Bern University
of Applied Sciences

AUDIO SEGMENTATION OF OPERA RECORDINGS

"7th Swiss Open Cultural Data Hackathon"



13. April 2021

TEAM



Prof. Dr. Eduard Klein

Senior Lecturer & Researcher in Software Engineering and Web Technologies

@ BFH



Michelle Schmid

Bachelor Thesis Student

@ BFH



Christian Stuber

Master Thesis Student

@ BFH

AGENDA

- Background
  - Vision
  - Project „Opera Forever“
  - State of Project
- Practical Implementation
  - Dataset „The Magic Flute“
  - Procedure Automated Segmentation
  - Feature Extraction
  - Results
  - Challenges
- Discussion







FEEDBACK CHANNEL

Feel free to add posts, ideas, questions, challenges …

https://padlet.com/kee14/5uakytncle7ffsje

BACKGROUND

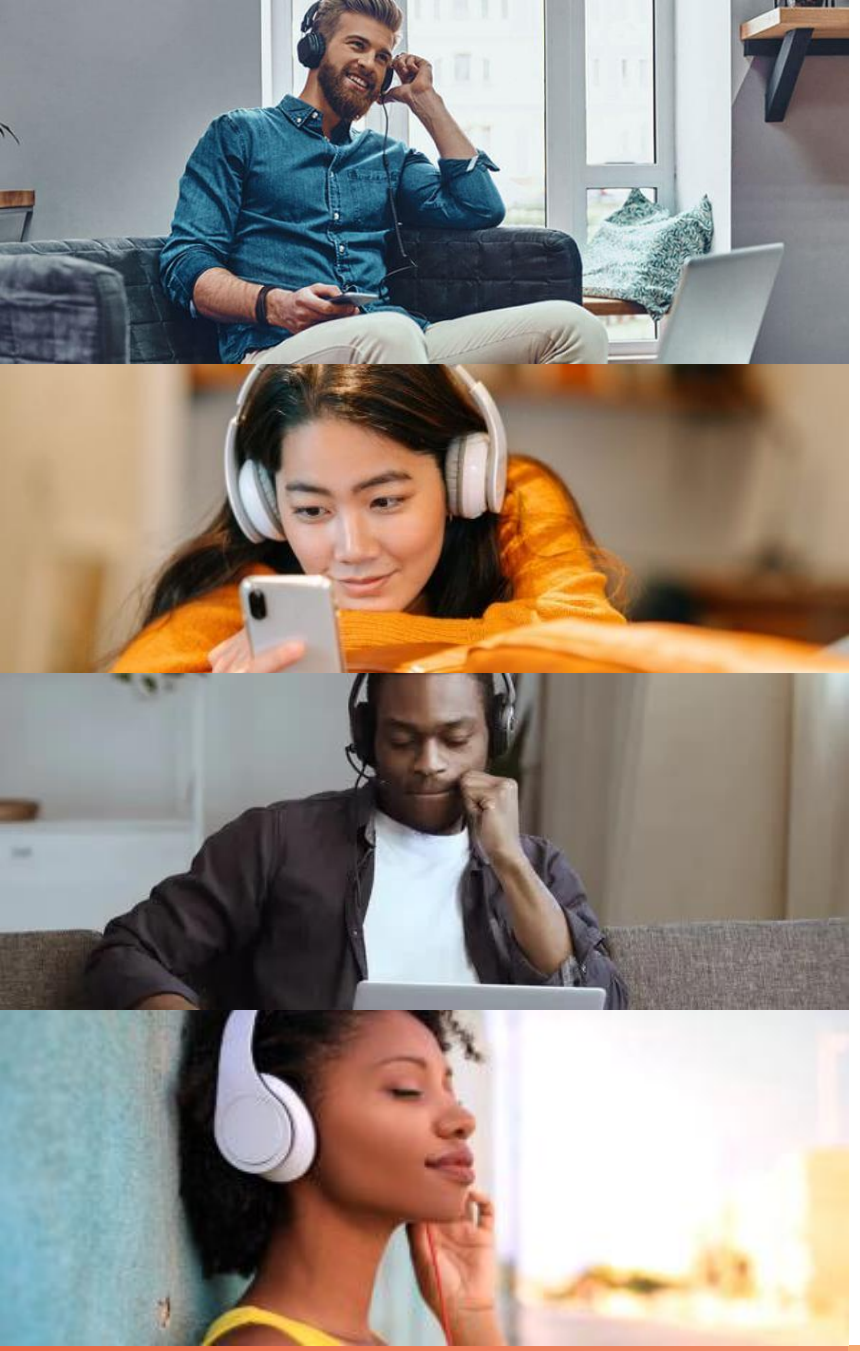
Vision

Project „Opera Forever“

State of Project







VISION

Value-Added Application for Music Listeners, Music Enthusiasts, Music Experts

- More than listening → a cultural and historical experience
- (Classical) Music Companion, browsing through a music collection resp. Segments
- Demo from GLAM mix'n'hack 2019: https://opera.now.sh

Not just music experience, but browsing related artefacts

- score, event tickets, booklets, …
- linking to web ressources

"social" features: rating, commenting, ranking, community, …

PROJECT „OPERA FOREVER“

Available collection of bootleg recordings ("Ehrenreich Collection", https://ehrenreich.bfh.science)

- 10'000 hours of music, mainly operas, recorded between 1965 – 2010

Goal: semantic enrichment of audio objects as "data base" for value-added applications

- Main challenge: splitting audio objects into segments, thus enabling browsing on segment level (example: Mozart's "Magic Flute" has 59 segments)

Manual segmentation process is time-consuming, tedious and cumbersome

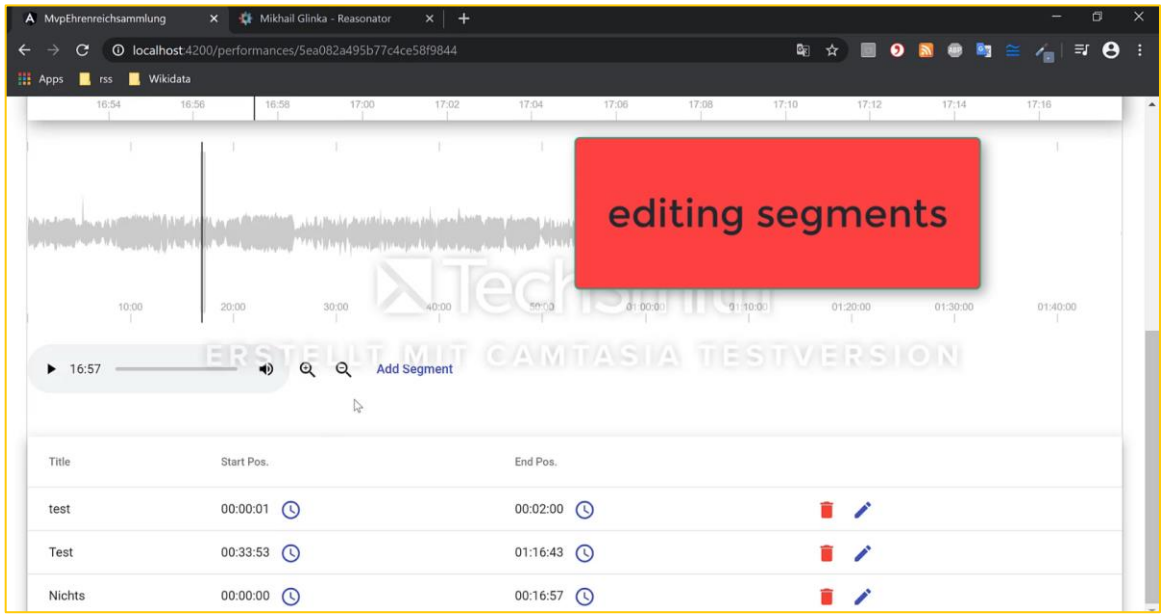
- ~ 1 hour for a 3h Opera, if reference segmentation is available)

Automation of segmentation process is crucial!

- What methods and to what extent is it possible to automize the segmentation process for the Ehrenreich Collection? (future question: for any collection?)

STATE OF THE PROJECT @BFH

- Master Thesis ([Hentze, Johanna 2020]) about a Segmentation Tool including links to related artefacts (e.g. to Wikipedia):
  "Einbindung und Nutzung von Kulturdaten in Wikidata im Zusammenhang mit der Ehrenreich-Sammlung"
- Case Study ([Schmid, Michelle 2021]) on building an experimental environment for audio analysis based on python/librosa library. Influenced by the project "Freischütz Digital" (https://freischuetz-digital.de)





ONGOING:

"Optimizing feature representations for automated segmentation within the Ehrenreich collection"
(MICHELLE SCHMID, BACHELOR THESIS)

"Automatisierte Erschliessung von Musiksammlungen"
(CHRISTIAN STUBER, MASTER THESIS)

PRACTICAL IMPLEMENTATION

Dataset „The Magic Flute"

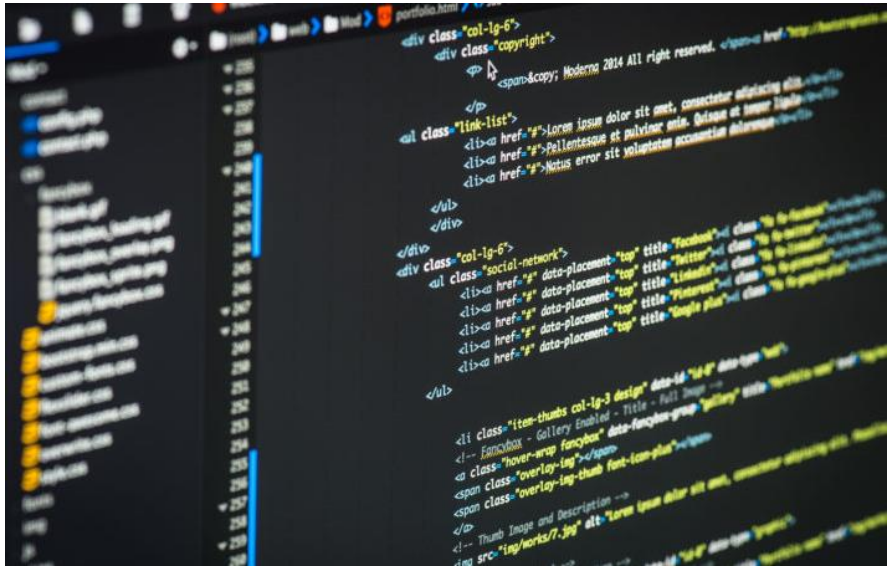
Procedure Automated Segmentation

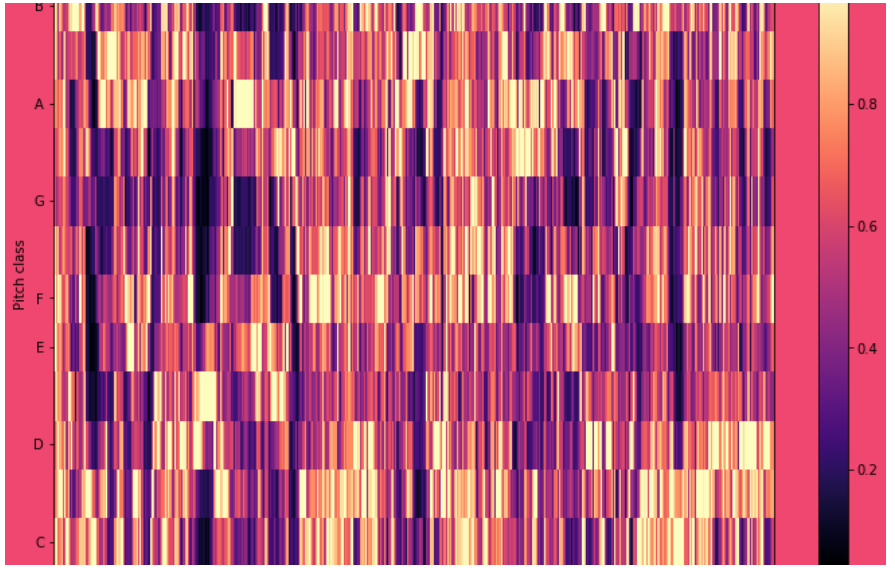
Feature Extraction

Dynamic Time Warping

Results

Challenges





DATASET „THE MAGIC FLUTE“



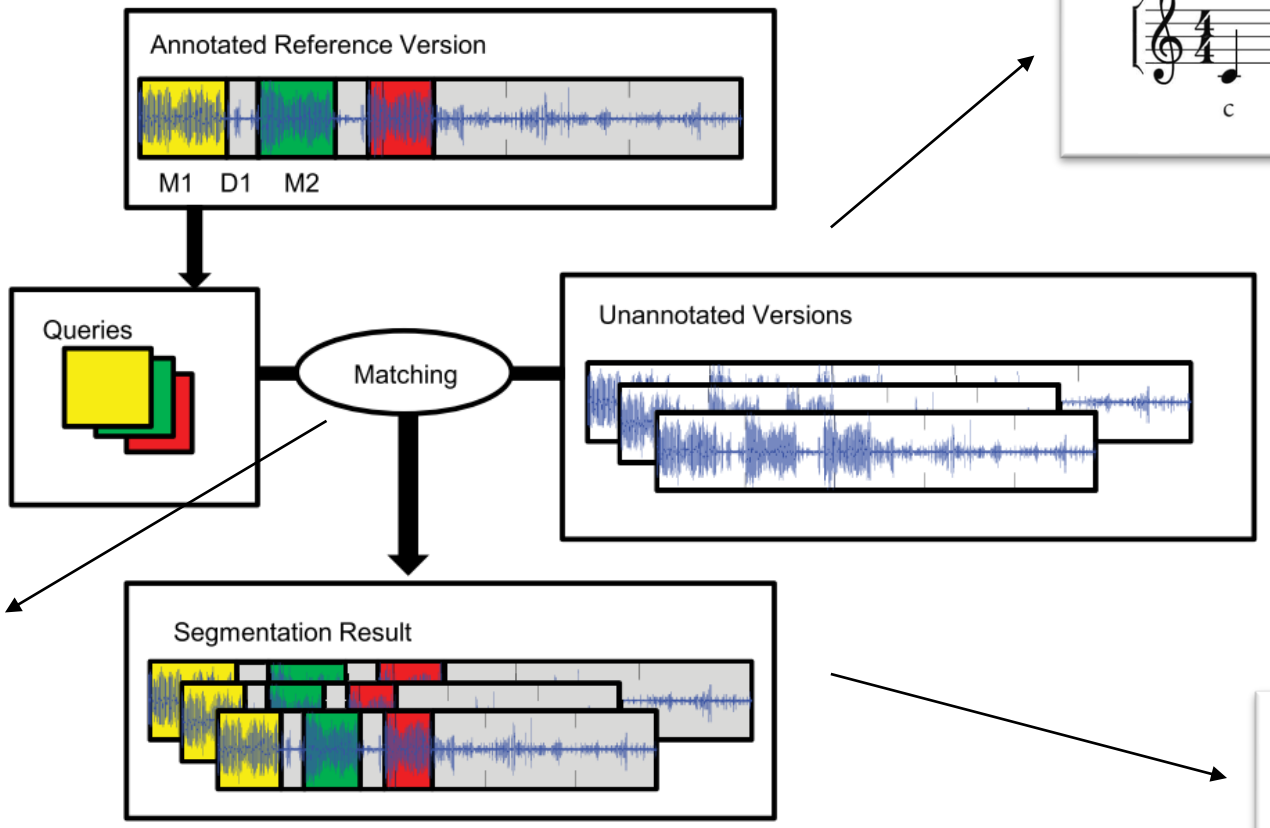
Segment types:

- instrumental
- speech
- singing
- example aria «queen of the night»

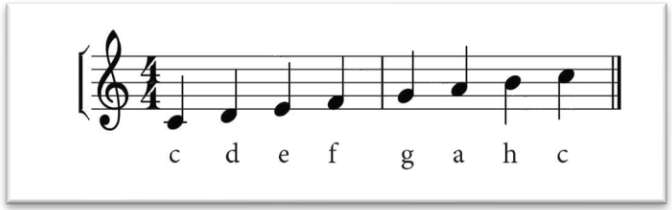
Spaces (scene changes, applaus…)

PROCEDURE „AUTOMATED SEGMENTATION“

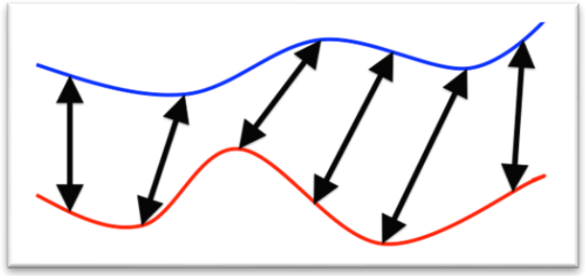
Librosa



Feature extraction (chroma)



Dynamic Time Warping



Relative overlap

$$\frac{|\alpha \cap \beta|}{|\alpha \cup \beta|} \in [0, 1]$$

Source: Prätzlich, Müller (2013) - A Case Study for Reference-Based Audio Segmentation of Operas

FEATURE EXTRACTION

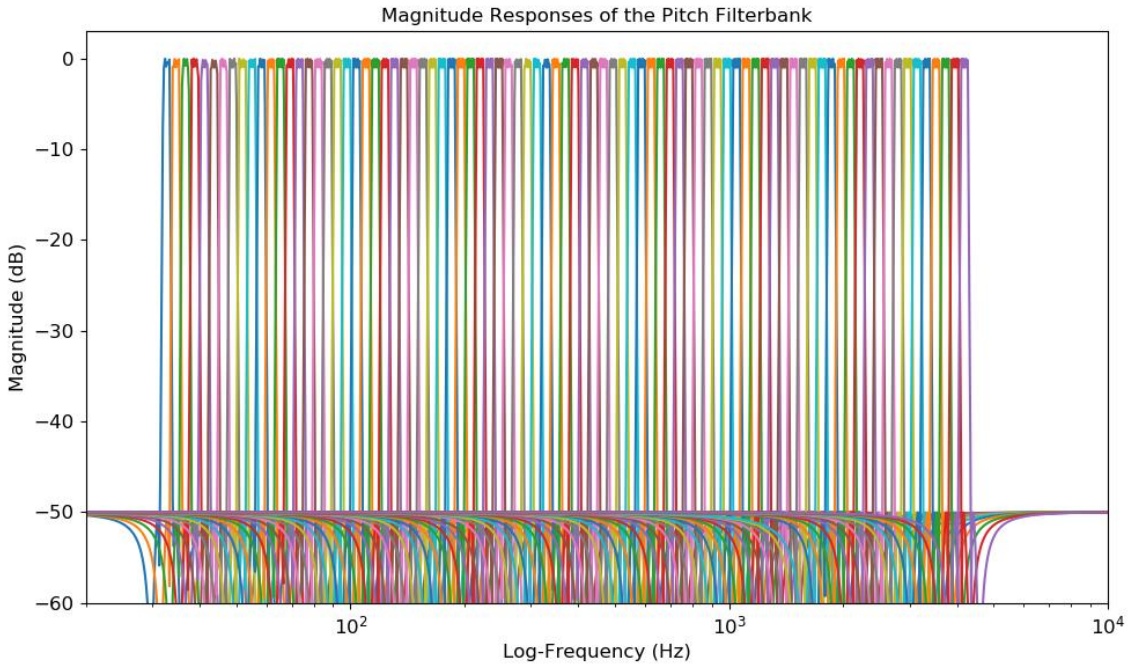
"If, for a given position in one audio recording, the corresponding position within another is to be determined, each recording must first be transformed into so-called feature representations"

Groundwork

Filter out less important information before feature extraction

Transform audio signal into frequency space

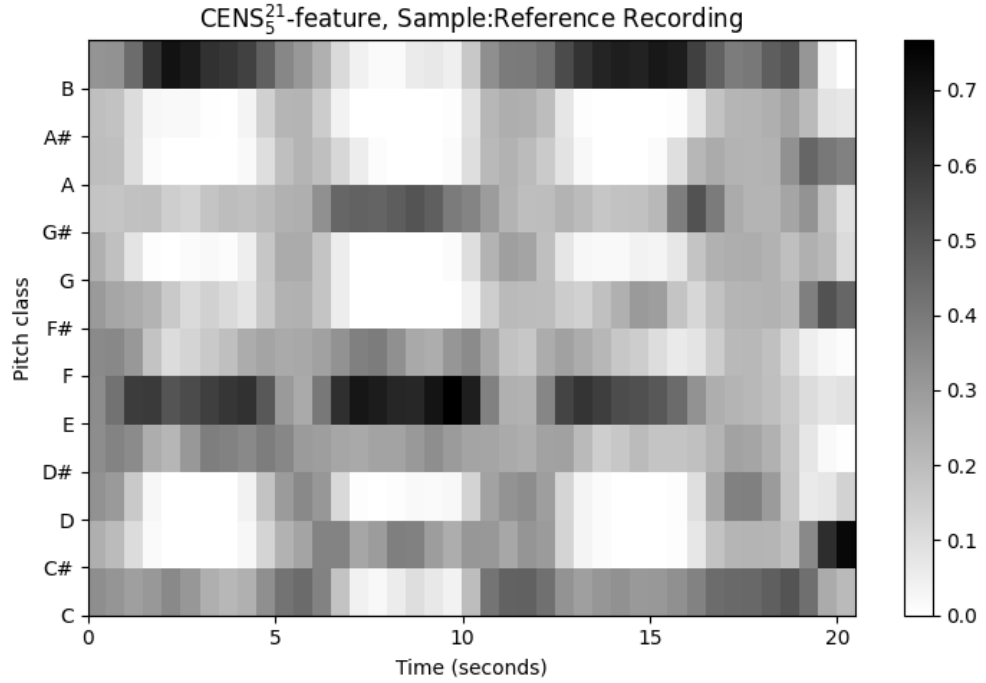
- Fast Fourier Transform (FFT)
- Bandpassfilter



Source: Librosa Documentation

FEATURE EXTRACTION

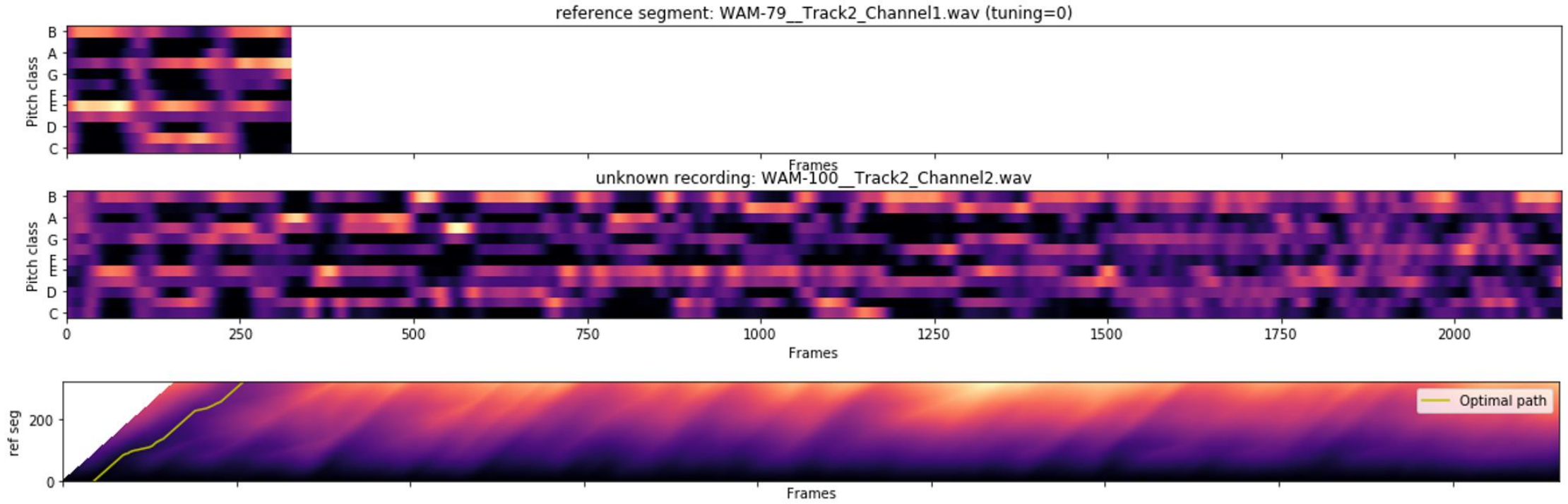
"If, for a given position in one audio recording, the corresponding position within another is to be determined, each recording must first be transformed into so-called feature representations"



Features

- Chromagrams capture melodic and harmonic characteristics of the music while being resilient to variations
- Chroma energy normalized statistics (CENS)
- Extracting statistics over quite large windows smooths out local deviations in articulation, tempo, and execution of note groups

DYNAMIC TIME WARPING



RESULTS

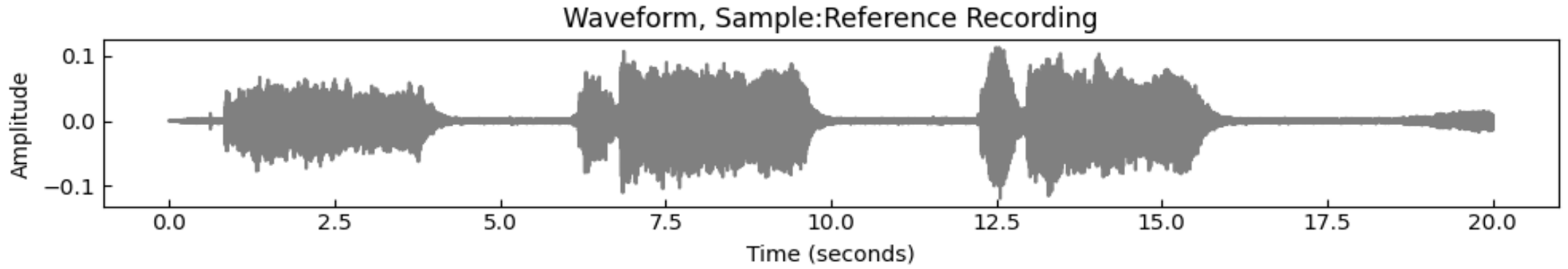
	scenario		
segment types	A0	A1	…
instrumental	0.24	0.81	
singing	0.20	0.63	
speech	0.05	0.08	
total (Ø)	0.15	0.46	

*values = average over 23 version of the magic flute

- A0 = base line scenario (no improvement)
- A1 = semitones implemented

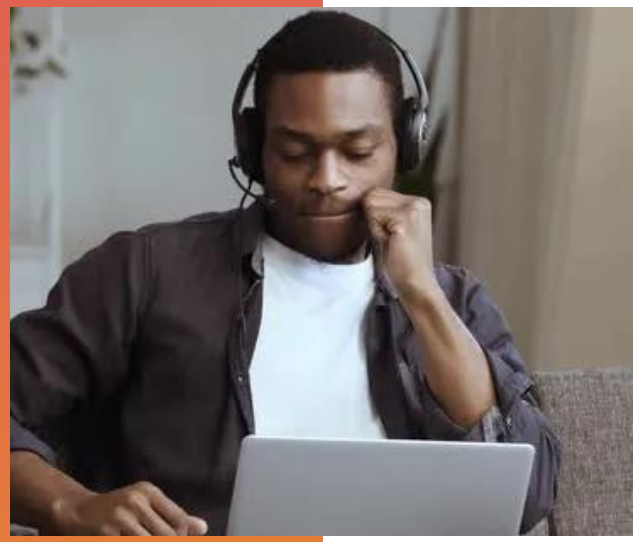
CHALLENGES

- Semitones
- Speech
- Noise
- Technical limitations (recording equipment, recording process, "digitization")
- Overall complexity of the topic



DISCUSSION





Are there any questions?



THANK YOU!

Eduard Klein,
Michelle Schmid,
Christian Stuber

eduard.klein@bfh.ch
michelle.schmid@students.bfh.ch
christian.stuber@students.bfh.ch