

## **РАЗРАБОТКА ИНСТРУМЕНТА АВТОМАТИЗАЦИИ ПРОТОКОЛИРОВАНИЯ СОВЕЩАНИЙ НА ОСНОВЕ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

**Козлов П. И.<sup>1</sup>**

**Научный руководитель – канд. техн. наук, доцент Лабковская Р. Я.<sup>1</sup>**

<sup>1</sup>Санкт-Петербургский государственный университет телекоммуникаций  
им. проф. М. А. Бонч-Бруевича  
kozlov.pi@sut.ru

### **Введение**

Цифровизация управленческих процессов и рост числа совещаний привели к существенному увеличению объема речевых данных, требующих преобразования в формализованные протоколы. На практике эта задача часто выполняется вручную: фиксируется стенограмма, выделяются решения, формируются поручения и сроки. Такой подход трудоемок, плохо масштабируется и зависит от человеческого фактора, особенно при многоспикерной речи, фоновых шумах и длительных обсуждениях. В связи с этим возникает научная задача разработки воспроизводимого инструмента, способного автоматически преобразовывать неструктурированный аудиопоток совещаний в структурированное представление, включающее участников, принятые решения и задачи.

Современные исследования показывают, что ключевые технологические компоненты данной задачи – автоматическое распознавание речи, диаризация и идентификация спикеров, а также анализ длинных диалогов – достигли высокого уровня зрелости, однако остаются разрозненными и требуют интеграции в единый конвейер. Крупные модели распознавания речи демонстрируют устойчивость к шумам и разнообразию речевых условий [3], методы построения эмбедингов спикеров обеспечивают надежную идентификацию говорящих [2], а исследовательские корпуса и задачи формирования краткого содержания совещаний подтверждают сложность обработки длительных многопользовательских диалогов и необходимость специализированных методов анализа [1], [4]. Поэтому актуальной является разработка интегрированного локального инструмента, объединяющего указанные технологии и формирующего структурированные выходные данные, пригодные для автоматизации протоколирования.

### **Основная часть**

Предлагаемое решение представляет собой локальный интеллектуальный инструмент, преобразующий аудиозапись совещания в структурированный протокол без использования внешних облачных сервисов. В основе лежит объединение в едином технологическом контуре автоматического распознавания речи на базе модели Whisper, диаризации спикеров с использованием ruannotate.audio, межсессионной идентификации участников на основе голосовых эмбедингов типа ECAPA-TDNN, а также семантического анализа текста с применением локальной языковой модели семейства Qwen. Архитектура системы реализована в виде модульного пайплайна с очередью задач и API-интерфейсом: аудиофайл проходит предварительную обработку и разбиение на фрагменты, после чего выполняется распознавание речи, формируется стенограмма с временными метками и, при необходимости, запускается диаризация в одном из режимов в зависимости от требований к точности и доступных вычислительных ресурсов.

Семантическая обработка реализована в виде гибридного подхода, сочетающего правила постобработки русского текста (нормализация дат, сроков и числовых выражений) и языковую модель для извлечения задач, обобщения содержания и ответов

на вопросы по стенограмме. Офлайн-развертывание, использование открытых моделей и адаптивный выбор вычислительно затратных этапов обеспечивают экономичность и конфиденциальность обработки, поскольку аудио- и текстовые данные не покидают локальный контур. Такая архитектура делает систему применимой как в крупных организациях, так и в командах с ограниченной ИТ-инфраструктурой и одновременно создает основу для дальнейшего развития, включая автоматический выбор режимов диаризации, накопительное обучение голосовых профилей и расширение методов извлечения управленческих решений и поручений.

### **Выводы**

Результаты исследования подтверждают практическую применимость разработанного инструмента для автоматизации протоколирования совещаний в локальном контуре организации. Реализованный пайплайн обеспечивает преобразование аудиозаписей в структурированные выходные данные, включая стенограмму, сегментацию по спикерам, краткое содержание обсуждения, перечень задач и сроки исполнения, что позволяет сократить трудозатраты на подготовку протоколов, снизить риск пропуска поручений и повысить управляемость процессов контроля исполнения. Наибольший эффект достигается при регулярной эксплуатации системы: по мере накопления локальной базы голосовых профилей повышается точность идентификации участников, а результаты обработки могут интегрироваться в существующие процессы документооборота и постановки задач в административных подразделениях, проектных офисах, учебных и исследовательских коллективах.

Для внедрения инструмента целесообразно использовать поэтапную схему испытаний: пилотное применение на ограниченном наборе совещаний с параллельным ручным протоколированием в качестве эталона, последующую оценку качества по прикладным метрикам (полнота фиксации поручений, корректность указания исполнителей и сроков, время подготовки протокола) и масштабирование после адаптации словарей и правил под предметную область организации. Практическая эксплуатация предполагает организацию локального контура хранения данных, использование типового шаблона протокола, периодический контроль качества распознавания и извлечения задач, а также сбор обратной связи от пользователей для донастройки системы, что позволяет применять инструмент как средство первичного формирования протоколов с последующей редакторской проверкой.

### **Литература**

1. Carletta J., Ashby S., Bourban S. et al. The AMI Meeting Corpus: A Pre-announcement // *Lecture Notes in Computer Science*. 2005.
2. Dawalatabad N., Ravanelli M., Cornell S. ECAPA-TDNN Embeddings for Speaker Diarization // *Proc. Interspeech*. 2021.
3. Radford A., Kim J. W., Xu T. et al. Robust Speech Recognition via Large-Scale Weak Supervision // *arXiv*. 2022. DOI: 10.48550/arXiv.2212.04356.
4. Zhong M., Yin D., Yu T. et al. QMSum: A New Benchmark for Query-based Multi-Domain Meeting Summarization // *Proceedings of NAACL*. 2021.