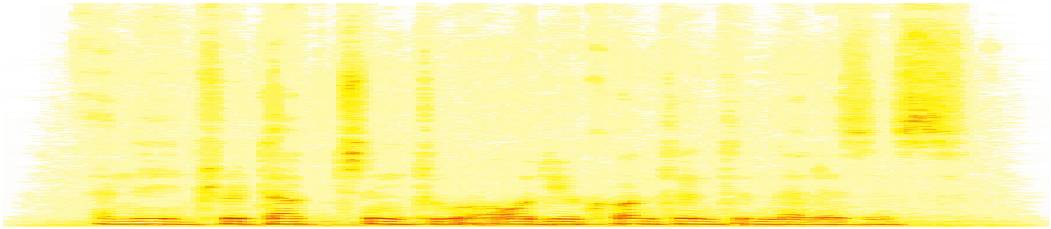
Introduction to Audio Content Analysis

Module 1.0: Introduction to MIR/ACA

alexander lerch



Georgia Tech Center for Music Technology
College of Design

introduction

overview

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Chapter 1 — Introduction: pp. 1–6

- **lecture content**
  - audio content analysis
  - typical applications
  - audio content
  - processing steps in a typical ACA system

- **learning objectives**
  - list goals and applications in ACA
  - discuss typical forms of content in an audio signal
  - describe the typical signal flow in an ACA system

introduction

overview

Georgia Tech Center for Music Technology
College of Design

corresponding textbook section

Chapter 1 — Introduction: pp. 1–6

- **lecture content**
  - audio content analysis
  - typical applications
  - audio content
  - processing steps in a typical ACA system

- **learning objectives**
  - list goals and applications in ACA
  - discuss typical forms of content in an audio signal
  - describe the typical signal flow in an ACA system

introduction

audio content analysis — terminology

Georgia Tech Center for Music Technology College of Design

- **goal**
  - analyze audio signal to extract information on musical content

- **terminology**
  - *music information retrieval (MIR)*:
    - analysis and retrieval of music data
    - includes both audio and symbolic data
  - machine listening & computer audition
    - focus on the recognition and understanding of music
  - computational auditory scene analysis (CASA)
    - focus on human perception & cognition, understanding of the auditory scene

introduction

audio content analysis — terminology

Georgia Tech Center for Music Technology
College of Design

- **goal**
  - analyze audio signal to extract information on musical content
- **terminology**
  - *music information retrieval (MIR)*:
    - analysis and retrieval of music data
    - includes both audio and symbolic data
  - machine listening & computer audition
    - focus on the recognition and understanding of music
  - computational auditory scene analysis (CASA)
    - focus on human perception & cognition, understanding of the auditory scene

introduction

audio content analysis — research field

Georgia Tech Center for Music Technology College of Design

- **interdisciplinary**
  - digital signal processing
  - machine learning / data mining
  - musicology
  - psycho-acoustics
  - ...
- ISMIR **community**
  - annual conferences
  - conference papers & Transactions
  - ISMIR-Community mailing list
  - MIREX: MIR Evaluation eXchange
- **related publication outlets**
  - *conferences*: ISMIR, ICASSP, ICME, SMC, DAFx, ACM MM, ...
  - *journals*: TISMIR, TASLP, Computer Music, JNMR, JAES, ...

ISMIR

www.ismir.net

introduction

audio content analysis — research field

Georgia Tech Center for Music Technology College of Design

- **interdisciplinary**
  - digital signal processing
  - machine learning / data mining
  - musicology
  - psycho-acoustics
  - ...
- ISMIR **community**
  - annual conferences
  - conference papers & Transactions
  - ISMIR-Community mailing list
  - MIREX: MIR Evaluation eXchange
- **related publication outlets**
  - *conferences*: ISMIR, ICASSP, ICME, SMC, DAFx, ACM MM, ...
  - *journals*: TISMIR, TASLP, Computer Music, JNMR, JAES, ...

ISMIR

www.ismir.net

introduction

audio content analysis — research field

Georgia Tech Center for Music Technology
College of Design

- **interdisciplinary**
  - digital signal processing
  - machine learning / data mining
  - musicology
  - psycho-acoustics
  - ...
- ISMIR **community**
  - annual conferences
  - conference papers & Transactions
  - ISMIR-Community mailing list
  - MIREX: MIR Evaluation eXchange
- **related publication outlets**
  - *conferences*: ISMIR, ICASSP, ICME, SMC, DAFx, ACM MM, ...
  - *journals*: TISMIR, TASLP, Computer Music, JNMR, JAES, ...

ISMIR

www.ismir.net

introduction

applications

Georgia Tech Center for Music Technology
College of Design

- **organization** in large databases
  - search & retrieval, classification, similarity
- **interfaces** to search and retrieval
  - fingerprinting, query-by-humming systems
- music **visualization**
  - symbolic (bars, harmony, score, . . .), similarity mappings
- adaptive **processing**
  - adaptive effect parametrization or algorithm selection
- adaptive **interaction**
  - playlist generation, recommendation

introduction

applications

Georgia Tech Center for Music Technology
College of Design

- **organization** in large databases
  - search & retrieval, classification, similarity
- **interfaces** to search and retrieval
  - fingerprinting, query-by-humming systems
- music **visualization**
  - symbolic (bars, harmony, score, . . .), similarity mappings
- adaptive **processing**
  - adaptive effect parametrization or algorithm selection
- adaptive **interaction**
  - playlist generation, recommendation

introduction

applications

Georgia Tech Center for Music Technology
College of Design

- **organization** in large databases
  - search & retrieval, classification, similarity
- **interfaces** to search and retrieval
  - fingerprinting, query-by-humming systems
- music **visualization**
  - symbolic (bars, harmony, score, . . .), similarity mappings
- adaptive **processing**
  - adaptive effect parametrization or algorithm selection
- adaptive **interaction**
  - playlist generation, recommendation

introduction

applications

Georgia Tech Center for Music Technology
College of Design

- **organization** in large databases
  - search & retrieval, classification, similarity
- **interfaces** to search and retrieval
  - fingerprinting, query-by-humming systems
- music **visualization**
  - symbolic (bars, harmony, score, . . .), similarity mappings
- adaptive **processing**
  - adaptive effect parametrization or algorithm selection
- adaptive **interaction**
  - playlist generation, recommendation

introduction

applications

Georgia Tech Center for Music Technology
College of Design

- **organization** in large databases
  - search & retrieval, classification, similarity
- **interfaces** to search and retrieval
  - fingerprinting, query-by-humming systems
- music **visualization**
  - symbolic (bars, harmony, score, . . .), similarity mappings
- adaptive **processing**
  - adaptive effect parametrization or algorithm selection
- adaptive **interaction**
  - playlist generation, recommendation

introduction
(commercial) examples

Georgia Tech Center for Music Technology
College of Design

- **recommendation**, playlist generation

- **fingerprinting**

- **score following**

- (multi-) **pitch detection**

audio content

sources

Georgia Tech Center for Music Technology
College of Design

what are the sources of (musical) audio content?

audio content

sources

Georgia Tech Center for Music Technology College of Design

what are the sources of (musical) audio content?

1. **score**:
   - definition of musical ideas
   - "blue-print" of the music
   - *examples*: melody, key, harmony, rhythmic patterns, ...

audio content

sources

Georgia Tech Center for Music Technology
College of Design

what are the sources of (musical) audio content?

1. **score**:
   - definition of musical ideas
   - "blue-print" of the music
   - *examples*: melody, key, harmony, rhythmic patterns, ...
2. **performance**:
   - unique acoustic rendition
   - information in the score is interpreted, modified, added to
   - *examples*: (micro-)tempo, dynamics, intonation, ...

audio content

sources

Georgia Tech Center for Music Technology
College of Design

what are the sources of (musical) audio content?

1. **score**:
   - definition of musical ideas
   - "blue-print" of the music
   - *examples*: melody, key, harmony, rhythmic patterns, ...
2. **performance**:
   - unique acoustic rendition
   - information in the score is interpreted, modified, added to
   - *examples*: (micro-)tempo, dynamics, intonation, ...
3. **production**:
   - aesthetic choices
   - editing & processing
   - *examples*: sound quality (EQ, microphone positioning), changes in timing and pitch

audio content

technical categories

Georgia Tech Center for Music Technology
College of Design

audio content can be structured into **5 technical basic categories:**

1. **timbral**: related to sound quality
   - *examples*: instrument(ation), playing technique, venue, audio processing, ...
2. **intensity-related**: related to musical dynamics
   - *examples*: accents, loudness, ...
3. **tonal**: related to pitch
   - *examples*: melody, chords, intonation, vibrato, ...
4. **temporal**: related to rhythm and tempo
   - *examples*: timing, meter, rhythmic patterns, ...
5. **statistical & technical**: related to signal properties
   - *examples*: amplitude distribution, number of zero crossings, ...

audio content

technical categories

Georgia Tech Center for Music Technology
College of Design

audio content can be structured into **5 technical basic categories:**

1. **timbral**: related to sound quality
   - *examples*: instrument(ation), playing technique, venue, audio processing, ...
2. **intensity-related**: related to musical dynamics
   - *examples*: accents, loudness, ...
3. **tonal**: related to pitch
   - *examples*: melody, chords, intonation, vibrato, ...
4. **temporal**: related to rhythm and tempo
   - *examples*: timing, meter, rhythmic patterns, ...
5. **statistical & technical**: related to signal properties
   - *examples*: amplitude distribution, number of zero crossings, ...

audio content
technical categories

Georgia Tech Center for Music Technology
College of Design

audio content can be structured into **5 technical basic categories:**

1. **timbral**: related to sound quality
   - *examples*: instrument(ation), playing technique, venue, audio processing, ...
2. **intensity-related**: related to musical dynamics
   - *examples*: accents, loudness, ...
3. **tonal**: related to pitch
   - *examples*: melody, chords, intonation, vibrato, ...
4. **temporal**: related to rhythm and tempo
   - *examples*: timing, meter, rhythmic patterns, ...
5. **statistical & technical**: related to signal properties
   - *examples*: amplitude distribution, number of zero crossings, ...

audio content

technical categories

Georgia Tech Center for Music Technology
College of Design

audio content can be structured into **5 technical basic categories:**

1. **timbral**: related to sound quality
   - *examples*: instrument(ation), playing technique, venue, audio processing, ...
2. **intensity-related**: related to musical dynamics
   - *examples*: accents, loudness, ...
3. **tonal**: related to pitch
   - *examples*: melody, chords, intonation, vibrato, ...
4. **temporal**: related to rhythm and tempo
   - *examples*: timing, meter, rhythmic patterns, ...
5. **statistical & technical**: related to signal properties
   - *examples*: amplitude distribution, number of zero crossings, ...

audio content

technical categories

Georgia Tech Center for Music Technology
College of Design

audio content can be structured into **5 technical basic categories:**

1. **timbral**: related to sound quality
   - *examples*: instrument(ation), playing technique, venue, audio processing, ...
2. **intensity-related**: related to musical dynamics
   - *examples*: accents, loudness, ...
3. **tonal**: related to pitch
   - *examples*: melody, chords, intonation, vibrato, ...
4. **temporal**: related to rhythm and tempo
   - *examples*: timing, meter, rhythmic patterns, ...
5. **statistical & technical**: related to signal properties
   - *examples*: amplitude distribution, number of zero crossings, ...

audio content

technical categories

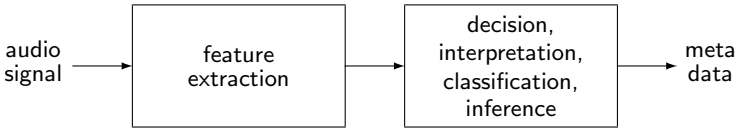
Georgia Tech Center for Music Technology
College of Design

audio content can be structured into **5 technical basic categories:**

1. **timbral**: related to sound quality
   - *examples*: instrument(ation), playing technique, venue, audio processing, ...
2. **intensity-related**: related to musical dynamics
   - *examples*: accents, loudness, ...
3. **tonal**: related to pitch
   - *examples*: melody, chords, intonation, vibrato, ...
4. **temporal**: related to rhythm and tempo
   - *examples*: timing, meter, rhythmic patterns, ...
5. **statistical & technical**: related to signal properties
   - *examples*: amplitude distribution, number of zero crossings, ...

audio content analysis

system overview



feature extraction

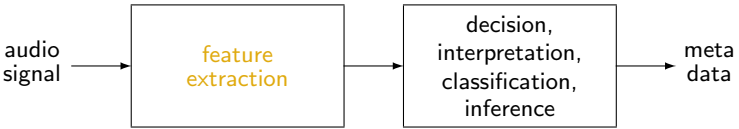
- dimensionality reduction
- meaningful representation

classification

- map or convert feature to comprehensible domain

audio content analysis

system overview



feature extraction

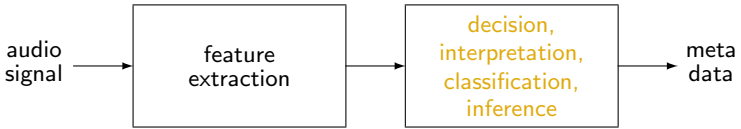
- dimensionality reduction
- meaningful representation

classification

- map or convert feature to comprehensible domain

audio content analysis

system overview



feature extraction

- dimensionality reduction
- meaningful representation

classification

- map or convert feature to comprehensible domain

summary

lecture content

Georgia Tech Center for Music Technology
College of Design

- music information retrieval and **audio content analysis**
  - aim at extracting data about the musical content from the music signal
- **audio content**
  - can relate to timbre, pitch, intensity, tempo and rhythm (but there is both lower level and higher level content)
  - is shaped by the musical ideas (score), the music performance, and the (studio) production
- the **flow chart of an ACA system** at its most fundamental level shows
  - a feature extraction step to extract meaningful descriptors
  - a classification or inference step to produce a "human" result