STUDI DAN ANALISIS KINERJA MPEG SURROUND PADA BITRATE 256 – 400 kbps

Amirul Luthfi, Fauzan Mustaqim dan Ikhwana Elfitri*

Jurusan Teknik Elektro, Fakultas Teknik, Universitas Andalas

*Corresponding author, e-mail : ikhwana@ft.unand.ac.id

Abstrak— MPEG Surround merupakan salah satu *spatial audio coding* yang populer digunakan. MPEG Surround umumnya digunakan pada *bitrate* rendah yaitu 64, 96, dan 160 kbps. MPEG Surround dapat digunakan pada *bitrate* tinggi, namun kualitas audio yang dihasilkan pada *bitrate* tersebut belum sepenuhnya diketahui. Paper ini menginvestigasi pengaruh kenaikan *bitrate* terhadap kenaikan kualitas audio. Hasil menunjukkan peningkatan kualitas audio yang cukup signifikan saat MPEG Surround digunakan pada *bitrate* tinggi. Dengan demikian MPEG Surround layak digunakan pada *bitrate* tinggi untuk mendapatkan kualitas audio yang lebih baik.

Kata Kunci : *Spatial Audio Coding* dan *MPEG Surround*

Abstract— MPEG Surround is one of the spatial audio codings that popularly used. MPEG Surround is generally used at a low bitrate such as 64, 96 and 160 kbps. MPEG Surround can be used at high bitrates, but the audio quality produced in the bitrate is not yet fully known. This paper investigates the effect of increase bitrate toward the increase of audio quality. The result showed a significant increase of the audio quality when the MPEG surround used in high bitrate. Accordingly, the MPEG Surround eligible for use in high bitrate to get better audio quality.

Keywords : *Spatial Audio Coding* dan *MPEG Surround*

Copyright © 2016 JNTE. All rights reserved

1. PENDAHULUAN

Teknologi *audio* dan *video* merupakan salah satu topik penelitian yang berkembang dalam beberapa tahun belakangan ini. Tuntutan akan kualitas *audio* dan *video* yang tinggi melahirkan beberapa standarisasi baru, salah satunya *Ultra High Definition Television* (UHDTV) yang dikenal juga sebagai *Super Hi-Vision* [1]–[3]. Teknologi UHDTV telah yang distandarisasi oleh *International Telecommunication Union* (ITU) ini, menggunakan video dengan resolusi tinggi untuk menghasilkan gambar yang lebih jernih, sementara itu untuk meningkatkan kualitas audio digunakan konfigurasi audio *multichannel* 22.2 sehingga *user* dapat mendengar *three-dimensional audio* (3D-Audio) [4],[5]. Kombinasi 3D-audio dan *high-resolution video* ini mampu memberikan *user* kesan audiovisual yang lebih nyata.

Dalam hal mentransmisikan 3D-audio, sangat penting untuk mempelajari *audio coding* guna menghemat *bandwidth* dari audio tersebut. *Motion Picture Expert Group* (MPEG) telah memperkenalkan MPEG-H 3D-Audio [6] yang

memiliki kemampuan untuk mentransmisikan audio 22 kanal seperti yang digunakan pada UHDTV.

Selain MPEG-H 3D-Audio, MPEG juga mengeluarkan MPEG Surround yang sangat populer dengan kemampuannya untuk mentransmisikan audio *multichannel* dengan hanya mengirimkan audio *downmix* berupa audio mono atau stereo [7]–[9]. Dengan menggunakan MPEG Surround, di sisi *encoder* audio *downmix* ditransmisikan bersamaan dengan informasi tambahan berupa *spatial parameter* dan *residual signal*. Dengan sistem yang demikian memungkinkan sinyal audio *multichannel* direkonstruksi kembali berdasarkan audio *downmix* dan informasi tambahan tersebut di sisi *decoder*.

Penggunaan MPEG Surround memberikan setidaknya dua kelebihan. Pertama, sistem ini memungkinkan pentransmisian audio *multichannel* dengan *bitrate* yang lebih rendah karena hanya mentransmisikan audio mono atau stereo saja. Kedua, secara konseptual MPEG Surround kompatibel dengan sistem audio yang lama (*backward compatibility*) sehingga, meskipun *user* tidak memiliki *codec* MPEG Surround, *user*

tetap dapat mendengarkan audio *downmix*-nya.

MPEG Surround umumnya dioperasikan dan diuji pada *bitrate* rendah, yaitu di-*bitrate* 64 dan 96 dan 160 kbps[10]. Sementara itu, MPEG Surround memungkinkan untuk digunakan pada *bitrate* yang lebih tinggi guna menghasilkan audio yang lebih jernih. Adapun kualitas audio yang dihasilkan MPEG Surround pada *bitrate* tinggi belum sepenuhnya diketahui. Berawal dari pemikiran tersebut, paper ini akan menginvestigasi kualitas audio MPEG Surround pada *bitrate* 256 hingga 400 kbps.

2. MPEG SURROUND

MPEG Surround merupakan *multichannel audio codec* yang bekerja dengan prinsip *Spatial Audio Coding*. MPEG Surround lahir karena tingginya permintaan untuk aplikasi *surround sound*. Adapun penyusunan standarisasi MPEG Surround dimulai pada tahun 2004 dan selesai pada musim gugur 2007 [9].

Gambar 1. Blok Diagram (a) Encoder dan (b) Decoder MPEG Surround

Blok diagram dari *encoder* dan *decoder* MPEG Surround ditunjukkan pada Gambar 1. Secara umum, MPEG Surround terdiri atas 4 bagian utama, yaitu *analysis filterbank*, *synthesis filterbank*, *encoding block*, dan *decoding block*. Perbedaan antara *encoder* dan *decoder* MPEG Surround hanya terdapat pada *encoding* dan *decoding block* saja, sedangkan *analysis* dan *synthesis filterbank* pada *encoder* maupun *decoder* tidak terdapat perbedaan.

Encoding dan *decoding block* merupakan bagian inti dari MPEG Surround yang mana pada bagian inilah proses kompresi audio

multichannel dilakukan. Sementara itu *Analysis filterbank* berfungsi untuk mendekomposisi sinyal audio pada domain *fullband* menjadi sinyal domain *subband* yang terdiri atas 71 *hybridband*. Sinyal domain *subband* inilah yang diolah oleh *encoding* dan *decoding block*. Selanjutnya hasil pengolahan ini diumpungkan ke *synthesis filterbank* guna mengembalikan audio ke domain *fullband*. Adapun pengolahan audio yang dilakukan pada domain *subband* ini meniru dari sistem pendengaran manusia yang juga bekerja pada domain *subband* guna meningkatkan kualitas audio.

Gambar 2. (a) R-OTT dan OTT Module
(b) R-TTT dan TTT Module

MPEG Surround menstandarisasi 2 pasang modul yang digunakan untuk men-*downmix* audio *multichannel* pada *encoding block* dan merekonstruksi kembali audio *multichannel* pada *decoding block* seperti yang ditunjukkan pada Gambar 2. Modul yang pertama adalah *One-to-Two* (OTT) yang berfungsi menyintesis kembali dua kanal audio berdasarkan satu kanal audio masukan, sementara itu pasangannya *Reverse One-to-Two* (R-OTT) digunakan untuk men-*downmix* dua kanal audio masukan menjadi audio satu kanal. Modul kedua bernama *Two-to-Three* (TTT) berfungsi untuk merekonstruksi kembali audio tiga kanal berdasarkan dua audio kanal audio masukan, dan pasangannya yang bernama *Reverse Two-to-Three* (R-TTT) didesain untuk mengkonversi tiga kanal audio masukan menjadi audio dua kanal. Untuk mengubah audio *multichannel* menjadi audio stereo *downmix* pada *encoder*, digunakan kombinasi beberapa R-OTT dan R-TTT module, sementara itu untuk mengubahnya menjadi mono *downmix* digunakan beberapa modul R-OTT saja seperti yang ditunjukkan pada Gambar 3. Untuk menyintesis kembali audio

multichannel pada *decoder* dilakukan proses yang merupakan kebalikan dari *tree structure* pada *encoder* menggunakan OTT dan TTT *module*. Pada paper ini, hanya MPEG Surround mono *downmix* menggunakan OTT dan R-OTT *module* yang akan dirinci secara mendetil, sementara itu untuk memahami lebih lanjut mengenai TTT dan R-OTT *module* dapat dilihat pada [9].

Gambar 3. *Tree Structure* untuk MPEG Surround konfigurasi 5-1-5

Selain men-*downmix* dua kanal audio masukan, R-OTT *module* juga mengekstrak *spatial parameter* untuk masing-masing *subband* yaitu *Channel Level Difference* (CLD) yang menyatakan perbedaan level antara dua kanal audio masukan dan *Inter-Channel Coherence* (ICC) yang menggambarkan ukuran kemiripan kedua sinyal masukan. *Spatial parameter* dalam domain *subband* ini selanjutnya ditransformasikan ke dalam *parameter band* yang berjumlah lebih kecil dibandingkan *subband* untuk menghemat pentransmisian. Adapun jumlah *parameter band* yang umum digunakan pada penelitian ini adalah 20 *parameter band*. Selain itu, agar memungkinkan rekonstruksi penuh bentuk gelombang audio pada *decoder* menggunakan OTT *module*, R-OTT juga mengekstrak *residual signal* yang merupakan selisih antara gelombang audio masukan dan gelombang audio yang disintesis menggunakan *spatial parameter*.

2.1. R-OTT Module

Jika $x_{1,s}[n]$ dan $x_{2,s}[n]$ merupakan merupakan audio masukan R-OTT *module*, maka *spatial parameter* ICC dan CLD untuk masing-masing *parameter band* b dapat dihitung dengan persamaan berikut,

$$ICC_b = 10 \log_{10} \left(\frac{\sigma_{x_{1,b}}^2}{\sigma_{x_{2,b}}^2} \right) \quad (1)$$

$$CLD_b = \text{Re} \left\{ \frac{\sum_n \sum_{s=s_b}^{s_{b+1}-1} x_{1,s}[n] x_{2,s}^*[n]}{\sigma_{x_{1,b}} \sigma_{x_{2,b}}} \right\} \quad (2)$$

dengan $\sigma_{x_{1,b}}^2$ dan $\sigma_{x_{2,b}}^2$ yang merupakan energi dari kedua sinyal *input* pada masing-masing *parameter band* b . Untuk sinyal pertama dan kedua pada *subband* s , $x_{1,s}[n]$ dan $x_{2,s}[n]$, kedua energi sinyal dapat dihitung dengan

$$\sigma_{x_{1,b}}^2 = \sum_n \sum_{s=s_b}^{s_{b+1}-1} x_{1,s}[n] x_{1,s}^*[n] \quad (3.a)$$

$$\sigma_{x_{2,b}}^2 = \sum_n \sum_{s=s_b}^{s_{b+1}-1} x_{2,s}[n] x_{2,s}^*[n] \quad (3.b)$$

dengan $x_{1,s}^*$ dan $x_{2,s}^*$ merupakan kompleks konjugate dari kedua audio masukan dan s_b merupakan *subband* pertama dari *parameter band* b .

Selanjutnya R-OTT *module* akan membuat sinyal *downmix* untuk masing-masing *subband* dengan persamaan

$$y_s[n] = \frac{x_{1,s}[n] + x_{2,s}[n]}{\varepsilon_{1,b} + \varepsilon_{2,b}} \quad (4)$$

yang mana konstanta energi ε_1 dan ε_2 digunakan untuk menjamin penjumlahan energi dari kedua sinyal *input* sama dengan energi sinyal *downmix* [9]. Kemudian, untuk mengompensasi distorsi yang disebabkan oleh proses *downmixing*, sebuah *residual signal*, $r_s[n]$ diekstrak berdasarkan sinyal *downmix* dan *spatial parameter* dengan persamaan sebagai berikut

$$r_{1,s}[n] = x_{1,s}[n] - \varepsilon_{1,b} y_s[n] \quad (5.a)$$

$$r_{2,s}[n] = x_{2,s}[n] - \varepsilon_{2,b} y_s[n] \quad (5.b)$$

$$r_s[n] = r_{1,s}[n] = r_{2,s}[n] \quad (5.c)$$

Berdasarkan pengalaman kami pada [11]–[16], *residual signal* pertama $r_{1,s}[n]$ dan *residual signal* kedua, $r_{2,s}[n]$ tidak identik sama, oleh sebab itu sinyal residu yang ditransmisikan adalah rata-rata dari kedua *residual signal* menggunakan persamaan

$$r_s[n] = \frac{r_{1,s}[n] + r_{2,s}[n]}{2} \quad (6)$$

Dengan mentransmisikan *residual signal* ke sisi *decoder* akan memungkinkan rekonstruksi penuh sinyal audio *multichannel* dengan menerapkan kebalikan dari persamaan (5).

2.2. OTT Module

Untuk menyintesis kembali dua kanal audio, OTT module terlebih dahulu mengestimasi konstanta energi ε_1 dan ε_2 menggunakan persamaan berikut ini (indeks dari *subband* dan *parameter band* tidak ditampilkan pada persamaan guna menyederhanakan penulisan)

$$\varepsilon_1 = \lambda_1 \cos(\alpha + \beta) \quad (7.a)$$

$$\varepsilon_2 = \lambda_2 \cos(\alpha + \beta) \quad (7.b)$$

yang mana penjumlahan dari kuadrat variabel λ_1 dan λ_2 bernilai 1

$$\lambda_1^2 + \lambda_2^2 = 1 \quad (8)$$

dengan λ_2 yang dapat dihitung dengan

$$\lambda_2 = \sqrt{\frac{1}{1 + 10^{\frac{CLD}{10}}}} \quad (9)$$

Sementara itu, variabel α dikomputasikan sebagai

$$\alpha = \frac{1}{2} \cos^{-1}(ICC) \quad (10)$$

dan variabel β dihitung berdasarkan nilai α , λ_1 , dan λ_2 dengan

$$\beta = \tan \left[\left(\frac{\lambda_2 - \lambda_1}{\lambda_2 + \lambda_1} \right) \tan^{-1}(\alpha) \right] \quad (11)$$

Selanjutnya, OTT module akan menyintesis dua kanal audio menggunakan

$$x_1[n] = \varepsilon_1 y[n] + r[n] \quad (12.a)$$

$$x_2[n] = \varepsilon_2 y[n] + r[n] \quad (12.b)$$

berdasarkan sinyal *downmix*, $y[n]$, sinyal residu, $r[n]$, dan konstanta energi ε_1 dan ε_2 (yang dihitung dari *spatial parameter*) yang dikirim oleh *encoder*.

3. PENGUJIAN

Pada penelitian ini, digunakan 5 jenis audio 5 kanal seperti yang ditunjukkan pada tabel 1. Adapun audio tersebut akan dikodekan menggunakan MPEG Surround pada *bitrate* 256, 288, 320, 352, 384, dan 416 kbps, dengan demikian terdapat 30 audio yang akan diuji dengarkan. Adapun 30 audio tersebut akan diujikan ke 20 orang pendengar menggunakan standar pengujian ITU-R BS.111-6 [17].

Tabel 1. Jenis Audio yang Diujikan

Audio	Deskripsi
Tepuk Tangan	Audio orang bertepuk tangan
Tertawa	Audio beberapa orang tertawa
Band Akustik	Bunyi gendang diiringi suara pria
Berita	Pria dan wanita membaca berita
Musik Klasik	Musik klasik dengan vokal wanita

Rekomendasi ITU-R BS.111-6 merupakan standarisasi pengujian audio yang umumnya digunakan untuk menilai kualitas kompresi audio yang tingkat penurunan kualitasnya sangat kecil. Standarisasi ini menggunakan metode *the double-blind triple-stimulus with hidden reference* yang mana masing-masing audio yang diuji akan didengarkan dengan 3 stimulus. Stimulus pertama (atau stimulus A) berisi audio asli yang nantinya akan menjadi referensi pengujian. Sementara itu, stimulus kedua dan ketiga (stimulus B dan C) berisi audio kompresi dan audio referensi (*hidden reference*) yang nantinya akan diberi nilai pada skala 1 sampai 5 bergantung pada tingkat gangguan kualitas audio terhadap stimulus A. Adapun tingkatan tersebut yaitu nilai 5 saat tidak terdengar gangguan, nilai 4 saat terdengar gangguan namun tidak mengganggu, nilai 3 saat terdengar gangguan dan sedikit mengganggu, nilai 2 saat terdengar gangguan dan mengganggu, dan nilai 1 saat terdengar gangguan yang sangat mengganggu.

Nilai yang diberikan oleh penguji selanjutnya akan dinormalisasi dengan memperhatikan nilai rata-rata dan standar deviasi menggunakan persamaan berikut ini [17]

$$Z_i = \frac{(x_i - x_{si})}{s_{si}} s_s + x_s \quad (13)$$

dimana

Z_i : Hasil normalisasi

x_i : Nilai subyek i

x_{si} : Nilai rata-rata subyek i
 x_s : Nilai rata-rata semua subyek
 s_s : Standar deviasi untuk semua subyek
 s_{si} : Standar deviasi untuk subyek i

Nilai hasil normalisasi tersebut selanjutnya dikonversi menjadi nilai SDG dengan persamaan

$$SDG = Z_{original} - Z_{compressed} \quad (14)$$

yang mana $Z_{original}$ merupakan nilai audio asli yang telah dinormalisasi sedangkan $Z_{compressed}$ merupakan nilai audio yang diuji. Dengan demikian nilai SDG akan berkisar pada rentang -4 hingga 0. Adapun deskripsi tingkatan gangguan atau penurunan kualitas audio yang ditunjukkan oleh nilai SDG dapat dilihat pada Tabel 2.

Tabel 2. Deskripsi Kualitas Audio berdasarkan Nilai SDG

Nilai SDG	Deskripsi
0	Tidak terdengar gangguan
-1	Terdengar gangguan namun tidak mengganggu
-2	Terdengar gangguan dan sedikit mengganggu
-3	terdengar gangguan dan mengganggu
-4	terdengar gangguan yang sangat mengganggu

4. HASIL DAN PEMBAHASAN

Hasil yang ditunjukkan pada Gambar 4 memperlihatkan grafik nilai SDG untuk pengode MPEG Surround. Pada grafik tersebut ditampilkan grafik SDG rata-rata yang merupakan nilai SDG rata-rata dari lima audio yang diuji beserta dua buah sampel nilai SDG untuk audio yang berbeda. Adapun grafik SDG dari kedua audio ini diambil untuk melihat pengaruh jenis audio terhadap kualitas audio yang dikodekan.

Pada Gambar 4 dapat dilihat, berdasarkan nilai SDG rata-rata kualitas audio yang dihasilkan berada pada level -1 hingga 0 (*impeceptible*), yang mana pada level tersebut gangguan atau penurunan kualitas audio tidak dapat dirasakan. Selain itu grafik tersebut juga menunjukkan pada *bitrate* tinggi, kenaikan *bitrate* tidak berpengaruh signifikan terhadap peningkatan kualitas audio. Secara keseluruhan nilai SDG untuk masing-masing sampel audio juga mengalami hal serupa dan juga dengan peningkatan nilai SDG yang tidak signifikan. Namun hal tersebut tidak terjadi pada audio tepuk tangan yang mana untuk mendapatkan audio dengan kualitas *imperceptible* setidaknya dibutuhkan *bitrate* sebesar 288 kbps. Pada *bitrate* tersebut kenaikan nilai SDG juga cukup signifikan yaitu meningkat sebesar 0,86 dibandingkan dengan *bitrate* 256 kbps.

Gambar 4. Grafik Nilai SDG MPEG Surround

5. KESIMPULAN

Paper ini memperlihatkan pengaruh kenaikan *bitrate* terhadap kenaikan kualitas audio yang dihasilkan pengkode MPEG Surround pada *bitrate* tinggi. Hasil menunjukkan pada sebagian audio kenaikan *bitrate* berpengaruh cukup signifikan saat *codec* digunakan pada *bitrate* dibawah 320 kbps. Dengan demikian MPEG Surround dapat digunakan pada *bitrate* tinggi untuk menjamin pendengar mendapat kualitas yang lebih baik.

DAFTAR PUSTAKA

- [1] Y. Shishikui et al., "High Performance video-codec for Super Hi-Vision," in *Proceedings of IEEE*, 2013, pp. 130–139.
- [2] E. Nakasu, "Super Hi-Vision on the horizon: A future TV system that conveys an enhanced sense of reality and presence," *IEEE Consumer Electronics Magazine*, pp. 36–42, 2012.
- [3] T. Ito, "Future television - super hi-vision and beyond," in *Proc. IEEE Asian Solid State Circuits Conference*, 2010.
- [4] T. Sugimoto, Y. Nakayama, and S. Oode, "Bitrate of 22.2 multichannel sound signal meeting broadcast quality," in *Proc. 137th AES Convention*, 2014.
- [5] K. Hamasaki, K. Hiyama, and R. Okumura, "The 22.2 Multichannel Sound System and Its Application," in *Proc. 118th AES Convention*, 2005.
- [6] J. Herre et al., "MPEG-H Audio—The New Standard for Universal Spatial/3D Audio Coding," *J. Audio Eng. Soc.*, vol. 62, no. 12, pp. 821–830, 2014.
- [7] J. Herre et al., "MPEG Surround - The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding," *J. Audio Eng. Soc.*, vol. 56, no. 11, pp. 932–955, 2008.
- [8] J. Hilpert and S. Disch, "The MPEG Surround Coding Standard [Standard in a nutshell]," *IEEE Signal Processing Magazine*, pp. 148–152, 2009.
- [9] J. Breebaart et al., "Background, concept, and architecture for the recent MPEG surround standard on multichannel audio compression," *J. Audio Eng. Soc.*, vol. 55, no. 5, pp. 331–351, 2007.
- [10] "Report on MPEG Surround Verification Test," 2007.
- [11] I. Elfitri, A. Luthfi, and Fitrilina, "R-TTT Module with Modified Residual Signal for Improving Multichannel Audio Signal Accuracy," in *Proceedings of 2015 International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT)*, Bandung, Indonesia, October 2015.
- [12] I. Elfitri, B. Gunel, and A. Kondo, "Multichannel Audio Coding Based on Analysis by Synthesis," in *Proceedings of the IEEE*, 2011, pp. 657–670.
- [13] I. Elfitri, R. Kurnia, and 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, October 2014.
- [14] I. Elfitri, M. Muharam, and M. Sobirin, "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, October 2014.
- [15] I. Elfitri, X. Shi, and A. Kondo, "Analysis by synthesis spatial audio coding," *IET Signal Process.*, vol. 3, no. 1, pp. 1–8, 2013.
- [16] I. Elfitri, H. D. Laksono, and A. Permana, "Balanced-Delay Filterbank for Closed-Loop Spatial Audio Coding," in *Proc. of 2015 Int. Seminar on Intelligent Technology and Its Applications (ISITIA)*, Surabaya, Indonesia, May 2015.
- [17] ITU-R, "Method for Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems," Recommendation ITU-R BS.1116-1, 1997.

Biodata Penulis

Amirul Luthfi, lahir pada tanggal 5 Desember 1993 di Kerinci. Penulis merupakan alumni Jurusan Teknik Elektro, Fakultas Teknik, Universitas Andalas dan kini sedang melanjutkan studi strata 2 di Program Studi Pasca Sarjana Teknik Elektro Universitas Andalas.

Fauzan Mustaqim, lahir 05 Mei 1990 di Pariaman. Kuliah Teknik Elektro, Fakultas Teknik Universitas Andalas, Padang, Sumatera Barat. Menyelesaikan studi strata 1 pada tanggal 5 mei 2014. Merupakan bungsu dari 9 saudara.