Computational mechanisms for the motion processing in visual area MT

(yet another model for the area V1-MT)

Hueihan Jhuang*, Thomas Serre[+], Tomaso Poggio*

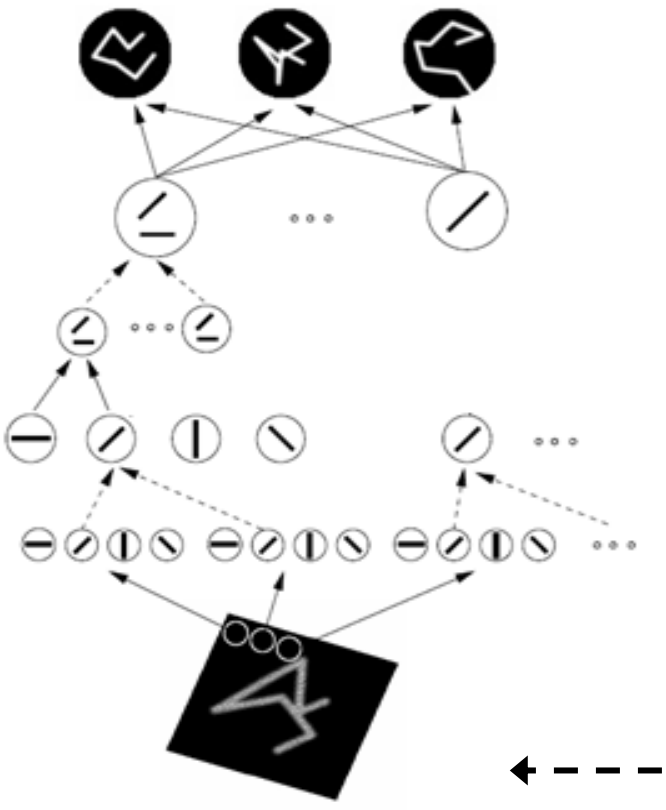
***MIT, [+] Brown University**

Nov. 9, 2010

Outline

- **The model**
- Experiments
  - Pattern - component MT cells (Movshon , Adelson, Gizzi & Newsome, 1985)
  - Local direction integration of MT cells (Majaj, Carandini & Movshon, 2007)
  - Speed tuned MT cells (Priebe , Lisberger &Movshon, 2006)

A general class of hierarchical models



max over space and scale.

Linear combination *(with nonlinear normalization)* of C1

max over space and scale

Linear combination *(with nonlinear normalization)* over space

Weighted sum

(selectivity/ complexity)

MAX (invariance)

(Riesenhuber & Poggio, 1999)

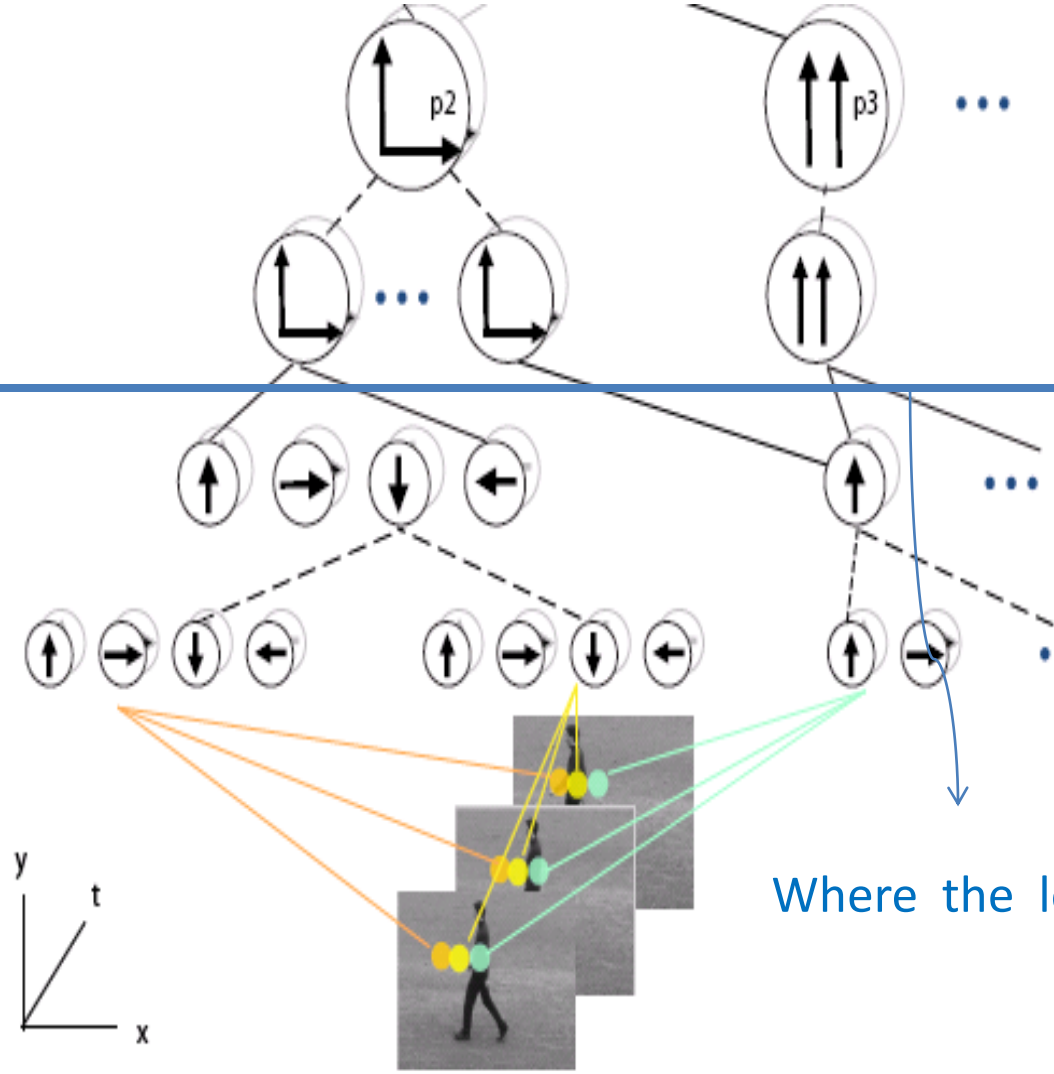
More hierarchical models : Fukushima 1980; Lecun et al, 1998; Wersing & Koerner, 2003, 2005; Amit & Mascaro, 2004;Ranzato et al, 2007)

Apply to object recognition (Serre 2005, 2007; Bileschi & Wolf, 2006; Mutch & Lowe, 2006,2008)

Matching to the ventral stream (Kouh & Poggio 2004; Serre 2006; Cadieu , et. al., 2007)

Apply to biological motion recognition (Giese & Poggio, 2003; Casile & Giese 2005; Sigala et al, 2005)

The model



C2 max over space and scale.

S2 Linear combination *(with nonlinear normalization)* over direction and space

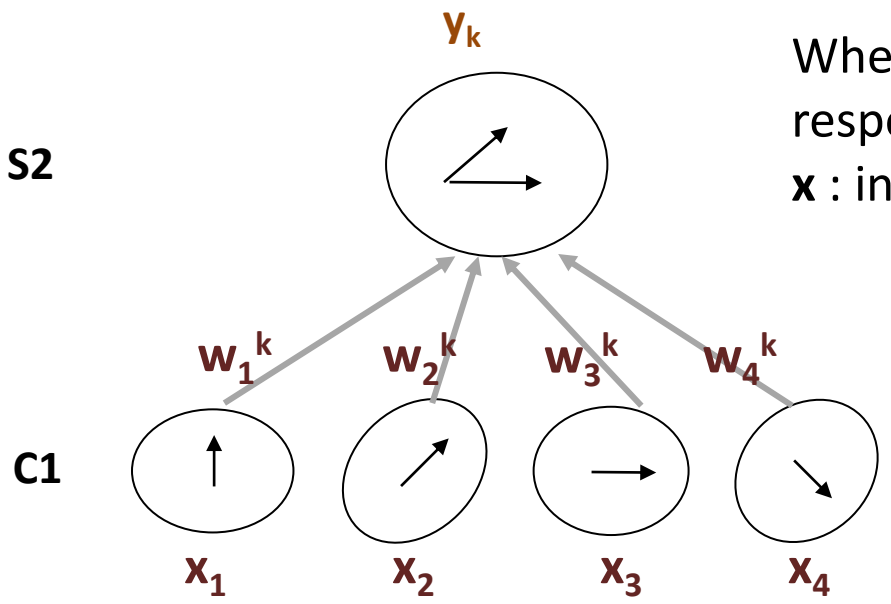
C1 max over space and scale

S1 Linear combination *(with nonlinear normalization)* over space and time

Where the learning from training data occurs

S1 units are 3D Spatio-temporal Gabor filters with nonlinearity

The connection between C1 and S2 layer



S2 weights and responses

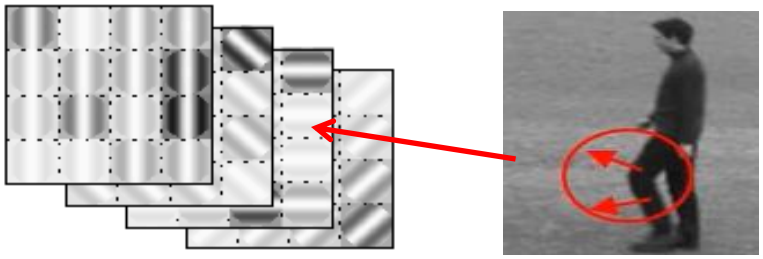
Where the weights **w** are extracted from C1 responses of training videos,
x : input c1 responses

$$y_k = \frac{\mathrm{x} \cdot \mathrm{w}^k}{||\mathrm{x}|| \times ||\mathrm{w}^k||}$$

w^k denote the k-th stored template

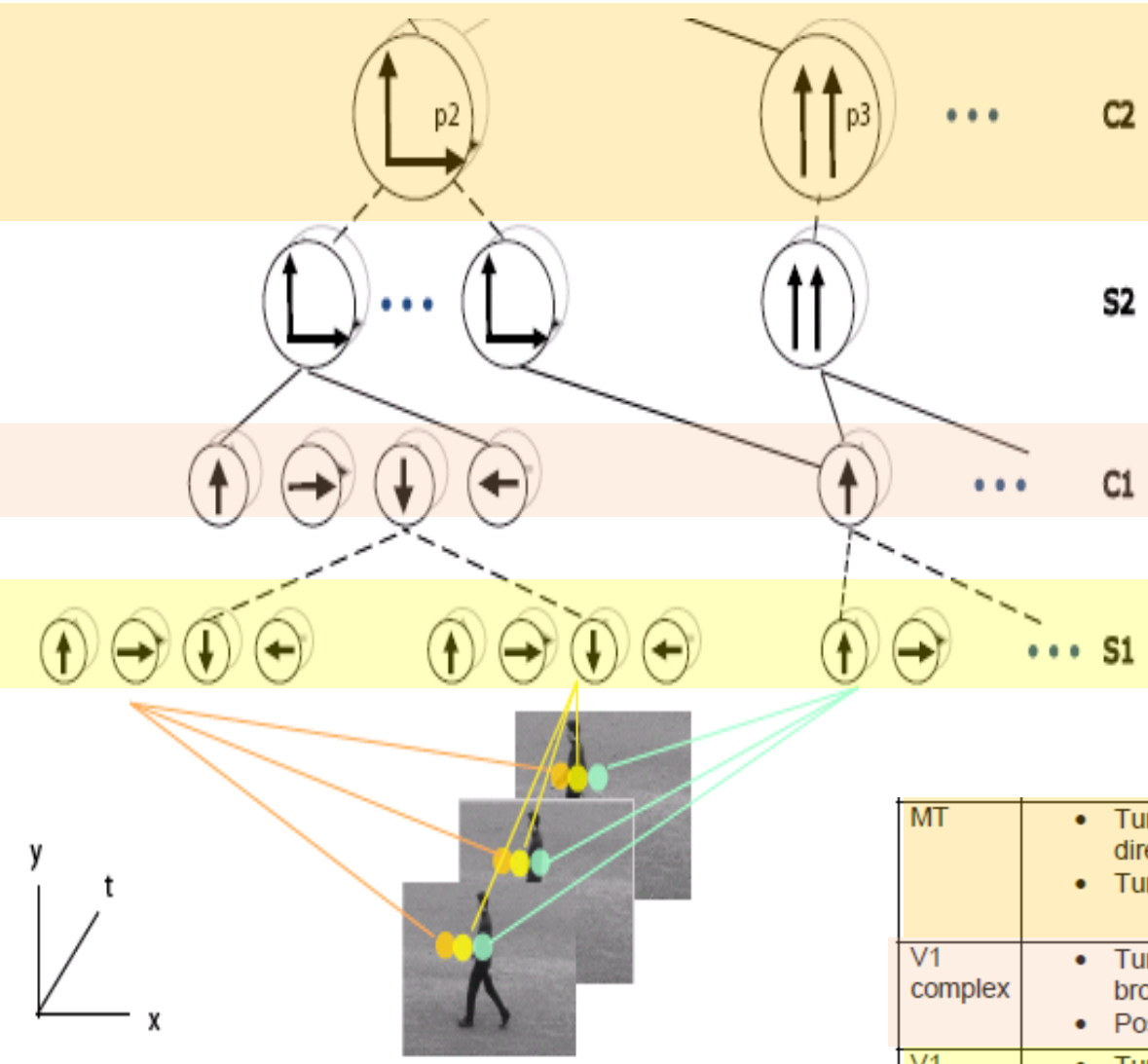
x denote a C_1 patch

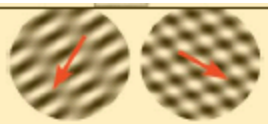
computer-vision viewpoint: S2 layer does template matching



See the previous talk by Tan et al. for more about template matching in S2 layer

The possible visual areas that match the model layers



MT	• Tuned to pattern or component directions • Tuned to speed	
V1 complex	• Tuned to component directions and a broad range of spatial frequency • Position invariance	
V1 simple	• Tuned to spatial and temporal frequency • Tuned to component directions	

Comparison of C1 units with V1 cells

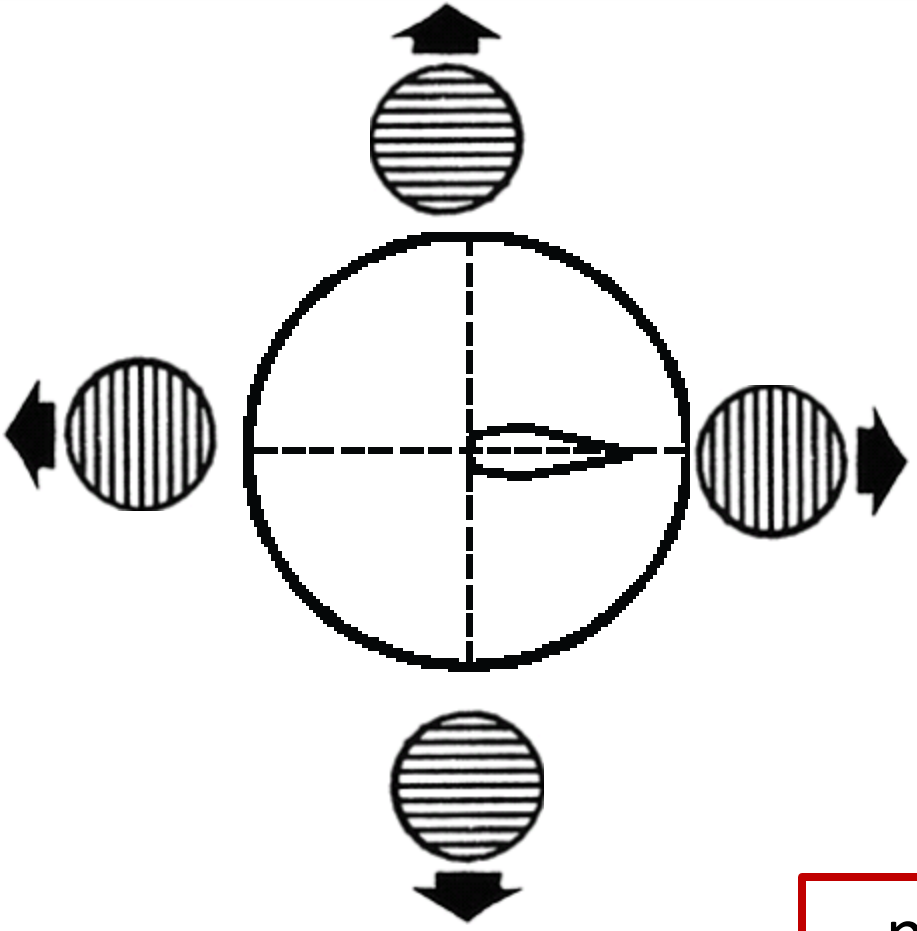
		C1 units	V1 cells
RF size	Range	0.63-3.8^{o}	0.6° - 4° (Mikami et al, 1986)
	mean	2.2°	2.2° (Mikami et al, 1986)
Preferred Spatial frequency	Range	0.3-6.1 cycle/°	0.5-8 cycle/° (Foster et al, 1986)
	Mean	1.7 cycle/°	2.2 cycle/° (Foster et al, 1986)
Spatial frequency BW	Range	0.8-3.1 octave	0.8-3.0 octave (Foster et al, 1986)
	Mean	1.5 octave	1.8 octave (Foster et al, 1986)
Preferred temporal frequency	Range	0.5 – 15.4 cycle/s	0.5-12cycle/s (Foster et al, 1986)
	Mean	4.4 cycle/s	3.7 cycle/s (Foster et al, 1986)
Temporal frequency BW	Range	0.94-4.6 octave	
	Mean	2.2 octave	2.9 octave (Foster et al, 1986)

Dynamics of V1 RF (DeAngelis, Ohzawa & Freeman, 1995)
Local Geometry in the visual system (Koenderink & Van Doorn, 1987)
Gaussian derivative filters as V1 directional selective cells (Simoncelli & Heeger, 1997)

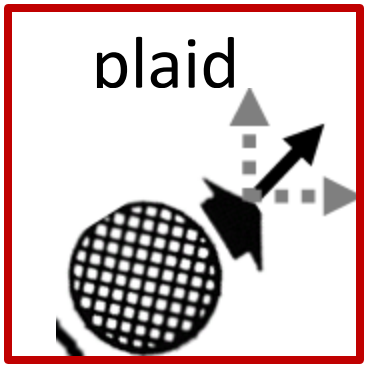
Outline

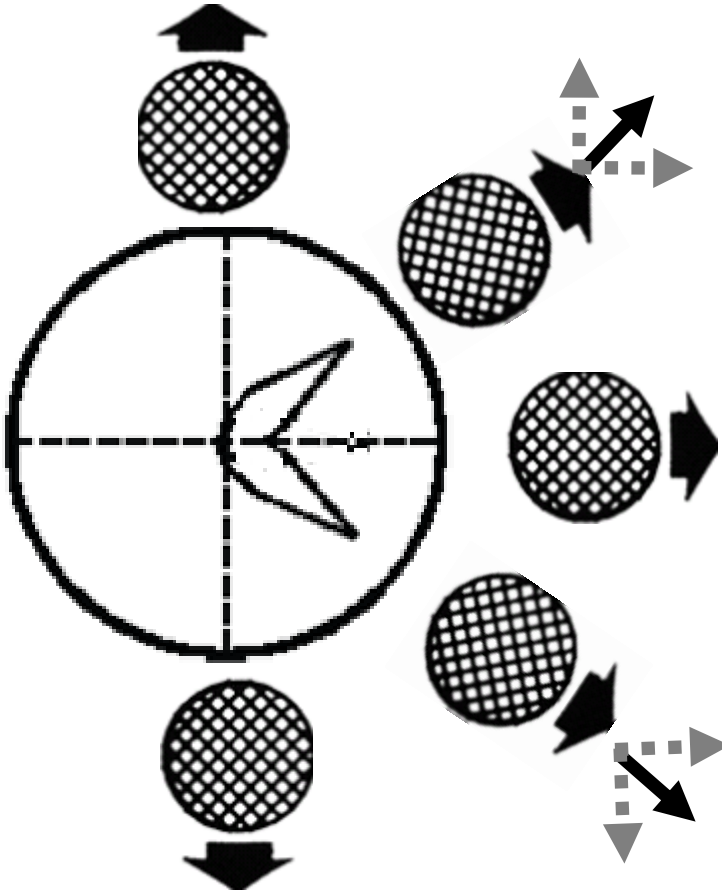
- The model
- Experiments
  - **Pattern - component MT cells (Movshon , Adelson, Gizzi & Newsome, 1985)**
  - Local direction integration of MT cells (Majaj, Carandini & Movshon, 2007)
  - Speed tuned MT cells (Priebe , Lisberger &Movshon, 2006)

Direction tuning curve of a typical V1 cell and an ideal MT component cell



grating

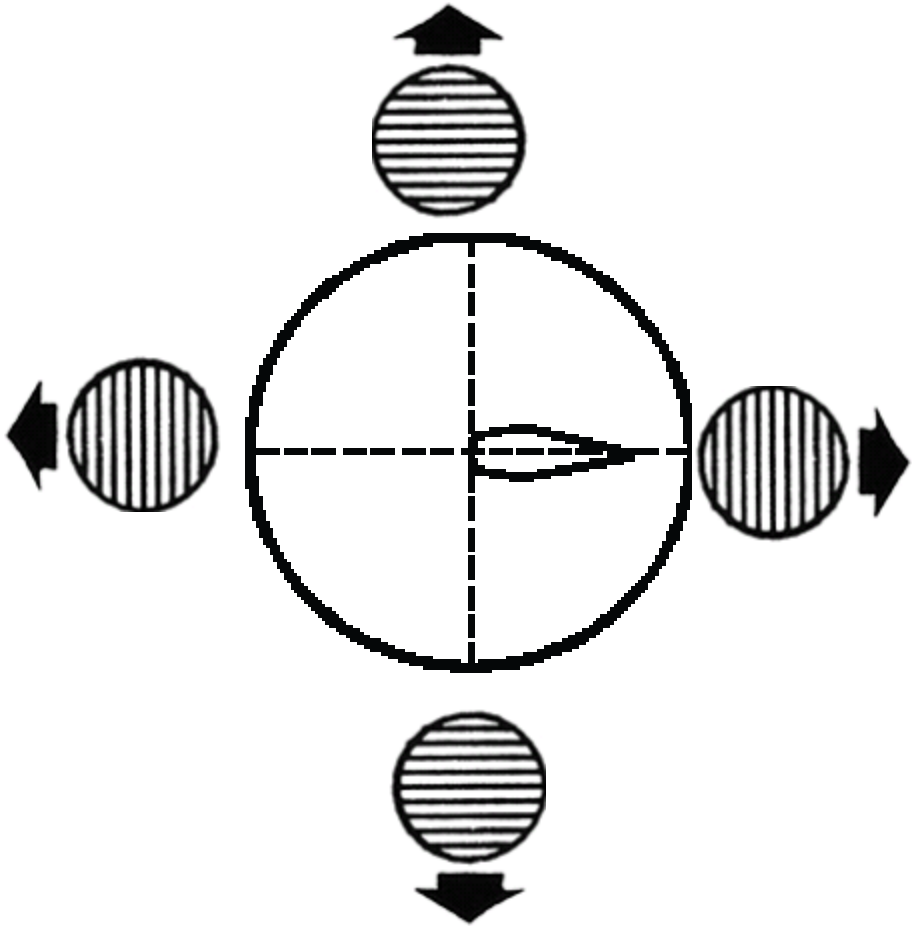


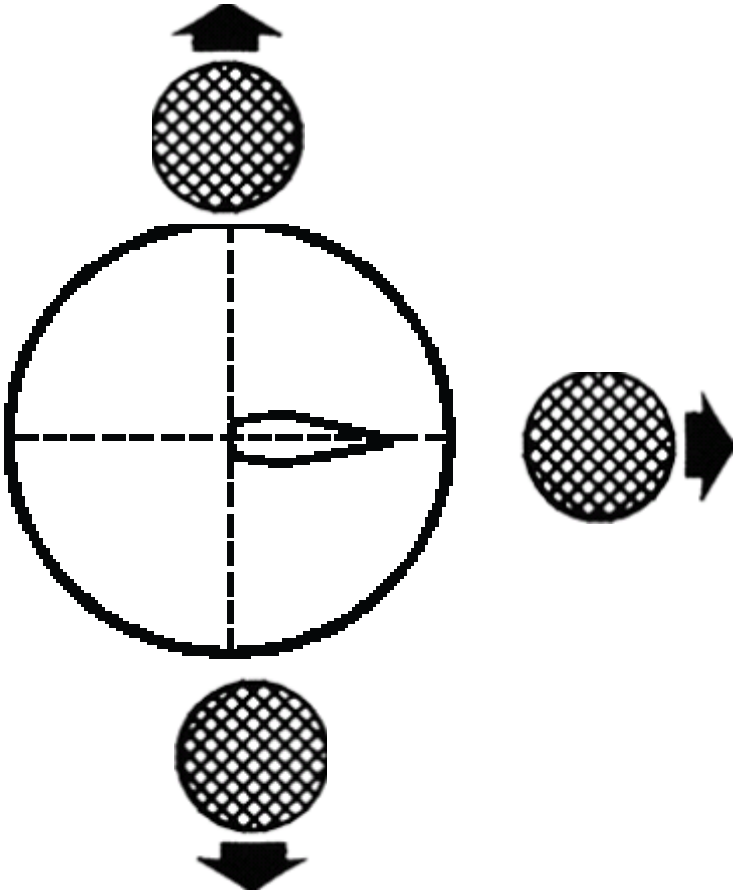


plaid

(Movshon , Adelson, Gizzi & Newsome, 1985)

Direction tuning curve of an ideal MT pattern cell

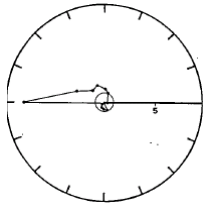
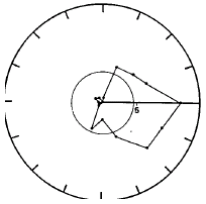
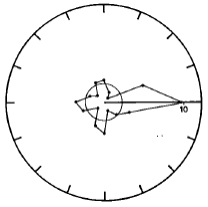
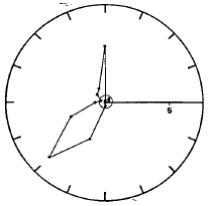
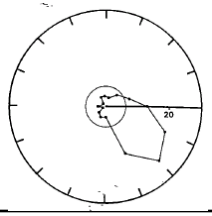
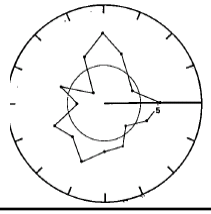




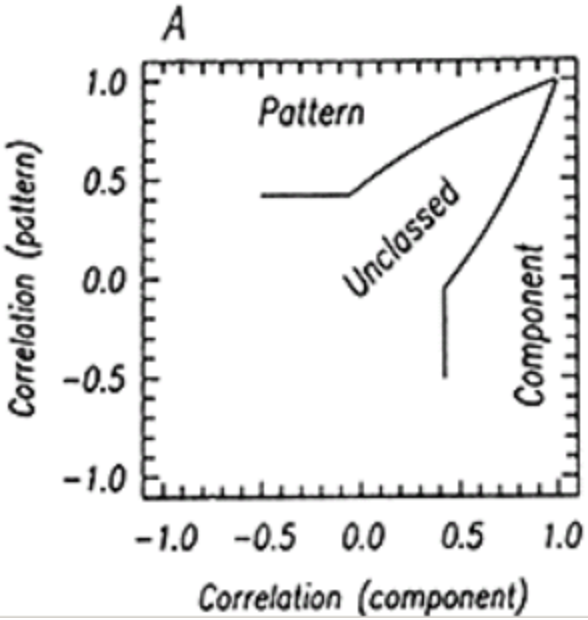
grating

(Movshon , Adelson, Gizzi & Newsome, 1985)

Direction tuning curve of typical MT cells

stimulus	MT component cells	MT pattern cells	Unclassified cells
Response to grating			
Response to plaid			

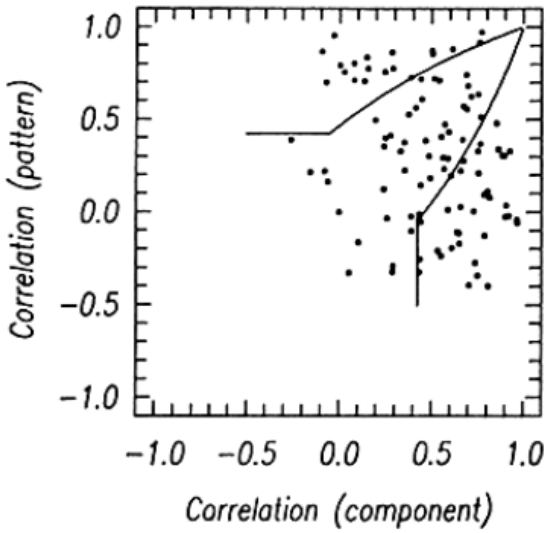
(Rodman & Albright, 1989)



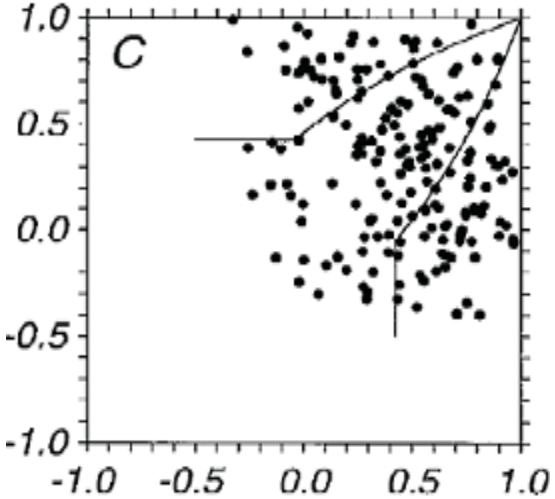
(Movshon , Adelson, Gizzi & Newsome, 1985)

Comparing pattern sensitivity of C2 units with MT cells

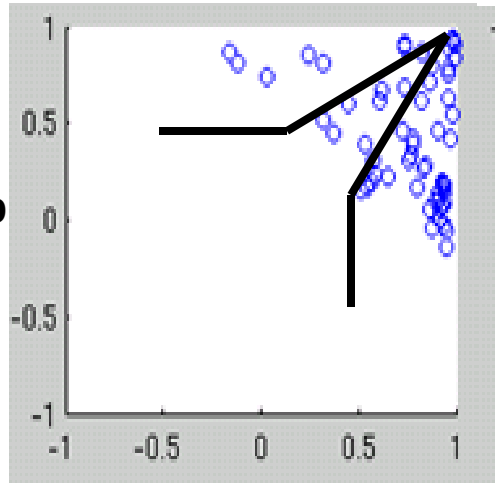
MT cells



MT cells



C2 units



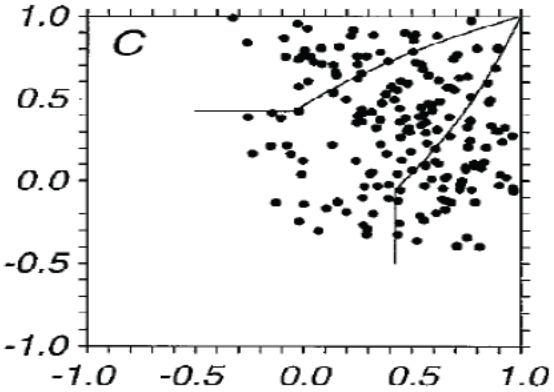
(Movshon , Adelson, Gizzi & Newsome, 1985)

(Einhauser et al, 2002)

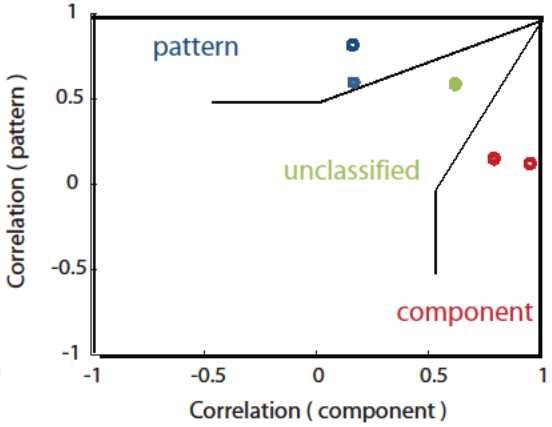
Other models that model MT cells as linear combination (plus some non-linearity) of V1 cells (Sereno, 1989; Wang 1997; Simoncelli & Heeger, 1997; Rust et al, 2006; Perrone 2002, 2004, 2006; Cadieu & Olsdausen, 2008; Tsui et al, 2010)

What makes a pattern cell and what affects the pattern sensitivity ?

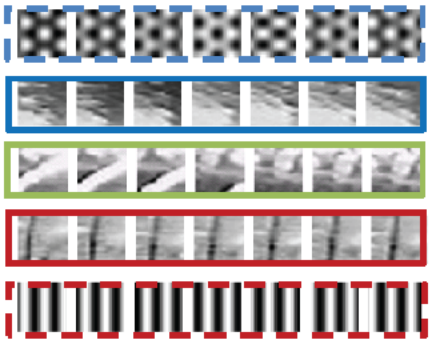
What visual stimulus makes a pattern cell:



(a)

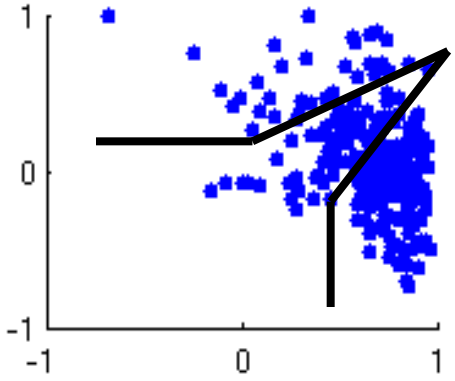


(b)

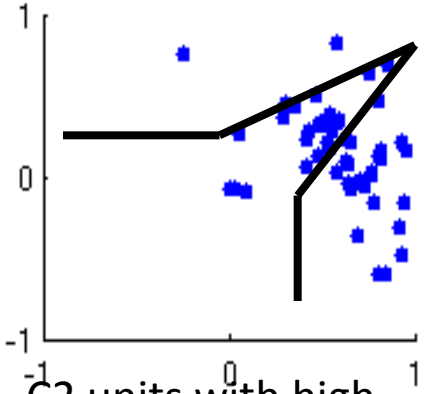


(c)

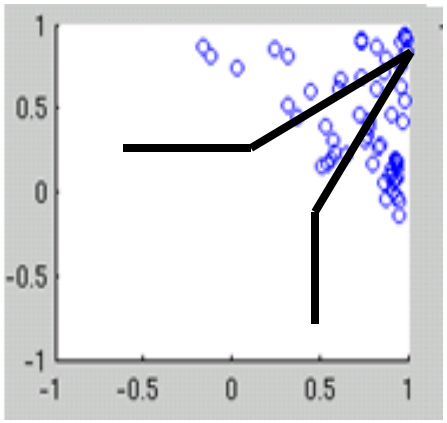
Some factors that affect Pattern sensitivity :



All the C2 units



C2 units with high Direction index

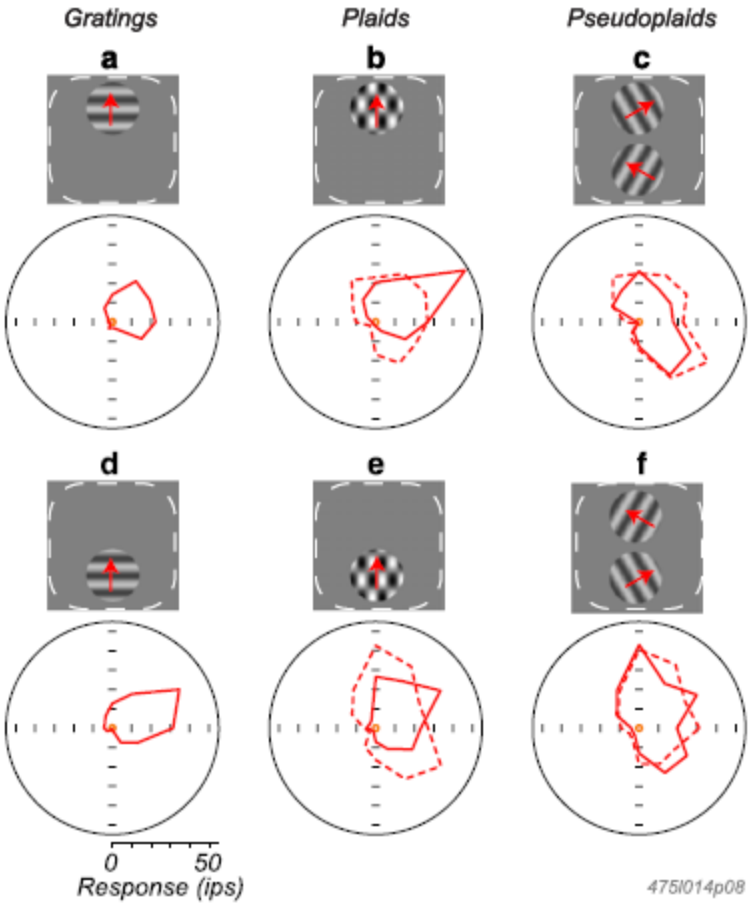


Different videos

Outline

- The model
- Experiments
  - Pattern - component MT cells (Movshon , Adelson, Gizzi & Newsome, 1985)
  - **Local direction integration of MT cells (Majaj, Carandini & Movshon, 2007)**
  - Speed tuned MT cells (Priebe , Lisberger &Movshon, 2006)

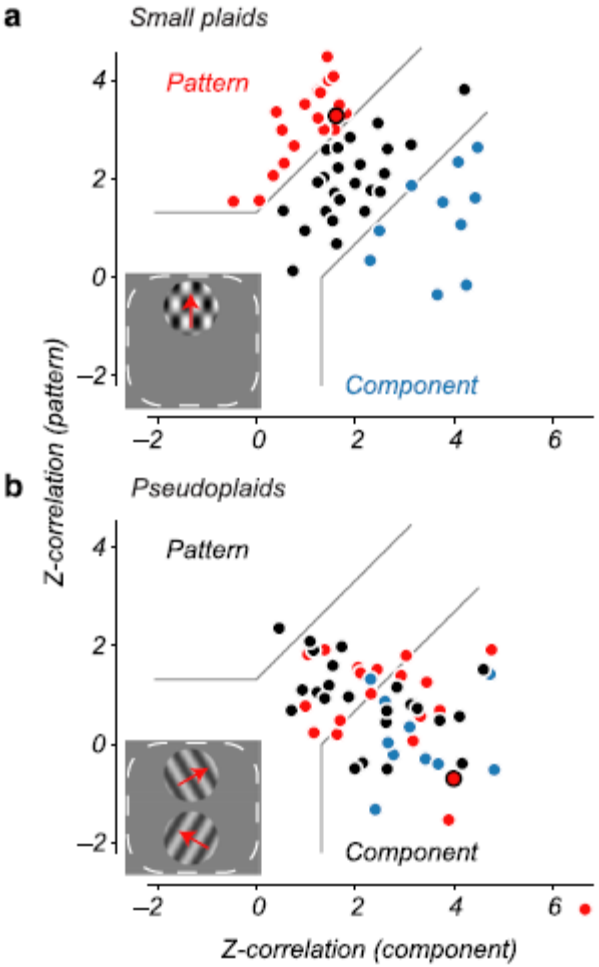
Directional integration in MT is local



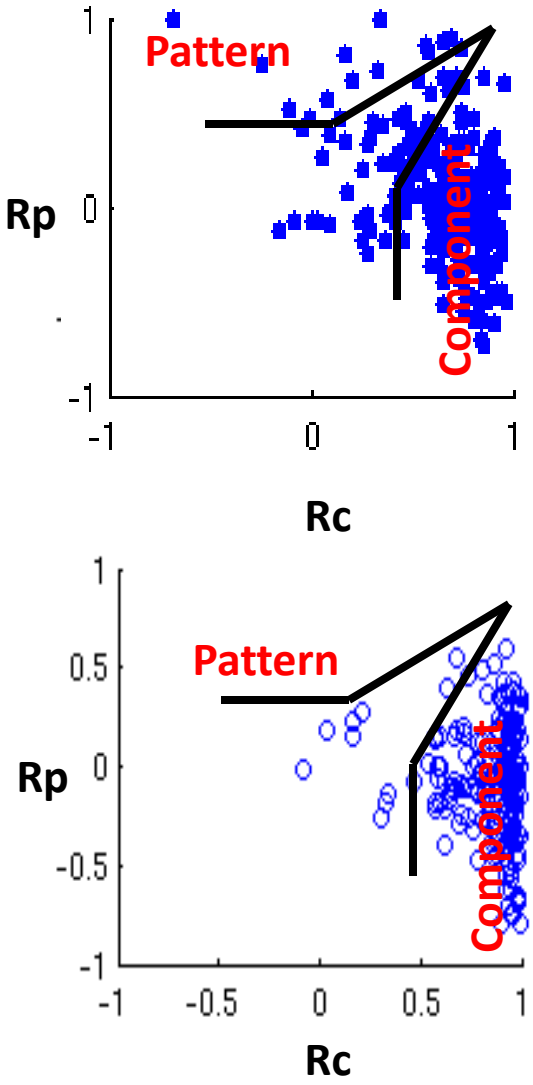
(Majaj, Carandini & Movshon, 2007)

Comparing pattern sensitivity of C2 units with MT cells

MT cells

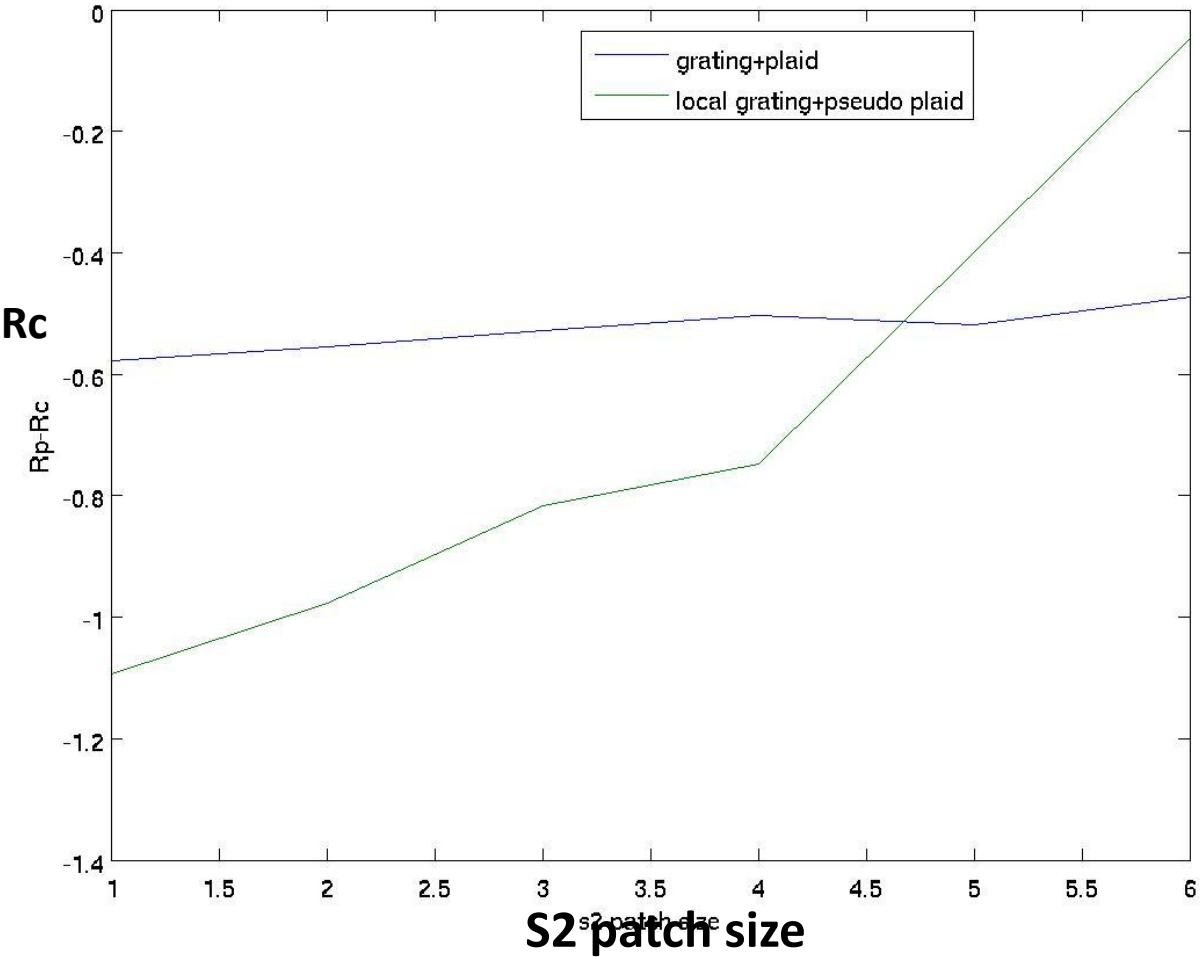


C2 units



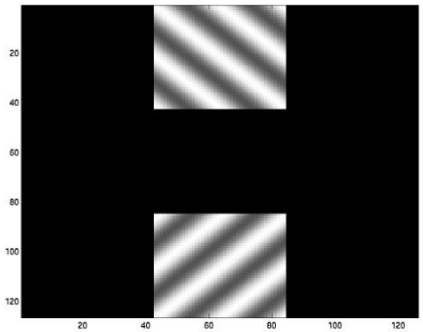
(Majaj, Carandini & Movshon, 2007)

Pattern sensitivity as a function of S2 unit size



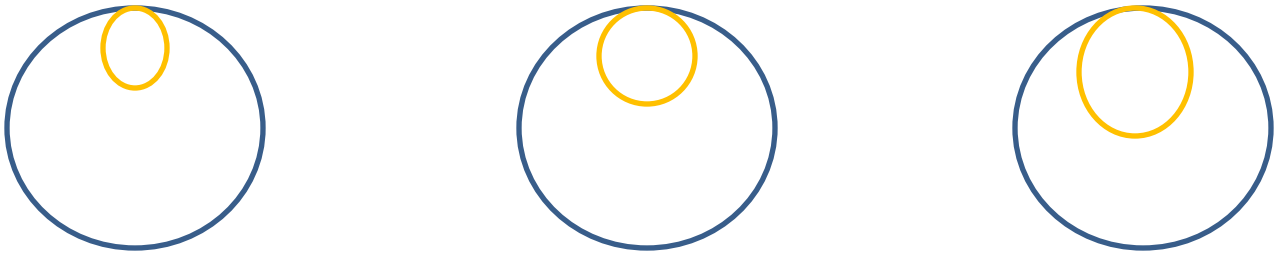
Rp-Rc

S2 patch size



C2

S2



Outline

- The model
- Experiments
  - Pattern - component MT cells (Movshon , Adelson, Gizzi & Newsome, 1985)
  - Local direction integration of MT cells (Majaj, Carandini & Movshon, 2007)
  - **Speed tuned MT cells (Priebe , Lisberger &Movshon, 2006)**

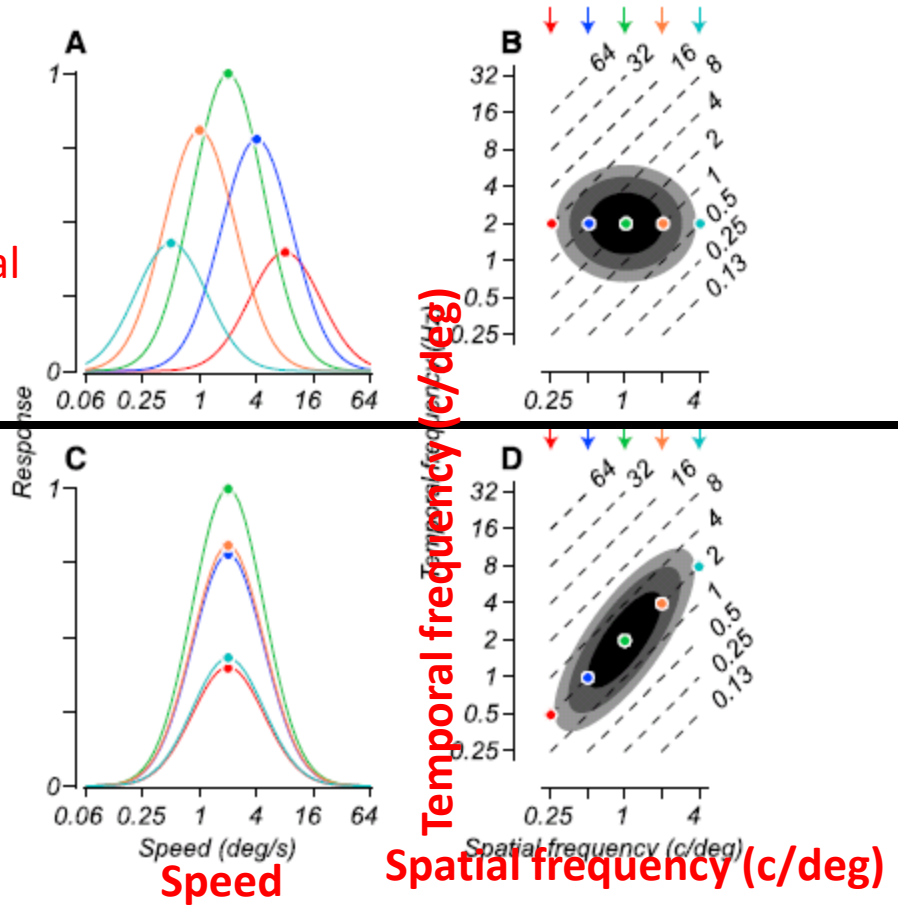
Tuning for temporal frequency vs tuning for speed

- V1 simple cells
- Some V1 complex cells
- Some MT cells

Cells whose peak speed changes with the spatial Frequency of the stimulus
(tuned to temporal frequency of the stimulus)

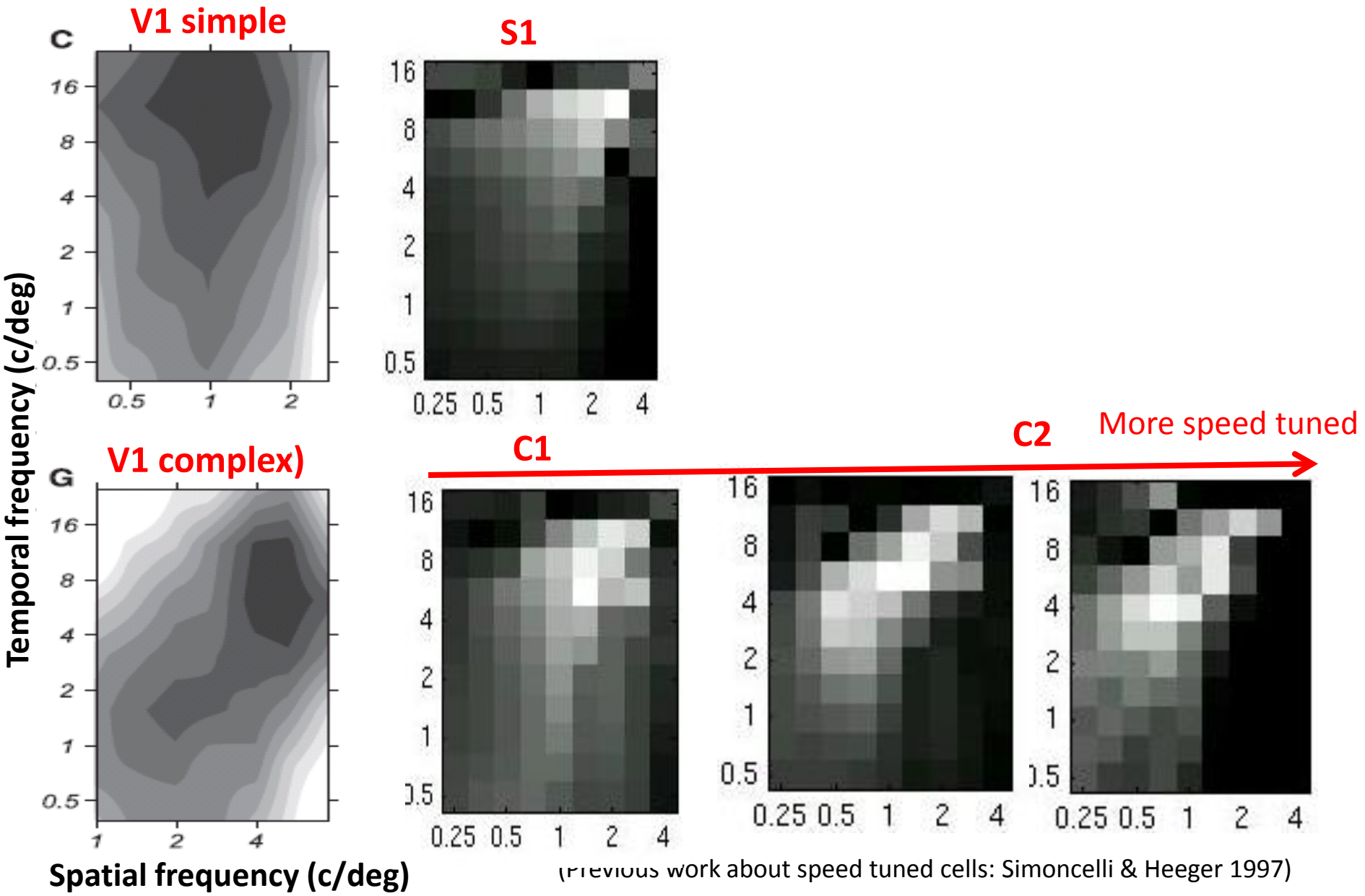
- Some V1 complex cells
-- some MT cells

Cells that tuned to speed
Independent of spatial frequencies of stimulus



(Priebe , Lisberger, &Movshon, 2006)

Comparing tuning for Spatiotemporal Frequency of S1/ C1/ C2 with V1 simple/V1 complex /MT



(Previous work about speed tuned cells: Simoncelli & Heeger 1997)

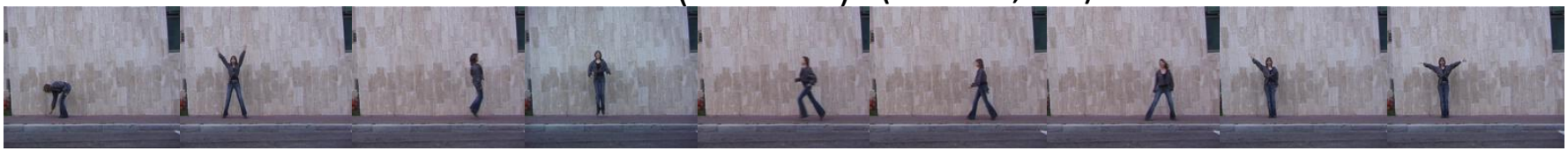
Where does this model stand ?

Ref / Previous V1-MT model	Many ref.	Many ref.	Movshon et al 85	Rodman Albright 87	Snowden et al 91	Okamoto et al 99	Priebe et al 06	Majij et al 07	
Properties	Direction tuning	RF size	Pattern cell	Speed tuning	Motion opponency	bi-modal direction tuning	Speed tuned	Local directional integration	Learning
Sereno89	Y								Y
Wang 97	Y	Y		Y					Y
Simoncelli Heeger 97	Y	Y	Y	Y	Y	Y			
Rust et al 06	Y		Y						
Perrone 02,04,06	Y	Y	Y	Y				Y	
Cadieu & Olshausen 08	Y		Y						Y
Tsui, et al 2010	Y	Y	Y						
This model	Y	Y	Y	Y	Y	Y	Y	Y	Y

More models for dynamics in the V1-MT JK Tsotsos , S. Grossberg

MT like units Our model on action recognition tasks

Weizmann Human action (9 classes) **(Blank et al, 2005)**



bend jack jump pjump run walk side wave1 wave2

KTH Human actions (6 classes) **(Schuldt et al, 2004)**

walk run jog box wave clap

	Dollar et al	Jhuang et al
KTH Human	81.3 %	**91.6 %**
UCSD Mice	75.6 %	**79.0 %**
Weiz. Human	86.7 %	**96.3 %**

(Jhuang et al, 2007)

methods	accuracy(%)
Our method	**87.60**
Jhuang et al. [8]	**91.70**
Nowozin et al. [15]	87.04
Niebles et al. [14]	81.50
Dollár et al. [4]	81.17
Schuldt et al. [17]	71.72
Ke et al. [9]	62.96

(Wang & Mori, 2008)

Name	Percentile	Ref
LTP	90.1%	
LBPabs	83.8	5.5.1
LBPdir	79.6	5.5.1
Schindler	92.7%	[15]
Laptev	91.8	[11]
Jhuang	91.7%	[8]
Niebles	81.5%	[12]
Dollár	81.2%	[4]
Schüldt	71.7%	[16]

(Yeffet & Wolf, 2009)

Large Human Action Dataset (50 classes)

	Laptev et al	Jhuang et al
HMDB	20.7 %	**23.0 %**

(in prep. 2010)

Our model is..

- Learning weights from natural stimulus (as opposed to fitting)
- Useful for real-world action recognition tasks
- Model population of cells, not only single cells
- Up to date (Majaj et al, 2007; Priebe et al, 2006)