The Implementation of Multi-Channel Audio Codec for Real-Time Operation

실시간 처리를 위한 멀티채널 오디오 코덱의 구현

Jin Woo Hong*
홍 진 우*

ABSTRACT

This paper describes the implementation of a multi-channel audio codec for HDTV. This codec has the features of the 3/2-stereo plus low frequency enhancement, downward compatibility with the smaller number of channels, backward compatibility with the existing 2/0-stereo system(MPEG-1 audio), and multilingual capability. The encoder of this codec consists of 6-channel analog audio input part with the sampling rate of 48 kHz, 4-channel digital audio input part and three TMS320C40 DSPs. The encoder implements multi-channel audio compression using a human perceptual psychoacoustic model, and has the bit rate reduction to 384 kbit/s without impairment of subjective quality. The decoder consists of 6-channel analog audio output part, 4-channel digital audio output part, and two TMS320C40 DSPs for a decoding procedure. The decoder analyzes the bit stream received with bit rate of 384 kbit/s from the encoder and reproduces the multi-channel audio signals for analog and digital outputs. The multi-processing of this audio codec using multiple DSPs is ensured by high speed transfer of data between DSPs through coordinating communication port activities with DMA coprocessors. Finally, some technical considerations are suggested to realize the problem of real-time operation, which are found out through the implementation of this codec using the MPEG-2 layer II audio coding algorithm and the use of the hardware architecture with commercial multiple DSPs.

요 약

본 논문은 저비트율을 갖는 고품질의 HDTV용 멀티채널 오디오 코덱의 구현에 대해 기술한다. 이 코덱은 저주파수 효과 채널을 포함한 최대 3/2 스테레오 채널 구성, 최대 채널 구성보다 낮은 채널 구성과의 호환성, 기존 2채널 스테레오 시스템과의 호환성(MPEG-1 오디오), 그리고 다중 대화 채널 등을 제공하는 특징을 갖는다. 구현한 멀티채널 오디오 코덱의 인코더는 3개의 DSP(TI의 TMS320C40)로 구성되었고, 최대 48KHz 샘플링율과 16비트의 부호화를 갖는 5.1 채널의 아날로그 및 AES/EBU, IEC 958등의 포맷을 갖는 스테레오 2채널의 디지털 오디오를 입력으로 받아 지각 심리음향 모델을 사용하여 압축한후 384Kbps의 비트 스트림으로 전송하는 특징을 가지며, 디코더는 2개의 DSP로 구성되어 있고, 384Kbps로 입력되는 비트 스트림을 받아 최대 5.1 채널의 아날로그 및 2개의 2채널 스테레오의 디지털 오디오 신호로 출력시키는 특징을 갖는다. DSP를 이용한 다중처리는 DMA를 통한 통신포트를 이용한 DSP들간의 고속 데이터 전송에 의해 이루어진다. 끝으로, 멀티채널 오디오 코덱의 구현을 통하여 나타난 실시간 처리를 위해 고려해야할 기술적 사항을 제안한다.

I. INTRODUCTION

With the development of HDTV(High Definition Television) with its improved resolution and increased size, improved audio performance rather than the existing one is desired. To provide more realistic audio rendering, multi-channel audio coding system using more channels than the existing system with two channels is necessary. But the storage on expensive storage media or transmission over channels with limited capacity has led to develop the generic audio

*ETRI Acoustic Communication Section
한국전자통신연구소 음향통신연구실
접수일자 : 1995년 6월 6일

coding system for many applications dealing with digitally coded audio data and requiring low data rates with high quality[1][2]. Most of the audio coding systems are, therefore, considering the parameters of audio quality at different bit rates, sensitivity to transmission bit errors, complexity, and coding delay.

This paper describes the implementation of a multichannel audio codec with low bit rate for HDTV using the multiple DSPs(Digital Signal Processors). This codec consists of an encoder located in the transmitting part and a separated decoder of the receiving part, and can compress the 5.1 channel audio signals with the bit rate of about 3,848kbits/s into that of 384kbits/s. This codec has the features of the 3/2-stereo plus LFE(Low Frequency Effect), downwards compatibility with a smaller number of channels, backward compatibility with the existing 2/0 stereo system(MPEG-1 audio : Moving Picture Experts Group), and multilingual capability. In case of the encoder, three TMS320C40 DSPs with a pipelined configuration, which can ensure high speed transfer of data between DSPs, are used for generating a formatted digital audio bitstream. It is, also, possible to connect six analog audio inputs or four digital audio inputs through three stereo A/D converters or two stereo digital audio interface receivers, respectively. A decoder with two TMS320C40 DSPs releases the compressed audio signals from the formatted bitstream and can generate the same audio outputs as the encoder through three D/A converters or two stereo digital audio transmitters. The variable channel configurations are selected by the control of toggle switches in the encoder and the decoder.

MPEG 2 layer II audio coding algorithm applied to this codec, but, requires intensive computing power for the complicated calculations and the iterative processes, and then real-time operation of a codec is very difficult, specially for encoder[4][5]. It must be, therefore, considered a special hardware configuration and a software design scheme because of this constraint. Some technical considerations for real-time operation which are found out through the implementation, are suggested.

II. ESTABLISHMENT OF CODEC SPECIFICATION

Figure 1 shows the fundamental concept of the multichannel audio codec for HDTV. Receiving side is different in physical quantity of the audio signal from sending side, but both sides equal to in subjective audio quality. Table 1 shows the requirement

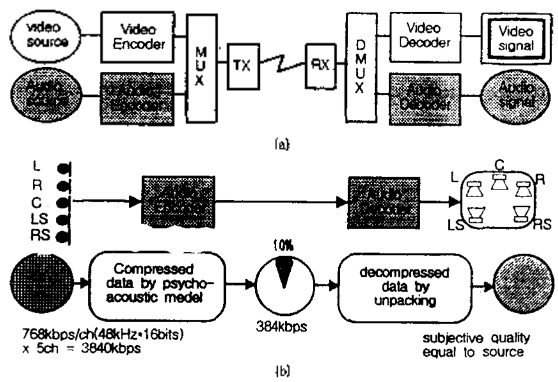


Figure 1. A fundamental concept of the multi-channel audio codec :
(a) for codec system of HDTV, (b) for audio codec.

Table 1. The specifications of the multi-channel audio codec.

items	requirement specifications	review matters
input/output interface	· analog audio : 5.1 channels · digital audio : 4 channels	· selecting coding mode by switch control
coding standard	· IEC/ISO 13818-3 International Standard	
bitstream format	· MPEG-2 bitstream format	· adaptation to changed bitstream
coding algorithm	· based on MPEG-2 layer II	· considering improved coding algorithm
channel configuration	· one of 3/2, 2/2, 2/0+2/0, 2/0 channel configuration managements	· selected configuration by switch control
multi-lingual channel	· use of 2/0 channel	· considering use of other multi-lingual channel later
bandwidth of audio signal	· 15Hz~20kHz	
sampling frequency	· 48kHz basically	· connecting input of 44.1kHz and 32kHz through an external frequency converter
bits/sample for input	· 16 bits	· considering above 16 bits
target quality	· fitness in assessment criteria of ITU-R Rec. 562	
bit rate	· 384kbit/s basically	· considering format of variable bit rates

specifications of the codec being considered in this paper. Multi-channel audio is composed of a center channel C and two surround channels LS(left surround), RS(right surround), in addition to the basic left and right stereo channels and an optional sub woofer channel.

It is referred to as 3/2-stereo(3 front/2 surround channels). Sub-woofer channel, corresponding to the sampling frequency which is equal to that of the main channels divided by a factor of 96, is capable of handling signals in the range from 15Hz to 120Hz[3][6].

Table 1 shows the specifications of the multi-channel audio codec.

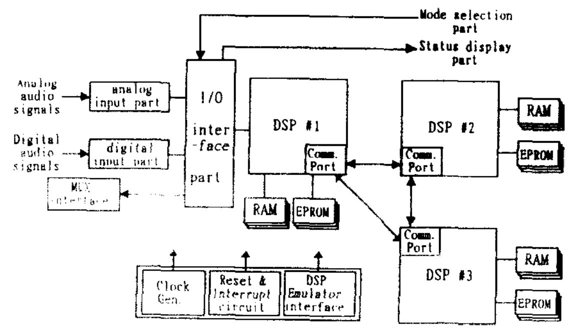
III. IMPLEMENTATION OF MULTI-CHANNEL AUDIO CODEC

3.1 Hardware

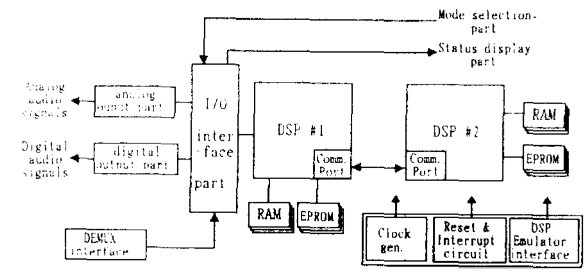
The structures of this codec for a multi-channel audio coding are designed within one board(size: 280mm×230mm) to shelve into HDTV system rack. An each unit using an architecture like the commercial DSP board is based on the Texas Instruments floating point DSP, 40MHz TMS320C40 with 25ns of access time(no wait states) and 50ns of execution cycle. A block diagram of the multi-channel audio codec is shown in Figure 2. The main feature of this audio codec is to apply the multi-processing technique by a multiple DSPs configuration which is ensured by high speed transfer of data between DSPs through coordinating communication port activities with DMA(Direct Memory Access) coprocessors[8].

The encoder part(Figure 2a) consists of 6 channel analog audio inputs, 4 channel digital audio inputs, and three DSPs for encoding procedure. The encoder implements a compression of multi-channel audio using a human perceptual psychoacoustic model and has the bit rate reduction to 384 kbit/s. The decoder part(Figure 2b) consists of 6-channel analog audio outputs, 4-channel digital audio outputs, and two DSPs for decoding procedure. The decoder analyzes the bit stream received with bit rate of 384 kbit/s from the encoder and reproduces the multi-channel audio signals for analog and digital outputs.

The audio signal of each channel is sampled at 48kHz sampling frequency, using 16 bits A/D and D/A. In the encoder, the digital data of one frame corresponding to 1152 PCM samples unit(24ms at 48 kHz) are buffered into the memory which can store digital data of 2 channels in DSP #1. The digital audio data read by DSP #1 are passed through DSP #2 and DSP #3 according to processing procedures, and then an encoded audio bitstream with a formatted structure controlled by DSP #2 and DSP #3 returns back to DSP #1 for sending to the decoder.



(a) Encoder



(b) Decoder

Figure 2. The block diagram of multi channel audio codec.

The data communications between DSPs are realized by coordinating communication port activity with CPU and DMA of TMS320C40. DMA channels are used in split mode including primary channel and

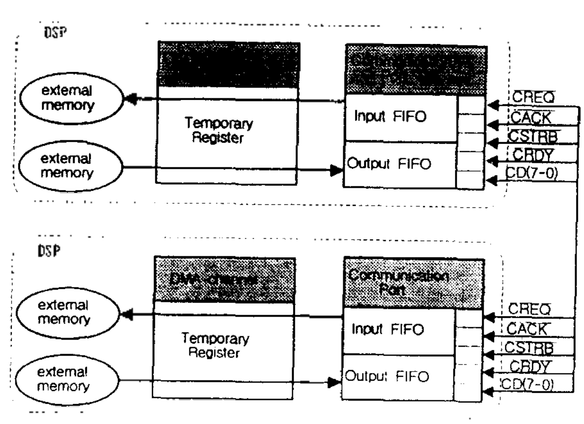


Figure 3. DMA configuration in split mode.

auxiliary channel as shown in Figure 3. Primary channel indicates to read data from a location in the external or internal memory map and write it to a communication port. Auxiliary channel means to receive data from a communication port and write it to a location in the memory map.

The mode selection part which consists of three toggle switches performs the function of channel selection and the status display part of LEDs indicates the status of selected channels and codec activities. The control logic and the decoding logic for peripheral interface are designed to a FPGA.

3.2 Software[1][3][9]

The block diagram of software flow for this multi channel audio codec is shown in Figure 4. Using a sampling frequency of 48kHz, the polyphase filterbank divides the audio signal into 32 subbands with a constant bandwidth of 750Hz. In each subband 36 successive samples are combined to three blocks with 12 samples. This is equivalent to a duration of 24msec. In each block the absolute peak value is determined and quantized as a scalefactor with a dynamic range of 120dB. To reduce the bitrate for the scalefactors, three scalefactors of each subband of one frame are considered together and classified into certain scalefactor patterns. Depending on the pattern, one, two, or three scalefactors are transmitted with an additional scalefactor select information consisting of 2 bits per subband.

The psychoacoustic model calculates the minimum masking threshold which is necessary to determine the noticeable noise level for each band. To compensate for the lack of accuracy of the spectrum analysis of filterbank, a 1024-point FFT is used in parallel to the process of filtering. The output of the FFT is used to determine the relevant tonal and non-tonal of audio signal. The individual masking thresholds for each masker above the absolute masking threshold are calculated dependent on frequency position, loudness level, and tonality. All the individual masking thresholds, including the absolute threshold are added to the so-called global masking threshold. For each subband, the minimum value of this masking curve is determined. Finally, the difference between the maximum signal level and the minimum masking threshold, called signal-to-mask ratio(SMR), is used in the bit or noise allocation to determine the actual quantizer level in each subband for each block.

The bit allocation procedure is determined by minimizing the tonal noise-to-mask ratio over every subband and the whole frame. This procedure is an iterative process where, in each iteration step the number of quantizing levels of the subband that has the greatest benefit is increased with the constraint that the number of bits used does not exceed the number of bits available for that frame.

Decoder separates the bit allocation information with scalefactor and 12 successive samples of each subband signal from the received multiplex signal. The reconstruction process to obtain again PCM audio is characterized by filling up the data format of

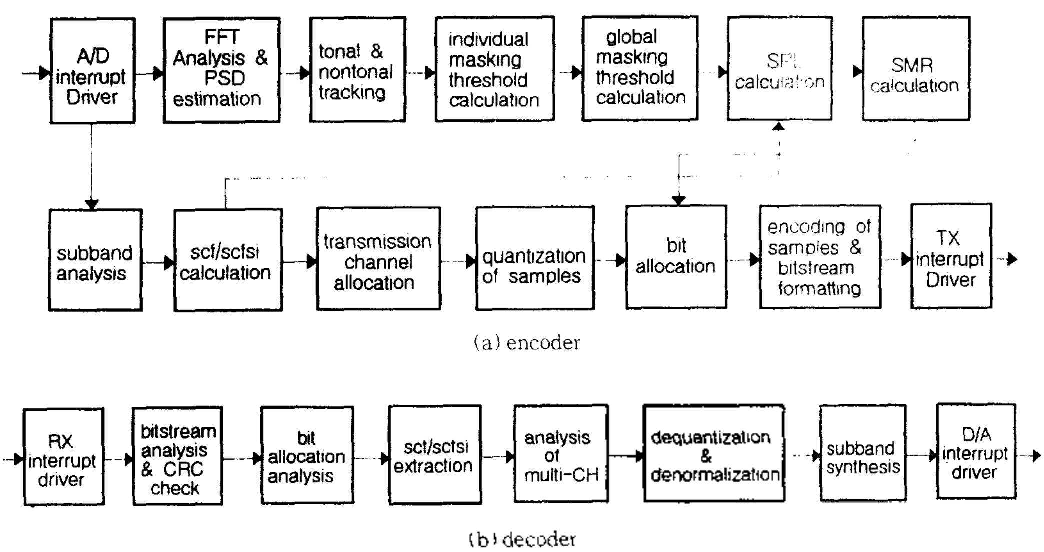


Figure 4. The block diagram of software flow.

the subband samples regarding the scalefactor and bit allocation for each subband and frame. The synthesis filterbank reconstructs the complete broad band audio signal with a bandwidth of up to 21kHz.

3.3 *Quality test through simulation*

A rendering environment for simulation and test of multi-channel audio is shown in Figure 5. For audio accompanying HDTV, the three front loudspeaker channels ensure sufficient directional stability and clarity of the picture related frontal images. Center channel can guarantee importance for most of the dialogue at any location of the listener.

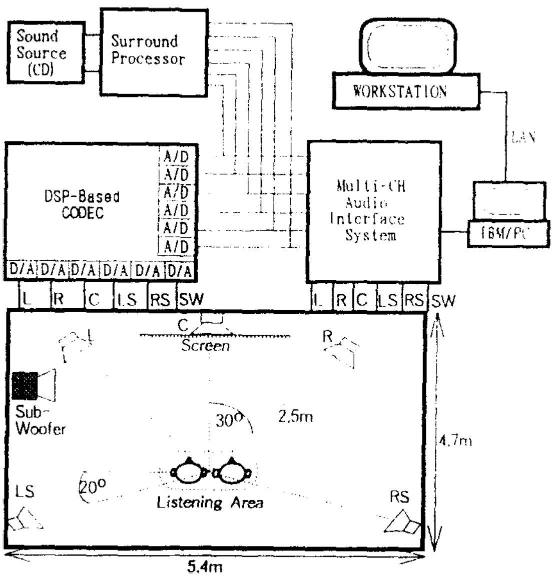


Figure 5. A rendering environment for simulation and test of multi-channel audio.

One pair of surround loudspeaker channels(LS, RS) allows improved realism of auditory ambience. The purpose of optional sub-woofer channel to supply a low frequency enhancement is to enable listeners to extend the low frequency content of the reproduced programme in terms of both frequency and level.

For the simulation, the audio signals of two channels with CD quality are reproduced to six channels from a digital surround processor(Lexicon CP-3), and then recorded into the hard disk of multi-channel audio interface system, using IBM/PC. These recorded data are transferred to SPARC workstation through LAN and coded by the encoding and the decoding software developed.

The compressed final data are sent into IBM PC and displayed at the listening room in which six loudspeakers are located like Figure 5. The informal subjective listening test and the paired comparison test are carried out for quality assessment.

Ⅳ. SUGGESTIONS FOR REAL-TIME OPERATION

In case of 48kHz sampling and 384kbit/s bit rate, the input to output processing time for one frame of 1152 samples of PCM audio data must be less than 24ms. The MPEG-2 layer II audio coding algorithm applied to this multi-channel codec, but, requires many computing times for the complicated calculations and the iterative processes, and then it is very difficult to operate software on real-time, specially for encoder. It must be considered a special hardware configuration and software design scheme because of this constraint. The divided parallel software architecture is applied to this codec in addition to hardware development with a pipelined multiple DSPs.

4.1 Hardware design using a pipelined DSP architecture

A block diagram of the suggested hardware design considering real time operation is shown in Figure 6. This design scheme is termed here a pipelined DSPs architecture which consists of a serial connection between DSPs. In addition, this architecture has input/output FIFOs for increasing the execution time of DSP and a shared memory for faster transfer of DSPs communication. 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, and DSP #3 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 total execution time for one frame from input audio data to output bitstream can be extended up to 72msec even if it has transmission delay.

Figure 7 shows the devided time chart and the processing schedule for each DSP in encoder, organized by using the pipelined DSPs structure. The overall processing software of one frame is devided into 3 steps which are assigned to each DSP, step-1 for

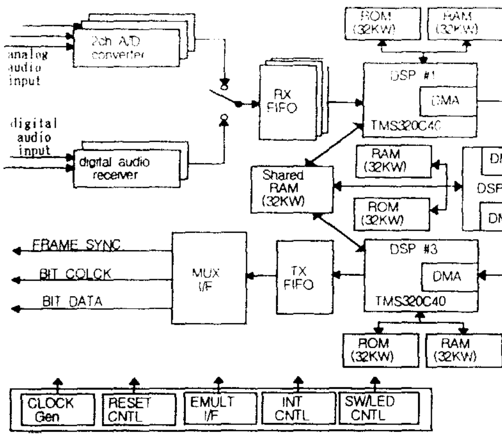


Figure 6. An example of a pipelined DSP structure with I/O FIFO(in case of encoder).

DSP #1, step-2 for DSP #2, and step-3 for DSP #3. Each DSP must, that is, finish the assigned processing itself for incoming frame within 24ms, and transfer the rest processing to next DSP. This technique can extend the limited time of 24ms to 72ms through some delays.

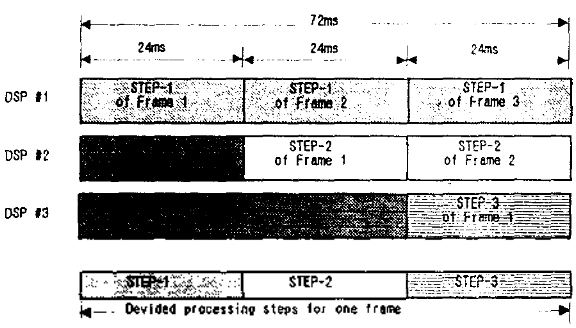


Figure 7. Divided time chart for each DSP in encoder processing.

4.2 Practical use of the configuration of the target DSP system

The TMS320C40 DSP, which is a main controller of target system, has the main features and many benefits, and the configuration of the 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 multichannel audio codec, real-time operation requires that unnecessary processor overheads be kept to a minimum.

Most systems use TMS320C40 interrupts to handle the transmission of bitstream and sample I/O. The overhead of handling the interrupts increases the processor loading not only due to the execution time of the interrupt code, but also by disrupting the on-chip instruction cache. This effects 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, is active only for external memory fetches. In addition, the use of the RPTS instruction with 4 instruction cycles in case of TMS320C40 is not recommended because this instruction sequence is not interruptable and cause the CPU to miss interrupts.

Another tip can be produced in the input/output buffering. In a typical system, audio sample input/output is performed through a direct interface to A/D and D/A converters. An additional circular buffer in memory can be used to DMA audio sample from the interface port using the TMS320C40 on-chip DMA controller. This allows the size of the circular buffer to be independent of the size of the blocks of samples required by the codec and makes the buffer management considerably more flexible. The execution time overhead of copying the samples from the circular buffer is very small.

4.3 Consideration in software design and coding

It is very important to use the features of the target processor that can help in producing faster and/or shorter programs. The following steps must be applied in generating a optimized code for increasing the execution speed : 1) write the application in the C language and debug the program, 2) estimate if it runs in real-time, 3) if it doesnt, identify places where most of the execution time is spent, 4) optimize these areas by writing assembly language routines that implement the functions, 5) call the routines from the C program as C functions.

To optimize assembly code, especially, the suggestions presented as followings can be used : 1) use delayed branches, and delayed subroutine call and return, 2) apply the repeat single/block construct, 3) use parallel instructions and maximize the use of registers, 4) use the cache and internal memory instead of external memory, 5) avoid pipeline conflicts.

The use of look-up tables for the arithmetic functions and the logic designs for the calculation algorithms are, also, suggested for saving the execution time.

V. CONCLUSION AND FUTURE DIRECTION

The multi-channel audio codec which has a compression rate of one-tenth, and consists of an encoder and a decoder, was presented with the implementation of hardware and software based on multiple DSPs. The some suggestions for hardware and software considering real time operating were, also, described. The pipelined DSPs architecture with I/O FIFO can extend the limited processing time even if some delay occurs.

This prototype codec can reduce a transmission bit rate of about 3,848kbit/s by 384kbit/s using perceptual coding of psychoacoustic modeling and has the characteristics of the 3/2-stereo plus LFE, downwards compatibility with a lower number of channels, backward compatibility with the existing 2/0-stereo system, and multilingual capability. The software implementation of algorithm, the development of hardware, and the simulation of multi-channel audio codec except for real-time processing had been finished and The audio quality of simulation result was good enough for listening test. The suggested considerations are identified as a powerful method for real-time operation and an optimization is, currently, progressing.

The applications of multi channel audio codec are in various fields except multi channel audio for HDTV. For example, computer-based multimedia, DAB(Digital Audio Broadcasting), audio recording and presentation, and realistic telecommunications can be considered.

The further work includes the improvement of the audio quality, the quality assessment method for multi-channel audio, and the implementation of ASIC for this codec.

Acknowledgements

The author wishes to thank his partners in the project of MPEG-2 audio codec development, in particular Mr. Daiyoung Jang and Mr. Seonghan Kim who have greatly contributed to hardware design and software coding respectively. Special thanks for careful considerations are due to Dr. Seonghoon Kang, head of acoustic communication section.

REFERENCES

1. 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.
2. 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., 1992.
3. ISO/IEC 13818-3, Coding of moving pictures and associated audio : Audio, 1994.
4. D. Wiese and G. Stoll, "Bitrate Reduction of High Quality Audio Signals by Modeling the Ears Masking Thresholds," The 89th AES Convention, LA, Preprint No. 2970, Sep., 1990.
5. Ken N. Linson, "Architectural Issues for Parallel DSP Audio Systems," AES UK DSP Conference Proceeding, London, pp.21~36, Sep., 1992.
6. 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., 1993.
7. S. Shlien, "Guide to MPEG-1 Audio Standard," IEEE Transactions on Broadcasting, Vol. 40, No. 4, pp.206~218, Dec., 1994.
8. Texas Instruments, TMS320C4x User's Guide, 1991.
9. J. W. Hong, "Analysis of ISO/MPEG 1 Audio Coding Method," ETRI Electronics and Telecommunications Trends, Vol. 10, No. 2, pp.191~201, Jul., 1995. (in Korean)
10. CCIR SG, "Proposed Universal Multi-channel Coding System Compatible with ISO 11172-3," Delayed Contribution, Document 10-1/12-E, Nov., 1992.

▲ Jin-Woo Hong

Jin-Woo Hong was born in Yu-Joo, Korea, on April 15, 1959. He received the B.S. and M.S. degree in Electronic Engineering in 1982 and 1984, and the Ph.D degree in Computer Engineering from Kwang-Woon University, Seoul, in 1993. Since 1984 he has been with Telecommunications Technology Division in ETRI, where he is now a senior member of technical staff. His research interests include the ISDN, electric acoustics, and transmission quality.