Eye Movements While Watching a Video: Comparisons Across Viewer Groups

R.B. Goldstein, E. Peli, S. Lerner, G. Luo

Schepens Eye Research Institute, Harvard Medical School, Boston, MA

Vision Science Society 2004

Abstract (Updated)

Eye movements while watching video: comparisons across viewer groups

Identifying the area of interest (AOI) on a video frame may be necessary: (1) to develop a television magnifying aid for people with low vision; (2) to implement some data compression schemes; and (3) to transform images for rendering on devices with small display areas. We determined the AOI in video frames by recording and analyzing the eye movements of 4 groups (Young Male-YM, Young Female-YF, Older Male-OM and Older Female-OF) of 5 subjects each while they watched 4-8 minute video segments of 6 movies. Within-group AOI (temporal and spatial coincidence) was marked on a frame if valid eye position data was available for at least 4 of the 5 subjects and the bivariate contour ellipse area (BVCA) (k=1, P=63.2%) was < 9 deg^2 (the screen was 27x15 deg). Within group, temporal coincidence occurred for a minimum of 58% of frames in all movies.

Within-group 9 deg^2 AOI rates ranged from 39% (YF) to 72% (OM) of all video frames (chance coincidences ranged from 11% to 32%). Between age and gender groups, the AOIs were < 3 deg apart at least 65% of the time. This was not due solely to the AOIs being clustered around the center of the screen, since they were within 3 deg of the center of the screen an average of 54% of the time. For a ≤ 25 deg^2 BVCA, the AOIs were within 5 deg of the screen center at least 91% of the time, accounting for almost all of the time that the AOIs were within 5 deg of each other.

Conclusion: The high level of AOI coincidence within and across age groups, even away from the screen's center, suggests that a single AOI might be appropriate for varied audiences in many applications. Analysis of the frames where group AOIs differ might be of interest in determining what type of visual or content categories account for difference between gender and age group AOIs.

Background

We want to:

- **Develop a television magnifying aid for people with low vision**
- Implement some data compression schemes
- Transform images for rendering on devices with small display areas

To do that, we need to know where people look when watching video. In particular, do they look at the same thing (i.e. in the same area on the screen)?

For static images, the scanpath[1,2] depends on

- Gender and age characteristics of the observers
- Task they are asked to perform
- Types of images

Area of Interest (AOI) - the most important part of a video frame – where people look

In watching a movie video:

- To what degree do people look at the same or different areas?
- Is there a consistency within gender and age groups in AOI?
- Does the AOI vary among groups?
- Does it vary with video content?

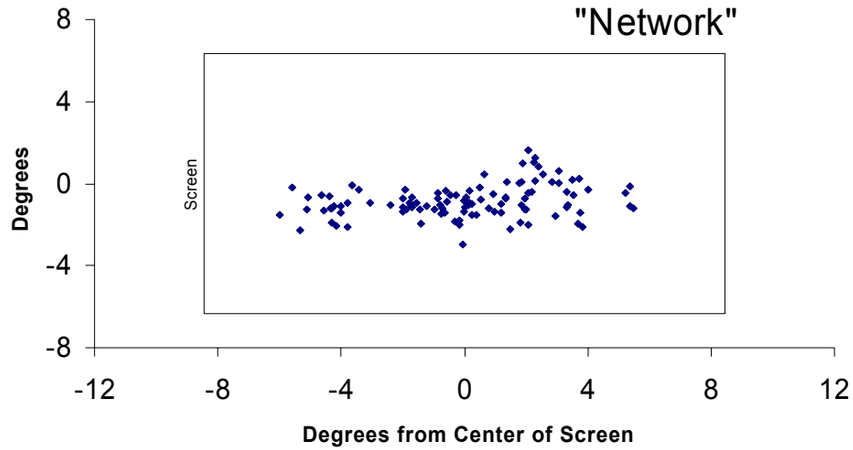
We have found little prior work on eye movements while watching video, but did find a suggestion that there might be differences between gender groups[3]

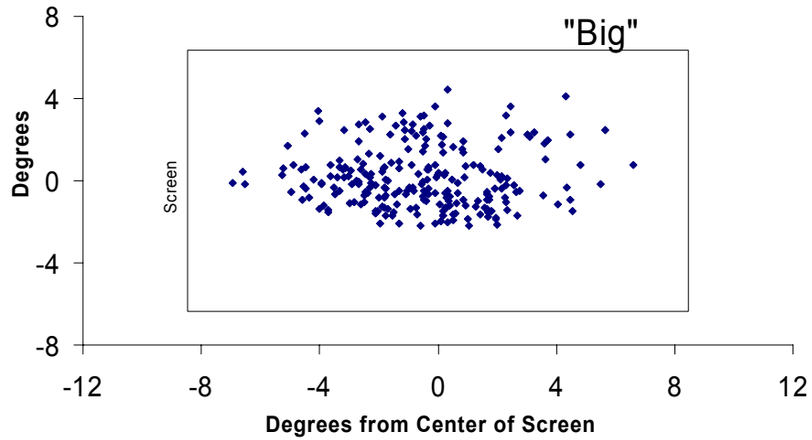
How to determine AOI

1. Should we record eye movements from a single observer?
   - –Some data is lost due to blinks and loss of tracking
2. Should we record a single observer watching multiple times?
   - –Possibly different viewing strategy with repeat viewing (different study)
3. We recorded eye movements from several observers
   - –Different observers may look at different objects

We merged these multiple eye-coordinate files into groups based on age and gender

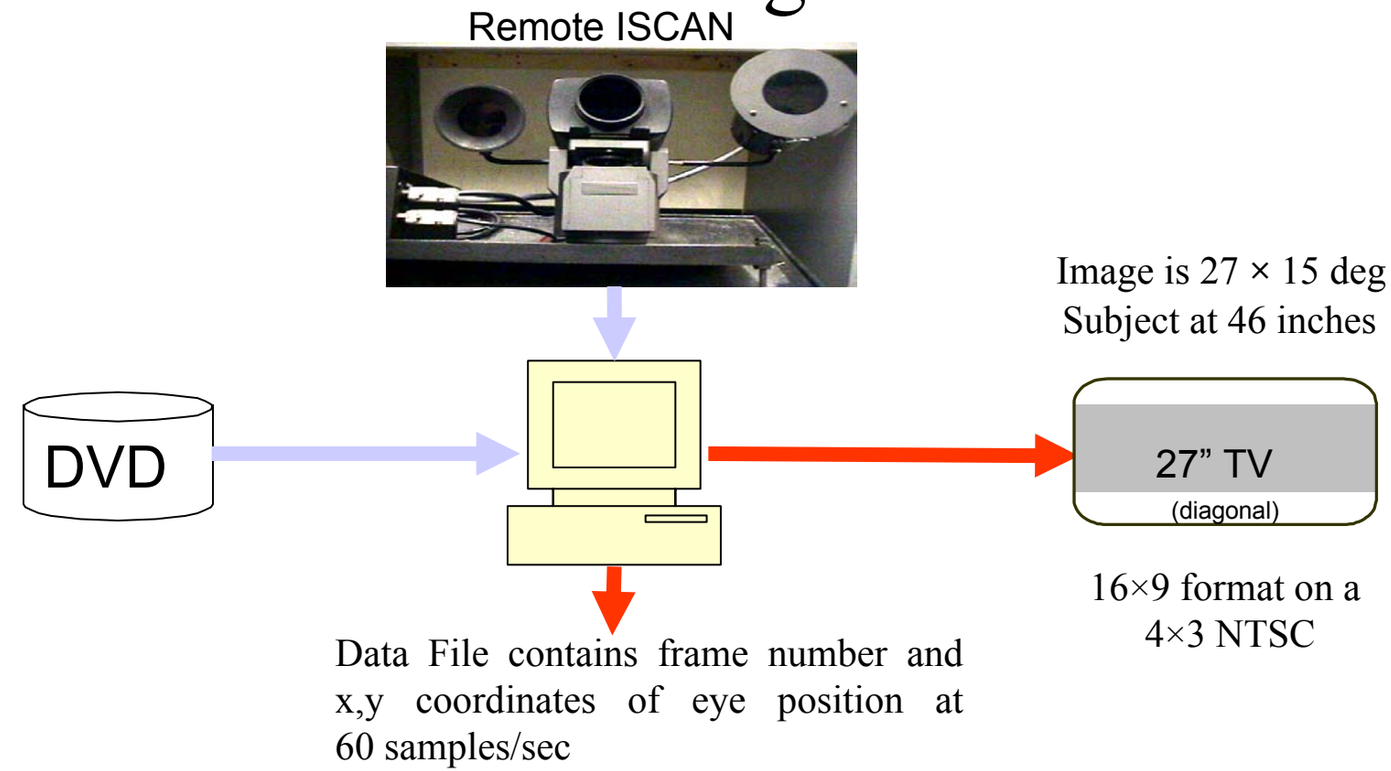
Eye movements are not limited to the center of the screen[4]





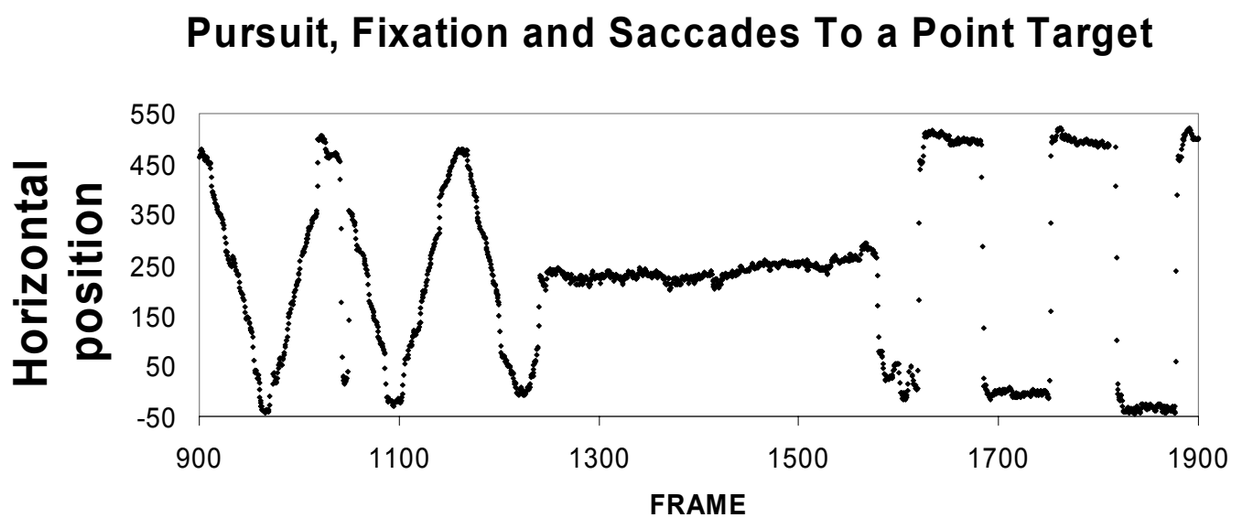
Fixation groups from 2 clips used in this (and prior study with smaller screen area) show wide distribution across the screen

Recording



Three Types of Calibrations

- ISCAN Internal 5-Point Calibration
- Pre-Segment and Post-segment Calibrations to relate ISCAN data to screen coordinates
- Preliminary recordings of pursuits, fixations and saccades to known targets to validate setup



Subject Groups and Videos

N=5 for every group	Age Average (range)	VA Average (range)	Education Level Median (range)	Hours of TV Per Week Average (range)	Refractive Correction
Young Male	24 (26-36)	20/13 (20/11 – 20/17)	BS (HS – BS)	10 (1 – 25)	Glasses: 3
Young Female	25 (18-29)	20/15 (20/14 – 20/17)	BS (HS – PhD)	11 (1 – 30)	Contacts: 2 Glasses: 1
Old Male	58 (42-66)	20/16 (20/14 – 20/18)	PhD (BS – PhD)	8 (3 – 20)	Glasses: 2
Old Female	58 (51-62)	20/14 (20/11 – 20/20)	MS (HS – PhD)	13 (2 – 20)	Glasses: 2

Category	Title	Time
Talk Show	Quiz Show (1994)	6:40
Romance	Shakespeare in Love (1998)	7:06
Sports	Any Given Sunday (1999)	4:12
Documentary	Blue Planet (2001)	8:14
News	Network (1976)	4:02
Comedy	Big (1988)	6:29
	Total (min:sec)	37:29

Observers watched each video only once. Some had seen a few of the movies in the distant past

Eye Movement Pre-processing Filter

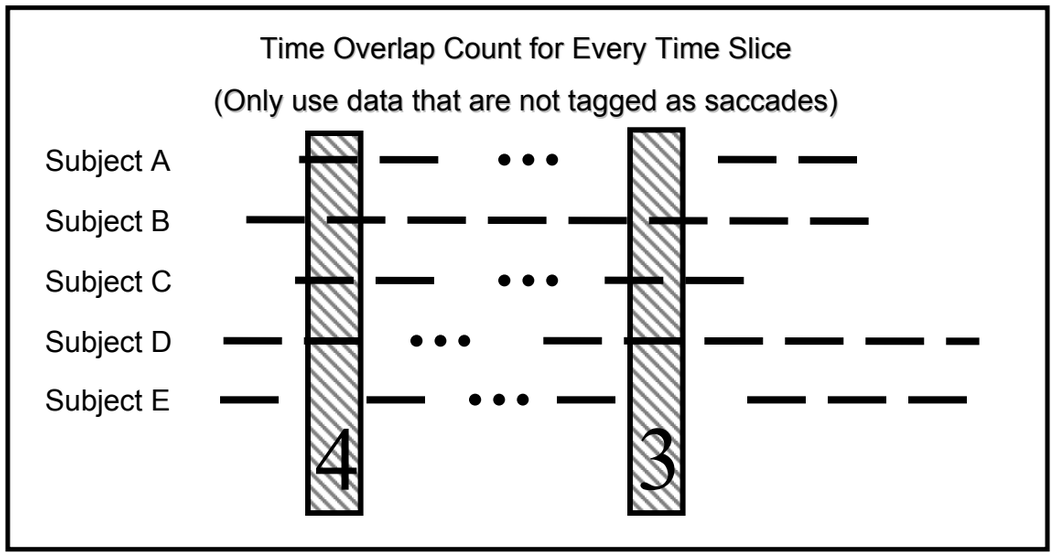
- Filter Purpose
  - Remove zero data (due to blinks, other loss of tracking)
  - Reject questionable data due to DVD-ISCAN timing interactions (DVD only provides a time-stamp at most at 2 Hz)
  - Reject data due to pupil diameters out of expected range
  - Apply calibrations to data
    - Convert data to visual angle and calculate eye movement velocities
  - Tag data as fixations, smooth pursuits or saccades (not used here)
  - Calculate statistics on fixation and smooth pursuit sections (not used here)

- Filter Criteria
  - Made very “liberal” here to provide maximum yields for the merge process
    - Velocity threshold for saccades set at 200 deg/sec
    - Pupil diameters allowed to be almost the entire range possible from ISCAN
    - As few as 2 samples (33ms) in fixation and pursuit groups are allowed
  - The following process (see panels 7 and 8) will tend to eliminate questionable data as they will not contribute to a high coincidence rank

Temporal and Spatial Coincidence

Temporal Coincidence Count

Samples are taken at 60 Hz, about one sample per video field, but not synchronized. Each sample is allocated to a field

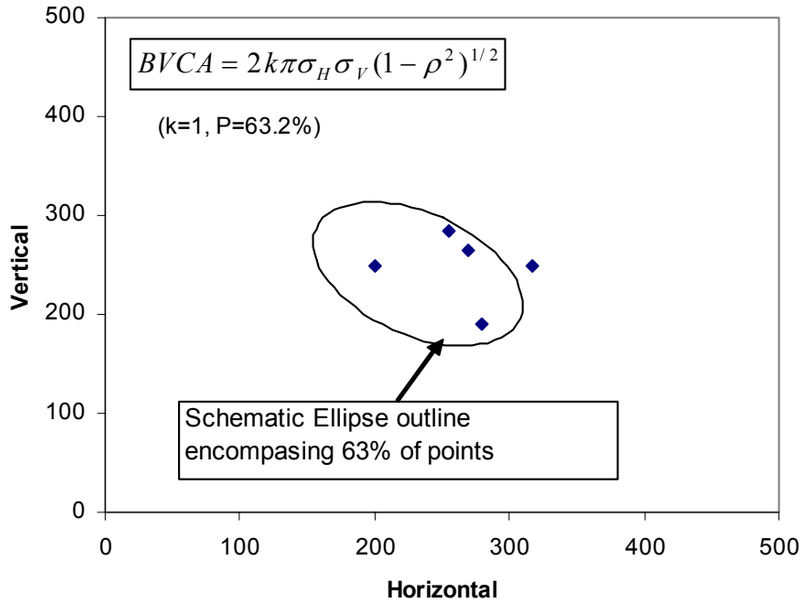


Horizontal line segments represent samples for which valid eye position data was available. A spatial coincidence must FIRST be a temporal coincidence. Temporal coincidence counts of 4 and 3 are shown for 2 non-consecutive fields.

Spatial Coincidence Count

The spatial count ***for every video field*** is based on the Bivariate Contour Ellipse Area (BVCA) statistic[5]

[σ_x and σ_y are the standard deviations in the x and y directions. ρ is Pearson correlation coefficient]



Individual points are removed one at a time (in arbitrary order) until the BVCA falls below the criterion (9 deg^2 or 25 deg^2). Spatial coincidence count is N (number of observers) minus the number of points that had to be removed. Spatial coincidence of 4 is shown

Coincidence Rank (CR_i)

T = Temporal coincidence count

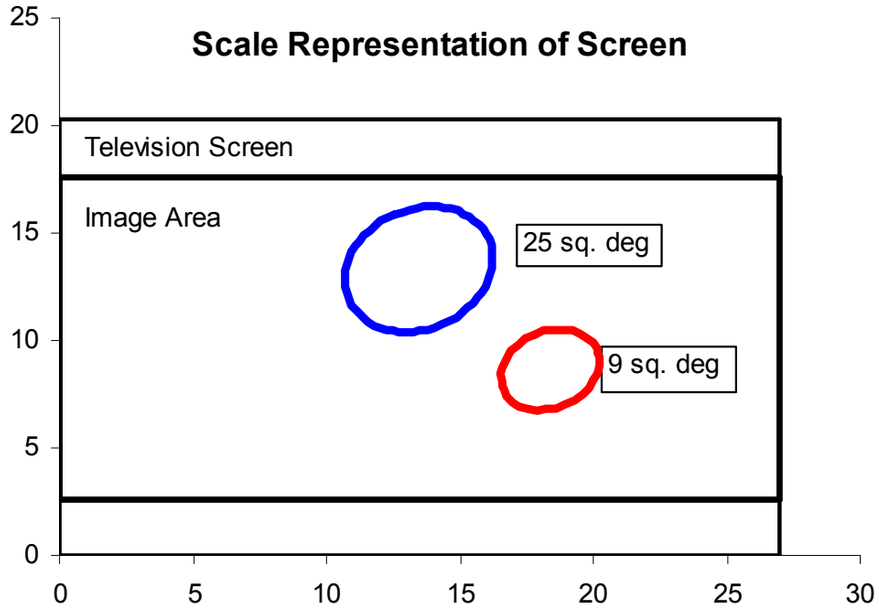
P = Spatial coincidence count

$$CR_i = N\frac{P^2}{T}$$

5-Subject Coincidence Ranks		
Temporal Coincidence	Spatial Coincidence (within Time)	Coincidence Rank
5	5	25
4	4	20
5	4	16
3	3	15
4	3	11.3
2	2	10

CR is increased if the prior adjacent (in time) record has a higher CR

CR formula is extensible to different numbers of subjects and time overlaps



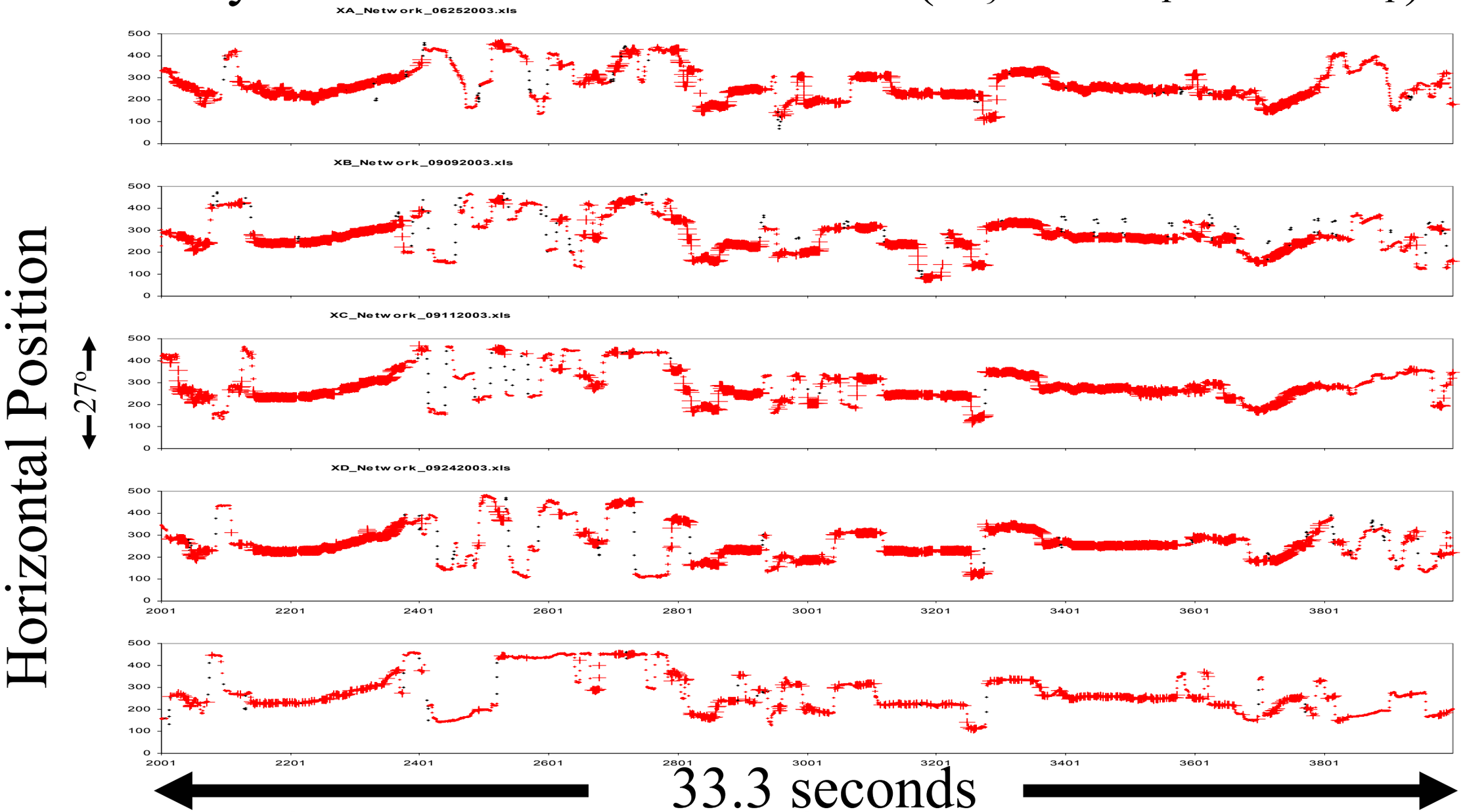
9 deg^2 is 2% of the image area

25 deg^2 is 6% of the image area

CR is dependent upon which criterion is used (CR_9 or CR_{25})

AOI is defined when $CR_i \geq 16$

Eye Position Traces With AOI (CR_9-OM Group-Network Clip)



The size of each marker is monotonically related to the Coincidence Rank. Black points are rejected data. Incidents of spatial coincidence, probable scene changes (where all subjects change fixations simultaneously), pursuits of moving targets, and saccades are evident in these traces

Pre-Processing Filter Yields

FILTER YIELDS		% Accepted Records from Filter The 5 subjects with the best yields are chosen For Inclusion in the Group Analyses						
Group	**ID**	**Sunday**	**Big**	**Blue**	**Network**	**Quiz**	**Shakes**	**Avg.**
OldMale	OM1	92%	93%	90%	91%	91%	91%	91%
	OM2	91%	89%	85%	85%	86%	85%	87%
	OM3	95%	94%	94%	92%	93%	92%	94%
	OM4	94%	93%	93%	92%	93%	92%	93%
	OM5	83%	86%	80%	87%	86%	87%	85%
	OM6	91%	91%	90%	89%	91%	89%	90%
YoungMale	YM1	85%	91%	90%	90%	90%	90%	89%
	YM2	89%	86%	87%	88%	87%	88%	88%
	YM3	80%	82%	79%	82%	80%	82%	81%
	YM4	92%	87%	89%	92%	88%	92%	90%
	YM5	85%	89%	91%	93%	93%	93%	90%
	YM6	88%	88%	88%	88%	86%	88%	88%
OldFemale	OF1	87%	88%	88%	87%	87%	87%	87%
	OF2	91%	77%	86%	89%	87%	89%	86%
	OF3	91%	57%	84%	71%	42%	71%	69%
	OF4	87%	90%	91%	91%	89%	91%	90%
	OF5	89%	91%	91%	90%	89%	90%	90%
	OF6	89%	87%	88%	87%	87%	87%	88%
	OF7	87%	87%	87%	88%	85%	88%	87%
YoungFemale	YF1	81%	80%	80%	83%	75%	83%	80%
	YF2	90%	89%	88%	90%	88%	90%	89%
	YF3	79%	82%	78%	79%	80%	79%	79%
	YF4	87%	87%	82%	80%	89%	80%	84%
	YF5	83%	82%	86%	84%	85%	84%	84%
	YF6	91%	91%	92%	90%	90%	90%	91%

Lowest yield subjects in every group rejected to keep 5 in each group

- Young Female group had lowest overall yields (perhaps caused by the use of contact lenses – see panel 5)
- Lowest average yield was greater than 80%

Within Group Results (Temporal and Spatial Coincidence)

	Percent Temporal Coincidence (4 or 5 Subjects Have Data at Same Time)						Average	Multiple
	Sunday	Big	Blue	Network	Quiz	Shakes		
OldMale	73%	73%	72%	70%	71%	69%	71%	73%
YoungMale	67%	69%	69%	69%	70%	71%	69%	69%
OldFemale	72%	73%	72%	72%	71%	73%	72%	72%
YoungFemale	69%	69%	59%	67%	68%	59%	65%	66%
Average	70%	71%	68%	70%	70%	68%		70%

9 Sq. Deg. Criteria	Percent of **The Above Temporal Coincidences** With AOI (CR ≥ 16)						Average	% of Clip	Multiple
	Sunday	Big	Blue	Network	Quiz	Shakes			
OldMale	76%	64%	71%	76%	68%	76%	72%	51%	26%
YoungMale	47%	38%	52%	59%	51%	74%	53%	37%	19%
OldFemale	78%	55%	61%	71%	59%	76%	67%	48%	32%
YoungFemale	44%	27%	21%	47%	36%	57%	39%	25%	12%
Average	61%	46%	51%	63%	53%	71%	58%		22%

25 Sq. Deg. Criteria	Percent of **The Above Temporal Coincidences** With AOI (CR ≥ 16)						Average	% of Clip	Multiple
	Sunday	Big	Blue	Network	Quiz	Shakes			
OldMale	95%	87%	95%	91%	90%	94%	92%	66%	71%
YoungMale	84%	77%	88%	86%	83%	93%	85%	59%	56%
OldFemale	96%	83%	91%	90%	86%	96%	90%	65%	76%
YoungFemale	83%	63%	62%	75%	69%	86%	73%	48%	48%
Average	90%	78%	84%	86%	82%	92%	85%		63%

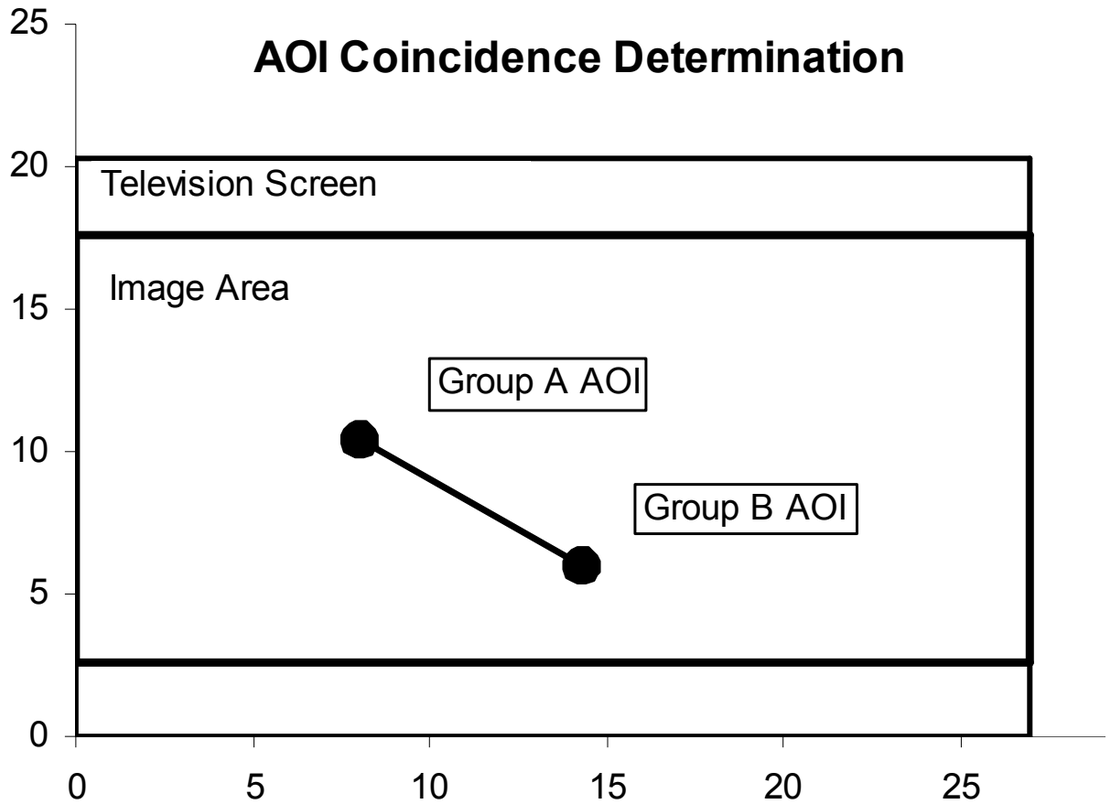
"Multiple" is a merging of DIFFERENT video clips to determine the chance coincidence rate

As expected, the temporal coincidence rate of each group is similar to the chance rate

Between-Group Results (AOI Coincidence)

	For those AOI's that have temporal coincidence, the percent that have spatial coincidence			
	OM	YM	OF	YF
OM		88.6%	94.6%	90.0%
YM	70.3%		88.0%	84.5%
OF	85.3%	71.2%		93.2%
YF	75.9%	65.0%	81.2%	
	Blue is the 5 degree criterion results			
	Yellow is the 3 degree criterion results			

	Percent of AOIs with spatial coincidence that were within 3° or 5° of the center of the screen			
	OM	YM	OF	YF
OM		96.0%	95.8%	92.9%
YM	53.1%		95.5%	92.0%
OF	53.1%	54.6%		91.6%
YF	53.6%	53.5%	54.4%	
	Blue is the 5 degree criterion results			
	Yellow is the 3 degree criterion results			



If the distance between the two Group AOI positions is less than the criterion, there is AOI coincidence

Summary

- Between all groups, when the AOIs had temporal coincidence, they were less than 3 degrees apart from 65% to 85% of the time (panel 12, table 1)
- The high level of AOIs and AOI coincidences between groups, even away from the screen center, suggests that a single AOI might be appropriate for varied audiences in many applications
- Within groups, the Old Male group had the highest AOI rate and the Young Female group had the lowest AOI rate (panel 11). The AOI rate was 2-3 times the chance coincidence rate (for 9 deg^2 criteria)
- The Young Female group had lowest overall yields, perhaps due to the use of contact lenses
- Analysis of the frames where group AOIs differ might be of interest in determining what type of visual or content categories would account for such difference between observers and among gender and age group AOIs

Conclusion: People tend to look at the same things in videos, independent of age and gender

References

1. S. Josephson and M. E. Holmes, "Attention to repeated images on the World-Wide Web: another look at scanpath theory," Behav. Res. Methods Instrum. Comput. **34**, 539-548 (2002).
2. L. W. Stark, K. Ezumi, T. Nguyen, R. Paul, G. Tharp and H. I. Yamashita, Visual search in virtual environments. In: B. E. Rogowitz, ed. Human Vision, Visual Processing, and Digital Display III/Human Perception, Performance, and Presence in Virtual Environments. San Jose, CALIFORNIA: The International Society for Optical Engineering, 1992:577-589
3. V. Tosi, L Mecacci, E. Pasquali, Scanning eye movements made when viewing film: preliminary observations. Int. J. Neuroscience 29(1-2) 47-52 (1997) .
4. R. B. Goldstein, E. Peli, H. Apfelbaum and R. Hier. Dynamic control of magnified image for low vision observers. 2002 Annual Meeting Abstract and Program Planner [on CD-ROM]. Association for Vision and Ophthalmology. Presentation #3803. 2002
5. M. D. Crossland, G. S. Rubin. The use of an infrared eye tracker to measure fixation stability. Optom. Vision Sci, 79, 735-739 (2002)

Supported by NIH grants EY05957, EY12890