

РАЗРАБОТКА ПРОТОТИПА МОБИЛЬНОГО ИНТЕЛЛЕКТУАЛЬНОГО ПРИЛОЖЕНИЯ ДЛЯ ГЕНЕРАЦИИ ДЕТСКИХ АУДИОСКАЗОК НА ОСНОВЕ ТЕХНОЛОГИЙ СИНТЕЗА РЕЧИ

Агеев Д. С.

Научный руководитель – канд. техн. наук, доцент Добренко Н. В.

Университет ИТМО

Lokki0900@mail.ru

Введение

В последнее время наблюдается активный рост рынка детского цифрового контента, однако существующие решения для озвучивания сказок либо используют заранее записанные голоса дикторов, что ограничивает вариативность, либо применяют синтезированную речь низкого качества, не учитывающую особенности детского восприятия [1]. Актуальность работы обусловлена потребностью в создании инструмента, позволяющего генерировать персонализированные аудиосказки с разнообразными голосами персонажей без привлечения профессиональных актёров озвучивания. В рамках данной работы представлен прототип мобильного приложения, использующего современные технологии синтеза речи для решения этой задачи.

Основная часть

Проведён сравнительный анализ современных технологий синтеза речи, включая коммерческие облачные API (Yandex SpeechKit, Google Cloud Text-to-Speech, VoiceMaker) и открытые локальные модели (Silero TTS, Coqui TTS, RVC). Теоретической базой для анализа послужили работы, рассматривающие различные методы синтеза — от классического формантного и конкатенативного до современных нейросетевых подходов [2, 3]. Критериями оценки выступили: качество и естественность синтеза, вариативность доступных голосов (пол, возраст, тембр), возможность тонкого управления параметрами (темп, интонация), стоимость использования при интеграции в мобильное приложение, а также требования к вычислительным ресурсам на целевой платформе. Важным аспектом являлась возможность объективной оценки качества синтезируемой речи, что требует применения как субъективных метрик (MOS), так и автоматизированных подходов [4, 5]. По результатам анализа обоснован выбор оптимального технологического стека, обеспечивающего баланс между качеством генерации, гибкостью настройки и возможностью локального выполнения на мобильном устройстве без постоянного подключения к сети. На основе выбранного решения разработана архитектура мобильного приложения, включающая модули обработки текста, синтеза речи с управлением голосовыми параметрами и воспроизведения. Реализован прототип для платформы Android, позволяющий пользователю задавать характеристики голоса для различных персонажей сказки. Проведено тестирование прототипа, включающее оценку производительности на мобильных устройствах и субъективную оценку качества синтезированной речи.

Выводы

В ходе работы разработан и протестирован прототип мобильного приложения для генерации детских аудиосказок с использованием технологий синтеза речи. Реализованная система позволяет создавать вариативные голоса персонажей, управляя ключевыми акустическими параметрами. Результаты тестирования подтверждают применимость выбранного подхода для создания персонализированного детского контента. Дальнейшие направления работы включают расширение набора управляемых параметров, улучшение качества синтеза за счёт тонкой настройки моделей и проведение пользовательских исследований с целевой аудиторией [6].

Литература

1. Ludusan B., Wagner P. Speech, laughter and everything in between: A modulation spectrum-based analysis // Proceedings of the 10th International Conference on Speech Prosody 2020. – Tokyo: ISCA, 2020. – P. 995-999. – DOI: 10.21437/speechprosody.2020-203. – URL: <https://pub.uni-bielefeld.de/record/2943658> (дата обращения: 25.02.2026).
2. Хлопенкова А. Ю., Гришунов С. С., Белов Ю. С. Алгоритмы преобразования текста в речь на основе различных форм синтеза // Научное обозрение. Технические науки. – 2021. – № 5. – С. 20-23. – URL: <https://science-engineering.ru/ru/article/view?id=1365> (дата обращения: 25.02.2026).
3. Tan X., Qin T., Soong F., Liu T. A survey on neural speech synthesis // arXiv preprint arXiv:2106.15561. – 2021. – URL: <https://arxiv.org/abs/2106.15561> (дата обращения: 25.02.2026).
4. Белоножко П. Е., Белов Ю. С. Анализ работы вокодера на базе модели WaveRNN в системе преобразования текста в речь // Научные технологии в приборостроении и развитии инновационной деятельности в вузе: материалы Всероссийской научно-технической конференции. Т. 1. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2024. – С. 347-350.
5. Репников П. О. Модель оценки TTS метрик: Выпускная квалификационная работа (магистратура). – М.: НИУ ВШЭ, 2025. – 52 с. – URL: <https://www.hse.ru/edu/vkr/1049018247> (дата обращения: 25.02.2026).
6. Дятлов И. В. Влияние данных на качество TTS систем: Выпускная квалификационная работа (бакалавриат). – М.: НИУ ВШЭ, 2023. – 48 с. – URL: <https://www.hse.ru/edu/vkr/837215299> (дата обращения: 25.02.2026).