

Веб-приложение для распознавания
музыкальных инструментов и
преобразования аудиозаписей

Авторы:

Бухараев Алим

Прохоров Юрий

Савелов Михаил

Яушев Фарух

Руководитель:

Бабичев Сергей Леонидович

Содержание

1 Техническое задание

1.1 Аннотация

Данный проект — курсовая работа вышеперечисленных студентов второго курса ФУПМ (факультета управления и прикладной математики) МФТИ. Он состоит из 3 основных обособленных частей:

1. Библиотека и API на C++ для обработки аудиозаписей.
2. API в виде расширения PHP для обработки аудиозаписей.
3. Пользовательский интерфейс в виде веб-приложения.

Об установке можно прочитать ниже. Ссылка на репозиторию Github:

<https://github.com/suchusername/sounds>

Изначально, у нас была идея написать программу, способную распознавать музыкальные инструменты на аудиозаписи. Причина связана с тем, что эта задача не имеет точных результатов, и мы хотели получить свое собственное частное решение. Затем мы решили расширить функционал нашего приложения, добавив наиболее популярные запросы, и реализовать его как легко доступный пользователю веб-сервис. Таким образом, основные цели проекта:

- Повышение командных навыков.
- Реализация приложения по популярной тематике.
- Применение знаний языка C++.
- Изучение новых инструментов и технологий.

Пользователю предлагается загрузить свой файл в формате WAV на сервер. После этого он получает некоторую статистику о файле и выбирает, какое действие он желает выполнить с ним или сразу с несколькими файлами. Среди возможных функций: предсказание музыкального инструмента по аудиозаписи, слияние файлов, их обрезка, усиление громкости, ускорение и другие преобразования. Пользователь имеет возможность отправить сообщение о том, угадала ли сеть, а аудиозапись может быть использована далее для дальнейшего обучения нашей или каких-либо других моделей. Наконец, преобразованный файл пользователь может прослушать и загрузить обратно на свое устройство.

При разработке приложения использовались следующие технологии:

- Язык C++ стандарта 11 для написания серверной составляющей (ядра).
- Набор инструментов для написания расширения PHP на C++.
- Python 3 и библиотеки Keras, NumPy, SciPy для реализации нейронной сети.
- Средства PHP, HTML5, CSS и Javascript для разработки пользовательского интерфейса.

1.2 Цели

Главной мотивацией работы является его выполнение в качестве курсового проекта по информатике студентов второго курса ФУПМ МФТИ: Бухараева Алима, Прохорова Юрия, Савелова Михаила и Яушева Фаруха. Руководитель проекта — Бабичев Сергей Леонидович.

Другая важная цель — реализация классификатора мелодий по играющему музыкальному инструменту. Нами был проведен анализ уже существующих средств распознавания музыкальных инструментов, и, как оказалось, тема не является очень популярной. Мы же считаем, что данный этап предобработки аудиозаписей принесет большую пользу для дальнейшего их анализа (например, распознавания играющей композиции, удаления некоторых тональностей или инструментов), поэтому мы попробовали внести свой вклад в решение данной задачи.

Еще одна задача проекта — это разработка доступного пользовательского интерфейса для простейшей обработки аудиофайлов. Мы не нашли сайта, который представляет собой "онлайн-редактор аудиозаписей", и в связи с этим решили дополнить нашу программу этими функциями и предоставить ее как веб-сервис.

Наконец, важной для нас целью является обретение опыта работы в команде и грамотное распределение обязанностей. Кроме того, сюда же входит получение опыта приложения полученных знаний о языке C++ к реальной задаче, изучение новых полезных для практического применения технологий и инструментов.

1.3 Требования

В данном разделе перечислены все задачи и подзадачи, которые мы поставили перед собой в рамках выполнения настоящей проектной работы.

1. Библиотека и API на C++.

- Разработка объектно-ориентированной внутренней структуры библиотеки.
- Побайтовый разбор файла формата WAV.
- Реализация функционала
 - Вывод статистики о файле.
 - Изменение громкости.
 - Ускорение аудиозаписи.
 - Слияние аудиозаписей.
 - Предсказание музыкального инструмента.
 - Обрезка аудиозаписи.
- Универсальный пользовательский интерфейс с вышеперечисленными функциями.

2. Нейронная сеть.

- Подготовка и нарезка датасета для обучения модели.
- Поиск оптимального способа предобработки аудиозаписей.
- Моделирование наиболее эффективной архитектуры классификатора, различающего музыкальные инструменты.
- Обучение классификатора на подготовленных данных.
- Анализ результатов.

3. Расширение RНР.

- Настройка Apache веб-сервера и установка необходимых инструментов для реализации RНР-расширения.
- Универсальный пользовательский интерфейс с вышеперечисленными функциями.

4. Веб-приложение.

-

1.4 Распределение обязанностей

Член команды	Задача
Бухараев Алим	Изучение сверточных и рекуррентных нейронных сетей
	Поиск большого числа мелодий, пригодных для использования в качестве обучающей выборки
	Проектирование и обучение модели-классификатора
Прохоров Юрий	Реализация анализатора файла формата .WAV
	Спектральный анализ аудиозаписей с помощью дискретного преобразования Фурье
	Различные преобразования аудиозаписей
	Реализация программного интерфейса расширения PHP
Савелов Михаил	Изучение PHP, HTML5, JavaScript и CSS
	Моделирование пользовательского интерфейса
Яушев Фарух	Изучение PHP, HTML5, JavaScript и CSS
	Моделирование пользовательского интерфейса

2 Документация

2.1 Структура директорий в репозитории

В этом и следующем разделах приведено описание репозитории проекта на GitHub.

Audios

В данную директорию загружаются все пользовательские файлы, там же создаются новые промежуточные аудиофайлы.

Classes

В этой директории находятся реализации методов всех классов на C++. Каждому классу соответствует свой файл. Классы, отвечающие преобразованиям аудиозаписей вынесены в отдельную папку *Classes/Queries*.

Documentation

В этой директории находится данный документ и его .tex файл. Также в папке *Documentation/Papers* присутствуют статьи на разные тематики, которые использовались при написании этой проектной работы.

Instrument_classifier

В этой директории находятся следующие файлы:

- *launch.py*
Содержит реализацию нейронной сети с использованием библиотеки Keras. Этот же файл отвечает за получения предсказания на конкретной аудиозаписи. Нейронную сеть можно явно использовать следующим образом:

```
cd Instrument_classifier
python3 launch.py path/to/file.wav
```

- *my_model_*.h5*
Различные версии модели.

root

Данная директория является корневой директорией Apache веб-сервера. В ней содержатся следующие объекты:

- **.php*
Файлы формата .php — ответы сервера на запросы клиента. Они генерируют html код, который видит пользователь.
- *sounds_update.sh*
Shell-скрипт, который делает новую сборку расширения PHP, переустанавливает его и перезапускает Apache веб-сервер.
- *logs*
Текстовый файл с отладочной информацией.
- *root/css, root/js*
Эти папки содержат файлы соответствующего формата, необходимые для реализации дизайна и пользовательского интерфейса веб-страницы.
- *root/sounds*
Эта папка содержит все файлы, необходимые для создания расширения PHP. Среди них:
 - *config.m4*
Файл с параметрами для компиляции расширения PHP.

- *php_sounds.h*
Задание интерфейса PHP-расширения.
- *sounds.cc*
Реализация методов PHP-расширения.

2.2 Конкретные файлы в репозитории

config.h

В этом файле содержатся все константы, используемые во всех файлах.

server.h

В этом файле содержится объявление всех классов и их методов. Также там присутствует описание функций программного интерфейса.

sounds_cpp_api.cc

Реализация методов программного интерфейса.

main.cc

Этот файл предназначен только для тестирования различных функций.

2.3 Документация: библиотека C++

В этом разделе перечислен весь функционал библиотеки с указаниями, как им пользоваться. Также к некоторым методам приведены примеры использования.

Информация об аудиозаписи

```
vector<double> sounds_info(const string &name);
```

`name` — полный путь к файлу относительно текущей директории.

Возвращает вектор из 6 элементов:

- [0] — Размер файла в байтах.
- [1] — Число аудиоканалов (1 — моно, 2 — стерео).
- [2] — Sample rate (количество значений амплитуды каждую секунду), обычно 44100.
- [3] — Bit depth (число битов на каждый sample).
- [4] — Количество sample'ов в аудиофайле.
- [5] — Продолжительность аудиофайла (в секундах).

Если элемент вектора [0] = -1, то ошибка.

Распознавание музыкальных инструментов

```
vector<double> sounds_classify(const string &name);
```

`name` — полный путь к файлу относительно текущей директории.

Возвращает вектор из 5 элементов с вероятностями музыкальных инструментов:

- [0] — аккордион
- [1] — пианино
- [2] — скрипка

[3] — гитара

[4] — флейта

Если элемент вектора [0] = -1, то ошибка.

Обрезка аудиозаписей

```
string sounds_crop(const string &name, const string &new_name,
                  double left = 0, double right = -1);
```

`name` — полный путь к файлу относительно текущей директории

`new_name` — полный путь с новым названием преобразованного файла

`left = 0` — левая граница обрезки в миллисекундах (default: начало файла)

`right = -1` — правая граница обрезки в миллисекундах (default: конец файла)

Если `left` ≤ 0 или \geq длины файла, то левая граница — начало файла. Если `right` ≤ 0 или \geq длины файла, то правая граница — конец файла. По умолчанию, границы — начало и конец файла соответственно.

Обрезает аудиозапись и сохраняет ее в указанную директорию под выбранным пользователем именем.

Возвращает строку:

- "OK" — успешно
- `string` — строка с текстом ошибки, если неуспешно

Изменение громкости

```
string sounds_volume(const string &name, const string &new_name,
                    double k, double left = 0, double right = -1, bool smooth = false);
```

`name` — полный путь к файлу относительно текущей директории

`new_name` — полный путь с новым названием преобразованного файла

`k` — множитель громкости (0 — тишина, 1 — без изменений)

`left = 0` — левая граница в миллисекундах (default: начало файла)

`right = -1` — правая граница в миллисекундах (default: конец файла)

`smooth = false` — плавное изменение громкости (линейное, в течение 2 секунд)

Если `left` ≤ 0 или \geq длины файла, то левая граница — начало файла. Если `right` ≤ 0 или \geq длины файла, то правая граница — конец файла. По умолчанию, границы — начало и конец файла соответственно.

Если `k` < 0 , то ошибка, если `k` = 0, то тишина, если $0 < k < 1$, то аудиозапись становится тише, если `k` > 1 , то аудиозапись становится громче. **Внимание:** если выбрать `k` очень большим (обычно ~ 5), то для нормального звучания не хватает битовой глубины, и звук становится обрывистым (что иногда звучит забавно).

Изменяет громкость выбранного отрезка аудиозаписи и сохраняет ее в указанную директорию под выбранным пользователем именем.

Возвращает строку:

- "OK" — успешно
- `string` — строка с текстом ошибки, если неуспешно

Изменение скорости аудиозаписи

```
string sounds_speed(const string &name, const string &new_name,
                   double mult);
```


name — полный путь к файлу относительно текущей директории
new_name — полный путь с новым названием преобразованного файла
mult — множитель скорости аудиозаписи (1 — без изменений)

Если **mult** = 0, то ошибка. Если $0 < \text{mult} < 1$, то аудиозапись становится медленнее в $1/\text{mult}$ раз. Если **mult** = 1, то без изменений. Если **mult** > 1, то аудиозапись ускоряется в **mult** раз. Если **mult** < 0, то аудиозапись проигрывается в обратную сторону (инвертируется).

Изменяет скорость аудиозаписи и сохраняет ее в указанную директорию под выбранным пользователем именем.

Возвращает строку:

- "OK" — успешно
- **string** — строка с текстом ошибки, если неуспешно

Слияние аудиозаписей

```
string sounds_merge(const string &left, const string &right,  
                   const string &new_name);
```

left — полный путь к первому файлу относительно текущей директории
right — полный путь ко второму файлу
new_name — полный путь с новым названием преобразованного файла

Сливает две аудиозаписи в одну и сохраняет ее в указанную директорию под выбранным пользователем именем.

Возвращает строку:

- "OK" — успешно
- **string** — строка с текстом ошибки, если неуспешно

2.4 Документация: расширение RHP

В этом разделе перечислен весь функционал расширения RHP с указаниями, как им пользоваться. Также к некоторым методам приведены примеры использования. Здесь написано то же самое, что и в разделе по библиотеку C++, но в терминах RHP.

Информация об аудиозаписи

```
sounds_info($name);
```

name — полный путь к файлу относительно текущей директории.

Возвращает массив из 6 элементов:

- [0] — Размер файла в байтах.
- [1] — Число аудиоканалов (1 — моно, 2 — стерео).
- [2] — Sample rate (количество значений амплитуды каждую секунду), обычно 44100.
- [3] — Bit depth (число битов на каждый sample).
- [4] — Количество sample'ов в аудиофайле.
- [5] — Продолжительность аудиофайла (в секундах).

Если элемент массива [0] = -1, то ошибка.

Распознавание музыкальных инструментов

```
sounds_classify($name);
```

`name` — полный путь к файлу относительно текущей директории.

Возвращает массив из 5 элементов с вероятностями музыкальных инструментов:

[0] — аккордион

[1] — пианино

[2] — скрипка

[3] — гитара

[4] — флейта

Если элемент массива [0] = -1, то ошибка.

Обрезка аудиозаписей

```
sounds_crop($name, $new_name, $left = 0, $right = -1);
```

`name` — полный путь к файлу относительно текущей директории

`new_name` — полный путь с новым названием преобразованного файла

`left = 0` — левая граница обрезки в миллисекундах (default: начало файла)

`right = -1` — правая граница обрезки в миллисекундах (default: конец файла)

Если `left` ≤ 0 или \geq длины файла, то левая граница — начало файла. Если `right` ≤ 0 или \geq длины файла, то правая граница — конец файла. По умолчанию, границы — начало и конец файла соответственно.

Обрезает аудиозапись и сохраняет ее в указанную директорию под выбранным пользователем именем.

Возвращает строку:

- "OK" — успешно
- `string` — строка с текстом ошибки, если неуспешно

Изменение громкости

```
sounds_volume($name, $new_name, $k, $left = 0, $right = -1, $smooth = False);
```

`name` — полный путь к файлу относительно текущей директории

`new_name` — полный путь с новым названием преобразованного файла

`k` — множитель громкости (0 — тишина, 1 — без изменений)

`left = 0` — левая граница в миллисекундах (default: начало файла)

`right = -1` — правая граница в миллисекундах (default: конец файла)

`smooth = False` — плавное изменение громкости (линейное, в течение 2 секунд)

Если `left` ≤ 0 или \geq длины файла, то левая граница — начало файла. Если `right` ≤ 0 или \geq длины файла, то правая граница — конец файла. По умолчанию, границы — начало и конец файла соответственно.

Если `k` < 0 , то ошибка, если `k` = 0, то тишина, если $0 < k < 1$, то аудиозапись становится тише, если `k` > 1 , то аудиозапись становится громче. **Внимание:** если выбрать `k` очень большим (обычно ~ 5), то для нормального звучания не хватает битовой глубины, и звук становится обрывистым (что иногда звучит забавно).

Изменяет громкость выбранного отрезка аудиозаписи и сохраняет ее в указанную директорию под выбранным пользователем именем.

Возвращает строку:

- "OK" — успешно

- **string** — строка с текстом ошибки, если неуспешно

Изменение скорости аудиозаписи

```
sounds_speed($name, $new_name, $mult);
```

name — полный путь к файлу относительно текущей директории
new_name — полный путь с новым названием преобразованного файла
mult — множитель скорости аудиозаписи (1 — без изменений)

Если **mult** = 0, то ошибка. Если $0 < \text{mult} < 1$, то аудиозапись становится медленнее в $1/\text{mult}$ раз. Если **mult** = 1, то без изменений. Если **mult** > 1, то аудиозапись ускоряется в **mult** раз. Если **mult** < 0, то аудиозапись проигрывается в обратную сторону (инвертируется).

Изменяет скорость аудиозаписи и сохраняет ее в указанную директорию под выбранным пользователем именем.

Возвращает строку:

- "OK" — успешно
- **string** — строка с текстом ошибки, если неуспешно

Слияние аудиозаписей

```
sounds_merge($left, $right, $new_name);
```

left — полный путь к первому файлу относительно текущей директории
right — полный путь ко второму файлу
new_name — полный путь с новым названием преобразованного файла

Сливает две аудиозаписи в одну и сохраняет ее в указанную директорию под выбранным пользователем именем.

Возвращает строку:

- "OK" — успешно
- **string** — строка с текстом ошибки, если неуспешно

3 Результаты

3.1 Нейронная сеть

Для работы с аудиофайлами обычно используются рекуррентные нейронные сети. Тем не менее, решение задачи классификации музыкальных инструментов мы решили свести к задаче классификации изображений и использовать для этого более простой тип нейронных сетей — свёрточные.

Исходный аудиофайл разбивается на фрагменты длиной в одну секунду, после чего для каждого из них строится спектрограмма. Все такие спектрограммы подавались на вход нейронной сети, которая по картинке определяла, что это за инструмент. Архитектура модели была тщательно подобрана, что подтверждается высоким процентом корректных ответов на тестовых данных.