

# 7 лучших API преобразования речи в текст для повышения доступности

10.05.2024

Технология преобразования речи в текст переживает бум и получает все более широкое распространение. Причиной может быть значительный прогресс в области распознавания речи для повышения точности, доступности и дешевизны. По данным исследования, 79 % респондентов назвали экономию времени одним из преимуществ использования решений для преобразования речи в текст. В 2020 году объем мирового рынка распознавания речи составит около 10 миллиардов долларов США. Сегодня организации и частные лица производят больше контента, используют голосовые команды для управления приложениями и устройствами, применяют чат-боты. Именно в этом им могут помочь речетекстовые API, которые помимо диктовки и перевода позволяют создавать письменный текст.

## Лучшие API для преобразования речи в текст в 2024 году

### Rev

Расшифровка и распознавание речи в реальном времени с помощью Rev API. Он позволяет транслировать речь в текст в прямом эфире для создания субтитров. Он обслуживает многие отрасли:

- **Средства массовой информации и развлечения:** Повышает доступность транслируемого контента или прямых трансляций в Интернете.
- **Образование:** Повышает доступность вебинаров, мероприятий

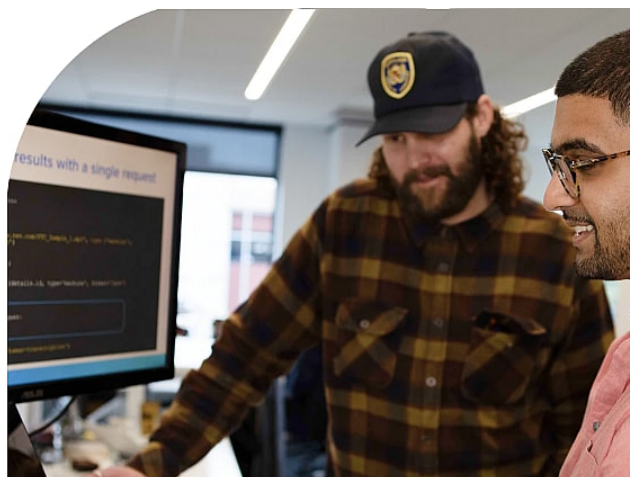
и лекций.

- **Центры обработки вызовов и аналитика:** Обучение торговых агентов и расшифровка звонков.
- Она также обслуживает другие отрасли, транскрибируя **тренинги, мероприятия и встречи** в режиме реального времени.

## Easily integrate Rev's Speech-to-text APIs into your product or workflow.

Start building

Contact Sales →

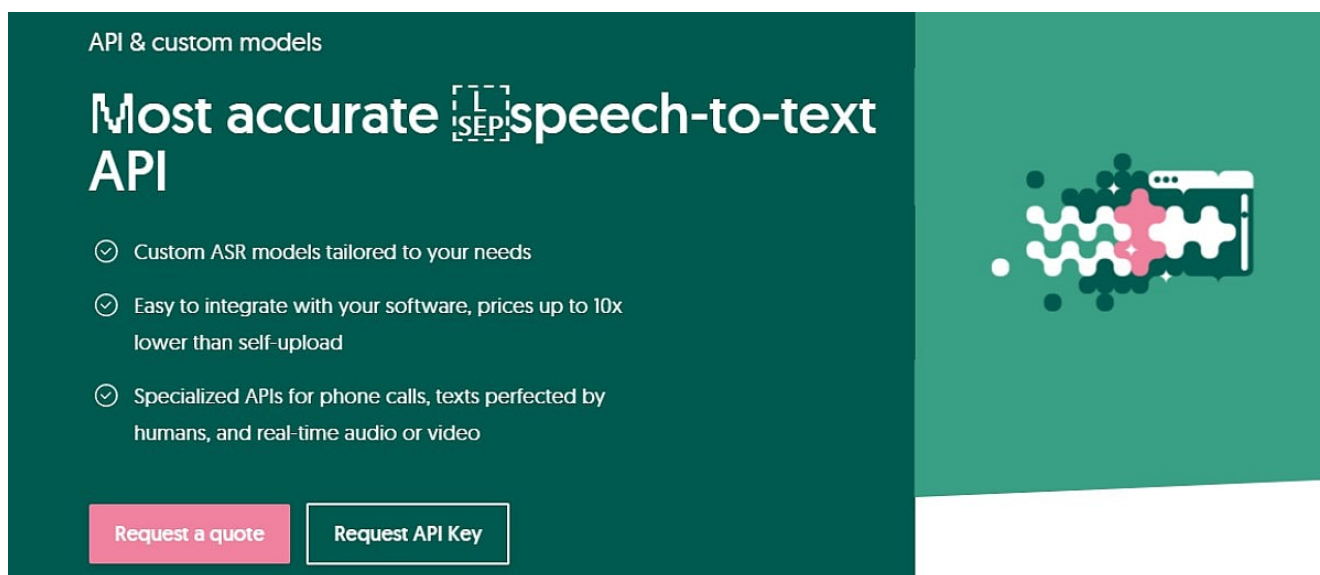


Rev охватывает почти все основные английские языки по всему миру и обеспечивает наилучший результат вне контекста, независимо от того, кто говорит. Он создает субтитры в режиме реального времени с минимальной задержкой и использует естественные языки для создания высокоточной, учитывающей контекст, полностью пунктуационной и читабельной транскрипции. Вы можете использовать отраслевые названия, терминологию и многое другое для повышения точности расшифровки. Кроме того, программа отфильтровывает около 600 оскорбительных слов из титров и позволяет отслеживать время начала и окончания каждого слова. Легко внедрите решения для преобразования речи в текст в своих приложениях и устраняйте барьеры в общении.

## Amberscript

Получите самый точный и один из лучших на рынке API для преобразования речи в текст – Amberscript. Он предоставляет пользовательские модели ASR в соответствии с вашими потребностями и позволяет легко интегрировать их в ваше программное обеспечение для воспроизведения в реальном времени

аудио- и видеофайлов, текстов, отточенных человеком, и телефонных звонков. Автоматизируйте рабочие процессы и расшифровывайте широкий спектр видео и аудио с помощью API преобразования речи в текст Amberscript. Он передает файлы на сервер ASR и возвращает их в выбранном вами формате. Он доступен на 80+ языках и поддерживает автоматическую пунктуацию, метки диктора, автоматическую обводку, временные метки, двухканальный звук и другие форматы видео/аудио файлов.



The image is a screenshot of the Amberscript website. The background is a dark teal color. At the top left, it says 'API & custom models'. The main heading is 'Most accurate [SEP] speech-to-text API', where '[SEP]' is a small icon of a speech bubble with a red cross. Below the heading are three bullet points, each with a checkmark icon: 'Custom ASR models tailored to your needs', 'Easy to integrate with your software, prices up to 10x lower than self-upload', and 'Specialized APIs for phone calls, texts perfected by humans, and real-time audio or video'. At the bottom left, there are two buttons: a pink one that says 'Request a quote' and a white one with a black border that says 'Request API Key'. On the right side of the screenshot, there is a graphic of a smartphone with a speech bubble and a red cross, surrounded by white and pink dots.

В формат XML/JSON можно включить такую информацию, как время начала и конца слова, указания на вопросы, баллы уверенности, пунктуацию и т. д. Amberscript делает аудиозапись доступной в формате .doc/.txt, экспортируемой с изменениями/без изменений диктора и временных меток. Amberscript поддерживает такие форматы, как EBU-STL и VTT, чтобы помочь с автоматическими субтитрами. Вы также можете индивидуально определять настройки внешнего вида субтитров. Он сочетает в себе новейшие научные, языковые и технологические знания для разработки моделей, ориентированных на конкретного пользователя, для различных случаев использования. После настройки он улучшает распознавание речи для:

- Акустические среды
- Разные акценты
- Адаптация словарного запаса для распознавания специальных терминов, названий продуктов и аббревиатур

- Адаптация к языкам, специфичным для конкретных областей, таких как здравоохранение, технологии, физика, политика и т. д.

Попробуйте Amberscript бесплатно. Получите больше преимуществ, заплатив \$10 за один час загрузки видео или аудио.

## Speech-to-Text от Google Cloud

Используйте мощный API для точного преобразования речи в текст с помощью решения Speech-to-Text от Google Cloud. Оно обеспечивает превосходный пользовательский опыт, транскрибируя вашу речь с точными подписями. Кроме того, оно помогает улучшить качество обслуживания благодаря информации, полученной и расшифрованной в ходе взаимодействия с клиентами. Вы можете применять передовые алгоритмы нейронных сетей Google с глубоким обучением для автоматического распознавания речи. В нем также предусмотрена функция настройки модели, позволяющая экспериментировать, управлять и создавать пользовательские ресурсы. Кроме того, вы можете гибко развернуть систему распознавания речи в облаке или в локальной сети. Передовая технология Google Cloud помогает распознавать специфические термины с помощью подсказок. Она автоматически преобразует произнесенные числа в годы, валюты, адреса и другие классы. Вы даже можете выбрать одну из моделей для конкретного домена, чтобы получить особые требования к качеству в зависимости от сервиса. Кроме того, решение Google Cloud для преобразования речи в текст предоставляет простой в использовании пользовательский интерфейс, позволяющий экспериментировать с аудиозаписями и пробовать различные конфигурации для достижения точности и качества. Кроме того, вы можете использовать решение для преобразования речи в текст в своих частных центрах обработки данных, чтобы иметь полный контроль над инфраструктурой и речевыми данными. Они предлагают бесплатный 60-минутный уровень. После этого вы будете платить за 15 секунд аудио. Сделайте следующий шаг прямо сейчас и

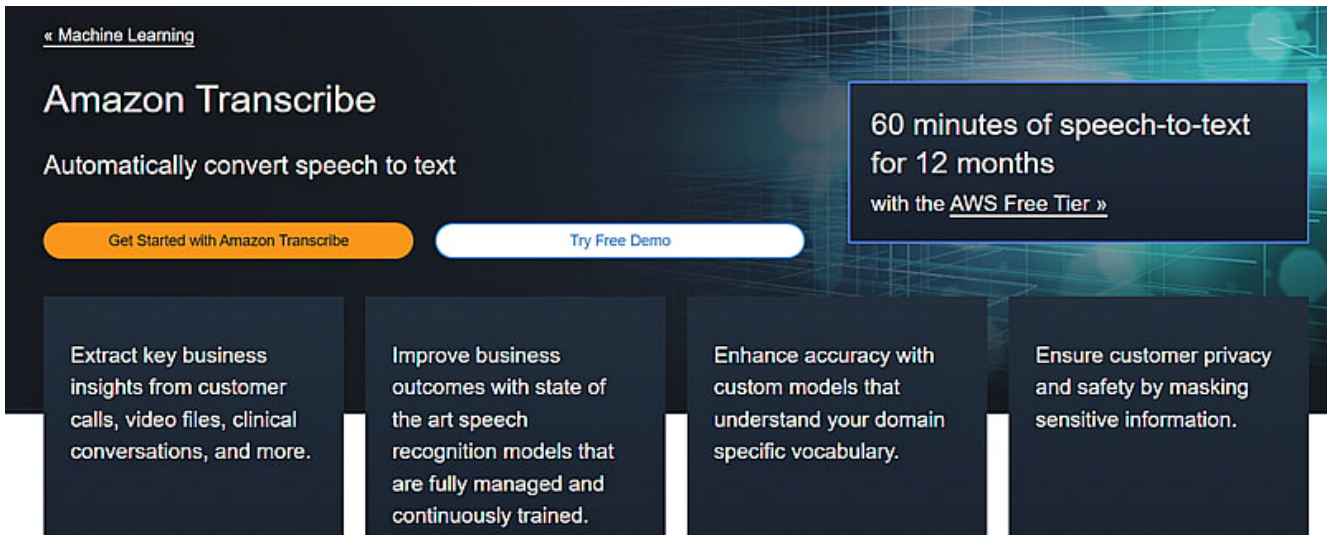
попробуйте все функции бесплатно.

## AssemblyAI

API-интерфейсы преобразования речи в текст AssemblyAI помогают автоматически конвертировать аудио- и видеофайлы и аудиопотоки в текст и правильно его понимать. Новейшие модели искусственного интеллекта обеспечивают преобразование речи в текст AssemblyAI, а его аудиоинтеллект позволяет определять темы, модерировать контент и обобщать его. Интегрируйте простой API в свои системы за считанные минуты и воспринимайте аудио без ошибок. Вы можете создавать надежные приложения с такими функциями, как обнаружение сущностей, редактирование PII, анализ настроения и многое другое. Кроме того, вы сможете автоматически расшифровывать видео- и аудиофайлы с высочайшей точностью и извлекать из данных важные сведения, включая настроения, конфиденциальный контент, темы и многое другое. Он предлагает только модель ценообразования с оплатой по мере роста. Стоимость основной транскрипции составляет \$0,00025/секунду, а аудиоинформации – \$0,000167/секунду. Начните прямо сейчас бесплатно и воспользуйтесь передовой технологией.

## Amazon Transcribe

Amazon Transcribe – это сервис автоматического распознавания речи (ASR), позволяющий разработчикам легко добавлять в свои приложения функцию преобразования речи в текст. Этот прорывной инструмент использует потенциал передовых алгоритмов машинного обучения, предлагая транскрипцию с непревзойденным уровнем точности на 100 и более языках, а также чрезвычайно гибкую модель ценообразования с оплатой по факту.



Кроме того, у него есть индивидуальные версии, разрабатывающие Amazon Transcribe Medical строго по оптимальным стандартам, соответствующим требованиям медицинской транскрипции. Благодаря повышенной конфиденциальности данных пациентов, а также безопасности в режиме реального времени и соответствию требованиям HIPAA, этот сервис остается идеальным решением для расшифровки аудиозаписей.

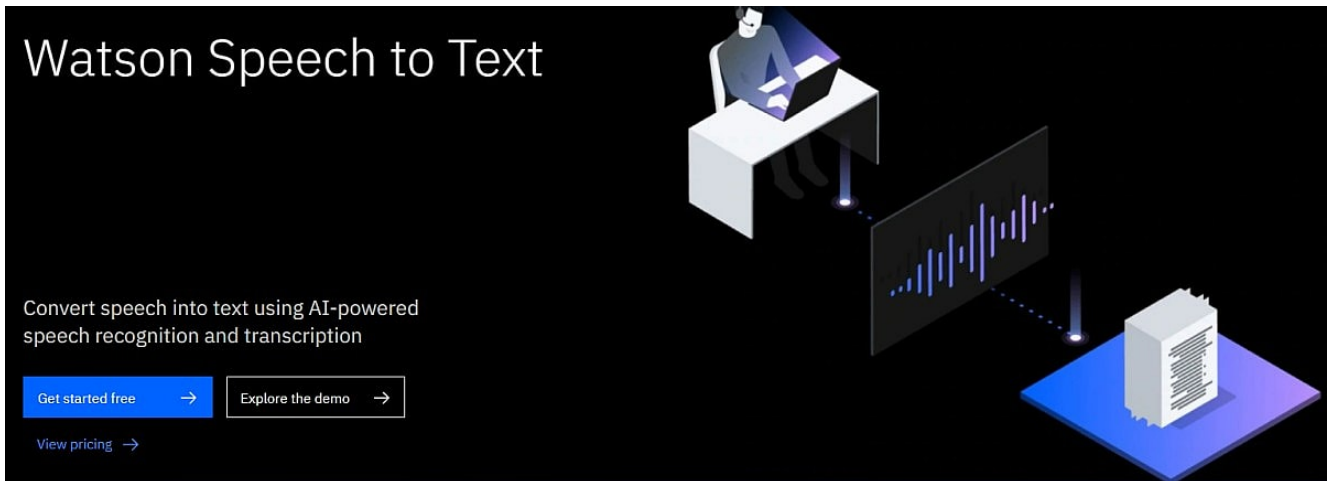
□ Плюсы	□ Минусы
На основе генеративного искусственного интеллекта	Это может быть дорого для крупных проектов
Многоязычная поддержка	Ограниченная настройка
Транскрипция в режиме реального времени	
Аналитика звонков	

**Сколько стоит Amazon Transcribe?** При использовании Amazon Transcribe вы платите по мере необходимости, в зависимости от количества секунд расшифровки аудиозаписей в месяц. Бесплатный уровень предлагает до 60 минут в месяц в течение первого года после регистрации.

## IBM Watson

IBM Watson Speech to Text предлагает решения для транскрипции и распознавания речи на основе искусственного интеллекта. Они

обеспечивают точное и быстрое распознавание речи на разных языках для различных сфер применения, таких как самообслуживание клиентов, речевая аналитика, помощь агентам и т. д.



Подобно человеку, он внимательно слушает разговор, транскрибирует аудиозапись, получает релевантный контент и точно выдает идеальный ответ. Вы можете обучить Watson предпочтительному языку домена и характеристикам аудио и развернуть решение для преобразования речи в текст на любой облачной платформе, включая частную, гибридную, публичную, многоцветную или локальную. Интегрируйте решение с вашими приложениями, чтобы постоянно получать точные результаты. Вы также можете использовать решение для акустических и языковых тренировок. Вы получите предварительно обученные речевые модели, обучение модели, тонкую настройку функций, низкую задержку, аудиодиагностику, промежуточную транскрипцию, интеллектуальное форматирование, фильтрацию слов и выделение пятен. Начните преобразовывать речь в текст бесплатно в течение 500 минут в месяц. Платите 0,01 доллара в минуту, чтобы настроить речевые модели и повысить точность.

## Scriptix

Scriptix предлагает облачный сервис преобразования речи в текст, а его специализированные модели генерируют наилучшие результаты для вашего контента. Он поможет вам превратить голосовые данные в текст для облегчения доступа, анализа и

поиска информации. Правительства, телекоммуникационные компании, СМИ и здравоохранение используют транскрипцию для улучшения цифрового присутствия.

## Speech to text models *built by you.*

Cloud based speech to text service. Process in real-time or batch mode. We customize our models to generate the best results for your content.

GET STARTED →

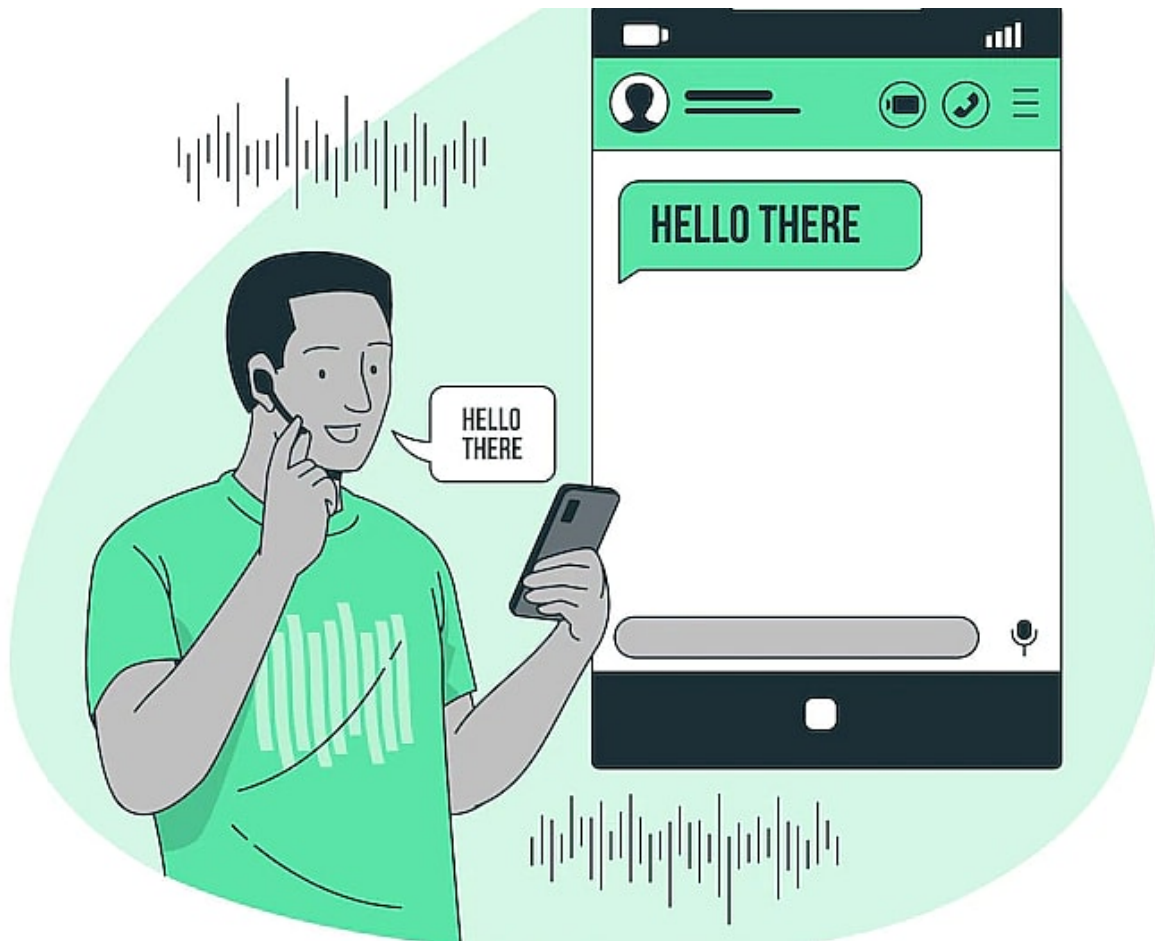


Независимо от того, нужен ли он вам для небольших объемов транскрипции или субтитров, Scriptix обладает множеством преимуществ. Вы получите баллы доверия, временные метки, обработку в реальном времени, пунктуацию, многоканальную обработку, поддержку различных файлов и многое другое. Он доступен на тринадцати языках, включая арабский, английский, французский, итальянский, шведский, немецкий, голландский, датский, фламандский, норвежский и другие. Интегрируйте API speech-to-text в свои приложения и испытайте лучшее.

## Что такое API преобразования речи в текст?

Speech-to-text или распознавание речи – это технология транскрибирования устных слов или аудиоконтента в текст. Для этого используются приложения, API, инструменты и другие программные решения. Он использует машинное обучение и искусственный интеллект для обнаружения закономерностей в звуковых волнах и точной транскрипции.

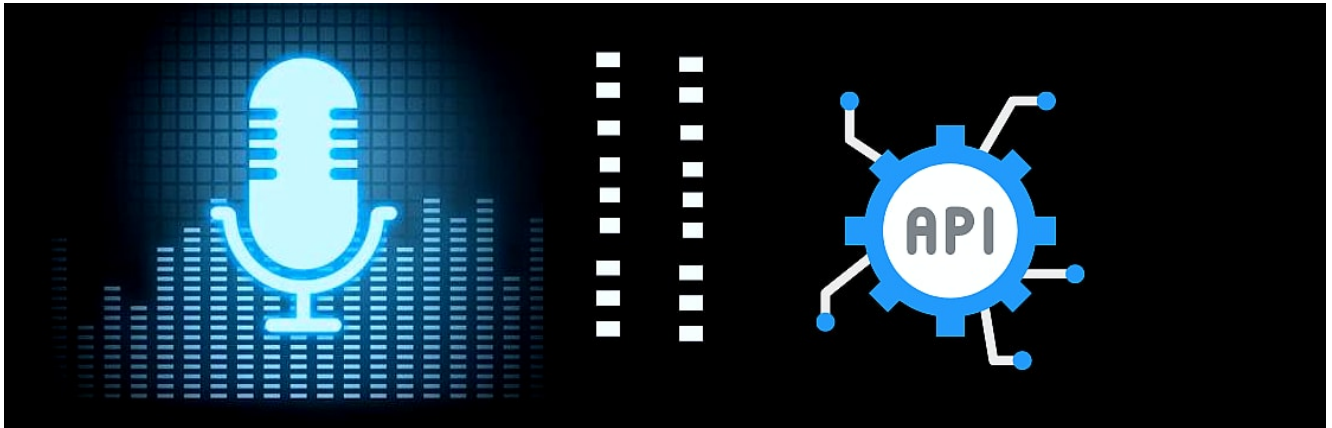




Некоторые особенности рече-текстовых API:

- Поддержка нескольких языков, кроме английского
- Принимайте различные аудиосигналы, включая файлы, хранящиеся на компьютере и в облаке, микрофоны и т. д.
- Обнаружение параграфов
- Этикетки для динамиков
- Пользовательский словарь
- Обнаружение темы
- Автоматическая обводка и пунктуация
- Фильтрация ненормативной лексики и многое другое

**Зачем использовать API преобразования речи в текст?**



Рече-текстовые API имеют массу преимуществ.

## **Повышает производительность и эффективность**

Ручной набор длинных текстов для статей, документации, презентаций и т. д. отнимает много сил. Вместо этого вы можете использовать API преобразования речи в текст, чтобы надиктовать свои слова и получить их в виде текста. Это облегчит вашу работу и ускорит рабочий процесс, давая при этом необходимый отдых рукам.

## **Надежный**

Использование хорошего API преобразования речи в текст обеспечивает превосходную точность. Таким образом, вы можете положиться на эти решения для создания документов и бумаг с более быстрым временем выполнения и меньшим количеством ошибок. Это также поможет вам работать в режиме многозадачности. Поэтому всегда выбирайте высокоточный API преобразования речи в текст, например Rev, который обеспечивает точность 84 %.

## **Экономия времени**

Написание тяжелого текста вручную требует не только усилий, но и времени. Как известно, говорить быстрее, чем писать, поэтому использование API speech-to-text значительно сэкономит ваше время. Это также очень полезно для профессионалов, чья

скорость письма низкая или средняя. Таким образом, вы сможете быстрее отправлять свои работы и посвящать сэкономленное время другим продуктивным занятиям.

## **Помогает людям с ограниченными физическими возможностями**

Люди с определенными физическими недостатками, такими как дислексия, травмы и т. д., могут столкнуться с трудностями при использовании обычных устройств и форматов ввода, таких как клавиатура. Использование рече-текстовых API поможет им вводить слова голосом, не набирая их вручную. Это облегчит их трудности и повысит производительность труда.

## **Где используются речевые API?**



Речевые API очень помогают во многих сценариях. Вот некоторые из них:

### **Автоматизированная диктовка**

Если вы создатель контента, писатель или любой другой человек, которому нужно набирать длинный текст, вам поможет API speech-to-text. Вместо того чтобы набирать каждое слово вручную, вы

можете использовать API, чтобы надиктовать свои слова, и он создаст для вас письменный текст.

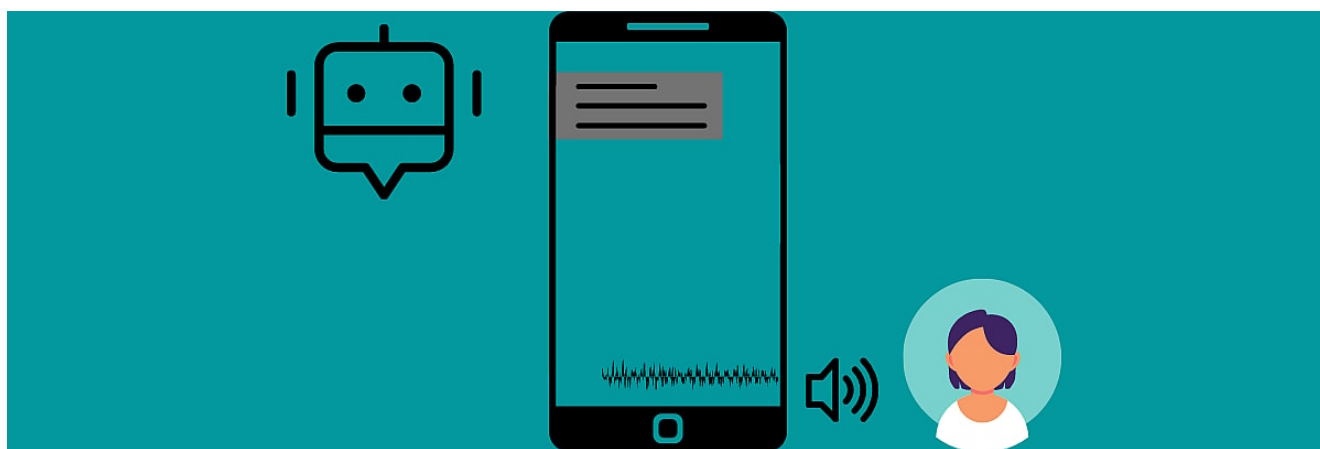
## Голосовые команды

Вы можете запускать некоторые действия голосом, используя API преобразования речи в текст. Например, ввод запросов голосом и выбор пункта меню.

## Умный помощник

Рече-текстовые API используются в умных помощниках, таких как Alexa, Siri и т. д., для управления бытовой техникой, веб-приложениями, автомобилями и т. д. Это позволит создать командно-административный или естественный интерфейс для поисковых запросов.

## Чатботы



Чат-боты активно используются на веб-сайтах и в приложениях для помощи посетителям и пользователям в решении их вопросов. Поэтому, если вы создаете приложение для чат-ботов, вы можете использовать API преобразования речи в текст, чтобы пользователи могли задавать вопросы голосом, взаимодействуя с ботами.

## Перевод

Рече-текстовые API оснащены функциями голосового перевода и

поддержки нескольких языков, чтобы помочь пользователям общаться с другими пользователями, говорящими на разных языках. Многие рече-текстовые API поддерживают широкий спектр глобальных языков, что обеспечивает беспрепятственную глобальную коммуникацию.

## **Обнаружение смешанных языков**

Даже если вы используете несколько языков при диктовке, с помощью API преобразования речи в текст вы сможете легко создавать документы. Многие из них могут распознавать смешанные языки, автоматически определяя разговорные языки и транскрибируя слова должным образом, не требуя, чтобы вы говорили только на одном языке во время транскрибирования.

## **Транскрипция для колл-центров**

Call-центрам может потребоваться запись разговоров между агентами и конечными пользователями в процессе поддержки клиентов, продаж и т. д. Это может понадобиться для проведения аудита или контроля качества. Поэтому, если вам нужна помощь в этом вопросе, API speech-to-text могут помочь, отправляя аудиозаписи в пакетном режиме для расшифровки.

## **Заключение**

Использование рече-текстовых API – разумный и экономически эффективный выбор по сравнению с созданием собственной системы транскрипции. Хорошо то, что большинство из перечисленных API не стоят целого состояния, так что попробуйте и посмотрите, что подойдет для вашего случая.