

УДК 004.934.2

РАЗРАБОТКА МОДЕЛИ ДЛЯ АВТОМАТИЧЕСКОЙ ОЦЕНКИ КАЧЕСТВА РЕЧИ В ЗАДАЧЕ TTS

Хо Куанг Чунг (Университет ИТМО)

Научный руководитель – к.ф.-м.н, Рыбин С. В.
(Университет ИТМО)

Введение. Оценка качества является важным шагом в современных системах синтеза речи. В этом исследовании представлен обзор последних достижений в области автоматизированной оценки качества синтезированной речи, представленных в конкурсе VoiceMos 2023 Challenge.

Цель исследования заключается в применении последних научных достижений для создания точной и надежной модели оценки качества синтезированной речи.

Основная часть. С 2022 года исследования в области автоматической оценки качества синтезированной речи проводятся преимущественно в рамках конкурса VoiceMOS Challenge. В конкурсе VoiceMOS Challenge 2023 года основное внимание уделялось прогнозированию MOS (Mean Opinion Score) оценки голоса для трех различных сценариев: обыкновенного голоса без шума, певческого голоса и голоса на фоне шума [1].

На конкурсе было представлено несколько перспективных методик: использование информации о слушателях, выбор наиболее репрезентативных признаков, повышающих прогнозирующую способность модели для множества различных голосовых сценариев [1]. Был также предложен ряд новых модельных архитектур. В этом исследовании рассматривается методика извлечения признаков модели, которая ранее эффективно применялась для решения задач распознавания речи (ASR, Automatic Speech Recognition). На основе этого подхода разработана модель с дополнительной интеграцией спектральных речевых признаков.

Выводы. Полученные экспериментальные результаты показывают, что точность разработанной модели приближается к точности лучших на данный момент моделей на основе SSL (Self-Supervised Learning) [2]. Объединение экстракторов признаков из моделей ASR и SSL станет многообещающим подходом в будущем.

Список использованных источников:

1. E. Cooper, W. C. Huang, Y. Tsao [и др.] The Voicemos Challenge 2023: Zero-Shot Subjective Speech Quality Prediction for Multiple Domains // материалы семинара IEEE по автоматическому распознаванию и пониманию речи, — Тайвань, декабрь 2023 г. URL: <https://ieeexplore.ieee.org/document/10389763> (дата обращения: 15.02.2024).
2. E. Cooper, W.-C. Huang, T. Toda, J. Yamagishi. Generalization ability of MOS prediction networks // материалы конференции IEEE 2022 по акустике, речи и обработке сигналов — Сингапур, май 2022 г. URL: <https://ieeexplore.ieee.org/document/9746395> (дата обращения: 15.02.2024).