# Tutorial on Music Information Retrieval for Audio Signals

George Tzanetakis
Computer Science Department
Princeton University
35 Olden Street
Princeton, NJ 08544, USA
+1 609 258 1798

gtzan@cs.princeton.edu

## 1. OBJECTIVES

The main objective of this tutorial is to provide an overview of the current status of music information retrieval (MIR) for audio signals. The intended audience are people with a technical background who are interested to learn the main approaches and current status of MIR for audio signals. An important part of the intended audience would consist of researchers who have a background in symbolic MIR and/or musicology and music cognition and are interested to learn more about the similarities/differences of audio MIR to their respective fields. Demonstrations of several of the described algorithms and techniques will be part of the tutorial presentation.

The goals of this tutorial are:

1. To specify the main topics of audio MIR and describe the current state of the art in solving each topic.

2. To provide an overview of the necessary background such as signal processing and pattern recognition techniques.

3. To describe the main principles and building blocks that can be used to design and build audio MIR systems.

4. To demonstrate several techniques with specific examples.

5. To identify new directions and challenges.

### 1.1 Instructor's biography

George Tzanetakis is expected to receive his PhD degree in Computer Science from Princeton University in May 2002. His thesis titled "Manipulation, Analysis, and Retrieval Systems for Audio Signals" deals with algorithms and tools for audio information retrieval with special emphasis on musical signals. He has published papers dealing with various aspects of audio information retrieval such as feature extraction, segmentation, classification, beat analysis and various graphical user interfaces for browsing large audio collections.

### 1.2 Intended audience

Mainly researchers in symbolic MIR, musicology and music cognition who want to learn more about audio MIR. Basic Signal Processing and Machine Learning concepts will be covered in the tutorial. The tutorial should also be interesting for anyone who wants to get an overview of the current state of the art in audio MIR. Although the material will mostly target non-audio MIR researchers there will be some interesting new ideas and techniques also for audio MIR people.

### 1.3 Course materials

Handouts of the presentation slides will be provided as well as an extended annotated bibliography of papers and textbooks relevant to MIR for audio signals.

## 2. TUTORIAL OUTLINE

### 2.1 Introduction-Motivation

### 2.2 Representation

In this section various representations of audio signals that can be used for audio MIR will be described. More specifically proposed ways to represent timbral texture, rhythmic structure, and pitch content will be reviewed.

#### 2.2.1 Timbral texture feature extraction

Features based on Short Time Fourier Transform (STFT), Mel-Frequency Cepstral Coefficients (MFCC), Wavelets, Linear Principal Components (LPC), MPEG audio filterbank are some of the techniques that will be covered.

#### 2.2.2 Beat analysis

Review of various automatic beat detection and analysis algorithms that can be applied to audio data such autocorrelation based methods and onset-based methods.

#### 2.2.3 Pitch analysis

Analysis of harmonic content based on pitch histograms. Multiple pitch detection.

#### 2.2.4 Polyphonic transcription

Discussion of current state in polyphonic transcription. Although still far from being solved in the general case the proposed techniques potentially can provide valuable information to audio MIR.

### 2.3 Analysis

#### 2.3.1 Segmentation

Segmentation by timbral-texture. Techniques based on Hidden Markov Models (HMM), Clustering, and abrupt-change detection will be described.

#### 2.3.2 Classification

Review of standard pattern recognition classification algorithms. Various types of classification:

- Music Speech
- Male Female Voice
- Singinng detection
- Musical Genre Classification
- Artist classification

#### 2.3.3 Query-by-example content-based retrieval

Techniques for query-by-example.

#### 2.3.4 Thumbnailing

Methods for automatic thumbnailing: clustering-based, chroma-based, segmentation-based.

### 2.3.5 Fingerprinting

Definition of the problem and constraints, description of some proposed techniques.

### 2.3.6 Playlist generation

## 2.4 Interaction

Various graphical user interfaces can be used to browse, visualize and interact with large audio collections. In this section, several proposed ideas for such interfaces will be covered.

### 2.4.1 Viewers

### 2.4.2 Browsers

### 2.4.3 Monitors

### 2.4.4 Content-aware editors

### 2.4.5 Query interfaces

## 2.5 Discussion

In this section the main challenges that are facing audio MIR today will be highlighted. In addition issues related to the evaluation of audio MIR algorithms and systems will be discussed. The section and tutorial will end with some ideas for future directions for research.

### 2.5.1 Challenges

### 2.5.2 Evaluation

### 2.5.3 Future Directions