



FICHA TÉCNICA • CAPTURA • VOZ STT

Transcrever reuniões palavra por palavra?

De agora em diante, cada gravação volta como texto.

A Voz STT transforma gravações de reuniões, chamadas e ditados em texto. Detecta o idioma sozinha, e o nível Ultra separa os falantes. Pensada para equipes de escritório, vendas e projetos, no app e pela API.

EM RESUMO

Arquivos de áudio e vídeo de até 50 MB, detecção automática do idioma, três níveis de qualidade, separação de falantes no Ultra e marcações de tempo por palavra, SRT e VTT via API.

01 · DO QUE SE TRATA

Primeiro a reunião, depois a escrita? Agora o texto chega sozinho.

A ata da reunião de projeto precisa sair hoje, e a gravação dura uma hora. Com a Voz STT você carrega o arquivo e recebe o texto de volta, separado por participante no Ultra, pronto para copiar ou traduzir.

ENTRADA

Áudio (MP3, WAV, OGG, FLAC, M4A) e vídeo (MP4, WebM, AVI, MOV) de até 50 MB

IDIOMAS

Detectados automaticamente; 100 códigos de idioma via API, 15 idiomas para escolher no app

SAÍDA

Transcrição em texto simples; no Ultra, um diálogo por falante; baixável como TXT

ACESSO

App em

Media AI › Voice Studio

, API REST e MCP

**Idioma detectado automaticamente**

A Voz STT detecta sozinha o idioma da gravação. Se você souber, pode defini-lo: via API, a partir de 100 códigos de idioma.

Três níveis de qualidade

O Basic devolve texto simples rapidamente, o Premium acrescenta segmentos com marcações de tempo por palavra, e o Ultra separa também os falantes. O nível fica em **Qualidade da transcrição**.

Separar os falantes

No Ultra, a Voz STT detecta os participantes e mostra a transcrição como diálogo. Renomeie os falantes com um clique.

Vocabulário

Com **Vocabulário técnico**, a Voz STT reconhece com mais precisão os termos técnicos das suas bases de conhecimento e listas de preços. Pela API, você pode acrescentar as próprias palavras-chave prioritárias a partir do Premium.

Legendas via API

Pela API, o Premium e os níveis superiores devolvem SRT e VTT, além de marcações de tempo por palavra, e o Ultra acrescenta RTTM para análise de falantes.

Passagem para o Translate Studio

Usar no **Translate Studio** entrega a transcrição diretamente ao Translate Studio, sem copiar e colar.

02 · COMO FUNCIONA

Solte o arquivo, escolha um nível. Use a transcrição.

No Voice Studio, são quatro passos. Pela API REST e pelo MCP, você envia arquivos de áudio diretamente dos seus próprios processos.

1

Abrir o Voice Studio

Em **Media AI > Voice Studio**, escolha a aba **Fala para texto (STT)**. A partir da página inicial, **Ferramentas Rápidas > Media > Fala para Texto** também leva você até lá.

2

Soltar o arquivo

Arraste o arquivo de áudio ou vídeo para o campo. Deixe o idioma em Automático ou escolha um.

3

Escolher nível e opções

Em **Qualidade da transcrição**, escolha Basic, Premium ou Ultra. No Ultra, você pode definir o número de falantes em **Reconhecimento de interlocutores**.

4

Usar a transcrição

Copie-a, salve-a como arquivo de texto com **Baixar como TXT** ou entregue-a ao Translate Studio com um clique.

DADOS TÉCNICOS

FUNCIONALIDADE	DETALHES
Formatos de entrada	MP3, WAV, OGG, FLAC, M4A, MP4, WebM, AVI, MOV; no app até 50 MB, um arquivo por execução
Idiomas	Deteção automática; 100 códigos de idioma via API, 15 opções fixas no app
Níveis de qualidade	Basic (padrão da API), Premium, Ultra (padrão do app)
Saída via API	output_format json, srt e vtt a partir do Premium, rttm no Ultra
Separação de falantes	no Ultra, min_speakers e max_speakers opcionais, de 1 a 20
API REST	POST /job/add/paperoffice_voice___stt com audio_file
MCP	po_stt_transcribe em /mcp-full

03 · NA PRÁTICA

Reunião, ditado, vídeo de treinamento. Cada um como texto pronto.

Ata de uma reunião de projeto

O responsável do projeto carrega a gravação no Ultra, renomeia Speaker 1 e 2 para os respectivos nomes e copia o diálogo para a ata.

Ditado do campo

Um topógrafo dita no celular na estrada e carrega o arquivo M4A no escritório. O Basic devolve o texto simples para o relatório, e o vocabulário conhece os nomes dos produtos.

Legendas para um vídeo de treinamento

O desenvolvimento envia o vídeo via API com `quality=premium` e `output_format=srt` e recebe de volta o arquivo de legendas com marcações de tempo.

PERGUNTAS FREQUENTES

Quais idiomas a Voz STT reconhece?

O idioma é detectado automaticamente. Via API, você pode defini-lo a partir de 100 códigos de idioma, e o app oferece 15 idiomas como opções fixas.

A Voz STT consegue distinguir os falantes?

Sim, no Ultra. A transcrição aparece como diálogo, e os falantes podem ser renomeados. Via API, você obtém também o número de falantes, as sobreposições e o RTTM.

Consigo legendas em SRT ou VTT?

Sim, via API a partir do nível Premium com `output_format=srt` ou `vtt`. No app, você salva a transcrição como arquivo TXT.

Como ela é faturada?

Por uso, conforme o nível de qualidade. Os preços atuais estão na página de preços.

COMEÇAR

Deixe sua próxima gravação ser transcrita automaticamente.

Faça login na sua conta PaperOffice, abra **Media AI > Voice Studio** e escolha **Fala para texto (STT)**. Solte o arquivo, escolha um nível e inicie a transcrição.

paperoffice.ai/pt/ai-voice-stt-voz-para-texto-transcricao-reconhecimento-stt



MAIS SOBRE O MÓDULO

paperoffice.ai/pt/ai-voice-stt-voz-para-texto-transcricao-reconhecimento-stt

PREÇOS

paperoffice.ai/en/pricing

CONTATO

paperoffice.ai/pt/contato