UNIVERSIDAD DE EXTREMADURA

Escuela Politécnica

Grado en Ingeniería de Sonido e Imagen en Telecomunicación

Trabajo Fin de Grado

DETECCIÓN DE LA TONALIDAD EN AUDIO MUSICAL

Paula Hernández Salado

Noviembre, 2017

UNIVERSIDAD DE EXTREMADURA

Escuela Politécnica

Grado en Ingeniería de Sonido e Imagen en Telecomunicación

Trabajo Fin de Grado

DETECCIÓN DE LA TONALIDAD EN AUDIO MUSICAL

Autor: Paula Hernández Salado

Tutor: Alberto Gómez Mancha

ÍNDICE GENERAL DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

RESUMEN

El objetivo primordial de este trabajo es desarrollar software que permita analizar una pista digital de música y estime la tonalidad de la obra musical.

En la música occidental, el sistema tonal es fundamental. La tonalidad de cualquier obra musical viene determinada por las notas musicales que le dan forma. Hay que destacar que la tonalidad influye en el carácter de la pieza, por lo que se trata de una herramienta de enorme utilidad para, por ejemplo, la clasificación de música por las emociones que sugiere.

El estudio de la tonalidad de un archivo de audio es un tema altamente complejo, atacado desde el punto de vista de la física del sonido, ya que depende de muchos factores melódicos, armónicos, etc.

Por ello se presentan los términos músico-teóricos básicos que sustentan el Trabajo Fin de Grado y las herramientas básicas para el análisis y procesamiento de señales musicales, aplicadas al problema que supone la detección de la tonalidad de una obra musical.

Se han desarrollado una serie de algoritmos que, a partir de la información extraída del procesamiento digital de la señal digital, detectan la tonalidad de la obra. Se describen los experimentos y las pruebas realizadas para determinar la precisión de los prototipos desarrollados.

Finalmente se muestran y evalúan los resultados obtenidos al aplicar los distintos prototipos software implementados, las conclusiones destacadas, las contribuciones, limitaciones del proyecto y los trabajos futuros.

1. INTRODUCCIÓN

El objetivo de este Trabajo fin de Grado es investigar y desarrollar algoritmos que detecten la tonalidad de una obra musical analizando directamente un archivo de sonido.

El análisis y extracción de características musicales a partir del procesamiento digital del sonido es un campo muy interesante que presenta múltiples aplicaciones. Se pueden implementar detectores del estilo musical, ritmo, afinación e incluso de sensaciones provocadas en el ser humano. Además toda esa información puede complementarse entre sí para el desarrollo de transcriptores musicales, buscadores de canciones, etc.

El congreso MIREX (*Music Information Retrieval Evaluation eXchange*), organizado por la universidad de Illinois en Urbana-Champaign y celebrado anualmente desde 2005, se dedica a la evaluación de recuperación de muchos tipos de información musical. Se caracteriza por contar con una comprometida comunidad investigadora a nivel internacional; prueba de ello es la existencia de la ISMIR (*International Society of Music Information Retrieval*).

Esta disciplina se ocupa de múltiples tareas como la clasificación de audio por género, el seguimiento del ritmo o la estimación de acordes en el audio y la estimación de la tonalidad entre otros. MIREX utiliza la plataforma wiki para dejar constancia de los avances realizados en cada convocatoria anual y donde a su vez se proponen los nuevos trabajos para la siguiente (Stephen Dwonie, 2017).

En la música popular y clásica occidental, la tonalidad de una obra musical viene determinada por la organización jerárquica de las notas que la forman. Así, dependiendo de la tonalidad, aparecerán ciertas notas e intervalos más frecuentemente que otros. La tonalidad influye en el carácter de la pieza. Por ejemplo, muchas piezas tristes están en una tonalidad menor.

La detección de la tonalidad en una obra musical es un problema complejo porque depende de muchos factores melódicos y armónicos y puede ir cambiando a lo largo del tiempo.

En algunas situaciones es muy importante conocer la tonalidad de una pieza musical. Por ejemplo, un *disk-jockey* debe mezclar canciones con tonalidades similares para que las transiciones sean suaves y agradables al oído.

En este Trabajo Fin de Grado se han aplicado los conocimientos sobre el tratamiento de señales digitales y el análisis acústico de la señal de audio al problema de la detección de la tonalidad de una obra musical.

Tras un estudio de las bases teóricas del procesamiento de señal y de los conceptos musicales relacionados con la tonalidad, se han estudiado y probado diversas alternativas para la obtención, a partir de la señal de audio, de características útiles en la detección de la tonalidad (principalmente, las intensidades de las frecuencias a lo largo del tiempo).

A continuación se han desarrollado varios prototipos capaces de reconocer tonalidades en muestras de audio musicales. Los resultados de cada prototipo han servido para modificar el algoritmo de detección y probar otras alternativas con el fin de mejorar el porcentaje de acierto en el reconocimiento de la tonalidad.

También se han programado los procedimientos para analizar un gran número de audios con cada uno de los prototipos, obteniendo los resultados necesarios para poder compararlos.

En el siguiente apartado de la memoria de este Trabajo Fin de Grado se detallan los objetivos de este trabajo.

En el capítulo 3 se presentan la terminología y los fundamentos básicos de la música occidental tonal para explicar el concepto de tonalidad. También se exponen los principios del análisis de sonidos y su tratamiento digital.

En el cuarto capítulo se describe la metodología llevada a cabo en el desarrollo de dicha investigación.

En el quinto capítulo se explican cada uno de los prototipos desarrollados, mostrando las pruebas ejecutadas así como las mejoras efectuadas. Se muestran también los pasos a seguir en la implementación de las funciones en Matlab para la extracción de las características explicadas en el capítulo de antecedentes.

Para terminar, en el capítulo 6, se exponen y evalúan los resultados obtenidos aplicando los prototipos software implementados. Se procederá a realizar una comparación de los resultados con otro software existente que permite la extracción de la tonalidad de un archivo de audio. Por último, ligado al anterior, se extraerán las conclusiones acerca del funcionamiento del software desarrollado. Además, se harán propuestas de mejora para futuras ampliaciones de este proyecto y para futuras líneas de investigación.

2. OBJETIVOS

Los objetivos principales del presente proyecto se detallan a continuación:

I. Analizar las técnicas de procesado más comunes relativas a la extracción de información musical.
II. Estudiar los parámetros que intervienen en la extracción de información musical en un audio, y determinar los valores más adecuados para nuestro estudio.
III. Desarrollar varios prototipos para detectar la tonalidad de una obra musical de forma eficiente, intentando optimizar al máximo su tiempo de respuesta y pudiendo extenderse a base de datos completas de ficheros de audio.
IV. Desarrollar esos algoritmos de forma que se puedan ejecutar en cualquier ordenador convencional de forma eficaz.
V. Elaborar un *data set* (base de datos) compuesto por muestras de audio etiquetadas con sus respectivas tonalidades para poder comparar los distintos prototipos desarrollados.

VI. Definir el protocolo de uso, estableciendo y describiendo, de forma clara y detallada, los pasos a seguir para la obtención de datos tonales en los distintos algoritmos implementados.

VII. Establecer su fiabilidad comprobando experimentalmente la bondad de los resultados obtenidos.

VIII. Realizar una evaluación de los resultados que se obtienen al ejecutar los prototipos según el porcentaje de precisión del sistema al detectar la tonalidad, así como una comparación de esos prototipos con otro software existente.

3. ANTECEDENTES

En este apartado del desarrollo del proyecto pasamos a describir la terminología musical necesaria para explicar el concepto de tonalidad, así como los conceptos básicos de la física del sonido y los modelos matemáticos funcionales que los describen, dando una breve explicación del papel que desempeñan en la música.

3.1 CONCEPTOS MUSICALES Y TERMINOLOGÍA

Existen muchas formas de definir la música, pero se puede decir que es el arte de ordenar sensible y lógicamente una combinación coherente de sonidos con el fin de crear una determinada emoción en el oyente (Moscoso, 2015).

Los griegos proyectaron su idea de sistema tonal como un método musical fundamentado en un efecto de la propia naturaleza física de los sonidos, estudiando los armónicos.

3.1.1 SONIDO

Desde un punto de vista físico, el sonido es una vibración que se propaga en un medio elástico. Para que se produzca sonido se requiere la existencia de un cuerpo vibrante, denominado foco, y de un medio elástico que transmita esas vibraciones,

que se propagan por él constituyendo lo que se denomina onda sonora (Gardey y Pérez Porto, 2010).

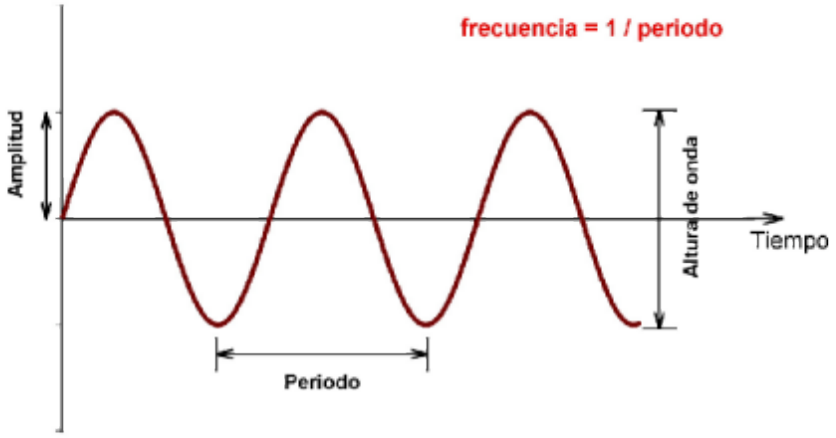


Figura 1. Onda sonora (Martín, 2007)

Un sonido se diferencia de otro por sus características de percepción, las cuales son: su intensidad (fuerza con que se percibe), que puede ser fuerte o débil; su tono (marca la frecuencia o número de vibraciones por segundo que produce el cuerpo que vibra), y puede ser grave o agudo; y por último, su timbre (cualidad que nos permite distinguir entre dos o más sonidos producidos por distintas fuentes sonoras).

La mayoría de los cuerpos vibrantes oscilan a diferentes frecuencias simultáneas, conocidas como parciales. La mayoría de los parciales son múltiplos enteros de la frecuencia fundamental y se denominan armónicos.

La señal resultante de sumar todas esas ondas sinusoidales de distintas frecuencias que se producen a la vez ya no es sinusoidal y puede no ser periódica. En la figura siguiente se muestra un ejemplo.

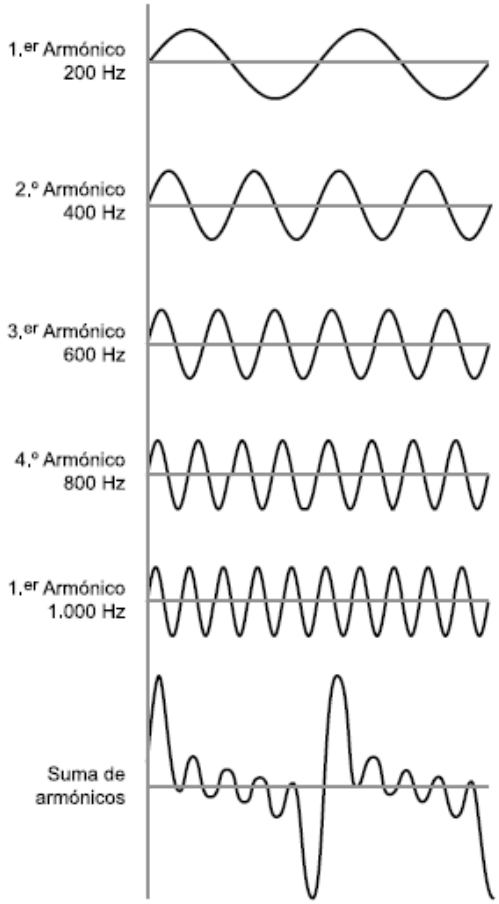


Figura 2. Serie de armónicos (Vilanova Ángeles)

3.1.2 FRECUENCIA

La frecuencia del sonido hace referencia a la cantidad de veces que vibra el aire que transmite un sonido en un periodo de tiempo de un segundo.

Según el sistema internacional (SI), la frecuencia se mide en Hertzios (Hz), unidad llamada originalmente ‹‹ciclos por segundos››, ya que como su propio nombre indica, hace referencia al número de repeticiones de un patrón por unidad de tiempo o a la frecuencia de un fenómeno repetido una vez por segundo.

$$f = \frac{1}{s} = 1\ Hz$$

Los sonidos que percibimos como agudos tienen una frecuencia mayor que los sonidos graves. Teóricamente, el espectro audible por el ser humano es de 20 Hz a 20 kHz.

La gama de frecuencias fundamentales con las que se trabaja en la música es de 27 Hz a 4186 Hz (frecuencias que se corresponden con la primera y última nota del piano). La región superior, (hasta los 20 kHz) está dedicada para los parciales o armónicos de los tonos agudos, que dan el timbre de los diferentes instrumentos.

Aquellas frecuencias fuera de este rango no se suelen utilizar para hacer música, ya que no todas las personas son capaces de percibir esos sonidos o pueden no resultar agradables a la percepción humana.

3.1.3 AMPLITUD

La amplitud es la cantidad de energía acústica que contiene un sonido, es decir, lo fuerte o suave de un sonido. Se corresponde con el valor máximo, tanto positivo como negativo, que puede llegar a adquirir la onda sinusoidal (Vilanova Ángeles).

La intensidad viene determinada por la potencia, que a su vez está determinada por la amplitud y nos permite distinguir si el sonido es fuerte o débil.

No hay que confundir amplitud con volumen o potencia acústica, aunque lo que sí que es cierto es que cuanto más fuerte es el sonido percibido, mayor amplitud tiene la onda sinusoide, porque se ejerce una presión mayor en el medio.

Habitualmente nos referiremos a la amplitud de una onda sonora empleando su medida en decibelios (dB), aunque también puede venir expresada en milibares y pascales. Los sonidos que percibimos deben superar el umbral auditivo (0 dB) y no llegar al umbral de dolor (140 dB).

3.1.4 NOTAS

En la rama de la física del sonido, una nota musical hace alusión a un sonido determinado por una vibración cuya frecuencia fundamental (tono) es constante.

La notación musical podemos definirla como el conjunto de signos convencionales por los cuales se indican los sonidos de la música y su interpretación. Los sistemas de notación han variado según los periodos y los géneros de música a lo largo de la historia.

Las notas que forman el sistema musical occidental se representan con palabras o símbolos cuya serie se repite cíclicamente conforme aumenta la frecuencia: Do, Re, Mi, Fa, Sol, La, Si.

En los países de habla inglesa, se emplea notación anglosajona, también conocida por notación alfabética ya que utiliza siete letras, de la A a la G (de La a Sol) para nombrar las notas musicales (Iglesias González y Robles Ojeda, 1999).

Sistema de notación europea:	La	Si	Do	Re	Mi	Fa	Sol
Sistema de notación anglosajona:	A	B	C	D	E	F	G

Tabla 1. Sistemas de notación europea y anglosajona

En el presente trabajo emplearemos indistintamente las notaciones anglosajona y española. Priorizaremos la española a la hora de realizar la mayoría de explicaciones, pero cuando nos situemos en el contexto de los programas desarrollados emplearemos la notación anglosajona con mayor asiduidad.

Las notas de la escala musical tienen unas frecuencias de sonido determinadas. La diferencia entre dos notas consecutivas puede ser de un tono o un semitono. Un tono está compuesto de dos semitonos; por tanto, entre dos notas separadas por un tono, podemos incluir otra nota que divide esta distancia en dos semitonos. Para designar estas notas utilizamos las alteraciones, las cuales se clasifican en tres tipos:

- Sostenido (#): la nota afectada asciende un semitono.
- Bemol (*b*): la nota afectada desciende un semitono.
- Becuadro (♮): anula el efecto del sostenido o bemol.

La mejor referencia visual del sistema musical occidental se puede encontrar en el teclado del piano. Las siete notas antes descritas vienen representadas por las teclas blancas. Sin embargo, en el intervalo de las teclas blancas también hay cinco teclas negras, éstas se nombran en relación con la nota blanca más próxima utilizando las alteraciones pertinentes.

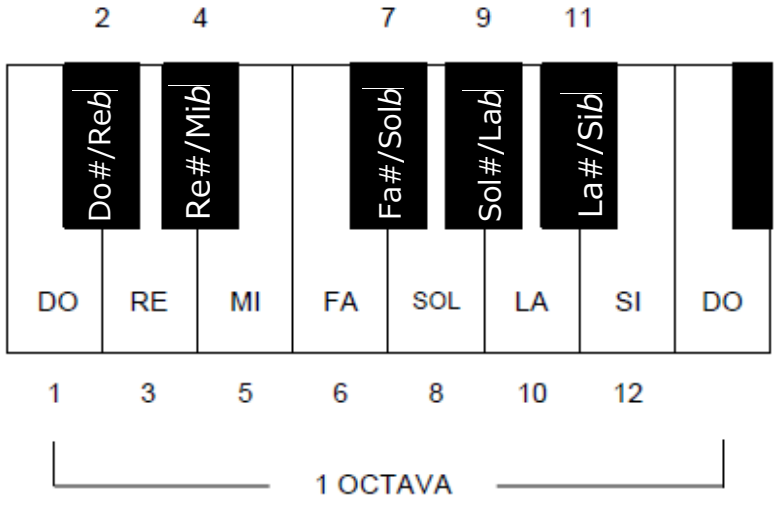


Figura 3. Escala del piano. Notas naturales (asociadas al color blanco de las teclas). Notas alteradas (correspondientes a las teclas de color negro).

3.1.5 OCTAVAS

La distancia entre dos notas se denomina intervalo musical. La octava es el intervalo más simple que separa dos sonidos cuyas frecuencias fundamentales tienen

una relación 2:1. Por ejemplo, el La_4 (A5 en inglés) con una frecuencia fundamental de 440 Hz está situada una octava por encima respecto al La_3 (A4) con 220 Hz. Como podemos observar la frecuencia de la nota separada por un intervalo de octava respecto a otra del mismo nombre se duplica.

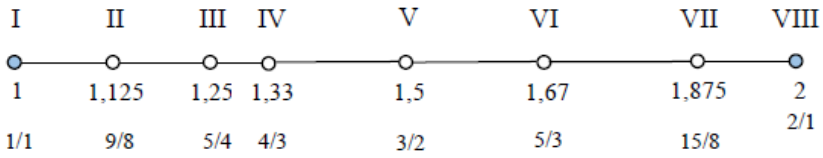


Figura 4. Relación entre las frecuencias de la escala diatónica, situadas sobre la recta real (Segura Sogorb, 2015, p.57)

El término octava es utilizado para designar por tanto una misma nota pero de una serie siguiente o de una anterior (Anchuela Arnalte, 2013).

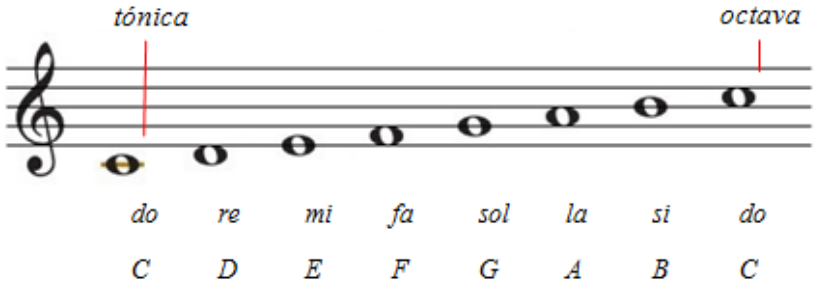


Figura 5. Octava de la tonalidad de Do Mayor

No existen dos notas más parecidas musicalmente que una y su respectiva octava, y por ello reciben el mismo nombre. Para saber a qué octava pertenece una nota se pone un índice numérico a su derecha y toda la serie de doce notas, de Do a Si, lleva el mismo índice. Cuanto más alto sea este número, la nota será más aguda y cuanto más bajo sea este número la nota será más grave. Así por ejemplo, Do_3 (C3 en inglés) es más grave que Do_6 (C6 en inglés).Las notas con la octava más baja tienen el índice 0 y las más altas el índice 10 (López y Molina, 2007).

3.1.6 SISTEMA TEMPERADO O TEMPERAMENTO IGUAL

El temperamento igual o sistema temperado es el sistema musical que se impuso en la música occidental.

En el sistema temperado la octava se divide en 12 notas. El temperamento igual consiste en establecer una misma distancia entre las notas de una octava. Así pues, en este sistema musical, la distancia entre dos notas adyacentes se denomina semitono (ST). A continuación veremos cómo deducir la razón de ser matemática del sistema temperado, calculando el intervalo de semitono en este sistema.

La frecuencia de cada nota se obtiene de multiplicar la nota anterior por una razón, ya que el intervalo entre nota y nota no deja de ser una proporción de frecuencias, siendo todas ellas iguales en una escala logarítmica. Así pues, la relación de frecuencias correspondientes a un semitono temperado se corresponde con $\sqrt[12]{2} = 1{,}05946$ (Castro, 2009).

Los problemas de afinación en instrumentos con intervalos fijos como el piano hizo construir una escala en la que el intervalo entre dos notas consecutivas fuese siempre el mismo (un semitono). Para ello se fijó la frecuencia de una nota de referencia, a partir de la cual poder deducir todas las otras. La nota y frecuencia escogidas fueron el La_4 (A5 en inglés) a 440 Hz (nota La de la escala central del piano).

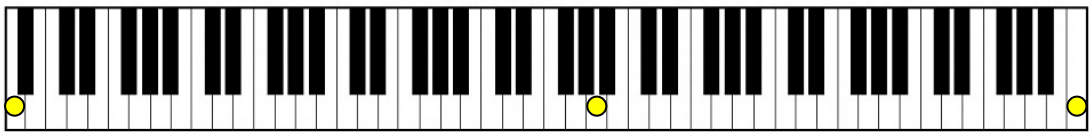


Figura 6. Rango de frecuencias dentro de un piano.

Para hallar la proporción correspondiente a cualquier intervalo de la escala temperada, tan sólo tendremos que elevar $\sqrt[12]{12}$ al número de semitonos (ST) que contiene dicho intervalo en cualquier dirección con respecto a la nota de referencia.

$$F_0nota = F_0ref\ \sqrt[12]{2^n}$$

Donde F_0nota es la frecuencia fundamental de la nota a calcular, F_0ref es la frecuencia fundamental de la nota de referencia (440 Hz) y n indica el

desplazamiento de semitonos desde la nota a calcular a la nota La de referencia (Molina García, 2010).

Esta relación es obtenida sabiendo que al aumentar una octava, una nota queda multiplicada por dos (relación 2:1) y, además, la octava se divide en doce partes de forma logarítmica.

3.1.7 ESCALAS

Por escala musical entendemos un conjunto de notas correlacionadas subiendo o bajando que mantienen una relación musical propia y unificada. Las escalas son utilizadas para la composición de melodías (Piston, 1987).

Cada escala tiene una estructura que la caracteriza y que viene dada por la distancia entre sus grados. Cada grado se identifica mediante un número romano, del I al VII, e indica la posición de la nota dentro de la escala. Esta distancia puede ser de un semitono o de un tono (dos semitonos), donde el semitono es la distancia mínima que puede ser establecida entre dos notas consecutivas.

Comúnmente, en la música occidental, se emplea un subconjunto de sonidos conocido por el nombre de "*escala diatónica*", que emplea 7 de las 12 notas que forman una octava y posee dos formas básicas con intervalos definidos: la escala diatónica mayor y la escala diatónica menor (Piston, 1987).

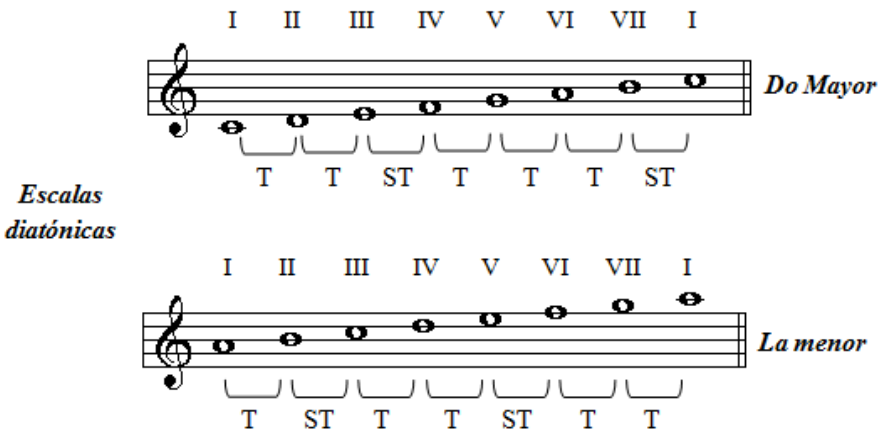


Figura 7. Escalas diatónicas de Do Mayor y La menor

Básicamente, en el presente trabajo, nos referiremos a las dos escalas mencionadas anteriormente, cuyas estructuras y distancias de tono (T) o semitono (ST) entre sus grados se pueden observar en la siguiente tabla:

Escalas diatónicas	Estructura						
	I-II	II-III	III-IV	IV-V	V-VI	VI-VII	VII-I
Mayor natural	T	T	ST	T	T	T	ST
Menor natural	T	ST	T	T	ST	T	T

Tabla 2. Patrón de tonos y semitonos de la escala diatónica mayor y diatónica menor

La nota tomada como base de la escala, correspondiente al grado I, es la nota principal y se denomina tónica (Tomasini, p. 16). Se dice entonces, que esa escala está en la tonalidad cuyo nombre viene determinado por el primer grado o tónica.

La modalidad hace referencia a la elección específica de los sonidos con relación a una tónica particular, por lo que se ocupa de los diferentes tipos de escalas. Las escalas mayores o menores son las modalidades específicas más familiares (Piston, 1987) y constituyen el eje de la inmensa mayoría de la música actual, ya sea procedente de la tradición clásica o de los artistas de rock, reggae, jazz, folk, hip hop, blues, o pop entre otros muchos.

Estas modalidades presentan caracteres diferentes. La escala mayor posee una sonoridad optimista, brillante, feliz y ligera, mientras que la escala menor presenta un sonido más emotivo, triste, melancólico y oscuro.

Las dos escalas vistas hasta ahora (Figura 7) no tienen ninguna nota alterada. Sin embargo, si queremos formar una escala Mayor o menor sobre otras notas tendremos que utilizar alteraciones para que tengan el esquema de tonos y semitonos necesario, de tal forma que cada una de las escalas posee un número concreto de alteraciones (Iglesias Castro, 2013).

Existen otros tipos de escalas que no son útiles en términos de clasificación de la tonalidad, pero que pueden considerarse como la raíz de la cual derivan las escalas

diatónicas. Hablamos entonces de la escala cromática, compuesta por un total de 12 sonidos, los naturales (Do, Re, Mi, Fa, Sol, La, Si) y los alterados (Do#/Re*b*, Re#/Mi*b*, Fa#/Sol*b*, Sol#/La*b*, La#/Si*b*), es decir, aquellos que se ven afectados por las alteraciones sostenido/bemol.

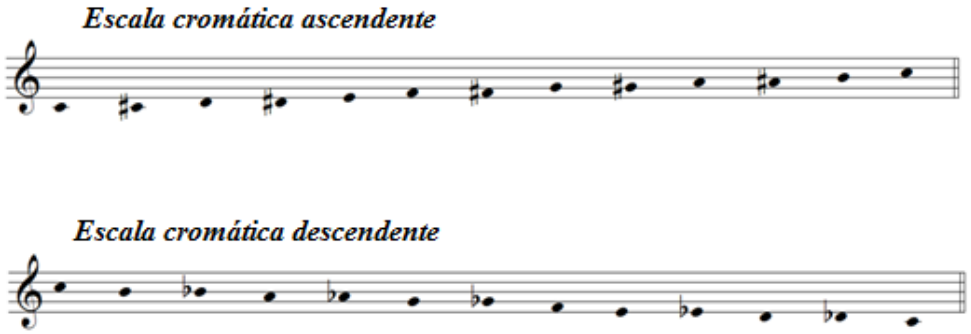


Figura 8. Escala cromática ascendente y descendente

3.1.8 TONALIDAD

La música tonal occidental está basada en siete notas (doce sonidos o clases de alturas), llamados grados, los cuales se agrupan siguiendo un sistema denominado tonalidad en torno a un sonido que actúa como centro de gravedad y que recibe el nombre de tónica.

La escala que forma la tonalidad Do Mayor es tan solo un ejemplo, el más básico de todos, ya que sus notas se corresponden con las teclas blancas del piano, pero también podemos construir nuestras propias tonalidades a partir de cualquier nota musical siempre y cuando mantengamos las mismas distancias o, mejor dicho, la misma distribución de intervalos de tonos y semitonos entre las notas que forman la escala.

La forma más fácil de entender el concepto de tonalidad la podemos encontrar gráficamente en el denominado círculo de quintas, empleado de múltiples formas prácticas en el arte musical. En la figura adjunta podemos observar 12 tonos mayores y 12 menores. Además, éstos se encuentran estrechamente relacionados entre sí dentro del mismo círculo, puesto que para cada tonalidad mayor (en el exterior)

existe una tonalidad menor (en el interior) con idéntica armadura (notas alteradas), diferenciándose tan solo en la nota tónica.

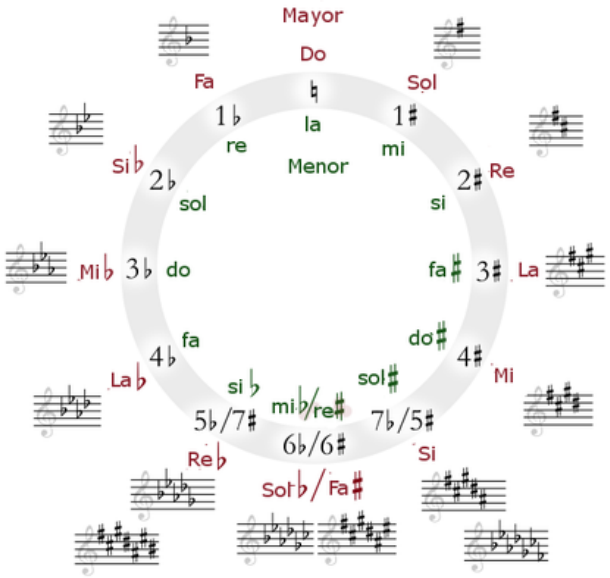


Figura 9. Círculo de quintas. Tonalidades mayores y sus respectivas menores que se conforman según el número de alteraciones sostenido/bemol (Martínez Salanova, 2015).

Se habla entonces de tonos relativos (relativo mayor y relativo menor), denominación que se puede extender a los acordes. Por ejemplo la tonalidad de La menor es el relativo menor de la tonalidad Do Mayor, por que las tonalidades a las que dan nombre respectivamente comparten armadura (Gil Pérez, Iglesias González y Robles Ojeda, 1998).

En cada una de las 24 posibles tonalidades se despliega, desde una tónica dada, el patrón de tonos y semitonos del modelo diatónico y se dan las mismas relaciones de alturas que en otras tonalidades.

Debido a la doble nomenclatura sostenido/bemol se da pie de nuevo a la enarmonía, es decir, a que existan tonalidades y acordes con sonido equivalente. Por tanto, las notas que pueden recibir dos nombres se denominan enarmónicas y el contexto en que se usan es lo que determina qué nombre es el apropiado.

3.2 EXTRACCIÓN DE CARACTERÍSTICAS MUSICALES DE LA SEÑAL DE AUDIO

En esta sección se presentan los modelos matemáticos funcionales que describen los conceptos de la física del sonido y que tienen relevancia en el presente trabajo.

Mencionamos algunos de los procedimientos para interpretar datos de una señal de audio como el análisis en el dominio temporal y espectral, modelos que nos permiten seleccionar qué información es extraída y de qué forma se analiza.

3.2.1 ANÁLISIS EN EL DOMINIO TEMPORAL

Una señal de audio se representa gráficamente de forma simple en el dominio temporal u oscilograma. Éste nos permite efectuar el estudio y caracterización de la estructura de una señal, representando las variaciones de la amplitud de la muestra de audio (después de haber sido muestreada) a lo largo del tiempo, para el análisis de propiedades como la sonoridad, la amplitud y la duración (Gil Fernández, 1987).

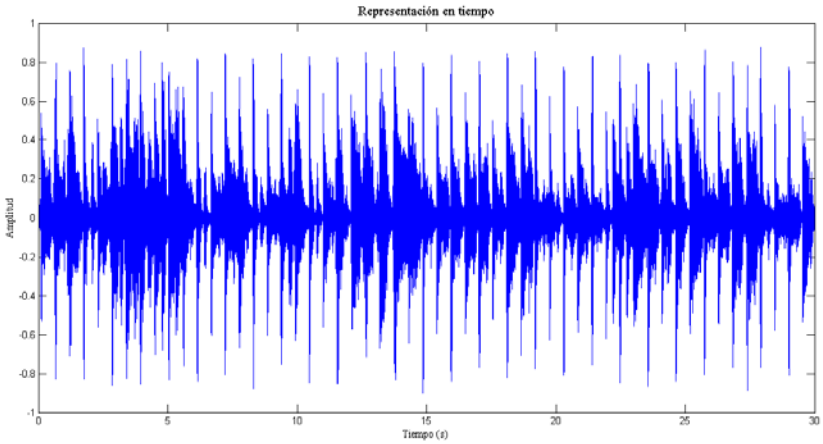


Figura 10. Análisis de la pista disco32.wav en el dominio temporal

Se trata por tanto de un conjunto de técnicas que posibilitan la descripción y manipulación de señales cuyo cómputo varía en el tiempo. En el caso del sonido, la amplitud representa la variación de la presión atmosférica respecto al eje temporal (Hernández de Miguel, 2012).

Por lo general la amplitud de una señal de audio, se representa a partir de un valor 0, que identifica la posición de equilibrio o dicho de otro modo, expresa el valor medio de la presión, y se extiende hasta el valor de máxima amplitud, dando forma a la onda sonora (Gómez Gutiérrez. 2009).

La amplitud de una muestra de audio cambia tantas veces por segundo que es imposible extraer de ella una información útil de cara al diseño del software. Por tanto, no proporciona la información contenida en una señal de cara al diseño del software, ya que hay ciertas propiedades que no podemos analizar en el dominio temporal y que sí se pueden estudiar en el dominio frecuencial.

3.2.2 LA TRANSFORMADA DISCRETA DE FOURIER

A través del análisis de principios fundamentales de la música como la melodía y la armonía de una señal de audio, se pretende mostrar la importancia de la Transformada de Fourier (FT) como base del entendimiento de la música empleando modelos físicos matemáticos.

La Transformada de Fourier permite descomponer el sonido en senos y cosenos de diferentes frecuencias y amplitudes. En las aplicaciones de ingeniería y tratamiento de señales resulta más práctico considerar el proceso de manera discreta y no continua, ya que los sistemas de adquisición de datos no pueden obtener ni analizar la totalidad de la información. Esta representación está basada en señales exponenciales complejas y se denomina *Transformada Discreta de Fourier*, nombre que se le da a la FT cuando se aplica a una señal digital (discreta) en vez de a una analógica (continua).

$$x[k] = \sum_{n=0}^{N-1} x[n] e^{-\left(2\pi n \frac{k}{N}\right) j} = \sum_{n=0}^{N-1} x[n] {W_N}^{nk},$$
$$k = 0, 1, \dots, N-1 .$$

Transformada discreta de Fourier (LezanaIllesca, 2005).

- N: Número de muestras en x[n] (incluyendo los ceros).
- x[n]: Señal de prueba discreta (con índice n=Enésima muestra original).
- x[k]: Espectro en función de la frecuencia discreta (con índice k=késimo término de la DFT).
- $e^{-(2\pi nk\,/N)j}$: Factor de fase(W_N).

La Transformada Discreta de Fourier, por tanto, es el nombre dado a la Transformada de Fourier cuanto se aplica a una señal digital discreta en el dominio del tiempo y frecuencia y definida para secuencias de duración finita. A pesar de que es una transformada capaz de ser implementada en una computadora, su eficiencia es muy baja, especialmente cuando la longitud de la secuencia es larga (Ortiz Arreygue, 2014).

Como alternativa a la Transformada Discreta de Fourier (DFT) se implementa el algoritmo de la transformada rápida de Fourier (FFT), el cual realiza los mismos cálculos que la DFT pero de manera mucho más rápida gracias a su recursividad. El algoritmo pone algunas limitaciones en la señal y en el espectro resultante. Solo puede ser aplicada cuando el número de muestras de la señal es una potencia de dos ($N = 2^m$, donde $m = 1,2,3,\ldots$). Un cálculo de la FFT toma aproximadamente $N \cdot \log_2(N)$ operaciones, mientras que DFT toma aproximadamente N^2 operaciones, así que la FFT es significativamente más rápida y por tanto da como resultado un algoritmo más eficiente (Franco García, 2015).

El algoritmo FFT consigue simplificar el cálculo de la DFT y disminuir los errores de redondeo, aprovechando la simetría y periodicidad del término W_N según Gentleman y Sande (1966), ya que contribuyen a la redundancia de la DFT.

$$W_N{}^{nk} = e^{-j\frac{2\pi}{N}nk} = \cos\left(\frac{2\pi}{N}\right) - j \cdot sen\left(\frac{2\pi}{N}\right)$$

$$W_N{}^{n+N} = W_N{}^{n}$$

$$W_N{}^{n+\frac{N}{2}} = -W_N{}^{n}$$

El análisis frecuencial mediante FFT es muy versátil y nos puede proporcionar una gran cantidad de información útil. Explicado de forma sencilla, con el análisis del método de Fourier buscamos reducir fragmentos de la señal de audio en sus componentes espectrales para poder extraer de ellas las diversas frecuencias presentes en la señal y sus valores de amplitud.

3.2.3 ANÁLISIS EN EL DOMINIO FRECUENCIAL

El termino dominio frecuencial se emplea para describir el análisis matemático o de señales con respecto a la frecuencia en vez del tiempo. Mientras el dominio temporal muestra el cambio de la señal a lo largo del tiempo, el dominio de la frecuencia muestra cuánto de la señal hay en el rango de frecuencias.

La representación frecuencial captura por tanto las características espectrales de una señal de audio. Una señal sinusoidal pura viene representada por una sola componente frecuencial. Este caso es ideal, pero las señales de audio que analizaremos a lo largo del proyecto contendrán sonidos mucho más complejos y difíciles de analizar (Casado García, 2011).

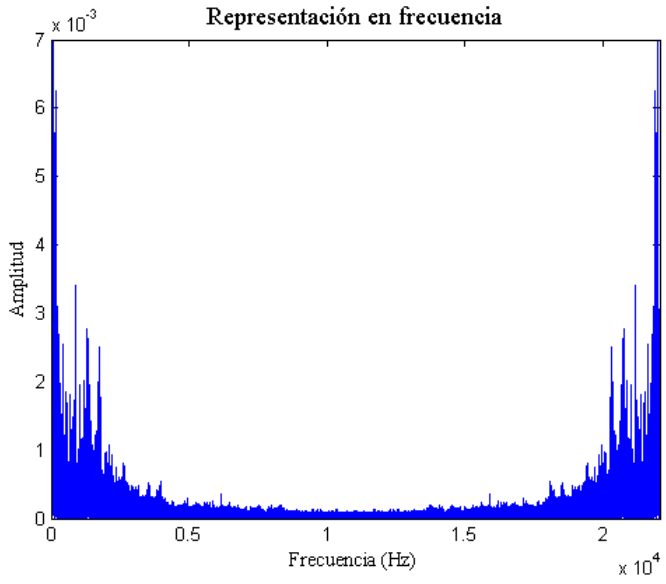


Figura 11. Análisis de la pista disco32.wav en el dominio frecuencial

El contenido frecuencial de una señal de audio puede representarse de diversas formas. Una forma estándar es la de dibujar cada parcial (componente frecuencial) como una línea en el eje x. La altura de cada línea se correspondería con la potencia o amplitud de cada componente frecuencial. De este modo obtendremos las frecuencias existentes en la señal analizada.

La información contenida en la frecuencia es indispensable ya que en la representación temporal la información comprendida en la señal no es del todo clara. El espectro en frecuencia de una señal nos permite analizar ciertas propiedades que aportan una mayor información de la muestra de audio que la que puede ofrecernos un oscilograma, como el timbre, la frecuencia fundamental, la sonoridad y la amplitud (Gil Fernández, 1987).

En la forma de onda de una muestra de sonido hay muchas frecuencias presentes. Al conjunto de frecuencias fundamentales o tonos contenidos en la señal le acompañan la multitud de frecuencias armónicas, cuyas amplitudes son menores que la de la onda sinusoidal y tienden a cero (Pérez, 2008). Por ello, cualquier frecuencia puede denominarse parcial, sea o no múltiplo de una frecuencia fundamental. De hecho, muchos sonidos no tienen una fundamental clara (Gómez Gutiérrez, 2009).

Cualquier muestra de sonido puede ser convertida en este dominio mediante un operador matemático denominado Transformada de Fourier, convirtiendo la señal de audio en una combinación lineal de exponenciales complejas armónicamente relacionadas mediante las series de Fourier, es decir, descomponiendo esa señal compleja en otras más simples (Basso, pg 78, 2001).

Este dominio nos da la amplitud de las partes reales e imaginarias de las sinusoidales que tenemos que sumar en cada frecuencia. Aplicando el valor absoluto a la transformada conseguimos realizar la raíz cuadrada de la parte real y la parte imaginaria de la señal que expresa el valor de la energía proporcionada por cada frecuencia (para mostrar solo la magnitud y no la parte real e imaginaria). Esto tiene su explicación haciendo uso extensivo de la fórmula de Euler.

$$A\cos(2\pi f) = \frac{1}{2}A^{i2\pi f} + \frac{1}{2}A^{-i2\pi f}$$

$$A[k] = |x[k]| = \sqrt{Re(x[k])^2 + Im(x[k])^2}$$

Este tipo de representación gráfica presenta ventajas sobre el oscilograma, ya que manifiesta ciertas características de la señal, pero se trata de un modelo que no muestra información sobre el tiempo, por lo que la interpretación resulta útil exclusivamente al llevar a cabo un análisis de puntos específicos de la secuencia estudiada.

3.2.4 ESPECTROGRAMA O SONOGRAMA

Hasta el momento hemos visto el espectro de una sola ventana. Para estudiar la evolución de la muestra de audio analizada debemos calcular el espectro de tramas *enventanadas* de la señal, a lo que denominamos espectrograma. Una alternativa fácil para interpretar este proceso consiste en emplear el color como una tercera dimensión.

Un espectrograma o sonograma es una representación gráfica de la evolución de los valores de amplitud (intensidad) o de la energía del contenido frecuencial de una señal de audio según ésta varía a lo largo del tiempo (Colomer Blasco, 2016). Esta herramienta la utilizamos para analizar propiedades como la sonoridad, la duración, el timbre y la amplitud (intensidad) (Gil Fernández, 1987).

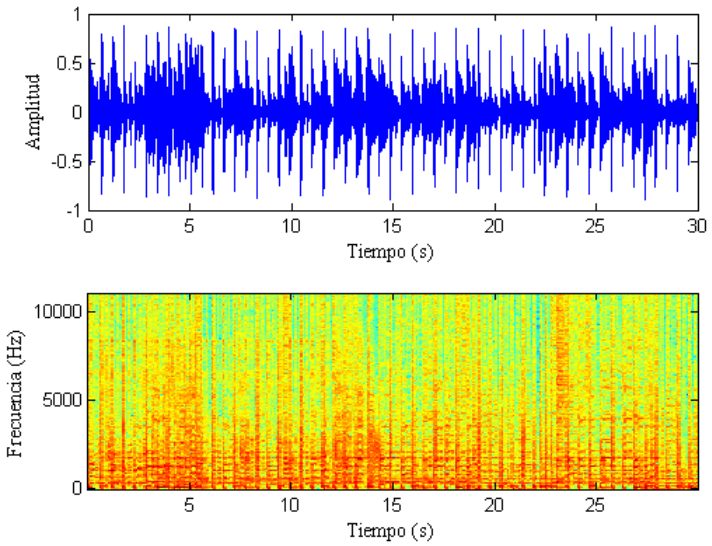


Figura 12. Espectrograma de la muestra de audio disco32.wav

Teóricamente la Transformada de Fourier calcula el espectro de una señal infinita y estacionaria. Sin embargo, en la práctica, las muestras de audio analizadas son señales finitas y cambiantes (no estacionarias). Consecuentemente debemos buscar un procedimiento capaz de analizar las señales en instantes de tiempo determinados. Una forma simple se basa en la utilización de una función ventana $w[n]$ con una anchura fija de muestras.

Al espectrograma de señales muestreadas, de duración finita y variantes respecto al tiempo, se le aplica, como se describe a continuación, la Transformada de Fourier de corta duración (STFT), dividiendo la señal en pequeñas porciones o fragmentos denominados tramas de análisis.

El funcionamiento básico de la STFT comprende los siguientes pasos:

- Con la función ventana, encuadrar la señal alrededor de un instante y calcular sobre ella la FT.
- Posteriormente, trasladar la función ventana de modo que se solapen un número pequeño de muestras (OVERLAP) con la trama anterior a fin de asegurar la continuidad temporal, cubriendo una nueva porción de la señal a la que volvemos a calcular la FT.
- Este proceso se repite hasta haber cubierto la totalidad de la señal.

La STFT aplica la Transformada Discreta de Fourier (DFT) a cada segmento *enventanado*. Su expresión para señales discretas viene dada por:

$$X[k,r] = \sum_{n=0}^{N-1} x[n]w[n-rI]e^{-j\frac{2\pi kn}{N}} \quad , \qquad k = 0, 1, \dots, N-1$$

Transformada de Fourier de Corta Duración (Segura Sogorb, pg. 33, 2015).

- N: Número de muestras en x[n].
- $x[n]w[n-rI]$: Señal enventanada, donde r es el número entero de ventana e I el tamaño (en muestras) del desplazamiento de ventana.
- $X[k,r]$: Espectro de la señal *enventanada*.

- $e^{-(2\pi nk/N)j}$: Factor de fase (W_N) .

Así pues, un espectrograma da como resultado una serie de valores que se corresponden con las amplitudes de las distintas frecuencias. El módulo de éstas son las que se codifican en una gama de colores (Bernal, Gómez y Bobadilla, pg. 94). La intensidad por tanto se traduce en una variación de grosor y de valor en la escala de color. La magnitud cuadrada de la STFT origina por tanto el espectrograma de la señal.

Por tanto, un espectrograma computa el contenido frecuencial de un determinado número muestras extraídas por medio de una ventana temporal y posteriormente las representa en tres dimensiones. Esta operación se repite sucesivamente a lo largo de la señal, desplazando en el tiempo la ventana definida inicialmente para tomar muestras dispares en cada trama *enventanada* de la señal de audio analizada. Una vez puesta en ventana la señal, se calcula la STFT de las muestras contenidas en la trama a estudiar. Cada trama obtenida del cálculo de la STFT se indexa de forma consecutiva en una matriz. Dicha matriz es la que da forma al espectrograma y permite representar la variación del espectro de la señal y la energía en función del tiempo.

De este modo, en el procesamiento de la señal, se pueden aprovechar las características producidas por la concentración de la energía en dos dimensiones, tiempo y frecuencia, en vez de solo una, tiempo o frecuencia. El único inconveniente es que, una vez fijada la función ventana, las resoluciones de dichas características son inversamente proporcionales, por lo que la resolución temporal se puede mejorar sólo a expensas de sacrificar resolución en frecuencia y viceversa.

3.2.5 VENTANA

El primer paso para la creación de un espectrograma consiste en realizar un análisis por tramas. Para ello es necesario aplicar una ventana $w[n]$ que seleccione un número determinado de muestras a procesar (WINDOW). La ventana no es más que una función que dista de cero en un rango limitado de tiempo y es usada

frecuentemente en el análisis y procesado de señales de audio con el objetivo de evitar discontinuidades en los inicios y fines de las tramas analizadas.

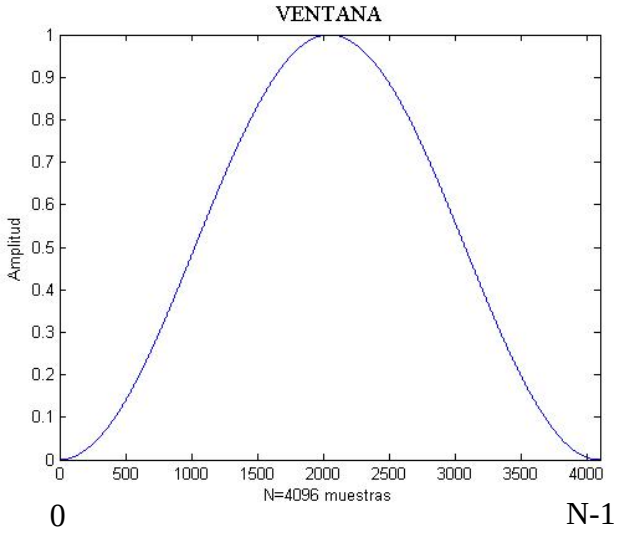


Figura 13. Ventana de longitud N= 4096 muestras

Cómo se ha mencionado en el apartado anterior tenemos dos parámetros con los que podemos jugar: el intervalo de tiempo entre ventanas y su longitud (igual a la longitud de la FFT). Con el tiempo entre ventanas podremos influir sobre la resolución temporal, es decir, si aplicamos un OVERLAP muy bajo perderemos los fenómenos ocurridos en el intervalo de tiempo donde las ventanas tiene mínima amplitud, mientras si aplicamos un OVERLAP muy elevado haremos un promedio de distintos intervalos de tiempo para una misma muestra con lo que quedará un espectrograma más suavizado.

Con respecto a la longitud de la ventana, dependiendo del tamaño que utilicemos para el análisis de Fourier, tendremos diferentes niveles de resolución del espectrograma. Si aplicamos una ventana demasiado pequeña no seremos capaces de distinguir los diferentes armónicos si están muy juntos en el espectrograma. Por el contrario, con una ventana muy grande obtendremos el efecto inverso, es decir, un espectrograma muy detallado pero a costa de incrementar el tiempo de cálculo necesario para la operación. Generalmente esta longitud es una potencia de 2 por razones de eficiencia en el cálculo de la DFT.

Nos encontramos por tanto con un doble compromiso que debemos equilibrar, ya que si aumentamos la resolución temporal disminuimos la resolución frecuencial y viceversa.

Según su forma, existen diversos prototipos de *enventanado*. No hay un tipo de ventana por excelencia; cada uno de los existentes presenta una serie de ventajas e inconvenientes. La elección de uno u otro va a depender de las operaciones realizadas o del tipo de señales que se traten. Por lo tanto, debemos elegir aquella función ventana que cause menor distorsión en la señal de audio analizada.

A continuación se muestra una comparativa de las distintas opciones de *enventanado* más utilizadas en el procesamiento de audio digital.

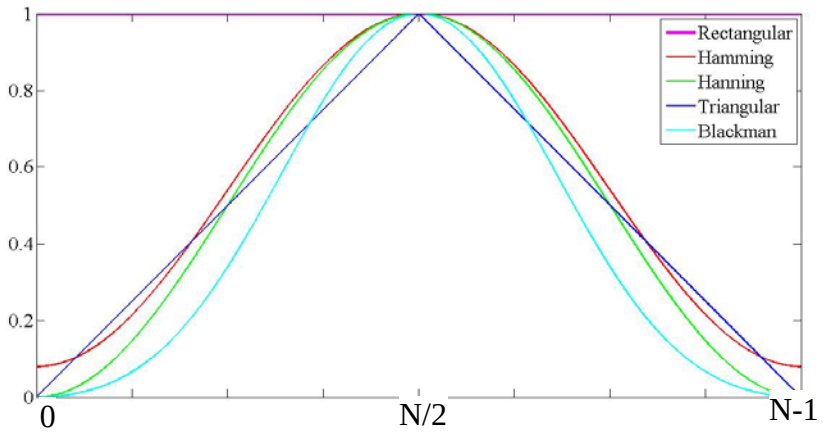


Figura 14. Diferentes formas de las funciones ventana con longitud N muestras

Si la señal de entrada está formada por una combinación de tonos simples, conviene aplicarle un tipo de ventana con un lóbulo central estrecho con la finalidad de que localice bien las frecuencias contenidas en la señal, como la venta rectangular. En cambio, para señales ruidosas y cambiantes en las que la energía no está del todo clara en ciertas frecuencias como son las señales musicales, una ventana con el lóbulo central más ancho como la Hamming será más eficaz.

3.2.6 CLASES DE ALTURAS Y CHROMA

La percepción humana del tono es periódica en el sentido de que dos tonos que difieren por una octava son percibidos como similares en "color". Si estos tonos

o notas, difieren por una o varias octavas se pueden separar en dos componentes, que se denominan altura de tono y *chroma*.

En la tonalidad, un componente crucial es la relación de octava referida a la altura del tono. El modelo realizado por Roger Shepard para dar claridad al concepto de altura tonal es uno de los más representativos ya que permite estudiar las diferencias de alturas absolutas, del color tonal, es decir, ilustra la relación que existe entre los intervalos de octavas. A este modelo se le conoce por el nombre de ‹‹hélice de alturas››, la cual sitúa los tonos en la superficie de un cilindro formado por dos dimensiones (Musiki, 2017).

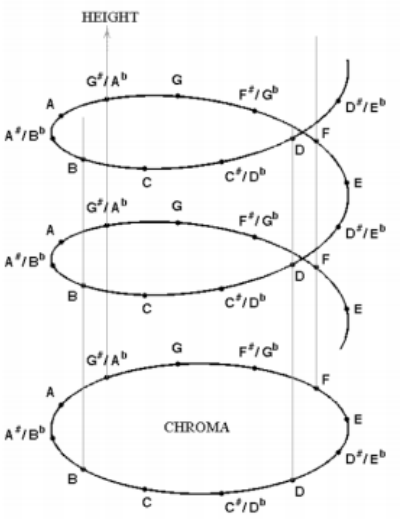


Figura 15. Hélice de alturas (Segura Sogorb, 2015)

Dos alturas iguales separadas por una octava se ubican en la misma línea vertical. Por tanto, comparten el mismo ángulo en el círculo de croma, tienen diferente dimensión, poseen el mismo nombre y forman el único intervalo en que los armónicos coinciden exactamente. Por ello son considerados musical y perceptualmente equivalentes. Como podemos observar en la Figura 15, la escala cromática se ubica en la línea curva ascendente (Segura Sogorb, 2015).

Como segunda dimensión, el *chroma,* referido a la posición del sonido dentro de la octava. Hace referencia a qué clase de altura está sonando de entre las disponibles en la escala cromática. Suponiendo que el sistema utilizado se basa en el temperamento

igual de doce tonos, el conjunto vendrá representado por doce valores cromáticos, uno por cada nota musical.

$$C = \{A, A\#, B, C, C\#, D, D\#, E, F, F\#, G, G\#\}$$

Si enumeramos los valores de croma de 0 a 11, donde 0 se corresponde al valor cromático de la nota A (La), 1 al de la nota A♯ (La♯) y así sucesivamente hasta 11, correspondiente con la nota G♯ (Sol ♯), podemos identificar el conjunto de valores cromáticos obtenidos.

$$c = \{0,1,2,3,4,5,6,7,8,9,10,11\}$$

Una clase de tono se define por todas aquellas notas musicales que comparten el mismo *chroma*. Por ejemplo, la clase de tono que corresponde al chroma c = 0 (A) consiste en el conjunto {…, 55, 110, 220, 440, 880,…} que son las frecuencias correspondientes a las notas musicales {..., A2, A3, A4, A5, A6,...}.

3.2.7 CHROMA FEATURES O CHROMAGRAM

Como hemos visto en el apartado 3.2.4, el espectrograma es capaz de representar la evolución temporal del espectro de una señal. Sin embargo, existe un refinamiento de éste, es decir, una mejora, capaz de ajustar aun más su aplicación al análisis y procesamiento de señales musicales a la que denominamos *chromagram* o cromagrama.

Fujishima se propuso describir el espectro de audio mediante una representación denominada “Perfiles de clases de tono o PCP”, conocida actualmente por el nombre de *chroma features* o características de croma. Se trata de una representación cromática para señales musicales (Fujishima, 1999).

Todo el espectro de frecuencias se dobla en una octava, es decir, en 12 compartimentos o dimensiones que representan cada uno de los 12 semitonos en los que se puede dividir una octava. Cada dimensión muestra la energía de cada una de

las doce notas musicales o tonos, sin tener en cuenta la octava en la que se encuentra (Escobar Zamora, 2014).

A partir del cálculo de la Transformada Discreta de Fourier (3.2.2) de una muestra de audio y de la evolución de la energía en un conjunto de bandas de frecuencias, se obtiene el vector *chroma features*, cuyas frecuencias centrales están estrechamente relaciones con las clases de alturas ($A, A\#, B, C, C\#, \dots, G\#$), obteniendo así un vector para cada instante de tiempo denominado *chroma vector* o vector de croma, el cual describe la variación de la energía asociada a las 12 clases de alturas a lo largo del tiempo (Gold, Morgan y Ellis, pg. 569, 1999).

Para la obtención y cálculo de las características de croma, dado un espectrograma como representación musical de un archivo de audio, cuya señal se ve influenciada por un *enventanado* de longitud WINDOW, el proceso se basa en agregar toda la información de *chroma* contenida en un único valor. La aplicación de una ventana de cierta longitud (WINDOW) nos permite extraer de ésta el contenido de cada una de las doce notas musicales o tonos de la señal *enventanada*. La representación de tiempo-croma resultante es lo que se conoce como *chromagram*.

Como hemos podido ver en apartados anteriores, el resultado de representar gráficamente una señal de audio mediante un espectrograma, se puede ver como una matriz de dimensión NxM, donde cada fila muestra los N índices de frecuencia y cada columna, las M ventanas en el tiempo. El *chromagram* se obtiene del resultado de multiplicar una matriz de *chroma* C de dimensión 12xN con el espectrograma de la señal de audio que se quiera analizar (Figura 16).

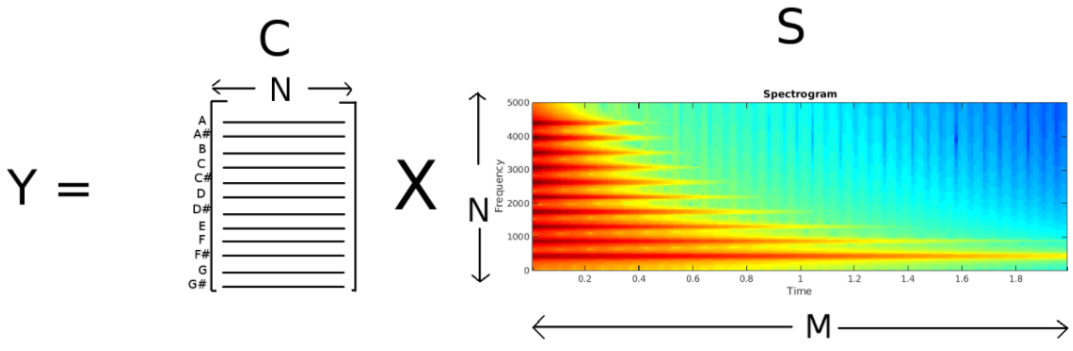


Figura 16. Cálculo de la matriz que da forma al *chromagram* (Tralier)

Como resultado obtenemos una matriz de elementos reales, de dimensión 12xM, compuesta por una fila para cada clase de altura y una columna para cada muestra de tiempo analizada, en la que se muestra la medida de fuerza de cada una de las doce notas posibles contenida en cada *enventanado* de la señal en el espectrograma sobre todas las octavas (Figura 18 b).

a)

b)

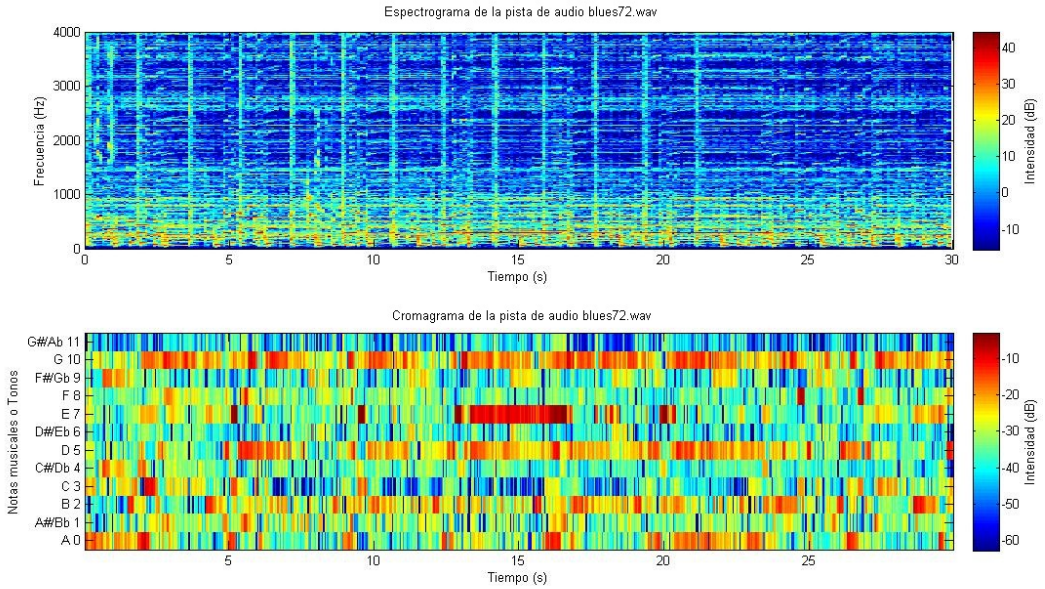


Figura 17. a) Espectrograma de la muestra de audio blues72. b) *Chromagram* de la muestra de audio blues72 (los colores que tienden a rojo denotan mayor energía)

Así como utilizamos el espectrograma para inferir propiedades sobre la distribución de la energía de una señal sobre la frecuencia y el tiempo, el *chromagram* puede usarse para inferir propiedades sobre la distribución conjunta de la intensidad de la señal sobre las variables tiempo y *chroma*.

Esta es una poderosa herramienta para el análisis musical, capaz de capturar características melódicas y armónicas de la música que son robustas frente a cambios en el timbre y la instrumentación.

Dado que, en música, notas separadas por intervalos de octava son percibidos como similares, saber la distribución de *chroma* incluso sin la frecuencia absoluta, es decir, la octava original, puede proporcionar información musical útil sobre el audio e

incluso puede revelar la similitud musical percibida cosa que no es aparente en el espectro original.

4. METODOLOGÍA

En este apartado se presentan las fases seguidas para la realización del Trabajo Fin de Grado y una estimación del tiempo empleado para su desarrollo. En una primera fase de exploración surgió el interés de realizar un Trabajo Fin de Grado relacionado con el análisis musical.

Indagando en el campo de estudio para demarcar el objeto de investigación de dicho trabajo, nos llevó a seleccionar una serie de TFG sobre temas relacionados con el análisis y procesamiento musical. Hay numerosos sitios web en los que podemos disponer de trabajos desarrollados por otros investigadores para extraer información sobre el análisis de una señal de audio, sus frecuencias, la tonalidad en la que se encuentra etc.; a ellos acudimos para extraer ideas sobre cómo realizar nuestro trabajo.

Algunos de los TFG seleccionados trataban temas como: detector automático de acordes (Segura Sogorb, 2015), diseño de un afinador musical (Chafchalaf Peña, 2013), extracción de datos tonales a partir del habla, transcripción musical, detector de escalas musicales para violín (Campos Salas, 2014), etc. Éstos dieron lugar a la actual propuesta: “Un detector de tonalidad de audio musical”.

En una segunda fase de desarrollo, una vez delimitado el campo de estudio, se elaboró el *data set* utilizado para probar el software de detección de la tonalidad de un archivo de audio específico, así como el desarrollo del sistema, definiendo sus alcances y limitaciones.

En los inicios de esta fase, se decidió ir programando distintos prototipos de manera que se pudiera ir probando diversas alternativas, con la finalidad de obtener una mejora del sistema con cada uno de ellos. Los prototipos se estructuran en dos partes: la primera de ellas tiene que ver con el análisis musical y la estimación de parámetros

de la Transformada de Fourier, siendo la segunda la que establece una estrecha relación con el análisis de la tonalidad.

La principal herramienta utilizada para la detección de la tonalidad de un fichero de audio fue MatLab, muy utilizada en el ámbito de las ingenierías, y que extrae, mediante algoritmos numéricos robustos, resultados validados y aceptados. Su fortaleza se debe a la capacidad que el programa tiene para resolver fácilmente problemas numéricos complejos de forma eficiente. Así hemos podido automatizar la extracción de características a partir de los ficheros de audio, generando la información en archivos de texto para su posterior estudio y análisis.

En una tercera fase de evaluación se estableció la fiabilidad del programa. Para ello se realizaron pruebas a medida que se iban desarrollando los prototipos con el objetivo de asegurar que dichos algoritmos estaban bien implementados.

Tras validar el algoritmo, se realizan las pruebas de eficacia del prototipo, para obtener como resultado el nivel de precisión que presenta el sistema. Dichas ejecuciones se realizaron con los mismos *data sets* para posteriormente realizar una comparación de todos los prototipos en la detección de la tonalidad.

Los resultados, además, han sido comparados con los resultados de *KeyFinder* (Sha'ath, 2011), un programa existente que permite la extracción de la tonalidad de un fichero de audio. A continuación, utilizando los resultados calculados a partir de los *data sets*, se procedió a realizar una clasificación de la eficacia de los prototipos desarrollados. Finalmente se ha realizado un resumen de las conclusiones tomadas en base a los resultados obtenidos en la elaboración del proyecto y se indican las posibles mejoras para los prototipos, analizando los problemas que han ido surgiendo a lo largo del desarrollo del trabajo.

La memoria de este trabajo se ha ido completando según se han obtenido los resultados requeridos en cada apartado y en las últimas semanas se ha unificado y finalizado por completo.

En el desarrollo de dicha investigación se han empleado alrededor de 350 horas de trabajo, según se detalla en la tabla siguiente.

Definición del proyecto	7 horas
Planificación	2 horas
Elaboración del índice	2 horas
Búsqueda de artículos de interés	30 horas
Análisis del problema	45 horas
Elaboración del *data set*	9 horas
Implementación de los algoritmos	86 horas
Estimación de parámetros	12 horas
Pruebas iniciales	15 horas
Pruebas de los algoritmos	50 horas
Desarrollo de la memoria	80 horas
Elaboración de la presentación	12 horas

Tabla 3. Estimación del tiempo empleado para el desarrollo del Trabajo Fin de Grado

5. IMPLEMENTACIÓN Y DESARROLLO

En este apartado del desarrollo del proyecto se detalla la construcción de la base de datos (*data set*) con la que se han evaluado los prototipos desarrollados. En el *data set* elaborado es necesario que exista información clasificada en dos grupos, por un lado el nombre del archivo de audio y por otro la tonalidad original de la obra musical.

Del mismo modo, se lleva a cabo una explicación de cómo son extraídas las características en las señales de audio pertenecientes al *data set* elaborado, utilizando la herramienta de software matemático Matlab. La explicación de cada una sigue un modelo desde el punto de vista matemático, cuyo objetivo es comprender el porqué se llevan a cabo ciertas operaciones. Así mismo, se muestran las pruebas realizadas para la obtención de los valores de los parámetros que hacen el software más

eficiente. Por último se expone el modelo de funcionamiento de cada prototipo desarrollado.

5.1 DATA SET

Para realizar las verificaciones experimentales paso a paso de cada prototipo, hemos elaborado una base de datos compuesta por un total de 24 pistas (una por cada tonalidad), cada una de 30 segundos de duración, una frecuencia de muestreo de 22050 Hz, Mono (un solo canal), 16-bits (tamaño de la muestra) y en formato.wav.

Los distintos prototipos están preparados para leer la información sobre los archivos que deben procesar de un fichero de texto. Este fichero (con extensión .txt) tiene, en cada línea, cuatro datos: el nombre del fichero que contiene la pista de audio, el nombre de la tonalidad, su modo (mayor o menor) y el coeficiente que identifica a esa tonalidad.

Los géneros musicales escogidos fueron el jazz, el blues y el rock, estilos musicales relacionados entre sí debido al hecho de que se originaron en el sur de Estados Unidos. El estilo del jazz se encuentra marcado por el uso del saxofón, piano, corneta y trombón. En síntesis, el jazz es más instrumental en comparación con el blues, que es sobre todo vocal. La música rock se nutrió fuertemente del blues e incorporó influencias del jazz. El rock se centra en la guitarra eléctrica, aunque incluye una fusión entre jazz y blues ya que utiliza cantante, bajo, batería y, algunas veces, instrumentos de teclado como el órgano y el piano.

Este conjunto de datos se han seleccionado de un *data set* mayor obra de Tzanetakis (2001).

Desafortunadamente la base de datos fue recogida gradualmente y muy rápido en su investigación por lo que los archivos de audio no poseen títulos y lógicamente ningún permiso de copyright. Los archivos fueron recolectados en los años 2000-2001 de una variedad de fuentes incluyendo CDs personales, radio, grabaciones de micrófono, con el fin de representar una variedad de condiciones de grabación.

Fue utilizada para investigaciones como *la clasificación de géneros musicales de señales de audio*, desarrollada por el mismo G. Tzanetakis y P. Cook (2002).

Conforme a esto, decidimos aprovechar parte de esta base de datos, ya que ha sido utilizada y validada por otros investigadores por lo que constituye una referencia sólida con la que contrastar los resultados obtenidos mediante el programa que hemos desarrollado.

Por esa misma razón elaboramos otros dos *data sets* con la finalidad de que fueran los seleccionados a la hora de evaluar la precisión final del sistema, uno de ellos que solo contuviera archivos de audio del género blues y otro de ellos exclusivamente de música disco. Ambos *data sets* fueron construidos a partir del conjunto de datos de Tzanetakis (2001).

5.2 PRUEBAS PREVIAS Y MATLAB

En los inicios de dicho Trabajo Fin de Grado se ha realizado un proceso de estimación de los parámetros más adecuados en las funciones declaradas en el software de Matlab, para obtener las características que se desean extraer de las señales de audio de la forma más eficientemente posible. A continuación se muestran las pruebas realizadas:

Para someter a cualquier muestra de audio contenido en el *data set* elaborado al análisis y procesamiento digital empezaremos por convertir dicha señal al dominio temporal.

El comienzo es sencillo, el eje de tiempos se define mediante un vector con componentes a intervalos regulares, cuyo patrón sigue la estructura *"inicio:incremento:fin"*. La dimensión de éste espacio vectorial será (1,N), con una longitud de N muestras. Para simplificar la creación del vector tiempo, especificaremos una primera entrada que se corresponderá con el inicio de la muestra de audio, es decir, con el instante 0 de la pista. Presentará un incremento que vendrá dado por el periodo (Ts) de la propia señal, equivalente a la tasa de muestreo de la señal de audio o, dicho de otro modo, a la inversa de la frecuencia (Fs) a la que se

muestrea la señal. Para concluir, la última entrada de dicho vector vendrá dada por el instante final de la muestra analizada. Éste valor será el mismo para todas las muestras y se corresponderá con un valor de 30 segundos aproximadamente.

Se han elegido solo los 30 primeros segundos porque es al principio (y también al final) de una obra donde se presentan los motivos musicales en la tonalidad de referencia. Posteriormente, se suelen ir produciendo modulaciones a otras tonalidades para terminar en la tonalidad original.

Al leer cualquier archivo de audio mediante la función *wavread(),* el software específico Matlab devuelve los datos muestreados (amplitudes) de dicha señal en forma de matriz de dimensión (N-1,1) (siendo N la longitud del vector tiempo). Para elaborar el diagrama en dos dimensiones (tiempo, amplitud), la matriz que contiene las amplitudes debe tener dimensiones tales que una de ellas sea igual a la longitud del vector. Como en nuestro caso, el número de filas de la matriz (N-1) dista de uno de la longitud del vector (N), basta con representar el vector ignorando su primer término, de este modo la función que representa gráficamente la señal en el dominio del tiempo (t(2:end),A) traza cada columna de la matriz A frente al vector t.

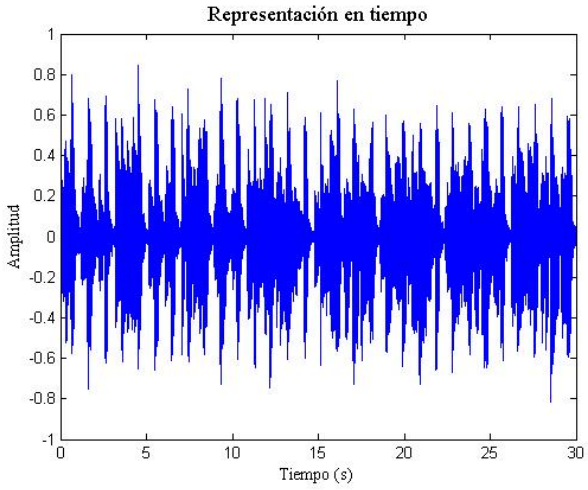


Figura 18. Análisis de la pista blues3.wav en el dominio temporal.

De esta forma, obtenemos una representación gráfica de la evolución en amplitud respecto al tiempo de la muestra de audio, que permite el estudio y caracterización de la estructura de la señal en tiempo.

Del mismo modo, cualquier muestra de sonido puede ser convertida al dominio frecuencial.

La función *fft()* que evalúa la Transformada Rápida de Fourier en Matlab resulta más eficiente cuando calcula la transformada de Fourier de una señal cuya longitud (número de muestras) es una potencia de 2 $(2,4,8,\ldots,2^n)$. Por este motivo, computaremos la potencia de 2 superiormás próxima al número de muestras que contiene el audio, ya que hacerlo puede acelerar el cálculo de la FFT cuando la longitud de la señal no es una potencia exacta de 2.

Por ejemplo, si el archivo tiene 6 muestras, el número que le pasaremos a la función FFT será 8, para que sea más eficiente y tenga mejor rendimiento. A este número lo llamaremos NFFT.

Si la longitud de la señal que utilizamos es mayor que NFFT, el programa la trunca para calcularla, y, por tanto, perdemos muestras de la señal, cosa que no queremos. Por el contrario, si es menor pero no es una potencia de dos, se añaden ceros hasta alcanzar la potencia de dos más cercana para su cálculo. Sin embargo, esto no afectará al resultado ya que las muestras que faltan son interpretadas como 0. De esta manera no añaden ningún contenido frecuencial al archivo original (Figura 19). El hecho de añadir más ceros, es decir, que el NFFT sea muy grande, aumenta el rango de frecuencias visualizadas, se suaviza el espectro, pero no aumenta la resolución, ya que ésta depende exclusivamente de la longitud de la ventana.

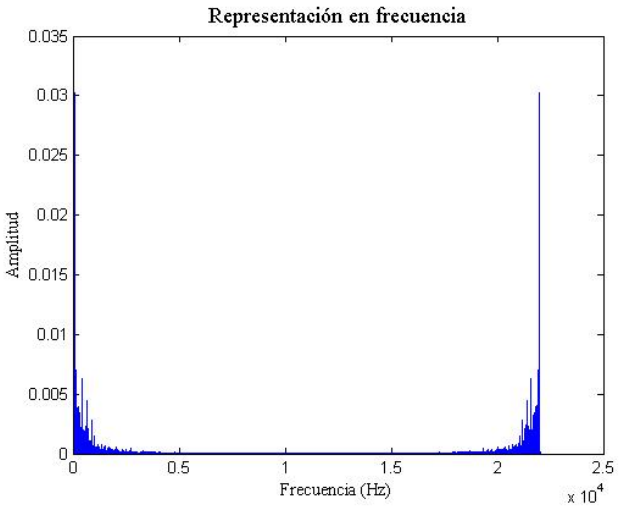


Figura 19. Análisis de la pista blues3.wav en el dominio frecuencial

Hay que considerar que la FFT entrega los resultados escalados por el número de total de muestras, así que para encontrar la verdadera magnitud de la frecuencia debemos dividir previamente entre el número total de muestras contenido en el archivo de audio para normalizar los valores.

Para finalizar, hay que construir el eje de coordenadas para posteriormente representar gráficamente los datos obtenidos. Las frecuencias que devuelve la FFT van desde 0 hasta la mitad de la frecuencia de muestreo (Fs/2). Por tanto, necesitamos generar un vector que vaya desde 0 hasta Fs/2. El número de elementos que este vector debe contener es NFFT/2 -1, ya que como podemos observar en la Figura 19, la señal se repite en el extremo opuesto debido a que la DFT genera una imagen espejo del espectro de frecuencia, por lo que solo queremos visualizar la mitad.

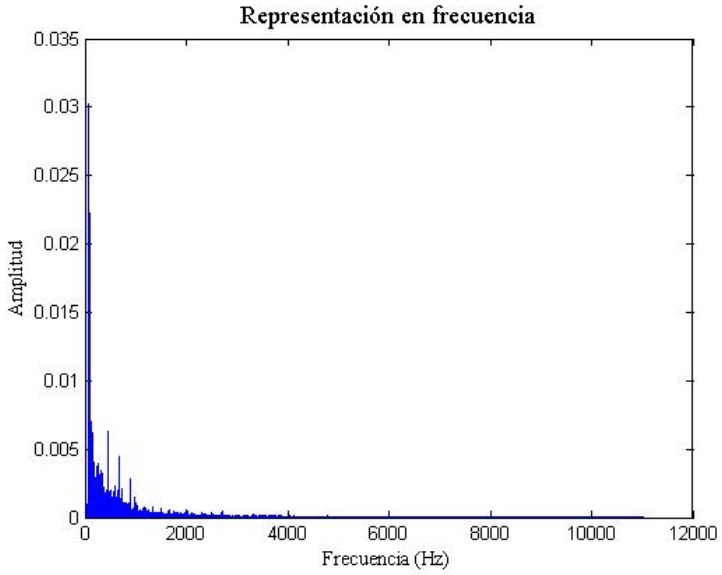


Figura 20. Análisis de la pista blues3.wav en el dominio frecuencial.

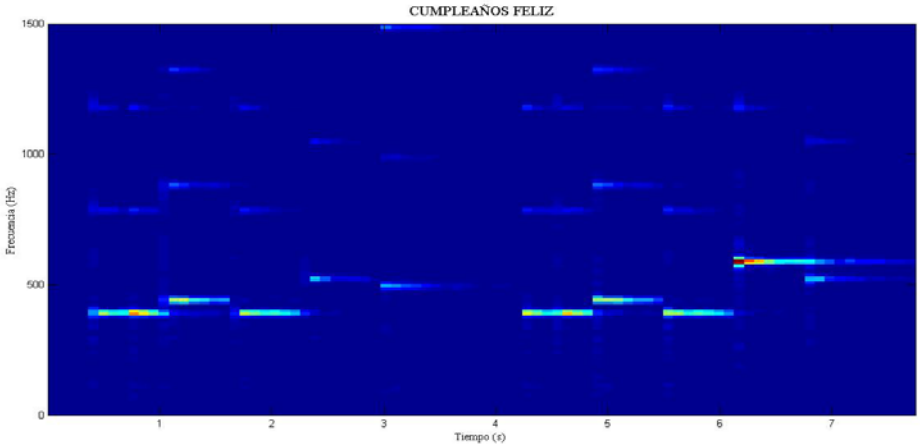
Como ya estudiamos en la parte teórica, uno de los problemas de la transformada de Fourier (FFT) es el régimen temporal, que no es capaz de distinguir en el instante de tiempo en el que ocurren los fenómenos aunque identifica perfectamente las frecuencias. Por esta razón utilizamos la aplicación de la STFT mediante el espectrograma.

En el software específico MatLab existe una función que calcula y representa gráficamente el espectrograma de una muestra de audio, *‹‹spectrogram()››*, cuyos parámetros de entrada incluyen la señal de audio, la frecuencia de muestreo (Fs) y el número de muestras sobre la que queremos realizar la Transformada Discreta de Fourier (NFFT), parámetro muy destacado por su influencia en el número de segmentos en los que se dividirá el espectro de señal muestreada, lo que determinará la resolución de la misma. Números muy grandes de NFFT aumentan el coste de procesamiento y el tamaño de los vectores de resultados, lo que a su vez aumenta la complejidad de cálculos posteriores.
Como parámetros de salida devuelve cuatro argumentos, [S,F,T,P]. Cada columna de S contiene una estimación del contenido de frecuencia a corto plazo de la señal x. El tiempo aumenta a través de las columnas de S, de izquierda a derecha. Así mismo, devuelve un vector de frecuencias cíclicas, F, expresado en términos de la frecuencia de muestreo, Fs, y un vector de instantes de tiempo, T, en el que se calcula el espectrograma. También devuelve una matriz, P, que contiene una estimación de la densidad espectral de potencia (PSD) o el espectro de potencia de cada segmento. Esta última será la matriz utilizada para la obtención del vector croma (más adelante), la cual evita realizar operaciones con números complejos a diferencia de la matriz de amplitud y fase.

Adicionalmente, tenemos como parámetro de entrada la tipología, la longitud de la ventana (WINDOW) como entero y el número de muestras superpuestas entre segmentos adyacentes (*I*=NOVERLAP). Como el parámetro WINDOW es escalar, la señal se divide en segmentos o tramas de longitud WINDOW, por lo que NOVERLAP debe ser menor que el número de muestras de la ventana. Si no se especifica, la ventana por defecto es del tipo Hamming.

Finalmente se obtiene un parámetro de salida que se corresponde con la matriz que da forma a la imagen que devuelve la función espectrograma. Si observamos la Figura 21, se muestra el análisis del sonido de una muestra de audio. Cada nota que compone la partitura de la canción deja una traza en el eje vertical del espectrograma. Las duraciones de cada nota se pueden analizar en el eje horizontal.

a)



b)

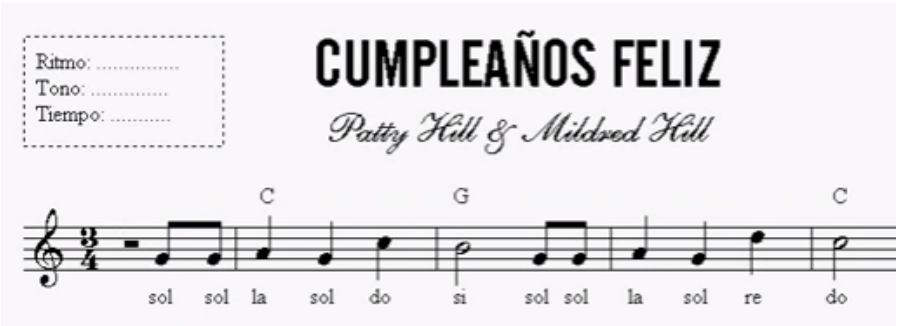


Figura 21. a) Espectrograma de un fragmento de la canción “Cumpleaños feliz” revela información sobre las notas que la componen. b) El pentagrama con la representación escrita de ese mismo fragmento

La representación del espectrograma presenta un problema, debido al principio de incertidumbre de Heisenberg. La resolución de este método depende del tamaño de la ventana, y como éste es constante, no se puede estimar con precisión el parámetro del tiempo y la frecuencia de forma simultánea.

El efecto de aumentar de tamaño la ventana de un espectrograma hace que el número de frecuencias que se estudian sea mayor. A costa de disminuir la distancia entre ellas obtenemos una mejor precisión. La razón matemática es que, cuanto más ancha sea la ventana, más estrecha resulta la función Sinc, que es la transformada de la ventana, produciendo una menor distorsión al espectro de la función original, como se puede apreciar en la Figura 22.

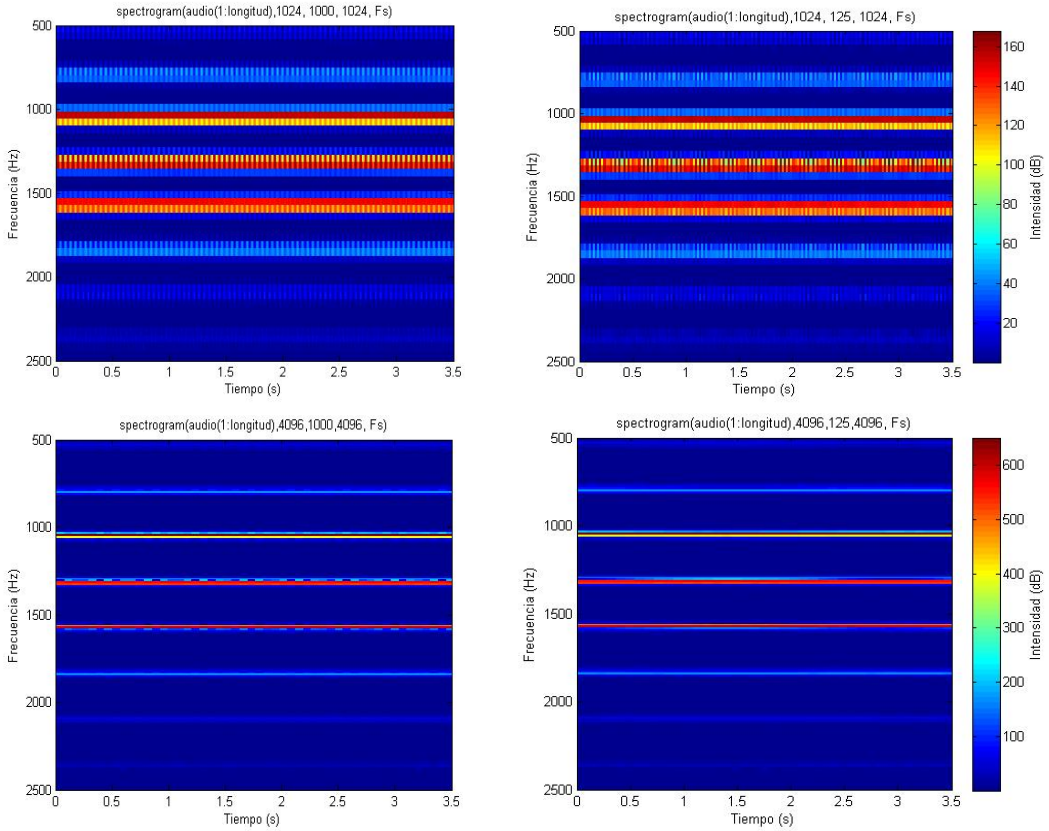


Figura 22. Espectrogramas del acorde Do Mi Sol con diferentes parámetros de entrada

Para suavizar el problema se utiliza un desplazamiento de la ventana menor que el tamaño de esta (NOVERLAP<WINDOW). De este modo se producen saltos menores y, consecuentemente, la resolución será más fina. Frecuentemente el desplazamiento viene definido en términos relativos al tamaño de la ventana. Hablamos por tanto de porcentaje de solapamiento, descrito por el porcentaje del número de muestras de la ventana que se desplaza por la muestra de audio analizada.

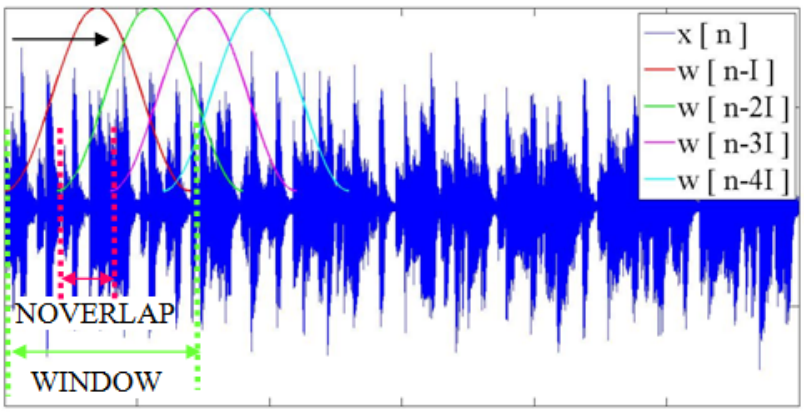


Figura 23. Avance de una ventana con el 65 % de solapamiento aproximadamente

Como las señales musicales que vamos a analizar presentan frecuencias muy juntas entre sí, lo que más nos interesa es utilizar un ancho de banda de la ventana grande ya que obtenemos una elevada resolución frecuencial, así como un solapamiento de muestras pequeño que suavice el espectro en tiempo

El hecho de aplicar una ventana a una señal, está fundamentado en realizar la convolución de la respuesta en frecuencia de la señal con la transformada de la ventana, luego observando los espectros de la Figura 16, concluimos que la ventana rectangular abarcará un rango de frecuencias mucho mayor que la Hamming pero el espectrograma, aún obteniendo mucha resolución, resulta más difícil de interpretar. Por el contrario, la ventana Hamming se centra solo en las frecuencias fundamentales y las aísla mucho más en el espectrograma, produciendo una menor distorsión.

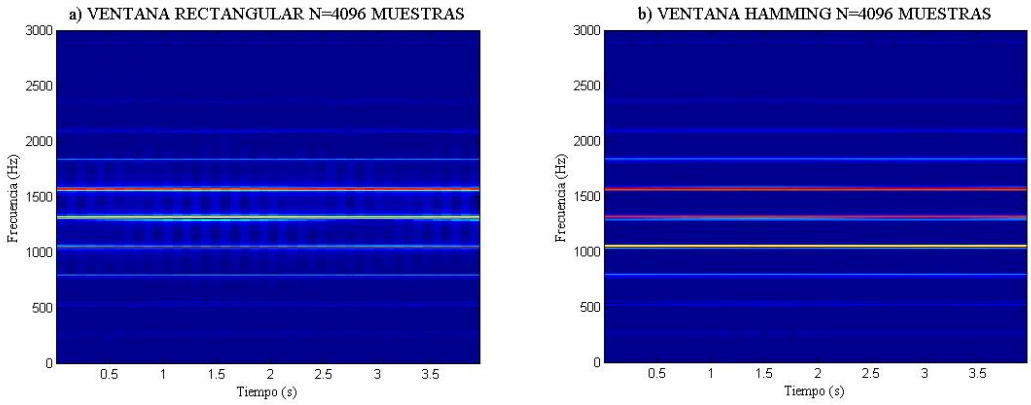


Figura 24. a) Espectrograma del acorde de Do Mayor con ventana rectangular. b) Espectrograma del acorde de Do Mayor con ventana Hamming. La muestra de audio analizada se trata de un sonido armónico, pues en el espectrograma se aprecia claramente un patrón de bandas en el eje vertical separadas aproximadamente por el mismo valor de frecuencias.

El software MatLab no dispone de ninguna función que calcule y represente el *chromagram* de una muestra de audio, por ello utilizamos la función encontrada en internet, obra de Ellis (2006). En ella se modificaron los valores de los parámetros que intervienen y que han sido explicados anteriormente.

Como ejemplo para poder entender de una forma más clara la información contenida en el cromagrama de una muestra de audio como segunda representación frecuencial,

utilizamos acordes, ya que el procedimiento para su reconocimiento se basa en el tipo de representación cromática. Como podemos observar en la Figura 20, aquellas clases de altura que componen cada acorde son la que disponen de mayor nivel de intensidad, representadas en color rojo, destacando sobre las demás.

a)

b)

c)

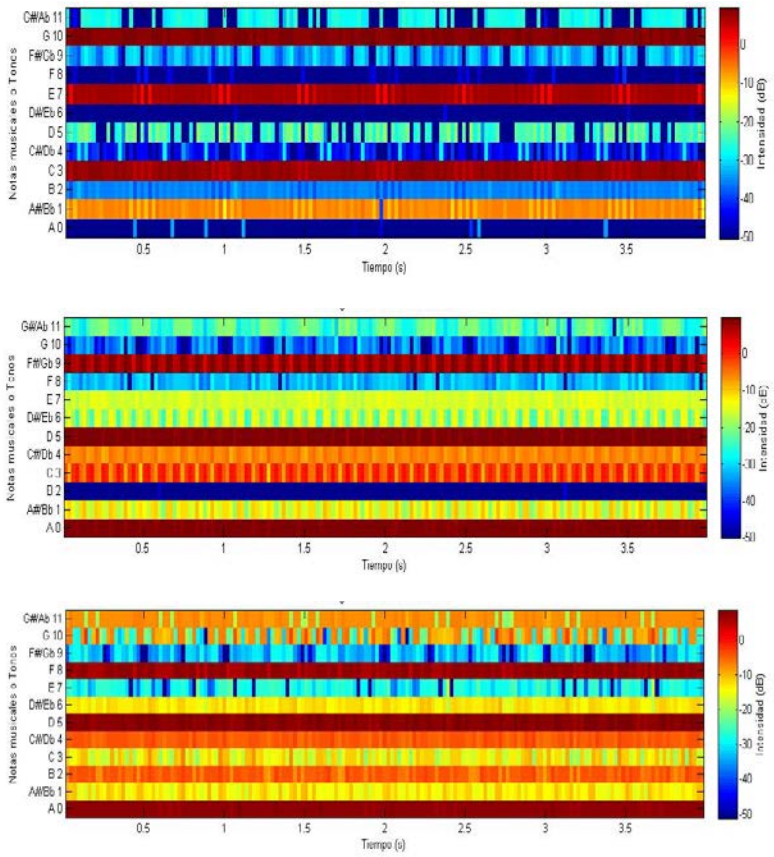


Figura 25. Matrices de croma o cromagramas, mostrando una serie de acordes por medio de la intensidad de la clase de altura o ‹‹*Pitch Class*›› en el instante de tiempo t. a) Acorde Mayor Do Mi Sol. b) Acorde Mayor Re Fa# La. c) Acorde Menor Re Fa La

5.3 PROTOTIPOS

La tonalidad de una obra música no viene definida por el acorde inicial ni el final, sino que viene determinada por las notas, las alteraciones que están sonando y las relaciones entre ellas.

Hay 24 posibles tonalidades que se indexan de acuerdo a un criterio común para facilitar su consulta y análisis como: Letra mayúscula, tonalidad mayor; letra minúscula tonalidad menor.

A	A#	B	C	C#	D	D#	E	F	F#	G	G#
0	1	2	3	4	5	6	7	8	9	10	11
a	a#	b	c	c#	d	d#	e	f	f#	g	g#
12	13	14	15	16	17	18	19	20	21	22	23

Tabla 4. Tonalidades

Los prototipos implementados se han construido como funciones en Matlab para procesar la información contenida en los *data sets* elaborados a partir de ficheros de textos que incluyen el nombre de la pista de audio y su tonalidad original.

Los resultados obtenidos tras la ejecución de cada prototipo se almacenan en un fichero de texto *.txt* que posteriormente será extraído en un documento Excel para su evaluación.

5.3.1 TRABAJO PREVIO A LOS PROTOTIPOS

Los archivos .wav que contienen audio y que se encuentran almacenados en los *data sets* elaborados se abren en el software Matlab y son decodificados, extrayendo la información contenida en ellos. A cada señal musical se le aplica un filtro predeterminado, para permitir una reducción en las frecuencias muestreadas.

Como la frecuencia de muestreo de los audios es de 22050 Hz, nos permite capturar frecuencias hasta el límite de la frecuencia de Nyquist (mitad de Fs) de 11025 Hz, aunque nos alejamos del límite de percepción humana, aún es considerado un valor elevado desde el punto de vista musical, consecuentemente aplicamos el filtro.

Se trata de un filtro *butterword*, diseñado para producir la respuesta más plana que sea posible hasta la frecuencia de corte, a partir de la cual la respuesta en frecuencia cae a razón de 6n dB por octava por debajo de la banda de paso, donde n es el orden

del filtro. Dicho con otras palabras, este tipo de filtros mantiene constante las amplitudes del espectro del audio analizado en la banda de frecuencias de paso.

Para un análisis de 8 octavas, el filtro que hemos implementado es de orden 10, paso-bajo con una frecuencia de corte de 4 kHz, ya que este tipo de filtros en realidad a lo que se dedica es a eliminar ruidos externos, es decir ayuda a evitar la contaminación de las frecuencias de los armónicos producidos por las notas fundamentales que componen el audio musical.

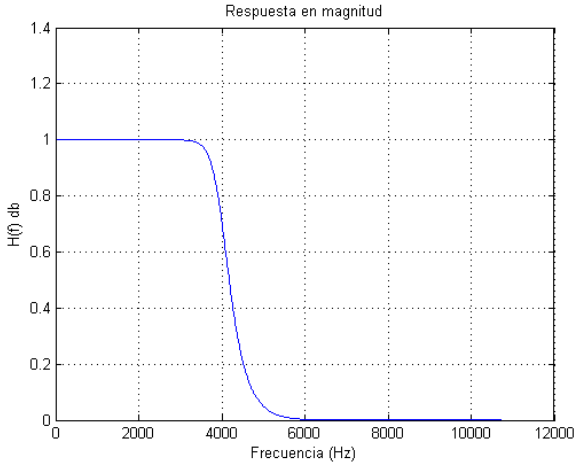


Figura 26. Respuesta en magnitud del filtro paso-bajo *butterword.* El eje vertical denota la magnitud (dB), y el eje horizontal denota la frecuencia (Hz).

El contenido espectral por encima de los 4 kHz se recorta, ya que los armónicos de la región de frecuencias altas no contribuyen a la detección de la tonalidad, ya que esta viene dada por las frecuencias bajas.

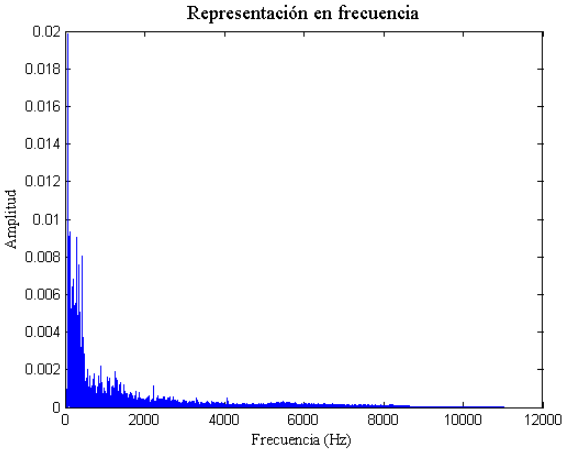


Figura 27. Espectro de frecuencias del archivo de audio original blues87.wav.

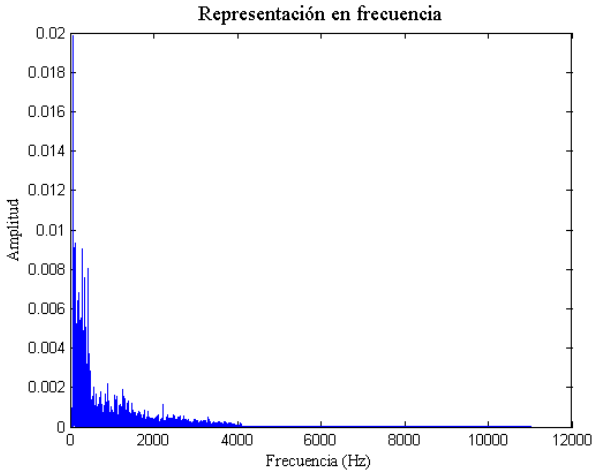


Figura 28. Espectro de frecuencias del archivo de audio blues87.wav tras pasar por el filtro predeterminado.

5.3.2 PROTOTIPO 1

En este primer método desarrollado nos centramos en la nota tónica, es decir aquella que se repite con mayor frecuencia.

Una idea simple como primera aproximación es la de almacenar todas las *chroma features* de la obra musical analizada en un único *chroma vector*. Esto se consigue sumando las características de *chroma* obtenidas en cada instante de tiempo en el que se define la señal de audio.

El procedimiento planteado se basa en obtener el espectro de la señal de audio analizada, a partir del cual sacaremos el vector de características de *chroma*. Contrastaremos los resultados con los dos modelos espectrales considerados más eficientes en el procesamiento digital de una señal de música explicados anteriormente: el espectrograma y *chromagram*.

El siguiente paso es localizar el valor máximo dentro del *chroma vector*, considerando el nombre de la nota que se corresponde con el valor máximo como el tono o tónica, el cual dará nombre a la tonalidad detectada por el algoritmo implementado.

Por último analizaremos el tercer grado respecto de la nota tónica para determinar si la tonalidad en la que se encuentra pertenece al modo mayor o al modo menor. Esto lo sabremos partiendo de los conocimientos previos adquiridos: si se encuentra en la modalidad mayor, el valor del vector croma perteneciente a la nota que dista dos tonos de la fundamental o tónica será mayor que la que dista de tono y medio de ésta. Si la tonalidad es menor, el valor del vector croma situado a tono y medio será mayor.

5.3.3 PROTOTIPO 2

Como una segunda aproximación un poco más compleja, conociendo la tónica mediante el procedimiento empleado en el prototipo 1, el siguiente paso es encontrar la escala diatónica incrustada en la obra musical analizada.

Para ello nos basamos en la idea de coincidencia de tono, a partir del cálculo del coeficiente de correlación entre las características de *chroma* y el modelo o estructura de tonos y semitonos de las escalas diatónicas. Por ejemplo, si encontramos que la nota tónica es La entonces generamos dos modelos, una para la clave $LaMayor\ (A)$ y otro para la clave $Lamenor\ (a)$.

$$y_{CMayorkey} = [1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 1]$$
$$y_{cmenorkey} = [1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0]$$

El primero de los índices se corresponde con la nota La, el segundo para la nota $La\#$,… y el último para la nota $Sol\#$.

La expresión que utilizamos para calcular el coeficiente de correlación es la siguiente:

$$R(x,y) = \frac{\sum_{k=1}^{12}(x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{12}(x_k - \bar{x})^2 \sum_{k=1}^{12}(y_k - \bar{y})^2}}$$

Donde x es el *chroma vector* y es el modelo de la estructura de tonos y semitonos una vez conocida la tónica. Para obtener los modelos correctos de escalas diatónicas,

deberemos desplazar los valores de los vectores $y_{CMayorkey}$ y $y_{cmenorkey}$ de forma circular el número de posiciones k, que se corresponde con el valor numérico o índice de la tónica detectada. Por ejemplo, si la tónica detectada es $Re\#$ (con índice 6), el número de posiciones que habría que desplazar ambos vectores sería $k = 6$:

$$y_{CMayorkey} = [0\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ ]$$
$$y_{cmenorkey} = [0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1]$$

Si el número de la nota tónica viene dado como $0 \leq k \leq 11$, nosotros solo tenemos que comparar los coeficientes de la correlación entre $R(x, y_{CMayorkey})$ (tonalidad mayor) y $R(x, y_{cmenorkey})$ (tonalidad menor). Si $R(x, y_{CMayorkey}) > R(x, y_{cmenorkey})$, decimos que la pieza musical está en la tonalidad Mayor k. De lo contrario, estará en la tonalidad menor $k + 12$.

5.3.4 PROTOTIPO 3

La nota tónica no tiene por qué ser la que presente más energía dentro de una pieza de música (la que más aparece).

Por tanto la tonalidad de una obra musical puede venir determinada por la presencia de la tercera o quinta (en su gran mayoría) del acorde de tónica. Debido a esto, se procede a construir un nuevo prototipo que se encarga de ver si la primera posición del *chroma vector* (ya ordenado de mayor a menor) la ocupa bien la fundamental, la tercera o la quinta respecto a la tónica.
Se presentan por tanto 3 posibles casos: en el primero de ellos, a priori, la tónica detectada es la correcta; en el segundo, la primera posición del chroma vector la ocupa la quinta respecto a la fundamental y, en el último de ellos, es la tercera la que ocupa la primera posición de dicho vector.

En el primer caso, a partir de la nota que presenta mayor suma de densidad espectral dentro del *chroma vector*, es decir, aquella que se encuentra en la primera posición del vector, se obtiene si entre la segunda y la tercera posición del chroma vector se encuentran o bien la tercera (mayor y/o menor) respecto a la fundamental, la quinta o

ambas. Si alguna de estas condiciones se cumple, suponemos que la tónica detectada es la correcta. En caso contrario, hay otras dos posibilidades.
La primera de ellas es que la quinta respecto a la fundamental pudiera encontrarse en la primera posición del *chroma vector*. En dicho caso, se obtiene si la fundamental (7 semitonos por debajo) respecto a la nota detectada aparece entre la segunda o tercera posición del chroma vector. En caso de acierto, la tónica detectada será aquella nota que se corresponde con un intervalo de siete semitonos por debajo de la que ocupaba la primera posición del *chroma vector*.

Si esta condición no se cumple, pasamos a la última posibilidad, en la que la tercera respecto a la fundamental puede encontrarse en la primera posición del *chroma vector*. En este caso realizaremos el mismo procedimiento descrito anteriormente pero ahora observando si entre la segunda o tercera posición de dicho vector se encuentra la fundamental (3 o 4 semitonos por debajo) respecto a esta. En el caso de que se encuentre la nota separada a un intervalo de tercera menor por debajo de la que ocupaba la primera posición del vector, pasará a ser ésta la nota tónica detectada; por el contrario, si la nota encontrada es aquella separada por un intervalo de tercera mayor por debajo, ésta será la tónica detectada por el prototipo.

5.3.5 PROTOTIPO 4

Una vez expuesta la idea de los prototipos anteriormente implementados nos disponemos a realizar una mezcla de dos de ellos para ver si se puede establecer una mejora del software detector de tonalidad.
Como cuarto prototipo vamos a aprovechar dos de las herramientas utilizadas y explicadas en prototipos anteriores. Para ello en este prototipo obtendremos la tónica a partir del método implementado en el prototipo anterior y calcularemos la escala diatónica que le pertenece a partir del cálculo del coeficiente de correlación, método llevado a cabo en el prototipo 2.

5.3.6 PROTOTIPO 5

En una obra musical se van produciendo modulaciones de la tonalidad a lo largo del tiempo. Así, aunque la obra tenga una tonalidad determinada, ciertos pasajes pueden estar en una tonalidad diferente.

Por este motivo, intentar detectar la tonalidad de toda la pieza a partir de un único análisis del espectro de toda la señal, puede no ser adecuado. Por ello como una quinta aproximación, vamos a segmentar en trozos el espectro para detectar la tonalidad de cada trozo y posteriormente ver si alguna de las tonalidades detectadas se repite más.

El proceso de segmentación del audio consiste simplemente en dividir el espectro en N segmentos de igual longitud a lo largo del tiempo. Posteriormente cada división será analizada mediante el procedimiento explicado en el prototipo 3.

La tonalidad detectada en cada uno de los segmentos será almacenada en un vector, del cual finalmente se calculará la moda para obtener aquella tonalidad que más se repite.

5.3.7 PROTOTIPO 6

Como una última aproximación pensamos en realizar un nuevo prototipo que pudiera obtener la tonalidad del audio detectado a partir del cálculo del coeficiente de correlación del audio procesado a partir de todas y cada una de las doce notas, y por tanto de las 24 tonalidades posibles.

Una vez obtenido el resultado para cada tonalidad, el valor obtenido del cálculo del coeficiente de correlación es almacenado en una matriz de dimensión 2x12, una posición para cada tonalidad.

Una vez completada la matriz se obtiene qué tonalidad presenta mayor coeficiente de correlación. Debemos tener en cuenta que este valor estará repetido, ya que la

disposición de los valores de los vectores $y_{CMayorkey}$ e $y_{cmenorkey}$ son iguales, y estos se corresponden con las tonalidades relativas. Ejemplo: Do Mayor y La menor.

$$y_{Do\ Mayor} = [0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ ]$$
$$y_{La\ menor} = [0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 1]$$

6. RESULTADOS Y DISCUSIÓN

El método de evaluación de los prototipos desarrollados es el utilizado por el congreso MIREX (Stephen Downie, 2017) para la evaluación de los programas de detección de tonalidad. Este método consiste en asignar una puntuación a la calidad de la detección obtenida de cada muestra de audio, sumar estos valores y hallar el promedio para obtener la precisión del sistema en porcentaje.

El análisis de errores se centra en una comparación de la tonalidad detectada por el software implementado con la tonalidad original de la obra musical almacenada en el *data set.* Este procedimiento determina cómo de próxima ha sido la detección realizada. Una detección se considera cercana si mantiene las siguientes relaciones con respecto a la tonalidad original: dista un intervalo de quinta perfecta, es el relativo mayor/menor o bien confunde el modo (mayor por menor o viceversa). A la detección realizada de forma correcta se le asigna un punto, mientras que a las detecciones incorrectas se les asignan fracciones de un punto de acuerdo con la siguiente tabla:

Relación con la tonalidad correcta	**Puntos**
Misma	1
Quinta perfecta	0.5
Relativo mayor/menor	0.3
Paralelo mayor/menor	0.2
Otros	0

Tabla 5. Método de evaluación para la detección de la Tonalidad en MIREX, usado como referencia para la evaluación de nuestros métodos.

Para la obtención de los resultados finales, tras la implementación de los prototipos, únicamente hemos utilizado *data sets* de dos géneros musicales: el blues y la música disco, dos estilos que presentan características musicales muy distintas.

A continuación se presentan los resultados obtenidos en cada prototipo, extrayendo las características de cada obra musical a partir del espectrograma y del *chromagram*, para continuar con una comparativa de los resultados.

Prototipo 1

El prototipo 1 trata de realizar la detección de la tónica extraída como aquella nota con mayor densidad, calculando su modo a partir del intervalo de tercera respecto a ésta que presente mayor densidad

Relación con la tonalidad correcta	**PROTOTIPO 1**
Misma	22
Quinta Perfecta	13
Relativo mayor/menor	2
Paralelo mayor/menor	19
Otros	42
Porcentaje de aciertos	33.57 %

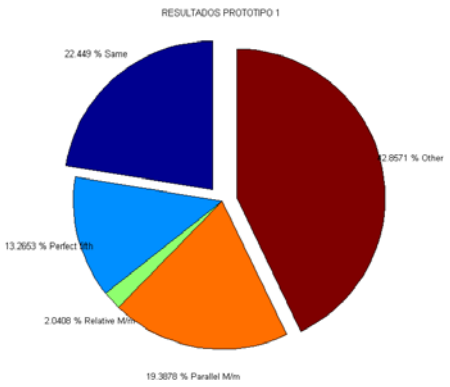


Figura 29. Resultados obtenidos del prototipo 1 con el *data set* de música blues y mediante el espectrograma

Aunque solo se reconocen correctamente un 22.4% de las tonalidades, se supera el 50% si sumamos el reconocimiento de tonalidades relacionadas con la real.

Prototipo 2

Como una segunda aproximación se ha implementado una pequeña modificación sobre el prototipo anterior: el modo (mayor o menor) se selecciona según el que presente un mayor coeficiente de correlación calculado a partir de la tónica seleccionada como en el prototipo anterior.

Relación con la tonalidad correcta	PROTOTIPO 2
Misma	22
Quinta Perfecta	13
Relativo mayor/menor	3
Paralelo mayor/menor	19
Otros	41
Porcentaje de aciertos	33.88 %

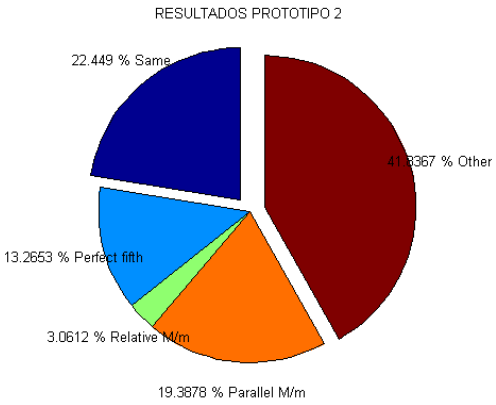


Figura 30. Resultados obtenidos del prototipo 2 con el *data set* de música blues y mediante el espectrograma

Aunque el nivel de aciertos en dicho caso se corresponde con el del prototipo anterior, se puede observar que la precisión del sistema aumenta, esto se debe al aumento del reconocimiento de tonalidades cercanas a la original.

Prototipo 3

Como un tercer método posible, y a diferencia del primer prototipo implementado, en este caso es la tónica la que se calcula de forma diferente. Dicho algoritmo busca si la nota detectada con mayor densidad pudiera ser la tercera o quinta respecto a la tónica que queremos conseguir.

Relación con la tonalidad correcta	PROTOTIPO 3
Misma	23
Quinta Perfecta	10
Relativo mayor/menor	3
Paralelo mayor/menor	28
Otros	34
Porcentaje de aciertos	35.2 %

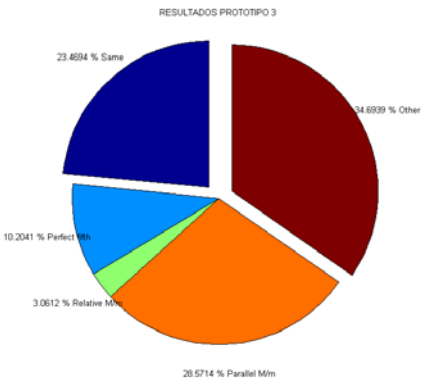


Figura 31. Resultados obtenidos del prototipo 3 con el *data set* de música blues y mediante el espectrograma.

En algunos casos este prototipo ya presenta una mayor precisión del sistema, con menor porcentaje de errores detectados. Esto puede ser debido a que se analizan fragmentos intermedios de las obras musicales en los que aparece con menor frecuencia la tónica.

Prototipo 4

El cuarto prototipo aprovecha el método de cálculo de la nota tónica del prototipo anterior, sin embargo el modo lo obtiene a partir del coeficiente de correlación desarrollado en el segundo prototipo.

Relación con la tonalidad correcta	**PROTOTIPO 4**
Misma	22
Quinta Perfecta	13
Relativo mayor/menor	3
Paralelo mayor/menor	19
Otros	41
Porcentaje de aciertos	33.88 %

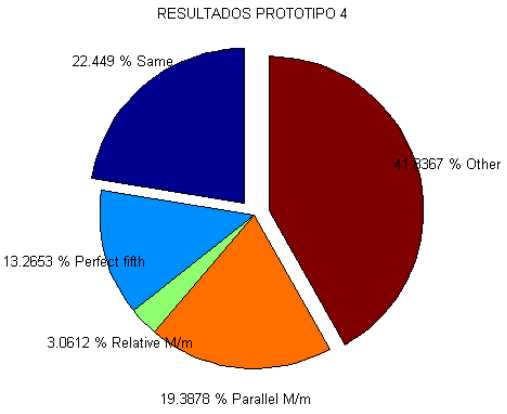


Figura 32. Resultados obtenidos del prototipo 4 con el *data set* de música blues y mediante el espectrograma.

En una valoración global de la precisión del sistema respecto a los resultados obtenidos, podríamos deducir que se trata del prototipo más eficiente hasta el momento. En dicho caso presenta los mismos resultados que el prototipo 2 pero esto se debe a que se está comprobando la eficiencia del sistema a partir del análisis de archivos musicales difíciles como son los pertenecientes al género blues.

Prototipo 5

En el quinto prototipo en cuestión, se ha probado realizar una segmentación de las dos herramientas utilizadas para el análisis de las señales musicales, el

espectrograma y el *chromagram*, a partir de cada uno de los segmentos se aplica el tercer prototipo y se obtiene la tonalidad de moda entre las detectadas.

Relación con la tonalidad correcta	**PROTOTIPO 5**
Misma	23
Quinta Perfecta	7
Relativo mayor/menor	5
Paralelo mayor/menor	25
Otros	38
Porcentaje de aciertos	33.67 %

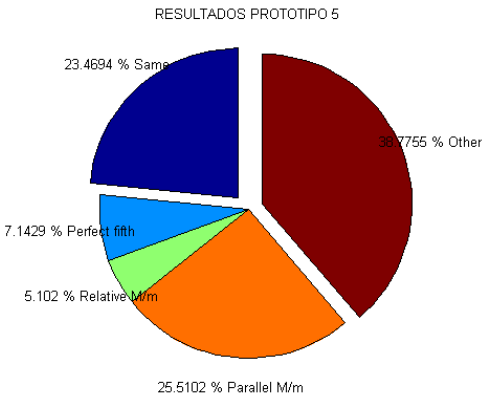


Figura 33. Resultados obtenidos del prototipo 5 con el *data set* de música blues y mediante el espectrograma.

Este método se ha pensado con la idea de que pudiera ser más eficiente teniendo en cuenta las modulaciones que pueden surgir a lo largo de cualquier obra musical, sin embargo los resultados no han sido nada óptimos.

Prototipo 6

Como última aproximación, el prototipo 6 detecta la tonalidad de obra calculando el coeficiente de correlación para cada una de las 24 tonalidades posibles.

Relación con la tonalidad correcta	**PROTOTIPO 6**
Misma	2
Quinta Perfecta	21
Relativo mayor/menor	2
Paralelo mayor/menor	2
Otros	71
Porcentaje de aciertos	13.78 %

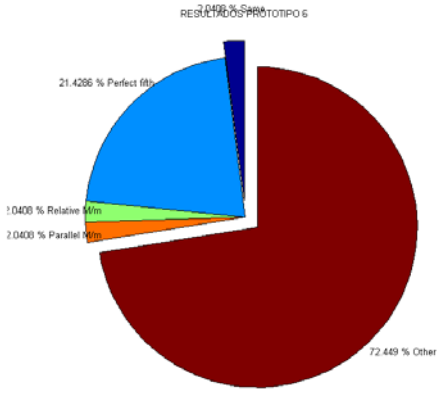


Figura 34. Resultados obtenidos del prototipo 6 con el *data set* de música blues y mediante el espectrograma.

Hasta que no lo desarrollamos y comprobamos el funcionamiento del algoritmo, no nos dimos cuenta que el mayor valor de ellos siempre se presenta tanto en el relativo mayor como en el menor, pues la disposición de los vectores $y_{CMayorkey}$ e $y_{cmenorkey}$ es la misma. A pesar de esto si resultara ser un método más eficiente que los anteriores, debería presentar una precisión del sistema aun mayor, ya que aunque la tonalidad detectada no fuera la correcta, la que la complementaría sería su respectivo relativo cuya precisión se pone en una puntuación de 0.3 puntos.

KeyFinder

Una vez introducidos nuestro data set de blues utilizado para la evaluación del sistema en el software libre *KeyFinder* (Sha'ath, 2011) obtuvimos los siguientes resultados.

Relación con la tonalidad correcta	***KeyFinder***
Misma	44
Quinta Perfecta	4
Relativo mayor/menor	0
Paralelo mayor/menor	22
Otros	28
Porcentaje de aciertos	51.43 %

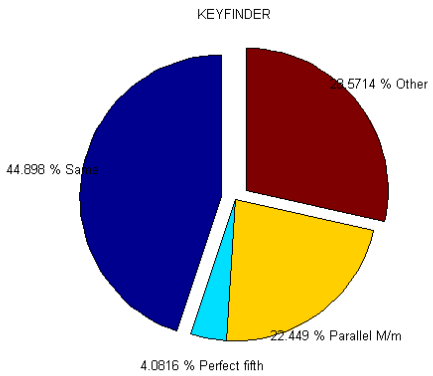


Figura 35. Resultados obtenidos del software *KeyFinder*

Podemos observar que la precisión de *KeyFinder* para el caso del *data set* del género blues es mucho mayor que nuestro software desarrollado. Presenta casi el doble de detecciones acertadas y aproximadamente el mismo número de errores para el mejor de los casos. Sin embargo podemos observar que no se trata de un sistema lo suficientemente eficiente (Sha'ath, 2011) ya que si analizamos los aciertos por separado no se supera ni el 50 % de ellos, por lo que dicho método se podría mejorar aún más.

Resumen

A continuación se resumen los resultados obtenidos extrayendo las características de cada obra musical a partir del espectrograma y seguidamente mediante el *chromagram* de cada *data set* mencionado anteriormente:

Relación con la tonalidad correcta	**P 1**	**P 2**	**P 3**	**P 4**	**P 5**	**P 6**	***KeyFinder***
Misma	22	22	23	22	23	2	44
Quinta Perfecta	13	13	10	13	7	21	4
Relativo mayor/menor	2	3	3	3	5	2	0
Paralelo mayor/menor	19	19	28	19	25	2	22
Otros	42	41	34	41	38	71	28
Precisión del sistema	33.57 %	33.88 %	35.2 %	33.88 %	33.67 %	13.78 %	51.43 %

Tabla 6. Resultados de la detección del *data set* de música blues mediante el espectrograma

Relación con la tonalidad correcta	**P 1**	**P 2**	**P 3**	**P 4**	**P 5**	**P 6**	***KeyFinder***
Misma	7	12	8	15	2	1	44
Quinta Perfecta	29	25	16	15	2	12	4
Relativo mayor/menor	0	0	2	2	12	2	0
Paralelo mayor/menor	20	15	35	28	23	3	22
Otros	42	46	37	38	59	80	28
Precisión del sistema	26.02 %	28.06 %	24.08 %	29.29 %	11.43 %	8.37 %	51.43 %

Tabla 7. Resultados de la detección del *data set* de música blues mediante el *chromagram*

Relación con la tonalidad correcta	**P 1**	**P 2**	**P 3**	**P 4**	**P 5**	**P 6**	***KeyFinder***
Misma	3	4	2	4	3	3	8
Quinta Perfecta	3	3	3	3	2	2	4
Relativo mayor/menor	1	1	1	1	1	2	3
Paralelo mayor/menor	3	2	4	2	3	0	2
Otros	11	11	11	11	12	14	4
Precisión del sistema	25.71 %	29.52 %	21.9 %	29.52 %	23.33	21.9 %	53.81 %

Tabla 8. Resultados de la detección del *data set* de música disco mediante el espectrograma

Relación con la tonalidad correcta	**P 1**	**P 2**	**P 3**	**P 4**	**P 5**	**P 6**	***KeyFinder***
Misma	6	6	8	8	2	5	8
Quinta Perfecta	7	8	6	7	4	5	4
Relativo mayor/menor	1	1	1	1	1	5	3
Paralelo mayor/menor	0	0	0	0	2	1	2
Otros	7	6	6	5	12	5	4
Precisión del sistema	46.67 %	49.05 %	53.81 %	56.19 %	22.38 %	43.81 %	53.81 %

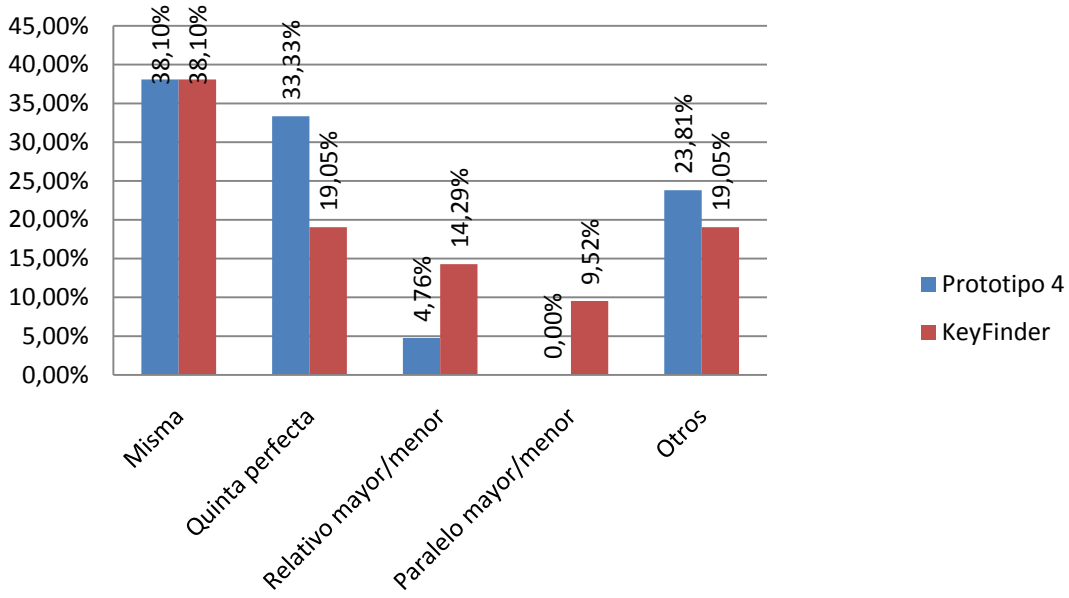
Tabla 9. Resultados de la detección del *data set* de música disco mediante el *chromagram*

En general, los resultados que hemos obtenido en nuestros prototipos son peores que los del software *KeyFinder*, pero no en todos los casos.

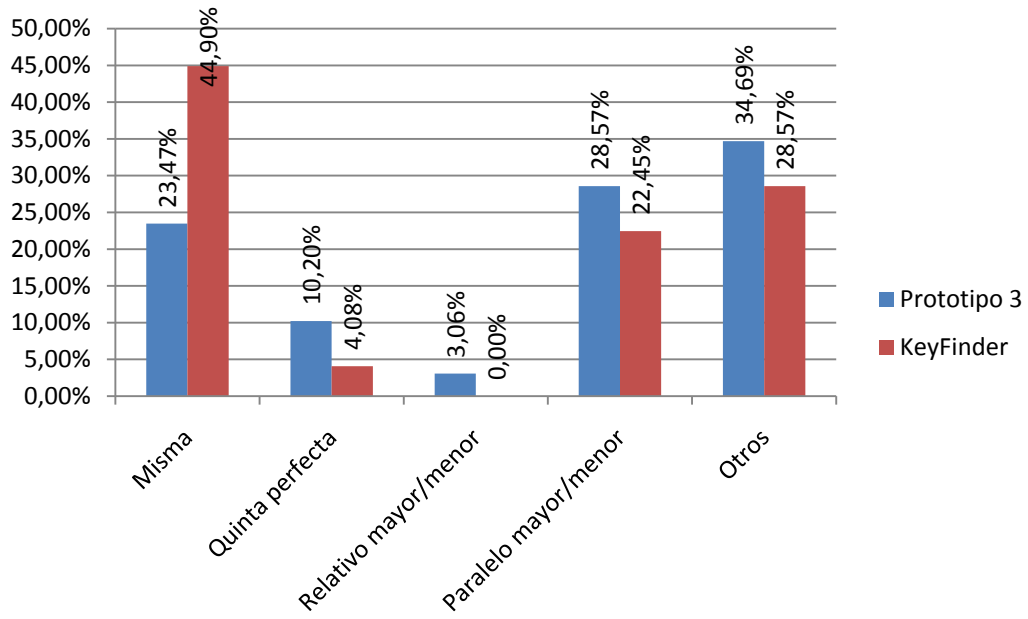
Para el *data set* de música blues se obtienen mejores resultados en el caso de la extracción de características realizada por el espectrograma, mientras que en los

archivos de audio de música disco los resultados mejores se obtienen mediante la extracción de sus características por medio del *chromagram.*

A la vista de los resultados, si nos fijamos en el porcentaje de aciertos entre nuestro software implementado y el que ya disponemos (*KeyFinder*) podemos observar buenas estimaciones para archivos musicales perteneciente al género disco, incluso llegando la precisión a superar en un 2.38 % el caso del prototipo 4 implementado a partir del *chroma vector* extraído del *chromagram.*



Por el contrario la utilización de archivos musicales del estilo blues dificultan notablemente la detección. Esto es debido a que la música blues es considerada como música improvisada en la que continuamente se expone a cambios de tonalidad, por lo que resulta difícil determinar cuál es la que predomina. En este caso nos encontramos con los mejores resultados tras implementar el prototipo 3 extrayendo las características a partir del espectrograma aunque, como dijimos anteriormente, respecto al *KeyFinder* los resultados obtenidos son mucho peores, reduciendo su porcentaje de aciertos en un 16.23 %.



Si solo nos fijamos en el número de aciertos, es decir, aquellas tonalidades detectadas que se corresponden con la tonalidad original del archivo de audio, en *KeyFinder*el valor de aciertos mediante la base de datos de blues está en 44, mientras que entre nuestros prototipos el valor más alto de aciertos es 23, significativamente inferior al obtenido en *KeyFinder*. En el caso de música disco el valor máximo de aciertos coinciden entre ambos software.

7. CONCLUSIONES

Este apartado final reúne las conclusiones tomadas tras el desarrollo del Trabajo Fin de grado y presenta las oportunidades que surgen para seguir trabajando en el problema que abarca el tema en cuestión.

El resultado principal del Trabajo Fin de Grado es el software desarrollado, que muestra cierto éxito en la detección de tonalidades de música disco, presentando resultados significativamente más eficientes que para otros géneros musicales como el blues.

El aspecto menos satisfactorio del trabajo es el tamaño del conjunto de datos. Hubiese sido un punto a favor poder disponer de un *data set* de música clásica, ya que se trata de un estilo musical en el que la tonalidad tiene gran importancia, a diferencia de otros géneros como el blues o el jazz en el que la improvisación desempeña un papel mayor y las tonalidades de las obras están menos marcadas.

En términos referidos a la recuperación de información musical, este trabajo investiga una serie de alternativas eficientes y flexibles como son los espectros de señales de audio, utilizados muy frecuentemente en otras herramientas de análisis y procesado de señales musicales. Los resultados muestran que tanto los métodos espectrales como los de *chroma* pueden utilizarse para la detección de la tonalidad.

Los prototipos desarrollados son aptos para ejecutarse en ordenadores sin requerimientos especiales de hardware y no necesitan una gran potencia de cálculo. Las versiones actuales funcionan aceptablemente en un ordenador portátil normal.

Tras las pruebas realizadas a lo largo del desarrollo del trabajo descrito en los capítulos precedentes hemos podido recopilar las conclusiones e ideas más destacadas.

El estudio de la tonalidad de un archivo de audio es un tema altamente complejo, atacado desde el punto de vista de la física del sonido. De los parámetros sonoros, altura, duración, intensidad y timbre, solo la altura del sonido resulta determinante para el ordenamiento dentro del sistema tonal.

En determinadas muestras de audio encontramos elementos que no hagan sonar ninguna nota de la escala de la tonalidad (por ejemplo, los instrumentos de percusión o efectos sonoros añadidos a la obra musical). Todos estos elementos originan alta energía en frecuencias que no guardan relación con las notas que componen la escala que da forma a la tonalidad de una canción por ello podemos prescindir de la información que aportan en altas frecuencias mediante la aplicación de un filtro.

Como hemos observado el análisis de señales mediante STFT está limitado a un campo de resolución y según la señal que queramos estudiar debemos utilizar un tipo

de ventana u otro en cada caso, es decir es importante saber de inicio algo de la señal a estudiar, ya que si sabemos con certeza que la señal que estamos trabajando posee componentes de frecuencias bien separadas entre sí, podríamos sacrificar resolución frecuencial, y dar una mejor resolución temporal. Pero si la señal a utilizar no posee estas características podemos tener problemas a la hora de estudiar dicha señal, y por lo tanto el método de análisis por STFT queda limitado en este aspecto, por lo tanto podemos concluir con que la STFT posee algunas problemas o limitaciones a la hora de estudiar la señal.

El desarrollo de distintos prototipos por separado permite la mejora independiente de cada uno de ellos con el fin de alcanzar un reconocimiento más eficaz.

Los muchos parámetros que controlan los algoritmos o prototipos desarrollados son personalizables por el usuario. Esto puede permitir la adaptación a determinados géneros musicales, ya que como hemos visto en cierta medida el conjunto de datos de música blues presenta en su gran mayoría errores en la detección según los parámetros que fueron escogidos.

La detección automática de la tonalidad de un audio es una herramienta de enorme utilidad sobre todo para la posible clasificación de música por las emociones creadas, ya que la tonalidad en la que está escrita una composición influye muy notablemente en el color tímbrico, y por tanto, en el clima, atmósfera o sensación que se quiera causar. Así mismo la importancia que supone sobre todo para aquellos músicos o DJs el conocimiento de las tonalidades de las canciones que emplean para mezclar música sin producir sonidos desagradables.

En determinadas ocasiones puede haber complicación en la detección de la tonalidad de una obra musical debido a circunstancias que se pueden dar como los cambios de tonalidad durante el desarrollo de una obra, que ésta esté escrita en una tonalidad complicada por el número de alteraciones que engloba, que el archivo de audio no tenga la calidad deseada, que el fragmento analizado sea intermedio a la obra musical y por tanto no aparezcan lo suficiente las notas que dan nombre a la tonalidad de ésta, que las señales no contengan suficientes muestras para que se produzca un entrenamiento deseado etc.

Las tasas de aciertos han sido ligeramente inferiores a las del software *KeyFinder*. Desafortunadamente, estos algoritmos no han mejorado significativamente los resultados generales de *KeyFinder*.

Como posibles mejoras y trabajos futuros enfocados en el desarrollo de un sistema como el implementado:

- Se puede recurrir a la utilización de otras herramientas de análisis que no se traten de la Transformada de Fourier, como por ejemplo la Transformada Wavelet u otras técnicas a estudiar.
- La utilización de banderas para denotar las frecuencias, es decir colocar las notas en 0 o 1 según estas cumplan un valor de amplitud considerable en el muestreo del dominio espectral.
- El desarrollo de una herramienta gráfica que nos permita seleccionar el valor de los parámetros que intervienen en cada prototipo implementado, ya que de estos depende la precisión de la detección para cada *data set* utilizado.
- La eliminación de información redundante en las señales de audio analizadas como los posibles ruidos o batidos producidos que no se aprecian auditivamente pero que sin embargo se muestran en las representaciones espectrales.
- La elaboración de prototipos aplicando técnicas más complejas como por ejemplo a partir del aprendizaje automático empleando redes neuronales capaces de clasificar la tonalidad de la señal de audio, etc.

Finalmente podemos concluir con que nos encontramos en un campo de trabajo muy abierto donde se pueden aplicar herramientas muy diversas para mejorar los resultados obtenidos.

8. REFERENCIAS

AL-MAJDALAWI ÁLVAREZ, A. 2006. *"Las escalas musicales. El origen de la escala musical"*. Universidad de Valladolid, 2006 [Consulta: 2 Septiembre 2017]. Disponible en: https://www.lpi.tel.uva.es/~nacho/docencia/ing_ond_1/trabajos_05_06/io2/public_html/index.html

ANCHUELA ARNALTE, C. 2013. *Armonizador para instrumento musical*. EET, Terrassa, 20 Diciembre 2013 [Consulta: 30 Agosto 2017]. Disponible en: http://upcommons.upc.edu/handle/2117/88201

BASSO, G. 2001. *Análisis Espectral: ‹‹ La Transformada de Fourier en la música ››*.Ediciones Al Margen. Editorial de la Universidad Nacional de La Plata (ed), Buenos Aires, 2ª edición, Mayo 2001. ISBN: 950-34-0150-X.

BERNAL, J; GÓMEZ, P.; BOBADILLA, J. *Una visión práctica del uso de la Transformada de Fourier como herramienta para el análisis espectral de la voz.* Universidad Politécnica de Madrid [Consulta: 25 Septiembre 2017]. Disponible en: http://stel.ub.edu/labfon/sites/default/files/EFE-X-JBernal_PGomez_JBobadilla-FFT_una_vision_practica_herramienta_para_el_analisis_espectral_de_la_voz.pdf

CAMPOS SALAS, M. 2014. *Algoritmo de detección de escalas musicales para violín.* Universidad de Costa Rica, Facultad de Ingeniería (Escuela de Ingeniería eléctrica). Ciudad universitaria "Rodrigo Facio", Diciembre 2014 [Consulta: 2 Agosto 2017]. Disponible en: http://eie.ucr.ac.cr/uploads/file/proybach/pb0894t.pdf

CASADO GARCÍA, M.E. 2011.*Acústica en el dominio de la frecuencia.* Escuela de Ingenierías. Edificio Tecnológico, Campus de Vegazana, s/n 24071. León España, 13 Noviembre 2011. [Consulta: 27 Julio 2017]. Disponible en: http://mecg.es/archivos/La%20ac%C3%BAstica%20en%20el%20dominio%20de%20la%20frecuencia.pdf

CASTRO, A.M. 2009. *Música y matemáticas: la afinación temperada.* Blog Enchufa 2, 12 Agosto 2009 [Consulta: 10 Octubre 2017]. Disponibles en: https://www.enchufa2.es/archives/musica-y-matematicas-la-afinacion-temperada.html

CHAFCHALAF PEÑA, D.A. 2013. *Diseño de un afinador musical digital por medio de MatLab*. Guatemala, Junio 2013 [Consulta: 30 agosto 2017]. Disponible en: http://biblioteca.usac.edu.gt/tesis/08/08_0342_EO.pdf

COLABORADORES DE MUSIKI. *Musiki: Altura tonal, Hélice de Shepard.* Musiki, repositorio de información musical de libre acceso, 2 Octubre 2017 [Consulta: 8 Octubre 2017]. Disponible en: http://musiki.org.ar/Altura_tonal

COLABORADORES DE WIKIPEDIA. *Wikipedia:ChromaFeatures.* Wikipedia, la enciclopedia libre, 11 Agosto 2017 [Consulta: 8 Octubre 2017]. Disponible en: https://en.wikipedia.org/wiki/Chroma_feature

COLABORADORES DE WIKIPEDIA. *Wikipedia: Transformada de Fourier Discreta.* Wikipedia, la enciclopedia libre, 22 Septiembre 2016 [Consulta: 22 Septiembre 2017]. Disponible en: https://es.wikipedia.org/wiki/Transformada_de_Fourier_discreta

COLOMER BLASCO, L. 2016. Acústica musical, *Capitulo 10. Análisis espectral de los sonidos musicales, Apartado 4. El espectrograma*, 5 Febrero 2016. [Consulta: 29 Julio 2017]. Disponible en: http://cursodeacusticamusical.blogspot.com.es/p/acerca-del-curso-de-ac.html

ELLIS, D. 2006. *Chroma Features Analysis and Synthesis.* Electrical Engineering, Columbia University, 21 November 2006 [Consulta: 25 Septiembre 2017]. Disponible en: http://www.ee.columbia.edu/~dpwe/resources/matlab/chroma-ansyn/

ESCOBAR ZAMORA, M.T. 2014. *Densidad espectral de potencia.* Guatemala, 17 Julio 2014 [Consulta: 18 Septiembre 2017]. Disponible en: https://es.scribd.com/document/234195022/Densidad-Espectral-de-Potencia

FRANCO GARCÍA, A. 2015. *Transformada Rápida de Fouerier (I).* Grado en Ingeniería de Energías renovables. Dpto. Física Aplicada I. Universidad del País Vasco, 2 Marzo 2015 [Consulta: 21 Septiembre 2017]. Disponible en: http://www.sc.ehu.es/sbweb/fisica3/datos/fourier/fourier_1.html

FUJISHIMA, T. 1999. *Realtime Chord Recognition of Musical Sound: A System Using Comon Lisp Music.* ICMA International Computer Music Conference, pp. 464-467 [Consulta: 30 Octubre 2017]. Disponible en: http://www.music.mcgill.ca/~jason/mumt621/papers5/fujishima_1999.pdf

GARDEY, A.; PÉREZ PORTO, J. 2010. *Definición de Sonido.* Actualizado en 2012 [Consulta: 25 Octubre 2017]. Disponible en: https://definicion.de/sonido/

GENTLEMAN, W. M.; SANDE, G. 1966. *Fast Fourier transforms – forfun and profit*. FallJointComputer Conf. AFIPS Proc., vol. 29, Washington D.C., Spartan, 1966 Consulta: 21 Septiembre 2017]. Disponible en: http://cis.rit.edu/class/simg716/FFT_Fun_Profit.pdf

GIL FERNÁNDEZ, J. 1987. Los sonidos del lenguaje: ‹‹Visualización de las características acústicas del habla en los documentos de análisis acústico››. Editorial Síntesis. Madrid, 2014 ed. ISBN 9788477380054.

GIL PÉREZ, M.; IGLESIAS GONZÁLEZ, J.; ROBLES OJEDA, G. 1998. *Lenguaje Musical Rítmico IV*. Edición Si bemol S.L. Torre del Mar (Málaga): Mayo 1998. ISBN: 84-95262-34-7.

GOLD, B.; MORGAN, N.; ELLIS, D. 1999. *Speech and Audio SignalProcessing: Processing andPerception of Speech and Music.*2nd Edition, Septiembre 2011 ISBN: 978-0-470-19536-9.

GÓMEZ GUTIÉRREZ, E. 2009. *Síntesi i Processament del So* I. Departament de Sonologia, Escola Superior de Musica de Catalunya, 28 de Septiembre de 2009. [Consulta: 27 Julio 2017]. Disponible en: http://www.dtic.upf.edu/~egomez/teaching/sintesi/SPS1/Tema3-Representacion.pdf

HERNÁNDEZ DE MIGUEL, J.M. 2012. *Estructura y características de la señal acústica, estudio y caracterización*. [Consulta: 29 Julio 2017]. Disponible en: https://jmhweb.wordpress.com/docencia/master-en-biologia-de-la-conservacion-ucm/estructura-y-caracteristicas-de-la-senal-acustica-estudio-y-caracterizacion/

IANUBRAE, T. 2014.*Sistema musical: Sistema temperado, "Temperamento Igual"*. [Consulta: 12 Octubre 2017]. Disponible en: http://lamusicadelasgalaxias.blogspot.com.es/2014/08/sistema-temperado-temperamento-igual.html

IGLESIAS GONZÁLEZ, J.; ROBLES OJEDA, G. 1999. *Lenguaje Musical Rítmico VI*. Edición Si bemol S.L. Torre del Mar (Málaga): 1º ed., Junio 1999. ISBN: 84-95262-12-6.

LESTER, J. 1989. *Enfoques analíticos de la música del siglo XX*. Ediciones Akal S.A. 2005 ed. Tres cantos (Madrid): W.W. Norton & Company, Inc. 1989. ISBN: 978-84-460-1692-2.

LEZANA ILLESCA, P. 2005. *Lectura 5: Transformada Rápida de Fourier*. Laboratorio de procesamiento digital de señales, ELO-385. Universidad Técnica Federico Santa María, Ayudante: Muratt Rodríguez, J., Valparaíso, 20 Mayo 2005 [Consulta: 4 Septiembre 2017]. Disponible en: http://www2.elo.utfsm.cl/~elo385/docs/Biblio/Lab3/Lectura_FFT.pdf

LÓPEZ, A.; MOLINA, E. 2007. *Cuaderno de teoría 5*. Enclave Creativa Ediciones S.L. Instituto de Educación Musical (IEM). Madrid: 1º ed., Junio 2007. ISBN: 978-84-96350-78-6.

MARTÍN, G. 2007. *Acústica: Descripción de la onda sonora.* La Guía, Física, 14 Diciembre 2007 [Consulta: 27 Octubre 2017]. Disponible en: https://fisica.laguia2000.com/acustica/descripcion-de-la-onda-sonora

MARTÍNEZ SALANOVA, E. 2015. *La música y las matemáticas: generando las notas musicales.* Revista científica Comunicar (versión electrónica), 23 Octubre 2015. ISSN: 1988-3293) Disponible en: https://revistacomunicar.wordpress.com/author/educomunicacioncreativa/page/11/

MOLINA GARCÍA, F.J. 2010. Telecomunicación (Sonido e Imagen), *"Desarrollo de una aplicación de secuenciado MIDI en MatLab"*, Gandía 2010. [Consulta: 31 Julio 2017]. Disponible en: https://riunet.upv.es/bitstream/handle/10251/9525/memoria.pdf

MOSCOSO, D. 2015. *La Música.* 12 de Abril de 2015 [Consulta: 30 Octubre 2017]. Disponible en: http://danielamoscoso1b.blogspot.com.es/2015/04/la-musica.html

ORTIZ ARREYGUE, M.A. 2014.*Transformada Discreta de Fourier y Transformada Rápida de Fourier.* Ingeniería Eléctrica, Señales y sistemas. 12 Mayo 2014 [Consulta: 21 Septiembre 2017]. Disponible en: https://es.scribd.com/doc/223469466/Transformada-Discreta-de-Fourier

PÉREZ, L.2008. *Los armónicos.*5 Agosto 2008 [Consulta: 19 Septiembre 2017]. Disponible en: http://www.monografias.com/usuario/perfiles/lia_perez/monografias

PISTON, W. 1987. *Armonía*. Mundimúsica Ediciones, S.L, 2008. 5ª ed. Alcorcón (Madrid): W.W. Norton &Company, Mark de Voto, *TuftsUniversity*. ISBN: 978-84-936631-1-7

SEGURA SOGORB, M. 2015. *Estimación automática de acordes para uso en transcripción musical a partir de audio.* Universidad de Alicante, Septiembre 2015 [Consulta: 4 Septiembre 2017]. Disponible en: http://grfia.dlsi.ua.es/repositori/grfia/degreeProjects/5/memoria.pdf

SHA'ATH, I. 2011. *Software KeyFinder.* Key estimation software for Djs. [Consulta: 10 Ocutubre 2017]. Disponible en: http://www.ibrahimshaath.co.uk/keyfinder/KeyFinder.pdf

SHA'ATH, I. 2011. *Estimation of key in digital music recordings.* Department of Computer Science and Information Systems. Birkbeck College, University of London [Consulta: 10 Octubre 2017]. Disponible en: http://www.ibrahimshaath.co.uk/keyfinder/

STEPHEN DOWNIE, J. 2017. *MIREX: Music Information Retrieval Evaluation eXchange*. The International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL) [Consulta: 15 Octubre 2017]. Disponible en: http://www.music-ir.org/mirex/wiki/MIREX_HOME

TOMASINI, M.C. 2006. El *fundamento matemático de la escala musical y sus raíces pitagóricas*. Colección C&T, Número 6. Universidad de Palermo, 2006 [Consulta: 30 Agosto 2017]. Disponible en: http://www.palermo.edu/ingenieria/downloads/CyT6/6CyT%2003.pdf

TRALIER, C. *Musical Pitches and ChromaFeatures.* [Consulta: 9 Octubre 2017]. Disponible en: http://www.ctralie.com/Teaching/ECE381_DataExpeditions_Lab1/

TZANETAKIS, G. 2001. *DATA SET*. [Consulta: 1 Septiembre 2017]. Disponible en: https://github.com/alexanderlerch/gtzan_key

TZANETAKIS, G.; COOK, P. 2002. *Musical genre classification of audio signals.* Published in IEEE Transactions on Speech and Audio Processing (Volume:

10, Issue: 5, Jul 2002). Page(s): 293 – 302, 07 November 2002. Print ISSN: 1063-6676.

VELASCO LORENZO, C.E. 2015. *Matemáticas, música y letra.* Valladolid, Julio 2015 [Consulta: 2 Septiembre 2017]. Disponible en: https://uvadoc.uva.es/bitstream/10324/15015/1/TFM-G%20492.pdf

VILANOVA ÁNGELES, S. *Análisis de audio.* UOC (Universitat Oberta de Catalunya) [Consulta: 25 Octubre 2017]. Disponible en: http://docplayer.es/6733555-Analisis-de-audio-santiago-vilanova-angeles-pid_00184754.html

9. ANEXOS

ENUMERACIÓN DE CLAVES O TONALIDADES

$$Mayor = 0 \dots \dots 11$$
$$Menor = 12 \dots \dots 23$$

Etiquetas

A	A#	B	C	C#	D	D#	E	F	F#	G	G#
0	1	2	3	4	5	6	7	8	9	10	11

a	a#	b	c	c#	d	d#	e	f	f#	g	g#
12	13	14	15	16	17	18	19	20	21	22	23