



# Introduction to **Audio Content Analysis**

Module 1.0: Introduction to MIR/ACA

alexander lerch

# introduction

## overview

### corresponding textbook section

#### chapter 1

#### ■ lecture content

- audio content analysis
- typical applications

#### ■ learning objectives

- list goals and applications in ACA
- understand the general development of the field
- differentiate various fields related to ACA



# introduction

## overview

### corresponding textbook section

#### chapter 1

#### ■ lecture content

- audio content analysis
- typical applications

#### ■ learning objectives

- list goals and applications in ACA
- understand the general development of the field
- differentiate various fields related to ACA



# introduction

## content in audio signals

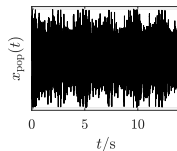
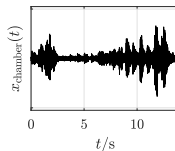
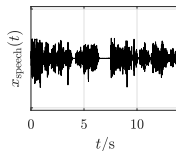
### examples for audio signal content

#### ■ speech

- text information
- speaker
- recording environment
- ...

#### ■ music

- melody
- harmony
- structure
- instruments
- mood
- genre
- ...



# introduction

## audio content analysis — goals

### Audio Content Analysis

The field of Audio Content Analysis (ACA) aims at designing and applying algorithms for the **automatic extraction of content information from the raw (digital) audio signal**.

This enables content-driven and content-adaptive services which describe, categorize, sort, retrieve, segment, process, and visualize the signal and its content.

# introduction

## audio content analysis — research fields

### ■ **speech** analysis

- speech recognition
- speech emotion recognition
- ...

### ■ **urban sound** analysis

- noise pollution monitoring
- audio surveillance
- ...

### ■ **industrial sound** analysis

- monitoring the state of mechanical devices (engines, etc.)
- monitoring the health of livestock
- ...

### ■ **musical audio** analysis

- music transcription
- music classification

# introduction

## audio content analysis — research fields

### ■ **speech** analysis

- speech recognition
- speech emotion recognition
- ...

### ■ **urban sound** analysis

- noise pollution monitoring
- audio surveillance
- ...

### ■ **industrial sound** analysis

- monitoring the state of mechanical devices (engines, etc.)
- monitoring the health of livestock
- ...

### ■ **musical audio** analysis

- music transcription
- music classification

# introduction

## audio content analysis — research fields

### ■ **speech** analysis

- speech recognition
- speech emotion recognition
- ...

### ■ **urban sound** analysis

- noise pollution monitoring
- audio surveillance
- ...

### ■ **industrial sound** analysis

- monitoring the state of mechanical devices (engines, etc.)
- monitoring the health of livestock
- ...

### ■ **musical audio** analysis

- music transcription
- music classification



# introduction

## audio content analysis — research fields

### ■ **speech** analysis

- speech recognition
- speech emotion recognition
- ...

### ■ **urban sound** analysis

- noise pollution monitoring
- audio surveillance
- ...

### ■ **industrial sound** analysis

- monitoring the state of mechanical devices (engines, etc.)
- monitoring the health of livestock
- ...

### ■ **musical audio** analysis

- music transcription
- music classification

# introduction

## audio content analysis — research fields

### ■ **speech** analysis

- speech recognition
- speech emotion recognition
- ...

### ■ **urban sound** analysis

- noise pollution monitoring
- audio surveillance
- ...

### ■ **industrial sound** analysis

- monitoring the state of mechanical devices (engines, etc.)
- monitoring the health of livestock
- ...

### ■ **musical audio** analysis

- music transcription
- music classification

# introduction

## musical audio vs. other audio

### music ...

- is a **wide band** signal  
unlike many other audio signals
- comprises both **tonal and noise** components  
like most audio signals
- combines **multiple sound sources**  
unlike speech, like urban sound
- is a **poly-timbral** mixture  
unlike industrial sound
- sources are **harmonically related and synchronous**  
unlike other multi-source signals
- has a highly structured language that is **abstract**  
unlike speech

# introduction

## musical audio vs. other audio

### music ...

- is a **wide band** signal  
unlike many other audio signals
- comprises both **tonal and noise** components  
like most audio signals
- combines **multiple sound sources**  
unlike speech, like urban sound
- is a **poly-timbral** mixture  
unlike industrial sound
- sources are **harmonically related and synchronous**  
unlike other multi-source signals
- has a highly structured language that is **abstract**  
unlike speech

# introduction

## musical audio vs. other audio

### music ...

- is a **wide band** signal  
unlike many other audio signals
- comprises both **tonal and noise** components  
like most audio signals
- combines **multiple sound sources**  
unlike speech, like urban sound
- is a **poly-timbral** mixture  
unlike industrial sound
- sources are **harmonically related and synchronous**  
unlike other multi-source signals
- has a highly structured language that is **abstract**  
unlike speech

# introduction

## musical audio vs. other audio

### music ...

- is a **wide band** signal  
unlike many other audio signals
- comprises both **tonal and noise** components  
like most audio signals
- combines **multiple sound sources**  
unlike speech, like urban sound
- is a **poly-timbral** mixture  
unlike industrial sound
- sources are **harmonically related and synchronous**  
unlike other multi-source signals
- has a highly structured language that is **abstract**  
unlike speech

# introduction

## musical audio vs. other audio

### music ...

- is a **wide band** signal  
unlike many other audio signals
- comprises both **tonal and noise** components  
like most audio signals
- combines **multiple sound sources**  
unlike speech, like urban sound
- is a **poly-timbral** mixture  
unlike industrial sound
- sources are **harmonically related and synchronous**  
unlike other multi-source signals
- has a highly structured language that is **abstract**  
unlike speech

# introduction

## musical audio vs. other audio

### music ...

- is a **wide band** signal  
unlike many other audio signals
- comprises both **tonal and noise** components  
like most audio signals
- combines **multiple sound sources**  
unlike speech, like urban sound
- is a **poly-timbral** mixture  
unlike industrial sound
- sources are **harmonically related and synchronous**  
unlike other multi-source signals
- has a highly structured language that is **abstract**  
unlike speech



# introduction

## audio content analysis — related terms and areas

### ■ terminology

- *Music Informatics*
  - ▶ overarching term for nearly everything with music and computers
- *Music Information Retrieval (MIR)*:
  - ▶ analysis and retrieval of music data
  - ▶ includes audio, symbolic, and other data
  - ▶ might also cover other tasks (source separation, generation)
- *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

### ■ interdisciplinary

- digital signal processing
- machine learning / data mining
- musicology
- music psychology
- ...

### ■ 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

# introduction

## audio content analysis — research field

### ■ interdisciplinary

- digital signal processing
- machine learning / data mining
- musicology
- music psychology
- ...

### ■ 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

# introduction

## audio content analysis — research field

### ■ interdisciplinary

- digital signal processing
- machine learning / data mining
- musicology
- music psychology
- ...

### ■ 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

# introduction

## audio content analysis — history

### ■ historic

- mechanical devices

### ■ expert systems

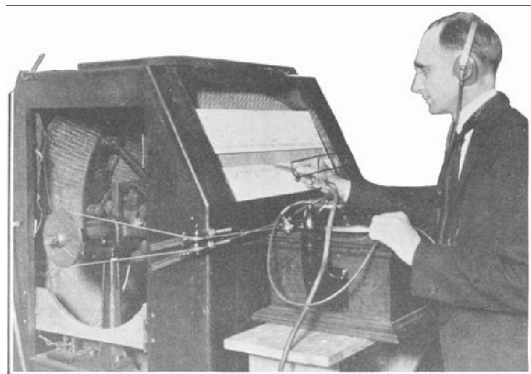
- rule-based approaches

### ■ data-driven **traditional ML systems**

- feature design plus ML engine

### ■ deep neural networks

- role of expert knowledge diminishes



The tonoscope for analyzing the pitch of the tones on a disk phonograph record

# introduction

## audio content analysis — history

### ■ historic

- mechanical devices

### ■ expert systems

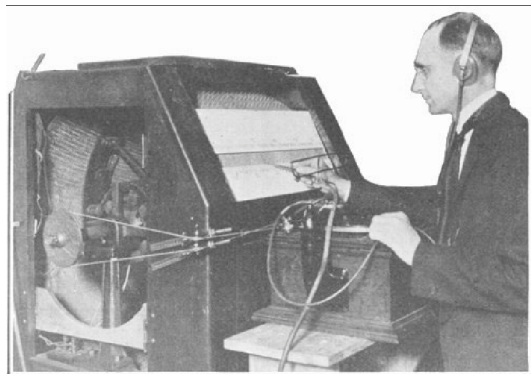
- rule-based approaches

### ■ data-driven **traditional ML systems**

- feature design plus ML engine

### ■ deep neural networks

- role of expert knowledge diminishes



The tonoscope for analyzing the pitch of the tones on a disk phonograph record

# introduction

## audio content analysis — history

### ■ historic

- mechanical devices

### ■ expert systems

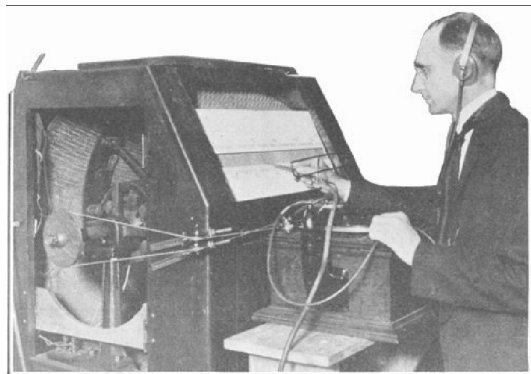
- rule-based approaches

### ■ data-driven **traditional ML systems**

- feature design plus ML engine

### ■ deep neural networks

- role of expert knowledge diminishes



The tonoscope for analyzing the pitch of the tones on a disk phonograph record

# introduction

## audio content analysis — history

### ■ historic

- mechanical devices

### ■ expert systems

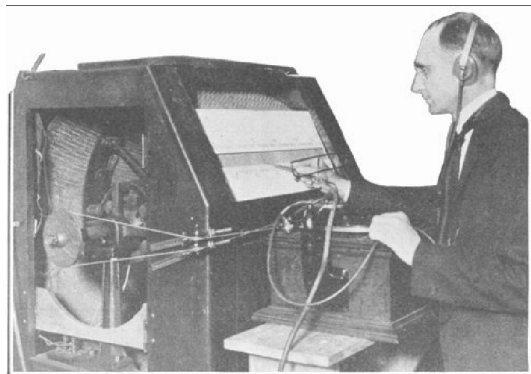
- rule-based approaches

### ■ data-driven **traditional ML systems**

- feature design plus ML engine

### ■ deep neural networks

- role of expert knowledge diminishes



The tonoscope for analyzing the pitch of the tones on a disk phonograph record



# introduction

## applications

### ■ music browsing and music discovery

- search & retrieval, similarity, interfaces (e.g., QBH)

### ■ music consumption

- creative music listening

### ■ music production

- adaptive parametrization, enhancements of creative process

### ■ music education

- musically intelligent software tutoring

### ■ generative music

- interactive soundtracks (games, video)

# introduction

## applications

### ■ music browsing and music discovery

- search & retrieval, similarity, interfaces (e.g., QBH)

### ■ music consumption

- creative music listening

### ■ music production

- adaptive parametrization, enhancements of creative process

### ■ music education

- musically intelligent software tutoring

### ■ generative music

- interactive soundtracks (games, video)

# introduction

## applications

- **music browsing and music discovery**
  - search & retrieval, similarity, interfaces (e.g., QBH)
- **music consumption**
  - creative music listening
- **music production**
  - adaptive parametrization, enhancements of creative process
- **music education**
  - musically intelligent software tutoring
- **generative music**
  - interactive soundtracks (games, video)

# introduction

## applications

- **music browsing and music discovery**
  - search & retrieval, similarity, interfaces (e.g., QBH)
- **music consumption**
  - creative music listening
- **music production**
  - adaptive parametrization, enhancements of creative process
- **music education**
  - musically intelligent software tutoring
- **generative music**
  - interactive soundtracks (games, video)

# introduction

## applications

- **music browsing and music discovery**
  - search & retrieval, similarity, interfaces (e.g., QBH)
- **music consumption**
  - creative music listening
- **music production**
  - adaptive parametrization, enhancements of creative process
- **music education**
  - musically intelligent software tutoring
- **generative music**
  - interactive soundtracks (games, video)

# introduction

## (commercial) application examples

### ■ recommendation, playlist generation



lost.fm

PANDORA

### ■ fingerprinting



### ■ score following



### ■ (multi-) pitch detection

melodyne



# summary

## lecture content

### ■ audio content analysis

- aims at extracting data about the (musical) content from the music signal
- interdisciplinary field

### ■ related areas

- music information retrieval, machine listening, ...

### ■ applications cover music ...

- discovery, consumption, production, education, generation

