



The Story Behind the RWC Music Database: An Interview with Masataka Goto

EDITORIAL

MEINARD MÜLLER 

STEFAN BALKE 

MASATAKA GOTO 

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

 jubiquity press

ABSTRACT

The Real World Computing (RWC) Music Database has been a widely used and valuable resource in music information retrieval (MIR) research for over two decades, offering high-quality audio recordings and comprehensive annotations for a variety of MIR tasks. In light of recent efforts to transform the RWC into an open-domain and community-driven dataset, this editorial presents a revised transcript of an interview with Masataka Goto, one of the original creators of the RWC. Conducted on February 18, 2025, by Stefan Balke and Meinard Müller, the interview explores the origins of the RWC, the motivations behind its development, the challenges involved in its production and annotation, and its long-term impact on MIR research. The conversation also reflects on the evolution of dataset practices in MIR and discusses future directions for open, collaborative dataset initiatives. This transcript has been carefully edited for clarity while preserving the authenticity of the questions and responses. Beyond its immediate relevance to the RWC, the interview provides a unique historical perspective on MIR, capturing firsthand insights from a key figure whose work has significantly shaped the field. By making this interview available, we aim to contribute to broader conversations on dataset sustainability, open access, and the role of community-driven research in MIR.

CORRESPONDING AUTHOR:

Meinard Müller

International Audio
Laboratories Erlangen,
Erlangen, Germany

meinard.mueller@audiolabs-erlangen.de

KEYWORDS:

Real World Computing, RWC,
Music Information Retrieval,
History of MIR, Sustainability,
Open Access

TO CITE THIS ARTICLE:

Müller, M., Balke, S., & Goto, M.
(2025). The Story Behind the
RWC Music Database:
An Interview with Masataka
Goto. *Transactions of the
International Society for Music
Information Retrieval*, 8(1), 154–
164.
DOI: <https://doi.org/10.5334/tismir.261>

1 INTRODUCTION

The Real World Computing (RWC) Music Database has been a fundamental resource in music information retrieval (MIR) research for more than two decades (Goto, 2004; Goto et al., 2002, 2003). This database provides high-quality audio recordings across multiple genres, including popular, classical, and jazz music (see Figure 1 for an overview). The RWC has been widely used in tasks such as music structure analysis (Goto, 2003a; Paulus and Klapuri, 2009; Wang et al., 2022), beat tracking (Böck et al., 2016; Cheng and Goto, 2023; Grosche and Müller, 2011), chord recognition (Cho and Bello, 2011; Korzeniowski and Widmer, 2016; Wu and Li, 2019), automatic music transcription (Dessein et al., 2010; Itoyama et al., 2006; Rynänen and Klapuri, 2005), and music synchronization (Ewert et al., 2009; Joder et al., 2011). In addition to its diverse audio collection, the RWC includes aligned Musical Instrument Digital Interface (MIDI) encodings and various complementary annotations, such as beat, structure, and chord annotations, making it an important benchmark for MIR research (Goto, 2006a).

Originally, the RWC audio material was distributed on physical media and offered for purchase at a nominal cost, while the annotations were made available separately through static websites. In response to the growing demand for open resources, efforts are underway to revisit and transform the dataset into a community-driven MIR corpus, ensuring its long-term accessibility and relevance. A key milestone in this transition is the planned release of the RWC dataset under a Creative Commons license, making it freely available for research purposes. This shift greatly enhances the dataset's usability, fostering broader adoption and collaboration within the MIR community.

Beyond making the dataset openly accessible, this transformation also involves addressing key aspects such as dataset maintenance and the expansion of annotations through collaborative efforts and ensuring reproducibility through platforms like GitHub (GitHub, Inc., San Francisco, CA, USA). Hosting the audio recordings on free and long-term archiving platforms such as Zenodo (European Organization for Nuclear Research, Geneva, Switzerland) will facilitate better data management while fostering a more open and transparent research environment. By shifting toward a community-driven approach, the RWC will not only preserve its historical significance but also encourage future research and collaboration within MIR.

As part of this initiative, we conducted an interview with Masataka Goto, the key figure who conceived the idea for the RWC database and led its creation, to gain deeper insights into its historical development and future perspectives.¹ His reflections provide valuable context not only on the evolution of the RWC but also on broader

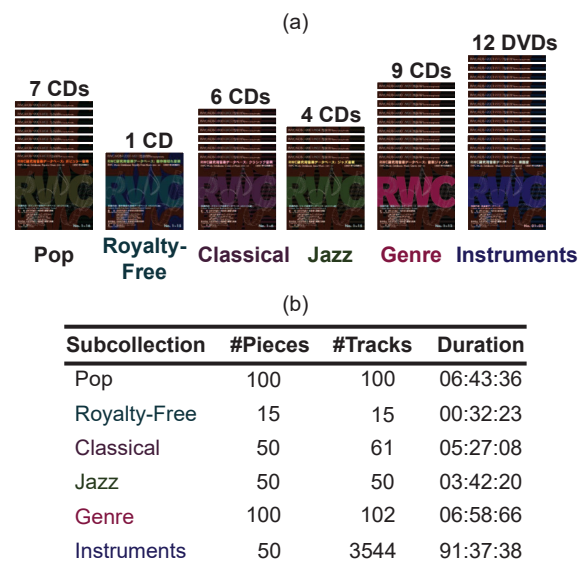


Figure 1 Overview of the RWC Music Database and its subcollections. **(a)** Number of discs per subcollection. **(b)** Summary of musical content regarding the number of pieces (or instruments), tracks, and duration (hh:mm:ss).

advancements in the field of MIR. Furthermore, the interview highlights the RWC's role in shaping MIR research while discussing the opportunities and challenges of transforming it into a truly open and community-driven resource.

The remainder of the editorial is structured as follows. Section 2 describes the interview methodology, outlining the preparation process, format, and key discussion topics. Section 3 presents a revised transcript of the interview with Masataka Goto, providing insights into the creation, impact, and future of the RWC dataset. Finally, Section 4 summarizes the key takeaways and discusses the broader implications for community-driven dataset development in MIR.

2 INTERVIEW METHODOLOGY

This section outlines the methodology of the interview conducted with Masataka Goto, including its setup, structure, and processing, as well as the key themes explored.

2.1 INTERVIEW SETUP AND FORMAT

The interview was conducted online using the video-conferencing software Zoom (Zoom Communications, San Jose, CA, USA). It took place on February 18, 2025, and lasted approximately two hours, including briefing and debriefing. The interview was carried out by Stefan Balke and Meinard Müller, following a semi-structured format designed to elicit both factual insights and personal reflections. The discussion adopted a question-and-answer approach embedded within a conversational tone, which encouraged Masataka Goto to share both personal anecdotes and technical perspectives. The interviewers took turns posing questions to maintain a



Figure 2 Screenshots from the interview with Masataka Goto (second and fourth in the first row), conducted by Stefan Balke (first in the first row) and Meinard Müller (third in the first row).

dynamic and balanced exchange, while the conversation was moderated to ensure comprehensive coverage of all key thematic areas.

Figure 2 shows a selection of screenshots from the interview session.

2.2 INTERVIEW STRUCTURE AND TOPICS

The interview followed a historical narrative, covering key aspects of the RWC Music Database in chronological order. The discussion was organized into the following thematic areas:

- **Personal Beginnings:** Masataka Goto's early experiences and his introduction to the field of music processing.
- **Early Days of MIR:** Insights into the early development phase of MIR research, highlighting key developments and milestones.
- **Idea Behind RWC:** The inspiration for the RWC Music Database, the challenges faced, and the development process.
- **Production and Annotation:** Overview of the dataset's production, including audio recordings and annotation creation.
- **Impact and Applications:** The role of RWC in advancing MIR research, its influence on various research domains, and its widespread adoption.
- **Lessons Learned and Future Directions:** Reflections on key challenges, achievements, and lessons from the RWC's journey, along with perspectives on its sustainability and the future of open, community-driven datasets in MIR.

2.3 USAGE AND PROCESSING OF INTERVIEW MATERIAL

The recorded interview served as a primary resource for this editorial, with the main outcome being a carefully

refined transcript. To ensure accuracy, clarity, and coherence, the following principles were applied:

- The transcript was carefully edited to enhance readability while preserving the authenticity of the conversation.
- To improve the flow, similar or related questions from the interviewers were merged, and the distinction between the individual interviewers, Stefan Balke and Meinard Müller, was intentionally removed.
- Selected excerpts from the interview have been used in supplementary materials, such as video excerpts for academic outreach.²
- The final version of the transcript was shared with Masataka Goto for review and approval prior to publication.

3 INTERVIEW

3.1 PERSONAL BEGINNINGS

Interviewer: Hello Masataka, we sincerely appreciate you taking the time for this interview. It is an honor to have this conversation with you.

Masataka: It's my pleasure.

Interviewer: As one of the pioneers in MIR, your contributions have been truly influential. Could you take us back to your early days as a Master's student? When did your interest in MIR and music processing first begin?

Masataka: Even before my Master's program, back in junior high school, I was fascinated by using a personal computer to play music with a programmable sound generator and frequency modulation synthesis. Since 1985, I have also enjoyed covering music with a MIDI-controlled electronic keyboard. Since I couldn't transcribe music by ear, I began imagining a system that could do it automatically. I attempted to use a fast fourier transform (FFT) for music analysis in high school but failed

to create such a system. However, that early curiosity later led me to music analysis and computational methods for understanding music.

Then, as an undergraduate student at Waseda University in 1992, I joined a lab that focused on parallel computing. Professor Muraoka,³ my advisor, was an inspiring person. He asked every student what they wanted to do. I told him I was interested in music analysis and asked if I could work on it. He responded, “If it’s an original idea that no one has realized yet, and you believe the technology has potential users, then you can pursue it.” However, this was before the World Wide Web, so I couldn’t just use a search engine to find information. I had to visit the university library and spend a week checking research papers and journals. Eventually, I realized that no one had worked on polyphonic drum transcription. I proposed this idea to my professor, and he encouraged me to pursue it.

When I started my Master’s in 1993, my professor suggested using the Fujitsu AP1000, a massive parallel computer with 64 CPUs, for music analysis. At that time, using such an expensive computing resource for music analysis was highly uncommon. I found this challenge exciting and proposed an idea for audio-based beat tracking, as most researchers in the 1990s were focused on MIDI or monophonic melody lines. No one was analyzing audio recordings of polyphonic pop music the way we do today. In 1993, computing spectrograms of 10–20 songs took hours, but with parallel computing, I could compute the FFT in real time using just five processors.

3.2 EARLY DAYS OF MIR

Interviewer: How was the research landscape in MIR and music processing in those days?

Masataka: In 1992, there was a research community for music information research in Japan, which later evolved into SIGMUS⁴ in 1993. Since I was the only person working on music in my lab, I attended one of its meetings and, while presenting printed papers on related studies, I asked senior researchers if I could be the first to work on polyphonic drum transcription, as it had not been covered in these papers. They confirmed that no one had done this before and encouraged me to continue.

Interviewer: Which proceedings did you consult back then? ISMIR⁵ didn’t exist yet.

Masataka: The International Computer Music Conference (ICMC)⁶ was the primary venue. The International Conference on Acoustics, Speech, and Signal Processing (ICASSP)⁷ was focused more on speech, not music. I was lucky that ICMC 1993 was held in Tokyo, Japan, during my first year of graduate school. That was when I started attending international conferences.

Interviewer: Who were your heroes at the time? Were there leading experts in the field?

Masataka: Keiji Hirata was a significant figure; he founded the SIGMUS society. Internationally, Roger Dannenberg was well-known. After reading ICMC papers, I knew of him, but he didn’t know me at the time. ICMC was where I first met such pioneers. From 1993–1998, I attended ICMC every year.

Interviewer: Was it difficult to get your work published at ICMC or other international conferences?

Masataka: My biggest challenge was translating my work into English, as my English skills were poor. However, I was fortunate that my research was considered original, so my papers were accepted at ICMC every year from 1994–1998.

Interviewer: How did you evaluate your systems back then? Were there test datasets or established best practices?

Masataka: Evaluation was informal, as researchers typically demonstrated results and provided a few examples in their papers. There were no standardized benchmarks. After working on beat tracking for several years, I realized I needed a dataset with annotations to compare improvements objectively. I manually annotated beat positions, spending weeks creating a dataset. There were no annotation tools like Sonic Visualiser,⁸ so I had to develop my own audio editor for annotation.

In 1997, I wrote one of the first papers on MIR evaluation, titled “Issues in Evaluating Beat-tracking Systems” (Goto and Muraoka, 1997). It was an early attempt to establish an evaluation framework for music systems. However, I couldn’t share my dataset due to copyright restrictions on commercial music.

3.3 IDEA BEHIND THE RWC

Interviewer: That sets the stage for your next big step, the first ideas behind the RWC Music Database, right?

Masataka: Yes, but before that, I completed my PhD in 1998 and joined my current lab at the National Institute of Advanced Industrial Science and Technology (AIST)⁹ in Tsukuba. I started working on fundamental frequency (F0) estimation for melody and bass lines, including melody extraction. Again, I needed annotated data, so I developed another annotation tool and spent several days annotating F0 contours on commercial music. I published my melody extraction paper at ICASSP (Goto, 2000). As I published more on evaluation, I saw the need for copyright-cleared datasets for benchmarking, leading to the idea for the RWC Music Database.

Interviewer: So the motivation was to share data with the community?

Masataka: Exactly. The dataset originated from a Japanese government-funded project called “Real World Computing (RWC).”¹⁰ The project aimed to advance computing technology. As our lab contributed to this project, I was given the financial opportunity to create a dataset for music research. Initially, it was intended for our own

use, but I decided to build a dataset that could be used by the entire research community.

Interviewer: The RWC dataset includes various genres, such as classical, pop, and jazz music. How did you decide what to include?

Masataka: Initially, I focused on pop music, but I realized a comprehensive dataset should include other genres. Since I wasn't deeply familiar with jazz and classical music, I consulted experts from the Japanese SIGMUS community. Their advice helped shape the recorded contents, making it a collaborative effort.

Interviewer: How did you approach the design, for example?

Masataka: I asked Dr. Keiji Hirata for advice on jazz music, considering what should be included if we had the resources to produce 50 jazz pieces. We decided to include seven different instrumentations of five standard-style original jazz pieces. I also sought advice from Dr. Yuzuru Hiraga on classical music, and we decided to include original recordings of 50 famous pieces with a rich variety of instrumentation and styles.

Interviewer: Then you expanded the dataset with more genres?

Masataka: Yes, after developing four databases for popular music, royalty-free music, classical music, and jazz music (Goto et al., 2002), I decided to develop two more databases: the Music Genre Database and the Musical Instrument Sound Database (Goto et al., 2003). We again carefully designed the recorded contents through discussions.

3.4 PRODUCTION AND ANNOTATION

Interviewer: Were there considerations regarding copyright clearance, and how did this influence your selection of musical pieces?

Masataka: We originally considered two approaches. First, we could pay for licenses to use existing commercial music, which might grant permission for research purposes. However, this approach had limitations—we might not be able to freely distribute the recordings. Due to this restriction, I chose the second approach: creating everything from scratch to ensure full rights for researchers.

Interviewer: That must have been a challenging task. How did you approach this while ensuring high-quality musical pieces and recordings?

Masataka: Yes, it was challenging. We needed the music pieces and recordings to sound professional, so we relied on a music production company to carry out the production. Our goal was to have professional musicians perform specifically for research purposes.

Interviewer: At that time, did people understand the concept of a research music database?

Masataka: No, the idea was novel. I had to explain that we are researchers aiming to advance music technologies for music cultures, and that we need music

databases performed by professional musicians for research purposes, not for commercial purposes.

Interviewer: This must have been a major undertaking. I imagine finding musicians willing to contribute was challenging. How did you convince them to participate?

Masataka: We did not find them ourselves. Instead, the music production company knew or found professional musicians for us. The musicians understood the research purposes and were financially compensated, as it was not voluntary work.

Interviewer: The orchestral pieces were performed by a well-known professional orchestra, the Tokyo City Philharmonic Orchestra. That must have been costly and demanding.

Masataka: Yes, it was expensive. The music production company booked the orchestra and rented a professional concert hall for several days, ensuring high-quality recordings with the help of professional engineers.

Interviewer: As for popular music and related genres, what role did MIDI play in the production?

Masataka: I formulated a concept that balanced human performances with MIDI-controlled performances. For example, in popular music, a higher proportion of acoustic instruments played by humans is desirable. However, MIDI-controlled digital instruments are also commonly used. Considering budget constraints, I decided that guitars should primarily be performed live by human musicians, while bass and drums would also include a certain proportion of live performances. The rest would be handled by MIDI-controlled instruments. This distinction is documented on the RWC website (Goto et al., 2002).¹¹

Interviewer: Nowadays, MIDI is often used as a basis for generating audio annotations in MIR research. In the RWC dataset, were the MIDI files perfectly aligned with the audio recordings?

Masataka: Not always. Even for pop music, MIDI files were not perfectly aligned because they were transcribed by ear. In Japan, the karaoke industry created MIDI files in this way, and professional transcribers from that industry created our MIDI files. That's why we later released the AIST Annotations (Goto, 2006a), which include time-synchronized MIDI files that were manually aligned using a tool specifically developed for this purpose.

Interviewer: After the release of the RWC dataset and its MIDI files, additional annotations—such as beat and structure—were introduced later on. Did you plan these annotations from the beginning, or did they emerge naturally through subsequent research?

Masataka: When designing RWC, I did not initially plan to create extensive annotations myself but hoped to develop them in collaboration with other researchers. However, through my research, I created the necessary annotations and later shared them with the community. As I worked on beat and melody annotations, I recognized their importance for research. Audio

alone was insufficient for evaluation, so I began creating annotations around 2001. They were initially intended for internal use, but I later made them available to the community.

Interviewer: Given the lack of specialized annotation software at the time, how did you produce your annotations, and how did you ensure their correctness and accuracy?

Masataka: First, I developed a custom annotation tool.¹² Furthermore, I hired a music college graduate to assist with annotations. While I annotated some parts myself, I couldn't do everything alone.

Interviewer: What methods did you use for verification?

Masataka: The annotation editor I developed provided both visual and auditory feedback. For beat annotations, click sounds were overlaid on beat and downbeat positions to check alignment so that we could verify timing by ear. To refine the F0 annotations, I used signal-processing techniques to create two types of audio playback: a melody-only version generated from the F0 and harmonics and a karaoke-style version with the melody canceled. If artifacts were heard, it indicated errors in the F0 annotations, and we could improve them.¹³

Interviewer: Despite your efforts, there could be imperfections in annotations. How did you feel about them, and how did you handle corrections?

Masataka: I did my best, but imperfections were unavoidable. Since this was the very beginning of that kind of annotation, sharing something was far more important than achieving perfection, right? Even if it was not perfect, I believed the community would refine it over time.

Interviewer: So the annotation process and software development evolved into a research area of its own. I assume this helped lay the foundation for your later work on interfaces and software like Songle (Goto et al., 2011; Goto and Dannenberg, 2019), right?

Masataka: Yes, that's right. Look at my papers on chorus detection (Goto, 2003c, 2006b)—for example, I explicitly mentioned the development of the annotation editor. Even before the AIST Annotations, in the early 2000s, I described the annotation process in my papers. The development of these annotation methods itself was an important research contribution.

Interviewer: I have a quick question about the names “RWC” and “AIST.” Everyone refers to the RWC dataset [as such], so why are the annotations called “AIST Annotations” instead of “RWC Annotations”?

Masataka: Because the annotation process was no longer funded by the RWC, as the project ended in March 2002. Since I was working at AIST, it was natural for me to refer to the annotations as the “AIST Annotations” at that time.

3.5 IMPACT AND APPLICATIONS

Interviewer: The RWC dataset and its annotations have had a lasting impact.

Masataka: Yes, after the release of the RWC database, the research community found it valuable. At conferences, I received a lot of positive feedback, and that motivated me to continue contributing additional annotations.

Interviewer: I first discovered the RWC dataset around 2005, and it was a real eye-opener for me. At the time, I was a postdoc at Bonn University and knew I needed access to the audio data. So, I asked my professor, Michael Clausen, to order the complete CD collection—all the subcollections. I still have the receipt and the order form, signed by him. One of the key aspects that made RWC so valuable was not just the annotations but also the availability of the audio data. The RWC was the first dataset to provide both at scale, making it a significant resource. How did you approach distributing the dataset?

Masataka: Yes, I agree. While annotations are useful, direct access to the original audio is essential for research. Without it, audio researchers couldn't do much. Sharing metadata, annotations, or audio features such as mel-frequency cepstral coefficients (MFCCs) alone wasn't sufficient.

As for the distribution of audio data, the internet was too slow at the time, and the dataset was too large for online sharing. Even though it contained only 315 musical pieces in waveform, I wanted to provide raw, uncompressed audio at 44.1 kHz—without MP3 compression. The only feasible way to distribute it worldwide was through physical CDs, so we began duplicating and shipping them.

Interviewer: I kept the receipt to ensure I could prove that we legally purchased the RWC and avoid any legal disputes with you and your lawyers. [Laughing.]

Masataka: Oh, wow! Yes, that's nice. [Also laughs.]

Interviewer: You've shared valuable insights into the RWC's production and annotation. How did it shape your research career, and how did the MIR community respond?

Masataka: Since the RWC was designed for community sharing, I expected it to be used by other researchers. However, when we started working on it in 2000, I had not yet attended ISMIR conferences and had no idea that such a wonderful research community would emerge.

When I attended ISMIR for the first time in 2002 and presented the RWC database, the positive reception from the community was incredibly encouraging. It made me realize the value of our work and motivated me to further develop annotations for the community.

After its distribution, the RWC was widely adopted by researchers in the MIR field. I was always delighted to see it cited in research papers. Sometimes I would even search for “RWC” in conference proceedings to count how many papers referenced it. Seeing that impact is truly rewarding.

Interviewer: The Beatles dataset (Harte and Sandler, 2005) was a key resource for harmony analysis. At one point, it was nearly impossible to publish a paper on the

topic without using it. What kinds of research tasks did the RWC help drive?

Masataka: Many research topics such as structure analysis and beat tracking were driven by the RWC. Without accessible datasets, students might abandon research ideas, but with audio and annotations available, they can start immediately. The RWC helped launch numerous research projects in audio and music analysis. For the first 5–10 years after its release, the RWC was widely used. As more datasets became available, researchers had more options, but for the first decade, it remained a leading resource.

Interviewer: Yes, the RWC was one of the first benchmark datasets for a variety of standard MIR tasks. What are some of the most creative or unexpected applications of the RWC that you've come across?

Masataka: That's a tough question since there were many. I don't recall a specific example, but when we created the RWC in 2000–2001, I never imagined the breadth of research it would be used in. Music analysis was still in its early stages, yet the dataset found many unexpected uses beyond its original intent. It was exciting to see such diverse applications.

Interviewer: We reviewed the literature on this topic and found it fascinating to see how the RWC dataset has inspired a wide range of work—such as MedleyDB,¹⁴ evaluations using re-synthesized components of the RWC,¹⁵ and creative projects like Matthew Davies' mashup interface.¹⁶

Masataka: Oh, yes, mashups! When we created the RWC around 2000, I never imagined it would be used for that. It was intended for music analysis, but it later became valuable for creative MIR tasks. The same happened with vocal synthesis. When Hatsune Miku and VOCALOID¹⁷ appeared in 2007, we started a singing synthesis project to make Hatsune Miku sing more naturally (Nakano and Goto, 2009). We wanted to create demo videos, but we couldn't use commercial music.

We selected several tracks from the RWC Music Database and swapped human singing with computer-generated vocals. The synthesized singing was created by Tomoyasu Nakano. He provided me with an audio file, and I mixed it for demo videos. When we posted the first demo video on Nico Nico¹⁸ in April 2008, we got a huge response. We really enjoyed it!

Interviewer: What tools did you use for mixing?

Masataka: At that time, we used Pro Tools (Digital Audio Workstation, Avid Technology, Burlington, Massachusetts, United States) for RWC, so I used Pro Tools to remix the Hatsune Miku vocal with accompaniments. Using the RWC database for singing synthesis demonstrations was definitely a fun and unexpected application.

Interviewer: Recently, the training of generative models has raised concerns about copyright infringement and

the unauthorized use of data. Have you encountered any misuse of RWC or potential copyright violations?

Masataka: No, I haven't observed that. The dataset was distributed to researchers who pledged to use it responsibly. We trusted the research community, and I don't recall any specific cases of misuse.

3.6 LESSONS LEARNED AND FUTURE DIRECTIONS

Interviewer: Almost all approaches we use today are highly data-intensive, requiring very large datasets. In this context, the RWC might be considered relatively small. What are your thoughts?

Masataka: Nowadays, many researchers have created and published datasets, and the importance of these resources is highlighted at ISMIR conferences. As we see database papers every year, the contribution of datasets to the research community is now widely recognized.¹⁹ While the RWC was highly significant 20 years ago, its role has evolved with the emergence of many alternative great datasets. However, its lasting value lies in providing shared audio data.

Interviewer: How has the relevance of the RWC dataset evolved over the years in the MIR community?

Masataka: The research community has added many annotations to the RWC, keeping it useful. It is useful for many research topics, but not for all. For studies requiring a large number of music tracks, the RWC would be too small. Fortunately, other datasets have emerged to meet such needs. Around the year 2000, creating and sharing music datasets was not as common as it is today. At that time, the RWC played a crucial role in demonstrating the importance of dataset creation and annotation. Today, dataset research is widely recognized.

Interviewer: As you mentioned, the RWC is not broad enough for training large systems. However, as a controlled test dataset for evaluation, it remains valuable. How can the community adapt or extend the RWC to better meet today's research needs? Do you have any suggestions?

Masataka: The current distribution of the RWC is outdated, as it still relies on compact discs, whereas modern datasets should be accessible online. While the dataset could be extended, doing so should be a collaborative effort. In the original RWC paper, I expressed the hope that other researchers would contribute further, either by creating annotations or expanding the dataset. No single researcher can do everything—this is the strength of the research community.

Interviewer: There have been discussions about an "RWC 2.0"—a publicly accessible version under an open license. Do you see it as a viable approach? How can community involvement help sustain the RWC?

Masataka: Yes. At last year's ISMIR conference, Geoffroy Peeters asked me to share excerpts of RWC audio for

hands-on and educational use in his tutorial, and I was happy to agree. While the RWC dataset may seem small by today's standards, it continues to serve as a valuable resource for tutorials, university courses, and concrete examples. Copyright-cleared datasets are essential for both research and education, and the community plays a key role in sustaining the RWC's relevance. Recognizing its continued importance—and following your [Meinard's] strong recommendation—I finally decided to make the RWC audio freely available online, with the support of your team.

Interviewer: The RWC has been an influential dataset for 25 years, shaping MIR research through its audio, annotations, and collaborations. Looking back, do you feel the effort and challenges were worth it? What aspect of the RWC are you most proud of?

Masataka: Yes, it was absolutely worth it! Seeing the RWC referenced in research papers and presentations has been incredibly rewarding. It was never just about creating a dataset but about contributing something meaningful to the community. Every time someone mentioned the RWC, it was a great moment for me. Developing the RWC was a long and challenging process, but knowing it has been widely used makes all the effort worthwhile. And it was never just my work—it was a collective effort. Many people contributed to its creation and spread and have continued to build on it over the years. I am very proud of that.

Interviewer: Looking ahead 20 years, what do you hope to see for the RWC and other datasets? How should they evolve to meet the needs of future research?

Masataka: I hope the RWC stays sustainable and valuable. If the research community maintains it, it can remain relevant and useful beyond my lifetime. My biggest concern is ensuring it continues without relying on a single person for the next 20, 50, or even 100 years. *[Said with a smile.]*

Interviewer: I'm sure this will happen. The RWC is a great dataset, a lifelong mission, and an incredible achievement. Thank you for this insightful interview and for sharing your journey, insights, and vision.

Masataka: Thank you very much.

4 CONCLUSIONS

Through this interview, we had the unique opportunity to hear directly from Masataka Goto about the creation, impact, and evolution of the RWC Music Database. By sharing the motivations, challenges, and decisions behind the RWC, this conversation provides valuable insights into the role of datasets in shaping MIR research. Beyond its history, the interview also explored broader issues like dataset sustainability, community involvement, and the move toward open access—topics that remain highly relevant today.

We hope this interview is not only interesting for those familiar with the RWC but also meaningful for the wider MIR community. As one of the first large-scale datasets in MIR, the RWC played a key role in advancing research over the past 25 years. However, its long-term relevance depends on how we, as a community, adapt it to new challenges, such as improving metadata and ensuring accessibility. The shift toward community-driven dataset curation offers a chance not only to sustain the RWC but also to set an example for future datasets. More broadly, this discussion highlighted the need for shared, standardized, and openly available research resources (McFee et al., 2019; Serra, 2014). Looking ahead, we believe the research community will play a crucial role in keeping datasets like the RWC alive—supporting innovation, reproducibility, and collaboration in MIR.

ACKNOWLEDGMENTS

Meinard Müller and Stefan Balke are funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) under grant no. 500643750 (MU 2686/15-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich–Alexander–Universität Erlangen–Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS. We gratefully acknowledge the thoughtful feedback and helpful suggestions provided by Simon Dixon, Florence Levé, and Bob Sturm.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHORS' CONTRIBUTIONS

MM and SB planned and conducted the interview. MM edited the interview transcription and wrote the initial version of the paper. MG provided the interview content. All co-authors of this paper actively participated in the preparation of the final manuscript.

NOTES

1. Similar interviews were frequently published in the *Computer Music Journal*—see, for example, Boulanger et al. (1990) and Roads and Minsky (1980).
2. A curated 10-minute excerpt of the full interview is available on YouTube: https://youtube.com/playlist?list=PLpaL3fT5fH2ohNCY3nw_k63H7ozkixvp5&si=WuLXyfiscVXaa0Yn.
3. Professor Yoichi Muraoka was a prominent figure in computer science, particularly known for his contributions to parallel computing and information processing.
4. SIGMUS (est. 1993, <https://sigmus.jp>) is the Special Interest Group on Music and Computers within the Information Processing Society of Japan.
5. The first ISMIR conference, then called the International Symposium on Music Information Retrieval, was held in the year 2000.

6. The ICMC, founded in 1974, was and still is a leading conference on computer music, electronic music, and computational music analysis.
7. The ICASSP is one of the most prestigious conferences in signal processing, first held in 1976.
8. Sonic Visualiser is an open-source software application, extensively used in the MIR community, designed for visualizing, analyzing, and annotating audio signals (Cannam et al., 2010).
9. The AIST (est. 2001) is one of Japan's largest public research institutes, covering a wide range of research topics. When Masataka Goto joined in 1998, it was called the Electrotechnical Laboratory, which was reorganized into AIST along with other labs in 2001.
10. The RWC initiative was a national project that began in 1992 and lasted for 10 years; see (Otsu, 1993).
11. <https://staff.aist.go.jp/m.goto/RWC-MDB/>.
12. The tool was referred to as a “multipurpose music-scene labeling editor” in Goto (2003b).
13. These verification methods are explained in Goto (2006a).
14. MedleyDB is a dataset of multitrack recordings with isolated instrument stems, designed for research in source separation, instrument recognition, and other MIR tasks (Bittner et al., 2014).
15. For example, the pYIN algorithm was evaluated on re-synthesized melody tracks (Mauch and Dixon, 2014).
16. A mashup is a musical composition that blends vocals from one song with the instrumental or harmonic elements of another; see (Davies et al., 2014).
17. Hatsune Miku is a virtual singer and the most famous use of VOCALOID, a voice-synthesis software by Yamaha that generates singing from text and melody input. Created by Crypton Future Media, her voice has become a global icon, appearing in original songs, concerts, and multimedia projects.
18. Niconico is the most popular Japanese video-sharing platform. The original post can be found at <https://www.nicovideo.jp/watch/sm3128145>, and it was later uploaded to YouTube at <https://youtu.be/msEN5bbIgbC>.
19. This is also reflected in the Dataset Track of *Transactions of the International Society for Music Information Retrieval*, which was established to provide a formal venue for publishing and documenting datasets—promoting transparency, reproducibility, and accessibility within the MIR community.

AUTHOR AFFILIATIONS

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

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

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

REFERENCES

- 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., 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).
- Boulanger, R., Mathews, M., Vercoe, B., and Dannenberg, R. (1990). Conducting the MIDI orchestra, Part 1: Interviews with Max Mathews, Barry Vercoe, and Roger Dannenberg. *Computer Music Journal*, 14(2), 34–46.
- 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).
- Cheng, T., and Goto, M. (2023). Transformer-based beat tracking with a low-resolution encoder and a high-resolution decoder. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Milan, Italy (pp. 466–473).
- 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.
- 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).
- 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).
- Goto, M. (2000). A robust predominant-F0 estimation method for real-time detection of melody and bass lines in CD recordings. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Vol. 2, pp. 757–760).
- Goto, M. (2003a). 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. (2003b). Music scene description project: Toward audio-based real-time music understanding. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)* (pp. 231–232).
- Goto, M. (2003c). SmartMusicKIOSK: Music listening station with chorus-search function. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)* (pp. 31–40).
- Goto, M. (2004). Development of the RWC music database. In *Proceedings of the International Congress on Acoustics (ICA)* (pp. 553–556).
- Goto, M. (2006a). AIST Annotation for the RWC Music Database. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)* (pp. 359–360).
- Goto, M. (2006b). A chorus section detection method for musical audio signals and its application to a music

- listening station. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5), 1783–1794.
- Goto, M., and Dannenberg, R. B.** (2019). Music interfaces based on automatic music signal analysis: New ways to create and listen to music. *IEEE Signal Processing Magazine*, 36(1), 74–81.
- 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).
- Goto, M., and Muraoka, Y.** (1997). Issues in evaluating beat tracking systems. In *Working Notes of the IJCAI-97 Workshop on Issues in AI and Music-Evaluation and Assessment* (pp. 9–16).
- Goto, M., Yoshii, K., Fujihara, H., Mauch, M., and Nakano, T.** (2011). Songle: A web service for active music listening improved by user contributions. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Miami, Florida, USA (pp. 311–316).
- 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.
- Harte, C., and Sandler, M. B.** (2005). Automatic chord identification using a quantised chromagram. In *Proceedings of the Audio Engineering Society Convention*, Barcelona, Spain.
- Itoyama, K., Kitahara, T., Komatani, K., Ogata, T., and Okuno, H. G.** (2006). Automatic feature weighting in automatic transcription of specified part in polyphonic music. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, Victoria, Canada (pp. 172–175).
- Joder, C., Essid, S., and Richard, G.** (2011). A conditional random field framework for robust and scalable audio-to-score matching. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(8), 2385–2397.
- 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).
- 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.
- Nakano, T., and Goto, M.** (2009). VocalListener: 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).
- Otsu, N.** (1993). Real world computing program - overview. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, Nagoya, Japan (pp. 1065–1068).
- 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.
- Roads, C., and Minsky, M.** (1980). Interview with Marvin Minsky. *Computer Music Journal*, 4(3), 25–39.
- 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).
- Serra, X.** (2014). Creating research corpora for the computational study of music: The case of the CompMusic project. In *Proceedings of the AES International Conference on Semantic Audio*, London, UK.
- 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).
- Wu, Y., and Li, W.** (2019). Automatic audio chord recognition with MIDI-trained deep feature and BLSTM-CRF sequence decoding model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(2), 355–366.

TO CITE THIS ARTICLE:

Müller, M., Balke, S., & Goto, M. (2025). The Story Behind the RWC Music Database: An Interview with Masataka Goto. *Transactions of the International Society for Music Information Retrieval*, 8(1), 154–164. DOI: <https://doi.org/10.5334/tismir.261>

Submitted: 1 April 2025 **Accepted:** 17 May 2025 **Published:** 19 June 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.