Different Audio Coding Techniques

Ms. Sabale Amruta V and Dr. Sarwade Nisha

***ABSTRACT*: In the recent years, large scale audio data transfer by remote computing and the development of massive storage and retrieval systems have witnessed a tremendous growth. So, the audio compression is required. If the redundancy from the audio signal is removed then compression can be done. Two types of compression are there, lossy and lossless. different audio coding technique such as mp3, advanced audio coding (aac), aac+ are used for compression of audio signal. The compression algorithms help to reduce the bandwidth requirements and also provide a level of security for the data being transmitted.**

Keywords— Audio Compression, Lossy and Lossless Data Compression, MP3, Advanced Audio Coding (AAC), AAC+.

I: INTRODUCTION

Portable electronic devices such as smart mobile phones, digital cameras and digital audio devices with audio players and recorders have been attractive now a days particularly due to prevalence of MP3 audio files. MP3 is the popular name of MPEG-1 layer-3 audio. Moreover, the so-called MP3's successor, MPEG-2 Advanced Audio Coding (AAC), finalized as an international standard in 1997 which was developed to achieve a higher quality than that of previous coder that is MP3. AAC reaches the same sound quality as MP3 at about 70% of the bit rate. This way more compression is done in AAC as compare with MP3.

High quality audio compression has found its way in many applications. Early research on audio has translated into standardization efforts of ISO/IEC and ITU-R 10 years ago. In the last couple of years, Internet audio broadcasting has come in powerful category of this type of high quality applications. These techniques become more and more popular in many parts of the world because of the business for the music industry.

This paper is organized as follows. In Section II, An overview of Audio Signal discussed. In Section III, survey of different Audio Coding techniques described. Finally, conclude this paper in Section IV.

Ms. Sabale Amruta V is a M.TECH.[Electronics and Telecommunication] Student, Department Of Electrical Engineering, VJTI, Mumbai, Maharashtra and **Dr. Sarwade Nisha** is working as Asst. Professor, Department Of Electrical Engineering, VJTI, Mumbai, Maharashtra, Emails: amrutas1990@gmail.com**,** npsarwade@vjti.org.in

II: AUDIO SIGNAL-AN OVERVIEW

Audio signal is the signal with frequency range of 20 Hz to 20 KHz. Human speeches and other musical component's sounds are merged together and it is called as audio signal. Broadcast of audio used 16-bit PCM encoding at 44.1 kHz, such an application would require a 1.4 Mbps [1] channel for a stereo signal (44.1KHz*16bit=705.6 Kbps for Mono audio signal). Since the beginning of the twentieth century, the art of sound coding, transmission of audio signal, recording of audio signal, and also the mixing and re-production of it has been constantly evolving. Starting from the mono-phonic technology, technologies on multichannel audio have been extended to include stereophonic, quadraphonic, 5.1 channels, 7.1 channels etc. Compared with the traditional mono or stereo audio signal the multichannel audio provides end users with a better experience and becomes more and more appealing to music producers. So, an efficient coding scheme is needed for the storage and transmission of multichannel audio and this subject has attracted a lot of attention now a days.

There are several multichannel audio compression algorithms. Dolby AC-3 and MPEG Advanced Audio Coding (AAC) are the two most prevalent perceptual digital audio coding systems. Dolby AC-3 is the 3rd generation of digital audio compression systems from Dolby Laboratories and has been used as the audio standard for High Definition Television systems. It is capable of providing transparent audio quality at 384 kb/sec for 5.1 channels. This 5.1 channel technology is come in the categories of multichannel audio. In MPEG family there are lots of different algorithms which can be used for compression of audio file. MP3 is most popular technique used for audio compression which supports to only up to two channels (stereo coding). There are also multichannel audio compression algorithms. Among that AAC is currently the most powerful multichannel audio coding algorithm of the MPEG family. It can support up to 96 audio channels. These low bit rate multichannel audio compression algorithms are utilized transform coding techniques to remove statistical redundancy within each channel of multichannel audio file. The audio compression is possible with different audio coding techniques. The audio file must be compressed without reducing the quality. Table 1 summarized brief history of MPEG family.

Table 1 Brief history of MPEG audio standards

Year	Standards	Sampling rate KHz	Bit rate Kbits/sec	Channels
1992	MPEG-1 Layer I	32, 44.1, 48	32 – 448	1-2
1992	MPEG-1 Layer II	32, 44.1, 48	32-384	1-2

1992	MPEG-1 Layer III	32, 44.1, 48	32-320	1-2
1994	MPEG-2 Layer I	32, 44.1, 48	32-448	1-5.1
		16, 22.05, 24	32-256	
1994	MPEG-2 Layer II	32, 44.1, 48	32-384	1-5.1
		16, 22.05, 24	8-160	
1994	MPEG-2 Layer III	32, 44.1, 48	32-384	1-5.1
		16, 22.05, 24	8-160	
1997	MPEG-2 AAC	8, 11.025, 12, 16 22.05, 24, 32, 44.1 48, 64, 88.2, 96	8-64 K/Ch	1-96
1998	MPEG-4 AAC	8, 11.025, 12, 16 22.05, 24, 32, 44.1 48, 64, 88.2, 96	8-64 K/Ch	1-96
2003	MPEG-4 HE AAC	8, 11.025, 12, 16 22.05, 24, 32, 44.1 48, 64, 88.2, 96	8-64 K/Ch	1-96

III: SURVEY OF DIFFERENT AUDIO TECHNIQUES

1) Sum-Difference Stereo Transform Coding:

The coder architecture for this technique was explained by Johnston and Ferreira [2] and it is shown in Fig. 1. There are four basic blocks for all perceptual coders. In the case of Perceptual Audio Coder (PAC), the filter bank block is implemented by an MDCT (Modified Discrete Cosine Transform) with the optional window switching abilities. The psychoacoustic analysis provides a noise threshold for the L (left), R (Right), M (Sum), and S (Difference) channels, as may be appropriate, for both the normal MDCT window and the optional shorter windows. The thresholds for the left and right channels THR_L and THR_R are calculated. This two thresholds are compared where the thresholds vary between left and right by less than 2dB, then the coder is switched into M/S mode, i.e. the left signal for that given band of frequencies is replaced by M = (L+R)/2 and the right signal replaced by S = (L-R)/2.

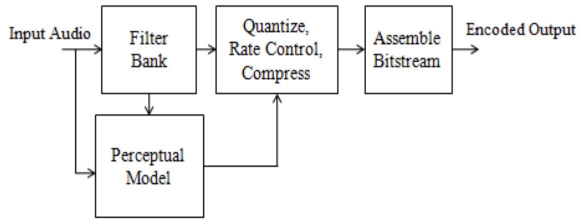

Fig 1 .Block Diagram of Perceptual Audio Coder (PAC) [2]

The method gives a substantial improvement over the dual monophonic case. The use of stereo redundancy in a time and frequency varying way results in a substantial increase in encoded signal quality.

2) Improving Joint Stereo Audio Coding by Adaptive Inter-channel Prediction:

The above MS-stereo coding [1] does not achieve any improvement for the class of most critical test sequences. In MS-stereo coding only the statistical dependencies between two samples of the left and right channels of signals are considered. A stereophonic sound signal is characterized by level differences as well as phase or time delay between the left and right channel signals. So, an appropriate and efficient stereo redundancy reduction technique has to be taken into account. So, for the improvement of it adaptive inter-channel predictor (AICP) is used, which compensates a possible phase or time delay and exploits more than one value of the cross-correlation function between the left and right channels of a sound signal [3]. From successive samples of input signal $x(n)$ in one channel the estimate of the actual sample of signal $y(n)$ in the other channel is calculated,

$$\hat{y}(n) = \sum_{k=0}^{K} a_k \cdot x(n - d - k) \qquad ...1$$

Where, k is the predictor order a_k is the predictor coefficients and d is a delay for compensation of phase or time delay between the two signals. The prediction error is then,

$$e(n) = y(n) - \hat{y}(n) \qquad ...2$$

Compared to the variance of $y(n)$, the variance of the prediction error $e(n)$ is reduced. Therefore, a bit rate reduction is achieved by coding and transmitting $e(n)$ instead of $y(n)$. And the prediction gain is given by the ratio of the variances.

$$G = \frac{E[y^2(n)]}{E[e^2(n)]} = \frac{\sigma_y^2}{\sigma_e^2} \qquad ...3$$

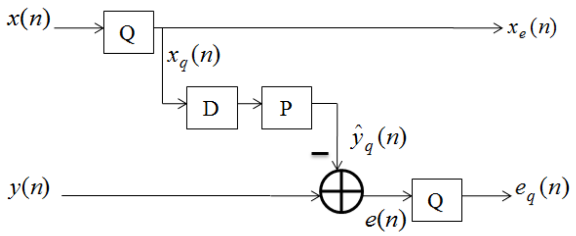

Fig 2(a) .Encoder block diagram of AICP [3]

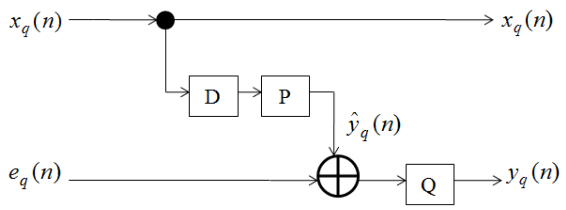

Fig 2 (b) .Decoder Block diagram of AICP [3]

Fig. 2(a) and Fig. 2(b) show the encoder and decoder block diagram of AICP with quantization. Where, $x(n)$ and $y(n)$ represent the samples of the left and right channel of a stereophonic sound signal respectively and Q is the Quantizer, D is the Delay and P is the Predictor. The Sum difference method does not achieve any improvement of some critical audio file as told earlier, this improvement is done by AICP.

3. Scalable Audio Coder Based on Quantizer Units of MDCT Coefficients:

A scalable codec has been constructed by using transform coding and the basic modules of scalable coder (encoder and decoder). The basic module is a quantizer that can quantize MDCT (Modified DCT) coefficients transformed from a variety of frequency regions. This module works at bitrates of more than 8 kb/s. Also the scalable structure can be changed according to the input signals. In the scalable codec described here, the input-output signals are **monaural** and the sampling frequency is 24 kHz. The total bit rate of this scalable codec is more than 8 kb/s [4].

The basic module is mainly constructed from a Twin VQ (Transform-domain Weighted Interleave Vector Quantization) codec. It is type of transform coding. Transform coder is used in audio coding. This module is a quantizer for the MDCT coefficients. Here, 4-layer scalable codec as shown in Fig. 3. This codec uses four basic modules with input sampling frequency of 24 kHz [4]. This is a 4-layered scalable codec, but it is possible to make any number of layers using these basic modules. The basic module for first layer (#1) has a fixed range of input or output frequency and other basic modules (#2, #3, #4) have a variable range of input or output frequency with each frequency band width fixed. The frequency point information of each basic module is added to the coded bit stream.

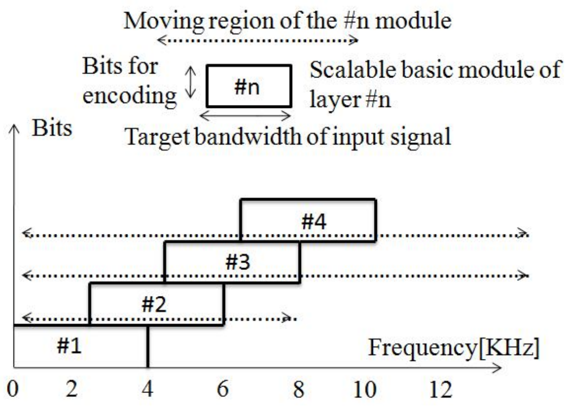

Fig 3 .Hierarchical Structure of Scalable Codec [4]

In Fig. 3, for example, the input 4-kHz signal is quantized in #1 module first and then its quantization error is again quantized in #2 and #3 modules. Here also one can change the width (frequency band width), height (number of bits for each frequency), and position (target frequency) of the each module.

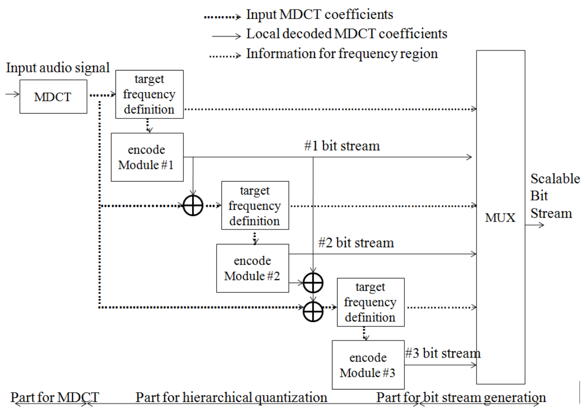

Fig 4 .Structure of Scalable Encoder [4]

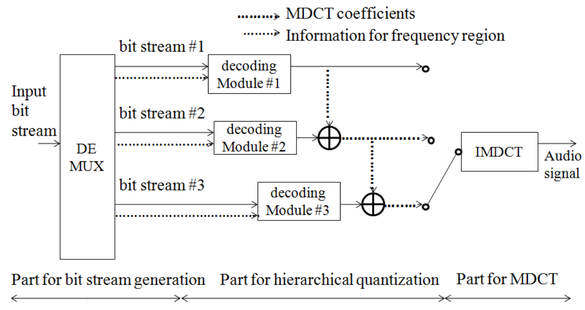

Fig 5 .Structure of Scalable Decoder [4]

Fig. 4 shows the Structure of Scalable Encoder. It has parts for MDCT, hierarchical quantization and bit stream generation. In the MDCT, the input time-series signals are transformed to MDCT coefficients according to its nature by particular transform points. And the scalable decoder has 3 parts for analyzing the bit stream, hierarchical re-quantization and Inverse MDCT. The Decoder is totally reversing then Encoder. The Decoder is shown in Fig. 5.

Subjective quality evaluation tests, for musical sound sources, showed that its sound quality is better than MPEG-layer3 codec at 8, 16 and 24 kb/s when scalable codec is constructed of 8-kb/s basic modules [4].

4. A Study of Why Cross Channel Prediction is not Applicable to Perceptual Audio Coding:

There has been a question that whether the compression rate of a multichannel perceptual audio coder can be increased by applying the cross channel linear prediction (LP) in the time domain or not. There exists correlation between two channels which can be removed by using cross channel linear prediction. Hence, theoretically, by coding the prediction residual instead of the original signal, the coder should achieve a higher compression ratio, at least without considering the requirements of human perception.

There is considerable correlation between the channel pairs C-L, C-R, L-R, Ls-Rs, L-Ls, and R-Rs. The M/S stereo coding and intensity stereo coding in the MPEG AAC can highly remove the correlation between the pairs L-R and Ls-Rs. But the rest of the channel pairs, it seems that one can take good advantage of the correlation between C and L channel and between C and R channel. Means one can get some coding gain by using C to predict L and R.

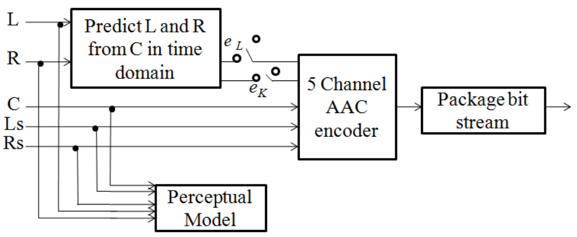

Fig 6 .Prediction-incorporated AAC coder[5]

But Problem is with low bit-rate coder. If the bit-rate is <32 kb/s [5] then cross channel prediction is effectively reducing the signal energy and coding bits in low frequency bands is also reducing, but increase the coding bits in high frequency bands. So, the increase in high frequency bits exceeds the bit reduction at low frequencies, which results in a net increase of coding bits required. So, overall improvement is not done and this is the reason why cross channel prediction is not applicable to perceptual audio coding.

5. MPEG-1 Layer-III (MP3):

MP3 is a lossy compression technique it means some audio information is certainly lost by using this compression technique. This loss can be noticed because the compression technique tries to control it. MPEG-1 audio describes three layers of audio coding with the following properties:

- Mono or stereo audio channels.
- Sample rate 32 kHz, 44.1 kHz or 48 kHz.
- Bit rates from 32 kbps to 448 kbps [6].

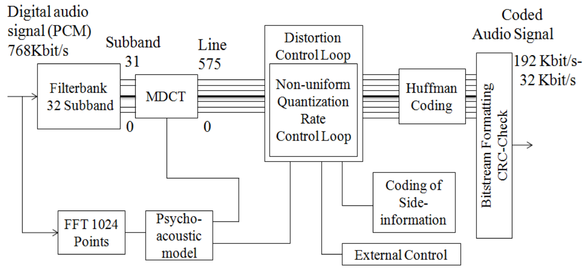

Fig 7 .Block diagram of MP3 Encoder[6]

Fig. 7 shows the block diagram of MP3 encoder. There are two filter banks in a MPEG audio algorithm, namely a filter bank and a hybrid polyphase/MDCT filter bank. The input PCM samples are simultaneously given to a filter bank and a psychoacoustic model. This Filter bank splits the signal into 32 equal sub bands in frequency domain and psychoacoustic model takes the signal spectrum as input and it determines the ratio of signal energy to masking threshold for each sub band. For better frequency resolution the 32 sub bands are further divided into 576 frequency lines by the MDCT. Here MDCT used is 12 point (short) or 36 point (long) with 50 % overlap and the type of MDCT (long or short window) is determined by the window switching algorithm [6]. In Layer-3, these coder partitions are roughly equivalent to the critical bands of human hearing system. If the quantization noise can be below the masking threshold for each coder partition, then the compression result should be indistinguishable from the original signal.

The signal to masking ratio (SMR) which is calculated by the psychoacoustic model is used by the quantizer to determine the number of bits that should be allocated for the quantization of the sub band coefficients. Here the quantization is done by the power-law quantizer. The quantized values are coded by Huffman coding. Then Finally the Huffman coded values are formed into a bit stream. A bit stream formatter is used to assemble the whole bit stream. The encoded bit stream consists of quantized and coded spectral coefficients with some side information like bit allocation information and quantizer step size information.

6. A Multi-Channel Audio Compression Method with Virtual Source Location Information (VSLI) for MPEG-4 SAC:

Binaural cue coding (BCC) was introduced as an efficient representation method for MPEG-4 SAC (Spatial Audio Coding). But when bit-rate is low then the spectrum of BCC output signals degrades with respect to the perceptual level. The system here is estimates VSLI (virtual source location information) [7] as the side information. The VSLI is the angle representation of spatial images between channels on playback layout as shown in Fig. 8.

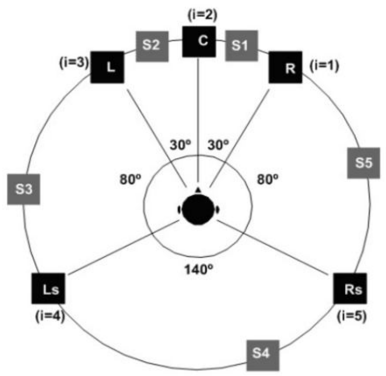

Fig 8 .Playback speaker layout of VSLI Method [7]

The concept of BCC is the separation of the information relevant for the spatial perception of multi-channel audio signals. BCC represents multi-channel signals as a mono/stereo audio signals and BCC parameters. The mono/stereo audio signal is the sum signal getting from all sound sources which are to be part of the spatial image of the multi-channel signals.

VSLI gives the angle information of sound image. VSLI encoder shown in Fig. 9 estimates angle information of sound image in each band and each frame with multi-channel gain information and VSLI decoder (Fig. 10) reconstructs the each channel gain with the angle information. The per-band and per-frame angle information is called Virtual Source Location Information (VSLI).

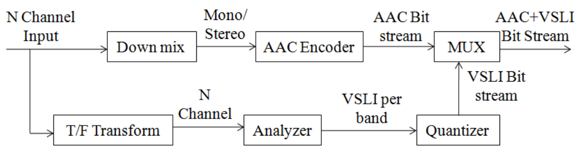

Fig 9 .Schematic diagram of VSLI-based SAC encoder [7]

The VSLI is azimuth information. It represents the geometric spatial information between power vectors of inter-channel frequency bands rather than power ratio such as conventional cue like ICLD (Inter Channel Level Difference). In VSLI-based SAC system, the entire source channels are transformed (DFT). And the audio input signal in each channel is divided into the non-overlapping partitions of time-frequency plane.

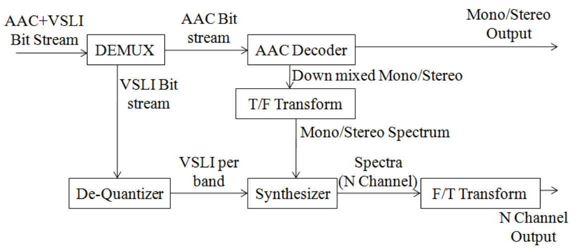

Fig 10 .Schematic diagram of VSLI-based SAC decoder [7]

By comparing the conventional ICLD-based BCC system, the VSLI-based SAC system reduces the bit-rate about 5 % [7].

IV: Conclusion

Different audio coding techniques are discussed in this paper and it can be concluded that since 1992 there are many technique implemented for compression of the Audio file. The Sum-difference audio coding method [2] gives the good results for stereo files. But for some critical audio file it does not give better results and this problem is solved by AICP method [3]. Using 8 Kb/s basic module one can get the better sound quality than the MPEG-layer-3 with different (8 Kb/s, 16 Kb/s 24 Kb/s) scalable bit-rates. Also seen that why cross channel is not used for prediction of left and right channel [5]. Above all techniques are for either Mono or Stereo files (Supporting to only 1-2 channels). VSLI method gives the 5% bit-rate reduction than the conventional ICLD (Inter Channel Level Difference) method.

REFRENCES

[1] Peter Noll, “MPEG Digital Audio Coding,” *IEEE Signal Processing Magazine*, pp. 59-81, September 1997.

[2] J. D. Johnston and A. J. Ferreira, “Sum-difference stereo transform coding,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 2, pp. 569–572, Mar. 1992.

[3] H. Fuchs, “Improving joint stereo audio coding by adaptive inter-channel prediction,” *IEEE Conference on Applications of Signal Processing to Audio and Acoustics*, pp. 39–42, Oct. 1993.

[4] A. Jin, T. Moriya, T, Norimam, M. Tsushima and T. Ishikawa, “Scalable Audio Coder Based on Quantizer Units of MDCT Coefficients,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 2, pp. 897-900, Mar. 1999.

[5] S. Kuo and J. D. Johnston, “A study of why cross channel prediction is not applicable to perceptual audio coding,” *IEEE Signal Processing Letters*, vol. 8, No. 9, pp.245-247, Sept. 2001.

[6] Mark Kahrs and Karlheinz Brandenburg, “Applications Of Digital Signal Processing To Audio And Acoustics,” Kluwer Academic Publishers, 2002.

[7] Han-gil Moon, Jeong-il Seo, Seungkwon Baek, and Koeng-Mo Sung, “A Multi-Channel Audio Compression Method with Virtual Source Location Information for MPEG-4 SAC,” *IEEE Transactions on Consumer Electronics*, Vol. 51, No. 4, pp.1253-1259, Nov. 2005.