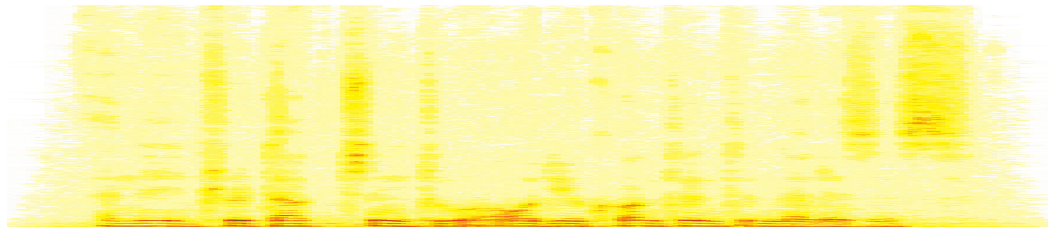


Introduction to Audio Content Analysis

Module 8.1: Musical Genre Classification

alexander lerch



introduction

overview

corresponding textbook section

Chapter 8: Musical Genre, Similarity, and Mood (pp. 151–155)

● 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



introduction

overview

corresponding textbook section

Chapter 8: Musical Genre, Similarity, and Mood (pp. 151–155)

● 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
- related fields:
 - 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
- related fields:
 - 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
- related fields:
 - speech-music classification
 - instrument recognition
 - artist identification
 - music emotion recognition

musical genre classification

applications

- large music databases:
 - annotation
 - sorting, browsing, retrieving
- recommendation systems
- automatic playlist generation
- mashup generation

musical genre classification

genre: definition

what is musical genre



musical genre classification

genre: definition

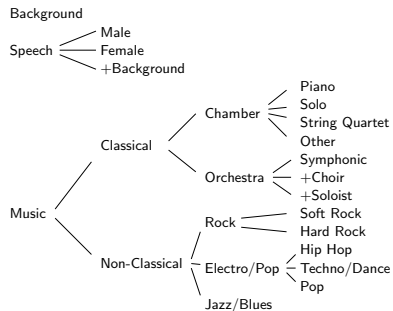
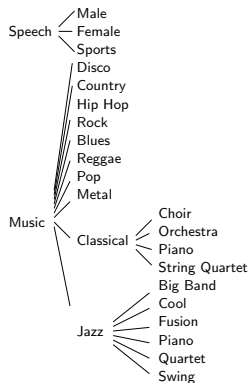
what is musical genre



- clusters of musical similarity?
- hard to answer in general, there are many **systematic problems**
 - 1 **non-agreement on taxonomies**
 - 2 **genre label scope**: song, album, artist, piece of a song
 - 3 **ill-defined genre labels**: geographic (*indian music*), historic (*baroque*), technical (*barbershop*), instrumentation (*symphonic music*), usage (*christmas songs*)
 - 4 **taxonomy scalability**: genres and subgenres evolve over time
 - 5 **non-orthogonality**: several genres for one piece of music

musical genre classification

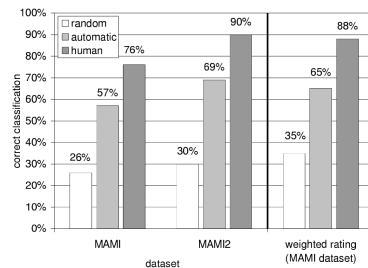
genre: taxonomy examples



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}

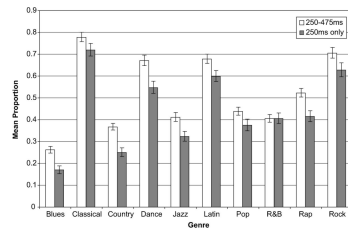
¹S. Lippens, J.-P. Martens, T. D. Mulder, *et al.*, "A Comparison of Human and Automatic Musical Genre Classification," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, 2004.

²R. 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?



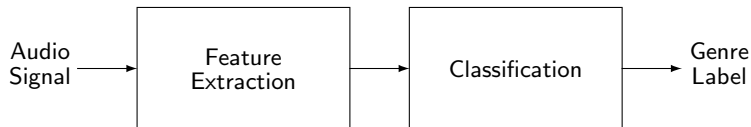
plots from^{1,2}

¹S. 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.

²R. 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

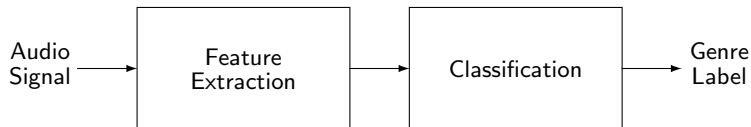
- dimensionality reduction
- meaningful representation

2 classification

- map or convert feature to comprehensible domain

musical genre classification

overview



1 feature extraction

- dimensionality reduction
- 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
- (robust to noise)

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
- (robust to noise)

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
 - (robust to noise)

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

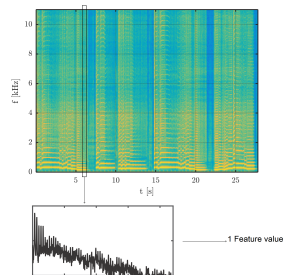
instantaneous features

- 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

1 extract **instantaneous features**



- 2 compute **derived features** (derivative, filtered)
- 3 compute **long term features** & subfeatures per texture window
- 4 compute **subfeatures** per file
- 5 **normalize** subfeatures
- 6 (select or) **transform** subfeatures
- 7 feature vector → **classifier input**

musical genre classification

feature extraction

- 1 extract **instantaneous features**
- 2 compute **derived features** (derivative, filtered)
- 3 compute **long term features** & subfeatures per texture window
- 4 compute **subfeatures** per file
- 5 **normalize** subfeatures
- 6 (select or) **transform** subfeatures
- 7 feature vector → **classifier input**

musical genre classification

feature extraction

- 1 extract **instantaneous features**
- 2 compute **derived features** (derivative, filtered)
- 3 compute **long term features** & subfeatures per texture window
- 4 compute **subfeatures** per file
- 5 **normalize** subfeatures
- 6 (select or) **transform** subfeatures
- 7 feature vector → **classifier input**

musical genre classification

feature extraction

- 1 extract **instantaneous features**
- 2 compute **derived features** (derivative, filtered)
- 3 compute **long term features** & subfeatures per texture window
- 4 compute **subfeatures** per file
- 5 **normalize** subfeatures
- 6 (select or) **transform** subfeatures
- 7 feature vector → **classifier input**

musical genre classification

feature extraction

- 1 extract **instantaneous features**
- 2 compute **derived features** (derivative, filtered)
- 3 compute **long term features** & subfeatures per texture window
- 4 compute **subfeatures** per file
- 5 **normalize** subfeatures
- 6 (select or) **transform** subfeatures
- 7 feature vector → **classifier input**

musical genre classification

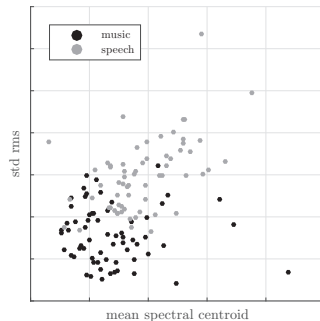
feature extraction

- 1 extract **instantaneous features**
- 2 compute **derived features** (derivative, filtered)
- 3 compute **long term features** & subfeatures per texture window
- 4 compute **subfeatures** per file
- 5 **normalize** subfeatures
- 6 (select or) **transform** subfeatures
- 7 feature vector → **classifier input**

musical genre classification

feature extraction

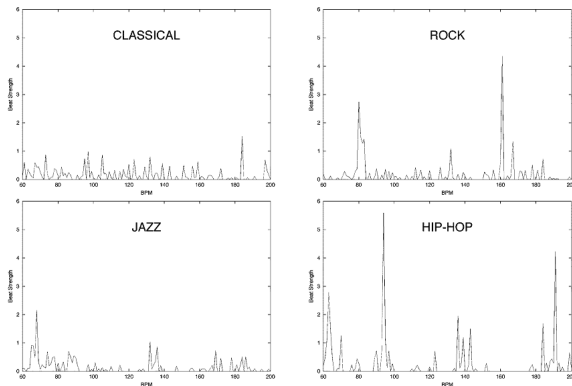
- 1 extract **instantaneous features**
- 2 compute **derived features** (derivative, filtered)
- 3 compute **long term features** & subfeatures per texture window
- 4 compute **subfeatures** per file
- 5 **normalize** subfeatures
- 6 (select or) **transform** subfeatures
- 7 feature vector → **classifier input**



musical genre classification

long term features 1/2

derived from beat histogram³

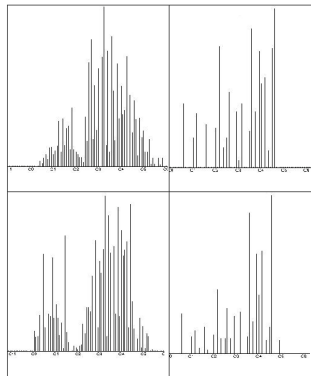


³G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, Jul. 2002, ISSN: 1063-6676. DOI: [10.1109/TSA.2002.800560](https://doi.org/10.1109/TSA.2002.800560).

musical genre classification

long term features 2/2

derived from pitch histogram or pitch chroma⁴



⁴G. 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 feature examples

- **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 feature examples

- **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 ranges: 10 classes \Rightarrow 50–80%
- note: results vary largely between datasets
 - 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 ranges: 10 classes \Rightarrow 50–80%
- note: results vary largely between datasets
 - 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 ranges: 10 classes \Rightarrow 50–80%
- note: results vary largely between datasets
 - ill-defined genre boundaries
 - non-uniformly distributed classes
 - overfitting through songs from same album or artist
 - ...

musical genre classification

speech/music classification baseline example

- 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 each
 - *mean* of Spectral Centroid

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

- *standard deviation* of RMS

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

- 4 represent each file as 2-dimensional vector

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

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 each
 - *mean* of Spectral Centroid

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

- *standard deviation* of RMS

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

- ④ represent each file as 2-dimensional vector

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

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 each
 - *mean* of Spectral Centroid

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

- *standard deviation* of RMS

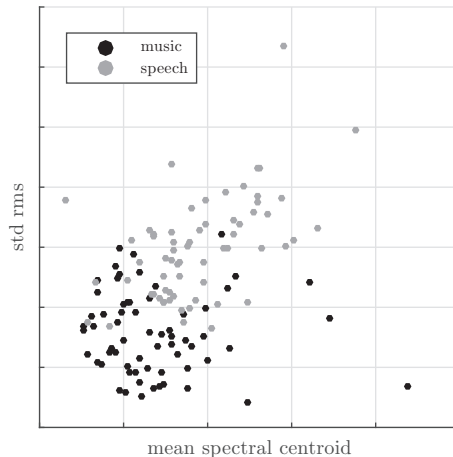
$$\sigma_{RMS} = \sqrt{\frac{1}{N} \sum_{\forall n} (v_{RMS}(n) - \mu_{RMS})^2}$$

- ④ represent each file as 2-dimensional vector

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

musical genre classification

speech/music classification example: features 2/2



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
- procedure: Leave-One-Out-Cross-Validation

musical genre classification

speech/music classification example: results (kNN)

- **confusion matrix:**

	speech	music	# files
speech	53	11	64
music	10	54	64

- ⇒ **classification rate:**

$$\frac{53 + 54}{64 + 64} = 83.59\%$$

- single feature classification results

- Spectral Centroid: 64.84%
- RMS: 76.56%

musical genre classification

speech/music classification example: results (kNN)

- **confusion matrix:**

	speech	music	# files
speech	53	11	64
music	10	54	64

- \Rightarrow **classification rate:**

$$\frac{53 + 54}{64 + 64} = 83.59\%$$

- single feature classification results

- Spectral Centroid: 64.84%
- RMS: 76.56%

musical genre classification

speech/music classification example: results (kNN)

- **confusion matrix:**

	speech	music	# files
speech	53	11	64
music	10	54	64

- \Rightarrow **classification rate:**

$$\frac{53 + 54}{64 + 64} = 83.59\%$$

- **single feature classification results**

- Spectral Centroid: 64.84%
- RMS: 76.56%

summary

lecture content

- **musical genre**
 - ill-defined, subjective, no general agreement
- **MGC: features**
 - ① from all possible categories as all categories might depend on genre
- **MGC: classifier**
 - ① any classifier works, and most have been used

