Dhirubhai Ambani University
(Formerly known as DA-IICT)

Topic: Feature Extraction

Course: Programming Lab
Course Code- PC503

Dr. Ankit Vijayvargiya
Assistant Professor
Room No. 4205, Faculty Block 4
Email: ankit_Vijayvargiya[at]dau.ac.in
Phone: 079-68261628(O), 7877709590(M)

Dhirubhai Ambani University
Formerly known as DA-IICT

Feature Extraction

- Feature extraction transforms raw data into meaningful representations.
- It reduces dimensionality while preserving key information.
- Enables machine learning models to work effectively.
- Different approaches for 1D (signals) and 2D (images) data.
- The algorithm doesn’t learn representations automatically. Most traditional ML algorithms (like Linear Regression, SVM, Random Forests, XGBoost, k-NN, etc.) require meaningful, hand-crafted features as input.

Feature Extraction

When we need to extract the features

- Feature extraction is **necessary** when your raw data isn't directly usable by the algorithm, i.e., it needs to be converted into a numerical, structured form.

- **Examples by data type:**

  Images: You might extract features like color histograms, edges (using SIFT, HOG), or texture statistics before feeding them to the model.
  Text: You extract features using TF-IDF, Bag-of-Words, or word embeddings (like Word2Vec vectors).
  Audio: You compute features such as MFCCs, chroma features, or spectral contrast.
  Time series/signals: You calculate statistical features (mean, variance, autocorrelation, FFT coefficients, etc.).

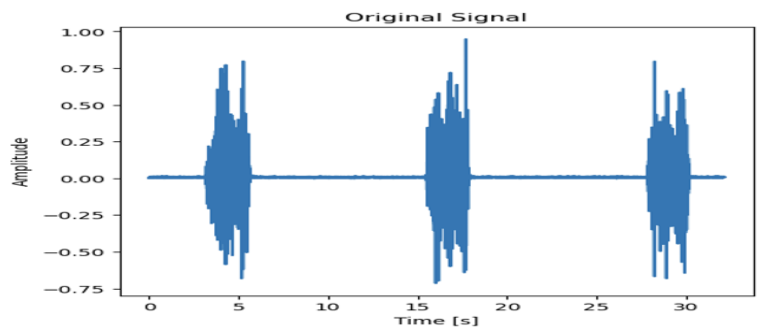
- **Reason: Algorithms like SVM or Random Forests can't "see" patterns in raw pixels or raw audio waves, they need those transformed, meaningful numbers.**

Dhirubhai Ambani University
Formerly known as DA-IICT

Feature Extraction

Feature Extraction with 1D data

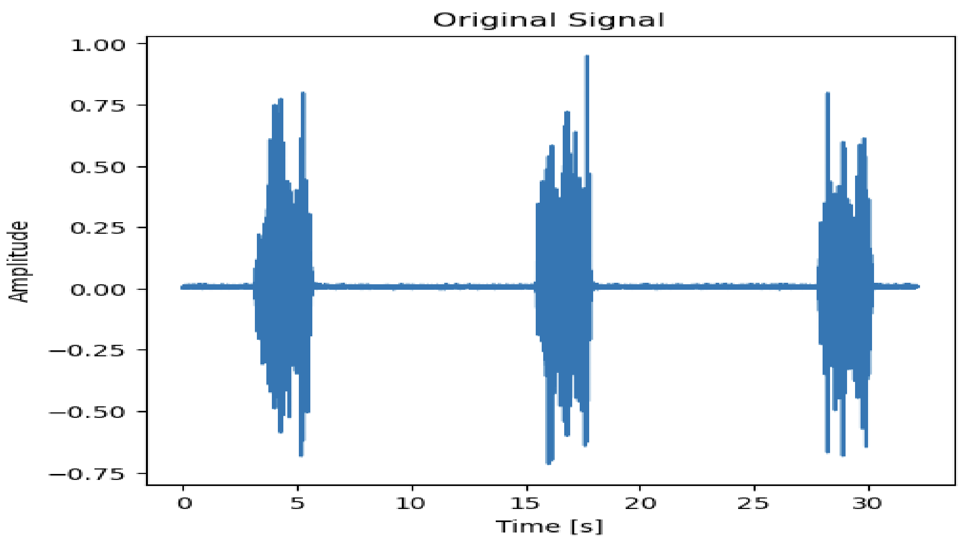
- After a 1D signal has been **segmented** into regions of interest (for example, time intervals or specific events), signal representation and description in a form suitable for further processing is very important.

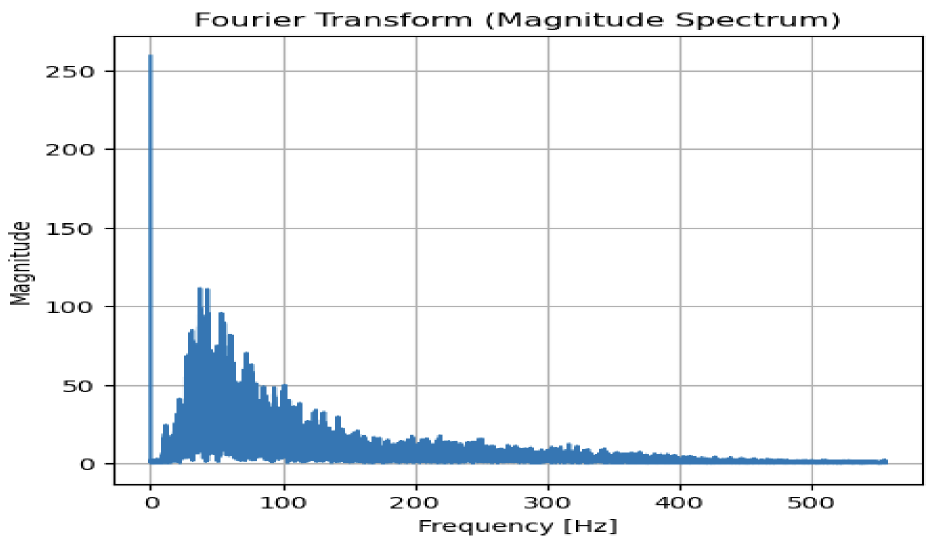


- Representing a segment can be done in two ways: (1) in terms of its external characteristics and (2) in terms of its internal characteristics
    - An external representation is chosen when the primary focus is on the overall shape or envelope of the signal, such as the waveform pattern, peaks, or slope changes.
    - An internal representation is preferred when the primary focus is on local properties of the segment, such as amplitude statistics, energy, frequency content, or time–frequency features.

Feature Extraction

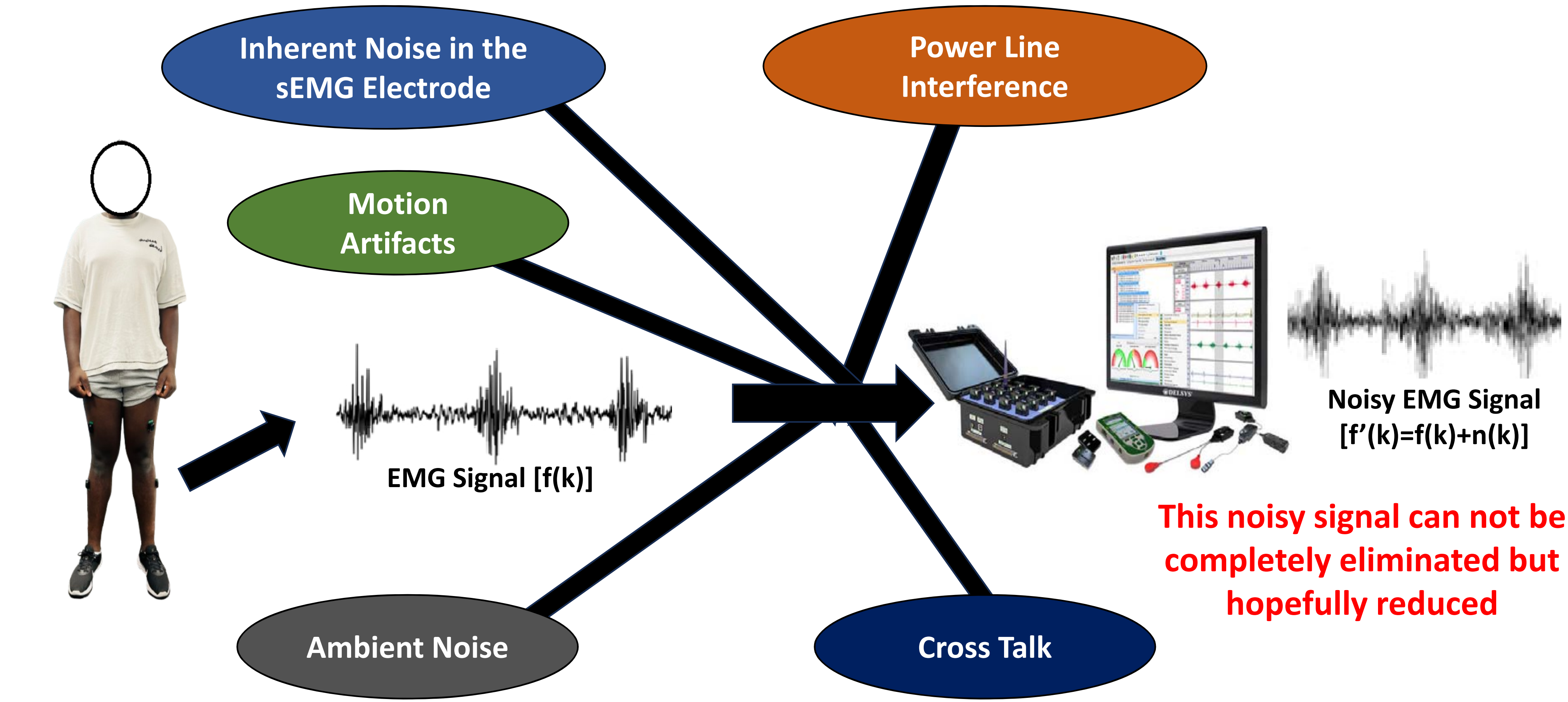
Feature Extraction in 1D data:





Dhirubhai Ambani University
Formerly known as DA-IICT

Denoising



DCU

Insight

Feature Extraction

Feature Extraction in 1D data:

Domain	Description	Common Features / Techniques	Examples / Applications
Time Domain	Uses the raw signal (amplitude vs. time). Captures temporal statistics and shape.	Mean, variance, RMS, zero-crossing rate, skewness, kurtosis, energy.	ECG signal analysis, vibration monitoring, speech energy features.
Frequency Domain	Represents signal in frequency space using transforms. Captures periodicities.	FFT (Fast Fourier Transform), Power Spectral Density (PSD), mean freq, median freq, spectral centroid, bandwidth.	Audio frequency analysis, fault diagnosis, EEG rhythm analysis.
Time–Frequency Domain	Shows how frequency content changes over time. Useful for non-stationary signals.	Short-Time Fourier Transform (STFT), Wavelet Transform, Continuous Wavelet Transform (CWT).	Speech spectrograms, transient detection, biomedical signal analysis.

Dhirubhai Ambani University
Formerly known as DA-IICT

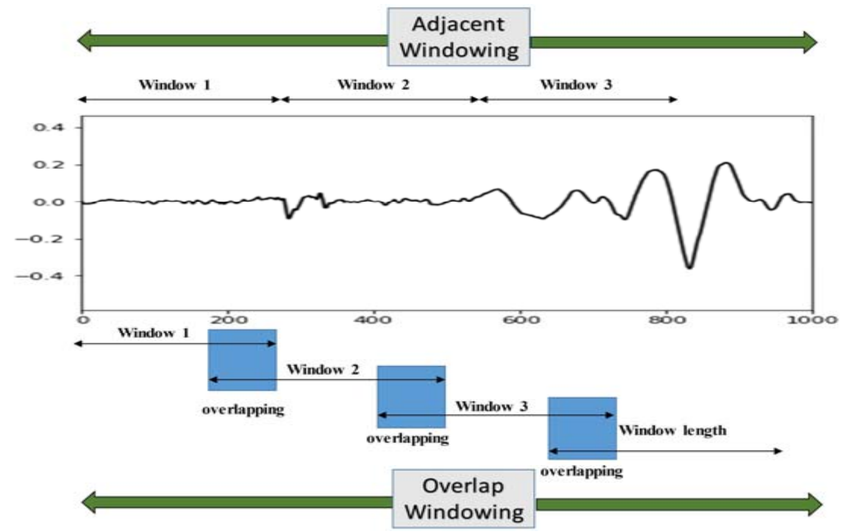
Feature Extraction

Signal Segmentation

- The nature of most signals is **stochastic**, meaning they exhibit random variations over time. Therefore, it is often preferable to **segment the signal into smaller regions of interest** after denoising, rather than processing the entire signal at once.

- To implement the segmentation process, the windowing method is utilized. **Two discrete methods of windowing** are present to accomplish the process of segmentation: **adjacent windowing and overlapping windowing.**

- Two critical factors to consider for data segmentation are the **windowing technique and the data length.**



Feature Extraction

Feature Extraction with 2D data (Images)

- After an image has been segmented into regions of interest, image representation and image description in a form suitable for further processing of the image is very important.

- Representing a region can be done in two ways. (1) in terms of its external characteristics and (2) in terms of its internal characteristics.

- An external representation is chosen when the primary focus is on shape analysis. An internal representation is preferred when the primary focus is on regional properties like color and texture.

Feature Extraction

Feature Extraction in 2D (images) data:

Domain	Description	Common Features / Techniques	Examples / Applications
Spatial (Time) Domain	Works on raw pixels and spatial gradients. Captures shape, edges, and intensity patterns.	HOG (Histogram of Oriented Gradients), LBP (Local Binary Patterns), Zoning, Moments (Hu, Zernike).	Object recognition, character recognition (MNIST), texture analysis.
Frequency Domain	Represents image in terms of spatial frequencies (global texture).	2D FFT, DCT, Gabor filters, Spectrum energy.	Image compression (JPEG), periodic texture analysis, filtering.
Space–Frequency (Time–Frequency Analogy)	Captures localized frequency information — both where and what frequencies.	2D Wavelet Transform, Gabor Wavelets, Multi-resolution analysis.	Medical imaging, texture classification, fault detection in images.

Dhirubhai Ambani University
Formerly known as DA-IICT

Dimensionality Reduction: