Design and Development of T-DMB Multichannel Audio Service System Based on Spatial Audio Coding

Yong Ju Lee, Jeongil Seo, Seungkwon Beack, Daeyoung Jang,
Kyeongok Kang, Jinwoong Kim, and Jin Woo Hong

In this paper, a terrestrial digital multimedia broadcasting (T-DMB) multichannel audio broadcasting system based on spatial audio coding is presented. The proposed system provides realistic multichannel audio service via T-DMB with a small increase of data rate as well as backward compatibility with the conventional stereo-based T-DMB player. To reduce the data rate for additional multichannel audio signals, we compress the multichannel audio signals using the sound source location cue coding algorithm, which is an efficient parametric multichannel audio compression technique. For compatibility, we use the dependent property of an elementary stream descriptor, and this property should be ignored in a conventional T-DMB player. To verify the feasibility of the proposed system, we implement the T-DMB multichannel audio encoder and a prototype player. We perform a compatibility test using the T-DMB multichannel audio encoder and conventional T-DMB players. The test demonstrates that the proposed system is compatible with a conventional T-DMB player and that it can provide a promisingly rich audio service.

Keywords: T-DMB, DMB, multichannel audio.

Manuscript received Sept. 24, 2008; revised Apr. 8, 2009; accepted May 26, 2009.

The work was supported by the IT R&D program (2008-F-011, Development of Next Generation DTV Core Technology) of KEIT&KCC&MKE, Rep. of Korea.

Yong Ju Lee (phone: +82 42 860 1672, email: draball@etri.re.kr), Jeongil Seo (email: seoji@etri.re.kr), Seungkwon Beack (email: skbeack@etri.re.kr), Daeyoung Jang (email: dyjang@etri.re.kr), Kyeongok Kang (email: kokang@etri.re.kr), Jinwoong Kim (email: jwkim@etri.re.kr), and Jin Woo Hong (email: jwhong@etri.re.kr) are with Broadcasting & Telecommunications Convergence Research Laboratory, ETRI, Daejeon, Rep. of Korea.

doi:10.4218/etrij.09.0108.0557

I. Introduction

Since digital signal processing technologies and multimedia technologies have been rapidly developing, rich media such as high definition (HD) video and multichannel audio contents are rapidly spreading. The evolution of television from black-and-white to digital high-definition TV (HDTV) in color clearly represents this trend. The same trend is also found in audio systems, that is, the 5.1- and 7.1-channel audio representation systems have evolved out of mono/stereo audio systems.

Multichannel audio can provide a more substantial sound field than stereo audio because multiple loudspeakers are arranged beside and behind the listener as well as in front of the listener as shown in Fig. 1, which shows the 5.1 channel loudspeaker arrangement specified in ITU-R [1]. For this reason, the latest multimedia contents such as DVD movies and DVD audio are authorized in a 5.1 channel audio format. Moreover, HDTV broadcasting specifications were established to support 5.1-channel audio.

The main contexts for multichannel audio have been cinemas and home theater systems. Recently, the number of automobiles with well-designed multichannel audio presentation systems has been increasing, enabling users to enjoy multichannel audio in an automobile environment. With the emergence and growing popularity of automotive entertainment systems and multichannel audio content, particularly when they are coupled with video playback systems, audio signal processing technologies supporting multichannel audio representation have become a significant and growing part of the current automotive environment [2].

The development of multimedia coding technologies and mobile device implementation technologies makes it possible to

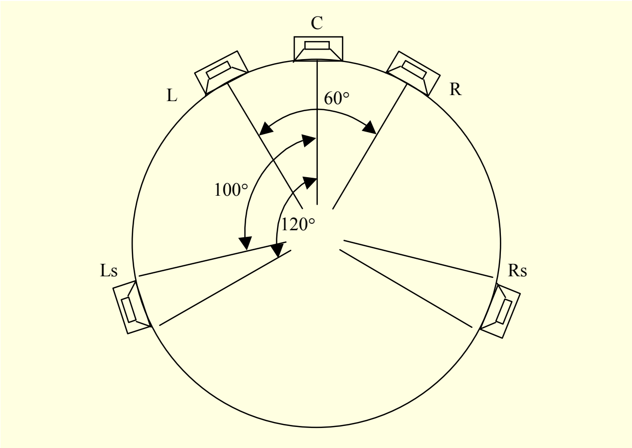


Fig. 1. Multichannel audio (5.1 ch) loudspeaker arrangement.

serve a new multimedia broadcasting service over a mobile environment. Digital multimedia broadcasting (DMB), digital video broadcasting-handheld (DVB-H), and MediaFLO were recently proposed for mobile multimedia broadcasting service [3]-[5]. In particular, DMB provided the first commercial digital mobile video broadcasting service of its kind in the world. The performance targets of DMB are providing VCD (video CD) quality video and FM radio quality audio. Because most mobile multimedia players, which are the main target terminals of DMB, do not have a multichannel audio representation environment, the DMB system only supports mono or stereo audio service. There are various types of DMB players, such as a cellular-phone-embedded type, PC-mounted type, portable-multimedia-player (PMP)-embedded type, and so on. As the number of people who want to see DMB service in an automobile is increasing, the DMB player is rapidly being embedded in PMPs and GPS navigators. Although the current DMB service only supports mono and stereo audio, since the number of automobiles with a well-designed multichannel audio representation system is increasing for DVD playback, it will be a valuable research topic to work towards providing multichannel audio service in the DMB environment.

In this paper, terrestrial DMB (T-DMB) multichannel audio broadcasting is described. There are two dominant issues we have to consider when developing the T-DMB multichannel audio service system. One issue is the narrow bandwidth of T-DMB transmission, and the other issue is how to preserve backward compatibility with conventional T-DMB players. To solve these two issues, we used a highly efficient parametric multichannel audio coding technology based on a spatial audio coding (SAC) algorithm, which is compatible with the stereo audio systems. The bit rate of additional data for reconstruction of the multichannel audio signal from the stereo audio signal is below 20 kbps. We also designed a flexible elementary stream (ES) description mechanism for signaling a multichannel audio service. The proposed mechanism can conceal the side information for multichannel audio from a conventional DMB player [6].

The remainder of this paper is organized as follows. In section II, we give an overview of the T-DMB standard. In section III, we introduce the concept of sound source location cue coding (SSLCC), which is a main algorithm to represent side information for multichannel audio. In sections IV and V, we describe a multichannel audio service system over T-DMB and its test results. Finally, we summarize and conclude this paper in section VI.

II. Overview of T-DMB Standards

The digital audio broadcasting (DAB) system can provide a reliable and multiplexed digital audio broadcasting service including data for mobile devices and portable and fixed players with a simple non-directional antenna [7]. During the last decade, the DAB system has demonstrated its stable functionality for mobile reception of a signal up to 200 km/h. On the basis of this mobile reception property, the technical challenge of transmitting multimedia signals for mobile broadcasting has been attempted since 2001 in Korea. The requirement for this challenging DMB system is to provide high quality multimedia services with interactive data for mobile reception circumstances.

This technical challenge has been successfully realized and was demonstrated to the public through an on-air service in December 2003. On the basis of this technical success, the T-DMB system specification was published as a Korean domestic standard in August 2004. In terms of worldwide usages, the T-DMB system was technically approved by the WorldDAB forum in November 2004 and finally published as an ETSI standard in June 2005 [3], [8].

To meet the requirements of T-DMB, international standards for multimedia service and a more robust channel coding scheme were applied to the traditional DAB system. As the DAB system was designed mainly for digital audio broadcasting, technologies for audio visual (AV) services such as compression and synchronization are required for T-DMB. MPEG-4 AVC|ITU-T H.264 for video compression and MPEG-4 Bit-Sliced Arithmetic Coding (BSAC)/High Efficiency Advance Audio Coding (HE-AAC) for audio compression are used [9], [10]. The DMB system also defines the interactive data service functionality in order to provide additional information suitable for a display size and to prepare convergent services between broadcasting and telecommunications. This functionality is realized by MPEG-4 systems technology [6]. For the transmission of a multimedia signal, an individual signal including AV and interactive data is

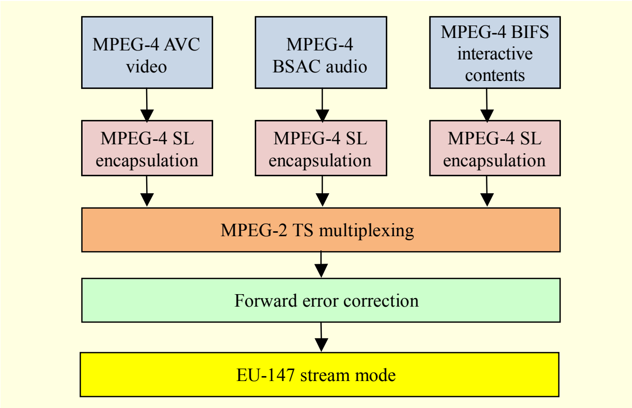


Fig. 2. Multimedia specification in T-DMB.

synchronized by MPEG-4 systems synchronization layer (SL) packetization and is multiplexed into an MPEG-2 transport stream (TS) as shown in Fig. 2.

The T-DMB system selects MPEG-4 BSAC or HE-AAC for audio compression. BSAC is one of the MPEG-4 general audio coding tools based on the perceptual coding approach used in the MPEG-2 and MPEG-4 AAC schemes. The compression tools of BSAC are similar to those of AAC except for the lossless coding algorithm; therefore, the coding efficiency of BSAC is almost the same as that of AAC. HE-AAC, in other words, aacPlus, is the combination profile of two MPEG audio technologies composed of AAC and spectral band replication (SBR). The SBR tool in the HE-AAC profile improves the performance of a low bit rate audio codec by increasing the audio bandwidth. Thus, the HE-AAC profile provides significantly better audio quality than AAC at a lower bit rate (under 48 kbps). Both BSAC and HE-AAC are defined as audio compression schemes in the T-DMB specification of ETSI.

Conventionally, the overall bandwidth of the AV stream itself should be lower than 512 kbps to provide two AV programs in one DAB ensemble in a Korean commercial service. Therefore, the bandwidth for an audio signal is limited to 128 kbps in the current T-DMB standard. In this regard, the current T-DMB system provides only mono and stereo audio service [8].

III. Overview of SSLCC

To provide a multichannel audio service compatible with a conventional T-DMB service, a new multichannel audio coding technology that is compatible with stereo audio systems is needed. The SSLCC algorithm developed by ETRI is a multichannel audio coding technology based on an SAC algorithm and is used for the proposed multichannel audio T-DMB system.

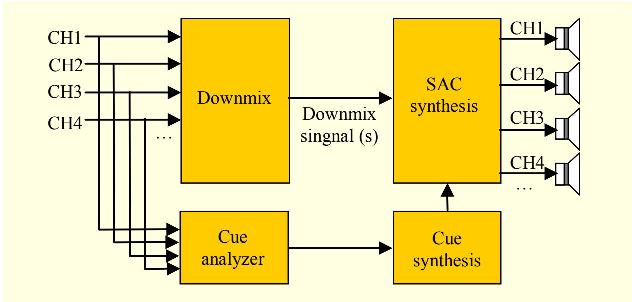


Fig. 3. Generic structure of SAC.

1. Spatial-Cue-Based Multichannel Audio Coding

It is commonly known that high bit rates should be requested when encoding multichannel audio signals proportional with an increase in the target number of audio channels when each channel is encoded individually by a legacy audio codec. For many years, audio coding technology has focused on reducing time-domain redundancy. Nowadays, compression techniques are being further developed to remove spatial redundancy by the introduction of a spatial parametric coding scheme. The spatial parametric coding scheme makes it possible to remarkably improve the compression performance of audio data and even reduce the number of audio channels. The spatial-cue-based multichannel audio coding technology is only one of the branches of the parametric coding scheme, but it is the most representative in terms of coding efficiency. The basic idea of spatial-cue-based multichannel audio coding, or simply SAC, is to estimate the spatial cues reflecting the spatial characteristics among different audio channels and then to encode them instead of directly encoding each channel.

Figure 3 shows a schematic diagram of SAC. The spatial cues are estimated by analyzing the input multichannel audio signals in a cue analyzer. Then, the multichannel signals are downmixed to mono or stereo signals. Because the compression of the downmixed audio signals is not within the scope of SAC, the downmixed audio signals can be encoded with conventional stereo audio coders such as HE-AAC, BSAC, MP3, and so on. Thus, SAC can be adapted to various multimedia systems using various stereo audio codecs.

Currently, as a result of considerable effort towards an improvement of the SAC scheme, MPEG Surround is regarded as a concrete multichannel coding scheme with high efficiency [11]. It shows significant performance improvement over a legacy multichannel audio coding scheme, in terms of the total bit rates required for each coding scheme.

MPEG Surround has the following three kinds of spatial cues:

- Channel level difference (CLD): The power ratio parameter

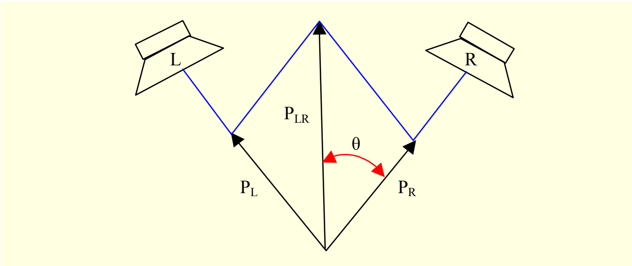


Fig. 4. Example of VSLI cue estimation in a pair of loudspeakers.

between two channels is represented as a logarithmic value (dB).

- Channel prediction coefficient (CPC): The prediction coefficient parameter is used to reconstruct three audio channel signals from two-channel downmixed signals.
- Inter channel correlation (ICC): This parameter describes a correlation or coherence between two audio channel signals.

Each spatial cue contributes toward reconstructing the corresponding multichannel signals. CLD has the main role in estimating the spectral structure of each multichannel signal, CPC is applied particularly to the case of stereo downmixing transmission in order to help the up-mixing of stereo signals into three output channel signals, and ICC is used to determine the overall spatial wideness of an audio scene.

Even though CPC and ICC also have a pivotal role in reconstructing a multichannel audio scene, CLD is the primary cue to reproduce multichannel signals because of its ability to redraw the spectral shapes of each channel signal from the given downmixed signals.

2. Sound Source Location Cue Coding

The method to estimate and synthesize CLD is straightforwardly understandable as it is related to the spectral power gain estimation. The main concern is that the power gain accuracy is easily degraded by the quantization process.

To alleviate quantization distortion, an alternative spatial cue to replace CLD was introduced in [12]. This virtual source location information (VSLI) cue is an angle representation based on virtual sound source location. To obtain VSLI, the virtual source position between adjacent loudspeakers is first estimated in each sub-band. A schematic example is depicted in Fig. 4.

The amplitude of vectors can be obtained from two adjacent channel signals, and the corresponding position (angle) is calculated using each loudspeaker position. If these two signals are downmixed, the vector of the downmixed signal can be obtained from

$$S_v = A_L \times P_L + A_R \times P_R, \tag{1}$$

where S_v is the vector of a downmixed signal that can also be represented as the form $P_{LR}\measuredangle\theta$ as in Fig. 4, and A_L and A_R are complex values of the corresponding position of a virtual loudspeaker (that is, $A_L = \cos 30° + j\sin 30°$).

Table 1. Power panning angle calculation.

Synthesized panning angles	Synthesized power gain factor
$\theta_1 = \frac{\mathrm{LHa_b} - \mathrm{LSa_b}}{-110° - \mathrm{LSa_b}} \times 90°$	$F_{\mathrm{C,b}} = \sin\theta_2 + \sin\theta_4$
$\theta_2 = \frac{\mathrm{LSa_b} - 30°}{0° - 30°} \times 90°$	$F_{\mathrm{L,b}} = \cos\theta_1 \cos\theta_2$
$\theta_3 = \frac{\mathrm{RHa_b} - \mathrm{RSa_b}}{110° - \mathrm{RSa_b}} \times 90°$	$F_{\mathrm{Ls,b}} = \sin\theta_1$
$\theta_4 = \frac{\mathrm{RSa_b} + 30°}{0° + 30°} \times 90$	$F_{\mathrm{R,b}} = \cos\theta_3 \cos\theta_4$
	$F_{\mathrm{Rs,b}} = \sin\theta_3$

A parametric multichannel audio codec based on the alternative spatial cue can be derived from the concept of the VSLI cue. We call this SSLCC [12], [13]. The coding procedure of SSLCC is similar to that shown in Fig. 3, but the analysis and synthesis parts of spatial cues are newly designed to adopt the VSLI. Under the assumption of a stereo downmixed signal transmission, four VSLI parameters are estimated from the input five-channel signals in the analysis:

$$LHv_b = A_C \times M_{C,b} / \sqrt{2} + A_L \times M_{L,b} + A_{Ls} \times M_{Ls,b}, \tag{2}$$

$$RHv_b = A_C \times M_{C,b} / \sqrt{2} + A_R \times M_{R,b} + A_{Rs} \times M_{Rs,b}, \tag{3}$$

$$LSv_b = A_L \times M_{L,b} + A_{Ls} \times M_{Ls,b}, \tag{4}$$

$$RSv_b = A_R \times M_{R,b} + A_{Rs} \times M_{Rs,b}. \tag{5}$$

where LHv_b and RHv_b are the left and right half-plane vectors; LSv_b and RSv_b are the left and right subsequent vectors of a 5.1 channel layout; subscript b is an index of the sub-band; subscripts L, Ls, C, R, and Rs (represented below as ch) denote the channel position in a 5.1 channel configuration; A_{ch} is the complex value of the loudspeaker position corresponding to ch; and $M_{ch,b}$ is the input signal power of the sub-band b at channel position ch, which is calculated by

$$M_{ch,b} = \sum_{n=B_b}^{B_{b+1}-1} \left| S_{ch,n} \right|. \tag{6}$$

The transmitted side information is represented as angles $\theta_{\mathrm{LHv}}, \theta_{\mathrm{LSv}}, \theta_{\mathrm{RHv}}, \theta_{\mathrm{RSv}}$, which are obtained using (2) to (6) simply by the arctangent law, $\theta_{\mathrm{LHv}} = \arctan(\mathrm{LHv}_b)$. Then, a uniform quantization scheme and Huffman coding can be applied to that information in order to represent a bitstream. On the decoder side, the spatial cue synthesizer converts the angle

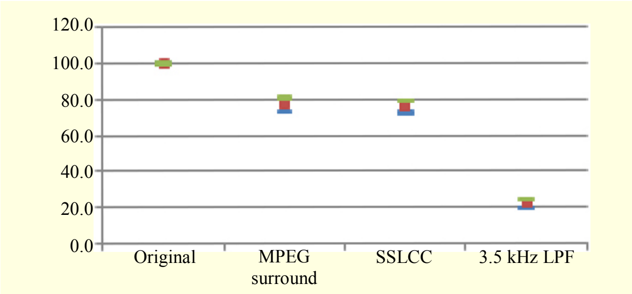


Fig. 5. Listening test results of SSLCC.

information into the power gain factor corresponding to each channel signal. The synthesizing procedure can be summarized as in Table 1. The equations of synthesized panning angles (that is, $\theta_1, \theta_2, \theta_3, \theta_4$) are derived using the constant power panning law, and each channel power gain factor (that is, $F_{C,b}, F_{L,b}, F_{R,b}, F_{Ls,b}, F_{Rs,b}$) can be attained using the panning angles. The scope of the description in this paper is focused on the stereo downmix transmission. The detailed procedure in a mono transmission can be found in [11].

3. Performance Evaluation

To verify the compression performance of SSLCC, we conducted a listening test. Eight experienced listeners used the MUSHRA blind test method to relatively rank the items compared to a known unencoded reference. For the listening test, we used four multichannel audio items (applse, ARL_applause, indie2, poulenc) among eleven multichannel audio items which were used to evaluating the performance of a multichannel audio codec in MPEG Surround standardization [14], [15]. The four items are known to be difficult to properly encode.

The results of the listening test are shown in Fig 5. The audio quality of MPEG Surround is slightly better than that of SSLCC. But the scores overlap in the 95% confidence interval, so it can be said that the performance of SSLCC is similar to that of MPEG Surround.

IV. Design of Multichannel Audio Service System over T-DMB

The main properties of the proposed multichannel audio T-DMB system are that it needs a very low additional bit rate for multichannel audio service, and it is backward compatible to a conventional T-DMB system. To achieve this, we used SSLCC and the dependency property of an elementary stream descriptor (ESD). In the previous section, we gave an overview of the SSLCC algorithm. In this section, we describe the transmission mechanism of a VSLI cue and the functionality of

```
ObjectDescriptor {
    ObjectDescriptorID 3
    esDescr {    // video ES
        ES_Descriptor {
            ES_ID 3
            muxInfo muxInfo {
                fileName "test_01.avc"
                streamFormat AVC
            }
            decConfigDescr DecoderConfigDescriptor {
                streamType 4
                …
            }
            slConfigDescr SLConfigDescriptor {
                ...
            }
        }
    }
}
ObjectDescriptor {
    ObjectDescriptorID 4
    esDescr {    // Stereo audio ES
        ES_Descriptor {
            ES_ID 4
            muxInfo muxInfo {
                fileName "test_01.sac"
                streamFormat BSAC
            }
            decConfigDescr DecoderConfigDescriptor {
                streamType 5
                bufferSizeDB 15060000
                objectTypeIndication 0x40
                decSpecificInfo DecoderSpecificInfoString {
                    info "obsolete string"
                }
            }
            slConfigDescr SLConfigDescriptor {
                ...
            }
        }
    }
}
```

Fig. 6. Example of an OD structure for T-DMB.

dependent ESD in more detail. We also describe the multichannel audio service system over T-DMB using these techniques.

1. Signaling and Packetizing of Multichannel Audio Signal

As previously described, the SSLCC encoding scheme converts a multichannel audio signal into a downmixed signal and side information. Since the downmixed signal is either a mono or stereo audio signal, it can be compressed by the BSAC standard. Then, it is packetized into an MPEG-2 TS through consecutive SL packetizing, PES packetizing, and TS packetizing procedures, just like in a conventional T-DMB system. Side information is generated at every audio frame as in an audio stream, so we can packetize the side information into an MPEG-2 TS through the same procedure as that used for the downmixed signal. However, the side information is not an elementary stream the T-DMB system supports, so it should be transmitted as a private stream. For this reason, an additional signaling method is needed to identify the side information. We used the dependent property of ESD for the signaling of side

```
ObjectDescriptor {
    ObjectDescriptorID 3
    esDescr {    // video ES
        ES_Descriptor {
            ES_ID 3
            muxInfo muxInfo {
                fileName "test_01.avc"
                streamFormat AVC
            }
            decConfigDescr DecoderConfigDescriptor {
                streamType 4
                …
            }
            slConfigDescr SLConfigDescriptor {
                ...
            }
        }
    }
}
ObjectDescriptor {
    ObjectDescriptorID 4
    esDescr {    // define downmix audio ES
        ES_Descriptor {
            ES_ID 4
            muxInfo muxInfo {
                fileName "test_01.sac"
                streamFormat BSAC
            }
            decConfigDescr DecoderConfigDescriptor {
                streamType 5
                bufferSizeDB 15060000
                objectTypeIndication 0x40
                decSpecificInfo DecoderSpecificInfoString {
                    info "obsolete string"
                }
            }
            slConfigDescr SLConfigDescriptor {
                ...
            }
        }
    }
    esDescr {    // define side information
        ES_Descriptor {
            ES_ID 6
            streamDependeceFlag TRUE    // define dependency
            muxInfo muxInfo {
                fileName "test_01.ssl"
                streamFormat SSLCC
            }
            decConfigDescr DecoderConfigDescriptor {
                streamType 5
                bufferSizeDB 15060000
                objectTypeIndication 0x40
                decSpecificInfo DecoderSpecificInfoString {
                    info "obsolete string"
                }
            }
            slConfigDescr SLConfigDescriptor {
                ...
            }
        }
    }
}
```

Fig. 7. Example of an OD structure for a T-DMB multichannel audio service.

information.

Since a T-DMB system uses the MPEG-4 systems standard [6], the initial object descriptor (IOD), object descriptor (OD), and binary information for scene (BIFS) are transmitted to a player for signaling MPEG-4 contents. The OD has the key information about the properties of an individual object, such as the stream type, ESD, and so on. Figure 6 shows an example of an OD that has one video object and one audio object used in a T-DMB system.

There are ESDs within an OD which convey all information related to a particular elementary stream. An ESD has the property of dependency. When the "streamDependenceFlag" field of an ESD is set to "TRUE," the ES is dependent on other ESs. In the MPEG-4 systems standard, there are many profiles and levels for various application environments. Some profiles support the dependency of an ES, but others do not. In the case of T-DMB, a simple profile is adopted regarding the complexity of the systems. Because a simple profile does not support the dependency of an ES, an ES that is described as a dependent ES is ignored by a conventional (stereo) T-DMB system.

Figure 7 is an example of an OD that has one video object and two audio objects: one audio object is for the downmixed audio signal and the other audio object is for the side information used in the proposed T-DMB multichannel audio system.

In a T-DMB multichannel audio system, we describe the side information as a dependent ES to the main stereo audio signal, and we interpreted the dependent ES as side information in order to reconstruct a multichannel audio signal. When a conventional T-DMB stream that does not contain side information is delivered to the T-DMB multichannel audio player, the SSLCC decoder does not work, and a stereo audio signal will be played. Thus, the T-DMB multichannel audio player is forward compatible with a conventional T-DMB system. When a T-DMB multichannel audio stream is delivered to a conventional T-DMB player, which does not have the SSLCC decoder, the side information is ignored and wasted because the side information is described as a dependent ES. Thus, the T-DMB multichannel audio system is backward compatible to a conventional T-DMB system.

2. System Design

Using SSLCC and the dependency property of ESD, we designed the T-DMB multichannel audio system. In this section, we describe the T-DMB encoding system and T-DMB multichannel audio encoding system, and the differences between the two.

A. T-DMB Encoding System

The T-DMB encoding system receives analog video and an audio signal and makes them an MPEG-2 TS as per the T-DMB standards. The structure of the T-DMB encoding system is shown in Fig. 8.

An AVC encoder encodes a video signal into a video ES as per the AVC standard. A BSAC encoder encodes a mono or stereo audio signal into an audio ES as per the BSAC standard.

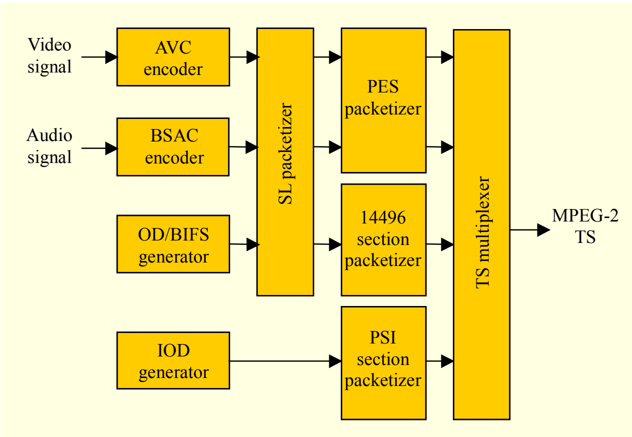


Fig. 8. Structure of T-DMB encoder system.

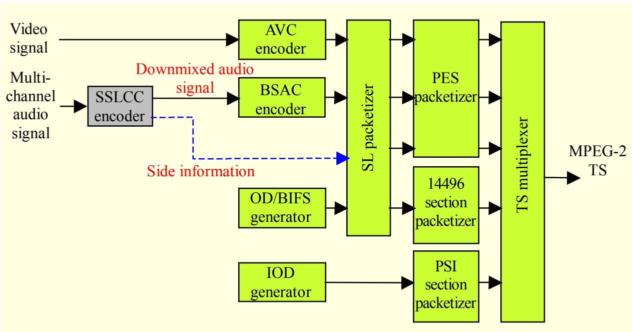


Fig. 9. Structure of the T-DMB multichannel audio broadcasting encoding system.

An OD/BIFS generator and IOD generator generate the OD, BIFS, and IOD information for signaling of the DMB stream. The SL packetizer, PES packetizer, 14496 section packetizer, PSI section generator, and TS multiplexer packetize the video ES, audio ES, and OD, BIFS data into an MPEG-2 TS.

B. T-DMB Multichannel Audio Encoding System

The structure of the multichannel audio encoding system over T-DMB is shown in Fig. 9.

The differences compared to a conventional T-DMB encoding system are that the T-DMB multichannel audio encoding system has an SSLCC encoder and a data path for side information from the SSLCC encoder to the SL packetizer.

The SSLCC encoder converts a multichannel audio signal into a stereo downmixed audio signal and side information. The downmixed stereo audio signal is encoded by the BSAC encoder as in a conventional T-DMB encoder. Side information is packetized into an MPEG-2 TS after SL packetizing and PES packetizing. The OD, which is generated by an OD/BIFS generator, has the ESD for the side information, whose streamDependenceFlag is set to 'TRUE'.

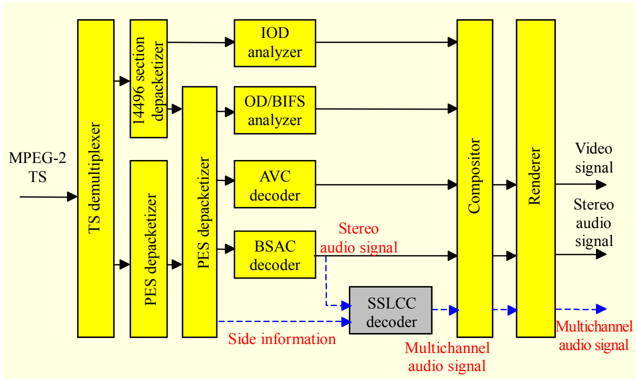


Fig. 10. Structure of T-DMB multichannel audio player.

C. T-DMB Multichannel Audio Player

A T-DMB multichannel audio player has an additional function for decoding multichannel audio compared to a conventional T-DMB player. Thus, the structure of a T-DMB multichannel audio player is a little different from a conventional one. Figure 10 presents the structure of a T-DMB multichannel audio player.

The decoding process of the multichannel audio signal in the player is carried out as follows. First, the audio ES and side information are acquired from the received MPEG-2 TS by TS demultiplexing, PES depacketizing, and SL depacketizing. The audio ES is converted into a stereo audio signal by BSAC decoding processing. The SSLCC decoder reconstructs a multichannel audio signal using the stereo audio signal and side information.

V. Experiments

To verify the proposed T-DMB multichannel audio system, we implemented the encoder and player. The proposed system uses a DVD player as a multichannel sound source and an RF generator for real-time broadcasting. One of the effective environments for a T-DMB multichannel audio service is considered to be an automobile. Therefore, we equipped the player in an automobile as well as a laboratory to examine and verify the T-DMB multichannel audio service. Figure 11 represents the test environment, and the detailed descriptions about the test are as follows.

1. T-DMB Multichannel Audio Transmission

For testing and verification of the T-DMB multichannel audio system, we developed a real-time T-DMB multichannel audio encoding system based on a PC. Because there are few audio sound cards that can process 5-channel analog audio signals, we used a multichannel audio interface apparatus whose input interface is analog audio and whose output

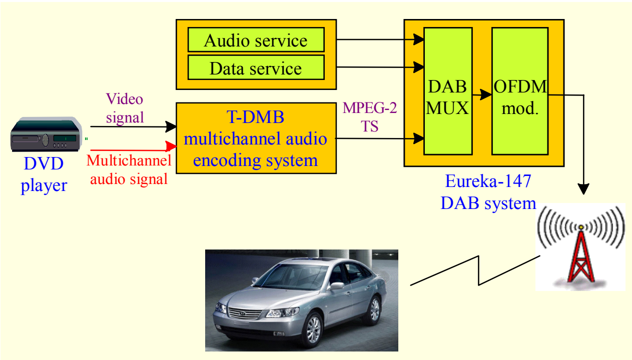


Fig. 11. T-DMB multichannel audio service test environments.

Fig. 12. Screen shot of the T-DMB multichannel audio encoder in run mode.

interface is IEEE1394.

Because many DVD movies contain a 5.1 channel audio signal, we used DVD movies as test content. A DVD player was used to play the DVD movies. The outputs of the DVD player were a composite of video signal and analog 5.1 channel audio signals. The PC-based encoder received the composite video signal directly from the DVD player and digitized audio samples from a multichannel audio interface apparatus. It then encoded them into an MPEG-2 TS.

Figure 12 shows a screen shot of the T-DMB multichannel audio encoding system.

In the T-DMB multichannel audio encoding system, it is possible to control the bit rate of the video and audio ESs. The following table shows the bit rate allocation of these experiments.

Since an ES should be packetized into an MPEG-2 TS through SL packetizing and PES packetizing, the TS rate is higher than the ES rate. In particular, in the case of SSLCC side information, the TS rate is four-times higher than the ES rate. This is because one transport packet (the size of a transport packet is fixed at 188 bytes) does not have more than one access unit, and the access unit of SSLCC side information is

Table 2. Bit rate allocation.

Classification	ES rate (kbps)	TS rate (kbps)
Video (AVC)	300	360
Audio (BSAC)	54	65
SSLCC side information	15	65
PAT	1	5
PMT	1	5
OD	3	5
BIFS	3	5
Summary	377	510

about 43 bytes. Thus, more than 100 bytes of dummy data are inserted in the TS packetizing process. In the case of SSLCC side information, it can be said that it is an inefficient transport method. Therefore, it should be improved for more efficient transmission.

An MPEG-2 TS is delivered to a commercial transmitter that has the functions of ensemble multiplexing and an RF generator. The transmitter receives the MPEG-2 TS through a UDP protocol, and it performs Reed Solomon coding, interleaving, and so on. Finally, it generates an RF signal on air.

2. T-DMB Multichannel Audio Player

As previously mentioned, we equipped an automobile with the multichannel audio representation environment. Figure 13 shows the arrangement of loudspeakers in the automobile. Although the speaker configuration in the automobile is not fit to the general 5.1ch speaker configuration, we did not use any signal processing algorithm to compensate for this. Instead, we made the automobile have a similar sound field by controlling the gain of each channel heuristically.

We implemented the PC-based T-DMB multichannel audio software player, and embedded the player in an automobile. The software player can parse the dependent ESD and contains the SSLCC decoding module. It uses a USB-type commercial T-DMB tuner module to receive T-DMB RF signal. Using this software player, we could receive the T-DMB multichannel audio signal on air and display video and multichannel audio. Figure 14 shows screen shots of the player.

3. Test for Compatibility

For verification of backward compatibility with a conventional T-DMB system, we executed a receiving test using a commercial T-DMB player that did not contain a multichannel audio decoder. A PDA-type T-DMB player and a cell-phone T-DMB player received the T-DMB multichannel

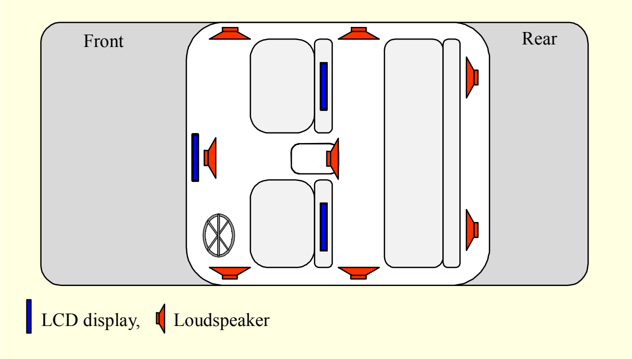
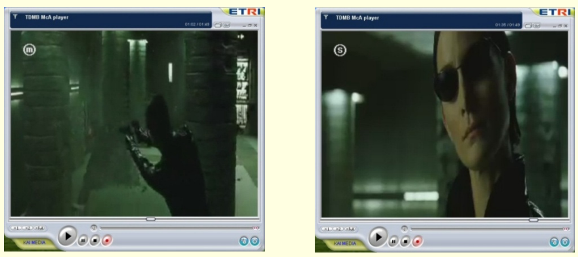


Fig. 13. Loudspeaker arrangement in an automobile.



(a) Multichannel audio playing mode (b) Stereo audio playing mode

Fig. 14. Screen shots of the T-DMB multichannel audio player.

Fig. 15. Snapshot of the PDA-type commercial T-DMB player receiving the T-DMB multichannel audio signal.

audio signal and displayed video and stereo audio as well. Figures 15 and 16 show snapshots of the PDA-type player and cell-phone-type player, respectively, receiving the T-DMB multichannel audio signal.

From this experimental broadcasting test, we could verify that the proposed T-DMB multichannel audio service system can provide multichannel audio service that is compatible with a conventional T-DMB system.

VI. Summary and Conclusion

The development of multimedia coding technologies and

Fig. 16. Snapshot of the cell-phone-type commercial T-DMB player receiving the T-DMB multichannel audio signal.

mobile device implementation technologies makes it possible to serve a new multimedia broadcasting service over a mobile environment. Although the current DMB service only supports mono and stereo audio, as the number of automobiles with a well-designed multichannel audio representation system is increasing for DVD playback, providing a multichannel audio service in an automobile DMB environment will be a valuable research topic.

In this paper, we proposed a T-DMB multichannel audio service that is an advanced service with multichannel audio via T-DMB. The proposed system requires only a small bit rate increment for multichannel audio service, and is compatible with conventional T-DMB services. To achieve this, we used a highly efficient parametric multichannel audio coding technology, SSLCC, and a dependent ESD mechanism.

To verify the proposed service, we implemented a real-time encoder and a player, and had an experimental broadcasting test. We confirmed that the proposed T-DMB multichannel audio system can provide a multichannel audio service with compatibility to a conventional T-DMB system.

On the other hand, we found that some technical issues still remain for commercial application of the proposed system. One of the issues is efficient packetizing of the side information. Because an MPEG-2 transport packet is not efficient for packetizing a very low bit rate elementary stream such as side information, the TS bit rate of the side information is much higher than an ES bit rate. One of the other issues is the signaling method of the side information when it is delivered while maintaining the MPEG-2 systems standard. There is no clear definition for signaling side information in the MPEG-2 systems standard; therefore, research and standardization regarding the packetizing and signaling of side information are needed for commercialization of the proposed service.

References

[1] ITU-R BS.775-2, “Multichannel Stereophonic Sound System

With and Without Accompanying Picture," Jan. 2006.
[2] B. Crockett, M. Smithers, and E. Benjamin, "Next Generation Automotive Research and Technologies," *120th AES Conference*, May 2006.
[3] "ETSI TS 102 428 Digital Audio Broadcasting (DAB); DMB Video Service; User Application Specification," ETSI, June 2005.
[4] Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H), ETSI EN 302 304 V1.1.1 (2004-11), European Telecommunications Standards Institute DVB-H.
[5] Qualcomm mediaFLO homepage, http://www.qualcomm.com/mediaflo/.
[6] ISO/IEC 14496-1, Information Technology: Coding of Audio-Visual Object, Part 1; Systems, Nov. 2002.
[7] "EN 300 401 Radio Broadcasting System: Digital Audio Broadcasting (DAB) to Mobile, Portable, and Fixed Receivers," ETSI, Jan. 2006.
[8] "ETSI TS 102 427 Digital Audio Broadcasting (DAB): Data Broadcasting-MPEG-2 TS Streaming," ESTI, July 2005.
[9] ITU-T Rec. H.264 | ISO/IEC 14496-10 "Information Technology: Coding of Audio-Visual Objects, Part 10: Advanced Video Coding."
[10] ISO/IEC 14496-3: 2001 "Information Technology: Coding of Audio-Visual Objects, Part 3: Audio."
[11] ISO/IEC 23003-1, "Information Technology: MPEG Audio Technologies: Part 1: MPEG Surround, Feb. 2007.
[12] S.K. Beack et al., "Angle-Based Virtual Source Location Representation for Spatial Audio Coding," *ETRI Journal*, vol. 28, no. 2, Apr. 2006, pp.219-222.
[13] H.G. Moon et al., "A Multichannel Audio Compression Method with Virtual Source Location Information for MPEG-4 SAC," *IEEE Trans. Consum. Electron.*, vol. 51, no. 4, Nov. 2005, pp.1253-1259.
[14] ISO/IEC JTC1/SC29/WG11 (MPEG), "Procedures for the Evaluation of Spatial Audio Coding Systems," Document N6691, Redmond, July 2004.
[15] J. Breebaart et al., "MPEG Spatial Audio Coding / MPEG Surround: Overview and Current Status," *Proc. 119th AES Convention*, New York, USA, Oct. 2005, Preprint 6447.

Yong Ju Lee received the BS and MS degrees in electronics from Kyungpook National University, Daegu, Korea, in 1999 and 2001, respectively. Since 2001, he has been a member of research staff with Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, where he has been involved in developing an interactive data broadcasting system and a terrestrial digital multimedia broadcasting system. His research interests include intelligent broadcasting and digital audio signal processing.

Jeongil Seo received his BE, MS, and PhD degrees in electrical engineering and computer science from Kyungpook National University, Daegu, Korea, in 1994, 1996, and 2005, respectively. He was a member of engineering staff with the Laboratory of Semiconductor, LG-Semicon, Cheongju, Korea, from 1998 to 2000. Since 2002, he has been with ETRI, Daejeon, Korea, where he is a senior researcher in the Broadcasting and Telecommunications Convergence Media Research Department. His research interests are digital audio processing, room acoustics, real-time audio codec systems, and interactive 3D audio broadcasting systems.

Seungkwon Beack received the BS degree in electrical engineering from Hankuk Aviation University, Korea, in 1999, and the MS degree in electrical engineering from Information and Communication University, Korea, in 2001. He is currently pursuing the PhD at Information and Communication University. He has been a member of research staff with ETRI, Daejeon, Korea. His research interests are in the fields of audio and speech signal processing, spatial audio processing, and multi-channel signal processing.

Daeyoung Jang received the BE degree in electronic engineering from Pukyong National University, Busan, Korea, in 1991, and the MS and PhD degrees in computer science from Paichai University, Daejeon, Korea, in 2000 and 2008, respectively. He has been with ETRI, Daejeon, Korea, since 1991, and he is now a principal member of engineering staff. He had researched electro-acoustics for telecommunications and broadcasting, and he has worked on the development of MPEG-1, 2, and 4 audio systems. He is currently working on the development of interactive audio technologies for realistic broadcasting and telecommunications.

Kyeongok Kang received his BS and MS degrees in physics from Pusan National University, Busan, Korea, in 1985 and 1988, respectively, and his PhD degree in electrical engineering from Hankuk Aviation University, Seoul, Korea, in 2004. He has been with ETRI since 1991, and he is now a principal member of engineering staff and the leader of the Media Application Research Team. His major interests are in low-bitrate audio coding; audio signal processing, including 3-dimensional audio and personalized broadcasting based on MPEG-7; and TV-Anytime related issues.

Jinwoong Kim received the BS and MS degrees from Seoul National University, Seoul, Korea, in 1981 and 1983, respectively. He received the PhD degree in electrical engineering from Texas A&M University, Texas, USA, in 1993. He has been working with ETRI since 1983 and is now a principal member of engineering staff in the Broadcasting and Telecommunications Media Research Division. He is currently a 3DTV project leader, and his major interests are digital broadcasting system, video coding, multimedia processing, and 3DTV.

Jin Woo Hong received the BS and MS degrees in electronic engineering from Kwangwoon University, Seoul, Korea, in 1982 and 1984, respectively. He also received the PhD in computer engineering from the same university in 1993. Since 1984, he has been with ETRI in Daejeon, Korea, as a principal member of engineering staff, where he is currently a director of the Broadcasting and Telecommunications Convergence Media Research Department. From 1998 to 1999, he worked at Fraunhofer Institute in Erlangen, Germany, as a visiting researcher. From 1984 to the present, he has participated in many projects including the development of the TDX electronic exchange, ISDN system, and ISDN terminal; development of the digital broadcasting system and multichannel audio codec system; and development of digital broadcast content protection and management. His current research interests include audio and speech signal processing, multimedia framework, and broadcasting service technology.