

Análisis de Audios Mediante Espectrogramas y MFCCs

Diana Cordero

Marzo 2025

1. Introducción

El análisis de señales de audio es fundamental en diversas aplicaciones de inteligencia artificial, como el reconocimiento de voz, clasificación de sonidos ambientales y síntesis de audio. En este trabajo, se utilizan técnicas de procesamiento de señales para extraer y visualizar características importantes de audios mediante espectrogramas y coeficientes cepstrales en frecuencia de Mel (MFCCs).

2. Espectrogramas y sus Métricas

El espectrograma es una representación gráfica de la potencia de las frecuencias en una señal a lo largo del tiempo. En este trabajo, se utilizó la Transformada Rápida de Fourier (STFT) para generar los espectrogramas de los audios. Las métricas clave incluyen la frecuencia y la potencia, donde las frecuencias bajas corresponden a sonidos graves y las altas a sonidos agudos. La potencia de cada frecuencia está representada en una escala logarítmica en dB, lo que permite observar con mayor precisión las variaciones de las frecuencias a través del tiempo.

3. MFCCs y sus Métricas

Los coeficientes cepstrales en frecuencia de Mel (MFCCs) son una representación compacta de la información espectral de una señal de audio. Estos coeficientes se calculan a partir de la transformada de Fourier seguida de una transformación a una escala Mel, que se ajusta mejor a la percepción humana del sonido. Los MFCCs son muy utilizados en tareas de clasificación

de audio debido a su capacidad para reducir la dimensionalidad de la señal manteniendo las características espectrales más relevantes. En este análisis, se calcularon los primeros 13 coeficientes de MFCC para cada archivo de audio.

4. Descripción del Dataset

Para este estudio se utilizó el dataset ESC-50, un conjunto de datos que contiene 2000 archivos de audio en 50 categorías distintas, como sonidos de la naturaleza, maquinaria, animales y voz humana. Cada archivo tiene una duración de 5 segundos y está muestreado a 44.1 kHz.

5. Metodología

El análisis se realizó en Google Colab usando Python y las bibliotecas Librosa y Matplotlib. Se seleccionaron aleatoriamente 4 audios del dataset para representar distintas categorías. A continuación, se procesaron de la siguiente manera:

1. Carga y visualización de la forma de onda de cada audio.
2. Generación del espectrograma utilizando la Transformada Rápida de Fourier (STFT).
3. Cálculo de los MFCCs para analizar características perceptivas del sonido.

En la Figura 1 se presentan las formas de onda de los audios seleccionados, mostrando la variabilidad en amplitud y duración.

En la Figura 2, se muestran los espectrogramas, los cuales revelan la distribución de frecuencias a lo largo del tiempo. Se observa cómo distintos sonidos presentan patrones característicos de intensidad y frecuencia.

Finalmente, en la Figura 3 se presentan los coeficientes MFCCs, los cuales permiten extraer información relevante para tareas de clasificación de audio.

6. Conclusión

El análisis de señales de audio mediante espectrogramas y MFCCs permite visualizar información clave sobre la estructura del sonido. Estos métodos son esenciales para aplicaciones de reconocimiento y clasificación de audio en inteligencia artificial.

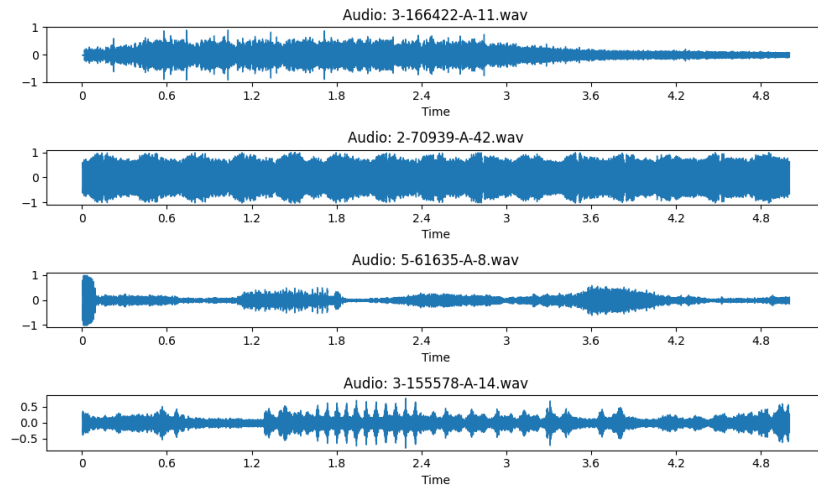


Figura 1: Formas de onda de los audios seleccionados.

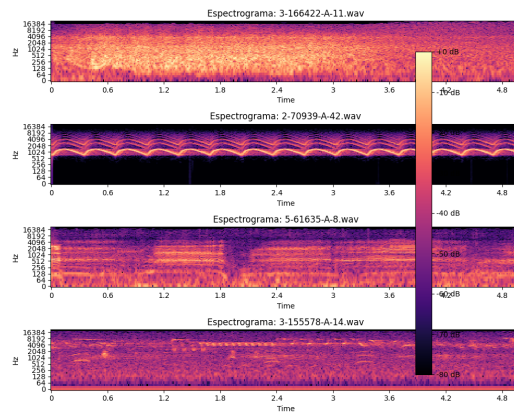


Figura 2: Espectrogramas de los audios seleccionados.

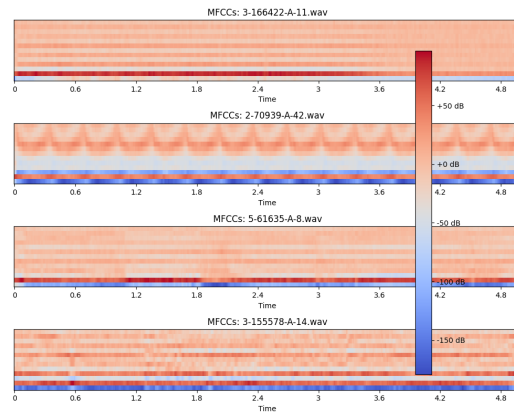


Figura 3: Coeficientes MFCCs de los audios seleccionados.

7. Referencias

- ESC-50 dataset: <https://github.com/karolpiczak/ESC-50>
- Tercero, P. B., Buil, P. (2015). Narrativa audiovisual y espectrogramas. Dialnet. <https://dialnet.unirioja.es/servlet/articulo?codigo=6698278>