Práctica: Clasificación de sonidos.

Reconocimiento de Voz

Escamilla Resendiz Aldo

11 de marzo de 2025

Índice general

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**, un algoritmo 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 implementación de la práctica, se emplearon las siguientes herramientas:

- **Python**: Lenguaje de programación principal para la manipulación y análisis de los datos de audio.
- **Librosa**: Biblioteca especializada en procesamiento de señales de audio, utilizada para cargar archivos, calcular espectrogramas y extraer características acústicas.
- **Matplotlib**: Biblioteca de visualización que permitió graficar la forma de onda, espectrogramas y segmentaciones temporales.
- **Soundfile**: Utilizada para la lectura y escritura de archivos de audio en formato WAV.

2.2 Metodología de desarrollo

La metodología implementada se dividió en varias etapas clave:

1. **Obtención de datos**: Se grabaron frases específicas en formato WAV, asegurando calidad de audio adecuada y frecuencia de muestreo constante.
2. **Procesamiento de datos**:
   - Carga del audio con `Librosa`.
   - Visualización de la forma de onda para identificar la estructura temporal.
   - Segmentación del audio en bloques de 100 ms para su posterior clasificación.
3. **Implementación de la metodología**:
   - Transformada rápida de Fourier (STFT) para obtener espectrogramas de banda ancha y estrecha.
   - Conversión a escala logarítmica para visualizar mejor la energía de las frecuencias.
   - Análisis de los patrones acústicos en los espectrogramas para correlacionar con eventos fonéticos específicos.

2.3 Código en Python

```python
# Escamilla Reséndiz Aldo - 2022630761
# Práctica Representación del habla e dominios del tiempo y frecuencia
# Reconocimiento de voz

import numpy as np
import matplotlib.pyplot as plt
import librosa
import librosa.display
import soundfile as sf

audio_path = 'escuela.wav'
y, sr = librosa.load(audio_path, sr=None)

plt.figure(figsize=(14, 5))
librosa.display.waveshow(y, sr=sr)
plt.title('Forma de onda (Sin divisiones)')
plt.xlabel('Tiempo (s)')
plt.ylabel('Amplitud')
plt.show()

segment_duration = 0.1
samples_per_segment = int(segment_duration * sr)

plt.figure(figsize=(14, 5))
librosa.display.waveshow(y, sr=sr)
plt.title('Forma de onda (Dividida en segmentos de 100 ms)')
plt.xlabel('Tiempo (s)')
plt.ylabel('Amplitud')

total_duration = len(y) / sr
for t in np.arange(0, total_duration, segment_duration):
    plt.axvline(x=t, color='r', linestyle='--', alpha=0.7)

plt.show()

D = librosa.amplitude_to_db(np.abs(librosa.stft(y)), ref=np.max)
plt.figure(figsize=(14, 6))
librosa.display.specshow(D, sr=sr, x_axis='time', y_axis='log')
plt.colorbar(format='%+2.0f dB')
plt.title('Espectrograma de banda ancha')
plt.show()

D_narrow = librosa.amplitude_to_db(np.abs(librosa.stft(y, n_fft=2048, hop_length=512)), ref=np.max
    )
plt.figure(figsize=(14, 6))
librosa.display.specshow(D_narrow, sr=sr, x_axis='time', y_axis='log')
plt.colorbar(format='%+2.0f dB')
plt.title('Espectrograma de banda estrecha')
plt.show()
```

3. Resultados gráficos de los audios

A continuación, se presentan las representaciones para cinco archivos de audio:

3.1 Audio 1 - La casa es grande.

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

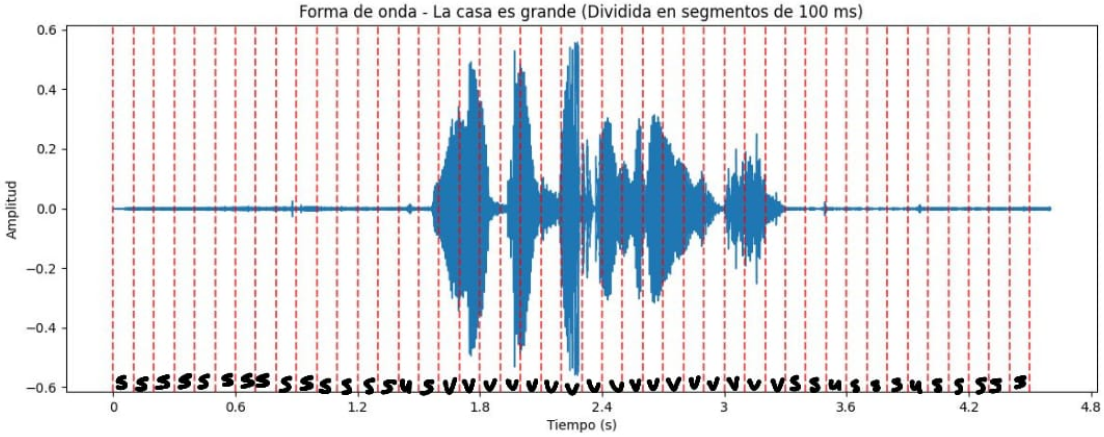

Figura 1: Forma de onda del audio: La casa es grande, con segmentación de 100 ms

- **Forma de onda de la señal de voz completa**

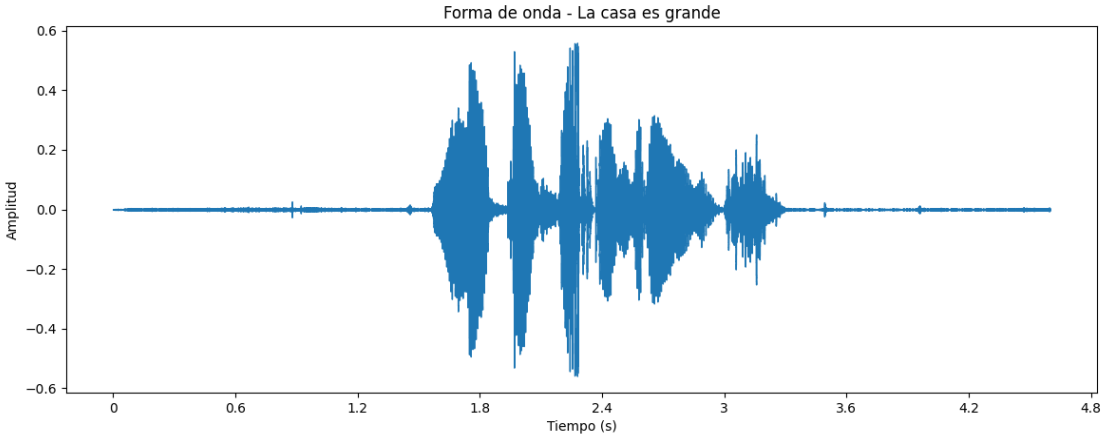

Figura 2: Forma de onda del audio: La casa es grande, con segmentación de 100 ms

- **Espectrograma de banda ancha**

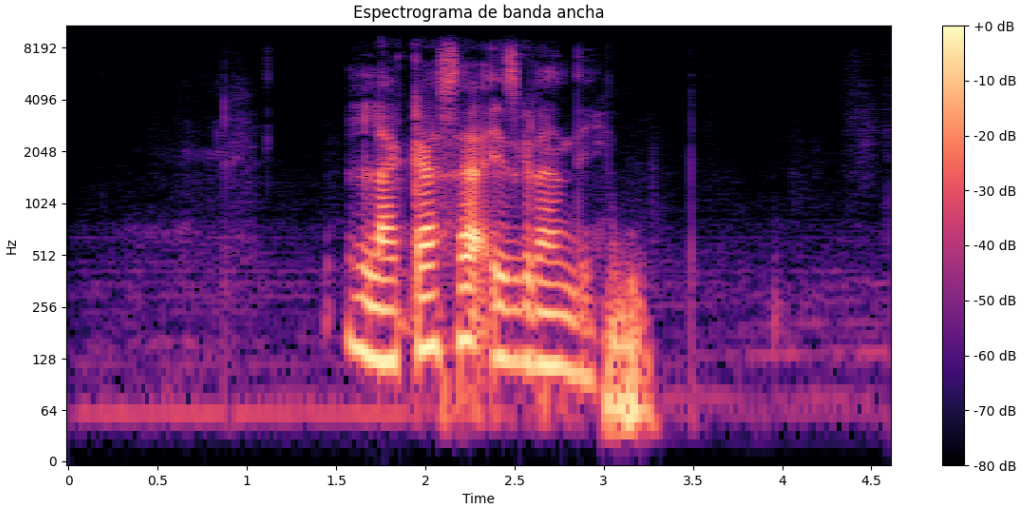

Figura 3: Espectrograma de banda ancha del audio: La casa es grande

- **Espectrograma de banda ancha**

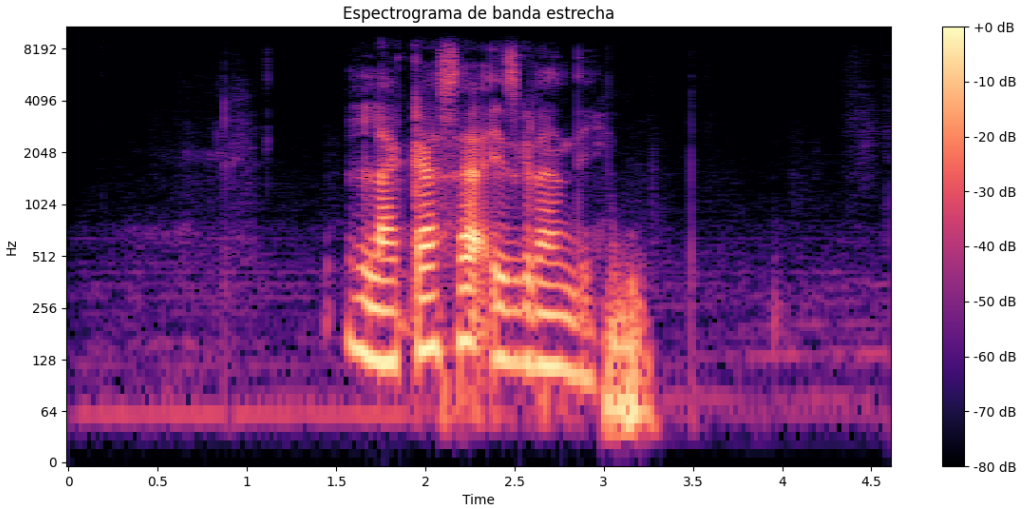

Figura 4: Espectrograma de banda estrecha del audio: La casa es grande

3.2 Audio 2 - Mi perro se salió de casa.

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

Figura 5: Forma de onda del audio: Mi perro se salió de casa

- **Forma de onda de la señal de voz completa**

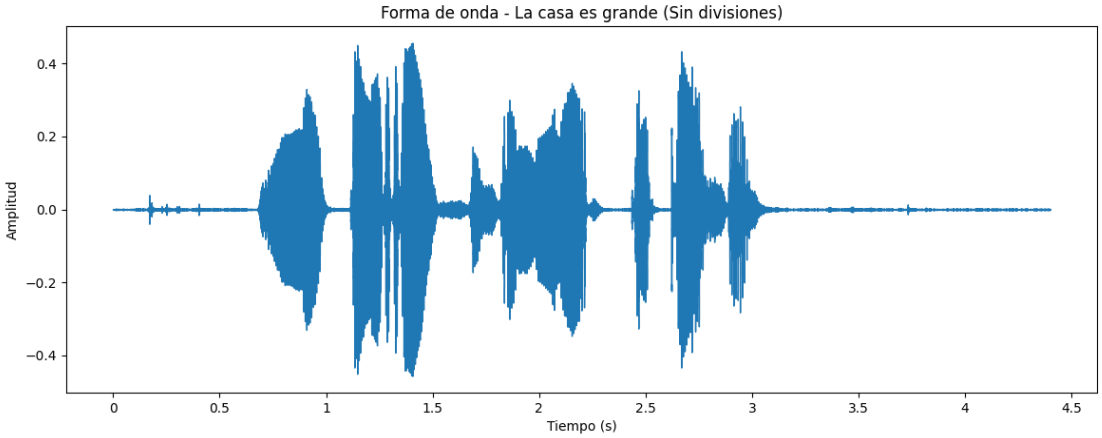

Figura 6: Forma de onda del audio: Mi perro se salió de casa, con segmentación de 100 ms

- **Espectrograma de banda ancha**

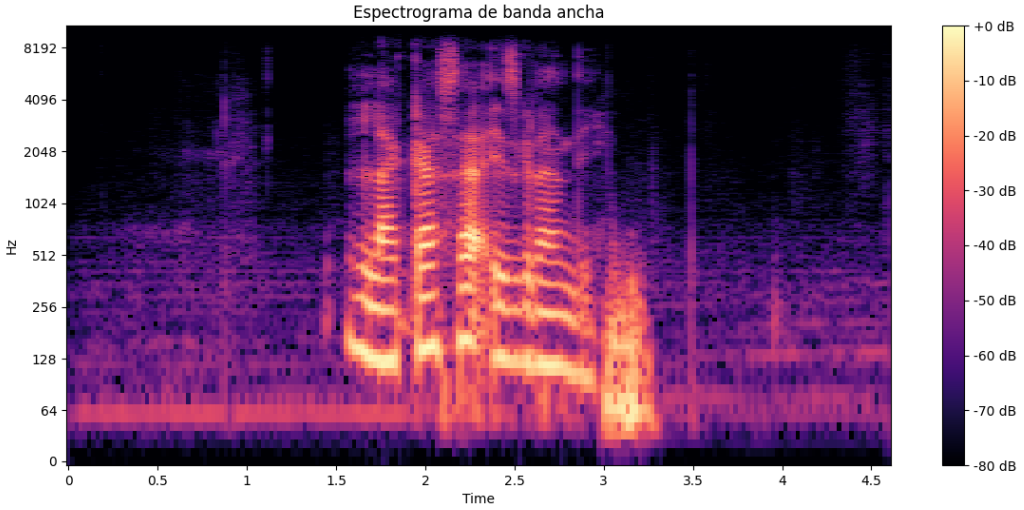

Figura 7: Espectrograma de banda ancha del audio: Mi perro se salió de casa

- **Espectrograma de banda estrecha**

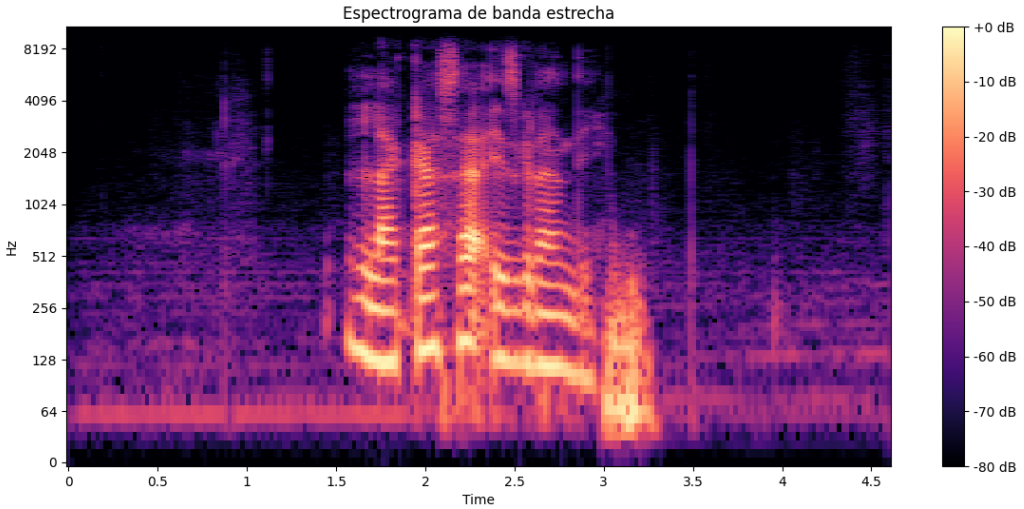

Figura 8: Espectrograma de banda estrecha del audio: Mi perro se salió de casa

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

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

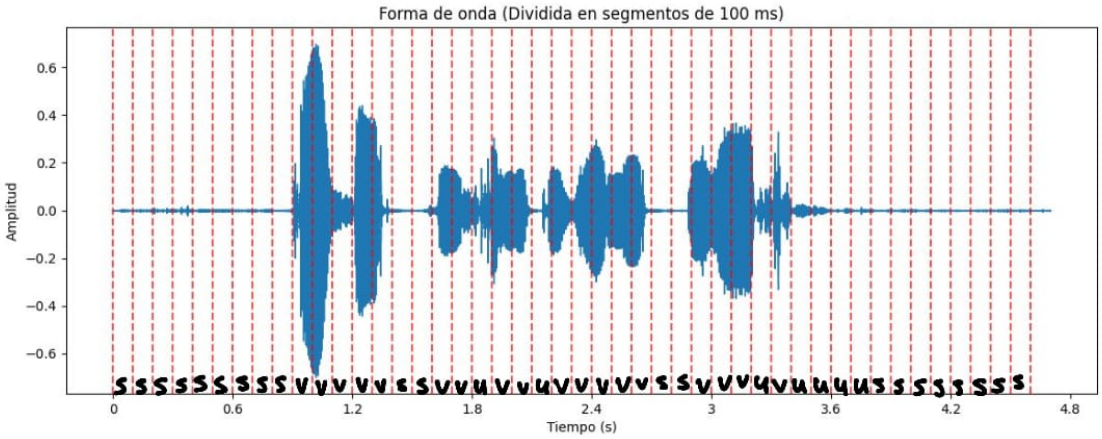

Figura 9: Forma de onda del audio: Casa es a lo que llamo hogar.

- **Forma de onda de la señal de voz completa**

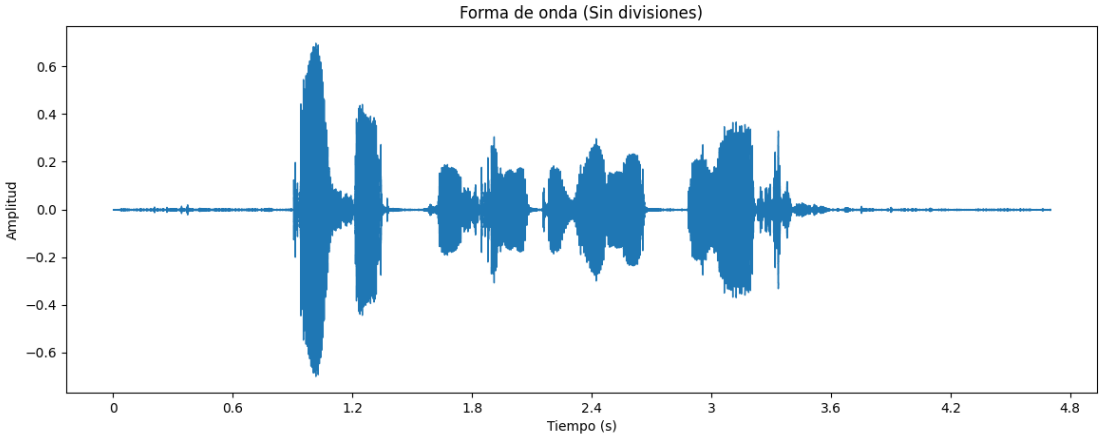

Figura 10: Forma de onda del audio: Casa es a lo que llamo hogar., con segmentación de 100 ms

- **Espectrograma de banda ancha**

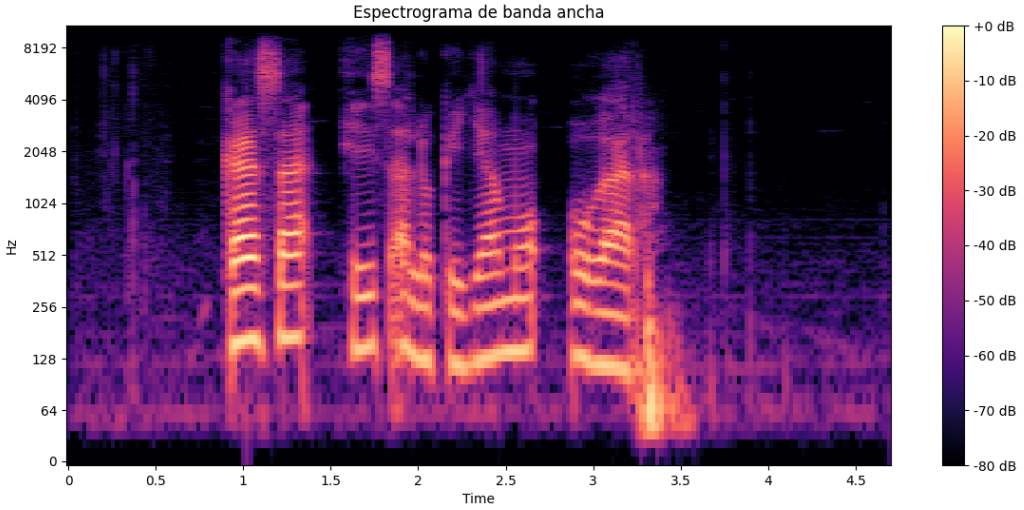

Figura 11: Espectrograma de banda ancha del audio: Casa es a lo que llamo hogar.

- **Espectrograma de banda estrecha**

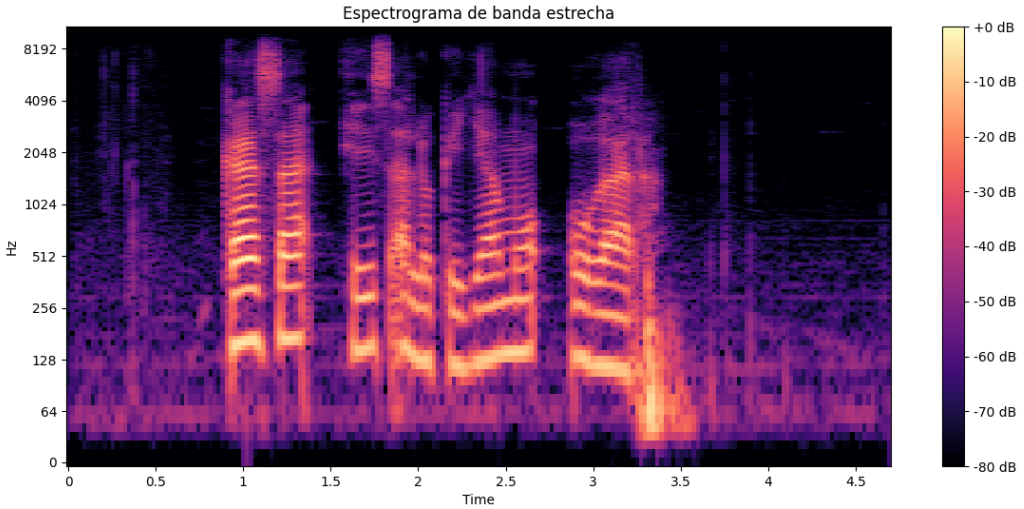

Figura 12: Espectrograma de banda estrecha del audio: Casa es a lo que llamo hogar.

3.4 Audio 4 - Los niños juegan en el parque.

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

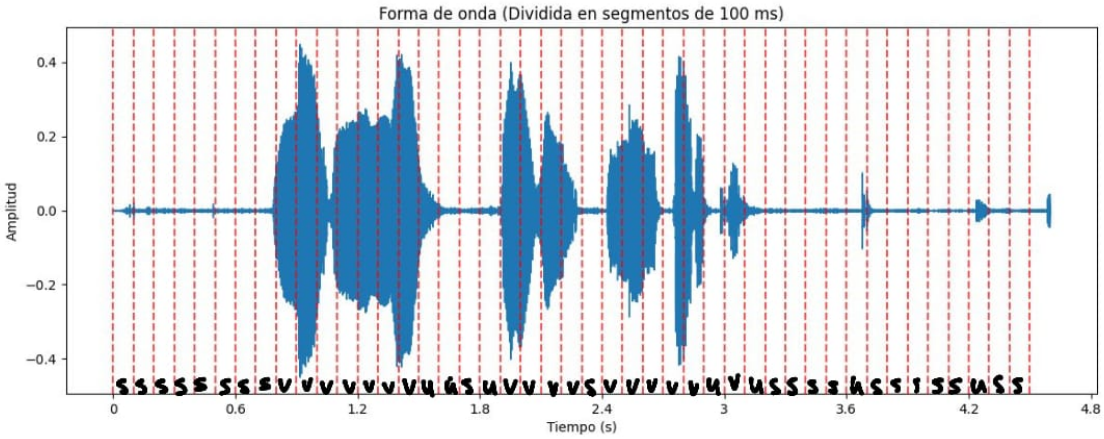

Figura 13: Forma de onda del audio: Los niños juegan en el parque.

- **Forma de onda de la señal de voz completa**

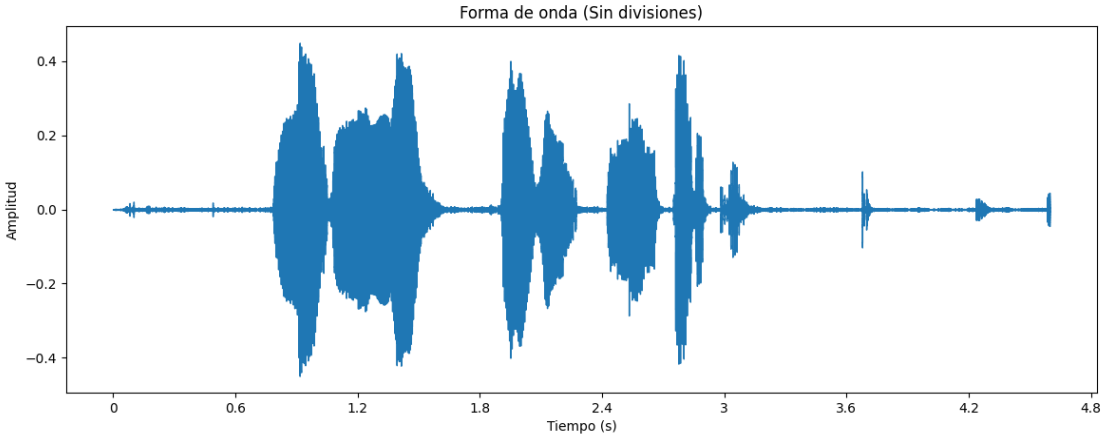

Figura 14: Forma de onda del audio: Los niños juegan en el parque, con segmentación de 100 ms

- **Espectrograma de banda ancha**

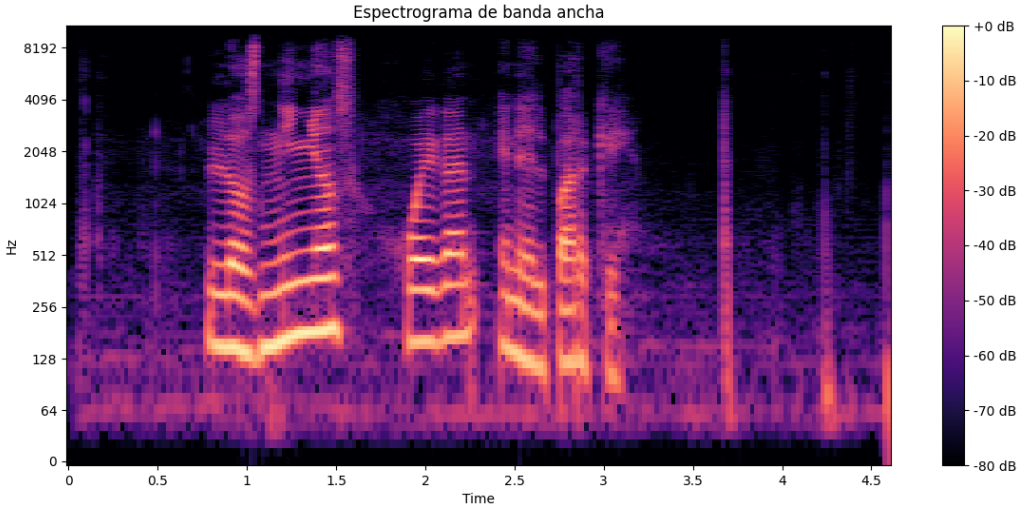

Figura 15: Espectrograma de banda ancha del audio: Los niños juegan en el parque.

- **Espectrograma de banda estrecha**

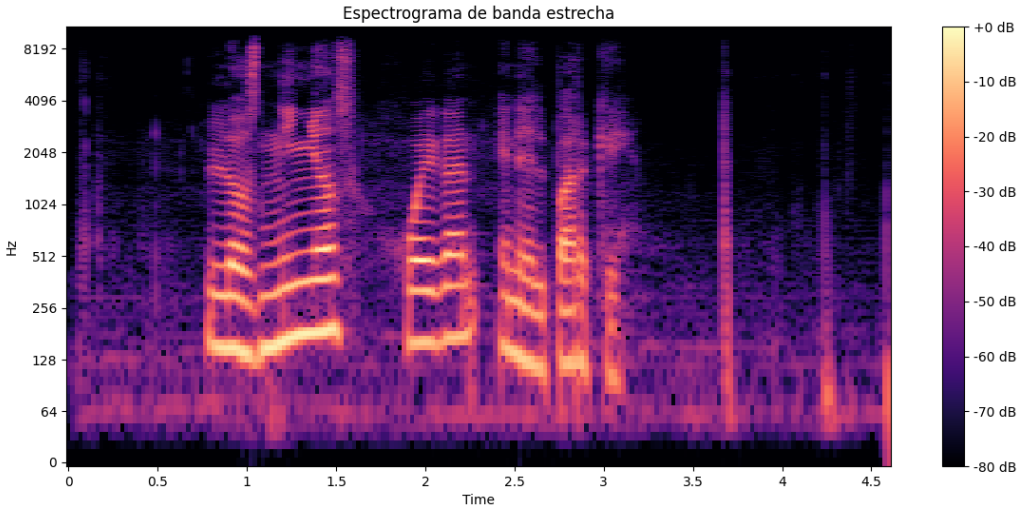

Figura 16: Espectrograma de banda estrecha del audio: Los niños juegan en el parque.

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)**

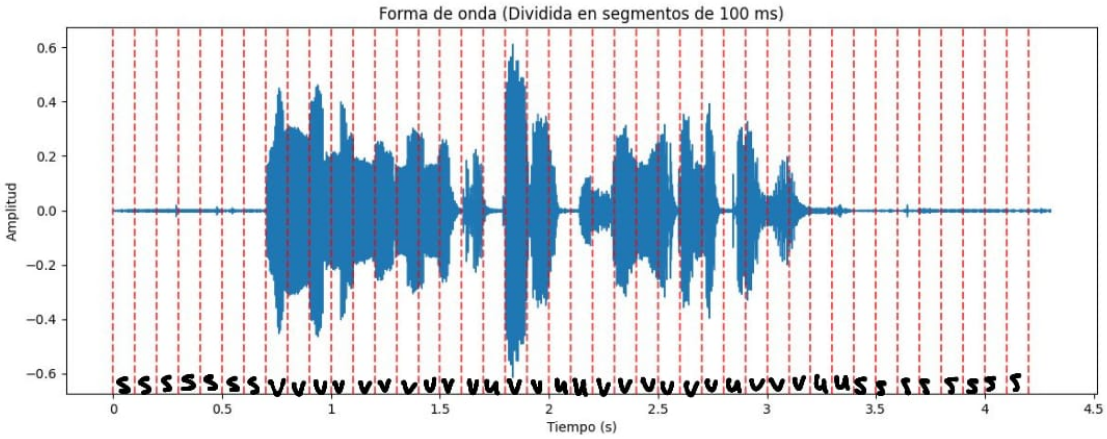

Figura 17: Forma de onda del audio: En la mañana me preparo y salgo a la escuela.

- **Forma de onda de la señal de voz completa**

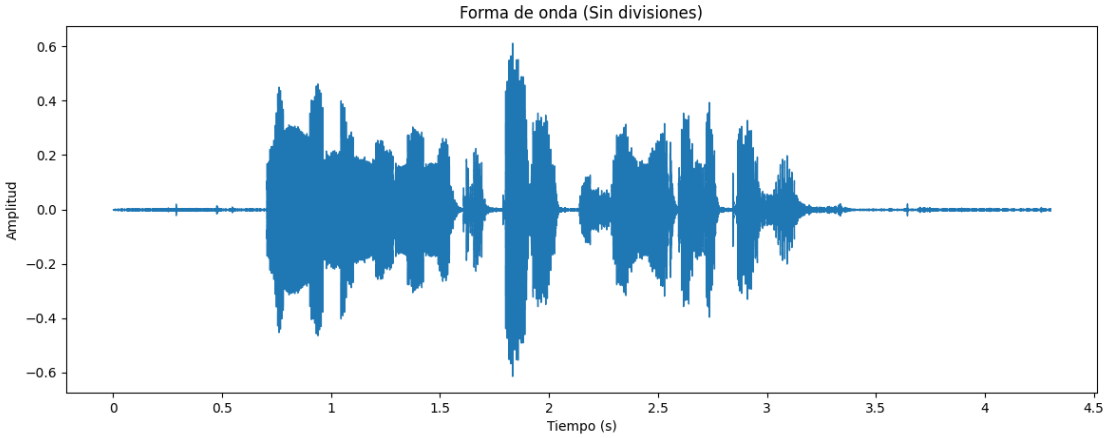

Figura 18: Forma de onda del audio: En la mañana me preparo y salgo a la escuela, con segmentación de 100 ms

- **Espectrograma de banda ancha**

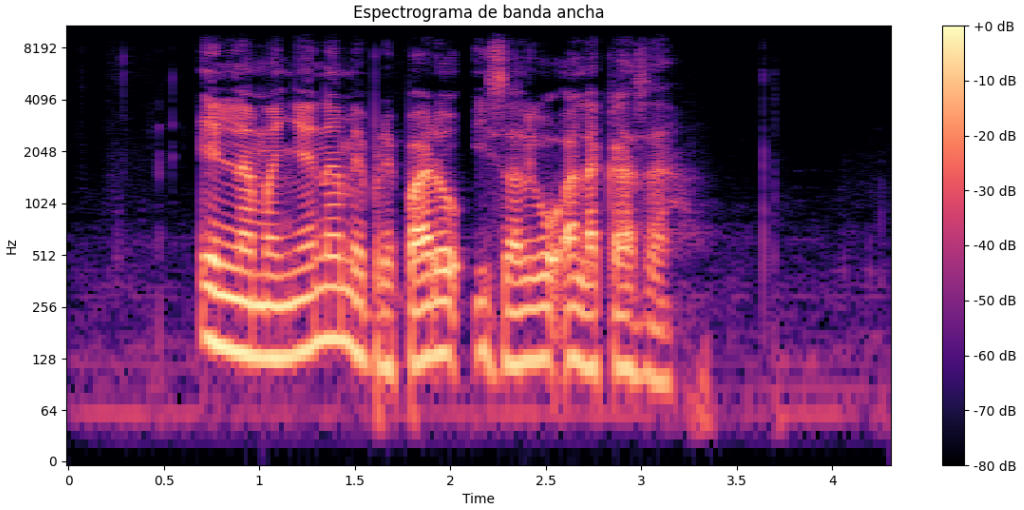

Figura 19: Espectrograma de banda ancha del audio: En la mañana me preparo y salgo a la escuela.

- **Espectrograma de banda estrecha**

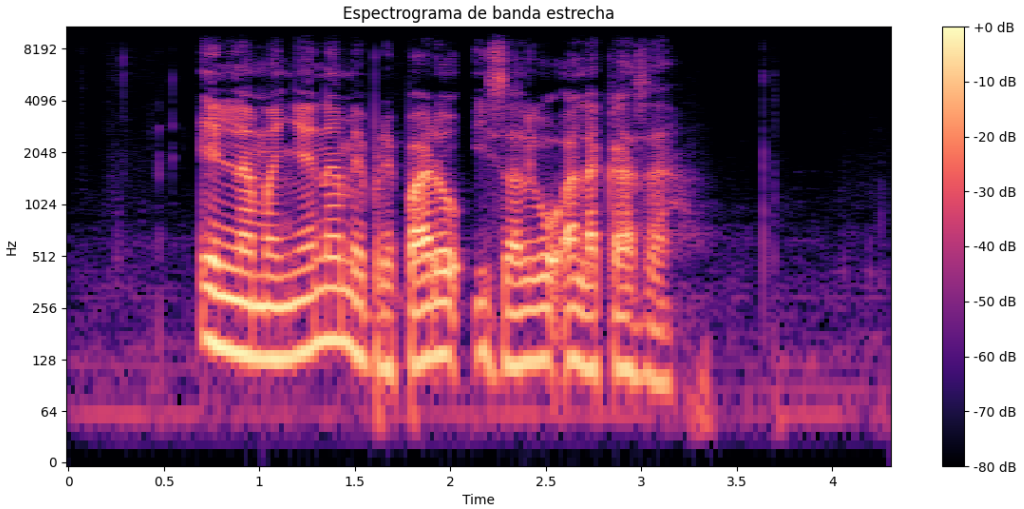

Figura 20: Espectrograma de banda estrecha del audio: En la mañana me preparo y salgo a la escuela.

4. Conclusiones

Los resultados obtenidos para los audios permiten observar cómo varía la energía en diferentes frecuencias a lo largo del tiempo. Esto es útil para identificar patrones de habla, como la aparición de formantes, que son concentraciones de energía en ciertas frecuencias y constituyen características fundamentales de los sonidos del habla.

Las áreas más oscuras en el espectrograma indican frecuencias con mayor energía, que podrían corresponder a sonidos vocálicos o consonánticos sonoros. Por otro lado, las áreas más claras representan frecuencias con menor energía, asociadas a pausas o sonidos fricativos.

Asimismo, en la forma de onda se observa la variación de amplitud a lo largo del tiempo, donde los picos corresponden a sonidos más fuertes (como las vocales) y los valles a sonidos más suaves o pausas.

Si bien las gráficas presentan similitudes para los cinco audios analizados, la comparación entre ellas permite apreciar que los espectrogramas de banda ancha y banda estrecha proporcionan información complementaria. Los espectrogramas de banda ancha resultan útiles para una visión general de la estructura del habla, mientras que los de banda estrecha permiten un análisis más detallado de las frecuencias y los armónicos.

Por otro lado, las formas de onda son esenciales para comprender la dinámica temporal del habla. Sin embargo, la ausencia de divisiones en los ejes puede dificultar una interpretación precisa de los eventos acústicos.

Gracias al desarrollo de esta práctica, se logró una mejor comprensión de las representaciones del habla en los dominios del tiempo y la frecuencia. Esto permitirá un análisis más profundo de las señales de voz en futuros proyectos de reconocimiento de voz, ya que el estudio de las formas de onda y los espectrogramas (tanto de banda ancha como de banda estrecha) proporciona una visión detallada de la estructura del habla en términos de amplitud, tiempo y frecuencia.

Bibliografía

[1] L. R. Rabiner and R. W. Schafer, *Theory and Applications of Digital Speech Processing*. Upper Saddle River, NJ, USA: Pearson, 2010.

[2] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2009.

[3] B. Gold, N. Morgan, and D. Ellis, *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*, 2nd ed. Hoboken, NJ, USA: Wiley, 2011.

[4] K. N. Stevens, *Acoustic Phonetics*. Cambridge, MA, USA: MIT Press, 1998.

[5] H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis of Speech," *Journal of the Acoustical Society of America*, vol. 87, no. 4, pp. 1738-1752, 1990, doi: 10.1121/1.399423.