The modified discrete cosine transform (MDCT)

The MDCT is a variant of the DCT used for audio coding. It takes $2N$ samples $(x_0, \dots, x_{2N-1})$ as input and produces N coefficients $(X_0, \dots, X_{N-1})$ as output:

$$X_k = \sum_{n=0}^{2N-1} x_n w_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right] \qquad k \in [0, N-1]$$

where w_n is a necessary edge attenuation window, with appropriate properties. For example, the function is often used:

$$w_n = \sin\left[\frac{\pi}{2N}\left(n + \frac{1}{2}\right)\right]$$

The inverse transform is obtained by:

$$y_n = \frac{2}{N} w_n \sum_{k=0}^{N-1} X_k \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right] \qquad n \in [0, 2N-1]$$

The definition refers to real numbers, so in the implementation you can use the double type, but the input samples will be signed 16-bit integers. To then quantize the coefficients, these can be stored in signed 32-bit integers.

The inverse transform cannot reconstruct the original $2N$ values, but by superimposing the input sample windows at 50%, it is possible to add (always with 50% overlap) the reconstructed values, obtaining the original signal (with any very small errors due to the rounding).

To allow the data of the first and last window to be transformed and then reconstructed, the input signal is extended with a window of N zeros before and after.

Exercise 1

You are given a file “test.raw”, with audio samples of a mono signal stored as little endian 16 bits signed integers:

1. Measure the entropy of the original signal
2. Quantize the samples by dividing them by a Q value (use 2600).
3. Measure the entropy of the quantized signal
4. Rebuild the data by de-quantizing and saving it (output_qt.raw)
5. Calculate the error at each sample as the difference from the original signal and save it (error_qt.raw)

At this point you can use Audacity by loading the dequantized and the error signals: listening to them together you have the original signal, listening to them separately you can hear the result of the quantization or the error.

6. Run the MDCT transform of the input file, with a window of N = 1024 samples
7. Quantize the coefficients by dividing them by a Q value (use 10000).
8. Measure the entropy of the quantized coefficients
9. Rebuild the data by de-quantizing
10. Anti-transform the reconstructed coefficients and save the obtained signal (output.raw)
11. Calculate the error committed to each sample as a difference from the original signal and save it (error.raw)

At this point you should have verified that the entropy of the two quantized signals is practically the same. Which of the two is better? Why?