AUDIO COMPRESSION USING DCT

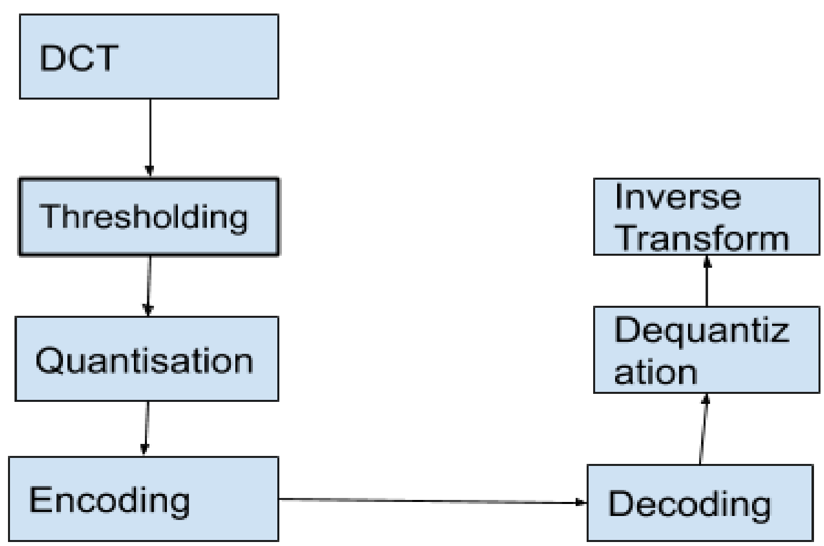
-Keshav Bansal

INTRODUCTION

The main objective is to represent the signals in a lesser number of bits. The reduction should be done in a way such that the audio is acceptable and almost all the audio is retained and it can later be decoded into the same original signal. The amount of data to be transmitted needs to be minimised. After transforming the signal using DCT it is encoded using Run length encoding and Huffman Compression techniques. Performance Analysis is done on the basis of Peak Signal to Noise Ratio (PSNR) and Normalized Root Mean Square Error (NRMSE) .

TRANSFORM CODING

First, the original signal can be transformed using Discrete Wavelet Transform and Discrete Cosine Transform. This doesn't compress it but it gives information about the signal and it can then be compressed after removing the lesser conspicuous signals and after quantisation using encoding techniques like Huffman and Run length encoding.



Discrete Cosine Transform

Using DCT(discrete cosine transform), reconstruction is very easy. This property makes DCT suitable for compression.

DCT of 1-D sequence x (n) of length N is given by:

$$X(m)= \frac{\sqrt{2}}{\sqrt{N}} C_m \sum_{m=0}^{N-1} x(n) \cos\left[\frac{(2n+1)m\Pi}{2N}\right]$$

Where m= 0,1,2,......N-1

The Inverse DCT is given by:

$$X(n)= \frac{\sqrt{2}}{\sqrt{N}} \sum_{m=0}^{N-1} Cm\, x(m) \cos\left[\frac{(2n+1)m\Pi}{2N}\right]$$

The output vector is twice as long as the DFT vector.

Thresholding

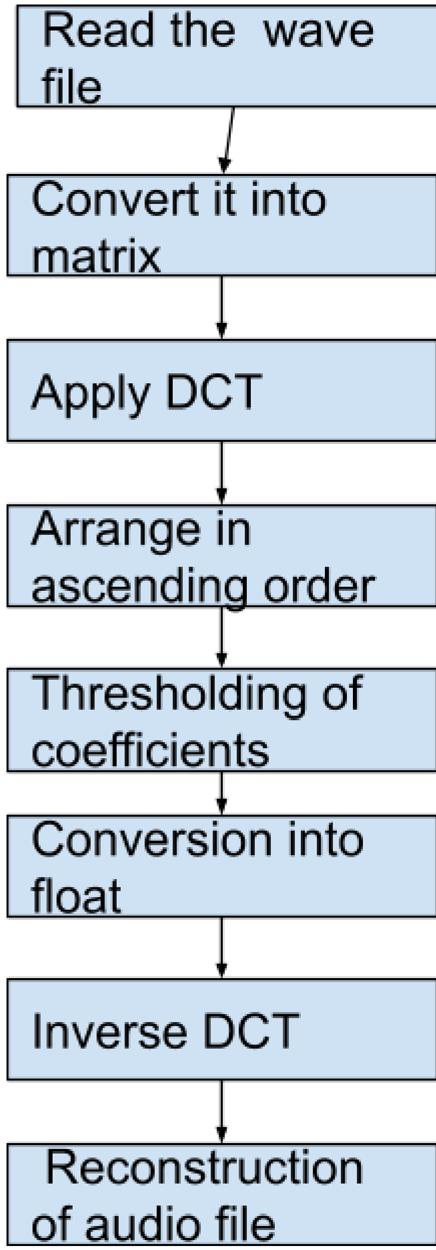
After the coefficients are received from different transforms, thresholding is done. Very few DCT coefficients represent 99% of signal energy; hence Thresholding is calculated and applied to the coefficients. Coefficients having values less than threshold values are removed.

Quantization

It is a process of mapping a set of continuous valued data to a set of discrete valued data. The aim of quantization is to reduce the information found in threshold coefficients. This process makes sure that it produces minimum errors.

Encoding

We use different encoding techniques like Run Length Encoding and Huffman Encoding. Encoding method is used to remove data that are repetitively occurring. In encoding we can also reduce the number of coefficients by removing the redundant data. Encoding can use any of the two compression techniques, lossless or lossy. This helps in reducing the bandwidth of the signal hence compression can be achieved. The compressed speech signal can be reconstructed to form the original signal by decoding followed by dequantization and then performing the Inverse transform methods. This would reproduce the original signal.



Performance Analysis

1. **Compression Factor**

$$CF = \frac{\textit{Length of original signal}}{\textit{Length of compressed signal}}$$

2. Peak Signal To Noise Ratio

$$PSNR = 10 \log_{10} \frac{NX^2}{\|X - X'\|^2}$$

Where N is the length of reconstructed signal, X is the maximum absolute square value of signal x and ||x-x`||2 is the energy of the difference between the original and reconstructed signal.

3. Normalized Root Mean square error

$$NRMSE = \sqrt{\frac{\sum_N (X(n) - X'(n))^2}{\sum_N (X(n) - u_x(n))^2}}$$

Here, X(n) is the speech signal, x'(n) is a reconstructed speech signal and $\mu_x(n)$ is the mean of speech signal.