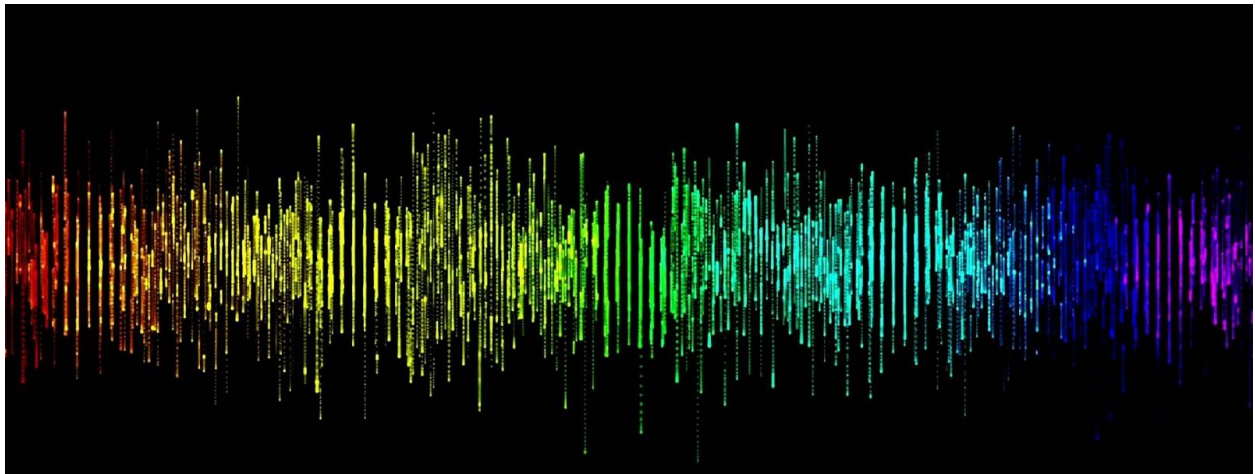


Propuesta de TP Final de AID

Exploración sonora utilizando algoritmos de clustering



Leandro Garber

Lic. en Artes Electrónicas con orientación en sonido

Como Licenciado en Artes Electrónicas de la UNTREF mi mayor interés en relación con la maestría es poder utilizar las técnicas que vimos en proyectos que se encuentren en el campo de las Artes.

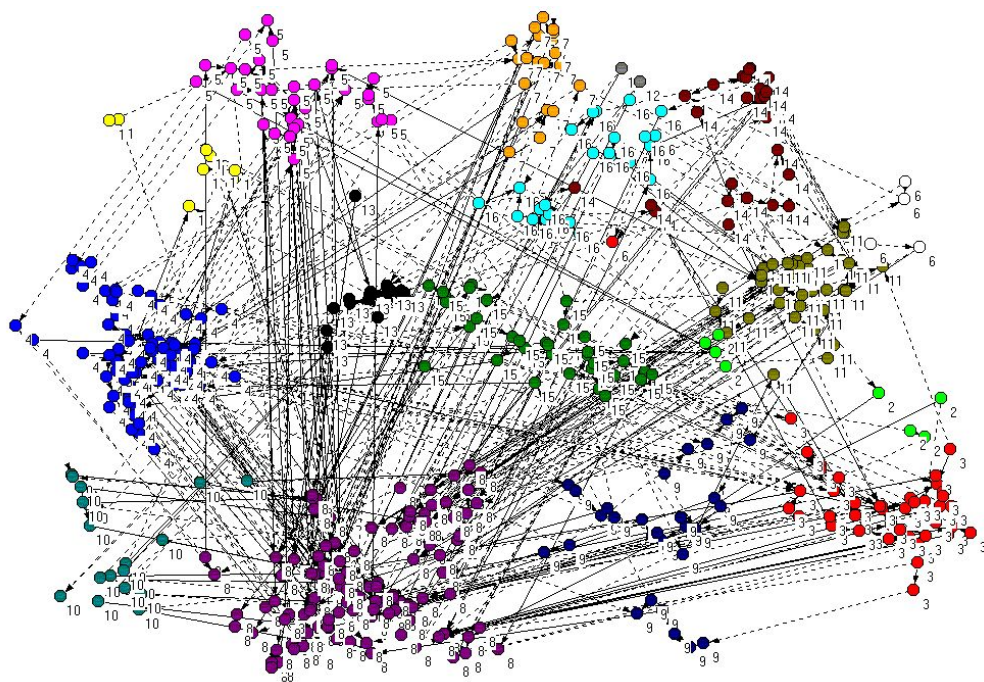
Estas últimas semanas estuve probando si era factible una idea que se me había ocurrido utilizando como dataset **señal digital de audio**. Felizmente, las pruebas superaron las expectativas.

Estoy separando canciones en pedazos de 1 segundo dando como resultado un **dataset de 22050 variables** y tantas filas como segundos tenga la canción. De momento probé con 4 canciones diferentes. Para el dataset final la idea es meter muchas canciones (o grabaciones, no necesariamente canciones) en el mismo dataset.

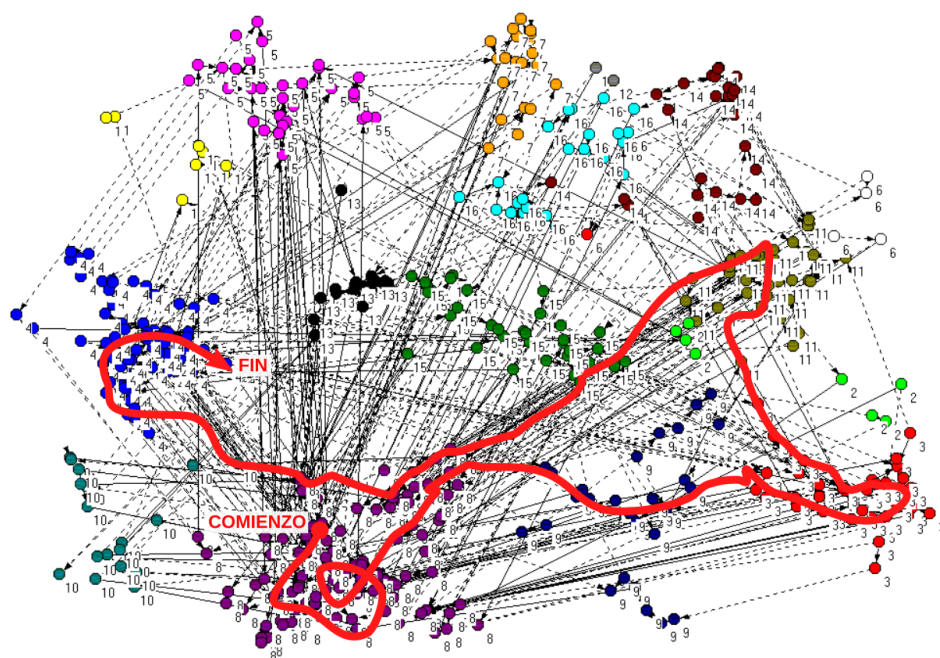
El **objetivo** es más bien exploratorio: utilizar **técnicas de clustering** para agrupar estos pedazos de canción (o grabaciones) y analizar los conglomerados resultantes en busca de un sentido musical: ver si pueden detectar la estructura de la canción (verso, estribillo, etc), ver si pueden separar armónicamente o en base a la amplitud, ritmo, etc.

Utilizando **escalamiento multidimensional**, el resultado de estos procedimientos puede visualizarse como un mapa donde los pedazos de grabaciones similares se encuentran más cercanos y los más distintos más lejanos. Ver desde ahí si es factible crear un **sistema de composición algorítmica** definiendo heurísticas de cómo recorrer este mapa (que también puede ser pensado en términos de grafo).

*El sistema de composición queda fuera de los alcances del trabajo, el objetivo es **analizar la factibilidad**.*



EJEMPLO DE GRAFO RESULTANTE



EJEMPLO DE RECORRIDO EN GRAFO

La **síntesis del proceso** entonces sería:

1. Dividir una grabación o conjunto de grabaciones en pedazos de igual longitud temporal (por ej. 1 segundo) dando como resultado un **dataset** de 22050 variables (en el caso de 1 segundo) por tantas filas como segundos de las grabaciones (esto será alrededor de 200 para las primeras pruebas y alrededor de 2000 para las últimas)
2. Utilizar **componentes principales** para reducir la dimensionalidad
3. Utilizar **clusters jerárquicos** y analizar el **dendograma** resultante para visualizar las aglomeraciones posibles. Escuchar las diferentes combinaciones con el fin de interpretar cómo está agrupando el algoritmo.
4. Utilizar **escalamiento multidimensional** para poder visualizar los clusters como grafos con el fin de analizar la posibilidades de composición musical algorítmica.

Las **pruebas** serían:

- Probar diferentes distancias y algoritmos de cluster jerárquico y elegir el más adecuado basándose en la escucha.
- Probar distintas grabaciones para ver si el algoritmo de cluster elegido se comporta de manera similar
- Utilizar como dataset no sólo una grabación sino una variedad de ellas y realizar el mismo procedimiento

Las **técnicas utilizadas** serían:

- Análisis de componentes principales
- Clustering jerárquico
- Escalamiento multidimensional

Bibliografía relacionada:

Burka, Zak, "Perceptual audio classification using principal component analysis" (2010). Thesis. Rochester Institute of Technology. ([LINK](#))