

**РАЗРАБОТКА СИСТЕМЫ ГОЛОСОВОГО УПРАВЛЕНИЯ ДЛЯ  
АВТОМАТИЗАЦИИ ПРОЦЕССА ВИДЕОМОНТАЖА В ADOBE PREMIERE PRO**

**Трифонов В.Н.** (Университет ИТМО), **Меженин А.В.** (ИТМО)

**Научный руководитель – к.т.н., доцент Меженин А.В.**  
(Университет ИТМО)

**Введение.** В работе рассматриваются особенности разработки и внедрения системы голосового управления процессом видеомонтажа в приложение Adobe Premier Pro. Особое внимание уделяется методам программного взаимодействия с графическим интерфейсом видеоредактора, а также требованиям к управляющим системам на основе распознавания речи в области создания творческого контента. Разработанная система голосового управления показала свою эффективность при решении задач, связанных с вызовом навигационных панелей и управлением инструментами, однако для выполнения с её помощью творческой обработки видеофрагментов необходимо существенное улучшение системы в отношении распознавания вольно-интерпретируемых команд. Согласно опросам, создатели цифрового контента рассматривают использование голосового ввода, для поиска редко используемых команд, в работе с инструментами и цветом. В качестве аналога можно рассматривать решение в области использования голосового управления для создания графического контента в Adobe Photoshop [1].

**Основная часть.** Предлагаемое решение основано на технологиях распознавания речи [2] и Python-библиотеке Pymiere [3], которая позволяет автоматизировать управление процессом создания и редактирования видеоконтента в программе Adobe Premiere Pro. Основные возможности библиотеки pymiere включают управление проектами, монтажом, автоматический запуск скриптов, взаимодействие с медиаресурсами и интеграцию с другими библиотеками. Для тестирования эффективности спроектированной системы были рассмотрены сценарии решения базовых задач видеомонтажа. Тестирование системы проводилось в жилом помещении. Тестовые сценарии представляли собой предоставление на вход системе голосового управления как отдельных команд, например, «обрезать видео с 5 по 10 минуту», так и цепочек команд, скомпонованных внутри текста одного голосового сообщения – например, «выдели первую видеодорожку, затем обрежь её до 10 минуты». Во всех рассмотренных сценариях система голосового управления смогла интерпретировать и выполнить каждую пользовательскую команду менее чем за 1.5 секунды, а минимальное качество распознавания команд составило 90%.

**Выводы.** Предлагаемое решение - система голосового управления показала свою работоспособность при поиске необходимых панелей меню и инструментов редактирования.

**Список использованных источников:**

1. Yea-Seul Kim, Mira Dontcheva, Eytan Adar, and Jessica Hullman. 2019. Vocal shortcuts for creative experts. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems(CHI '19). 1–14. <https://doi.org/10.1145/3290605.3300562>
2. Конференция HighLoad++: Распознавание речи: как сделать Speech-to-Text своими руками // <https://highload.ru/siberia/2018/abstracts/3585>
3. Python for Premiere Pro. Pypi: vosk // <https://pypi.org/project/pymiere/>