INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

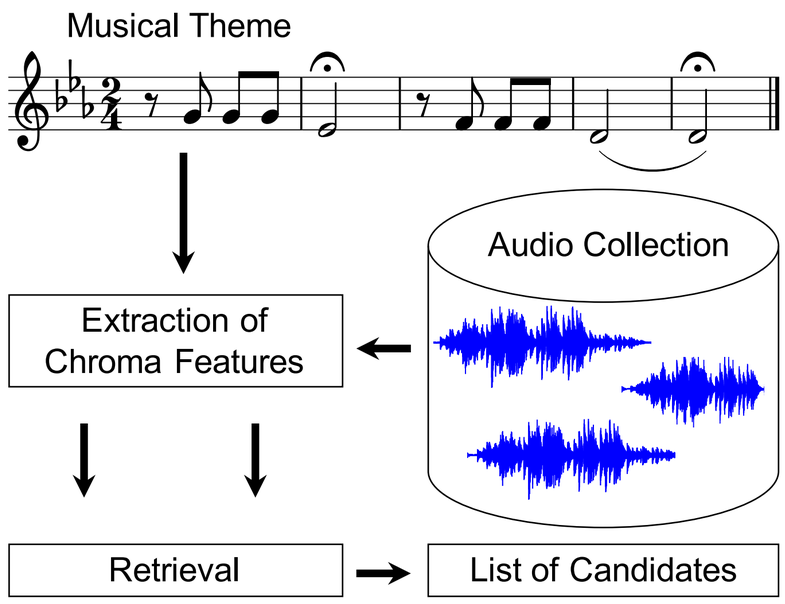
RETRIEVING AUDIO RECORDINGS USING MUSICAL THEMES

Stefan Balke, Vlora Arifi-Müller, Lukas Lamprecht, and Meinard Müller

Abstract

In 1948, Barlow and Morgenstern released a collection of about 10000 themes of well-known instrumental pieces from the corpus of Western Classical music [1]. These monophonic themes (usually four bars long) are often the most memorable parts of a piece of music. In this contribution, we report on a systematic study considering a cross-modal retrieval scenario. Using a musical theme as a query, the objective is to identify all related music recordings from a given audio collection. By adapting well-known retrieval techniques, our main goal is to get a better understanding of the various challenges including tempo deviations, musical tunings, key transpositions, and differences in the degree of polyphony between the symbolic query and the audio recordings to be retrieved. In particular, we present an oracle fusion approach that indicates upper performance limits achievable by a combination of current retrieval techniques.

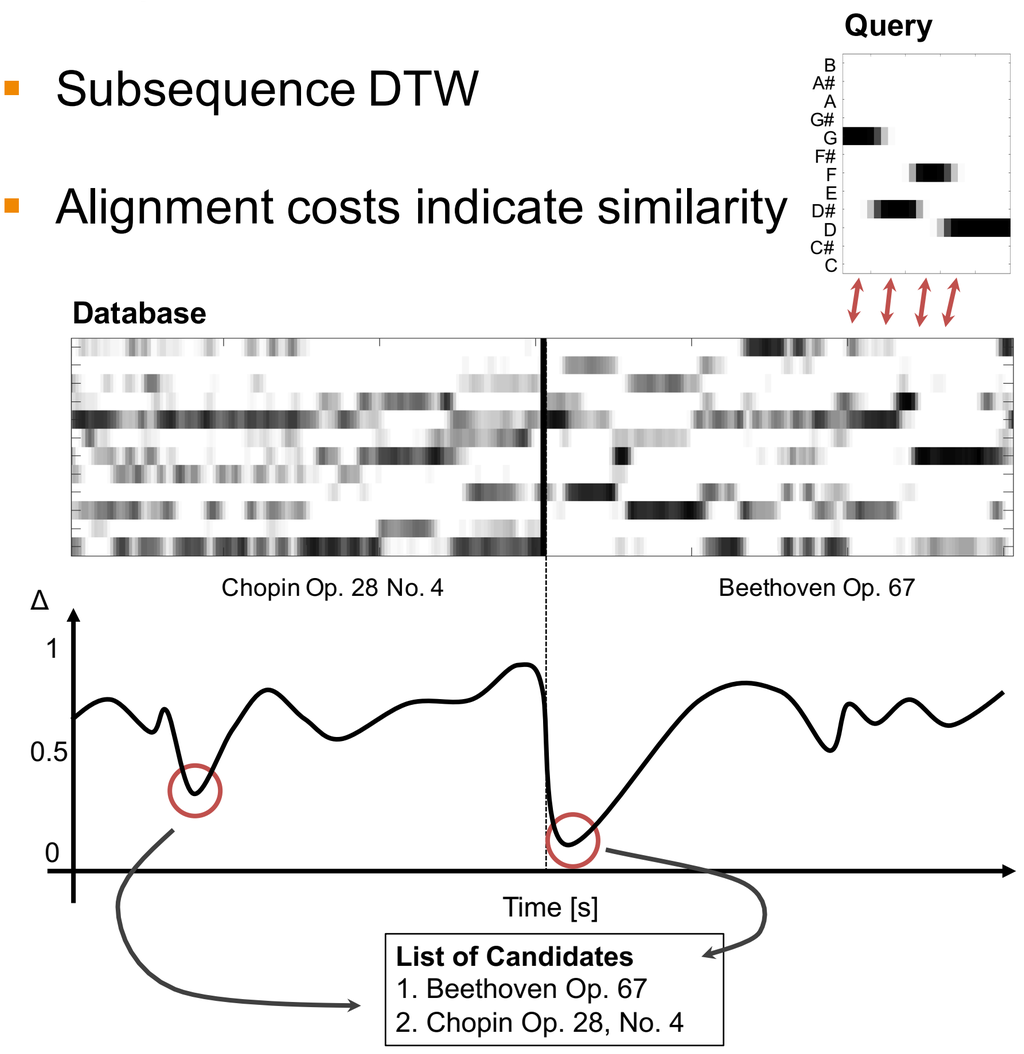
Task & Challenges



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

Retrieval Pipeline

- Compensate for local and global tempo differences
- Subsequence DTW
- Alignment costs indicate similarity



Literature

[1] Harold Barlow and Sam Morgenstern, A Dictionary of Musical Themes, Crown Publishers, 3. edition, 1975.
[2] Jacob T. Schwartz and Diana Schwartz, The electronic dictionary of musical themes, http://www.multimedialibrary.com/barlow/
[3] Meinard Müller, Fundamentals of Music Processing, Springer.

Acknowledgments

This work has been supported by the German Research Foundation (DFG MU 2682/5-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and the Fraunhofer-Institut für Integrierte Schaltungen IIS.

Experimental Setting

- **Goal:** Learn about lower and upper performance boundaries of the used method.
- **Method:** Step-wise providing more prior knowledge.

  [Tu] Tuning information [Tr] Transposition information

  [Ql] Query length [Df] Dominant feature
- **Evaluation measure:** Top-K match

Test Dataset “Small”

- **Queries:** 177 monophonic themes
- **Database:** 100 recordings (~11 h)

More prior knowledge

Top-K	1	5	10	20
Baseline	45.2	62.1	70.1	76.8
Tu	46.9	64.4	72.9	81.9
Tr	52.0	68.9	79.1	87.6
Tu+Tr	53.7	72.3	83.1	91.0
Tu+Tr+Ql	68.4	79.1	88.1	93.2
Tu+Tr+Ql+Df	37.3	57.6	67.8	74.6
Oracle Fusion	72.3	84.7	92.1	97.7

- Compensation for tuning and tempo differences has high impact on results.
- Oracle Fusion indicates upper performance limit with these methods.

Test Dataset “Medium”

- **Queries:** 2046 monophonic themes
- **Database:** 1113 recordings (~120 h)

Top-K	1	5	10	20	50	100	200	500
Tu+Tr+05s	14.9	21.8	25.8	29.2	35.5	43.0	54.1	76.1
Tu+Tr+10s	18.3	25.1	28.3	32.6	38.7	46.1	56.1	76.2
Tu+Tr+15s	13.6	19.5	22.7	26.1	31.6	38.9	49.7	72.4
Oracle	25.0	34.1	39.0	43.5	51.0	59.6	70.2	86.9

- Query length is replaced by fixed durations (5, 10, 15 s).
- Differences in the degree of polyphony remains a major problem.

Application Scenario & Further Challenges