Introduction to Lossless Audio Encoding

Lossless Encoding is a method of encoding that allows no loss of data while converting to and from encoded data to normal data.

Lossless Audio Encoding

Q: Why special encoding methods for audio?

A : Lossless encoding is not only done in audio, video etc. We actually have a ton of genric lossless algorithms like gzip, RAR etc. Which obviously can be used to compress audio as well, but we know that digital audio is in a specific wave form which is often used as an advantage to create these algorithms. So, in short "We use special encoding methods for audio because we can save more data.".

Audio Codecs specify a set of tested methods and instructions to follow, to actually implement lossless encoding. Any operations an encoder does should be well defined in the audio specifications and bit-exact definition should be provided so as to make sure that encoder follows what the codec intends it to.

Q : How is Encoding done?

A : First the audio is divided into further parts i.e frames which contains a certain number of samples. Then different encoding methods are employed on these frames which is what we'll discuss in this article.

Q : How many samples are there in a frame?

A : Actually it depends on the codec, Some codecs specify a fixed size but let us divide each frame into sub-frames, some codecs specify a set of possible sizes but we have to chose one size for the whole file and some codecs let us change the size in each frame.

Example :-

in Raw PCM the size is set to 1 i.e we'll use 1 sample per frame.

Encoding

There are a number of ways in which different codecs are doing lossess encoding and in this tutorial we'll try to cover the basics of some of them.

1. The encoder checks if the data has repetition of a specific byte whole over the frame, this the most simple case, and then the encoder has a way to trigger this signal i.e 'This frame is all value x'. if this isn't the case the encoder moves on to the next method
2. Then the encoder finds corellation between the different channels, for eg: in stereo the left and right channels have a lot of simmilarities, For stereo its typically called mid/side.
3. Linear Prediction

At the most rudimentry level Linear Prediction (LP), is to assume the sample is the same as the last one. The difference between the predicted sample and the actual sample becomes what we call the residual sample.

Types :-

o Linear Prediction with an order of 1
  This involves 1 coeficiant. This coeficiant can be 1,2 etc. If its 1 we simply miltiply the previous sample by 1 and subtract it from the current sample to get the residual sample, same goes for 2, 3 etc.
o Linear Prediction with and order of 2
  This involves 2 coeficiants. The algorithms can be deminsified in the following points :-
o First the algorithm checks for 2 samples ago, multiply it by the 2nd coeficiant and add it to the running total
o Then it checks for the previous sample, multiply it with the 1st coeficant and it to the running total.
o Thirdly, it subtracts the running total from the present sample to get the residual.

Q : What does the codec do with the intial samples i.e 1st sample in LP with an order of 1 and 1st and 2nd sample in LP with an order of 2?

A : Again that depends on the codec, FLAC encodes the first N samples as normal PCM (where N is the order of Linear Prediction Filter), and then it starts LP encoding from origion + maximum order (where maximum order is the order of samples for the LP encoding), but with ALS its totally different ALS actually does a progressive prediction at the beginning, so the first sample is Raw-PCM (0-order), 2nd sample uses 1st order , 3rd order uses 2nd order ... upto max order.

4. Pitch Prediction

In signals that are very tonal (samples look more like a sine wave), so they have almost same wavelengths and amplitudes (resemblence to a sine wave), in this case the sample 1 wavelength in past has a better resemblence and is likely to be a better predictor of the data, so instead of looking previous N samples we use these, the encoder can then have some functionality to analyse this data and see if this is beneficial and then usually codes the distance in the past and maybe some scale factor to compensate for changing loudness.

Some Lossless codecs also includes Checksum data to verify if the encoded audio is same as that of original. So, normally there is a large checksum (normally md5) of the original audio transmitted in the stream so that the user can check the decoded result.

Conclusion

Lossless encoding is a vast topic and this tutorial not even covers 50% of it, What we have tried in this tutorial is to learn some basic foundation of lossless encoding and how is it done.