

Performance Evaluation of Audio Coding by Amalgam AAC and FLAC Audio codec using MDCT and INTMDCT Algorithm

M. Davidson Kamala Dhas, and R. Priyadharsini

Abstract—The MDCT and IntMDCT Algorithm is widely utilized in Audio coding. By lifting scheme or rounding operation IntegerMDCT is evolved from Modified Discrete Cosine Transform. This method acquire the properties of MDCT and contribute excelling invertibility and good spectral mean. In this paper we discuss about the audio codec like AAC and FLAC using MDCT and Integer MDCT algorithm and to find which algorithm shows better Compression Ratio(CR).The confines of this task is to hybridizing lossy and lossless audio codec with diminished bit rate but with finer sound quality. Certainly the quality of the audio is figure out by Subjective and Objective testing which is in terms of MOS (Mean opinion square), ABx and some of the hearing aid testing methodology like PEAQ(Perceptual Evaluation Audio Quality) and ODG(Objective Difference Grade)is followed. Execution measure, that is Compression Ratio(CR) and Sound Pressure Level (SPL) is approximated.

Keywords—Audio Codec, Advanced Audio Coding (AAC), Free Lossless Audio Codec (FLAC),Modified Discrete Cosine Transform (MDCT),Integer Modified Discrete Cosine Transform (IntMDCT), Mean Opinion square (MOS), Perceptual Evaluation Audio Quality (PEAQ)

I. INTRODUCTION

THE Integer MDCT is develop by an Integer estimation of Modified discrete cosine change utilizing lifting plan, it is reversible [1]. In Analysis technique MDCT can be appropriated as a sampled filter bank with 50%overlapped windows. This analysis is very advantageous to ease the blocking artifacts that depreciate the reconstruction of transform audio coders [2].Subsampling operation is performed to achieve critical sampling in frequency domain, and subsampling operation is by and by eliminated in time domain by Overlap and add operation [4]. IntMDCT transform can build the gap between lossless and perceptual audio coding. Integers transforms can be precisely converted from their prototypes by restoring given rotation with three lifting steps or rounding operations [6].The scope of this work is to design AAC and FLAC audio codec (lossy and lossless) with better sound quality and with minimum bit rate .The audio quality is estimated by performance measures such as SPL and CR and with the help of listeners the audibility is measured by two testing process i.e,Subjective and Objective testing.

M.Davidson Kamala Dhas is with ECE,Mepco Schlenk Engineering College, Sivakasi, India (Sivakasi,India mdavid@mepcoeng.ac.in).

R.Priyadharsini, is with ECE, Mepco Schlenk Engineering College, Sivakasi, India (Priyadharsini0294@gmail.com).

IntMDCT transform can build the gap between lossless and perceptual audio coding. Integers transforms can be precisely converted from their prototypes by restoring given rotation with three lifting steps or rounding operations [6].The scope of this work is to design AAC and FLAC audio codec (lossy and lossless) with better sound quality and with minimum bit rate .The audio quality is estimated by performance measures such as SPL and CR and with the help of listeners the audibility is measured by two testing process i.e,Subjective and Objective testing.

II. RELATED WORK

KARLHEINZ BRANDENBURG [12] shows about MP3.Advanced Audio Codec has been constructed to compress the audio (music) and its aim is to overcome the deprivation faced in the MPEG layer. The reproduced sound signal gets same as that of original signal. The reconstructed audio signal receives same as that of original signal.

International Telecommunication Union BS.1187-1, [16] with the MPEG,the moving picture master group exhorted the testing approach for estimating the sound quality pursues two form to be specific fundamental and propelled variant of the apparent sound signal.

Rajesh Kumar [14] characterizes the Mean Opinion Square for objective and Subjective listening utilizing two diverse datasets. Advanced version uses the filterbank used model and the basic version uses the Fast Fourier transform based model. Rongshan Yu,Susanto Rahardja,Lin Xiao,Chi Chung Ko,[7] describes an algorithm called an Advanced Audio Zip. Hence, it produces a lossless audio coding .It provides a better CR.In [5] proposed a method known as LD TDAC analysis (Low delay time domain aliasing cancellation) and faster implementation mapped in synthesis filter bank.These are designed by Codec methods.

III. DEFINITION AND BASIC FACTS

A. Lifting scheme

Orthogonal block transform are disintegrated into given rotations,

$$\begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \quad (1)$$

In lifting Scheme, rot of 2*2 lattices is utilized in the foundation of wavelet changes. Lifting steps is meant as the fundamental structure of lifting scheme.where 'z' is called as the lifting coefficient.

$$L_z = \begin{bmatrix} 1 & 0 \\ z & 1 \end{bmatrix} \quad (2)$$

The lifting step L_z design two value (x_1, x_2) to,

$$L_z(y_1, y_2) = (y_1, y_2 + zy_1) \quad (3)$$

Inverse lifting step is,

$$L_z^{-1} = L_z = \begin{bmatrix} 1 & 0 \\ -z & 1 \end{bmatrix} \quad (4)$$

Rounding function[.] can be combined into the lifting step L_z and the appeared integer lifting step is ,

$$L_{z,[\cdot]}(y_1, y_2) = (y, y_2 + [zy_1]) \quad (5)$$

By taking inverse lifting step of equation (5)

$$L_{z,[\cdot]}^{-1} = (y_1, y_2 + [zy_1]) \quad (6)$$

A lifting step dissipation for a matrix $\begin{bmatrix} z & b \\ c & d \end{bmatrix}$ with b not equal to 0 and the determinant 1 is shown below,

$$\begin{bmatrix} z & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ d-1 & 1 \end{bmatrix} \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ z-1 & 1 \end{bmatrix} \quad (7)$$

By equation (7) given rotation can be access as,

$$\begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} = [X][Y][Z] \quad (8)$$

$$[X] = \begin{bmatrix} 1 & \frac{\cos \alpha - 1}{\sin \alpha} \\ 0 & 1 \end{bmatrix} \quad (9)$$

$$[Y] = \begin{bmatrix} 1 & 0 \\ \sin \alpha & 1 \end{bmatrix} \quad (10)$$

$$[Z] = \begin{bmatrix} 1 & \frac{\cos \alpha - 1}{\sin \alpha} \\ 0 & 1 \end{bmatrix} \quad (11)$$

The Forward transform of Modified discrete cosine transform is given by,

$$X_K = \sqrt{\frac{2}{N}} \sum_{n=0}^{2N-1} x_n \cos \frac{\pi}{4N} \left(n + \frac{1}{2} + N \right) \left(K + \frac{1}{2} \right) \quad (12)$$

$$k = 0, 1, \dots, \frac{N}{2} - 1$$

Multiply matrix equation (9) with (12) results in (13) as follows,

$$X_K = \sqrt{\frac{2}{N}} \sum_{n=0}^{2N-1} x_n \cos \frac{\pi}{2N} \left(n + \frac{1}{2} + N \right) \left(K + \frac{1}{2} \right) \quad (13)$$

$$k = 0, 1, \dots, \frac{N}{2} - 1$$

Multiplying the matrix equation (10) with (13) and operating the rotation in π values, results in (14)

$$X_K = \sqrt{\frac{2}{N}} \sum_{n=0}^{2N-1} x_n \cos \frac{\pi}{N} \left(n + \frac{1}{2} + N \right) \left(K + \frac{1}{2} \right)$$

$$k = 0, 1, \dots, \frac{N}{2} - 1 \quad (14)$$

Multiply equation (11) with (14) results in (15)

$$X_K = \sqrt{\frac{2}{N}} \sum_{n=0}^{2N-1} x_n \cos \frac{2\pi}{N} \left(n + \frac{1}{2} + N \right) \left(K + \frac{1}{2} \right) \quad (15)$$

$$k = 0, 1, \dots, \frac{N}{2} - 1$$

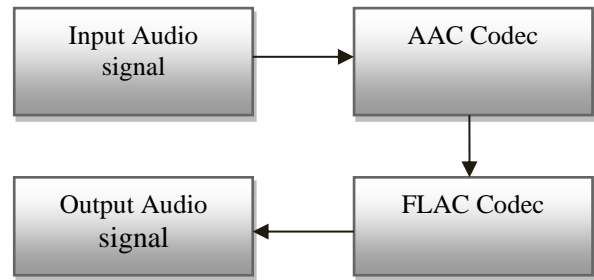
Thus the IntMDCT build upon the Lifting scheme/rotations act on MDCT, without introducing any artifacts the integer approximation can be inverted. So Integer MDCT retaining perfect reconstruction

The inverse transform is given as,

$$Y_n = \sqrt{\frac{2}{N}} \sum_{n=0}^{2N-1} x_k \cos \frac{2\pi}{N} (2n + 1 + N) (K + 1) \quad (16)$$

$$n = 0, 1, \dots, N - 1$$

III. DESIGN FLOW



The above figure has the following steps, Input audio signal with the sampling frequency of 44.1 KHz and it is taken in .wav format, then the signal is fed toward Adaptive Audio Coding(AAC) audio codec which found in encoder and decoder. The reconstructed audio signal gets same as that of original signal. At that point the signal will be changed from time to recurrence domain using transform. INTMDCT is a altered version of MDCT. Output of AAC audio codec is added with the Free Lossless Audio Codec(FLAC). Finally the codec produces a compressed signal that is output audio signal.

A. Proposed AAC

AAC is an Adaptive Audio Coding legitimate for lossy propelled audio compression. Compared to MP3, AAC achieves higher sound quality at lower bitrate. As a part of MPEG-4 and MPEG-2 specification adaptive audio codec has been regulated by IEC and ISO. HE-AAC (High Efficiency Audio Coding) which is a part of AAC and MPEG-4. Audio is also accepted into DAR (Digital Radio Standards) like DAB+ (Digital Audio Broadcasting) and Digital Radio Mondiale (DRM). This audio codec supports 48 full BW upto 96 KHz that is channels in one stream with 16 low frequency effects [19] up to 16 data streams. High accuracy for transient signals uses a block size of 120 or 128 samples.

ENCODING PROCEDURES

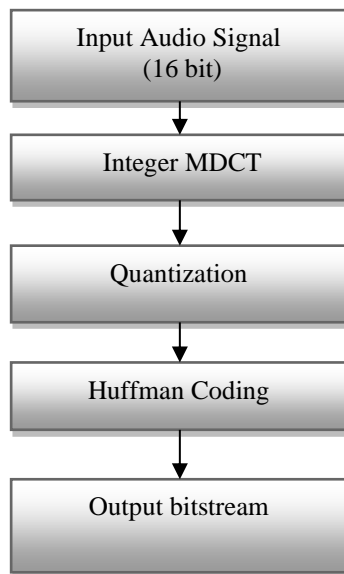


Fig.1. Block diagram of Encoding process of AAC

In AAC audio codec, provide input in the form of samples. To find the frequency components use Integer MDCT as a filter. Find the power coefficients i.e. envelope of the Integer MDCT and quantize using scale factor based upon the coefficients. Then based upon the codebooks the coefficients are separated, and using Huffman coding code the coefficients as bitstreams.

DECODING PROCEDURES

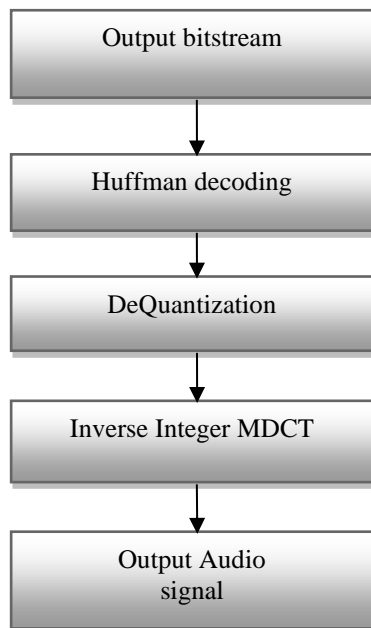


Fig. 2. Block diagram of Encoding process of AAC

Using Huffman coding the bitstreams are coded back into power coefficients then based on codebook the coefficients are merged together. Dequantize the coefficients using scale factor as step measure. To convert frequency component into time domain component use Inverse Integer MDCT. Finally reconstruct the output in the form of samples .

B. Proposed FLAC

Free Lossless Audio Codec (FLAC) is a sound pressure format as MP3. It is a lossy compression, it can be easily discard the audio information. Therefore human can't able to hear the clear audio but FLAC is a lossless audio codec, without any distortion the audio signal will be compressed with better sound audibility that is if we slit an CD track to a FLAC file, that file will maintain the same quality of the original like .wav format. Compared to lossy audio codec, FLAC will take only less amount of space i.e. better compression. It is a most generally utilized lossless audio codec.

ENCODING PROCEDURES

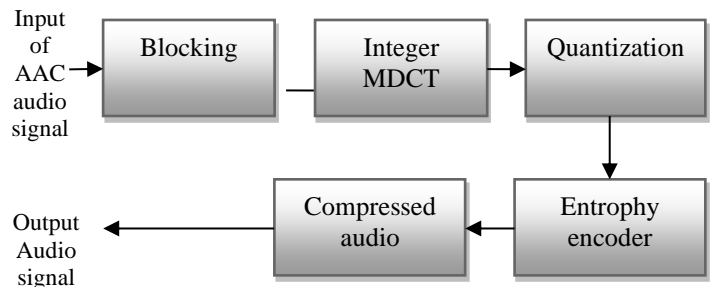


Fig. 3. Block diagram of Encoding process of FLAC

AAC audio codec output is taken as the input of FLAC. During blocking stage the data is splitted into 8*8 blocks of pixels, and block of frame is encoded separately, for example Mono audio signal is splitted into blocks with each having one subblock and a stereo audio signal has left, right, and center channels. Then FLAC performs Interchannel decorrelation which has multiple channels. The number of bits per sample is diminished. Thus the audio signal is packed with less bitrate

STEPS IN DECODING PROCESS

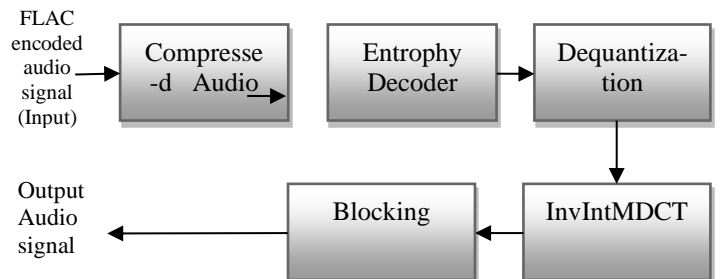


Fig. 4. Block diagram of Decoding process of FLAC

FLAC encoded audio signal output is taken as the input of Decoding process that is number of output bitstream in compressed audio is decoded using entrophy decoder. Then the

numbers of samples are quantized. Finally the signals are separated into block of pixels. By taking Inverse Integer MDCT the audio signal is compressed with minimum number of bitrate and it is reduced with high fidelity music.

IV. PERFORMANCE MEASURE

Execution Measure gives the investigation between the original and the reconstructed sound signal. In this paper two types of execution are evaluated that is Sound Pressure Level(SPL) and Compression Ratio(CR) test i.e. Subjective and Objective testing.

A. Sound Pressure Level(SPL) Test

To represent Sound pressure of a sound relative to reference pressure, sound Pressure Level uses a logarithmic scale and it is measured in terms of decibels(dB). At the threshold of human hearing is 0 dB, and it has same as reference pressure i.e. 2×10^{-5} and SPL is denoted as L_p . shown in eqn(17)

$$SPL = 20 \log \left(\frac{p}{p_{ref}} \right) \quad (17)$$

Where 'p' is the Sound Pressure of Original/Compressed signal and p_{ref} Sound pressure of the reference signal.

V. EVALUATION OF AUDIO QUALITY

Audio quality can be tested using two types of testing process that is subjective and objective testing. Objective testing means evaluating process and subjective testing mean listening test.

A. Subjective Test

It is a listening test and it can be done using MOS(Mean Opinion Square) and ABx test. Here the original signal and the reconstructed signal are played to the listener so that the listener can identify the quality of the audio signal.

1. ABx Test

It is a method of comparing choices of sensory stimuli to identify and find the difference between them. Sample A be the first reference, Sample B be the second reference. Hence, X is selected inconstantly selected A or B and find the difference between them.

2. Mean Opinion Score

It is an subjective estimation, audience would listen the melody and see the sound quality and score it. The MOS is expressed as a number of 1 to 5. Where 5 is a highest perceived audio quality, 1 is the lowest perceived audio quality, 4 is perceptible but annoying, 3 is 'Fair' that is slightly annoying, 2 is annoying, and 1 is very annoying.

In Objective testing PEAQ measurement method recommended International Telecommunication union. PEAQ measures the ODG(Objective Difference Grade) between the original and the reconstructed signal. In Objective testing five level impairment for Perceptual Evaluation of Audio Quality (PEAQ). The table are shown below,

TABLE I
MEAN OPINION SCORE

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly Annoying
2	Poor	Annoying
1	Bad	Very Annoying

TABLE II
FIVE LEVEL GRADE FOR PEAQ TEST

LEVEL	ODG
0	Imperceptible
-1	Perceptible but not annoying
-2	Slightly annoying
-3	Annoying
-4	Very Annoying

VI. RESULTS & DISCUSSIONS

To perform audio coding a set of original audio files are selected. The audio dataset are shown below in Table 3. The audio dataset has a similar length and same arrangement of sampling frequency of about 44.1KHz. inside term of < 1min and 1 min. Input audio signals are in .wav format. It is passed through the AAC audio encoder and then by applying transform some of the coefficients are produced. Hence, AAC works only on single channel.

The audio files of 16bit are converted to 8 bit and then the AAC audio signal is compressed and it is fed into FLAC audio codec, use INTMDCT as a filter to find the envelope of power coefficients and based upon the coefficients quantize using Scale factor. Separate the quantised coefficients using Huffman codebook and the output bitstream will be obtained then decode the encoded audio signal.

Table III demonstrates Audio database with same duration. In Fig:5,7 shows the Input Audio Signal and Fig. 6,8 shows the compressed audio signal combined with both the audio codec(AAC+FLAC). Then the compressed audio signal is compressed from MB to KB or the compressed ratio is decreased. While combining two audio codec the signal provides good compression ratio.

TABLE III
AUDIO DATASET WITH SAME TIME DURATION

File Name	Duration	File Size
1.	59sec	5.32 MB
2.	59sec	5.32 MB
3.	59sec	5.32 MB
4.	59sec	5.32MB
5.	59sec	5.32 MB
6.	59sec	5.32MB

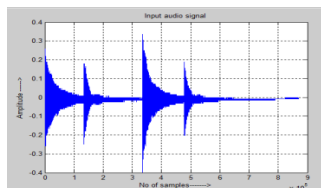


Fig. 5. Input Signal(1.)

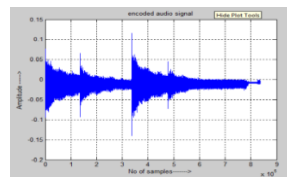


Fig. 6. Output of Combined signal(AAC+FLAC)

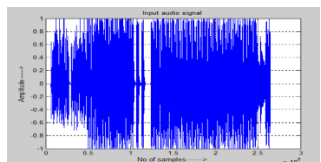


Fig. 7. Input Signal(2.)

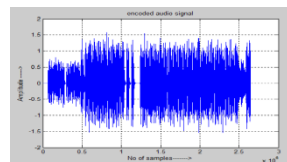


Fig. 8. Output of Combined signal(AAC+FLAC)

TABLE IV
AUDIO DATASET AFTER COMBINING TWO AUDIO CODEC'S (INTMDCT)

File Name	Duration	File Size
1.	58sec	120 KB
2.	56sec	125 KB
3.	57sec	121 KB
4.	56sec	115 KB
5.	56sec	121 KB
6.	57sec	122 KB

TABLE V
AUDIO DATASET AFTER COMBINING TWO AUDIO CODEC'S (MDCT)

File Name	Duration	File Size
1.	49sec	143KB
2.	51sec	142 KB
3.	48sec	162 KB
4.	36sec	143 KB
5.	56sec	138 KB
6.	43sec	131 KB

VII. PERFORMANCE MEASURE

A. Compression Ratio

Compression ratio is calculated using the following formulas, that is defined as the ratio of original signal to the compressed signal.

$$CR = \frac{\text{uncompressed signal}}{\text{compressed signal}} \quad (18)$$

TABLE VI
COMPRESSION RATIO OBTAINED BY AAC+FLAC CODEC (INTMDCT)

S.No	Original Size	Compressed Size	Compression Ratio
1.	5.32 MB	120 KB	44:1
2.	5.32 MB	125 KB	43:1
3.	5.32 MB	121 KB	44:1
4.	5.32MB	115 KB	46:1
5.	5.32 MB	121 KB	44:1
6.	5.32MB	122 KB	44:1

TABLE VII
COMPRESSION RATIO OBTAINED BY TWO AUDIO CODEC'S (MDCT)

File Name	Original Size	Compressed Size	Compression Ratio
1.	5.32 MB	143KB	37:1
2.	5.32 MB	142 KB	37:1
3.	5.32 MB	162 KB	33:1
4.	5.32MB	143 KB	37:1
5.	5.32 MB	138 KB	39:1
6.	5.32MB	131 KB	40:1

TABLE VIII
COMPARISON TABLE FOR MDCT AND INTMDCT

S.No	MDCT	INTMDCT
1.	37:1	44:1
2.	40:1	43:1
3.	33:1	44:1
4.	37:1	46:1
5.	39:1	44:1
6.	40:1	44:1

Hence, by comparing MDCT transform with INTMDCT transform, Integer MDCT has better compression ratio compared with MDCT. Where maximum CR for INTMDCT is 46:1 and minimum is 43:1.

TABLE IX
SOUND PRESSURE LEVEL COMBINED BY TWO AUDIO CODEC'S

S.No	Original Signal(.wav)	Compressed Signal (AAC+FLAC)
1.	77.861	78.181
2.	79.192	79.568
3.	88.306	89.491
4.	83.712	85.818
5.	76.432	80.5691
6.	73.231	74.842

C. Evaluation Test : Evaluation Test has both Objective and Subjective evaluation is used to measure the audio quality.

❖ Mean Opinion Score

❖ ABx Test

❖ PEAQ Test(Perceptual Evaluation of Audio Quality)

1. Mean Opinion Score: It is an listening test,based upon the audio quality listener has to give their opinion .It has to grades from 5 to 1

2. ABx Test: Listener should find out the randomly selected sample.Depending upon this result is evaluated.

TABLE XI
ABX TEST RESULTS

S.NO	Identified(YES/NO)
Listener 1	YES
Listener 2	NO
Listener 3	YES
Listener 4	NO
Listener 4	NO

TABLE XII
PEAQ RESULTS

File Name	ODG Level	Impairment
1.wav	-2.135	Perceptible but not annoying
2.wav	-1.918	Perceptible but not annoying
3.wav	-1.123	Slightly annoying
4.wav	-1.932	Slightly annoying
5.wav	-1.413	Slightly annoying

In ABx Evaluation 40 number of listeners are taken.Original signal and the Compressed audio signal is played to the listened ,if they had identified correctly it is denoted as 'YES' and if they had not identified the difference of orginal and the compressed audio signal that is deonoted as'NO'.Out of 40 listeners 16 have identified the difference and 24 have not identified the difference between the samples.

3. PEAQ Test In Perceptual Evaluation of Audio Quality, ODG shows the difference between the Original and the compressed audio signal are shown in below table as five level impairment.

Table13 shows by measuring the original sample and the compressed sample contains a slight difference between them. Hence, the audio quality of the signal with minimum bitrate and with better audio quality.

VIII. CONCLUSION

In audio coding and audio signal analysis, IntMDCT is for the most part utilized in audio compression.Audio codec uses Integer MDCT transform to compress an audio signal with diminished bit rate but with finer audio quality and two codec's of AAC and FLAC that is lossy and lossless audio codec is used.The variance between the original and the compressed audio signal is measured by means of Compression Ratio and Sound Pressure Level (SPL).Maximum compression obtained by the combination of two different audio codec is 47:1.From the full test,it is concluded that AAC and FLAC audio codec described using IntMDCT and MDCT.Compared to MDCT,Integer MDCT provides better sound quality.Memory space basic for compressed signal is less contrasted with uncompressed signal.

REFERENCES

- [1] Ralf Geiger, Jurgen Herre, et al., "Int-MDCT- a link between perceptual and lossless Audio Coding", IEEE Trans. Sig. Process., vol.52, no.9, Oct 2004.
- [2] Yoshikazu Yokotani, Ralf Geiger, Gerald Schuller, Soontorn Oraintara and K. R. Rao, "Lossless Audio Coding Using the Int MDCT and Rounding Error Shaping" IEEE Transactions on Speech and Audio Processing, submitted: October 2004, revised: july 2005.
- [3] Ralf Geiger, Thomas Sporer, Jurgen Koller, Karlheinz Brandenburg, "Audio Coding based on Integer Transforms".
- [4] Yoshikazu Yokotani and Soontorn Oraintara, "Lossless Audio Compression Using Integer Modified Discrete Cosine Transform" International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2003) Awaji Island, Japan, December 7-10, 2003.
- [5] Ravi K. Chivukula, Yuriy A. Reznik, Senior Member, et.al.,"Fast Algorithms for Low-Delay TDAC Filterbanks in MPEG-4 AAC-ELD", IEEE Trans. Audio, Speech, Lang. Process., vol.22, no.12, Dec 2014.
- [6] Dissertation zur Erlangung des akademischen Grades Doktor Ingenieur, "Audio Coding Based on Integer Transforms".
- [7] Yoshikazu Yokotani, Ralf Geiger, et al.," Lossless audio coding using the Int-MDCT and Rounding Error Shaping", IEEE Trans. Signal. Process, vol.16, no.2, Feb.2006.
- [8] Huang HaiBin, RAHARDJA Susanto, YU Rongshan, LIN Xiao, "A Fast Algorithm of Integer MDCT for Lossless Audio Coding".

- [9] Rongshan Yu, Susanto Rahardja, et al. "A Fine Granular Scalable to Lossless Audio Coder", IEEE Trans. Audio, Speech, Lang. Process, vol. 14, no.4, Dec. 2006.
- [10] Ralf Geiger, Jurgen Herre, et al., "Fine grain scalable perceptual and lossless audio coding based on Int-MDCT", IEEE Trans. Signal. Process., vol.16, no.2, Feb.2005.
- [11] Te Li, Susanto Rahardja, et al., "On Integer MDCT for Perceptual Audio Coding", IEEE Trans. Audio, Speech, Lang. Process., vol.15, no.8, Nov. 2007.
- [12] KARLHEINZ BRANDENBURG, "MP3 AND AAC EXPLAINED", Fraunhofer Institute for Integrated Circuits FhG-IIS A, Erlangen, Germany.
- [13] M. T. Ali, M. Saleem, "Improved Audio Quality at 48 Kbits/s for MPEG-4 AAC"
- [14] Rajesh Kumar, "Comparison of Subjective and Objective Speech Quality Assessment for Different Degradation and Noise Conditions", International Conference on Signal Processing and Communication, vol. 14 March 2015, pp 261-266.
- [15] X. Maitre, "7 kHz audio coding within 64 kbits/s" IEEE Journal on Selected Areas in Communication Vol.6 no.2, February 1988.
- [16] P. Kabal, "An Examination and Interpretation of ITU-R BS: 1387 Perceptual Evaluation of Audio Quality", Telecommunication and Signal Processing Laboratory.
- [17] Jürgen Herre and Martin Dietz, "Mpeg-4 High-Efficiency AAC coding"
- [18] Rainer Huber and Birger Kollmeier, PEMO-Q—A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception, IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, no. 6, November 2006