Age Differences in Online Processing of Video: An Eye Movement Study

Heather L. Kirkorian, **Daniel R. Anderson**, and **Rachel Keen**
Department of Psychology, University of Massachusetts at Amherst

Abstract

Eye movements were recorded while 62 one-year-olds, four-year-olds, and adults watched television. Of interest was the extent to which viewers looked at the same place at the same time as their peers because high similarity across viewers suggests systematic viewing driven by comprehension processes. Similarity of gaze location increased with age. This was particularly true immediately following a cut to a new scene, partly because older viewers (but not infants) tended to fixate the center of the screen following a cut. Conversely, infants appear to require several seconds to orient to a new scene. Results are interpreted in the context of developing attention skills. Findings have implications for the extent to which infants comprehend and learn from commercial video.

The American Academy of Pediatrics recommends that children under 2 years of age have no exposure to screen media including television and videos (American Academy of Pediatrics, 1999; 2010). Despite these recommendations, over one third of very young children currently watch more than one hour of television per day (Rideout & Hamel, 2006). Although many programs and videos targeting this young audience make claims about educational value either explicitly through marketing or implicitly with names such as Baby Einstein (Garrison & Christakis, 2005), research suggests that children under two years of age suffer from a video deficit whereby they learn substantially less from symbolic media (such as video) than from equivalent real-life experiences (see Anderson & Pempek, 2005; Troseth, 2010). For example, infants and toddlers are more likely to imitate target behaviors demonstrated by a live model than by that same model on video (e.g., Barr & Hayne, 1999; Hayne, Herbert, & Simcock, 2003).

Most experimental studies of very young children's comprehension of video have focused on infants' ability to demonstrate learning. Few studies have focused on age differences in online processing of video. The present study recorded eye movements by infants, preschool-age children, and adults as they watched television. The goal was to determine whether there are systematic age differences in online processing of video as evidenced by eye movements during video viewing.

Correspondence concerning this article should be addressed to Heather L. Kirkorian, Human Development and Family Studies, University of Wisconsin, 1305 Linden Drive, Madison, WI 53706. kirkorian@wisc.edu.
Heather L. Kirkorian is now at University of Wisconsin at Madison.
Rachel Keen is now at University of Virginia.

Development of Attention to Video

Early theories of children's attention to television posited that formal features such as cuts and movement elicit automatic orienting responses and that, as a result, children's television viewing is void of understanding or learning (Singer, 1980). In contrast, Anderson and Lorch (1983) proposed that comprehension primarily drives children's attention to television: Children choose to watch content that is understandable to them and to not watch incomprehensible content. Huston and Wright (1983) proposed a theory that subsumes both ideas: Formal features and comprehension each play an important role in driving attention, with the role of both changing during development and with media experience. This theory distinguishes between perceptually salient and informative formal features. Perceptually salient features include sound effects, fast movement, and rapid pacing whereas informative features include dialogue, narration, and story arc. In the Huston and Wright model, attention in very young children is initially drawn to perceptually salient formal features but, with age and experience, viewers habituate to these and differentially attend to informative features. Thus the pattern of attention to television should change qualitatively over the first few years of life.

The Huston and Wright (1983) model of attention to television has aspects that are analogous to Ruff and Rothbart's (1996) more general theory of attention development. This latter theory posits that there are two systems of attention. The first is present in infancy and is referred to as the orienting/investigative system of attention. In this primary system, attention is driven by perceptually salient features in the environment that elicit orienting responses. The infant has little active control over attention and quickly habituates to novel stimuli. By the end of the first year of life, the second system of higher-level controls begins to emerge. Attention begins to come under cognitive control as the infant can actively attend to a stimulus after habituating to initial novelty. Goals, plans, and comprehension schemes thus become integral to the higher-level attention system. The system of cognitive control continues to develop through the preschool years and is attributed in part to maturation of the brain, particularly frontal areas.

Together these theories suggest that visual attention to television may be driven by different mechanisms in very young children than in older children and adults. This developmental shift from bottom-up to top-down processes is due in part to changes in cognitive control and may also reflect increasing knowledge about conventions of video production through experience with video.

Eye Movements during Video Viewing

Most studies of eye movements during video viewing are limited to adult populations. Several such studies demonstrate that adult video viewers tend to look at the same parts of the screen at the same time (Dorr, Martinez, Gegenfurtner, & Barth, 2010; Goldstein, Woods, & Peli, 2007; Mital, Smith, Hill, & Henderson, 2010; Stelmach, Tam, & Hearty, 1991; Tosi, Mecacci, & Pasquali, 1997). This is consistent with the hypothesis that top-down, systematic attention processes drive adults' eye movements when watching video as there may be a variety of perceptually salient features on screen but only one focal point that is intended by the producer and is consistent with story arc. However, very little research compares eye movements across a wide age range. One study comparing eye movements by young versus older adults during video viewing suggests that, at least in adulthood, individual differences in the location of fixations decreases with age (Goldstein et al., 2007). Some research on eye movements to static scenes (e.g., line drawings) suggests that children's (relative to adults') fixations are more spatially variable and less likely to center on important regions (Mackworth & Bruner, 1970; Whiteside, 1974).

There have been only a few published studies of children's video viewing that utilize eye-tracking methodologies. Flagg (1978) observed 4- and 6-year-old children's eye movements while watching *Sesame Street*. The primary motivation for the research was that most studies of television viewing were interested in after-the-fact recall and behavior and that relatively little was known about moment-to-moment information processing during television viewing. One measure of interest was the spatial variability of fixations over time (i.e., number of screen regions that received a fixation). Flagg found no differences by age or sex. She proposed that there were no age differences in her study because television as a medium organizes information for the viewer (e.g., directs attention to important regions). Given this interpretation, one might expect that a younger viewer who is less experienced with video and less capable of following features such as dialogue and narration would show less predictable patterns of eye movements.

One study attempted to determine early developmental changes in attention to animated stories by observing eye movements. Takahashi (1991) recorded toddlers' (one to three years) eye movements by judging where the children were looking on the screen using video recordings of children's faces during the viewing sessions. Takahashi reported more frequent eye movements by older children during narrative segments than during the introduction segment whereas infants demonstrated the reverse. The author concluded that dynamic movements during the introduction attracted attention in one-year-olds but that three-year-old children were relatively more interested in the narrative. These conclusions are consistent with the Huston and Wright (1983) theory that younger children's attention is more influenced by salient formal features, whereas older children's attention is guided more by informative formal features in service of comprehension activities.

More direct evidence for Huston and Wright's (1983) theory comes from a recent study by Frank, Vul, and Johnson (2009) who recorded infants' and adults' fixations to brief (4-sec) video clips. They found that the relative importance of semantic information (i.e., faces) over perceptually salient features (e.g., visual contrast) as predictors of visual attention increased from 3 to 9 months and was even greater in an adult comparison group. Moreover, they found that the spatial variability (i.e., individual differences) of fixations to video frames decreased as a function of age, suggesting a greater reliance on systematic, top-down attention processes in older viewers. Together these findings suggest that eye-tracking methodologies can be used to observe developmental changes in online processing of video.

Overview of the Current Study

The current study recorded eye movements by infants, preschool-age children, and adults during video viewing. The video stimulus was a 19.5-min segment of *Sesame Street*, a professionally produced children's program. The goal was to determine whether there are age differences in the spatial variability of fixations across viewers (i.e., the extent to which viewers of the same age look at the same part of the screen at the same time). High consistency in the location of fixations is consistent with systematic, top-down attention processes. If comprehension processes become more important with age, older viewers should deploy visual attention systematically, thus looking at the same part of the screen at the same time. This study extended Frank et al. (2009) by using naturalistic stimuli (complete vignettes affording story arc development and integration across distinct camera shots) and by observing age groups known to vary in ability to demonstrate learning from video (infants, preschoolers, and adults).

Method

Participants

Participants were 25 one-year-olds (10 female, range 11 to 15 months), 20 four-year-olds (8 female, range 4 years 2 months to 4 years 9 months), and 17 adults (15 female, range 19 to 28 years). An additional 10 one-year-olds, 12 four-year-olds, and 9 adults were excluded from analyses because of technical problems (e.g., poor calibration) or participant-related complications (e.g., glasses interfering with eye-tracking, falling asleep). One-year-olds and four-year-olds were selected in an effort to observe children both within and well beyond the age range defined by Anderson and Pempek (2005) as demonstrating a video deficit. These authors proposed that children begin to overcome the video deficit between 18 and 30 months of age. An adult group was included for two reasons. First, adults allow for some comparison between the current study and previous research which largely focuses on adults. Second, four-year-olds have limited ability to comprehend video sequences that incorporate complex video editing techniques (Smith, Anderson, & Fischer, 1985) whereas adults provide data from experienced and competent video viewers.

Names and addresses for potential child participants in Western Massachusetts were obtained from a database of local birth records. Each family received a letter describing the project and a follow-up phone call one week later. A total of 165 families were contacted for this project with a response rate of about 40%. All children received a small gift for participating. Adult participants were recruited from the University of Massachusetts-Amherst student population in exchange for Psychology course credit.

Of the final sample of children, one was identified by a parent as Black/African-American, two were identified as Asian/Pacific Islander, and two were identified as Latino(a); the other 40 were identified as White/Caucasian. As a proxy for socio-economic status, parents reported the number of years of education completed by each parent with 12 years indicating a high-school diploma. The average number of years per family was 16.85 (range 12–26). Of the final sample of adults, three were identified as Asian/Pacific Islander and one as Latino(a); the other 13 were White/Caucasian.

Parents reported home television exposure by completing a one-week schedule to reflect times when children were typically in the room while the television was on. Of those times, parents indicted whether this was background exposure (i.e., children in the room but not watching) or foreground exposure (i.e., children watching a child-directed program). For one-year-olds, parents reported 2.54 hrs of television exposure per day on average (range 0–8.29 hrs); infants were exposed to more background television (mean 2.27 hrs per day, range 0–8.29 hrs) than foreground television (mean .27 hrs per day, range 0–2.5 hrs). Parents of four-year-olds reported 2.01 hrs of television exposure per day on average (range 0–7.71hrs); there was less reported background television exposure (mean .38 hrs per day, range 0–2.21 hrs) than foreground exposure (mean 1.63 hrs per day, range 0–5.93 hrs per day).

Video Stimuli

The calibration video consisted of small, animated images alternating between the top-left and bottom-right corners of the screen for 4 sec per image. The stimulus video was a 19.5-min segment of *Sesame Street*, a children's television program whose audience includes infants, toddlers, and preschool-age children. Two calibration checks were presented between distinct vignettes, and a third calibration check was presented at the end of the video. During the calibration checks, an animated *Sesame Street* character (Elmo) appeared for 4 seconds at 5 locations on the screen (corners and center).

Apparatus

The eye-tracking apparatus consisted of two cameras mounted on a tripod: the eye camera and the video head-tracking camera. The tripod was located between the participant and the stimulus display screen (30.50 cm from the television). The eye camera was the Applied Science Laboratories (ASL) Eye-Trac 6000 Series, a near-infrared corneal reflection system with remote pan/tilt optics. Effective accuracy is claimed to be .25° visual angle (about .25 sq cm at a distance of 50 cm) and temporal resolution is 60 Hz (one data point per video field or 60 data points per second). Fixations were averaged across 4 fields per manufacturer recommendation; that is, each data point was the average of the current gaze coordinate and the previous three. The head-tracking camera was the ASL VH2 model that uses face-recognition software to locate the viewer's head in space. The pan/tilt eye camera used information from the video head-tracker to follow the participant's head movements and minimize data loss.

Setting

The study took place in a laboratory space on the University of Massachusetts-Amherst Campus. A curtain hung from the ceiling to separate the experimenter in the control area from the viewing area (1.93 m × 2.54 m) where participants watched the video.

The stimulus display screen, a 66 cm television set (standard 330 line resolution screen, 4:3 aspect ratio), was centered along one short wall. The visible image on the display screen was 40.60 cm high and 54.60 cm wide. Given the dimensions of the screen and typical distance from the viewer (100 cm), the video image subtended approximately 23° visual angle vertically and 30.5° visual angle horizontally.

For children, the viewing area contained either a highchair (for infants) or a chair with a booster seat (for preschoolers). The chair was positioned approximately 65 cm from the eye-tracking apparatus for optimum focus and 100 cm from the television screen. Adult participants sat in a small chair designed for children so that their height and viewing angle were approximately the same as those for child participants. The mean auditory volume of the stimulus was 55 dB (averaged over 9 samples taken at 2-minute intervals at the height and distance of a typical participant) with a maximum of approximately 64 dB.

There was a chair to the right of the participant for parents and a lamp in the corner of the room to the far right of the stimulus display screen. Dark curtains hung along the walls on all sides of the viewing area and around the video display to minimize distraction and focus attention on the television screen.

In the control area on the other side of the curtain divider, the experimenter used a stimulus display computer and a DVD player to present the eye-camera calibration stimulus and the video stimulus, respectively. A switcher allowed the experimenter to toggle between the two displays as needed. A second computer was used as the eye-tracker user interface. This computer was used to capture the participant's eye with the eye camera, calibrate the video head-tracker and eye camera, and save data files of eye movements. A Sony DV-Mini recording deck was used to record the scene with overlaid eye cursor for subsequent videotape coding. An ASL Digital Frame Overlay was also used which overlaid the digital frame from the ASL Control Unit onto the scene video. This allowed the experimenter to coordinate the data file from the user interface computer with the video record from the DV-Mini tape.

Eye Camera Calibration

A two-point (top-left and bottom-right) calibration procedure was used for all participants. While this may have resulted in a calibration that was less accurate than one utilizing more than two points, this inaccuracy should not vary systematically across age groups. In other studies in the same laboratory, this procedure has resulted in equally accurate calibration across age groups (i.e., non-significant differences in horizontal and vertical variability of the samples averaged to calculate gaze location for the two calibration points). For infants, parents were asked to draw their children's attention to the target points. Older children were asked to "play a guessing game" with the experimenter by identifying the images on the screen as they appeared. Adult participants were simply asked to look at each image as it appeared on the screen.

Procedure

Upon entering the study room, the participant was seated in front of the video display screen. Children were placed in either the highchair or booster seat with appropriate safety restraints. A parent remained in the room with the child at all times. After explaining the procedure to the adult and answering any remaining questions, the parent or participant was asked to provide written informed consent. Parents were asked to refrain from directing his or her child's attention to any particular area on the screen once the session began.

After calibrating the eye-tracker, the experimenter used the switcher to present the stimulus video from the DVD player. The experimenter began recording both the data file on the user-interface computer and the video record on the DV recording deck. After the stimulus video began, the experimenter maintained data collection by ensuring that the head-tracker was enabled and that the eye camera was focused on the eye. In the event that a participant became particularly fussy during stimulus presentation, the experimenter provided him or her with a small toy in an attempt to keep him or her calm for the entire duration of the study. This was required for all but one infant and none of the older participants. Though this may have resulted in lower overall attention to the video on the part of the infants, the pattern of results did not change. Analyses on only the first 3 min of video (i.e., before any child was given a toy) did not produce results different from those found for the entire video, nor did analyses on only the first 40 sec of the video (i.e., before infants' overall attention to the screen differed from that of older viewers), thus analyses presented here reflect usable data from the entire stimulus video.

Data Reduction

Final calculations excluded fixations by any subject that were preceded by at least 500 ms of lost data which may have indicated a return from a look away from the screen. These fixations were excluded because the first fixation to the screen after a look away may be aimed indiscriminately at the center of the screen or at a random location and may not reflect processing of video content. Moreover, the only frames included in final calculations were those receiving fixations from at least 8 participants in each of the three age groups resulting in 4,215 video frames. This was a further attempt to compensate for age differences in overt attention to the screen.

Defining fixations—The first step of data reduction was to define fixations using ASL's data analysis package, Eyenal. This software uses a fixation algorithm in which the user defines fixation criteria such as minimum and maximum movement thresholds and minimum fixation duration. The resulting fixation file contains information such as start time, duration, and horizontal and vertical gaze coordinates for each fixation.

A fixation began with 6 consecutive fields during which each set of vertical and horizontal gaze coordinates were within .5° visual angle of each other. Once begun, a fixation continued until there were 12 consecutive fields of missing data or 3 consecutive fields in which the location of gaze was greater than 1° visual angle from the average coordinate for that fixation. Lastly, outliers within a fixation were excluded when calculating the location of that fixation. Any single gaze coordinate that exceeded a distance of 1.5° visual angle from the fixation average was excluded from calculation of the average coordinate for that fixation.

Bivariate contour ellipse area (BVCEA)—One set of analyses was conducted on the spatial area of the best-fit bivariate contour ellipse surrounding at least 61.33% of the data points for each age group for each frame. The equation used to calculate BVCEA for each frame, which has been used by others to quantify spatial variability in adults' fixations to video (Goldstein et al., 2007), was:

$$2k\pi\sigma_H\sigma_V(1 - p^2)$$

where σ_H is the standard deviation of the horizontal gaze coordinates, σ_V is the standard deviation of the vertical gaze coordinates, p is the product-moment correlation between horizontal and vertical values, and k is the enclosure; k is calculated as:

$$p = 1 - e^{-k}$$

where p is the proportion of points included in the ellipse. In the current study, k = .95 such that 61.33% of data points were required to be included in the best-fit ellipse. This is similar to the threshold used by Golstein and colleagues.

Clusters of fixations—Another set of analyses was based on clusters of fixations defined as fixations falling within 6° visual angle of each other. The algorithm used here to define clusters was adapted from Stelmach et al. (1991). This algorithm calculated the Cartesian distance between each pair of points for a given frame, created lists of points for each subject that were within 6° visual angle, and then created lists of pairs of points that were in conflict (i.e., those two points were within 6° visual angle of a third point but were not within 6° visual angle of each other). One subject was selected at random. For any pair of points within that subject's list that was in conflict, each subject within that pair was removed from the potential cluster list and summed squared residuals were calculated. The subject resulting in the greatest decrease in summed squared residuals was removed from the potential cluster. After resolving all conflicts in that list, the first cluster was created. The second cluster began by randomly selecting a subject not already in the first cluster, and the same algorithm was followed until all fixations were identified as a member of a cluster or not close enough to any other points to fall within a cluster (i.e., forming a cluster of one fixation). The same procedure was followed for each video frame.

Fixations to Center Following Cut—Also of interest was the extent to which viewers fixate the center of the screen following a cut to a new shot. Others have found that adult viewers center fixations following these transitions to new scenes (LeMeur, LeCallet, & Barba, 2007; Mital et al., 2010; Tosi et al., 1997; Tseng, Carmi, Cameron, Munoz, & Itti, 2009). There were 173 cuts in the video used here. Each fixation was characterized by timing (i.e., whether it was the very first fixation initiated following a cut for a subsequent fixation within that cut) and centering (i.e., whether or not the fixation fell within 2.5° visual angle from the center of the screen).

Generating Random Fixation Patterns—For every observed fixation, a matched fixation was randomly selected from the pool of all observed, usable fixations in that age group. Thus observed data for each age group could be compared to a set of fixations that were not systematically related to each other or to the content on the screen.

Results

Analyses excluded fixations that may have been the first fixation following a look away from the screen (i.e., those preceded by at least 500ms of missing data) and were limited to video frames during which at least 8 participants in each age group were fixating the screen. These were attempts to control for more frequent looks away from the screen by infants. Because the pattern of results did not differ when analyzing data from the entire 19.5-min video, the first 3 min of video only (i.e., before any child received a toy), and the first 40 sec of the video only (i.e., before infants' attention differed from that of older viewers), findings presented below represent data from the entire 19.5-min video.

General Characteristics of Fixations

Descriptive statistics for the general characteristics of fixations are presented in Table 1. Percent of the entire video during which fixations were recorded (including those preceded by 500ms missing data) increased across age groups, $F(2, 58) = 67.30$, $p < .001$, *Cohen's f* = 1.46. Post-hoc t-tests with Bonferroni correction indicated that each age group differed significantly from the others [$t(43) = 7.09$, $p < .001$ for one-year-olds versus four-year-olds, $t(40) = 11.30$, $p < .001$ for one-year-olds versus adults, and $t(35) = 4.45$, $p < .001$ for four-year-olds versus adults].

There was also a significant main effect of age for average duration of fixations, $F(2, 58) = 5.11$, $p = .009$, *Cohen's f* = .36. One-year-olds had significantly longer fixations than did four-year-olds [$t(43) = 3.13$, $p = .003$] whereas adults were intermediate and not significantly different from the other age groups [$t(39) = .71$, $p = .482$ for one-year-olds and $t(34) = 2.12$, $p = .040$ for four-year-olds]. It should be noted that the speed of the camera used in the present study is not sufficiently high to identify smooth pursuit eye movements (as opposed to fixations separated by saccades) which sometimes occur when viewing a moving target (e.g., Kremenitzer, Vaughan, Kurtzbert, & Downling, 1979; Richards & Holley, 1999). The durations of fixations reported here may be biased by averaging over smooth pursuit movements and should be interpreted with caution, but findings on the location of fixations reported below remain relatively unaffected.

Spatial Variability of Fixations

The goal of the present study was to determine to what extent television viewers look at the same place on the screen at the same time indicating employment of systematic, top-down attention processes guided by story arc and production features (e.g., cuts between shots). Figure 1 plots fixations for all viewers in each age group to each of three separate video frames as examples of frame-by-frame data. Three measures of spatial variability were used in the current study: the best-fit bivariate contour ellipse area (BVCEA; measured in ° visual angle^2), the number of clusters found (defined as fixations falling within 6° visual angle of each other), and the percent of viewers whose fixations were included in the dominant (i.e., most populated) cluster. The first dependent variable, BVCEA, measures overall scatter in the data, but it treats similarly a very disperse set of fixations and two tight clusters of fixations that are far apart (e.g., see the lower right panel of Figure 1). The other dependent variables account for this possibility in different ways. For example, the BVCEA accurately reflects decreasing scatter with age in the first row of panels in Figure 1 but may not show differences across age for the third row: one ellipse encapsulating data points for adults

would be large despite relatively low individual differences. The second dependent variable, number of clusters, would correctly reveal age differences for all frames in Figure 1; however, it would not distinguish between adult data in the second and third rows as both would form two clusters. The third dependent variable, percent of individuals fixating the dominant cluster, would build on the number of clusters in these two frames by revealing more consistency in adult data to the second frame (all but one participant fixating the girl) than in adult data to the third frame (participants equally divided between the two characters). Descriptive statistics for spatial variability are presented in Table 2 as a function of age; descriptive statistics for random fixation pattern data are also provided for comparison.

The unit of analysis was video frame. Hierarchical linear modeling (HLM) was employed because the individual video frames were not independent of each other (i.e., came from the same or similar shots). Age was a repeated measure within frame and was included at Level-1 in the form of two dummy codes representing one-year-olds and adults. Time into the shot (i.e., seconds elapsed since the most recent camera cut) varied between frames; this was centered at the first video frame following a cut and was included as a Level-2 predictor for each of the three age groups. Thus the Level-1 intercept represents four-year-olds at the start of the shot, and the Level-2 slope for the intercept reflects change over time in four-year-olds' fixations. Significant intercepts for Level-1 predictors (i.e., dummy codes for one-year-olds and adults) reflect significant differences from four-year-olds at the start of the shot, and significant slopes at Level-2 for these predictors reflect change over time in the difference from four-year-olds. See Raudenbush and Bryk (2002) for a detailed description of the HLM technique.

Bivariate Contour Ellipse Area (BVCEA)—Figure 2 plots fitted lines from the HLM model for BVCEA as a function of time into the shot and age group. Not surprisingly, four-year-olds' average BVCEA was significantly different from zero at the start of the video as indicated by a significant Level-1 intercept [$B = 45.44$ (SE = 2.22), $t(323) = 20.48$, $p < .001$]. At the start of the shot, BVCEA increased significantly with age as indicated by a significantly negative intercept for one-year-olds and a significantly positive intercept for adults [$B = 23.57$ (SE = 3.47), $t(323) = 6.78$, $p < .001$, and $B = −15.20$ (SE = 1.83), $t(323) = −8.29$, $p < .001$, for one-year-olds and adults, respectively]. Moreover, age differences decreased as a function of time into the shot. Specifically, four-year-olds' average BVCEA increased over time and one-year-olds' BVCEA decreased over time [$B = −.76$ (SE = .38), $t(323) = −2.01$, $p = .045$, and $B = .47$ (SE = .21), $t(323) = 2.30$, $p = .022$, for one-year-olds and four-year-olds, respectively]. Adults' average BVCEA increased over time at the same rate as did that of four-year-olds as indicated by a non-significant Level-2 slope [$B = .06$ (SE = .18), $t(323) = .31$, $p = .759$]. It is important to note that although infants' fixations are significantly more scattered than those of older viewers (at least at the beginning of individual shots), they are still much less variable than a sample from a random fixation pattern (see Table 2).

Number of Clusters Found—Figure 3 plots fitted lines from the HLM model for number of clusters per frame as a function of time into the shot and age group. The pattern of results for number of clusters paralleled that of BVCEA. At the start of the shot, the number of clusters found for four-year-olds was significantly greater than zero [$B = 2.41$ (SE = .05), $t(323) = 47.06$, $p < .001$]. Variability decreased with age insofar as infants' fixations formed significantly more clusters than did four-year-olds, whereas adults' fixations formed significantly fewer clusters [$B = .29$ (SE = .06), $t(323) = 4.59$, $p < .001$, and $B = −.38$ (SE = .05), $t(323) = −6.95$, $p < .001$, for one-year-olds and adults, respectively]. However, infants' fixations produced fewer clusters on average than did random fixation patterns with the same number of fixations (see Table 2), and again age differences decreased over time into

the shot insofar as one-year-olds' fixations formed fewer clusters and four-year-olds' fixations formed more clusters as time increased [$B = -.02$ ($SE = .006$), $t(323) = -3.17$, $p = .002$, and $B = .01$ ($SE = .005$), $t(323) = 2.97$, $p = .003$, for one-year-olds and four-year-olds, respectively]. Adults' average number of clusters increased over time at the same rate as did that of four-year-olds [$B = .003$ ($SE = .005$), $t(323) = .68$, $p = .500$].

Percent of Subjects in Dominant Cluster—The final dependent variable was the percent of subjects per age group in the dominant (i.e., most populated) cluster. Figure 4 shows fitted lines from the HLM model as a function of time into the shot and age group. The pattern of results paralleled that of the other two dependent variables. Naturally four-year-olds average was significantly greater than zero [$B = 77.40$ ($SE = 1.05$), $t(323) = 73.60$, $p < .001$]. The representativeness of the dominant cluster increased with age such that one-year-olds' average at the start of the shot was significantly less than that of four-year-olds, and adults' average was significantly more than that of four-year-olds [$B = -11.02$ ($SE = 1.31$), $t(323) = -8.40$, $p < .001$, and $B = 7.52$ ($SE = 1.00$), $t(323) = 7.53$, $p < .001$, for one-year-olds and adults, respectively]. Once again, even infants' show much more consistency than a random fixation pattern (see Table 2), and age differences decreased as a function of time into the shot insofar as four-year-olds' dominant cluster became less representative and one-year-olds' dominant cluster became more representative [$B = .52$ ($SE = .12$), $t(323) = 4.46$, $p < .001$, and $B = -35$ ($SE = .09$), $t(323) = -4.18$, $p < .001$, for one-year-olds and four-year-olds, respectively]. Adults' average changed at the same rate as that of four-year-olds [$B = .08$ ($SE = .10$), $t(323) = .81$, $p = .421$).

Fixating the Center of the Screen Following a Cut

The final dependent variable of interest was the probability of fixating the center of the screen (within a radius of 2.5° visual angle) immediately following a cut to a new scene. Mean probability of fixating the center of the screen as a function of timing (first fixation following a cut, subsequent fixation in the shot) and age is presented in Figure 5. Viewers in the two older groups were significantly more likely to fixate the center of the screen immediately following a cut than subsequently in the shot [$t(19) = 7.01$, $p < .001$, and $t(16) = 5.16$, $p < .001$, for four-year-olds and adults, respectively]. One-year-olds were equally likely to fixate the center of the screen regardless of timing [$t(24) = .56$, $p = .581$].

Discussion

The goal of this study was to describe spatial variability of fixations during video viewing across three age groups. With increasing age, viewers tended to fixate the same part of the screen as their peers suggesting greater employment of systematic, top-down attention processes. This supports the common observation that video comprehension increases with age (see Kirkorian & Anderson, 2008) and is consistent with recent findings on infants' and adults' fixations to 4-second animated segments (Frank et al., 2009) and research with adults watching longer videos (Dorr et al., 2010; Golstein et al., 2007; Mital et al., 2010; Stelmach et al., 1991; Tosi et al., 1997). It is important to note that less spatial variability with age is representative of the majority of video frames. For example, when considering BVCEA for the 4,215 video frames during which at least 8 participants within each age group were fixating the screen (excluding fixations preceded by 500ms or more of missing data), one-year-olds' fixations were quantitatively more variable than were four-year-olds' and adults' for 65% and 77% of frames, respectively. Four-year-olds' fixations were more variable than were adults' for 70% of frames. Thus the trend for spatial variability to decrease with age is generalizable to the majority of video frames and is not limited to a specific type of shot (e.g., close-up of a face).

These changes in eye movements with age likely reflect a common and consistent understanding of what is being viewed, insofar as viewers will tend to look at the same things at the same time when they are able, as intended by the producer of the video, to comprehend the action, dialogue, and narration. The relatively more variable fixation patterns of one-year-olds, in contrast, likely reflect little sequential or linguistic comprehension of this excerpt from *Sesame Street*. It is worth noting that if infants' visual attention is organized on the basis of perceptually salient features (Frank et al., 2009; Huston & Wright, 1983; Ruff & Rothbart, 1996), then there appears to be little moment to moment consistency across children in control of fixations by those features. Nonetheless, infants' fixations were much more consistent than a random fixation pattern. Infants' visual attention may be less rule-governed than that of older viewers, but their fixations are far from randomly distributed.

Moreover, despite overall decreases in spatial variability with age, these age differences did change over the course of time into a shot. Specifically, age differences were greatest immediately following a cut to a new scene. As time progressed, infants became more similar to each other in their fixations. This finding suggests that processing transitions to new shots may be particularly difficult for infant viewers. This supports research demonstrating that children's ability to comprehend video editing techniques increases with age (Smith et al., 1985), that very young children's attention may be insensitive to such transitions (Pempek, Kirkorian, Stevens, Lund, Richards, & Anderson, 2010), and that experienced video viewers readily process these transitions without conscious effort (Smith & Henderson, 2008). Moreover, fixations by older viewers were most consistent immediately following a cut to a new scene, and variability increased as time progressed into the shot. This is apparently due to older viewers' tendency to fixate the center of the screen immediately following a cut, perhaps using such editing techniques strategically to guide attention to important content. Others have found that adults often center their gaze following cuts in video, likely in response to producers' tendency to film with important content in the center of the screen (LeMeur et al., 2007; Mital et al., 2010; Tosi et al., 1997; Tseng et al., 2009).

Together these findings are consistent with the hypothesis that there is a video deficit in children under two years of age (Anderson & Pempek, 2005) insofar as infants appear to process video in a less systematic way than do older viewers. Among other things, in order to comprehend standard video productions, infants must acquire the ability to inhibit attention to irrelevant formal features, to recognize informative visual and auditory features, and to integrate information appropriately across shot boundaries (e.g., Anderson, Fite, Petrovich, & Hirsch, 2006; Pempek et al., 2010; Smith et al., 1985). As indexed by their eye movements, they have apparently not accomplished this by one year of age, or at the very least they may require several seconds to orient to a new scene. This study demonstrates that eye tracking is a useful technique for studying online processing of video, particularly with infants. Future studies of this sort may have implications for the extent to which video can be comprehensible to, and thus educational for, very young children.

Acknowledgments

References

American Academy of Pediatrics, Committee on Public Education. Media education. Pediatrics. 1999; 104:341–342. [PubMed: 10429023]

American Academy of Pediatrics, Committee on Public Education. Media education. Pediatrics. 2010; 126:1012–1017.

Anderson DR, Fite KV, Petrovich N, Hirsch J. Cortical activation while watching video montage: An fMRI study. Media Psychology. 2006; 8:7–24.

Anderson, DR.; Lorch, EP. Looking at television: Action or reaction?. In: Bryant, J.; Anderson, DR., editors. Children's understanding of TV: Research on attention and comprehension. New York: Academic Press; 1983. p. 1-34.

Anderson DR, Pempek TA. Television and very young children. American Behavioral Scientist. 2005; 48:505–522.

Barr R, Hayne H. Developmental changes in imitation from television during infancy. Child Development. 1999; 70:1067–1081. [PubMed: 10546335]

Dorr M, Martinetz T, Gegenfurtner K, Barth E. Variability of eye movements when viewing dynamic natural scenes. Journal of Vision. 2010; 10(10):1–17.

Flagg, BN. Children and television: Effects of stimulus repetition on eye activity. In: Senders, JW.; Fisher, DF.; Monty, RA., editors. Eye movements and the higher psychological functions. Hillsdale, NJ: Erlbaum; 1978. p. 279-291.

Frank MC, Vul E, Johnson SP. Development of infants' attention to faces during the first year. Cognition. 2009; 110:160–170. [PubMed: 19114280]

Garrison, MM.; Christakis, DA. A teacher in the living room?: Educational media for babies, toddlers, and preschoolers. Menlo Park, CA: The Henry J. Kaiser Family Foundation; 2005.

Goldstein RB, Woods RL, Peli E. Where people look when watching movies: Do all viewers look at the same place? Computers in Biology and Medicine. 2007; 37:957–964. [PubMed: 17010963]

Hayne H, Herbert J, Simcock G. Imitation from television by 24- and 30-month-olds. Developmental Science. 2003; 6(3):254–261.

Huston, AC.; Wright, JC. Children's processing of television: The informative functions of formal features. In: Bryant, J.; Anderson, DR., editors. Children's understanding of television: Research on attention and comprehension. New York: Academic Press, Inc; 1983. p. 35-68.

Kirkorian, HL.; Anderson, DR. Learning from educational media. In: Calvert, SL.; Wilson, BJ., editors. The Blackwell Handbook of Children, Media, and Development. Boston, MA: Blackwell; 2008. p. 319-360.

Kremenitzer JP, Vaughan HG, Kurtzberg D, Dowling K. Smooth-pursuit eye movements in the newborn infant. Child Development. 1979; 50:442–448. [PubMed: 487885]

Le Meur O, Le Callet P, Barba D. Predicting visual fixations on video based on low-level visual features. Vision Research. 2007; 47:2483–2498. [PubMed: 17688904]

Mackworth NH, Bruner JS. How adults and children search and recognize pictures. Human Development. 1970; 13:149–177. [PubMed: 5469121]

Mital PK, Smith TJ, Hill R, Henderson JM. Clustering of gaze during dynamic scene viewing is predicted by motion. Cognitive Computation. 2010; 3:2–24.

Pempek TA, Kirkorian HL, Richards JE, Anderson DR, Lund AF, Stevens M. Video comprehensibility and attention in very young children. Developmental Psychology. 2010; 46:1283–1293. [PubMed: 20822238]

Raudenbush, SW.; Bryk, AS. Hierarchical linear models: Applications and data analysis methods. 2nd ed.. Thousand Oaks, CA: Sage Publications; 2002.

Richards JE, Holley FB. Infant attention and the development of smooth pursuit tracking. Developmental Psychology. 1999; 35:856–867. [PubMed: 10380875]

Rideout, VJ.; Hamel, E. The media family: Electronic media in the lives of infants, toddlers, preschoolers, and their parents. Menlo Park, CA: The Henry J. Kaiser Family Foundation; 2006.

Ruff, HA.; Rothbart, MK. Attention in early development: Themes and variations. New York: Oxford University Press; 1996.

Singer, JL. The entertainment function of television. Hillsdale, NJ: Lawrence Erlbaum; 1980. The power and limits of television: A cognitive-affective analyses.

Smith R, Anderson DR, Fischer C. Young children's comprehension of montage. Child Development. 1985; 56:962–971. [PubMed: 4042754]

Smith TJ, Henderson JM. Edit blindness: The relationship between attention and global change blindness in dynamic scenes. Journal of Eye Movement Research. 2008; 2:1–17.

Stelmach L, Tam WJ, Hearty P. Static and dynamic spatial resolution in image coding: An investigation of eye movements. Human Vision, Visual Processing, and Digital Display II. 1991; 1453:147–152.

Takahashi N. Developmental changes of interests to animated stories in toddlers measured by eye movement while watching them. International Journal of Psychology in the Orient. 1991; 34:63–68.

Tosi V, Mecacci L, Pasquali E. Scanning eye movements made when viewing film: Preliminary observations. International Journal of Neuroscience. 1997; 92:47–52. [PubMed: 9522254]

Troseth GL. Is it life or is it Memorez? Video as a representation of reality. Developmental Review. 2010; 30:155–175.

Tseng PH, Carmi R, Cameron IGM, Munoz DP, Itti L. Quantifying centre bias of observers in free viewing of dynamic natural scenes. Journal of Vision. 2009; 9(7):1–16.

Whiteside JA. Eye movements of children, adults, and elderly persons during inspection of dot patterns. Journal of Experimental Child Psychology. 1974; 18:313–332. [PubMed: 4429624]
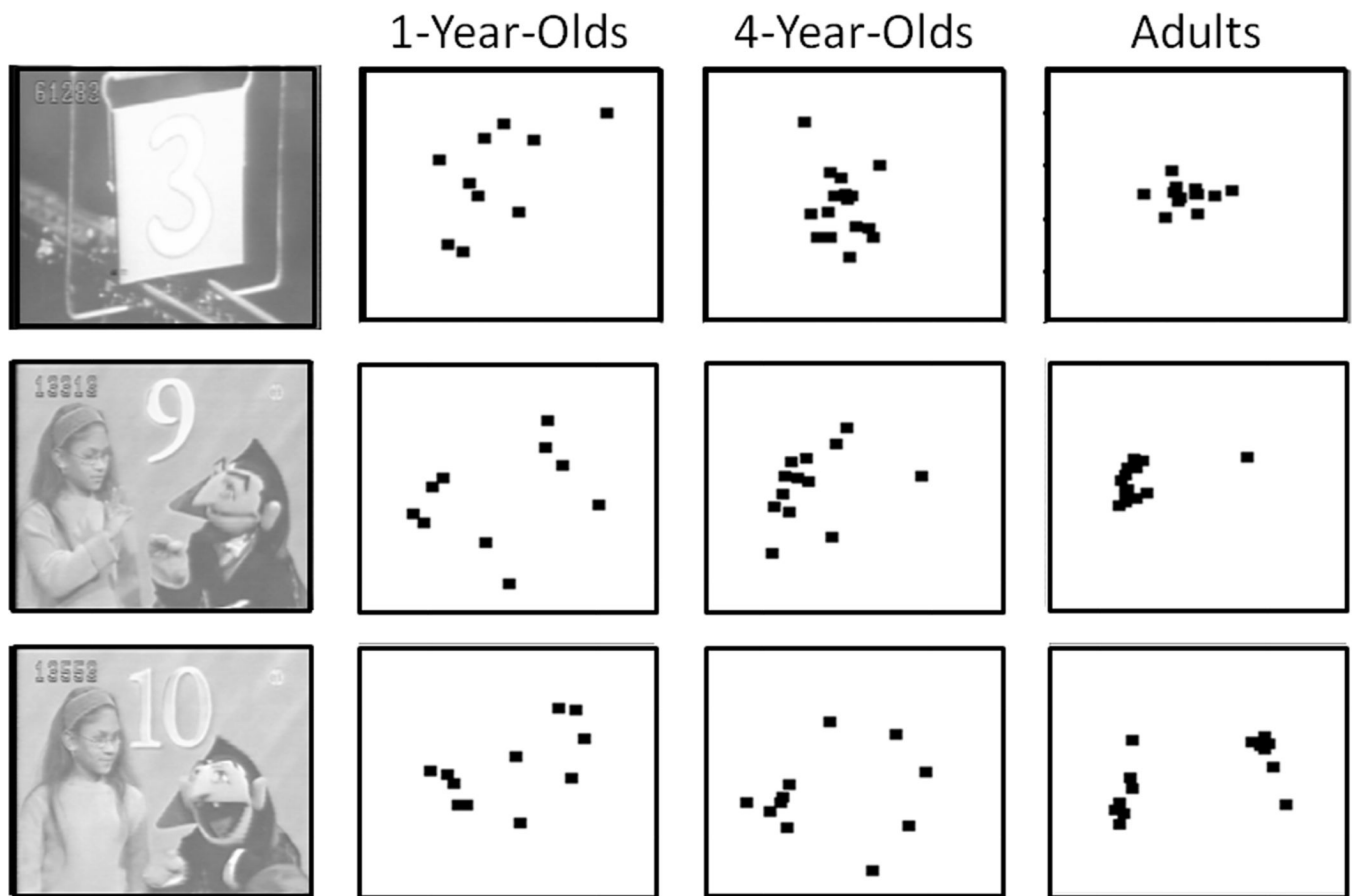
Figure 1.
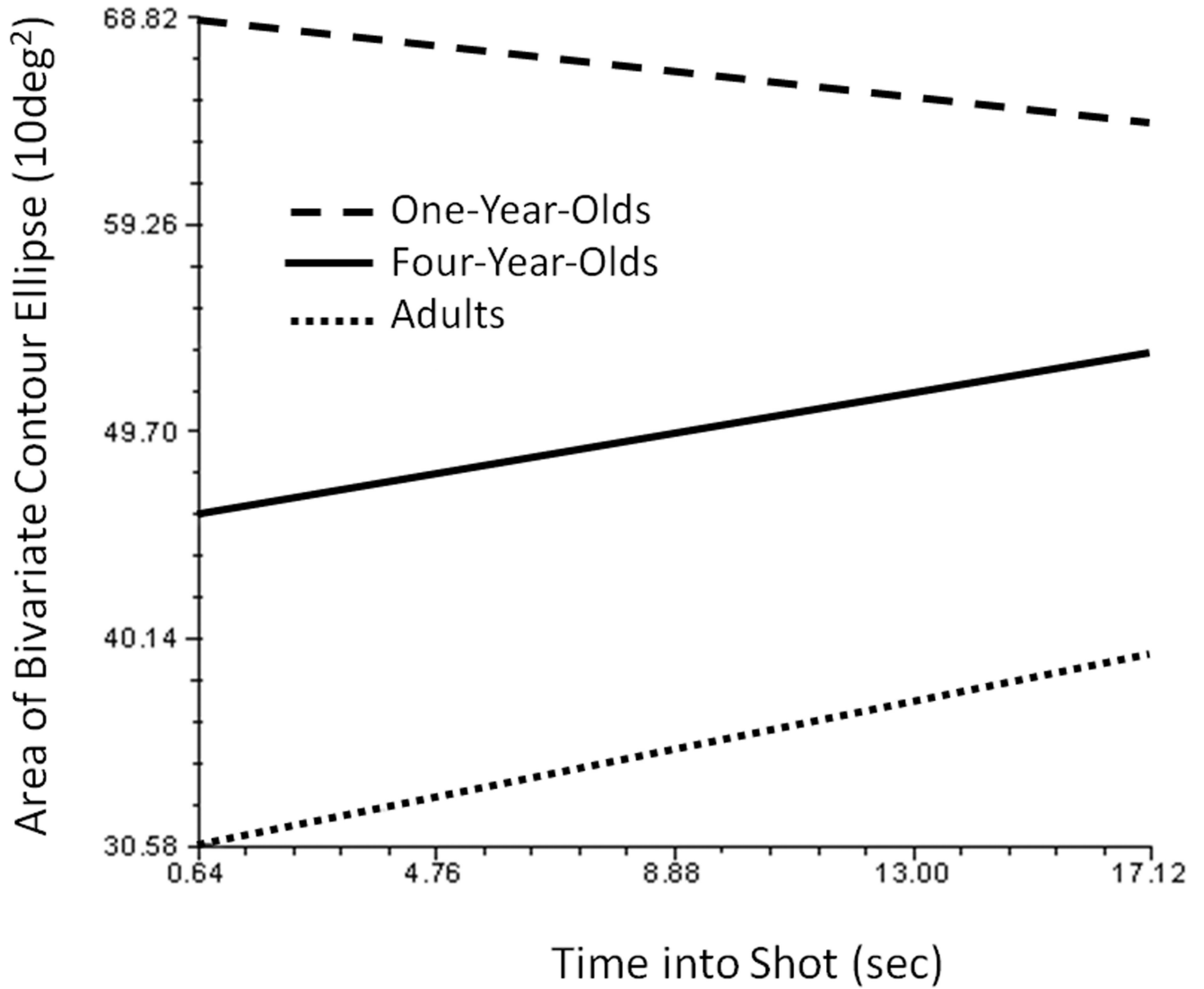Fixations for all participants in each age group to three different video frames.

Figure 2.
Mean best-fit bivariate contour ellipse area (BVCEA; 10° visual angle^2) as a function of
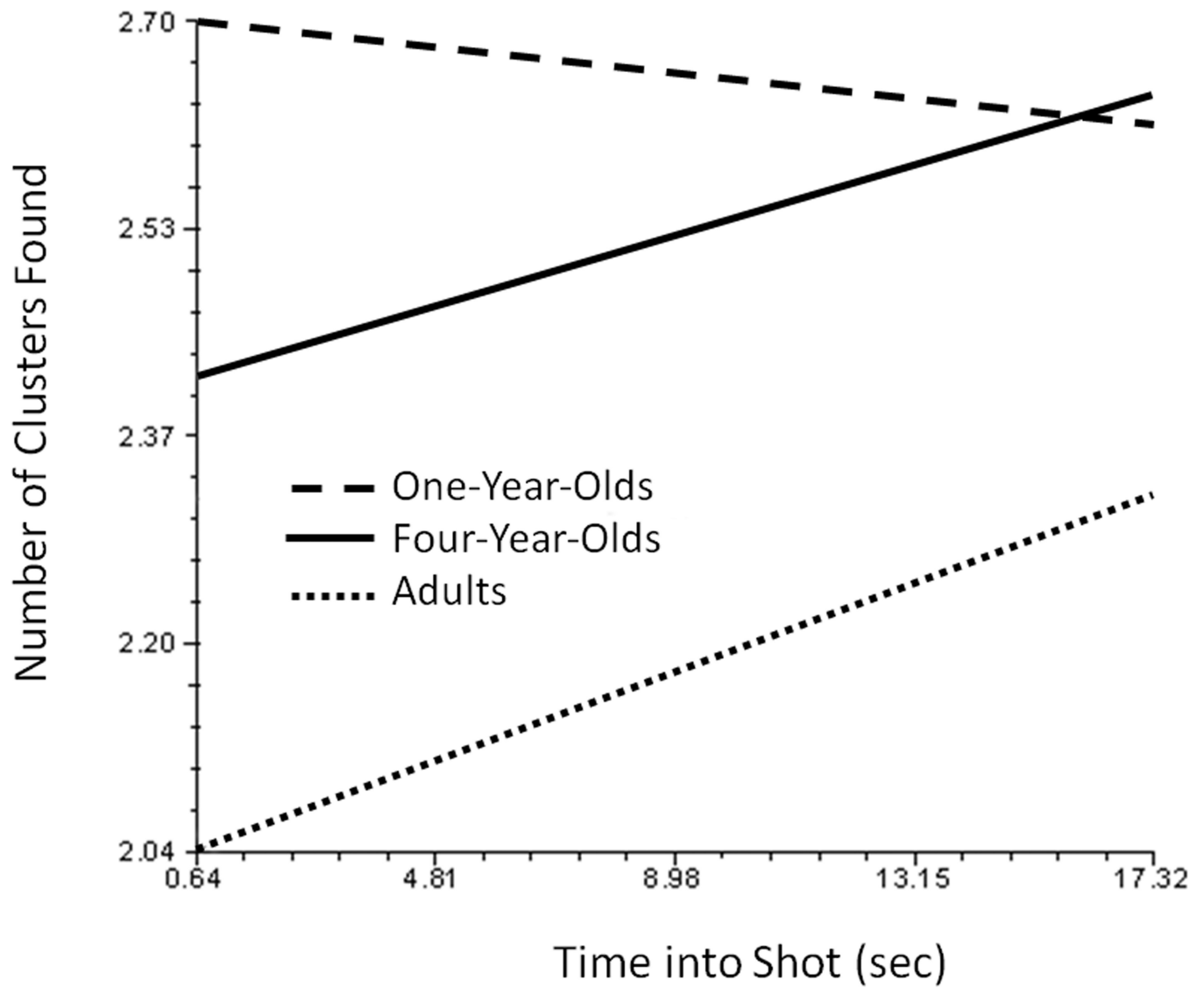time since the start of the shot and age group.

Figure 3.
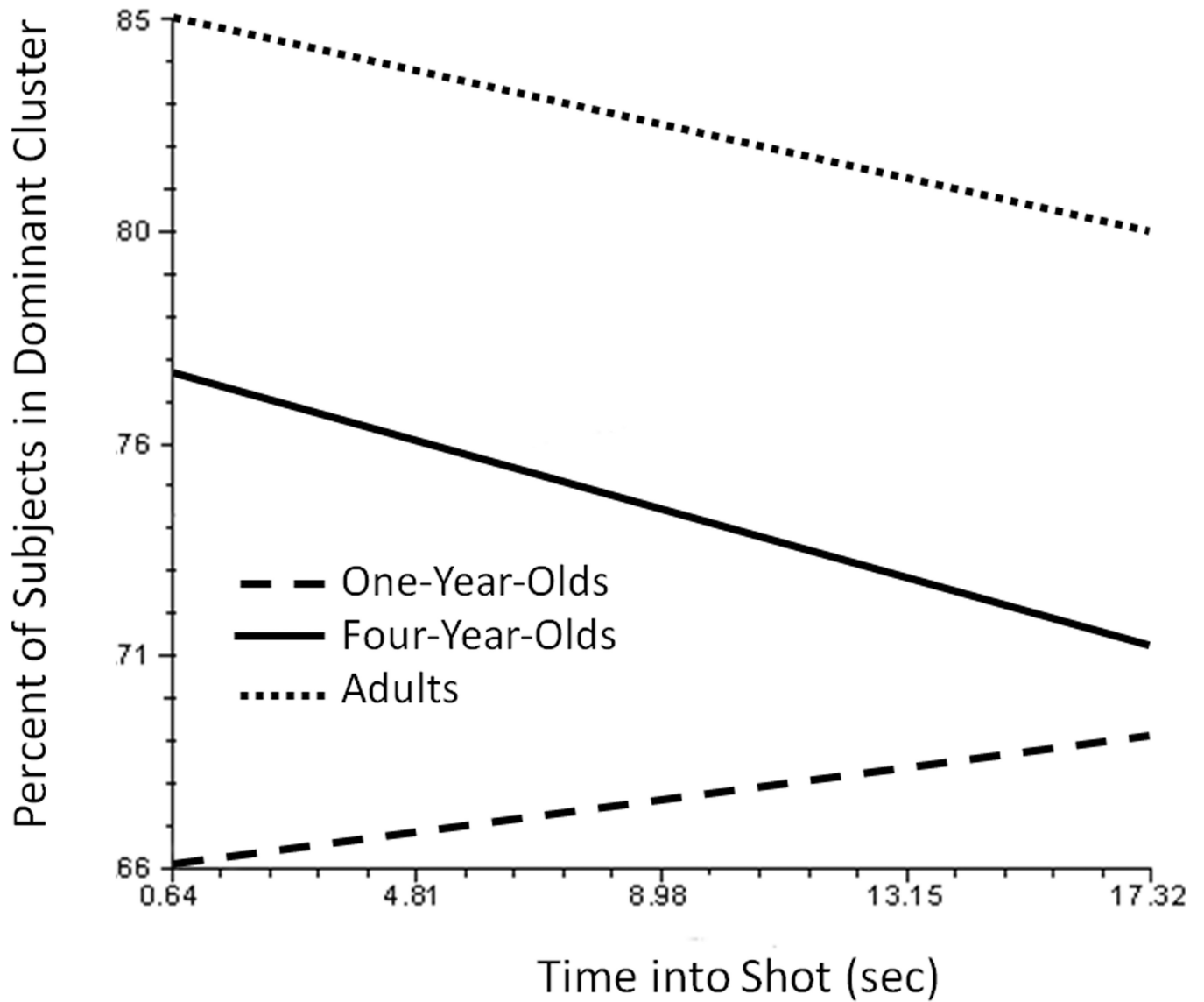Mean number of clusters found per frame as a function of time since the start of the shot and age group.

Figure 4.
Mean percent of viewers whose fixations landed in the dominant cluster as a function of time since the start of the shot and age group.
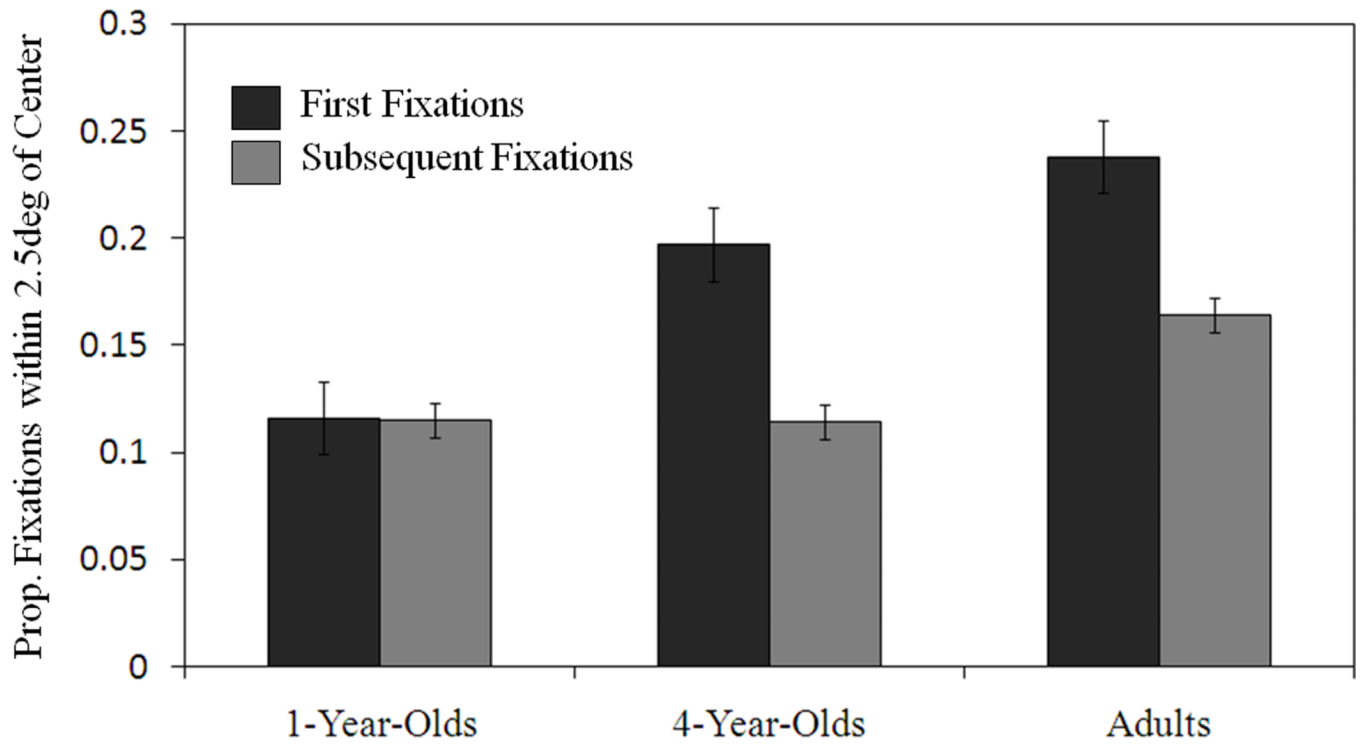
Figure 5.
Mean proportion of fixations that landed within 2.5° visual angle of the center of the screen as a function of timing (first fixation following a cut, subsequent fixation) and age. Bars represent +/− one standard error.

Table 1

Means (Standard Errors) for General Characteristics of Fixations as a Function of Age

	1 year	4 Years	Adult	Average
Percent Fixation	20.36 (2.76)	49.70 (3.09)	70.28 (3.45)	46.78 (1.80)
Fixation Duration (ms)	402.85 (12.55)	343.96 (14.03)	388.59 (15.69)	378.47 (8.17)

Note. Dependent variables are percent of the entire video during which accurate eye movement data was collected and average fixation duration (milliseconds) for each age group.

Table 2

Means (Standard Errors) for Three Measures of Spatial Variability of Fixations as a Function of Age for Observed and Randomly Grouped Data

	1 Year	4 Years	Adult
BVCEA			
Observed	61.35 (.62)	46.22 (.54)	32.07 (.40)
Random	90.17 (.54)	86.17 (.50)	72.16 (.40)
Number Clusters			
Observed	2.61 (.01)	2.46 (.01)	2.13 (.01)
Random	3.26 (.01)	3.39 (.01)	3.33 (.01)
Biggest Cluster			
Observed	69.18 (.26)	76.06 (.25)	83.63 (.23)
Random	57.10 (.20)	58.60 (.21)	62.53 (.20)

Note. Dependent variables are the best-fit bivariate contour ellipse area (BVCEA; $10°$ visual angle^2), the number of clusters found, and the percent of participants whose fixation landed in the dominant cluster. Randomly grouped data were generated by randomly selecting from observed fixations within each age group.