David Sears
MUMT 621
10/23/09
jAudio

Introduction

jAudio is a software application intended to solve a critical issue in the music information retrieval community: providing a standard method for the extraction of features from audio recordings. The programmers for jAudio bundled a GUI interface and a command-line interface written in java with a collection of extraction algorithms, reflecting the authors' desire to make jAudio immediately accessible to a broad community of researchers.

In this paper I will outline a few of the related extraction systems currently found in the MIR community in order to provide a basis for comparison for jAudio's design and implementation. I will then discuss some of jAudio's unique features: namely, its use of multidimensional and metafeatures, it's solution to the calculation of dependencies within a list of features, and its extensibility. As a departure from jAudio's MIR-intended applications, I will finally provide a psycho-physiological application in which changes in a number of features extracted from a musical stimulus by jAudio can predict physiological responses in listeners.

Design

There have been a number of feature extraction systems designed over the last several years, each with a number of strengths and weaknesses. Marsyas, developed by George Tzanetakis, is an open-source feature extraction program written in C++. Although the authors of jAudio note that Marsyas is efficient and is integrated into a general classification system, one drawback is its complicated interface for controlling extracted features (McEnnis 2005). CLAM, a system developed by the Music Technology Group at Pompeu Fabra University, is an analysis/synthesis system implemented in C++. While its GUI user interface makes CLAM accessible, it was not intended for extracting features for classification problems. Sonic Visualiser, the most recent extraction system to appear in the community, is an open source visualization tool for extracted features from audio. Although its GUI interface and visualization methods make Sonic Visualizer extremely accessible and a powerful method for visualizing extracted data, like CLAM, it isn't intended for classification problems, meaning that the source and method for extracting features implemented in Sonic Visualizer are not immediately transparent to the researcher.

jAudio is a feature extraction system developed by Cory McKay and Daniel McEnnis at McGill University. It is an open source program that provides both a GUI and a command line interface to make extraction both transparent and accessible.

jAudio was written in Java because of its cross-platform portability, and jAudio supports multiple output formats, including the XML format of the ACE system, and the ARFF format (McEnnis 2005).

Unique Features

jAudio possess a number of unique features, including metafeatures, the resolution of dependencies, and extensibility. In addition to its core of 26 implemented features, jAudio provides 5 metafeatures: derivative, running mean, standard deviation, derivative of running mean, derivative of standard deviation (McKay in progress). Metafeatures are feature templates that can be applied to all of the core features to automatically generate derivative features. The appeal of such an addition is that researchers can automatically smooth their data, consider rate of change, or determine the amount of variation in a given feature over time, all by checking a box.

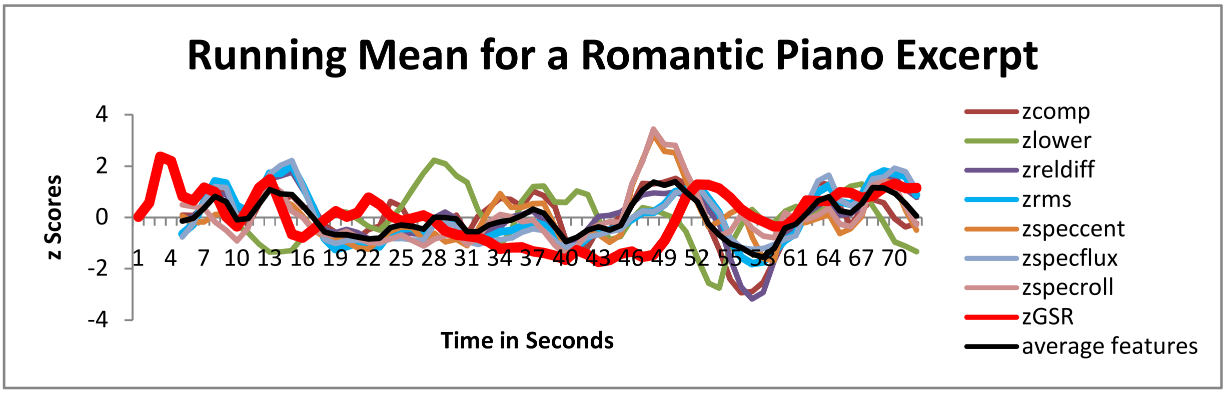
One possible difficulty researchers face during the extraction of a number of features from audio is that the extraction of one feature often depends on or necessitates the extraction of information from another feature. In order to solve the feature dependency problem, jAudio reorders the execution of feature calculations such that every feature's calculation is executed only after all of its dependencies have been executed. The user doesn't have to specify the order of feature extraction (McEnnis 2006).

Finally, jAudio includes a plugin folder that is automatically searched for compiled Java classes indicated in the configuration file (McEnnis 2006). It has therefore been designed so that researchers can add new features to jAudio without recompiling after every addition.

In addition to these features, jAudio also provides all of the necessary pre-processing options one expects in dealing with audio: input for multiple audio formats, output in multiple formats, normalization, downsampling, and the ability to change window size and window overlap so that the researcher can specify whether to have an averaged data point for a feature across the entire excerpt, or a sample of data points for some specific feature that reflect the change in that feature over time.

A Psycho-Physiological Application

Although jAudio's primary purpose is to provide a feature extraction system in order to solve classification problems, a topic relevant to the MIR community, its accessibility makes jAudio equally useful as a standalone extraction method for studying how the acoustic signal relates to analytical, perceptual, or psychophysiological domains of interest. One particular use of jAudio, for example, might be to determine how well changes in a host of different acoustic features might predict the arousal responses of listeners over time.



	RMS	Relative difference	Spectral centroid	Low energy	compactness	Spectral flux	Spectral roll off
GSR	.451	.254	.279	-.336	-.353	.501	.446

The figure above shows the running mean of standardized scores for 7 extracted features against galvanic skin response (GSR), a measure of psychologically-induced electrodermal activity located on the surface of the skin. The table beneath it indicates the cross correlation values at a lag of 1 second. Of the features selected, Spectral Flux and RMS Amplitude best predicted the GSR response of listeners.

Conclusion

jAudio is an extraordinarily useful and accessible feature extraction system. Although its feature extraction algorithms are transparent, the visualization tools could be substantially improved. On a personal note, I would also be interested to see jAudio implement psycho-acoustically motivated features such as loudness, roughness, or sharpness, thereby making jAudio useful as a method for studies involving perception and psychoacoustics.

References

McEnnis, D., C. McKay, and I. Fujinaga. 2006. jAudio: Additions and improvements. *Proceedings of the International Conference on Music Information Retrieval*. 385–6.

McEnnis, D., C. McKay, and I. Fujinaga. 2006. Overview of OMEN. *Proceedings of the International Conference on Music Information Retrieval*. 7–12.

McEnnis, D., C. McKay, I. Fujinaga, and P. Depalle. 2005. jAudio: A feature extraction library. *Proceedings of the International Conference on Music Information Retrieval*. 600–3.

McKay, C. 2010 (in press). Automatic music classification with jMIR. Ph.D. Dissertation. McGill University, Canada.