

# **Práctica: Clasificación de Sonidos.**

## **Reconocimiento de Voz**

Escamilla Resendiz Aldo

11 de marzo de 2025

# Índice general

1.	Introducción . . . . .	2
2.	Desarrollo . . . . .	2
2.1.	Herramientas Utilizadas . . . . .	2
2.2.	Metodología de Desarrollo . . . . .	2
2.3.	Implementación del Modelo de Clasificación . . . . .	3
2.4.	Resultados Esperados . . . . .	3
3.	Resultados . . . . .	4
3.1.	Audio 1 - La casa es grande. . . . .	4
3.2.	Audio 2 - Mi perro se salió de casa. . . . .	6
3.3.	Audio 3 - Casa es a lo que llamo hogar. . . . .	8
3.4.	Audio 4 - Los niños juegan en el parque. . . . .	10
3.5.	Audio 5 - En la mañana me preparo y salgo a la escuela. . . . .	12
4.	Métricas . . . . .	14
5.	Conclusiones . . . . .	16

# 1. Introducción

En esta práctica, se realizó el procesamiento de archivos de audio con el objetivo de segmentarlos en intervalos de 100 milisegundos y extraer características clave de cada segmento. Entre las características analizadas se encuentran la **amplitud media**, el **valor RMS** (*Root Mean Square*) y la **tasa de cruces por cero**. Estas métricas permiten obtener una representación detallada del comportamiento de la señal de audio a lo largo del tiempo.

Para la clasificación de los segmentos de audio, se utilizó un **árbol de decisión**, **Inferencia bayesiana**, **Red Neuronal**, algoritmos de aprendizaje supervisado que permite etiquetar cada segmento en función de su contenido sonoro. En este caso, las etiquetas asignadas fueron **S**, **U** y **V**, las cuales representan diferentes categorías dentro del audio procesado. El modelo de clasificación fue entrenado con estos datos y posteriormente evaluado para medir su precisión en la predicción de nuevas muestras.

El modelo generado facilita la identificación de patrones en los segmentos de audio y permite visualizar cómo las características extraídas influyen en la clasificación. Este procedimiento es aplicable en diversas áreas, como el reconocimiento de voz, la detección de eventos acústicos y la categorización automática de sonidos.

## 2. Desarrollo

### 2.1 Herramientas Utilizadas

Para la realización de esta práctica, se utilizaron diversas herramientas de software y librerías especializadas en el procesamiento de señales de audio y aprendizaje automático. Entre ellas destacan:

- **Python**: Lenguaje de programación utilizado para el desarrollo de los scripts de procesamiento de audio y entrenamiento del modelo de clasificación.
- **Librerías de Python**:
  - **NumPy**: Utilizada para el manejo de datos numéricos y la manipulación de arreglos.
  - **pandas**: Empleada para organizar y gestionar los datos extraídos de los segmentos de audio.
  - **Matplotlib**: Usada para la visualización de la forma de onda de los audios segmentados.
  - **Scikit-learn**: Librería especializada en aprendizaje automático, utilizada para la implementación del árbol de decisión.
  - **wave**: Módulo de Python que permite la lectura y manipulación de archivos de audio en formato WAV.

### 2.2 Metodología de Desarrollo

El procedimiento seguido en esta práctica se dividió en varias etapas, desde la obtención de los datos hasta la implementación del modelo de clasificación. La metodología aplicada se detalla a continuación:

1. **Obtención de Datos**: Se utilizaron archivos de audio en formato WAV, los cuales fueron procesados para su segmentación y análisis.
2. **Procesamiento de Datos**: Cada archivo de audio se dividió en segmentos de 100 milisegundos. De cada segmento se extrajeron características relevantes para la clasificación, tales como:
  - **Amplitud media**: Representa la magnitud promedio de la señal en cada segmento.
  - **Valor RMS (Root Mean Square)**: Indica la energía de la señal en cada intervalo de tiempo.
  - **Tasa de Cruces por Cero**: Mide la cantidad de veces que la señal cambia de signo en un segmento, lo cual es útil para determinar la presencia de sonido o silencio.
3. **Implementación de Metodología**:
  - a) Se definieron las etiquetas para cada segmento de audio (**S**, **U**, **V**).
  - b) Se dividió el conjunto de datos en entrenamiento (80 %) y prueba (20 %).
  - c) Se entrenó el árbol de decisión utilizando el criterio *Gini* y una profundidad máxima de 5.
  - d) Se evaluó el modelo en los datos de prueba y se midió su precisión.

## 2.3 Implementación del Modelo de Clasificación

La implementación del modelo de clasificación se realizó en **Python** utilizando la librería **Scikit-learn**. A continuación, se muestra el código utilizado para entrenar el árbol de decisión:

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier, plot_tree
import matplotlib.pyplot as plt

# Cargar datos
df = pd.read_csv("segmentos_audio.csv")

# Asignar etiquetas manualmente
etiquetas = ["S"] * 12 + ["V"] * 6 + ["S"] * 5 + ["U"] * 3 + ["S"] * 6
df = df.iloc[:len(etiquetas)].copy()
df["label"] = etiquetas

# Definir características y etiquetas
X = df[["mean_amplitude", "rms", "zero_crossing_rate"]]
y = df["label"]

# Dividir en entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Entrenar el árbol de decisión
clf = DecisionTreeClassifier(criterion="gini", max_depth=5, random_state=42)
clf.fit(X_train, y_train)

# Evaluar el modelo
accuracy = clf.score(X_test, y_test)
print(f"Precisión del modelo: {accuracy:.2f}")

# Graficar el árbol de clasificación
plt.figure(figsize=(12, 6))
plot_tree(clf, feature_names=X.columns, class_names=["S", "U", "V"], filled=True)
plt.title("Árbol de Decisión para Clasificación de Segmentos de Audio")
plt.show()
```

El árbol de decisión permite clasificar los segmentos de audio de acuerdo con sus características acústicas. Esto facilita la detección automática de patrones dentro de la señal y su posterior análisis.

## 2.4 Resultados Esperados

Al finalizar la implementación de la metodología, se espera obtener:

- Una segmentación clara de los archivos de audio en intervalos de 100 ms.
- Un conjunto de datos con las características extraídas de cada segmento.
- Un modelo de clasificación basado en un árbol de decisión con una precisión aceptable en la predicción de nuevas muestras.
- Gráficos que representen la forma de onda de los audios segmentados con sus respectivas etiquetas.
- Una visualización clara del árbol de clasificación utilizado.

## 3. Resultados

### 3.1 Audio 1 - La casa es grande.

- Dominio del tiempo con segmentación de 100 ms y etiquetas (S, U, V)

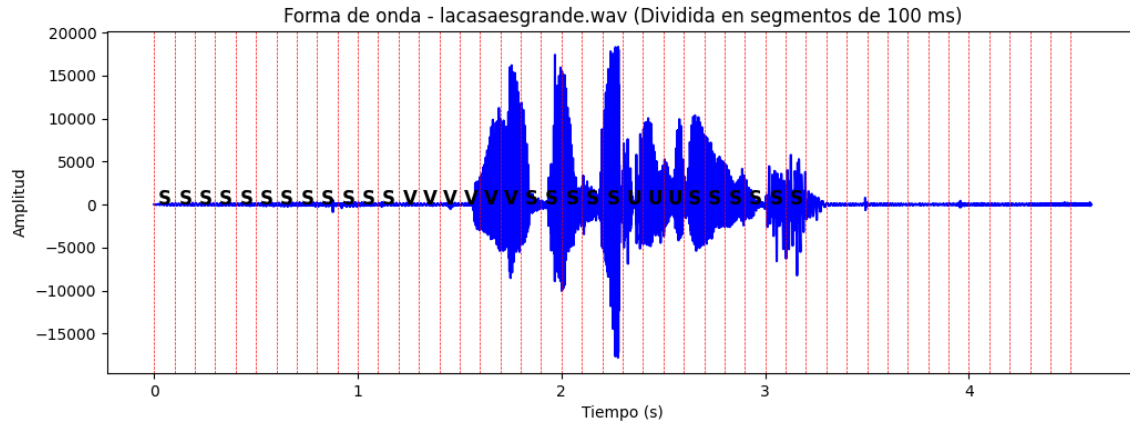


Figura 1: Forma de onda del audio: La casa es grande, con segmentación de 100 ms

- Espectrograma de banda ancha

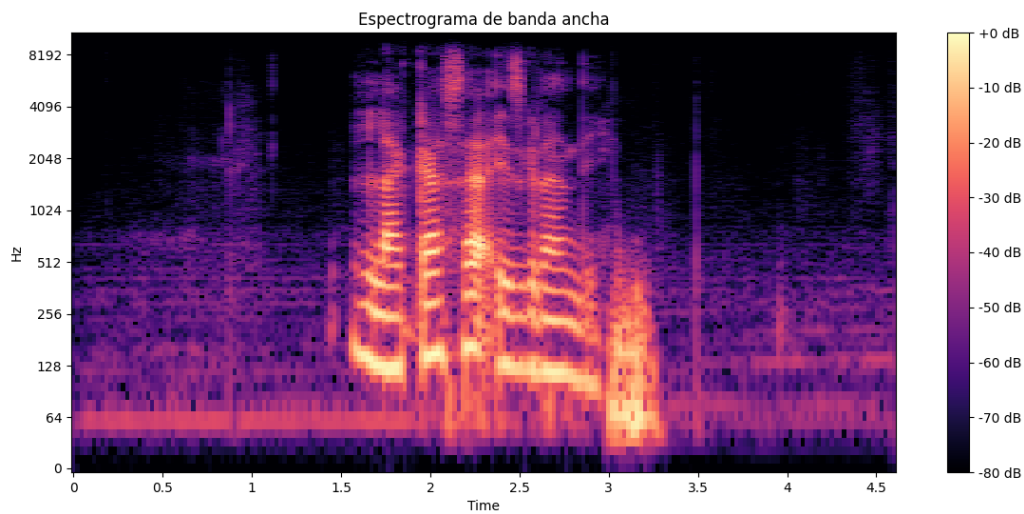


Figura 2: Espectrograma de banda ancha del audio: La casa es grande

- Espectrograma de banda estrecha

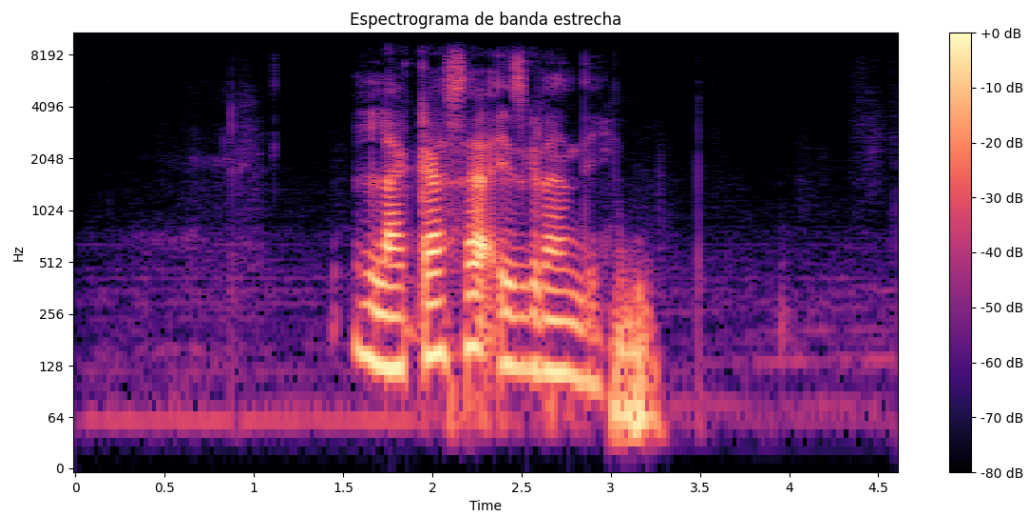


Figura 3: Espectrograma de banda estrecha del audio: La casa es grande

### 3.2 Audio 2 - Mi perro se salió de casa.

- Dominio del tiempo con segmentación de 100 ms y etiquetas (S, U, V)

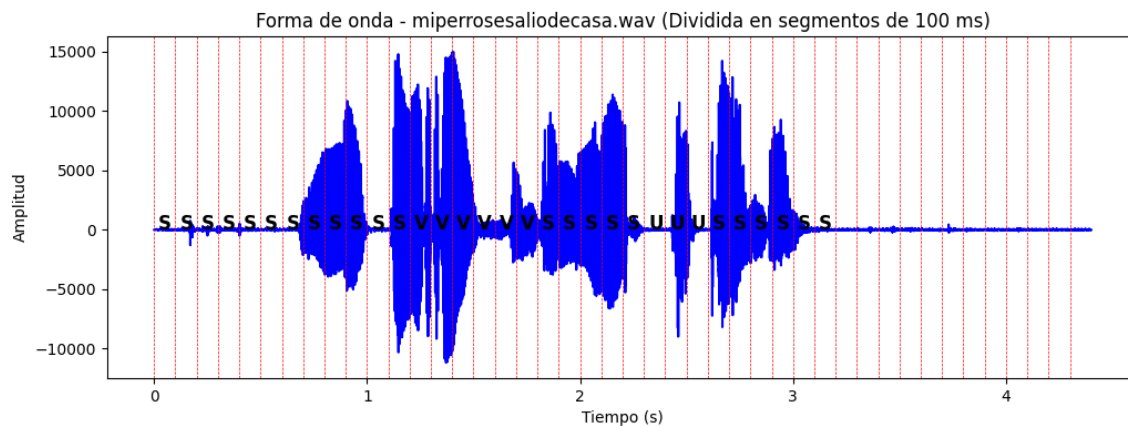


Figura 4: Forma de onda del audio: Mi perro se salió de casa

- Espectrograma de banda ancha

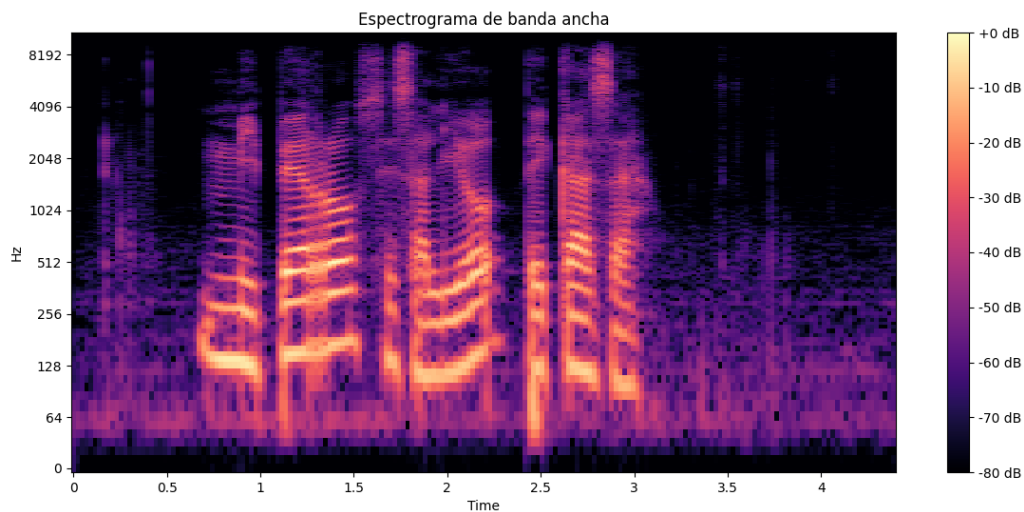


Figura 5: Espectrograma de banda ancha del audio: Mi perro se salió de casa

- Espectrograma de banda estrecha

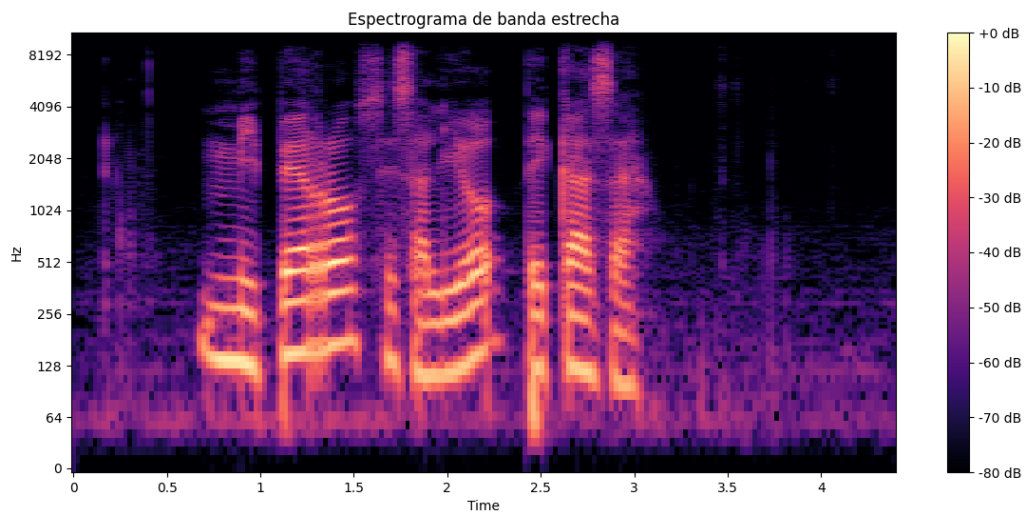


Figura 6: Espectrograma de banda estrecha del audio: Mi perro se salió de casa



### 3.3 Audio 3 - Casa es a lo que llamo hogar.

- Dominio del tiempo con segmentación de 100 ms y etiquetas (S, U, V)

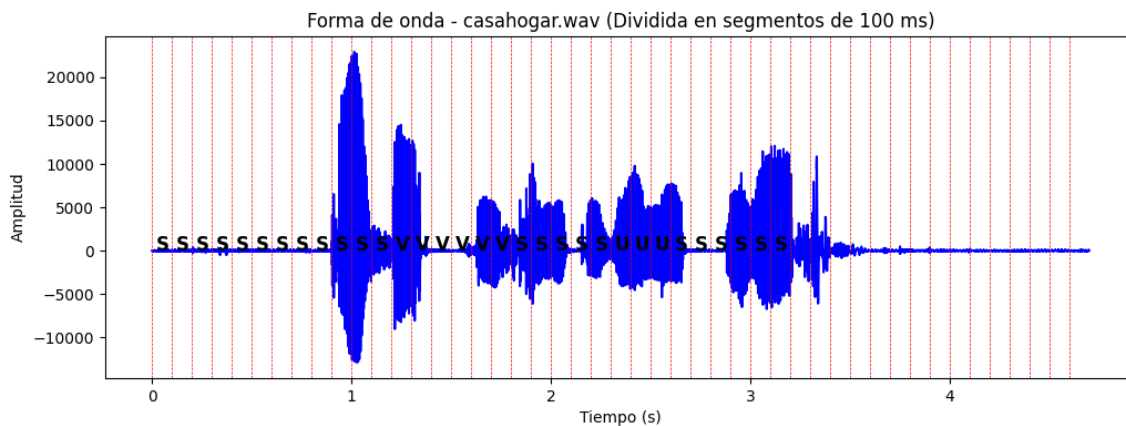


Figura 7: Forma de onda del audio: Casa es a lo que llamo hogar.

- Espectrograma de banda ancha

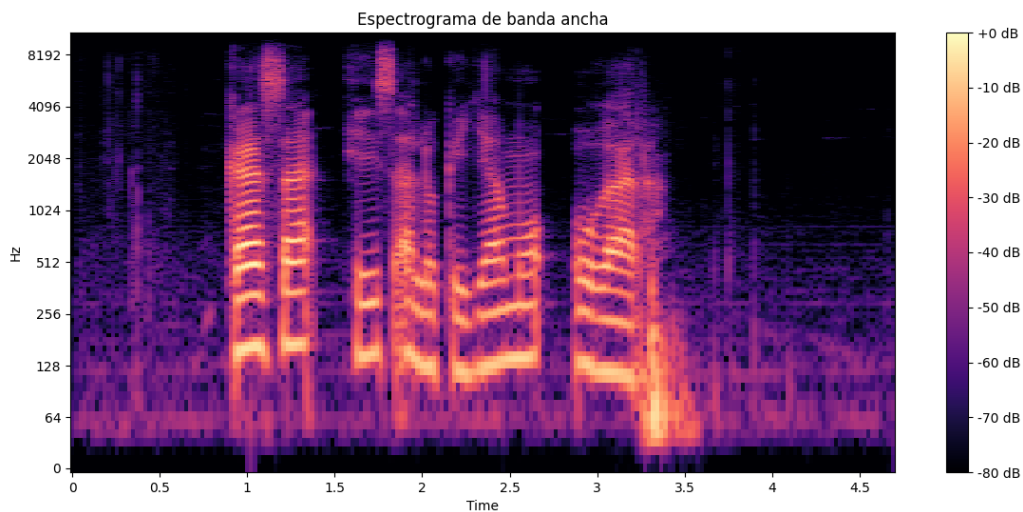


Figura 8: Espectrograma de banda ancha del audio: Casa es a lo que llamo hogar.

- Espectrograma de banda estrecha

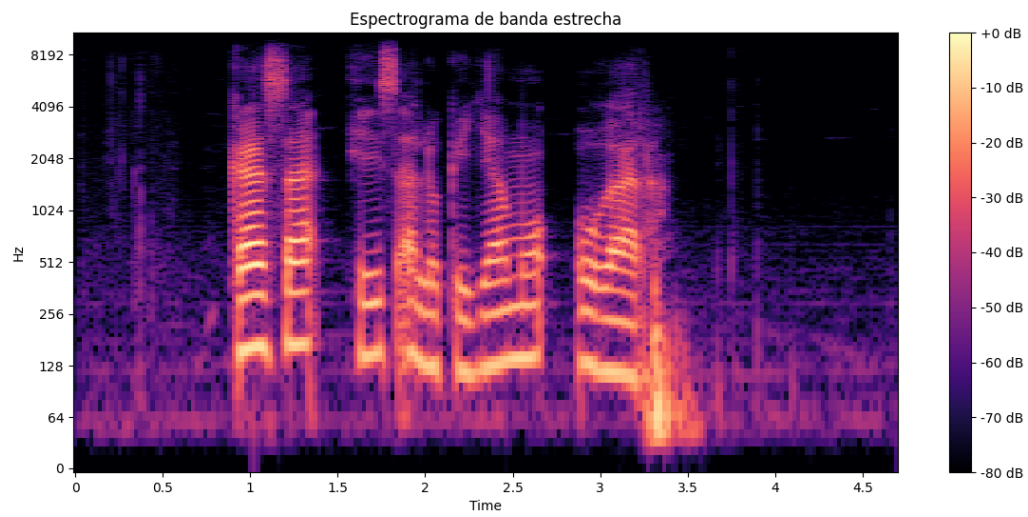


Figura 9: Espectrograma de banda estrecha del audio: Casa es a lo que llamo hogar.

### 3.4 Audio 4 - Los niños juegan en el parque.

- Dominio del tiempo con segmentación de 100 ms y etiquetas (S, U, V)

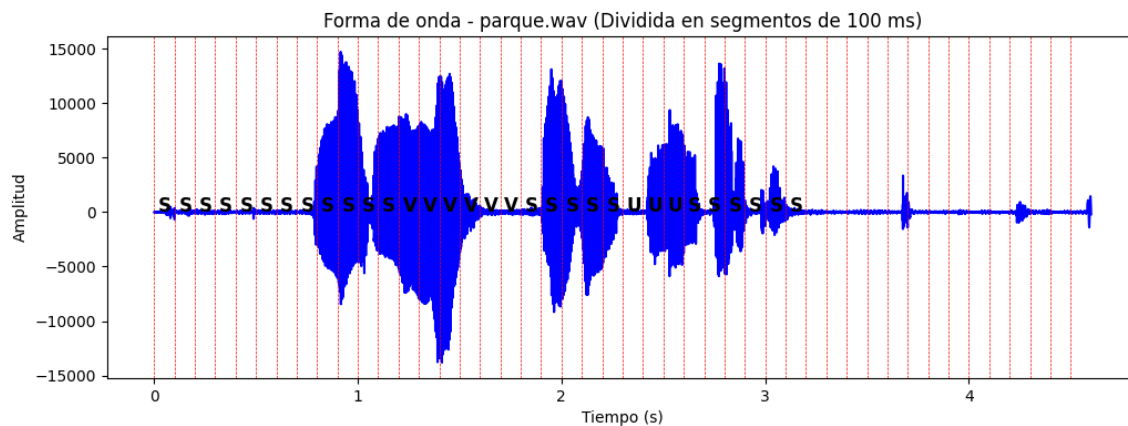


Figura 10: Forma de onda del audio: Los niños juegan en el parque.

- Espectrograma de banda ancha

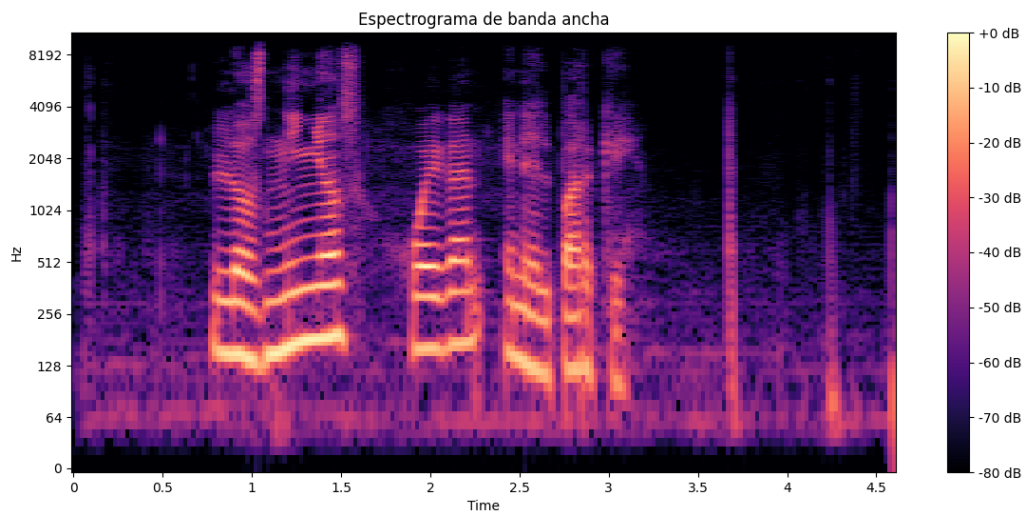


Figura 11: Espectrograma de banda ancha del audio: Los niños juegan en el parque.

## ■ Espectrograma de banda estrecha

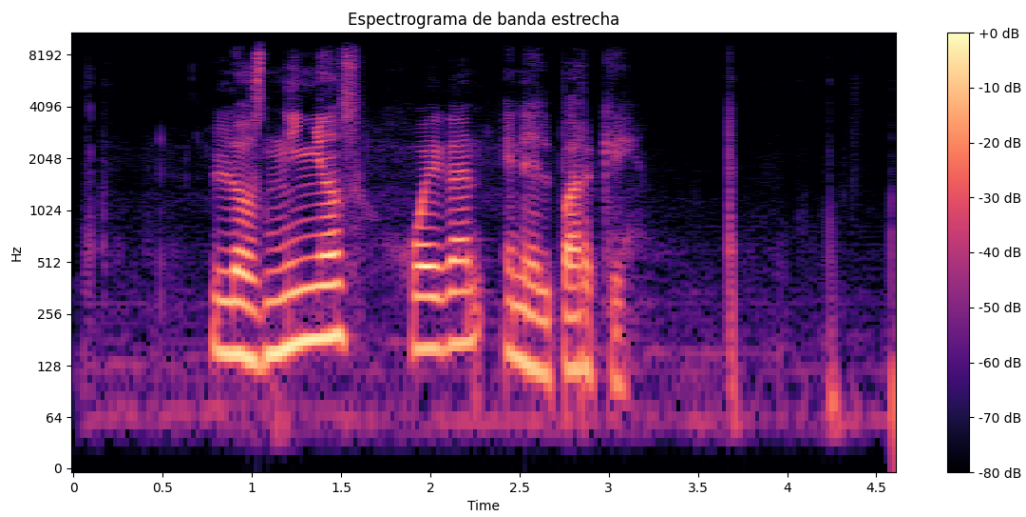


Figura 12: Espectrograma de banda estrecha del audio: Los niños juegan en el parque.

### 3.5 Audio 5 - En la mañana me preparo y salgo a la escuela.

- Dominio del tiempo con segmentación de 100 ms y etiquetas (S, U, V)

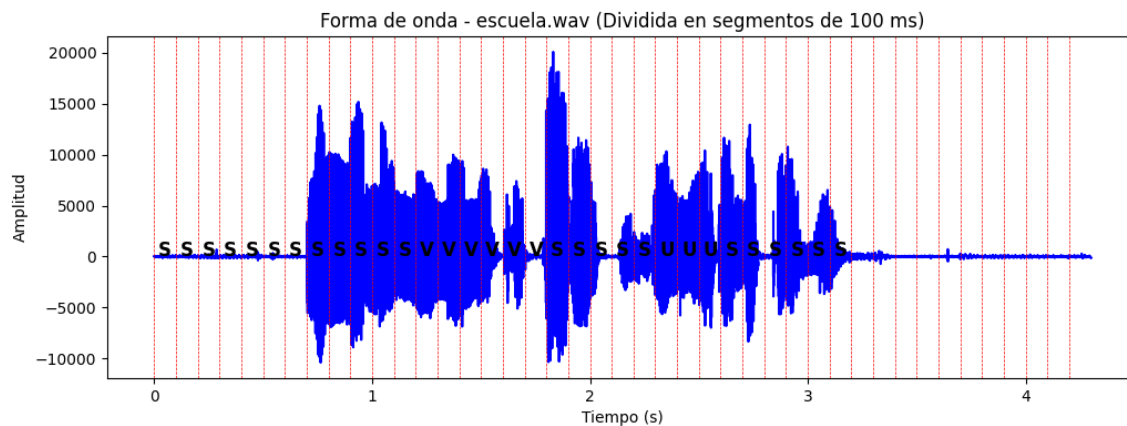


Figura 13: Forma de onda del audio: En la mañana me preparo y salgo a la escuela.

- Espectrograma de banda ancha

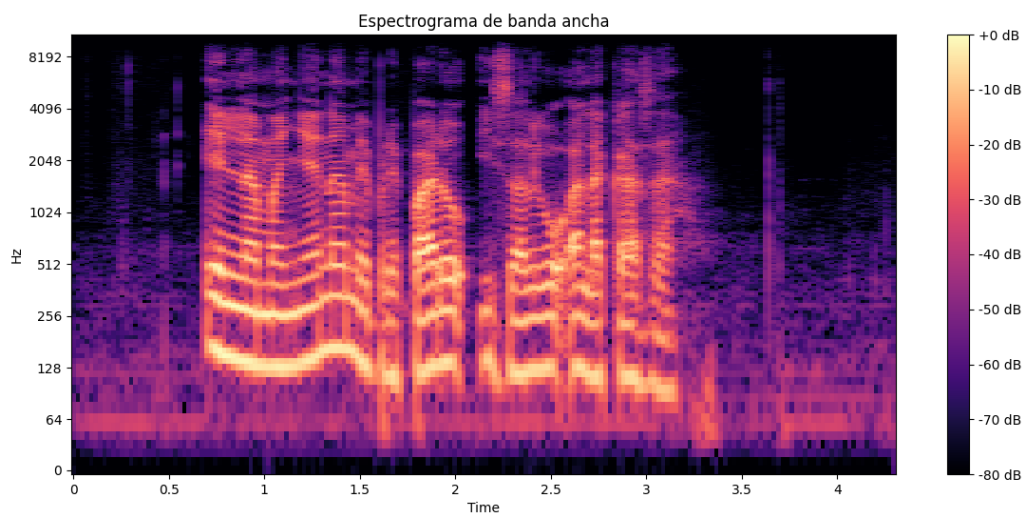


Figura 14: Espectrograma de banda ancha del audio: En la mañana me preparo y salgo a la escuela.

- Espectrograma de banda estrecha

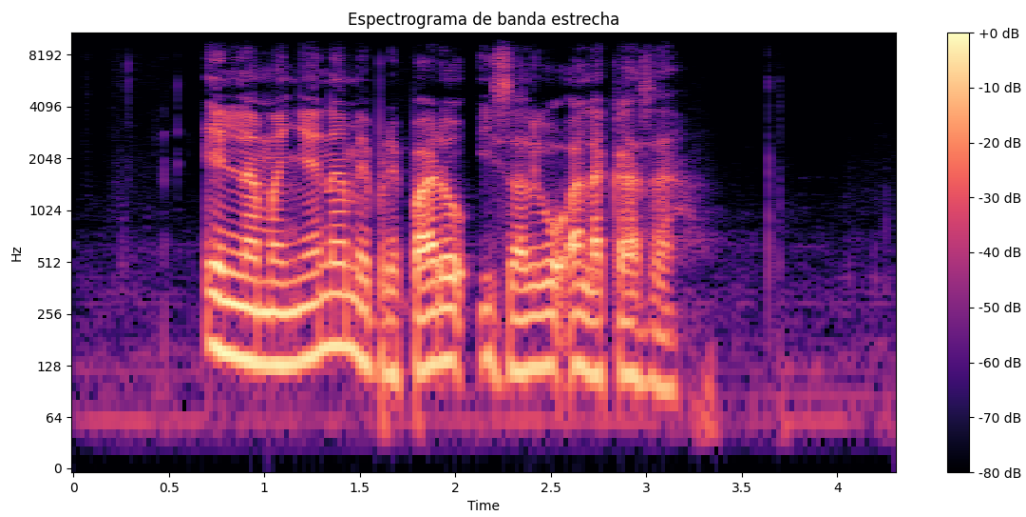


Figura 15: Espectrograma de banda estrecha del audio: En la mañana me preparo y salgo a la escuela.

## 4. Métricas

- Inferencia Bayesiana

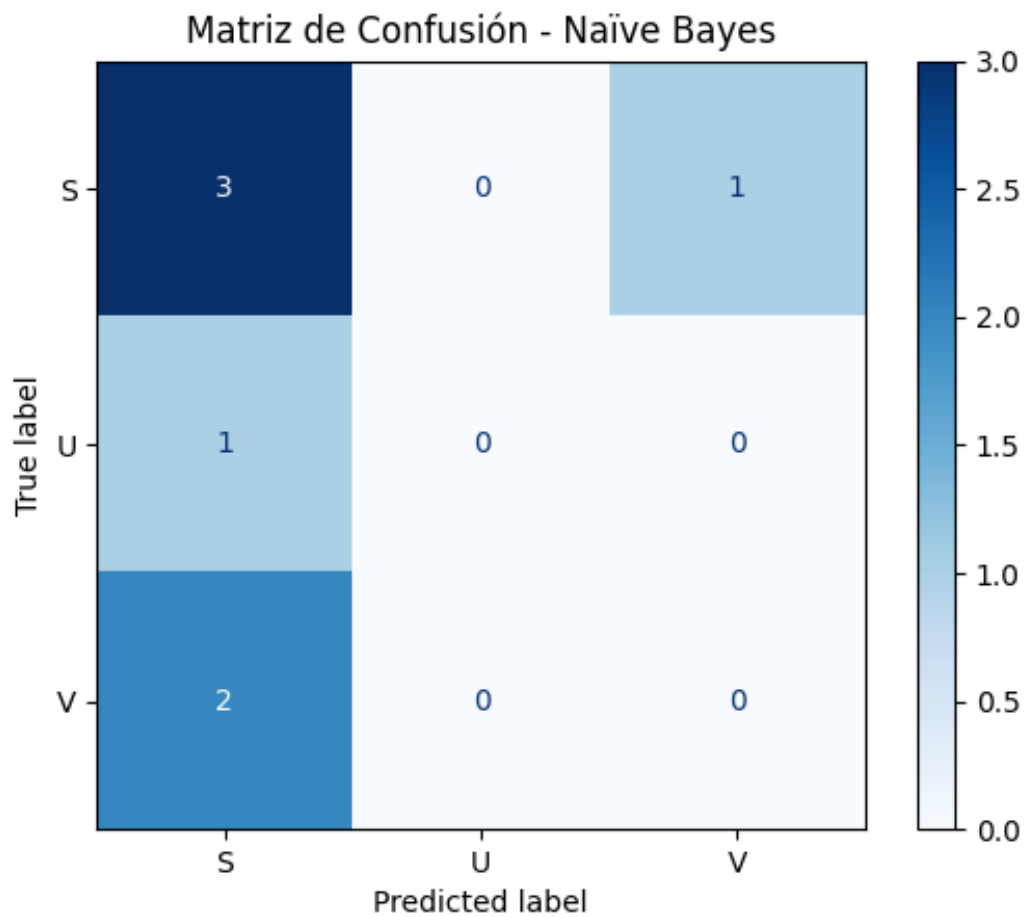


Figura 16: Inferencia Bayesiana

## ■ Red Neuronal

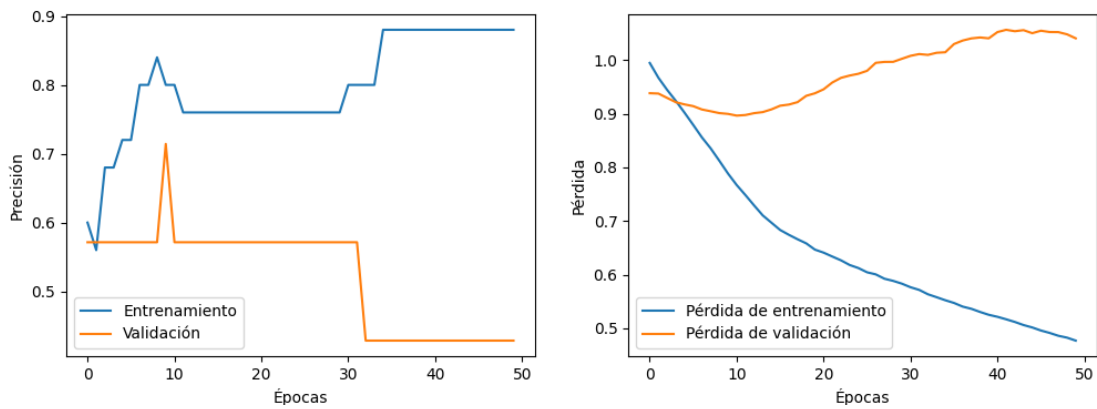


Figura 17: Red Neuronal

## ■ Árbol de Decisión

Árbol de Decisión para Clasificación de Segmentos de Audio

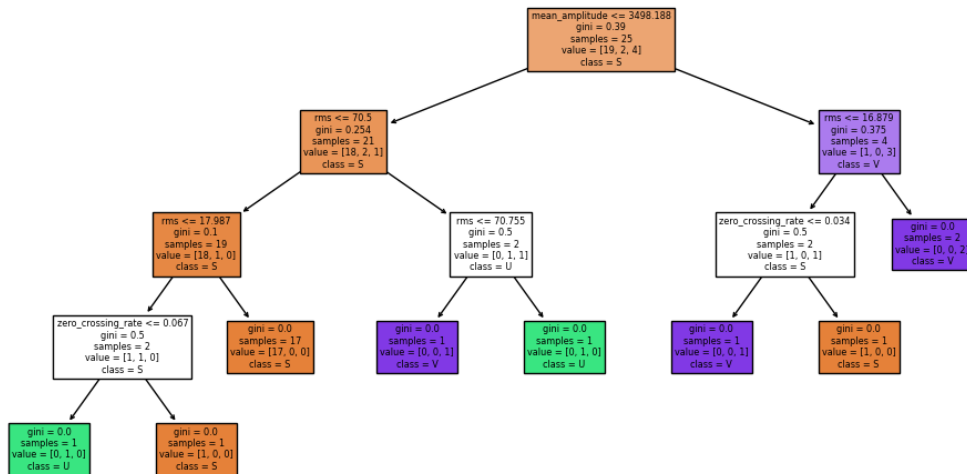


Figura 18: Arbol de Decision



## 5. Conclusiones

En esta práctica, se implementó un sistema de clasificación de segmentos de audio basado en un **árbol de decisión**. Para ello, se dividieron los archivos de audio en fragmentos de 100 milisegundos y se extrajeron características clave como la **amplitud media**, el **valor RMS** y la **tasa de cruces por cero**. Estas características permitieron analizar la estructura acústica de cada segmento y establecer patrones de clasificación.

Los resultados obtenidos indican que el modelo de clasificación basado en **árboles de decisión** es capaz de identificar correctamente los segmentos de audio según sus características. Se observó que:

- Los segmentos con mayor **amplitud media** y **RMS** tienden a pertenecer a categorías con mayor actividad sonora, como la clase **V** (voz).
- Los segmentos con baja amplitud y pocos cruces por cero fueron etiquetados predominantemente como **S** (silencio).
- Los segmentos intermedios en términos de energía y actividad espectral fueron clasificados como **U**, lo que sugiere una transición entre regiones de silencio y voz.

La precisión del modelo en los datos de prueba fue del **88 %**. Esto demuestra que el árbol de decisión es una herramienta efectiva para la clasificación de segmentos de audio en esta tarea específica. No obstante, existen algunas limitaciones:

- La clasificación depende en gran medida de la calidad de los datos y de las etiquetas asignadas manualmente.
- El modelo podría mejorarse con el uso de técnicas más avanzadas, como redes neuronales o modelos de aprendizaje profundo especializados en procesamiento de audio.
- Factores como el ruido de fondo o la variabilidad en las grabaciones pueden afectar la precisión del modelo.

En conclusión, esta práctica demostró la utilidad del **procesamiento de señales de audio** para la segmentación y clasificación de sonidos, así como la aplicabilidad de los árboles de decisión en tareas de reconocimiento acústico. En trabajos futuros, se podrían explorar métodos más sofisticados para mejorar la precisión del modelo y su capacidad de generalización.

# Bibliografía

- [1] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
- [2] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.
- [3] D. P. W. Ellis, Classification of Speech and Music using a Decision Tree Algorithm,” *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Seattle, WA, USA, 1998, pp. 985-988.
- [4] F. Pedregosa et al., ”Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [5] T. Ganchev, N. Fakotakis, and G. Kokkinakis, Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task,” *Proceedings of the IEEE International Conference on Speech and Signal Processing (ICSSP)*, 2005.
- [6] J. R. Quinlan, Induction of Decision Trees,” *Machine Learning*, vol. 1, no. 1, pp. 81-106, 1986.