AVS-P3: Algorithm and Implementation

Zhang Tao, Dong Yuxi and Zhang Wen
School of Electronic and Information Engineering, Tianjin University, Tianjin, China
Email: zhangtao@tju.edu.cn

***Abstract*—The third part of the AVS family, AVS–P3 is an audio coding standard of China own intellectual property. It is drafted according to the requirement of the advanced audio compression technology with high quality and efficiency in the field of DAB, DVB, DSM, ISM and MMC. In this paper, main algorithm modules of AVS-P3 are analyzed in detail, including Transient Detection, MDCT, Multi-Resolution Analysis in Frequency Domain, Quantization, Post Quantization Square Polar Stereo Coding, and Context Based Coding. Next, AVS P-3 audio Codec is implemented on DSP platform. The demo system could capture original audio input, encode/transmit/decode with AVS-P3 format, and play back in real time. Finally, the result of the subjective test is presented, its quality is a little better than the quality of LAME MP3.**

***Index Terms*—Audio Coding, AVS, Codec, Algorithm**

I. Introduction

For a long time, China has no independent intellectual property rights of technical standard on digital audio. The digital audio and video industry development has been nagged by the problem of the huge amount of foreign patent fee. Therefore, establishing our own technical standard as soon as possible to respond positively admits of no delay.

On May 25, 2002, the preparatory meeting of the "Digital Audio and Video coding Standardization (AVS) Working Group"was held in the Chinese Academy of Sciences Institute of Computing Technology. The decision was made to set up the "AVS Working Group".

On June 11, 2002, Department of Science and Technology of the Ministry of Information Industry published the announcement of the establishment of "AVS Working Group" on the "China Electronics News", and welcomed the domestic relevant units of the industry, academia, research, and users of the digital audio, video to actively participate in, collecting the first group of members. On June 21, 2002, "AVS Working Group" was established in Beijing. The representatives of the first group of the members, such as Beijing Fuguo digital technology Co., Ltd., Huawei Technologies Co., Ltd., Tsinghua University, Microsoft Research Asia, Institute of Automation Chinese Academy of Science, attended the meeting. The meeting assigned Pan Xingde, Hu Ruimin, Chen Yuanzhi as the audio group conveners.

On August 23 and 24, 2002, the first meeting of the "AVS Working Group" was held in Beijing. The meeting decided to combine this working group with the MPEG-China, with the title of the two organizations retaining for domestic and international use respectively. During this meeting, the audio group received five proposals, and set the Group's annual plan basically.

In September, 2005, the 14th meeting of the AVS Working Group was held in Shanghai. When it comes to the audio, the Working Draft of AVS-P3 (two-channel audio) had entered into the verification stage. In December, 2005, the 15th meeting of the AVS Working Group was held in Qingdao. AVS-P3 was extended to support multi-channel, and Final Committee Draft document[1] was contributed, and the reference code was optimized. The Fuguo Company was responsible for optimizing codec for formal test by Academy of Broadcasting Planning before December 28.

On January 19, 2006, the Ministry of Information Industry formally applies for the AVS audio as the national standard. The AVS audio standard contained the patented technology of Fuguo, Samsung, and others. Initially, the AVS standard group also included Sony, Panasonic and other more than 30 foreign enterprises, they could contribute their patented technology into AVS Audio "patent pool". The principle of the AVS standard group was to contribute their own patented technology to national standard with very low fee. Many members were unwilling to abandon their own benefits, eventually they withdrew from the AVS Audio Working Group. The quality of the AVS Audio Coding could reach or better than the MPEG-2 AAC.

In 2008, the AVS-P3 audio standard applied for the national standard again. So far, the AVS audio standard has passed 90-day publicity period, waiting for accept by the national standard committee.

This paper is a review of AVS-P3 algorithm and implementation. In second section, main algorithm modules of AVS-P3 are analyzed. In the third section, the encoder and decoder are imported to a DSP platform, and optimization is fulfilled to achieve real-time encoding and decoding. Finally, the subjective test of the AVS working group shows that the result is impressive and the quality is a little better than LAME MP3.

II. AVS-P3 Algorithm Analysis

A. Introduction

AVS is drafted in order to adapt the requirement of the advanced audio compression technology with high quality and efficiency in the field of DAB, DVB, Digital Storage Media (DSM), Internet Streaming Media (ISM),

doi:10.4304/jsw.5.8.858-865

and Multi-Media Communication (MMC). The AVS standard[2] is mainly applied in the following fields, cable TV, DAB, DVB-Territorial, mutual storage multimedia, multimedia service of email sorting network, and communication in real time which comprises audio and video meeting and visible telephone.

AVS-P3 supports the mono-channel, dual-channel, and multi-channel PCM audio signal with the sampling rate from 8 to 96 kHz as the input signals, and the range of output bit rate is 16-96kbps per channel. Meanwhile, fine scalability of bit stream is supported. The step of the scalable bit rate is 1kbps when the encoding rate is below 79kbps/ch, while the step of the scalable bit rate is a little over 1kbps when the encoding rate is 79kbps-96kbps/ch. Transparent coding quality could be achieved at 64kbps/ch.

B. Structure of AVS-P3 Encoder and Decoder

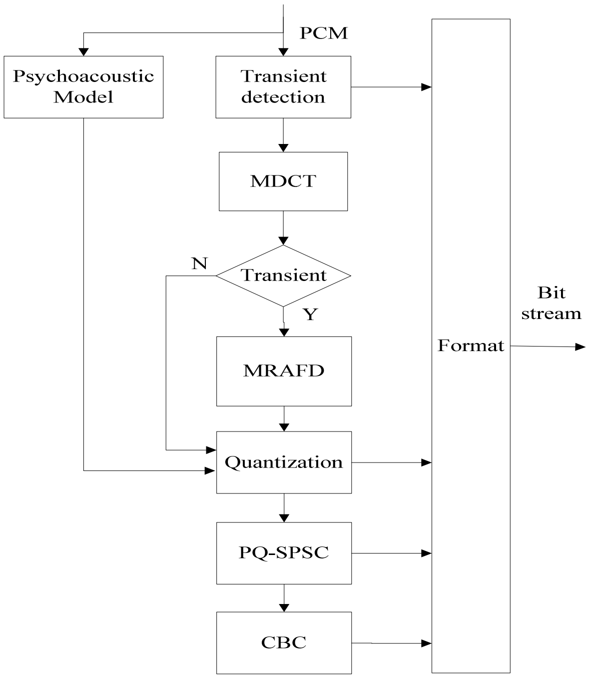

Figure 1. The diagram of AVS-P3 encoder

The diagram of AVS-P3 encoder is shown in Fig. 1. In the transient detection module, the input PCM data are classified to be a transient signal or a stationary signal. Then, the PCM data are transformed from time domain to frequency domain using 2048-sample Modified Discrete Cosine Transform (MDCT). According to the result of the transient detection, the Multi-Resolution Analysis in Frequency Domain (MRAFD) is employed (for transient signal) or not (for stationary signal). Next, the frequency coefficients are non-linear quantized. During the quantization process, the scale factor is determined by the Signal-to-Mask Ratio (SMR) based on the psycho-acoustic model. Finally, after the Post Quantization Square Polar Stereo Coding (PQ-SPSC) and the Context Based Coding (CBC), the bit stream is formatted. Wherein, the MRAFD module and the PQ-SPSC module are optional. The decoder diagram of AVS-P3 is just the verse, as shown in Fig.2 [3, 4].

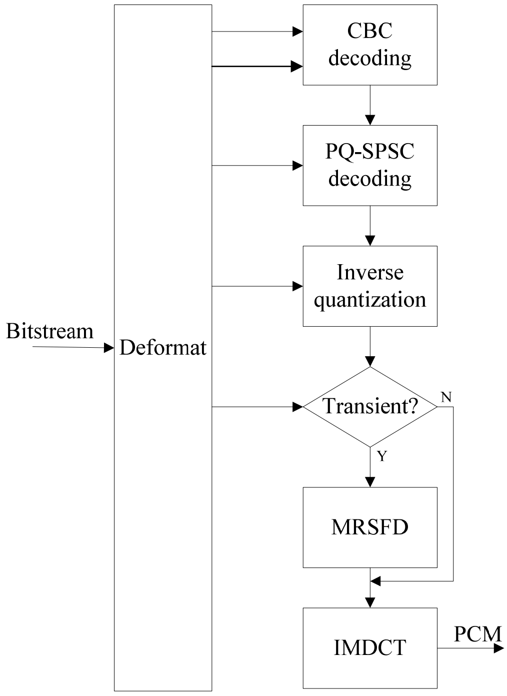

Figure 2. The diagram of AVS-P3 decoder

C. The main algorithms of AVS-P3

1. Transient dection

A dual-level judgment method based on the energy in time domain and unpredictability in frequency domain is recommended in AVS-P3 transient diction module. The procedure is shown in Fig. 3. In Fig.3, E_SWITCH and P_SWITCH are energy and unpredictability thresholds.

2. Transform from time to frequency domain

In the AVS-P3 standard, MDCT with Time Domain Alias Cancellation (TDAC) characteristic is employed to transform the signal from time domain to frequency domain in order to reduce the relevant of each component in the signal [5]. The energy of the signal could be concentrated on a few frequency coefficients, so the encoding efficiency is enhanced. The length of the transform is 2048 samples.

The formula of MDCT is:

$$F_j(m) = \sum_{i=0}^{2N-1} w(i) X_j(i) \cos\frac{(2i+1+N)(2m+1)\pi}{4N} \qquad (1)$$
$$m = 0, ..., N-1$$

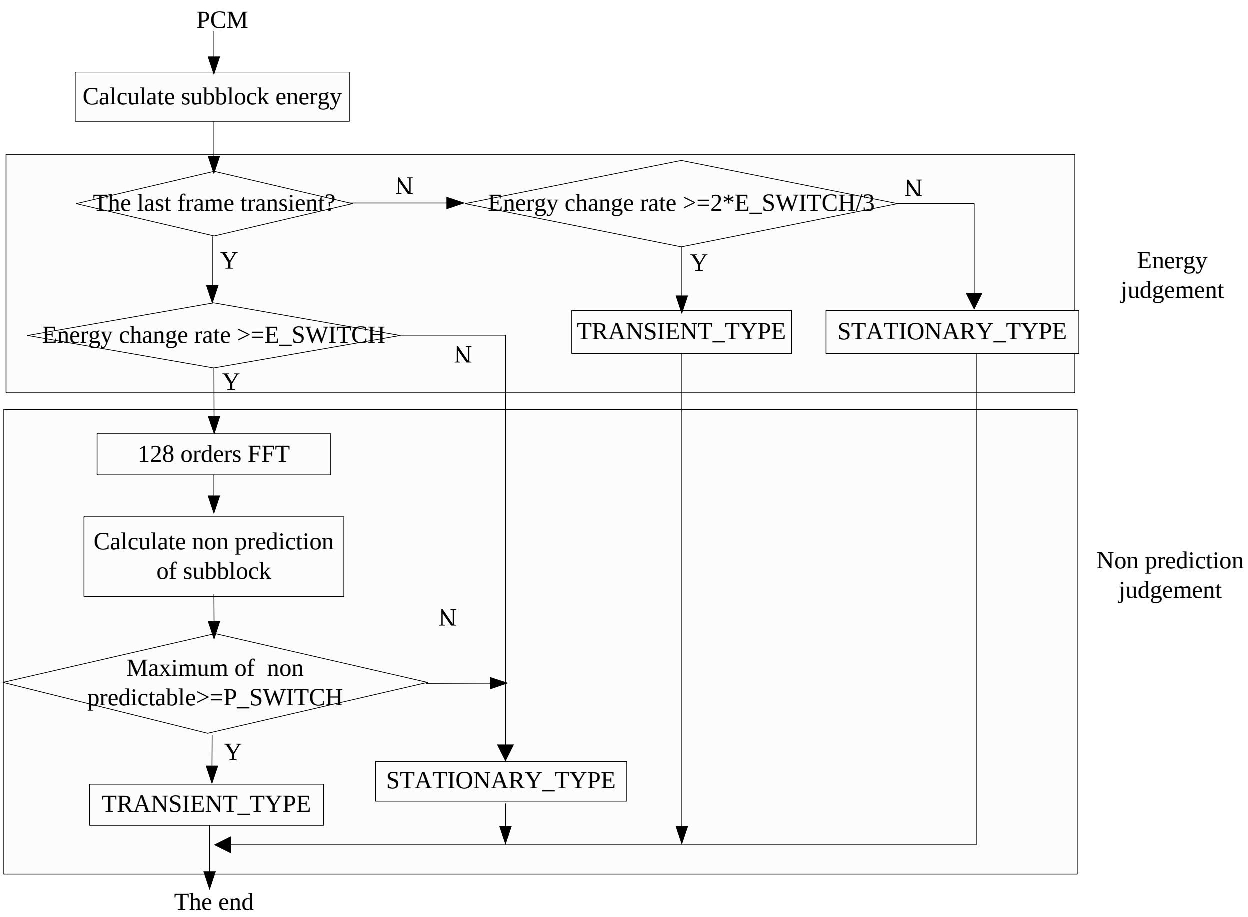

Figure 3. The diagram of the transient detection module in AVS-P3 encoder

A stationary signal waveform and its waveform in MDCT domain are shown in Fig. 4 and Fig. 5 respectively.

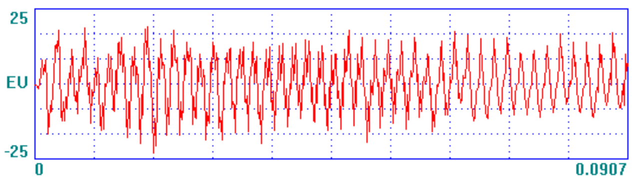

Figure 4. A stationary signal waveform in time domain

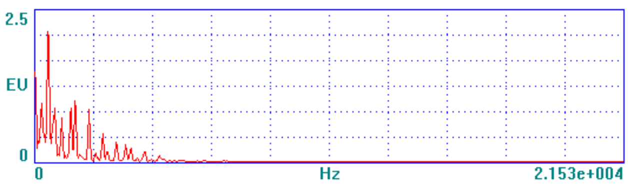

Figure 5. The waveform of the signal in MDCT domain

3. Multi-Resolution Analysis in Frequency Domain

According to the signal vary level in time domain, the signal can be divided into stationary signal and transient signal [6, 7]. As for the transient signal, after MDCT, the energy of the signal could not be well concentrated on a few frequency coefficients. In addition, the distribution of the frequency coefficients tends to be smooth so that the encoding efficiency is reduced and the pre-echo effect will appear when the bit rate is not enough. After the MRAFD, the resolution of the signal in time domain can be enhanced to make the energy concentrate on a few coefficients. A waveform of a transient signal frame after MDCT and MRAFD is shown in Fig. 6

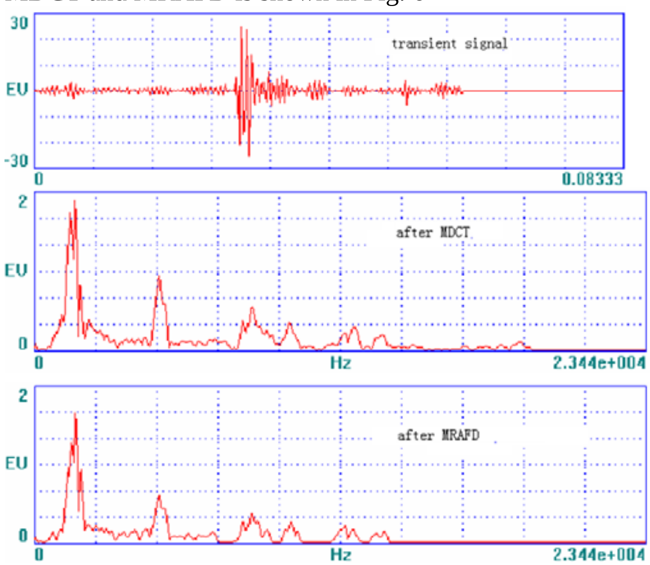

Figure 6. The MRAFD waveform of transient signal

4. Quantization

Non-linear quantization is used in Quantization module to the frequency coefficients. The formula is:

$$freq\_Q(k) = \text{int}((abs(freq(k)))^{3/4} * 2^{(3/16)*(sf)} + MAGIC\_NUMBER) \quad (2)$$

When the frequency coefficients are small, the small quantization step is used, and when the frequency coefficients are large, the large quantization step is used. The essence of quantization is to find the appropriate scale factor to make sure that the quantization noise could be masked by other components, so that the noise could not be heard by ears.

When the scale factor is determined, a global gain will be added on the original scale factor to do the normalization to all the scale factors so that the compression ratio could be enhanced further. Both the global gain and the scale factor could provide 1.5 dB ($20*\log 2^{1/4}$) quantization step.

5. Post Quantization-Square Polar Stereo Coding

PQ-SPSC module could further reduce the redundancy of the two channels to enhance the compression efficiency. PQ-SPSC takes the scale factor band as the minimum unit, and either the stereo coding is used to all the quantized frequency coefficients, or just part of quantized, or the stereo coding is not used. The formula of PQ-SPSC is:

$$\begin{pmatrix} m \\ a \end{pmatrix} = \begin{cases} \begin{pmatrix} l \\ \text{abs}(l-r) \end{pmatrix}, & \text{abs}(l) > \text{abs}(r) \\ \begin{pmatrix} r \\ -\text{abs}(l-r) \end{pmatrix}, & else \end{cases} \quad (3)$$

Wherein, l and r represent the data of left channel and right channel in frequency domain respectively, while m and a represent the corresponding amplitude and phase respectively, as illustrated in Fig. 7.

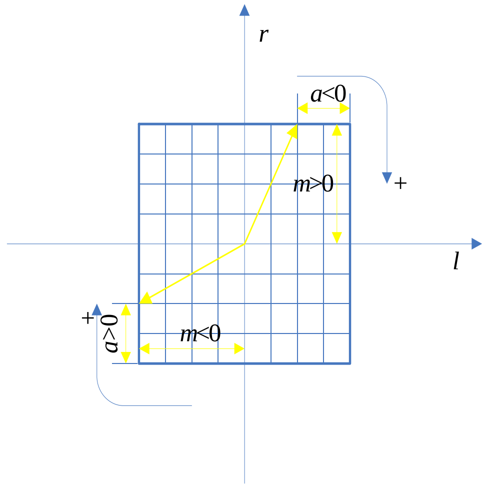

Figure 7. Diagram of the square polar coordinate transformation

6. Context Based Coding

CBC is used to do the entropy coding to the frequency coefficients after the quantization. CBC is an entropy coding technique with high efficiency and scalability[8].

The quantized frequency coefficients are the input signals of the CBC. According to different frequencies, the frequency coefficients can be divided into several frequency bands which are not overlapped with each other, and these bands are divided into coding layers. The maximum is 100 coding layers limited in AVS-P3. The information in each layer band is encoded respectively, and the information in the previous layer band is referred in the encoding of the current layer band. Thus, CBC encoder can be viewed as a combination of several sub-encoders which encode the information in current coder layer referring to the previous layer band, as shown in Fig. 8.

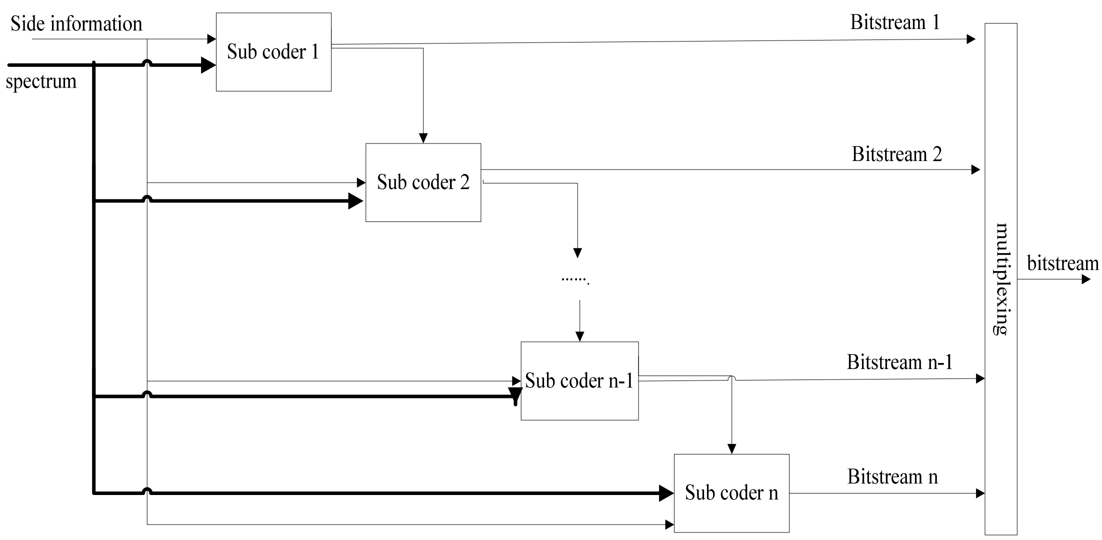

Figure 8. The diagram of CBC encoder

In each coder layer, the frequency coefficients are organized in bit layer to realize the scalability of the bit stream. The vector obtained by combining the bit layer is encoded by Huffman coding, so that the transmitting bit rate is reduced without distortion. As for the four frequency coefficients in each group, the method to combine bit layer is shown in Fig. 9.

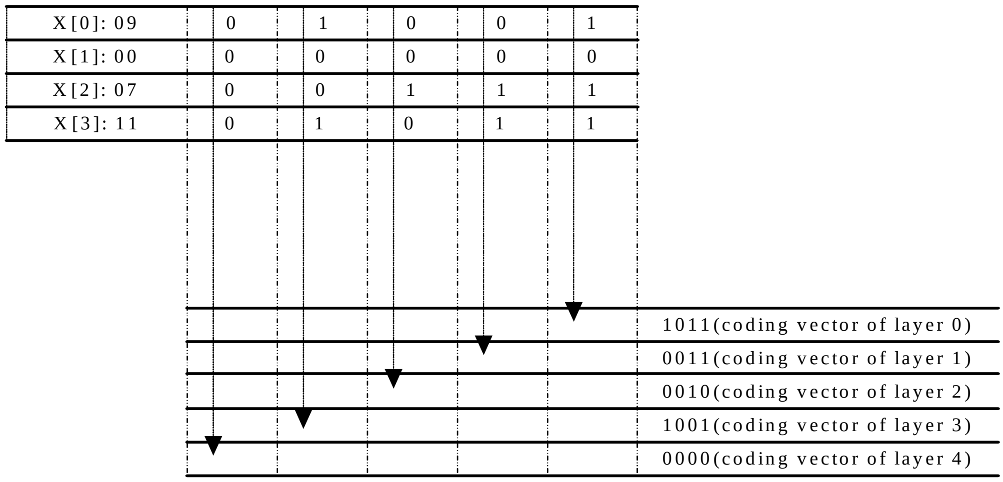

Figure 9. Bit layer combination method

The coder layer comprises basic layer and enhance layer. The frequency coefficients in the basic layer could guarantee the basic audio quality while the enhance layer could further improve the audio quality. The more layers there are, the better the audio quality is. The division of the coder layer is shown in Fig. 10.

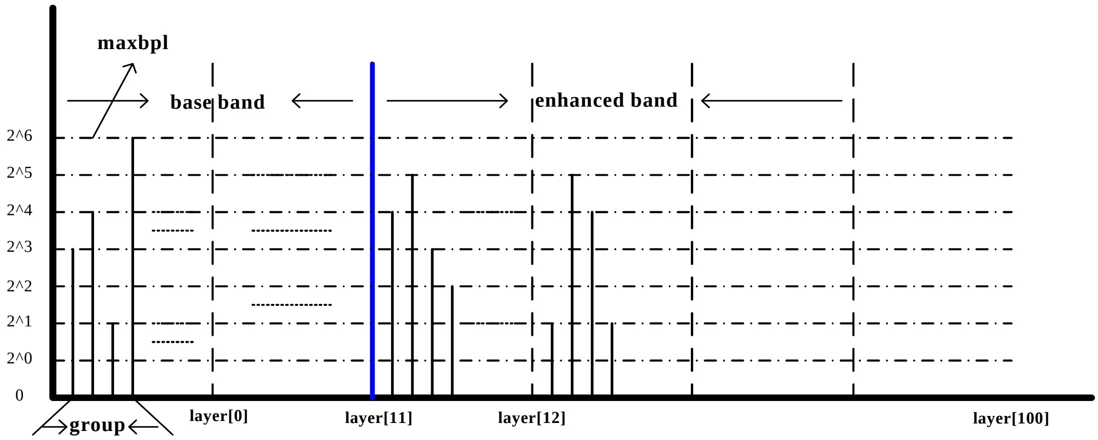

Figure 10. The coder layer division diagram

III. AVS-P3 CODEC IMPLEMENTATION ON DSP

A. System Design

AVS-P3 Codec system is shown in Fig. 11. The audio material is captured and quantized by the audio codec,TLV320AIC23B, and then are sent to the DM642 [9] which will encode and transmit the AVS-P3 bit stream in real-time. The DSK5416[10] receives the bit stream in real-time, decodes the bit stream, and plays back the decoded audio. This system supports 8 ~ 96 kHz sampling rate and 16 ~ 96 kbps/ch bit rate, and supports the mono and stereo encoding, and also supports LINE and MICPHONE level input and output with transparent audio quality.

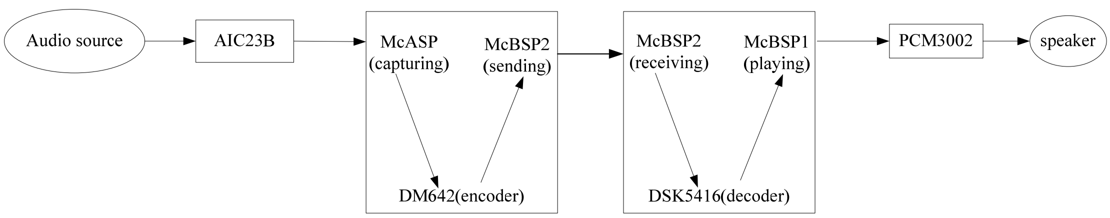

Figure 11. The AVS audio encoding and decoding system

B. AVS-P3 encoder Implementation on DM642

1. The algorithm optimization

(1) MDCT module

The N samples MDCT is calculated by the N/4 samples FFT. Here, the MDCT of 2048 samples is calculated by the FFT of 512 samples. The complexity comparison of MDCT module before and after optimization is shown in Table 1.

(2) MRAFD module

The optimization method in MDCT module could be shared by MRAFD because 128 16-sample MDCT is used in MRAFD. The complexity comparison of this module before and after optimization is shown in Table 2.

(3) Quantization module

The quantization module employs the non-linear quantization which has the exponent calculation, and it is time-consuming to realize on DSP platform. Thus, the exponent calculation is transformed to looking-up table.

(4) CBC module

The optimization is divided into two parts, codebook optimization and the method of looking-up table optimization.

Three kinds of codebooks are required in CBC module, side information codebook, scale factor codebook, and frequency coefficients codebook. Each codebook comprises two arrays, one is the code array and the other is the code-length array. In addition, all the elements in the arrays are presented by 32 bits. In fact, the information of code and code-length into the high position and low position of the 32 bit data respectively, so that the whole codebook size is just a half of the original one. Besides, many one dimensional codebooks could be duplicated in frequency coefficients codebook, codesize could be reduced significantly by defining a pointer array. The codebook size before and after optimization is shown in Table 3.

Frequency coefficient huffman codebook employs a method of three dimensional looking-up. In fact, on the same code layer, the codebook index of the first dimensional is the same, and the pointer array mentioned above is added. Therefore, the searching method of the codebook is simplified to one dimensional so Huffman coding efficiency is improved significantly.

TABLE I. THE COMPLEXITY COMPARISON OF MDCT MODULE

The comparison of two methods	Multiply times	Add times
MDCT without fast algorithm	2097152	2096128
MDCT with fast algorithm	11264	16896

TABLE II. THE COMPLEXITY COMPARISON OF MRAFD MODULE

The comparison of two methods	Multiply times	Add times
without fast algorithm	18432	15360
with fast algorithm	8192	6144

TABLE III. CBC CODEBOOK SIZE OPTIMIZATION RESULT

	Before optimization(kB)	After optimization(kB)
code table of side info	0.24	0.12
code table of scale factor	1	0.5
Code table of frequency coefficients	31.5	6.2

2. The optimization based on DM642

AVS-P3 encoder is optimized based on the properties of the platform DM642 [11, 12]. The optimization is as follows.

(1) The complex C program is replaced by intrinsics function;
(2) The data packing technique is employed;
(3) The loop body in the program is modified to realize software pipeline optimization;
(4) The optimization options of the compiler are opened.

After optimization, the cycle consuming of AVS-P3 encoder before and after optimization on the platform DM642 is shown in Table 4

TABLE IV. THE CYCLE CONSUMING COMPARISON BEFORE AND AFTER OPTIMIZATION ON DM642

Modules	Cycles (before optimization)	percent	Cycles (after optimization)	percent
Transient detection	31462	2.1%	2120	0.062%
MDCT	2643821	17.74%	529899	15.54%
MRAFD	3174084	21.3%	680882	19.97%
Quantization	4794662	32.18%	1316582	38.62%
Stereo coding	126478	8.5%	18437	5.41%
CBC	4128598	18.18%	861099	20.398%
Total	14899105	100%	3409019	100%

C. Implementation of AVS-P3 Decoder on DSK5416

Because the encoder and the decoder program of AVS-P3 are similar, similar optimization methods are used. The cycle consuming comparison of AVS –P3 decoder on DSK5416 before and after optimization is shown in Table 5.

TABLE V. THE CYCLE CONSUMING COMPARISON OF DECODER BEFORE AND AFTER OPTIMIZATION

Modules	Cycles (before optimization)	Percent	Cycle (after optimization)	Percent
IMDCT	2036891	37.0%	90562	17.54%
MRSFD	1118445	20.34%	71467	13.84%
Inverse Quantization	148530	2.7%	63628	13.32%
Stereo decoding	38654	0.7%	39223	7.60%
CBC decoding	2156322	39.26%	251478	47.7%
Total	5498842	100%	516358	100%

III. SUBJECTIVE TEST

In order to verify AVS-P3 performance, an internal subjective test is carried out by AVS working group. Fig. 12 and Fig.13 show the result of this subjective test [13].

According to the result, at 64kbps mono, AVS-P3 V.S. LAME MP3, the CMOS average score is 0.35. At 128kbps stereo, the CMOS average score is 0.13333. Thus, in the same coding condition, the quality of AVS-P3 is a little better than the quality of LAME MP3.

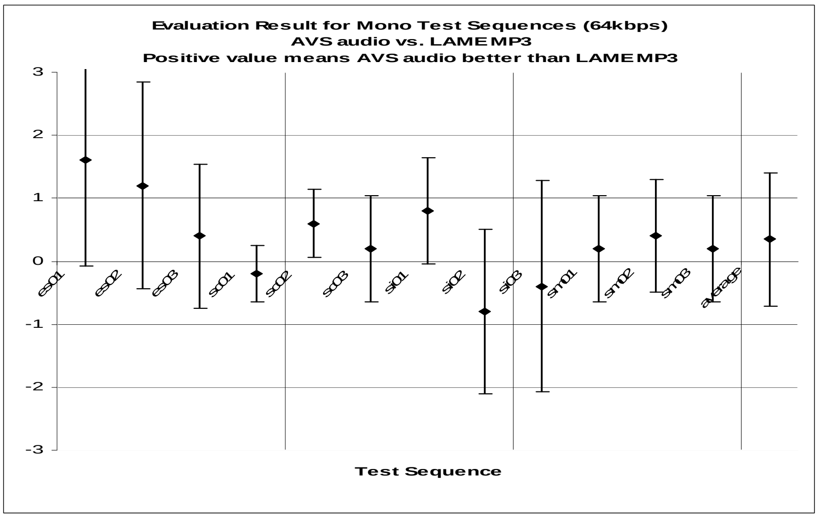

Figure 12. CMOS test result of 64kbps mono

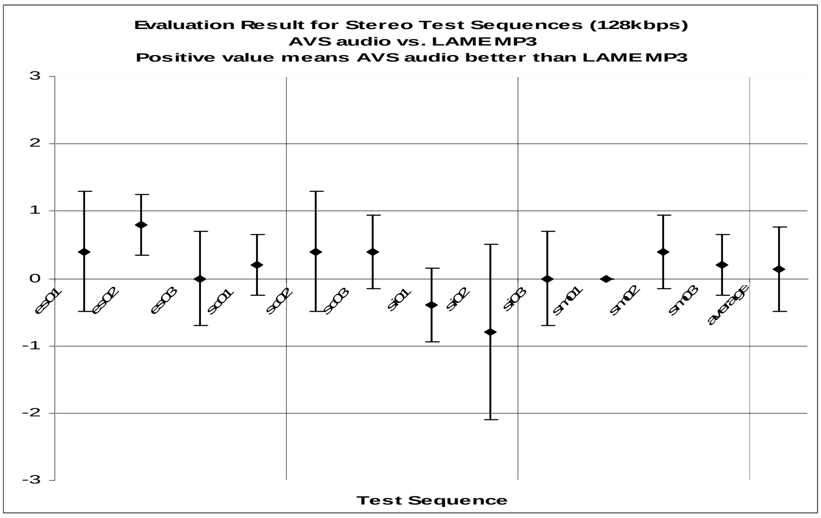

Figure 13. CMOS test of 128kbps stereo

IV. CONCLUSION

AVS-P3 is the third part of AVS which has China own intellectual property rights. It is a very good solution for high quality audio compression. It is mainly applied in high resolution digital broadcasting, high-density laser digital storage medium, wireless broad band multimedia communication and broad band internet streaming media application.

This paper introduces the development on AVS-P3 standard. A detailed analysis of the algorithms of AVS-P3 is provided. And then, a real-time demo system of AVS-P3 codec is designed based on DM642 EVM and DSK5416. Finally the subjective test conduced by AVS work group verified that AVS-P3 could meet the requirement of high quality audio coding.

ACKNOWLEDGMENT

Wang Wei, He Liqing, Kong Xianglan, Wang Li and He Jialin have also made contributions to this project. The authors wish to thank all of them. This work was supported by Electronic Information Industry Fund, MIIT, and the Tianjin University & Texas Instruments DSP Joint Lab.

REFERENCES

[1] Information Technology—Advanced Audio and Video coding Standard, Part 3 Audio.

[2] RuiMin Hu, Yong Zhang, Haojun Ai, “Digital audio compression technology and AVS audio standard research”, Intelligent Signal Processing and Communication Systems, 2005. ISPACS 2005. Proceedings of 2005 International Symposium on 13-16 Dec. 2005.

[3] Peter Paint, “Perceptual Coding of Digital Audio”, PROCEEDINGS OF THE IEEE, VOL. 88, NO. 4, P451-513, APRIL 2000.

[4] Noll, P., “MPEG digital audio coding”, Signal Processing Magazine, IEEE Volume 14,Issue 5,Sept.1997.

[5] A.M. Karshenas, M.W. Dunnigan, B.W. Williams, ” A MODIFIED STRUCTURE FOR MULTI-RESOLUTION ANALYSIS OF FREQUENCY DOMAIN SELF-TUNING RANDOM VIBRATION CONTROL”, UKACC International Conference on CONTROL '96,2-5 September 1996, Conference Publication No. 427 0 IEE 1996

[6] Tao Zhang, Wei Wang, Qiang Yao, “On the Pre-echo Control Method in Transient Signal Coding of AVS Audio”，International Conference on Audio, Language and Image Processing，2008.

[7] Boyer, R.; Essid, S., “Transient modeling with a frequency-transform subspace algorithm and ‘transient + sinusoidal’ scheme”, Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on Volume 2, 1-3, Page(s):865 - 868 vol.2, July 2002

[8] W Li, “Fine granularity scalability in MPEG-4 for streaming video[C]”, InISCS 2000, Geneva, Switzerland, June 2005

[9] Texas Instruments, TMS320DM642 Datasheet, 2006-07

[10] Texas Instruments, TMS320VC5416 Fixed-Point Digital Signal Processor, 1999-03

[11] Texas Instruments, C6000 Optimizing Compiler User's Guide, 2004-05

[12] Texas Instruments, CCS v3.0 Getting Started Guide, 2004-05

[13] AVS_M1710, The new revise of the AVS audio test report, Dec 2005.

Zhang Tao was born in Heilongjiang Province, China in 1975.4. He obtained his dual B.S. in Electronic Engineering major and Management Engineering at Tianjin University in 1998, and his M.S., and Ph’D in signal and information processing from Electronic and Information Engineering School at Tianjin University in 2001 and 2004.

He has worked in research and application of audio and video processing at Tianjin University since 2003. He was charged with many projects on audio/video algorithm research and applications, include projects cooperation with Huawei company, Telegene Semi., and projects funded by Guangdong Science and Technology Department and Ministry of Industry and Information Technology, China. Since 2003 He has been the director of Tianjin University and Texas Instrument DSP Joint Lab. Now he is an associated professor in Tianjin University. His current research interests include signal processing, digital audio processing and DSP system application.

He is a member of Audio Engineering Society, and a senior member of Chinese Institute of Electronics. And he serves as the vice-chairman of AVS audio group.

Dong Yuxi was born in Shenyang Province, China in 1986. He received his B.S. degree in Electronic Engineering major at Tianjin University in 2009. Now he is reading for his master degree in signal and information processing at Tianjin University. His research interests include audio processing and DSP system application.

Zhang Wen was born in Jiangsu Province, China in 1987. She received her B.S. degree in Electronic Engineering major at Tianjin University in 2009. Now she is reading for her master degree in signal and information processing at Tianjin University. Her research interests include signal processing and audio coding.