Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03), London, UK, September 8-11, 2003

M/S CODING BASED ON ALLOCATION ENTROPY

Chi-Min Liu, Wen-Chieh Lee, and Yo-Hua Hsiao

Department of Computer Science and Information Engineering
National Chiao Tung University, Hsinchu, Taiwan, ROC
cmliu@csie.nctu.edu.tw

ABSTRACT

Main/Side (M/S) coding is a module deigned to reduce the channel irrelevancy and redundancies for multichannel coding in current audio standards. However, the success of this module relies on the careful design on four aspects: M/S decision, psychoacoustic model for the M/S channels, bits allocated between the M/S channels, and bit allocation in the M/S channels. This paper presents an efficient M/S method well designed from the four aspects. We conduct the experiments through the MP3 encoder to show the better quality and complexity than the Lame.

1. INTRODUCTION

For the current audio coding, M/S (Middle/Side) coding is the kernel technique to efficiently reduce the redundancy and irrelevancy in stereo channels [1]. For channel numbers greater than two, a way used in current standards like MPEG2-Layer III [2] and MPEG4-Layer IV [3][4] is to separate the channels into pairs and then apply the M/S coding to each pair. This paper presents an efficient M/S coding technique for the audio coding standard.

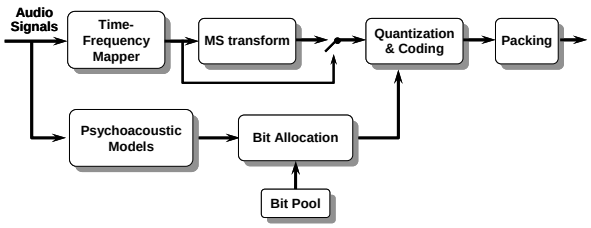


Figure 1 *Block diagram of M/S coding.*

M/S coding is an extended perceptual audio coding which includes the M/S transform module to transform L/R signals to M/S signals. Figure 1 illustrates the block diagram of a perceptual coding with M/S transform. The audio signal is segmented into overlapped blocks and transformed into frequency domain through the time-frequency mapper. The L/R signals are then transformed to M/S signals if there is coding gain. These signals are then quantized and coded with the parameters decided by the bit allocation. The psychoacoustic model analyzes the signal contents and calculates the associated perceptual resolution on the human hearing systems. According to the perceptual resolution and the available bits, the bit-allocation decides the suitable quantization manner to fit the bit rate. The packing module packs all the coded information with a format specified by standards. There are mainly four issues in the M/S coding. The first issue is on the psychoacoustic model for M/S signals. Since what we hear is the L/R signals instead of the "virtually" M/S signals, the existing psychoacoustic models, which are derived for the perceptual effects from the sounds, are not directly applicable to the M/S signals. The second issue is on the decision to code the signals based on either the stereo data or the M/S data. The decision concerns with the measurement on the coding gain from M/S coding over L/R coding. The third issue is on the available bits allocated to the two M/S channels. For stereo channels, the available bits are equally allocated to the stereo channels. The allocation policy should be modified to fit the coding merits in the M/S channels. The forth problem is bit allocation in the M/S signals according to a rate-distortion metric.

On the M/S coding, Johnston et. al. [8][9] have outlined the experimental and theoretic framework. The M/S coding used in the most representative mp3 encoder, Lame, has been developed based on the research [9] with some subtle parameter tuning. This paper presents a new M/S coding method on Lame to show the better quality and lower complexity over the existing method. The new method is derived with a new perceptual criterion referred to as the allocation entropy (AE). The AE is different from the perceptual entropy (PE) which calculates the minimum amount of bits that has to be transmitted to have the transparent audio quality [8][9]. The PE has not reflected the bits required for the cases where the transparent quality is not achievable under limited bit rates. For coding, it is always the objective to have the best bit-constrained quality instead of the transparent quality. The AE can well reflect the bits to have the graceful degradation and have put into consideration the bandwidth-proportional noise-shaping criterion [10].

2. M/S CODING

M/S stereo coding is derived from the L/R stereo signals in frequency domain to have the coding merits. Let us take for example the M/S coding in MPEG1 Layer III.

$$M[k] = \frac{L[k] + R[k]}{\sqrt{2}}, \tag{1}$$

and

$$S[k] = \frac{L[k] - R[k]}{\sqrt{2}}, \tag{2}$$

where *L[k]* and *R[k]* are the L/R (Left/Right) frequency lines and *M[k]* and *S[k]* the M/S (Middle/Side) frequency lines. The alternative representation leads to the design issues on M/S

decision, psychoacoustic model for the M/S channels, bits allocated between the M/S channels, and bit allocation in the M/S channels.

The psychoacoustic model for M/S signals in Lame is developed based on the research [9]. The M/S coding has gained much better performance than the approach in MPEG reference code [1]. The M/S signals are fed into the psychoacoustic model which is the same model in L/R coding to have the masking threshold. Then, the final masking threshold is determined by including masking level difference (MLD) [7]. The psychoacoustic model consists of the 1024-point FFT, the grouping into the partition bands, the applying of the excitation functions, and the evaluation of final signal-to-masking ratio.

In the psychoacoustic model, the perceptual entropy *(PE)* [9] which reflects the minimum bits required to have the transparent quality is evaluated for the left, right, middle, and side channels. The *PE* is defined as

$$PE = \sum_{g=0}^{1}\sum_{i=0}^{P} W_i * \log_{10}(SMR_i), \tag{3}$$

where P is the number of total partition band, W_i is the number of spectrum lines in each partition. The partition band has a bandwidth proportional to the critical bandwidth. The signal-to-masking ratio SMR_i is defined as

$$SMR_i = \begin{cases} E_i / T_i & , if\,(E_i > T_i) \\ 0 & , if\,(E_i \le T_i) \end{cases}, \tag{4}$$

where E_i and T_i are the spectrum energy and the masking thresholds in the partition band *i*. In MPEG Layer III, the frame is the unit to have the MS coding and each frame consists of two granules labeled with index *g* in (2).

The coding merits of MS coding can be obtained if the bits required for M/S is lower than L/R coding. The *PEs* can be evaluated for the left, right, middle, and side signals. The MS coding can be decided by comparing the *PEs* from the L/R and the M/S channels. In addition to the *PEs*, Lame also includes the masking ratios. Figure 3 illustrates the process to decide the MS coding in Lame.

The method used in Lame is to allocate the available bits according to the masking thresholds ratio between the M/S channels. However, the masking threshold alone can not reflect the required bits. After decoding the available bits in the two M/S channels, the rate-distortion loop decides the bits used to allocate to the scalefactor bands to have the best quality.

3. NEW M/S CODING METHOD

This paper develop the MS coding based on the allocation entropy *(AE)* instead of the *PE*. Although *PE* can well represent the theoretical bits required to have the transparent quality. However, it does not reflect the coding scheme adopted in the encoder. Also, the objective of encoder is to have the best quality under limited bits instead of achieving the transparent quality.

3.1 M/S Psychoacoustic Model

The psychoacoustic model is trying to simulate the human hearing system and to give the proper masking thresholds for quantization. The masking model for L and R channels is already constructed in standard. However, it seems not so reasonable to put on the same procedure to M and S channels. Although there are some assumptions in speaker location to have sensible explanation on M signals, the applicability to the S signals and the ear phones needs to be investigated. Besides, the complexity of the psychoacoustic models has contributed to a factor for more than 15% in the L/R coding. The additional complexity from the M/S psychoacoustic models leads to high overhead on the M/S coding.

To derive the masking thresholds /S channels, consider the reconstructed left and right channels as

$$L'[k] = \frac{M'[k] + S'[k]}{\sqrt{2}} \tag{5}$$

and

$$R'[k] = \frac{M'[k] - S'[k]}{\sqrt{2}}, \tag{6}$$

where *M'[k]* and *S'[k]* are the M/S frequency lines requantized from the decoder. Due to the quantization errors, we can rewrite reconstructed signals as

$$L'(k) = L(k) + N_L(k) = \frac{M(k)+S(k)}{\sqrt{2}} + \frac{N_M(k)}{\sqrt{2}} + \frac{N_S(k)}{\sqrt{2}} \tag{7}$$

and

$$R'(k) = R(k) + N_R(k) = \frac{M(k)-S(k)}{\sqrt{2}} + \frac{N_M(k)}{\sqrt{2}} - \frac{N_S(k)}{\sqrt{2}}, \tag{8}$$

where $L'(k)$ and $R'(k)$ denote the reconstructed spectrums for the left and right channels. $N_L(k)$ and $N_R(k)$ are the associated noise for the left and right channels, respectively. The variances of $N_L(k)$ and $N_R(k)$ should be below the masking thresholds of *L* and *R* signals to have the transparent quality. The variance with respect to the partition bands should be constrained by

$$\sigma^2_{N_L(i)} = \frac{\sigma^2_{N_M(i)}}{2} + \frac{\sigma^2_{N_S(i)}}{2} \le T_{L(i)} \tag{9}$$

and

$$\sigma^2_{N_R(i)} = \frac{\sigma^2_{N_M(i)}}{2} + \frac{\sigma^2_{N_S(i)}}{2} \le T_{R(i)}. \tag{10}$$

The sufficient conditions satisfying the inequalities (9) and (10) are

$$\sigma^2_{N_M(i)} \le Min(T_{L(i)}, T_{R(i)}) \tag{11}$$

and

$$\sigma^2_{N_S(i)} \le Min(T_{L(i)}, T_{R(i)}). \tag{12}$$

The thresholds can be used to replace the thresholds derived directly from M/S frequency lines. Next section shows that the replacement leads to much worse quality. To check the problems, let us define the allocation entropy which reflects the bits required to have the best quality as follows

$$AE = \sum_{g=0}^{1}\sum_{j=0}^{S} W_j * \log_{10}(SMR_j) \tag{13}$$

and

$$SMR_j = \begin{cases} \dfrac{E_j}{T_j B_j} & , if\,(E_j > T_j) \\ 0 & , if\,(E_j < T_j) \end{cases}, \tag{14}$$

where index *j* is the scalefactor band instead of the partition band *i* in (3). B_j is effective bandwidth proposed in [10].

According to [10], the bit allocation criterion leads to the optimum graceful degradation when noise is higher than masking threshold is to have a noise proportion to effective bandwidth. The effective bandwidth is derived from the critical band with bandwidth about one-third to one-forth of the critical bandwidth. In general, the higher spectrum bands often have wider effective bandwidth and should have a higher noise shape. Also, the *AE* is evaluated with the unit of the scalefactor bands to match directly the units in quantization and encoding process. The masking thresholds on the scalefactor bands can be derived directly from the masking thresholds for the left and right channels with respect to the scalefactor bands.

$$T_{M(j)} = T_{S(j)} \le Min(T_{L(j)}, T_{R(j)}). \quad (15)$$

3.2 M/S Switching

The *AE* for the left, right, middle, side channels can be obtained by directly applying (14). The coding gains can be evaluated by comparing the *AEs* for M/S and L/R channels. However, two factors introduce a discount on the factor. First, the sufficient conditions in (11) and (12) indicate that the decision is too conservative to reflect the coding merits. Second, the L/R channels are usually allocated equal bits. Hence, the direct summation on the *AEs* of the left and right channels can not reflect the situation where two channels require different bits. To include the two situations, we put a discount on the *AEs* from M/S channels.

3.3 Available Bits in the M/S channels

Now that we have obtained the allocation entropy, the available bits between the two M/S channels can be derived directly as

$$Bit_M = \frac{AE_M}{AE_M + AE_S} * B \quad (16)$$

and

$$Bit_S = \frac{AE_S}{AE_M + AE_S} * B, \quad (17)$$

where *B* represents the total available bits of a frame. The detail allocation process can then integrated with the bandwidth-proportional bit allocation in [10] to shape the noise.

4. EXPERIMENTS

Figure 2 illustrates the objective score of the various M/S stereo coding based on the system [10] and Lame 3.88. Here we have adopted for objective quality measure the PEAQ (perceptual evaluation of audio quality) which is the recommendation system by ITU-R Task Group 10/4. The objective difference grade (ODG) is the output variable from the objective measurement method. The ODG values should range from 0 to -4, where 0 corresponds to an imperceptible impairment and -4 to an impairment judged as very annoying. The tracks used are the same as that in [10]. In Figure 2, the first two columns denote the Lame without and with MS coding, respectively. The third and fourth columns are the system [10] without and with the M/S scheme in Lame. The fourth column is the NCTU-Lame with masking thresholds replaced with the one evaluated by (11) and (12). The fifth and sixth columns are the system [10] using the proposed MS coding without and with discounting on the switching condition. In Figure 2, the worst ODG and the best ODG are also depicted for each scheme.

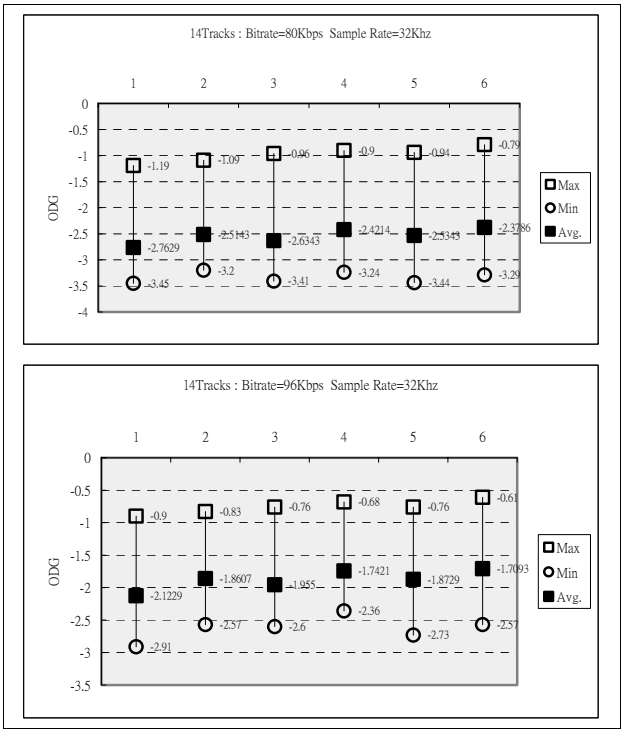
The MS coding has reduced the gaps between the worst score and the best score. Figure 2 illustrates the ODGs for various encoding conditions. The bit rates ranging from 80 kbps to 162 kbps are included in Figure 2. Also, the 14 tracks are changed into 50 tracks to have the intensive tests. All the above experiments conclude that the new M/S coding scheme can perform better than either the M/S coding in Lame applying the direct psychoacoustic model to the M/S signals or the M/S coding based on *PE*.

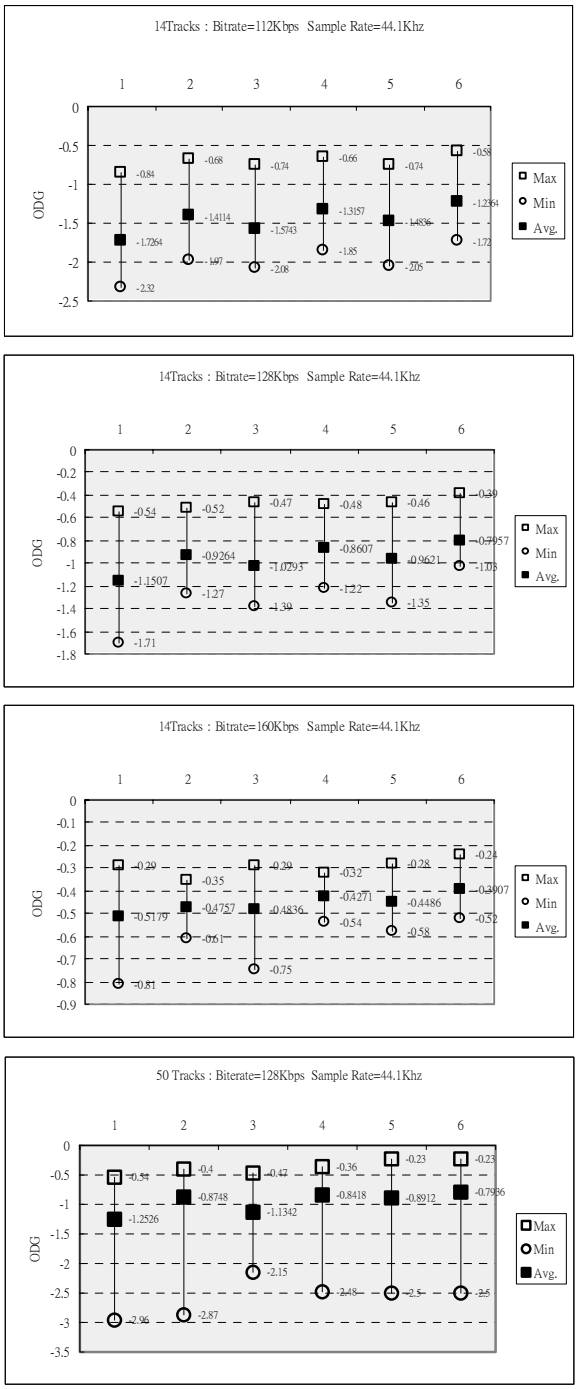
5. CONCLUSION

This paper has presented a new M/S coding based on the allocation entropy (*AE*). The *AE* is different from *PE* in the following aspects:

- The *AE* reflects the coding schemes leading to the best quality under constrained bit rates instead of the transparent quality.
- The *AE* includes the noise-shaping proportional to the effective bandwidth to reflect the bit allocation method used in [10].
- The *AE* has the basic units using the scalefactor bands instead of the partition bands to match the quantizers used in audio standards.

The method can provide better performance than that in Lame and also has the lower complexity due to the applying of the psychoacoustic model calculation only to the L/R channels not to M/S channels.





1. Lame 3.88: without MS coding
2. Lame 3.88: with MS coding
3. NCTU Lame: without MS coding
4. NCTU Lame: with MS coding
5. NCTU Lame: with new MS coding but without discounting
6. NCTU Lame: with new MS coding and discounting

Figure 2 *Objective measurements through the ODGs for different M/S coding methods.*

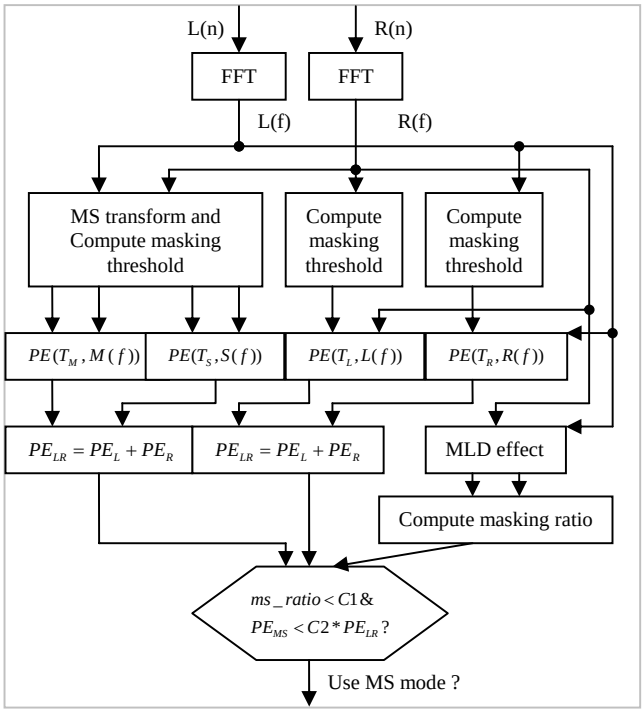


Figure 3 *Flow chart of the Lame M/S switching.*

7. REFERENCES

[1] ISO/IEC 11172-3:1993, "Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to 1.5Mbit/s, Part 3: Audio".

[2] ISO/IEC 13818-3:1997, "Information Technology–Generic Coding of Moving Pictures and Associated Audio, Part3: Audio".

[3] ISO/IEC 13818-7:1997, "Information Technology–Generic Coding of Moving Pictures and Associated Audio, Part7: MPEG-2 Advanced Audio Coding, AAC".

[4] ISO/IEC 14496-3:1999, "Information Technology–Coding of .Audiovisual objects, Part3: Audio".

[5] LAME, website http://www.mp3dev.org/mp3/ .

[6] J.D. Johnston, "Estimation of Perceptual Entropy Using Noise Masking Criteria", *ICASSP*, 1988, pp.2524-2527.

[7] J.D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *Proc. IEEE J. Select. Areas Commun.*, vol. 6, Feb. 1988, pp.314-323.

[8] J.D. Johnston, "Perceptual Transform Coding of Wideband Stereo Signals," *ICASSP*, 1989, pp.1993-1996.

[9] J.D. Johnston and A.J. Ferreira, "Sum-Difference Stereo Transform Coding," *ICASSP*, 1992, pp.569-571.

[10] C.M. Liu, W.J. Lee, and R.S. Hong, "A Bandwidth-Proportional Noise-Shaping Criterion and the Associated Fast Bit Allocation Method for Audio Coding," *Digital Audio Effect (DAFX-02)*, Sep. 2002, pp. 26-28.

[11] ITU Radiocommunication Study Group 6, "Draft Revision to Recommendation ITU-R BS.1387- Method for objective measurements of perceived audio quality".

[12] J. Herre and E. Eberlein, "Combined Stereo Coding," *93rd AES Convention 1992*, October1-4, San Francisco.