Brain Computer Interaction

Feature Extraction

Course Instructors

Dr. Sreeja S R

Assistant Professor

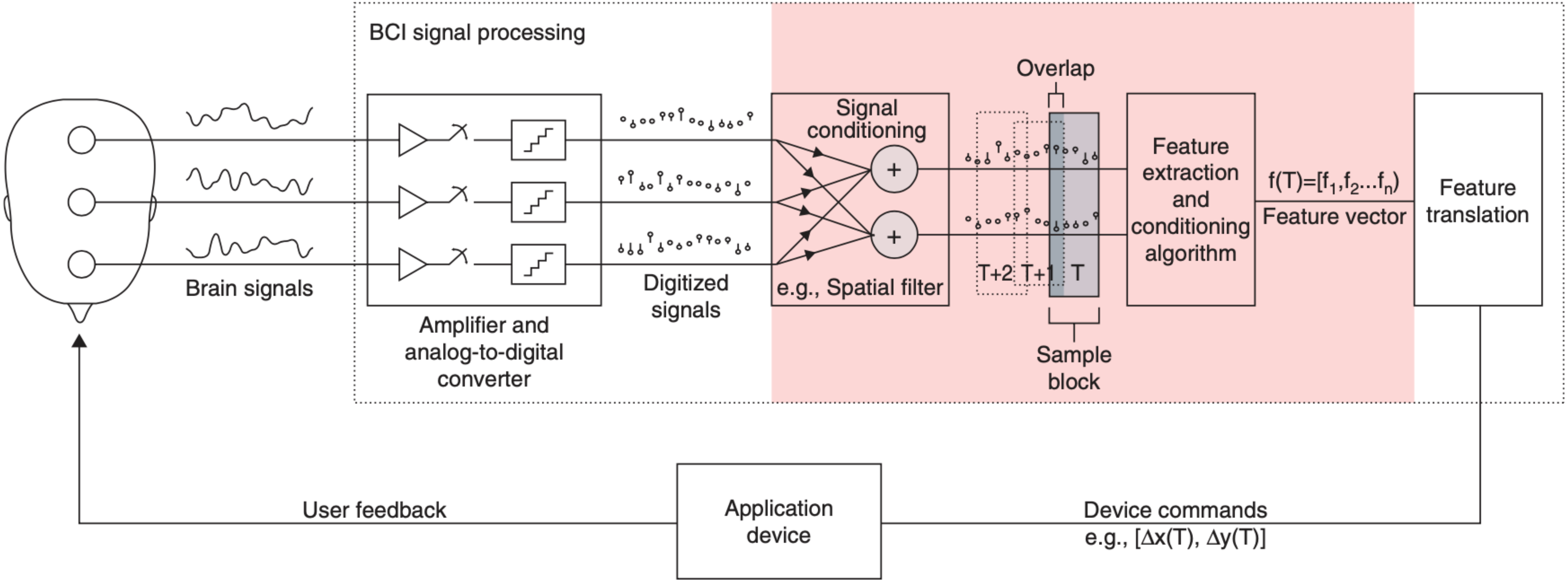
Indian Institute of Information Technology

IIIT Sri City

Features

- The purpose of a BCI is to **detect and quantify characteristics of brain signals** that indicate what the user wants the BCI to do, to translate these measurements in real time into the desired device commands, and to provide concurrent feedback to the user.
- The brain-signal characteristics used for this purpose are called ***signal features*, or simply *features***.
- ***Feature extraction*** is the process of distinguishing the pertinent signal characteristics from extraneous content and representing them in a compact and/or meaningful form, amenable to interpretation by a human or computer.

Overall structure of a BCI



Feature vector

- A **fundamental signal feature** is simply a direct measurement of the signal. They usually provide limited relevant information about typically complex brain signals.
- Thus, it is more common for BCIs to use features that are **linear or nonlinear combinations, ratios, statistical measures, or other transformations of multiple fundamental features** detected at multiple electrodes and/or multiple time points.
- Such complex features, if selected appropriately, can reflect the user’s desires more accurately than the fundamental features themselves.
- Most features used in BCI applications are based on **spatial, temporal, and/or spectral analyses** of brain signals or the relationships among them.
- Furthermore, in order to determine the user’s wishes as accurately as possible, most BCIs extract a number of features simultaneously. This set of features is referred to as a ***feature vector***.

Feature vector

To be effective for BCI applications, a feature should have the following attributes:

- its spatial, temporal, spectral characteristics, and dynamics can be precisely characterized for an individual user or population of users
- it can be modulated by the user and used in combination with other features to reliably convey the user's intent
- its correlation with the user's intent is stable over time and/or can be tracked in a consistent and reliable manner

BCI SIGNAL PROCESSING - Fourier Analysis

- Much of signal-processing theory is rooted in Fourier analysis, which transforms a time-domain (i.e., time on the x-axis) signal into its equivalent frequency-domain (i.e., frequency on the -axis) representation.
- The primary utility of Fourier analysis is to decompose a signal into individual sinusoidal components that can be isolated and evaluated independently.
- Using Fourier analysis, practically any signal can be accurately represented as the sum of a number (possibly an infinite number) of amplitude-scaled and time-shifted sinusoids at specific frequencies.
- In order to model these signals, it is necessary to properly adjust the phase and the magnitude of each sinusoid.

Fourier Analysis

- For an arbitrary signal x(t), the magnitude (scale) and phase (shift) of the sinusoid at each frequency [ω(radians) = 2πf (Hz)] required to represent an arbitrary signal can be determined from the Fourier transform:

$$
\begin{aligned}
X(\omega) &= \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt = \int_{-\infty}^{\infty} x(t)[\cos\omega t + j\sin\omega t]dt \\
&= \underbrace{\int_{-\infty}^{\infty} x(t)\cos\omega t dt}_{a(\omega)} + j\underbrace{\int_{-\infty}^{\infty} x(t)\sin\omega t dt}_{b(\omega)} \\
&= a(\omega) + jb(\omega)
\end{aligned}
$$

The magnitude and phase for each sinusoidal component are given as:

$$\textit{Magnitude}: \quad |X(\omega)| = \sqrt{a^2(\omega) + b^2(\omega)}$$

$$\textit{Phase}: \quad \theta = \arg(X(\omega)) = \tan^{-1}\left(\frac{b(\omega)}{a(\omega)}\right)$$

THE THREE STEPS OF FEATURE EXTRACTION

The process of feature extraction is discussed here as a three-step procedure:

- signal conditioning to reduce noise and to enhance relevant aspects of the signals
- extraction of the features from the conditioned signals
- feature conditioning to properly prepare the feature vector for the feature-translation stage

FIRST STEP: SIGNAL CONDITIONING

- The first step of feature extraction is called signal conditioning or preprocessing.
- This step enhances the signal by preemptively eliminating known interference (i.e., artifacts) or irrelevant information, and/or by enhancing spatial, spectral, or temporal characteristics of the signal that are particularly relevant to the application.
- It is common to have some prior knowledge about the general signal characteristics relevant for a particular application, and this knowledge is used in conditioning.

FIRST STEP: SIGNAL CONDITIONING

Signal conditioning can include a number of different procedures that can primarily be categorized as:

- frequency-range prefiltering
- data transformation and normalization
- spatial filtering
  - Data independent spatial filtering (*common- average reference* and *surface Laplacian spatial filters)*
  - Data dependent spatial filtering (PCA, ICA and CSP)
- removal of environmental interference and biological artifacts

Frequency-Range prefiltering

- Digital filters are central to digital signal processing. They modify the frequency content of a digital signal by attenuating some frequencies (or frequency ranges) and amplifying others. Each successive sample of a digitized signal is passed through a digital filter to produce a new value as the output.

  - Low pass filter (higher frequencies are attenuated and lower frequencies are preserved)
  - High pass filter (very specific ranges of signal frequencies can be amplified, attenuated, preserved, and/or eliminated)
  - Bandpass filter (A band-pass filter preserves signal power within a specified continuous frequency range, while attenuating signal power outside of this range)
  - Notch filter/Band-stop filter (A notch filter is the converse of a bandpass filter; it attenuates signal power within a specified continuous frequency range, while preserving signal power outside of this range)

Data Transformation

Data transformation are needed to bring uniformity to the data. In addition, it can be used to scale the data to a preferred range. Following are few methods as data transformation.

- Change of origin
- Change of scale
- Change of origin and scale
- Decimal scaling
- Min-Max normalization
- Standard normalization

1. Change of Origin:

- Arbitrarily choose a constant. If sample values are integers, an integer constant is preferred.
- Shift data point by subtracting (or adding) the chosen constant from each sample observation.
- This technique is useful when data values are large, and variability is not so large.

Lemma 6.1:

a) The range of original data is preserved by a change of origin transformation.
b) If $\bar{x}$ is the old mean and c is the chosen constant, then the new mean of the transformed data is $\bar{x} = c + \bar{x}'$

Example:

Apply change of origin method to the following data and calculate the old and new means.

$$x = \{115, 128, 110, 104, 133\}$$

Let us arbitrarily chose the constant 120 and subtract this from each value in x to get the transformed data x' as

$$x' = \{-5, 8, -10, -16, 13\}$$

The new mean $\bar{x}' = \frac{\sum x_i'}{5} = -2$

The mean of the original data is

$$\bar{x} = c + \bar{x}' = 120 - 2 = 118$$

2. Change of scale:

- This method is used to shorten the range of large numbers or lengthen the range of very small numbers.
- Chose (arbitrarily) a non-zero constant c. If c is less then the minimum of the observation, each value will be transformed to a value greater than 1. On the other hand, if it is greater than the maximum of the observation, then each value will be transformed to a value less than 1.
- If the value is between min and max of the sample, then the transformed values lie on the real line (positive real line if all x_i's are positive)
- If all values are small fractions, we may multiply a large constant to scale them up and vice-versa.

3. Change of origin and scale:

- This is the most frequently used method to standardize data values. Depending upon the constants used to change the origin and scale, a variety of transformation intervals can be obtained.

Example: A sample in the range (a, b) can be transformed to a new interval (c, d) by the following transformation.
Let x is the original and y is the transformed value. Then

$$y = c + \frac{(d - c)}{(b - a)} \times (x - a)$$

[Prove that all values in the range (a, b) are mapped onto the range (c, d)].

4. Min-Max Normalization:

- Min-Max normalization performs a linear transformation on the original data.
- Suppose, min_A and max_A are the minimum and maximum values of an attribute A. Min-Max normalization maps a value, v of A to v' in the range $min_A{}'$ and $max_A{}'$ using the following transformation:

$$v' = \frac{(v - min_A)}{(max_A - min_A)} \times (max'_A - min'_A) + min_A{}'$$

If $[min_A{}' \ , \ max_A{}'] = [0,1]$, then it is a special case of Min-max normalization.

5. Standard Normalization:

- This transformation is so called because it is extensively used in statistics in standardizing normal scores.
- Here, the origin is changed using the mean of the sample, and the scale is changed using the standard deviation of the sample.

$$\upsilon' = \frac{(\upsilon - \bar{A})}{\sigma_A}$$

where $\bar{A}$ and σ_A are the mean and std for the attribute A.

- This method is also alternatively termed as z-score normalization (zero-mean normalization) and the transformed values are called z-scores.

5. Standard Normalization:

Example: Given X = {32, 80, 56, 75, 69, 26, 44, 50}. Apply the standard normalization.

Here, the mean $\bar{A} = 54$, $\sigma_A^2 = 390$. Thus the z-scores are

$\upsilon' = \frac{(\upsilon - \bar{A})}{\sigma_A}$ = {-1.1140133 1.3165612 0.1012739 1.0633763 0.7595545 -1.4178351 -0.5063697 -0.2025479}

Note: It is very interesting to note that z-scores will always lie in the interval [-3, +3]

6. Decimal Scaling:

This data transformation method is same as the "change of scale" method by either scale up or scale down.

$$\upsilon' = \frac{\upsilon}{10^j}$$

Here j represents movement of decimal points. Decided based on the maximum value in the data.

SECOND STEP: EXTRACTING THE FEATURES

- ***BLOCK PROCESSING***
  - For most BCI applications, it is highly desirable for the processing to occur in real time. Prior to feature extraction, the incoming signal samples are commonly segmented into consecutive, possibly overlapping, sample blocks.
  - A feature vector is created from the signal samples within each individual sample block. The feature vectors from the successive sample blocks are then fed to the translation algorithm, which produces a device command or user feedback corresponding to each sample block or corresponding to sets of consecutive sample blocks.
  - For efficient online implementation, the length and overlap of these sample blocks should fit the relevant temporal dynamics of the signal, the feature-extraction method, the nature of the application, and the concurrent user feedback, as well as the available processing power.
    - E.g., BCI cursor control
    - P300 response

SECOND STEP: EXTRACTING THE FEATURES

- ***TIME (TEMPORAL) FEATURES***

1. Peak-Picking and Integration
   - Peak-picking simply determines the minimum or maximum value of the signal samples in a specific time block (usually defined relative to a specific preceding stimulus) and uses that value (and possibly its time of occurrence) as the feature(s) for that time block.
   - The signal can be averaged or integrated over all or part of the time block to yield the feature(s) for the block. Some form of averaging or integration is typically preferable to simple peak-picking, especially when the responses to the stimulus are known to vary in latency and/or when unrelated higher-frequency activity is superimposed on the relevant feature

SECOND STEP: EXTRACTING THE FEATURES

- ***TIME (TEMPORAL) FEATURES***

2. Correlation and Template-Matching
   - The similarity of a response to a predefined template might also be used as a feature.

SECOND STEP: EXTRACTING THE FEATURES

Statistical features

Sl. No	Features	Short description
1	MEAN	Mean value
2	STD	Standard deviation
3	MAX VALUE	Maximum positive amplitudes
4	MIN VALUE	Maximum negative amplitudes
5	SKEWNESS	a measure of asymmetry of the distribution
6	KURTOSIS	a measure of flatness of the distribution
7	MEDIAN	the middle value of a set of ordered data

SECOND STEP: EXTRACTING THE FEATURES

Interval or period analysis features.

Sl. No	Features	Short description
8	LINE LENGTH	Line length
9	MEAN VV AMPL	Mean of vertex-to-vertex amplitudes
10	VAR VV AMPL	Variance of vertex-to-vertex amplitudes
11	MEAN VV TIME	Mean of vertex-to-vertex times
12	VAR VV TIME	Variance of vertex-to-vertex times
13	MEAN VV SLOPE	Mean of vertex-to-vertex slope
14	VAR VV SLOPE	Variance of vertex-to-vertex slope
15	ZERO CROSSING	Number of zero crossings in a signal
16	MIN MAX NUMBER	Number of local minima and maxima
17	COEFF OF VARIATION	a statistical measure of the deviation of a variable from its mean, standard deviation divided by mean
18	AMPL RANGE	The difference between the maximum positive and maximum negative Amplitude values

SECOND STEP: EXTRACTING THE FEATURES

Features derived from the first and second derivative.

Sl. No	Features	Short description
19	1^{st} DIFF MEAN	Mean value of the first derivative of the signal
20	1^{st} DIFF MAX	Maximum value of the first derivative of the signal
21	2^{nd} DIFF MEAN	Mean value of the second derivative of the signal
22	2^{nd} DIFF MAX	Maximum value of the second derivative of the signal

SECOND STEP: EXTRACTING THE FEATURES

The Hjorth parameters

Sl. No	Features	Short description
23	HJORTH 1	Ability
24	HJORTH 2	Mobility $(\sigma x' / \sigma x)$
25	HJORTH 3	Complexity $\frac{(\sigma x'' / \sigma x''}{(\sigma x' / \sigma x)}$

NOTE:

$\sigma x'$ is the standard deviation of the first derivative of the signal
$\sigma x''$ is the standard deviation of the second derivative of the signal

SECOND STEP: EXTRACTING THE FEATURES

- ***FREQUENCY (SPECTRAL) FEATURES***
- Much brain activity manifests itself as continuous amplitude- and frequency-modulated oscillations. Therefore, it is often advantageous to accurately track these changes in the frequency domain. Although the Fourier transform is the most common method for converting from the time domain to the frequency domain, there are several alternatives that have characteristics that are particularly desirable given specific constraints or specific objectives. These include:
  - band power
  - fast Fourier transform (FFT)
  - autoregressive (AR) modeling

SECOND STEP: EXTRACTING THE FEATURES

FFT-based features calculated from the EEG spectra.

Sl. No	Features	Short description
26	FFT DELTA	0.1 - 3 Hz
27	FFT THETA	3 - 7 Hz
28	FFT ALPHA	7 - 12 Hz
29	FFT BETA	12 - 30 Hz
30	FFT GAMMA	30 - 40 Hz
31	FFT WHOLE	0.1 - 40 Hz

SECOND STEP: EXTRACTING THE FEATURES

FFT-based Spectral Features.

Sl. No	Features	Short description
32	FFT DT RATIO	*DELTA / THETA*
33	FFT DA RATIO	*DELTA / ALPHA*
34	FFT TA RATIO	*THETA / ALPHA*
35	FFT DTA RATIO	*(DELTA + THETA) / ALPHA*
36	FFT SEF	Spectral edge frequency (95 % of the total spectral power resides)
37	FFT SP-ROLL OFF	The frequency below which 85 % of the total spectral power resides

SECOND STEP: EXTRACTING THE FEATURES

Wavelet based Features.

Sl. No	Features	Short description
38	MIN WAV VALUE	Minimum value
39	MAX WAV VALUE	Maximum value
40	MEAN WAV VALUE	Mean value
41	MEDIAN WAV VALUE	Median value
42	STD WAV VALUE	Standard deviation
43	SKEWNESS WAV VALUE	Skewness
44	KURTOSIS WAV VALUE	Kurtosis
45	WAV BAND	Relative energy

SECOND STEP: EXTRACTING THE FEATURES

Wavelet based Features.

Sl. No	Features	Short description
46	ENTROPY SPECTRAL WAV	The spectral entropy
47	1^{st} DIFF WAV MEAN	Mean value of the 1*st* derivative
48	1^{st} DIFF WAV MAX	Maximum value of the 1*st* derivative
49	2^{nd} DIFF WAV MEAN	Mean value of the 2*nd* derivative
50	2^{nd} DIFF WAV MAX	Maximum value of the 2*nd* derivative
51	ENERGY PERCENT WAV	Percentage of the total energy of a detail/approximation
52	WAV ZERO CROSSING	Zero crossing
53	WAV COEFF OF VARIATION	Coefficient of variation
54	WAV TOTAL ENERGY	Total Energy

SECOND STEP: EXTRACTING THE FEATURES

Other Features.

Sl. No	Features	Short description
55	ENTROPY SPECTRAL	The spectral entropy
56	ENTROPY SHANNON	The Shannon entropy
57	MAX ABS XCORR EEG-EEG	Maximum positive amplitude of auto-correlation or cross-Correlation function
58	MEAN ABS XCORR EEG-EEG	Mean value of auto-correlation or cross-correlation function

THIRD STEP: FEATURE CONDITIONING

- The distributions and the relationships among the features can have a significant effect on the performance of the translation algorithm that follows feature extraction. These effects depend on the characteristics of the particular translation algorithm.

  - ***NORMALIZATION***
  - ***LOG-NORMAL TRANSFORMS***
  - ***FEATURE SMOOTHING***
  - ***PCA AND ICA***

Thank You!