String Instrument Recognition

Jian-Lung(Larry) Wu

Media Art and Technology Dept., UCSB University

1-650-796-6024

f7488405@gmail.com

ABSTRACT

In this paper, the design and implementation of the string instrument recognizer are presented. The Hidden Markov Model is used for modeling the input signal. To find out the sequence of four hidden states (Attack, Decay, Sustain, and Release) in the Hidden Markov Models, the Viterbi algorithm is used. The best matched string instrument model for the input signal is the model with the highest probability of the Viterbi path.

KEYWORD

Music Instrument Recognition, Hidden Markov Model (HMM), Music Information Retrieval.

1. INTRODUCTION

Music instrument recognition is the process of retrieving the features of the input signal and identifying this signal by comparing its features to every music instrument's features in database. There is a scoring system to determine which instrument the input signal belongs to. In our case, the input string instrument signal is decomposed into four states such as attack, decay, sustain, and release (ADSR). The features used for recognition are the mean and variance value of each state. To find out the best separation of each state is based on the properties in each time frame of the input signal. This process is similar to find out the most likely sequence of hidden states in the Hidden Markov Model. Therefore, the Hidden Markov Model is used for modeling the input signal. The Viterbi algorithm, a dynamic programming algorithm, is used for finding the most likely sequence of hidden states in the Hidden Markov Model. In other words, the Viterbi algorithm is used for finding the best separation of the four states (ADSR) in the input signal.

The rest of the paper is organized as follows. The technique of music instrument recognition is given in section 2. In section 3, conclusion is discussed. Future works are in section 4.

2. STRING INSTRUMENT RECOGNITION

2.1 Overview

Figure 1 shows the overview of recognition method. At first, users need to build the instrument-models by recording the instruments' sound as training samples. After finishing all instrument-models building, these instrument -models could be used in the scoring system. Scoring system will determine which instrument the test sample belongs to by finding the maximum probability of Viterbi paths.

2.2 Training System

There are three options of learning algorithm for the implementation of the training system. 1) Neural network-based training. 2) Vector quantization-based training. 3) Hidden Markov model (HMM)-based training. The Neural Network algorithm is capable of solving much more complicated recognition tasks, but do not scale as well as HMMs when it comes to large

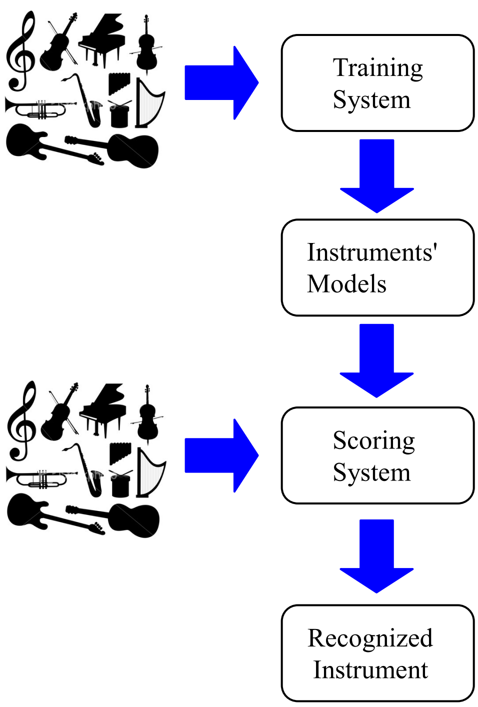

Fig.1. Recognition method

music instruments database. Rather than being used in general-purpose music instrument recognition applications, Neural Network can handle low quality, noisy data and player independence case. The vector quantization algorithm is originally a quantization technique often used in lossy data compression in which the basic idea is to code or replace with a key, but it could also be used in music instrument recognition. The shortage is that when a new instrument is added to the database, the centroids of previous music instruments need to be recalculated. For the presented training system, HMM is used for modeling the training examples of string instruments. One reason is that the signal of a string instrument could be viewed as a piece-wise stationary signal or a short-time stationary signal. That is, one could assume in a short-time in the range of 10 milliseconds, string instrument signal could be approximated as a stationary process. The string instrument signal could thus be thought as a Markov model for many stochastic processes (known as states). Another reason is because the models can be trained separately and are simple and computationally feasible to use.

Observations of HMM are sets of 12 cepstral coefficients from each time frame. The reason of using cepstral coefficients is explained as following. As we know, the string instrument signal is generated by the convolution of the vibration of the string (excitation) and the impulse response of the string instrument body (resonance). Convolution in time domain is the multiplication in frequency domain. The log of the spectrum will make the multiplication become addition, because the log of a product of two numbers is the sum of the logarithms of those numbers. The "inverse Fourier transform of the log of the spectrum", as known as ceptral coefficients, contain the separated excitation signal and the separated impulse response of the string instrument body. The impulse response of the string instrument body will locate on the low quefrency part. As a result, the first twelve ceptral coefficients which well represent the music instrument are used as the observation to find out the sequence of hidden states in the Hidden Markov Model.

Generally, the string instrument signal has four states such as Attack, Decay, Sustain, and Release, so the number of HMM states can be assumed as four. To find out the sequence of hidden states in the Hidden Markov Models, the Viterbi algorithm is used. Mean and variance calculated from each state are used for the observation density which is simple diagonal covariance Gaussian with a single mixture component. The Viterbi path with maximum probability is almost converged by iterating the training processes five times. The state transition probabilities are disregarded by assuming they are all the same. The mean and variance for each state are saved as a string instrument's model, because they are needed in the scoring system.

2.3 Scoring System

The same procedure of training system is used for the scoring system. The difference is that the Viterbi algorithm is only run through one time. The best matched string instrument model for the input signal is the model with the highest probability of Viterbi path. At the same time, Viterbi algorithm also finds the optimal sequence of the four states (ADSR) of the input signal. The best matched model will be the recognition result.

3. CONCLUSION

If the training example is used as a test example, the recognition rate is 100%. That means all training examples can be recognized correctly. If the same string instrument plays the note one tone lower than the training example, the recognition rate is 78%. If the same string instrument plays the note one octave lower than the training example, the recognition rate is 47%. However, if these new test samples are added into the training process and produce new models for each string instrument, the recognition rate will become 89%. In other words, the more training examples are added to the training system, the better recognition rate the scoring system will have. Figure 2 shows the models which current training system has.

Name	Covered Note
Acoustic Guitar -Nylon String	G5, E5, C5
Acoustic Guitar -Steel String	G5, E5, C5
Banjo	G5, E5, C5
Electric Guitar -Clean	G5, E5, C5
Electric Guitar -Distortion	G5, E5, C5
Electric Guitar -Muted	G5, E5, C5
Koto	G5, E5, C5
Sitar	G5, E5, C5

Fig.2. Words V.S. Functions

4. FUTURE WORK

In the recognition procedure, the input signal is decomposed into four different states (Attack, Decay, Sustain, and Release). Therefore, various frequency domain modifications can be performed on each state to make the input signal sound like other types of string instruments. The other work could be done in the future is to port the Matlab codes to C++. One reason is that the efficiency will be improved a lot. Another reason is about the practicality of the application, because Multithread programming and Win32 API are available in C++ but not in Matlab.

5. REFERENCE

[1] Lawrence R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE, 77 (2), p. 257–286, February 1989.

[2] What HMMs can do, Jeff Bilmes, U. Washington Tech Report, Feb 2002