

## SPATIAL AUDIO CODING

Ikhwana Elfitri

Jurusan Teknik Elektro Unand, E-mail : [ikhwana@ft.unand.ac.id](mailto:ikhwana@ft.unand.ac.id)

### ABSTRAK

Pada paper ini diulas pengertian dasar dan prinsip operasi dari spatial audio coding sebagai salah satu bidang penelitian yang baru berkembang. Hasil-hasil penelitian terbaru beserta metode yang dikembangkan dibahas seperti MPEG Surround, Spatial Audio Scene Coding (SASC), Directional Audio Coding (DirAC). Kelebihan dan kekurangan setiap metode dianalisis dan dijadikan sebagai sebuah bahan studi perbandingan untuk mengembangkan metode terbaru.

### 1. PENDAHULUAN

*Speech coding* dan *audio coding*, telah menjadi salah satu topik penelitian yang berkembang pesat dalam beberapa tahun terakhir ini. Aplikasi dan implementasi dari hasil-hasil penelitian di bidang ini, juga telah banyak dirasakan dan bermanfaat secara luas di masyarakat umum sekalipun. Salah satu aplikasi yang digunakan secara luas adalah *Voice over IP* (VoIP), baik yang bersifat *private* ataupun yang *free* untuk publik seperti *Yahoo Messenger* (YM) dan *Skype*.

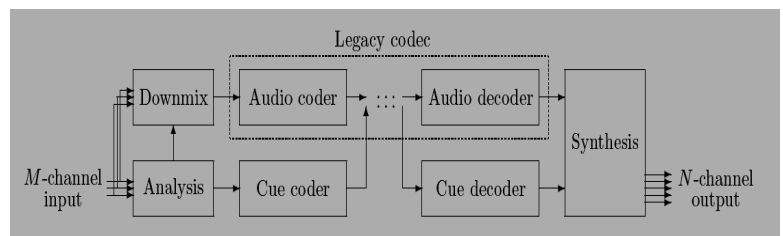
Untuk komunikasi suara, YM dan Skype telah memberikan kualitas komunikasi yang memuaskan, bahkan mungkin telah melebihi kualitas komunikasi suara melalui PSTN. Tetapi, animo dan keinginan untuk mendapatkan kualitas yang lebih baik lagi terus meningkat terutama untuk salah satu fitur yang disediakan oleh YM dan Skype, dan saat ini sangat banyak dimanfaatkan masyarakat, yaitu *teleconference*. Hal inilah yang kemudian menjadi salah satu motivasi untuk terus mengembangkan riset dan penelitian di bidang pengkodean suara ini. Suatu dialektika telah muncul disini, yaitu hasil riset dan penelitian telah membawa masyarakat untuk menjalani hidup dengan lebih mudah dan menyenangkan. Sebaliknya, keinginan dan motivasi untuk mendapatkan kemudahan dan kesenangan juga telah memicu para akademisi dan peneliti untuk melakukan penelitian lebih dalam lagi.

Salah satu area penelitian yang muncul dan berkembang pesat untuk menjawab tingginya kebutuhan konsumen ini adalah *spatial audio coding*. Karena itulah, dalam paper ini akan dibahas dan dikaji hal-hal mendasar yang terkait dengan topik penelitian ini serta disampaikan perkembangan penelitian terbaru yang telah diraih dan dipublikasikan secara luas.

### 2. PENGERTIAN SPATIAL AUDIO CODING

*Spatial audio coding* biasanya juga dikenal dengan *multichannel audio coding* atau *3D (3-dimension) audio coding*. Walaupun jika dilihat dari sisi bahasa ada sedikit perbedaan dari ketiga terminologi ini, secara prinsip dapat didefinisikan bahwa *spatial audio coding* adalah sistem untuk merepresentasikan beberapa sinyal audio (*multiple audio signals*) secara efisien untuk tujuan transmisi dan penyimpanan data.

Model yang paling umum dari spatial audio coding dapat dijelaskan dengan merujuk kepada gambar 1. M sinyal audio di-downmix menjadi satu sinyal dan kemudian diproses dengan *existing audio coder*. Ini berarti bahwa semua tipe audio coder dapat digunakan untuk pengkodean sinyal downmix ini. Model seperti ini juga memungkinkan sistem mempunyai kemampuan *backward compatibility*, artinya compatible dengan audio coder yg digunakan untuk sinyal downmix. Blok analysis berfungsi untuk mengekstrak spatial parameter untuk selanjutnya ditransmisikan sebagai side information. Sedangkan blok synthesis pada decoder berfungsi merekonstruksi kembali sinyal audio dari sinyal downmix menggunakan spatial parameter yang diterima.

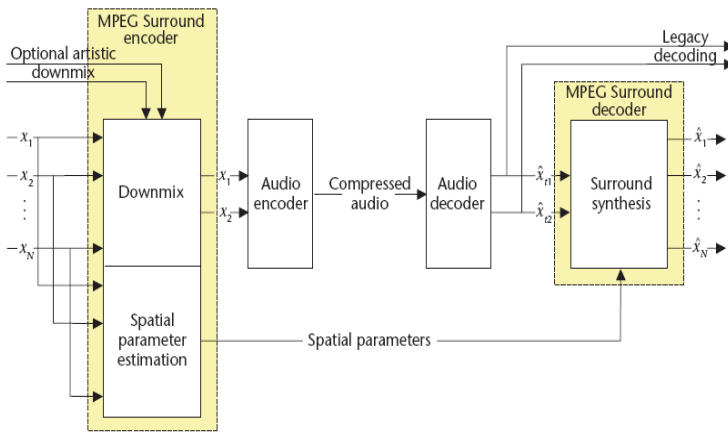


Gambar 1. Diagram blok spatial audio coding[1].

3. MPEG SURROUND (MPS)

Dengan tingginya permintaan pasar (*market demand*) untuk aplikasi surround sound, MPEG telah mengeluarkan CFP (*Call for Proposal*) pada tahun 2004 untuk penyusunan standard bagi spatial audio coding ini. Tahun 2007, penyusunan standard ini selesai dan populer dengan nama MPEG Surround (MPS).

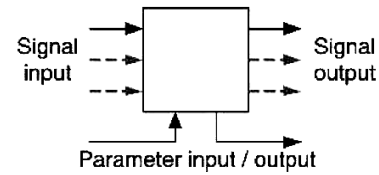
Blok diagram MPEG Surround dapat dilihat pada gambar 2. Seperti model yang umum pada spatial audio coding, semua sinyal audio yang masuk ke encoder dijadikan satu sinyal downmix, kemudian dikodekan menggunakan audio coder yang sudah tersedia. Ini berarti MPEG Surround punya kemampuan backward compatibility.



Gambar 2. Blok diagram MPEG Surround [2]

Selain sinyal downmix, semua sinyal audio yang masuk ke encoder juga direpresentasikan dengan spatial parameter. Untuk sistem dasar, MPEG Surround menggunakan 4 spatial parameter yaitu : channel level differences (CLDs), inter-channel coherences (ICCs), channel prediction coefficients (CPC) dan residual errors. Ke-empat parameter ini dikembangkan dari prinsip sistem pendengaran manusia (*human spatial hearing*). Pada dasarnya, parameter ini tidak mendeskripsikan *spatial position* dari sinyal audio, melainkan mengambil *interchannel relationship* dari semua sinyal audio. Karena itu, jumlah sinyal audio yang direproduksi pada decoder haruslah sama dengan jumlah sinyal audio yang diterima encoder.

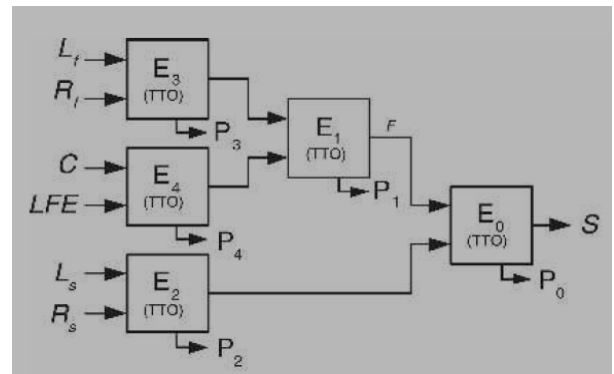
MPEG Surround bekerja dengan menggunakan dua blok dasar yaitu one-to-two (OTT) encoder-decoder dan two-to-three (TTT) encoder-decoder.



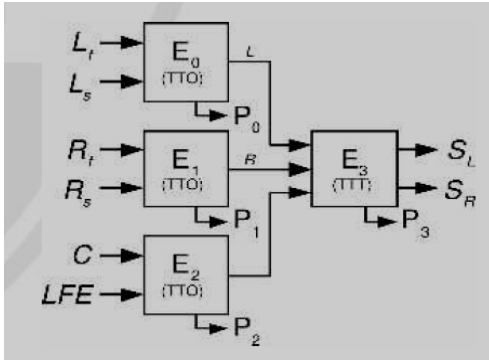
Setiap konfigurasi input-output dibentuk dari kombinasi kedua blok dasar ini. Gambar 3 memperlihatkan bentuk *generic* dari kedua blok dasar MPEG Surround ini.

Gambar 3. Blok dasar MPS[2]

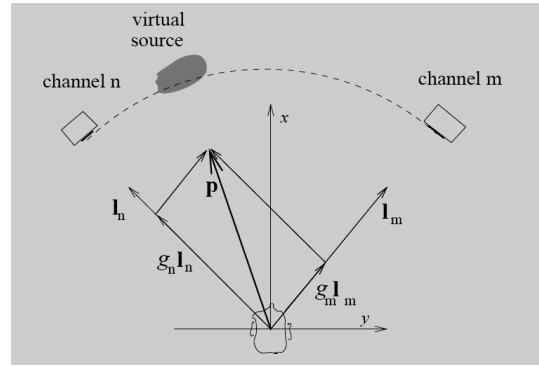
OTT encoder berfungsi membentuk satu sinyal downmix dari dua sinyal audio. Sebaliknya OTT decoder merekonstruksi kembali 2 sinyal audio dari satu sinyal downmix. TTT encoder berfungsi membentuk dua sinyal downmix (stereo) dari 3 sinyal audio. TTT decoder berfungsi untuk mengembalikan 3 sinyal audio dari 2 sinyal downmix. Sebagai contoh bentuk implementasi dari konfigurasi input-output menggunakan blok dasar ini dapat dilihat konfigurasi sistem dari 5 sinyal audio menjadi satu sinyal (mono downmix) pada gambar 4 dan konfigurasi sistem dari 5 sinyal audio menjadi dua sinyal (stereo downmix). Dengan kombinasi blok dasar ini, MPEG Surround dapat membentuk sinyal mono/stereo downmix, berapapun jumlah sinyal audio yang dijadikan input sinyal pada encoder.



Gambar 4. Konfigurasi Encoder untuk membentuk sinyal mono downmix[2].



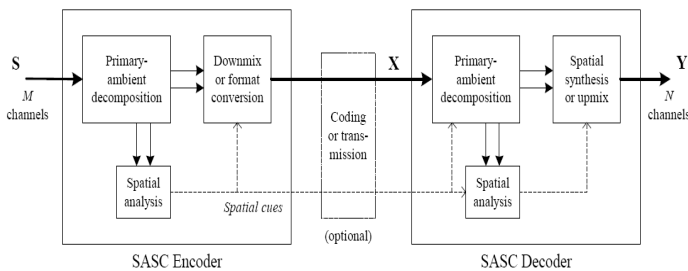
Gambar 5. Konfigurasi encoder untuk membentuk sinyal stereo downmix[2].



Gambar 7. Prinsip kerja VBAP[3]

4. SPATIAL AUDIO SCENE CODING (SASC)

SASC dikembangkan oleh beberapa peneliti dari Creative Advanced Technology Center, USA. Prinsip dasar pengkodean ini adalah menangkap audio scene dan merepresentasikannya dalam bentuk sinyal monodownmix dan direction vector. Pada decoder, audio direproduksi kembali dengan menggunakan vektor *base amplitude panning* (VBAP)[3]. Gambar 6 memperlihatkan blok diagram dari SASC. M channel sinyal masukan dipisahkan menjadi sinyal *primary* dan *ambient*. Selanjutnya kedua sinyal tersebut diproses dengan algoritma yang sama.



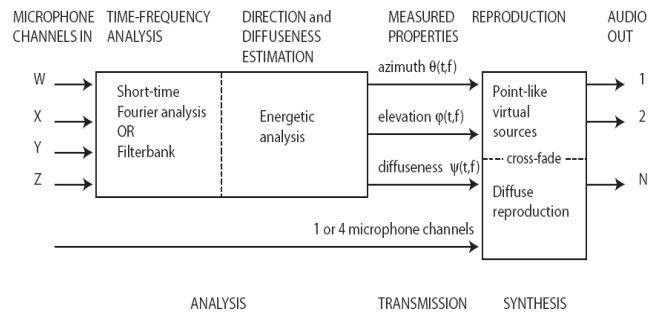
Gambar 6. Blok diagram SASC[4]

SASC dikembangkan berdasarkan sistem reproduksi audio berbasis vektor (VBAP). Pada gambar 7 dapat dilihat bahwa VBAP akan menghasilkan sinyal audio yang seolah-olah berasal dari suatu sumber virtual. Untuk menghasilkan sinyal audio yang berasal dari suatu sumber virtual, maka sinyal audio yang dijadikan input speaker ditentukan amplitudonya oleh suatu *direction vektor*. Oleh karena itu, pada SASC encoder sinyal audio dari semua channel dijadikan satu sinyal downmix dan posisi sumbernya direpresentasikan dengan sebuah vektor. Untuk dapat merepresentasikan sejumlah sinyal audio menjadi sebuah vektor, SASC menggunakan prinsip vektor gerzon.

Dengan prinsip operasi seperti ini, SASC dapat digunakan dengan format input dan output yang berbeda. SASC dapat menerima sejumlah M channel sinyal audio, merepresentasikannya dalam bentuk direction vektor dan kemudian mentransmisikan ke decoder. Decoder selanjutnya dapat merekonstruksi kembali sinyal audio ini untuk menghasilkan N channel audio.

5. DIRECTIONAL AUDIO CODING (DirAC)

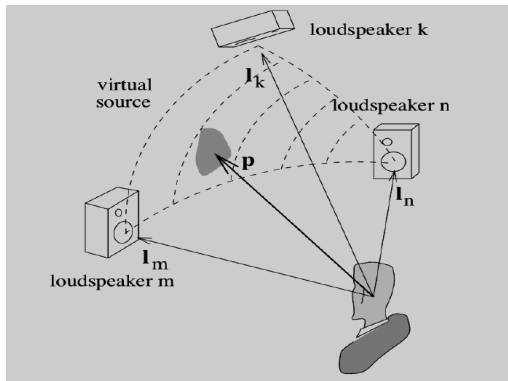
DirAC merepresentasikan sinyal audio dalam bentuk direction vector, seperti pada SASC. Yang membedakannya adalah, DirAC menerima sinyal audio dalam bentuk sinyal dari suatu *microphone array*. Sinyal ini dianalisis untuk menghasilkan sudut elevasi dan azimuth dari direction vector serta *diffuseness* untuk merepresentasikan non-directional sound. Direction vector ini ditransmisikan ke decoder bersama-sama dengan 1 atau 4 sinyal microphone. Diagram blok DirAC ditampilkan pada gambar 8.



Gambar 8. Blok diagram DirAC[5]

Sedikit berbeda dengan SASC, DirAC merekonstruksi sinyal audio dengan *triple wise amplitude panning*[6]. Artinya, sebuah sumber virtual dihasilkan dari 3 speaker aktif, sedangkan SASC

menghasilkan sumber virtual dari 2 speaker aktif (*pair-wise amplitude panning*).



Gambar 9. Triple wise VBAP[7]

## 6. ANALISA PERBANDINGAN

Satu-satunya spatial audio coder yang telah distandardkan dan diimplementasikan adalah MPS. Karena itu, banyak sekali fitur pada MPS yang menjadi kelebihannya. Diantaranya adalah kemampuan untuk menerima sinyal *artistic downmix* yang dihasilkan oleh pre-encoder. Disamping itu MPS juga dapat beroperasi dengan jangkauan resolusi frekuensi dan resolusi waktu yang luas yang menghasilkan kemampuan untuk mengirimkan spatial audio dalam range bit rate yang sangat beragam. MPS juga dapat digunakan untuk menghasilkan sinyal surround audio dari headphone dengan menggunakan teknologi *binaural rendering systems*[8].

Disamping memiliki banyak kelebihan, ada satu kelemahan pada MPS yaitu pada dasarnya coder ini tidaklah mengambil spatial posisi dari sinyal audio yang akan dikodekan, tetapi hanya mengekstrak beberapa parameter yang menggambarkan relasi antar channel. Dengan demikian MPS tidak cocok digunakan untuk mengirimkan directional sound. MPS lebih tepat digunakan untuk mengirimkan *non-directional sound*.

SASC dikembangkan untuk mendapatkan solusi atas kelemahan MPS. Dengan prinsip *interchannel relationship*, jumlah channel sinyal audio pada input MPS haruslah sama dengan jumlah channel audio pada decodernya. Pada SASC, hal ini tidak diharuskan, karena itu SASC juga dapat digunakan untuk konversi sinyal audio dari M channel menjadi N channel audio. Oleh karena itu, SASC dikenal dengan istilah *format independent spatial audio coder*.

Kelebihan lain dari SASC adalah konsep pengkodeannya dikembangkan dari metode analysis dan synthesis yang memiliki konsistensi. Artinya, secara matematis, sinyal audio yang dihasilkan dari decoder SASC dapat dianalisis kembali untuk

menghasilkan sinyal audio dengan direction vector yang sama.

Kelemahan SASC justru ada pada prinsip gerzon vector yang digunakan untuk merepresentasikan multiple sinyal audio menjadi satu sinyal audio. Dengan prinsip gerzon vector ini, direction vector yang dihasilkan akan memunculkan potensi error yang besar, karena vektor ini dihasilkan dari resultan sejumlah vektor. Faktor lainnya yang juga menjadi kekurangan SASC adalah ketidakmampuannya untuk merepresentasikan *non-directional sound*. Semua sinyal audio pada SASC diasumsikan sebagai directional sound. Sedangkan pada decodernya, non directional sound dihasilkan hanyalah untuk mendapatkan efek radial dari sinyal audio.

DirAC mempunyai solusi atas kelemahan SASC ini. Dengan mengambil *diffuseness* dari sinyal audio, DirAC dapat menghasilkan non directional sound dengan kualitas yang baik. Hanya saja, hal ini baru dapat dihasilkan jika semua sinyal microphone (4 sinyal) dikirimkan ke decoder. Hal ini tentu saja akan membutuhkan bit rate yang tinggi. Kelemahan DirAC lainnya adalah keharusannya untuk menerima sinyal audio dalam bentuk sinyal microphone.

## 7. KESIMPULAN

Pada makalah ini telah ditampilkan tiga spatial audio coder yang baru dikembangkan yaitu MPEG Surround, Spatial Audio Scene Coding (SASC) dan Directional Audio Coding (DirAC). Prinsip kerja setiap coder dibahas dan dianalisa perbandingan antara ketiganya.

## DAFTAR PUSTAKA

- [1] Goodwin, M. M. Dan Jot, J-M., *A frequency Domain Framework for Spatial Audio Coding Based on Universal Spatial Cues*, the 120<sup>th</sup> AES Convention, France, 2006
- [2] Breebart, J., Hotho, G., Schuijers, E., Oomen, W., De Par, S. V., *Background Concept and Architecture for the Recent MPEG Surround Standard on Multichannel Audio Compression*, J. Audio Eng. Soc., Vol 55 no 5, 2007
- [3] Pulkki, V., *Compensating Displacement of Amplitude-Panned Virtual Sources*, AES 22<sup>nd</sup> Int. Conf. On Virtual, Synthetic and Entertainment Audio, Finland, 2002
- [4] Jot, J-M., Merimaa, J., Goodwin, M. M., Krishnaswamy, A., Laroche, J., *Spatial Audio Scene Coding in a Universal Two-Channel 3-D Stereo Format*, the 123<sup>rd</sup> AES Convention, New York, 2007

- [5] Pulkki, V., Faller, C., *Directional Audio Coding : Filterbank and STFT-based Design*, the 120<sup>th</sup> AES Convention, Paris, 2006
- [6] Pulkki, V., Karjalainen, M., *Multichannel Audio Rendering Using Amplitude Panning*, IEEE Signal Processing Magazine, May 2008
- [7] Vilkamo, J., *Spatial Sound Reproduction with Frequency Band Processing of B-Format Audio Signals*, Master Thesis, Helsinki University of Technology, May 2008
- [8] Herre, J. Kjørling, K., Breebart, J., Faller, C., Disch, S., Purnhagen, H., Koppens, K., Hilpert, J., Roden, J., Oomen, W., Linzmeier, K., Chong, K. S., *MPEG Surround-The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding*, J. Audio Eng. Soc., Vol 56 no 11, November 2008.