Lecture 10: Audio Coding

Introduction

Topic of this lecture was audio coding. Audio coding is a way of representing various audio content for different purposes such as storing. Audio coding techniques can be classified into two main different types - lossless and lossy. I will give a brief overview of both approaches and provide explanations what their purpose is with an implementation example of decorrelation using polynomial predictors of various degrees.

Lossless Audio Coding

Lossless audio coding relies on reducing the size of audio signal using redundancy reduction. This means that these approaches are based on various statistical methods that can be used to store the data more compactly while not losing any information and therefore being able to perfectly recreate it afterwards. The process of "assembling" the signal back into its original form is referred to as decoding. Lossless audio coding has many different purposes such as archiving original recordings, studio grade recording, digital music distribution (e.g. various streaming services), etc. It mainly consists of three main steps - framing, decorrelation and entropy encoding. Framing involves dividing the audio signal into smaller time frames. Decorrelation ensures that the redundancy in the signal is removed. Main two approaches for decorrelation are linear predictive model which stores the predictor coefficients and error signal, and linear transform model which stores the transform coefficients and the error signal. Lastly, entropy encoding is performed to encode the signal efficiently, e.g. we want to use short code for common sample values and a longer code for more infrequent sample values.

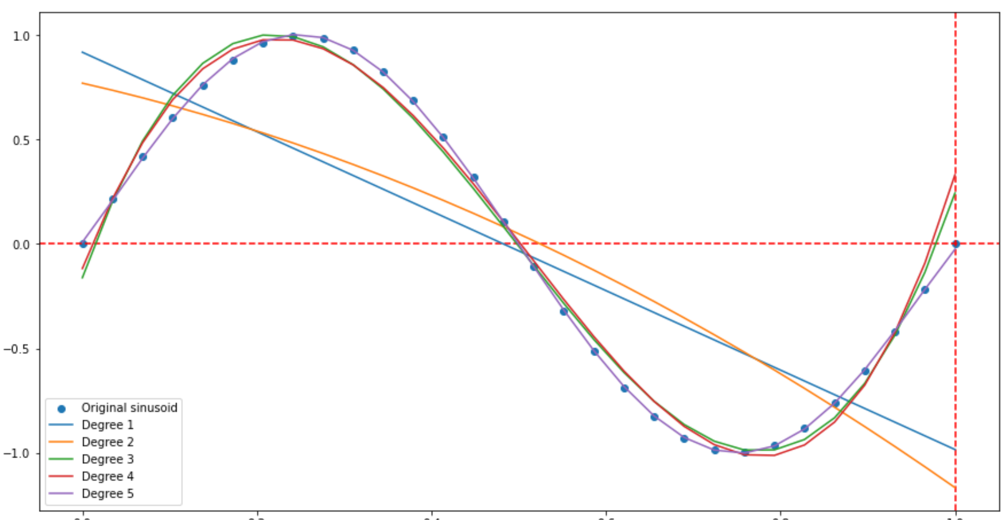

Figure 1: Example of polynomial prediction of a sinusoid signal sample using different polynome degrees.

Lossy Audio Coding

As the name suggests, contrary to lossless audio coding techniques, lossy techniques involve reducing audio signal's size with losses. As a result, lossy codecs can often achieve significantly higher compression ratios when compared to losselss. Most common use cases for lossy audio coding are smartphones, streaming services, digital radio, movie soundtracks, etc. Goal of lossy audio codecs is to encode audio in such way to maximize compression while simultaneously minimizing audible impact for the end user. This is achievable by utilizing various well-known psychoacoustic phenomena such as absolute threshold of hearing, masking phenomenon, critical bandwidth, etc. This approach is called subband audio coding, also known as perceptual audio coding. In particular, we can utilize knowledge about limits of human hearing (20 Hz to 20 kHz range with 1 - 5 kHz being the most sensitive) and use that to discard unnecessary frequencies. Another and most important psychoacoustic concept is masking - this means that presence of certain frequencies can impair our ability to hear and distinguish some other frequencies. This is commonly used as it allows us to discard certain frequencies while retaining the same, indistinguishable listening experience. We can visualize masking by utilizing masking curves. Frequency ranges around the masker frequency where the masking curve remains flat are called critical bandwidths and are measured in a unit called Bark. It is important to note that when using these techniques, SNR (signal-to-noise ratio) is not a good indicator as it is usually fairly high even though audio quality is perceived as good.