Kuleshov Effect in VR: Exploring the Storytelling Possibility of Binocular Rivalry in Head Mounted Display

Stereopsis is not the only perception one could get when they put on the Head Mounted Display (HMD) for the Virtual Reality (VR) experience. This paper gives an overview of all the possible cognitive results (stereoscopy, binocular rivalry, and binocular suppression) and their applications under different HMD using scenarios. Among all of the three results, the paper tries to discuss the unrealism storytelling possibility of binocular rivalry, a unique phenomenon when two different images are presented to both eyes of the user.  A comparative research is conducted between the binocular rivalry and Kuleshov Effect in the motion picture. The experiment result indicates that dichoptically viewing two images could actually yield to more meanings than watching the two images in isolation under the binocular rivalry scenario.

# Introduction

Virtual Reality is not a new invention in the digital age. Stereoscopy technique has been used to create the illusion of depth and space since two hundred years ago[1]. During the development of virtual reality, the VR pioneers have not stopped trying to deliver the more realistic experience. Bringing people to where they will never get the chance to go has become the goal for nearly every VR company one can simply tell that by paying attention to the number of appearances of the phrase “being there” in the press about VR. Such mainstream visions of VR bring new questions into play: Is re-creating the reality the ultimate capability of digital media? Quoting McLuhan’s hackneyed aphorism “the medium is the message”[2], the art movement going beyond the realism that liberates people from the sensing restriction and gives people a new perception of the reality could be found in any medium (e.g., the impressionism after the realism in painting) but not yet in VR. Because  stereoscopy is still the fundamental basis of any current VR technology (and is going to maintain this way in the foreseeable future), one of the approaches to the unrealism VR could be related to binocular rivalry - the unique epistemic phenomenon occurring in dichoptic presentation (one image is presented to one eye and a very different image is presented to the other) of stereoscopy.

Because our eyes are horizontally separated, each eye has its own perspective of the world, and thus both eyes receive slightly different images. Stereopsis is the perception of depth that is constructed based on the difference between these two retinal images. The brain fuses the left and right images and, from retinal disparity, i.e., the distance between corresponding points in these images, extracts relative depth information. When it comes to the stereoscopic three dimensional (3D) displays, usually a pair of goggles is used to present one image one eye and a very slightly different image to the other, the brain can fuse these two images into a single perception and yield stereopsis. However, to perfectly create the 3D aspects of a virtual scene, the two images presented need to be precisely adjusted according to human’s binocular vision, otherwise, our brain will extract the wrong depth information that contradicts to our epistemology [3]. When the pair of images is not set up correctly (the asymmetrical stereoscopy), dichoptically viewing those images pairs may produce competition in the form of binocular rivalry: perception alternates between different images presented to each eye[4]. Although the mechanisms behind binocular rivalry are still debated[5], previous research[6]; [7]; [8] has proven that, asymmetrical stereopsis will lead to three different epistemic perception, depending on the different levels of visual stimuli of the images: (1) **stereopsis**, when the images are still similar enough for the brain to extract the depth information; (2) **binocular rivalry**, when the two images are too different fuse by the brain and they have similar sensory eye dominances; (3) **binocular suppression** (only one of the images is seen while the other is hidden) when one image leads to a significant eye dominance [9] than the other[10].

## 

## Stereopsis & Binocular Suppression in Asymmetrical Stereoscopy

Other than the traditional symmetrical stereoscopy that is widely adopted in education and entertainment industry -3D TV, 3D movies, VR etc. - the stereopsis yielded by the asymmetrically viewing process has been researched in many aspects. Such cognitive phenomenon has been used in 3D content streaming where one image is transmitted in full resolution while the other is compressed. The stereopsis will still be achieved and the missing pixels in the compressed image will be filled by the corresponding ones from the full resolution image by the brain when viewing them stereoscopically[11]. This technique will preserve the 3D viewing quality while preventing the bandwidth to be doubled in 3D content compared to the 2D streaming. Another applicable situation of the asymmetrical stereopsis is the monocular, transparent (a.k.a “see-thru”) head-mounted display that project 2D information to only one transparent goggle of the headset [12]. Binocular opaque HMDs are useful for immersive virtual reality applications while monocular transparent displays are preferred when interacting with the world while looking at the display[13]. Specific potential applications could be found mostly in Augmented Reality field, providing users with additional visual information or visual navigation aids[14]. However, although the two images are shown to the eye asymmetrically, as long as the brain infuse them together into the single vision, there is no difference between the asymmetrical stereoscopy and the symmetrical one on the cognitive level. Therefore, the storytelling potential remains the same.

Binocular suppression happens when one image is extremely “interesting” to the brain and gets all of its attention in dichoptic viewing. As a result, the visual stimulus from the other image is completely blocked[15]. The stereoscopical watching experience under such situation will be no different than watching a flat image with one eye closed.

## Storytelling with Binocular Rivalry

Binocular rivalry is a stage in between stereopsis and binocular suppression dichoptic presentation. When the images are too different to be fused and, at the same time, are both similarly “interesting”, the brain will find difficulty in choosing which image it should see and constantly switch focus between them. For example, when an image of a house is presented to one eye and an image of a face to the other, then subjective experience alternates between the house and the face.  There have been many empirical studies of binocular rivalry but the data they produce are conflicting and it is very difficult to give them an unequivocal interpretation. A number of proposals have been made but the neurocognitive mechanism that explains this visual effect remains unresolved [16]; [17]; [18]. Some research tried to look at this phenomenon from the cognitive perspective, holding the belief that binocular rivalry is an epistemic response from the brain to a seemingly incompatible stimulus condition where two distinct objects occupy the same spatiotemporal location.

In his famous paper[19], Eisenstein believes that the nature of the art is the conflict between natural existence and creative tendency, and montage creates such conflict by pictorially placing two immobile images (or extendedly, two shots) next to each other. Eisenstein’s montage theory is arguably considered as the extended explanation and application of the Kuleshov Effect [20] by which viewers derive more meanings from the interaction of two sequential shots than from a single shot in isolation. Eisenstein argues that, when two pictures (or two shots) in the motion picture are sequentially aligned, they, like the stereoscopic effect that is created by human perception, are actually superimposed on top of each other in the audience’s brain – the conflicts of thoughts are formulated by the conflicts of shots. In other words, the conflicted thoughts from people’s brain are conditioned and derived by the combination of shots created by filmmakers. By manually changing the shots and cutting pace, filmmakers are able to direct how people’s ideas are conflicted and sequentially formed into a new meaning. In this light, although the binocular rivalry is still about the conflicts of consciousnesses, this phenomenon is autonomously controlled by people’s brain and free from the control of others. Contrary to montage in motion pictures, although the images under the dichoptic presentation are still selected by others, how the conflicts would happen is completely determined by the viewer’s brain itself. However, there is not enough research about whether and how this neurological autonomy would affect the perception result of the combination of two images. Would dichoptically viewing two images yield more meaning than watching the two images in isolation?  If it would, will the meaning perceived by the brain during binocular rivalry differ from the interpretation from watching the same pair of images sequentially in motion pictures?

# Experiment

## Questions & Hypothesis

To study the storytelling possibility of the mystical binocular rivalry, we first need to validate that, like the Kuleshov Effect in motion pictures, dichoptically viewing two images could yield more meanings than watching the two images in isolation. Under such hypothesis, the additional question could be formed: is the new meaning created in the binocular rivalry different from that in the montage? To verify the hypothesis, a between-subject experiment is conducted.

## 

## Apparatus

The state-of-art commercial VR headset (Oculus Rift) that has 1080×1200 resolution per eye, a 90 Hz refresh rate and 110° field of view was used in user testing. Each participant was asked to adjust the lenses spacing and height to their most comfortable level, so they could see the arguably clearest stereoscopic VR content in the current major consumer market.

## 

## Participants

43 college students from New York University, USA (26 female; 17 male; age 21- 31) participated in the experiment. All participants have used the VR headset before. None of them knew the concept of binocular rilvary.

## Materials

The testing application is made in Unity. In the application, users start in a completely dark VR environment where a 1.2\*1.5-meters-size image is placed 1 meter in front of them. The locomotion in this VR environment matches the physical movement of users in the real world: they could freely move and rotate both their heads and bodies to see the image from different angles and positions. The Unity project was built to the Oculus Runtime Executable file via Oculus Utilities for Unity (Version 1.25) and the final visual was automatically rendered into split-screen stereo with distortion correction for each eye via the Oculus Compositor. (<https://developer.oculus.com/documentation/pcsdk/latest/concepts/dg-render/>)



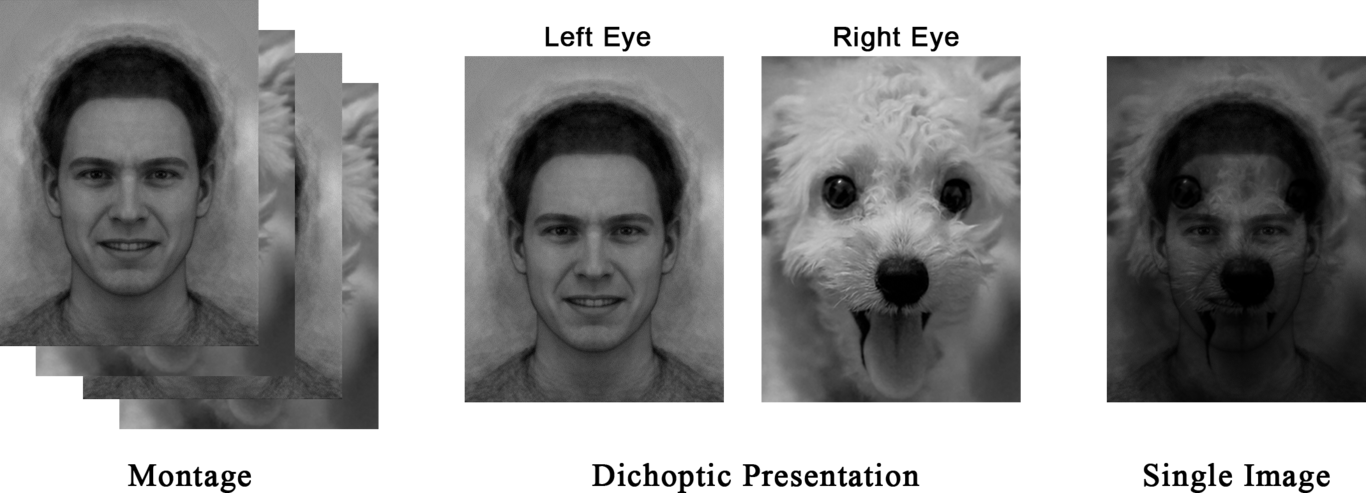
Images used in experiment

As shown in Figure 1, four images were used in the experiment. Image (a) is from the emotion perception task developed by Penton-Voak et al [21]. The researchers composited prototypical happy images of 20 individual male faces showing a happy facial expression and the same 20 faces showing an angry expression and created the image of a statistically approved neutral face. The original images came from the Karolinska Directed Emotional Faces [22].  Image (b) is a royalty free photo of a puppy that was used with the intention to arouse the happiness emotion according to Barratt’s retest of the Kuleshov Effect [20]. Images (c) is produced by multiple blending [23] image (a) and (b), and image (d) is composed of only random noise. All images are converted into grayscale by reducing the level of saturation to make sure that image (a) and (b) share a similar eye significance factor[24].

## Method

As shown in figure 2, three types of visual stimulus – (1) montage (2) dichoptic presentation and (3) a single image–were presented to the participants in each individual trial in the VR environment. A figure of a man with neutral facial expression could be seen in each visual stimulus and the noise image (4) from Figure 1 was shown to the participants in the 3-second-interval between each stimulus.

**The montage** was the looping sequence composed of the image (a) and image (b) from Figure 1 that lasted 3 seconds individually. The image (a) was always shown first and the sequence would cut to the image (b) after 3 seconds. The duration of the image was set according to the estimated Average Shot Length (ASL) in mainstream Hollywood films of between 3 and 4 s [25].



Three types of visual stimuli in each trial

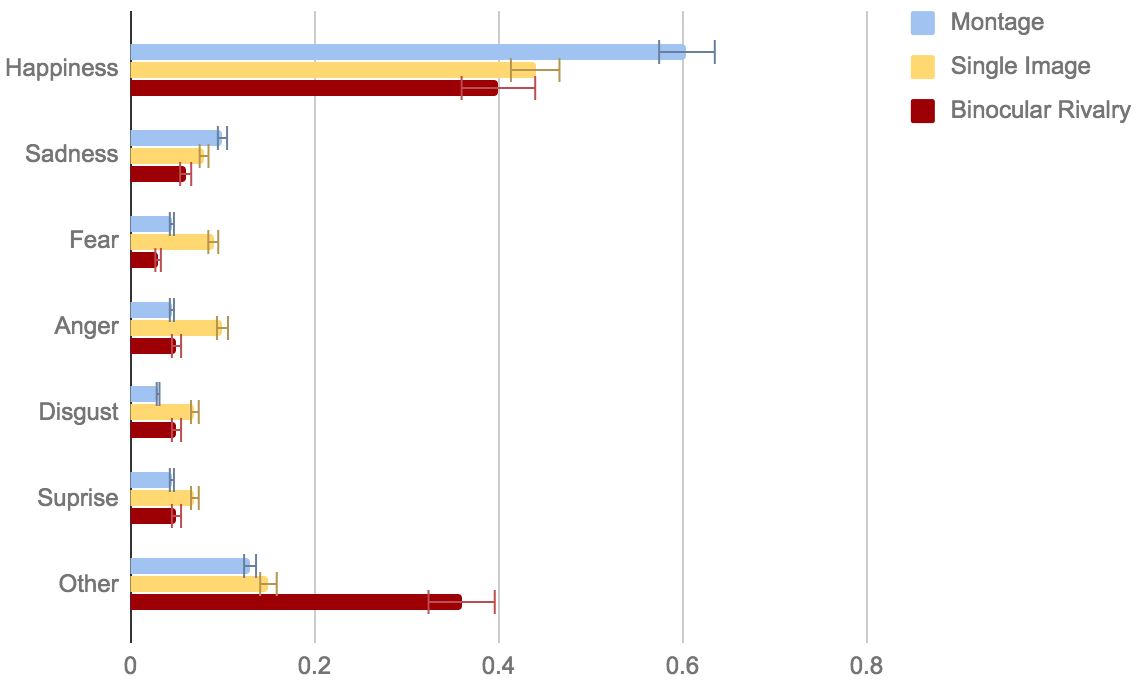
**The dichoptic presentation**was also composed of the image (a) and (b) from Figure 1, but they were presented in static. Image (a) was only presented to users’ left eye while image (b) was only presented to users’ right eye. The two images were positioned at the exact the same position, rotation, and scale in the VR environment.

**The Single Image** was the static presentation of the image (c) from Figure 1.

The experiment began with a verbal instruction followed by a written instruction shown in the VR environment when participants put the VR headset on. Participants were told that they would be presented with 3 types of visual stimulus and they needed to categorize the man’s emotion into 7 options: “happiness”, “disgust”, “sadness”, “surprise”, “fear”, “anger” and “other” from the menu below the image in VR. The previous six categories were demonstrated by researchers as the six “primary” emotions[26]. Participants were asked to choose “other” if they thought the man’s emotion in the current visual stimulus could not fall into any of the six options above (e.g., this man has a mixed emotion or his emotion cannot be recognized.) The three types of the visual stimulus were presented to participants in a random order, and participants were asked to make their choices after carefully watching the visual for at least 12 seconds (the time duration for the montage to loop twice). Participants were asked to subjectively describe their personal feelings about the stimuli after choosing the man’s emotion for all of the three visual stimuli.

# 

## Result



Relative frequency of emotion categories by different types of visual stimulus. The error bar shows the 95% confidence interval.

 According to the subjective description of participants after the user testing, 39 out of 43 participants experienced the binocular rivalry phenomenon when dichoptically viewing the puppy image and human image at the same time.  A summary of the emotion under the three stimuli chosen by the 39 participants who experienced the binocular rivalry is presented in Figure ???. The bars show how often each of the 7 options was chosen under different conditions. For all three stimuli, the participants tended to choose the target emotion (happiness) that was supposed to be created in Kuleshov Effect more frequently than other alternative options. If perceiving the dog image had no effect on people’s interpretation of the man’s emotion, each of the 7 options should be selected with similar possibilities that were supposed to be around 0.167 (1/7). But as shown in Figure ???, the option “happiness” in montage, single image, and the binocular rivalry was chosen with the 0.605, 0.443 and 0.407 frequencies respectively. The 95% confidence intervals of the percentage of happiness in the three conditions were also above the 0.167 baseline. However, the option “other” under the binocular rivalry view condition was also selected more than the average with a frequency of 0.368. On the contrary, when the participants viewed the montage and the single image, no other options were chosen significantly more often than 16.7% except happiness.

As for other worth-mentioning subjective feedback, when viewing the montage of the man and the dog in VR, most participants stated that they felt the man changed to the dog. This subjective feeling was different from that in the Kuleshov Effect experiment in 2-dimensional motion pictures where people would more likely think, like in continuity editing, that the man was looking at the dog [20]. In the single image viewing scenario, most participants could spot that it was an image composed of a man and a dog overlaying together.

# 

# Discussion & Conclusion

Before going to the discussion, a few limitations of the experiment need to be declared. Researchers have already proven that the dichoptic presentation would create the visual fatigue and the visual discomfort to the eyes [27]. However, this paper fails to discuss the influence of such fatigue and discomfort to the result of the user testing. Moreover, because the user study is designed as the single-trial between-subject experiment, it is prone to noise in the data. More trails with different images need to be introduced in the future.

Based on the results of the current experiment, bold interpretations could be made that (1) dichoptically viewing two images could actually yield to more meanings than watching the two images in isolation under the binocular rivalry scenario, and (2) it would affect people’s perception differently than viewing the same pair of images sequentially.

Interestingly, for nearly 40% of the participants, the new idea created in the binocular is similar to the thought they get when watching the same content in motion picture as a montage. There are usually two types of rationale for the Kuleshov Effect: one is that the two shots are related in the purely cognitive level (i.e., the man and the dog exist independently and are only connected in the audience’s mind)[(Eisenstein 1949)](https://www.authorea.com/users/216969/articles/287818-exploring-the-storytelling-possibility-of-binocular-rivalry-in-head-mounted-display#eisenstein1949dialectic) while others believe that the shots are actually connected spatially or/and temporally (i.e., the man is looking at the dog because of the continuity editing effect)[28]. Although Barrate et al. has proved that the Kuleshov Effect does exist in continuity editing, the subjective feedback and the statistic results of the current experiment in the montage condition show that the Kuleshov Effect could still exist on the cognitive level. Indeed, the continuity editing is hard to achieve in VR by simply swapping the content, it does not illustrate that montages cannot be used in VR and, if used, they could still create some new meanings in storytelling.

Since “other” is another option frequently chosen by participants in binocular rivalry, which did not appear in the other two viewing conditions, the following interpretations could be inferred. Combining Eisenstein’s argument – the interpretation of the montage comes from the conflict of the original thoughts derived from the shots – people often choosing “other” is probably due to the reason that no confirmed idea idea was initially created. In other words, during the binocular rivalry, people may not *read* (cognitively perceive) any meanings from either the left eye image or the right eye image even they are constantly *seeing* them both. This inference is in line with the epistemological theory that our brain is a predictive machine and people’s perception is actually a hypothesis with the 100% possibility [4]. Hohwy et al. argue that in the binocular rivalry, it is because that none of the two hypotheses made by the brain (1. I am looking at a man; 2. I am looking at a dog) could reach the 100% possibility, people are *seeing* the two images contantly switching. The binocular rivalry will stop if  the brain thinks one hypothesis is 100% possible: a perception is formed (one of the images is *read.*) However, there is another interpretation of the high percentage of “other” in the participants’ choices. As the original 6 options of people’s basic emotions are from the research based on animals that actually live in the real world, a dog-man does not exist on the earth after all. It is probably because that a completely new perception is created during the binocular rivalry, and none of the “old” emotion could describe it. Future research is needed to explain why the participants chose “other” frequently.

# References

J. Crary, “Géricault the Panorama, and Sites of Reality in the Early Nineteenth Century”, *Grey Room*, vol. 9, pp. 5–25, Oct. 2002.

[2]S. Levine and M. McLuhan, “Understanding Media: The Extensions of Man”, *American Quarterly*, vol. 16, no. 4, p. 646, 1964.

[3]F. Crick and C. Koch, “The Problem of Consciousness”, *Scientific American*, vol. 267, no. 3, pp. 152–159, Sep. 1992.

[4]J. Hohwy, A. Roepstorff, and K. Friston, “Predictive coding explains binocular rivalry: An epistemological review”, *Cognition*, vol. 108, no. 3, pp. 687–701, Sep. 2008.

[5]J. W. Brascamp, P. C. Klink, and W. J. M. Levelt, “The ‘laws’ of binocular rivalry: 50 years of Levelt’s propositions”, *Vision Research*, vol. 109, pp. 20–37, Apr. 2015.

[6]R. Blake, Y. Yang, and H. R. Wilson, “On the coexistence of stereopsis and binocular rivalry”, *Vision Research*, vol. 31, no. 7-8, pp. 1191–1203, Jan. 1991.

[7]K. Lee and S. Lee, “3D Perception Based Quality Pooling: Stereopsis Binocular Rivalry, and Binocular Suppression”, *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 533–545, Apr. 2015.

[8]R. L. GREGORY, “Seeing in Depth”, *Nature*, vol. 207, no. 4992, pp. 16–19, Jul. 1965.

[9]“Ocular Dominance and Binocular Retinal Rivalry”, in *Visual Optics and the Optical Space Sense*, Elsevier, 1962, pp. 409–417.

[10]R. Blake, D. H. Westendorf, and R. Overton, “What is Suppressed during Binocular Rivalry?”, *Perception*, vol. 9, no. 2, pp. 223–231, Apr. 1980.

[11]J. Wang, K. Zeng, and Z. Wang, “Quality prediction of asymmetrically distorted stereoscopic images from single views”, in *2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014.

[12]R. S. Laramee and C. Ware, “Rivalry and interference with a head-mounted display”, *ACM Transactions on Computer-Human Interaction*, vol. 9, no. 3, pp. 238–251, Sep. 2002.

[13]S. Feiner, B. MacIntyre, T. Höllerer, and A. Webster, “A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment”, *Personal Technologies*, vol. 1, no. 4, pp. 208–217, Dec. 1997.

[14]J. J. Ockerman and A. R. Pritchett, “Preliminary investigation of wearable computers for task guidance in aircraft inspection”, in *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*.

[15]H. R. Wilson, “Binocular contrast stereopsis, and rivalry: Toward a dynamical synthesis”, *Vision Research*, vol. 140, pp. 89–95, Nov. 2017.

[16]F. Tong, M. Meng, and R. Blake, “Neural bases of binocular rivalry”, *Trends in Cognitive Sciences*, vol. 10, no. 11, pp. 502–511, Nov. 2006.

[17]L. N.K, “On the neural mechanisms of binocular rivalry”, *Frontiers in Human Neuroscience*, vol. 2, 2008.

[18]F. Sengpiel, “The neuron doctrine of binocular rivalry”, in *The Constitution of Visual Consciousness*, John Benjamins Publishing Company, 2013, pp. 167–186.

[19]S. Eisenstein, “A dialectic approach to film form”, *Film form: Essays in film theory*, pp. 45–63, 1949.

[20]D. Barratt, A. C. Rédei, Åse Innes-Ker, and J. van de Weijer, “Does the Kuleshov Effect Really Exist? Revisiting a Classic Film Experiment on Facial Expressions and Emotional Contexts”, *Perception*, vol. 45, no. 8, pp. 847–874, Apr. 2016.

[21]I. S. Penton-Voak, J. Thomas, S. H. Gage, M. McMurran, S. McDonald, and M. R. Munafò, “Increasing Recognition of Happiness in Ambiguous Facial Expressions Reduces Anger and Aggressive Behavior”, *Psychological Science*, vol. 24, no. 5, pp. 688–697, Mar. 2013.

[22]D. Lundqvist, A. Flykt, and A. Öhman, “Karolinska Directed Emotional Faces”. American Psychological Association (APA), 1998.

[23]“Pixels”, in *Fundamentals of Digital Image Processing*, John Wiley & Sons Ltd, 2011, pp. 49–83.

[24]D. A. Leopold, D. L. Sheinberg, and N. K. Logothetis, “What is rivalling during binocular rivalry?”, *American Journal of Ophthalmology*, vol. 122, no. 2, p. 293, Aug. 1996.

[25]J. E. Cutting, K. L. Brunick, J. E. DeLong, C. Iricinschi, and A. Candan, “Quicker faster, darker: Changes in Hollywood film over 75 years”, *i-Perception*, vol. 2, no. 6, pp. 569–576, 2011.

[26]C. Darwin, *The Expression of the Emotions in Man and Animals*. Cambridge University Press, 2009.

[27]M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, “Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review”, *Journal of Imaging Science and Technology*, vol. 53, no. 3, p. 030201, 2009.

[28]C. Plantinga, “: The Reality of Illusion: An Ecological Approach to Cognitive Film Theory . Joseph D. Anderson.”, *Film Quarterly*, vol. 50, no. 4, pp. 64–65, Jul. 1997.