Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

The influence of exit-entrance cuts on visual attention during movie viewing

An eye tracking experiment

Gilles van den Eede

MSc 09/2019-010

The influence of exit-entrance cuts on visual attention during movie viewing:

An Eye tracking experiment

Gilles Van den Eede

Snr: 1260843

Master Thesis

Communication and Information Sciences

Faculty of Humanities and Digital Sciences

Tilburg University

Supervisor: Dr. R. Cozijn

Second reader: Dr. R. Koolen

September 2019

Index

Abstract

Film editing is a critical post-production phase that consist of rearranging and manipulating shots to create a continuous story with the idea of providing the spectators with a particular experience. This creation of flow must be done in spite of the differences in space, time and information.

Film editors have a broad gamma of editing techniques at their disposal. In this thesis, the focus is on continuity editing, which aims to create a smooth transition and to lead the attention of the audience from one scene to the next. In particular, the focus is on a technique called ‘exit-entrance cut’ and the aim is to investigate whether this technique is well suited to cause a sense of continuity.

To determine the extent of the influence on perceived continuity and visual attention, an experiment is created in which exit-entrance cuts from existent movies are used, but where the entrance is replaced by an icon. The icon is shown either congruently (in line with the direction of motion) or incongruently (at the opposite side of the direction of motion) with the expectation that is created by the exit. Eye tracking and a recognition task are employed to monitor whether the icon was perceived. The pattern of the data suggests that the icons are looked at longer during the congruent trials than during the incongruent trials. However, the effect is not as strong as hypothesized. Further analysis shows that the icon that is located congruently is seen more often than the one that is located incongruently, which suggests that the exit-entrance cut is able to influence the visual attention of viewers. Further research is suggested to ascertain that exit-entrance cuts guide the visual attention of viewers.

Introduction

A film is a visual narrative made up of scenes which are a series of shots and cuts. Shots are the images that are filmed by cameramen under instructions of the director. A connection of multiple shots that are part of the same plot point form a scene. When one shot ends and another begins, they are connected by a cut. There are ways to tie shots together such as by making use of a cross-cut or dissolve. Cross-cutting occurs when shots of one person rapidly alternate with shots of another, creating the implication that both events are occurring at the same time. A dissolve is a smooth visual transition from one shot to the next. The first scene becomes more and more transparent while simultaneously the second scene becomes clearer.

When a film is well edited, it will seem like a continuous and unbroken story. This is comparable to how individuals usually see. In natural situations, the brain strings together different shots that are broken by eye blinks and saccades, which are rapid eye movements to find a new point of fixation. If the flow of the scene is interrupted during film viewing, the viewer may become distracted and lose attention. Therefore, an important task of editing is to create continuity and manipulate the viewers' focus on what they should look at (Dmytryk, 1984).

Ultimately, shots need to be brought together into a perceived fluent sequence which the spectator sees as a continuous story, without noticing how it was created. One of the first instances of editing can be seen in the seven-minute film Cinderella, produced by Georges Méliès in 1899, where a connection was established between twenty separate shots to provide a consistent story across aspects such as time and location. The idea that the previous shot is built upon by the next is known as continuity (Dmytryk, 1984).

Another way to grab the attention of viewers is to shorten the time between cuts. A fast-paced action movie is often easier to focus on than a slow-moving drama. Slow cinema features static camerawork, less cuts and not a lot of motion (Dwyer & Perkins, 2018). When a cut occurs, it often increases the attention of the viewer (Loschky, Larson, Magliano, & Smith, 2015). A slow movie, however, allows a viewer's thoughts to trail off.

In this work, one specific editing technique that is used to guide the attention of the audience, namely the exit-entrance cut, is examined in detail (Smith, 2012). The most common example of the exit-entrance cut is when an actor walks off screen to the left, only to re-emerge in the next shot on the right after a cut. When properly done, this type of cut comes across to the viewer as a smooth and continuous action (Dmytryk, 1984). Moreover, this cut guides the attention of viewers and creates expectations: when an actor or subject is in a continuous motion and disappears off screen on one side, the viewer expects that they will reappear on the other side. In this study, an experimental framework is set up to investigate if it is possible to observe the effect this type of cut has on the visual attention of viewers. This is done with an eye tracking experiment, as the movement of the eyes are an indicator for the visual attention. The experiment should provide an answer to the question: “Does the focal point leaving the screen create an expectation of what the next shot will bring, and will viewers anticipate it?”

Attentional Theory of Cinematic Continuity

When a movie maintains continuity, it makes the viewing process effortless and the editing unnoticed. But what exactly does continuity entail? Continuity itself signifies a logical sequence or cohesion. Cinematic continuity is established when the viewer experiences a “flow” in the movie between shots. As stated before, this flow can refer to a coherence of space and time, but also to a coherence of characters and their interactions, goals and causal relations (Zacks, 2013). This illusion of continuity is created in the mind of the viewer, it aids to make the viewing process effortless and the editing unnoticed. In order to infer what the viewer is experiencing it is important to keep three questions in mind:

1. What are the viewers attending to?
2. What are they perceiving?
3. What are they expecting?

Through their gaze, viewers are connecting and collaborating with movies and therefore, it is through the observation of their eye movements that these questions can be answered, and continuity can be measured (Anderson, 1998; Berliner & Cohen, 2011; Smith, 2012).

The Attentional Theory of Cinematic Continuity identifies the essential role that visual attention has in perceiving continuity across cuts and how expectations can be created prior to a cut and matched across a cut. Smith (2012) argues that the viewer is an active participant as they seek out information on screen, build new expectations about future events, and are able to follow objects across cuts. As such, the Attentional Theory of Cinematic Continuity is based on three stages: attending to a shot, cuing attention across a cut, and matching expectations after a cut (Smith, 2012).

When viewers are shown two shots in succession and the second shot matches the expectation that the first shot created, this is known as *a priori* continuity. It entails that no explanation is required to process what has occurred. As opposed to this, there is a concept that is known as *a posteriori* continuity. This type of continuity signifies that the expectations, which links one shot with the next, are not met *a priori*. When such a cut occurs, the audience will look for an explanation and try to preserve the continuity at the second shot, to connect it to the previous shot (Smith, 2012).

Certain editing rules have been developed in Hollywood in order to preserve *a priori* continuity. An example of this is the shot reverse shot, in which shots alternate between two or more characters who are engaged in an interaction such as a conversation. Guidelines have also been developed about what should be avoided in order to preserve continuity, such as not crossing the axis of action. This entails that if characters are filmed from one side, the shot should not cross their line of interaction and jump to the other side, creating a movement larger than 180°. Making such a jump can be confusing as it changes both the direction of the action, and the direction in which the characters are looking. For an exit-entrance cut, the direction of motion should be the same across both shots, which automatically enforces the 180° rule (Smith, 2006). A sequence that can be confusing for the viewer is the jump cut. This type of cut happens if the position of the camera between shots changes less than 30 degrees. The impression that results from this cut is that it seems as if the character 'jumped' (Smith & Henderson, 2008; Smith, 2012).

Kachkovski, Vasilyev, Kuk, Kingstone, and Street (2019) have investigated how distracting and displeasing violations toward the 180° rule can be. Their participants were largely able to identify such violations, and they did not find them significantly distracting. More importantly, the violation seemed to have no effect on the enjoyment of the scene. This raises the question whether the mainstay editing techniques that are known as Hollywood style are as crucial to making a movie as many directors seem to believe.

Another mainstay editing technique used to preserve cinematic continuity is the match-action cut. In this type of cut, an action is set in and the viewer builds up an expectation about its ending. Continuity will be experienced when the shot is followed by a shot of the action reaching its goal or destination. An example is when a ball is thrown and the next shot shows the ball in the air flying across the sky, after which it ends up in the hands of a person at the end of its trajectory (Smith, 2012). A character's gaze or pointing can also create an expectation for what the next shot should bring. Viewers will become curious what the character is gazing or pointing at. *A priori* continuity is perceived if the following shot does in fact show the gazed or pointed at object or character. This is known as a gaze cue cut (Smith, 2012).

Smith and Martin-Portugues Santacreu (2017) have investigated whether viewers are able to notice the occurrence of cuts when looking at match-action cuts. They were mainly interested in the motion just before a cut (pre-cut motion), and the motion straight after a cut (post-cut motion). Removing post-cut motion greatly increased the time it took viewers to realize a cut had occurred. While removing pre-cut motion, which made the character go from no motion in one shot to sudden movement in the next, did not have an effect. This contrasted with their hypothesis that pre-cut motion was required to hold viewers' attention while the cut occurred, for the participants to miss the cut. Eye tracking showed that pre-cut motion followed by a termination of the motion while no post-cut motion was present, was jarring for the participants. This was shown by a saccade that the eyes made as the cut occurred, as the motion the viewers were following did not continue. In the current research the post-cut motion of the exit-entrance cut will be shown, however only briefly, for approximately four frames.

One possible reason why continuity editing is successful is that it recreates actions that are familiar to the real-world (Anderson, 1998). An example of this can be found in cutting on sound. In this instance, a sound can be heard during a scene, after which the origin of the sound is shown. An example is when a doorbell is heard, and the next shot shows a person waiting at the front door. Like in the real-world, hearing a sound would make you look up in its direction to try and figure out what caused it. However, there are three major differences between films and the real world. First, films have the ability to rapidly switch between locations after a cut. Secondly, films can make bits and pieces from different locations and even different times feel connected. Lastly, films can remove sections that are deemed as redundant to the narrative. For example, if a character gets into a car it is not necessary to show the whole ride, simply showing the arrival at the destination in the next shot is enough to maintain continuity (Münsterberg & Griffith, 2004).

When the audience is surprised by a change from one shot to another, they can become confused which can cause a sense of discontinuity. As explained above, the continuity editing rules intend to minimize confusion by preparing the viewer and guiding their attention. Their attention can be guided by several cues such as off-screen sounds, gazing and pointing, following the action in the scene, and when objects of focus leave the scene, such as during an exit-entrance cut. If a cut occurs while the viewers' attention is captured by the cue, the cut will not be noticed, and continuity will be experienced (Smith & Henderson, 2008; Smith, 2012).

Swenberg and Eriksson (2018) believe that the Hollywood editing style is important to reduce the cognitive load of viewers. They have found that cognitive load increases for viewers when there is a lack of continuity. The increase in cognitive load was deduced by interpreting pupil dilation and an increase in saccade frequency of the viewers. This was not really an issue in their short experiment, however if viewers must constantly actively deduce what is happening it could become a problem for the viewing experience. Cognition and attention will be discussed in more detail below.

Visual attention

Individuals perceive many things in the world in parallel, for example, when walking through a store they can see the different products, the special offers and the queue at the checkout. Through a cognitive selection process that is called selective attention, they are able to filter the stimuli that will reach their consciousness. In other words, selective attention is what is used to decide on which sensory input to focus, and which information is ignored. While watching a movie, the sensory input is visual, which is why in the context of film editing it is referred to as visual attention. For movie directors, it is essential that the audience focuses their attention on the relevant information. The information storage, or cognitive load, is most often referred to as working memory. Here, the relevant information is retained, and irrelevant information is suppressed. Furthermore, this is what individuals use, among other things, to make decisions. The concept working memory is not only the temporary storage of information, it can also manipulate the information it receives (Olivers, 2008; Vos, 2018).

To understand movies, both working memory and visual attention are required, allowing the processing of images, while using prior knowledge to process what is occurring (Walther, Rutishauser, Koch, & Perona, 2005). Humans can reduce the amount of data to more easily digest relevant information (Rutishauser, Walther, Koch, & Perona, 2004). To do this, attention is not focussed on an entire shot at once, but focus on multiple different parts, one after the other, in order to extract relevant information (Sharma, Kiros, & Salakhutdinov, 2015).

When eyes rest on one position this is known as a fixation. The area that is being fixated holds the persons interest, which is why fixations are related to visual attention. In order to fixate on a different point, the eyes move which is known as a saccade. During a saccade, which takes 20-50ms on average, it is difficult to register new information, which is known as saccadic suppression (Rayner, 1998). This happens because the human eye moves with a speed as high as 500 degrees per second, making it possible to undershoot or overshoot the area intended to focus on. During a fixation, viewers can see everything that is in the foveal or central region of their eyes. The area that they register through the fovea is no bigger than a thumbnail that is held at arm's length. Visual acuity decreases by 80% outside of the fovea. Registering something 2° away from the fovea, in the region known as the parafovea, causes the resolution to decrease by 70%, and by 90% if the object is 20° separated from the fovea, in the periphery (Rayner, 1998).

The human eyes move so that the object that they are attending to is located in the fovea. Humans can recognize objects within 2.6° of the area upon which they are fixating. This means that when a high amount of detail is needed, viewing should process through the fovea. In contrast, when lower amounts of detail are needed, information can be acquired and stored from part of the periphery. Another aspect is that what viewers see in their periphery can influence their decision making in future fixations (Nelson & Loftus, 1980). This type of information gain can be a part of covert attention, which will be explained below.

With every eye movement new phase of visual processing begins. It is not possible to see a whole scene of a movie in detail at once. Therefore, viewers must move their eyes in order to process the part they are interested in (Smith, 2013). Processing something while looking at it is known as overt attention. As opposed to this, covert attention is used when processing power is spent on a point in space independent of the location of the gaze. Overt attention can be measured by using an eye tracker. Covert attention is not easily measured, however covert shifts in attention rarely occur except before a saccade. In other words, when viewers prepare to shift their gaze to a new location, their brain is already allocating resources to that location before the movement of the eyes (Nelson & Loftus, 1980; Smith, 2013).

The more viewers can fixate on a shot, the greater their attention and the higher their recognition of the shot and its individual aspects will be. Therefore, longer exposure time for a shot leads to more recognition if it is accompanied by more fixations. To test this, Loftus (1972) conducted a recognition memory test for pictures. He found that the recognition accuracy of a still that was present for 125ms was only 15% of the image, as opposed to a still that was displayed for two seconds that yielded an accuracy of 90%. Similarly, Intraub (1981) showed participants a sequence of twelve pictures where each picture was displayed for either 114ms, 172ms, or 258ms. Afterwards, the participants were respectively able to recognize 19%, 49%, and 58% of the pictures in the sequence.

Eye tracking

There are many ways to monitor what the audience is experiencing while watching a movie, such as asking for a self-report, testing their memory afterwards, recording their heart rate and even measuring their brain activity with functional Magnetic Resonance Imaging (fMRI). A real-time measurement of how the audience is both watching and processing a film, which is less invasive and more accessible, is eye tracking.

Eye tracking measures the eye movements of viewers as they gaze at visual stimuli such as a screen or a panorama. One of the first studies of eye tracking, performed in 1965 by Yarbus, was done by attaching a minuscule lens to the eye (Tatler, Wade, Kwan, Findlay & Velichkovsky, 2010). The viewer’s head was clamped so they could not move, and a beam of light shone on the lens. Modern eye tracking is not as invasive. The current technique is based on the infrared light that reflects off the human eye in a specific pattern. This results in two measurements: the corneal reflection and the pupil position. The corneal reflection is a small glint on the cornea, which is the transparent front part of the eye. If viewers look in different directions, the reflection will remain in the same position relative to the infrared light source. This measurement can therefore show whether the head is changing positions. Moreover, the pupil position can be determined as it does not reflect the infrared light. By combining both measurements, it is possible to establish how the pupil is moving with regards to the corneal reflection. This allows movement of the pupil to be attributed to the movement of the eye and be disassociated from head movements.

To determine what viewers focus on when using an eye tracker, it is useful to investigate attentional synchrony. This entails the clustering of the gazes of viewers who are fixating on the same area at the same time. Moving images almost always contain moments of attentional synchrony, in which at least 80% of the viewers are looking at the same area at the same time. However, during professionally composed movies, the amount of times this happens is greater than in naturalistic videos. In other words, directors and editors can guide the attention of their audience (Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Smith, 2013).

After the occurrence of a cut, the attention of viewers increases. During eye tracking this is shown as viewers having an active gaze and sharing the most attentional synchrony right after a cut. The number of saccades reduces linearly for up to four seconds after a cut, after which it remains at a low frequency. The frequency of saccades is generally higher when there are multiple focal points present such as faces, when these focal points are offset from the center of the screen, or during shorter shots. This could be an indication that shots have a limited amount of information and when the viewer has exhausted that information the shot is no longer interesting. Therefore, it is up to the editor to cut to new information or to reframe old information in order to hold the viewers' attention. When editing is successful, viewers' saccade frequency and attentional synchrony remain at their highest throughout the viewing experience (Smith & Henderson, 2008; Smith, 2013).

Loschky et al. (2015) agree that Hollywood style editing leads to a great amount of attentional synchrony. They postulate that filmmakers exert a lot of control over the audience's attention and refer to this as the Tyranny of Film. This is showcased by the fact that viewers' attention resets to the middle of the screen every time a cut occurs. Additionally, in their experiment they established that adding additional context to short clips further reduces the cognitive load that is required to process continuous film material.

Eye tracking has also shown that individuals prioritize faces and moving objects (Tatler et al., 2010). However, it is not yet determined if that is because these are often important to the narrative. Moreover, visual features such as color or contrast do not seem to draw attention unless the objects are also of interest to the viewer. Motion seems to be one of the strongest factors at influencing visual attention, especially when contrasted with a static background (Wolfe & Horowitz, 2004). When viewers are focused on movement, the editor has the opportunity to make a cut that will be smooth and invisible (Smith, 2013).

When viewers get a different viewing task it is possible that their viewing priorities get shifted (Yarbus, 1967). Smith & Mital (2011) asked their participants to determine in what location the scene they were watching took place. They found that their gaze actively avoided people and motion. As their viewing task was completed, the participants immediately returned their focus to motion.

Unless a shot is constructed with multiple centers of interest, shots will hold a center bias (Goldstein, Woods, & Peli, 2007). This means that viewers will spend a substantial amount of time focused on the middle of the screen. The attentional synchrony after a cut often takes place at the screens center. This can partly be explained due to a preference of positioning important objects in the middle. However, the off-center placement of important objects can be useful to lead the attention before a cut and after a cut. An example of this can be found in the exit-entrance cut, which will be explained below (Smith, 2013).

Exit-entrance cut

A B

Figure 1: Example of an exit-entrance cut. The exit part of the exit-entrance cut is shown in A, where the actors are leaving at the left side of the screen. The entrance part is shown in B, where the actors re-enter at the right side of the screen (Back to the Future 3).

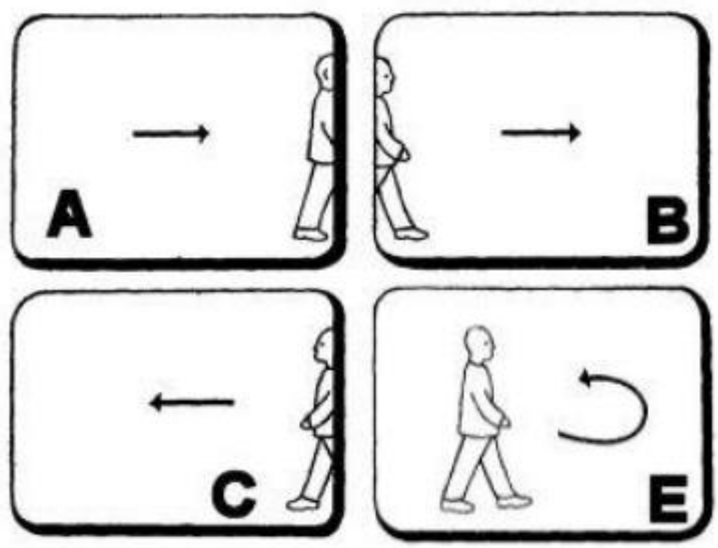


Figure 2: Continuity across an exit-entrance cut. A cut from shot A to B is acceptable. A cut from A to C is seen as discontinuous unless shot E occurs after A. (Smith, 2006, p.28).

This thesis is a continuation of similar research concerning match-action cuts (Van de Schepop, 2018) and gaze cue cuts (Eekelaar, 2019). Instead it focuses on a different editing technique, namely the exit-entrance cut, of which an example is shown in Figure 1. This type of cut can occur when a character leaves the screen or exits a place on screen. If the next shot shows the character re-entering the screen, the editing can be defined as an exit-entrance cut (Smith, 2012). Exit-entrance cuts were chosen as they are motion based, which is an important focus of visual attention. Moreover, they are readily observable through eye tracking, which is useful to analyze viewing behaviours (Smith & Matil, 2011; Wolfe & Horowitz, 2004). Furthermore, exit-entrance cuts show a natural continuation of motion and meet the expectation of the viewer, thus resulting in *a priori* continuity (Smith, 2012). Like match-action cuts and gaze cue cuts, exit-entrance cuts guide the attention of viewers and establish continuity.

The exit-entrance cut is a specific instance of a match-action cut. Figure 2 explains the general idea of such a cut. It is an excellent tool to create the illusion of continuous motion by employing the correct side of the screen for re-entry that is consistent with the direction of the motion. The sudden onset of movement of a character draws the attention towards both the character and their destination. For this reason, it is not a detriment to the perceived continuity if, for example, the background is inconsistent from shot to shot (Smith, 2012).

The cut away from the exit scene should occur when the eyes of the actor are no longer visible. The following entrance shot should be made 3 to 5 frames after the eyes of the actor have re-entered the screen (Dmytryk, 1984). Four frames take about 167ms at 24 frames per second which is less than the time it takes for a viewer to perform a saccadic eye movement (Fischer & Ramsperger, 1984; Rayner, 1998). How come this type of cut works? The actor that the viewer is focusing on, leaves the screen. Then, the viewer no longer has a focal point and is looking at the edge of the screen. To resolve this situation, the eyes of the viewer make a saccade towards the center of the screen and continue to the opposite side where the actor is expected to enter the screen in the next shot. This process happens in the fraction of a second and during the movement from side to side, the viewer will not have seen anything with clarity (Dmytryk, 1984). The attention is reset towards the middle of the screen because this is where the most important objects are expected to be displayed (Goldstein, Woods, & Peli, 2007; Smith, 2013).

Smith (2012) argues that if an object moves from the left to the right past the viewers' field of view and they wish to track it, the first thing the human eye does is to look to the right. Afterwards, the viewers' head would have to catch up by rotating to the right, after which their eyes move to the left again for their gaze to remain in the same location. This compensation is like rebounding to the screen center after an object of interest leaves the shot. Moreover, when the object reappears in the next shot it provides an analogue for the normal head rotation.

In order to process specific parts of a film, fixations are required. Viewers are unable to process information as their eyes are moving. Therefore, the cut to the other side often occurs while a viewer's eyes are moving to the other side during an exit-entrance cut. The viewer will have missed the cut, which results in a smooth and natural viewing experience (Dmytryk, 1984).

Van de Schepop (2018) has conducted research on whether match-action cuts can capture the attention of the audience. To test this, he showed participants several clips. After each match-action cut in the clip, the screen cut to black and an object was shown. This object was either located in line with the motion present in the match-action cut, or it was located on the opposite side. He found that viewers tend to follow the motion of the match-action cut and have a certain expectation of what should be shown after the cut. This became apparent as the participants were better at recognizing objects that were in line with the direction of the motion (congruent) than the objects that were at the opposite side of the direction of the motion (incongruent).

This research was replicated with clips that made use of a gaze cue cut instead of the match-action cut (Eekelaar, 2019). Following a gaze cue cut, an icon was shown congruently or incongruently and for a display time of 200ms or 800ms. This resulted in the same design as the research by Van de Schepop (2018). Eekelaar found that participants were better at recognizing and looked longer at objects in the congruent condition, as opposed to the incongruent condition. However, this was only true for the 800ms condition. In the present paper, a similar research will be conducted. Instead of match-action cuts or gaze cue cuts, this research will use clips that feature an exit-entrance cut.

Research question

The focus of this research is to investigate how individuals watch movies and how film directors can influence their viewers. Specifically, the editing technique exit-entrance cut is investigated and how it can guide the eye movements of the audience. This leads to the following research question:
"What is the influence of an exit-entrance cut on the visual attention of viewers?"

Exit-entrance cuts create the expectation that an object or person will reappear on the side of the screen that is opposite of where they left. In order to test whether the attention of the audience follows this principle, an experiment is created. After displaying the exit component of an exit-entrance cut, the screen turns black and an object will be shown on screen. This object is either placed on the side that follows the rules of an exit-entrance cut (congruent), or on the side where the person or object is not expected to reappear (incongruent). Furthermore, the objects will be shown for either 200ms or 800ms, which are the same display times used by Van de Schepop (2018) and Eekelaar (2019). The 200ms condition will test whether viewers are correctly estimating where a person or object should reappear after an exit cut. In this case, the viewers' covert attention shifts to the expected location allowing the participants to have seen the image when it appears. The 800ms condition will serve as a comparison, in which the viewers have time to notice the image and fixate on them through overt attention. In this condition, the congruency of the object should be of less importance, as 800ms allows for enough time to notice the object.

It is expected that an object shown in the congruent condition will have a greater viewing time of the object and a greater object recognition compared to an object shown in the incongruent condition. Therefore, the following hypotheses have been postulated:

H1: Object recognition scores will be higher for objects displayed at the congruent location after an exit cut than at the incongruent location.

H2: Object recognition scores will be higher for objects that are displayed for 800ms than for 200ms.

H3: Viewing times will be longer for objects displayed at the congruent location after an exit cut than at the incongruent location.

H4: Viewing times will be longer for objects that are displayed for 800ms than for 200ms.

Furthermore, it is expected that the differences between the congruent and incongruent locations are greater in the 200ms condition than in the 800ms condition. Even though 200ms is a very short time, it is possible to fixate on the objects in the congruent condition. Thus, a difference in viewing time, however small, is still expected.

Method

An eye tracking experiment was conducted to determine the effects of exit-entrance cuts on the visual attention. 20 clips were shown which were accompanied by an object that was either displayed opposite of where the exit took place (congruent) or at the same location (incongruent). The object was shown for 200ms or 800ms. The experiment consisted of a 2x2 design with two display durations and congruencies.

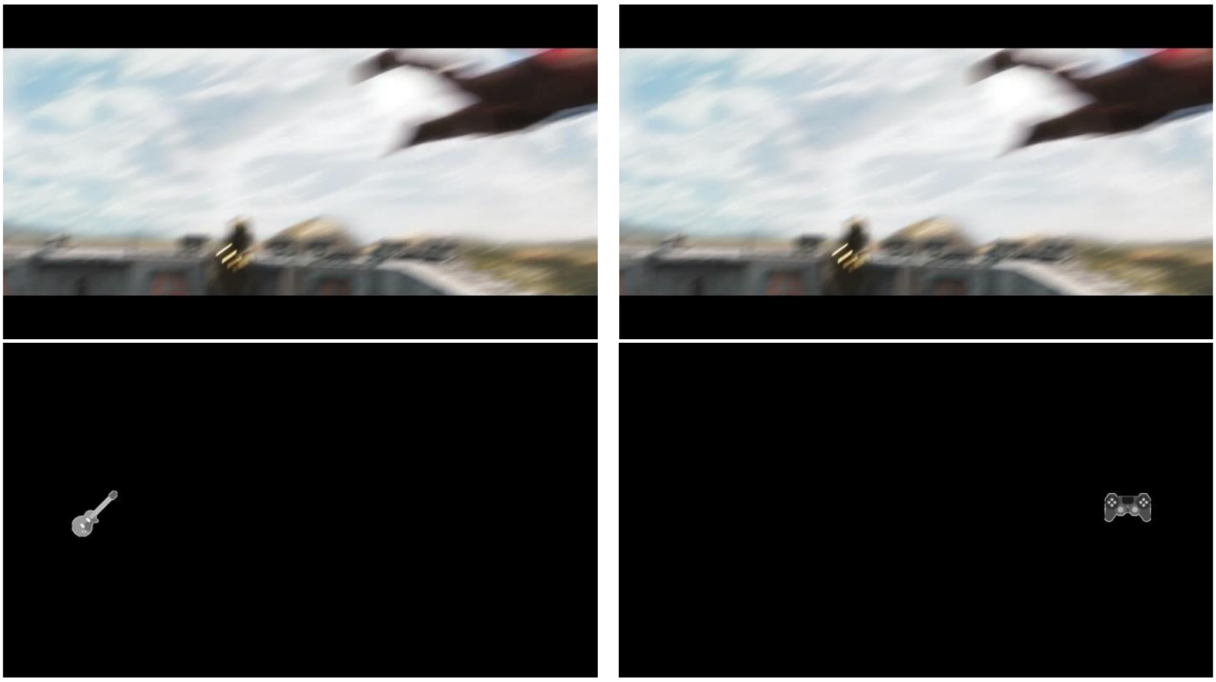
Afterwards the participants were asked to complete a recognition task to determine which objects they had seen.

Participants

53 participants were sampled via the Participants Pool of Tilburg University. The data of 13 participants were excluded as their eye tracking data was partially or fully missing. The data of 24 men and 16 women were analysed with a mean age of 21.7 years old (SD = 3.24). The

minimum age was 18 and the maximum was 32 years old. The students received 0.5 points of credit in the Participants Pool. Nine individuals indicated that they wore either contact lenses or glasses during the eye tracking experiment. Lastly, 26 participants indicated that they watched series or films at least twice a week.

Materials



A B

Figure 3: The onset of the exit and the icon shown at the congruent location A and at the incongruent location B (Black Panther).

The experiment featured 20 clips from professionally edited movies and series such as Blade Runner, The Godfather, The Haunting of Hill House, and Ben Hur. The clips contained an exit-entrance cut and had an average display time of 20 seconds with a frame rate of 24 frames per second. 11 clips featured an exit that occurred on the right side of the screen. The clips were merged into one continuous movie. Four frames after the exit part of the exit-entrance cut occurred, the film cut to black. While the entrance part was barely shown in the experiment, it was still important to find actual exit-entrance cuts as this meant that the film director meant to lead the attention of the viewer from one scene to the next. Moreover, the exit-entrance cut had to be continuous, meaning that the entrance had to occur on the opposite side. In order to find the clips, around 50 movies and series had been surveyed in detail, either fully or at increased speed. Exit-entrance cuts that occurred in the middle of the screen were not included as they are not suitable for eye-tracking experiments. After the exit part occurred in the experiment, an object was shown on the side where the re-entrance was expected to occur (congruent) or on the side of the exit itself (incongruent). This is displayed in Figure 3. The four frames gave the participants the opportunity to shift their gaze in response to the exit, that is, to the side where they expected something to appear. Twenty objects in grey tones with a resolution of 124x124 were gathered from an online icon database (www.iconfinder.com). The icons were unrelated to the clip that they accompanied, as to avoid creating a connection between them and resulting in them being easier to remember, such as when a knife icon appears after a violent scene. Before every new clip, a break of one second was implemented, 500ms of a black screen and 500ms of a fixation cross. The clips and the icons can be found in Appendix A and B.

Design

Two independent within-participants variables were present in the experiment: Congruency, whether the object was shown congruent or incongruent with expectations and Timing, either 200ms or 800ms. The 20 clips were equally distributed over the conditions, which resulted in five clips for each combination of variables. Four lists of the experiment were created. The lists all had the same sequence of movie clips but differed in which objects were shown and in which combination of variables. Participants were randomly assigned to a list.

Apparatus

To conduct the eye tracking experiment in the DCI lab at Tilburg University, an SMI RED 250 eye tracker with a speed of 250Hz and an accuracy of 0.5 degrees was used. For the audio component of the clips, the participants used a Sennheiser Headset. They had to be seated correctly in accordance with the calibration by moving the desktop monitor to create approximately 65 cm between them and the eye tracker. To record the eye tracking, a Shuttle XPC mini desktop with a monitor with a resolution of 1680x1050 was used. The program SMI Experiment Center 3.7 was used to record the eye tracking data while SMI Begaze 3.7 was used for the visual data inspection.

Instrumentation

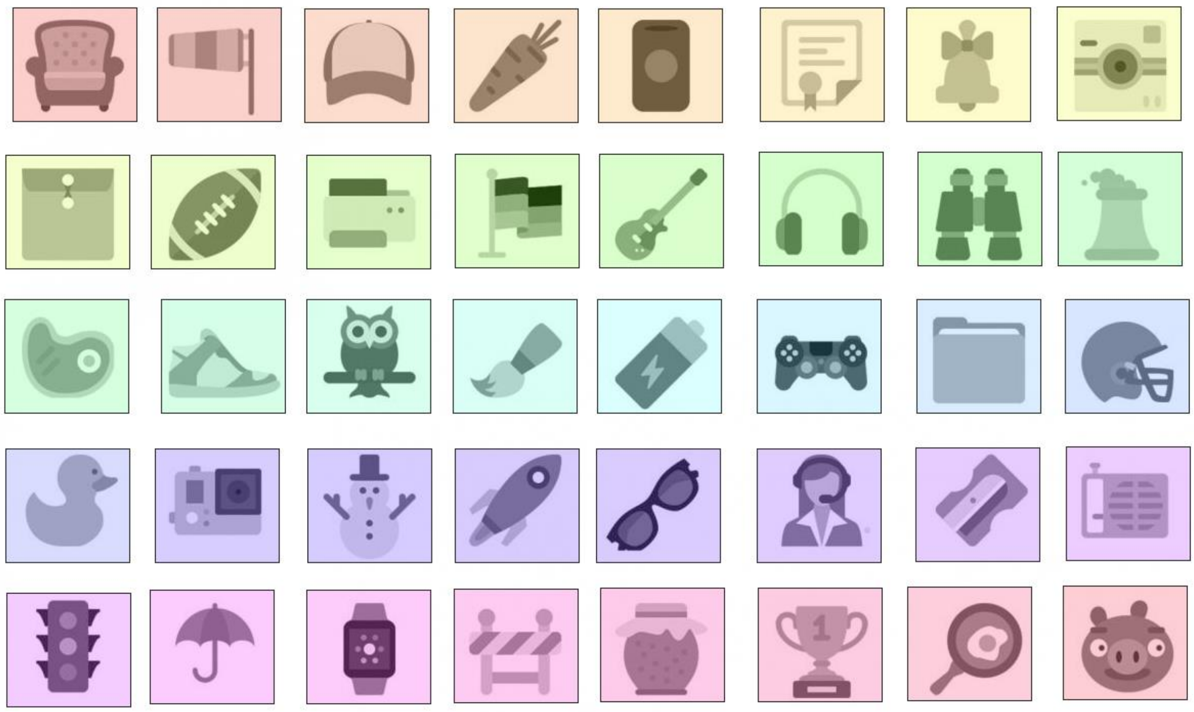


Figure 4: The icons as they were displayed in the recognition task. 20 of these icons were present in the experiment, while the other 20 were used as a distraction.

After viewing the clips, a demographical questionnaire was presented to the participants. The questions concerned their age, whether they wore lenses or glasses and how often they watched movies and TV series. They can be found in Appendix C. The experiment ended with a recognition task, like the one used by Van de Schepop (2018). The task consisted of 40 images, of which only 20 appeared in the experiment. These images are shown in Figure 4. The participants were asked to select the ones they had seen. Images that were highlighted and were indeed a part of the experiment were rated as correct. The number of correctly scored objects showed how many objects the participants had seen and memorized.

Procedure

Before the experiment, the participants were properly seated, and the eye tracking device was calibrated. A 9-point calibration was used to ensure that the eye tracking device monitored the eye movements accurately. Afterwards, the participants were told to pay close attention and that some questions would be asked afterwards, not specifying what the questions were concerning. During the experiment, 20 clips were shown that feature an exit-entrance cut. After each clip, the screen turned black and a drawing of an object was shown somewhere on the screen. This object was shown for either 200ms or 800ms and congruently or incongruently with expectations.

After the eye tracking experiment, the participants were asked to complete a recognition task to determine which objects they have seen. The results of this task were complementary to the results of the eye tracking. Between the recognition task and the showing of the clips, the participants were asked to fill out a demographical questionnaire, which was meant as a distraction task.

Data Analysis

A B

Figure 5. Onset of the exit of the first clip accompanied by the icon displayed after the clip at the congruent location A and the incongruent location B (Back to the Future 3).

To analyze the eye movements, the data were segmented. The part of the clip that was used consisted of four frames before the onset of the exit part of the exit-entrance cut, until the object disappeared. As stated, the object was on screen for either 200ms or 800ms and was either congruent or incongruent with expectations. This resulted in a list of 20 segments, one for each clip. A time code was created for every version of the movie. The eye tracking data of every participant were then segmented using the time codes with the program SMImovie. Afterwards, these data were monitored with the software *Fixation* (Cozijn, 2006). Each segment that was created was accompanied by a screenshot, of which an example is shown in Figure 5. This screenshot showed the onset of the exit part of the exit-entrance cut, as well as the icon that was shown after the clip on the same location as during the experiment. These two parts of the screenshot were defined as the Areas of Interest (AOI’s). The program

Fixation (Cozijn, 2006) assigned the fixations of participants to either one of two AOI's, or to nothing. The assignments were inspected manually. If a fixation was an under- or overshoot close to the AOI of the object, it was assigned to that area. The viewing time per AOI was computed as the sum of the fixations on the AOI in milliseconds. The resulting data file was statistically analyzed with SPSS.

The recognition scores were computed using Excel. This resulted in a score out of five for every participant and every condition. These scores were analyzed in SPSS.

Results

A repeated measures ANOVA was conducted with two factors: Congruency and Timing. The analysis was performed over the data of 40 participants that watched 20 clips accompanied by 20 icons. This was done for both the eye tracking data and the recognition data. In 172 cases, the data showed that there were no fixations on the AOI. These cases were removed from the analysis. Furthermore, the data were analysed over participants (F_1) and over items (F_2). This was done for both eye tracking data and the recognition data. For the recognition test, two objects did not have a correct position in the randomization of objects and were excluded from the analysis over items.

Recognition

Table 1: Means and standard deviations of the participant analysis of object recognition scores (out of 5) as a function of Congruency (congruent and incongruent) and Timing (200ms and 800ms).

	Exit-entrance cut			
Display Time	Congruent		Incongruent	
	M	*SD*	*M*	*SD*
200ms	1.05	.20	.88	.14
800ms	2.95	.18	3.23	.20

The mean recognition scores of the participant data analysis are shown in Table 1.
The analysis of the recognition test showed an effect of Timing (F_1 (1,39) = 191.034, $p <$ 0.01, $\eta^2 = .830$, F_2 (1,17) = 329.389, $p < 0.01$, $\eta^2 = .758$). Significantly more objects were recognized and remembered from the 800ms trials than from the 200ms trials.

There was no effect of Congruency (F_1 (1,39) = .124, $p = .726$, F_2 (1,17) = .222, $p =$ .686). However, the pattern of the data suggested that Items were recognized more often if they were displayed congruently rather than incongruently. This was only the case when Items were displayed for 200ms.

No interaction effect was found between Congruency and Timing (F_1 (1,39) = 2.773, p = .104, F_2 (1,17) = 2.722, $p = .324$).

Eye movements

Table 2: Means and standard deviations of the participant analysis of the duration of fixations (in ms) as a function of Congruency (congruent and incongruent) and Timing (200ms and 800ms).

	Exit-entrance cut			
Display Time	Congruent		Incongruent	
	M	*SD*	*M*	*SD*
200ms	254	55	254	47
800ms	448	35	399	36

The mean viewing times of the participant analysis are shown in Table 2.

The analysis of eye movements showed an effect in Timing (F_1 (1,102) = 15.015, $p <$ 0.01, $\eta^2 = .128$, F_2 (1,51) = 14.134, $p < 0.01$, $\eta^2 = .217$). The participants fixated significantly more on objects during the 800ms trials than during the 200ms trials.

There was no effect of Congruency (F_1 (1,102) = .315, $p = .576$, F_2 (1,51) = 1.588, $p =$ .213). However, the pattern of the data suggested that Items seemed to be looked at longer during the congruent trials than during the incongruent trials. This was only true when Items were displayed for 800ms.

No interaction effect was found between Congruency and Timing (F_1 (1,102) = .305, p = .582, F_2 (1,51) = 1.065, p = .307).

Discussion

The research question posed in this work concerns the experimental testing of concepts embedded in the Attentional Theory of Cinematic Continuity (AToCC), which distinguishes three stages for perceiving cinematic continuity: attending to a shot, cuing attention across a cut, and matching expectations after a cut (Smith, 2011). The exit-entrance cut, which is the subject of this work, is an editing technique that should be able to cue attention across a cut and match the expectation after a cut. To test this statement and to measure the viewers' visual attention, an eye-tracking study was combined with a recognition task. Viewing times and recognition scores were gathered after the occurrence of an exit-entrance cut. More precisely, after the exit part, an icon was shown for either 200ms or 800ms and displayed congruently or incongruently with the expectations created by the exit.

The results of the experiment did not definitively show that the visual attention of viewers is captured by the exit-entrance cut. The findings are for both the data over participants (F1), and over items (F2). Both the eye tracking data and the recognition scores did not display that viewers actively anticipated what would happen next. In other words, the exit part of the exit-entrance cut was not able to guide the attention and *a priori* continuity was not established. However, increasing the time that the icon was displayed increased both the recognition scores and the viewing times, which is in line with research conducted by Loftus (1972). This study showed that viewing an object for a longer time increases the recognition accuracy for that object.

The results regarding congruency are unexpected, because previous research has shown that other types of cues do influence the visual attention of the viewer, for instance, the match action cut (Van de Schepop, 2018) and, very recently, the gaze cue cut (Eekelaar, 2019). For the gaze cue cut, the findings showed that the attention was only successfully matched after a cut when the display time was 800ms as opposed to 200ms. In the current research, the increased display time did not significantly influence the results when comparing congruent and incongruent trials. This indicates that the attention was not successfully matched after a cut. Moreover, the exit-entrance cut was not able to cue the visual attention across a cut. Perhaps this type of cut is not as suited to guide the attention of viewers and maintain a sense of continuity as the AToCC seems to suggest (Smith, 2012).

One explanation for these results is that during the incongruent trials, the icon appeared in the same location as where the exit portion of the exit-entrance cut occurred. This could have impacted the study, as the scene transition occurred in the same area as the display of the incongruent icons. Viewers could have been following a character's motion and rested their eyes where the motion stopped (Wolfe & Horowitz, 2004). Furthermore, icons that were located at the incongruent location were always located substantially closer to the exit than those located at the congruent location. This means that participants had more time to view the icons during the incongruent condition. For a display time of 200ms this could have greatly affected the results, as a viewing time of 172ms results in a recognition accuracy that is 30% higher than viewing an icon for 114ms (Intraub, 1981). This means that the small difference to perform a saccade to the same side of the screen (incongruent condition) or to the other side of the screen (congruent condition) could have had a great effect on recognition.

Another explanation of these unexpected results is that exit-entrance cuts are not as embedded into reality as match-action cuts and gaze cue cuts (Anderson, 1998). Like in real life, no context is necessary to assume the continuation of motion during a match-action cut. Additionally, in real life, individuals are interested in what someone is looking at, similar to the visual attention that gaze cue cuts provoke (Smith, 2012). While exit-entrance cuts are not difficult to understand, they have no similarity with real life. They are more similar to scenes where a character gets into their car, and arrives at their destination in the next scene (Münsterberg & Griffith, 2004). In order to establish *a priori* continuity, these types of cuts require context to allow the viewer to infer where the character will go as they leave the shot. This context was not provided during the experiment and might be necessary for the viewers expectation to be properly cued across a shot.

To see whether other characteristics of the clips in this study may have influenced the results, the clips were submitted to a closer inspection. It turned out that in some clips, the exit seemed to indicate the end of the scene, which ties into the fact that there might have been a lack of context to the characters actions. This means that it was conceivable that the participants did not interpret every exit-entrance scene as such. When the exit of the scene occurred, the scenes could be mistakenly viewed as being concluded. This could have caused the motion to not appear as continuous and the expectation to not be successfully cued across the cut (Smith, 2011). For the viewer, it could seem as if a new scene would start for which they did not have a clear and prior expectation. In other words, the exit was not interpreted as a cue. In these cases, their attention was not guided towards the entrance. When the entrance occurred, they realized that the scene was not over, and they had to shift their attention to the location of the entrance. This is an example of *a posteriori* continuity. Clips that featured a clear exit in the middle of a scene, participants expected that there would be a re-entrance at the location indicated by the exit cue. This is an example of *a priori* continuity (Smith, 2012).

To inspect whether this might have been the case, additional analyses were conducted. Subsequently, clips that could be interpreted as scene endings were excluded. This applied to seven clips. This additional analysis showed some similarities with the previous analysis. The participants spent more time looking at the icons when they were displayed for 800ms (F_1: $M = 426$, $SD = 48.6$; F_2: $M = 436$, $SD = 29.7$) than when they were displayed for 200ms (F_1: $M = 273$, $SD = 42.4$; F_2: $M = 266$, $SD = 30.5$)($F_1(1,80) = 5.663$, $p < 0.01$, $\eta^2 = .066$, $F_2(1,33) = 15.941$, $p < 0.01$, $\eta^2 = .326$). Similarly, participants were better at recognizing icons that were displayed for 800ms (F_1: $M = 2.02$, $SD = .13$; F_2: $M = 3.78$, $SD = .58$) than those that were on screen for only 200ms (F_1: $M = .69$, $SD = .11$; F_2: $M = 1.33$, $SD = .26$)($F_1(1,39) = 95.592$, $p < 0.01$, $\eta^2 = .710$, $F_2(1,17) = 107.556$, $p < 0.01$, $\eta^2 = .497$). Furthermore, the recognition scores did not significantly differ when the icons were shown congruently (F_1: $M = 1.33$, $SD = .14$; F_2: $M = 2.44$, $SD = .42$) or incongruently (F_1: $M = 1.39$, $SD = .12$; F_2: $M = 2.67$, $SD = .46$) ($F_1(1,39) = .131$, $p = .720$, $F_2(1,17) = .155$, $p = .699$). These are the results which are in line with the analysis prior to the exclusion of the clips. However, the results of the eye tracking data over items revealed that icons that were displayed congruently (F_1: $M = 398$, $SD = 47.7$; F_2: $M = 396$, $SD = 30.5$) were looked at longer than those that were displayed incongruently (F_1: $M = 302$, $SD = 43.4$; F_2: M = 306, $SD = 29.7$)($F_1(1,80) = 2.214$, $p = .141$, $F_2(1,33) = 4.504$, $p < 0.05$, $\eta^2 = .120$). The reason that the same results were not found for the data over participants, could be related to the fact that only 13 clips were analysed.

These results seem to indicate that the exits that were excluded could indeed be interpreted as scene endings and failed to cue the attention of viewers across a cut. Editing techniques are used to establish continuity and they are employed in service of the narrative

(Smith, 2011). Perhaps more so than for match-action cuts and gaze cue cuts, exit-entrance cuts require context so the viewer can more properly anticipate what the next scene should bring. This narrative context was not present in the short clips that were used during the experiment, while they should be present during normal film viewing. Moreover, the results indicate that a mean duration of fixation on the icon of either 300ms or 400ms does not significantly influence the ability to recognize objects. The research of Intraub (1981) has shown that recognition accuracy increases substantially when the different viewing times are 114ms and 172ms. Perhaps a bigger difference in viewing time is necessary to be able to significantly recall more objects than the 100ms difference between 300ms and 400ms. It can be concluded that, similar to match-action cuts and gaze cue cuts, exit-entrance cuts also create an expectation for the viewer, one that is in line with the actions that occur on screen.

Implications

This research adds to the work conducted by Van de Schepop (2018) and Eekelaar (2019), creating a collection of research on how well-known continuity editing techniques are able to influence visual attention. While not all results were in line with the hypotheses, they do support AToCC by providing evidence for *a priori* continuity (Smith, 2012). Moreover, this study supplements the knowledge about continuity editing by suggesting that exit-entrance cuts are best at maintaining continuity when they fit into the context or when they can function on their own without being mistaken for a scene transition. Film makers could use this information to better decide when to take full advantage of the exit-entrance cut.

Exit-entrance cuts can be a useful tool to lead the visual attention from one shot to the next, and to keep the viewer focused on the focal point that is performing the movement. Therefore, an interesting secondary application of editing techniques could be in the field of advertisement, as this domain is concerned with capturing the viewers visual attention. Since it has been demonstrated that the highest level of attentional synchrony occurs following a cut, this appears to be the best moment for visual attention to be led (Smith & Henderson, 2008; Smith, 2013). However, it should be underlined that ideally context is required, which is often lacking in advertising, and that the exit-entrance should not be mistaken for a scene transition.

Limitations

Smith and Martin-Portugues Santacrue (2017) found that it was difficult for viewers to determine that a cut had occurred when the post-cut motion of a match-action cut was removed. It is possible that the motion that was displayed for only four frames was too short for participants to notice that a cut had occurred. However, even without a post-cut motion, a cut should still be expected for an exit-entrance cut as an actor has just left the screen. Additionally, four frames last about 167ms, while a saccade takes 20-50ms (Fischer & Ramsperger, 1984; Rayner, 1998). If the viewers' attention was guided towards the entrance part of the exit-entrance cut, they should have had time to move their eyes. Moreover, if their eyes were in the process of moving from side to side, they should have missed the icon that was displayed in the incongruent condition (Dmytryk, 1984).

The clips that were used during the experiment always featured an entrance that occurred at the side of the screen. Furthermore, the exit never took place centrally, as these types of exits would not be very interesting to monitor with eye tracking. As a result, the viewer often had to saccade from one side of the screen, all the way to the other side in order to see the icon that was displayed in a location congruent with the expectation. As opposed to the experiment conducted with match-action cuts, 200ms might not have been enough time for viewers to perform a saccade from one side of the screen to the other and recognize or fixate on the object. Moreover, Intraub (1981) found that displaying a sequence of 12 pictures for 258ms each, resulted in participants being able to recognize 58% of the pictures. When icons are displayed for 200ms, in a sequence of 20 icons with clips in between, the recognition accuracy would be even lower. This could provide an argument for why the additional analysis did not show significant results for the data of the recognition task. Here, the data over participants showed a recognition accuracy of around 40% in the 800ms condition, while the recognition accuracy of the 200ms was around 14%. When the amount of recognition is that low, it is difficult to find that the congruent condition has a positive effect on the recognition task. In other words, the recognition accuracy was low in general which indicates that the task could have been too difficult, which could have affected the results.

Future research

To recapitulate, this research suggests that the exit-entrance cut is able to cue attention across a cut and to match the expectation of the viewer after a cut. However, the results of the study did not unequivocally show that visual attention is influenced by the exit-entrance cut. Therefore, future research is required to further investigate the impact of this cut. For follow-up research it is suggested to counteract the occurrence of the exit at the same location as the incongruent icon, by not showing the incongruent icons opposite of the congruent location. Hence, if a viewer's gaze remained at the location of the exit, their gaze would no longer be misattributed to the incongruent icon. For example, if an exit occurs at the right side and is followed by an icon located centrally at the bottom of the screen, the viewer would have no reason to linger their gaze on the location of the exit. This would also reduce the distance between performing a saccade to the congruent location and to the incongruent location.

Another solution is to not make use of icons. Instead, the exit-entrance cut could be shown fully, and the time it takes to make a saccade to the opposite side could be monitored. A baseline would have to be established for how fast viewers must look at the entrance for it to be attributed to the manipulation of the exit-entrance cut. In order to create this baseline, the time it takes for viewers to look at the entrance part when it is preceded by an exit would have to be compared with the time it takes to look at the entrance part without a previous manipulation.

Additionally, it is important that the clips are clear, unambiguous exit-entrance cuts that cannot be interpreted as the ending of a scene. Moreover, an fMRI could be used to further investigate the cognitive aspect and to allow exit-entrance cuts to occur in the middle of the screen. Another suggestion for future research is to ask the participants whether they experienced continuity and noticed a cut. If the participants' attention is properly cued across a cut, they will have experienced continuity and will not have seen the cut, which would be further evidence that visual attention is influenced by the exit-entrance cut (Smith & Henderson, 2008; Smith, 2012). Moreover, similar to the experiment by Swenberg and Eriksson (2018), interpreting pupil dilation and monitoring saccade frequency are tools that could be used to assess whether or not the viewer is experiencing discontinuity.

Exit-entrance cuts do guide the attention of viewers, but the results obtained here show that the impact of these cuts is not as pronounced as it is for match-action cuts and gaze cue cuts. This type of cut has substantial differences from the match-action cut and gaze cue cut and therefore a different research method is suggested. With some adjustments to the experimental layout, it is expected that a stronger effect of congruency will be found. To that end, it would be an interesting idea to direct a short film. At least 20 exit-entrance cuts should be included in the film, ones that cannot be confused with the conclusion of a scene. Some entrances should be congruent to the expectations created by the exit, while others should occur at the same side as the exit, as a part of the incongruent condition. As Kachkovski et al. (2019) have established, changing the axis of action should not be too much of a distraction. Several entrances could also occur at the top or bottom of the screen, while the exit occurred on the right or left side of the screen. Moreover, a comparison could be made between how fast individuals look at an entrance when it was introduced by an exit as opposed to an entrance which occurs on its own, with no build-up. Eye tracking could be used to analyze how quickly the viewers' eyes fixate on the occurrence of the entrance, and whether it can be attributed to the influence of the exit-entrance cut. It is also possible to include a recognition task in the experiment by having the actor hold up a sign with an icon for a short time after they re-enter the screen. Lastly, this method has the added benefit of providing a narrative and context for viewers, so they can more easily process the film (Loschky et al., 2015). A reduction of cognitive load reduces the amount of confusion which in turn increases the likelihood that viewers will experience a sense of continuity (Swenberg & Eriksson, 2018).

References

Anderson, J. D. (1998). *The reality of illusion: An ecological approach to cognitive film theory*. Carbondale, IL: SIU Press.

Berliner, T., & Cohen, D. J. (2011). The illusion of continuity: Active perception and the classical editing system. *Journal of Film and Video*, *63*, 44-63.

Cozijn, R. (2006). Het gebruik van oogbewegingen in leesonderzoek. *Tijdschr. Taalbeheersing*, *28*, 220-232.

Dmytryk, E. (1984). *On film editing: An introduction to the art of film construction*. City: Routledge.

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*, 1-17.

Dwyer, T., Perkins, C., Redmond, S., & Sita, J. (Eds.). (2018). *Seeing Into Screens: Eye Tracking and the Moving Image*. Bloomsbury Publishing USA.

Eekelaar, J. (2019). *Do you see what I see? The influence of gaze cue on visual attention*. (Master's thesis), Tilburg University, Tilburg, Netherlands.

Fischer, B., & Ramsperger, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Experimental Brain Research*, *57*, 191-195.

Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine*, *37*, 957-964.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 604-610.

Kachkovski, G. V., Vasilyev, D., Kuk, M., Kingstone, A., & Street, C. N. H. (2019). Exploring the Effects of Violating the 180-Degree Rule on Film Viewing Preferences. *Communication Research*. https://doi.org/10.1177/0093650219838959

Münsterberg, H., & Griffith, R. (2004). *The film: A psychological study*. Mineola, NY: Dover Publications.

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 391.

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology*, *3*, 525-551.

Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would Jaws do? The tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PloS one*, *10*, e0142474.

Olivers, C. N. (2008). Interactions between visual working memory and visual attention. *Frontiers in Bioscience*, *13*, 1182-1191.

Rutishauser, U., Walther, D., Koch, C., & Perona, P. (2004). Is bottom-up attention useful for object recognition? In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*(Vol. 2, 2-2). IEEE

Sharma, S., Kiros, R., & Salakhutdinov, R. (2015). Action recognition using visual attention. In: *NIPS workshop on Time Series (2015).*

Smith, T. J. 2006. *An Attentional Theory of Continuity Editing*. Ph.D diss., University of Edinburgh, Edinburgh, UK.

Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, *6*, 1-27

Smith, T. J. (2013). Watching you watch movies: Using eye tracking to inform film theory. In: *Shimamura, A. P. (2013). Psychocinematics: Exploring cognition at the movies*. Oxford:Oxford University Press

Smith, T.J., & Henderson, J. (2008). Attentional synchrony in static and dynamic scenes. *Journal of Vision*, *8*, 773-773.

Smith, T. J., & Martin-Portugues Santacreu, J. Y. (2017). Match-action: The role of motion and audio in creating global change blindness in film. *Media Psychology*, *20*, 317-348.

Smith, T. J., & Mital, P. K. (2011). Watching the world go by: Attentional prioritization of social motion during dynamic scene viewing. [conference abstract]. *Journal of Vision, 11*, 478.

Swenberg, T., & Eriksson, P. E. (2018). Effects of continuity or discontinuity in actual film editing. *Empirical Studies of the Arts*, *36*, 222-246.

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception*, *1*, 7-27.

Van de Schepop, E. (2018). *Ready, set, match-action! The influence of a match-action cut on visual attention* (Master's thesis). Tilburg University, Tilburg, Netherlands.

Vos, S. (2018). *Auditory Spatial Attention under Cognitive* Load (Doctoral dissertation). Ghent University, Ghent, Belgium

Walther, D., Rutishauser, U., Koch, C., & Perona, P. (2005). Selective visual attention enables learning and recognition of multiple objects in cluttered scenes. *Computer Vision and Image Understanding*, *100*, 41-63.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*, 495.

Yarbus, A. L. (1967). *Eye Movements and Vision. New York*: Plenum Press.

Zacks, J. M. (2013). Constructing event representations during film comprehension. *In: Shimamura, A. P. (2013). Psychocinematics: Exploring cognition at the movies.* Oxford:Oxford University Press

Appendix A

Table 5: The distribution of objects, display times and congruency across four movie versions. A = Congruent & 200ms. B = Congruent & 800ms. C = Incongruent & 200ms. D = Incongruent & 800ms.

Clip	V1	Object	V2	Object	V3	Object	V4	Object
Back to the Future 3	A	Owl	B	Snowman	C	Meat	D	Sharpener
Black Panther	B	Ball	C	Controller	D	Rocket	A	Guitar
Superman (clip 3)	D	Carrot	A	Go pro camera	B	Pan	C	Meat
Dr. No	C	Snowman	D	Guitar	A	Duck	B	Rocket
Ben Hur	D	Controller	A	Meat	B	Letter	C	Go pro camera
Vertigo	A	Pan	B	Owl	C	Guitar	D	Snowman
Dial M for Murder (clip 2)	B	Guitar	C	Rocket	D	Go pro camera	A	Jar
Never Say Never	A	Rocket	B	Ball	C	Owl	D	Ball
Superman (clip 2)	C	Duck	D	Cap	A	Snowman	B	Duck

Back to the Future 3 (clip 2)	D	Meat	A	Carrot	B	Cap	C	Bell
Diamonds are Forever	C	Cap	D	Duck	A	Controller	B	Shoe
Dial M for Murder	B	Go pro camera	C	Letter	D	Battery	A	Cap
Head Above Water	C	Headset	D	Shoe	A	Ball	B	Battery
Superman (clip 1)	B	Jar	C	Battery	D	Jar	A	Brush
Guns of Navarone	C	Brush	D	Pan	A	Bell	B	Controller
Haunting of Hill House (episode 1)	D	Letter	A	Headset	B	Carrot	C	Pan
Diamonds are Forever (Clip 2)	B	Sharpener	C	Jar	D	Brush	A	Letter
Twilight	A	Battery	B	Bell	C	Shoe	D	Headset
Blade Runner	D	Bell	A	Sharpener	B	Headset	C	Carrot

The Birds	A	Shoe	B	Brush	C	Sharpener	D	Owl

Appendix B

Table 6: The clips that were used in the experiment. What movie they were taken from, the time when the clip occurs in the movie, when the exit takes place and at which location.

Clip	Start clip	Onset of exit	Location of exit
Back to the Future 3	0:45:26	0:46:00	Bottom left side
Black Panther	1:48:14	1:48:30	Right side
Superman (clip 3)	1:28:33	1:28:43	Right side
Dr. No	0:03:00	0:03:11	Right side
Ben Hur	2:20:45	2:21:19	Right side
Vertigo	0:28:53	0:28:56	Left side
Dial M for Murder (clip 2)	1:29:33	1:30:20	Left side
Never Say Never	1:47:22	1:47:50	Bottom left side
Superman (clip 2)	1:11:32	1:11:56	Bottom side
Back to the Future 3 (clip 2)	1:24:09	1:24:19	Right side
Diamonds are Forever (clip 2)	1:00:46	1:01:09	Right side
Dial M for Murder	0:42:55	0:43:07	Left side
Head Above Water	0:16:21	0:16:43	Right side
Superman (clip 1)	0:53:01	0:53:31	Right side

Guns of Navarone	1:09:06	1:09:21	Left side
Haunting of Hill House (episode 1)	0:04:27	0:04:40	Left side
Diamonds are Forever	0:21:46	0:22:00	Right side
Twilight	0:03:10	0:03:29	Left side
Blade Runner	0:16:39	0:17:02	Right side
The Birds	0:59:32	0:59:43	Right side

Appendix C

The questions that were asked during the distraction task.

1. Please specify your gender:
    - Man
    - Woman
2. What is your age?
3. How many times a week do you watch a movie and/or series on average?
    - Less than once a week
    - Once a week
    - Twice a week
    - Three times a week
    - More than three times a week
4. Did you wear either contact lenses or glasses during the experiment?
    - Yes
    - No