

# 다중 영상 및 텍스트 동기화를 고려한 Music Player MAF 의 확장 포맷 연구

양찬석\*, 임정연\*\*, 김문철\*\*\*  
한국정보통신대학교

## A study on Extensions to Music Player MAF for Multiple JPEG images and Text data with Synchronization

Chansuk Yang\*, Jeongyeon Lim\*\*, Munchurl Kim \*\*\*  
Information Communication University

E-mail : \*hewu@icu.ac.kr, \*\* jylim @ icu.ac.kr, \*\*\* mkim @ icu.ac.kr

### Abstract

The Music Player MAF Player Format of ISO/IEC 23000-2 FDIS consists of MP3 data, MPEG-7 metadata and one optional JPEG image data based on MPEG-4 File Format. However, the current Music Player MAF format does not allow multiple JPEG image data or timed text data. It is helpful to use timed text data and multiple JPEG images in the various multimedia applications. For example, listening material for the foreign language needs an additional book which has text and images, the audio contents which can get image and text data can be helpful to understand the whole story and situations well. In this paper, we propose the detailed file structure in conjunction with MPEG-4 File Format in order to improve the functionalities, which carry multiple image data and text data with synchronization information between MP3 data and other resources.

### I. 서론

개인용 미디어 단말로서 MP3 플레이어는 음악 파일을 감상하기 위한 중요한 휴대 가전이 되었으며, 이미지 정보나 텍스트 정보를 표현할 수 있는 기능들이 추가 되고 있다. 이처럼 변화하는 멀티미디어 서비스에 발맞추어 MPEG 에서는 Multimedia Application Format (MAF) (ISO/IEC 23000)에 대한 표준화 활동을 진행하고 있다. 그 첫 번째 결과물로, 2005 년 10 월 국제표준(ISO: International Standard) 규격이 완성되었다 [1]. Music Player MAF 는 MPEG-1 Audio Layer III(MP3) 을 중심으로 MPEG-7 MDS 메타데이터 정보를 결합한 파일 포맷으로, 풍부한 메타데이터 정보와 더불어 선택적으로 하나의 JPEG 이미지를 포함한 형태로 이루어져 있다. 그러나, 현재의 Music Player MAF 파일 포맷은 단지 한 장의 영상 데이터와 ID3 에 대한 메타데이터만을 포함시킴으로써 다양한 멀티미디어

서비스를 제공하기에는 많은 한계점을 갖고 있다. 그래서 본 논문에서는 Music Player MAF 에 제안된 세가지 파일 포맷에 관해 기술 하고, Music Player MAF 기능을 확장 하여 복수의 JPEG 이미지와 자막 정보를 추가하여 MP3 오디오 데이터를 보다 확장된 형태로 활용 할 수 있는 Music Player MAF 의 확장 포맷에 대한 연구 결과를 제시하고, 관련 현재의 MPEG 표준화 기술 동향에 대해 기술한다.

### II. Music Player MAF 파일 포맷

Music Player MAF 파일 포맷은 ISO/IEC 2003 MPEG-A 의 서브 파트의 하나로, MP3 오디오 파일과 더불어, MP3 에 포함된 노래 제목이나, 연주자에 관한 ID3 정보를 MPEG-7 메타데이터로 변환 하고 필요할 경우 추가적인 메타데이터 정보를 포함 할 수 있도록 구성된 파일 포맷이다.



그림 1 Music Player MAF 의 기본 내용

Music Player MAF 파일 포맷은 MPEG-4 파일 포맷을 기반으로 정의 되어있다. MPEG-4 파일 포맷은 손쉽게 여러 종류의 미디어를 포함하고 관리할 수 있도록 객체 지향적인 BOX 요소들의 집합으로 구성되어 있다. 각각의 BOX 들은 여러 필드 값들로 이루어져있으며, 각기 독립적으로 해석되어 고유한 기능을 수행할 수 있다.

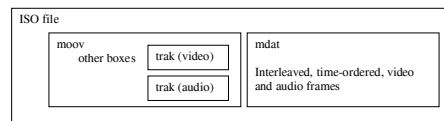


그림 2 MPEG-4 파일 타입의 기본 구조

본 연구는 정보통신부 및 정보통신연구진흥원의 디지털미디어 연구소 지원사업의 연구결과로 수행되었음

현재의 Music Player MAF 파일 포맷은 세가지 형태로 정의 되어 있다. 첫째, MPEG-4 파일 타입을 기반으로,

하나의 MP3 오디오 트랙과 MPEG-7 메타데이터 포함하는 포맷이 정의 되어 있다 (Type 1 Music Player MAF). 둘째, MPEG-21 파일 타입을 기반으로 하나의 MP3 오디오 트랙과 MPEG-21 메타데이터를 포함하는 파일 포맷이 정의 되어 있다 (Type 2 Music Player MAF). 마지막으로 MPEG-21 파일 타입을 기반으로 복수의 트랙을 포함하는 형태의 파일 구조가 정의되어 있다 (Type 3 Music Player MAF) [1].

### III. Music Player MAF 파일 포맷의 확장

본 절에서는 사용자 시나리오를 통해 현재 Music Player MAF 의 기능상의 제한 점을 분석하고, 기존의 세가지 형태의 Music Player MAF 파일 포맷에 대해 각각의 기능을 확장시키기 위한 방안에 대해 설명 한다.

#### 3.1 사용자 시나리오

현재 MP3 플레이어는 단순히 음악 파일을 재생하는데 그치지 않고, 조그만 액정 화면을 통해 이미지 뷰어나 텍스트 브라우징 기능들을 기본으로 제공하고 있다. 이에 따라, MP3 음악 파일과 이미지 혹은 텍스트 정보를 결합하여 표현 할 수 있는 파일 포맷에 대한 사용자들의 요구가 증대되고 있다. 예를 들어, 영어 학습 교재로서 MP3 파일을 사용하고 있으나, 실제 재생되는 문장이나 단어의 철자나 회화 내용의 배경이 되는 그림 등은 표현 되지 않아, 항상 별도의 책자를 소지해야 하는 단점이 있었다.

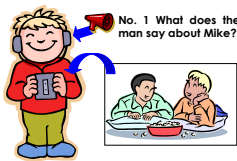


그림 3 Music Player MAF 를 이용한 영어 교재

또한, 들을 수 있는 뉴스나 이야기 책 같은 콘텐츠에 적절한 그림을 관련 시킬 수 있다면, 사용자들이 해당 콘텐츠를 훨씬 쉽게 이해할 수 있을 것이다. 일반적인 음악 파일을 들을 경우에도, 해당 음악과 관련된 가수의 이미지나 해당 앨범의 표지 그림 등을 함께 표현해 줄 수 있을 것이다. 현재의 Music Player MAF 파일 포맷은 복수의 JPEG 이미지나 텍스트 정보를 MP3 오디오와 동기화 시키는 방안을 포함하고 있지 않다. 이에, Music Player MAF 의 파일 구조를 확장하여 다 수의 이미지와 자막 정보를 MP3 데이터와 동기화 시켜 포함 하는 방법을 제안한다.

#### 3.2 Music Player MAF 파일 구조의 확장

기존의 Music Player MAF 파일 구조를 기본으로, 복수의 JPEG 이미지 데이터와 텍스트 데이터를 MP3 오디오 데이터와 동기화 할 수 있는 파일 구조는 다음과 같다. 첫째, 포함되는 모든 미디어의 위치 정보와 크기 정보는 meta 박스 안의 iloc 박스를 이용하여 표현한다. 둘째, 시간 정보가 필요한 미디어들은 각각의 trak 박스를 통해 선언한 후, 미디어 간의 동기화 정보는 trak 박스 안의 stb1 박스나 edts 박스 안에 필요한 시간 정보를 추가한다.

#### 3.2.1 Type 1 Music Player MAF 의 확장

Type1 Music Player MAF 는 다음의 방법을 통해 확장 가능하다. 기존의 파일 포맷에 독립적인 하나의 meta 박스를 추가 한 후, 해당 meta 박스 안의 iloc/iinf 박스를 통해 포함된 미디어들의 파일 상의 위치 와 크기, 미디어 타입 정보 등을 표현한다. 시간정보를 포함하는 미디어들은 각각 별도의 trak 박스를 통해 관리하며, trak\_ID 와 item\_ID 를 연결 지어 하나의 미디어를 표현한다. 또한 다양한 Metadata 를 포함 하기 위해, meta 박스 내부의 xml 박스를 사용 한다.

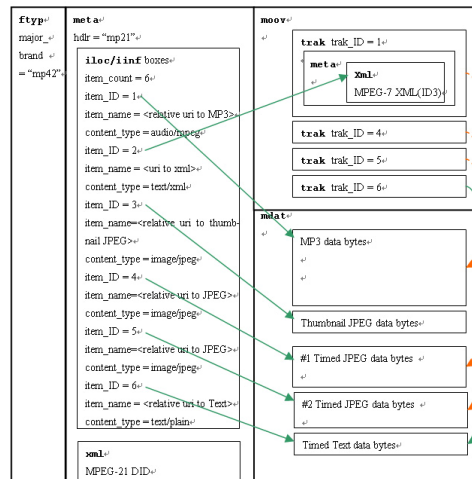


그림 4 Type1 Music Player MAF 의 확장된 구조

그림 4 는 Type 1 Music Player MAF 의 확장 구조의 한 예를 나낸다. 미디어 구성 요소는 meta 박스를 통해 총 6 개의 아이টে으로 표현되어 있다. 각 아이টে은 하나의 리소스에 해당되며 mdat 박스의 리소스를 가리킨다. MP3 오디오 데이터는 moov 박스의 “trak\_ID=1”에 의해 기술되며, 두 장의 JPEG 이미지 데이터는 “trak\_ID=4”, “trak\_ID=5”에 의해 동기화를 위한 타임 정보를 가지고 있다. 텍스트 데이터는 “trak\_ID=6”에 의해 MP3 오디오 데이터에 대한 동기화 정보를 포함하고 있다.

#### 3.2.2 Type2 Music Player MAF 의 확장

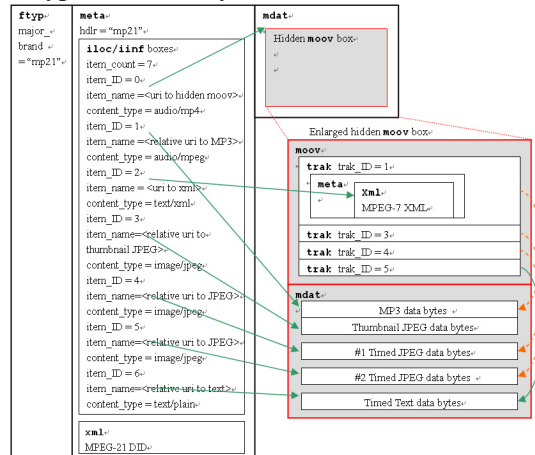


그림 5 Type2 Music Player MAF 의 확장된 구조

Type 2 Music Player MAF 는 하나의 MP3 오디오 트랙데이터와 한 장의 JPEG 이미지 데이터 그리고 ID-3 에 대응되는 MPEG-7 MDS 메타데이터의 구성 정보를 MPEG-21 DID 메타데이터로 구성된 파일 구조를 정의한다. 확장된 Type2 Music Player MAF 는 기존의 Type2 Music Player MAF 와 전체적인 파일 구조는 동일 하나, 몇 가지의 차이점이 있다. 우선 mdat 박스 안에 감추어진 moov 박스 내부에 복수의 JPEG 이미지와 자막 정보를 추가함에 따라, 그에 상응하는 trak 박스들을 추가 한다. 또한 이렇게 추가된 trak 박스들을 참조하기 위하여, 기존의 iloc 박스 안에 하나의 아이템을 추가하여 숨어있는 moov 박스의 실제 파일 상에서의 위치를 표현하도록 하였다.

### 3.2.3 Type3 Music Player MAF 의 확장

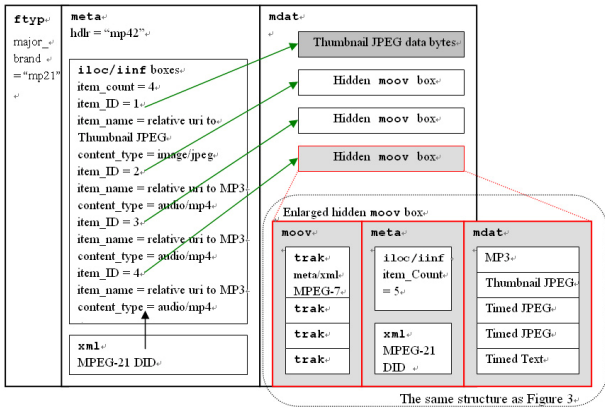


그림 6 Type3 Music Player MAF 의 확장된 구조

Type3 Music Player MAF 는 하나 이상의 MP3 오디오 파일을 포함하기 위해 고안된 포맷으로 큰 변화 없이 다음의 방법을 통해 기능을 확장 할 수 있다. 우선 포함된 여러 MP3 를 대표하기 위한 JPEG 이미지를 추가 한다. Type3 Music Player MAF 가 개별적인 곡이 아닌 하나의 앨범을 표현하기 위한 파일 포맷이라 생각할 때 이 JPEG 이미지는 앨범 자켓의 이미지와 같은 정보를 표현하기 위하여 사용될 수 있다. 그리고, iloc 박스 내의 item 들이 각각의 개별적인 미디어들을 일일이 가리키는 대신, 각각의 숨겨진 moov 박스들을 가리키게 한다. 이 때, 숨은 moov 박스는 앞서 제안된 Type1 Music Player MAF 확장 구조를 사용하며, 해당 moov 박스내의 iloc 박스를 이용하여, 포함되어있는 실제 미디어들의 위치 정보를 표현한다. 이것은 iloc 박스 내에는 아이템간의 계층적인 관계를 표현할 수 없기 때문에, 하나의 JPEG 이미지가 단순히 하나의 item 으로 표현 될 경우, 어떠한 MP3 오디오에 종속되어 있는 미디어 인지 표현 할 수 없기 때문이다.

## IV. 확장된 Music Player MAF 의 기능

Music Player MAF 파일을 재생하기 위해서는, 포함되어있는 각각의 미디어들의 위치와 크기와 더불어, 시간 정보가 포함된 미디어의 경우 해당 동기화 정보를 알아내야 한다. 이번 절에서는 확장된 Music Player MAF 파일 포맷이 어떠한 방법으로 각 미디어들의 위치

정보와 시간 정보를 표현하고 이를 활용하는지에 대해 설명한다.

### 4.1 확장 Music Player MAF 의 미디어 위치 표현

본 논문에서 제안된 확장된 Music Player MAF 에서는 미디어의 위치를 표현하기 위해 meta 박스 안에 존재하는 iloc 박스를 사용한다.

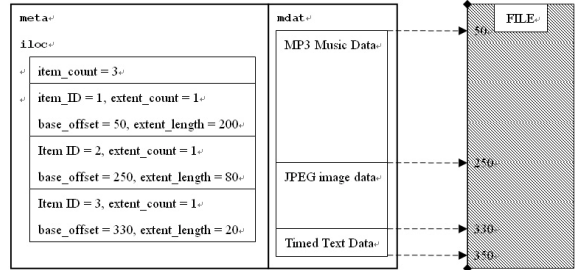


그림 7 확장된 Music Player MAF 의 미디어 위치 표현

그림 7 은 iloc 박스를 이용하여 실제 미디어 위치를 표현하는 과정을 보여준다. item\_count 필드는 포함된 미디어의 개수를 나타낸다. item\_ID 필드는 해당 아이템을 상징하는 아이디를 나타낸다. extent\_count 필드는 해당 미디어가 몇 개의 조각으로 나누어져 파일 상에 위치하고 있는지를 나타내며, Music Player MAF 에서는 미디어의 크기가 상대적으로 작기 때문에, 대부분 하나의 조각만이 존재 한다. 그러므로 base\_offset 필드는 실제 미디어의 파일 상의 위치를 표현하며, 해당 미디어의 크기는 extent\_length 필드에 나타나게 된다.

### 4.2 확장 Music Player MAF 의 동기화 정보 표현

각 미디어의 동기화 정보를 표현하기 위해서는 trak 박스 내의 elst 박스나 stbl 박스가 사용된다. 우선 다수의 JPEG 이미지 데이터와 MP3 오디오간의 동기화 정보는 elst 박스를 통해 표현한다. elst 박스는 네 종류의 필드를 가지며, entry\_count 는 elst 박스에 포함된 동기화 정보의 개수를, segment\_duration 은 해당 미디어의 플레이 시간을, media\_time 은 해당 미디어의 시작 시간을 표현한다.

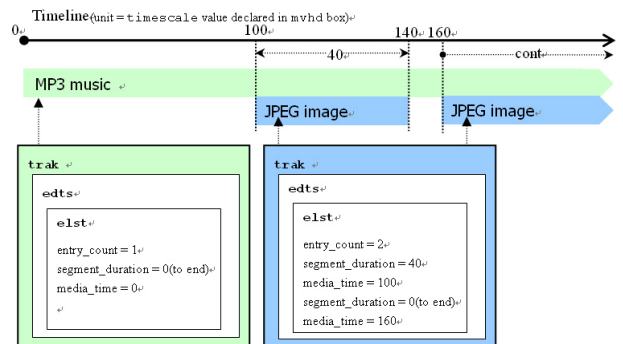


그림 8 확장된 Music Player MAF의 JPEG 시간정보 표현

그림 8 은 JPEG 이미지 데이터의 MP3 오디오 트랙에 동기화 하기 위해 elst 가 실제로 어떻게

사용되고 있는지를 보여준다. 그림 8 의 JPEG 이미지는 elst 박스 안에 두 개의 시간 정보를 갖고 있으며 이에 따라 해당 Music Player MAF 파일이 재생되는 중 100~140 시간 단위 동안 재생된 후 160 시간 단위에서부터 파일 재생이 종료될 때까지 다시 한번 화면 상에 출력된다.

자막파일의 경우 각각의 문장들은 서로 다른 시간에 화면 상에 출력되게 된다. MPEG-4 파일 포맷은 이렇게 최소한의 시간 단위를 갖는 미디어의 집합을 sample 로 표현 한다. 즉, 자막 파일은 여러 개의 sample 단위로 이루어진 미디어라고 생각 할 수 있으며, 이러한 sample 미디어의 동기화 정보를 표현하기 위해서는 stb1 박스가 사용된다.

그림 9 는 실제로 stb1 박스를 이용하여 자막파일의 동기화 정보를 표현하는 한 예를 나타낸다. 하나의 자막 파일이 세 개의 sample 로 구성되어 있으며, stts 박스를 통해 각 자막 파일이 재생되는 시간과 재생 길이를 표현하며, stsz 박스와 stco 박스를 이용하여 해당 sample 의 실제 위치를 가리키고 있다.

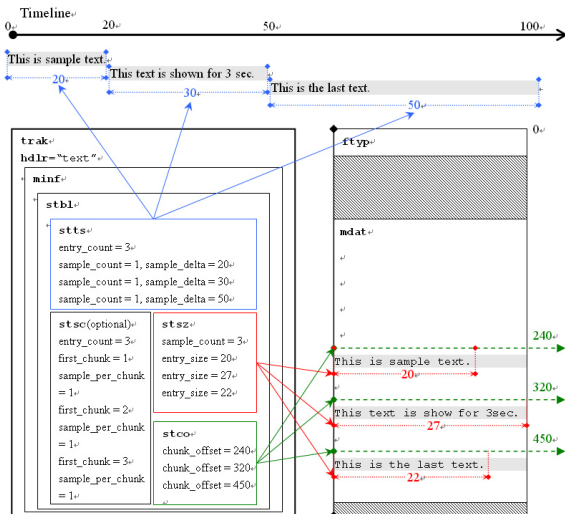


그림 9 확장된 Music Player MAF 의 자막 시간 정보 표현

### V. 실험 결과

본 논문에서 제안된 파일 구조의 실제 동작을 보이기 위해, Type 1 Music Player MAF 확장 파일을 재생하기 위한 Music Player MAF 재생기를 구현하였다. 그림 10 은 실제 구현된 재생기를 통해 재생되는 각각의 콘텐츠를 보여 준다.

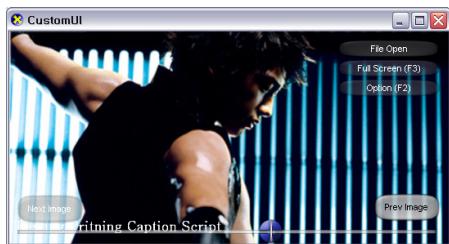


그림 10 Music Player MAF 를 이용한 포토 뮤직 콘텐츠

그림 10 에서 MP3 오디오 트랙에 동기화 되어 각 이미지 데이터가 해당 시점에 동기화 되어 디스플레이

된다.

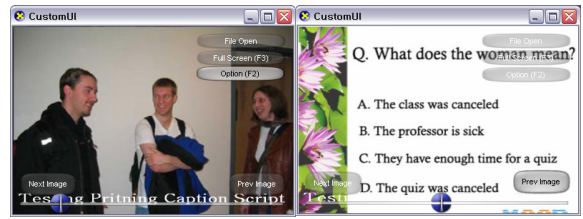


그림 11 Music Player MAF 확장 포맷을 이용한 영어 학습 콘텐츠의 재생

그림 11 은 영어 Music Player MAF 확장 포맷에 기반한 영어 학습 교재 콘텐츠의 재생 예를 나타낸다. 그림 11 의 왼쪽 그림은 설정된 장면에서의 대화 모습을 JPEG 이미지로 표현하고 있으며, 대화 내용은 동기화된 형태의 텍스트 데이터로 화면에 디스플레이 되고 있다. 한편, 그림 10 의 오른쪽 그림은 왼쪽의 대화 내용에 대한 문제를 JPEG 이미지 데이터 형태로 표현하고 있다.

### VI. 결론

ISO/IEC 23000-2 Music Player MAF 는 MP3 dheldh 데이터와, ID-3 그리고 한 장의 JPEG 이미지를 메타데이터로한 파일 포맷을 정의 한다. 이러한 현재의 규격은 제한된 미디어 수용으로 인해 개인형 미디어 단말에서 보다 풍부한 멀티미디어 응용을 많은 제약점을 가지고 있다. 따라서 본 논문에서는 현재의 Music Player MAF 규격을 확장하는 구조를 제안하여, 다수의 JPEG 이미지 데이터와 텍스트를 MP3 에 동기화하여 개인형 미디어 단말에서 보다 풍부한 멀티미디어 응용을 가능하게 하였다. 향후, Music Player MAF 확장 포맷에 기반한 콘텐츠를 저작할 수 있는 저작툴 연구와 재생기를 구현할 예정이다.

### 참고문헌

- [1] FDIS, MPEG Music Player Application Format, ISO/IEC JTC1/SC 29/WG 11/N7156, Busan, Korea, April, 2005.
- [2] ISO Base Media file format incorporating amendment 1, ISO/IEC JTC2/SC29/WG11/N6596-B, Redmond, USA, June, 2004.
- [3] Fernando Pereira and Touradj Ebrahimi, "The MPEG-4 book," IMSC press, 2002.
- [4] Jeongyeon Lim and Munchurl Kim, "Synchronization of Multiple JPEG data to MP3 tracks in Music Player MAF Player Format," ISO/IEC JTC2/SC29/WG11/MPEG2005/M12396, Poznan, Poland, 2005.
- [5] Chansuk Yang, Jeongyeon Lim and Munchurl Kim, "Extensions to Music Player MAF Player Format for Multiple JPEG images and Text data with Synchronizations to MP3 data," ISO/IEC JTC1/SC29/WG11/MPEG05/M12589, Nice, France, 2005.