



RWC Revisited: Towards a Community-Driven MIR Corpus

DATASET ARTICLE

STEFAN BALKE JOHANNES ZEITLER VLORA ARIFI-MÜLLER BRIAN MCFEE TOMOYASU NAKANO MASATAKA GOTO MEINARD MÜLLER 

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

ABSTRACT

The Real World Computing (RWC) Music Database has been a cornerstone of Music Information Retrieval (MIR) research for over two decades, offering high-quality recordings across multiple genres, including popular, classical, and jazz music. Beyond its extensive audio collection, the dataset is enriched by aligned Musical Instrument Digital Interface (MIDI) encodings and complementary annotations, including beat, structure, and chord labels, making it a valuable resource for music structure analysis, beat tracking, chord recognition, automatic transcription, and music synchronization. Originally, the RWC audio material was distributed on physical media and made available for purchase at a nominal price. A significant development, announced and initiated with this paper, is the release of the RWC dataset under a Creative Commons license, making it freely accessible for research purposes. This transition significantly enhances the dataset's usability and supports broader adoption within the MIR research community. We outline the steps taken to enable this release and share a vision for transforming RWC into a community-driven resource that promotes open research and collaboration. With the audio recordings now hosted on Zenodo, we also discuss strategies for dataset maintenance, annotation expansion, and reproducibility through collaborative platforms such as GitHub. This shift promotes transparency and inclusivity, helping to ensure the dataset's continued relevance for cutting-edge MIR research. We further revisit the historical significance of the RWC dataset, incorporating insights from an interview with its original creator, Masataka Goto, and provide an overview of its current applications and future potential. In summary, by embracing an open and community-supported approach, we aim not only to renew the dataset's impact and preserve its legacy within the MIR community but also to shed light on broader best practices for open, collaborative, and sustainable research infrastructures.

CORRESPONDING AUTHOR:

Stefan Balke

International Audio Laboratories Erlangen,
Germany
stefan.balke@audiolabs-
erlangen.de

KEYWORDS:

open data, data curation,
community dataset

TO CITE THIS ARTICLE:

Balke, S., Zeitler, J., Arifi-Müller, V., McFee, B., Nakano, T., Goto, M., & Müller, M. (2026). RWC Revisited: Towards a Community-Driven MIR Corpus. *Transactions of the International Society for Music Information Retrieval*, 9(1), 21–35.
DOI: <https://doi.org/10.5334/tismir.326>

]u[ubiquity press

1 INTRODUCTION

The Real World Computing (RWC) Music Database has played a foundational role in Music Information Retrieval (MIR) research for over two decades (Goto, 2004; Goto et al., 2002, 2003). Its carefully curated audio recordings—covering popular, classical, jazz, and other genres—together with aligned Musical Instrument Digital Interface (MIDI) files and a variety of annotations (e.g., beat, structure, and chords (Goto, 2006)), have supported a broad range of tasks, including structure analysis (Paulus and Klapuri, 2009; Wang et al., 2022), beat tracking (Böck et al., 2016), chord recognition (Korzeniowski and Widmer, 2016), automatic music transcription (Dessein et al., 2010), and synchronization (Ewert et al., 2009).

Released at a time when freely accessible music data were scarce due to copyright constraints, the RWC Music Database addressed a critical need by providing royalty-free, original content under clear legal conditions. Distributed on physical compact discs (CDs) at moderate cost, it enabled reproducible research and quickly became a widely used benchmark in the MIR community. The dataset includes five main sub-collections for music—Pop (RWC-P), Royalty-Free (RWC-R), Classical (RWC-C), Jazz (RWC-J), and Genre (RWC-G)—which together include 328 tracks totaling approximately 23 hours of audio, as illustrated in Figure 1.¹ A sixth sub-collection, the Instrument Database (RWC-I), features isolated recordings of 50 instruments but is not within the scope of this work. Over time, RWC evolved into a de facto standard and became integral to evaluation campaigns such as MIREX,² with over 1,700 citations to its core publications (see Figure 2).

This paper marks a pivotal moment in the history of RWC: for the first time, the audio recordings are being made openly available under a Creative Commons license. Hosted on Zenodo in lossless quality and paired with a GitHub repository for collaborative development, the RWC dataset is transitioning into a fully open and community-supported resource.

Beyond the dataset release itself, this paper has two main objectives: it documents and contextualizes the transition of the RWC dataset into the open research domain, and it offers a broader reflection on the role of open, community-driven data practices within the MIR field. By revisiting the origins and impact of RWC—including insights from an interview with its original creator, Masataka Goto—we examine what it takes to sustain a dataset over decades and how collaborative engagement can help renew and extend its relevance. In doing so, this paper aims to serve both as a practical roadmap for similar initiatives and as a case study demonstrating how legacy resources can be revitalized through open and inclusive collaboration.



Figure 1 Overview of the RWC Music Database. Illustration includes a comic-style portrait of Masataka Goto, generated using OpenAI's DALL-E model via ChatGPT.

In the remainder of this paper, we first provide historical background on the creation and evolution of the RWC Music Database in Section 2. We then situate RWC within the broader landscape of open-source and open-data practices in MIR in Section 3. Sections 4 and 5 offer detailed insights into the structure of the dataset's audio and MIDI components, respectively. In Section 6, we explore strategies for collaborative dataset development and outline opportunities for future community contributions. Finally, in Section 7, we summarize our key findings and reflect on broader implications for sustaining and advancing community-driven MIR resources.

2 RWC BACKGROUND

This section outlines the historical foundations of the RWC Music Database, drawing on insights from an interview conducted by Meinard Müller and Stefan Balke with Masataka Goto on February 18, 2025. We highlight the key developments and motivations behind the creation of the RWC dataset, as well as its evolving role within the MIR community since its initial release. For a more personal and in-depth account of the motivations, design choices, and early development of the RWC dataset, we refer readers to the edited transcript of the interview (Müller et al., 2025).

In the 1990s, Masataka Goto recognized a pressing need for annotated musical data—first to evaluate his work on beat tracking as a PhD student, and later for melody extraction and fundamental frequency (F0) estimation as a researcher at the Electrotechnical Laboratory (later the National Institute of Advanced Industrial Science and Technology (AIST)). Faced with limited resources, he manually annotated beats and F0 contours of commercial recordings using custom tools

| Year | Comment | Sub.-Coll. |
|---------|--|------------|
| 2000/01 | Dataset design and recordings | |
| 2002 | RWC presentation (Goto et al., 2002) | {P,R,C,J} |
| 2003 | RWC extension (Goto et al., 2003) | {G,I} |
| 2006 | AIST Annotations (Goto, 2006) | |
| 2009 | Music alignment (Ewert et al., 2009) ³ | C |
| 2011 | Chord annotations (Cho and Bello, 2011) ⁴ | P |

Table 1 Key developments in the evolution of the RWC Music Database.

Abbreviations: P = Pop, R = Royalty-Free, C = Classical, J = Jazz, G = Genre, I = Instruments.

he had developed. These early challenges—especially data scarcity and the difficulty of reproducible evaluation—led to the idea of creating a publicly accessible, copyright-cleared dataset for computer-driven music research. This vision aligned with Japan’s RWC program, a national effort aimed at advancing intelligent information processing. Supported by this program, Goto and his team developed the RWC Music Database, designed to provide high-quality, professionally produced audio recordings across diverse musical genres. The dataset was created to support the emerging field of Music Information Processing, a precursor to what is now known as MIR.

Significant milestones of the RWC database and related publications are listed in [Table 1](#). In the years 2000 and 2001, the RWC dataset was mainly designed by Masataka Goto. For certain genres, such as jazz and classical music, external experts Keiji Hirata and Yuzuru Hiraga were consulted, respectively, to help in the selection of suitable pieces. However, rather than licensing commercial music, Goto decided to commission original recordings to ensure full copyright clearance. With the help of a music production company, he created a dataset of recordings by professional musicians and ensembles—including the Tokyo City Philharmonic Orchestra—covering a broad repertoire ranging from pop and jazz to classical music and instrumental sounds. Following the production phase, the first RWC collections—Pop, Royalty-Free, Jazz, and Classical—were introduced at ISMIR 2002 ([Goto et al., 2002](#)). The Genre and Instruments collections were subsequently presented at ISMIR 2003 ([Goto et al., 2003](#)).

Beyond audio, Goto recognized the importance of annotations for advancing research and prepared MIDI files that were created after music production by professional music transcribers from the karaoke industry. Although the audio recordings were distributed on CDs, these MIDI files and lyric text files have been distributed online from the beginning. However, the MIDI files are not necessarily aligned with the audio, and the original

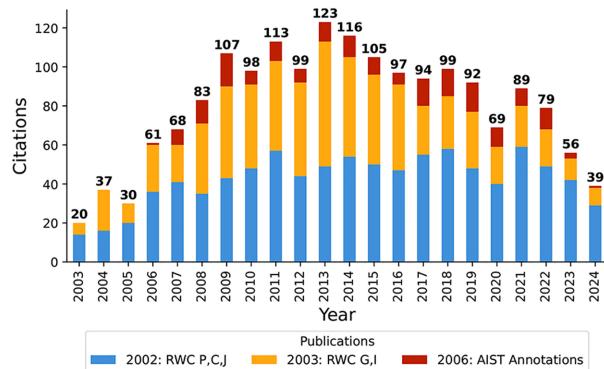


Figure 2 Citation counts from 2003–2024 for [Goto et al. \(2002\)](#), [Goto et al. \(2003\)](#), and [Goto \(2006\)](#), respectively.

RWC dataset does not contain annotations other than lyrics and track-level metadata. Goto therefore developed custom annotation tools and collaborated with a music college graduate to manually align the MIDI files with the audio and to annotate beats, structural boundaries, and F0 contours. These efforts culminated in the AIST Annotations, which complemented the RWC recordings and became widely used benchmarks for MIR evaluation ([Goto, 2006](#)). In subsequent years, the MIR community continued to enhance the dataset through additional contributions—providing, for example, high-resolution audio–MIDI alignments for the classical collection ([Ewert et al., 2009](#)) and chord annotations for popular music ([Cho and Bello, 2011](#)).

The RWC dataset had a profound impact on the MIR field, providing one of the first large-scale resources that combined professionally produced audio with rich, structured annotations. At a time when internet-based distribution was not yet feasible, the dataset was made available on physical CDs, ensuring global accessibility. Goto’s efforts not only laid the groundwork for modern MIR research practices but also inspired the development of new tools, methods, and user interfaces. The strong and immediate response from the international research community—especially following its presentation at ISMIR 2002—underscored the significance of RWC and encouraged Goto to continue contributing additional annotations and tools in support of open, reproducible music research. As shown in [Figure 2](#), the three core publications describing the RWC collections, along with the AIST Annotations, have been cited in 1,774 publications between 2003 and 2024.⁵

A selection of key publications that build upon and make extensive use of the RWC dataset is summarized in [Table 2](#). These works illustrate the dataset’s broad impact across the MIR field. First, the RWC database has been applied to a wide variety of core MIR tasks, demonstrating its versatility. Second, the publications span leading venues, including conferences such as ISMIR, ICASSP, and WASPAA, as well as prominent journals. Third, the

| Task @ Venue | Sub-Collections | Citation |
|--|-----------------|---------------------------------|
| Chorus-Section Detection @ ICASSP | P | (Goto, 2003) |
| Drum Sound Detection @ ISMIR | {P,I} | (Yoshii et al., 2004) |
| Musical Interfaces @ ISMIR | {P,R,C,J,G} | (Goto and Goto, 2005) |
| Singer Identification @ ISMIR | P | (Fujihara et al., 2005) |
| Instrument Identification @ ISMIR | {C,I} | (Kitahara et al., 2005) |
| Drum Loop Retrieval @ CBMI | P | (Gillet and Richard, 2005) |
| Audio Retrieval @ ISMIR | {P,R,C,J} | (Bertin and de Cheveigné, 2005) |
| Music Source Separation @ GRETSI | C | (Vincent and Gribonval, 2005) |
| Audio Melody Extraction @ WASPAA | G | (Ryynänen and Klapuri, 2005) |
| Automatic Mixing @ AXMEDIS | P | (Katayose et al., 2005) |
| Drum Sound Detection @ ICASSP | P | (Yoshii et al., 2006) |
| Lyrics Alignment @ ISM | P | (Fujihara et al., 2006) |
| Pitch Estimation @ ICASSP | P | (Fujihara et al., 2006) |
| Music Structure Analysis @ ISMIR | {P,J} | (Bruderer et al., 2006) |
| Musical Interfaces @ ISMIR | J | (Hamanaka, 2006) |
| Audio Coding @ TASLP | {C,J} | (Derrien et al., 2006) |
| Automatic Music Transcription @ ISMIR | P | (Ryynänen and Klapuri, 2006) |
| Genre Classification @ ISMIR | RWC | (Reed and Lee, 2006) |
| Music Structure Analysis @ AMCM | P | (Paulus and Klapuri, 2006) |
| Pitch Estimation @ TASLP | {C,J} | (Kameoka et al., 2007) |
| Singing Voice Retrieval @ ISMIR | P | (Fujihara and Goto, 2007) |
| Thumbnail Image Generation @ ISMIR | G | (Yoshii and Goto, 2008) |
| Singing Synthesis @ SMC | P | (Nakano and Goto, 2009) |
| Beat Tracking @ TASLP | {P,C,J} | (Grosche and Müller, 2011) |
| Singer Identification @ ISMIR | P | (Lagrange et al., 2012) |
| Music Source Separation @ ICML | P | (Yoshii et al., 2013) |
| Singing Voice Detection @ ISMIR | P | (Lehner et al., 2013) |
| Pitch Estimation (pYIN) @ ICASSP | P (synth.) | (Mauch and Dixon, 2014) |
| Music Mashup @ TASLP | P | (Davies et al., 2014) |
| Chord Estimation @ ISMIR | P | (Zhou and Lerch, 2015) |
| Beat Tracking @ ISMIR | P | (Böck et al., 2016) |
| Melody Harmonization @ ISMIR | P | (Tsushima et al., 2017) |
| Lyrics Transcription @ ICASSP | P | (Nishikimi et al., 2019) |
| Singing Voice Separation @ WASPAA | P | (Nakano et al., 2019) |
| Audio Declipping @ TASLP | {P,J,C} | (Gaultier et al., 2021) |
| Singing Voice Extraction @ Electronics | P | (Gao et al., 2021) |
| Lyrics Generation @ ISMIR | P | (Watanabe and Goto, 2023) |
| Music Generation @ IJCAI | P | (Lin et al., 2024) |

Table 2 Selection of key works building upon the RWC Music Database.

Popular Music collection (P) emerges as the most frequently used subset—likely due to its extensive set of available annotations. Finally, the selection reflects the progression of MIR research itself, from early signal processing techniques like melody extraction to source separation methods based on Non-Negative Matrix Factorization (NMF) and more recent applications of deep neural networks, such as for beat tracking. The continued use of RWC in current studies—including work on lyrics processing and music generation—underscores its lasting relevance within the community.

3 OPEN MUSIC DATA

The MIR field has made remarkable strides toward greater transparency—driven not only by broader trends in open-source software and open data, but also through its own community-led efforts to develop, share, and maintain tools and datasets. On the software side, Python libraries such as `librosa` (McFee et al., 2015) and `madmom` (Böck et al., 2016) have provided researchers with reliable, well-documented implementations that serve as foundations for a wide range of applications. Educational resources like the `fmp-notebooks` (Müller and Zalkow, 2019), with their extensive explanations and visualizations, have further lowered the entry barrier and deepened understanding across the community. Additionally, the `mir-eval` library filled a critical gap by offering standardized evaluation metrics to support reproducible experiments (Raffel et al., 2014).

Open datasets are equally essential. While open-source code supports reproducibility, true replicability—particularly in machine learning—requires access to the original data. Without it, researchers can only approximate prior results, which undermines both transparency and the reliability of scientific findings.

In its early years, the MIR field lagged behind other research domains in terms of openly accessible datasets. Copyright restrictions often prevented the distribution of audio recordings, posing a major challenge for tasks that require direct access to sound—such as transcription, recommendation, or music generation. In contrast, other fields benefited early on from public data repositories such as the UCI Machine Learning Repository,⁶ which supported benchmarking and empirical research from the outset. Offering hundreds of datasets across diverse domains—including biology, finance, healthcare, social science, and games—the UCI repository provided researchers with immediate access to structured, ready-to-use data for algorithm development and evaluation.

The release of the RWC Music Database marked a pivotal moment for the MIR community. Although not fully open—since the audio was not in the public domain and had to be purchased on physical media—it was licensed

for research use and remained easily accessible, making it a widely adopted benchmark for tasks such as beat tracking, structural segmentation, and chord recognition. This balance of legal clarity and practical usability enabled consistent evaluation and reproducibility, establishing RWC as a foundational resource. It also served as a model and inspiration for later datasets.

The Million Song Dataset, for example, supported large-scale recommendation research without providing audio, instead offering precomputed features and listening histories (Bertin-Mahieux et al., 2011). In contrast, MedleyDB included full audio access, focusing on multi-track recordings and melody annotations (Bittner et al., 2014), parts of which were later incorporated into MUSDB18 for source separation (Rafii et al., 2017). Other specialized datasets soon followed, including MAPS and MAESTRO, which support research in piano transcription (Emiya et al., 2010; Hawthorne et al., 2019). Genre-specific collections such as the Wagner Ring Dataset (Weiß et al., 2023) and Schubert’s Winterreise (Weiß et al., 2021) either provide links to commercial CD recordings or include carefully curated older recordings for which copyright has expired, making them freely available for research purposes.

Several of these datasets gained prominence through open challenges, where research teams submitted and evaluated their systems on shared tasks. These challenges often culminated in workshops—either as standalone events or integrated into larger conferences such as ISMIR. Within the MIR community, MIREX (as previously mentioned) has played a central role by offering standardized benchmarking tasks, including melody extraction and beat tracking (Downie, 2008). Following a period of reduced activity, MIREX was recently revitalized with updated tasks, such as music generation. Other focused competitions—like the Sound/Music Demixing Challenge—have also contributed significantly to progress, particularly in music source separation (Fabbro et al., 2024).

4 RWC DATASET: AUDIO

Audio recordings are the core component of the RWC Music Database. As part of this work, we announce the official release of the original RWC audio material under a Creative Commons license (CC BY-NC 4.0).⁷⁸ All recordings are now freely available for research purposes and hosted on Zenodo.⁹ Crucially, the distributed files stem from the original master tracks used in the CD production—not from consumer-ripped copies. This distinction is essential for ensuring high fidelity and timing accuracy. Ripping software may introduce inconsistencies such as compression artifacts and time shifts, which can compromise tasks involving synchronization, alignment, or precise timing analysis. By providing the original

| ID | Sub-Collection | #Pieces | #CDs | #Tracks | Dur. |
|----------|--------------------|---------|------|---------|----------|
| RWC-P | Popular Music | 100 | 7 | 100 | 6:43:36 |
| RWC-R | Royalty-Free Music | 15 | 1 | 15 | 0:32:23 |
| RWC-C | Classical Music | 50 | 6 | 61 | 5:27:08 |
| RWC-J | Jazz Music | 50 | 4 | 50 | 3:42:20 |
| RWC-G | Genre Mix | 100 | 9 | 102 | 6:58:26 |
| Σ | | 315 | 27 | 328 | 23:23:55 |

Table 3 Overview of the five sub-collections of the RWC Music Database.

Abbreviations: P = Pop, R = Royalty-Free, C = Classical, J = Jazz, G = Genre.

master-quality recordings, this release serves as an authoritative audio reference that enables reproducible MIR research and promotes consistent, high-quality dataset usage across the community.

4.1 STRUCTURE AND DESIGN

The RWC Music Database is organized into six sub-collections: Pop (P), Royalty-Free (R), Classical (C), Jazz (J), Genre (G), and Instruments (I). [Table 3](#) provides an overview of the first five sub-collections, which form the core focus of our current initiative. The sixth sub-collection, Instruments, consists of a range of isolated notes from 50 instruments. Since this sub-collection does not contain complete performances, we do not consider it in this work.⁹ Collectively, the first five sub-collections include 328 tracks, totaling approximately 23 hours of audio. The largest among them are Pop and Genre, with 100 and 102 tracks, respectively—each contributing nearly 7 hours of material. Classical and Jazz follow, with 61 and 50 tracks, while Royalty-Free includes 15 tracks totaling around 30 minutes. The discrepancy between the number of tracks and the number of pieces arises from defining a ‘piece’ as a distinct musical work. This distinction is particularly relevant in the Classical collection, where a single work—such as a symphony—may span multiple tracks corresponding to its individual movements. These special cases are covered by our naming convention: Each file begins with a prefix indicating its sub-collection, followed by a three-digit number corresponding to the piece number (e.g., `RWC_P014`). If a piece consists of multiple parts, these are indicated by a capital letter (e.g., `RWC_C024A`, `RWC_C024B`, and `RWC_C024C`). Audio files, MIDI files, and extra annotation files, such as lyrics, follow these patterns to allow for easy and structured access to corresponding files.

All tracks in the RWC Music Database were specifically performed and recorded for inclusion in the collection. To avoid copyright restrictions, the dataset includes only

public domain works or newly commissioned compositions created exclusively for the RWC project. The RWC-P subset includes entirely original works, featuring contributions from 25 composers, 30 lyricists, and 23 arrangers. In RWC-G, 73 pieces were newly composed, while RWC-J contains 46 original compositions by 4 composers and 1 lyricist.

The RWC sub-collections were curated with distinct design objectives that reflect their respective musical domains. RWC-P aims to capture a range of mainstream pop styles. Of its 100 pieces, 80 are sung in Japanese and emulate the sound of 1990s Japanese pop charts, while the remaining 20 are in English, inspired by American pop music from the 1980s. In contrast with RWC-P, which consists of 100 newly composed songs, RWC-R includes 15 well-known public-domain songs, including 10 traditional songs with English lyrics and 5 Japanese children’s songs. The RWC-C collection was designed to offer variety across instrumentation, historical periods, compositional styles, and composers. RWC-J is structured into three parts: the first features five jazz pieces, each recorded with seven different instrumentations (for a total of 35 tracks); the second includes nine pieces illustrating stylistic diversity, such as modal and free jazz; and the third presents six fusion-style tracks. Finally, RWC-G spans 10 musical genres (popular, rock, dance, jazz, Latin, classical, marches, world, vocals, and traditional Japanese music), further divided in 33 sub-genres,¹⁰ each represented by at least three pieces, with one additional a cappella piece rounding up the collection.

4.2 STATISTICAL OVERVIEW

[Figure 3](#) shows the distribution of track durations across the five RWC sub-collections. RWC-P tracks have a mean duration of 4:02 (mm:ss), ranging from 00:52 to 10:08. RWC-J exhibits a slightly higher average of 4:26, with durations between 2:23 and 7:39. RWC-G tracks average 4:06, with a similar range to RWC-P tracks. RWC-R contains the shortest tracks overall, with a mean of 2:09 and durations from 1:46 to 2:56. RWC-C spans the widest range, from 00:50 to 17:59, and shares the highest average duration (4:26) with RWC-J. To support vocal-specific research tasks such as melody extraction, singing voice detection, source separation, or lyrics alignment, [Figure 3](#) highlights the tracks containing singing. Vocal content varies considerably across the RWC sub-collections: all 100 tracks in RWC-P include vocals, compared to 52 in RWC-G, 15 in RWC-R, 6 in RWC-C, and only 2 in RWC-J. These differences reflect the distinct stylistic and functional goals that guided the design of each collection.

As a final statistic, [Figure 4](#) presents the tempo distribution for the RWC-P collection. Tempos range from 62 to 200 beats per minute (BPM), with an average of approximately 110 BPM. The distribution reveals three prominent

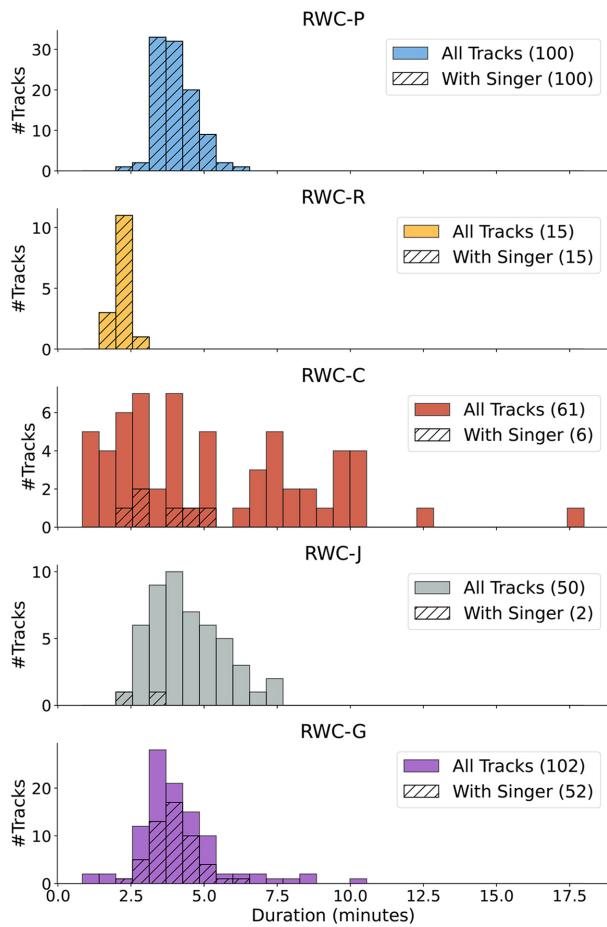


Figure 3 Track duration distributions for the five RWC sub-collections. Tracks containing singing voice are highlighted with black overlays.

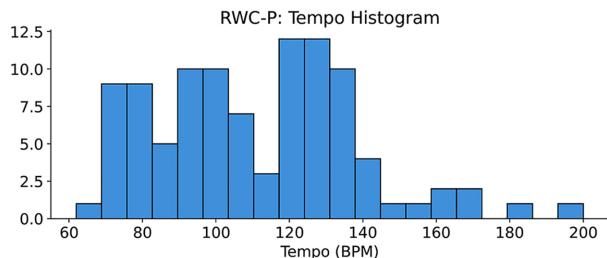


Figure 4 Tempo distribution of the RWC Popular Music sub-collection, showing the tempo range in beats per minute (BPM) across all tracks.

clusters centered around 80 BPM, 100 BPM, and 130 BPM. While the other sub-collections do not include explicit tempo annotations in their metadata, approximate tempo estimates can be derived from the beat annotations available in the AIST annotation set.

5 RWC DATASET: MIDI

From its inception, the RWC Music Database was designed to include both audio and symbolic representations, with each track complemented by a corresponding MIDI file. These MIDI files were not created during the original

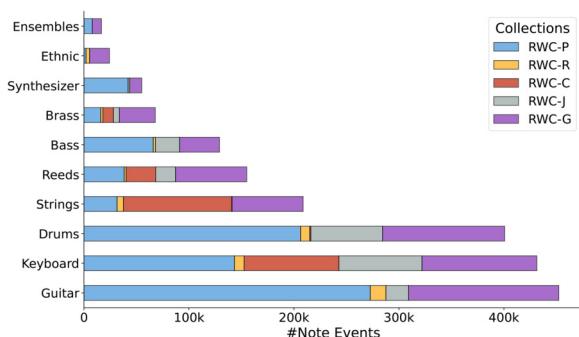


Figure 5 Number of note events per collection and instrument group (top 10).

recording sessions but were transcribed afterwards by professionals from Japan’s karaoke industry, following the standard practices of the time. While this approach ensured broad symbolic coverage, it also led to temporal misalignments between audio and MIDI, limiting their suitability for tasks requiring precise timing.

To address this limitation, Masataka Goto later released the AIST Annotations, which include a set of manually aligned MIDI files created using dedicated editing tools.¹¹ While the alignment process was labor-intensive, it substantially improved temporal accuracy and enhanced the dataset’s usefulness for a broad range of audio-symbolic MIR tasks, such as score-informed source separation, automatic transcription, and performance analysis. Beyond these applications, the MIDI data also enables statistical analysis of instrument usage across the dataset. For example, Figure 5 shows the number of note events for the 10 most common instrument groups, based on the instrument labels provided in the MIDI files.

It has long been known that alignment issues persisted—particularly in the Classical sub-collection (RWC-C), where audio and MIDI often exhibit significant temporal discrepancies. Moreover, our recent inspection revealed that similar misalignment problems exist across other sub-collections as well. In response, we initiated a systematic effort to refine and realign all MIDI files in the RWC dataset, combining automated synchronization algorithms with careful manual verification to enhance the correspondence between audio and symbolic timelines.

In the following subsections, we detail this alignment initiative. Beyond improving the dataset’s accuracy and utility for downstream MIR applications, our work reflects a broader vision: establishing a transparent, community-driven corpus that supports collaborative contributions—such as corrections, validations, and extended annotations—through reproducible workflows and open infrastructure.

As part of this re-release, we provide both the original MIDI files from AIST and the temporally adjusted versions. Together, they constitute a core component of the

RWC dataset and offer a solid foundation for future annotation and research efforts.

5.1 ALIGNMENT ISSUES

During our inspection of the RWC Music Database, we identified three main types of alignment issues between the audio recordings and their corresponding MIDI files, which we describe below in order of increasing complexity.

The first issue involves a global offset, where the MIDI file is uniformly shifted in time relative to the audio. Such offsets can result from differences in audio sources (e.g., ripped vs. master files), decoding artifacts (e.g., MP3 compression), or inconsistencies introduced during processing. Although relatively easy to correct, even small offsets can compromise accuracy in time-sensitive applications.

The second issue arises when, in addition to a global offset, the overall performance speed of the audio and MIDI differs by a constant factor. That is, the playback tempo in the MIDI file is consistently slower or faster than in the audio, leading to an increasing temporal drift over time. While a perfect alignment corresponds to a time-scaling factor of 1.0, we observed deviations typically ranging from 0.97–1.03, resulting in misalignments of several seconds in longer recordings. We refer to this as linear time scaling (LTS) case, which necessitates correcting both the global offset and the constant tempo mismatch.

The third and most complex issue involves non-linear timing deviations between the audio and MIDI, typically caused by tempo fluctuations in expressive performances. Unlike a constant tempo mismatch, these deviations change over time and require a non-linear warping of the MIDI timeline to accurately align with the audio. We refer to this as non-linear time warping (NTW) case. Such timing irregularities are particularly common in the Classical collection (RWC-C) and in Western classical pieces within the Genre collection (RWC-G), though they may also occur in other sub-collections.

5.2 ALIGNMENT CORRECTION

To address the three types of alignment issues, we first manually annotated the onset time of the first and the offset time of the last note event in each audio recording.¹² These anchor points served as stable temporal references and were used to reduce boundary artifacts during synchronization.

Based on these annotations, we computed an initial alignment between the audio and MIDI files using the `SyncToolbox` (Müller et al., 2021), a high-resolution synchronization framework tailored to Western music. It combines multiscale dynamic time warping (MsDTW), memory-restricted DTW, and onset-enhanced alignment strategies (Ewert et al., 2009; Prätzlich et al., 2016). By integrating chroma-based and onset-based features, this approach produces alignment paths that are both robust

and temporally precise. The manually annotated anchor points were used to guide and stabilize the alignment, particularly at the beginning and end of each track, where we found that automatic methods often introduce errors.

We used the resulting alignment paths to categorize each audio–MIDI pair into either the LTS case or the NTW case. In the LTS scenario, a perfect alignment between the audio and MIDI should follow a linear function characterized by an offset and a constant slope. Based on this observation, we applied least-squares fitting to approximate each alignment path with a linear function. To reduce the influence of boundary artifacts, we excluded the first and last 10% of the alignment path from the fitting process. Approximations with a mean absolute error below a predefined threshold (Mean Square Error (MSE) empirically set to 0.01) were classified as LTS cases, while those exceeding the threshold were considered NTW cases.

For LTS cases, we adjusted the MIDI timeline using the linear fit, which involved first extrapolating the global offset, then applying a uniform time scaling based on the fitted slope. For NTW cases, we used the initial alignment path obtained from the `SyncToolbox` to non-linearly warp the MIDI file. In practice, applying the linear adjustment in LTS cases proved more robust and accurate than relying on the more general NTW strategy, which is better suited for performances with expressive tempo fluctuations.

Figure 6a illustrates a representative LTS case, where the alignment path closely follows the linear fit. The offset at the beginning reflects a delayed initial MIDI onset, while the slope below one suggests that the audio performance was slightly faster than the MIDI. Figure 6b shows an NTW case, where the initial alignment deviates substantially from the linear fit, reflecting local tempo fluctuations that cannot be captured by a global scaling. Finally, Table 4 summarizes the number of MIDI files classified as LTS or NTW cases according to our heuristic criteria. For example, in the RWC-P subset, 94 MIDI files are identified as LTS cases, of which 65 required almost no adjustment and 29 involved a time stretch. In contrast, all 61 MIDI files in the RWC-C subset are NTW cases.

5.3 ALIGNMENT VERIFICATION

To verify the accuracy of the MIDI alignments, we employed sonification using the `libsoni` toolbox (Özer et al., 2024). MIDI events were rendered and played alongside the corresponding audio, enabling direct auditory comparison. Verification of all tracks through listening was carried out by the authors of this article, with particular attention given to the beginning and ending passages of each recording.¹³

Overall, the alignment corrections yielded reliable results, especially for tracks classified as LTS cases. However, for a number of tracks—particularly those involving soft-onset instruments such as strings—the alignment produced by the `SyncToolbox` proved less accurate,

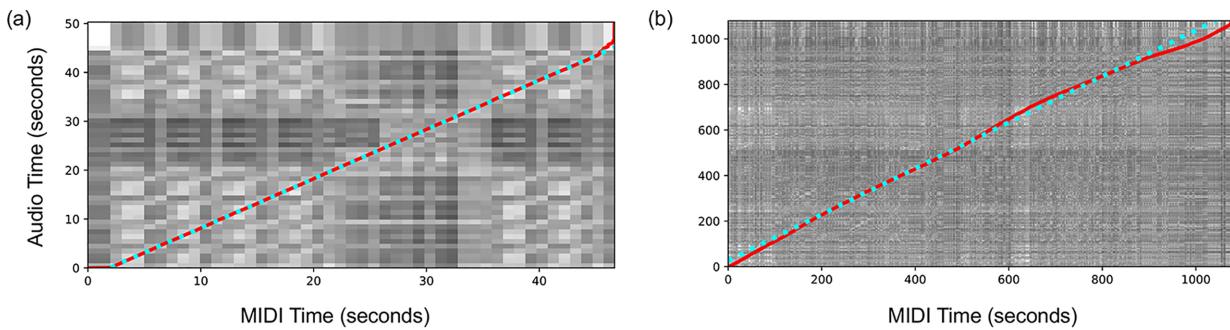


Figure 6 Two audio-MIDI alignment examples illustrating linear time scaling (LTS) and non-linear time warping (NTW) cases. The background shows the chroma-based cost matrix; the red line indicates the alignment path computed with the SyncToolbox (Müller et al., 2021), while the dotted cyan line shows the corresponding linear fit. **(a)** LTS case: Musette in D major by Bach, performed on harpsichord (RWC_C024C). **(b)** NTW case: Prelude and Liebestod from Wagner's Tristan und Isolde, performed by a symphony orchestra (RWC_C009).

| ID | #Tracks | LTS | | NTW |
|-------|---------|---------|---------|-----|
| | | ≤ 50 ms | > 50 ms | |
| RWC-P | 100 | 65 | 29 | 6 |
| RWC-R | 15 | 13 | 2 | – |
| RWC-C | 61 | – | – | 61 |
| RWC-J | 50 | 6 | 5 | 39 |
| RWC-G | 102 | 27 | 16 | 59 |

Table 4 Classification of MIDI files into LTS (Linear Time Scaling) and NTW (Non-linear Time Warping) cases based on heuristic criteria, summarizing the number of files per category. The two LTS columns indicate whether applying a time-scaling factor of 1.0 results in an accumulated alignment error of ≤ 50 ms (indicating tempo agreement between audio and MIDI) or > 50 ms (indicating a tempo mismatch requiring time scaling to adapt the MIDI). The > 50 ms threshold reflects a typical tolerance used in time-critical tasks such as onset detection.

e.g., solo cello recording in RWC_C041. In such cases, we manually annotated additional anchor points between audio and MIDI to stabilize and improve the alignment. Human intervention was especially valuable in performances featuring expressive timing variations or ambiguous onset cues.

Despite our best efforts, some limitations persist. In particular, tracks from the Classical (RWC-C) and Genre (RWC-G) collections may still contain transcription inaccuracies or musically ambiguous passages. In contrast, the Popular Music collection (RWC-P) generally offers the highest synchronization quality, and the Royalty-Free collection (RWC-R) also proved to be highly reliable.

While the listening-based verification was carried out with care, it remains informal and inherently subjective. In the process, various note-level discrepancies between the MIDI and corresponding audio were identified, including missing, incorrect, or additional notes. Many of these issues stem from limitations in the original MIDI annotations or transcription artifacts. Correcting them requires meticulous manual effort and lies beyond the scope of

this initiative. A more systematic and quantitative evaluation will be necessary in future work, ideally tailored to the specific needs and constraints of individual application contexts.

To facilitate ongoing refinement, we encourage community contributions such as corrections, improvements, and extended annotations. These can be integrated into the versioning infrastructure described in the next section. For instance, the recording RWC_P009 begins with a drum pick-up beat that contains transcription errors in the MIDI—an issue that could be addressed through community engagement. Further discussion of the community's role follows below.

6 TOWARDS COMMUNITY-DRIVEN CURATION

Early MIR datasets were typically developed by individual labs or through institutional funding. Today, improved tooling and infrastructure—such as version-controlled repositories, standardized annotation formats, and open contribution models—make collaborative, community-driven maintenance feasible. Instead of static releases, modern datasets can evolve with input from a broad research community. In this context, the concept of an 'RWC 2.0' has emerged: a unified effort to consolidate disparate RWC versions, improve alignment between symbolic and audio content, and standardize annotations—all under a clear, permissive license that supports sustainable community involvement.

Community-maintained datasets offer both opportunities and challenges. Open collaboration broadens the contributor base, improving data quality, enabling new annotations, and fostering tool development across the ecosystem. However, such distributed efforts require careful coordination: data curation must follow shared standards, and all changes should be versioned and reviewed to prevent inconsistencies or regressions. Projects like `mirdata` already support this approach by offering a unified interface to many datasets, along with

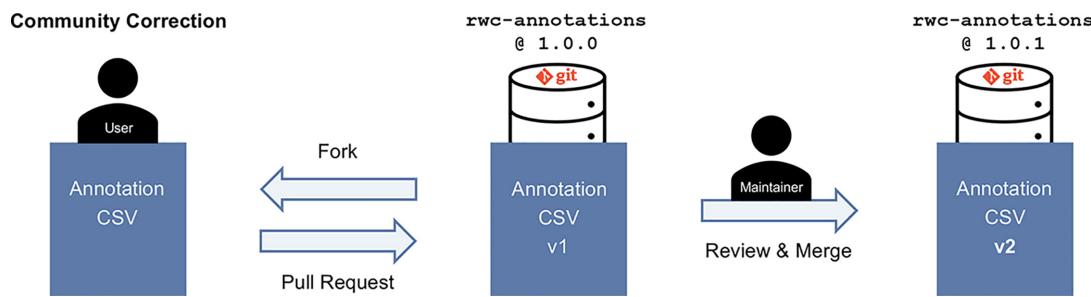


Figure 7 Illustration of the workflow for incorporating community-provided corrections into the centralized annotation repository.

tools for download management and data validation (Bittner et al., 2019).

Inspired by software engineering and open-source development, the MIR field is now well-equipped for robust, transparent, and community-driven data curation. At the core of the updated RWC dataset are the original master recordings, which were used for the original compact disc production and are now hosted on Zenodo within a dedicated community collection.⁹ To preserve historical context and ensure continued access, a copy of the original RWC website is archived alongside the audio. Complementing this, a GitHub-based infrastructure manages annotations, including MIDI files, beat and melody annotations, and other metadata.¹⁴ This setup supports long-term storage, version control, collaborative contributions, and transparent change tracking, aligning the dataset with best practices in open and sustainable research (McFee et al., 2019).

The GitHub infrastructure, serving as the main hub for community contributions, begins with the `rwc-annotations-archive` repository.¹⁵ This repository mirrors annotation files from the original RWC website, including MIDI and other data. However, these legacy materials vary in format and are often poorly documented, making them difficult to interpret and reuse. New community contributions, such as time-aligned lyrics, are first added to the `rwc-annotations-archive`, which acts as a central collection point for both original and newly created resources.¹⁶

A second repository, `rwc-annotations`, hosts cleaned and standardized versions of the annotations.¹⁴ Here, files from the `rwc-annotations-archive` are converted into a consistent, well-documented format to support long-term accessibility and usability. For pragmatic reasons—and with awareness of potential limitations—we adopt a simple comma-separated values (CSV) format, chosen for its human-readability, ease of parsing, and broad compatibility with text editors and spreadsheet tools. The structure and semantics of each CSV file are carefully documented to ensure sustainability and transparency.

The conversion from the `rwc-annotations-archive` to the cleaned annotations is intended as a one-time process. Dedicated scripts transform the data into the standardized format before inclusion in

`rwc-annotations`. Once migrated, backward compatibility with the archive is no longer maintained—encouraging future work to build on the cleaned format. This workflow balances openness to community contributions with the need for a consistent, well-documented annotation set.

Community-suggested corrections to the annotations are tracked in the `rwc-annotations` repository via GitHub features such as issues and pull requests. Figure 7 illustrates this workflow: a user forks the repository, applies changes, and submits a pull request, which is then reviewed by maintainers. Revisions may be requested before approval and merging. Drawing on experience from building the ChoraleBricks dataset (Balke et al., 2025), continuous integration tests help prevent format errors (e.g., invalid values) and annotation mistakes (e.g., duplicate fundamental frequency entries). Following software development best practices, a textual changelog summarizes all changes. To support reproducibility and citation, we recommend regular versioned releases, using a semantic versioning scheme.¹⁷

Beyond infrastructure and versioning, the sustainability of a dataset ultimately depends on active community participation. It is this collective effort that keeps the resource relevant, accurate, and broadly useful. Engagement fosters shared ownership, encourages diverse contributions, and helps uncover issues that might otherwise go unnoticed. A living dataset relies not just on well-designed systems but also on a committed, collaborative user base that continuously uses, critiques, and improves it.

7 CONCLUSIONS

The re-release of the RWC Music Database under an open license marks a significant milestone in the evolution of community-driven research infrastructure in MIR. By making the original master-quality audio freely available and establishing a modern, version-controlled annotation ecosystem, we aim to transform RWC from a static historical resource into a living, community-driven MIR corpus.

Our efforts not only preserve the legacy and continued relevance of the RWC dataset but also serve as a practical case study in revitalizing research assets through collaborative, open-source principles. The GitHub-based

infrastructure supports transparent development workflows, facilitates community contributions, and enables reproducible research by linking data, code, and changelogs in a cohesive ecosystem. The decision to separate archival data from cleaned, standardized annotations helps to ensure long-term maintainability without sacrificing openness to experimentation and iteration.

More broadly, this initiative reflects a growing maturity in the MIR field—one where open data, shared infrastructure, and sustained community engagement are increasingly recognized as essential pillars of scientific progress. It also underscores the importance of historical awareness: by revisiting and improving upon foundational datasets like RWC, we not only enhance current research practices but also honor the work of those who laid the groundwork for the field. Whether the RWC dataset will ever be truly complete remains an open question. Initial ideas for an ‘Everything Corpus’ were outlined by [Gotham et al. \(2025\)](#). Even without a formal evaluation, the original RWC release already followed many of the principles described there. With the re-release, especially data availability will be further improved and made much more accessible for future generations of researchers.

Looking ahead, we invite the MIR community to actively shape the continued development of the RWC dataset. Our initiative marks only the beginning. Inconsistencies in the MIDI alignments have already been identified and corrected, suggesting that further refinements remain—an opportunity for community engagement. Whether by proposing corrections, contributing new annotations, improving tools, or exploring novel use cases, researchers and developers can help ensure that RWC evolves into a robust, versatile, and inclusive resource for years to come.

ACKNOWLEDGEMENTS

The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS. Part of this work was initiated through conversations during the Dagstuhl Seminar 24302 ‘Learning with Music Signals: Technology Meets Education.’

FUNDING INFORMATION

SB and MM were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant number 500643750 (MU 2686/15-1).

COMPETING INTERESTS

The authors have no competing interests to declare. MM is member of the editorial team and have recused himself from any editorial involvement with this article.

AUTHORS’ CONTRIBUTIONS

The re-release of the dataset was led by SB and MM, encompassing conception, coordination, execution, and documentation. MG contributed the original data and provided valuable historical context. TN supported its long-term dissemination. The refinement of the audio-MIDI alignment was carried out by VAM, JZ, and SB. BM provided insights into open-source and open-data best practices, as well as insights into community-led initiatives. All co-authors actively participated in the preparation of the final manuscript.

DATA ACCESSIBILITY

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

- **Dataset:** The RWC dataset is available under a CC BY-NC 4.0 license on Zenodo: <https://zenodo.org/communities/rwc-music/>.
- **Annotations:** Structured and cleaned annotations are available on GitHub: <https://github.com/rwc-music/rwc-annotations>.

NOTES

1. A detailed description is given in [Section 4](#).
2. MIREX (Music Information Retrieval Evaluation eXchange) is an annual community-driven benchmarking initiative that evaluates state-of-the-art algorithms across a wide range of MIR tasks.
3. <https://audiolabs-erlangen.de/resources/MIR/SyncRWC60/>, accessed May 27, 2025.
4. <https://github.com/tmc323/Chord-Annotations>, accessed May 27, 2025.
5. Sources: https://scholar.google.com/citations?view_op=view_citation&user=BkZggZkAAAAJ&citation_for_view=BkZggZkAAAAJ:u5HHmVD_uO8C, https://scholar.google.com/citations?view_op=view_citation&user=BkZggZkAAAAJ&citation_for_view=BkZggZkAAAAJ:d1gkVwhDp0C, https://scholar.google.com/citations?view_op=view_citation&user=BkZggZkAAAAJ&citation_for_view=BkZggZkAAAAJ:LkGwnXOMwfC, accessed February 10, 2025.
6. <https://archive.ics.uci.edu/>, accessed May 27, 2025.
7. <https://creativecommons.org/licenses/by-nc/4.0/>.
8. The choice of the ‘CC BY-NC 4.0’ license best aligns with the original licensing approach and most accurately reflects the intended purpose of the RWC dataset—as a resource for research, including use within corporate research settings. More detailed information is available in the corresponding Zenodo data repository.
9. The data are available online at <https://zenodo.org/communities/rwc-music/>, accessed October 2025.
10. For details, we kindly refer to the original publication ([Goto et al., 2003](#)).
11. We kindly refer the reader to [Table 1](#) in [Goto \(2006\)](#) for an overview on the included annotations.
12. Note that we manually checked the offset times. However, this point in time can only be approximated.
13. During verification, we found that the MIDI files for `RWC_G025` and `RWC_G026` were accidentally swapped in the original RWC Music Database and need to be corrected.
14. <https://github.com/rwc-music/rwc-annotations>, accessed October 2025.

15. <https://github.com/rwc-music/rwc-annotations-archive>, accessed October 2025.
16. Keeping the original annotations in a separate repository is based on practical experience. As the archive may grow in size, storing these files separately helps keep the cleaned annotations lightweight and maintain a small memory footprint. However, changing this design is up to the community.
17. <https://semver.org/>, accessed May 2025.

AUTHOR AFFILIATIONS

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

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

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

Brian McFee  <https://orcid.org/0000-0001-6261-9747>
Music and Audio Research Laboratory, New York University, New York, USA

Tomoyasu Nakano  <https://orcid.org/0000-0001-8014-2209>
National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

Masataka Goto  <https://orcid.org/0000-0003-1167-0977>
National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

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

REFERENCES

Balke, S., Berndt, A., and Müller, M. (2025). ChoraleBricks: A modular multitrack dataset for wind music research. *Transaction of the International Society for Music Information Retrieval (TISMIR)*, 8(1), 39–54.

Bertin, N., and de Cheveigné, A. (2005). Scalable metadata and quick retrieval of audio signals. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, UK (pp. 238–244).

Bertin-Mahieux, T., Ellis, D. P. W., Whitman, B., and Lamere, P. (2011). The million song dataset. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* Miami, Florida, USA (pp. 591–596).

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., Korzeniowski, F., Schlüter, J., Krebs, F., and Widmer, G. (2016). madmom: A new Python audio and music signal processing library. In *Proceedings of the ACM International Conference on Multimedia (ACM-MM)*, Amsterdam, The Netherlands (pp. 1174–1178).

Böck, S., Krebs, F., and Widmer, G. (2016). Joint beat and downbeat tracking with recurrent neural networks. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, New York City, New York, USA (pp. 255–261).

Bruderer, M. J., McKinney, M., and Kohlrausch, A. (2006). Structural boundary perception in popular music. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada (pp. 198–201).

Cho, T., and Bello, J. P. (2011). A feature smoothing method for chord recognition using recurrence plots. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, Florida, USA (pp. 651–656).

Davies, M. E. P., Hamel, P., Yoshii, K., and Goto, M. (2014). AutoMashUpper: Automatic creation of multi-song music mashups. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(12), 1726–1737.

Derrien, O., Duhamel, P., Charbit, M., and Richard, G. (2006). A new quantization optimization algorithm for the mpeg advanced audio coder using a statistical subband model of the quantization noise. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 14(4), 1328–1339.

Dessein, A., Cont, A., and Lemaitre, G. (2010). Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, The Netherlands (pp. 489–494).

Downie, J. S. (2008). The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research. *Acoustical Science and Technology*, 29(4), 247–255.

Emiya, V., Bertin, N., David, B., and Badeau, R. (2010). MAPS - A piano database for multipitch estimation and automatic transcription of music. Research report.

Ewert, S., Müller, M., and Grosche, P. (2009). High resolution audio synchronization using chroma onset features. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan (pp. 1869–1872).

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.

Fujihara, H., and Goto, M. (2007). A music information retrieval system based on singing voice timbre. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Vienna, Austria (pp. 467–470).

Fujihara, H., Goto, M., Ogata, J., Komatani, K., Ogata, T., and Okuno, H. G. (2006). Automatic synchronization between lyrics and music cd recordings based on viterbi alignment of segregated vocal signals. In *IEEE International Symposium on Multimedia (ISM)*, Los Alamitos, California, USA (pp. 257–264).

Fujihara, H., Kitahara, T., Goto, M., Komatani, K., Ogata, T., and Okuno, H. G. (2005). Singer identification based on accompaniment sound reduction and reliable frame selection. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, UK (pp. 329–336).

Fujihara, H., Kitahara, T., Goto, M., Komatani, K., Ogata, T., and Okuno, H. G. (2006). F0 estimation method for singing voice in polyphonic audio signal based on statistical vocal model and viterbi search. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France (pp. 253–256).

Gao, Y., Zhang, X., and Li, W. (2021). Vocal melody extraction via hrnet-based singing voice separation and encoder-decoder-based f0 estimation. *Electronics*, 10(3), 1–14.

Gaultier, C., Kitić, S., Gribonval, R., and Bertin, N. (2021). Sparsity-based audio declipping methods: Selected overview, new algorithms, and large-scale evaluation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29, 1174–1187.

Gillet, O., and Richard, G. (2005). Indexing and querying drum loops databases. In *Proceedings of the International Workshop on Content-Based Multimedia Indexing (CBMI)*.

Gotham, M., Bemman, B., and Vatolkin, I. (2025). Towards an ‘everything corpus’: A framework and guidelines for the curation of more comprehensive multimodal music data. *Transactions of the International Society for Music Information Retrieval*, 8(1), 70–92.

Goto, M. (2003). A chorus-section detecting method for musical audio signals. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Hong Kong, China (pp. 437–440).

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

Goto, M. (2006). AIST Annotation for the RWC Music Database. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada (pp. 359–360).

Goto, M., and Goto, T. (2005). Musicream: New music playback interface for streaming, sticking, sorting, and recalling musical pieces. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, UK (pp. 404–411).

Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R. (2002). RWC Music Database: Popular, classical and jazz music databases. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Paris, France (pp. 287–288).

Goto, M., Hashiguchi, H., Nishimura, T., and Oka, R. (2003). RWC Music Database: Music genre database and musical instrument sound database. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Baltimore, Maryland, USA (pp. 229–230).

Grosche, P., and Müller, M. (2011). Extracting predominant local pulse information from music recordings. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(6), 1688–1701.

Hamanaka, M. (2006). Music scope headphones: Natural user interface for selection of music. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada (pp. 302–307).

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.

Kameoka, H., Nishimoto, T., and Sagayama, S. (2007). A multipitch analyzer based on harmonic temporal structured clustering. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3), 982–994.

Katayose, H., Yatsui, A., and Goto, M. (2005). A mix-down assistant interface with reuse of examples. In *Proceedings of the International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution (AXMEDIS)*, (pp. 1–8).

Kitahara, T., Goto, M., Komatani, K., Ogata, T., and Okuno, H. G. (2005). Instrument identification in polyphonic music: Feature weighting with mixed sounds, pitch-dependent timbre modeling, and use of musical context. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, London, UK (pp. 558–563).

Korzeniowski, F., and Widmer, G. (2016). Feature learning for chord recognition: The deep chroma extractor. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, New York City, New York, USA (pp. 37–43).

Lagrange, M., Ozerov, A., and Vincent, E. (2012). Robust singer identification in polyphonic music using melody enhancement and uncertainty-based learning. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)*, Porto, Portugal.

Lehner, B., Sonnleitner, R., and Widmer, G. (2013). Towards light-weight, real-time-capable singing voice detection. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Curitiba, Brazil (pp. 53–58).

Lin, L., Xia, G., Zhang, Y., and Jiang, J. (2024). Arrange, inpaint, and refine: Steerable long-term music audio generation and editing via content-based controls. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 7690–7698).

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).

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.

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (2015). Librosa: Audio and music signal analysis in Python. In *Proceedings the Python Science Conference*, Austin, Texas, USA (pp. 18–25).

Müller, M., Balke, S., and Goto, M. (2025). The story behind the RWC Music Database: An interview with Masataka Goto. *Transaction of the International Society for Music Information Retrieval (TISMIR)*, 8(1), 1–10.

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.

Müller, M., and Zalkow, F. (2019). FMP Notebooks: Educational material for teaching and learning fundamentals of music processing. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, The Netherlands (pp. 573–580).

Nakano, T., and Goto, M. (2009). VocaListener: A singing-to-singing synthesis system based on iterative parameter estimation. In *Proceedings of the 6th Sound and Music Computing Conference (SMC)* (pp. 343–348).

Nakano, T., Yoshiii, K., Wu, Y., Nishikimi, R., Lin, K. W. E., and Goto, M. (2019). Joint singing pitch estimation and voice separation based on a neural harmonic structure renderer. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (pp. 160–164).

Nishikimi, R., Nakamura, E., Fukayama, S., Goto, M., and Yoshiii, K. (2019). Automatic singing transcription based on encoder-decoder recurrent neural networks with a weakly-supervised attention mechanism. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brighton, UK (pp. 161–165).

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), 06524.

Paulus, J., and Klapuri, A. P. (2006). Music structure analysis by finding repeated parts. In *Proceedings of the 1st ACM Audio and Music Computing Multimedia Workshop*, Santa Barbara, California, USA (pp. 59–68).

Paulus, J., and Klapuri, A. P. (2009). Music structure analysis using a probabilistic fitness measure and a greedy search algorithm. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6), 1159–1170.

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).

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., Mimirakis, S. I., and Bittner, R. (2017). The MUSDB18 corpus for music separation.

Reed, J., and Lee, C. (2006). A study on music genre classification based on universal acoustic models. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada (pp. 89–94).

Ryynänen, M., and Klapuri, A. (2005). Polyphonic music transcription using note event modeling. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA (pp. 319–322).

Ryynänen, M., and Klapuri, A. (2006). Transcription of the singing melody in polyphonic music. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada (pp. 222–227).

Tsushima, H., Nakamura, E., Itoyama, K., and Yoshiii, K. (2017). Function-and rhythm-aware melody harmonization based on tree-structured parsing and split-merge sampling of chord sequences. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China (pp. 502–508).

Vincent, E., and Gribonval, R. (2005). Construction d'estimateurs oracles pour la séparation de sources. In *XXe colloque GRETSI (traitement du signal et des images)*, Louvain-la-Neuve, Belgium.

Wang, J.-C., Hung, Y.-N., and Smith, J. B. L. (2022). To catch a chorus, verse, intro, or anything else: Analyzing a song with structural functions. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 416–420).

Watanabe, K., and Goto, M. (2023). Text-to-lyrics generation with image-based semantics and reduced risk of plagiarism. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy (pp. 398–406).

Weiβ, 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. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 6(1), 135–149.

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

Yoshii, K., and Goto, M. (2008). Music Thumbnailer: Visualizing musical pieces in thumbnail images based on acoustic features. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Philadelphia, Pennsylvania, USA (pp. 211–216).

Yoshii, K., Goto, M., and Okuno, H. G. (2004). Automatic drum sound description for real-world music using template adaptation and matching methods. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain (pp. 184–191).

Yoshii, K., Komatani, K., Ogata, T., Okuno, H., and Goto, M. (2006). An error correction framework based on drum pattern periodicity for improving drum sound detection. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toulouse, France (pp. 237–240).

Yoshii, K., Tomioka, R., Mochihashi, D., and Goto, M. (2013). Infinite positive semidefinite tensor factorization for source separation of mixture signals. In *International conference on machine learning (ICML)* (pp. 576–584).

Zhou, X., and Lerch, A. (2015). Chord detection using deep learning. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Málaga, Spain (pp. 52–58).

TO CITE THIS ARTICLE:

Balke, S., Zeitler, J., Arifi-Müller, V., McFee, B., Nakano, T., Goto, M., & Müller, M. (2026). RWC Revisited: Towards a Community-Driven MIR Corpus. *Transactions of the International Society for Music Information Retrieval*, 9(1), 21–35. DOI: <https://doi.org/10.5334/tismir.326>

Submitted: 21 July 2025 **Accepted:** 10 November 2025 **Published:** 13 February 2026

COPYRIGHT:

© 2026 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.