Introduction to Audio Content Analysis

Module 2.0: Audio Content Analysis Process

alexander lerch

introduction
overview

Georgia Tech | **Center for Music Technology**
College of Design

corresponding textbook section

Chapter 2

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

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

introduction
overview

Georgia | Center for Music
Tech | Technology
College of Design

corresponding textbook section

Chapter 2

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

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

audio content
sources

Georgia | Center for Music
Tech | Technology
College of Design

what are the sources of (musical) audio content?

audio content
sources

what are the sources of (musical) audio content?

1 **score/composition**:
- 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/composition**:
- 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 | **Center for Music**
Tech | **Technology**
College of Design

what are the sources of (musical) audio content?

1 **score/composition**:
- 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
categories

Georgia **Center for Music**
Tech **Technology**
College of Design

audio content can be structured into **4 basic categories:**

1 **tonal**: related to pitch
  - *examples*: melody, chords, intonation, vibrato, . . .

2 **timbral**: related to sound quality
  - *examples*: instrument(ation), playing technique, venue, audio processing, . . .

3 **intensity-related**: related to musical dynamics
  - *examples*: accents, loudness, . . .

4 **temporal**: related to rhythm and tempo
  - *examples*: timing, meter, rhythmic patterns, . . .

audio content
categories

audio content can be structured into **4 basic categories:**

1 tonal: related to pitch
  • *examples*: melody, chords, intonation, vibrato, . . .

2 timbral: related to sound quality
  • *examples*: instrument(ation), playing technique, venue, audio processing, . . .

3 intensity-related: related to musical dynamics
  • *examples*: accents, loudness, . . .

4 temporal: related to rhythm and tempo
  • *examples*: timing, meter, rhythmic patterns, . . .

audio content
categories

audio content can be structured into **4 basic categories:**

1 tonal: related to pitch
- *examples*: melody, chords, intonation, vibrato, . . .

2 timbral: related to sound quality
- *examples*: instrument(ation), playing technique, venue, audio processing, . . .

3 intensity-related: related to musical dynamics
- *examples*: accents, loudness, . . .

4 temporal: related to rhythm and tempo
- *examples*: timing, meter, rhythmic patterns, . . .

audio content
categories

audio content can be structured into **4 basic categories:**

1 **tonal**: related to pitch
  • *examples*: melody, chords, intonation, vibrato, . . .

2 **timbral**: related to sound quality
  • *examples*: instrument(ation), playing technique, venue, audio processing, . . .

3 **intensity-related**: related to musical dynamics
  • *examples*: accents, loudness, . . .

4 **temporal**: related to rhythm and tempo
  • *examples*: timing, meter, rhythmic patterns, . . .

audio content
categories

audio content can be structured into **4 basic categories:**

1 **tonal**: related to pitch
- *examples*: melody, chords, intonation, vibrato, . . .

2 **timbral**: related to sound quality
- *examples*: instrument(ation), playing technique, venue, audio processing, . . .

3 **intensity-related**: related to musical dynamics
- *examples*: accents, loudness, . . .

4 **temporal**: related to rhythm and tempo
- *examples*: timing, meter, rhythmic patterns, . . .

audio content
categories

audio content can be structured into **4 basic categories:**

1 tonal: related to pitch
  - *examples*: melody, chords, intonation, vibrato, . . .

2 timbral: related to sound quality
  - *examples*: instrument(ation), playing technique, venue, audio processing, . . .

3 intensity-related: related to musical dynamics
  - *examples*: accents, loudness, . . .

4 temporal: related to rhythm and tempo
  - *examples*: timing, meter, rhythmic patterns, . . .

other non-musical content descriptions: e.g., statistical, technical

audio content analysis
system overview
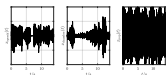
audio
signal → feature extraction
input representation → decision,
interpretation,
classification,
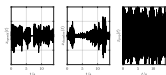inference → meta
data

feature representation

- compact and non-redundant
- task-relevant
- easy to analyze

classification

- map or convert feature to
comprehensible domain

audio content analysis
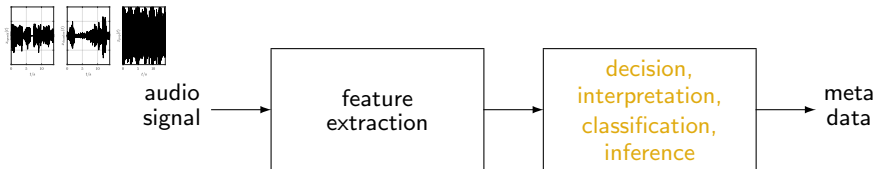## system overview

feature representation

- compact and non-redundant
- task-relevant
- easy to analyze

classification

- map or convert feature to comprehensible domain

audio content analysis
system overview

Georgia Tech | Center for Music Technology
College of Design

feature representation
- compact and non-redundant
- task-relevant
- easy to analyze

classification
- map or convert feature to comprehensible domain

summary
lecture content

Georgia Tech | Center for Music Technology
College of Design

- **audio content**
    - is shaped by the musical ideas (score), the music performance, and the (studio) production
    - can relate to timbre, pitch, intensity, tempo and rhythm (but there is both lower level and higher level content)

- 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