DESIGN OF THE COUPLING SCHEMES FOR THE AC-3 CODER IN STEREO CODING

Chi-Min Liu, Member, IEEE, Wen-Chieh Lee and Shyh-Yan Juang
Department and Institute of Computer Science and Information Engineering
National Chiao Tung University, Hsinchu, 30050, Taiwan
Email: cmliu@csie.nctu.edu.tw

Abstract — **When applying the Dolby AC-3 coder for the stereo music compression, the coupling scheme that combines the two channels stereo audio signals in high frequency into one channel is the key technology for the Dolby AC-3 to achieve the bit rates lower than 96x2 kbits/sec while preserving high stereo audio quality. This paper proposes four coupling methods for the AC-3 encoder. These four methods vary with the complexity and performance. These four methods are compared through both subjective and objective tests. These four coupling methods are also combined with the dithering scheme and examined through subjective and objective tests. The result shows that the dithering scheme can effectively ease the coupling artifacts and enhance the audio quality.**

I. INTRODUCTION

THE coupling scheme, which applies the low perceptual sensitivity of the stereo signals in high frequency to audio compression, is the key technology to achieve near transparent quality at the bit rates below 96x2 kbits/sec. The principle of the coupling scheme is derived from the stereo irrelevancy from the auditory systems. The stereo irrelevancy expresses that the ability of the human auditory system to resolve the exact location of audio sources decreases with frequency. As stated in [2] and [3], the localization of the stereophonic image for the frequencies above 2 kHz is determined by the signal envelope instead of the signal fine structures. Following the stereo irrelevancy, the AC-3 coder has developed the coupling scheme to achieve efficient compression. However, the standard draft [1] illustrates the decoupling process for the decoder and leaves unmentioned the coupling process for the encoder. This paper proposes and compares four coupling methods for the coupling process of the AC-3 encoder.

Fig. 1 illustrates the block diagram for the coupling process in the Dolby AC-3. The audio sequences in stereo signal pairs are individually transformed into spectral lines and grouped into vectors referred to as the coupling bands. Fig. 1 shows the coupling process for one band corresponding to the same frequency range in a stereo signal pair. The bands from the left and the right channels are coupled through the coupling block in Fig. 1. The coupling process produces four outputs: the coupling vector or band C_{band}, the two coordinate values (s_L, s_R) and a phase flag p. The coupling band C_{band} is quantized and packed into the AC-3 bit stream. In this manner, the bands from the left and the right channels have been reduced into one band to achieve data reduction. The decoder multiplies the left coordinate (or the right coordinate with negative if the phase flag is on) with the coupling band to reconstruct the left band (or right band). For the coupling process, the design criterion for the encoder is to provide appropriately the four coupling information such that the stereo signal bands can be reconstructed with good listening quality.

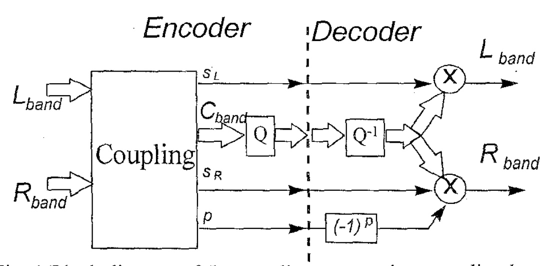


Fig. 1 Block diagram of the coupling process in a coupling band of the Dolby AC-3 codec

As mentioned above, the sensitivity of the stereophonic image for the frequencies above 2 kHz is determined by the signal envelope instead of the signal fine structures. The coupling scheme in AC-3 keeps the audio contents through the coupling band C_{band}, and preserves the envelope through the two coordinates (s_L, s_R). Since the

Manuscript received June 17, 1998

two bands have been reduced to one coupling band, it is impossible to reconstruct without loss the original two bands from the single band. Hence the design objective of the coupling is to keep envelope of the two bands through the coupling coordinates and minimizes the loss of the audio content through the coupling band. The coupling scheme is similar to the intensity coding in MPEG-1/2 audio coding. We have applied the Karhuner-Loeve transform to the intensity scheme to achieve the above objective [4]. The AC-3 has a higher potential to achieve a better performance than the intensity stereo in MPEG because of the two additional options: the phase flag and the dithering scheme. On these potential, this paper proposes four coupling methods for the AC-3. Section III gives the subjective and objective comparison for these four methods.

II. FOUR PROPOSED COUPLING METHODS

We developed four methods for the coupling scheme. These four methods differ in the complexity and the associated fidelity concepts as illustrated in Fig. 2-Fig. 5. Considering the SUM algorithm in Fig. 2, the coupling vector C_{band} is evaluated by summing the band signals R_{band} and L_{band} in the left and the right channels. For energy preservation, the two coordinate values (s_L, s_R) are calculated from the square root of the energy ratio for the R_{band} and C_{band}, and the ratio L_{band} and C_{band}. The phase flag P is fixed to be 0 in this method. The detailed algorithm of the SUM algorithm is illustrated as follows:

Encoding process for the SUM algorithm

1. *The phase flag evaluation process*
   $P_{band}=0$.
2. *The summation process*
   $C_{band}=L_{band}+R_{band}$.
3. *The coordinates evaluation process*
   $s_L=Energy(L_{band})^{0.5}/Energy(C_{band})^{0.5}$
   $s_R=Energy(R_{band})^{0.5}/Energy(C_{band})^{0.5}$

   where $Energy(S_{band}) = \sum_{bin\ in\ band} S_{bin}^2$. (1)

For the NORM_SUM algorithm in Fig. 3, the coupling vector C_{band} is calculated by summing the energy-normalized signals $R_{band}/Energy(R_{band})^{0.5}$, $L_{band}/Energy(L_{band})^{0.5}$. The two coordinate values (s_L, s_R) and the phase flag p are decided in the same way as the SUM algorithm. The NORM_SUM algorithm indicates that the larger value of L or R will not dominate during the summation process as the SUM algorithm. The detailed algorithm of the NORM_SUM algorithm is illustrated as follows:

Encoding process for the NORM_SUM algorithm

1. *The phase flag evaluation process*
   $P_{band}=0$.
2. *The summation process*
   $C_{band}=$
   $L_{band}/Energy(L_{band})^{0.5}+R_{band}/Energy(R_{band})^{0.5}$
3. *The coordinates evaluation process*
   $s_L=Energy(L_{band})^{0.5}/Energy(C_{band})^{0.5}$
   $s_R=Energy(R_{band})^{0.5}/Energy(C_{band})^{0.5}$
   where $Energy(S_{band})$ *is defined in (1).*

The KLT_MSE algorithm in Fig. 4 directly applies the Karhuner-Loeve (KL) transform to the coupling process in AC-3. The KL transform and the inverse KLT for N=2 can be viewed as the rotation matrix

$$\begin{bmatrix} I \\ E \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix}; \begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} I \\ E \end{bmatrix} \quad (2)$$

where L and R are signals of the left and right channels, and I and E are transformed intensity and error channel. The rotation angle α for the KL transform can be evaluated from

$$\tan(2\alpha) = \frac{2c_{lr}}{c_{ll} - c_{rr}}; \ -\frac{\pi}{2} \le \alpha < \frac{\pi}{2} \quad (3)$$

where C_{ll} and C_{rr} are the autocorrelation coefficients of the left and the right channels. C_{lr} is the cross-correlation coefficient of the left and the right channels. In least mean square error sense between decoded signals and input signals, the error channel is ignored and the KLT matrix becomes

$$\begin{bmatrix} I \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} L \\ R \end{bmatrix}; \begin{bmatrix} L \\ R \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} I \\ 0 \end{bmatrix}. \quad (4)$$

From (4), the coordinates of left and right channels for the KLT_MSE algorithm are $\cos\alpha$ and $\sin\alpha$, and the coupling vector can be obtained by $L_{band}\cos\alpha + R_{band}\sin\alpha$. In order to embed into the AC-3, the coordinates in AC-3 allow only positive values. Thus, by the phase modifier flag p, the coordinates of left and right channels and the coupling vector are changed to $\cos\alpha$, $\sin\alpha(-1)^p$ and $L_{band}\cos\alpha+R_{band}\sin\alpha(-1)^p$. From above, the KLT_MSE algorithm ensures the least mean square error of the original coupling vector and decoded coupling vector even the signals of the left and the right channels are negatively correlated. The detailed KLT_MSE algorithm is demonstrated as follows:

Encoding process for the KLT_MSE algorithm

1. *The rotation angle evaluation process*
   The rotation angle α is defined in (2).
2. *The phase flag evaluation process*
   $p=\begin{cases}1 & \text{if } \sin(\alpha)<0\\ 0 & \text{otherwise}\end{cases}$
3. *The summation process*
   $C_{band}=L_{band}\cos\alpha+R_{band}\sin\alpha(-1)^p$.
4. *The coordinates evaluation process*
   $s_L=\cos\alpha$
   $S_R=\sin\alpha(-1)^p$.

For the KLT_ENG algorithm in Fig. 5, a compromise between the SUM and KLT_MSE algorithm is considered. The two coordinate values (s_L, s_R) are decided from the square root of the energy ratio for *the* R_{band} and C_{band}, and the energy ratio for L_{band} and C_{band}. The detailed algorithm of the KLT_ENG algorithm is shown as follows:

Encoding process for the KLT_ENG algorithm

1. *The rotation angle evaluation process*
   The rotation angle α is defined in (2).
2. *The phase flag evaluation process*
   $p=\begin{cases}1 & \text{if } \sin(\alpha)<0\\ 0 & \text{otherwise}\end{cases}$
3. *The summation process*
   $C_{band}=L_{band}\cos\alpha+R_{band}\sin\alpha(-1)^p$.
4. *The coordinates evaluation process*
   $s_L=Energy(L_{band})^{0.5}/Energy(C_{band})^{0.5}$
   $s_R=Energy(R_{band})^{0.5}/Energy(C_{band})^{0.5}$
   where $Energy(S_{band})$ is defined in (1).

Among them, the methods in Fig. 4 and Fig. 5 are developed based on the KL transform. The KLT can minimize the square-errors during the coupling of two bands into one band. However, the KLT also leads to higher complexity than the other two methods.

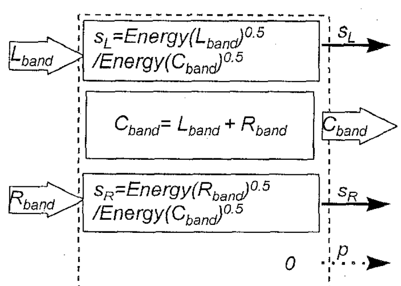


Fig. 2 The SUM algorithm for the coupling process

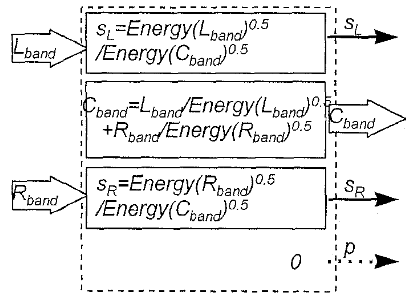


Fig. 3 The NORM_SUM algorithm for the coupling process

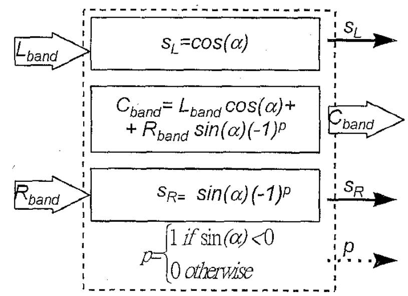


Fig. 4 The KLT_MSE algorithm for the coupling process

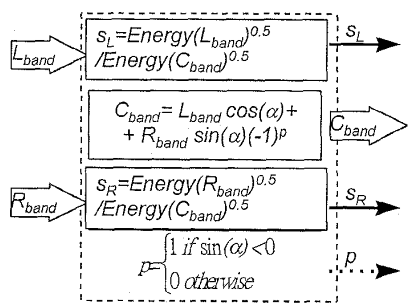


Fig. 5 The KLT_ENG algorithm for the coupling process

III. EXPERIMENTS ON THE COUPLING METHODS

The performances of the four coupling methods are compared through objective tests and subjective tests. A total of nine 20 sec stereo audio songs including vocal, symphony, piano and so on are taken as the materials for testing. The detailed descriptions of the test materials are listed in Table 1. The objective measure is verified by the segmental noise-to-masking ratio (NMR) value defined by averaging the NMR values in each coupling band in each frame as

$$NMR_{seg}=\frac{1}{F}\sum_{f}\left(\frac{1}{B}\sum_{b}SMR_{f,b}-SNR_{f,b}\right)$$

where the *SMR* stands for the signal-to-masking ratio in dB, the *SNR* for the signal-to-noise ratio in dB, F for the total audio frames, f for the frame number, B for the total coupling bands, and b for the coupling band number. Negative values of the NMR_{seg} indicate that the noise of the coded signal is inaudible, and larger negative values of NMR_{seg} indicate the noise may be more inaudible. The coupling scheme is performed in the range of 3.14 KHz to 12.45 KHz. The coupling methods are performed under high bit rate and the exponents are transmitted with D15 mode for six times in a frame. Table 2 illustrates the testing results. The results indicate that the KLT_MSE and KLT_ENG algorithm can have better NMR_{seg} values than the SUM and NORM_SUM algorithm. The SUM and NORM_SUM algorithms cause coupling artifacts and poor NMR_{seg} values due to signal cancellation when L_{band} and R_{band} are negatively correlated. We further consider the encoding for the bit rate at 128 kbits/s and the exponents strategy D15 is transmitted once per frame. The test results are summarized in Table 3 which indicates the order of the performance being the KLT_MSE, KLT_ENG, SUM and NORM_SUM algorithm.

In the subjective test under the critical bit rate at 128 kbits/s, the same test materials in Table 1 are evaluated. The results of the listening test show the order of the quality performance of the four coupling methods is the KLT_ENG, SUM, NORM_SUM, KLT_MSE algorithm. Although the excellent performance of the objective test, the KLT_MSE algorithm gives poor subjective performance due to some ringing noise. The noise may be due to the discontinuous coordinates across different bands in the KLT_MSE algorithm. To sum up, the KLT_ENG algorithm gives high performances on both objective and subjective tests because it takes the advantages from the KLT_MSE algorithm on the signal preservation and the SUM algorithm on the energy preservation.

IV. DITHERING ON THE COUPLING BANDS

In AC-3, dithering scheme is to add white noise to the coded bands in the decoding process. For low bit rate audio coding, quantization leads to the noises which are correlated with signals. Such a correlation is very sensitive for the human hearing systems. Especially, the coupling scheme can also lead to the artifacts as mentioned in last section. Dithering can reduce the artifacts from either the quantization or the coupling process. The four coupling methods presented in last section are examined through subjective tests when the dithering in AC-3 is applied. In our subjective listening test for the SUM and NORM_SUM algorithm, the dithering can significantly reduce the coupling noise. As a result, the quality from the KLT_ENG, SUM, and NORM_SUM algorithm become indistinguishable when the dithering is applied.

V. CONCLUDING REMARKS

In this paper, four coupling methods for the AC-3 encoder have been introduced. These four methods vary with the complexity and performance. Both subjective and objective tests have been conducted and demonstrated the performance of the KLT_ENG algorithm is better than other algorithms. We have also demonstrated that the dithering scheme gives great improvement on the quality of the coupling methods. With the dithering scheme, the performance of the four coupling methods is similar and the algorithm with low complexity will be more essential.

REFERENCES

[1] J. C. McKinney, R. Hopkins, "Digital audio compression standard (AC-3)," *Advanced television system committee*, Dec. 1995.

[2] D. H. Teh, A. P. Tan, "An Improved Stereophonic Coding Scheme Compatible to the ISO/MPEG Audio Coding Algorithm," *ICCS 1992*, pp. 437-441.

[3] J. Herre, K. Brandenburg, D. Lederer, "Intensity Stereo Coding," *96th AES Convention*, Feb. 1994, Preprint 3799.

[4] C. M. Liu and J. C. Liu, "A new intensity stereo coding scheme for MPEG1 audio encoder- layer I and II," *IEEE Trans. on Consumer Electronics*, Vol. 42, Aug. 1996, pp. 535-539.

Test song	Description
Symphony	The Choral symphony (Choral part)
Piano	Pure and clear piano
Violin	Violin playing from low to high frequency
Flute	Clear flute sound
Woman	Pure woman vocal song
Pipe	Pure pipe sound
man	Man vocal song; country music song
Violoncello	Violoncello sound in low frequency
Drum	Pure pipe sound & sudden and loud drum

Table 1 Testing audio segments and their descriptions

Algorithms	SUM		NORM_SUM		KLT_MSE		KLT_ENG	
D15 6 times	left	right	left	right	left	right	left	right
Symphony	-2.19	-2.75	-2.61	-0.95	-5.07	-7.17	-3.82	-6.18
Piano	-6.99	1.21	-5.72	1.29	-10.1	-6.01	-9.22	-4.72
Violin	5.90	7.81	5.74	10.2	1.42	-1.67	2.72	-0.66
Flute	-4.23	2.89	0.74	2.31	-10.1	-1.36	-9.49	0.02
Woman	1.17	8.35	1.26	9.23	0.45	1.17	1.36	1.96
Pipe	-12.4	-11.2	-12.1	-10.8	-12.9	-15.5	-12.4	-15.0
man	-2.91	16.5	-2.75	16.5	-3.19	-3.94	-2.97	-3.72
Violoncello	-8.61	-9.99	-9.56	-8.20	-8.23	-12.7	-7.35	-12.2
Drum	5.88	5.27	6.87	6.42	4.33	3.56	5.59	4.80

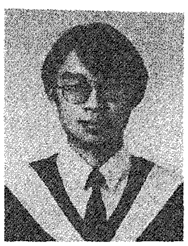
Table 2 NMR_{seg} values for the four proposed coupling methods under high bit rate with D15 mode 6 times per frame

Algorithms	SUM		NORM_SUM		KLT_MSE		KLT_ENG	
D15 1 times	left	right	left	right	left	right	left	right
Symphony	40.4	39.8	40.0	40.3	39.8	38.8	41.0	39.5
Piano	34.8	37.8	34.8	37.9	33.8	33.5	34.3	34.2
Violin	37.5	39.8	37.1	40.3	35.2	34.7	36.2	34.1
Flute	33.3	33.7	34.7	33.4	32.7	31.5	32.8	32.0
Woman	36.5	37.6	36.0	37.8	36.4	35.3	37.2	35.8
Pipe	33.3	34.4	33.3	34.4	33.0	33.1	33.4	33.5
man	35.9	44.0	35.4	44.0	36.0	36.0	36.2	36.6
Violoncello	36.2	36.3	35.3	36.3	36.6	36.1	37.1	36.2
Drum	36.9	37.3	37.4	37.3	36.3	36.5	37.0	37.6

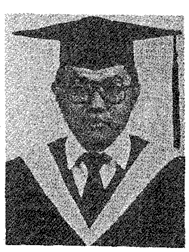
Table 3 NMR_{seg} values for the four proposed coupling methods under the bit rate of 128 kbits/sec with D15 mode once per frame

Chi-Min Liu received the B.S. degree in electrical engineering from Tatung Institute of Technology, Taiwan, R.O.C. in 1985, and the M.S. degree and Ph. D. degree in electronics from National Chiao Tung University, Hsinchu, Taiwan, in 1987 and 1991, respectively.

He is currently an Associate Professor of the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests include video/audio compression, speech recognition, radar processing, and application-specific VLSI architecture design.



Wen-Chieh Lee was born in Toayuan, Taiwan in Oct. 1972. He received the B. S. degree from the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan in 1995. He is currently a Ph. D. Candidate of the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan. His research interests are audio compression and real-time computer architecture.



Shyh-Yan Juang received the B.S. degree from the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan in 1995. He graduated from the Department of Computer Science and Information Engineering, National Chiao Tung University, Hsinchu, Taiwan in 1997. He is currently in the military service. His research interests are audio compression and real-time computer architecture.