Proyecto

DEEP FEELINGS

ÍNDICE

01. Objetivos

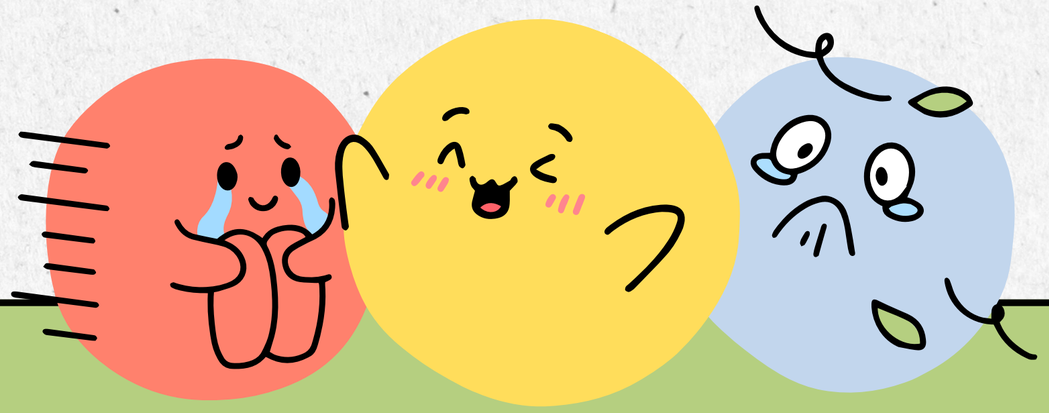
02. Descripción de la base de datos

03. Preprocesamiento de la base

04. Análisis exploratorio

05. Modelo

06. Post Procesamiento



OBJETIVO GENERAL

Identificar emociones a través de su valencia, excitación y dominación (VAD) de personas peruanas mediante un modelo de Deep Learning de redes convolucionales

USACH

OBJETIVOS ESPECÍFICOS

1. Explorar la base de datos para obtener una visión general de los datos disponibles.
2. Evaluar los métodos de extracción de características para seleccionar las técnicas más adecuadas.
3. Investigar los métodos de preprocesamiento de datos para mejorar la calidad y consistencia de los datos.
4. Implementar dos modelos con diferentes características para comparar su rendimiento.
5. Comparar y seleccionar el mejor modelo para asegurar la mayor efectividad en el análisis de datos.
6. Analizar los features, ver como se relacionan con los VAD y cómo afectan al modelo de CNN

DESCRIPCIÓN DE LA BASE DE DATOS

ESCORPUS-PE

80 HABLANTES
44 HOMBRES
36 MUJERES

3764

01. DEBATE PRESIDENCIAL 2021	638
02. TESTIMONIO DE HUILCA F. Y FLORES M.	210
03. TESTIMONIO LA CLASIFICACIÓN COPA DEL MUNDO 2018	199
04. REPORTE DE LA TRAGEDIA EN AREQUIPA EN 1996	116
05. REPORTE DE HELADAS EN AREQUIPA	230
06. REPORTE DE HUAYNACOTAS EN AREQUIPA	166
07. ENTREVISTA A PILOTO DE LA FAP POR ATAQUE OVNI ...	561
08. TESTIMONIO DE MARIA ANTONIETA QUISPE	326
09. ENTREVISTA "DERECHOS SEXUALES Y REPRODUCTIVOS ..."	279
10. INFORME DE LA MUERTE DE MARYORI EN HUANCAYO	103

ESCORPUS-PE

189	11. CHARLA TEDXCHARACATO DE MONICA HUERTA	
108	12. REPORTE DE FIN DE SEMANA EN AREQUIPA	
	13. TESTIMONIO DE CIPRIANA HUAMANI	454
	14. REPORTAJE DE RESTAURANT PERUANO EN MEXICO	185

NOMENCLATURA

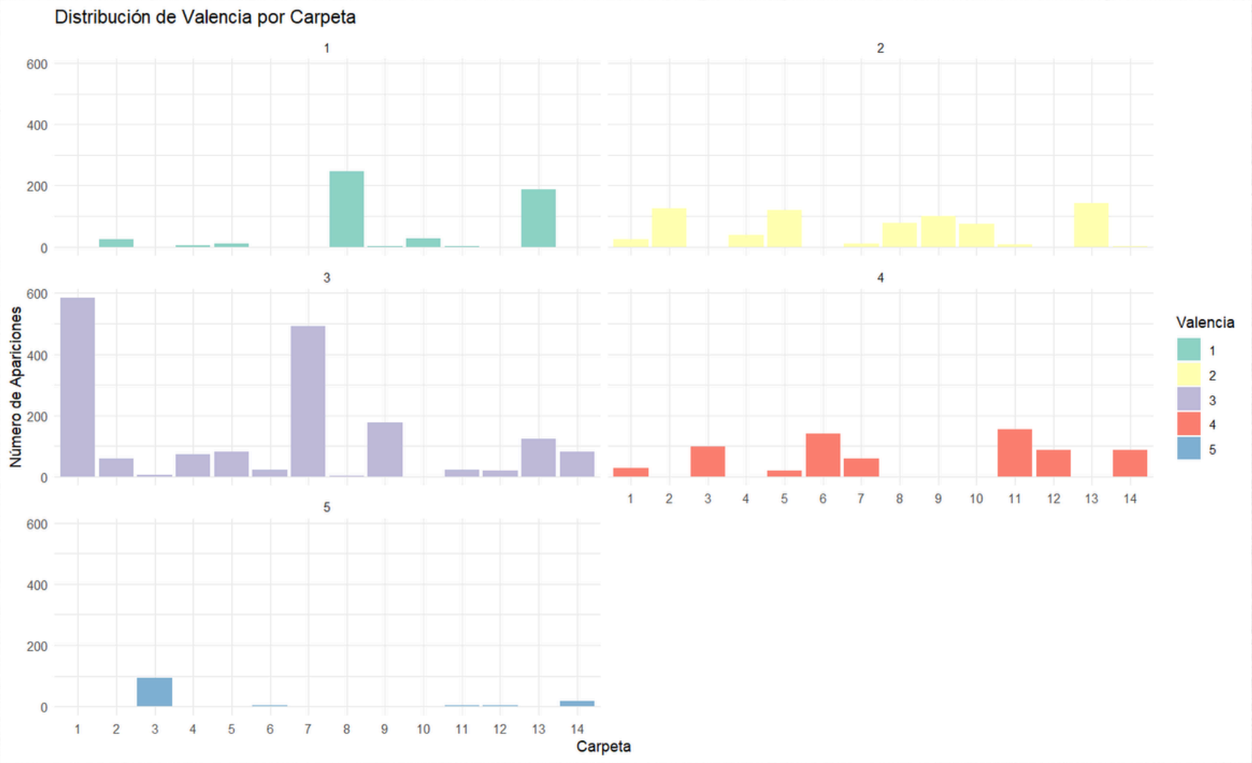
Audio[N]_[M]-

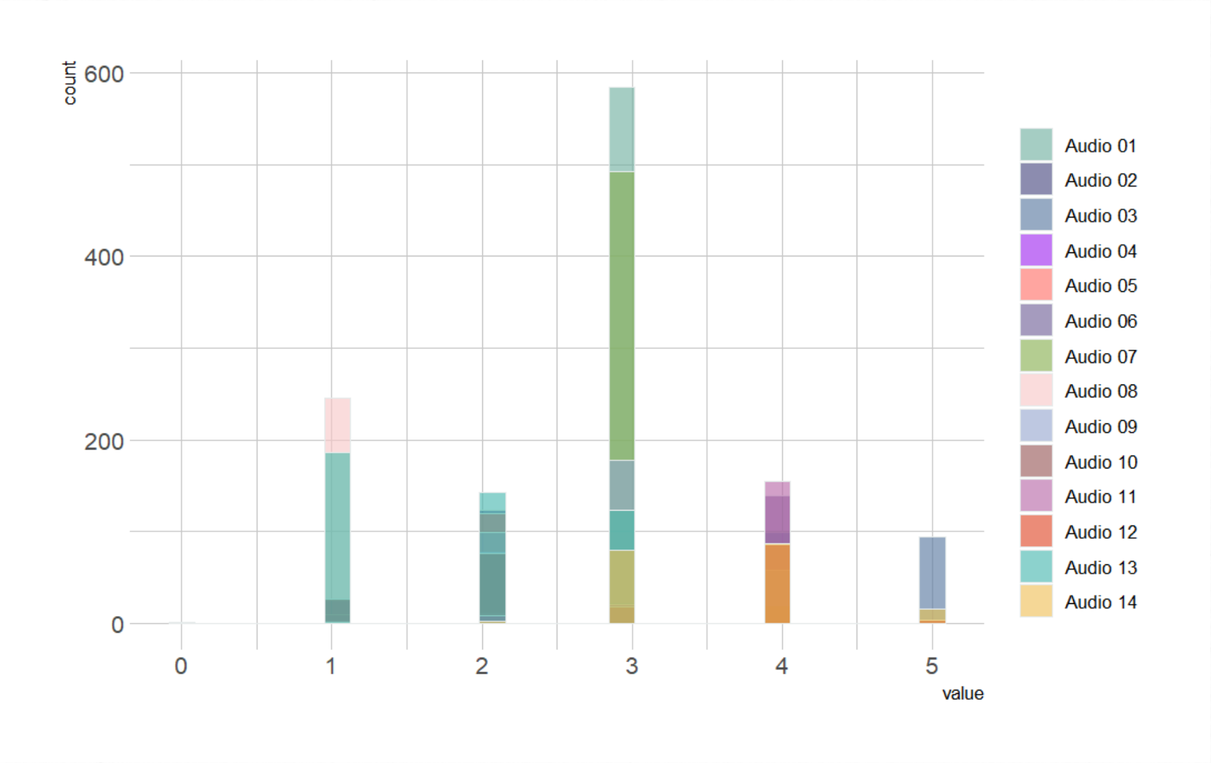
- [N]: [1-14]
- [M]: N° de audio en la carpeta

[1-5] - [1-5] - [1-5]

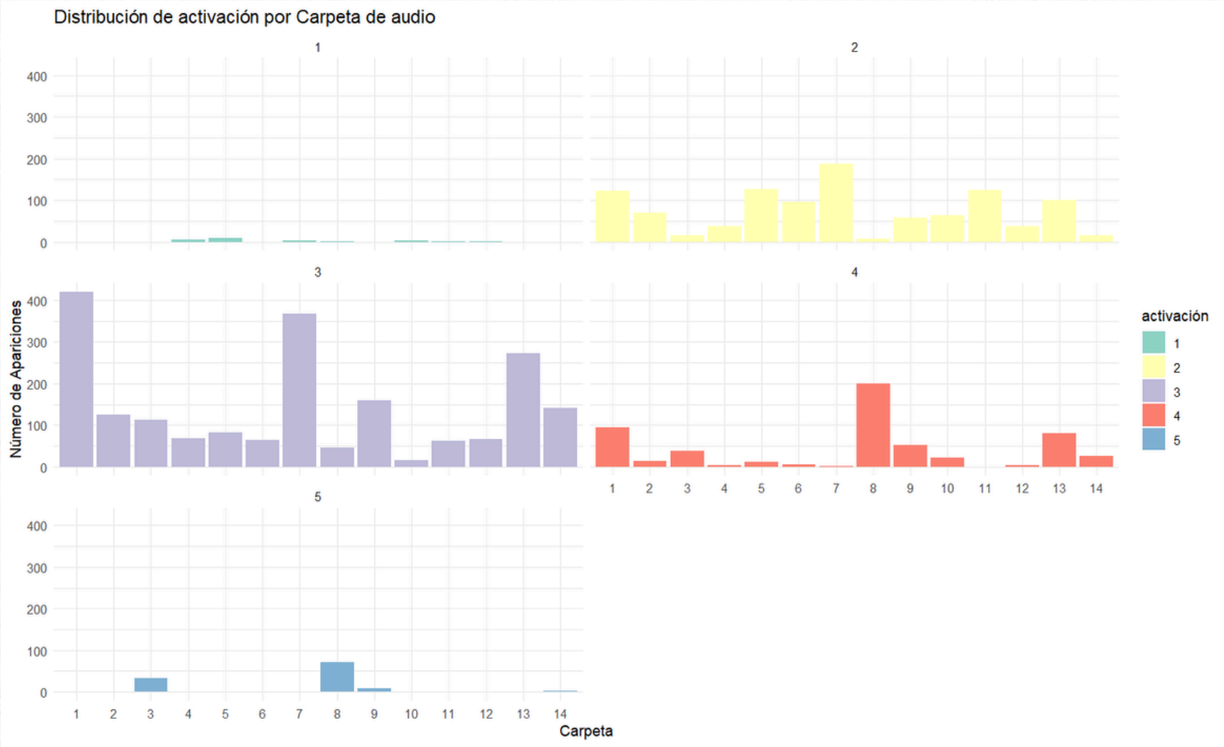
- **VALENCIA**: Si tiene connotación positiva o negativa
- **EXCITACIÓN**: Grado de intensidad de la persona al hablar
- **DOMINANCIA**: Grado de poder o control al hablar

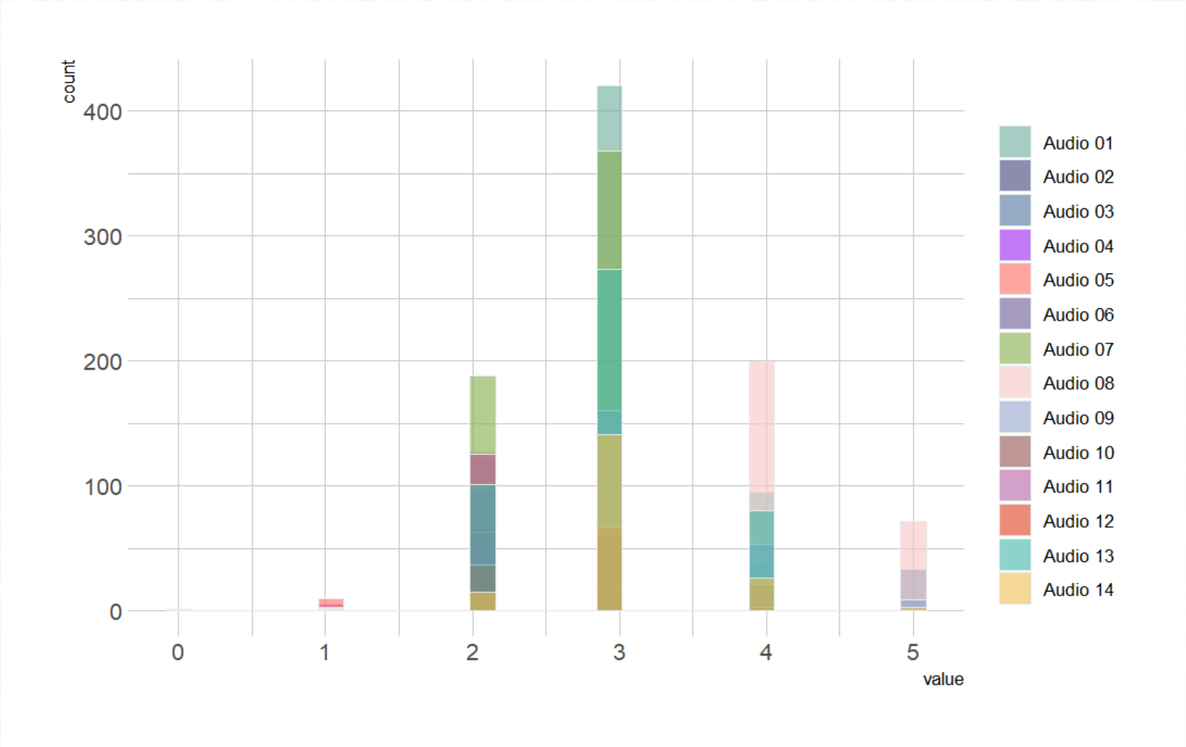
VALENCIA


Distribución de Valencia por Carpeta
1
2
3
4
5
Número de Apariciones
Carpeta
Valencia
1
2
3
4
5

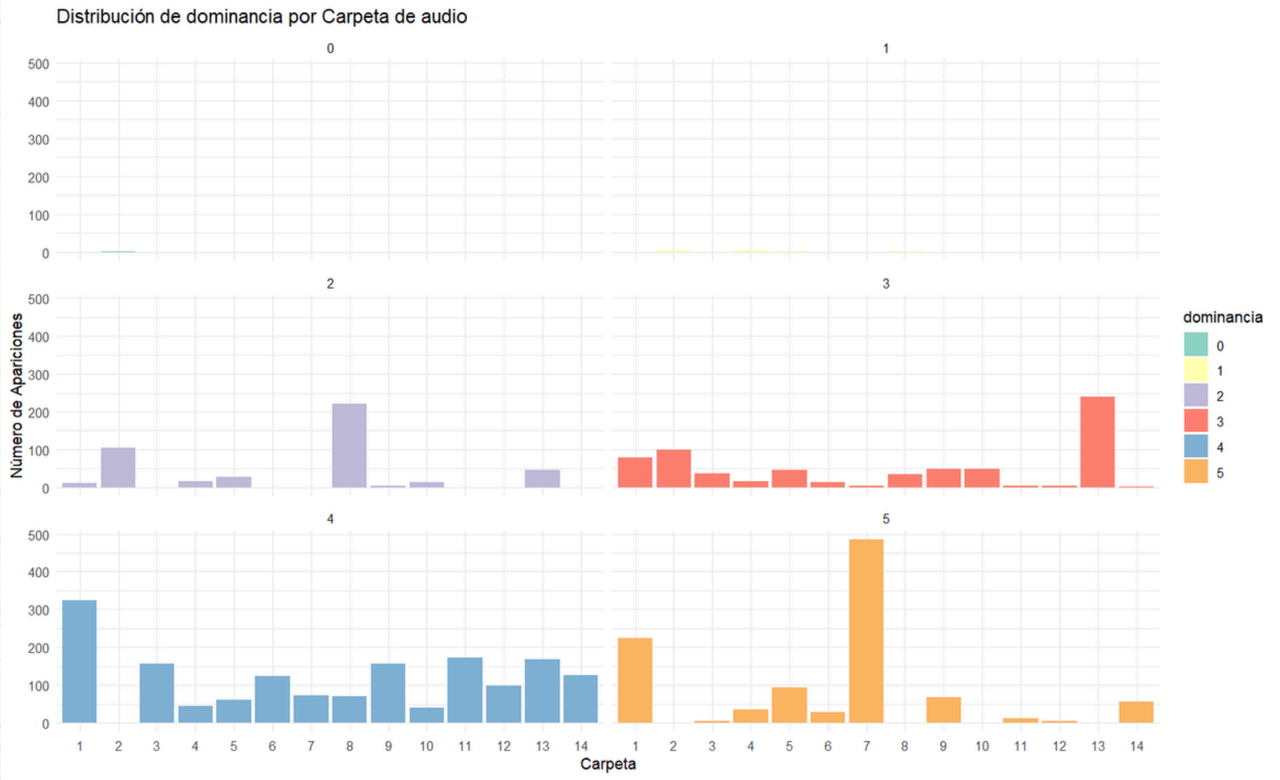

count
value
Audio 01
Audio 02
Audio 03
Audio 04
Audio 05
Audio 06
Audio 07
Audio 08
Audio 09
Audio 10
Audio 11
Audio 12
Audio 13
Audio 14

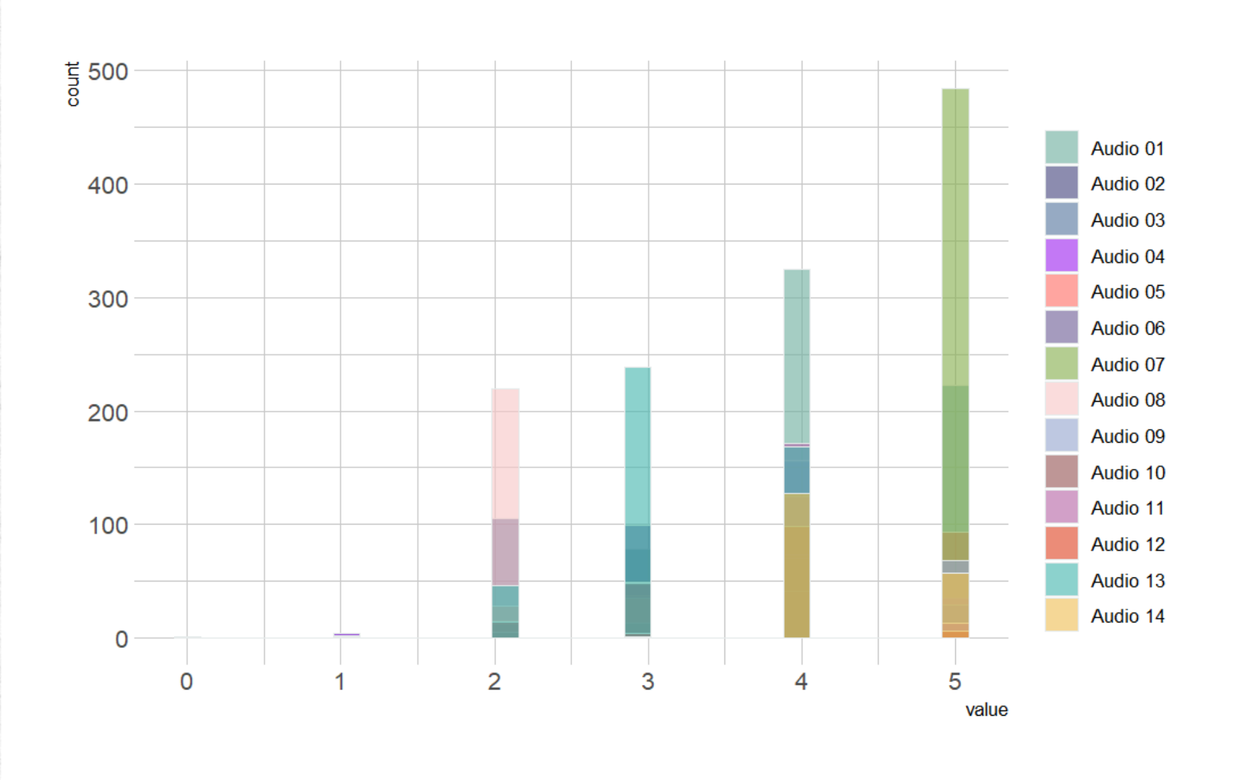
ACTIVACIÓN


Distribución de activación por Carpeta de audio
1
2
3
4
5
Número de Apariciones
Carpeta
activación
1
2
3
4
5


count
value
Audio 01
Audio 02
Audio 03
Audio 04
Audio 05
Audio 06
Audio 07
Audio 08
Audio 09
Audio 10
Audio 11
Audio 12
Audio 13
Audio 14

DOMINANCIA


Distribución de dominancia por Carpeta de audio
0
1
2
3
4
5
Número de Apariciones
Carpeta
dominancia
0
1
2
3
4
5


count
value
Audio 01
Audio 02
Audio 03
Audio 04
Audio 05
Audio 06
Audio 07
Audio 08
Audio 09
Audio 10
Audio 11
Audio 12
Audio 13
Audio 14

PREPROCESAMIENTO

Spectral Gating

Método para reducir ruido de fondo, Funciona estimando un umbral de ruido para cada banda de frecuencia de esa señal/ruido. Ese umbral se utiliza para calcular una máscara que silencia el ruido por debajo del umbral de frecuencia variable

Relleno y Recorte

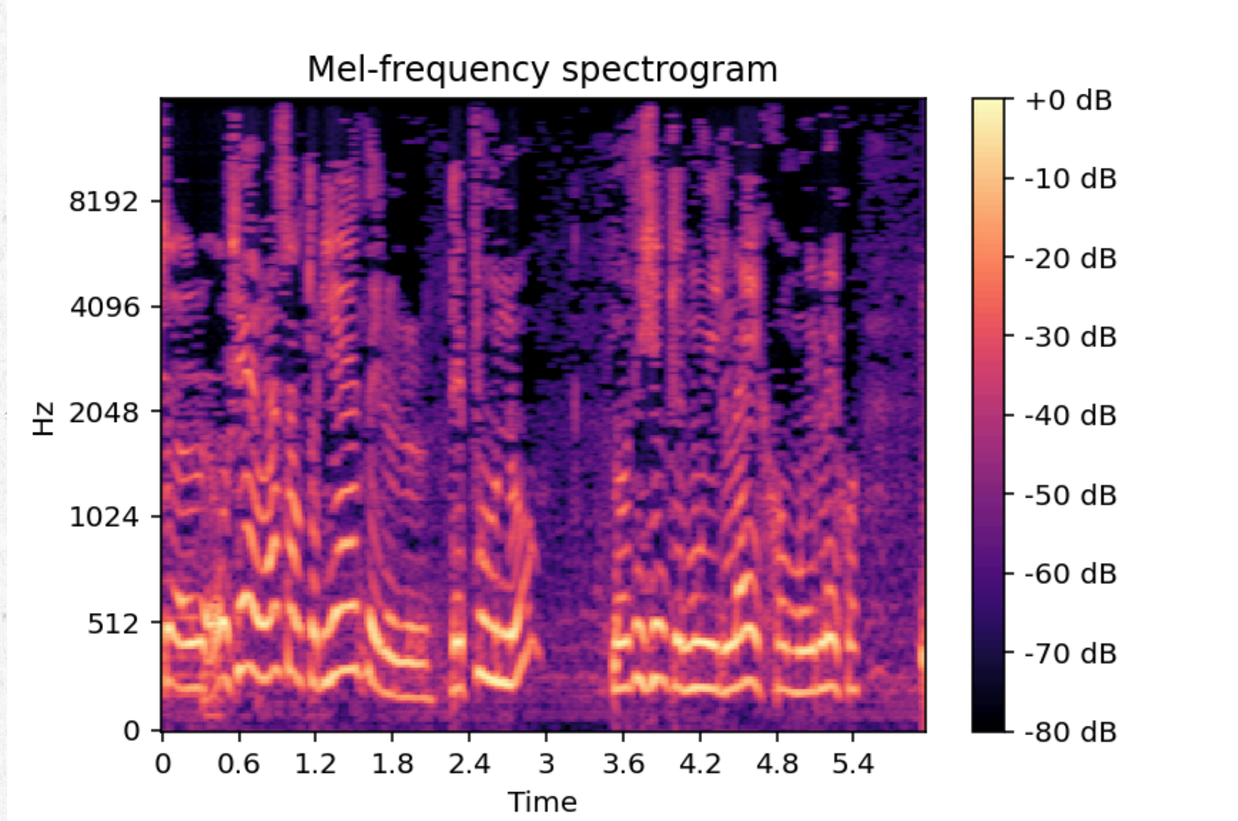
Necesario para que todas las matrices tengan la misma dimensión(12x60).
se trata de rellenar matriz con ceros hasta alcanzar 60 columnas. o cortar ka natriz hasta que tenga 60 columnas

normalización

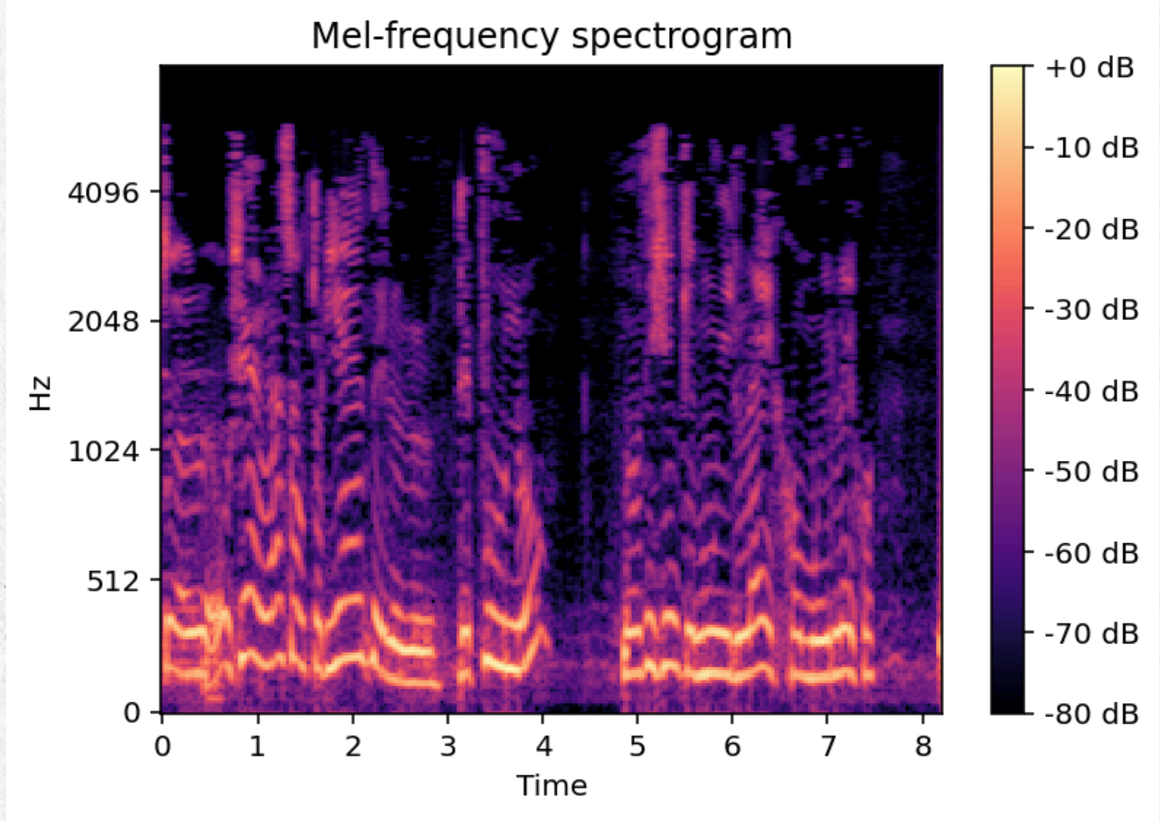
Aplicado a los mfcc antes de entrar al modelo

$$S_{Ni} = \frac{S_i - \mu}{\sigma}$$

SPECTRAL GATING

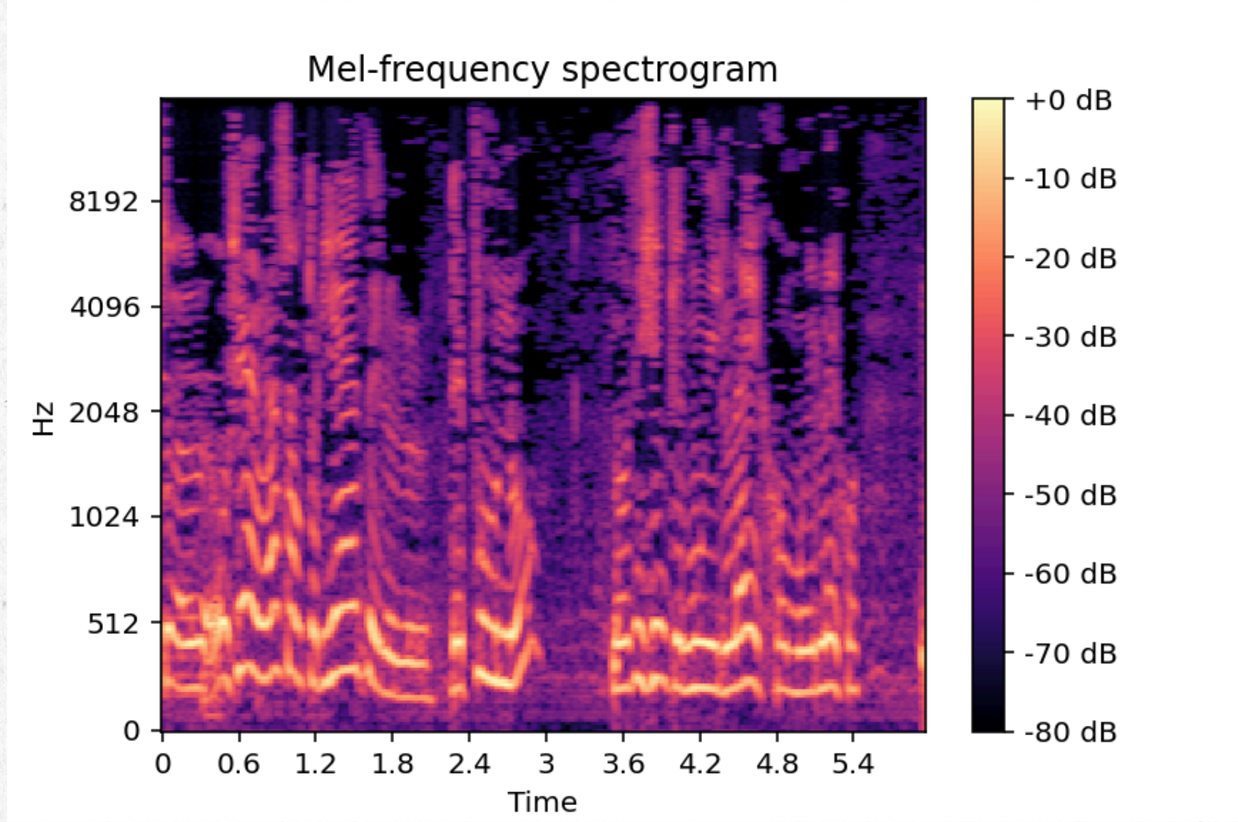


AUDIO ORIGINAL

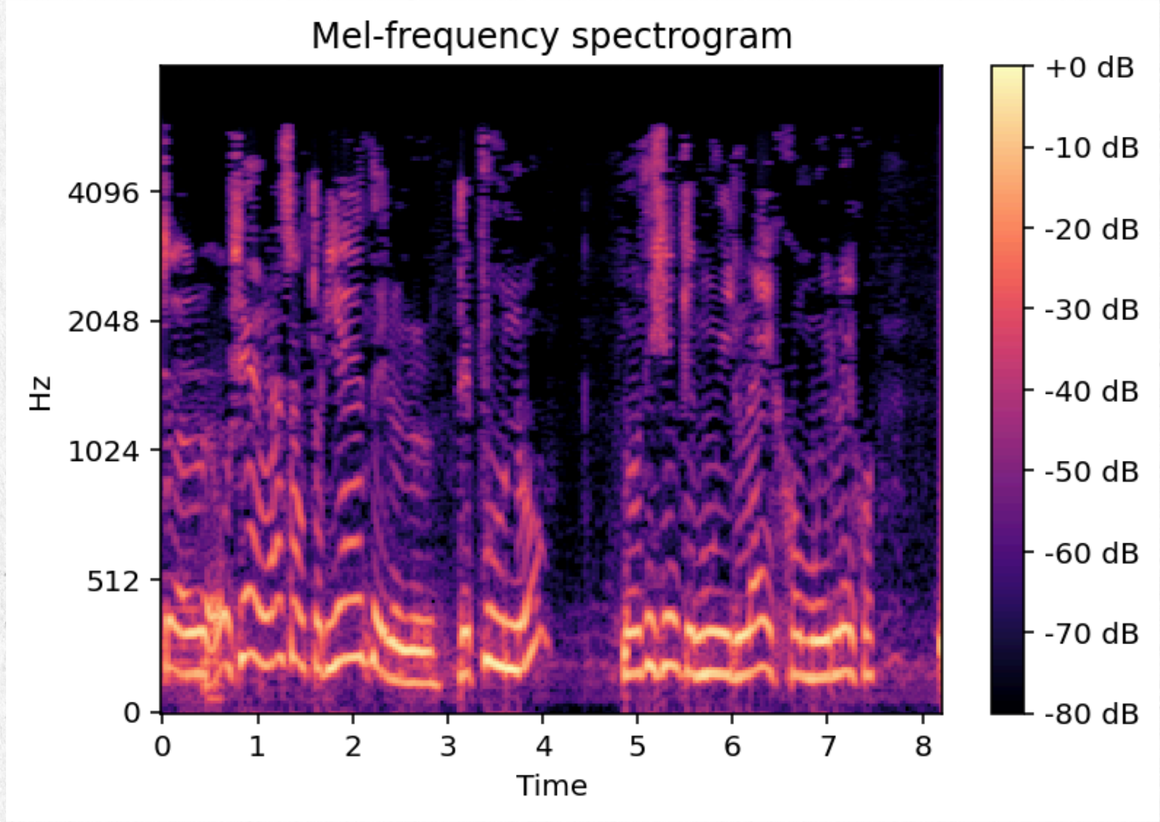


AUDIO PROCESADO

SPECTRAL GATING

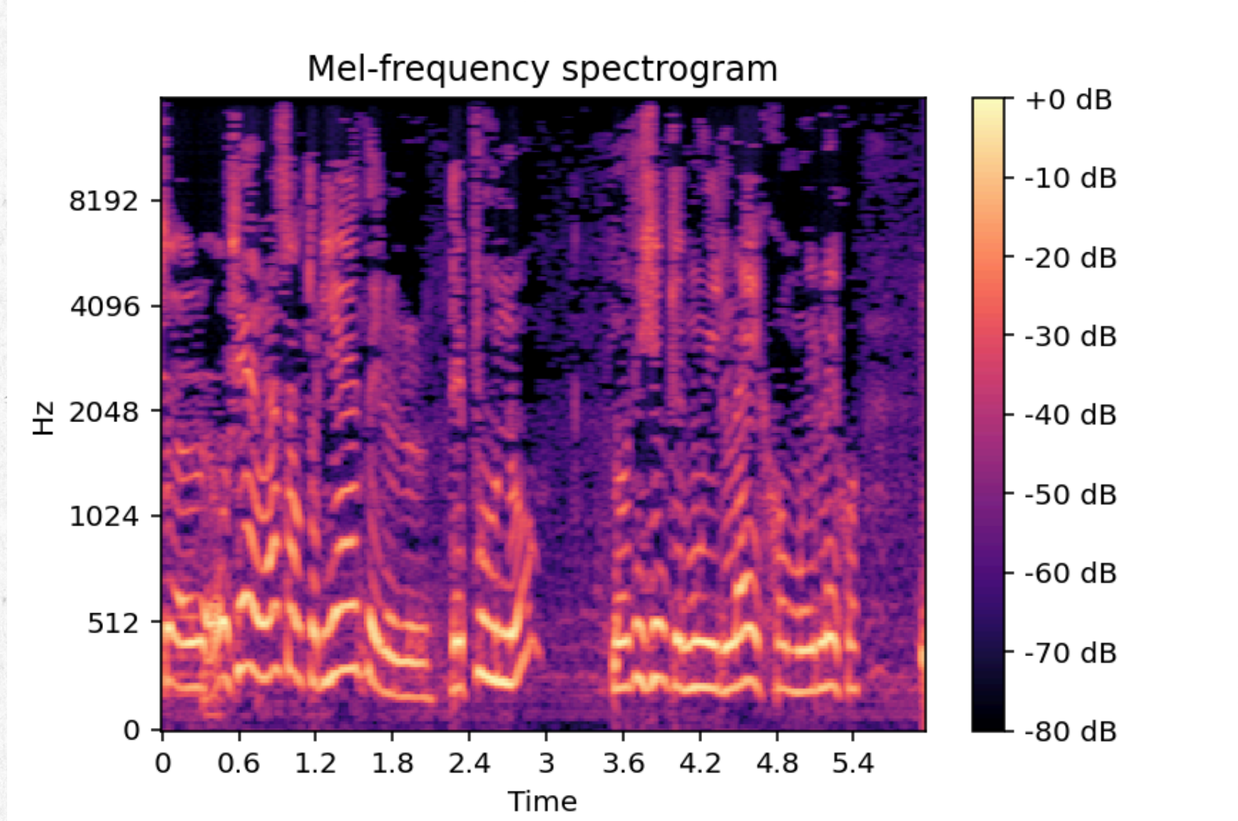


AUDIO ORIGINAL

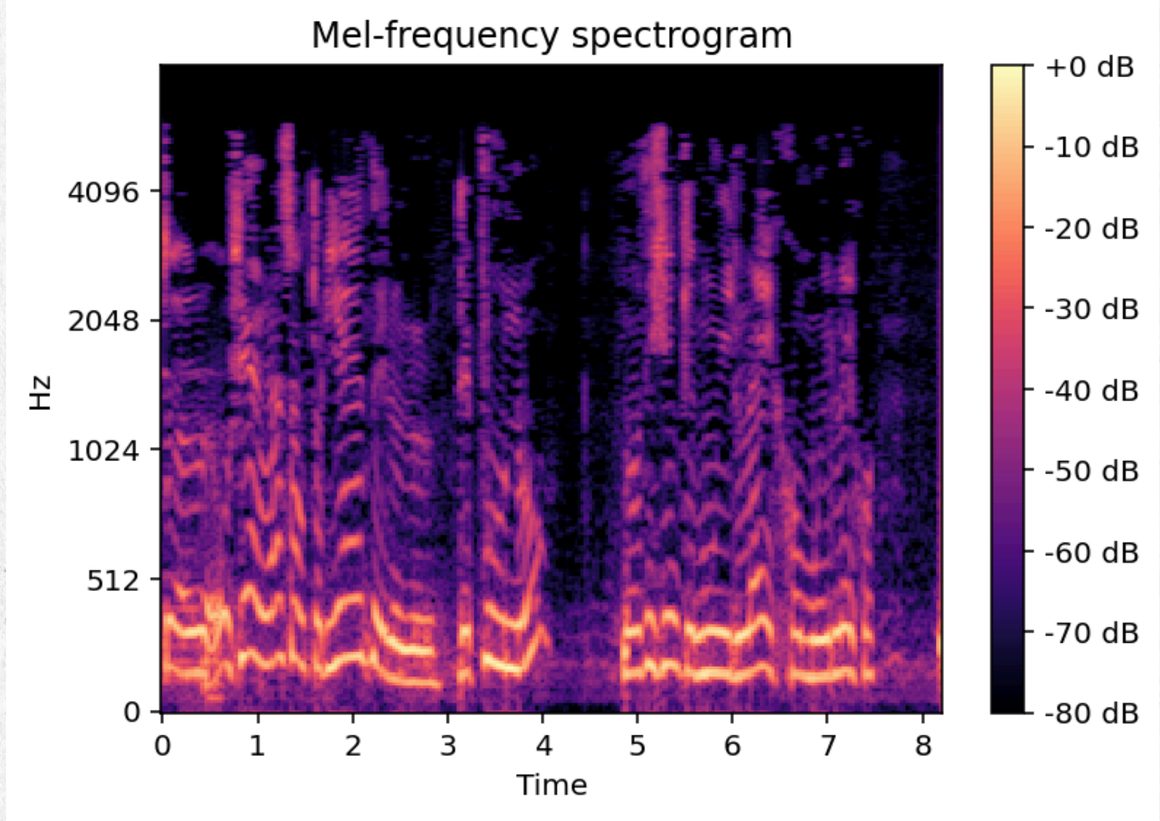


AUDIO PROCESADO

SPECTRAL GATING



AUDIO ORIGINAL



AUDIO PROCESADO

DESCRIPCIÓN DE CARACTERISTICAS PARA ANÁLISIS EXPLORATORIO

Zero Crossing Rate (ZCR)

Es una tasa de cuántas veces la señal cruza el eje horizontal (valor cero) en un intervalo de tiempo. Identifica sonidos de alta frecuencia.

Spectral Centroid

Es una medida que indica el "centro de masa" del espectro de una señal de audio. Representa la frecuencia promedio ponderada por la magnitud de las componentes espectrales. Identifica timbres, caracterización del brillo del sonido.

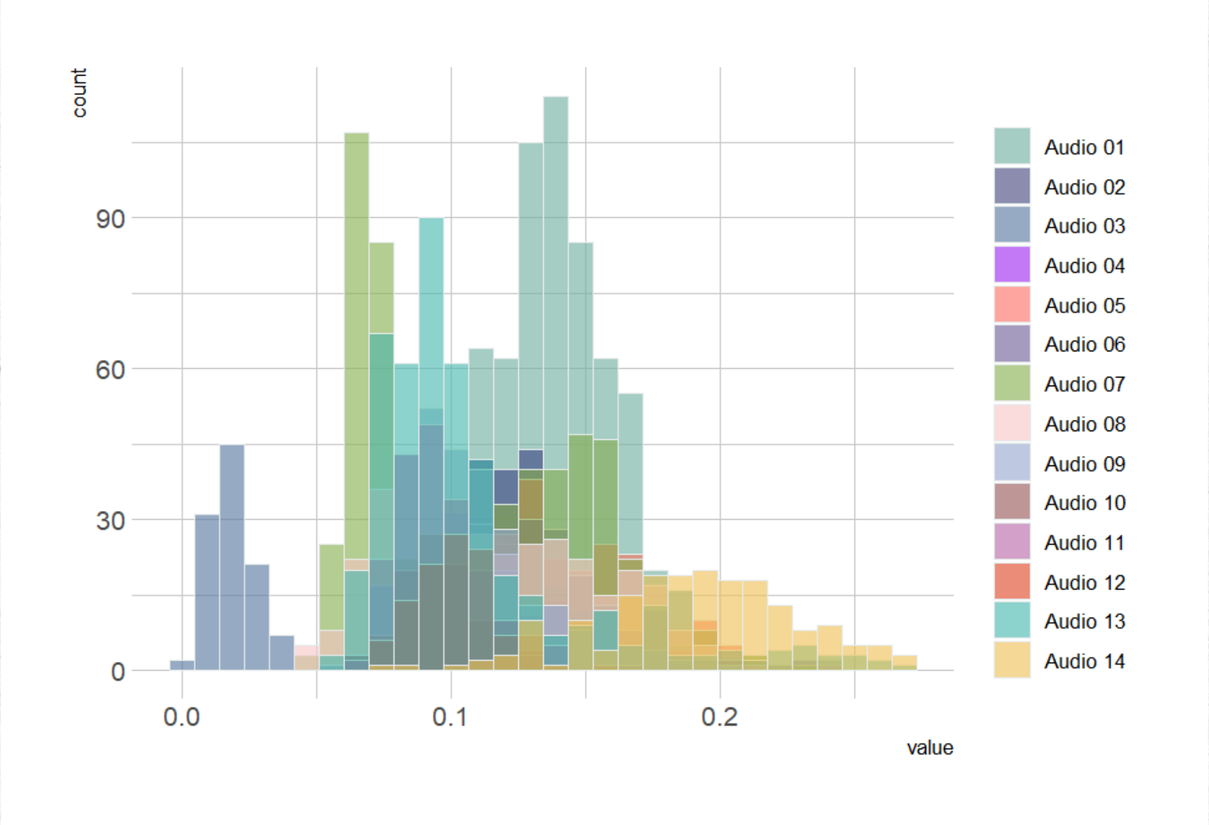
Mel-Frequency Ceptral Coefficients (MFCC)

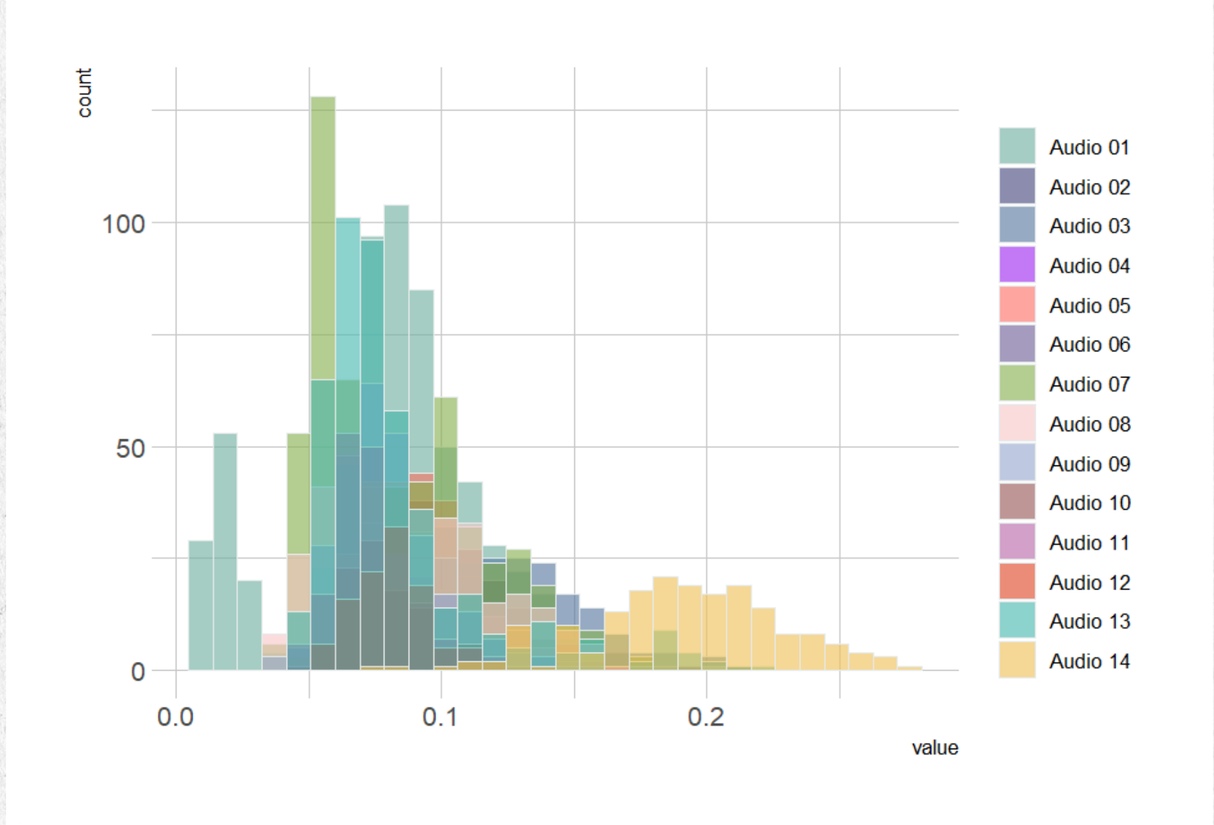
Son características que representan cómo suena una señal de audio en la escala Mel, la cual se asemeja a cómo los humanos perciben las frecuencias sonoras. Capturan las propiedades importantes del sonido.

Root Mean Square (RMS)

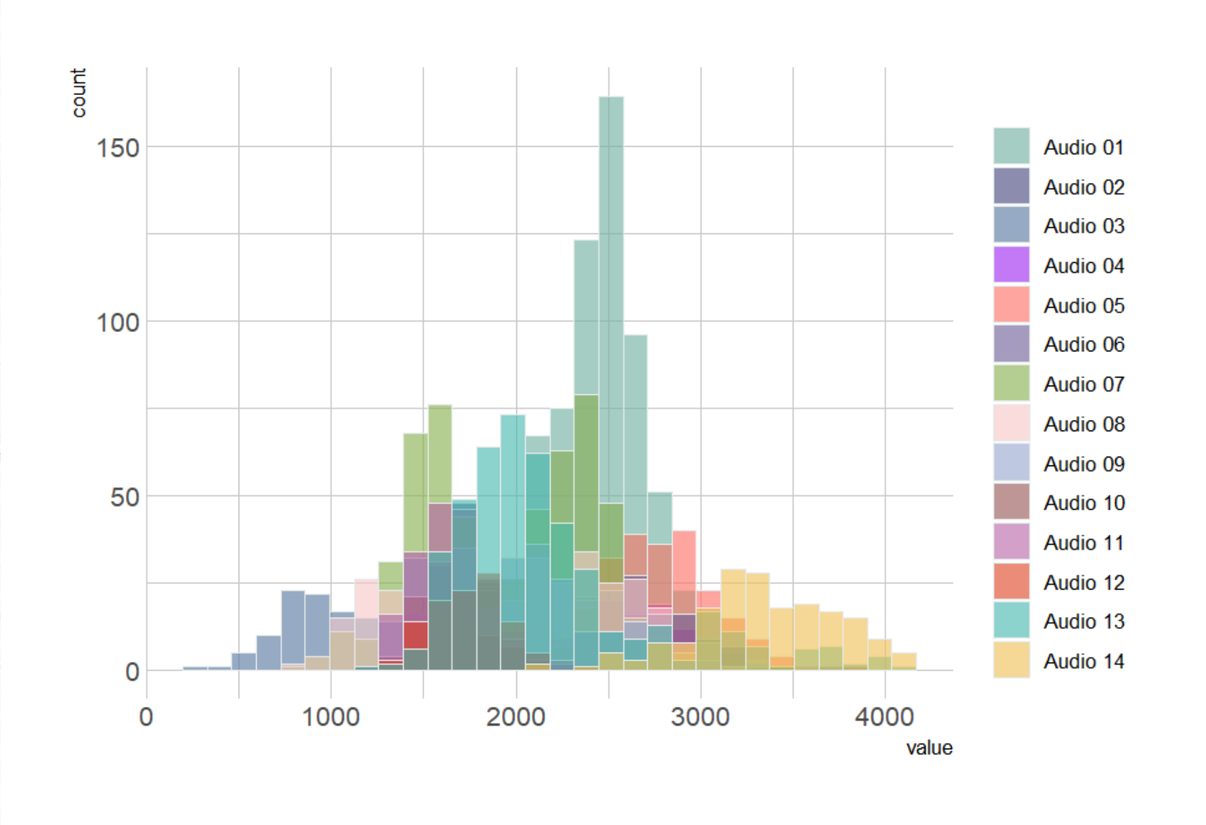
Medida de la energía de la señal de audio. Se calcula como la raíz cuadrada de la media de los cuadrados de los valores de la señal. Mide la intensidad o volumen del audio.

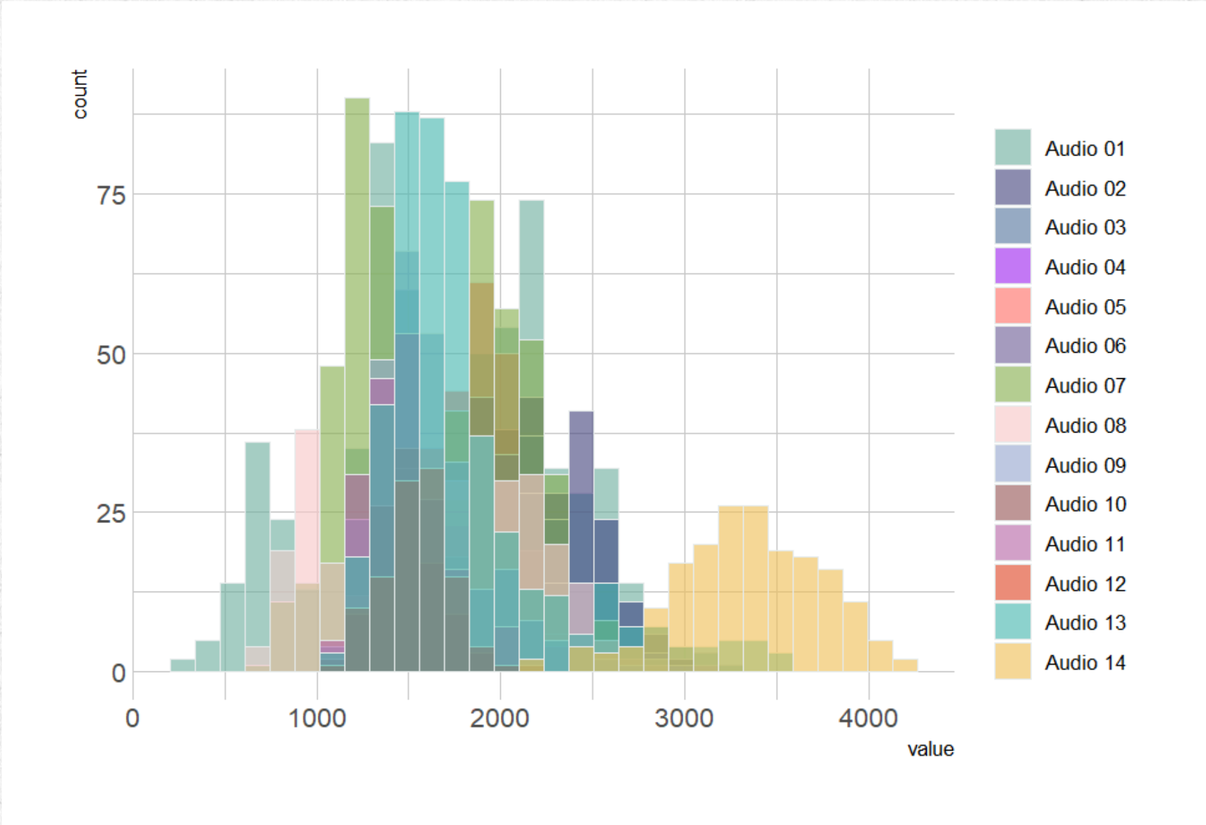
ZERO CROSSING RATE (ZCR)



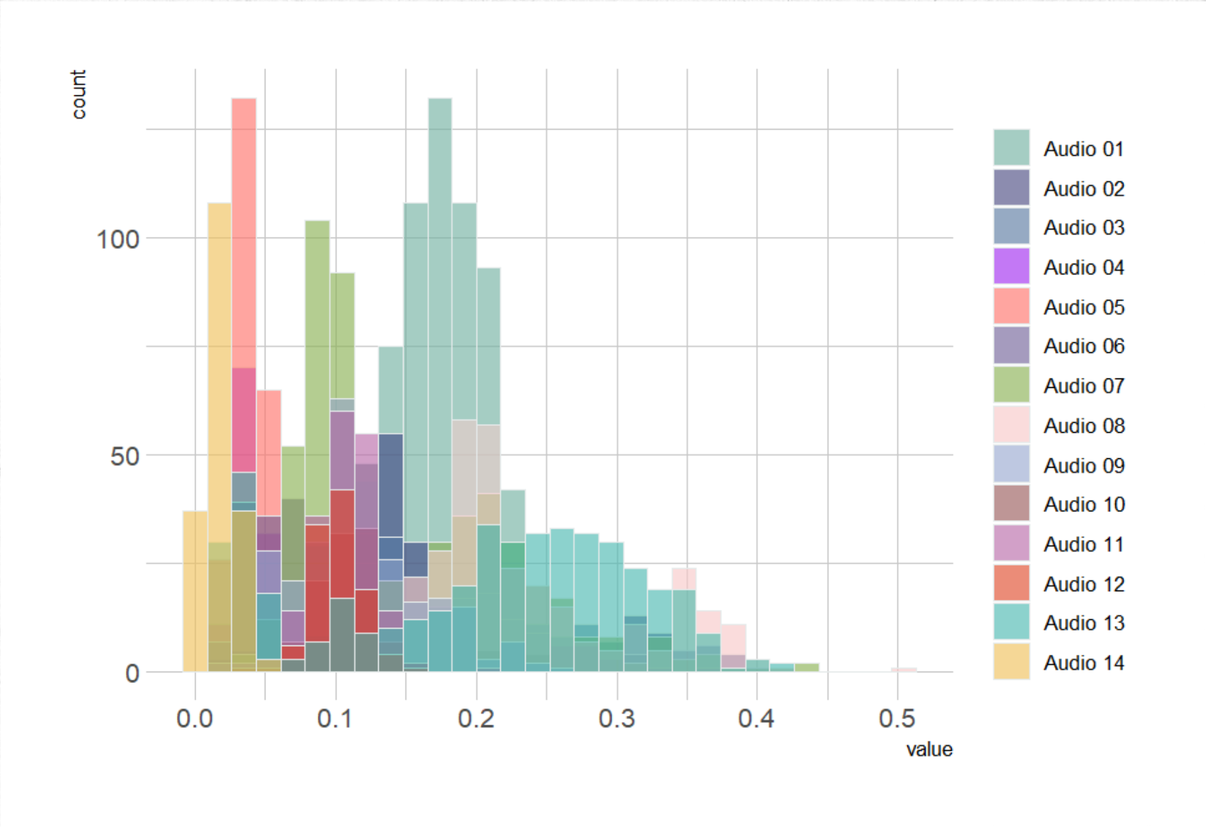


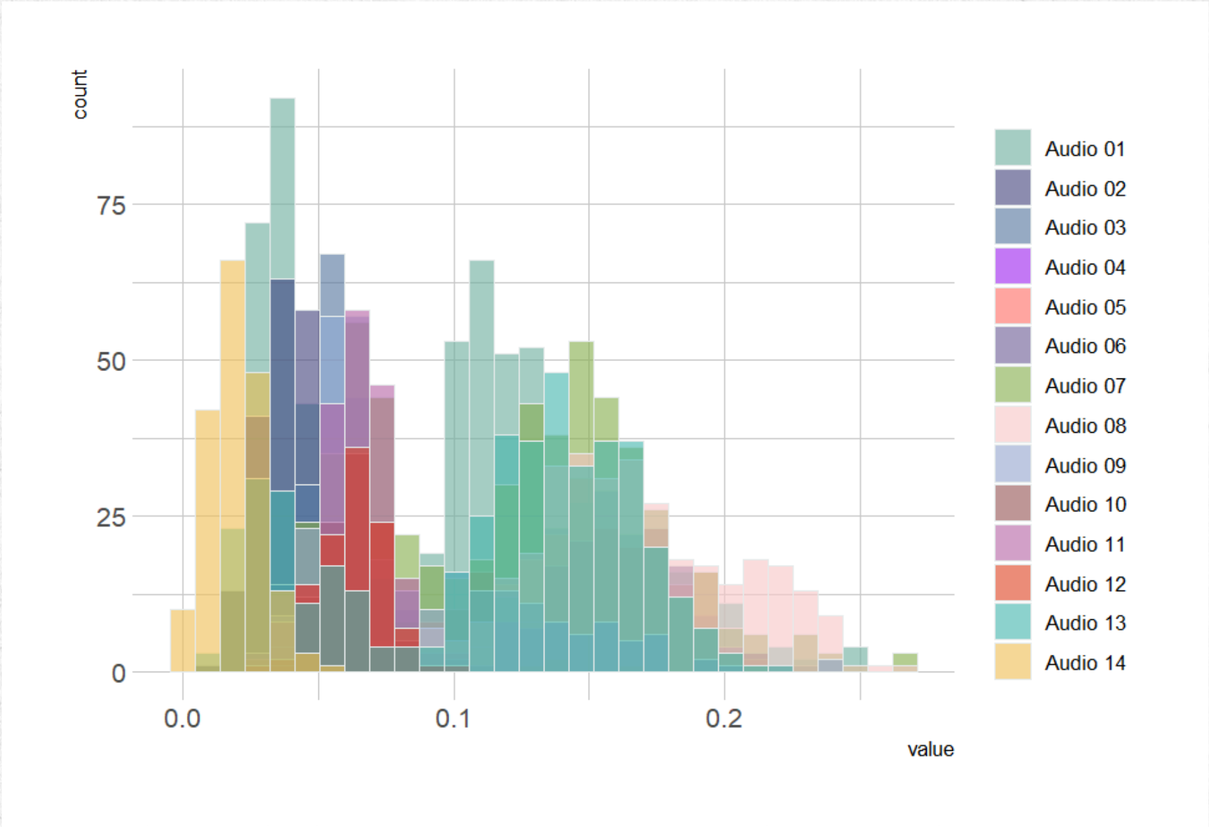
SPECTRAL CENTROID





ROOT MEAN SQUARE (RMS)

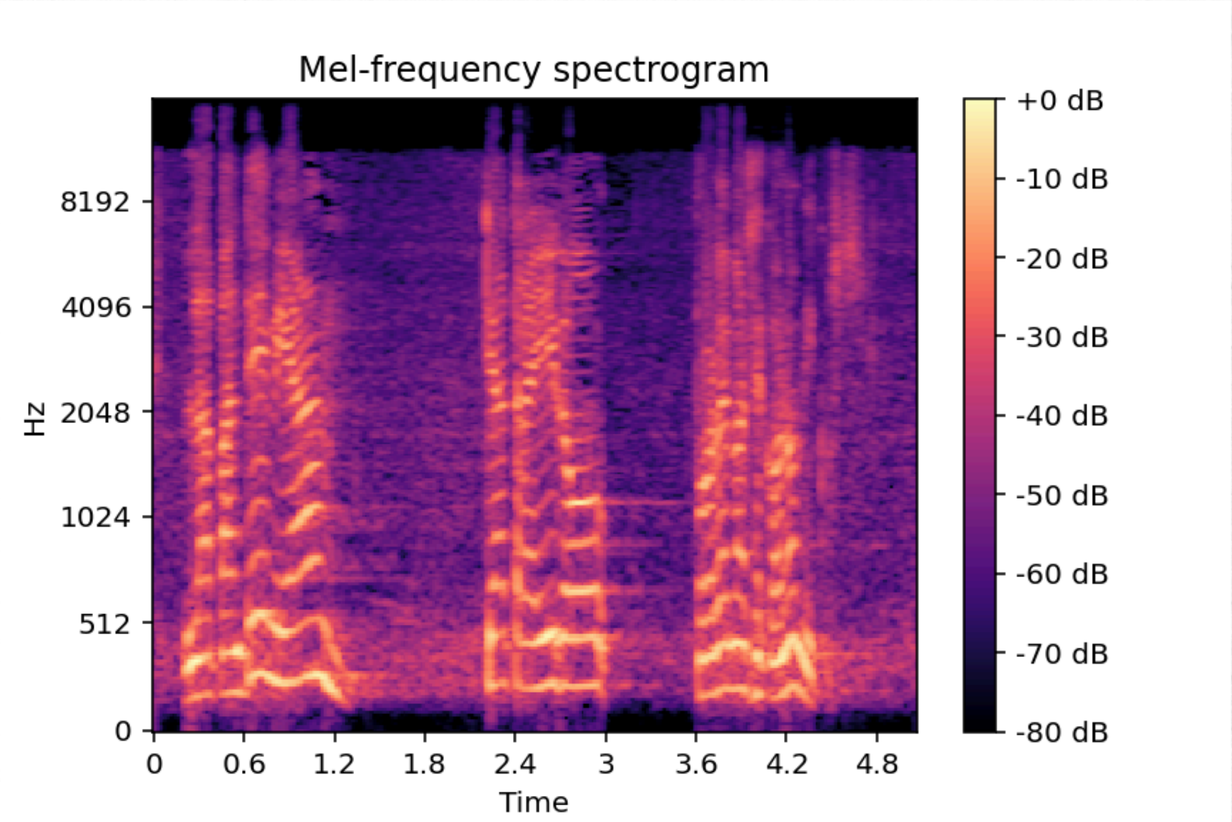




ESPECTOGRAMA DE MEL

El espectrograma es una representación visual de como varían las frecuencias y ocupa las transformadas de Fourier a corto plazo mientras el de Mel aplica la escala de Mel $\longrightarrow$ $2595 \times log_{10}(1 + \frac{Hz}{700})$

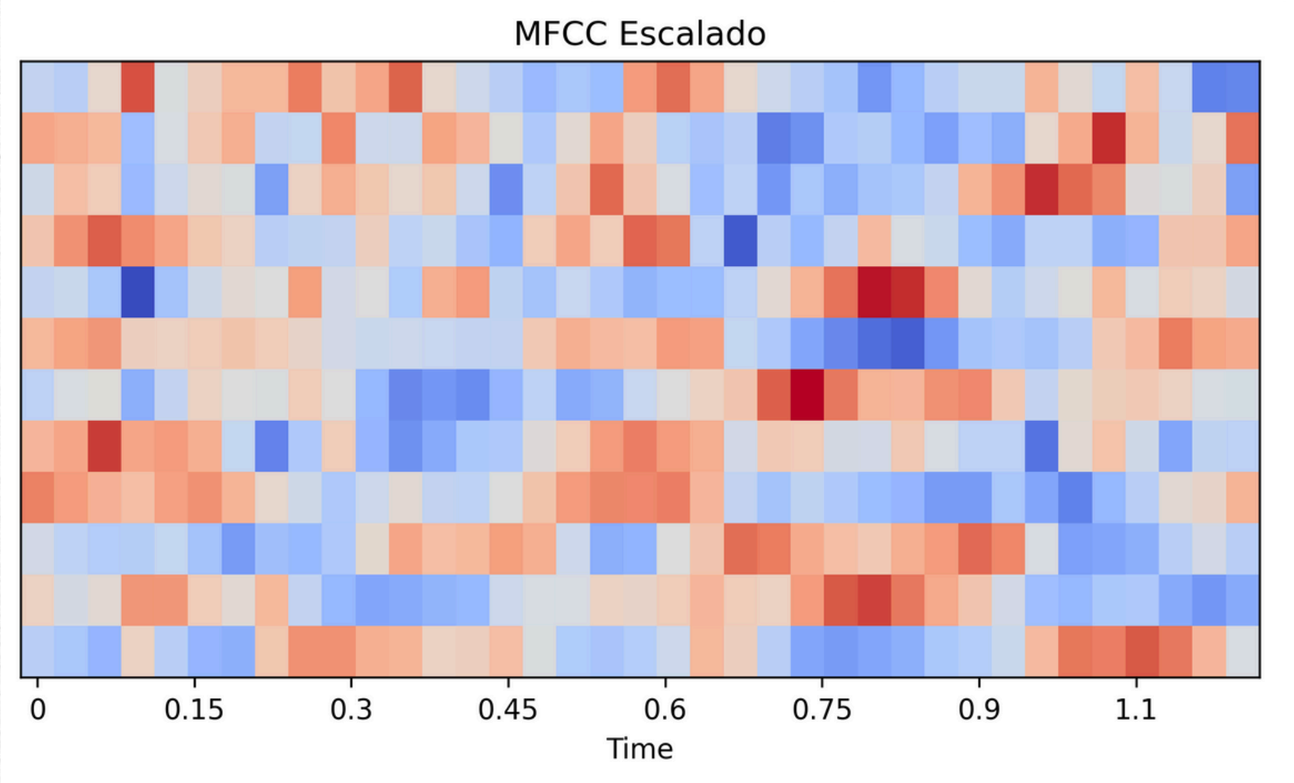
Que es una escala diseñada específicamente para imitar la forma en que los humanos percibimos el sonido. Luego se aplica una serie de filtros de Mel y la salida de cada filtro es la suma ponderada de las energías de las frecuencias que caen dentro del rango del filtro.



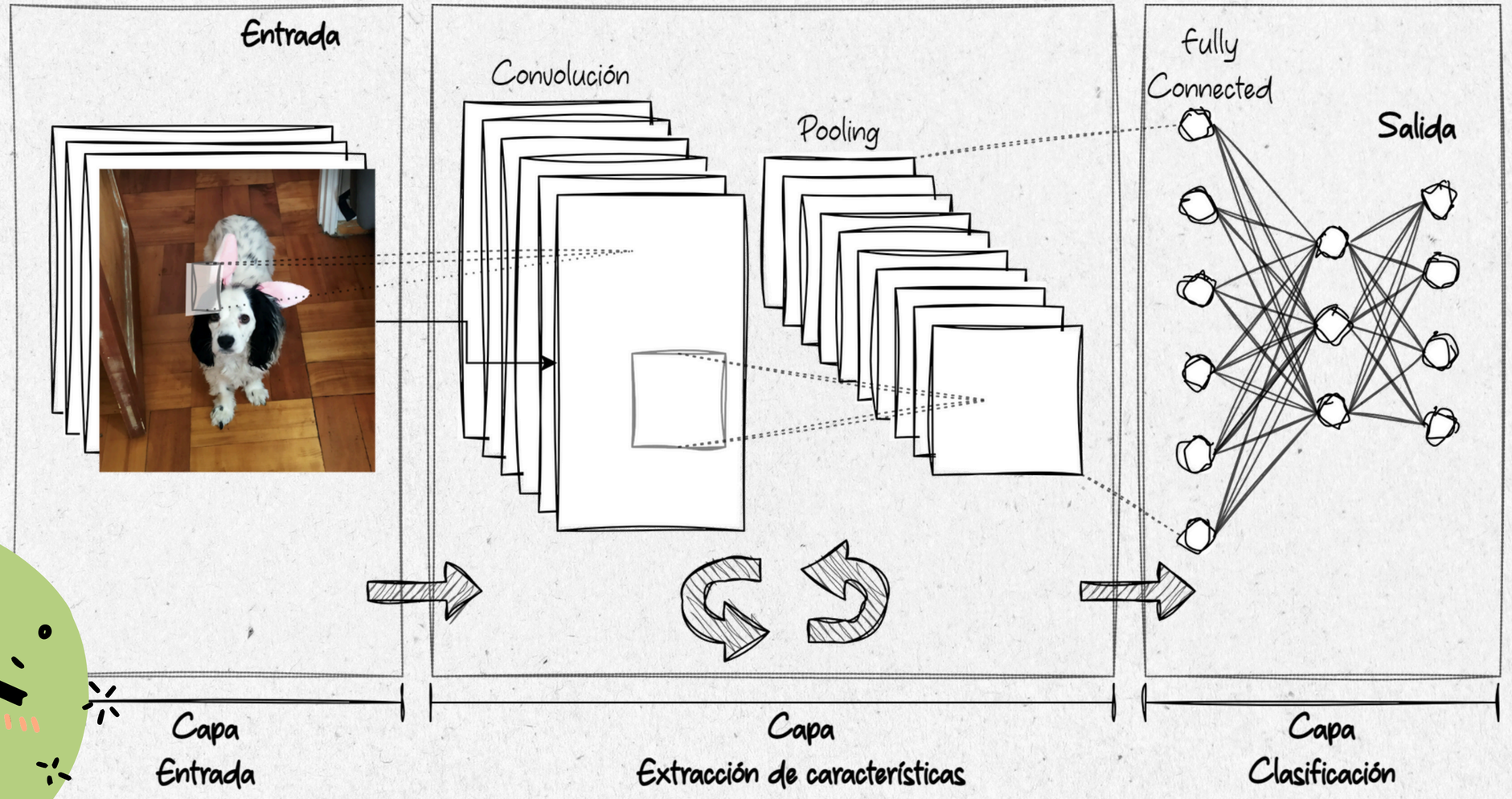
MFCC

Los MFCC son una forma de representar el espectro de una señal de audio en un formato que es más compacto y que captura las características perceptualmente relevantes para el reconocimiento de voz y otros análisis acústicos.

Básicamente, a lo anterior se le aplica logaritmo de las amplitudes y la transformada discreta del coseno (DCT) para obtenerlos coeficientes cepstrales, que son una representación compacta y robusta de las características del audio.



REDES CONVUOLUCIONALES



** Imagen rescatada de Fran Plaza. (n.d.).

REDES CONVOLUCIONALES

Son un tipo especial de redes neuronales para procesar datos que tienen una topología en forma de cuadrícula conocida

01

Convolución

Proceso mediante el cual un filtro (o kernel) se aplica a la imagen de entrada para extraer características específicas, como bordes, texturas y patrones.

02

Max pooling

Técnica de reducción de dimensionalidad que se utiliza para disminuir el tamaño espacial del mapa de características, conservando las características más importantes y reduciendo la complejidad computacional.

03

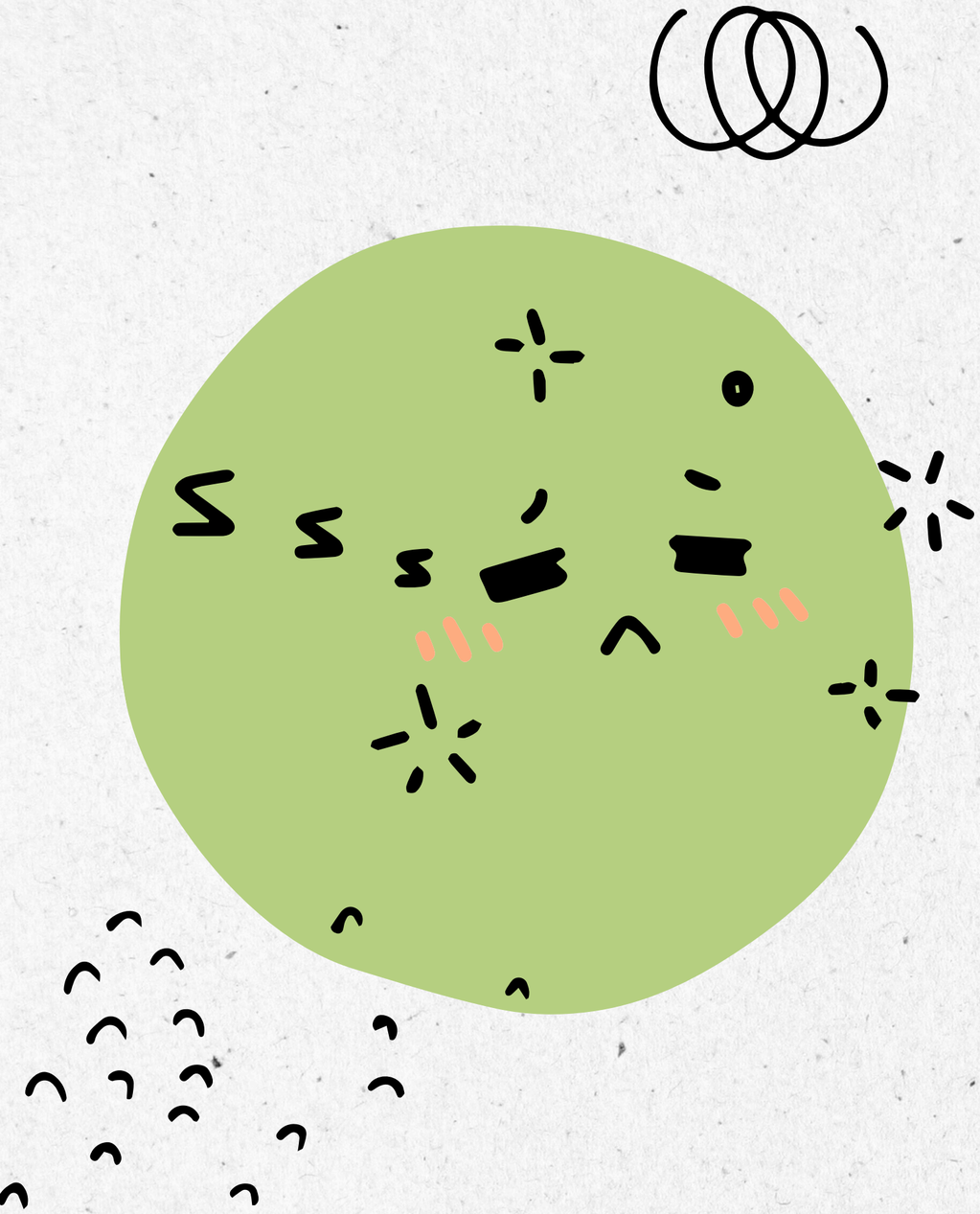
Flattening

Proceso de convertir un mapa de características bidimensional (o multidimensional) en un vector unidimensional.

04

Dropout rate

Corresponde al porcentaje de neuronas que se desactivan en cada entrenamiento



POST PROCESAMIENNTO

FUNCIÓN DE SALIDA

$$\text{softplus}(x) = \ln(1 + e^x)$$

FUNCIÓN DE PERDIDA

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

ARQUITECTURA DEL MODELO

```python
import tensorflow as tf
from tensorflow.keras import layers, models

input_shape = (20, 60, 1)
model = models.Sequential([
    layers.Conv2D(64, kernel_size=(3, 3), activation='relu', input_shape=input_shape),
    layers.Dropout(0.2),
    layers.MaxPooling2D(pool_size=(2, 2)),

    layers.Conv2D(128, kernel_size=(3, 3), activation='relu'),
    layers.Dropout(0.2),
    layers.MaxPooling2D(pool_size=(2, 2)),

    layers.Conv2D(256, kernel_size=(3, 3), activation='relu'),
    layers.Dropout(0.2),

    # Aplanar los datos
    layers.Flatten(),

    layers.Dense(3, activation=lambda x: tf.keras.activations.softplus(x) + 1)
])
```

Model: "sequential_9"

Layer (type)	Output Shape	Param #
conv2d_27 (Conv2D)	(None, 18, 58, 64)	640
dropout_27 (Dropout)	(None, 18, 58, 64)	0
max_pooling2d_18 (MaxPooling2D)	(None, 9, 29, 64)	0
conv2d_28 (Conv2D)	(None, 7, 27, 128)	73,856
dropout_28 (Dropout)	(None, 7, 27, 128)	0
max_pooling2d_19 (MaxPooling2D)	(None, 3, 13, 128)	0
conv2d_29 (Conv2D)	(None, 1, 11, 256)	295,168
dropout_29 (Dropout)	(None, 1, 11, 256)	0
flatten_9 (Flatten)	(None, 2816)	0
dense_9 (Dense)	(None, 3)	8,451

Total params: 378,115 (1.44 MB)
Trainable params: 378,115 (1.44 MB)
Non-trainable params: 0 (0.00 B)

MUCHAS GRACIAS

Por su atención

C1: CAPTURA VARIACIONES A GRAN ESCALA EN EL ESPECTRO DE FRECUENCIA, RELACIONADAS CON LA FORMA GENERAL DEL ESPECTRO.

C2 A C4: ESTOS COEFICIENTES TIENDEN A CAPTURAR INFORMACIÓN SOBRE LA FORMA DEL ESPECTRO EN FRECUENCIAS MEDIAS A ALTAS, COMO LA RESONANCIA DE LA VOCAL (FORMANTES EN EL HABLA).

C5 A C8: REPRESENTAN CARACTERÍSTICAS MÁS FINAS DEL ESPECTRO, INCLUYENDO VARIACIONES EN LAS FORMANTES Y OTROS DETALLES ESPECTRALES.

C9 A C12: ESTOS COEFICIENTES CAPTURAN DETALLES DE ALTA FRECUENCIA Y PUEDEN SER SENSIBLES A RUIDOS O CAMBIOS SUTILES EN EL ESPECTRO QUE NO SON TAN PERCEPTIBLES A NIVEL HUMANO PERO PUEDEN SER IMPORTANTES PARA DISTINCIONES MÁS FINAS EN EL RECONOCIMIENTO DE SONIDOS.