

El oído del agente: qué es el STT

Datos de esta lección

Fase	F1 — Cimientos, seguridad y arquitectura del sistema
Módulo	M1.2 — Modelos mentales: radiografía de un agente telefónico
Puntos de profesionalidad al superar el test	15

Objetivo práctico

Al terminar esta lección podrás explicar qué es el STT (reconocimiento de voz), qué papel cumple en un agente telefónico y qué factores hacen que transcriba mejor o peor lo que dice la persona que llama.

Resultado que obtendrás

Un modelo mental claro de la primera pieza del flujo de una llamada: cómo la voz humana se convierte en texto que el LLM puede procesar.

Dónde encaja dentro del sistema final

En el sistema TELÉFONO → AGENTE → CEREBRO → VOZ → AUTOMATIZACIÓN → CALENDARIO, el STT es el «oído»: convierte lo que dice la persona que llama en texto, antes de que ese texto llegue al LLM (el cerebro).

Conocimientos previos

- Lección anterior: «El cerebro del agente: qué es un LLM».

Herramientas necesarias

- Un navegador con NVDA activo, en modo de exploración.

Posible coste

Esta lección es conceptual y no requiere ningún gasto. El coste real del STT en producción se calcula en el módulo 2.1, «Gestión financiera y eliminación de riesgos».

Modelo mental

Piensa en el oído humano y en una secretaria que escribe exactamente lo que oye, palabra por palabra, mientras alguien habla por teléfono. Esa secretaria no interpreta el significado de lo que se dice: sólo transcribe el sonido en texto, con la mayor fidelidad posible. Esa es exactamente la función del STT dentro del sistema.

Explicación

STT son las siglas de *Speech To Text* (voz a texto). Es la tecnología que convierte el audio de la voz humana en texto escrito, en tiempo real durante una llamada. El STT es una pieza separada del LLM: primero se transcribe lo que dice la persona (STT), y sólo después ese texto se envía al LLM para que decida qué responder.

La calidad de la transcripción depende de varios factores: la claridad del audio (ruido de fondo, calidad de la línea), el acento y la velocidad al hablar, el vocabulario específico del sector (nombres propios, términos técnicos), y el idioma configurado. Un STT mal configurado —por ejemplo, en el idioma equivocado— puede hacer que toda la conversación falle desde el primer segundo, aunque el LLM y la voz de salida (TTS) estén perfectamente configurados. Por eso, en el método de diagnóstico por descarte de este curso (módulo 4.1), comprobar la transcripción es uno de los primeros pasos cuando una llamada «no entiende nada» de lo que dice la persona.

Vocabulario nuevo

STT

Piénsalo así: El oído y la secretaria que transcribe literalmente lo que escucha.

Speech To Text: tecnología que convierte audio de voz en texto escrito.

Ejemplo: Cuando llamas a un agente y dices «quiero pedir cita para el jueves», el STT convierte ese audio en el texto «quiero pedir cita para el jueves» antes de que el LLM lo procese.

Preparación

No se requiere preparación técnica.

Procedimiento paso a paso

1. **Contexto:** Vas a localizar el término «STT» en el Glosario acumulativo de esta plataforma y a comprobar el enlace que te lleva de vuelta a esta lección.

Acción de teclado: Abre el Glosario (/glosario.php) y pulsa **H** hasta el encabezado de nivel 2 «S», y después hasta el encabezado de nivel 3 «STT».

Respuesta esperada de NVDA: NVDA debería anunciar «STT, encabezado de nivel 3» al llegar.

Qué significa: Confirmas que el término está documentado y accesible por navegación de encabezados, sin depender de desplazamiento visual.

Acción siguiente: Sigue leyendo con la flecha abajo hasta el enlace «Aparece por primera vez en...» y confirma que apunta a esta misma lección.

2. **Contexto:** Comprueba que ese enlace funciona y te devuelve al punto correcto.

Acción de teclado: Con el foco sobre el enlace «Aparece por primera vez en...», pulsa **Enter**.

Respuesta esperada de NVDA: NVDA debería anunciar el título de esta lección, «El oído del agente: qué es el STT», al cargar la página.

Qué significa: El glosario y las lecciones están enlazados entre sí en ambas direcciones, lo que te permite moverte por el curso sin perder el hilo.

Acción siguiente: Vuelve atrás con el navegador si quieres seguir explorando otros términos del glosario.

Errores frecuentes y diagnóstico

El agente responde de forma completamente incoherente a lo que la persona ha dicho.

Cómo localizarlo: Antes de sospechar del LLM (el cerebro), conviene comprobar la transcripción real que ha generado el STT: si la transcripción ya está mal, el LLM está respondiendo correctamente a un texto equivocado.

Solución: Revisar en los registros (logs) de la plataforma de voz qué texto exacto se transcribió, y compararlo con lo que realmente se dijo. Se explica en detalle en el módulo 4.1, «Logs sin saturación».

Comprobación final

Explica con tus palabras la diferencia entre el STT y el LLM: uno transcribe, el otro decide qué responder. Si puedes dar un ejemplo de un fallo que sería culpa del STT y otro que sería culpa del LLM, la comprobación es positiva.

Qué acabas de conseguir

Sabes distinguir la primera pieza del flujo de la llamada (el oído) del resto del sistema, lo que te permitirá diagnosticar con más precisión cuándo algo falla.

Ejercicio práctico

Piensa en tres situaciones reales que podrían dificultar una buena transcripción por STT (por ejemplo, una persona llamando desde un lugar ruidoso) y anota cómo podrías comprobar, sin herramientas técnicas todavía, si el problema viene de ahí.

Resumen de lo imprescindible

- STT significa Speech To Text: convierte voz en texto.
- Es una pieza separada del LLM: primero se transcribe, después se interpreta.
- La calidad del audio, el acento, la velocidad y el idioma configurado afectan a la transcripción.
- Ante una respuesta incoherente del agente, comprobar la transcripción es un paso de diagnóstico temprano.

No necesitas aprender todavía

Todavía no necesitas saber qué proveedor concreto de STT usa cada plataforma (Vapi, Retell): eso se explica cuando configures cada una, en la Fase 2.

Estoy preparado para continuar si puedo...

- Explicar qué es el STT y diferenciarlo del LLM.
- Explicar por qué comprobar la transcripción es un paso de diagnóstico temprano.
- Nombrar al menos dos factores que afectan a la calidad de una transcripción.

Test de comprobación

Para registrar el resultado y obtener Puntos de profesionalidad, realiza este test

desde la versión web de la lección. Las opciones se muestran aquí sin señalar la respuesta correcta.

1. Un agente responde con una frase que no tiene ninguna relación con lo que la persona que llama acaba de decir. Siguiendo el método de diagnóstico por descarte de este curso, ¿qué pieza tiene sentido revisar primero?
 - a. El calendario de reservas, porque cualquier fallo empieza siempre por ahí.
 - b. La transcripción generada por el STT, para comprobar si el LLM está respondiendo correctamente a un texto ya erróneo.
 - c. La automatización con n8n o Make, porque siempre es la causa de una respuesta incoherente.
2. Un agente atiende llamadas correctamente durante todo el día, pero falla sistemáticamente con clientes que llaman desde un taller mecánico con mucho ruido de fondo. ¿Qué pieza del sistema es más probable que esté detrás de ese patrón?
 - a. El STT, porque el ruido de fondo dificulta una transcripción fiable de la voz.
 - b. El TTS, porque la voz de salida del agente suena distinta según el ruido de quien llama.
 - c. El calendario, porque los talleres mecánicos suelen tener horarios distintos.
3. ¿Por qué el STT y el LLM se consideran piezas separadas del sistema, en lugar de una única pieza que «entiende» directamente el audio?
 - a. Porque primero hace falta convertir el audio en texto (STT) y sólo después ese texto se envía al LLM para decidir la respuesta; son dos tareas distintas con posibles fallos distintos.
 - b. Porque son exactamente la misma tecnología con dos nombres distintos según la plataforma.
 - c. Porque el LLM nunca recibe texto, sólo procesa audio directamente.

Fuentes

- [File transcription \(Speech to text\) — OpenAI \(consultado el 2026-08-30\)](#)

Última actualización de este contenido: **2026-02-10**.