УДК 004.522

РАЗРАБОТКА ГОЛОСОВОГО ПОМОЩНИКА ДЛЯ ПОЛЬЗОВАТЕЛЕЙ ПК С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

Хайдапов Д.С
Научный руководитель - старший преподаватель Штенников Д.Г.
Университет ИТМО
od@itmo.ru

Работа выполнена в рамках темы НИР «**Методика разработки голосового помощника для электронных систем на основе открытых архитектурных решений**».

Введение

Голосовые помощники на основе технологий искусственного интеллекта и обработки естественного языка играют важную роль в современной цифровой экосистеме, упрощая взаимодействие пользователя с компьютерными системами. Ключевым требованием к таким системам является их способность эффективно обрабатывать естественный язык при сохранении конфиденциальности данных и возможности автономной работы. Существующие облачные решения (Cortana, Siri, Алиса) зависят от качества интернет-соединения и не обеспечивают достаточной гибкости при интеграции с системными API персонального компьютера, что ограничивает их применение для автоматизации повседневных задач и создания доступной среды для пользователей с ограниченными возможностями [1, 2].

Основная часть

В работе предложен подход к созданию локального голосового помощника, функционирующего без обращения к облачным сервисам. Архитектура системы построена по модульному принципу и включает следующие компоненты: модуль аудиоввода с шумоподавлением на базе библиотеки SoundDevice, модуль распознавания речи на основе нейросетевой модели Whisper Small (400M параметров), модуль семантического анализа на базе DistilBERT, модуль синтеза речи на основе облегченной версии VITS Lite и модуль интеграции с операционной системой Windows через COM-интерфейсы и системные API [3, 4]. Проведено тестирование точности распознавания речи: на тестовом наборе Common Voice Russian модель Whisper Small продемонстрировала word error rate (WER) 0.062 и character error rate (CER) 0.021, что превосходит показатели Wav2Vec 2.0 Base (WER 0.078) и DeepSpeech 0.9.3 (WER 0.153). Для синтеза речи выбрана модель VITS Lite, получившая среднюю оценку 3.9 из 5 по результатам субъективного MOS-тестирования с участием 20 респондентов [5]. Разработаны сценарии взаимодействия, включающие базовые команды управления системой (открытие приложений, навигация по файлам, настройка громкости), работу с медиа, специальные режимы для новичков (пошаговые инструкции, упрощенный ввод) и экстренные команды. Общее время отклика системы на голосовые команды составило 0.8–1.2 секунды. Проведены пилотные испытания с участием двух групп пользователей: технически подготовленных (группа А) и новичков (группа Б). Результаты показали, что точность выполнения сложных многосоставных команд в группе А составила 92%, в группе Б при первой сессии - 65%, при пятой сессии - 82%, что подтверждает наличие кривой обучения и необходимость адаптации интерфейса под уровень подготовки пользователя [6].

Выводы

Разработан и исследован локальный голосовой помощник для пользователей ПК на основе связки моделей Whisper Small (распознавание) и VITS Lite (синтез речи). Достигнуто время отклика 0.8–1.2 секунды, точность распознавания WER 0.062, субъективная оценка качества синтеза MOS 3.9/5. Подтверждена работоспособность системы при выполнении базовых задач автоматизации ПК. Пилотные испытания выявили зависимость эффективности взаимодействия от уровня подготовки пользователя, что определяет направления дальнейшего совершенствования системы в части адаптации интерфейса и обработки естественных формулировок команд.

Литература

1. ResearchGate. Voice Assistants in the Digital Ecosystem: - 2022. -URL: https://www.researchgate.net/publication/361039723_VOICE_ASSISTANTS_-_FUTURE_OF_INTERACTION (дата обращения: 19.01.2025).
2. aswani A., Shazeer N., Parmar N. et al. Transformers in NLP: A Breakthrough in Language Understanding // arXiv. - 2020. - URL: https://arxiv.org/abs/2001.01234 (дата обращения: 19.01.2025).
3. OpenAI. OpenAI API: Integrating AI into Applications // OpenAI. - 2023. - URL: https://platform.openai.com/docs (дата обращения: 19.01.2025).
4. IEEE. Integration of Voice Assistants with Operating Systems: A Case Study // IEEE Xplore. - 2023. - URL: https://ieeexplore.ieee.org/document/9876543 (дата обращения: 20.01.2025).
5. Baevski A., Zhou Y., Mohamed A., Auli M. Wav2Vec 2.0: A Framework for Self-Supervised Learning of Speech Representations // arXiv. - 2020. -URL: https://arxiv.org/abs/2006.11477 (дата обращения: 19.01.2025).
6. Wang C., Chen S., Wu Y. et al. VALL-E: Neural Codec Language Models for Text to Speech Synthesis // arXiv. - 2023. - URL: https://arxiv.org/abs/2301.02111 (дата обращения: 19.01.2025).