Análisis y clasificación de géneros musicales

Ariel Rojas, Juan Espinosa y Mauro Chavez

Departamento de Física. Facultad de Ciencias Exactas y Naturales (FCEyN), Universidad de Buenos Aires

Laboratorio de Datos

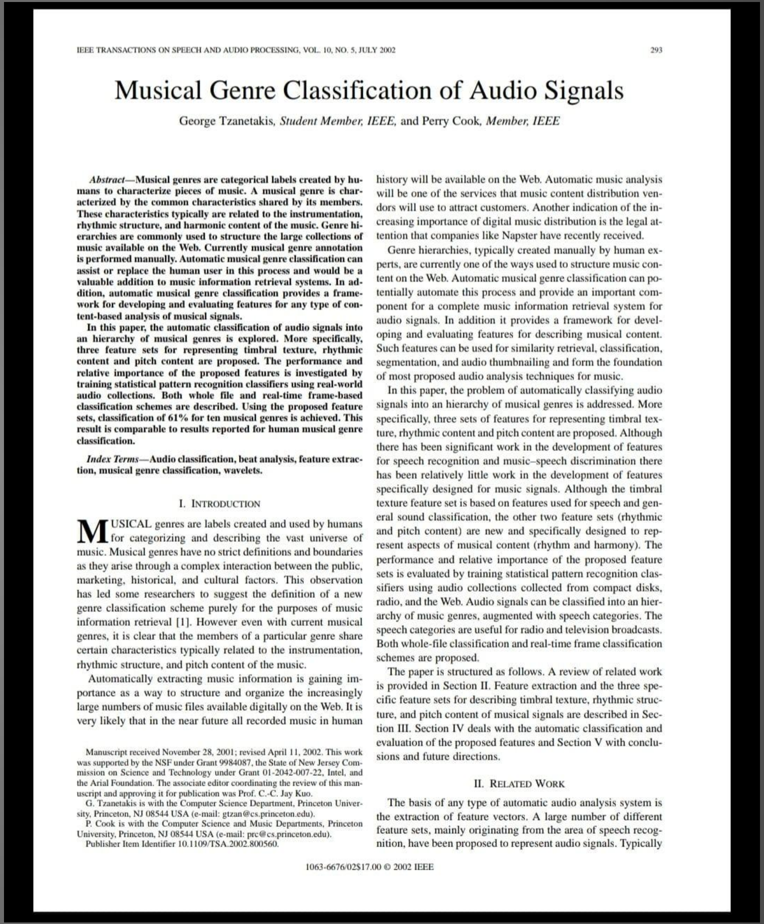
Caraterísticas del DataSet

- 10 géneros
- 100 audios cada uno
- Cada audio de 30’’
- Balanceado
- Visualización (Mel Espectrogramas)
- Features Extraídos

https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification

Género	Cantidad
blues	100
jazz	100
metal	100
pop	100
reggae	100
disco	100
classical	100
hiphop	100
rock	100
country	100

https://ieeexplore.ieee.org/document/1021072



Musical Genre Classification of Audio Signals

George Tzanetakis, *Student Member, IEEE,* and Perry Cook, *Member, IEEE*

***Abstract*—Musical genres are categorical labels created by humans to characterize pieces of music. A musical genre is characterized by the common characteristics shared by its members. These characteristics typically are related to the instrumentation, rhythmic structure, and harmonic content of the music. Genre hierarchies are commonly used to structure the large collections of music available on the Web. Currently musical genre annotation is performed manually. Automatic musical genre classification can assist or replace the human user in this process and would be a valuable addition to music information retrieval systems. In addition, automatic musical genre classification provides a framework for developing and evaluating features for any type of content-based analysis of musical signals.**

In this paper, the automatic classification of audio signals into an hierarchy of musical genres is explored. More specifically, three feature sets for representing timbral texture, rhythmic content and pitch content are proposed. The performance and relative importance of the proposed features is investigated by training statistical pattern recognition classifiers using real-world audio collections. Both whole file and real-time frame-based classification schemes are described. Using the proposed feature sets, classification of 61% for ten musical genres is achieved. This result is comparable to results reported for human musical genre classification.

***Index Terms*—Audio classification, beat analysis, feature extraction, musical genre classification, wavelets.**

I. Introduction

MUSICAL genres are labels created and used by humans for categorizing and describing the vast universe of music. Musical genres have no strict definitions and boundaries as they arise through a complex interaction between the public, marketing, historical, and cultural factors. This observation has led some researchers to suggest the definition of a new genre classification scheme purely for the purposes of music information retrieval [1]. However even with current musical genres, it is clear that the members of a particular genre share certain characteristics typically related to the instrumentation, rhythmic structure, and pitch content of the music.

Automatically extracting music information is gaining importance as a way to structure and organize the increasingly large numbers of music files available digitally on the Web. It is very likely that in the near future all recorded music in human history will be available on the Web. Automatic music analysis will be one of the services that music content distribution vendors will use to attract customers. Another indication of the increasing importance of digital music distribution is the legal attention that companies like Napster have recently received.

Genre hierarchies, typically created manually by human experts, are currently one of the ways used to structure music content on the Web. Automatic musical genre classification can potentially automate this process and provide an important component for a complete music information retrieval system for audio signals. In addition it provides a framework for developing and evaluating features for describing musical content. Such features can be used for similarity retrieval, classification, segmentation, and audio thumbnailing and form the foundation of most proposed audio analysis techniques for music.

In this paper, the problem of automatically classifying audio signals into an hierarchy of musical genres is addressed. More specifically, three sets of features for representing timbral texture, rhythmic content and pitch content are proposed. Although there has been significant work in the development of features for speech recognition and music–speech discrimination there has been relatively little work in the development of features specifically designed for music signals. Although the timbral texture feature set is based on features used for speech and general sound classification, the other two feature sets (rhythmic and pitch content) are new and specifically designed to represent aspects of musical content (rhythm and harmony). The performance and relative importance of the proposed feature sets is evaluated by training statistical pattern recognition classifiers using audio collections collected from compact disks, radio, and the Web. Audio signals can be classified into an hierarchy of music genres, augmented with speech categories. The speech categories are useful for radio and television broadcasts. Both whole-file classification and real-time frame classification schemes are proposed.

The paper is structured as follows. A review of related work is provided in Section II. Feature extraction and the three specific feature sets for describing timbral texture, rhythmic structure, and pitch content of musical signals are described in Section III. Section IV deals with the automatic classification and evaluation of the proposed features and Section V with conclusions and future directions.

II. Related Work

The basis of any type of automatic audio analysis system is the extraction of feature vectors. A large number of different feature sets, mainly originating from the area of speech recognition, have been proposed to represent audio signals. Typically

Manuscript received November 28, 2001; revised April 11, 2002. This work was supported by the NSF under Grant 9984087, the State of New Jersey Commission on Science and Technology under Grant 01-2042-007-22, Intel, and the Arial Foundation. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. C.-C. Jay Kuo.

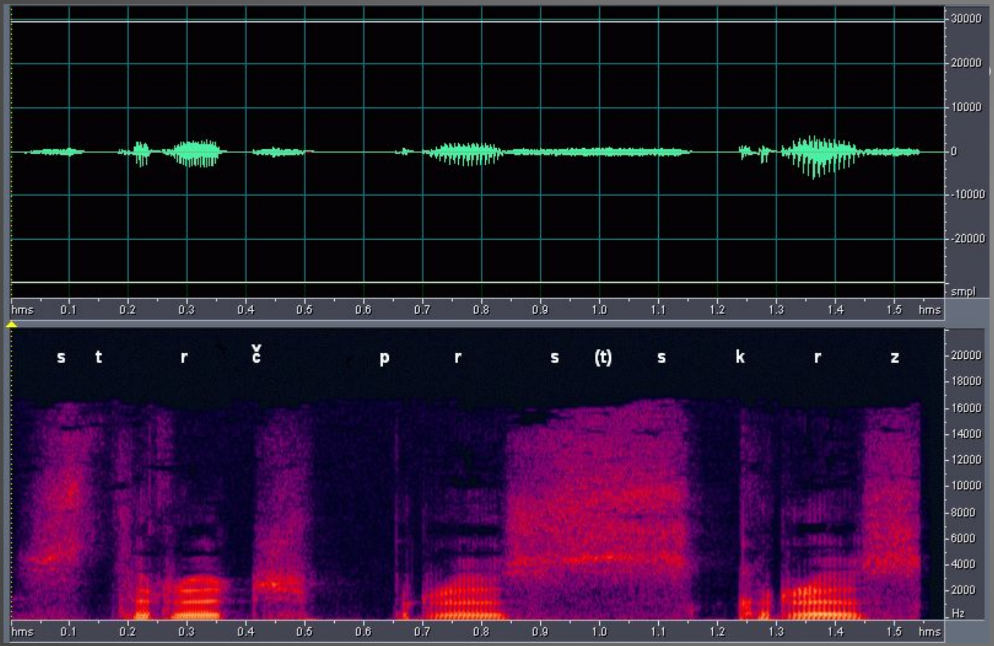
G. Tzanetakis is with the Computer Science Department, Princeton University, Princeton, NJ 08544 USA (e-mail: gtzan@cs.princeton.edu).

P. Cook is with the Computer Science and Music Departments, Princeton University, Princeton, NJ 08544 USA (e-mail: prc@cs.princeton.edu).

Publisher Item Identifier 10.1109/TSA.2002.800560.

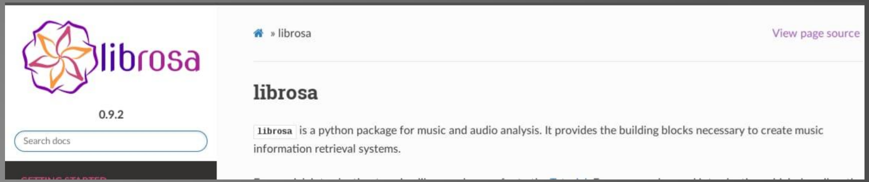
Extracción de Features: ¿Cómo Trabajamos con Audios?

Señal en función del tiempo



Espectrograma

Librería de análisis de audio



Transformada de Fourier Rápida

s:

$$\mathbf{STFT}\{x(t)\}(\tau,\omega) \equiv X(\tau,\omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-i\omega t}\,dt$$

Features(MFCC)

- Flujo Espectral: Es una medida de cómo el espectro está cambiando localmente.

- Spectral Rolloff: Es la frecuencia R_t por la que, debajo de ella, se encuentra el 85% de la magnitud del espectro.

- Spectral Centroid: Es el centro de gravedad de la magnitud del espectro, da una idea del shape del espectro.

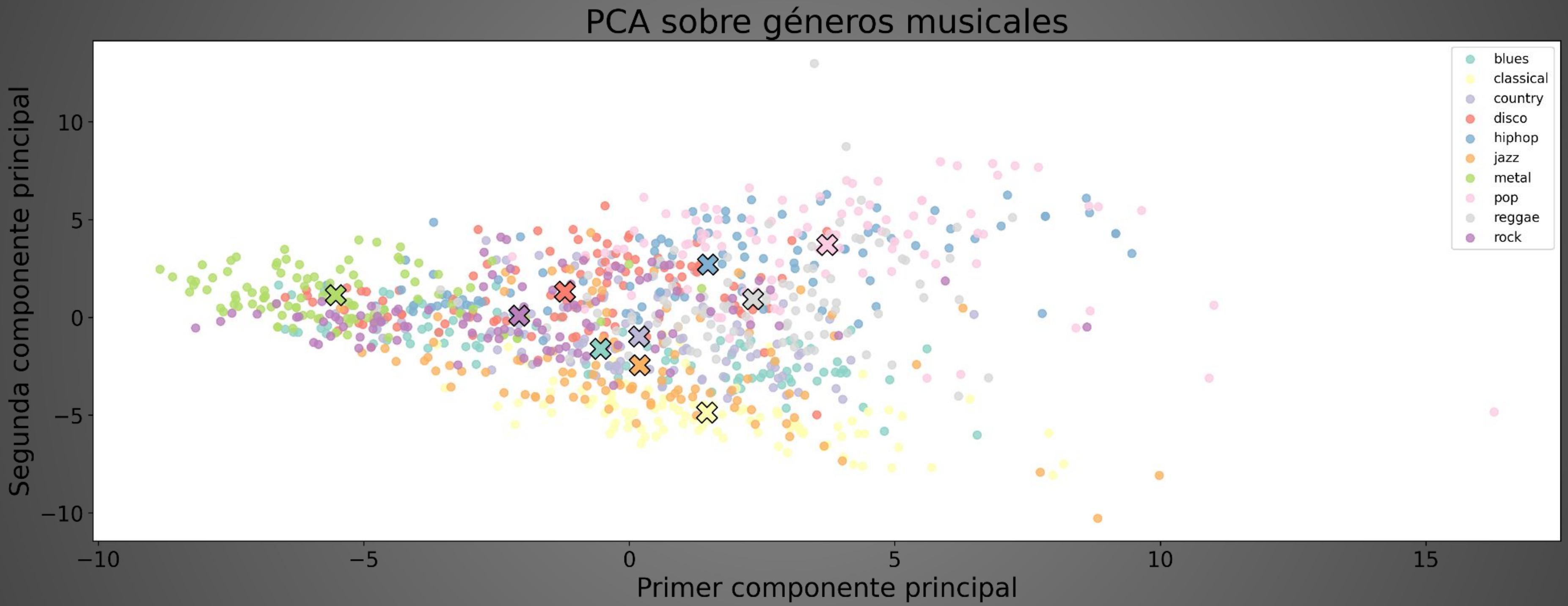
Features(MFCC)

- Low-energy Feature: Promedio de analysis windows con menos energía que el promedio de texture window. ●●●

- Timbral Texture Features Vector: Combinación de los otros features, es un vector de 19 dimensiones.

- Time Domain Zero Crossings: Da una idea del ruido de la señal.

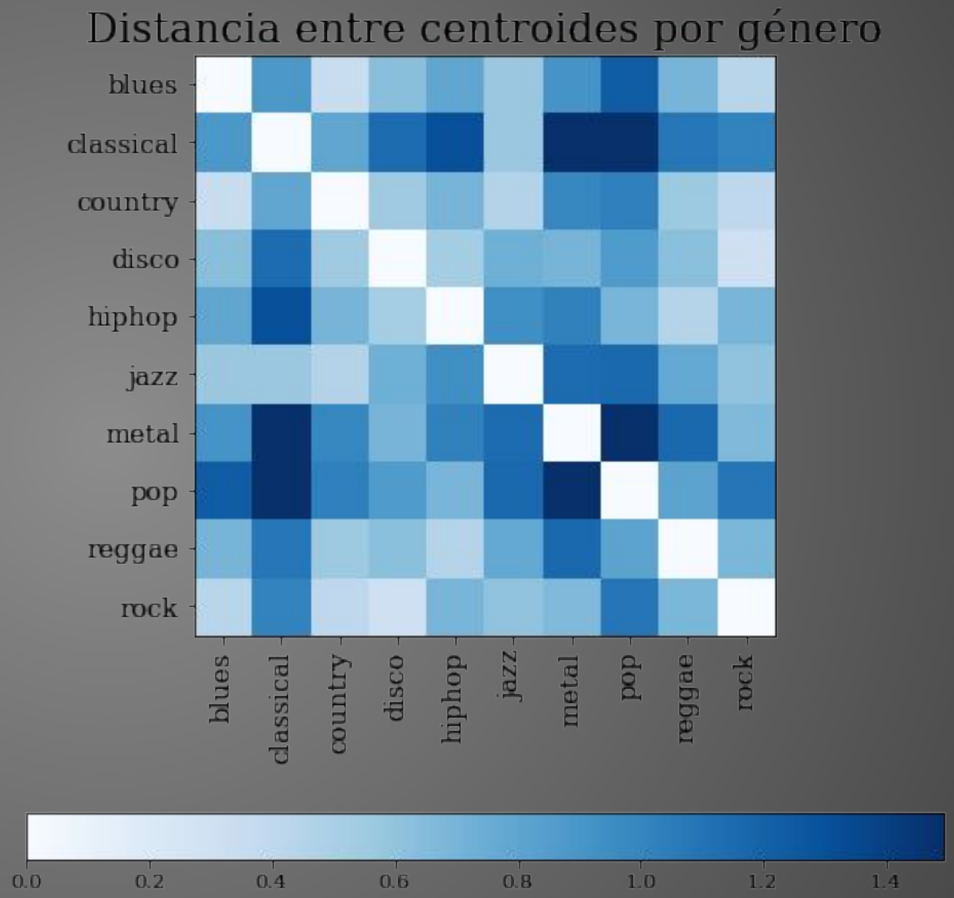
PCA


PCA sobre géneros musicales
Segunda componente principal
Primer componente principal
blues
classical
country
disco
hiphop
jazz
metal
pop
reggae
rock
10
5
0
−5
−10
−10
−5
0
5
10
15

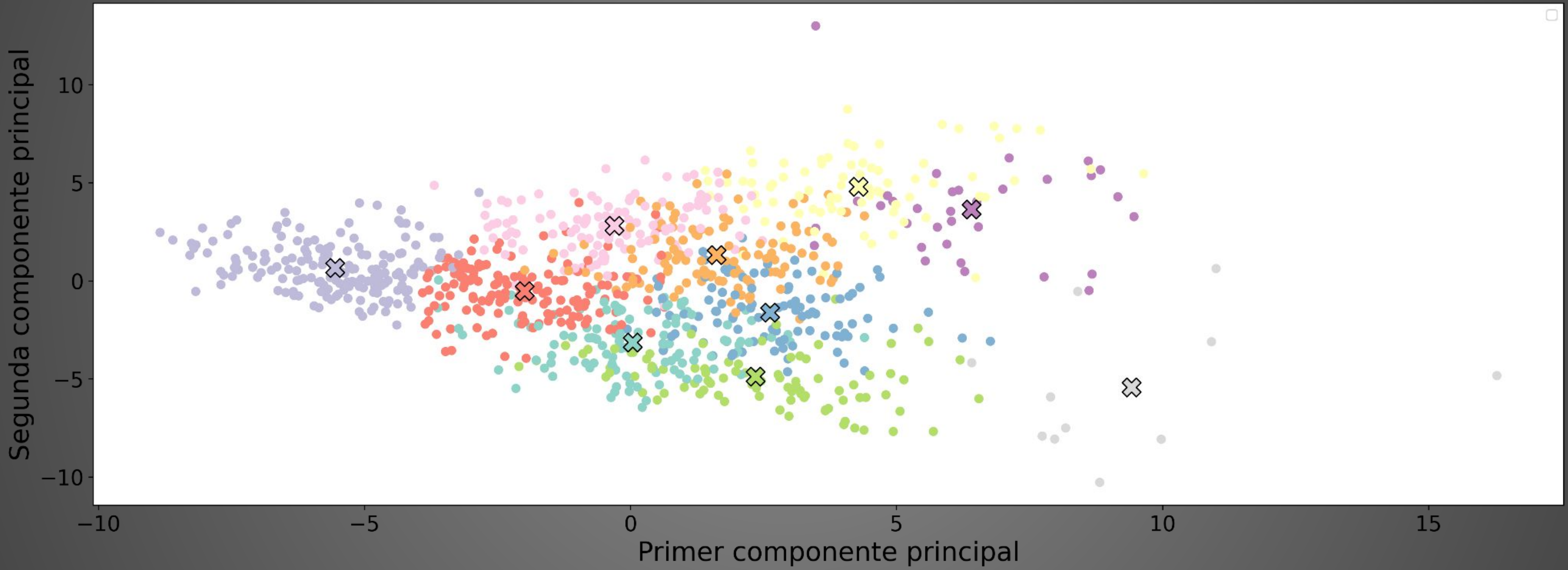
PCA

Distancia Máxima:
Classical-pop

Distancia Mínima:
disco-rock



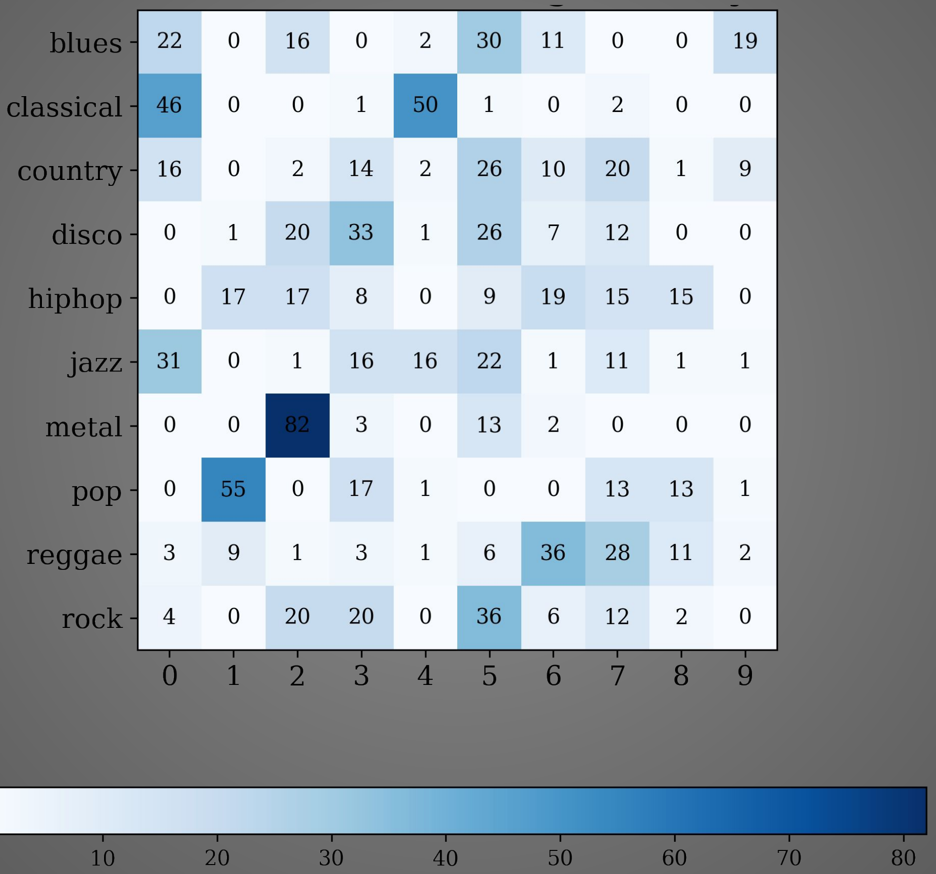
K-Means



Segunda componente principal
10
5
0
−5
−10
−10
−5
0
5
10
15
Primer componente principal

K = 10

K-Means


blues 22 0 16 0 2 30 11 0 0 19
classical 46 0 0 1 50 1 0 2 0 0
country 16 0 2 14 2 26 10 20 1 9
disco 0 1 20 33 1 26 7 12 0 0
hiphop 0 17 17 8 0 9 19 15 15 0
jazz 31 0 1 16 16 22 1 11 1 1
metal 0 0 82 3 0 13 2 0 0 0
pop 0 55 0 17 1 0 0 13 13 1
reggae 3 9 1 3 1 6 36 28 11 2
rock 4 0 20 20 0 36 6 12 2 0
0 1 2 3 4 5 6 7 8 9
0 10 20 30 40 50 60 70 80

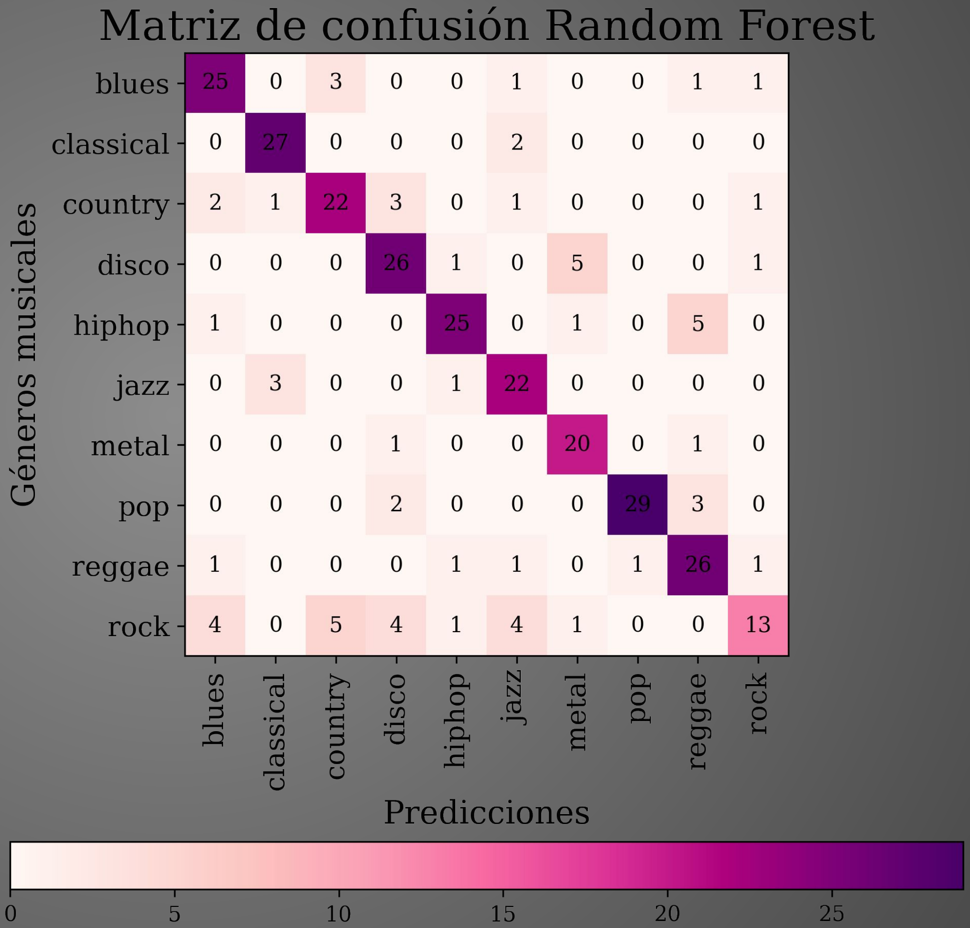
Modelos de clasificación

Modelo	ACC: Test Set	ACC: Train Set
KNN	0.757	0.663
SVM Lineal	0.377	0.423
SVM RBF	0.763	0.957
Árbol de decisión	0.523	0.616
Random Forest	0.783	0.998

Random Forest

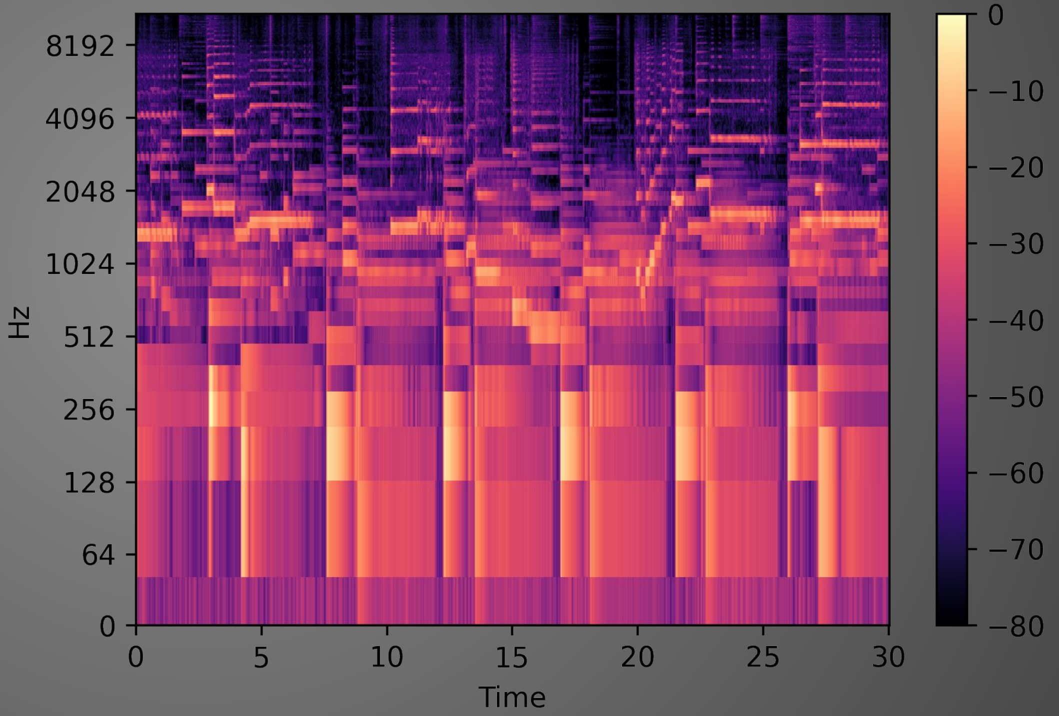
Accuracy:

- SetTrain: 0.998
- TestTrain: 0.783



Modelos de regresión a partir de mel-espectrogramas

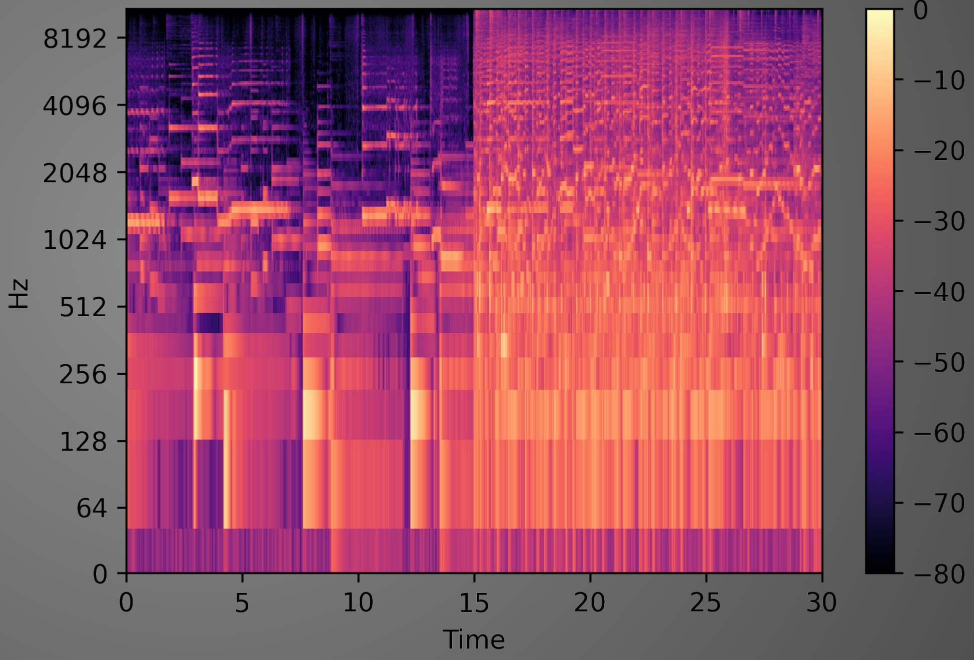
- X: primeros 15 segundos
- y: últimos 15 segundos



Modelos de regresión a partir de mel-espectrogramas

KNN:

- Jazz:
- Clásica:




Primeros segundos
Ultimos segundos

Conclusiones

- Hay géneros musicales que son más fácilmente clasificables que otros.
- 20 años después del paper que inspiró nuestro trabajo, conseguir un score más alto es razonablemente realizable.
- El dominio de los problemas se puede cambiar (Audio -> imagen)
- La música generada por regresión es inquietante.

¡Muchas gracias por su atención!

¿Preguntas?