MTS AI

Audiogram

Платформа распознавания
и синтеза речи

AUDIOGRAM — ПЛАТФОРМА РАСПОЗНАВАНИЯ И СИНТЕЗА РЕЧИ НА БАЗЕ НЕЙРОННЫХ СЕТЕЙ И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ.

Audiogram создан для разработчиков сервисов и услуг, а также для конечных клиентов, которым требуется распознавание больших объёмов аудиофайлов и генерация звукового контента: телеком-операторам, банкам, СМИ и другим компаниям.

Audiogram позволяет в реальном времени и в офлайн-режиме автоматически преобразовывать речь в текст, и наоборот, озвучивать текст выбранным голосом, с определенной интонацией и ударениями.

На базе платформы, разработанной MTS AI, можно создавать различные сервисы для работы с речью. В том числе:

- распознавание речи и озвучивание голосовых помощников;
- создание аудиокниг;
- автоматическая генерация субтитров для видео;
- транскрибация аудио в текст;
- запись аудиосообщений синтезированным голосом.

Audiogram может поставляться в качестве программного обеспечения или как облачный сервис. Платформа легко интегрируется с другими решениями MTS AI, в том числе с речевой аналитикой.

1. ОБЗОР РЫНКА ТЕХНОЛОГИЙ РАСПОЗНАВАНИЯ И СИНТЕЗА РЕЧИ

Тенденции рынка технологий распознавания и синтеза речи

Во всем мире растет потребность в программном обеспечении, способном понимать и воспроизводить человеческий голос, а также общаться с пользователями.

Драйверами развития подобных технологий стали следующие факторы:

- спрос на удаленное обслуживание клиентов в ритейле, медицине, телекоме и других отраслях;
- стремление бизнеса повысить эффективность коммуникаций с клиентами и увеличить скорость обработки запросов аудитории;
- внедрение технологии распознавания речи в потребительские товары: смартфоны, ноутбуки, планшеты и устройства для умного дома, и увеличение количества устройств с голосовым управлением;
- растущая потребность в голосовой аутентификации в приложениях и устройствах.

Объем мирового рынка распознавания речи и голоса вырастет с $9,4 млрд США в 2022 году до $28,1 млрд США к 2027 году. Среднегодовой рост составит в среднем 24,4%.

$9,4 млрд

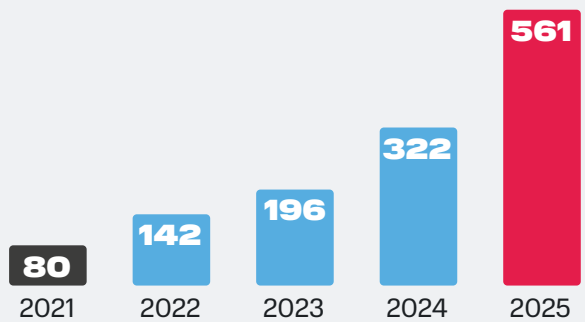
Глобальный рынок распознавания речи в 2022 году

$28,1 млрд

Глобальный рынок распознавания речи в 2027 году

В России рынок разговорного ИИ, в том числе технологий распознавания и синтеза речи тоже идет вверх. И в ближайшие годы эта тенденция сохранится.

Объем рынка разговорного AI в России: 2021-2025



Объем рынка в 2025 году

$561 млн
₽41 млрд

Рост рынка с 2021 по 2025 год

600%

Интерес к подобным решениям проявляют представители разных компаний:

- Крупные корпорации с выручкой более $1 млрд запускают пилоты по созданию виртуальных ассистентов и ботов.
- Среднему бизнесу нужны кастомизируемые решения под конкретную потребность.
- Малый бизнес заинтересован в коробочных продуктах, требующих минимальной адаптации, и в сервисной поддержке со стороны партнеров.

Большой объем заказов на решения в области разговорного ИИ поступает от государства — на рынке существует несколько игроков, которые практически полностью специализируются на госзаказах.

Бизнес-эффекты от внедрения продуктов, созданных с помощью аналогов Audiogram:

в 2 раза

увеличение скорости информирования клиентов

16-25%

рост продаж

96%

повышение точности распознавания запросов пользователей

Дополнительные преимущества:

50%

уменьшение нагрузки на контакт-центры

20%

снижение затрат на оплату труда сотрудникам

20%

увеличение повторных продаж

Источник данных: кейсы компаний Наносемантика, Fonemica, исследования Just AI, MarketsandMarkets

2. ПРЕИМУЩЕСТВА AUDIOGRAM

Audiogram – платформа распознавания и синтеза речи на базе нейронных сетей и методов машинного обучения. Она создана для разработчиков сервисов и услуг, а также для конечных клиентов, которым требуется распознавание больших объёмов аудиофайлов и генерация звукового контента: телеком-компаниям, банкам, СМИ и другим.

С помощью Audiogram бизнес-заказчики могут создавать сервисы по преобразованию речи в текст, и, наоборот, текста в речь в реальном времени и в офлайн-режиме, создавать голосовых ботов, генерировать субтитры и озвучивать тексты выбранным голосом с определенной интонацией и другими параметрами.

Мультиотраслевая модель

Уникальная модель распознавания речи может использоваться в любых сферах (в ритейле, телекоме, банках и т.д.) и не требует дополнительного обучения

Дообучение доменных моделей

Возможность адаптировать модель распознавания речи под специфическую лексику бизнес-заказчика (маркетинговые названия, специфические термины) за 14 дней на основе 300 часов аудиозаписей

Продвинутые функции синтеза речи

Создайте голос своего бренда, быстро и качественно озвучивайте художественные книги и ролики с помощью автоматической расстановки ударений и интонаций

Высокое качество работы

Audiogram распознает речь в разных шумовых условиях: платформа поддерживает диалог с клиентами, когда они разговаривают очень тихо или находятся в местах, где есть посторонние звуки

Легкая интеграция с системами заказчика

Платформа поддерживает взаимодействие с внешними приложениями при помощи gRPC API и протоколов UniMRCP и SIP для интеграции с телефонией

Гибкое лицензирование

Оплата возможна по принципу pay as you go за минуту распознавания аудио и за синтез каждого миллиона символов. Также предоставляются пакетные тарифы и доработки системы под домен клиента за отдельную плату.

3. КАК БИЗНЕС МОЖЕТ ИСПОЛЬЗОВАТЬ AUDIOGRAM?

Разработчики голосовых ботов и умных ассистентов:

- озвучивание ответа бота или ассистента синтезированным голосом, неотличимым от человеческого, благодаря чему пользователю приятно и удобно общаться с ботом;
- использование готовой мультиотраслевой модели распознавания речи, с помощью которой бот или ассистент может поддерживать диалог с пользователем на любую тему.

СМИ и другие создатели контента:

- автоматическая генерация субтитров для видео;
- озвучивание статей и видеоматериалов;
- озвучивание навигации по сайту для людей с ослабленным зрением;
- перевод в текст аудио- и видеоматериалов — интервью и конференции с одним или несколькими участниками.

EdTech-компании:

- расшифровка аудио- и видеозаписей лекций;
- создание субтитров к обучающим видео;
- озвучивание роликов для курса;
- озвучивание статей.

Издательства и электронные библиотеки:

- быстрое озвучивание художественной, научно-популярной и другой литературы для создания аудиокниг.

КАК БИЗНЕС МОЖЕТ ИСПОЛЬЗОВАТЬ AUDIOGRAM?

Социальные сети и мессенджеры:

- транскрибация голосовых сообщений;
- перевод текстовых сообщений в звуковые;
- автоматическая генерация субтитров для видео.

Кол-центры:

- распознавание речи пользователей и генерация ответов бота синтезированным голосом;
- оперативное изменение умного голосового меню (IVR) без привлечения диктора;
- транскрибация звонков.

Разработчики ПО и видеоигр:

- внедрение функции распознавания и синтеза речи в приложения и программы для конечных пользователей;
- аудионавигация для пользователей;
- генерация субтитров для видео;
- озвучивание персонажей компьютерных игр.

Транспортные и торговые объекты

(аэропорты, вокзалы, метрополитен, торговые центры, магазины):

- озвучивание рекламных и информационных сообщений для привлечения внимания пассажиров и клиентов;
- генерация аудиоподсказок для комфортной навигации посетителей.

4. КЕЙСЫ

4.1. Внедрение Audiogram в продукт МТС «Защитник»

Заказчик обратился в MTS AI за решением для записи и транскрибации звонков, а также для поддержания диалога со спамерами. Его нужно было интегрировать в сервис по защите клиентов от телефонного спама и нежелательных звонков.

MTS AI предложила использовать для этой цели Audiogram.

Что было сделано?

Audiogram интегрировали с внутренними системами МТС.

- Платформу синтеза и распознавания речи развернули на CPU (центральных процессорах) в контуре заказчика;
- Audiogram подключили к программному обеспечению контакт-центра МТС с помощью протокола gRPC, который упрощает обмен сообщениями между клиентами и внутренними службами.

Функции платформы были кастомизированы.

- Добавлен модуль расстановки знаков препинания, чтобы конечным пользователям было проще воспринимать сообщения;
- Платформу научили переводить числительные в цифровые значения (например, триста двадцать пять, сказанные словами, в 325);
- Реализован сервис «Антимат» по преобразованию нецензурной лексики в специальные символы.

Результат

Благодаря интеграции Audiogram, компания МТС получила для своего продукта «Защитник» необходимые опции:

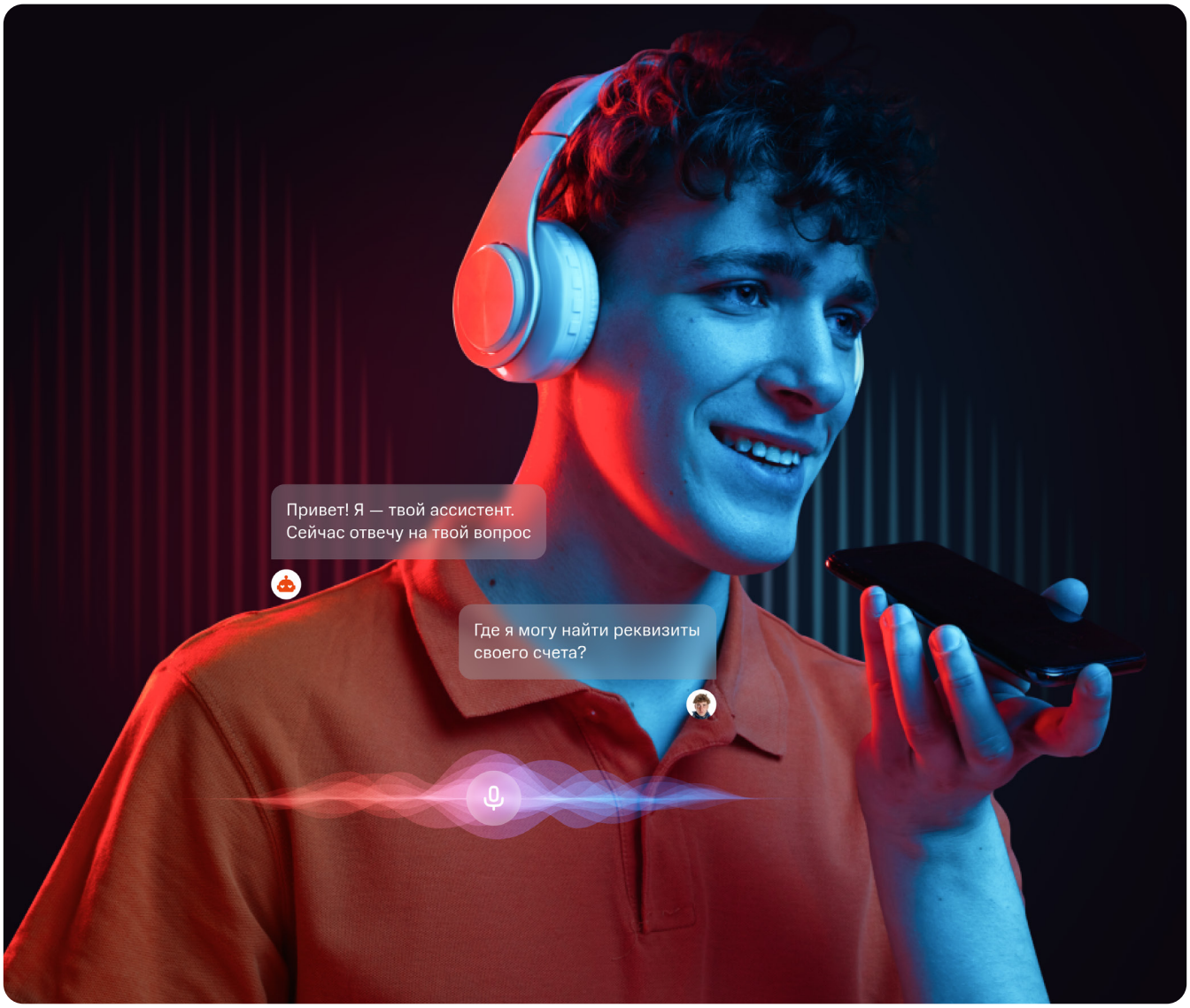
- запись и точная расшифровка спам-звонков;
- дослушивание сообщения до конца.

Клиентам приходят смс-сообщения с текстом разговора и указанием категории спама, при этом абоненты МТС могут быть уверены, что важная информация не потеряется, поскольку они получат информацию о звонке и содержании разговора.

Бизнес-эффект от внедрения Audiogram в «Защитника» МТС:

- Число пользователей услуги выросло более чем в три раза за 5 месяцев;
- Коэффициент возврата инвестиций (ROI) составил 73%;
- Повышение лояльности пользователей за счет снижения количества нежелательных разговоров.

Интеграция Audiogram заняла около 5 месяцев и стоила 2,1 млн рублей
с учетом разработки кастомизированных сервисов.
Продукт запущен в коммерческую эксплуатацию и продолжает развиваться.



4.2. Использование Audiogram для озвучивания книг МТС Библиотеки

У МТС появилась задача увеличить количество аудиокниг в библиотеке и сократить время на их подготовку, чтобы пользователи не ждали появления новинок от трех до 6 месяцев и не переставали пользоваться сервисом.

МТС обратилась в MTS AI для поиска способа создавать звуковой контент без привлечения диктора, аренды студии, затрат на обработку звука и т.д.

MTS AI предложила использовать возможности Audiogram.
Платформа позволяет озвучивать художественные тексты с необходимыми интонациями и ударениями. По желанию заказчика может быть использован один из четырех голосов: женский и три мужских.

Что было сделано?

- Автоматизирован процесс по созданию аудиокниг из электронных версий изданий в распространенном формате EPUB;
- Усовершенствована модель синтеза речи: улучшены интонации, характерные для литературных текстов, в том числе вопросительные, расстановка ударения*.

Результат

Audiogram позволил МТС Библиотеке оптимизировать процесс подготовки аудиокниг:

- Время на подготовку аудиоверсии электронного издания сократилось с нескольких месяцев до часа – 30 минут;
- Качество озвучания осталось на высоком уровне.
  Между синтезированным голосом и дикторской озвучкой пользователи выбирали первый вариант**.

По результатам эксперимента, запущен MVP для озвучки 300-500 художественных книг в МТС Библиотеке.

* Согласно исследованию, проведенному MTS AI, интонации и ударения Audiogram расставляет лучше, чем лидеры отрасли: Яндекс.Читалка и spichki.org.
** Опрос был проведен MTS AI среди пользователей МТС Библиотеки.

4.3. Создание с помощью Audiogram ИИ-оператора для контакт-центра МТС

Компания МТС обратилась в MTS AI с предложением создать сервис по обслуживанию клиентов в контактных центрах параллельно с текущим IVR. Целью этого было снизить нагрузку на операторов и повысить качество обслуживания абонентов.

Разработчики MTS AI с помощью Audiogram создали голосового помощника, отвечающего на звонки клиентов, с функцией синтеза и распознавания речи.

Что было сделано?

- Подключили чат-бот, созданный на платформе JAICP, к АТС.
- Модули распознавания и синтеза речи Audiogram подключили к программному обеспечению и к чат-боту контакт-центра МТС с помощью UniMRCP.
- Настроили детектор активности речи (voice activity detector – VAD) – алгоритм, предназначенный для различения интервалов активной речи и пауз.

Результат

Эксперимент по обслуживанию клиентов экосистемы МТС с помощью ИИ-оператора голосового помощника запущен в двух регионах России. Голосовой помощник принимает входящие звонки, транскрибирует их, отправляет в бот и синтезирует устный ответ.

- Голосовой помощник обработал более 200 тысяч обращений с начала года.
- Улучшение сервисного обслуживание привело к повышению лояльности клиентов на 17-20%.

В дальнейшем планируется подключить к эксперименту новые регионы и другие направления: банк, цифровые продукты, обслуживание абонентов стационарной телефонной сети, и перевод проекта в коммерческую эксплуатацию.

Внедрение голосового помощника заняло около 7 месяцев с учетом проведения пилота.

5. ОСНОВНАЯ ФУНКЦИОНАЛЬНОСТЬ

ASR (Automatic Speech Recognition) — автоматическое распознавание речи

- Потоковое преобразование речи в текст, которое позволяет транскрибировать аудиозапись в реальном времени и получать результаты в текстовом формате.
- Файловое преобразование речи – асинхронное транскрибирование речи в текст для больших объемов аудиофайлов или аудиоархивов.

В Audiogram доступно два типа моделей:

- доменная, которая позволяет эффективно распознать речь в области медицины, телекома и финансов;
- общая модель с повышенным потреблением ресурсов, подходящая для применения в любой сфере в широком диапазоне шумности.

TTS (Text-to-Speech) – преобразование текста в речь

- Озвучивание текста женским или одним из трех мужских синтезированных голосов.
- Автоматическая ML разметка, для литературной озвучки книг и статей.
- Платформа поддерживает язык разметки синтеза речи SSML, что позволяет добиваться более естественного звучания с помощью управления интонацией, скоростью, ударениями и другими параметрами.

Вспомогательные сервисы:

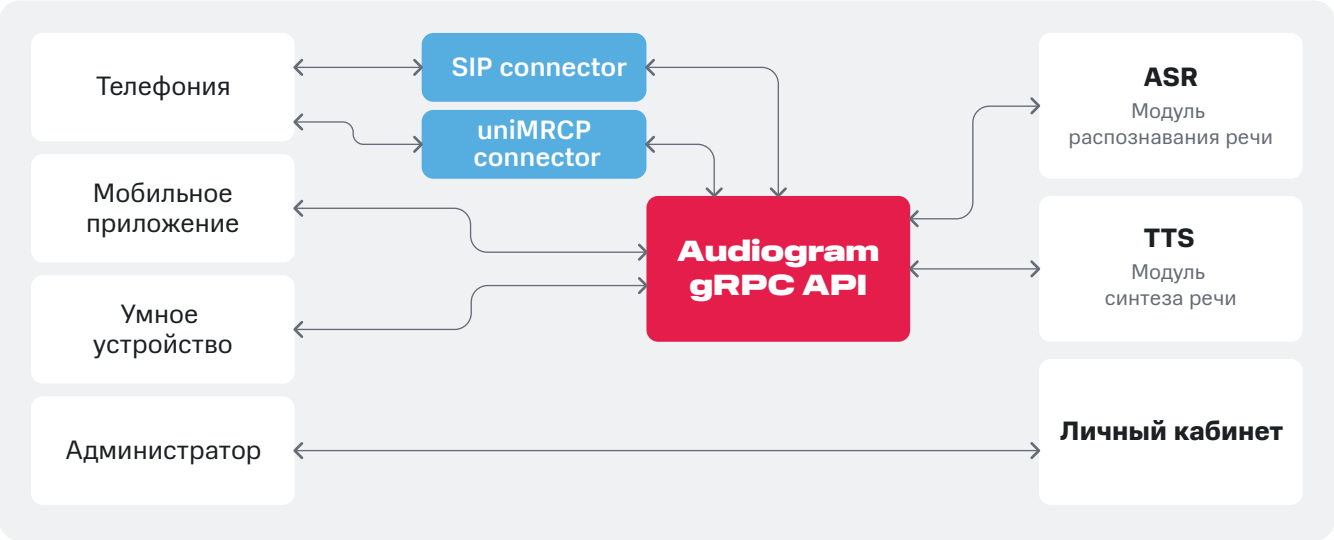
- сбор статистики по использованию платформы;
- биллинг – формирование счета на основании статистики использования услуг платформы и тарифов;
- сервисы коннекторов для поддержания взаимодействия с внешними приложениями.

6. ПРОГРАММНЫЕ КОМПОНЕНТЫ

6.1. Описание программных компонентов

Платформа **Audiogram** создана по принципу микросервисной архитектуры.

Audiogram предоставляется пользователю как API, через которое он может взаимодействовать с платформой напрямую, так и набор коннекторов для преобразования в другие протоколы: SIP connector, UniMRCP connector, REST gateway.



On premise и cloud-версии Audiogram

MTS AI может предоставлять платформу Audiogram в двух форматах:

ON PREMISE

Заказчик получает дистрибутив ПО и лицензию для установки на своих серверах.

CLOUD

Решение будет развернуто в облаке, на мощностях MTS AI, и заказчик получит только ссылку на API для работы и ссылку на личный кабинет.

Никакого дополнительного оборудования ему устанавливать не нужно.

Разработчики рекомендуют компаниям, для которых важно обеспечивать конфиденциальность данных клиентов (например, для банков и телеком-компаний), выбирать вариант on premise. Таким образом, вся информация не будет выходить за пределы компании.

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ МОДУЛЕЙ ГОЛОСОВОЙ ПЛАТФОРМЫ

6.2. Модуль ASR

Общие характеристики:

- интеграция с системами клиента с помощью стандартного протокола передачи данных gRPC;
- задержка ответа от 500 мс*;
- интеграция с платформой IP-телефонии Asterisk через программный модуль SIP connector;
- интеграция с программным обеспечением для контакт-центров Genesis с помощью ПО UniMRCP connector;
- поддерживаемые форматы аудио: WAV PCM 16bit, WAV MULAW, WAV ALAW;
- доступный язык — русский.

Режимы работы:

Файловый

Подходит для распознавания одноканального аудио небольшого размера, ответ будет направлен по окончании передачи аудио, задержка ответа – не менее длины самого аудио

Потоковый

Позволяет в рамках одного соединения отправлять аудиофрагменты и получать результаты, в том числе промежуточные результаты распознавания, задержка ответа – от 500 мс

Распознавание длинных аудио

Дает возможность распознавать длинные многоканальные аудиозаписи, скорость ответа зависит от длины аудио

У MTS AI в наличии есть языковые модели для таких тематик, как телеком, медицина, и общая разговорная модель

Общая модель распознавания речи:

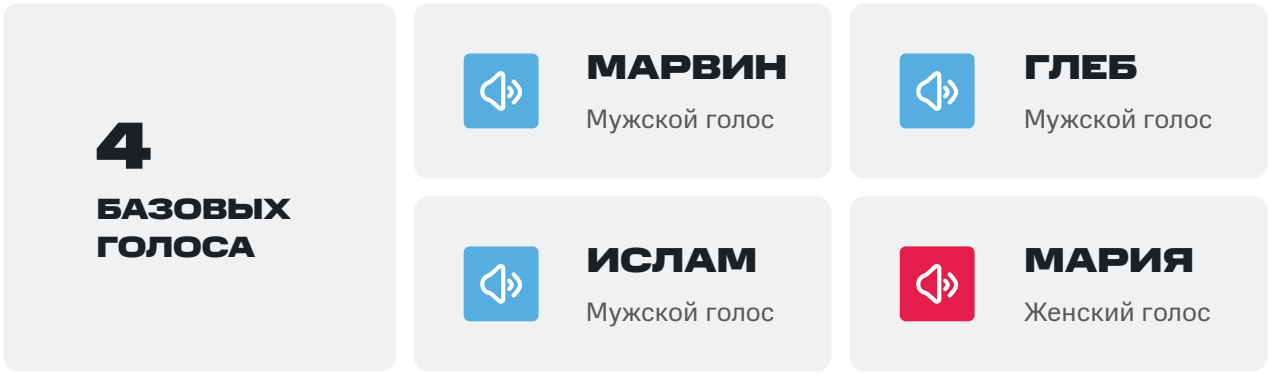
- не требуется дообучения;
- возможность запустить в работу «из коробки»;
- более высокая точность распознавания, чем у доменной модели.

Доменная модель распознавания речи:

- ориентирована на пользователя из конкретной сферы: телекома, ритейла, медицины, образования и так далее;
- требуется дообучение для новых предметных областей;
- для дообучения необходимо более 200 часов аудиоматериала.

6.3 Модуль TTS

Общие характеристики:



- разметка SSML для точечного управления синтезом, позволяющая корректировать интонацию, скорость, ударения и другие параметры, ставить акценты во фразах ботов;
- задержка ответа от 500 мс*;
- доступный язык — русский;
- автоматическая разметка для художественной озвучки.

*Итоговая задержка зависит от количества символов.

7. РЕКОМЕНДУЕМЫЕ МОДЕЛИ РАЗВЕРТЫВАНИЯ

7.1. ASR

Процессор	Оперативная память	Объем жесткого диска
2x GPU Nvidia V100 16Gb, 64 vCPU 2.3GHz	160 Gb RAM	1.5 Tb HDD

В доменной модели ASR каждый сервер с описанной выше конфигурацией может обрабатывать одноканальные аудио с разной пропускной способностью:

- в файловом режиме пропускная способность составляет 700 RTF (Real-Time Factor) в 1000 одновременных потоков с потреблением оперативной памяти 512 Gb;
- в потоковом режиме — 500 одновременных потоков с потреблением 128 Gb оперативной памяти. Для пятисекундного аудио задержка составляет 0.271 сек; для десятисекундного — 0.343 сек; для пятнадцатисекундного — 0.364 сек.

В общей модели ASR каждый сервер с описанной выше конфигурацией может обрабатывать одноканальные аудио:

- в файловом режиме с пропускной способностью 80 RTF (Real-Time Factor), потребляя при этом 512 Gb оперативной памяти;
- данные об утилизации в потоковом режиме будут предоставлены по окончании продуктивизации.

7.2 TTS:

Процессор	Оперативная память	Объем жесткого диска
1x GPU Nvidia V100 16Gb, 32 vCPU 2.3GHz	64 Gb RAM	1.7 Tb HDD

На текущей модели TTS каждый сервер с описанной выше конфигурацией синтезирует аудио со скоростью 290 символов в секунду.

7.3. Вспомогательные модули:

Модуль	Процессор	Оперативная память	Объем жесткого диска
Личный кабинет	4 vCPU 2.3GHz	8 Gb RAM	256 Gb HDD
Сервис статистики	8 vCPU 2.3GHz	16 Gb RAM	2 Tb HDD

В предлагаемой схеме серверы для ASR/TTS можно горизонтально масштабировать, исходя из ожидаемой нагрузки:

- gRPC API gateway масштабируется путем увеличения vCPU и RAM пропорционально увеличению нагрузки;
- Модули «Личный кабинет», «Сервис статистики» также масштабируются путем увеличения vCPU и RAM.

8. СРАВНЕНИЕ С КОНКУРЕНТАМИ

Разработчики MTS AI постоянно совершенствуют платформу Audiogram, добавляя новые функции.

Чтобы познакомиться с актуальным конкурентным анализом, сканируйте QR-код.

9. ЛИЦЕНЗИРОВАНИЕ

Платформа Audiogram представляет два вида тарифов:

- pay as you go;
- пакетный.

Контакты для запроса цен и демо

Чтобы запросить информацию о тарифах, получить доступ к демо и дополнительную информацию, обращайтесь по адресу **sales@mts.ai,** либо сканируйте QR-код и переходите на сайт MTS AI.


MTS AI