



데이터시트 · 캡처 · 음성 STT

회의 내용을 한 마디도 빠짐없이 직접 타이핑하시나요? 이제 모든 녹음이 텍스트로 돌아옵니다.

음성 STT는 회의, 통화, 받아쓰기 녹음을 텍스트로 바꿉니다. 언어를 스스로 인식하며, Ultra 단계에서는 화자를 구분합니다. 사무실, 영업, 프로젝트 팀을 위해 앱과 API에서 사용할 수 있습니다.

한눈에 보기

최대 50 MB의 오디오 및 영상 파일, 자동 언어 인식, 세 가지 품질 단계, Ultra의 화자 구분, API에서 단어 타임스탬프, SRT, VTT를 제공합니다.

01 · 개요

회의가 먼저, 타이핑은 나중에? 이제 텍스트가 저절로 도착합니다.

프로젝트 회의록을 오늘 보내야 하는데 녹음은 한 시간짜리입니다. 음성 STT에 파일을 업로드하면 텍스트를 돌려받습니다. Ultra에서는 참석자별로 나뉘어 있어 바로 복사하거나 번역할 수 있습니다.

입력

오디오(MP3, WAV, OGG, FLAC, M4A)와 영상(MP4, WebM, AVI, MOV), 최대 50 MB

언어

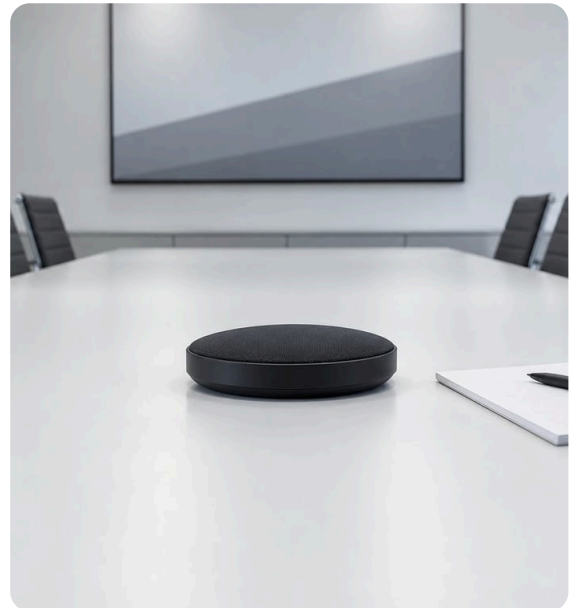
자동 인식; API에서는 언어 코드 100개, 앱에서는 15개 언어 중 선택

출력

일반 텍스트 녹취록, Ultra에서는 화자별 대화 형식; TXT로 내려받기

접근

Media AI > 음성 스튜디오
아래의 앱, REST API, MCP



언어 자동 인식

음성 STT는 녹음의 언어를 스스로 인식합니다. 언어를 알고 있다면 API에서 언어 코드 100개 중 설정할 수 있습니다.

세 가지 품질 단계

Basic은 빠른 일반 텍스트를, Premium은 단어 타임스탬프가 있는 구간을 추가하고, Ultra는 화자까지 구분합니다. 단계는 전사 품질 에서 선택합니다.

화자 구분

Ultra에서 음성 STT는 참석자를 인식하고 녹취록을 대화 형식으로 보여 줍니다. 화자 이름은 클릭 한 번으로 바꿀 수 있습니다.

전문 용어 사전

전문 용어집 을 사용하면 음성 STT가 지식 베이스와 가격표의 전문 용어를 더 정확하게 인식합니다. API에서는 Premium부터 직접 핫워드를 추가할 수 있습니다.

API로 자막 만들기

API에서 Premium 이상은 SRT, VTT와 단어 타임스탬프를 반환하며, Ultra는 화자 분석용 RTTM을 추가합니다.

Translate Studio로 이어짐

Translate Studio에서 사용 을 누르면 녹취록이 복사하거나 붙여넣을 필요 없이 바로 Translate Studio로 넘어갑니다.

02 · 작동 방식

파일을 끌어다 놓고 단계를 고르면, 녹취록을 활용하십시오.

Voice Studio에서는 네 단계로 진행합니다. REST API와 MCP를 통해 자체 프로세스에서 오디오 파일을 바로 보낼 수 있습니다.

1

Voice Studio 열기

Media AI > 음성 스튜디오 에서 음성 텍스트 변환 (STT) 탭을 선택하십시오. 시작 화면에서는 빠른 도구 > 미디어 > 음성 텍스트 변환 로도 이동할 수 있습니다.

2

파일 끌어다 놓기

오디오 또는 영상 파일을 입력란으로 끌어 놓으십시오. 언어는 자동으로 두거나 직접 고르십시오.

3

단계와 옵션 고르기

전사 품질 에서 Basic, Premium 또는 Ultra를 선택하십시오. Ultra에서는 대화 상대 인식 에서 화자 수를 설정할 수 있습니다.

4

녹취록 활용하기

복사하거나, TXT로 다운로드 로 텍스트 파일로 저장하거나, 한 번의 클릭으로 Translate Studio에 넘기십시오.

기술 데이터

기능	세부 정보
입력 형식	MP3, WAV, OGG, FLAC, M4A, MP4, WebM, AVI, MOV; 앱에서는 최대 50 MB, 실행당 파일 하나
언어	자동 인식; API에서는 언어 코드 100개, 앱에서는 고정 선택 15개
품질 단계	Basic(API 기본값), Premium, Ultra(앱 기본값)
API 출력	output_format json, srt, vtt는 Premium부터, rttm은 Ultra에서
화자 구분	Ultra에서 선택 사항으로 min_speakers와 max_speakers를 1부터 20까지 설정
REST API	audio_file과 함께 POST /job/add/paperoffice_voice___stt
MCP	/mcp-full의 po_stt_transcribe

03 · 실제 활용

회의, 받아쓰기, 교육 영상. 각각 완성된 텍스트로.

프로젝트 회의록

프로젝트 책임자가 Ultra에서 녹음을 업로드하고, 화자 1과 2의 이름을 실제 이름으로 바꾼 뒤 대화 내용을 회의록에 복사합니다.

현장에서 받아쓰기

측량 담당자가 이동 중에 휴대폰으로 받아쓰고, 사무실에서 M4A 파일을 업로드합니다. Basic은 보고서용 일반 텍스트를 반환하고, 어휘 사전은 제품명을 알고 있습니다.

교육 영상 자막

개발팀이 quality=premium과 output_format=srt를 지정해 영상을 API로 보내고, 타임스탬프가 있는 자막 파일을 돌려받습니다.

자주 묻는 질문

음성 STT는 어떤 언어를 인식하나요?

언어는 자동으로 인식됩니다. API에서는 언어 코드 100개 중 설정할 수 있으며, 앱에서는 15개 언어를 고정 선택지로 제공합니다.

음성 STT가 화자를 구분할 수 있나요?

예, Ultra에서 가능합니다. 녹취록이 대화 형식으로 표시되며 화자 이름을 바꿀 수 있습니다. API에서는 화자 수, 겹치는 발화, RTTM도 받을 수 있습니다.

자막을 SRT나 VTT로 받을 수 있나요?

예, Premium 단계부터 API에서 output_format=srt 또는 vtt로 받을 수 있습니다. 앱에서는 녹취록을 TXT 파일로 저장합니다.

요금은 어떻게 청구되나요?

사용량 기준이며 품질 단계에 따라 다릅니다. 현재 요금은 가격 페이지에서 확인하실 수 있습니다.

시작하기

다음 녹음을 정리된 글로 받아 보십시오.

PaperOffice 계정에 로그인하고 **Media AI** > **음성 스튜디오** 을 연 뒤 **음성 텍스트 변환 (STT)** 를 선택하십시오. 파일을 끌어다 놓고, 단계를 고르고, 받아쓰기를 시작하십시오.

paperoffice.ai/ko/ai-voice-stt-eumseong-tekseuteu-byeonhwan-insig-stt



모듈 자세히 보기

paperoffice.ai/ko/ai-voice-stt-eumseong-tekseuteu-byeonhwan-insig-stt

요금

paperoffice.ai/en/pricing

문의

paperoffice.ai/ko/munui