VR Edits:

The Influence of Transition Type on Spatiotemporal Judgments in Virtual Reality

Adam Hersko-RonaTas

COGNITIVE NEUROSCIENCE HONORS THESIS

April 16, 2018

Bill Warren

Submitted in partial fulfillment of requirements for the degree of Bachelors of Science with Honors in the Department of Cognitive, Linguistic, and Psychological Sciences.

Acknowledgments

My contributions to this thesis project range from conceiving the question, reading background literature, creating the displays, running subjects, analyzing the data, and compiling the results. However, I must extend a very special thanks to my advisor Bill Warren. For without his guidance, patience, and curiosity, I may have lost my wits in navigating this thesis. He lent his expertise in suggesting readings, crafting the experimental design, and analyzing the data. I am also indebted to the Virtual Environment Navigation Lab at Brown University for providing me with the resources necessary to conduct my experiment. The graduate students in the lab – Trent Wirth, Britt Baxter, and Joey Bai – have generously offered me their help and sage advice throughout the last few semesters. I appreciate the ideas film scholar András Kovács imparted to me early on in this process from which I drew inspiration to pursue this line of questioning. And of course, I'd like to thank my parents Ákos Róna-Tas and Judit Hersko for their love, support, and granting me the opportunity to pursue my passions and interests at a university level.

Table of Contents

Abstract

With the introduction of virtual reality (VR) a new question arises: When the viewer is no longer looking at a screen but rather at a surrounding, immersive environment, how might edits – conceived in the context of cinema – carry over to this new medium? There are two schools of thought with regard to the perceptual influence of edits. The narrative editing principle maintains that narrative structure provides the source of continuity viewers perceive when watching an edited sequence of images, while the formal editing principle argues that the cuts themselves specify some form of information about the spatiotemporal relationship between two shots. Since VR provides the viewer with an added immersive/spatial element nonexistent in traditional movie-watching, the perception of space and time is expected to more closely resemble event perception. Using a novel experimental design, this study sought to examine the effects of seven basic edit types (cut, fade-to-black, dissolve, wipe, portal, dolly, dolly through aperture) on participants' judgments on distance travelled and time elapsed across a transition between two virtual environments in a head-mounted VR display. Participants reported felt changes in distance and time using magnitude estimation. The results demonstrated no significant effect of edit type on distance judgments; however, a marginally significant effect was found for time judgments. No correlation was found between the reported rankings of edits for distance and time judgments. These findings are consistent with both the narrative and formal editing principles since edits appear to specify some temporal information, while spatial information may be attributed to a higher-order narrative understanding of the images being presented.

Introduction

Whether in the comfort of a home, a plane, a hotel lobby, or a theater, screens pervade our daily lives. Often these screens stream movies, shows, or miscellaneous programming. We take our ability to perceive these rapid sequences of images spliced together in succession for granted, yet the moving image only entered the collective conscious just over a century-and-a-half ago. Moreover, with time, filmmakers discovered various ways to manipulate the image and the sequences in which they were shown in order to craft stories that take place over varying expanses of space and time. The primary tool in this effort was editing: its purpose being the ability to guide the viewer's attention by framing a specific part of the narrative world through a virtual "window" (the screen) in a particular sequence using a palette of cuts to transition from one shot to another.

However, with the introduction of virtual reality (VR) a new question arises: When the viewer is no longer looking at a screen but rather at an immersive environment, how might such transitions translate (if at all) to this nascent medium? Let's take a step back to discuss the basis of film perception. The visual system evolved in a spatiotemporally continuous world, allowing us to perceive motion, transformations, and events (Gibson, 1986). Film successfully makes use of these low-level visual mechanisms through the rapid presentation of still images, creating the illusion of motion and detecting short-range transformations between complex film frames (Smith, 2010). But the illusion of motion is not what makes cinema special. What primarily differentiates reality and film is the expansion and compression of time to control narrative. This is manifested in the editing process by ordering shots into sequences. These sequences – filmed at different places and times as well as narratively taking place at different places and times – generate a cohesive flow of discrete scenes and events. To achieve narrative continuity across shots, editors choose from a palette of cuts that convey particular spatiotemporal transitions that I will detail below. For present purposes, I have broken-down seven basic edits (for a visual representation, see *Figure 1*):

a) A cut: an immediate jump from one scene to the next.
b) A fade-to-black: the first scene fades to black then fades back to the second scene.
c) A dissolve: the first scene blends directly into the second scene.
d) A wipe (think *Star Wars*): a vertical line of transition slides across the field-of-view.
e) A portal (think *2001: A Space Odyssey*): moving through an "impossible" passage or a wormhole.
f) A dolly: passively moving through undistorted space from one scene to the next.
g) A dolly through aperture: identical to the previous edit, but moving through a narrow opening (e.g. a door).

These edits were developed in the context of cinema. When watching a film on a screen, the viewer looks *through* a "window" at a scene with varying degrees of focal length (i.e. how zoomed in the vantage is) and transitions occur between scenes on the other side of this "window." When viewing VR in a head-mounted display, however, the viewer is immersed *in* the scene as a point within a surrounding 360° optic array. In other words, VR eliminates the clear separation of viewer from a framed environment. Therefore, it follows that a scene in VR is more congruous with how we visually perceive events in day-to-day life. Editing attempts to explicitly segment narrative events in the virtual world. In the real world, however, events are ongoing and segmentation occurs simultaneously at multiple timescales (Kurby & Zacks, 2008). For example, when make a sandwich for lunch, that is an event in itself. Yet the act of retrieving ingredients from a cabinet or washing your hands is a sub-event that has no clear boundaries for

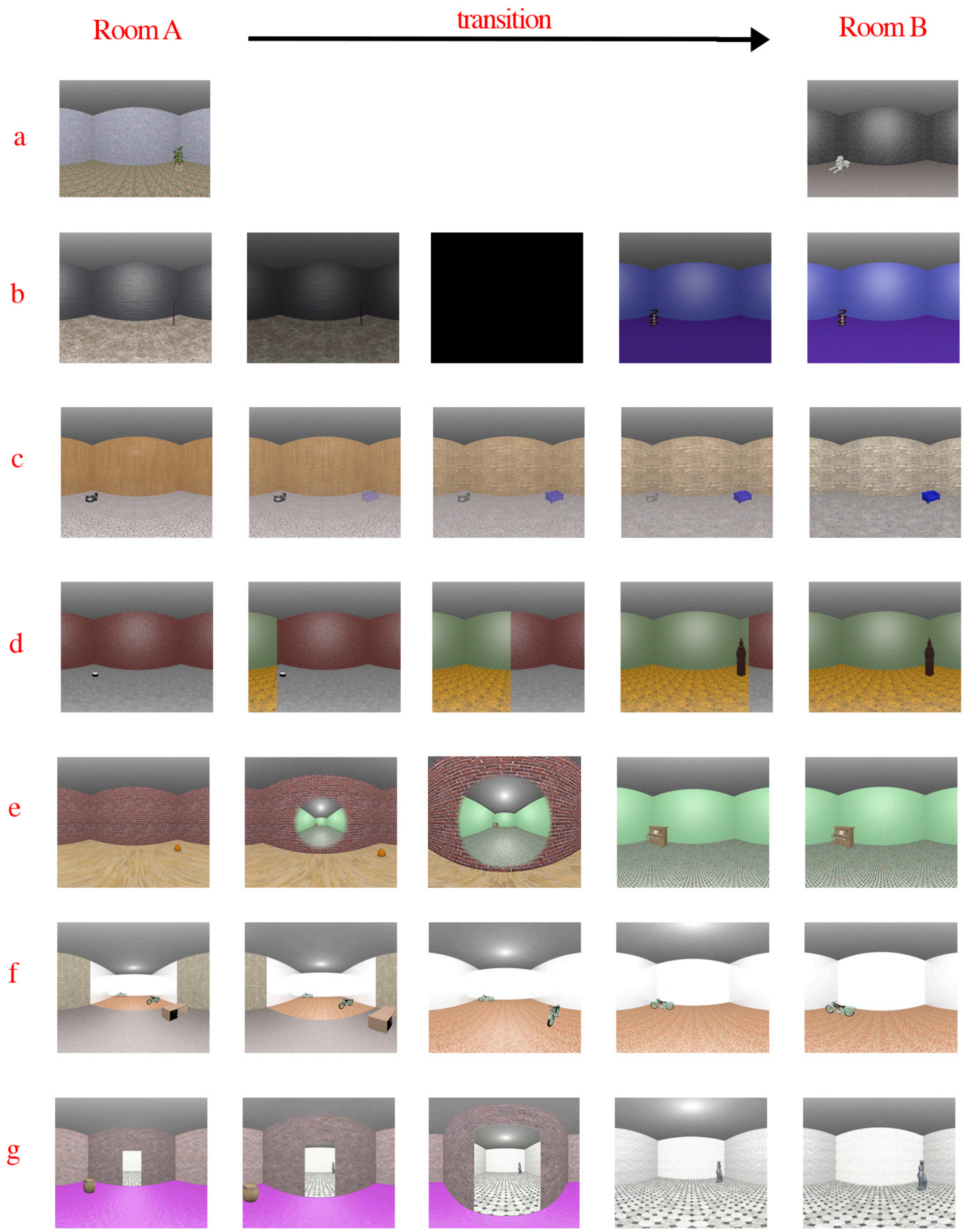

Figure 1. Sequential frames from this study's displays. The edits used to transition from one environment to another are a) cut, b) fade-to-black, c) dissolve, d) wipe, e) portal, f) dolly, g) dolly through aperture.

when/where it ends and another begins. As a result, placing transitions between two scenes in VR creates event boundaries where we would not expect them in a real, continuous environment.

Yet somehow viewers are able to piece together subscenes from presented visuals to appropriately extract a narrative. A film can be broken down into a hierarchy of units: frame, shot, subscene, scene, sequence, act, and totality of the film. In an experiment by Cutting et al. (2012), a numeric code was used to indicate the agreement among the subjects on whether a shot reflected the beginning of a new event or not. The experimenters also coded for nine independent variables: transition type, shot duration, previous shot duration, numeric representation of shot scale (e.g. close-up, medium, extreme long shot, etc.), previous shot's scale, relative change in the amount of motion, luminance, and color from previous to current shot. Overall there was strong agreement among subjects about where events began and ended (accounting for the error of picking adjacent shots; some marked the shot after the expected event boundary while others just before). Cutting maintains that subscenes are the integral level for parsing structure and that these events are primarily based on the lower-level physical attributes of the single stream of shots (2012). Understanding the characters' intentions and actions is not wholly necessary to this process of basic event comprehension. But to grant these events meaning within the greater context of the narrative, we are highly dependent on comprehending character intention; therefore, physical variables of the visual stream are crucial in guiding the top-down parsing process for us to achieve a full grasp of the story (2012).

These experiments demonstrate that perception of continuity in film operates at two levels. Low-level visual processing mechanisms make it so that viewers are unaware of cuts between shots and perceive upcoming action. Editing conventions serve to smooth over the ever-changing visual array, making it easier for the director to convey new vantage points and move them through time. However, higher-level processing mechanisms ensure that these visual cues can guide viewer attention, reinforcing their inferences about space and time, and helping assemble a cohesive narrative. All of these can successfully take place without necessarily having to construct a detailed spatiotemporal representation of the diegesis (the narrative world depicted on screen). The key assumption of Smith's (2012) "Attentional Theory of Cinematic Continuity" is that viewers do not – and should not – construct a detailed spatiotemporal representation of a depicted scene. Such effortful cognition is redundant for the perception of most important elements in a cinematic narrative. Editing a scene so that it allows the perception of continuity is not about empowering the audience to assemble a detailed representation of the space. Instead it is about enabling the viewer to shift their attention to the audiovisual details currently relevant to them and the story (2012). However, still, cuts in film can have spatially dislocating effects, especially if the cut fails to preserve the line of action specifying low-level visual cues like direction of motion on-screen (Amorim, 2002). A cut in VR is likely even more jarring since scene recognition from various vantages has provided evidence that spatial orientation is viewpoint-dependent (2002).

With VR the surrounding, moving images specify the spatial relationship between objects in the diegesis. The viewer is placed in the environment rather than being shown a sequence of framed segments from that environment. Moreover, attention is more difficult to manipulate since the creator cannot force the viewer to look in any particular direction. Therefore an edit in VR can move the viewer in the virtual world, but it cannot guide attention to a specific detail in it. So what kind of information can these edits specify (if any at all) and how might they affect the viewer's judgments (if at all) of the spatiotemporal relationship between one scene and another in a VR narrative? Views on the influence of editing vary. The two main schools of thought on continuity are the constructivist view (Smith, 2010; Magliano & Zacks, 2011) and the ecological view (Anderson, 1996; Smith et al., 2012, Cutting et al., 2012).

The constructivist view supports the "narrative editing principle," which arose in the mid-20th century during the French New Wave. Filmmakers like Jean-Luc Godard maintained that cuts should be disruptive and serve as "punctuations" to the content; the narrative would be sufficient to carry the viewer across disruptive transitions (Gianetti, 2011). Meanwhile, the ecological view maintains that the "formal editing principle" – adopted by Hollywood and considered to be the "classical" style of editing – ensures that edits provide sufficient information to specify the spatiotemporal transition between scenes. This mode of editing keeps transitions seamless and the viewers "blind" to them. Filmmakers like Karel Reisz maintained that cuts should be "invisible" to minimize disruption to the viewer's perception (2011).

Smith (2010) offers a constructivist account of associating two shots by means of the cut. The edit-blindness experienced by viewers is aided by using cuts preceded by motion onset, cuing viewers to a sequence of motion. Smith's experiments in attentional synchrony demonstrate that this intrinsic tendency to follow a line of action can be used to guide attention between frames (2010). An updated viewpoint of editing convention maintains that perceptual inconsistencies from shot-to-shot are superseded by narrative structure and that the formal editing principle – with its emphasis on seamless transitions – is not absolutely pivotal for viewers' understanding. Higher-order, narratively defined breakpoints are where this perceptual segmentation occurs, not cuts.

However, it appears as though transitions themselves possess some lower-level information too (even if perhaps the higher-order narrative information is capable of overriding any visual inconsistencies). Magliano & Zacks (2011) found that their subjects reported cuts in film as indicating a continuous flow of events (or merely a short temporal ellipsis when the sound or action did not carry across from one shot to the next), whereas a fade or a dissolve signified a shift in scenes or a leap in narrative timing. Might this be attributable to convention?

A third, and less supported view, maintains that viewers learn to interpret the arbitrary relations between two shots and thus rely on convention to glean meaning (if any) from an edit. In other words, transitions are symbolic. Yet experiments conducted by Schwan & Ildirar (2010) demonstrate that film has developed to suit our visual and attentional needs, not the other way around. In their experiment, subjects in rural Turkey with no prior experience watching films were able to comprehend cuts across events, ellipses, and edits matched on action (2010).

While cinema was always capable of manipulating the viewer's perception of space and time, the primary question remains whether the spatiotemporal relation between scenes in VR is interpreted based on semantic content (aligned with the narrative editing principle) or visually specified by the transition (aligned with the formal editing principle)? In VR, is it the higher-order storyline that carries a bulk of the perceptual weight or do the transitions themselves provide the glue with which viewers are better capable of perceiving the story? If the relationship between two shots is visually specified we would expect viewers to make consistent judgments about the distance travelled and time elapsed during a transition, regardless of whether a storyline is present.

This experiment set out to determine whether viewers' judgments of how much distance and time elapses in the narrative world across edit boundaries would be affected by the added factor of "presence" – the sense of being spatially grounded in an environment. My hypothesis follows the formal editing principle's way of interpreting cuts in assuming the transitions themselves specify some spatial and temporal relationship between shots and scenes. I hypothesize that each type of edit will be located on a spatial dimension that specifies the perceived distance travelled between two virtual spaces. Furthermore, each type of edit will be located on a temporal dimension that specifies the perceived time elapsed between two virtual spaces. The order of transitions on these dimensions will differ for distance and time (for a visual representation of this ranking, see *Figure 2*.).

In this experiment, subjects viewed a series of ambiguously spaced rooms in a virtual head-mounted display. One virtual room would transition into another virtual room using one of the aforementioned seven transition types. Subjects were then asked to provide a magnitude estimation judgment for how much distance they felt was travelled or how much time had elapsed between the two virtual scenes.

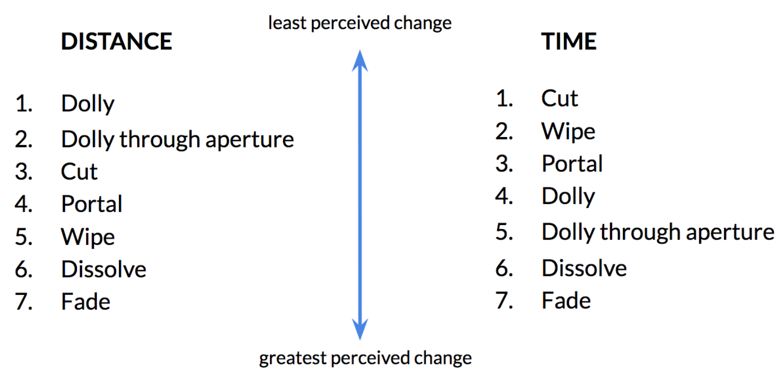

Figure 2. The hypothesized relative ordering of transitions based on the perceived amount of change in distance and time. The ordering is based on both existing literature and speculation based on my editing experience. The dolly is thought to indicate the least amount of distance travelled since that transition is an unbroken, continuous event where distance is specified. Dolly through aperture is a close second since passing through doorways can induce event segmentation and perhaps skew distance and time judgments as being larger (Radvansky & Copeland, 2006). Cuts are likely to indicate a slightly larger jump in distance since the relationship between spaces is not visually specified there; however it implies a continuous flow of events and is therefore hypothesized to indicate the least amount of perceived change in time (Magliano & Zacks, 2011). Wipe and portal were both speculative. Dissolve signifies a shift in scenes or a larger leap in narrative timing and is therefore placed last (2011). Fade (since it dips to black first), may indicate a larger jump for both distance and time judgments since there is a clearer demarcation between the first scene and the next.

Method

Subjects

The subjects were 11 naïve, paid, student volunteers at Brown University. Initially 12 students were tested, but one subject's data was removed due to large outliers in their estimates.

Displays

The displays consisted of 2 rooms (randomly paired from 14 total rooms) with 1-of-the-7 transitions linking them. These random room pairings occurred before they were assigned a transition. The 14 rooms (10m x 10m x 6m) each had different wall and floor textures. Each was lit using the same virtual lighting. To help differentiate the rooms, two of a randomly selected object (e.g. a house-plant, motorbike, marble statue, etc.) was placed in each room with one copy in front and the other behind the participant. In each display, the camera began in the center of the first room and ended in the center of the second.

Each display lasted 8 seconds. Transitions that were movement-based (dolly, dolly through aperture, and portal) all lasted 6 seconds so the camera to move from the center of one room to the center of the next. Transitions that were graphical (dissolve, fade, and wipe) all lasted 3 seconds. Cuts, by their nature, were instantaneous and took no time at all.

The virtual rooms and objects in them were generated using MAXON's Cinema 4D 18.1 graphics software and its embedded model and material packages. Cineversity's CV Toolbox (namely its CV-VRCam plugin) allowed for the creation of a virtual camera that could render equirectangular and omni-directional stereo frames within Cinema 4D (see Figure 3). These frames were placed in a sequence and post-processed in Adobe After Effects 2018. These sequences were exported as .mp4s using the H.264 video codec in Adobe Media Encoder 2018.

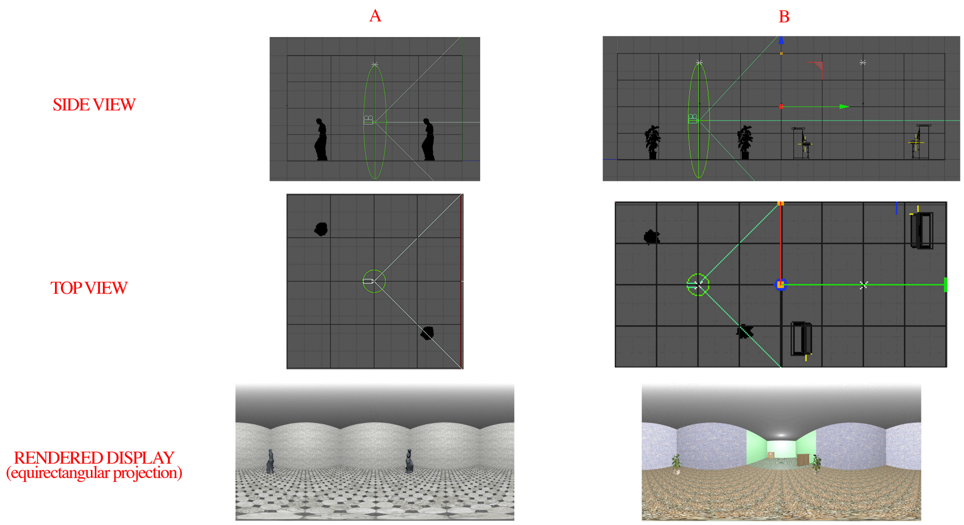

Figure 3. Room arrangements and camera configurations differed depending on edit type. For the cut, dissolve, fade, portal, and wipe conditions, the room displayed after the transition was not visible to the participant before the transition took place (configuration A). For dolly and dolly through aperture, the second room was visible prior to the transition by nature of the edit type (configuration B). For visual reference, see the rendered equirectangular (stretched 360° panorama) displays above.

Equipment

The pre-rendered displays were presented using Vizard 5, running on an MSI VR ONE 7RE-065US Backpack PC and presented to the viewer in an Oculus CV1 head-mounted display. The displays were viewed while sitting on a waist-high stool and were updated based only on head orientation (not head position).

Design

The independent variable in this experiment was edit type (7 total; cut, dissolve, dolly, dolly through aperture, fade, portal, wipe). The dependent variable was each subject's magnitude estimation of distance or time. With 7 transitions, 7 repetitions, and 2 judgments (distance/time), there were a total of 98 trials per session.

Procedure

Each session was broken into two blocks (distance judgments/time judgments). Block order was counterbalanced, with half the participants beginning with distance judgments and the other half beginning with time judgments. Subjects were seated on a waist-high stool while wearing the head-mounted display and instructed that they were about to watch a virtual reality movie. In each trial, characters – such as themselves – would travel from one scene to the next. In this imagined world there were many rooms that were not necessarily adjacent. They could be in the same building or in a different city, with no fixed configuration. Their task was to verbally indicate how far it felt they had travelled or how much time it felt had passed in this imagined world between the two rooms. They were instructed to use a magnitude estimation scale by assigning a positive, proportional number to the perceived distance/time changed. The dolly condition served as the *standard* or *modulus* since the dolly transition was unbroken and the distance/time was visually specified; this was labeled a ’10.’ Travelled distances or times that seemed to be 10 times that, for example, would be reported as ‘100’; those that seemed to be 1/10 of that would be reported as '1'. Trials were randomized within blocks so no transition was repeated before all 7 had been displayed. After the experiment ended, subjects were debriefed on nature of the study and answered a series of questions about the types of edits they could recall as well as any patterns they noticed in their responses.

Data Analysis

Because of the subjective nature of the task, participants used magnitude estimation to report their judgments. This way, each participant could establish a personal ratio scale. To equalize these individual scales across subjects, each subject's ratings were transformed so the subject mean was equal to the group mean (Moskowitz, 1977; Stevens, 1975). First, the geometric means of each subject's data and the entire data set were computed. Then each subject's ratings were multiplied by the ratio of the grand mean to that subject's mean, so that the geometric means for each participant's data and the entire data set were equated. This preserved the ratio scale, which would have been lost had the data been normalized instead.

Results

Distance

A repeated-measures ANOVA was performed on the equalized distance judgments. There was no significant effect of transitions on distance judgments [$F(6, 5)=.479$, $p=.802$]. A Spearman's rank order correlation was performed on the expected and observed ranking of edits. There was a moderate negative correlation between them, $r_s = . -.5357$, implying that contrary to what was expected, dolly was ranked on the higher end of the judged distance scale while wipe and portal were ranked on the lower end (see Figure 4). These results must be taken with a grain of salt though since the distance judgments had high variance and the ANOVA was not significant.

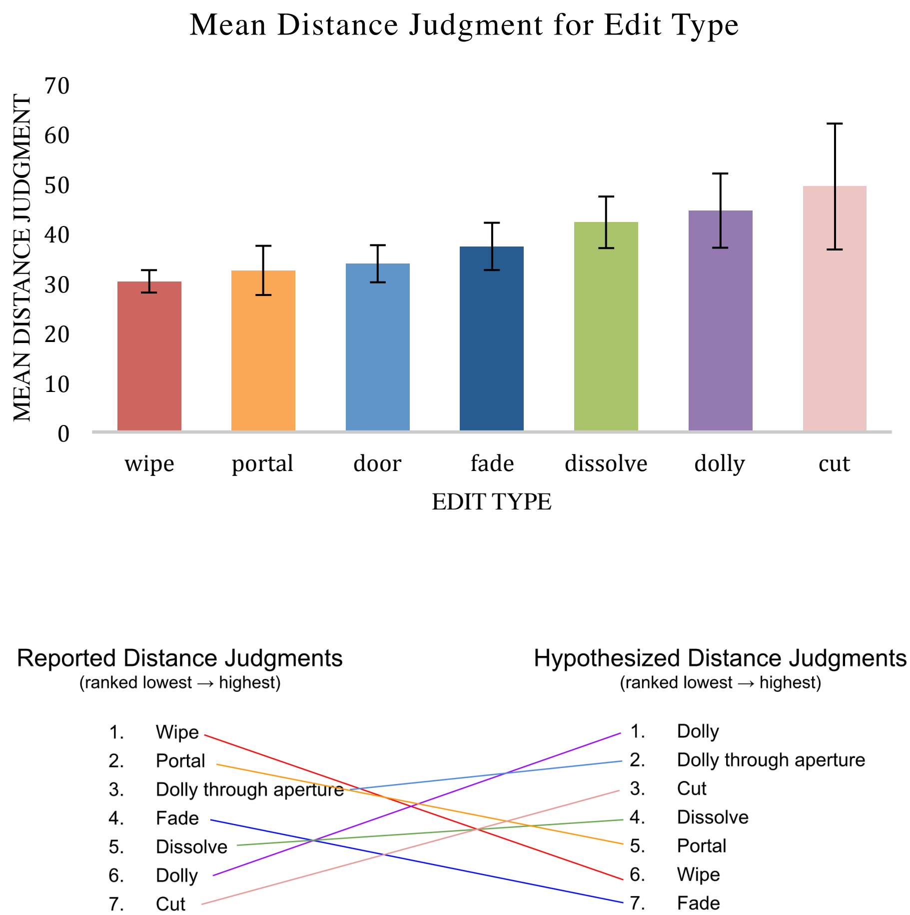

Figure 4. A) The mean distance judgments for edit type ordered from lowest to highest. B) The resulting rank order was moderately, negatively correlated with the expected order.

Time

A repeated-measures ANOVA was performed on the equalized time judgments. There was a marginally significant effect of transitions on time judgments [F(6, 5)=3.869, p=.080]; however, Bonferroni-corrected pairwise comparisons were not significant. A Spearman's rank order correlation was performed on the expected and observed ranking of edits. There was a strong positive correlation between them, r_s =.9643, meaning the rank orders were almost identical except that the dolly and dolly through aperture conditions were switched (see Figure 5.)

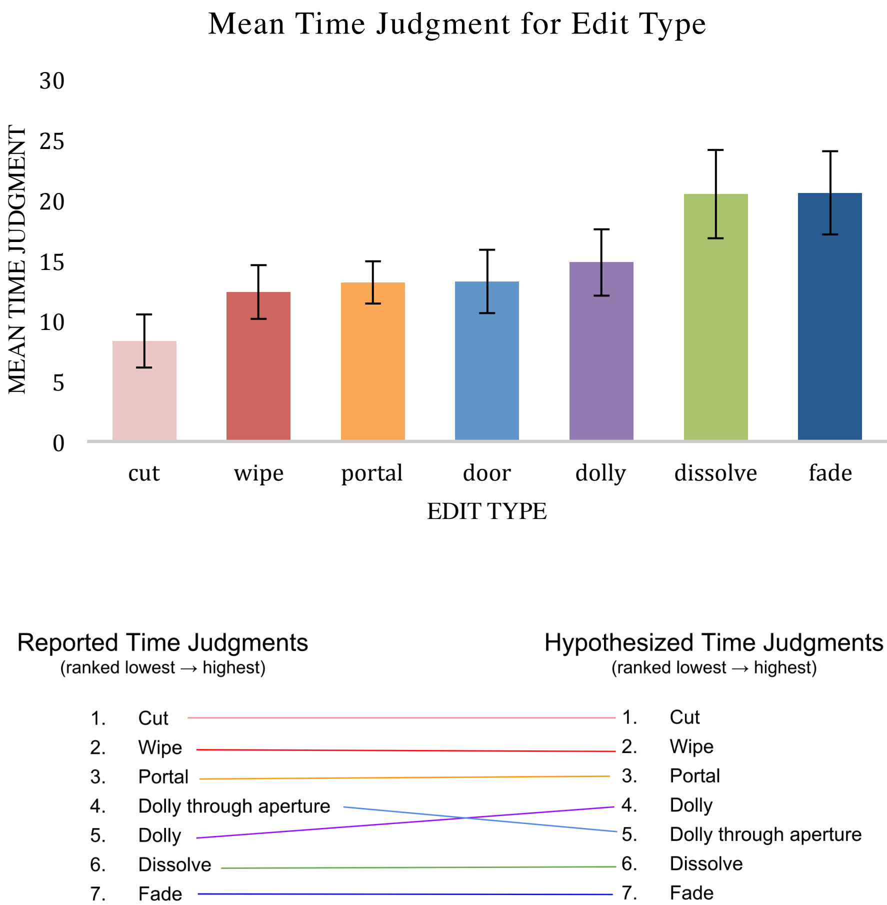

Figure 5. A) The mean time judgments for edit type ordered from lowest to highest. B) The resulting rank order was strongly correlated with the expected order.

Distance vs. Time

A Spearman's rank order correlation was run to assess the relationship between the ranking of reported distance judgments and the ranking of reported time judgments. There was a weak correlation between them, r_s=.1071, implying that edit type did not influence distance and time judgments in the same way (see Figure 6.).

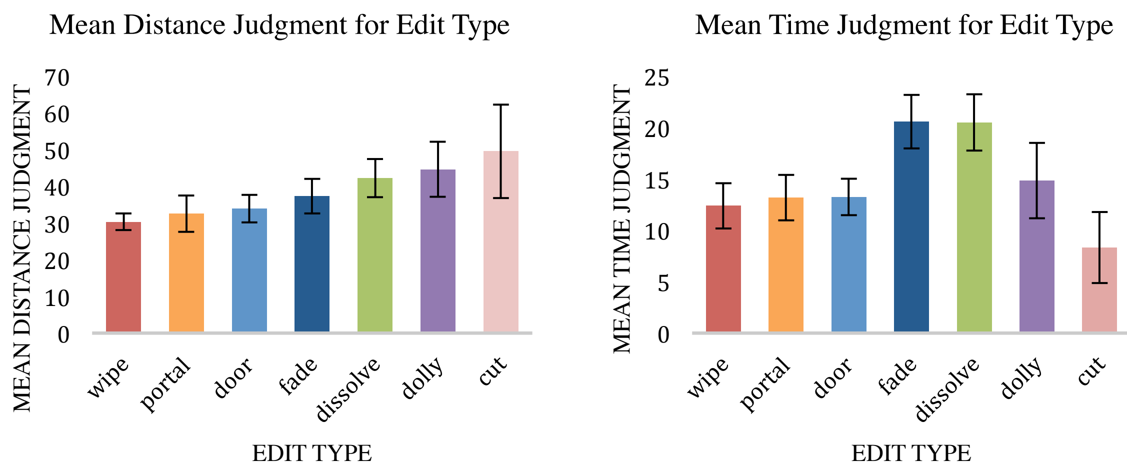

Figure 6. A side-by-side comparison of the reported rank orderings for mean distance and time judgments based on the ordering for distance. No reliable correlation can be noted.

Discussion

In this experiment we found no support that transitions have any significant effect on distance judgments (p=.802). However, we did a marginally significant effect of transitions on time judgments (p=.080). Furthermore, it is clear there is no relationship between the ordering of edits based on distance and time judgments. This is consistent with the hypothesis that an edit does not necessarily specify the same change in both distance and time. Again, this interpretation should be taken with a grain of salt given the high variance in responses for the distance condition.

Yet it appears there is preliminary evidence that elapsed time is perceptually specified by transition. The cut provided the shortest time estimates while the fade and dissolve provided the longest time estimates. This is consistent with the time hypothesis that cuts are likely to indicate a continuous flow of events while fades and dissolves indicate a larger jump in narrative timing (Magliano & Zacks, 2011). It also offers some support for the formal editing principle in that edits themselves bear perceptual information that influences understanding of timing.

There is no evidence, however, that distance travelled is specified by transition. This suggests that narrative content or higher-order contextual details in the environment may be needed. Participants' answers to the post hoc questions revealed that some of them actively sought semantic clues in the displays when making both distance and time judgments, even after being explicitly instructed that the materials and objects in the environment were merely to help differentiate the rooms. Seeking such clues in the environment to construct a higher-order 'narrative' of how these ambiguous spaces relate to one another provides evidence for the constructivist notion that narrative can help disambiguate spatial relations between two shots (Smith, 2012). Another demonstration of this arises from a key component of the editing process: the juxtaposition of images, which can in turn provide contextual framing. For example, placing a shot of the protagonist's neutral expression after a shot of a casket being lowered into a grave would lead the viewer to assume this reaction is directly related to the object and action depicted prior – that is, one of mournfulness (Bordwell, 2002). Therefore it appears as though both the lower-level visual cues offered by edit placement and higher-order semantic inferences made by the viewer influence narrative perception. *A priori* spatiotemporal continuity through the cut is established by low-level visual cues, but an active reconstruction of the relationship between shots is primarily an illusion pieced together *a posteriori* by the viewer using higher order processes (Smith, 2012). So a robust theory of cinematic continuity must highlight this reciprocal relationship between the viewer and the media-form itself.

One criticism of the experimental design is the duration of the edits since they varied by nature and could not be completely equated. A cut is instantaneous, whereas the fade, dissolve, and wipe took 3 seconds; dollies and the portal lasted 6 seconds. But a 6-second dissolve is unusually long while a 3-second dolly is nauseatingly fast. I argue that this discrepancy is not highly influential given that the length of the transitions in real-time did not align with their ultimate rank ordering. Fade and dissolve ratings were higher than those for the portal and dolly conditions even thought the edits themselves lasted half the time.

The primary limitation of this study is that it was statistically underpowered. A post hoc power analysis demonstrated that approximately 30 subjects would be necessary to provide a reliable result, given the observed effect sizes. A power analysis could not be performed beforehand because this was an original experimental design and the effect size was not known in advance. Moreover, the lack of agency in the virtual world meant subjects were unable to move about the virtual space. This meant that motion parallax from head movement was missing, but also that no spatiotemporal relationship could be established with the virtual environment by the perceiver through their own bodily action. After all, presence – a key factor for immersion (Bates, 1992; Loomis, 2016) – is strongly contingent on a combination of self-location and the set of perceived possible actions in a given environment (Hartmann et al., 2016).

Given the abstract nature of this study's task, it is also possible that participants were unable to generate reliable magnitude estimations and therefore gave arbitrary responses; however, this was accounted for by offering participants a set of 7 practice trials (one for each edit type) before the experiment began to help them calibrate their personal ratio scale. If repeated, this experiment would likely be better served by a forced-choice methodology. That way, two edit types could be displayed back-to-back and subjects would be instructed to report which one felt "further" or "later," for instance. Moreover, it would have been helpful to have subjects rank the edits themselves post hoc based on their own intuitions.

Future research should explore the relationship between edits on narrative perception in VR since this study attempted to isolate edits and remove any influence of narrative. For example, one could create a mystery VR experiment where the subject is virtually transported to various rooms in a building shortly after a murder has taken place. Each room would have a different potential murder weapon and/or suspect. The subject would then answer questions about who they thought had committed the murder and by what means, given their/its proximity to the crime scene. By randomizing what transition is used to move the viewer from the scene of the murder to peripheral rooms in the same building, one could examine whether the edit type had any significant effect on subject's accusations. The primary issue with a study like this, however, is that it is difficult to operationalize a concept as abstract as 'narrative' or even separate the semantic value certain spaces have (e.g. a kitchen and a dining room are likely to be interpreted as being closer together than a garage and a bedroom). Even a narrative distilled to its most basic elements (e.g. characters, setting, conflict) will vary wildly in its interpretation between subjects simply because story-comprehension is inseparable from personal histories, biases, and temporary mental states (Genereux & Mckeough, 2007). Viewers can extrapolate meaning from even the most meaningless stimuli (Shermer, 2008). Therefore any effect on narrative interpretation may not be attributable to edits alone in such a study.

In conclusion, I speculate that both lower-level visual and higher-order narrative information both contribute to the interpretation of spatiotemporal transitions. Edits themselves can specify some temporal information while spatial comprehension is more likely to be specified by visual markers in the environment or an active understanding of the presented narrative. I further speculate that such higher-order cues can eclipse lower-level ones in instances of apparent confusion or contradiction. It will be curious to see if a new editing 'language' emerges in this recently evolved media landscape where a more robust sense of spatial presence is introduced.

References

Albers, D., & Gallege, S. (2011). Perception in Film and Animation. Retrieved December 10, 2016, from http://pages.cs.wisc.edu/~dalbers/perception/film.html

Alsobrook, L. (2015). The spaces of narrative consciousness: Or, what is your event? *Technoetic Arts Technoetic Arts, 13*(3), 239-244. doi:10.1386/tear.13.3.239_1

Amorim, M., Thibault, G., Droulez, J., & Berthoz, A. (2002). The price that cinematography is paying to spatial cognition.

Anderson, J. (1996) The reality of illusion: An ecological approach to cognitive film theory. S. Illinois Univ. Press.

Bates, J. (1992). Virtual Reality, Art, and Entertainment. *Presence: Teleoperators and Virtual Environments, 1*(1), 133-138. doi:10.1162/pres.1992.1.1.133

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*. doi:10.1038/nrn3122

Bordwell, D. (2002). Intensified Continuity Visual Style in Contemporary American Film. *Film Quarterly,55*(3), 16-28. doi:10.1525/fq.2002.55.3.16

Carter, E. J., Sharan, L., Trutoiu, L., Matthews, I., & Hodgins, J. K. (2010). Perceptually motivated guidelines for voice synchronization in film. *ACM Transactions on Applied Perception,7*(4), 1-12. doi:10.1145/1823738.1823741

Cutting, J. E., Brunick, K. L., & Delong, J. E. (2011). The Changing Poetics of the Dissolve in Hollywood Film. Empirical Studies of the Arts, 29(2), 149-169. doi:10.2190/em.29.2.b

Cutting, J. E., Brunick, K. L., & Candan, A. (2012). Perceiving event dynamics and parsing Hollywood films. Journal of Experimental Psychology: Human Perception and Performance, 38(6), 1476-1490. doi:10.1037/a0027737

Delong, J. E., Brunick, K. L., & Cutting, J. E. (2013). Film Through the Human Visual System. The Social Science of Cinema, 123-137. doi:10.1093/acprof:oso/9780199797813.003.0005

Diemer, J., Alpers, G. W., Peperkorn, H. M., Shiban, Y., & Mahlberger, A. (2015). The impact of perception and presence on emotional reactions: A review of research in virtual reality. Frontiers in Psychology Front. Psychol., 6. doi:10.3389/fpsyg.2015.00026

Diettrich, O. (2001, January). Virtual Reality and Cognitive Processes. In *Virtual Reality: Cognitive Foundations, Tech. Issues & Philosophical Implications* (pp. 203-219).

Genereux, R., & Mckeough, A. (2007). Developing narrative interpretation: Structural and content analyses. *British Journal of Educational Psychology,77*(4), 849-872. doi:10.1348/000709907x179272

Germeys, F., & D'Ydewalle, G. (2005). The psychology of film: Perceiving beyond the cut. Psychological Research, 71(4), 458-466. doi:10.1007/s00426-005-0025-3

Giannetti, L. (2011) Understanding movies. (12th Ed.). Englewood Cliffs, NJ: Prentice Hall.

Gibson, J. J. (1986). Motion Pictures and Visual Awareness. In *The Ecological Approach to Visual Perception* (pp. 292-306). New York, NY: Psychology Press.

Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., . . . Sacau, A. M. (2016). Spatial Presence Experience Scale. *PsycTESTS Dataset,28*(1). doi:10.1037/t48962-000

Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008). Enhanced Intersubject Correlations during Movie Viewing Correlate with Successful Episodic Encoding. Neuron, 57(3), 452-462. doi:10.1016/j.neuron.2007.12.009

Hirose, Y. (2010). Perception and memory across viewpoint changes in moving images. *Journal* *of Vision, 10*(4), 1-19. doi:10.1167/10.4.2

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Science, 12*(2), 72-79.

Loomis, J. M. (2016). Presence in Virtual Reality and Everyday Life: Immersion within a World of Representation. Presence: Teleoperators and Virtual Environments, 25(2), 169-174. doi:10.1162/pres_a_00255

Magliano, J. P., & Zacks, J. M. (2011). The Impact of Continuity Editing in Narrative Film on Event Segmentation. *Cognitive Science,35*(8), 1489-1517. doi:10.1111/j.1551-6709.2011.01202.x

Mobbs, D., Weiskopf, N., Lau, H. C., Featherstone, E., Dolan, R. J., & Frith, C. D. (2006). The Kuleshov Effect: The influence of contextual framing on emotional attributions. *Soc Cogn Affect Neurosci Social Cognitive and Affective Neuroscience, 1*(2), 95-106. doi:10.1093/scan/nsl014

Moskowitz, H. R. (1977). Magnitude Estimation: Notes on what, how, when, and why to use it. Journal of Food Quality, 3, 195-227.

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition,34*(5), 1150-1156. doi:10.3758/bf03193261

Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? *Memory & Cognition, 28*(2), 214-223. doi:10.3758/bf03213801

Schwan, S., & Ildirar, S. (2010). Watching Film for the First Time: How Adult Viewers Interpret

Perceptual Discontinuities in Film. *Psychological Science, 21*(7), 970-976. doi:10.1177/0956797610372632

Shermer, M. (2008). Patternicity. *Scientific American, 299*(6), 48-48. doi:10.1038/scientificamerican1208-48

Smith, T. J. (2012). The Attentional Theory of Cinematic Continuity. Projections, 6(1), 1-27. doi:10.3167/proj.2012.060102

Smith, T. J. (2010) Film (cinema) perception. In: Goldstein, B.E. (ed.) Encyclopedia of Perception. Thousand Oaks, U.S.: Sage. ISBN 9781412940818.

Smith, T. J., & Henderson, J. M. (2008). Edit Blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research, 2*(2), 6th ser., 1-17. doi:DOI 10.16910/jemr.2.2.6

Smith, T. J., Levin, D. T., & Cutting, J. E. (2012). A window on reality: Perceiving edited moving images. Current Directions in Psychological Science, 21(2), 107-113.

Stevens, S. S. (1975). Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. New York City, NY: John Wiley & Sons.

Warren, W. H. (2005). Direct Perception: The View From Here. *Philosophical Topics, 3*(1), 335-361. doi:10.4135/9781412972000.n115

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*(2), 273-293. doi:10.1037/0033-2909.133.2.273

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2007). Human Brain Activity Time-Locked to Narrative Event Boundaries. *Psychological Science, 18*(5), 449-455. doi:10.1111/j.1467-9280.2007.01920.