



FICHA TÉCNICA • CAPTURAR • VOZ STT

¿Transcribir las reuniones palabra por palabra?

A partir de ahora, cada grabación vuelve como texto.

Voz STT convierte en texto las grabaciones de reuniones, llamadas y dictados. Detecta el idioma por sí mismo, y el nivel Ultra separa a los hablantes. Pensada para equipos de oficina, ventas y proyectos, en la app y mediante API.

DE UN VISTAZO

Archivos de audio y vídeo de hasta 50 MB, detección automática del idioma, tres niveles de calidad, separación de hablantes en Ultra y marcas de tiempo por palabra, SRT y VTT mediante API.

01 · DE QUÉ TRATA

¿Primero la reunión y luego escribir? Ahora el texto llega solo.

El acta de la reunión de proyecto tiene que salir hoy, y la grabación dura una hora. Con Voz STT sube el archivo y recibe el texto de vuelta, separado por participante en Ultra, listo para copiar o traducir.

ENTRADA

Audio (MP3, WAV, OGG, FLAC, M4A) y vídeo (MP4, WebM, AVI, MOV) de hasta 50 MB

IDIOMAS

Detección automática; 100 códigos de idioma por API, 15 idiomas para elegir en la app

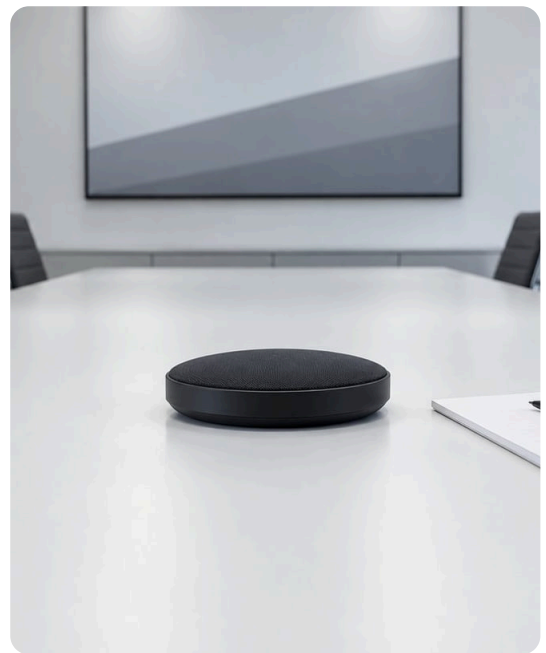
SALIDA

Transcripción en texto plano; en Ultra, como diálogo por hablante; descarga en TXT

ACCESO

App en

Media AI › Voice Studio
, API REST y MCP

**Idioma detectado automáticamente**

Voz STT detecta por sí mismo el idioma de la grabación. Si lo conoce, puede fijarlo; por API, entre 100 códigos de idioma.

Tres niveles de calidad

Basic devuelve texto plano rápido, Premium añade segmentos con marcas de tiempo por palabra, Ultra además separa a los hablantes. El nivel está en **Calidad de transcripción**.

Separar a los hablantes

En Ultra, Voz STT detecta a los participantes y muestra la transcripción como diálogo. Cambie el nombre de los hablantes con un clic.

Vocabulario

Con **Vocabulario técnico**, Voz STT reconoce con más precisión los términos técnicos de sus bases de conocimiento y listas de precios. Por API puede añadir sus propias palabras clave desde Premium.

Subtítulos mediante API

Por API, desde Premium se devuelven SRT y VTT más las marcas de tiempo por palabra, y Ultra añade RTTM para el análisis de hablantes.

Paso a Translate Studio

Usar en **Translate Studio** pasa la transcripción directamente a Translate Studio, sin copiar ni pegar.

02 · CÓMO FUNCIONA

Suelte el archivo, elija un nivel. Use la transcripción.

En Voice Studio son cuatro pasos. Mediante API REST y MCP puede enviar archivos de audio directamente desde sus propios procesos.

1

Abrir Voice Studio

En **Media AI > Voice Studio**, elija la pestaña **Voz a texto (STT)**. Desde la página de inicio, **Herramientas rápidas > Media > Voz a texto** también lo lleva allí.

2

Soltar el archivo

Arrastre el archivo de audio o vídeo al campo. Deje el idioma en Auto o elíjalo.

3

Elegir nivel y opciones

En **Calidad de transcripción**, elija Basic, Premium o Ultra. En Ultra puede definir el número de hablantes en **Reconocimiento de interlocutores**.

4

Usar la transcripción

Cópiela, guárdela como archivo de texto con **Descargar como TXT** o envíela a Translate Studio con un clic.

DATOS TÉCNICOS

CARACTERÍSTICA	DETALLES
Formatos de entrada	MP3, WAV, OGG, FLAC, M4A, MP4, WebM, AVI, MOV; en la app hasta 50 MB, un archivo por ejecución
Idiomas	Detección automática; 100 códigos de idioma por API, 15 opciones fijas en la app
Niveles de calidad	Basic (predeterminado en API), Premium, Ultra (predeterminado en la app)
Salida por API	output_format json, srt y vtt desde Premium, rttm en Ultra
Separación de hablantes	En Ultra, min_speakers y max_speakers opcionales de 1 a 20
API REST	POST /job/add/paperoffice_voice___stt con audio_file
MCP	po_stt_transcribe en /mcp-full

03 · EN LA PRÁCTICA

Reunión, dictado, vídeo de formación. Cada uno como texto terminado.

Acta de una reunión de proyecto

El responsable del proyecto sube la grabación en Ultra, cambia el nombre de Hablante 1 y Hablante 2 por sus nombres y copia el diálogo en el acta.

Dictado desde el terreno

Un topógrafo dicta en su teléfono por la carretera y sube el archivo M4A en la oficina. Basic devuelve el texto plano para el informe, y el vocabulario conoce los nombres de los productos.

Subtítulos para un vídeo de formación

Desarrollo envía el vídeo por API con `quality=premium` y `output_format=srt` y recibe de vuelta el archivo de subtítulos con marcas de tiempo.

PREGUNTAS FRECUENTES

¿Qué idiomas reconoce Voz STT?

El idioma se detecta automáticamente. Por API puede fijarlo entre 100 códigos de idioma, y la app ofrece 15 idiomas como opciones fijas.

¿Puede Voz STT distinguir a los hablantes?

Sí, en Ultra. La transcripción aparece como diálogo, y los hablantes pueden renombrarse. Por API también obtiene el número de hablantes, los solapamientos y el RTTM.

¿Puedo obtener subtítulos en SRT o VTT?

Sí, por API desde el nivel Premium con `output_format=srt` o `vtt`. En la app guarda la transcripción como archivo TXT.

¿Cómo se factura?

Por uso, según el nivel de calidad. Los precios actuales están en la página de precios.

EMPEZAR

Deje que su próxima grabación se transcriba por usted.

Inicie sesión en su cuenta de PaperOffice, abra **Media AI > Voice Studio** y elija **Voz a texto (STT)**. Suelte el archivo, elija un nivel e inicie la transcripción.

paperoffice.ai/es/ai-voice-stt-voz-a-texto-transcripcion-reconocimiento-stt



MÁS SOBRE EL MÓDULO

paperoffice.ai/es/ai-voice-stt-voz-a-texto-transcripcion-reconocimiento-stt

PRECIOS

paperoffice.ai/es/precios

CONTACTO

paperoffice.ai/es/contacto