2
DEATH METAL	0	0	2	3	2	1	1	0	0	5	2	13	0	0	1	2	0	4	3	2	0
GOTHIC METAL	0	0	0	2	1	0	5	2	3	2	2	2	20	3	0	0	0	3	2	2	2
METAL	3	0	1	6	3	0	2	0	2	4	2	0	3	12	3	0	1	6	6	1	3
METALCORE	1	0	4	6	4	2	1	0	2	3	2	3	3	6	5	1	0	1	6	3	4
POWER METAL	2	0	2	5	2	0	1	1	1	7	2	2	0	3	3	10	0	6	1	2	2
FOLK ROCK	0	1	2	2	2	0	3	0	2	2	1	2	1	0	0	0	5	2	1	2	0
PUNK ROCK	1	0	1	5	3	2	3	1	4	3	0	1	2	3	0	0	1	12	1	3	0
ROCK	2	0	1	1	0	1	2	0	4	2	1	0	1	2	0	1	0	4	15	1	0
ROCK & ROLL	2	0	3	4	3	1	3	0	2	7	1	2	0	0	1	1	0	0	3	10	2
FLAMENCO	1	0	4	6	2	0	4	1	2	4	4	2	2	2	1	1	1	2	2	1	5

*Tabla 14. Matriz de confusión de los géneros musicales normalizando la matriz de características e igualando el número de canciones de cada género.*

Introduciendo estas mejoras, se puede comprobar con la matriz de confusión que ahora la clasificación está distribuida entre todos los géneros y el porcentaje de acierto aumenta notablemente para la mayoría de los géneros musicales. Sigue siendo un porcentaje bajo pero es aceptable ya que la base de datos cuenta con muchos géneros distintos.

### 6.3 AGRUPACIÓN DE GÉNEROS MUSICALES.

Para seguir intentando mejorar el porcentaje de acierto del clasificador se ha optado por combinar géneros musicales que, para el oído humano, puedan resultar similares o tengan alguna característica en común.

De esta forma reducimos el número de géneros musicales de la base de datos de 21 a 11.

Las combinaciones que se han hecho han sido las que se muestran en la tabla 15.

<i>Géneros finales</i>	<i>Géneros originales</i>
<b>CLASICAL</b>	CHOIR
	LYRIC
	ORCHESTRA
<b>JAZZ/BLUES</b>	JAZZ
	BLUES
<b>METAL</b>	METAL
	METALCORE
	DEATH METAL
	POWER METAL
	GOTHIC METAL
<b>ROCK</b>	ROCK
	PUNK ROCK
	FOLK ROCK
	ROCK & ROLL

Tabla 15. Combinación de géneros musicales similares.

Después de reducir el número de géneros musicales, hay que volver a asignar las etiquetas. Esta asignación se muestra a continuación, en la tabla 16.

<i>Género musical</i>	<i>Etiqueta</i>
CLASICAL	1
ELECTRONIC	2
FOLK	3
HIP HOP	4
JAZZ/BLUES	5
LATIN	6
POP	7
WORLD	8
METAL	9
ROCK	10
FLAMENCO	11

*Tabla 16. Nuevos géneros musicales con sus respectivas etiquetas.*

Con estos cambios realizados, y manteniendo la normalización de la matriz de características y la compensación del número de canciones entre los distintos géneros musicales, se van presentar los resultados obtenidos.

Primero, en la tabla 17, se muestra el porcentaje de acierto del clasificador para cada género musical.

<i>Género musical</i>	<i>% acierto</i>
CLASICAL	17,14%
ELECTRONIC	11,43%
FOLK	14,29%
HIP HOP	32,00%
JAZZ/BLUES	31,43%
LATIN	14,29%
POP	21,43%
WORLD	14,29%
METAL	31,43%
ROCK	7,14%
FLAMENCO	14,89%

*Tabla 17. Porcentaje de acierto para cada género normalizando la matriz de características y equiparando el número de elementos de cada género después de combinar géneros.*



A continuación, en la tabla 18, se muestra es la probabilidad de acierto media, el MSE media y el coeficiente de correlación medio después de combinar los géneros musicales, con la matriz de características normalizada y manteniendo la compensación entre géneros.

Probabilidad de Acierto Media	19,08 %
MSE media	16,44845
Coficiente de Correlación Medio	0,009786

*Tabla 18. Probabilidad de acierto media, coef de correlación medio y MSE media del clasificador al introducir la matriz de características normalizada y la base de datos compensada después de combinar géneros musicales.*

Por último, en la tabla 19, se muestra la matriz de confusión para los nuevos géneros de la base de datos si se normaliza la matriz de características y se equipara el número de canciones de cada género musical.

	CLASICAL	ELECTRONIC	FOLK	HIP HOP	BLUES/JAZZ	LATIN	POP	WORLD	METAL	ROCK	FLAMENCO
CLASICAL	12	3	2	4	16	0	12	2	14	4	1
ELECTRONIC	5	8	1	4	16	2	8	3	16	4	3
FOLK	4	9	9	2	11	2	9	3	9	4	1
HIP HOP	5	2	0	16	9	1	6	4	4	2	1
BLUES/JAZZ	7	3	3	5	22	2	10	3	9	4	2
LATIN	4	1	0	7	11	6	2	3	8	0	0
POP	6	2	2	3	12	3	15	6	17	2	2
WORLD	6	6	1	6	13	2	8	10	14	4	0
METAL	7	3	4	3	11	2	9	5	22	2	2
ROCK	4	3	6	2	19	5	9	3	12	5	2
FLAMENCO	3	5	1	5	10	0	1	3	10	2	7

*Tabla 19. Matriz de confusión de los géneros musicales normalizando la matriz de características e igualando el número de canciones de cada género.*

En este caso no se aprecia una mejora notable en el porcentaje de acierto del clasificador, de hecho, disminuye ligeramente. Aun así, el porcentaje es aceptable por el mismo motivo que en el caso anterior.

Con el clasificador funcionando de forma correcta, queda preguntarse qué ocurriría dentro de la agrupación, es decir, se puede comprobar que tal funciona el clasificador para géneros similares o con características en común.

En el siguiente apartado se mostrarán los resultados que se obtienen al clasificar canciones de géneros parecidos.

## 6.4 CLASIFICACIÓN DE GÉNEROS SIMILARES.

Como en el apartado anterior se hizo una agrupación de géneros similares para comprobar si el porcentaje de acierto del clasificador mejoraba, ahora se comprobará qué tal funciona el clasificador dentro de esa agrupación. Se ejecutará el clasificador para los géneros CLASICAL, JAZZ/BLUES, METAL y ROCK respectivamente. Para este apartado se utilizará el número original de archivos de audio de cada uno de los géneros musicales a menos que éste sea muy superior al del resto. Si esto último ocurre, se equiparán las canciones de todos los géneros.

El primer género en ser clasificado es el perteneciente a música clásica. Éste se subdivide en tres subgéneros (orchestral, lyric y choir) a los que se les asignarán las etiquetas 1, 2 y 3 respectivamente.

Después de ejecutar el clasificador para estas características, manteniendo eso sí, la normalización de la matriz de características en todo momento, se presentan los resultados obtenidos.

- **CLASICAL**

Para este caso se ha tenido que reducir de 141 a 50 las canciones del subgénero *ORCHESTRAL*.

En la tabla 20 se va a mostrar el porcentaje de acierto para cada subgénero dentro del género *CLASICAL*.

<i>Género musical</i>	<i>% acierto</i>
<i>ORCHESTRAL</i>	56,00%
<i>LYRIC</i>	50,00%
<i>CHOIR</i>	43,75%

Tabla 20. Porcentaje de acierto para los subgéneros del género *CLASICAL*

A continuación, en la tabla 21, se muestra es la probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género *CLASICAL*.

Probabilidad de Acierto Media	50,77 %
MSE media	0,995726
Coeficiente de Correlación Medio	0,097034

Tabla 21. Probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género *CLASICAL*.

Por último se muestra en la tabla 22 la matriz de confusión para los subgéneros contenidos en el género *CLASICAL*.

	ORCHESTRA	LYRIC	CHOIR
ORCHESTRA	28	8	14
LYRIC	11	25	14
CHOIR	8	10	14

Tabla 22. Matriz de confusión para los subgéneros del género *CLASICAL*.

- **JAZZ/BLUES**

En la tabla 23 se va a mostrar el porcentaje de acierto para cada subgénero dentro del género *JAZZ/BLUES*.

Género musical	% acierto
JAZZ	76,60%
BLUES	80,65%

Tabla 23. Porcentaje de acierto para los subgéneros del género *JAZZ/BLUES*.

A continuación, en la tabla 24, se muestra es la probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género *JAZZ/BLUES*.

Probabilidad de Acierto Media	50,77 %
MSE media	0,995726
Coficiente de Correlación Medio	0,097034

*Tabla 24. Probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género JAZZ/BLUES.*

Finalmente, se muestra en la tabla 25 la matriz de confusión para los subgéneros contenidos en el género *JAZZ/BLUES*.

	JAZZ	BLUES
JAZZ	36	11
BLUES	6	25

*Tabla 25. Matriz de confusión para los subgéneros del género JAZZ/BLUES.*

- **METAL**

Como en los casos anteriores, primero se muestra el porcentaje de acierto para cada subgénero contenido en el género metal en la tabla 26.

<i>Género musical</i>	<i>% acierto</i>
<i>GOTHIC METAL</i>	25,49%
<i>DEATH METAL</i>	68,29%
<i>METAL</i>	22,41%
<i>METALCORE</i>	21,05%
<i>POWER METAL</i>	40,38%

*Tabla 26. Porcentaje de acierto para los subgéneros del género METAL.*

A continuación, en la tabla 27, se muestra es la probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género *METAL*.

Probabilidad de Acierto Media	33,63 %
MSE media	3,275792
Coficiente de Correlación Medio	0,057399

Tabla 27. Probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género *METAL*.

Finalmente, se muestra en la tabla 28 la matriz de confusión para los subgéneros contenidos en el género *METAL*.

	GOTHIC METAL	DEATH METAL	METAL	METALCORE	POWER METAL
GOTHIC METAL	13	21	3	3	11
DEATH METAL	1	28	0	5	7
METAL	9	18	13	6	12
METALCORE	10	16	9	12	10
POWER METAL	2	23	1	5	21

Tabla 28. Matriz de confusión para los subgéneros del género *METAL*.

- **ROCK**

En la tabla 29 se va a mostrar el porcentaje de acierto para cada subgénero dentro del género *ROCK*.

Género musical	% acierto
FOLK ROCK	64,29%
PUNK ROCK	36,96%
ROCK	52,63%
ROCK & ROLL	62,22%

Tabla 29. Porcentaje de acierto para los subgéneros del género *ROCK*

A continuación, en la tabla 30, se muestra es la probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género *ROCK*.

Probabilidad de Acierto Media	52,82 %
MSE media	1,337097
Coeficiente de Correlación Medio	0,248986

Tabla 30. Probabilidad de acierto media, el MSE media y el coeficiente de correlación medio de la clasificación de los subgéneros del género *ROCK*.

Por último, se muestra en la tabla 31 la matriz de confusión para los subgéneros contenidos en el género *ROCK*.

	FOLK ROCK	PUNK ROCK	ROCK	ROCK & ROLL
FOLK ROCK	18	3	1	6
PUNK ROCK	11	17	9	9
ROCK	3	9	20	6
ROCK & ROLL	2	11	4	28

Tabla 31. Matriz de confusión para los subgéneros del género *ROCK*.

Terminado este experimento, se puede comprobar que el porcentaje de acierto aumenta considerablemente a la hora de clasificar subgéneros con características en común. Para los géneros de *CLASICAL* y *JAZZ/BLUES* el acierto es muy alto, por lo que se demuestra que el clasificador funciona correctamente.

Los resultados más bajos obtenidos para los géneros de *METAL* y *ROCK*, se debe a que en estos géneros, las características extraídas son muy similares, lo que dificulta mucho la diferenciación entre los distintos subgéneros. Este caso es particular, ya que para el oído humano todos los subgéneros dentro de estos géneros son muy difíciles de diferenciar. Hay numerosas bases de datos por internet que difieren mucho a la hora de clasificar este tipo de géneros.

## 7. CONCLUSIONES

Con este trabajo fin de grado se ha intentado solventar el problema que existía a la hora de clasificar los distintos géneros musicales, ya que se venía haciendo de forma manual y era un proceso lento y costoso. Para conseguirlo se han seguido estos pasos:

- Primero se buscó una base de datos ya etiquetada que sirviese como referencia. A esta base de datos se le añadieron más géneros musicales, obteniendo así una base de datos más completa y variada.
- Sobre esta base de datos se aplicaron técnicas de procesamiento de señal para obtener distintas características tímbricas de las canciones que servirán para ser analizadas posteriormente y serán los patrones que se utilizarán para la clasificación.
- Con el conjunto de características ya extraídas, se implementa un clasificador mediante una SVM, que ha sido la encargada de clasificar los distintos géneros musicales.
- Dado que los resultados del primer experimento que se realizó fueron bastante malos, se proponen mejoras y soluciones que mejoraron notablemente el porcentaje de acierto del clasificador.
- Con todo esto, finalmente se consiguió obtener unos resultados bastante aceptables dado el alto número de géneros musicales de los que estaba compuesta la base de datos. También se ha comprobado que el resultado mejora mucho si se reduce el número de géneros musicales o para clasificar los subgéneros que hay dentro de un mismo género.

Durante el desarrollo del trabajo se han tenido que ir haciendo modificaciones con el fin de buscar un mejor funcionamiento del clasificador. Al principio la base de datos estaba descompensada, lo que hacía que el clasificador no funcionase correctamente, concentrando la clasificación de las canciones en los géneros que más archivos de audio tenían. Por eso se optó por compensar el número de canciones de todos los géneros, reduciendo el número de archivos de estos géneros.

Otro de los cambios que se hicieron, fue el de normalizar la matriz de características de todas las canciones. Esto se hizo porque para trabajar con una SVM es recomendable que los datos que se introducen estén normalizados entre los valores de -1 y 1. Con esto se mejoraron notablemente los resultados, además de disminuir el tiempo de ejecución.

Con todo esto, los resultados mejoraron considerablemente, teniendo en cuenta los del primer experimento pero aun así el porcentaje de acierto obtenido es bajo, aunque aceptable. Por eso se optó por comprobar cómo funciona el clasificador a la hora de clasificar subgéneros dentro de otro. El resultado de este experimento fue bastante satisfactorio, obteniendo porcentajes de acierto mucho más altos.





## 8. LÍNEAS FUTURAS

- Diseñar una interfaz gráfica: de esta forma podría desarrollarse para distintas plataformas y facilitar su funcionamiento. Otra de las razones de realizar una interfaz gráfica, sería la de poder hacérselo llegar a las personas y convertirlo en una herramienta de uso cotidiano.
- Utilización de otros algoritmos para la clasificación: la existencia de tantos algoritmos con los que se puede entrenar una red neuronal nos da la opción de poder elegir otros algoritmos para desempeñar esta función. Uno de ellos, que da muy buenos resultados, podría ser el algoritmo ELM (Extreme Learning Machine).
- Búsqueda de nuevas características: las características que se han utilizado para la clasificación se basan en el timbre. Existen distintas formas de clasificar, basándose en el ritmo o en la melodía.
- Diseñar una base de datos más compensada que mejore los resultados.
- Otra línea futura de estudio basándose en este trabajo, sería la investigación sobre la identificación vocal o la diferenciación de instrumentos musicales en una canción.



# BIBLIOGRAFÍA

- [1] Bonomo Laynez, D. (2012). Extracción de Características. En D. Bonomo Laynez, *Sistemas de verificación automática de locutor*. Universidad de Sevilla.
- [2] L. R. Rabiner and B. H. Juang. (2003). *Fundamentals of Speech Recognition (2003)*.
- [3] *Mathworks*. Obtenido de <http://www.mathworks.com/matlabcentral/fileexchange/32849-htk-mfcc-matlab/content/mfcc/mfcc.m>
- [4] MIREX. (2005). *Music Information Retrieval Evaluation eXchange*. Obtenido de [http://www.music-ir.org/mirex/wiki/2005:Audio\\_Genre](http://www.music-ir.org/mirex/wiki/2005:Audio_Genre)
- [5] Music Technology Group (2004). *Universitat Pompeu Fabra*. Obtenido de [http://www.iaa.upf.edu/mtg/ismir2004/contest/Training\\_Tracks1.tar.tgz](http://www.iaa.upf.edu/mtg/ismir2004/contest/Training_Tracks1.tar.tgz)  
[http://www.iaa.upf.edu/mtg/ismir2004/contest/Training\\_Tracks2.tar.tgz](http://www.iaa.upf.edu/mtg/ismir2004/contest/Training_Tracks2.tar.tgz)
- [6] Tzanetakis, G., & Cook, P. (2002). *Automatic Musical Genre Classification*. *IEEE transactions on speech and audio processing*, Vol.10, N°5.
- [7] Zhang Y., & Wu, L. (2012). *Classification of Fruits Using Computer Vision and a Multiclass Support Vector Machine*. School of Information Science and Engineering, Southeast University, Nanjing 210096, China.
- [8] Máquinas de Vectores de Soporte. *Wikipedia*. Obtenido de [https://es.wikipedia.org/wiki/M%C3%A1quinas\\_de\\_vectores\\_de\\_soporte](https://es.wikipedia.org/wiki/M%C3%A1quinas_de_vectores_de_soporte)
- [9] *Redes de Neuronas Artificiales*. Obtenido de <http://www.lab.inf.uc3m.es/~a0080630/redes-de-neuronas/index.html>
- [10] *Mathworks*. Obtenido de <http://es.mathworks.com/help/bioinfo/ref/crossvalind.html>
- [11] Betancourt, G.A., (2005). *Las Máquinas de Soporte Vectorial (SVMs)*. Universidad Tecnológica de Pereira.
- [12] Guaus í Termens, E., (2009). *Audio content processing for automatic music genre classification, descriptors, databases and classifiers*. Universitat Pompeu Fabra.
- [13] Fernández, F., y Chávez, F. *Clasificación de género musical mediante sistemas basados en reglas difusas*. Universidad de Extremadura.
- [14] Red Neuronal Artificial. *Wikipedia*. Obtenido de [https://es.wikipedia.org/wiki/Red\\_neuronal\\_artificial](https://es.wikipedia.org/wiki/Red_neuronal_artificial)