



ChoraleBricks: A Modular Multitrack Dataset for Wind Music Research

DATASET ARTICLE

STEFAN BALKE AXEL BERNDT MEINARD MÜLLER 

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

Ju[ubiquity press

ABSTRACT

Concert band and wind music are deeply embedded in society and play a significant role in the cultural landscape of many countries, including Germany and Austria, particularly within the amateur music scene. However, this type of music, as well as research on wind and brass instruments in general, remains largely overlooked in the field of music information retrieval (MIR). In this paper, we address this underexplored area by introducing ChoraleBricks, a framework featuring multitrack recordings of ten different chorales, each comprising four musical parts: soprano, alto, tenor, and bass. At its core, ChoraleBricks provides isolated recordings of individual parts performed by a diverse selection of wind instruments, including flute, oboe, clarinet, trumpet, saxophone, baritone horn, trombone, and tuba. These isolated recordings act as building blocks or “bricks” that can be modularly superimposed to create full mixes with varying instrumentation. In addition, ChoraleBricks provides sheet music, time-aligned symbolic music representations, conducting videos, and reference annotations such as fundamental frequencies and note events. The framework is further enhanced by Python software tools that support parsing, mixing, annotation, and modular combination of the recorded audio material. With all multimedia and software components available as open-source, ChoraleBricks provides a versatile framework for generating and augmenting datasets for polyphonic wind music. It supports systematic experimentation and facilitates evaluation across various research topics, including multi-pitch estimation, note transcription, audio alignment, and music education applications.

CORRESPONDING AUTHOR:

Stefan Balke

International Audio
Laboratories Erlangen,
Germany

stefan.balke@audiolabs-erlangen.de

KEYWORDS:

Wind Music, Brass,
Woodwinds, Multitrack,
Multi-Modal

TO CITE THIS ARTICLE:

Balke, S., Berndt, A., & Müller, M. (2025). ChoraleBricks: A Modular Multitrack Dataset for Wind Music Research. *Transactions of the International Society for Music Information Retrieval*, 8(1), 39–54. DOI: <https://doi.org/10.5334/tismir.252>

1 INTRODUCTION

Music information retrieval (MIR) has seen significant advancements in recent years, with a focus on a wide array of musical genres, instruments, and analysis techniques, such as in singing voice separation in popular music or automatic music transcription (AMT) for piano recordings; see, e.g., [Peeters and Richard \(2021\)](#) and [Richard et al. \(2023\)](#). However, wind instruments, despite their prominent role in various musical traditions, remain underrepresented in many MIR studies. For instance, wind instruments are an integral part of the modern symphonic orchestra alongside strings and percussion and take on an even more prominent role in ensembles such as big bands. In popular music, many songs feature a wind section composed of instruments such as trumpets, saxophones, and trombones, with some songs, such as *Sir Duke* by Stevie Wonder,¹ being characterized by their wind arrangements. Most often, the neglect of these instruments is not due to any deliberate oversight but rather the scarcity of data. However, this lack of data limits the generalizability and applicability of MIR methods, particularly for tasks such as source separation, alignment, or instrument recognition. To address this gap, we introduce the ChoraleBricks dataset.

ChoraleBricks is a collection of multitrack recordings specifically designed to feature wind instruments in a controlled and flexible way. The dataset consists of ten chorales, each composed of four distinct musical parts, namely soprano (S), alto (A), tenor (T), and bass (B). These parts differ in pitch range and in their musical roles. For instance, the soprano usually presents the main melody, whereas the bass is responsible for the lowest register, usually providing the root or third of the chord. For each part, we have multiple recordings with different instruments. A key feature of the recordings is that each part was recorded individually, ensuring no crosstalk from other instruments. This feature provides researchers with the flexibility to combine different instruments in a modular fashion, allowing for systematic experimentation and evaluation across a range of research questions. In total, 13 different instruments were recorded, resulting in 193 tracks and 2 hours and 10 minutes of isolated multitrack audio recordings. [Figure 1](#) depicts the concept of the ChoraleBricks collection, showing three different instrument combinations (mixes) for the same chorale. In [Figure 1a](#), soprano and alto parts are played by trumpets, tenor by trombone, and bass by a tuba. In [Figure 1b](#), the mix consists of clarinets for soprano and alto, bassoon for tenor, and french horn for the bass. In [Figure 1c](#), we are combining parts from the first two mixes to yield a third mix—as simple as stacking bricks.

The ChoraleBricks chorales stem from German church music, a tradition integrating wind ensembles since the 16th century. The Protestant *Posaunenchor*—amateur

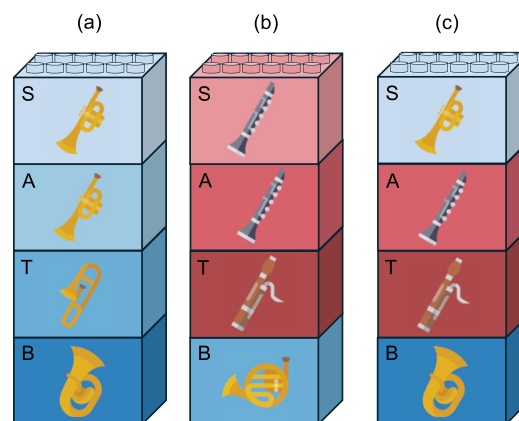


Figure 1 Schematic visualization of the ChoraleBricks concept for generating different mixes from the same chorale. **(a)** Mix consisting solely of brass instruments. **(b)** Mix primarily featuring woodwinds. **(c)** Mix combining two brass and two woodwind instruments.

brass ensembles featuring instruments such as trumpet, trombone, and tuba—evolved into its modern form in the early 20th century under influences such as the Moravian Church and Johannes Kuhlo ([Niemann, 2006](#)). These ensembles, reading in concert pitch (i.e., non-transposed C notation), adapt seamlessly to organ and choir repertoire and can therefore readily accompany community singing during church services. Today, the tradition remains lively, with a wealth of original compositions and arrangements spanning sacred and secular works across all styles, genres, and eras being published annually in individual scores and scorebook series.

The ChoraleBricks dataset is a carefully designed and versatile resource, offering value not only to the MIR community but also to amateur musicians and music education programs. It includes isolated audio tracks, sheet music in both MEI and MusicXML format, conducting videos, and additional annotations. These annotations provide editorial metadata for the tracks (e.g., player IDs and microphone types), precise alignments linking audio to sheet music, and manually corrected fundamental frequency (F_0) trajectories with derived note events. Following open science practices described by [McFee et al. \(2019\)](#), we provide data,² reference implementations in Python,³ and extensive documentation⁴ under suitable Open Access licenses.

The remainder of the paper is structured as follows. [Section 2](#) situates the presented work within the context of existing literature. [Section 3](#) details the creation process of the ChoraleBricks collection, including work selection, track recording, and data curation. [Section 4](#) provides quantitative statistics on the collection's multitrack audio material and discusses the potential of track permutations to expand the dataset's size. [Section 5](#) details the released software and accompanying websites. Finally, [Section 6](#) summarizes the paper and outlines ideas and a research road map for utilizing this new dataset.

2 RELATED WORK

2.1 WIND MUSIC RESEARCH

The historical development of brass and woodwind music in Europe began with ancient military signals and evolved through the Renaissance into organized bands for civic and ceremonial functions (Hofer, 1992). In the 18th and 19th centuries, professional military bands and the popular brass band movement emerged. This genre has evolved from its origins to develop its own distinct tradition, reflecting Europe's rich musical heritage. In particular, brass and woodwind music is an integral part of contemporary German culture, with military, church-related, and purely civilian concert bands and ensembles performing frequently. With a focus on the history and relevance of American wind music, the development from military music to modern wind music in the 20th century is described by Battisti (2018). A significant milestone in this development was the formation of the Eastman Wind Ensemble (EWE) by Frederick Fennell in 1952. The EWE emphasized flexible instrumentation, allowing for a wide repertoire ranging from classical transcriptions to contemporary works. In Chapter 24, Battisti (2018) states: "All new technological devices and instruments should be used to promote wind bands/ensembles and their music. Every iPhone and iPad is a concert hall."

From a scientific perspective, wind music is studied by small yet active research communities around the world, focusing mainly on musicological aspects. Two examples of such communities are the Internationale Gesellschaft zur Erforschung und Förderung der Blasmusik (IGEB)⁵ and the World Association for Symphonic Bands and Ensembles (WASBE).⁶ At universities, research on concert band and wind music is primarily conducted at musicological institutes or within performance studies. A notable research institute is the Pannonische Forschungsstelle/Internationales Zentrum für Blasmusikforschung,⁷ which focuses on wind music research and houses one of the most prolific libraries dedicated to wind music in the world. Additionally, it serves as the headquarters of the aforementioned IGEB.

2.2 MULTITRACK DATASETS

In the past, some multitrack datasets have been released for research purposes, e.g., MUSDB18 (Rafii et al., 2017), MedleyDB (Bittner et al., 2014), or MoisesDB (Pereira et al., 2023). However, most of these datasets concentrate on popular and rock music, where brass instruments and woodwinds usually play a minor role. An exception is the Cadenza Woodwind dataset, which was released during Cadenza's Second Open Machine Learning Challenge (CAD2) for the task of rebalancing classical music ensembles (Roa Dabike et al., 2024).⁸ In this case, the audio recordings in the dataset stem from a synthetic data generation pipeline, an emerging trend which could be

observed in other datasets as well, such as in EnsembleSet (Sarkar et al., 2022) or the CocoChorales dataset (Wu et al., 2022). Further symbolic encodings for chorales can be found in the Chorale Corpus (Gerhardt and Kirsch, 2024).⁹ Although high amounts of data can be generated, the versatility in timbre, articulation, and modulation of some instruments (e.g., trumpets) is not yet comparable to real recordings.

In the following, we discuss the most influential works for this paper and the ways in which the ChoraleBricks dataset is continuing the existing line of research. On a conceptual level, our work on ChoraleBricks has been inspired by three influential datasets. First, Duan and Pardo (2011) created the dataset Bach10 by recording ten Bach chorales using violin, clarinet, saxophone, and bassoon. Each part was recorded sequentially, starting with the violin, while successive players listened to the previously recorded parts to align their tempo and dynamics. Second, the Operation Beethoven dataset by Kaiser et al. (2023), involved an orchestra recording a reference performance of the first movement of Beethoven's 4th Symphony (approximately 10 minutes). Each instrumental section then recorded their parts individually using the reference track as a guide, although the recordings were not strictly isolated. Finally, University of Rochester Multi-Modal Music Performance (URMP) by Li et al. (2019) features 44 pieces with various instruments, totaling 1.2 hours of isolated multitrack recordings. The methodology included a prerecorded video of a conductor accompanied by a pianist. Subsequent musicians were provided with this video and the piano audio as a guide track to ensure alignment during their recordings.

These datasets significantly influenced ChoraleBricks' design and methodology. While our dataset builds upon Bach10's approach, we integrate methodological insights from URMP, such as alignment strategies, to enhance the recording process. The key contribution of ChoraleBricks lies in its diversity of instruments, enabling dynamic creation of new mixes with varying complexity, thus broadening its utility for downstream tasks.

3 DATASET CREATION PROCESS

The dataset creation process, illustrated in Figure 2, consists of four main steps: selecting the works, recording conducting videos, sequentially recording individual tracks, and curating the data with accompanying annotations. In the following sections, we provide a detailed explanation of each step and share insights gained during the process.

3.1 WORK SELECTION

The first step involves selecting the works to include in the dataset. In ChoraleBricks, each work is a four-part

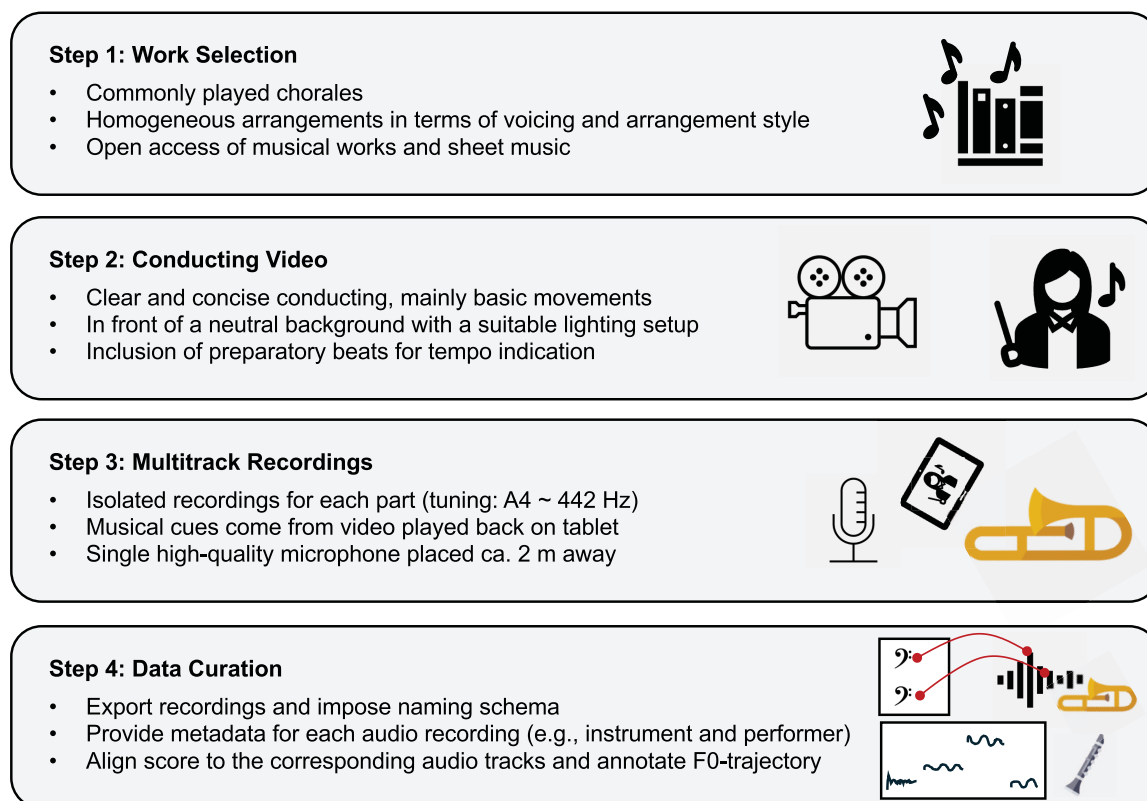


Figure 2 Overview of the creation process of the ChoraleBricks collection outlining the four main steps: work selection, conducting videos, recording of the individual tracks, and the data curation.

chorale consisting of soprano (S), alto (A), tenor (T), and bass (B) parts, with the soprano voice presenting the melody. During the work selection process, we focused on three key aspects: First, the chosen works should be representative of the chorale literature, particularly within the German tradition. Second, the arrangements should exhibit a homogeneous style and a comparable level of complexity—short and accessible enough for amateur musicians to perform. Finally, and crucially, the arrangements must be eligible for publication under an Open Access license.

A common resource in chorale music literature is the chorale accompaniment book, which typically provides four-part chorales corresponding to the entries in a related hymn book. One such accompaniment book is the *Neues Thüringer Choralbuch* (English: *New Thuringian Chorale Book*), originally compiled by Rudolf and Erhard Mauersberger (Mauersberger and Mauersberger, 1955).¹⁰ Although initially designed for organists, some of the chorale arrangements included in the book are derived from choir literature. The movements adhere to a strict arrangement with four monophonic parts. The counterpoint is predominantly homophonic, meaning all parts follow the rhythm of the melody, which (referred to as *cantus firmus* in Latin) is always in the soprano. The pitch ranges are comfortably singable by an SATB choir and playable by brass ensembles such as a *Posaunenchor* or woodwind groups. While the *Neues Thüringer Choralbuch* was initially created to accompany two now-outdated

hymn books (*Deutsches Evangelisches Gesangbuch* and *Evangelisches Kirchengesangbuch*), its chorale arrangements remain relevant. The melodies frequently appear in modern hymn books, and the four-part chorales continue to represent today's practices, offering flexible instrumentation options for a variety of ensembles.

Given the stylistic homogeneity of the arrangements in this particular chorale book (Baroque style with predominantly homophonic counterpoint), along with their relatively low technical skill requirements and flexibility for various instrumentations, we chose this book as the foundation for our dataset. For ChoraleBricks, we selected ten works from this book as a representative subset, as shown in Table 1. The selected works date back to the 16th, 17th, and 18th centuries and include compositions by renowned composers such as Bach, Telemann, and Vulpinus, among others.

Figure 3 presents a score rendition of a typical chorale from the dataset (“Auf, auf, mein Herz, mit Freuden” by Johann Crüger (CR1)). In the score, the first and second voices (soprano and alto) are notated in the treble clef on the first and second staff lines, while the third and fourth voices (tenor and bass) are notated in the bass clef.

In ChoraleBricks, the musical scores are provided in multiple formats. First, we use the MusicXML format. As shown in Figure 3, each voice is placed on a separate staff line, and all repetitions are expanded by duplicating the respective measures. These MusicXML encodings

form the foundation for symbolic music representations utilized in the audio-to-score alignment processes discussed later in the paper.

Additionally, the scores are encoded in the MEI symbolic music format (version 5.0), established as the international standard for digital scholarly music editions

Number	ID	Name	Year
1	AN1	Anonymous: “Aus meines Herzens Grunde”	1598
2	BA1	Bach, J. S.: “Ich steh’ an Deiner Krippe hier”	1736
3	CR1	Crüger, J.: “Auf, auf, mein Herz, mit Freuden”	1647
4	DR1	Drese, A.: “Jesu, geh voran”	1698
5	GE1	Gesius, B.: “Befiehl Du Deine Wege”	1603
6	GE2	Gesius, B.: “Du Friedensfürst, Herr Jesu Christ”	1601
7	JA1	Jan, M.: “Du großer Schmerzensmann”	1668
8	TE1	Telemann, G. P.: “Der lieben Sonne Licht und Pracht”	1730
9	VU1	Vulpis, M.: “Die helle Sonn leucht’ jetzt herfür”	1609
10	VU2	Vulpis, M.: “Christus, der ist mein Leben”	1609

Table 1 An overview of the musical works included in the ChoraleBricks collection.

Notes: The composers of the melodies are specified, while the four-part harmonizations were composed by Rudolf Mauersberger (Mauersberger and Mauersberger, 1955).

(Hankinson et al., 2011; Rol, 2023). MEI’s capabilities for annotating critical editions, analytical data, and variances across different versions of the *Neues Thüringer Choralbuch* make it ideal for this project and potential future expansions.

The graphical scores, such as the one shown in Figure 3, were rendered using MuseScore v4.0+ (Muse Group, 2022). In the original sheet music, the soprano and alto parts were combined as two voices in a single staff line in the treble clef, while the tenor and bass parts were combined in the bass clef. Although most musicians are accustomed to reading from the full score during recording sessions, the additional MusicXML encoding, derived from MEI, provides a flexible and standardized format for computational tasks.

3.2 RECORDING OF THE CONDUCTING VIDEO

A key insight from the URMP dataset preparation was that conducting videos effectively ensured synchronization across sequentially recorded tracks (Li et al., 2019). Building on this approach, we created conducting videos for all ten chorales. The videos feature a clear and accessible conducting style (performed by the first author) and tempi consistent with common performance practices. Each video begins with an inviting gesture and a smile toward the musician, typically followed by one measure with preparatory beats to establish the tempo. The chorales conclude with a fermata to signal the ending.

Figure 4a illustrates the technical setup for the video recordings. Various backgrounds, including white walls and sheets, were tested before settling on a gray paper backdrop commonly used in photography. Lighting was provided by a single light source positioned behind a

J. Crüger: Auf, auf, mein Herz, mit Freuden

Figure 3 “Auf, auf, mein Herz, mit Freuden” (CR1), melody by Johann Crüger, harmonization by Rudolf Mauersberger, rendered with MuseScore.

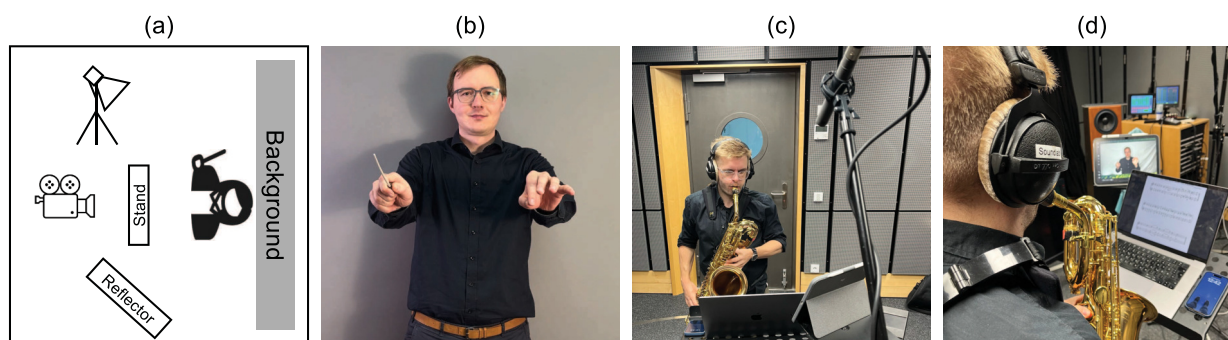


Figure 4 Impressions of the recording setup used to create the conducting videos and multitrack audio material. **(a)** Illustration of the setup used for the conducting videos. **(b)** A screenshot from one of the conducting videos. **(c)** A photo from a recording session in Erlangen featuring a baritone saxophone player. **(d)** A rear view of the same session, showcasing the setup with sheet music displayed on a laptop and the conducting video on a tablet. Faces in images are shown with consent.

diffuser on the right side, complemented by a reflector on the left to soften shadows. The recordings were made using a standard smartphone (Apple iPhone 13) mounted on a tripod approximately 2 meters in front of the conductor. Figure 4b presents a screenshot from one of the conducting videos as an example. It is worth noting that the lighting and background setup evolved during the video recordings, resulting in minor inconsistencies in appearance across the videos.

Although not entirely consistent, we decided to include the conducting videos in the dataset for two key reasons. First, they document the recording process and may offer a research avenue, such as analyzing the relationship between conducting tempo and performed tempo. Second, the videos could prove highly valuable for expanding the dataset in the future, enabling other musicians and researchers to contribute, potentially through web-based experiments.

3.3 RECORDING OF THE INDIVIDUAL TRACKS

The third step, depicted in Figure 2, is central to the dataset creation process: recording the individual tracks. As noted earlier, the primary objective of the ChoraleBricks dataset is to provide completely isolated recordings of each voice. To achieve this, musicians were recorded sequentially—a method that, while effective for ensuring data modularity, deviates significantly from traditional musical practices and the way this repertoire is typically performed. Recognizing this limitation, we carefully navigated the trade-off between developing an analytically robust and highly modular dataset and maintaining the natural musicality of the performances.

As for ChoraleBricks, we prioritized the flexibility afforded by a controlled recording process and setup. Figures 4c and 4d illustrate the arrangement used during the recording sessions. The conducting video was played back via a digital audio workstation (Apple Logic Pro, v11.1.1) and displayed on an iPad as a secondary

screen, ensuring synchronization between the video and audio during the recording process.

For tuning, we aimed for $A_4 = 442$ Hz, a standard frequency commonly used in modern wind ensembles and concert bands. Recordings were captured with a high-quality condenser cardioid microphone (Schoeps MK 4), positioned approximately 2 meters from the musician. This setup aimed to capture the sound as perceived by an audience or conductor rather than the dry sound typically produced by close-up or clip microphones. The microphone's angle was adjusted on the basis of the instrument, following recommendations from experience and relevant literature (Albrecht, 2017).

The sessions were conducted in three different environments: a rehearsal room (in Bödexen), a recording studio (in Erlangen), and an office (in Höxter). All locations featured acoustically controlled conditions, free from the excessive reverberation often found in churches or similar spaces.

In most chorales, the soprano and tenor voices were typically recorded first, often performed by the trumpet or flügelhorn. In some cases, however, the bass voice was recorded first. This was the case when a baritone horn player recorded the initial tracks of a chorale. Due to its pitch range, it is possible to record all four voices with the baritone horn (e.g., in VU2). Both strategies provided a reference for subsequent musicians, who could listen to the previously recorded tracks while following the conductor's movements during their own performance. This approach aligns with observations from the Bach10 dataset, where the initial player effectively “sets the scene” (Duan and Pardo, 2011). Similar to the methodology used in URMP (Li et al., 2019), the inclusion of a conducting video in our setup served as a valuable visual cue, ensuring consistent timing across all recordings.

Musicians reported that, as the number of prerecorded tracks increased, it became easier to play along with the audio. Once enough tracks were available, they could

request specific mixes or receive tailored ones on the basis of their preferences. A common preference was for mixes featuring similar instruments. For example, saxophone players found it easier to match intonation with other reed instruments, while brass players favored mixes dominated by brass instruments. Although not systematically documented, these preferences align with intuitive musician behavior: Familiar instruments provide more comfortable references for both recording sessions and live performances.

3.4 DATA CURATION

In the fourth step, as illustrated in Figure 2, the data were curated and enriched with additional annotations, including score-based audio annotations. This section details the post-processing steps, annotation procedures, and alignment processes, ensuring high-quality data preparation and comprehensive metadata to support diverse applications in MIR (see also Section 5).

3.4.1 Post-processing

For the audio recordings, variations in recording parameters, such as room acoustics, instruments, and microphone distances, introduced slight differences in perceived loudness. To address this bias, we adhered to the European Broadcasting Union (EBU) R 128 broadcast standard, normalizing all audio recordings to -23 loudness units full scale (LUFS) (EBU, 2023). A limiter set to -1 dB ensured that no tracks clipped after normalization. All recordings were exported as WAV files with a sampling frequency of 44.1 kHz. The normalization process was performed using the digital audio workstation Reaper (v7.30),¹¹ and the details of this step are documented on the accompanying website. We offer additional metadata such as the recording date, location, microphone used, performer identifier, and additional comments (e.g., “all notes played an octave higher than notated”) for each track.

For the conducting videos, we provide the files in 1080p full high-definition (HD) resolution at 30 frames per second, encoded using the H.264 format with the tool HandBrake (v1.9.0).¹² Each video is accompanied by the offset to the corresponding audio recording, enabling users to synchronize their individual mixes with the conducting video using tools such as ffmpeg.

3.4.2 Annotations

For annotations, we provide manually corrected fundamental frequency (F_0) trajectories and aligned note events (similar to a piano-roll representation) for each track.

F_0 Annotations

The F_0 trajectories were annotated using Sonic Visualiser (v5.0.1) and the pYIN VAMP plugin (v3) (Cannam et al.,

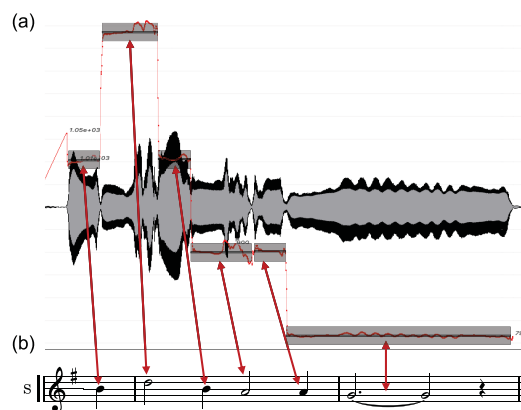


Figure 5 Last three measures of AN1. (a) Screenshot from Sonic Visualiser displaying the waveform with F_0 trajectories (red) and the note track (gray). (b) Corresponding sheet music of the soprano voice, with red arrows indicating the alignment between the two modalities.

2010; Mauch and Dixon, 2014), except for the tuba, where a different procedure was required owing to its very low pitches (details provided below). Figure 5a illustrates an example of the resulting annotations for a flute track. On the last note, regular amplitude modulations in the waveform and frequency modulation in the extracted F_0 trajectory (red curve) can be observed, producing a characteristic vibrato typical of many flute recordings. The F_0 trajectory was derived using the “pYIN: Smoothed Pitch Track” function with standard settings.

Additionally, a note track (Figure 5a, gray boxes), similar to a piano-roll representation, was extracted using the “pYIN: Notes” function with the same settings and adjustments as for the F_0 trajectories. Built-in sonification was used to verify both the F_0 trajectory and the note tracks. Manual corrections were necessary for the note tracks, particularly when consecutive notes were not accurately identified or when note onsets were audibly delayed.¹³ In cases of F_0 trajectory errors (primarily octave errors), a melodic range spectrogram was employed to manually correct the trajectory on the basis of the spectrogram data. For the tuba recordings, where low-note extraction quality was unsatisfactory, a different method was used. Note events were first extracted using Sonic Visualiser. Then, a salience-based F_0 estimator was adapted to incorporate constraint regions defined by the note events (Müller and Zalkow, 2021).¹⁴ Finally, both the F_0 trajectories and note tracks were exported as CSV files for each track.

The high quality of these manually corrected and verified annotations makes them well suited for tasks such as assessing pitch estimation methods in terms of pitch accuracy. The sonification process proved instrumental in detecting and rectifying voicing errors. Additionally, the pYIN implementation operated at a resolution of 20 cents, followed by parabolic interpolation, surpassing the 50-cent resolution commonly employed by evaluation

toolkits such as `mir_eval` (Raffel et al. 2014). However, two potential confounding factors should be considered when using them as a reference for evaluation. First, initial annotations were generated using the probabilistic YIN (pYIN) algorithm, which may introduce a bias toward this approach. Second, determining note onsets and durations is inherently ambiguous, and the annotations were adjusted by ear when necessary. This process could introduce uncertainty in note boundaries, potentially affecting evaluation metrics focused on the voicing capabilities of pitch estimators.

Aligned Score Annotations

In audio-to-score alignment, the task is to link recorded performances in physical time (given in seconds) to sheet music in musical time (given in measures), as described by Müller (2015). This process is often labor-intensive when performed manually. ChoraleBricks simplifies it by leveraging F_0 -annotated note events and monophonic note sequences derived from the sheet music. This enables a direct mapping between the F_0 annotations and note events extracted from MusicXML files using the `music21` library (Cuthbert and Ariza, 2010).

Figure 5 illustrates the annotation process: The note events from the F_0 annotations (Figure 5a, gray boxes) served as the basis for the performed notes in the audio recordings. Frequency values were converted into MIDI note numbers. On the sheet music side (Figure 5b), the corresponding note events were extracted from the “unrolled” MusicXML encoding (Figure 3) using the Python library `music21`. Once the number of note events was confirmed to match between the audio and sheet music, both sequences were linked (Figure 5, red arrows). Finally, pitch alignment between the audio and score was verified through sonifications using `libsoni` (Özer et al., 2024).

The aligned scores were exported to a CSV file, with each line representing a note event specified by its physical time in the audio and its measure position, encoded as a real number with a fixed precision of three digits. The integer component indicates the measure, while the fractional component represents the relative position within the measure. For instance, the note events from Figure 5b correspond to (start:end): (0.833:1.000), (1.000:1.333), (1.333:1.500), (1.500:1.833), (1.833:2.000), and (2.000:2.833).

This approach, however, does not retain connections to the original MusicXML note objects. To address this, we provide a second export using the Music Performance Markup (MPM) format (Berndt, 2022), an extension to symbolic music formats such as MEI and MusicXML with additional performance properties. Combined with the MPM Toolbox¹⁵ that provides built-in functionalities for alignment to audio recordings, these exports enable the creation of synthesized renditions with fine-grained control over performance parameters such as timing, dynamics, and articulation.

Each track in the ChoraleBricks collection includes a corresponding alignment file containing sheet music information and its link to the audio. The precise mapping between F_0 -annotated note events and the corresponding MusicXML-derived note sequences guarantees consistency across the dataset. The annotations were subjected to thorough manual verification, including sonification for detecting alignment errors. Verification of the alignments is subject to future investigations; however, we assume that alignment deviations stay below 50 milliseconds. While the alignments share the same limitations regarding note onset and duration ambiguities as the F_0 annotations do, their integration with ChoraleBricks’ modular mixing capabilities makes the dataset well suited for a variety of systematic experiments.

4 DATASET STATISTICS

We now provide an overview of the ChoraleBricks collection, detailing its statistical characteristics and overall structure. As outlined in Table 1, the collection features ten pieces by eight different composers. The earliest piece, AN1, dates back to 1598, while the most recent, BA1, was composed in 1736. These works were performed by 11 musicians using 13 distinct instruments. The performers, aged between 25 and 54 years (average age: 39.91 years), each had at least ten years of experience with their instrument and ensemble playing, for example, in orchestras. Although none of the performers were professional instrumentalists, they are considered skilled and dedicated amateur musicians.

Table 2 summarizes the musical properties of the ChoraleBricks collection and indicates the available parts for each chorale and instrument. Most arrangements are in sharp keys (6), with fewer in flat keys (2) or in concert key (2). The pieces feature a variety of time signatures, all based on a quarter-note metric, such as 3/4 or 6/4. Some pieces, such as CR1, exhibit frequent time signature changes, likely reflecting adaptations to modern notation. The collection includes pieces ranging from 9 to 19 measures in length, with durations between 16 and 59 seconds. Assuming one representative mix per piece, the collection amounts to a total duration of 6 minutes and 42 seconds.

ChoraleBricks includes a total of 193 tracks, distributed among the voice parts as follows: 62 soprano tracks, 57 alto tracks, 28 tenor tracks, and 46 bass tracks. The pieces with the highest number of tracks are CR1 and GE1 (21), while VU1 has the fewest (17). On average, each piece features 19.3 tracks. Regarding instruments, the baritone horn recorded the most tracks (24), while the trombone contributed the fewest (8). On average, each instrument is represented by 14.9 tracks. Notably, 11 out of the 13 instruments, including the flute and trumpet, have at least one track recorded for every piece. The combined duration of all unique tracks is 2 hours and 10 minutes.

ID	Properties				Available Parts per Instrument														Σ	#Ens.
	Key	TS	#Meas.	Dur.	fl	ob	eh	cl	bcl	as	bs	tp	fh	bar	fho	tb	tba			
AN1	B♭	4/4	14	00:57	S	S	A	SA	B	SA	TB	SA	SA	STB			B	18	280	
BA1	D	4/4	17	00:45	S	S	A	SA	B	SA	TB	SA	SA	TB	AT	B	B	20	540	
CR1	C	Div.	13	00:48	S	S	A	SA	TB	SA	TB	SA	SA	TB	T	TB	B	21	750	
DR1	G	3/4	12	00:32	S	S	A	SA	B	SA	TB	SA	SA	TB	AT	B	B	20	540	
GE1	C	4/4	19	00:45	S	S	A	SA	B	SA	TB	SA	SA	TB	AT	TB	B	21	720	
GE2	B♭	Div.	15	00:34	S	S	A	SA	TB	SA	B	SA	SA	ATB	AT		B	20	504	
JA1	G	Div.	11	00:40	S	S	A	SA	B	SA	TB	SA	SA	TB		B	B	18	300	
TE1	G	4/4	19	00:59	S	S	A	SA	B	SA	B	SA	SA	TB	AT		B	18	288	
VU1	D	6/4	9	00:16	S	S	A	SA	B	SA	TB	SA	SA	TB			B	17	240	
VU2	D	4/4	9	00:26	S	S	A	SA	B	SA	TB	SA	SA	SATB		B	B	20	420	
Σ			138	06:42	10	10	10	20	12	20	18	20	20	24	11	8	10	193	4582	

Table 2 Overview table of ChoraleBricks indicating the available parts for each chorale (ID) and instrument.

The column #Meas. specifies the measure count from the unrolled sheet music, including incomplete measures such as upbeat measures. The column #Ens. specifies the number of possible distinct four-voice ensemble mixes for a given chorale. Woodwinds: fl = flute, ob = oboe, eh = English horn, cl = clarinet, bcl = bass clarinet, as = alto saxophone, bs = baritone saxophone. Brass: tp = trumpet, fh = flugelhorn, bar = baritone horn, fho = French horn, tb = trombone, tuba = tuba. Div. = diverse; Dur. = duration; Ens. = ensemble; Meas. = measure; TS = time signature.

The modularity of the ChoraleBricks collection is a key strength, enabling various configurations of tracks to form valid four-voice ensembles. The last column of Table 2 (#Ens.) indicates the number of possible four-voice ensembles for each chorale. For instance, one ensemble might include flute (S), clarinet (A), baritone saxophone (T), and bass clarinet (B), while another could feature trumpet (S), flügelhorn (A), baritone horn (T), and tuba (B). The number of possible combinations ranges from 240 (VU1) to 750 (CR1), with an average of 458.2 valid four-voice ensembles per piece. Overall, this results in a total duration of 52 hours and 18 minutes of distinct audio that can be rendered across the various ensemble configurations. However, it is important to note that replacing a single instrument within an ensemble will result in a version with significant similarity to the original rendition. Consequently, the primary purpose of ChoraleBricks may not be to serve as a stand-alone dataset for training machine learning models but rather as a supplementary dataset to enhance data diversity or as an out-of-domain testbed.

Figure 6 illustrates the pitch distributions across all tracks for the soprano (S), alto (A), tenor (T), and bass (B) voices. The density curves demonstrate notable overlap in pitch ranges among the voices. The lowest pitch is MIDI pitch 28 (G_3), performed by the tuba in the bass voice, while the highest is MIDI pitch 86 (D_6), played by the flute in the soprano voice. The standard concert pitch, A_4 , commonly used for tuning, corresponds to MIDI pitch 69. The

average pitches for the four voices are 69.75 (S), 62.88 (A), 57.56 (T), and 47.04 (B). The soprano and bass voices span broader pitch ranges compared with the alto and tenor voices, reflecting the intentional decision to record flute tracks an octave higher and tuba tracks an octave lower than written. This approach enhances the dataset by extending the instruments' pitch ranges.

5 ACCOMPANYING SOFTWARE

Along with the data, we present another major contribution of this work: a Python-based software toolbox. Its core functionality is to act as an object-relational mapper for the dataset, meaning that all data entities are linked and represented as classes. Figure 7a illustrates the main components in a class diagram. In the chosen class structure, the SongDB class contains all Songs, each Song consists of multiple Tracks, and each Track provides access to the corresponding audio files and annotations, such as F_0 trajectories and note events. Metadata, such as player_id or microphone, are integrated through class properties.

The toolbox includes functionality for ensemble mixing, allowing users to randomly combine available tracks to form ensembles covering all four voices. As illustrated in Figure 7b, users can also apply optional gain adjustments to each voice for customization. For advanced features, such as panning or additional audio effects,

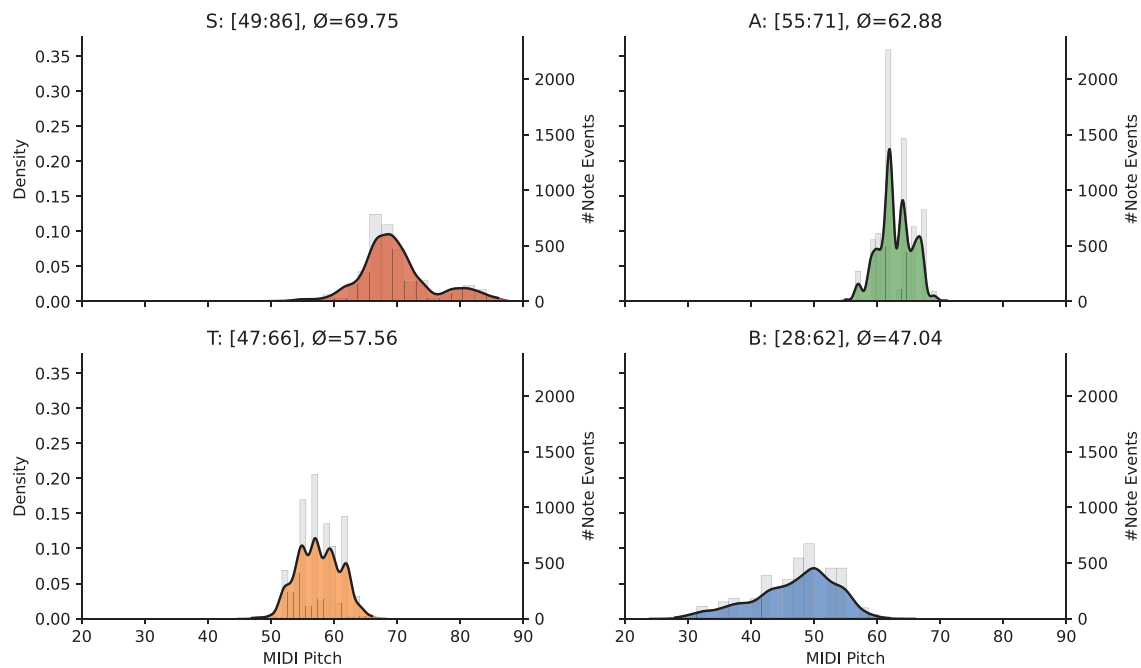


Figure 6 Pitch distributions visualized as smoothed density curves for each voice part (SATB). Background bars indicate the actual number of observations in the recorded tracks. Additionally, the pitch range and average pitch are specified for each part.

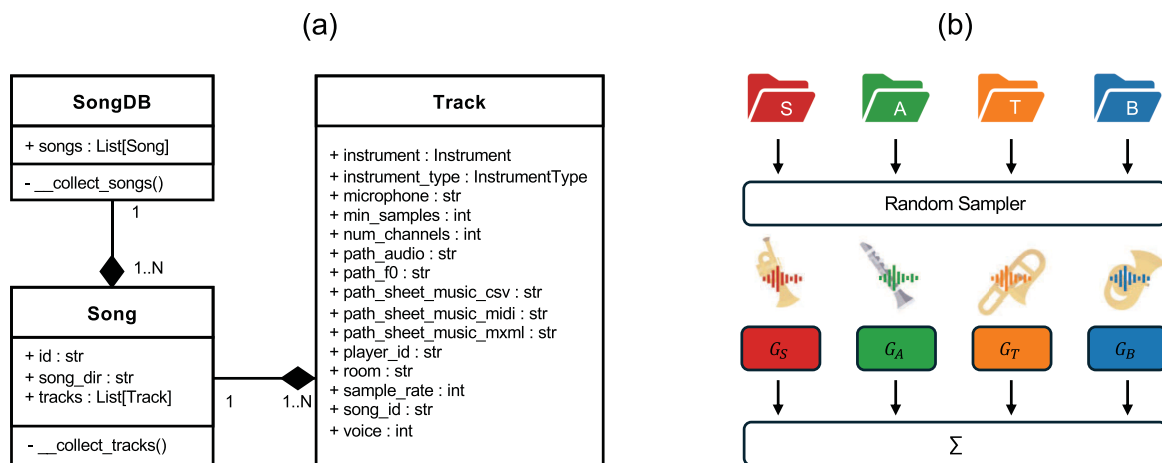


Figure 7 Overview of the Python package accompanying ChoraleBricks: **(a)** Class diagram illustrating the central toolbox components. **(b)** Visualization of random ensemble mixing using gain factors G_S , G_A , G_T , and G_B .

the code serves as a foundation for custom implementations. The toolbox is equipped with comprehensive documentation and example scripts for common tasks, such as iterating through tracks for F_0 evaluation or reproducing dataset statistics described in Section 4.

To ensure reliability, the code includes systematic unit tests and integrity checks for the dataset itself. These tests range from basic validations, such as verifying file existence and correct CSV headers, to advanced checks, such as matching audio file durations and detecting duplicates in F_0 trajectories (an issue encountered when exporting F_0 tracks from Sonic Visualiser). During the dataset creation process, these tests were instrumental in substantially reducing errors, such as accidental exports of incorrect annotation panes in Sonic Visualiser or incorrectly named files.

In addition to the Python toolbox, the dataset can be accessed and explored through an accompanying website. This platform provides an overview of the available songs, including representative mixes for each piece. Users can also create custom mixes for individual songs directly on the site. Figure 8 illustrates the web player interface for the song “Auf, auf, mein Herz, mit Freuden” (CR1). Below the main playback controls such as play and stop, the player provides a list of all available tracks for the song. Each track is categorized by its voice and labeled with the corresponding instrument name. Using the checkbox on the left side, users can select tracks and create their own mixes. In Figure 8, the selected tracks are “1. Alto Saxophone,” “2. Alto Saxophone,” “3. Baritone horn,” and “4. Tuba.”

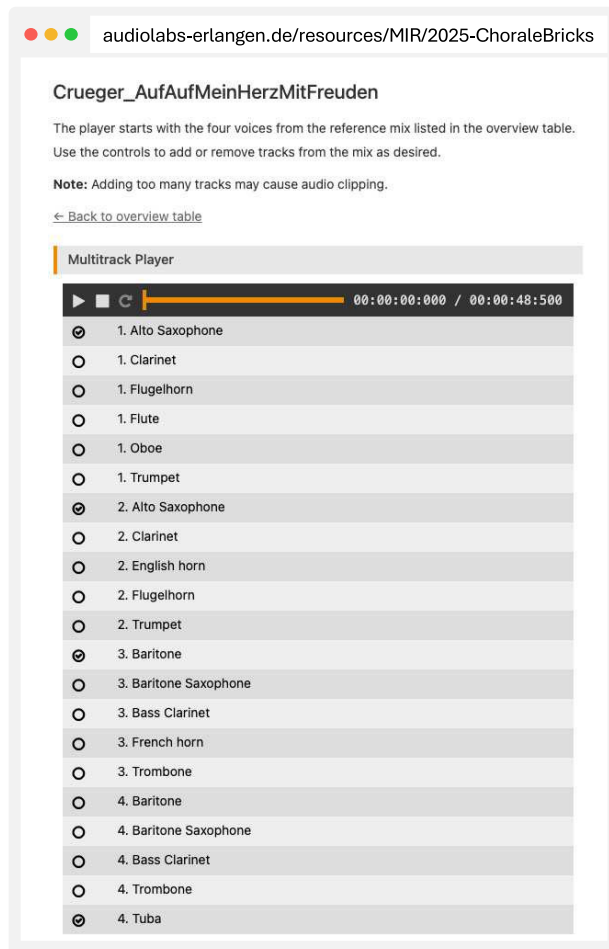


Figure 8 Screenshot of the web interface for creating individual mixes. The multitrack player is using trackswitch.js (Werner et al., 2017).

6 CONCLUSIONS AND POTENTIAL APPLICATIONS

In this article, we introduced ChoraleBricks, a versatile dataset that addresses a significant gap in MIR research by focusing on the underrepresented category of wind music and its associated instruments. We emphasized the cultural significance of wind music, which is deeply embedded in community and tradition, particularly within amateur and ensemble settings. By offering isolated multitrack recordings, symbolic sheet music representations, manually curated F_0 annotations, and audio-to-score alignments, ChoraleBricks serves as a novel and valuable resource for research and analysis.

The dataset also includes conducting videos, enabling extensions and further studies, such as performance synchronization and conducting analysis. Supported by a Python toolbox and an intuitive web interface, ChoraleBricks provides streamlined access and facilitates systematic experimentation. The raw data are hosted on Zenodo, and the accompanying code is available on GitHub, ensuring wide availability and accessibility for the research community.

We conclude this article by emphasizing the broad potential of ChoraleBricks for MIR research and beyond. Its modular design and diverse resources make it a valuable tool for advancing fundamental MIR tasks, fostering innovation in music education and generation, and supporting studies in cultural heritage and traditions.

6.1 F_0 ESTIMATION AND TRANSCRIPTION

F_0 estimation and automatic music transcription (AMT) are foundational areas in MIR, with state-of-the-art methods often relying on neural-network-based approaches (Kim et al., 2018). However, applications such as real-time intonation monitoring require low-latency estimates, where traditional methods such as YIN (de Cheveigné and Kawahara, 2002) or the sawtooth waveform inspired pitch estimator (SWIPE) (Camacho and Harris, 2008) remain viable owing to their simplicity and efficiency. In such contexts, robustness and efficiency are just as important as accuracy.

ChoraleBricks, with its clean, isolated recordings, provides a robust framework for testing F_0 estimators under diverse conditions, such as simulated crosstalk in ensemble performances. Additionally, the dataset can mitigate biases inherent in annotations (e.g., pYIN-based errors or subjective onset/offset discrepancies) through synthetic approaches using sinusoidal models (Salamon et al., 2017), enabling more objective evaluations across different F_0 estimation methods.

In AMT, early research primarily targeted piano recordings, driven by large datasets such as MIDI and Audio Edited for Synchronous Tracks and Organization (MAESTRO) (Hawthorne et al., 2019). Recent advancements, including transformer-based models (Maman and Bermano, 2022), have broadened transcription applications to other instruments. However, their performance on diverse datasets such as URMP (Li et al., 2019) remains limited. ChoraleBricks, with its varied instrumentation and modular design, offers an essential resource for evaluating and enhancing AMT models, addressing transcription challenges across a broader spectrum of instruments.

6.2 INTONATION ANALYSIS

In ensemble performances, achieving rhythmic precision and balanced intonation is crucial for a cohesive sound. While ensembles tune their instruments to a reference pitch (e.g., $A_4 = 442$ Hz) at the start of a session, factors such as temperature and humidity, particularly for woodwind instruments, can lead to intonation shifts during rehearsals or concerts. Players adjust their tuning dynamically through subtle variations in their playing techniques rather than altering the instrument itself.

Harmonic context also necessitates pitch adjustments, as notes in a major triad (e.g., root, third, or fifth) often require fine-tuning to align with harmonic

expectations. Studies such as that of Howard (2007) have shown intonation drifts in vocal ensembles, often due to modulations, with singers tending to non-equal temperament tunings. Similar analyses could be conducted for wind ensembles using the ChoraleBricks dataset.

ChoraleBricks' modular design enables the exploration of various instrument combinations and harmonic complexities, offering opportunities to define or learn cost measures for intonation quality (Schwär et al., 2021). Such measures could help identify well-intonated ensembles and investigate phenomena, such as whether larger ensembles achieve better intonation than smaller ones—an observation often noted in amateur settings.

6.3 AUDIO ALIGNMENT

Audio-score alignment, along with related problems such as score following and computer accompaniment, has been extensively studied (Dannenberg and Raphael, 2006; Müller et al., 2019). While much of this research focuses on piano music, orchestral and opera music pose unique challenges owing to their complexity. In particular, inconsistencies arising from musical and performance-specific factors further complicate the alignment of score elements, such as notes or measures, with corresponding audio positions. These asynchronies between orchestral and vocal elements are commonly found in opera recordings (Weiß et al., 2016).

In this context, the ChoraleBricks dataset offers valuable opportunities for studying these challenges by providing alignments at the track level, allowing for evaluations at the instrument level. This level of detail enables researchers to explore alignment behaviors more thoroughly, such as identifying which instruments an alignment approach tends to prioritize, for example, trumpets in the soprano voice or tuba in the bass voice. Such analyses yield important insights into the performance and refinement of alignment methods, contributing to advancements in audio-to-score alignment research.

6.4 MUSIC EDUCATION

Music education apps such as Yousician¹⁶ and flowkey¹⁷ have become popular tools for beginners, offering features such as just-in-time feedback while practicing. Most apps focus on widely used instruments such as piano, guitar, and bass. Some, such as Songs2See,¹⁸ also support wind instruments. For more advanced aspects, such as intonation and sound quality, specialized tools such as Korg Cortosia analyze tone properties, including pitch and timbre stability (Bandiera et al., 2016).

The ChoraleBricks dataset could expand these educational approaches by facilitating interactive training environments for ensemble playing, potentially through a web interface. For example, students could practice specific voices from chorales and receive feedback on accuracy and timing. As proficiency increases, they could transition to play-along scenarios with customizable

challenges, such as changing instruments, adding reverberation effects, or simulating tuning shifts. Additionally, user-generated recordings could extend the dataset with crowdsourced performances, creating a valuable resource for evaluating F_0 estimation and transcription methods.

6.5 MUSIC SOURCE SEPARATION

Music source separation (MSS) seeks to isolate and extract individual audio components (e.g., vocals, drums, bass, guitar) from a mixed recording, allowing the creation of separate tracks for each source. Recent progress in MSS has been driven by efforts such as the Sound Demixing Challenge (Fabbro et al., 2024) and datasets such as MUSDB18 (Rafii et al., 2017) and MoisesDB (Pereira et al., 2023). While MSS systems excel with Rock and Pop instruments, current research aims to achieve universal source separation to address instrument-specific limitations (Watcharasupat and Lerch, 2024). With its diverse instrumentation and isolated tracks, the ChoraleBricks dataset provides a valuable resource for advancing and evaluating MSS techniques.

6.6 MUSIC GENERATION

Recent advancements in generative models, such as denoising diffusion probabilistic models (DDPM) (Hawthorne et al., 2022; Maman et al., 2024), leverage large datasets to synthesize audio directly from symbolic music representations. To reduce data requirements while enhancing interpretability, model-based deep learning combines traditional knowledge-driven methods with data-driven techniques within a differentiable computing framework (Richard et al., 2024). An example is the differentiable digital signal processing (DDSP) framework (Engel et al., 2020), which integrates digital signal processing (DSP) elements into deep learning pipelines and has gained prominence in musical sound synthesis.

ChoraleBricks contributes to advancing music generation, particularly for wind instruments, by offering high-quality, modular recordings and precise per-track note-wise annotations. These resources support the development of generative models capable of synthesizing realistic performances from symbolic music data. For instance, ChoraleBricks may facilitate research using DDPMs, DDSP, and hybrid frameworks, helping to address confounding factors and enabling innovative ensemble generation with detailed control over timbre, articulation, and dynamics.

ACKNOWLEDGEMENTS

We thank all the participating musicians who spend their time contributing to this dataset, namely (in alphabetical order) Heike Balke (oboe, english horn), Stefan Balke (trumpet, flugelhorn), Hans-Ulrich Berendes

(French horn), Axel Berndt (baritone horn), Nicole Krois (clarinet), Melanie Lange (bass clarinet), Adrian Maiworm (baritone horn, trombone, tuba), Kristina Mengersen (flute), Bettina Quest (alto saxophone), Simon Schwär (alto saxophone, baritone saxophone), and Thomas Specht (French horn). Furthermore, we want to thank Aida Amiryan-Stein for encoding the original sheet music in MEI. Last, but not least, a special thanks to Peter Meier and Simon Schwär for their contributions to the Python tooling and for being the first users of the dataset.

The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS.

FUNDING INFORMATION

This work was funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation) under grant number 500643750 (MU 2686/15-1) and under Grant No. 555525568 (MU 2686/18-1). The music encoding is supported by KreativInstitut.OWL: a consortium consisting of Ostwestfalen-Lippe (OWL) University of Applied Sciences and Arts, Detmold University of Music, and Paderborn University, funded by the Ministry of Economic Affairs, Industry, Climate Action and Energy of the State of North Rhine-Westphalia, Germany.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHORS' CONTRIBUTIONS

SB, AB, and MM designed and outlined the dataset. AB selected the works and provided the sheet music encodings. SB conducted the audio and video recordings, processed the data, implemented the Python toolbox, created the website, and wrote the initial version of the paper. MM substantially contributed to the writing of the article. All co-authors of this paper actively participated in the preparation of the final manuscript.

DATA ACCESSIBILITY

To foster reproducibility and further research, we provide access to the ChoraleBricks dataset and related resources.

- **Dataset:** The ChoraleBricks dataset, including multitrack recordings, annotations, and related metadata, is available under a Creative Commons Attribution 4.0 International (CC-BY 4.0) license on Zenodo: <https://doi.org/10.5281/zenodo.15081741>.

- **Code:** The source code for processing and analyzing the dataset, as well as implementation details, is available under a Massachusetts Institute of Technology (MIT) license on GitHub: <https://github.com/stefan-balke/choralebricks>.
- **Demos:** Interactive demonstrations showcasing dataset applications can be accessed at: <https://audiolabs-erlangen.de/resources/MIR/2025-ChoraleBricks>.

NOTES

1. Stevie Wonder. "Sir Duke." *Songs in the Key of Life*, 1976.
2. <https://doi.org/10.5281/zenodo.15081741>
3. <https://github.com/stefan-balke/choralebricks>
4. <https://audiolabs-erlangen.de/resources/MIR/2025-ChoraleBricks>
5. <https://www.igeb.net/>
6. <https://www.wasbe.online/>
7. <https://institut-oberschuetzen.kug.ac.at/>
8. <https://cadenzachallenge.org/>
9. <https://github.com/Chorale-Corpus>
10. MEI encodings of the complete book are available at <https://github.com/axelberndt/Neues-Thueringer-Choralbuch-digital>.
11. <https://www.reaper.fm/>
12. <https://handbrake.fr/>
13. These errors occurred frequently, particularly when instruments were played with soft onsets, such as the baritone saxophone. However, the Sonic Visualiser interfaces facilitated efficient error correction.
14. Reference implementation and further description: <https://audiolabs-erlangen.de/resources/MIR/FMP/C8/C8S2-FundFreqTracking.html>.
15. <https://github.com/axelberndt/MPM-Toolbox>, visited Nov. 2024.
16. <https://yousician.com>, visited Dec. 2024.
17. <https://www.flowkey.com>, visited Dec. 2024.
18. <https://www.songquito.com>, visited Dec. 2024.

AUTHOR AFFILIATIONS

Stefan Balke  <https://orcid.org/0000-0003-1306-3548>
International Audio Laboratories Erlangen, Germany

Axel Berndt  <https://orcid.org/0009-0006-2807-592X>
Paderborn University, Germany

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

REFERENCES

- Albrecht, C.** (2017). *Der Tonmeister*. (2nd ed.). Schiele & Schön.
- Bandiera, G., Picas, O. R., Tokuda, H., Hariya, W., Oishi, K., and Serra, X.** (2016). Good-sounds. org: A framework to explore goodness in instrumental sounds. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, New York City, New York, USA.
- Battisti, F. L.** (2018). *The new winds of change: The evolution of the contemporary American wind band/ensemble and its music*. The New Winds of Change, Hal Leonard Corporation. 978-1-57463-474-7.

- Berndt, A.** (2022). The music performance markup format and ecosystem. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India. <https://doi.org/10.5281/zenodo.5624429>
- 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).
- Camacho, A., and Harris, J. G.** (2008). A sawtooth waveform inspired pitch estimator for speech and music. *Journal of the Acoustical Society of America*, 124(3), 1638–1652.
- 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).
- Cuthbert, M. S., and Ariza, C.** (2010). Music21: A toolkit for computer-aided musicology and symbolic music data. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, The Netherlands (pp. 637–642).
- Dannenbergh, R. B., and Raphael, C.** (2006). Music score alignment and computer accompaniment. *Communications of the ACM, Special Issue: Music Information Retrieval*, 49(8), 38–43.
- de Cheveigné, A., and Kawahara, H.** (2002). YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America (JASA)*, 111(4), 1917–1930.
- Duan, Z., and Pardo, B.** (2011). Soundprism: An online system for score-informed source separation of music audio. *IEEE Journal of Selected Topics in Signal Processing*, 5(6), 1205–1215.
- EBU.** (2023). R 128—loudness normalisation and permitted maximum level of audio signals. Technical report. European Broadcasting Union. <https://tech.ebu.ch/docs/r/r128.pdf>
- Engel, J., Hantrakul, L., Gu, C., and Roberts, A.** (2020). DDSP: Differentiable digital signal processing. In *Proceedings of the International Conference on Learning Representations (ICLR)*, Virtual. <https://openreview.net/forum?id=B1x1ma4tDr>
- Fabbro, G., Uhlich, S., Lai, C.-H., Choi, W., Martínez-Ramírez, M., Liao, W., Gadelha, I., Ramos, G., Hsu, E., Rodrigues, H., Stöter, F., Défossez, A., Luo, Y., Yu, J., Chakraborty, D., Mohanty, S., Solovyev, R., Stempkovskiy, A., Habruseva, T., . . . Mitsufuji, Y.** (2024). The sound demixing challenge 2023—music demixing track. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 7(1), 63–84. <https://doi.org/10.5334/tismir.171>
- Gerhardt, K., and Kirsch, M.** (2024). Choralbücher from Northern Germany c. 1800—a Dataset for Studies in Hymnology, Music Culture and Figured Bass. *Transactions of the International Society for Music Information Retrieval*, 7(1), 2514–3298. <https://transactions.ismir.net/articles/10.5334/tismir.191> <https://doi.org/10.5334/tismir.191>
- Hankinson, A., Roland, P., and Fujinaga, I.** (2011, October). The music encoding initiative as a document-encoding framework. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, Florida, USA (pp. 293–298).
- Hawthorne, C., Simon, I., Roberts, A., Zeghidour, N., Gardner, J., Manilow, E., and Engel, J. H.** (2022). Multi-instrument music synthesis with spectrogram diffusion. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 598–607). <https://archives.ismir.net/ismir2022/paper/000072.pdf>
- 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. <https://openreview.net/forum?id=r1YRjC9F7>
- Hofer, A.** (1992). *Blasmusikforschung: Eine kritische Einführung* (1st ed.). Wissenschaftliche Buchgesellschaft Darmstadt. 3-534-11083-8
- Howard, D. M.** (2007). Intonation drift in a capella soprano, alto, tenor, bass quartet singing with key modulation. *Journal of Voice*, 21(3), 300–315.
- Kaiser, U., Mestemacher, I., and Vieregg, M.** (2023). Operation Beethoven. Ein Projekt der Open Music Academy in Kooperation mit der Hofkapelle München. <https://openmusic.academy/revs/JwXCJwcEyAXsJDNsU7DtN3>
- Kepper, J., and Roland, P. D.** (2023). *The music encoding initiative guidelines version 5.0*. Music Encoding Initiative. <https://doi.org/10.5281/ZENODO.10040258>
- Kim, J. W., Salamon, J., Li, P., and Bello, J. P.** (2018). CREPE: A convolutional representation for pitch estimation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada (pp. 161–165). <https://doi.org/10.1109/ICASSP.2018.8461329>
- Li, B., Liu, X., Dinesh, K., Duan, Z., and Sharma, G.** (2019). Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. *IEEE Transactions on Multimedia*, 21(2), 522–535.
- 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)*, Baltimore, Maryland, USA (pp. 14918–14934).
- Maman, B., Zeitler, J., Müller, M., and Bermano, A. H.** (2024). Performance conditioning for diffusion-based multi-instrument music synthesis. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Seoul, South Korea (pp. 5045–5049).
- Mauch, M., and Dixon, S.** (2014). pYIN: A fundamental frequency estimator using probabilistic threshold distributions. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy (pp. 659–663).

- Mauersberger, R., and Mauersberger, E.** (1955). *Neues Thüringer Choralbuch* (1st ed.). Evangelische Verlagsanstalt Berlin.
- McFee, B., Kim, J. W., Cartwright, M., Salamon, J., Bittner, R. M., and Bello, J. P.** (2019). Open-source practices for music signal processing research: Recommendations for transparent, sustainable, and reproducible audio research. *IEEE Signal Processing Magazine*, 36(1), 128–137.
- Müller, M.** (2015). *Fundamentals of Music Processing—Audio, Analysis, Algorithms, Applications* (pp. 1–480). Springer Verlag. <https://doi.org/10.1007/978-3-319-21945-5>
- 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>, <https://ieeexplore.ieee.org/document/8588416/>
- Müller, M., and Zalkow, F.** (2021). libfmp: A Python package for fundamentals of music processing. *Journal of Open Source Software (JOSS)*, 6(63), 1–5, 3326. <https://doi.org/10.21105/joss.03326>, 2021_MuellerZalkow_libfmp_JOSS.pdf, <https://github.com/meinardmueller/libfmp>
- Muse Group.** (2022, December). *MuseScore v4.0*. <https://musescore.org>
- Niemann, N.** (2006). *Bläserklang im Gottes-Dienst—Ein Streifzug durch 3000 Jahre Gotteslob*. Verein zur Förderung der Posaunenchorarbeit in der Evangelisch-lutherischen Landeskirche in Braunschweig e.V., Germany.
- Özer, Y., Brütting, L., Schwär, S., and Müller, M.** (2024). libsoni: A Python toolbox for sonifying music annotations and feature representations. *Journal of Open Source Software (JOSS)*, 9(96), 1–6. <https://doi.org/10.21105/joss.06524>
- Peeters, G., and Richard, G.** (2021). *Deep Learning for Audio and Music, Multi-faceted Deep Learning: Models and Data* (pp. 231–266). Springer International Publishing. https://doi.org/10.1007/978-3-030-74478-6_10
- Pereira, I., Araújo, F., Korzeniowski, F., and Vogl, R.** (2023). MoisesDB: A dataset for source separation beyond 4-stems. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy (pp. 619–626).
- Raffel, C., McFee, B., Humphrey, E. J., Salamon, J., Nieto, O., Liang, D., and Ellis, D. P. W.** (2014). MIR_EVAL: A transparent implementation of common MIR metrics. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, Taiwan (pp. 367–372).
- Rafii, Z., Liutkus, A., Stöter, F., Mimitakis, S. I., and Bittner, R.** (2017). The MUSDB18 corpus for music separation. <https://doi.org/10.5281/zenodo.1117372>
- Richard, G., Lostanlen, V., Yang, Y.-H., and Müller, M.** (2024). Model-based deep learning for music information research: Leveraging diverse knowledge sources to enhance explainability, controllability, and resource efficiency. *IEEE Signal Processing Magazine*, 41(6), 51–59.
- Richard, G., Smaragdis, P., Gannot, S., Naylor, P. A., Makino, S., Kellermann, W., and Sugiyama, A.** (2023). Audio signal processing in the 21st century: The important outcomes of the past 25 years. *IEEE Signal Processing Magazine*, 40(5), 12–26. <https://doi.org/10.1109/MSP.2024.3415569>, <https://ieeexplore.ieee.org/document/10819669>
- Roa Dabike, G., Cox, T. J., Miller, A. J., Fazenda, B. M., Graetzer, S., Vos, R. R., Akeroyd, M. A., Firth, J., Whitmer, W. M., Bannister, S., Greasley, A., and Barker, J. P.** (2024). The cadenza woodwind dataset: Synthesised quartets for music information retrieval and machine learning. *Data in Brief*, 57, 2352–3409. <https://doi.org/10.1016/j.dib.2024.11199>, <https://www.sciencedirect.com/science/article/pii/S2352340924011612>
- Salamon, J., Bittner, R. M., Bonad, J., Vicente, J. J. B., Gómez, E., and Bello, J. P.** (2017). An analysis/synthesis framework for automatic f_0 annotation of multitrack datasets. In *Proceedings of International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China (pp. 71–78). <https://doi.org/10.5281/zenodo.1415588>, MDB-melodysynth
- Sarkar, S., Benetos, E., and Sandler, M.** (2022). EnsembleSet: A new high quality dataset for chamber ensemble separation. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Bengaluru, India (pp. 625–632).
- Schwär, S., Rosenzweig, S., and Müller, M.** (2021). A differentiable cost measure for intonation processing in polyphonic music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 626–633).
- Watcharasupat, K. N., and Lerch, A.** (2024). A stem-agnostic single-decoder system for music source separation beyond four stems. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, San Francisco, CA, USA.
- 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). <https://doi.org/10.5281/zenodo.1417449>
- Werner, N., Balke, S., Stöter, F.-R., Müller, M., and Edler, B.** (2017). trackswitch.js: A versatile web-based audio player for presenting scientific results. In *Proceedings of the Web Audio Conference (WAC)*, London, UK. 2017 WernerBSME TrackswitchJSPlayer WAC.pdf
- Wu, Y., Gardner, J., Manilow, E., Simon, I., Hawthorne, C., and Engel, J.** (2022). The Chamber ensemble generator: Limitless high-quality MIR data via generative modeling. <http://arxiv.org/abs/2209.14458>, 2024-03-18, arXiv

TO CITE THIS ARTICLE:

Balke, S., Berndt, A., & Müller, M. (2025). ChoraleBricks: A Modular Multitrack Dataset for Wind Music Research. *Transactions of the International Society for Music Information Retrieval*, 8(1), 39–54. DOI: <https://doi.org/10.5334/tismir.252>

Submitted: 17 January 2025 **Accepted:** 27 March 2025 **Published:** 30 April 2025

COPYRIGHT:

© 2025 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 <https://creativecommons.org/licenses/by/4.0/>.

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