File Format Design for Interactive Music Service

Inseon Jang, Jeongil Seo, and Kyeongok Kang

This letter describes a file format designed for interactive music service where users are able to recompose music based on their own preferences. The proposed design utilizes the flexible features of the ISO base media file format (ISO/IEC 14496-12, ISO-BMFF). The techniques offered in the proposed file format enable compact storage and efficient preset management, allowing users to recompose music with ease and convenience. Additionally, an implementation scheme for an interactive music player is shown along with its results.

Keywords: File format, interactive music service, preset.

I. Introduction

As digital technology has advanced, IT services have also changed from being one-sided and passive to interactive and user-participant. The success of user-participant services, represented by user generated content (UGC), shows that users are no longer only passive consumers of multimedia content, but are active subjects able to create their own content and share it with others. These recent trends have brought about an increasing need to allow users interested in music production to listen to music based on their own preferences. As a result, research has been carried out for a system that will allow users with individual creativity to interact with music content, for example, the interactive reflective musical system introduced by Pachet at Sony [1]. Also, as a new digital music concept, MPEG-4 object-based technology has been developed for interactive music service [2], [3]. The MPEG-4 standard allows users to interact with audiovisual objects in a scene. It also provides highly efficient coding tools for audiovisual objects [4], as well as a standardized framework for various media types such as text, pictures, animation, and 3D objects.

Beginning in the year 2000, commercial interactive music services have been launched in various countries, for example, U-MYX in the UK, MXP4 and iKlax [5] in France, and Music2.0 in Korea. These types of commercial interactive music service provide various rich interactive functionalities to users based on multitrack music content packaging with pre-defined mixing information. However, they use their own file formats, allowing no interoperability between current interactive music contents. This has made it difficult to vitalize the global interactive music market.

In this letter, we propose a new file format that exploits the latest features and functionalities of different interactive music services, providing them with interoperability. The proposed design is defined based on the ISO base media file format (ISO/IEC 14496-12, ISO-BMFF) [6], which is a flexible file format standard for combining multimedia data in a variety of ways, and thus it supports packaging interactive music content in a compact form that is optimized for interactive music service. Using the proposed file format, a user can record his/her own preset in the form of an interactive music file, which is then delivered to a related server in order to publish it or share it with others. Additionally, the proposed file format was adopted in the interactive music application format (IM AF) [7], which is a standardized MPEG-A multimedia application format (MAF). Note that MAF is an MPEG-A application under ISO/IEC 23000, whose main purpose is to specify multimedia application formats with MPEG and non-MPEG standard technologies beyond file formats, and to fast track them to the multimedia industry. IM AF involves formatting media data related to interactive music service, especially multiple audio tracks with interactive data (for example, presets and rules [8], [9]), and storing such data in an ISO-BMFF. The final draft of the international standard of IM AF was submitted at the 91st MPEG meeting and published in July 2010.

The organization of this letter is as follows. The next section

Manuscript received Apr. 16, 2010; revised July 8, 2010; accepted Aug. 11, 2010.

This work was supported by the IT R&D program of KEIT&KCC&MKE, Rep. of Korea [KI001932, Development of Next Generation DTV Core Technology].

Inseon Jang (phone: +82 42 860 5791, email: jinsn@etri.re.kr), Jeongil Seo (corresponding author, email: seoji@etri.re.kr), and Kyeongok Kang (email: kokang@etri.re.kr) are with the Broadcasting & Telecommunications Convergence Research Laboratory, ETRI, Daejeon, Rep. of Korea.

doi:10.4218/etrij.11.0210.0129

presents a brief introduction of an interactive music service. Section III describes the proposed file format. In section IV, we illustrate the implementation scheme and its performance results. Finally, we conclude in section V.

II. Interactive Music Service

General music content is created through recording, mixing, and mastering processes by music producers or engineers. A producer records a number of audio tracks, which are usually recorded separately for efficient production, and generates a final (stereo) track for distribution through mixing and mastering processes. Because all audio tracks are already mixed as one stereo track, users can control only the overall volume, and thus are not able to listen to a specific instrument or vocal.

Interactive music content consists of audio tracks prior to conventional mixing and preset data. A preset is pre-defined mixing information on multiple audio tracks and provides information such as the number of audio tracks involved in the preset, the related audio track IDs, and the volume gains. Through a preset, a producer is able to create several versions of a particular piece of music, such as karaoke, rhythmic, or his/her own mixed versions.

By means of an interactive music player, users can select and control the audio tracks in an interactive music file, along with their volume, directly. Or, users can select one preset among the presets stored in the file, so that they can also listen to music easily and conveniently based on their own preferences. Furthermore, they can make their own music UGC by controlling and remixing the interactive contents.

Also, other media data can be additionally stored and played, allowing users to enrich their interactive experience; for example, song lyrics or images related to the song or artist can be displayed.

III. File Structure for Interactive Music Content

Resources within an interactive music file format are wrapped into a single ISO-BMFF. ISO-BMFF is designed to contain timed media information for presentation in a flexible, extensible format that facilitates the interchange, management, editing, and presentation of the media. ISO-BMFF is structured in an object-oriented manner to enable the simple decomposing of a file into a series of objects that have their own names, sizes, and defined specifications according to their intended purposes.

Figure 1 illustrates a file structure for interactive music content. This type of file structure mainly consists of `ftyp`, `moov`, and `mdat` boxes. The type or brand of file is set in the `ftyp` box. The `moov` box contains multiple `trak` boxes and one `meta` box to provide information for the collection of resources in the `mdat` box.

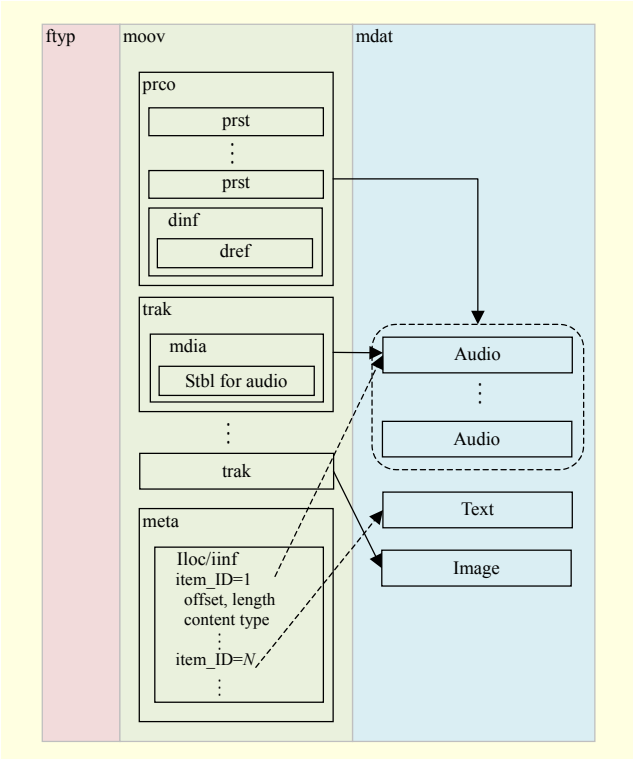

Fig. 1. File structure for interactive music content.

This file structure also includes specific boxes for storing preset information. Preset information is related to the presentation, that is, audio playback, and its characteristics are distinguished from the descriptive metadata. For compact storage and efficient preset parsing, it is more efficient to store preset information in the newly defined preset container box (`prco`) inside the `moov` box for presentation metadata rather than in the `meta` box for descriptive metadata.

The `prco` box contains fields for general information, such as the number of presets and the default preset ID, which indicates the preset activated at the initial condition without any user interaction. Also, one or more preset boxes (`prst`) and only one `dinf` box are included in the `prco` box. Each `prst` box contains specific pre-defined mixing information, such as preset ID, track IDs involved in the preset, the preset type, playback volume gain, and preset name.

Also, the `dinf` box inside the `prco` box contains reference information that declares the location of the interactive music file where audio tracks involved in the preset are stored. Originally, the `dinf` box in the ISO-BMFF specification was stored in the `minf` box or `meta` box, and is used to store the data reference objects that declare the location of the media information on a track. This concept has been extended, allowing a preset to be exported as a separate file. The preset file can also refer to an external interactive music file even if

multiple audio tracks are not stored in the same interactive music file where the preset is stored. Due to the large size of an interactive music file, the preset file can be effectively applied toward exporting, sharing, or publishing a user-defined preset as a UGC in an interactive music service. The data reference declaring the location of an external file is generally described as URL or URN. For an interactive music service, an International Standard Recording Code as a URN is appropriate for providing interoperability between other interactive music contents.

IV. Interactive Music Player Architecture

The architecture of an interactive music player is shown in Fig. 2. The player consists of a file parser, preset parser, decoding modules, metadata parsing module, and mixer. The process of playing an interactive music file is as follows.

First, an interactive music file is loaded into the player. The file parser module initially confirms the file type format, which is the value stored inside the `ftyp` box. The player then continues to read the `moov` box and parse the data stored inside the `meta` and `trak` boxes. The information obtained from parsing the data will be used to locate the data inside the `mdat` box. If a user selects a preset on the player, only the relevant audio tracks are decoded and mixed according to the volume gains which are parsed from the `prst` box. If the user controls each audio track directly on the player, only the selected audio tracks are decoded and mixed with the volume gains controlled by the user. Timed text and image data are decoded in parallel with the audio data. Synchronized with audio playback, timed text data for lyrics can be displayed on a thumbnail image.

Figure 3 shows an example of an interactive music player. The player supports general functions, such as random access and play/pause, as in a conventional music player. After loading an interactive music file selected by the user, the player shows a user-mix page that includes buttons for playback selection and slide bars for volume control of multiple audio tracks. Using these functions, the user can select audio tracks and control their volume (see Fig. 3(a)). Also, in the preset-mix page, the user can choose one of the presets created by a producer, and the player will show the specific preset information, such as audio tracks and their volume gains, according to the user's preset choice (see Fig. 3(b)).

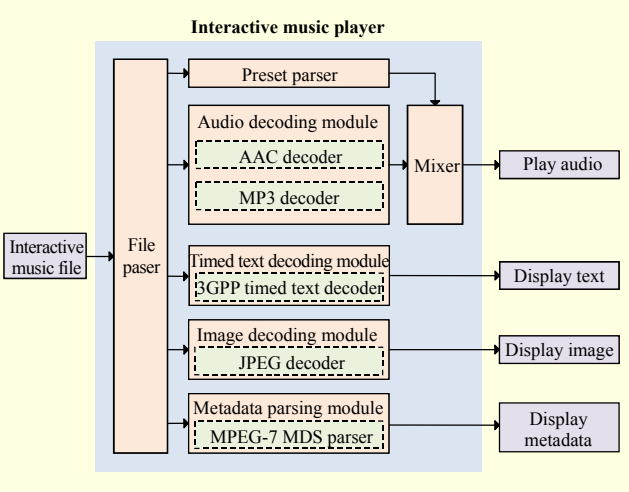

Fig. 2. Interactive music player architecture.

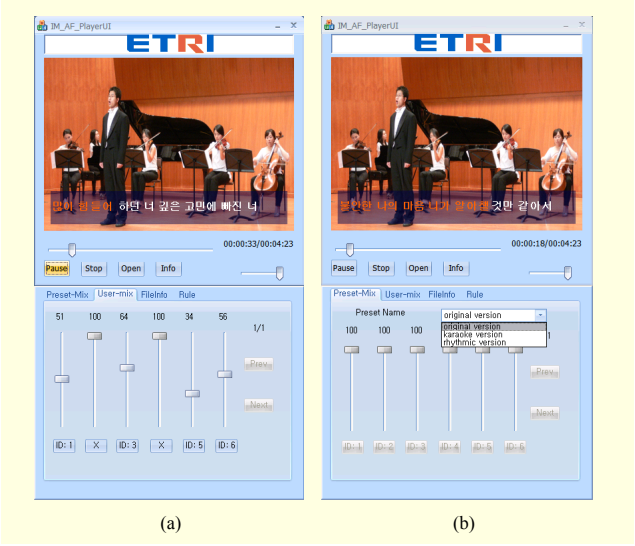

Fig. 3. Example of an interactive music player: (a) user's direct control and (b) preset selection.

V. Conclusion

In this letter, we suggested a file format for an interactive music service and showed the implementation of an interactive music player. The proposed file format uses the flexible features of ISO-BMFF, and thus provides interoperability among interactive music content, supporting the compact storage and efficient management of presets as well as other resources. Currently, MPEG has standardized the interactive music application format (IM AF) [7] to provide and guarantee interoperability between different interactive music content. An international standard for IM AF will be published in 2010, and a standardization of its conformance file and reference software is in progress [10]. As the proposed file format technology utilizes IM AF, it is expected to enable commercial interactive music service based on service-specific and interoperable file formats.

References

[1] I. Deliège and G. Wiggins, *Musical Creativity: Multidisciplinary*

Research in Theory and Practice, Psychology Press, 2006.

[2] T.J. Lee et al., "An Object-Based 3D Audio Broadcasting System for Interactive Services," *Proc. 118th AES Convention*, May 2005.

[3] T.J. Lee et al., "Personalized Audio Broadcasting System through the Terrestrial-DMB System," *Proc. IEEE Int. Conf. Consum. Electron.*, 2007.

[4] ISO/IEC Std. 2002, *Information Technology – Coding of Audio-Visual Objects – Part 1: Coding of Audio-Visual Objects*, ISO/IEC 14496-1.

[5] F. Gallot et al., "iKlax: A New Musical Audio Format for Interactive Music," *Proc. Int. Computer Music Conf.*, Aug. 2008.

[6] ISO/IEC Std. 2005, *Information Technology – Coding of Audio-Visual Objects – Part 12: ISO Base Media File Format*, ISO/IEC 14496-12.

[7] ISO/IEC JTC 1/SC29/WG11 w11158, *Text of ISO/IEC FDIS 23000-12 Interactive Music AF*, MPEG, Feb. 2010.

[8] ISO/IEC JTC 1/SC29/WG11 m16131, *Constraints Specifications for IM AF*, MPEG, Jan. 2009.

[9] ISO/IEC JTC 1/SC29/WG11 m16134, *A Proposal for Constraints Representation Method for IM AF*, MPEG, Jan. 2009.

[10] ISO/IEC JTC 1/SC29/WG11 w11160, *Text of ISO/IEC 23000-12 PDAM 1 Conformance and Reference Software*, MPEG, Feb. 2010.