

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра дискретной математики и информационных технологий

**РАЗРАБОТКА ANDROID-ПРИЛОЖЕНИЯ ДЛЯ
РАСПОЗНАВАНИЯ ЖЕСТОВ И ПОДДЕРЖКИ
КОММУНИКАЦИИ ЛЮДЕЙ С НАРУШЕНИЯМИ СЛУХА И
РЕЧИ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 421 группы
направления 09.03.01 — Информатика и вычислительная техника
факультета КНиИТ
Бабачкова Рауля Егвньевича

Научный руководитель
старший преподаватель

П. О. Дмитриев

Заведующий кафедрой
доцент, к. ф.-м. н.

Л. Б. Тяпаев

Саратов 2026

ВВЕДЕНИЕ

Актуальность темы исследования обусловлена необходимостью создания доступных и удобных средств коммуникации для людей с нарушениями слуха и речи. Несмотря на развитие цифровых технологий и широкое распространение мобильных устройств, проблема эффективного взаимодействия между пользователями жестового языка и слышащими людьми остаётся актуальной. Современные методы компьютерного зрения и машинного обучения позволяют реализовывать системы распознавания жестов непосредственно на мобильных устройствах, что открывает возможности для создания мобильных приложений, обеспечивающих преобразование жестового ввода в текстовую и звуковую форму.

Дополнительным фактором актуальности является развитие технологий мультимодального взаимодействия, позволяющих объединять различные способы ввода и вывода информации в рамках единой программной системы. Использование жестового ввода, синтеза речи, текстового взаимодействия и интеллектуальных диалоговых механизмов позволяет существенно расширить возможности коммуникации пользователей и повысить доступность цифровых сервисов для людей с ограниченными возможностями здоровья.

Целью работы является разработка Android-приложения для распознавания жестов, поддержки коммуникации и обучения пользователей жестовому взаимодействию.

Для достижения поставленной цели были поставлены следующие задачи:

- Изучить предметную область.
- Изучить проблему и аналоги.
- Изучить процессы и этапы разработки в среде Android Studio.
- Сформировать требования к разработке.
- Разработать мобильное приложение.
- Выполнить тестирования программы.

Объектом исследования является процесс организации коммуникации пользователей с применением жестового ввода на мобильных устройствах.

Предметом исследования являются методы и средства распознавания жестов руки на платформе Android, а также программные технологии, обес-

печивающие преобразование жестового ввода в текстовую и звуковую форму и поддержку обучающих сценариев взаимодействия.

В ходе выполнения работы использовались методы анализа предметной области, сравнительного анализа существующих программных решений, проектирования программных систем, а также методы компьютерного зрения и разработки мобильных приложений. Практическая реализация выполнена с использованием языка программирования Kotlin и среды разработки Android Studio. Для решения задачи распознавания жестов использовались технологии CameraX и MediaPipe, для хранения данных — Room, для озвучивания сообщений — TextToSpeech, а для реализации интеллектуального диалогового режима — OpenRouter API.

Бакалаврская работа состоит из введения, трёх глав, заключения, списка использованных источников и приложений. В первой главе рассмотрены предметная область и инструментальные средства разработки, используемые при создании мобильного приложения. Во второй главе проведён анализ существующих программных решений для поддержки коммуникации людей с нарушениями слуха и речи и сформулированы требования к разрабатываемой системе. В третьей главе представлены проектирование, реализация и тестирование мобильного приложения. В заключении приведены основные результаты выполненной работы и сформулированы выводы по результатам исследования, а также перспективы дальнейшего развития проекта.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В первом разделе работы была рассмотрена предметная область, связанная с разработкой мобильных приложений для поддержки коммуникации людей с нарушениями слуха и речи, а также проведён анализ технологий и инструментальных средств, необходимых для реализации системы распознавания жестов на мобильных устройствах.

Одним из наиболее перспективных направлений развития современных информационных технологий является компьютерное зрение, представляющее собой совокупность методов и алгоритмов, предназначенных для получения и анализа визуальной информации. Технологии компьютерного зрения активно используются в системах видеонаблюдения, медицинской диагностике, робототехнике, дополненной реальности и интеллектуальных пользовательских интерфейсах. Особое значение данные технологии имеют для создания средств коммуникации, позволяющих распознавать жесты и преобразовывать их в удобную для восприятия форму.

В рамках работы была рассмотрена задача распознавания жестов руки с использованием камеры мобильного устройства. Анализ существующих подходов показал, что классические методы обработки изображений, основанные на анализе контуров, границ объектов и поиске локальных признаков, обладают рядом ограничений при работе с жестами. Точность таких методов существенно зависит от условий освещения, положения руки относительно камеры и особенностей окружающего фона.

Для решения поставленной задачи были исследованы современные программные библиотеки и фреймворки, предназначенные для обработки изображений и распознавания объектов. Особое внимание было уделено библиотекам OpenCV и MediaPipe. OpenCV предоставляет широкий набор инструментов для обработки изображений и компьютерного зрения, однако требует самостоятельной реализации логики выделения признаков и интерпретации положения кисти пользователя.

В качестве основного средства распознавания жестов была выбрана библиотека MediaPipe. Данное решение позволяет в реальном времени определять положение кисти и формировать набор ключевых точек руки, характеризующих расположение ладони и пальцев. Использование готовой модели распознавания руки позволило сосредоточиться на реализации логики клас-

сификации жестов и обеспечить возможность работы приложения непосредственно на мобильном устройстве без необходимости постоянного обращения к удалённым вычислительным ресурсам.

Для разработки мобильного приложения была выбрана платформа Android как наиболее распространённая мобильная операционная система. Реализация программного продукта осуществлялась в среде Android Studio с использованием языка программирования Kotlin. Выбор Kotlin обусловлен его официальной поддержкой со стороны Google, современными средствами разработки и высокой степенью интеграции с Android SDK.

В процессе исследования были также рассмотрены дополнительные технологии, необходимые для реализации отдельных функциональных возможностей приложения. Для работы с камерой устройства использовалась библиотека CameraX, обеспечивающая удобный доступ к видеопотоку и поддержку различных моделей мобильных устройств. Для хранения пользовательских данных была выбрана библиотека Room, позволяющая организовать локальную базу данных с использованием архитектурных компонентов Android. Озвучивание сообщений реализовано средствами TextToSpeech, а для поддержки интеллектуального режима обучения и сценарных диалогов использовался сервис OpenRouter API, предоставляющий доступ к современным языковым моделям.

Проведённый анализ предметной области и существующих технологических решений позволил сформировать требования к разрабатываемой системе и определить программно-технический стек, необходимый для создания мобильного приложения. Выбранный набор технологий обеспечил возможность реализации распознавания жестов в реальном времени, организации мультимодального взаимодействия пользователей и поддержки обучающих функций в рамках единого мобильного приложения.

Во втором разделе работы был проведён анализ программных решений, предназначенных для поддержки коммуникации людей с нарушениями слуха и речи. Целью анализа являлось выявление существующих подходов к организации взаимодействия пользователей, определение сильных и слабых сторон аналогов, а также формирование требований к разрабатываемому мобильному приложению.

Для проведения сравнительного анализа были выбраны приложения,

ориентированные на преодоление коммуникативных барьеров между людьми с нарушениями слуха и слышащими пользователями. В качестве критериев сравнения рассматривались возможности распознавания речи, озвучивания текстовых сообщений, использования готовых фраз, поддержки жестового языка, наличия обучающих функций, автономности работы и удобства пользовательского интерфейса.

Одним из рассмотренных решений стало приложение «Яндекс.Разговор». Основной функционал данного приложения заключается в преобразовании устной речи собеседника в текстовую форму и озвучивании сообщений, вводимых пользователем. Решение обеспечивает удобное двустороннее взаимодействие между слышащими и слабослышащими людьми, однако не поддерживает распознавание жестового языка и не содержит средств обучения жестам.

Также было рассмотрено приложение Deaf Pad Pro, предназначенное для текстового общения пользователей. Программа предоставляет удобный интерфейс для обмена сообщениями и может использоваться в повседневных ситуациях общения. Вместе с тем функциональность приложения ограничивается текстовым взаимодействием и не предусматривает использование технологий компьютерного зрения и распознавания жестов.

Анализ приложения «Переводчик глухих» показал, что данное решение ориентировано преимущественно на преобразование речи в текст и обратно. Приложение позволяет упростить бытовое общение между пользователями, однако не предоставляет возможностей для изучения жестового языка и не использует визуальный жестовый ввод.

Отдельное внимание было уделено приложению «Переводчик ЖЯ», связанному с использованием жестового языка. Несмотря на ориентацию на соответствующую предметную область, данное решение в большей степени выполняет справочные функции и не обеспечивает полноценного распознавания жестов пользователя в реальном времени посредством камеры мобильного устройства.

Проведённый анализ показал, что существующие программные решения обычно сосредоточены на отдельных аспектах коммуникации: преобразовании речи в текст, озвучивании сообщений, использовании шаблонных фраз либо предоставлении справочной информации по жестовому языку. При этом

большинство рассмотренных приложений не поддерживает комплексный подход, объединяющий распознавание жестов, обучение пользователей, текстовое и голосовое взаимодействие в рамках единой системы.

На основании результатов анализа были сформулированы требования к разрабатываемому приложению. Было определено, что создаваемая система должна обеспечивать распознавание жестов в режиме реального времени с использованием камеры мобильного устройства, преобразование распознанных жестов в текстовую форму, озвучивание сообщений, работу с готовыми фразами, поддержку локального чатового взаимодействия, а также наличие обучающих режимов для изучения и практического применения жестов.

Таким образом, проведённый анализ существующих решений подтвердил актуальность разработки собственного мобильного приложения, объединяющего возможности коммуникации и обучения. Полученные результаты были использованы при формировании требований к системе и определении её функциональной структуры.

В третьем разделе работы рассматриваются вопросы проектирования и реализации мобильного приложения, предназначенного для распознавания жестов, поддержки коммуникации и обучения пользователей жестовому взаимодействию.

В процессе разработки были сформированы требования к системе, определена её архитектура, а также реализованы основные функциональные возможности приложения.

На основе анализа предметной области и существующих программных решений были сформированы функциональные и нефункциональные требования к разрабатываемому приложению.

К функциональным требованиям относятся: обеспечение распознавания жестов пользователя в реальном времени с использованием камеры мобильного устройства; преобразование распознанных жестов в текстовую форму; озвучивание сформированных сообщений средствами синтеза речи; поддержка системы готовых фраз для ускорения коммуникации; реализация локального текстового чата между пользователями; наличие обучающего режима для изучения жестов; поддержка интеллектуального диалогового режима с использованием внешнего API.

К нефункциональным требованиям относятся: обеспечение высокой про-

изводительности и минимальной задержки при обработке видеопотока; возможность автономной работы приложения без постоянного подключения к сети Интернет (за исключением ИИ-режима); обеспечение безопасности хранения пользовательских данных; расширяемость архитектуры и возможность добавления новых жестов и функциональных модулей.

Сформированные требования послужили основой для проектирования архитектуры программной системы.

Архитектура приложения построена по модульному принципу, что обеспечивает гибкость, расширяемость и удобство сопровождения программного продукта.

В состав системы входят следующие основные модули: модуль захвата видеопотока на основе библиотеки CameraX, обеспечивающий получение изображения с камеры устройства; модуль обработки изображения и выделения ключевых точек руки на базе MediaPipe; модуль классификации жестов, выполняющий интерпретацию положения пальцев и определение соответствующего жеста; модуль формирования текстовых сообщений на основе распознанных жестов; модуль синтеза речи, реализованный с использованием TextToSpeech; модуль хранения данных, основанный на библиотеке Room, обеспечивающий работу с локальной базой данных; модуль пользовательского интерфейса, реализующий взаимодействие пользователя с системой; модуль интеллектуального взаимодействия, использующий OpenRouter API для реализации сценарных диалогов.

Взаимодействие модулей организовано таким образом, что данные, полученные с камеры, последовательно обрабатываются и преобразуются в текстовую и звуковую форму, обеспечивая мультимодальное взаимодействие пользователя с приложением.

В рамках работы были реализованы ключевые функциональные возможности приложения.

Распознавание жестов осуществляется с использованием библиотеки MediaPipe, которая позволяет в реальном времени определять положение кисти и формировать набор ключевых точек. На основе полученных данных выполняется классификация жестов и их сопоставление с заранее заданным набором.

Формирование текстовых сообщений реализовано путём последователь-

ного накопления распознанных жестов и преобразования их в символы и слова. Это позволяет пользователю формировать осмысленные фразы.

Озвучивание сообщений осуществляется с помощью встроенного механизма TextToSpeech, что обеспечивает возможность голосового воспроизведения введённого текста.

Локальный чат реализован с использованием базы данных Room и позволяет пользователям обмениваться текстовыми сообщениями без подключения к сети Интернет.

Система готовых фраз предоставляет пользователю возможность быстрого выбора часто используемых выражений, что существенно ускоряет процесс коммуникации.

Обучающий режим предназначен для освоения жестов и включает в себя тренировку распознавания и практическое взаимодействие пользователя с системой.

Интеллектуальный режим реализован с использованием OpenRouter API и позволяет организовать диалоговое взаимодействие пользователя с системой в рамках заданных сценариев, что расширяет возможности обучения и коммуникации.

Таким образом, реализованное приложение объединяет функции распознавания жестов, текстового и голосового взаимодействия, а также интеллектуального обучения в рамках единой программной системы.

Тестирование приложения проводилось с использованием различных устройств и сред выполнения, включая эмулятор Android, а также реальные мобильные устройства Samsung Galaxy S9 и Google Pixel 7.

В ходе тестирования оценивалась корректность распознавания жестов, стабильность работы приложения, а также производительность системы при обработке видеопотока.

Результаты тестирования показали высокую точность распознавания жестов на различных устройствах и представлены в таблице 1.

Таблица 1 – Результаты тестирования распознавания жестов

Среда тестирования (Устройство)	Корректно распознано
Samsung Galaxy S9	31 из 33 (93,94%)
Android-эмулятор	27 из 33 (81,82%)
Google Pixel 7	32 из 33 (96,97%)

Полученные результаты свидетельствуют о корректной работе алгоритмов распознавания и устойчивости системы к различным условиям выполнения.

В ходе тестирования была проведена оценка временных характеристик работы системы, включая время построения скелета руки и время классификации жестов.

Анализ показал, что на современных мобильных устройствах задержка обработки видеопотока составила около 0,7 секунды, а задержка распознавания жестов составила около 0,2 секунды, что обеспечивает комфортное взаимодействие пользователя с приложением.

Сравнение результатов на различных устройствах показало, что производительность системы зависит от вычислительных возможностей устройства, однако в целом остаётся на приемлемом уровне для практического использования.

Проведённое тестирование подтвердило работоспособность разработанного приложения, корректность распознавания жестов и соответствие системы сформированным требованиям. Полученные результаты свидетельствуют о возможности практического применения разработанного программного продукта для поддержки коммуникации и обучения пользователей жестовому взаимодействию.

ЗАКЛЮЧЕНИЕ

В ходе выполнения выпускной квалификационной работы была достигнута поставленная цель — разработано мобильное Android-приложение, предназначенное для распознавания жестов, поддержки коммуникации и обучения пользователей с нарушениями слуха и речи.

Для достижения цели были последовательно решены все поставленные задачи:

1. Исследована предметная область компьютерного зрения, алгоритмов распознавания жестов и технологий мультимодального человеко-машинного взаимодействия.
2. Выполнен анализ существующих программных решений для помощи глухим и слабослышащим людям, выявлены их функциональные ограничения и определена ниша разрабатываемого приложения.
3. Обоснован и выбран оптимальный технологический стек (Kotlin, CameraX, MediaPipe, Room, TextToSpeech, OpenRouter API), а также сформированы подробные функциональные и нефункциональные требования к системе.
4. Спроектирована модульная гибридная архитектура и программно реализовано приложения, успешно объединившее функции визуального распознавания жестов, локального чата, системы готовых фраз и интеллектуального сценарного обучения.
5. Проведено комплексное тестирование продукта на различных устройствах, подтвердившее высокую точность классификации жестов и комфортное время отклика интерфейса.

Практическая значимость исследования заключается в создании готового к использованию программного инструмента, который решает проблему коммуникативного барьера путем объединения жестового, текстового и голосового каналов связи в единой среде. В отличие от большинства аналогов, приложение не требует постоянного доступа к сети для выполнения своих базовых функций и предоставляет уникальные интерактивные режимы обучения с использованием языковых моделей.

Перспективы дальнейшего развития проекта включают:

- расширение встроенной базы жестов и поддержку русского жестового языка на уровне распознавания целых слов;

- использование продвинутых нейросетевых архитектур для классификации сложных динамических жестов;
- реализацию защищенной облачной синхронизации пользовательских данных, словарей и истории диалогов между различными устройствами;
- расширение библиотеки обучающих сценариев и глубокую интеграцию контекстно-зависимых ИИ-ассистентов.

Бакалаврская выполнена в полном объеме, все поставленные задачи успешно решены. Разработанное мобильное приложение полностью работоспособно и может применяться в повседневной практике для обеспечения безбарьерной коммуникации и изучения основ жестового языка.

Основные источники информации:

- 1 Szeliski R. Computer Vision: Algorithms and Applications. – 2nd ed. – Cham: Springer, 2022. – 947 p.
- 2 Методы компьютерной обработки изображений / под ред. В. А. Сойфера. – 2-е изд., испр. – М.: Физматлит, 2003. – 784 с.
- 3 Хайкин С. Нейронные сети: полный курс. – 2-е изд. – М.: Вильямс, 2006. – 1104 с.
- 4 Рассел С., Норвиг П. Искусственный интеллект: современный подход. – 2-е изд. – М.: Вильямс, 2006. – 1408 с.
- 5 Lugaresi C., Tang J., Nash H., McClanahan C., Uboweja E., Hays M., Zhang F., Chang C.-L., Yong M. G., Lee J., Chang W.-T., Hua W., Georg M., Grundmann M. MediaPipe: A Framework for Building Perception Pipelines // arXiv preprint arXiv:1906.08172. – 2019.
- 6 Zhang F., Bazarevsky V., Vakunov A., Tkachenka A., Sung G., Chang C.-L., Grundmann M. MediaPipe Hands: On-device Real-time Hand Tracking // arXiv preprint arXiv:2006.10214. – 2020.
- 7 OpenRouter. OpenRouter API Documentation. – [Электронный ресурс]. – URL: <https://openrouter.ai/docs> (дата обращения: 2.04.2026).
- 8 Android Developers. CameraX overview. – [Электронный ресурс]. – URL: <https://developer.android.com/media/camera/camerax> (дата обращения: 3.04.2026).
- 9 Android Developers. Save data in a local database using Room. – [Электронный ресурс]. – URL: <https://developer.android.com/training/data-storage/room> (дата обращения: 21.03.2026).