Pontificia Universidad Javeriana Cali
Facultad de Ingeniería.
Ingeniería de Sistemas y Computación.
Anteproyecto de Grado.

Reconocimiento, segmentación y extracción de la estructura musical en canciones populares

Juan Camilo Arevalo Arboleda

Director: Dr. Gerardo Mauricio Sarria

13 de Noviembre del 2016

Santiago de Cali, 13 de Noviembre del 2016.

Señores
Pontificia Universidad Javeriana Cali.
Dr. Andres Navarro Newball
Director Carrera de Ingeniería de Sistemas y Computación.
Cali.

Cordial Saludo.

Por medio de la presente me permito informarle que el estudiante de Ingeniería de Sistemas y Computación Juan Camilo Arevalo Arboleda (cod: 199384) trabaja bajo mi dirección en el proyecto de grado titulado "Reconocimiento, segmentación y extracción de la estructura musical en canciones populares".

Atentamente,

Dr. Gerardo Mauricio Sarria

Santiago de Cali, 13 de Noviembre del 2016.

Señores
Pontificia Universidad Javeriana Cali.
Dr. Andres Navarro Newball
Director Carrera de Ingeniería de Sistemas y Computación.
Cali.

Cordial Saludo.

Me permito presentar a su consideración el anteproyecto de grado titulado "Reconocimiento, segmentación y extracción de la estructura musical en canciones populares" con el fin de cumplir con los requisitos exigidos por la Universidad para llevar a cabo el proyecto de grado y posteriormente optar al título de Ingeniero de Sistemas y Computación.

Al firmar aquí, doy fe que entiendo y conozco las directrices para la presentación de trabajos de grado de la Facultad de Ingeniería aprobadas el 26 de Noviembre de 2009, donde se establecen los plazos y normas para el desarrollo del anteproyecto y del trabajo de grado.

Atentamente,

Juan Camilo Arevalo Arboleda
Código: 199384

Resumen

La segmentación de la estructura de las canciones resulta interesante para ayudar y guiar a un nuevo oyente o inexperto en la musica, partiendo y seccionando la canción en partes que desea escuchar o simplemente para tener clara las secciones para una futura interpretación sin la necesidad de tener que repetir muchas veces una pieza musical. Este proyecto se enfoca en desarrollar un método para encontrar, seccionar y extraer las secciones de las canciones ¨populares¨.

Palabras Clave: Musica, Alineación de Secuencias, Canciones populares, Salsa, Sección de canciones.

Índice general

Introducción

Las canciones están compuestas por 3 componentes: la melodía, el ritmo y el armonía. El ritmo define la duración de sonidos y sus silencios, lo cual sirve para definir la melodía, la cual es el elemento mas perceptible de una composición y es lo principal en el momento de las composiciones ya que esta es un conjunto de notas dentro de una escala, que acorde a dicha escala, se define la armonía, que se refiere a un acompañamiento, el uso de notas simultaneas para formar un acorde y dar una mejor experiencia al oyente.

Muy pocas piezas musicales son una progresión continua de la armonía y melodía en toda la duración de la canción, por lo general, en las canciones populares, la armonía y la melodía se desarrollan dentro de una sección. La estructura de las composiciones populares es la forma en que sus secciones se han organizado generalmente, de manera repetitiva para crear toda una pieza musical, esto con el motivo de dar una apropiada experiencia eficaz al oyente y facilitar la composición seccionando las canciones.

Las estructuras en las canciones populares usualmente están dadas por introducciones, versos, pre-coros, coros, solos instrumentales, puentes entre secciones y las salidas del tema. En el caso de los coros es muy peculiar ya que generalmente es la sección mas representativa de las canciones por que guardan la misma estructura en cada repetición que tenga y es la sección mas repetida, y que por tanto contiene las características mas relevantes de las canciones. Aun que estas estructuras son comunes de encontrar, no podemos generalizarlo para todas las canciones, ya que existen canciones cuya estructura es muy variada y sus partes no guardan ninguna correlación con otras de la misma pieza musical, como es en el caso de las composiciones en el sub-genero "power metal", por tal motivo el énfasis de este proyecto se dará en composiciones de canciones populares, es decir, para algunos géneros musicales que son generalmente llamativos para el publico, que son el rock, pop y en nuestro caso particular, la salsa.

En lo general, la estructuración de las piezas de una canción es de los últimos procesos en la composición la cual le termina de dar forma a las canciones, y que de manera inversa, para la transcripción o interpretación de las canciones es el primer proceso a realizar. Para una persona normal buscar estas secciones y determinar una estructura al escuchar una canción, resultaría una tarea confusa de identificar, y para un musico, hacer la correcta identificación de estas partes llevaría a tener que escuchar gran porción de estas piezas musicales y en algunas composiciones, llevaría a hacer varias revisiones a la canción puesto que su estructura podría tener distintas variaciones. La automatización de esta tarea de manera eficiente y correcta ahorraría tiempo en la construcción de partituras o para hacer la debida reescritura de una canción.

Capítulo 1

Descripción del Problema

1.1. Planteamiento del Problema

El concepto de interacción humano computadora refiere a las diversas formas y interfases con las que un humano pueda compartir e intercambiar información con un sistema de computo por medio de dispositivos (pantallas, joysticks, teclados, touch, etc.), en el área de los videojuegos se vuelve un área de estudio fundamental para lograr la mejor inmersión de un usuario en un entorno virtual, de manera que este tenga una mayor interacción con un mundo recreado por medio de contenido multimedia, lo que podríamos llamar como una realidad virtual, buscando que dichos dispositivos cubran la mayoría de nuestros sentidos para tener una interacción implícita y una inmersión sensorial en tiempo real.

Lo que es considerado como la primera generación de realidad virtual tiene muy bien trabajado la visualización 3D, como por ejemplo un juego de escalada de una montaña en primera persona usando como dispositivo interfaz el oculus rift, dicho hardware permite a la persona tener una vista en 3D dependiendo de la inclinación de la cabeza del usuario muestra el panorama correspondiente en el mundo virtual, en la que el jugador asciende los picos mas altos de cada continente equipado por 2 hachas[DPC+14]. Hasta el momento, los vídeo juegos se han centrado en hacer la parte visual lo mas fiable posible, pero de lograr una verdadera inmersión se debe buscar abarcar muchos mas sentidos del ser humano.

Además del contenido visual, el audio es otro factor importante en la inmersión, ya que acompañado de un componente visual, dará una buena recreación en la que se puede dar una espacialización mucho mas fina del usuario en el mundo virtual. Lo mas fiable que existe en sonido tanto como para cine y vídeo juegos, es el sonido envolvente 7.1, la se basa en dar una calidad sonora de una fuente de audio con canales adicionales provenientes de dispositivos que emitan sonido en un plano horizontal con un radio de 360 grados, esto adicionalmente requiere de hardware especializado para poder tener dicha fiabilidad de sonido. En los videojuegos existen software que emulan un sonido envolvente 7.1 con unos audífonos promedio o de la marca del fabricante a su preferencia, como por ejemplo el Razer Surround Personalized 7.1 Gaming Audio Software.

Aun que se logre emular el sonido envolvente 7.1, no es lo suficientemente preciso como para poder dar la ubicación exacta de un elemento en el mundo virtual, no logra definir la altura, distancia y ángulo de donde proviene la fuente de sonido, es decir, una verdadera espacialización 3D del audio. Por ello es importante automatizar estas propiedades auditivas desde el mismo motor del vídeo juego, para que este pueda ser usado por todos los usuarios con dispositivos accesibles para ellos y que de una espacialización del sonido mejor que el sonido envolvente. Existen herramientas que pueden hacer la simulación de dicha espacialización , una de ellas es HRIR, desarrollada en

Pure Data [Vil15], pero no está implementada para ser usada desde un motor de vídeo juegos, por tal motivo, es imperativo crear la interfaz entre esta herramienta y un motor de vídeo juegos con un gran numero de usuarios, el cual seria Unity.

1.1.1. Formulación

¿Como desarrollar una interfaz que pueda compartir la espacialización de las fuentes de sonido entre Unity y Pure Data?

1.1.2. Sistematización

1.2. Objetivos

1.2.1. Objetivo General

Desarrollar un plug-in entre Pure Data y Unity para lograr espacialización en 3D de las fuentes de sonido configuradas en Unity.

1.2.2. Objetivos Específicos

1.3. Justificación

Debido a que muchos de los emuladores de espacialización en 3D no logran ser muy precisos, en cuanto a una distancia desde la fuente de sonido aloyente.

1.4. Delimitaciones y Alcances

Puesto que dependemos de HRIR para hacer la espacialización de sonido, el Plug-In estará limitado a sus restricciones, las cuales son que la distancia Max. seria de 160 cm, y debido a que nos centraremos en la espacialización de la fuente de audio, se ignorara temas como la reververacion, delays y otros efectos que pueda tener una fuente de sonido en cualquier esceneario posible.

1.4.1. Entregables

- Plug-in para unity con el cual pueda usar la herramienta HRIR en Pure Data para dar una espacializacion en 3D de las fuentes de audio en el juego.
- Documento referente al trabajo de grado.

Capítulo 2

Desarrollo del Proyecto

2.1. Marco de Referencia

2.1.1. Áreas Temáticas

- H.5.5 Sound and Music Computing
- B.2.4 High-Speed Arithmetic
- F.2.2 Nonnumerical Algorithms and Problems

2.1.2. Marco Teórico

La estructura musical es la organización coherente del material utilizado por el artista, mas lo abstracto de la misma le hace al compositor la tarea mas difícil. La causa por la cual el compositor se preocupa por mantener un molde de la estructura de su obra, es que a pesar de salirse de moldes establecidos, una obra tendrá que apoyar su estructura en una explicación coherente de ella, para así darle sentido a su organización y no sea esta hecha al azar, pero siempre dar una importancia a que dicha estructura sea coherente con el objetivo de la musica y por tanto unifique el resultado[C+61].

Para la estructura musical, existe el principio de la repetición, la cual sirve de apoyo a la musica por el medio de diversas interpretaciones del mismo principio y es el recurso mas ampliamente desarrollado que ha dado pie a gran numero de obras. Existe también el principio contrario de la no repetición que es usado para piezas breves por su difícil tratamiento. En las canciones populares, con tal de llegar al oyente, se usa el principio repetitivo por secciones, las cuales generalmente están dadas por:

- **Introducción:** Se encuentra siempre al inicio de la obra, la idea es preparara al oyente y guiarlo hacia un tema o melodía principal o simplemente a la primera parte donde entra la voz o instrumental ejecutando acordes principales de la obra en una pequeña melodía.
- **Tema o estrofa:** Están compuestas por 2 o mas versos, particularmente de 4 o 6, se le llama tema por que la letra y el acompañamiento armónico puede cambiar, mientras que la armonía se mantiene, al momento de cambiar la melodía se da a entender que es un nuevo tema.
- **Interludio:** Mayoritariamente instrumental que une dos partes dela canción creando una conexión armónica entre ellas, en ocasiones suele repetir la melodía de la introducción y es utilizada mayormente como puentes entre secciones.

- **Estribillo o coro:** Es la sección, generalmente, mas representativa de la canción. La letra y la melodía se repite dos o mas veces en la obra y es la parte donde el compositor expresa la idea principal de la canción con respecto a la letra y melodía.
- **Solo instrumental:** Es la sección diseñada para destacar uno o mas instrumentistas, en algunas composiciones pueden existir 2 o mas solos, se desarrollan sobre los acordes de la estrofa, del estribillo o el interludio.
- **Coda o outro:** Es la parte final de la composición que con frecuencia repite el interludio o la introducción con unas cuantas variaciones.

Para poder buscar estas secciones de una manera eficiente, como lo antes propuesto en un trabajo previo[AMMAL15], la idea sera usar la transformada rápida de fourier (FFT), que nos permite para una porción de la onda de la canción, cambiar el dominio de tiempo a frecuencia de la señal y poder extraer la frecuencia mas dominante en esa parte y poder asociar dicha frecuencia a una nota musical[BGK11] para poder agrupar la canción por unas 255 partes aprox. que representen las notas musicales y no tener millones de puntos en una onda. Teniendo dichas agrupaciones se procede a hacer uso de un algoritmo de alineación de secuencias para encontrar las sub-secuencias que se repitan con otra sección de la canción[HDT03], para luego ser comparadas a su vez con otras secciones que cumplan un patron y revisar que la correlación entre ellas sean mayor al 65 % para determinar que hemos encontrado mas de una repetición de una sección[AMMAL15].

2.1.3. Trabajos Relacionados

En lo general, los trabajados se han enfocado en la extracción del coro puesto que idealmente se busca extraer la sección mas representativa de las piezas musicales. Un primer acercamiento para encontrar un estribillo de una canción es el uno de Bartsch y Wakefield [BW01]. Este enfoque supone que "Partes fuertemente repetidas de una canción corresponden al estribillo, o parte importante de una canción ". Así que el algoritmo primero segmenta la canción en marcos de datos de audio, y para cada trama calcula un vector de características con los tonos musicales. Luego se calcula la correlación entre Cada par de vectores y construye una matriz de similitud. El Algoritmo finaliza analizando las diagonales de la matriz y seleccionando la sección similar más larga. Masataka Goto en [Got06] propuso un algoritmo llamado RefraiD. Este algoritmo extrae primero un vector de características (el Chroma vector) de cada trama de la canción. Cada elemento del vector croma corresponde a uno de los 12 semi tonos musicales (C, C , D, D , E, F, F , G, G , A, A y B). Entonces, las similitudes entre los vectores cromáticos se calculan, las secciones repetidas se enumeran y agrupan analizando su relación en toda la canción. Finalmente, el grupo de coros es seleccionado con la medida más alta de otros grupos de secciones repetidas. Por otro lado Yeh et al. En [YLLT10] propuso otra algoritmo para extraer coro basado en una generación de mapa de color. La idea es similar a la de Goto, para cada cuadro de la canción, los vectores de características se extraen calculando la energía de intensidad, banda alta y banda baja en la dominio de la frecuencia. Los vectores de características se asignan a la R, G, B dominio del espacio de color para obtener el mapa de color de la canción y poder representar la estructura, las regiones con distribución de

color similar se agrupan. Entonces, el coeciente cepstrales en las frecuencias de Mel (MFCCs) se extraen de cada región obtenida del mapa de colores como la característica para la clasificación de las secciones de verso o coro.

2.2. Metodología

2.2.1. Tipo de Estudio

Este proyecto se enmarca en un esquema de estudio tipo científico, pues la idea es encontrar un método eficiente en la búsqueda y extracción de secciones en una canción probándolo en una base de datos de canciones de salsa.

2.2.2. Actividades

La idea del trabajo es una refinación de un estudio previo ya realizado en el cual se especializaba en el reconocimiento de la parte mas representativa en las canciones de salsa(citarme) y extenderlo para que pueda reconocer mucho mas que los coros en las canciones.

[1]Analizar y estudiar en que secciones en la alineación de secuencias se pueden encontrar las secciones de una canción. Implementar la búsqueda y extracción de estas secciones. Realizar la prueba con una muestra representativa de la base de datos de canciones de salsa (375 canciones). Iterar y re ajustar parámetros en las correlaciones y la búsqueda de secciones.

2.3. Resultados Esperados

Se espera que al final un software pueda encontrar todas las secciones de una canción con una precisión que sea mayor al 70 % y que dicho método sea eficiente.

2.4. Recursos

2.4.1. Humanos

Dr. Gerardo Sarria y profesor en la Universidad Javeriana Cali quien con el grupo de investigación DESTINO de la misma universida, a investigado sobre la caracterización computacional de las canciones de salsa, enfocado para realizar la minería de datos en un gran conjunto de canciones y así construir un sistema que modela este género musical y reconoce y clasifica viejas y nuevas canciones de salsa usando técnicas de aprendizaje de máquina.

Alfredo Moreno, estudiante y musico egresado del EAFIT Medellin, Colombia, quien también trabaja con el Dr. Sarria en el proyecto de investigación para caracterizar computacional de las canciones de salsa

2.4.2. Técnicos

Se necesitara de un Mac para hacer pruebas pequeñas y posteriormente para poder seccionar un conjunto grande de canciones, lo recomendable seria usar el cluster de la universidad para dicha tarea.

. Bibliografía

[AMMAL15] C. Arévalo, G. M. S. M., M. J. Mora, and C. Arce-Lopera. Towards an efficient algorithm to get the chorus of a salsa song. In *2015 IEEE International Symposium on Multimedia (ISM)*, pages 258–261, Dec 2015.

[BGK11] Gavin M Bidelman, Jackson T Gandour, and Ananthanarayan Krishnan. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, 23(2):425–434, 2011.

[BW01] Mark A Bartsch and Gregory H Wakefield. To catch a chorus: Using chroma-based representations for audio thumbnailing. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 15–18. IEEE, 2001.

[C+61] Aaron Copland et al. *Cómo escuchar la música*. 1961.

[DPC+14] Tristan Dufour, Vincent Pellarrey, Philippe Chagnon, Ahmed Majdoubi, Théo Torregrossa, Vladimir Nachbaur, Cheng Li, Ricardo Ibarra Cortes, Jonathan Clermont, and Florent Dumas. Ascent: A first person mountain climbing game on the oculus rift. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play*, CHI PLAY '14, pages 335–338, New York, NY, USA, 2014. ACM.

[Got06] Masataka Goto. A chorus section detection method for musical audio signals and its application to a music listening station. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1783–1794, 2006.

[HDT03] Ning Hu, Roger B Dannenberg, and George Tzanetakis. Polyphonic audio matching and alignment for music retrieval. *Computer Science Department*, page 521, 2003.

[Vil15] Julián Villegas. Locating virtual sound sources at arbitrary distances in real-time binaural reproduction. *Virtual Reality*, 19(3):201–212, 2015.

[YLLT10] Chia-Hung Yeh, Yu-Dun Lin, Ming-Sui Lee, and Wen-Yu Tseng. Popular music analysis: chorus and emotion detection. *Proc APSIPA ASC*, 2010.