LEARNING DIARY – AUDIO PROCESSING

My experience after lecture 4,5 and 6.

In my personal observation, lecture 4 is quite practical and friendly. From lecture 4, I have studied "Dynamic Range Control" and "Equalizer". I had heard about dynamic range in imaging, but it was nice to see that concept in audio. I says it was friendly because I could find it easily during daily life, especially every time I use Zoom, there are configurations such as: automatically adjust microphone volume and suppress background noise. I was aware of the technique behind those options. However, the lecture helped me dig deeper into the problem and solutions.

For example, it was not just limit the input audio to prevent it from being too loud (high dB), or eliminate input audio if it is too quiet (considered as noise), there were different input regions in processing (Limiter, Compressor, Expander, Noise Gate). Each stage had its own coefficient to scale the input signal if the input is in its processing region (controlled by combination system).

Equalizer is much more common, often implemented in audio player app or equalizer device. From my understanding, equalizer filters the input signal with different frequency, then allow us to adjust the amplitude of it. IIR and FIR were introduced as techniques behind that. There was also peak filter used in graphic equalizer.

Lecture 5 explained different approaches for sound synthesis. Among those, I find FM the easiest to understand. Because shortwave radio is much better in travelling through long distance so, FM uses a carrier which is a shortwave radio to transmit the signal. The modulation signal is injected inside frequency of carrier. Both result output wave with the frequency vary through time (DC part is carrier freq and varying frequency part is the modulation signal with a scale called modulation index).

Other techniques are a little bit difficult, until I started working on the exercise, things became clearer. In those, we take every short sample of signal to analyze and reconstruct the signal in different ways. Then we can perform time stretching, pitch shifting. Neural audio generation uses deep learning to encode and decode signal. Output signal can be decided based on past signal samples. However, I find this part quite brief and I am not sure to understand it deeply.

What I find most interesting in lecture 6 is audio coding, which helped me understand the parameters of MP3 standard (128kbit/s and Fs 44100Hz) or different Bluetooth codec such as SBC, AAC, Qualcomm AptX or Sony LDAC. I have been using my Bluetooth headphone with AptX but until now I have knowledge how these parameters are important.

From my understanding, one important factor of audio coding is noise, distortion. Other methods are introduced: masking, perceptual audio coding. Quantization and its error were an easy part, because it is used in any purpose that requires ADC. Filter banks are used to separate input signal into multiple components, which then can be processed independently. Some of perceptual model were introduced. To be honest, this part contains quite amount of information to remember.