



Facultad de Ingeniería
Escuela de Ingeniería de Sistemas y Computación

Mauricio Gaona
mauricio.gaona@correounivalle.edu.co

Profesor

2025-I



Agenda

- Reconocimiento del habla
- Librerías - Servicios - Plataformas
 - Asignación a cada Grupo un API (Taller 2)





¿Qué es el Reconocimiento del Habla?

Reconocimiento de voz se refiere al proceso de identificar y distinguir la **voz de una persona** en particular.

El reconocimiento del habla es la capacidad de un sistema para entender y procesar el lenguaje hablado, convirtiéndolo en texto o acciones.

Aplicaciones:

- Asistentes virtuales (Siri, Google Assistant)
- Comandos por voz en dispositivos y vehículos
- Dictado en procesadores de texto
- Transcripción de entrevistas, conferencias, etc.
- Comandos para aplicaciones de IA





Proceso de Reconocimiento del Habla

1. Captura de Audio: El micrófono del dispositivo captura las ondas sonoras.
2. Preprocesamiento: Filtrado de ruido y segmentación de señales.
3. Conversión a Texto: Uso de modelos de lenguaje para convertir el audio en texto.
4. Postprocesamiento: Ajuste y corrección del texto para mejorar precisión.



Componentes Técnicos Clave

- **Modelos Acústicos:** Transforman ondas sonoras en representaciones digitales.
- **Modelos de Lenguaje:** Predicen secuencias de palabras para mejorar la precisión.
- **Gramáticas:** Limitan las palabras o frases que el sistema puede reconocer.
- **Modelos de Redes Neuronales:** Utilizados en sistemas avanzados para mejorar la capacidad de entendimiento.



Retos del Reconocimiento del Habla

- **Ruido de fondo:** Dificulta la captación clara de la voz.
- **Acentos y dialectos:** Variaciones en la pronunciación pueden reducir la precisión.
- **Homófonos:** Palabras que suenan igual, pero tienen significados diferentes (ej. 'caza' vs 'casa').
- **Contexto:** Dificultad para entender el significado sin un contexto adecuado.



Usos en la Vida Cotidiana

- Educación: Transcripción automática de clases.
- Automatización en atención al cliente: Centros de llamadas automáticos que identifican intenciones del usuario y brindan respuestas rápidas, también análisis de sentimientos y detección de emociones en clientes.
- Automatización del hogar: Control por voz de iluminación, temperatura, seguridad y electrodomésticos inteligentes.
- Asistentes virtuales: Siri, Alexa, Google Assistant, Cortana. Permiten ejecutar comandos por voz como reproducir música, controlar dispositivos, consultar el clima, y responder preguntas generales.
- Salud y retail : Reconocimiento de voz en sistemas médicos y comercio minorista para registros rápidos.
- Automóviles (control de vehículos): Comandos de voz para controlar funciones del vehículo.
- Accesibilidad: Asistentes de voz para personas con discapacidades motoras.
- Automatización de procesos y productividad: Controlar aplicaciones y software mediante comandos de voz para realizar tareas repetitivas de forma más eficiente.



Reconocimiento del habla

Crear una Aplicación con reconocimiento del habla

1. Captura del Audio

- Se graba la voz del usuario mediante un micrófono o dispositivo compatible.
- La calidad del audio es clave para una buena precisión en el reconocimiento.

2. Reconocimiento de Voz (Speech-to-Text)

- Convierte el audio capturado a texto utilizando un motor de reconocimiento de voz.
- Herramientas comunes: Google Speech-to-Text, Whisper (OpenAI), Microsoft Azure Speech Service, Amazon Transcribe, etc

3. Procesamiento del Lenguaje Natural (NLP)

- Analiza el texto obtenido para identificar la intención del usuario y extraer el comando o la acción requerida.
- Herramientas: ChatGPT (OpenAI), GPT-4, Dialogflow, IBM Watson Assistant, Azure Language Understanding.



Reconocimiento del habla

Crear una Aplicación con reconocimiento de voz

4. Lógica de Ejecución

- Ejecuta la acción o tareas correspondientes según el comando identificado.
- Puede involucrar consultas a bases de datos, interacción con dispositivos físicos o servicios web externos.

5. Generación de Respuesta (Texto)

- Genera una respuesta textual acorde a los resultados obtenidos en el paso anterior.
- Puede usar modelos LLM como ChatGPT, GPT-4.5, Gemini, etc.

6. Síntesis de Voz (Text-to-Speech) (Opcional)

- Convierte la respuesta textual generada en audio usando motores de texto a voz.
- Herramientas comunes: Google Text-to-Speech, Amazon Polly, Microsoft Azure Text-to-Speech.

7. Entrega de la Respuesta

- Retorna la respuesta al usuario en formato texto formateado (pantalla) o audio (altavoz) según la interfaz diseñada.



Librerías - Servicios - Plataformas

- **1. Amazon Polly** - <https://aws.amazon.com/polly/>
- **2. Azure Speech Services** - <https://azure.microsoft.com/en-us/services/cognitive-services/speech/>
- **3. Dialogflow** - <https://dialogflow.com/>
- **4. Google Cloud Speech-to-Text** - <https://cloud.google.com/speech-to-text/>
- **5. Amazon Transcribe** - <https://aws.amazon.com/transcribe/>
- **6. Watson Speech to Text de IBM** - <https://www.ibm.com/products/speech-to-text>
- **7. Whisper (OpenAI) y whisper-large-v3-local**
- **8. ElevenLabs**
- **9.** <https://platform.openai.com/docs/guides/text-to-speech>
- **10.** <https://play.ht/>

Reconocimiento del habla

Grupo

Librerías - Servicios - Plataformas

Realizar un MVP que haga uso de reconocimiento del habla.

Ruleta



¿Preguntas?



Proyecto Integrador II

Economy of the
European Union

Gracias

