Effects of Size and Motion Parallax on Perceived Depth in Two-Dimensional Virtual Reality Displays

Citation
Park, Jungyeon. 2020. Effects of Size and Motion Parallax on Perceived Depth in Two-Dimensional Virtual Reality Displays. Bachelor's thesis, Harvard College.

Permanent link
https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37364695

Terms of Use

Share Your Story

Effects of Size and Motion Parallax on Perceived

Depth in Two-Dimensional Virtual Reality Displays

A thesis presented

by

Jungyeon Park

to

The Department of Computer Science

in partial fulfillment of the requirements

for a degree of

Bachelor of Arts with Honors

in the field of

Computer Science

Harvard University

Cambridge, Massachusetts

March 30, 2020

Statement of Research

This research was conducted in the Engert Lab in the Department of Molecular and Cellular Biology at Harvard University, under the direct supervision of Professor Florian Engert. Work in the Engert Lab started January 2019, and continued alterations to the project throughout May 2019 resulted in the writing of this thesis. Research during the summer of 2019 was supported by Harvard College Research Program (HCRP), and term-time work was credited through CS91r and CPB99A/B.

This project was conceived by Professor Engert and jointly designed by him and me. The idea of the computer programmed virtual image was derived from Johnny Lee's work on Wii Remote Virtual Reality (Lee 2008). All experiments were programmed by myself, and 6 of the 9 total subjects participated in the experiment under my guidance. The remaining 3 were completed remotely under the guidance of Tyler Griggs, to whom I communicated the description of the experiment. All figures were created by myself, unless otherwise noted.

Effects of Size and Motion Parallax on Perceived Depth in Two-Dimensional Virtual Reality Displays

Jungyeon Park

Professor Florian Engert, Thesis Advisor

Abstract

Depth perception refers to the ability to perceive distance between objects, and the visual features from which such conclusions are derived are called depth cues. Depth is also the primary distinguishing factor between two-dimensional images or videos and reality. Virtual reality, with the goal of creating an illusion of reality, employs depth perception in many of its images through depth cues. In particular, binocular depth cues are frequently used due to the vividness they portray. However, technology to create or display binocularly disparate images are inaccessible and hard to develop. Here, we program a two-dimensional virtual reality display to compare the depth portrayals of two powerful monocular cues, size and motion parallax. Results indicate that the effect of size cue dominates that of motion parallax. Furthermore, results show that a 10-second decision about depth based on intuition is more accurate than a longer deliberation based on calculations and logic, particularly when faced with a complex display.

Acknowledgments

I would like to thank first and foremost Florian Engert, my thesis advisor, mentor, the Principal Investigator of my lab, and the most cheerful professor I've ever met. Without his help and guidance, this thesis would not have been possible. Thank you for all your mentorship, as well as for all the very enjoyable conversations we have had in Biolabs, whether about the thesis or other issues.

I would also like to thank my computer science professor and thesis advisor, Steven Gortler, for his help and guidance. Particularly in developing the program to perform the experiments, his classes, textbook, and office hours were essential.

Through 3 years of being a Chemical and Physical Biology concentrator, Professor Dominic Mao has gone above and beyond his role as the concentration advisor and has become a very valuable mentor for me. Thank you for believing in what I can do and reminding me what is most important in learning. I would also like to thank Irina Cashen for being so understanding and helpful with all the coordination for MCB-CPB events.

Through 2 years of being a Computer Science concentrator, I have learned from many professors, but I would like to particularly thank Professor Boaz Barak for his teaching and mentorship. Learning from him in the classroom, as well as from conversations outside of it, has been extremely valuable and shaping as I started to navigate the world of Computer Science later than most of my peers.

Last but not least, I would like to thank my friends who volunteered as participants and/ or offered me a place and companionship as I struggled to find housing as an international student after leaving Harvard abruptly: Tyler Griggs, Luke van Gog, Francesca Hess, Micah Meekins, Mariana Kessler, and April Kessler.

Contents

List of Figures

Chapter 1

Introduction

1.1 Virtual Reality and Vision

Will virtual reality ever so closely resemble our physical world that it becomes indistinguishable? Has that day already come? A 1992 definition of virtual reality (VR) that remains relevant to this day describes VR as a real or simulated environment in which a perceiver experiences presence in an environment by means of a communication medium (Steuer 1992). Under this definition, handwritten letters and phone calls are all modes of VR, along with the more recent technologies of 3D films and head-mounted displays (HMD). The primary difference between the former and latter modes of VR lies in the level of vividness and interactivity that can be mediated through the communication medium. The latter modes of VR, which are the technology we commonly think of when referring to VR, create illusions for our senses to vividly experience a virtually-created 3D world.

The history of modern virtual reality technologies date back to early 1960s, starting with an invention by Morton Heilig. His invention entitled "Sensorama Simulator" was a VR video arcade that provided a multimodal experience across vision, sound, smell, and touch. Later VR inventions included HMDs developed by NASA called "VIVED", which

uses liquid crystal display (LCD) that can be worn on the head, the first sensing gloves called "Data Gloves", and a 3D audio system invented by Fisher and Elizabeth Wenzel (Burdea & Coiffet 2003).

As can be seen by the above inventions, approaches to create illusions of all five senses (visual, tactile, auditory, gustatory, and olfactory) have existed for some time. However, as vision dominates the human sensory perception, vision has been and continues to be the primary target for sensory illusion to enhance vividness (Peli 1995). Compared to other sensory illusions, visual illusions' potential for highly vivid three-dimensional environments has led to further investigation and development of vision-centered VR technologies.

Early vision-centered virtual reality media such as paintings and photographs allow the viewer to experience the presence of the portrayed entities without their physical presence. These media were further developed into videos, through which a large number of images can be shown every second, allowing for an illusion of a dynamic, continuous environment in lieu of many static images. Most recently, technology has added the perception of depth to videos by employing various depth cues (further discussed in the next section) in 3D Computer-Generated Imagery (CGI). Such depth cues create two-dimensional images that add the illusion of third dimensionality to VR displays.

1.2   Depth Cues

A visual feature, such as size or linear perspective, that relays information about depth is known as a depth cue (Howard 2002). In 3D CGI, depth is portrayed to the viewer by the employment of several depth cues that interact with each other. When information is available from two or more cues, they are often combined to enhance the perception

of depth or resolve an ambiguity of a cue. When the information from depth cues are conflicting, a conclusion is drawn on depth by weighting the cues or ignoring one (Howard 2002).

Depth cues can be organized into monocular cues and binocular cues. As the name suggests, monocular cues are depth cues that require just one eye for information to be delivered, whereas binocular cues require both eyes. Within each category, there are several types of depth cues. A categorization of 21 depth cues by Howard can be seen in Figure 1.1 (Howard 2002).

Monocular cues can be divided into static and dynamic cues, depending on whether they require motion of the viewer or the visual stimuli for the depth cue to function. A few well-known and intuitively comprehensible monocular static cues are linear perspective, image size, and occlusion. An illustration of these cues are explained in Figure 1.2 using the famous album photo of Beatles's Abbey Road as example.

Whereas dynamic cues cannot be demonstrated with images on a paper, motion parallax can be easily enacted in reality. Pick up the screen or paper on which this thesis is being read so that there is another object behind the thesis that is also within your visual field. Then, with only one eye open, move your head left and then right, while taking note of both the thesis and the chosen object farther away. You will notice that the object farther away from you moves left to right *less* than the thesis which is closer to you. This is a demonstration of motion parallax, which is defined as "relative motion of the images of object points at different distances that is caused by motion of the observer or of the object point" (Howard 2002).

Motion parallax is capable of conveying very detailed information about depth to the viewers. In fact, it has been shown that motion parallax alone can "produce a reliable,

Monocular

Static

Dynamic

Optic flow

Motion parallax

Accretion deletion

Binocular

Convergence

Static

Changing

Binocular disparity

Position disparity

Orientation disparity

Temporal disparity

Monocular occlusion

Perspective

Linear

Image size

Texture gradient

Height in field

Overlap

Interposition

Transparency

Lighting

Shading

Shadow

Aerial effects

Optical haze

Haze

Focusing

Image blue

Accommodation

Figure 1.1: "Sources of information for the perception of distance and relative depth" from Seeing in Depth, Volume I (Howard 2002)

This diagram depicts a tree categorizing known depth cues. The biggest categories are monocular and binocular cues, which are determined by the number of eyes required to utilize the depth cue. Then, they are further divided into subcategories such as static cues and dynamic cues, which are determined by whether motion is necessary to perceive the depth cue. As can be seen in the diagram, there are more known monocular depth cues than binocular ones.

Figure 1.2: Illustration of Static Depth Cues through Abbey Road Album Photo

Through this image of the Beatles, monocular cues such as linear perspective, image size, and occlusion can be illustrated. Linear perspective is the idea that parallel lines converge in the far distance. This is visible from the cement road that converges into a point behind the members of the Beatles.

Image size portrays depth as closer objects appear larger than distant objects. This is visible in the image as the size of John Lennon is larger than the size of cars. This indicates that John Lennon is closer to the viewer than the cars.

Lastly, occlusion is a monocular static depth cue that illustrates depth when two objects are overlapping. The object visible on top of the other object at the site of overlap is the closer one. In the image, the buildings in the upper right corner are hidden by the trees. This indicates that the trees are closer to the viewer than the buildings.

consistent, and unambiguous impression of relative depth in the absence of all other cues to depth and distance" (Rogers & Graham 1979). Beyond relative depth, accurate estimations of absolute depth through motion parallax has shown to be achievable through training (Ferris 1972).

Within the set of binocular cues, the two primary types are convergence and binocular disparity. Convergence refers to the amount of eye rotation needed to focus on the object, and binocular disparity refers to the "difference arising from viewing a scene from two vantage points" (Howard & Rogers 2002). In other words, binocular cues tell distance by focusing on the differences of the images each of our eyes perceives due to their separate location.

Traditionally, binocular cues have been considered to be more powerful as depth indicators than monocular cues (Qian 1997). Furthermore, binocular cues dominate much of depth perception research, as can be seen by the fact that half of the chapters in the classic depth perception literature "Seeing in Depth Volume 2: Depth Perception" is about binocular disparity, whereas the other half of the volume is divided among the rest of the depth cues (Howard & Rogers 2002).

For this reason, much of the depth portrayal in advanced VR displays, such as HMDs or 3D films, focus on binocular displays utilizing binocular disparity. However, there are several problems with limiting binocular disparity as the primary depth cue for VR displays.

First, there is significant binocular stress induced by binocular displays. When individuals use a commercially available HMD that suits the average interpupillary distance of the subject population for 10 minutes, there were clear signs of induced binocular stress (Mon-Williams et al. 1993). Another experiment in which subjects viewed a display that required constant ocular focus with changing vergence eye movements for 10 minutes were

sufficient to cause deficits of binocular vision (Mon-Williams & Wann 1998).

The second problem is the difficulty of creating CGI containing distinct 2D images for each eye. Unlike taking photographs, constructing a video that portrays depth with binocular disparity requires special technology and time-consuming expertise. This difficulty of creating 2D images with binocular disparity leads to inaccessibility of 3D CGI.

Lastly, devices required to perceive binocular disparity on a 2D display, such as HMDs or polarized glasses for 3D films, are not universally accessible and are often uncomfortable to wear. This limitation restricts the accessibility and quality of VR experiences.

1.3 Aims and Hypothesis

To address these issues, this study focused on investigating depth cues alternative to binocular disparity that can be utilized in 3D CGI. This study aimed to find a monocular depth cue that provides similarly accurate distance information while avoiding the aforementioned issues that arise from primary depth cues requiring binocular accommodations.

Previous research indicates that binocular disparity is most relevant in depth perception within an arm's length (Haase & Wermann 1985) and not crucial for distances farther than 40m (Schmidt 1966). For objects at farther distances, monocular cues such as occlusion, relative size, motion parallax, and linear perspective influence depth more predominantly (Bauer et al. 2001).

This is demonstrated by studies involving activities that require accurate depth judgments from a distance, such as flying a plane or driving a car. Past research shows that pilots are able to land a flight monocularly as accurately as they do binocularly (Grosslight

et al. 1978). Further, research on truck drivers revealed that monocular drivers are no worse than binocular drivers in terms of most day-to-day driving functions (McKnight et al. 1991).

Given the overlap in depth information that certain monocular depth cues and binocular disparity are able to portray, this study aims to discover the monocular cues capable of creating as powerful depth illusions. The aims and hypotheses of this study are as follows:

Aim 1 Create a computer-generated 2D display that represents depth only using monocular cues, whose average depth can be altered by the test subjects.

Aim 2 Compare the accuracy of depth information portrayed by each monocular depth cue, and determine most effective monocular cue.

Hypothesis 1 Motion parallax is a key monocular depth cue that will provide most accurate depth information.

Hypothesis 2 When motion parallax is not present, size will function as the most powerful depth cue.

Chapter 2

Methods

In this study, a 2D computer generated image is created on the computer screen. Subjects were asked to complete a task that requires precise depth perception, and each set of trials involved a different set of depth cues so that task performance represented the amount of information each depth cue provides.

2.1   Participants and Apparatus

9 subjects participated in the experiments (mean age 32.5, SD 16.2, 7 female). All participants reported normal to corrected-to-normal vision and no color blindness. Participants were asked to complete the task to the best of their ability while relying on their intuition. Participants were given a total of 96 trials, each of which had to be completed during the 10 second time limit.

Before completing the above experiment, 3 of the 9 subjects had previously completed a second experiment. In this second experiment, the three subjects were given identical

trials with the same task, but without a time limit. Subjects were encouraged to take as much or as little time as they needed to complete each task.

Experiments were presented on a display of 13.3 inch screen with 2560 x 1600 resolution. Stimulus presentation and response recording were controlled by Python 2.7.16 using Psychopy 3.2.4 (Peirce 2007). Psychopy was used for rendering the display, as it is the most widely used Python library used for psychophysics experiments. All subjects started the experiments with their head location 50cm away from the screen.

The general design of the two-dimensional virtual reality display was derived from Lee's Wii-Remote Virtual Reality (Lee 2008). A video recording of his work can be viewed at tinyurl.com/lee-display, and the repository for my work can be found at github.com/falltwenty/Thesis.

A sprite-graphic-like rendering of two-dimensional objects was used in this study in lieu of shaded cube-like objects or perspective projections to eliminate confounding variables and narrow the depth cues involved in the perception tasks.

2.2 Experiment Overview

The study consisted of a total of 96 trials for each participant. 48 trials were static experiments, and the remaining 48 were dynamic experiments. Static experiments were trials in which motion parallax is not present as a depth cue. Dynamic experiments are trials in which motion parallax is available through head tracking of the participants.

The 48 trials for each static and dynamic experiments were composed of 4 different dart location settings and 2 different dart size settings. For each of the 8 setting combinations, there were 6 trials (Figure 2.1). The displays shown in the 24 trials under each size and

motion parallax setting were pre-configured and identical across subjects. The 24 trials were chosen for displays without darts on or too close to the display borders, from a pool of randomly generated configurations.

In each trial, the screen showed two displays side-by-side. Each display showed 4 darts and 4 white lines connecting the center of each dart to the center of the display on a black background (Figure 2.2). Each display resembles a window through which four identical darts *behind* the screen can be seen. Although the radii of the actual darts are the same, because the darts are positioned at different depths, their radii *on* the screen vary.

During the experiment, subjects can control the distance of the darts using the keyboard. The up keystroke pushes back all 4 darts of the left display by 1 unit, and the up keystroke pulls all 4 darts toward the screen by 1 unit. For the sake of expedience, the keystroke "R" pushes back all 4 darts of the left display by 5 units, and "F" pulls forward by 5 units. The left and right keystrokes work identically for darts on the right display, respectively pushing and pulling the darts by 1cm.

The task for the subjects was identical for all 96 trials: match the average depth of the left display with that of the right display. To avoid the training effect confounding results, subjects were given sufficient explanation and opportunity to practice before starting the experiment. They were encouraged to primarily alter the left display and to only alter the right display to calibrate their perception of depth. The pair of displays at the 10-second mark was saved as the subject's response. In the second experiment in which trials were untimed, participants pressed the key "N" to move onto the next trial when they felt that the task was completed. All participants continued until 96 trials were completed.
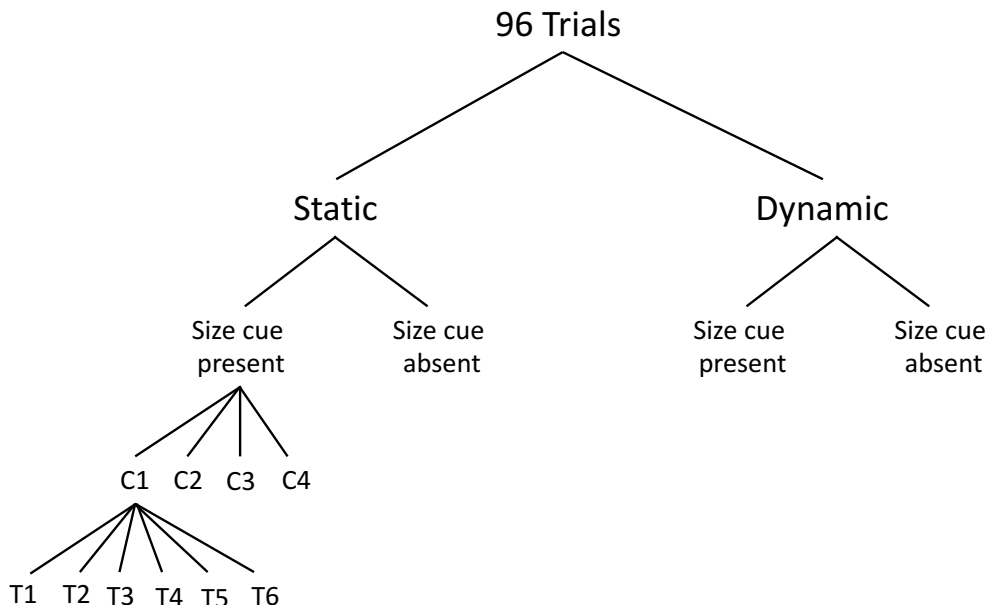
Figure 2.1: Abridged Diagram of Experiment Trials

This abridged tree diagram represents the setup of the experiment trials. The tree is symmetrical, although parts of it have been abbreviated to avoid clutter in the above image. T1 in the diagram refers to trial 1, and C1 refers to configuration 1.

Of the 96 total trials, 48 are static, and the other 48 are dynamic. For each static or dynamic experiment, image size is available as a depth cue in 24 trials but not in the other 24. In each of the 24 trials, there are 4 configurations of dart locations. For each configuration, there are 6 separate trials. A description of each configuration is available in Section 2.5.
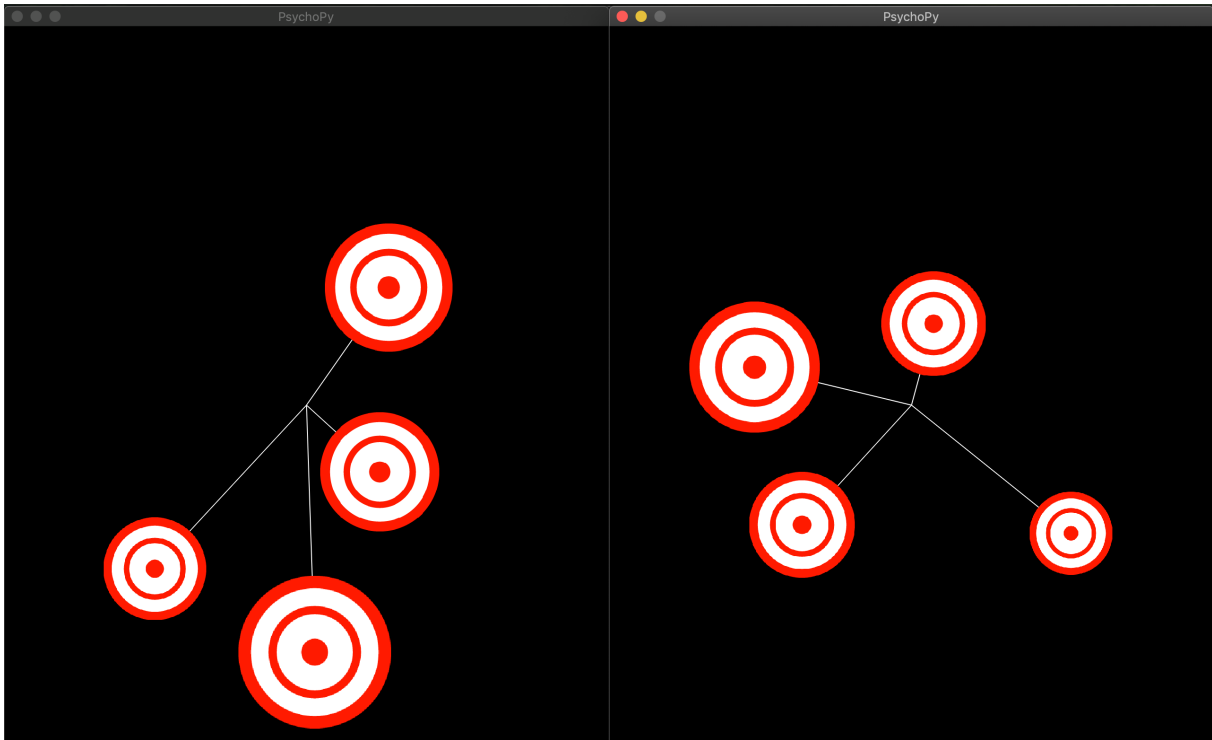
Figure 2.2: Visual Display of a Trial

This is a screenshot of one of the 96 trials. Each trial has two displays side-by-side, and each display has a black background, four red-and-white darts, and four white perspective lines connecting the dart centers to the display center. Each display serves as a window to view the four darts that are some distance behind the screen on which the experiment is conducted. The location and size of the darts of the left and right display vary every trial.

2.3   Static Experiment

In static experiments, participants were given a pair of visual displays in each trial, similar to Figure 2.2. The task was to increase or decrease the distance behind the screen for the darts on the left display to match the average distance of the darts on the right display.

The 4 depth cues that were available for perception in the static experiment and changed as z was altered by the participants are: height in field cue from the location of darts, image size cue from the dart size, occlusion cue from the overlapping of lines or darts, and perspective cue from the white lines.

When distance was increased, the darts moved incrementally closer to the horizon center point, the dart radius became smaller, the area of the occluded overlap changed accordingly, and the perspective lines shortened as the darts moved closer to the horizon. On the other hand when distance was decreased, darts moved away from the center point, darts grew in size, the occluded area changed to accommodate the change in location and size of the darts, and the perspective lines lengthened.

2.3.1   Dart Location

In this subsection, we discuss how dart locations were determined behind and on the screen.

A coordinate system C is defined for ease of communication about where darts are located. The origin of C is the center of the computer screen. The x axis is the horizontal line parallel to both the keyboard and the participant. The positive x axis points towards the right side of the viewer. The y axis is the vertical line perpendicular to the keyboard, and the positive y axis points towards the ceiling. The z axis is the horizontal line parallel to the keyboard and perpendicular to the participant. Positive z axis points towards the

back of the screen.

Given this coordinate system, location of the dart on the screen is determined using the 3-dimensional coordinate of the dart's center behind the screen (x_0, y_0, z_0). This coordinate represents the physical location of the dart should it actually exist behind the monitor. Then, considering the viewing distance and angle of the participant, the 3D coordinates are translated onto 2D coordinates on the screen, $(x_s, y_s, 0)$. This coordinate represents the location of the darts on the plane $z = 0$, should the screen function as a window through which the darts behind can be viewed. Figure 2.3 shows a bird's-eye view diagram of a subject viewing a dart on the screen (black dot) that is a representation of a "real" dart behind the screen (blue dot). The geometric calculations translating the 3D coordinate into the 2D coordinate is described in Figure 2.3.

From the diagram depicted in Figure 2.3, one can notice that the farther the blue dot is located (large z_0), the closer the black dot will be to the center of the screen (smaller x_s). This viewing experience is representative of the depth cue of height in field. Height in field is a monocular cue that represents distance of an object by its height relative to the whole visual field. In visual fields like Figure 2.2 where the horizon is visible, the closer an object is to the horizon, the farther away the object appears to be.

2.3.2 Dart Size

In this subsection, we discuss how dart sizes were determined behind and on the screen.

Dart size behind the screen is identical for all darts in all trials. Then for each trial, dart size on the screen is determined using the distance between the dart and the screen (z_0 value), as well as the "real" radius of the dart behind the screen, referred to in this paper
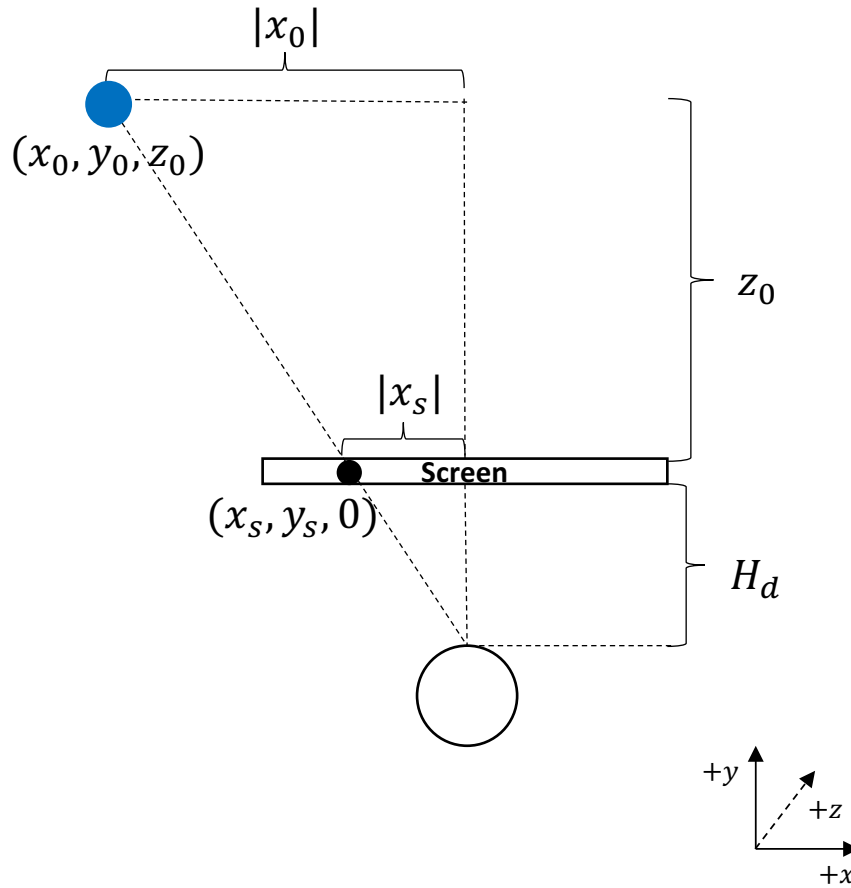
Figure 2.3: Subject's viewing angle of the stimuli from a bird's-eye view

The blue dot illustrates the "real" dart located z_0 units behind the screen, and the black dot illustrates the representation of the blue dot on the screen. The dart centered on coordinate (x_0, y_0, z_0) is centered on $(x_s, y_s, 0)$ on the screen, where x_s and y_s are determined using the ratios $x_0 : x_s = (z_0 + H_d) : H_d$ and $y_0 : y_s = (z_0 + H_d) : H_d$

as the absolute radius. For instance, in Figure 2.3, the blue dot is represented by a smaller black dot on the screen.

Just as x_s and y_s are determined by the original center coordinate as well as head distance, on-screen radius of the dart (r_s) is determined by the absolute radius (r_{abs}) and the head distance. Similarly to the ratio provided in Figure 2.3, $r_{abs} : r_s = (z_0 + H_d) : H_d$.

2.3.3 Occlusion

Occlusion happens within the visual field when two or more darts or lines occupy the same coordinates on the screen. Then, the dart or line with a smaller z_0 value, the one closer to the screen, is shown on the screen whereas the other one remains undisplayed for the coordinates that are overlapped. Occlusion provides non-ambiguous information about relative depth between objects, as the object shown is always closer than the object hidden. However, occlusion is unable to provide any information on the actual distance of the object from the viewer, also called absolute depth (Palou & Salembier 2011).

2.3.4 Linear Perspective

The monocular depth cue of linear perspective is created using the white lines that connect the center of each dart with the center of the display. All of the drawn lines converge on the center point, giving the illusion that the center is the horizon. This enhances the perception that the darts are located at different depths behind the screen, connected to a single point much farther in the distance.

2.4   Dynamic Experiment

In dynamic experiments, motion parallax is available as a depth cue in addition to the depth cues available in the static experiments. As the viewer moves, the relative velocities of objects at different distances portray depth information. In this experiment, movement of the viewer is tracked using the internal webcam of the computer on which this experiment was run. The webcam lies on the same plane as the screen. Using OpenCV, the rectangle that outlines the face of the subject is found and tracked (Bradski & Kaehler 2008). We call this rectangle the face box.

The dimensions of the face box represents the distance of the subject from the screen. At the initial distance of 50cm, the face box dimensions are remembered as the initial width and height. When the subject moves her head towards or away from the screen, the face box dimensions change. By calculating the product between 50cm and the ratio $\frac{newwidth}{initialwidth}$, we are able to find the distance of the subject from the screen.

The coordinates of the center of the face box represents the left, right, up, and down movement of the subject parallel to the screen. The distance of the face box's movement side-to-side or up-and-down is calculated by tracking the change of the face box center coordinates. The calculation also takes into account the distance between the subject and the screen. Figure 2.4 shows the diagram for the geometry of this calculation.

2.5   Size Settings and Configurations

For each static and dynamic experiment, there are 8 settings, created from combination of 2 size settings and 4 dart configurations. The two size settings are presence or absence of size
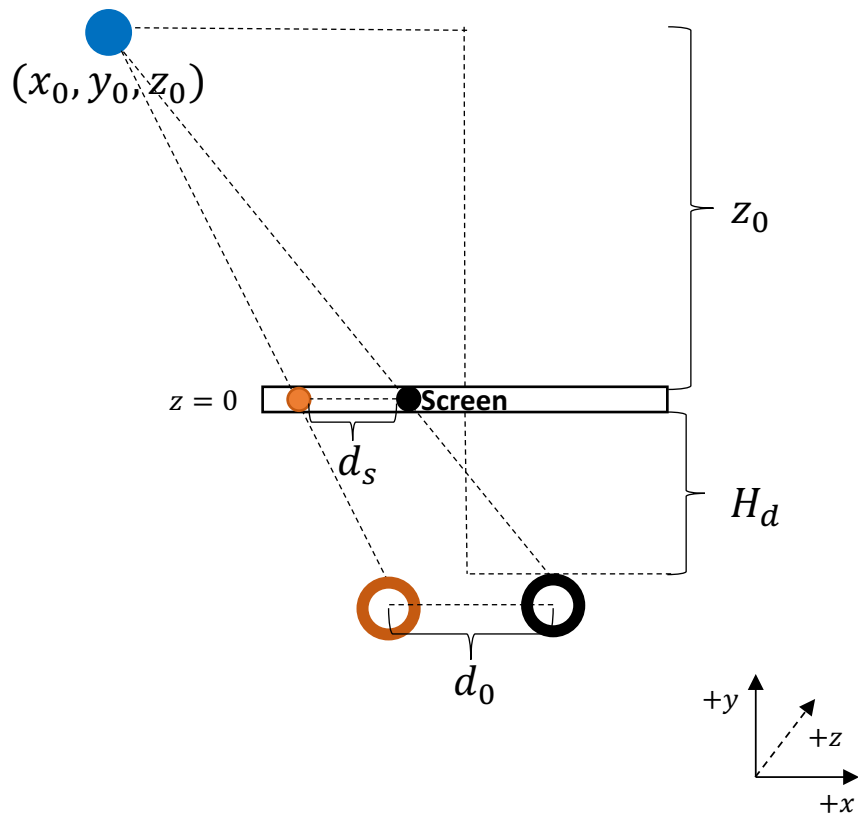
Figure 2.4: Subject movement and the corresponding visual stimuli location change from a bird's-eye view

The blue dot illustrates the "real" dart located z_0 units behind the screen, and the black dot illustrates the representation of the blue dot on the screen to the black individual. When the black individual moves to the left d_0 units in the position of the orange individual, the blue dot represented on the screen also shifts to the left. The resulting new location of the dart representation at $z = 0$ is the orange dot. d_s is determined using the ratios $d_0 : d_s = (z_0 + H_d) : z_0$. The same calculation applies to movement to the right, up, or down parallel to the screen.

cues. When size cue is present, the size of the darts change to reflect the z value of darts altered by participants using the keyboard. When size cue is absent, the size of the darts remain the same while all other aspects of the visual field, such as length of perspective lines and degree of motion parallax, accurately reflect the altered dart depth values.

The four dart configurations distinguish tasks according to their complexity and corresponding difficulty. For ease of communication, let's call the list of x-coordinates (x_0 not x_s) of the 4 darts in the left display as $X_L = [x_{l1}, x_{l2}, x_{l3}, x_{l4}]$ and those of the right as $X_R = [x_{r1}, x_{r2}, x_{r3}, x_{r4}]$. Y_L, Y_R, Z_L, and Z_R are defined in the same way. The first configuration is where $X_L = X_R$ and $Y_L = Y_R$. $Z_L \neq Z_R$, but $z_{l1} = z_{l2} = z_{l3} = z_{l4}$ and $z_{r1} = z_{r2} = z_{r3} = z_{r4}$. (Figure 2.5(a))

In the second configuration, $X_L = X_R$, $Y_L = Y_R$, and $Z_L = Z_R + \triangle z$, where $\triangle z$ is a constant that is the same for all four darts but unique for each trial. (Figure 2.5(b)) In the third configuration, $X_L = X_R$, $Y_L = Y_R$, but $Z_L \neq Z_R$. Also, each of the list elements in Z_L or Z_R are chosen independently. (Figure 2.5(c)). In the fourth configuration, all $x, y,$ and z coordinates of both the left and right display are chosen independently of each other. (Figure 2.5(d))
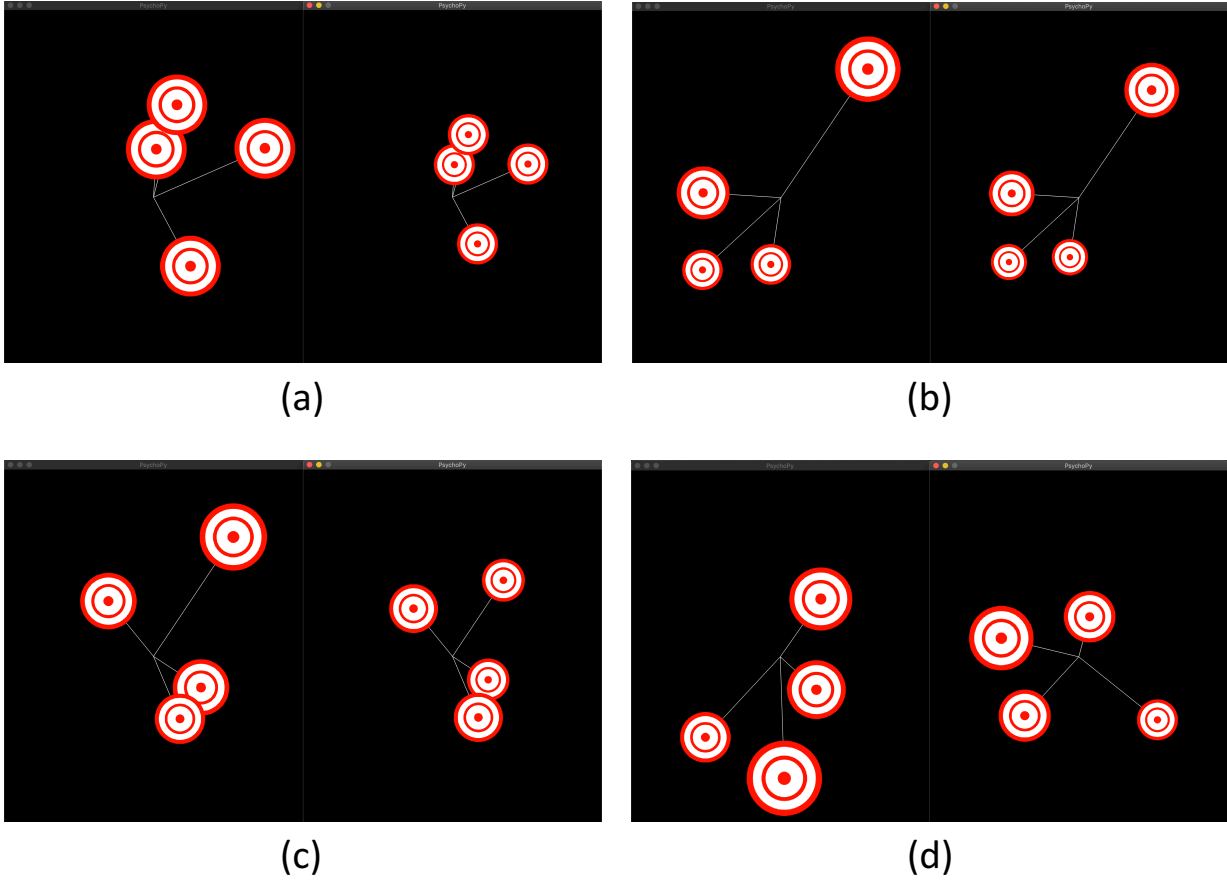
Figure 2.5: Example trials for each configuration

An illustration of a trial in configuration 1 is (a); 2 is (b); 3 is (c); 4 is (d).

Chapter 3

Results

3.1 Depth Cue Comparison

Four monocular depth cues were present in the experiments: image size, motion parallax, linear perspective, and occlusion. Linear perspective and occlusion were present in all trials, whereas image size was present in half of the trials and absent in the other half. Motion parallax also had the same distribution of trials.

The effectiveness of depth cues in depth perception was quantified using percentage error. Percentage error refers to the amount of error relative to the depth of the display. In other words, if the correct display depth was D_c and the subject reported depth was D_r, percentage error would equal $\frac{|D_c - D_r|}{D_c}$. The ratio is used in lieu of the absolute discrepancy $|D_c - D_r|$ to account for the narrow changes that occur at far distances.

Figure 3.1 represents the percentage errors of trials involving depth cues of image size against those with motion parallax. The blue bars represent data from trials in which the respective cue was present, and orange bars represent data from trials with the respective
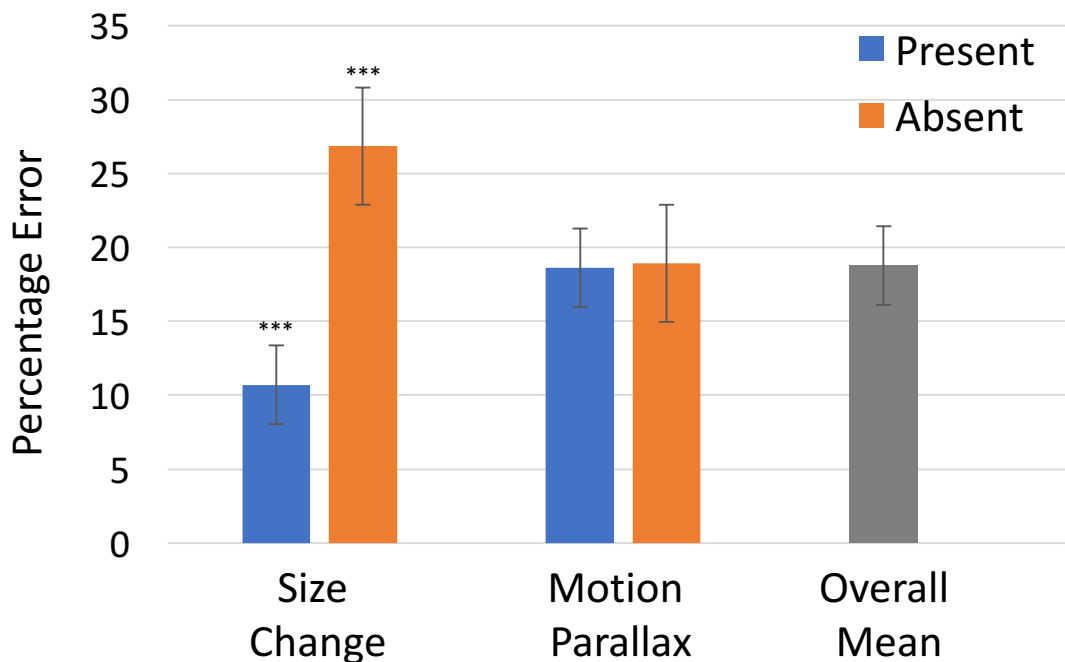
Figure 3.1: Impact of monocular cues on depth perception accuracy

The first pair of graphs represent the mean percentage depth error in trials where the size cue is present or absent. The blue represents the trials in which the cue is present, and orange represents those in which the cue is absent. The second pair in the graph represents the percentage error in trials with or without motion parallax as a cue.

A t-test for two-samples assuming equal variances was conducted to verify the statistical significance of the results. For the comparison between the overall mean and when size cue is present, the p-value is 7.16E-14. For when size cue is absent, p-value is 1.50E-10.

cue absent. The gray bar on the far right represent the mean of percentage error over all trials (18.8). Compared to the overall mean, one can see that the presence or absence of motion parallax does not alter the accuracy of depth perception trials.

However, with regards to the cue of size change, the difference in accuracy between trials with or without it is very visible and statistically significant. During trials in which the size cue was present, participants decreased their percentage error by 8.06 percentage points. On the other hand, during trials without the size cue, the error percentage increased by 8.08 percentage points. This is a 43% change in accuracy relative to the overall mean and clear evidence that size cue influenced the depth perception task at hand far more than motion parallax did.

Next, the percent errors for the 4 sets of trials in which size change and motion parallax were present or absent were analyzed (Figure 3.2). The first two bars respectively represent the mean percent errors of trials in which both SC and MP cues were present and when only SC was present. The latter two bars respectively represent the mean percent errors of trials in which only MP was present and when neither were present. The trials with SC present reported higher accuracy regardless of the presence of MP, and trials with SC absent reported lower accuracy regardless of the presence of MP.

To further understand the accuracy differences arising from different sets of trials, percent errors for the 4 groups of trials were analyzed for each configuration (Recall Section 2.5's desecription of each configuration). In figure 3.3, the trends of the four bars in each of the four configurations generally follow that of Figure 3.2. One noticeable deviation of the trend is the nearly statistically significant decrease in percent error (increase in accuracy) between trials with only SC and trials with both SC and MP are both present (p-value = 0.055). The percent error for the first set of trials is 10.2, whereas that for the second set is
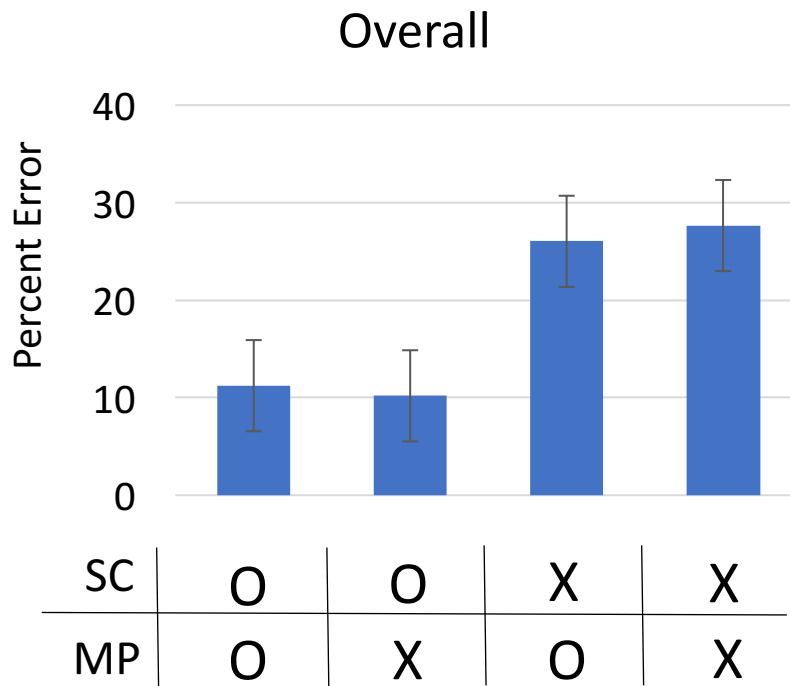
Figure 3.2: Percent error of depth perception by configuration and depth cues

Each of the bars in this graph depict percent errors of trials with a particular set of depth cues. SC stands for size change, and MP is motion parallax. The first bar represents percent error of trials where both cues are present. The second bar represents percent of trials where only size cue is present, and the third bar is that of those where only motion parallax is present as a cue. The last bar is where neither are present. Error bars represent one standard deviation.
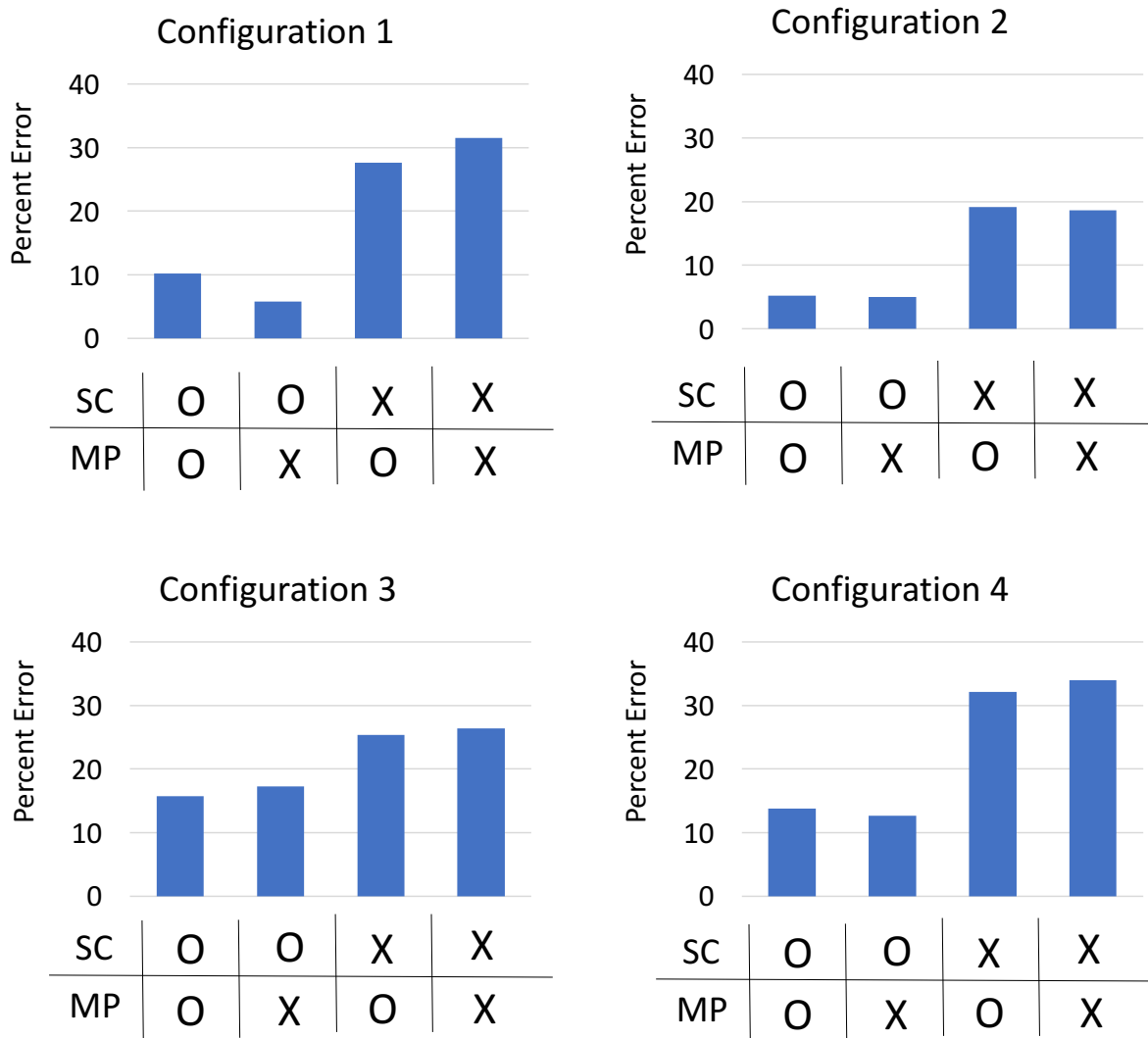
Figure 3.3: Percent error of depth perception by configuration and depth cues

SC is short for size change, and MP is short for motion parallax. Each of the four graphs represent data from trials in the specified configuration. As in Figure 3.2, each bar in the graph represents the percent error for trials completed using a specific set of depth cues.

5.8.

Another noticeable trend is that for all four configurations, the third trial with only MP cue always performed equally or slightly better than the fourth trial, in which both depth cues were absent. Overall, configuration 2 has the smallest mean percent error for all trials among the four configurations (12.0).

3.2 Task Inaccuracy in Longer Trials

3 of the 9 participants additionally completed a second set of experiments. The experimental setup of the second experiment, such as the task and configurations, was identical, but participants were given no time restrictions and were encouraged to take as much or as little time as needed for each trial. Excluding self-imposed breaks, participants took on average 48 minutes to complete the experiment. Compared to the 19 minutes that a timed experiment with the same number of trials requires, the trial duration increased by 153%.

Figure 3.4 shows the change in percent error for each configuration that occurred between trials without time restriction and timed trials. Y-axis represents the (percent error without time restriction) - (percent error with time restriction). For configuration 1, there is a slight decrease in error for trials without time restriction. However, for configurations 2, 3, and 4, depth percent error was greater when participants were given more time to complete the same task, surprisingly. This increase in error without time restriction is particularly noticeable in configuration 4.

Lastly, Figure 3.5 shows the change in percent error in the second experiment, with respect to each configuration and size cue presence settings. A very noticeable element of this graph is the magnitude of the blue bar graph. The blue bar graph represents the
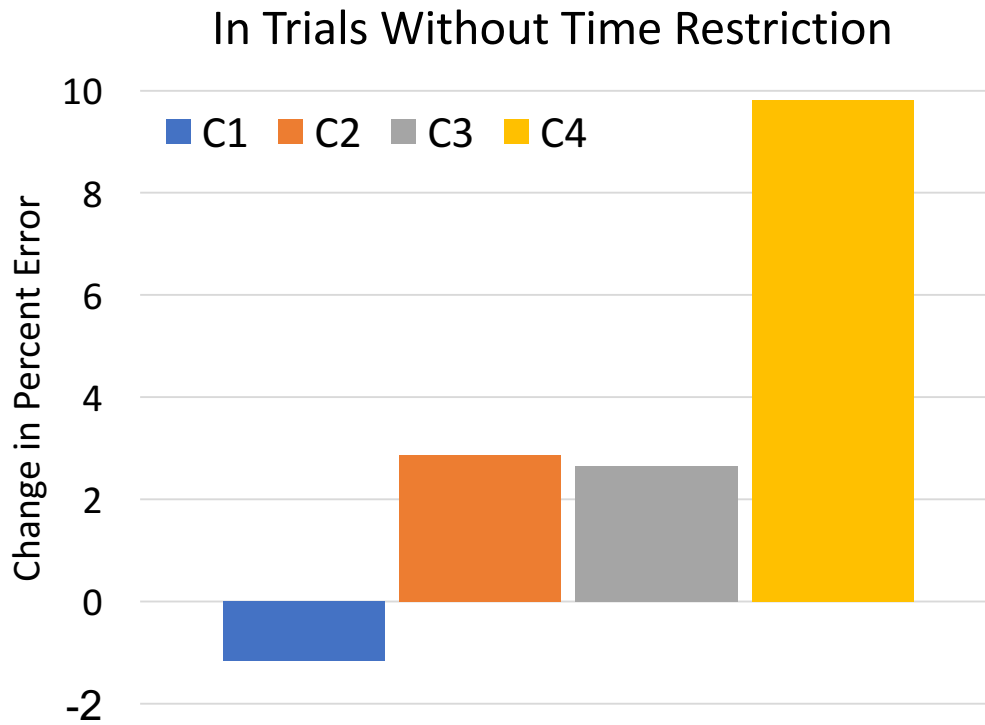
In Trials Without Time Restriction

Figure 3.4: Inaccuracy of depth perception in untimed trials

Change in percent error refers to the change in the second experiment relative to the first experiment. In other words, y-axis represents (percent error in untimed trials) - (percent error in timed trials). For configuration 2, 3, and 4, the positive values represent increased error during the untimed trials, relative to the 10-second quick trials. Configuration 1 shows a slight increase in accuracy in the untimed trials.
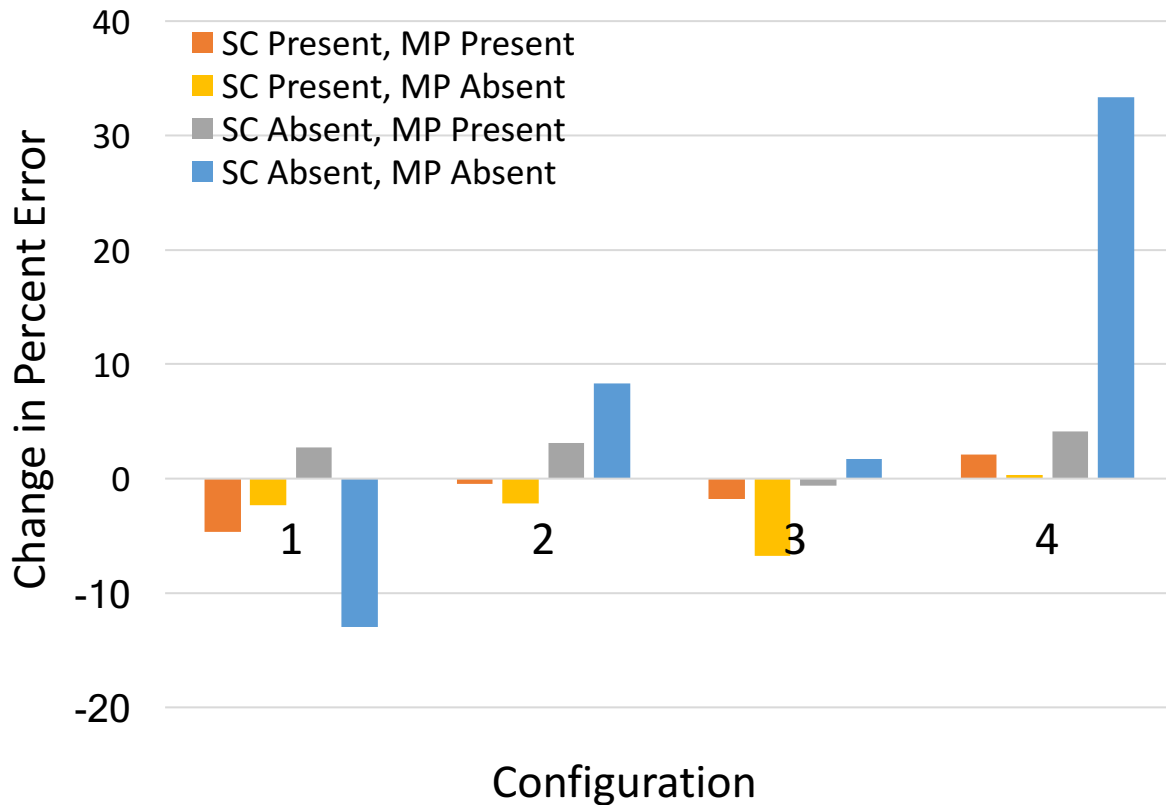
Figure 3.5: Distribution of the change in accuracy in untimed trials

The graph above represent the change in accuracy when the trials are completed untimed, relative to their previous accuracy as a 10-second trial. The four clustered bars show data for each configuration, and the four colors show data for the depth cue settings. The blue bar represents change in percent error for all trials without SC and MP in the particular configuration.

change in percent error for trials in which both SC and MP were absent.

In configurations 2 and 4, error increased significantly in untimed trials relative to timed trials, whereas error decreased significantly in configuration 1. For configurations 1, 2, and 4, the change in percent error identified in Figure 3.4 arises not equally from all trials, but primarily from the trials in which both SC and MP cues are absent.

Chapter 4

Discussion

4.1 Interpreting Results

There are three main conclusions that can be drawn from the results of this study. First, size change affects the accuracy of the depth perception task significantly, whereas motion parallax does not. This can be clearly seen in Figure 3.1, as the presence or absence of size change respectively decreases or increases the error against the mean. For motion parallax, presence nor absence affects the mean error rate. From Figure 3.2, one can see that this is true for trials in which both cues were present or absent at the same time. Only the presence of size cue is an accurate indicator of depth perception accuracy.

The second conclusion is that motion parallax can serve different roles depending on which context they are employed in. As can be seen in the first two bars in the graph for configuration 1 in Figure 3.3, the elimination of motion parallax and only using the size cue decreases error. Due to the design of configuration 1, the size cue is sufficient to complete the task of matching the two display planes, and motion parallax rather serves

as a distraction during the 10 seconds subjects have to complete the task. This finding is consistent in the untimed experiment (Figure 3.5), where the negative yellow bar in Configuration 1 show that with more time, participants perform better.

On the other hand, when neither motion parallax nor size cue is available, the addition of motion parallax cue enhances accuracy. This can be seen in Figure 3.3, although further collection of data and statistical analysis is necessary to firmly draw this conclusion.

Lastly, the comparison between timed and untimed trials shows the surprising accuracy that quick intuition bestows in difficult trials. In untimed trials, participants took more than twice as long to complete each trial, making their best effort to align the planes as precisely as possible. However, compared to the timed 10-second trials, error rates in the untimed trials were higher by as much as 9 percent points (Figure 3.4). As the configuration of the darts became more difficult (from C1 to C4), the effect of decreased accuracy in trials with more time and effort is more salient.

The effect of decreased accuracy among untimed trials, particularly in difficult trials, can be seen in Figure 3.5 as well. Of the four depth cue combination settings, the blue setting in which neither SC or MP are present is the setting in which it is the most difficult to successfully complete the depth equalizing task. As can be seen in configurations 2 and 4, subjects perform particularly poorly in this most difficult trial setting when they are given as much time as needed. These results indicate that in trials that are difficult to gauge depth, quick intuition is far more precise than trying to logic through the display for a longer period of time.

Overall, both aim 1 of creating a two-dimensional virtual reality display and aim 2 of comparing monocular cues were accomplished in this study. However, the two hypotheses were proven to be false, as size cue conveyed more information about depth than motion

parallax.

4.2   Applications

If a convincing VR display could be developed using only monocular cues, the application of VR can expand dramatically without the need for additional equipment or complicated software. Because the size and motion parallax cues rely solely on the distance of the object from the viewer, the only equipment required to implement these cues on a 2D image is a front-facing camera. Because most hand-held devices, such as smartphones or tablets, come equipped with front-facing cameras, distance can be measured using the video footage and then integrated into size or motion parallax calculations.

The implementation of a simple but vivid VR 2D display will allow for further interaction with images or videos online, while also avoiding the hassle or nausea accompanying many of the current technologies involving binocular cues. Furthermore, a simple interactive VR display on widely available devices can be used for education, particularly in fields where spatial orientation is important, such as engineering, anatomy, or organic chemistry.

4.3   Future Work

This study delineated the effect of SC and MP monocular cues in perceiving depth on 2D digital displays. Further research on the depth portrayal of other monocular cues, such as contrast or aerial perspective, on 2D displays will shed light on effective cue combinations that does not require binocular cues. Moreover, a study on the depth portrayal by binocular

cues on 2D digital displays with a similar approach can further delineate whether a very convincing VR is achievable using only monocular cues.

In terms of the 2D computer-generated display, its depth illusion can be enhanced by integrating perspective projection matrix and processing the subject face tracking footage to include roll, pitch, and yaw.

References

Bauer, A., Dietz, K., Kolling, G., Hart, W., & Schiefer, U. 2001, Graefe's archive for clinical and experimental ophthalmology, 239, 400

Bradski, G., & Kaehler, A. 2008, Learning OpenCV: Computer vision with the OpenCV library (" O'Reilly Media, Inc.")

Burdea, G. C., & Coiffet, P. 2003, Virtual reality technology (John Wiley & Sons)

Ferris, S. H. 1972, Journal of experimental psychology, 95, 258

Grosslight, J., Fletcher, H. J., Masterton, R. B., & Hagen, R. 1978, Human Factors, 20, 27

Haase, W., & Wermann, O. 1985, Klinische Monatsblätter für Augenheilkunde, 186, 337

Howard, I. P. 2002, Seeing in depth, Vol. 1: Basic mechanisms. (University of Toronto Press)

Howard, I. P., & Rogers, B. J. 2002, Seeing in depth, Vol. 2: Depth perception. (University of Toronto Press)

Lee, J. C. 2008, IEEE pervasive computing, 7, 39

REFERENCES

McKnight, A. J., Shinar, D., & Hilburn, B. 1991, Accident Analysis & Prevention, 23, 225

Mon-Williams, M., & Wann, J. P. 1998, Human Factors, 40, 42

Mon-Williams, M., Warm, J. P., & Rushton, S. 1993, Ophthalmic and Physiological Optics, 13, 387

Palou, G., & Salembier, P. 2011, in 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, 1093–1096

Peirce, J. W. 2007, Journal of neuroscience methods, 162, 8

Peli, E. 1995, Optics and Photonics News, 6, 28

Qian, N. 1997, Neuron, 18, 359

Rogers, B., & Graham, M. 1979, Perception, 8, 125

Schmidt, I. 1966, PA: Society of Automotive Engineers

Steuer, J. 1992, Journal of communication, 42, 73