2.5. Acoustic features

The data for this project consists of audio cough recordings. These audio recordings are converted into spectrogram images so that image processing techniques can be applied. A spectrogram is an image that captures the frequencies and amplitudes of audio signals using a Discrete Fourier Transform (DFT). Figure 2.15 shows how an audio signal is converted into a spectrogram.

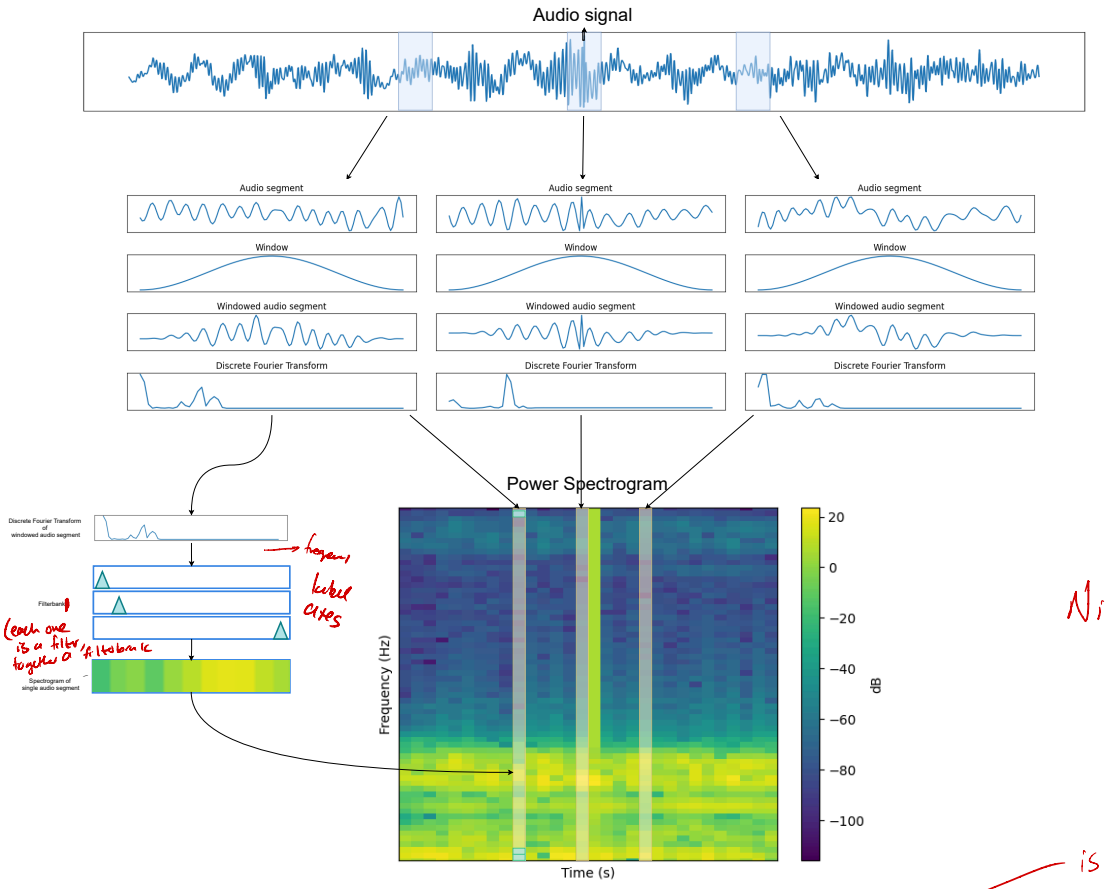

Figure 2.15: The process to convert a raw audio signal to a spectrogram. The audio signal is split into a segments. Each segment is multiplied with a window function. A Discrete Fourier Transform (DFT) is applied to each of the windowed segments. The DFT is multiplied with filterbanks and the result is stacked to form a spectrogram image.

The audio signal is divided into audio segments. A window is created using a windowing function such as a Hamming window. The audio segment is multiplied by the window to minimise the frequency leakage between overlapping signals. A Fast Fourier Transform is then applied to the windowed audio signal to indicate the probabilities that a certain frequency is present in that audio signal. Each FFT of the audio segments are then multiplied with filterbanks to summarise the information. Finally, the resulting values are then stacked together to form a contour plot with time on the x axis, frequency on the y axis and the magnitude (in decibels) on the z axis. The contour plot is shown as

a 2D image with the colours indicating the magnitude ~~of the~~ frequencies.

A mel spectrogram is a spectrogram where the linear ~~filterbanks are~~ replaced by mel ~~filterbanks~~ using the mel-scale as shown in Equation 2.20. The mel scale is used to mimic how humans perceive sounds. It is based on a logarithmic scale and is calculated such that each unit sounds equal in pitch difference.

$$m = 1127 \times \log(1 + \frac{f}{700}) \tag{2.20}$$

Mel-frequency cepstral coefficients (MFCCs) are ~~created~~ by applying a Discrete Cosine Transform (DCT) to the mel ~~spectrograms~~. MFCCs capture uncorrelated information from audio signals that are useful to machine learning models.

2.6. Conclusion

This chapter discussed various theoretical concepts needed to understand contrastive learning and the terminology used throughout this project. The next chapter is a literature review of the existing work done in the field of TB classification and contrastive learning.