Connexions module: m13196

FUTURE WORK IN MUSICAL RECOGNITION*

Patrick Kruse Kyle Ringgenberg Yi-Chieh Wu

This work is produced by The Connexions Project and licensed under the Creative Commons Attribution License †

Abstract

A list of future ideas for the musical recognition project.

A number of changes and additions to this project would help it to scale better and be more statistically accurate. Such changes should help the project to handle more complex signals and operate over a larger number of musical instruments.

1 Improving the Gaussian Mixture Model

To improve the statistical accuracy, the Gaussian Mixture Model used in this project must improve. The features of this model help determine its accuracy, and choosing appropriate additional features is a step towards improving the project. These features may include modeling additional temporal, spectral, harmonic and perceptual properties of the signals, and will help to better distinguish between musical instruments. Temporal features were left out of this project, as they are difficult to analyze in polyphonic signals. However, these features are useful in distinguishing between musical instruments. Articulation, in particular, is useful in distinguishing a trumpet sound, and articulation is by its very nature a temporal feature.

Additionally, more analysis of what features are included in the Gaussian Mixture Model is necessary to improve the statistical accuracy. Too many features, or features that do not adequately distinguish between the instruments, can actually diminish the quality of the output. Such features could respond to the environment noise in a given signal, or to differences between players on the same instrument, more easily than they distinguish between instruments themselves, and this is not desirable. Ideally, this project would involve retesting the sample data with various combinations of feature sets to find the optimal Gaussian Mixture Model.

2 Improving training data

As training data for this experiment, we used chromatic scales for each instrument over its entire effective range, taken in a single recording session in a relatively low noise environment. To improve this project, the GMM should be trained with multiple players on each instrument, and should include a variety of music

^{*}Version 1.3: Dec 14, 2005 10:14 pm -0600

 $^{^\}dagger \rm http://creative commons.org/licenses/by/2.0/$

Connexions module: m13196 2

- not just the chromatic scale. It should also inlude training data from a number of musical environments with varying levels of noise, as the test data that later is passed through the GMM can hardly be expected to be recorded under the same conditions as the training recordings.

Additionally, the training of the GMM would be improved if it could be initially trained on some polyphonic signals, in addition to the monophonic signals that it is currently trained with. Polyphonic training data was left out of this project due to the complexity of implementation, but it could improve the statistical accuracy of the GMM when decomposing polyphonic test signals.

3 Increasing the scope

In addition to training the GMM for other players on the three instruments used in this project, to truly decode an arbitrary musical signal, additional instruments must be added. This includes other woodwinds and brass, from flutes and double reeds to french horns and tubas, to strings and percussion. The GMM would likely need to extensively train on similar instruments to properly distinguish between them, and it is unlikely that it would ever be able to distinguish between the sounds of extremely similar instruments, such as a trumpet and a cornet, or a baritone and a euphonium. Such instruments are so similar that few humans can even discern the subtle differences between them, and the sounds produced by these instruments vary more from player to player than between, say, a trumpet and a cornet.

Further, the project would need to include other families of instruments not yet taken into consideration, such as strings and percussion. Strings and tuned percussion, such as xylophones, produce very different tones than wind instruments, and would likely be easy to decompose. Untuned percussion, however, such as cymbals or a cowbell, would be very difficult to add to this project without modifying it, adding features specifically to detect such instruments. Detecting these instruments would require adding temporal features to the GMM, and would likely entail adding an entire beat detection system to the project.

4 Improving Pitch Detection

For the most part, and especially in the classical genre, music is written to sound pleasing to the ear. Multiple notes playing at the same time will usually be harmonic ratios of one another, either thirds, or fifths, or octaves. With this knowledge, once we have determined the pitch of the first note, we can determine what pitch the next note is likely to be. Our current system detects the pitch at each window without any dependence on the previously detected note. A better model would track the notes and continue detecting the same pitch until the note ends. Furthermore, Hidden Markov Models have been shown useful in tracking melodies, and such a tracking system could also be incorporated for better pitch detection.