



Proyecto C.

Reconocimiento de Voz

Sobre el proyecto

Colección

Personas que hablan comandos de una sola palabra para un total de 20 palabras diferentes en inglés:

“Yes”, “No”, “Up”, “Down”, “Left”, “Right”, “On”, “Off”, “Stop”, “Go”, “Zero”, “One”, “Two”, “Three”, “Four”, “Five”, “Six”, “Seven”, “Eight”, and “Nine”.

Para ayudar a distinguir palabras no reconocidas, hay diez palabras auxiliares, que la mayoría de los hablantes solo dijeron una vez.

“Bed”, “Bird”, “Cat”, “Dog”, “Happy”, “House”, “Marvin”, “Sheila”, “Tree”, and “Wow”.

// 105,829 Archivos de audio
Formato .wav
1 seg de grabación en promedio
6 audios de ruido


Organización

 down

 eight


 five


 follow


 forward


 four


 go


 0a2b400e_nohash_0.wav


 0a7c2a8d_nohash_0.wav


 0a196374_nohash_0.wav

 0ab3b47d_nohash_0.wav

 0ac15fe9_nohash_0.wav

 0ac15fe9_nohash_1.wav

 0ac15fe9_nohash_2.wav

 0b7ee1a0_nohash_0.wav

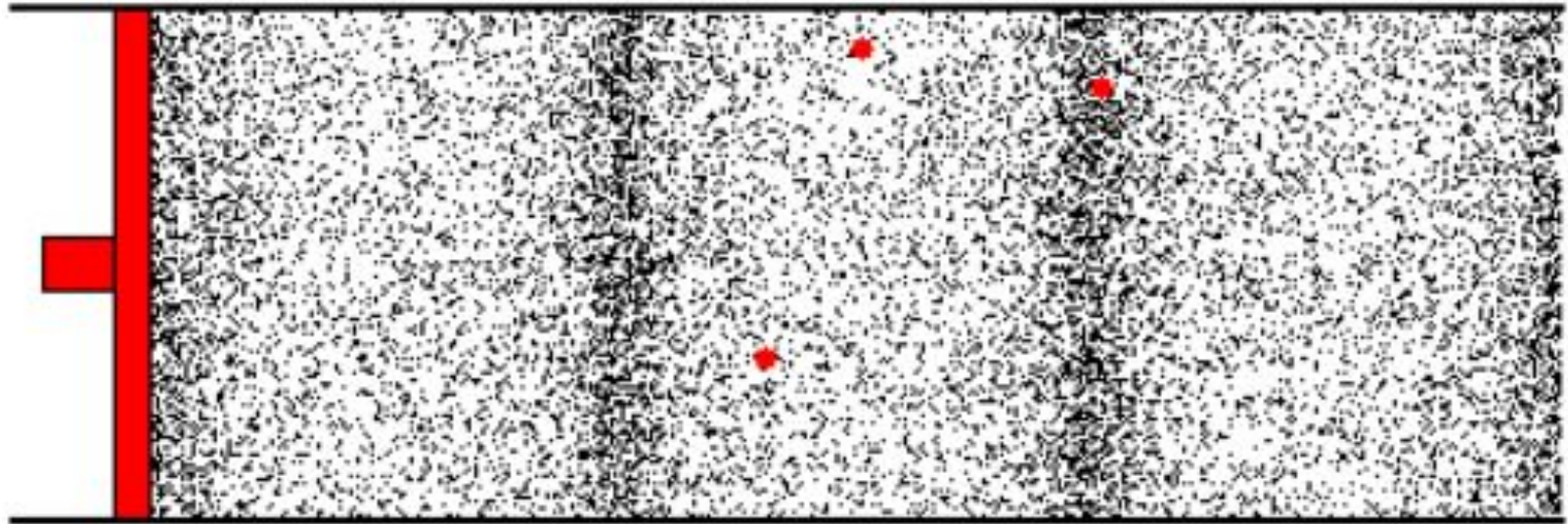
Tratamiento

- Grabado por personas de muchas partes del mundo.
- Único requisito fue lugar cerrado.
- No se exige calidad de audio.
- Originalmente se capturaron en formato Ogg Vorbis y luego WAVE.

Introducción al mundo del procesamiento de audio

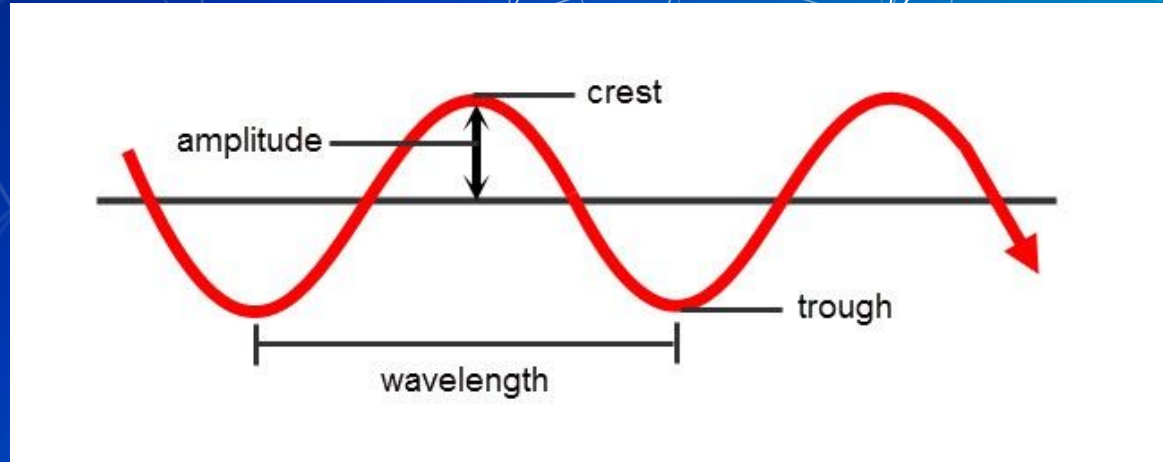
¿Qué es una señal de audio?

Longitudinal Wave



Parámetros de una señal de audio

Amplitud: Máximo desplazamiento de una molécula de aire desde su posición de reposo.

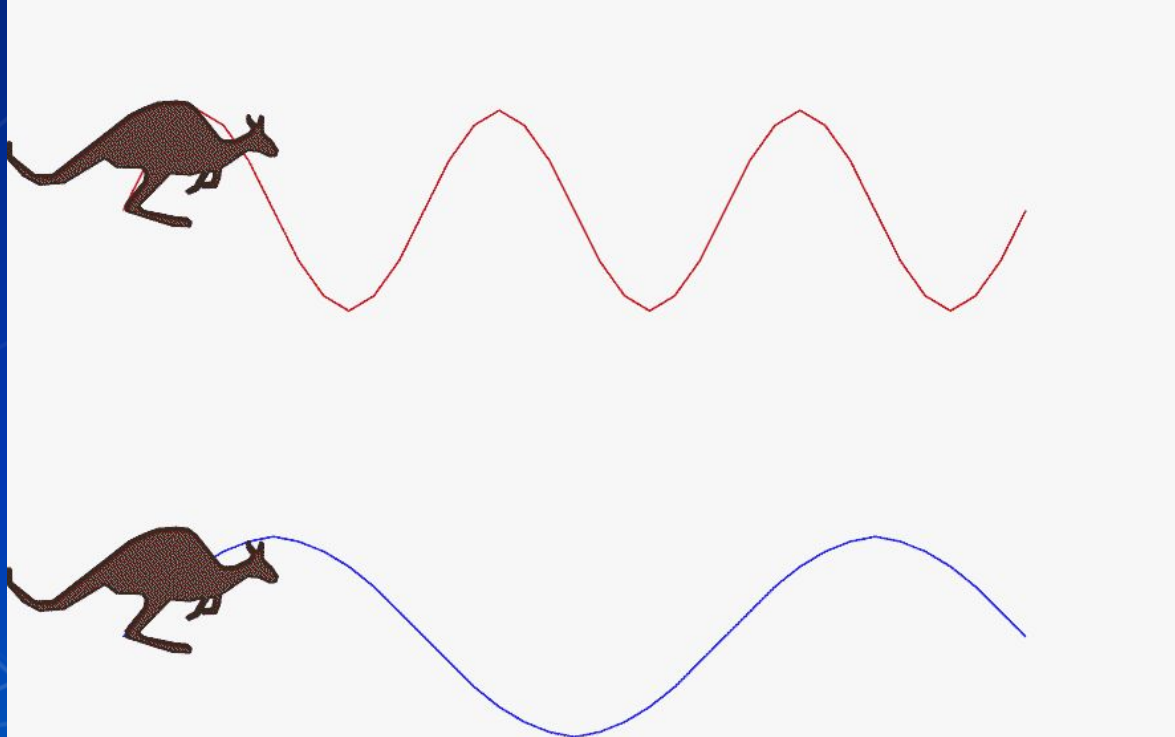


Longitud de la onda (M): la distancia entre 2 crestas sucesivas.

Parámetros de una señal de audio

Periodo(seg): Tiempo entre 2 puntos en un segundo.

Frecuencia(Hz): Se refiere a qué tan rápido está cambiando una señal durante un período de tiempo.

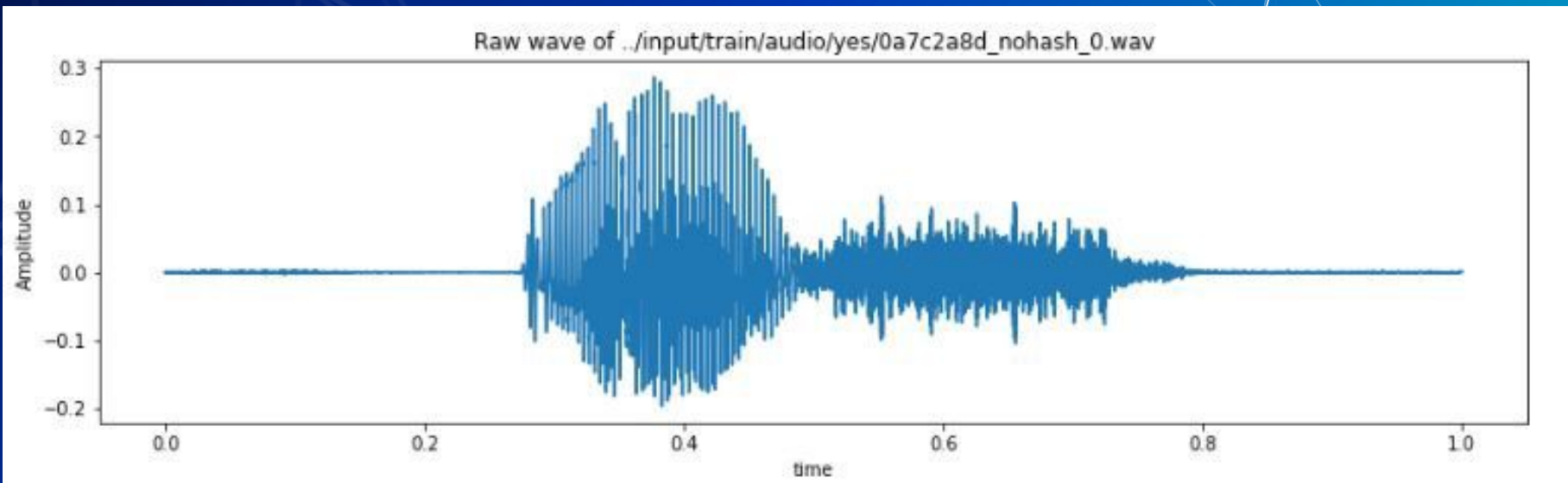




Representación gráfica de audios

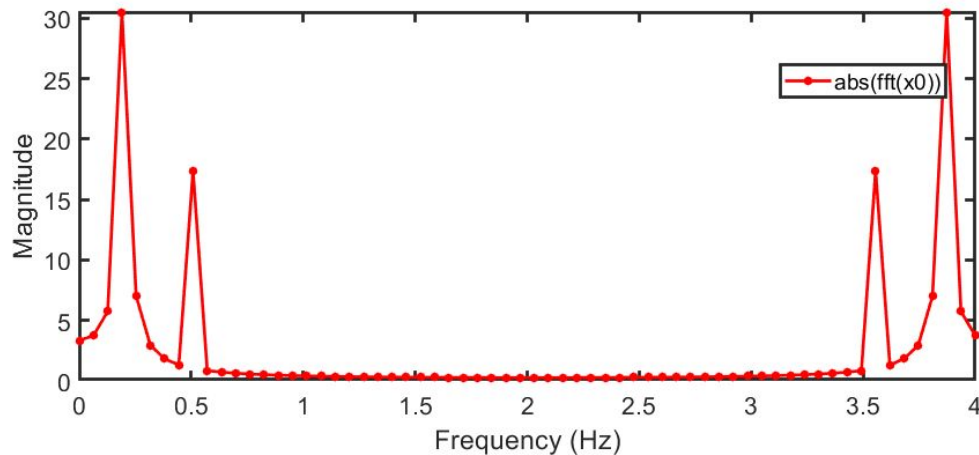
Visualización de señal de audio

La señal observada en el dominio de series temporales.
Tiempo vs Amplitud.



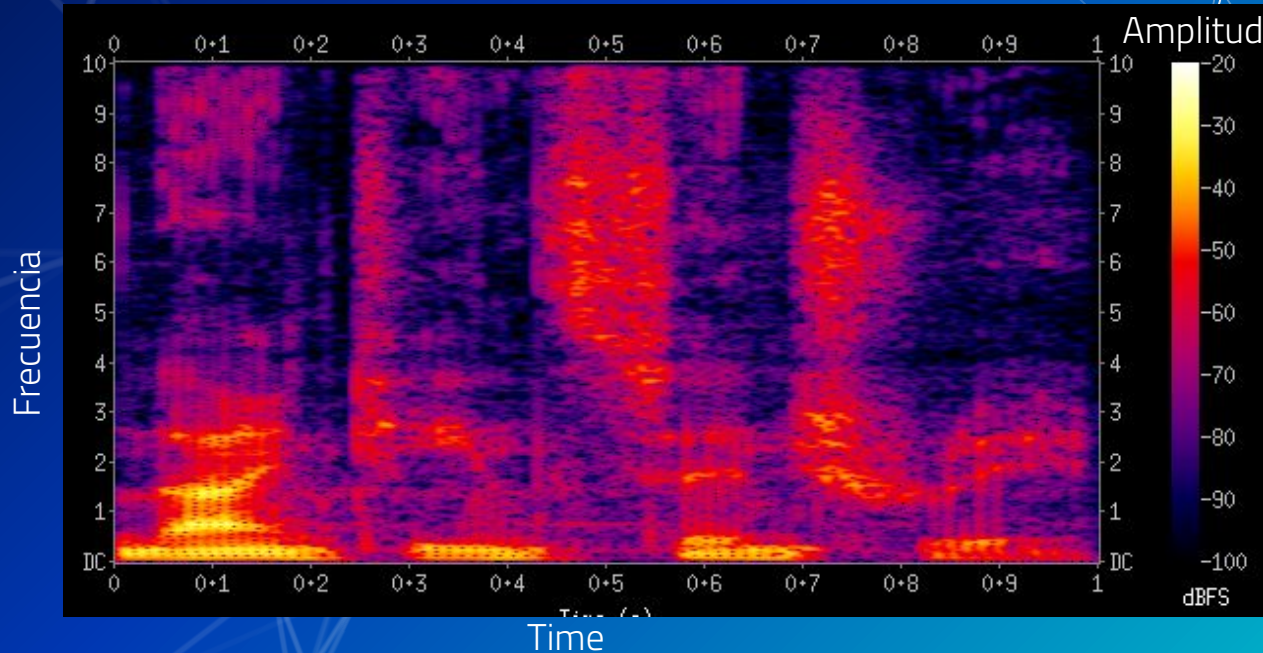
Fast Fourier transform FFT

Transformada matemática que nos permite convertir señales en su contenido frecuencial



Espectrograma

Representación visual 2D del sonido entre tiempo y frecuencia, donde cada punto en la gráfica representa la amplitud de una frecuencia en un tiempo en términos de intensidad de color.





Código

https://github.com/josearangos/Voice_Recognition