# New Interactive TV Service Model based on the MPEG-4 System

Jongho Kim and Jechang Jeong
Division of Electrical and Computer Engineering, College of Engineering,
Hanyang University,
17 Haengdang-dong, Seongdong-gu, Seoul, Korea
Tel. +82-2-2292-6457, Fax.: +82-2-2281-9912
e-mail : {angel, jjeong}@ece.hanyang.ac.kr

**Abstract:** In this paper, a new interactive TV service model is proposed. The MPEG-4 system is specified for composing and managing various object streams including user interactions. The data broadcasting model supporting user interactions is designed using MPEG-4 system in our proposal. We evaluate possibility of proposed service model using simulation player. This player supports MPEG-2 TS which contains MPEG-2 video and AC-3 audio streams as a main service and MPEG-4 system data as interactive services as welle as user specific EPG information, and XML data, etc as supplemetary services. The player also supports a multi-channel environment. The synchronization between audio and visual data is achieved by DTS and PTS in TS.

## 1. Introduction

With the beginning of digital broadcasting service, necessity of interactive service has been increased. Interactive broadcasting service is very attractive to both service providers and users. For user's side, it offers auxiliary information to users quickly and simply; and for service provider's side, it maximizes efficiency of service. Current interactive TV services mainly offer ability to access Web pages to obtain information related to TV programs. In these services, a web page is displayed on TV screen along with the main picture, and the user can send or receive information about the current TV program through this web page. Inherently these services are the same as web services in computer world. In these types of services, internet connection is required to deliver the web contents and transmit the user's interactions. In fact, interactive TV service should enable users to interact with the TV program by exerting some control over the contents of the program being viewed.

On the other hand, ISO/IEC 14496 (MPEG-4) was standardized for various applications of audio/visual data. Especially, ISO/IEC 14496-1 (MPEG-4 System) [1] provides tools for interactive TV services. This standard introduces scene description techniques used in computer graphics, and offers interactivity to objects of the scene. In this paper, we propose a new interactive TV service model based on the ATSC system and MPEG-4 System standard. By implementation a digital television receiver model based on the ATSC and MPEG-4 System standard, we verify a hardware-based receiver model with low cost, and offer a flexible test resource that can convert its platform according to change of interactive service models.

The rest of this paper is organized as follows. In section 2, we give an outline of the structure of this player and describe standards as well as specifications related to this player. In section 3, we describe our proposal to realize a new interactive TV service and implementation issues. In section 4, we present the result of our method for interactive broadcasting service. Conclusions and further research issues are presented in section 5.

## 2. Structure of the Player

In this paper, using specifications and standards related to interactive TV service, a player is designed and implemented in order to simulate an interactive TV receiver system. This player is modeled on a home digital television receiver or a digital set-top box. Both European and American standards for digital TV (DVB and ATSC) [2, 3] employ the ISO/IEC 13818-1 (MPEG-2 System) [4] for the packetization and multiplexing of audio, visual, and data streams. Therefore it is possible to play an MPEG-2 TS (Transport Stream) that contains audio/visual data for a main service and the other data for supplementary services.

For a main service, the visual stream is coded using ISO/IEC 13818-2 (MPEG-2 video) [5], and the audio stream is coded using Dolby AC-3. For supplementary services, we use the MPEG-4 System to transfer the audio/visual scene related to the main picture; and we use XML to deliver the other information like news, advertisements, and weather information, etc. By adoption of PSIP that is defined in ATSC, this player provides channel information and EPG data to users, and in addition, there are simple applications using JAVA Xlet in TS private section. The basic block diagram is shown in Fig. 1.

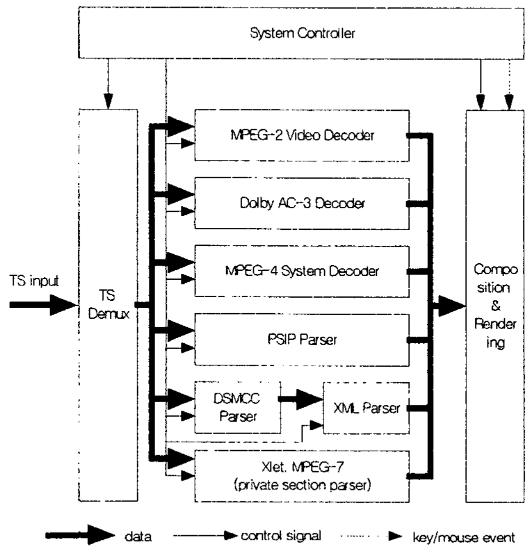


Figure 1. Structure of proposed player

## 2.1 MPEG-2 Transport Stream

The MPEG-2 System addresses combining of one or more elementary streams of audio and video as well as other data, into a single or multiple streams which are suitable for storage (PS; Program Stream) or transmission (TS; Transport Stream).

TS is packetized by 188 bytes fixed for the advantages of effective transmission and error resilience, etc. Each packet begins with a 4 bytes prefix, which contains a 13 bit PID (Packet ID), will be described in next paragraph. Packet header is followed by adaptation field which contains PCR (Program Clock Reference) for synchronization of the system, and payload which contains coded media data. Since TS includes one or more programs, PSI (Program Specific Information) tables that enable demultiplexing of programs by decoders are carried in the transport stream. There are four PSI tables:

- Program Association Table (PAT)
- Program Map Table (PMT)
- Conditional Access Table (CAT)
- Network Information Table (NIT)

These tables contain the necessary and sufficient information to demultiplex and present programs. The PAT provides the correspondence between a program number and the PID value of the TS packets which carry the program definition. The PMT specifies, among other information, which PIDs, and therefore which elementary streams are associated to form each program. The CAT shall be present if scrambling is employed. The NIT is optional and its contents are not specified by MPEG-2 system. TS syntax diagram is showm in Fig. 2.

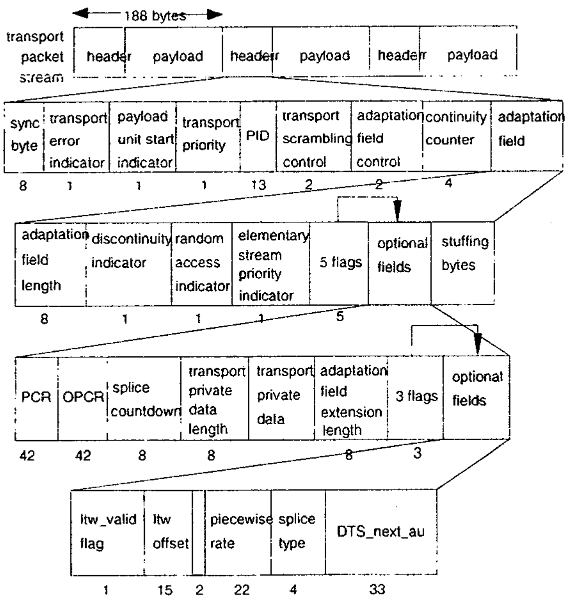


Figure 2. Syntax diagram of MPEG-2 TS

## 2.2 MPEG-4 System

An MPEG-4 scene consists of audio/visual media objects. Data for each streaming media objects are conveyed within one or more elementary streams. A major task of MPEG-4 system architecture is to describe relations of various audio/visual components of an MPEG-4 scene. Fig. 3 shows the MPEG-4 terminal architecture, which presents the relationship and hierarchical structure of each component defined in the MPEG-4 standard. A special feature compared with existing standards, is the description scheme. The description has two levels. On a structural level, it must be known how the media objects are arranged in space and time. This scene description is accomplished by binary format for scene (BIFS), which has its roots in VRML. This scheme provides the interactivity and we propose the system using this point [1, 6]. On a lower level, it is only of interest how various streams that contain the compressed MPEG-4 data related each other, how they are configured and synchronized, and where they are located. In addition, some information about the content, including intellectual property information, is desirable. This process of description is accomplished by the media object description (OD) framework [1, 7]. These special features and object-based processing make possible the interaction with users, which is proposed as a new broadcasting service model in this paper.

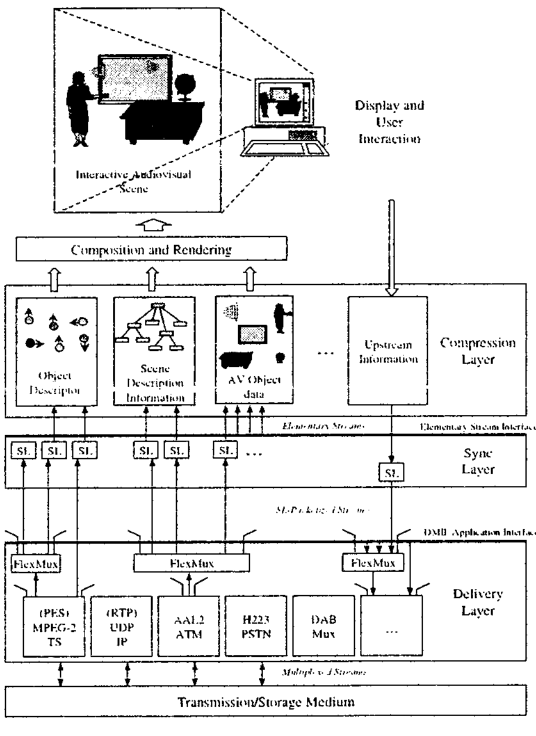


Figure 3. Terminal architecture of MPEG-4

## 2.3 Other Data

In this paper, some supplementary data are contained for various and flexible simulation of actual digital television services.

**PSIP data**: offer EPG (Electronic Program Guide) data to users. These are based on the ATSC specification [8]. In addition, we propose user specific EPG information can be stored in the receiver. As shown in Fig. 4, preferences of users are databased by age, genre, and time, etc. This idea can provide the model of intelligent data broadcasting service.

**XML data**: offer various kinds of information such as news, movies, advertisements, stock market information, and weather information, etc. Because of asynchronous non-streaming data, these are transmitted through DSM-CC data carousel protocol.

**Xlet data**: offer simple JAVA applications. Simply, for simulation of various situations, Xlet data provide a simple game and a calendar, etc. These are transmitted to the player in TS private section.

As well as these data, the e-mail and web browsing service are also provided to user. And in the near future, MPEG-7 data can be used for rich services.

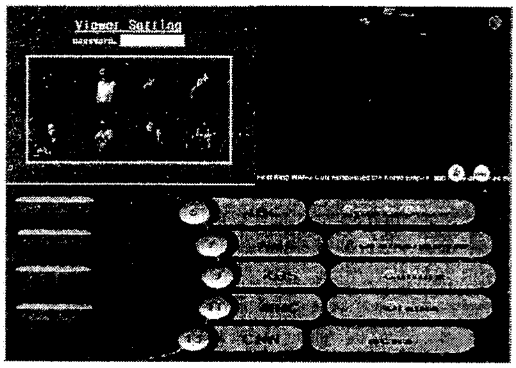

Figure 4. An example of user specific EPG information

## 3. Implementation Issues

For simulating our proposed service model, a player which can play all of the data described in previous section is implemented. All processing of the player is managed by system controller module. Buffers are used for data exchange between each module, and all modules are synchronized by the system clock. Since this player is designed for the digital TV receiver using software, each module is operated by threads. For composition of many pictures such as MPEG-2 video picture, MPEG-4 system pictures, and supplementary pictures, the overlay techniques of Microsoft DirectX are used. In addition, appropriate PAT and PMT parser in order to play multi-channel TS is also implemented. The MPEG-4 system decoder is implemented using MPEG-4 system refernce software, IM1-2D ver.5.0, that can play a MPEG-4 system data supporting interaction with users. For synchronization of audio/visual data, time stamps (DTS; Decoding Time Stamp, PTS; Presentation Time Stamp) are used and implemented as Fig. 5. These time stamps are specified as follwings (1) and (2):

$$PTS(k) = (\frac{((system\_clock\_frequency \times tp_n(k))}{300})\%2^{33} \quad (1)$$

where $tp_n(k)$ is the presentation time of presentation unit Pn(k), Pn(k) is the k-th presentation unit in elementary stream n.

$$DTS(j) = (\frac{((system\_clock\_frequency \times td_n(j))}{300})\%2^{33} \quad (2)$$

where $td_n(j)$ is the decoding time of access unit An(j), An(j) is the j-th access unit in elementary stream n.

**scenario**: For simulation, there are three channels; one is the Korean historical drama, another is a movie, and the other is a car CF. These are coded by MPEG-2 video and AC-3 audio for main pictures. Each channel has the MPEG-4 sub-pictures related to main pictures and activated by user interaction. XML data provide following information: news, stock market information, weather information, movie, play, and advertisement and these are also activated by user interaction. For presenting user specific EPG information, simply, html files are used. Applications using JAVA Xlet are simple ping-pong game, calendar, and calculator.

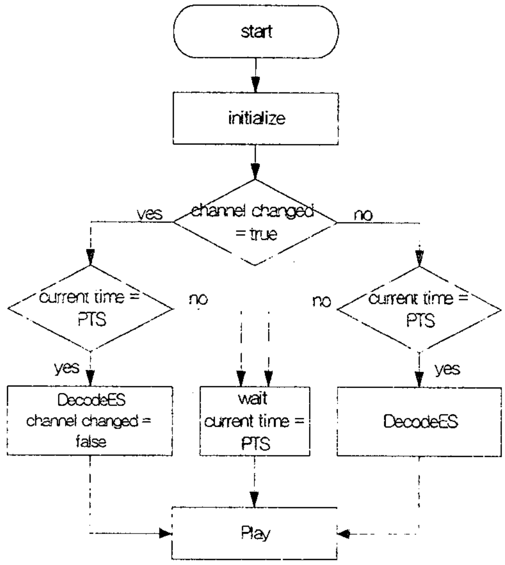


Figure 5. Decoding process using time stamps

**MPEG-4 system data**: In this paper, the key point is interactive data broadcasting using MPEG-4 system data. We use JPEG images related to main pictures such as the characters, synopses, scenerics, and other data even though the MPEG-4 system data can contain various types of objects. Each JPEG objects are presented on the screen by user interaction spported by BIFS of MPEG-4 system. When channel changed, MPEG-4 system data is also changed to the corresponding changed main picture. These results are shown in the next section.

## 4. Simulation Results

In this section, results of the interactive TV simulation player implemented to support multi-channel TS are presented. The main picture has the standard definition resolution and MPEG-4 system data has the same size. All kinds of pictures are overlaid. Only main picture's audio data is presented. When channel changed, the player checks the PIDs and corresponding data change is occurred. Except the main pictures, all of the supplementary data is activated by user interaction. Because of software implementation, synchronization between audio and video is a little out of accord, so that we plan to find appropriate method as will be described in next section. Some resultant capture images are shown below.

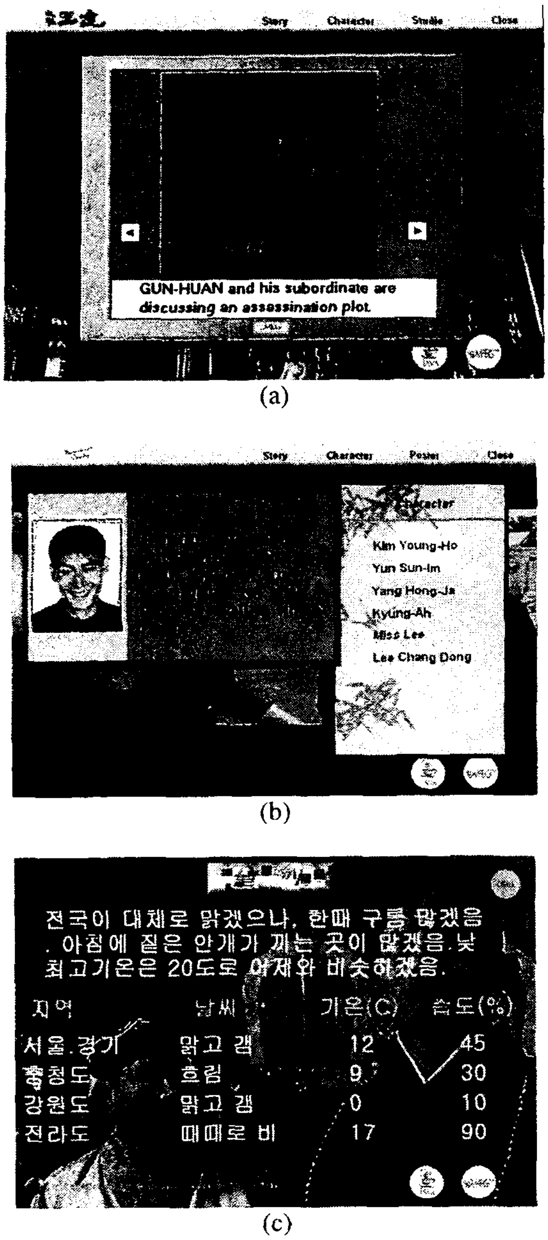


Figure 6. Typical scenes that consist of main and supplementary pictures

(a) Korean historical drama and related data (ch. 1)
(b) A movie (Peppermint) and related data (ch. 2)
(c) A scene that XML data (weather information) is presented (ch. 1)

## 5. Conclusions and Further reaserch issues

In this paper, a new interactive broadcasting service model using the MPEG-4 System is proposed. Specially, software implementation provides advantages of the flexibility in designing various applications with low cost. Also, we verify that the MPEG-4 System is a powerful tool for interactivity in broadcasting environment instead of html. In addition, for various situations user specific EPG data, information data using XML, and JAVA Xlet applications are also presented.

The model introduced in this paper provides the framework for a new interactive TV system. In our future research, we plan to stand more intelligent and almost automatic user specific data broadcasting model and more appropriate synchronization method.

## References

[1] ISO/IEC 14496-1, "Coding of audio-visual objects: Systems," International Standard, 1998.
[2] ETS 300 802, "Network-independent protocols for DVB interactive services," ETSI, 1997.
[3] ATSC Document A/53, "ATSC Digital Television Standard," ATSC, 1995.
[4] ISO/IEC 13818-1, "Generic coding of moving pictures and associated audio: Systems," International Standard, 1994.
[5] ] ISO/IEC 13818-2, "Generic coding of moving pictures and associated audio: Video," International Standard, 1994.
[6] Julien Signes, et al, "MPEG-4's binary format for scene description," *Signal Processing: Image Comm.* Vol. 15, pp. 321-345, Jan. 2000.
[7] Carsten Herpel, "Elementary Stream management in MPEG-4," *IEEE Trans. On Circuits and Systems for Video Technology*, Vol. 9, no. 2, Mar. 1999.
[8] ATSC Document A/65, "Program and System Information Protocol for Terrestrial Broadcast and Cable," ATSC, 1997.