music genre recognition

karin kosina

kyrah@gnu.org

systems in motion

- music genre recognition by analysis of texture
- kyrah's masters thesis at fh hagenberg
- system for the automatic recognition of music genres, based only on the sound signal
- no meta-data, no DB lookup,. . .
  only based on sound properties
- question: how do you do it?

two steps:

- feature extraction
  calculate numerical representation of audio data
  choosing the right features is crucial!

- classification
  use output of feature extraction as basis for classification

- ▷ no toy data – use arbitrary polyphonic signals

- no toy data – use arbitrary polyphonic signals
- advanced listening is not important

- no toy data – use arbitrary polyphonic signals
- advanced listening is not important
- music theory is not important

[design principles]

- no toy data – use arbitrary polyphonic signals
- advanced listening is not important
- music theory is not important
- notes are not important

- no toy data – use arbitrary polyphonic signals
- advanced listening is not important
- music theory is not important
- notes are not important
- be aware of limitations of “perceptual” modeling

- no toy data – use arbitrary polyphonic signals
- advanced listening is not important
- music theory is not important
- notes are not important
- be aware of limitations of “perceptual” modeling
- describe genres in terms of typical members

- no toy data – use arbitrary polyphonic signals
- advanced listening is not important
- music theory is not important
- notes are not important
- be aware of limitations of “perceptual” modeling
- describe genres in terms of typical members
- use information available in sound signal

music files cannot be compared directly
instead: calculate feature representation, i.e.
essential information needed to differentiate classes

“feature vector” – point in n-dimensional space
classification based on distance

features used in mugrat:

- music texture features (short-time spectral change)
- beat-related features (rhythm and beatedness)

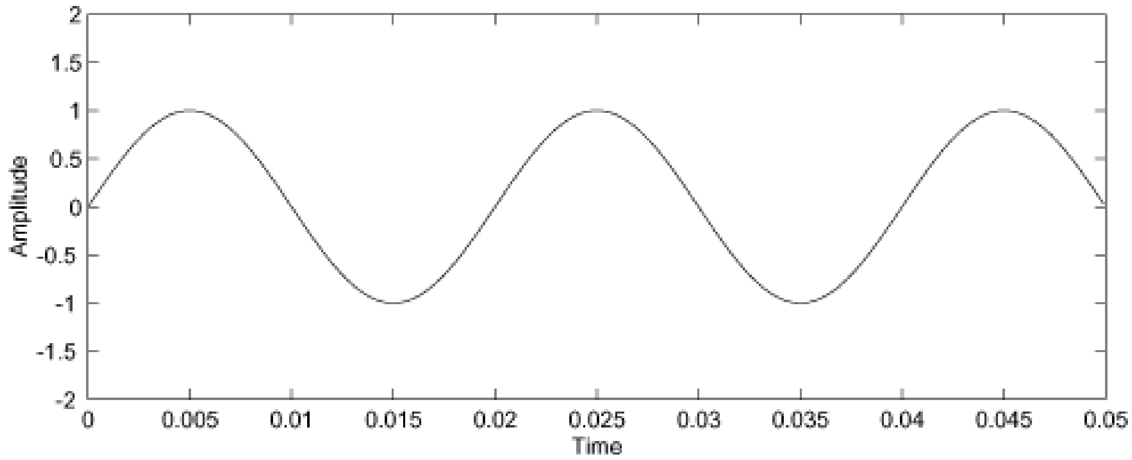
spectral attributes;
mean and variance to capture short-time spectral change

- spectral centroid
- rolloff
- flux
- zero-crossing rate

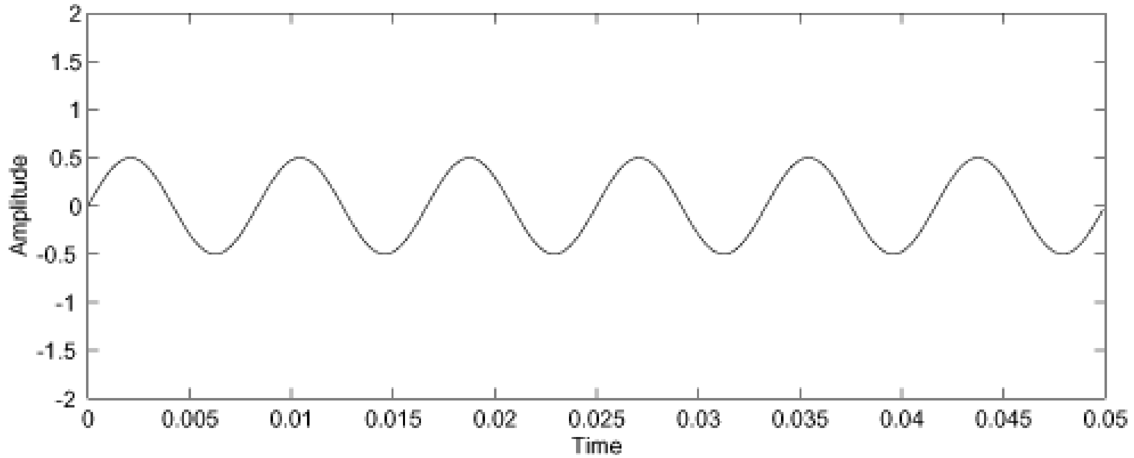
feature set based on:

George Tzanetakis, Georg Essl, and Perry Cook. Automatic Musical Genre Classification of Audio Signals. In: Proceedings International Symposium for Audio Information Retrieval (ISMIR), Princeton, N.J., October 2001.

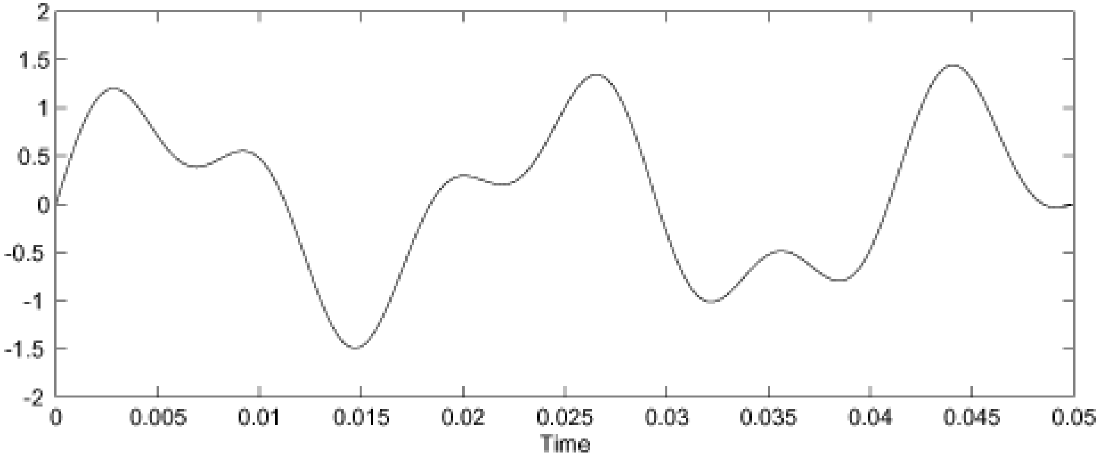
[background :: sine superposition :: $a = \sin(2\pi 50t)$]

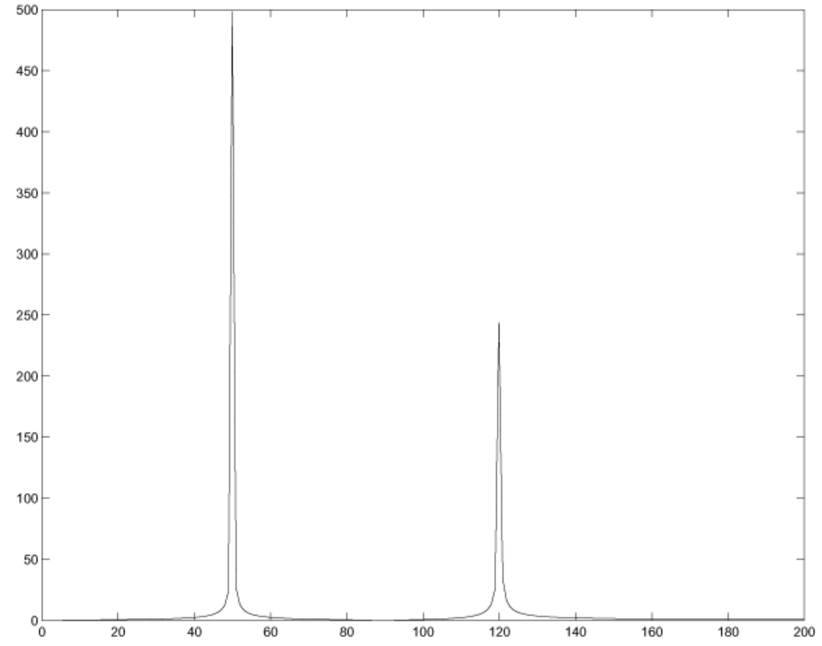


[background :: sine superposition :: $b = 0.5\sin(2\pi 120t)$]

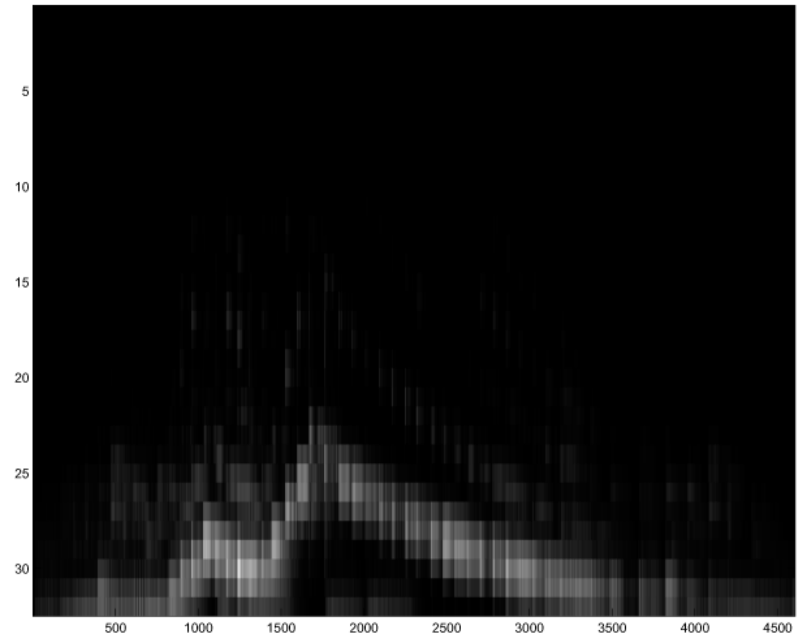


[background :: sine superposition :: $c = a + b$]

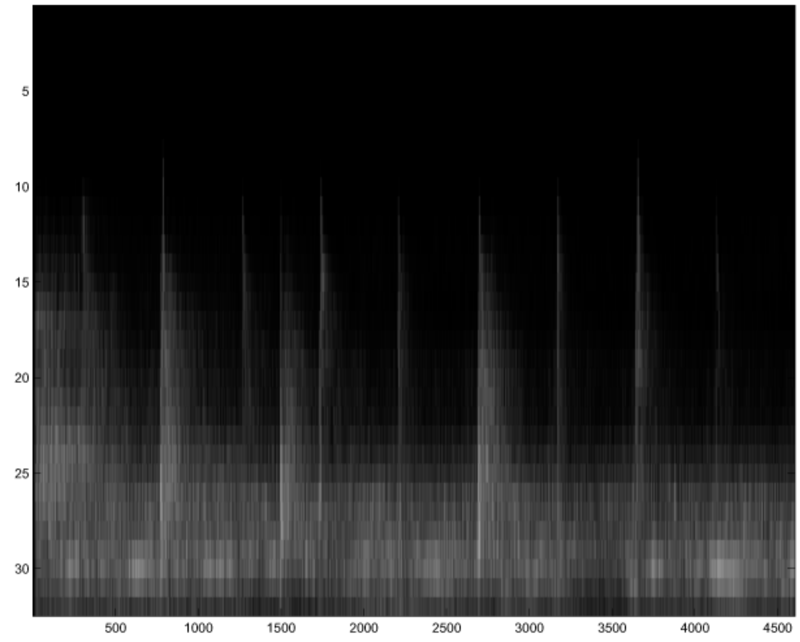



500
450
400
350
300
250
200
150
100
50
0
0
20
40
60
80
100
120
140
160
180
200

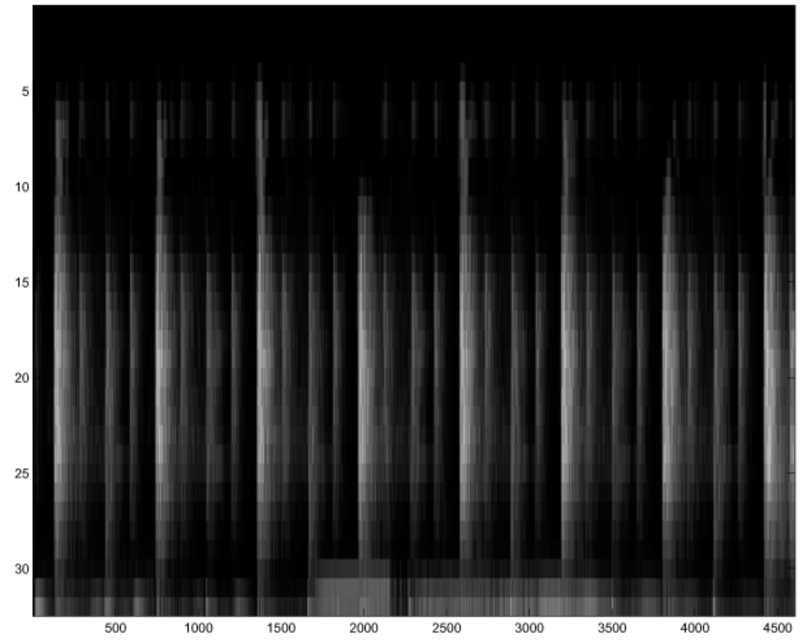
[classical spectrogram]



[metal spectrogram]



[dance spectrogram]



- spectral centroid

balancing point of spectrum
measure of spectral shape
associated with spectral brightness

$$C = \frac{\sum_{n=1}^{N} M_t[n] \cdot n}{\sum_{n=1}^{N} M_t[n]}$$

▷ rolloff

measure of spectral shape
frequency R corresponding to
$r\%$ of the magnitude distribution, so that

$$\sum_{n=1}^{R} M_t[n] = r \cdot \sum_{n=1}^{N} M_t[n]$$

in mugrat prototype $r = 80\%$

- flux

  measure of local spectral change

$$F = \sum_{n=1}^{N} \left(N_t[n] - N_{t-1}[n]\right)^2$$

- zero-crossing rate

zero-crossing:
successive samples in a digital signal have different signs
measure of the noisiness of a signal
time domain feature!

$$Z = \sum_{n=1}^{N} |s(x[n]) - s(x[n-1])|$$

DWT, envelope extraction, autocorrelation, beat histogram generation: main beat (strength and BPM), second-strongest beat, relationship of these two, general beatedness

- relative amplitude
  of first and second beat histogram peak
- ratio of amplitude second peak / first peak
- period of the first and second peak in BPM
- sum of the histogram
  (indication of beat strength)

genres defined in terms of typical members

do i know songs that sound like this one?

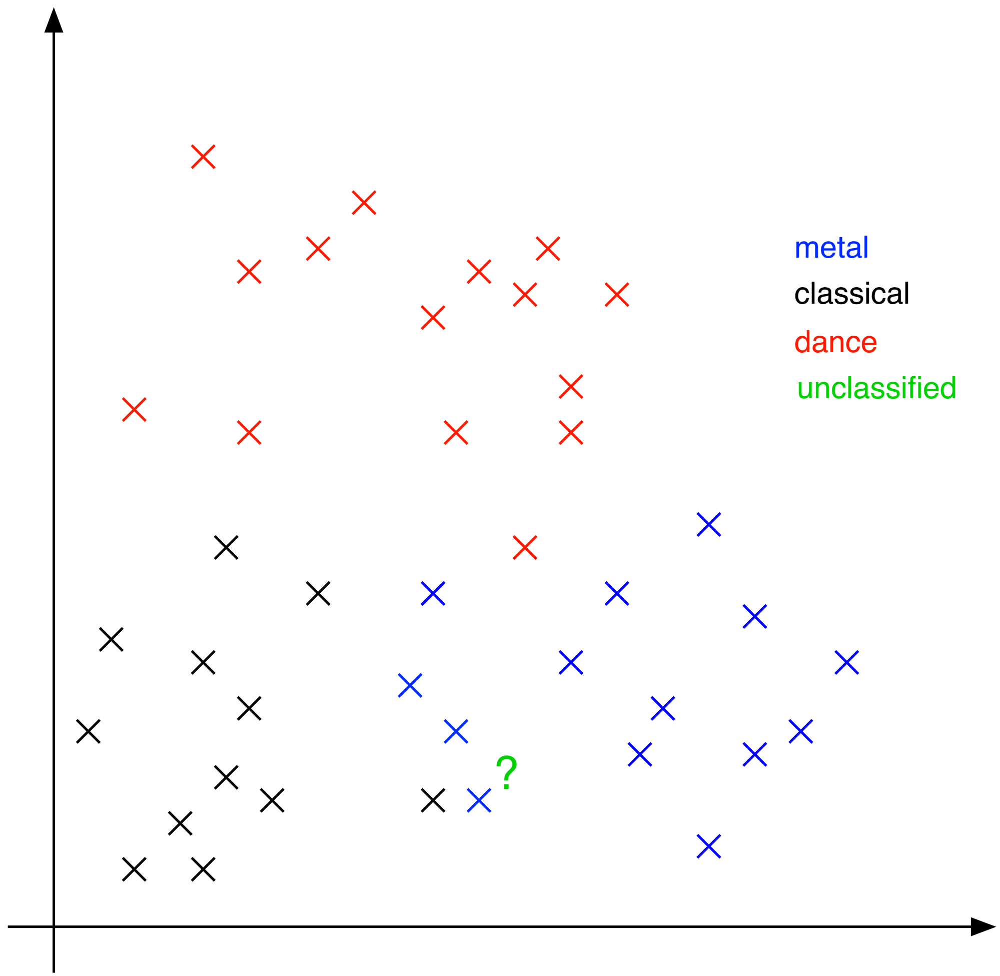
genres defined in terms of typical members
do i know songs that sound like this one?

feature extraction == abstraction
use “standard” machine learning techniques

genres defined in terms of typical members
do i know songs that sound like this one?

feature extraction == abstraction
use “standard” machine learning techniques

k-nearest-neighbour classification:
data items == points in feature space
labels of songs that are close to test instance,
weighted by distance


metal
classical
dance
unclassified
?

3 genres:

- metal
- dance
- classical

189 test songs (63, 65, 61)

88.36% accuracy

remember:

information is not knowledge; knowledge is not wisdom;
wisdom is not truth; truth is not beauty; beauty is not love;
love is not music; music is the best

(frank zappa)