

Soy una IA y oigo voces!

Ponentes

Adrián García

Roberto Sánchez

¡Bienvenidos!

Abril 2024

Qué es Speech-to-Text

Contexto

- Gran volumen de información en audio
 - Reuniones
 - Llamadas a servicios de atención al cliente
 - Audios de películas, ...
- Necesidad de procesamiento automatizado

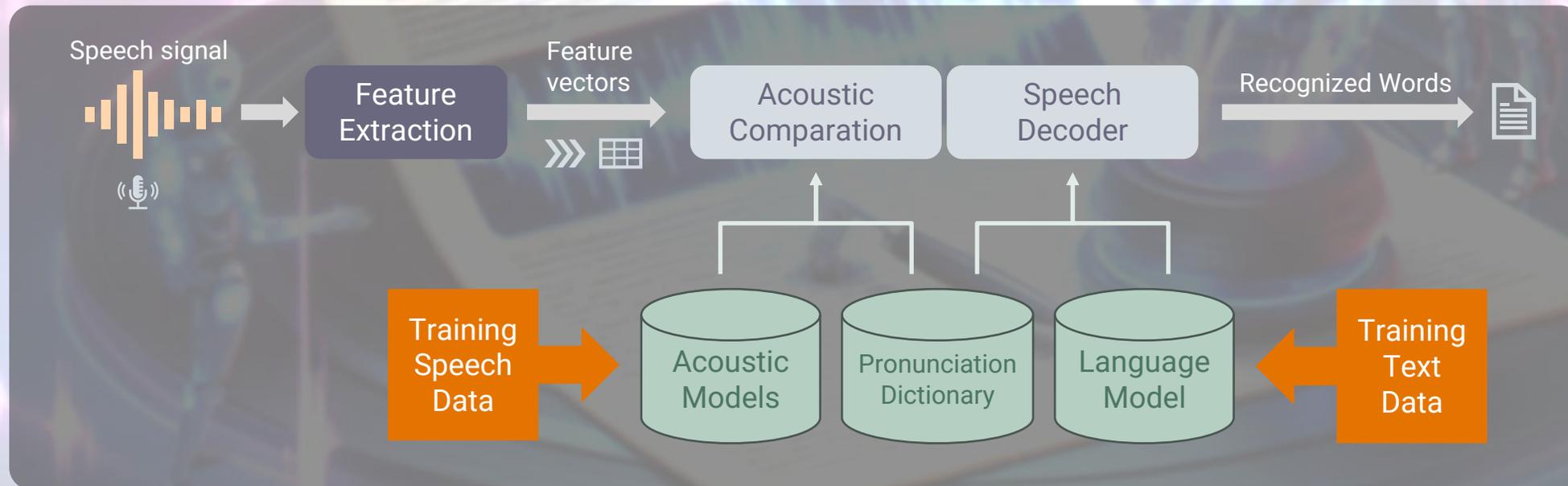
Tecnología

- ASR
- Diarización
- Identificación de locutores
- Análisis de sentimiento
- NLP

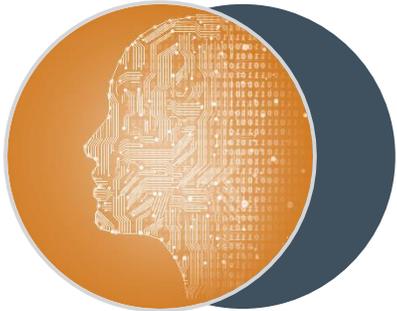
ASR (Automatic Speech Recognition)

Tecnología que permite a las máquinas entender y transcribir el habla humana en texto.

Utiliza algoritmos avanzados para extraer características de señales de audio y reconocer patrones lingüísticos que permiten transcribir el habla.



ASR (Automatic Speech Recognition)



Aplicaciones

- Telefonía: Seguimiento y análisis de llamadas
- Plataformas de Video: Subtitulación en tiempo real y asincrónica
- Monitoreo de Medios: Detectar menciones de marcas y otros temas de interés
- Reuniones Virtuales: Transcripciones precisas y la capacidad de analizar este contenido para generar ideas clave y acciones

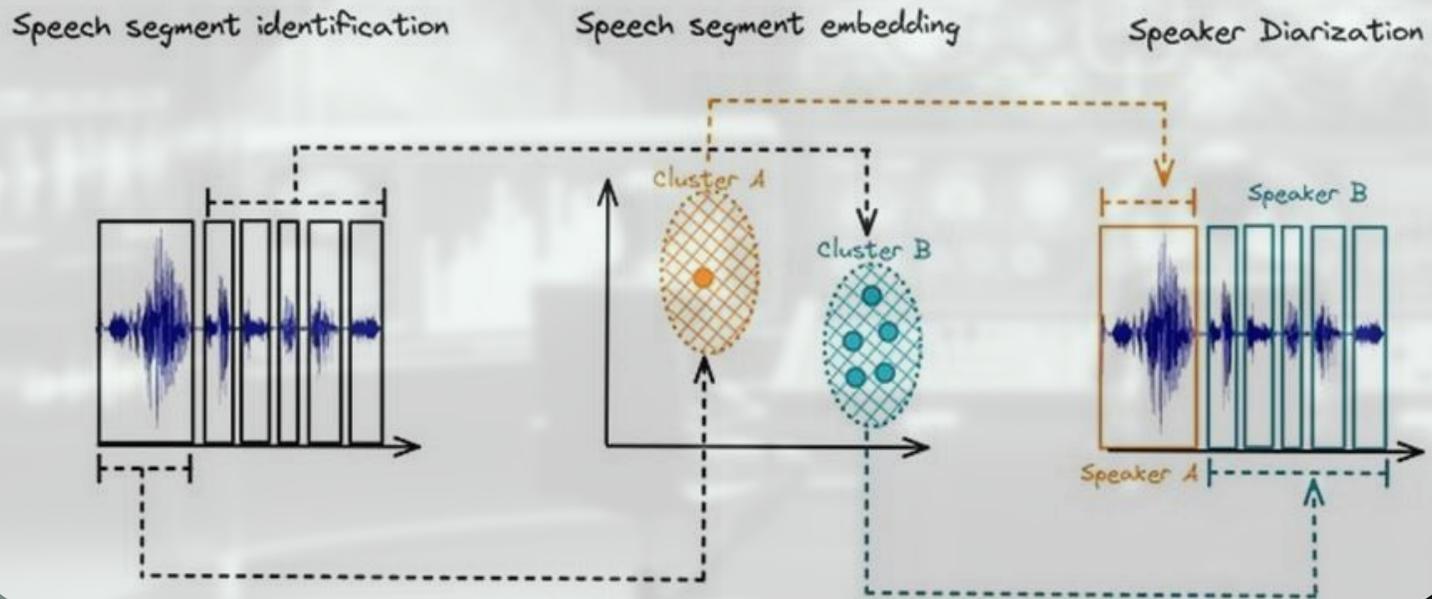


Desafíos

- Variabilidad del habla y acentos
- Entornos Ruidosos
- Idiomas complejos
- Mezcla de idiomas en la misma frase
- Múltiples interlocutores y solape entre las intervenciones
- Reconocimiento de nombres propios y términos específicos

Diarización de hablantes (speaker diarization)

Proceso que permite diferenciar los locutores en cada momento en una conversación.



Aplicaciones

Transcripción de reuniones

Transcripción de llamadas telefónicas

Generación de subtítulos para Videos

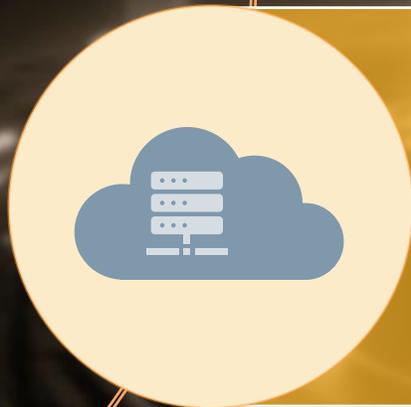
...

Diarización de hablantes (speaker diarization)



Open Source

- Sin coste por uso, pero necesitan recursos locales
- Curva de aprendizaje mayor, pero capacidad de ajuste
- Privacidad de los datos
- Ejemplos: PyAnnote, Kaldi, NVIDIA NeMO



API Cloud

- Coste por uso
- Fácil de implementar, pero funcionalidades estándar
- Transmisión de datos a servidor remoto
- Ejemplos: AssemblyAI, Picovoice (Falcon Speaker Diarization), Deepgram

Identificación de locutores (Speakers recognition)

Asociar cada intervención a locutores concretos, previamente registrados

Perfilado las características de cada potencial hablante

Utilizar los segmentos del proceso de diarización

Se calcula la probabilidad por cada segmento de pertenecer a cada hablante

Cada segmento se asocia al hablante con mayor probabilidad

Aplicaciones

Análisis de reuniones de negocios

Servicios de seguridad y vigilancia

Personalización de servicios

...

Procesado de los textos generados (NLP)

Análisis de Sentimientos

Extracción de Entidades Nombradas

Desambiguación de sentido por contexto

Resumen Automático

Traducción Automática

LLM

La aparición de los LLM ha supuesto una gran revolución para el procesado de textos

- Mayor comprensión del texto analizado
- Generación de texto coherente y fluida
- Capacidad para realizar múltiples tareas sin entrenamiento específico



NLP: El Procesamiento del Lenguaje Natural es una rama de la IA cuyo objetivo es permitir que las máquinas entiendan, interpreten, generen y respondan al lenguaje humano de una manera valiosa y útil.

Caso práctico: “Speech-toText PoC”

ASR

Diarización

Id. locutores

Resumen (acta)

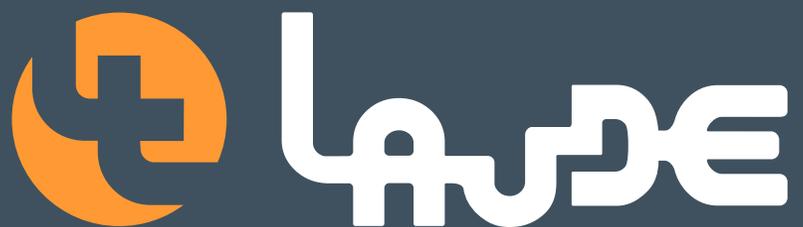
RAG (chat)



Speech-to-Text PoC

<https://s2t.ai.laude.tech/>





Soy una IA y oigo voces!

Ponentes

Adrián García

Roberto Sánchez

¡Gracias!

Abril 2024