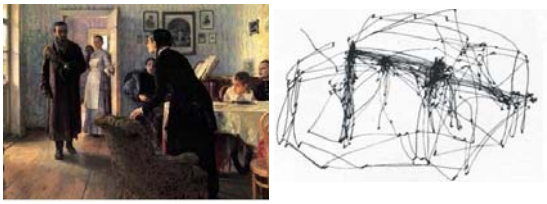
Active Vision (in Humans)

Christoph Rasche
Justus Liebig Universität
Germany

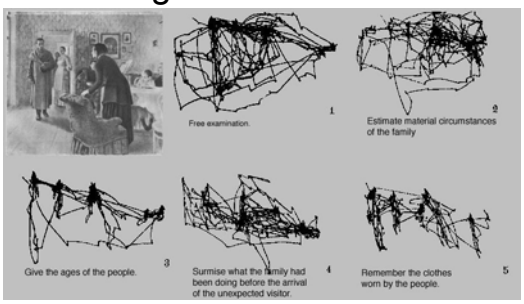
Seeing

Yarbus 67



- Saccades (rapid shifts in gaze position): 3-4 times/second
- Attentional shifts: several/between saccades: fastest: 33ms
- Also referred to as **overt** and **covert** attentional shifts

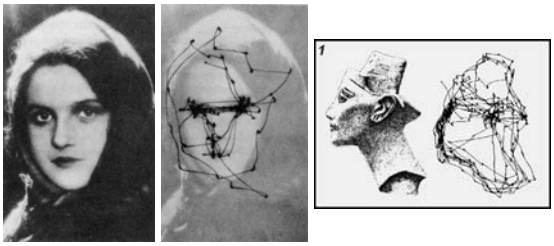
Cognitive Influence



→ endless amount of information ('a picture is worth a 1000 words')
→ scene representation allows precise saccades (top-down)

More Famous Examples

Yarbus 67

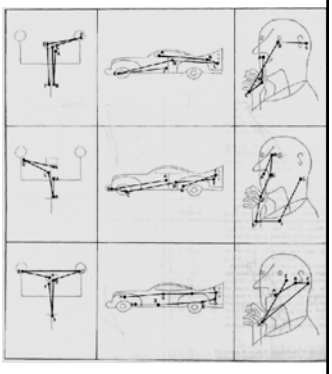


Scan Path

Noton & Stark 71

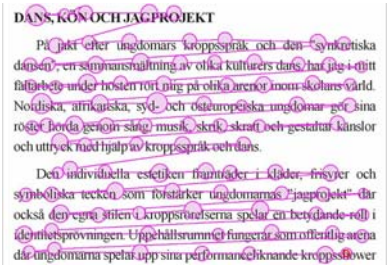
Idea: viewer uses a memorized sequence of fixations to obtain object/scene understanding

Today: limitedly accepted.



Nevertheless: clusters of fixations!
→ no explanations so far

Reading


DANS, KÖN OCH JAGPROJEKT

På jakt efter ungdomars kroppsspråk och den synkretiska dansen, en sammanställning av olika kulturers dans har jag i mitt fältarbete under hösten rört mig på olika arenor inom skolans värld. Nordiska, afrikanska, syd- och östeuropeiska ungdomar ger sina röster borda genom sång, musik, skrik, skratt och gestaltar känslor och uttryck med hjälp av kroppsspråk och dans.

Den individuella estetiken framträder i kläder, frisyrer och symboliska tecken som förstärker ungdomarnas "jagprojekt" där också den egna stilen i kroppsrörelserna spelar en betydande roll i identitetsprövningen. Uppehållsrummet fungerar som offentlig arena där ungdomarna spelar upp sina performancelknande kroppsflower

Forward saccade: ca. 7-9 letters (adjusted to fontsize!)
Reverse saccades
Landing point: avg. 3rd, 4th letter (→ precise saccadic target selection)

Eye Trackers

- modern ones based on video cameras
- calibration required
- typical accuracy: 0.5 – 2.0 degrees
- maximum accuracy: ~0.3 deg (head fixated)

Other Types of Eye Movements

- **Saccades** (so far)
- **Smooth pursuit**: following a moving object

- for stabilization during self motion:
  VOR: vestibulo-ocular reflex (sense of balance)
  OKR: optokinetic reflex (flow field)

→ can combine to *Nystagmus:*
saw-tooth pattern (sitting in train and watching the landscape)

A Fixation Zoomed

Fixational eye-movements:

Microsaccades Drift Tremor

Scientific debate: purposely or accidental?

Saccadic Flight Limitedly Ballistic

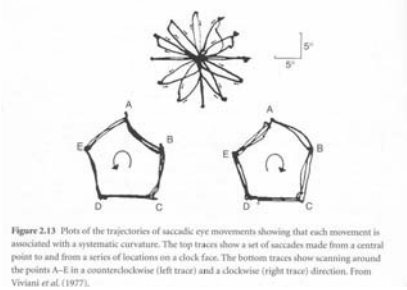
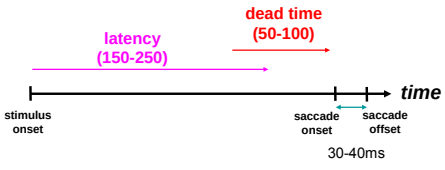


Figure 2.13 Plots of the trajectories of saccadic eye movements showing that each movement is associated with a systematic curvature. The top traces show a set of saccades made from a central point to and from a series of locations on a clock face. The bottom traces show scanning around the points A–E in a counterclockwise (left trace) and a clockwise (right trace) direction. From Viviani *et al.* (1977).

saccadic flight a trajectory. In rare cases: flight rerouting!
→ system is highly dynamic

Timing

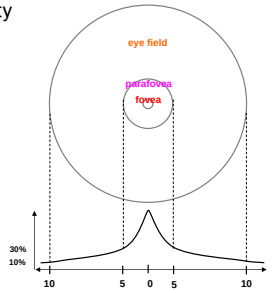


- latency and dead time overlap: while this saccadic target is being computed, the subsequent saccade target is already programmed too!
  → blazing flexibility

- Saccades required for survival? Not necessarily.

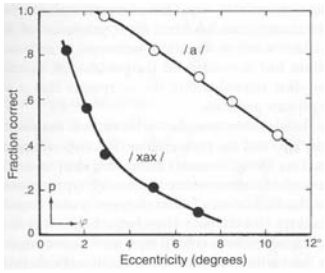
Peripheral Acuity Ranges

	degs	acuity
• Fovea	2	
• Parafovea	10	30%
• Eye field	40	10%
• Head field	180	



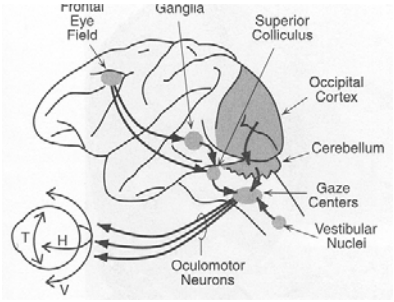
- Saccadic amplitudes [degs]:
  - max: ca. 30
  - object analyzing: ca. 5-10
  - scene orienting: ca. 10-20

Recognition in Periphery



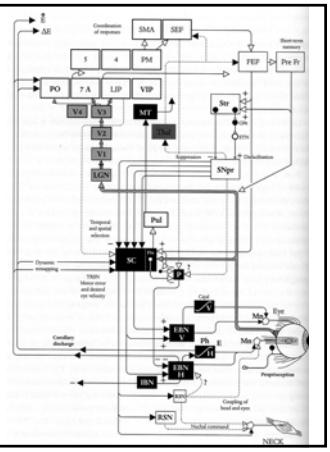
→ isolated objects recognized better
(short words are skipped during reading)

Brain Areas Involved



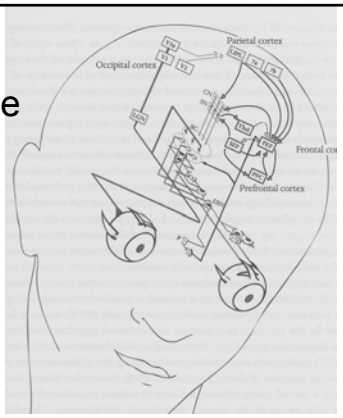
→ all over!

The more Elaborate Version



The Illustrative Version

Berthoz



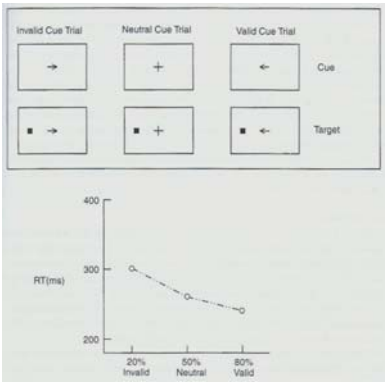
Attentional Shifts

[covert shifts]

- Preattentive [= distributed attention]
  - fast, automatic, massively parallel
- Postattentive [= attentive,focused attn]
  - slow, serial
- Definition of attention is difficult
  your's is as good as anyone else's
- Metaphors
  - Spot-light
  - Zoom lens
  - Object versus region based

Cueing

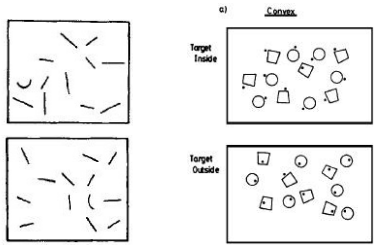
Posner



→ Spotlight

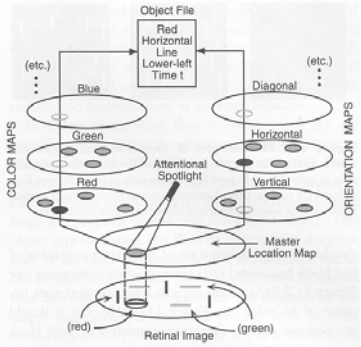
Visual Pop-Out

Treisman



→ parallel/serial distinction
Criticism: many pop-outs possible.

Treisman's Feature Integration Theory

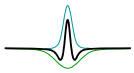


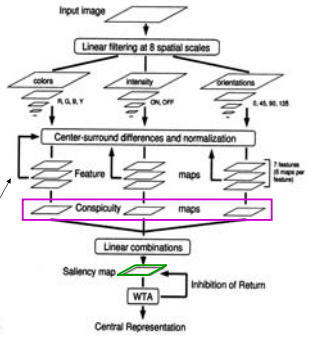
Saliency Map

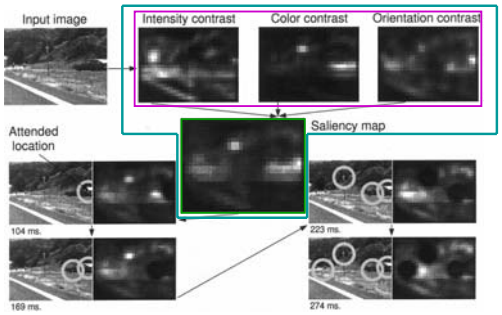
Itti & Koch

Goal: to mimic a bottom-up driven visual search of attentional and saccadic shifts

Band-Pass Filter: DOG (difference of Gaussians)



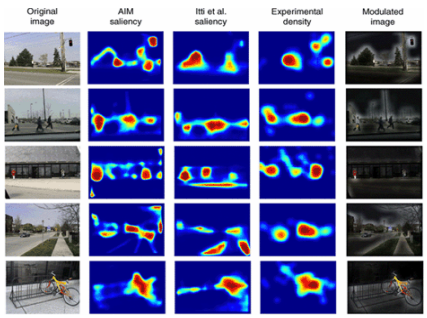




- popular in psychology/neuroscience due to attempt to connect to visual search studies and neurophysiology
- limitedly used in computer science due to time-intensive DOG filtering

Information Maximization

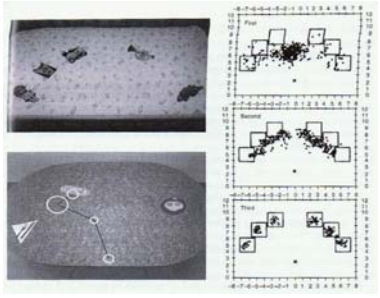
Bruce & Tsotsos



maybe just that?

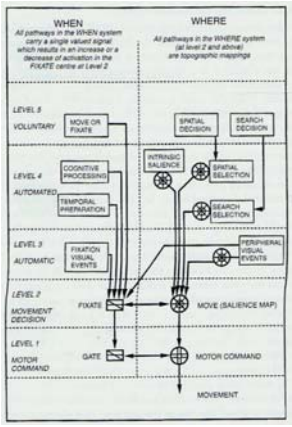
Zoom Lens Attention

Zelinsky

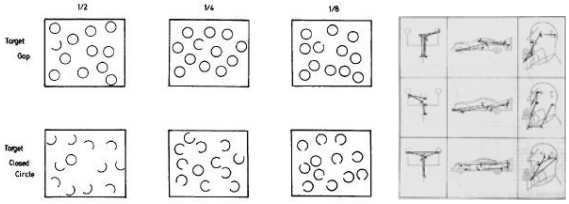


A Framework For Saccade Generation

Findlay & Gilchrist



Still Unexplained



some still try: e.g. Fourier Transform

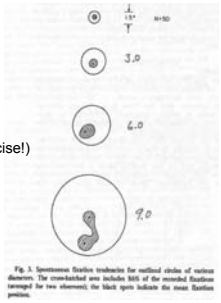
Saccadic Target Selection

Richards and Kaufmann

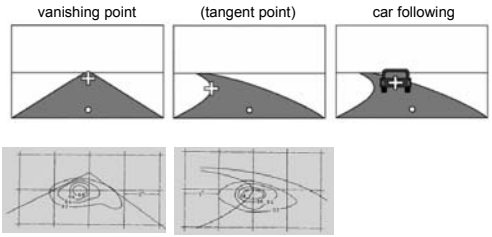
3 observers:

→ preferred landing positions (relatively precise!)

→ also unexplained



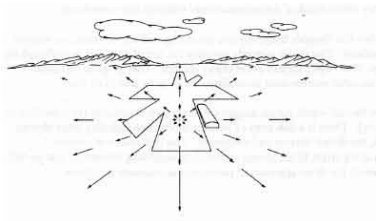
Gaze During Driving



When reacting: fixation duration increases.
But why? To exploit attentional shifts?

Optic Flow

Gibson 50



→ *Focus of Expansion* [FoE] can serve as a direct pointer of where to steer to (may be so in the air)

The Challenge on Ground

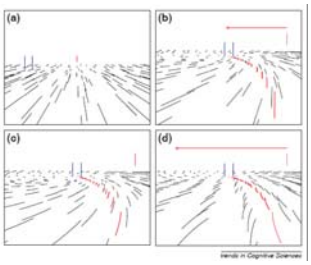
optic flow

\+
eye-/head movements

→ potential masking by retinal flow
a) how is OF recovered (ER)
b) OF used at all?

Retinal Flow Patterns…

Wann & Land



→ could be even useful/necessary for steering
- debate: both flows are used in some way
- criticism: too noisy during driving

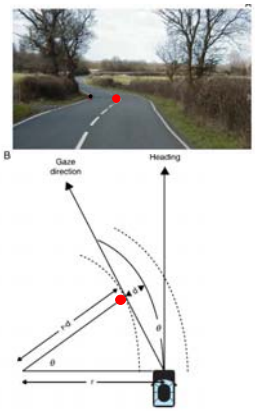
Another Potential Steering Mechanism

Donges 78, Land&Horwood 95

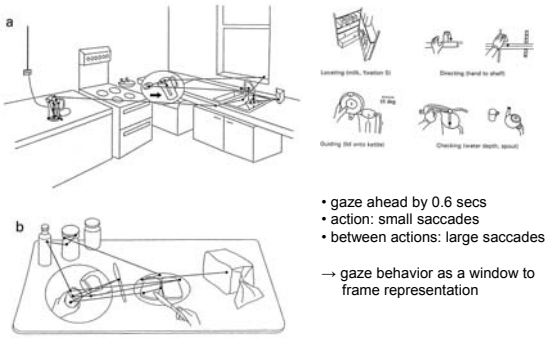
- estimation of curvature by observing tangent point (apex of the bend)

criticism: humans can not estimate curvature so accurately

visual system may use multiple cues



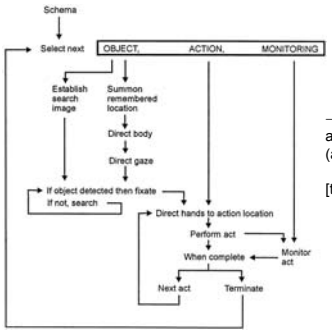
Gaze During Daily Activities Land



- gaze ahead by 0.6 secs
- action: small saccades
- between actions: large saccades

→ gaze behavior as a window to frame representation

Action Control



→ sequential (almost continuous) access of pieces of representation (a stream?)

[that's why AI failed]

Diagnostic

Duchowski

- *Neuroscience*
  - Attentional Neuroscience, Brain Imaging
- *Psychology*
  - Reading, Scene Perception (Perception of Art/Film), Visual Search, Natural Tasks, Auditory Language Processing, Speech Production
- *Industrial Engineering, Human Factors*
  - Aviation, Driving, Visual Inspection (food, X-rays)
- *Marketing/Advertising*
  - Copy Testing, Print Advertising

Human-Computer Interaction

[exploiting gaze]

- The study of interaction between people (users) and computers
- = MMI (man-machine interaction), = CHI
- (Human Factors: emphasizes human behavior)
- (Ergonomics: emphasizes comfort of use)

Gaze-Controllable Interface?

- Traditional input:
  keyboard, mouse → awkward at times
- Toward *advanced user interfaces*
  → include speech, gaze
- Dream: selection by gaze (since mid 80's)
  (requires an eye-tracker)
  - speed (gaze is always first)
  - comfort (reducing repetitive strain inj.)

However

Jacob 90

- How to 'emulate' a click?
  → blinks, dwell time: Midas Touch Problem
- Gaze (landing) is inaccurate:
  → difficult to select small icons quickly:
  Space Problem

→ not suitable as a pure output modality
(thus, no wide-spread use so far)

Eye-Typing for Disabled People

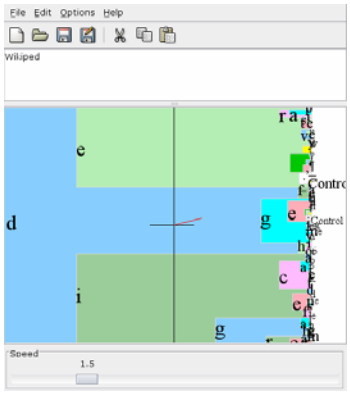
But why difficult? See experiments before…???

Dasher

[see wikipedia]

letters flying in from the right side

does not require high precision



Gestures

Drewes, Schmidt, 2007

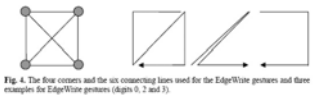
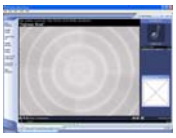


Fig. 4. The four corners and the six connecting lines used for the EdgeWrite gestures and three examples for EdgeWrite gestures (digits 0, 2 and 3).



- they are relative – no absolute position necessary, therefore:
  → no high accuracy of tracker required
  → no need for calibration
  → no Midas-Touch problem

  → still slow

Gaze-Assistive Systems

- Gaze not used to steer but just recorded in order to make display adjustments

MAGIC Pointing

Zhai, Morimoto, Ihde 99

- Principal: To place mouse cursor near gaze position. Cursor movement to be completed by hand movement.
- Goal: To save on duration for hand movement

liberal conservative

- Manual and Gaze-Input Cascaded

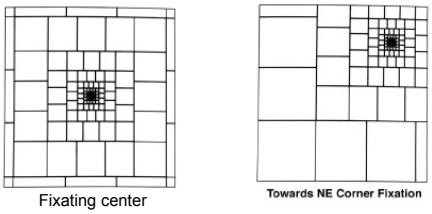
Large Screens

Problem: refresh slow

→ gaze-contingent refresh:

- *high-resolution* refresh at the center of fixation,
- *low-resolution* refresh in the periphery

Gaze-Contingent Displays



→ research: is it comfortable?

Recommended Reading

- **Active Vision: The Psychology of Looking and Seeing**
  John M Findlay and Iain D Gilchrist
- **Vision Science: Photons to Phenomenology**
  Stephen E Palmer