

Spatiotemporal mechanisms for detecting and identifying image features in human vision

Peter Neri & David J. Heeger

Department of Psychology, Stanford, USA

Nature Neuroscience : August 2002.

Scope

- To explore the visual system mechanisms for selecting salient features in the environment.
- How attention is directed to selected salient features.
- And what further processing is performed that finally leads to object recognition.

- ① Is *feature detection* based on *contrast energy* ?
- ② Are the mechanisms responsible for detection also responsible for identification of the targeted features ?
- ③ Can a simple mathematical model capture/summarize the underlying early sensory mechanisms ?

INTRODUCTION

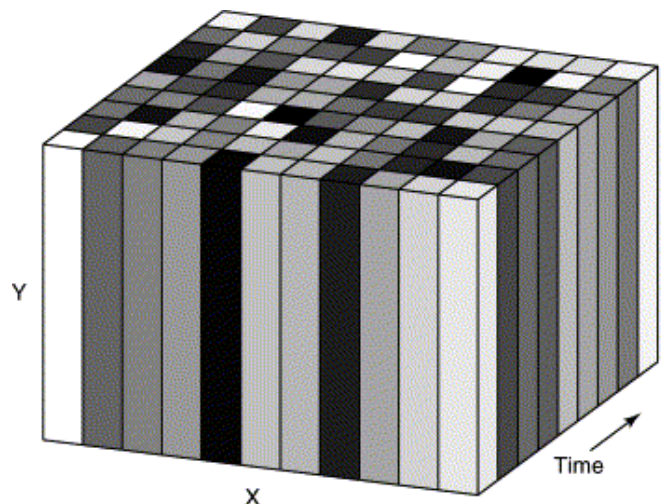
Traditional attempts are based on *psychophysical experiments*, in which the *detectability* or *identifiability* of some fixed targets is measured as a function of target brightness (or some other target attributes).

Randomizing the presentation of stimuli is a classic strategy employed to avoid unwanted *history-dependence* in the responses.

Stochastic stimuli paradigms

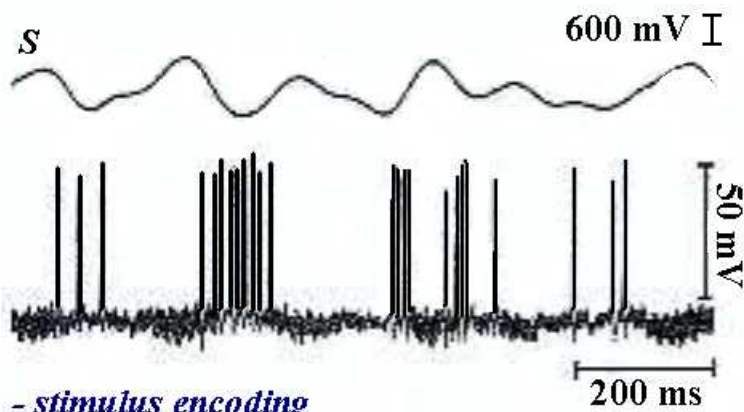
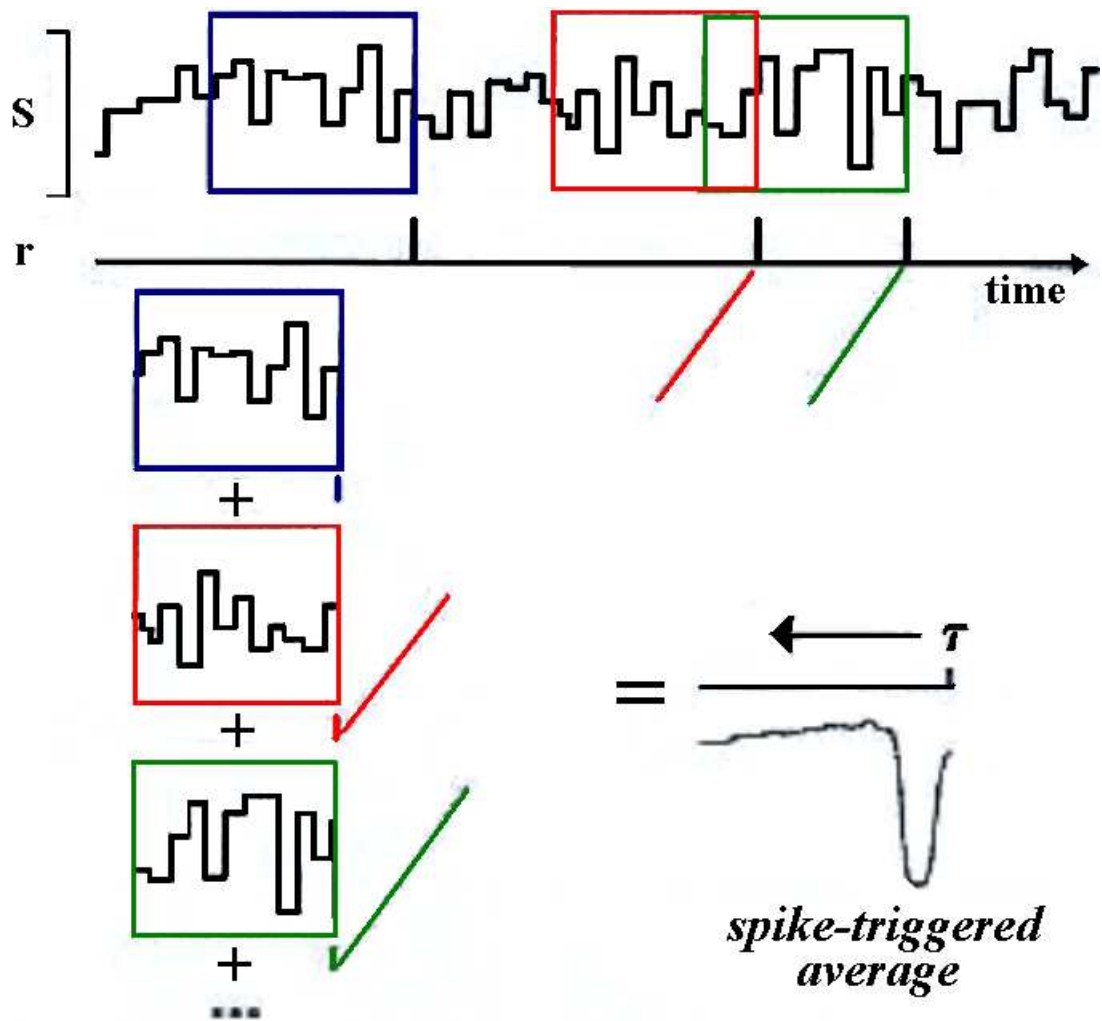
- (i) Stimuli are drawn randomly from an ensemble and presented in rapid succession.
- (ii) The stimuli are labelled based on subject's response (e.g. 'yes/no') and grouped accordingly.
- (iii) The properties of these 'response-triggered' groups are analyzed.
- (iv) The stimuli are stored enabling "simulated experiments" for the test/validation of different mathematical models

Each stimulus was a movie:
9 frames, 11 bars/frame,
the intensity of each vertical bar
was a random number.
The target was placed at the bar in the
center of the stimulus block.

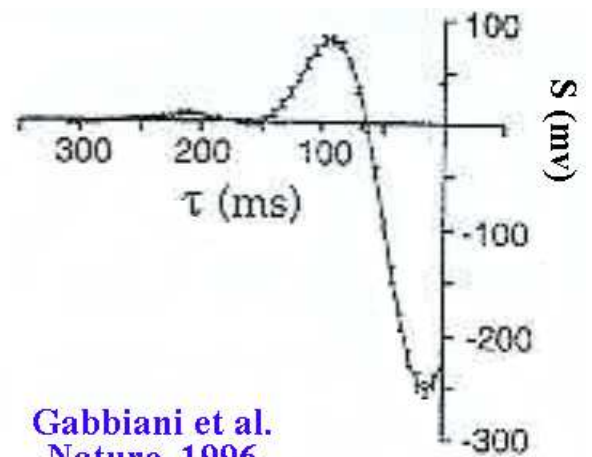


The data-analysis was based on *Noise-Image classification*, a psychophysical variant of *noise-based reverse correlation*

REVERSE CORRELATION

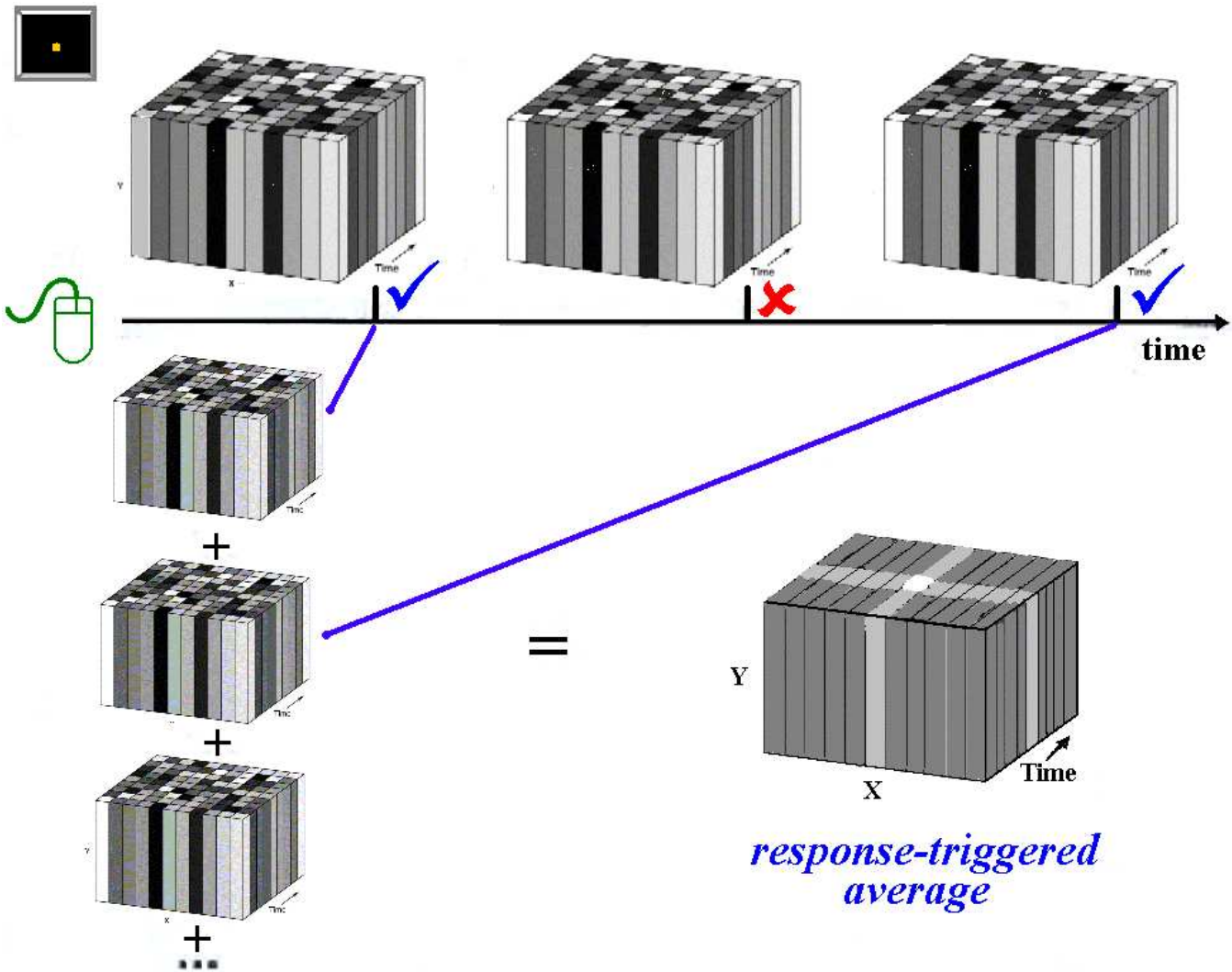
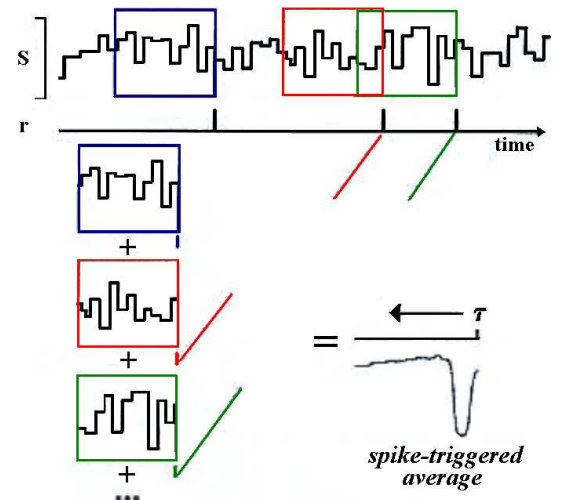


- stimulus encoding
in weakly electric fish



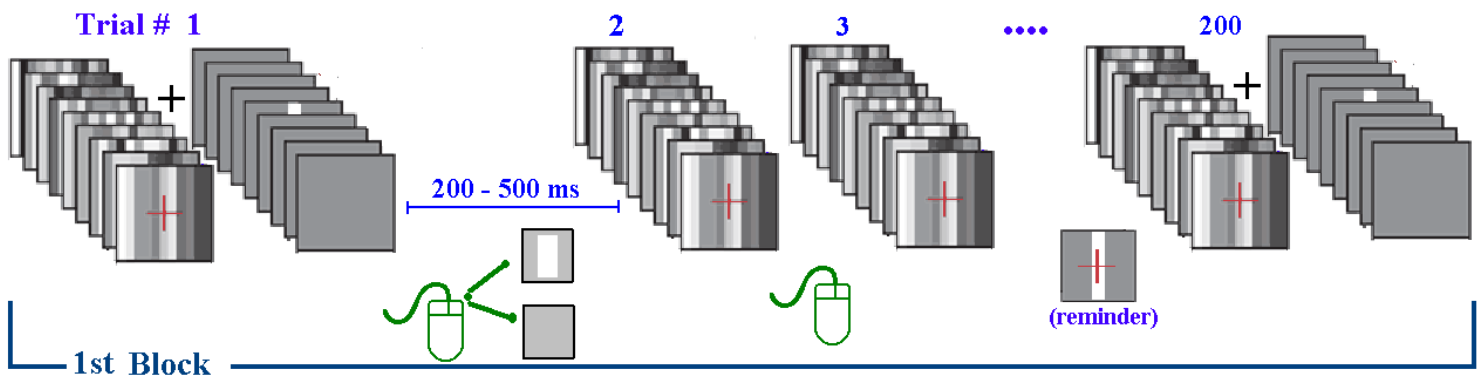
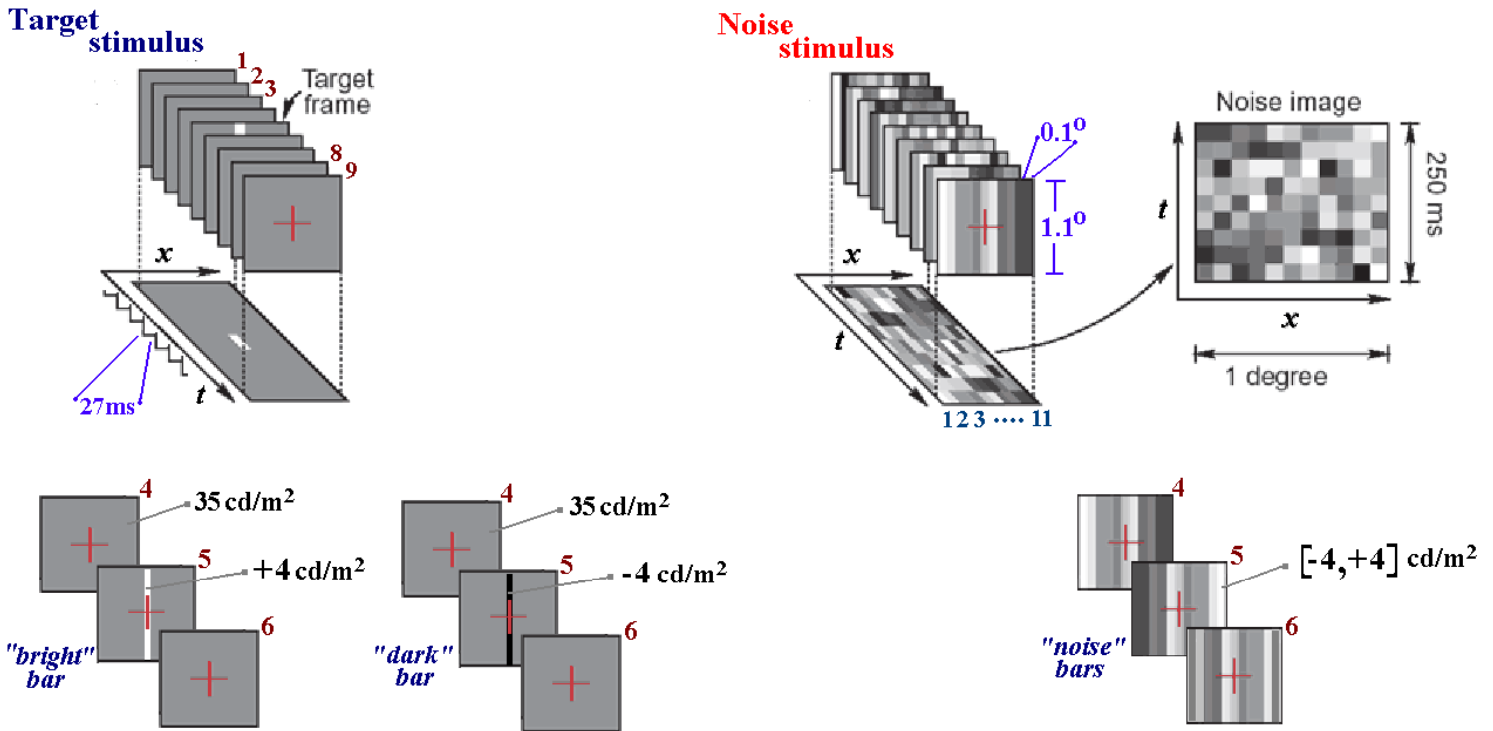
Gabbiani et al.
Nature, 1996

NOISE-IMAGE CLASSIFICATION



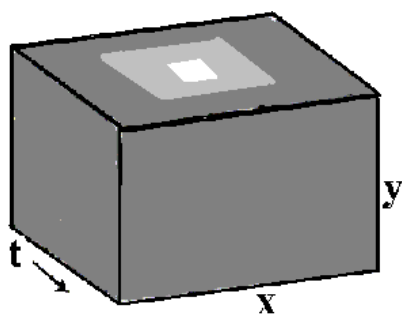
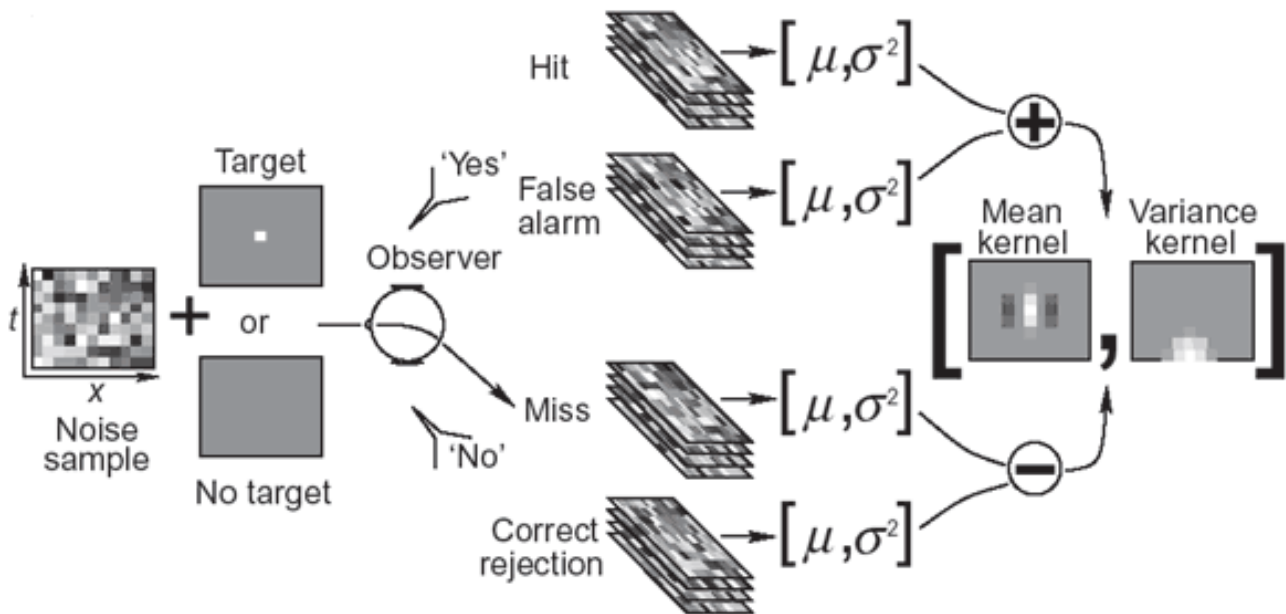
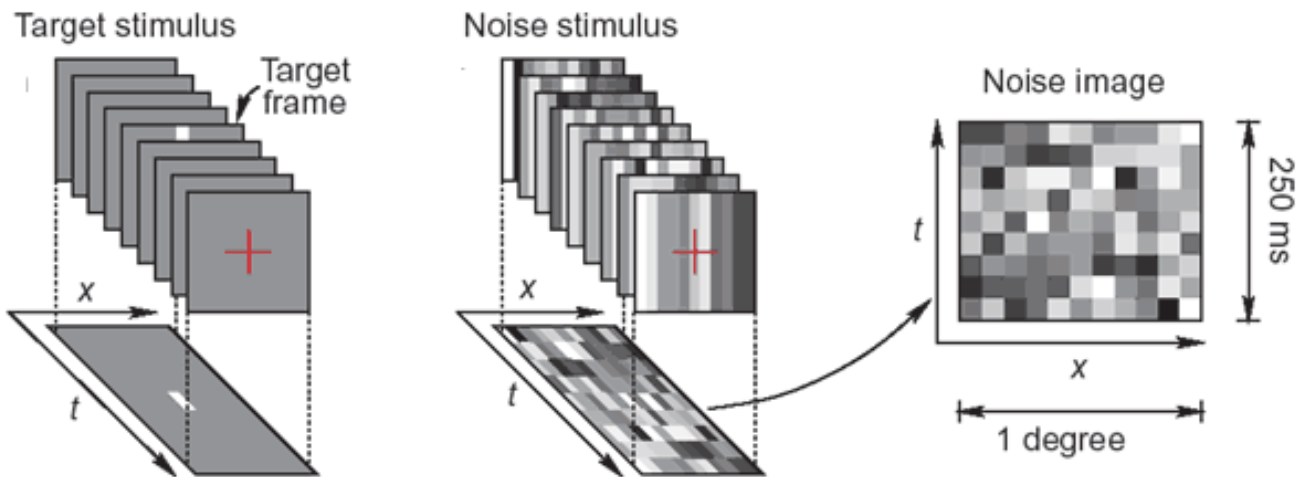
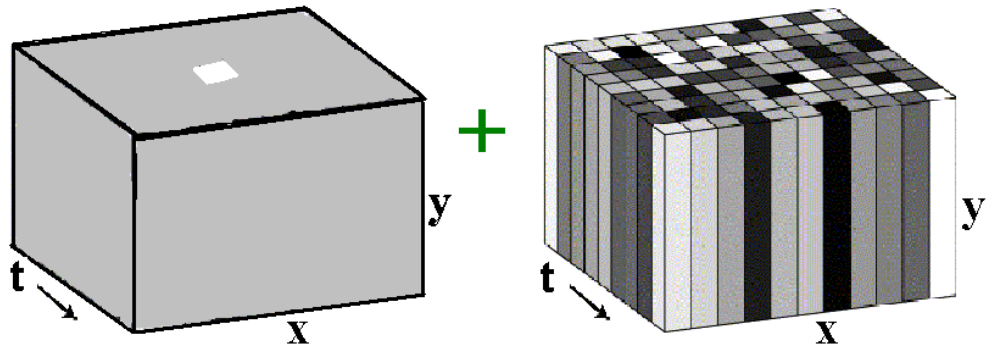
EXPERIMENT I

A forced-choice decision task.
 Subjects had to report the presence or absence
 of the target bar (for a given bar polarity)

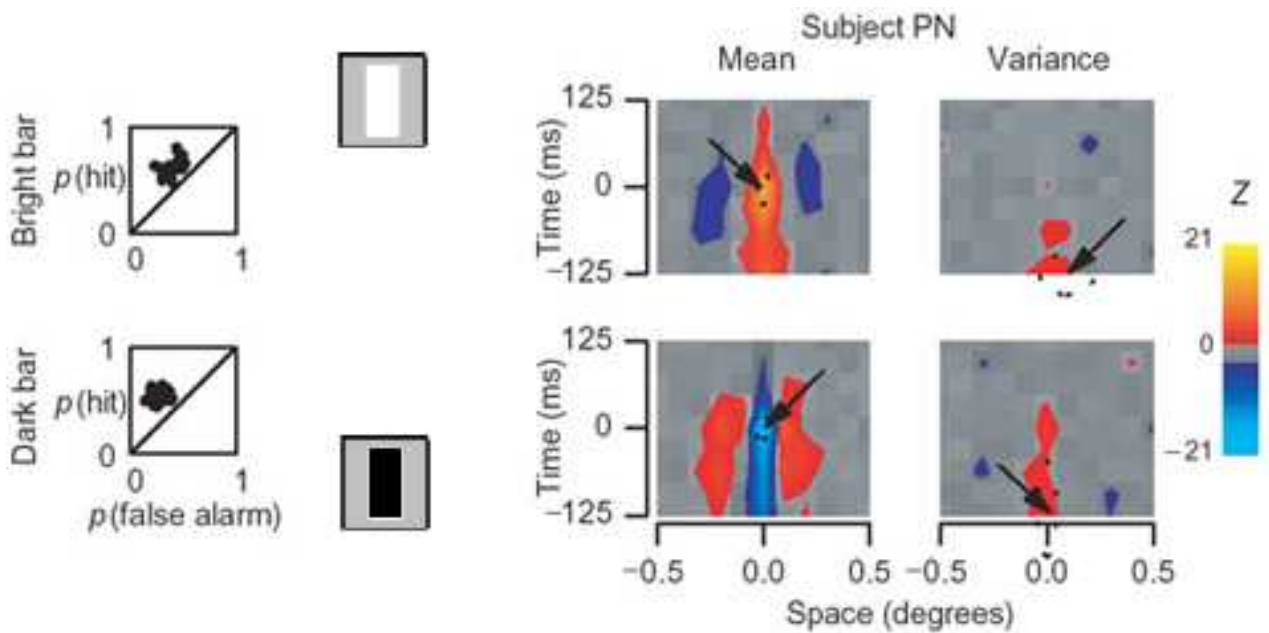


- For each polarity, 5000 trials per subj. were applied in blocks of 200.
- A "reminder-frame" every 25 trials.
- No feedback was given.

Analyzing the “noise cubes” from all the stimuli (movies) based on subject’s perception



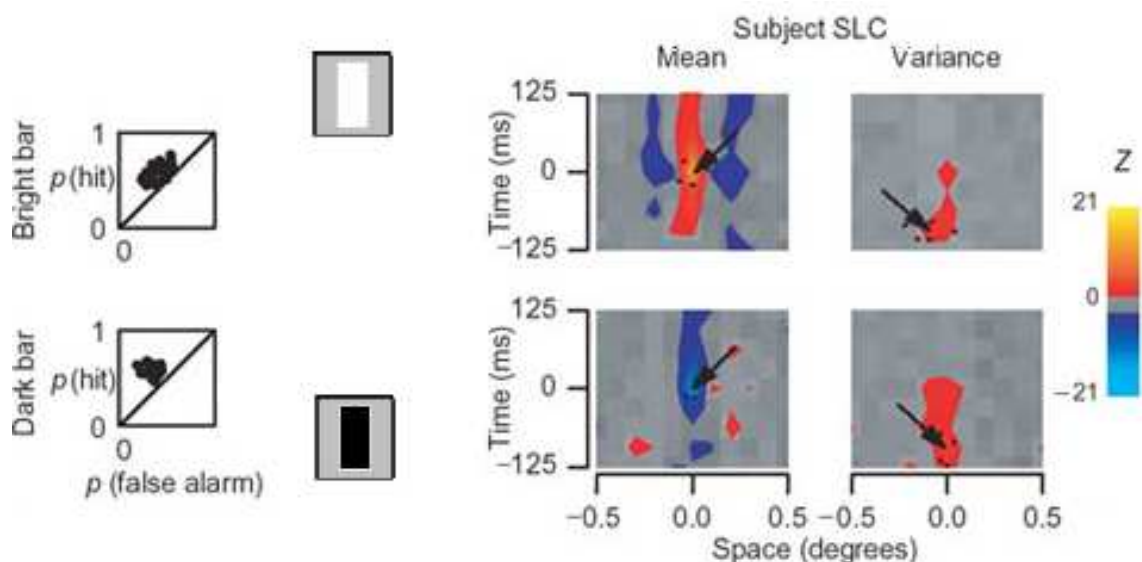
A spatiotemporal “receptive-field” is constructed the deviation of which from being random (unstructured) is quantified via bootstrapping



The mean kernels have a **positive**(**negative**) peak at the spatiotemporal location of the **bright**(**dark**) target and **negative**(**positive**) flanks across the x dim.

‘yes’ responses were more frequent than ‘no’ when there was excess of noise intensity at the location of the target.

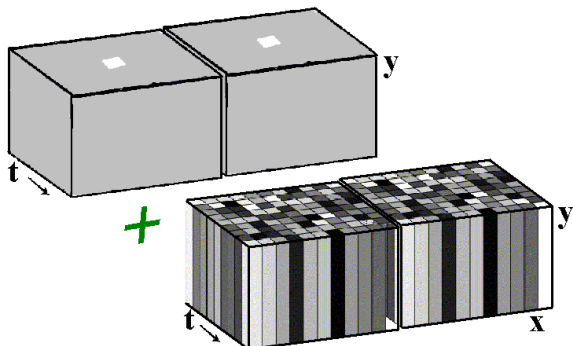
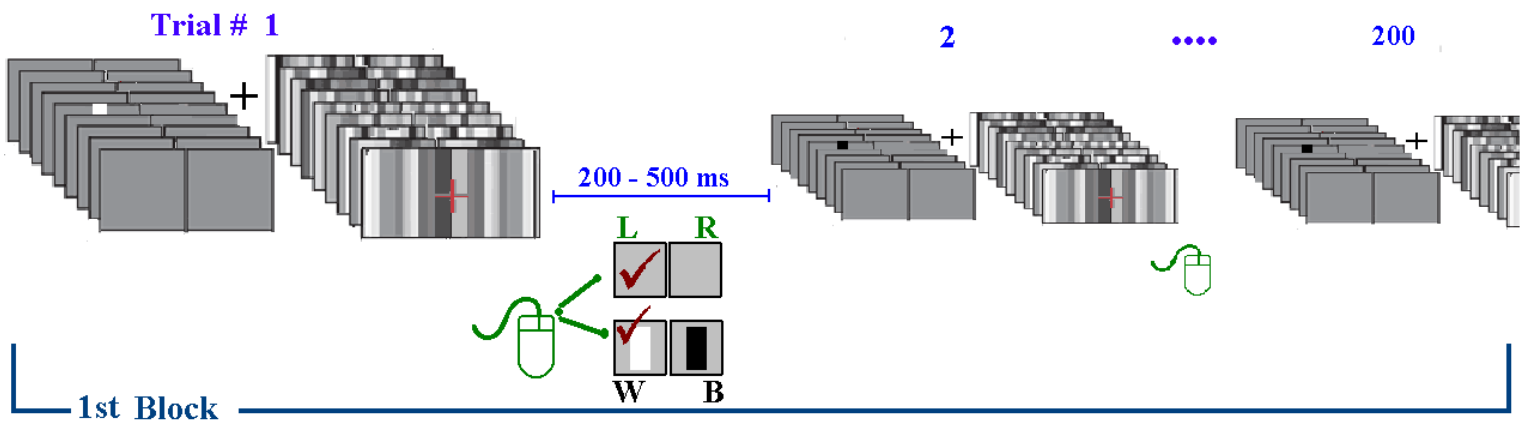
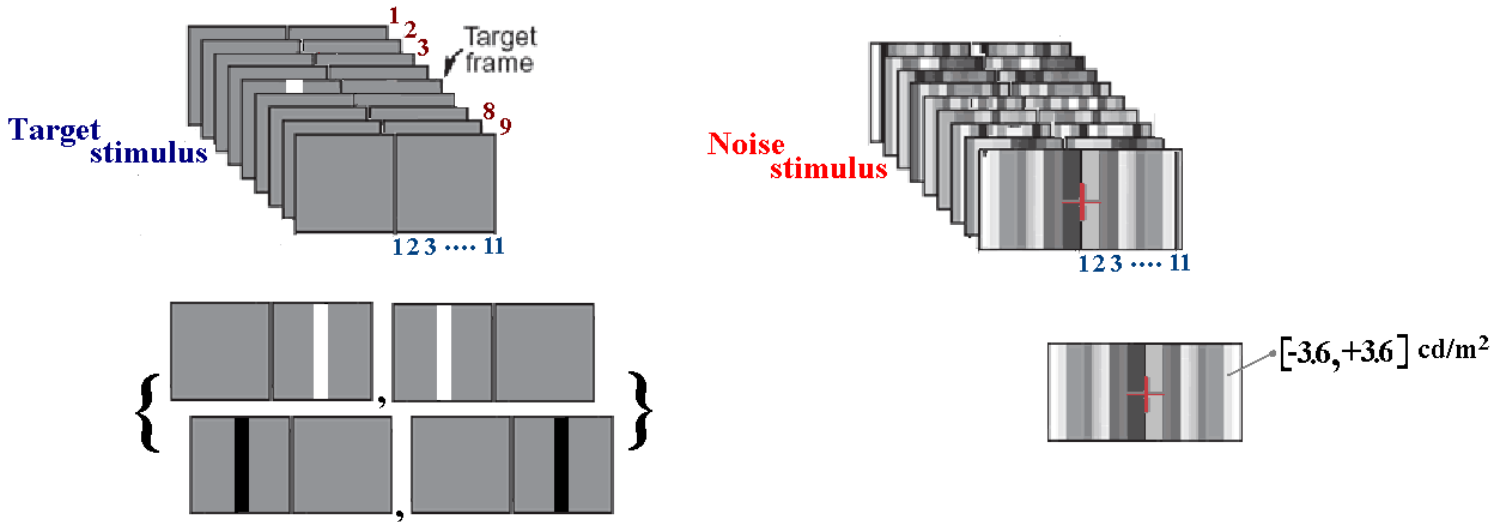
The variance kernels have a **positive** peak confined to a period **~100 ms** preceding the (bright & dark) target



➡ The temporal offset between peaks in the **mean** and **variance** kernels suggests that the two kernels probably reflect different mechanisms : **Detection** (early/contrast) & **Identification** (target polarity)

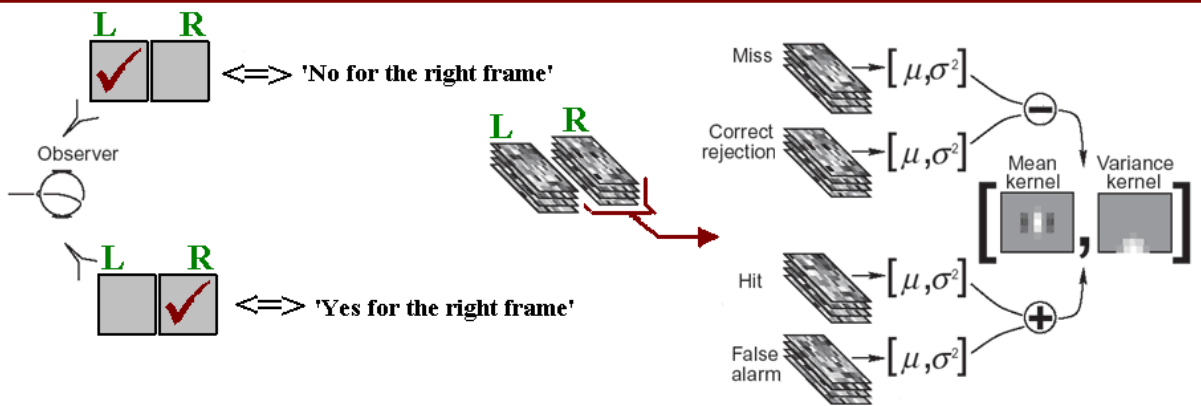
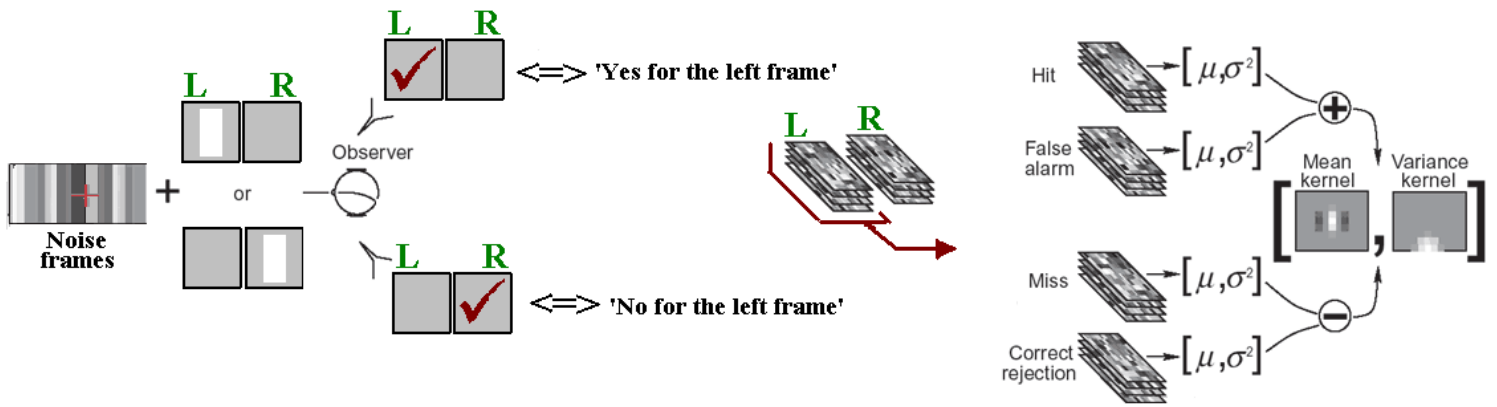
EXPERIMENT II

The target bar appeared bright or dark,
 and either to the left or to the right of fixation.
 For each trial, subjects had to report
 both the location and the polarity of the target.

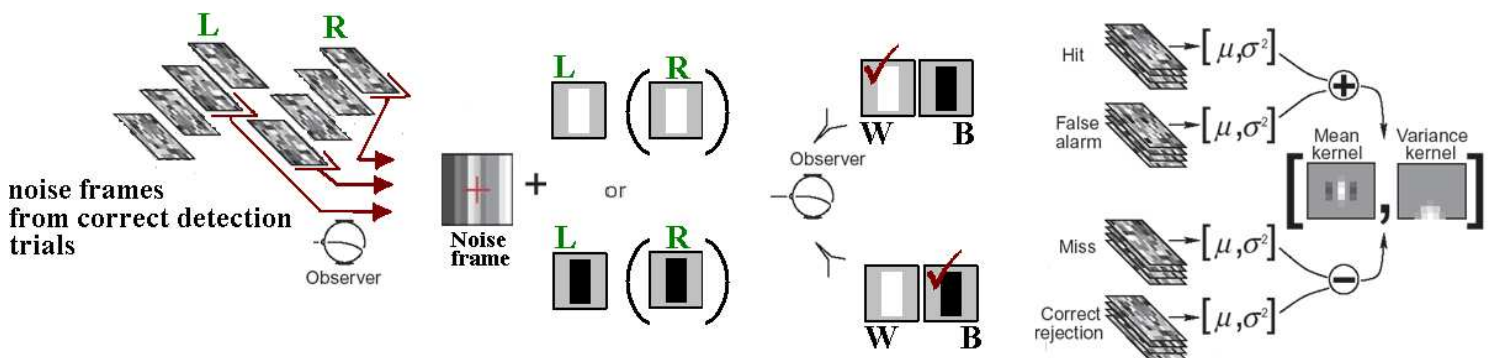


By analyzing
 the pairs of “noise cubes”
 from all the 2-frame-width stimuli (movies),
 based on subject’s decision
 for the **location/identity** of the target,
 a *mean kernel* (and a *variance kernel*)
 was constructed corresponding only to the
detection/identification task

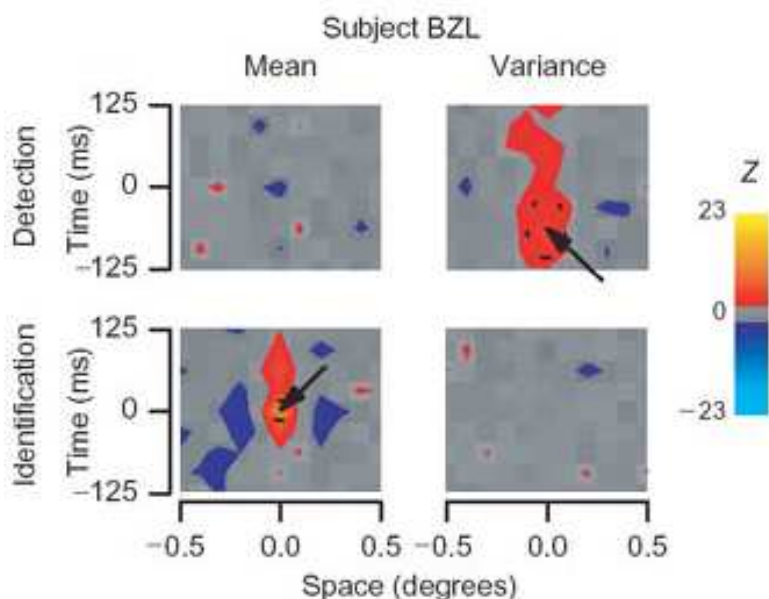
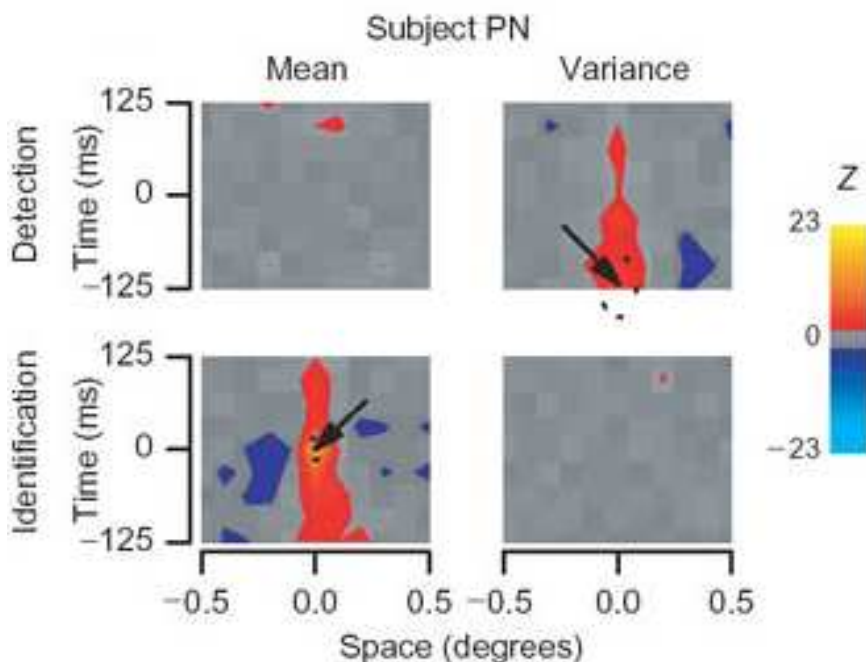
Detection task :



Identification task :



Detection generated pronounced modulations in the *variance-kernel*,
but not in the *mean-kernel*.
Identification the opposite.



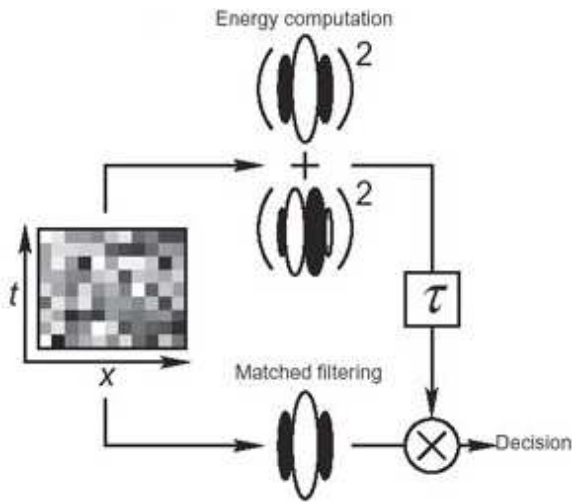
Detection and **Identification** task can be disassociated
both in terms of timing and strategy.

Detection judgements were not differentially dependent on polarity,
and influenced by the variance early in the stimulus.

Identification judgements were differentially dependent on polarity.

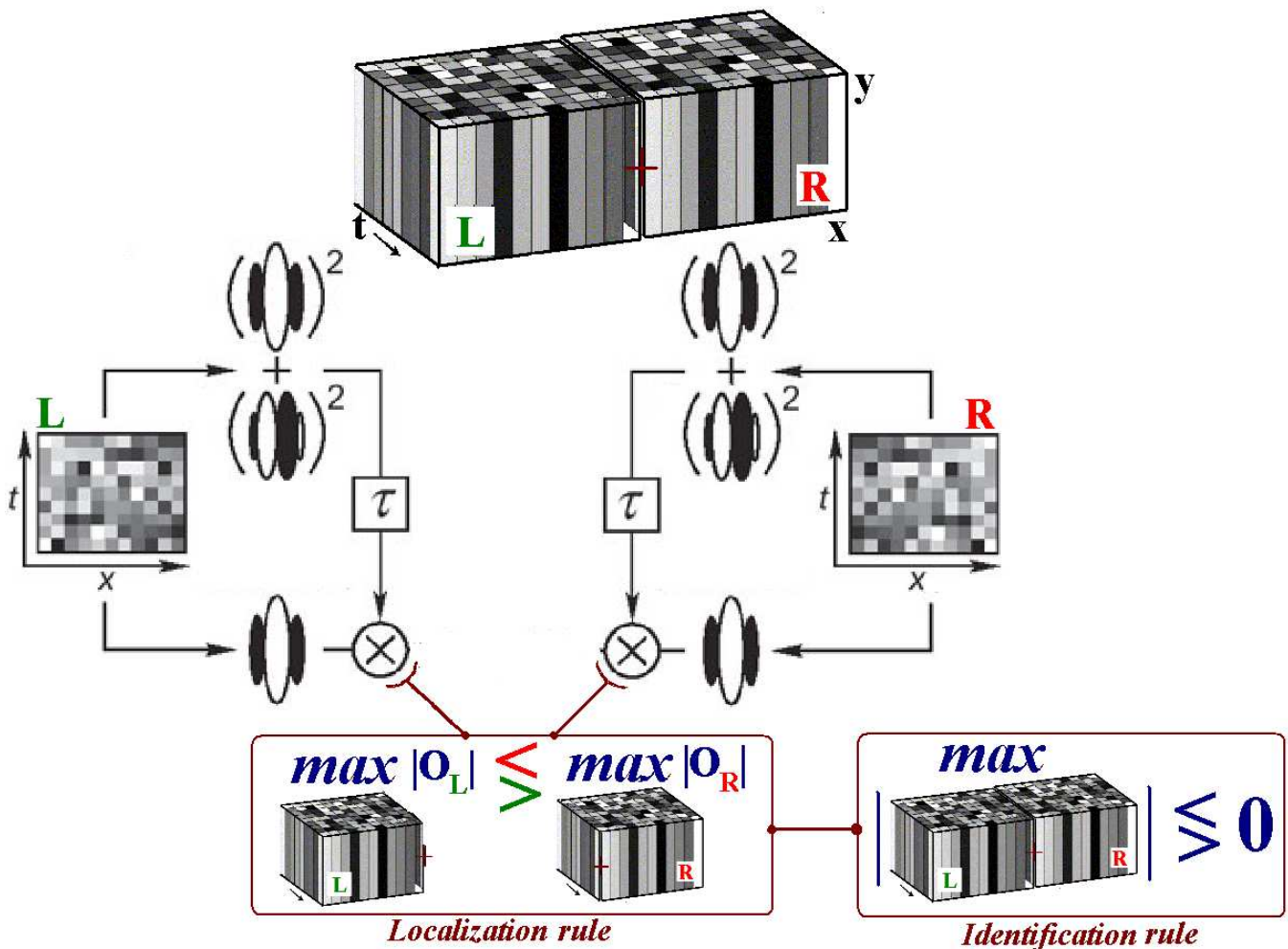
MODELING

Detection and localization depend on extracting contrast energy from the stimulus, while identification depends on matched filtering.

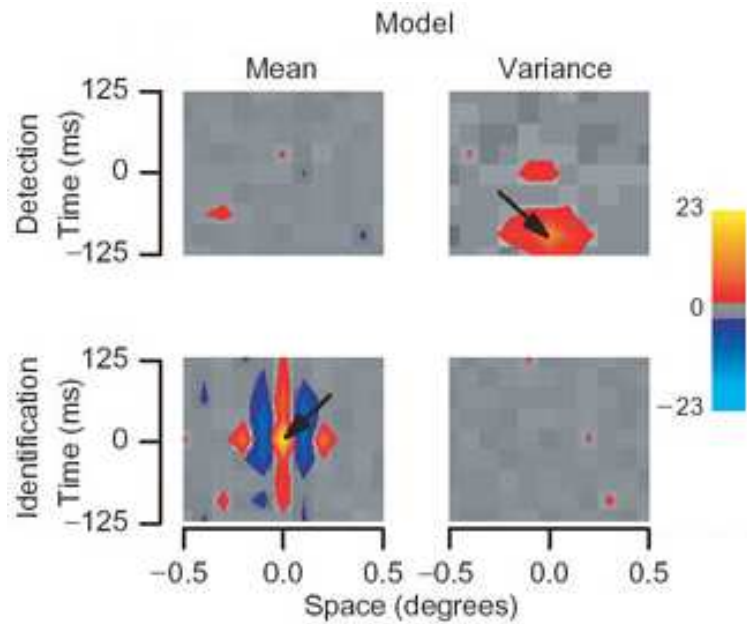


The responses of the matched filters are modulated, after a delay (100 ms), by the extracted contrast energy so that detection leads to identification.

Energy responses: *quadrature pair* of linear filters (*odd & even*)

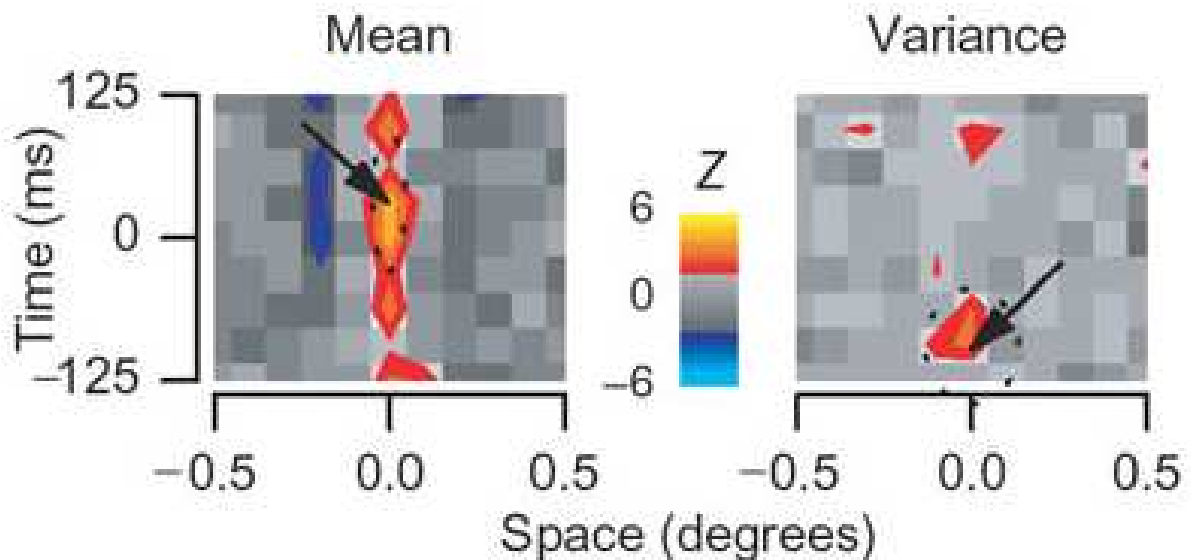


After “running” the experimental procedure
using the model on the same (stored) stimuli



FINAL JUSTIFICATION

The trials corresponding to False Alarms from the 1st experiment, were analyzed separately:



The result provided a further justification for their interpretation and ruled out alternative possibilities.

CONCLUSIONS

- ◆ The ‘*mean kernel*’ shows a *center-surround* type organization
 - ◆ The “*bar-receptors channel*” (Hubel & Wiesel, 1962).
- ◆ The ‘*variance kernel*’ spreads over the earliest frames
 - ◆ separate ‘*attentional*’ mechanisms triggered by early high-contrast signals.
- ◆ A non-optimal strategy is followed by the subjects
 - ◆ the experiment probed built-in mechanisms, probably useful in many real-life contexts
- ◆ “*automatic, exogenous, bottom-up attentional capture by high contrast cues*”

DISCUSSION

- ① Stochastic Stimuli
- ② Modeling