BPSD: A Coherent Multi-Version Dataset for Analyzing the First Movements of Beethoven's Piano Sonatas

TISMIR

DATASET ARTICLES

ubiquity press

JOHANNES ZEITLER
CHRISTOF WEIß
VLORA ARIFI-MÜLLER
MEINARD MÜLLER

*Author affiliations can be found in the back matter of this article

ABSTRACT

This paper introduces the Beethoven Piano Sonata Dataset (BPSD), a multi-version dataset focusing on the first movements of Beethoven's 32 piano sonatas. Recognized as pivotal works in classical music, Beethoven's piano sonatas have profoundly shaped Western classical music, holding a significant place in cultural history. The BPSD includes sheet music in different machine-readable formats and audio recordings from 11 performances, with 4 of them being in the public domain and freely accessible for research purposes. A key feature of BPSD is its coherence, ensuring alignment of all versions on a unified musical timeline and enforcing consistent structures through careful editing of both score and audio representations. The focus and main motivation for the design choices made in BPSD are on the technical and computational level. In particular, BPSD facilitates the assessment of algorithmic approaches in tasks like harmony analysis, structure analysis, music transcription, beat and downbeat estimation, and score following. The dataset's coherence makes it an ideal platform for systematically training and evaluating deep learning methods, shedding light on their robustness and uncovering data biases across different data splits using cross-version strategies for evaluation. To ease applicability for computational approaches, the BPSD is based on various simplifications that may be disputable from a musicological perspective. Rather than providing novel musicological annotations, the main conceptual contribution of BPSD with its measure annotations is to provide a framework for transferring existing annotations from the symbolic to the audio domain. We hope that, as such, BPSD is also useful for the systematic analysis and exploration of Beethoven's piano sonatas, providing insights into their influence on the development of harmony and structure in Western classical music. Beyond research applications, the dataset also holds educational potential, aiding in the preparation and presentation of Beethoven's work to a broader audience through interactive multimedia experiences. This paper delivers a comprehensive overview of the BPSD, highlighting its potential for computational musicology and outlining future research directions.

CORRESPONDING AUTHOR:

Johannes Zeitler

International Audio Laboratories Erlangen, Germany

johannes.zeitler@audiolabs-erlangen.de

KEYWORDS:

dataset; multi-version; multimodal; music processing; music synchronization; computational musicology

TO CITE THIS ARTICLE:

Zeitler, J., Weiß, C., Arifi-Müller, V., & Müller, M. (2024). BPSD: A Coherent Multi-Version Dataset for Analyzing the First Movements of Beethoven's Piano Sonatas. *Transactions of the International Society for Music Information Retrieval,* 7(1), 195–212. DOI: https://doi.org/10.5334/tismir.196

1 INTRODUCTION

The rise of digital technology has brought about significant developments in computer science, recently with a remarkable success of data-driven methods utilizing deep learning (DL) techniques. This progress has reinforced the importance of comprehensive, systematic, and reliable datasets. In the field of Music Information Retrieval (MIR), carefully curated datasets have become indispensable resources, playing a crucial role in advancing tasks such as harmony analysis (Pauwels et al., 2019; Weiß et al., 2020b), beat and downbeat estimation (Böck et al., 2019), structure analysis (Nieto et al., 2020), and music transcription (Benetos et al., 2019). From a technical perspective, high-quality and well-controlled datasets are essential for evaluating and understanding DL methods, providing insights into their robustness, generalization capabilities, and uncovering potential data biases. For an illustrative study in this direction, particularly for the task of local key estimation, we refer the reader to (Weiß et al., 2020b). Beyond the development and evaluation of analysis methods, such datasets enable musicological corpus studies, which allow for systematic analyses and explorations across entire corpora, thus going beyond individual pieces while contributing to an objective methodology for musicological research (Mauch et al., 2015; Nakamura and Kaneko, 2019; Serra, 2014).

As a novel resource for such endeavors, we introduce the Beethoven Piano Sonata Dataset (BPSD), a multi-version corpus focusing on the first movements of Ludwig van Beethoven's 32 piano sonatas.[1] These works rank among the most popular pieces in the classical music repertoire, holding a special place in Western cultural history (Cooper, 2017) and being performed and adapted innumerable times. Beethoven's piano sonatas expand, advance, and further develop the traditional sonata form (Hepokoski and Darcy, 2006), explore innovative harmonic progressions (Damschroder, 2016), and showcase an expressive range and dramatic intensity that elevated the genre and heavily influenced later composers (Tovey, 1931). Our BPSD contributes valuable resources on a work cycle of high musicological relevance, collecting, unifying, and providing a rich set of musical data and annotations. Regarding the primary material (*raw data*), the BPSD includes diverse representations of the sonata movements across various modalities. This encompasses sheet music in different machine-readable formats and audio recordings from 11 performances, with 4 of them being in the public domain (in the EU) and freely accessible for research purposes (see Figure 1 for an overview). Beyond the raw data, our dataset comprises secondary material in the form of carefully curated and linked *annotations*, encompassing measures, beats, note events, global and local keys, absolute and relative chords, and structural elements. Starting with annotations specified on a musical timeline (given in measures) based on sheet music, we employed semi-automated approaches that leverage music synchronization techniques to transfer these annotations to the physical timeline of the recordings (given in seconds). As a result, BPSD includes version-specific annotations for both the sheet music and the 11 audio recordings, forming a systematic and complete three-dimensional data tensor, as illustrated in Figure 1.

Regarding its practical applicability in a computational context, a key feature of the BPSD is its *coherence*, evident not only in the various annotations provided for all versions but also in its unifying approach to musical timeline and structures. Note that depending on the representation and performance, versions of the same piece of music may exhibit different structures, resulting in significant temporal inconsistencies. For example, in sheet music, repeats are notated using repeat signs, possibly with alternative first and second endings. While the enumeration of measures involves assigning a unique number to each measure in sequential order, this order is violated when there are repeats, complicating matters from an algorithmic perspective. Furthermore, different performances may add or omit (e.g., playing the exposition of a sonata form only once) repetitions. Even worse, particularly in older live recordings, there may be severe playing errors, introducing and leaving out several measures.

To address these issues, we ensure temporal coherence in the BPSD by selecting a reference version for all movements and carefully adjusting musical timelines. In particular, we unfold repetitions in musical scores, enumerate measures contiguously, and edit audio recordings by suitably copying or deleting certain parts or sections. This process results in all versions adhering to the same musical timeline, each measure having a unique identifier, and alignments becoming well-defined without gaps across versions. This uniformity greatly enhances the usability of the BPSD for computational approaches, cross-version evaluation, visualization, navigation, and other applications.

The remaining sections of this paper are organized as follows. In Section 2, we discuss related work with a particular focus on other resources conceptually or thematically associated with the BPSD. Section 3 outlines the overall organization of the BPSD and introduces its primary musical material. Section 4 presents an enhanced music synchronization approach used as our central method for aligning different versions and transferring annotations between them. Section 5 delves into the annotations and their main properties. Furthermore, Section 6 explores the potential of the BPSD by discussing three concrete application scenarios, and Section 7 concludes the paper by suggesting possible research directions. The dataset is accessible through a version-controlled repository on the Zenodo platform, and its DOI is 10.5281/zenodo.10847702.[2]

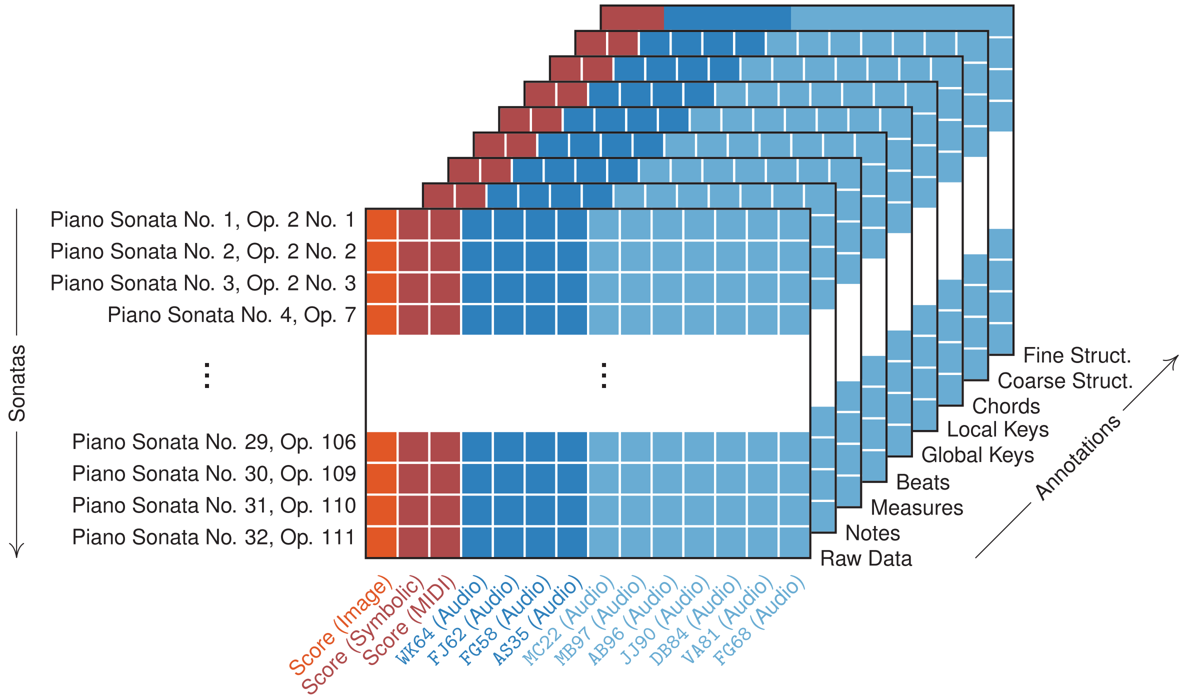

Figure 1 Schematic overview of the BPSD. For the first movements of all 32 piano sonatas, the dataset comprises raw data in different representations (*versions*) such as score images, symbolic score representations, and audio recordings of different performances. For the different versions, we provide time-aligned annotations of measure positions, beats, global and local keys, chords, and structural elements. The score and the four audio versions indicated in dark blue are in the public domain (EU).

2 RELATED DATASETS AND RESOURCES

The MIR community has provided a diverse range of research datasets, which constitute an important basis for advancing music-related computational research. Specifically, in the realm of data-driven methods, datasets play a crucial role as an essential component for training and evaluating DL models in MIR research. The Real World Computing (RWC) music database (Goto, 2004) stands out as one of the first larger music datasets designed specifically for research purposes. It encompasses various genres, including popular, jazz, and classical music, providing both audio recordings and symbolic Musical Instrument Digital Interface (MIDI) representations synchronized through the production processes or the application of alignment techniques. More recent examples include the multitrack dataset MedleyDB (Bittner et al., 2014) designed for source separation and other MIR applications and the Erkomaishvili dataset (Rosenzweig et al., 2020) suitable for ethnomusicological research, both featuring high-quality annotations. In his thoughtful essay, Serra (2014) makes a distinction between unstructured yet annotated *datasets* (or *test corpora*) and curated *research corpora*, identifying five essential criteria (purpose, coverage, completeness, quality, reusability) for a research corpus. The BPSD, along with several others mentioned in this article, can be considered a research corpus based on these criteria. For a comprehensive overview of diverse publicly accessible datasets in the field of MIR, we refer to (Bittner et al., 2019).[3]

In the subsequent discussion, we focus on datasets revolving around piano music, many of which are specifically tailored to automatic music transcription (AMT), see, e.g., (Benetos et al., 2019). These datasets provide audio recordings of musical pieces along with corresponding symbolic encodings that are synchronized with the recordings. Examples include the MIDI Aligned Piano Sounds (MAPS) (Emiya et al., 2010), Saarland Music Data (SMD) (Müller et al., 2011), and Maestro datasets (Hawthorne et al., 2019). Hybrid acoustic-digital player pianos (with MIDI interfaces) or software synthesizers are utilized for these datasets to achieve precise alignments between symbolic and audio representations. The Aligned Scores and Performances (ASAP) dataset is a comprehensive resource that comprises digital musical scores aligned with MIDI and audio performances (Foscarin et al., 2020). Peter et al. (2023) made further refinements to ASAP, incorporating note-wise alignments between scores and performances. MusicNet (Thickstun et al., 2017) is another important dataset comprising chamber music audio recordings (mostly involving the piano) with aligned MIDI information. Finally, we want to mention the Piano Concerto Dataset (PCD) (Özer et al., 2023), which features excerpts of separate piano and orchestra tracks, making it a valuable resource not only for AMT and but also for source separation.

While the BPSD includes piano music, it differs from the previously mentioned piano datasets in two significant ways. Firstly, it focuses on a specific corpus of musical relevance. Secondly, it includes sheet music in various machine-readable formats, along with several carefully aligned performances of the same musical works. This characteristic defines it as what we refer to as a *multi-version* dataset.

Similar multi-version datasets, based on other repertoires, have influenced the design and structure of the BPSD. One notable example is the Schubert Winterreise Dataset (SWD), a multimodal collection featuring various representations and annotations of Franz Schubert's song cycle *Winterreise* (Weiß et al., 2021a). The SWD comprises sheet music in different machine-readable formats, audio recordings of nine performances, and annotations such as musical keys, chords, or structural parts. This dataset enables a systematic study of Schubert's musical language and style in *Winterreise*, also allowing for comparisons of annotations across different annotators and hierarchy levels. Similarly, the Wagner Ring Dataset (WRD) constitutes a multi-modal and multi-version resource focused on the extensive opera cycle *Der Ring des Nibelungen* by Richard Wagner (Weiß et al., 2023). It includes 16 recorded performances of the complete *Ring* cycle, with annotations aligned to a symbolic score representation based on a public-domain piano reduction. The dataset also comprises annotations on measures, key and time signatures, scenes, and singing voice regions.

In all three datasets (SWD, WRD, and our BPSD), a combination of manual work and semi-automated approaches has been employed for creating and transferring annotations between different versions, bridging the symbolic and audio domains. As detailed in Section 6, the multi-version property of datasets creates numerous research opportunities by utilizing the diverse information sources provided through sheet music and the various performances, reflecting a range of different perspectives on the same musical work. This not only facilitates the study of performance aspects (Lerch et al., 2020) but also enables the investigation of algorithmic approaches for various tasks, including automatic music transcription (Benetos et al., 2019), optical music recognition (Calvo-Zaragoza et al., 2020), music synchronization (Müller et al., 2021), cross-modal retrieval (Müller et al., 2019), chord recognition (Pauwels et al., 2019), structure analysis (Nieto et al., 2020), and pattern discovery (Meredith, 2016), just to name a few.

We conclude this section by discussing resources specifically related to our Beethoven scenario, focusing on the first movements of his piano sonatas. In (Chen and Su, 2018), the authors introduce the Beethoven Piano Sonata with Function Harmony (BPS-FH) dataset, which contains expert annotations on chords and harmonic functions for the sonata movements, referring to a symbolic encoding. These annotations include chord symbols and various interrelated chord functions, such as local keys and modulations, chord inversions, secondary chords, and chord quality. Continuing this work, the BPS-Motif dataset described in (Hsiao et al., 2023) offers note-level annotations of motives (characteristic parts of melodic themes) and their occurrences in the sonata movements. These annotations are well-suited for tasks such as motif retrieval and repeated pattern discovery. Finally, we want to mention the work presented in (Jiang and Müller, 2013), which introduces automated methods for analyzing music recordings in sonata form. This approach is evaluated using audio recordings of the Beethoven corpus, incorporating structure annotations following Tovey's analysis of the Beethoven sonatas (Tovey, 1931). Some of these annotations also serve as the foundation for the annotations provided by the BPSD, which have undergone careful revision, extension, and transfer to all versions.

Beyond the piano sonatas, Neuwirth et al. (2018) released a symbolic dataset of Beethoven's string quartets including all movements as symbolic scores together with harmony and phrase annotations. This dataset built the basis for a detailed statistical analysis of harmony (Moss et al., 2019). Furthermore, there is a plenty of other corpora on piano music such as Mozart's piano sonatas (Hentschel et al., 2021), as well as specific corpora on harmony (Gotham et al. 2023b; Hentschel et al., 2023) or structure (Gotham and Ireland, 2019), which contain numerous piano works including Beethoven's piano sonatas. All of these datasets focus on symbolic scores and expert-level annotations while not including audio recordings or multiple versions. Additional related work on datasets and applications is discussed in subsequent sections of this article.

3 DATA ORGANIZATION AND PRIMARY MUSICAL MATERIAL

In this section, we provide an overview of the overall organization of the BPSD. In particular, we outline the global structure, introduce folder and filename conventions, discuss timeline conventions, describe the primary material (score and audio data), and summarize structural modifications for ensuring coherence.

3.1 FOLDER STRUCTURE

The BPSD is organized into four folders: `0_RawData`, `1_Audio`, `2_Annotations`, and `3_Scripts`. The folder `0_RawData` contains the raw audio files, directly extracted from the respective CDs without undergoing any structural modifications. Additionally, this folder includes sheet music representations in various formats. When necessary, editing operations are applied (see Section 3.6), and the modified audio files are stored in the folder `1_Audio`, where every version of a sonata adheres to a consistent structure. All annotations (see Section 5) are gathered in the folder `2_Annotations`. The subdirectories within this folder are named to indicate the type of annotation. Folders with the prefix `ann_score_` contain annotations with a musical timeline (given in measures) as specified by the score. The alignment of these annotations with individual audio versions results in annotations with a physical timeline (given in seconds), which are then stored in folders with the prefix `ann_audio_`. The directory `3_Scripts` contains all Python scripts used in creating the dataset. For a detailed folder structure overview of the BPSD, please refer to Table 1.

3.2 FILENAMING CONVENTION

We encode score-based data using filenames in the format `Beethoven_workID.ext`, where `workID` indicates the work by opus and movement number, and `ext` signifies the type of annotation file. For example, the local key annotations based on a musical timeline for the first movement of Op. 53 (Waldstein Sonata) is stored at `2_Annotations/ann_score_localkey/` with the filename `Beethoven_Op053-01.csv`. For performance-related data, we append the performance ID to the filename. For example, an audio recording of the same sonata played by Wilhelm Kempff in 1964 is stored in `1_Audio/` with the filename `Beethoven_Op053-01_WK64.wav`.

3.3 MUSICAL AND PHYSICAL TIMELINES

The concept of 'time' in music can generally be approached in two ways. There is the concept of *musical time*, specified by the composer in terms of measures and beats, which remains independent of the performance's individual tempo shaping. In contrast, the events in a performed version of the piece can be defined in *physical time*, measured in seconds.

For the BPSD, we adopt the following convention for encoding time positions on the musical timeline. We utilize a decimal representation with the integer measure position before the decimal point and three digits after the decimal point to encode relative positions within a measure. As an example, in a 3/4 time signature, we encode the position of the start of the third beat in measure 137 as 137.666. Physical time is consistently represented in seconds with a precision of three decimal places.

3.4 SCORE DATA

We offer score data in various formats, starting with a scanned version of the 1952 edition by Henle, which was chosen being in the public domain accessible at IMSLP (stored at `score_pdf_scan`). Using this version as a basis, we create sheet music for each movement using the music notation software 'Sibelius' (`score_sibelius_repetitions`), adhering to the measure numbering of the original score. Additionally, we unfold the notated repetitions in Sibelius to achieve a continuous measure count (`score_sibelius_unfolded`). Both the original and unfolded versions are then exported as MusicXML and PDF files, stored at `score_xml_repetitions`, `score_xml_unfolded`, `score_pdf_repetitions`, and `score_pdf_unfolded`, respectively.

To enable playback of the symbolic scores in synthesizers, we export the unfolded sonatas in the MIDI format (`score_MIDI`). To eliminate the need for a MIDI or MusicXML parser and ensure the preservation of measure numbering, we also export all individual notes of a sonata in `.csv` files (in `ann_score_note`). These files contain the start and end of a note event specified on a musical timeline, the MIDI pitch number, the current time signature, and accents or playing styles like *staccato* or *sforzato*.

3.5 AUDIO RECORDINGS

The BPSD comprises 11 complete performances of the first movements of Beethoven's 32 piano sonatas. Four of these performances (`AS35`, `FG58`, `FJ62`, `WK64`) are in the public domain in the EU[4] and are included with the dataset, while the remaining seven performances are commercially available. The recordings by Jank are available at IMSLP[5], while all other recordings are uniquely identified using their European Article Number (EAN). We provide additional metadata such as MusicBrainz IDs, when available. An overview of all audio versions in the BPSD is presented in Table 2. To ensure consistency, we convert all audio files to mono WAV format with a sampling rate of 22050 Hz.

3.6 STRUCTURAL MODIFICATIONS

In the original recordings forming the basis of the BPSD, we identified structural differences that resulted in notable inconsistencies in the musical timelines. These differences include the absence of repetitions in the exposition, additional repetitions of the development and recapitulation, and discrepancies in the number of measures played due to performance errors.

While concepts such as 'Measure Maps' (Gotham et al., 2023a) have been introduced for transferring symbolic annotations between editions with different structures, we choose to apply structural modifications directly at the audio level. This approach ensures that audio and annotations are inherently coherent at every step of the processing pipeline and can be immediately used in data-driven approaches without the need for ad-hoc modifications.

To this end, we designated the performances by Daniel Barenboim (`DB84`) as the reference version for all movements to ensure temporal coherence in the BPSD. It is noteworthy that Friedrich Gulda follows the same performance structure convention in both `FG58` and `FG67`. Barenboim generally adheres to the notated score and closely follows the model often referred to as sonata form ("Sonatenhauptsatzform"), beginning with an exposition

Folder name	Content
`- 0_RawData`	Raw audio and symbolic data
`| - audio_ripped`	Audio files as ripped from the CD
`| - WK64`	
`| ...`	
`| - FG67`	
`| - score_pdf_scan`	Scanned score from IMSLP
`| - score_pdf_repetitions`	Symbolic score in PDF format with repeat signs
`| - score_pdf_unfolded`	Symbolic score in PDF format with unfolded repetitions
`| - score_sibelius_repetitions`	Symbolic score in Sibelius format with repeat signs
`| - score_sibelius_unfolded`	Symbolic score in Sibelius format with unfolded repetitions
`| - score_xml_repetitions`	Symbolic score in MusicXML format with repeat signs
`| - score_xml_unfolded`	Symbolic score in MusicXML format with unfolded repetitions
`| - score_midi`	MIDI export of the symbolic score
`- 1_Audio`	Audio files with coherent structure
`- 2_Annotations`	Annotations with musical and physical timelines
`| - ann_score_note`	Note events with start and end given in musical time
`| - ann_score_chord`	Harmony annotations given in musical time
`| - ann_score_localkey`	Local key annotations given in musical time
`| - ann_score_globalkey`	Global key annotations
`| - ann_score_structureFine`	Fine structure annotations given in musical time
`| - ann_score_structureCoarse`	Coarse structure annotations given in musical time
`| - ann_audio_note`	Note events with start and end given in physical time
`| - ann_audio_midi`	Note events in physical time in MIDI format
`| - ann_audio_beat`	Beat annotations given in physical time
`| - ann_audio_measure`	Measure annotations given in physical time
`| - ann_audio_startEnd`	Start and end of audio recordings (for removing silence/applause) given in physical time
`| - ann_audio_syncInfo`	Alignment tuples for converting between musical and physical timeline
`| - ann_audio_modifications`	Annotations for structural modifications of recordings
`| - ann_audio_chord`	Harmony annotations given in physical time
`| - ann_audio_localkey`	Local key annotations given in physical time
`| - ann_audio_structureFine`	Fine structure annotations given in physical time
`| - ann_audio_structureCoarse`	Coarse structure annotations given in physical time
`- 3_Scripts`	Python scripts to convert raw data into the structured format

Table 1 Overview of the folder structure of the BPSD. Score-based folders contain files named in the format `Beethoven_workID.ext`, while audio-based folders contain files in the format `Beethoven_workID_performerID.ext`.

ID	Performer	Year	Label	EAN Code	Orig. Dur.	Final Dur.
WK64	Wilhelm Kempff	1964	Deutsche Grammophon	028944796629	03:18:26	03:45:31
FJ62	Fritz Jank	1962	Instituto Piano Brasileiro	available at IMSLP	03:35:13	03:41:26
FG58	Friedrich Gulda	1958	Decca	028948514519	03:34:00	03:34:00
AS35	Artur Schnabel	1935	Warner Classics	0190295975050	03:31:03	03:33:35
MC22	Muriel Chemin	2022	Odradek	855317003615	04:08:22	04:05:11
MB97	Malcolm Bilson et al.	1997	Claves	7619931970721	03:52:23	03:46:08
AB96	Alfred Brendel	1996	Philips	028941257529	03:54:34	03:52:28
JJ90	Jeno Jando	1990	NAXOS	730099150224	03:41:06	03:39:14
DB84	Daniel Barenboim	1984	Deutsche Grammophon	028941375926, 028941376626	03:58:37	03:58:37
VA81	Vladimir Ashkenazy	1981	London Records	028944370621	03:48:16	03:46:27
FG67	Friedrich Gulda	1967	Amadeo	028947687610	03:25:02	03:25:02
				Total	40:47:08	41:07:45

Table 2 Overview of audio recordings in the BPSD. The upper four performances with identifiers `WK64`, `FG58`, `FJ62`, and `AS35` are in the public domain and freely accessible within the BPSD. All remaining recordings are commercially available and can be identified using the EAN code. Durations are presented in the format hh:mm:ss.

and its repetition, followed by a development, and concluding with a recapitulation. Only when explicit deviations from this form are notated in the sheet music, such as repeat signs for the recapitulation and development parts with alternative first and second endings and/or a closing coda, does Barenboim precisely follow the structure notated in the score. An overview of the reference structure for all sonatas underlying our dataset is provided in the last column of Table 3. For a more detailed discussion of the sonata form, see Section 5.3 and (Hepokoski and Darcy, 2006; Neuwirth, 2021).

Using `DB84` as a reference, we edited the other performances to precisely follow the same structure. For this purpose, we applied three types of modifications: *cut* (removing a part completely), *copy* (duplicating a part and inserting it at a different time position), and *insert silence* (introducing silence in exceptional cases for short periods when a copy operation was not possible due to missing audio material for alternative endings). Annotations for these modifications are provided in separate CSV files for each recording that requires adjustment, available in `ann_audio_modifications`. All edit operations are automatically applied using a Python script. The finalized audio recordings are stored in the folder `1_Audio`.

As a result, the editing process guarantees coherence across all versions, adhering to the same musical timeline with a unique identifier for each measure, thus establishing well-defined alignments without gaps. Additionally, we ensured that the unfolded musical scores (Section 3.4) also adhere to the same timeline conventions. Finally, we note that all annotations are exclusively specified for the modified music recordings and unfolded scores.[6]

4 MUSIC SYNCHRONIZATION

Thanks to a coherent musical timeline across all versions of a movement, we can utilize standard music synchronization techniques to temporally align all score and audio representations. This approach enables us to initially specify annotations based on a shared musical timeline and subsequently transform this timeline to align with the physical timelines of specific performances. In this section, we summarize the music synchronization techniques used for creating the BPSD. While adhering to the standard high-resolution synchronization approach provided by the Sync Toolbox (Müller et al., 2021), we introduce a variant that further improves temporal accuracy by incorporating recent music transcription techniques (Section 4.1). Furthermore, we present experiments assessing the synchronization accuracy (Section 4.2) and subsequently detail the procedure for aligning musical and physical timelines by transferring measure and score-related information between versions (Section 4.3). For a detailed description of the transcription and synchronization approach we refer to (Zeitler et al., 2024).

No.	Work ID	Name	Key	Mean Dur.	Min. Dur.	Max. Dur	Meas.	Structure
01	Op002No1-01		F:min	03:47	03:22 (AS35)	04:33 (WK64)	200	E-E-D-R
02	Op002No2-01		A:maj	07:04	06:23 (FG67)	07:45 (MC22)	452	E-E-D-R
03	Op002No3-01		C:maj	10:15	09:47 (FG58)	11:25 (MC22)	347	E-E-D-R-C
04	Op007-01	Grand Sonata	Eb:maj	08:17	07:27 (AS35)	08:58 (MC22)	497	E-E-D-R-C
05	Op010No1-01		C:min	05:33	04:41 (AS35)	06:13 (MC22)	388	E-E-D-R
06	Op010No2-01		F:maj	05:38	05:03 (FG67)	06:14 (VA81)	268	E-E-D-R
07	Op010No3-01		D:maj	06:59	06:26 (FJ62)	07:53 (JJ90)	467	E-E-D-R-C
08	Op013-01	Pathétique	C:min	08:56	08:06 (FG58)	09:57 (MC22)	431	I-E-E-D-R-C
09	Op014No1-01		E:maj	06:35	05:31 (VA81)	07:25 (AB96)	222	E-E-D-R-C
10	Op014No2-01		G:maj	07:06	05:49 (AS35)	07:56 (AB96)	263	E-E-D-R-C
11	Op022-01		Bb:maj	07:26	06:43 (AS35)	08:36 (MC22)	267	E-E-D-R
12	Op026-01	Funeral March	Ab:maj	08:01	06:51 (FG67)	10:02 (AS35)	219	T-V1-V2-V3-V4-V5
13	Op027No1-01	Son. q. u. fant.	Eb:maj	05:12	04:36 (AB96)	05:42 (FG58)	106	An-Al-T1
14	Op027No2-01	Moonlight	C#:min	06:01	04:58 (AS35)	07:28 (FG58)	69	P1-P2-P3-C
15	Op028-01	Pastoral	D:maj	09:58	08:58 (FJ62)	11:39 (MC22)	622	E-E-D-R-C
16	Op031No1-01		G:maj	06:23	05:44 (FG58)	07:19 (MC22)	435	E-E-D-R-C
17	Op031No2-01	Tempest	D:min	08:27	06:49 (FG58)	09:52 (MC22)	320	E-E-D-R-C
18	Op031No3-01	The Hunt	Eb:maj	08:29	07:53 (FG67)	09:07 (MB97)	341	E-E-D-R-C
19	Op049No1-01	Easy Sonata	G:min	04:35	03:41 (JJ90)	05:17 (MB97)	143	E-E-D-R-C
20	Op049No2-01	Easy Sonata	G:maj	04:37	04:19 (FJ62)	05:10 (MC22)	174	E-E-D-R
21	Op053-01	Waldstein	C:maj	10:38	09:25 (FG67)	11:36 (MC22)	387	E-E-D-R-C
22	Op054-01		F:maj	05:38	04:58 (AS35)	06:13 (MC22)	154	M1-Tr1-M2-Tr2-M3-C
23	Op057-01	Appassionata	F:min	09:35	07:35 (FG67)	10:39 (DB84)	262	E-D-R-C
24	Op078-01	A Thérèse	F#:maj	07:04	06:20 (FG58)	08:18 (MC22)	206	I-E-E-D-R-D-R
25	Op079-01	Cuckoo	G:maj	04:40	03:58 (AS35)	05:12 (MC22)	372	E-E-D-R-D-R-C
26	Op081a-01	Les adieux	Eb:maj	07:04	06:00 (FG67)	07:50 (DB84)	308	I-E-E-D-R-C
27	Op090-01		E:min	05:35	04:34 (FG67)	06:19 (MB97)	245	E-D-R-C
28	Op101-01		A:maj	04:00	03:35 (WK64)	04:29 (DB84)	102	E-D-R-C
29	Op106-01	Hammer-klavier	Bb:maj	11:06	08:54 (AS35)	13:04 (DB84)	530	E-E-D-R-C
30	Op109-01		E:maj	03:46	03:14 (WK64)	04:19 (DB84)	99	E-D-R-C
31	Op110-01		Ab:maj	06:33	06:00 (FJ62)	07:33 (DB84)	116	E-D-R-C
32	Op111-01		C:min	09:05	08:20 (AS35)	10:04 (VA81)	209	I-E-E-D-R-C

Table 3 Overview of the first movements of Beethoven's 32 Piano Sonatas. The table displays information including the work ID, trivial name (if applicable), global key, mean, minimum, and maximum duration of available recordings (see Table 2), number of measures, and the coarse structure. All durations are presented in the format mm:ss.

4.1 SYNCHRONIZATION APPROACH

Music synchronization refers to a procedure that temporally aligns different versions of the same underlying piece of music (Müller, 2021). Depending on the versions, there are various variants, including score-audio synchronization for aligning a score representation with an audio recording and audio-audio synchronization for aligning two audio versions. In our scenario, we employ a synchronization procedure provided by the Sync Toolbox (Müller et al., 2021), which is based on a multi-scale and multi-resolution variant of dynamic time warping (DTW) (Prätzlich et al., 2016). The synchronization pipeline utilizes framewise chroma representations (capturing harmonic and melodic properties) as input features and incorporates additional onset activations (capturing note beginnings) for refinement, see also (Özer et al., 2022).

In the case of score data, chroma and onset information can be directly extracted from the symbolic representation and encoded in the form of a piano-roll representation. As for audio data, one may use signal processing techniques for extracting chroma and onset features (Müller and Ewert, 2011). As an alternative, we convert the audio recordings into piano-roll representations by applying state-of-the-art music transcription techniques based on DL. In particular, we use a model based on the Onsets and Frames architecture proposed by Hawthorne et al. (2018). Following Maman and Bermano (2022), we further improve this model by including a fine-tuning step on unaligned pairs of audio and score representations from the BPSD. This results in highly accurate transcription results, particularly for piano music, which can be integrated in the previous synchronization pipeline to improve the alignment results.

4.2 SYNCHRONIZATION ACCURACY

We now assess the accuracy of our synchronization pipeline. Since reference alignments for the BPSD are unavailable, we employ heuristics similar to (Prätzlich and Müller, 2016). To achieve this, we calculate measure estimates for all audio versions in two different ways. In the first case, we manually annotate the measure positions for all `WK64` recordings (see also Section 5.1) and then apply audio-audio synchronization to transfer these annotations to all other audio versions. In the second case, we apply score-audio synchronization to align the symbolic score with all audio versions. As all note events in the symbolic score are associated with a position on the musical timeline (see Section 3.3), we directly obtain estimates of measure positions for the audio versions as long as there is a note onset on the measure's first beat. If this is not the case, we linearly interpolate between the two neighboring note events to estimate the measure position.

In our heuristics, we assume that the synchronization accuracy is high if the audio-audio estimates closely match the score-audio estimates for all measure positions. Although, strictly speaking, this is only a necessary condition and not a sufficient one, these heuristics serve as a reliable indicator of accuracy. For each pair (audio-audio, audio-score) of measure estimates, we calculate the absolute error. In Table 4, we present the mean, median, and 95% confidence interval for these errors on the basis of all audio versions in the BPSD. For example, we obtain an overall mean error of 25 ms and a median error of 17 ms. Considering only measure estimates that align with note onsets, these numbers slightly improve to 18 ms and 16 ms, respectively. Recall that the measure positions for `WK64` were annotated manually (see also Section 5.1). Thus, the absolute errors reported for `WK64` provide an indicator of the performance of audio-score synchronization alone (along with inaccuracies in the manual annotation process that, for more complex music, can reach a level of up to 100 ms (Weiß et al., 2016)). While yielding slightly better results, the errors on `WK64` remain in the same order of magnitude as the other results.

Overall, we may conclude that the synchronization yields accurate alignment results with errors in the order of 20 to 30 ms (assuming that it behaves in a similar fashion within measure positions). The 95% confidence interval of about 40 to 50 ms provides a more generous estimate of the accuracy, which needs to be taken into account when working with automatically transferred annotations (see Section 5).

4.3 ALIGNING MUSICAL AND PHYSICAL TIMELINES

Using the music synchronization method described before, we now describe the process of aligning the musical and physical timelines across all performances. In the context of Western classical music, audio-audio synchronization often proves to be more robust than score-audio synchronization. This is attributed to the fact that input features are computed from data within the same domain in the audio-audio case.

Motivated by this and following (Weiß et al., 2021a, 2023), we adopt a multi-step approach to align the musical and physical timelines. In the initial step, audio-audio synchronization is applied to transfer measure annotations from the `WK64` recordings to all other recordings. To mitigate boundary artifacts resulting from silence or applause at the beginning or end of a recording, we manually annotated the start and end points of each recording (`ann_audio_startEnd`). These time positions serve as anchor points in the synchronization approach (see (Prätzlich et al., 2016) for an explanation of the anchor point concept).

In the second step, using the transferred measure positions as anchor points, we apply score-audio synchronization. This maps all note onsets and offsets from the

	All Measures			Measures With Note Onset		
Version	Mean	Median	95% Conf.	Mean	Median	95% Conf.
WK64	20	13	40	14	11	40
FJ62	25	19	60	19	18	45
FG58	23	16	41	17	14	40
AS35	25	15	54	18	13	40
MC22	27	20	63	21	20	60
MB97	30	20	60	20	20	47
AB96	28	18	46	18	17	40
JJ90	24	17	43	16	16	40
DB84	29	19	60	19	18	54
VA81	24	17	56	17	16	41
FG67	25	9	40	17	8	40
All	25	17	53	18	16	40

Table 4 Accuracy of synchronization approaches. The table presents absolute errors between measure estimates obtained from audio-audio synchronization (based on manually annotated measure positions for `WK64`) and score-audio synchronization. Mean, median, and the 95% confidence interval for all measures (left side) and for only those measures with a note onset (right side) are reported. All values are given in milliseconds.

musical timeline to the physical timeline, providing a representation of all symbolic note events precisely aligned with an audio version (available in `ann_audio_note` as a list of note events and in `ann_audio_midi` as a playable MIDI file for each track).

In the third step, the aligned note events serve as the basis for mapping musical time to physical time. Recall from Section 3.3 that each note onset of the score is encoded by a real-valued position on a continuous measure axis, while the aligned position in an audio version is a real-valued number encoding physical time in seconds. This information is stored as tuples '(musical time, physical time)' on a discretized time grid in `ann_audio_syncInfo` and is utilized for every transfer from a score-based version to an audio-based version or vice versa. We use linear interpolation between note events to obtain a continuous mapping between positions on the musical timeline and a physical timeline.

It is crucial to emphasize that our approach transfers note information from the score, aligning the musical timeline of the score with the physical timeline of the performance. Consequently, the annotations are not tailored to individual deviations within each performance, especially playing errors. Additionally, our synchronization approach is not designed to handle arpeggios or trills, which are treated as simultaneous events. Hence, our note annotations should be regarded not as a direct transcription of the audio recording but as a temporal adaptation of the score information.

5 ANNOTATIONS

In this section, we provide an overview of the annotations included in the BPSD. These annotations cover measure and beat positions, harmonic segments (local key and chord), and structural elements (on a coarse and fine level). For each movement, the annotations are available with respect to the musical timeline (for the score version) and with respect to performance-specific physical timelines (for all audio versions). For an overview, we refer to Figure 2.

5.1 MEASURE AND BEAT ANNOTATIONS

As already mentioned in Section 4, we manually annotated measure positions for `WK64`, which serves as our reference version for measure annotations. The reasons for selecting `WK64` as a reference include its availability in the public domain, its high audio quality, and Kempff's choice of a relatively slow tempo, which simplifies the manual annotation process. In this process, a musically trained listener annotated all `WK64` recordings using the Sonic Visualiser software (Cannam et al., 2010). To ensure high quality, these annotations were processed with a sonification tool and cross-checked by an independent and experienced audio annotator. The experiments in Section 4.2 suggest that the annotation accuracy is within the order of or better than the synchronization accuracy of 20 to 30 ms, which is better than obtained for complex orchestral music in (Gadermaier

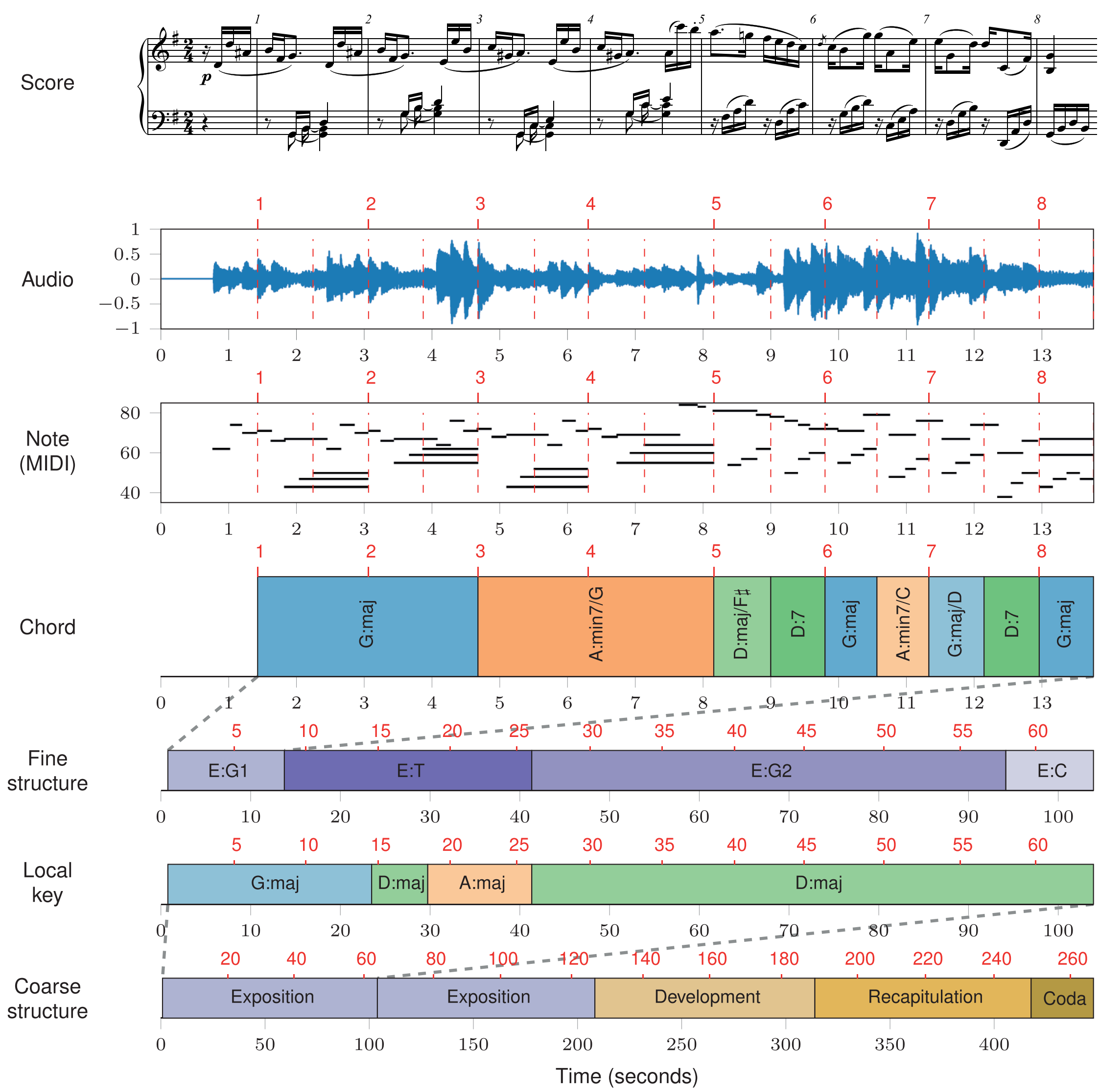

Figure 2 Overview of various annotations in the BPSD illustrated using the first measures of the Sonata Op. 14, No. 2 in G Major. Measure positions are marked with red ticks, while beat positions are indicated by red dashed lines.

and Widmer, 2019; Weiß et al., 2016). The measure annotations for the other audio versions were then obtained using audio-audio synchronization techniques and stored in `ann_audio_measure`.

To derive beat annotations for the score, we initially generated a list of all beats in a movement by dividing each measure by the number of beats specified in the time signature. Subsequently, utilizing the alignment of musical and physical timelines (see Section 4.3), we transferred the beat positions to the physical timelines of the audio versions, storing them in `ann_audio_beat`.

5.2 HARMONY ANNOTATIONS

The BPSD includes diverse harmony-related annotations in the form of absolute and relative chord, local key, and global key annotations. For the chord annotations, we started with the score-based annotations provided by the BPS-FH dataset (Chen and Su, 2018). We adjusted the original timelines, initially specified on a quarter note grid, to our measure-based musical timelines, represented by real-valued measure counts. Additionally, we made structural modifications to ensure coherence with the reference musical timeline when necessary. Finally, these annotation were transferred to the audio versions using the alignment information. As a first type of chord annotations, we include the high-level chord labels from the original BPS-FH dataset consisting of chord functions (as scale degrees in Roman numeral notation) relative to the local key. To facilitate the usage of BPSD for tasks like audio-based chord estimation, we additionally opt for a more pragmatic encoding.[7] Therefore, we choose to provide absolute chord information (root notes specified as pitch classes) in the encoding scheme proposed by Harte et al. (2005). Enhancing the applicability of chord annotations for various tasks, we represent chords at different levels of detail. This includes shorthand notation (e.g., C:7/E), extended form (e.g., C:(3,5,b7)/E), major/minor with inversion (e.g., C:maj/E), major/minor (e.g., C:maj), and a numerical identifier of the latter (e.g., 1). Chord annotations with the start and end of each chord in musical and physical time are provided in `ann_score_chord` and `ann_audio_chord`, respectively.

In a similar manner, we include the local key annotations from the BPS-FH dataset (Chen and Su, 2018), adapting and transferring them accordingly for the audio representations. The local key annotations, along with their start and end times in both musical and physical timelines, are accessible in `ann_score_localkey` and `ann_audio_localkey`, respectively.

Finally, for the global key annotations, we deviate from our convention of having one annotation file per movement and performance. Instead, we provide a single table in `ann_score_globalkey`, containing the global keys of the first movements of all 32 piano sonatas.

5.3 STRUCTURE ANNOTATIONS

As a last type of annotation, we include structural information, primarily intended for the purposes of overview and navigation. Strictly following the analysis by Tovey (1931), we include structural boundaries in two levels of granularity. In particular, we provide boundary annotations for the coarse structure (such as Exposition, Development, Recapitulation, etc.) and fine sub-structures based on thematic material used (such as First Group, Second Group, Transition phases, etc.), as visualized in Figure 2 and described in the following section.

The sonata form is a prominent structural concept in Western classical music and can be categorized into five types (Hepokoski and Darcy, 2006). In Beethoven's piano sonatas, the type 3 sonata ('Sonatenhauptsatzform') is predominantly used, usually containing a (repeated) exposition, a development, and a recapitulation part, occasionally starting with an introduction and concluding with a coda. Following the analysis by Tovey (1931), we provide this coarse structural information in relation to the musical timelines in `ann_score_structureCoarse` and the performance-specific physical timelines in `ann_audio_structureCoarse`.

In addition to the overarching sonata form structure, one often encounters specific substructures within the exposition and recapitulation. Typically, the exposition introduces the primary thematic material of the movement through two tonally contrasting subject groups. The first subject group is usually in the global key (home key) of the movement, while the second subject group is in the dominant key (for sonatas in major) or the relative key (for sonatas in minor). These subject groups are commonly linked by a modulating transition, and the exposition frequently concludes with an additional closing theme or codetta. The recapitulation mirrors the subparts of the exposition but incorporates a significantly different local key progression. Also, it often includes prolonged sections, introduces new material, and incorporates local modulations.

Consistent with the analysis by Tovey (1931), we offer structure annotations on a more detailed level in `ann_score_structureFine` for the score and in `ann_audio_structureFine` for the audio recordings. Comparable annotations were employed in the study by Jiang and Müller (2013) to assess a computational approach for analyzing music recordings in sonata form.

While we are aware of the limitations of Tovey's analysis, assuming a rather outdated view on sonata form, the great level of detail and descriptiveness ('bar-by-bar') is an advantageous property in our case since it allows for deriving clear marks of reference that facilitate navigation and overview.

We suggest a careful interpretation of structure annotations, as they often capture tendencies rather than strict rules. In practice, numerous exceptions and modifications to the sonata form exist, and the utility of the sonata form as an oversimplifying model is a subject of debate among many researchers, as discussed in (Kleinertz, 2016; Neuwirth, 2021).

6 APPLICATION SCENARIOS

In this section, we outline several use cases to highlight the research potential of the BPSD in fields such as MIR and computational musicology. Some of these studies draw upon prior research that utilized similar datasets.

6.1 LEARNING PITCH-CLASS REPRESENTATIONS

Pitch-class or chroma representations extracted from audio recordings play a fundamental role in various MIR tasks. Traditional chroma features, computed using signal processing methods, are often impaired by timbral properties like overtones or vibrato, resulting in an only rough correspondence to the pitch classes indicated by a musical score. To overcome this, Weiß et al. (2021b) employed a DL approach for learning transcription-like pitch-class representations using synchronized score-audio pairs from classical music. Providing aligned training pairs, the BPSD proves to be a valuable resource for training such DL-based models. An example of an audio-based pitch-class representation learned with a CNN-based model and trained exclusively on the BPSD is illustrated in Figure 3.

As another central feature of BPSD, its coherent structure enables the exploration of different systematic training-test splits, including the cross-work split (across movements) and the cross-version split (across performances). Introduced in (Weiß et al., 2020b), such an approach was applied to study the generalization properties of data-driven methods for local key estimation using the Schubert Winterreise Dataset (Weiß et al., 2021a).

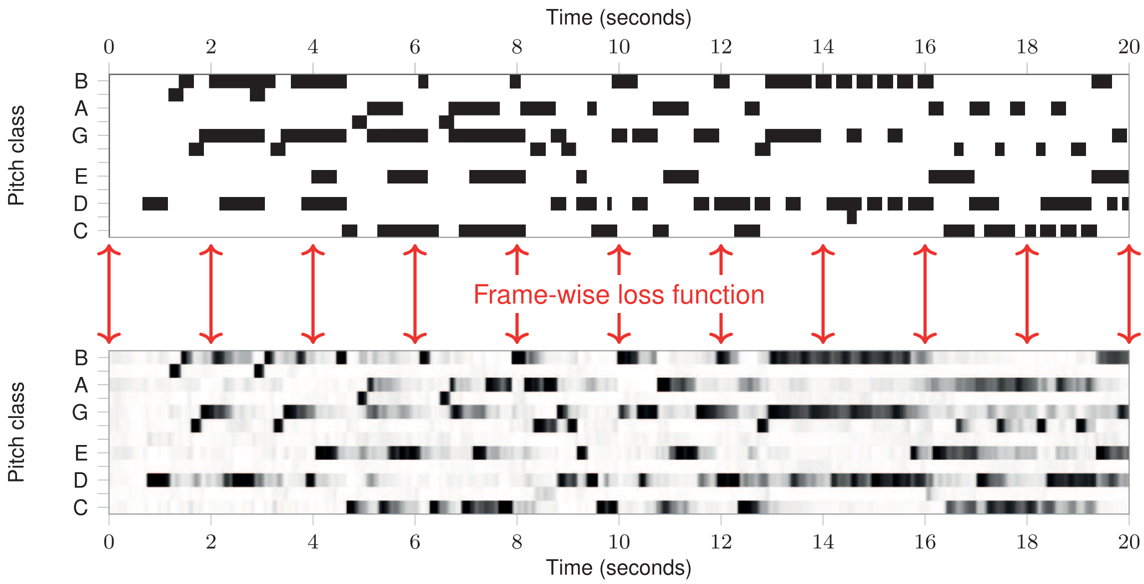

Figure 3 Synchronized score-audio training pair for learning pitch-class representations using a frame-wise loss function.

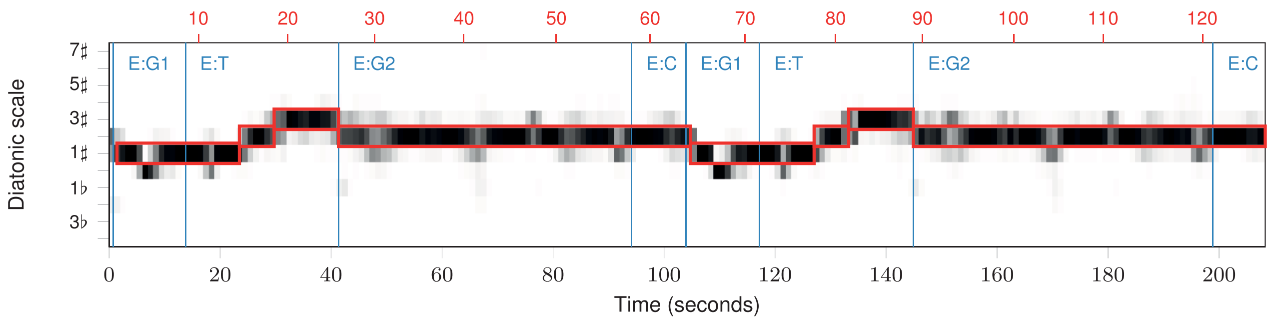

Figure 4 Visualization of a time-diatonic representation derived from the `WK64` recording of the first movement of the Piano Sonata Op. 14 No 2 in G Major. The local-key reference annotations are indicated by the overlaid red rectangles.

6.2 LOCAL KEY ANALYSIS

The aforementioned study leads us directly to our next use case, where we consider the task of estimating local keys-a task that can be seen as conducting harmony analysis on a mid-level time scale. Instead of explicitly extracting and numerically evaluating such information, it is often insightful to visualize harmonic structures and leave the final interpretation to a music expert. One such approach is described and applied to Beethoven's sonatas in (Weiß et al., 2020a), where relevant pitch content with respect to the 12 diatonic scales is extracted from an audio recording and visualized in the form of a time-diatonic representation.

Figure 4 shows such a visualization of a time-diatonic representation derived from a `WK64` recording of the Piano Sonata Op. 14 No 2 in G Major. Note that the visualization is organized along the recording's physical timeline. The reference annotations for local keys, indicated by overlaid red rectangles, assist in evaluating the estimation results. The automatically computed time-diatonic representation offers a clear overview of the general harmonic progression, emphasizing the typical key structure of the sonata form.

The figure also demonstrates that the inclusion of additional information, such as the start times of musical form sections or the position of measures, can further enrich the visualization. Such annotations are particularly helpful when using audio-based visualizations in combination with musical scores for scrutinizing and gaining a deeper understanding of the phenomena revealed by the visualizations.

6.3 CHORD RECOGNITION

In our third use case, we consider the task of chord recognition, which can be seen as a detailed analysis of harmony (Pauwels et al., 2019). In the early cross-version study by Konz et al. (2013), the authors showed that chord recognition results can be stabilized by simultaneously performing this task for several audio recordings and then merging the results using a fusion approach (e.g., majority voting).

The BPSD proves ideal for such a cross-version study, encompassing scores and multiple recordings for all movements. Figure 5 illustrates this methodology, showcasing chord recognition results for all 11 performances of the initial measures of the Piano Sonata Op. 31 No. 1

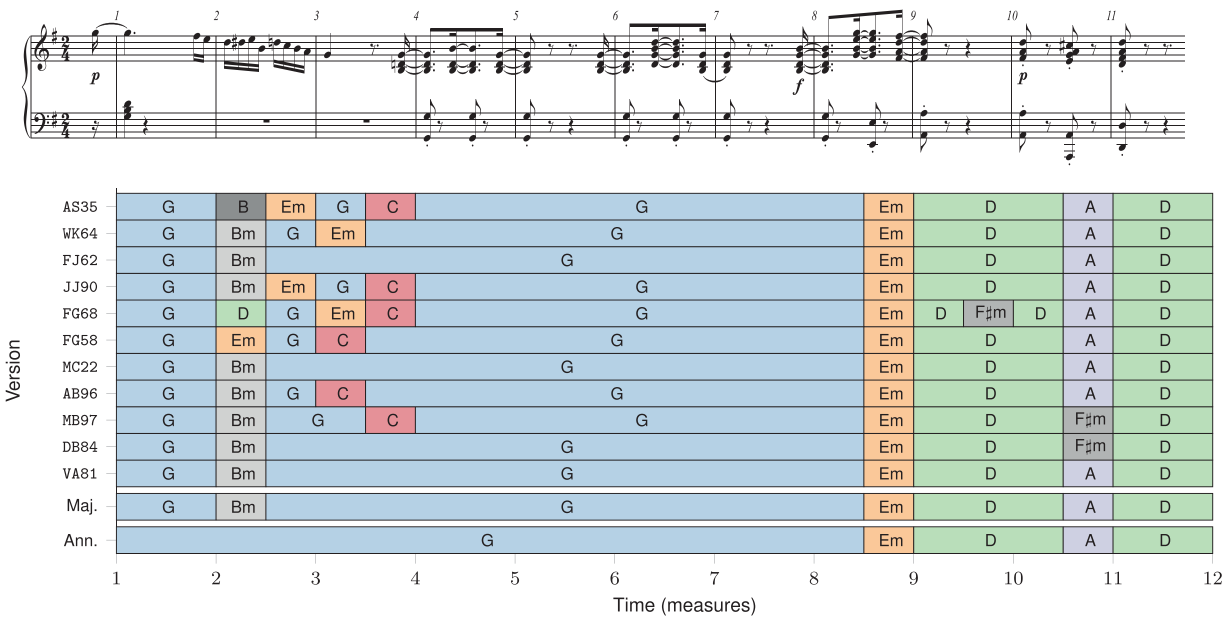

Figure 5 Cross-version chord recognition for the initial measures of the Piano Sonata Op. 31 No. 1 in G major. The results are presented for all 11 performances, alongside the majority vote and the reference annotations.

in G major, alongside the majority vote and reference annotations. In this example, leveraging the improved pitch-class representations outlined in Section 6.1, we employed a simple template-based chord recognition approach, simplifying to the 24 major and minor triads. Furthermore, we transformed the performance-dependent physical timelines of the chord recognition results into a unified musical timeline based on alignment annotations. Finally, we consolidated the results via majority voting, resulting in increased stability and improvement compared to individual recognition outcomes.

7 CONCLUSIONS

This paper presented the carefully curated and annotated multi-version BPSD, centered on the first movements of Beethoven's piano sonatas. We provided a comprehensive overview of the data, addressing our methodologies for curation, processing, correction, and annotation. The dataset includes symbolic scores in various formats and over 40 hours of audio data from 11 performances, with 4 being in the public domain accessible for research purposes. Secondary materials encompass alignments, measure and beat positions, chord and local keys, and structure annotations for all versions.

We estimate that the creation of the BPSD required more than 2,000 hours in total, involving tasks such as symbolic score creation (ca. 500 hours), annotation creation and adjustment (ca. 500 hours), programming (ca. 500 hours), and further processing and refinement (ca. 500 hours), excluding the actual research work with the data.

A key feature is the dataset's coherence, achieved by enforcing unified musical time for all versions and ensuring annotation consistency. Leveraging this coherency (admittedly sometimes achieved through simplification and musicologically debatable assumptions), the BPSD is designed to be technically convenient and can be readily utilized for training and testing machine learning approaches. This includes the investigation of data-driven approaches for chord recognition, local key estimation, measure and beat tracking, structure analysis, and music transcription to evaluate their capabilities for generalization across works, versions, and modalities.

Building upon previous studies, we highlighted various research opportunities provided by the dataset. Specifically, we hope that the BPSD will foster further engagement in the musicological discourse on Beethoven's piano sonatas and their performances by offering a framework that allows for the easy transfer of musicological annotations, typically defined on a musical timeline, to a multitude of audio recordings. In this context, we believe that the development of interactive software and visualization applications holds significant potential for interdisciplinary research and educational applications. In conclusion, we aspire for the BPSD to be a valuable resource and a source of inspiration, offering compelling research opportunities in the fields of music information retrieval, computational musicology, and digital humanities at large.

ACKNOWLEDGEMENTS

We express our gratitude to all team members, student assistants, and colleagues who played pivotal roles in the data curation process and annotation work. Special mentions among the numerous contributors include Harald Grohganz, Celina Hüttner, Nanzhu Jiang, and Michael Kohl.

FUNDING INFORMATION

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) within the project 'Computational Analysis of Harmonic Structures' under No. 252013209 (MU 2686/7-2), the project 'Learning with Music Signals: Technology Meets Education' under No. 500643750 (MU 2686/15-1), and the Emmy Noether research group 'Computational Analysis of Music Audio Recordings: A Cross-Version Approach' under No. 531250483 (WE 6611/3–1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHORS' CONTRIBUTIONS

Compiling the BPSD involved an extensive workflow, encompassing the preparation and processing of raw data, typesetting, manual corrections, annotation creation and transfer, as well as the utilization of the data for research. This complex process underwent multiple cycles of revision and correction, requiring a collaborative effort from a considerable number of individuals within the research team. Given the collective nature of the work, it is challenging to attribute specific contributions to individual team members. All co-authors of this paper actively participated in both the data curation process and the preparation of the final manuscript.

NOTES

1. This selection follows the convention of most editions and CD compilations to not include the early sonatas WoO 47 (*Kurfürstensonaten*), which have been subject to a computational study by Klauk et al. 2021.
2. https://doi.org/10.5281/zenodo.10847702
3. https://mirdata.readthedocs.io/en/latest/
4. Please note that these recordings might not be in the public domain in other countries outside the EU and Switzerland.
5. https://imslp.org/wiki/Category:Jank,_Fritz
6. We are aware that such modifications are unusual from an editor's perspective or for performance analysis. However, regarding the BPSD mainly as a resource for studying algorithmic approaches to MIR and computational musicology, we value the structural consistency across versions as favorable. Please note that original audio files are still accessible in the folder `0_RawData` and the scripts provided along with the data ensure reproducibility and bridge the link back to the original data
7. See, for instance, the mir-eval library (https://github.com/craffel/mir_eval).

AUTHOR AFFILIATIONS

Johannes Zeitler https://orcid.org/0000-0003-2171-7679
International Audio Laboratories Erlangen, Germany

Christof Weiß https://orcid.org/0000-0003-2143-4679
Center for Artificial Intelligence and Data Science, University of Würzburg, Germany

Vlora Arifi-Müller https://orcid.org/0009-0006-3477-7143
International Audio Laboratories Erlangen, Germany

Meinard Müller https://orcid.org/0000-0001-6062-7524
International Audio Laboratories Erlangen, Germany

REFERENCES

Benetos, E., Dixon, S., Duan, Z., and Ewert, S. (2019). Automatic music transcription: An overview. *IEEE Signal Processing Magazine*, *36*(1), 20–30. https://doi.org/10.1109/msp.2018.2869928

Bittner, R. M., Fuentes, M., Rubinstein, D., Jansson, A., Choi, K., and Kell, T. (2019). Mirdata: Software for reproducible usage of datasets. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR*), Delft, The Netherlands, pp. 99–106.

Bittner, R. M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., and 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*), Taipei, Taiwan, pp. 155–160.

Böck, S., Davies, M. E. P., and Knees, P. (2019). Multitask learning of tempo and beat: Learning one to improve the other. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR*), Delft, The Netherlands, pp. 486–493.

Calvo-Zaragoza, J., Hajič, Jr., J, and Pacha, A. (2020). Understanding optical music recognition. *ACM Computing Surveys*, *53*(4). https://doi.org/10.1145/3397499

Cannam, C., Landone, C., and Sandler, M. B. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the International Conference on Multimedia*, Florence, Italy, pp. 1467–1468.

Chen, T., and Su, L. (2018). Functional harmony recognition of symbolic music data with multi-task recurrent neural networks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR*), Paris, France, pp. 90–97.

Cooper, B. (2017). *The Creation of Beethoven's 35 Piano Sonatas*. Routledge.

Damschroder, D. (2016). *Harmony in Beethoven*. Cambridge University Press.

Emiya, V., Badeau, R., and 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. https://doi.org/10.1109/tasl.2009.2038819

Foscarin, F., McLeod, A., Rigaux, P., Jacquemard, F., and Sakai, M. (2020). ASAP: A dataset of aligned scores and performances for piano transcription. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 534–541).

Gadermaier, T., and Widmer, G. (2019). A study of annotation and alignment accuracy for performance comparison in complex orchestral music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, The Netherlands, pp. 769–775.

Gotham, M., Hentschel, J., Couturier, L., Dykeaylen, N., Rohrmeier, M., and Giraud, M. (2023a). The 'measure map': An inter-operable standard for aligning symbolic music. In **M. E. Thomae** (Ed.), *Proceedings of the International Conference on Digital Libraries for Musicology (DLfM)* (pp. 91–99). ACM.

Gotham, M., and Ireland, M. (2019). Taking form: A representation standard, conversion code, and example corpora for recording, visualizing, and studying analyses of musical form. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, The Netherlands, pp. 693–699.

Gotham, M., Micchi, G., López, N. N., and Sailor, M. (2023b). When in rome: A meta-corpus of functional harmony. *Transaction of the International Society for Music Information Retrieval (TISMIR)*, *6*(1), 150–166. https://doi.org/10.5334/tismir.165

Goto, M. (2004). Development of the RWC music database. In *Proceedings of the International Congress on Acoustics (ICA)* (pp. 553–556).

Harte, C., Sandler, M. B., Abdallah, S., and Gómez, E. (2005). Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, London, UK, pp. 66–71.

Hawthorne, C., Elsen, E., Song, J., Roberts, A., Simon, I., Raffel, C., Engel, J. H., Oore, S., and Eck, D. (2018). Onsets and frames: Dual-objective piano transcription. In *Proceedings of the International Society for Music Information Retrieval Conference, (ISMIR)*, Paris, France, pp. 50–57.

Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C. A., Dieleman, S., Elsen, E., Engel, J. H., and 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.

Hentschel, J., Neuwirth, M., and Rohrmeier, M. (2021). The annotated mozart sonatas: Score, harmony, and cadence. *Transaction of the International Society for Music Information Retrieval (TISMIR)*, 4(1), 67–80. https://doi.org/10.5334/tismir.63

Hentschel, J., Rammos, Y., Moss, F. C., Neuwirth, M., and Rohrmeier, M. (2023). An annotated corpus of tonal piano music from the long 19th century. *Empirical Musicology Review*, *18*(1), 84–95. https://doi.org/10.18061/emr.v18i1.8903

Hepokoski, J., and Darcy, W. (2006). *Elements of Sonata Theory. Norms, Types, and Deformations in the Late-Eighteenth-Century Sonata*. Oxford University Press.

Hsiao, Y., Hung, T., Chen, T., and Su, L. (2023). BPSMotif: A dataset for repeated pattern discovery of polyphonic symbolic music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 281–288).

Jiang, N., and Müller, M. (2013). Automated methods for analyzing music recordings in sonata form. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, Brazil. (pp. 595–600).

Klauk, S., Kleinertz, R., Weiß, C., and Müller, M. (2021). Seitensatz versus Mittelsatz: Expositionen in Beethovens frühen Klaviersonaten zwischen zeitgenössischer Theorie und computergestützter Analyse. In **S. Hohmaier** (Ed.), *Jahrbuch 2017 des Staatlichen Instituts für Musikforschung (SIM) - Preußischer Kulturbesitz* (pp. 271–300). Schott Music.

Kleinertz, R. (2016). Streichquartette von Joseph Haydn und Wolfgang Amadeus Mozart im Spiegel der Sonatentheorie Francesco Galeazzis. In **C. Speck** (Ed.), *The String Quartet from 1750 to 1870: From the Private to the Public Sphere* (Vol. 26, pp. 297–316). Brepols.

Konz, V., Müller, M., and Kleinertz, R. (2013). A cross-version chord labelling approach for exploring harmonic structures-a case study on Beethoven's Appassionata. *Journal of New Music Research*, *42*(1), 61–77. https://doi.org/10.1080/09298215.2012.750369

Lerch, A., Arthur, C., Pati, A., and Gururani, S. (2020). An interdisciplinary review of music performance analysis. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, *3*(1), 221–245. https://doi.org/10.5334/tismir.53

Maman, B., and Bermano, A. H. (2022). Unaligned supervision for automatic music transcription in the wild. In *Proceedings of the International Conference on Machine Learning (ICML)* (pp. 14918–14934).

Mauch, M., MacCallum, R. M., Levy, M., and Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, *2*(5). https://doi.org/10.1098/rsos.150081

Meredith, D. (Ed.). (2016). *Computational Music Analysis*. Springer.

Moss, F. C., Neuwirth, M., Harasim, D., and Rohrmeier, M. (2019). Statistical characteristics of tonal harmony: A corpus study of Beethoven's string quartets. *PLOS ONE*, *14*(6). https://doi.org/10.1371/journal.pone.0217242

Müller, M. (2021). *Fundamentals of Music Processing Using Python and Jupyter Notebooks* (2nd ed.). Springer Verlag.

Müller, M., Arzt, A., Balke, S., Dorfer, M., and Widmer, G. (2019). Cross-modal music retrieval and applications: An overview of key methodologies. *IEEE Signal Processing Magazine*, *36*(1), 52–62. https://doi.org/10.1109/msp.2018.2868887

Müller, M., and Ewert, S. (2011). Chroma Toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, Florida, USA, pp. 215–220.

Müller, M., Konz, V., Bogler, W., and Arifi-Müller, V. (2011). Saarland music data (SMD). In *Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, Florida, USA.

Müller, M., Özer, Y., Krause, M., Prätzlich, T., and Driedger, J. (2021). Sync Toolbox: A Python package for efficient, robust, and accurate music synchronization. *Journal of Open Source Software (JOSS)*, *6*(64), 3434:1–4. https://doi.org/10.21105/joss.03434

Nakamura, E., and Kaneko, K. (2019). Statistical evolutionary laws in music styles. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-52380-6

Neuwirth, M. (2021). A myriad of exceptions? challenging the sonata-form typology for the classical repertoire. *Music Analysis*, *40*(3), 312–352. https://doi.org/10.1111/musa.12186

Neuwirth, M., Harasim, D., Moss, F. C., and Rohrmeier, M. (2018). The Annotated Beethoven Corpus (ABC): A dataset of harmonic analyses of all Beethoven string quartets. *Frontiers in Digital Humanities*, *5*, 16. https://doi.org/10.3389/fdigh.2018.00016

Nieto, O., Mysore, G. J., Wang, C., Smith, J. B. L., Schlüter, J., Grill, T., and McFee, B. (2020). Audio-based music structure analysis: Current trends, open challenges, and applications. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, *3*(1), 246–263. https://doi.org/10.5334/tismir.78

Özer, Y., Ištvánek, M., Arifi-Müller, V., and Müller, M. (2022). Using activation functions for improving measure-level audio synchronization. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India, pp. 749–756.

Özer, Y., Schwär, S., Arifi-Müller, V., Lawrence, J., Sen, E., and Müller, M. (2023). Piano Concerto Dataset (PCD): A multitrack dataset of piano concertos. *Transaction of the International Society for Music Information Retrieval (TISMIR)*, *6*(1), 75–88. https://doi.org/10.5334/tismir.160

Pauwels, J., O'Hanlon, K., Gómez, E., and Sandler, M. B. (2019). 20 years of automatic chord recognition from audio. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, The Netherlands, pp. 54–63.

Peter, S. D., Cancino-Chacó, C. E., Foscarin, F., McLeod, A. P., Henkel, F., Karystinaios, E., and Widmer, G. (2023). Automatic note-level score-to-performance alignments in the ASAP dataset. *Transaction of the International Society for Music Information Retrieval (TISMIR)*, *6*(1), 27–42. https://doi.org/10.5334/tismir.149

Prätzlich, T., Driedger, J., and Müller, M. (2016). Memory-restricted multiscale dynamic time warping. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, pp. 569–573.

Prätzlich, T., and Müller, M. (2016). Triple-based analysis of music alignments without the need of ground-truth annotations. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, pp. 266–270.

Rosenzweig, S., Scherbaum, F., Shugliashvili, D., Arifi-Müller, V., and 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. https://doi.org/10.5334/tismir.44

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.

Thickstun, J., Harchaoui, Z., and Kakade, S. M. (2017). Learning features of music from scratch. In *Proceedings of the International Conference on Learning Representations (ICLR)*, Toulon, France.

Tovey, D. F. (1998/1931). *A Companion to Beethoven's Pianoforte Sonatas*. The Associated Board of the Royal Schools of Music.

Weiß, C., Arifi-Müller, V., Krause, M., Zalkow, F., Klauk, S., Kleinertz, R., and Müller, M. (2023). Wagner Ring Dataset: A complex opera scenario for music processing and computational musicology. *Transaction of the International Society for Music Information Retrieval (TISMIR)*, *6*(1), 135–149. https://doi.org/10.5334/tismir.161

Weiß, C., Arifi-Müller, V., Prätzlich, T., Kleinertz, R., and Müller, M. (2016). Analyzing measure annotations for Western classical music recordings. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, New York, USA, pp. 517–523.

Weiß, C., Klauk, S., Gotham, M., Müller, M., and Kleinertz, R. (2020a). Discourse not dualism: An interdisciplinary dialogue on sonata form in Beethoven's early piano sonatas. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Montréal, Canada. pp. 199–206.

Weiß, C., Schreiber, H., and Müller, M. (2020b). Local key estimation in music recordings: A case study across songs, versions, and annotators. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *28*, 2919–2932. https://doi.org/10.1109/taslp.2020.3030485

Weiß, C., Zalkow, F., Arifi-Müller, V., Müller, M., Koops, H. V., Volk, A., and Grohganz, H. (2021a). Schubert Winterreise dataset: A multimodal scenario for music analysis. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, *14*(2), 25:1–18. https://doi.org/10.1145/3429743

Weiß, C., Zeitler, J., Zunner, T., Schuberth, F., and Müller, M. (2021b). Learning pitch-class representations from score-audio pairs of classical music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 746–753. Online.

Zeitler, J., Maman, B., and Müller, M. (2024). Robust and accurate audio synchronization using raw features from transcription models. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, San Francisco, USA.

TO CITE THIS ARTICLE:
Zeitler, J., Weiß, C., Arifi-Müller, V., & Müller, M. (2024). BPSD: A Coherent Multi-Version Dataset for Analyzing the First Movements of Beethoven's Piano Sonatas. *Transactions of the International Society for Music Information Retrieval, 7*(1), 195–212. DOI: https://doi.org/10.5334/tismir.196

Submitted: 28 March 2024 **Accepted:** 12 August 2024 **Published:** 19 September 2024