

УДК 004.934

**РАЗРАБОТКА МОДЕЛИ ДЛЯ РАЗДЕЛЕНИЯ РЕЧИ ДИКТОРОВ НА ОСНОВЕ
TS-VAD**

Капранова В.О. (Университет ИТМО), **Астапов С.С.** (Университет ИТМО),
Кабаров В.И. (Университет ИТМО)

Научный руководитель – ведущий научный сотрудник, Попов Д.В.
(ООО ЦРТ-Инновации)

В работе рассматривается модификация метода TS-VAD для улучшения качества diarизации.

В современном мире очень востребованы системы, позволяющие разделить реплики нескольких дикторов из аудиозаписей. Такая задача осложняется тем, что в разговоре нескольких людей довольно часто встречаются наложения реплик. Для решения этой задачи были разработаны методы, такие как сквозная нейронная diarизация (EEND) и методы на основе предварительно вычисленных характеристик интересующего говорящего (Target-Speaker ASR, Speaker Beam, Voice Filter, GSS и другие). При этом у всех методов есть существенный недостаток, а именно использование какой-либо заранее известной информации о говорящих (голосовые характеристики) или о самой аудиозаписи (разметка). В данном докладе будет рассмотрен метод Target-Speaker Voice Activity Detection (TS-VAD) и его модификация, позволяющий решить задачу diarизации и не использующий априорную информацию о записи и её участниках, что является преимуществом перед другими методами.

Рассматривается возможность модификации метода TS-VAD с целью улучшения качества разделения реплик в аудиозаписях. Проверка качества работы модифицированного метода производится на реальных записях разговоров нескольких человек в естественной шумовой обстановке с наложением речи и с доступной эталонной разметкой. Для сравнения используются начальная версия TS-VAD.

В рамках исследования была создана модификация метода TS-VAD и проведён сравнительный анализ с исходной версией данного метода. Модификация метода привела к улучшению качества по сравнению с начальной версией.