

RWC Music Database: Popular, Classical, and Jazz Music Databases

Masataka Goto

National Institute of Advanced Industrial Science and Technology (AIST)
& "Information and Human Activity", PRESTO, JST
IT, AIST, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

Takuichi Nishimura

National Institute of Advanced Industrial Science and Technology (AIST)
CARC, AIST, 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, Japan

Hiroki Hashiguchi

Mejiro University
4-31-1 Naka-Ochiai, Shinjuku-ku,
Tokyo 161-8539, Japan

Ryuichi Oka

University of Aizu
Aizu-Wakamatsu, Fukushima 965-8580, Japan

ABSTRACT

This paper describes the design policy and specifications of the *RWC Music Database*, a music database (DB) that is available to researchers for common use and research purposes. Various commonly available DBs have been built in other research fields and have made a significant contribution to the research in those fields. The field of musical information processing, however, has lacked a commonly available music DB. We therefore built the *RWC Music Database* which contains four original DBs: the *Popular Music Database* (100 pieces), *Royalty-Free Music Database* (15 pieces), *Classical Music Database* (50 pieces), and *Jazz Music Database* (50 pieces). Each consists of originally-recorded music compact discs, standard MIDI files, and text files of lyrics. These DBs are now available in Japan at a cost equal to only duplication, shipping, and handling charges (virtually for free), and we plan to make them available outside Japan. We hope that our DB will encourage further advances in musical information processing research.

1. INTRODUCTION

We believe that research into musical information processing will be advanced if music databases (DBs) become available that can be used by various researchers. The main purposes and advantages of such commonly available DBs can be summarized as follows:

- Researchers will be able to use the DBs as a common benchmark for comparing and evaluating various methods related to musical information processing. The lack of common music DBs available worldwide for research purposes at almost no cost has made it difficult to establish benchmarks (evaluation frameworks) for much of the research done regarding musical information processing.
- The DBs will accelerate the progress of various forms of research using statistical methods. Recent progress in the use of statistical methods in other research fields such as speech recognition has been largely due to the availability of large DBs.
- Researchers will be able to use the DBs for research publication and presentation without conventional copyright restrictions. It has been difficult to demonstrate research using copyrighted musical pieces that will be included in, for example, conference videos and CD-ROMs.

Although there is an enormous amount of music available on commercially distributed compact discs, it is difficult to use this music for the above purposes because of copyright issues. Commonly available DBs with copyright-cleared pieces are therefore essential to encourage the healthy development of this field of research.

Various commonly available DBs have been built in other research fields since the importance and significance of such DBs have been widely recognized. In the research field of speech information

processing, for example, many DBs (corpora) have been built. Furthermore, recognition of the need for commonly available DBs has led to the creation of the Linguistic Data Consortium (LDC) in the USA and the European Language Resources Association (ELRA) to support the development and sharing of resources. There are also several DBs in the field of image processing. On the other hand, in the field of musical information processing, there is no commonly available music DB containing hundreds of musical pieces.

We therefore have built a music DB — the *RWC (Real World Computing) Music Database* — that gives researchers freedom of common use and research use. (The DB is intended to be used only for research purposes.) The *RWC Music Database* is composed of four original DBs: the *Popular Music Database*, *Royalty-Free Music Database*, *Classical Music Database*, and *Jazz Music Database*. In the following sections, we describe the design policy and provide an overview of each of the four DBs.

2. OVERVIEW OF THE RWC MUSIC DATABASE

We had to address various design, trade-off, and copyright issues in building the *RWC Music Database*. In this section, we discuss some of the most important issues.

- Contents of the database (DB)
An ideal music DB would contain many richly varied musical pieces, in various genres, of the highest quality possible. For practical purposes, though, we had to build our DB under production resource constraints such as our budget and available time. We therefore took up three major music genres — popular, classical, and jazz — and tried to include as many realistic pieces as possible in a way that reflected the complexity and diversity of real-world music. For example, as well as ensuring that various styles, moods, tempi, and lengths were represented, we also included as many professional composers, lyric writers, arrangers, singers, and players as our resources allowed. To achieve sound quality as high as that of commercially distributed compact discs, we used professional digital equipment for all recording, mixdown, and mastering processes.
- Copyrights of musical pieces
To make our DB available to researchers around the world, we had to obtain all necessary copyrights and neighboring rights for research purposes.¹ We therefore included 215 pieces (for the four DBs) that were all originally performed and recorded for the purpose of inclusion in the DB. For the *Popular Music Database*, we included 100 pieces that were originally composed and arranged in modern popular music styles (the lyrics were also originally written). For the *Royalty-Free Music Database*, we included 15 public-domain traditional popular-music pieces that were originally arranged for this DB. For the *Classical Music Database*, we selected 50 well-known public-domain pieces.

¹Note that our DB is not copyright-free even if it is available for free: we reserve all necessary copyrights and neighboring rights. All users of the DB must submit the user agreement form to the general manager of the *RWC Music Database* (Masataka Goto, contact: m.goto@aist.go.jp).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2002 IRCAM – Centre Pompidou

Table 1. List of music compact discs for distribution (Popular, Royalty-Free, Classical, and Jazz Music Databases).

Contents (Version)	# of discs	Catalog number	# of pieces	Piece number
Popular Music Database (Original Version: <u>Mixed</u>)	7	RWC-MDB-P-2001-M01 – M07	100	No. 1 – 100
Royalty-Free Music Database (Original Version: <u>Mixed</u>)	1	RWC-MDB-R-2001-M01	15	No. 1 – 15
Classical Music Database (Original Version: <u>Mixed</u>)	6	RWC-MDB-C-2001-M01 – M06	50	No. 1 – 50
Jazz Music Database (Original Version: <u>Mixed</u>)	4	RWC-MDB-J-2001-M01 – M04	50	No. 1 – 50

Catalog number: RWC-MDB-[Contents]-[Year]-[Version][Volume No.], Contents: the first letter, Year: Made in 2001, Version: Mixed

For the *Jazz Music Database*, we included 50 pieces where four well-known public-domain pieces were originally arranged and the other 46 pieces were originally composed and arranged.

- Standard MIDI files (SMFs)

We prepared transcribed SMFs for all 215 pieces. These were stored in SMF format 1 (multiple tracks) and conform to the GS format. Given audio signals, most of them were transcribed by ear. For music genres such as popular and jazz where there are typically no detailed scores, these SMFs can be used as effective substitutes for scores. Even for classical music with scores, SMFs that can be freely used for research purposes are valuable. The lyrics of songs are provided as text files.

We used music compact discs (CD-DA: Compact Disc - Digital Audio) as the medium for distributing the audio signals of the DB pieces to researchers. The list of the compact discs and their catalog numbers are shown in Table 1. Each piece has a unique “piece number” numbered in consecutive order from 1 within each DB. The volume number (the last two digits of the catalog number) is used only for putting pieces onto the compact discs and should not be used for reference: a piece should be referred to by the piece number for research use (e.g., RWC-MDB-P-2001 No. 28).

2.1 Popular Music Database

The *Popular Music Database* consists of 100 songs — 20 songs with English lyrics performed in the style of popular music typical of songs on the American hit charts in the 1980s, and 80 songs with Japanese lyrics performed in the style of modern Japanese popular music typical of songs on the Japanese hit charts in the 1990s. All 100 songs with vocals were originally produced in as rich a variety as our resources allowed. The songs were recorded by 148 people including 25 composers, 30 lyric writers, 23 arrangers, and 34 singers. As a result of our attempts to achieve a good male-female balance in the 100 songs and to include songs by vocal groups, there are 50 songs by 15 male singers, 44 songs by 13 female singers, and 6 songs by 6 vocal groups.

2.2 Royalty-Free Music Database

The *Royalty-Free Music Database* consists of 15 songs, 10 well-known standard popular songs with English lyrics and 5 well-known children’s songs with Japanese lyrics. All 15 public-domain songs were originally arranged and recorded. This DB was built to contain well-known popular songs, while the *Popular Music Database* contains only original popular songs. The songs were recorded by 16 people, including two arrangers and three singers.

2.3 Classical Music Database

The *Classical Music Database* consists of 50 pieces:

- Symphony: 4
- Concerto: 2
- Orchestral: 4
- Chamber: 10
- Solo: 24
- Vocal: 6

All 50 public-domain pieces were originally recorded for our DB (not all movements were recorded: a certain movement was selected and recorded for several categories such as symphony and concerto). These pieces were selected to represent a rich variety of instrumentation, style, period, composer, and mood. We did not intend to produce a mere anthology of well-known musical pieces: we tried to include pieces that have been previously used in research or have interesting aspects from a research viewpoint. The pieces were recorded by 115 people including a philharmonic orchestra (72 players with 1 conductor), 16 pianists, and 4 violinists.

2.4 Jazz Music Database

The *Jazz Music Database* consists of 50 pieces:

- Instrumentation variations: 35 (5 pieces × 7 instrumentations)
- Style variations: 9
- Fusion (crossover): 6

All 50 pieces were originally produced for our DB, except for the composition and lyric writing of four style-variation pieces. First, the instrumentation-variation pieces were recorded to obtain different arrangements of the same piece: five standard-style jazz pieces were originally composed and then performed in modern-jazz style using seven instrumentations: 1) piano solo, 2) guitar solo, 3) duo (vibraphone + piano, flute + piano, or piano + bass), 4) piano trio, 5) piano trio + trumpet or tenor saxophone, 6) octet (piano trio + guitar + alto saxophone + baritone saxophone + two tenor saxophones), and 7) piano trio + vibraphone or flute. Second, the style-variation pieces were recorded to represent various styles of jazz. The nine pieces, which include four well-known public-domain pieces, consist of vocal jazz (two), big band jazz (two), modal jazz (two), funky jazz (two), and free jazz (one) pieces. Finally, the fusion pieces were recorded to obtain music that combines elements of jazz with other styles such as popular, rock, or latin. All the pieces were recorded by 53 people, including four composers and one lyric writer.

3. CONCLUSION

The building and sharing of commonly available databases (DBs) will clearly make an important contribution to the research into musical information processing. With the four DBs that compose the *RWC Music Database*, researchers can now use copyright-cleared musical pieces for each stage of problem finding, problem solution, implementation, evaluation, and presentation.

The *RWC Music Database* was built in fiscal 2000 and 2001 by the RWC Music Database Sub-Working Group (chair: Masataka Goto) in the Real World Computing Partnership (RWCP) funded by the Ministry of Economy, Trade and Industry of Japan [1, 2]. While our DB was built for general purposes related to musical information processing and was designed independently of the ISMIR 2001 resolution (on the need to create standardized MIR test collections), it is consistent with the resolution and can provide useful test sets for various forms of music-related research.

We plan to make our DB available for researchers around the world. In the future, it will be necessary to add various annotations to the DB pieces in cooperation with other researchers. We hope that our DB will be widely used worldwide, and that various other DBs will follow, thus expediting progress in this field of research.

4. ACKNOWLEDGMENTS

We thank everyone who has made this DB project possible. Yuzuru Hiraga and Keiji Hirata devotedly assisted us in designing the Classical and Jazz Music Databases, respectively. Satoru Hayamizu, Hideki Asoh, and Katunobu Itou advised us concerning the development and distribution of the DB. This project has also been supported by many parties involved in the RWC project. The C MUSIC Corporation carried out the production of all the pieces.

5. REFERENCES

- [1] M. Goto *et al.*, “RWC Music Database: Popular music database and royalty-free music database (in Japanese),” *IPSI SIG Notes 2001-MUS-42-6*, vol. 2001, no. 103, pp. 35–42, 2001.
- [2] M. Goto *et al.*, “RWC Music Database: Classical music database and jazz music database (in Japanese),” *IPSI SIG Notes 2002-MUS-44-5*, vol. 2002, no. 14, pp. 25–32, 2002.