

Multimedia Processing Techniques for Retrieving, Extracting, and Accessing Musical Content

PhD Defense

Stefan Balke





International Audio Laboratories Erlangen

Vision



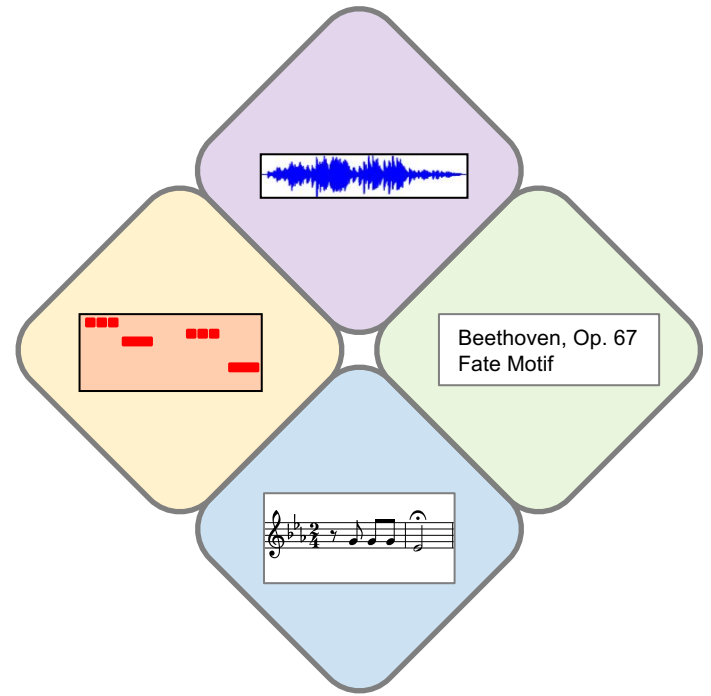
YouTube

Clifford BROWN and Max ROACH
GEORGE MORROW HAROLD LAND RICHELIE POWELL
DELANE
FARSIAN THOROUGHFARE
THE BLUE WALE
DARBOOD
JOY SPRING
JORDU
WHAT ARE WE DOING FOR

Head In     Head Out

Talk Outline

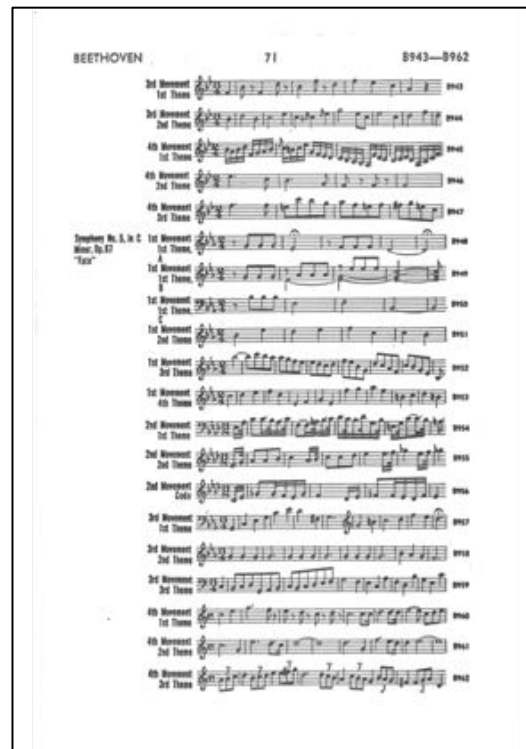
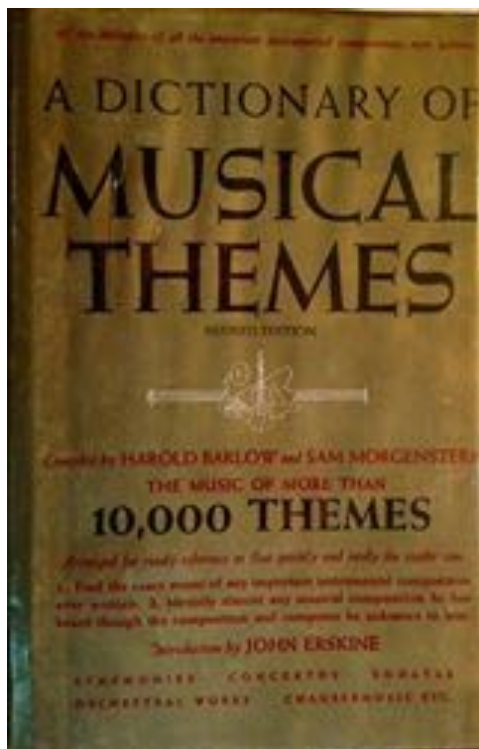
- Retrieval of Musical Themes
- Extraction of Predominant Musical Voices
- Web-Based Technologies for Accessing Musical Content



Retrieval of Musical Themes

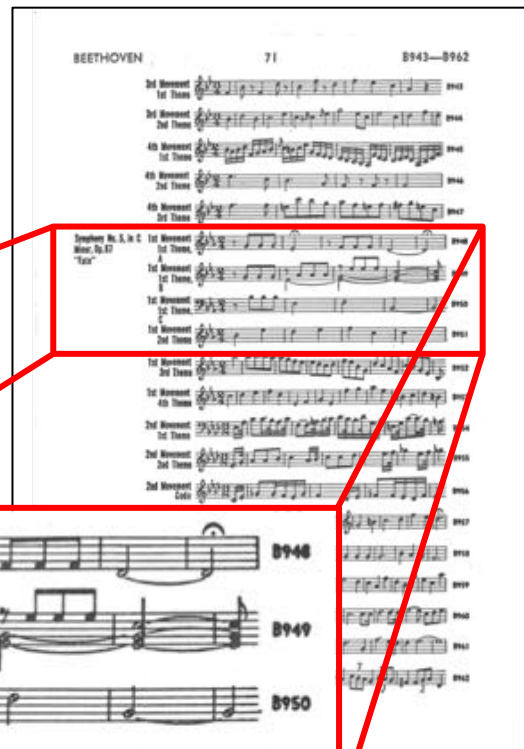
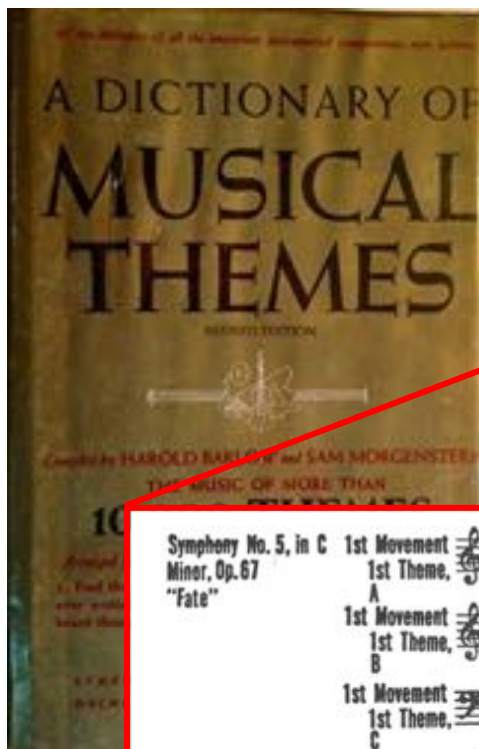
A Dictionary of Musical Themes

Harold Barlow and Sam Morgenstern



A Dictionary of Musical Themes

Harold Barlow and Sam Morgenstern



Symphony No. 5, in C Minor, Op. 67 "Fate"

1st Movement	1st Theme, A	B948
1st Movement	1st Theme, B	B949
1st Movement	1st Theme, C	B950
1st Movement	2nd Theme	B951

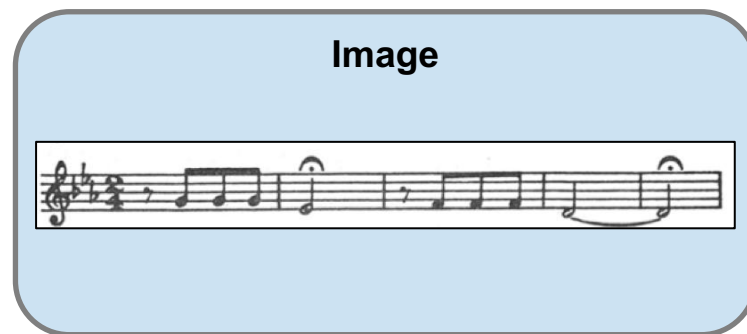


A Dictionary of Musical Themes

Datasets

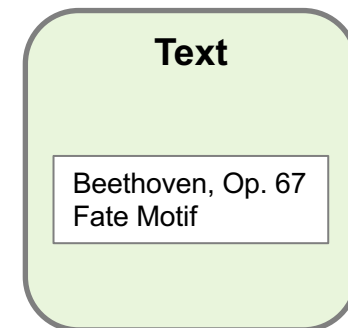
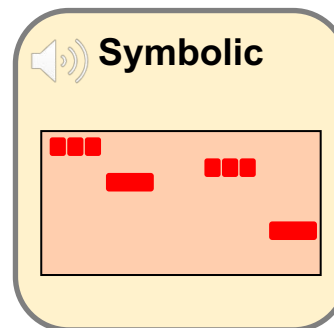
Image

- H. Barlow, S. Morgenstern:
A Dictionary of Musical Themes
- 10,000 Themes from Western classical music



MIDI/Text

- Electronic Dictionary of Musical Themes
- Corresponding MIDI files plus metadata

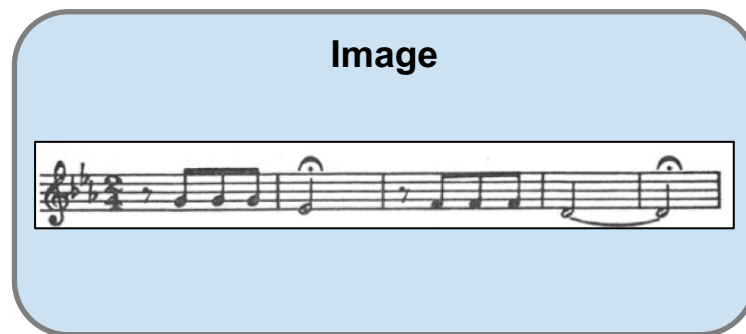


A Dictionary of Musical Themes

Datasets

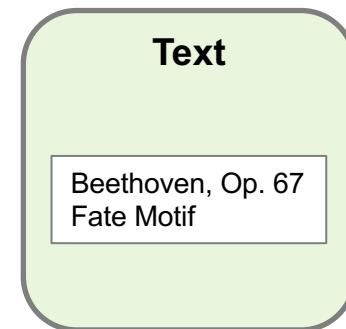
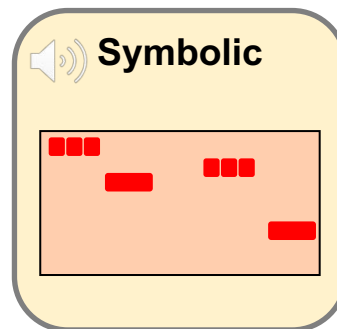
Image

- H. Barlow, S. Morgenstern:
A Dictionary of Musical Themes
- 10,000 Themes from Western classical music



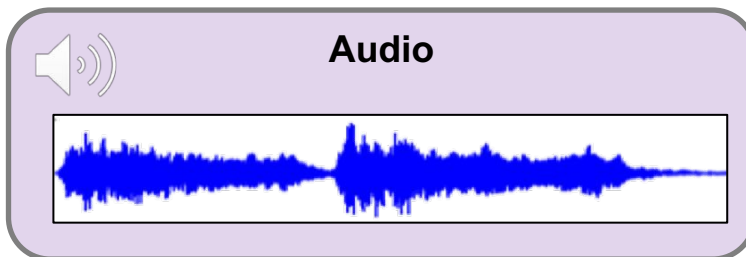
MIDI/Text

- Electronic Dictionary of Musical Themes
- Corresponding MIDI files plus metadata



Audio

- Performances of the musical works

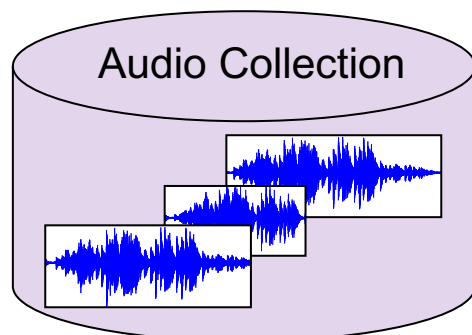
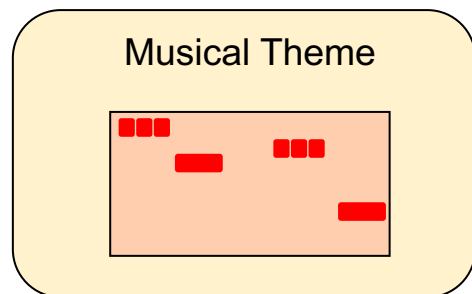


A Dictionary of Musical Themes

Audio-based Retrieval

[Balke16]

- **Query:** Musical theme
- **Goal:** Retrieve audio recording



Challenges

- **Cross-modality**
Symbolic vs. audio data
- **Tuning**
Deviations from standard tuning
- **Transposition**
Played key vs. written key
- **Tempo**
Local & global tempo deviations
- **Polyphony**
Monophonic query vs. polyphonic audio

A Dictionary of Musical Themes

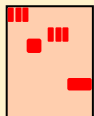
Retrieval Pipeline



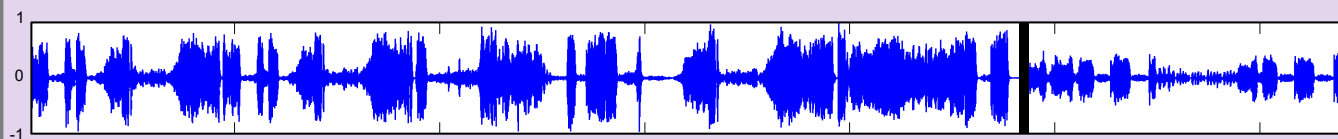
Beethoven, 5th Symphony

Brahms, Hungarian Dance

Musical Theme



Audio Collection



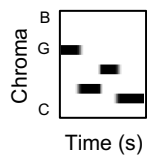
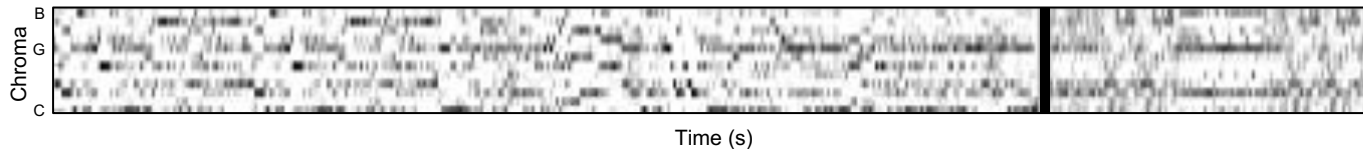
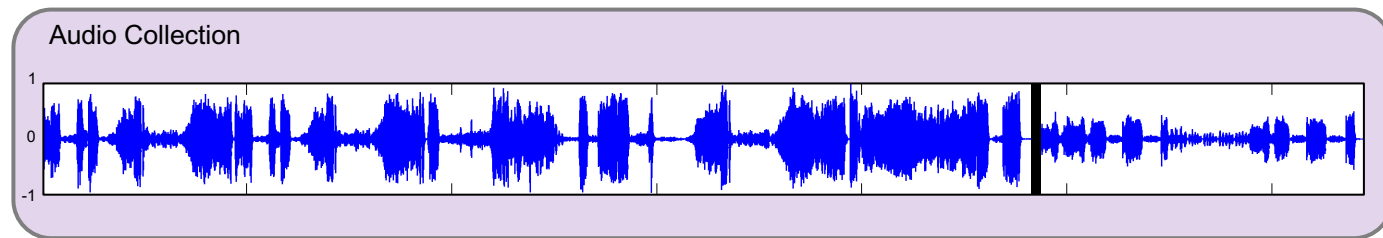
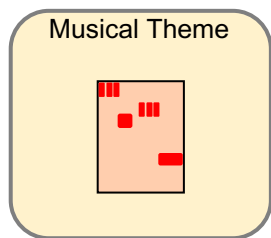
A Dictionary of Musical Themes

Retrieval Pipeline



Beethoven, 5th Symphony

Brahms, Hungarian Dance



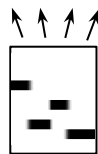
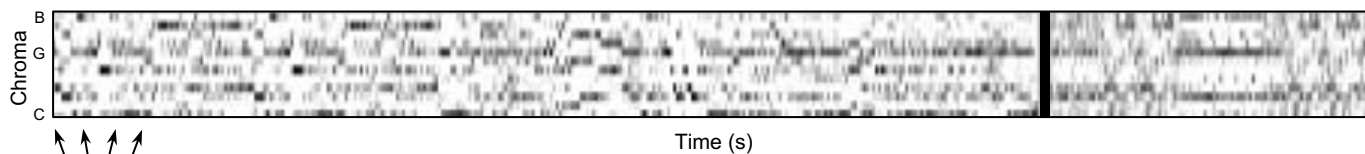
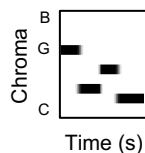
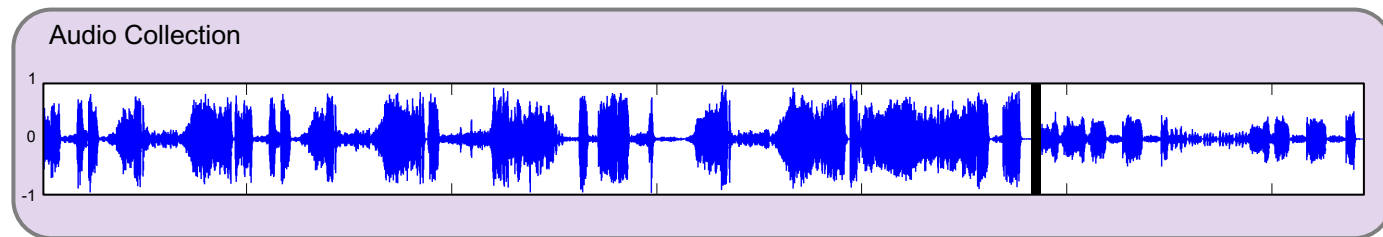
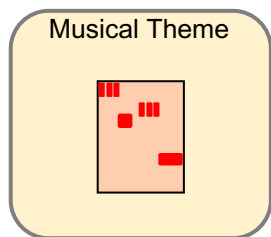
A Dictionary of Musical Themes

Retrieval Pipeline

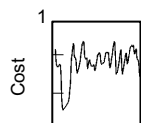


Beethoven, 5th Symphony

Brahms, Hungarian Dance



Matching Function



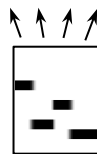
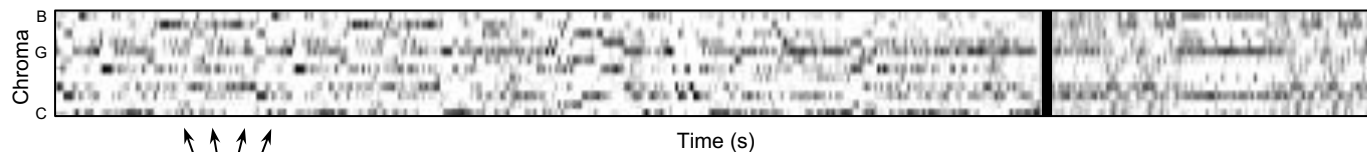
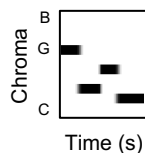
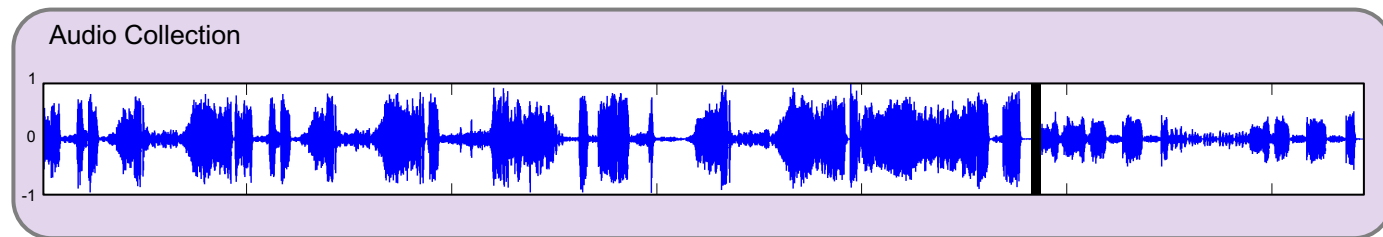
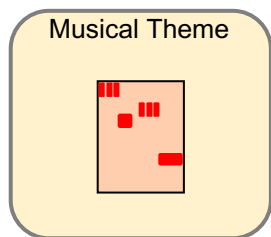
A Dictionary of Musical Themes

Retrieval Pipeline

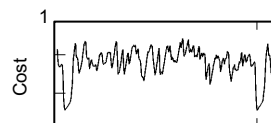


Beethoven, 5th Symphony

Brahms, Hungarian Dance



Matching Function



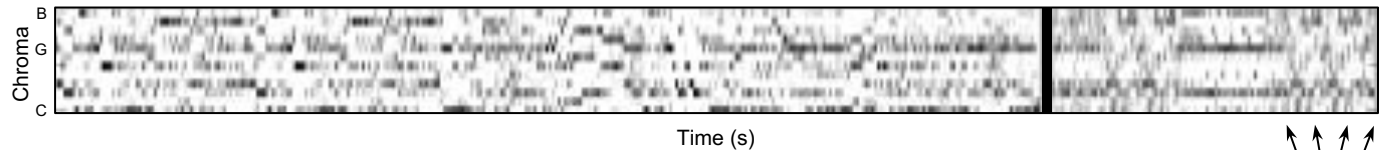
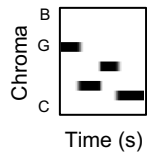
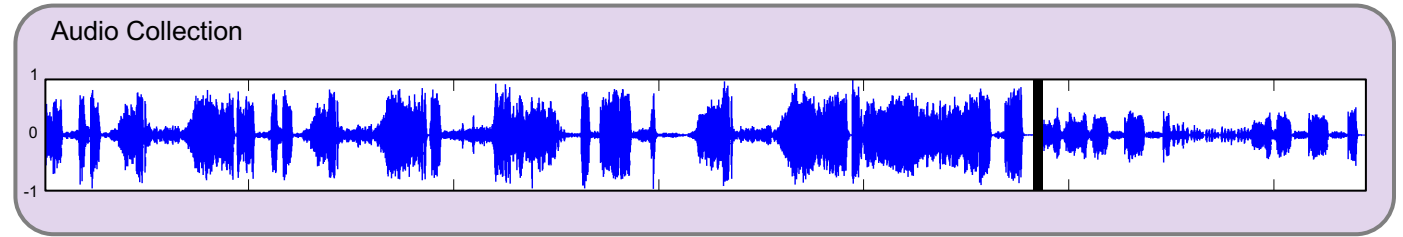
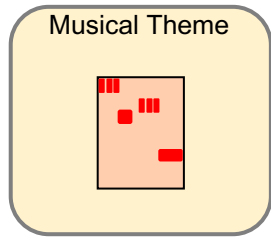
A Dictionary of Musical Themes

Retrieval Pipeline

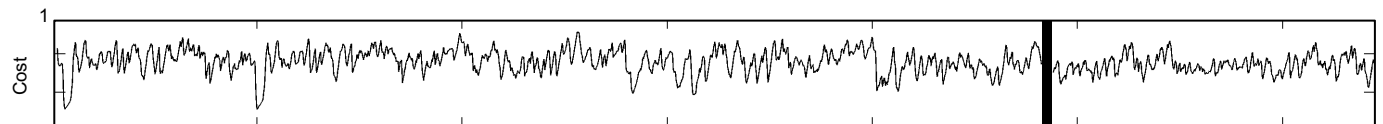


Beethoven, 5th Symphony

Brahms, Hungarian Dance



Matching Function



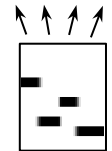
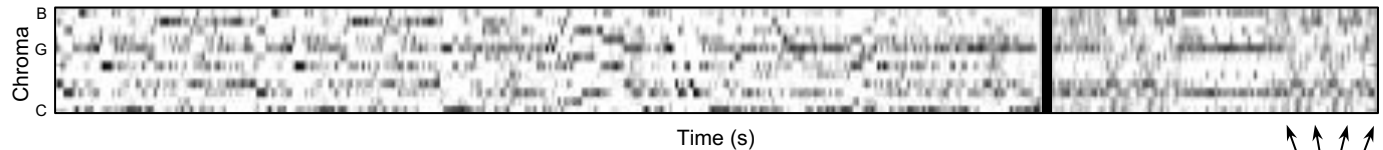
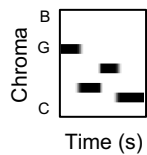
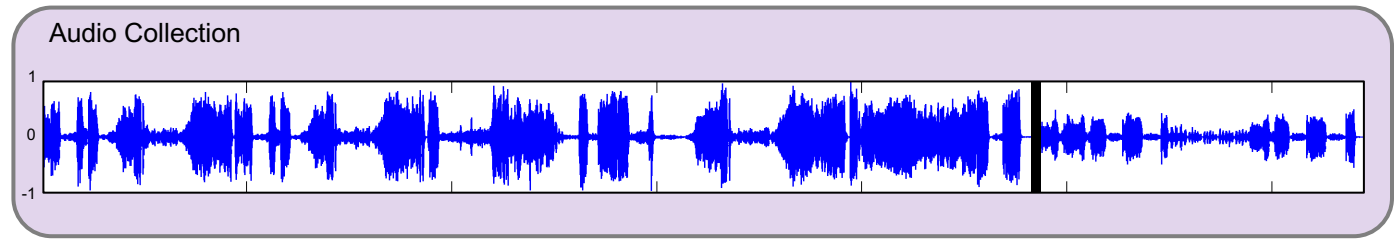
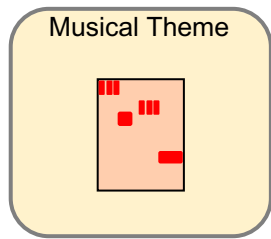
A Dictionary of Musical Themes

Retrieval Pipeline

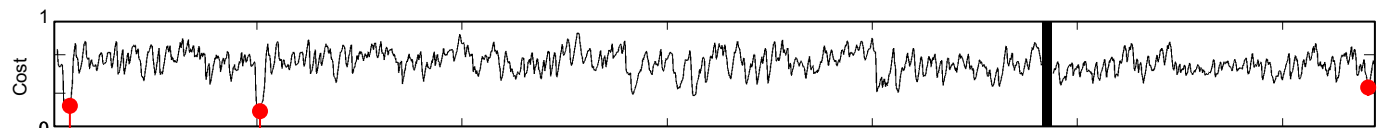


Beethoven, 5th Symphony

Brahms, Hungarian Dance



Matching Function



Original Theme Repetition

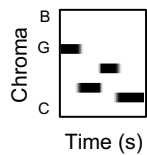
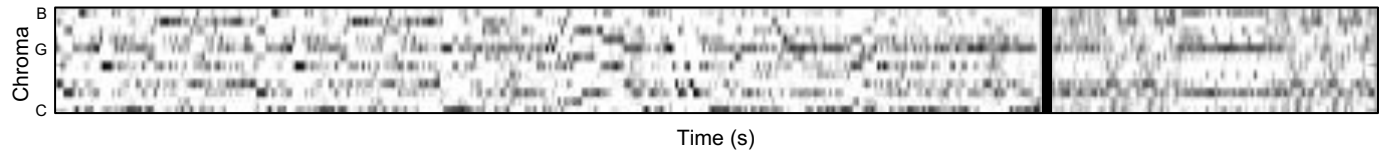
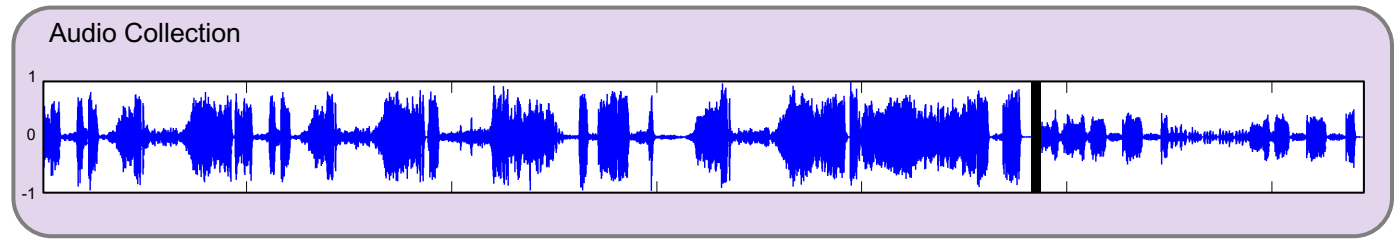
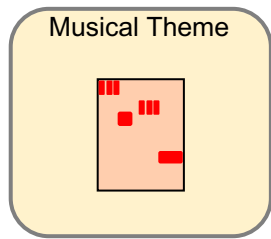
A Dictionary of Musical Themes

Retrieval Pipeline



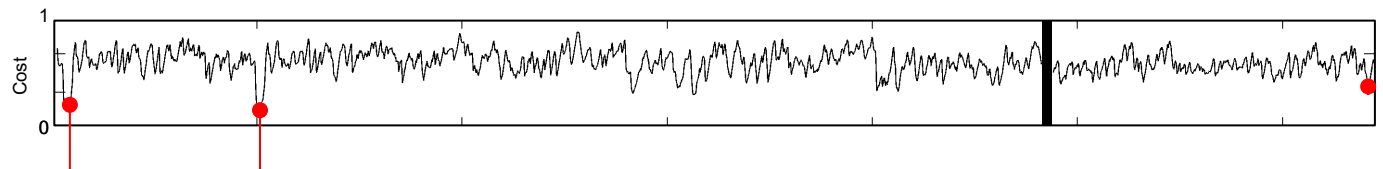
Beethoven, 5th Symphony

Brahms, Hungarian Dance



1. Beethoven Op. 67
2. Brahms, Hungarian Dance

Matching Function



Original Theme Repetition

A Dictionary of Musical Themes

Retrieval Experiment I

#Queries: 177 Themes

#Database: 100 Tracks (~11 h)

A Dictionary of Musical Themes

Retrieval Experiment I

#Queries: 177 Themes

#Database: 100 Tracks (~11 h)

	Top-1	Top-20
Baseline	45.2	76.8

A Dictionary of Musical Themes

Retrieval Experiment I

#Queries: 177 Themes

#Database: 100 Tracks (~11 h)

	Top-1	Top-20
Baseline	45.2	76.8
+ Tuning	46.9	81.9

A Dictionary of Musical Themes

Retrieval Experiment I

#Queries: 177 Themes

#Database: 100 Tracks (~11 h)

	Top-1	Top-20
Baseline	45.2	76.8
+ Tuning	46.9	81.9
+ Transposition	53.7	91.0

A Dictionary of Musical Themes

Retrieval Experiment I

#Queries: 177 Themes

#Database: 100 Tracks (~11 h)

	Top-1	Top-20
Baseline	45.2	76.8
+ Tuning	46.9	81.9
+ Transposition	53.7	91.0
+ Query Length	68.4	93.2

A Dictionary of Musical Themes

Retrieval Experiment II

#Queries: 2046 Themes

#Database: 1113 (~120 h)

A Dictionary of Musical Themes

Retrieval Experiment II

#Queries: 2046 Themes

#Database: 1113 (~120 h)

	Top-1	Top-20	Top-50
Tuning + 10 s	18.3	29.2	46.1

A Dictionary of Musical Themes

Retrieval Experiment II

#Queries: 2046 Themes

#Database: 1113 (~120 h)

	Top-1	Top-20	Top-50
Tuning + 10 s	18.3	29.2	46.1
Transp. + Query Length *)	39.5	66.9	76.1

*) Results from a recent study together with Frank Zalkow.

A Dictionary of Musical Themes

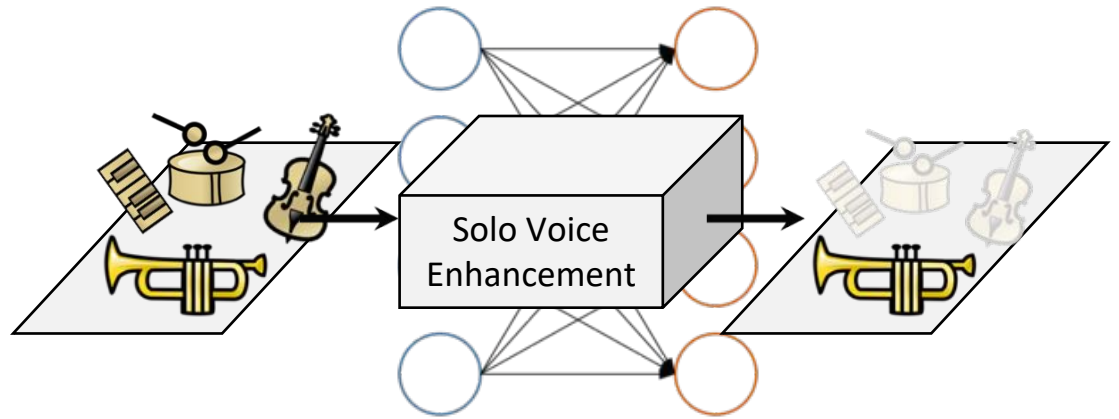
Retrieval Experiment II

#Queries: 2046 Themes

#Database: 1113 (~120 h)

	Top-1	Top-20	Top-50
Tuning + 10 s	18.3	29.2	46.1
Transp. + Query Length *)	39.5	66.9	76.1
+ Predominant Melody *)	61.2	81.8	86.7

*) Results from a recent study together with Frank Zalkow.



Extraction of Predominant Musical Voices

Extraction of Predominant Musical Voices

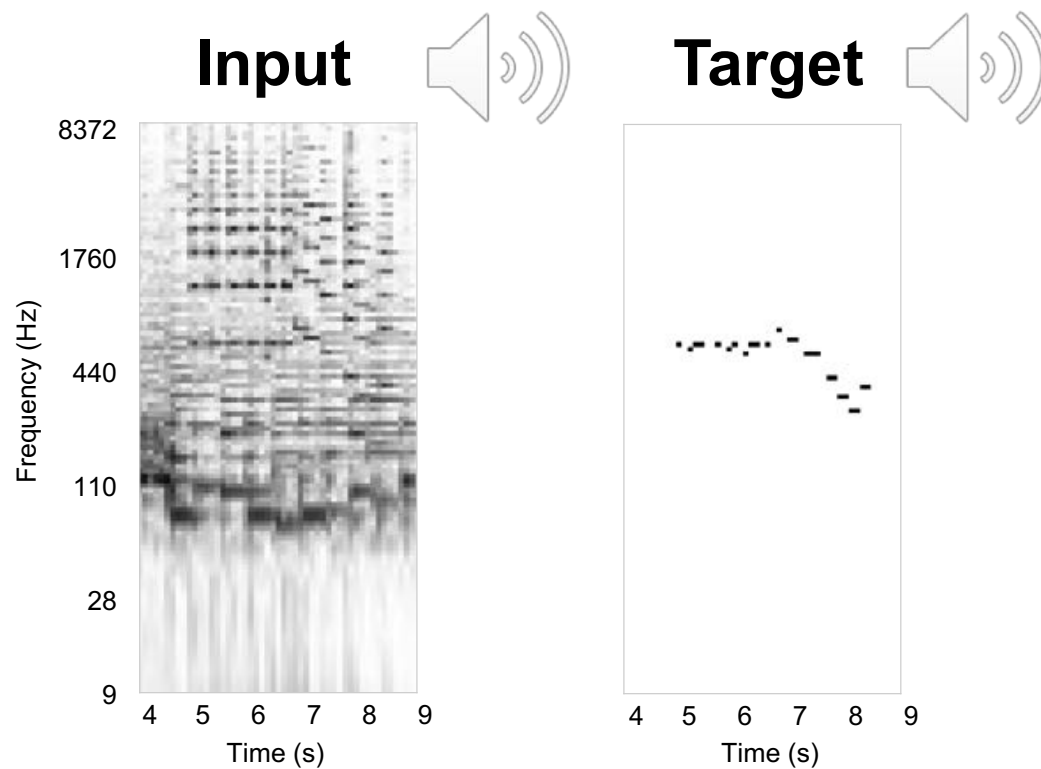
Predominant Melody Extraction

1. Model-based approach [Salamon13, Bosch16]
2. Data-driven approach [Bittner15, Rigaud16, Bittner17]

Our Data-Driven Approach [Balke17]

Estimate “monophonic” time-frequency representation from a “polyphonic” audio recording using Deep Neural Networks (DNNs).

DNN Training



Dataset

Weimar Jazz Database (WJD)



- 299 transcribed jazz solos of monophonic instruments
- ca. 10 h of annotated music

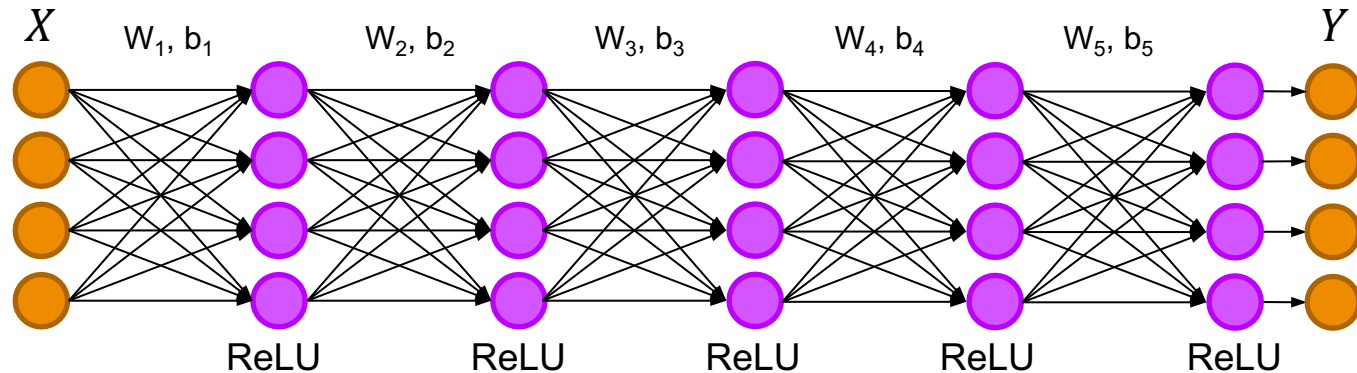
Thanks to the Jazzomat research group: M. Pfeiderer, K. Frieler, J. Abeßer, and W.-G. Zaddach

DNN Architecture

[Balke17]

$X := \text{Input}$, $Y := \text{Output}$, $T := \text{Target}$, $L := \text{Loss}$

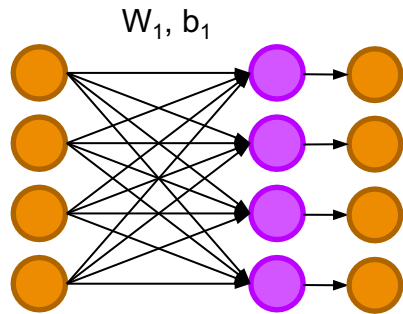
$L = \text{MSE}(X, Y)$



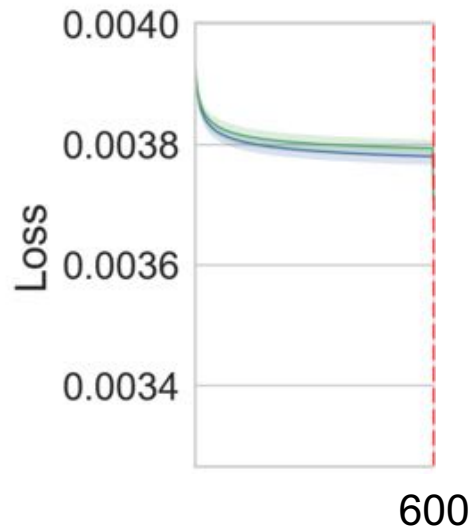
Dimensions: 120 120 120 120 120 120 120

- Basic DNN with 5 fully-connected layers.
- Training is applied layer-wise [Bengio06, Uhlich15].

Layer-Wise Training



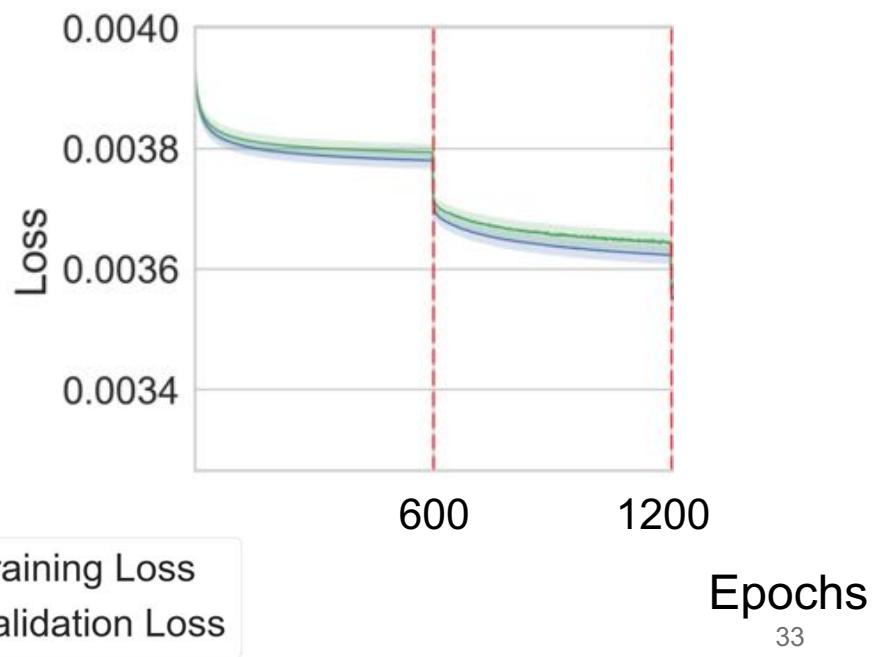
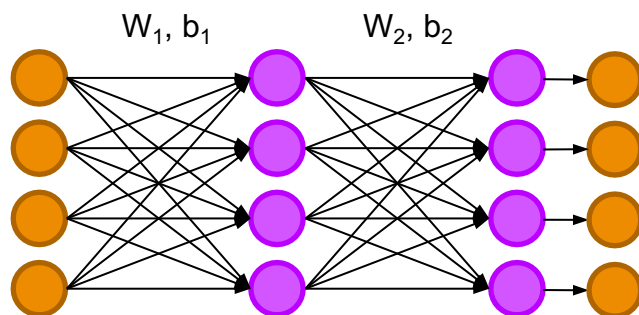
- Initialize weights and bias with Linear Least Squares (LLS)
- Train 600 epochs ...



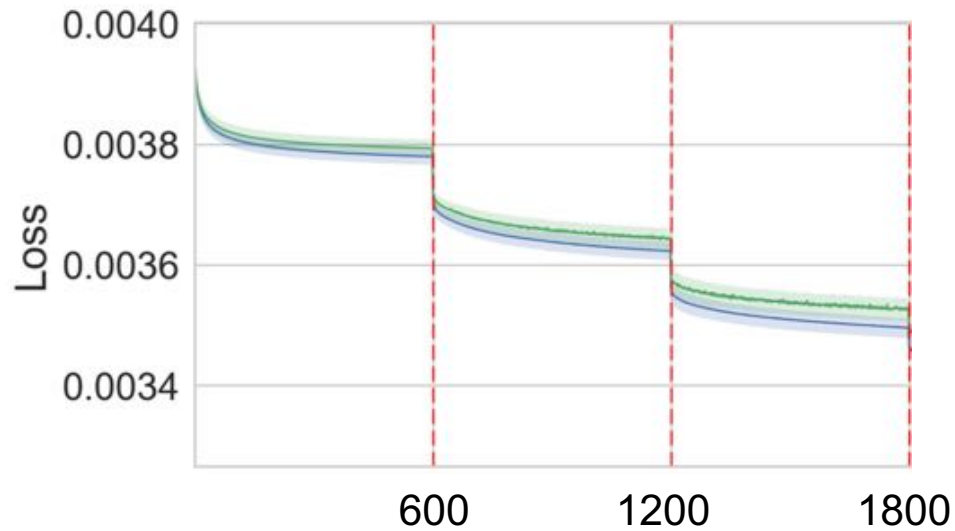
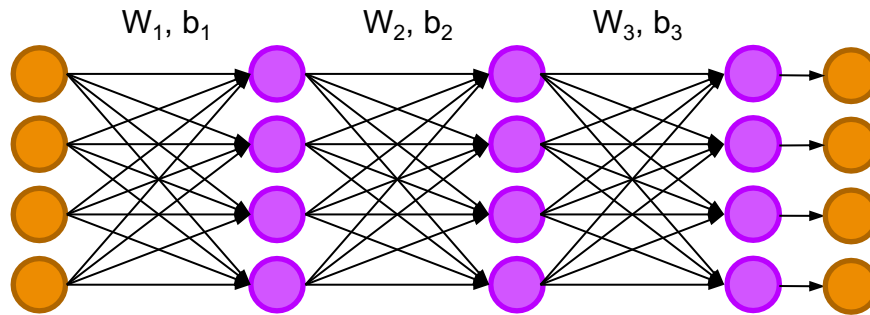
— Training Loss
— Validation Loss

Epochs

Layer-Wise Training



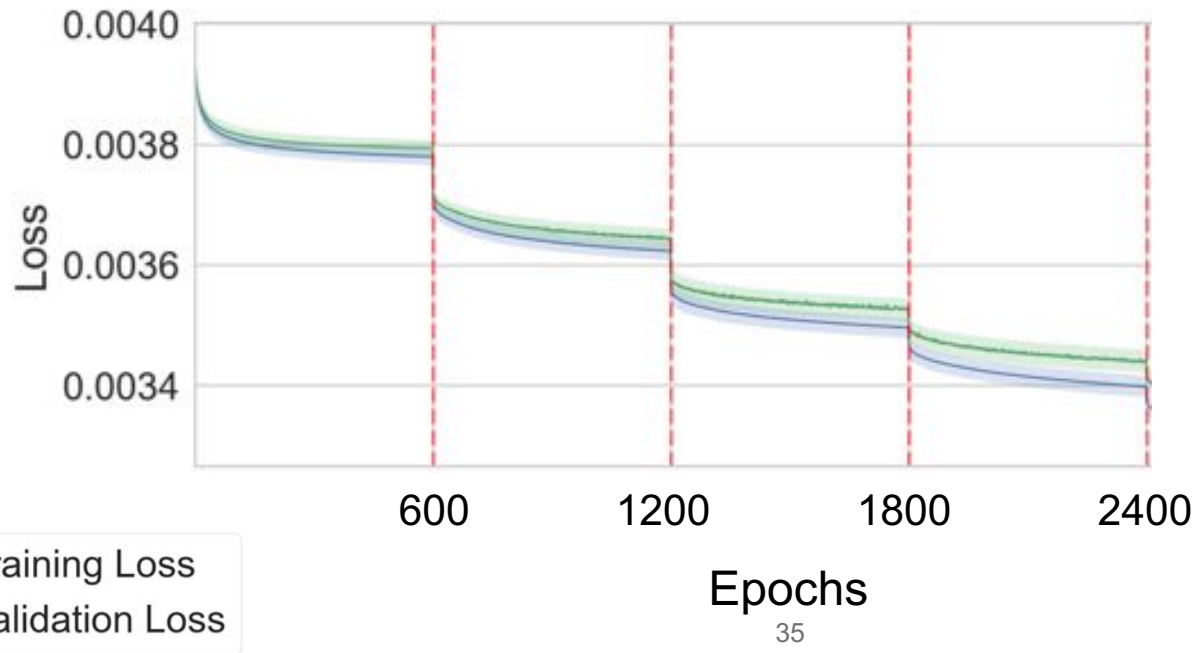
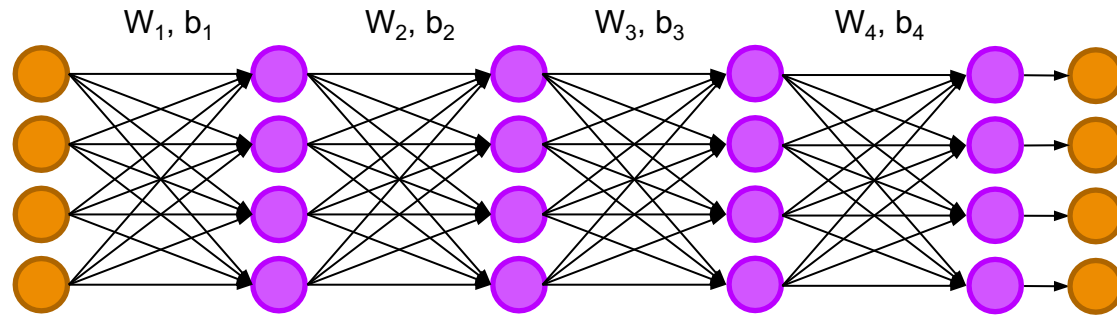
Layer-Wise Training



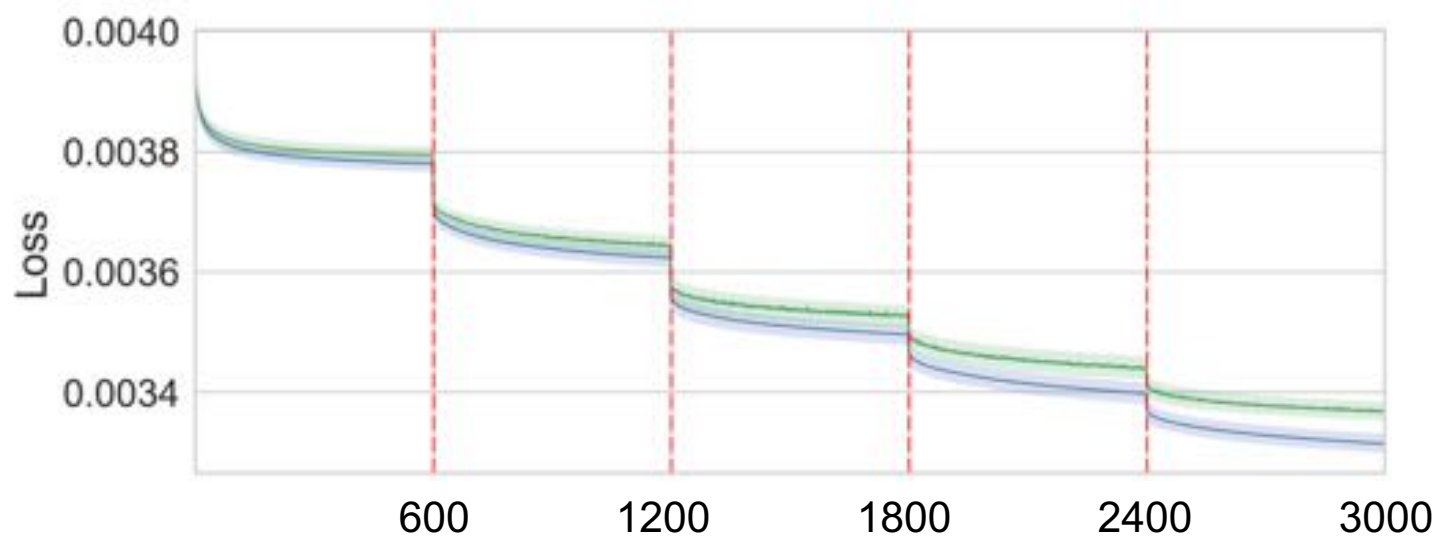
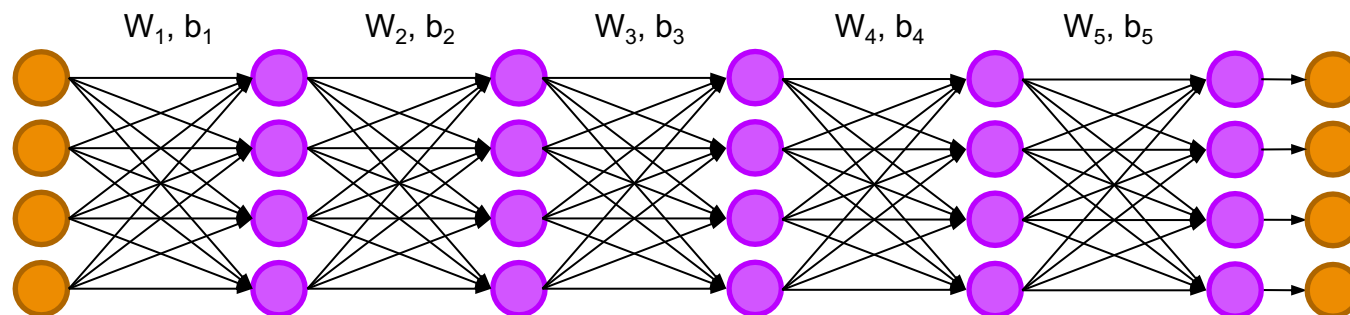
— Training Loss
— Validation Loss

Epochs

Layer-Wise Training



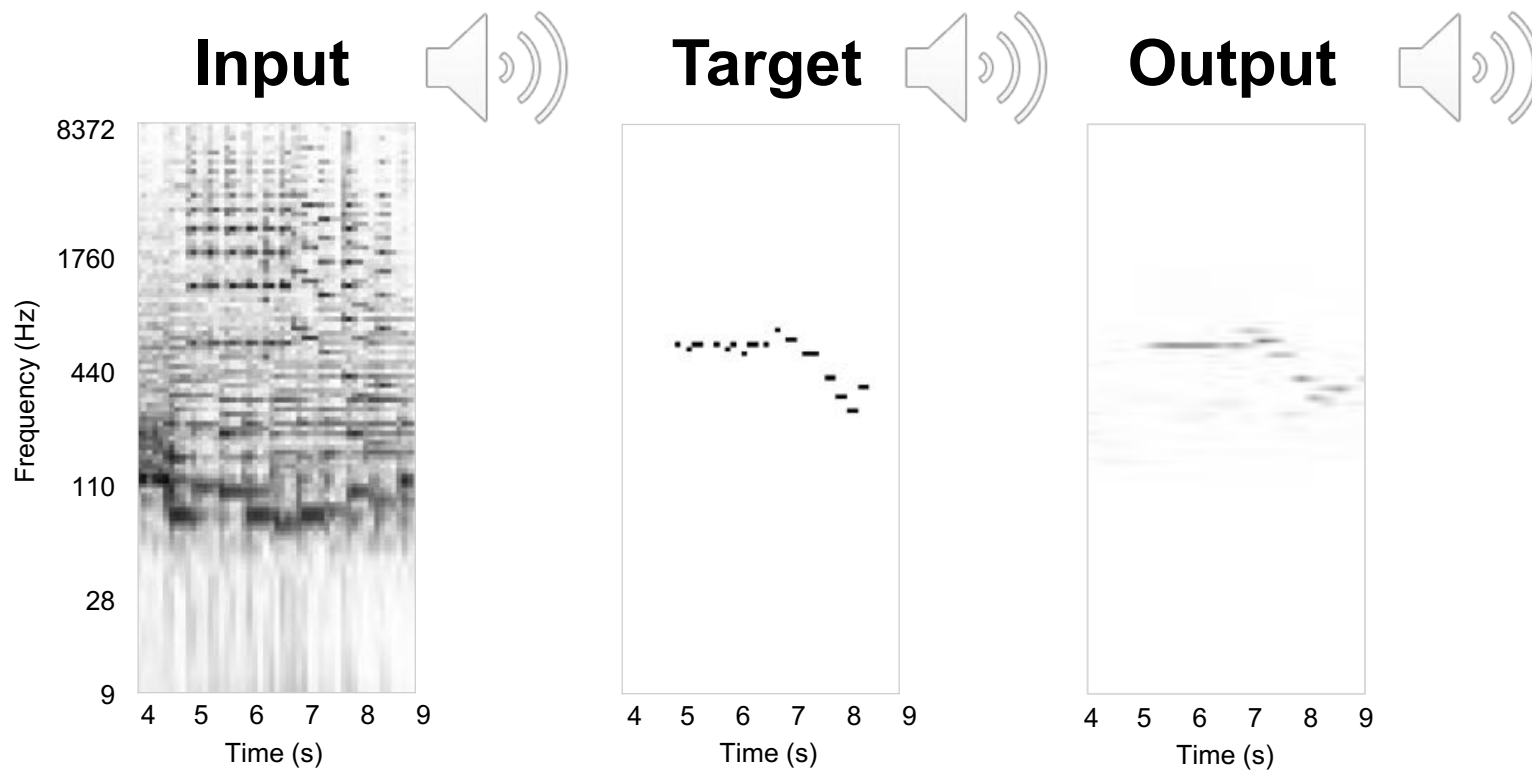
Layer-Wise Training



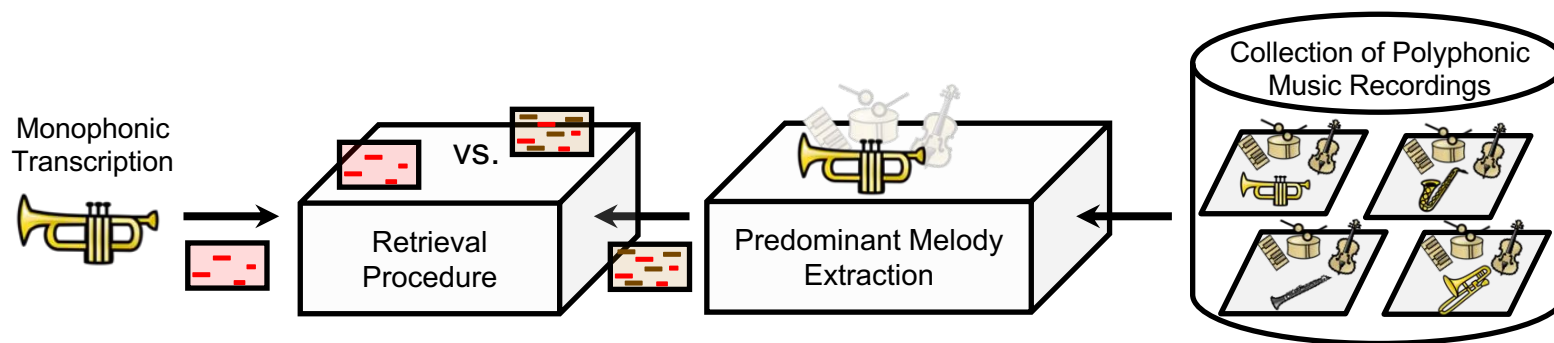
— Training Loss
— Validation Loss

Epochs

Qualitative Evaluation



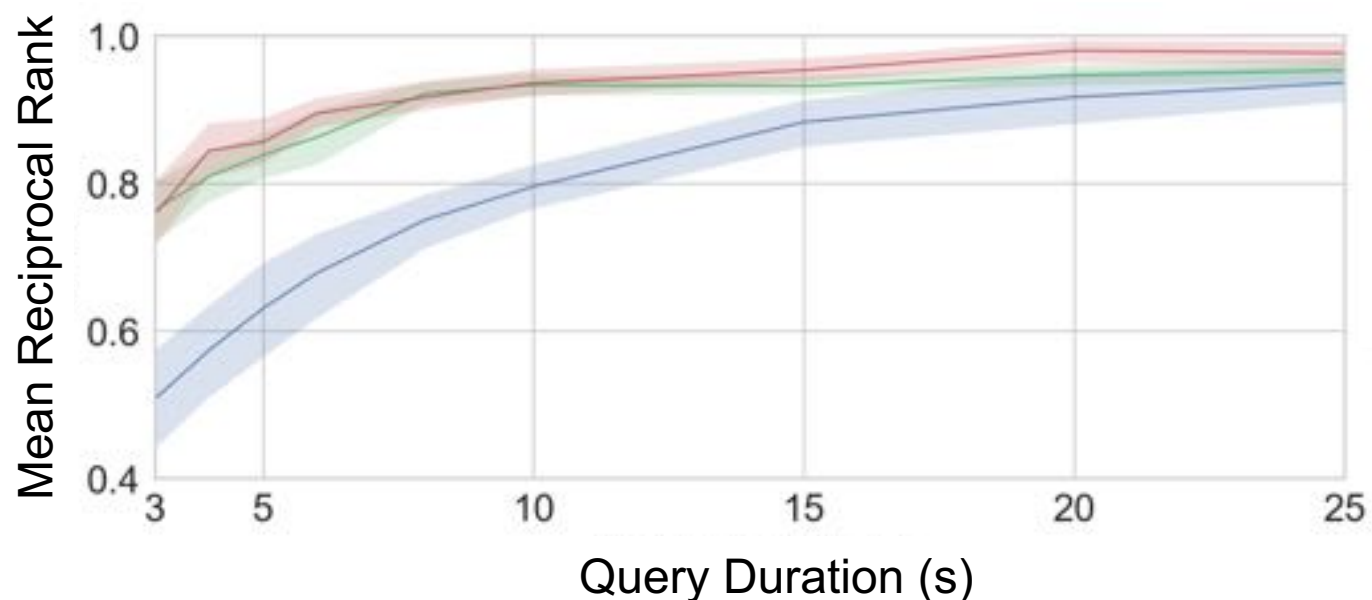
Experiment: Jazz Music Retrieval

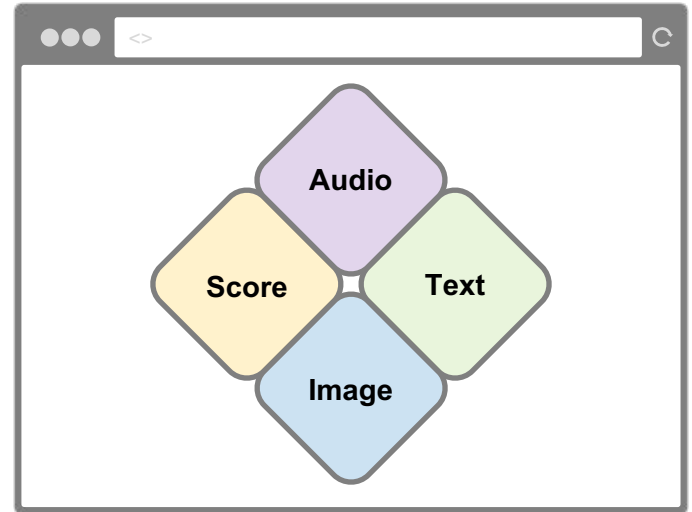


Experiment: Jazz Music Retrieval

Results

- Baseline** Chroma-based matching [Mueller15]
- Melodia** Quantized F0-trajectory [Salamon13]
- DNN**





Web-Based Technologies for Accessing Musical Content

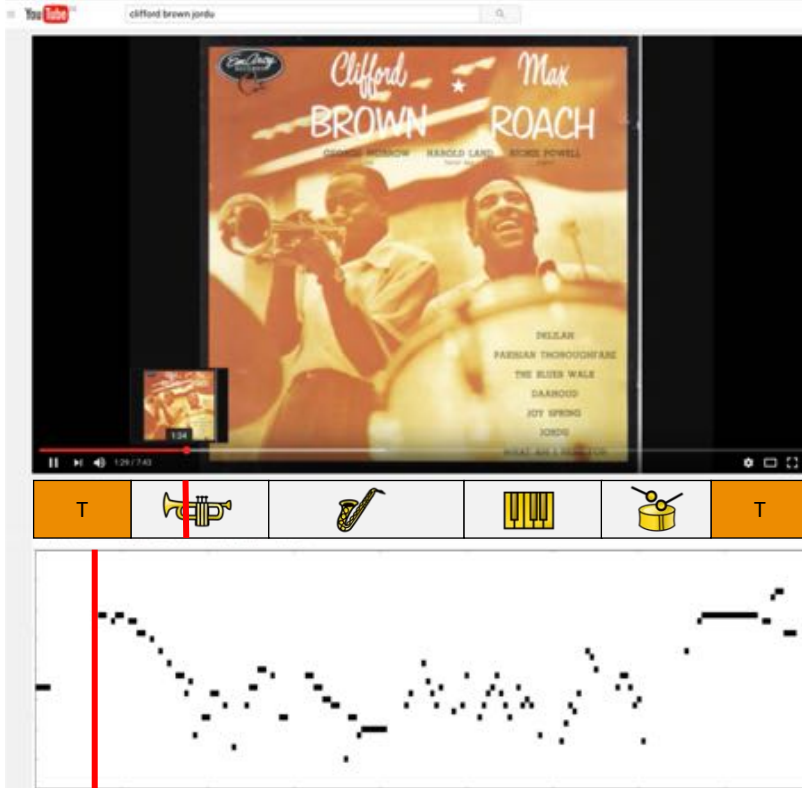
Technologies for Accessing Musical Content

[Balke18]

The image shows a YouTube video player interface. The video title is "Clifford Brown and Max Roach". The video player includes a progress bar, a volume icon, and a search bar. Below the video player is a control bar with icons for different instruments: a trumpet, a saxophone, a piano, and a drum set. A red vertical line is positioned over the trumpet icon. Below the control bar is a waveform visualization of the audio content, with a red vertical line indicating the current playback position.

Technologies for Accessing Musical Content

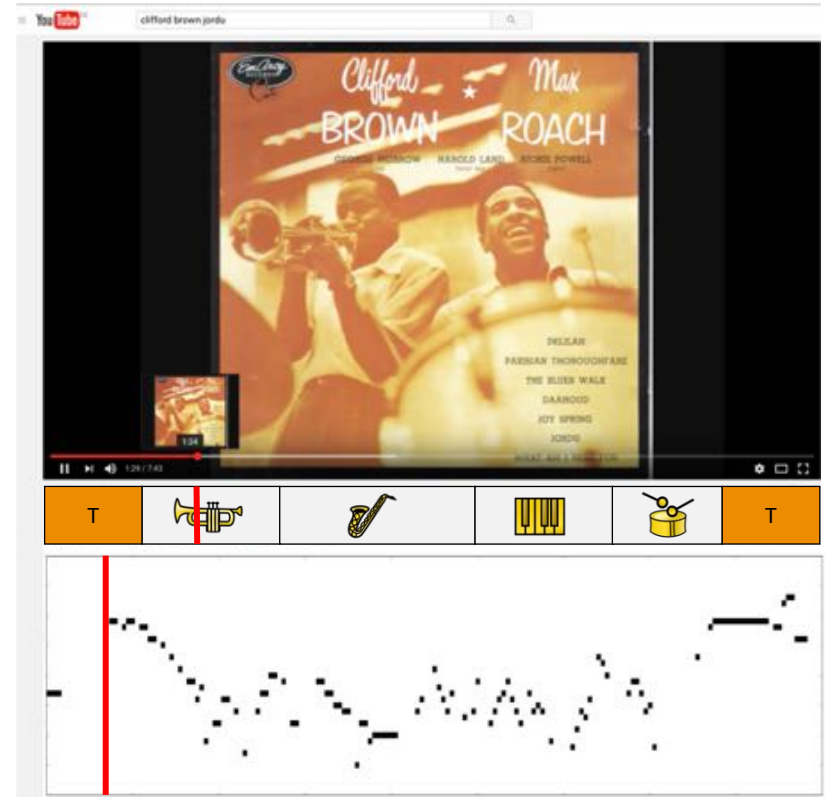
[Balke18]



YouTube video player interface showing a video titled "Clifford BROWN Max ROACH". The video player includes a search bar, a play button, and a progress bar. Below the video player is a control bar with icons for different instruments: a trumpet (highlighted with a red vertical line), a saxophone, a piano, and a drum set. Below the control bar is a waveform visualization of the audio content.

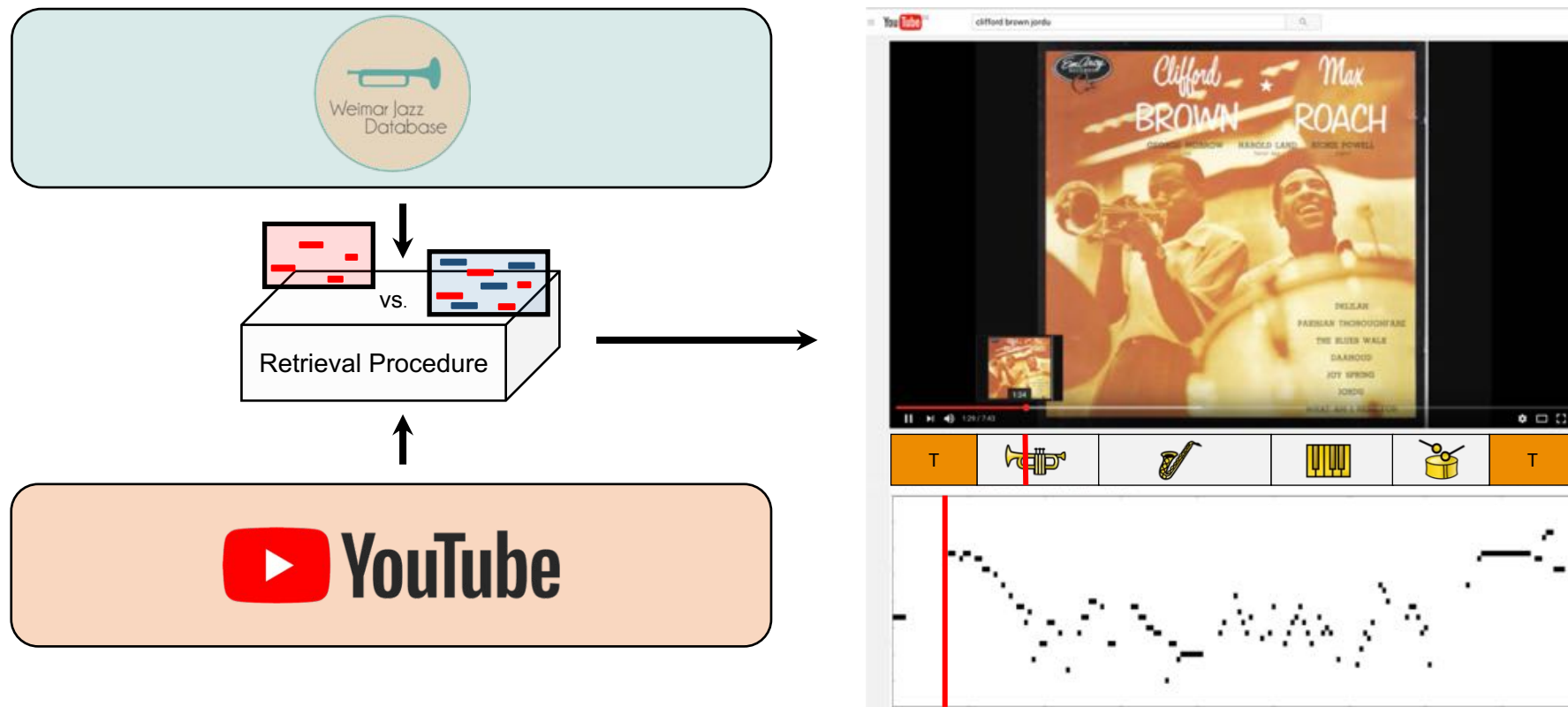
Technologies for Accessing Musical Content

[Balke18]



Technologies for Accessing Musical Content

[Balke18]



The screenshot shows a web browser window with the address bar displaying `mir.audiolabs.uni-erlangen.de`. The browser's navigation bar includes back, forward, and refresh buttons. The website's navigation menu contains links for `Solos`, `Soloists`, `Compositions`, `Tracks`, `Records`, `Statistics`, `Downloads`, and `About`.

The main content area features a large heading: **Welcome to JazzTube**, where the 'J' is stylized with a musical note. Below this is a grey box with the text: **Jazz + YouTube = JazzTube**. The text inside the box reads: "JazzTube lets you explore the annotations from the [Weimar Jazz Database \(WJD\)](#) in intuitive and interactive ways. Experience an enriched listening experience by exploring famous jazz recordings along with solo transcriptions and further information about the featured soloists. Get started by checking out [Clifford Brown's solo on Jordis!](#)"

Below the grey box is a section titled **JazzTube features...** with a bulleted list:

- ...456 Solos
- ...343 Tracks
- ...197 Records
- ...307 Compositions
- ...more statistical

The next section is titled **Acknowledgments** and contains the following text:

JazzTube is publicly funded research (DFG MU 2586/9-1 and DFG PF 669/7-1) and serves as an accompanying demonstrator for the article mentioned below.

If you are interested in technical details or want to refer to this work in your publications, please use the following citation:

Stefan Balke, Christian Dittmar, Jakob Abeller, Klaus Frieier, Martin Pfeiderer, and Meinard Müller
Bridging the Gap: Enriching YouTube Videos with Jazz Music Annotations
 Under review, 2017.

At the bottom left, there is a copyright notice: © 2017. At the bottom right, there are two logos: **jazzomat** and **AUDIO LABS**.

