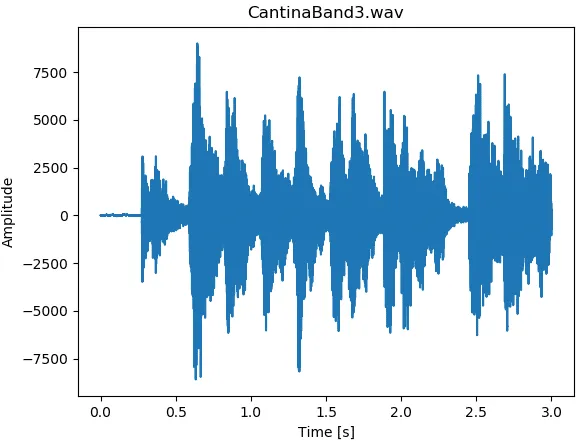
Audio signals can be represented in various ways, each offering different insights into the characteristics of the sound. Two common representations are the raw time series and time-frequency decompositions.

**Raw Time Series**

A raw time series is the **direct recording of the audio signal over time.** This representation shows how the **amplitude of the sound wave changes over time.** It is a one-dimensional signal where the **x-axis represents time**, and **the y-axis represents the amplitude of the sound wave**. Here are some key points:

**Amplitude**: The height of the wave at any given point in time, representing the sound pressure level. It is measured in Decibels (dB) and Sound pressure level (SPL)



* **Too Low**: Amplitude values that are too low may result in inaudible audio or audio with poor signal-to-noise ratio.
* **Too High**: Amplitude values that are too high may cause clipping, resulting in distortion. Clipping occurs when the amplitude exceeds the maximum value the system can handle.

**Sampling Rate**: The number of times an audio signal is measured (sampled) per second. It is typically measured in Hertz (Hz).

The sampling rate determines the resolution and quality of the digital audio. Higher sampling rates capture more detail from the original analog signal, resulting in better sound quality.

44.1 kHz It means 44,100 samples are taken per second, 48 kHz It means 48,000 samples are taken per second and so on. **Human Hearing Range**: Typically 20 Hz to 20,000 Hz (20 kHz).

### Time-Frequency Decompositions

Time-frequency decompositions provide a way to analyze how the frequency content of an audio signal evolves over time. Instead of just looking at the amplitude over time, these methods break down the signal into its constituent frequencies at each point in time. Common time-frequency decomposition methods include:

 \***Short-Time Fourier Transform (STFT)**: This technique divides the signal into short overlapping segments and applies the Fourier transform to each segment. The result is a spectrogram, a two-dimensional representation with time on one axis, frequency on the other, and intensity (amplitude) represented by color or grayscale.

 **Wavelet Transform**: This method uses wavelets (short oscillatory functions) to analyze the signal at different scales. It provides a multi-resolution analysis, which is useful for capturing both short-term and long-term patterns in the audio signal.

 **Mel-Frequency Cepstral Coefficients (MFCCs)**: Often used in speech processing, MFCCs are derived from the Fourier transform and are designed to approximate the human ear's response to different frequencies. They represent the short-term power spectrum of the sound.

**Frequency** is the rate at which a sound wave oscillates. It represents how many times the wave repeats (cycles) in one second. It is measured in Hertz (Hz), where 1 Hz equals one cycle per second. Frequency determines the **pitch** of a sound. Higher frequencies correspond to higher pitches (e.g., a whistle), while lower frequencies correspond to lower pitches (e.g., a drum beat).

**Applying CNNs to Spectrograms**

Convolutional Neural Networks (CNNs) are highly effective for image processing due to their ability to capture spatial hierarchies and patterns. Spectrograms, being 2D images, provide a natural domain for applying CNNs to audio processing. Here’s why and how:

1. **Spectrograms as Images**:
   * Since spectrograms are 2D representations with time and frequency as axes and intensity representing magnitude, they can be treated similarly to grayscale images where pixel intensity corresponds to magnitude.
2. **CNNs for Audio**:
   * **Feature Extraction**: CNNs can learn to identify patterns in the spectrogram that correspond to audio features, such as harmonics, pitch, and rhythms.
   * **Hierarchical Patterns**: Similar to how CNNs identify edges, textures, and objects in images, they can learn hierarchical patterns in spectrograms, from simple frequency components to complex sound structures.