universität wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

"Selective Attention in Videos –
Known Features versus Bottom Up Processes"

verfasst von / submitted by

Raphael Antonio Seywerth, MSc

angestrebter akademischer Grad /
in partial fulfilment of the requirements for the degree of

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2015 / Vienna, 2015

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 298
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Diplomstudium Psychologie
Betreut von / Supervisor:	Prof. Dr. Ulrich Ansorge

Abstract

With the widespread use of mobile phones capable of producing short video clips and easy to use software to compose short films, more and more people create and share videos. A lot of research is done for features that capture the attention of viewers, but very few analyse attention after cinematic cuts. Two theories: the top down, where known features guide attention, and the bottom up theory, seeing attention as stimulus-driven, are being challenged for interrelations.

In this paper two eye tracking experiments are conducted to analyse the horizontal fixations of viewers, the fixations' latency periods as well as their duration during the presentation of short video clips of natural scenes. By using sequences of two clips each, separated by a short cut, three different conditions: same, novel and shifted content were designed to test for preferences of novel against already known content.

In the first experiment, preferences for already known content are found between 251 and 1,500 ms in the clips after the cut ($p < .01$). Furthermore, significant reduced latency periods are found for fixations on known content. Since the task used for this experiment focuses on the relation of each two clips within a sequence, a modified task is used for the second experiment. The second experiment makes use of memorised clips that are to be searched for among the clips in each sequence. Slightly reversed effects are found here. Between 501 and 1,250 ms after the cut, preferences for novel content are seen ($p < .01$).

These results show a highly task specific attention process. Top down processes enable viewers to automatically focus on known as well as on novel content depending on their task. Usage of cuts and shifted conditions, as previously used in different settings, proved to be efficient for this research. Further research should focus more on varying tasks. Video makers have to take into account that the viewing experience can be, and sometimes should be, guided by instructions that could be as well part of the video themselves.

Acknowledgements

I like to express my appreciation and thanks to my supervisor Ulrich Ansorge whose research is the base for my works research question. Further thanks goes to Christian Valuch who also supported the experiments that were implemented for this study as well as all others who helped along the way.

Table of Contents

List of Figures

List of Tables

I Introduction

Due to technical advances in mobile phones, video software and sharing platforms anybody can become a film maker. While there is a lot of research focused on what captures the attention of people in supermarkets, advertising (Higgins, Leinenger, & Rayner, 2014) and very specific search tasks in laboratory environments, underlying processes of visual preferences in videos of natural environments get less attention. Such areas might get more and more into focus as technologies like head-mounted displays for virtual realities or other applications, or other attention-aware visual display technologies (Ferscha, Paradiso, & Whitaker, 2014) advance.

What captures our attention and why? Over the last years various theories to this question have been developed. Two of the most competitive seeming theories are the top-down or goal-driven and the bottom-up or stimulus-driven theory. Research shows that those theories do not seem to be eliminating one another in every aspect. Furthermore, ways to combine both theories have been suggested by Carmi and Itti (2006) and others. Enns, Los, Olivers, Schreij and Theeuwes (2014) write that top-down and bottom-up influences can occur automatically, but instead of interactive relationships only additive ones can be found. This suggests that both forms of attention capture operate independently. Enns et al. (2014) show that reaction time results may disguise underlying interactions of both influences and propose that both types of capture influence a shared spatial orienting mechanism. Their results also show that bottom-up as well as top-down attention capture can take place on subsequent saccades, implying both bias and signal being present simultaneously.

Much of the current visual attention research tests the principle of image salience and its guiding of human eye movements. The usual experiments use items shown completely within the participant's visual field. In a search task the top-down knowledge can guide attention, although, sometimes the most conspicuous item may capture attention, regardless of the observer's task. Chapman, Foulsham, Kingstone and Nasiopoulos (2014) argue that one way to bring visual search closer to real life is to

increase the complexity of both: targets and presented background. A dominant framework for modeling the selection of different items within a scene remains a feature-based approach, which was exemplified by Itti and Koch's (2000) computational model of visual salience. Chen and Zelinsky showed (2006) that bottom-up visual saliency seems to have little predictive power in realistic search tasks. Top-down expectations seem to dominate in real-world scenes and it is criticized that static research does not represent the dynamic and task-driven nature of vision in the real world. A study of Chapman et al. (2014) aims to determine the extent to which the principle of feature-driven or bottom-up attentional selection can be generalized to gaze allocation in a real-world search task where the observer is free to move his or her head and body. Using a mobile eye tracker and participants sent to a room with mailboxes, they report that bottom-up salience of the target envelope had no effect on the overall time taken to search for the target. Nevertheless, the salient target was more likely to be fixated and found once it was within the central visual field. Top-down knowledge of the target appearance had a larger effect, reducing the need for multiple head or body movements resulting in earlier fixation of the target. Chapman et al. (2014) conclude that they are able to demonstrate that principles from visual search in the laboratory also work in tasks within natural environments. Their results show that apart from possible application to real-world scenarios, the actual search task can have a large impact on the observer's reaction time.

Many studies exist that use photographs of natural scenes and record human gaze behavior to understand where attention is directed. Participants fixate more and longer on single objects during a scene-memory task than during a visual search task (Castelhano & Henderson, 2009), suggesting a supportive role of fixations for memorisation of natural scenes. Other research shows that humans are also able to recognize scenes at single glances or, on the other hand, are prone to dynamic overwriting of scenes (Simons & Rensink, 2005). Ansorge, Becker and Valuch (2013) show that particular objects or locations fixated during a learning period were repeatedly fixated during a recognition period by the same participant, and that recognition was facilitated by looking at previously fixed locations compared to a free-viewing condition.

Valuch, Ansorge, Buchinger, Patrone and Scherzer (2014) provide further evidence that attention is more strongly attracted by repeated visual content than by novel or surprising content using within scene and between scene cuts in videos. By using sports videos with deliberately inserted cuts, two conditions were created: between scene cuts with major visual changes and within scene cuts showing the same scene, action and actors. Two videos were presented at the same time on the screen. The target video was visually marked at the beginning and the participants were instructed to relocate their gaze if the video switched location. By analysing the saccadic reaction times (Deubel & Schneider, 1996), they find a quicker reorientation of attention if visual content is repeated after within scene cuts. Valuch et al. (2014) argue that this conflicts with the assumption that novel information is the best predictor of attention in videos as stated by Itti and Baldi (2009). Further, they remark that the advantage of a faster reorientation is only given at a short time frame following cuts where viewers search for familiar visual content. Since Itti and Baldi (2009) used a single video display, a recreation of the findings of Valuch et al. (2014) in a similar setting can be seen as a starting point for the experiments in this resarch paper.

Eye movements are described as being either saccades, where eyes are oriented towards regions of interest, or as stable fixations, during which the processing of visual information at that location takes place (Hoffman & Subramaniam, 1995). Even though acuity is highest at the central point of fixation, the fovea centralis, extrafoveal processing is used to process contextual information as well as to guide saccades to the next fixation. The importance of peripheral information processing is demonstrated by Castelhano and Henderson (2007). Hoffman and Subramaniam (1995) analyzed saccadic eye movements and concluded that saccades executed to peripheral locations in the visual field involve orienting of attention to these locations prior to the executions of the saccades. Together with similar findings for a relationship between attention and pursuit eye movements reported by Kowler (1985) and the inability to orient attention to one location while simultaneously executing a saccade to another location, they suggest that orienting of attention is an essential component for the preparation or the execution of saccades. This

suggests that fixations in between saccades, when the eye movement is stopped for a period of time, can be used to determine what particular areas observers visually attend to.

There are limitations of the representational capacity of perception. Attentional selection is therefore often described as an active process. Folk and Remington (1998) as well as others demonstrated that distractors, not sharing identifying features with the target being searched for, are ignored, while similar distractors capture attention. Thus, attentional capture is said to be contingent on whether a visual stimulus possesses a goal-relevant feature. Various studies also show that this attentional selection can be restricted to different precise domains, enabling participants to efficiently select the targets. Moore and Weissman (2010) showed that, even in a demanding condition by using a rapid serial visual presentation task, selection is efficient. These findings were described as precise template hypothesis. Anderson (2014) on the other hand, shows that observers are unable to utilize a target template to efficiently discriminate the target from a nontarget feature when presented successively in time and argues that this failure to select the target, otherwise highly effective, points to nontarget rejection as an explaining theory instead of target matching.

The present study takes a more similar approach to the experiment's setting of Itti and Baldi (2009) with a single video shown on screen, utilising novel, shifted and same conditions like shown by Ansorge et al. (2013) and cuts in short video clips (Valuch et al., 2014) to examine known features and surprise as predictors for attention by tracking eye movements of participants with different tasks in two experiments.

II Present Study

This study makes use of high definition video clips by dividing each clip into two smaller clips two-thirds in size. By showing these smaller clips fullscreen, it is possible to present a continued content on one half of the screen, by shifting either to the left or right side after a cut, while the other half changes to novel content, thus utilising a simple aproach to novel versus known content presentation. Scenes in a natural environment are used as content that is shown in a fullscreen computer setting. Two eye tracking experiments with differing tasks for the participants, were performed to test for differences and to show boundaries in known features versus novel features guiding attention in videos.

Figure 1 shows single frames as examples of the 240 video clips created and used for both experiments in this study. These frames are randomly selected and already cut to their final relative size.

III Experiment 1

The first experiment was used to test against the hypothesis of novel or surprising influences as best or equally good predictors for attention, compared by known features, in short video clips of natural environment settings.

III.I Method

Participants. Eye movements of 24 naive observers were recorded. The additional data of one participant was excluded due to calibration problems. The participants were students, whose ages were between 18 and 23 years (average of 19.8 years). All stated normal or corrected-to-normal vision. Informed consent was obtained from all participants prior to the experiments. Most students received test participation credits for University courses for their collaboration. Each subject watched 160 sequences of two short video clips each in one of the three main test conditions.

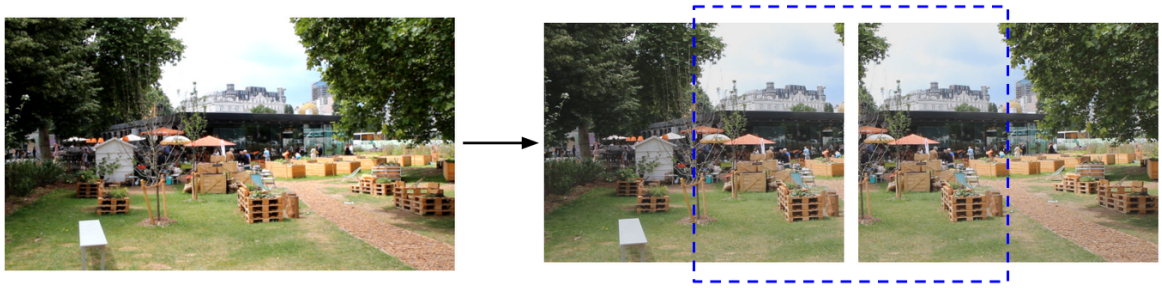


Figure 2 shows a single frame of source clip #3955 at the runtime of five seconds and the processed two target clips: left and right side with their overlapping halves.

Stimuli. The videos were cut to a duration of 10 seconds each. The motives were chosen to be at a wide angle with as little as possible artificial distractors like written advertisement signs or brand names. A tripod was used without any automatic or manual rotations or tilts to offer a steady picture. Further, most part of the pictures had to be in focus resulting in little blur within the videos. Movement in various areas of the picture

was anticipated like working people, moving cars or animals. The clips were created in and around the city of Vienna, Austria. From the high definition source videos with a size of 1,920 × 1,088 pixels parts to two-thirds with a size of 1,280 × 1,024 pixels (5:4 ratio) were shown from either the left or the right side of the picture. The output height was centered and cropped. 240 different videos were shot as source for the used 320 clips, 160 of which were pre-cut and 160 post-cut, as well as additional demonstration clips.

Setup. Eye movement data was recorded using an EyeLink 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 9-point calibration. Every time the clips changed, the exact timestamp was saved to the eye tracking data file, which allowed analysing the latency of the first fixation to the target video as well as the following ones for each video separately with millisecond precision. Recalibration checks were made after every tenth sequence. The stimuli were displayed on a 19 inch color CRT monitor (38 cm width) with a resolution of 1,280 × 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 and the Eyelink toolbox (Brainard, 1997; Pelli, 1997). Viewing distance to the monitor was 72 cm supported by chin and forehead rests.

Procedure and design. 160 sequences in random order with two clips each were presented to every participant after a demonstration including six sequences. The participants were instructed to watch and enjoy those impressions and that there are some questions in between to keep their attention. Additionally, it was said that it is not a problem if they did not answer correctly.

Half (80) of the sequences included a completely novel second video clip, while the other half continued either with the whole clip or shifted, meaning the continuation of half of the clip on the other side of the clip that was presented before the cut. Within these, shifted clips were shifted either to the right (20) or to the left (20) and the last 40

showed the same content onward. Thus, in the second condition 50 percent of the clips content was presented a second time but without temporal repetition. The same condition basically showed one video cut into two halves in time. Following 40, 80 and 120 sequences a break could be taken.

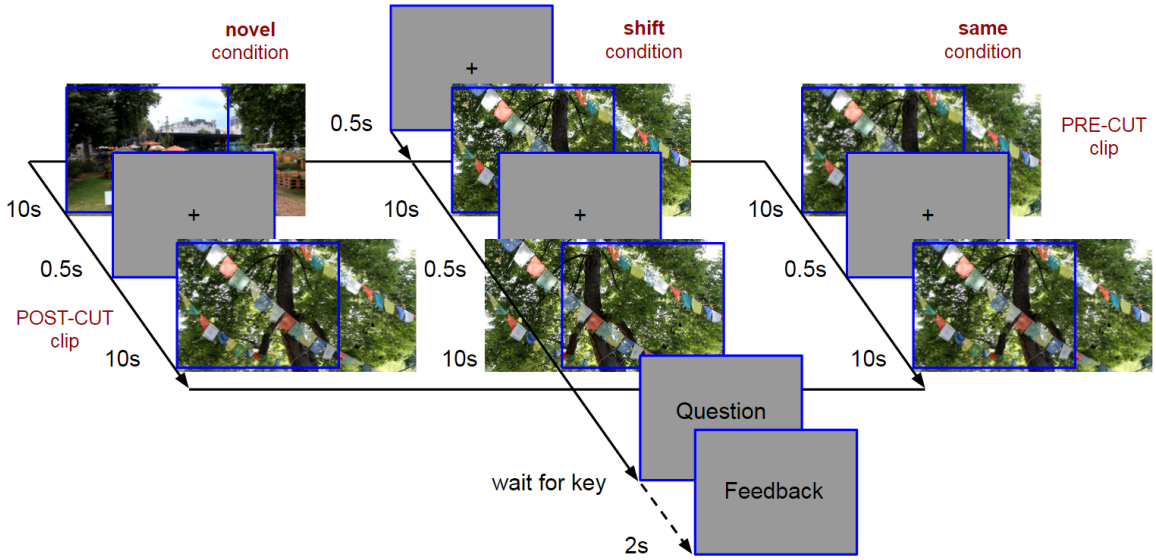


Figure 3 illustrates the process of one trial and the used sequence conditions: novel, shifted or same content video clips. Each clip is continued for 10 seconds (*POST-CUT clip*) after showing a fixation cross (cut) for 0.5 seconds following the pre-cut clip (*PRE-CUT clip*). Four different sets were created, as the sequences were randomised within each subject as well as balanced between the conditions resulting in pairs to six participants sharing sequences with clips in the same conditions.

After every clip sequence, a question was asked to keep the attention of the observer. This question asked for key number 2 on a standard keyboard, if two clips with different content were shown. In case of similar content, meaning parts of the clips showing the same content or both clips showing the same content, key number 8 had to be pressed. Therefore, the correct answers were balanced between the two keys. The original question can be found in the appendix. Feedback was only given for incorrect answers implicating a delay of 2 seconds. The completion of a single sequence took in average 25

seconds, the total test completion time being about 80 minutes.

Analysis. The area of interest (AOI) in this experiment was the video clips overlapping parts, more to the point: the side on which fixations were made within the different conditions. Fixations within two degrees (Henderson, 2003) from the horizontal middle of the videoframes were ignored in both directions (to the left and to the right) for certain statistical analyses to account for possible calibration inaccuracies.

$$\frac{\text{clip width px}}{2 \times \left(\arctan\left(0.5 \times \frac{\text{screen width cm}}{\text{distance to screen cm}} \right) \times \text{deg} \right)} \times 2 \qquad \frac{1{,}280}{\arctan\left(0.5 \times \frac{38}{72} \right) \times 57.296} \approx 86.587$$

Figure 4. Subtracting 87 pixels as 2 degrees correction leaves a horizontal AOI of 553 pixels. These defined the AOI side for left shifted clips on the right side (727 – 1,280 pixels) and for right shifted clips on the left side (1 – 553 pixels). Therefore, 80 clips can be analysed on their right- and 80 on the left-hand side. See figure 2 for a right shifted example.

Fixations were extracted using the edfImport library (Pastukhov, 2010). Since movements within the clips questioned the use of fixations as appropriate measure, preanalyses with all tracked data coordinates were made showing insignificant differences. As time limit the first five seconds into the second clip were chosen. For the determination of the viewing direction prior to the cut, the last five seconds in the pre-cut clip were evaluated. Subsuming, eye movement data was preprocessed in MATLAB and statistical tests were run in R (Cabakoff, 2011). For all statistical tests, α was set at 0.01.

III.II Results

At first, the data was analysed for subject and trial related deviations. The mean for incorrect answers to the question following each of the 160 clip sequences was 1.46 percent (sd=1.18). For all subjects the average response time was below 2.41 seconds. With one exception where an instructed pause was made, all slow response times were

below 7.95 seconds. The 16 clips that are going to be used for the memorising task in the second experiment were excluded to provide a better comparability. For details to these clips see experiment 2 and figure 8. All other trials were used for the following analyses when not stated differently.

III.II.I Fixation frequencies

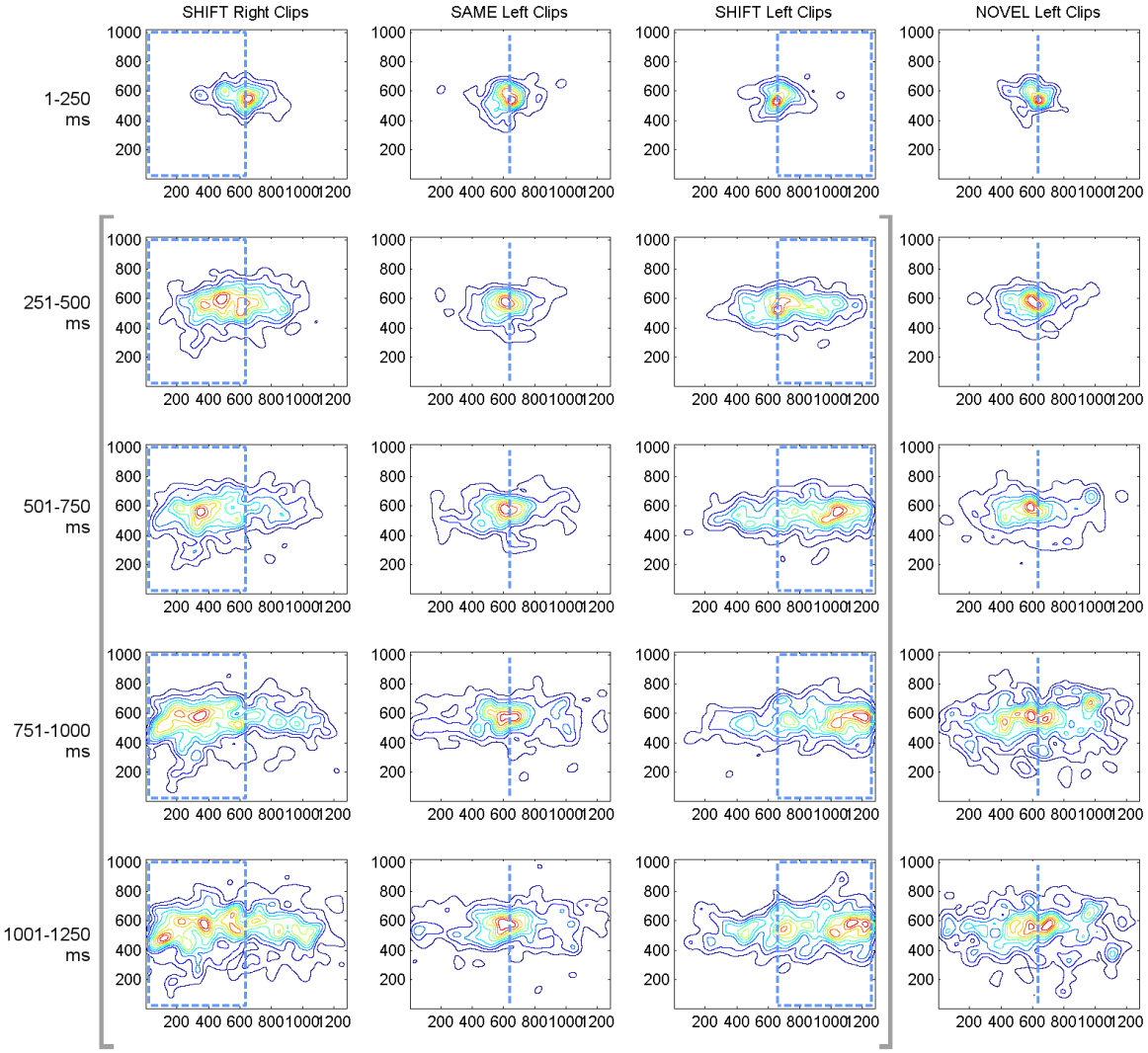


Figure 5 shows heat maps for the fixations within the first 1,250 ms in the four test conditions. The heat maps use the same coordinates (1,280 x 1,024) as the presented clips. Fixation data from the right shifted clips are shown in the first column, the left AOI containing repeated content. In the other columns the clips cut on the left side are shown

in the conditions: same or continuous content, shifted left and novel content. The time segments are post-cut - after the cut. Deviations of a center bias, as seen in the same and novel conditions, can be seen in the rows three to five (from 501 to 1,250 ms) in both shift conditions.

Within a five second time frame the participants made an average of 13.24 fixations (sd=3.56). For the statistical analysis the first three seconds are split into 250 ms segments. Within 250 ms 0.68 fixations were made in average per participant depending on the time past after the cut. In the first 250 ms the fewest fixations were made (mean=0.37), followed by the 250 ms were the most fixations were made (mean=0.98). T tests with Holm-Bonferroni correction were used to calculate the statistical significances. Since each clip was only shown as either left shifted or right shifted, both conditions are compared separately. Means for the horizontal fixations as well as the sum of their durations for each condition and participant can be found in the appendix.

Pairwise comparison using t tests with Holm-Bonferroni correction

		Time after the cut in ms											
		1-250	-500	-750	-1000	-1250	-1500	-1750	-2000	-2250	-2500	-2750	-3000
right	Shift/Novel	1,000	**,001** *	**,000** *	**,000** *	**,000** *	**,000** *	,011	,104	,109	,877	1,000	1,000
	Shift/Same	,959	**,000** *	**,000** *	**,000** *	**,000** *	**,005** *	,185	1,000	1,000	1,000	,972	1,000
	Novel/Same	1,000	,366	,860	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
left	Shift/Novel	,511	**,000** *	**,000** *	**,000** *	**,000** *	**,000** *	,035	,138	,026	,289	,070	,022
	Shift/Same	1,000	**,000** *	**,000** *	**,000** *	**,000** *	**,000** *	,022	,536	,094	,063	,059	**,001** *
	Novel/Same	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Table 1. The results of the pairwise comparison using T tests with Holm-Bonferroni correction show that average fixations on the horizontal axis differ significantly (* $p < .01$) for the shifted compared with both same and novel conditions in both right frame and left frame clips within a timespan of 251 to 1,500 ms after the cut. As it can be seen in the next figure, both sets of clips (right and left) show similar differences away from the center of the videoframes. Additionally, fixation means within 2,750 to 3,000 ms show also significant (* $p < .01$) differences in the left frame clips when shift and same conditions are compared. Only half of the novel trials are used in these comparisons. This was to ensure the use of the same amount of data since these trials were doubled for the balancing of the correct answers to the trial questions.

To estimate the effect sizes, correlation coefficients were calculated for the significant conditions. See the appendix for the used calculation method. These were medium (r = 0.41 to 0.47) for fixations within 251 ms and 500 ms (Shift/Novel right) and 1,251 ms to 1,500 ms (Shift/Same right) and large for the other significant time frames and conditions (r = 0.50 to 0.93). For mean values and deviations, see also appendix: Descriptive statistics experiment 1.

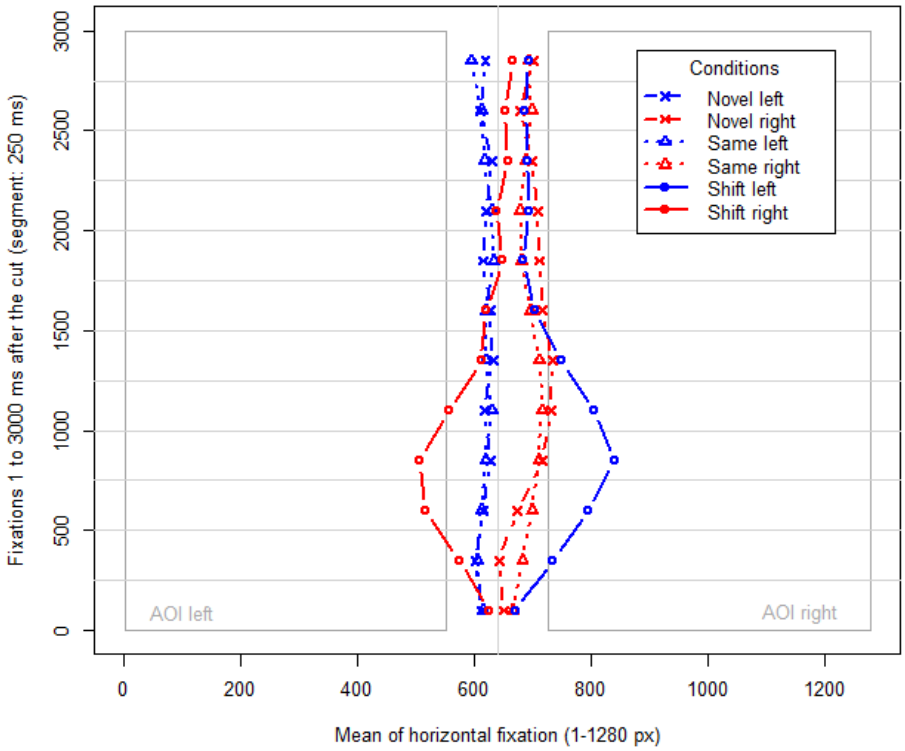


Figure 6 gives insights into the directions of the participant's mean horizontal fixations. The AOI on the right side attracted more fixations in the shifted left frame clips (blue line) - where the right side was presented a second time - in the time frame of 251 to 1,500 ms. A similar increase in fixations is found in the left AOI when shifted right frame clips (red line) were shown. The fixations in the same and novel conditions are more consistent within 2 degrees off the clips centers in both directions. Like stated at the statistical comparison before, only half of the novel trials are used in this figure to visualise the same amount of data as received for the other conditions.

III.II.II Latency of first fixation

As second dependent variable the latency of the first fixation on either AOI was analysed. Since it took up to 4,997 ms for the longest fixation to reach a specific AOI, outliers were removed using 1.5 times the interquartile range as limit one time only. Additionally, values of sides that were not fixated within the 5 seconds analysed or happened within 2 degrees from the center were not used. As a result data from 9557 fixations or 20.9 percent were excluded, taking participants on average 865 ms to attend to the AOI on each side at least once (sd=714).

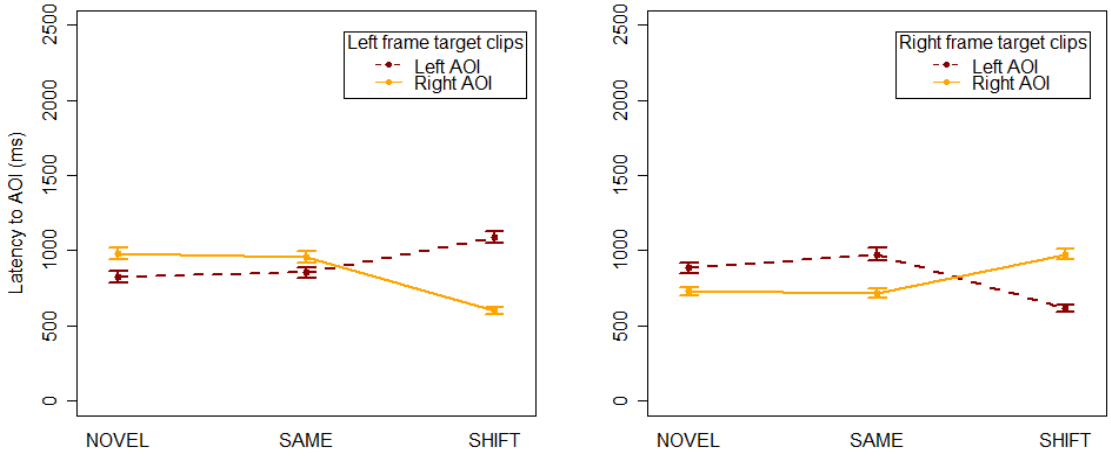
Pairwise comparison using Wilcoxon signed rank test with Holm-Bonferroni correction and descriptive statistics

target clips		left	right
Right frame	Shift – Novel	.000 *	.000 *
	Shift – Same	.000 *	.003 *
	Novel – Same	.331	.509
Left frame	Shift – Novel	.012	.000 *
	Shift – Same	.007 *	.000 *
	Novel – Same	.509	.640

left	m	sd		right	m	sd
**	617 ms	500 ms	Shift		976 ms	703 ms
	976 ms	759 ms	Same		715 ms	621 ms
	885 ms	696 ms	Novel		730 ms	531 ms
	1088 ms	766 ms	Shift	**	600 ms	451 ms
	853 ms	687 ms	Same		955 ms	714 ms
	826 ms	682 ms	Novel		980 ms	681 ms

Table 2 lists the comparisons of the average latencies until the first fixation into the left or right AOI was made. Significant differences were found for both shift to same comparisons as well as for shift to novel comparisons for the right target clips and the right AOI for the left frame clips (* $p < .01$). Although, the left AOI of the left target clips would also be significant for the shift and novel comparison at $p < .05$. Importantly, the latencies are shorter in the shift conditions only for the repeatedly shown side (**). In the right frame clips the repeatedly shown AOI was fixated in average at 617 ms and in the left frame clips at 600 ms. The mean latencies of the novel, thus not repeated, side in the shift conditions are among the highest values with 976 ms for the right side in the right shifted clips and 1088 ms for the left side in the left shifted clips.

The following figure visualises the compared data of table 2 showing the interactions of the three main conditions. Right and left frame target clips were separated to show distinct differences.



In figure 7 the latencies to both sides of the AOI in the post-cut clip (*Right AOI* as solid line, *Left AOI* as dashed line) are shown for the left frame and right frame clips separately. For the repeatedly shown side in both clip sets, e. g. the right AOI in the left frame clips and the left AOI in the right frame clips, lower latencies can be seen. The scale includes 0 to 2,500 ms.

IV Experiment 2

The second experiment tests if the results of the preceding experiment can be replicated when the participant's search task is changed to a memory task over the whole duration of every testing. Instead of evaluating the post-cut clips in reference to the pre-cut clips like in experiment 1, the used task is referring to pre-cut clips only secondarily together with the post-cut clips in reference to memorised clips at the start of every testing. The hypothesis is that, either the results can be replicated displaying an effect stronger than the influence of these tasks, or different results can be seen, suggesting that the actual task plays an even more important role.

Figure 8 displays single frames of the 16 randomly selected video clips for memorising. These clips were used for all participants but according to the conditions of each participant's clip set.

IV.I Method

Participants. In the second experiment eye movements of 24 naive observers, none of which took part in the first experiment, were recorded. The additional data of eight participants was excluded: four due to calibration problems, the other four to retain balanced sets. The participants were students, whose ages were between 19 and 32 years (average of 24.6 years). As in the first experiment the participants stated normal or corrected-to-normal vision, gave informed consent and received test participation credits for University courses. This time, each subject watched the 160 sequences of two short video clips each following 16 of those clips that were to be memorised in advance. The setup was the same as in the first experiment.

Stimuli. The same source video clips as in the first experiment were used. The duration was shortened to five seconds for each clip using the same parts that were analysed in the first experiment. Thus, the first five seconds of the post-cut clip were used, whereas the last five seconds of the pre-cut clip were used. The 16 clips used for the memorising task were chosen randomly but balanced over the conditions including the left and right shifted parts. Each of these clips was shown for five seconds and followed by a fixation cross for 1.5 seconds at the beginning of each testing.

Procedure and design. Following a demonstration including two clips to memorise and six example sequences, every participant was shown the same 16 clips to memorise, followed by 160 sequences in random order with two clips each. The participants were instructed to watch those impressions and that their task was to guess the question's answers after each pre-cut, post-cut sequence. Additionally, it was remarked that it is not a problem if incorrect answers were given.

Since the sequences stayed the same but the playback time was shortened, only one break could be taken, after half of the trials. Following every clip sequence up to two questions were asked to keep the attention of the observer. For the original questions see the section: Statistical evaluation details in the appendix. The first question was to press key 8 if there was a memorised clip among the shown sequence or key 2 if new clips were seen. Was there a memorised clip with a correct answer the second question followed. Key 8 had to be pressed if the first (pre-cut) clip, key 2 if the second (post-cut) clip was memorised and thus shown before. Were both clips seen before (same condition) both answers were correct. Due to the use of only 16 memory clips, the correct answers were not balanced between the two keys. Feedback was only given for incorrect answers implicating a delay of 2 seconds. The completion of a single sequence took in average 14 seconds, the total test completion time beeing around 60 minutes.

Analysis. The definition of AOIs and the main dependent variables are the same as in experiment 1. For the extraction of the fixations the same methods were used and for

all statistical tests α was set at 0.01. The same time limit of five seconds was applied, matching the whole clips duration in this experiment.

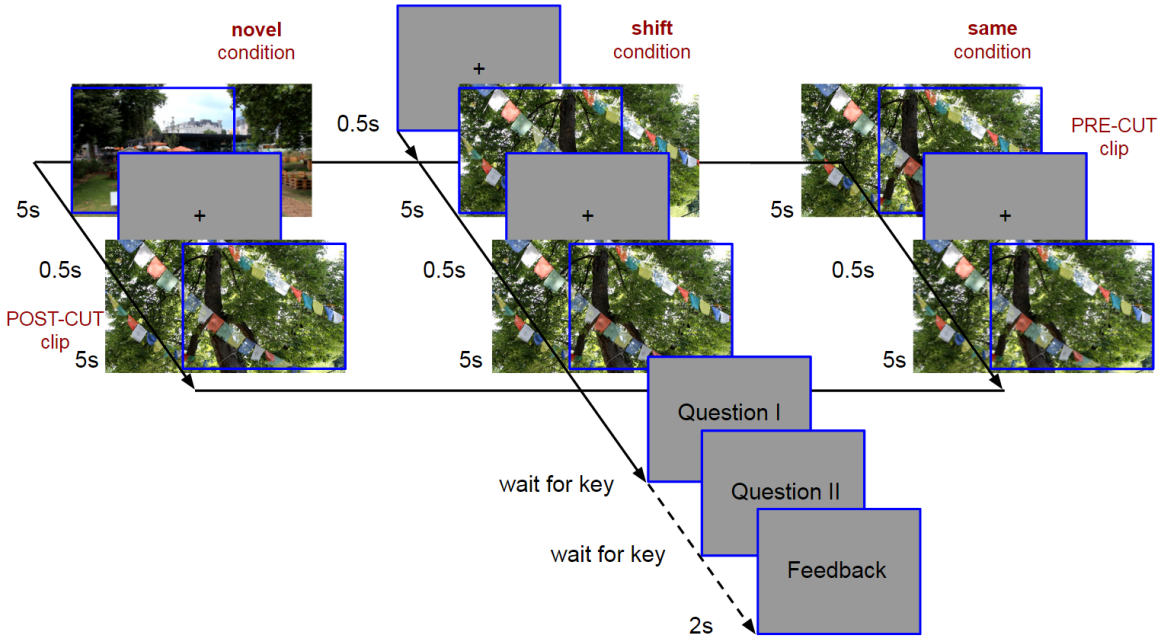


Figure 9 illustrates the trial procedure used in the second experiment for a right shifted clip. The conditions: novel, shifted and same content video clips were reused in the same sets as in the first experiment with the difference that only 5 seconds prior to the cut and 5 seconds after the cut were used from the presentation of the clips.

IV.II Results

Trial and subject related deviations were analysed. The mean for incorrect answers to the first question following each of the 160 clip sequences was 13 percent (sd = 3.89). It should be noted that, to keep the amount of clips to exclude low, the correct answers were not balanced unlike in experiment 1. For all subjects the average response time was below 0.98 seconds. All response times were below 27.6 seconds. The 16 clips used for the memorising task were excluded while all other trials were used for the following analyses.

IV.II.I Fixation frequencies

The frequencies of the fixations were analysed in the fashion of the previous experiment. After a visualisation of the experiment's data and its statistical evaluation for significances, experiment data from the first experiment will be added into the final analysis of both experiments.

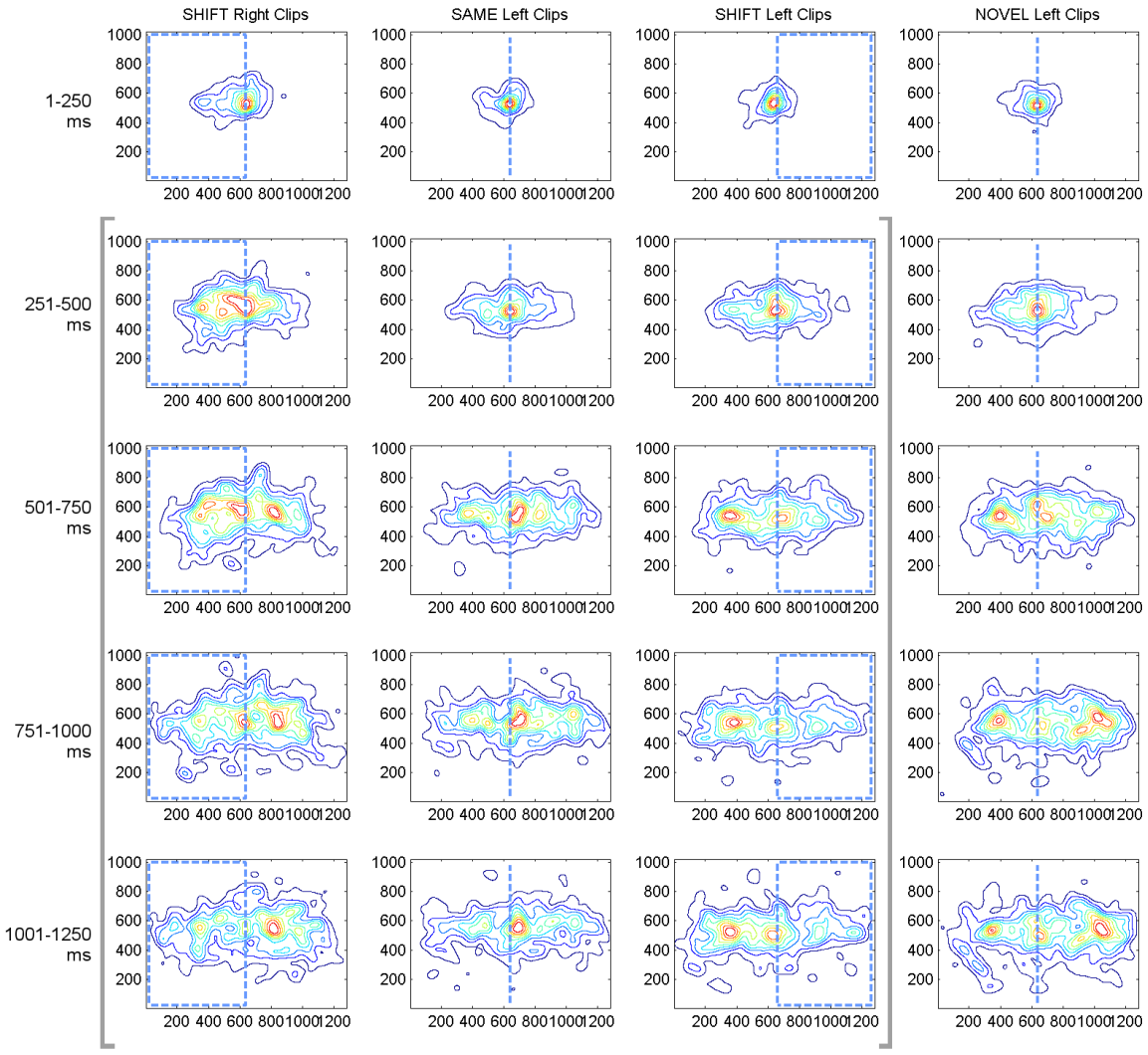


Figure 10 visualises fixations within the first 1,250 ms of the post-cut clips, in the test conditions, as heat maps including matching coordinates (1,280 x 1,024 pixels) with the presented video clips. AOIs are marked in dashed blue rectangles. This figure can be compared to figure 5 in the first experiment since the same trial data, although from different subjects, is used. The heat maps contain data from 20 trials, minus the ones

including the memory clips, from each participant in the conditions: shifted right, as well as same, shifted and novel of the shifted left clip set.

Within the five seconds time frame the participants made an average of 13.27 fixations. For the statistical analysis the first three seconds were split into 250 ms segments. Within 250 ms 0.7 fixations were made in average per participant depending on the time past, after the cut. Like in the first experiment, in the first 250 ms the fewest fixations were made (mean=0.45), followed by the 250 ms were the most fixations were made (mean=1.0). T tests with Holm-Bonferroni correction were used to calculate the statistical significances. Both clip sets, left and right, were compared separately. All means for the horizontal fixations as well as the sum of their durations for each condition and participant can be found in the appendix at: Descriptive statistics experiment 2.

Pairwise comparison using t tests with Holm-Bonferroni correction

		Time after the cut in ms											
		1-250	-500	-750	-1000	-1250	-1500	-1750	-2000	-2250	-2500	-2750	-3000
right	Shift/Novel	1,000	1,000	**,008** *	**,004** *	**,002** *	,028	,202	,467	1,000	1,000	1,000	1,000
	Shift/Same	1,000	1,000	,891	,671	,030	,014	,016	,915	1,000	1,000	1,000	1,000
	Novel/Same	1,000	1,000	,891	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
left	Shift/Novel	1,000	1,000	,034	**,000** *	**,006** *	,135	1,000	1,000	1,000	1,000	1,000	1,000
	Shift/Same	1,000	1,000	,166	,129	,288	,135	,647	1,000	1,000	1,000	1,000	1,000
	Novel/Same	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Table 3. The results of the tests show that fixations on the horizontal axis differ significantly (* $p < .01$) for shifted compared with novel conditions in the left frame clips within a time frame of 751 to 1,250 ms and for the right frame clips from 501 to 1,250 ms after the cut. In the next figure, the directions of the significant fixation mean values can be seen.

Next, effect sizes were calculated with correlation coefficients for the significant conditions. These were medium ($r = 0.39$ to 0.45) for fixations within 501 and 1,250 ms, exept for the results in the left shift to novel comparison from 751 to 1000 ms, which can be seen as a large effect ($r = 0.53$).

Overall latency data did not show similar differences within the conditions as calculated in experiment 1 and are therefore not reported here. The next comparison of interest were the shifted conditions in both experiments. To give a figurativ overview, the fixation data of the shift conditions of both experiments are represented in the next figure.

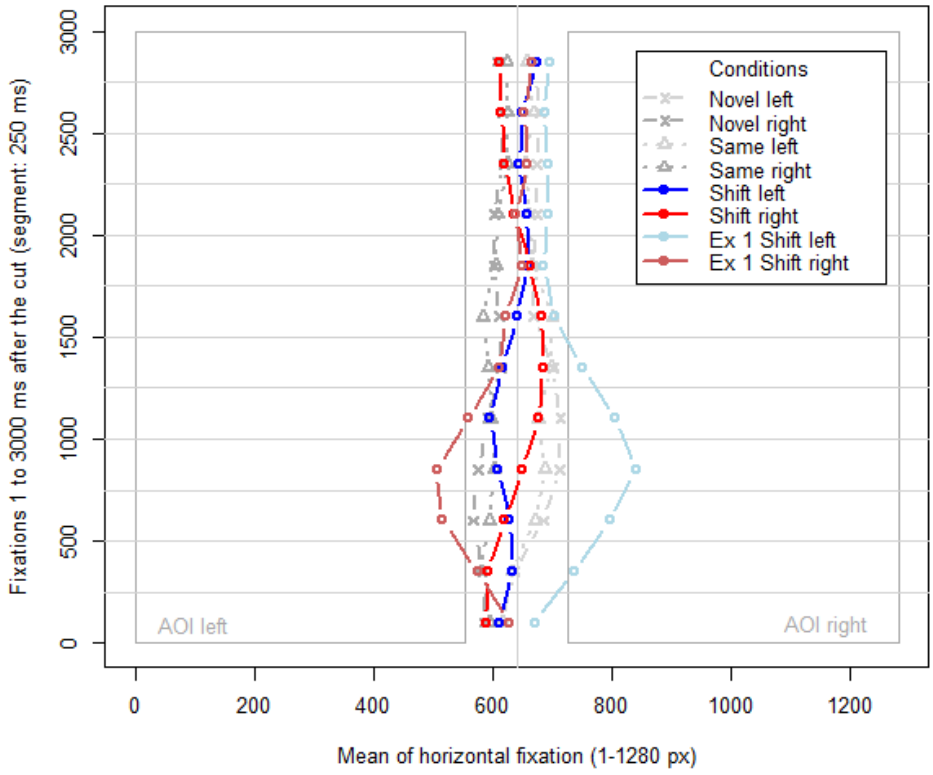


Figure 11 can be compared with figure 6 from the first experiment. The average horizontal fixations are drawn along the time frames from 1 ms to 3000 ms in 250 ms segments. Additionally to the conditions in both left and right clips, the fixations of the shifted conditions from the first experiment are shown (*Ex 1 Shift left* in lightblue color and *Ex 1 Shift right* in lightred). The tendency of the significances found in experiment 2 are in the opposite direction compared with experiment 1. The means from the shifted right conditions are higher with 619, 650 and 676 px compared with 567, 575 and 593 px of the novel condition whereas the shifted left values are with 608 and 596 px lower than 713 and 714 px in the novel condition. These values suggest a preference to the novel AOI instead of the already known, repeatedly shown AOI in this experiment.

IV.II.II Comparison of the experiments shift conditions

The next part of this study was a comparison of both experiments data as they use the exact same stimuli and conditions but differ in the tasks the participants were given. This was to analyse the influence of the actual task for keeping the participant's attention and find out possible limitations. The following table lists the results of the statistical comparisons. As a mental anchor point for the descriptive data: at 640 pixels was the exact center of the horizontal axis used. This analysis complements the already found significant differences in fixation means in both experiments.

Comparison using Wilcox test with Holm-Bonferroni correction

		Time after the cut in ms											
		1-250	-500	-750	-1000	-1250	-1500	-1750	-2000	-2250	-2500	-2750	-3000
shift left	Ex 1/Ex 2	,047	**,000**	**,000**	**,000**	**,000**	**,001**	,203	1,000	,589	,389	,740	1,000
shift right	Ex 1/Ex2	,058	,384	**,000**	**,000**	**,000**	**,005**	,103	1,000	,791	,323	,197	,019
shift left	Ex 1 sd	149,8	232,6	285,7	302,0	312,6	323,6	323,4	326,4	333,5	336,5	330,9	323,3
	Ex 1 m	670,2	**735,8**	**796,4**	**840,4**	**805,0**	**749,0**	703,3	684,3	693,5	692,3	686,5	694,5
	Ex 2 m	611,7	**633,2**	**628,4**	**607,9**	**595,6**	**617,1**	640,0	660,4	656,3	645,0	649,9	674,7
	Ex 2 sd	116,1	204,3	261,6	292,2	302,5	314,3	309,1	305,9	305,9	295,1	298,0	292,6
shift right	Ex 1 sd	142,8	223,1	273,0	308,5	323,0	333,9	338,5	342,0	339,8	334,3	333,3	326,3
	Ex 1 m	626,4	574,5	**515,6**	**505,6**	**557,9**	**612,1**	621,4	649,1	639,0	658,1	653,0	665,9
	Ex 2 m	587,7	592,1	**618,9**	**649,7**	**676,0**	**685,5**	681,1	662,0	636,3	619,9	614,1	612,1
	Ex 2 sd	133,3	201,7	250,3	283,3	297,5	294,4	285,4	297,4	299,8	299,5	295,9	287,0

Table 4 lists the between subject comparisons of the shift conditions in both experiments using Wilcox tests. Insights into the direction of these differences can be also found in the previous chart (figure 11). Below the p-values, the mean values as well as the standard deviations are listed for each 250 ms segment. Significant results ($p < .01$) are listed in boldface in both, the comparison and the descriptive analysis data. The fixations in both shift conditions, left and right, differ significantly from 501 to 1500 ms post-cut. Additionally, the left shifted clips differ also from 251 to 500 ms post-cut. The mean values in experiment 1 (*Ex 1*) show with 736 to 840 px as either higher (*shift left*) or with 506 to 612 px lower values (*shift right*) than the frame center. This is in contrast to the values of experiment 2 (*Ex 2*), which are with 596 to 633 px (shift left) and with 619 to 686 px much closer to the center of the video frames.

IV.II.III Visualisation and influence of fixations in pre-cut clip

At last, further visualisations were made to show the interactions of the different main conditions (shift, same, novel) along the timeline. To take the last fixations on the first clip into account, 2,000 ms prior to the cut - the last fixations in the pre-cut clips - were analysed. Only those trials were taken into account, where fixations were made counting to either the left AOI pre-cut, when more than 50 percent of the fixations were spent on this side, or to the right AOI pre-cut, if more than 50 percent of the fixations happened on the right AOI side of the pre-cut clip.

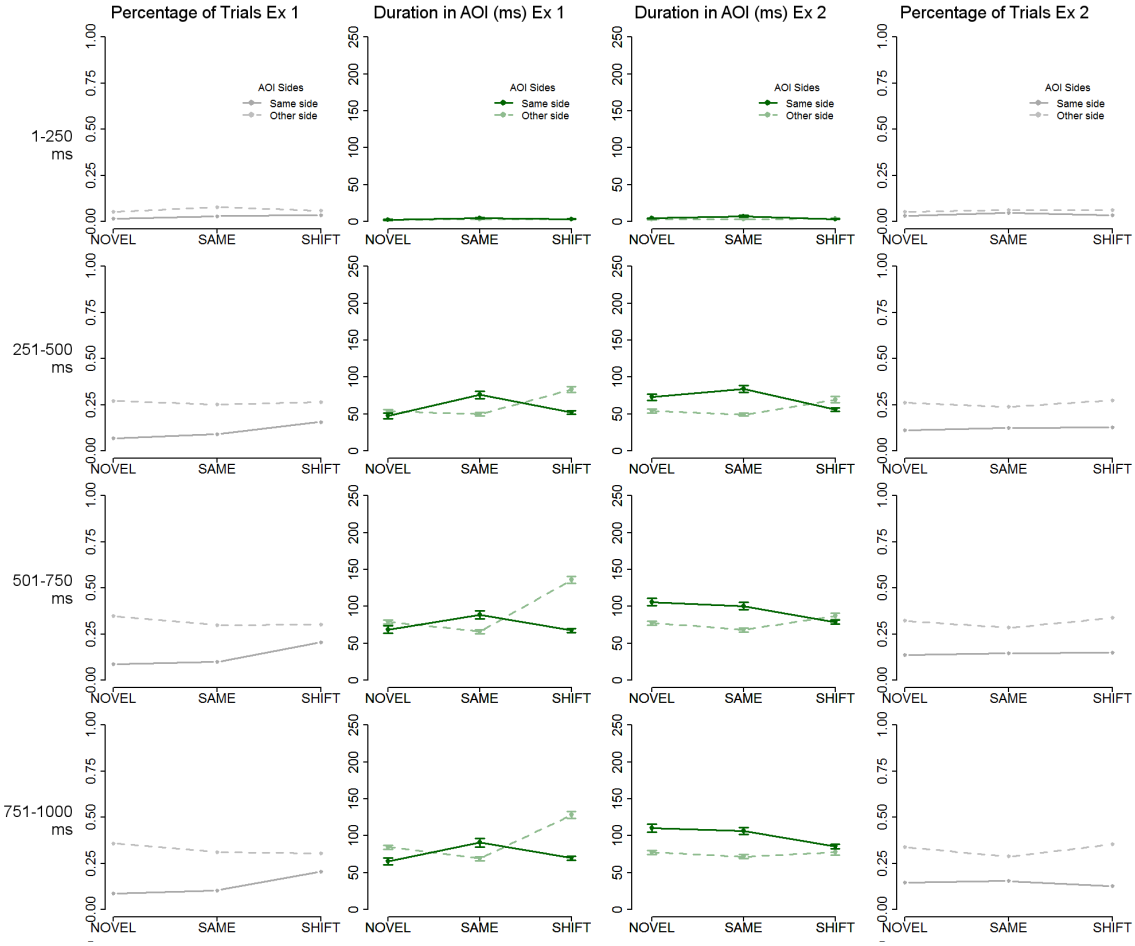


Figure 12 shows the first 1,000 ms in segments of 250 ms each. In the center columns the average duration spent in each condition can be found for the first experiment (*Ex 1)* and the second experiment (*Ex 2*). Next to the durations are the percentages of the trials included for each condition and time frame.

Differences in the novel conditions as well as the shift conditions can be seen between experiment 1 and 2, whereas similar results are found for the average duration of the fixations in the AOI, for the same condition. The following figure shows with 1,001 to 2,000 ms the data for the next second in the post-clip data.

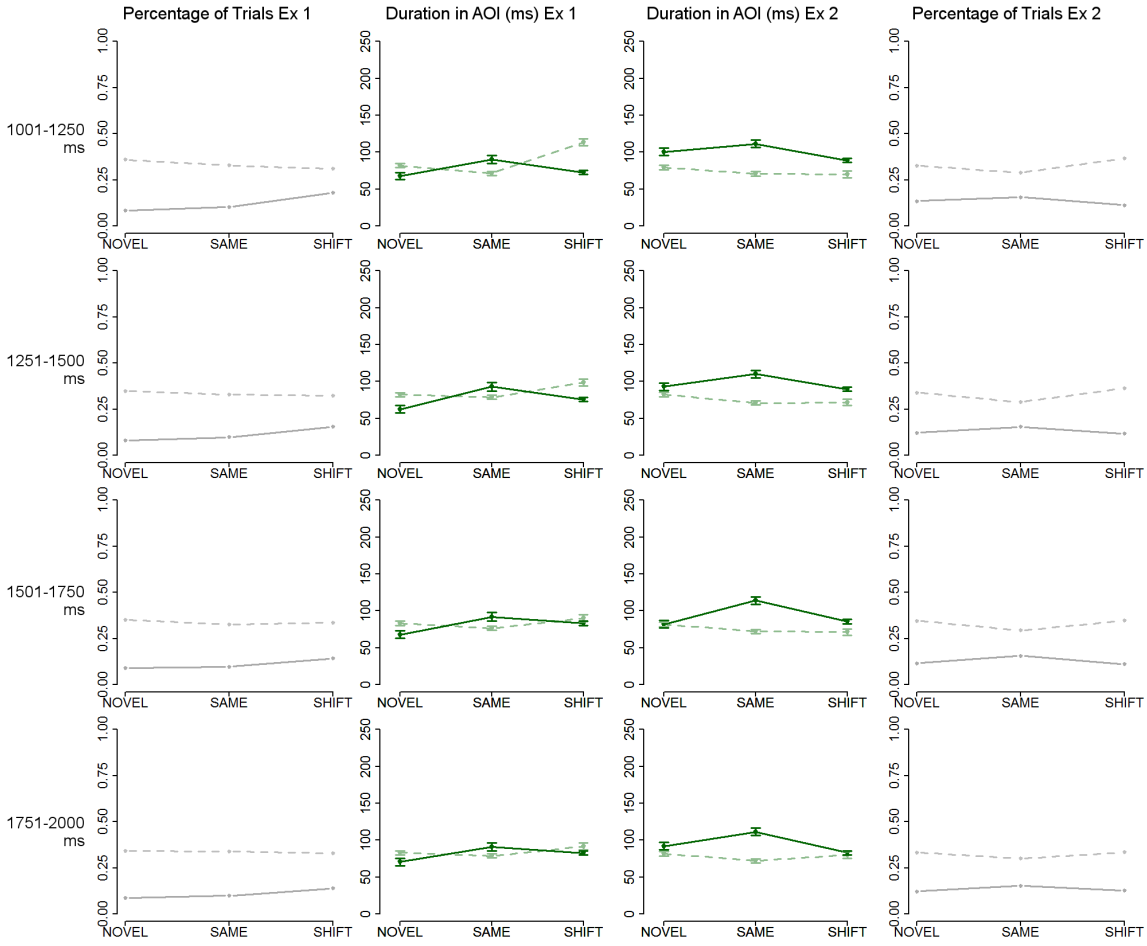


Figure 13. For additional labels: see previous figure. It is important to note that the other AOI side (dashed lines, *Other side*) in this and in figure 12 includes only data where the right AOI was watched in the right clips pre-cut, and then switched to the left AOI post-cut, or where the right AOI was watched post-cut, after looking at the left AOI in the pre-cut clip for the left clips, representing continued content for the shift conditions (*SHIFT*). All other durations in the shift condition count to the same AOI side (solid lines, *Same side*). For the novel and same conditions, only durations of fixations where the right AOI was focused in the pre-cut as well as the post-cut clip for left clips, and where the

left AOI was focused in the pre-cut and post-cut clip for right clips, belong to the same AOI category. Nevertheless, there is no continued content for the novel conditions. The rest of the novel and same condition durations count towards the other AOI side data. Thereby lies the focus for these figures on the same side (solid line) for the novel and same conditions, but on the other side (dashed line) for the shift conditions. The percentages indicate that more trials fell in the other AOI side condition, which has to be noted for the shift condition.

Since these last visualisations include only a small percentage of all trial data, they are used for descriptive purposes and generation of further questions only. Removed trials include: trials with memory clips shown, trials where only center (not AOI) fixations in pre-clips happened up to 2,000 ms prior to the cut, as well as trials where only center fixations were recorded in the post-clips' 3,000 ms analysed.

V General Discussion

In two experiments the horizontal fixations of participants, the fixations' latency periods as well as their duration were analysed during the presentation of short video clips shot in natural environments. Sequences of two clips each, separated by a short cut, were used in the three different conditions: same, novel and shifted content. By using short clip runtimes of 10 seconds in the first and five seconds in the second experiment for each pre- and post-cut clip, and the use of the same clips in different conditions, further influences were tried to be kept under control.

In the first experiment significant differences for the average participant's fixations when viewing continued, already known, against novel, for the first time presented, content were found between 251 and 1,500 ms after the cut. Both: known and novel content were shown and analysed in shifted clips where half of the clip (one side) included novel and the other half already seen content. Those differences were controlled by the same condition that showed similar differences to the shifted but no significant differences to the novel condition. By seeing effects within the first 1,500 ms using a within subject design for two different sets of videos (left and right shifted content), we can assume that attention was at first directed to known content, following a goal-driven approach. The preference for fixations at the shift conditions continued video parts (AOI) show similar results as already shown for photographs of natural scenes (Ansorge, Becker, & Valuch, 2013). Furthermore, the significant differences in the latencies of the first fixations support the findings of Valuch et al. (2014) who reported faster reaction times in within scene versus between scene cut video sequences. The results also support the findings of Carmi and Itti (2006) who showed that perceptual memory was utilized across 7 consecutive saccades within 2.5 seconds, even in the absence of visually guided actions. Certain influences, like surprise (Itti & Baldi, 2009) may be used to explain parts of the directed attention to novel content.

One open question was that of the influence of the used task which could be seen as simple search task within each sequence by the question used for keeping the

participants' attention. Stark and Ellis (1981) showed that the same fixations on specific objects or locations made during a learning phase would be repeated during image or scene recognition. On the other hand, recognition might be accomplished with a single fixation (Helmholtz, 1867). Thus, for analysing the tasks influence, a second experiment was conducted. It was designed to ensure the comparability between the different clip sets, right and left as well as to find limitations for the used search task. To shorten the duration of the experiment, only the five seconds of each clip, that were to be analysed, were used. The task was changed to a search task where each clip had to be compared to memorised clips shown at the beginning of the experiment. The results of the first experiment could not be found in the second experiment. Unexpectedly, even slightly reversed effects were found. Between 501 and 1,250 ms after the cuts, significant preferences for novel content can be seen. This shows a much higher influence of the nature of the task itself, that has to be done by the participants, then expected.

In a more detailed analysis including the fixations of the pre-cut clip, similar preferences to the same side, reported in other papers (Foulsham & Underwood, 2008; Noton & Stark, 1971), were found in the same versus novel condition in the first experiment (see figure 12). More precisely, when continued content was shown, the attention was directed to the side showing that content again. Contrary, in the second experiment the side of the attention pre-cut was more likely predicting the side in both: novel and same condition.

For more insights into the interaction between such bottom up processes and this goal directed approach, further analysis is needed. Post hoc tests for between subject differences and differences between clips could lead to insights into what other aspects catch the viewers' attention. Two factors are imaginable in this scenario: an amount of surprise coming from the video clips content, and a specific state of awareness of the perceiving person. These are but two factors that could enable a more detailed prediction for the visual attention of viewers.

At last, significant video differences could exist through their composition alone. Pereira and Castelhano (2014) analysed interactions of scene context and object contents in computer generated scenes finding differences for the amount of objects presented. Further analysis for two separate sets of videos, containing many or few objects could be done. Alternatively, the experiment could be repeated with left shifted changed to right shifted video sources and right shifted changed to left shifted output clips. This would not only lead to a higher sample size, but also make both shift conditions more comparable to each other.

VI Conclusions

This study shows that viewing video clips can be a highly task specific process for attention even on a very low, seemingly automatic level. Further, the usage of cuts and shifted conditions proved to be efficient for this research. More experiments are needed to find out further details about the settings and tasks in which there are influences of known or similar content in videos and cuts within those. A free viewing condition in comparison could also contribute to the findings, albeit would be harder to hold valid. Memorising clips deserves also more attention and could be taken into account as second part to an experiment where clips were shown before in a free viewing condition or similar.

An example of a follow-up study could start by preselecting the clips used here for their averagely provoked fixations. Those with evenly distributed fixations on the left and right video side, could be used in both conditions to make the right and left shifted clips better comparable. Furthermore, a quarter of the clips could be used as memorised clips or already known content by splitting an experiment in two halves. This would lead to two different tasks within the same goup of participants and thus, a better comparison by another within subject analysis of two or more shifted conditions.

At last, video makers have to take into account that the viewing experience can be guided and influenced by instructions. Moreover, even though focusing attention at novel content might be seen as stimulus-driven, parts of it might be just one form of automatic goal-directed attention to differences like also shown by Chapman et al (2014). Sometimes guiding attention might be especially necessary, e. g. when specific information is needed to get across to the viewer. Nevertheless, the instructions might be as well part of the video themselves. Future studies are to be expected as the entertainment and advertisement branches advance and video sharing platforms grow.

VII References

Anderson, B. A. (2014). On the Precision of Goal-Directed Attentional Selection. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1755-1762.

Ansorge, U., Becker, S. I., & Valuch, Ch. (2013). Priming of fixations during recognition of natural scenes. *Journal of Vision*, *13*, 1-22.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436. http://psychtoolbox.org.

Carmi, R., & Itti, L. (2006). The Role of Memory in Guiding Attention during Natural Vision. Neuroscience Program. *Journal of Vision, 6*, 898-914.

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12*, 1017–1040.

Chapman, C., Foulsham, T., Kingstone, A., & Nasiopoulos, E. (2014). Top-Down and Bottom-Up Aspects of Active Search in a Real-World Environment. Canadian *Journal of Experimental Psychology, 68*, 8-19.

Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research, 46*, 4118–4133.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*, 1827-1837.

Enns, J. T., Los, S. A., Olivers, Ch. N., Schreij, D., & Theeuwes, J. (2014). The Interaction Between Stimulus-Driven and Goal-Driven Orienting as Revealed by Eye Movements. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 378-390.

Ferscha, A., Paradiso, J., & Whitaker, R. (2014). Attention management in pervasive computing. *Pervasive Computing, IEEE*, *13*, 19-21.

Folk, C. L., & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: Evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 847–858.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision, 8*, 1–17.

Helmholtz, H. (1867). Handbuch der physiologischen optik. *Allgemeine Enzyklopädie der Physik* (Vol. 9). Leipzig, Germany: Voss.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences, 7*, 498–504.

Higgins, E., Leinenger, M., & Rayner, K. (2014). Eye movements when viewing advertisements. *Frontiers in Psychology, 5*, ArtID 210.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics, 57*, 787-795.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research, 49*, 1295–1306.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*, 1489–1506.

Moore, K. S., & Weissman, D. H. (2010). Involuntary transfer of a top-down attentional set into the focus of attention: Evidence from a contingent attentional capture paradigm. *Attention, Perception, & Psychophysics, 72*, 1495–1509.

Noton, D., & Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11*, 929–942.

Pastukhov, A. (2011). edfImport library v1.0. http://kobi.nat.uni-magdeburg.de/edfImport

Pereira, E. J., & Castelhano, M. S. (2014). Peripheral guidance in scenes: The interaction of scene context and object content. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 2056-2072.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Kabacoff, R. I. (2011). *R IN ACTION – Data analysis and graphics with R*. Manning Publications Co. Shelter Island, NY.

Kowler, E. (1985). Smooth eye movements as indicators of selective attention. In M. I. Posner & O. S. M. Marin (Eds.), *Mechanisms of attention: Attention and performance XI*, 285-300. Hillsdale, NJ: Erlbaum.

Salomon, D. (2004). *Data compression: The complete reference.* New York, NY: Springer.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences, 9*, 16–20.

Stark, L. W., & Ellis, S. R. (1981). Scanpaths revisited: Cognitive models direct active looking. *In Eye movements: Cognition and visual perception (1981)*, 193–226.

Valuch, C., Ansorge, U., Buchinger, S., Patrone, A., & Scherzer, O. (2014). The Effect of Cinematic Cuts on Human Attention. In: *TVX '14: Proceedings of the 2014 ACM International Conference on Interactive Experiences for Television and Online Video.* ACM, New York, NY, USA.

A Appendix

A.I Statistical evaluation details

Outlier calculation method. To remove outliers, the following R statistics method was used with a maximum of 1.5 times the interquartile range.

```
remove_outliers <- function(x, na.rm = TRUE) {
  qnt <- quantile(x, probs=c(.25, .75), na.rm = na.rm);
  H <- 1.5 * IQR(x, na.rm = na.rm);
  y <- x;
  y[x < (qnt[1] - H)] <- NA;
  y[x > (qnt[2] + H)] <- NA;
  y;
}
```

Correlation coefficients calculation method. This method was used in R statistics to calculate the correlation coefficients for estimating effect sizes. The variables d$avFix, d$cond, cond1, cond2 and len have to be set accordingly.

```
# d <- read.csv(...)
# len <- length of cond1+cond2
t <- pairwise.t.test(d$avFix, d$cond, p.adjust.method="holm", paired = TRUE)
pvalue <- t$p.value[cond1, cond2]
r <- (qnorm(pvalue/2) * -1) / sqrt(len)
```

Experiment recorded participant data. The following values were recorded for each participant before the experiment started.

subject_no	the continuous participant id for anonymisation purposes
set_no	set of clips to use (1-4) deciding the clip/condition combination
subject_age	the participant's age in years
subject_gender	the participants gender (1=male, 2=female)
subject_glasses	if glasses were worn (1=no, 2=glasses, 3=contact lenses)
eyeTracked	which eye was tracked (1=left, 2=right)

Experiment Questions. Stated are the directive used in experiment 1 after each clip sequence and the up to two questions following the sequences in experiment 2, as well as their possible answers. These questions are in german language.

Experiment 1 – Directive:

"Bitte drücken Sie die Taste 2 wenn es sich um verschiedene Clips handelt."

"Drücken Sie die Taste 8, sollten die jeweiligen Clips gleich oder ähnlich sein."

Experiment 2 – Questions:

"Befand sich ein eingeprägter Clip unter den beiden Clips?"

Taste 8 für JA

Taste 2 für NEIN

"Befand sich ein eingeprägter Clip unter den beiden Clips?"

Taste 8 für den ersten Clip

Taste 2 für den zweiten Clip

Descriptive statistics experiment 1. These tables state the mean horizontal fixation values in pixels (*fix, m*), the standard deviations (*sd*) and the total of fixation duration in ms (*dur*) for the three conditions (*same, novel, shift*) in both, left and right frame, clips.

	left frame clips						right frame clips					
	same		novel		shift		shift		novel		same	
subj	fix	dur	fix	dur	fix	dur	fix	dur	fix	dur	fix	dur
1	610,3	43615	652,4	37353	768,2	40659	553,6	40796	724,9	43988	701,2	41851
2	669,6	43149	653,1	36278	861,1	43948	636,9	45742	704,8	42128	788,9	39572
3	649,0	40513	623,3	36726	710,7	39641	648,9	39427	719,7	36516	626,6	40245
4	654,7	37096	729,0	36340	865,8	36910	620,0	39240	702,9	37161	759,1	37956
5	646,0	37050	633,0	36930	798,7	37373	598,0	39529	786,0	40370	762,8	41340
6	661,7	42986	580,7	36964	711,1	41598	570,7	44742	674,4	41249	678,0	44754
7	591,0	42715	580,3	38568	591,9	42584	550,3	41353	710,4	38704	581,2	41337
8	604,5	39046	606,8	36679	796,5	39304	534,3	39532	675,1	40521	640,7	40745
9	564,3	43541	657,8	39825	792,2	40788	661,5	40777	772,8	43420	776,6	42361
10	617,6	40870	587,4	35380	680,8	40744	714,6	40930	731,1	40111	732,0	41532
11	603,5	45092	584,4	39681	735,9	42538	587,7	43692	711,9	39876	598,7	43158
12	597,3	36409	659,1	33750	717,4	34843	601,7	39510	685,4	33437	667,0	28783
13	658,9	42481	604,5	40229	805,2	42725	562,3	42762	694,2	43506	748,4	42167
14	628,2	45658	641,0	36408	673,3	43694	593,3	43824	718,7	42256	725,0	45365
15	620,1	47453	608,9	40434	727,3	45258	673,2	46814	702,5	41653	667,3	48064
16	584,4	42177	583,5	39524	631,4	40759	608,8	40634	685,7	43029	715,1	43096
17	601,3	41515	588,5	39095	809,0	39802	519,3	41061	689,7	44098	672,9	42601
18	683,5	41785	651,2	37901	757,3	42522	584,4	41778	716,8	42959	714,3	44096
19	578,4	46012	648,3	38419	661,8	44041	681,8	44800	657,4	38218	649,0	44284
20	621,8	39779	593,1	37401	781,0	39124	616,7	40786	653,0	40798	737,1	41908
21	597,8	34597	678,6	32810	700,4	38117	608,5	38866	743,9	34800	671,3	33900
22	644,8	44002	620,6	34290	669,8	43683	656,2	43389	689,6	40020	700,4	43553
23	583,4	40029	579,4	37846	663,9	41022	598,2	42382	659,2	38182	645,8	41992
24	561,0	40481	563,0	39651	737,1	41428	511,6	40447	626,0	40059	770,1	40923
	618,1	41585,5	**621,2**	37436,8	**735,3**	40962,7	**603,9**	41783,9	**701,5**	40294,1	**697,1**	41482,6

	left frame clips									right frame clips								
	same			novel			shift			shift			novel			same		
t	m	sd	dur	m	sd	dur	m	sd	dur	m	sd	dur	m	sd	dur	m	sd	dur
1	615,2	152,2	11357	613,7	108,9	8422	670,2	149,8	11238	626,4	142,8	13359	651,4	156,9	12233	667,4	184,6	10810
2	606,6	215,1	77112	601,5	190,9	70608	735,8	232,6	73259	574,5	223,1	79889	642,8	232,5	72681	682,9	231,9	78738
3	613,8	251,9	87564	615,9	251,9	81665	796,4	285,7	91013	515,6	273,0	92268	674,4	270,1	85632	700,2	273,4	89439
4	619,3	290,8	86517	627,6	289,2	79586	840,4	302,0	88041	505,6	308,5	90034	716,9	298,4	84427	710,6	298,4	86581
5	630,5	298,7	88395	618,0	297,1	79349	805,0	312,6	88595	557,9	323,0	88878	732,2	296,6	87436	716,8	307,6	87244
6	620,1	305,6	90596	632,6	300,2	79406	749,0	323,6	88706	612,1	333,9	89973	734,5	293,5	88020	712,6	311,6	93154
7	620,6	302,6	90872	627,9	300,6	83221	703,3	323,4	89779	621,4	338,5	90118	717,7	297,0	86830	696,2	317,9	91852
8	634,5	296,6	91472	616,3	298,9	82655	684,3	326,4	90526	649,1	342,0	91688	712,6	299,1	89196	681,5	310,6	93131
9	629,9	298,7	94053	620,8	293,9	83176	693,5	333,5	89736	639,0	339,8	90893	708,9	302,5	90940	679,8	313,4	91061
10	618,3	307,3	93655	629,8	307,8	85409	692,3	336,5	91355	658,1	334,3	92452	699,8	299,8	90290	689,5	318,9	90941
11	612,7	301,8	93190	610,4	311,1	82739	686,5	330,9	89929	653,0	333,3	91622	679,4	301,9	89741	698,7	307,9	92003
12	595,7	302,9	93268	619,0	304,0	82246	694,5	323,3	90928	665,9	326,3	91639	701,9	304,3	89633	694,2	301,5	90629
	618,1	**277,0**	83171	**619,4**	**271,2**	74874	**729,3**	**298,4**	81925	**606,5**	**301,6**	83568	**697,7**	**279,4**	80588	**694,2**	**289,8**	82965

The last row averages over every condition. Data within the first 3,000 ms post-cut is included. The first table shows values for each of the 24 participants, the second for each of the 250 ms time frames, used and analysed of the first experiment.

Descriptive statistics experiment 2. In the following table the mean horizontal fixation values in pixels (*fix, m*), the standard deviations (*sd*) and the total of fixation duration in ms (*dur*) for the three conditions (*same, novel, shift*) in both, left and right frame, clips.

	left frame clips						right frame clips					
	same		novel		shift		shift		novel		same	
subj	fix	dur	fix	dur	fix	dur	fix	dur	fix	dur	fix	dur
1	616,7	40070	646,9	38786	651,2	40153	656,2	41858	607,0	42601	588,5	43136
2	667,8	41775	664,6	36943	584,7	43874	647,8	38605	596,2	40945	680,4	42305
3	712,3	40290	778,5	38190	758,8	35161	691,1	41004	598,2	37298	645,5	43082
4	747,8	41672	744,8	40571	596,5	40570	683,2	41452	610,0	45727	608,3	42266
5	784,8	46440	759,3	41953	733,7	46263	680,7	45821	601,9	41862	675,3	46616
6	730,6	45550	660,5	35798	663,8	40534	638,1	45013	675,8	41753	629,6	45014
7	695,6	41895	683,0	41379	633,8	40365	678,3	40594	680,2	40720	572,4	42466
8	624,5	39580	619,6	42233	627,7	41461	602,7	41516	588,2	41350	589,7	44383
9	601,1	45168	741,0	41091	671,6	45391	597,4	43317	637,7	40349	692,3	45683
10	646,2	44981	617,1	37727	585,8	46014	621,1	47167	520,6	42908	602,5	47189
11	721,4	39759	641,4	37462	551,0	44191	633,8	40155	654,1	45080	638,5	42293
12	573,5	36683	604,0	41534	596,7	41485	625,5	42775	588,9	40623	515,4	42351
13	549,5	43222	713,6	38115	606,0	44327	624,9	44239	508,0	40384	514,2	43852
14	630,4	44851	685,8	37282	550,5	44761	626,4	46133	543,4	44235	511,3	46868
15	635,8	39415	655,5	40877	659,2	39800	661,8	42288	610,8	39379	651,5	39231
16	661,2	41931	698,0	39385	675,8	40005	634,0	41593	611,2	41415	627,1	41616
17	611,6	40763	707,3	38580	645,4	42020	687,9	35146	547,4	36617	596,2	43046
18	680,2	44792	665,5	38607	649,0	47263	626,3	44395	572,6	43245	642,7	45693
19	724,6	42413	660,5	40606	656,6	40641	676,1	40006	614,6	40883	620,2	41100
20	648,9	40189	668,5	42150	667,3	42137	532,4	40725	598,1	42591	573,9	44070
21	652,4	47841	685,5	39675	605,8	45965	615,3	46185	533,1	41299	584,6	45287
22	704,9	47880	684,9	38279	580,2	45577	664,6	45407	556,7	42549	534,7	46690
23	723,4	26937	647,6	26873	723,0	28448	576,7	31770	672,1	31380	616,8	25164
24	687,0	44107	659,5	41426	650,5	42561	640,6	43285	642,7	46460	570,5	43382
	668,0	42008,5	**678,9**	38980,1	**638,5**	42040,3	**638,5**	42102,0	**598,7**	41318,9	**603,4**	43032,6

	left frame clips									right frame clips								
	same			novel			shift			shift			novel			same		
t	m	sd	dur	m	sd	dur	m	sd	dur	m	sd	dur	m	sd	dur	m	sd	dur
1	616,9	146,6	16307	613,5	125,9	13681	611,7	116,1	14839	587,7	133,3	14082	586,8	124,6	13610	594,3	139,7	15599
2	631,8	206,9	79036	636,9	196,8	73932	633,2	204,3	78651	592,1	201,7	81779	582,0	192,9	79822	581,2	201,6	82255
3	671,3	257,6	90411	684,5	253,2	86462	628,4	261,6	92373	618,9	250,3	92486	566,9	232,1	93078	594,6	243,7	93685
4	688,0	289,0	90485	712,6	287,2	86046	607,9	292,2	91824	649,7	283,3	92863	575,4	263,2	90044	602,2	277,2	92493
5	676,0	301,8	90663	713,7	310,0	85703	595,6	302,5	92116	676,0	297,5	90330	592,7	285,4	89699	598,4	300,2	93191
6	698,9	298,5	90275	700,2	303,8	84459	617,1	314,3	91577	685,5	294,4	90560	614,0	301,9	89899	592,8	305,7	93845
7	699,9	306,1	92908	667,8	302,0	82262	640,0	309,1	90862	681,1	285,4	89722	611,1	304,8	88756	582,9	301,2	93144
8	666,6	308,8	91489	665,8	309,2	83376	660,4	305,9	91320	662,0	297,4	90761	602,0	300,9	89607	604,3	299,5	91682
9	660,6	304,8	91310	675,1	310,3	84528	656,3	305,9	91348	636,3	299,8	91833	601,8	300,8	88778	609,0	299,4	93116
10	655,0	298,0	90606	673,9	297,2	84480	645,0	295,1	92031	619,9	299,5	91379	618,1	300,9	89560	624,1	302,9	93780
11	668,1	295,8	91389	679,1	291,0	84658	649,9	298,0	91441	614,1	295,9	92103	615,6	306,5	88404	623,6	294,9	95321
12	656,4	290,0	93325	669,7	294,9	85935	674,7	292,6	90585	612,1	287,0	92551	607,5	304,6	90396	623,2	297,5	94672
	665,8	**275,3**	84017	**674,4**	**273,5**	77960	**635,0**	**274,8**	84081	**636,3**	**268,8**	84204	**597,8**	**268,2**	82638	**602,5**	**272,0**	86065

The last row averages over every condition. In the first table, all data within the first 3,000 ms post-cut is included. It shows values for each of the 24 participants, whereas the second table shows values for each of the 250 ms time frames of the second experiment.

A.II Zusammenfassung

Mit der zunehmenden Verbreitung von Mobiltelefonen, deren Möglichkeiten Kurzvideos aufzunehmen und günstiger, sowie einfach bedienbarer Videoschnittsoftware, stieg die Zahl der Menschen die Videos erstellen und teilen stark an. Aktuelle Forschung bezieht sich häufig auf Eigenschaften und Merkmale die Aufmerksamkeit von Zusehern auf sich ziehen, selten jedoch auf die Analyse von Aufmerksamkeit nach Filmschnitten. Zwei Theorien werden aktuell auf deren Zusammenhänge geprüft: die Top-down Theorie, wo angenommen wird, dass bekannte Merkmale Aufmerksamkeit lenken, und die Bottom-up Theorie, wo Aufmerksamkeit als Reiz getrieben verstanden wird.

In dieser Arbeit werden zwei Eye-tracking Experimente vorgestellt. Analysiert wurden neben den horizontalen Fixationen von Zusehern, die Latenzzeiten der Fixationen sowie deren Dauer während des Sehens von kurzen Videos aus natürlichen Umgebungen. Es wurden Sequenzen zu je zwei Videos, getrennt durch einen Schnitt, in den drei Bedingungen: gleicher (same), neuer (novel), verschobener (shifted) Inhalt dargestellt, um auf Präferenzen von neuen versus bereits bekannten Inhalten zu prüfen.

Im ersten Experiment werden Präferenzen für bekannte Inhalte, zwischen 251 und 1.500 ms nach dem Schnitt, gefunden ($p < .01$). Weiters können signifikant verkürzte Latenzzeiten für bekannte Inhalte aufgezeigt werden. Nachdem die Aufgabe der Zuseher jeweils fokusiert auf die Beziehung zwischen den beiden Videos einer Sequenz waren, wurde ein zweites Experiment mit einer unterschiedlichen Aufgabe entwickelt. Im zweiten Experiment wurden Videos in einem Memory Abschnitt vorab angesehen. Die Aufgabe bezog sich anschließend auf die Suche nach diesen Videos unter anderen Videos in den Sequenzen. Hier wurden, im Gegensatz zum vorigen Experiment, leicht gegensätzliche Effekte gefunden. Zwischen 501 und 1.250 ms nach dem Schnitt wurden Präferenzen für neue Inhalte gefunden ($p < .01$).

Diese Resultate zeigen einen starken Einfluss der Aufgabe an der eigentlichen Aufmerksamkeitssteuerung. Top-down Prozesse ermöglichen es Zusehern, aufgrund der gestellten Aufgabe, automatisch auf bereits bekannten wie auch auf neue Inhalte zu

fokussieren. Die Verwendung von kurzen Videos, Schnitten und einer Bedingung mit verschobenen Inhalt, erwies sich als effizient und passend für diese Analyse. Zukünftige Forschung könnte unter Anderem stärker unterschiedliche Aufgaben verwenden, in jedem Fall aber berücksichtigen. Die Produzenten von Videos sollten berücksichten, dass das Zuseherlebnis und die Aufmerksamkeit von Sehern, durch Instruktionen stark beeinflusst werden kann und in manchen Fällen dies auch sollte. Diese Instruktionen könnten allerdings auch Teil des Videos selbst sein.

A.III EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Wien, am 23. November 2015

(Raphael Seywerth)

Raphael A. Seywerth, Msc **Curriculum Vitae**

References

Software architect, since 12/2009
Architecture and implementation of Java web services
From portal prototype to production in multiple countries

Psychology internship, 11/2014 – 01/2015
As part of the 240 hour internship of the University of Vienna
Execution of an industrial psychological scientific study
Measurement of engagement in social media

Software developer, 09/2005 – 12/2008
Development of modules for a time management system

Education

2009 –	University of Vienna, Study of Psychology Majors in industrial psychology and educational psychology
2006 – 2008	University of Applied Sciences, Technikum Wien, Vienna **MSc in Engineering**, Telecom- and Internet Technologies
2004 – 2005	University of Derby, United Kingdom **BSc (Hons) Computer Studies**, with Second Class Honours
2003 – 2004	Paramedic and Ambulance Driver Training During fulfilment of the alternative service for the Red Cross
1997 – 2002	Federal Higher Technical Institute, HTBLVA Spengergasse, Vienna Department of IT and Organisation, completion with **Matura**

Publications/Presentations

Seywerth R., & Wolf A. (2011) *Studie zu den psychologischen Unterschieden von Smartphonenutzern und Nichtnutzern*, Department of Psychology, University of Vienna.

Seywerth R. (2008) *MASSAI - Social Awareness and Beyond*, University for Applied Sciences, Technikum Wien, Master's Programme Telecommunications and Internet Technologies.

Seywerth R., & Mayerdorfer T., Digital Identity with MASSAI - Mobile Applications based on Social Smart Identities, CMG-AE/FITCE/GIT IT- and Telecom symposium 2007.

Seywerth R. (2005) *The Human Brain - A Masterpiece of Artificial Intelligence*, Bachelor's thesis in Computer Studies, University of Derby.