Memory and Computationally Efficient Psychoacoustic Model for MPEG AAC on 16-bit Fixed-point Processors

Shih-Way Huang, Liang-Gee Chen
Grad. Inst. of Elect. Eng. and Dept. of Electrical Engineering,
National Taiwan University,
Taipei, Taiwan, ROC
shihway@video.ee.ntu.edu.tw

Tsung-Han Tsai
Dept. of Electrical Engineering
National Central University,
Chung-Li, Taiwan, ROC
han@ee.ncu.edu.tw

***Abstract*—When MPEG AAC encoders are implemented on 16-bit fixed-point processors, the constraints of low complexity and limited data word length become challenges especially on the critical algorithm, Psychoacoustic Model (PAM). This paper proposes a very efficient PAM design concerning the reduction of memory requirement and computational complexity while maintaining the audio quality.**

I. INTRODUCTION

Portable electronic devices with audio playback and recording have prevailed. 16-bit fixed-point processors are preferred for the devices because of its low cost and power. However, when the audio encoder is implemented on a 16-bit fixed-point processor, there are two challenges especially on the critical algorithm, Psychoacoustic Model (PAM) [1]. One challenge is the calculation of complex operations that large computation or memory (look-up table) is required, and the other is the limited word length to represent the audio signal energy and masking threshold with sufficient precision that is 48 bit in theory. Our previous work [2] deals with the first problem and proposes methods to reduce the computational complexity with look-up tables and modified MDCT-based PAM from [3]. However, memory requirement such as table size and data memory storage and bandwidth are not concerned together. Dealing with the second challenge, [4] states that 32-bit logarithmic representation is enough for those data. However, it is still difficult to implement these data on 16-bit processors.

This paper presents a very efficient PAM design by logarithmic based algorithm and 16-bit logarithmic representation for the energy and masking threshold data. These approaches not only reduce the computational complexity with smaller look-up tables but also reduce the data memory storage and bandwidth while maintaining the audio quality.

II. PAM REVIEW

Fig. 1 illustrates the block diagrams of MPEG AAC (above) and PAM (below) based on [1]. PAM calculates a masking threshold, which is the maximum distortion energy masked by the signal energy for each coding partition. Moreover, it also predicts a block type, which determines the block length used in other parts of the encoder. Meanwhile, Modified Discrete Cosine Transform (MDCT) transforms input audio samples in time domain into spectrums in frequency domain. The frequency spectrums are then transferred to Spectral Processing (SPP), which includes some tools to remove redundancies such as Temporal Noise Shaping (TNS) and joint coding. The spectrums are then non-uniformly quantized based on the masking threshold and available number of bits to minimize the audible quantization error in the Quantization Loop (Q Loop).

PAM can be divided into two parts. One part is T/F transform, and the other part is Threshold Generation. T/F transform is composed of Steps 1-2. Threshold Generation is composed of Steps 3-13. The algorithms of these steps are reviewed. In Steps 1-2, PAM normalizes the time-domain samples as input and transforms them into frequency-domain spectrums of real part *r(w)* and imaginary part *i(w)* by FFT. Real-part spectrums are used to calculate the partitioned energy and imaginary-part spectrums are used to calculate the weighted unpredictability measure *c(b)* in Steps 3-4. In Step 5, both partitioned energy and unpredictability are convolved with the spreading function in order to estimate the effects across the partitioned bands. Tonality index is estimated in Step 6 to indicate whether a signal is tonal-like. Signal-to-Noise Ratio (SNR) is calculated in Step 7 and the masking partitioned energy threshold *nb(b)* is calculated in Steps 8-10 to estimate the masking curve. Perceptual Entropy (PE) is calculated in Steps 11-12 to determine the block type used in the MDCT, Q loop, and SPP. Block type decision requires detecting whether there is a transient signal

0-7803-8834-8/05/$20.00 ©2005 IEEE.

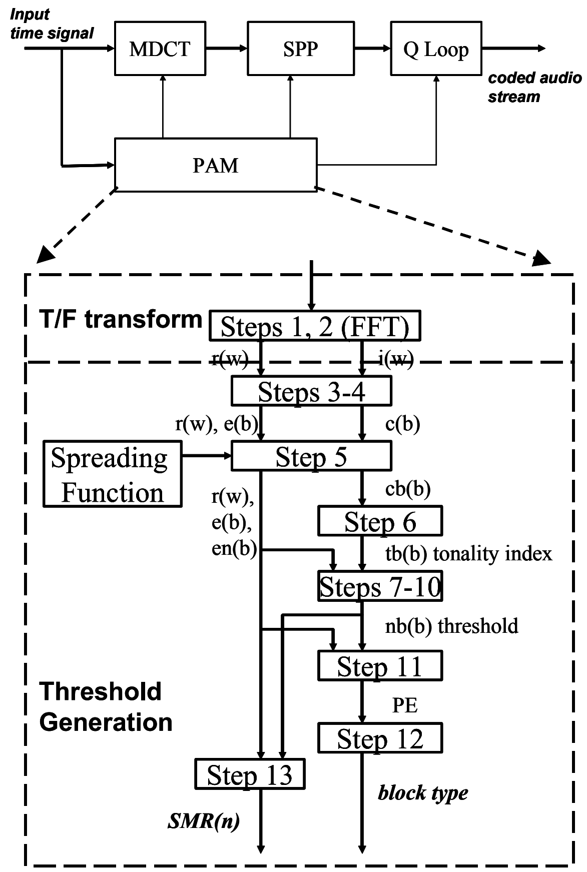

Figure 1. Block diagrams of MPEG AAC (above) and its critical algorithms PAM (below) based on [1]

in the frame. Finally, Signal-to-Mask Ratio (SMR) is computed in Step 13 as output. *w*, *b*, and *n* indicate indices in the spectral line domain, the threshold calculation partition domain, and the coder scalefactor band domain, respectively.

III. The Challenges of Implementation

There are two challenges when PAM is implemented on the 16-bit fixed-point processors. One is the calculation of complex functions, and the other is the limited word length.

A. Complex function:

There are many complex operations in PAM to calculate the masking thresholds, such as arctangent, sine, cosine, logarithmic, power, square root, and division operations. In conventional approaches, formulas such as Taylor series are usually used to approximate the complex functions. However, this would take many computational cycles especially when high order Taylor series are required for high precision. Besides, the division operation also takes many computational cycles. In [5], it takes more than 24 cycles to calculate a division for a typical 24-bit fixed-point processor. Another common approach makes use of a look-up table that stores pre-computed values. Nevertheless, this will also cost large on-chip memory because many tables are required for those functions.

B. Limited word length:

In Threshold Generation, many data represent energies and masking thresholds. Because the word length is about 24 bits for a spectrum to meet the precision, 48-bit word length is required for the energy and masking threshold data. Long word length of these data is a challenge for a 16-bit fixed-point processor. Besides, large memory to store and transfer the data is required. If word length is decreased to reduce the memory usage, the precision would be lowered and the audio quality would degrade, too. This is a trade-off.

IV. Proposed Schemes

A. Only-LOG-based algorithm

In order to implement the complex function efficiently, we propose the only-LOG-based algorithm. The idea is to transform the entire complex functions into only logarithms and corresponding power functions. First, the original Fast Fourier Transform (FFT) (Step 2) is replaced with Modified Discrete Cosine Transform (MDCT) to reduce complexity; tonality indices (Steps 3-6) are calculated by Spectral Flatness Measure (SFM) instead of the original Unpredictability. This method is first proposed by [3] and then modified by [2]. Here, we follow the same algorithmic flow as [2]. In this way, we can employ logarithmic operations in SFM instead of the original complex operations in Unpredictability. Besides, [6] also mentions that the memory usage will reduce because the spectrums of the two previous frames are not required to calculate the Unpredictability. Secondly, as for the calculations of masking threshold (Steps 7-13), the original logarithmic and power operations are rescheduled. This method not only simplifies the original operations of power and multiplication but also converts the original division operations in perceptual entropy (PE) and inverse signal-to-mask ratio (ISMR) into subtractions. Table I summaries these methods. Therefore, only a table of logarithmic and corresponding power is required; both look-up tables and computational complexity are reduced.

B. 16-bit logarithmic representation of energy and masking threshold

The characteristic of the logarithmic representation is the mapping of the completely dynamic range. Besides, multiplications and divisions in the original domain correspond to the summations and subtractions in the logarithmic domain. Because the previous scheme transforms the original PAM into logarithm-based PAM, all the energies and masking thresholds are stored in logarithmic formats accordingly and naturally. Note that the energies *e(b)* in Step 5 are still stored in the original domain in order to convolve with the spreading function, but this overhead is low. In addition, we find the precision is much less important

TABLE I. THE ORIGINAL PAM VS. THE PROPOSED ONLY-LOG-BASED PAM

Algorithms	**Original PAM [1]**		**Proposed only-LOG-based PAM**	
	Equation	*Complex function*	*Equations*	*Complex function*
Step 1				
Step 2	FFT (spectrums from a+bi to magnitude and phase)	arctan	MDCT	
Step 3				
Step 4	Unpredictability c(w)	sin, cos, square root, division	SFM(b)	log10
Step 5			log_en(b)=log10(en(b))	log10
Step 6	tb(b)=-0.299-0.43loge(cb(b))	loge		
Step 7				
Step 8	bc(b)=10^(-SNR(b)/10)	power of 10	log_bc(b)=-SNR(b)/10	
Step 9	nb(b)=en(b)*bc(b)		log_nb(b)= log_en(b)+log_bc(b)	
Step 10				
Step 11	PE=sigma(width* log10(nb(b)/(e(b)+1))	log10, division	PE=sigma(width* (log_nb(b)-log_e(b)))	
Step 12				

Step 13	ISMR(n)=npart(n)/epart(n)	division	log_epart(n)=log10(epart(n));
			log_ISMR(n)= log_npart(n)-log_epart(n);
			ISMR(n)=10^(log_ISMR(n));
				power of 10, log10

if the dynamic range is maintained. Therefore, word length is analyzed to find the acceptable bits for precision (fractional bits). Table II displays simulation results of sound quality degradation when those values are stored as 16-bit logarithmic formats in different audio files. Quality degradation is in ODG [7] and NMR. Positive value represents that the sound quality of proposed encoder is worse. The results show that 16-bit fixed-point logarithmic format where 5 bits are for dynamic range is sufficient for the audio signal energy and masking threshold to maintain the quality. By this scheme, memory storage and bandwidth are reduced a lot.

V. PERFORMANCE

In order to quantify the performance of the proposed design, we model the computational complexity and memory storage and bandwidth. Because the computation of spreading function and its convolution is large and it will affect the comparison, the spreading function in all algorithms are assumed to optimize in the same method as [8] for fair comparison. Here, the psychoacoustic parameters in the model are assumed with CD-quality input (16 bits per sample at sampling rate 44100 Hz) for general use.

In the model of computational complexity, one multiplication or one addition is counted as one operation. The calculation of complex function is by approximated formulas or look-up tables. Here, second-order Taylor series for each special function are estimated for simplicity, and one division is assumed to take 24 operations (cycles). Table III displays the results of computational complexity. The unit of computation complexity is Millions of Operations Per Second (MOPS). In the first column (not the table head), the original algorithm with approximated formulas requires the most operations. The second column shows that complexity is dramatically reduced by look-up tables of large memory. The third column shows the computational complexity can be reduced a lot with our previous work [2]. This work, the last column, further reduces the required table with almost same complexity.

In the model of memory storage and bandwidth, only those in Threshold Generation are calculated because energy and masking threshold data are only in the block. The word length of those data in the original and [2] is 48 bits, whereas that in the proposed schemes is 16 bits (in logarithmic format). Table IV shows the results of memory requirement. It is obvious that the proposed schemes require the least memory for storage and bandwidth. From the two tables, therefore, the proposed design is the most efficient one in computational complexity and memory requirement.

TABLE II. QUALITY DEGRADATION BY 16-BIT LOGARITHMIC FORMAT OF ENERGY AND MASKING THRESHOLD

Audio files	*Quality degrade (ODG)*	*Quality degrade (NMR)*
castanets	-0.01	0.01
speech	-0.01	-0.01
velvet	0.01	0.00
frer07_1	-0.01	-0.01
trpt21_2	0.00	0.00
horn23_2	0.00	0.00
gspi35_2	-0.01	-0.04
quar48_1	0.01	0.00
spfe49_1	-0.02	0.01

VI. CONCLUSIONS

The paper presents a memory and computationally efficient PAM design, which is very important especially when an AAC encoder is implemented on a 16-bit fixed-point processor. The technique is to modify and reschedule the flow of PAM so that the original various complex operations are replaced with only logarithm and corresponding power operations. Therefore, it can not only reduce the computational complexity but also decrease required look-up tables. Besides, 16-bit logarithmic format is sufficient to represent the audio signal energy and masking threshold while maintaining almost the same audio quality. This scheme effectively reduces the required data storage and transfer in the calculation of masking threshold.

REFERENCES

[1] MPEG. Information technology – Coding of audio-visual objects – Part 3: Audio, International Standard IS 14496-3, ISO/IEC JTC1/SC29 WG11, 1999.

[2] S. Huang, T. Tsai, and L. Chen, "A low complexity design of psycho-acoustic model for MPEG-2/4 advanced audio coding," *IEEE Tran. on Consumer Electronics*, Vol. 50, No. 4, Nov. 2004, pp. 1209-1217.

[3] Y. Takamizawa, T. Nomura, M. Ikekawa, "High-quality and processor-efficient implementation of an MPEG-2 AAC encoder," in *Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 985 –988.

[4] M. Gayer, M. Lohwasser, M. Lutzky, "Implementing MPEG Advanced Audio Coding and Layer-3 encoders on 32-bit and 16-bit fixed-point processors," preprint 5978, *AES 115th Convention*, New York, Oct. 10-13, 2003.

[5] J. Hilpert, M. Braun, M. Lutzky, S.Geyersberger and R. Buchta", Implementing ISO/MPEG-2 Advanced Audio Coding in realtime on a fixed point DSP," preprint 4822, *AES 105th Convention*, San Francisco, California, USA, Sep. 26-29, 1998.

[6] E. Kurniawati, C. Lau, B. Premkumar, J. Absar, S. George, "New implementation techniques of an efficient MPEG Advanced Audio Coder," *IEEE Trans. on Consumer Electronics*, Vol. 50, No. 2, May 2004, pp. 655-665

[7] ITU-R Recommendation BS. 1387: "Method for objective measurements of perceived audio quality," July 2001

[8] S. Huang, T. Tsai, L. Chen, "Memory reduction technique of spreading function in MPEG AAC encoder," in *Proc. of the 7th Int. Conference on Digital Audio Effects*, Naples, Italy, October 5-8, 2004

TABLE III. COMPARISON OF COMPUTATIONAL COMPLEXITY AND REQUIRED LOOK-UP TABLES

Algorithm	Original [1]	Original [1]	MDCT-based [2]	Proposed
Calculation of complex function	Approximated formula (2nd order Taylor series)	Look-up table	Look-up table	Look-up table
Computational complexity (MOPS)	24.6	12.9	6.5	6.3
Required tables	None	Square root Arctangent Division Sine Cosine Power of 10 log10, ln	Power of 10 log10 Division	Power of 10 log10

TABLE IV. COMPARISON OF DATA MEMORY STORAGE AND BANDWIDTH IN THRESHOLD GENERATION

Algorithm	Original	MDCT-based [2]	[4]	Proposed
Word length of data for energy and masking threshold (bit)	48	48	32	16
Memory storage of intra-frame data (Byte)	17446	8554	N/A	3506
Memory storage of inter-frame data (Byte)	35340	7692	N/A	4612
Memory bandwidth (MB/s)	14.9	7.1	N/A	4.6