Cover Song Identification

Amy Ruskin
MUMT 621

The task: search a collection & retrieve all of the different versions based on the same underlying musical work

Query \ Document		Textual	Symbolic	Acoustic		Image
				Monophonic	Polyphonic	
Textual		e.g., retrieving lyrics via keywords	e.g., retrieving MIDI or Humdrum music via keywords	e.g., retrieving solo trumpet music via keywords	e.g., retrieving popular songs via keywords	e.g., retrieving scanned sheet music via keywords
Symbolic		e.g., retrieving lyrics via example MIDI music	e.g., using one MIDI music to retrieve other MIDI versions	e.g., retrieving solo trumpet music via example MIDI music	e.g., retrieving popular songs via example MIDI music	e.g., retrieving scanned sheet music via example MIDI music
Acoustic	Mono-phonic	e.g., retrieving lyrics via humming	e.g., retrieving MIDI music via humming	e.g., retrieving solo trumpet music via humming	e.g., retrieving popular songs via humming	e.g., retrieving scanned sheet music via humming
	Poly-phonic	e.g., checking the source of a pre-recorded popular song	e.g., retrieving MIDI versions of a pre-recorded popular song	e.g., retrieving solo trumpet versions of a pre-recorded popular song	e.g., retrieving original/cover versions of a popular song (the problem investigated in this study)	e.g., retrieving scanned sheet music of a popular song
Image		e.g., checking the source of a song via scanned sheet music	e.g., retrieving MIDI music from scanned sheet music	e.g., retrieving solo trumpet versions from scanned sheet music	e.g., retrieving a symphony from scanned sheet music	e.g., checking similar music via scanned sheet music

Problem categories in music retrieval (Tsai et al. 2008, 1671)

What counts as a cover?

Traditionally: Artist X plays a song by Artist Y

For our purposes: “any new version, performance, rendition, or recording of a previously recorded track” (Serrà et al. 2010)

- Some types of versions:
  - remaster
  - instrumental
  - live performance
  - acoustic
  - demo
  - duet
  - medley
  - remix
  - quotation

Cover versions vs. original works

- What can **change**:
  - timbre
  - tempo
  - timing
  - structure
  - key
  - harmonization
  - lyrics
  - additional noise
- What should stay the **same**:
  - (long stretches of) melody/tonal content

(Serrà et al. 2010)

Image source

Image source

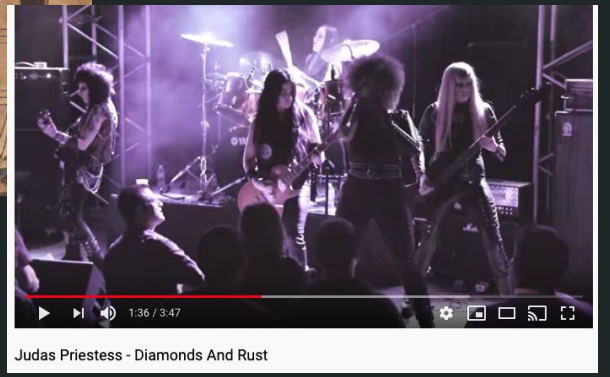

Image source

Usefulness and applications

- Commercial
  - Musical rights management
  - Search, retrieval, and organization of large digital music collections
- Research insights
  - Music similarity
  - Music cognition
- Music consumers
  - “valuable and fun” (Serrà et al. 2010)

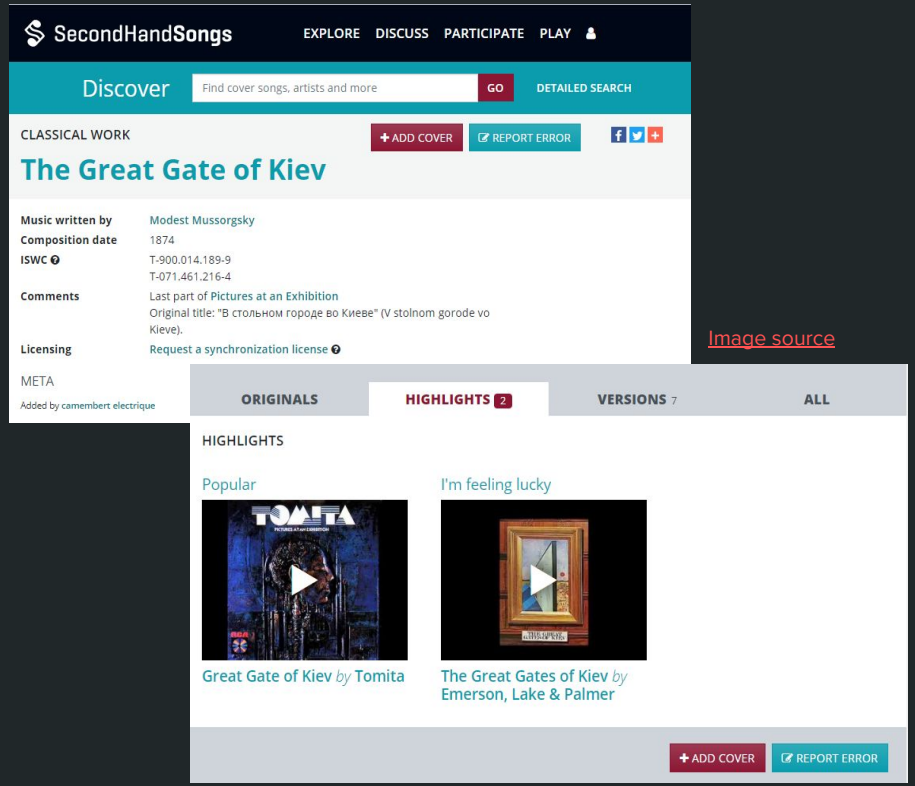

Image source

General approach

1. Feature extraction
    - (Usually) tonal or harmonic content
    - Most common: Pitch Class Profiles (PCP) or chroma features
2. Feature post-processing
    - Key invariance
    - Tempo invariance
    - Structural invariance
3. Similarity estimation
    - Finding and quantifying “matches”
    - May be tied to post-processing methods

Reference(s)	Extracted feature	Key invariance	Tempo invariance	Structure invariance	Similarity computation
Ahonen & Lemstrom [2]	Chords	Relative changes			NCD
Bello [4]	Chords	*K* transpositions	DP		Edit distance
Egorov & Linetsky [20]	PCP	OTI	DP	DP	Match length
Ellis et al. [21, 23]	PCP	All transpositions	Beat		Cross-correlation
Foote [25]	Energy + Spectral		DP		DTW
Gómez & Herrera [28]	PCP	Key estimation	DP		DTW
Gómez et al. [29]	PCP	Key estimation	DP	Repeated patterns	DTW
Izmirli [35]	Key templates		DP		DTW
Jensen et al. [36]	PCP	All transpositions	Fourier transform		Frobenius norm
Jensen et al. [37]	PCP	2D autocorrelation	2D autocorrelation		Euclidean distance
Kim et al.[38, 39]	PCP + Delta PCP	All transpositions			Dot product
Kim & Perelstein [40]	PCP	Relative changes	HMM		MLSS
Kurth & Muller [41]	PCP	All transpositions	Temporal comp./exp.	Sequence windowing	Dot product
Lee [43]	Chords	Key estimation	DP		DTW
Marolt [49]	Melodic	Key estimation	DP	Repeated patterns	Cross-correlation
Marolt [50]	Melodic	2D spectrum	Beat + 2D spectrum	Sequence windowing	Euclidean distance
Müller et al. [53]	PCP		Temporal comp./exp.	Sequence windowing	Dot product
Nagano et al. [55]	PBFV	All transpositions	Beat + DP	Seq. windowing + DP	Match length
Sailer & Dressler [68]	Melodic		Relative changes		Edit distance
Serrà et al. [74, 76]	PCP	OTIs	DP	DP	Match length
Tsai et al. [78, 79]	Melodic	*K* transpositions	DP		DTW
Yang [89]	Spectral		DP	Linearity filtering	Match length

Some approaches to cover song identification (Serrà et al. 2010, 321)

Approaches for scalability

- Locality-sensitive hashing (LSH)
  - Index similar objects using similar hash values
  - Reduces complexity of retrieval from database (Marolt 2008)
  - Applied to chroma features, melodic fragments, chord profiles, etc.
- Reduce number of features
  - Project features into lower dimensional space (Bertin-Mahieux and Ellis 2012)
  - Only compute features for short, representative excerpts (Silva et al. 2018)
- Reduce number of observations (size of candidate pool)
  - Database pruning (Osmalskyj et al. 2013)

Evaluation of CSI systems

- **Task:** Query song ➔ ranked list of answers
- Some measures of assessing quality of returned list:
  - Mean average precision (MAP)
  - R-precision
  - Precision or recall at different rank levels (e.g., P@10)
  - F-measure

(Serrà et al. 2010)

A refresher:

- **precision** = # relevant songs returned / # songs returned
- **recall** = # relevant songs returned / # relevant songs in dataset
- **F-measure** = (potentially weighted) harmonic mean of precision and recall

Cover song datasets

- SecondHandSongs dataset (SHS)
  - Training: 12,960 songs in 4,128 cliques
  - Testing: 5,236 songs in 726 cliques
  - Audio features from EchoNest API (proprietary algorithms)
- Covers80
  - 80 cliques, 2 songs per clique
  - Audio files
- YoutubeCovers
  - 50 cliques, 7 songs per clique
  - Pre-computed audio features
- Mazurkas
  - 2,914 recordings of 49 Chopin Mazurkas (41 to 95 songs per clique)
- MIREX audio cover song identification task
  - Data not publicly available, time-restricted
- Da-TACOS (DaTAset for COver Song Identification and Understanding)
  - Cover Analysis subset: 10,000 songs in 5,000 cliques (2 songs per clique)
  - Benchmark subset: 13,000 songs in 1,000 cliques (13 songs each), 2,000 songs not in cliques (acting as noise in data)
  - Low- and mid-level audio features computed with open source libraries

(Yesiler et al. 2019)

Challenges of Cover Song Identification

- Task made up of complex sub-tasks
  - Inherits challenges of beat-tracking, melody extraction, audio segmentation, etc.
  - Propagation of errors from estimations in feature extraction and processing pipeline (Marolt 2008; Silva et al. 2016)
- Scalability
  - Majority of papers rely on quadratic algorithms to compare each pair of songs
  - Trade-off between scalability and robustness (Yesiler et al. 2019)
- Lack of standardization w.r.t. evaluation metrics and datasets
  - Relationship between mix of genres/types of covers in dataset and complexity of task (Serrà et al. 2010)
  - Difficult to compare results (Serrà et al. 2010; Yesiler et al. 2019)
- Different assumptions about what makes a cover
  - No large-scale studies quantifying changes in musical attributes between cover versions (Yesiler et al. 2019)

References

Bertin-Mahieux, Thierry, and Daniel P.W. Ellis. 2012. “Large-Scale Cover Song Recognition Using the 2D Fourier Transform Magnitude.” In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, 241–6.

Marolt, Matija. 2008. “A Mid-level Representation for Melody-based Retrieval in Audio Collections.” In *IEEE Transactions on Multimedia* 10 (8): 1617–1625.

Osmalskyj, Julien, Sebastien Pérard, Marc Van Droogenbroeck, and Jean-Jacques Embrechts. 2013. “Efficient Database Pruning for Large-Scale Cover Song Recognition.” In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 714–718. https://doi.org/10.1109/ICASSP.2013.6637741

Serrà, Joan, Emilia Gómez, and Perfecto Herrera. 2010. “Audio Cover Song Identification and Similarity: Background, Approaches, Evaluation, and Beyond.” In *Advances in Music Information Retrieval*, edited by Zbigniew W. Raś and Alicja A. Wieczorkowska, 307–332. Berlin; Heidelberg: Springer-Verlag.

Silva, Diego F., Chin-Chia M. Yeh, Gustavo E.A.P.A Batista, and Eamonn Keogh. 2016. “SiMPle: Assessing Music Similarity Using Subsequences Joins.” In *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR 2016)*, 23–29.

Silva, Diego Furtado, Felipe Falcão, and Nazareno Andrade. 2018. “Summarizing and Comparing Music Data and Its Application on Cover Song Identification.” In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR 2018)*, 732–739.

References

Tsai, Wei-Ho, Hung-Ming Yu, and Hsin-Min Wang. 2008. “Using the Similarity of Main Melodies to Identify Cover Versions of Popular Songs for Music Document Retrieval.” *Journal of Information Science & Engineering* 24 (6): 1669–87.

Yesiler, Furkan, Chris Tralie, Albin Correya, Diego F. Silva, Philip Tovstogan, Emilia Gómez, and Xavier Serra. 2019. “Da-TACOS: A Dataset for Cover Song Identification and Understanding.” In *Proceedings of the 20th Conference of the International Society for Music Information Retrieval (ISMIR 2019)*, 327–34.