

SAARLAND MUSIC DATA (SMD)

Meinard Müller, Verena Konz
Saarland University and MPI Informatik
meinard@mpi-inf.mpg.de

Wolfgang Bogler, Viora Arifi-Müller
Hochschule für Musik Saar
w.bogler@hfm.saarland.de

1. COLLABORATION

Computers have become an indispensable tool for storing, processing, analyzing, and generating music. The field of Music Information Retrieval (MIR) is a relatively young research discipline with the objective to develop technologies and interfaces that allow users to access and explore music in all its different facets. Being an interdisciplinary area, MIR brings together experts from a multitude of research and application fields ranging from information science, audio engineering, computer science, to musicology, music theory, and library science. Having a collaboration between the *Max-Planck-Institut für Informatik* (MPII) and the *Hochschule für Musik Saar – University of Music* (HFM), our goal is to establish a platform where computer scientists and musicians can explore and discuss the application of computer-based methods in music analysis, performance analysis, and music education.

2. DATASET

The objective evaluation and comparison of various techniques is crucial for the scientific progress in applied fields such as music information retrieval. Here, the availability of common datasets are of foremost importance. As one important part of our collaboration, we have set up a dataset referred to as *Saarland Music Data* (SMD), which contains royalty free music. This dataset is freely available on the web at:

<http://www.mpi-inf.mpg.de/resources/SMD/>



This dataset is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0

Unported License. It is freely available for research purposes and can be shared and remixed under the same license.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2011 International Society for Music Information Retrieval.

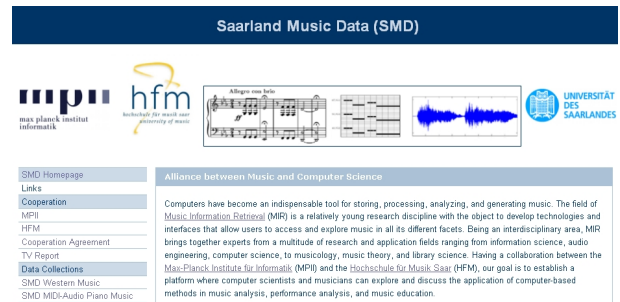


Figure 1. Website of Saarland Music Data (SMD)

The SMD dataset currently consists of two music collections. The first collection contains MIDI-Audio pairs of piano music and is described in Section 2.1. The second collection consists of audio recordings in various instrumentations and is described in Section 2.2.

2.1 SMD MIDI-Audio Piano Music

In the first collection of Saarland Music Data, one finds audio recordings along with perfectly synchronized MIDI files for 50 pieces or movements from the Western piano music literature. In particular, it contains compositions by Bach, Bartok, Beethoven, Brahms, Chopin, Haydn, Liszt, Mozart, Rachmaninoff, Ravel and Skryabin, see Figure 2.

The pieces were performed by students of the Hochschule für Musik Saar on a hybrid acoustic/digital piano (Yamaha Disklavier). The Disklavier allows for capturing key and pedal movements of the piano while playing. This information, which can be stored in a MIDI file, yields an accurate annotation of the corresponding audio recording in the form of a symbolic description of all played musical note events. The SMD MIDI-Audio pairs constitute a valuable dataset for various music analysis tasks such as music transcription, performance analysis, music synchronization, audio alignment, or source separation.

All performances were recorded in the studios of the Hochschule für Musik Saar, played by students of piano classes of different levels, on a Yamaha Disklavier model DCFIIISM4PRO. Using two cardioid-condenser microphones fixed over the resonating body of the piano, all performances were directly recorded into Steinberg Cubase

Database			
No.	Filename	Audio	MIDI
1	Bach_BWV849-01_001_20090916-SMD	mp3	mid
2	Bach_BWV849-02_001_20090916-SMD	mp3	mid
3	Bach_BWV871-01_002_20090916-SMD	mp3	mid
4	Bach_BWV871-02_002_20090916-SMD	mp3	mid
5	Bach_BWV875-01_002_20090916-SMD	mp3	mid
6	Bach_BWV875-02_002_20090916-SMD	mp3	mid
7	Bach_BWV888-01_008_20110315-SMD	mp3	mid
8	Bach_BWV888-02_008_20110315-SMD	mp3	mid
9	Bartok_S2080-01_002_20110315-SMD	mp3	mid
10	Bartok_S2080-02_002_20110315-SMD	mp3	mid
11	Bartok_S2080-03_002_20110315-SMD	mp3	mid
12	Beethoven_Op027No1-01_003_20090916-SMD	mp3	mid
13	Beethoven_Op027No1-02_003_20090916-SMD	mp3	mid
⋮			
42	Mozart_KV265_006_20110315-SMD	mp3	mid
43	Mozart_KV398_002_20110315-SMD	mp3	mid
44	Rachmaninoff_Op036-01_007_20110315-SMD	mp3	mid
45	Rachmaninoff_Op036-02_007_20110315-SMD	mp3	mid
46	Rachmaninoff_Op036-03_007_20110315-SMD	mp3	mid
47	Rachmaninov_Op039No1_002_20090916-SMD	mp3	mid
48	Ravel_JeuxD'Eau_008_20110315-SMD	mp3	mid
49	Ravel_ValseesNoblesEtSentimentales_003_20090916-SMD	mp3	mid
50	Skryabin_Op008No8_003_20090916-SMD	mp3	mid

Figure 2. SMD MIDI-Audio Piano Music Collection

4. Except for trimming the beginnings and ends of the recordings, no further post-processing (filters, effects) was applied to the musical material. From each Cubase project, an audio file (44.1 kHz, stereo) as well as a synchronized standard MIDI file (SMF) were exported. The audio files were then converted into MP3 files (192 kbit/s) encoded with the LAME MPEG Audio Layer III (MP3) encoder. A correctly decoded MP3 file should result in a WAV file with a temporal synchronization accuracy of 10 ms on the note onset level compared to the respective MIDI file. The overall temporal accuracy is limited by internal delays of the Disklavier and the speed of sound.

2.2 SMD Western Music

In the second collection of Saarland Music Data, one finds audio recordings of 200 pieces or movements from the Western classical music repertoire. Among others, it contains compositions by Bach, Beethoven, Berg, Brahms, Chopin, Debussy, Fauré, Mozart, Poulenc, Rachmaninoff, Ravel, Respighi, Schubert, Schumann and Tchaikovsky. The collection contains mainly piano music, chamber music for different instruments (e. g. violine, cello, flute, clarinet, bassoon, horn), Klavierlieder (songs with piano), and some orchestral music. All pieces were performed by students or staff members of the Hochschule für Musik Saar between the years 2004 and 2010 under different recording conditions.

Database			
No.	Filename	Instrumentation	Audio
1	Bach_BWV848-01_100_20040203-SMD	piano	mp3
2	Bach_BWV848-02_100_20040203-SMD	piano	mp3
3	Bach_BWV853-01_100_20040203-SMD	piano	mp3
4	Bach_BWV1001-01_101_20080527-SMD	violin	mp3
5	Bach_BWV1001-02_101_20080527-SMD	violin	mp3
6	Bach_BWV1004-01_102_20080707-SMD	violin	mp3
7	Bach_BWV1004-02_102_20080707-SMD	violin	mp3
8	Bach_BWV1004-03_102_20080707-SMD	violin	mp3
9	Bach_BWV1004-04_102_20080707-SMD	violin	mp3
10	Bach_BWV1042-02_114_20100627-SMD	violin, orchestra	mp3
11	Bach_BWV1056-01_115_20101108-SMD	piano	mp3
12	Beethoven_Op007-03_116_20101108-SMD	piano	mp3
13	Beethoven_Op007-04_116_20101108-SMD	piano	mp3
⋮			
192	Schumann_Op102-05_128_20100609-SMD	duett piano cello	mp3
193	Schumann_Op105-01_103_20100609-SMD	duett piano violin	mp3
194	Schumann_Op105-02_103_20100609-SMD	duett piano violin	mp3
195	Schumann_Op105-03_103_20100609-SMD	duett piano violin	mp3
196	Schumann_Op132-01_133_20100609-SMD	clarinet, viola, piano	mp3
197	Schumann_Op132-02_133_20100609-SMD	clarinet, viola, piano	mp3
198	Schumann_Op132-03_133_20100609-SMD	clarinet, viola, piano	mp3
199	Schumann_Op132-04_133_20100609-SMD	clarinet, viola, piano	mp3
200	Tchaikovsky_Op062_114_20100627-SMD	cello, orchestra	mp3

Figure 3. SMD Western Music Collection

3. NAMING CONVENTIONS FOR FILES

For the audio and MIDI files, the following naming convention is used:

```
Composer_Work_Performer_Version.mp3
Composer_Work_Performer_Version.mid
```

Here, the field `Composer` denotes the last name of the composer, the field `Work` the opus number (or similar) of the piece and possibly the number of the movement (using the extension `-xx`), the field `Performer` a three digit identifier of the performer (orchestra, conductor, soloist, and so on), and the field `Version` the recording date in the form `yyyymmdd` and possibly other information ending with `-SMD`. For example,

```
Beethoven_Op057-02_013_20090221-SMD.mp3
```

denotes an MP3 recording of Beethoven's piano sonata Opus 57, second movement, played by pianist 013 recorded on February 21, 2009.

Acknowledgement. This dataset is the result of the work by many people over the last three years. First of all, we thank all piano students and musicians of the Hochschule für Musik Saar for their contributions. Furthermore, we thank Prof. Thomas Duis and Fedele Antonicelli for their support. We wish to thank Peter Grosche and Philip Thelen for their technical support as well as Thomas Prätzlich, Jonathan Driedger, Nanzhu Jiang and Zhe Zuo for helping to clean up the data and to set up the website. This work has been supported by the Cluster of Excellence on Multimodal Computing and Interaction at Saarland University.