

---

## DATASET

# MTD: A Multimodal Dataset of Musical Themes for MIR Research

Frank Zalkow, Stefan Balke, Vlora Arifi-Müller and Meinard Müller

---

Musical themes are essential elements in Western classical music. In this paper, we present the Musical Theme Dataset (MTD), a multimodal dataset inspired by “A Dictionary of Musical Themes” by Barlow and Morgenstern from 1948. For a subset of 2067 themes of the printed book, we created several digital representations of the musical themes. Beyond graphical sheet music, we provide symbolic music encodings, audio snippets of music recordings, alignments between the symbolic and audio representations, as well as detailed metadata on the composer, work, recording, and musical characteristics of the themes. In addition to the data, we also make several parsers and web-based interfaces available to access and explore the different modalities and their relations through visualizations and sonifications. These interfaces also include computational tools, bridging the gap between the original dictionary and music information retrieval (MIR) research. The dataset is of relevance for various subfields and tasks in MIR, such as cross-modal music retrieval, music alignment, optical music recognition, music transcription, and computational musicology.

---

**Keywords:** Musical Themes; Western Classical Music; Sheet Music; Audio; MIDI; Score-Audio Alignments

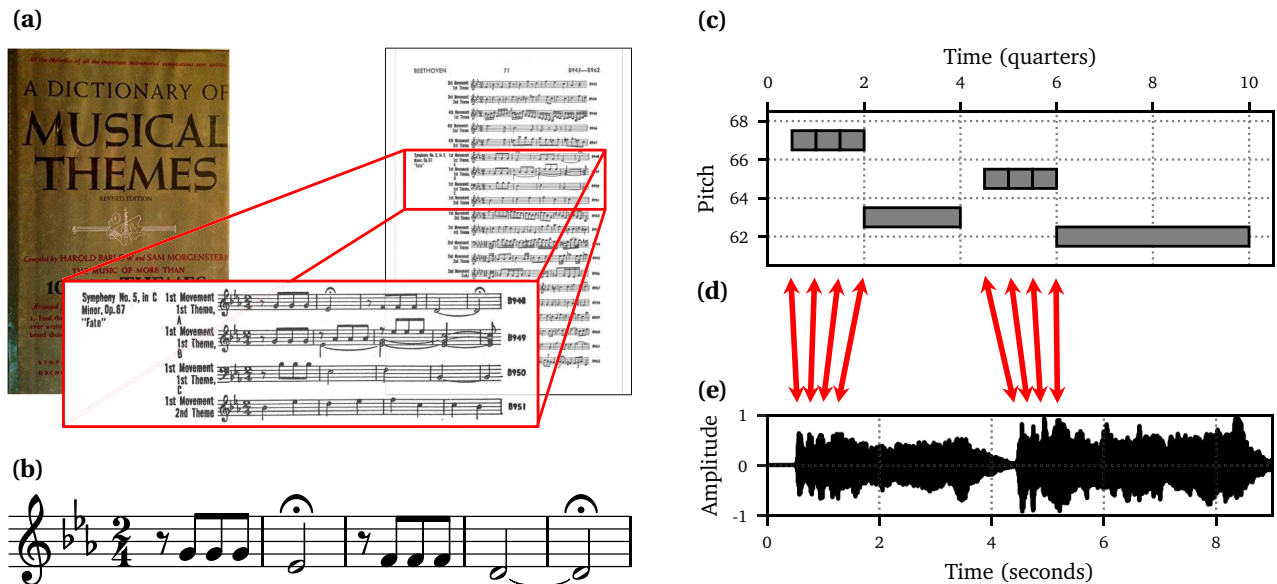
---

## 1. Introduction

Western classical music is largely based on musical themes (Drabkin, 2001). Such a theme is a musical idea used to build a composition (or a part of it). Often, this idea is a prominent melody that is announced in the first measures and is easily recognizable by the listener. Throughout the musical piece, usually, the theme recurs in the form of repetitions and variations. Barlow and Morgenstern compiled “A Dictionary of Musical Themes” that first appeared in 1948 (Barlow and Morgenstern, 1975). In this book (referred to as BM in the following), the authors listed nearly ten thousand musical themes from instrumental musical works. **Figure 1a** shows a page from the BM dictionary and a detailed view of the first theme of Beethoven’s Symphony No. 5. John Erskine explains in the book’s introduction that “the ten thousand themes [...] do not encompass the entire literature of music, but they do include practically all the themes which can be found in compositions that have been recorded.” Though this statement from 1948 may not be valid today, it shows how ambitious the book was perceived at the time.

This paper describes a multimodal dataset, called MTD (Musical Theme Dataset), that is inspired by the original BM dictionary. Using a subset of 2067 themes of the BM book, we digitized and extensively augmented the material, and provide several digital representations of the

musical themes. Beyond graphical sheet music (**Figure 1b**, similar to the BM dictionary), the dataset encompasses symbolic encodings (**Figure 1c**) of the themes in various formats (MIDI, MusicXML, CSV). As one main contribution of the MTD, we annotated the occurrences of the themes in audio recordings and provide snippets from these recordings corresponding to the annotated occurrences (**Figure 1e**). We also provide machine-readable metadata concerning the composer, work, recording, and musical characteristics. As another major component, we manually time-aligned the symbolic encodings to the audio snippets (**Figure 1d**). We provide alignment information as well as modified symbolic encodings that are synchronized to the audio versions. These links can be seen as note-level annotations of the audio material, yielding valuable fine-grained reference annotations for tasks such as audio transcription and cross-modal retrieval. A special feature of the dataset is that the themes are monophonic, while they usually appear in a polyphonic context in the recordings. The difference in the polyphony of the modalities allows for studying tasks related to melody extraction and source separation. All the modalities are easily accessible through our web-based interfaces. In addition, we provide basic tools for parsing, visualizing, converting, and processing the various data modalities. In this way, we bridge the gap between the printed BM book and music information retrieval (MIR) research. We also provide our custom tools for manually aligning the symbolic and audio versions to enable the users of the MTD to continue expanding the dataset.



**Figure 1:** The BM book and various derived data modalities from MTD. **(a)** Original book. **(b)** Clean sheet music engraving of a musical theme. **(c)** Piano roll representation. **(d)** Alignment data. **(e)** Waveform of audio snippet.

We refer to the dataset as multimodal because it contains representations on various semantic levels (symbolic, audio, image). Other meanings of the term “modality” may instead refer to sensations like vision, touch, hearing, and kinematics (Timmers et al., 2015). In this article, we do not use the term in the latter way.

The MTD is relevant to the MIR community in various ways. The correspondences between the different modalities (symbolic, audio, image) can be used for cross-modal retrieval (Müller et al., 2019). Actually, preliminary versions of what have become the MTD already have been used for music retrieval experiments (Balke et al., 2016; Zalkow et al., 2019; Zalkow and Müller, 2020). The manual alignments constitute valuable material for automatic music alignment (Joder et al., 2013; Müller et al., 2004; Arzt and Lattner, 2018). The graphical sheet music can be used for optical music recognition (Rebelo et al., 2012; Byrd and Simonsen, 2015; Calvo-Zaragoza et al., 2020). The dataset can also be useful for music transcription (Benetos et al., 2019) or melody extraction (Salamon et al., 2014). Finally, aspects of polyphony could be interesting for the subfield of computational musicology (Volk et al., 2011).

This article is structured as follows. In Section 2, we discuss related datasets and summarize the MIR literature that is related to aspects of the BM book. Then, in Section 3, we address the role of musical themes in Western classical music and give some examples from the BM book. As the main contribution of this article, we describe in Section 4 the MTD and its modalities, and then introduce in Section 5 our web-based interfaces and tools. Finally, to illustrate the potential of the MTD, we discuss in Section 6 possible applications and future work.

## 2. Related Datasets and Literature

### 2.1 Datasets

A diverse range of research datasets has been published by the MIR community. The first larger music dataset compiled specifically for research purposes is the RWC

music database (Goto, 2004). More recent examples are the multitrack dataset MedleyDB (Bittner et al., 2014) or the Erkomaishvili dataset for ethnomusicological research (Rosenzweig et al., 2020).<sup>1</sup> Serra (2014) discussed the role of datasets for the MIR community. Specifically, he distinguishes between rather unstructured datasets and curated research corpora. In particular, he introduces five criteria (purpose, coverage, completeness, quality, reusability) that are essential for a research corpus. In that sense, the MTD and many other datasets mentioned in our article, can be regarded as research corpora. In the following, we discuss some datasets that are more closely related to the MTD.

Most of the BM themes have been available as symbolic versions (MIDI) at a website called *The Multimedia Library*, developed by Jacob and Diana Schwartz. Unfortunately, the page is now offline and the MIDI files have been withdrawn due to copyright reasons. Currently, the page is only reachable with the Way-back Machine without access to the MIDI files.<sup>2</sup> The dataset was denoted as *The Electronic Dictionary of Musical Themes* (EDM). While the EDM yields MIDI files for all ten thousand BM themes, the MTD provides a wealth of different representations and tools for two thousand of these themes.

There are related datasets containing main melody annotations, such as the Orchset (Bosch et al., 2016) for orchestral music recordings, and MedleyDB (Bittner et al., 2014) mainly for popular music recordings. Datasets with main melody annotations also exist for purely symbolic music (Simonetta et al., 2019). Although a musical theme can occur as a main melody, the concepts are not identical. A main melody is a salient element in an excerpt of music and does not necessarily play an important role in the composition. In contrast to that, a theme is a musical idea that can occur in different ways within the musical context (see Section 3 for various examples). Furthermore, a theme is important throughout the musical piece because it typically recurs in the form of repetitions and variations in

the course of the composition. Another distinction of the MTD is its diverse instrumentation, which is not restricted to orchestral music but also contains instrumental solo pieces and chamber music.

Several datasets for automatic music transcription (AMT) provide audio recordings of musical pieces and symbolic encodings that are synchronous with the recordings. In particular, many datasets focus on piano music. Examples are the MAPS database (Emiya et al., 2010), the SMD (Müller et al., 2011), or the Maestro dataset (Hawthorne et al., 2019). For these AMT datasets, the alignments between the symbolic and audio representations are obtained by using hybrid acoustic/digital player pianos. In contrast to that, the MTD contains audio from commercial recordings that are manually aligned to the symbolic encodings. Another AMT dataset is MusicNet (Thickstun et al., 2017), which provides audio recordings and symbolic encodings of Western classical music pieces in diverse solo and chamber music instrumentations. For this dataset, the alignments between audio and symbolic representations have been created fully automatically using dynamic time warping. A further related dataset is MSMD containing MIDI representations, graphical sheet music, and synthesized audio for classical piano pieces with note-level alignments between the modalities (Dorfer et al., 2018). Though rather designed for sheet music retrieval and score-following, it can also be used for AMT. Unlike the MTD, AMT datasets (such as MAPS, SMD, Maestro, MusicNet, and MSMD) contain all note events of the polyphonic pieces. In contrast, we provide the note events of the monophonic themes, even if they appear in a polyphonic context. While the mentioned datasets are better suited for polyphonic music transcription, the MTD is more appropriate, e.g., for melody estimation, or for studying questions concerning musical themes, saliency, and polyphony.

Other related datasets are used for the MIR task of query-by-humming (see Section 2.2 for more details). For example, the MTG-QBH dataset (Salamon et al., 2013) contains monophonic recordings of melody excerpts from amateur singers without symbolic representations. Given a set of polyphonic audio recordings where the same melodies occur, one can compare the a cappella recordings with the polyphonic recordings.

## 2.2 Literature

The BM dictionary contains a book index for finding the musical themes in the dictionary. To get the index term for a theme, one first transposes it to C (C major for major keys and C minor for minor keys). The first pitches from the theme without octave information are then used as an index term. This concept (pitches without octave information) is similar to the concept of pitch classes. However, for the index terms, one keeps enharmonic spellings, which is not always done for pitch classes. For example, let us consider the first theme of Beethoven's Fifth Symphony (see **Figure 1**), which does not need to be transposed because it already is in the key of C minor. Its index term in the book is (G, G, G, Eb, F, F). In the book, the index term is paired with a theme identifier, which allows the reader to quickly find the page of the theme.

This indexing scheme influenced several algorithmic approaches to musical search. There exist many different ways to realize an automated search engine for melodies and themes. One may categorize search scenarios according to the modalities used for the query and the database. **Table 1** shows such a categorization, where a single reference for each category is provided as an example. An early example for symbolic-symbolic search is Themefinder,<sup>3</sup> which provides a web-based interface for searching musical themes or incipits (Kornstädt, 1998). A more recent example is the online music catalog RISM,<sup>4</sup> where a diatonic incipit search used to be offered, using index terms similar to the BM index, but without accidentals (Diet and Gerritsen, 2013).

An undertaking similar to the BM book is the dictionary by Parsons (1975). In this book, compared to Barlow and Morgenstern (1975), the indexing technique for musical melodies is further developed. The index term for a theme is here defined by its contour, where one specifies for each note if its pitch goes up (u), down (d), or repeats (r) compared to the previous note. The symbol \* denotes the beginning of the sequence. For our Beethoven example the index term is (\*, r, r, d, u, r, r, d). Researchers from the field of MIR investigated the benefits and limitations of the indexing schemes by Barlow and Morgenstern (Berman et al., 2006) and Parsons (Uitdenbogerd and Yap, 2003). Prechelt and Typke (2001) used the Parsons code in their query-by-humming system called Tuneserver. In this system, the user specifies a query by whistling or humming a melody. This query is then transcribed and compared with the symbolic melodies of the dictionary using the Parsons code. In this retrieval scenario, they used audio-based queries and a database of symbolic themes. There is also a kind of opposite retrieval scenario, where the query consists of a symbolic encoding of a musical theme, which is then used to identify music recordings that contain this theme (Balke et al., 2016; Zalkow et al., 2019; Zalkow and Müller, 2020). We will describe this scenario in further detail in Section 6. Other retrieval scenarios use audio representations for both queries and database documents (Salamon et al., 2013).

The BM themes have also been used in the MIR community for classifying composers (Pollastri and Simoncelli, 2001) and for segmenting themes in polyphonic symbolic music (Meek and Birmingham, 2001). London (2013) examined the BM book in a meta-study on building representative music corpora and considered it a useful collection. Of course, the BM dictionary also influenced and inspired research outside music retrieval research.

**Table 1:** Categorization of theme/melody retrieval scenarios, according to modalities.

Query	Database	Example Literature
symbolic	symbolic	Kornstädt (1998)
audio	symbolic	Prechelt and Typke (2001)
symbolic	audio	Balke et al. (2016)
audio	audio	Salamon et al. (2013)

Examples are a theoretical study about musical intervals (Vos and Troost, 1989), psychological music articles (Simonton, 1980, 1991), and a book about the origin and evolution of speech and music (Changizi, 2011).

### 3. Musical Themes in Western Classical Music

As Drabkin (2001) describes, a theme is “the musical material on which part or all of a work is based.” Going even beyond that, Reti (1951) describes the thematic process in a composition as its main form-building element, creating unity even across multiple movements of a work. The musical term “theme” originates from the 16th century. As we understand it today, a theme is a musical idea that conveys a sense of “completeness” and “roundedness,” in contrast to the shorter and more basic musical motif. An essential aspect of Western classical music is the repetition and variation of thematic material throughout a musical work. In musicology, it is not well defined whether a theme is only the monophonic melodic line, independent from the polyphonic context in which it occurs, or the entire polyphonic section (Drabkin, 2001). However, in this article, we always consider a theme as being monophonic. Sometimes, it can be subjective whether a melodic line is a theme, and even musicologists may not agree upon this. In our context, we use the BM dictionary as reference to identify musical themes.

Musical themes can have different degrees of prominence or salience within their polyphonic contexts. In the following, we have a look at four different examples with a decreasing degree of salience. The first example is again the famous “Fate motif” from Beethoven’s Fifth Symphony. **Figure 2a** shows a piano transcription of the section where the theme occurs. Although the name suggests that it is a motif, it is classified as a theme in the BM book. One might consider the first four notes as a motif, while the theme consists of two motif statements at different diatonic pitches. In this example, the theme is played by all instruments, in different octaves. As a consequence, only a single pitch class is present at a time. In the BM dictionary, the pitches of the upper staff constitute the theme (colored in red in **Figure 2a**). As a second example, we consider the second theme in the first movement of Beethoven’s Piano Sonata Op. 2, No. 2 (see

**Figure 2b**). Here, a main melody appears with a harmonic accompaniment, which is a typical situation for a musical theme. The sixteenth notes of the accompaniment present a minor triad (E, G, B) in the first half and a diminished triad (F#, A, C) in the second half. The theme is still prominent since it contains the highest pitches and is the only melodic line in this section. The third example is the second theme in the first movement of Schubert’s Piano Sonata D 960 and is shown in **Figure 2c**, where the red noteheads indicate the theme. This theme is less prominent because, first, it is in a middle voice and, second, the upper voice also is a melodic line, though with less independence. The fourth example (**Figure 2d**) is the beginning of the first piece of the suite *Images* by Debussy. This is a complex case because two different themes are overlapping. One theme is played by the right hand (upper staff, colored in red), and another short theme is played with the left hand (lower staff, colored in blue). In the BM book, both themes are referenced with a single identifier. However, for MTD, we gave two identifiers for the respective themes. Furthermore, the lower-staff theme again illustrates that there is no strict boundary between a musical theme and a motif. One may argue that this basic three-note sequence is to be regarded as a motif instead of a theme. However, Barlow and Morgenstern classified it as a theme in their dictionary.

### 4. Dataset

As our main contribution, in this section, we describe the MTD dataset. We start by discussing the origins of the MTD in the BM book and the EDM (Section 4.1). Next, we explain our collected metadata (Section 4.2). We then describe the various symbolic encodings of the MTD (Section 4.3) and the audio recordings (Section 4.4). The alignment between the symbolic and audio representations is then explained in Section 4.5. Finally, we summarize the directory and file structure of the MTD (Section 4.6).

#### 4.1 BM and EDM

The BM book, which contains nearly ten thousand themes, is considered a good starting point for building a representative corpus of classical music (London, 2013).

Figure 2 consists of four musical examples labeled (a) through (d). Each example shows a piano transcription of a specific musical theme. (a) shows the first theme of Beethoven's Symphony No. 5 in C minor, Op. 67, first movement, marked *ff*. (b) shows the second theme of Beethoven's Piano Sonata No. 2 in A major, Op. 2, No. 2, first movement, marked *espressivo* and *sf*. (c) shows the second theme of Schubert's Piano Sonata in Bb major, D 960, first movement, marked *pp* and *crzsc.*. (d) shows the beginning of the first piece of the suite *Images* by Debussy, marked *pp*, featuring two overlapping themes: one in the right hand (upper staff, colored in red) and one in the left hand (lower staff, colored in blue).

**Figure 2:** Various themes from the MTD. **(a)** Beethoven: Symphony No. 5 in C minor, Op. 67, first movement (piano transcription), first theme. **(b)** Beethoven: Piano Sonata No. 2 in A major, Op. 2, No. 2, first movement, second theme. **(c)** Schubert: Piano Sonata in Bb major, D 960, first movement, second theme. **(d)** Debussy: *Reflets dans l'eau* (*Images*, Book 1, L 110, No. 1), two themes.



The book lists the themes along with an identifier and basic information about the composer and the musical work. *The Electronic Dictionary of Musical Themes* (EDM, see Section 2.1) contains the BM themes as MIDI files, which are named by a one- to four-digit number. The enumeration does not follow the BM book's order, which makes it hard to link the EDM collection to the BM book.

The first step in processing the EDM files was to identify the corresponding themes in the BM book. We then re-enumerated the MIDI files strictly in the order of the BM book. This new number is used as an identifier for the MTD. Because our dataset contains not all themes, but a subset, the list of the MTD identifiers has gaps.

The MTD was initially designed as a test scenario for cross-modal retrieval experiments. Then the dataset was successively broadened by adding more modalities, metadata, and alignment data. This makes the dataset a valuable testbed for various MIR tasks. Since the preparation of the modalities is labor-intensive, we restricted ourselves to 2067 themes by well-known composers. When selecting the MTD themes, we were guided by several practical considerations rather than musical guidelines. We preferred sets of themes from the BM book that correspond to complete work cycles, e.g., Beethoven's complete piano sonatas. In particular, we considered musical works contained in comprehensive CD album collections (e.g., Brilliant Classics' "Complete Edition" of works by Beethoven), such that a single album collection covers many themes of the MTD (see also Section 4.4). Furthermore, we selected works from the standard repertoire, where many performances are easily available, such that users of the MTD can add further audio occurrences to the dataset. Due to these considerations, the MTD is not musically balanced in a stricter sense (e.g., in terms of periods or genres). Even though this imbalance may be problematic for musicological studies, the variety

of themes in the MTD is useful for MIR applications (such as described in Section 6).

For the 2067 themes, we provide complete coverage of all modalities and metadata, described in the following subsections.

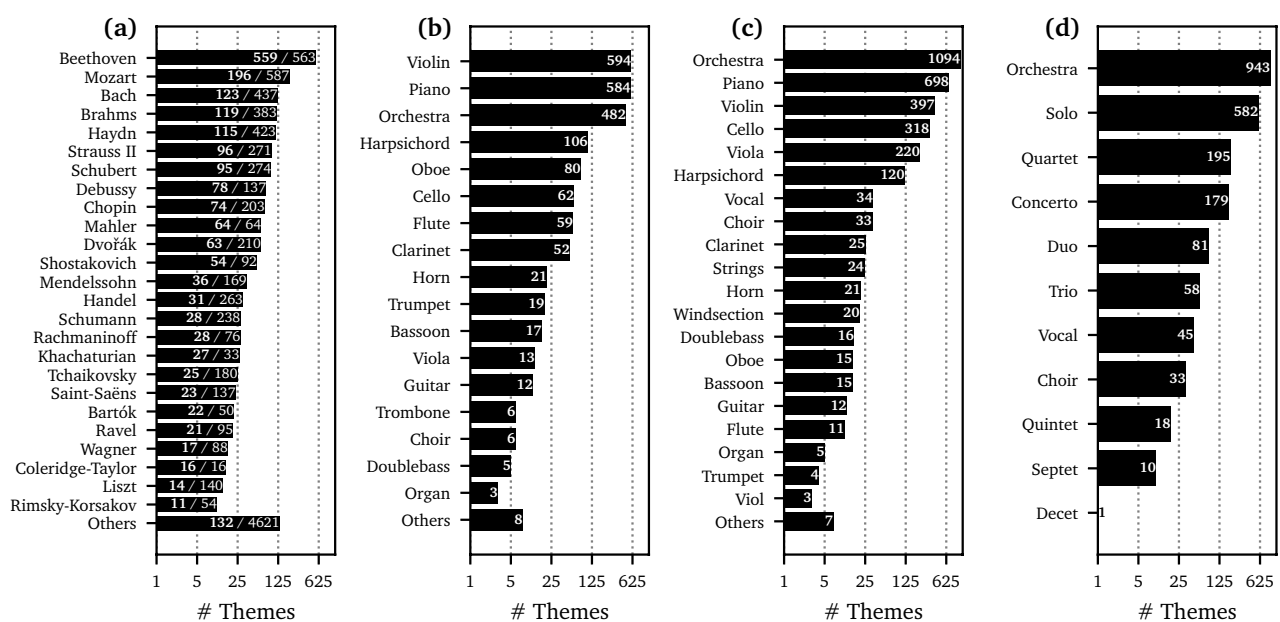
#### 4.2 Metadata

In the MTD, we provide some detailed metadata on the composer, work, recording, and musical characteristics of the themes, which are described in the following.

As a main contribution, we identified the catalog number for each musical work that contains a theme. Consistent catalog-based work information is not specified in the BM book. For example, for Beethoven's Fifth Symphony, the Opus number 67 is given in the BM book. But as another example, for all works by J. S. Bach, the BM book does not provide a catalog number, such as BWV 1046 for the first Brandenburg Concerto. In the case of the well-known BWV catalog of works by Bach from 1950, this was impossible because it was first published after the BM book from 1948. For our MTD work identifiers, we always use standard work catalogs when available. These identifiers relate to the movement level for multi-movement works. For example, the work identifier for the first movement in Bach's first Brandenburg Concerto is BWV1046-01.

Overall, we have 54 composers in our dataset. **Figure 3a** shows a bar graph of the number of themes per composer. In this figure, we only show composers with more than ten themes. We see that the most prominent composer of the dataset is Beethoven, with 559 themes. The second most common composer is Mozart, with 196 themes, followed by Bach, Brahms, and Haydn with a little more than 100 themes each.

Our metadata also contains several annotations regarding musical instrumentation. We show the distribution of these annotations in the **Figure 3b-d**. The bar graphs



**Figure 3:** Various bar graphs for metadata of the themes. Numbers of themes on horizontal axes (logarithmic) are shown per (a) composer (non-bold numbers after the slash indicate total number of themes in BM book), (b) theme instrumentation, (c) work instrumentation, and (d) ensemble type.

show the number of themes per instrumentation of the theme melody, per instrumentation of the entire work, and per ensemble type. The BM book specifies the instrumentation only to a certain degree. For example, a keyboard work by J. S. Bach could be played by a piano or by a harpsichord. **Figure 3b** shows the instrumentation of the themes as they occur in our selected recordings (the recordings are explained in Section 4.4). Note that a theme can be played by more than one instrument. That is why the overall bar graph count amounts to more than 2067 themes. We see that the dominating theme instruments are violin, piano, and orchestral tutti. But there are also several themes played by other instruments, such as harpsichord, oboe, cello, flute, or clarinet. It may be surprising that the choir appears as a category because the BM book only covers instrumental musical works. After the original BM book, Barlow and Morgenstern also published a separate dictionary of vocal music (Barlow and Morgenstern, 1976), which we did not use for the MTD. However, in the BM book there are a few exceptions, such as the choral finale of Beethoven’s Ninth Symphony.

**Figure 3c** shows a bar graph of the instrumentation of the musical works. This instrumentation now refers to the entire piece of music and not just to the instruments which play the themes. For example, the third theme in Beethoven’s Fifth Symphony is played by the horn, but the instrumentation of the corresponding work is the entire orchestra. Finally, **Figure 3d** shows a bar graph of the ensemble type. For nearly half of the themes (943) the ensemble is the full orchestra, and more than a quarter (582) of them are from solo pieces. Another quarter of the themes are played by other types of ensembles, such as quartets or duos.

As a further musical characteristic, we also provide annotations for the musical texture of the music segments where the themes appear. Studying texture is a challenging topic on its own (Giraud et al., 2014). Following the textbook by Benward and Saker (2009), we annotate the texture of the themes according to the standard categories of monophony, homophony, and polyphony. A monophonic texture consists of a single melodic line (possibly doubled by octaves). A homophonic texture is made up of a melody and an accompaniment. A polyphonic texture comprises two or more independent melodic lines. According to Benward and Saker (2009), there is also the fourth category of homorhythmic texture with a similar rhythm in all voices. However, for our annotations, we include cases of this category into homophony. Of course, more than one texture can appear in a single theme. For example, the beginning of a fugue often starts by presenting a musical theme (called the subject in the case of fugues) in a monophonic way. Still, the texture turns polyphonic as soon as another voice joins in (which can be before the end of the subject). In such cases, we assign multiple categories. Even though our coarse categorizations have to be taken with care, they may serve, e.g., as a guiding principle when evaluating MIR applications using the MTD. One may decide on a different category for border cases, but the annotations should be appropriate for unambiguous cases.

**Table 2** shows all metadata fields of the MTD with a short description of each. The upper part of the table shows various identifiers, and the lower part descriptive metadata fields. Some of the entries relate to the audio recordings, which are described in Section 4.4.

#### 4.3 Symbolic Encodings

We provide various symbolic encodings of the musical themes. As one main contribution, we newly engraved the themes with the sheet music editor Sibelius. We denote this encoding by SCORE. Beyond the Sibelius files, our dataset contains exports to PDF, MIDI, MusicXML, and CSV. Among these formats, the only non-standard music format is our CSV representation, which encodes each note’s start, duration, and pitch in a simple way.

While building the MTD dataset, we worked with the original EDM files parallel to engraving the SCORE

**Table 2:** Overview of all metadata contained in the MTD.

Field	Description
MTDID	Identifier, used in the MTD
BMID	Identifier, from original BM book
EDMID	Identifier, used in the EDM
ComposerID	Identifier, based on composer’s name
WorkID	Identifier, usually based on catalog number
PerformanceID	Identifier, based on main performer of recording
CollectionID	Identifier, based on album collection
LabelID	Identifier, based on recording label
WC MID	Internal ID for audio recording
MusicBrainzID	MusicBrainz release ID for album collection
ComposerBirth	Composer’s year of birth
ComposerDeath	Composer’s year of death
WorkTitle	Sub-title, nickname, or non-numeric title for musical work
ThemeLabelBM	Label for theme, from original BM book
ThemeInstruments	Instrument (s) playing the theme
WorkInstruments	Instrument (s) of the musical work
Ensemble	Ensemble type
Polyphony	Indication of musical texture
NameCD	CD name in album collection
NameTrack	Track name in the CD of the album collection
StartTime	Start time of theme occurrence in audio recording
EndTime	End time of theme occurrence in audio recording
MidiTransposition	Pitch transposition difference between recording and symbolic encoding
Comment	Textual comment

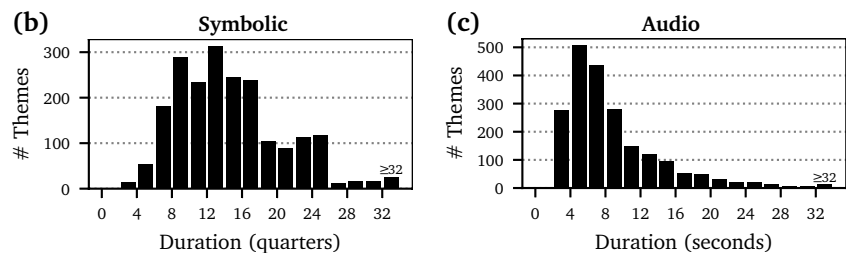
versions. For this reason, we also provide the original EDM files of the MTD themes (denoted by EDM-orig). The EDM files often contain errors, e.g., wrong pitches and rhythms, or missing and incorrect ornaments. As one of our contributions, we consistently corrected all wrong pitches in the MIDI files. We also fixed some rhythm and ornament errors without being comprehensive here. We provide our corrected EDM files (denoted by EDM-corr) additionally to the original ones.

In some cases, we also found errors in the original engraving of the themes, where the BM book unintentionally deviates from the corresponding musical work's score. In these cases, we corrected the errors (for SCORE and EDM-corr) to be consistent with the score.<sup>5</sup> A general principle of the MTD is that a theme is always a monophonic note sequence. As for the BM book, however, there are few exceptions, where the theme is notated in a polyphonic way. For these examples, we decided on a monophonic representation of the theme (for EDM-corr).<sup>6</sup> We also always assume that a theme is a continuous sequence of notes and rests. In almost all cases, this assumption is fulfilled in the BM dictionary. However, for a few instances in the BM book, a single identifier is used to denote two themes. In these cases, the themes are either separated by a gap of several measures or overlapping in time (we discussed such a situation in Section 3, **Figure 2d**). For these exceptions, we deviate from the BM book and give individual identifiers for the themes.<sup>7</sup>

Both EDM-orig and EDM-corr are symbolic representations that do not focus on the sheet music layout. In contrast, our new SCORE engravings are created with Sibelius and can be used for graphical purposes.

We now discuss some statistical aspects of the symbolic encodings of the MTD themes. **Figure 4a** shows some general statistics of the dataset. The average duration of the 2067 themes is 14.2 quarter notes. The entries for the audio representations will be explained in Section 4.4. **Figure 4b** shows a histogram of the themes' durations (based on EDM-corr), measured in quarter notes. Most themes have a length of 6 to 18 quarter notes. However, a few themes have a short duration of below four quarter notes or a long duration of more than 30 quarter notes. Of course, in different time signatures, quarter notes have different meanings. In the bar graph of **Figure 5**, we show the number of themes for different time signatures. In the case of multiple time signatures for a single theme, we only use the first one for this figure.<sup>8</sup>

<b>(a)</b>	
# Themes	2067
# Composers	54
<b>Symbolic</b>	
$\Sigma$ Dur. (quarters)	29359.8
$\varnothing$ Dur. (quarters)	$14.2 \pm 6.0$
<b>Audio</b>	
$\Sigma$ Dur. (hh:mm:ss)	4:57:48.3
$\varnothing$ Dur. (seconds)	$8.6 \pm 5.6$



**Figure 4:** Dataset overview. **(a)** Main statistics of dataset. Average information ( $\varnothing$ ) given as mean  $\pm$  standard deviation. **(b)** Histogram of theme durations in quarter notes and **(c)** in seconds.

#### 4.4 Audio Recordings

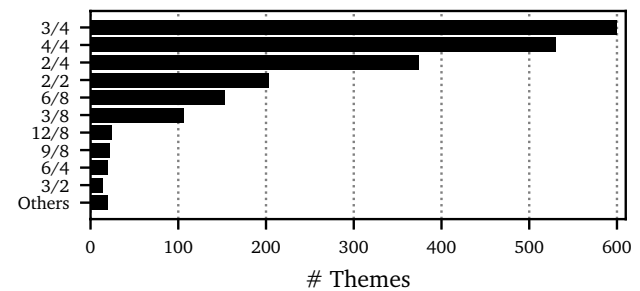
For each musical work of the MTD themes, we selected a recording contained in a comprehensive CD album collection, where a single collection typically covers many themes in the MTD. Overall, the 2067 occurrences are to be found in 61 album collections. To identify these collections, we provide MusicBrainz IDs (Swartz, 2002) in our metadata. The full list of MusicBrainz IDs is also available as a MusicBrainz user collection.<sup>9</sup>

For each theme, we decided on one prominent occurrence in the recording. Typically this is the first occurrence of the theme. Then, we annotated the beginning and end of the theme occurrence in the recording. We generated audio excerpts corresponding to these occurrences, which are a central component of MTD. Since these excerpts are concise, we consider providing the audio files as fair use. As shown in the table of **Figure 4a**, the occurrences have an average duration of 8.6 seconds. **Figure 4c** shows a histogram of the durations of the audio snippets. This distribution is strongly skewed, with most themes having a duration between 4 and 6 seconds.

We observed that several theme occurrences are transposed compared to the versions in the BM book. The reason for this may be that the entire recording is transposed or that the reference for the BM book is a different occurrence of the theme. We consistently annotated the transposition differences in semitones.

#### 4.5 Alignment Data

As a further main contribution, we manually aligned the symbolic themes (EDM-corr) to the audio recordings. To support this process, we created a custom web interface for editing the alignment and listening to the result in the form of a superposition of the audio snippet and a synthesized version of the aligned theme (more details in



**Figure 5:** Time signatures in the MTD.

Section 5). This sonification helped us to check the edited alignment. We created alignment paths that consist of pairs of corresponding time points in the audio and the symbolic representations, respectively. They enabled us to synchronize the symbolic representations (EDM-corr) to the audio recordings. We additionally provide these modified symbolic representations (denoted by EDM-alig).

#### 4.6 Directory and File Structure

The structure of the MTD with regard to its modalities yields a natural directory structure. There are several directories on the top level, each containing a different data modality for all themes. **Table 3** lists the directories of the MTD. All files inside the directories are consistently named using the MTD, composer, and work identifiers, e.g., MTD1066\_Beethoven\_Op067-01.

### 5. Interfaces and Tools

We provide access to the MTD in three different ways.<sup>10</sup> The first way is to download an archive with the raw data (see also **Table 3**). The second way is a website that presents the different data modalities of the dataset. Third, we provide a Jupyter notebook containing Python code for parsing, visualizing, and sonifying the data. In this section, we introduce the website and the Jupyter notebook. Furthermore, we also describe our custom tool for aligning the symbolic and audio representations.

On the website's start page, we list all 2067 themes of MTD in a table, see **Figure 6a**. For each theme, there is a dedicated subpage. The subpages can be accessed by clicking on the MTD ID in the table. For example, in **Figure 6a**, we highlighted the link for the MTD ID 1066. **Figure 6b** shows a screenshot of the subpage of this

theme. On the top, we display three variants of graphical sheet music (EDM-orig, EDM-corr, and SCORE) along with corresponding MIDI playback buttons.<sup>11</sup> Note that the images for EDM-orig and EDM-corr are generated from MIDI (using the software MuseScore), which is not meant for graphical sheet music rendering. Even though the corresponding piano roll representations are accurate, the graphical rendition may not be musically meaningful. However, the images of the SCORE versions are of consistent quality because we directly engraved the sheet music using Sibelius. Furthermore, we offer two versions of the theme's occurrence in an audio recording: the first version is the respective audio excerpt, and the second one is a mixture of the same excerpt and a sonification of the aligned theme (EDM-alig). Finally, we show a table with the metadata.

The website is an easy way to explore the MTD, but it is static. In contrast to that, our Jupyter notebook allows for interaction with the data. We build upon standard Python packages such as `pretty_midi` (Raffel and Ellis, 2014) and `librosa` (McFee et al., 2015), and use the Jupyter framework, which is common in the MIR community (Müller and Zalkow, 2019). **Figure 6c** shows a screenshot of the notebook. In this part of the notebook, we first load the audio snippet for the first theme of Beethoven's Fifth Symphony. Then, we compute a spectral representation with logarithmic frequency spacing (Schörkhuber and Klapuri, 2010). We visualize the spectral matrix as a grayscale image. Because the frequency bandwidth corresponds to a semitone, the frequency axis can also be used as the pitch axis of a piano roll representation. Using the aligned symbolic music encoding (EDM-alig), we can superimpose a piano roll visualization (red color) on the image. Thus, we highlight the spectral bins that coincide with the fundamental frequencies of the theme's notes.

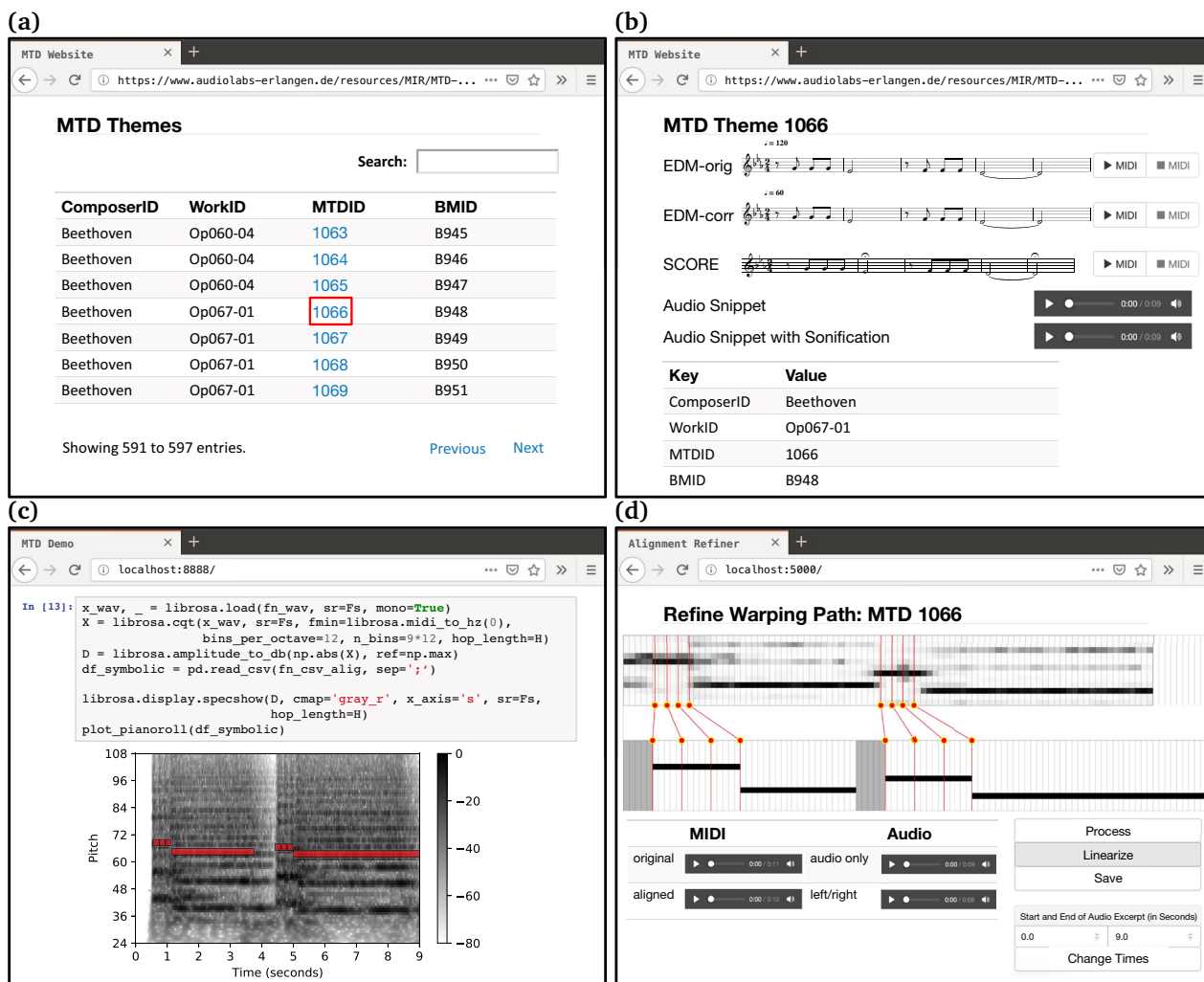
We also provide our custom alignment tool for manually creating the alignments between the audio snippets and the MIDI files. This web tool is not a general-purpose product, but its source code may be useful for the MTD users who want to further refine the alignments or add further themes to the dataset. The tool is a client-server application that uses the Python web development package `Flask`<sup>12</sup> for its back end and the Javascript canvas library `Fabric.js`<sup>13</sup> for the graphical elements in the web front end. **Figure 6d** shows a screenshot of this web tool, which visualizes chromagrams for the audio snippet (upper) and the MIDI file (lower). In between, there are red lines that are connected to the chromagrams by red dots. These red dots indicate the time positions of the theme's note onsets in the representations of the respective chromagrams. In this way, the connecting lines define an alignment path between both representations.

The red dots are fixed in the MIDI version and flexible in the audio version. By moving the red dots in the audio version, the user can specify the onsets' time positions and, therefore, change the alignment. The alignment for all time points not related to note onsets is then obtained by linear interpolation. After clicking on the button for processing, a sonification is automatically generated that helps to evaluate the overall alignment accuracy. Finally,

**Table 3:** Overview of the MTD directory structure.

Directory	Description	Format
data_EDM-orig_CSV		CSV
data_EDM-orig_IMG	Original EDM files	PDF
data_EDM-orig_MID		MIDI
data_EDM-corr_CSV		CSV
data_EDM-corr_IMG	Corrected EDM files	PDF
data_EDM-corr_MID		MIDI
data_EDM-alig_CSV	Aligned EDM files	CSV
data_EDM-alig_MID		MIDI
data_SCORE_CSV		CSV
data_SCORE_IMG		PDF
data_SCORE_MID	SCORE files	MIDI
data_SCORE_SIB		Sibelius
data_SCORE_XML		MusicXML
data_ALIGNMENT	Alignment data	CSV
data_AUDIO	Audio snippets	WAV
data_META	Metadata	JSON





**Figure 6:** Screenshots of our web-based interfaces. (a) Overview table of web page. (b) Subpage for the theme with MTD ID 1066. (c) Jupyter notebook. (d) Alignment tool.

the alignment can be saved as a CSV file (same format as data\_ALIGNMENT in the MTD). We also provide further Python scripts that use this CSV file to generate the time-aligned symbolic formats (data\_EDM-alig\_CSV and data\_EDM-alig\_MID).

There are many tasks where temporal annotations of high resolution are beneficial. As future work, our alignment tool may be modified to be useful for other applications, such as bird song recognition (Morfi et al., 2019) and audio event detection (Gemmeke et al., 2017).

## 6. Applications and Future Work

Due to its multimodal nature, the MTD can be useful for many MIR tasks. In this section, we discuss some possible applications and indicate future work directions.

Being inspired by the BM dictionary, our dataset offers links between the sheet music in the printed book and new digital engravings. This is an interesting testbed for optical music recognition (OMR) (Rebello et al., 2012; Byrd and Simonsen, 2015; Calvo-Zaragoza et al., 2020). On the one hand, the monophonic themes constitute a relatively simple OMR scenario. On the other hand, the difference between modern versions and old engravings from the 1940s can be challenging. Balke et al. (2015) already

presented a study using OMR for the printed BM book. They report on retrieval experiments, where they aimed at finding relevant MIDI files in the EDM collection. The queries were generated using OMR and OCR processing of the BM book.

The new engravings of the MTD go along with symbolic representations. The correspondences between these symbolic representations and the audio occurrences open up possibilities for exciting cross-modal retrieval scenarios. For example, in one such task, using a theme’s symbolic encoding as a query, the aim is to identify all relevant audio recordings that contain an occurrence of the query theme in an audio database. In this retrieval scenario, main challenges are the differences in modality (symbolic vs. audio) and musical characteristics (monophonic vs. polyphonic). Several studies approached this task, and some already used preliminary versions of what have become the MTD (Balke et al., 2016; Zalkow et al., 2019; Zalkow and Müller, 2020). Using a retrieval approach based on enhanced chroma features and local alignment techniques, Zalkow et al. (2019) reported results, where they found a relevant recording as the top match for roughly 75 percent of the MTD themes in a database of 1114 audio files. As future work, the audio occurrences

and the aligned symbolic versions of the MTD may serve as a training set for data-driven approaches to further improve on these results.

Another task that also involves symbolic and audio representations is to estimate the fundamental frequency ( $F_0$ ) contours of the themes in the audio occurrences. This task is closely related to melody estimation (Salamon et al., 2014). A typical approach for this task is to first compute a pitch salience function, which is a time–frequency representation where the melody’s  $F_0$  components are enhanced, and other components are attenuated. In the second step,  $F_0$  contours are tracked in the salience representation. Bittner et al. (2017) proposed a machine learning strategy to compute a salience representation using a fully convolutional neural network. Using such techniques, one may learn a salience function for musical themes employing the audio occurrences and the aligned symbolic representations of the MTD as a training set.

In even more challenging scenarios, one may aim to find direct correspondences between the audio recordings and the sheet music images, without utilizing symbolic representations. If the alignment of these modalities is performed online, the application is also known as score following. There are first works to learn representations for score following with data-driven approaches in an end-to-end fashion (Dorfer et al., 2016). The correspondences between sheet music images and audio excerpts of the MTD can be used as additional training and test data for such approaches. A further challenge here is the monophonic–polyphonic discrepancy between the score–audio pairs.

A general problem in data-driven approaches is the need for aligned training data. The correspondences between audio excerpts and symbolic encodings in the MTD may serve different purposes since they are both weakly-aligned (theme level) and strongly-aligned (note level). For example, the MTD may serve as a testbed to develop and evaluate alignment approaches within deep learning frameworks. An example of such an approach is the connectionist temporal classification (CTC) loss (Graves et al., 2006), which can be used to train a neural network with weakly aligned data. Stoller et al. (2019) used this loss to align music recordings to textual lyrics, where they only used weakly aligned audio–lyrics pairs for training. In another study, the CTC loss was used for OMR of monophonic music (Calvo-Zaragoza and Rizo, 2018). In this context, one may use the weak and strong alignments of the MTD to develop and test CTC-based learning approaches within a challenging musical scenario. A first study using the CTC loss with the MTD was presented by Zalkow and Müller (2020).

As another application, the MTD alignments between the symbolic scores and audio recordings can be used for detailed analyses of the music performances (Lerch et al., 2019), in particular in terms of tempo and timing (Dixon, 2001). For example, MIR researchers used alignment information to compute tempo curves that visualize the tempo change throughout the performance of a musical piece (Müller et al., 2009). A web tool for tempo comparison was created by Peachnote.<sup>14</sup> Using

the MTD, one may analyze the tempo characteristics of performances of musical themes.

Given the rich metadata of the MTD, the dataset may be valuable for various music recommendation and classification tasks. Examples are composer classification (Verma and Thickstun, 2019) and instrument identification (Essid et al., 2006).

In summary, the MTD offers a rich and diverse cross-modal dataset for music processing. The dataset may trigger future research directions to further explore the potential of musical themes and multimodality for MIR research.

## Notes

- <sup>1</sup> A list of datasets for MIR is to be found at <http://ismir.net/resources/datasets/>.
- <sup>2</sup> <https://web.archive.org/web/20160209045946/http://www.multimedialibrary.com/barlow/index.asp>.
- <sup>3</sup> <http://www.themefinder.org>.
- <sup>4</sup> <http://www.rism.info>.
- <sup>5</sup> Out of 2067, this affects 30 themes, namely the ones with the MTD-IDs 0770, 1033, 1109, 1143, 1484, 1501, 1737, 1742, 1788, 2609, 2619, 2966, 3944, 4287, 4305, 5323, 5566, 5753, 6008, 6840, 7636, 7670, 8111, 8137, 8141, 8355, 8549, 8560, 9130, and 9516-2.
- <sup>6</sup> Out of 2067, this affects 18 themes, namely the ones with the MTD-IDs 0261, 0389, 0411, 0429, 0433, 0435, 0765, 0957, 1035, 1067, 1162, 2583, 2597, 6803, 7232, 7718, 8470, and 8780.
- <sup>7</sup> This is the case for 9 BM IDs, which we split into 18 MTD IDs, namely 0739-1/2, 0755-1/2, 0768-1/2, 2255-1/2, 2332-1/2, 2338-1/2, 2631-1/2, 9164-1/2, and 9516-1/2.
- <sup>8</sup> Out of 2067 themes, 16 have multiple time signatures, namely the ones with the MTD-IDs 0579, 0585, 3622, 5332, 5337, 5354, 5364, 6777, 7230, 8052, 8131, 8147, 8150, 8782, 8794, and 8795.
- <sup>9</sup> <https://musicbrainz.org/collection/57faf42e-045e-415d-9f1d-dd5d0699b689>.
- <sup>10</sup> <https://www.audiolabs-erlangen.de/resources/MIR/MTD>.
- <sup>11</sup> The MIDI playback is realized with the Javascript libraries MidiPlayerJS (<https://github.com/grimmdude/MidiPlayerJS>), and soundfont-player (<https://github.com/danigb/soundfont-player>).
- <sup>12</sup> <https://flask.palletsprojects.com/>.
- <sup>13</sup> <http://fabricjs.com/>.
- <sup>14</sup> <https://www.tuttitempi.com>.

## Acknowledgments

This work was supported by the German Research Foundation (DFG MU 2686/11-1, DFG MU 2686/12-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS. We thank Lena Krauß, Lukas Lamprecht, Anna-Luisa Römling, and Quirin Seilbeck for helping us with the annotations.

## Competing Interests

The authors have no competing interests to declare.

## References

- Arzt, A., & Lattner, S.** (2018). Audio-to-score alignment using transposition-invariant features. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 592–599. Paris, France.
- Balke, S., Achankunju, S. P., & Müller, M.** (2015). Matching musical themes based on noisy OCR and OMR input. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 703–707. Brisbane, Australia. DOI: <https://doi.org/10.1109/ICASSP.2015.7178060>
- Balke, S., Arifi-Müller, V., Lamprecht, L., & Müller, M.** (2016). Retrieving audio recordings using musical themes. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 281–285. Shanghai, China. DOI: <https://doi.org/10.1109/ICASSP.2016.7471681>
- Barlow, H., & Morgenstern, S.** (1975). *A Dictionary of Musical Themes*. Crown Publishers, Inc., revised edition.
- Barlow, H., & Morgenstern, S.** (1976). *A Dictionary of Opera and Song Themes*. Crown Publishers, Inc., revised edition.
- Benetos, E., Dixon, S., Duan, Z., & Ewert, S.** (2019). Automatic music transcription: An overview. *IEEE Signal Processing Magazine*, 36(1), 20–30. DOI: <https://doi.org/10.1109/MSP.2018.2869928>
- Benward, B., & Saker, M.** (2009). *Music in Theory and Practice*. McGraw Hill, 8th edition.
- Berman, T., Downie, J. S., & Berman, B.** (2006). Beyond error tolerance: Finding thematic similarities in music digital libraries. In *Proceedings of the European Conference on Digital Libraries (ECDL)*, pages 463–466. Alicante, Spain. DOI: [https://doi.org/10.1007/11863878\\_44](https://doi.org/10.1007/11863878_44)
- Bittner, R. M., McFee, B., Salamon, J., Li, P., & Bello, J. P.** (2017). Deep salience representations for F0 tracking in polyphonic music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 63–70. Suzhou, China.
- Bittner, R. M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., & Bello, J. P.** (2014). MedleyDB: A multitrack dataset for annotation-intensive MIR research. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 155–160. Taipei, Taiwan.
- Bosch, J. J., Marxer, R., & Gómez, E.** (2016). Evaluation and combination of pitch estimation methods for melody extraction in symphonic classical music. *Journal of New Music Research*, 45(2), 101–117. DOI: <https://doi.org/10.1080/09298215.2016.1182191>
- Byrd, D., & Simonsen, J. G.** (2015). Towards a standard testbed for optical music recognition: Definitions, metrics, and page images. *Journal of New Music Research*, 44(3), 169–195. DOI: <https://doi.org/10.1080/09298215.2015.1045424>
- Calvo-Zaragoza, J., Hajič, J., Jr, & Pacha, A.** (2020). Understanding optical music recognition. *ACM Computing Surveys*, 53(4). DOI: <https://doi.org/10.1145/3397499>
- Calvo-Zaragoza, J., & Rizo, D.** (2018). End-to-end neural optical music recognition of monophonic scores. *Applied Sciences*, 8(4). DOI: <https://doi.org/10.3390/app8040606>
- Changizi, M.** (2011). *Harnessed: How Language and Music Mimicked Nature and Transformed Ape to Man*. BenBella Books.
- Diet, J., & Gerritsen, M.** (2013). Encoding, searching, and displaying of music incipits in the RISM-OPAC. In *Proceedings of the Music Encoding Conference (MEC)*, pages 11–14. Mainz, Germany.
- Dixon, S.** (2001). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30, 39–58. DOI: <https://doi.org/10.1076/jnmr.30.1.39.7119>
- Dorfer, M., Arzt, A., & Widmer, G.** (2016). Towards score following in sheet music images. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 789–795. New York, USA.
- Dorfer, M., Hajič, J., Jr, Arzt, A., Frostel, H., & Widmer, G.** (2018). Learning audio-sheet music correspondences for cross-modal retrieval and piece identification. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 1(1), 22–31. DOI: <https://doi.org/10.5334/timsir.12>
- Drabkin, W.** (2001). Theme. In *Grove Music Online*. Oxford University Press. DOI: <https://doi.org/10.1093/gmo/9781561592630.article.27789>
- Emiya, V., Badeau, R., & David, B.** (2010). Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6), 1643–1654. DOI: <https://doi.org/10.1109/TASL.2009.2038819>
- Essid, S., Richard, G., & David, B.** (2006). Instrument recognition in polyphonic music based on automatic taxonomies. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1), 68–80. DOI: <https://doi.org/10.1109/TSA.2005.860351>
- Gemmeke, J. F., Ellis, D. P. W., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M., & Ritter, M.** (2017). Audio Set: An ontology and human-labeled dataset for audio events. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 776–780. New Orleans, Louisiana, USA. DOI: <https://doi.org/10.1109/ICASSP.2017.7952261>
- Giraud, M., Levé, F., Mercier, F., Rigaudière, M., & Thorez, D.** (2014). Towards modeling texture in symbolic data. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 59–64. Taipei, Taiwan.
- Goto, M.** (2004). Development of the RWC music database. In *Proceedings of the International Congress on Acoustics (ICA)*, pages 553–556.



- Graves, A., Fernández, S., Gomez, F. J., & Schmidhuber, J.** (2006). Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, pages 369–376. Pittsburgh, Pennsylvania, USA. DOI: <https://doi.org/10.1145/1143844.1143891>
- Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C.-Z. A., Dieleman, S., Elsen, E., Engel, J., & Eck, D.** (2019). Enabling factorized piano music modeling and generation with the MAESTRO dataset. In *Proceedings of the International Conference on Learning Representations (ICLR)*. New Orleans, Louisiana, USA.
- Joder, C., Essid, S., & Richard, G.** (2013). Learning optimal features for polyphonic audio-to-score alignment. *IEEE Transactions on Audio, Speech & Language Processing*, 21(10), 2118–2128. DOI: <https://doi.org/10.1109/TASL.2013.2266794>
- Kornstädt, A.** (1998). Themefinder: A web-based melodic search tool. *Computing in Musicology*, 11.
- Lerch, A., Arthur, C., Pati, A., & Gururani, S.** (2019). Music performance analysis: A survey. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 33–43. Delft, The Netherlands.
- London, J.** (2013). Building a representative corpus of classical music. *Music Perception*, 31(1), 68–90. DOI: <https://doi.org/10.1525/mp.2013.31.1.68>
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O.** (2015). Librosa: Audio and music signal analysis in Python. In *Proceedings the Python Science Conference*, pages 18–25. Austin, Texas, USA. DOI: <https://doi.org/10.25080/Majora-7b98e3ed-003>
- Meek, C., & Birmingham, W. P.** (2001). Thematic extractor. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*. Bloomington, Indiana, USA.
- Morfi, V., Bas, Y., Pamula, H., Glotin, H., & Stowell, D.** (2019). Nips4bplus: A richly annotated birdsong audio dataset. *PeerJ Computer Science*, 5, e223. DOI: <https://doi.org/10.7717/peerj-cs.223>
- Müller, M., Arzt, A., Balke, S., Dorfer, M., & Widmer, G.** (2019). Cross-modal music retrieval and applications: An overview of key methodologies. *IEEE Signal Processing Magazine*, 36(1), 52–62. DOI: <https://doi.org/10.1109/MSP.2018.2868887>
- Müller, M., Konz, V., Bogler, W., & Arifi-Müller, V.** (2011). Saarland Music Data (SMD). In *Late-Breaking and Demo Session of the 12th International Conference on Music Information Retrieval (ISMIR)*. Miami, USA.
- Müller, M., Konz, V., Scharfstein, A., Ewert, S., & Clausen, M.** (2009). Towards automated extraction of tempo parameters from expressive music recordings. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 69–74. Kobe, Japan.
- Müller, M., Kurth, F., & Röder, T.** (2004). Towards an efficient algorithm for automatic score-to-audio synchronization. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 365–372. Barcelona, Spain.
- Müller, M., & Zalkow, F.** (2019). FMP notebooks: Educational material for teaching and learning fundamentals of music processing. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 573–580. Delft, The Netherlands.
- Parsons, D.** (1975). *The Directory of Tunes and Musical Themes*. S. Brown.
- Pollastri, E., & Simoncelli, G.** (2001). Classification of melodies by composer with hidden Markov models. In *Proceedings of the International Conference on WEB Delivering of Music (WEDELMUSIC)*, pages 88–95. Florence, Italy. DOI: <https://doi.org/10.1109/WDM.2001.990162>
- Prechelt, L., & Typke, R.** (2001). An interface for melody input. *ACM Transactions on Computer-Human Interaction*, 8(2), 133–149. DOI: <https://doi.org/10.1145/376929.376978>
- Raffel, C., & Ellis, D. P. W.** (2014). Intuitive analysis, creation and manipulation of MIDI data with pretty\_midi. In *Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*. Taipei, Taiwan.
- Rebelo, A., Fujinaga, I., Paszkiewicz, F., Marcal, A. R. S., Guedes, C., & Cardoso, J. S.** (2012). Optical music recognition: State-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, 1(3), 173–190. DOI: <https://doi.org/10.1007/s13735-012-0004-6>
- Reti, R.** (1951). *The Thematic Process in Music*. The Macmillan Company.
- Rosenzweig, S., Scherbaum, F., Shugliashvili, D., Arifi-Müller, V., & Müller, M.** (2020). Erkomaishvili Dataset: A curated corpus of traditional Georgian vocal music for computational musicology. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 3(1), 31–41. DOI: <https://doi.org/10.5334/tismir.44>
- Salamon, J., Gómez, E., Ellis, D. P. W., & Richard, G.** (2014). Melody extraction from polyphonic music signals: Approaches, applications, and challenges. *IEEE Signal Processing Magazine*, 31(2), 118–134. DOI: <https://doi.org/10.1109/MSP.2013.2271648>
- Salamon, J., Serrà, J., & Gómez, E.** (2013). Tonal representations for music retrieval: From version identification to query-by-humming. *International Journal of Multimedia Information Retrieval*, 2(1), 45–58. DOI: <https://doi.org/10.1007/s13735-012-0026-0>
- Schörkhuber, C., & Klapuri, A. P.** (2010). Constant-Q transform toolbox for music processing. In *Proceedings of the Sound and Music Computing Conference (SMC)*, pages 322–329. Barcelona, Spain.
- Serra, X.** (2014). Creating research corpora for the computational study of music: The case of the CompMusic project. In *Proceedings of the AES International Conference on Semantic Audio*. London, UK.
- Simonetta, F., Chacón, C. E. C., Ntalampiras, S., & Widmer, G.** (2019). A convolutional approach to melody line identification in symbolic scores. In



*Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 924–931. Delft, The Netherlands.

- Simonton, D. K.** (1980). Thematic fame, melodic originality, and musical zeitgeist: A biographical and transhistorical content analysis. *Journal of Personality and Social Psychology*, 38(6), 972–983. DOI: <https://doi.org/10.1037/0022-3514.38.6.972>
- Simonton, D. K.** (1991). Emergence and realization of genius: The lives and works of 120 classical composers. *Journal of Personality and Social Psychology*, 61(5), 829–840. DOI: <https://doi.org/10.1037/0022-3514.61.5.829>
- Stoller, D., Durand, S., & Ewert, S.** (2019). End-to-end lyrics alignment for polyphonic music using an audio-to-character recognition model. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 181–185. Brighton, UK. DOI: <https://doi.org/10.1109/ICASSP.2019.8683470>
- Swartz, A.** (2002). MusicBrainz: A semantic web service. *IEEE Intelligent Systems*, 17(1), 76–77. DOI: <https://doi.org/10.1109/5254.988466>
- Thickstun, J., Harchaoui, Z., & Kakade, S. M.** (2017). Learning features of music from scratch. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*. Toulon, France.
- Timmers, R., Dibben, N., Eitan, Z., Granot, R., Metcalfe, T., Schiavio, A., & Williamson, V.** (2015). Introduction to the proceedings of ICMEM 2015. In *Proceedings of the International Conference on the Multimodal Experience of Music (ICMEM)*. Sheffield, UK.
- Uitdenbogerd, A. L., & Yap, Y. W.** (2003). Was Parsons right? An experiment in usability of music representations for melody-based music retrieval. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*. Baltimore, Maryland, USA.
- Verma, H., & Thickstun, J.** (2019). Convolutional composer classification. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 549–556. Delft, The Netherlands.
- Volk, A., Wiering, F., & Kranenburg, P. V.** (2011). Unfolding the potential of computational musicology. In *Proceedings of the International Conference on Informatics and Semiotics in Organisations (ICISO)*, pages 137–144. Leeuwarden, The Netherlands.
- Vos, P. G., & Troost, J. M.** (1989). Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception*, 6(4), 383–396. DOI: <https://doi.org/10.2307/40285439>
- Zalkow, F., Balke, S., & Müller, M.** (2019). Evaluating salience representations for cross-modal retrieval of Western classical music recordings. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 331–335. Brighton, United Kingdom. DOI: <https://doi.org/10.1109/ICASSP.2019.8683609>
- Zalkow, F., & Müller, M.** (2020). Using weakly aligned score-audio pairs to train deep chroma models for cross-modal music retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 184–191. Montréal, Canada.

**How to cite this article:** Zalkow, F., Balke, S., Arifi-Müller, V., & Müller, M. (2020). MTD: A Multimodal Dataset of Musical Themes for MIR Research. *Transactions of the International Society for Music Information Retrieval*, 3(1), pp. 180–192. DOI: <https://doi.org/10.5334/tismir.68>

**Submitted:** 23 July 2020

**Accepted:** 21 September 2020

**Published:** 21 October 2020

**Copyright:** © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

**U[** *Transactions of the International Society for Music Information Retrieval* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 