COSC421: Neural Models of Language

Lecture 1: Introduction / The visual system

Alistair Knott

Dept. of Computer Science, University of Otago

Course organisation

I'm Ali: email alik@cs.otago.ac.nz, phone 8299, office 254

Course webpage: http://www.cs.otago.ac.nz/cosc421/

Assessment: 60% internal, 40% exam.

Assignment	Topic	Handed out	Due	Worth
1	Phonological working memory	Week 5	Week 7	20%
2	Word forms	Week 6	Week 9	20%
3	Infant word-meaning learning	Week 7	Week 11	20%

For each assignment, there will be programming and essay options.

Textbook

The textbook for the course is a book I wrote:

- A Knott: Sensorimotor Cognition and Natural Language Syntax. MIT Press, 2012.
- The E-book is free from the library website: http://site.ebrary.com/lib/otago/Doc?id=10617471

Each lecture is based on a portion of the book.
You should do the readings associated with each lecture.

Aim of the course

The big question we'll be asking:

- How does human language work?

Aim of the course

The big question we'll be asking:

- How does human language work?

There are lots of ways of investigating this question.

Aim of the course

The big question we'll be asking:

- How does human language work?

There are lots of ways of investigating this question.

- **Linguists** develop theories of syntax, which attempt to model particular human languages.

Aim of the course

The big question we'll be asking:

- How does human language work?

There are lots of ways of investigating this question.

- **Linguists** develop theories of syntax, which attempt to model particular human languages.
- **Psycholinguists** run experiments on speakers and hearers, which investigate the processes of sentence interpretation and generation.

Aim of the course

The big question we'll be asking:

- How does human language work?

There are lots of ways of investigating this question.

- **Linguists** develop theories of syntax, which attempt to model particular human languages.
- **Psycholinguists** run experiments on speakers and hearers, which investigate the processes of sentence interpretation and generation.
- **Neurolinguists** explore how these processes work in the brain.

Aim of the course

The big question we'll be asking:

- How does human language work?

There are lots of ways of investigating this question.

- **Linguists** develop theories of syntax, which attempt to model particular human languages.
- **Psycholinguists** run experiments on speakers and hearers, which investigate the processes of sentence interpretation and generation.
- **Neurolinguists** explore how these processes work in the brain.
- **Computational linguists** implement theories of syntax, and of sentence interpretation/generation.

Aim of the course

The big question we'll be asking:

- How does human language work?

There are lots of ways of investigating this question.

- **Linguists** develop theories of syntax, which attempt to model particular human languages.
- **Psycholinguists** run experiments on speakers and hearers, which investigate the processes of sentence interpretation and generation.
- **Neurolinguists** explore how these processes work in the brain.
- **Computational linguists** implement theories of syntax, and of sentence interpretation/generation.
- **Developmental linguists** study babies learning language.

Aim of the course

In 421, we'll focus on a particular question:

- How is language implemented in the human brain?

Aim of the course

In 421, we'll focus on a particular question:

- How is language implemented in the human brain?

Why focus on this question?

Aim of the course

In 421, we'll focus on a particular question:

- How is language implemented in the human brain?

Why focus on this question?

- No-one really knows how language works.

Aim of the course

In 421, we'll focus on a particular question:

- How is language implemented in the human brain?

Why focus on this question?

- No-one really knows how language works.
- But we know that the brain can do it.

Aim of the course

- How is language implemented in the human brain?

Aim of the course

- How is language implemented in the human brain?

We'll approach this question from three perspectives.

Aim of the course

- How is language implemented in the human brain?

We'll approach this question from three perspectives.

1. **Embodied cognition.**
   How does language interface with the sensorimotor system?

Aim of the course

- How is language implemented in the human brain?

We'll approach this question from three perspectives.

1. **Embodied cognition.**
   How does language interface with the sensorimotor system?
2. **Computational modelling.**
   We'll be looking at computer models of sensorimotor and language processing.

Aim of the course

- How is language implemented in the human brain?

We'll approach this question from three perspectives.

1. **Embodied cognition.**
   How does language interface with the sensorimotor system?
2. **Computational modelling.**
   We'll be looking at computer models of sensorimotor and language processing.
3. **Theoretical syntax.**
   We'll be asking how theoretical models of syntax relate to brain mechanisms.

Aim of the course

1. **Embodied cognition**

'Embodied' approaches to cognition investigate high-level cognitive faculties (e.g. language) by looking at how they interface with the **sensorimotor** (**SM**) system.

Aim of the course

1. **Embodied cognition**

'Embodied' approaches to cognition investigate high-level cognitive faculties (e.g. language) by looking at how they interface with the **sensorimotor** (**SM**) system.

- An evolutionary argument:
  We had low-level SM capabilities first, and higher-level faculties evolved on top of these. So perhaps the organisation of higher-level faculties reflects their SM origins.

Aim of the course

1. **Embodied cognition**

'Embodied' approaches to cognition investigate high-level cognitive faculties (e.g. language) by looking at how they interface with the **sensorimotor** (**SM**) system.

- An evolutionary argument:
  We had low-level SM capabilities first, and higher-level faculties evolved on top of these. So perhaps the organisation of higher-level faculties reflects their SM origins.
- An argument about starting points:
  We know a lot about SM processing in the brain. If we want to find out about high-level cognitive faculties, we should focus on how they interface with the SM system.

Aim of the course

1. **Embodied cognition**

'Embodied' approaches to cognition investigate high-level cognitive faculties (e.g. language) by looking at how they interface with the **sensorimotor** (**SM**) system.

- An evolutionary argument:
  We had low-level SM capabilities first, and higher-level faculties evolved on top of these. So perhaps the organisation of higher-level faculties reflects their SM origins.
- An argument about starting points:
  We know a lot about SM processing in the brain. If we want to find out about high-level cognitive faculties, we should focus on how they interface with the SM system.
- An argument about meaning…

An argument about meaning

Q: How should we represent the meaning of the following piece of text?

> *The dog, a distinguished and unusual collie with blue eyes, and long silver-grey fur blotched with black and white, came bounding into the room.*

An argument about meaning

Q: How should we represent the meaning of the following piece of text?

> *The dog, a distinguished and unusual collie with blue eyes, and long silver-grey fur blotched with black and white, came bounding into the room.*

How can you represent this properly without referring to what dogs look like? & how they move, how their fur feels...

Aim of the course

2 **Computational modelling**

Processing in the brain is very complex. It's hard to state models of neural processes in words or diagrams.

- Expressing models of neural processes as computer programs allows us to *simulate* the processes being modelled.
- Computer simulations also force us to be *explicit* about our model.

I'll be discussing computational models of the SM system, and also of language.

Aim of the course

3. **Theoretical syntax**

The most sophisticated models of language are the ones made by theoretical linguists.

Aim of the course

3. **Theoretical syntax**

The most sophisticated models of language are the ones made by theoretical linguists.

- These make no reference to how the brain represents or processes language.

Aim of the course

3. **Theoretical syntax**

The most sophisticated models of language are the ones made by theoretical linguists.

- These make no reference to how the brain represents or processes language.
- But if they're good models of language, it should be possible to *interpret* them as describing the brain's language mechanism.

A concrete starting point

This course is going to be structured around a single example event: one in which a man grabs a cup.

This event can be studied from two points of view:

- Linguists study the syntax and semantics of sentences which *report* the event (e.g. *The man grabbed a cup*).
- Psychologists study how the event can be recognised, and how the described action can be performed.

We will explore to what extent these areas of study overlap.

The language/sensorimotor interface

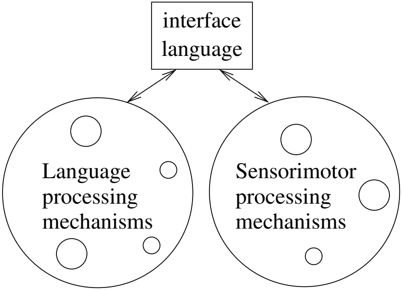
Clearly we can *convert* sensorimotor representations into linguistic ones (and vice versa).

- Because we can talk about what we see;
- Because we can execute verbal instructions.

Question: how much work is involved in this conversion?

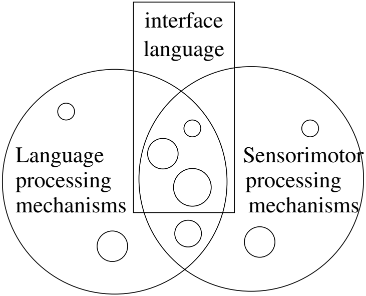
Two suggestions

Language and SM processing are **modules**



Semantic representations **abstract away** from details of SM processing

Language and SM processing **share mechanisms**



Semantic representations **retain** details of SM processing

Motivations for the ‘shared mechanisms’ hypothesis

A priori arguments: good, but vague

- Evolutionary arguments
- Parsimony
- ‘Embodied’ accounts of meaning

Motivations for the 'shared mechanisms' hypothesis

A priori arguments: good, but vague

- Evolutionary arguments
- Parsimony
- 'Embodied' accounts of meaning

Empirical experiments: quite nice evidence about *single words*...

Motivations for the ‘shared mechanisms’ hypothesis

A priori arguments: good, but vague

- Evolutionary arguments
- Parsimony
- ‘Embodied’ accounts of meaning

Empirical experiments: quite nice evidence about *single words*. . .

- Pulvermüller *et al.* (1999) ERP study: concrete nouns activate visual cortices; action verbs activate motor/premotor cortices

Motivations for the ‘shared mechanisms’ hypothesis

A priori arguments: good, but vague

- Evolutionary arguments
- Parsimony
- ‘Embodied’ accounts of meaning

Empirical experiments: quite nice evidence about *single words*. . .

- Pulvermüller *et al.* (1999) ERP study: concrete nouns activate visual cortices; action verbs activate motor/premotor cortices
- Pulvermüller *et al.* (2005) MEG study: leg and mouth verbs activate corresponding areas of motor cortex

Motivations for the ‘shared mechanisms’ hypothesis

A priori arguments: good, but vague

- Evolutionary arguments
- Parsimony
- ‘Embodied’ accounts of meaning

Empirical experiments: quite nice evidence about *single words*. . .

- Pulvermüller *et al.* (1999) ERP study: concrete nouns activate visual cortices; action verbs activate motor/premotor cortices
- Pulvermüller *et al.* (2005) MEG study: leg and mouth verbs activate corresponding areas of motor cortex
- Pulvermüller *et al.* (2005): TMS over hand and leg areas differentially affects processing of hand and leg verbs

Compositionality in language and SM cognition

It's not so surprising that individual words access SM representations.

A more interesting question: what's the relationship between **combinatorial mechanisms** in language and in the SM domain?

Compositionality in language and SM cognition

It's not so surprising that individual words access SM representations.

A more interesting question: what's the relationship between **combinatorial mechanisms** in language and in the SM domain?

- There must be mechanisms for combining primitive SM concepts representing actions and objects to create event representations.

Compositionality in language and SM cognition

It's not so surprising that individual words access SM representations.

A more interesting question: what's the relationship between **combinatorial mechanisms** in language and in the SM domain?

- There must be mechanisms for combining primitive SM concepts representing actions and objects to create event representations.
- Q: Do these bear any resemblance to syntactic mechanisms in language?

Methodology 1: starting with concrete sentences

On this course, we'll start by looking at sentences which describe events we can apprehend (or bring about) using 'well-understood' SM mechanisms.

We know a lot about the SM mechanisms involved in:

- attention to / categorisation of concrete objects
- control/perception of reach-to-grasp actions

So *The man grabbed a cup* is a good place to start.

Methodology 2: comparing syntactic and SM models

What we'll do:

Methodology 2: comparing syntactic and SM models

What we'll do:

- Develop a SM model of the processes involved in grabbing a cup, or in observing a cup-grabbing event
(motivated from SM psychology / neuroscience)

Methodology 2: comparing syntactic and SM models

What we'll do:

- Develop a SM model of the processes involved in grabbing a cup, or in observing a cup-grabbing event
(motivated from SM psychology / neuroscience)
- Develop a syntactic model of the cup-grabbing sentence
(motivated on purely linguistic grounds)

Methodology 2: comparing syntactic and SM models

What we'll do:

- Develop a SM model of the processes involved in grabbing a cup, or in observing a cup-grabbing event
  (motivated from SM psychology / neuroscience)
- Develop a syntactic model of the cup-grabbing sentence
  (motivated on purely linguistic grounds)
- Look for **formal similarities** between these models.

Methodology 2: comparing syntactic and SM models

What we'll do:

- Develop a SM model of the processes involved in grabbing a cup, or in observing a cup-grabbing event
  (motivated from SM psychology / neuroscience)
- Develop a syntactic model of the cup-grabbing sentence
  (motivated on purely linguistic grounds)
- Look for **formal similarities** between these models.

If there are nontrivial similarities, this is support for the 'shared mechanisms' hypothesis.

An objection

Surely syntactic reps can't be closely related to SM mechanisms, because syntax is orthogonal to concreteness.

The man grabbed a cup
The company acquired a subsidiary...

An objection

Surely syntactic reps can't be closely related to SM mechanisms, because syntax is orthogonal to concreteness.

The man grabbed a cup
The company acquired a subsidiary. . .

Linking syntax to SM mechanisms forces us to adopt a particular account of the semantics of abstract sentences, in which abstract concepts are grounded in concrete SM ones.

Another objection

Surely syntactic reps can't be closely related to SM mechanisms, because SM mechanisms are *universal*, but syntactic structures (on the surface) are not.

Another objection

Surely syntactic reps can't be closely related to SM mechanisms, because SM mechanisms are *universal*, but syntactic structures (on the surface) are not.

To maintain a strong version of the shared mechanisms hypothesis, we must adopt a syntactic theory that posits an 'underlying' level of syntactic structure:

- which is reasonably invariant over translation;
- which represents the meaning of sentences.

Another objection

Surely syntactic reps can't be closely related to SM mechanisms, because SM mechanisms are *universal*, but syntactic structures (on the surface) are not.

To maintain a strong version of the shared mechanisms hypothesis, we must adopt a syntactic theory that posits an 'underlying' level of syntactic structure:

- which is reasonably invariant over translation;
- which represents the meaning of sentences.

This idea of an 'underlying' level of syntactic structure is a cornerstone of the syntactic theories developed by Noam Chomsky.

Summary of the course

Part 1: I'll introduce a model of the SM processes involved in experiencing a reach-to-grasp action.

Part 2: I'll introduce models of neural representations of words in the brain, and about how infants learn words.

Part 3: I'll introduce two influential syntactic theories: one from Chomsky (**Minimalism**), and one influenced more by recent cognitive science (**construction grammar**).

Part 4: I'll discuss the relationship between syntax and the sensorimotor system, and propose a model of how infants learn syntax.

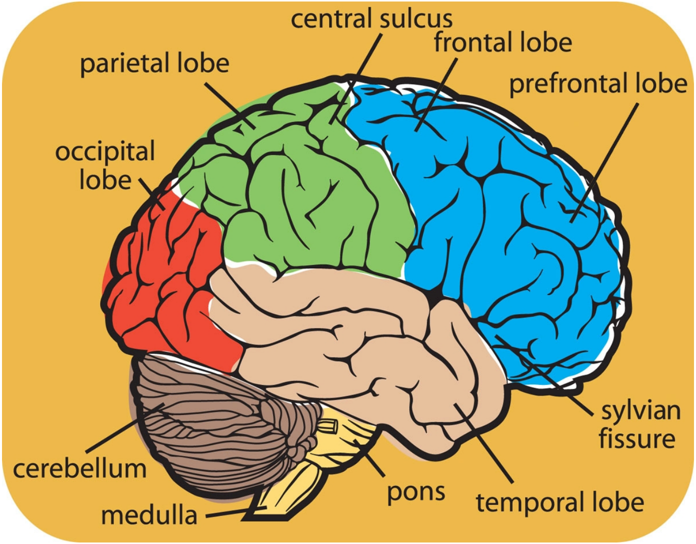
Course overview

Week	Topic
1	Intro / Visual object classification and visual attention
2	Execution of reach-to-grasp actions
3	Recognition of reach-to-grasp actions
4	Sequential structure of reach-to-grasp actions
5	Working memory for reach-to-grasp actions
6	Language networks in the brain
7	How infants learn words: data and models
8	Syntax I: Chomsky's Minimalism
9	Syntax I (continued)
10	Syntax II: Empiricist models of syntax
11	A sensorimotor interpretation of Minimalist syntax
12	Neural network models of syntax
13	Summary / overspill

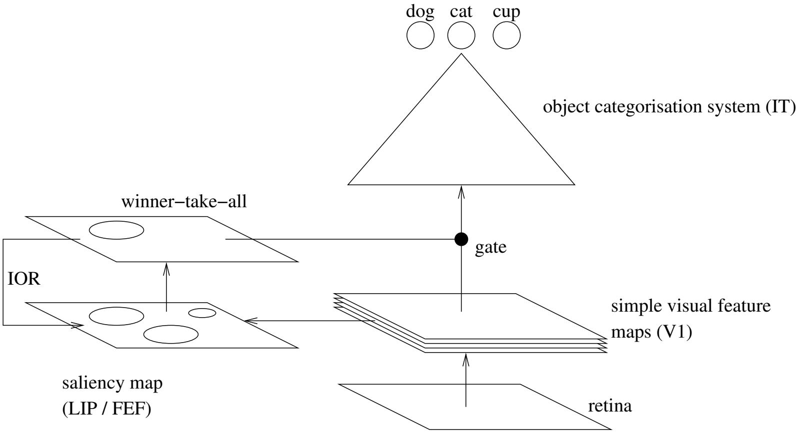
Outline of today's lecture

1. Introduction to the visual system
2. Early visual processing
3. The object classification pathway
4. The attentional pathway

Areas of the brain



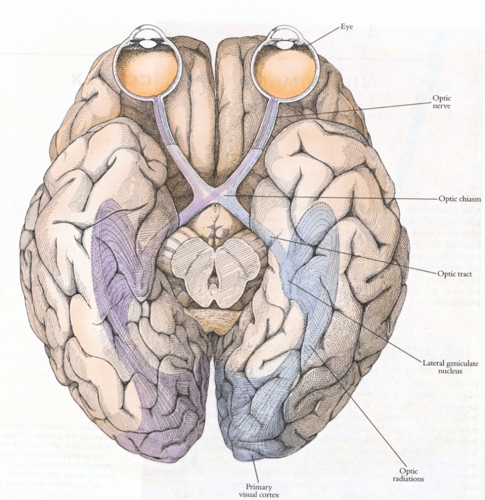
Neural pathways involved in perception of objects



Outline of today's lecture

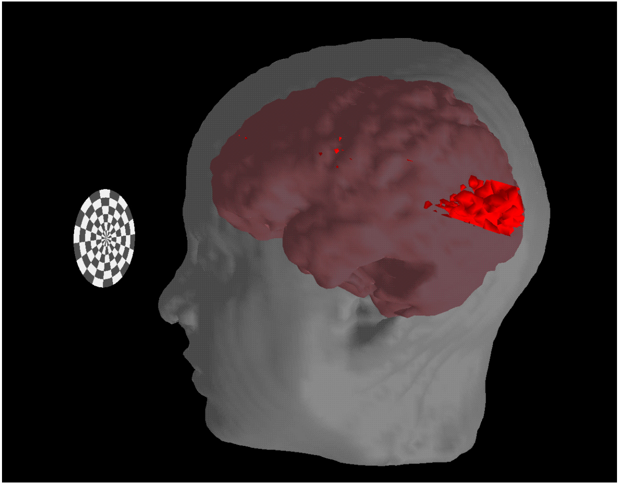
Early visual processing

Information from the retina is transmitted via the lateral geniculate nucleus to **primary visual cortex** (V1) in the occipital lobes.



V1 in humans

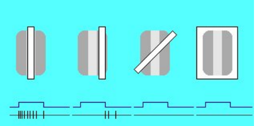
Here's activity from an fMRI scan generated while a (human) subject watches a simple visual pattern:



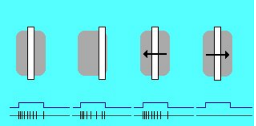
Simple and complex cells in V1

The structure of cells in V1 was discovered by Hubel and Wiesel (1968), using single-cell recordings.

- Cells in V1 are organised **retinotopically**. They compute a range of simple **feature maps** over the retina.
- A **simple cell** responds best to a stimulus with a particular orientation, and a particular size, at a particular point on the retina.



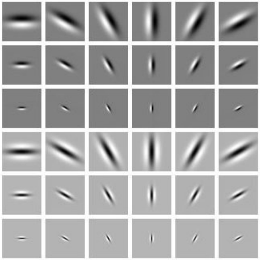
- A **complex cell** responds best to a stimulus with a particular orientation, size & motion within a (small) area of retina.



The kind of filters computed by V1

Difference-of-Gaussian filters ('blob detectors'):

Oriented Gaussian filters ('line detectors'):



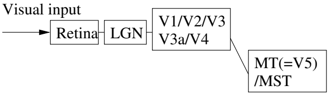
Projections from V1

V1 projects to V2, which is also retinotopically organised (though at a coarser granularity).

V1 and V2 project to several other more specialised retinotopic areas.

- V3 cells are sensitive to orientation and binocular disparity (Adams and Zeki, 2001) but not to colour (Baizer, 1982).
- V4 cells are sensitive to simple shapes (Cadieu *et al.*, 1998).
- MT (=V5) and MST cells are sensitive to motion (Maunsell and van Essen, 1983).

To give a very simple summary:



Outline of today's lecture

The object classification pathway

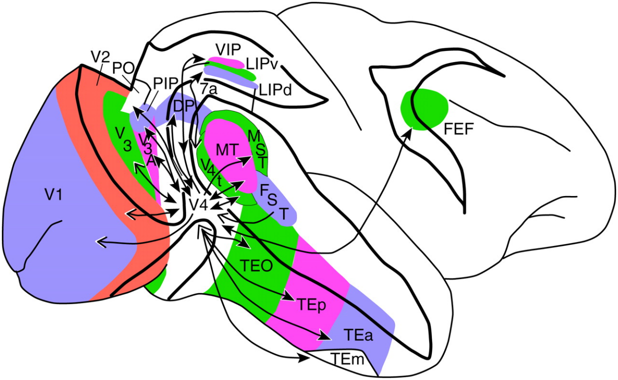
There is a specialised pathway for categorising *complex object shapes*: the kind of information which is necessary to identify the *type* of an object, or to recognise individual objects.

- The pathway receives input mainly from V4, and involves the inferotemporal (IT) areas TEO and TE.

The object classification pathway

There is a specialised pathway for categorising *complex object shapes*: the kind of information which is necessary to identify the *type* of an object, or to recognise individual objects.

- The pathway receives input mainly from V4, and involves the inferotemporal (IT) areas TEO and TE.



Characteristics of cells in the classification pathway

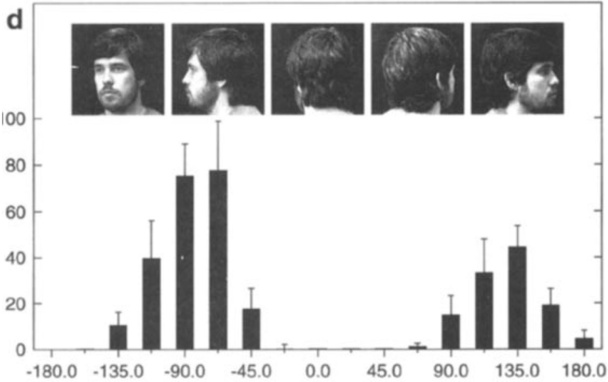
As we progress along the pathway from TEO to TE, cells respond to progressively **more complex stimuli**, and have progressively **wider receptive fields** (Tanaka, 1993).

The cells at the end of the pathway respond to fairly complex shapes. Logothetis *et al* (1995) trained monkeys to discriminate between 'paperclip' stimuli:

They found individual cells in TE which responded selectively to particular stimuli. (Most were selective to particular *views*, but some responded to multiple views.)

Responses to biological shapes in IT

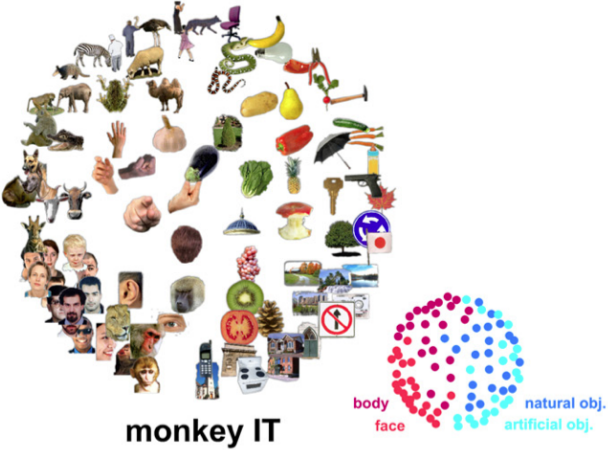
Many IT cells respond selectively to ‘biological’ shapes, such as faces and hands.



Here’s a cell that responds selectively to particular views of faces (Logothetis and Sheinberg, 1995).

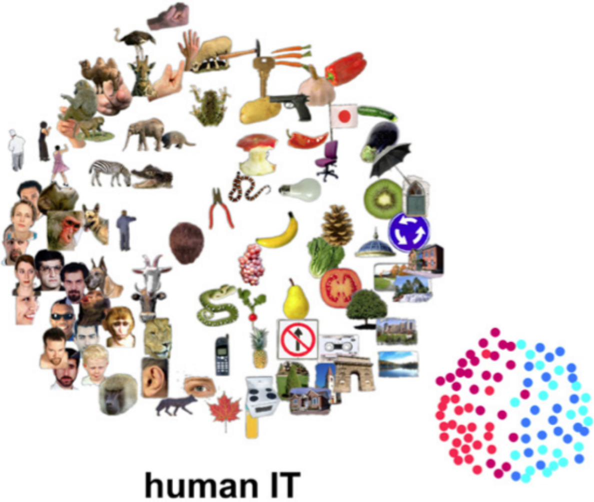
Responses to object categories in IT

Kriegeskorte *et al.* (2008) showed monkeys pictures of objects of different kinds and analysed the fMRI signals generated in IT. When these were organised by similarity, clusters of categories could be seen:



The object classification pathway in humans

Kriegeskorte *et al.* (2008) found a similar result in humans:



The object classification pathway in humans

It's possible to train classifiers that *predict* object type from brain activity (see e.g. Sudre *et al.*, 2012).

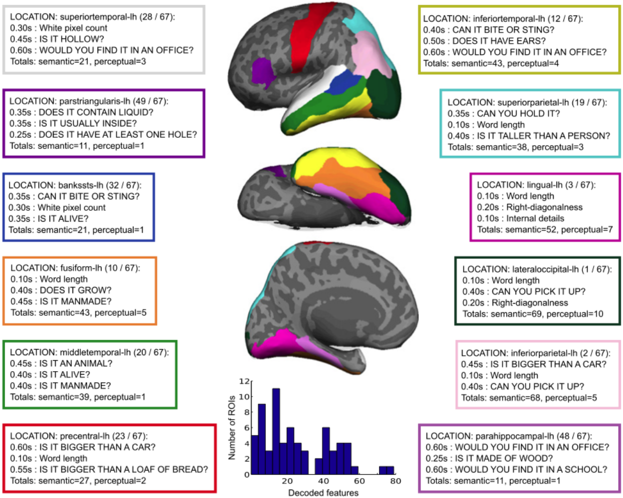
- The training objects are represented to the classifier as vectors of binary semantic features: e.g.

Can you hold it?	Is it taller than a person?	Is it alive?	Is it an animal?	Can it bend?
0	1	1	1	0
Is it hairy?	Does it have feelings?	Does it grow?	is it manmade?	Was it invented?
1	0	0	1	1

- The classifier learns to map vectors of brain activity onto feature complexes.
- After training, it can predict the features associated with unseen objects.

The object classification pathway in humans

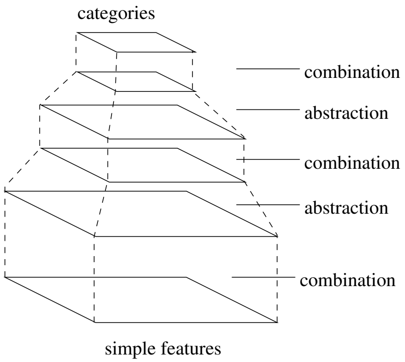
The neural areas that best predict different semantic features in Sudre *et al.*'s study. (Note: they're not all in IT!)



A model of the visual categorisation system

The categorisation system is often modelled as a **convolutional NN**. Le Cun & Bengio (1995); Riesenhuber & Poggio (1999); Krizhevsky *et al.* (2012)

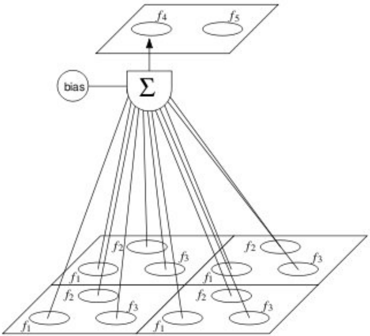
- Its input is a map of simple visual features.
- Each layer takes a map of features and returns a map of combined features.
- To avoid a combinatorial explosion, each layer also abstracts over space.



This is a reasonably good model of cells in the IT pathway.

A model of the visual categorisation system

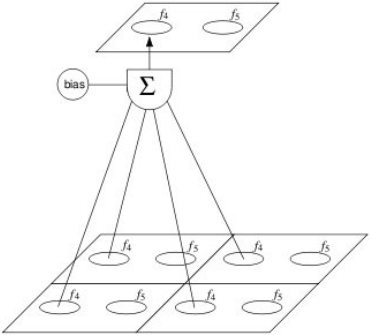
Combination layers look like this:



Shared weights are used, to simulate training at each retinal location.

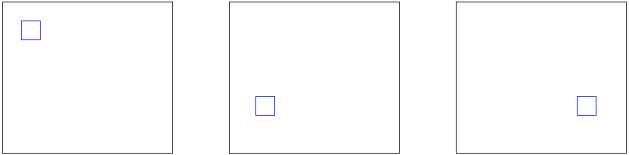
A model of the visual categorisation system

Abstraction layers look like this:

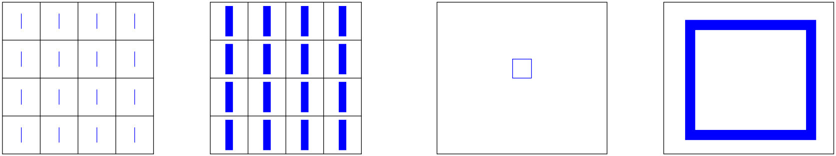


Translation and scale invariance of a convolutional NN

The abstraction operations allow an object to be categorised anywhere on the retina.

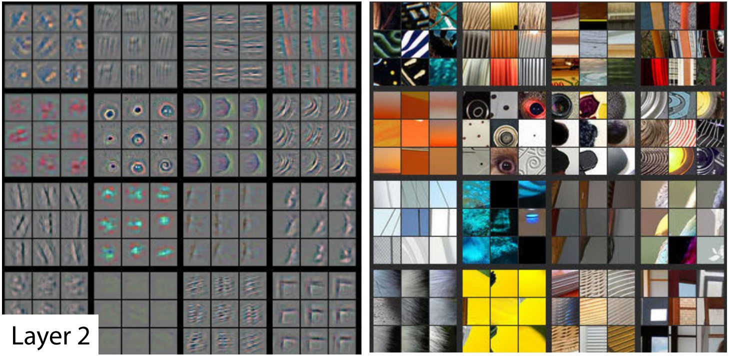


Input feature maps of different spatial frequencies allow an object to be categorised at a range of sizes.



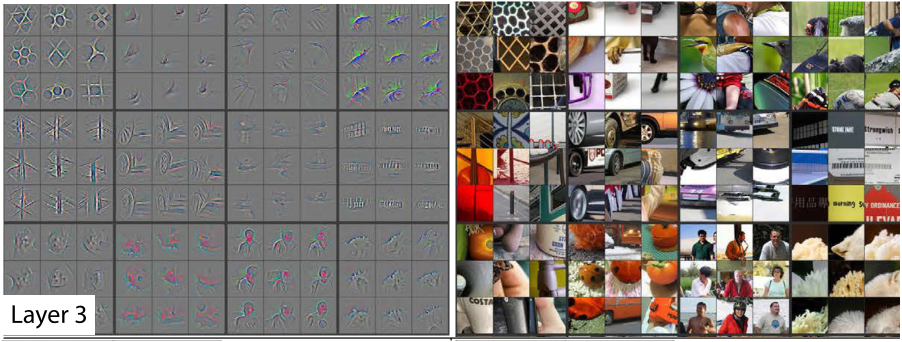
A modern convolutional NN

Here's a way of visualising the responses of units in different layers of a convolutional NN for object classification. (Zeiler and Fergus, 2014)



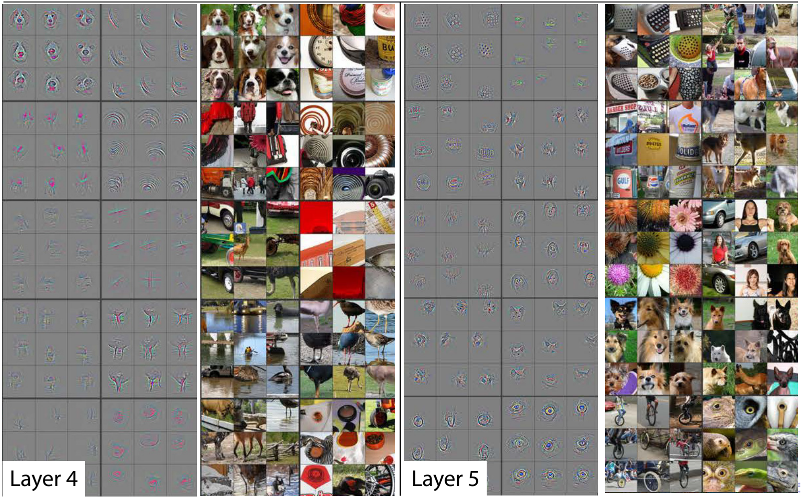
A modern convolutional NN

Here's a way of visualising the responses of units in different layers of a convolutional NN for object classification. (Zeiler and Fergus, 2014)



A modern convolutional NN

Here's a way of visualising the responses of units in different layers of a convolutional NN for object classification. (Zeiler and Fergus, 2014)

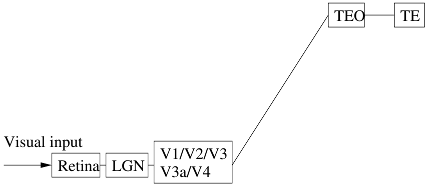


Top-down influences on object categorisation

Object categorisation is not only driven by perceptual information from the retina. There are top-down influences as well, which relate to the observer's *expectations*.

- Some expectations are general, and relate to the type of scene which the observer is looking at. These can be referred to under the umbrella term **priming**.
- Other expectations are specific, and relate to the particular objects which the observer has recently been looking at.

The object classification pathway



Outline of today's lecture

The attentional pathway

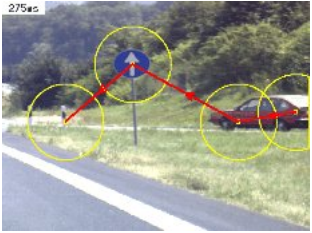
Cells at the end of the object classification pathway respond to specific shapes, but abstract away from retinal location. So how do we know *where* a classified object is?

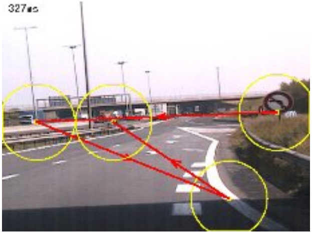
- There's a separate visual pathway in **parietal** cortex which represents the location of objects.
  Posterior parietal cortex is involved in converting retinal locations into locations in a *motor coordinate system*. (See Lecture 2.)
- There's another visual area called the **frontal eye fields** which is involved in controlling eye movements (and maybe other attentional operations), which also represents the location of objects on the retina.

Visual attention

The basic idea:

- The visual field is full of stimuli. We need to focus in on the most important ones.
- In the attentional pathway, we produce a *map* of the most important locations to attend to.
- Then we attend to them one by one.

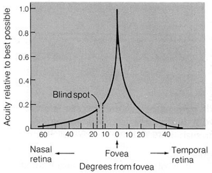




Overt attentional actions: saccades

Our eyes are designed to ‘focus’ on one location at a time.

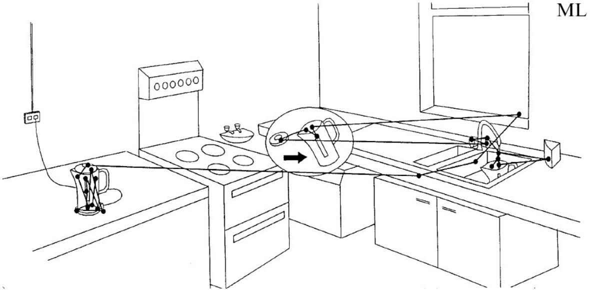
- The retina has a **fovea** at the centre, where visual acuity is hugely higher than in the periphery.



- The fovea ‘sees’ only 2 degrees of the visual field, but it contains about half the photoreceptors on the retina.
- We perceive the world by directing the fovea at a series of different locations.
- Eye movements are called **saccades**: we make around 3 a second, the whole of our waking lives.

Overt attentional actions: saccades

Here's a summary of the saccades recorded while an observer made a cup of tea (Land *et al.*, 1999):



Idea: the main job of peripheral vision is to build a map of important locations to attend to.

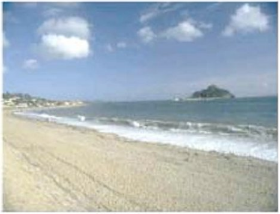
The saliency map

What counts as an ‘interesting location’?

- There are bottom-up things: e.g. *local contrast*, *movement*.
- There are also top-down things: e.g. low-level features of ‘something you are looking for’.

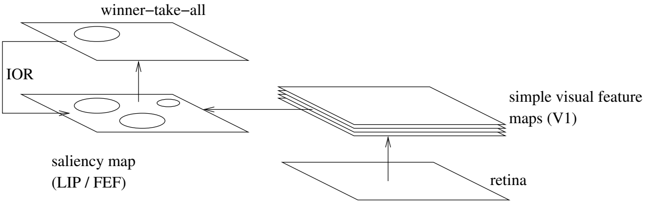
Itti *et al.* (1998) implemented a simple saliency map function.

- Its input comes from ‘early vision’: it’s a set of maps of simple visual features, at different orientations and scales.
- It computes local contrast in all these maps, and sums the results.



The saliency map

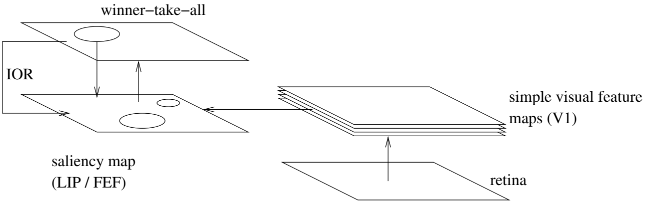
The saliency map projects to a **winner-take-all** (**WTA**) map, where active regions compete against one another, and the 'most salient region' is selected.



- After a time, an active region in the WTA map *inhibits* its corresponding region in the saliency map.
- This biases competition in the saliency map, and a new region becomes the winner.
- In this way, each of the salient regions is selected in turn.

The saliency map

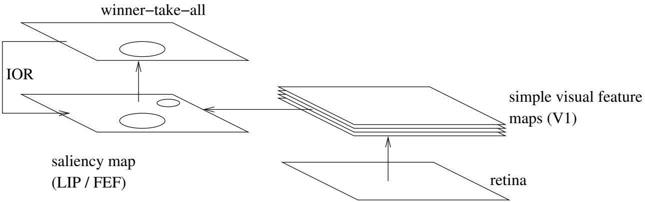
The saliency map projects to a **winner-take-all** (**WTA**) map, where active regions compete against one another, and the 'most salient region' is selected.



- After a time, an active region in the WTA map *inhibits* its corresponding region in the saliency map.
- This biases competition in the saliency map, and a new region becomes the winner.
- In this way, each of the salient regions is selected in turn.

The saliency map

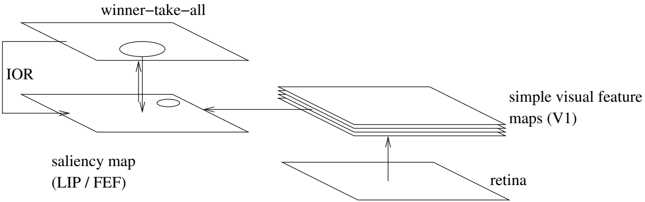
The saliency map projects to a **winner-take-all** (**WTA**) map, where active regions compete against one another, and the 'most salient region' is selected.



- After a time, an active region in the WTA map *inhibits* its corresponding region in the saliency map.
- This biases competition in the saliency map, and a new region becomes the winner.
- In this way, each of the salient regions is selected in turn.

The saliency map

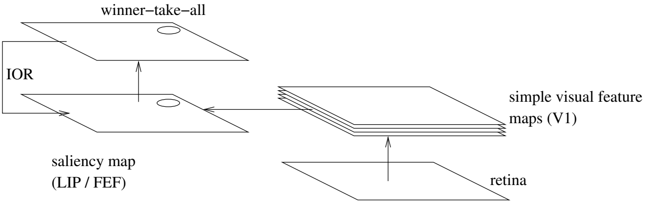
The saliency map projects to a **winner-take-all** (**WTA**) map, where active regions compete against one another, and the 'most salient region' is selected.



- After a time, an active region in the WTA map *inhibits* its corresponding region in the saliency map.
- This biases competition in the saliency map, and a new region becomes the winner.
- In this way, each of the salient regions is selected in turn.

The saliency map

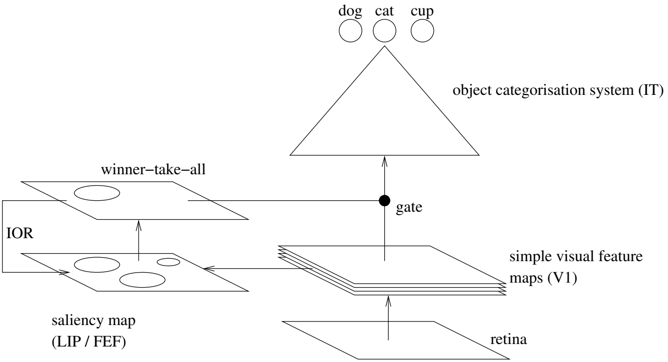
The saliency map projects to a **winner-take-all** (**WTA**) map, where active regions compete against one another, and the ‘most salient region’ is selected.



- After a time, an active region in the WTA map *inhibits* its corresponding region in the saliency map.
- This biases competition in the saliency map, and a new region becomes the winner.
- In this way, each of the salient regions is selected in turn.

The saliency map

Finally, we assume the most salient location *gates input to the classification system*.

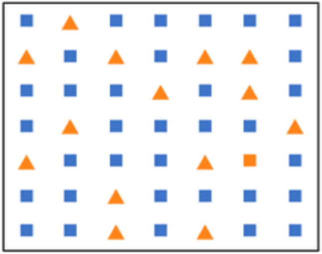


Now only the object which is attended to is classified.

Experimental evidence for visual attention

A lot of the evidence comes from **visual search** experiments.

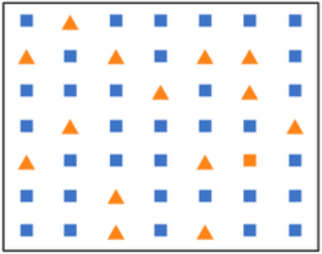
- The subject hunts for a target in a field of distractors.
- If the target is similar to the distractors, search time is proportional to the number of distractors.



Experimental evidence for visual attention

A lot of the evidence comes from **visual search** experiments.

- The subject hunts for a target in a field of distractors.
- If the target is similar to the distractors, search time is proportional to the number of distractors.
- If the target can be distinguished from the distractors by a ‘simple visual feature’, it ‘pops out’. . .



Is there a saliency map in the brain?

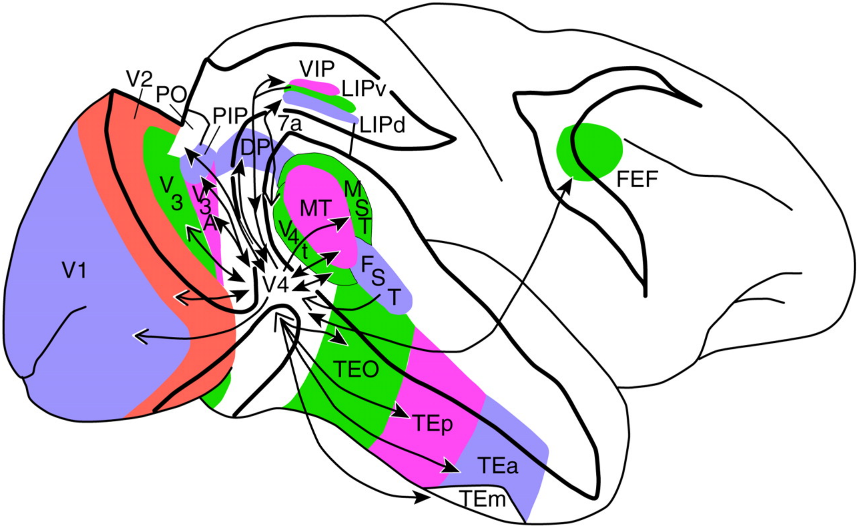
One area of parietal cortex—**lateral intraparietal cortex** (**LIP**)—seems a good candidate.

- LIP neurons respond more to 'newly appearing' stimuli (Gottlieb *et al.*, 1998).
- *Some* LIP neurons respond equally well to a salient stimulus whether the agent has to saccade *towards* it or *away from* it (Kusunoki *et al.*, 2000). This suggests they don't just encode saccades.

Another good candidate is the frontal eye fields.

- FEF cells respond better to stimuli which 'pop out' of visual displays (Bichot *et al.*, 2001).
- *Some* FEF cells encode salience, not saccades (Schall, 2004).

LIP and FEF


V2
PO
PIP
VIP
LIPv
LIPd
7a
DP
V3
V3A
MT
MST
V4t
FST
V4
V1
TEO
TEp
TEa
TEm
FEF

Overt and covert attention

LIP and FEF both influence eye movements.

- They project to the **superior colliculus**, which generates eye movements.

But we are also able to attend to peripheral objects *covertly*, without eye movements.

- Covert attention must involve gating retinal input to the classification system.
- Stimulating cells in FEF has been shown to modulate activity in the corresponding area of V4 (Moore and Armstrong, 2003).
- TMS of the human FEF improves detection of near-threshold stimuli (Grosbras and Paus, 2003).

Top-down influences on attention

At any given time, an agent has a particular **cognitive set**—one or more 'tasks' which s/he is actively pursuing.

- Cognitive sets are represented in **prefrontal cortex** (**PFC**).
- Some tasks are attentional—these are called **search tasks**.

During a search task, the target being searched for appears to be represented in PFC (Hagesawa *et al.*, 2000).

PFC also projects to FEF and LIP. And cells in these areas respond in a task-specific manner (Kusunoki *et al.* 2000; Bichot *et al.*, 2001).

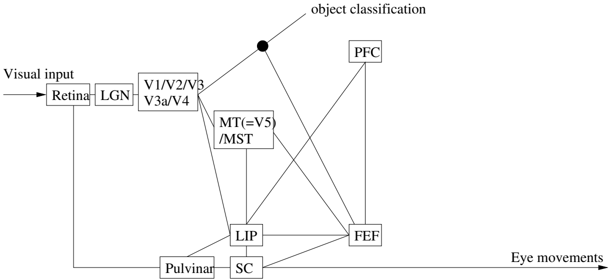
Top-down actions in visual search

Say we're searching for an object—a coke can. We can implement this by imposing two top-down biases:

- In one area of PFC, we can impose a **bias on the saliency maps** in FEF and LIP, to give preference to 'coke-can-like stimuli'.
- In another area of PFC, we can evoke a **representation of the search target** (a coke can), to be matched against incoming IT patterns.

Inhibition-of-return will cycle through the salient objects until a match is found, or until all items have been attended to.

The visual attention pathway



Summary

The model so far:

- Early vision represents simple visual features
- The object categorisation pathway combines visual features, and abstracts away from retinal location
- The attention pathway represents salient items in the visual field, and allows the agent to attend to these one by one (either overtly, or covertly).

Reading for today: Ch1; Ch2 Sections 1–4

Next week: visuomotor control (reaching and grasping).