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

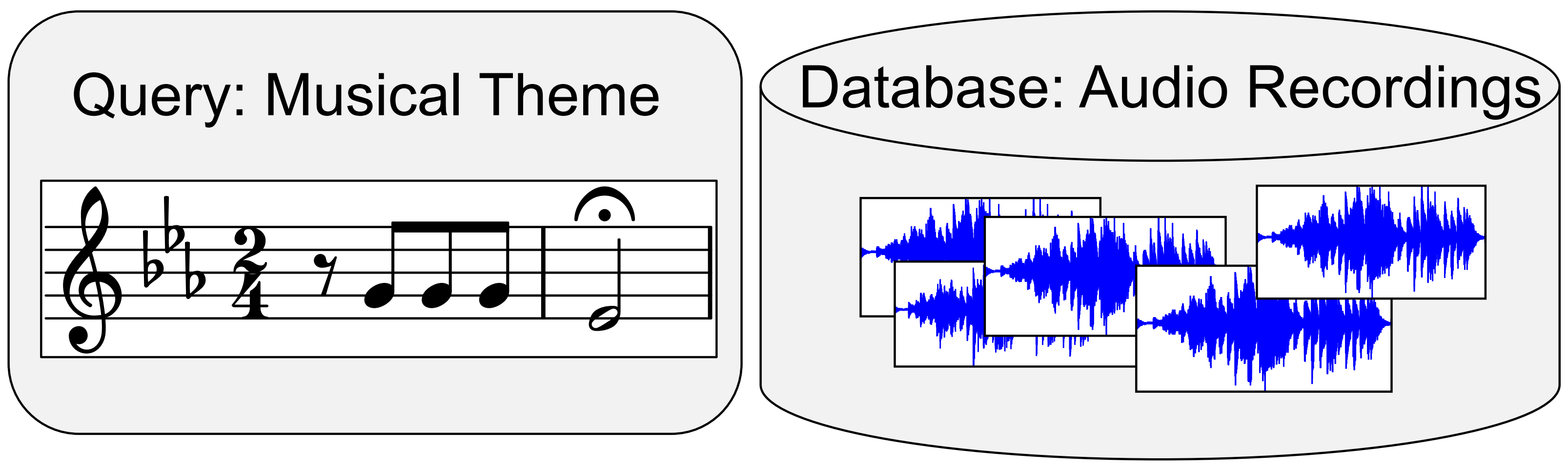
Evaluating Salience Representations for Cross-Modal Retrieval of Western Classical Music Recordings

Frank Zalkow, Stefan Balke, and Meinard Müller

Abstract

In this contribution, we consider a cross-modal retrieval scenario of Western classical music. Given a short monophonic musical theme in symbolic notation as query, the objective is to find relevant audio recordings in a database. A major challenge of this retrieval task is the possible difference in the degree of polyphony between the monophonic query and the music recordings. Previous studies for popular music addressed this issue by performing the cross-modal comparison based on predominant melodies extracted from the recordings. For Western classical music, however, this approach is problematic since the underlying assumption of a single predominant melody is often violated. Instead of extracting the melody explicitly, another strategy is to perform the cross-modal comparison directly on the basis of melody-enhanced salience representations. As our main contribution, we evaluate several conceptually different salience representations for our cross-modal retrieval scenario. Our extensive experimental results, which have been made available on a website, comprise more than 2000 musical themes and 100 hours of audio recordings.

Cross-Modal Retrieval

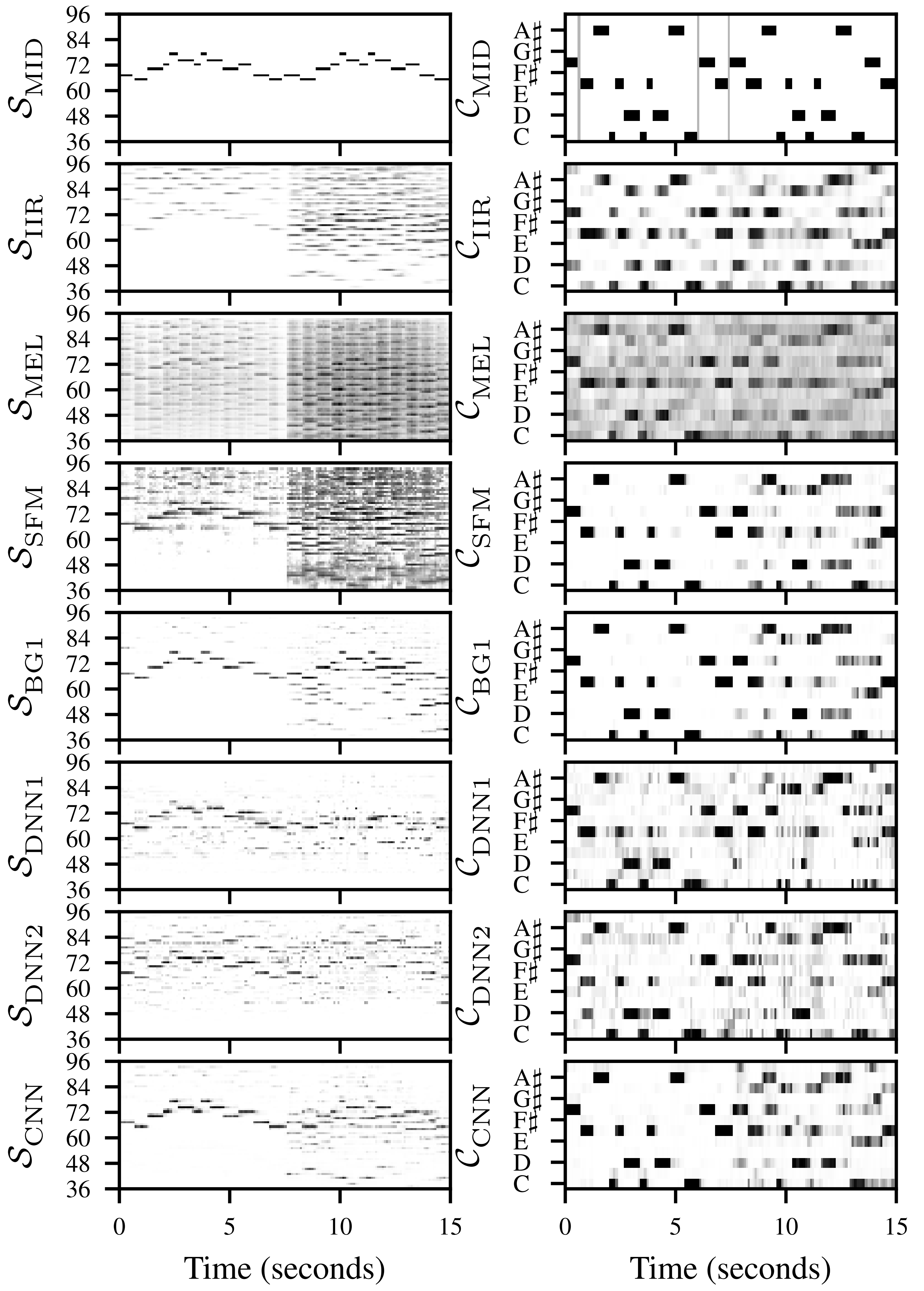

Query: Musical Theme

Database: Audio Recordings

- **Retrieval**
  - Subsequence Dynamic Time Warping with chroma feature sequences [7]
- **Challenges** [1]
  - 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
- **Data Set**
  - Barlow-Morgenstern data set [3]
  - MIDI themes and corresponding audio recordings

	#	Mean Dur.	Total Dur.
Queries	2045	00:00:09	05:00:03
Database	1114	00:06:25	119:15:19

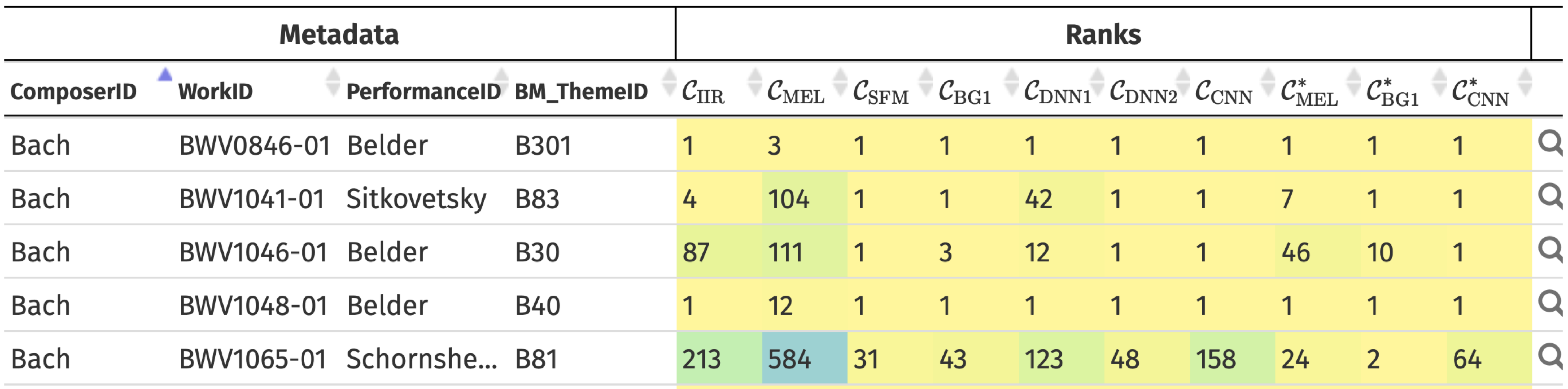
Feature Representations



$\mathcal{S}_{\mathrm{MID}}$	Midi
$\mathcal{S}_{\mathrm{IIR}}$	Log. filterbank [8]
$\mathcal{S}_{\mathrm{MEL}}$	Melodia [9]
$\mathcal{S}_{\mathrm{SFM}}$	Source-filter-model [6]
$\mathcal{S}_{\mathrm{BG1}}$	Bosch [5]
$\mathcal{S}_{\mathrm{DNN1}}$	DNN (trained on jazz) [2]
$\mathcal{S}_{\mathrm{DNN2}}$	DNN (trained on classical) [2]
$\mathcal{S}_{\mathrm{CNN}}$	CNN [4]

- **Main insights:**
  1. Computing salience representation before extracting chroma features helps!
  2. Melody extraction is not beneficial!

Qualitative Evaluation

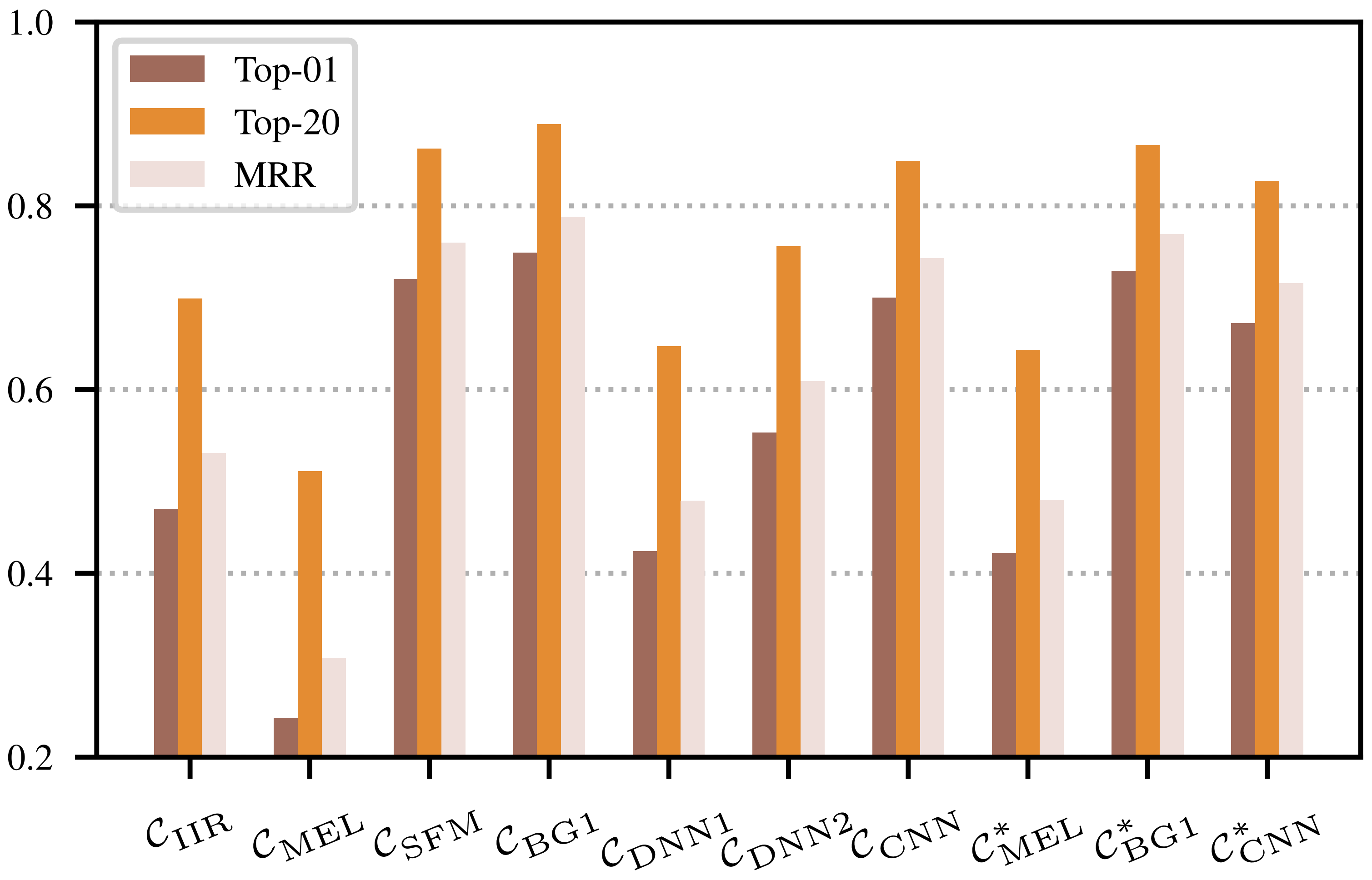


Metadata				Ranks									
ComposerID	WorkID	PerformanceID	BM_ThemeID	c_{IIR}	c_{MEL}	c_{SFM}	c_{BG1}	c_{DNN1}	c_{DNN2}	c_{CNN}	c^{*}_{MEL}	c^{*}_{BG1}	c^{*}_{CNN}
Bach	BWV0846-01	Belder	B301	1	3	1	1	1	1	1	1	1	1
Bach	BWV1041-01	Sitkovetsky	B83	4	104	1	1	42	1	1	7	1	1
Bach	BWV1046-01	Belder	B30	87	111	1	3	12	1	1	46	10	1
Bach	BWV1048-01	Belder	B40	1	12	1	1	1	1	1	1	1	1
Bach	BWV1065-01	Schornshe...	B81	213	584	31	43	123	48	158	24	2	64

Website:

Quantitative Evaluation

- Standard retrieval evaluation measures: Top-01, top-20, mean-reciprocal rank (MRR)
- About 28% more queries achieve rank 1 for BG1, comparing to the baseline approach IIR

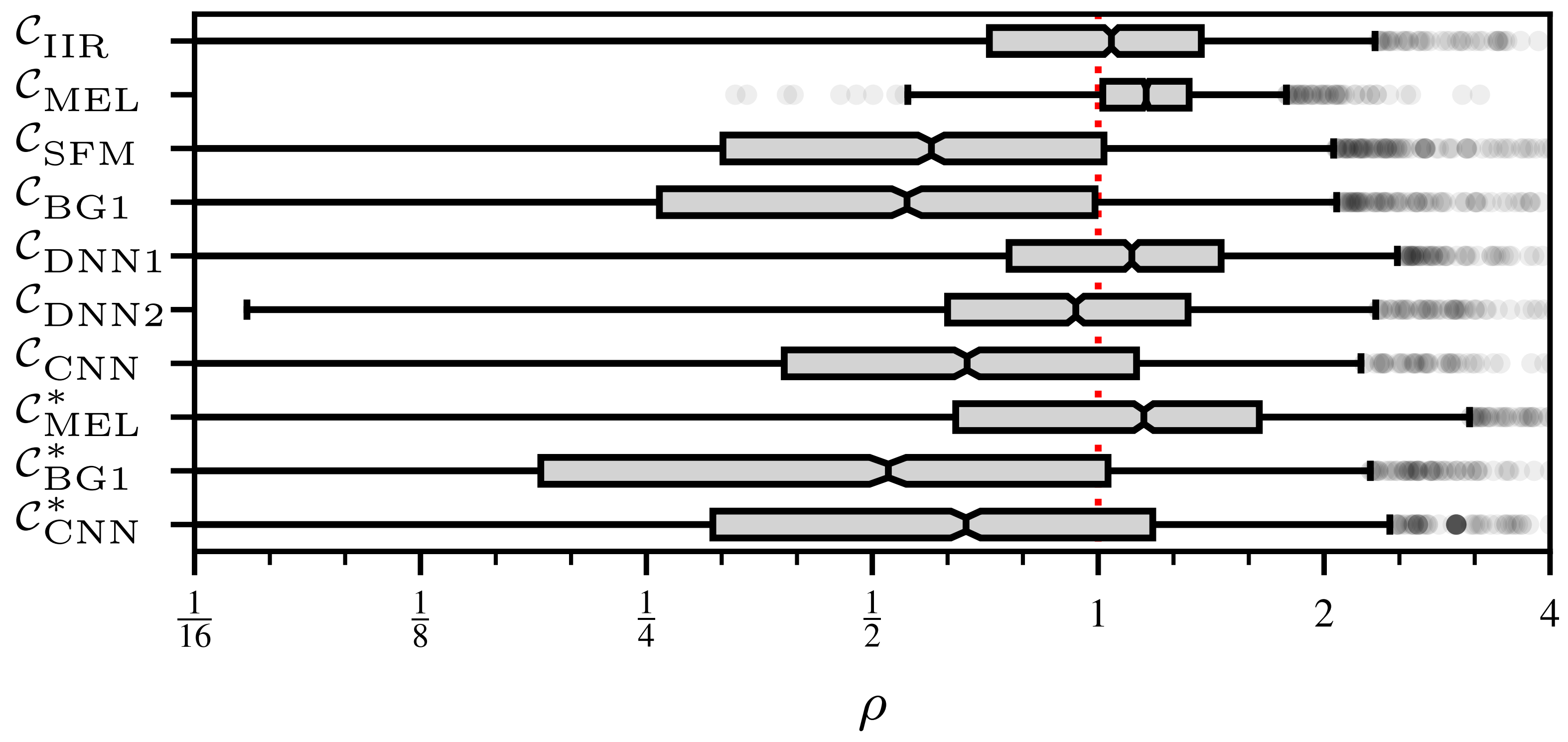


Matching Quality

- Insights in matching quality beyond rank-based evaluation measures
- Separation indicator ρ

$$\rho = \frac{\mathrm{cost}(\text{relevant document})}{\mathrm{cost}(\text{first non-relevant document})}$$

- $\rho < 1$ if relevant document on rank 1
- $\rho > 1$ otherwise
- Small ρ means high matching quality
- Boxplots of ρ values for all representations



References

[1] S. Balke, V. Arifi-Müller, L. Lamprecht, and M. Müller, Retrieving Audio Recordings Using Musical Themes, IEEE ICASSP 2016.

[2] S. Balke, C. Dittmar, J. Abeßer, and M. Müller, Data-Driven Solo Voice Enhancement for Jazz Music Retrieval, IEEE ICASSP 2017.

[3] H. Barlow and S. Morgenstern, A Dictionary of Musical Themes, Crown Publishers, 3rd edition, 1975.

[4] R. Bittner, B. McFee, J. Salamon, P. Li, and J. Bello, Deep Salience Representations for F0 Tracking in Polyphonic Music, ISMIR 2017.

[5] J. Bosch and E. Gómez, Melody Extraction Based on a Source-Filter Model Using Pitch Contour Selection, SMC 2016.

[6] J.-L. Durrieu, B. David, and G. Richard, A Musically Motivated Mid-Level Representation for Pitch Estimation and Musical Audio Source Separation, IEEE JSTSP 2011.

[7] M. Müller, Fundamentals of Music Processing, Springer.

[8] M. Müller and S. Ewert, Chroma Toolbox: MATLAB Implementations for Extracting Variants of Chroma Based Audio Features, ISMIR 2011.

[9] J. Salamon and E. Gómez, Melody Extraction from Polyphonic Music Signals Using Pitch Contour Characteristics, IEEE TASLP 2012.

Acknowledgments

This work has been supported by the German Research Foundation (MU 2686/11-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS.