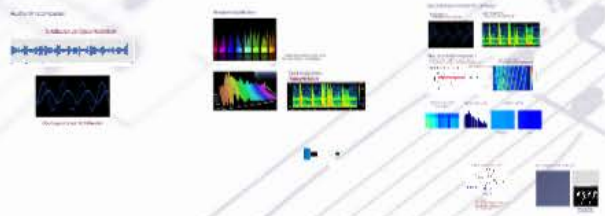
W1 - SOUND & AUDIOVISUALS

1 Music Analysis



2 Computing Music Videos

Alexander Schindler

Next Generation Content Management Systems
Information Management
Austrian Institute of Technology

Alexander.Schindler@ait.ac.at

Thomas Lidy

Institute of Software Technology and Interactive Systems
TU Wien

lidy@ifs.tuwien.ac.at

Andreas Rauber

Institute of Software Technology and Interactive Systems
TU Wien

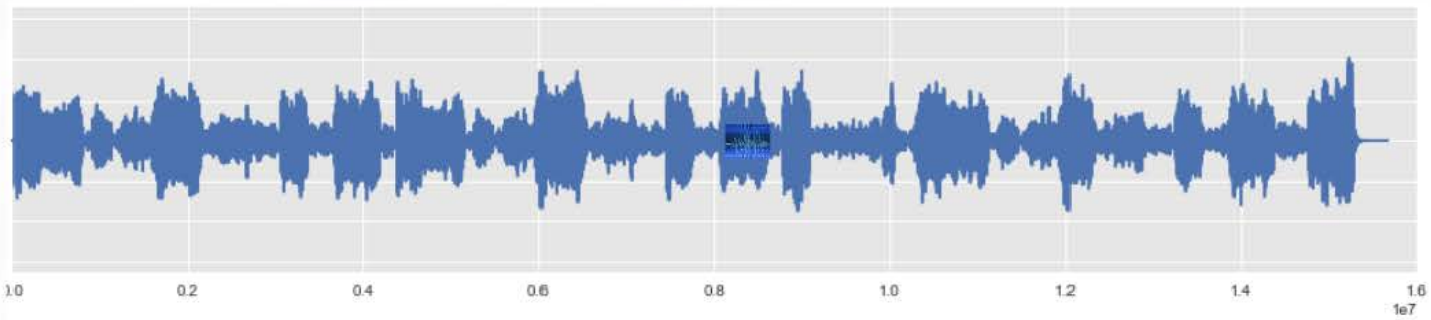
rauber@ifs.tuwien.ac.at

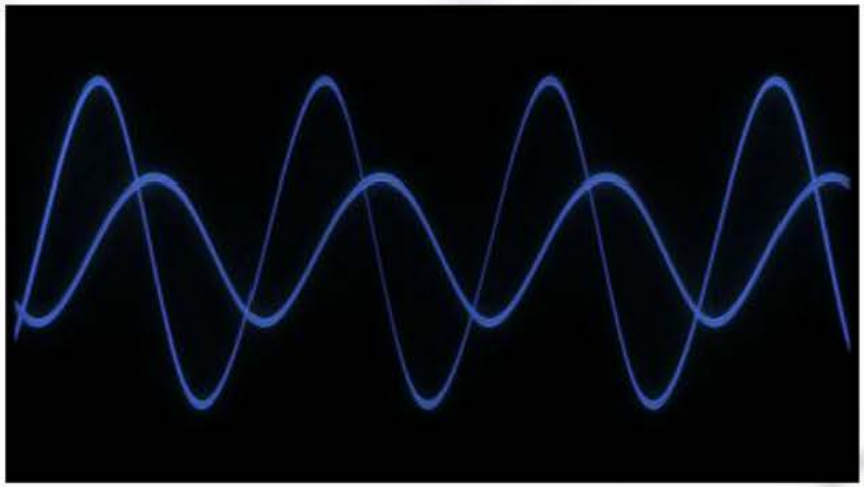
AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

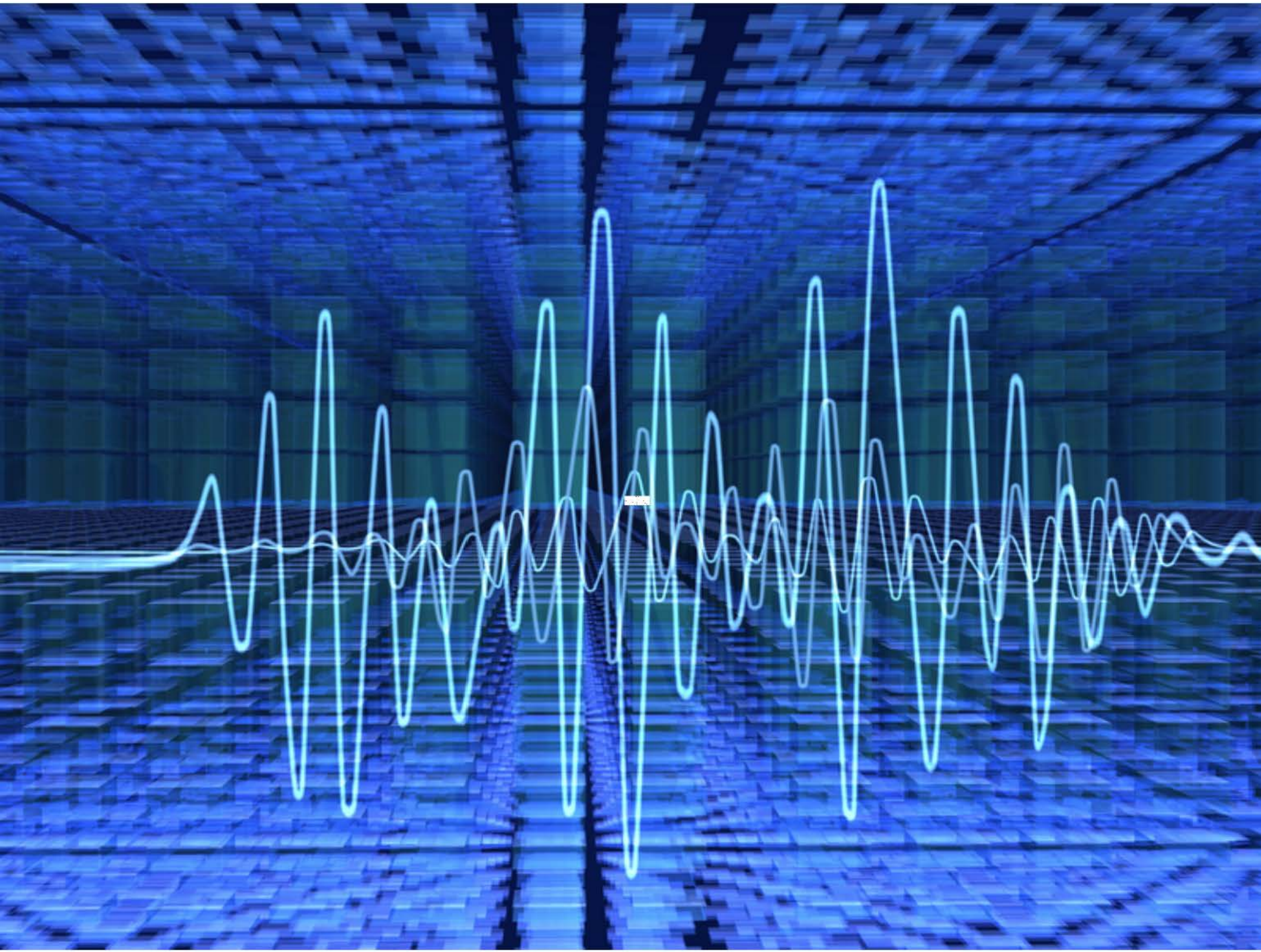
Audio im Computer

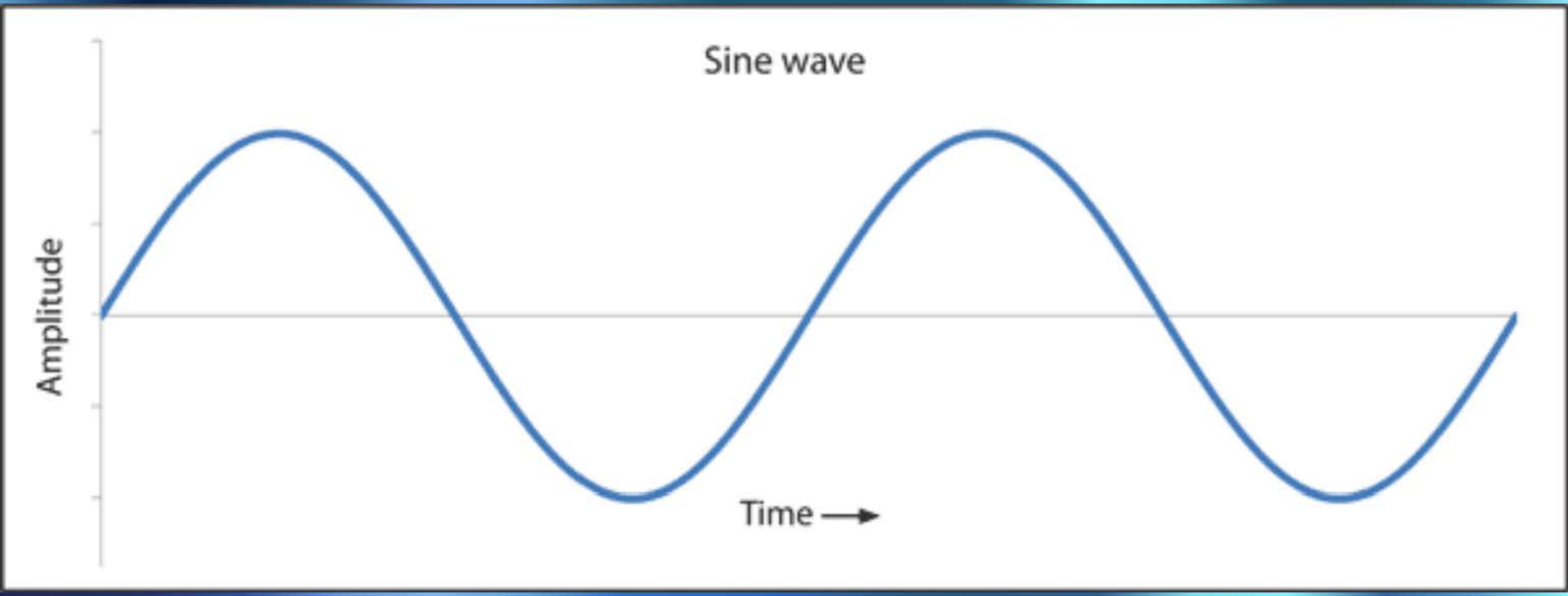
Schallwellen als digitale Wellenform





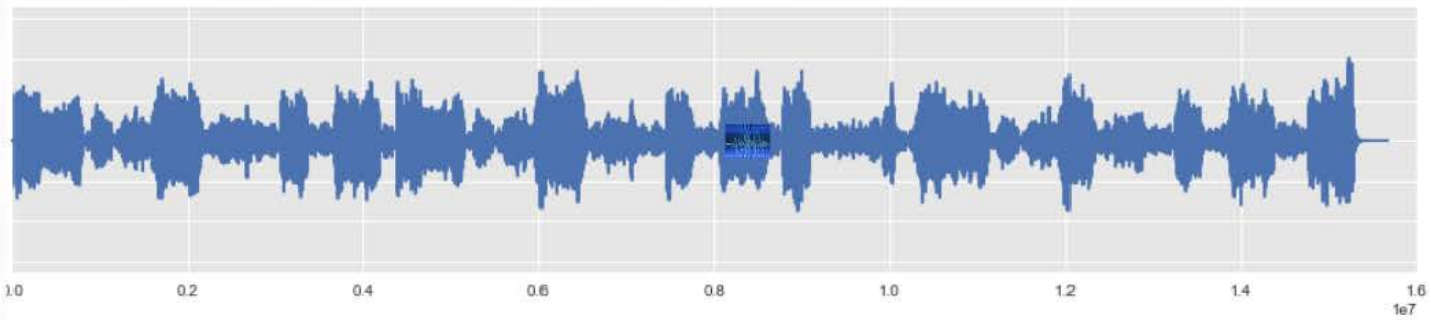
Überlagerung von Schallwellen

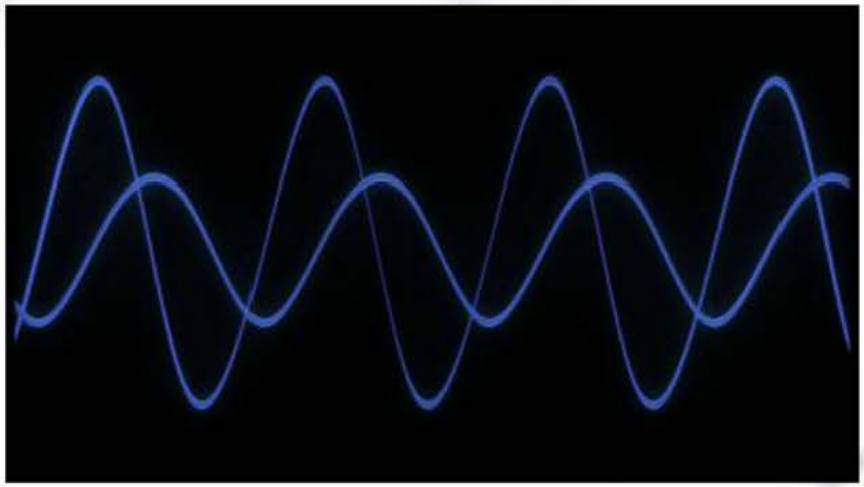



Sine wave
Amplitude
Time

Audio im Computer

Schallwellen als digitale Wellenform



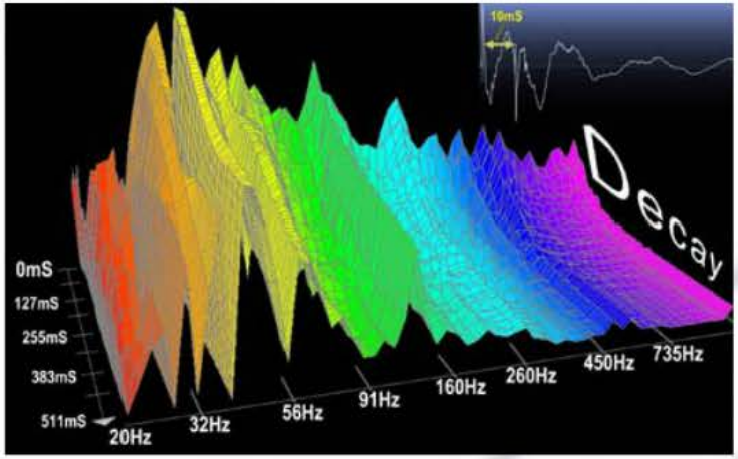


Überlagerung von Schallwellen

Frequenzspektrum

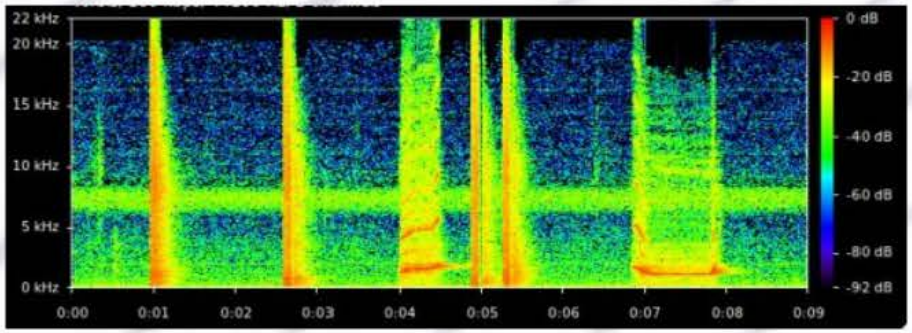
20 Hz 20.000 Hz

= Stärke des Sounds (Amplitude)
für verschiedene Frequenzen

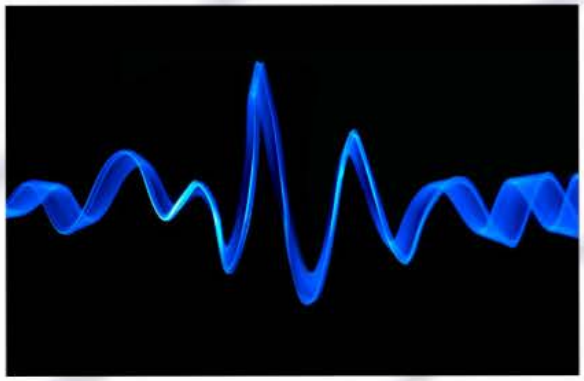


Spektrogramm

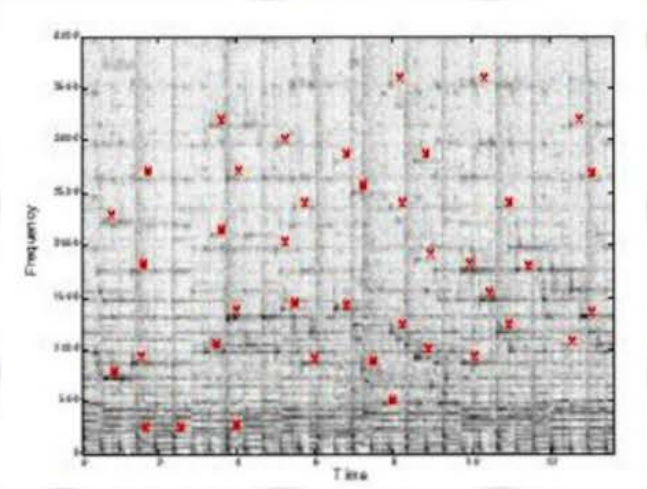
= Frequenzspektrum über die Zeit



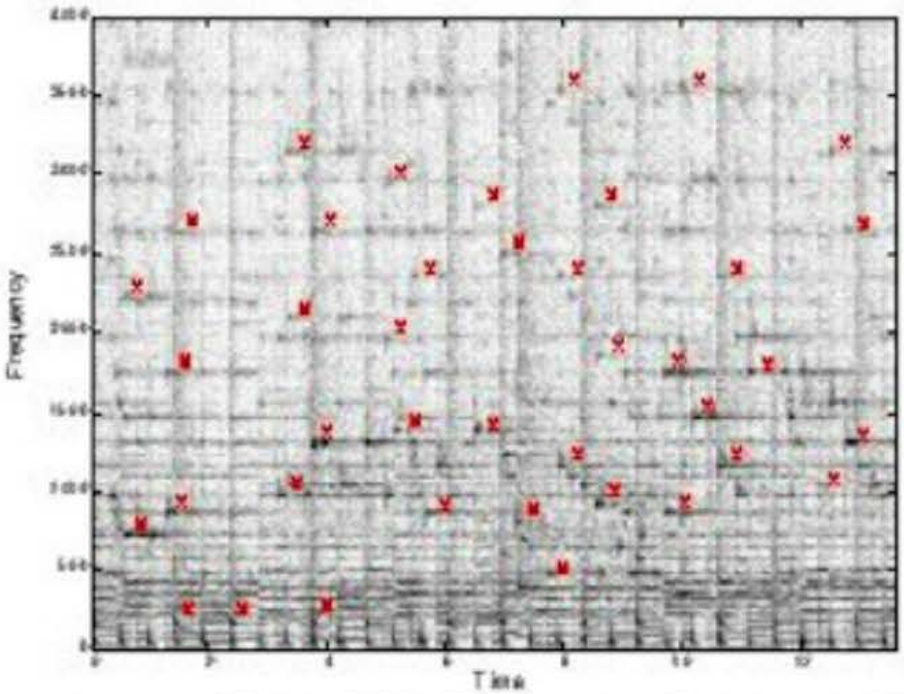
Wie funktioniert Shazam?



Spektrogramm



Spektrogramm



Shazam findet im Spektrogramm markante Punkte

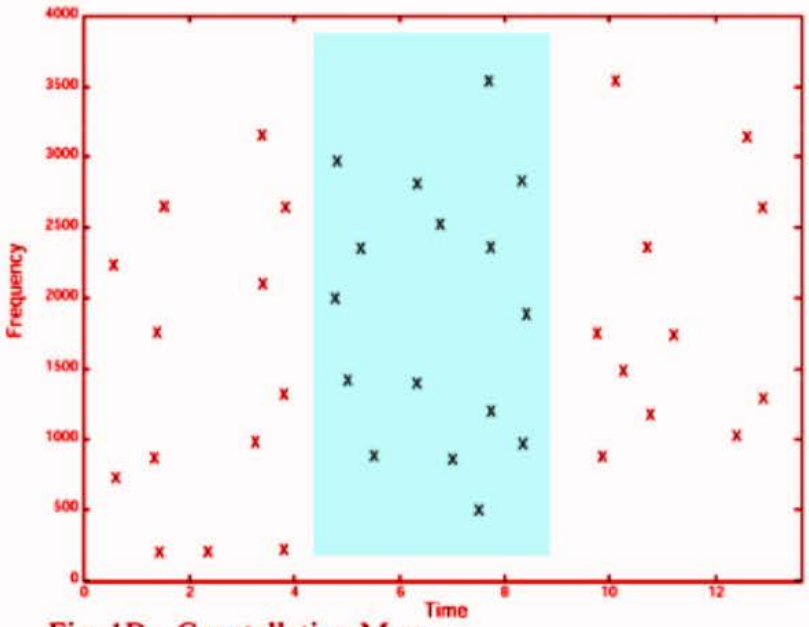


Fig. 1B - Constellation Map

Der Song mit dem

Shazam findet im Spektrogramm markante Punkte

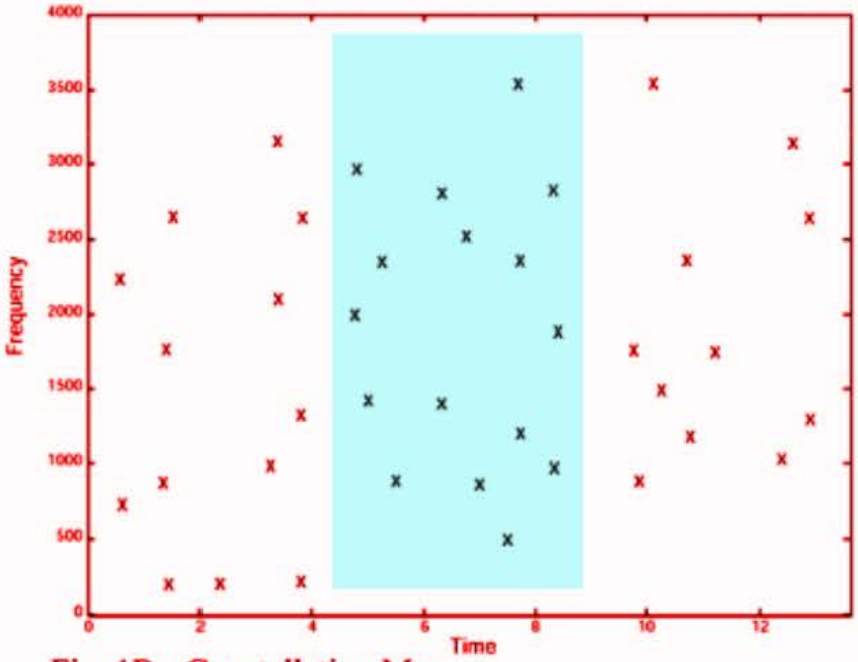
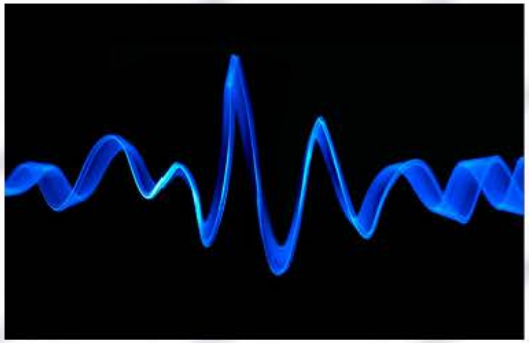
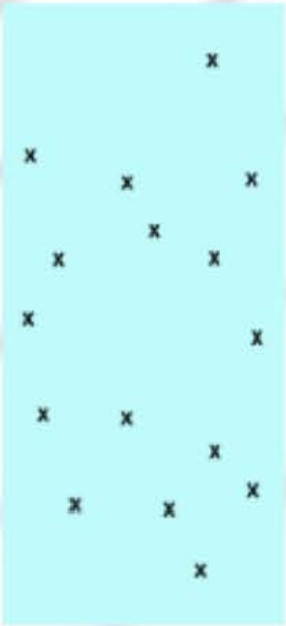


Fig. 1B - Constellation Map

Beim Abfragen eines Songs mit dem Handy werden diese Punkte verglichen



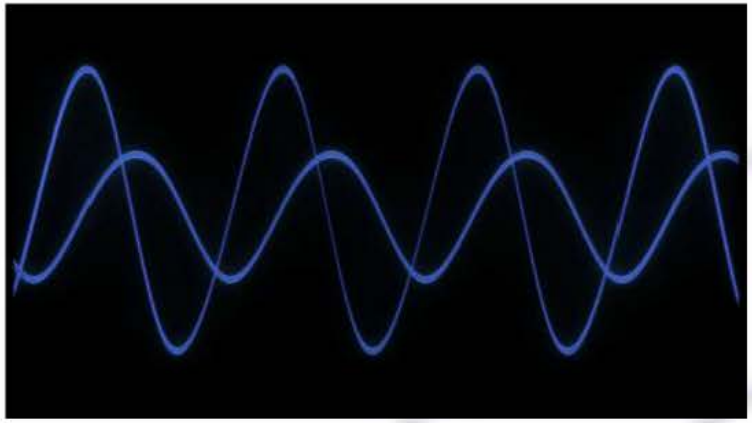


Der Song mit dem besten Treffer wird angezeigt

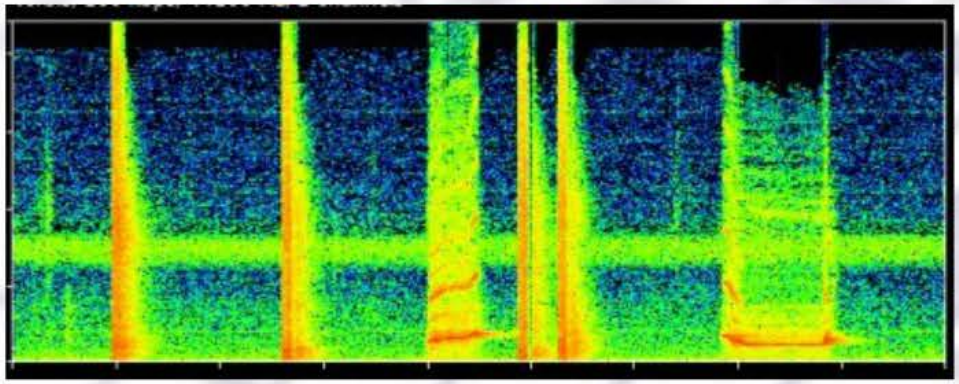
Der Song mit dem besten Treffer wird angezeigt

Darstellung von Audio im Computer

Wellenform
= Überlagerung von Schallwellen



Spektrogramm
= Frequenzspektrum über die Zeit

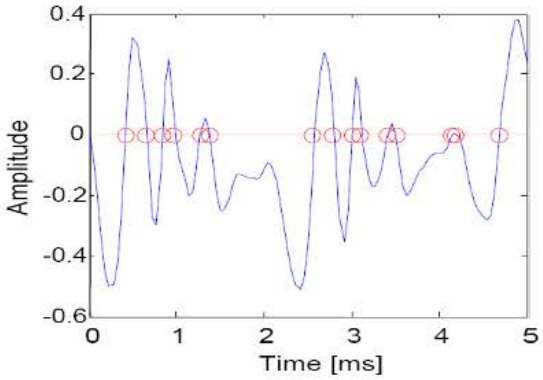


Was sind Audio Features?

Information über Rhythmus, Genre, Tonart,
Instrumente etc. die der Computer analysieren kann

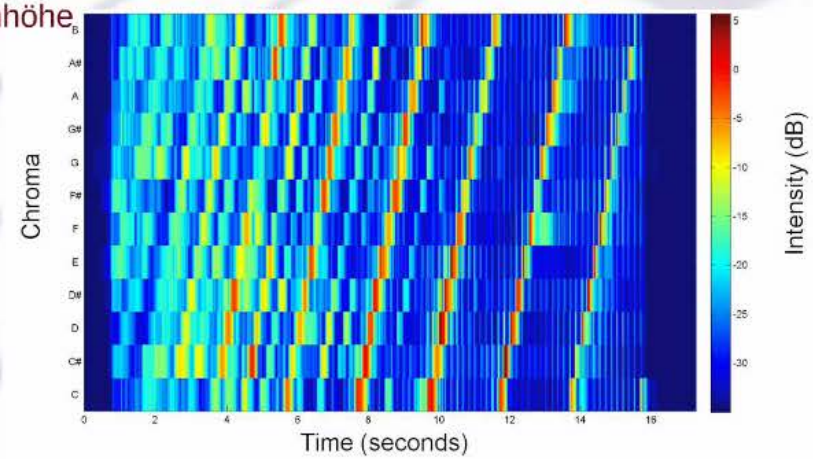
Zero Crossing Rate (ZCR) = 3/ms

mißt die ungefähre Tonhöhe



= 15 / 5ms

Chroma Features
bildet die Frequenzen auf Noten ab

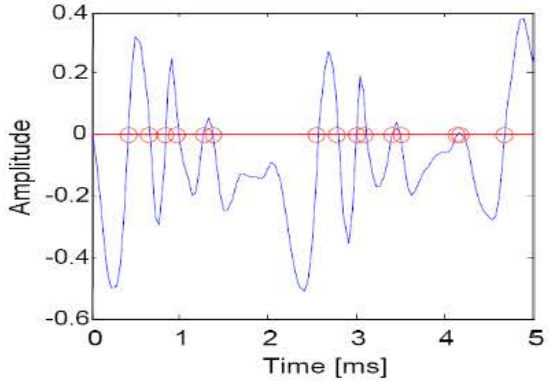


Was sind Audio Features?

Information über Rhythmus, Genre, Tonart,
Instrumente etc. die der Computer analysieren kann

Zero Crossing Rate (ZCR) = 3/ms

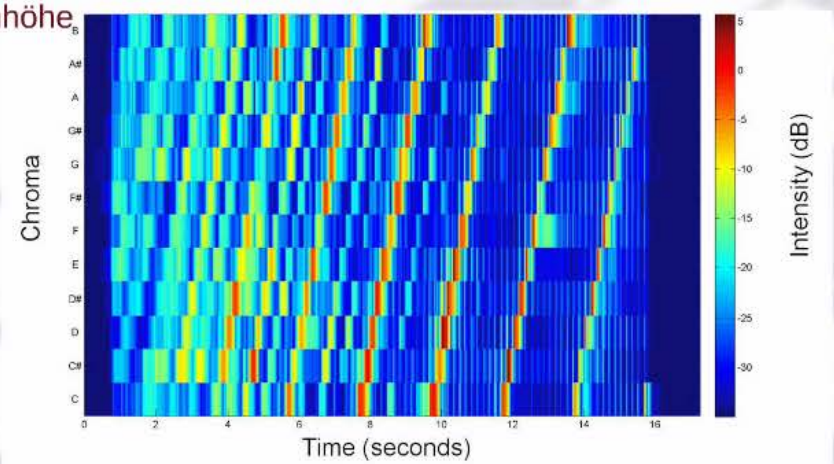
mißt die ungefähre Tonhöhe



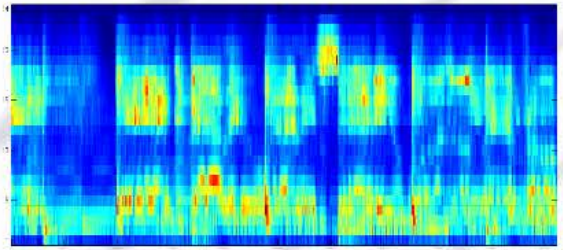
= 15 / 5ms

Chroma Features

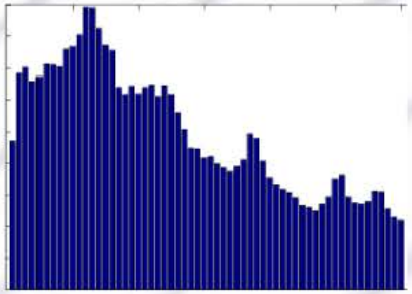
bildet die Frequenzen auf Noten ab



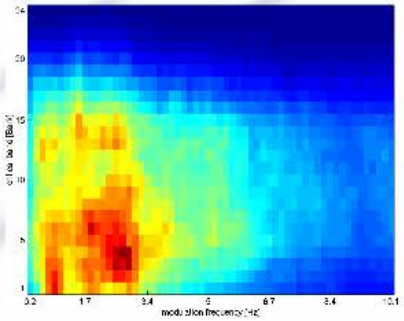
Statistical Spectrum Descriptor



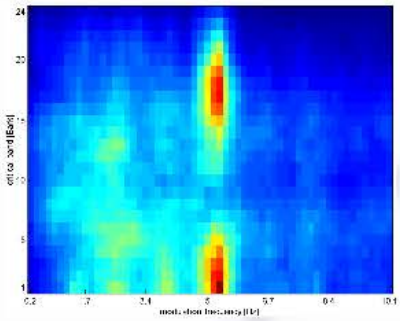
Rhythm Histogram



Rhythm Pattern

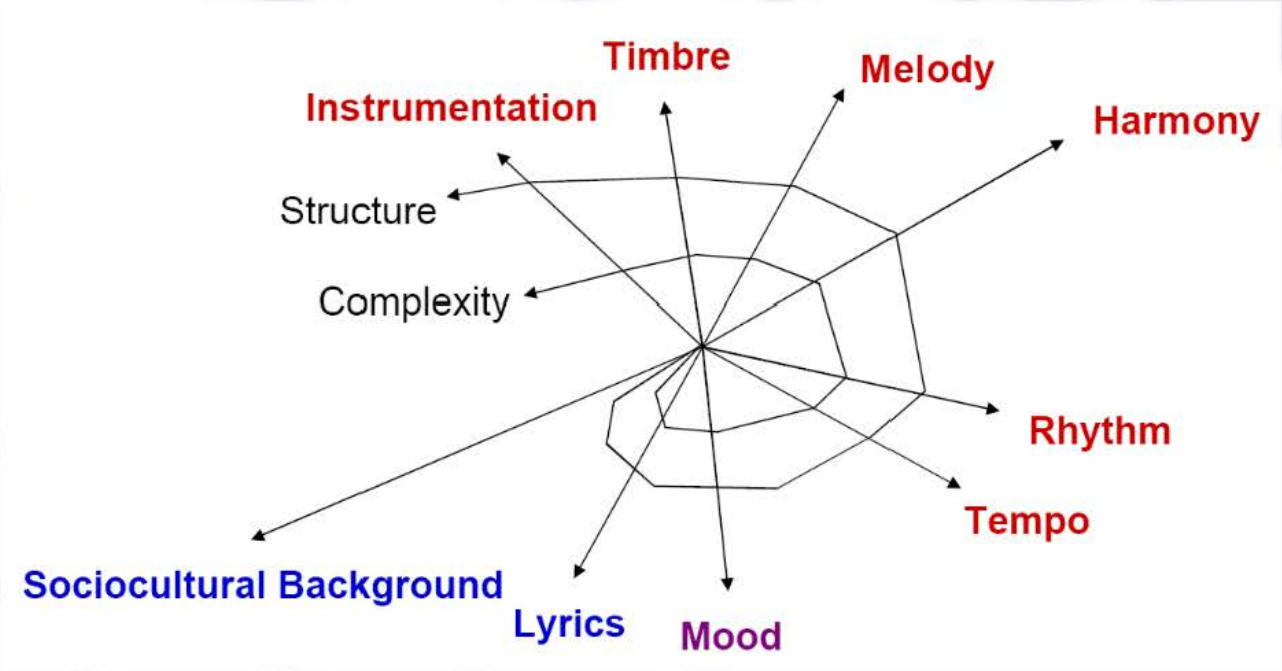


Classical



Rock

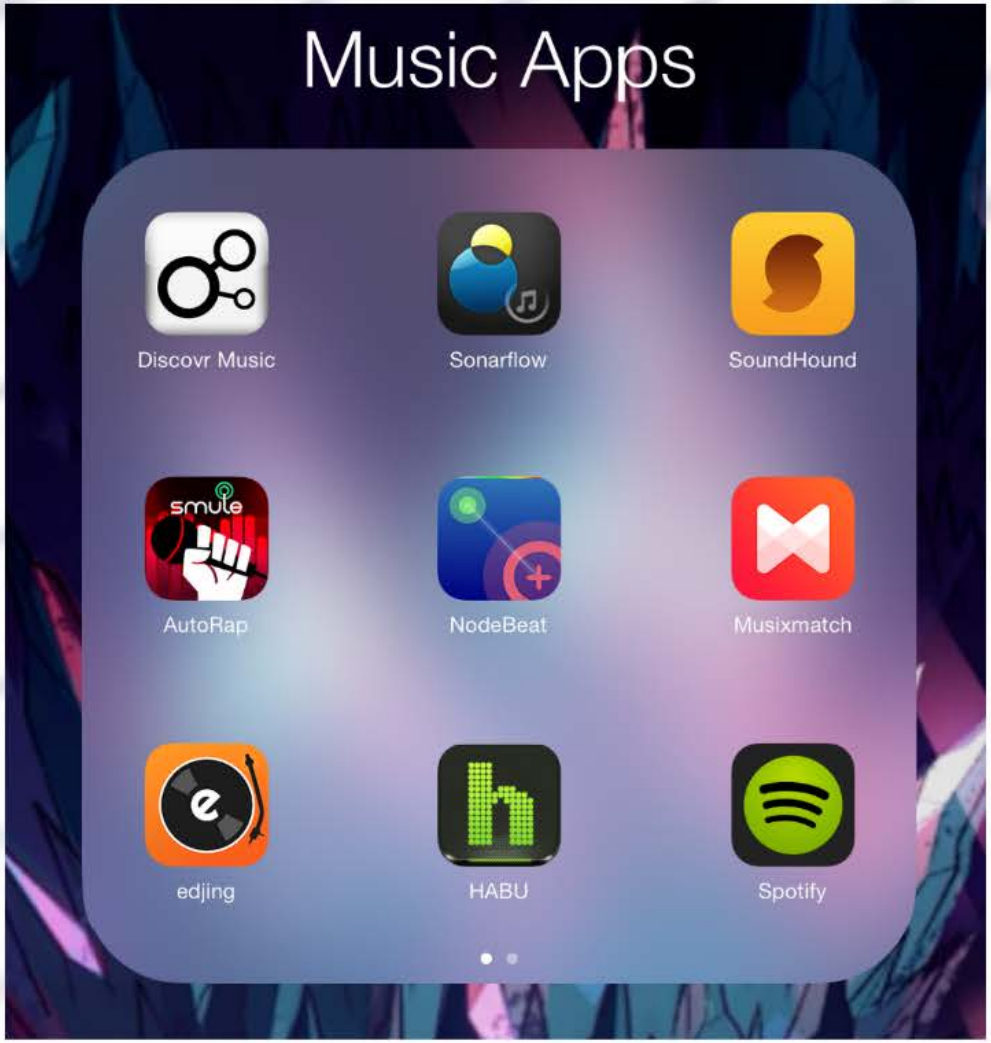
What can we analyze from Music?

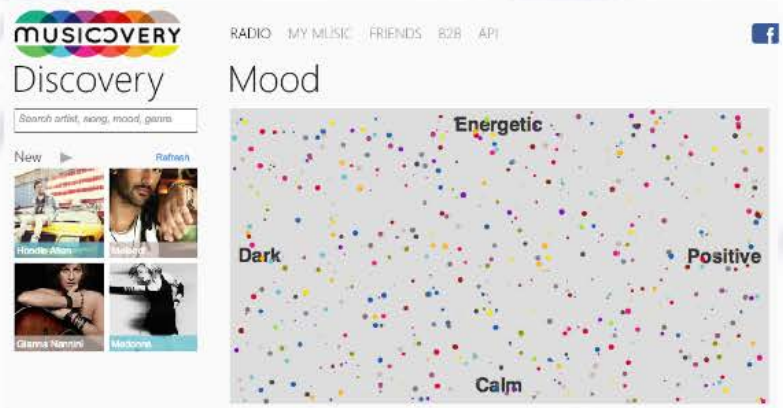


What for?

- Musik suchen
- Musik empfehlen
- Genre erkennen, automatisch kategoriesieren
- Genius Vorschläge in iTunes
- automatische Playlisten in Spotify & Co.
- Musikvideos generieren

Die besten Apps zum Musik entdecken



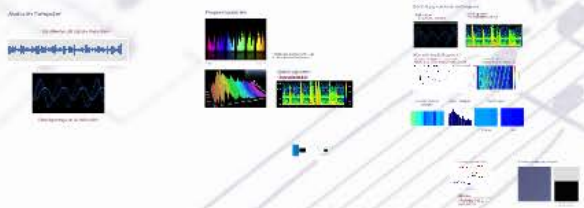


http://musicovery.com

http://songrium.jp/map/

W1 - SOUND & AUDIOVISUALS

1 Music Analysis



2 Computing Music Videos

Alexander Schindler

Next Generation Content Management Systems
Information Management
Austrian Institute of Technology

Alexander.Schindler@ait.ac.at

Thomas Lidy

Institute of Software Technology and Interactive Systems
TU Wien

lidy@ifs.tuwien.ac.at

Andreas Rauber

Institute of Software Technology and Interactive Systems
TU Wien

rauber@ifs.tuwien.ac.at

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Visual Music

"The Look of Music"

Cover-Art as a visual mnemonic to the music

Visual Language of Music

Music Genres

How to identify Music without Sound?

Album Covers provide in

"The Look of Music"

Cover-Art as a visual mnemonic to the music

Mood of a record

Metal

Country

Album Covers provide information about:

Performing Artist

Mood of a record

Metal

Country

Dance

Visual Language of Music

Guess the music genre from the video

Computing

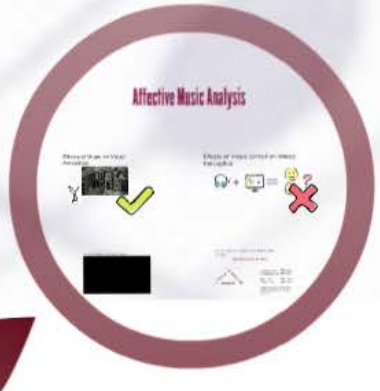
Music Video

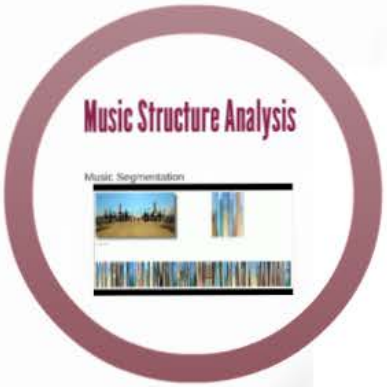
Music Recommendation

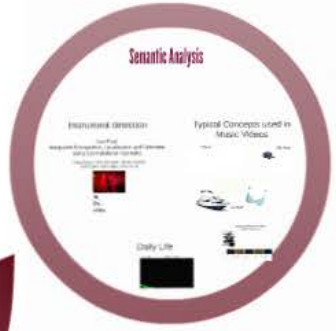
Although, it's so easy to see!

Videos

Visual Music Analysis

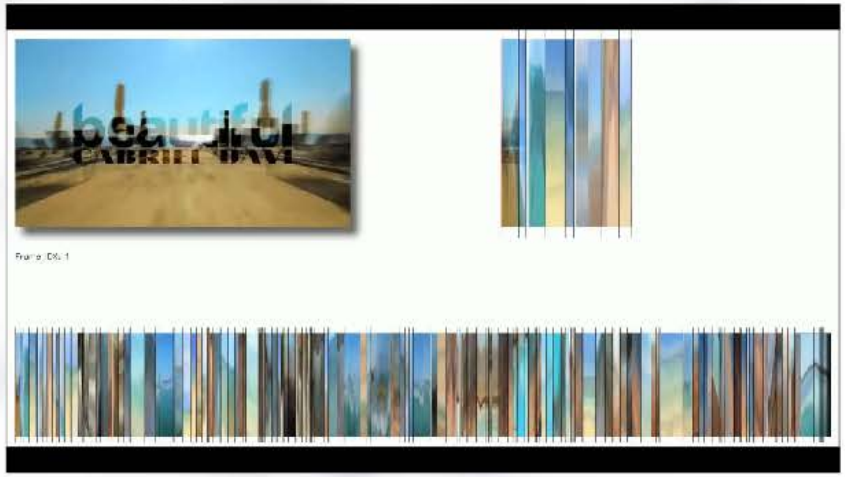




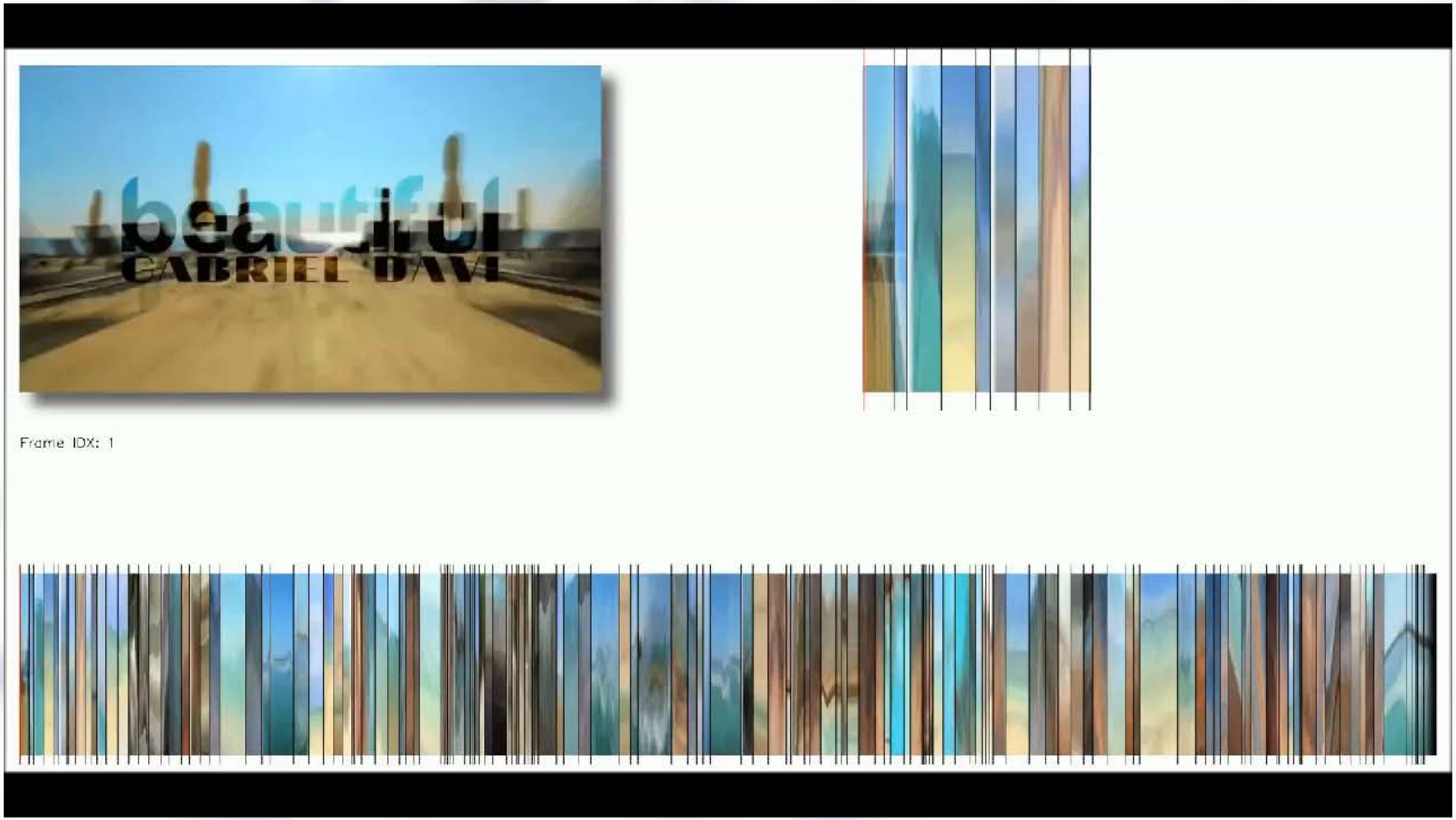


Music Structure Analysis

Music Segmentation

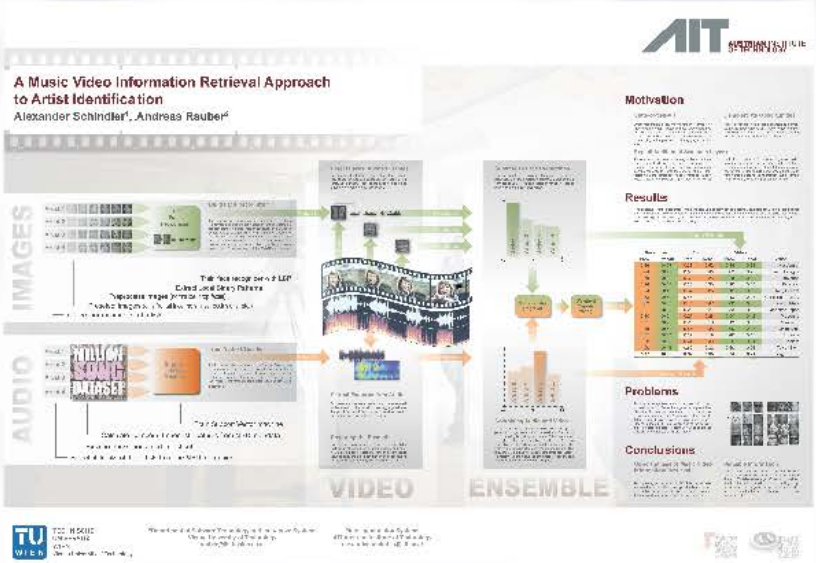


Music Segmentation



Music Indexing

Artist Identification



A Music Video Information Retrieval Approach to Artist Identification

Alexander Schindler[1], Andreas Rauber[2]

AUSTRIAN INSTITUTE OF TECHNOLOGY

IMAGES

Artist 1
Artist 2
Artist 3
Artist 4

LBP Face Recognizer

Build Face Recognizer

Train face recognizer with LBP

Extract Local Binary Patterns

Preprocess images (normalize, crop faces)

Preselect images (only frontal face, no smiles, occlusions, etc.)

Collect portrait images of artists

AUDIO

Artist 1
Artist 2
Artist 3
Artist 4

MILLION SONG DATASET

Support Vector Machine

Train Audio Classifier

Train Support Vector machine

Calculate Temporal Echonest Features from MSD metadata

Exclude tracks contained in test set

Select all tracks of the artists from the MSD for training

VIDEO

Detect Faces in Video Frames

Predict

Prediction

Prediction

Prediction

Prediction

Extract Features from Audio

Preparing the Ensemble

ENSEMBLE

Summarize Face Predictions

Artist 1 Artist 2 Artist 3 Artist 4

Bootstrap Aggregation

Weighted Majority Voting

Artist 1 Artist 2 Artist 3 Artist 4

Combining Audio and Video

Motivation

State-of-the-Art

Demonstrate Opportunities

Exploit Additional Semantic Layers

Results

Video Results

Ensemble		Audio		Video		
Prec.	Recall	Prec.	Recall	Prec.	Recall	Artist
0.38	0.57	0.33	0.62	0.34	0.33	Aerosmith
0.64	0.45	0.50	0.45	0.62	0.25	Avril Lavigne
0.55	0.32	0.35	0.23	0.28	0.42	Beyonce
0.24	0.27	0.28	0.36	0.20	0.04	Bon Jovi
0.34	0.42	0.32	0.33	0.16	0.17	Britney Spears
0.33	0.50	0.46	0.71	0.18	0.43	Christina Aguilera
0.62	0.63	0.41	0.41	0.00	0.00	Foo Fighters
0.27	0.19	0.22	0.24	0.33	0.14	Jennifer Lopez
0.30	0.34	0.27	0.28	0.50	0.12	Madonna
0.35	0.20	0.20	0.10	0.12	0.50	Maroon 5
0.58	0.44	0.65	0.35	1.00	0.18	Nickelback
0.26	0.14	0.29	0.19	0.40	0.10	Rihanna
0.28	0.60	0.44	0.40	0.28	0.21	Shakira
1.00	0.16	0.60	0.32	0.50	0.06	Taylor Swift
0.47	0.38	0.37	0.38	0.34	0.21	avg

Audio Results

Problems

Conclusions

Opportunities of Music Video Information Retrieval

Valuable Information

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

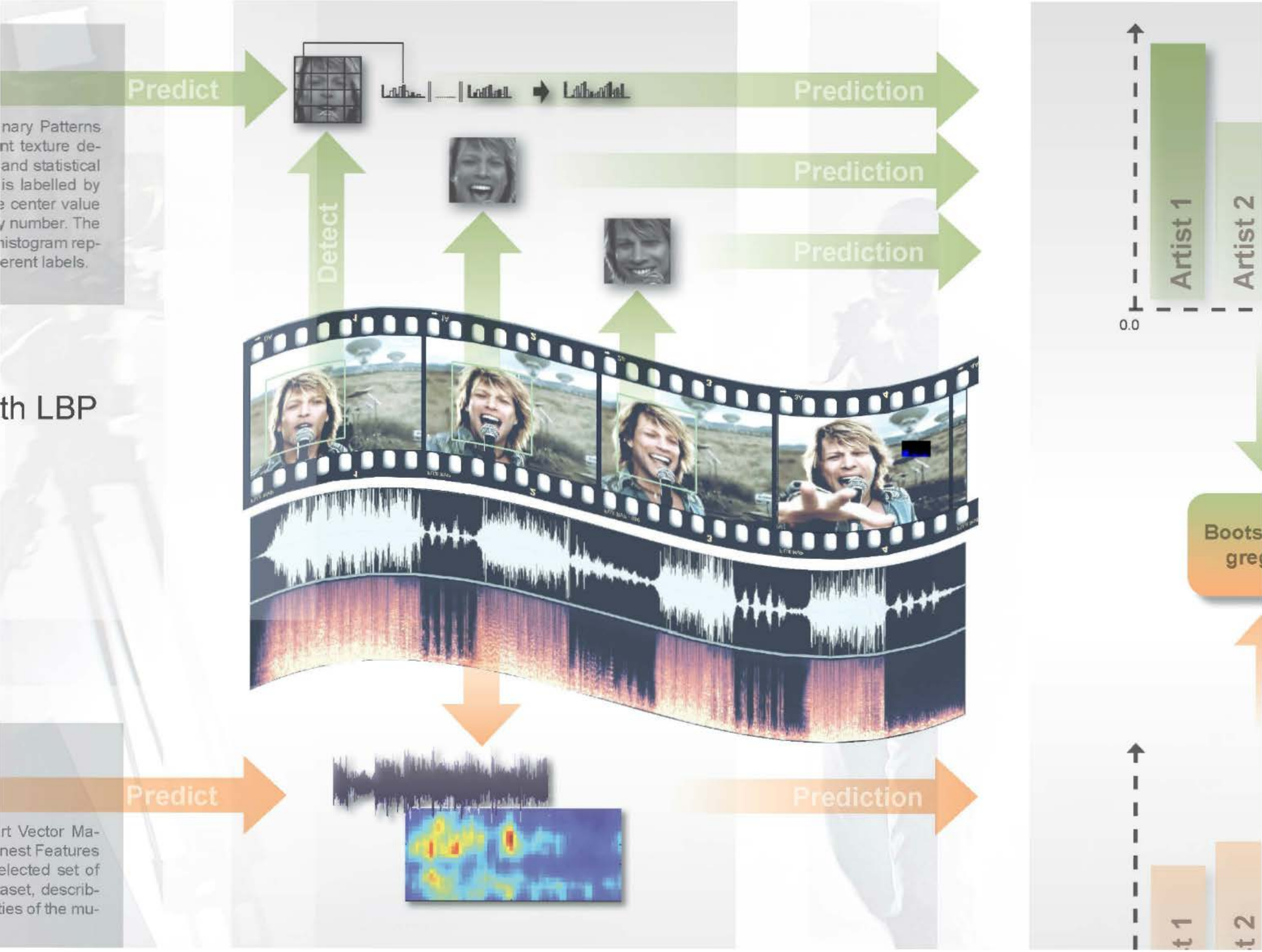
[1]Department of Software Technology and Interactive Systems
Vienna University of Technology
rauber@ifs.tuwien.ac.at

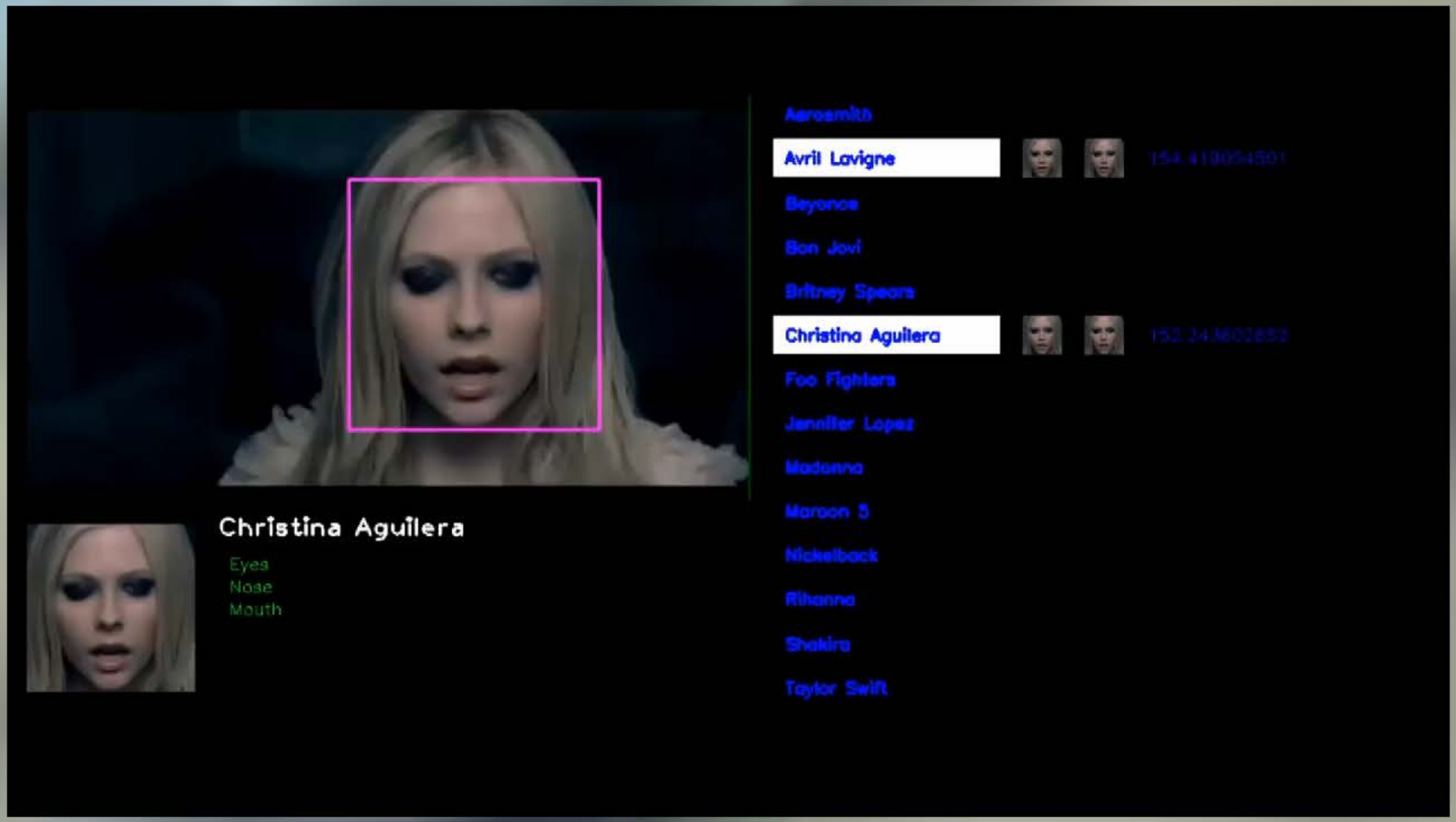
[2]Intelligent Vision Systems
AIT Austrian Institute of Technology
alexander.schindler@ait.ac.at

nary Patterns
nt texture de-
and statistical
is labelled by
e center value
y number. The
histogram rep-
erent labels.

th LBP

rt Vector Ma-
nest Features
elected set of
aset, describ-
ties of the mu-




Christina Aguilera
Eyes
Nose
Mouth
Aerosmith
Avril Lavigne
Beyonce
Bon Jovi
Britney Spears
Christina Aguilera
Foo Fighters
Jennifer Lopez
Madonna
Maroon 5
Nickelback
Rihanna
Shakira
Taylor Swift

Affective Music Analysis

Effects of Music on Visual Perception

Effects of Visual Stimuli on Music Perception

Emotion Recognition in Music Videos

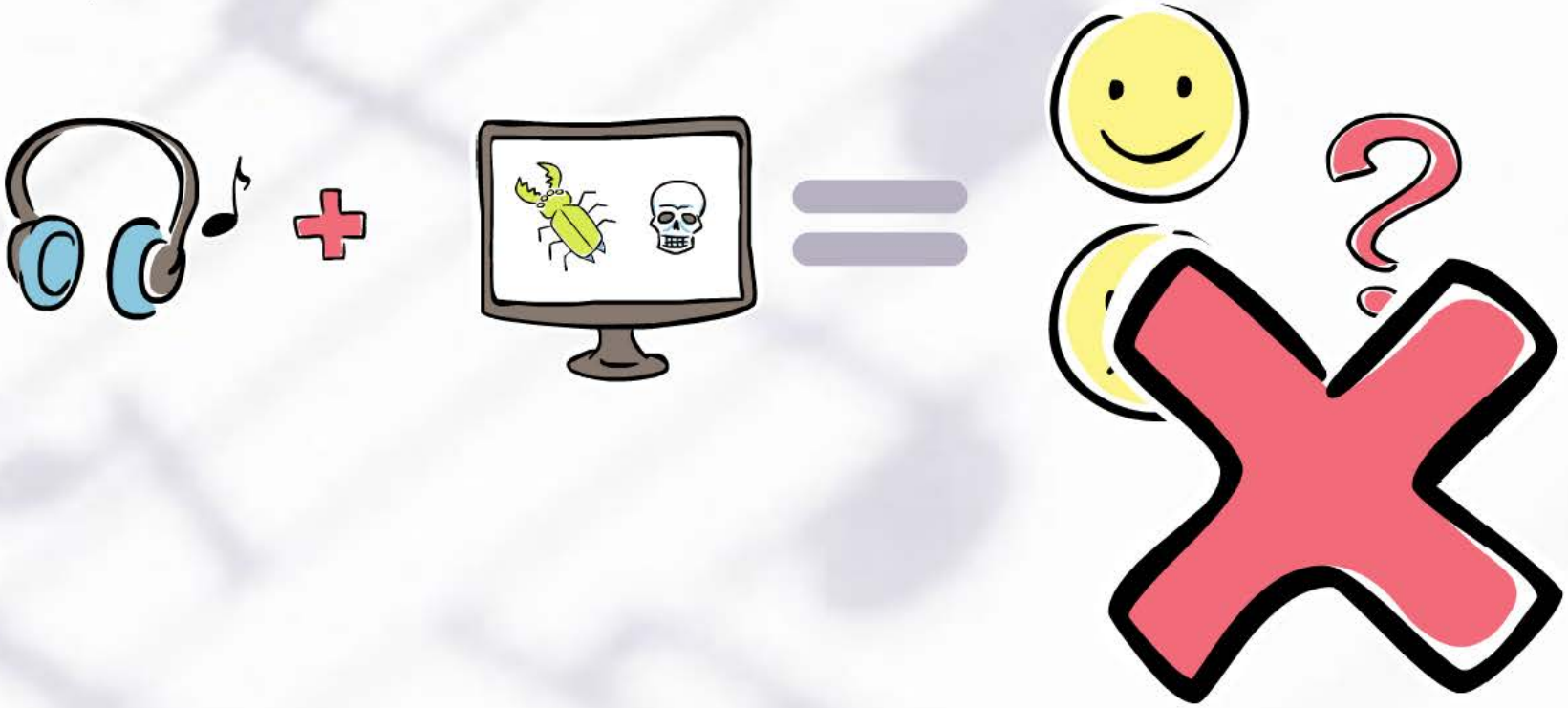
Music-color associations are mediated by emotion

Stephen E. Palmer, Karen B. Schloss, Zena Xu, Glad R. Prado-León

- Saturated, light — Major
- Unsaturated, dark — Minor
- Blue, Red — Slow
- Yellow, Green — Fast
- High correlations between emotions based on music and color

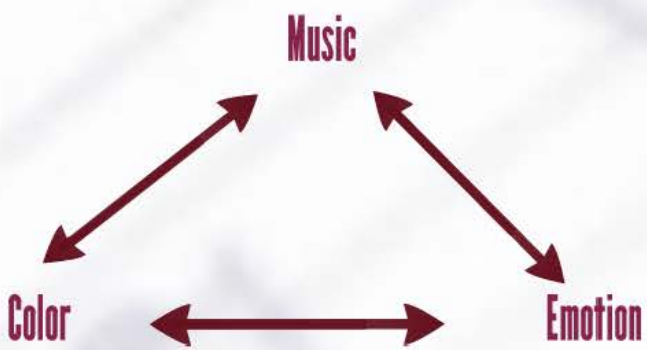
Effects of Music on Visual Perception

Effects of Visual Stimuli on Music Perception



Music–color associations are mediated by emotion

Stephen E. Palmer, Karen B. Schloss, Zoe Xu, Lilia R. Prado-Leónb



- Saturated, light ⇨ Major
- Unsaturated, dark ⇨ Minor

- Blue, Red ⇨ Slow
- Yellow, Green ⇨ Fast

- High correlations between emotions based on music and color

Emotion Recognition in Music Videos

Semantic Analysis

Instrument detection

OverFeat:
Integrated Recognition, Localization and Detection using Convolutional Networks

Xiang Zhang, Pierre Sermanet, Michael Mathieu, David Eigen, Rob Fergus, Yann LeCun

Typical Concepts used in Music Videos

Rock

Hip Hop

Dance

Deep Learning for Video understanding

Daily Life

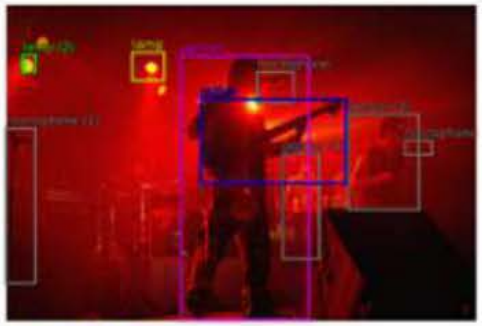
Fan Shirt

Album Cover

Instrument detection

OverFeat:
Integrated Recognition, Localization and Detection using Convolutional Networks

Xiang Zhang, Pierre Sermanet, Michael Mathieu, David Eigen, Rob Fergus, Yann LeCun



Groundtruth:
drum
lamp
lamp (2)
guitar
person
person (2)
person (3)
microphone
microphone (2)
microphone (3)

Typical Concepts used in Music Videos

Rock

Hip Hop

Dance

Deep Learning for Video understanding

stage	0.3162
electric guitar	0.1169
bassoon	0.0649
accordion	0.0611
drumstick	0.0386
microphone	0.0313
marimba	0.0276

mosquito net	0.0932
wardrobe	0.0857
brassiere	0.0815
shower curtain	0.0471
candle	0.0400
plastic bag	0.0204
hoopskirt	0.0187

maillot	0.2745
bolo tie	0.0732
Windsor tie	0.0550
letter opener	0.0486
brassiere	0.0390
bikini	0.0384
bassoon	0.0364

lumbermill	0.1925
tow truck	0.1215
harvester	0.1152
thresher	0.0513
jeep	0.0484
half track	0.0473
pickup truck	0.0460

wig	0.4399
neck brace	0.0577
chimpanzee	0.0418
hair spray	0.0375
orangutan	0.0366
cloak	0.0267
Windsor tie	0.0236

Synset	Example Images
Abaya	
Micro-phone	
Cowboy Hat	
brassiere, bra	
Spotlight	
Wig	
Stage	

stage	0.3162
electric guitar	0.1169
bassoon	0.0649
accordion	0.0611
drumstick	0.0386
microphone	0.0313
marimba	0.0276

mosquito ne
wardrobe
brassiere
shower curt
candle
plastic bag
hoopskirt

	0.3162
	0.1169
	0.0649
	0.0611
	0.0386
	0.0313
	0.0276

mosquito net	0.0932
wardrobe	0.0857
brassiere	0.0815
shower curtain	0.0471
candle	0.0400
plastic bag	0.0204
hoopskirt	0.0187

maillot
bolo tie
Windsor tie
letter opene
brassiere
bikini
bassoon

	0.0932
	0.0857
	0.0815
	0.0471
	0.0400
	0.0204
	0.0187

maillot	0.2745
bolo tie	0.0732
Windsor tie	0.0550
letter opener	0.0486
brassiere	0.0390
bikini	0.0384
bassoon	0.0364

lumbermill
tow truck
harvester
thresher
jeep
half track
pickup truck

	0.2745
	0.0732
	0.0550
	0.0486
	0.0390
	0.0384
	0.0364

lumbermill	0.1925
tow truck	0.1215
harvester	0.1152
thresher	0.0513
jeep	0.0484
half track	0.0473
pickup truck	0.0460

wig
neck brace
chimpanze
hair spray
orangutan
cloak
Windsor tie

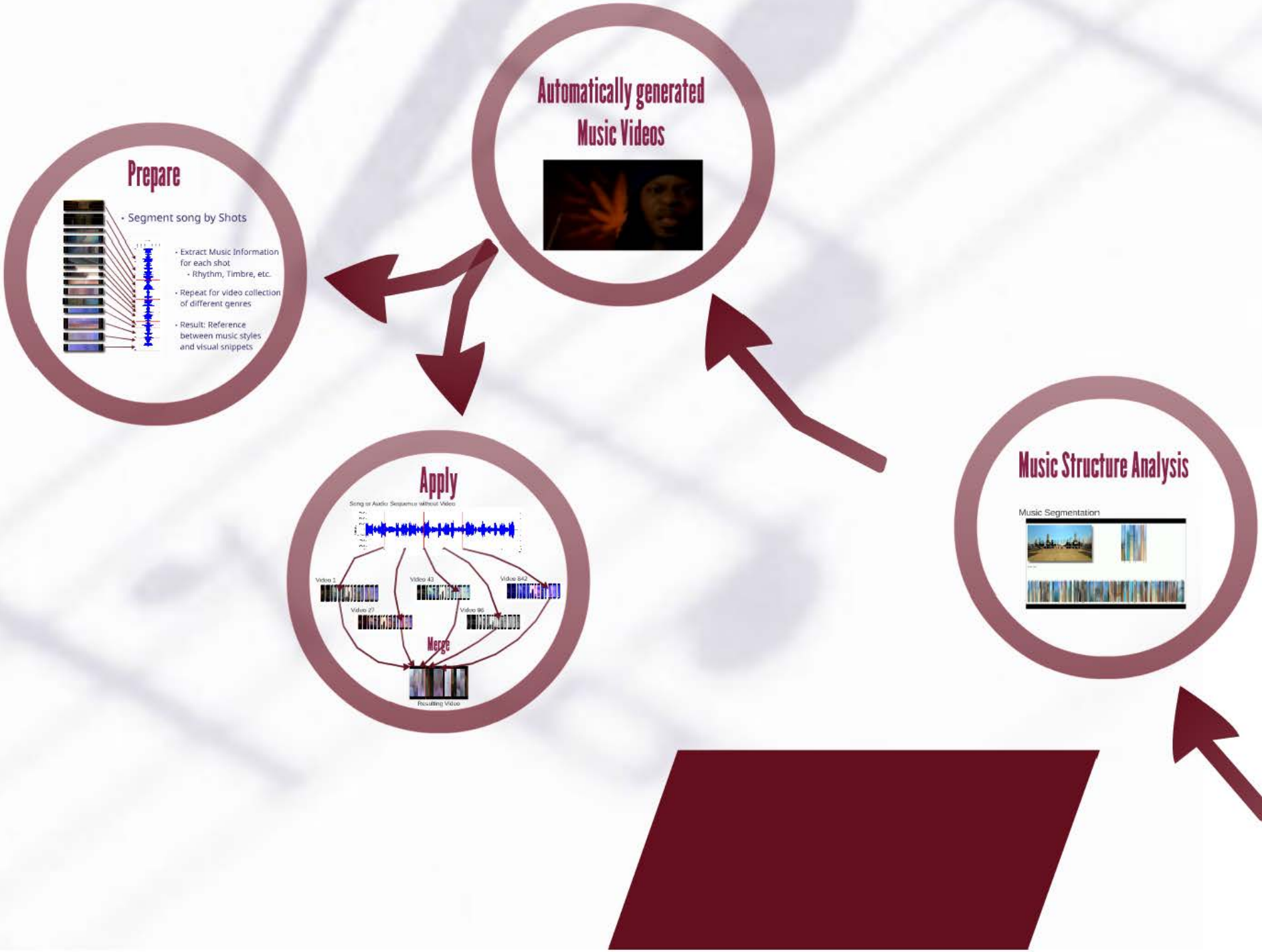
	0.1925
	0.1215
	0.1152
	0.0513
	0.0484
	0.0473
	0.0460

wig	0.4399
neck brace	0.0577
chimpanzee	0.0418
hair spray	0.0375
orangutan	0.0366
cloak	0.0267
Windsor tie	0.0236

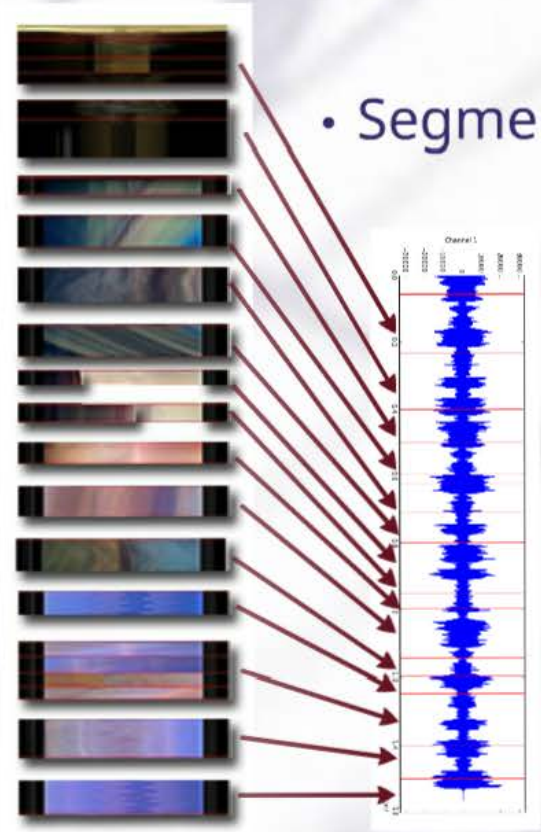
Daily Life

Fan Shirt

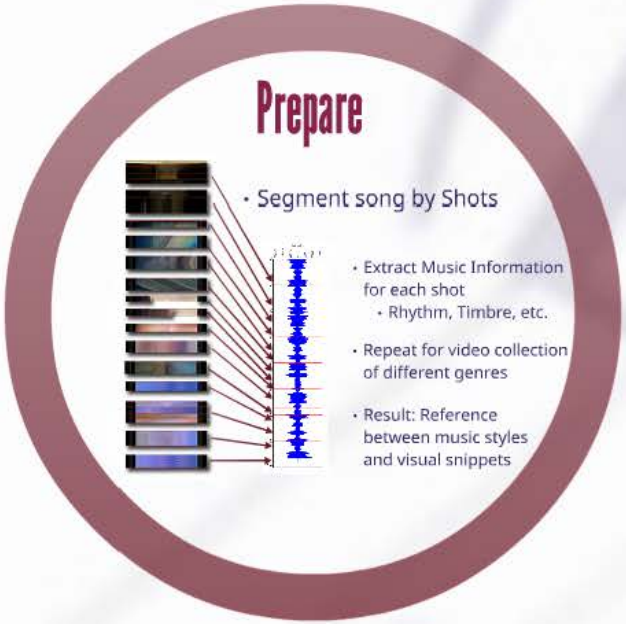
Album Cover


Automatically generated Music Videos
Prepare
• Segment song by Shots
• Extract Music Information for each shot
• Rhythm, Timbre, etc.
• Repeat for video collection of different genres
• Result: Reference between music styles and visual snippets
Apply
Song or Audio Sequence without Video
Video 1
Video 43
Video 842
Video 27
Video 96
Merge
Resulting Video
Music Structure Analysis
Music Segmentation

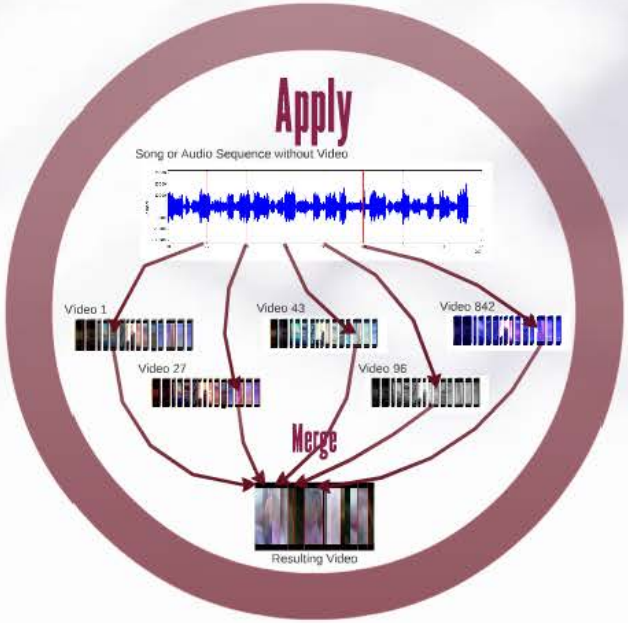
Prepare

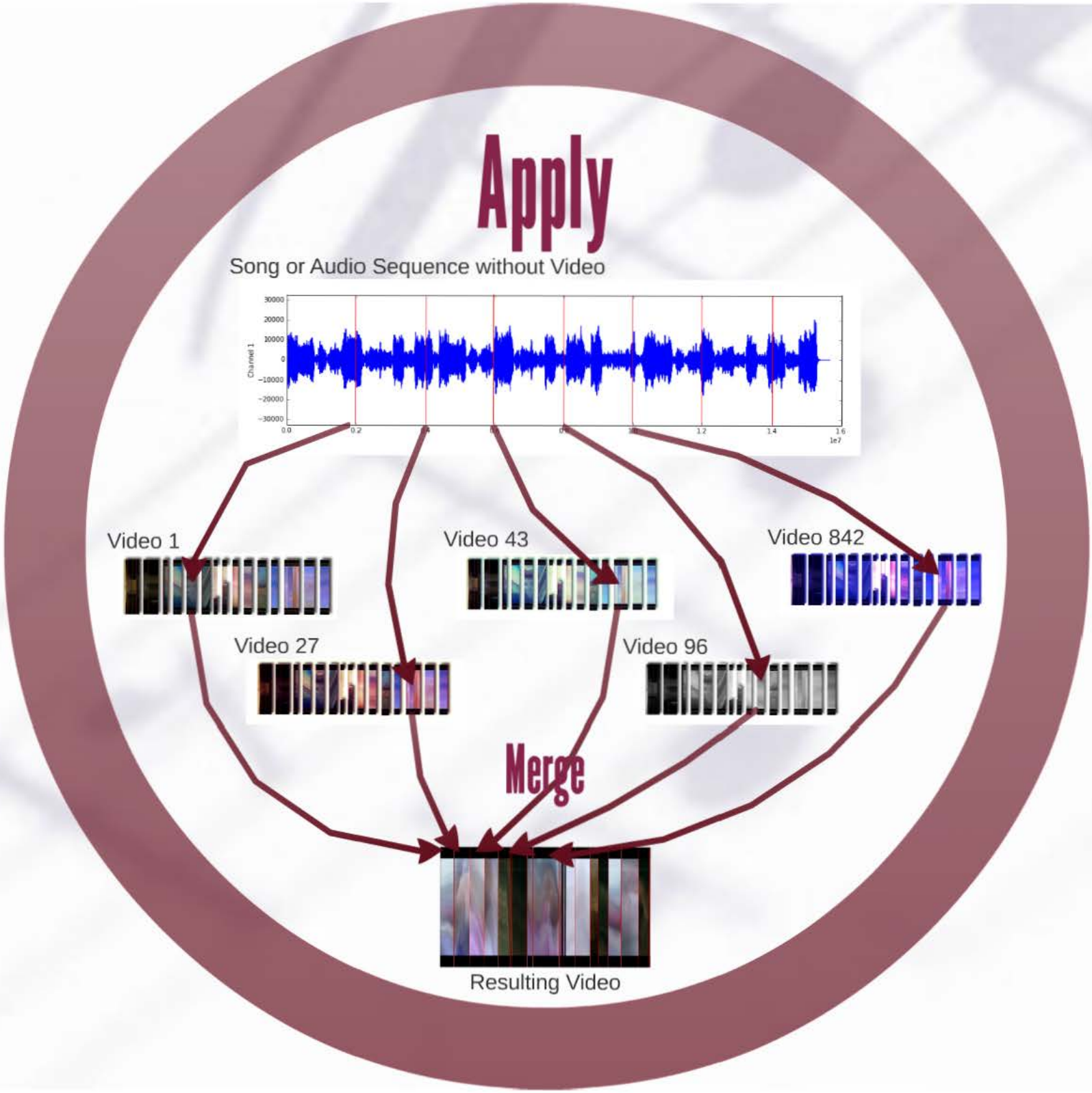


- Segment song by Shots
  - Extract Music Information for each shot
    - Rhythm, Timbre, etc.
  - Repeat for video collection of different genres
  - Result: Reference between music styles and visual snippets


Prepare
• Segment song by Shots
• Extract Music Information for each shot
• Rhythm, Timbre, etc.
• Repeat for video collection of different genres
• Result: Reference between music styles and visual snippets

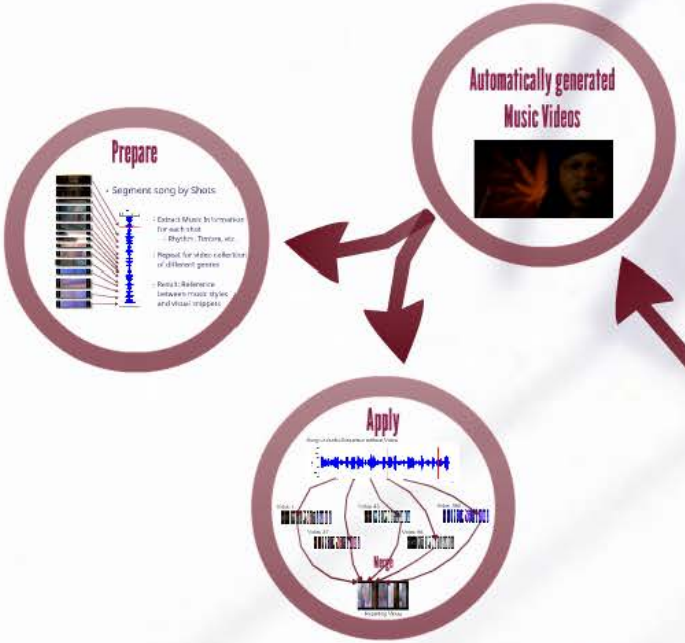
Automatically generated Music Videos

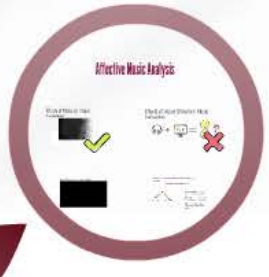

Apply
Song or Audio Sequence without Video
Video 1
Video 43
Video 842
Video 27
Video 96
Merge
Resulting Video

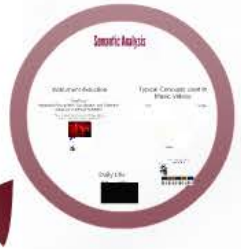

Apply
Song or Audio Sequence without Video
Channel 1
Video 1
Video 43
Video 842
Video 27
Video 96
Merge
Resulting Video

Music Videos

Visual Music Analysis



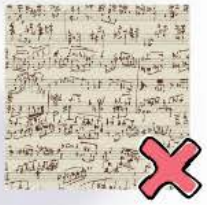




Symbolic Music

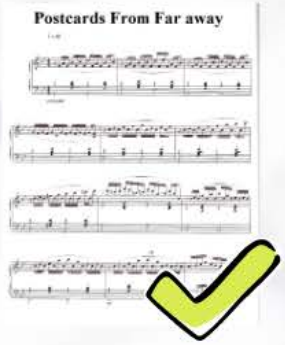
Handwritten Music Scores Processing

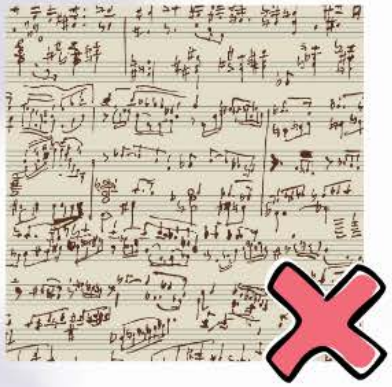
Optical Music Recognition (OMR)



Handwritten Music Scores Processing

Optical Music Recognition (OMR)





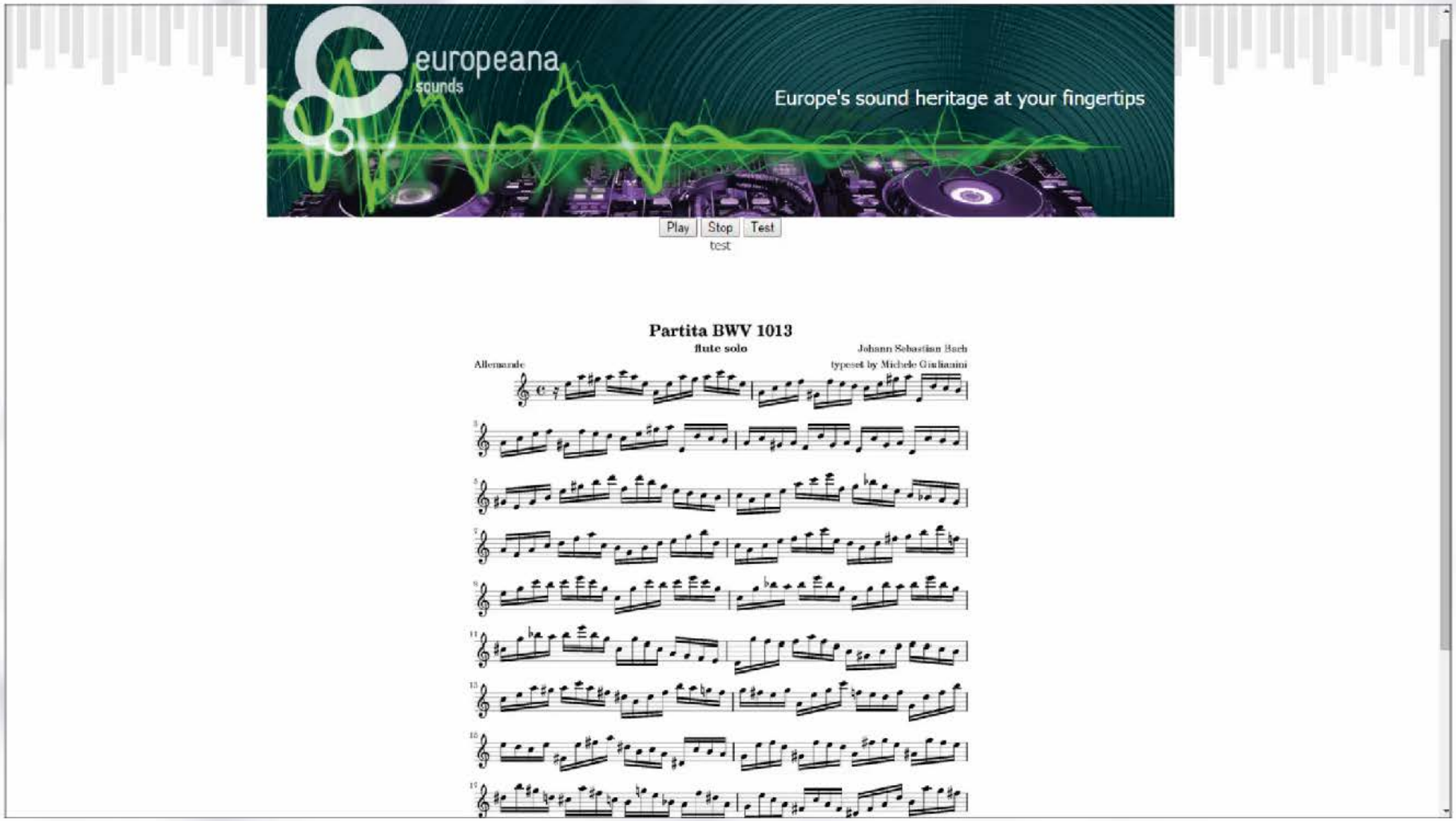
Symbolic Music

Handwritten Music Scores Processing

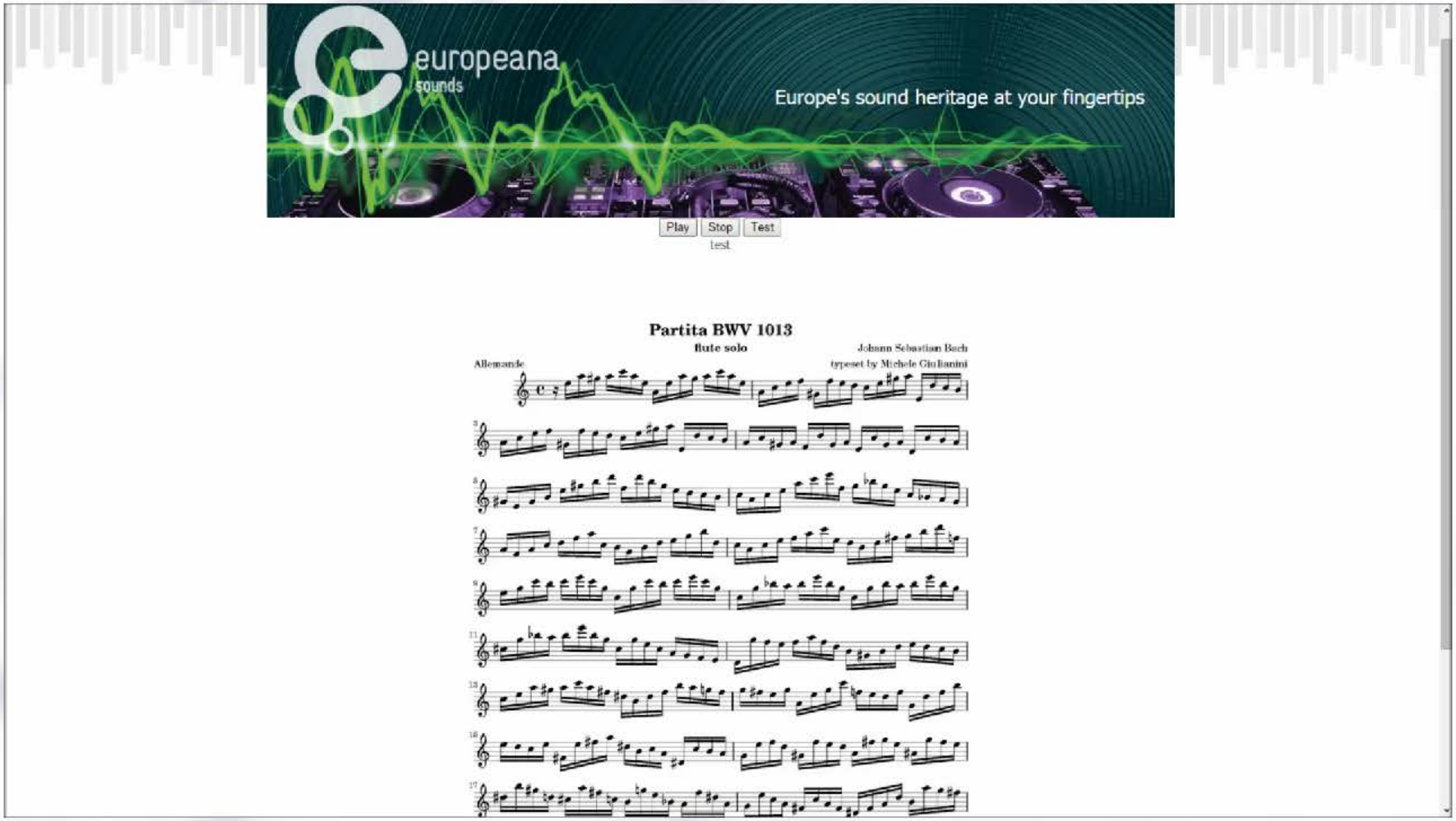
Optical Music Recognition (OMR)

europeana sounds

Score-Following Web Interface



Score-Following Web Interface



Fragen?

Kontaktiert uns:

Alexander Schindler

Next Generation Content Management Systems
Information Management
Austrian Institute of Technology

Alexander.Schindler@ait.ac.at

Thomas Lidy

Institute of Software Technology and Interactive Systems
TU Wien

lidy@ifs.tuwien.ac.at

http://ifs.tuwien.ac.at/mir

AIT AUSTRIAN INSTITUTE OF TECHNOLOGY TOMORROW TODAY

TU WIEN TECHNISCHE UNIVERSITÄT WIEN Vienna University of Technology