Image Retrieval Using Eye Movements

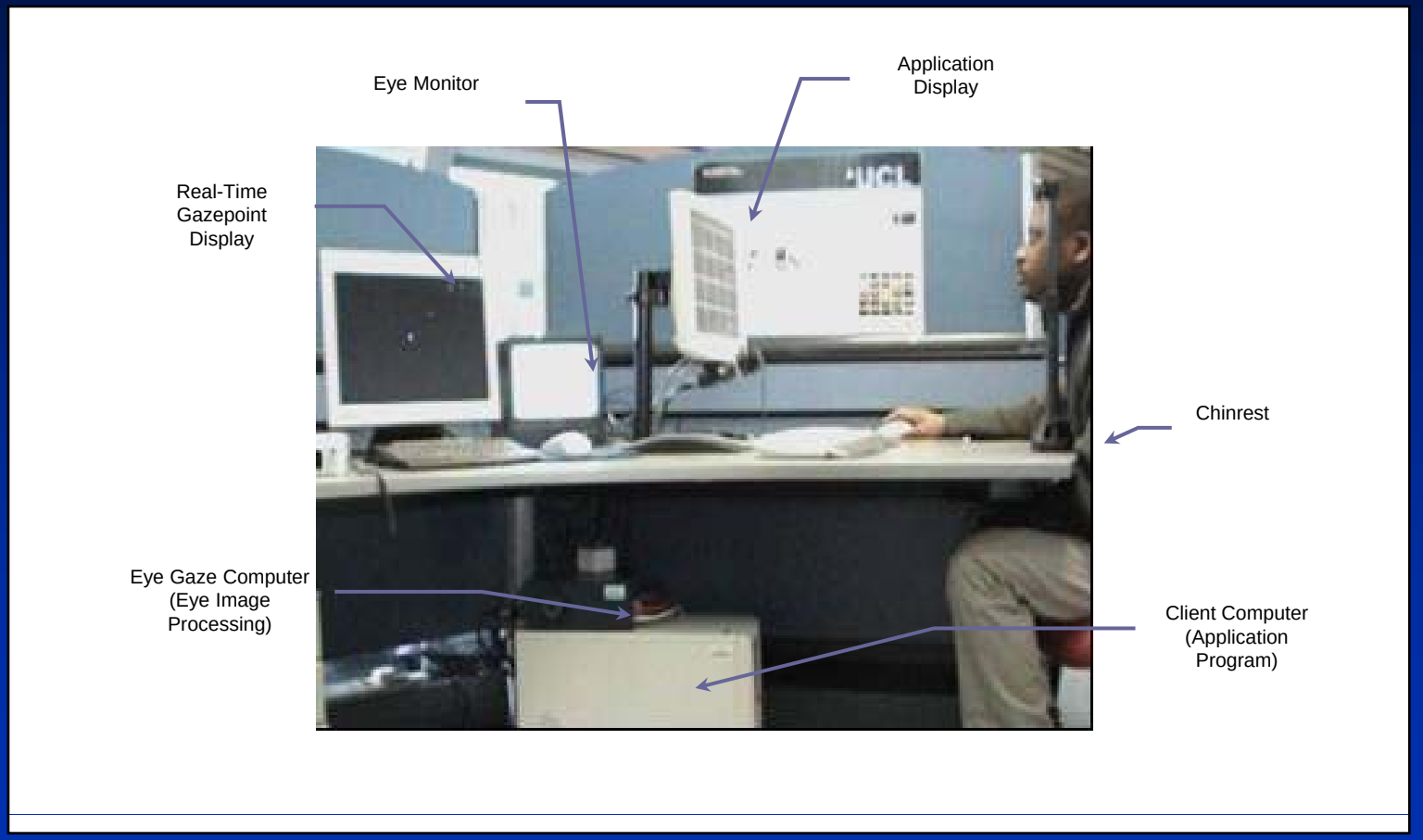
Fred Stentiford & Wole Oyekoya

University College London

Outline

1. Eye Movement Behaviour
2. Image Identification
3. Image Search
4. Conclusions & Future Work

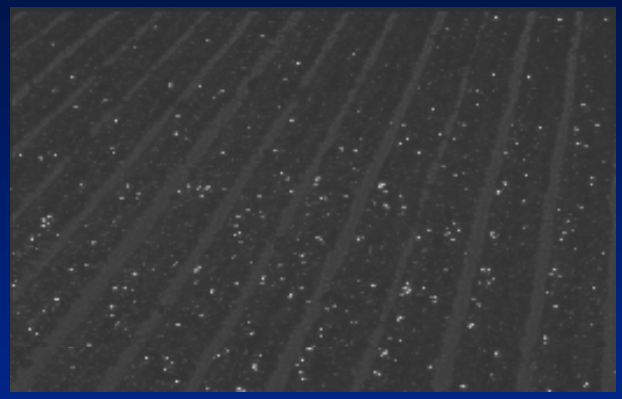
Eye Tracking System



Eye Movement Behaviour

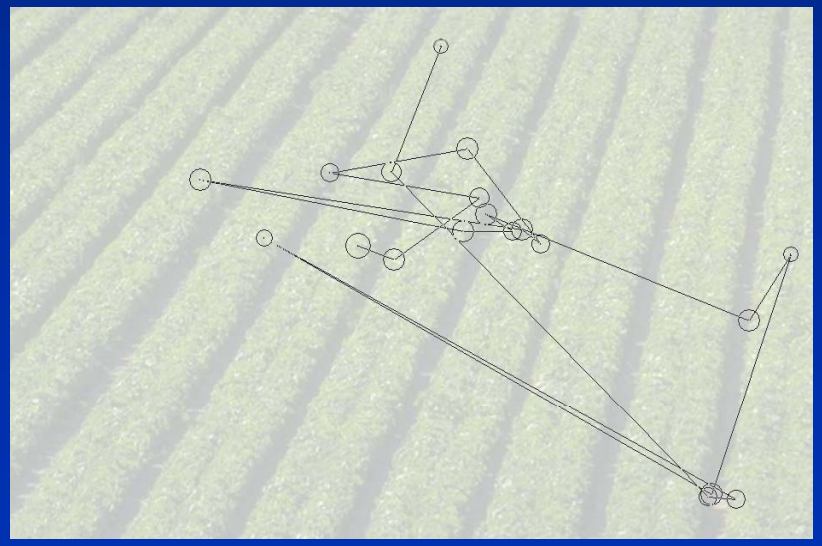
image

saliency map



no
ROI

fixation and saccade map



Eye Movement Behaviour

image

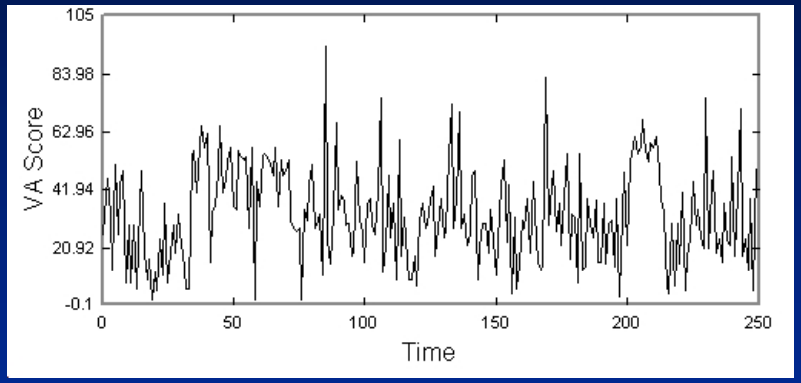
saliency map

clear
ROI

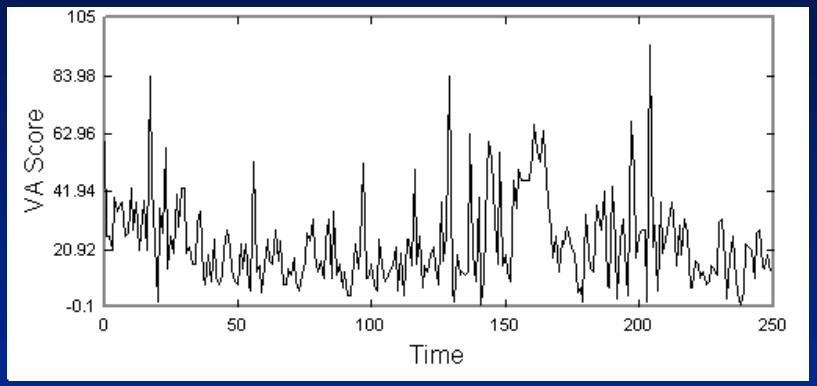
fixation and saccade map

Eye Movement Behaviour – no ROI

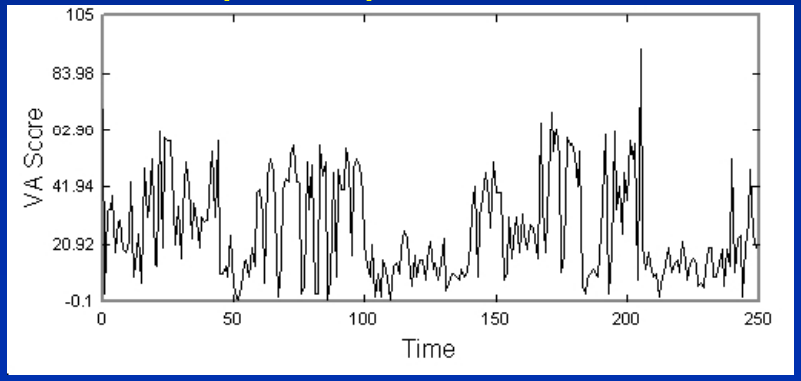
participant A



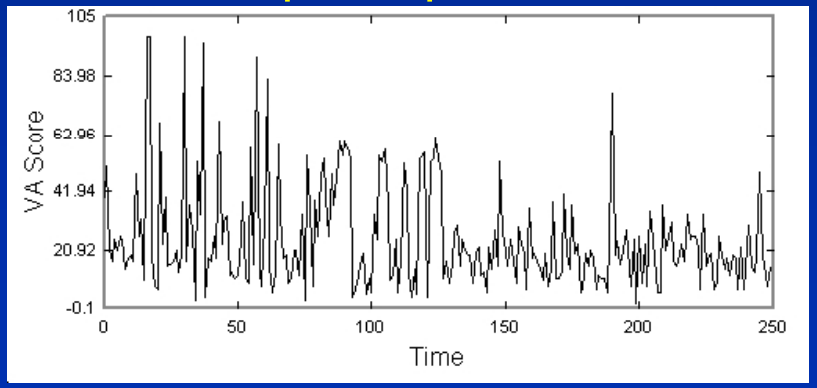
participant B



participant C

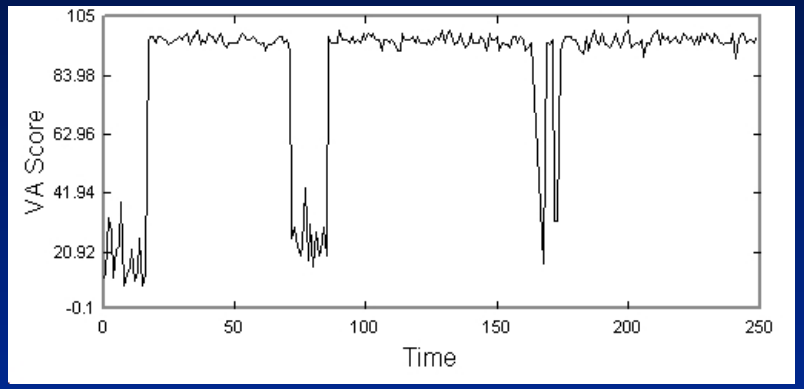


participant D

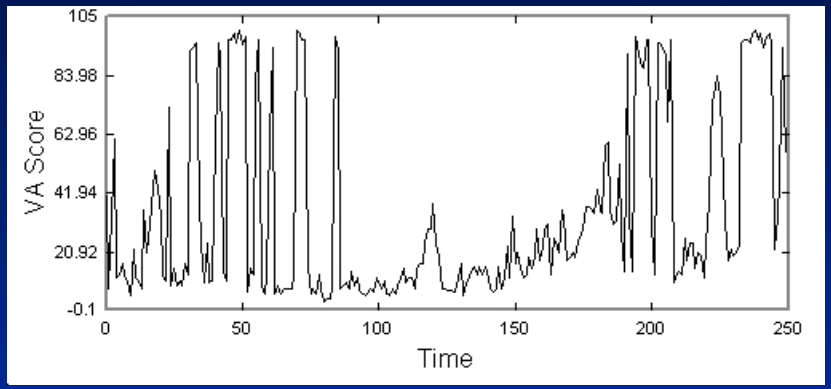


Eye Movement Behaviour – clear ROI

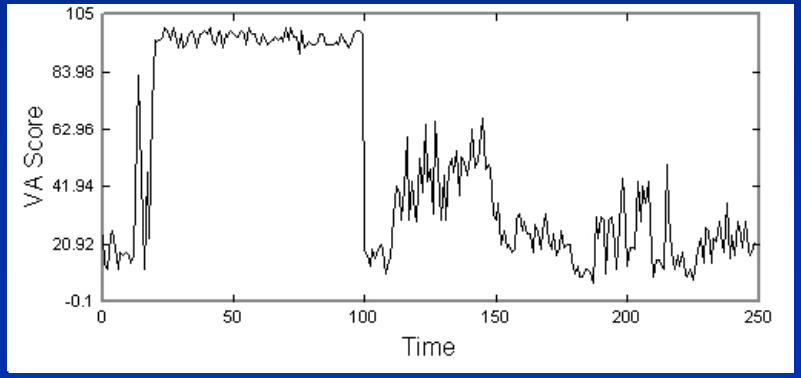
participant A



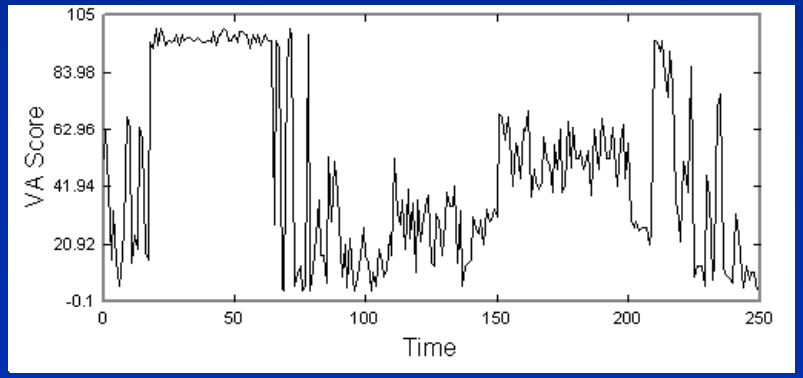
participant B



participant C



participant D



Variance of Attention Measure

		Image Variance	Participants			
			A	B	C	D
Unclear ROI	Image1	298	325	193	333	532
	Image2	500	479	496	328	629
	Image3	175	389	175	365	197
Obvious ROI	Image4	443	741	687	1094	857
	Image5	246	1432	1453	1202	1466
	Image6	378	1246	1226	862	1497

Time Fixating Salient Regions (ms)

Images		Participants			
		A	B	C	D
Unclear ROI	1	40	60	20	140
	2	580	420	500	400
	3	100	0	40	20
Obvious ROI	4	2820	2340	2420	1280
	5	3680	1480	2220	1960
	6	4240	980	1620	1240

Findings

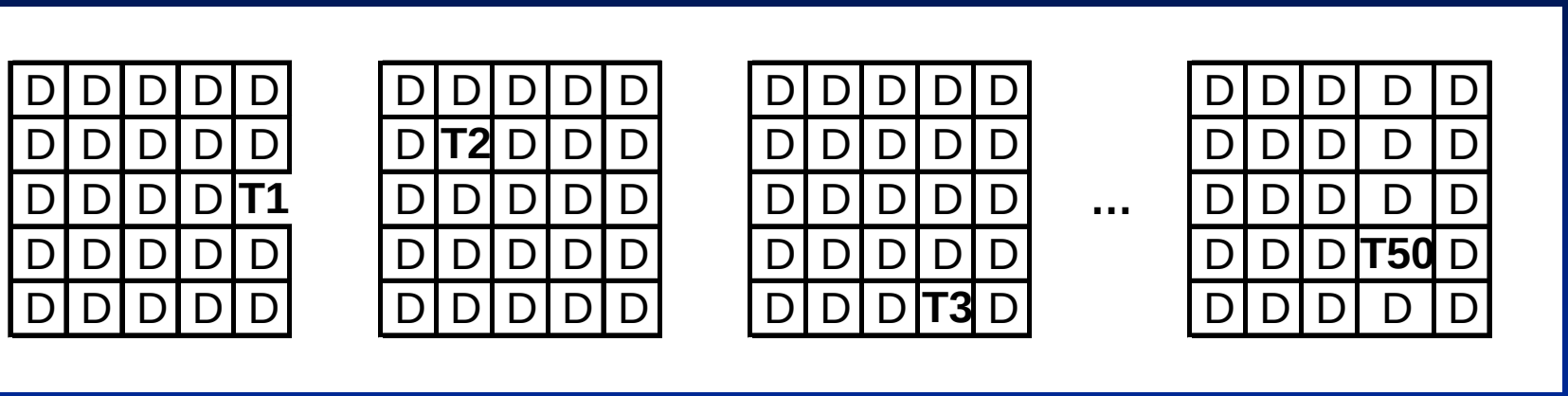
- No special fixation sequence although many look at salient regions first
- Very salient regions inspected frequently and compared with background

Eye vs Mouse for Image Identification

target image

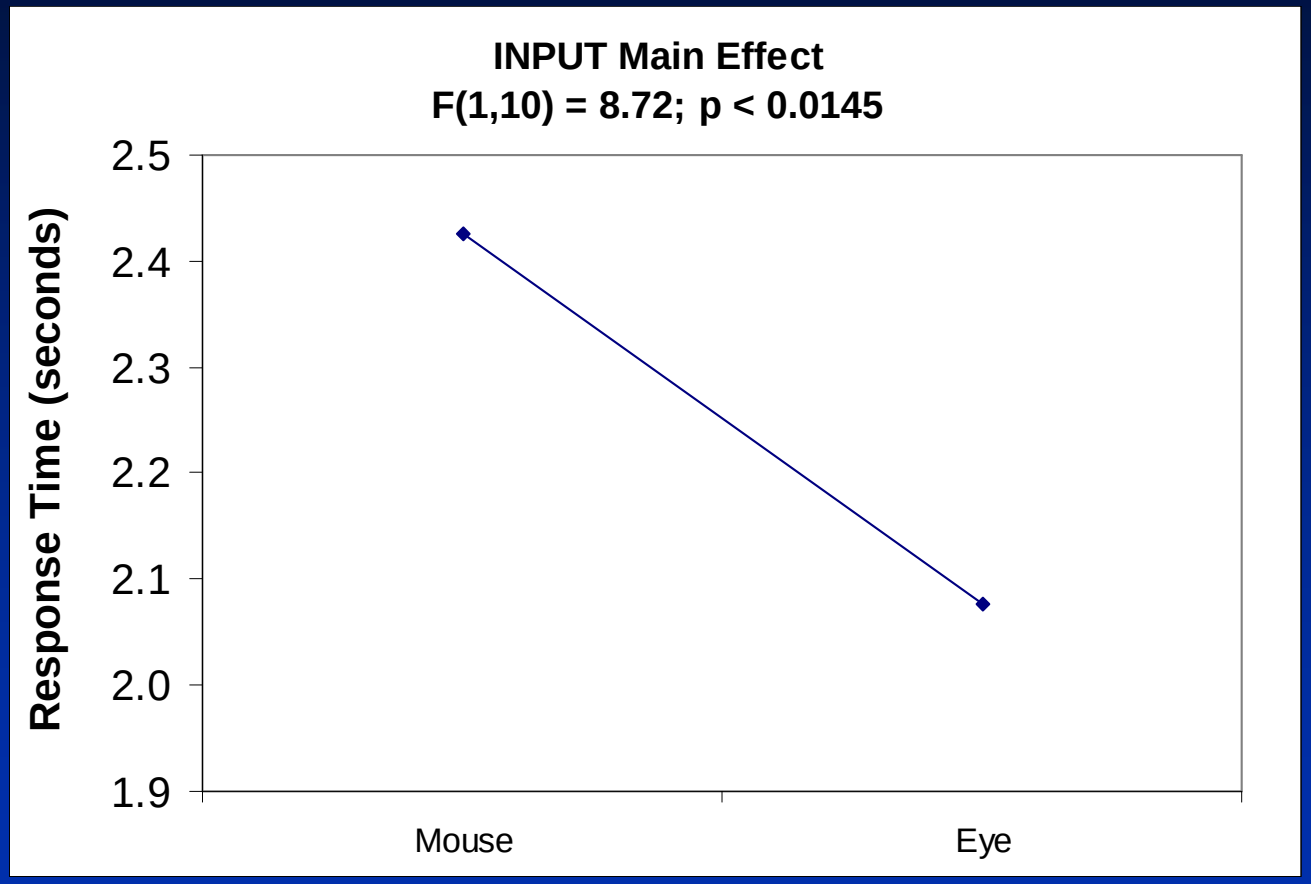
1. Mouse click
2. Fixation > 40ms

Screen Display Sequence



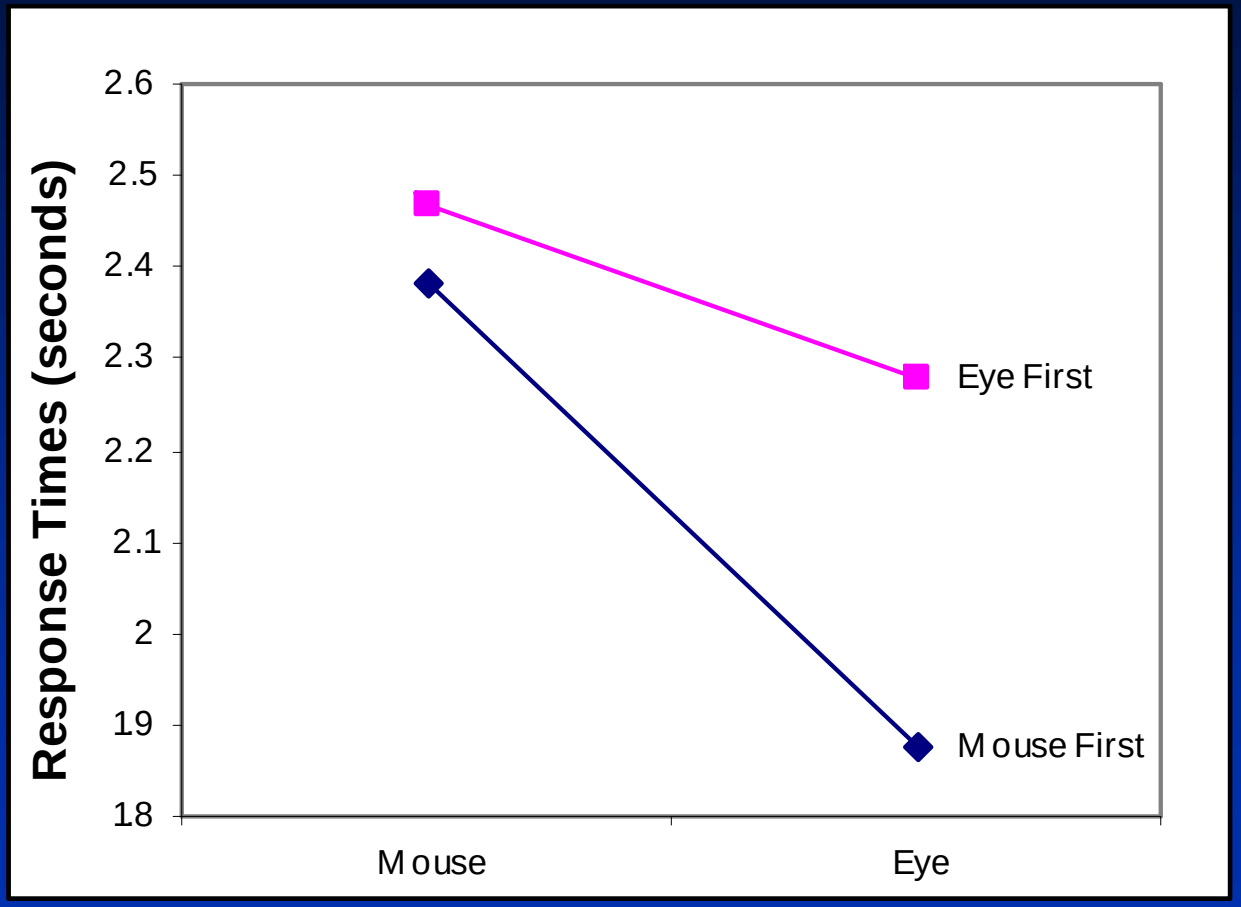
D = distractor Tn = target image

Eye vs Mouse Response Times



12 participants

Eye vs Mouse Response Times

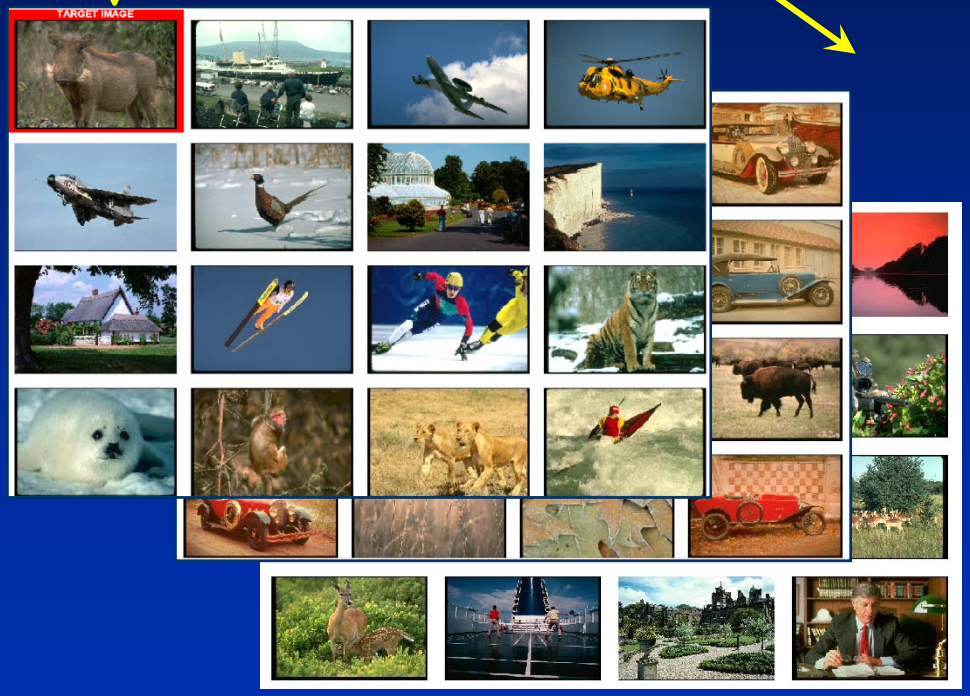


6 participants in each group

Image Search Task

target image

steps to target



target image

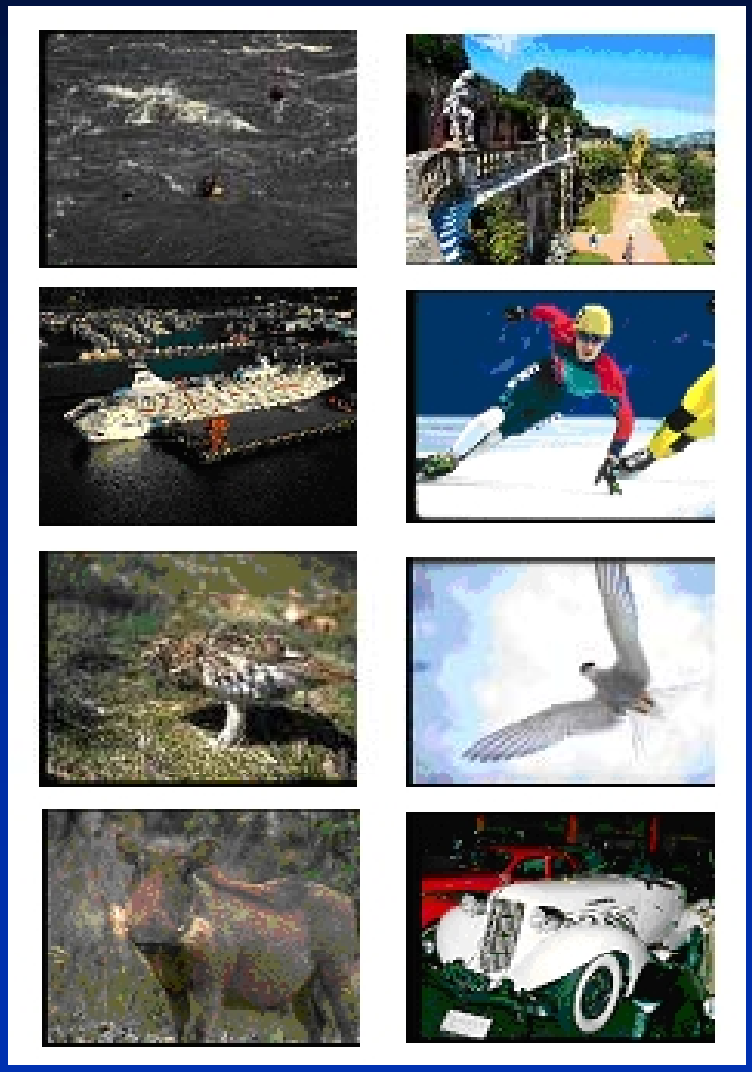
1000 images
13 participants

Image Selection

- Gaze selection of an image is determined by the sum of all fixations of 80ms or more on that image exceeding a threshold.
- Two thresholds 400ms and 800ms
- Successive sets of 15 images are retrieved based on their similarity with selected image.
- Performance compared with images randomly retrieved
- Participants not told what determines screen changes

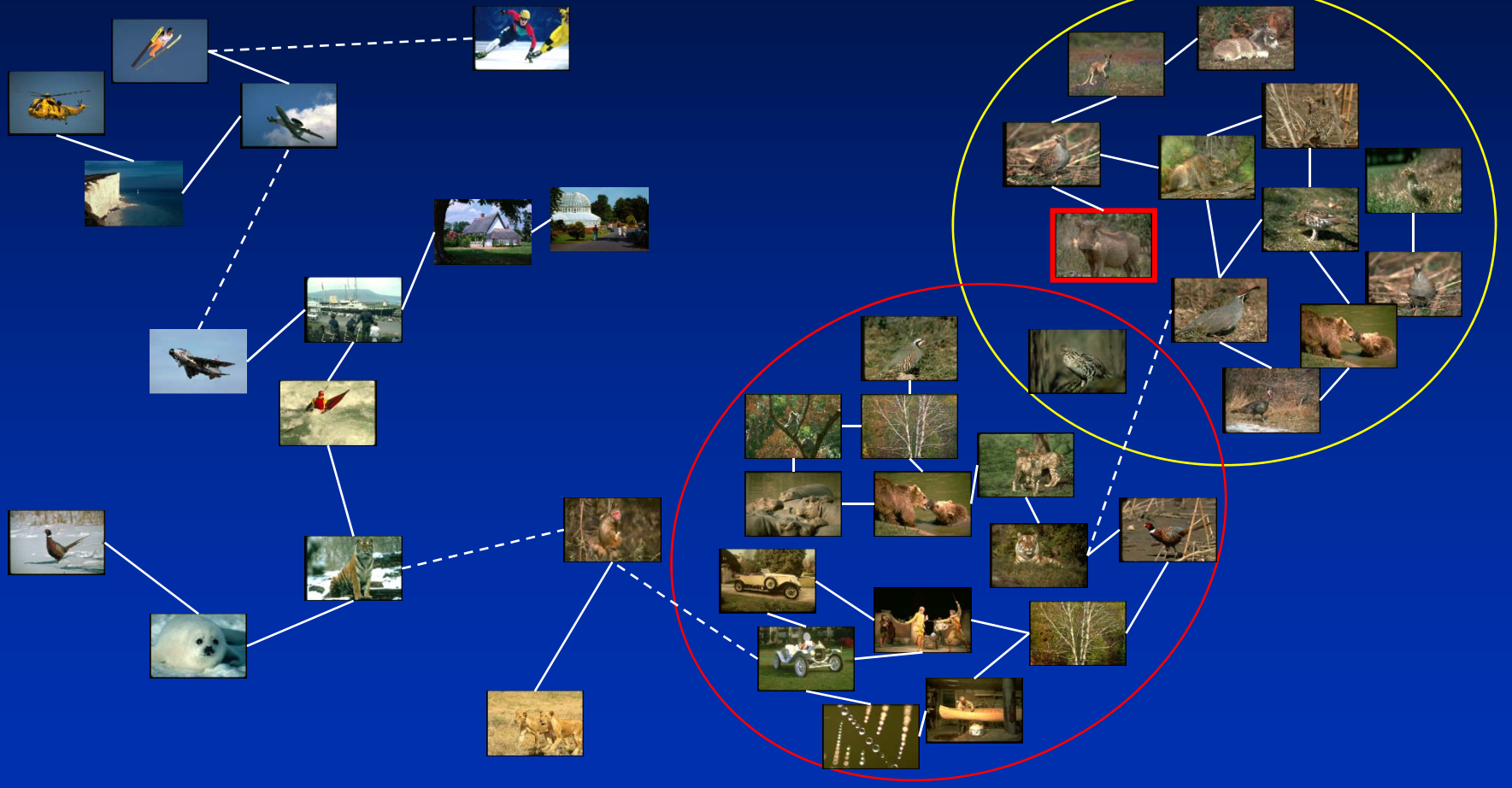
Target Images

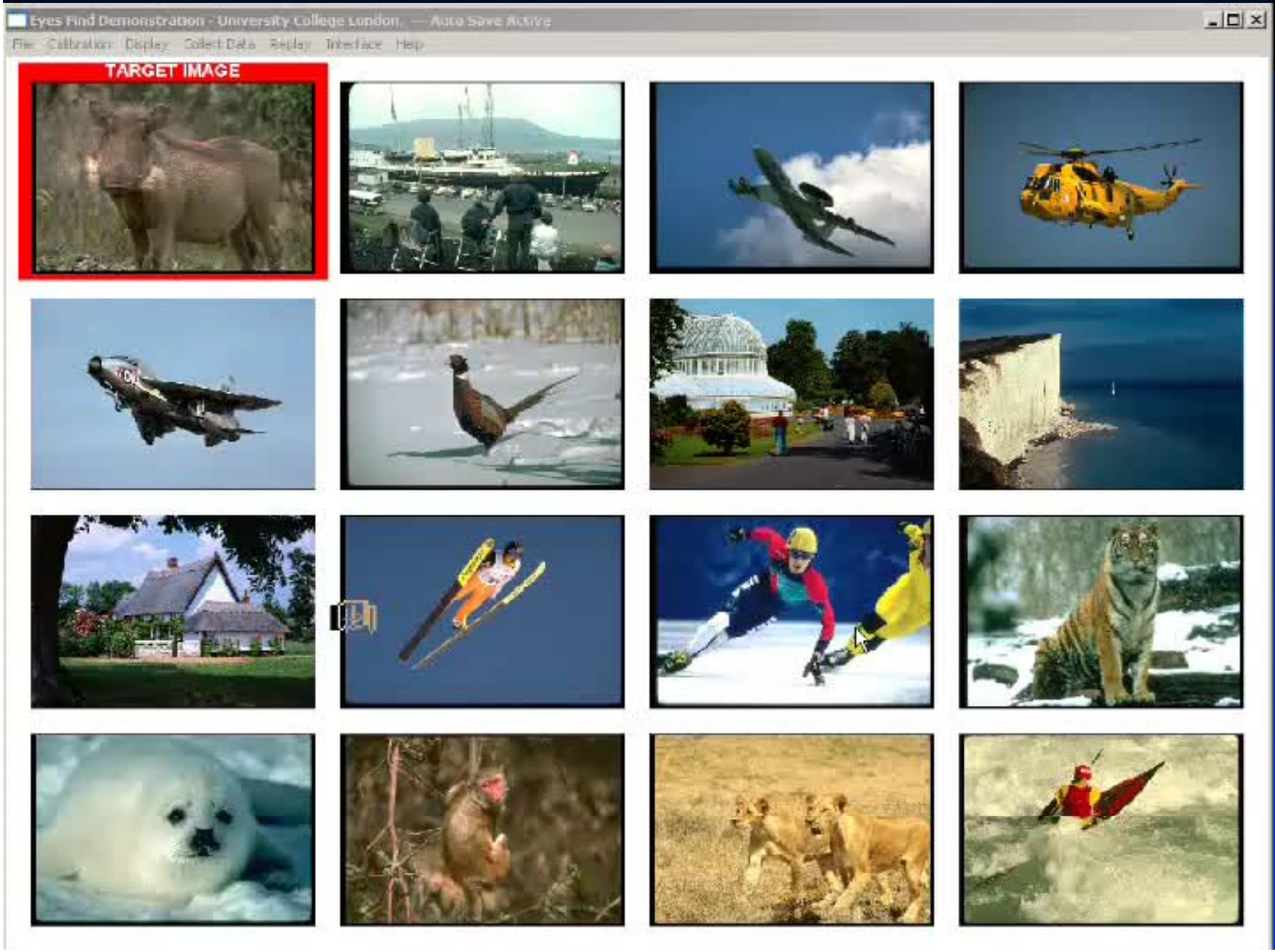
easy to find



hard to find

Similarity Links




Eyes Find Demonstration - University College London — Auto Save Active
File Calibration Display Collect Data Replay Interface Help
TARGET IMAGE

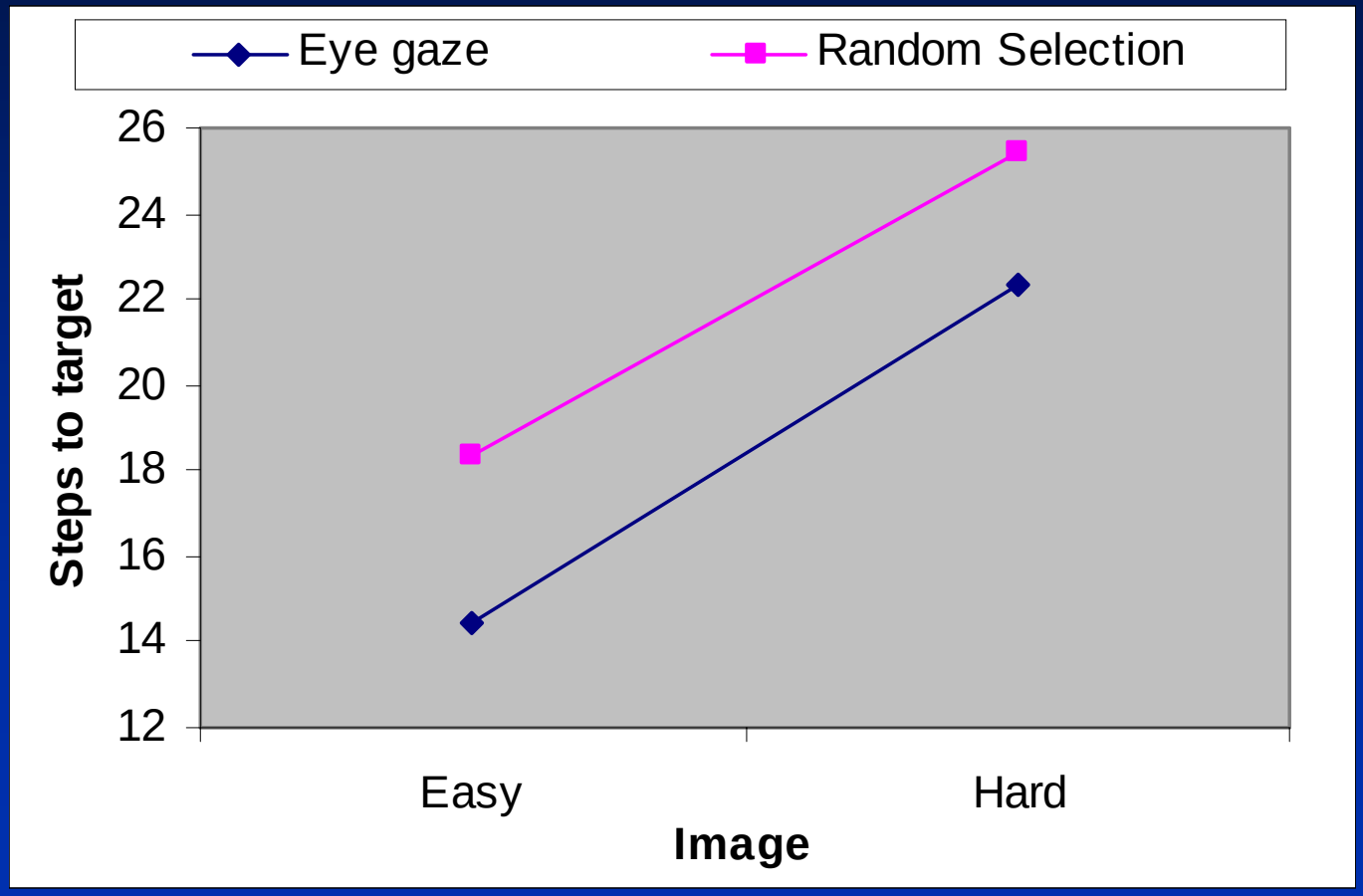
Results

Selection Mode	Image Type	Steps to target
Eye gaze	Easy-to-find	14
		15
	Hard-to-find	23
		21
Random selection	Easy-to-find	20
		16
	Hard-to-find	25
		26

13 participants
8 sessions

Main effect: Eye gaze 18 steps
Random 22 steps
$p < 0.037$

Results – Easy vs Hard Images



Other Selection Criteria

Fixation Threshold	Steps to target	Time to target (seconds)	Average Time per display	Fixation Numbers	Average Fixation Numbers per display
300ms	17	17.9	1.081	53	3
400ms	18	28.1	1.630	86	5
Revisit	16	37.7	2.352	99	6
Revisit/400ms	17	24.0	1.470	72	4

24 participants
8 sessions

Main effect: fixation threshold not significant

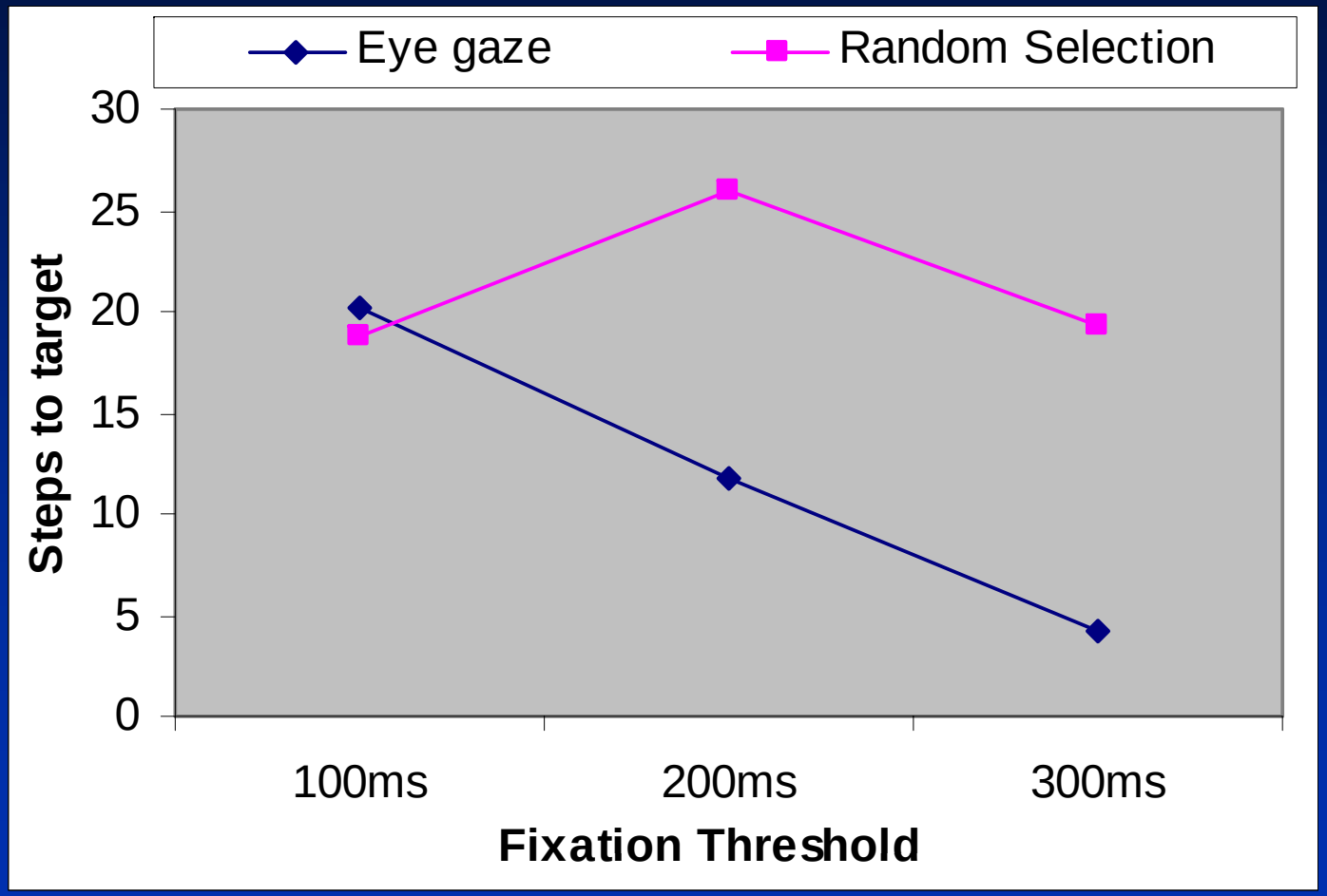
Results - Lower Fixation Thresholds

Fixation Threshold	Steps to target	Time to target (seconds)	Average Time per Display	Fixation Numbers	Average Fixation Numbers per Display
100ms	20	8.0	0.394	20	1
200ms	12	7.0	0.634	18	2
300ms	4	5.2	1.139	17	3

6 participants
3 sessions

Significant differences between random and 200ms + 300ms.

Results - Lower Fixation Thresholds



Conclusions

- Eye tracking can be faster than tactile interfaces for visual tasks
- Eye tracking interfaces are feasible for fast image search
- Pre-attentive vision plays a part in very rapid search

Future Work

- Further study of human visual behaviour
- Use of higher performance similarity measures
- Application to browsing large collections of photos/videos
- Shared interaction