A Performance Assessment of Real-time Multichannel Audio Codec

*Sunghan Kim, *Daeyoung Jang, and *Jinwoo Hong

Abstract

In this paper, we describe a real-time implementation of a multi-channel audio codec system that is based on the MPEG-1 audio algorithm. The major feature of this system is that it has a flexible multi-DSP system that can be adapted for various applications with using up to four TMS320C40 DSPs. The purpose of this paper is to present the problems of the system and is to describe the optimized methods to solve the problems in the view of hardware and software. Our audio codec is composed of an encoder an a decoder system and the bit rate of bitstream is up to 384 kbps. Fast input/output interfaces, DSP overloads, and inter-DSP communications methods with high speed are considered in multi-DSP H/W. Also, to run real-time in S/W, optimizing methods of algorithm are considered. After implementation of system, the subjective assessment method, and 'triple stimulus/hidden reference/double blind' that recommended by ITU-R TG10/3 is adopted for the quality of our system. All test items except one are awarded difference grades(diffgrade) better than −1. From the results, multi-channel audio system can be used for HDTV service.

Ⅰ. Introduction

Multimedia system, digital TV and HDTV provide us new information services with high quality, recently. According to the demands of big screen size and high quality in video, audio system also requires high resolution. To give audio services with high quality, data transfer rate of 768 kbps(48kHz * 16bits) is needed per channel. In audio service of HDTV, it requires 3.84Mbps in data transfer rate so that it needs to apply compression methods of MPEG-1(ISO/IEC 11172-3) to transmit data [1].

Recently, based on DSP technology, the complicated audio signal algorithm is possible to implement for the operation of real-world service. Various DSP architectures are, therefore, considered for use in real-time signal processing applications. For example, parallel processing architecture, pipelined processing architecture, and array processor methods are considered for the selection of multi-processor DSP architecture [2]. Most of applications in multi-DSP require a fast operation of input/output interfaces, DSP overloads as little as possible, and high-speed inter-DSP communications. Especially, the system which require variable configurations is necessary to design to make it possible to change the structure into another application according to user requirements.

The system architecture with multi-DSP, therefore, must be determined that which hardware configuration and software model will be most efficient for the user's application to run real-time operation. The most hardware configurations are the topology of interconnection among multi-DSP, memory locality, input/output processing performance, and synchronization. To design a audio codec system for HDTV service, it is necessary to the processing time into how much processing power is necessary to archive the software purposes. It is necessary to consider the best structure of H/W and S/W to run in realtime mode on target hardware board if we use the non real-time algorithm. So, our research concentrated on what is a flexible hardware system and how to optimize the non-real time MPEG-1 algorithm to improve the performance of overall system.

The multi-channel audio codec system adopts the MPEG-1 layer Ⅱ compression method. The system receives analog data from 6 channels and digital data from 4 channels. Data transmission rate is up to 384 kbps. The multi-channel configuration is composed of the two independent stereo units. The audio codec system is implemented using a flexible multi-DSP system so that it also can be utilized to various applications as well as HDTV.

In next section, we will describe the concepts and techniques that we have experience in designing and implementing the MPEG multi-channel audio system for HDTV.

Ⅱ. An Implementation of multi-channel audio codec

In this section, major features of multi-channel audio codec

*Realistic Telecommunications Section, ETRI

Manuscript Received : November 1, 1997.

system are described from the viewpoint of hardware configuration and software functions. And the basic concepts of software and hardware design are considered. In hardware, a flexible multi-DSP architecture is considered to optimize DSP configuration. In software, each method is explained that how to optimize the function to run real time among the three DSPs.

2.1 Design of Multi-DSP System

The audio compression algorithm which applied to multichannel audio codec requires a lot of computing time for the complicated calculations and the iterative processes, and then it is very difficult to operate the software-based algorithm on real-time, especially for the encoder of MPEG audio codec. In case of 48kHz sampling rate and 384kbit rate per sec, the whole processing time for one frame of 1152 samples of PCM audio data has to be less than 24ms [3]. Hence one has to consider a special hardware configuration and software design scheme because of this constraint. The divided parallel software structure is adopted for the codec in addition to hardware development with a parallel or a pipelined multiple DSPs architecture.

In case of one approach that consists of the maximum pipelined DSPs for data processing of one frame, the digital data of one frame are buffered into input FIFO which can store digital data of 2 channels. The input audio data read by controlling the input FIFO are passed through DSP #1, DSP #2, DSP #3, and DSP #4 according to processing procedures, and then encoded audio bitstream with a formatted structure goes into the output FIFO for sending to the decoder. By using this architecture in case of 48kHz sampling frequency and 16 bits coding, the available total execution time for one frame from input audio to output bit stream permitted for the real-time operation can be extended up to 96msec in spite of some delay.

The divided time chart and the processing schedule for each DSP in encoder, organized by using the pipelined DSPs structure, are applied. The overall processing software of one frame is divided into 4 steps which are assigned to each DSP, i.e., step-1 for DSP #1, step-2 for DSP #2, and so on. Each DSP must, that is, finish the assigned processing itself for incoming frame within 24ms, and transfer the rest processing to the nex DSP. This technique can extend the limited time of 24ms to 96ms with some delays.

Using another approach of parallel DSPs, the audio data of one frame stored in DSP #1 through the input FIFO are transferred to DSP #2, DSP #3 and DSP #4 simultaneously. Also, DSP #2 and DSP #3 transfer their processing results to DSP #1 and DSP #4, and then DSP #4 to DSP #1. Finally, encoded audio bit stream with a formatted structure goes into the output FIFO for sending to the decoder by DSP #1. By using this architecture in case of 48kHz sampling frequency and 16 bits conding, execution time for one frame from input audio data to output bit stream for the real-time operation can be processed within the total execution time of 24msec. One of these configurations and the number of DSP can be made by user's choice according to the characteristics of application area.

The TMS320C40 DSP, which is a main controller of target system, has the main features and many benefits, and the configuration of our target system supporting the software naturally affects its execution efficiency. In particular the I/O configuration and the use of available internal memory and the on-chip instruction cache all contribute significantly. Because of the computational complexity of the multi-channel audio codec, unnecessary processor overheads have to be kepts as little as possible for the real-time operation.

Most systems use TMS320C40 interrupts to handle the transmission of bit stream and sample I/O. The handling of the interrupts increases the overloading of processor, not only due to the execution time of the interrup code, but also by disrupting the on-chip instruction cache. These effect the performance of code loops that could otherwise execute without memory accesses for instruction fetches. This difficulty can be avoided by placing handlers for frequently occurring interrupts in internal memory. This minimizes their execution time and also prevents them from disturbing the instruction cache.

Using the memory boot loader as the other technique is very efficient to save the access time of executable code. Generally, RAM with zero wait-state has faster access time than ROM. This method is achieved by loading the executable code within ROM to RAM and by executing the executable code within RAM.

DSP hardware has been developed for the applications that required for real-time processing. Also, various multiprocessor DSP architecture, for example, parallel processor structure, pipelined processor structure and array processor structure are the representative architectures for the complex computation.

A block diagram of the hardware design considering real-time operation, is shown in Figure 1. This design scheme is termed here a parallel or a pipelined DSPs architecture which consists of a parallel or a serial con-

nection between DSPs[4]. In DSP system, the transfer methods of data and control signal are important factors. These are under the influence of the functional and data partitioning which are allocated among the DSP. Generally, in case of parallel processing, distributed memory or shared memory or shared memory can be used for data communications. TMS320C40 DSP has the exclusive communication ports and this provides for simple circuit design. But, there are some problems of the 5M word/sec restriction of bi-directional transfer speed and this bring about much transfer time. Also, there is shared memory for transferring data among DSPs. This system has input/output FIFOs for the efficient I/O process.

Fig 2 shows the transfer delay in Multi-channel audio

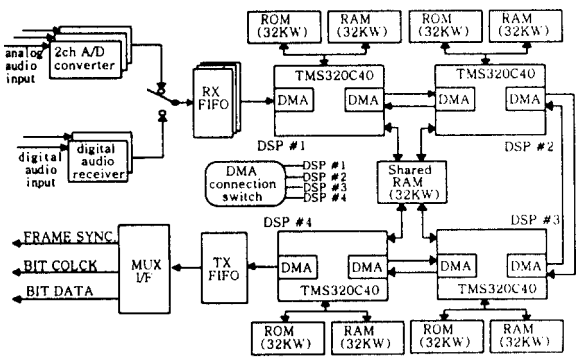
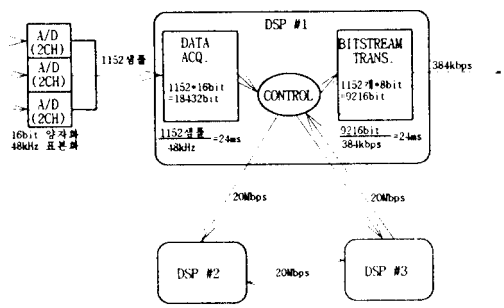


Figure 1. An example of a parallel or a pipelined DSP architecture(in case of encoder).



(Reference)
1) Data transfer time from DSP #1 to DSP #2
(1152 sample * 3 block + 1word) * 32bit = 110,624bit / 20Mbps = 5.5ms
2) Data transfer time from DSP #1 to DSP #3
(1152 sample * 3 block + 1 word) * 32bit = 110,624bit / 20Mbps = 5.5ms
3) Data transfer time from DSP #2 to DSP #3
(32sb * 7ch) * 32bit = 7,168bit / 20Mbps = 0.36ms
4) Data transfer time from DSP #1 to DSP #2
(32sb * 7ch) * 32bit = 7,168 / 20Mbps = 0.36ms
5) Data transfer time from DSP #3 to DSP #1
1152 * 32bit = 36,864bit / 20Mbps = 1.84ms

Figure 2. The processing delay time in Multi-channel audio encoder H/W

encoder. In encoder, DSP #1 transfer to DSP #2 and DSP #3, DSP #2 transfer to DSP #3, DSP #3 transfer to DSP #2 and DSP #1. The transfer methods are performed by DMA(Direct Memory Access) in all DSPs. If the synchronization problems of data transfer happened, the delay processing time can be maximum to 1.01ms(DSP #1 → DSP #2/DSP #3 : 0.69ms, DSP #2 → DSP #3 : 0.045ms, DSP #3 → DSP #2 : 0.045ms, DSP #3 → DSP #1 : 0.23ms)

2.2 An Implementation of Algorithm

MPEG audio algorithm uses the psycho-acoustic modelling, which is called perceptual coding, to compress audio signals. The basic idea behind perceptual coding of high quality digital audio coding signal is to hide the quantization noise below the signal-dependent thresholds of hearing. The psycho-acoustic model calculates the minimum masked threshold that is necessary to determine the just noticeable noise level for each band in the filter bank. A calculation of the actual masked threshold is necessary for a perceptual coding system. The algorithm contains a various a various mathematical functions with computational complexity. The properties of like these have an influence on the consideration of using DSP. At first, we implement the psycho-acoustic model with high level language for many mathematical functions and many looping processes in modules. High level language does not fully utilize the performance of DSP structure. FFT routing and subband filtering include very many iteration loops and bit manipulation operations. Therefore, these operations consumed timing in C language. So we replaced FFT routine and subband filtering module with assembly routine. And we use matrixing operation which provided in DSP architecture[9]. In general, processing time for mathematical functions like sine or cosine is consuming in time. This problem can be solved by using proper look-up table. And the size of table is concerned with resolution of results, it is important to decide proper size of table. In case of looping processes in modules, we analyzed whether module contains parallelism or not. Also, it is necessary to analyze inter-connection among each module. To minimize the processing time in modules, we considered proper processing scheduling between modules.

This sequential main program divided into 2 subblock for parallel processing, which is one for psycho-modeling module and the other is encoding module except psycho-modeling. Psycho-modeling module is performed in DSP #2 and the other module in DSP #3, respectively. Here, the number of ( ) is number of instruction cycle and period of one instruction cycle is 50 nsec.

Ⅲ. Performance Analysis of Multichannel Audio Codec

3.1 Execution Time

Finally, as an example, we applied the methods for optimization into psycho-modeling module in encoder program. Calculation of sound pressure level, extraction of tonal components, calculation of noise signal and masking threshold are considered. We described the methods to reduce the processing time in psycho-modeling below. Audio data performs by frame unit that is composed of 1152 samples. In psycho-modeling, it uses 1024 samples except the first and the last 64 samples. And it has delays of 192 samples to synchronize with subband filtering. So, we have sufficient 192 buffer size and define the first locations of buffer with starting address. These make it possible not to calculate the address. So, the time to calculate address is saved. Also, the bit reverse operation of data performs during data transfer by DMA, preliminary. FFT(Fast Fourier Transform) needs bit reverse operation. We use data of real value only, therefore, we do not need to calculate the imaginary data. And we use bit reverse operation that is provided in DSP. So, it is possible to reduce the time compared with FFT with complex data and bit-reverse operation. Here, it is necessary of window function to make it reverse bit address in advance. A calculation of SPL(Sound Pressure Level) calculates 512 energy with real and imaginary data which is gained from FFT and transforms into log based values. From 512 values, it calculates 27 energy value. So we can reduce the amount to use log function from 512 to 27. Extraction of tonal signal is to find the signal above 7 dB compared with neighbored signals. It extracts the useless period signals by using critical bands and absolute threshold values. We found important period signals in advance. It is necessary to make the table of absolute threshold values in case of getting rid of useless data. Noise calculation performs for critical bands and sums all of the energy except period signals. And then it calculates the energy and find the location of noise signal. We use the output of FFT. Output of FFT is absolute energy so that it can be possible to calculate by adding all the critical band values. This method has small computational complexity. Like calculation of tonal signal, it extracts useless noise signal in advance[9]. Masking threshold applies tonal signal and noise signal with masking functions. It is necessary to stop calculating if masker in frequency is bigger than minimum masking threshold in subbands. While we are seeking for minimum masking threshold for each subbands, we find the minimum value in advance, which is the masking threshold value in each subbands. Here, method for finding minimum masking threshold is simplified.

In Table 1, we compared the computation time of modules before and after using above modification. The results depict for the execution cycle and processing time.

Table 1. Calculation results after using optimization method

	Before Cycle Number/ Time(ms)	After Cycle Number/ Time(ms)	Results
Data input	197291/9.86	124251/6.21	37.0%
FFT, SPL Calculation	502164/25.11	90832/4.54	81.9%
Tonal labeling Noise labeling	43398/2.17 296223/14.81	21574/1.08 17201/0.86	88.6%
Threshold calculation SMR calculation	1316244/65.81 4823/0.24	340629/17.03	74.2%
Total	2486265/134.31	715761/35.79	71.2%

3.2 Quality Evaluation

After the implementation of our multi-channel audio codec, preliminary hearing test is performed in a listening room to assess the coded audio quality. Usually, the subjects who participate in the test we required to mark the perference by evaluating the perceptual difference only between the two codecs. To assess the quality of multichannel audio codec, the proper experiment design is necessary to obtain a reliable result. Test materials are made by using the reference signals and the reproduction signals. But, it assumes that subjects do not know the combination of test sequence. The subject listens a test sequence and put on a 5 kind of relative scores for the two signals. After that, the results come out a difference grade between the score of reproduced signal and that of reference signal. 10 subjects are participated in the assessment test, and they take a training course for the familiarity with the test items[7][8].

The subjective quality for the implemented audio codec were, in this paper, considered as evaluation items of multi-channel audio codec. The test materials for the subjective quality assessment of multi-channel audio codec are shown in Table 2.

Subjective quality test was implemented in the listening room which is designed according to the specifications of ITU-R TG 10-3 and EBU experts group GI/LIST. The listening room which satisfy the reference listening condition of MPEG audio group and has the characteristics

of reverberation time of 0.2sec, room area of 4.7m×6m×3m, initial reflection sound of −14dB, and background noise of 40dBA[5][6].

Table 2. Test materials for the subjective assessment.

Item	Contents	periods(msec)
1	Cinema music(Indiana Johnse)	13
2	Movie effect sound(Night Bird)	12
3	Movie effect sound(Ski Jump)	13
4	Commentary of tennis game	15
5	Male speech of English	8

An experimental environment is shown in figure 3. Original data and reproduced data are prepared for the experiment. The reproduced signal is recorded in DAT from the multi-channel audio codec system. And then, the data are stored in a PC and the sound level is edited by using a multi-channel editor.

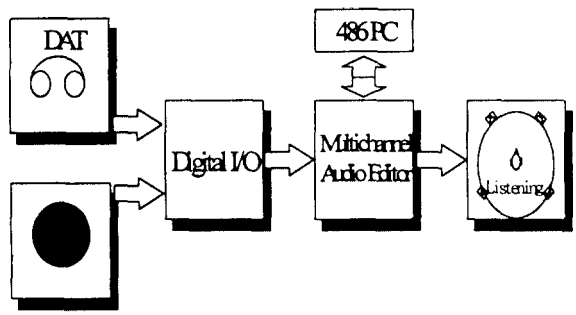


Figure 3. The experimental environment

Subjective test method used in this paper is "double blind triple stimulus with hidden reference" and continuous 5-grade impairment scale with corresponding differential score between original signal and coded signal. The differential grade by test result is calculated by the following equation.

Diff-Grade = Score of hidden reference signal-Score of coded test signal

Figure 4 shows the average differential grade between original signal and coded signal of multi-channel audio signal with 5 channels, 48kHz sampling rate, and 16bit coding. The average differential grade falls below −1 in case of 320kbit/s rate for 5 test materials and is −3 with annoying in 128kbit/s rate. The former is a compression rate of 12:1 and the latter 30:1.

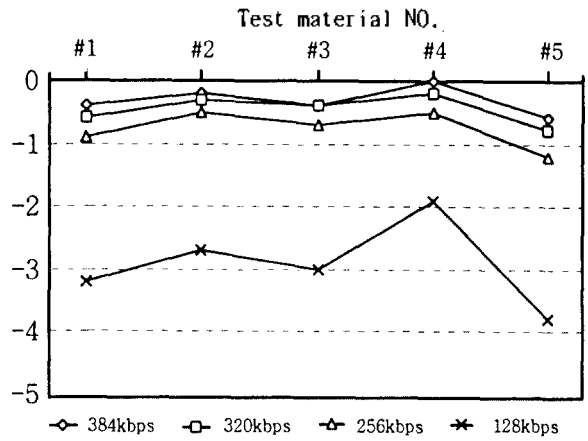


Figure 4. Diff-grade of test materials by bit rate.

Ⅳ. Conclusion

This paper described major features of multi-channel audio codec system. And it considered two aspects of software and hardware design. In hardware, a flexible multi-DSP architecture to optimize DSP configuration is researched and to verify a flexible configuration, multi-channel audio codec was developed, and then the evaluation result was described. To run real-time operation with 3 DSPs of TMS32C40 50MHz, the architecture can be configure with a pipelined and a parallel methods. In software, to implement the non-real time software, we adopt parallelism of algorithm, look-up table of mathematics functions and bit reverse operations that provided in DSP. And the results which have applied for MPEG-1 layer Ⅱ are shown in table of contents.

Finally, the subjective assessment is performed by "triple stimulus/double blind" method. In the subjective quality test, 10 subjects are participated in the test. The results show that the average differential grade between original signal and coded signal of multi-channel audio signal, 48kHz sampling rate, and 16bit coding falls below −1 in case of 320kbit/s rate for 5 test materials and is −3 with annoying in 128kbit/s rate. It means that the coded signal is hard to perceive the difference from the original one. From these results, we confirm that the quality of our codec is acceptable for broadcasting systems.

Acknowledgment

This work is supported by the Korea Telecom under the title of "A Development of HDTV transmission Technology."

References

1. D. Douglas Wilmarth, "A Hierarchical Parallel Processor for Digital Signal Processing Application," *Proceedings of ICSPAT '93*, pp. 1114-1121, Sep. 1993.
2. M. Platzner, etc., "Performance Measurements on a Multi-DSP Architecture with TMS320C40," *Proceedings of ICSPAT '93*, pp. 1144-1152, Sep. 1993.
3. S. Shlien, "Guide to MPEG-1 Audio Standard," *IEEE Transactions on Broadcasting*, Vol. 40, No. 4, pp. 206-218, Dec. 1994.
4. Texas Instruments, *TMS320C4x User's Guide : Compiler, Assembler, Debug Tool*, 1991.
5. ISO/IEC 11172-3, *Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s-Part 3 : Audio*, 1993.
6. ISO/IEC 13818-3, *Coding of moving pictures and associated audio : Audio*, 1994.
7. G. Stoll, G. Theile, S. Nielsen, "Extension of ISO/MPEG-Audio Layer II to Multi-Channel Coding : The Future Standard for Broadcasting, Telecommunication, and Multimedia Applications," *The 94th AES Convention*, Berlin, Preprint No. 3550, Mar. 16-19, 1993.
8. K. Brandenburg and G. Stoll, "The ISO/MPEG-Audio CODEC : A Generic Standard for Coding of High Quality Digital Audio," *The 92nd AES Convention*, Vienna, Preprint No. 2970, Mar. 24-27, 1992.
9. D. Wiese and G. Stoll, "Bit rate Reduction of High Quality Audio Signals by Modeling the Ears Masking Thresholds," *The 89th AES Convention*, LA, Preprint No. 2970, Sep. 21-25, 1990.

▲Sunghan Kim

Sunghan Kim received the B.S degree in computer engineering from Kwangwoon University in 1989, the M.S degree in 1991, respectively. Since 1991, Mr. Kim has been with Realistic Telecommunication Section of Human Interface Department of ETRI. His recent research is on MPEG-2 Audio software implementation for HDTV and have many interest in applications of digital signal processing like sound reproduction techniques for realistic telecommunication, MPEG-4 audio and MSDL. He is a member of the Acoustical Society of Korea.

▲Daeyoung Jang

Daeyoung Jang was born in Jumunjin, Gwangwon-do. He received the B.S degree in electronic engineering from Pukyung University in 1991. Since 1991, Mr. Jang has been with Realistic Telecommunication Section of Human Interface Department of ETRI. His recent research is on digital signal processing, MPEG-1 Audio S/W and H/W implementation. He is a member of the Acoustical Society of Korea.

▲Jinwoo Hong

Jinwoo Hong received the B.S degree in application electronic engineering in 1982, the M.S degree in 1984 and Ph.D. degrees in 1993 from computer engineering of Kwangwoon University, respectively. Since 1984, Dr. Hong has been with Realistic Telecommunication Section of Human Interface Department of ETRI. His recent research is on ISDN telephone, quality assessment of communication and MPEG audio. He is a member of the Acoustical Society of Korea.