ECE397 Project Report
— Instrument Recognition

Tianhao Yu, Ruohua Li

Abstraction

This project is programmed to recognize the type of the instrument from a pre-loaded * .wav file. Using FFT and Bayes algorithms, we are able to train the computer with wav files. The computer can then recognize the instrument used to perform the .wav file under test.

Our project is based on the Bayes theorem and corresponding algorithms. Brown(1999) mentioned in his article the Bayes decision rule such that the accuracy of judging would be less affected by single outliers. Similar methods and formula would be applied in our project in the decision-making part of the algorithm.

Objective

The goal of the project is to distinguish between multiple types of instruments based on the training data provided. The computer will be able to tell the difference between different training spectrums and, based on the training data set, calculate the probability that the test data is performed by a certain instrument. The final output would be the computer's guess of the type of instrument.

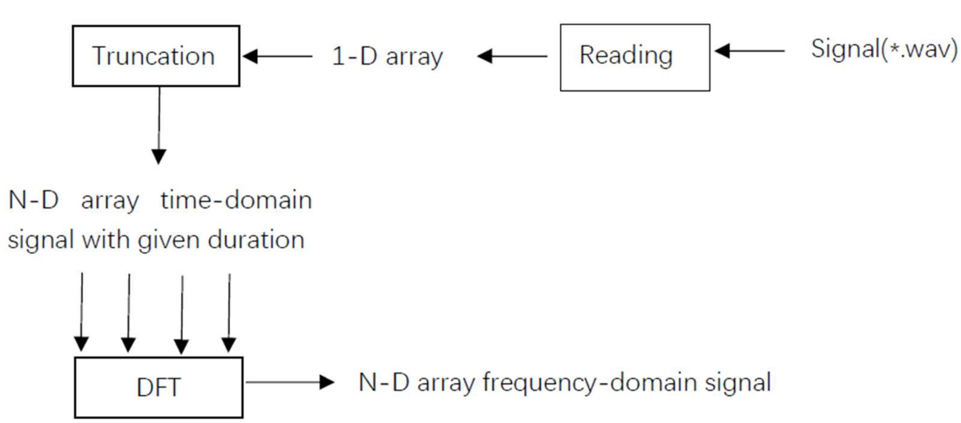
Approach & Algorithms

Pre-decision processing:

To prepare the training set for the machine to be able to recognize the instruments, we prepared a list of categorized *.wav files which will later will be used to set up the training set via Discrete Fourier Transform. After reading in the files, taking the unwanted effect of the transient parts of the audio files into consideration, we first truncate the array containing the data of a given audio into pieces with constant time durations then delete the parts in the front/back that are below the average amplitude value for each piece. After the truncation, we then take 1000-point DFT of each piece to generate the spectrum we finally used for training.

To further enhance the program's capability to recognize correctly, we also ask it to learn the variance of different types of instruments. So when the program is trying to decide which instrument was used to perform the test signal, it will first take the variance of the test signal and categorize the test

Basic work flow for processing of test input/training set:



signal and give every instrument in the category an advantage on initial probability.

Bayes Net Decision:

The training and deciding part of the algorithm is based on the Bayes Theorem.

Bayes Theorem:

Bayes' theorem is stated mathematically as the following equation:[2]

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

where A and B are events and $P(B) \neq 0$.

- $P(A \mid B)$ is a conditional probability: the likelihood of event A occurring given that B is true.
- $P(B \mid A)$ is also a conditional probability: the likelihood of event B occurring given that A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B independently of each other; this is known as the marginal probability.

from Wikipedia Bayes' theorem

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)\times P(A)}{P(B)} = P(B|A) \times \frac{P(A)}{P(B)} \propto P(B|A)$$

P(A|B) (The probability of A given B) is proportional to P(B|A) (the probability of B given A). The probability of an instrument given a certain spectrum observed is proportional to the probability of the spectrum observed given it is performed by the instrument. Based on this theorem, we can deduce the type of the instrument given the spectrum produced. Therefore, we can train the computer to recognize spectrums of a certain kind of instrument using the corresponding instrument solo *.wav files (training data set).

The training process contains counting the occurrence of a certain spectrum in training data, smoothing data, and calculating probabilities. The occurrence of spectrum allows the computer to get hold of the basic distribution of P(instru|spec). However, the training data is limited--if the test file contain spectrum we have never seen, the occurrence would return 0 and thus make the probability 0. A single observation should never change the result significantly. This 0 probability would make the cumulative probability 0 even if the rest of them are high probabilities, which would spoil our outcome. Therefore, we smooth the data so that the "unknown sounds" are assigned with a low value instead of 0. Finally, we normalize the values to make them valid probabilities.

For a given input test file, we process it using the same FFT process as the training set. The spectrums of the file, based on the statement above, can be used to deduct the instruments. To make the judgement more accurate, we take multiples of the probabilities (Πp_i) to reduce the effect of possible outliers or "unknown sounds". Another potential problem here is the downflow of data. Since the probabilities are all fractions, the multiple would get smaller as calculation goes on. To prevent downflow, we take logs of the probabilities and sum them up during our actual algorithm.

$$\Sigma \log(Pi) = \log(\Pi Pi)$$

The final output is decided based on the log numbers. The larger the value is, the more likely it would be the correct answer. We then return the corresponding name of the instrument based on the values.

Result

```
piano correction rate =  90.0 %
guitar correction rate =  80.0 %
flute correction rate =  90.0 %
correction rate =  85.71428571428571 %
```

The outcome of our recognizing program is decent: it reached a correction rate of 85.71% with most regular solo files successfully recognized.

test type	flute test	piano test	guitar test
number of test	10	15	10
correction rate	90.0%	90.0%	80.0%

The problem lies in piano wav files with higher pitches and guitar files. The high frequency would be more likely to be mistakenly recognized. The accuracy is lowered when two pitches are simultaneously played. Also, wind and string instruments are different in their spectrums and can be distinguished by their patterns, which is not fully accomplished in the recognizing process.

Limitations & Future Perspective

Currently there are various limitations of the program, especially around training set. The training set used for this version of the program are mainly composed of single pitched, fixed-duration signals, which could not fully take advantage of Bayes' deciding algorithm in increasing the probability of correct recognition. To fully exploit the value of Bayes deciding algorithm, the next step would be introducing real solo music pieces in diverse categories in large scale. There is also another problem with processing. For now the processing algorithm deployed is Fast Fourier Transform, which fails to take time scale into consideration, causing problems when comparing signals with different durations and unstable signals. Solving this problem requires a different algorithm for the program such as Wavelet Transform. Also we are currently relying on the difference in variance to recognize different categories of instrument(e.g. wind instruments versus string instruments). Without this pre-processing, the machine is sometimes unable to differentiate between two instruments that belong to different categories, which can be easily accomplished by human ears. So in the future, this feature will need to be improved.

Reference

Judith C. Brown (1999). Computer identification of musical instruments using pattern recognition with cepstral coefficients as features. *The Journal of the Acoustical Society of America 105*, 1933.

Bayes Theorem (n.d.). Retrieved from https://en.wikipedia.org/wiki/Bayes%27_theorem