

Authoring Tool of Musical Slide Show MAF Contents

*Muhammad Syah Houari Sabirin **Munchurl Kim

Laboratory for Multimedia Computing, Communications and Broadcasting,
School of Engineering, Information and Communications University

*houari@icu.ac.kr **mkim@icu.ac.kr

Abstract

The Musical Slide Show MAF, which currently being standardized by MPEG, conveys the concept of combining several established standard technologies in a single file format. It defines the format of packing up MP3 audio data, along with MPEG-7 Simple Metadata Profile and MPEG-21 Digital Item Declaration metadata; with JPEG images and optional text, and synchronizes them all together to create a slideshow of JPEG image data associated to MP3 audio data during the audio playback. The implementation of Musical Slide Show MAF can be a music karaoke file where users can sing along while listening to the music, view the JPEG slideshow and reading the lyrics; or a story-telling file where users can listen to the narrated story by looking at the related illustration slideshow of the story. In this paper we present the tool to producing the Musical Slide Show MAF contents. Regardless the knowledge of user on the MAF file format, the authoring tool simplify the manner of packaging several multimedia contents into single file.

1. Introduction

As the capability of digital media player becomes more sophisticated these days, the need of multimedia-rich content to be played in such device is also becoming somewhat a necessity. Along with that, the manner of authoring a digital multimedia content such as audio or image data is also in-line with the improvement of the player device. Many advanced technology had been or currently being standardized and developed such as the most popular audio format from the Moving Picture Experts Group (MPEG): MPEG-1/2 Layer 3 (MP3) and the Joint Picture Experts Group (JPEG) image data.

However, to create new content that consists of more than one type of multimedia data (e.g. an audio data with an image data in one file), developing new method or standard might take more time and effort. In this sense, MPEG is currently standardized the method of combining

existing multimedia standard into one package called MPEG Multimedia Application Format (MAF). Musical Slide Show MAF, as one of several type of MAFs currently being standardized, defines the method of combining one MP3 audio data with several JPEG images and optional text data; and the synchronization among them.

In this paper, we present the tool of authoring a Musical Slide Show MAF content, which enable users to simply package the aforementioned multimedia data, regardless the user's knowledge on the file format, and creating its respective MPEG-7 and MPEG-21 metadata, into a single file. The structure of this paper is as follows: we firstly explain the architecture of Musical Slide Show MAF in section 2 and its file format section 3. The method of authoring Musical Slide Show MAF following the architecture and file format is explained in section 4. Section 5 presents the result of authoring tool in an application. Finally,

section 6 concludes this paper.

2. MPEG Musical Slide Show MAF

2.1. System Architecture

MPEG is developing a new standard of selecting the existing technologies from its all published standards and combining them in a single standard. This standard, the "Multimedia Application Format" or MAF, is created by selecting readily tested and verified tools taken either from the MPEG standard itself or additional technologies originating from other organizations and put it all together in one file (a MAF file). Utilizing this approach, MPEG is aiming to provide solutions to reduce the effort involved in selecting MPEG (or other organization) technologies and combining them to meet the specific needs of an application area.

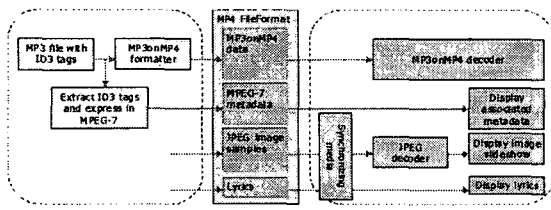


Figure 1. System architecture of MSS MAF

The MPEG Musical Slide Show MAF, which is currently at the working draft status, is an extension to the existing Music Player MAF and Photo Player MAF under development. The Music Player MAF is designed simply to contain MP3 audio data with its associated MPEG-7 metadata and JPEG still image for cover art. The Photo Player MAF combines JPEG still images with MPEG-7 metadata. The Musical Slide Show MAF uses both MP3 audio data with its MPEG-7 metadata and combination of several JPEG still images. As an extension, it allowed the synchronization of JPEG still images and optional text (lyrics) during the MP3 playback using time information synchronization. This section will describe the system architecture of Musical Slide

Show MAF and how can it synchronizes the resources (media) during the MP3 playback.

The system architecture of Musical Slide Show MAF is based on that of Music Player MAF. The extension made to the Music Player MAF is the allocation of JPEG still image and the addition of text / lyrics. As shown in Figure 1, archiving Musical Slide Show MAF is architectonically the same with archiving Music Player MAF (the description of Music Player MAF is can be found in reference [3]). To archiving Musical Slide Show MAF we take MP3 audio file, extract its ID3 tags into MPEG-7, and put it all together with JPEG images and lyrics inside MP4 file format. Moreover, instead of allocating JPEG images as one media, here we can allocate several JPEG image as a collection of samples in one media chunk. To play the Musical Slide Show MAF, it has the same system architecture to that of Music Player MAF, with an addition of synchronization module for displaying JPEG images and lyrics, which can be done prior to JPEG decoding.

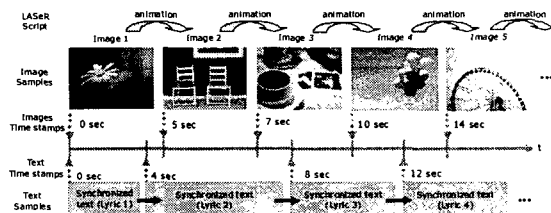


Figure 2. Media synchronization

2.2. Synchronization of Resources

To have the JPEG images and lyrics displayed in the proper timestamp during the MP3 playback, the synchronization of those media can be done by keeping the timeline information for each media. The timeline information is can be stored in movie presentation box in the file format As shown in Figure 2, the synchronization of media is based on the time duration assigned to each media during the authoring process. This is possible since the system architecture allocates the JPEG and lyrics

as two distinguish resource, where each resource can have its own sample table definition. In the figure, we have an example of synchronizing five images and four baits of lyrics with the MP3 with different timestamp

3. File Format

This section presents the structures of Musical Slide Show MAF. As aforementioned, the resources within both MAF's are wrapped in a single MPEG-4 File Format [1] which derives from the specification of ISO Base Media File Format [4]. This format is designed to contain timed media information for a presentation in a flexible, extensible format that facilitates the interchange, management, editing, and presentation of the media. The file is structured in an object-oriented manner to enable simplicity of decomposing file into series of objects that have their own names, sizes, and defined specifications according to its purpose.

The Musical Slide Show MAF was originally proposed to enhance the functionality of Music Player MAF in terms of multiple JPEG images and lyrics with their synchronization to MP3 audio tracks. It is being considered MAF under developments in MPEG. The Musical Slide Show MAF allows for multiple JPEG images and lyrics to be associated in a synchronized way to an MP3 audio track. Also, metadata about the JPEG and MP3 information can be described based on a subset of MPEG-7 MDS schema. Therefore, based on the Musical Slide Show MAF content, the MP3 audio can be play backed with the JPEG images synchronized. Also the information about JPEG images and MP3 audio can also browsed when required.

As shown in Figure 3, the structure of Musical Slide Show MAF shall contain one meta box in moov box to provide the information for the collection of resources in the mdatbox. This

information is a combination of iloc/iinf box and MPEG-21 DID metadata. The DID Resource element refers for the offset and size (length) for resources in the iloc box, so the MAF application can point to the physical location of the resource inside the mdat box based on these information. On the other hand, the associated MPEG-7 metadata of MP3 audio shall be placed in the meta box inside the respective trak box of MP3 data.

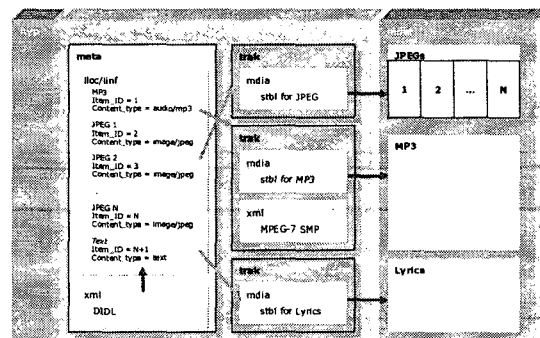


Figure 3. MSS MAF file format

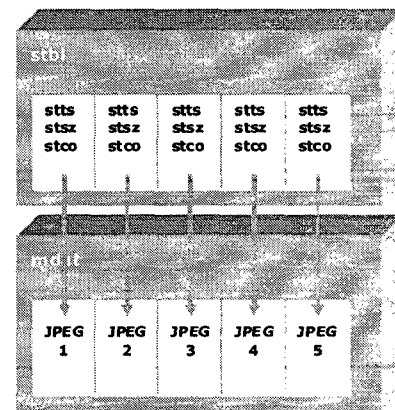


Figure 4. Allocating several JPEGs as a collection of samples

The synchronization information, as described previously in Section 2, is stored in sample table (stbl) box in each track for JPEG samples and lyrics. Figure 4 shows the example of allocating JPEG samples and referring them in sample table box. Each JPEG image is considered as a sample, where the timing information for the slide show

presentation (slide show duration) is stored in *stts* box, the image size is stored in *stsz* box, and the image location (offset) is located in the *stco* box. This allocating method is also applied to the lyrics, where each bait of lyrics can be considered as a sample, consists of a collection of strings.

There are two possible rendering modes of presenting the resources in Musical Slide Show MAF: basic mode and enhanced mode. The rendering modes are respect to the ability of the MAF player to handling the animation effect for the JPEG image transition during the playback [6]. While the basic mode the resources (MP3, JPEG images, and text) are rendered by only using the synchronization information from the *stbl* box, the enhanced mode uses LAsER script to render JPEG images regardless the timing information of JPEG in *stbl* box.

In both basic mode and enhanced mode, it is possible to format the synchronized text (timed text), in respect to the media content, in order to the text really resembles the content. Proposed definitions in formatting text are[7]: supporting all languages, supporting character set and glyphs, defining font and sizes, supporting text rendering, and supporting dynamic highlighting. In the basic mode, the timed text is formatted following the 3GPP definition, while in the enhanced mode the the formatting is defined using LAsER[7].

4. Authoring Tool

Based on aforementioned system architecture and file format of the Musical Slide Show MAF, we can now implement the MAF authoring tool to create a package file consists of one MP3 audio data, several JPEG images, one optional text data, MPEG-7 and MPEG-21 metadata. As shown in Figure 5, the authoring tool basically consists of three modules: resource loader, synchronization module, and writer module [5]. There is also a graphical user interface (GUI) module for users to

control the application.

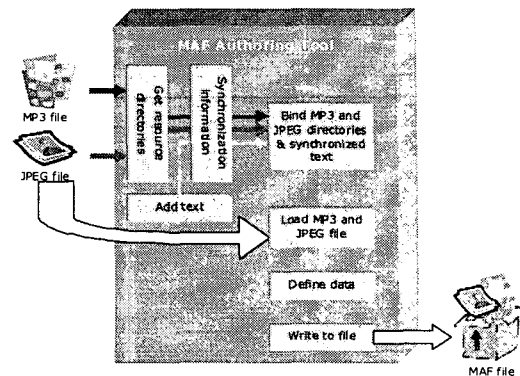


Figure 5. System architecture of the authoring tool

Table 1. Pseudo-code of the authoring tool

```
// add resources
Add_MP3();
  Extract_ID3();
Add_JPEG();

// in GUI
Synchronize_Data();

// begin defining variables for boxes'
parameters
Initialize_parameters();
Get_resource_data(); // duration/size
Define_box_parameters_values () {
  ftyp;
  moov;
  mvhd;
  trak;
  stbl;
  stsd;
  stts;
  ...
meta;
xml;
}
Write_DIDL_metadata();
Accumulate_boxes_size();
Calculate_resources_offsets();

// write the file
Write_data();
Allocate_resources_in_mdat();
```

The resource loader and writer modules are simple module that read and write data from/into

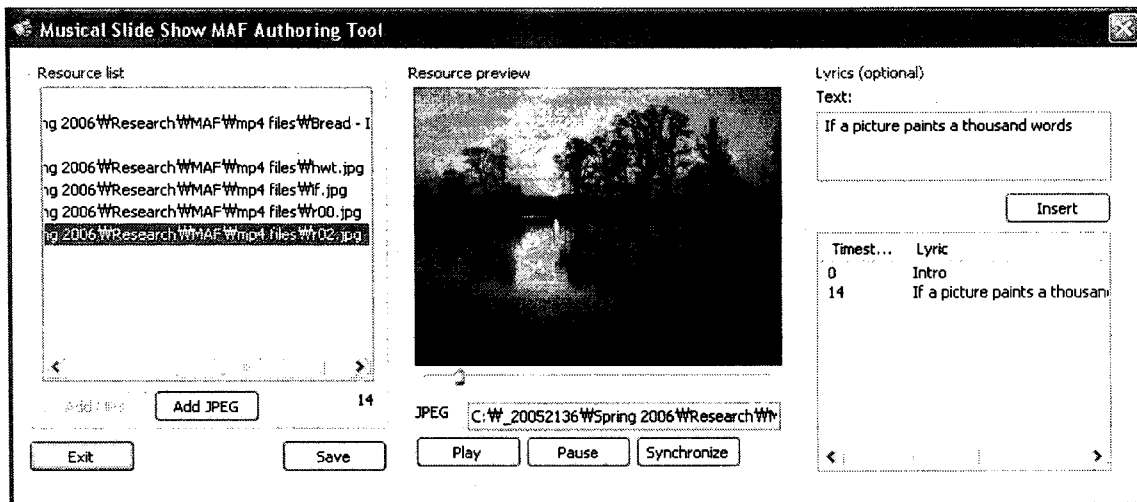


Figure 6. Musical Slide Show MAF authoring tool application

file. Resource loader enables user to load two types of resources via the GUI: one MP3 file and one or more JPEG images, and send the file directory information into the synchronization module. The writer module wrap up the files based on the information established in the synchronization module, and write it to the file.

The implementation of Musical Slide Show MAF definitions is in the synchronization module. It first reads the synchronization information defined by user via the GUI, bind the MP3 and JPEG data, and optional text data into the synchronization information, and finally defines the information respective to the parameters of the boxes in the file format based on the binding and synchronization information. In this paper, we are using basic mode of Musical Slide Show MAF rendering mode and unformatted synchronized/timed text. Therefore, the synchronization of the resources is fully based on the information in the `stbl` box.

In detail, what is actually performed in the synchronization module is as follow:

1. Firstly, it reads the directories of MP3 file and one or more JPEG files. For MP3 file, we extract the ID3 data and write it into MPEG-7 metadata.

2. User defines the synchronization information in the GUI by indicating the timestamp, that is, the time when one JPEG image will be displayed during a period of time while playing the MP3

audio data. Similar method also applied for the optional text data (for example, the lyrics of a song).

3. The modules read the file and synchronization information and define it as a value for the parameters in `meta`, `iloc`, `iinf`, `xml`, and `stbl` boxes; except for the values regarding the position (offsets) of the resources inside the `mdat` box (this will be defined later after the resources have been allocated inside the `mdat` box, in next step). This module will also automatically parse the ID3 metadata from the MP3 file to be stored in `meta` box.

4. The media data then allocated in the `mdat` box. Based on their offsets and sizes, we can define the offset values in the `stbl` box. These information then passed to the write module to finally write the data into a single package as a Musical Slide Show MAF content.

Table 1 shows the pseudo-code for the algorithm used for the authoring tool.

5. Result

The result on the implementation of Musical Slide Show MAF is the authoring tool application of the MAF. As shown in Figure 6, the application consists of three parts: resource list, resource preview, and the lyrics input.

The resource list is used to display the resources (that is, the MP3 audio and JPEG images) that has been loaded to the application. It

shows the list of the directory of the files. The resource preview is used to see the JPEG images loaded into the application, and hear the MP3 audio. In this part, users can perform the synchronization among MP3, JPEG and the text by pressing the "Synchronization" button during the playback of the MP3. The lyrics input area is used to write the text or lyrics into the MAF file.

The following procedure [5] applied to create Musical Slide Show MAF content using the authoring tool:

1. Add one MP3 file using "Add MP3" button.

2. Add a JPEG file using "Add JPEG" button. Repeat this step to add more JPEG files

3. Select the MP3 file from the list

4. Select the first image to be displayed from the Image list, and press "Synchronize" to bind the JPEG to the MP3 file and synchronized it at timestamp 0

5. Synchronizing another JPEG file can be done by either scroll the slider control, or play the audio by pressing "Play" button and press "Pause". Select another JPEG file, and press "Synchronize". Repeat this step to synchronize another image

6. To add optional text/lyrics, it can be done by either scroll the slider control, or play the audio by pressing "Play" button and press "Pause". Write down the text in the provided textbox, and press "Insert". The text just inserted is listed in the table. Repeat this step to add more text/lyrics

7. Finally, press "Save" button to define a name for the new MAF file.

The result of the authoring tool (i.e. the Musical Slide Show MAF content) should be able to be played on any MAF player that following the system architecture and file format specified before. Our prototype of MAF player, available for Microsoft Windows operating system for both PC and PDA shows the results. Figure 7 and 8 show the Musical Slide Show MAF contents being played in different devices.



Figure 7. MSS MAF player for PC

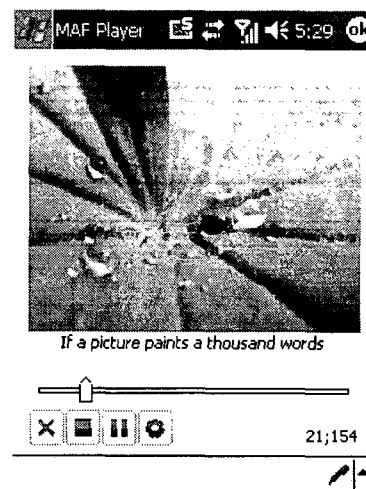


Figure 8. MSS MAF player for PDA

6. Conclusions

In this paper we presented the method of implementing the Musical Slide Show MAF system architecture and file format into an authoring tool application. The authoring tool allows user to create Musical Slide Show MAF contents without any knowledge on the MAF file format. Users only have to know what MP3 data they want to hear, want JPEG images they want to see, and in what time point during the MP3 playback they want to see the images and the lyrics/text.

There are many improvements still can be done for the authoring tool, such as improvements on the user interface, synchronizing method, adding animation as specified by the enhanced rendering mode, timed text formatting, etc. The basic

technique of authoring MAF content mentioned in this paper can be expanded to have an application that is suitable to be commercialized.

7. References

1. ISO/IEC JTC 1/SC 29/ WG 11 (MPEG), Working Draft 1.0, Musical Slide Show MAF, ISO/IEC JTC 1/SC 29/WG 11/N8131, Montreux, Switzerland, April 2006
2. ISO/IEC JTC 1/SC 29/WG 11 (MPEG), White Paper on MPEG-A, ISO/IEC JTC 1/SC 29/WG 11 MPEG2005/N7068, Busan, Korea, April 2005
3. ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Defect Report of ISO/IEC 23000-2 Music Player MAF, ISO/IEC JTC 1/SC 29/WG 11 N7946, Bangkok, Thailand, January 2006
4. ISO/IEC 14496-12: Information technology – Coding of audio-visual objects – Part 12: ISO base media file format, ISO/IEC 2005
5. ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Contribution to Reference Software of ISO/IEC 23000-4: MPEG Musical Slideshow Application Format, ISO/IEC JTC 1/SC 29/WG 11 M13673, Klagenfurt, Austria, July 2006
6. ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Text of ISO/IEC CD 23000-4 Musical Slide Show MAF, ISO/IEC JTC 1/SC 29/WG 11/N8397, Klagenfurt, Austria, July 2006.
7. ISO/IEC JTC 1/SC 29/WG 11 (MPEG), Proposed timed text format for Musical Slide Show MAF, ISO/IEC JTC 1/SC 29/WG 11/M14001, Hangzhou, China, October 2006.