

Летняя международная XXXI молодежная Школа-конференция по параллельному программированию

Определение пола по голосу

Участник:

Дроздов Э.К., НГТУ,
ФПМИ, 3 курс

Руководитель:

Компания Dasha.AI

02.07.18 – 13.07.18

г. Новосибирск

Содержание

- Задача
- Схема решения
- Детали реализации
- Результаты

Задача

- Дан фрагмент аудиофайла с записью человеческого голоса. Требуется определить пол говорящего.

Решение. Общая схема.

- Голоса мужчин и женщин различаются по высоте
- Фундаментальная частота – характеристика высоты голоса.

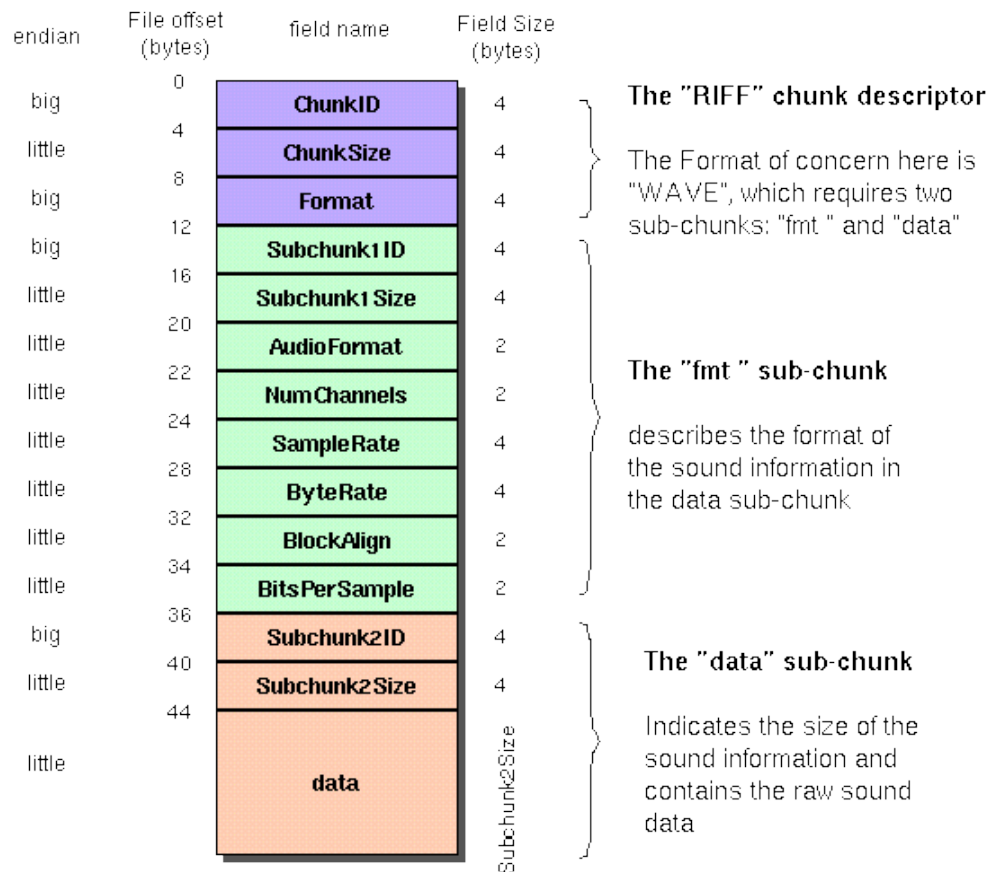
Решение. Общая схема.

- Аудио-данные → сигнал → спектр → кепстр
→ фундаментальная частота →
предполагаемый пол

Детали решения

The Canonical WAVE file format

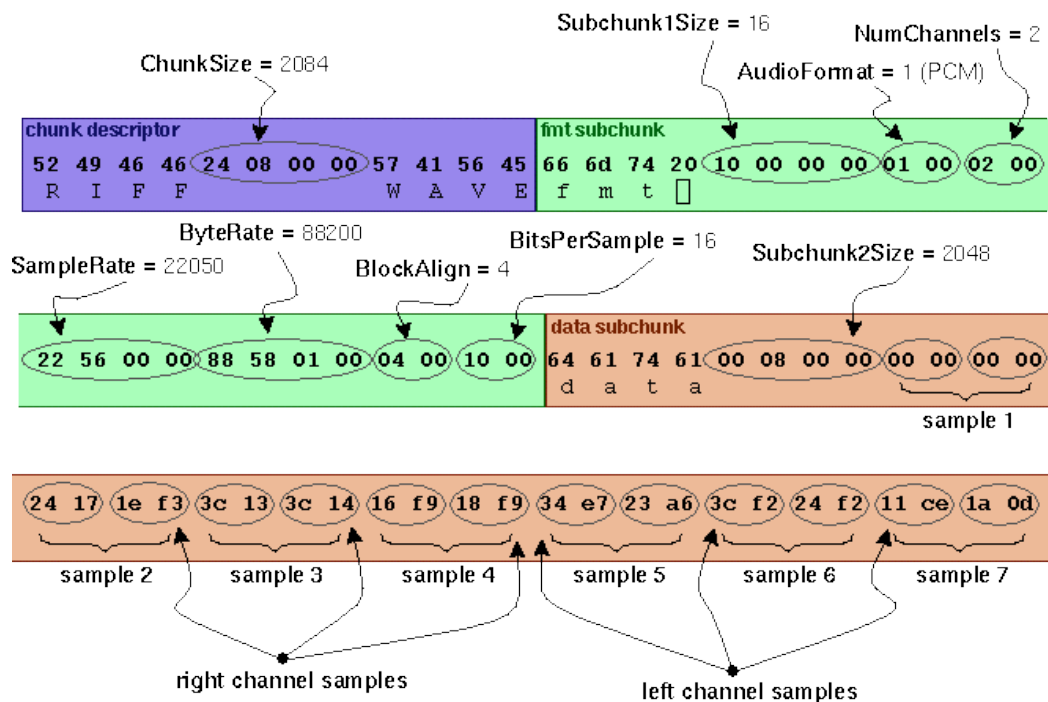
- Структура *.wav-файла



Детали решения

- Структура *.wav-файла

Многоканальность



Детали решения

- Разложение сигнала в спектр
- Преобразование Фурье

$$(x_1, x_2, \dots, x_n) \rightarrow \begin{bmatrix} x_1 & x_2 & \dots & x_{n-m} \\ x_2 & & & x_{n-m+1} \\ \vdots & & & \vdots \\ x_m & x_{m+1} & \dots & x_n \end{bmatrix} \quad M - \text{размер окна}$$

N – длина сигнала в семплах

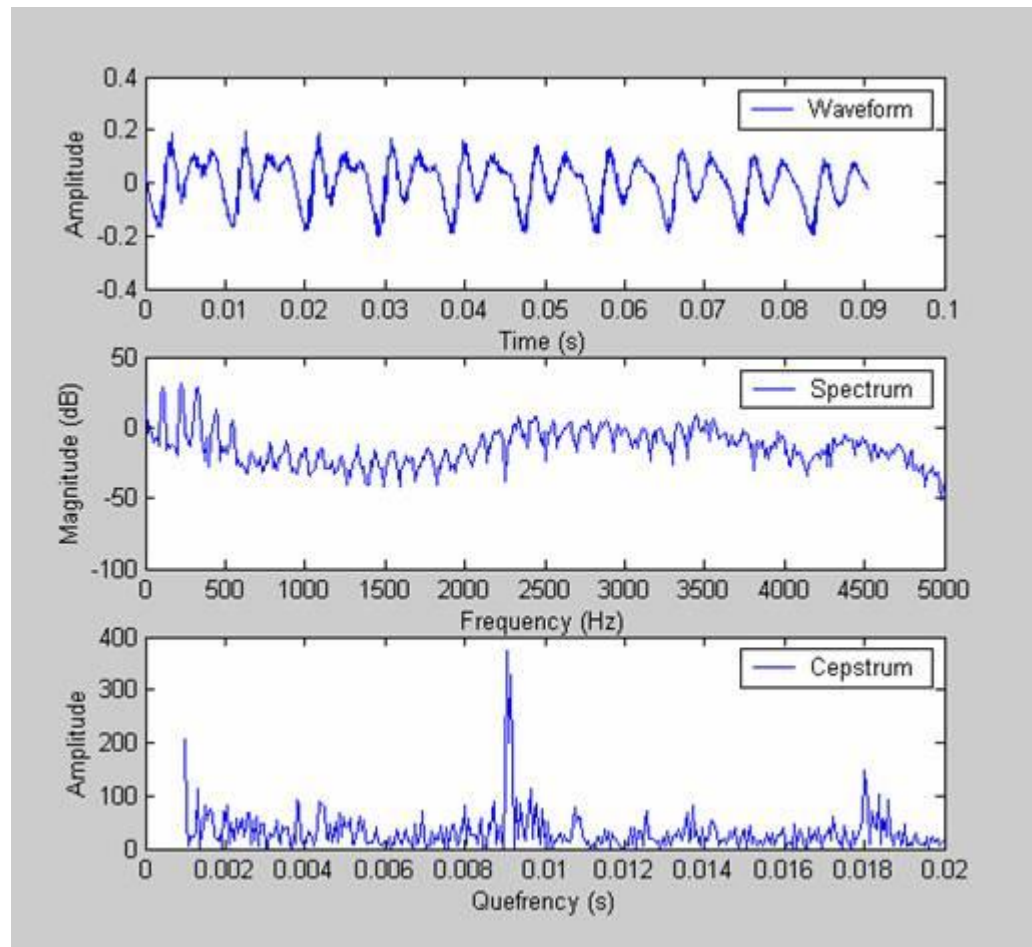
Детали решения

- Кепстр $C = IDTF(\log(S))$

S – матрица спектров,

DTF – обратное

преобразование Фурье



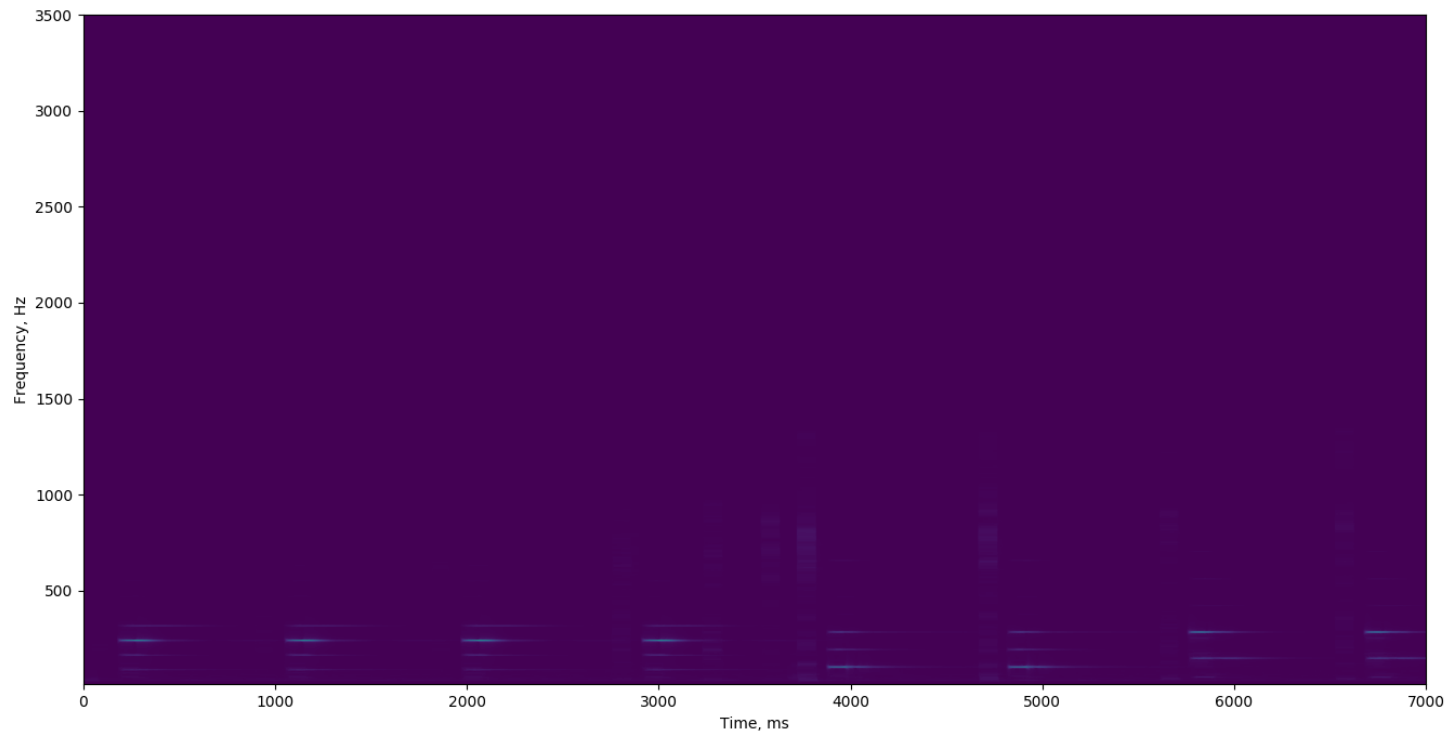
Детали решения

- Фундаментальная частота
- Мужской диапазон: 85 – 155 Гц
- Женский диапазон: 165 – 255 Гц

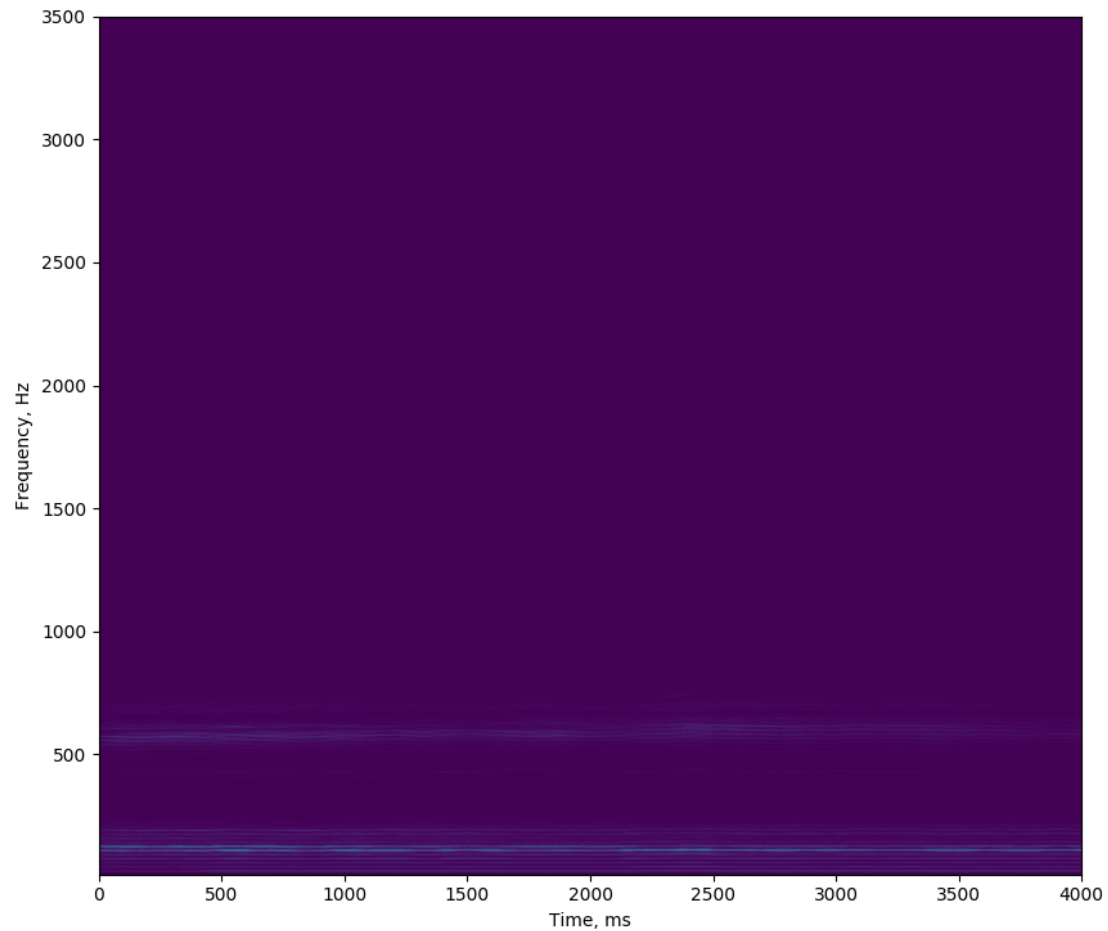
Инструменты

- Python 3.0
 - NumPy
 - Matplotlib
 - SciPy (тестирование)
 - Wave

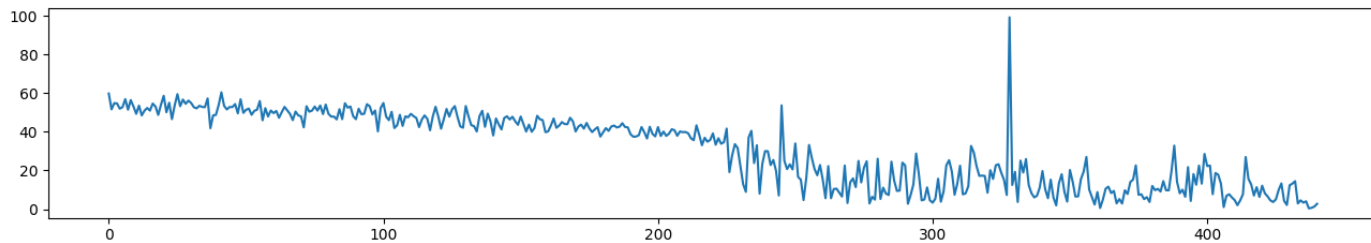
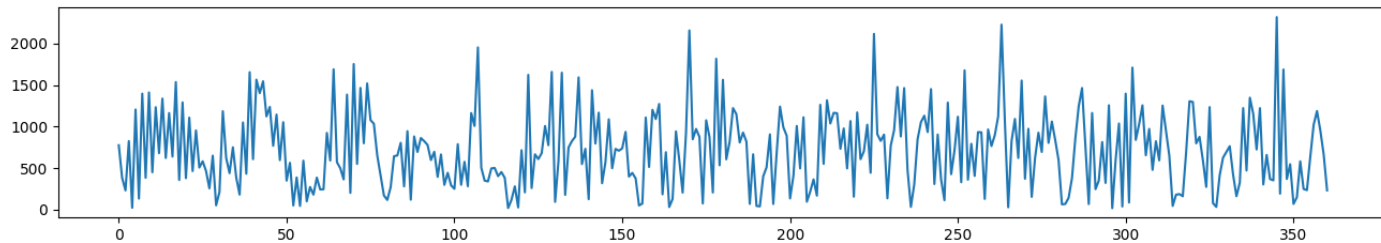
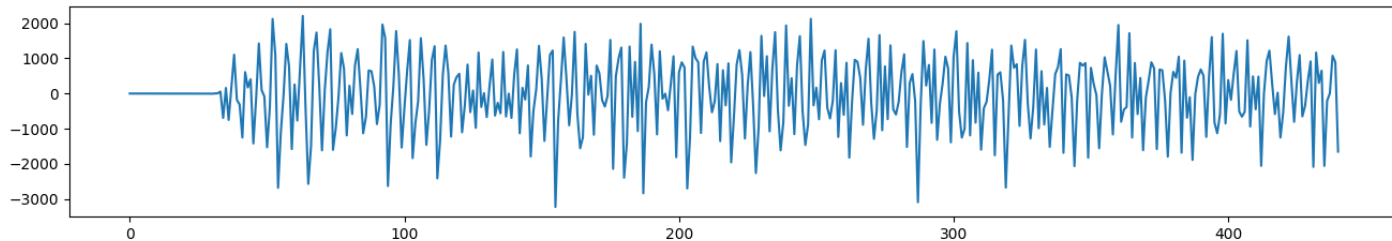
Результат



Результат



Результат



Результат

- Фундаментальная частота для записи женского голоса: 203 Гц
- Для мужского: 152 Гц

Дальнейшие планы

- Калибровка осей
- Нормализация сигнала, удаление шумов
- Работа в режиме реального времени

Спасибо за внимание