

AIST Annotation for the RWC Music Database

Masataka Goto

National Institute of Advanced Industrial Science and Technology (AIST)
IT, AIST, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
m.goto@aist.go.jp

Abstract

In this paper, we introduce our activities regarding the manual annotation of the musical pieces of the RWC Music Database. Although the RWC Music Database is widely used, its annotated descriptions are not widely available. We therefore annotated a set of music-scene descriptions consisting of the beat structure, melody line, and chorus sections. We call this *AIST Annotation*. We also manually synchronized standard MIDI files with the corresponding audio signals at the beat level. We hope that the AIST Annotation will contribute to further advances in the field of music information processing.

1. Introduction

Manual music annotation for musical pieces in a music collection is a laborious task, but it is important because annotated descriptions are useful for training statistical models and for evaluating or benchmarking various systems including automatic music annotation systems. In this paper, music annotation is defined as a process to label events and scenes in musical pieces with correct descriptions regarding the beat structure, melody line, and chorus sections. Since manual annotation is so important, various studies have been reported, including efforts to build special editors for annotating musical pieces by hand [1, 2, 3], work on manual annotation of music collections [4], and discussions of methodologies and file formats for annotation [5, 6].

The RWC Music Database [7], a copyright-cleared music database (DB) available to researchers as a common foundation for research, is already widely used worldwide, but annotated descriptions for its musical pieces are not widely available. The database is distributed as 27 music compact discs (CDs) containing 315 musical pieces and 12 DVD-ROM discs containing 29.1 Gbytes of monaural sound files of 50 instruments. For all 315 musical pieces, standard MIDI files (SMFs) and text files of lyrics (for songs) are also provided on the WWW, but the SMFs are not synchronized with the original audio signals.

To enhance the usefulness of the RWC Music Database, we have made a continuous effort to manually annotate its musical pieces since August, 2001. The following sections

describe our annotated descriptions and issues regarding the annotation and distribution of those descriptions.

2. AIST Annotation

Table 1 lists music scene descriptions [8] for musical pieces we have already annotated in the RWC Music Database. Using our multipurpose music-scene labeling editor (described in [8]), a music college graduate with absolute pitch annotated the pieces with the following descriptions.

2.1. Beat Structure

The hierarchical beat structure consists of the quarter-note level represented as the temporal position of each beat and the measure level annotated by labeling the beginning of each measure on the corresponding beat.¹

Two techniques facilitated this annotation. First, when the audio signal of a track before mixdown² for a musical piece included metronome clicks that were given to musicians to keep the tempo in recordings, its track was analyzed by using a simple amplitude-based event detection method. Beat positions were thus initialized with the detected events and each position was then manually checked and adjusted on the editor while watching the waveform and listening to audio playback with clicks at beat positions as well as short playback excerpts before or after a beat position. Second, given the annotated beat positions and a time-signature assumption, the beginning of all measures after the current cursor position of the editor was automatically labeled.

2.2. Melody Line

The melody line is represented as the temporal trajectory of the fundamental frequency (F0). The F0 is measured in hertz and the discrete time step is 10 ms. For time steps where the melody line is absent, the F0 is set to 0 Hz. Note that the melody line is not represented as a series of either musical notes or MIDI note numbers.

As we did for the beat annotation, the melody line was also initialized with the F0 estimated on a melody track before mixdown when available. The F0 values were graphically set and adjusted on the editor while watching the spectrogram with the melody line and listening to the melody playback generated using the amplitude of harmonics of the

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.

© 2006 University of Victoria

¹When the time signature is 4/4, for example, the beginning of measures is labeled on every four beats.

²Audio signals of tracks before mixdown were stored in the music production system used for the database development [7]. Those signals are usually not available for commercially distributed copyrighted music.

Table 1. List of music scene descriptions annotated in five of six component databases (DBs) in the RWC Music Database.

Component DB in the RWC Music Database	Beat structure	Melody line	Chorus sections	Audio-synchronized SMF
Popular Music DB (RWC-MDB-P-2001 Nos. 1–100)	V	V	V	V
Royalty-Free Music DB (RWC-MDB-R-2001 Nos. 1–15)	V	V		
Classical Music DB (RWC-MDB-C-2001 Nos. 1–50)	V		V	
Jazz Music DB (RWC-MDB-J-2001 Nos. 1–50)	V		V	
Music Genre DB (RWC-MDB-G-2001 Nos. 1–100)	V		V	

currently labeled F0 as well as the melody-cancelled background playback.

2.3. Chorus Sections

The chorus (refrain) sections, which are the most representative thematic sections of a musical piece, are represented as a list of the beginning and end points of every chorus section. When the music structure is obvious, a musical piece is manually segmented into sections and every section is labeled with a section name of the music structure, such as *intro*, *verse A*, *verse B*, *pre-chorus*, *chorus A*, *chorus B*, *post-chorus*, *bridge A*, *bridge B*, and *ending*.

By making the most of the beat-structure annotation, the beginning and end points of each section were easily specified on beat positions while moving the cursor only on beat positions, watching both global and local views of labeled sections, and listening to the audio playback in units of measure or section. It was also useful to highlight each section with a color corresponding to the labeled section name, especially when showing the entire piece in the global view.

2.4. Audio-Synchronized Standard MIDI File (SMF)

We have worked on synchronizing each SMF with the audio signal of the corresponding musical piece. Although the SMFs in the RWC Music Database were transcribed by ear and might not correspond to original scores, they can still be considered a potential source of informative annotated descriptions.³ Using the annotated beat positions of audio signals, it is not difficult to synchronize those positions with beat positions in an SMF and generate a synchronized tempo track for the SMF. But since the beat positions around the introduction and ending of a piece sometimes do not match straightforwardly, the editor had to include a function to edit their positions on a wave or MIDI-piano-roll display. The editor also supported interactive and synchronized audio/MIDI playback during editing.

3. Issues When Sharing Annotated Descriptions

To make annotated descriptions for sound files ripped from CDs available for researchers around the world, an important issue is how to synchronize their temporal axes because different CD drives and ripping software have different temporal offsets or gaps at the beginning of sound files ripped from the same CD. We solve this issue by measuring a gap

³For example, the onset times of drum sounds were extracted from the synchronized SMFs of RWC-MDB-P-2001 and used as the ground-truth annotation for the Audio Drum Detection contest in the Music Information Retrieval Evaluation eXchange (MIREX) 2005.

— a silent period where absolute waveform values are below a threshold — at the beginning of a sound file and storing a very short excerpt of the waveform right after the gap as a signature. We can share its length and the signature, together with annotated descriptions, for the synchronization. This technique can also be used for sharing the annotation for copyrighted music without sharing the music itself.

All descriptions are stored in text files and can easily be converted to any file format such as XML and CSV. Each time step or section (temporal region) is represented, in a separate text file line, as a pair consisting of its absolute time (with temporal resolution of 10 ms) and values/words.

4. Conclusion

We have described *AIST Annotation*, the annotation for the RWC Music Database, which is done for the purposes of statistical learning, evaluation, comparison, shared benchmarking, systematic technology improvement, and so on. We plan to distribute the annotated descriptions on the RWC Music Database home page. We sincerely hope that these descriptions will be widely used⁴ and will contribute to our research field. Although our annotation is not perfect and may include errors, we hope that researchers around the world will also contribute by adding and improving annotated descriptions in various ways and will share their additions and improvements, thus expediting progress in this field of research.

References

- [1] P. Herrera, *et al.*, “MUCOSA: A music content semantic annotator,” in *ISMIR*, 2005, pp. 77–83.
- [2] X. Amatriain, *et al.*, “The CLAM annotator: A cross-platform audio descriptors editing tool,” in *ISMIR*, 2005, pp. 426–429.
- [3] “Sonic visualiser,” <http://www.sonicvisualiser.org/>.
- [4] K. Tanghe, *et al.*, “Collecting ground truth annotations for drum detection in polyphonic music,” in *ISMIR*, 2005, pp. 50–57.
- [5] M. Lesaffre, *et al.*, “Methodological considerations concerning manual annotation of musical audio in function of algorithm development,” in *ISMIR*, 2004, pp. 64–71.
- [6] C. Harte, *et al.*, “Symbolic representation of musical chords: A proposed syntax for text annotations,” in *ISMIR*, 2005, pp. 66–71.
- [7] M. Goto, “Development of the RWC Music Database,” in *ICA*, 2004, pp. I-553–556.
- [8] —, “Music scene description project: Toward audio-based real-time music understanding,” in *ISMIR*, 2003, pp. 231–232.

⁴The AIST Annotation can also be used for development (training) data sets for future MIREX contests.