

**СИНТЕЗ ПЕВЧЕСКОГО ГОЛОСА
В СИСТЕМАХ MUSIC INFORMATION RETRIEVAL**

Абросимов К.И. (Университет ИТМО)

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

Введение. В работе рассматриваются модели синтеза певческого голоса в MIR системах, их достоинства и недостатки, а также актуальность и значимость задачи в современном мире. Особое внимание в работе уделяется методам и инструментам аугментации данных, так как отсутствие качественных открытых данных – основная проблема в данной задаче.

Music Information Retrieval (MIR) – это область вычислительного музыковедения, которая активно развивается в современном мире. Создаются современные MIR системы, позволяющие решать такие задачи, как автоматическая музыкальная транскрипция, распознавание эмоций в музыке, алгоритмическая композиция, синтез звучания инструментов, синтез певческого голоса и многое другое [1].

Основная часть. Одной из интереснейших задач на стыке речевых и музыкальных технологий является синтез певческого голоса. Актуальность данной задачи обусловлена несколькими факторами. Одним из основных является неизбежность смерти исполнителя, однако, при наличии хороших аудиозаписей, правильной разметки данных, можно построить модель, которая позволит создавать новые музыкальные произведения для любимого исполнителя. Но самая главная сложность задачи заключается в отсутствии открытых качественных данных [2], поэтому аугментация – это ключевой инструмент в данных системах. В работе рассматриваются современные нейросетевые модели синтеза певческого голоса [3], их достоинства и недостатки, применение различных вокодеров и анализ результатов, а также методы и инструменты аугментации данных.

Выводы. В результате анализа моделей и проведения экспериментов были найдены качественные предобученные модели, на основе которых будут производиться эксперименты по улучшению качества звучания с помощью различных механизмов и инструментов, а также реализованы некоторые методы аугментации для задачи синтеза певческого голоса.

Список использованных источников:

- [1] Michael Fingerhut (2004). "Music Information Retrieval, or how to search for (and maybe find) music and do away with incipits", IAML-IASA Congress, Oslo (Norway), August 8–13, 2004
- [2] Soonbeom Choi, Juhan Nam. "A MELODY-UNSUPERVISION MODEL FOR SINGING VOICE SYNTHESIS". — 2022. — Vol. arXiv: 2110.06546. — version 2. Access mode: <https://arxiv.org/pdf/2110.06546.pdf>
- [3] Yu Gu, Xiang Yin, Yong hui Rao, Yuan Wan, Benlai Tang, Yang Zhang, Jitong Chen, Yuxuan Wang, Zejun Ma. "ByteSing: A Chinese Singing Voice Synthesis System Using Duration Allocated Encoder-Decoder Acoustic Models and WaveRNN Vocoders". — 2021. — Vol. arXiv: 2004.11012. — version 2. Access mode: <https://arxiv.org/pdf/2004.11012.pdf>