Ambient and focal visual processing of naturalistic activity

Michelle L. Eisenberg   Department of Psychology, Washington University in St. Louis, St. Louis, MO, USA

Jeffrey M. Zacks   Department of Psychology, Washington University in St. Louis, St. Louis, MO, USA

When people inspect a picture, they progress through two distinct phases of visual processing: an *ambient*, or exploratory, phase that emphasizes input from peripheral vision and rapid acquisition of low-frequency information, followed by a *focal* phase that emphasizes central vision, salient objects, and high-frequency information. Does this qualitative shift occur during dynamic scene viewing? If so, when? One possibility is that shifts to exploratory processing are triggered at subjective event boundaries. This shift would be adaptive, because event boundaries typically occur when activity features change and when activity becomes unpredictable. Here, we used a perceptual event segmentation task, in which people identified boundaries between meaningful units of activity, to test this hypothesis. In two studies, an eye tracker recorded eye movements and pupil size while participants first watched movies of actors engaged in everyday activities and then segmented them into meaningful events. Saccade amplitudes and fixation durations during the initial viewings suggest that event boundaries function much like the onset of a new picture during static picture presentation: Viewers initiate an ambient processing phase and then progress to focal viewing as the event progresses. These studies suggest that this shift in processing mode could play a role in the formation of mental representations of the current environment.

Introduction

Visual processing of a static image is characterized by two stages: ambient and focal processing (Trevarthen, 1968). Ambient visual processing typically occurs during the first few seconds of viewing and involves short fixations and long saccades as people extract information from peripheral vision. After a few seconds, viewers shift to focal processing, which involves longer fixations and shorter saccades as viewers extract more detailed information from central vision (Pannasch, Helmert, Roth, Herbold, & Walter, 2008; Unema, Pannasch, Joos, & Velichkovsky, 2005). Although this shift in visual processing modes has been studied extensively while people view static visual images, there has been much less research into whether people shift back and forth from ambient to focal processing while watching dynamic visual stimuli such as movies. We therefore conducted two studies to determine whether the same processes that occur during viewing of static images are maintained during viewing of dynamic stimuli.

Evidence that people shift from ambient to focal visual processing during the course of viewing a static picture has accumulated over the past century. In 1935 Buswell (as cited in Unema et al., 2005) found that the duration of fixations increased over the time course of viewing a picture, and Antes (1974) found that fixation duration increased and saccade amplitude decreased over time as participants viewed a static image. Later, Irwin and Zelinsky (2002) found that fixation duration increased over 15 s of viewing a static image. Unema et al. (2005) found that during the first few seconds of viewing a static image of a room, participants displayed larger saccade amplitudes and shorter fixation durations, whereas after a few seconds of viewing, participants displayed smaller saccade amplitudes and longer fixation durations. Unema et al. (2005) suggested that the larger saccade amplitudes and shorter fixation durations during the early viewing period represented ambient processing and that the smaller saccade amplitudes and longer fixation durations during the later viewing period represented focal processing. Providing further support for the distinction between ambient and focal visual processing, Velichkovsky, Joos, Helmert, and Pannasch (2005) found that viewers were more likely to remember objects when fixation duration was long and when fixation duration was short but the subsequent saccade amplitude was short, compared to when fixation duration was short and the subsequent saccade

amplitude was long. They suggested that long fixations and small saccade amplitudes represent focal processing of detailed information about the objects, allowing for better subsequent recognition of these objects. Furthermore, in a series of four studies using naturalistic scenes, participants displayed clear evidence of a shift from ambient to focal processing, with longer saccades and shorter fixations during the early viewing period and shorter saccades and longer fixations during the later viewing period (Pannasch et al., 2008).

The limited evidence regarding ambient and focal processing in dynamic stimuli suggests that viewing patterns display a similar shift during the first few seconds of viewing. Velichkovsky, Rothert, Kopf, Dornhöfer, and Joos (2002) studied patterns of eye movements while participants engaged in a driving simulation task. They found that when participants detected a hazard, fixation duration increased, suggesting that participants engaged in more focal processing at the times when detailed information was necessary to avoid an accident. Providing further evidence that such shifts occur during viewing of dynamic stimuli, Smith and Mital (2013) compared viewing of dynamic to static scenes and found that participants displayed shorter fixations and larger saccade amplitudes during the first second of viewing of both types of scenes and then shifted to longer fixations and shorter saccade amplitudes during the remainder of viewing.

In contrast, Krejtz, Szarkowska, Krejtz, Walczak, and Duchowski (2012) found the opposite viewing pattern in children watching animated educational videos. They found that in a 4-s period during which a specific concept was presented, children displayed longer fixations and shorter saccades during the first two seconds of viewing and then shifted to shorter fixations and longer saccades during the third second of viewing. The authors suggested that this shift represents a shift from a focal to an ambient processing mode while viewing the material related to the concept. This pattern is the opposite of the one we have described previously. One possibility is that because this segment was taken from the middle of their educational video, it was not necessary for the children to engage in ambient exploratory processing when the 4-s period began. In addition, the authors focused on only one time period during the movie rather than investigating whether the children shifted back and forth between ambient and focal processing throughout the movie.

When studying shifts in processing mode during viewing of static images, the onset of processing is easy to identify: It is the time when the stimulus is first presented. Given this reference point, it is possible to demarcate early and late viewing periods in which ambient and focal processing are likely to occur.

However, it is more challenging to identify early and late viewing periods in dynamic stimuli such as movies or live action. Although the beginning of a movie provides one starting point, there are not necessarily other clear points during which an ambient mode of processing would be expected. Several studies have examined viewing patterns around edits, at which one continuous run of the camera ends and another begins, as shifts were hypothesized to appear at these locations. Multiple studies have found that within the first few hundred milliseconds of an edit, gaze shifts toward the center of the screen and then quickly shifts to an examination of the periphery (Carmi & Itti, 2006a, 2006b; Mital, Smith, Hill, & Henderson, 2011), suggesting a shift to ambient processing after the edit. In addition, Pannasch (2014) found that edits during movies of a single actor walking through indoor locations shifted processing from an ambient to a focal mode. However, edits do not occur during naturalistic viewing of live action. Therefore, a different approach to examining shifts between ambient and focal processing is necessary to provide further insight into whether shifts occur during continuous dynamic scenes.

One way to identify points in movies at which viewers might shift between focal and ambient processing is to assess *event segmentation*. Event segmentation is the psychological process of parsing an ongoing stream of behavior into meaningful chunks, or events. It can be measured with a unitization task, in which participants are instructed to press a button every time they believe one meaningful unit of activity has ended and another has begun (Newtson, 1973). Previous research has found that the locations of these button presses are reliable within and across participants (Zacks, Speer, Vettel, & Jacoby, 2006). According to *Event Segmentation Theory* (Zacks, Speer, Swallow, Braver, & Reynolds, 2007), people experience event boundaries as a side effect of predictive processing during ongoing perception. This account proposes that perception and comprehension systems continuously make predictions about the course of events on the time-scale of a few seconds, and monitor the accuracy of those predictions. Prediction is abetted by a working memory representation of the current event, called an *event model*. When features of the environment change, prediction error spikes and the event model is updated. For example, when watching someone make breakfast, people often identify an event boundary when the actor switches from making toast to making eggs, as this change in activity requires people to update their working memory representation of the actor's goal. Because people must update their working memory representation at event boundaries, it is likely that they engage in exploratory eye movements at these points—in other words, that they switch to an ambient processing mode in order to capture as much

information from the new event as possible. After this rapid visual exploration, people likely switch to focal processing of the details of the environment in order to make valid predictions about future activity. Focal processing then likely continues until the next event boundary, at which time this updating process starts anew.

Smith, Whitwell, and Lee (2006) found preliminary evidence for shifts from focal to ambient processing around event boundaries. They found a decrease in saccade frequency a few hundred milliseconds before event boundaries followed by an increase about 100 milliseconds after event boundaries. However, the authors did not provide a time course of the shift throughout multiple seconds preceding and following event boundaries.

The two studies discussed below therefore utilized the event segmentation task to investigate the multi-second time course of shifts in processing mode while participants watched movies of actors engaging in neutral, everyday activities. Participants were hypothesized to shift from focal to ambient processing for the few seconds following event boundaries and then return to focal processing. Specifically, we hypothesized that during viewing of naturalistic movie stimuli, saccade amplitude would increase and fixation duration would decrease during the few seconds following event boundaries and then would return to baseline. In addition, because pupil size has been found to increase during exploratory processing (Jepma & Nieuwenhuis, 2011) and because a preliminary study found effects of event boundaries on pupil diameter (Swallow & Zacks, 2006), we hypothesized that pupil size would increase at event boundaries.

It is important to note that because the event segmentation task requires participants to press a button to identify event boundaries, it is likely that the processes involved in button pressing would contaminate any measures of saccade amplitude and fixation duration obtained during this task. However, there is evidence that people engage in event segmentation whenever they view dynamic stimuli, even when they are completely naïve to the event segmentation task. For example, Zacks et al. (2001) found brain activity during passive viewing of naturalistic movie stimuli that was time-locked to the event boundaries the same participants identified during a second viewing of the stimuli. Therefore, in the studies reported here, we measured saccade amplitude, fixation duration, and pupil size during an initial passive viewing phase. We then asked participants to segment the movies into meaningful events, and time-locked the oculomotor measures to the event boundaries each participant identified. We hypothesized that, during passive viewing, shifts from focal to ambient processing would be time-locked to event boundaries.

Materials and methods

Participants

For the first study, participants were recruited from the Washington University subject pool. Participants were given either course credit or $10 per hour for their time. Thirty-two participants completed the study, but four were dropped from analysis due to inability to calibrate eye tracker (two), self reported lazy eye on side being tracked (one), and participant's decision to withdraw partway through the session (one). Data from 28 participants were included in the analyses reported here (age range: 18–25; mean age: 20.6; 50% male, 50% female). For the second study, participants were recruited from the Volunteer for Health participant registry, which is maintained by the Washington University School of Medicine. Thirty-two participants completed the study, but seven participants were dropped from analysis because of failure to follow instructions (one), difficulty obtaining accurate measurements using the eye tracker (five), and technical error leading to loss of data (one). Data from 25 participants were included in the analyses for this study (age range: 22–50; mean age: 34; 32% male, 68% female). Both studies received approval from the Washington University Human Research Protection Office and were conducted in accordance with the Declaration of Helsinki.

Materials

Three movies of actors performing everyday events were used in each study. In the first study, participants watched movies of an actor making copies to put together a binder (349 s), sweeping the floor (329 s), and changing a car tire (342 s). In the second study, participants viewed movies of an actor making breakfast (329 s), preparing for a party (376 s), and planting plants (354 s). The movies were filmed from a fixed, head-height perspective, with no pan or zoom. See Figure 1 for example frames from these movies.

Behavioral tasks and oculometric measures

In each study, participants first passively watched three movies of an actor completing an everyday activity. After participants finished passively watching all of these movies, they watched each movie two additional times while completing the event segmentation task. For this task, they were instructed to push a button whenever they believed one natural and meaningful unit of activity ended and another began.

Figure 1. Example frames from movies used in the first study (Top row) and the second study (Bottom row).

Participants segmented each movie at a coarse grain ("push the button whenever you believe that a large meaningful unit of activity has ended") and at a fine grain ("push the button whenever you believe that a small meaningful unit of activity has ended"). Although participants were not given any examples of how to perform this task, an illustrative example of a coarse unit would be "put the new tire on the car," and examples of fine units would be "screw in first nut" and "screw in second nut." The order of fine versus coarse segmentation was counterbalanced across participants, and movie order was fixed across participants. Participants were not given any information about the event segmentation task, nor were they told that they would later perform the event segmentation task, until after they completed the passive viewing task. This precaution ensured that they would not covertly make button press decisions during passive viewing.

Fixation duration, saccade amplitude, and pupil size were obtained from the right eye using an infrared camera (EyeLink 1000; SR Research Ltd., Mississauga, ON, Canada) that sampled at 1000 Hz. The camera was mounted on the SR Research Desktop Mount. A chin/forehead rest was used to minimize head motion during the tasks. The camera was positioned 52 cm from the top of the forehead rest. Stimuli were presented on a 19-in. (74 cm) monitor (1440 × 900 resolution, viewing distance of 58 cm from the forehead rest, viewing angle of 38.6°). The oculomotor measures were obtained during passive viewing as well as during the fine and coarse segmentation tasks. Data Viewer software (SR Research Ltd., Mississauga, ON, Canada) was used to export the data into text files which were then exported into R (R Core Team, 2014) for analysis. Raw pupil data was extracted from Data Viewer sample reports, which resulted in pupil measurements for each millisecond of data collection. Fixation duration and saccade amplitude were extracted from Data Viewer fixation reports, which included the fixation duration of each fixation and the saccade amplitude of the saccade immediately following each fixation.

Data analysis

All analyses reported here were conducted after data collection from both studies was complete. Fixations were identified using SR-Research's algorithm (Eyelink User Manual, 2010), and default values were not modified. Fixation durations were trimmed to remove all fixations that were less than 40 ms. In addition, fixations less than 100 ms were removed if they occurred on either side of a blink. This removed 6.5% of the fixations for Study 1 and 3.7% of the fixations for Study 2. This left 163,042 fixations in the analyses for Study 1 and 175,951 fixations in the analyses for Study 2. Because blinks resulted in missing raw pupil data, a linear interpolation algorithm was used to interpolate across missing raw sample pupil data due to blinks. Stretches of missing data longer than 600 ms were not interpolated, as these long periods of missing data were not likely to represent typical blinks and because linear interpolation was less likely to accurately capture pupil dynamics during these long stretches of missing data. Saccades were also identified using SR-Research's algorithm (Eyelink User Manual, 2010), and only saccades that did not contain blinks were used to

calculate saccade amplitude. Saccade amplitude was calculated by determining the visual angle between the start and end points of each saccade. Then, fixation durations, saccade amplitudes, and pupil sizes were binned into 1000 ms bins by calculating the mean fixation durations, saccade amplitudes, and pupil sizes falling within each bin. When fixations were longer than 1000 ms, bins were assigned the fixation duration of the next ending fixation. This means that when fixations were longer than 1000 ms, multiple time bins were assigned the fixation duration associated with that fixation. To remove potential artifacts from loss of eye-tracker calibration over the course of the five- to six-minute movies, fixation duration, saccade amplitude, and pupil size data were detrended by regressing out time and using the residuals in subsequent analyses.

We used mixed-effects models (lme4 package: Bates, Maechler, Bolker, & Walker, 2014; lmerTest package: Kuznetsova, Brockhoff, & Christensen, 2014) to test the hypotheses that saccade amplitude, fixation duration, and pupil size would show phasic changes during the periods around event boundaries. Mixed-effects models were chosen to capture the nesting of observations on movies and on subjects (in other words, to explicitly model the random variance attributable to both differences between movies and differences between people). Subject and movie were both entered into the models as random effects. (See Raudenbush and Bryk, 2002, for a comprehensive reference on mixed-effects models.)

We created predictor variables to model the effects of event boundaries for the ten seconds before and after event boundaries. The following steps were taken to create these predictor variables: (a) For each 1000 ms bin, participants either had a 0, meaning that they did not identify an event boundary during that bin, or a 1, meaning that they did identify an event boundary during that bin; (b) separate binary variables were created based on each participant's unique segmentation data; and (c) predictor variables were generated by shifting the resulting binary event boundary variable to estimate the effects of event boundaries from 10 s before an event boundary to 10 s after an event boundary. For example, if a participant segmented during the 100th time bin, the original binary event boundary variable would have a 1 in the 100th time bin. To create the predictor variable modeling the response occurring 1 s after this event boundary, the original binary event boundary variable would be shifted forward by one place, and the new predictor variable estimating this response would have a 0 in the 100th time bin and a 1 in the 101st time bin.

This resulted in 21 predictor variables for fine segmentation and 21 predictor variables for coarse segmentation. These event boundary predictor vari-

ables were entered into the models as fixed effects. The parameter estimates associated with these 21 predictors can be interpreted as the model's estimate of the time course of the response to an event boundary, controlling for the baseline value, effects of nuisance variables, and other nearby event boundaries. In other words, the parameter estimates are statistically controlled "average" responses at each time point.

Each participant's own segmentation data were used to determine the locations of fine and coarse boundaries for the event boundary variables. There was good agreement across individuals in the locations of event boundaries (see Supplementary materials).

Eye movements and pupil diameter are known to be driven by low-level visual features; for example, ambient brightness strongly affects the contraction of the pupil (Crawford, 1936). To control for these effects, we performed visual feature analyses on all stimuli using iLab visual saliency software (Itti, 2012), which provided measures of color conspicuity, motion conspicuity, orientation conspicuity, blue/yellow contrast, motion up, motion down, motion left, motion right, temporal flicker, intensity contrast, orientation 0°, orientation 45°, orientation 90°, orientation 135°, red/green contrast, mean saliency, variance of saliency, and luminance. The iLab software outputted data for every frame of the movies, and the resulting raw data were binned into 1000 ms bins by averaging across the data for each bin. The binned measures were then entered into the mixed-effects models as covariates to control for their effects on the dependent measures.

Finally, saccade amplitude, fixation duration, and pupil size were entered as the dependent variables in separate mixed-effects models. Because we were interested in the time course of these three oculometric measures during passive, coarse segmentation, and fine segmentation viewing conditions, our investigation resulted in a total of nine mixed-effects models. The three models of the passive viewing runs included predictor variables for both fine and coarse event boundaries ($2 \times 21 = 42$ predictors); the models for the coarse segmentation runs included predictor variables for only coarse segmentation (21 predictors); and the models for the fine segmentation runs included predictors for only fine segmentation (21 predictors). All of the mixed-effects models also included subject and movie as random effects.

To test whether time relative to event boundaries that were identified during active segmentation predicted each oculometric measure during passive viewing, we first conducted omnibus tests that compared the full model for that measure to a reduced model that omitted either the fine or coarse segmentation predictors. To illustrate, to determine whether fine event boundaries identified during the active event segmentation task predicted changes in saccade amplitude

	Fine segmentation		Coarse segmentation	
	Boundaries	Lengths (seconds)	Boundaries	Lengths (seconds)
Study 1				
Binder	33.57 (24.34)	14.26 (8.12)	9.18 (6.93)	47.88 (25.78)
Sweeping	26.50 (15.7)	14.93 (6.79)	8.57 (5.69)	44.01 (21.63)
Changing tire	24.61 (18.91)	19.11 (11.23)	6.57 (3.36)	56.48 (29.16)
Study 2				
Breakfast	30.92 (19.1)	15.01 (9.38)	13.6 (9.98)	34.11 (24.14)
Party	34.24 (24.00)	18.77 (15.18)	11.6 (13.35)	45.58 (23.95)
Planting plants	28.44 (17.28)	17.40 (11.29)	10.56 (9.66)	48.53 (27.40)

Table 1. Mean number of boundaries and event lengths identified for each movie. *Notes*: Numbers in parentheses are standard deviations.

during passive viewing, we removed the fine event boundary predictor variables from the model and compared this reduced model to the full model that included both fine and coarse event boundary predictor variables. A significant difference in the predictive utility of these two models suggested that fine event boundaries predicted variability in saccade amplitude. The approach for the fine and coarse segmentation runs was similar, except that the fine segmentation runs included only fine event boundary predictor variables and the coarse segmentation runs included only coarse boundary predictors.

When the omnibus test suggested that there was a significant difference between the full and reduced models, we described the effect by examining the uncorrected t statistics for partial slope estimates for the individual time points. An estimate of zero suggested that the dependent variable did not differ from the statistically controlled average (hereafter referred to as the baseline) at the time point, whereas an estimate greater than or less than zero suggested that the dependent variable increased or decreased relative to the baseline at that time point. Time points far away from an event boundary served as the statistical baseline, as it was unlikely that time points nine or ten seconds before or after event boundaries would influence the dependent variables. (See the caption of Figure 2 for an illustration of how to interpret the model estimates.)

Results

For both studies, participants identified fewer event boundaries when they were asked to segment at a coarse grain than when they were asked to segment at a fine grain. (See Table 1 for descriptive statistics for the number of event boundaries identified during each movie for the two studies.)

Passive viewing

For Study 1, during passive viewing, the mixed-model omnibus test of the effect of event boundaries on saccade amplitude was significant for saccade amplitude for fine ($\chi^2 = 47.07$, $df = 21$, $p < 0.001$), but not coarse ($\chi^2 = 20.08$, $df = 21$, $p = 0.52$) event boundaries. As shown in Figure 2 (top row), examining the estimates for the fixed effects of time point suggested that saccade amplitude decreased at 8 s before fine boundaries, returned to baseline, and then decreased again during the 3 s before fine event boundaries. Saccade amplitude returned to baseline at fine event boundaries. For the second study, as shown in Figure 3 (top row), the saccade amplitude effect during passive viewing resembled that seen in Experiment 1, but the effects of coarse ($\chi^2 = 26.42$, $df = 21$, $p = 0.19$) and fine ($\chi^2 = 27.82$, $df = 21$, $p = 0.14$) boundaries were not significant. Figure 2 displays the fixed effect estimates and standard errors for the event boundary predictor variables for Study 1, and Figure 3 displays the estimates and standard errors for Study 2.

The omnibus test for fixation duration during passive viewing in Study 1 was significant for coarse ($\chi^2 = 58.7$, $df = 21$, $p < 0.001$) but not fine ($\chi^2 = 6.5$, $df = 21$, $p = 0.99$) boundaries. For the fixed effects of time point, fixation duration during passive viewing decreased at around four seconds before coarse event boundaries and then remained depressed through five seconds after event boundaries, after which it returned to baseline. For Study 2, the omnibus tests for fixation duration (coarse: $\chi^2 = 11.9$, $df = 21$, $p = 0.94$; fine: $\chi^2 = 5.4$, $df = 21$, $p = 0.99$) during passive viewing were not significant. (See Tables S1 and S4 for the effects of the visual variables on saccade amplitude and fixation duration during passive viewing.) Pupil size did not change significantly at event boundaries during passive viewing for either study. (See Tables S1 and S3 for the effects of the visual variables on pupil size during passive viewing.)

To assess whether controlling for the low-level visual features of the movies could be masking effects of event
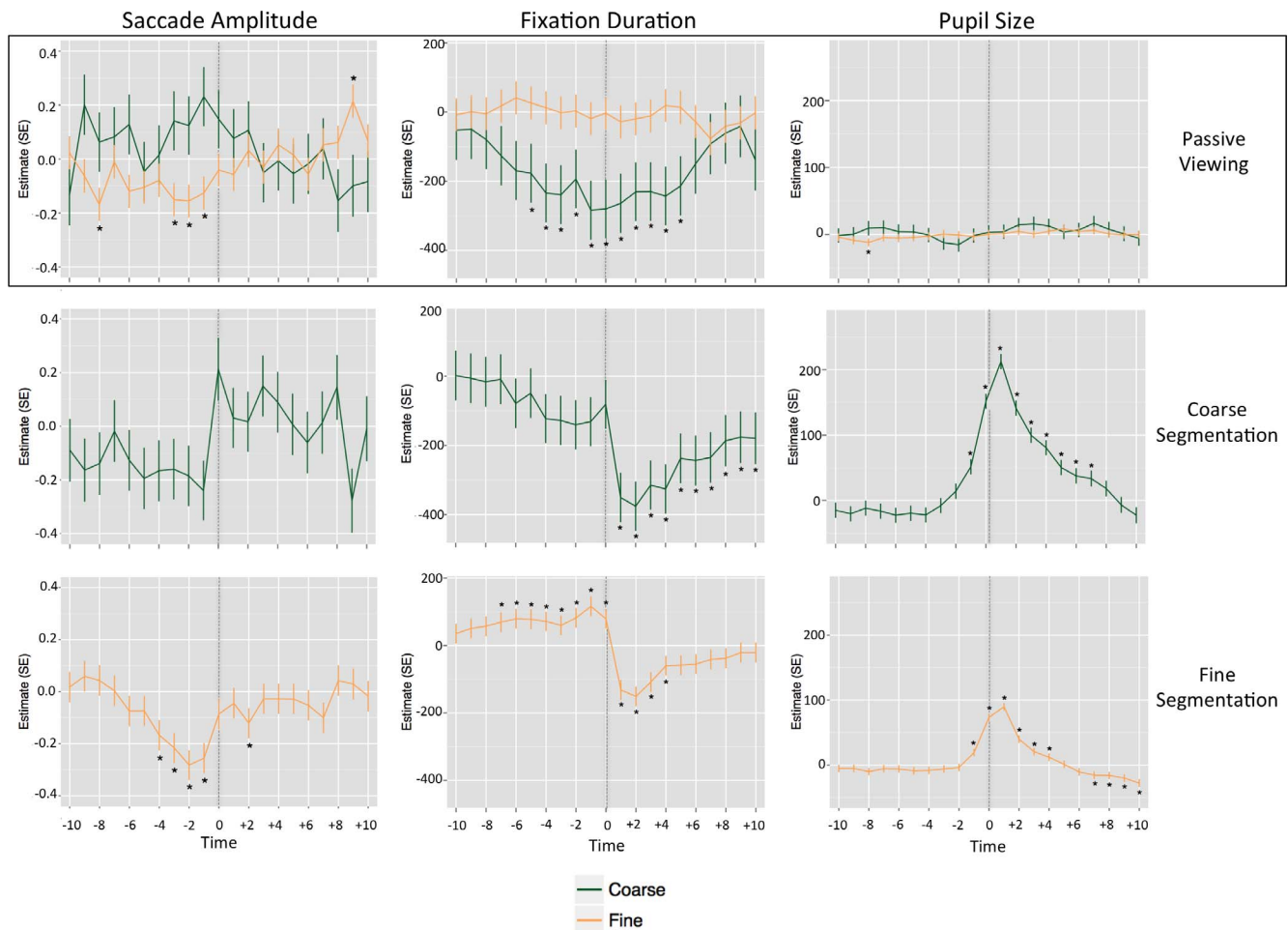
Figure 2. (Top row) Saccade amplitude, fixation duration, and pupil size during passive viewing time-locked to coarse and fine boundaries in Study 1. (Middle row) Saccade amplitude, fixation duration, and pupil size during fine segmentation time-locked to fine boundaries. (Bottom row) Saccade amplitude, fixation duration, and pupil size during coarse segmentation time-locked to coarse boundaries. In this figure, the *x* axes represent the time points before, during, and after event boundaries. The dotted lines are at time 0—the time of the event boundaries. The *y* axes represent the estimated time course of response from the mixed-effects models (see *Data analysis*). An estimate of 0 suggests that there is no effect of an event boundary for that time point. A positive estimate indicates that the value of the dependent variable increased at that time point, and a negative estimate suggests that the value of the dependent variable decreased at that time point, relative to the model's estimate of the baseline. To illustrate, for saccade amplitude during passive viewing, time-locked to fine boundaries (top left, yellow line) the estimate for saccade amplitude 1 s before an event boundary was −0.13°. This means that saccade amplitude decreased by 0.13° below baseline 1 s before participants identified event boundaries. The visual variable covariates were included in each model but are not displayed in this Figure. (*: *t* score > 2)

boundaries on eye movements during passive viewing, we fit mixed effects models as described above but without controlling for the visual features. The results were quite similar to those from the models that included the visual features. The only substantive difference was that for Study 2, the omnibus test for the effects of fine boundaries on saccade amplitude during passive viewing was significant ($\chi^2 = 41.77$ $df = 21$, $p = 0.004$). Saccade amplitude was significantly below baseline at fine event boundaries and 7 s after.

The random effect of movie was not significant for Study 1 for saccade amplitude ($\chi^2 = 0$, $p = 1$) during passive viewing, but was significant for Study 2 ($\chi^2 =$

9.92, $p = 0.002$). The random effects of movie on fixation duration during passive viewing were not significant for Study 1 ($\chi^2 = 0$, $p = 1$) or Study 2 ($\chi^2 = 2.84$, $p = 0.09$).

Active segmentation task

For Study 1, during active segmentation there was a significant omnibus effect on saccade amplitude of fine boundaries ($\chi^2 = 88.7$, $df = 21$, $p < 0.001$), but not coarse boundaries ($\chi^2 = 31.9$, $df = 21$, $p = 0.06$). For Study 2, there was significant omnibus effect on
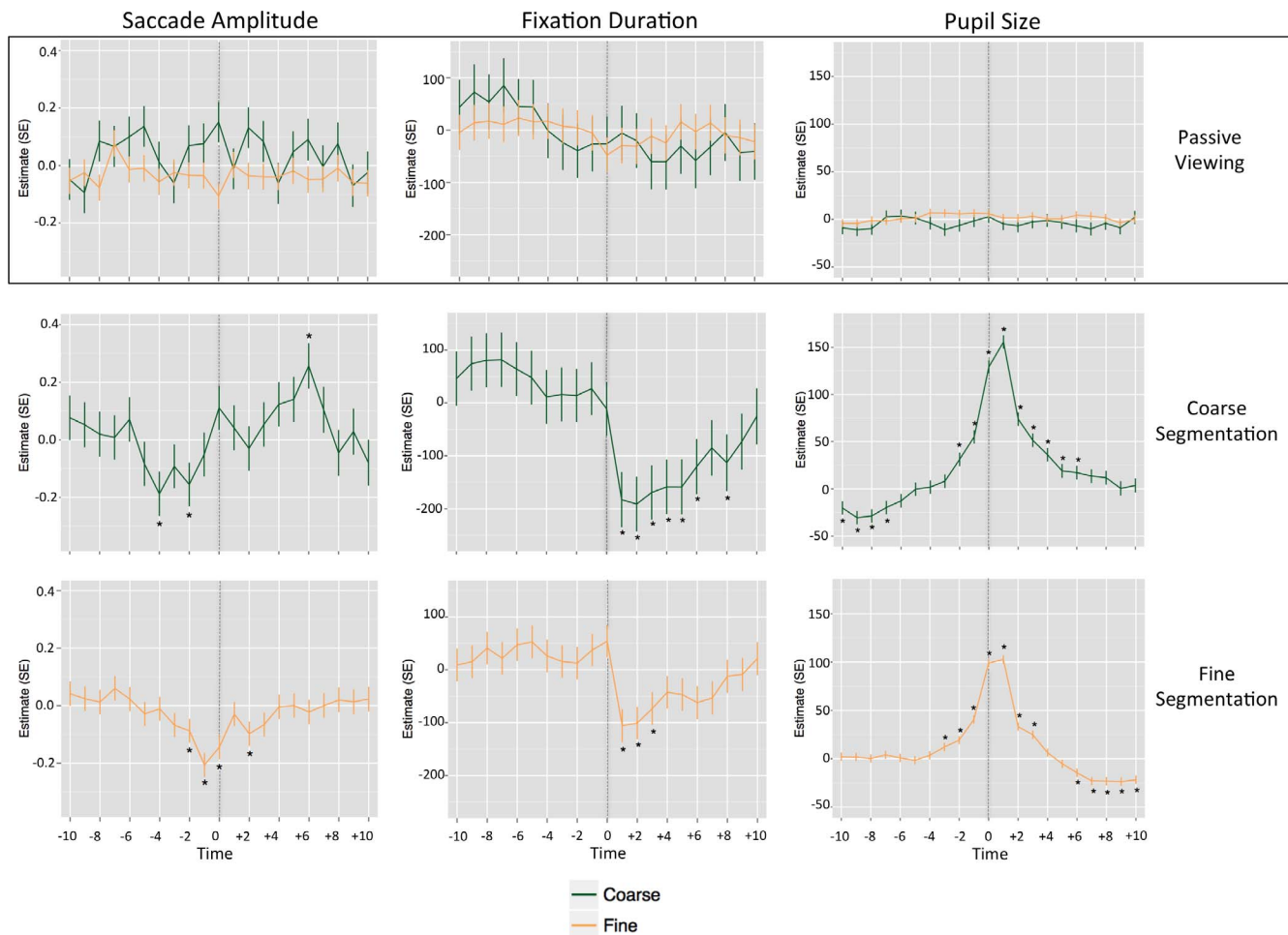
Figure 3. (Top row) Saccade amplitude, fixation duration, and pupil size during passive viewing time-locked to coarse and fine boundaries in Study 2. (Middle row) Saccade amplitude, fixation duration, and pupil size during fine segmentation time-locked to fine boundaries. (Bottom row) Saccade amplitude, fixation duration, and pupil size during coarse segmentation time-locked to coarse boundaries. (*: $t > 2$)

saccade amplitude of both coarse ($\chi^2 = 35.7$, $df = 21$, $p = 0.02$) and fine boundaries ($\chi^2 = 47.4$, $df = 21$, $p < 0.001$). Overall, saccade amplitude decreased prior to detection of both types of boundary. (See Tables S2, S3, S5, and S6 for the effects of the visual variables on saccade amplitude during coarse and fine segmentation.)

During active segmentation for both studies, there were significant omnibus effects on fixation duration of coarse boundaries (Study 1: $\chi^2 = 142.64$, $df = 21$, $p < 0.001$; Study 2: $\chi^2 = 69.8$, $df = 21$, $p < 0.001$) and fine boundaries (Study 1: $\chi^2 = 124.72$, $df = 21$, $p < 0.001$; Study 2: $\chi^2 = 40.8$, $df = 21$, $p = 0.006$). Fixation duration decreased after coarse and fine boundaries and then slowly recovered. (See Tables S2, S3, S5, and S6 for the effects of the visual variables on fixation duration during coarse and fine segmentation.)

Finally, during active segmentation, the effects on pupil size of coarse boundaries (Study 1: $\chi^2 = 756.6$, $df = 21$, $p < 0.001$; Study 2: $\chi^2 = 921.5$, $df = 21$, $p < 0.001$) and fine boundaries (Study 1: $\chi^2 = 444.5$, $df = 21$, $p < $

0.001; Study 2: $\chi^2 = 1008.4$, $df = 21$, $p < 0.001$) were significant. Pupil size showed large increases around coarse and fine event boundaries. (See Tables S2, S3, S5, and S6 for the effects of the visual variables on pupil size during coarse and fine viewing.)

Discussion

These results indicate that fixation duration and saccade amplitude dynamically vary during the seconds preceding and following event boundaries during viewing of naturalistic movie stimuli. Specifically, we found that during passive viewing, saccade amplitude decreased just prior to event boundaries, and fixation duration decreased around coarse event boundaries that were identified during active segmentation. Although these effects were not significant in the second

study, the general patterns replicated across the two studies.

Comprehending dynamic activity requires people to make predictions about the future and update working memory representations when these predictions fail. Previous research has found that predictions are less accurate around event boundaries (Zacks, Kurby, Eisenberg, & Haroutunian, 2011) and that people tend to identify event boundaries at times when activity features change (Magliano, Miller, & Zwaan, 2001; Zacks, Speer, & Reynolds, 2009; Zacks, Swallow, Vettel, & McAvoy, 2006), suggesting that people must gather as much information as possible about the changes in the environment at event boundaries. It is possible that as features in the environment begin to change and predictions become less accurate, people explore the local changes happening near the site of the previous activity. Then, as the actor begins to interact with a new object, saccade amplitude increases and fixation duration decreases, allowing for exploration of a larger portion of the screen. It is sensible that this exploratory visual processing would best allow for a fast influx of information to allow for adaptive updating of working memory. Once broad information about the new activity has been gathered, focal processing of detailed information about the current activity would allow for a more in-depth understanding of the current activity and better predictions of future activities.

The results of the studies discussed here provide support for this hypothesis, as participants were more likely to look at smaller regions of the screen prior to event boundaries and spent less time looking at each individual region during the few seconds preceding and following event boundaries. In addition, participant looking patterns were generally at baseline during times that were within events (farther from event boundaries), meaning that they did not differ from the statistically controlled average when event boundaries were farther away in time. On the other hand, contrary to our hypothesis, saccade amplitude did not increase significantly above baseline at or immediately following event boundaries. However, overall, these results suggest that people engage in exploratory processing while updating their working memory representation of a new activity. They then return to focal processing and focus on details of the activity for longer periods of time in order to make accurate predictions of future activity.

Another possible explanation for the finding that fixations were shorter around event boundaries is that the distribution of visual information around the scene is likely larger around event boundaries. For example, if an actor switches from making toast to making eggs, which involves walking across the screen, shorter fixations are likely necessary to follow the actor's

movements. On the other hand, this simple visual processing mechanism also would predict an increase in saccade amplitude around event boundaries, a prediction that was not supported by the data presented here. In addition, because working memory was not tested in this study, it is not possible to determine whether simple visual processing or working memory updating led to the shift from focal to ambient processing found in this study. However, previous studies have tested working memory around event boundaries and have found that working memory is, in fact, updated at event boundaries (e.g., Radvansky, Tamplin, & Krawietz, 2010), providing further support for the hypothesis that working memory updating and the need to form accurate predictions about future activity drive the change from focal to ambient processing around event boundaries.

These results also extend Smith et al.'s (2006) finding that saccade frequency decreased a few hundred milliseconds before event boundaries and increased through 100 milliseconds after event boundaries, as we found evidence for exploratory processing over a longer time scale. In addition, we used movies that were as naturalistic as possible, without any edits, extending Pannasch's (2014) findings for edited movies to movies of naturalistic, everyday activities.

Because the random effect of movie was significant for Study 2, the differences in saccade amplitude between the two studies could be associated with an effect of movie. However, it is unlikely that all of the differences in the results for the two studies were due to an effect of movie, especially given that the overall patterns in the data were similar across the two studies. Instead, it is possible that differences in the participant populations, or potentially random selection of participants for the two studies, led to a replication of the overall pattern of results, but a lack of significant findings in the second study.

Although we have focused mainly on the results obtained during passive viewing, there were also a number of significant effects during active segmentation: Saccade amplitude decreased immediately before both fine and coarse event boundaries, fixation duration decreased immediately after boundaries, and pupil size increased around boundaries. However, these effects are not pure measures of the correlates of event boundaries per se; they also reflect task-specific processing. In this task, one must make a perceptual judgment that an event boundary has occurred and plan and execute a motor response to report this judgment, and these operations likely have substantial effects on the eyes. The pupillary results from the active segmentation runs are consistent with previous eye-tracking studies of perceptual-motor task performance (e.g., Richer & Beatty, 1985; Richer, Silverman, & Beatty, 1983). On the other hand, the fixation duration

results from the active segmentation runs are in contrast with previous eye-tracking studies, which have found that making a button press inhibits saccades for a few hundred milliseconds, and therefore results in longer fixations (e.g., Huestegge, 2011; Loschky, McConkie, Yang, & Miller, 2005). As discussed above, in the present study, fixation duration decreased immediately after the button was pressed to indicate the presence of an event boundary. One possible explanation is that the short time course of the inhibition of eye movements is overwhelmed by the much longer time course of the shift to ambient processing around event boundaries. It is also possible that performing the active event segmentation task focuses participants' attention on the event boundaries, thereby amplifying the effects of event boundaries on the oculometric responses.

In addition, the effects seen during passive viewing were not just smaller than those during active segmentation, but also more smeared out in time. This was certainly due, in part or in full, to the fact that the data from the passive viewing condition was time locked to the event boundaries identified during the active viewing conditions. Temporal smearing would therefore be expected to the extent that participants experience event boundaries slightly differently during each viewing.

Finally, the fact that all the effects reported here were observed while controlling for low-level visual features of the movies, including luminance, motion, color, and other properties, suggests that low-level visual features were not sufficient to account for the changes in saccade amplitude and fixation duration around event boundaries. Indeed, the reported effects may slightly underestimate the true response due to these controls. Therefore, it is likely that the need to update working memory representations when activity becomes less predictable accounts for the switch to exploratory processing around event boundaries.

Conclusion

The results presented here provide support for the hypothesis that viewers update working memory representations when predictions become less accurate, and that this update entails a shift from focal to exploratory visual processing. This phasic shift could be an important component of cognitive control during the comprehension of everyday activity. It unfolds on the timescale of a few seconds and is well positioned to regulate the formation of a mental representation that underlies the experience of the conscious present.

Acknowledgments

References

Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, *103*, 62–70.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. Retrieved from http://CRAN.R-project.org/package=lme4

Carmi, R., & Itti, L. (2006a). The role of memory in guiding attention during natural vision. *Journal of Vision*, *6*(9):4, 898–914, doi:10.1167/6.9.4. [PubMed] [Article]

Carmi, R., & Itti, L. (2006b). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research*, *46*, 4333–4345, doi.org/10.1016/j.visres.2006.08.019.

Crawford, B. H. (1936). The dependence of pupil size upon external light stimulus under static and variable conditions. *Proceedings of the Royal Society of London. Series B*, *121*, 376–395.

Eyelink User Manual. (2010). Version 1.52. Mississauga, ON, Canada: SR Research Ltd.

Huestegge, L. (2011). The role of saccades in multitasking: towards an output-related view of eye movements. *Psychological Research*, *75*(6), 452–465.

Itti, L. (2012). iLab Neuromorphic Vision C++

Toolkit (iNVT). Available at http://ilab.usc.edu/toolkit/. Accessed November 23, 2013.

Irwin, D. E., & Zelinsky, G. J. (2002). Eye movements and scene perception: Memory for things observed. *Perception & Psychophysics*, *64*, 882–895, doi:10.3758/BF03196793.

Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: Evidence for the adaptive gain theory. *Journal of Cognitive Neuroscience*, *23*, 1587–1596, doi:10.1162/jocn.2010.21548.

Krejtz, I., Szarkowska, A., Krejtz, K., Walczak, A., & Duchowski, A. (2012). Audio description as an aural guide of children's visual attention. In *Proceedings of the symposium on eye tracking research and applications* (pp. 99–106). New York: ACM.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-11. Retrieved from http://CRAN.R-project.org/package=lmerTest

Loschky, L., McConkie, G., Yang, J., & Miller, M. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition*, *12*(6), 1057–1092.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, *15*, 533–545.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, *3*(1), 5–24, doi:10.1007/s12559-010-9074-z.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*, 28–38.

Pannasch, S. (2014). Characteristics of ambient and focal processing during the visual exploration of dynamic stimuli. *Journal of Vision*, *14*(10): 1208, doi:10.1167/14.10.1208. [Abstract]

Pannasch, S., Helmert, J. R., Roth, K., Herbold, A-K., & Walter, H. (2008). Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, *2*(2), 1–19.

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Radvansky, G. A., Tamplin, A. K., & Krawietz, S. A. (2010). Walking through doorways causes forgetting: Environmental integration. *Psychonomic Bulletin & Review*, *17*, 900–904, doi:10.3758/PBR.17.6.900.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Thousand Oaks, CA: Sage Publications.

Richer, F., & Beatty, J. (1985). Pupillary dilations in movement preparation and execution. *Psychophysiology*, *22*, 204–207.

Richer, F., Silverman, C., & Beatty, J. (1983). Response selection and initiation in speeded reactions: A pupillometric analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 360–370.

Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, *13*(8):16, 1–24, doi:10.1167/13.8.16. [PubMed] [Article]

Smith, T. J., Whitwell, M., & Lee, L. (2006). Eye movements and pupil dilation during event perception. In *Proceedings of the 2006 symposium on eye tracking research & applications* (pp. 48). New York, NY: ACM. doi:10.1145/1117309.1117333.

Swallow, K. M., & Zacks, J. M. (2006). Hierarchical grouping of events revealed by eye movements. *Abstracts of the Psychonomic Society*, *11*, 81.

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, *31*, 299–337, doi:10.1007/BF00422717.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, *12*, 473–494, doi:10.1080/13506280444000409.

Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. *Proceedings of the XXVII Conference of the Cognitive Science Society*, 2283–2288.

Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*, 145–156, doi:10.1016/S1369-8478(02)00013-X.

Zacks, J., Braver, T., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*, 651–655. Retrieved from http://dcl.wustl.edu/pubs/ZacksNatNeuro01.pdf

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*, 4057–4066.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology. General*, *138*, 307–327, doi:10.1037/a0015305.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*, 273–293, doi:10.1037/0033-2909.133.2.273.

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. L. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging*, *21*, 466–482, doi:10.1037/0882-7974.21.3.466.

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neutral correlates of event perception. *Brain Research*, *1076*(1), 150–162, doi:10.1016/j.brainres.2005.12.12.