A NEW INTENSITY STEREO CODING SCHEME FOR MPEG1 AUDIO ENCODER—LAYERS I AND II+

Chi-Min Liu and Jen-Chang Liu
Department of Computer Science and Information Engineering
National Chiao Tung University, Hsinchu, 30050, Taiwan
E-Mail: cm1iu@csie.nctu.edu.tw

Abstract

In MPEG 1 audio compression standard, the intensity stereo coding scheme is a key technology to achieve a bit rate lower than 2x128 kbit/s for stereophonic signals while preserving high audio quality. However, the intensity coding scheme has difficulties in signal cancellation and scalefactor consistency. This paper shows that the Karhunen-Loève (KL) transform can be applied to the intensity coding for solving the two problems. In this paper, we present two methods to implement the Karhunen-Loève transform in the MPEG 1 layers I and II. Among the two methods, one is fully compatible with the MPEG1 while the other need a slight modification on the MPEG coding process. All the two methods are tested to have better audio quality than the technique indicated by the MPEG 1 layers I and II.

1. Introduction

With the increasing demands on high fidelity audio signals, the MPEG 1 [1] audio compression standard is receiving a wide range of applications in the field of transmission and storage. MPEG 1 consists of three layers, and a higher layer has more complex computation and provides better audio quality at low bit rate compression. When the two channels of stereo signals are separately coded, layers I and II have a transparency quality at bit rates higher than 2x128 kbit/s [2][3]. For bit rate lower than 2x96 kbit/s, coding impairments may be audible. Intensity stereo coding can be applied to improve the quality of these stereophonic signals at low bit rates.

The intensity stereo coding scheme is based on the psychoacoustic effects of human auditory system at high frequency range. MPEG 1 intensity stereo coding adopted these effects by the computationally simplest way. Several addressed problems of the original MPEG1 intensity stereo coding and modification can be found in [4][5]. In [6], the idea of KL (Karhunen-Loève) transform has been considered to analyze the data redundancy between the stereo channels. Also, the authors have suggested the applying of the transform to intensity coding. In this paper, we propose two methods to implement the KL transform in the MPEG 1 layers I and II.

The rest of the paper is organized as follows: Section 2 describes two problems in the MPEG 1 layer I, II intensity stereo coding. Section 3 shows the KL transform and illustrates the potential to ease these two problems. Section 4 presents two methods to implement the KL transform in the intensity coding of MPEG 1. Section 5 gives a brief conclusion.

2. The Intensity Stereo Coding of MPEG 1

The principle of intensity stereo coding is derived from the property of stereo irrelevancy of the human auditory system. As stated in [4][5], the localization of the stereophonic image within a critical band for frequencies approximately above 2 kHz is determined by the temporal envelope, instead of the temporal fine structure. The intensity stereo coding in MPEG1 retains the envelope information through the scalefactors in left and right channels, and the fine structure through the jointed normalized samples. Fig. 1 shows the block diagram for the coding process.

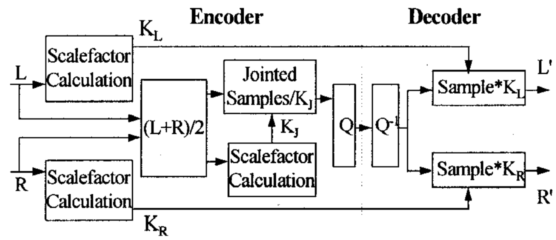


Fig. 1 Intensity stereo coding of MPEG1 in a high frequency band.

+ This work was supported in part by Acer Laboratories Inc. under contract C85098.

Manuscript received June 10, 1996

Consider the block diagram in Fig. 1, two problems arising from the process. The first problem is on the consistency between the scalefactors in the encoder and the decoder. As shown in Fig. 1, the signals from the left and right channel are summed together and jointly scaled by a scalefactor K_J, while the decoders utilize the scalefactors K_R and K_L to rescale the decoded samples. There is no direct relation between the K_J and the pair (K_R, K_L). Hence, the decoder and the encoder do not have consistent scalefactors. The second problem concerns with the signal cancellation in the summation process. During the summation process, when the signals in the left and the right channels do not have the same signal sign, the signals from the two channels will be mutually canceled and it is hard to reconstruct the canceled information. The researches in [4][5] try to ease these problems by modifying the transmitted scalefactors. Such an approach can ease the problem of the consistency of scalefactors, but can not provide help on the signal cancellation problem. This paper presents an approach to modify both the scalefactor calculation and the summation manner to ease the above two problems.

3. The Karhunen-Loève Intensity Stereo Coding Scheme Compatible to MPEG1

Our method is based on the KL transform presented in [6], which is modified to be compatible with the intensity coding in MPEG1 layers I and II. It is well known that KL transform can de-correlate the input data and achieve the best energy compaction properties. The KL transform matrix for two channel inputs is shown below.

$$R = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}, \quad -\frac{\pi}{2} \le \alpha \le \frac{\pi}{2} \qquad (1)$$

Here the angle α can be evaluated from

$$\tan(2\alpha) = \frac{2C_{lr}}{C_{ll} - C_{rr}} \qquad (2)$$

where C_{ll}, C_{rr} are the autocorrelation coefficient of left-channel and right-channel. And C_{lr} is the cross-correlation coefficient of the left and right channel. The derivation of equation (2) is straight forward from the definition of KL transform.

On the embedding of KL transform into MPEG1 intensity stereo coding, we first note that there are three components transmitted in the intensity mode: the joint samples, left and right scalefactors. In the KL transform, we discard the channel with smaller energy, and transmit the other channel, denoted as intensity channel, as the joint samples. Then the inverse transform appears as follows.

$$\begin{bmatrix} L' \\ R' \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} i \\ 0 \end{bmatrix} = \begin{bmatrix} i\cos\alpha \\ i\sin\alpha \end{bmatrix} \qquad (3)$$

From equation (3), we may say that the left and right scalefactor are $\cos\alpha$ and $\sin\alpha$ respectively. In fact, for the one-channel intensity signals, we can prove that $\cos\alpha$ and $\sin\alpha$ are the scalefactors that minimize the mean square error between the reconstructed two-channel signals and the original two-channel signals. Because the joint samples are normalized by the joint scalefactor before quantization in the encoder, we must re-normalized the decoded joint samples by multiplying the joint scalefactor in the decoder. Thus we must multiply the transmitted left and right scalefactor $\cos\alpha$ and $\sin\alpha$ by a joint scalefactor K_J. The block diagram of the our new intensity stereo coding based on KL transform is shown in Fig. 2.

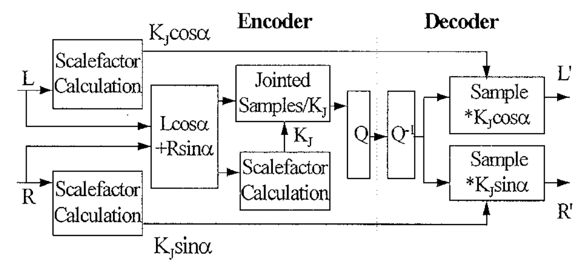


Fig. 2 The new intensity coding based on Karhuner-Loève transform in a high frequency band.

The KL transform is free from the scalefactor consistence problem mentioned in Section 2 since that scalefactors are obtained directly from the multiplying of the K_J and $\cos\alpha$ or $\sin\alpha$ as indicated in (3). For the signal cancellation problem, the KL transform has found the angle which minimizes the mean square errors and hence minimizes the cancellation effects.

On the embedding the algorithm into MPEG1, two issues need to be considered. The first issue is that the table used to quantize the scalefactors in MPEG1 allows only positive value. Since that the α

1.57079632679490	1.50825556499841	1.44546849562683	1.38217994061949
1.32014066445877	1.28569988654215	1.21988898320382	1.18639955229926
1.13432729805998	1.08292117925460	1.04719755119660	1.00423199617382
0.94842783823988	0.89666582012758	0.85495826656979	0.81275556136866
0.76879354899128	0.72273424781342	0.67513153293703	0.63571128540130
0.59740641664535	0.56007530622658	0.50536051028416	0.45281659474493
0.41834638644347	0.35542120169022	0.31782370392788	0.25268025514208
0.22053326092083	0.15689287102046	0.09388787510752	0.03125508849950

Table 1. Angle quantization table (list only positive 32 angles, the negative angles are with negative sign)

and hence scalefactor in the right channel may be negative, some modifications on the above KL transform need to be considered. We will propose two methods to resolve this problem in Section 4.

The second issue is on the energy normalization problem. As described above, the intensity stereo coding has the signal cancellation problem. Although the new technical has minimized the cancellation through the KL transform, the problem still exists. One resultant effect is the reduced energy of the decoded audio. The effect can be eased by scaling up the left and right scalefactor such that the left and right decoded samples have the same energy as the encoded samples.

4. Simulation and Test Results

4.1 Implementation methods

As mentioned in Section 3, direct applying of our modified intensity stereo coding algorithm to MPEG 1 has negative angle problem; that is, the negative angle can not be represented by the original scalefactor table. Here we propose two implementation methods for simulation.

In the first method, when the angle α is positive, we perform our intensity stereo coding algorithm as shown in Fig. 2; when the angle is negative, we perform the original MPEG 1 intensity stereo coding. In this way, the method can be totally compatible to the MPEG 1 standard in the sense that the same decoder as MPEG1 can be used to decode the bitstreams encoded by the method. However, the presented method has sacrificed parts of the potential of the KL transform. This method is denoted as compatible coding method in the following simulation.

In the second method, we transmit the joint scalefactor K_J and the angle α to approximate the KL transform indicated in Fig. 2. The joint scalefactor is quantized as six bits based on the look-up table designed for the scalefactors in MPEG1. The rotation angle α is also quantized as six bits but based on Table 1. The table shows the 32 positive quantized angles that are used to quantize the legal angles ranging from 0 to $\pi/2$. The negative angles have the same values but negative signs. Table 1 is designed to have a uniform quantization effects on the terms cos α and sin α such that the effective scalefactor $K_J\cos\alpha$ and $K_J\sin\alpha$ have the similar quantization effects as that in the scalefactors of MPEG1. This method can approximate the KL transform under the same bit rate as MPEG 1, but a slight modification on the decoder is required to decode the bitstreams encoded by the method. This method is denoted as noncompatible coding method in the following simulation.

4.2 Objective Quality Evaluation

Traditional objective quality measurements, such as SNR, are not suitable for the quality assessment of the perceptual coding systems, because they did not consider the psychoacoustic effects used in such systems. Several objective quality measurements are proposed over the past years [7], but none is taken as a widely acceptable standard. This is due to our knowledge of the human auditory system is still not enough. Moreover, these proposed measurements are valid only for monophonic signals, and failed to measure the quality of stereophonic signals.

Although no specific measurement standard can be used, there are still several commonly used measurements in the literature, such as Segmental-SNR, and NMR (Noise-to-Masking Ratio) [6][8]. Here, to check the quality, we adopt the analysis results from the psychoacoustic model in MPEG1 to measure the difference between the original signal and the decoded signal. The block diagram of the measurement system is shown in Fig. 3.

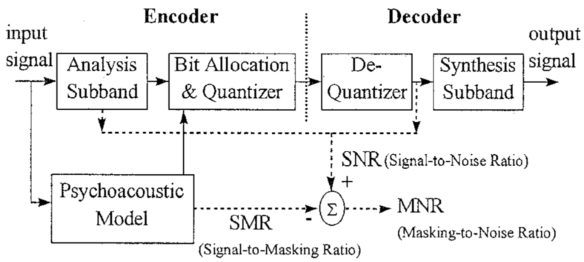


Fig. 3 The block diagram of objective evaluation system.

We extract the $SMR_{k,f}$ (Signal-to-Masking Ratio) information from the psychoacoustic model of MPEG 1 encoder for subband k and frame f. The SMR(k,f) indicates the noise levels that the quantized noise is inaudible. We also calculate the $SNR_{k,f}$ (Signal-to-quantization-Noise Ratio) values in subband k and frame f. Then we have a MNR value for each subband in every frame.

$$MNR_{k,f}(\text{dB}) = SNR_{k,f}(\text{dB}) - SMR_{k,f}(\text{dB}) \quad (4)$$

Note that a positive MNR value means that the sound signals mask the quantization noise, and no audible impairment can be heard. On the contrary, a negative MNR means audible impairment. The average MNR for an audio sequence is calculated as

$$MNR = \frac{1}{F}\sum_{f}\left(\frac{1}{K_f}\sum_{k} MNR_{k,f}\right), \quad (5)$$

where K_f is the number of subbands that have positive SMR in frame f, which means these subbands are necessary to be coded, in a frame, and F is the total number of frames in an audio sequence. The measurement may be reasonable in the sense that the psychoacoustic model in MPEG 1 can well catch the human auditory system. We will calculate average MNR values of the left and right channel for each test sequence in the following simulation.

4.3 Simulation Results

We use different music instruments for evaluation. The test sequences are directly read from CD. The contents of each test sequence are shown in Table 2. We select the compressed bit rate 2x64 kbits/s for tests. Also, the psychoacoustic model 1 in MPEG 1 is adopted for the objective quality measure. The MNR results of implementation in MPEG 1 layers I and II are shown in Table 3 and Table 4, respectively. All the test results show that the two KL intensity coding methods can have a lower MNRs than the original MPEG intensity coding method. Among the two KL intensity coding methods, the noncompatible coding method can have a better performance than the compatible one.

No.	Abbreviated	time	contents
1	Carmen	5.86 sec	orchestra
2	Songs	5.00 sec	female song
3	Huqin	10.03 sec	huqin solo
4	Drum	5.52 sec	drum solo
5	Violin	17.54 sec	violin solo
6	Orchestra	5.00 sec	orchestra
7	Guitar	15.46 sec	guitar, castanets

Table 2. Test sequences.

Methods / Test	Original MPEG	Compatible Coding	Nonompatible Coding
1. Carmen	-9.2147 -9.7622	-7.4974 -9.5023	-6.0669 -8.6928
2. Songs	-20.0317 -20.0534	-19.3518 -19.3164	-18.8416 -18.7605
3. Huqin	-6.3173 -6.2085	-6.1051 -5.5827	-5.9114 -5.3510
4. Drum	-24.0901 -23.9808	-23.5307 -23.6609	-23.1951 -23.4160
5. Violin	-8.7437 -7.6760	-8.4205 -6.8404	-6.7604 -5.0461
6. Orchestra	-13.0901 -14.0977	-12.9893 -13.4846	-11.9143 -12.4876
7. Guitar	-11.3882 -9.7984	-11.1488 -8.9466	-10.9694 -8.6443

Table 3. MNR (dB) values in layer I. In each box, the upper value is for the left channel, the lower value is for the right channel.

Methods / Test	Original MPEG	Compatible Coding	Noncompatible Coding
1. Carmen	-0.5985 -1.3170	-0.2276 -1.2783	0.6296 -0.7510
2. Songs	-7.3448 -7.3165	-6.5771 -6.4914	-5.6519 -5.6685
3. Huqin	-1.2521 -1.2330	-0.8507 -0.5945	-0.7297 -0.7566
4. Drum	-5.1192 -5.2201	-4.5126 -4.6360	-3.9989 -4.8985
5. Violin	-3.0766 -2.1584	-2.9204 -1.7142	-1.5388 -0.3412
6. Orchestra	-6.3791 -6.7642	-5.7489 -6.3137	-4.1817 -5.0613
7. Guitar	-4.4968 -3.6040	-4.0042 -2.8042	-3.8239 -2.7585

Table 4. MNR (dB) values in layer II. In each box, the upper value is for the left channel, the lower value is for the right channel.

5. Concluding Remarks

Intensity stereo coding in MPEG1 exploits the stereo irrelevancy property to achieve better audio quality at low bit rates. However, the intensity coding scheme has difficulties in signal cancellation and scalefactor consistency. This paper have shown that the Karhunen-Loève (KL) transform can be applied to the intensity coding for solving the two problems. In this paper, we have present two methods to implement the Karhunen-Loève transform in the MPEG 1 layers I and II. Among the two methods, one is fully compatible with the MPEG1 while the other need a slight modification on the MPEG coding process. All the two methods have been tested to have better audio quality than the technique indicated by the MPEG 1 layers 1 and II.

References

[1] Part 3 of ISO/IEC Committee Draft 11172 on "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s," Mar 1992.

[2] T. Rydén, et. al., "The SR report on the MPEG/Audio subjective listening tests in Stockholm April/May 1991", ISO/IEC JTC1/SC2/WG11 MPEG 91/010, May 1991.

[3] H. Fuchs, "Results of the ISO/MPEG subjective listening tests in Hannover", ISO/IEC JTC1/SC2/WG11 MPEG 91/331, November 1991.

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

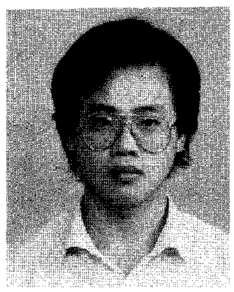
[5] J. Herre, K. Brandenburg, and D. Lederer, "Intensity Stereo Coding," 96th AES Conv. Feb. 1994, Preprint 3799.

[6] R. G. v. d. Waal and R. N. J. Veldhuis, "Subband Coding of Stereophonic Digital Audio Signals," IEEE ICASSP 1991, pp. 3601-3604.

[7] K. Brandenburg, "Perceptual Models for the Prediction of Sound Quality of Low Bit-Rate Codecs", AES 12th Inter. Conf., 1993.

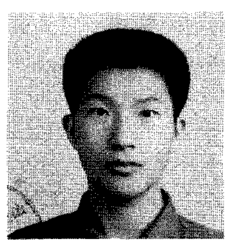
[8] M. Keyhl, J. Herre, Th. Schmidmer, "NMR Measurements of Consumer Recording Devices Which Use Low Bit-Rate Audio Coding", 94th AES Conv., Berlin 1993, preprint 3616.

Biographies



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.



Jen-Chang Liu was born in Keelung, Taiwan, in 1972. He graduated from the Department of Computer Science and Information Engineering, National Chiao Tung University (NCTU), Taiwan, R.O.C., in 1994. He is now a master student in the Institute of Computer Science and Information Engineering, NCTU. His current research interests are audio compression and artificial intelligence.