

Министерство образования Республики Беларусь

Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.896

Лось  
Павел Викторович

Модели и средства генерации субтитров в режиме реального времени

## **АВТОРЕФЕРАТ**

на соискание степени магистра  
по специальности 1-40 80 06 «Искусственный интеллект»

Научный руководитель  
Голенков Владимир Васильевич  
доктор технических наук, профессор

Минск 2023

## ВВЕДЕНИЕ

В настоящее время все больше людей предпочитают взаимодействовать в онлайн-среде, в том числе в процессе обучения, работы и коммуникации с близкими. Однако для людей с ограниченными возможностями восприятия аудиоинформации, таких как люди с нарушениями слуха, этот процесс может быть затруднительным. В этой связи создание систем генерации субтитров в режиме реального времени имеет большое значение, поскольку позволяет сделать аудиоинформацию доступной для людей с ограниченными возможностями слуха.

В данной диссертации рассматриваются современные технологии распознавания речи и генерации субтитров в режиме реального времени, а также предлагаются подходы для повышения качества и скорости работы систем. Будут рассмотрены различные алгоритмы и методы машинного обучения, используемые в системах распознавания речи и генерации субтитров, а также применение облачных технологий для обеспечения масштабируемости и доступности системы.

Целью данной диссертации является разработка системы, которая могла бы помочь людям с ограниченными возможностями восприятия аудиоинформации путем отображения субтитров к таковой в режиме реального времени. Разработка должна учитывать современные технологии и потребности пользователей. Результаты работы могут быть использованы в различных областях, включая технологии обучения, медиа и коммуникации.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Диссертация «Модели и средства генерации субтитров в режиме реального времени» является исследованием и разработкой системы распознавания речи и генерации субтитров в режиме реального времени. Автором была проведена обзорная аналитика существующих решений в области распознавания речи и генерации субтитров, были выявлены их преимущества и недостатки.

На основе анализа были сформулированы требования к системе и предложена архитектура, основанная на использовании WebSocket, FastAPI, React и Twilio Programmable Video. Также была использована технология Krisp для удаления шума из аудио потока и Google Cloud Services для обработки аудио и генерации перевода.

Основными задачами являлись:

- анализ существующих подходов и технологий распознавания речи;
- анализ существующих решений по распознаванию речи;
- проектирование системы распознавания речи для отображения субтитров в режиме реального времени;
- разработка и тестирование системы.

Разработанная система была успешно протестирована и показала высокую точность распознавания речи и генерации субтитров в режиме реального времени. Результаты исследования и разработки могут быть полезны для людей с нарушением слуха, а также для улучшения опыта использования видеозвонков и видеоконференций.

Научная новизна данной работы заключается в предложенной архитектуре и принятых на этапе разработки решениях.

Общий объем магистерской диссертации составляет 42 страницы, включая 16 иллюстраций, 3 таблицы, библиографический список из 19 наименований, 1 приложение.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы. В работе рассматриваются подходы к построению системы распознавания речи для отображения субтитров в режиме реального времени на основе существующих решений. Спроектированы сама система и намечены возможные пути ее развития.

В первой главе произведен анализ существующих подходов и технологий распознавания речи, исследовано историческое развитие технологии. Кратко описаны принципы работы систем, основанных на простых алгоритмах, методах динамического программирования, описано использование скрытых марковских моделей. Проанализировано использование искусственного интеллекта в современных системах автоматического распознавания речи. Описаны способы использования ИИ в процессе распознавания, а также этапы обучения нейронных сетей, которые используются в этой технологии. Также произведен анализ существующих решений по распознаванию речи, а именно: Amazon Transcribe, Google Speech-to-Text, Microsoft Azure Speech Services и Symbal.ai. Сделан вывод о том, что технологии распознавания речи прошли долгий путь развития и, благодаря возможности использования искусственного интеллекта, стали доступны широкой публике. Также сделан вывод, что хотя современные системы распознавания речи все более совершенны, они не лишены недостатков. Например, все еще затруднено распознавание речи в условиях посторонних шумов.

Во второй главе спроектирована архитектуры системы распознавания речи для отображения субтитров в режиме реального времени, описаны ее модули и их взаимодействие друг с другом. Описаны ключевые отличия данного решения от уже существующих. Определены требования, которые должна соблюдать система, а именно: высокая точность распознавания, быстрое действие, мультиязычность, отказоустойчивость, совместимость, надежность, доступность, простота использования, расширяемость. Предложено базовое решение, которое должно покрыть вышеупомянутые требования, а также возможные пути его развития. Например, намечено расширение системы в виде перевода распознанной речи на язык получателя в режиме реального времени. Также продемонстрирована диаграмма деятельности, которая позволяет подробно изучить, как происходит процесс распознавания речи в режиме реального времени и последующего отображения субтитров.

В третьей главе приведены результаты разработки и тестирования системы, сделаны выводы об эффективности примененных методов. Описаны используемые технологии. Продемонстрирован используемый интерфейс, его функционал и возможности. Также более детально описано как происхо-

дит взаимодействие между клиентом и сервером. Проведено тестирование разработанной системы, проверено соответствие сформированным на этапе проектирования требованиям. В ходе тестирования системы были произведены подсчеты WER метрики для определения эффективности системы. Результаты тестирования показали, что разработанная система работает стабильно и эффективно, соответствует требованиям и может быть использована для генерации субтитров в режиме реального времени.

## ЗАКЛЮЧЕНИЕ

В ходе работы были изучены существующие подходы и технологии распознавания речи, исследовано историческое развитие технологии. Были проанализированы существующие на рынке готовые решения по распознаванию речи, проведен сравнительный анализ таковых.

Сформулированы требования к системе, а именно: высокая точность распознавания, быстродействие, мультиязычность, отказоустойчивость, совместимость, надежность, доступность, простота использования, расширяемость. На основе сформулированных требований спроектирована архитектура системы, предложены возможные варианты развития.

Используя результаты проектирования была разработана система системы распознавания речи для отображения субтитров в режиме реального времени. В ходе испытания системы не было выявлено критических недостатков или проблем системы, которые могли бы негативно повлиять на опыт использования системы конечным пользователем.

## СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

[1] Лось, П. В. Маршрутизация в интеллектуальных системах на базе технологии OSTIS / П. В. Лось // Информационные технологии и управление : материалы 57-ой научной конференции аспирантов, магистрантов и студентов по направлению 2, Минск, 19-23 апреля 2021 года / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2021. – С. 13.

[2] Лось, П. В. Принципы представления семантических моделей в графовых базах данных / Лось П. В. // Информационные технологии и системы 2021 (ИТС 2021) = Information Technologies and Systems 2021 (ITS 2021) : материалы международной научной конференции, Минск, 24 ноября 2021 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2021. – С. 90–91.