COMS21202: Symbols, Patterns, and Signals

Part 3

Representations, Transformations, and Features

Majid Mirmehdi

m.mirmehdi@bristol.ac.uk

COMS21202: Symbols, Patterns, and Signals

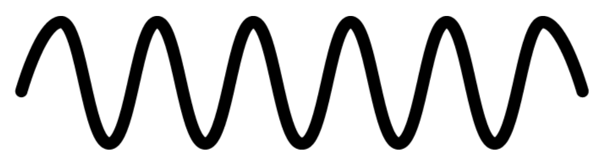
SPS: What comes next!?

- Data representations
- Transformations
- Feature extraction

- Fourier Space Analysis
- Convolutions
- Principal Component Analysis

This Lecture:

- Overview
- Intro to Signals



Analog Signal

Digital Signal

EXTREMELY IMPORTANT WARNING!!!

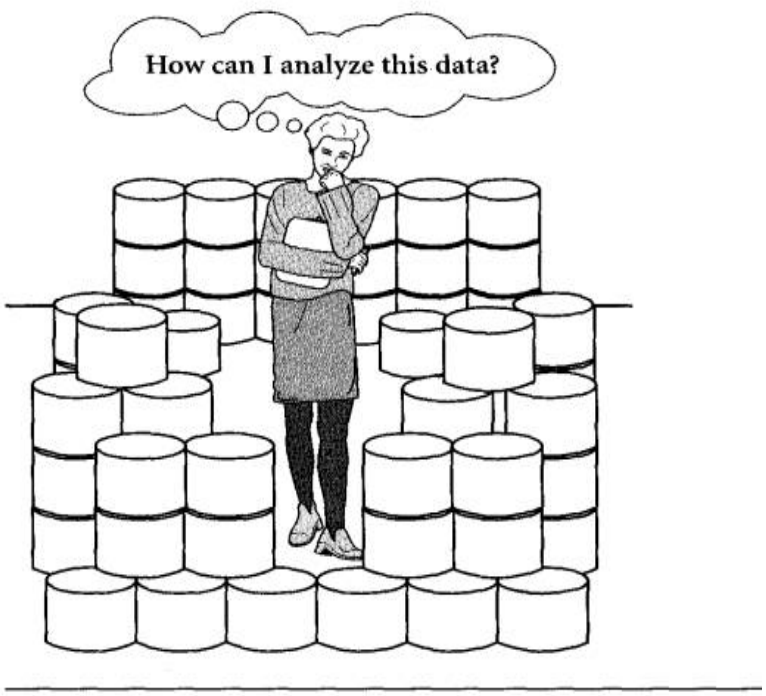
- **My part of the SPS unit was taught by someone else last year.**
- **This means for the academic year 2018-2019.**
- **The content for that year are either completely different or have a different focus to what I normally teach.**
- **This means the lecture notes for that year are different.**
- **This means the problem classes for that year are different.**
- **This means the exam questions for that year are different.**

IGNORE ANY MATERIALS OR EXAM QUESTIONS FROM 2018-19 ACADEMIC YEAR RELATED TO THE THIRD PART OF SPS!

SPS: The Story So Far

The sorts of ways we wish to manipulate and analyze data:

- Data modelling
- Classification and recognition
- Clustering and segmentation
- Estimation and detection



What next?

Example: Image Categorization (Indoors or Outdoors?)

Majid's part of SPS: How to represent the data and then extract features from it

Supervised Training

Training Data

Image Features → Classifier Training → Trained Classifier

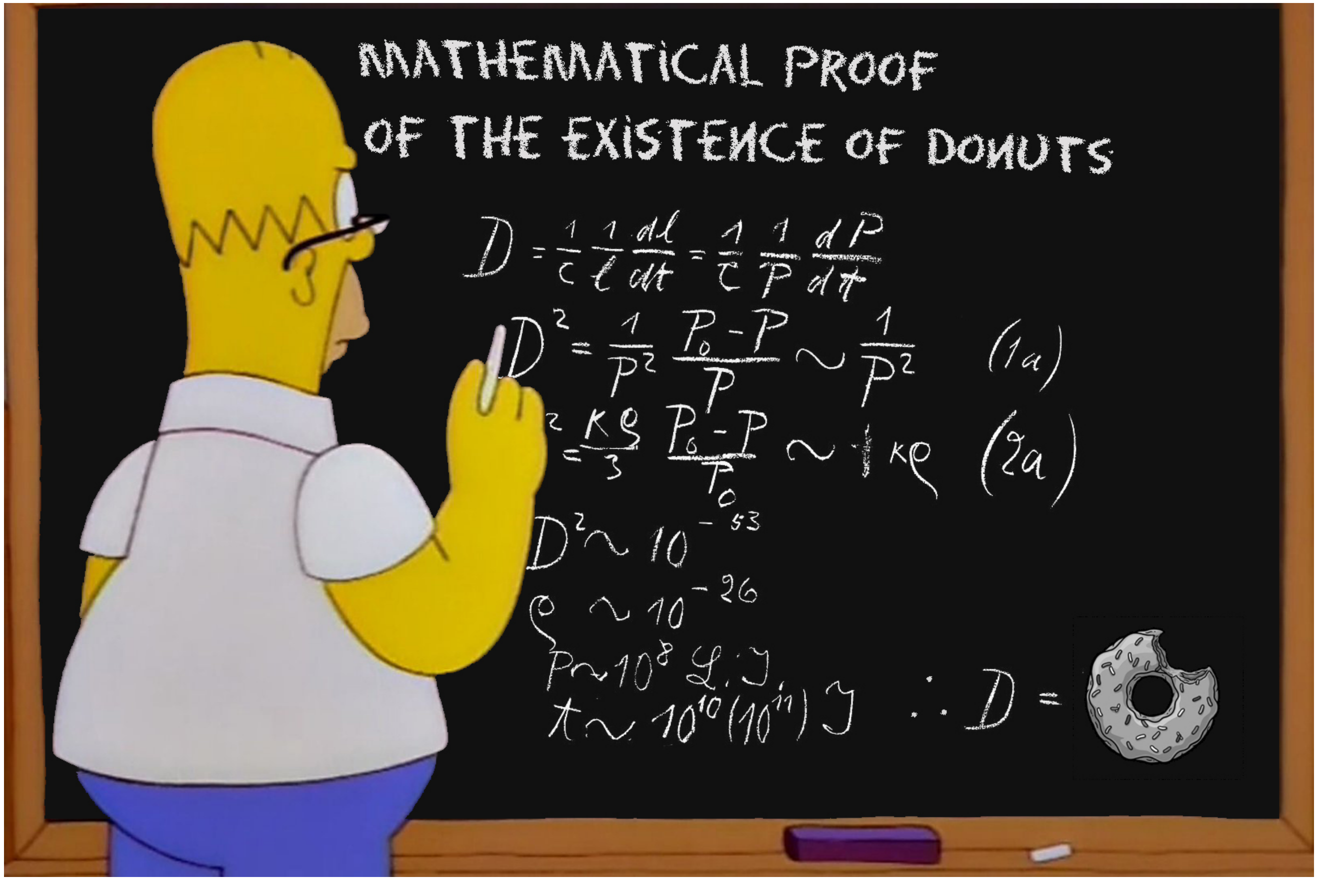
Training Labels → Classifier Training

Testing

Test Data

Image Features → Trained Classifier → Prediction *Outdoor*

Maths: *nothing scary!*



Representing Data

To manipulate data properly we may have to represent it in a different way. ***Why?***

- Sometimes we need to look at data in a different way.
- Sometimes we need to alter it to prepare it for the next stage of processing or data analysis. Because:
  - It is noisy (errors or outliers),
  - It is missing values,
  - It contains redundancies,
  - It contains inconsistencies
  - It reveals its substance or begins to make sense

Representing Data

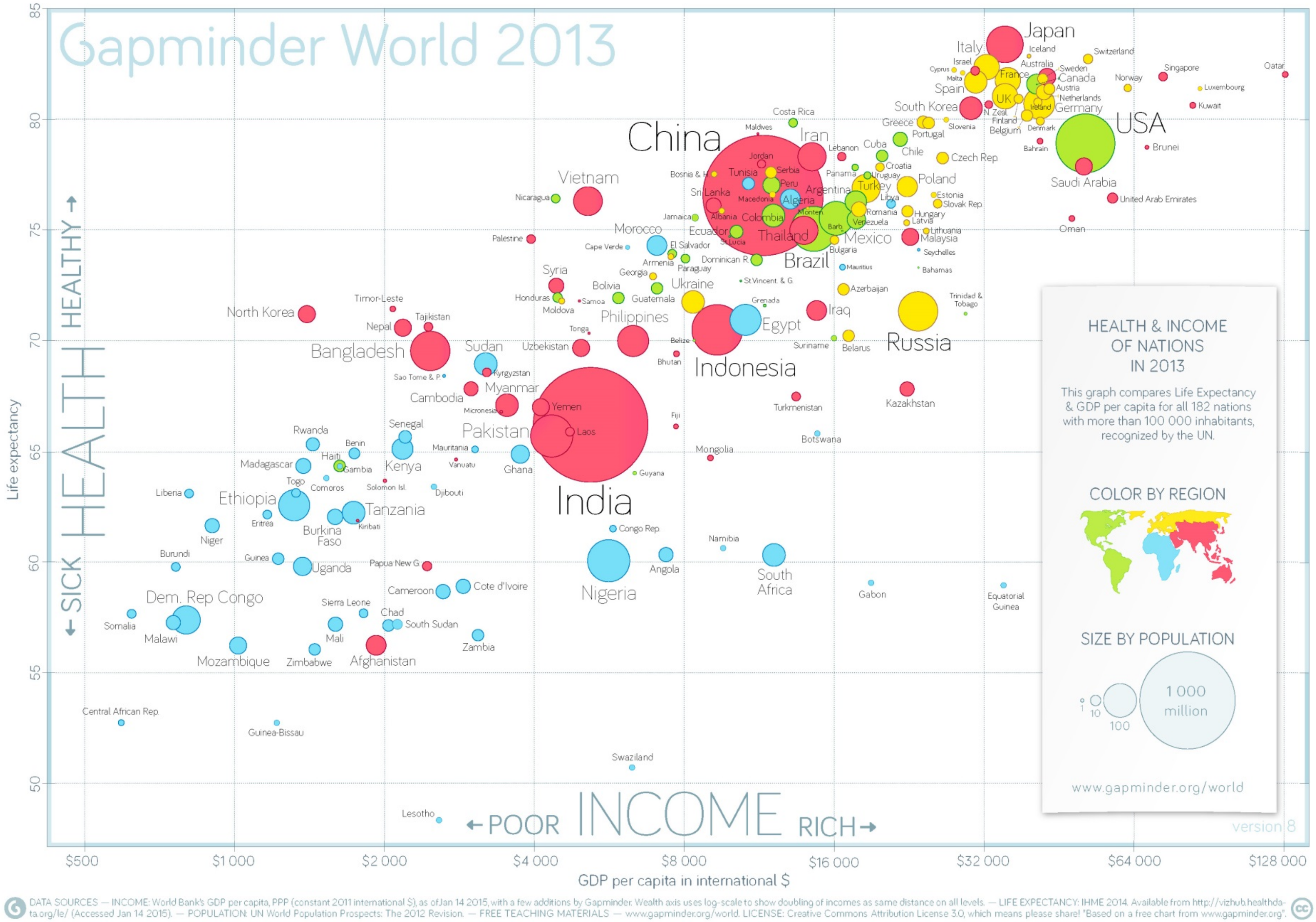
To manipulate data properly we may first **pre-process** it:

- **Data cleaning**: a process that removes noisy and inconsistent data
- **Data integration**: where multiple data sources may be combined (also known as Data Fusion)
- **Data selection**: where data relevant to the analysis task are retrieved, filtered, extracted

Then we are ready for data representation:

- **Data transformation**: where data are transformed, reduced or consolidated into forms appropriate for alternative representation and/or further analysis.

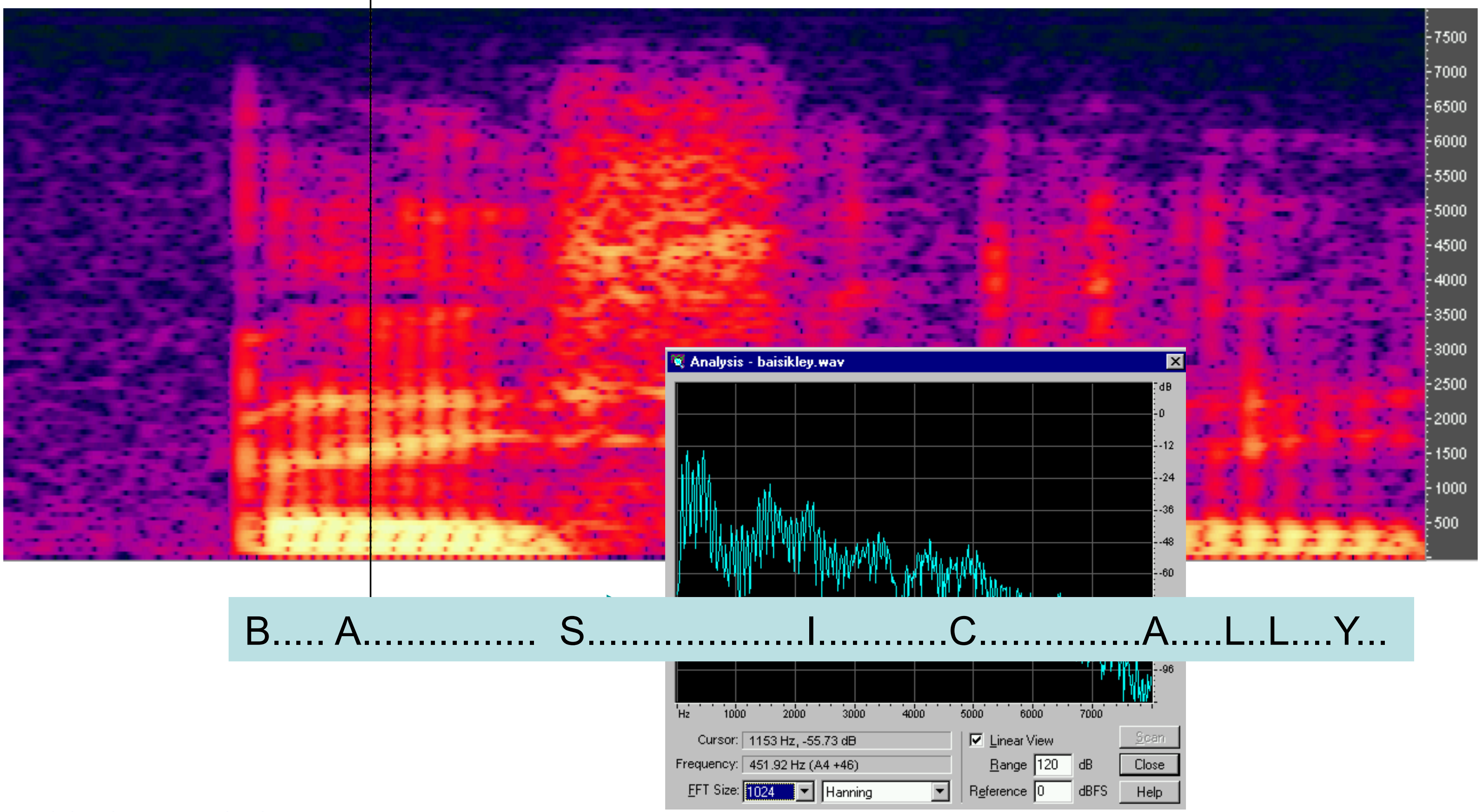
Visualizing Data



Go to https://www.gapminder.org/tools/ to play!

Frequency Domain Data Analysis

Spectrogram: Representation of time, frequency and amplitude



Spatial Domain Data Analysis: Cleaning/Clearing up Data

Sometimes we may manipulate data just so we (humans) can *see* the data better.

Noisy Gene Sequence:

GGATACAWCTTTAGAG

Cleaned Gene Sequence:

GGATACAACTTTAGAG

Spatial Domain Data Analysis: Feature Detection

Edge Detection

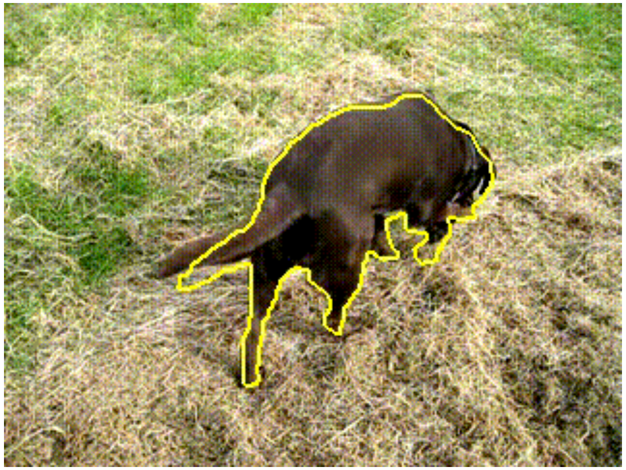
Blob Detection

sigma = 6.126

Features help simplify the problems

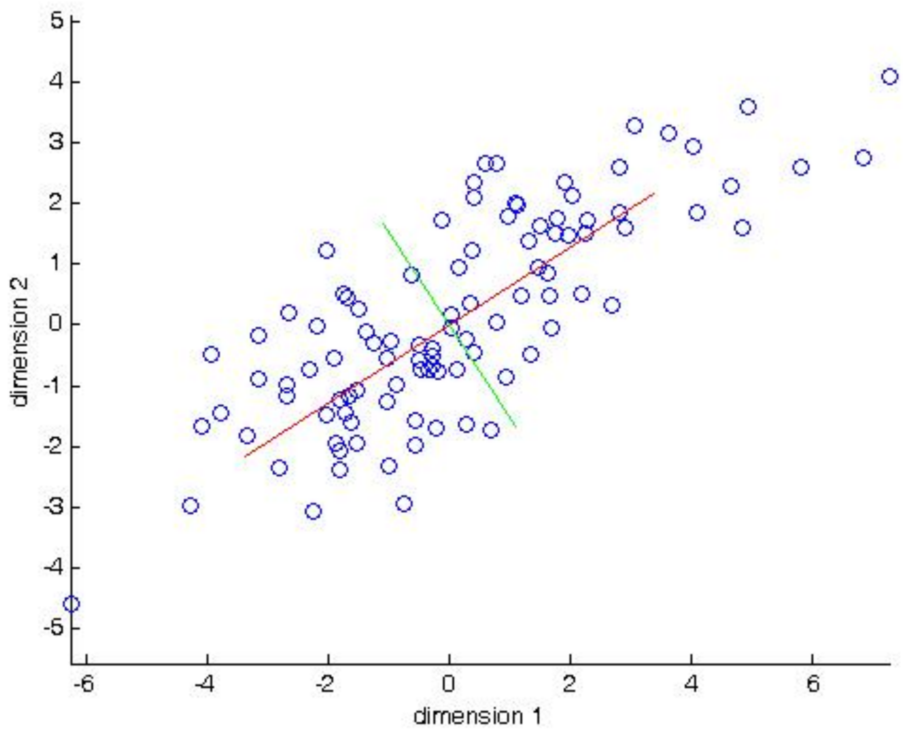
- Even “impoverished” motion data can evoke a strong percept

- Some tracking examples



Principal Component Analysis

The two principal eigenvectors demonstrate the orthogonal directions of maximum variation in the data.



Before:

$$\mathbf{C} = \begin{pmatrix} 0.258 & 0.314 \\ 0.314 & 0.403 \end{pmatrix}$$

After:

$$\mathbf{C} = \begin{pmatrix} 0.518 & 0 \\ 0 & 0.174 \end{pmatrix}$$

Signals and Functions

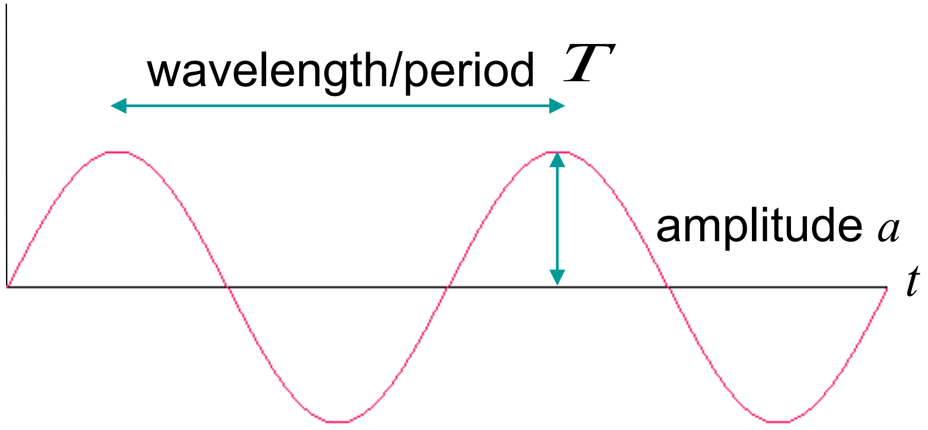
A signal is a physical quantity that is a function of one or more independent variable(s), such as space and/or time.

Data from a Gene pool

Position of a car in a video sequence

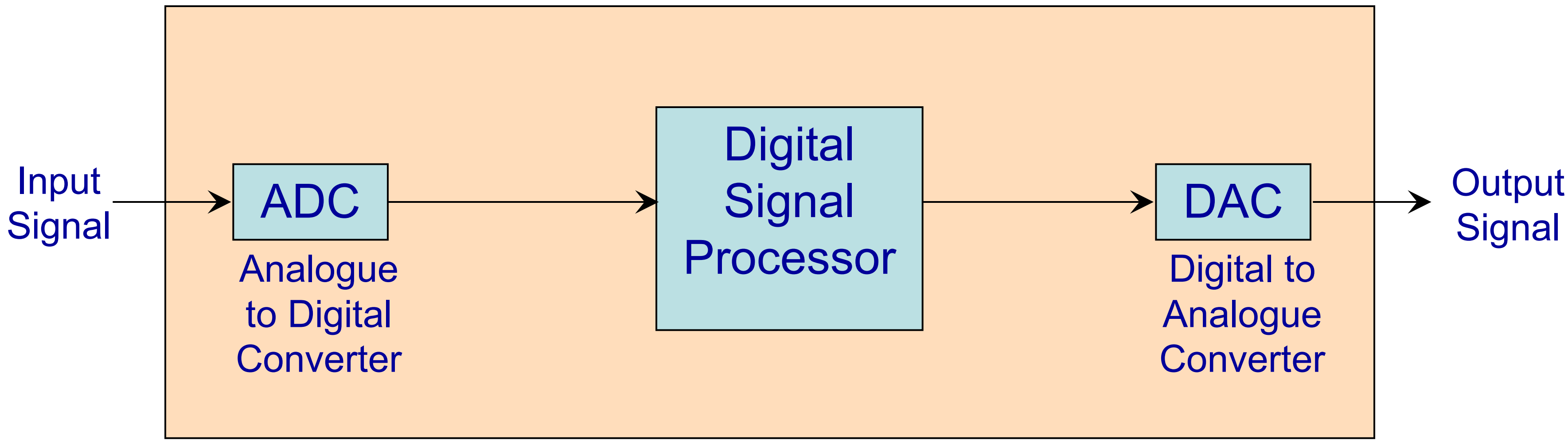
Example signals:

1D signal: $f(t)$
2D signal: $f(x,y)$
3D signal: $f(x,y,t)$ etc.



What is DSP?

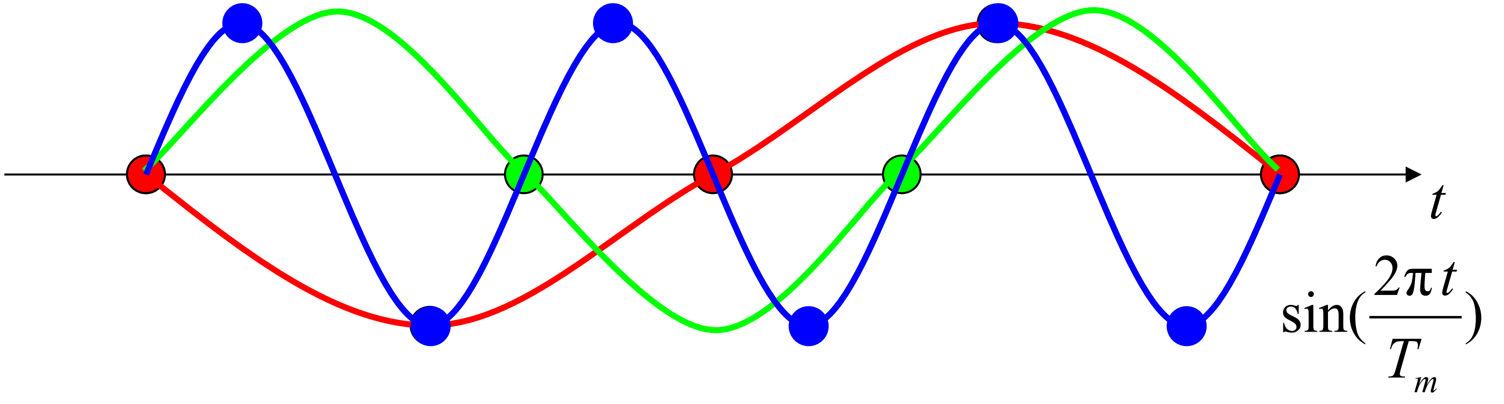
- **Digital Signal Processing** – the processing or manipulation of signals using digital techniques



Shannon's Sampling Theorem

"An analogue signal containing components up to some maximum frequency ***u*** *(Hz) may be completely reconstructed by regularly spread samples, provided the sampling rate is at least* ***2u*** *samples per second"*

Also referred to as the Nyquist criterion: sampling frequency should be at least twice the highest spatial frequency.

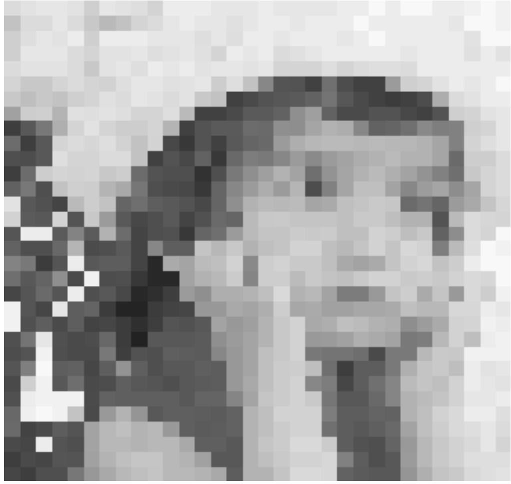


Sampling

The effect of sparser sampling…is ALIASING

256 x256

64x64



32x32

Anti-aliasing achieved by filtering to remove frequencies above Nyquist limit.

Quantization

This results from representing a continuously varying function $f(x)$ with a discrete one using quantization levels

16 levels

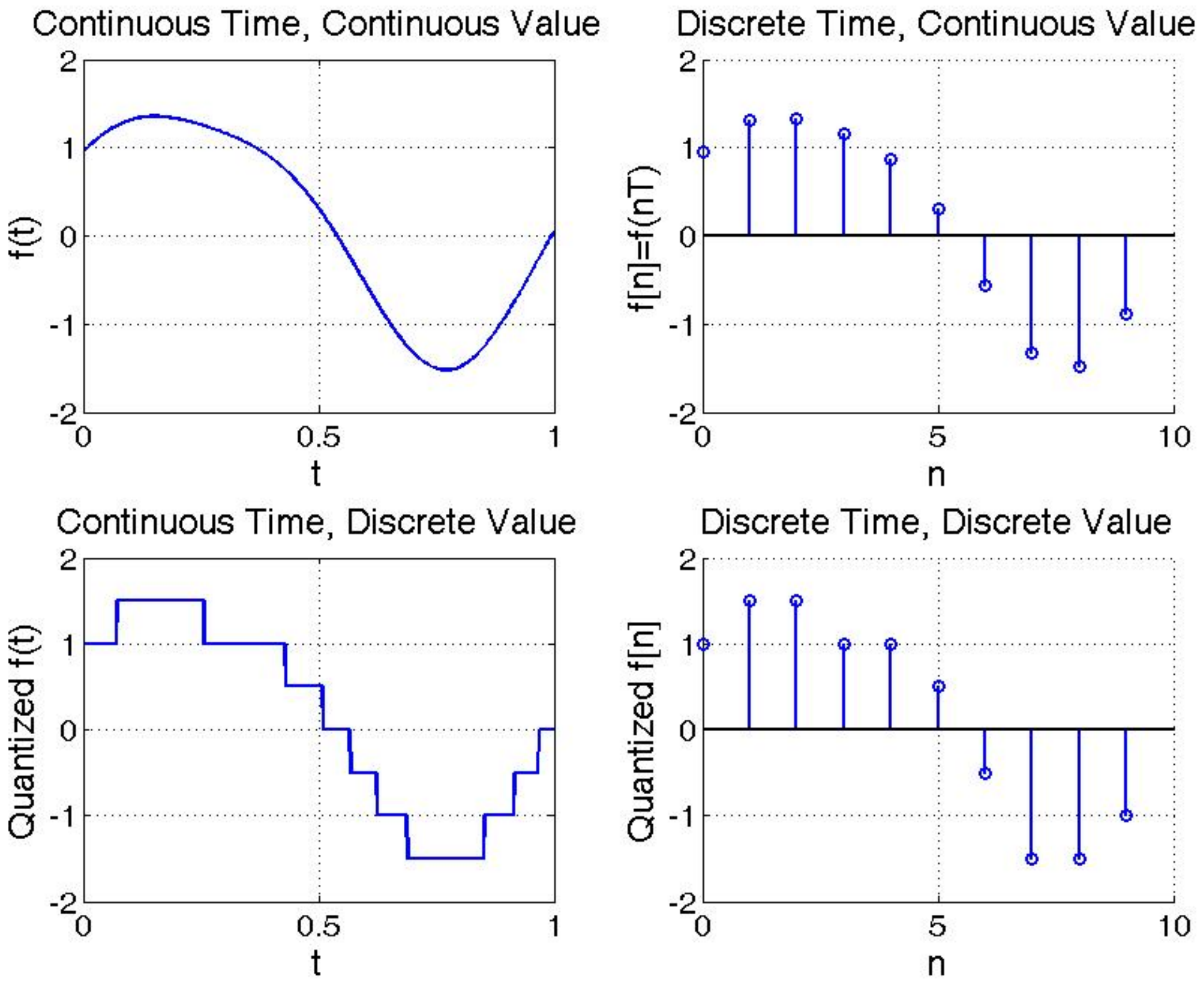
6 levels

2 levels

- Matlab code:

```
F = imread('romina.gif');
[X, map] = gray2ind(F, 16);   // 2, 6, or 16
imview(X, map);
```

Signal Processing



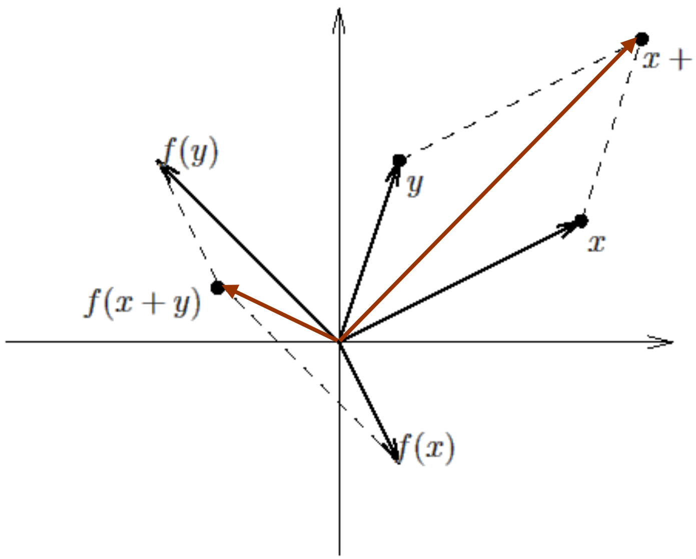
Linear Systems

- For a linear system: output of the linear combination of many input signals is the same linear combination of the outputs → *superposition*

A function f is linear if

- $f(x + y) = f(x) + f(y)$
- $f(\alpha x) = \alpha f(x)$

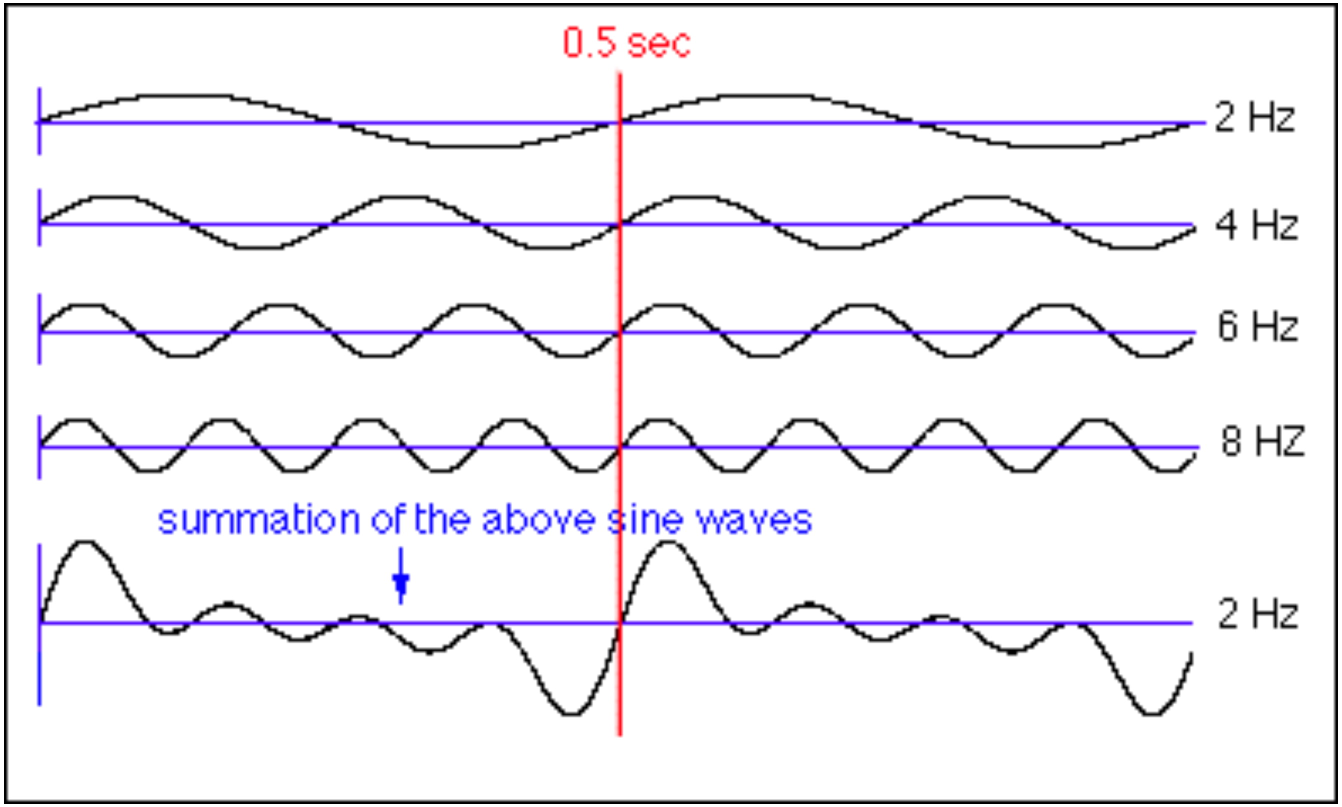
i.e., superposition holds.



Linearity allows us to decompose our input into smaller, elementary objects. Output is the sum of the system's response to these basic objects.

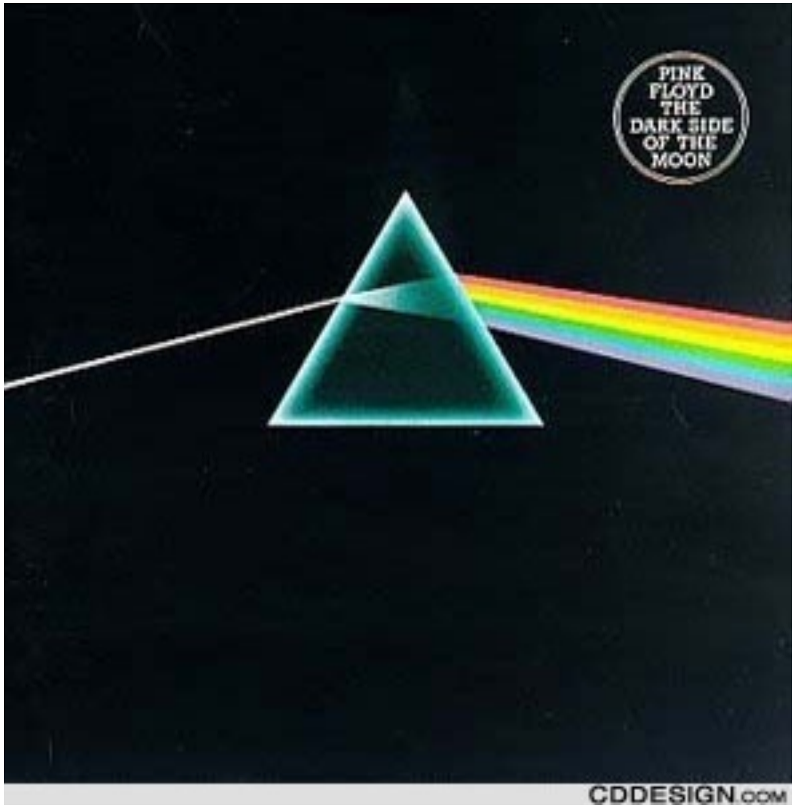
Linear Systems

- For a linear system: output of the linear combination of many input signals is the same linear combination of the outputs → *superposition*



Example: White Light?

White light is made up of variable wavelengths of each component color.



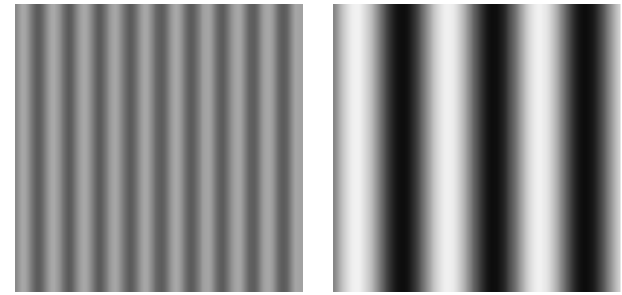
AAAAATAAAAA

0000001000000

$$\delta[n] = \begin{cases} 0, n \neq 0 \\ 1, n = 0 \end{cases}$$

Basic signals…

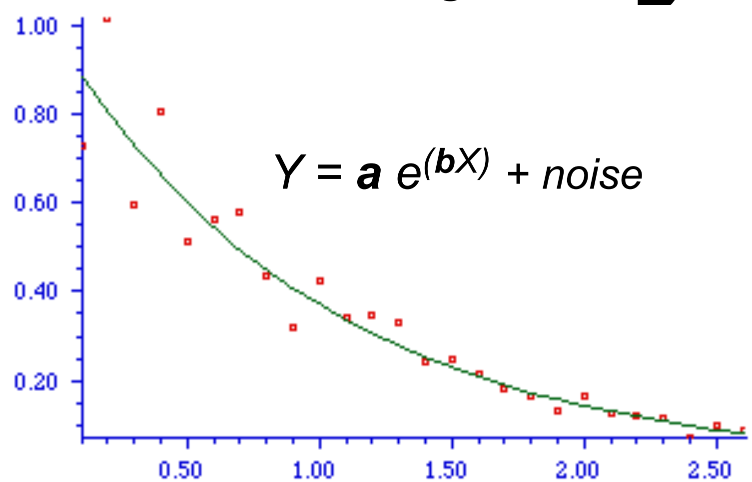
*x = sin(t) = sin(t+2**π**)*

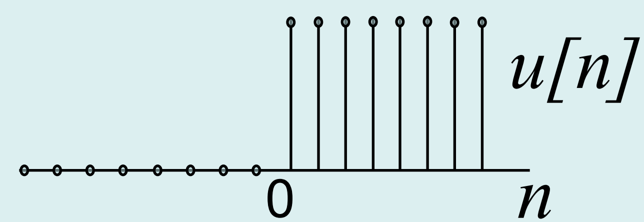


Some basic signals:

- **Unit impulse signal**
- **Unit step signal**
- **Exponential signal**
- **Periodic signal**

All signals can be represented by these basic signals!





$$u[n] = \begin{cases} 0, n < 0 \\ 1, n \geq 0 \end{cases}$$

Overview of next few lectures

Topics covered:

- Fourier Series
- 1D and 2D Fourier Transform
- Convolution
- Feature Selection and Extraction
- PCA