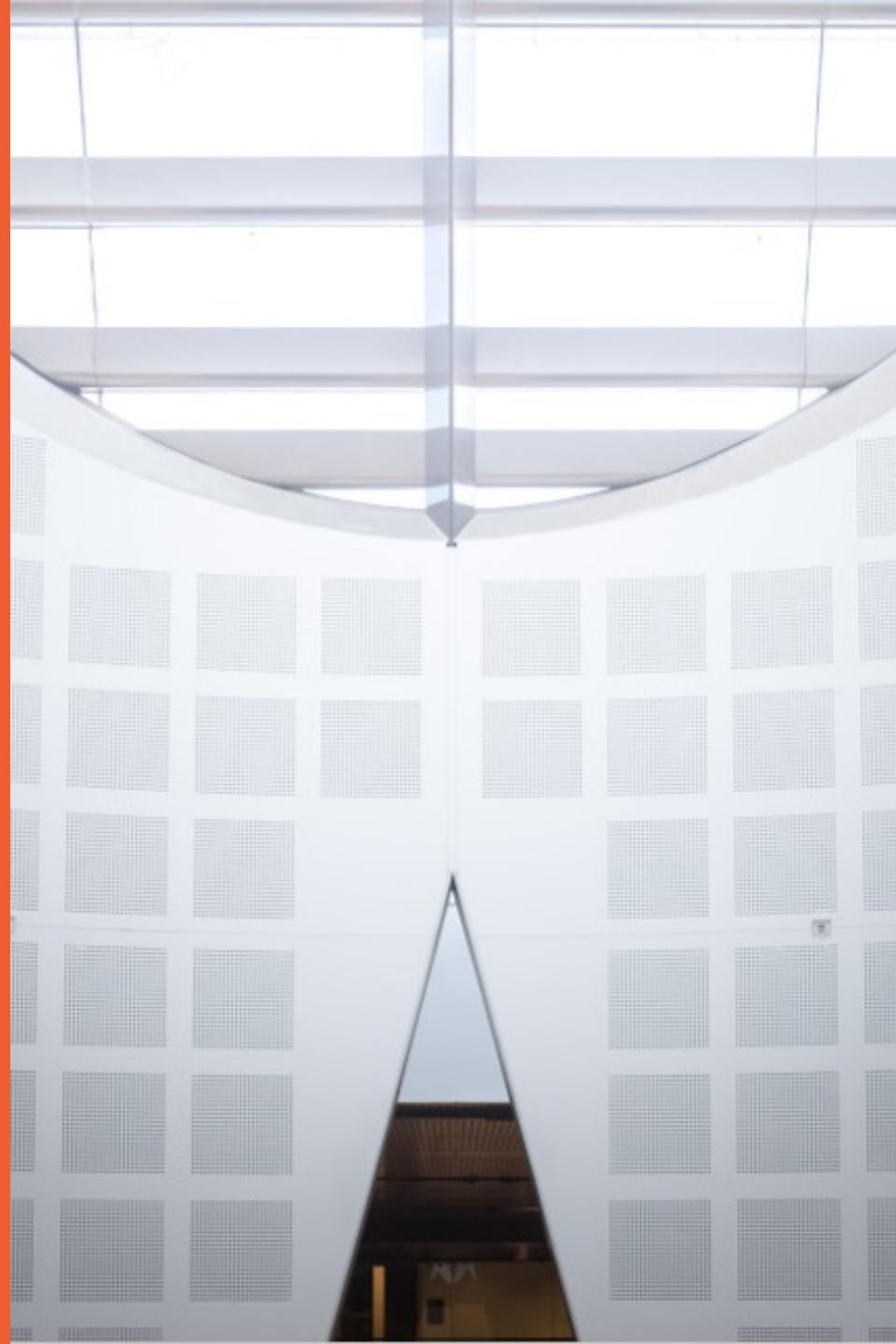


Biomedical Signal Processing

Dr Chang Xu

School of Computer Science

Reference: Healthcare Data Analytics, Chapter 5



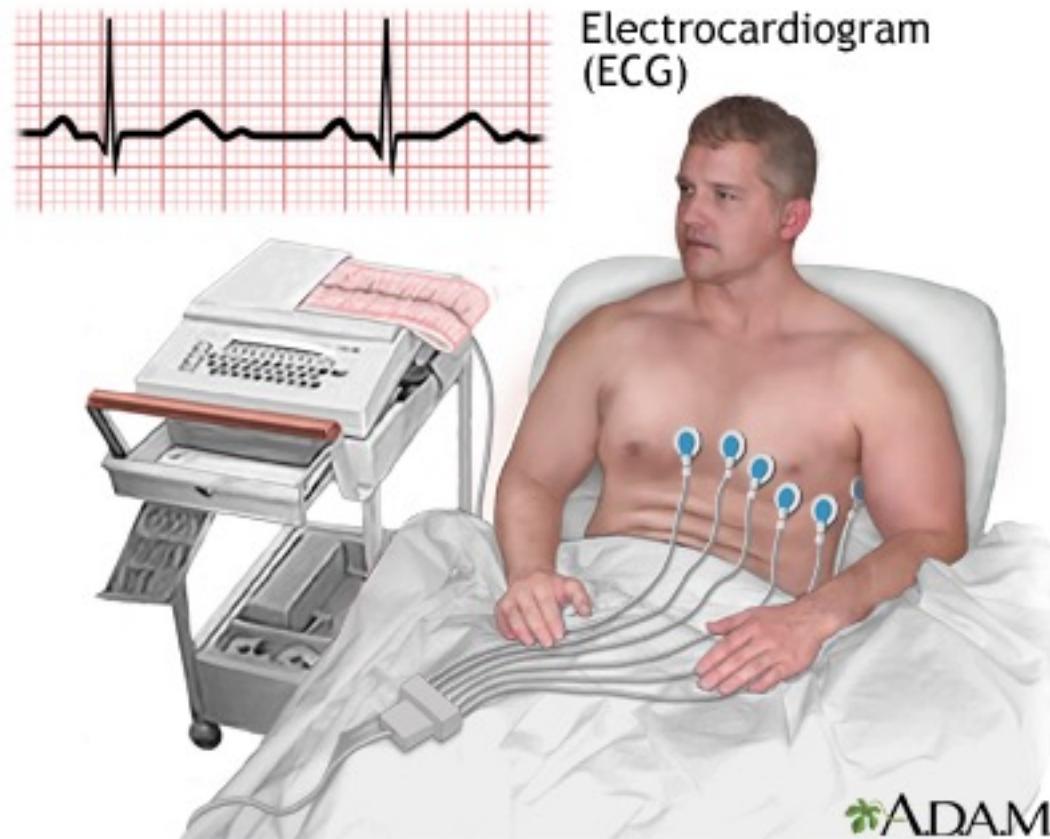
Biopotentials recording

- At the cellular level, the movement of ions like K^+ , Na^+ , Ca^{2+} and Cl^- determines the presence of biopotentials at the level of the cellular membrane. When a stimulus arrives on the membrane, an action potential is generated and is transmitted to the neighboring cells, spreading within the entire tissue or just within some parts of the tissue, depending on the presence of inhibitory channels. At the macroscopic level, the sums of all the action potentials generate a biopotential.
- By placing electrodes on the human body, in specific configurations, a projection of the investigated biopotentials on the measurement direction may be recorded. Basically, every human body tissue has associated an electromagnetic field, which can be theoretically measured.
- However, usually the biopotentials have a very low power (amplitude of μV), thus just a few can be recorded: the electrical activity generated
 - by the cardiac tissue, the electrocardiogram (ECG),
 - by the brain, the electroencephalogram (EEG),
 - by the skeletal muscles, the electromyogram (EMG),
 - by the uterus, electrohysterogram (EHG),
 - by the retina, the electroretinogram (ERG),
 - by the stomach muscles, the electrogastrogram (EGG),
 - by the fetal cardiac tissue, the fetal electrocardiogram (fECG).



Electrocardiogram (ECG)

- **Origin:** Electrocardiogram signals are generated as a consequence of electrical activity of the heart. .
- **Recording methods:** These signals are typically recorded on the skin of the human body.



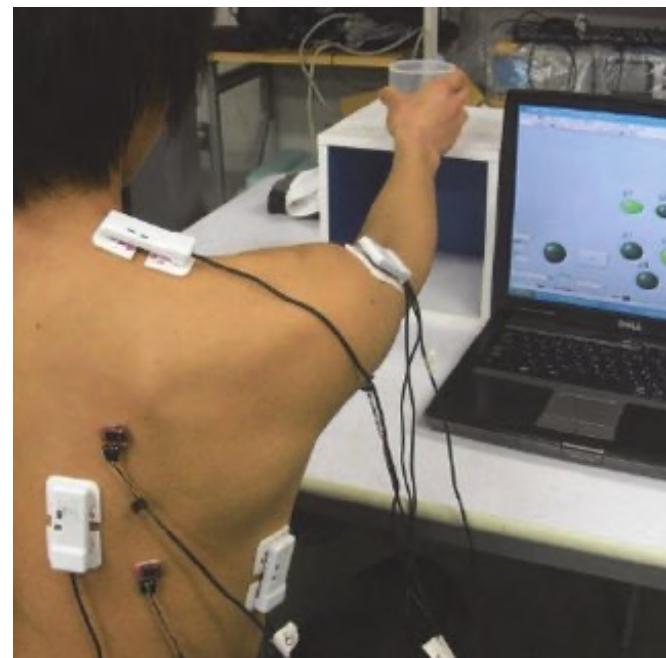
Electroencephalogram (EEG)

- **Origin:** The electroencephalogram (EEG) represents the electrical activity occurring at the surface of the brain.
- **Recording methods:** EEG recorded from the surface of the scalp is in major part generated by the synchronous activity of neurons on the cerebral cortex
- **Characteristic:** EEG is characterized by a good temporal resolution on a submillisecond scale, but is poor in terms of spatial resolution.



Electromyogram (EMG)

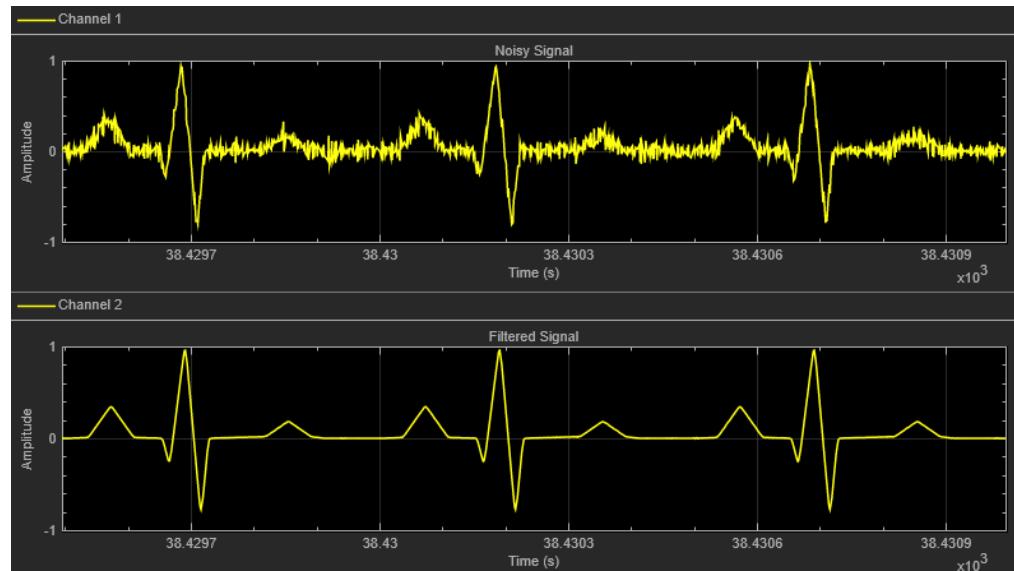
- **Origin:** Spatial-temporal summation of the motor unit action potentials of all the active motor units gives rise to an electromyogram (EMG) of the muscle.
- **Recording methods:** EMG can be recorded by placing needle electrodes on the surface of the body.



Noisy collected biopotential signals

All these signals contain diagnostic information about the health status of the source tissue, which can be extracted by applying different signal processing methods.

Nevertheless, the raw signals recorded after the measurement cannot be used directly to evaluate the health status of the source tissue because it contains also noise, which in most of the cases hampers the extraction of diagnostic information.



Introduction

- Many times the signals acquired from the sensor need to be teased out from the raw data so that meaningful information or features could be extracted.
 - This needs the application of signal processing algorithms so that a constituent waveform feature could be separated in the presence of noise.
- The processing and interpretation of physiological signals sometimes exhibit challenges because of the low signal-to-noise ratio (SNR) or because of interdependency of the physiological systems
 - The goal of this lecture is to present an overview of various signal processing techniques used for processing biomedical signals.

Adaptive Filtering

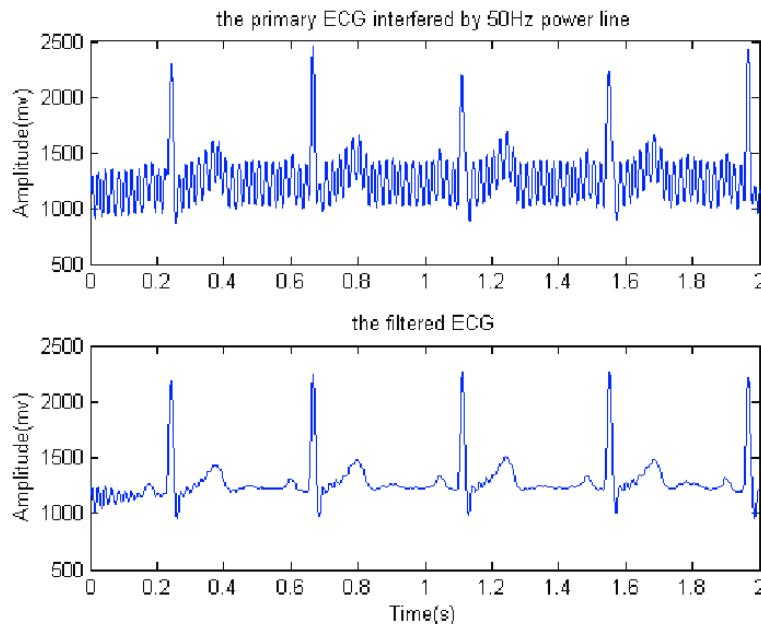
ECG signal Analysis

The ECG waveform is corrupted with several sources of noise and before any feature could be extracted, a proper signal conditioning is necessary.

- (a) Power line interference
- (b) Electrode contact noise
- (c) Motion artifact
- (d) Muscle contraction (electromyographic, EMG)
- (e) Baseline drift and ECG amplitude modulation with respiration
- (f) Instrumentation noise generated during signal acquisition
- (g) Electrosurgical noises, and many other less significant noises

Power Line Interference

- Power line interference consists of 60 Hz/50 Hz pick up, depending upon where the instrument is operated (United States or Europe/Asia).
- By properly shielding the cables as well as the device, effects due to AC interference can be minimized to a certain extent. However, these alternating currents still manage to interfere with the signals of interest by flowing through the system ground, thereby getting picked by the tissue or electrode

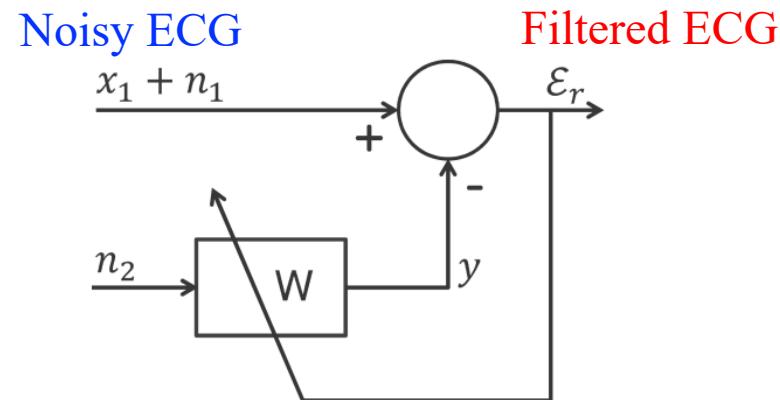


Since signal to noise ratio (SNR) of electrocardiogram (ECG) signal is very low, the interfering frequency at 50 Hz may overwhelm the source signal.

Least Mean Square Adaptive Filter

- main signal is the noise corrupted ECG signal $(x_1 + n_1)$, where n_1 is the additive noise and x_1 is the pure ECG signal.
- n_2 is the reference signal. The requirement is that the noise n_2 is correlated in some way to noise n_1 .
- Minimizing the filter error output ε_r , that is the best least-squares estimate of the signal x_1

$$\varepsilon_r^2 = (x_1 + n_1 - y)^2 = (x_1 + n_1)^2 - 2y(x_1 + n_1) + y^2$$



$$y(t) = \sum_{t=0}^{T-1} w(t) n_2(t)$$

- With the filter W trained, it can be applied to filter new data with similar noise.

Least Mean Square Adaptive Filter

Filter error

$$\begin{aligned}\varepsilon_r^2 &= (x_1 + n_1 - y)^2 = (x_1 + n_1)^2 - 2y(x_1 + n_1) + y^2 \\ &= y^2 - 2y(x_1 + n_1)\end{aligned}$$

Calculate the gradient

$$\frac{\partial \varepsilon_r^2}{\partial w} = 2\varepsilon_r \frac{\partial \varepsilon_r}{\partial w} = 2\varepsilon_r (-n_2) = -2\varepsilon_r n_2$$

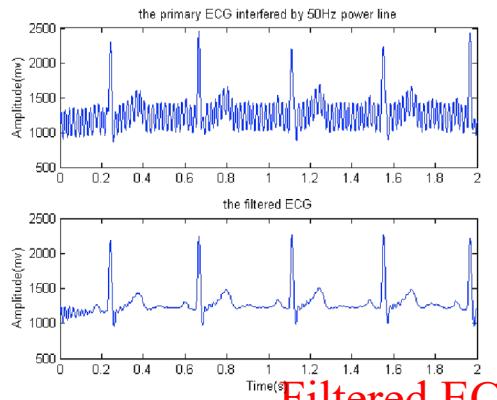
The update equation (gradient descent) is

$$w \leftarrow w + 2\mu\varepsilon_r n_2$$

Step 1: Filtering, calculate y

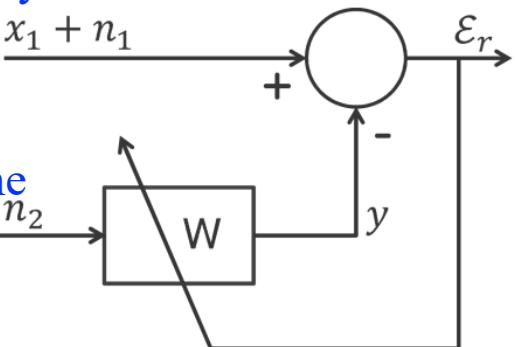
Step 2: Error formation, calculate $\varepsilon_r = (x_1 + n_1 - y)$

Step 3: Coefficient update, $w \leftarrow w + 2\mu\varepsilon_r n_2$



Noisy ECG

Filtered ECG



$$y(t) = \sum_{t=0}^{T-1} w(t) n_2(t)$$

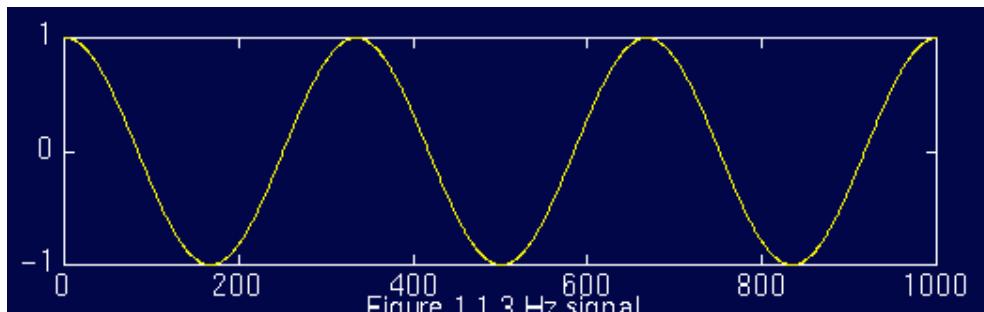
μ is the step-size control parameter

Wavelet Transform

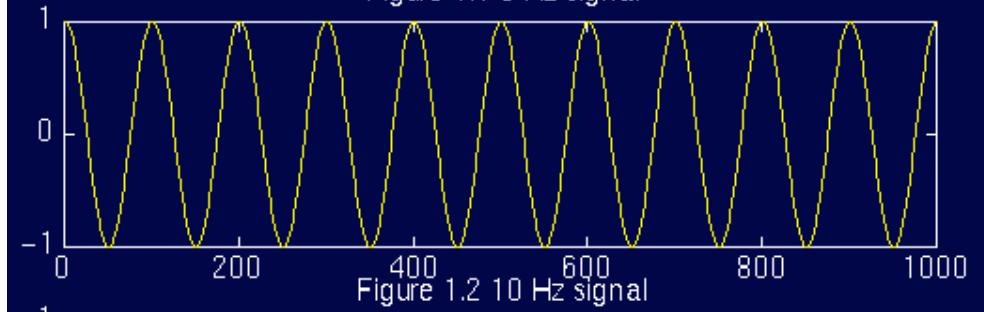
The frequency

The frequency is measured in cycles/second, or with a more common name, in "Hertz".

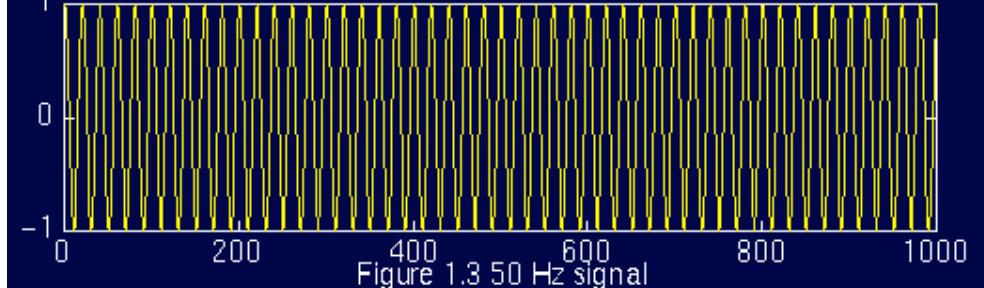
For example the electric power we use in our daily life in the AU is 50 Hz. This means that if you try to plot the electric current, it will be a sine wave passing through the same point 50 times in 1 second.



a sine wave at 3 Hz



a sine wave at 10 Hz

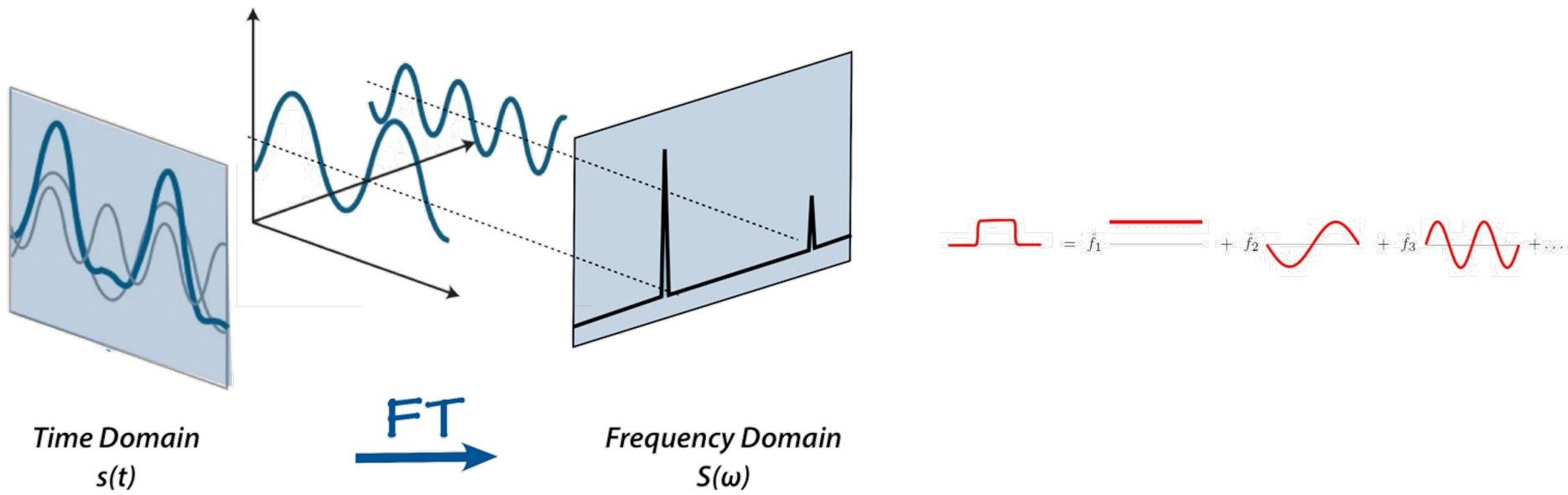


a sine wave at 50 Hz

Fourier transform

How do we measure frequency, or how do we find the frequency content of a signal?

The answer is **FOURIER TRANSFORM (FT)**.



If the FT of a signal in time domain is taken, the frequency-amplitude representation of that signal is obtained. In other words, we now have a plot with one axis being the frequency and the other being the amplitude. This plot tells us how much of each frequency exists in our signal.

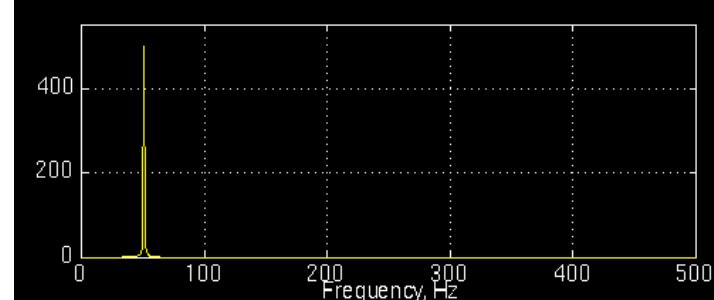
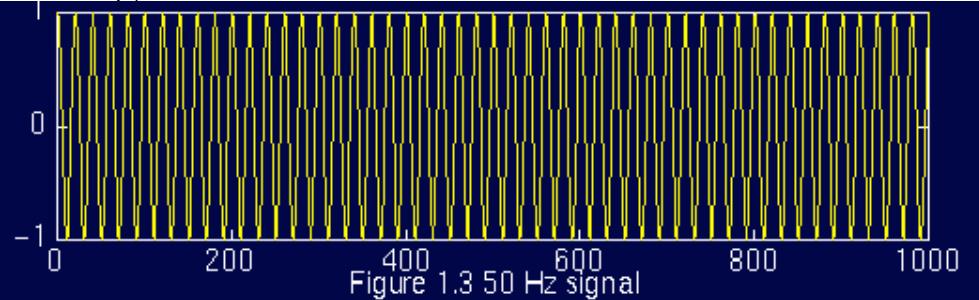


Figure 1.3 50 Hz signal

Why do we need the frequency information?

Often times, the information that cannot be readily seen in the time-domain can be seen in the frequency domain.

Suppose we are looking at an ECG signal (ElectroCardioGraphy, graphical recording of heart's electrical activity). The typical shape of a healthy ECG signal is well known to cardiologists. Any significant deviation from that shape is usually considered to be a symptom of a pathological condition.

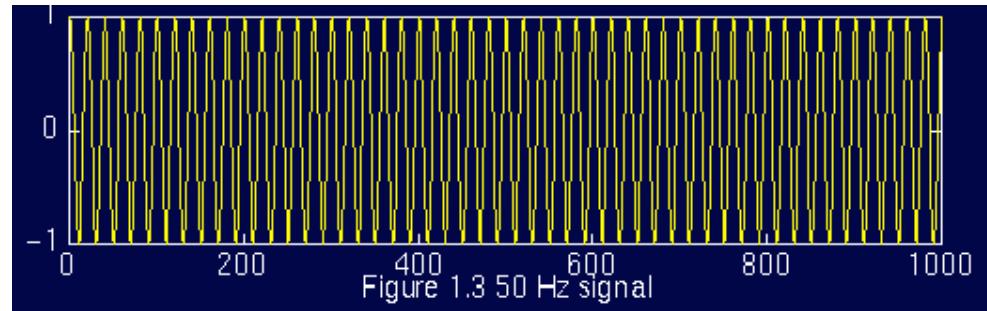
This pathological condition, however, may not always be quite obvious in the original time-domain signal.

Recently, the new computerized ECG recorders/analyzers also utilize the frequency information to decide whether a pathological condition exists. A pathological condition can sometimes be diagnosed more easily when the frequency content of the signal is analyzed.

Disadvantages of FT

Recall that the FT gives the frequency information of the signal, which means that it tells us how much of each frequency exists in the signal, but it does not tell us **when in time** these frequency components exist. This information is not required when the signal is so-called **stationary** .

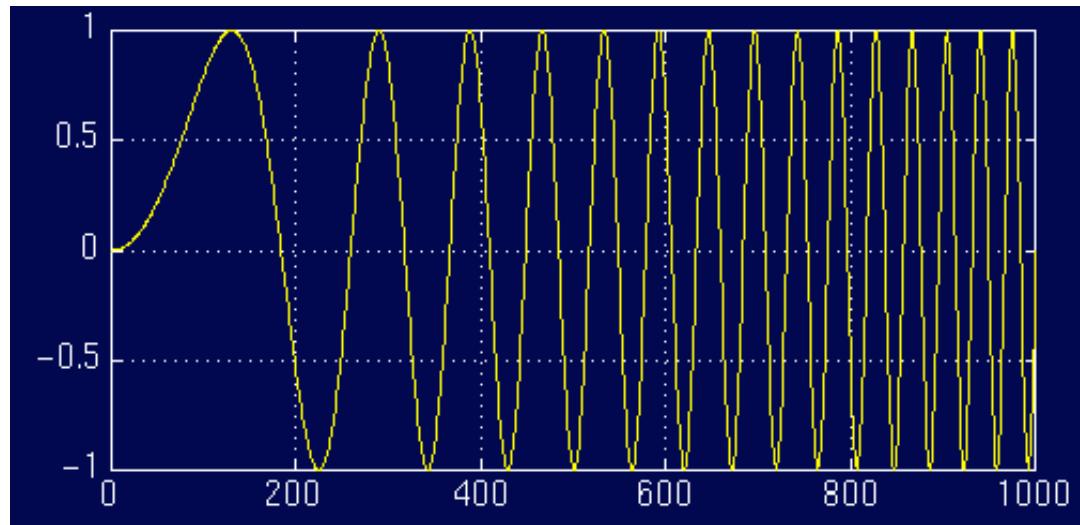
Signals whose frequency content do not change in time are called stationary signals . In other words, the frequency content of stationary signals do not change in time, e.g., for the electric current that we use in our houses (50Hz).



In this case, one does not need to know at what times frequency components exist , since all frequency components exist at all times !!! .

How about a non-stationary signal?

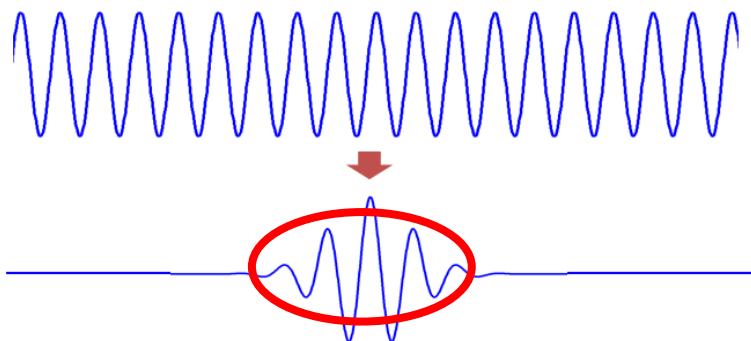
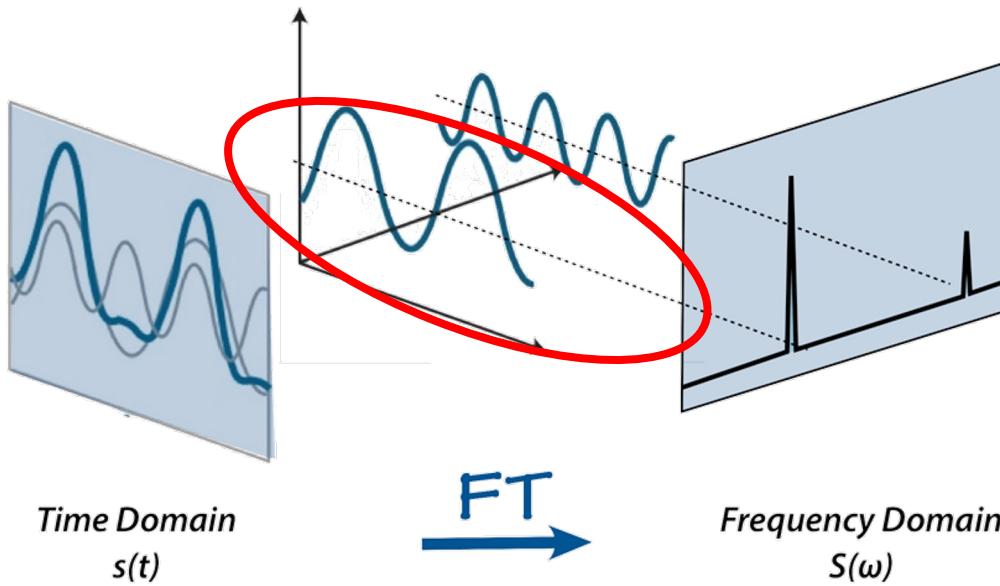
A signal whose frequency constantly changes in time.



FT gives what frequency components (spectral components) exist in the signal.
Nothing more, nothing less.

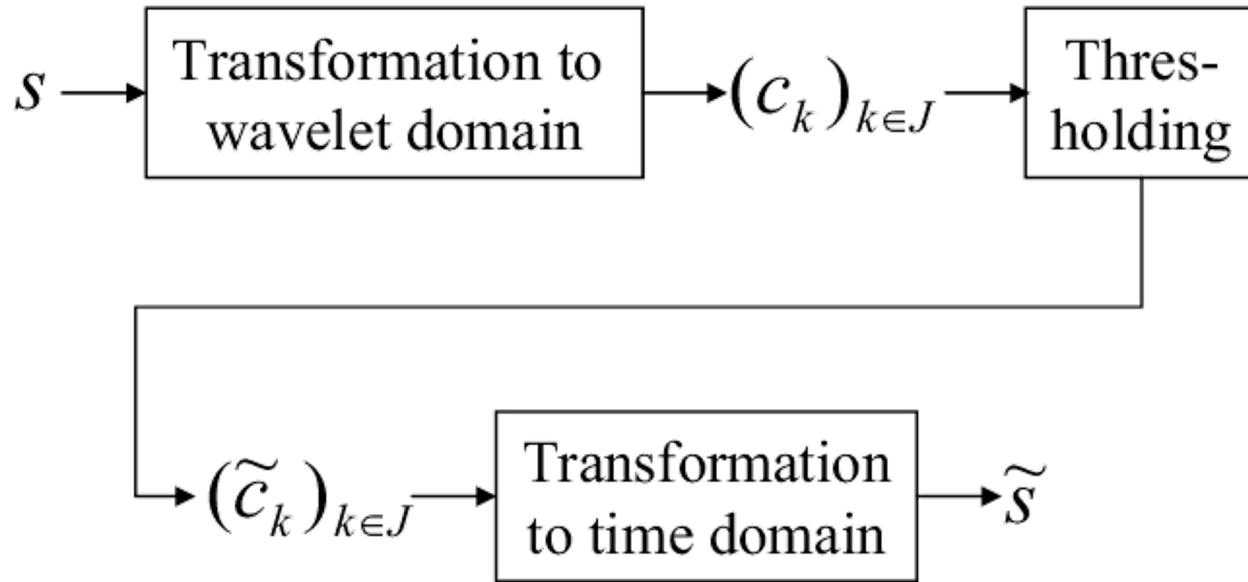
Wavelet transform (WT) can help with the non-stationary signal.

Wavelet transform (WT)



The term **wavelet** means a **small wave**. The smallness refers to the condition that this (window) function is of finite length (compactly supported).

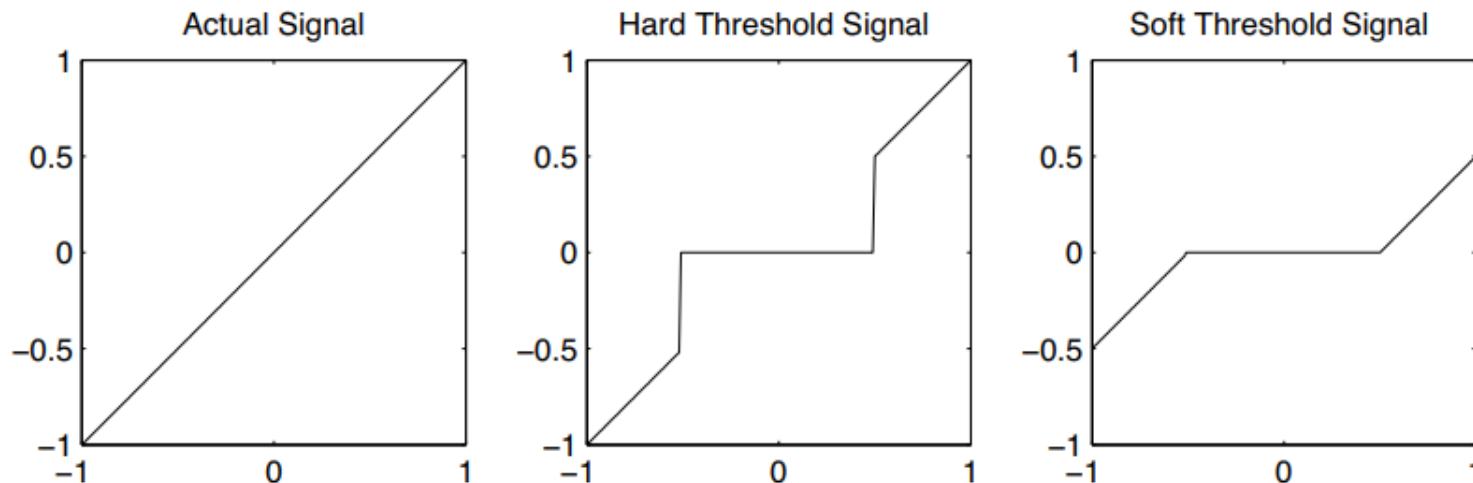
Denoising with Wavelet



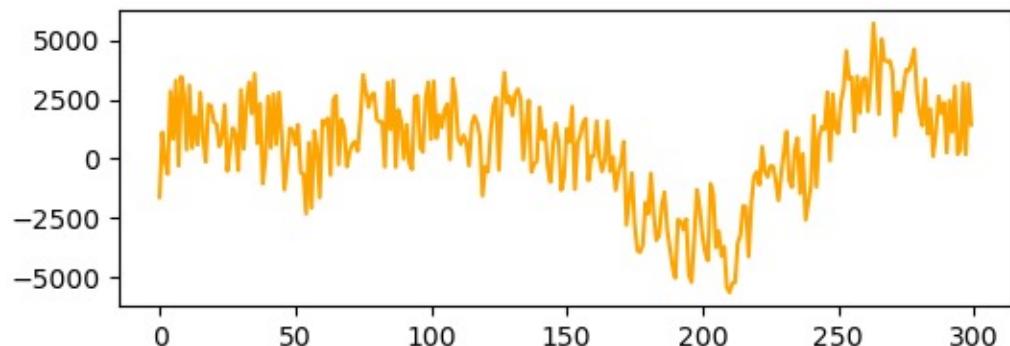
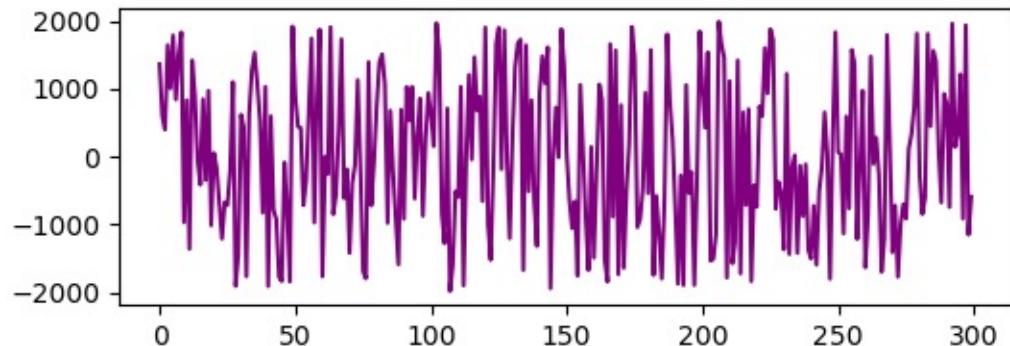
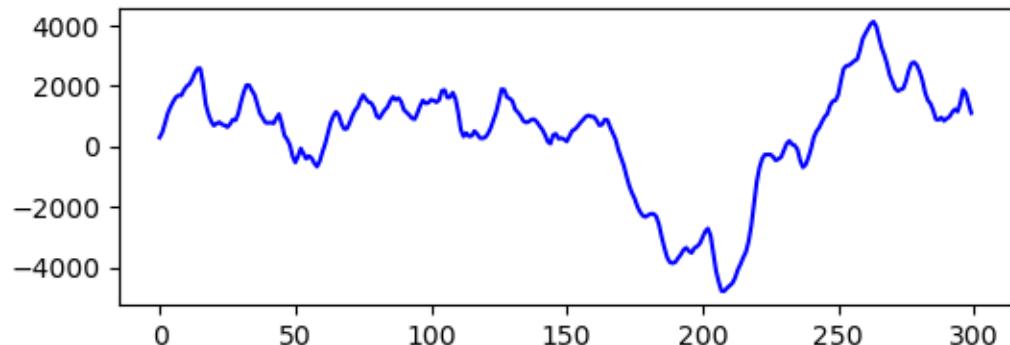
- When DWT is applied to a noisy ECG signal, the noise-free component of the signal (true signal) will be concentrated in a small number of **larger coefficients**, while the noise will be distributed as **smaller coefficients**.
- By applying simple threshold to the smaller coefficients and by performing the inverse wavelet transform (IDWT), noise-free ECG reconstruction can be obtained

Denoising with Wavelet

- The selection of a thresholding scheme is a problem of significant interest in wavelet thresholding. In general, the thresholding methods are categorized into two types, namely the hard thresholding and the soft thresholding.
 - The hard threshold function tends to have bigger variance and is unstable (sensitive to even small changes in the signal)
 - In contrary, soft thresholding function is much more stable than hard thresholding and tends to have a bigger bias due to the shrinkage of larger wavelet coefficients.

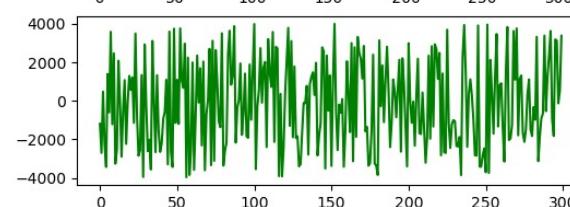
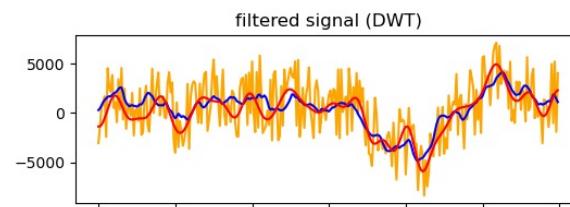
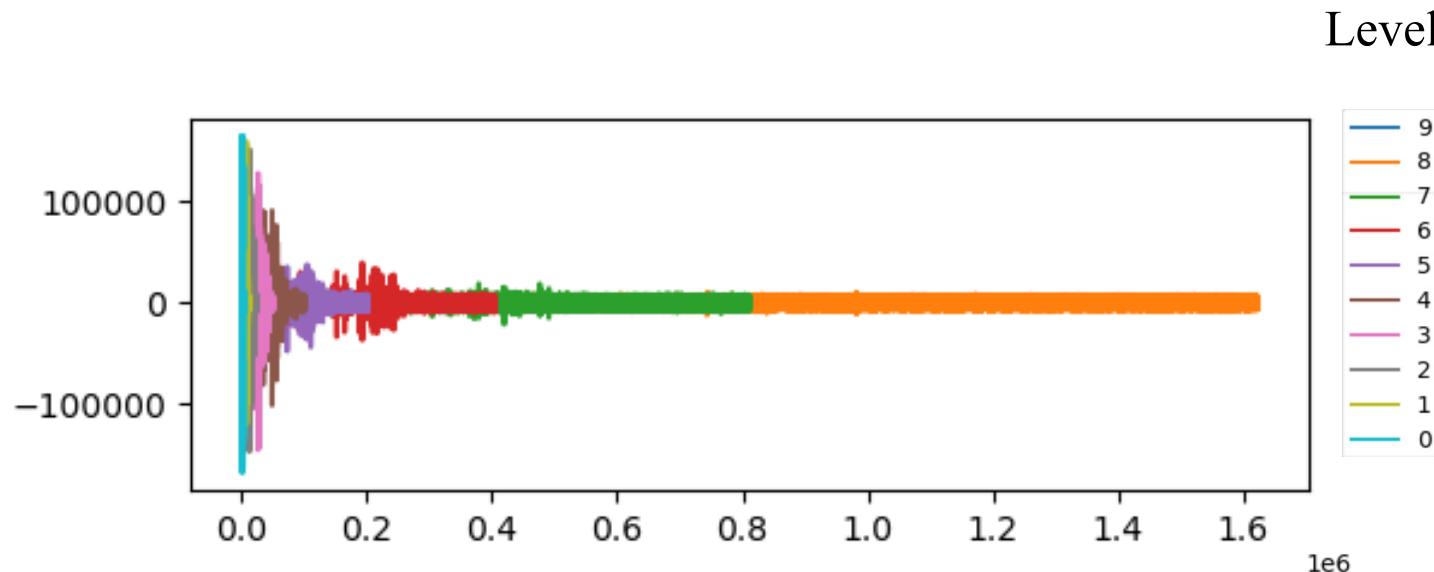


Demo of denoising with Wavelet



Demo of denoising with Wavelet

- Frequency density of extracted wavelets and the results

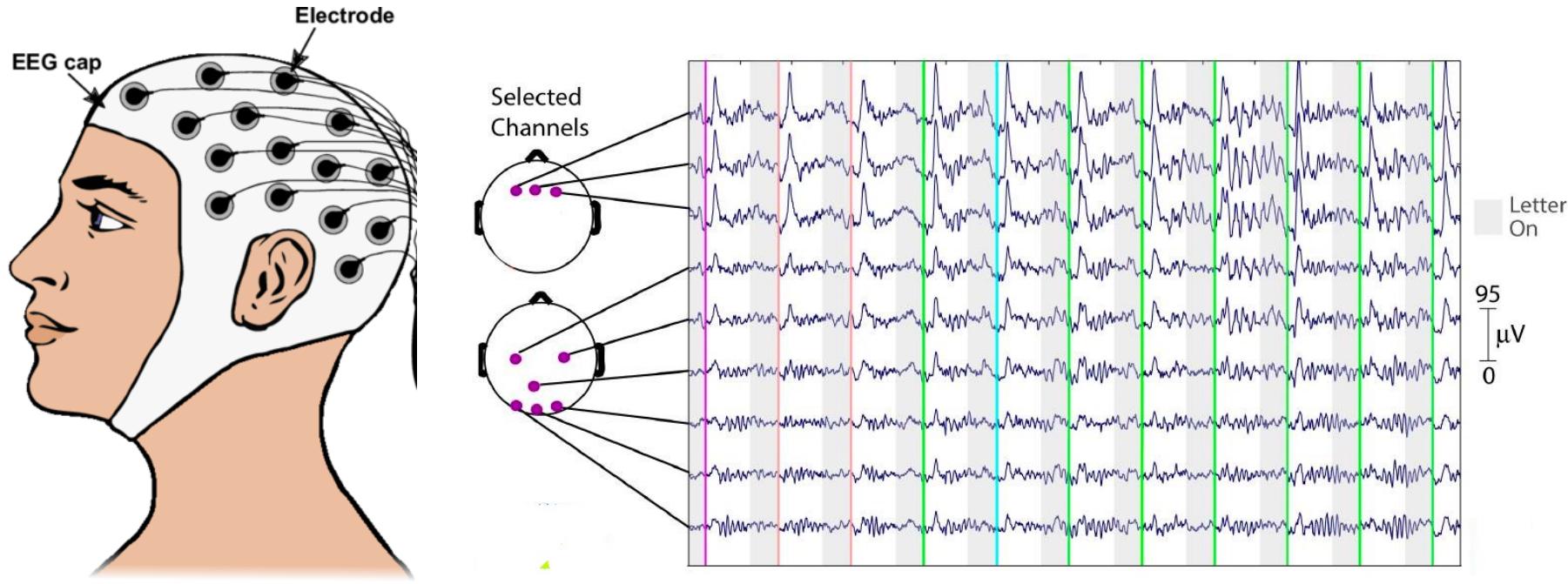


Red: recovered signal
Blue: ground-truth signal

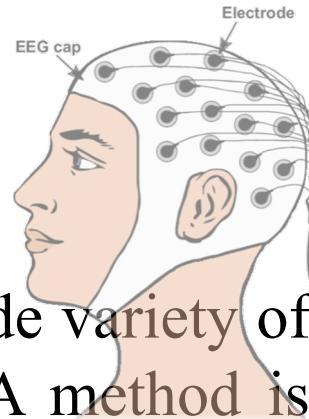
Independent Component Analysis

Removing Artifacts from EEG

Contamination of EEG activity by eye movements, blinks, muscle, heart and line noise is a serious problem for EEG interpretation and analysis.



How to remove artifacts from EEG recordings?

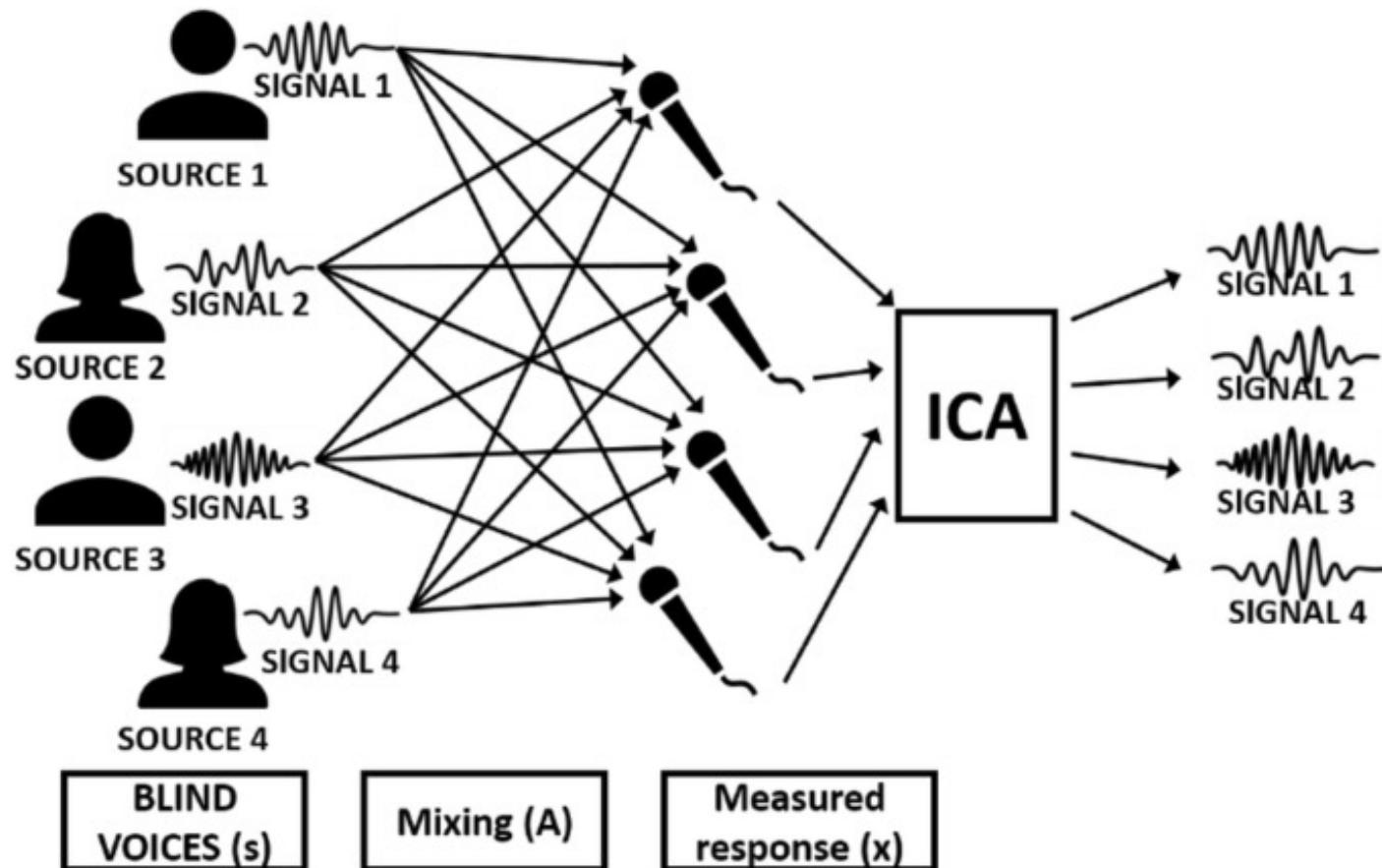


Independent Component Analysis (ICA)

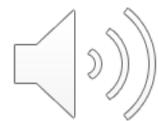
ICA-based artifact correction can separate and remove a wide variety of artifacts from EEG data by linear decomposition. The ICA method is based on the **assumptions** that the time series recorded on the scalp:

- are spatially stable **mixtures** of the activities of temporally independent cerebral and artifactual sources, that
- the summation of potentials arising from different parts of the brain, scalp, and body is **linear** at the electrodes, and that
- propagation delays from the sources to the electrodes are negligible.

Explanation of ICA technique, using the cocktail party problem as an example.



(Guitar)



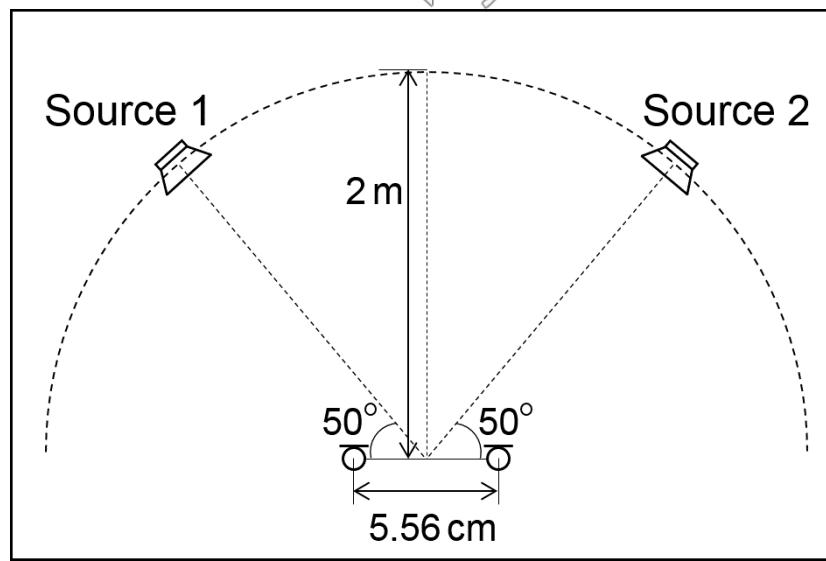
(Vocals)



ICA

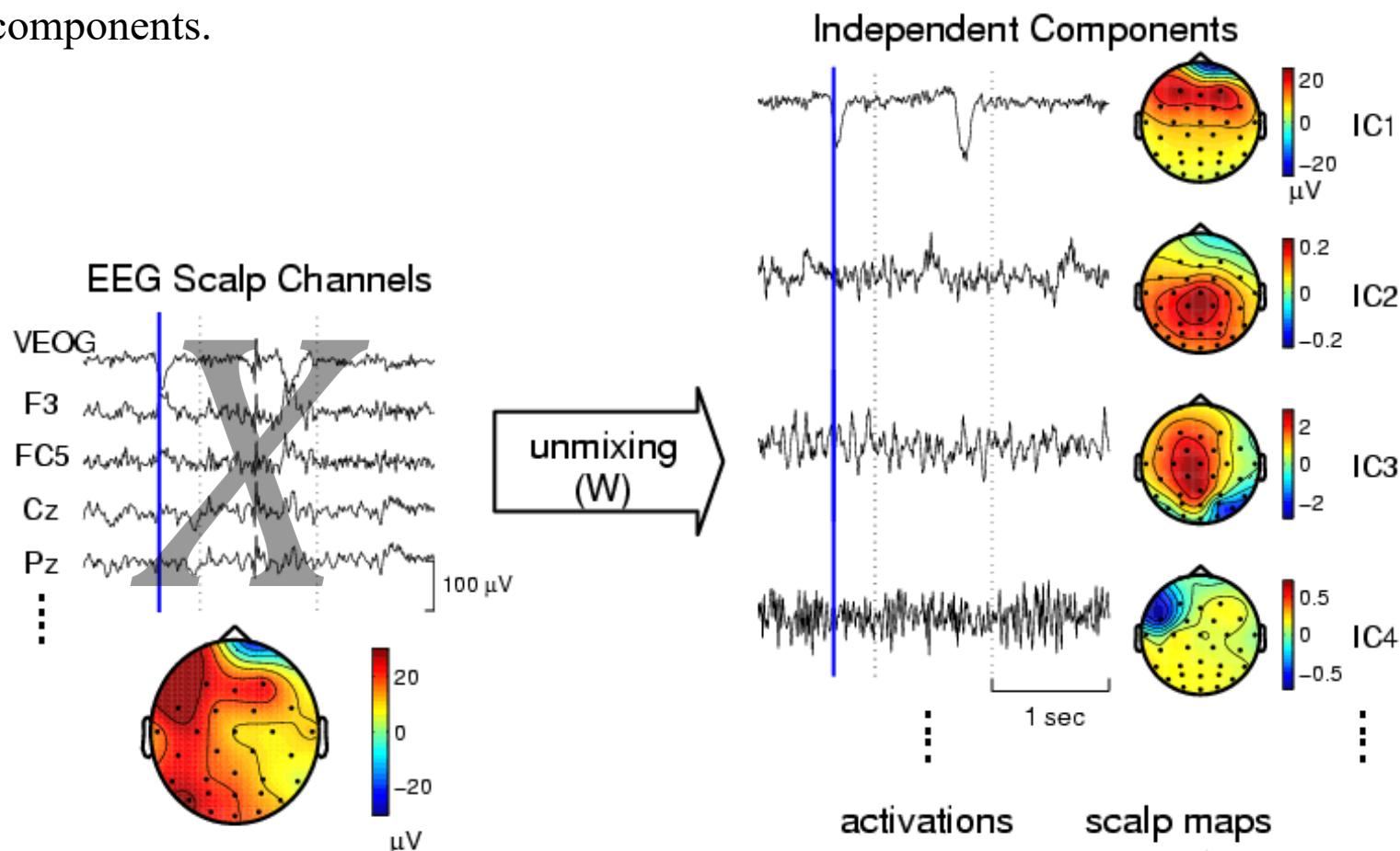


(Guitar, Vocals)



ICA decomposition

ICA decomposition is to find an 'unmixing' matrix, W , which decomposes or linearly unmixes the multi-channel scalp data into a sum of temporally independent and spatially fixed components.

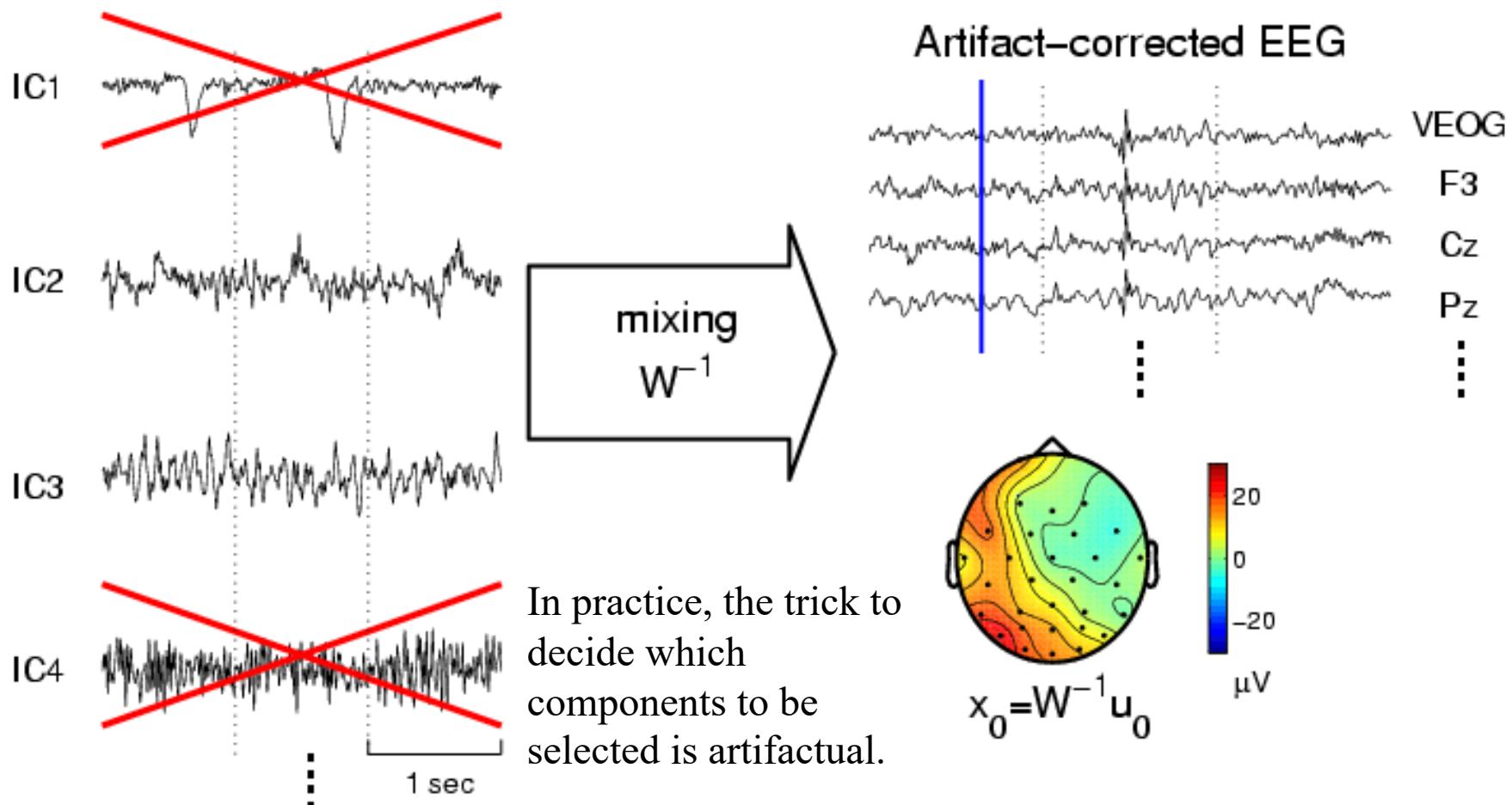


The rows of the input matrix, X , are EEG signals recorded at different electrodes and the columns are measurements recorded at different time points.

$$\begin{array}{ll} \text{activations} & \text{scalp maps} \\ (\mathbf{u} = \mathbf{WX}) & (\mathbf{W}^{-1}) \end{array}$$

The rows of the output data matrix, $\mathbf{U} = \mathbf{WX}$, are time courses of activation of the ICA components.

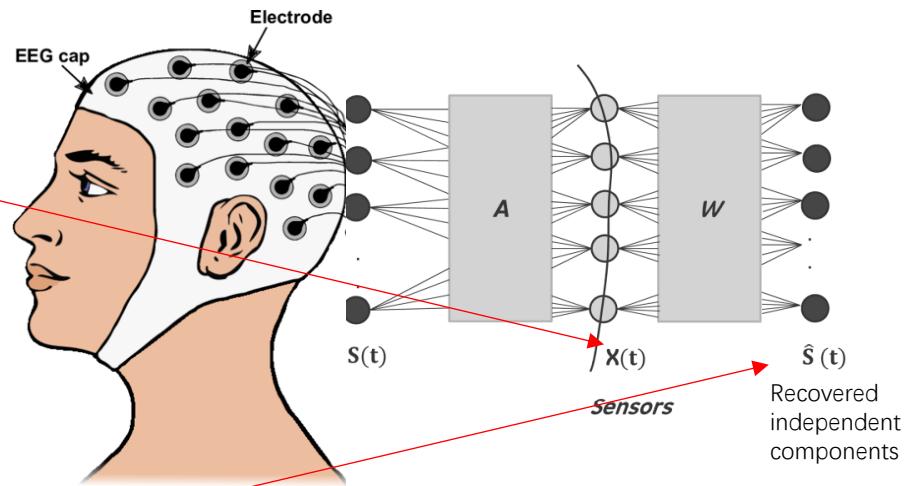
Summed Projection of Selected Components



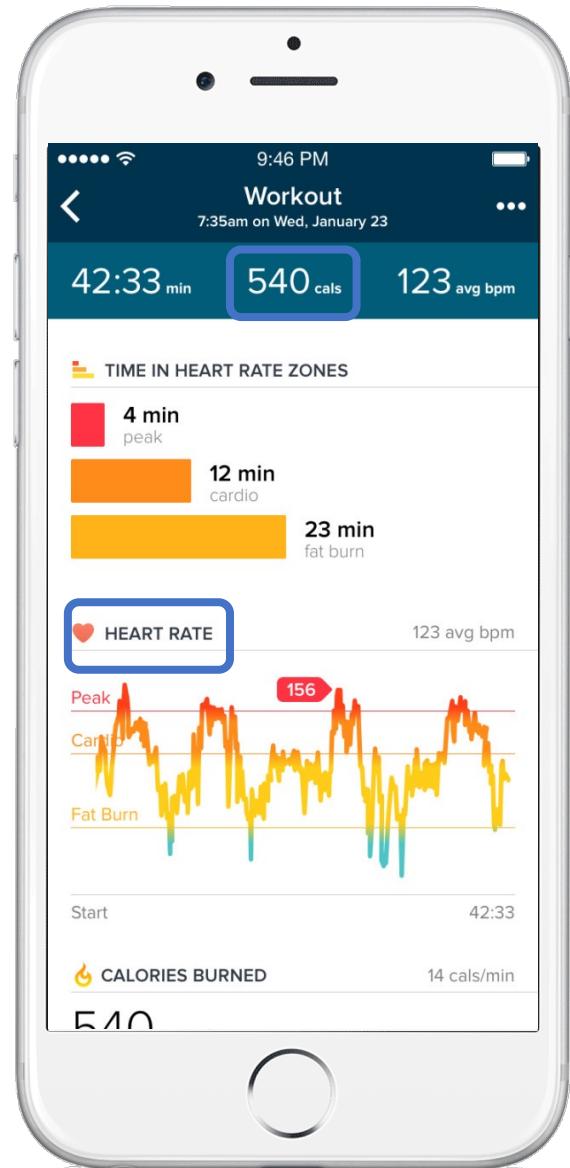
For each component, the amplitudes of scalp maps (given by the individually scaled color bars of the *right panel*) give the size of the component projections at the time point marked by the *vertical blue line*.

FastICA

- FastICA:
 - Input: X
 - Step 1: centering
 - Centering involves subtracting the signal with its mean value so as to make a zero mean variable. Let the result be \hat{X}
 - Step 2: whitening
 - Transforming the \hat{X} into \tilde{X} , so that its components are uncorrelated and their variances equal unity.
 - Step 3: optimize for **non-Gaussianity**
 - Relies on an iterative algorithm that finds W such that
 - $\hat{S} = W^T \tilde{X}$

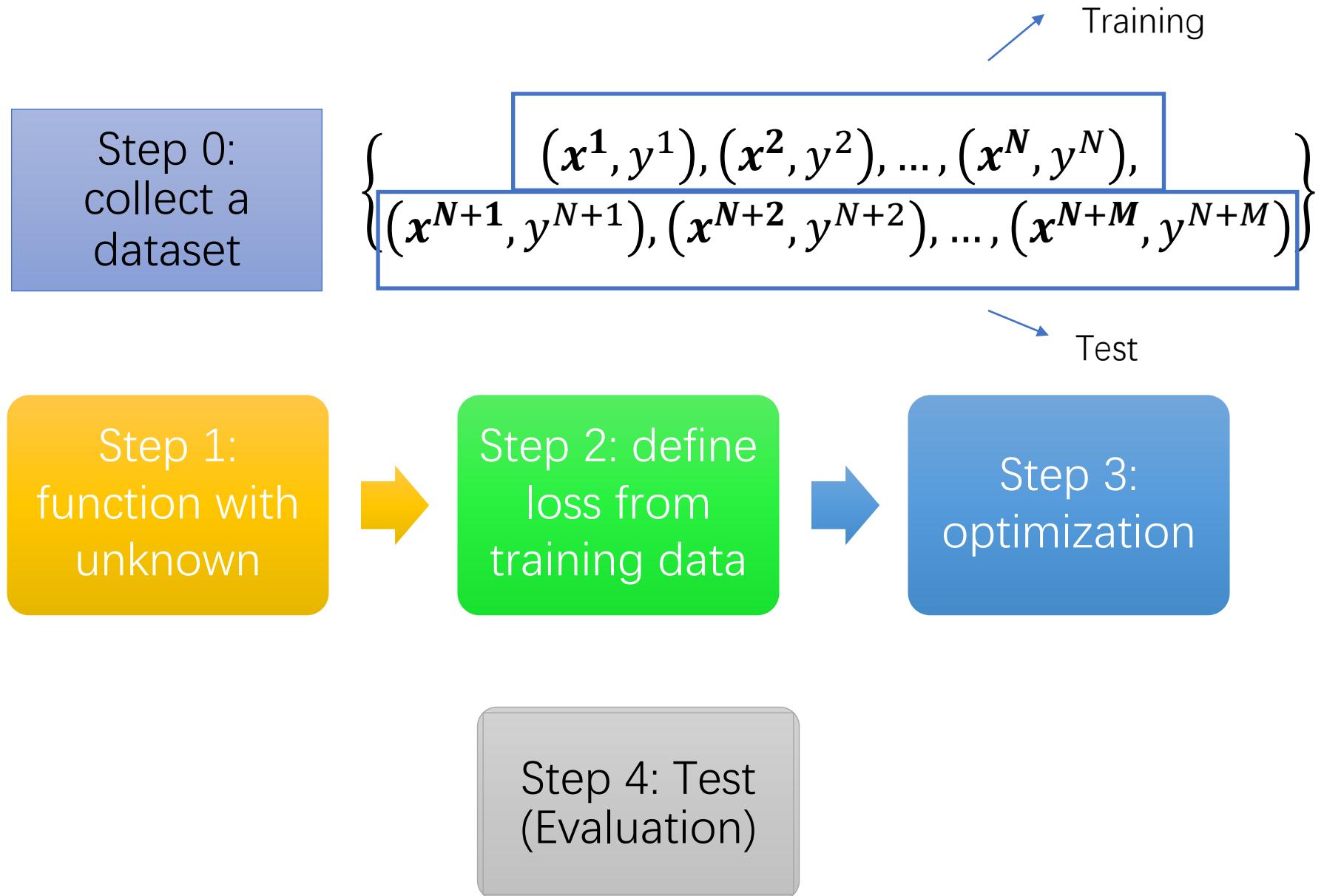


Estimate how many calories do you burn?



- This is a **regression** problem.
- Input data $x \in \mathbb{R}^4$:
 - Duration of exercise
 - Average heart rate
 - Your weight in kilograms
 - Your age
- Target output y :
 - Number of calories burned

Machine Learning is so simple

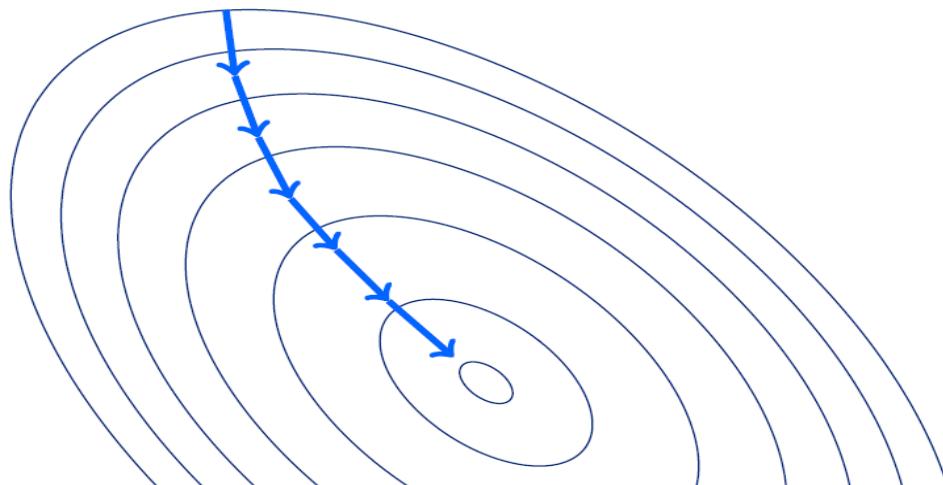


Gradient descent

- $J(\theta, \mathcal{X})$ is the **objective function**, θ are the parameters, \mathcal{X} is the training dataset. We want to optimize $J(\theta, \mathcal{X})$ by:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta, \mathcal{X})$$

where $\nabla_{\theta} J(\theta, \mathcal{X})$ is the gradient of objective function w.r.t. the parameters. η is the learning rate.



Gradient descent

- **Batch gradient descent:**

Compute the gradient for the **entire** training dataset.

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta, \mathcal{X}^{(1:end)})$$

- **Stochastic gradient descent:**

Compute the gradient for **each** training example.

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta, \mathcal{X}^{(i)})$$

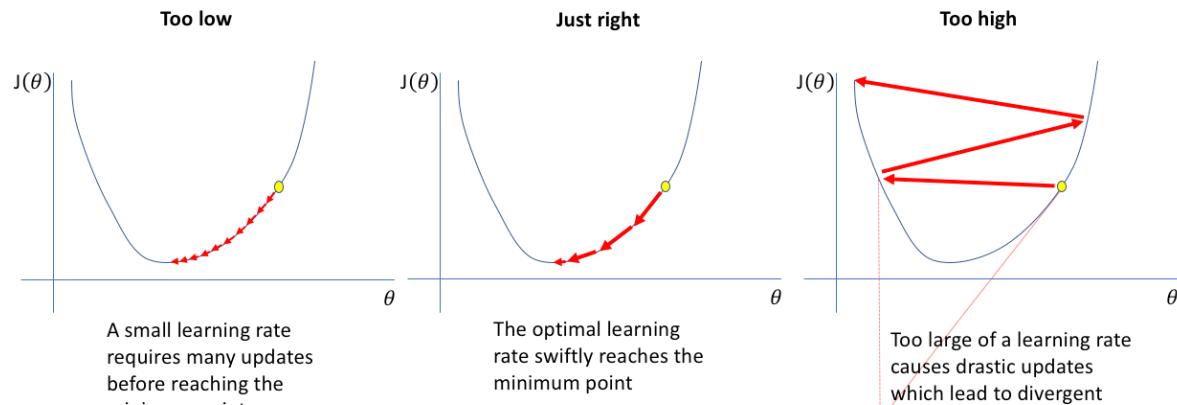
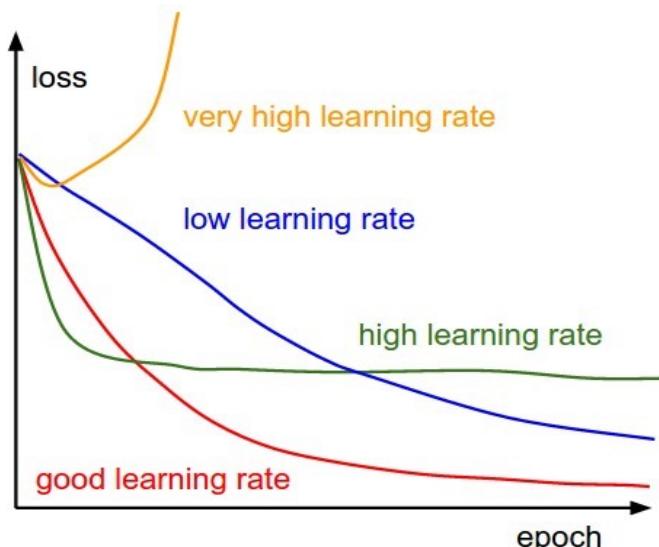
- **Mini-batch gradient descent:**

Compute the gradient for every **mini-batch** training examples.

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta, \mathcal{X}^{(i:i+n)})$$

Some Challenges For Gradient descent

- **Choosing a proper learning rate can be difficult.**
 - A learning rate that is too small leads to painfully slow convergence.
 - A learning rate that is too large can hinder convergence and cause the loss function to fluctuate around the minimum or even to diverge.



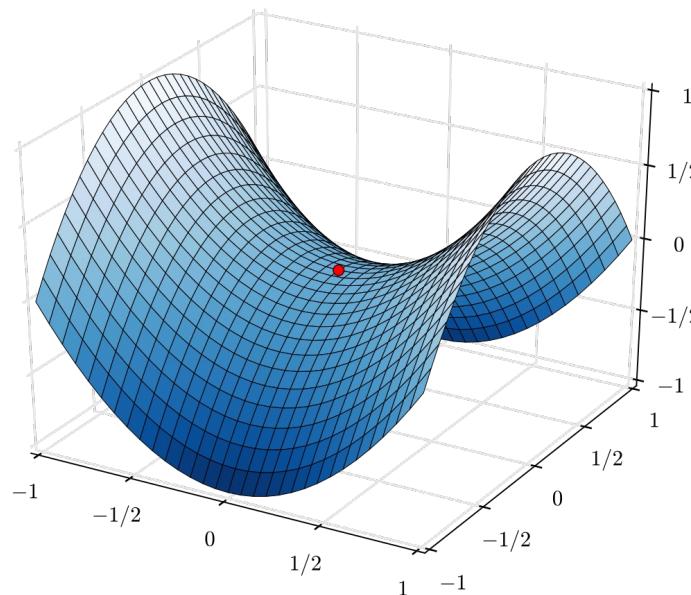
Some Challenges For Gradient descent

- **The same learning rate applies to all parameter updates.**
 - If our data is sparse and our features have very different frequencies, we might not want to update all of them to the same extent, but perform a larger update for rarely occurring features.

Some Challenges For Gradient descent

- **Easily get trapped in numerous saddle points.**

- Saddle points are points where one dimension slopes up and another slopes down. These saddle points are usually surrounded by a plateau of the same error, which makes it notoriously hard for SGD to escape, as the gradient is close to zero in all dimensions.



Overview of Different Optimisation Algorithms

