The Effect of Cinematic Cuts on Human Attention

Christian Valuch
Cognitive Science Research Platform, University of Vienna
christian.valuch@univie.ac.at

Ulrich Ansorge
Faculty of Psychology, University of Vienna
ulrich.ansorge@univie.ac.at

Shelley Buchinger
Faculty of Computer Science, University of Vienna
shelley.buchinger@univie.ac.at

Aniello Raffaele Patrone
Cognitive Science Research Platform, University of Vienna
aniello.patrone@univie.ac.at

Otmar Scherzer
Faculty of Mathematics, University of Vienna
otmar.scherzer@univie.ac.at

ABSTRACT

Understanding the factors that determine human attention in videos is important for many applications, such as user interface design in interactive television (iTV), continuity editing, or data compression techniques. In this article, we identify the demands that cinematic cuts impose on human attention. We hypothesize, test, and confirm that after cuts the viewers' attention is quickly attracted by repeated visual content. We conclude with a recommendation for future models of visual attention in videos and make suggestions how the present results could inspire designers of second screen iTV applications to optimise their interfaces with regard to a maximally smooth viewing experience.

Author Keywords

Attention; Eye tracking; Saccades; Editing; Continuity; Second screen applications; Evaluation

ACM Classification Keywords

H.1.2 [Models and Principles]: User/Machine Systems—human information processing; H.5.2 [Information Interfaces and Presentation (e.g. HCI)]: User Interfaces—evaluation/methodology; H.5.m. [Information Interfaces and Presentation (e.g. HCI)]: Miscellaneous

INTRODUCTION

Designing successful applications for online and interactive television (iTV) requires a proper understanding of the factors that determine the user's experience. Working towards this objective, HCI research has been using eye tracking as a means of evaluating user interfaces [11]. For instance, in multiple screen applications [1] users frequently shift their gaze between at least two locations [6]. The presence of the second screen can distract the viewer from the main content of the show [1]. Understanding which factors determine the viewer's attention in such situations would allow designers to optimize their applications in favor of a maximally smooth viewing experience. To investigate these questions on a more general level – with a broad range of applications, such as video coding [13], or continuity editing [14] – we looked at gaze shifts after cinematic cuts. Human attention is closely related to eye movements. *Saccades* - abrupt gaze shifts between two locations - are a direct consequence of shifting attention to a new location [8]. Accordingly, by looking at the properties of saccades, it is possible to formulate and test theories about attention.

Current models of human attention and gaze behavior in videos emphasize the role of novelty, or *Bayesian surprise*. They assume that visual content that is maximally dissimilar from the viewer's prior visual experience is the best predictor of human attention and gaze direction. Indeed, eye tracking confirmed that human gaze direction in continuous videos is better explained by Bayesian surprise than by alternative models [7]. However, this is not necessarily true for cuts within edited videos. Existing evidence suggests that attention is attracted by repeated visual features in situations where location correlations between two successive images are low [2, 9].

Edited videos frequently contain hard cuts, i.e. visual discontinuities that require shifting attention from one location to another because object locations are uncorrelated across the cut. Moreover, making sense of narratives and content across cuts implicitly requires deciding whether the post-cut scene is a continuation of the pre-cut scene [14]. Here, *within scene cuts (WSCs)* continue with the same scene from a different angle; *between scenes cuts (BSCs)* continue with a different scene (see Figure 1). Orienting attention to repeated visual features could enable viewers' quick and efficient recognition of content that connects the cut images (in the case of WSCs).

The Present Study

We tested the hypothesis that after cuts, attention is more strongly attracted by repeated visual content than by novel, or surprising content. We conducted an eye tracking experiment, in which participants had to watch and keep their gaze on a video that was shown next to another, irrelevant video. Both videos contained hard cuts and unforeseeably kept or switched their locations at the cuts. This manipulation created a low correlation of object locations as is typical of cuts. Presenting two videos side by side also allowed us to measure influences of repeated versus novel content on saccades,

TVX 2014, June 25–27, 2014, Newcastle Upon Tyne, UK.

ACM 978-1-4503-2838-8/14/06 ...$15.00.
http://dx.doi.org/10.1145/2602299.2602307

during which attention and eye movements are tightly coupled [8]. If locations switched, participants had to saccade to the new location of the video they were instructed to follow (similar to shifting gaze between two screens).

We analyzed *saccadic reaction time (SRT)* as a measure of viewers' re-orienting of attention to the post-cut scene after a location switch. Following our hypothesis, we predicted shorter SRT after cuts where much visual content was repeated (WSCs, or cuts with high image-image similarity) and longer SRT after cuts where less visual content was repeated (BSCs, or cuts with low image-image similarity).[1] In the remaining sections of this paper we give details on our method, results, and discuss implications for further research and improvements of iTV applications.

Figure 1. Example cuts. (A) Within scene cut (WSC). (B) Between scenes cut (BSC). Screenshots derived from videos by QParks.com, available under CC BY 3.0 at vimeo.com/89901459 and vimeo.com/89248621.

METHOD

Participants

Forty-two students (34 female) with a mean age of 23 years took part in an eye tracking experiment. Informed consent was obtained from all participants.

Stimuli

We used 20 sports videos in which we deliberately inserted new cuts. Each video showed the same sport throughout (e.g., *skiing*). Videos were edited in pairs, resulting in ten sets of two videos. The sport in the first video was always different from the sport in the second video (e.g., *skiing* vs. *surfing*). Cuts always occurred simultaneously in both videos. Average video duration was 2.5 minutes and the complete set contained 212 cuts. Cuts were assigned to either a WSC or a BSC condition. Whenever major visual changes, e.g. in scenery, actors, or ongoing actions occurred with the cut, the cut was coded as a BSC. In contrast, cuts that connected two images showing the same scene, action, and actors were coded as WSCs. Figure 1 shows examples. We assumed that more visual content is repeated after WSCs than after BSCs.

[1]This prediction is the opposite of that of the Bayesian surprise model which generally predicts a shorter SRT for less similar than for more similar image content.

To validate this, we compared the similarity of color histograms of the last pre-cut and the first post-cut frame and, based on this measure, assigned each cut to a *High similarity* or a *Low similarity* condition. We used color similarity because color contributes to gaze and attention preferences for repeated information [9], allows visual recognition after location and/or perspective shifts [15], and conveys information useful for cut detection [5].

Apparatus

Gaze data were recorded using an EyeLink 1000 Desktop Mount eye tracker (SR Research Ltd., Kanata, Ontario, Canada) at a sampling rate of 1000 Hz. The eye tracker was calibrated to each viewer's dominant eye using a 5-point calibration. Every time the videos switched locations, the exact timestamp was saved to the eye tracking data file, which allowed analyzing the latency of the first saccade to the target video with millisecond precision. Stimuli were displayed on a 19-in. color CRT monitor (Sony Multiscan G400) with a resolution of 1,280 $\times$ 1,024 pixels and a refresh rate of 60 Hz. The experimental procedure was implemented in MATLAB (MathWorks, Natick, MA, USA) using the Psychophysics Toolbox [3, 10] and the Eyelink toolbox [4]. Viewing distance to the monitor was 72 cm supported by chin and forehead rests. The viewable screen area subtended 28° $\times$ 21°. The apparent size of the 400 $\times$ 300 pixel videos was 8.75° $\times$ 6.15° and they were shown vertically centered at a horizontal eccentricity of 6.56°.

Procedure and Design

The experiment consisted of 20 blocks in which two videos were presented on the screen. Importantly, participants were instructed to view only one of the videos (the target video) while ignoring the other (the distractor video). At the beginning of each block, the starting location of the target video was announced by a green rectangle. Participants were informed that the videos switched locations at random intervals, and instructed to relocate their gaze as fast as possible to the target video's new location once the videos had switched locations. Throughout the experiment, each block was presented twice so that either of the videos was serving as the target in the first and as the distractor in the second half of the experiment (or vice versa).

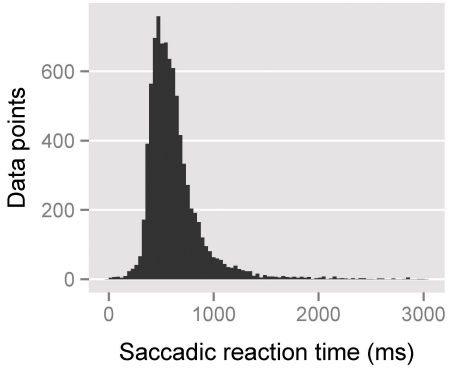

Figure 2. Distribution of valid saccadic reaction times (i.e. the latencies of the first saccade to the target video after a location switch).

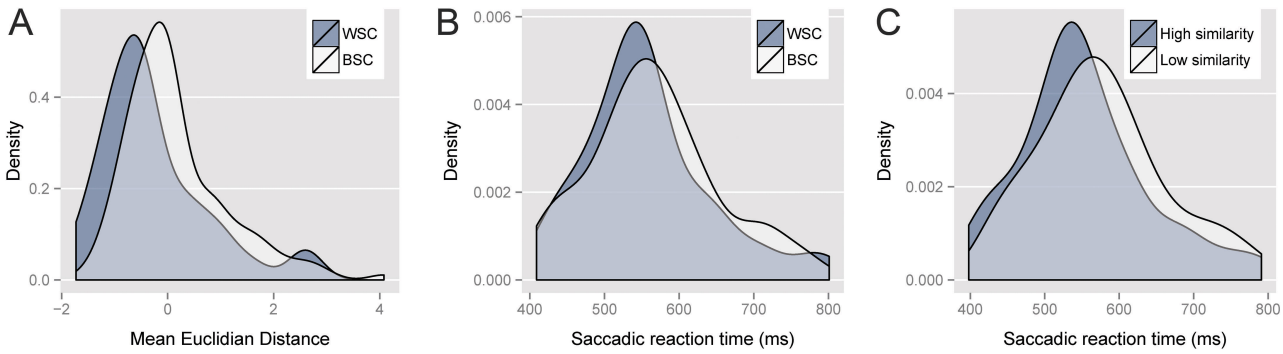

Figure 3. Results. (A) Distribution of mean Euclidian distances (z-transformed) of RGB color histograms of the last pre-cut and the first post-cut frame as a function of cut category. Values below 0 represent higher similarity, values above 0 represent lower similarity. (B) Distribution of individual median SRT as a function of cut category. (C) Distribution of individual median SRT as a function of color histogram similarity across the cut.

Data Analysis

Saccades were identified as sample periods where the change in gaze direction was larger than 0.1°, eye movement velocity exceeded 30°/s, and acceleration exceeded 8000°/s^2. The main dependent variable was SRT, defined as the latency of the first saccade towards the target video after the videos switched locations. SRT was analyzed as a function of the type of cut (WSC vs. BSC) in the target video, and the similarity of RGB color histograms across cuts (High similarity vs. Low similarity) – for further details see Stimuli and Results. We expected shorter SRTs (faster gaze relocation) after WSCs than BSCs. Similarly, we expected shorter SRTs after High similarity than after Low similarity cuts.

Gaze data were preprocessed in MATLAB and statistical tests were run in R [12]. Out of 8,904 collected data points (i.e. 212 cuts for each of the 42 participants), 8,397 (94.3 %) contained valid SRTs and were subjected to statistical analyses. Data were excluded if no saccade to the target video was identified within a time-window of 3 s after the location switch or if gaze was already at the new location shortly ahead of the switch. Figure 2 depicts the distribution of valid SRTs. Individual median SRTs per condition were tested for within-participant differences by t-tests. We report Pearson correlation coefficients as measures of effect sizes. For all statistical tests, we set α at 0.05.

RESULTS

Image Similarity Across Cuts

To validate that more visual content is repeated after WSCs than BSCs, we calculated the mean Euclidian distance of the RGB color histograms between the final pre-cut and the first post-cut frame. For better interpretability, we z-transformed these values, so that values below 0 represent higher similarity (indicated by the smaller Euclidian distance), and values above 0 represent lower similarity (indicated by the greater Euclidian distance). A Welch two sample t-test indicated significantly higher color similarity in WSCs than BSCs, $t(148.3) = 2.86$, $p<.01$, $r = .23$ (see also Figure 3A).

Saccadic Reaction Time After Location Switches

In a first analysis, we tested whether the a priori categories of WSCs and BSCs could explain any variance in SRTs. Using a paired t-test, we found that median SRT was on average 9 ms shorter in WSCs than BSCs, $t(41) = -2.03$, $p<.05$, resulting in a medium-sized effect of $r = .30$ (see also Figure 3B).

For a second analysis, we categorized the cuts into either a *High similarity* or a *Low similarity* condition, depending on whether the z-transformed similarity measure for these cuts was below or above 0. Again, we tested for significant differences in SRTs between these conditions. A paired t-test of median SRTs confirmed that on average SRTs were 23 ms shorter after High similarity than after Low similarity cuts, $t(41) = -6.83$, $p<.001$, representing a large effect of $r = .73$ (see also Figure 3C).

DISCUSSION

Our data suggest that after cuts viewers are able to re-orient their attention more quickly if visual content is repeated from the pre-cut scene: Following WSCs or High similarity cuts, saccades to the target video were initiated significantly faster than after BSCs, or Low similarity cuts. Results confirmed viewers' preference for repeated features during reorienting after cuts with low object-position correlations. The following limitations apply.

First, our results seem to conflict with the assumption that novel or surprising information is the best predictor of attention and gaze direction in videos [7]. However, we argue that an advantage for repeated information characterizes only a short time frame following cuts. During this period, viewers search for familiar visual content for deciding whether the previous scene continues, or not. Soon after, a preference for novel or surprising information should take over but future models should account for the effect of cuts on attention, too.

Second, in an effort to precisely measure the speed of attentional orienting after cuts we presented two videos simultaneously. This enabled us to elicit and record saccades of comparable start/end points for each cut. This is good because saccades are valid reflections of attention. However, the surprise model was supported during viewing of single videos. Viewing single videos is a situation that we ultimately also want to understand. Therefore, future research should aim to replicate our findings under single video viewing conditions.

Third, motivated by previous research [2, 9, 5, 15], we validated the stronger repetition of visual content across WSCs as compared to BSCs based on color similarity only. However, other descriptors that do not rely on color might also sufficiently explain the observed differences in SRTs. Also, we are unable to isolate color-repetition effects operating on a short timescale from the viewers' long-term knowledge about object-associated colors that possibly contributed to the color repetition effect (e.g., the knowledge that *snow* is *white*). These questions are open to debate and should be studied in future experiments, possibly by including control conditions with black and white videos.

Implications for iTV Applications

To conclude, we think a preference for repeated visual content applies in all situations in which the location of objects is uncorrelated across successive views. This is relevant for improving user interfaces in iTV. To give just one example, with second screen applications a second screen showing information that is visually unrelated to the main screen might distract the viewer [1]. Following from the present study, we would recommend that designers of second screen applications should include visual elements that repeat across both screens to minimize the time necessary for shifting attention between the two screens and assure a maximally smooth user experience. Even more interesting applications could become possible once eye tracking becomes widely available in consumer electronics. Then, it will be possible to dynamically adapt the content on a second device based on what was just looked at on the primary screen. Finally, we would like to stress that the methods presented in this paper can be easily adapted to study the effects of particular second screen iTV applications on human attention.

CONCLUSION

Our paper presents evidence that after cinematic cuts viewers quickly re-orient their attention to visual content that is repeated from the pre-cut scene. A preference for repeated visual content after cuts should be incorporated into models of human attention which currently assume that novelty or Bayesian surprise is the best predictor of human attention and gaze direction in videos. We also discussed implications of our results for the improvement of iTV applications.

ACKNOWLEDGMENTS

We thank the reviewers for their excellent comments on a previous version of the paper, as well as Melanie Szoldatics and Heide Maria Weißenböck for assistance with data collection, and gratefully acknowledge grant CS11-009 from the Wiener Wissenschafts-, Forschungs-, und Technologiefonds (WWTF, Vienna Science and Technology Fund) to Ulrich Ansorge, Shelley Buchinger, and Otmar Scherzer. We also wish to express our special gratitude to Dr. Anton Luger.

CORRESPONDENCE

Correspondence should be addressed to Christian Valuch, Cognitive Science Research Platform, University of Vienna, Liebiggasse 5, 1010 Wien, Austria.
Email: christian.valuch@univie.ac.at.

REFERENCES

1. Basapur, S., Mandalia, H., Chaysinh, S., Lee, Y., Venkitaraman, N., and Metcalf, C. Fanfeeds: evaluation of socially generated information feed on second screen as a tv show companion. In *Proc. EuroITV 2012*, ACM Press (2012), 87–96.
2. Becker, S. I. Can intertrial effects of features and dimensions be explained by a single theory? *Journal of Experimental Psychology: Human Perception and Performance 34*, 6 (2008), 1417–1440.
3. Brainard, D. The psychophysics toolbox. *Spatial Vision 10*, 4 (1997), 433–436. `http://psychtoolbox.org//`.
4. Cornelissen, F. W., Peters, E. M., and Palmer, J. The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers 34*, 4 (2002), 613–617.
5. Gargi, U., Kasturi, R., and Strayer, S. Perfomance characterization of video-shot-change detection methods. *Circuits and Systems for Video Technology, IEEE Transactions on 10*, 1 (2000), 1–13.
6. Holmes, M., Josephson, S., and Carney, R. Visual attention to television programs with a second screen application. In *Proc. Eye Tracking Research and Applications*, ACM Press (2012), 397–400.
7. Itti, L., and Baldi, P. Bayesian surprise attracts human attention. *Vision Research 49*, 10 (2009), 1295–1306.
8. Kowler, E., Anderson, E., Dosher, B., and Blaser, E. The role of attention in the programming of saccades. *Vision Research 35*, 13 (1995), 1897–1916.
9. Maljkovic, V., and Nakayama, K. Priming of pop-out: I. Role of features. *Memory & Cognition 22*, 6 (1994), 657–672.
10. Pelli, D. G. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision 10*, 4 (1997), 437–442.
11. Poole, A., and Ball, L. J. Eye tracking in HCI and usability research. In *Encyclopedia of Human-Computer Interaction*, C. Ghaoi, Ed. 2006.
12. R Core Team. *R: A language and environment for statistical computing*. 2012. `http://R-project.org/`.
13. Salomon, D. *Data compression: The complete reference*. Springer, 2004.
14. Smith, T. J., Levin, D. T., and Cutting, J. A Window on Reality: Perceiving Edited Moving Images. *Current Directions in Psychological Science 21*, 2 (2012), 107–113.
15. Swain, M. J., and Ballard, D. H. Color indexing. *International Journal of Computer Vision 7*, 1 (1991), 11–32.