

Proyecto 2.

Reconocimiento de voz
Transformada de Fourier



Integrantes:

Maximiliano Uribe
Dante Valenzuela
Nicolas Jara
Fernando Guerra
Jose Luis Calderon
Boris Berrios

Asignatura: Matemáticas IV

Profesor: Eric Jeltsch

Índice

1	Introducción
2	Objetivos del Proyecto
3	Estructura del Proyecto
4	Este Proyecto
5	Metodo matemático
6	¿Cómo trabajaremos con python?
7	Conclusión
8	github
9	Bibliografía

1 Introducción

La voz humana es una de las formas más naturales, personales y complejas de comunicación. Cada individuo posee un timbre, tono e intensidad característicos, determinados por las frecuencias presentes en su emisión vocal. El análisis de estas propiedades permite estudiar y clasificar distintos tipos de voces, información que resulta fundamental en áreas como la música, la ingeniería acústica y el reconocimiento de patrones biométricos.

En este proyecto se propone utilizar Python y las transformadas de Fourier para analizar las señales de voz capturadas mediante un micrófono. A partir de dicho análisis, se busca identificar las características propias de cada voz, tales como la frecuencia fundamental, los armónicos y la energía de la señal.

El objetivo final es aplicar esta información al desarrollo de un sistema de reconocimiento de voz capaz de verificar la identidad de un usuario y permitir el acceso a una página web o archivo privado. De esta manera, se combina el análisis digital de señales con la seguridad informática, explorando el potencial de la voz como una herramienta de autenticación biométrica.

2 Objetivos del Proyecto

- 1 El objetivo principal de este proyecto es analizar la voz de una persona utilizando Python y las transformadas de Fourier, con el propósito de desarrollar un sistema de reconocimiento de voz que permita proteger el acceso a un objeto privado o archivo.
- 1 A través de este análisis, se busca identificar y diferenciar distintos tipos de voz, comparando las características de cada registro con un patrón original. De esta manera, el sistema podrá validar únicamente la voz del usuario autorizado, garantizando la correcta autenticación mediante voz.
- 2 Finalmente, se plantea como objetivo complementario que, manteniendo la funcionalidad del reconocimiento por voz, el sistema sea capaz de abrir una página web con acceso controlado mediante autenticación vocal.
- 3
- 5

3 Estructura del Proyecto

Procesos

1. Conexión del micrófono.
2. Obtención y recepción de los datos hacia Python.
3. Aplicación de la Transformada de Fourier a los datos de voz.
4. Análisis de las frecuencias para identificar tono, timbre y energía.
5. Comparación de audios y reconocimiento de voz.

Métodos

1. **Investigación:** recopilar información sobre el funcionamiento del micrófono y la Transformada de Fourier.
2. **Procesamiento digital:** utilizar Python con bibliotecas como `numpy` y `matplotlib` para realizar la FFT y visualizar resultados.
3. **Simulación y prueba:** probar distintas voces para verificar la respuesta del sistema.

4. **Análisis comparativo:** comparar los espectros de los dos audios para hacer un análisis y su respectiva verificación.

4 Este Proyecto

Para poder hacer realidad el objetivo de este proyecto primero debemos conocer algunos conceptos como:

4.1 ¿Que es una señal de voz?

Es una representación acústica de los sonidos producidos por la voz humana, que presentan variaciones de presión que se propagan a través del aire como ondas sonoras. Puede capturarse y convertirse en una señal eléctrica a través de un micrófono para procesamiento, transmisión y reproducción. En forma mas visual se puede ver como este mapa:

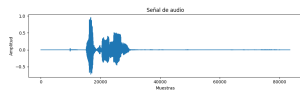


Figure 1: Caption

4.2 ¿Qué características son importantes de ella?

El timbre: es la cualidad que nos permitirá diferenciar una voz de otra, incluso si tienen el mismo tono y volumen. Está relacionada con la complejidad de la onda sonora y la resonancia.

El ritmo: Es la cadencia y el patrón temporal del habla, incluyendo pausas y la duración de las sílabas.

El volumen: Se relaciona con la energía de la señal y puede variar según lo que intente transmitir.

La duración: longitud del sonido al igual que la velocidad del habla

La entonación: Es la variación del tono de la voz a lo largo del tiempo y es crucial para la expresividad y la emoción.

4.2.1 Características temporales:

Energía de la señal: indica la potencia de la señal en un momento dado.

Tasa de cruce por cero: el número de veces que la forma de onda de la señal cruza el eje horizontal en un intervalo de tiempo.

4.2.2 Características espectrales:

Frecuencia fundamental: Determina la altura del tono y es el resultado de la frecuencia de vibración de las cuerdas vocales.

Espectro: la distribución de la energía de la señal a través de las diferentes frecuencias que la componen.

4.3 ¿En que se inspiró para crear este proyecto?

Este proyecto se inspiró de un juguete antiguo de un diario secreto de la marca MATTEL en el cual su característica más llamativa es que se le puede asignar un comando de voz como contraseña que tiene como objetivo abrir dicho diario con su contenido adentro (siendo más una caja que se puede abrir con voz que un diario en sí)



Figure 2: Caption

4.4 ¿Cómo es que funciona en comparación a nuestro proyecto?

El funcionamiento en cuestion tiene que ver con la comparacion de grabaciones de audio, de forma mas explicada, cuando se pida grabar el audio por primera vez, se entra este registrado en la memoria del pedirá nuevamente la contraseña por voz, una vez que esta se escuche, se comprara con la grabación original para

ver si coinciden, si llegasen a coincidir, rogramar, entonces desde ahí cuando se quiera volver a ingresar, se desbloqueara lo que sea que haya guardado (en este caso, nuestro sitio web) en caso de no coincidir, pues simplemente rechazara la contraseña por voz que se haya dado en ese momento

5 Metodo matemático

5.1 ¿Qué fórmulas ocuparemos y en que?

En este proyecto deseamos implementar en específico una sola fórmula con respecto a Fourier la cual es la Transformada rápida de Fourier (FFT), la cual en pocas palabras nos ayudará a encontrar las frecuencias y eliminar el ruido externo al que queremos.

5.2 ¿Qué es la transformada de fourier?

Es una herramienta matemática que descompone una señal o función en una suma de funciones sinusoidales (oscilación suave y periódica) de diferentes frecuencias, lo que permite ver la señal desde el dominio de la frecuencia en lugar del dominio del tiempo

$$F(w) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt$$

5.3 ¿Qué muestran sus resultados?

La Transformada de Fourier de una señal muestra sus componentes de frecuencia, descomponiendo la señal original en una suma de ondas senoidales. En el espectro resultante, el eje X representa la frecuencia y el eje Y la amplitud de cada frecuencia. De esta forma, permite ver qué frecuencias están presentes y con qué intensidad, transformando la señal del dominio del tiempo al dominio de la frecuencia.

Información oculta en el tiempo: en lugar de ver cómo cambia una señal en el tiempo, se ve la composición frecuencial, lo que permite un análisis más detallado.

Análisis de ruido y vibraciones: permite la identificación de ruido y vibraciones no deseados en una señal, facilitando su eliminación.

6 ¿Cómo trabajaremos con python?

La forma en que ocuparemos python sera de tres maneras siendo la 1ra parte una version prueba donde contamos con las opciones de **grabar audio**, **mejorar audio** (quitarle el ruido externo y mejorar la frecuencia de la voz), y por ultimo **comparar voces**.

La 2da manera sera ya intentando superar nuestro objetivo de encriptacion y acceso por reconocimiento de voz, siendo esta parte donde ya probamos **encriptar un archivo basico** (txt, word, otro) y que pida el reconocimiento de una voz ya determinada para el archivo siendo esta la misma abriendo el acceso.

La 3ra y ultima forma es en la **creación de una web** propia que solo se pueda acceder por medio de una voz encriptada u opcional una contraseña ya guardada, en esta web se podrá expandir a gusto de cada uno que quiera hacer su propia web y no quiere que cualquiera lo pueda ejecutar o así mismo para poder guardar mas cosas en ella.

6.1 ¿Cómo funciona la encriptación del archivo?

Cuando un usuario graba su voz, el archivo WAV se lee como datos binarios. Luego se calcula un hash SHA256 de esos datos. Ese hash se convierte en una clave de cifrado válida para Fernet, que es el método que se usa para cifrar el texto. La clave no es fija, depende del contenido exacto de la voz del usuario.

6.1.1 —Cifrar las notas—

Cada vez que el usuario escribe una nota, el texto se cifra usando esa clave derivada de su voz. El resultado es un archivo binario que no se puede leer directamente sin la clave correcta.

6.1.2 —Descifrar las notas—

Para leer la nota, el sistema vuelve a generar la clave a partir de la voz original del usuario. Si la voz coincide, la clave será la misma y el texto se puede descifrar. Si alguien intenta abrir el archivo sin la voz correcta, la clave será distinta y el contenido no podrá leerse.

6.1.3 —Seguridad básica—

Cada usuario tiene su propio archivo de notas cifradas. La seguridad se basa en que solo la voz original genera la clave correcta. El texto nunca se guarda en claro en el sistema; siempre está cifrado hasta que se autentica la voz.

En pocas palabras: la voz actúa como contraseña, y el texto sólo puede leerse si la voz coincide con la que se usó al cifrar.

6.2 ¿Cómo creamos la web?

La página web fue desarrollada utilizando una combinación de tres lenguajes de programación: **Python**, **HTML** y **JavaScript**.

El lenguaje **JavaScript** (JS) se empleó principalmente para la captura del audio desde el navegador, mientras que **HTML** se utilizó para estructurar la interfaz visual. Por su parte, **Python** actúa como el motor principal del sistema, ejecutando el programa, gestionando las solicitudes del servidor y manteniendo un registro de actividad en la terminal mientras la web está en uso.

Para el desarrollo se utilizaron diversas librerías de Python, especialmente orientadas al manejo, análisis y visualización de audio, además de facilitar la comunicación entre los distintos lenguajes. Entre las más destacadas se encuentran: NumPy, Flask, Librosa, SoundFile, Matplotlib y SciPy, entre otras.

6.2.1 ¿Cómo funciona?

La web cuenta con una página inicial que permite acceder mediante voz o contraseña, ambas dirigiendo al menú principal, el cual ofrece varias funciones: **Análisis:** muestra las gráficas correspondientes a la voz registrada.

Historial: presenta las veces que se ha accedido a la aplicación y el modo de acceso.

archivos protegidos: muestra los archivos que estén guardados dentro de la estructura de la web.

Configuración: permite cerrar sesión o cambiar la contraseña de acceso.

7 Conclusión

7.1 ¿Para qué nos sirve este proyecto en la vida cotidiana?

Esto nos puede servir para resguardar de mejor manera nuestros datos, e incluso puede servir como una clase de desbloqueo futuro para dispositivos diversos, solo imaginen que se pueda desbloquear con la voz tu celular o cualquier aparato y no solo verlo de modo llamativo sino que puede servir en momentos cruciales.

Se han llegado a crear sistemas automatizados por voz que permita controlar diversas funciones de hogares más subdesarrollados (Poniendo de ejemplo, varias funciones de asistentes virtuales como Alexa)

7.2 ¿Para qué nos sirve en el ámbito matemático?

Nos sirve para entender como se descomponen usando fourier algo tan comun para todos como es la voz y así mismo podriamos llevarlo a algo mas personal algun ruido que nos identifique y así mismo nos ayuda a poder conectar la matematica con lo que seria criptografia, programacion y el proceso digital de las señales.

en otras palabras, nuestro proyecto demuestra que la teoria matemática puede obtener soluciones tecno-

logicas reales dentro de nuestras vidas y no tan solo ver sus resultados dentro de una sala de estudio.

8 github

<https://github.com/maximiliano-uribe/praxis2-Reconocimiento-de-voz.git>

9 Bibliografia

- https://scholar.google.com/scholar?hl=es&as_sdt=0%2C5&q=A+software+tool+for+the+analysis+of+vocal+performances+m+howard&btnG=#d=gs_qabs&t=1761673852905&u=%23p%3DKvzcC4sV3fAJ
- <https://www.espaicoriveu.com/diferentes-tipos-de-voz-y-como-distinguirlos>
- <https://rua.ua.es/server/api/core/bitstreams/a631d1ee-5df5-4896-baee-da9ca6fd3e41/content>
- ChatGPT. (2025, noviembre 08-13). Respuesta a consulta sobre creación y corrección del código web. <https://chat.openai.com/>
- Manual de servicio de juguete MATTEL (referencia del proyecto).https://service.mattel.com/instruction_sheets/V4360.pdf