

УДК:004.934.2

ОБЗОР, АНАЛИЗ И СРАВНЕНИЕ МОБИЛЬНЫХ ОПЕРАЦИОННЫХ СИСТЕМ И КРОССПЛАТФОРМЕННЫХ ПРИЛОЖЕНИЙ ДЛЯ РАСПОЗНАВАНИЯ РЕЧИ

¹Б.Х. Ашурзода, ²А.С. Бобоев, ³А.А. Анварзода

¹Таджикский технический университет имени академика М.С. Осими,

²Университет инновации и цифровых технологий Таджикистана,

³Таджикский государственный университет права, бизнеса и политики

В статье рассматриваются современные кроссплатформенные технологии, применяемые для разработки мобильных приложений распознавания речи на таджикском языке. Особое внимание уделено особенностям работы с нейросетевыми моделями на мобильных устройствах, методам оптимизации и интеграции распознавания речи в многоплатформенные приложения. Приведены перспективы развития и основные направления исследований в данной области.

Ключевые слова: распознавание речи, таджикский язык, кроссплатформенная разработка, мобильные приложения, нейросетевые модели.

БАРРАСИ, ТАҲЛИЛ ВА МУҚОИСАИ СИСТЕМАҲОИ ОПЕРАТСИОНИИ МОБИЛӢ ВА ЗАМИМАҲОИ КРОССПЛАТФОРМА БАРОИ ШИНОХТИ НУТҚ

Б.Х. Ашурзода, А.С. Бобоев, А.А. Анварзода

Дар мақола технологияҳои муосири кросс-платформавӣ, ки барои таҳияи замимаҳои мобилии шинохти нутқ ба забони тоҷикӣ истифода мешаванд, баррасӣ мешавад. Диққати махсус ба хусусиятҳои кор бо моделҳои шабакаи нейронӣ дар дастгоҳҳои мобилӣ, усулҳои оптимизатсия ва ҳамгироии шинохти нутқ ба барномаҳои бисёрплатформавӣ дода мешавад. Дурнамои рушд ва самтҳои асосии тадқиқот дар ин самт оварда шудаанд.

Калидвожаҳо: шинохти нутқ, забони тоҷикӣ, таҳияи кросс-платформа, замимаҳои мобилӣ, моделҳои шабакаҳои нейрон.

REVIEW, ANALYSIS, AND COMPARISON OF MOBILE OPERATING SYSTEMS AND CROSS-PLATFORM SPEECH RECOGNITION APPLICATIONS

B.Kh. Ashurzoda, A.S. Boboev, A.A. Anvarzoda

The article discusses modern cross-platform technologies used to develop mobile applications for speech recognition in the Tajik language. Particular attention is paid to the features of working with neural network models on mobile devices, methods of optimization and integration of speech recognition into multi-platform applications. The development prospects and main areas of research in this area are given.

Keywords: speech recognition, Tajik language, cross-platform development, mobile applications, neural network models.

Введение

Современный мир стремительно движется в сторону повсеместного использования мобильных устройств, которые становятся основным инструментом взаимодействия человека с цифровыми технологиями. Распознавание речи — одна из ключевых технологий, обеспечивающих естественный и удобный интерфейс пользователя. Особенно актуально развитие систем распознавания речи для малоизученных и ресурсно ограниченных языков, таких как таджикский, для расширения доступа к цифровым сервисам.

Мобильные операционные системы, такие как Android и iOS, доминируют на рынке, однако обладают ограниченными вычислительными ресурсами, что ставит задачу оптимизации моделей распознавания речи и их адаптации под кроссплатформенные приложения. Технологии кроссплатформенной разработки (Flutter, React Native, Xamarin и другие) позволяют создавать универсальные приложения, однако интеграция сложных нейросетевых моделей распознавания речи в такие приложения сопряжена с рядом технических вызовов.

Цель данной статьи — проанализировать особенности мобильных операционных систем и технологий кроссплатформенной разработки в контексте интеграции и оптимизации моделей распознавания речи, а также рассмотреть основные вызовы и перспективы их решения для таджикского языка.

Задачи статьи включают обзор теоретических основ распознавания речи, анализ архитектур мобильных ОС и ограничений, а также рассмотрение возможностей и проблем при использовании кроссплатформенных технологий и оптимизации нейросетевых моделей для мобильных платформ.

Методология исследования

Исследование основано на сравнительно-аналитическом, функционально-системном и структурно-технологическом подходах. Сравнительный анализ использован для оценки архитектуры мобильных ОС и кроссплатформенных фреймворков по критериям производительности, гибкости и энергоэффективности. Функционально-системный подход позволил рассмотреть распознавание речи как элемент комплексной системы взаимодействия «пользователь — устройство — программная модель». Технологический анализ применён при изучении инструментов интеграции распознавания речи (API, SDK, ML-библиотек) на мобильных устройствах[5,7].

Анализ мобильных операционных систем

Мобильные ОС Android и iOS различаются по архитектуре и уровню открытости. Android — открытая ОС, основанная на ядре Linux, с поддержкой множества библиотек и API. iOS — закрытая система с высокой производительностью и оптимизацией под ограниченный набор устройств. Обе платформы предоставляют средства интеграции распознавания речи, включая TensorFlow Lite и Core ML[8].

Анализ кроссплатформенных технологий разработки

Кроссплатформенные технологии (Flutter, React Native, Xamarin) позволяют разрабатывать приложения из единой кодовой базы. Flutter обеспечивает высокую производительность и гибкость; React Native удобен для интеграции API; Xamarin тесно связан с .NET. Каждая технология имеет преимущества и ограничения при внедрении нейросетевых моделей распознавания речи. Ниже приведены основные показатели мобильных ОС и фреймворков (таблица 1).

Таблица 1 – Сравнительная таблица мобильных ОС и фреймворков

Платформа /Характеристика	Тип платформы	Язык разработки	Поддержка ML/AI	Офлайн-распознавание	Производительность	Гибкость интеграции
Android	Нативная ОС	Java, Kotlin	TensorFlow Lite, Vosk	Да	Высокая	Очень высокая
iOS	Нативная ОС	Swift, Objective-C	Core ML, Apple Speech	Частично	Очень высокая	Средняя
Flutter	Кроссплатформа	Dart	TensorFlow Lite, Vosk	Да	Высокая	Высокая
React Native	Кроссплатформа	JavaScript	Внешние модули	Ограничено	Средняя	Высокая
Xamarin	Кроссплатформа	C#	ML.NET, ONNX	Да	Средняя	Средняя

Результаты и обсуждение

Согласно показателям ОС, Android обеспечивает наибольшую гибкость, а iOS — энергоэффективность. Flutter и React Native демонстрируют оптимальное соотношение универсальности и производительности. Использование квантованных моделей и аппаратных ускорителей позволяет достичь высокой скорости распознавания речи даже на устройствах среднего уровня[3].

Теоретические основы автоматического распознавания речи

Автоматическое распознавание речи (APP) — это процесс преобразования речевого сигнала в текстовую информацию с помощью компьютерных алгоритмов и моделей. Основная задача APP — выделить и интерпретировать звуковые единицы речи, учитывая вариативность произношения, шумы и особенности языка.

Ключевые этапы распознавания речи включают:

Предобработку сигнала: фильтрация, нормализация, выделение характерных признаков (например, мел-частотных кепстральных коэффициентов — MFCC)[1,2].

Акустическое моделирование: связывает акустические признаки с фонемами или другими единицами языка.

Языковое моделирование: учитывает вероятностные связи между словами и фразами для повышения точности.

Декодирование: поиск наиболее вероятной последовательности слов на основе акустической и языковой моделей.

В историческом развитии APP выделяют несколько основных типов моделей:

Скрытые Марковские модели (Hidden Markov Models, HMM) — классический подход, основанный на статистическом моделировании последовательностей, активно применялся до середины 2010-х годов.

Нейросетевые модели (Deep Neural Networks, DNN) — современный подход, использующий глубокие нейронные сети для улучшения распознавания. В частности, архитектуры рекуррентных нейросетей (RNN), таких как LSTM (Long Short-Term Memory), успешно справляются с последовательной природой речи[4].

Трансформеры — более новая архитектура, основанная на механизме внимания (attention), которая показывает высокую эффективность в задачах обработки естественного языка и распознавания речи (например, модели Whisper от OpenAI).

Эти модели требуют больших вычислительных ресурсов, что становится проблемой при внедрении их на мобильные устройства с ограниченной производительностью и энергопотреблением. Поэтому для мобильных приложений применяются методы оптимизации, такие как квантование, прунинг (сокращение параметров), дистилляция знаний и использование специализированных аппаратных ускорителей.

Мобильные операционные системы: Android, iOS и др.

Мобильные операционные системы (ОС) являются программной средой, обеспечивающей управление аппаратными ресурсами мобильных устройств и выполнение прикладных программ. Наиболее распространёнными ОС на сегодняшний день являются Android и iOS, которые в совокупности занимают более 95% мирового рынка смартфонов.

Особенности Android

Android — открытая операционная система, разработанная компанией Google на базе ядра Linux.

Ключевые характеристики:

Гибкость: поддержка большого количества аппаратных конфигураций.

Модульность: возможность использования различных библиотек и фреймворков для интеграции распознавания речи.

API для работы с речью: Google Speech-to-Text API, Android SpeechRecognizer.

Особенности интеграции: доступ к низкоуровневой обработке аудио через AudioRecord и OpenSL ES.

Ограничения: высокая фрагментация устройств, различная производительность процессоров и объём оперативной памяти, что требует оптимизации моделей под конкретные классы устройств.

Особенности iOS

iOS — проприетарная операционная система компании Apple.

Ключевые характеристики:

Оптимизация под ограниченный набор устройств: высокая стабильность и предсказуемая производительность.

Средства разработки: Xcode, Swift, Objective-C.

API для распознавания речи: Apple Speech Framework.

Аппаратные ускорители: поддержка Core ML и Neural Engine для ускорения работы нейросетевых моделей.

Ограничения: закрытость платформы, жёсткие требования к приложениям, меньшая гибкость при использовании сторонних библиотек для обработки речи.

Другие мобильные ОС:

Хотя их доля на рынке мала, существуют и другие операционные системы:

HarmonyOS (Huawei) — ориентирована на экосистему устройств IoT и смартфонов Huawei.

KaiOS — облегчённая ОС для кнопочных телефонов с поддержкой некоторых веб-приложений и базовых функций распознавания речи.

Ограничения мобильных ОС для распознавания речи

Общие ограничения, характерные для всех мобильных платформ:

Вычислительные ресурсы: меньшая производительность по сравнению с настольными ПК и серверами.

Энергопотребление: работа нейросетевых моделей требует значительных энергозатрат, что может быстро разряжать аккумулятор.

Ограниченный объём памяти: необходимость сокращать размер моделей и использовать методы сжатия (quantization, pruning).

Реализация в офлайн-режиме: не всегда возможна работа без доступа к облачным сервисам.

Кроссплатформенные технологии позволяют создавать приложения, которые могут выполняться на различных мобильных операционных системах, таких как Android и iOS, из единой кодовой базы. Это существенно снижает трудозатраты на разработку и сопровождение программных продуктов[6].

Популярные кроссплатформенные фреймворки

Flutter (Google) — использует язык Dart, обеспечивает высокую производительность и нативный интерфейс[7,10].

React Native (Meta) — основан на JavaScript и React, позволяет использовать нативные модули.

Xamarin (Microsoft) — язык C#, интеграция с .NET, доступ к нативным API.

Qt — C++/QML, подходит для высокопроизводительных приложений.

Пример кода: Flutter + Vosk для распознавания речи (Android и iOS)

Ниже приведён упрощённый пример кроссплатформенного приложения на Flutter, использующего библиотеку vosk_flutter для офлайн-распознавания речи на таджикском языке[9].

Преимущества кроссплатформенного подхода для распознавания речи

1. Единая кодовая база — экономия времени и ресурсов.
2. Одинаковый пользовательский интерфейс на Android и iOS.
3. Возможность интеграции офлайн-моделей для работы без интернета.
4. Поддержка дополнительных языков через загрузку соответствующих моделей.

Особенности работы с нейросетевыми моделями на мобильных устройствах

Применение нейросетевых моделей распознавания речи на мобильных платформах имеет ряд технических особенностей, связанных с ограничениями вычислительных ресурсов, энергоэффективностью и необходимостью обеспечения высокой скорости обработки аудиосигнала в реальном времени.

1. Ограниченные ресурсы процессора и памяти

Мобильные устройства, в отличие от серверов или настольных ПК, обладают сравнительно меньшей производительной мощностью и объёмом оперативной памяти. Это требует:

- применения облегчённых архитектур нейронных сетей (MobileNet, EfficientNet, TinyML-модели);
- использования квантования весов (int8 вместо float32);
- оптимизации графа вычислений с помощью инструментов TensorFlow Lite или Core ML.

2. Энергоэффективность и тепловые ограничения

Нейросетевые вычисления, особенно при работе в режиме постоянного прослушивания ("always-on"), существенно нагружают процессор и графический ускоритель. Это ведёт к:

- повышенному энергопотреблению;
- перегреву устройства при длительных сессиях;
- необходимости применять механизмы динамического управления частотой процессора (Dynamic Voltage and Frequency Scaling, DVFS).

3. Аппаратное ускорение вычислений

Современные мобильные чипсеты (Qualcomm Snapdragon, Apple A-series, MediaTek) содержат специализированные блоки для выполнения операций машинного обучения:

- **NPU (Neural Processing Unit)** — оптимизирован для матричных умножений;
- **DSP (Digital Signal Processor)** — подходит для обработки аудиопотоков;
- **GPU (Graphics Processing Unit)** — используется для параллельных вычислений.

4. Форматы и конвертация моделей

Для интеграции нейросетевой модели в мобильное приложение необходимо:

- конвертировать её в формат, поддерживаемый платформой (TFLite для Android, Core ML для iOS);
- минимизировать размер файла модели (сжатие, обрезка слоёв);
- при необходимости — разбить модель на сегменты для загрузки по частям[5].

5. Обеспечение работы в офлайн-режиме

Для распознавания таджикской речи в условиях нестабильного интернета рекомендуется локальное выполнение модели:

- это повышает приватность данных;
- исключает задержки, связанные с сетевыми запросами;
- требует дополнительной оптимизации под малый объём памяти и вычислительные мощности.

Заключение

Проведённый анализ показал, что комбинация Flutter и Vosk является оптимальной для разработки кроссплатформенных систем распознавания речи на таджикском языке. Научная новизна заключается в систематизации подходов к интеграции распознавания речи в мобильные среды. Практическая значимость работы состоит в возможности применения результатов при создании образовательных и коммуникационных приложений, функционирующих в офлайн-режиме.

Особое внимание уделено особенностям работы с нейросетевыми моделями на мобильных устройствах, включая оптимизацию, квантование и аппаратное ускорение, что критически важно для обеспечения высокой скорости обработки речи при минимальных затратах ресурсов устройства.

1. **Кроссплатформенные технологии** позволяют разрабатывать приложения для распознавания речи одновременно под несколько мобильных ОС, снижая затраты времени и ресурсов.
2. **Оптимизация нейросетевых моделей** (квантование, уменьшение размера весов, использование GPU/NNAPI) обеспечивает работу систем распознавания в реальном времени даже на маломощных устройствах.
3. Для **таджикского языка** особую сложность представляет недостаток открытых и качественных речевых корпусов, что требует разработки собственных наборов данных и моделей.

4. Перспективным направлением является **интеграция распознавания речи с мобильными и облачными сервисами**, что позволит улучшить точность и масштабируемость решений.

Перспективы развития кроссплатформенных решений для распознавания речи на таджикском языке

В ближайшие годы ожидается значительное развитие технологий распознавания речи, включая:

- Создание и расширение **речевых корпусов таджикского языка**, что позволит обучать более точные модели[10].
- Разработка **гибридных решений**, совмещающих локальную обработку (для быстродействия) и облачные сервисы (для повышения точности).
- Повышение эффективности работы за счёт **новых архитектур нейросетей** (например, трансформеров и моделей семейства Whisper).
- Расширение возможностей кроссплатформенной разработки благодаря улучшенным фреймворкам и API, поддерживающим работу с нейросетевыми моделями на мобильных устройствах.

Рецензент: Мирзоев С.Х. – д.т.н., профессор кафедры информатики Таджикского национального университета.

Литература

1. Ашурзода Б.Х., Худойбердиев Х.А. Моделирование процесса распознавания речи в контексте таджикской язычной речи / Б.Х. Ашурзода, Х.А. Худойбердиев // Политехнический вестник. Серия Интеллект. Инновации. Инвестиции. - Душанбе. - 2022. № 2 (58) - С. 39-42. (на таджикском языке)
2. Ашурзода Б.Х. Применение алгоритма динамической трансформации временной шкалы для распознавания ключевых слов в звуковом потоке на таджикском языке / Б.Х. Ашурзода // Вестник технологического университета Таджикистана. - Душанбе. - 2022. № 3 (50). - С. 132-136.
3. Graves A., Mohamed A., Hinton G. Speech Recognition with Deep Recurrent Neural Networks // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – 2013. – pp. 6645–6649.
4. Jurafsky D., Martin J.H. Speech and Language Processing. – 3rd ed. – Pearson, 2023.
5. Povey D. et al. The Kaldi Speech Recognition Toolkit // IEEE 2011 Workshop on Automatic Speech Recognition and Understanding. – 2011.
6. Vosk API Documentation. – <https://alphacephei.com/vosk>
7. Google Cloud Speech-to-Text API Documentation. – <https://cloud.google.com/speech-to-text>
8. TensorFlow Lite Model Optimization Toolkit. – https://www.tensorflow.org/lite/performance/model_optimization
9. Flutter Documentation. – <https://flutter.dev>
10. Қосимов А. А. Муқоисаи асарҳои назмиву насрӣ дар асоси симои рақамии униграммаи рамзӣ ва ҳиҷо / А. А. Қосимов, Ҳ. А. Астанақулов, Ш. А. Бозоров // Паёми политехникӣ. Бахши: Интеллект, Инноватсия, Инвеститсия. – 2024. – №. 4(68). – Р. 45-49. – EDN KYNDQR.

МАЪЛУМОТ ОИДИ МУАЛЛИФОН - СВЕДЕНИЯ ОБ АВТОРАХ - INFORMATION ABOUT THE AUTHORS

TJ	RU	EN
Ашурзода Баҳром Хайриддин н.и.т., муаллими калон	Ашурзода Баҳром Хайриддин к.т.н., старший преподаватель	Ashurzoda Bahrom Khairiddin candidate of technical sciences, Senior Lecturer
ДТТ ба номи академик М.С. Осимӣ.	ТТУ имени академика М.С. Осими	TTU named after academician M.S.Osimi
E-mail: bahrom.91@mail.ru		
TJ	RU	EN
Бобоев Аҳмадҷон Сафаралиевич муаллими калон	Бобоев Ахмадджон Сафаралиевич старший преподаватель	Boboev Ahmadjon Safaralievich Senior Lecturer
Донишгоҳи инноватсия ва технологияҳои рақамии Тоҷикистон	Университет инновации и цифровых технологий Таджикистана	University of Innovation and Digital Technologies of Tajikistan
E-mail: ahmadjon94-26@mail.ru		
TJ	RU	EN
Анварзода Акмал Анвар докторанти (PhD) кафедраи «Технологияҳои иттилоотӣ коммуникатсионӣ ва барномарезӣ»	Анварзода Акмал Анвар докторант (PhD) кафедры «Информационнокоммуникационных технологий и программирования»	Anvarzoda Akmal Anvar doctoral Student (PhD) of the Department of Information and Communication Technologies and Programming,
Донишгоҳи давлатии ҳуқуқ, бизнес ва сиёсати Тоҷикистон	Таджикский государственный университет права, бизнеса и политики	Tajik State University of Law, Business and Politics
E-mail: akmal_dadaboev@mail.ru		