1 Audio Extraction Technique

During this week I worked on understanding more how maqams are defined and how oriental music works and is structured. I also worked on understanding how to extract important information from the audio file.

1. Extracted the STFT of the audio file.
2. Applied filtering to analyze the notes being played.
3. Attempted to classify the Maqam by using the peaks visible in the FFT.

1.1 Frequency Isolation

Let Y be the floating-point time series extracted from the music sample.

$$F_{db} = \mathrm{STFT}(Y) \in \mathbb{R}^{m\times n} \tag{1}$$

To be able to manipulate frequencies and determine which notes or frequencies are being played, we need to isolate the strong frequencies from those that are either reverb or other noise.

$$\begin{aligned}
\theta_{\alpha,n} &: \mathbb{R} \longrightarrow \mathbb{R}, & \theta_{\alpha,n}(x) &= \begin{cases} x & \text{if } x > \alpha \text{ dB}, \\ 0 & \text{otherwise}, \end{cases} \\
\Psi_{\alpha} &: \mathbb{R}^n \longrightarrow \mathbb{R}^n, & \Psi_{\alpha}(X) &= \big(\theta_{\alpha,n}(x_i)\big)_{i=1}^{n}, \\
\delta^{\mu,\sigma}_{k,\varepsilon,n} &: \mathbb{R}^n \longrightarrow \mathbb{R}^n, & \delta^{\mu,\sigma}_{k,\varepsilon,n}(X) &= X * \mathcal{N}_{\varepsilon}(\mu,\sigma).
\end{aligned} \tag{2}$$

$$F'_{mod} = \delta^{\mu,\sigma}_{k,\varepsilon,n}(F_{db}) \tag{3}$$

To approach this problem I started by trying to classify on each ΔT the potential notes being played first of all to be able to later on feed it into a generative model to generate music and also to analyze the Maqam and the frequency of playing an interval in a specific Maqam this could be helpful to later create some kind of Markov chain of notes played that could help in a global identification of the Maqam

2 Potential Problems

This technique could be inaccurate because it gives too much weight to small ΔT intervals. In particular, the Oud produces continuous spectra of frequencies rather than discrete, separated notes. A slight finger movement produces a vibrato effect, which can alter the detected frequency and reduce accuracy.

2.1 Possible Solutions

One way to improve the robustness of the method is to compute a global Fourier transform (FFT) of the entire audio signal to identify the most dominant frequencies. These can then be assigned as the main notes.

Additionally, considering only the first 30% and last 30% of the piece may increase accuracy, since the middle often contains modulations that change the maqam being played. In this case, to find the Maqam we're in I analyzed the FFT and took the main notes played and the diffrence of tones to detect the Maqam and what is planned is to given these informations to use this spectrogram F_{db} to understand the notes played at each instant and later count the frequency of playing a note after a note

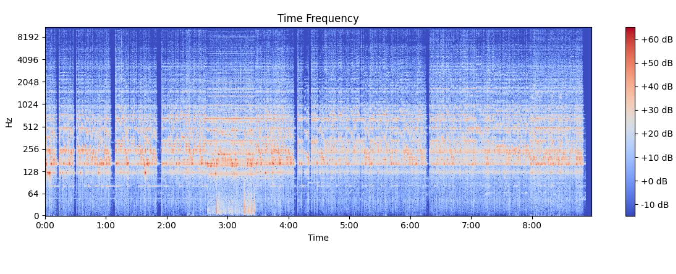

Figure 1: Spectrogram that I got

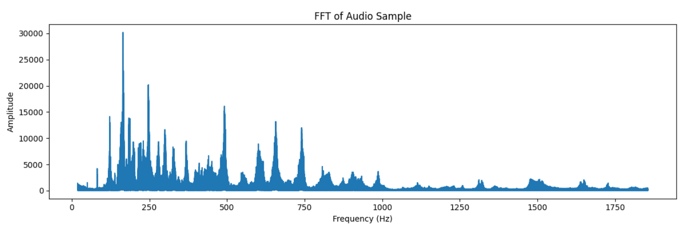

Figure 2: FFT of the whole piece