Vol: 6, No. 1, Maret 2017 ISSN: 2302 - 2949

REVIEWS ON TECHNOLOGY AND STANDARD OF SPATIAL AUDIO CODING

Ikhwana Elfitri*, Amirul Luthfi
Electrical Engineering Dept., Engineering Faculty, Universitas Andalas
*Corresponding author, e-mail : ikhwana@ft.unand.ac.id

Abstract—Market demands on a more impressive entertainment media have motivated for delivery of three dimensional (3D) audio content to home consumers through Ultra High Definition TV (UHDTV), the next generation of TV broadcasting, where spatial audio coding plays fundamental role. This paper reviews fundamental concept on spatial audio coding which includes technology, standard, and application. Basic principle of object-based audio reproduction system will also be elaborated, compared to the traditional channel-based system, to provide good understanding on this popular interactive audio reproduction system which gives end users flexibility to render their own preferred audio composition.

Keywords : spatial audio, audio coding, multi-channel audio signals, MPEG standard, object-based audio

1. INTRODUCTION

Various new technologies are being continuously invented and brought to the market. One of them is the technology of three dimensional (3D) audio, also called spatial audio [1], which is applied in many digital audio entertainment media such as Ultra High Definition TV (UHDTV), the next generation of TV broadcasting. For this UHDTV standard, multiple loudspeakers even more than 20 will be used to provide users with a realistic 3D audio perception. Every loudspeaker will be fed with single audio channel hence multi-channel audio signals will be required. Broadcasting companies such as NHK, Japan and BBC, UK, have actively participated in developing these audio chain technologies which include recording, transmission, and reproduction.

A key technology that plays fundamental role in spatial audio delivery is perceptual audio coding [2-5]. Based on knowledge on psycho-acoustic, perceptual audio coding has been developed to be capable of incredibly compressing the size of audio data in that the audio signal properties which would not detected by our hearing system are just removed. It started in around 1990s when MPEG-1 layer 3, known as MP3 [6-8], the first and most popular digital audio compression standard was introduced. Until recently, numerous audio coding techniques have been invented and standardized which include Spatial Audio Coding (SAC) [9-10], a technology to efficiently represent multi-channel audio signals.

This paper reviews the basic concept of spatial audio coding technologies and standards. The benefits of applying SAC technique is not only the capability to provide high compression ratio compared to conventional multi-channel audio coding approaches but also the opportunities to employ this technology to legacy broadcasting system. Supported by the modern digital signal processing method, MPEG Surround as an SAC-based standard has attracted much more attention due to its rich-functionalities such as binaural rendering and artistic stereo down-mix. Moreover, when object-based audio is introduced, users are offered an option to interact to update the audio scene composition and spatial characteristic of the rendered surround sound making the audio rendering system much more interesting. Teleconference, karaoke system, gaming, dialog enhancement, sports broadcasting, and music re-composition are among the applications that are highly recommended for applying this object-based technology [11-14].

Received date 2017-01-23, Revised date 2017-02-25, Accepted date 2017-02-28
DOI : 10.20449/jnte.v6i1.372

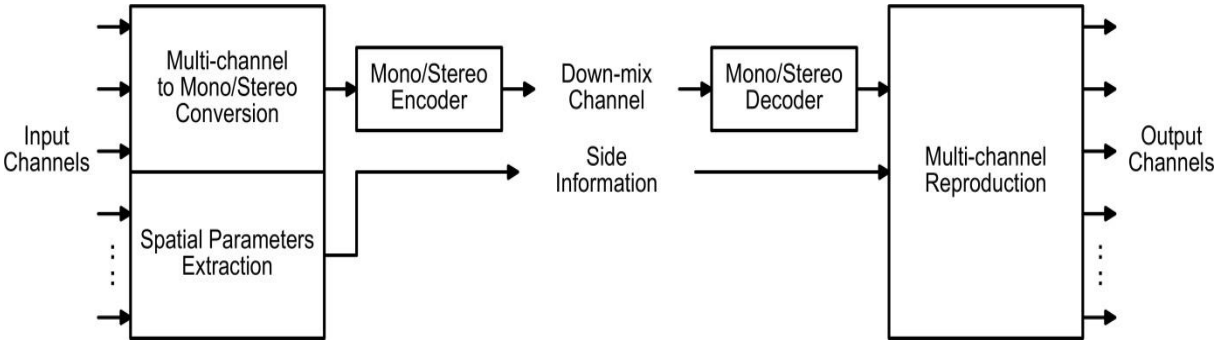

Figure 1. Diagram block of spatial audio coding. At the left side, multi-channel audio signals are represented as mono/stereo down-mixed signals to much reduce the audio data. Spatial parameters must be estimated to reconstructed multi-channel audio signals in decoder as shown at the right side of this figure.

2. PRINCIPLE OF SPATIAL AUDIO CODING

Spatial Audio Coding (SAC) is not a pure compression method. Instead, it can be considered as a technique to represent multichannel audio signals as a lower number of channels, such as one (mono) or two (stereo), while maintaining the spatial properties of the audio signals. The process to reduce the number of audio channels, as illustrated in Fig. 1, is typically termed as down-mixing. In order to be capable of reconstructing the original multichannel audio signals at a later time, a set of spatial parameters should be estimated and kept as additional data of the down-mix audio signal. For the purpose of transmission and storage, the down-mix signal must be encoded by an existing compression technique, such as MPEG-1 layer 3, Advanced Audio Coding (AAC) [15-18], and Universal Speech and Audio Coding (USAC) [19-22], while the spatial parameters can be considered as side information. This allows reconstruction of multichannel audio signals, when needed later, by extending the down-mix signal with guidance of the spatial parameters. However, it is not always necessary to reconstruct multichannel audio signals particularly when user equipment only supports the down-mix audio rendering where in this case the spatial parameters can be simply removed.

The ability to reduce the number of audio channels is the main advantage of this technology, facilitating less number of audio signals needed to be compressed. This is important especially when compared to the conventional multichannel audio coding approach where every channel needs to be encoded separately. It creates two opportunities: first, to represent audio data as a fewer bits as possible; second, to reproduce channel configuration that is different from the original multichannel format. More advantage of SAC is a backward compatibility which makes possible to gradually upgrade existing mono or stereo audio broadcasting systems to have multichannel audio content. Users with old equipment still enjoy mono or stereo audio content as usual while other users, who have new multichannel decoder, can render a more realistic spatial audio content.

Various approaches have been proposed to efficiently encode multichannel audio signals based on the principle of SAC. In general, based on the approach to extract spatial parameters, all of them can be classified as two groups: channel-relationship based and virtual source position based approaches. On one hand, approaches in the first group, which include Parametric Stereo (PS) [23-25], Binaural Cue Coding (BCC) [26-29], and MP3 Surround [30-32], make use of relationship among audio channels to represent spatial properties of the audio scene. Level and time differences as well as coherences, either between two particular channels or among multiple channels, are examples of the extracted spatial parameters. When recreating multichannel audio signals, the spatial parameters will be used to keep channel relationships remain the same as the original ones. On another hand, approaches in the second group associate a virtual audio source for the audio scene. The position of the virtual audio

source is represented by a direction vector which can be determined from any configuration of multichannel audio signals. Spatial Audio Scene Coding (SASC) [33-40] and Directional Audio Coding (DirAC) [41-46] are among the SAC approaches that can be included in the second group. To reproduce multichannel audio signals, Vector Based Amplitude Panning (VBAP) technique [47-50] is applied to reproduce sound field from virtual sources located in positions that given by the direction vector.

2.1. Psycho-Acoustic of Spatial Sound

Phenomenon on human hearing system, such as threshold in quiet and frequency/temporal masking, have been extensively exploited in earlier audio coding approaches. However, SAC technique takes into account more cues particularly with respect to human hearing ability to perceive spatial characteristic of sound wave. Since having two ears, human brain can detect inter-aural time and level differences of sound wave coming from a sound source position which is then made use to form perception on the position of the sound source. Based on these major cues, human can localize an incoming sound source without seeing even though more cues, such as head movement, must be considered for more complex situation.

2.2. Historical Perspective

If the terminology of multi-channel audio is generalized to include stereo (two channels) then the SAC technique started to develop with the introduction of parametric based encoding of stereo audio signals such as Intensity Stereo (IS) and Mid-Side (MS) stereo coding, leading to a more efficient Parametric Stereo (PS) technique which later adopted as an MPEG standard. Using Parametric Stereo approach, stereo audio signals are analyzed to compute three stereo parameters: Inter-channel Intensity Difference (IID), Inter-channel Phase Difference (IPD), and Inter-channel Coherence (IC), which are basically determined based on inter-aural time difference and inter-aural time difference of the human hearing system [51]. Then, stereo audio signals are down-mixed and encoded further by AAC standard. In an attempt to extend the approach to a larger number of audio channels, Binaural Cue Coding (BCC) and MP3 Surround have been proposed. BCC is more general where different audio encoder can be used to compress down-mix signals while MP3 Surround is specifically intended to extend the usage of MP3 encoder, as a widely used codec, to multi-channel audio configuration.

2.3. Proposed SAC Techniques

Numerous encoding methods have been proposed with different spatial parameters. Spatial Audio Scene Coding (SASC) extracts direction vector as spatial parameter making it possible as a three dimensional representation of spatial sound. For audio reproduction, Vector Based Amplitude Panning (VBAP) technique is applied to reproduce sound field from virtual sources located in positions that given by the direction vector. Another approach, called Directional Audio Coding (DirAC), is also proposed to use direction vector as spatial parameters, however, it includes technology of microphone array [52-55] to the process of audio recording and production. In addition, diffuseness is also transmitted as a parameter describing the ambient sound characteristics.

An approach trying to take advantage of panning technique has also been proposed for compressing multi-channel audio signals [56-60]. It can be considered as a method to squeeze a 360 degree sound field into a lower number such as 60 degree. This technique basically does not need spatial parameters especially for low bit rates operation. However, as reverse panning technique is used to reproduce multi-channel audio signals, the localization of audio sources seems to cause ambiguity.

2.4. Improving Reconstructed Audio Accuracy

Closed-loop SAC method [61-65] has been introduced to improve accuracy of the rendered audio signals. This method aims to improve further performance of any SAC technique by minimizing error and distortion caused by the processes of encoding and quantisation. It can basically be considered as a minimization method applied on top of SAC technique. To support this closed-loop configuration, a number approaches have also been proposed that include balanced-delay filter-bank, integrated residual coding, and frequency domain based SAC.

Analysis by Synthesis (AbS) concept [66-71], which has been widely applied in many

applications, has also been employed in this technique to perform a trial and error procedure to improve the quality of reconstructed audio signals. Even though waveform based error criteria is used to compare the reproduced and the original audio signals, however, since the AbS algorithm is carried out on top of perceptual-based system, improved perceptual quality has been reported.

2.5. Quality Measurement of Reproduced Audio Signal

In general, audio subjective test [72-79] is considered as the only valid method for assessing the quality of reconstructed audio signal. This is because objective test particularly conventional method, such as signal to noise ratio, cannot detect artifacts [80] introduced in perceptual audio coding. However, perceptual based objective test such as Perceptual Evaluation of Audio Quality (PEAQ) [81] is recommended as secondary assessment for perceptual audio coding. Moreover, as in real time systems assessment using subjective test is not possible, perceptual based objective test must be further developed especially for multi-channel audio reproduction.

3. MPEG STANDARD FOR ENCODING MULTI-CHANNEL AUDIO SIGNALS

In this section three international standards for encoding multi-channel audio signals will be discussed which include MPEG Surround [81-90], MPEG SAOC [91-92] and MPEG-H 3D Audio [93]. Key technologies applied in several MPEG standards are given in Table 1 where in general, it can be seen that MPEG Surround applies channel-based system, MPEG SAOC applies object-based system, and MPEG-H 3D Audio Coding applies three-dimensional audio reproduction technology. Furthermore hybrid filter-bank and decorrelator [94-95] are employed in MPEG standards. Filter-bank is very useful to decompose audio signal and to process audio signal in critical band, the same manner as it is done in our hearing system. With hybrid system, it is possible to provide sub-band signal with different frequency resolution. On the other hand, decorrelator helps to improve spatial effects of reproduced audio signals particularly for low bit rate implementation.

3.1. MPEG Surround

MPEG Surround standard, released in 2009, can be considered as the first international standard taking advantages of the concept of spatial audio coding. Multi-channel audio signals can be represented as mono, stereo, or 5.1 down-mix signals even though the use of stereo down-mix is mostly reported. Perceptual based spatial parameters, consisted of Channel Level Differences (CLD), Inter-Channel Coherence (ICC), and Channel Prediction Coefficient (CPC), are applied. In case of high bit rate operation is possible, residual signal can be produced and included in the spatial parameter bit-stream to compensate for error due to down-mix and up-mix processes enabling waveform reconstruction of the audio signals. Otherwise, MPEG Surround decoder is equipped with decorrelator to produce synthetic residual signal which is able to give more spatial effects in the reproduce audio signals. As an SAC-based standard, MPEG Surround is compatible to any legacy audio codec, for encoding the down-mixed signals. However, it is well-combined with High Efficiency AAC (HE-AAC) [96-97] since both MPEG Surround and HE-AAC use the same hybrid filter-bank to perform time-frequency decomposition of the audio signal.

Formal subjective tests show that high quality audio reproduction are achieved at very low bit rates such as 64 and 96 kb/s [98-99] while operation at higher bit rates up to 256 kb/s increases performance as reported in [100]. In terms of functionality, MPEG Surround also supports binaural rendering [101] where mobile users can enjoy multi-channel audio reproduction through headphone. In order to keep existing users, who have stereo decoder only without multi-channel decoder, receive the best stereo content, MPEG Surround encoder also has capability of producing artistic stereo down-mix signals.

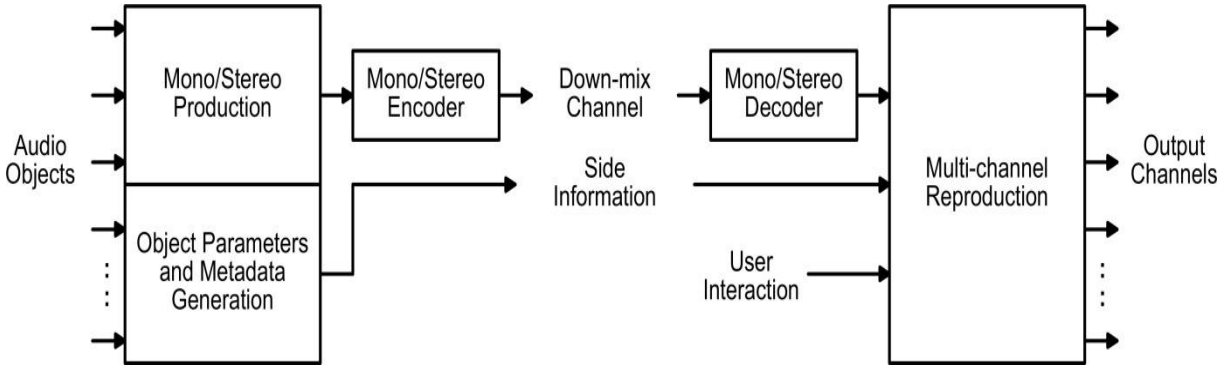

Figure. 2. Block diagram of spatial audio object coding technique, offering users capability to update and adjust audio composition.

3.2. MPEG Spatial Audio Object Coding

New approach for multi-channel audio reproduction is introduced with a concept of object-based audio [102-105]. It essentially differs from the conventional channel-based audio in that, as illustrated in Fig. 2, users are given opportunities to interact with the multi-channel reproduction system to update the reconstructed audio composition. For instance in the news broadcasting, users can adjust the anchor volume only while keeping the other background sounds remains as they were. This feature provides music composers to change existing music composition based on their own preferences. It also offers movie watchers and game players to update spatial effects of the reproduced sound with respect to the displayed video. In order to apply this approach, the encoder side needs to capture every audio object that is desired to be possible for adjustment in the decoder side. Moreover, object parameters as well as metadata, such as object audio source position, must be generated and transmitted to decoder. These parameters and metadata must be able to be utilized to reproduce multi-channel audio signals without any user helps at the decoder side.

MPEG SAOC standard is developed by exploiting the principle of object-based audio reproduction. Interestingly, MPEG SAOC makes use of MPEG Surround decoder to reproduce multi-channel audio reproduction by providing conversion technique from object parameters and metadata to channel-based spatial parameters. To improve performance and usability of MPEG SAOC, some approaches have been proposed. In [106], MPEG SAOC is combined with DirAC to provide spatial teleconference system while [107] introduces two-step coding structure to improve performance of every rendered audio object. To increase vocal removal performance of MPEG SAOC on music re-composition application, a harmonic information can also be transmitted to the decoder side [108-109].

3.3. MPEG-H 3D Audio Coding

This new MPEG standard is aimed at supporting more audio reproduction system, such as 10.2 [110] and 22.2 [111-115] audio system as well as larger number of loudspeakers when wave field synthesis (WFS) [116-117] technique is applied, thus, a true three dimensional (3D) audio reproduction becomes possible. On one hand, those larger number of loudspeaker setup are special since it is also intended to reproduce higher sound field than the listener position through the elevated loudspeakers. On another hand, WFS technique takes much more attention because it is able to reconstruct realistic audio wave-front by employing a large number of loudspeakers. Moreover, any audio format sent from encoder side can be reproduced to different loudspeaker set up. For encoding down-mixed signals, MPEG Unified Speech and Audio Coding (USAC) standard, that is more powerful in encoding both speech and audio signals, is recommended. Performance of this MPEG 3D standard has been tested at very high bit rates such as 512 and 1200 kb/s providing satisfying 3D audio perception.

Table 1. Summary of key technologies applied in MPEG spatial audio standards

No	Standard	Reproduction Mode	Key Features	Main Applications
1.	MPEG Surround	Channel-based audio	Binaural rendering Artistic stereo downmix	Audio broadcasting Teleconference
2.	MPEG SAOC	Object-based audio	MPEG Surround transcoder	Music recomposition Gaming, Karaoke
3.	MPEG-H 3D Audio	3D audio	Higher order ambisonic	Ultra HDTV

4. FUTURE DIRECTION

The future development of spatial audio coding seems to be directed to have a universal and scalable approach. Universal in terms of compatibility to various method and standards which include numerous audio reproduction setup, while scalable can be considered as its capability to transmit variable size of audio data which depends on the network condition and end user audio equipment.

Providing universal and scalable audio coding system is essential because there are standards currently available and implemented in different audio applications. For instance, Parametric Stereo standard is the most efficient technique intended to stereo application at a recommended stereo bit rate of 24 kb/s. However, increasing the bit rate seems to be not affecting in improving the audio quality making Parametric Stereo cannot compete with MP3 or AAC at bit rates higher than 24 kb/s. For 5.1 audio configurations, even though MPEG Surround can be the best option to operate at low bit rates such as 64 and 96 kb/s, transmitting 5.1 audio signals will be much more efficient when higher bit rates, such as 320 kb/s, is operated. In terms of this operating bit rate, the future SAC system is expected to be able to efficiently represent multi-channel audio signals at all bit rates and channel configurations avoiding the need to employ different audio coding standard. Modern SAC approach is also expected to consider the configuration of audio reproduction system. Home consumers prefer to have simple and portable audio equipment. Offering equipment with a large number of loudspeakers to home customers is apparently not the best choice. Considering this consumer demand various ways of flexible audio rendering system need to be addressed.

Another area requiring further exploration in spatial audio coding is low delay audio codec for conversation. Most of MPEG standard have relatively longer delay than required in a standard two way communication system. Thus, some approaches have been proposed for low delay audio codec such as in [118-119]. It is also interesting to look at approaches based on speech coding, such as in [110-125], to extend their capability to have full audio band and multi-channel speech codec.

5. CONCLUSION

This paper has presented an overview on technology of spatial audio coding. Basic concept, proposed techniques, as well as various spatial parameters have been discussed. Moreover, the principle of channel-based and object-based audio have been highlighted. MPEG standards that include MPEG Surround, MPEG SAOC, and MPEG-H 3D Audio Coding, developed for delivering spatial audio, have also been explored. At the end of the paper, future direction of spatial audio coding research potential is discussed with focus on universal and scalability as well as low-delay requirement.

ACKNOWLEDGMENT

This work was funded by the Ministry of Research, Technology and Higher Education, the Republic of Indonesia under the scheme of PUPT in 2017. The authors would like to thank the reviewers for their constructive comments and suggestions to improve the quality of the manuscript. Furthermore, the authors also thank Dr. Rahmadi Kurnia, Mr. Heru Dibyo Laksono and Mrs. Fitrilina, who were also work together in the research project, for meaningful suggestions during the writing up of the manuscript.

REFERENCE

[1] F. Rumsey, Spatial Audio, 2nd Edition, Focal Press, Oxford, England, 2001.

[2] D. Pan, A tutorial on MPEG/audio compression, IEEE Multimedia No. 2 (1995) 60–72.

[3] T. Painter, A. Spanias, Perceptual coding of digital audio, Proceedings of the IEEE 88 (4) (2000) 451–513.

[4] M. Bosi, R. E. Goldberg, Introduction to Digital Audio Coding and Standards, Springer, New York, USA, 2002.

[5] K. Brandenburg, C. Faller, J. Herre, J. D. Johnston, W. B. Kleijn, Perceptual coding of high-quality digital audio, Proceedings of the IEEE 101 No. 9 (2014) 1905–1919.

[6] K. Brandenburg, G. Stoll, Y. Dehery, J. Johnston, L. Kerkhof, E. Schroder, ISO-MPEG-1 audio: A generic standard for coding of high-quality digital audio, J. Audio Engineering Society 42 No. 10 (1994) 780–792.

[7] K. Brandenburg, MP-3 and AAC explained, Presented at AES 17th Int. Conf. on High Quality Audio Coding (September 1999).

[8] H. G. Musmann, Genesis of the MP3 audio coding standard, IEEE Transactions on Consumer Electronics 52 (2006) 1043–1049.

[9] J. Herre, C. Faller, S. Disch, C. Ertel, J. Hilpert, A. Hoelzer, K. Linzmeier, C. Spenger, P. Kroon, Spatial audio coding: Next-generation efficient and compatible coding of multi- channel audio, in: Proc. the 117th Convention of the Audio Engineering Society, San Fransisco, CA, USA, 2004.

[10] J. Herre, From joint stereo to spatial audio coding - recent progress and standardization, in: Proc. of the 7th Int. Conf. on Digital Audio Effects (DAFx'04), Naples, Italy, 2004.

[11] R. Oldfield, B. Shirley, J. Spille. An object based audio system for interactive Broadcasting, Presented at the 137th Convention of the Audio Engineering Society (October 2014).

[12] R. Oldfield, B. Shirley, D. Satongar. Application of object-based audio for automated mixing of live football broadcast, Presented at the 139th Convention of the Audio Engineering Society (November 2015).

[13] J. Jot, B. Smith, J. Thompson. Dialog Control And Enhancement in Object-Based Audio Systems, Presented at the 139th Convention of the Audio Engineering Society (November 2015).

[14] R. Bleidt, A. Borsum, H. Fuchs, and S. M. Weiss, "Object-Based Audio: Opportunities for Improved Listening Experience and Increased Listener Involvement," Motion Imaging Journal, SMPTE, vol. 124, no. 5, pp. 1–13, 2015.

[15] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, Y. Oikawa, ISO/IEC MPEG-2 Advanced Audio Coding, J. Audio Eng. Soc. 45 No. 10 (1997) 789–814.

[16] K. Brandenburg, M. Bosi, ISO/IEC MPEG-2 advanced audio coding: Review and applications", in: AES 103rd Convention, New York, USA, 1997.

[17] ISO/IEC, Information Technology - Generic coding of moving pictures and associated audio information, Part 7: Advanced Audio Coding, ISO/IEC 13818-7:2006(E), Int. Standards Organization, Geneva, Switzerland (2006).

[18] ISO/IEC, Information Technology - Coding of audio-visual objects, Part 3: Audio, ISO/IEC 14496-3:2009(E), International Standards Organization, Geneva, Switzerland (2009).

[19] S. Quackenbush, R. Lefebvre, Performance of MPEG unified speech and audio coding, Presented at the 131st Convention of the Audio Engineering Society (October 2011).

[20] ISO/IEC, Information Technology - MPEG Audio Technologies, Part 3: Unified Speech and Audio Coding, ISO/IEC 23003-3/FDIS, Int. Standards Organization, Geneva, Switzerland (2012).

[21] M. Nuendorf, et al, The ISO/MPEG unified speech and audio coding standard - consistent high quality for all content types and at all bit rates, Journal of Audio Engineering Society 61 (12) (2013) 956–977.
[22] E. Oh, M. Kim, Enhanced stereo algorithm in the unified speech and audio coding, in: AES 43rd International Conference, Pohang, Korea, 2011.
[23] J. Breebaart, S. van de Par, A. Kohlrausch, E. Schuijers, Parametric coding of stereo audio, EURASIP Journal Applied Signal Processing (2005) 1305–1322.
[24] E. Schuijers, J. Breebaart, H. Purnhagen, J. Engdegard, Low complexity parametric stereo coding, Presented at the 116th Convention of the Audio Engineering Society (May 2004).
[25] E. Schuijers, W. Oomen, B. den Brinker, J. Breebaart, Advances in parametric coding for high-quality audio, Presented at the 114th Convention of the Audio Engineering Society (Mar. 2003).
[26] F. Baumgarte, C. Faller, Why binaural cue coding is better than intensity stereo coding, Presented at the 112th Convention of the Audio Engineering Society (May 2002).
[27] F. Baumgarte, C. Faller, Binaural cue coding-part I: Psychoacoustic fundamentals and design principles, IEEE Trans. Speech Audio Process. 11 (6) (2003) 509–519.
[28] C. Faller, F. Baumgarte, Binaural cue coding-Part II: Schemes and applications, IEEE Trans. Speech Audio Process. 11 (6) (2003) 520–531.
[29] C. Faller, F. Baumgarte, Binaural cue coding-Part II: Schemes and applications, IEEE Trans. Speech Audio Process. 11 (6) (2003) 520–531.
[30] S. B. Chon, I. Y. Choi, H. G. Moon, J. Seo, K.-M. Sung, Virtual source location information for binaural cue coding, in: Proc. 123th AES Convention, New York, USA, 2005.
[31] [26] J. Herre, C. Faller, C. Ertel, J. Hilpert, A. Hoelzer, C. Spenger, MP3 Surround: efficient and compatible coding of multi-channel audio, Presented at the 116th Convention of the Audio Engineering Society (May 2004).
[32] B. Grill, O. Hellmuth, J. Hilpert, J. Herre, J. Plogsties, Closing the gap between the multichannel and the stereo audio world: Recent mp3 surround extensions, in: Proc. the 120th Convention of the Audio Engineering Society, Paris, France, 2006.
[33] H. Moon, A low-complexity design for an mp3 multichannel audio decoding system, IEEE Trans. on Audio, Speech, and Lang. Proc. 20 (1) (2012) 314–321.
[34] M. M. Goodwin, J.-M. Jot, A frequency domain framework for spatial audio coding based on universal spatial cues, Presented at the 120th Convention of the Audio Engineering Society (May 2006).
[35] M. M. Goodwin, J.-M. Jot, Analysis and synthesis for universal spatial audio coding, Presented at the 121th Convention of the Audio Engineering Society (Oct. 2006).
[36] J.-M. Jot, J. Merimaa, M. M. Goodwin, A. Krishnaswamy, J. Laroche, Spatial audio coding in a universal two-channel 3D stereo format, in: Proc. 123rd AES Convention, New York, USA, 2007.
[37] M. M. Goodwin, J.-M. Jot, Multichannel surround format conversion and generalized upmix, in: Proc. AES 30th Int. Conf., Saariselka, Finland, 2007.
[38] M. M. Goodwin, J.-M. Jot, Primary-ambient signal decomposition and vector-based localization for spatial audio coding and enhancement, in: Proc. IEEE Intl. Conf. Acoustic, Speech, Signal Processing, Honolulu, Hawaii, USA, 2007.
[39] M. M. Goodwin, J.-M. Jot, Binaural 3-D audio rendering based on spatial audio scene coding, in: Proc. 123rd AES Convention, New York, USA, 2007.

[40] M. M. Goodwin, J.-M. Jot, Binaural 3-D audio rendering based on spatial audio scene coding, in: Proc. 123rd AES Convention, New York, USA, 2007.
[41] M. M. Goodwin, J.-M. Jot, Spatial audio scene coding, Presented at the 125th Convention of the Audio Engineering Society (October 2008).
[42] V. Pulkki, C. Faller, Directional audio coding: Filter bank and STFT-based design, Presented at the 120th Convention of the Audio Engineering Society (May 2006).
[43] V. Pulkki, Directional audio coding in spatial sound reproduction and stereo up-mixing, in: Proc. Audio Engineering Society 28th Intl. Conf., Pitea, Sweden, 2006.
[44] V. Pulkki, M. Karjalainen, Multichannel audio rendering using amplitude panning, IEEE Signal Processing Mag. 25 (3) (2008) 118–122.
[45] J. Vilkamo, T. Lokki, V. Pulkki, Directional Audio Coding: Virtual microphone-based synthesis and subjective evaluation, Journal Audio Eng. Soc. 57 (9) (2009) 709–724.
[46] J. Ahonen, V. Pulkki, T. Lokki, Teleconference application and B-format microphone array for directional audio coding, in: Proc. AES 30th International Conference, Saariselka, Finland, 2007.
[47] V. Pulkki, Virtual sound source positioning using vector based amplitude panning, J. Audio Eng. Soc. 45 (6) (1997) 456–466.
[48] V. Pulkki, M. Karjalainen, V. Vesa, Localization, coloration and enhancement of amplitude-panned virtual sources, in: Proc. AES 16th International Conference, Rovaniemi, Finland, 1999
[49] V. Pulkki, Compensating displacement of amplitude panned virtual sources, in: Proc. AES 22th Int. Conf. on Virtual, Synthetic, and Entertainment Audio, Espoo, Finland, 2002.
[50] V. Pulkki, Compensating displacement of amplitude panned virtual sources, in: Proc. AES 22th Int. Conf. on Virtual, Synthetic, and Entertainment Audio, Espoo, Finland, 2002.
[51] V. Pulkki, T. Hirvonen, Localization of virtual sources in multi-channel audio reproduction, IEEE Transaction on Speech and Audio Processing 13 no. 1 (2005) 105–119.
[52] J. Blauert, Spatial Hearing, The Psychophysics of Human Sound Localization, MIT Press, Cambridge, MA, 2001.
[53] R. S-Amling, F. Kuech, M. Kallinger, G. D. Galdo, J. Ahonen, V. Pulkki, Planar microphone array processing for the analysis and reproduction of spatial audio using directional audio coding, Presented at the 124th Convention of the Audio Engineering Society (May 2008).
[54] F. Kuech, M. Kallinger, R. S-Amling, G. del Galdo, J. Ahonen, V. Pulkki, Directional audio coding using planar microphone arrays, in: Proc. Hands-free Speech Communication and Microphone Arrays (HSCMA), Trento, Italy, 2008.
[55] M. Kallinger, F. Kuech, R. S-Amling, G. del Galdo, J. Ahonen, V. Pulkki, Enhanced direction estimation using microphone arrays for directional audio coding, in: Proc. Hands-free Speech Communication and Microphone Arrays (HSCMA), Trento, Italy, 2008.
[56] J. Ahonen, V. Pulkki, F. Kuech, M. Kallinger, R. S-Amling, Directional analysis of sound field with linear microphone array and applications in sound reproduction, in: Proc. 124th AES Convention, Amsterdam, The Netherlands, 2008.
[57] B. Cheng, C. Ritz, I. Burnett, Advances in Multimedia Information Processing 2006, Springer, Berlin, Heidelberg, 2006, Ch. Squeezing the Auditory Space: A New Approach to Multichannel Audio Coding, pp. 572–581.
[58] B. Cheng, C. Ritz, I. Burnett, Principles and analysis of the squeezing approach to low bit rate

spatial audio coding, in: Proc. IEEE Intl. Conf. Acoustic, Speech, Signal Processing, Honolulu, Hawaii, USA, 2007.

[59] B. Cheng, C. Ritz, I. Burnett, A spatial squeezing approach to Ambisonic audio compression, in: Proc. IEEE Intl. Conf. Acoust. Speech, Signal Process., Las Vegas, Nevada, USA, 2008

[60] E. Cheng, B. Cheng, C. Ritz, I. Burnett, Spatialized teleconferencing: Recording and squeezed rendering of multiple distributed sites, in: Proc. Australian Telecom. Network and Appl. Conf., Adelaide, Australia, 2008.

[61] I. Elfitri, R. Kurnia, Fitrilina, Investigation on objective performance of closed-loop spatial audio coding, in: Proc. of 2014 Int. Conf. on Information Tech. and Electrical Eng., Jogjakarta, Indonesia, 2014.

[62] I. Elfitri, R. Kurnia, D. Harneldi, Experimental study on improved parametric stereo for bit rate scalable audio coding, in: Proc. of 2014 Int. Conf. on Information Tech. and Electrical Eng., Jogjakarta, Indonesia, 2014.

[63] I. Elfitri, M. Muharam, M. Shobirin, Distortion analysis of hierarchical mixing technique on MPEG surround standard, in: Proc. of 2014 Int. Conf. on Advanced Computer Sciences and Information System, Jakarta, Indonesia, 2014.

[64] I. Elfitri, H. D. Laksono, A. Permana, Balanced-delay filterbank for closed-loop spatial audio coding, in: Proc. of 2015 Int. Conf. on Intelligent Tech. and Its Applications, Surabaya, Indonesia, 2015.

[65] I. Elfitri, A. Luthfi, Fitrilina, BR-TTT module with modified residual signal for improving multichannel audio signal accuracy, in: Proc. of 2015 Int. Conf. on Autiomation, Cognitice Science, Optics, Micro Electro-Mechanical Syatem, and Information Tech., Bandung, Indonesia, 2015.

[66] I. Elfitri, X. Shi, A. M. Kondoz, Analysis by synthesis spatial audio coding, IET Signal Processing 8 (1) (2014) 30–38.

[67] I. Elfitri, B. Gunel, A. M. Kondoz, Multichannel audio coding based on analysis by synthesis, Proc. of the IEEE 99 (4) (2011) 657–670.

[68] P. Eisert, Model-based camera calibration using analysis by synthesis techniques, in: Proc. of vision, modeling, and visualization, Erlangen, Germany, 2002.

[69] Z. Yang, M. Jia, C. Bao, W. Wang, An analysis-by-synthesis encoding approach for multiple audio objects, in: Proc. of 2015. Asia-Pasific Signal and Information Processing Association Annual Summit and Conference, Hongkong, 2015.

[70] X. Zeng, C. Ritz, J. Xi, Encoding Navigable Speech Sources: A Psychoacoustic-Based Analysis-by-Synthesis Approach, IEEE Transactions on Audio, Speech, and Language Processing 21 no. 1 (2013) 29–38

[71] X. Zeng, C. Ritz, J. Xi, Encoding Navigable Speech Sources: Analysis-by-Synthesis Approach, in: Proc. of 2012 IEEE Int. Conf. On Acoustics, Speech, and Signal Processing, Kyoto, Japan, 2013

[72] ITU-R, Method for Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems, Recommendation ITU-R BS.1116-1 (1997).

[73] ITU-R, Method for Subjective Assessment of Small Impairments in Audio Systems, Recommendation ITU-R BS.1116-2 (2014).

[74] ITU-R, Method for Subjective Assessment of Small Impairments in Audio Systems, Recommendation ITU-R BS.1116-3 (2015).

[75] ITU-R, Method for Subjective Assessment of Intermediate Quality Level of Coding Systems, Recommendation ITU-R BS.1534 (2001).

[76] ITU-R, Method for Subjective Assessment of Intermediate Quality

Level of Coding Systems, Recommendation ITU-R BS.1534-1 (2003).

[77] ITU-R, Method for Subjective Assessment of Intermediate Quality Level of Audio Systems, Recommendation ITU-R BS.1534-2 (2014).

[78] ITU-R, Method for Subjective Assessment of Intermediate Quality Level of Audio Systems, Recommendation ITU-R BS.1534-3 (2015).

[79] B. Cheng, C. Ritz, I. Burnett, Advances in Multimedia Information Processing-PCM 2007, Springer, Berlin, Heidelberg, 2007, Ch. Encoding Independent Sources in Spatially Squeezed Surround Audio Coding, pp. 804–813.

[80] C. Liu, H. Hsu, W. Lee, Compression artifacts in perceptual audio coding, IEEE Trans. on Audio, Speech, and Lang. Proc. 16 (4) (2008) 681–695.

[81] ITU-R, Method for Objective Measurements of Perceived Audio Quality, Recommendation ITU-R BS.1387-1 (2001).

[82] S. Quackenbush, J. Herre, MPEG Surround, IEEE Trans. On Multimedia 12 (4) (2005) 18–23.

[83] J. Breebaart, J. Herre, C. Faller, J. Roden, F. Myburg, S. Disch, H. Purnhagen, H. G, M. Neusinger, K. Kjorling, W. Oomen, MPEG spatial audio coding/ MPEG surround: Overview and current status, Presented at the 11th Convention of the Audio Engineering Society (October 2005).

[84] J. Breebaart, G. Hotho, J. Koppens, E. Schuijers, W. Oomen, S. V. de Par, Background, concepts, and architecture for the recent MPEG Surround standard on multichannel audio compression, J. Audio Eng. Soc. 55 (2007) 331–351.

[85] J. Herre, H. Purnhagen, J. Breebaart, C. Faller, S. Disch K. Kjorling, E. Schuijers, J. Hilpert, F. Myburg, The reference model architecture for MPEG spatial audio coding, Presented at the 118th Convention of the Audio Engineering Society (May 2005).

[86] J. Herre, et al., MPEG Surround - The ISO/MPEG standard for efficient and compatible multichannel audio coding, ournal Audio Engineering Society 56 (11) (2008) 932–955.

[87] J. Hilpert and S. Disch, "The MPEG Surround audio coding standard [Standards in a nutshell]," IEEE Signal Processing Mag., vol. 26, no. 1, pp. 148–152, Jan. 2009.

[88] ISO/IEC, "Information Technology - MPEG Audio Technologies, Part 1: MPEG Surround," ISO/IEC 23003-1:2007(E), International Standards Organization, Geneva, Switzerland, 2007.

[89] S. Samsudin, E. Kurniawati, and S. George, "A direct MPEG surround encoding scheme for surround sound recording with coincidence microphone techniques," in AES 55th International Conference, Helsinki, Finland, August 2014.

[90] C. Tournery, C. Faller, F. Kuech, and J. Herre, "Converting stereo microphone signals directly to MPEG surround," in Proc. 128th AES Convention, London, UK, May 2010.

[91] J. Engdegard et al., "Spatial audio object coding (SAOC)-The upcoming MPEG standard on parametric object based audio coding," Presented at the 124th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, May 2008.

[92] L. Terentiev, C. Falch, O. Hellmuth, J. Hilpert, W. Oomen, J. Engdegard, and H. Mundt, "SAOC for gaming-the upcoming MPEG standard on parametric object based audio coding," in Proc. AES 35th Int. Conference, London, UK, Feb. 2009.

[93] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H Audio The new standard for universal spatial/3D audio coding," J. Audio Eng. Soc., vol. 62, no. 12, pp. 821–830, 2015.

[94] J. Engdegard, H. Purnhagen, J. Roden, and L. Liljeryd, "Synthetic ambience in parametric stereo coding," Presented at the 116th Convention of the Audio

Engineering Society, Berlin, Germany, May 2004.

[95] D. P. Chen, H. F. Hsiao, H. W. Hsu, and C. M. Liu, "Gram-schmidt-based downmixer and decorrelator in the MPEG surround coding," Presented at the 128th Convention of the Audio Engineering Society, London, UK, May, 2010.

[96] M. Wolters, K. Kjorling, D. Homm, and H. Purnhagen, "A closer look into MPEG-4 high efficiency AAC," in Proc. the 115th Convention of the Audio Engineering Society, New York, USA, October 2003.

[97] J. Herre and M. Dietz, "MPEG-4 high-efficiency AAC coding," IEEE Signal Proc. Mag., vol. 25, no. 3, pp. 137–142, 2008.

[98] A. Mason, D. Marston, F. Kozamernik, and G. Stoll, "EBU test of multichannel audio codecs," Presented at the 122nd Convention of the Audio Engineering Society, Vienna, Austria, May, 2007.

[99] D. Marston, F. Kozamernik, G. Stoll, and G. Spikofski, "Further EBU test of multichannel audio codecs," Presented at the 126th Convention of the Audio Engineering Society, Munich, Germany, May, 2009.

[100] J. Roden, J. Breebart, J. Hilpert, H. Purnhagen, E. Schuijers, J. Koppens, K. Linzmeier, and A. Holzer, "A study of the MPEG Surround quality versus bit-rate curve," Presented at the 123rd Convention of the Audio Engineering Society, New York, USA, Oct. 2007.

[101] J. Breebaart, J. Herre, L. Villemoes, C. Jin, K. Kjorling, J. Plogsties, and J. Koppens, "Multi-channel goes mobile: MPEG Surround binaural rendering," in Proc. the AES 29th Int. Conference, Seoul, Korea, September 2006.

[102] J. Herre and S. Disch, "New concepts in parametric coding of spatial audio: From SAC to SAOC," in Proc. IEEE Int. Conf. on Multimedia and Expo, San Fransisco, CA, USA, Oct. 2007.

[103] J. Herre and L. Terentiv, "Parametric coding of audio objects: Technology, performance, and opportunities," Presented at the 42nd Int. Conference: Semantic Audio, Ilmenau, Germany, July 2011.

[104] S. Gorlow, E. A. P. Habets, and S. Marchand, "Multichannel object-based audio coding with controllable quality," in Proc. 2013 IEEE Int. Conf. Acoustics, Speech and Signal Proc., Vancouver, Canada, June 2013.

[105] S. Fug, A. Holzer, C. Borb, C. Ertel, M. Kratschmer, and J. Plogsties, "Design, coding, and processing of metadata for object-based interactive audio," in Proc. 137th AES Convention, Los Angeles, USA, Oct. 2014.

[106] J. Herre, C. Falch, D. Mahne, G. del Galdo, M. Kallinger, and O. Thiergart, "Interactive teleconferencing combining spatial audio object coding and DirAC technology," Presented at the 128th Convention of the Audio Engineering Society, London, UK, May 2010.

[107] K. Kim, J. Seo, S. Beack, K. Kang, and M. Hahn, "Spatial audio object coding with two-step coding structure for interactive audio service," IEEE Trans. on Multimedia, vol. 13, no. 6, pp. 1208–1216, December 2011.

[108] J. Park, K. Kim, M. Hahn, Vocal removal from multichannel audio using harmonic information for karaoke Service, IEEE Transactions on Audio, Speech, and Language Processing 21 no. 4 (2013) 798–805

[109] J. Park, K. Kim, M. Hahn, "Harmonic Elimination structures for Karaoke Mode in Spatial Audio Object Coding Scheme," in Proc. 2011 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, USA, Jan 2011.

[110] S. Kim, Y. Lee, and V. Pulkki, "New 10.2-channel vertical surround system (10.2-vss); comparison study of perceived audio quality in various multichannel sound systems with height loudspeakers," Presented at the 129th Convention of the Audio Engineering Society, San Fransisco, USA, November 2010.

[111] K. Hamasaki, T. Nishiguchi, R. Okumura, and Y. Nakayama, "Wide listening area with exceptional spatial sound quality of a 22.2 multichannel sound system," Presented at the 122nd Convention of the Audio Engineering Society, Vienna, Austria, May, 2007.

[112] T. Sugimoto, Y. Nakayama, and S. Oode, "Bitrate of 22.2 multichannel sound signal meeting broadcast quality," in Proc. 137th AES Convention, Los Angeles, USA, Oct. 2014.
[113] T. Nishiguchi and et al, "Production and live transmission of 22.2 multichannel sound with ultrahigh-definition TV," in Proc. the 122nd AES Convention, Vienna, Austria, May 2007.
[114] K. Matsui and A. Ando, "Binaural reproduction of 22.2 multichannel sound with loudspeaker array frame," in Proc. the 135th AES Convention, New York, USA, Oct. 2013.
[115] K. Hamasaki, "The 22.2 multichannel sounds and its reproduction at home and personal environment," in AES 43rd International Conference, Pohang, Korea, Sep. 2011.
[116] A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," J. Acoust. Soc. Am., vol. 93, no. 5, pp. 2764–2778, May 1993.
[117] G. Theile, "Wave filed synthesis-a promising spatial audio rendering concept," in Proc. of the 7th Int. Conf. on Digital Audio Effects (DAFX'04), Naples, Italy, Oct. 2004.
[118] J.-M. Valin, T. B. T. C. Montgomery, and G. Maxwell, "A high-quality speech and audio codec with less than 10-ms delay," IEEE Trans. on Audio, Speech, and Language Proc., vol. 18, no. 1, pp. 58–67, January 2010.
[119] R. Chivukula, Y. Reznik, Y. Hu, V. Devarajan, and M. Lakshman, "Fast algorithms for low-delay TDAC filterbanks in MPEG-4 AAC- ELD," IEEE/ACM Trans. on Audio, Speech and Language Processing, vol. 22, no. 12, pp. 1701–1712, 2014.
[120] R. V. Cox, D. C. Neto, C. Lamblin, and M. H. Sherif, "ITU-T coders for wideband, superwideband, and fullband speech communication," IEEE Communications Magazine, vol. 47, no. 10, pp. 106–109, October 2009.
[121] V. Eksler and M. Jelinek, "Coding of unquantized spectrum sub-bands in superwideband audio codecs," in Proc. IEEE Intl. Conf. Acoust. Speech, Signal Processing Prague, Czech Republic, May 2011.
[122] B. Geiser et al., "Candidate proposal for ITU-T super-wideband speech and audio coding," in Proc. IEEE Intl. Conf. Acoust. Speech, Signal Processing Taipei, Taiwan, Apr. 2009.
[123] Y. Hiwasaki et al., "G.711.1: A wideband extension to ITU-T G.711," in Proc. 16th European Signal Processing Conference (EUSIPCO-2008), Lausanne, Switzerland, Aug. 2008.
[124] B. Kovesi, S. Ragot, C. Lamblin, L. Miao, Z. Liu, and C. Hu, "Re-engineering ITU-T G.722: Low delay and complexity superwideband coding at 64 kbit/s with G.722 bitstream watermarking," in Proc. IEEE Intl. Conf. Acoust. Speech, Signal Processing Prague, Czech Republic, May 2011.
[125] S. Ragot et al., "ITU-T G.729.1: An 8-32 kbit/s scalable coder interoperable with G.729 for wideband telephony and voice over IP," in Proc. IEEE Intl. Conf. Acoust. Speech, Signal Processing Honolulu, Hawaii, USA, Apr. 2007.

About the Authors

Ikhwana Elfitri, is a Senior Lecturer in the Dept. of Electrical Engineering, Universitas Andalas, Padang, Indonesia. He has been an invited speaker in both CMOSETR 2015 in Vanouver, Canada, and WCSM 2016, Singapore. His research interest includes audio signal processing, microstrip antenna design, and digital communication system.

Amirul Luthfi, received the B.Sc. degree in Electrical Engineering from Andalas University (Unand), Padang, Indonesia. Currently, he is taking M.Sc. degreee in Electrical Engineering Department at Andalas University. He is a Student Member of the Institute of Electrical and Electronics Engineer (IEEE) since 2015.