

Procesamiento Digital de Señales de Audio

Instituto de Ingeniería Eléctrica - UdelaR

Proyecto de fin de curso - 2022

- La realización del proyecto es en grupos de dos estudiantes.

1. Introducción

Este documento describe la propuesta de proyecto de fin de curso para los estudiantes de grado. Se aborda una serie de problemas de extracción de información musical (Music Information Retrieval, MIR) a partir de grabaciones de audio, sobre un cierto corpus de música, utilizando algunas de las técnicas estudiadas durante el curso. La información obtenida tiene diversas aplicaciones, como por ejemplo en la recomendación automática de música.¹

1.1. Corpus musical

Se analiza música del grupo *El Cuarteto de Nos*². En particular, dos de sus discos: *Otra Navidad en las Trincheras* de 1994 y *El Cuarteto de Nos* de 2004. Tienen un total de 36 canciones, varias de las cuales aparecen en ambos discos.

1.2. Objetivos, etapas y alcance

En el proyecto se busca estudiar una forma de reconocer versiones de una misma canción y alinearlas temporalmente. Para ello, primero se calcula una representación intermedia mediante el *cromagrama* [1]. Luego, utilizando esta representación intermedia se aplica la técnica de Dynamic Time Warping (DTW) para alinear secuencias temporales [1]. Finalmente, dado el alineamiento entre dos versiones de la misma canción, se genera audio sincronizado usando técnicas de deformación temporal, como *phase vocoder*.

El proyecto está organizado en dos etapas. En la primera, se aborda el problema de identificar versiones de una misma canción a través de la comparación de la distancia DTW resultante del alineamiento de dos archivos de audio. Una vez identificadas dos versiones de una misma canción, en la segunda etapa se busca generar archivos de audio alineados para ser reproducidos de forma sincronizada.

Se recomienda regular el alcance de las tareas a lo largo del proyecto de forma incremental. Por ejemplo, se sugiere comenzar por calcular la distancia DTW solo entre dos archivos de audio, que incluso pueden ser fragmentos elegidos manualmente. Una vez analizadas las dificultades asociadas a resolver la tarea

¹<http://ismir2018.ircam.fr/pages/events-tutorial-11.html>

²https://es.wikipedia.org/wiki/El_Cuarteto_de_Nos

y su costo computacional, se recomienda aumentar el alcance y la complejidad. Algunos de los objetivos del proyecto se pueden lograr a nivel de prueba de concepto. Por ejemplo, comparar exhaustivamente todas las canciones es computacionalmente costoso, pero para mostrar y analizar resultados puede ser suficiente considerar solo un subconjunto relevante de canciones.

Se recomienda consultar los Notebooks utilizados en la clase sobre sincronismo de audio, así como el capítulo 3 del libro *Fundamentals of Music Processing* [1] y los Notebooks asociados.³ Pueden resultar de utilidad las bibliotecas `librosa`⁴, `synctoolbox`⁵ y `libtsm`⁶.

2. Primera etapa

A continuación se detallan las tareas de la primera etapa.

2.1. Tareas

1. Identificar manualmente versiones de una misma canción en distintos discos. Escuchar y analizar similitudes y diferencias entre ellas. Discutir sobre las dificultades que plantean las diferencias para resolver la tarea.
2. Seleccionar algunos fragmentos de audio correspondientes en cada archivo de audio. Calcular y comparar el *cromograma* de los fragmentos. Analizar el efecto de los diferentes parámetros en la representación obtenida. Identificar limitantes y analizar posibles mejoras de la representación.
3. Implementar un algoritmo para detectar versiones de una misma canción, usando *cromograma* como representación intermedia y una distancia calculada mediante DTW. Utilizar un umbral sobre la distancia.
4. Usar el algoritmo implementado para detectar automáticamente dos versiones de una misma canción. Analizar el impacto de los parámetros de la técnica de alineamiento en el resultado obtenido. Identificar limitantes del método de alineamiento y considerar posibles mejoras.
5. Analizar la viabilidad del enfoque para la detección automática de versiones. Discutir los resultados obtenidos y las dificultades encontradas.

3. Segunda etapa

A continuación se detallan las tareas de la segunda etapa.

3.1. Tareas

1. Elegir dos versiones de una misma canción para las cuales las técnicas de identificación automática funcionan razonablemente. Analizar y describir

³<https://www.audiolabs-erlangen.de/resources/MIR/FMP/C3/C3.html>

⁴<https://librosa.org/>

⁵<https://github.com/meinardmueller/synctoolbox>

⁶<https://github.com/meinardmueller/libtsm>

sus diferencias más importantes, como la utilización de distintos instrumentos, la velocidad de interpretación, secciones diferentes, etc.

2. Hallar la correspondencia temporal entre ambas versiones usando las técnicas de la etapa anterior. Analizar el impacto de los parámetros del algoritmo en el alineamiento y comparar con los resultados de la etapa anterior.
3. Representar de manera alineada los espectrogramas de las dos versiones. Puede resultar apropiado calcular la representación para tramos de menor duración. Comparar y analizar diferencias.
4. Transformar el audio de una de las versiones para sincronizarla con la otra, usando una implementación adecuada del *phase-vocoder*.
5. Analizar el resultado del alineamiento de forma cualitativa. Proponga medidas objetivas para evaluar el alineamiento.

3.2. Entrega final

Consiste en un informe de no más de 10 páginas del trabajo realizado durante todo el proyecto. Se deben describir los experimentos realizados para las dos etapas del proyecto, incluyendo un análisis y discusión de los resultados obtenidos que se consideren relevantes. Se sugiere hacer referencia a los resultados cuantitativos y a los cualitativos. No es necesario describir en detalle las técnicas utilizadas, salvo lo que sea pertinente para la discusión de los resultados experimentales. Se deberá entregar también el código. Siempre que sea posible, se sugiere aprovechar implementaciones existentes de las técnicas aplicadas.

3.3. Defensa

Cada equipo tendrá 15 minutos para hacer una presentación oral de su trabajo, de forma presencial. El objetivo principal es discutir los resultados cualitativos y cuantitativos de los experimentos realizados. Se valorará el análisis de esos resultados, vinculado con los conceptos y técnicas vistas en el curso. No es necesario incluir una descripción detallada de las técnicas utilizadas en la presentación. Las preguntas de los docentes pueden aludir a detalles de estas técnicas.

Referencias

- [1] M. Müller, *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer, 1st ed., 2015. Ch. 3 - Music Synchronization.