How to Cite:
Raj, V. A., & Dhas, M. D. K. (2022). Analysis of audio signal using various transforms for enhanced audio processing. *International Journal of Health Sciences*, *6*(S2), 12989–13001. https://doi.org/10.53730/ijhs.v6nS2.8890

Analysis of audio signal using various transforms for enhanced audio processing

Mr. V. Arun Raj
Department of ECE, Mepco Schlenk Engineering College (Autonomous), Sivakasi
Corresponding author email: arunraj@mepcoeng.ac.in

Dr. M. Davidson Kamala Dhas
Department of ECE, Mepco Schlenk Engineering College (Autonomous), Sivakasi
Email: mdavid@mepcoeng.ac.in

***Abstract*---**Audio Signals are the portrayal of sounds. It changes with respect to frequencies rather than time, and it shows more information in the frequency domain. So it is much appropriate to evaluate in the frequency domain rather than the time domain. By using different transforms like DFT, DST, DCT, MDCT, Integer MDCT, the time domain audio signal can be converted into a frequency domain signal. The signal is reconstructed to analyze the features like mean square error, Signal to noise ratio, Peak signal to noise ratio between the original and reconstructed signal. Other features like energy, entropy, zero crossing rates (ZCR) were also considered for the evaluation. In this paper, different audio file formats were taken for interpretation. It includes wave file, mp3 file, m4a file, aac file, where wave file is in uncompressed format and mp3, m4a, aac are in compressed format. These compressed files come under lossy compression. The above-mentioned features are used for applications like music information retrieval (MIR). MIR includes onset detection, pitch detection and to measure the noise and loudness of the music.

***Keywords*---**discrete fourier transform (DFT), discrete sine transform (DST), discrete cosine transform, modified discrete cosine transform (MDCT), integer modified discrete cosine transform.

Introduction

Representation of sound is called as audio signals. The Frequency range of audio signals is 20 Hz to 20 kHz. Nowadays, the audio contents are available through a more number of channels. The computation process of the audio signal will be reduced to a certain extent if the audio signal is represented in the frequency domain. To extract useful information available in the audio signal with plenty of

data needs a system that is capable of analyzing, evaluating and processing the audio content present in the signal. In both vibrating and pure tones, Fourier transforms persists and sinusoids a major part in musical sounds. By using the butterfly structure, amplitude, and phase information can't be exploited. But in this structure, change in one bit might change all the bits at the time of computation. Due to the properties of Discrete Cosine Transform, audio signals are considered as sparse signals in the spectral domain.

For Discrete Fourier Transform, a small amount of overlap between adjacent frames is abundant. To remove the artifacts, the maximum overlap is required. Hence, 50% overlap is needed for Modified Discrete Cosine Transform (MDCT) and Integer MDCT. For representing and handling sounds, a parametric test has a major effect. The quality and nature of the audio signal depending on the performance metrics.

Block Diagram

Audio signal processing is a subdivision of signal processing. Preprocessing steps are mandatory in audio signal processing. Framing and windowing are the necessary preprocessing steps in audio signals. The time domain signal is converted into a frequency domain signal with the help of transforms. It includes DFT, DST, DCT, MDCT, INT-MDCT. Features can be extracted both in the time domain and frequency domain. The signal should be reconstructed to compare its parameters with the original signal. To reconstruct the signal, an inverse transform must be applied where the frequency domain signal is again converted into a time domain signal. It is followed by re-windowing and de-framing. The obtained result will be the reconstructed signal of the input.

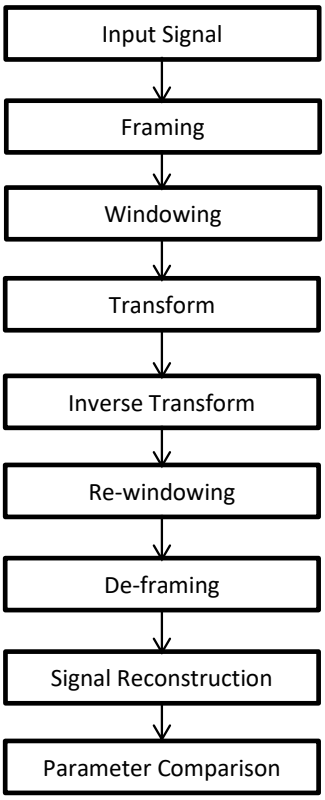

Fig.1 Flow diagram

Preprocessing Steps

Framing

Audio Signal is a non-stationary signal. Within a window of 20ms, it seems to be a stationary signal. So, it is necessary to cut down the signal into N number of frames. This is called as framing. Since the signal is divided into a number of frames, it is easy to process and analyze the signal. It is essential to consider the overlapping between the frames. If framing is done without overlapping, there will be a loss of information in the signal. For analysis, we have divided the signal with the frame length of 256.

Windowing

The process of taking a small subset from a large signal or from a large subset is called as windowing. The window function is a type of mathematical function i.e., it is zero-valued outside the selected interval, it maximum at middle and symmetric around the middle of the interval. Since the edges of the framed signal have some harmonics, windowing also used to tear down the edges of the framed signal. We have used different windows for our analysis. It includes a Hamming

window, a Hanning window, a Sine window, a Kaiser window, a rectangular window, and trapezoidal window.

Hamming window

Hamming window is used to remove the large side lobes. By optimization, side lobes which are near to the main lobe are strongly shaped in this window.
The mathematical expression of the Hamming window is given as,

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \qquad (1)$$

In equation (1), n denotes the sample values varies from 0 to (N-1)

Sine window

Because of its spectral property and reconstruction property, the sine window is basically used before the implementation of transforms. Autocorrelation of the sine window gives a function called as Bohman window. With its simpler design, it gives good frequency domain behavior.
Sine window is expressed as,

$$w(n) = \sin\left(\pi\left(\frac{n+0.5}{N}\right)\right) \qquad (2)$$

Implementation of Transforms

The necessity of transforming a signal from time domain to frequency domain is to get direct visibility of signal characteristics and for easy extraction. Initially, based on symmetric property and phase information property DFT is used for audio signal processing. The coefficients of DCT and DST are real positive integers whereas DFT coefficients are real and complex. So, DCT is effective for audio signal processing. MDCT and Int-MDCT have better reconstruction.

Discrete Fourier Transform

Discrete Fourier Transform is a type of discrete transform. It converts a discrete time data sets into a discrete frequency representation with the same length. Here, the input should be a finite sequence. Usually, FFT (Fast Fourier Transform) is used for the implementation of DFT. FFT is based on the divide and conquer method. In DFT, the input and the output coefficients are complex.
The mathematical expression for DFT is given as

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-j2\pi kn}{N}} \qquad (3)$$

In equation (3), x (n) denotes the input signal with N samples; X (K) is the frequency domain representation of the input. Inverse DFT has to be applied to obtain the original signal. It is expressed as

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{\frac{j2\pi kn}{N}} \qquad (4)$$

Discrete Sine Transform

DST explicit finite input data sequence into a sum of sine functions fluctuating at various frequencies. DST is related to discrete transforms, but it differs from DFT by using only a real matrix. The mathematical expression for DCT is given as,

$$X(k) = \sum_{n=0}^{N-1} x(n)\sin(\frac{\pi}{N}(n+\frac{1}{2})(K+1)) \qquad (5)$$

In order to obtain the original signal inverse DST is used. It is expressed as

$$x(n) = \sqrt{\frac{2}{N}}\sum_{k=0}^{N-1} X(k)\sin(\frac{\pi}{N}(n+\frac{1}{2})(K+1)) \qquad (6)$$

In equations (3) and (4), k denotes the transformed coefficient number and n denotes the coefficients in the time domain.

Discrete Cosine Transform

DCT explicit finite input data sequence into a sum of cosine functions fluctuating at various frequencies. DCT is related to discrete transforms, but it differs from DFT by using only real numbers. The most general alternative for DCT is type-II DCT. The mathematical expression for DCT is given as,

$$X(k) = \sum_{n=0}^{N-1} x(n)\cos(\frac{\pi}{N}(n+\frac{1}{2})K) \qquad (7)$$

Where x (n) is the input with N samples is transformed into X (k) in the frequency domain with N samples. The inverse DCT is given as,

$$x(n) = \sqrt{\frac{2}{N}}\sum_{k=0}^{N-1} X(k)\cos(\frac{\pi}{N}(n+\frac{1}{2})K) \qquad (8)$$

In equations (7) and (8), k denotes the transformed coefficient number and n denotes the coefficients in the time domain.

Modified Discrete Cosine Transform

MDCT is a type of lapped transform. It is based on the DCT IV algorithm. In addition, it is added with the property of being lapped i.e., It performs on the successive blocks of an enormous dataset. Since it is a lapped transform it overlaps between the adjacent blocks with a 50% overlap. MDCT is expressed as

$$X(k) = \sum_{n=0}^{2N-1} x(n)\cos(\frac{\pi}{N}(n+\frac{1}{2}+\frac{N}{2})(K+\frac{1}{2})) \qquad (9)$$

Where x (n) is the input with N samples is transformed into X (k) in frequency domain with N samples. The inverse MDCT is expressed as

$$x(n) = \frac{2}{N}\sum_{k=0}^{2N-1} X(k)\cos(\frac{\pi}{N}(n+\frac{1}{2}+\frac{N}{2})(K+\frac{1}{2})) \qquad (10)$$

In equations (9) and (10), k denotes the transformed coefficient number and n denotes the coefficients in time domain.

Integer modified Cosine Transform

Integer modified cosine transform is implemented to remove the Time Domain Aliasing Cancellation (TDAC). Since it is derived from the MDCT, most of the properties of the MDCT are inherited by Int-MDCT. It is done by carrying out integer approximation on the obtained result of the MDCT. Integer approximation is performed by rounding or lifting scheme operations.

Audio Parameters

From the audio signal, the parameters are extracted for our analysis. Parameters are taken to ensure the perfect reconstruction of the signal. Following are parameters that we have taken

Mean Square Error

Mean square error is also called as mean square deviation. The error is calculated between the original and reconstructed signal. MSE measures the average of the squared error.

$$MSE = \frac{1}{N}\sum_{i=0}^{N}(x_i - y_i)^2$$

Where x is the input signal and y is the reconstructed signal. N denotes the number of samples in the input signal.

Energy

To measure the loudness of the audio energy is used. For a good audio signal, energy must be high. It is calculated by the following equation,

$$E(i) = \sum_{n-1}^{L} \left| x_i(n) \right|^2$$

Here, L is the length of the frame. It indicates that energy is dependent on the length of the frame.

Power

It is defined as the normalization of the energy by the length of the frame. It is expressed as

$$P(i) = \frac{1}{L}\sum_{n-1}^{L} \left| x_i(n) \right|^2$$

Entropy

The measure of abrupt change in the energy of an audio signal is called as Entropy i.e., the amount of input energy that is unavailable for the work. Entropy must be low for a signal. The following equation is used to compute the entropy.

$$H(i) = \sum_{j-1}^{k} e_j log_2(e_j)$$

Here, $e_{j=\frac{E_{subframe\ j}}{\sum_{k-1}^{K} E_{subframe\ k}}}$, j varies from 1….k. Here the input signal is divided into number of sub frames.

Zero Crossing Rate

The zero crossing rate (ZCR) is defined as the rate of sign changes of an audio signal. It is used to measure the noise present in the signal. It is given by

$$Z(i) = \frac{1}{2L}\sum_{n-1}^{L} \left| sgn[x_i(n)] - sgn[x_i(n-1)] \right|$$

Where $sgn[x_i(n)] = \begin{cases} 1, x_i(n) \geq 0 \\ 0, x_i(n) < 0 \end{cases}$

Results and Discussions

From the representation of the given block diagram as shown in Fig.1 Initially, an audio signal x(n) of .wav file format, with a better sampling rate is fed as an input which is shown in Fig.2

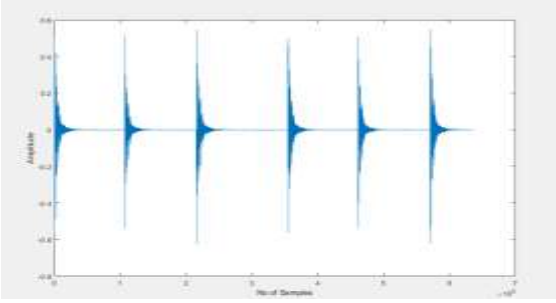

Fig.2 Input Signal

The input audio signal x (n) is subjected to framing. The audio signal after framing is represented in Fig 3.

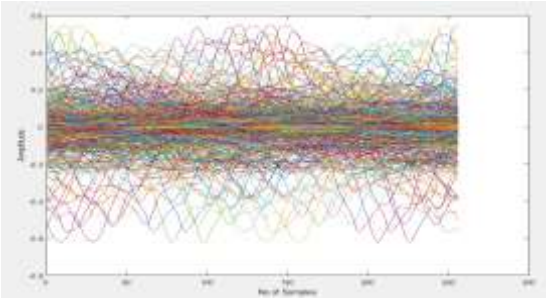

Fig.3 Framed Signal

The framed audio signal is subjected to windowing i.e., each frame is multiplied by the Hamming window in DFT. This is represented in Fig. 4

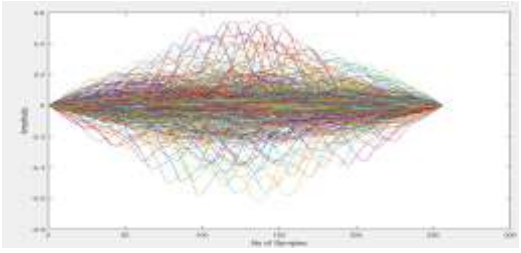

Fig.4

After the process of framing and windowing, the transform is applied and the transformed output of DFT is shown in Fig. 5

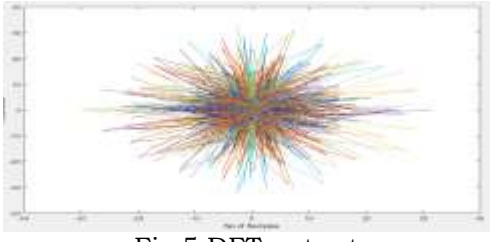

Fig.5 DFT output

Similarly, the transformed output of DST is represented in Fig.6

Fig.6 DST output

The following figure Fig.7 shows the transformed output of DCT

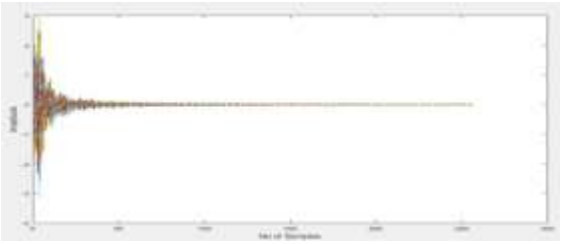

Fig.7 DCT output

The Int-MDCT representation of the input signal is given in Fig. 8

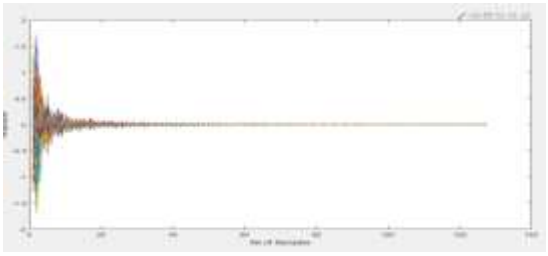

Fig. 8 MDCT output

The transformed output is subjected to Inverse transform and it is represented in Fig. 9

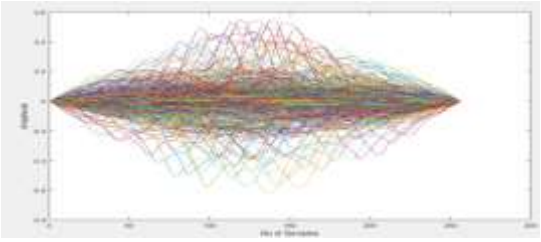

Fig.9 Inverse Transform

The inverse transformed is further subjected to windowing and it is represented in Fig. 10

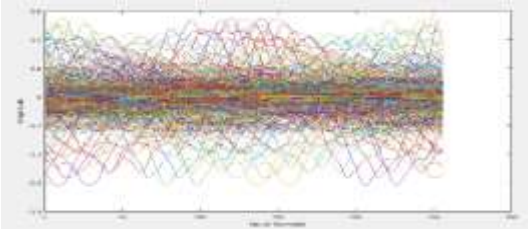

Fig. 10 Re-windowing

The windowed is de-framed to recover the original audio signal. It is given in Fig.11

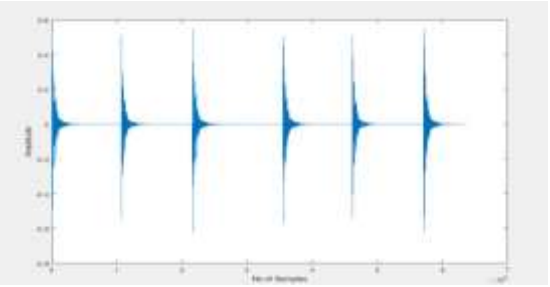

Fig. 11 Reconstructed signal

Similarly, the transformed signal was reconstructed for the other transforms (DST, DCT, MDCT, Integer MDCT) of different audio file formats. Table 1 gives the input energy, input power, input ZCR, and input entropy values for different file formats which include wave file, MP3 file, M4A file, and AAC file

Table I : Comparison of various audio formats

File Format	Input energy	Input power	Input ZCR	Input Entropy
.wav	944.09	0.3792	0.0229	7.7690
.mp3	854.01	0.3422	0.0178	8.1299
.m4a	944.09	0.3792	0.0228	7.7690
.aac	940.87	0.3769	0.0234	17.6984

Table II clearly depicts the Mean Square Error, zero crossing rate and entropy for several transforms of different audio formats.

Table II: Comparison of Audio Transforms

File Type	Transform	MSE	ZCR	SNR
.wav	DFT	6.887e-05	0.029	54.51
	DST	1.558e-34	0.031	64.78
	DCT	1.152e-34	0.028	65.04
	MDCT	1.046e-34	0.029	65.08
	IntMDCT	1.028e-34	0.033	65.14
.mp3	DFT	5.796e-05	0.037	55.99
	DST	1.449e-34	0.039	64.80
	DCT	1.045e-34	0.036	65.08
	MDCT	1.102e-34	0.035	65.04
	IntMDCT	8.759e-35	0.039	65.24
.m4a	DFT	6.910e-05	0.029	54.91
	DST	1.558e-34	0.031	64.78
	DCT	1.152e-34	0.028	65.04
	MDCT	1.102e-34	0.029	65.08

	IntMDCT	1.028e-34	0.031	65.14
.aac	DFT	6.535e-05	0.023	55.37
	DST	1.553e-34	0.023	64.78
	DCT	1.159e-34	0.023	65.03
	MDCT	1.102e-34	0.023	65.08
	IntMDCT	1.011e-34	0.023	65.15

Table III: Audio Transforms in terms of Power & Entropy

File Type	Transform	Energy	Power	Entropy
.wav	DFT	880.5	0.353	12.126
	DST	892.0	0.358	11.092
	DCT	901.8	0.362	10.636
	MDCT	912.3	0.366	10.986
	IntMDCT	917.5	0.368	10.975
.mp3	DFT	791.7	0.317	15.754
	DST	803.3	0.321	11.545
	DCT	812.7	0.325	11.055
	MDCT	822.9	0.329	11.378
	IntMDCT	828.1	0.331	11.370
.m4a	DFT	878.6	0.352	15.449
	DST	892.0	0.358	11.092
	DCT	901.8	0.362	10.636
	MDCT	912.3	0.366	10.975
	IntMDCT	917.3	0.368	10.825
aac	DFT	876.8	0.351	19.286
	DST	888.7	0.356	19.283
	DCT	898.5	0.359	19.282
	MDCT	909.2	0.364	19.282
	IntMDCT	910.4	0.366	19.264

Table III clearly shows the parameters such as energy, power, and entropy of the reconstructed signal. It is inferred that there is some error occurs between the original and reconstructed signal and the calculated error was minimum in Int-MDCTSince zero Crossing Rate has slight or negligible changes; it has minor impacts in transforms. Entropy gives the position of abrupt change in the signal.

Conclusion

From our analysis, Int-MDCT has a minimum error and better outcomes when compared to other transforms. Hence, it can be used for audio signal processing to get the best results. The impact of transforms on both the compressed and uncompressed audio signals remains the same. The performance of DCT is nearly equal to that of MDCT because of the usage of the sine window in DCT which is popularly used in MDCT.

The perfect reconstruction of an audio signal is possible in Integer MDCT since it is a real orthogonal lapped transform with a 50% overlap between adjacent blocks. Thus, the parameters which are extracted using Integer MDCT can be used for various audio processing applications like pitch detection, detection of Onset in music signals, etc.

Future Work

The extracted parameters can be used for music information retrieval. This work can also be extended to music instrument recognition and discrimination of speech and music.

References

1. Theodoros Giannakopoulos, Aggelos Pikrakis ,“Introduction to Audio Analysis: A MATLAB Approach,” Academic press, 2014
2. Emmanuel Ravelli, Gaël Richardand Laurent Daudet “Audio Signal Representations for Indexing in the Transform Domain,” IEEE Transactions on audio, speech, and language processing, vol. 18, no. 3, March 2010
3. Sylvain Marchand, "Fourier-based methods for the spectral analysis of musical sounds," Signal processing conference (EUSIPCO), 2013 proceedings of the 21st european , vol., no., pp.1,5, 9-13 September. 2013
4. R.G. Moreno-Alvarado, Mauricio Martinez-Garcia,” DCT-compressive Sampling of Frequencysparse Audio Signals,” Proceedings of the World Congress on Engineering 2011 vol II, wce 2011, July 6 - 8, 2011, London, U.K.
5. Shuhua Zhang, Weibei Dou, Huazhong Yang, "MDCT Sinusoidal Analysis for Audio Signals Analysis and Processing," Audio, speech, and language processing, IEEE Transactions on , vol.21, no.7, pp.1403,1414, July 2013
6. Dominique Fourer ,Sylvain Marchand, “Informed spectral analysis: audio signal parameter estimation using side information,” EURASIP Journal on Advances in Signal Processing, December 2013
7. R. R. Coifman, Y. Meyer, and V. Wickerhauser, “Wavelet analysis and signal processing,” in In Wavelets and their Applications. Citeseer,1992.
8. Vladimir Britnak, Pratnik Yip, Kamisetty R.Rao, “Discrete Cosine and Sine Transforms :General Properties, Fast Algorithms and Integer Approximations” Academic press, 2007
9. H.Malvar, “A Modulated Complex Lapped Transform and its Applications to Audio Processing,” in Proc.IEEE Int. Conf.Acoust.,Speech,Signal Process.(ICASSP ’99),March 1999, vol.3, pp.1421-1424.
10. C.Cheng, ”Method for estimating magnitude and phase in the MDCT domain,” in Proc. 116th AES Conv.,May 2004,pp.6091-6091,Audio Eng. Soc.
11. Mu-Huo Cheng and Yu-Hsin Hsu, “Fast IMDCT and MDCT Algorithms— A Matrix Approach,” IEEE Transactions on signal processing, vol. 51, no. 1, January 2003
12. Yaroslavsky, L., & Wang, Y., “ DFT, DCT, MDCT, DST and signal Fourier spectrum analysis,” EUPSICO 2000: European signal processing conference, pp. 1065-1068.

13. Yoshikazu Yokotani, Member IEEE, Ralf Geiger, Member IEEE, Gerald D.T.Schuller, Senior Member, IEEE & K.R.Rao,, “Lossless Audio Coding using the Int MDCT & the round error shaping”, IEEE trans. on Audio,Speech & Language Processing, Vol.14, No.6, Nov 2006.
14. Rongshan Yu, Member, IEEE, Susanto Rahardja, Lin Xiao and Chi Chung Ko, Senior Member, IEEE, “A Fine Granular Scalable to Lossless Audio Coder”, IEEE Trans. On Audio, Speech and Language Processing, Vol.14. No.4. July2006
15. Te Li, Student Member IEEE, Rongshan Yu, Member IEEE, Susanto Rahardja, Member IEEE, Soo Ngee Koh, Member IEEE, “On integer MDCT for Perceptual Audio Coding”, IEEE trans. on Audio,Speech & Language Processing, Vol.15, No.8, Nov 2007.