

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

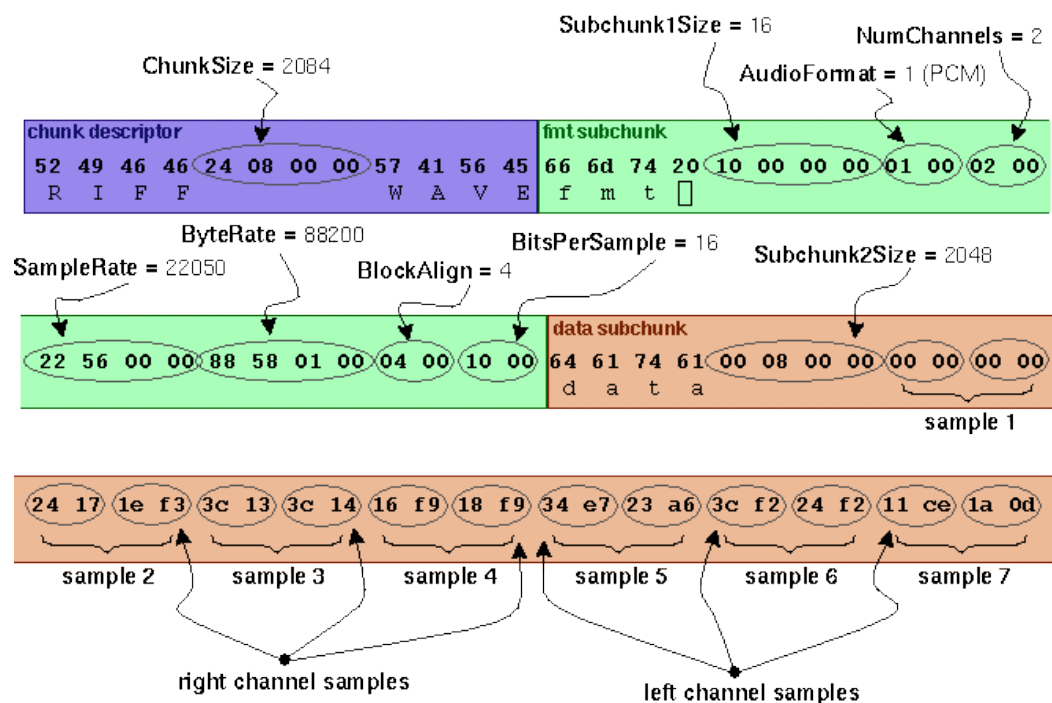
- Дан фрагмент аудиофайла с записью человеческого голоса. Требуется определить пол говорящего.

Решение

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

Детали решения

• Структура *.wav-файла



The Canonical WAV file format

endian

big

little

big

big

little

little

little

little

little

little

little

big

little

little

File offset
(bytes)

field name

Field Size
(bytes)

0	ChunkID	4
4	ChunkSize	4
8	Format	4
12	Subchunk1 ID	4
16	Subchunk1 Size	4
20	AudioFormat	2
22	NumChannels	2
24	SampleRate	4
28	ByteRate	4
32	BlockAlign	2
34	BitsPerSample	2
36	Subchunk2ID	4
40	Subchunk2 Size	4
44	data	Subchunk2Size

The "RIFF" chunk descriptor

The Format of concern here is "WAVE", which requires two sub-chunks: "fmt" and "data"

The "fmt" sub-chunk

describes the format of the sound information in the data sub-chunk

The "data" sub-chunk

Indicates the size of the sound information and contains the raw sound data

Детали решения

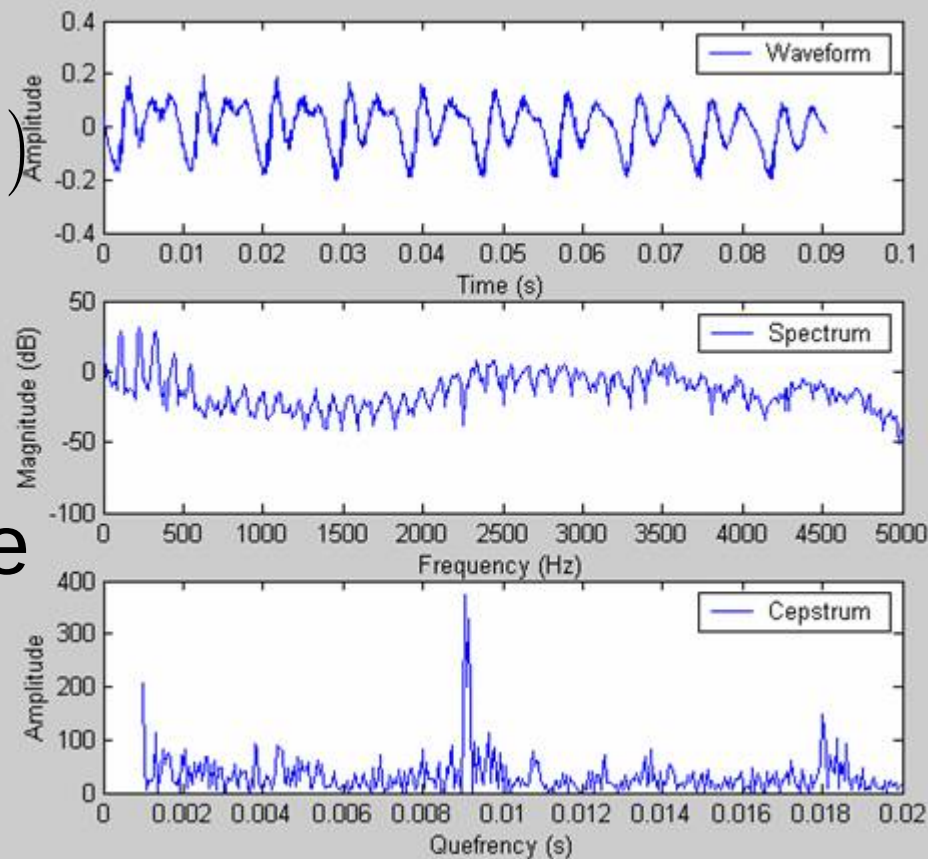
- Сигнал \rightarrow преобразование Фурье \rightarrow спектр

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

N – длина сигнала в семплах

Детали решения

- Кепстр $C = IDTF(\log(S))$
 S – матрица спектров,
 $IDTF$ – обратное
преобразование Фурье



Детали решения

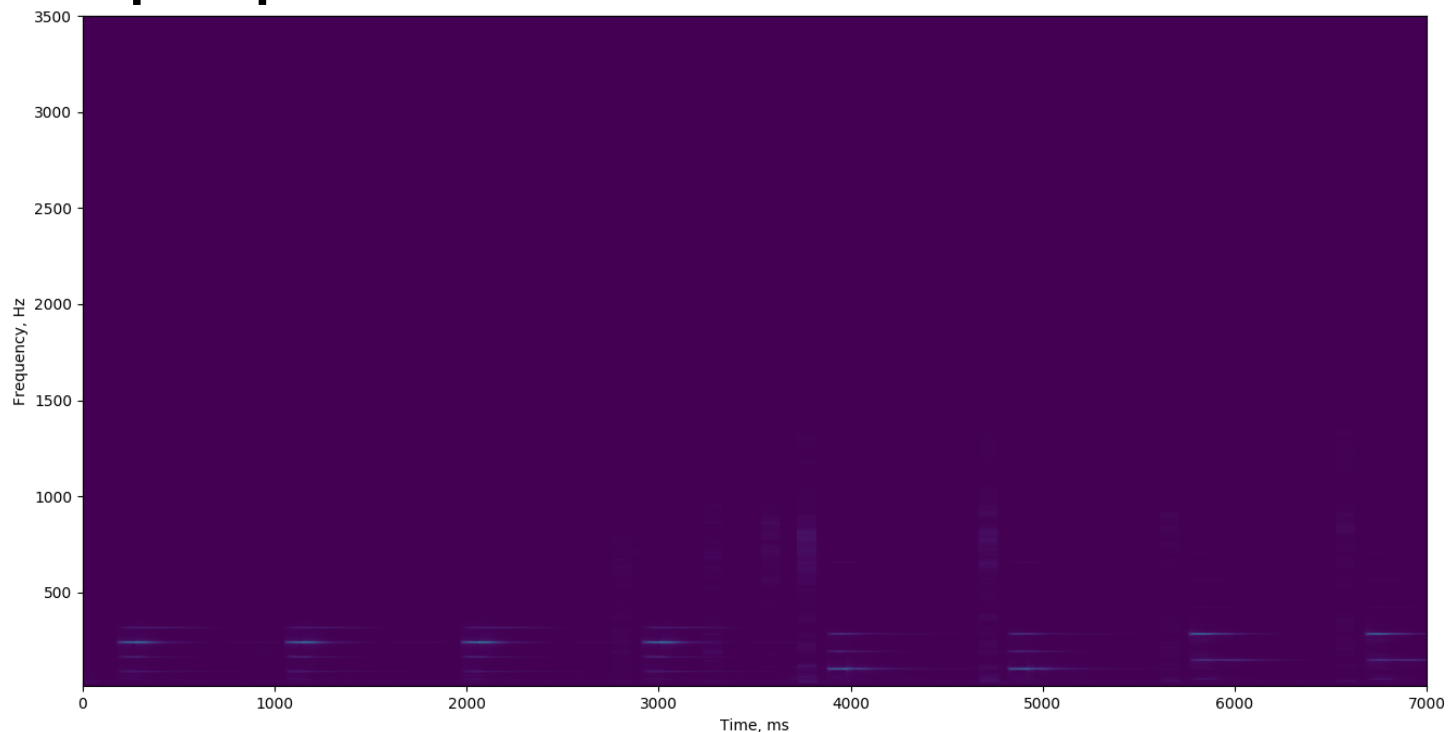
- Фундаментальная частота
- Мужской диапазон: 85 – 155 Гц
- Женский диапазон: 165 – 255 Гц

Инструменты

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

Результат

- Спектрограмма



Результат

- Графики

