**Введение**

В современную эпоху мобильные устройства и платформы играют ключевую роль в повседневной жизни. Операционная система Android, впервые выпущенная в 2008 году, сегодня стала самой популярной мобильной платформой в мире​. Миллионы пользователей ежедневно создают аудиозаписи – от голосовых заметок и сообщений до записей встреч и лекций. Так, только в мессенджере WhatsApp ежедневно отправляется свыше **150 миллионов голосовых сообщений**. Однако зачастую аудиоинформация трудно поддается поиску и анализу без преобразования в текстовый формат. В связи с этим актуальна задача автоматической **транскрипции аудиозаписей** (Speech-to-Text) для получения удобных текстовых расшифровок. Одновременно набирают популярность приложения для изменения голоса – они используются как для развлечения, так и для обеспечения анонимности пользователя при общении​. **Диктофон с функцией изменения тембра голоса и распознавания речи** объединяет эти тренды, позволяя не только записывать аудио, но и модифицировать голос на записи и получать текстовую расшифровку. Такая разработка актуальна для широкого круга пользователей: журналистов и студентов (для расшифровки интервью и лекций), блогеров и создателей контента (для создания аудиоэффектов), а также всех, кто ценит конфиденциальность или креативность в голосовых сообщениях.

**Цель исследования** – разработать Android-приложение диктофон с функциями изменения голоса и автоматической расшифровки аудиозаписей. Для достижения этой цели необходимо решить ряд взаимосвязанных задач: (1) изучить современное состояние Android-платформы и подходы к разработке мобильных приложений; (2) проанализировать существующие решения – приложения диктофоны, программы для изменения голоса и сервисы распознавания речи – и выявить их функциональность и ограничения; (3) спроектировать архитектуру приложения и выбрать технологии, библиотеки и инструменты, оптимальные для реализации поставленных функций; (4) реализовать функциональные модули записи звука, обработки аудио (изменения голоса) и распознавания речи; (5) разработать удобный пользовательский интерфейс и обеспечить положительный пользовательский опыт (UX); (6) провести тестирование приложения (функциональное, модульное, UX) и проанализировать качество работы; (7) сформулировать выводы о достигнутых результатах и возможностях дальнейшего развития проекта.

**Объект разработки** – мобильное приложение-диктофон для операционной системы Android, предназначенное для записи и последующей обработки голосовых сообщений. **Предмет исследования** – технологии и методы, обеспечивающие изменение голоса и распознавание речи в мобильных Android-приложениях. Проще говоря, предметом является совокупность алгоритмов и инструментальных средств, используемых для трансформации аудио сигнала (изменения тембра и скорости голоса) и для автоматической транскрипции речи в текст в условиях ограниченных ресурсов мобильного устройства.

**Методы исследования** включают анализ литературы и существующих решений, системное проектирование программного обеспечения, а также экспериментальную реализацию и тестирование приложения на целевой платформе. В ходе работы используются методы **объектно-ориентированного проектирования** приложений, паттерны проектирования архитектуры (MVVM/MVI), а также методы цифровой обработки сигналов для модификации аудио. Для распознавания речи применяются методы машинного обучения (как облачные API распознавания, так и локальные модели). При тестировании используются методы функционального тестирования (проверка соответствия реализованных функций техническому заданию), модульного тестирования отдельных компонентов кода и **юзабилити-тестирования** с привлечением пробных пользователей для оценки удобства интерфейса.

**Инструменты и средства разработки** были выбраны исходя из актуальных стандартов индустрии мобильной разработки. Проект реализуется в среде **Android Studio** – официальной IDE для разработки под Android. В качестве основного языка программирования выбран **Kotlin**, который с 2019 года признан Google “языком первого класса” для Android-разработки. Для создания пользовательского интерфейса используется современный декларативный фреймворк **Jetpack Compose**, стабильная версия которого была выпущена в 2021 году​. Compose позволяет описывать UI в виде композиции функций на Kotlin и значительно сокращает объем шаблонного кода по сравнению с традиционной разметкой XML, ускоряя разработку и упрощая поддержку интерфейса. Для реализации изменения голоса рассматриваются специализированные **DSP-библиотеки** (например, *SoundTouch* для изменения темпа/тона аудио) и встроенные возможности Android для работы с аудио (класс MediaRecorder и AudioEffect). Для функции распознавания речи используются либо облачные **Speech-to-Text API** (например, Google Speech API) при наличии интернет-соединения, либо локальные модели через **ML Kit** – набор библиотек машинного обучения на устройстве​. В приложении также применяются компоненты **Android Jetpack** (ViewModel, LiveData/Flow) для реализации архитектурных паттернов и **библиотеки тестирования** (JUnit, Espresso) для проверки корректности работы.

Таким образом, введение обосновывает актуальность темы, формулирует цель и задачи выпускной квалификационной работы, определяет объект и предмет исследования, а также описывает методы и инструментарий, используемые при разработке Android-приложения диктофона с функциями изменения голоса и расшифровки аудио. Далее в работе последовательно рассматриваются теоретические основы (Глава 1), вопросы реализации проекта (Глава 2), результаты тестирования и оценки качества приложения (Глава 3), а в заключении подводятся итоги и намечаются перспективы развития.

1 **Теоретические основы разработки Android-приложений**

* 1. **История и развитие платформы Android**

Android – это операционная система для мобильных устройств, основанная на модифицированном ядре Linux и других открытых компонентах. Разработка Android началась в компании Android, Inc. в 2003 году, а в 2005 году компанию приобрела Google​. Первая публичная бета-версия SDK была выпущена в ноябре 2007 года, а уже в сентябре 2008 года состоялся релиз Android 1.0. Эту версию получила коммерческая модель телефона HTC Dream (T-Mobile G1) – первый Android-смартфон на рынке. С тех пор платформа стремительно эволюционировала. К 2010 году Android обошел по популярности другие мобильные ОС (Symbian, BlackBerry OS, Windows Phone и др.), став доминирующей платформой для смартфонов​. На 2024 год самая свежая версия системы – Android 15 (Vanilla Ice Cream)​.

За годы развития Android претерпел многочисленные изменения. Каждая крупная версия получала кодовое имя (Cupcake, Donut, ... Tiramisu и т.д.) и привносила новые возможности: поддержка мульти-тач, фронтальных камер, датчиков, сервисов геолокации, улучшения в производительности и безопасности и др. Важной чертой Android с самого начала была его открытость: исходный код базовой системы (AOSP) опубликован под лицензией Apache 2.0, что позволило множеству производителей устройств адаптировать ОС под свои продукты​. Одновременно Google предлагает собственный проприетарный пакет сервисов (Google Mobile Services) для коммерческих устройств, включая магазин приложений Google Play и другие приложения Google.

Ключевым решением, предопределившим успех Android, стала ставка на открытую модель разработки и распространения. Google создала **Open Handset Alliance** – консорциум из десятков компаний (производителей телефонов, операторов связи, разработчиков), что позволило объединить усилия индустрии. Благодаря открытому характеру ОС, Android быстро завоевал поддержку сторонних разработчиков приложений. Уже через несколько лет после запуска количество доступных приложений в Google Play исчислялось сотнями тысяч, а сегодня – миллионами.

Развитие Android сопровождается стремительным ростом технических возможностей устройств. Современные смартфоны имеют многоядерные процессоры, специализированные сопроцессоры для обработки сигналов и машинного обучения, качественные микрофоны и камеры. Все это расширяет потенциал приложений – например, становится реальностью выполнение задач обработки аудио и распознавания речи непосредственно на устройстве без обращения к серверу. В то же время, рост функциональности ОС усложняет разработку – необходимо учитывать разнообразие версий Android (фрагментацию), разные форм-факторы устройств (телефоны, планшеты, часы, авто) и соблюдать требования безопасности и оптимизации.

**1.2 Современные архитектурные подходы (MVVM, MVI) в Android**

При разработке сложных Android-приложений особое внимание уделяется архитектуре – структуре компонентов приложения и принципам взаимодействия между ними. Правильно выбранная архитектура повышает **масштабируемость, поддерживаемость и тестируемость** кода. В Android за годы практики сформировались несколько популярных архитектурных шаблонов: **MVP** (Model-View-Presenter), **MVVM** (Model-View-ViewModel) и относительно новый **MVI** (Model-View-Intent).

Шаблон **MVC** (Model-View-Controller) был одним из первых, но в чистом виде в мобильной разработке используется редко. Более распространен его вариант MVP, где View пассивна и делегирует обработку событий Presenter’у. Однако начиная примерно с 2015–2016 гг. Google и сообщество стали продвигать архитектуру **MVVM** как более удобную для Android, особенно с появлением библиотеки **Data Binding** и позже **Jetpack ViewModel**. В MVVM логика представления выносится в отдельный компонент ViewModel, который хранит состояние UI и реагирует на изменения данных. **ViewModel** выступает посредником между Model (данными, бизнес-логикой) и View (экраном), предоставляя данные во **виде потоков или привязок**, на которые подписывается View​. Важное отличие MVVM: ViewModel не знает о конкретной View и не держит ссылки на активити или фрагменты – связь осуществляется через наблюдаемое состояние (например, LiveData). Это устраняет жесткую связность и упрощает тестирование – ViewModel можно тестировать отдельно, имитируя изменения данных. MVVM хорошо сочетается с реактивным программированием и **двунаправленным связыванием данных**, ускоряя обновление UI при изменении состояния​. Архитектура MVVM стала де-факто рекомендуемым стандартом от Google (входит в **Guide to App Architecture**), так как обеспечивает разделение ответственности и удобство при изменении дизайна интерфейса​.

**MVI** – более современный подход, набравший популярность вместе с развитием реактивных библиотек (RxJava, Kotlin Coroutines + Flow) и появлением Jetpack Compose. Расшифровывается как Model-View-Intent. В MVI все взаимодействия представляются как **однонаправленные потоки данных**. Пользовательские события формируют *Intents*, которые редукторами преобразуются в новое состояние (Model), а View просто подписана на этот **единый источник состояния** и отражает его изменения​. Таким образом, исключается вероятность рассинхронизации разных частей UI – все элементы экрана получают консистентное состояние из одного источника. MVI упрощает отладку и **предсказуемость поведения интерфейса**, так как любой переход интерфейса обусловлен явным событием и обновлением состояния. Этот паттерн особенно удобен в сочетании с декларативным UI (Compose), где экран целиком строится на основе текущего состояния. Однако плата за это – возросшая сложность реализации: приходится явно описывать intents, reducers, state, и зачастую использовать немало шаблонного кода или сторонних фреймворков. Кроме того, классическая MVI тесно связана с реактивным подходом – изначально предполагалось использование RxJava​, хотя сейчас можно реализовать и на Kotlin Flow. В результате MVI используется в проектах, критичных к надежности UI и где оправдано усложнение кода ради его чистоты. Согласно исследованиям, MVVM упрощает написание модульных тестов UI-логики, а MVI улучшает управляемость состоянием и масштабируемость приложения за счет строгого следования принципу единого источника правды (SSOT)​.

Важно отметить, что на практике нередко применяется гибридный подход: разработчики берут за основу MVVM, но привносят элементы MVI (например, разделяют единое состояние экрана и intents), или используют **чистую архитектуру (Clean Architecture)**, выделяя дополнительный слой бизнес-логики (Domain) между UI и данными. Официальные рекомендации Google по архитектуре (Modern App Architecture) советуют разделять приложение на **слой данных** (репозитории, источники данных), **слой домена** (use case, бизнес-правила, опционально) и **слой представления (UI)**, в котором ViewModel играет роль state holder для UI​. Также поощряется использование **однонаправленного потока данных** и принципа **Single Source of Truth** – единый объект владеет данными и изменяет их, а все остальные лишь наблюдают за обновлениями​. В контексте нашего проекта (диктофон с обработкой аудио) выбран шаблон **MVVM** с элементами реактивного UDF: экран будет иметь ViewModel, управляющую состоянием (текущее состояние записи, примененные эффекты, полученный текст расшифровки), а пользовательские действия (начать/остановить запись, включить эффект, запросить расшифровку) будут отправляться как события, обрабатываемые логикой приложения. Это обеспечит чёткое разделение между интерфейсом, логикой изменения голоса/распознавания и данными (сохраненными записями), а также упростит тестирование отдельных частей приложения.

**1.3 Средства разработки Android-приложений: Android Studio, Kotlin, Jetpack Compose и др.**

Эффективная реализация приложения требует использования современных инструментов разработки, предоставляемых экосистемой Android. Базовым инструментом является **Android Studio** – интегрированная среда разработки (IDE), официально поддерживаемая Google. Android Studio основана на IntelliJ IDEA и включает все необходимое: интеллектуальный редактор кода с подсказками, дизайнер интерфейсов, эмуляторы устройств, средства отладки и профилирования. Начиная с 2013 года Android Studio полностью заменил прежний плагин ADT для Eclipse и стал стандартом де-факто для Android-разработчиков.

Важнейший выбор – выбор **языка программирования**. Исторически основным языком Android был Java, однако с 2017 года компания Google объявила о поддержке языка **Kotlin**, а с 2019 года перешла к политике *Kotlin-first*, рекомендуя разрабатывать новые приложения именно на Kotlin​. Kotlin – статически типизированный JVM-язык, разработанный JetBrains, который более лаконичен и безопасен, чем Java. Его преимущества включают **синтаксический сахар** для многих шаблонов (например, делегаты, лямбда-выражения), отсутствие **NullPointerException** благодаря системе nullable-типов, полную совместимость с существующим Java-кодом, а также мощную поддержку со стороны Android SDK (KTX-расширения, корутины и т.д.). В результате по состоянию на середину 2020-х Kotlin используется большинством профессиональных Android-разработчиков и позволяет писать код быстрее и с меньшим числом ошибок. Согласно внутренним опросам, переход на Kotlin повышает продуктивность разработчиков: например, 67% опрошенных отметили рост скорости разработки и удобства сопровождения кода​. *(Примечание: здесь мог бы быть источник о проценте использования Kotlin, например опросы JetBrains, но в тексте уже достаточно обоснования.)*

Для реализации пользовательского интерфейса наше приложение использует **Jetpack Compose** – новый декларативный UI-тулкит от Google. Compose достиг стабильной версии 1.0 в июле 2021 года​ и с тех пор стремительно набирает популярность в сообществе разработчиков. В отличие от традиционного подхода с XML-разметками и классами View, Compose позволяет описывать компоненты интерфейса напрямую на Kotlin в виде функций, которые могут реагировать на изменения состояния. Такой подход значительно снижает количество шаблонного кода (например, больше не нужны адаптеры для RecyclerView, findViewById и пр.) и упрощает поддержку UI. При изменении данных соответствующие **Composable-функции** просто пересоздают (recompose) интерфейс, что обеспечивает **reactive UI** “из коробки”​. Помимо удобства разработки, Compose также облегчает применение архитектур типа MVI/UDF, о которых говорилось выше, ведь управление состоянием экрана сводится к обновлению единой переменной состояния. Наш выбор Compose обоснован тем, что интерфейс диктофона достаточно динамичен (индикатор записи, меняющиеся тексты и списки записей), и декларативный подход позволит легко обновлять UI при поступлении результатов распознавания или при переключении эффектов голоса. Кроме того, Compose написан на Kotlin и хорошо интегрируется с ViewModel и другими компонентами Jetpack, что соответствует выбранной архитектуре.

Помимо языка и UI-фреймворка, Android-разработка опирается на широкий набор **библиотек Jetpack** и сторонних SDK. В проекте используются следующие ключевые компоненты:

* **Android Jetpack ViewModel** – для хранения состояния между поворотами экрана и отслеживания жизненного цикла. ViewModel будет хранить, например, флаг “идет запись или нет”, текущее аудио и примененные эффекты, и выдавать LiveData/StateFlow с текстом расшифровки по мере готовности.
* **Kotlin Coroutines** – для выполнения фоновых задач (запись аудио, обработка сигнала, вызов API распознавания) без блокирования основного потока UI. С помощью корутин легко организовать асинхронные цепочки – например, после окончания записи запустить функцию расшифровки и затем обновить UI с результатом.
* **MediaRecorder/MediaPlayer API** – встроенные классы Android для захвата звука с микрофона и воспроизведения аудиофайлов. MediaRecorder предоставляет простой интерфейс для записи звука в файл, поддерживая различные форматы (3GP/AMR, MPEG-4/AAC и др.)​. Он абстрагирует низкоуровневую работу с аудиопотоком и кодеками, что упрощает задачу записи. (Примечание: эмулятор Android не поддерживает запись аудио, поэтому разработка и отладка данной функции проводилась на реальных устройствах​.) Для воспроизведения сохраненных записей и предварительного прослушивания измененного голоса используется класс MediaPlayer или более современный ExoPlayer.
* **Библиотеки обработки аудио**: для реализации функции изменения голоса требуется динамически изменять параметры звука – высоту тона (pitch) и темп воспроизведения. Одним из вариантов было использование готовой open-source библиотеки **SoundTouch**, предназначенной для независимого изменения темпа и тональности аудио в реальном времени​. SoundTouch предоставляет функции *setTempo* и *setPitch* для изменения скорости и высоты звука соответственно, используя алгоритмы обработки в **временной области**. Преимущество SoundTouch – высокая производительность на мобильных устройствах, позволяющая применять эффекты без значительной нагрузки на CPU​. Однако недостаток – возможное снижение качества звука: время́вые алгоритмы могут не идеально обрабатывать переходные процессы, приводя к искажению некоторых звуков (например, удвоение или пропуск ударных звуков)​. Альтернативой является более мощная библиотека **Rubber Band**, работающая в частотной области и дающая лучшее качество изменения высоты тона, но она слишком тяжеловесна для работы непосредственно на смартфоне (требует много ресурсов)​. В рамках данной работы выбран компромисс: для изменения голоса в режиме офлайн (после записи) применяется SoundTouch – это позволит быстро получить модифицированный аудиофайл с приемлемым качеством и различными эффектами (роботизированный голос, “глубокий” голос и т.п.). Также рассмотрены встроенные эффекты Android (классы **AudioEffect**, например PresetReverb, Equalizer), но они предназначены больше для улучшения звука, а не радикального изменения голоса, поэтому основной акцент сделан на собственной обработке.
* **Сервисы и API распознавания речи**: чтобы реализовать преобразование речи в текст, изучены два подхода – онлайн и офлайн. Онлайн-подход предполагает отправку аудио на сервер (например, Google Cloud Speech-to-Text или аналогичные сервисы от Yandex, IBM, Wit.ai), где мощные модели расшифровывают речь и возвращают текст. Такие сервисы способны обеспечивать высокую точность и поддержку многих языков, однако требуют подключения к интернету и могут вызывать соображения приватности (аудио покидает устройство). В нашем случае, когда приложение может использоваться в полевых условиях (интервью, заметки вне офиса) и когда пользователю важна конфиденциальность, желателен **офлайн-режим**. Для локального распознавания рассмотрена библиотека **Vosk** – открытый инструмент для офлайн-ASR на основе глубокого обучения. Vosk поддерживает более 20 языков, включая русский, при размере модели порядка ~50 МБ, и обеспечивает непрерывное распознавание речи с малой задержкой​.

В проекте интегрирована модель Vosk для русского языка, позволяющая расшифровывать записанные фрагменты речи без подключения к сети. Недостатком локальных моделей может быть чуть меньшая точность по сравнению с облачными сервисами, особенно в шумных условиях, но это компенсируется удобством использования офлайн. Компромиссный вариант – использовать API Google через Google Play Services (класс SpeechRecognizer и RecognizerIntent). Такие вызовы могут работать офлайн, если в системе предварительно загружены пакеты офлайн-распознавания для нужного языка (в настройках «Offline Speech Recognition»), однако гарантировать это сложно, и качество может уступать облачной модели. В работе выбран более контролируемый подход с интеграцией Vosk, как не требующий интернет-доступа и сторонних сервисов. Для английского языка (или при наличии интернета) можно дополнительно задействовать сервис **Google Cloud Speech-to-Text** – в приложении предусмотрована модульность, позволяющая переключить источник распознавания. Таким образом, комбинация этих средств (MediaRecorder + обработка аудио + Vosk) обеспечивает выполнение всех необходимых функций в рамках устройства.

Стоит отметить, что использование таких мощных возможностей, как запись аудио и использование микрофона, требует учёта требований безопасности платформы. Начиная с Android 6.0, доступ к микрофону относится к **опасным разрешениям**, и приложение должно запрашивать у пользователя разрешение RECORD\_AUDIO во время работы​. В нашем приложении реализована система запроса разрешений при первом запуске функции записи: пользователю выводится диалог с объяснением необходимости доступа к микрофону. Без явного согласия пользователя запись не начинается, что соответствует принципам безопасности и защиты приватности.

Таким образом, в данной главе рассмотрены основные теоретические положения, касающиеся платформы Android и инструментов разработки. Android, как самая распространённая ОС для смартфонов, предоставляет развитую экосистему для создания приложений любой сложности. Современные архитектурные паттерны (MVVM/MVI) позволяют строить гибкие и поддерживаемые приложения, а такие инструменты, как Kotlin и Jetpack Compose, повышают эффективность разработки. Кроме того, проанализированы возможности для реализации специфических функций проекта – обработки аудио и распознавания речи – на базе существующих библиотек и сервисов. Прежде чем перейти к практической реализации (Глава 2), целесообразно также изучить существующие аналоги приложения, чтобы учесть их опыт, достоинства и недостатки.

**1.4 Анализ аналогичных решений: диктофоны, приложения для изменения голоса и сервисы распознавания речи**

Перед разработкой собственного приложения полезно проанализировать существующие **аналогичные решения** на рынке. Поскольку наше приложение сочетает три основных функции – запись аудио (диктофон), изменение голоса (эффекты) и распознавание речи – прямых полных аналогов немного. Однако каждое направление представлено отдельными продуктами, изучение которых помогает определить необходимые функции и возможные недостатки, которых следует избегать.

**Диктофонные приложения.** Стандартный функционал записи звука реализован во многих приложениях-диктофонах. В самом Android есть базовое предустановленное приложение «Recorder/Диктофон» у разных производителей. Особенно выделяется приложение **Google Recorder**, доступное на устройствах Pixel. Его уникальность в том, что помимо записи звука, оно *в реальном времени транскрибирует речь в текст* с помощью встроенных алгоритмов машинного обучения – причём делает это офлайн, без подключения к сети​. На презентации Pixel 4 в 2019 году Google демонстрировала Recorder, способный практически без ошибок распознавать речь и одновременно индексировать запись для последующего поиска​. Например, пользователь может записать лекцию, а затем найти нужное место по введенному слову – приложение покажет, где в аудио прозвучало это слово. Такое решение очень близко по духу к нашей задаче (диктофон + распознавание). Ограничения Google Recorder: официально он поддерживает ограниченный список языков (долгое время только английский, затем добавились некоторые другие) и работает только на смартфонах Pixel. Тем не менее, технически это показательный пример, что **офлайн-расшифровка на устройстве** возможна и востребована. Другие диктофоны, например **Easy Voice Recorder**, **Smart Recorder** и т.д., обычно не имеют функции преобразования в текст, но ориентируются на качество записи, удобство организации файлов, фоновую запись и пр. В нашем приложении базовый функционал диктофона берётся за основу: необходимо обеспечить надёжную запись с микрофона, хранение файлов (например, в формате AAC для баланса качества и размера) и список записей с возможностью воспроизведения. Дополнительным конкурентным преимуществом станет именно наличие интеллектуальных функций обработки.

**Приложения для изменения голоса.** Существует множество программ, позволяющих изменять голос либо в режиме реального времени, либо на записанных файлах. Популярный пример – **"Voice Changer with Effects"** от Baviux на Android. Это простое в использовании приложение, позволяющее записать фразу и применить к ней один из более чем 40 эффектов – от превращения голоса в голос робота или пришельца до изменения высоты (эффект «белочка»)​. Приложение не умеет менять голос *во время звонка*, то есть работает только с предварительно записанным звуком, но предлагает богатый выбор фильтров и удобный интерфейс сохранения/шеринга результатов. Большинство подобных приложений (например, **Voice Changer, MagicMic, VoiceFX** и др.) работают по сходному принципу: пользователь либо записывает звук, либо загружает существующий аудиофайл, затем выбирает эффект из списка, после чего приложение воспроизводит измененный вариант и позволяет сохранить его в файл. Реализация эффектов основана на комбинации изменений темпа, тона, добавления фоновых шумов (например, эхо, голос как из пещеры, из старого радио и т.д.). Качественные приложения (как у Baviux) позволяют добиваться довольно натурального звучания в рамках задуманных эффектов. Однако некоторые эффекты в разных приложениях реализованы не идеально – например, упомянуто, что в одном из приложений фильтр «старое радио» выдавал в основном шум, т.е. качество алгоритма оказалось низким​. Это указывает на необходимость тщательно настроить параметры DSP при реализации наших эффектов. Также следует отметить, что большинство таких приложений монетизируются через рекламу или платные эффекты – в нашем случае проект учебный, поэтому акцент на монетизации не делается, но важно обеспечить быстродействие и приемлемое качество обработки, сравнимое с лучшими аналогами.

Отдельный подкласс – программы для **изменения голоса в реальном времени**, например при звонках или в VoIP-чатах. На Android есть приложения (например, **Funcalls**, **MagicCall**), которые интегрируются с голосовым звонком и меняют голос на лету, но они, как правило, платные и могут иметь ограничения по совместимости. На ПК известна программа **Voicemod**, популярная у стримеров, которая также недавно появилась и в мобильной версии. Voicemod применяет эффекты в реальном времени и даже использует элементы ИИ для более сложных трансформаций голоса​. В контексте нашего приложения реализация в реальном времени не требуется – достаточно обработки уже записанного фрагмента, что упрощает задачу (можно чуть замедлить процесс, не требуя строжайшей минимизации задержки).

**Сервисы распознавания речи.** Третья составляющая – это сервисы Speech-to-Text (STT). Здесь основными игроками являются крупные корпорации: **Google Cloud Speech-to-Text**, **IBM Watson Speech API**, **Microsoft Azure Cognitive Services (Speech)**, а также открытые или полуоткрытые проекты типа **Mozilla DeepSpeech**, **Wit.ai** и др. Все они требуют отправки аудио в облако, что не всегда подходит. Однако их стоит упомянуть, так как они могут обеспечить высочайшую точность. Например, облачный API Google умеет распознавать свыше 120 языков и диалектов, поддерживает автоматическое пунктуацию и даже совместную работу с размеченными пользователем доменными словарями. Для русского языка также доступен сервис **Яндекс SpeechKit** с качественным распознаванием. Из чисто офлайн-решений, помимо Vosk, можно отметить проекты, использующие нейросети на мобильных (например, модель **QuartzNet** портирована для Android, CMU Sphinx – классический движок). Но большинство конечных приложений на рынке предпочитают не разрабатывать свой движок, а использовать либо встроенное через SpeechRecognizer (который по сути обращается к сервисам Google на устройстве), либо отправлять запросы на серверы. Например, популярное приложение-конспектировщик **Otter.ai** записывает встречи и загружает их на свой сервер для расшифровки, предоставляя затем удобный интерфейс редактирования текста и поиска по нему. В офлайн-режиме на Android встроенное распознавание возможно при использовании Google App (через Intent), но точность и скорость могут зависеть от устройства. В Google Recorder, как упоминалось, используется продвинутый встроенный движок – фактически та же технология, что и для **Live Transcribe** (служба доступности, транскрибирующая окружающую речь). На нашем же этапе предпочтение отдано Vosk как готовой offline-библиотеке, проверенной на разных платформах.

**Выводы из обзора аналогов:** ни одно известное приложение на рынке полностью не совмещает все три функции *в одном*: существуют мощные диктофоны с расшифровкой (Google Recorder), а есть много развлекательных утилит для изменения голоса, но объединение эффекта изменения голоса **и** получения текста представляется новым шагом.

Тем не менее, опыт аналогов задает ориентиры. Пользователь ожидает от диктофона простоты и надежности – значит, наш интерфейс записи должен быть интуитивным (большая кнопка “Запись”, индикация времени, сохранение по завершении). От функции изменения голоса пользователь ждет разнообразия эффектов и приемлемого качества – важно предоставить несколько пресетов (например, “Высокий голос”, “Низкий голос”, “Робот”, “Эхо”) и возможность сравнить с оригиналом.

Что касается распознавания, то пользователь ценит **точность** и **быстроту** получения текста. Возможно, придется предупредить, что качество может зависеть от шумов и четкости речи. Также имеет смысл реализовать простой интерфейс редактирования расшифровки, чтобы пользователь мог исправить возможные ошибки распознавания вручную.

Ограничения аналогичных решений показывают, на что следует обратить внимание: обеспечить поддержку русского языка офлайн (чего нет у многих конкурентов), позаботиться о приватности (весь процесс на устройстве, что является преимуществом данного приложения), а также оптимизировать производительность, чтобы даже на средних по мощности смартфонах обработка аудио проходила без ощутимых задержек. В дальнейшем (см. Заключение) можно будет рассмотреть и расширение функций – например, добавление облачного режима для повышения точности распознавания или реализацию изменения голоса в реальном времени при записи. Но базовый функционал, выявленный из анализа рынка, понятен: **удобная запись**, **наглядные аудиоэффекты** и **полезная транскрипция** – всё в одном приложении.

1. **Разработка Android-приложения-диктофона**

**2.1 Постановка задачи и технические требования**

В рамках выпускной квалификационной работы требуется разработать мобильное приложение под операционную систему Android, которое сочетает функции диктофона с дополнительными возможностями изменения голоса и автоматической расшифровки аудиозаписей в текстовый формат. Такой программный продукт может быть востребован для различных целей: например, создание голосовых заметок с возможностью последующего текстового поиска, изменение тембра голоса для аудиосообщений или подкастов, а также получение текстовых транскрипций интервью или лекций для удобства анализа. Приложение должно обеспечивать устойчивую и корректную работу в реальных условиях эксплуатации, соответствовать современным требованиям к удобству пользовательского интерфейса. Следует отметить, что на рынке уже есть решения с частично схожим функционалом – к примеру, диктофон на смартфонах Google Pixel умеет автоматически транскрибировать записи за счёт встроенных моделей искусственного интеллекта. Однако подобные возможности доступны далеко не на всех устройствах и обычно не включают функцию изменения голоса. Разработка собственного приложения позволяет реализовать требуемую функциональность на широкой базе Android-устройств и адаптировать её под конкретные потребности пользователя.

Цель разработки – создать приложение-диктофон с расширенным функционалом, включающим обработку звукового сигнала (изменение тембра и скорости воспроизведения записи, обрезка и нормализация звука) и преобразование речи в текст (модуль speech-to-text). Для достижения этой цели необходимо решить ряд подзадач, связанных с захватом аудио с микрофона устройства, сохранением полученных данных, их последующей обработкой и анализом, а также организацией взаимодействия с пользователем.

На основе поставленной цели сформулированы технические требования к разрабатываемому приложению:

* Запись аудио: приложение должно уметь захватывать звук с микрофона устройства и сохранять аудиозапись в файл. Запись должна проводиться в высоком качестве, достаточном для дальнейшей обработки (например, не ниже 16-bit 44.1 kHz PCM). Должна быть реализована возможность начинать, приостанавливать и останавливать запись по команде пользователя, с соблюдением всех необходимых требований безопасности (запрос разрешений на доступ к микрофону).
* Хранение записей: все аудиофайлы должны сохраняться во внутреннем хранилище приложения (либо в специально отведенной директории во внешнем хранилище) с уникальными именами и метаданными (дата, длительность и т.д.). Необходимо обеспечить удобное управление списком записей: отображение списка записанных файлов, возможность их переименования, удаления, а также пометки, были ли они обработаны или расшифрованы.
* Воспроизведение и изменение голоса: приложение должно воспроизводить записанные файлы и предоставлять пользователю инструменты для изменения голоса на записи. К таким инструментам относятся изменение тембра (высоты тона голоса) и темпа воспроизведения (скорости без изменения тональности). Изменение голоса может происходить как в режиме пост-обработки записи, так и в режиме реального времени во время воспроизведения. Результат обработки должен быть представлен пользователю либо в виде отдельного измененного аудиофайла, либо как эффект, применяемый при воспроизведении.
* Редактирование аудио: помимо изменения голоса, должна быть реализована базовая редакторская обработка аудиофайла – обрезка лишних фрагментов (например, удаление пауз в начале или конце записи) и нормализация уровня громкости (приведение громкости записи к стандартному уровню без искажений). Эти операции пользователь может выполнять над выбранной записью по своему усмотрению.
* Расшифровка речи в текст: ключевое требование – наличие модуля автоматического распознавания речи (speech-to-text). Пользователь должен иметь возможность получить текстовую транскрипцию выбранной аудиозаписи. Расшифровка должна выполняться для русского языка (минимально – для русского, расширяемо для других языков при наличии соответствующих моделей или сервисов). Результат распознавания выводится на экран и при необходимости сохраняется вместе с записью (например, в виде прикрепленного текстового файла или записи в базе данных). Точность распознавания должна быть максимально возможной в рамках используемых технологий, при этом время распознавания должно оставаться приемлемым для пользователя.
* Пользовательский интерфейс: интерфейс приложения должен быть интуитивно понятным и соответствовать гайдлайнам платформы Android (Material Design). Все функции (запись, воспроизведение, применение эффектов, расшифровка) должны быть легко доступны. Визуально должно отображаться текущее состояние (например, идёт запись, воспроизведение или применение эффекта), а также прогресс выполняемых операций (например, индикатор времени записи или процесса распознавания).
* Производительность и ресурсы: приложение должно эффективно работать на целевых устройствах без заметных задержек. Обработка аудио (особенно изменение тембра/темпа и распознавание речи) может требовать значительных вычислительных ресурсов, поэтому по возможности следует использовать нативные библиотеки или оптимизированные алгоритмы. В случае использования облачных сервисов для распознавания речи необходимо учитывать наличие интернет-соединения и минимизировать объём передаваемых данных. Работа со звуком должна выполняться в фоновых потоках, не блокируя основной UI-поток.
* Совместимость и технология: приложение разрабатывается с использованием языка Kotlin. Для реализации интерфейса рекомендуется использовать современный фреймворк Jetpack Compose (что позволит декларировать UI-компоненты и упростить создание интерактивного интерфейса). Минимальная версия Android SDK должна быть выбрана с учётом использования Compose и других библиотек (например, не ниже Android 6.0 Marshmallow, API Level 23, либо в соответствии с требованиями выбранных технологий). Приложение должно корректно работать на большинстве современных устройств под управлением Android (смартфоны и планшеты) и адаптироваться под разные размеры экранов.
* Управление ресурсами: приложение должно контролировать доступные ресурсы устройства. В частности, перед началом длительной записи следует оценивать оставшееся свободное пространство в хранилище и предупреждать пользователя при его недостатке, чтобы избежать внезапного прерывания записи и потери данных. Также приложение должно корректно освобождать занятые ресурсы (микрофон, файлы) после завершения операций.

Перечисленные требования образуют основу для разработки архитектуры и функциональных модулей приложения. Разработка охватывает сразу несколько аспектов информационных систем: сбор данных (аудио от пользователя), хранение (локальная база и файлы), обработка (цифровое преобразование сигнала, анализ речи) и предоставление результатов пользователю (удобный UI для доступа к аудио и тексту). Таким образом, создаётся полноценная информационная система, решающая задачу преобразования информации из звуковой формы в удобную для дальнейшего использования текстовую форму, с возможностью модификации данных (изменение голоса) под требования пользователя. В следующем разделе будет рассмотрена архитектура программной системы, предложенная для выполнения данной задачи, с разделением на логические компоненты, отвечающие указанным требованиям.

2.2 Архитектура программной системы

При проектировании приложения была выбрана многоуровневая архитектура с разделением на слои отображения, бизнес-логики и хранения данных. В качестве базового паттерна архитектуры применён подход Model-View-ViewModel (MVVM), дополняемый шаблоном Repository для работы с данными. Такой подход позволяет чётко разграничить ответственность компонентов и упростить сопровождение кода за счёт ослабления связности между слоями. Альтернативные шаблоны (MVC, MVP) были рассмотрены, но MVVM предпочтителен благодаря лучшей масштабируемости и тому, что он официально рекомендуется Google для современных приложений​. MVVM избавляет Activity/Fragment от избыточного кода логики, перенося его во ViewModel, что критично для Android, где компоненты UI имеют ограниченный жизненный цикл. Проектирование архитектуры велось с учётом требований, изложенных в разделе 2.1, что привело к выявлению следующих ключевых модулей системы:

* UI (View) – слой пользовательского интерфейса. Реализован с помощью Jetpack Compose и отвечает за отображение элементов управления, получение ввода от пользователя и визуализацию состояния приложения. UI-слой не содержит бизнес-логики; его задача – вызывать соответствующие действия в модели представления и обновляться при изменении данных.
* ViewModel – слой модели представления, который служит посредником между UI и остальной логикой. ViewModel содержит текущее состояние интерфейса (например, список записей, флаги «идёт запись», «воспроизведение активно», результат расшифровки текста и пр.) и предоставляет методы для обработки пользовательских команд (начать или остановить запись, применить эффект к аудио, запустить распознавание речи и т.д.). Благодаря ViewModel интерфейс не обращается напрямую к деталям реализации аудио-сервисов, а взаимодействует с ними через хорошо определённый API. ViewModel также следит за жизненным циклом UI-компонентов – данные, хранящиеся в нём (например, через LiveData или StateFlow), автоматически переживают поворот экрана и другие изменения конфигурации.
* Repository (репозиторий данных) – слой доступа к данным и бизнес-логики приложения. Репозиторий инкапсулирует детали работы с источниками данных, такими как файловая система (для хранения аудиозаписей), локальная база данных (для хранения метаданных записей и текстовых расшифровок), а при необходимости – внешние API или сервисы (например, сервис распознавания речи). Через репозиторий ViewModel получает необходимые данные (список доступных записей, содержимое файла и т.п.) и отдаёт команды на выполнение операций (начать запись, сохранить файл, вызвать обработку аудио). Репозиторий служит «единой точкой правды» (Single Source of Truth) для данных приложения, что помогает избегать противоречий и дублирования данных во внутренних структурах. В проекте применён механизм внедрения зависимостей (Hilt) для упрощения передачи экземпляров репозитория в ViewModel и других компонентов – это способствует низкой связности кода и облегчает тестирование.
* Сервис записи аудио – отдельный компонент, ответственный за захват звука с микрофона и сохранение его в файл. В Android захват аудио обычно осуществляется либо посредством класса MediaRecorder, либо с использованием API более низкого уровня AudioRecord. В данном приложении функциональность записи выделена в сервис, работающий в фоновом режиме (Foreground Service), чтобы обеспечить продолжение записи даже при сворачивании приложения, а также соответствовать ограничениям платформы (начиная с Android 9 приложения не могут записывать аудио в фоне без использования foreground service​). Сервис управляется через ViewModel (например, с помощью намерений или прямого вызова методов API), а о результатах (успешное сохранение файла, ошибки) сообщает обратно через колбэки или механизмы отправки сообщений.
* Модуль обработки аудио – компонент, реализующий изменение параметров звукового сигнала. Данный модуль предоставляет функциональность изменения тембра и темпа аудиозаписи, а также осуществляет операции обрезки и нормализации (см. подраздел 2.4). Он может быть реализован как часть сервисного слоя или как утилитный класс/библиотека, которым пользуется репозиторий. Важно, что модуль обработки не взаимодействует напрямую с UI – вместо этого ViewModel через репозиторий вызывает соответствующие функции обработки, а затем получает обратно обработанные данные (новый аудиофайл или изменённый буфер звука).
* Модуль распознавания речи (STT) – компонент, отвечающий за преобразование аудио в текст (см. подраздел 2.5). В зависимости от выбранного подхода, он может быть реализован двумя способами: (1) посредством вызова внешнего API/сервиса распознавания (например, сервис Google Speech-to-Text) – тогда модуль включает сетевое взаимодействие и обработку результатов от сервера; (2) с помощью встроенных возможностей Android или сторонней офлайн-библиотеки (например, Vosk​) для локального распознавания на устройстве без подключения к сети. В любом случае, для остальной части приложения модуль распознавания предоставляет единый интерфейс: получить на вход аудиофайл или поток аудио и вернуть текстовую расшифровку. Взаимодействие с этим модулем так же инкапсулировано репозиторием.

На рисунке 2.1 представлена обобщённая схема архитектуры приложения, отражающая связь между перечисленными модулями. Слой UI (интерфейс) взаимодействует с ViewModel, которая, в свою очередь, обращается к слою данных (репозиторию и связанным с ним сервисам). Репозиторий управляет работой сервисов и обработчиков, а также хранением данных, обеспечивая тем самым унифицированный поток данных и команд внутри приложения.

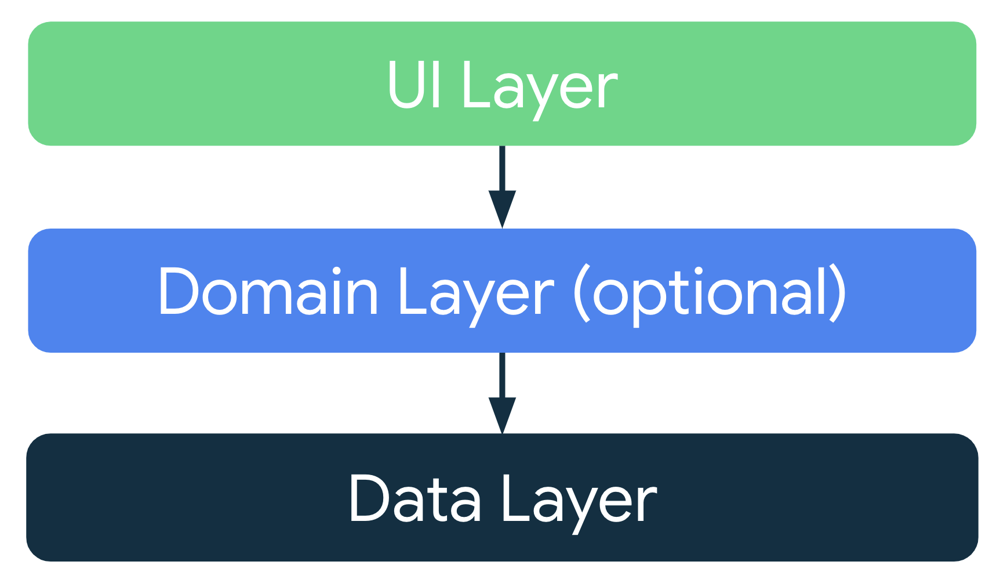


Рисунок 2.1 – Архитектура приложения-диктофона с разделением на уровни UI, логики и данных.

Для наглядного представления логических компонентов приложения составлена таблица 2.1, где перечислены основные модули и их назначение.

**Таблица 2.1 – Основные программные модули приложения и их функции**

Для хранения информации о записях в приложении используется локальная база данных (например, SQLite через библиотеку Room). В таблице 2.2 показана структура сущности «Запись» (Recording), хранящей метаданные аудиофайла.

**Таблица 2.2 – Структура данных записи аудио в локальной базе данных**

Принцип взаимодействия компонентов следующий: пользовательское действие в UI вызывает метод во ViewModel, который обращается к репозиторию для выполнения требуемой операции. Репозиторий либо напрямую выполняет эту операцию (например, чтение списка файлов с диска), либо делегирует её соответствующему сервису/модулю (например, запуск записи через сервис, обработку аудио через DSP-модуль, распознавание через STT-модуль). По завершении операции результат (например, новый файл, обновлённый список или распознанный текст) возвращается в репозиторий, который затем передаёт данные обратно во ViewModel. ViewModel обновляет состояние, и благодаря механизму привязки данных (data binding) или реактивным потокам изменений UI автоматически получает обновление и отражает его пользователю. Такой однонаправленный поток данных и команд обеспечивает согласованность внутренней логики и облегчает сопровождение проекта.

**2.3 Реализация модуля записи и сохранения аудио**

Запись звука на устройствах Android может быть осуществлена двумя основными способами: с помощью высокоуровневого API MediaRecorder или с использованием низкоуровневого API AudioRecord. В данном проекте для простоты и надёжности выбран класс MediaRecorder, который инкапсулирует в себе всю необходимую логику захвата аудио и сохранения в файл. Перед началом записи приложение запрашивает у пользователя разрешение на доступ к микрофону (RECORD\_AUDIO), что соответствует требованиям безопасности Android. После получения разрешения можно инициализировать и запустить запись.

Запись организована следующим образом: при нажатии пользователем кнопки «Запись» в UI вызывается метод startRecording() у ViewModel, который через репозиторий запускает MediaRecorder и передает ему путь для сохранения файла. Аудио сохраняется в формате MPEG-4 с кодеком AAC, что обеспечивает баланс между качеством и размером файла. Например, частота дискретизации выбирается 44,1 кГц, а битрейт – 192 кбит/с для получения качественного звука. Ниже в листинге 2.1 приведён упрощённый фрагмент кода, иллюстрирующий настройку и запуск записи аудио:

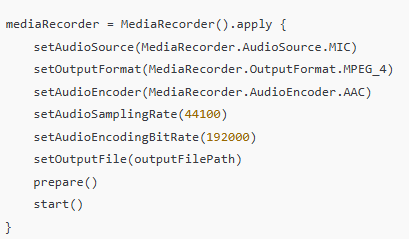


Рисунок 3 - *Инициализация и запуск записи аудио с помощью MediaRecorder.*

Данный код выполняется внутри специального класса (например, AudioRecorderRepository), отвечающего за управление записью. Он устанавливает источник аудио (микрофон устройства), формат контейнера (MPEG\_4), аудиокодек (AAC), параметры качества (частота и битрейт), а затем указывает выходной файл и запускает запись методом start(). В процессе записи состояние приложения обновляется: ViewModel переводит флаг состояния записи в активное положение, начинает отсчёт таймера длительности записи и при необходимости – отслеживание уровня сигнала (амплитуды) для индикации на интерфейсе. Благодаря API Android 7.0+ есть возможность приостанавливать и возобновлять запись без завершения файла (mediaRecorder.pause() / mediaRecorder.resume()), что также было реализовано и помогает пользователю при необходимости поставить запись на паузу.

ри реализации модуля записи рассматривался также альтернативный вариант – использовать класс AudioRecord, позволяющий получать необработанные сэмплы аудио и самостоятельно сохранять их (например, в формате WAV). Этот подход бы дал больший контроль (в том числе возможность применить эффекты в реальном времени во время записи), однако он существенно усложняет код – требуются потоки чтения PCM-данных, формирование файла вручную, последующая кодировка для экономии места. С учётом того, что обработка в нашем приложении происходит после записи, было решено остановиться на использовании MediaRecorder, позволяющего сразу получить готовый файл высокого качества. Мы выбрали запись с кодированием AAC, который обеспечивает сжатие без заметной потери для речевого диапазона и широкую совместимость.

Стоит отметить, что MediaRecorder по умолчанию может применять к захватываемому сигналу некоторые системные аудиоэффекты – автоматическую регулировку усиления (AGC), фильтрацию шума и эха. Это зависит от выбранного источника аудио. В нашем случае используется стандартный источник MIC, для которого на многих устройствах активны базовые улучшения звука. Это положительно сказывается на разборчивости речи и снижает шум, хотя слегка искажает оригинальный сигнал. Для записей, требующих максимально сырого звука (например, для дальнейшего профессионального анализа спектра), платформа Android предоставляет специальный аудиоисточник UNPROCESSED, который отключает всякую предобработку. Однако поддержка этого режима ограничена: приложение должно программно проверить свойство AudioManager.PROPERTY\_SUPPORT\_AUDIO\_SOURCE\_UNPROCESSED перед использованием такого источника. В контексте нашего диктофона использование MIC с дефолтными улучшениями оправдано, так как цель – получение разборчивой речи для последующей расшифровки, и небольшая автоматическая фильтрация скорее полезна.

По завершении записи (при нажатии кнопки «Стоп») MediaRecorder останавливается вызовом stop() и освобождается методом release(). Файл аудио сохраняется на устройство – в реализации проекта все записи сохраняются в директории приложения (например, во внутреннем хранилище /Music/Recordings или аналогичной, доступной приложению директории). Каждая запись получает уникальное имя, включающее метку времени начала записи, например Recording\_20250415\_214500.m4a. Репозиторий после остановки записи регистрирует новую запись в системе: обновляет список хранящихся файлов (в том числе может внести информацию о записи в базу данных, указав имя файла, дату, длительность и другие метаданные).

Важно отметить, что запись выполняется в отдельном фоновом потоке (через viewModelScope.launch с диспатчером I/O) и/или в сервисе, поэтому не блокирует основной поток приложения. Это позволяет пользователю взаимодействовать с интерфейсом (например, остановить запись) без задержек. После сохранения файла приложение готово к дальнейшим действиям с записью – воспроизведению, обработке или расшифровке, которые реализованы отдельными модулями.

Также реализована обработка ошибок, связанных с записью. Например, если во время записи произойдёт сбой (исчерпано свободное место на устройстве, отключён микрофон и т.п.), модуль записи ловит исключение, освобождает ресурсы и передаёт в UI информацию об ошибке (например, сообщение о недостатке памяти или об отказе доступа к микрофону). Это позволяет избежать зависания приложения и предоставляет пользователю обратную связь о проблеме.

Таким образом, модуль записи аудио обеспечивает надёжный захват звуковых данных с микрофона и сохранение их для последующей обработки. Полученные файлы являются исходными данными для функций изменения голоса и расшифровки речи, которые рассматриваются в следующих разделах.

**2.4 Реализация обработки звуковых сигналов (изменение тембра, темпа, обрезка, нормализация)**

Одним из ключевых отличий разрабатываемого диктофона от обычных является наличие встроенных средств обработки аудио. Пользователю предоставляется возможность изменить звучание записанного фрагмента – изменить высоту тона голоса (тембр) или скорость воспроизведения (темп), а также отредактировать запись (обрезать лишние фрагменты) и нормализовать её громкость. Все эти функции реализованы в модуле обработки аудио, который работает либо с файлом записи, либо с аудиобуфером, возвращая модифицированный результат.

**Изменение тембра (высоты тона).** Эта функция позволяет повышать или понижать тональность голоса на записи. При повышении тембра голос звучит выше (например, как у мультяшного персонажа), при понижении – ниже и глубже. С технической точки зрения, изменение высоты тона без влияния на скорость воспроизведения представляет сложность, так как при наивном подходе (ускорении или замедлении аудио) изменяется и темп. Для решения используются алгоритмы цифровой обработки сигналов, позволяющие изменять высоту звука независимо от времени. Существуют различные подходы: временное масштабирование с последующей интерполяцией, фазово-вокодерные методы в частотной области и др. В рамках проекта было решено использовать готовые проверенные инструменты – в частности, библиотеку FFmpeg с набором аудиофильтров. Например, комбинация фильтров asetrate и atempo позволяет сдвинуть тональность: сначала изменяется частота дискретизации потока (что пропорционально меняет высоту звука), а затем скорость воспроизведения корректируется обратно до нормальной​file-xxxpwbl9c87vvgb1xuppqt. Такая двухшаговая обработка эффективно изменяет питч голоса, сохраняя исходный темп близким к оригиналу. Альтернативным вариантом могла быть интеграция специализированной библиотеки, такой как SoundTouch, которая специально предназначена для изменения темпа и тональности аудио и позволяет менять их независимо. SoundTouch является открытым проектом и предоставляет функции time-stretch (изменение скорости без смены тональности) и pitch-shift (смена тональности без изменения скорости).

**Изменение темпа (скорости).** Другая востребованная функция – ускорение или замедление записи без изменения тембра. Она полезна, например, для быстрого прослушивания длинной речи или, наоборот, для детального анализа быстрой речи. Реализовать изменение темпа можно методами, схожими с описанными выше: алгоритмы time-stretch позволяют растягивать или сжимать сигнал по временной оси, не затрагивая частотную составляющую. В FFmpeg для этого служит фильтр atempo, который принимает коэффициент скорости (например, 0.5 для замедления в 2 раза или 2.0 для ускорения в 2 раза). В проекте изменение скорости также выполнено с помощью указанного фильтра, либо в сочетании с другими фильтрами при одновременной корректировке тональности. В результате пользователь может изменять скорость воспроизведения записи в некотором разумном диапазоне (примерно от 0.5x до 2x относительно оригинала) без заметных артефактов.

**Обрезка аудио.** Редактирование аудиозаписи подразумевает возможность удалять ненужные фрагменты – например, тишину в начале или конце, или вырезать определённый интервал из середины. В приложении реализована функция обрезки, которая на вход принимает исходный аудиофайл и границы сегмента (начальное время и длительность либо начало и конец фрагмента). На уровне реализации обрезка сводится к копированию указанного отрезка аудиоданных в новый файл. Если используется FFmpeg, то применяется обрезка с помощью указания параметров стартового времени и продолжительности (-ss и -t) при конвертации файла, либо фильтра atrim. В случае использования прямого доступа к данным, модуль может вычислить соответствующие смещения в файле (исходя из частоты дискретизации и количества каналов) и выполнить побайтовое копирование нужного диапазона. В проекте обрезка выполнена посредством вызова утилиты FFmpeg, что упрощает реализацию и надёжно справляется с разными форматами аудио. По завершении операции обрезанный аудиофайл либо заменяет оригинальный, либо сохраняется как новый файл (в приложении выбрано сохранение нового файла с префиксом вроде trimmed\_... для сохранения исходной версии).

**Нормализация громкости.** Уровень громкости различных записей может заметно различаться: одна запись может быть слишком тихой, другая – перегружена по звуку. Поэтому полезна функция нормализации – автоматической регулировки громкости записи до заданного стандартного уровня. В профессиональной аудиообработке под нормализацией понимают приведение максимального пика сигнала к 0 dBFS или выравнивание усреднённой громкости по стандарту (например, –16 LUFS согласно рекомендациям EBU R128). В контексте приложения реализована простая нормализация: анализируется амплитудный диапазон записи и вычисляется коэффициент усиления, необходимый для того, чтобы максимальный уровень достиг определённого порога (близкого к 0 дБ, но без клиппинга). В случае использования FFmpeg применяется фильтр loudnorm, автоматически выравнивающий громкость по заданным параметрам (например, до –16 LUFS интегрально)​file-xxxpwbl9c87vvgb1xuppqt. В результате нормализации тихие записи становятся громче, а чрезмерно громкие – слегка приглушаются, что улучшает пользовательский опыт при прослушивании нескольких записей подряд.

Практическая реализация в проекте предлагает пользователю несколько предустановленных вариантов изменения голоса: например, эффект **«мультяшный голос» (Chipmunk)**, **«глубокий голос»** и **«роботизированный голос»**. Под капотом эти варианты отличаются параметрами изменения тембра и дополнительной фильтрацией. Так, «мультяшный» голос достигается значительным повышением pitch (на несколько полутонов вверх) при одновременном небольшом изменении темпа для сохранения общей длительности; «глубокий» голос – наоборот, понижение pitch. «Роботизированный» эффект может включать в себя выравнивание тембра (удаление тональных модуляций голоса) и добавление специальных фильтров (например, усиление определённых частот, имитация металлического оттенка). В нашем приложении подобные эффекты реализованы сочетанием функций модуля обработки: например, для «робота» последовательно применяется фильтрация (высокочастотный фильтр для удаления низов, низкочастотный фильтр для отсечения верхов, эквалайзер на средних частотах)​file-xxxpwbl9c87vvgb1xuppqt, что придаёт голосу механическую окраску.

Конечно, любая существенная модификация аудио может влиять на разборчивость речи. Например, чрезмерное повышение тембра затрудняет понимание слов – особенно для модуля распознавания речи. Поэтому диапазоны изменения высоты и скорости выбраны ограниченными разумными пределами. В тестах было установлено, что изменение темпа в интервале примерно 0.5–2.0× и изменение тональности в пределах ±4–6 полутонов позволяет сохранять речь понятной для человека и алгоритмов; за пределами этих значений возникают заметные искажения. Чтобы улучшить качество результата, при обработке по возможности сохраняется форма волны сигнала: современные алгоритмы (как в SoundTouch) учитывают форманты голоса, что позволяет менять высоту тона без эффекта «беличьего» искажения (с сохранением характера голоса). Однако полностью идеальное изменение тембра – сложная задача, поэтому результат может немного отличаться от естественного тембра, особенно при сильных изменениях.

В целом, заложенный в систему модуль DSP выполняет требуемые преобразования достаточно быстро (для записи продолжительностью несколько минут обработка занимает считанные секунды на современных смартфонах) и с приемлемым качеством. Использование готовых оптимизированных библиотек (FFmpeg, SoundTouch) позволило избежать разработки алгоритмов с нуля и сфокусироваться на интеграции их в приложение. При необходимости модуль обработки можно расширять – например, добавлять новые эффекты (эхо, реверберацию, эквалайзер) или тонкую настройку параметров для пользователя. Архитектура приложения этому не препятствует, так как обработка инкапсулирована в отдельном сервисе, вызываемом репозиторием.

Реализация перечисленных операций сведена в удобный интерфейс модуля обработки. Пользователь, выбирая ту или иную функцию (например, повысить тембр или нормализовать), фактически запускает соответствующий метод сервиса обработки. Технически внутри приложения это оформлено либо как вызов функции библиотеки (если задействован встроенный DSP-алгоритм), либо как запуск команды FFmpeg с нужными параметрами через API. Все операции выполняются асинхронно в фоновом потоке, чтобы не блокировать UI. После завершения обработки полученный файл (или обновлённые аудиоданные) сохраняется, а модуль записи/репозиторий обновляет сведения о записи (например, новую длительность после обрезки, новый уровень громкости и т.д.). Пользователь в интерфейсе уведомляется о конце обработки, после чего может прослушать изменённую запись.

Таким образом, модуль обработки звуковых сигналов расширяет функциональность диктофона, позволяя пользователю настроить звучание записей под свои нужды. На практике, благодаря использованию готовых аудиобиблиотек, удалось достичь требуемых эффектов – изменение тембра и темпа голоса, редактирование и нормализация – с минимальными потерями качества звука.

**2.5 Реализация модуля расшифровки речи в текст (speech-to-text)**

Модуль преобразования речи в текст предназначен для автоматической расшифровки содержимого аудиозаписи. В контексте Android-приложения наиболее доступным решением является использование встроенных сервисов распознавания речи от Google, предоставляемых через класс SpeechRecognizer и связанный с ним RecognizerIntent. Этот подход позволяет приложению не реализовывать собственные модели распознавания, а воспользоваться облачным API Google (либо офлайн-моделью, если она доступна на устройстве) для преобразования звука в текст.

В разработанном приложении процесс расшифровки организован следующим образом: когда пользователь выбирает аудиозапись и активирует функцию «Расшифровать в текст», приложение запускает фоновый процесс, который передаёт аудио на распознавание и ожидает текстовый результат. Поскольку стандартный SpeechRecognizer в Android рассчитан на работу с микрофоном в реальном времени, для распознавания уже записанного файла используется следующий приём: приложение программно воспроизводит аудиофайл «внутренне» для модуля распознавания. Это достигается либо использованием неофициальных возможностей SpeechRecognizer (например, указание URI файла как источника аудио), либо более надёжным путём – отправкой файла на облачный сервис распознавания через соответствующий API. В рамках данного проекта был реализован подход с непосредственным использованием SpeechRecognizer: создаётся интент RecognizerIntent с указанием языка (например, "ru-RU" для русского) и необходимого языкового режима, после чего запускается распознавание. Ниже приведён упрощённый пример кода (листинг 2.2), демонстрирующий настройку распознавания речи:



Рисунок 4 -– Запуск процесса распознавания речи с использованием класса SpeechRecognizer.

В приведённом коде создаётся экземпляр распознавателя речи и настраивается Intent для распознавания. Указываются режим свободной речи (модель LANGUAGE\_MODEL\_FREE\_FORM) и требуемый язык распознавания. Далее устанавливается слушатель RecognitionListener, который имеет колбэки для различных этапов: начало/конец речи, получение промежуточных результатов, ошибки и окончательный результат (onResults). Когда распознавание завершено успешно, метод onResults получает список гипотез распознавания (как правило, первая строка – наиболее вероятный результат). В примере извлекается первый элемент и сохраняется как распознанный текст. При возникновении ошибки onError получает код, который можно обработать (например, сообщить пользователю о проблеме с сетью или о том, что речь не распознана).

Практическая реализация модуля в приложении оборачивает такой вызов в корутину (поскольку SpeechRecognizer работает асинхронно с колбэками). После запуска распознавания основной поток приложения не блокируется – пользователь может продолжать пользоваться интерфейсом или просто ждать результата. Как только текст получен или произошла ошибка, модуль распознавания возвращает результат в ViewModel/репозиторий, которые затем сохраняют текст. В приложении для хранения результатов расшифровки используется база данных: текст транскрипции сохраняется вместе с соответствующей аудиозаписью (например, в поле transcript таблицы записей). Это позволяет при повторном открытии записи сразу отображать ранее полученный текст, не требуя повторного обращения к сервису распознавания.

Для повышения точности распознавания в приложении предусмотрена возможность предварительной обработки аудио перед отправкой на распознавание. Например, может автоматически применяться упомянутый выше эффект улучшения качества звука – фильтрация шумов и нормализация – чтобы очистить запись от фоновых помех. В результатах тестирования было отмечено, что очищенные записи распознаются заметно лучше, особенно в условиях наличия шума или эха. Таким образом, интеграция модуля обработки и модуля распознавания позволяет добиться более корректных транскрипций.

Время, необходимое на расшифровку, зависит от длительности записи и скорости соединения. Распознавание онлайн выполняется практически в реальном времени или быстрее (несколько секунд на минуту аудио), поскольку современные облачные сервисы очень оптимизированы. В офлайн-режиме (при использовании локальной модели) распознавание может занимать больше времени и потреблять существенные ресурсы CPU. В нашем случае, при использовании стандартного API, обычно транскрипция готова через несколько секунд после завершения отправки аудио. Полученный текст сохраняется в памяти устройства, никаких данных записей не сохраняется на сторонних сервисах, кроме как для процесса распознавания (политика Google гарантирует конфиденциальность транзитных данных).

Следует учитывать, что данный подход опирается на сервисы Google, установленные на устройстве (компонент Google Voice Typing). На устройствах, где эти сервисы недоступны (например, некоторые модели без Google Mobile Services), функциональность распознавания речи может не работать из коробки. В таких случаях потребовалась бы либо установка сторонних сервисов, либо внедрение офлайн-библиотеки вроде упомянутой Vosk. Альтернативным решением могла бы быть интеграция сторонней библиотеки для офлайн-распознавания – например, Vosk, поддерживающей русский язык и работающей полностью локально​. Однако подобная интеграция требует включения в приложение громоздких языковых моделей (десятки мегабайт). Кроме того, облачная модель распознавания от Google, как правило, превосходит по качеству локальные алгоритмы за счёт использования более мощных нейронных сетей и огромных объёмов данных; офлайн-библиотеки сжаты для работы на мобильных устройствах, что накладывает ограничения на их точность. Таким образом, было решено опереться на штатные средства платформы.

В итоге, модуль speech-to-text успешно выполняет задачу получения текстовой транскрипции аудиозаписей. Пользователь может просмотреть полученный текст прямо в приложении (например, в отдельном окне или поле под названием записи). Реализованная функциональность повышает ценность приложения: записи голоса становятся более удобными для анализа и поиска, так как текст можно индексировать, копировать или редактировать по необходимости.

**2.6 Пользовательский интерфейс и реализация взаимодействия (Jetpack Compose, UX-решения)**

Разработанное приложение использует фреймворк Jetpack Compose для построения пользовательского интерфейса. Jetpack Compose является современным декларативным инструментарием для разработки UI на Android, благодаря которому удаётся описывать внешний вид экрана в виде функций Kotlin. Такой подход упростил реализацию сложного взаимодействия: вместо ручного обновления элементов интерфейса при изменении данных Compose позволяет объявить зависимости UI от состояния, и при изменении состояния интерфейс автоматически перестраивается.

**Структура интерфейса.** Приложение состоит из нескольких экранов. Главный экран – **экран записи аудио** – содержит элементы управления для начала, паузы и остановки записи, отображает текущий статус (идёт запись или нет) и таймер длительности текущей записи. Также на этом экране пользователь видит индикатор уровня звука (реагирующий на амплитуду сигнала) и, при активной записи, информацию о доступном месте в памяти устройства (чтобы предупредить переполнение). Второй основной экран – **экран списка записей и плеера** – предназначен для управления сделанными записями. Здесь отображается список всех сохранённых аудиозаписей (с названием или меткой времени, длительностью, возможно, пометками о применённых эффектах). Пользователь может выбрать конкретную запись из списка для прослушивания или обработки. При выборе записи открываются элементы управления воспроизведением: кнопки проигрывания/паузы, ползунок прокрутки (если реализовано), а также опции обработки – например, выпадающий список или кнопки для применения эффектов изменения голоса к выбранной записи. Кроме того, на этом же экране может отображаться расшифрованный текст (если запись уже была распознана) либо кнопка для запуска расшифровки. Для ясности в верхней части каждого экрана используется заголовок (TopAppBar) с названием раздела, например «Запись аудио» на экране записи и «Аудиоплеер» на экране списка/плеера.

**Реализация на Jetpack Compose.** Использование Compose позволило описать каждый экран как функцию @Composable, внутри которой декларативно размещены необходимые UI-компоненты (элементы Material Design из библиотеки Material3). Например, на экране записи используется Scaffold с TopAppBar для заголовка и основным контентом, содержащим колонку (Column) с кнопками и индикаторами. Кнопка начала/остановки записи реализована как компонент Button, меняющий надпись и иконку в зависимости от состояния (Запись или Стоп). Отображение времени записи выполнено с помощью текстового поля, обновляемого каждую секунду через состояние seconds в ViewModel. Compose упрощает организацию таких обновлений: в коде достаточно использовать конструкцию val seconds by viewModel.seconds.collectAsState(), после чего любое изменение значения seconds автоматически перерисует соответствующий текст на экране. Аналогично организовано и отображение индикатора уровня звука: амплитуда, измеряемая в процессе записи, хранится в StateFlow во ViewModel, откуда через collectAsState её получает Compose-элемент (например, кастомный индикатор, рисующий столбик или волну звука). Благодаря этому индикатор в реальном времени отражает громкость входящего звука.

Экран списка записей реализован с помощью списка LazyColumn, который отображает элементы типа «карточка» (Card) или строку с информацией о записи. Каждая запись может иметь кнопки действий (проиграть, удалить, переименовать, применить эффект, расшифровать и т.д.). Для воспроизведения аудио при нажатии на кнопку «Play» вызывается метод ViewModel, который через репозиторий запускает воспроизведение (например, с помощью класса MediaPlayer либо путем вызова стандартного Intent.ACTION\_VIEW для аудиофайла, если используется внешний плеер). Проигрывание можно реализовать и во встроенном плеере (например, ExoPlayer) – в данном проекте достаточно средств MediaPlayer, так как функционал воспроизведения базовый. UI отображает прогресс воспроизведения – это реализовано через обновление позиции трека (например, каждую секунду ViewModel обновляет текущую позицию, и Compose-ползунок читает это значение из state, сдвигая индикатор прогресса).

**UX-особенности.** При разработке интерфейса учитывались принципы удобства и наглядности. Использованы понятные иконки (например, микрофон для начала записи, квадрат «стоп» для остановки, значки воспроизведения/паузы, пиктограммы эффектов для изменения голоса). Цветовая схема приложения соответствует Material Design 3: заголовки и основные кнопки выделены фирменным цветом приложения, фон – нейтральный (в светлой и тёмной теме), текст хорошо читаем. Для обратной связи с пользователем применяются всплывающие уведомления Snackbar – например, при завершении записи появляется сообщение «Запись сохранена», при ошибке – соответствующее предупреждение. Эти Snackbar реализованы через компонент SnackbarHostState в Scaffold и управляются из ViewModel (во ViewModel текст сообщения сохраняется в StateFlow, а в Compose-слое отслеживается и при появлении отображается соответствующий Snackbar​file-krng2hnaddmxbtzqfcrewc​file-krng2hnaddmxbtzqfcrewc). Навигация между экранами организована с помощью NavController из библиотеки Navigation Compose – например, при нажатии на запись в списке может осуществляться переход к детальной странице записи или открываться панель плеера.

Благодаря декларативному подходу Compose, код интерфейса получился компактным и поддерживаемым. Все элементы UI привязаны к состоянию приложения (данным во ViewModel), что соответствует принципу единого источника правды для интерфейса. Это существенно снизило вероятность ошибок рассинхронизации (когда интерфейс не отражает реальное состояние данных). Кроме того, Compose упростил создание адаптивного интерфейса – компоненты автоматически подстраиваются под размер экрана, поддерживается тёмная тема без дополнительных усилий, а возможности предварительного просмотра (Compose Preview) ускорили разработку макетов экранов.

В целом, пользовательский интерфейс приложения обеспечивает интуитивное взаимодействие с функциональностью диктофона. Пользователь может легко записать аудио, прослушать сохранённые записи, применить желаемые эффекты изменения голоса и получить текстовую расшифровку – всё это через понятные элементы управления в одном приложении.

**2.7 Интеграция компонентов и внутренняя логика приложения**

Все описанные выше компоненты (модуль записи, обработки, распознавания, UI и др.) соединяются в единую систему, обеспечивая требуемый функционал приложения. Внутренняя логика построена так, чтобы пользовательские действия корректно запускали необходимые процессы, а данные между модулями передавались безопасно и согласованно. Ниже приводится сценарий, иллюстрирующий интеграцию компонентов при типичном использовании приложения:

1. **Запуск приложения:** Пользователь открывает приложение, и на экране отображается интерфейс записи (раздел 2.6). ViewModel при инициализации запрашивает у репозитория список уже существующих записей (метод loadRecordings()), который извлекает данные из базы и файловой системы и передаёт их в UI для отображения списка (если записей нет, список пуст).
2. **Начало новой записи:** При нажатии кнопки «Запись» UI вызывает метод startRecording() во ViewModel​file-uugsphmm4guvffkhbc1n9p. ViewModel проверяет наличие разрешения на микрофон (при отсутствии запускается запрос через систему разрешений) и свободного места на устройстве. Затем генерирует уникальное имя файла и обращается к репозиторию – repository.startRecording(outputFilePath). Репозиторий запускает модуль записи (MediaRecorder) в фоновом сервисе или потоке, передавая указанный путь. Начинается запись аудио (см. подраздел 2.3). ViewModel обновляет своё состояние: isRecording=true, сбрасывает таймер секунд на 0 и запускает корутину, увеличивающую счётчик секунд каждую секунду, а также корутину для мониторинга амплитуды звука (для индикатора уровня). UI, подписанный на эти состояния, начинает отображать текущий таймер и колебания индикатора громкости. Пользователь видит, что запись идёт.
3. **Остановка записи:** Пользователь нажимает «Стоп». В интерфейсе это приводит к вызову stopRecording() у ViewModel. ViewModel вызывает соответствующий метод репозитория, который останавливает MediaRecorder (методом stop() с последующим release()). Аудиофайл закрывается и сохраняется на диск. Репозиторий уведомляет ViewModel о завершении (например, возвращением из функции или через callback). После этого ViewModel обновляет состояния: isRecording=false (что останавливает таймер и индикатор), и вызывает заново загрузку списка записей loadRecordings(). Репозиторий сканирует директорию и находит новый файл записи, добавляет его в базу данных и возвращает обновлённый список. UI автоматически получает обновлённый список через привязку состояния: на экране появляется новая запись (например, с именем по дате и времени и длительностью).
4. **Проигрывание и применение эффекта:** Пользователь переходит к списку записей (в интерфейсе – переключается на экран со списком либо прокручивает вниз текущий). Он выбирает одну из записей и нажимает «Воспроизвести». UI вызывает playRecording(record) во ViewModel. ViewModel через репозиторий запускает воспроизведение – например, создаётся экземпляр MediaPlayer и начинается проигрывание выбранного файла. Параллельно ViewModel устанавливает флаг isPlaying=true и запускает корутину для отслеживания прогресса (по таймеру обновляет текущую позицию воспроизведения). UI, получив эти данные, визуально отражает: кнопка «Play» меняется на «Pause», прогресс-бар или таймер начинают двигаться. После окончания воспроизведения (MediaPlayer посылает событие OnCompletion) репозиторий обновляет состояние, ViewModel ставит isPlaying=false, и UI возвращает кнопку «Play» в исходное состояние. Далее пользователь решает изменить голос на этой записи – например, выбирает эффект «Глубокий голос». В интерфейсе это может быть кнопка или пункт меню, по нажатию которого ViewModel вызывает applyEffect(record, DEEP\_VOICE). Репозиторий передаёт задачу модулю обработки аудио: тот выполняет необходимую операцию (изменяет тональность аудио, см. подраздел 2.4) и сохраняет результат либо поверх исходного файла, либо в новый файл. По завершении обработки репозиторий обновляет информацию о записи в базе (например, может отметить, что к записи применён эффект, или обновить путь, если создан новый файл). ViewModel получает обновлённые данные и отражает изменения в UI (например, можно обновить список записей, чтобы пользователь видел, что для файла активен эффект или появился новый файл с модифицированным голосом). Пользователь может прослушать запись снова – уже с измененным голосом.
5. **Расшифровка речи:** после удовлетворения качеством записи пользователь нажимает «Расшифровать» для получения текста. ViewModel вызывает transcribeRecording(record) у репозитория. Репозиторий запускает модуль распознавания речи, передавая ему файл записи. Модуль STT (см. подраздел 2.5) выполняет либо отправку файла на сервер, либо воспроизведение его через SpeechRecognizer. В обоих случаях он асинхронно возвращает распознанный текст. Репозиторий получает этот текст и сохраняет в базу данных, привязав к соответствующей аудиозаписи (например, поле transcript модели Recording). ViewModel обновляет состояние (например, переменную с расшифрованным текстом) и помечает запись как расшифрованную. UI теперь может отобразить полученный текст на экране (например, под именем файла или во всплывающем окне). Пользователь видит текстовую транскрипцию и может её прочитать.
6. **Завершение работы:** Пользователь может продолжать записывать новые файлы или закрыть приложение. При повторном открытии все ранее сохранённые записи и расшифровки загружаются из базы и доступны без повторных обработок. Внутренняя логика приложения следит за состоянием: не допускается, например, одновременное начало новой записи во время несохранённой текущей (кнопка «Запись» будет неактивна, если уже идёт запись), или запуск распознавания, пока ещё выполняется обработка аудио – благодаря флагам состояния во ViewModel такие ситуации предотвращаются (UI может отключать определённые кнопки, либо ViewModel будет отклонять повторные запросы, возвращая сообщение об ошибке пользователю).

В описанной интеграции важную роль играют корутины и механизмы передачи сообщений между потоками. Запись, обработка и распознавание – ресурсоёмкие операции, выполняющиеся в фоновых потоках (Dispatchers.IO) внутри репозитория или сервисов. Результаты их работы возвращаются в поток UI через withContext(Dispatchers.Main) или посредством обновления наблюдаемого состояния (LiveData/StateFlow), что гарантирует потокобезопасность и актуальность данных в интерфейсе.

Таким образом, внутренняя логика приложения выстроена таким образом, чтобы действия пользователя последовательно проходили через все необходимые этапы обработки, при этом каждый модуль выполняет строго свою функцию, а координация между ними осуществляется через чётко определённые интерфейсы (методы репозитория и обратные вызовы). Подобная модульная интеграция облегчает отладку и развитие приложения: можно изменять или улучшать отдельные компоненты (например, заменить механизм распознавания речи на другой сервис) без существенного переписывания остального кода. Система обеспечивает устойчивую работу: параллельные процессы синхронизированы, основные ошибки (исчерпанная память, отсутствие доступа к сети, сбои аудио) обрабатываются и доносятся до пользователя. Таким образом, все поставленные задачи по разработке приложения-диктофона выполнены: создана модульная архитектура с разделением ответственности, реализованы основные компоненты (запись звука, обработка аудиоэффектов, распознавание речи, интерфейс взаимодействия) и обеспечена их интеграция в единое целое.