



DATENBLATT · ERFASSEN · VOICE STT

Besprechungen Wort für Wort abtippen? Ab jetzt kommt jede Aufnahme als Text zurück.

Voice STT schreibt Aufnahmen von Besprechungen, Anrufen und Diktaten als Text. Die Sprache erkennt es selbst, in der Stufe Ultra trennt es die Sprecher. Für Büro, Vertrieb und Projektteams, in der App und per API.

AUF EINEN BLICK

Audio- und Videodateien bis 50 MB, automatische Spracherkennung, drei Qualitätsstufen, Sprechertrennung in Ultra und per API Zeitmarken je Wort, SRT und VTT.

01 · Worum es geht

Erst das Meeting, dann das Tippen? Jetzt kommt der Text von allein.

Nach dem Projektmeeting soll das Protokoll heute noch hinaus, und die Aufnahme dauert eine Stunde. Mit Voice STT laden Sie die Datei hoch und erhalten den Text, in Ultra getrennt nach Gesprächspartnern, bereit zum Kopieren oder Übersetzen.

EINGABE

Audio (MP3, WAV, OGG, FLAC, M4A) und Video (MP4, WebM, AVI, MOV) bis 50 MB

SPRACHEN

Automatisch erkannt; per API 100 Sprachcodes, in der App 15 Sprachen zur Auswahl

ERGEBNIS

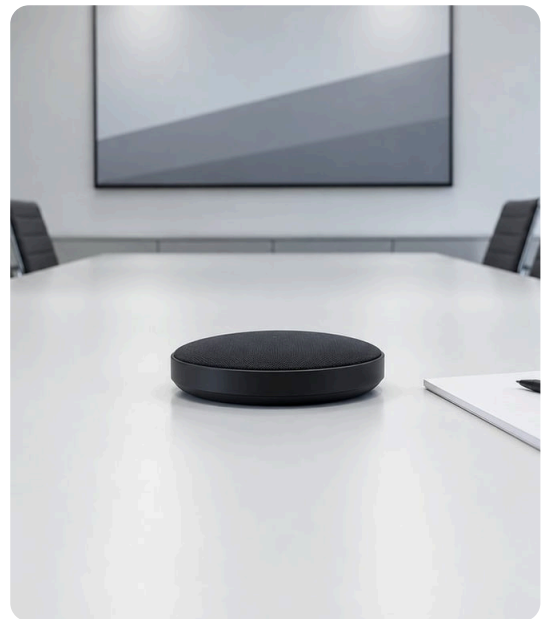
Transkript als Fließtext, in Ultra als Dialog je Sprecher; Download als TXT

ZUGANG

App unter

Media AI › Voice Studio

, REST-API und MCP

**Sprache automatisch erkannt**

Voice STT erkennt die Sprache der Aufnahme selbst. Wer sie kennt, gibt sie fest vor, per API aus 100 Sprachcodes.

Drei Qualitätsstufen

Basic liefert schnellen Fließtext, Premium ergänzt Segmente mit Zeitmarken je Wort, Ultra trennt zusätzlich die Sprecher. Die Stufe steht unter **Transkriptions-Qualität**.

Sprecher trennen

In Ultra erkennt Voice STT die Gesprächspartner und zeigt das Transkript als Dialog. Sprecher benennen Sie per Klick um.

Fachwortschatz

Mit **Fachwortschatz** erkennt Voice STT Fachbegriffe aus Ihren Wissensbasen und Preislisten genauer. Per API ergänzen Sie ab Premium eigene Hotwords.

Untertitel per API

Per API erhalten Sie ab Premium SRT und VTT sowie Zeitmarken je Wort, in Ultra zusätzlich RTTM für Sprecheranalysen.

Weiter ins Übersetzungsstudio

Mit **In Translate Studio verwenden** übergeben Sie das Transkript direkt an das Übersetzungsstudio, ohne Kopieren und Einfügen.

02 · SO FUNKTIONIERT ES

Datei ablegen, Stufe wählen. Transkript übernehmen.

Im Voice Studio sind es vier Schritte. Per REST-API und MCP senden Sie Audiodateien direkt aus Ihren eigenen Abläufen.

1

Voice Studio öffnen

Unter **Media AI > Voice Studio** wählen Sie den Reiter **Sprache zu Text (STT)**. Von der Startseite führt auch **Schnelltools > Media > Sprache zu Text** dorthin.

2

Datei ablegen

Ziehen Sie die Audio- oder Videodatei in das Feld. Die Sprache lassen Sie auf Automatisch oder wählen sie fest.

3

Stufe und Optionen wählen

Unter **Transkriptions-Qualität** wählen Sie Basic, Premium oder Ultra. In Ultra legen Sie unter **Gesprächspartner-Erkennung** optional die Zahl der Sprecher fest.

4

Transkript übernehmen

Kopieren, mit **Als TXT herunterladen** als Textdatei sichern oder mit einem Klick an das Übersetzungsstudio übergeben.

TECHNISCHE DATEN

MERKMAL	DETAILS
Eingabeformate	MP3, WAV, OGG, FLAC, M4A, MP4, WebM, AVI, MOV; in der App bis 50 MB, eine Datei je Lauf
Sprachen	automatische Erkennung; per API 100 Sprachcodes, in der App 15 Sprachen zur festen Wahl
Qualitätsstufen	Basic (Standard der API), Premium, Ultra (Standard der App)
Ausgabe per API	output_format json, srt und vtt ab Premium, rttm in Ultra
Sprechertrennung	in Ultra, optional min_speakers und max_speakers von 1 bis 20
REST-API	POST /job/add/paperoffice_voice___stt mit audio_file
MCP	po_stt_transcribe auf /mcp-full

03 · IM EINSATZ

Meeting, Diktat, Schulungsvideo. Jeweils als fertiger Text.

**Protokoll aus dem
Projektmeeting**

Die Projektleitung lädt die Aufnahme in Ultra hoch, benennt Sprecher 1 und 2 mit Namen und kopiert den Dialog ins Protokoll.

Diktat aus dem Außendienst

Ein Gutachter diktiert unterwegs ins Handy und lädt die M4A-Datei im Büro hoch. Basic liefert den Fließtext für den Bericht, der Fachwortschatz kennt die Produktnamen.

**Untertitel für ein
Schulungsvideo**

Die Entwicklung sendet das Video per API mit `quality=premium` und `output_format=srt` und erhält die Untertiteldatei mit Zeitmarken zurück.

HÄUFIGE FRAGEN

Welche Sprachen erkennt Voice STT?

Die Sprache wird automatisch erkannt. Per API geben Sie sie optional aus 100 Sprachcodes vor, in der App stehen 15 Sprachen zur festen Auswahl bereit.

Kann Voice STT Sprecher unterscheiden?

Ja, in Ultra. Das Transkript erscheint als Dialog, die Sprecher lassen sich umbenennen. Per API erhalten Sie dazu Sprecherzahl, Überlappungen und RTTM.

Bekomme ich Untertitel als SRT oder VTT?

Ja, per API ab Stufe Premium mit `output_format=srt` oder `vtt`. In der App sichern Sie das Transkript als TXT-Datei.

Wie wird abgerechnet?

Nach Verbrauch, abhängig von der Qualitätsstufe. Die aktuellen Preise stehen auf der Preisseite.

JETZT STARTEN

Lassen Sie Ihre nächste Aufnahme mitschreiben.

Melden Sie sich in Ihrem PaperOffice-Konto an, öffnen Sie **Media AI › Voice Studio** und wählen Sie **Sprache zu Text (STT)**. Datei ablegen, Stufe wählen, Transkription starten.

paperoffice.ai/de/ki-voice-stt-speech-to-text-spracherkennung-transkription-stt



MEHR ZUM MODUL

paperoffice.ai/de/ki-voice-stt-speech-to-text-spracherkennung-transkription-stt

PREISE

paperoffice.ai/de/preise

KONTAKT

paperoffice.ai/de/kontakt