

A DIRECT MPEG SURROUND ENCODING SCHEME FOR SURROUND SOUND RECORDING WITH COINCIDENT MICROPHONE TECHNIQUES

SAMUEL SAMSUDIN NG, EVELYN KURNIAWATI, AND SAPNA GEORGE

STMicroelectronics Asia Pacific, Pte. Ltd., Singapore

samsudin.ng@st.com

MPEG Surround provides an efficient representation of multi-channel audio signals by a set of down-mixed signals and spatial parameters. These parameters describe the inter-channel spatial properties, which are utilized by the decoder to reconstruct the multi-channel signals from the down-mixed signals. The encoding process includes time/frequency analysis-synthesis filtering and signal down-mixing; hence the computational requirement is highly dependent on the number of surround audio playback channels. Meanwhile, coincident microphone techniques offer compact microphone array geometry and low number of microphones to produce surround sound recording. Usually, the recorded signals are post-processed into multi-channel surround sound signals prior to encoding. In this paper, an MPEG Surround encoding scheme for coincident microphone recording is presented. The main feature is the direct derivation of the spatial parameters and down-mixed signals from the microphone array signals. It minimizes the dependency of the computational demand on the number of surround sound audio channels. Implemented on ARM-based embedded platform, the direct encoding scheme requires approximately 50% less memory and 40% less processing power than the reference encoding scheme.

INTRODUCTION

Surround sound playback system has gained an increasing popularity in consumers' home theatre setup. Handheld and mobile devices are catching up with this trend, offering capabilities including surround sound recording and playback. Some of the applications are portable sound recorder and mobile phone with surround sound recording capability. Such recording scenarios generally require compact microphone array within a limited mounting area. For storage and transmission of the recordings, there are limitations in terms of the data bandwidth availability, storage capacity and processing power.

Surround sound can be recorded efficiently using coincident microphone technique with compact multi-microphone array, and made highly feasible for handheld and mobile devices due to the availability of microphone capsules with very small dimension or even microelectromechanical systems (MEMS) microphone. Subsequently, the huge amount of surround sound data can be compressed efficiently with MPEG Surround spatial audio encoder [1][2].

This paper is structured as follows: Section 1 provides a summary of surround sound production using coincident microphone techniques. Section 2 reviews the MPEG Surround encoding scheme and Section 3 introduces the direct encoding scheme. Section 4 provides some comparison result of the direct scheme and the standard scheme. Lastly, Section 5 presents the overall conclusion.

1 COINCIDENT MICROPHONE TECHNIQUES

Coincident microphone techniques utilize the psychoacoustic principles of inter-aural level differences (ILD) to record and recreate the audio scene during surround sound playback. Such techniques require only a minimum of three first-order directional microphones, arranged in such a way that the polar patterns of these microphones coincide on the horizontal plane (most commonly by stacking them on top of each other). Some of the popular microphone setups for coincident surround recording are:

1. Double Mid/Side (DMS) array [3] which consists of a front-facing cardioid (mid-front), a side-facing bidirectional (side) and a rear-facing cardioid (mid-rear) microphones,
2. FLRB array [4] which consists of front (F), left (L), right (R), and rear (B) facing cardioid microphones, and
3. B-format microphone array [5] which consists of three or four microphones and additional signal processing to produce coincident B-format signals with omnidirectional (W), front-facing bidirectional (X) and side-facing bidirectional (Y) responses required for horizontal surround sound production.

Figure 1(a) shows the polar pattern of B-format microphone array signals. Weighted combination of these directional signals produces a *virtual microphone*

signal. By properly designing the weighting factors, unlimited number of virtual microphone signals with any first-order directivity pointing to any directions around the horizontal plane can be derived.

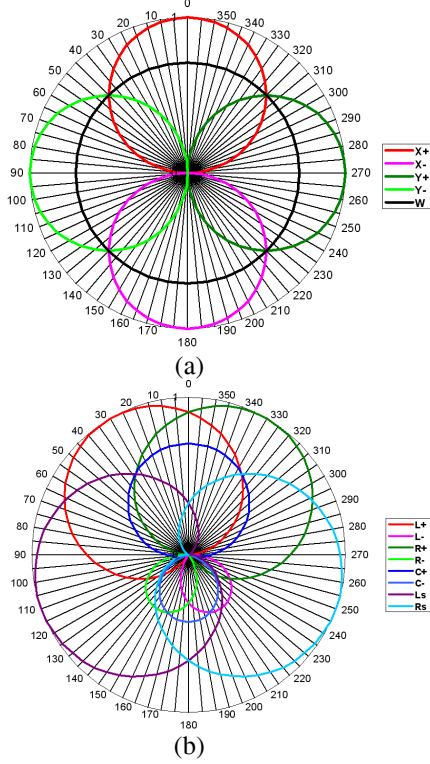


Figure 1: (a) Polar patterns of B-format microphone array signals, (b) ITU 5.0 surround sound polar pattern resulting from virtual-microphone matrixing of B-format signals

The surround sound signals can then be obtained by deriving one virtual microphone signal for each surround sound channel. In this context, the weighting factors to derive each surround audio channel's signal (this set of weighting factors will be subsequently referred to as the *channel coefficients*) are designed such that the resulting virtual microphone is pointing to the direction which corresponds to the location of the speaker in the surround playback configuration. For example, a surround channel C_i is derived from B-format signals and its channel coefficients (α_i , β_i , γ_i) according to

$$\mathbf{C}_i = \alpha_i \mathbf{W} + \beta_i \mathbf{X} + \gamma_i \mathbf{Y}. \quad (1)$$

Figure 1(b) shows the typical virtual-microphone polar pattern for a standard ITU 5.0 surround sound signals. In this example, the channel coefficients have been designed such that the virtual microphones for the center (C), left-front (L) and right-front (R) surround

channels possess supercardioid directivity and point to 0° and $\pm 30^\circ$, respectively, while the virtual microphones for the left-surround (Ls) and right-surround (Rs) surround channel possess cardioid directivity and point to $\pm 110^\circ$, respectively.

2 MPEG SURROUND ENCODING SCHEME

Multi-channel surround sound applications require huge amount of memory space and transmission bandwidth which scales up linearly with the number of surround sound channels. State-of-the-art perceptual or hybrid audio compression schemes such as MPEG-1 layer 3 (mp3) [6] and Advanced Audio Coder (AAC) [7] are able to compress monaural or stereo audio signals very efficiently. However for multi-channel signals, the required data rate scales up with the number of surround sound channels.

MPEG Surround (MPS) has been standardized as multi-channel audio compression scheme. It represents surround sound by a set of down-mixed signals (with a lower number of channels than the surround sound, eg. monaural or stereo downmix) and low-overhead spatial parameters that describe its spatial properties. The decoder is able to reconstruct the surround sound channels from the down-mixed signals and spatial parameters.

When combined with perceptual audio coders to compress downmix signals, MPS enables an efficient representation of surround sound which is compatible with the existing mono or stereo infrastructure. As the details of this encoding structure is beyond the scope of this document, only a brief overview is provided in the following paragraphs and interested readers may refer to [1], [2] and [8] for more details.

2.1 Encoder structure

A reference MPEG Surround audio encoder structure is shown in Figure 2. The time/frequency (T/F) analysis consists of an exponential-modulated quadrature mirror filterbank (QMF) filtering followed by a low-frequency filtering to increase the frequency resolution for the lower subbands. Together, this filtering scheme is referred to as *hybrid analysis filtering*. The filtering is performed on each surround sound channel to convert the time-domain audio signals into the subband-domain signal representations.

The multi-channel subband signals are passed to a spatial encoding stage which calculates the spatial parameters and performs signal downmixing. The down-mixed signals are synthesized back into the time domain and can be further compressed using any audio compression schemes. The spatial parameters are quantized and formatted according to the spatial audio syntax as defined in [1], and typically appended to the downmix-audio bitstream. Optionally, a set of residual signals are derived and coded according to AAC low-

complexity syntax. These coded signals can be transmitted in the spatial parameter bitstream to enable full waveform reconstruction at the decoder side.

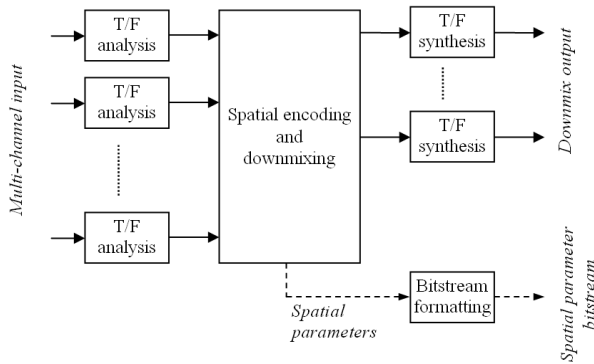


Figure 2: MPEG Surround encoder structure

2.2 Spatial encoding tree

The spatial encoding stage is realized as a tree structure, which consists of a series of two-to-one (TTO) and three-to-two (TTT) encoding blocks. The encoding tree structures for mono- and stereo-based encoding of 5.1 surround sound is illustrated in Figure 3(a) and (b).

TTO takes a subband-domain signal pair as input, calculates the signal energy and cross-correlation, and groups these values into several parameter bands with non-uniform frequency bandwidth. At each parameter band, the spatial parameters and downmix scalefactor are calculated. The subband-domain signal pair is mixed to derive the monaural and residual signals. The monaural (summed) signal is subsequently scaled by the downmix scalefactor, which is required to ensure overall energy preservation in the downmix signal. The residual (subtracted) signal is either discarded or coded for transmission in the spatial parameter bitstream. TTT performs similar operations but with three input signals and stereo down-mixed signals.

2.3 Output-downmix modification

In the stereo-based encoding mode, MPEG Surround coding scheme also provides the possibility to transmit matrix-compatible or 3D-stereo downmix instead of the standard stereo downmix. The transmission of matrix-compatible stereo downmix provides backward compatibility with legacy matrixed surround decoders, while 3D stereo downmix provides the advantage of binaural listening for existing stereo playback system. In the generic encoding scheme, such modified down-mixed signals are created by applying a 2x2 post-processing matrix which modifies the energy and phase of the standard stereo down-mixed signals. Upon receiving these downmixes, a standard MPEG Surround decoder is able to revert back to the standard stereo downmixes by applying the inverse of the post-

processing matrix, such that proper MPEG Surround decoding can be performed.

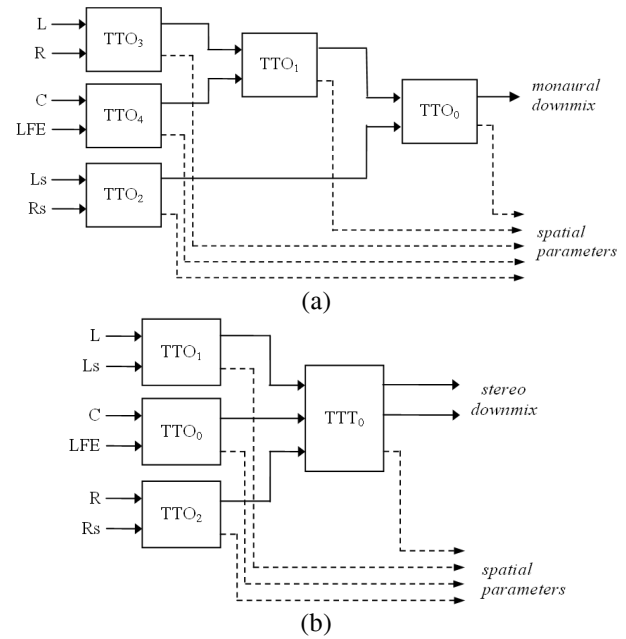


Figure 3: MPEG Surround tree structure for (a) mono-based encoding, and (b) stereo-based encoding

3 DIRECT ENCODING SCHEME

MPS hybrid analysis filtering is a computationally intensive scheme and it has to be performed on each of the surround audio channel. In the spatial encoding stage, the energy and cross-correlation calculation and subband signal downmixing contribute to substantial computational power as they have to be performed at each encoding block.

As the number of surround sound channel is increased, more TTO and/or TTT encoding blocks are required to encode the extra channels which increase the overall computational requirement of the encoder. Due to such structure, memory and computational requirement of MPS encoder is highly dependent on the number of surround audio channels.

The following sections described the direct encoding scheme. The development of the scheme was started in 2008 when MPEG Surround was just standardized. It has been granted US patent in 2012 [9].

3.1 Direct encoder structure

In coincident surround sound recording scheme, theoretically unlimited number of virtual microphone signals can be derived from a fixed number of microphone array signals. Hence, the same microphone array signals can be used to derive different surround audio signals for different playback configurations simply by changing the size and coefficients of the

channel-coefficients matrix. For example, a 5.0 and 7.0 surround sound signals can be derived from B-format signals by designing the corresponding 3-to-5 and 3-to-7 channel-coefficients matrixes. It can be seen that the required number of coincident microphone signals is independent of the number of surround channels.

Based on this observation, an MPEG Surround encoding scheme for coincident microphone recording scheme is presented. It derives the spatial parameters, residual signals, and down-mixed signals directly from the microphone array signals and the channel-coefficients matrix. The analysis filtering is performed on the microphone array signals; hence it is independent of the surround sound configuration.

Figure 4 shows the diagram of the direct encoding scheme. In the diagram and subsequent discussion, the commonly used three-channel coincident microphone technique is assumed. For simplicity of signal labelling, we refer to B-format signals (W, X and Y). However, the proposed scheme is applicable to other coincident recording techniques with any numbers of microphone signals which utilize virtual microphone matrixing to produce surround sound signals.

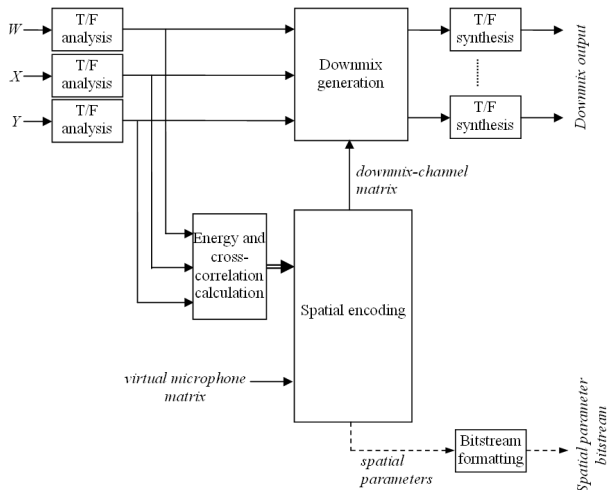


Figure 4: Direct MPEG Surround encoding scheme for coincident-microphone surround recording

3.2 Energy and cross-correlation calculation

At each hybrid analysis subband, signal energy of $\sigma_{W,b}^2$, $\sigma_{X,b}^2$, $\sigma_{Y,b}^2$ and cross-correlation $r_{WX,b}$, $r_{WY,b}$, $r_{XY,b}$ at each parameter band b are calculated according to

$$\sigma_{S_i,b}^2 = \sum_n \sum_{k=k_b}^{k_{b+1}-1} S_{i,k,n} \cdot S_{i,k,n}^* \quad (2)$$

$$r_{S_i S_j,b} = \text{Re} \left\{ \sum_n \sum_{k=k_b}^{k_{b+1}-1} S_{i,k,n} \cdot S_{j,k,n}^* \right\} \quad (3)$$

where S_i and S_j represents any of the microphone array signals, k_b refers to the subband index of the subband

boundary of parameter band b , $\text{Re}\{Z\}$ denotes the real part of complex signal Z , and $*$ denotes complex conjugation. This set of parameter-band signal energies and cross-correlations form a common input to all TTO and TTT encoding blocks in the spatial encoding stage.

If there are more than three microphone array signals, the cross-correlation has to be calculated for all possible pairs of the microphone signals. This may increase the complexity depending on the number of microphone signals. However, for the case of first order horizontal surround sound recording (without height), generally three microphones are adequate hence optimal for the encoding.

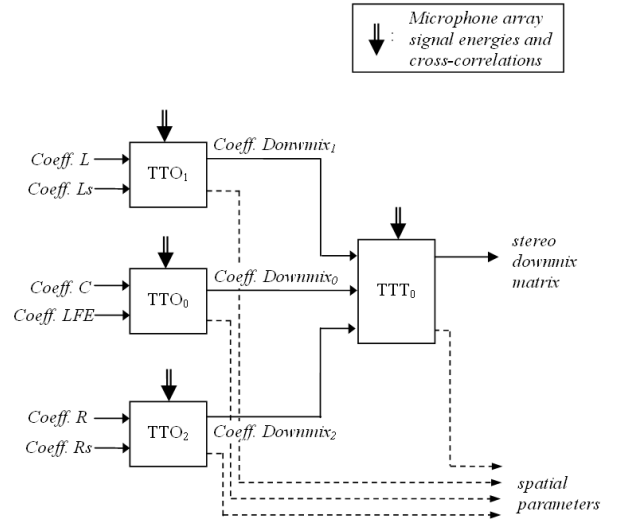


Figure 5: Encoding tree structure of the proposed MPEG Surround encoding scheme for stereo-based encoding scheme

3.3 Spatial encoding tree

The proposed spatial encoding tree structure for stereo-based encoding is shown in Figure 5. It can be seen that the encoding tree structure is similar to the tree structure as shown in Figure 3, but the actual input channel signals are represented by their respective input-channel coefficients in the virtual microphone coefficients matrix.

The advantage of this scheme is that the signal energy and cross-correlation calculation are only performed once on the microphone array signals, instead of multiple times at each encoding block. Effectively, the spatial parameters are calculated without deriving the actual surround sound channels.

Signal down-mixing at each TTO or TTT encoding block is represented by summation and scaling of the input channel-coefficients pair or triplet. Instead of the actual surround channel signals, only their respective

channel coefficients are navigated through the encoding tree.

3.3.1 TTO Encoding Block

Signal energy $\sigma^2_{C_1,b}$ and $\sigma^2_{C_2,b}$ of the actual TTO input channels C_1 and C_2 are calculated from their respective channel coefficients and the microphone array signal energy and cross-correlation. This is shown by expanding the virtual microphone operations according to

$$\begin{aligned}\sigma^2_{C_1,b} &= \sum_n \sum_{k=k_0}^{k_{b+1}-1} C_{i,k,n} \cdot C_{i,k,n}^* \\ &= \sum_n \sum_{k=k_0}^{k_{b+1}-1} (\alpha_{i,b} W_{k,n} + \beta_{i,b} X_{k,n} + \gamma_{i,b} Y_{k,n}) \\ &\quad \cdot (\alpha_{i,b} W_{k,n}^* + \beta_{i,b} X_{k,n}^* + \gamma_{i,b} Y_{k,n}^*) \\ &= \alpha_{i,b}^2 \sigma^2_{W,b} + \beta_{i,b}^2 \sigma^2_{X,b} + \gamma_{i,b}^2 \sigma^2_{Y,b} \\ &\quad + 2\{\alpha_{i,b} \beta_{i,b} r_{WX,b} + \alpha_{i,b} \gamma_{i,b} r_{WY,b} + \beta_{i,b} \gamma_{i,b} r_{XY,b}\}\end{aligned}\quad (4)$$

where i refers to the input channel index.

Using similar expansion technique, the cross-correlation between the pair of input channels $r_{C_1C_2,b}$ is calculated according to

$$\begin{aligned}r_{C_1C_2,b} &= \alpha_1 \alpha_2 \sigma^2_{W,b} + \beta_1 \beta_2 \sigma^2_{X,b} + \gamma_1 \gamma_2 \sigma^2_{Y,b} \\ &\quad + (\alpha_1 \beta_2 + \alpha_2 \beta_1) r_{WX,b} + (\alpha_1 \gamma_2 + \alpha_2 \gamma_1) r_{WY,b} \\ &\quad + (\beta_1 \gamma_2 + \beta_2 \gamma_1) r_{XY,b}\end{aligned}\quad (5)$$

This is possible because the virtual microphone signal matrixing is a linear operation hence the channel signal energy and cross-correlation can be calculated from the linear combination of the microphone array signal energy and cross-correlation.

From these values, the spatial parameters channel level difference (CLD), inter channel correlation (ICC) and downmix scalefactor g_b are then calculated according to [8]

$$CLD_b = 10 \log_{10} \frac{\sigma^2_{C_1,b}}{\sigma^2_{C_2,b}} \quad (6)$$

$$ICC_b = \frac{r_{C_1C_2,b}}{\sigma_{1,b} \sigma_{2,b}} \quad (7)$$

$$g_{C_1C_2,b} = \sqrt{\frac{\sigma^2_{C_1,b} + \sigma^2_{C_2,b}}{\sigma^2_{C_1,b} + \sigma^2_{C_2,b} + 2r_{C_1C_2,b}}} \quad (8)$$

Subsequently, the pair of input-channel coefficients are summed and scaled to derive the downmix channel-coefficients of the encoding block according to Equation 9 to derive the monaural downmix-channel coefficients. Similarly, the residual-signal coefficients can also be calculated by subtracting the input-channel

coefficients pair. The resulting residual-signal coefficients are used to derive the residual signal for the corresponding TTO block.

$$\begin{bmatrix} \alpha_{\text{Downmix}_{0,b}} \\ \beta_{\text{Downmix}_{0,b}} \\ \gamma_{\text{Downmix}_{0,b}} \end{bmatrix} = g_{C_1C_2,b} \begin{bmatrix} \alpha_{1,b} + \alpha_{2,b} \\ \beta_{1,b} + \beta_{2,b} \\ \gamma_{1,b} + \gamma_{2,b} \end{bmatrix} \quad (9)$$

3.3.2 TTT Encoding Block

Similar to TTO, signal energies $\sigma^2_{C_1,b}$, $\sigma^2_{C_2,b}$, $\sigma^2_{C_3,b}$ and cross-correlations $r_{C_1C_2,b}$, $r_{C_1C_3,b}$, $r_{C_2C_3,b}$ of the actual input channel triplet C_1 , C_2 , and C_3 are calculated. For TTT block operating in energy mode, the spatial parameter CLD₁ and CLD₂ are calculated according to

$$CLD_{1,b} = 10 \log_{10} \frac{\sigma^2_{C_1,b} + \sigma^2_{C_2,b}}{\frac{1}{2} \sigma^2_{C_3,b}} \quad (10)$$

$$CLD_{2,b} = 10 \log_{10} \frac{\sigma^2_{C_1,b}}{\sigma^2_{C_2,b}} \quad (11)$$

assuming that C_3 is the common channel which is attenuated by 3 dB and mixed to the other channels to derive the stereo output-downmix. Two downmix scalefactors $g_{C_1C_3,b}$ and $g_{C_2C_3,b}$ are calculated similar to Equation 8, taking care the 3 dB signal attenuation of input channel C_3 .

The input channel coefficients are mixed and scaled according to

$$\begin{bmatrix} \alpha_{\text{Downmix}_{1,b}} \\ \beta_{\text{Downmix}_{1,b}} \\ \gamma_{\text{Downmix}_{1,b}} \end{bmatrix} = g_{C_1C_3,b} \begin{bmatrix} \alpha_{1,b} + \frac{1}{2} \sqrt{2} \alpha_{3,b} \\ \beta_{1,b} + \frac{1}{2} \sqrt{2} \beta_{3,b} \\ \gamma_{1,b} + \frac{1}{2} \sqrt{2} \gamma_{3,b} \end{bmatrix} \quad (12)$$

$$\begin{bmatrix} \alpha_{\text{Downmix}_{2,b}} \\ \beta_{\text{Downmix}_{2,b}} \\ \gamma_{\text{Downmix}_{2,b}} \end{bmatrix} = g_{C_2C_3,b} \begin{bmatrix} \alpha_{2,b} + \frac{1}{2} \sqrt{2} \alpha_{3,b} \\ \beta_{2,b} + \frac{1}{2} \sqrt{2} \beta_{3,b} \\ \gamma_{2,b} + \frac{1}{2} \sqrt{2} \gamma_{3,b} \end{bmatrix} \quad (13)$$

For stereo-based encoding configuration, the direct scheme provides an advantage in terms of the derivation of matrix-compatible or 3D-stereo downmix [1][8]. The post-processing required to derive these downmixes can be implemented efficiently by pre-multiplying the 2x2 conversion matrix with the 3x2 stereo-downmix channel-coefficients matrix, practically adding no significant computational requirement for the generation of such modified downmix.

3.4 Down-mixed signals generation and modification

The last encoding block outputs the downmix channel-coefficients matrix which is used to derive the final down-mixed signals from the microphone array signals according to

$$\text{Downmix}_{i,k,n} = \alpha_{\text{Downmix}_{i,b}} W_{k,n} + \beta_{\text{Downmix}_{i,b}} X_{k,n} + \gamma_{\text{Downmix}_{i,b}} Y_{k,n} \quad (14)$$

The matrix dimension is 3x1 and 3x2 for mono- and stereo-based encoding configuration, respectively.

At any point in an encoding block, any signal mixing operations can be carried out by mixing the input-channel coefficients accordingly. For example, the residual signal for a TTO block can be obtained by subtracting and averaging the input-channel coefficients pair. The desired signal can then be derived by applying the resulting coefficients to the microphone array signals similar to Equation 14.

4 PERFORMANCE

In this section, the performance of the direct MPEG Surround encoding scheme is compared to the reference encoding scheme in terms of the differences in the generated output spatial parameter, downmix signals, and computational complexity.

4.1 Output spatial parameters and downmix

The framework to compare the output spatial parameters and downmix signals produced by the direct and reference encoding scheme is shown in Figure 6. For the reference encoding, the microphone array signal is first matrixed into 5.0 surround sound signals. These signals are passed to the encoder to extract the spatial parameters and down-mixed signals. The same microphone signals and virtual microphone matrix are passed to the direct encoder. The root mean square (rms) differences of the un-quantized spatial parameters produced by both encoders are calculated, and sample-by-sample comparison of the output downmix signals is performed.

Table 1 shows the *rms* analysis result for the un-quantized spatial parameters. The analysis was performed over 1084 audio frames with 20 parameter-band configuration. In MPEG Surround encoding, the Channel Level Difference (CLD) parameter is quantized using 30-level non-linear quantizer ranging from -150 dB to 150 dB with minimum quantization step of 2 dB, while the Inter-channel Coherence (ICC) parameter is quantized using 8-level non-linear quantizer ranging from -1.0 to 1.0 with minimum quantization step of 0.063.

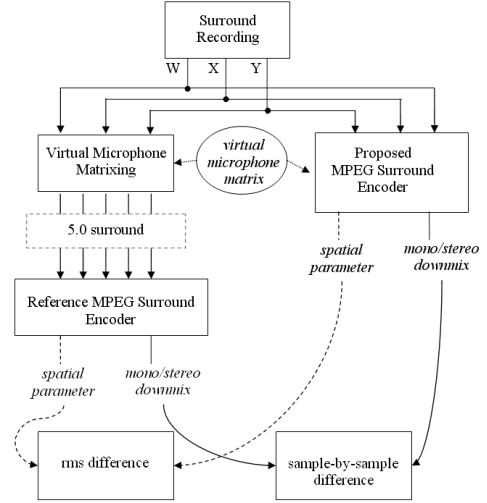


Figure 6: Framework for evaluation of spatial parameter and downmix output of the direct (proposed) vs reference MPEG Surround encoding scheme

It can be seen that the rms difference is way below half the minimum quantization step for both cases. Effectively, these differences do not affect the value of the quantized parameters sent as the spatial audio bitstream.

Encoding Mode	RMS Difference	
	CLD	ICC
5151 (mono-based)	0.0248209	0.002864
5152 (mono-based)	0.0241663	0.00267
525 (stereo-based)	0.0238245	0.002318

Table 1: Spatial parameter RMS difference of reference vs direct encoder

The sample-by-sample comparison of the down-mixed signals was performed on 30 second segment of the audio recording, where the samples are represented as 16-bit PCM data. The maximum difference obtained was 2 bits, where 98% of the differences were 1 bit. The resulting signal-to-noise ratio is 70.5 dB.

4.2 Complexity comparison

To compare the complexity of the reference and direct encoding scheme, the encoders have been implemented in C floating-point Window-based PC platform as well as ARM-based embedded platform.

The first evaluation was performed on PC platform with AMD Athlon 64 X2 2.4 GHz processor. Table 2 shows the encoding speed for a 30-second B-format recording. It can be seen that for the direct encoder, the encoding time has been reduced by at least 40%.

The second evaluation was performed on embedded platform with ARM Cortex A9 processor. The reference encoder was configured for 5.1 surround sound input

encoding (six playback signal channels). The direct encoder was configured for B-format input encoding (three microphone input signal channels) which output spatial parameters for a 5.1 decoding similar to the reference encoder. The direct encoding scheme requires approximately 50% less memory and 40% less processing power than the reference encoder.

Encoding Mode	Encoding Time (sec)	
	Reference Encoder	Direct Encoder
Mono-based (5151,5152)	29	16
Stereo-based (525)	31	19

Table 2: Comparison of encoding time for reference and proposed encoder on PC platform

5 CONCLUSIONS

An MPEG Surround encoding scheme for coincident surround sound recording has been presented. The main feature of the encoding scheme is the direct derivation of the spatial parameters and down-mixed signals from the microphone array signals, without the need to derive the actual surround sound signals. With the scheme, dependency of the memory and computational demand on the number of surround audio channels is eliminated. The required spatial parameters and down-mixed signals can still be fully derived without any compromise. Stereo-downmix modification can be integrated efficiently without adding significant computational requirements.

The overall computational demand is significantly lower than that required by the reference MPEG Surround encoder, saving at least 40% of the processing power.

REFERENCES

- [1] ISO/IEC 23003-1, "Information Technology – MPEG audio technologies – Part 1: MPEG Surround" (2007).
- [2] J. Herre, K. Kjørting, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Roden, W. Oomen, K. Linzmeier and K. S. Chong, "MPEG Surround – The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding", *Journal of the Audio Engineering Society*, vol. 56, no. 11, pp. 932-955 (2008).
- [3] H. Wittek, C. Haut and D. Keinath, "Double M/S – a Surround recording technique put to test", presented at the Tonmeistertagung 2006, viewed

23 April 2008 < http://www.schoeps.de/E-2004/PDFs/Schoeps_DoubleMS_Paper.pdf> (2006).

- [4] M. Sokol, H. La Torre, "Practical Surround Microphone Techniques", DPA Microphones workshop on mic techniques for multichannel audio, AES 24th International Conference, Banff, Canada (2003).
- [5] E. Benjamin, T. Chen, "The Native B-format Microphone: Part I", presented at the 119th AES Convention, New York, USA, preprint no. 6621 (2005).
- [6] ISO/IEC 11172-3, "Information Technology – Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1,5 Mbit/s, Part 3: Audio" (1993).
- [7] ISO/IEC 14496-3, "Information Technology – Coding of audio-visual objects, Part 3: Audio" (1999).
- [8] J. Breebaart, G. Hotho, J. Koppens, et. al., "Background, Concept, and Architecture for the Recent MPEG Surround Standard on Multichannel AudioCompression", *Journal of the Audio Engineering Society* vol. 55, no. 5, pp. 331-351 (2007).
- [9] Samsudin, S. George, "Low Complexity MPEG Encoding for Surround Sound Recordings", United States Patent #8,332,229.