



ТЕХНИЧЕСКОЕ ОПИСАНИЕ · ЗАХВАТ · VOICE STT

Печатать протоколы совещаний слово в слово? Теперь каждая запись возвращается текстом.

Voice STT превращает записи совещаний, звонков и диктовок в текст. Язык он определяет сам, а уровень Ultra разделяет говорящих. Создан для офисных, торговых и проектных команд, в приложении и через API.

КОРОТКО

Аудио- и видеофайлы до 50 МБ, автоматическое определение языка, три уровня качества, разделение говорящих в Ultra, а также временные метки слов, SRT и VTT через API.

01 · О ЧЁМ РЕЧЬ

Сначала совещание, потом печать? Теперь текст приходит сам.

Протокол проектного совещания нужно разослать сегодня, а запись длится час. С Voice STT вы загружаете файл и получаете текст обратно: в Ultra он разделён по участникам, его можно скопировать или перевести.

ВВОД

Аудио (MP3, WAV, OGG, FLAC, M4A) и видео (MP4, WebM, AVI, MOV) до 50 МБ

ЯЗЫКИ

Определяются автоматически; 100 кодов языков через API, 15 языков для выбора в приложении

РЕЗУЛЬТАТ

Расшифровка простым текстом, в Ultra — диалог по говорящим; скачивание в TXT

ДОСТУП

Приложение в разделе
 Media AI › Студия голоса
 , REST API и MCP



Язык определяется автоматически

Voice STT определяет язык записи сам. Если вы его знаете, его можно задать через API из 100 кодов языков.

Три уровня качества

Basic быстро возвращает простой текст, Premium добавляет сегменты с временными метками слов, Ultra также разделяет говорящих. Уровень указан в разделе [Качество транскрипции](#).

Разделение говорящих

В Ultra Voice STT определяет участников и показывает расшифровку как диалог. Переименовать говорящих можно одним кликом.

Словарь

[Специализированная лексика](#) повышает точность распознавания технических терминов из ваших баз знаний и прайс-листов. Через API свои ключевые слова (hotwords) добавляются начиная с Premium.

Субтитры через API

Через API Premium и выше возвращает SRT и VTT, а также временные метки слов, Ultra дополнительно добавляет RTTM для анализа говорящих.

Дальше в Translate Studio

[Использовать в Translate Studio](#) передаёт расшифровку прямо в Translate Studio, без копирования и вставки.

02 · КАК ЭТО РАБОТАЕТ

Загрузите файл, выберите уровень. Используйте расшифровку.

В Voice Studio это четыре шага. Через REST API и MCP вы отправляете аудиофайлы прямо из своих процессов.

1

Открыть Voice Studio

В разделе [Media AI](#) › [Студия голоса](#) выберите вкладку [Речь в текст \(STT\)](#) . Со стартовой страницы туда ведёт также [Быстрые инструменты](#) › [Media](#) › [Речь в текст](#) .

2

Загрузить файл

Перетащите аудио- или видеофайл в поле. Оставьте язык на «Авто» или выберите его вручную.

3

Выбрать уровень и параметры

В разделе [Качество транскрипции](#) выберите Basic, Premium или Ultra. В Ultra количество говорящих можно задать в разделе [Определение собеседников](#) .

4

Использовать расшифровку

Скопируйте её, сохраните как текстовый файл ([Скачать как TXT](#)) или передайте в Translate Studio одним кликом.

ТЕХНИЧЕСКИЕ ДАННЫЕ

ФУНКЦИЯ	ПОДРОБНОСТИ
Форматы ввода	MP3, WAV, OGG, FLAC, M4A, MP4, WebM, AVI, MOV; в приложении до 50 МБ, один файл за запуск
Языки	автоматическое определение; 100 кодов языков через API, 15 фиксированных вариантов в приложении
Уровни качества	Basic (по умолчанию в API), Premium, Ultra (по умолчанию в приложении)
Вывод через API	output_format: json, srt и vtt начиная с Premium, rttm в Ultra
Разделение говорящих	в Ultra, необязательные min_speakers и max_speakers от 1 до 20
REST API	POST /job/add/paperoffice_voice___stt с audio_file
MCP	po_stt_transcribe на /mcp-full

03 · НА ПРАКТИКЕ

Совещание, диктовка, учебное видео. Каждое — готовым текстом.

Протокол проектного совещания

Руководитель проекта загружает запись в Ultra, переименовывает Говорящего 1 и Говорящего 2 в имена участников и копирует диалог в протокол.

Диктовка из поездки

Геодезист надиктовывает текст на телефон по дороге и загружает файл M4A в офисе. Basic возвращает простой текст для отчёта, а словарь знает названия продуктов.

Субтитры для учебного видео

Разработка отправляет видео через API с quality=premium и output_format=srt и получает обратно файл субтитров с временными метками.

ЧАСТО ЗАДАВАЕМЫЕ ВОПРОСЫ

Какие языки распознаёт Voice STT?

Язык определяется автоматически. Через API его можно задать из 100 кодов языков, а в приложении предлагаются 15 языков для фиксированного выбора.

Может ли Voice STT различать говорящих?

Да, в Ultra. Расшифровка выглядит как диалог, говорящих можно переименовать. Через API вы также получаете число говорящих, перекрытия и RTTM.

Можно ли получить субтитры в SRT или VTT?

Да, через API начиная с уровня Premium с output_format=srt или vtt. В приложении расшифровку можно сохранить как файл TXT.

Как это тарифицируется?

По использованию, в зависимости от уровня качества. Актуальные цены указаны на странице с тарифами.

НАЧАТЬ РАБОТУ

Пусть следующую запись расшифруют за вас.

Войдите в свой аккаунт PaperOffice, откройте **Media AI** > **Студия голоса** и выберите **Речь в текст (STT)**. Загрузите файл, выберите уровень и запустите расшифровку.

paperoffice.ai/ru/ai-voice-stt-rech-v-tekst-transkriptsiya-raspoznavanie-stt



ПОДРОБНЕЕ О МОДУЛЕ

paperoffice.ai/ru/ai-voice-stt-rech-v-tekst-transkriptsiya-raspoznavanie-stt

ЦЕНЫ

paperoffice.ai/ru/tseny

КОНТАКТЫ

paperoffice.ai/ru/kontakt