# Introduction to **Audio Content Analysis**

module 12.2: musical genre classification

alexander lerch

Georgia Tech Center for Music Technology
College of Design

# introduction
## overview

### corresponding textbook section

section 12.2

- **lecture content**
  - musical genre
  - processing steps in basic genre classifiers
  - example: genre classification with a kNN

- **learning objectives**
  - discuss ambiguities in the definition of musical genre and the possible impact on automatic systems
  - describe the processing steps for traditional musical genre classifiers
  - implement your own music genre classifier with Matlab

## corresponding textbook section

section 12.2

- **lecture content**
  - musical genre
  - processing steps in basic genre classifiers
  - example: genre classification with a kNN

- **learning objectives**
  - discuss ambiguities in the definition of musical genre and the possible impact on automatic systems
  - describe the processing steps for traditional musical genre classifiers
  - implement your own music genre classifier with Matlab

# musical genre classification

## introduction



- one of the early/**seminal research topics** in MIR
- classic *machine learning* task
  - features → classification
- **related tasks**:
  - speech-music classification
  - instrument recognition
  - artist identification
  - music emotion recognition

# musical genre classification
## introduction

- one of the early/**seminal research topics** in MIR
- classic *machine learning* task
  - features $\rightarrow$ classification
- **related tasks**:
  - speech-music classification
  - instrument recognition
  - artist identification
  - music emotion recognition

# musical genre classification
## introduction

- one of the early/**seminal research topics** in MIR
- classic *machine learning* task
  - features $\rightarrow$ classification
- **related tasks**:
  - speech-music classification
  - instrument recognition
  - artist identification
  - music emotion recognition

# musical genre classification

## applications

- large music databases:
  - annotation
  - sorting, browsing, retrieving

- recommendation and music discovery systems
- automatic playlist generation
- improving downstream MIR tasks by using side information/conditioning

# musical genre classification

## applications

- large music databases:
  - annotation
  - sorting, browsing, retrieving
- recommendation and music discovery systems
- automatic playlist generation
- improving downstream MIR tasks by using side information/conditioning

# musical genre classification
## genre: definition

**what is musical genre**

# musical genre classification

## genre: definition



**what is musical genre**

- clusters of musical similarity?

→ hard to answer in general, there are many **systematic problems**

# musical genre classification

## genre: definition



### what is musical genre

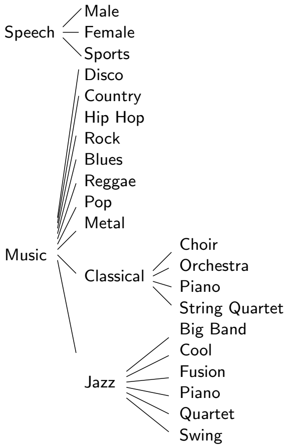
- clusters of musical similarity?

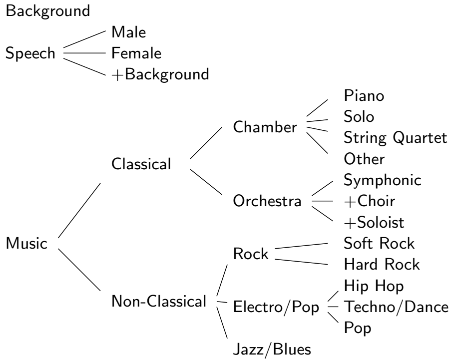
→ hard to answer in general, there are many **systematic problems**

1. **non-agreement on taxonomies**
   - e.g., AllMusic vs. Pandora
2. **genre label scope**
   - e.g., song, album, artist, piece of a song
3. **ill-defined genre labels**
   - e.g., geographic (*indian music*), historic (*baroque*), technical (*barbershop*), instrumentation (*symphonic music*), usage (*christmas songs*)
4. **taxonomy scalability**
   - e.g., genres and subgenres evolve over time
5. **non-orthogonality**
   - e.g., several genres for one piece of music

# musical genre classification

## genre: taxonomy examples



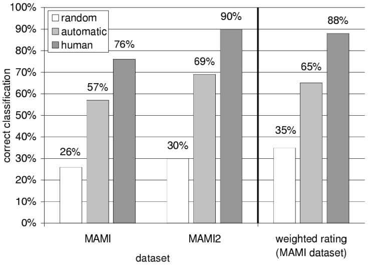



# musical genre classification

## observations with humans

1. human classification far from perfect: 75–90 % for limited set of classes
2. for many genres, humans need only a fraction of a second to classify

⇒ short time timbre features sufficient?




plots from[1],[2]

[1] S. Lippens, J.-P. Martens, T. D. Mulder, *et al.*, "A Comparison of Human and Automatic Musical Genre Classification," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, 2004.

[2] R. O. Gjerdingen and D. Perrott, "Scanning the Dial: The Rapid Recognition of Music Genres," *Journal of New Music Research*, vol. 37, no. 2, pp. 93–100, Jun. 2008, 00067, ISSN: 0929-8215.

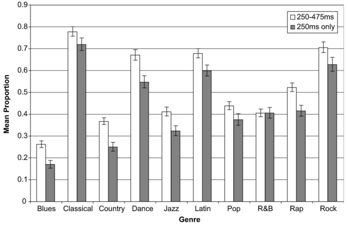
# musical genre classification

## observations with humans



1. human classification far from perfect: 75–90 % for limited set of classes
2. for many genres, humans need only a fraction of a second to classify

⇒ short time timbre features sufficient?



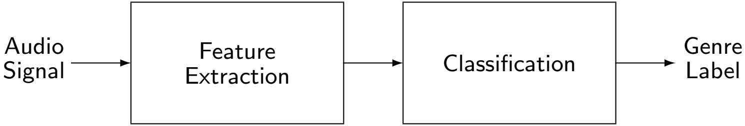

plots from[1],[2]

[1] S. Lippens, J.-P. Martens, T. D. Mulder, *et al.*, "A Comparison of Human and Automatic Musical Genre Classification," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, 2004.

[2] R. O. Gjerdingen and D. Perrott, "Scanning the Dial: The Rapid Recognition of Music Genres," *Journal of New Music Research*, vol. 37, no. 2, pp. 93–100, Jun. 2008, 00067, ISSN: 0929-8215.

# musical genre classification
## overview



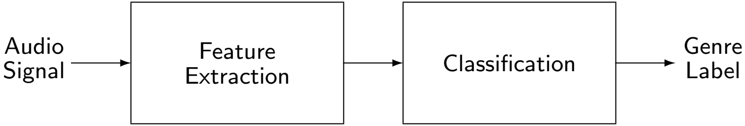



1. **feature extraction**
   - compressed, meaningful representation
2. **classification**
   - map or convert feature to comprehensible domain

# musical genre classification

## overview






1. **feature extraction**
   - compressed, meaningful representation
2. **classification**
   - map or convert feature to comprehensible domain

# musical genre classification

## feature categories

- **high level similarities**?
  - melody, hook lines, bass lines, harmony progression
  - rhythm & tempo
  - structure
  - instrumentation & timbre
- **technical feature categories**
  - tonal
  - technical
  - timbral
  - temporal
  - intensity
- **extracted features should be**
  - extractable (not: time envelope in polyphonic signals)
  - relevant (not: pitch chroma for instrument ID)
  - non-redundant
  - have discriminative power

# musical genre classification
## feature categories

- **high level similarities**?
  - melody, hook lines, bass lines, harmony progression
  - rhythm & tempo
  - structure
  - instrumentation & timbre
- **technical feature categories**
  - tonal
  - technical
  - timbral
  - temporal
  - intensity
- **extracted features should be**
  - extractable (not: time envelope in polyphonic signals)
  - relevant (not: pitch chroma for instrument ID)
  - non-redundant
  - have discriminative power

# musical genre classification

## feature categories

- **high level similarities?**
  - melody, hook lines, bass lines, harmony progression
  - rhythm & tempo
  - structure
  - instrumentation & timbre
- **technical feature categories**
  - tonal
  - technical
  - timbral
  - temporal
  - intensity
- **extracted features should be**
  - extractable (not: time envelope in polyphonic signals)
  - relevant (not: pitch chroma for instrument ID)
  - non-redundant
  - have discriminative power

# musical genre classification
## instantaneous features

- spectral features (**timbre**):
  Spectral Centroid, MFCCs, Spectral Flux, ...
- pitch features (**tonal**):
  pitch chroma distribution/change, ...
- rhythm features (**temporal**):
  onset density, beat histogram features, ...
- statistical features (**technical**):
  standard deviation, skewness, zero crossings, ...
- **intensity** features:
  level variation, number of "pauses", ...

# musical genre classification
## instantaneous features



- spectral features (**timbre**):
  Spectral Centroid, MFCCs, Spectral Flux, ...
- pitch features (**tonal**):
  pitch chroma distribution/change, ...
- rhythm features (**temporal**):
  onset density, beat histogram features, ...
- statistical features (**technical**):
  standard deviation, skewness, zero crossings, ...
- **intensity** features:
  level variation, number of "pauses", ...

# musical genre classification
## instantaneous features

- spectral features (**timbre**):
  Spectral Centroid, MFCCs, Spectral Flux, . . .
- pitch features (**tonal**):
  pitch chroma distribution/change, . . .
- rhythm features (**temporal**):
  onset density, beat histogram features, . . .
- statistical features (**technical**):
  standard deviation, skewness, zero crossings, . . .
- **intensity** features:
  level variation, number of “pauses”, . . .

# musical genre classification

## instantaneous features

- spectral features (**timbre**):
  Spectral Centroid, MFCCs, Spectral Flux, ...
- pitch features (**tonal**):
  pitch chroma distribution/change, ...
- rhythm features (**temporal**):
  onset density, beat histogram features, ...
- statistical features (**technical**):
  standard deviation, skewness, zero crossings, ...
- **intensity** features:
  level variation, number of "pauses", ...

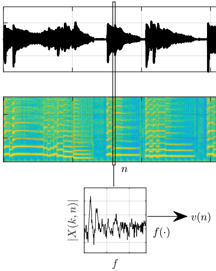
# musical genre classification
## instantaneous features

- spectral features (**timbre**):
  Spectral Centroid, MFCCs, Spectral Flux, ...
- pitch features (**tonal**):
  pitch chroma distribution/change, ...
- rhythm features (**temporal**):
  onset density, beat histogram features, ...
- statistical features (**technical**):
  standard deviation, skewness, zero crossings, ...
- **intensity** features:
  level variation, number of "pauses", ...

# musical genre classification
## feature extraction process



1. extract **instantaneous features**




2. compute **derived features** (derivatives etc.)
3. compute **long term features** & subfeatures per texture window or file
4. **normalize** subfeatures
5. (select or) **transform** subfeatures
6. feature vector → **classifier input**

# musical genre classification

## feature extraction process

1. extract **instantaneous features**
2. compute **derived features** (derivatives etc.)
3. compute **long term features** & subfeatures per texture window or file
4. **normalize** subfeatures
5. (select or) **transform** subfeatures
6. feature vector $\rightarrow$ **classifier input**

# musical genre classification

## feature extraction process

1. extract **instantaneous features**
2. compute **derived features** (derivatives etc.)
3. compute **long term features** & subfeatures per texture window or file
4. **normalize** subfeatures
5. (select or) **transform** subfeatures
6. feature vector $\longrightarrow$ **classifier input**

# musical genre classification

## feature extraction process

1. extract **instantaneous features**
2. compute **derived features** (derivatives etc.)
3. compute **long term features** & subfeatures per texture window or file
4. **normalize** subfeatures
5. (select or) **transform** subfeatures
6. feature vector → **classifier input**

# musical genre classification

## feature extraction process

1. extract **instantaneous features**
2. compute **derived features** (derivatives etc.)
3. compute **long term features** & subfeatures per texture window or file
4. **normalize** subfeatures
5. (select or) **transform** subfeatures
6. feature vector $\rightarrow$ **classifier input**

# musical genre classification

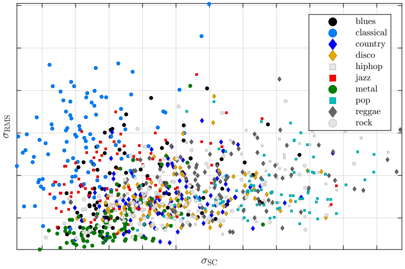
## feature extraction process

1. extract **instantaneous features**
2. compute **derived features** (derivatives etc.)
3. compute **long term features** & subfeatures per texture window or file
4. **normalize** subfeatures
5. (select or) **transform** subfeatures
6. feature vector → **classifier input**

# musical genre classification
## feature extraction process



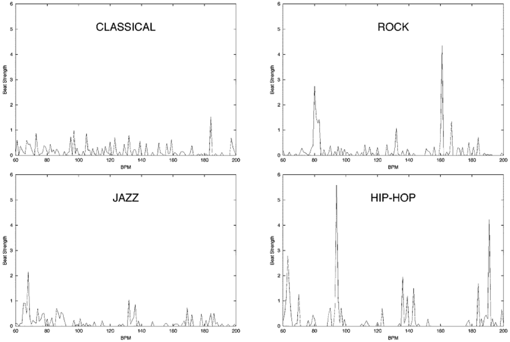
1. extract **instantaneous features**
2. compute **derived features** (derivatives etc.)
3. compute **long term features** & subfeatures per texture window or file
4. **normalize** subfeatures
5. (select or) **transform** subfeatures
6. feature vector → **classifier input**



# musical genre classification
## long term features 1/2

derived from beat histogram[3]

- statistical histogram features
- number and values of top maxima
- location (relation) of top maxima
- ...




[3] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, Jul. 2002, ISSN: 1063-6676. DOI: 10.1109/TSA.2002.800560.

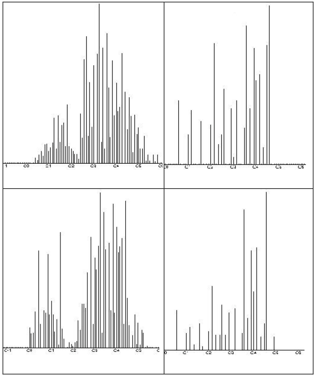
# musical genre classification

## long term features 2/2

Georgia Tech | Center for Music Technology College of Design

derived from pitch histogram or pitch chroma[4]

- statistical histogram features
- number and values of top maxima
- location (relation) of top maxima
- ...



[4] G. Tzanetakis, A. Ermolinskyi, and P. Cook, "Pitch Histograms in Audio and Symbolic Music Information Retrieval," in *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR)*, Paris, 2002.

# musical genre classification
## additional possible features

- **stereo features**
  - mid channel energy vs. side channel energy
  - spectral channel differences

- features at **higher semantic levels**:
  - tempo, structure, harmonic complexity, instrumentation

# musical genre classification
## additional possible features



- **stereo features**
  - mid channel energy vs. side channel energy
  - spectral channel differences

- features at **higher semantic levels**:
  - tempo, structure, harmonic complexity, instrumentation

# musical genre classification

## results

- classification results depend on training set, test set, and number of classes
- typical range: $\approx 10$ classes $\Rightarrow$ 50–80%

- main challenges
  - ill-defined genre boundaries
  - non-uniformly distributed classes
  - overfitting through songs from same album or artist
  - ...

# musical genre classification

## results

- classification results depend on training set, test set, and number of classes
- typical range: $\approx$ 10 classes $\Rightarrow$ 50–80%
- main challenges
    - ill-defined genre boundaries
    - non-uniformly distributed classes
    - overfitting through songs from same album or artist
    - ...

# musical genre classification
## results



- classification results depend on training set, test set, and number of classes
- typical range: $\approx$ 10 classes $\Rightarrow$ 50–80%

- main challenges
  - ill-defined genre boundaries
  - non-uniformly distributed classes
  - overfitting through songs from same album or artist
  - ...

# musical genre classification

## speech/music classification baseline example

binary classification task

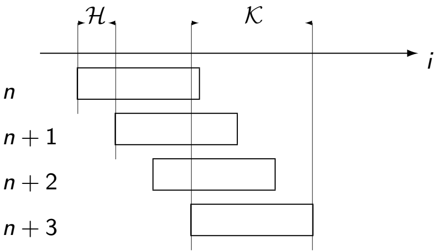
1. extract features
2. represent each file with its 2-dimensional feature vector
3. kNN to classify unknown audio files
4. evaluate classification performance

# musical genre classification

## speech/music classification example: features 1/2

for each audio file

1. split input signal into (overlapping) blocks



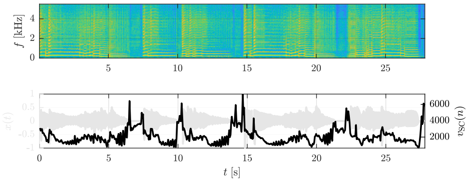

2. compute 2 feature series (spectral centroid, RMS)
3. aggregate feature series to one value per file
   - *mean* of Spectral Centroid $\mu_{\mathrm{SC}}$
   - *standard deviation* of RMS $\sigma_{\mathrm{RMS}}$
4. represent each file as 2-dimensional vector

# musical genre classification

## speech/music classification example: features 1/2

for each audio file

1. split input signal into (overlapping) blocks
2. compute 2 feature series (spectral centroid, RMS)




3. aggregate feature series to one value per file
    - *mean* of Spectral Centroid $\mu_{SC}$
    - *standard deviation* of RMS $\sigma_{RMS}$
4. represent each file as 2-dimensional vector

# musical genre classification

## speech/music classification example: features 1/2

for each audio file

1. split input signal into (overlapping) blocks
2. compute 2 feature series (spectral centroid, RMS)
3. aggregate feature series to one value per file
   - *mean* of Spectral Centroid $\mu_\mathrm{SC}$

$$\mu_\mathrm{SC} = \frac{1}{N}\sum_{\forall n} v_\mathrm{SC}(n)$$

   - *standard deviation* of RMS $\sigma_\mathrm{RMS}$

$$\sigma_\mathrm{RMS} = \sqrt{\frac{1}{N}\sum_{\forall n}\left(v_\mathrm{RMS}(n) - \mu_\mathrm{RMS}\right)^2}$$

4. represent each file as 2-dimensional vector

# musical genre classification

## speech/music classification example: features 1/2

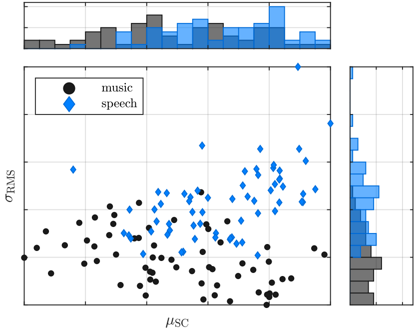
for each audio file

1. split input signal into (overlapping) blocks
2. compute 2 feature series (spectral centroid, RMS)
3. aggregate feature series to one value per file
   - *mean* of Spectral Centroid $\mu_{\mathrm{SC}}$
   - *standard deviation* of RMS $\sigma_{\mathrm{RMS}}$
4. represent each file as 2-dimensional vector

$$\left(\mu_{\mathrm{SC}}, \sigma_{\mathrm{RMS}}\right)^{\mathrm{T}}$$

# musical genre classification

## speech/music classification example: features 2/2




matlab source: plotFeaturespace.m

# musical genre classification

## speech/music classification example: training set

- use **dataset** annotated as speech and music:
  - requirements
    - large compared to number of features
    - representative for use case (diverse)
  - here (toy example):
    - 64 speech files
    - 64 music files
- extract the features for the dataset
  - centroid mean
  - rms std
- use 3NN classifier
- procedure: Leave-One-Out-Cross-Validation

# musical genre classification

## speech/music classification example: results (kNN)

- **confusion matrix**:

| | **speech** | **music** | # files |
|---|---|---|---|
| **gt speech** | **51** | 13 | 64 |
| **gt music** | 11 | **53** | 64 |

- $\Rightarrow$ **classification rate**:

$$\frac{53+54}{64+64} = 81.25\%$$

- single feature classification results
  - Spectral Centroid: 63.28%
  - RMS: 73.44%

matlab source: matlab/ExampleMusicSpeechClassification.m

# musical genre classification

## speech/music classification example: results (kNN)

- **confusion matrix**:

| | **speech** | **music** | # files |
|---|---|---|---|
| **gt speech** | **51** | 13 | 64 |
| **gt music** | 11 | **53** | 64 |

- $\Rightarrow$ **classification rate**:

$$\frac{53+54}{64+64} = 81.25\%$$

- single feature classification results
  - Spectral Centroid: 63.28%
  - RMS: 73.44%

matlab source: matlab/ExampleMusicSpeechClassification.m

# musical genre classification

## speech/music classification example: results (kNN)

- **confusion matrix**:

| | speech | music | # files |
|---|---|---|---|
| **gt speech** | **51** | 13 | 64 |
| **gt music** | 11 | **53** | 64 |

- ⇒ **classification rate**:

$$\frac{53 + 54}{64 + 64} = 81.25\%$$

- single feature classification results
  - Spectral Centroid: 63.28%
  - RMS: 73.44%

matlab source: matlab/ExampleMusicSpeechClassification.m

# summary

## lecture content



- **musical genre**
  - ill-defined, subjective, no general agreement
  - some human agreement
- **MGC: features**
  - from all possible categories as all categories might depend on genre
  - timbre seems most meaningful feature
- **MGC: classifier**
  - any classifier works, and most have been used
- **MGC: standard baseline**
  1. MFCCs
  2. SVM