



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

– **TELECOM** ESCUELA
TÉCNICA **VLC** SUPERIOR
DE INGENIERÍA DE
TELECOMUNICACIÓN

SEGMENTACIÓN DE LA ESTRUCTURA MUSICAL MEDIANTE TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación

Noelia Aroca Martínez

Tutor: José Javier López Monfort

Curso 2019 - 2020
Valencia, 10 Noviembre de 2020

Resumen

Una canción suele estar estructurada en distintas partes, las cuales pueden ser identificadas claramente por el oyente en la mayoría de casos.

En general, la música se basa en partes normalmente denominadas como: *Introducción*, *Sección* (según el estilo puede ser *estrofa y/o estribillo*), *Interludio* o *Puente* y *Final*; que se van intercalando durante la canción.

Encontrar estas secciones de forma automática puede tener diversas aplicaciones en el mundo del multimedia y el espectáculo. A lo largo del trabajo se han tratado técnicas de segmentación automática utilizando diversos métodos de procesado de audio y de aprendizaje de máquinas (“machine learning”) con el objetivo de avanzar en este propósito. Dada la complejidad de los diferentes estilos musicales, el proyecto se ha centrado en un estilo musical determinado que es la música electrónica de baile (EDM). Este es uno de los géneros con una de las estructuras más sencillas y, por tanto, más fáciles de segmentar del amplio abanico de géneros musicales existentes.

Para llevar esto a cabo, se ha entrenado y ajustado un sistema de “machine learning”. Una de las tareas donde más esfuerzo se ha puesto ha sido al estudio estructural de este estilo musical, la definición de las partes y la implementación de una gran base de datos de muchas canciones etiquetadas manualmente, para entrenar un algoritmo de detección y así lograr cometer una tasa de acierto alta.

Una vez elaborada la base de datos, se ha trabajado en un software de clasificación básico basado en umbrales de parámetros, para verificar que dicho algoritmo detecta correctamente las secciones de la música y comete el mínimo error posible.

Los resultados obtenidos han sido muy alentadores, obteniendo tasas de acierto superiores al 90% con bajas falsas alarmas, y por tanto haciendo más viable el objetivo último de crear un sistema automático que pueda ser implementado y explotado comercialmente en el sector audiovisual, principalmente en espectáculos en directo, discotecas, etc., para sincronizar luces, animaciones 3D y otras efectos visuales con la música.

Resum

Una cançó sol estar estructurada en distintes parts, les quals poden ser identificades clarament per l'oient en la majoria de casos.

En general, la música es basa en parts normalment denominades com: Introucció, Secció (segons l'estil pot ser estrofa i/o tornada), Interludi o Pont i Final; que es van intercalant durant la cançó.

Trobar aquestes seccions de manera automàtica pot tindre diverses aplicacions en el món del multimèdia i l'espectacle. Al llarg del treball, s'han tractat tècniques de segmentació automàtica utilitzant diversos mètodes de processament d'àudio i d'aprenentatge de màquines (“machine learning”) amb l'objectiu d'avançar en aquest propòsit. Donada la complexitat dels diferents estils musicals, el projecte s'ha centrat en un estil musical determinat que és la música electrònica de ball (EDM). Aquest és un dels gèneres amb una de les estructures més senzilles i, per tant, més fàcils de segmentar de l'ampli ventall de gèneres musicals existents.

Per a portar açò a cap, s'ha entrenat i ajustat un sistema de “machine learning”. Una de les tasques on més esforç s'ha posat ha sigut a l'estudi estructural d'aquest estil musical, la definició de les parts i la implementació d'una gran base de dades de moltes cançons etiquetades manualment, per a entrenar un algorisme de detecció i així aconseguir cometre una taxa d'encert alta.

Una vegada elaborada la base de dades, s'ha treballat en un programari de classificació bàsic basat de llindars de paràmetres, per a verificar que aquest algorisme detecta correctament les seccions de la música i comet el mínim error possible.

Els resultats obtinguts han sigut molt encoratjadors, obtenint taxes d'encert superiors al 90% amb baixes falses alarmes, i per tant fent més viable l'objectiu últim de crear un sistema automàtic que pugui ser implementat i explotat comercialment en el sector audiovisual, principalment en espectacles en directe, discoteques, etc., per a sincronitzar llums, animacions 3D i uns altres efectes visuals amb la música.

Abstract

A song is usually structured in different parts which can be clearly identified by the listener in most cases.

In general, the music is based on parts normally called as: Intro, Section (depending on the style it can be verse and/or chorus), Interlude or Bridge and Final; that are blended during the song.

Finding these sections automatically can have several applications in the field of multimedia and entertainment. Throughout the work, automatic segmentation techniques have been discussed using various methods of audio processing and machine learning with the aim of advancing in this purpose. Given the complexity of the different musical styles, the project has focused on a particular musical style which is electronic dance music (EDM). This is one of the genres with one of the simplest structures and, therefore, easiest to segment from the wide range of existing musical genres.

To carry out this, a “machine learning” system has been trained and adjusted. A greater effort has been required in the tasks of structural study of this musical style, the definition of the parts and the implementation of a large database of many manually tagged songs, to train a detection algorithm and thus achieve commit a high hit rate.

Once the database has been developed, a basic classification software based on parameter thresholds was used to verify that this algorithm correctly detects sections of the music and makes the least possible error.

The results obtained have been very encouraging, acquiring success rates higher than 90% with low alarm failures, and therefore making the ultimate objective of creating an automatic system that can be implemented and commercially exploited in the audiovisual sector feasible, mainly in live shows, nightclubs, etc., to synchronize lights, 3D animations and other visual effects with the music.

Índice

1.	Introducción	6
2.	Objetivos del proyecto	7
3.	Metodología del trabajo	8
3.1.	Gestión del proyecto	8
3.2.	Distribución en tareas	9
3.3.	Diagrama temporal	10
4.	Desarrollo del trabajo	11
4.1.	Segmentación musical	11
4.2.	Segmentación musical en música electrónica	15
4.2.1.	Estructura de la música electrónica	15
4.2.2.	Términos en EDM	16
4.2.3.	Sub-géneros de EDM	16
4.3.	Etiquetado manual de canciones	18
4.3.1.	Selección de eventos a etiquetar en EDM	18
4.3.2.	Extensión de esta estructura al pop/rock	19
4.3.2.	Herramienta software de etiquetado	19
4.4.	Observaciones y particularidades encontradas en la música EDM	25
4.5.	Clasificación	25
4.5.1.	Funcionamiento del software de clasificación	25
4.5.2.	Extracción de las características	26
4.5.3.	Clasificación	27

5.	Resultados del trabajo	28
5.1.	Obtención de los umbrales	28
5.2.	Validación del algoritmo con el conjunto de validación	29
6.	Conclusiones y trabajo futuro	31
6.1.	Conclusiones	31
6.2.	Trabajo futuro	31
7.	Bibliografía	32

1. Introducción

Una canción o una composición musical, en sentido amplio, suele estar constituida por distintas partes más o menos diferenciadas que pueden ser identificadas claramente por el oyente en la mayoría de casos.

En general, en la música pop y contemporánea se suele tener una estructura basada en partes, las cuales normalmente se denominan: Intro, Verse, Chorus, Bridge, Outro que a medida que va progresando la canción, estas se van intercalando.

El verso (Verse) se repite unas cuantas veces y suele tener variaciones unos versos con otros. El estribillo (Chorus) también se repite varias veces y además, es el gancho que la gente recordará. Y por último, el puente (Bridge) es la parte donde se rompe el estribillo en algún momento durante la canción.

Manualmente, un humano podría identificar los eventos en el tiempo donde se producen los cambios en las secciones de la canción. Sin embargo, resultaría de mucha utilidad para ciertas aplicaciones relacionadas con el mundo del espectáculo que un sistema automático por ordenador fuera capaz de identificar dichos eventos.

Por este motivo, en los últimos años, diversos grupos de investigación han trabajado en técnicas de segmentación automática utilizando diversos métodos de procesamiento de audio y de aprendizaje de máquinas (“machine learning”).

No obstante, los resultados no son completamente fiables dado que la música es un arte muy complejo, y los distintos tipos de música, estilos y carácter de la misma han variado a lo largo de los siglos. No es lo mismo una pieza barroca que un tema de jazz, aunque los fundamentos musicales sean los mismos.

Debido a este problema que se nos plantea, si nos centramos en un estilo particular resultará más fácil el trabajo puesto que centraremos nuestros esfuerzos en un solo estilo. Este es el caso de este trabajo donde nos centraremos en la música electrónica de baile (EDM), la cual dentro del amplio abanico de estilos que hay en la música, sigue una de las estructuras más sencillas. Por tanto, para comenzar y ver cómo se desarrolla el trabajo es mejor empezar con un tipo de música sencillo.

Las aplicaciones de un sistema automático que segmente una canción en sus partes son muy diversas. Un ejemplo de mucho interés es la sincronización de luces y videoclips con la música, cambiando los mismos cuando cambia la sección de la canción. Una aplicación así, tendría un gran interés comercial, dado que puede sustituir a los operadores de luces o videoDJ en espectáculos donde se quiera disminuir el personal necesario.

Por consiguiente, para poder implementar un sistema de estas características hace falta entrenar y ajustar el sistema de “machine learning”, por lo que será necesario contar con bases de datos de muchas canciones que alimenten el algoritmo y así se cometa el menor error posible.

A lo largo de este trabajo iremos explicando todos estos detalles y cómo los hemos abordado en este proyecto.

2. Objetivos del proyecto

El objetivo de este proyecto será trabajar en el desarrollo de un sistema automático de segmentación musical basado en “machine learning”.

En particular nos centraremos en la música electrónica de baile (EDM – “Electronic Dance Music”), tanto por ser relativamente más fácil de segmentar que otros tipos de música como por el interés comercial que puede suscitar un algoritmo en este sector del ocio en la actualidad.

Para llevar este trabajo a cabo, se deben desarrollar los siguientes objetivos parciales organizándolo así en distintas partes:

- 1) Definición de qué tipos de eventos de cambio de sección se producen en la música EDM.
- 2) Compilación de una base de datos extensa de EDM etiquetada manualmente para que sirva de aprendizaje al sistema bajo los criterios del punto 1.
- 3) Elegir, probar y ajustar un método de detección/clasificación automática de los eventos.
- 4) Validar y comprobar la precisión del sistema desarrollado.

Este trabajo se ha centrado la mayor parte del tiempo en los puntos 1 y 2, aunque en la última parte también se ha trabajado en los puntos 3 y 4, en este último caso, con más ayuda del profesor tutor dada la complejidad de los mismos y el know-how necesario en “machine learning”.

3. Metodología del trabajo

3.1. Gestión del proyecto

La gestión de proyectos en definitiva es una serie de pasos a seguir o fases para planificar y dirigir un proyecto de la forma más correcta posible. Por tanto, las fases son las siguientes [1] [2]:

a) Análisis de viabilidad del proyecto

Es el paso principal ya que se evalúa si el proyecto puede seguir adelante o no.

En esta fase se valora si los beneficios que se pueden obtener con el proyecto son mayores a las inversiones, y si esto ocurre se sigue adelante.

En nuestro caso, puesto que el proyecto había comenzado por un grupo de ingenieros en la UPV, este paso ya se llevó a cabo.

b) Planificación del trabajo a realizar

Una vez se ha decidido continuar con el proyecto, pasamos a detallar todas las tareas que se deben realizar así como los recursos que serán necesarios.

Es importante detallar todo con claridad ya que si alguna parte es ambigua puede repercutir negativamente al proyecto. Para que esto no ocurra, es aconsejable:

- Analizar el alcance del proyecto.
- Estimar costes y recursos.
- Definir el plan del proyecto.

Por tanto, el alcance de nuestro proyecto es darle un uso comercial con motivo de utilizarlo en el ámbito audiovisual y del espectáculo, por ejemplo, para la automatización de luces en un concierto o festival.

En este caso, el alumno ha empleado numerosas horas en el estudio y el trabajo del proyecto durante las semanas desde Febrero de 2020, aunque por motivos de la Covid-19 surgieron problemas para continuar correctamente retrasándose así su entrega prevista para Septiembre del 2020. Por tanto, ya que los programas utilizados los han llevado a cabo ingenieros del grupo y el alumno tenía que aprender a usarlos así como entenderlos correctamente, no ha supuesto un coste determinado.

Por último, el objetivo del proyecto es realizar unas tareas determinadas para llegar al resultado final, es decir, se comienza por la segmentación de canciones creando una amplia base de datos; una vez obtenida dicha base de datos, se pasa a un programa de clasificación el cual sea capaz de identificar dichos segmentos y, por ello, poder crear un sistema automático. Tras todo este proceso, se pretende que el programa de clasificación al detectar una determinada etiqueta realice una función determinada en ese instante de mayor energía, es decir, si por ejemplo se usa para automatizar luces se podría encender una determinada luz en ese instante.

c) Ejecución del proyecto

En esta parte se trata de asignar las tareas planificadas dependiendo de los recursos, ejecutarlas y gestionar los posibles cambios que sean necesarios realizar.

Puesto que se trata de un proyecto que ya había comenzado, el software para el etiquetado de canciones ya estaba creado. Por tanto, la primera tarea de este trabajo es relacionar las etiquetas disponibles del programa con la estructura que tiene la música EDM.

Tras entender la relación que se ha determinado entre dicho estilo de música y el programa, se procede al etiquetado de canciones.

Por último, cuando se hayan realizado todas las etiquetas, estas se pasarían por el software de clasificación donde se visualizará y se comprobará que el sistema las detecta correctamente.

d) Seguimiento y control del trabajo

Se centra en comprobar si se está cumpliendo la planificación y así poder actuar de manera rápida y sencilla a los problemas que surjan. Por tanto, en esta parte se realizan:

- Seguimientos de tareas planificadas.
- Reuniones con el jefe/tutor para evaluar si el seguimiento es el correcto.
- Reuniones para explicación de nuevas tareas.

Así pues, durante el proceso del proyecto se han ido realizando reuniones para entender los programas a utilizar así como para comprobar que el etiquetado seguía los criterios definidos en un primer momento, y si no, modificar aquellas que no estuviesen correctamente.

Además, también se han realizado reuniones para verificar que el programa de clasificado detecta perfectamente cuándo sucede un evento o segmento en la canción y para observar los resultados obtenidos y si estos son correctos.

e) Cierre

Esta es la última fase en la que se verifica que todo el trabajo realizado está correctamente o se comete el menor error posible antes de cerrar el proyecto.

Además, se debe realizar un análisis de los resultados obtenidos con las estimaciones iniciales.

Así, en recopilación con todo lo anterior, los principales objetivos en la gestión de un proyecto son:

- Gestionar el inicio y evolución del proyecto.
- Controlar y buscar soluciones ante los problemas que vayan surgiendo.
- Verificar que se ha finalizado el proyecto correctamente.

3.2. Distribución en tareas

Es necesario organizar en tareas todo el proyecto para tratar de cometer un error mínimo. Las tareas distribuidas han sido:

- Buscar información sobre la segmentación musical de forma general.
- Buscar información de los términos y tecnicismos en música electrónica.
- Vincular el software de etiquetado con la música electrónica.
- Etiquetar unas 500 canciones para tener una base de datos amplia.
- Entender el funcionamiento del software de clasificación.
- Clasificar dichas etiquetas en el software indicado y verificar su resultado.

3.3. Diagrama temporal

FEBRERO	<ul style="list-style-type: none"> • Inicio TFG. • Reunión con el tutor. • Buscar información música EDM. • Familiarizarse con el programa de etiquetado. • Descargar música.
MARZO	<ul style="list-style-type: none"> • Realizar primeras etiquetas. • Revisión primeras etiquetas. • Covid-19 => No poder reunirse ni realizar el proceso de clasificación.
ABRIL	<ul style="list-style-type: none"> • Continuar con etiquetas. • Descargar música nueva. • Revisión de etiquetas por la aplicación Zoom.
MAYO	<ul style="list-style-type: none"> • Continuar con etiquetas. • Descargar música nueva. • Revisión de etiquetas por la aplicación Zoom.
JUNIO	<ul style="list-style-type: none"> • Continuar con etiquetas. • Descargar música nueva. • Revisión de etiquetas por la aplicación Zoom.
JULIO	<ul style="list-style-type: none"> • Continuar con etiquetas. • Descargar música nueva. • Revisión de etiquetas presencialmente.
AGOSTO	<ul style="list-style-type: none"> • Finalizar etiquetas.
SEPTIEMBRE	<ul style="list-style-type: none"> • Previsión de entrega del TFG (antes del Covid-19). • Revisión etiquetas. • Reunión para explicar el proceso de clasificación. • Comienzo de redacción del TFG.
OCTUBRE	<ul style="list-style-type: none"> • Pasar las etiquetas al programa de clasificación. • Ver posibles errores y solucionarlos. • Continuar con la redacción.
NOVIEMBRE	<ul style="list-style-type: none"> • Últimos retoques del TFG. • Entrega de TFG. • Exposición del TFG.

Tabla.1. Diagrama temporal.

Nota: Por motivo del Covid-19 se retrasó todo, por lo que el diagrama temporal llega a Noviembre cuando se preveía terminar el TFG en Septiembre.

4. Desarrollo del trabajo

En el mundo actual del ocio y del entretenimiento cada vez son más los demandantes de espectáculos donde la animación visual tenga una gran relevancia ya sea en discotecas, salas de conciertos, festivales, etc.

Debido al gran auge de la tecnología en el siglo XXI, es imprescindible pensar en aunar la tecnología disponible en nuestros días con la creciente demanda que se presenta. Por tanto, aunque en la industria audiovisual ya se le da la suficiente importancia a lo visual, se plantea poder realizar una animación visual que se diseñe con anticipación y se adapte al estilo de música que se reproducirá.

Este trabajo se centra en ampliar el uso de la tecnología mediante el desarrollo de un software de segmentación musical asequible que se utilizará en gráficos en movimiento 3D o automatizar luces en tiempo real para la industria de música EDM.

Para disminuir los costes y también el personal, se trata de automatizar la animación visual o las luces en tiempo real durante una sesión en un evento musical determinado. Esto implica que se debe desarrollar un análisis musical computacional para extraer la información más importante y útil para el contenido visual.

En esta parte se presenta todo el desarrollo llevado a cabo para hacer posible el objetivo comentado.

Primero, se habla de la segmentación o estructura musical de forma general y de forma específica para la música electrónica de baile (EDM).

Después, se comentan los distintos sub-géneros y términos utilizados en EDM para tratar de comprender la relación presente entre el software utilizado para el segmentado de canciones, así como la estructura de música electrónica.

El siguiente paso es mostrar la herramienta y el funcionamiento de la misma. Por ello, se pondrá como ejemplo una canción determinada para ver cómo se van realizando los eventos poco a poco y cómo se muestran todas las etiquetas una vez finalizada su segmentación.

Además, también se muestra otro ejemplo con música pop/rock para corroborar la similitud que tiene con la música EDM, puesto que ambos son estilos de música contemporánea.

Por último, se comenta el software de clasificación donde se introducen todas las etiquetas realizadas en las canciones para que este las detecte y se pueda llegar al objetivo de crear un sistema automático con técnicas de “machine learning”, así como comentar los resultados obtenidos y verificar si son correctos.

A continuación, se presenta todo de forma más detallada.

4.1. Segmentación musical

Una obra musical o canción tiene una estructura que la define. Para comprender la estructura de una canción tenemos que analizar la melodía que está presente en cada parte y analizar cuáles son las similitudes y diferencias entre sus partes.

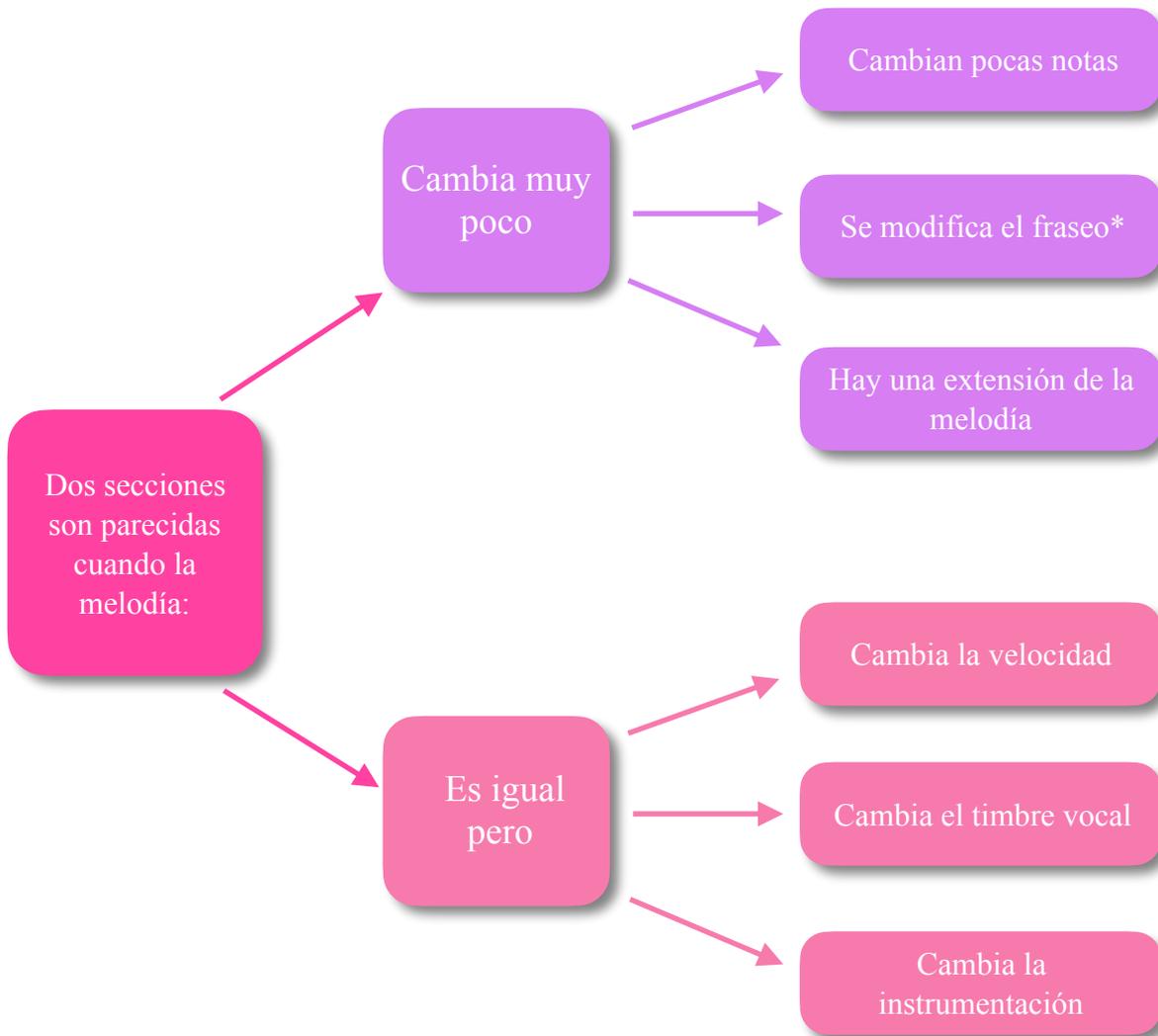
De forma general, se podría llamar a la primera parte “Introducción” ya que en esta parte se presenta la obra y permite poner en contexto al oyente, es decir, en este fragmento mayoritariamente se presenta el ritmo, los instrumentos que intervienen, género, estilo, compás, etc.

Tras la introducción (que no siempre tiene por qué estar presente), vendría la primera estrofa a la cual se le denomina A. Así, si la siguiente sección contiene la misma melodía, instrumentos, velocidad, ritmo, timbre, etc. también se le llamará A. En cambio, si no es así se le denomina B.

Como vemos, el nombre de un fragmento de la canción sigue el orden alfabético, por tanto, si dicho fragmento tiene una letra distinta a otro anterior es porque la melodía lo es.

Cuando una fracción es la misma pero varían en cierto modo como por ejemplo, la velocidad o algún instrumento; se dice que esa parte es parecida a una de las anteriores. En ese caso, a esta parte se le asigna la misma letra a la que se parece pero añadiéndole una comilla (') llamada “Prima”. Se añadirán tantas comillas como partes parecidas surjan en la canción (A, A', A'', ...).

Todo lo que se ha descrito se puede apreciar de forma más clara en el siguiente esquema (*Esquema.1*) [3].



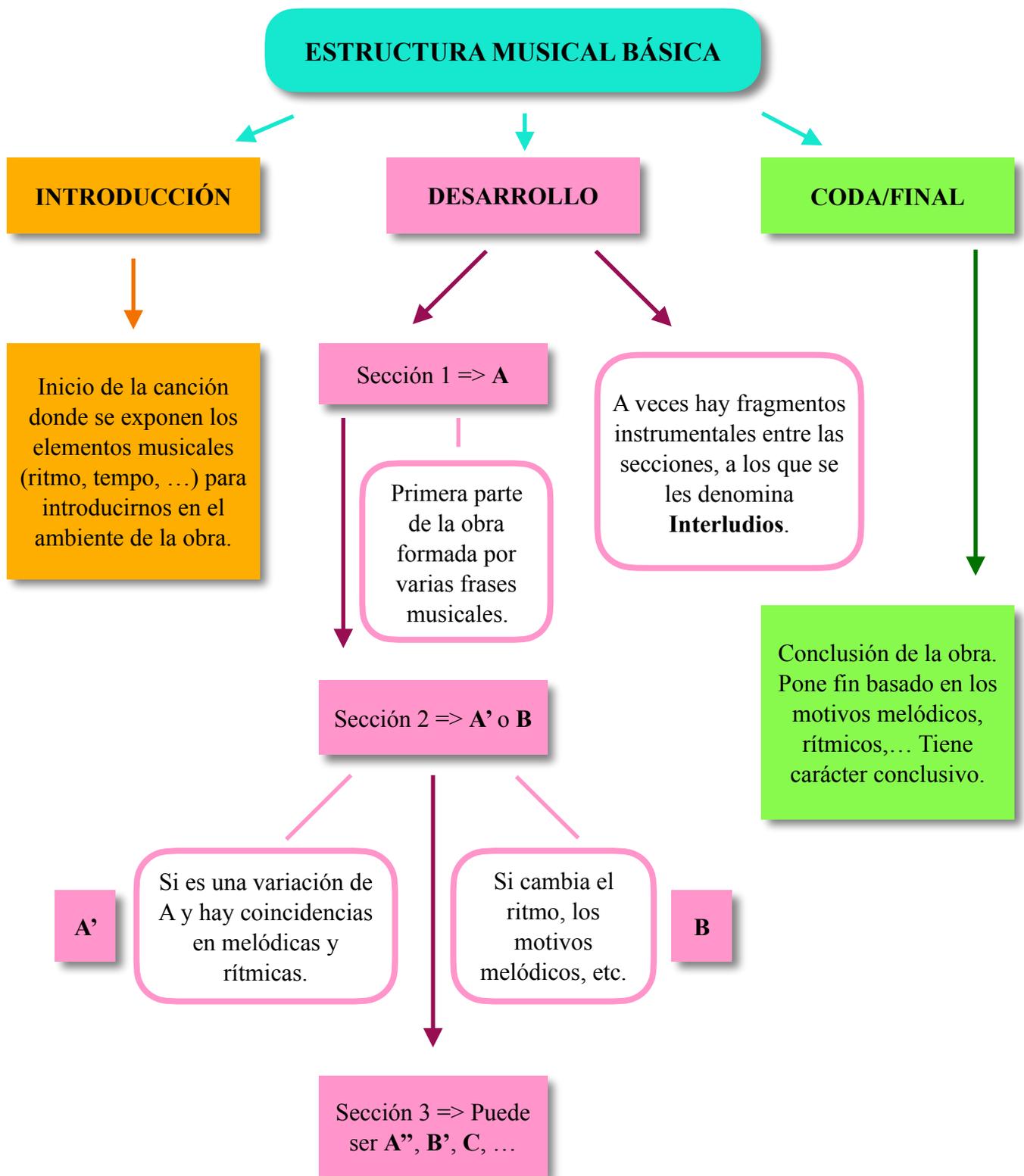
Esquema.1. Fracciones parecidas en una canción.

Una vez se ha analizado todo el desarrollo de la canción, es decir, la canción se organiza con una introducción opcional y unos fragmentos diversos o similares a los cuales se les asignan letras del abecedario, llegamos a la parte final de la obra musical.

Al igual que se puede tener una introducción también puede estar presente un final. En música, al final de una canción se le denomina “Coda”, “Final” o “Outro”.

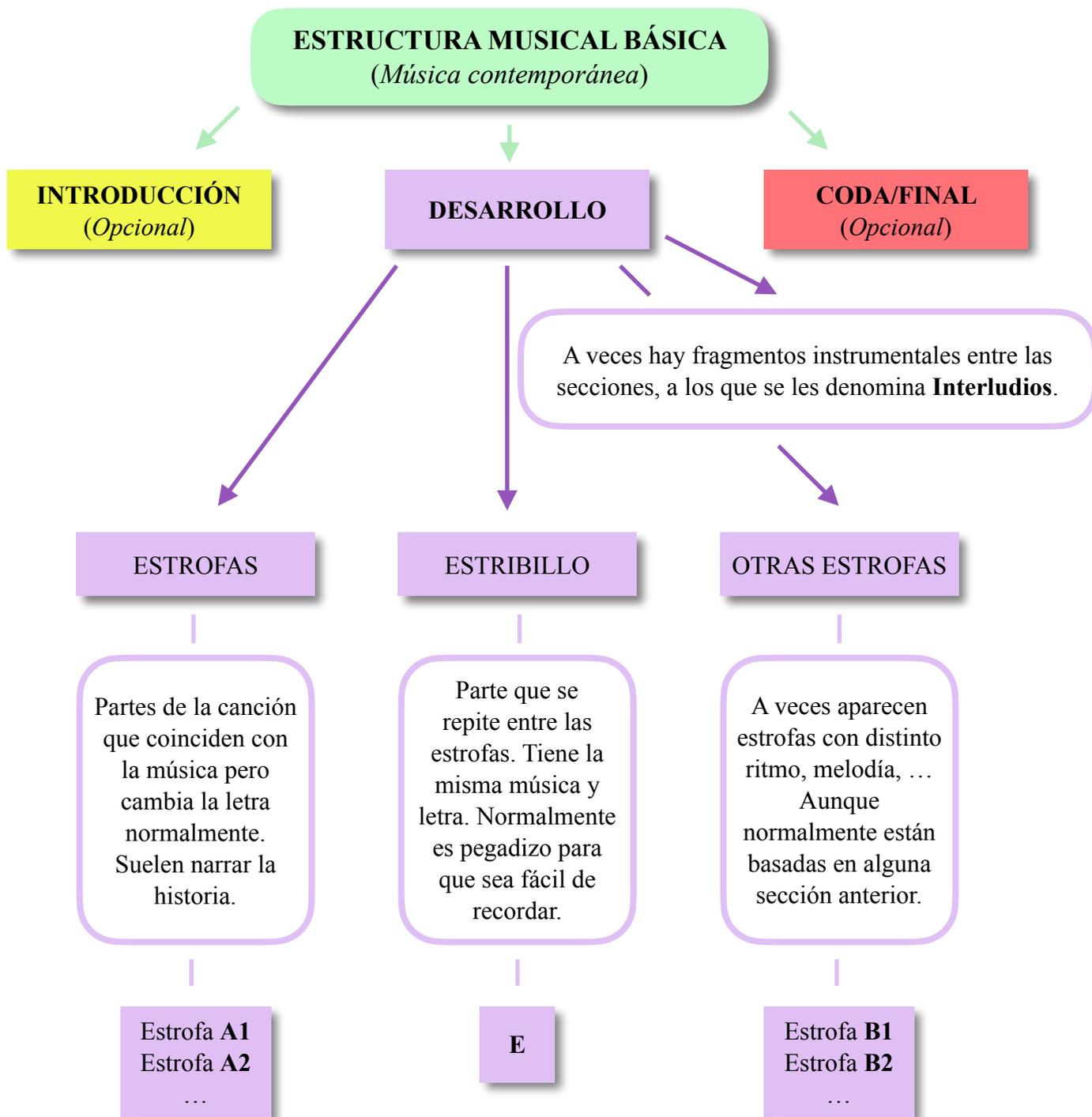
En resumen, se puede decir que se ha establecido cómo está organizada y estructurada una canción de forma general, pero para visualizar de forma más detallada y clara cómo se estructuran o segmentan las canciones se presenta a continuación un esquema en el cual se aprecian las diferentes partes descritas anteriormente (Esquema.2) [4].

*fraseo: Es la agrupación de notas tanto en la forma que está hecha su composición como en la interpretación de las mismas. También se le denomina frase puesto que no es más que una melodía [5].



Esquema.2. Estructura de una obra musical básica.

En música contemporánea la estructura es prácticamente la misma pero cambian algunas designaciones y hay algunas partes que son opcionales. Por ello, se muestra otro esquema en el que se pueden ver las diferentes partes presentes en este tipo de música (*Esquema.3*) [4].



Esquema.3. Estructura básica de música contemporánea.

Tanto la introducción como la parte final son opcionales pero es bastante común que ambas estén presentes en una canción, si estas no aparecen normalmente se debe a que se trata de un estilo más experimental o minimalista para sorprender al oyente.

Además, si se analiza y compara el Esquema.3 con el Esquema.2 es fácil apreciar que las principales diferencias se encuentran en el desarrollo de la obra.

Mientras que en un concepto más general de música los distintos fragmentos se denominan secciones y siguen un orden alfabético, en música contemporánea se distinguen los fragmentos dependiendo de cómo varían entre ellos, es decir, se denominan estrofas y estribillos, aunque también se utilizan las letras del abecedario para diferenciarlo.

4.2. Segmentación musical en música electrónica

La música electrónica sigue un patrón similar a la música pop/rock pero no es exactamente igual. Dependiendo del estilo musical, las características particulares que el género presente y la intención de la obra, influirá todo ello de una determinada manera en las elecciones del sistema de reconocimiento y segmentación.

4.2.1. Estructura de la música electrónica

La estructura de la música electrónica o EDM (Electronic Music Dance) de forma general se divide en [6] [7] [8]:

- **Intro**: Sirve como presentación de la canción marcando el ritmo que va a llevar. Es un fragmento opcional y se reproduce una sola vez al principio de la canción (aunque a veces hay ciertos elementos de la introducción que aparecen a lo largo de la canción). Normalmente suele ser algún elemento de percusión y algún elemento melódico que juntos introducen la sensación de la pista. Además, sirve de ayuda a los DJ's para hacer la transición a sus sets.
- **Breakdown**: Después de la "intro" suele seguirle el "breakdown" donde la batería/ritmo se apaga para anticipar al oyente. Se da una ruptura del ritmo para que la canción cambie de ritmo o melodía dentro de la misma estructura de la canción. Es la parte más calmada y pausada de la canción ya que los ritmos desaparecen y se les da más importancia a los instrumentos melódicos creando así una atmósfera tranquila.
- **Build up**: Es un "puente" donde la energía de la canción va aumentando para preparar al oyente para la parte con más energía de la canción. Esta sección parte desde el "breakdown" y se van añadiendo capas de instrumentos o sonidos para anticipar al oyente y mostrarle que se aproxima la parte más álgida de la obra. Suelen aparecer elementos como sintetizadores de tono ascendente y patrones de percusión acelerados.
- **Drop**: Suele ser la parte más importante de la canción, es como el estribillo en música pop. Cambia según el género, pero de forma general, es el punto más álgido y más energético de la canción en el cual se libera la tensión mostrada en el "build-up" y el beat se activa creando un ambiente de máximo alcance sonoro y atmósfera bailable. Se produce una acumulación de distintas melodías y sonidos que dan lugar al punto máximo sonoro de la canción.

Además, para nombrar las partes en música electrónica también se suelen utilizar las letras del abecedario como en música clásica:

- A: Sección que representa los versos de la canción o "Breakdown".
- B: Representa al "Drop".
- C: Es el puente de la canción o "Build up".

Una de las estructuras más comunes en EDM es: ABAB o ABCB.

Las partes descritas se identifican de una forma más compleja y menos predecible cuando se trabaja con música de DJ en directo, ya que las canciones no comienzan ni terminan de una forma clara. El DJ suele eliminar la introducción y el final para mezclar toda la música de una manera continua debido a que el objetivo de este género es que sea música para bailar y busca que los oyentes estén en la pista bailando el mayor tiempo posible. Por lo que, dificulta la detección y segmentación de las canciones en estos ambientes.

4.2.2. Términos en EDM

Otros términos que se utilizan en EDM y que son conveniente conocer son [8] [9]:

- Beats: pulsaciones rítmicas de una canción.
- Key: es el equivalente a los sostenidos.
- Bpm: beats por minuto, es decir, número de pulsaciones por minuto. Es la unidad empleada para medir el tiempo en música.
- Loop: una determinada sección que se repite de forma continuada. Hay loops de 4, 8, 16, 32 o 64 beats.
- Waveform: representación gráfica de la amplitud de un sonido respecto al tiempo.
- Sample: Pequeña porción de sonido que ha sido grabada previamente y que sirve para ser reutilizada a lo largo de la canción.

4.2.3. Sub-géneros de EDM

Dentro de la música electrónica hay distintos sub-géneros que son variaciones o mezclas de estilos pero que se diferencian unos de otros. Cada uno de esos sub-géneros dispone de unas reglas propias que les permite adaptarse de forma diferente a la estructura de música EDM que se comentaba en el apartado 4.2.

Esto complica significativamente el desarrollo del software para resolver el desafío del segmentado, y más aún si tenemos en cuenta que también dependerá por la subjetividad del sujeto que se encargue en segmentar las canciones, debido a que probablemente dos personas diferentes etiquetarían los segmentos de una forma distinta.

Por tanto, desde un punto de vista práctico cada segmento se puede considerar como una parte de la canción, la cual la audiencia pueda percibir, es decir, partes donde se detecte una rotura de la canción, capas de instrumentos o sonidos que se van añadiendo o quitando, etc.

Los distintos estilos de EDM más importantes que se pueden encontrar en la actualidad son los siguientes [10] [11]:

◆ HOUSE:

Este sub-género está influenciado por la música disco ya que imita su percusión con el golpe de bombo en cada beat, baterías electrónicas, samples, suelen llevar melodías vocales bastante presentes.

Es un tipo de música que no suele ser demasiado agresiva como su percusor (Progressive House), siendo más agradable en cuanto a las sensaciones y energía que proyecta.

Algunos ejemplos de este estilo son: “Waiting for love” de Avicii o “This is what you came from” de Calvin Harris feat. Rihanna.

◆ DUBSTEP:

Combina varios estilos como 2-step garage (usa ritmos irregulares), broken beat, reggae, etc.

Tiene un sonido abrumador y sigue unos patrones estructurales poco predecibles. Se debe a que tiene loops con líneas de bajo bastante presentes, patrones rítmicos y samples que se van cortando, voces muy ocasionales o incompletas.

Algunos ejemplos: “Bangarang” de Skrillex o “I can’t stop” de Flux Pavilion.

◆ **DRUM AND BASS:**

Se considera una ramificación de la música rave y se caracteriza por mezclar breaks muy acelerados con potentes líneas de bajo.

Algunos ejemplos son: “Peace Love and Unity” de DJ Hype o “Blood Sugar” de Pendulum.

◆ **TECHNO:**

Hay muchos estilos de techno pero el más popular es “Detroit Techno” que se basa en utilizar un sintetizador de una forma más experimental teniendo referencias al free jazz y al funk.

Algunos ejemplos: “Bring the Noise” de Benny Benassi o “Watch Out” de Alex Gaudino.

◆ **TRANCE:**

Es caracterizado por ser un estilo muy melódico ya que las melodías tienen mayor presencia que la línea de bajo o los patrones rítmicos. Una de las características más importantes de este sub-género es que tiene elementos que se van repitiendo a lo largo de la canción y es de los estilos más bailables.

Algunos ejemplos: “Insomnia” de Faithless o “Children” de Robert Miles.

◆ **TRAP:**

Uno de los sub-géneros más populares en la actualidad. Recibe influencias del hip-hop, house, dubstep, electro house, etc.

Tiene un contenido lírico muy presente, contiene múltiples pistas de sintetizadores agresivos, caja de ritmos y bajos sintetizados y también teclados que producen sonidos melancólicos. Otra característica importante es que su tempo suele ser de 140 bpm.

Algunos ejemplos: “Chasing Colors” de Marshmello o “Oh me, oh my” de Dj Snake.

◆ **HARDCORE:**

Es uno de los estilos más conocidos. Se caracteriza por tener pistas con un sonido pesado y con ritmos muy marcados, así como melodías pegadizas. En este tipo de música predomina la línea de bajo y refuerzo de graves.

Algunos ejemplos son: “Earthquake” de Art of Fighters o “Adagio for Strings” De Tiësto.

Aunque estos son algunos de los sub-géneros más importantes, a partir de estos estilos se han ido derivando en otros formando así una amplia cadena de diferentes estilos de música EDM.

4.3. Etiquetado manual de canciones

Tal como se describió en los objetivos, es necesario etiquetar manualmente una base de datos para poder entrenar un sistema automático y también para verificar que dicho sistema funciona con la suficiente precisión. El primer paso será definir las etiquetas que se van a utilizar.

Las distintas secciones de una canción presentarán técnicas de clasificación y procesamiento de señal para detectar los eventos importantes en la música EDM.

4.3.1. Selección de eventos a etiquetar en EDM

En función de todo el know-how comentado en el apartado 4.2, hemos seleccionado unos eventos determinados para etiquetar. Hemos simplificado los eventos anteriores con criterios de calidad para aplicar en entornos reales, buscando la coherencia y la facilidad de etiquetado manual.

Los eventos a utilizar quedan definidos como “Break In”, “Break Out”, “Layer”, “Crescendo”, “Decrescendo”, “Voice In”, “Voice Out” e “Instant”.

Sin embargo, como se ha expuesto anteriormente, en EDM los términos que se usan en cuanto a estructura son “Breakdown”, “Build up” y “Drop”.

Por tanto, para llegar a un consenso y que se sepa cómo se está etiquetando una canción se debe proponer un criterio de etiquetado.

El criterio que se ha seguido ha sido:

- **BREAK IN:** Este término se ha utilizado de forma similar al “Breakdown”, es decir, esta etiqueta sirve para indicar en el programa que comienza una zona con poca energía.
- **BREAK OUT:** Es lo contrario a BREAK IN, es decir, es el “Drop”. Así, cuando se utiliza esta etiqueta indica que comienza el punto álgido de la canción y, por tanto, ha acabado la parte menos energética.
- **LAYER:** Esta etiqueta sirve para indicar que se añade o se quita una capa, es decir, indica cuándo un instrumento nuevo o una melodía entra/sale de la canción. No indica un “Build up” como tal, pero sí cuando se va uniendo o quitando un sonido.
- **VOICE IN:** Plasma cuando entra la voz en la canción.
- **VOICE OUT:** Al contrario que VOICE IN, indica que la voz sale de la canción.
- **INSTANT:** Sirve para señalar el momento en el que se escucha un sonido determinado en un instante muy corto de tiempo.
- Además de estas etiquetas, el programa también presenta las etiquetas de CRESCENDO y DECRESCENDO pero como es difícil de etiquetar y no está muy claro cuándo se puede utilizar, se ha decidido no utilizarlas.
- Las casillas de “Load Song”, “Save Tags”, “Undo” y “Play” sirven para cargar la canción que se quiera, guardar las etiquetas realizadas, borrar la etiqueta anterior y reproducir/pausar la canción respectivamente.

4.3.2. Extensión de esta estructura al pop/rock

Aunque los objetivos del trabajo sean centrarse en la música EDM, durante el mismo se decidió ampliar los objetivos e incluir algunas canciones pop/rock.

Así, se analizó la posibilidad de utilizar la misma estructura que sigue la música EDM para música pop/rock y el resultado obtenido fue que efectivamente era posible, ya que tiene una similitud en cuanto a su segmentado se refiere, con el de la música EDM.

Esto hace posible que el software sea útil para diferentes géneros, si asignamos previamente las etiquetas del programa a los segmentos de la estructura de un estilo de música determinado.

Tal y como se expone en apartados anteriores, toda obra musical tiene una estructura determinada. Por lo que, es importante observar cómo se encuentra estructurado el género pop/rock para adaptarla a la estructura de la música EDM.

La gran mayoría de canciones siguen un patrón bastante sencillo. Dicho patrón es el que se definía como estructura en música contemporánea que se podía ver claramente en el *Esquema.3* [4].

De manera general se podría decir que gran parte de las canciones de género pop/rock están formadas mínimamente por versos o estrofas y estribillos. Después, la introducción, interludio (también se le suele denominar “Puente”) y final son partes opcionales pero cuando se usan suelen generar tensión o clímax en la canción.

Por tanto, tratando la música pop/rock como un sub-estilo de la música EDM, se llega a la siguiente conclusión:

- BREAK IN: Puesto que se asignaba a la parte menos energética de la canción, se ha decidido utilizarla para indicar el comienzo de una estrofa ya que en este género musical normalmente es la parte más relajada de la canción.
- BREAK OUT: Al igual que antes que se usaba para señalar la parte álgida de la canción, en música pop/rock la parte más importante suele ser el estribillo de la canción. Por tanto, esta etiqueta indica el estribillo.

El resto de etiquetas siguen el mismo criterio para este estilo de música.

4.3.2. Herramienta software de etiquetado

Con el objeto de facilitar el rápido etiquetado de canciones, en el grupo de investigación se ha desarrollado una interfaz gráfica de usuario en MATLAB que permite acelerar el proceso de etiquetado de una canción. Esta herramienta ha sido desarrollada por un programador del grupo siguiendo las necesidades del proceso.

Dicha herramienta permite al usuario cargar una canción, escuchar sus partes y decidir dónde hay un evento determinando qué tipo es.

Para ver esto de forma más detallada, a continuación se describe el funcionamiento y el proceso en detalle.

El primer paso es cargar el programa que se ha creado en la función de Matlab con nombre “vmTaggin.m”.

Una vez que se ha ejecutado, aparece una ventana (Fig.2) donde se encuentran todas las etiquetas disponibles que se pueden marcar en una canción determinada. Además, también aparece todo un espacio en blanco que servirá para visualizar la forma de onda de la canción seleccionada.

Posteriormente, el siguiente paso es cargar la canción pulsando “Load Song” y aparece la lista de canciones disponibles para etiquetar (Fig.3) y se selecciona la canción que se quiera segmentar.



Fig.2. Software de etiquetado de canciones.

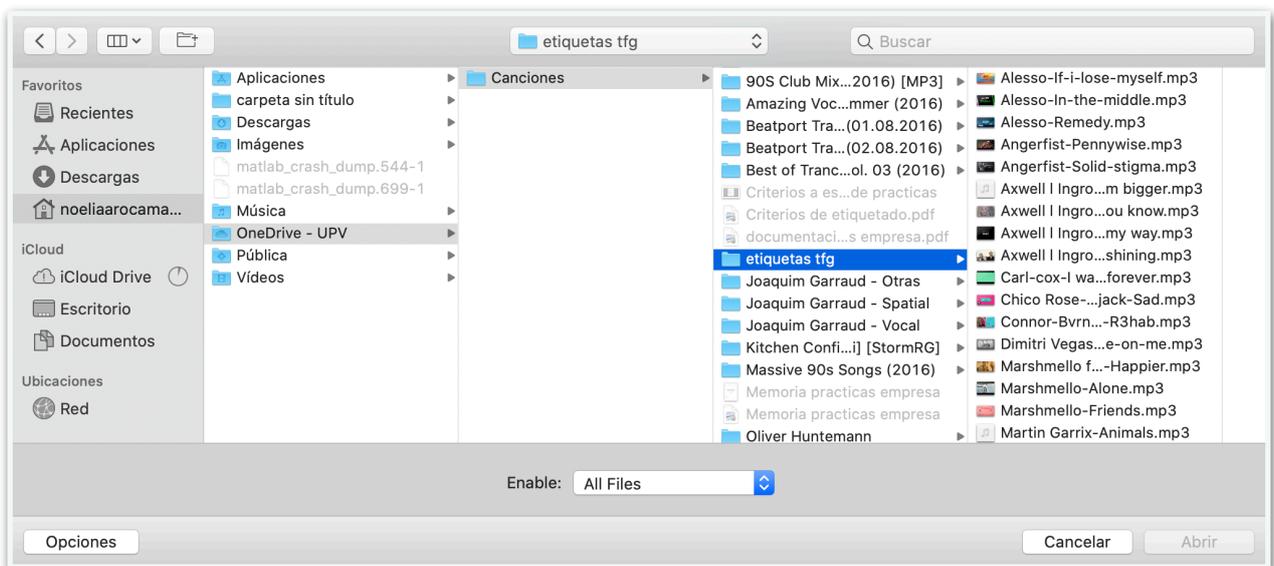


Fig.3. Selección de una canción.

Una vez se ha seleccionado qué canción se quiere proceder a etiquetar (en nuestro caso “Alesso-If-i-lose-myself.mp3”), aparece en el software la forma de onda de la canción escogida (Fig. 4).

A continuación, pulsamos “Play” para ir escuchando la canción. En el momento que percibamos un cambio en la música, volvemos a pulsar “Play” para pausar la canción y decidir qué etiqueta es la correspondiente a esa sección o instante.

Siguiendo el ejemplo de la canción anterior, se muestra a continuación como se van realizando las etiquetas a lo largo de la canción (Fig. 5).

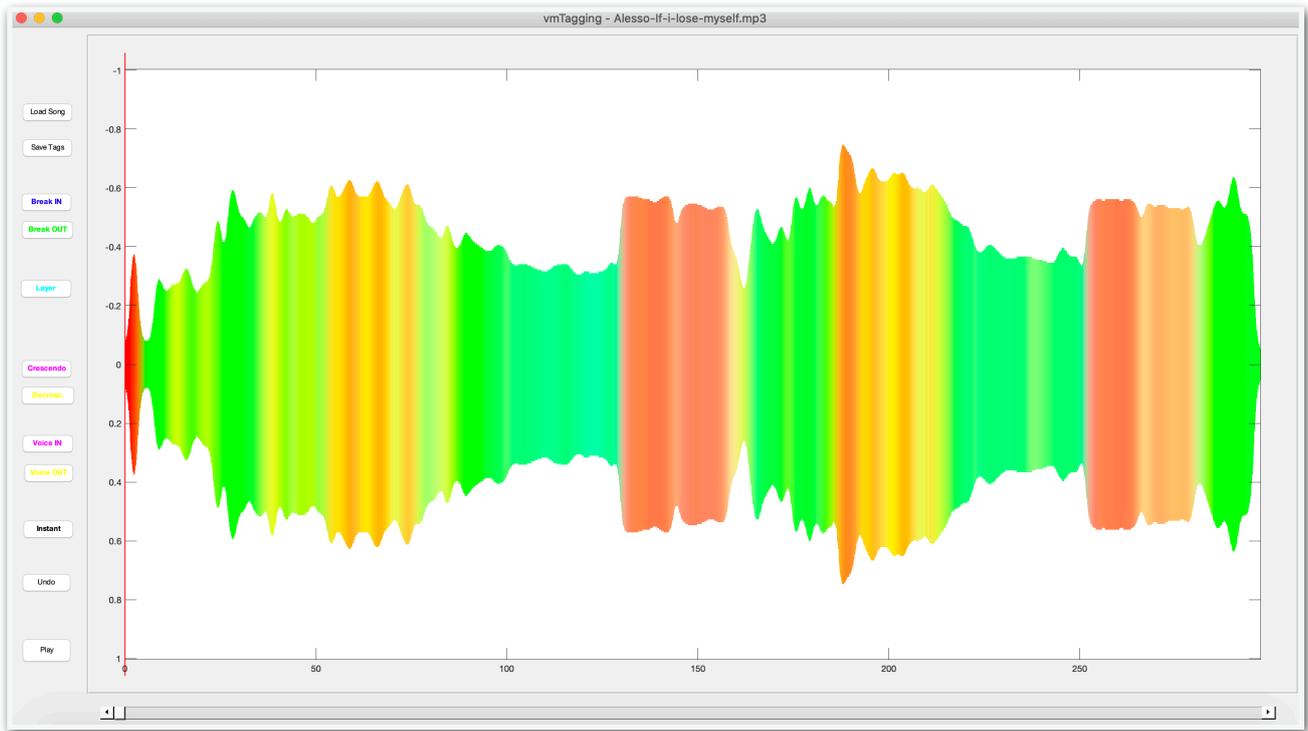


Fig.4. Forma de onda de la canción escogida para ser etiquetada.

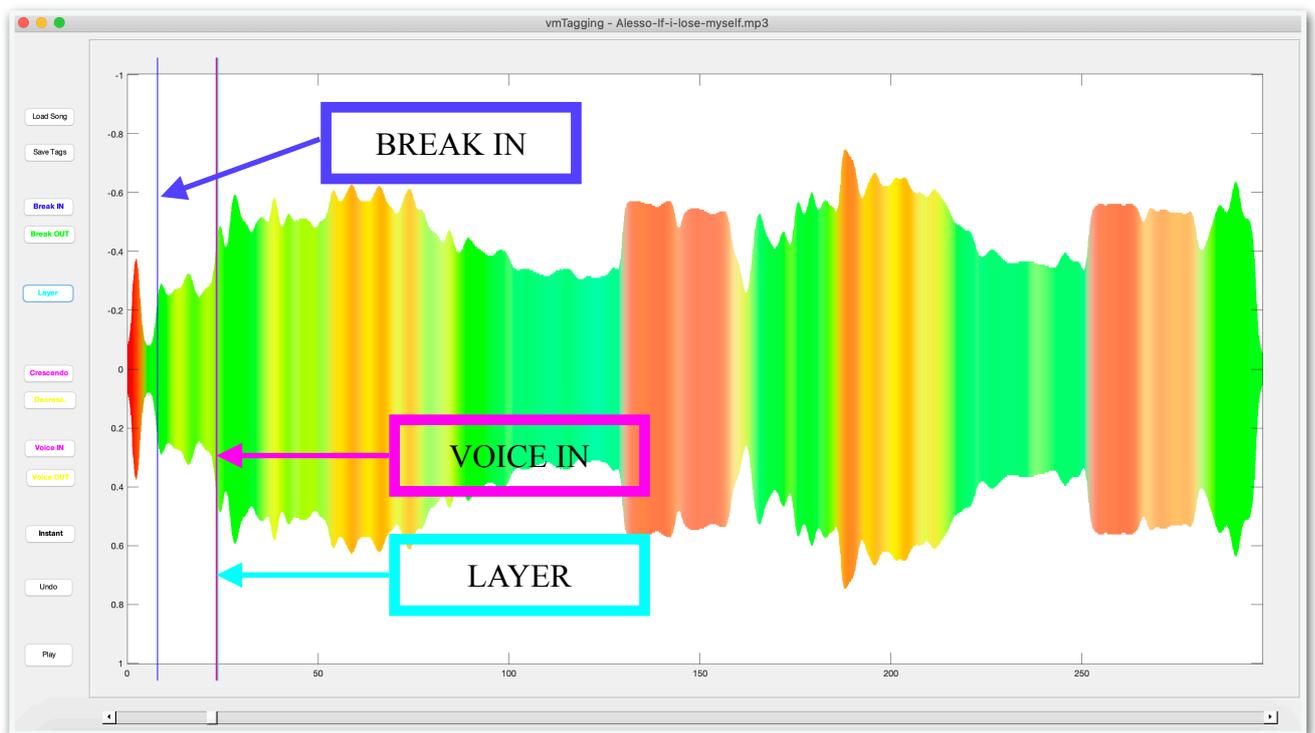


Fig.5. Etiquetado en una canción electrónica.

Observando la Fig.5, es fácil ver lo que sucede. Desde el momento que comienza la canción hasta que se produce el primer cambio, es la introducción. Después, hay un cambio y se ha elegido la etiqueta “Break In” porque sucede una parte poco energética en la canción. Tras el “Break In”, aparece una voz por lo que se pone la etiqueta “Voice In”. Pero, la voz también es un instrumento, por lo tanto, se ha añadido también una capa en la zona del “Break In”. Así, se ha decidido poner junto a “Voice In” la etiqueta de “Layer”.

Poco a poco se va recorriendo la canción y se va completando con etiquetas. La canción anterior con todas las etiquetas puestas quedaría de la siguiente manera (Fig.6).

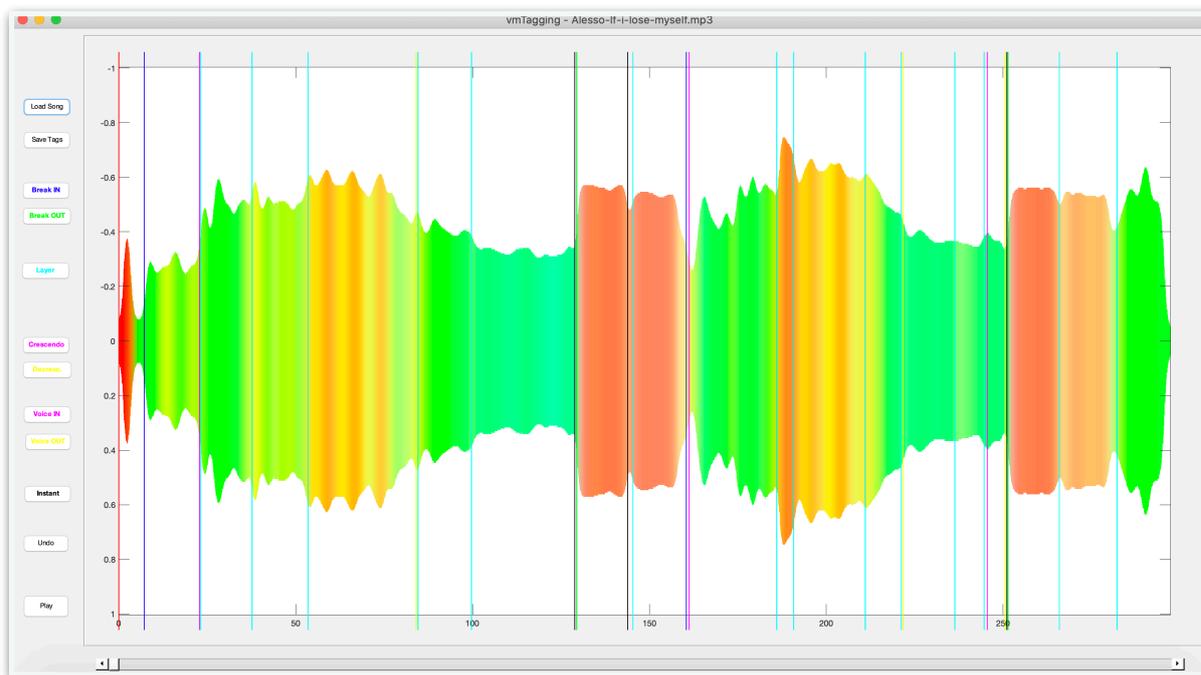


Fig.6. Etiquetado completo de una canción.

Si se observa la Fig.6, se puede ver que en una canción estándar de música electrónica se tiene una intro, dos zonas más pausadas o calmadas y dos partes en las que culmina la canción (“Break Out”). Además, también se puede decir que a lo largo de la canción van entrando y saliendo muchos instrumentos así como voces (“Layer”).

Por último, cuando se ha realizado todo el etiquetado de la canción que se ha escogido, se pulsa “Save Tags” para guardar las etiquetas realizadas. Si se accede a la carpeta donde está la canción seleccionada, se hallará un fichero “.txt”. En la Fig.7 se muestra el contenido de dicho fichero.

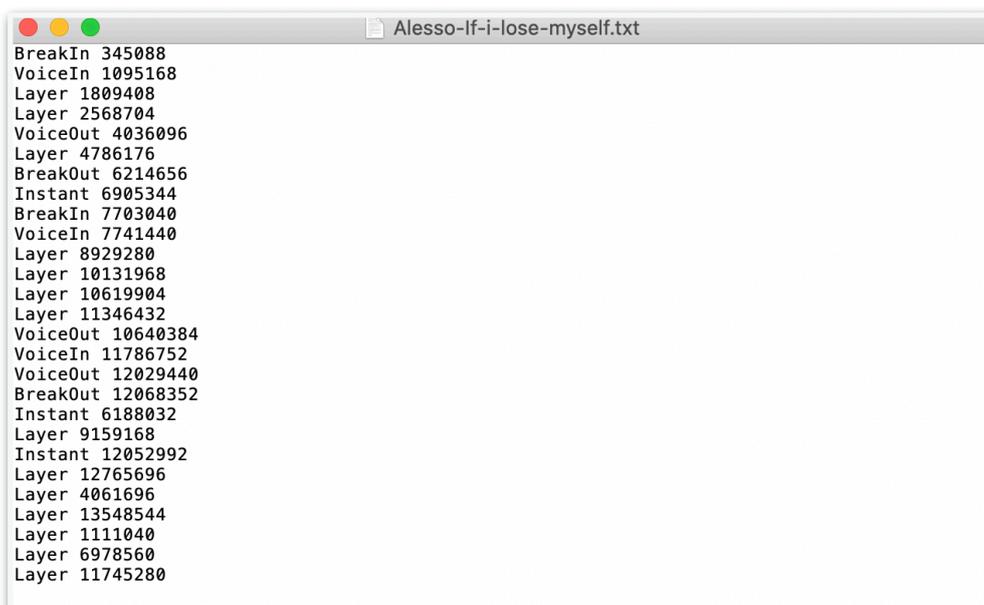


Fig.7. Etiquetas en su respectivo fichero “.txt”.

En la Fig.7 se muestran las etiquetas que se han realizado en la Fig.6. Se puede ver que aparece el nombre de la etiqueta al lado de un número. Dicho número es la muestra de la forma de onda a la que se le ha puesto una etiqueta.

Para averiguar cuál es ese instante en el tiempo, se divide el número de la muestra la cual se quiere saber su instante, entre la frecuencia de muestreo de la canción. Como es un archivo “.mp3”, su frecuencia de muestreo es 44100 Hz.

Veámoslo en el ejemplo de la primera etiqueta, cuyo número de la muestra es 345088:

$$instantemuestra1 = \frac{345088}{44100} = 7,825 \text{ s} \quad (1)$$

El instante de tiempo en el cual se ha puesto la etiqueta “Break In” es 7,825 segundos.

Esto se puede hacer con cualquiera de las muestras que se han etiquetado para saber el instante exacto en el tiempo.

Hay otras ocasiones en las que la música no cumple una estructura estándar y es una canción más experimental. En esos casos, hay veces que por ejemplo no hay “Break Out” o “Break In”, etc. A continuación, se muestra un ejemplo de este tipo de situaciones con la canción “Dominic Fike - 3 Nights.mp3” (Fig.8).

Observando la Fig.8, se puede ver que carece de partes donde hay “Break Out”. Esto se debe a que la canción es muy lineal y nunca hay una parte en la que la música se rompe y se perciba una clara diferencia de energía dando lugar al punto más álgido. Lo que ocurre en este tipo de canciones es que partiendo de su parte más calmada (Break In) se van añadiendo o quitando capas de instrumentos.

Como se puede ver en este ejemplo, la canción comienza con una introducción y posteriormente entra la parte del “Break In” donde también aparece una voz. Después de esto solo se van añadiendo o quitando capas de instrumentos hasta que desaparece también la voz (Voice Out) y acaba la canción.

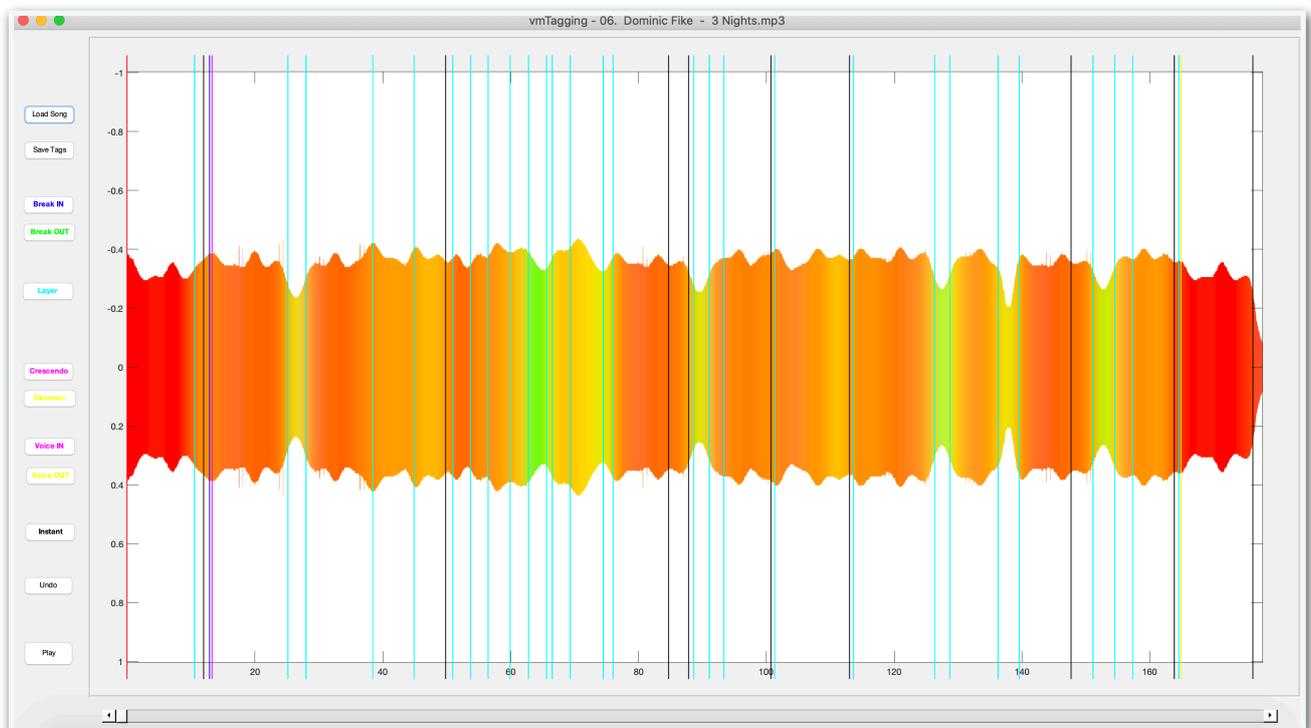


Fig.8. Etiquetado de una canción de música electrónica experimental.

Como se explicaba en el apartado 4.3.2, se ha decidido etiquetar algunas canciones del género pop/rock para ver la similitud en el proceso de segmentación con la música electrónica.

El proceso para etiquetar una canción de música pop o rock es el mismo que en el apartado anterior, pero a veces puede ser un proceso más complejo porque en el estilo mencionado se suelen emplear más instrumentos que van apareciendo o desapareciendo en la canción.

Es más sencillo comprobarlo si se observa la forma de onda así como la selección de elementos en una canción pop/rock determinada. Por ello, veámoslo con un ejemplo en la canción “queen - liar.mp3” (*Fig.9*).

Lo que se puede decir cuando se observa la *Fig.9* es que la canción es más complicada que una canción de música electrónica porque a simple vista se ve que se realizan muchos cambios en la música así como la cantidad de capas o instrumentos que se van añadiendo o van saliendo de la canción.

También, hay una introducción bastante larga porque no es hasta la muestra 4027904 o hasta el segundo 91,33 de la canción (siguiendo la fórmula (1)) cuando se produce la primera parte calmada de la canción o “Break In”.

Otra de las cosas que se pueden comentar es que se puede observar que hay muchos instantes donde sucede un determinado acontecimiento, es decir, de repente se oye un instrumento distinto en un instante muy corto de tiempo ya que hay muchas etiquetas de “Instant”.

Por tanto, aunque las etiquetas se hayan adaptado al tipo de música que no sea música electrónica se ve claramente que dependiendo del estilo de música que se elija, el proceso de etiquetado será más o menos complicado. Por ello, el estilo de música más sencillo para hacer un segmentado, poder llevar a cabo todo el proyecto y ver si es viable es la música electrónica. Por este motivo ha sido el principal género musical de este proyecto.

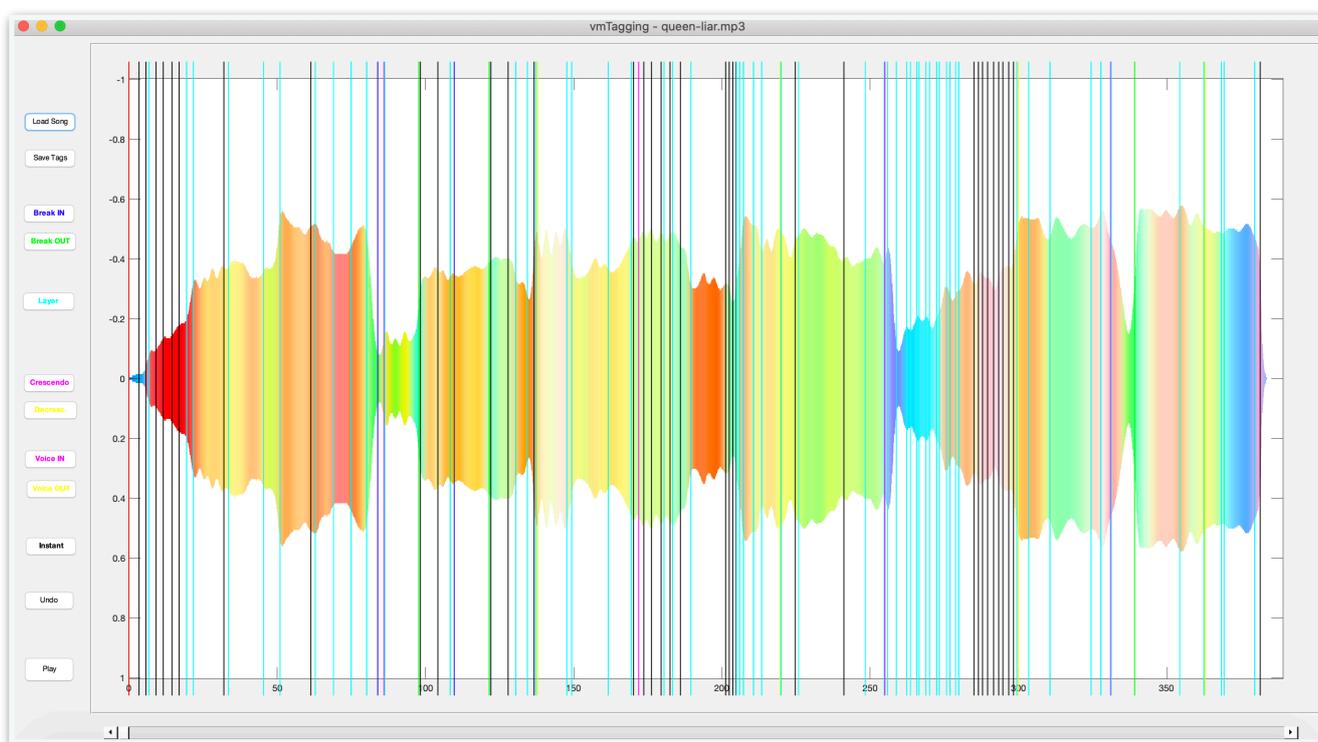


Fig.9. Etiquetado de una canción rock.

4.4. Observaciones y particularidades encontradas en la música EDM

A medida que se iba desarrollando el proyecto y se ponía atención a los sucesos que iban apareciendo en las diferentes canciones seleccionadas para segmentar, nos dábamos cuenta de que se podía llegar a una conclusión ya que aunque a veces las canciones proviniesen de sub-géneros distintos, estaban relacionadas entre sí dando lugar a ciertas particularidades que a grandes rasgos se presentaban. Por tanto, algunas de las particularidades que se han percibido y que son necesario comentar han sido:

- No suele haber letra o no es tan importante como en la música pop.
- No hay un estribillo como tal, si no que es más común presentar un loop con mucha energía que actúa como gancho de la canción.
- Una canción EDM se construye agregando y restando diferentes capas de instrumentos sobre la estructura del ritmo y conduciendo las emociones del oyente mediante diferentes trucos.
- No es necesario que se repitan las diferentes partes una y otra vez. Por el contrario, normalmente las partes son una evolución o modificación de partes anteriores.
- La identificación de esta estructura es más compleja y menos predecible cuando se trabaja con música de DJ en vivo. Las canciones no comienzan y terminan claramente porque el DJ elimina “Intros” y “Outros” para mezclar música continua, lo que dificulta la segmentación.
- Otra de las dificultades que se presentan es que hay diferentes sub-géneros dentro de la música EDM, como house, techno, dubstep, etc; y dentro de cada sub-género se cumplen unas reglas propias. Aunque como se ha explicado, no deja de haber particularidades comunes entre ellas.

4.5. Clasificación

Tras realizar el etiquetado de todas las canciones (en total se han etiquetado 462), el siguiente paso es procesarlas con el software de clasificación donde se verá cómo reconoce las etiquetas que se han hecho dentro de una determinada canción.

Para ello, se ha empleado un algoritmo de baja latencia para extraer dichos eventos realizados dentro de la estructura de una canción. Emplea un coste de computación bajo y es bastante eficiente en cuanto a funciones de detección se refiere.

Aunque la transformada de Fourier (FFT) se utiliza en técnicas de segmentación y clasificación, en este trabajo se han utilizado algoritmos cuya FFT no es necesaria, por lo que se ahorran costes computacionales así como la latencia. Sin embargo, aún así se obtiene la suficiente información para cumplir con los objetivos del proyecto.

4.5.1. Funcionamiento del software de clasificación

El primer paso es comprender cómo funciona todo el proceso que se lleva a cabo en el programa de clasificación de etiquetas. Por tanto, se ha creado un gráfico para entender de forma visual lo que ocurre. A continuación, se muestra el gráfico mencionado donde se explica de forma más clara (*Fig.10*).

De forma resumida (puesto que más adelante se muestra el proceso de forma más detallada), los pasos que se siguen según un análisis previo del programa son:

- a) Recibe una señal de audio a la entrada con una determinada forma de onda.
- b) Se realiza un inventariado con un tamaño de muestras de 735, si queremos saber los segundos aplicamos la fórmula (1), es decir, número de muestras entre 44100 que es la frecuencia de muestreo de la canción. Por tanto, 735 muestras entre 44100 son 16,6 milisegundos. Además, no hay solapamiento entre ventanas ya que tras realizar las primeras 735 muestras, la siguiente ventana comienza en 736.

- c) Posteriormente, esa señal se pasa por tres filtros aunque en el programa de Matlab aparecen cuatro (ya que es más fácil crear un Filtro Paso Bajo + Filtro Paso Alto construyendo así un Filtro Paso Banda). Por tanto, se pasa por un Filtro Paso Bajo, por un Filtro Paso Banda y por un Filtro Paso Alto.
- d) Tras realizar el filtrado, se realiza la energía de cada uno de lo obtenido, se calcula su logaritmo para pasarlo a dB y se normaliza para que en lugar de tener los resultados de 0 a -40dB, los tengamos entre 0 y 1.
- e) Así, obtendríamos las energías de cada parte. Sin embargo, como sigue siendo un proceso a una velocidad demasiado rápida para el oído humano, se vuelve a realizar otro filtrado para suavizar la señal pero sin modificar los puntos en los que la señal tenga un cambio brusco de energía.
- f) Esos cambios bruscos de energía son los instantes en los que sucede un evento en la canción segmentada. En este punto el programa aún no sabe diferenciar qué tipo de evento es el que sucede, pero sí sabe identificar dónde hay un evento.

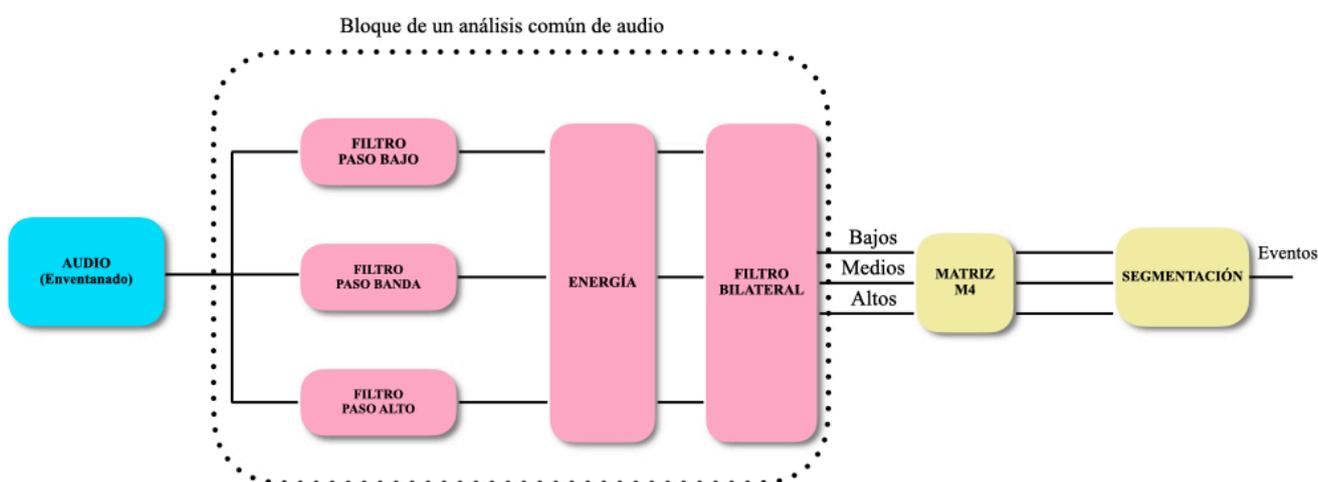


Fig.10. Gráfico del funcionamiento del programa de clasificación.

4.5.2. Extracción de las características

- FILTROS

La señal de audio se divide en tres bandas empleando filtros IIR de orden 4. La banda de frecuencias comprendida hasta 250 Hz corresponde al filtro 1 (filtro paso bajo), el filtro 2 (filtro paso banda) comprende las frecuencias entre 250 Hz y 4 kHz y, por último, el filtro 3 (filtro paso alto) es para frecuencias desde 4 kHz.

- ENERGÍA

Las tres bandas de frecuencia presentes (baja, media y alta) desemboca en una energía determinada para cada una de las ventanas calculándose y convirtiéndose a escala logarítmica.

- FILTRO BILATERAL

Para suavizar las características de energía obtenida de cada ventana con una latencia mínima se ha adaptado a campo del audio un algoritmo de procesamiento de imágenes denominado filtro bilateral.

Cuando una característica cambia de manera muy rápida en el tiempo para adaptarla para una animación o visualización, es necesario un suavizado. Utilizando un filtro bilateral, las salidas de las etapas anteriores son posibles de suavizar conservando los bordes.

El filtro bilateral se representa con $f(x)$ en la ecuación (2) y se resume como la combinación de un rango $s(f(\xi), f(x))$ y un filtro de dominio $c(\xi, x)$:

$$h(x) = k^{-1}(x) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi) c(\xi, x) s(f(\xi), f(x)) d\xi \quad (2)$$

Donde $k(x)$ es el factor de normalización:

$$k(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c(\xi, x) s(f(\xi), f(x)) d\xi \quad (3)$$

El filtro bilateral se ha aplicado a las 3 bandas de frecuencia.

- MATRIZ M4

Las frecuencias bajas, medias y altas se obtienen una vez se ha procesado el filtrado bilateral, por lo que después, se aplica una matriz de transformación M4 a los tres valores de energía para ponderarlos adecuadamente y proporcionar una mejor detección de las secciones cambiantes de la canción.

Por ello, se calcula la distancia euclidiana entre el vector de características actual y el anterior. El filtrado bilateral aplicado en las etapas anteriores garantiza cierta inmunidad a los pequeños cambios pero conservando los grandes, lo que mejora la detección.

La distancia se compara con un umbral establecido en el proceso de entrenamiento que se describe más adelante. Así, si se dispara uno de los umbrales anteriores, también se implementa un mecanismo para eliminar falsos positivos atendiendo a una distancia mínima entre eventos.

4.5.3. Clasificación

La principal fuente del algoritmo de segmentación empleado es el vector de características calculado en la etapa anterior. El proceso de segmentación sigue las siguientes etapas:

- 1) Calcula la distancia euclidiana entre el vector de características anterior y actual. Como se comentaba en el apartado anterior, el filtro bilateral que se aplica a las distintas etapas no afecta a los cambios pequeños de energía pero conserva los grandes, por lo que es muy útil para detectar los eventos.
- 2) La distancia mencionada en el punto anterior se compara con diferentes umbrales asociados con distintos eventos. Dichos umbrales se pueden ajustar dependiendo del estilo de música que se esté analizando y posteriormente se explica cómo se establecen.
- 3) Si uno de los umbrales se identifica, se evalúa un siguiente paso o etapa donde se trata de identificar con mayor precisión el evento, es decir, si es el inicio o el final de una ruptura, cambio de capa, etc. Como se hablaba anteriormente, se dispone de 3 bandas de frecuencia pero se analiza cada una de ellas por separado. También, se implementa un mecanismo de eliminación de falsos positivos utilizando una distancia mínima entre eventos.

- 4) Por último, una máquina de estados finitos mantiene el tipo de segmento que se ha aplicado a las partes de la canción e indica los cambios a otros estados en función de los eventos que se van desencadenando a lo largo de la canción.

5. Resultados del trabajo

Una vez definidos los parámetros a utilizar así como el algoritmo queda probarlo con toda la librería de canciones etiquetadas desarrollada en la sección 4.3.

5.1. Obtención de los umbrales

Es necesario hacer un ajuste muy preciso para establecer los umbrales mencionados. Este ajuste debe crearse utilizando muestras del estilo de música que vayamos a analizar.

Dividiremos el conjunto de canciones etiquetadas en dos: el conjunto de entrenamiento y el conjunto de validación. De las 460 canciones etiquetadas utilizaremos 200 para el conjunto de entrenamiento y 260 para la validación.

Finalmente, para el proceso de ajuste fino del algoritmo se optimizaron 4 parámetros: dos factores de ponderación (para la energía total y para la banda de baja frecuencia), el umbral de distancia entre características y vector y el tiempo mínimo entre eventos.

Se sigue la siguiente iteración:

- a) Toda la base de datos se plasma en una matriz de 4 dimensiones con cuatro valores equiespaciados para cada parámetro seleccionados dentro de un valor razonable obtenido por la pruebas preliminares. Esto totaliza $4^4 = 256$ posibles combinaciones de parámetros.
- b) La tasa de coincidencia y falsos positivos se calculan para cada combinación
- c) Cuando se obtienen los mejores resultados, la matriz se refina y se repite la tasa mencionada en el paso 2.
- d) El proceso se detiene cuando no se obtienen mejores resultados en varias iteraciones seguidas o cuando se ha alcanzado un número de interacciones máximo prefijado.

Un resultado del 100% de aciertos y de 0% de falsos positivos no puede llegar a obtenerse, por lo que el mejor resultado se calcula a partir de los casos en los que la tasa de coincidencia es superior al 90% y la tasa de falsos positivos es más baja.

En el gráfico que se muestra a continuación (*Fig. 11*), se plasman con un '*' las diferentes combinaciones de los 4 parámetros para el conjunto de entrenamiento. Cada '*' corresponde a un vector de 4 elementos de dichos parámetros. Se puede observar como es posible trazar una línea límite marcada en rojo a partir de la cual ya no es posible mejorar el algoritmo.

Dentro de los valores más próximos a la línea se puede escoger un vector que proporcione más aciertos o menos falsos positivos dependiendo de la aplicación o de lo que se desee.

En este proyecto, la tasa de falsos positivos es menos importante ya que, en la práctica encontrar cambios sutiles en la canción no causa alteraciones visuales en la animación o en las luces que se sincronicen, es decir, tienen más importancia los cambios grandes que son cambios de ruptura en la canción o añadido de capas.

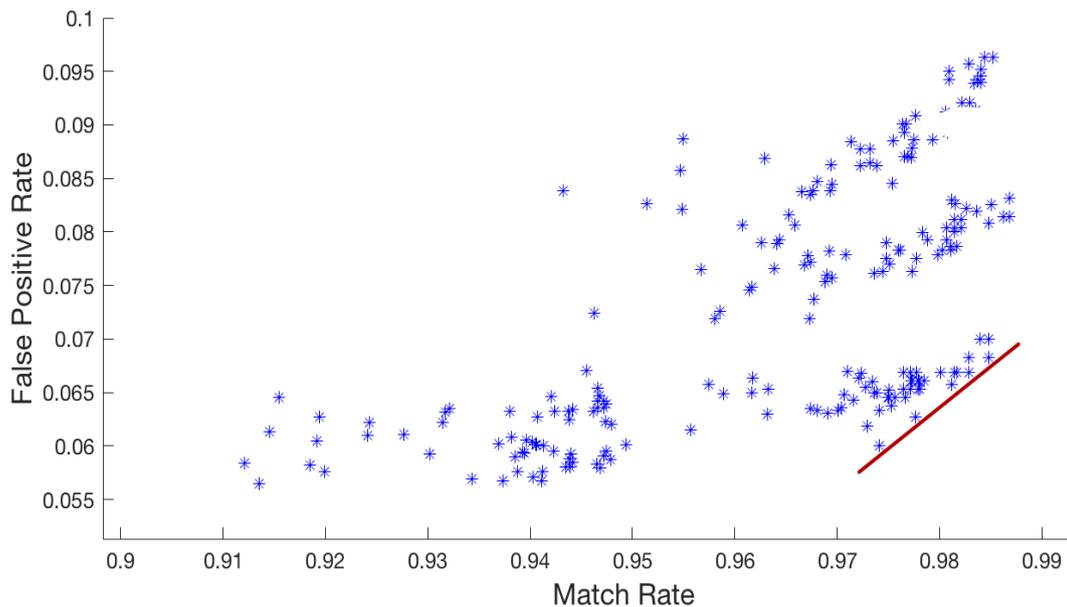


Fig.11. Tasa de acierto vs tasa de falsos positivos.

5.2. Validación del algoritmo con el conjunto de validación

Por último queda comprobar con el conjunto de validación que se obtienen resultados parecidos al conjunto de entrenamiento.

Dependiendo del vector de 4 parámetros que escojamos próximo a la línea roja los resultados también variarán con el conjunto de validación.

Hemos escogido los 4 vectores de parámetros más próximos a la línea roja (*Fig.12*) y hemos elaborado la siguiente tabla donde se ven los resultados de tasa de acierto y tasa de falsos positivos en el conjunto de entrenamiento y en el conjunto de validación (*Tabla.1*).

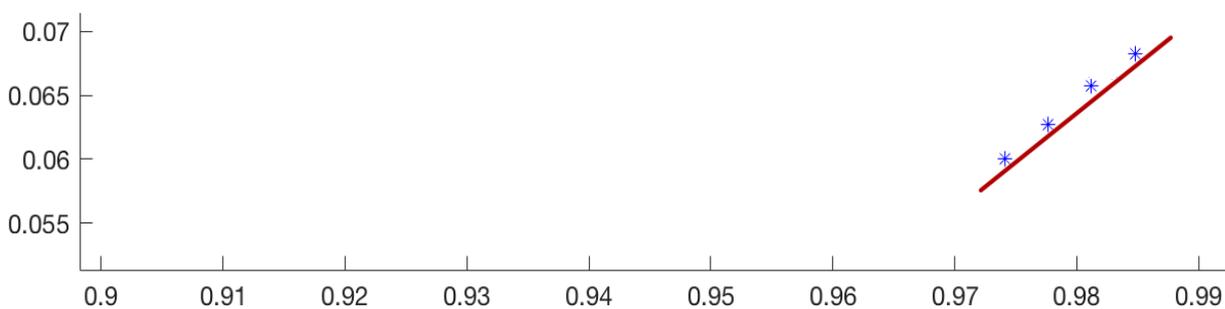


Fig.12. Vectores más eficientes para el conjunto de entrenamiento.

	% Acierto Conjunto de entrenamiento	% Falsa Alarma Conjunto de entrenamiento	% Acierto Conjunto de validación	% Falsa Alarma Conjunto de validación
Vector 1	97,4	6	95,9	6,4
Vector 2	97,9	6,2	96,1	6,9
Vector 3	98,1	6,5	96,7	7,1
Vector 4	98,5	6,7	96,8	7,9

Tabla.1 Tasa de acierto y tasa de falsos positivos en el conjunto de entrenamiento y el de validación.

Lo que se puede observar en la tabla es que los resultados con el conjunto de validación son solo ligeramente peores que con el conjunto de entrenamiento. Esto es normal en la mayoría de los sistemas de machine learning debido al efecto de sobreajuste (overfitting).

Siempre que la diferencia no sea muy grande el algoritmo será útil y generalizará bien.

6. Conclusiones y trabajo futuro

6.1. Conclusiones

En este proyecto se ha trabajado en el desarrollo de un sistema automático de segmentación musical basado en “machine learning”.

En particular nos hemos centrado en la música electrónica de baile (EDM), tanto porque nos pareció a priori más fácil de segmentar que otros tipos de música (como hemos corroborado a lo largo del desarrollo del proyecto) como por el interés comercial que puede suscitar un algoritmo en este sector del ocio en la actualidad.

A lo largo del trabajo se han desarrollado diferentes fases que han ido cumpliendo los objetivos marcados.

En una primera fase se estudió la música EDM definiendo su estructura y los tipos de eventos de cambio de sección que se producen en la misma y que han constituido el objetivo de segmentación en este trabajo.

Durante varios meses se realizó un meticuloso trabajo de compilación de una base de datos extensa de EDM intentando englobar todos los estilos, así como diferentes épocas de la misma, incorporando canciones de los 90, 2000, 2010 y 2020.

Después se etiquetó manualmente para que sirviera de aprendizaje al sistema de machine learning. Esta labor ha sido muy importante y compleja, ya que hay que tomar muchas veces decisiones que requieren un conocimiento musical y mis conocimientos musicales han sido muy útiles para esta labor.

Con la ayuda y soporte de los conocimientos del grupo de investigación donde se ha desarrollado este proyecto, se definió un algoritmo de clasificación que incluía una extracción de parámetros y una clasificación.

Con la base de datos se creó un subconjunto de entrenamiento para realizar el ajuste fino de los umbrales del clasificador para detectar los cambios de sección. Posteriormente, se validó con un subconjunto de validación.

Los resultados obtenidos son excepcionales consiguiendo tasas de detección superiores al 98%, con falsas alarmas de alrededor del 5%. Como hemos mencionado en el trabajo, las falsas alarmas no suponen un gran problema en esta aplicación, por lo que el sistema podría ser plenamente funcional en un entorno real de trabajo.

6.2. Trabajo futuro

En este trabajo nos hemos enfocado en un estilo de música concreto. Sin embargo, utilizando el conocimiento desarrollado se podría extender a otros tipos de música más compleja.

En cualquier caso, al haber etiquetado e incorporado en los conjuntos de entrenamiento y validación algunas canciones con un estilo menos EDM y más pop/rock ya hemos iniciado en cierta medida esta vía, pero quedaría mucho trabajo para incorporar canciones muy diferentes al EDM.

Otro aspecto a mejorar en el futuro, sería utilizar clasificadores basados en redes neuronales, que modelaran mejor los cambios de sección. Ello permitiría incorporar muchos más parámetros al conjunto de características, que finalmente se quedó en tan solo 4 parámetros por simplicidad, pero que el uso de redes neuronales nos permitiría extender y obtener resultados aún más eficientes.

7. Bibliografía

- Gestion proyectos

[1] <https://www.ticportal.es/glosario-tic/gestion-proyectos>

[2] <https://searchdatacenter.techtarget.com/es/definicion/Gestion-de-proyectos-definicion>

- Segmentación musical (Segmentos de una canción)

[3] <http://musifica.com/2013/11/15/que-es-la-forma-en-musica/>

[4] <https://www.mariajesusmusica.com/inicio/estructura-musical-dos-esquemas-sencillos-para-el-analisis-musical-en-clase>

[5] <https://es.wikipedia.org/wiki/Fraseo>

- Segmentación música electrónica

[6] <https://cymatics.fm/blogs/production/edm-song-structure>

[7] <https://majomontemayor.com/diccionario-de-musica-electronica/>

[8] <https://barbeofficial.wordpress.com/2015/12/07/terminos-mas-utilizados-dentro-de-la-musica-electronica/>

[9] <https://www.tutoca.com/la-estructura-de-la-musica-electronica/>

[10] <https://www.audioproduccion.com/los-4-principales-sub-generos-del-edm-identificarlos/>

[11] <https://www.complex.com/music/an-idiots-guide-to-edm-genres/>