Introduction to Neural Networks
U. Minn. Psy 5038

Representation of visual information

Neural codes and representation

Overview

- **Final home work**

- **Final projects**

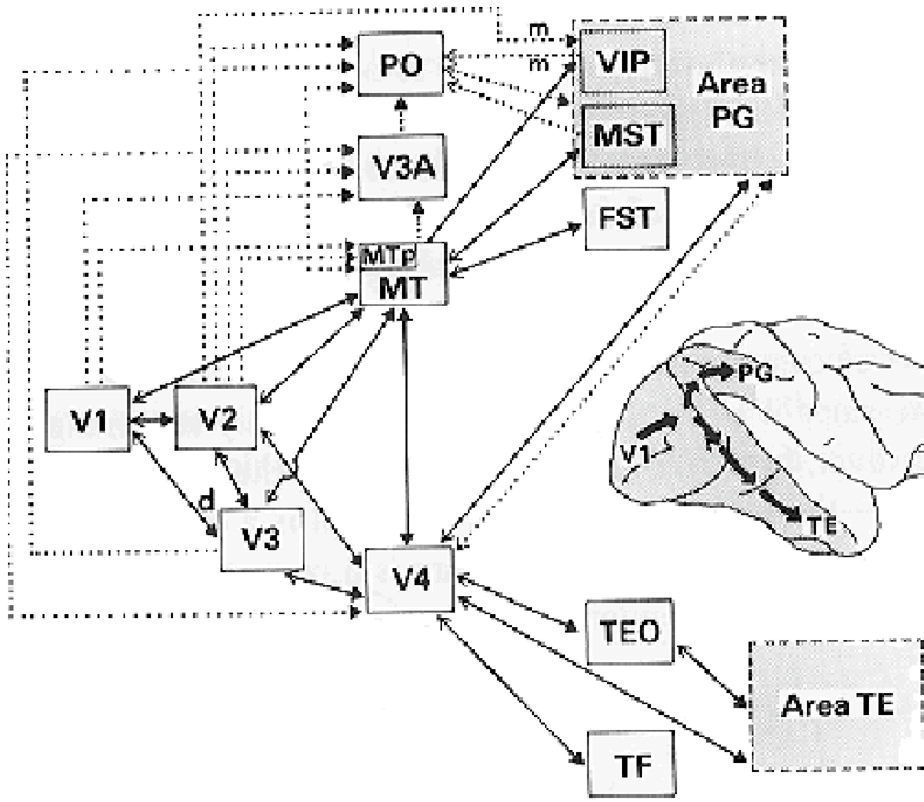
- **More on representation--tie with V1 orientation selectivity**

- **More on representation--contingent adaptation--tie with decorrelating neural networks**

Extra-striate cortical visual areas

V1 plus extra-striate areas, ~40-50% of primate cortex.

We've noted that cells in the primary visual cortical area send their visual information to an incredibly complex, and yet structured collection of extra-striate areas.



■ Connectivity between areas

Feedforward (ascending pathways) from mainly superficial layers (I,II, III) to layer IV of receiving areas. More diffuse feedback (descending pathways) from outside IV back to layers also outside IV (mainly I or VI). (~cm)

But there are also feedforward/feedback local circuits at a finer grain, within columns (~mm)

■ Two main large-scale functional pathways (involving mutliple areas and area-area connections)

What kind of information is represented in these areas? Any hypothesized function of striate cortex must eventually take into account what the information is to be used for.

Two primary functions of vision are: object perception and recognition--"within object" processing, and spatial processing (between object, and view-object relations). "Where" vs. "What"

■ Occipital-parietal pathway ("dorsal stream"): Spatial judgments, action ("where"/"how")?

V1,MT,MST,LIP,...

Viewer-centered computations

"where" or **"how"** or **"now"?**

Spatial computations,such as coordinate transformations for action

■ Occipital-temporal pathway ("ventral stream"): Object perception, recognition ("what")?

object perception, recognition pathway

V1, V2, V4, Posterior IT, Anterior IT, ...

view-independent, object-intrinsic, computations

Receptive field properties: Size grows: V1 = .5°-2°; TE (mean) = 26°

Response properties become more difficult to characterize. Earlier RF responses often show approximate additivity--response is determined by the combination of responses to shape and orientation of parts or features. Neural RFs in later areas (TE), more difficult to characterize. Respond to highly specific features, sometimes interpretable (faces), sometimes more abstract ("toilet brush-like thing").

Some information has to be discounted, and other information selected that is diagnostic for accurate classifiction.

Invariances required for recognition:

photometric: illumination level, direction, shadows

geometrical: translation, size, orientation in depth

category-related: levels of abstraction

Combinatorial problem of object representation and classification

grandmother cells, distributed codes, sparse codes

■ Behavioral evidence for distinction between ventral and dorsal pathway functions

Behavioral evidence: Double dissociation of function.

Monkeys

IT Lesions (Mishkin) showed impaired ability in matching tasks (non-match-to-sample task) but can do

spatial task.

Posterior parietal lesions: impaired ability to perform spatial task but can do matching

task.

Human patients.

Milner & Goodale's Patient D. F. (agnosic): Motor competence, good acuity, color, and motion perception. Normal intelligence and language comprehension.

But can't recognize objects or faces. Further, DF cannot judge orientation or size.

Motor interactions are close to normal. She can "post" a letter-like object into a slot, adjust grasp size correctly before touching the object.

How can she “post” letter into an arbitrarily oriented slot with her hand but cannot tell you in advance whether the orientation of the slot is vertical or horizontal!

(Other patients have been studied that show the opposite pattern of deficits)

Neural codes and representation: Demonstration of orientation adaptation

Let's return to the problem of representation at the level of quasi-homogeneous population of neurons, such as the collection of simple cells making up a hypercolumn or a collection of hypercolomns in V1. What might be the relationship between a perceptual judgment of a stimulus property (like orientation) and the receptive field properties of neurons in V1?

To motivate this problem and to introduce concepts of coarse coding, population or distributed codes, etc. (below), let's make a demo to study a well-known illusion involving adaptation.

Make stimuli

In[1]:=
```
width = 64;
grating[x_,y_,xfreq_,yfreq_] := Cos[(2. Pi)*(xfreq*x + yfreq*y)];
```

■ Left-slanted adapting grating

In[3]:=
```
xfreq = 4; theta = 0.8 * Pi / 2;
yfreq = xfreq / Tan[theta];

gleft = DensityPlot[grating[x, y, xfreq, yfreq], {x, 0, 1}, {y, 0, 1},
   PlotPoints → 64, Mesh → False, Frame → False, DisplayFunction → Identity];
```

■ Right-slanted adapting grating

In[6]:=
```
xfreq = 4; theta = 1.2 * Pi / 2;
yfreq = xfreq / Tan[theta];

gright = DensityPlot[grating[x, y, xfreq, yfreq], {x, 0, 1}, {y, 0, 1},
   PlotPoints → 64, Mesh → False, Frame → False, DisplayFunction → Identity];
```

■ Vertical test grating

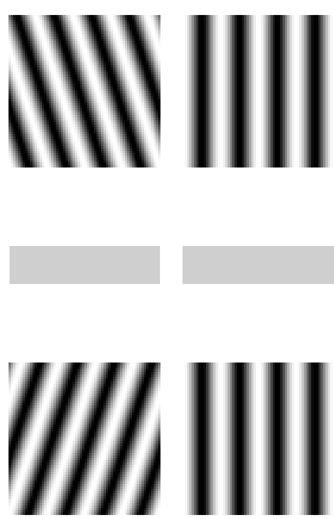
In[9]:=
```
xfreq = 4; theta = Pi / 2;
yfreq = xfreq / Tan[theta];

gvertical = DensityPlot[grating[x, y, xfreq, yfreq], {x, 0, 1},
   {y, 0, 1}, PlotPoints → 64, Mesh → False, Frame → False,
   DisplayFunction → Identity];
```

■ Gray fixation bar

In[12]:=
```
gbar = Graphics[{GrayLevel[.77], Rectangle[{0, .45}, {1, .5}]},
   AspectRatio → 1 / 4];
```

Test: Try it

In[13]:=
```
Show[GraphicsArray[{{gleft, gvertical}, {gbar, gbar}, {gright, gvertical}},
   DisplayFunction → $DisplayFunction ]];
```



What happens if you adapt with the left eye and test with the right eye?

Can this effect be explained in terms of changes to neurons in the retina? LGN?

Neural codes and representation

Can we explain orientation adaptation in terms of neural networks? To do this, we have to grapple with several questions: What are the "languages" of the brain? How is information represented? What is the information in a train of action potentials?

First some background concepts.

Firing rate

Firing rate correlates well with subjective intensity in sensory systems. What does it mean elsewhere in the brain?

"Labeled lines"

Suppose that when a particular cell fires it means something in particular, i.e. when neuron S fires at a rate of f spikes per second, then the animal must be looking at and recognizing its grandmother. Or when neuron T fires at a rate of g spikes per second, then the animal must be seeing a spot at location (x,y) of contrast 10%. Let's consider the implications.

A ganglion cell normally fires when stimulated by light coming from the upper left visual field. If the cell fires for any reason at all (e.g. you press on your eyeball), the fact that information is coming from this cell means "bright spot in upper left visual field". Similarly, for a pressure-sensitive cell on your finger tip. The identity of the cell that is firing represents information. Assuming a neuron has a "label" doesn't say much about how that label information gets passed around in the brain (other than by virtue of connectivity), but is useful for comparing behavioral/perceptual responses to neural measurements.

We noted that there are about 12 visual cortical areas in the primate (macaque) with topographic maps of the visual field. Spatial location is represented on the cortical surface. If cells are labeled lines for position, then excitation of a cell signals information about location. But extra-striate areas have increasingly larger receptive fields with a coarser representation of space (see below).

Over 100 years ago, William James (1890) proposed the thought experiment that if we could splice the nerves so that the excitation of the ear sends input to the brain's visual area, and vice versa, we would “hear the lightning and see the thunder”. A study has re-wired visual input to the auditory cortex of the ferret. Surgical methods can be used to hookup retinal fibers to the auditory thalamus in the ferret neonate. Retinal projections that become redirected to the auditory thalamus then have visually responsive cells in auditory thalamus and cortex, form a retinotopic map in auditory cortex and have visual receptive field properties with orientation selectivity like those in visual cortex (Sharma et al., 2000). But contrary to James, the auditory cortex of the ferret showed *plasticity*--it not only adjusted cells to become like the visual area "V1" in terms of receptive field properties, but further the behavior was consistent with the experience of sight being derived from visual inputs to their re-wired auditory cortex (von Melchner et al., 2000).

Distributed representations

Earlier in the course we discussed distributed vs. "grandmother" cell (or "local") representations of an object or event. Consider object memory. Suppose we have n neurons that can be active or inactive. In a grandmother cell representation, the activity of a single unit signals a unique object. There are strong theoretical arguments against a grandmother cell representation for objects--one needs a new neuron for every new object, i.e. there would be a single neuron whose firing would uniquely signal your "grandmother", hence the name. ("yellow volkswagen" detectors is another phrase of historical interest.). Representational capacity is n.

In a distributed representation, object identity is represented by the pattern. Advantages to distributed coding?

· Capacity: If there are m distinguishable levels for each neuron, the system can represent m^n objects.

• Similarity between two patterns can be represented in a graded way, via the correlation (e.g. dot product, or angle between them, or cosine).

...but how are decisions made? i.e. "this is or is not my grandma". (If it is by another layer that matches the distributed code to a template using a TLU, then we have in effect added grandmother code!)

Although there is no direct evidence in humans about neural responses to grandmothers, there is data on the response of neurons to Jennifer Aniston (and other famous people), see Quiroga et al. (2005).

Sparse (vs. fully) distributed representations

The latter case, m^n, would be a fully distributed system. But we noticed before that the cortex seems to be quiet on average--e.g. most V1 cells at any given moment are not firing. So maybe the truth is somewhere between, and an object is coded by a small population that is active for an event. This is *sparse coding*. We have n units, but an object is represented by the firing of p units, where 1<p<<n.

One possible advantage for sparse coding is that neurons that fire the most could mean something like "grandmother", the spread in the pattern of firing could represent uncertainty ("probably grandmother"). I.e., assume similarity corresponds to cortical distance, then if a few neurons are very active around the "grandmother" lines (and others are quiet), then it is almost certainly grandmother. However, if lots of other neurons are also active, although less so, then "well, it might be grandmother". This raises the possibilty that neural populations code for uncertainty as well as central tendencies, such as identity. See Pouget et al., (2000) and Knill & Pouget (2004) for a discussion of population coding in terms of representing probability distributions.

We discussed coding in terms of objects, but the issue is relevant for any kind of information. Later we'll talk about sparse coding of images. I.e. for any given image, say 256x256 8-bit pixels, what are the properties of V1 coding schemes that represent the image? The type of coding interacts with the statistical structure of the set of events to be encoded. If images were arbitrary, i.e. any image was equally likely (even TV snow), then we'd require representational space equal to the task (e.g. max representational capacity is 256x256x8 bits, or 2^(256x256x8) possible signals). But if there is statistical structure, we could get by with less. The space of natural images is much much smaller than 256x256x8 bits (Kersten, 1987). Further, it turns out that the Gabor set described in the previous lecture produces a sparse distributed code for natural images. (Related to wavelet signal compression methods). There is more than one cell activated for any given image, but the number is relatively small. Recall the discussion of Olshausen and Field in the previous lecture.

Coarse vs. fine coding

We saw in the last lecture that a neuron can be "tuned" to various features or dimensions of an input pattern or stimulus, e.g. for simple cells: position, orientation, spatial frequency, spatial phase, motion direction and speed, ocularity.

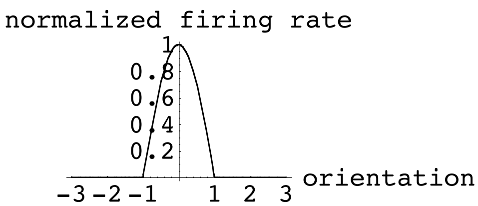
Features can be coarsely sampled (few detectors to span the range), and the receptive fields broad (so no empty regions). Broadly tuned cells mean that similar inputs to the cell's preferred input also fire the cell. *Coarse coding* with broad tuning functions result in overlap of the tuning. For example, an image could be sampled at only a few spatial locations, but if the receptive fields span a large regions of space, there will be sufficient overlap that any stimulus no matter how small with stimulate some of the neurons.

Fine coding means that the neurons finely or "densely" sample the feature space, typically with correspondingly narrower tuning functions and receptive fields that are more closely packed. Neurons are more closely tuned to the exact feature, and show little or no response to similar features.

How can information be represented given a coarse code?

Let's construct an hypothetical tuning function for some feature s (e.g. orientation of the above gratings) with tuning width w:

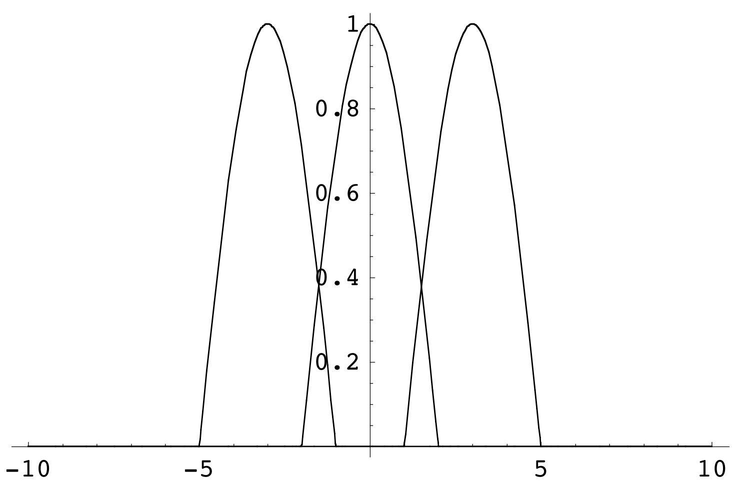
In[14]:=
```
truncCos[s_, w_] := If[-w / 2 < s < w / 2, Cos[Pi * (1 / w) * s], 0];
Plot[truncCos[s, 2], {s, -3, 3},
  AxesLabel → {"orientation", "normalized firing rate"}];
```



Try the above cell with various values of w.

Coarse coding

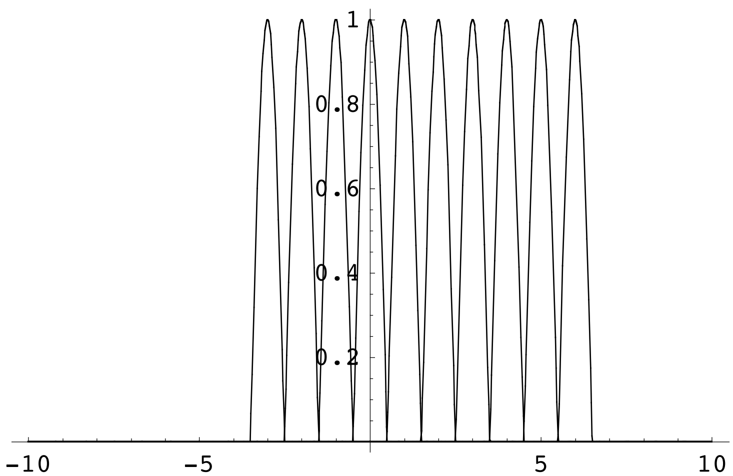
In[16]:=
```
w = 4;
R[i_, x_] := truncCos[x - i, w];
Plot[{R[-3, x], R[0, x], R[3, x]}, {x, -10, 10}];
```



E.g. wavelength coding in the retina, where humans have three types of cones, overlapping but each with different peak spectral sensitivities.

Finer coding

In[19]:=
```
w = 1.0;
R[i_, x_] := truncCos[x - i, w];
Plot[{R[-3, x], R[-2, x], R[-1, x], R[0, x], R[1, x], R[2, x],
   R[3, x], R[4, x], R[5, x], R[6, x]}, {x, -10, 10}];
```



Here we discretely sample more values (e.g. orientation or wavelength), but without overlap. There is almost always overlap between tuning functions.

Increase the degree of overlap in the above

E.g. positional coding by cones in retina. V1 shows finer coding for position than extra-striate topographic areas, like V2. Suggests that V1 is important for fine spatial tasks, such as vernier acuity. Orientation is another example. Although there isn't definitive evidence for discrete class as with cones for wavelength, if one assumed that cortical cells have a mean orientation bandwidth of about 15% (at half-width), a model might be chosen to have 12 "channels" or 12 classes of receptive fields each with a different orientation in order to span orientation space (=180/12).

Population vector coding

■ Definition

Receptive fields typically overlap (e.g. a bright spot at one location creates a neural point spread function, the "projective field"). E.g. a bar at one orientation will more or less activate cells within a certain feature range (+/- 15 deg in V1).

Cell S firing at a rate of g spikes per second may not tell us the orientation of the stimulus (because the rate per second also depends on other dimensions of the stimulus, such as contrast).

But suppose we have access to the responses of a bunch of neurons all "seeing" the same stimulus bar. How can information (e.g. about orientation) be extracted from this pattern of activity? We aren't going to answer this with a neural mechanism, but rather with an interpretive measure that an experimenter could employ, and for the moment leave the mechansim for speculation.

One can combine information across a population in terms of a "population vector".

Let x_k be a vector representing a stimulus feature (e.g. the kth 2-D position, or kth motion direction, or kth orientation, etc.). Let the firing rate of the ith cell be $R_i(x_k)$ in response to input x_k. Let the feature that produces the peak response of the ith cell be x_i^p --i.e. x_i^p is the ith cell's preferred feature, the one that fires the cell the most (the center of the tuning function, so $R_i(x_i^p)$ is the maximum firing rate).

Given an input feature x_k, the firing rate of the ith cell can be interpreted as the strength of its "vote" for its preferred feature.

So for the example of positional coding, position could be represented by the weighted average over the population of cells, each responding with various firing rates to x_k:

$$\mathbf{x} = \sum_i R_i(x_k)\, x_i^p \tag{1}$$

Analogous to computing the center of mass, or an average, we can normalize the estimate by the total activity:

$$\mathbf{x} = \sum_i R_i(x_k)\, x_i^p \Big/ \sum_i R_i(x_k) \tag{2}$$

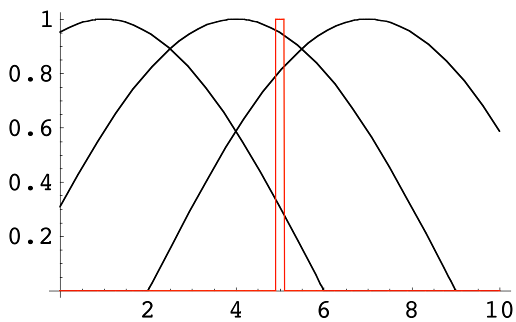
This measure has not only been applied to modeling sensory coding, but also in the motor system and cognitive processes involving direction of movement (Georgopoulos et al., 1993). In certain tasks the population vector can be measured in real neuronal ensembles and be seen to evolve in time consistent with behavioral measures (mental rotation, reach planning).

■ Illustration of population coding

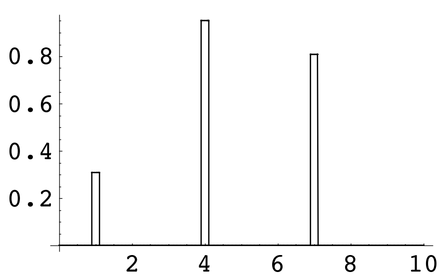
Simple example. Population vector will be 1-D, i.e. just a scalar. Coarse coding with only 3 units, and broad tuning with width = 10.

First we predict the response of the three neurons to a particular input. Suppose the input to the population is xinput = 5.

In[73]:=
```
xinput = 5; (*stimulus feature*)

w = 10;
R[i_, x_] := N[truncCos[x - i, w]];
input[x_] := If[Abs[x] < .1, 1, 0];
(* just for plotting purposes-- a narrow deltafunction-like pulse *)

prefx = {1, 4, 7}; (*preferred features for the 3 cells, i.e. rf centers*)

responses = Table[R[prefx[[i]], xinput], {i, 1, Length[prefx]}];

Plot[{R[prefx[[1]], x], R[prefx[[2]], x], R[prefx[[3]], x],
   input[x - xinput]}, {x, 0, 10},
  PlotStyle → {RGBColor[0, 0, 0], RGBColor[0, 0, 0], RGBColor[0, 0, 0],
    RGBColor[1, 0, 0]}];
```



In[78]:=
```
Plot[ {responses[[1]] input[x - prefx[[1]]],
   responses[[2]] input[x - prefx[[2]]],
   responses[[3]] input[x - prefx[[3]]]}, {x, 0, 10}, PlotPoints → 100];
```



Population vector response

Now we go backwards: from the three activity levels, can we say what the input was?

In[79]:=
```
xpop = responses.prefx / Apply[Plus, responses]
```

Out[79]= 4.72496

How could accuracy be improved?

Noise and quantization

Suppose position is coarsely coded by just two neurons, with peak sensitivity say at 0 and the other at position 1, but with broad receptive fields spanning both. Does this mean that position can't be represented accurately and reliably?

No. There are only 3 cone photoreceptor types for daytime vision (spanning 380 to 750 nanometers), but our ability to discriminate wavelength is on the order of nanometers. And we can distinguish thousands of colors.

How can you use the population vector idea to explain this? What if each neuron or photoreceptor can only reliably signal 2 levels? Then what is accuracy like?

Discussion or project question: Where might wavelength discrimination be best?

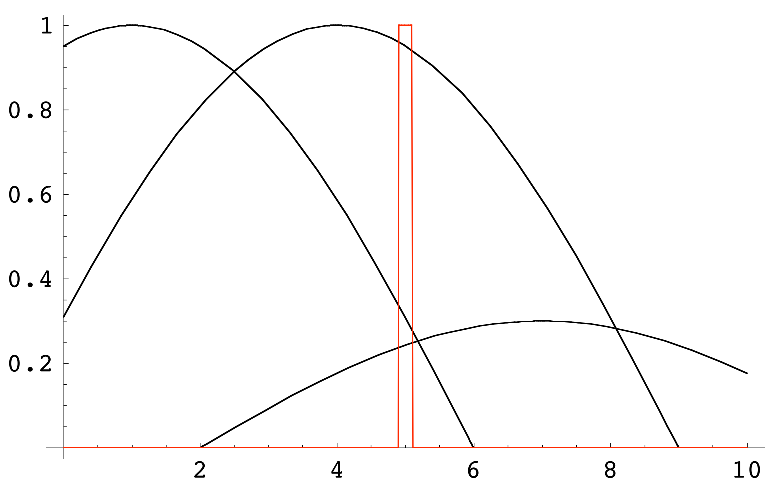
At the peak sensitivities? At the cross-over points? First imagine just one channel or tuning function.

Adaptation and population coding

Suppose that after viewing a stimulus at the preferred value for a channel, its response decreases (as with the orientation demo above). We'll model the change in response level due to adaptation our our three neurons by 0< adaptstrength[i] <1, where 0 means completely adapated so that there is no response, and 1 means that it was unaffected by the adaptation time.

In[80]:=
```
adaptstrength = {1, 1, 0.3};
xinput = 5; (*stimulus feature*)

w = 10;
R[i_, x_] := N[truncCos[x - i, w]];
input[x_] := If[Abs[x] < .1, 1, 0];
(* just for plotting purposes-- a narrow deltafunction-like pulse *)

prefx = {1, 4, 7}; (*preferred features for the 3 cells*)

responses = Table[R[prefx[[i]], xinput], {i, 1, Length[prefx]}];
responses = adaptstrength responses

Plot[{adaptstrength[[1]] * R[prefx[[1]], x],
   adaptstrength[[2]] * R[prefx[[2]], x],
   adaptstrength[[3]] * R[prefx[[3]], x], input[x - xinput]},
  {x, 0, 10},
  PlotStyle → {RGBColor[0, 0, 0], RGBColor[0, 0, 0], RGBColor[0, 0, 0],
    RGBColor[1, 0, 0]}];
```

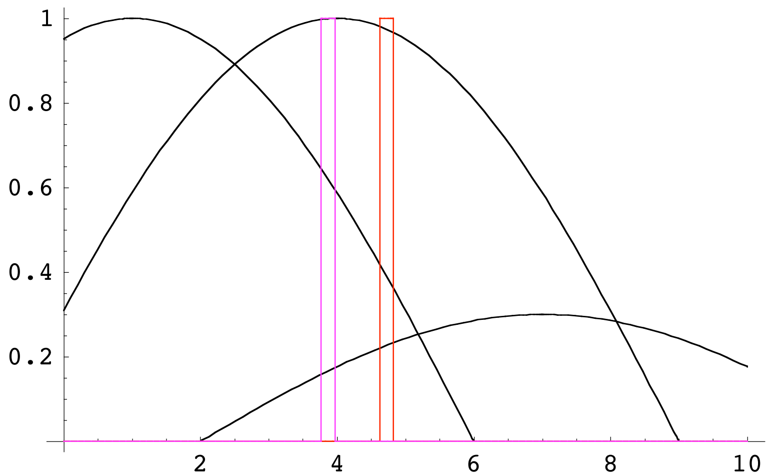
Out[87]= {0.309017, 0.951057, 0.242705}



In[89]:=
```
xpopadapt = responses.prefx / Apply[Plus, responses]
```

Out[89]= 3.86762

In[90]:=
```
Plot[{R[1, x], R[4, x], 0.3 * R[7, x], input[x - xpop], input[x - xpopadapt]},
  {x, 0, 10},
  PlotStyle → {RGBColor[0, 0, 0], RGBColor[0, 0, 0], RGBColor[0, 0, 0],
    RGBColor[1, 0, 0], RGBColor[1, 0, 1]}, PlotPoints → 100];
```



Is this population vector account the right explanation of orientation after-effect? (See Carandini, 2000).

But is firing rate the "code"? Timing

Information in the detailed timing of spikes? See F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek (1996).

Information in the temporal coherence across spiking ensembles? "Binding by synchrony", see Shadlen & Movshon.

Modeling large-scale neural systems & systems analysis

Much of the modeling of visual processing has been built on the tools that we've learned about.

- *generic feedforward neural network models*

But there are many aspects of functional brain modeling that require additional tools and ways of thinking.

■ Modeling information processing functions

*Neural representation and coding:

grandmother cells, distributed codes, sparse codes, coarse codes

binding problem, "binding by synchrony"

Large-scale architectures (e.g. Inter-area processing) -- what is represented in the various cortical visual areas?

*Feedback

Information processing roles of feedback

*Dynamical behavior

*Memory

Timing and sequences (e.g. speech, motor sequences)

Dynamical issues for real-time control, visuo-motor control

*Efficient coding, dimensionality reduction

*Handling uncertainty - Probabilistic models

■ Measuring and characterizing neural systems

Linear and non-linear systems analysis, statistical and stochastic processes analysis (time series),...

See "Spikes" by F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek (1996).

References

Barlow, H. B., & Foldiak, P. (1989). Adaptation and decorrelation in the cortex. In C. Miall, R. M. Durban, & G. J. Mitchison (Ed.), The Computing Neuron Addison-Wesley.

Barlow, H. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. Vision Research, 30(11), 1561-1572.

Carandini M. 2000. Visual cortex: Fatigue and adaptation. *Curr Biol* 10: R605-7

Fang, F., Murray, S. O., Kersten, D. J., & He, S. (2005). Orientation-tuned fMRI adaptation in human visual cortex. J Neurophysiol.

Field, D. J. (1994). What is the goal of sensory coding? Neural Computation, 6, 559-601.

Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B., & Massey, J. T. (1989). Mental Rotation of the Neuronal Population Vector. *Science, 243*, 234-236.

Kersten D. 1987. Predictability and redundancy of natural images. *J Opt Soc Am A* 4: 2395-400

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci, 27(12), 712-719.

Lee, C., Rohrer, W. H., & Sparks, D. L. (1988). Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature, 332*(6162), 357-360.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381, 607-609.

Pouget A, Dayan P, Zemel R. 2000. Information processing with population codes. *Nat Rev Neurosci* 1: 125-32

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature, 435(7045), 1102-1107.

F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek (1996). Spikes: Exploring the neural code (MIT Press, Cambridge).

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron, 24*(1), 67-77, 111-125.

Sharma, J., Angelucci, A., & Sur, M. (2000). Induction of visual orientation modules in auditory cortex. Nature, 404(6780), 841-847.

Sillito, A. M., Jones, H. E., Gerstein, G. L., & West, D. C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. Nature, 369, N. 9, 479-482.

von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature, 404*(6780), 871-876.