

Программное обеспечение для анализа и обработки речевых сигналов
Руководство пользователя

Киров, 2025 г.

АННОТАЦИЯ

Настоящий документ содержит основные положения и сведения, необходимые для работы программиста с программным обеспечением для анализа и обработки речевых сигналов.

Документ разработан согласно требованиям ГОСТ Р 59795–2021 и ГОСТ Р 59853–2021.

СОДЕРЖАНИЕ

1	Введение	
	Ошибка! Закладка не определена.	
1.1	Область применения	Ошибка! Закладка не определена.
1.2	Краткое описание возможностей.....	Ошибка! Закладка не определена.
1.3	Уровень подготовки пользователя	Ошибка! Закладка не определена.
1.4	Перечень эксплуатационной документации, с которыми необходимо ознакомиться пользователю	Ошибка! Закладка не определена.
2	Назначение и условия применения	5
2.1	Виды деятельности, функции, для автоматизации которых предназначено данное средство автоматизации	Ошибка! Закладка не определена.
2.2	Условия, при соблюдении которых обеспечивается применение средства автоматизации в соответствии с назначением	5
3	Подготовка к работе	
	Ошибка! Закладка не определена.	
3.1	Состав и содержание дистрибутивного носителя данных	10
3.2	Порядок загрузки данных программы и установки программы	Ошибка! Закладка не определена.
3.3	Порядок проверки работоспособности	Ошибка! Закладка не определена.
4	Описание операций	
	Ошибка! Закладка не определена.	
4.1	Общее описание технологического процесса ..	Ошибка! Закладка не определена.
4.2	Описание экранных форм и интерфейса	Ошибка! Закладка не определена.
4.3	Описание экранных форм и интерфейса	Ошибка! Закладка не определена.
5	Аварийные ситуации	15
5.1	Действия при ошибках загрузки аудиофайла.....	15
5.2	Действия при зависании приложения	15
5.3	Действия при потере данных	15
5.4	Действия при отказе оборудования.....	16
5.5	Действия в других аварийных ситуациях.....	16
6	Рекомендации по освоению	
	Ошибка! Закладка не определена.	
7	Рекомендации по использованию	
	Ошибка! Закладка не определена.	
7.1	Оптимизация работы приложения.....	Ошибка! Закладка не определена.
7.2	Работа с результатами анализа	Ошибка! Закладка не определена.

1 Общие положения

Настоящий документ содержит сведения, необходимые для работы программиста с программным обеспечением для анализа и обработки речевых сигналов.

1.1 Наименование программы

Полное наименование системы: Программное обеспечение для анализа и обработки речевых сигналов.

2 Назначение и условия применения

2.1 Назначение программы

Основным назначением приложения является автоматизация процессов анализа и обработки речевых сигналов, включая:

- автоматическую сегментацию речи на фонетические единицы;
- идентификацию фонетических структур и элементов речи;
- проведение спектрального анализа аудиозаписей;
- формирование отчётов и результатов для дальнейшего исследования;
- сравнение и анализ различных аудиозаписей.

2.2 Функции, выполняемые программой

Перечень функций, реализуемых программой:

- Импорт и воспроизведение аудиофайлов (WAV, MP3, AAC, FLAC);
- Нормализация и фильтрация аудиосигналов;
- Автоматическая сегментация речи на фонемы;
- Классификация фонем (гласные, согласные, ударные);
- Поиск и анализ звуковых сочетаний;
- Визуализация спектрограмм и волновых форм;
- Экспорт результатов (PDF, XML, JSON).

2.3 Сведения о технических и программных средствах, обеспечивающих выполнение программы

2.3.1 Требования к составу и параметрам комплекса технических средств

Средства аппаратного обеспечения, необходимые для функционирования программы, рассчитываются индивидуально под потребность решаемых задач, минимальные требования представлены в таблице.

Таблица 1 – Минимальные системные требования

Вид ресурса	Минимальные требования
Тип ЭВМ	Персональный компьютер или ноутбук
Процессор	Тактовая частота не менее 2.0 ГГц

Оперативная память (ОЗУ)	Не менее 8 ГБ
Накопитель (HDD/SSD)	Не менее 10 ГБ свободного места
Видеокарта	Интегрированная с поддержкой OpenGL 3.0
Монитор	Разрешение не ниже 1024×768
Устройства ввода	Клавиатура и мышь
Аудиооборудование	Встроенные или внешние динамики/наушники
Операционная система	Windows 11
Сетевое соединение	Локальное подключение (localhost)
Программные компоненты	Python 3.8+, Qt 5.15+, PostgreSQL 12+
Средства резервного копирования	Носитель ёмкостью не менее 32 ГБ (по необходимости)

Приложение должно использоваться в стабильных условиях окружающей среды без воздействия экстремальных температур и влажности.

2.3.2 Требования к общесистемному программному обеспечению

Перечень и описание Системного ПО, обеспечивающего корректную работу приложения, представлены в таблице 2.

Таблица 2 – Сведения о программном обеспечении

Вид ресурса	Минимальные требования
Операционная система	Windows 11 (64-разрядная)
Интерпретатор Python	Версия не ниже 3.8
Компилятор C++	Поддержка стандарта C++17 (MSVC 2019 или MinGW 8.1 и выше)
Qt Framework	Версия 5.15 и выше
PostgreSQL	Версия 12 и выше
Библиотека SpeechBrain	Совместимая версия (через <code>pip install speechbrain</code>)
Библиотеки NumPy и SciPy	Актуальные стабильные версии
Веб-фреймворк Flask или FastAPI	Версия не ниже 2.0

OpenGL	Поддержка версии 3.0 и выше
Средства архивирования	Любые стандартные утилиты ОС (7-Zip, встроенные архиваторы)
Браузер или HTTP-клиент (по необходимости)	Любой современный браузер (Chrome, Edge, Firefox)

Средства, обеспечивающие функционирование Системы:

Сервер приложения — локальный сервис `analysis_server.exe`, реализованный на Python, обеспечивающий выполнение вычислений, обработку аудиофайлов и взаимодействие с клиентом через REST API.

Клиентское приложение — модуль `SpeechAnalyzer.exe`, реализованный на C++/Qt, обеспечивающий интерфейс пользователя, визуализацию данных и обмен с сервером по протоколу HTTP (порт 8000).

База данных — локальное хранилище PostgreSQL, используемое для сохранения метаданных, параметров анализа и ссылок на файлы.

Файловое хранилище — локальный каталог, содержащий аудиофайлы, кэш спектрограмм и результаты экспорта.

Облачный узел Hugging Face Hub — внешний источник предобученных моделей (SpeechBrain, Wav2Vec2 и др.), обеспечивающий загрузку и обновление алгоритмов обработки речи при первом запуске.

Сетевой стек HTTP/HTTPS — обеспечивает взаимодействие между компонентами приложения и загрузку моделей по защищённому протоколу.

Взаимодействие всех компонентов системы реализуется в локальной клиент–серверной архитектуре, представленной на рисунке 1.

Компоненты работают автономно и не требуют постоянного подключения к внешним сервисам, за исключением первоначальной загрузки моделей.

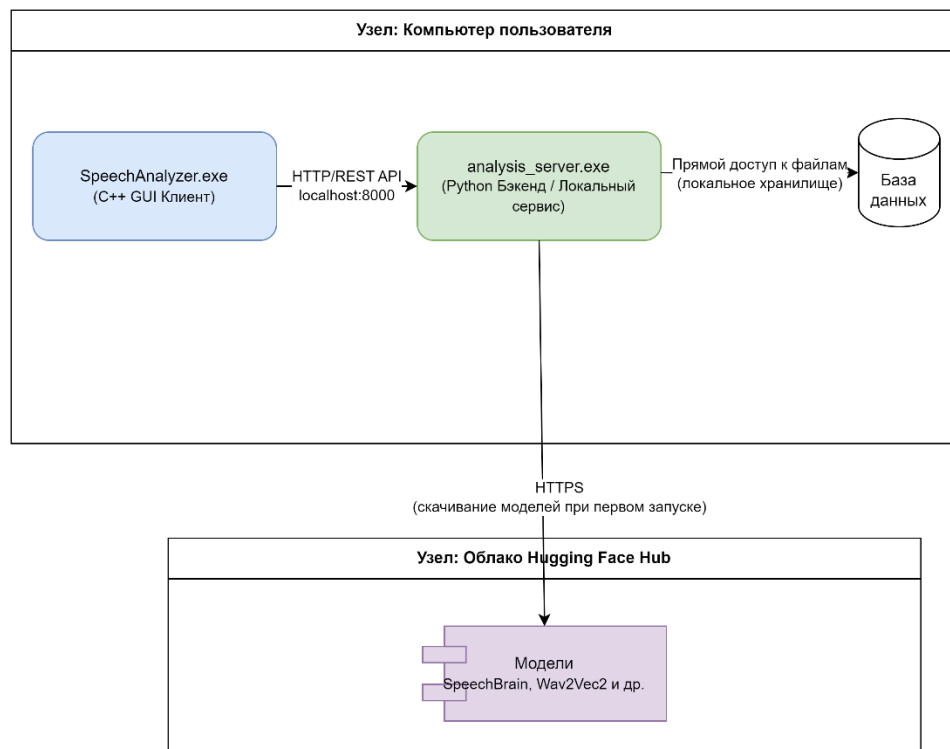


Рисунок 1 – Архитектурная схема функционирования системы

2.3.3 Требования к персоналу

Обязательные навыки программиста:

- Навыки работы с Linux/Windows;
- Знание стандартных возможностей ОС (командная строка, управление процессами);
- Основы администрирования систем.

Специальные навыки:

Python разработка:

- Стандартная библиотека Python;
- Работа с NumPy и SciPy для обработки сигналов;
- Принципы обработки аудио и цифровой обработки сигналов;
- Работа с REST API и JSON;
- Асинхронное программирование (asyncio, async/await).

C++ разработка:

- C++17 стандарт и выше;
- Qt Framework (сигналы, слоты, модель-представление);
- Многопоточное программирование;

- OpenGL базовые принципы.

Базы данных:

- SQL язык (SELECT, INSERT, UPDATE, DELETE);
- Проектирование реляционных БД;
- PostgreSQL специфика (PL/pgSQL, индексы, оптимизация);
- Работа с миграциями БД.

Смежные знания:

- Основы цифровой обработки сигналов (DSP);
- Основы фонетики и лингвистики;
- Спектральный анализ, преобразование Фурье;
- Машинное обучение (общие принципы);
- REST архитектура.

Желательные навыки

- Опыт работы с Audio API (PortAudio, ALSA);
- Знание Docker и контейнеризации;
- Опыт с CI/CD системами (Jenkins, GitLab CI);
- Опыт профилирования и оптимизации производительности;
- Знание инструментов для анализа речи (Praat, WaveForm Editor).

3 Характеристика программных средств программы

3.1 Описание основных характеристик программы

Программа SpeechAnalyzer предназначена для анализа и обработки речевых сигналов.

Она реализована как локальное клиент–серверное приложение, состоящее из:

- SpeechAnalyzer.exe — клиентское приложение (C++/Qt), обеспечивающее графический интерфейс пользователя, визуализацию данных и управление анализом;
- analysis_server.exe — серверный модуль на Python, реализующий REST API и выполняющий вычислительные операции;
- Базы данных PostgreSQL — для хранения параметров, результатов анализа и журналов;
- Файлового хранилища — для хранения аудиофайлов и экспортируемых результатов.

Компоненты взаимодействуют через протокол HTTP (порт 8000), а при первом запуске подключаются к облаку Hugging Face Hub для загрузки моделей (SpeechBrain, Wav2Vec2 и др.).

3.1.1 Временные характеристики программы

Среднее время запуска клиентского приложения — не более 5 секунд.

Анализ короткого аудиофайла (до 20 секунд) занимает 2–4 секунды.

При увеличении объёма данных время обработки растёт пропорционально длительности файла.

Передача данных между модулями по локальному REST API занимает не более 0,5 секунды на операцию.

3.1.2 Режим работы программы

Программа функционирует в интерактивном режиме: пользователь запускает анализ через GUI и получает результаты в виде графиков, таблиц и текстовых отчётов.

Также поддерживается пакетный режим — вызовы API из сторонних приложений или пользовательских скриптов (для автоматического анализа набора аудиофайлов).

3.1.3 Средства контроля правильности выполнения программы

Для обеспечения корректности функционирования реализованы следующие механизмы:

- проверка формата и целостности входных аудиофайлов;
- автоматическая валидация параметров анализа;
- логирование всех действий в файлы client.log и api.log;
- контроль целостности базы данных (проверка соединения и доступности таблиц);

- отображение системных сообщений об ошибках в интерфейсе GUI.

4 Обращение к программе

Программа запускается с помощью исполняемого файла SpeechAnalyzer.exe. Для управления программой предусмотрены следующие способы:

- интерактивный запуск пользователем через графический интерфейс;
- запуск с передачей параметров командной строки (при необходимости);
- загрузка конфигурационного файла с параметрами запуска.

Порядок вызова программы:

- Запуск файла SpeechAnalyzer.exe из меню "Пуск" или ярлыка на рабочем столе.
- Инициализация пользовательского интерфейса и загрузка стартового экрана с возможностью выбрать существующий проект или создать новый.
 - При запуске программа читает конфигурационный файл config.ini для настройки параметров работы.
 - Пользователь через GUI может загружать аудиофайлы, запускать анализ и использовать функции программы.
 - При необходимости программа выводит сообщения о статусах, ошибках и результатах операций в диалоговых окнах и областях уведомлений.

5 Входные и выходные данные

5.1 Характер и организация входных данных

Входные данные представляют собой аудиосигналы в цифровом формате, которые характеризуются следующими параметрами:

- Форматы файлов: WAV (без сжатия, PCM), MP3, AAC, FLAC и другие распространённые аудиоформаты с сжатием и без потерь.
- Частота дискретизации: от 8 кГц до 200 кГц, в программном обеспечении обычно поддерживается стандарт 44.1 кГц (CD качество) или выше.
- Разрядность (битность): 16, 24, 32 бита на сэмпл (или float 32-bit в обработке).
- Каналы: моно или стерео (программное обеспечение может объединять два моно-го трека для стерео).
- Размер данных зависит от длительности аудиозаписи и параметров кодирования.
- Структура входных данных организована в файлы с метаданными (формат, длительность, частота дискретизации), загружаемые в проект.

Данные проходят предварительную обработку: нормализацию уровня громкости, фильтрацию шумов, сегментацию на фонемы, что требует сохранения аудиоданных как в исходном, так и в промежуточных форматах.

5.2 Характер и организация выходных данных

Выходные данные формируются после обработки входящих аудиосигналов и содержат:

- Обработанные аудиофайлы с применённой нормализацией, фильтрацией шумов в формате WAV или других аудиоформатах.
- Результаты анализа речи, в том числе:
 - Список фонем с временными метками начала и конца;
 - Классификация фонем по типу (гласные, согласные, ударные фонемы);
 - Найденные звуковые сочетания и паттерны с указанием их позиции во времени.
- Визуализации спектрограмм и других акустических характеристик, экспортируемые в растровых форматах (например, PNG).
- Текстовые файлы отчётов и данных с возможностью дальнейшей обработки, обычно в форматах TXT, CSV, JSON.
- Логи работы и отчёты о ходе обработки и возможных ошибках.

6 Сообщения

В программном обеспечении для анализа и обработки речевых сигналов сообщения реализуются как элементы интерфейса, обеспечивающие информирование пользователя о ключевых состояниях и событиях системы. Сообщения подразделяются на следующие виды:

- Информационные сообщения — уведомляют пользователя об успешном выполнении операций или о текущем состоянии (например, "Проект успешно сохранён", "Аудиофайл загружен").
- Предупреждения — информируют о потенциальных проблемах, которые не блокируют работу, но требуют внимания (например, "Формат файла не совсем поддерживается, возможны ошибки").
- Ошибки — отображаются при возникновении критических сбоев или неверных действиях (например, "Ошибка загрузки файла", "Неверный формат аудиофайла").
- Аварийные сообщения — сопровождают серьезные проблемы, ведущие к остановке или невозможности продолжения работы (например, "Память заполнена, завершение операции").

Сообщения отображаются пользователю через диалоговые окна, строки статуса, всплывающие подсказки и лог-файлы. Формат сообщений стандартизирован по структуре: тип сообщения, код ошибки (если есть), текстовое описание, рекомендации по устранению.

7 Аварийные ситуации

В данном разделе настоящего руководства пользователя содержится информация о действиях пользователя в случаях возникновения аварийных ситуаций.

7.1 Действия при ошибках загрузки аудиофайла

Возможные ошибки:

- «Файл не поддерживается»;
- «Ошибка при чтении файла»;
- «Файл повреждён».

Действия:

- Убедиться, что загружаемый файл находится в одном из поддерживаемых форматов (WAV, MP3, AAC, FLAC);
- Проверить целостность аудиофайла (попробовать открыть его в другом приложении для воспроизведения);
- Если файл повреждён, использовать инструмент восстановления файлов;
- Попробовать перезагрузить приложение и повторить загрузку файла.

7.2 Действия при зависании приложения

При полном зависании приложения:

- Нажать комбинацию клавиш Ctrl+Alt+Delete для открытия диспетчера задач;
- Найти в списке «SpeechAnalyzer» или «Анализ речевых сигналов»;
- Выделить процесс и нажать кнопку «Завершить задачу»;
- Перезапустить приложение.

При медленной работе:

- Закрыть другие запущенные приложения;
- Проверить объём свободной памяти на диске (должно быть не менее 2 ГБ);
- Перезагрузить компьютер;
- Если проблема сохраняется, обратиться к системному администратору.

7.3 Действия при потере данных

При случайном удалении результатов анализа:

- Если результаты были экспортированы в отдельные файлы, восстановить их из резервной копии;
- Если проект не был сохранён, потерянные данные восстановить невозможно;
- При следующей работе регулярно сохранять проект (Ctrl+S).

7.4 Действия при отказе оборудования

При отказе дисплея:

- Перезагрузить компьютер;
- Проверить подключение монитора;
- Убедиться в исправности видеокарты.

При отказе динамиков при воспроизведении:

- Проверить громкость в системе (значок динамика на панели задач);
- Убедиться в подключении динамиков;
- Проверить значок звука в приложении;
- Обновить драйверы звуковой карты.

7.5 Действия в других аварийных ситуациях

В случае возникновения других аварийных ситуаций при работе с программой необходимо выполнить следующее:

- определить природу и причины аварийной ситуации;
- принять меры для предотвращения распространения аварийной ситуации и минимизации возможного ущерба;
- сообщить о произошедшем разработчику и заинтересованным сторонам;
- восстановить работоспособность системы и возобновить технологический процесс.