Istituto Italiano di Tecnologia iit

# Neural Coding

**Stefano Panzeri**

*IIT, Robotics, Brain and Cognitive Sciences Dept*
*Laboratory of Brain Signals Analysis*
*stefano.panzeri@iit.it*

## Plan of lectures

- Lecture 1(today): introduction to neural coding
- Lecture 2 (Friday): temporal coding
- Lecture 3 (Monday) A) Linear decoding – B) population codes
- Lecture 4 (Tuesday) : tutorial on mathematical techniques to study neural codes. A) Spectral analysis B) information analysis

2

## Lecture 1

## Introduction to Neural Coding

3

## Overview of lecture 1

- What is a code?
- Neural Codes: Spikes
- Neural Codes: Noise and probabilities
- Information Theory
- Decoding analysis
- Single neuron information transmission
- Times scales of single neuron information transmission: spike counts vs spike times

4

## Neuronal Coding

- How do neurons encode and transmit sensory information?
- What is the language ("code") used by the neurons to transmit information to one another?
- Which response variables encode the most information?
- How can a downstream neuronal population decode the cortical output?

5

## **Neuronal Coding/Decoding**

- Why is it important to understand how to decode a neural signal?
- To understand brain function: we cannot understand how the brain works if we do not know how its computing elements communicate
- To understand how to command neural prostheses and Brain-Machine Interfaces

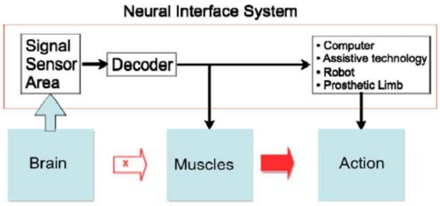



**Figure 1. Design of a Neural Interface System**
Disconnection of signals carrying movement intention to muscles is "bridged" by detecting signals that are either direct representations of movement or indirect substitutes for that signal. Each signal type requires the selection of a sensor and the area from which that signal is obtained. Neural signals are decoded into a command that is used to operate a range of devices to effect actions. In addition to physical systems, the signal could be routed to the muscles and, via muscle stimulators, reconnect the brain to the muscles to restore voluntary action.

## What is a code?

A code is transformation of a certain message by using another alphabet
For example, you can use your fingers to represent numbers


Use only thumb to code 0 or 1


2 fingers are used to encode 3 different numbers by the total amount of extended fingers


2 fingers are used with a more complex code to encode 4 different numbers. Here the position of the extended finger is also used to signal – this increases the capacity to encode information

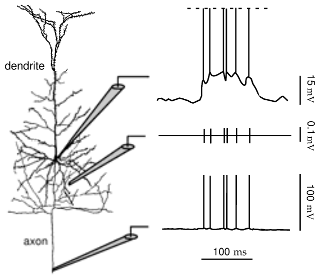


## Morse code

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | .- | K | -.- | U | ..- | 4 | ....- |
| B | -... | L | .-.. | V | ...- | 5 | ..... |
| C | -.-. | M | -- | W | .-- | 6 | -.... |
| D | -.. | N | -. | X | -..- | 7 | --... |
| E | . | O | --- | Y | -.-- | 8 | ---.. |
| F | ..-. | P | .--. | Z | --.. | 9 | ----. |
| G | --. | Q | --.- | 0 | ----- | Fullstop | .-.-.- |
| H | .... | R | .-. | 1 | .---- | Comma | --..-- |
| I | .. | S | ... | 2 | ..--- | Query | ..--.. |
| J | .--- | T | - | 3 | ...-- | | |

Telegrams were transmitted by electrical signals over wires by using Morse codes.
Morse codes are a correspondence between the pattern of electrical signals and the characters of the English alphabet.
There are two types of "symbols": long (_) and short (.) electrical pulses.
Not only the length of the individual signal is important, but also the timing of the individual signals (> for characters between different words)

## The neuronal code is a sequence of spikes

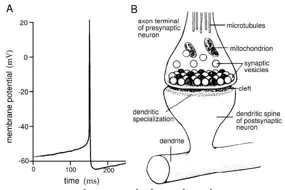



**Somatic electrode** – subthreshold membrane potential plus Action Potentials (spikes)

**Extracellular electrode**

**Axonal electrode**– subthreshold membrane potential are attenuated and only spikes propagate long distance

10



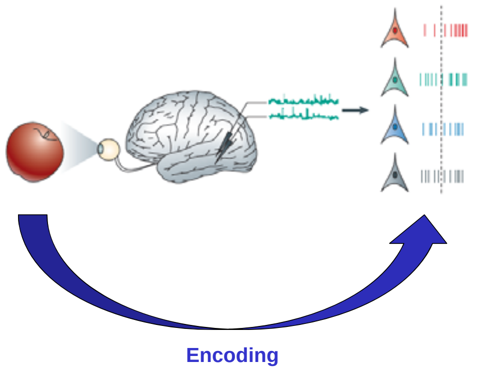

The synaptic vesicles in the axon terminal release neurotransmitters only when an action potential arrives from the presynaptic neuron

Thus, neurons communicate information only through action potentials, not through subthreshold membrane fluctuations.

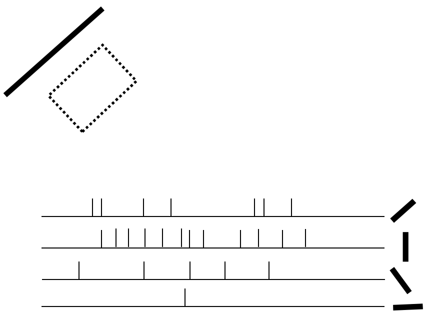
**Thus, the neuronal code consists of a time series of stereotyped action potentials**

11

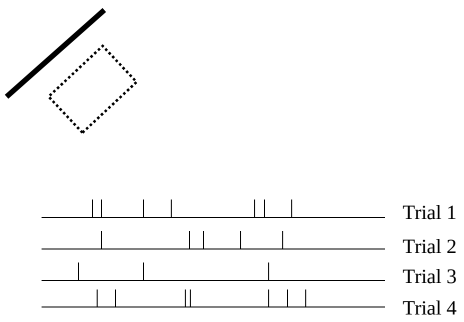
Rosetta Stone approach



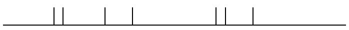

## Spike trains encode stimuli



## Single Neuron Variability

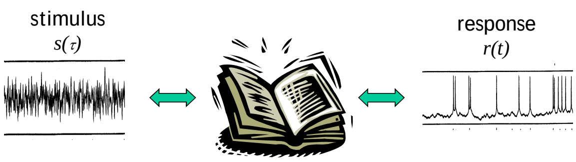


# Decoding brain activity



**??**

The dictionary for neural code is noisy…




Thus, the neuronal dictionary is probabilistic!

17

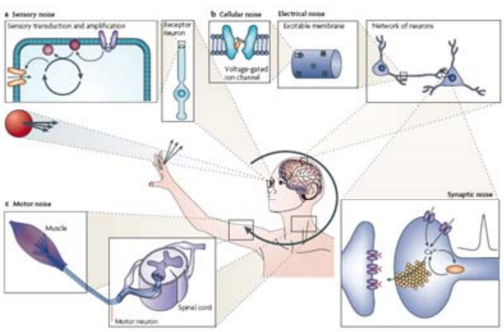
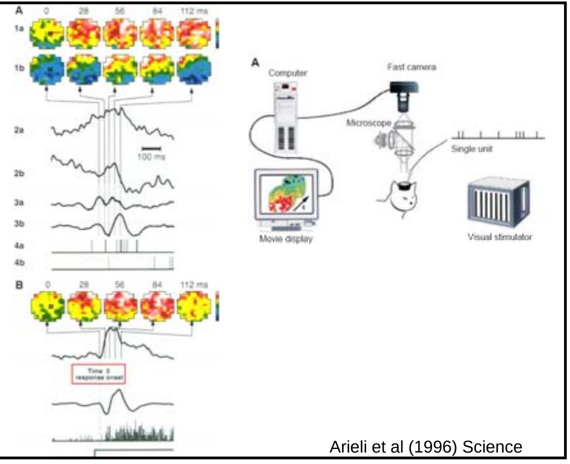
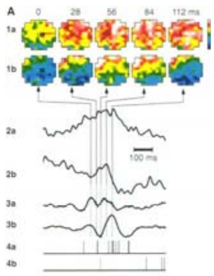



Figure 1 | **Overview of the behavioural loop and the stages at which noise is present in the nervous system.** **a** | Sources of sensory noise include the transduction of signals. This is exemplified here by a photoreceptor and its signal-amplification cascade. **b** | Sources of cellular noise include the ion channels of excitable membranes, synaptic transmission and network interactions (see BOX 2). **c** | Sources of motor noise include motor neurons and muscle. In the behavioural task shown (catching a ball), the nervous system has to act in the presence of noise in sensing, information processing and movement.



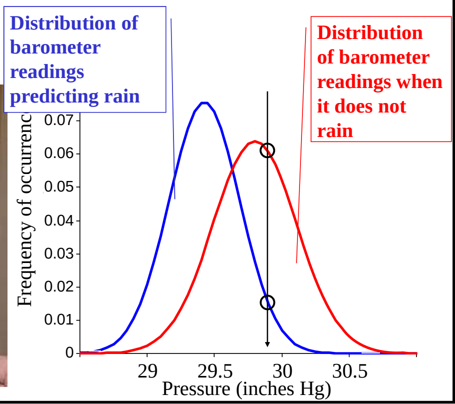

Arieli et al (1996) Science



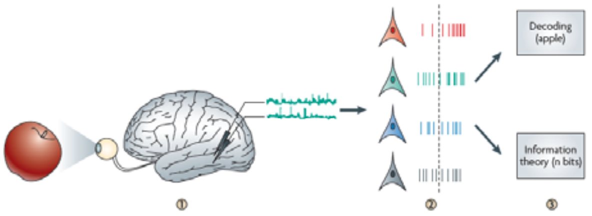

A response to an individual trial can be modeled as a sum of two components: the reproducible response and the ongoing network fluctuations

The effect of a stimulus may be likened to the additional ripples caused when throwing a stones in a wavy sea

## Probabilistic dictionary



Single-trial analyses of neural activity: information theory and decoding




Quian Quiroga and Panzeri (2009) Nature Reviews Neuroscience

## Information theory

Shannon, C. E. (1948). A mathematical theory of communication. Bell Sys. Tech. Journal 27:379-423, 623-656.

TM Cover and JA Thomas, Elements of Information Theory, John-Whiley 2006

PE Latham and Y Roudi (2008) Mutual Information – Scholarpedia article (freely accessible)

R Quian Quiroga and S Panzeri (2009) Extracting information from neuronal populations: informaiton theory and decoding approaches. Nature Reviews Neurosci

## What is Information Theory?

Information theory is a branch of mathematics that deals with measures of information and their application to the study of communication, statistics, and complexity.

It originally arose out of communication theory and is sometimes used to mean the mathematical theory that underlies communication systems and communication in the presence of noise.

Based on the pioneering work of Claude Shannon

## Source of Information

Information theory relies on the theory of PROBABILITY.

A source of information X produces a message out of a set of possible messages.

If there is only one possible message, then no information is transmitted by sending that message. The amount of information obtained from a message, and the difficulty in transmitting it, is related to its UNCERTAINTY (not to its meaning)

**Source X**
**P(x)**

$$\text{Source } X,\ P(x) \longrightarrow x_1\ x_2 \ldots\ldots x_n$$

## Entropy of a random variable X

**The first step to define information is to quantify uncertainty**
**The entropy H of a random variable x is a simple function of its probability distribution P(x):**

$$H(X) = -\sum_x P(x) \log_2 P(x)$$

**H(X) measures the amount of uncertainty inherent to a random variable, i.e. the amount of information necessary to describe it**

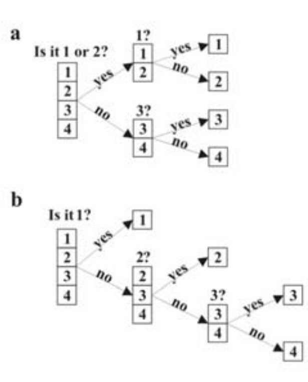
26

## Entropy and description length

**H(X) has a very concrete interpretation: Suppose x is chosen randomly from the distribution P(x) , and someone who knows the distribution is asked to guess which was chosen. If the guesser uses the optimal question-asking strategy -- which is to divide the probability in half on each guess by asking questions like "is x greater than .. ?", then the average number of yes/no questions it takes to guess lies between H(X) and H(X)+1.**

**This gives quantitative meaning to the "uncertainty" expressed by quantities like entropy: it relates to the number of yes/no questions it takes to guess a random variables, given knowledge of the underlying distribution and taking the optimal question-asking strategy**

27




Case a
All variable have equal P = ¼
H(X)=2
Avg number of questions = 2

Case b
P (1)= ½
P(2)=1/4
P(3)=P(4)=1/8
H(X)=1.75
Avg number of questions = 1.75

Taken from Nirenberg and Latham, PNAS 2005

## Communication Channel

A communication channel is made of a sender X (the source) which sends information and by a receiver Y which receives the message from the sender in some other format.
Transmission of information over a communication channel is subject to noise (e.g interference from other senders).
Mathematically, the channel is expressed by the "transition probability" P(y|x), the probability of receiving message y when message x was sent

$x_1\ x_2 \ldots\ldots x_n$ → **Channel P(y|x)** → $y_1\ y_2 \ldots\ldots y_n$

**Source P(x)** **Receiver P(y)**

## Conditional Entropy

If the communication channel works, after the observation of the message y our uncertainty about the message x is reduced.

Conditional entropy:

$$H(X|Y) = -\sum_y P(y) \sum_x P(x|y) \log_2 P(x|y)$$

Conditional Entropy measures the average residual uncertainty of message x when we know the value of y.

$$H(X|Y) \leq H(X)$$

## Shannon's Mutual Information

Entropy of the source: Average uncertainty of the source X

$$H(X) = -\sum_x P(x)\log_2 P(x)$$

Conditional Entropy: Average uncertainty of x after observing y

$$H(X\,|\,Y) = -\sum_y P(y)\sum_x P(x\,|\,y)\log_2 P(x\,|\,y)$$

Mutual Information:

$$I(X;Y) = H(X) - H(X\,|\,Y)$$

I(X;Y) is the average reduction in the uncertainty of X due to the knowledge of Y

$$\underset{\text{Source } P(x)}{\xrightarrow{x_1\ x_2 \ldots\ldots x_n}} \boxed{I(X;Y)} \underset{\text{Receiver } P(y)}{\xrightarrow{y_1\ y_2 \ldots\ldots y_n}}$$

**Info = k bits means that observing the neural response decreases uncertainty about the stimulus by 2^k**

**1 bit of info: After observing r we know s is here**

**2 bits of info: After observing r we know s is here**

**Before we observe the response the stimulus can be anywhere here**

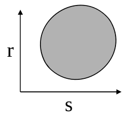
## Mutual Information and Correlation

$$I(X;Y) = \sum_{x,y} P(x,y)\log_2 \frac{P(x,y)}{P(x)P(y)}$$

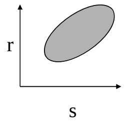
I is a measure of how the two stochastic variables X and Y can be predicted from one another.
Thus, it is a measure of how much X and Y are correlated.

If X and Y are uncorrelated, then P(x,y) = P(x)P(y) and I(X;Y)=0

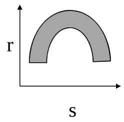
If y is any deterministc monotonic function of x, then I(X,Y) = maximal = H(X)
Thus, I(X;Y) is a non-linear measure of correlation




Zero information
Zero linear correlations



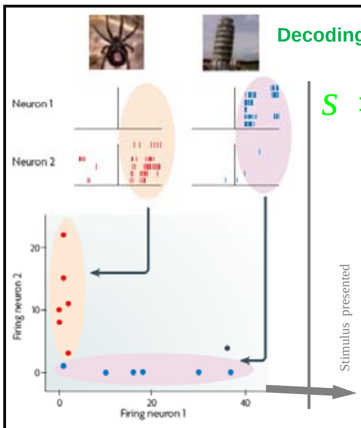

High information
High linear correlations

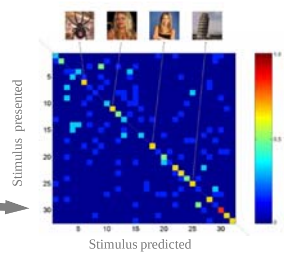



High information
Low linear correlations

## Decoding (stimulus reconstruction)

$$s \Rightarrow \underbrace{r \Rightarrow s^{p}}_{decoding}$$



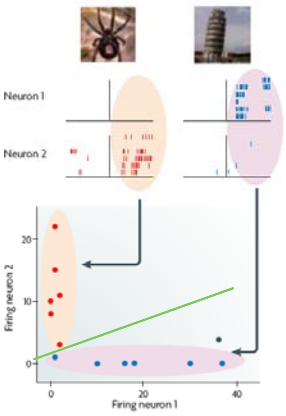

Quian Quiroga & Panzeri Nature Reviews Neurosci 2009

## Decoding algorithm using Bayes rule

$$P(s'|r) = \frac{P(s')P(r|s')}{P(r)}$$

To predict the stimulus sp that generated a given response r, we can chose the stimulus that maximizes the posterior probability of have caused r

$$s^{p} = \arg\max_{s'} P(s'|r)$$




## Decoding using clustering algorithms

Divide the response space into regions. When the response falls in a given region, assign it to a stimulus

Optimal boundaries

## Decoding using neural networks

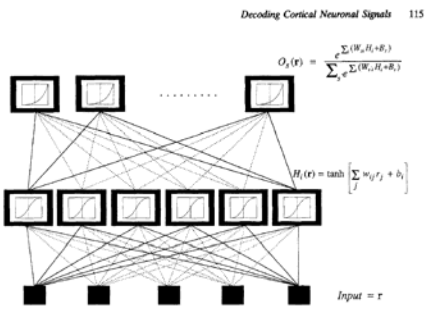
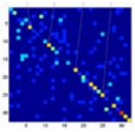



Fig. 2 Feed-forward neural network to decode neuronal signals. This example has 5 input units, 6 sigmoidal hidden units, and one normalized exponential output unit for each class.

Kjaer et al (1994) J Comput Neurosci

## Information in the confusion matrix

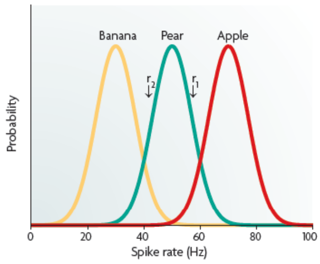
$$I(S;S^p) = \sum_{s,s^p} P(s,s^p)\log_2 \frac{P(s,s^p)}{P(s)P(s^p)}$$




I(S;$S^p$) quantifies the information than can be gained by a receiver that knows the true P(s|r) but only uses the information about which is the most likely stimulus

Because of data processing inequality:

$$I(S;S^p) \le I(S;R)$$




Neurons may convey information in other ways than reporting the most likely stimulus
If the receiver only uses the information about which is the most likely stimulus, it may miss out on important information even if it operates on the true P(s|r).

Quian Quiroga & Panzeri Nature Reviews Neurosci 2009

## Advantages of information theory

- It quantifies single-trial stimulus discriminability on a meaningful scale (bits)
- It is a principled measure of the correlation on a single trial basis between neuronal responses and stimuli
- It works even in non-linear situations
- It makes no assumption about the relation between stimulus and response – we do not need to specify which features of the stimulus activate the neuron and how they activate it
- Since it takes into account all ways in which a neuron can reduce ignorance, it bounds the performance of any biological decoder. Thus it can be used to explore which spike train features are best for stimulus decoding, without the limitations coming from committing to a specific decoding algorithm

## Advantages of decoding techniques

- Decoding algorithms are much easier to implement and compute
- Calculations are robust even with limited amounts of data
- Allow to test specific hypotheses on how downstream systems may interpret the messages of other neurons

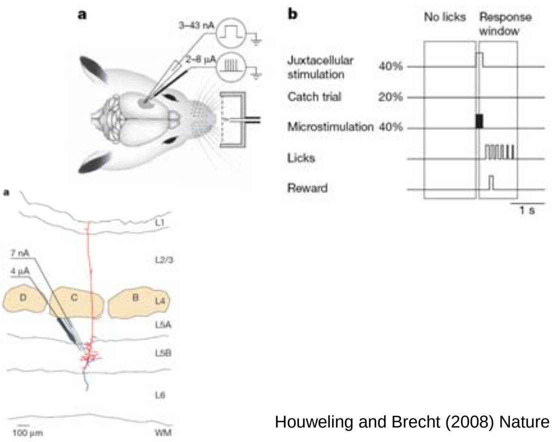
## Information coding by single neurons

There are millions of neurons in the brain

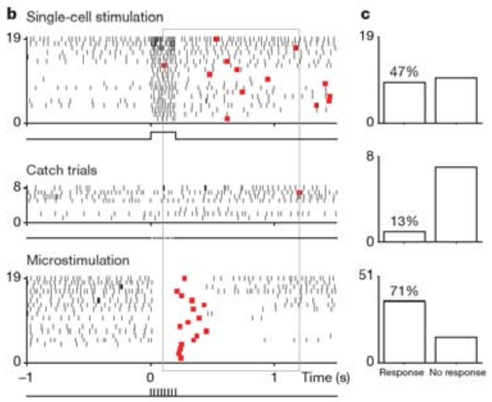
Since we cannot record from all of them, and since understanding a single neuron is already difficult, it is worth starting to **investigate how individual neurons carry information**

The output of single neurons has an impact on the behaviour of the animal; thus it should be taken seriously

43




Houweling and Brecht (2008) Nature




Houweling and Brecht (2008) Nature

## Population coding

Population coding deals with understanding how the brain may put together all the different bits of information carried by each individual neuron

46

## Questions on single neuron coding

Understand how the activity of single neurons codes information and contributes to behaviour

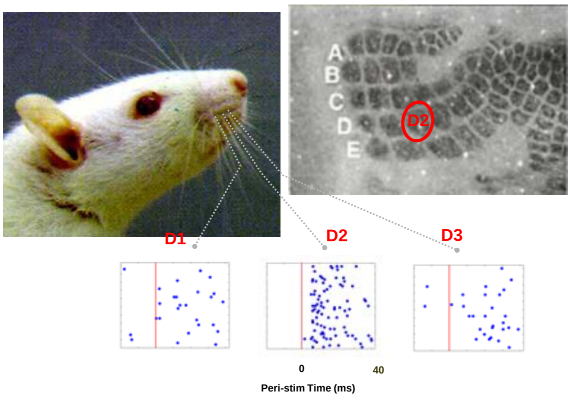
How is a stimulus transformed into a patterns of spikes?

What computational abilities does the downstream neuron need to decode the spike train?

Does the decoder need to register the incoming spike times with ms-precision?

47

## Coding of whisker location in rat S1 cortex

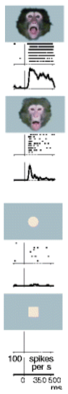


## The spike count hypothesis

The simplest hypothesis is that the identity of a stimulus is encoded by the number of spikes emitted by the neuron in response to the presentation of the stimulus.

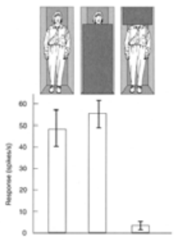
Example: face-selective neurons in Inferior Temporal (IT) cortex of monkeys.
They encode the identity or the emotion of a face by increasing the total number of emitted spikes **only when** certain faces are presented to the animal

49

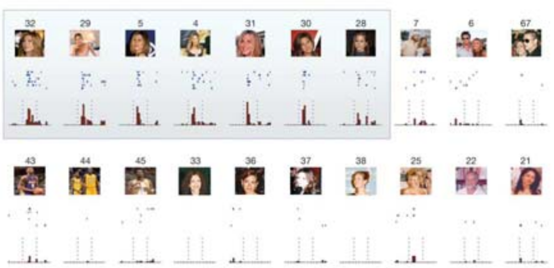


A face neuron responds only to faces (to some faces, not to all faces).
Coding of face identity

A face neuron does not respond to simple stimuli (such as spots of light etc) or when the face is removed from view



50



Spike counts encode Jennifer Aniston's identity in human temporal lobe

RQ Quiroga et al Nature 2005

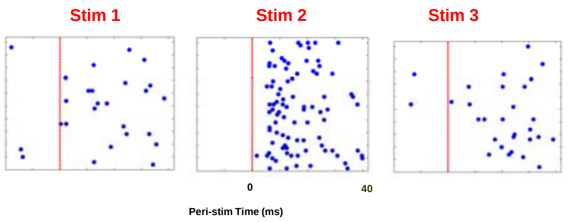
## The spike timing coding hypothesis

The spike timing coding hypothesis states that not only spike counts are important, but also the timing at which each spike was emitted.

Spike timing may convey additional information about the stimulus that is not conveyed by the spike counts alone.

52

## Spike timing code ?




## What is the advantage of a spike timing code?

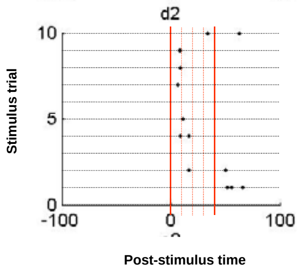
Spike timing can code more information than spike counts alone.

Example, use your fingers to represent numbers


2 fingers are used to encode 3 different numbers by the total amount of extended fingers


2 fingers are used with a more complex code to encode 4 different numbers. Here the position of the extended finger is also used to signal – this increases the capacity to encode information

54




| Spike count | Spike timing |
|---|---|
| 1 | 0001 |
| 1 | 1000 |
| 1 | 1000 |
| 1 | 1000 |
| 0 | 0000 |
| 1 | 0100 |
| 2 | 1100 |
| 0 | 0000 |
| 1 | 0100 |
| 0 | 0000 |
| 0 | 0000 |

Up to 5 numbers

$2^4 = 16$ numbers

55

**Tomorrow**

**Temporal codes (or: how the brain uses time to represent information)**

56