

---

*Actividad Académicamente Dirigida:*  
*Text to Speech*

---

PROCESAMIENTO DEL HABLA, VISIÓN E INTERACCIÓN MULTIMODAL

CURSO 2023 – 2024



ÁLVARO DOMÍNGUEZ MORA

RAÚL JIMÉNEZ SUÁREZ

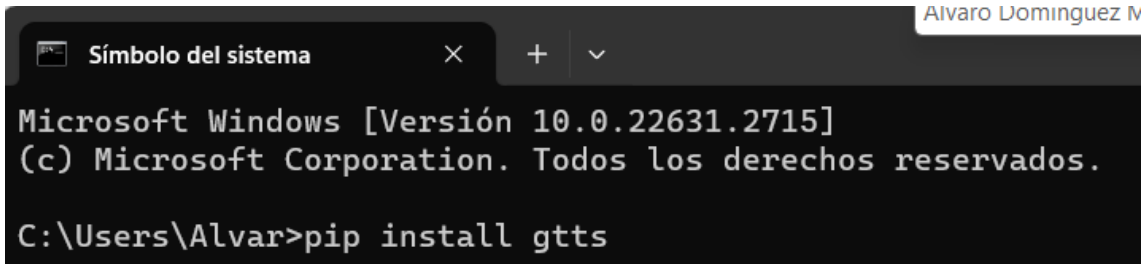
ÁNGEL BARROSO ROMERO

## Librería de Google gTTS

En primer lugar, vamos a comentar cuales son las librerías que hemos usado para la ejecución de nuestro programa. La librería de Google Text-to-Speech (gTTS) es una herramienta que nos permite convertir texto en voz mediante la tecnología de síntesis del habla. Esta librería, que ha sido desarrollada por Google, facilita la creación de aplicaciones y scripts que nos permite generar archivos de audio a partir de texto en varios idiomas.

En cuanto a su uso, vamos a explicar como se debe de realizar su instalación para la incorporación a nuestro proyecto.

Podemos hacerlo a través de la terminal de comandos, donde introduciremos: *pip install gtts*

A screenshot of a Windows command prompt window. The title bar at the top says 'Símbolo del sistema' and has window control buttons. The window content shows the Windows version '10.0.22631.2715' and the copyright notice '(c) Microsoft Corporation. Todos los derechos reservados.' The command prompt shows the user's location as 'C:\Users\Alvar' and the command being executed is 'pip install gtts'.

```
Símbolo del sistema
Microsoft Windows [Versión 10.0.22631.2715]
(c) Microsoft Corporation. Todos los derechos reservados.
C:\Users\Alvar>pip install gtts
```

Una vez instalada nuestra librería, procedemos a instalarla en nuestro script de Python, donde podremos crear ya los objetos gTTS, incorporándoles el texto que queremos convertir en voz.

Además, también podremos salvar nuestro archivo audio generado en formato MP3, u otro archivo compatible y la reproducción de este directamente desde la ejecución de nuestro script, para ello debemos importar la librería os.

Como anotaciones, la librería gTTS hace uso de servicios en la nube de Google para realizar la síntesis de voz, por lo tanto, para la ejecución de cualquier programa que haga uso de esta, deberá contar con una conexión a internet, ya que este texto se envía a los servidores de Google, donde se realiza la conversión a audio.

En la explicación de funcionamiento de nuestro script, veremos de forma detallada, como realizamos el uso de esta librería.

La tecnología de síntesis de voz de Google, también conocida como Google Text-to-Speech (TTS), es una tecnología avanzada que convierte el texto en voz. Así funciona:

1. **Redes neuronales:** Google TTS utiliza redes neuronales para generar voz. Estas redes son entrenadas con grandes cantidades de datos de voz anotados, lo que les permite aprender a generar voz que suena casi humana.
2. **Espectrograma:** El sistema de Google se basa en una primera red neuronal que traduce el texto a un espectrograma. Un espectrograma es una representación visual de las frecuencias de audio a lo largo del tiempo.
3. **Entonación:** Google TTS puede hablar con diferentes entonaciones. Puedes elegir entre tres entonaciones distintas: expresiva, algo expresiva y plana. La entonación plana es la entonación robótica típica de los sintetizadores de voz de hace décadas.
4. **Volumen:** Google TTS también te permite ajustar el volumen de la voz. Esto es útil si quieres que la voz se superponga a otros sonidos que puedan estar sonando en tu dispositivo<sup>2</sup>.
5. **Ajustes:** Para cambiar la entonación y el volumen en Google TTS, debes ir a sus ajustes. Normalmente, estos se encuentran en el apartado de Accesibilidad - Visión de tu dispositivo.

Es importante mencionar que la tecnología de síntesis de voz de Google está en constante evolución y mejora, gracias a la investigación y desarrollo continuos en el campo de la inteligencia artificial.

## **Librería Mixer**

Por otra parte tenemos la librería mixer que es parte del conjunto de herramientas proporcionadas por la biblioteca Pygame, diseñada entre otras cosas para el desarrollo de aplicaciones multimedia en Python, *pygame.mixer* se utiliza específicamente para manejar la reproducción de sonidos y música.

Para su instalación, podemos hacer como hemos realizado antes, en nuestra venta de comandos, introducimos: *pip install pygame*.

A continuación, nos vamos a nuestro proyecto y realizamos la importación de esta librería.

**From pygame import mixer**

Una vez importada, ya podemos hacer uso de esta para cargar archivos de sonidos, y también archivos de música, a través del método *play()*, podremos probar su reproducción, también tenemos la opción de pausar ( *pause()* ) y parar ( *stop()* ). Además, debemos también ajustar el volumen, al que queremos que se escuche nuestro audio.

Todas estas herramientas que nos permite usar la librería, vienen implementadas en nuestro código.

## **Funcionamiento de la práctica**

En primera instancia, definimos la función *textToSpeechAndWait* que será la que usaremos para la realización de la práctica. A continuación, pasamos el texto a Google Text to Speech que nos entregara un archivo mp3, creamos un numero aleatorio para el archivo mp3 que vamos a crear y generamos el nombre del archivo mp3 con dicho numero generado para guardarlo posteriormente.

El siguiente paso sería iniciar el mixer de sonido de la librería pygame, donde cargamos el archivo mp3 que tenemos guardado en nuestro equipo. Adaptamos el volumen a la cantidad que guardamos en la variable de configuración volumen.

Por último, iniciamos la reproducción del mp3 en el mixer, mientras el archivo se reproduce, paramos la reproducción de otros sonidos. Una vez termina la reproducción, quitamos el mixer y eliminamos el archivo mp3 en reproducción.

## Uso de la librería gtts

La librería gTTS (Google Text-to-Speech) es una biblioteca de Python que permite convertir texto en voz utilizando la tecnología de síntesis de voz de Google. En resumen, permite generar archivos de audio a partir de textos en varios idiomas y voces.

```
from gtts import gTTS

# Texto que deseas convertir a voz
texto = "Hola, ¿cómo estás?"

# Crear un objeto gTTS con el texto y el idioma
tts = gTTS(text=texto, lang='es')

# Guardar el archivo de audio
tts.save("saludo.mp3")
```

Como podemos observar, en este ejemplo, se convierte el texto “Hola, ¿cómo estás?” en un archivo de audio llamado “saludo.mp3” en español. Una de las características mas importantes que tiene es que podemos elegir el lenguaje mediante la variable “*lang*”. En nuestro caso, queremos hacerlo en español, así que le ponemos como valor ‘es’. Otro parámetro sería “*slow*”, el cual

nos permite que, si lo deseamos, la voz sea más lenta con ‘true’(lenta) o ‘false’ (más rápida). Un parámetro muy curioso sería el “*tld*”, el cual nos permite obtener variantes de acentos o entonaciones.

Otra librería que implementamos es la librería “os” la cual nos permite reproducir el audio directamente desde nuestro script. En nuestro caso, la usamos para eliminar el archivo .mp3 en reproducción.

La biblioteca adicional que hemos usado sería ‘*pygame*’, la cual nos permite reproducir el audio directamente desde nuestro script.

```
# Reproducir el archivo de audio
pygame.mixer.init()
pygame.mixer.music.load("saludo.mp3")
pygame.mixer.music.play()
```

Por último, hay que recalcar que la librería gTTS en sí misma es bastante simple de usar, pero si necesitas funcionalidades más avanzadas, como control sobre la velocidad de bits del audio o ajustes específicos de la voz, sería necesario exportar otras bibliotecas o servicios de texto a voz más especializados.