A multi-resolution sinusoidal model

Explanation and justification of the band edges and the window sizes

Sample #1: ASPMA orchestra (https://freesound.org/people/xserra/sounds/217543/)

The first sample is the orchestra sample from class. In the lows of this sound I see there are no well-defined attacks and the lowest sinusoids are around 100Hz. Since the time domain is not so important at the lows, I can take as big as a window to capture the low frequencies correctly. Higher up, in the mids, I see a lot of spectral information, but also some attacks are present here. So I will pick an all-round window size to capture a good frequency and time resolution. Lastly, in the highs, because of their nature, they are quickly changing frequencies, so I must have a smaller window, which is not a problem since we can capture a lot of detailed highs with smaller windows.
For this sound, I chose the bands to be $B_1 = (0, 1000)Hz$, $B_2 = (1000, 5000)Hz$, $B_3 = (5000, 22050)Hz$ and window sizes $M_1 = 4095$, $M_2 = 2047$, $M_3 = 1023$ with analysis FFT size N_i the highest power of 2 bigger than the corresponding M_i, thus $N_1 = 4096$, $N_2 = 2048$, $N_3 = 1024$.

Sample #2: Oriental orchestra (https://freesound.org/people/LawrieR/sounds/330955/)

The second sample is from an oriental orchestra, quite similar to the first orchestra, since in both samples, there are percussions and instruments of varying frequencies. In the lower spectrum, I see there are no well-defined attacks and the lowest sinusoids are around 125Hz. So, since I don't care so much about the time domain, I can take as big as a window to capture the fundamental and the rest of the low frequencies correctly. So my first band is $B_1 = (0, 500)Hz$. Higher up, in the mids at $B_2 = (500, 5000)Hz$, I see a lot of spectral information, but also some attacks are present here. So I will pick an all-round window size to capture a good frequency and time resolution. As with the first sound, I will further lower the window size for the 3rd band at $B_3 = (5000, 22050)Hz$. Respectively, I chose the window sizes to be $M_1 = 4095$, $M_2 = 2047$, $M_3 = 1023$ with the analysis FFT size N_i the highest power of 2 bigger than the corresponding M_i, thus $N_1 = 4096$, $N_2 = 2048$, $N_3 = 1024$.

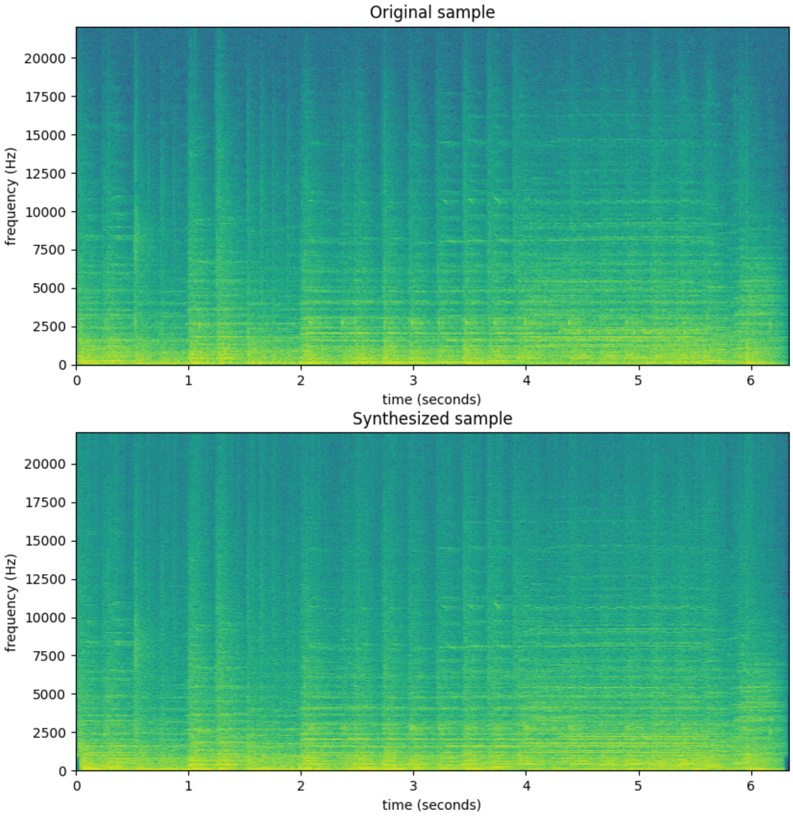
Observations about the advantages of a multi-resolution analysis

The advantage of a multi-resolution analysis is that by splitting the sound in different spectral bands, we can suppress the time-frequency tradeoff when choosing window sizes. Lower frequencies require higher window sizes to have more spectral resolution. As we go higher in the spectrum, we need less and less of a window size to have that resolution. As said in class, having a big window size, reduces the time resolution, which is crucial for the attacks of sounds. Since the attacks are mainly in higher frequencies, we can lower the window size in order to obtain this resolution without trading off the spectral resolution.

Challenges if to extend to HPR and HPS models

If we were to extend the model to the HPR and HPS, a challenge arises. Since the Harmonic model needs a fundamental frequency in order to generate the corresponding harmonics, if the fundamental frequency is in the lowest band, then the mids and highs would have to somehow know about this fundamental frequency. If the fundamental frequency was bouncing from lows to mids to lows and so on, then that would cause a lot of confusion on the algorithm.

Sample #1



Sample #2