

## Оглавление

Термины и сокращения .....	1
Введение .....	1
1 ОБЗОРНАЯ ЧАСТЬ.....	1
1.1 Анализ аналогов приложения. ....	1
1.2 Обзор технологий для распознавания голоса	<b>Ошибка! Закладка не</b>
определена.	
1.3 Обзор технологий для коррекции текста	<b>Ошибка! Закладка не</b>
определена.	
2 РАСЧЕТНО-КОНСТРУКТОРСКАЯ ЧАСТЬ .....	1
2.1 Определение требований к системе .....	1
2.2 Разработка структуры базы данных .....	1
2.3 Выбор языка и платформы разработки .....	1
2.4 Разработка архитектуры приложения .....	1
2.5 Разработка структуры серверной части приложения .....	1
2.6 Разработка структуры клиентской части приложения .....	1
2.7 Разработка структуры интерфейса взаимодействия пользователя с системой.	1
2.8 Разработка плана проведения тестирования и отладки.....	1
3 ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ .....	1
3.1 Реализация разработанных алгоритмов .....	1
3.2 Реализация базы данных .....	1
3.3 Тестирование и отладка .....	<b>Ошибка! Закладка не определена.</b>
3.4 Комплексная отладка и тестирование	<b>Ошибка! Закладка не определена.</b>
3.5 Оценка качества функционирования. ....	<b>Ошибка! Закладка не определена.</b>
3.6 Руководство пользователя.....	<b>Ошибка! Закладка не определена.</b>
Заключение.....	<b>Ошибка! Закладка не определена.</b>
Список используемых источников .....	<b>Ошибка! Закладка не определена.</b>
Приложения	



## Оглавление

Термины и сокращения .....	3
Введение.....	3
Обзорная часть .....	7
Анализ аналогов приложения .....	7
Обзор технологий для распознавания голоса .....	9
Обзор технологий для коррекции текста.....	17
Обзор мобильных операционных систем .....	18

# Термины и сокращения

ОВС – ограниченные возможности слуха

## Введение

От начала времен человечество двигалось по пути саморазвития и поиска инструментов для облегчения жизни. Первобытные люди придумали каменные инструменты, чтобы можно было возделывать землю, охотиться и строить дома. Затем появилась необходимость для инструментов вместо камня использовать другие материалы. Т. к. стало понятно, что камень сложно обрабатывать и он малоэффективен. Так появились металлы и технологии их добычи, обработки иковки. Одновременно с этим люди начали изучать целебные свойства веществ и способы их получения и использования. Так постепенно, с течением времени наши предки перешли от жилищ в пещерах к надежным деревянным домам. Конечно, технология постройки домов отличается для климатических зон. Так, если мы на севере строили толстые и надежные срубы, которые должны были защищать нас от лютых морозов. То народы на юге строили маленькие хлипкие домики только для того, чтобы защититься от дождя, ветра и диких животных, которые шастают в округе. Так же люди научились использовать огромное кол-во природных материалов. Медь и железо использовалось для создания инструментов и оружия. Для постройки жилищ, и во многих других сферах. Люди научились использовать целебные свойства трав что бы лечить болезни. И на этой основе было создано огромное кол-во трудов. Для облегчения жизни люди научились приручать и использовать животных. Люди научились разводить скот для добычи еды, вместо того что бы с палками бегать по лесу. Люди научились использовать животных для облегчения передвижения и перевозки грузов. Люди придумали корабли что бы плавать по рекам и в дальнейшем морям. Но несмотря на то что уже практически все аспекты жизни были продуманы, и так или иначе

облегчены, человечество все равно двигалось по пути развития. Технологии так или иначе развивались неумолимо, и все во благо облегчения жизни человека. Если начать рассматривать основные значимые этапы развития человечества, то станет ясно что «чудо» начало происходить с началом индустриального века. Если до этого, в древности и в средних веках люди открыли медь, железо, и мореплавание, то с началом века индустриализации люди начали создавать рабочие станки и в последствии фабрики. Эта необходимость была вызвана тем, что кол-во людей неуклонно росло даже несмотря на различные катаклизмы. Росло кол-во людей, а значит росло кол-во потребляемых товаров. Десятки лавочек-портных уже не могли обеспечить потребности крупных городов. Не говоря уже о продуктовой сфере. вследствие этого появилась необходимость увеличить масштабы производства и человечество начало создавать станки. В целом и в средние века были простейшие станки, например ткацкий или точильный. Но именно в эпоху индустриализации человек усложнил их ради облегчения собственной жизни. Станок стал воплощением человеческого гения. Он позволял реализовать сложную технологию производства и автоматизировать ее. В рамках фабрик появились производственные линии. Этот прорыв позволил человечеству покрыть все потребности в товарах практически любой сферы. Конечно, этот скачок принес и проблемы. На первых этапах сильно пострадала экология, и были определенные социальные проблемы. Но все равно, человек в очередной раз облегчил себе жизнь используя технологии. Дальше больше, человечество придумало автомобили. Хотя была возможность передвигаться на лошадях и в редких случаях на транспортных средствах на человеческой тяге. Но за лошадьми надо ухаживать и их надо где-то содержать. А технологическое развитие в это время неумолимо шло вперед. Появилась возможность сделать автомобиль. Ему больше не нужно было стойло с несколькими конюхами. Его не надо было кормить и обслуживать как ездовое животное. Он был «быстр» и мощен. В дальнейшем автомобили развивались одновременно с другими технологическими аспектами. Их начали

использовать в грузоперевозках. Появилось огромное кол-во разновидностей автомобилей, от небольших грузовичков, до огромных фур в дальнейшем. Автомобили были так же и средством роскоши. Но даже при этом их первостепенная идея заключалась в облегчении жизни. А именно в облегчении перемещения. Затем появилась авиация. Изначально она казалась бредом и затем чудом. Но очень скоро человек начал использовать ее для всего. Для перевозок, для перемещения, для войны в конце концов. И в это ж время, все другие технологические аспекты человеческой жизни параллельно развивались. Электричество, радиосвязь, топливо и многие другие технологии привели к тому, что современному человеку в 18-19 веке уже не надо было тратить огромное кол-во сил на выживание. Ему не надо было охотиться – он мог купить еду. Ему не надо было мерзнуть зимой в темноте, так как появилось централизованное отопление и электричество. Одежда стала доступнее. Благодаря технологическому развитию, человек избавил себя от многих трудностей. У человека стало появляться больше свободного времени. Конечно, это в очередной раз привело к очередным социальным проблемам, но любое развитие социума не может проходить гладко. Следующим важным этапом было появление цифровых технологий. Крайне сложно даже представить, как это повлияло на жизнь, и не возможно охватить все аспекты жизни которые эти технологии изменили. Технологический скачок получился просто грандиозным. У человечества появилась возможность еще сильнее автоматизировать производство, одновременно с повышением качества продукции, что раньше прямо противоречило друг другу. Невероятно облегчилась навигация. Человеку больше не надо смотреть на звезды, или полагаться на магнитные полюса земли. Благодаря этому появились автоматизированные пилотируемые системы. Средства связи стали проще и доступнее. В настоящее время мы можем позвонить другому человеку с другого конца света, и мы услышим ответ через несколько секунд, раньше это было чем-то невообразимым. Практически невозможно охватить все аспекты жизни, на которые повлияло появление цифровых технологий, но

наиболее важным аспектом оказалась медицина. Базовые болезни вроде простуды перестали вообще быть какой-либо проблемой. Появилась возможность более безопасно проводить сложные операции, а также дольше поддерживать жизнь тяжело больных людей. Технологический процесс производства лекарств так же упростился, и позволил производить более сложные вещества. Правда текущая ситуация с коронавирусом показала, что технологический уровень человечества в этой сфере все же недостаточно высок.

Таким образом главная цель технологического развития всегда состояла в упрощении жизни человека. И в наше время существует огромное кол-во технологий и методик для помощи людям с ограниченными возможностями. Технологии позволяют частично восстановить потерянные конечности или утраченный функционал. Но даже сейчас существуют сферы, в которых люди с ограниченными возможностями чувствуют себя некомфортно. Одна из таких сфер – образование. Несмотря на наличие специальных центров для людей с ограниченными возможностями и специальных методик обучения. Все равно есть образовательные программы, которые были бы интересны таким людям, но из-за ограничений они не могут ими воспользоваться. Как пример: люди с ограниченными возможностями слуха. Это люди, у которых частично снижена способность обнаруживать и понимать звуки. Иногда такие люди хотят освоить программу которой нет в специальном образовательном учреждении. Особенно если речь идет о каком-то дополнительном курсе. И так же существуют ситуации, когда подобные дети учатся в обычных образовательных учреждениях. И им остается либо полагаться на устройства повышающие способность понимать и обнаруживать звуки. Либо на визуальное восприятие информации. Очевидно, что не всегда презентация содержит достаточно информации, так как прежде всего она служит для иллюстрации того что рассказывается. Цель данного проекта заключается в том, что бы используя имеющиеся технологии позволить людям с ограниченными возможностями слуха как

можно более эффективно воспринимать информацию, а так же повысить качество коммуникации между ведущим курса и участниками курса с ограниченными возможностями слуха. Результатом должно стать мобильное приложение, предоставляющее функционал трансляции речи ведущего на мобильные устройства участников. Тем самым переводя речь ведущего в информацию, которую можно воспринимать визуально. Это особенно актуально в период пандемии коронавируса, так как все занятия проводятся в удаленном формате.

## Обзорная часть

### Анализ аналогов приложения

На данный момент не существует приложений, нацеленных конкретно на повышение эффективности образовательного процесса для людей с ограниченными возможностями слуха. Однако существует несколько приложений с подобным функционалом перевода речи собеседника в текст с последующим отображением на экране:

#### **Яндекс.Разговор**

Яндекс.Разговор – приложение, разработанное компанией Яндекс, для помощи людям с ограниченными возможностями слуха. Приложение переводит речь в текст и обратно. Придумано и разработано группой студентов из Московского физико-технического института с использованием технологии распознавания речи Yandex SpeechKit.

Все что говорит один из собеседников переводится в речь и отображается на экране собеседника. Так же есть возможность произносить вслух напечатанные слова.

Плюсы:

- Разрабатывается и поддерживается компанией Yandex



- Удобный интерфейс и хорошим дизайном
- Использует технологию Yandex SpeechKit
- Есть история диалогов
- Приложение бесплатное

Минусы:

- Одновременно можно вести диалог только с одним человеком
- Точность распознавания зависит от скорости речи и дикции собеседника. Выше скорость речи – ниже точность.
- Поддерживается только один язык: Русский
- Нет возможности экспортировать диалог
- Судя по отзывам большое кол-во ошибок.

## **RogerVoice**

RogerVoice – приложение разработанное компанией Rogervoice. Приложение предоставляет функционал с помощью, которого люди с ограниченными возможностями слуха могут разговаривать по телефону. Приложение использует собственную технологию распознавания речи, для перевода голоса собеседника в текст и отображения полученного текста на экране.

Плюсы:

- Собственная технология распознавания голоса
- Удобный интерфейс с отличным дизайном интерфейса
- Бесплатное приложение

Минусы:

- Не удалось найти информацию по поддерживаемым языкам
- Только один собеседник одновременно
- Нет локализации на русский язык
- Судя по отзывам большое кол-во критичных ошибок

## **Вывод**

Рассмотренные приложения невозможно использовать в образовательном процессе, так как каждое из них одновременно может работать только с одним собеседником. Такая схема может использоваться только при индивидуальном репетиторстве, но не при массовых занятиях. Так же приложение RogerVoice невозможно использовать, так как у него нет локализации на русский язык. В целом становится понятно что на данный момент нет приложений которые могли бы использоваться при обучении людей с ограниченными возможностями слуха.

## **Обзор технологий для распознавания голоса**

### **Yandex SpeechKit**

Yandex SpeechKit – технология для синтеза и распознавания речи разработана российской компанией Yandex (Яндекс). Компания с 2012 года занимается разработками в области распознавания естественной речи. Данная технология впервые была представлена в 2013 году на ежегодной технологической конференции Яндекса. Технология поддерживает такие платформы как: Android, IOS и windows phone. Так же технология активно используется в рест-серверных Web-приложениях. Технология активно используется в проектах компании Яндекс, таких как: голосовой помощник Алиса, Яндекс.Карты и др.

Технология работает по принципу рест-сервиса. Это значит, что для интеграции с Yandex SpeechKit необходимо проводить аутентификацию по токену. И использовать HTTP методы POST и GET для получения результата распознавания или синтеза речи. Само распознавание или синтез речи проходит на серверах Яндекса. Следовательно качество работы сервиса напрямую зависит от качества интернет-соединения и загруженности сервера.

Распознавание аудио происходит в несколько этапов:

1. Из полученного текста выделяются слова. Обычно существует несколько вариантов распознанного слова.
2. Варианты распознанного слова проверяются с помощью языковой модели. Модель проверяет, насколько согласуется новое слово со словами, распознанными ранее. Для каждого варианта указывается параметр доверия, отражающий насколько распознанное слово, подходит под контекст.
3. Распознанный текст Обрабатывается – числительные преобразуются в цифры, расставляются необходимые знаки препинания (например, дефисы) и т. д. Затем преобразованный текст отправляется в теле ответа.

Плюсы:

- Технология SpeechKit учитывает стилистические и лексические особенности устной речи, а также вероятности сочетания слов для повышения точности распознаваемого текста за счет контекста.
- Три возможных режима распознавания голоса:
  - Короткие аудио – используется для распознавания речи из короткого, до 1 минуты, одноканального аудио.
  - Длинное аудио – используется для распознавания речи из многоканального аудио, не ограниченной длины. Длина аудио напрямую влияет на скорость распознавания речи.
  - Потокное аудио – позволяет в рамках одного соединения отправлять аудиофрагменты и получать результаты. Используется для получения промежуточных результатов в реальном времени.
- Поддержка трех языков: Русский, Английский, Турецкий

- Премиум голоса для синтеза речи – набор различных голосов доступных для синтеза речи. Технология оценивает текст и подбирает интонации на основе контекста, характерные для человека
- Поддержка различных форматов аудио

Минусы:

- Так как распознавание или синтез речи происходит на серверах Яндекса, то качество и скорость работы сервиса напрямую зависят от качества интернет-соединения и загруженности серверов Яндекса.
- Невозможно проводить распознавание или синтез сразу на устройстве, что само по себе увеличивает время, между тем как была сказана фраза и был получен результат распознавания, так как нам необходимо ждать ответ от сервера.
- Поддерживается только три языка: Русский, Английский, Турецкий
- Сервис является платным. Ежемесячный тариф рассчитывается на основе типа и суммарной длительности аудио дорожек в месяц. Следовательно, необходимо детально просчитывать экономическую выгоду от использования этой технологии.
- Качество голосов для синтеза речи хуже, чем у конкурентов.

Исходя из вышеперечисленных данных технология Yandex SpeechKit выглядит крайне интересно. Она имеет достаточно большое количество весомых плюсов. Наиболее интересным вариантом распознавания речи является потоковое распознавание речи. Этот режим позволяет нам получать промежуточные результаты распознавания, не дожидаясь конца фразы или предложения. Данный формат как раз необходим и является ключевой особенностью, так как при проведении лекций или занятий необходимо распознавать текст в реальном времени небольшими отрывками, а не огромными кусками, чтобы участники могли быть в «потоке».

Ключевыми же минусами является рест-сервисная архитектура и наличие абонентской платы. Рест-сервисная архитектура на несколько секунд увеличивает суммарное время распознавания речи. При хорошем качестве интернет-соединения и приемлемой нагрузке на серверах Яндекса скорость отклика будет удовлетворительной, но если один из этих аспектов ухудшится, то, следовательно, увеличится время отклика. В свою очередь время отклика является критичным. Так же этот метод распознавания имеет ряд ограничений:

- Время между отправкой фрагментов аудио должно примерно совпадать с длительностью фрагментов. Нельзя отправлять сообщения слишком редко или слишком часто. Например, каждые 400мс необходимо отправлять аудио фрагмент длительностью 400мс. Если пауза между отправкой будет больше 5 секунд, текущее соединение будет закрыто, и нужно будет открывать его заново.
- Максимальная длительность аудио в рамках одной сессии – 5 минут. Затем сессию необходимо открывать повторно.
- Размер переданных аудиоданных не должен превышать 10 мегабайт.

Вторым минусом является наличие абонентской платы. Абонентская плата напрямую зависит от длительности обрабатываемых аудиоданных. Если рассматривать ситуацию, когда проводятся лекции 22 дня в месяц по 8 часов в день, ежемесячная плата находится в районе 5 тысяч рублей. Цифры примерные, но они отражают необходимость точного расчёта экономической целесообразности использования данной технологии.

В остальном же технология хорошо зарекомендовала себя в других сервисах от компании Яндекс.

## **Google Cloud Speech API**

Google Cloud Speech API – технология разработана компанией Google. Изначально технология была встроена в платформу Android и использовалась для поиска голосом, и диктовки. Это позволяло разработчикам использовать

данный функционал в своих приложениях. Но примерно в 2013-2014 годах компания Google закрыла доступ к этому сервису так как сильно увеличилась нагрузка на сервера Google. Позднее компания представила сервис Google Cloud Speech API который по факту был описанной и регламентированной версией предыдущей технологии.

Принцип работы схож с Yandex SpeechKit. Это тоже рест-сервис, с которым обмен информацией происходит посредством HTTP запросов. Сервис так же обладает рядом преимуществ, у сервиса явно выше качество распознавания текста, более высокого качества голоса для синтеза речи. Больше кол-во поддерживаемых языков. Но у сервиса существует один ключевой недостаток:

- Значительно более высокая цена по отношению к Yandex SpeechKit. Каждые 15 секунд распознавания аудио стоят примерно 0,006 USD. Если рассматривать сценарий, когда проводятся лекции 22 дня в месяц по 8 часов в день, итоговая сумма выходит больше 12 тысяч рублей за одно только устройство. Крайне высокая цена, использование данной технологии экономически не выгодно.

## **Android Speech API (Google Speech API)**

Android Speech API – технология для распознавания и синтеза речи разработанная компанией Google. Изначально разрабатывалась под платформу Android, была встроена и активно использовалась внутри приложений с возможностью голосового поиска или диктовки. На данный момент так же используется в голосовом помощнике от Google. В дальнейшем на ее основе была создана технология Google Cloud Speech API. У Android Speech API есть огромное и основное отличие от Google Cloud Speech API, эта технология является интегрированной в платформу Android. Следовательно распознавать и синтезировать речь можно сразу на устройстве, не будучи привязанным к качеству интернет-соединения или серверам на которых

происходит обработка. На данный момент эта технология активно используется во многих приложениях на платформе Android и крайне хорошо себя зарекомендовала.

Плюсы:

- Технология интегрирована в платформу Android, что позволяет нам распознавать и синтезировать речь прямо на устройстве. Такой подход уменьшает время отклика и исключает из алгоритма возможные проблемы с интернет-соединением и загрузкой серверов.
- Технология бесплатна. Все что необходимо для ее использования это непосредственно само устройство на платформе Android.
- Поддерживается огромное кол-во языков. Технология может распознавать и синтезировать речь практически на всех известных языках.
- Высокое качество голоса для синтеза речи.

Минусы:

- Нет потокового распознавания речи. Существует только один формат обработки аудио для распознавания речи, и его особенность в том что он распознает только полную аудио дорожку, которая получается либо при длительной паузе в речи либо при ручной остановке алгоритма.
- Поддерживается только одна платформа: Android. На других платформах таких как IOS или Windows Phone эта технология не поддерживается. Но на этих платформах существуют свои аналоги от компаний Apple и Microsoft.

Основным минусом данной технологии является отсутствие потокового распознавания речи. Для примера необходимо рассмотреть алгоритм распознавания:

1. Настраивается конфигурация для распознавания. Указываются необходимы лингвистически словари, указывается периоды и промежутки, настраивается устройство ввода.
2. Запускается алгоритм распознавания речи.
3. Алгоритм формирует аудио дорожку на основе поступающих данных с устройства ввода
4. Если речь прервалась, или пауза речи выше указанной, то алгоритм прекращает запись звука и формирования аудио дорожки и приступает к распознаванию.
5. Распознанный текст выдается в результат.

Данный алгоритм не подразумевает получение промежуточных результатов распознавания. В лучшем случае удастся получить результат по предложениям, но если человек с высоким темпом речи, то результат мы получим сильно позже. Если рассматривать процесс ведения лекций или семинаров, то такой вариант не подходит, так как необходимо получать результат в реальном времени что бы участники могли реагировать на события так же в реальном времени. При использовании данной технологии проблему с отсутствием потокового распознавания речи можно решить с помощью периодической остановки алгоритма. Так если мы с определенной периодичностью на шаге 4 алгоритма будем вызывать прерывание алгоритма, то мы циклически будем проходить пункты 2-5, тем самым фактически получая потоковое распознавание речи. Но у такого подхода есть ряд недостатков. Основной заключается в том, что слова распознаются и корректируются на основе контекста фразы. А при прерывании алгоритма контекст фактически прерывается и возникают ситуации, когда распознанные слова находятся не в том падеже или теряют окончания. Так же могут быть потеряны союзы. Эту проблему можно решить, используя инструменты коррекции текста, когда будут анализироваться распознанный текст из моментов времени  $t$  и  $t+1$  и на основе контекста этих двух составляющих корректировать текст.



В целом технология Android Speech API является наилучшим вариантом, так как она интегрирована в платформу и не зависит от качества сетевого соединения. Но корректная работа будет обеспечена только в том случае если будет решена проблема с ухудшением качества распознанного текста из-за потери контекста в потоковом распознавании.

## **Другие технологии**

Существует также некоторое количество сторонних библиотек с открытым исходным кодом вроде Pocketsphinx и других. Но они не могут рассматриваться в качестве потенциальных технологий так как работают по принципу реагирования на конкретную заранее заготовленную фразу.

## **Итог**

Из перечисленных выше технологий можно выделить две:

### **Yandex SpeechKit**

Плюсы:

- Наличие потокового распознавание речи
- Высокое качество распознавания
- Относительная простота интеграции
- Возможна поддержка различных мобильных платформ

Минусы:

- Зависимость от интернет-соединения и нагрузки на сервера сервиса
- Наличие весомой абонентской платы

## Android Speech API

Плюсы:

- Технология интегрирована в платформу Android, что в свою очередь убирает влияние качества интернет-соединения и загрузке серверов сервиса на скорость отклика при распознавании речи
- Не имеет абонентской платы.

Минусы:

- Повышенная сложность реализации. Так как изначально технология не поддерживает потоковое распознавание речи. Алгоритм распознавания требует доработки и постобработки полученного результата с помощью коррекционных инструментов.
- Поддерживает только платформу Android

Остальные технологии либо не удовлетворяют условиям, либо слишком дороги.

## Обзор технологий для коррекции текста

Алгоритмы распознавания текста имеют крайне высокую точность, но иногда результат все равно содержит ошибки, особенно если появляются сильные внешние шумы либо помехи. Для коррекции подобных ошибок можно использовать инструменты для коррекции текста. Ниже приведены некоторые из актуальных инструментов для коррекции.

### LanguageTool

LanguageTool – библиотека разработана немецкой компанией LanguageTooler GmbH в 2016 году. Интегрирована в библиотеку текстового поиска Lucene и поставляется вместе с ней.

Технология проста в интеграции и поддерживает большое кол-во языков. Распространяется по лицензии компании Apache, стоимость определяется от кол-во проверяемых ежемесячно слов. Библиотека является встраиваемой, что позволяет анализировать и исправлять текст непосредственно в приложении.

## **Яндекс. Спеллер**

Яндекс. Спеллер – сервис, разработанный компанией Яндекс. Является рест-сервисом следовательно запросы к сервису и получение ответа осуществляется через HTTP запросы методами. Ключевыми минусами сервиса является ограничение в 10 миллионов символов в сутки и зависимость скорости отклика от качества интернет-соединения и нагрузки на серверах сервиса.

## **Обзор мобильных операционных систем**

Android – изначально разрабатывалась компанией Android inc. которую затем выкупила компания Google. Релиз операционной системы состоялся 23 сентября 2008 года. Операционная система основана на ядре Linux и собственной реализации виртуальной машины Java от компании Google. Система получила огромное распространение и большее кол-во смартфонов на данный момент построены на операционной системе Android

IOS – операционная система была выпущена в 2007 году компанией Apple. IOS основана на ядре XNU и языке C, C++. Операционная система используются в основном на устройствах компании Apple, из-за чего популярность меньше чем у операционной системы Android.