

Arnes Respati Putri

16/392817/PA/17084

21 April 2019

Pattern Recognition Assignment

One-dimensional Data Recognition Implementation with Python Report

I. Introduction

The purpose of the system is to recognize five given commands. In this implementation, the system works such that it can recognize audio signal inputs as either one of five common genres of music, namely rock, blues, jazz, soul and pop. This system can then be used further for softwares like Spotify, Youtube, Apple Music, *etc.* The sample audio data, used for training, is taken manually by recording speech from four different people.

II. Implementation

The system is implemented in Python. There are several libraries that are used, namely numpy, scipy and matplotlib. The library `scipy.io.wavfile` is used for reading from and writing to the audio data.

There are several basic steps of one-dimensional signal processing. The steps are as follows.

1. Signal Normalization

a. Signal normalization

Normalization of the signal is necessary prior to comparing it with respect to other signals.

The result of before and after normalization can be seen in Figure 1 and Figure 2, respectively.

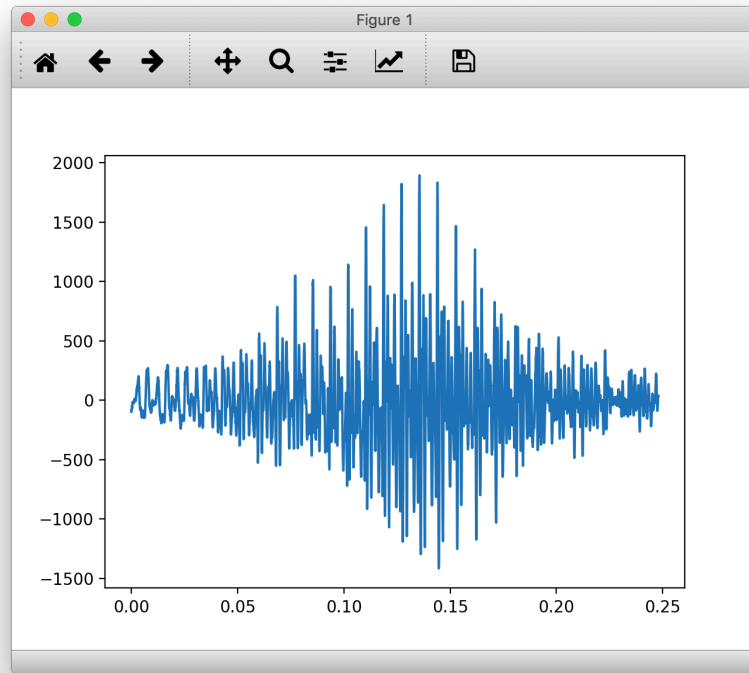


Figure 1. Before signal normalization.

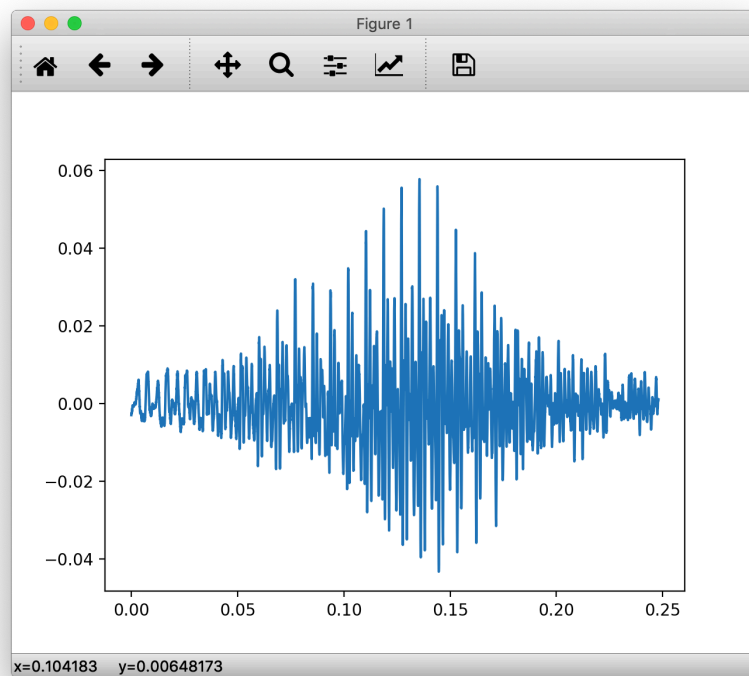


Figure 2. After signal normalization.

b. Time warping

Time warping is done so all audio data in the same genre or category have the same duration of time. The frequency that is used is 44100 Hz.

The result of before and after warping can be seen in Figure 3 and Figure 4, respectively.

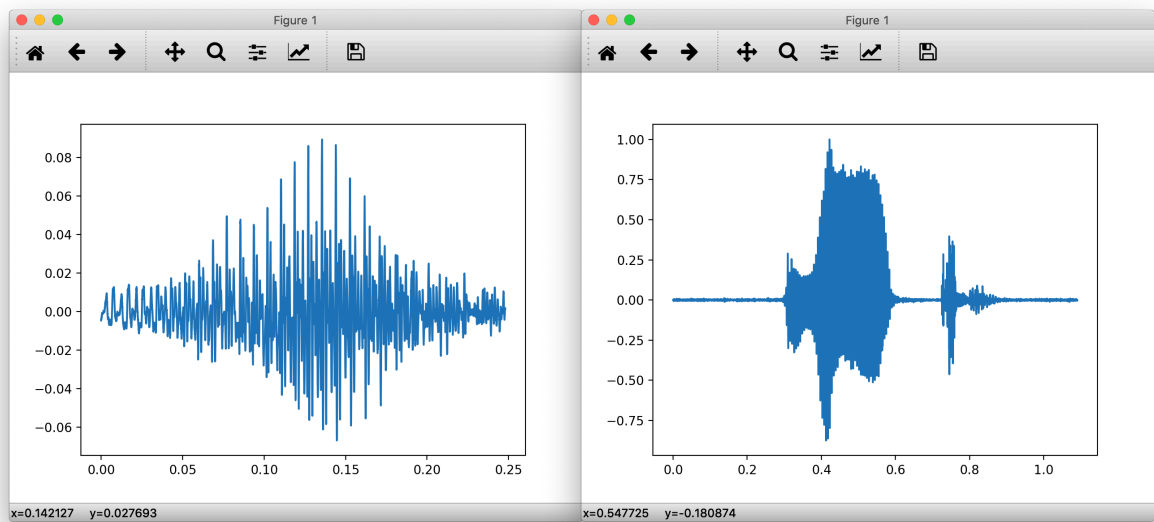


Figure 3. Two different signal from one category before warping.

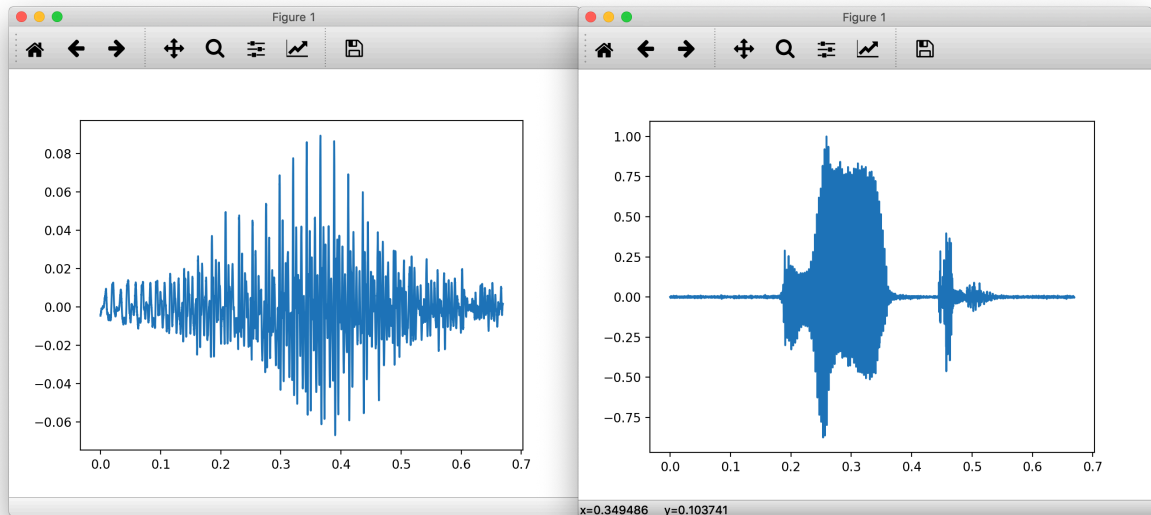


Figure 4. Two different signal from one category after warping.

2. Segmentation

The data in each category is divided into 10 segments. The segmentation program takes place directly inside the feature extraction method.

3. Feature Extraction

The feature that is being extracted in the implementation is the average energy, where the average of the audio signal in each segment is calculated.

4. Feature Matching

After the input data is read and processed based on the previous steps, it is being matched with the sample or the training data. This is done by finding the similarity of the input data and the training data. The similarity is obtained by calculating the euclidean distance of the average energy of each segment of the input data and the training data of all samples in each genre category.

III. Result

The result of the implementation is as follows.

```
(base) arness-MacBook-Pro:de
Speech 1 recognized as Pop
Speech 2 recognized as Blues
Speech 3 recognized as Rock
Speech 4 recognized as Soul
Speech 5 recognized as Rock
```

Figure 5. Result of the program code.

Input Data	Prediction Output	Actual Output
test_audio_1	Pop	Rock
test_audio_2	Blues	Blues
test_audio_3	Rock	Jazz
test_audio_4	Soul	Soul
test_audio_5	Jazz	Pop

Table 1. Result of implementation.

Table 1. above depicts the output of the prediction and the actual output of the test data. Out of five test data, two give the correct output giving the accuracy of 40%.

IV. Conclusion

The result of the feature matching can still be improved since in this implementation it only uses one feature and that the similarity measurement is only by calculating the euclidean distance. If there are more than one feature, the distance should be calculated by similarity measurement algorithm that is better to use for data with higher dimensionality, such as the Manhattan and Cosine algorithm. Furthermore, the lack of sample data also contributes to the model's accu-

racy. There are only five sample data in each category and have not fully represent the signal data of spoken words in each category.