# Multi-channel Audio Service in a Terrestrial-DMB System Using VSLI-Based Spatial Audio Coding

Jeongil Seo, Han-gil Moon, Seungkwon Beack, Kyeongok Kang, and Jae-Keun Hong

*ABSTRACT—Spatial audio coding (SAC) is an extremely high compact representation of encoded multi-channel audio material. This paper suggests a multi-channel audio service in the terrestrial digital multimedia broadcasting (T-DMB) system using a novel SAC tool, which is called a virtual source location information (VSLI)-based SAC tool. Intensive experiments are presented to evaluate the validity of the proposed VSLI-based SAC tool, and prototypical systems are also presented to demonstrate the reliability of the proposed multi-channel T-DMB system in real applications.*

*Keywords—Spatial audio coding, terrestrial digital multimedia broadcasting system, multi-channel audio.*

## I. Introduction

Due to the limitations of allowed frequency bands and the insufficient sound quality of the analog FM radio broadcasting system, the digital audio broadcasting (DAB) system was designed for the delivery of high-quality audio programs and data services [1]. DAB eliminates both interference and multi-path problems that occur in vehicles, and provides CD-like undistorted sound quality. As such, it is currently being implemented and exploited in many countries such as America, England, and Germany.

The terrestrial digital multimedia broadcasting (T-DMB) system is designed to provide multimedia broadcasting service in a mobile reception environment over a DAB system [2], [3]. To provide high-quality audio and video, T-DMB uses state-of-the-art technologies. MPEG-4 bit-sliced arithmetic coding (BSAC) and MPEG-4 advanced video coding (AVC) are used for compressing audio and video signals. As well, MPEG-4 and MPEG-2 systems are adapted to supply object-based interactivity and compact packetization. The overall bandwidth on the AV stream itself should be lower than 512 kbps for providing two AV programs in one DAB ensemble, so 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 services.

This paper describes a novel T-DMB system for a multi-channel audio service that allows for compatibility with the current T-DMB system for mono/stereo audio service. A novel SAC tool, which uses virtual source location information (VSLI) as key spatial cue parameters, is also suggested for compact representation of the multi-channel audio signal.

## II. VSLI-Based Spatial Audio Coding Tool

Binaural cue coding (BCC) is one of the most powerful SAC systems. BCC allows for the representation of multi-channel audio signals through the down-mix audio signal and additional spatial parameters, including inter-channel level differences (ICLDs), inter-channel time differences (ICTDs), and inter-channel coherences (ICCs) [4]. ICLD and ICTD describe a power difference and time difference between channels individually. ICTD describes a time difference between channels and is estimated by locating the maximum of the cross-correlation between channels. The BCC system is mainly composed of its analyzer and synthesizer. The BCC analyzer extracts spatial parameters (ICLD, ICTD, and ICC)

Manuscript received Mar. 17, 2005; revised June 24, 2005.

This work is being supported by the Ministry of Information and Communication, Republic of Korea, under the title of "SmarTV Technology Development."

Jeongil Seo (phone: +82 42 860 1571, email: seoji@etri.re.kr) and Kyeongok Kang (email: kokang@etri.re.kr) are with Digital Broadcasting Research Division, ETRI, Daejeon, Korea.

Han-gil Moon (email: fullmoon@acoustics.snu.ac.kr) is with Applied Acoustics Laboratory, Seoul National University, Seoul, Korea.

Seungkwon Beack (email: skbeack@icu.ac.kr) is the School of Electrical Engineering, ICU, Daejeon, Korea.

Jae-Keun Hong (email: jkhong@ee.knu.ac.kr) is with the School of Electrical Engineering & Computer Science, Kyungpook National University, Daegu, Korea.

from the multi-channel signal. The BCC synthesizer reconstructs a spatial image by restoring spatial localization cues when it generates multi-channel outputs from the down-mix signal. The BCC parameters can be transmitted with a rate of only a few kbps per channel. BCC can then provide multi-channel audio to users of current 2-channel stereo systems at a very small expense in bit-rate.

In this paper, a newly proposed virtual source location information (VSLI)-based SAC method is introduced. VSLI as a spatial cue parameter is an angle which represents the geometric spatial information between power vectors of inter-channel frequency bands rather than power ratios, such as a conventional cue, for example, the ICLD. The VSLI cue is extracted under the assumption that the playback layout of multi-channel loudspeakers is fixed. Also, the inter-channel power vectors form a spatial audio image between adjacent loudspeakers. Essentially, the ICLD and VSLI are similar spatial cue parameters for estimating the power of audio channels. However, there is a big difference in the quantization process. The ICLD theoretically will have infinite dynamic range to perfectly recover the original power ratio between channels, but the VSLI has a finite dynamic range because of the angle representation with the fixed speaker configuration. Therefore, the VSLI can provide better multi-channel sound quality than the ICLD in a limited bandwidth channel environment because the former is more robust to quantization noise than the latter.

In the proposed VSLI-based SAC system, one spatial image can be represented by one angle between adjacent loudspeakers. If the power of each channel is not zero at a certain frequency band, all spatial audio images can be realized by using at least five angles and power levels. However, the transmitted mono or stereo down-mix signal does not contain any power information for each angle. Only the power information about the transmitted signal is available, that is, global power in the case of a mono signal and half-plane powers in the case of a stereo signal. For this reason, each channel vector should be estimated from the global vector or half-plane vector.

The estimation procedure is simply implemented by a vector projection operation. In the case of a mono down-mix, the global vector is projected to the locations of both the left half-plane angle and right half-plane angle. The left and right half-plane vectors are then projected to the location of their subsequent angles. In the stereo down-mix case, each sub-band power of the stereo signal becomes the left and right half-plane powers, so the global vector information is not necessary in this case. Consequently, we can reconstruct five channel signals by utilizing the transmitted five angles: global angle (Ga), left half-plane angle (LHa), right half-plane angle (RHa), left subsequent angle (LSa), and right subsequent angle (RSa) as illustrated in Fig. 1.

A basic scheme of a VSLI-based SAC system is shown in Fig. 2. The VSLI analyzer receives N channels and performs a T/F (time to frequency) transform of the inputs. The spectra derived from the T/F transform for each of the channels are partitioned into 20 equivalent rectangular bandwidths (ERBs) approximating the ERB scale. The band energy vectors are then derived from the partitioned spectra, and the angles of the sound images of the bands are estimated by means of the energy vectors and the constant power panning law [5]. The VSLI synthesizer reconstructs N multi-channel gains per band by means of panning. With the gains-per-band and down-mix spectral values, the spectra of N channels are estimated. The final N channel audio outputs are then derived from the N channel spectra after an F/T (frequency to time) transform.

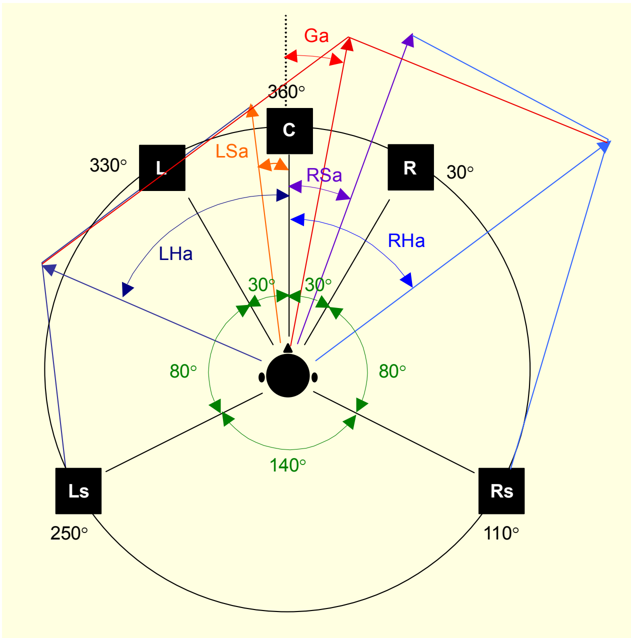


Fig. 1. Component angles of VSLI cues.

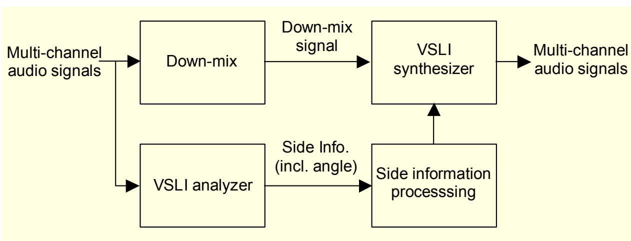


Fig. 2. A schematic block diagram of VSLI-based SAC.

## III. Multi-channel Audio Service in the T-DMB

The T-DMB system is designed to provide multimedia broadcasting service in a mobile reception environment over a DAB system. The bandwidth assigned to AV data is about 1.2 Mbps in one T-DMB channel. To provide two video services in one DAB channel, the bandwidth for the audio signal is limited to 128 kbps in the current T-DMB standard. In addition, the number of audio objects to be played is restricted to only one audio object in order to reduce the complexity of the receiver.

This paper suggests a novel T-DMB system for multi-channel audio services within a 128 kbps bandwidth while providing compatibility with the current T-DMB receiver for mono or stereo audio service. The VSLI-based SAC audio coder splits multi-channel audio signals into a down-mix audio signal and spatial cue parameters. The proposed T-DMB system uses the down-mix audio signal as a main audio elementary stream (ES) for providing compatibility with the legacy T-DMB system, and the spatial cue parameters are delivered in an additional audio ES for providing multi-channel audio service. The additional audio ES, which is compressed below 24 kbps, is defined as a dependent ES of the main audio ES. The MPEG-4 object descriptor (OD) structure can deal with two ESs as one object when there is dependency between them. However, the current T-DMB standard does not allow a dependent ES structure. The legacy T-DMB receiver can process only the main audio ES, so it will discard the additional audio ES, which includes spatial cue parameters. The proposed multi-channel audio T-DMB system makes multi-channel audio signals using the main audio stream and an additional audio stream.

Excepting audio encoding and decoding blocks, the proposed T-DMB system uses a legacy T-DMB system. A schematic diagram of the media processor for multi-channel audio service is shown in Fig. 3. The down-mixer makes a mono or stereo down-mix signal, and the MPEG-4 BSAC encoder makes a main audio ES. The VSLI-based SAC encoder extracts the so-called VSLI spatial cue parameters from multi-channel audio signals and makes an additional audio ES.

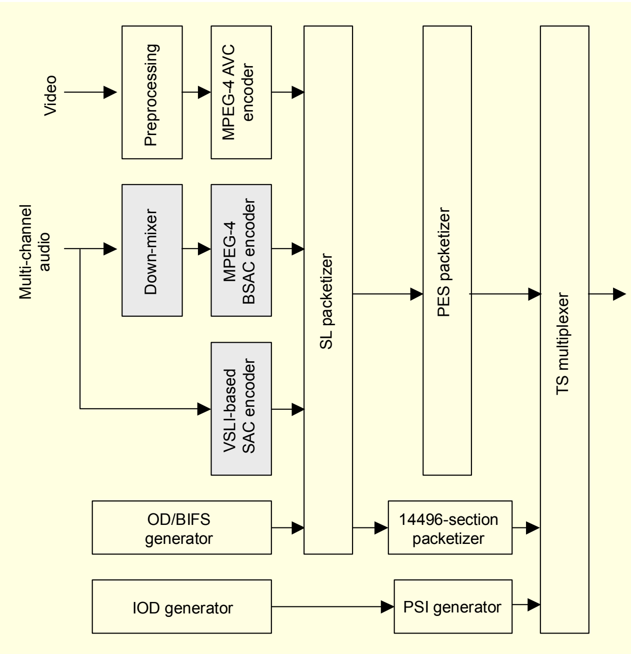

Fig. 3. Block diagram of T-DMB media processor for multi-channel audio service.

To show the validity of the proposed approach in a real application, we implemented a media processor and a player for the proposed system and experimented with a legacy DMB receiver. Multi-channel audio signals were split into a stereo down-mix signal and spatial cue parameters. The stereo down-mix signal was encoded at a rate of 96 kbps with the MPEG-4 BSAC encoder, and spatial cue parameters were encoded at a rate of 24 kbps with the VSLI-based SAC encoder. A video signal was encoded at a rate of 384 kbps with the MPEG-4 AVC encoder. The total bit-rate for the audio and video signal was 512 kbps including packet overhead.

## IV. Experiments on the VSLI-Based SAC System

For a set of eleven multi-channel (5-channel) items, which were offered by the MPEG audio sub-group [6], a subjective listening test was performed according to the MUSHRA test methodology [7]. These test items are listed in Table 1. The items have a resolution of 16 bits per sample at a sampling frequency of 44.1 kHz, and the length of all test items was less than 20 seconds. To obtain a fair assessment of the quality of the test items, the listening panels were composed of ten experienced listeners. Four systems, which are listed in Table 2, were used for this test. Listeners were asked to judge the basic audio quality of the versions of the test item in each trial.

Table 1. Test items.

| Index | Material | Description |
|---|---|---|
| A | Applause | Ambience |
| B | ARL_applause | Ambience |
| C | Chostakovitch | Music (back: direct) |
| D | Fountain_music | Pathological |
| E | Glock | Pathological |
| F | Indie2 | Movie sound |
| G | Jackson1 | Music (back: ambience) |
| H | Pops | Music (back: direct) |
| I | Poulenc | Music (back: direct) |
| J | Rock_concert | Music (back: ambience) |
| K | Stomp | Movie sound |

Table 2. Multi-channel audio systems under test.

| Classification | Description |
|---|---|
| Hidden reference | Original |
| Anchor | 3.5 kHz band-limited |
| Additional anchor | Software-based commercial Dolby ProLogic II encoder/decoder |
| Proposed 5-2-5 system | Stereo AAC codec (96 kbps) + VSLI-based SAC codec (24 kbps) |

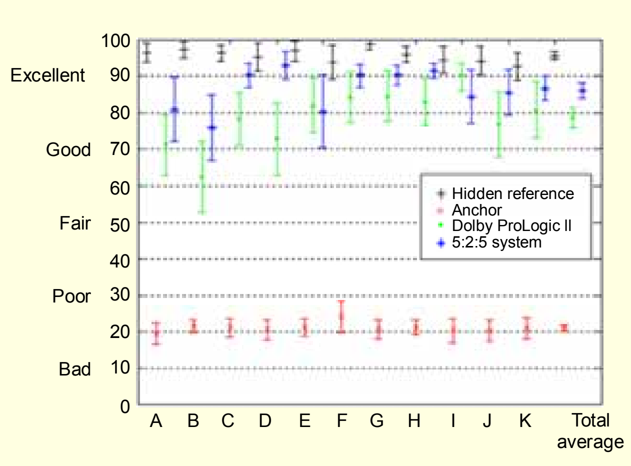


Fig. 4. Listening test results

The average scores and the 95% confidence intervals are shown in Fig. 4. The listening test results show that the proposed 5-2-5 system (stereo down-mix) performs significantly better than Dolby ProLogic II for three test items (C, D, H). The total average score shows that the proposed system is significantly better than Dolby ProLogic II. Furthermore, compared to the original hidden reference, the proposed system has less sound-quality degradation.

## V. Conclusion

This paper addressed the issue of multi-channel audio service on a T-DMB system. It was shown that the proposed VSLI-based SAC tool efficiently compressed multi-channel signals without significant quality degradation. As well, the prototypical media processor and player for multi-channel audio service in T-DMB were implemented and tested, yielding very promising results in terms of the validity of the proposed approach and its compatibility with legacy T-DMB systems.

On going standardization activities for MPEG-4 SAC are focused on how the side information for a multi-channel audio signal transmits with MPEG-4 AAC or MPEG-4 HE-AAC. The current standard documents restrict that the side information is multiplexed in the auxiliary data section of an AAC and HE-AAC bitstream. However, the T-DMB standard selected the MPEG-4 BSAC tool as a standard audio codec, so the MPEG-4 SAC standard should consider how to multiplex the side information with the BSAC stream. Moreover, it should also consider in the current standardization activities that the side information for multi-channel audio be independently delivered with a down-mixed audio bitstream to deploy SAC technologies into any kind of conventional mono/stereo audio codec.

## References


[1] ETSI EN 300 401, *Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to Mobile, Portable and Fixed Receivers*, 2000.

[2] Bong-Ho Lee, Kyu-Tae Yang, Young Kwon Hahm, Soo In Lee, and Chieteuk Ahn, “A Framework for MPEG-4 Contents Delivery over DMB,” *ETRI J.,* vol.26, no.2, Apr. 2004, pp. 112-121.

[3] Gwang Soon Lee, Byungjun Bae, Young Kwon Hahm, and Soo In Lee, “A Novel Method for Inserting an MPEG-2 TS into Ensemble in a DMB Transmission System,” *ETRI J.*, vol.26, no.6, Dec. 2004, pp. 653-656.

[4] C. Faller and F. Baumgarte, “Efficient Cue Coding Applied to Stereo and Multi-channel Audio Compression,” *Proc. AES 112th Conv*., Munich, Germany, May 2002.

[5] V. Pulkki, “Virtual Sound Source Positioning Using Vector Base Amplitude Panning,” *Proc. AES 113th Conv*., Los Angeles, CA, Oct. 2002.

[6] MPEG Audio Group, *Proc. the Evaluation of Spatial Audio Coding Systems*, ISO/IEC JTC1/SC29/WG11 N6691, Redmond, July 2004.

[7] ITU-R Recommendation, *Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA)*, ITU, BS. 1543-1, Geneva, 2001.