Activity Recognition

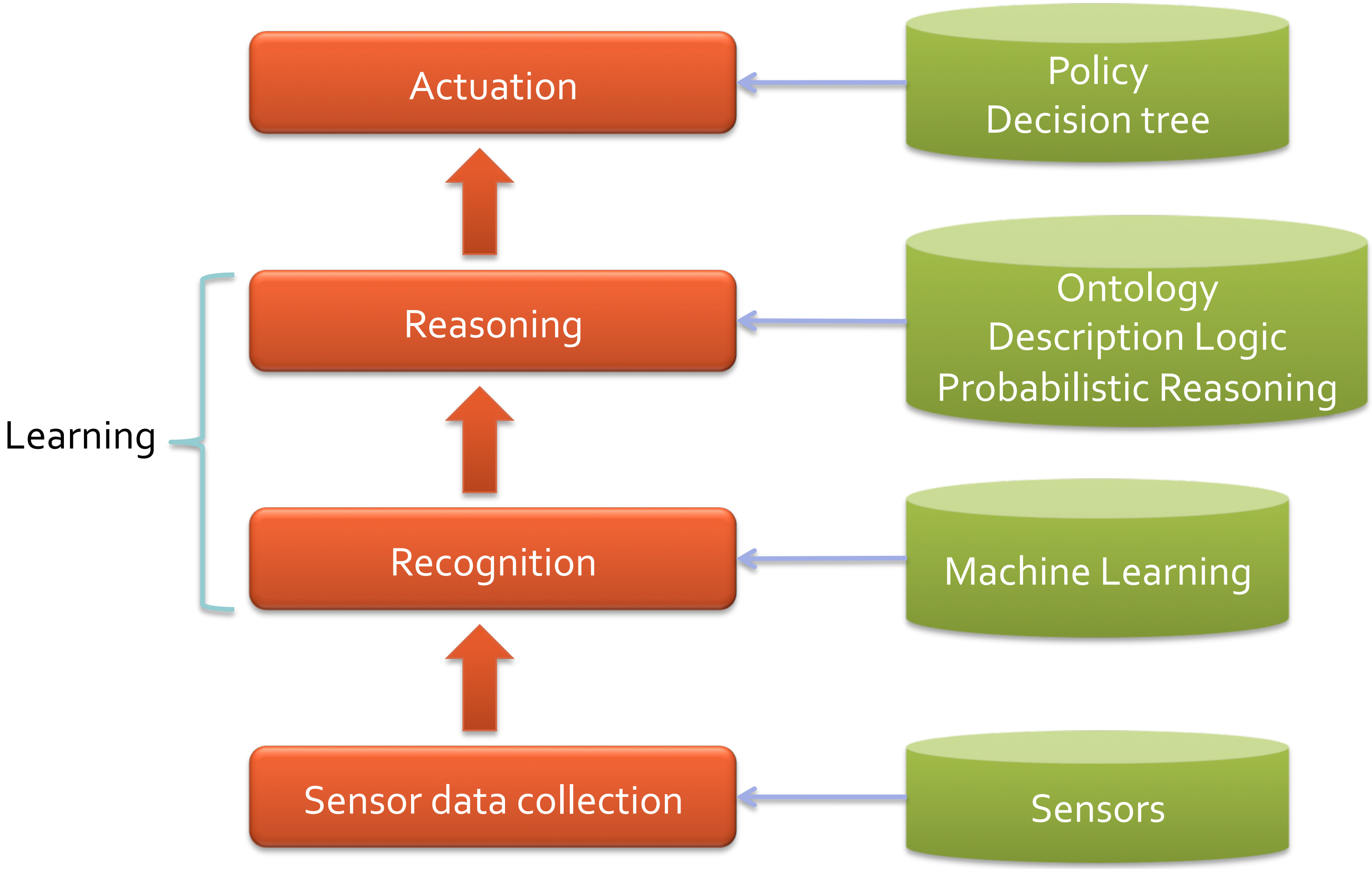
Enterprise Computing

Minho Shin

2011. 10. 5

REVIEW: MOBILE SENSING & ACCELEROMETER

Data processing of HMC



Supervised Learning

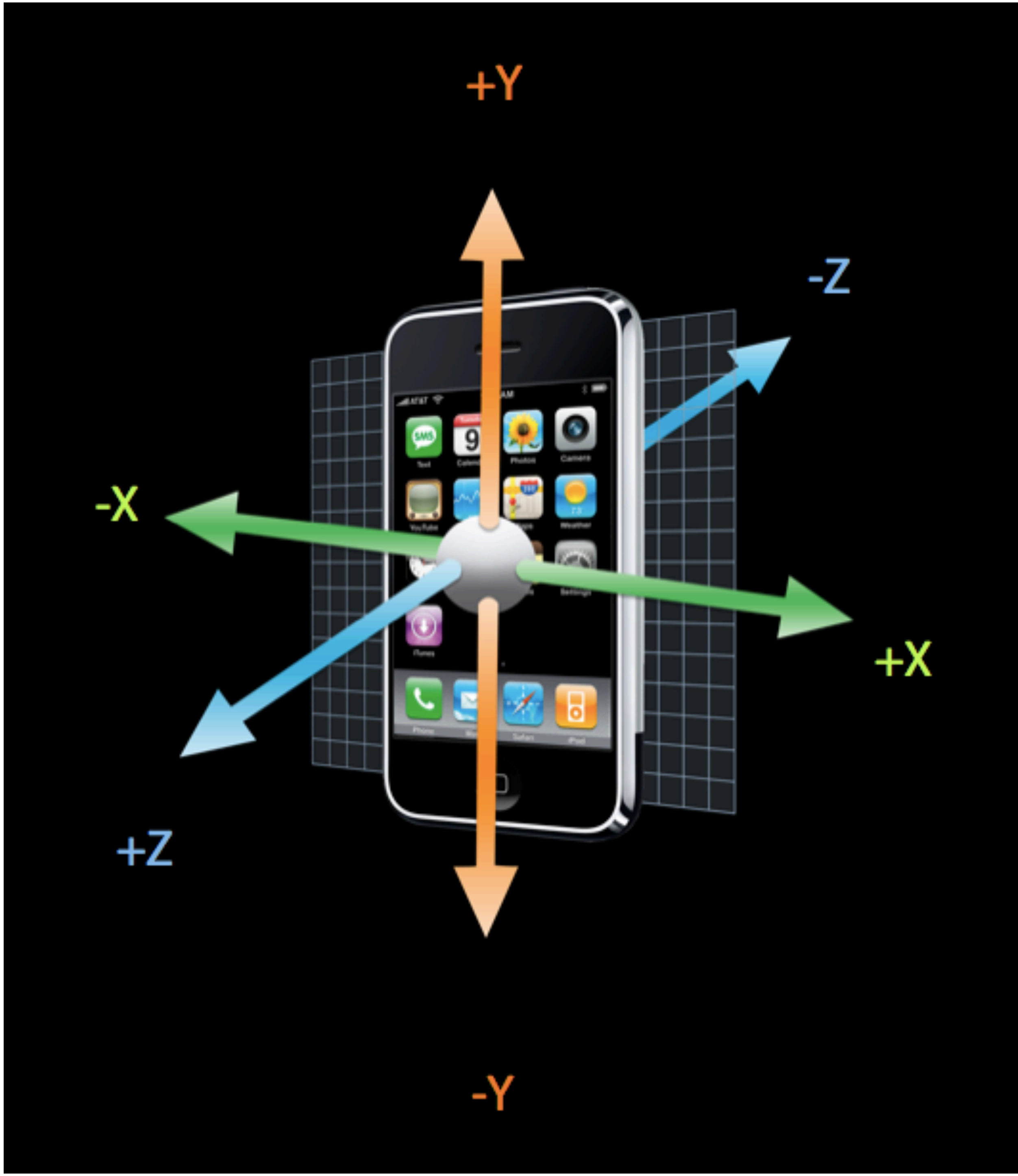
- Modeling → Training → Testing
- Modeling: decide a statistical model
- Training: Learn model parameters from training data
- Testing: Apply this model to the real data
- Data labeling
  - Supervised: all data is labeled
  - Semi-supervised: some data is labeled
  - Unsupervised: no data is labeled

Challenges: Sensing

- Continuous Sensing
  - Require multitasking and background processing
    - e.g., continuous accelerometer sampling
  - Heavy computation load
    - Interpreting audio data
  - Energy consuming
    - GPS reading requires a lot of energy (20 hrs down to 6 hrs)
    - Cloud-helped sensing
    - Duty cycling
    - Special processor architecture for continuous sensing (by Microsoft)

What is accelerometer?

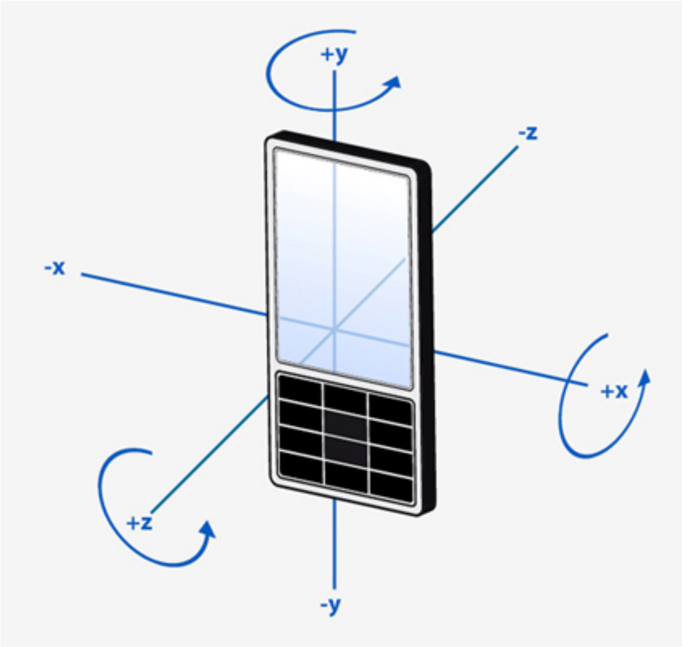
- An electromechanical device that measures acceleration
  - Static: gravity
  - Dynamic: moving, shaking, vibrating…
- Used for
  - Tilt: check the direction of the gravity
  - Acceleration to a particular direction
    - IBM/Apple use accelerometer to protect hard disk from scratch when falling
    - Launch an air bag in a car



rometer

What is Gyroscopic sensor?

- Measures the rotational movement around the three axes
- Combined with accelerometer, it senses detailed orientation
- Accel. vs Gyroscope
- Gyroscope & iPhone



What can we do with accelerometer?

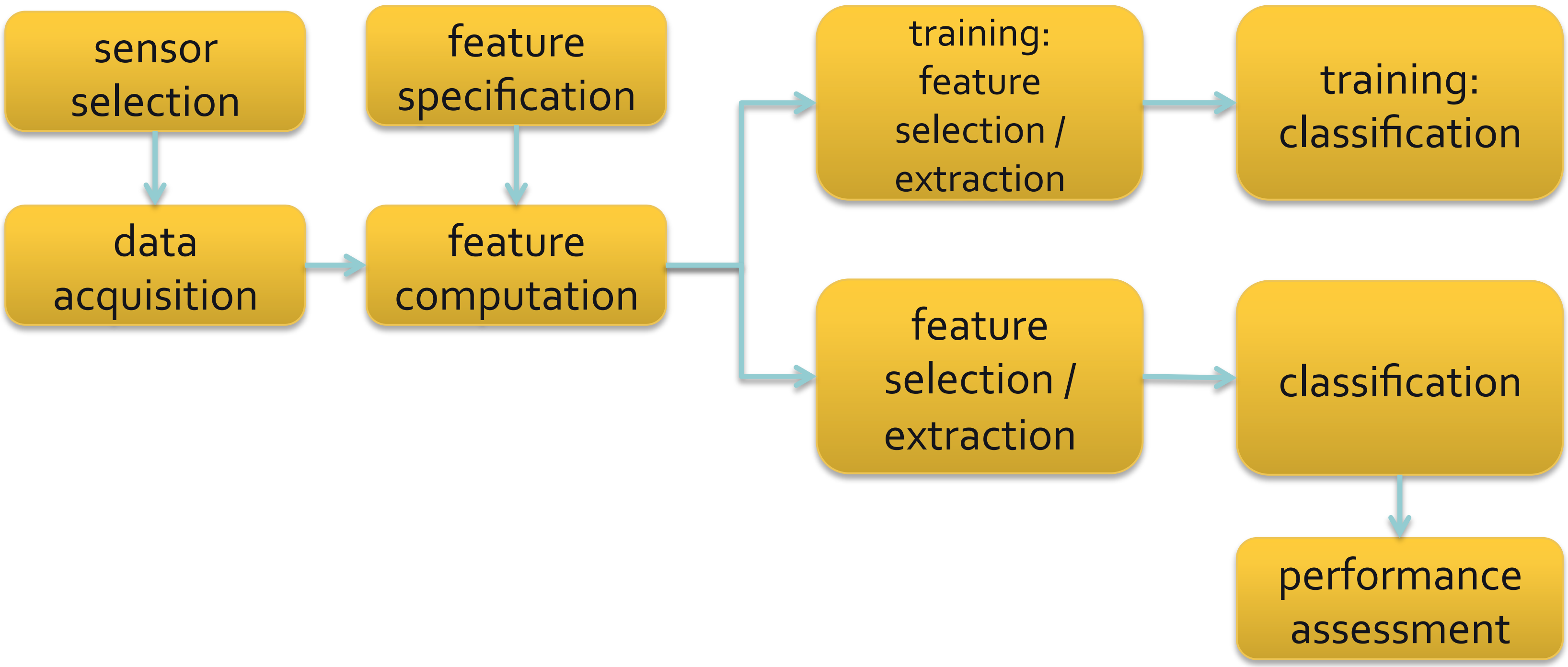
- Activity recognition
- Transportation recognition

Activity Recognition

- Physical activity
  - static posture: standing, sitting, lying
  - dynamic motions: walking, running, stair climbing, cycling
- Useful for
  - Bio-medical
    - metabolic energy expenditure
    - rehabilitation engineering: walking aid
  - Contextual knowledge
    - Human-computer interaction
    - Behavior prediction

Classification with supervised learning

- Classification: determine the type of activity

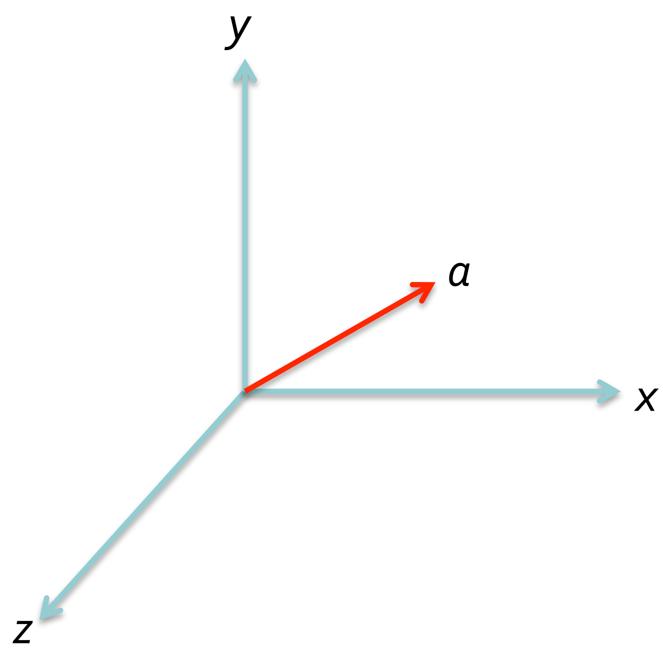


First step: Feature

- Raw data is not appropriate for analysis
- Feature: (statistical) characteristic of data
  - average, variance, ...
- Best feature-set depends on the problem

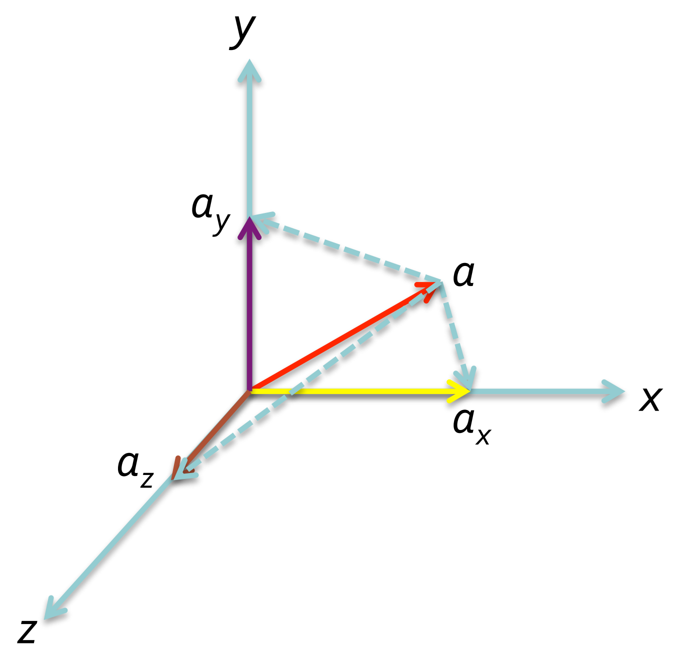
Accelerometer

- Measures an acceleration a



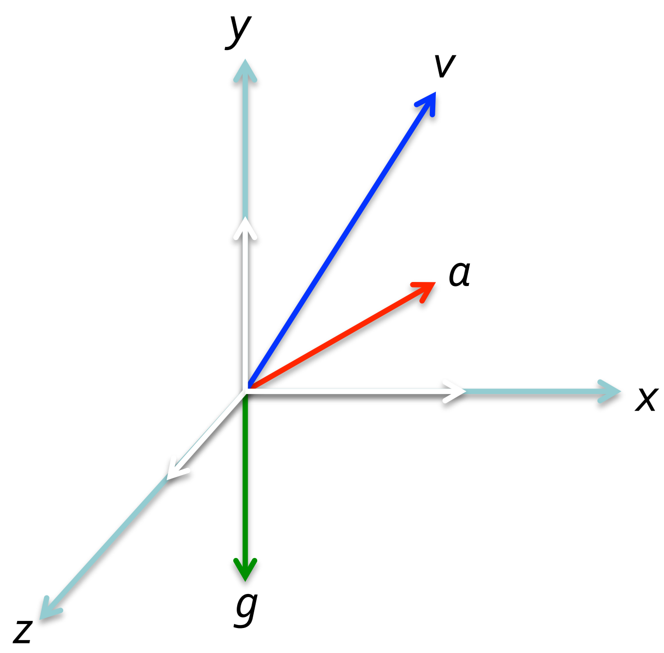
Accelerometer

- Measures an acceleration *a*
- Measures by the projections onto three axes a_x, a_y, a_z



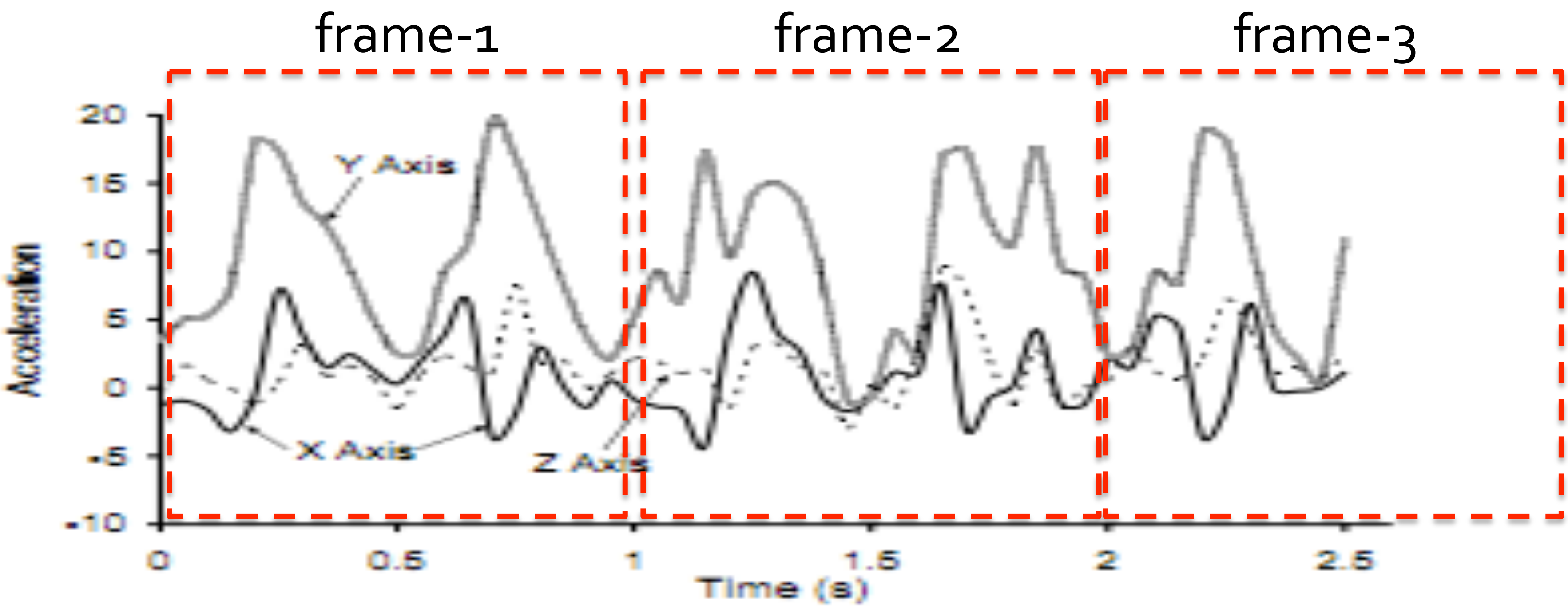
Accelerometer

- Measures an acceleration a
- Measures by the projections onto three axes a_x, a_y, a_z
- Acceleration is composed of (sum of)
  - Gravity (g)
    - constant force
    - low-frequency (DC component)
  - Movement (v)
    - temporary force
    - high-frequency (AC component)



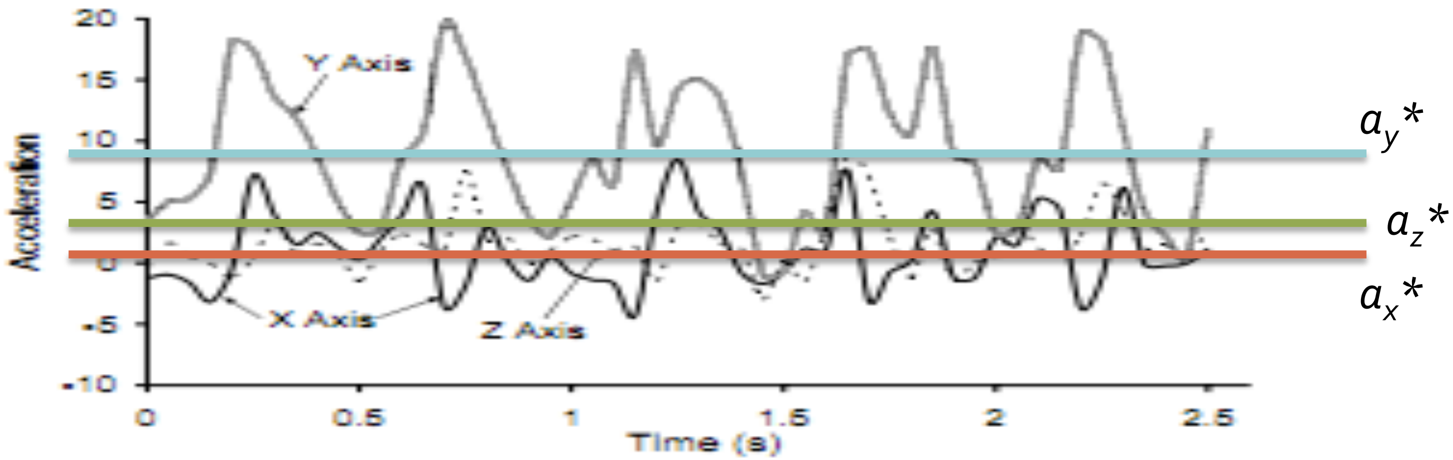
Data frame

- Features are evaluated within sliding windows with finite and constant width



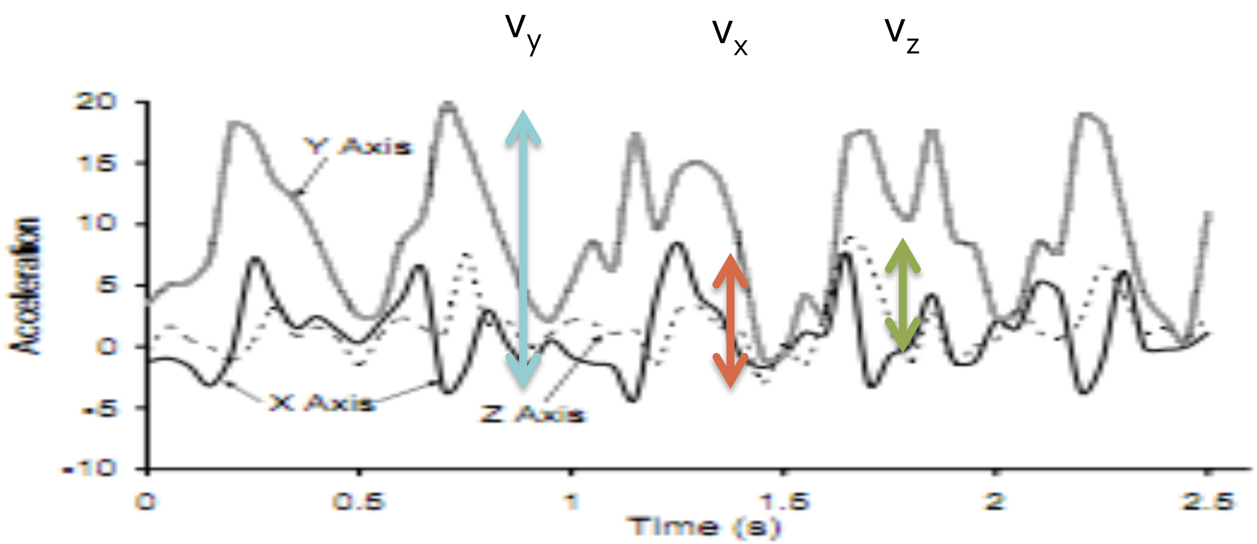
Features

- Signal average
  - (a_x^*, a_y^*, a_z^*), $a_x^* = \frac{a_1 + a_2 + \cdots + a_n}{n}$
  - the orientation w.r.t. gravity direction



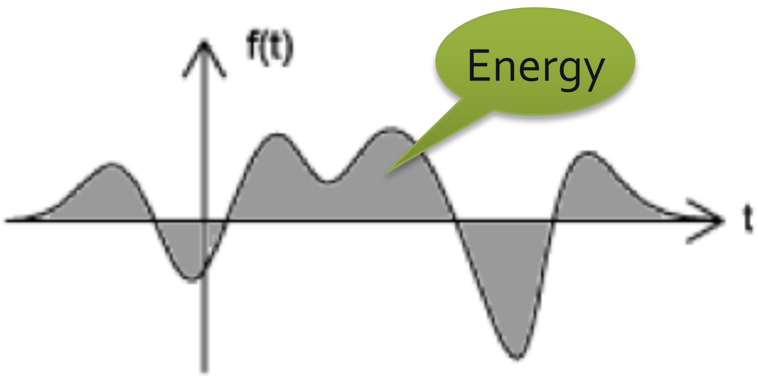
Features

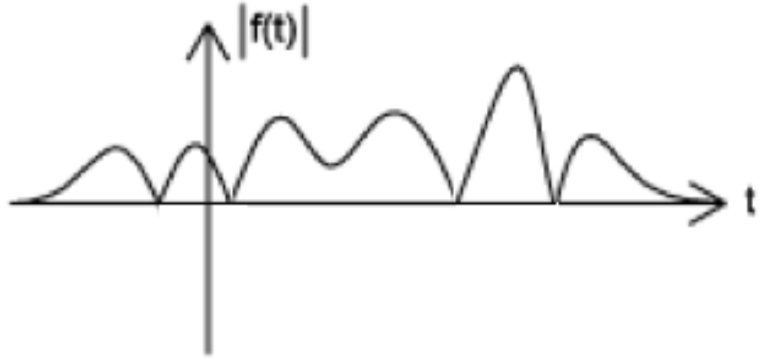
- Variance:
  - Deviation from the average
  - (v_x, v_y, v_z),
  - Level of instability



Features

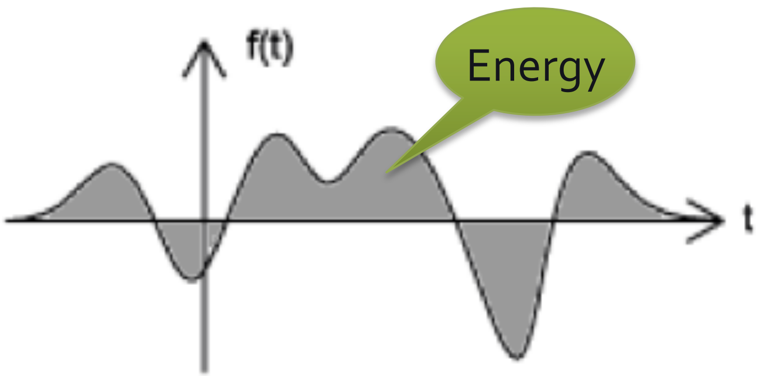
- Signal energy (in signal processing)
  - Area between the signal and the time axis
  - Negative integration reduce total energy!

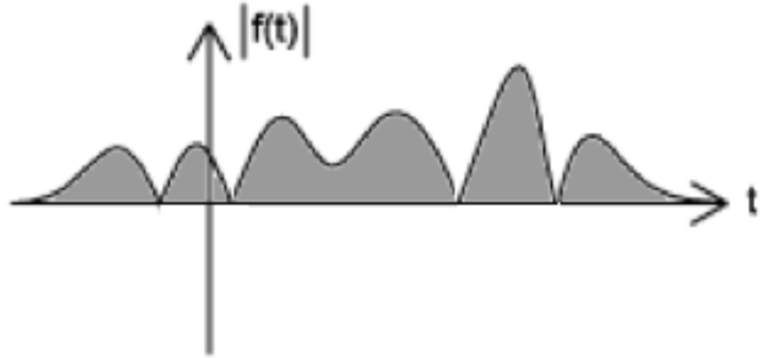




Features

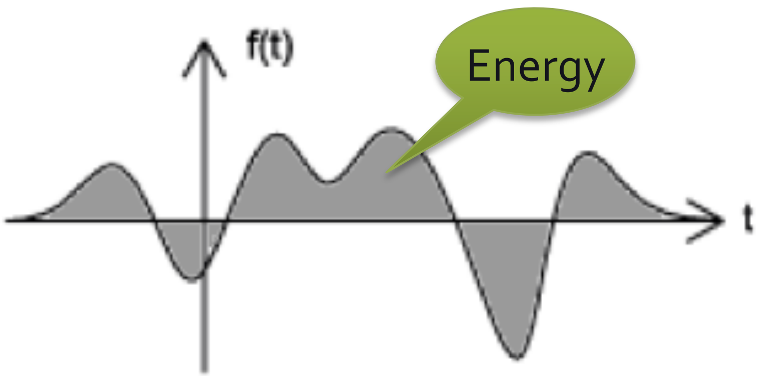
- Signal energy (in signal processing)
  - Area between the signal and the time axis
  - Integrate absolute value of f
  - Difficult to handle

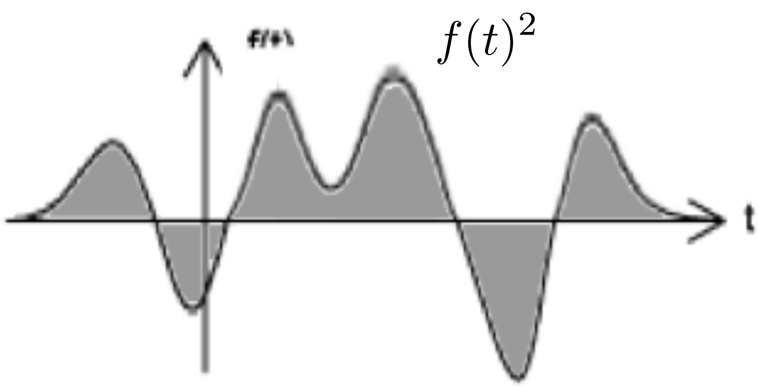




Features

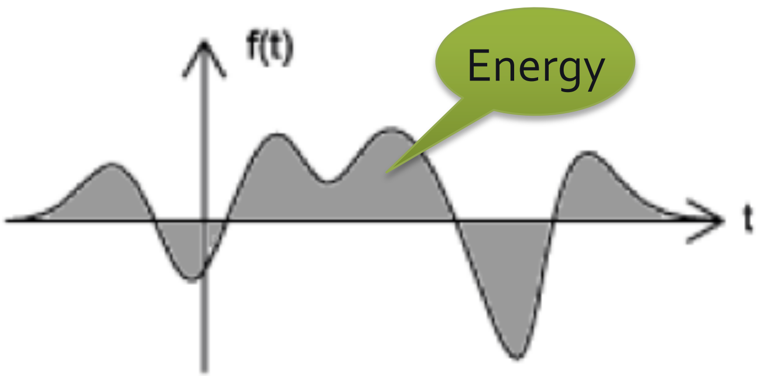
- Signal energy (in signal processing)
  - Area between the signal and the time axis
  - Integrate square of f
  -

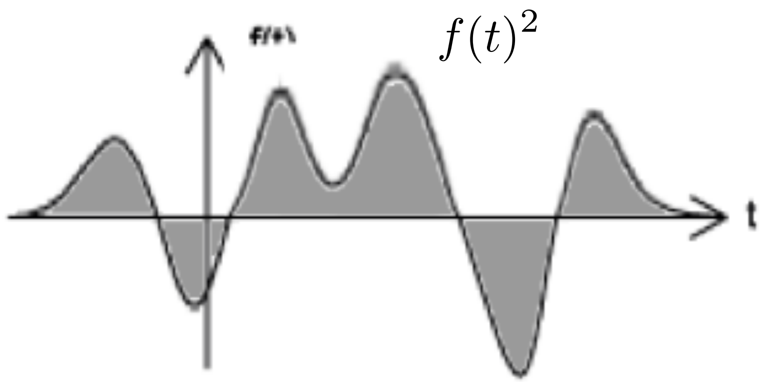




Features

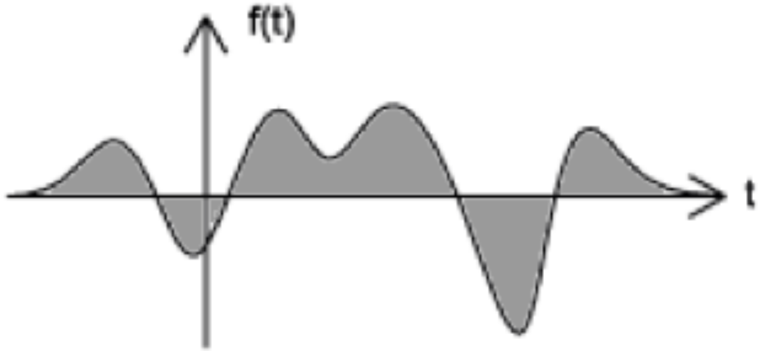
- Signal energy (in signal processing)
  - Area between the signal and the time axis
  - Integrate square of f
  -





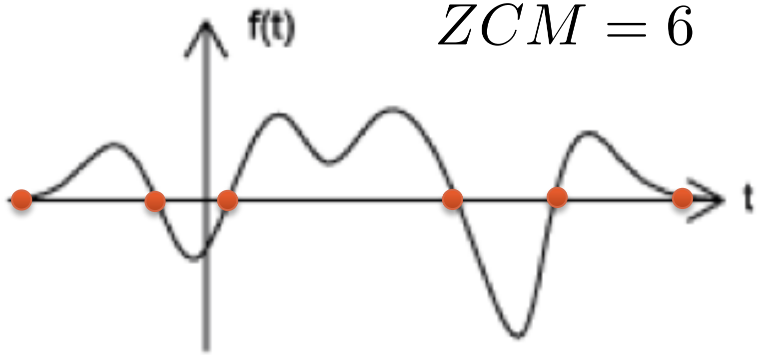
Actigraphy

- Activity level: Actigraphy
  - Measures activity level of a person
  - Used for diagnosing sleep disorder
  - Fitbit Tracker
  - WakeMate for iPhone
  - Three methods: ZCM, PIM, TAT
  - Proportional Integral Mode (PIM)

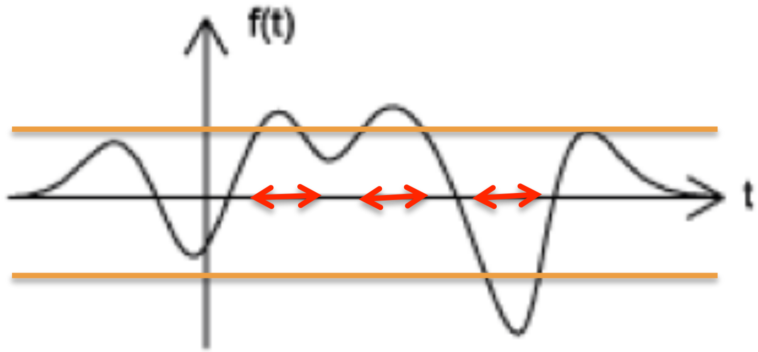


Actigraphy

- Zero Crossing Mode (ZCM)



- Time Above Threshold (TAT)

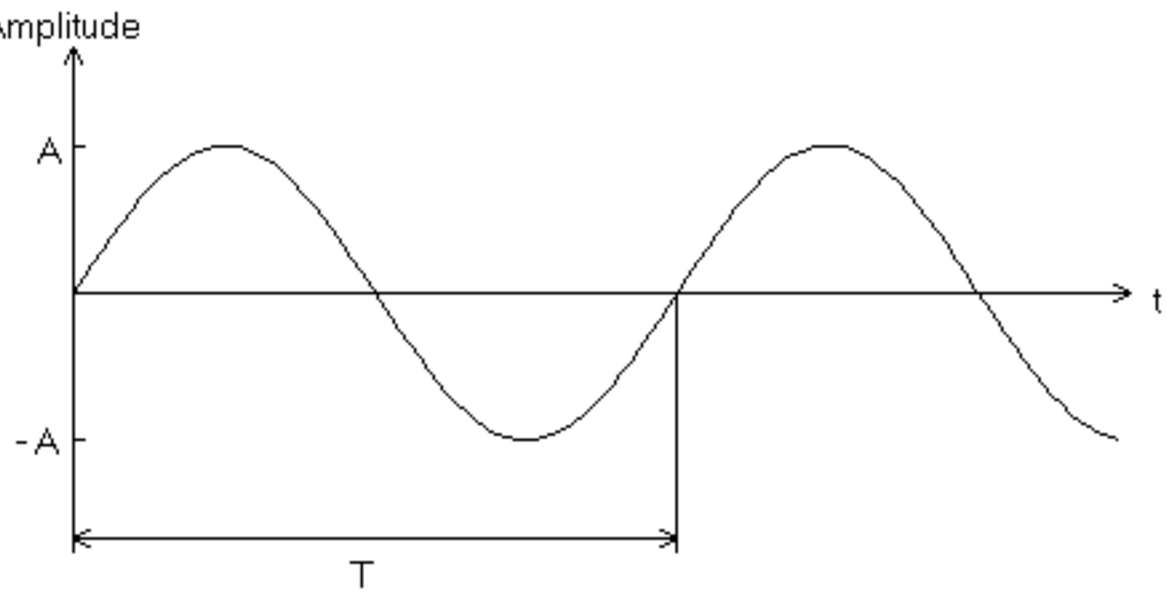


Features

- Frequency-domain entropy
  - Differentiate between walking and cycling
- What is *frequency domain*?

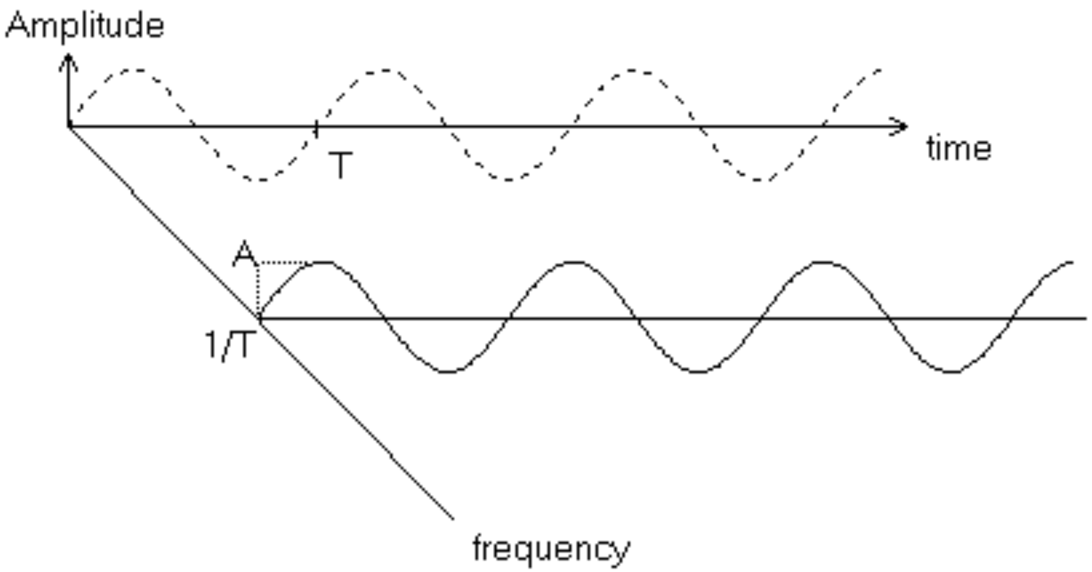
Representation of Signal

- Most common representation of signals and waveforms is in the *time domain*
- *Time plane*: a plane defined by *time* and *amplitude*

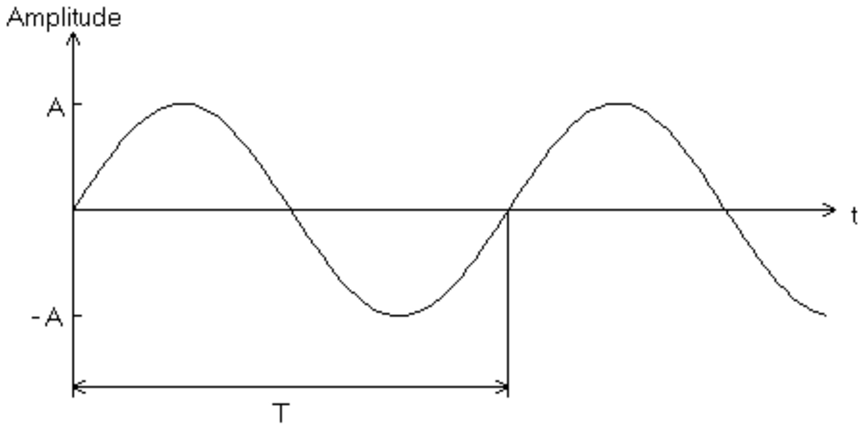


Frequency Axis

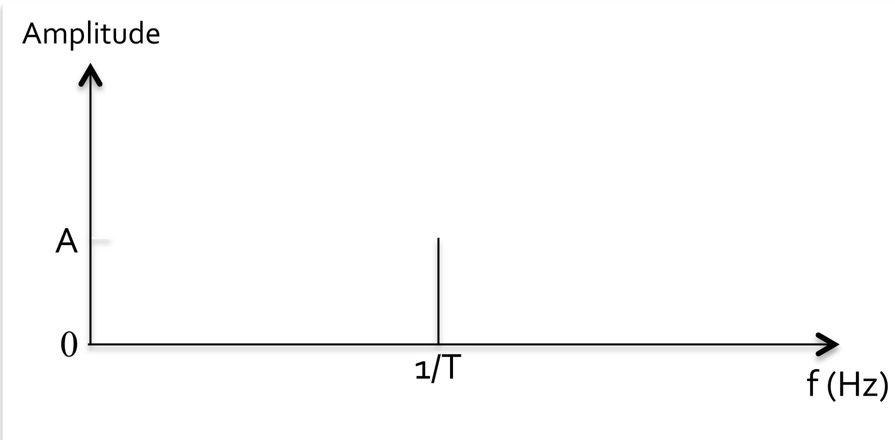
- We can add *frequency* axis to the time plane
- frequency & amplitude defines *frequency plane*



- Time domain representation is a projection of the signal (in time-freq-amp space) onto the time plane



- Frequency domain rep. is a projection of the signal (only positive amplitude part) onto the frequency plane



Frequency domain

- Frequency domain representation is just another way to represent a signal
- But it looses some information, i.e., phases
  - So, time domain representation has more information than freq. domain
- Fortunately, freq. domain provides useful information: Frequency Domain Analysis

Fourier Transform

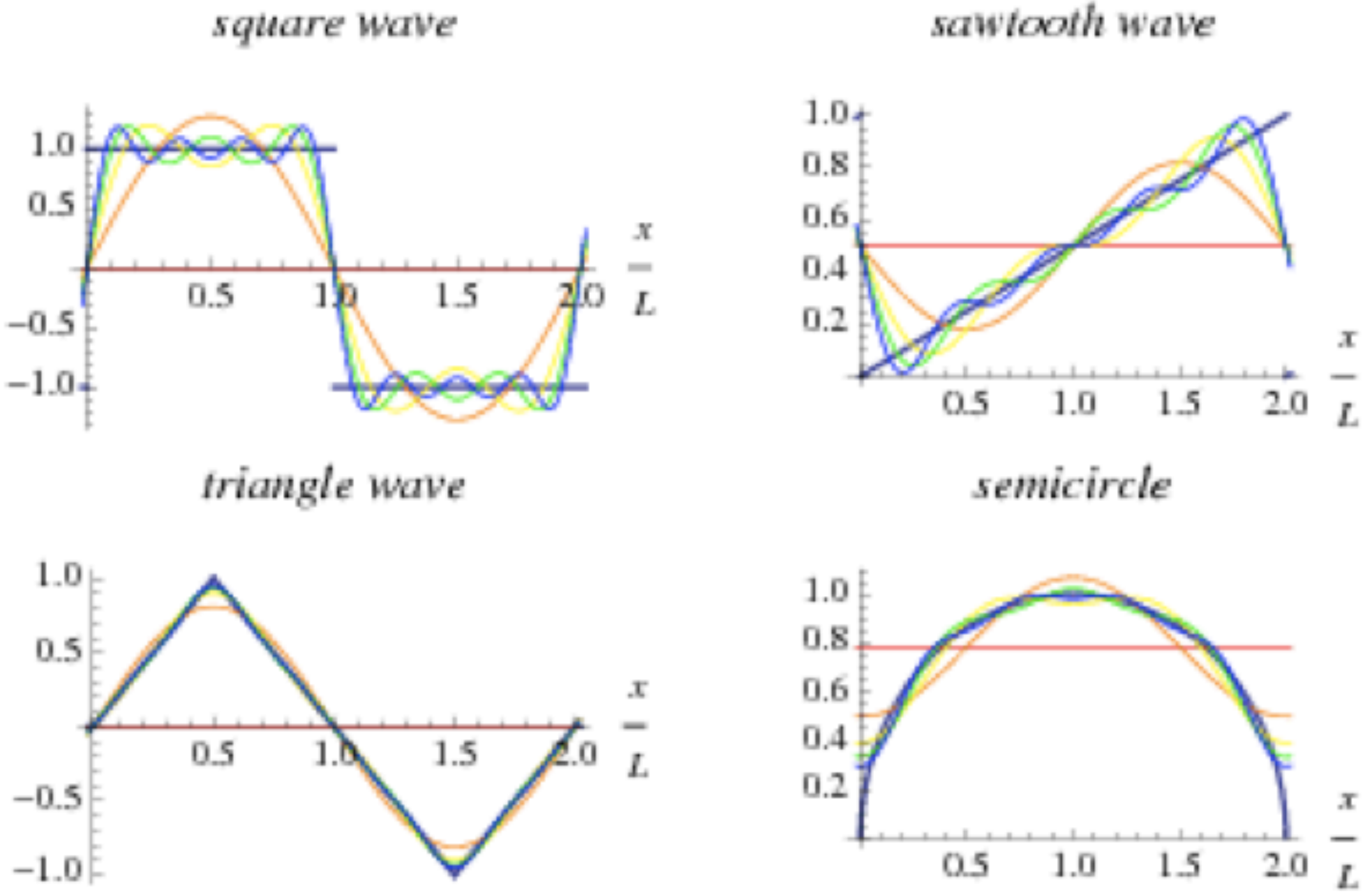
- How can we represent a general signal in frequency domain?
  - Fourier Transformation
- Fourier found that *any* signal can be represented by the sum of infinite number of *periodic* waveforms (or approximated by finite such waveforms)

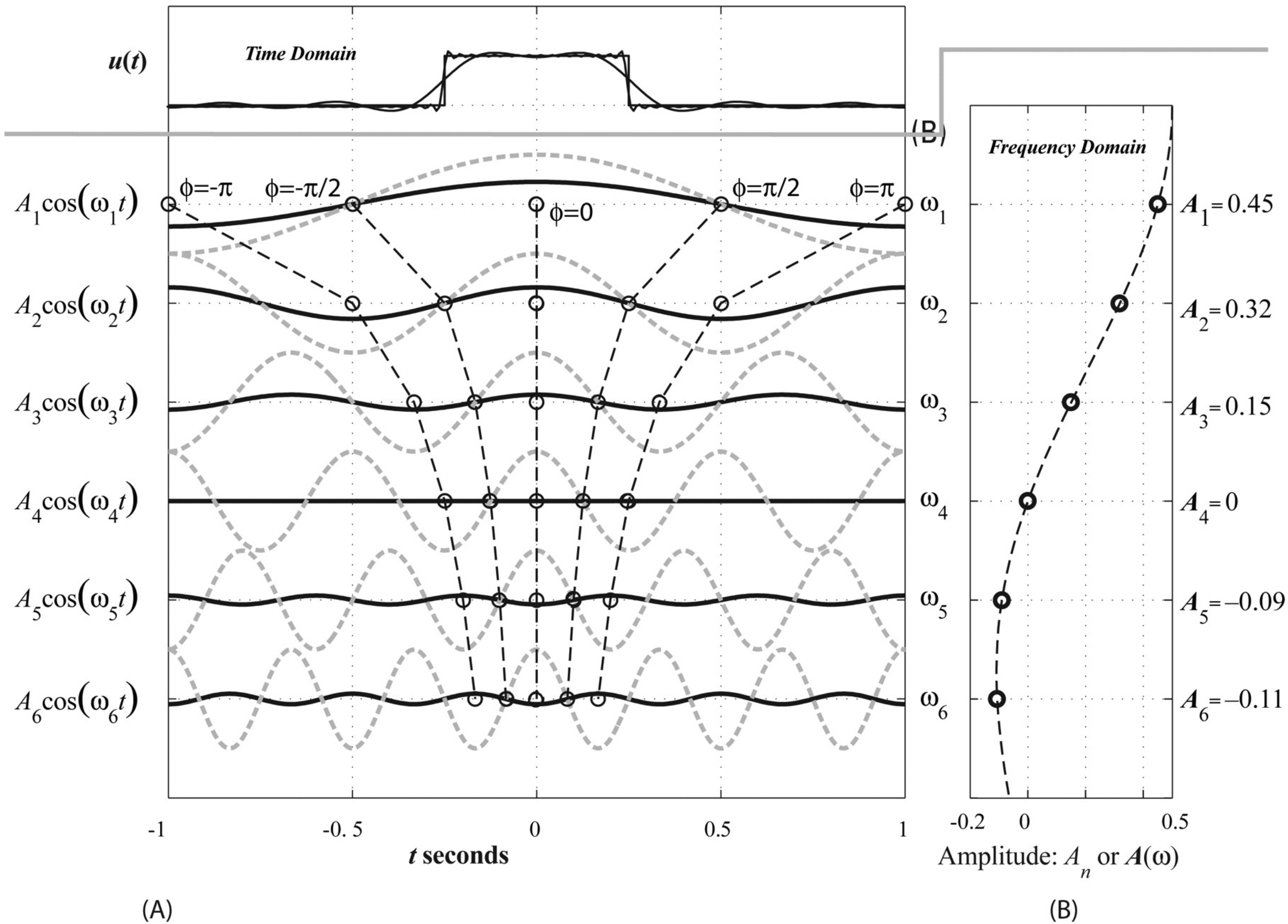
Fourier Series

- Joseph Fourier submitted a paper in 1807 to Academy of Sciences of Paris, describing Fourier Series, but rejected for lack of mathematical rigor. Later honored to him

$$F(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos n\omega_T t + b_n \sin \omega_T t)$$

- Using only a finite series (say the first *k* terms), we can approximate any signals with a finite sum of sin/cos waves

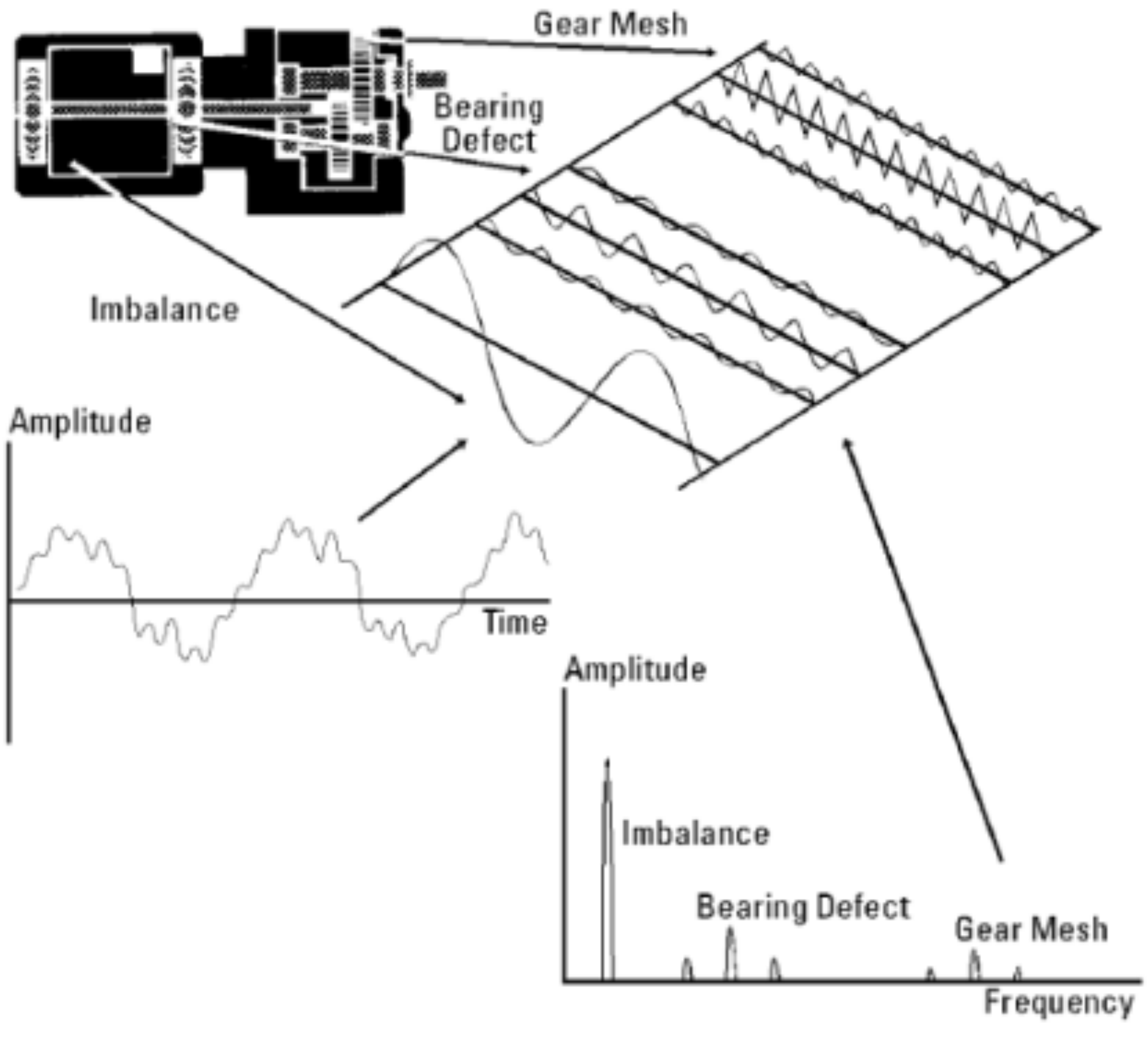

square wave
1.0
0.5
−0.5
−1.0
0.5
1.0
1.5
2.0
x
—
L
sawtooth wave
1.0
0.8
0.6
0.4
0.2
0.5
1.0
1.5
2.0
x
—
L
triangle wave
1.0
0.5
−0.5
−1.0
0.5
1.0
1.5
2.0
x
—
L
semicircle
1.0
0.8
0.6
0.4
0.2
0.5
1.0
1.5
2.0
x
—
L

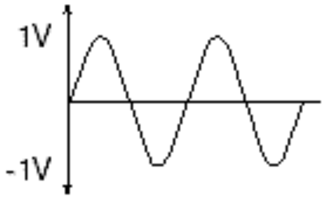
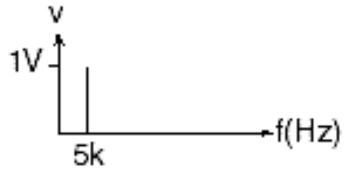
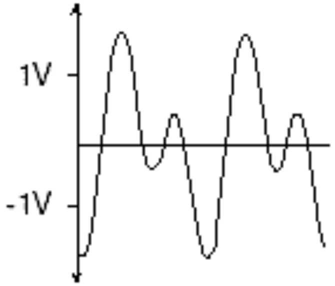
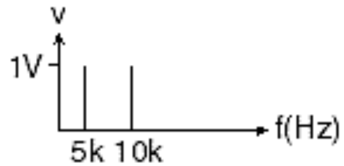
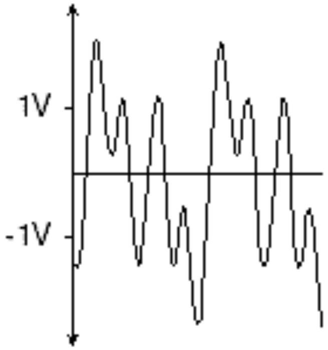
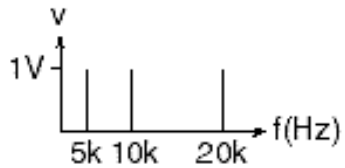
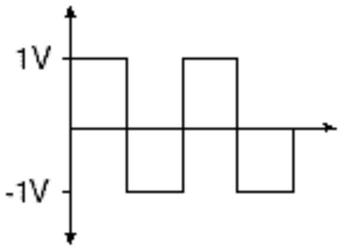
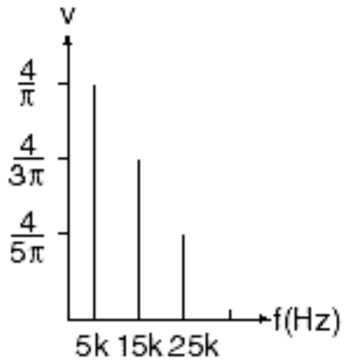


u(t)
Time Domain
Frequency Domain
A1cos(ω1t)
A2cos(ω2t)
A3cos(ω3t)
A4cos(ω4t)
A5cos(ω5t)
A6cos(ω6t)
ϕ=-π
ϕ=-π/2
ϕ=0
ϕ=π/2
ϕ=π
ω1
ω2
ω3
ω4
ω5
ω6
A1 = 0.45
A2 = 0.32
A3 = 0.15
A4 = 0
A5 = –0.09
A6 = –0.11
-1
-0. 5
0
0.5
1
t seconds
-0.2
0
0.5
Amplitude: An or A(ω)
(B)
(A)
(B)

Build Frequency Domain Rep. from Fourier Series

Build Frequency Domain Rep. from Fourier Series

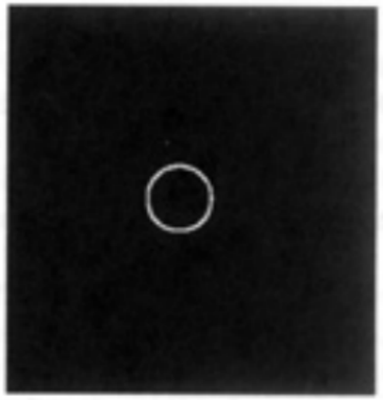


Description	Time Series	Fourier Expansion	Power Spectrum
A pure 5kHz sine wave measuring 1 volt peak		$v(t) = 1\sin(\omega 1)t$ $\omega 1 = 2\pi(5kHz)$	
A pure 5kHz and 10kHz sine wave, each measuring 1 volt peak, added together		$v(t) = 1\sin(\omega 1)t + 1\sin(\omega 2)t$ $\omega 1 = 2\pi(5kHz)$ $\omega 2 = 2\pi(10kHz)$	
A pure 5kHz, 10kHz, and 20kHz sine wave, each measuring 1 volt peak, added together		$v(t) = 1\sin(\omega 1)t + 1\sin(\omega 2)t + 1\sin(\omega 3)t$ $\omega 1 = 2\pi(5kHz)$ $\omega 2 = 2\pi(10kHz)$ $\omega 3 = 2\pi(20kHz)$	
A pure 5kHz square wave measuring 1 volt		$v(t) = \frac{4}{\pi}\sin(\omega 1)t + \frac{4}{3\pi}\sin(\omega 2)t + \frac{4}{5\pi}\sin(\omega 3)t\ldots$ $\omega 1 = 2\pi(5kHz)$ $\omega 2 = 2\pi(15kHz)$ $\omega 3 = 2\pi(25kHz)\ldots$	

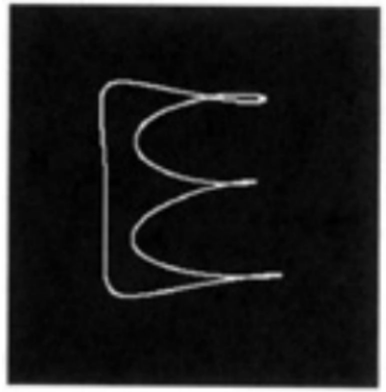
Application of FT

Original

k=3



k=21



k=61

k=201

k=401

Feature selection/extraction

- Curse of dimensionality
  - If dimension of features is high, classification becomes difficult
- Solution
  - Reduce the dimension of features
  - Feature selection
  - Feature extraction

Feature selection

- Given a feature set *F*, get a subset *G* of *F*
- Discarding features that are little helpful for classification
- But, finding the subset is exponentially expensive
  - For example, if $F=\{f_1, f_2, \ldots, f_d\}$ (d is F's dimension), for $m=1, 2, .., d$, the we have to check all subsets of F of size m
  - So, we use sub-optimal search algorithm instead of optimal, which is an exhaustive search
    - Branch-and-bound search
    - Sequential forward/backward search (SFS-SBS)
    - Sequential forward/backward floating search (SFFS-SBFS)

Feature selection: Suboptimal Algorithms

- Use sub-optimal search algorithm
  - Branch-and-bound search
  - Sequential forward/backward search (SFS-SBS)
  - Sequential forward/backward floating search (SFFS-SBFS)
- Sequential search algorithms
  - Iterative procedure
  - Add or remove some features at each step so that the new set leads to a better classification performance, measured by
    - Inter-class distance / intra-class distance
    - Analyze classifier output

Feature Extraction

- Idea: another data representation can be constructed in a subspace (less dimension) while keeping discriminative capability
- Lose physical meaning
- Example algorithms
  - PCA (Principle Component Analysis): transform features into small number of uncorrelated variables
  - ICA (Independent Component Analysis)
- Feature selection and extractions can be used together

Type of Classifiers

- Supervised & Unsupervised
  - Supervised: class membership of each feature vector is known
  - Unsupervised: Only the number of classes is known
- Single-frame & Sequential
  - Single-frame: Each frame is classified regardless of previous frames
  - Sequential: Each frame is classified in consideration of previous frames

Type of Classifiers

- Probabilistic & Geometric & Template matching
  - Probabilistic: feature vector $\boldsymbol{x}$ is classified to class C_{i*} if class-conditional PDF $p(\boldsymbol{x}|C_i)$ is maximized for $i=1, \ldots, C$
    - Optimal Bayesian classifier
    - Since class-conditional pdf is unknown, use suboptimal
      - naïve Bayesian, Logistic, Parzen, Gaussian Mixture Model (GMM)

Type of Classifiers

- Probabilistic & Geometric & Template matching
  - Geometric: Construct decision boundaries that divide feature space into classes
    - Artificial Neural Networks (ANN): iterative tessellation of feature space
    - k-NN/ Nearest Mean (NM): geometrical distance between feature vectors of from different classes
    - Support Vector Machine (SVM): construct boundaries maximizing the margins between nearest features relative to two distinct classes
    - Threshold-based classifier: careful handcrafting of thresholds

Type of Classifiers

- Probabilistic & Geometric & Template matching
  - Template matching: Based on similarity between data and templates obtained by training or defined by the designer
  - Binary classifier: Descend a binary decision tree from the root to leaves as refining the classification