

**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук
Департамент программной инженерии

СОГЛАСОВАНО
Доцент департамента больших
данных и информационного поиска,
кандидат компьютерных наук

УТВЕРЖДАЮ
Академический руководитель
образовательной программы
«Программная инженерия»
профессор департамента
программной инженерии,
канд. техн. наук

_____ М.Л. Каледин
«__» _____ 2025 г.

_____ Н.А. Павлочев
«__» _____ 2025 г.

Подп. и дата	
Инв. № дубл.	
Взам. инв. №	
Подп. и дата	
Инв. № подл	RU.17701729.12.10-01 12 01-1-ЛУ

Приложение для разделения аудио на основе нейронных сетей

Текст программы

ЛИСТ УТВЕРЖДЕНИЯ

RU.17701729.12.10-01 12 01-1-ЛУ

Исполнитель
Студент группы БПИ 223
_____/Иванов Г.Я./
«__» _____ 2025 г.

Москва 2025

УТВЕРЖДЕН
RU.17701729.12.10-01 12 01-1-ЛУ

Приложение для разделения аудио на основе нейронных сетей
Текст программы

RU.17701729.12.10-01 12 01-1

Листов 8

Инв. № подл	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата
RU.17701729.12.10-01 12 01-1-ЛУ				

Москва 2025

3
RU.17701729.12.10-01 12 01-1-ЛУ
Содержание

ВВЕДЕНИЕ.....	4
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ.....	6
Приложение 1	7
Лист регистрации изменений.....	8

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.12.10 -01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

ВВЕДЕНИЕ

В данном документе представлено описание структуры репозитория, в котором находится исходный код программы «Приложение для разделения аудио на основе нейронных сетей».

Репозиторий программы находится по ссылке: <https://github.com/prikokes/AudSep>.
Полный текст не приводится в связи с большим количеством строк кода.

Репозиторий имеет следующую структуру:

- 1) В папке `model_loaders` лежат все загрузчики моделей для разделения аудио:
 - `bs_roformer_loader.py` - загрузчик модели BS RoFormer [13]
 - `htdemucs_loader.py` - загрузчик модели HTDemucs [12]
 - `mel_band_roformer_loader.py` - загрузчик модели MelBand RoFormer
- 2) В папке `models` лежат все модели нейронных сетей:
 - `bs_roformer.py` - модель BS RoFormer [13] для разделения аудио
 - `htdemucs.py` - модель HTDemucs [12] для разделения аудио
 - `mel_band_roformer.py` - модель MelBand RoFormer [14] для выделения вокала
 - `attend.py` - модуль внимания для трансформеров
- 3) В папке `templates` лежат компоненты графического интерфейса:
 - `audio_player.py` - компонент аудиоплеера
 - `audio_separator_app.py` - основное приложение
- 4) В папке `utils` лежат утилиты для обработки аудио:
 - `demix_track.py` - утилита для разделения треков
 - `demix_track_demucs.py` - утилита для разделения треков с использованием HTDemucs [12]
- 5) В папке `configs` лежат конфигурационные файлы для моделей:
 - `config_bs_roformer.yaml` - конфигурация BS RoFormer [13]
 - `config_htdemucs_6stems.yaml` - конфигурация HTDemucs [12]
 - `config_vocals_mel_band_roformer_kj.yaml` - конфигурация MelBand RoFormer [14]

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.12.10 -01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

6) В папке `weights` хранятся веса предобученных моделей:

- `bs_roformer.ckpt` - веса BS RoFormer [13]
- `ht_demucs_v4.th` - веса HTDemucs [12]
- `melband_roformer.ckpt` - веса MelBand RoFormer [14]

7) В корне проекта находятся основные файлы

- `main.py` - точка входа в приложение
- `requirements.txt` - зависимости проекта
- `README.md` - описание проекта

8) В папке `docs` находится документация проекта

- Техническое задание
- Руководство оператора
- Программа и методика испытаний
- Пояснительная записка
- Текст программы

9) В директории `dist` находятся рабочие сборки приложения

10) В директории `notebooks` находятся файлы Jupyter Notebook с тестами моделей и производительности

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.12.10 -01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. ГОСТ 19.101-77 Виды программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
2. ГОСТ 19.102-77 Стадии разработки. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
3. ГОСТ 19.103-77 Обозначения программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.104-78 Основные надписи. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.105-78 Общие требования к программным документам. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. ГОСТ 19.201-78 Техническое задание. Требования к содержанию и оформлению. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
8. ГОСТ 19.603-78 Общие правила внесения изменений. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
9. ГОСТ 19.604-78 Правила внесения изменений в программные документы, выполненные печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
10. ГОСТ 15150-69 Машины, приборы и другие технические изделия. Исполнения для различных климатических районов. Категории, условия эксплуатации, хранения и транспортирования в части воздействия климатических факторов внешней среды. – М.: Изд-во стандартов, 1997.
11. ГОСТ 19.301-79 Программа и методика испытаний. Требования к содержанию и оформлению. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
12. Rouard S., Massa F., Défossez A. Hybrid transformers for music source separation //ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2023. – С. 1-5.
13. Lu W. T. et al. Music source separation with band-split rope transformer //ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2024. – С. 481-485.
14. Wang J. C., Lu W. T., Won M. Mel-Band RoFormer for Music Source Separation //arXiv preprint arXiv:2310.01809. – 2023.
15. ZFTurbo. Music-Source-Separation-Training [Электронный ресурс]. – Режим доступа: <https://github.com/ZFTurbo/Music-Source-Separation-Training> (дата обращения: 01.04.2025)

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.12.10 -01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

ТЕРМИНОЛОГИЯ

1. **Разделение аудио** — разделение аудиофайла с записанной музыкой на несколько дорожек с инструментами.
2. **drag-and-drop** — это интерфейсная техника, позволяющая пользователю перемещать элементы на экране с помощью мыши или сенсорного ввода. Пользователь "захватывает" элемент, перетаскивает его в нужное место и "отпускает". В данном приложении эта техника должна поддерживаться для звуковых файлов.
3. **solo/mute** — функции, используемые в аудиообработке и звуковом монтаже для изоляции или отключения отдельных дорожек. Solo - функция, которая позволяет прослушивать только выбранные дорожки, временно отключая все невыбранные. Mute - функция, которая полностью отключает звук на выбранной дорожке или источнике, позволяя сосредоточиться на других элементах микса без удаления дорожки.
4. **waveform** — графическое представление звукового сигнала, отображающее изменение амплитуды звука во времени.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.12.10 -01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата

[illegible]

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.12.10 -01 ТЗ 01-1				
Инв. № подл.	Подп. и дата	Взам. Инв. №	Инв. № дубл.	Подп. и дата