



データシート・キャプチャ・VOICE STT

会議の内容を一言一句書き起こしていますか？ これからは、すべての録音がテキスト で戻ってきます。

Voice STTは、会議、通話、口述の録音をテキストに変換します。言語は自動で判別され、Ultraティアでは話者を区別します。事務、営業、プロジェクトチームのために、アプリとAPIから利用できます。

一目で分かること

50 MBまでの音声・動画ファイル、言語の自動判別、三段階の品質ティア、Ultraでの話者の分離、そしてAPIで単語タイムスタンプ、SRT、VTT。

01・概要

まず会議、そのあとタイピング？ 今では、テキストが自動で届きます。

プロジェクト会議の議事録を今日中に出さなければなりません、録音は一時間あります。Voice STTでは、ファイルをアップロードするだけで、Ultraでは参加者ごとに分けられたテキストが戻ります。そのままコピーしたり翻訳したりできます。

入力

音声（MP3、WAV、OGG、FLAC、M4A）および動画（MP4、WebM、AVI、MOV）、50 MBまで

言語

自動判別。APIでは100の言語コードから指定可能、アプリでは15言語から選択

出力

ブレンテキストの文字起こし。Ultraでは話者ごとの対話形式。TXTとしてダウンロード

アクセス

Media AI ▶ ボイススタジオ
のアプリ、REST APIおよびMCP



言語を自動判別

Voice STTは、録音の言語を自動で判別します。言語が分かっている場合は指定でき、APIでは100の言語コードから選べます。

三段階の品質ティア

Basicは速いブレンテキストを返し、Premiumは単語タイムスタンプ付きのセグメントを加え、Ultraはさらに話者を分離します。ティアは **文字起こし品質** で設定します。

話者を分ける

Ultraでは、Voice STTが参加者を検出し、文字起こしを対話として表示します。話者の名前はワンクリックで変更できます。

語彙

専門用語 を使うと、Voice STTはお使いのナレッジベースや価格表にある専門用語をより正確に認識します。APIでは、Premiumからご自身のホットワードを追加できます。

APIで字幕

APIでは、Premiumティア以上でSRTとVTTに加え単語タイムスタンプを返し、Ultraでは話者分析用のRTTMを追加します。

翻訳スタジオへ

Translate Studioで使用する を使うと、文字起こしをコピーや貼り付けなしで翻訳スタジオにそのまま渡せます。

02・仕組み

ファイルを入れ、ティアを選ぶ。 文字起こしを使う。

Voice Studioでは四つの手順で完了します。REST APIとMCPでは、ご自身のプロセスから音声ファイルを直接送信できます。

1

Voice Studioを開く

Media AI › ボイススタジオ の下で **音声文字起こし (STT)** タブを選びます。スタートページからは **クイックツール › メディア › 音声認識** でも開けます。

2

ファイルを入れる

音声または動画ファイルを入力欄にドラッグします。言語は自動のままにするか、指定してください。

3

ティアとオプションを選ぶ

文字起こし品質 の下で、Basic、Premium、Ultraのいずれかを選びます。Ultraでは、**会話相手の識別** で話者数を設定できます。

4

文字起こしを使う

コピーするか、**TXTとしてダウンロード** でテキストファイルとして保存するか、ワンクリックで翻訳スタジオに渡せます。

技術データ

機能	詳細
入力形式	MP3、WAV、OGG、FLAC、M4A、MP4、WebM、AVI、MOV。アプリでは最大50 MB、実行ごとにファイル
言語	自動判別。APIでは100の言語コード、アプリでは15の固定選択肢
品質ティア	Basic（APIの既定）、Premium、Ultra（アプリの既定）
APIでの出力	output_format：json、Premiumからsrtとvtt、Ultraではrttm
話者の分離	Ultraでは、任意でmin_speakersとmax_speakersを1から20の範囲で指定
REST API	POST /job/add/paperoffice_voice___stt、audio_fileを指定
MCP	/mcp-full の po_stt_transcribe

03・実際の活用

会議、口述、研修動画。 それぞれ完成したテキストで。

プロジェクト会議の議事録

プロジェクトリーダーが録音をUltraでアップロードし、話者1と話者2をそれぞれの名前に変えて、対話を議事録にコピーします。

現場からの口述

測量士が移動中にスマートフォンに口述し、事務所でM4Aファイルをアップロードします。Basicが報告書用のプレーンテキストを返し、語彙には製品名が登録されています。

研修動画の字幕

開発チームは、quality=premium と output_format=srt を指定して動画をAPIで送り、タイムスタンプ付きの字幕ファイルを受け取ります。

よくある質問

Voice STTはどの言語を認識しますか？

言語は自動で判別されます。APIでは100の言語コードから指定でき、アプリでは15言語が固定の選択肢として用意されています。

Voice STTは話者を区別できますか？

はい。Ultraでは可能です。文字起こしは対話として表示され、話者の名前を変更できます。APIでは、話者数、重なり、RTTMも取得できます。

字幕をSRTまたはVTTで受け取れますか？

はい。APIでPremiumティア以上から、output_format=srt または vtt を指定します。アプリでは、文字起こしをTXTファイルとして保存します。

料金はどのように請求されますか？

利用量に応じて、品質ティアによって決まります。現在の料金は料金ページに掲載されています。

はじめに

次の録音は、テキストにまとめてもらいましょう。

PaperOfficeアカウントにサインインし、Media AI > ボイススタジオ を開いて音声文字起こし (STT) を選びます。ファイルを入れ、ティアを選び、文字起こしを開始してください。

paperoffice.ai/ja/ai-voice-stt-onsei-tekisuto-henkan-stt



モジュールの詳細

paperoffice.ai/ja/ai-voice-stt-onsei-tekisuto-henkan-stt

料金

paperoffice.ai/en/pricing

お問い合わせ

paperoffice.ai/ja/otoiawase