Perceptual interactions in visual depth perception: A quantitative EEG study

Jeremy Ma

Abstract

Percepts of different sensory modalities have been shown to interact with one another. Previous studies have qualitatively looked into the results of the interactions between stereo depth and specific pictorial depth cues, but failed to address the interaction themselves. My study will quantitatively investigate perceptual interactions between pictorial (two dimensional) and stereo (three dimensional) depth perception, the combination of which I term *combined depth perception*. Using a steady-state visually evoked potential (SSVEP) paradigm and a high density EEG net, the neural activity of eight subjects was recorded during the alternation and detection of different types of depth. I proposed and implemented the Relative Peak Strength variable, in order to quantitatively compare and plot the response strengths of each electrode. The perception of pictorial depth was observed to induce neural activity in the ventral stream, while stereo depth was observed to induce activity in the dorsal stream, suggesting these percepts have different functions when perceiving depth. The heat map for combined depth perception was significantly different from the mere sum of pictorial and stereo depth, suggesting combined depth behaves like a Gestalt, in which pictorial depth and stereo depth are not processed parallel to each other. Furthermore, heat maps of combined depth resembled the heat maps of stereo depth scenarios, suggesting stereo depth is the predominant type of depth perceived in combined depth. Pictorial depth was also observed to lower neural activity when comparing different amounts of depth; but has no major contributions when detecting depth. These results suggest that pictorial depth can have different roles on combined depth depending on the task; for example, pictorial depth can play a supplementary role when comparing depth, but have no influence when detecting depth. When pictorial depth is not able to supplement stereo depth, the increased amount of neural activity could be a crucial reason of visual fatigue when viewing stereoscopic displays. This connection between depth percepts and streams should be further investigated, as it can lead to a better understanding of the perceptual mechanisms underlying the reconstruction of the visual world.

We, as humans, have the ability to obtain conscious experience and make sense of the world through perception. The process of perception converts raw sensory information, called sensations, into meaningful forms, called percepts. One major goal of cognitive science is to understand and decode this process. However, this task is extremely difficult due to the large amount of information being combined and processed during perception (Hogendoorn, 2015).

The process of seeing and identifying objects is mediated through a process called visual perception, in which the brain analyzes light hitting the retina. In addition, there are other types of perception such as olfactory, auditory and haptic. Current studies have shown that these different modes of perception are not individual processes; instead they influence one another, affecting the resulting percept (Hidaka & Ide, 2015; Ide & Hidaka, 2013; Rosemann, Wefel, Elis & Fahle, 2016; R. Sekuler, A. Sekuler & Lau, 1997). A common example would be the ventriloquist effect, in which vision has dominance over auditory sensations, altering the apparent source of sound. The process of combining sensory information into a sensible percept is called multisensory integration and is widely studied among neuroscientists and cognitive scientists (Holmes & Spence, 2005; Shams, Kamitani & Shimojo, 2000).

The investigation of multisensory integration is challenging due to the large number of unknown cross-modal interactions. Some factors and interactions include alertness and attention from auditory information, tactile enhancement effects from haptic information (Gillmeister & Eimer, 2007) and a subject's innate bias towards specific senses. Due to the complexity of multisensory integration, this study would investigate the unimodal interaction between two visual percepts of depth – stereo (three dimensional) and pictorial (two dimensional) depth. Since these percepts are both visual and spatial output, there will be fewer confounding interactions than in multimodal perception. This study hopes to shed light on perceptual interactions during visual depth perception, in order to increase our understanding visual system.

Review of Literature

Visual depth perception is defined as the perception of three dimensions based on visual depth cues, or clues that provide information about depth; it is categorized into pictorial and stereo depth perception. Both percepts are simultaneously perceived in real life. In order to study the interactions between these percepts, they must first be individually quantified. A good way to do so is by identifying the neural activities that correlate to these percepts through physiological methods. My study used electroencephalography (EEG) as a means to detect neural activity.

Stereo Depth Perception

Stereo depth perception, or *stereopsis*, is the perception of realistic and actual three dimensions. In real life, this is the predominant perception of depth, which is suggested to play a major role in visuomotor actions (Levi, Knill & Bavelier, 2015). There are multiple hypotheses that explain stereopsis, the most predominant one being the Visual Parallax Hypothesis.

Visual Parallax Hypothesis. The Visual Parallax Hypothesis was the first hypothesis developed to explain stereopsis; it suggests that stereopsis is induced by visual parallax (Wheatstone, 1838). Visual parallax is the presentation of different perspectives of the same scene. The main type of visual parallax in humans is binocular disparity, which is the slight difference in images between the left and right eye (Figure 1). By fusing these two images, we are able to perceive stereo depth. My study would simulate binocular disparity using red-green anaglyphs.

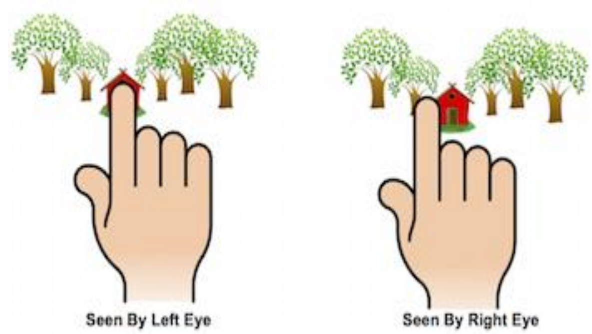

Figure 1. Demonstration of binocular disparity. (Study.com)

Physiological studies. Previous studies have been conducted to assess neural activity during stereo depth perception, most of which used visually evoked potentials (VEP) (Akay, 2009; Liu, Meng, Wu & Dang, 2013). VEP is a common EEG technique that involves presenting a stimulus and recording brain activity simultaneously in order to measure the response evoked in the brain by the stimulus. Due to its vulnerability to artifacts and latency, EEG paradigms, especially VEP require a large number of trials and careful control in order to maintain a reasonable signal-to-noise ratio (SNR).

VEP paradigms were used by both Akay (2009) and Liu et al. (2013) to measure neural activity during the viewing of stereoscopic stimuli. Each study found responses related to depth perception in both occipital and parietal regions of the brain around 120ms to 320ms after stimulus onset. Other studies suggested that stereo depth is processed near the extrastriate and primary visual cortices, however induced neural activity was found to be widespread throughout the brain (Backus, Fleet, Parker & Heeger, 2001; Parker, Smith & Krug, 2016).

Pictorial Depth Perception

As compared to stereo depth, pictorial depth is depth from a two dimensional plane as seen in pictures (Figures 2a, 2b). Pictorial depth perception is induced by pictorial depth cues, which are two dimensional depth cues that do not require both eyes to be perceived.

Pictorial depth cues. Pictorial depth cues are also called monocular cues since only one eye is required to perceive the depth within the stimuli. For example, shading is the use of light and shadows to convey the sense of depth (Figure 2a), while perspective is the phenomenon of converging parallel lines due to the lines subtending different visual angles (Figure 2b). This study would use shading and perspective cues to induce the perception of pictorial depth, while using an EEG to observe neural activity.

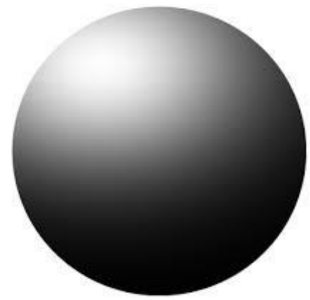

Figure 2a. Shading as a depth cue.

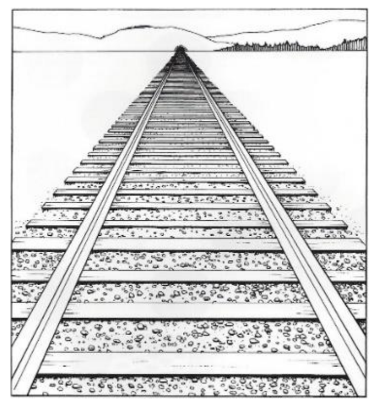

Figure 2b. Perspective as a depth cue (guidetodrawing.com).

Physiological studies. Hou, Pettet, Vildavski & Norcia (2006) looked at neural correlates of pictorial depth, specifically shading. They found a negative VEP in respond to pictorial depth around occipital and parietal lobes between 200ms and 400ms after stimulus onset. When compared with physiological studies on stereo depth, stereo and pictorial depth seemed to be processed in nearby neural pathways, however pictorial depth takes more time to process and perceive.

Combined Depth Perception

In real life, pictorial depth cues and binocular disparity are both present in our daily experiences, which leads to the simultaneous perception of stereo and pictorial depth. However, there is no term for this combined percept, so I will call it *combined depth perception*. There were studies that qualitatively investigated the combination of stereo depth and specific pictorial depth cues. They all show that the impression of depth was strengthened when both stereo and pictorial cues were present, however they failed to address the mechanisms behind the combination (Bulthoff & Mallot, 1988; Johnston, Cumming & Parker, 1993; Schiller, Slocum, Jao, & Weiner, 2011). In this study, I propose possible interactions between pictorial and stereo depth percepts based on two different theories: the structuralistic approach and the Gestalt approach.

Structuralist approach. The structural approach suggests that percepts are solely formed by combining sensations. For example, the concept of structuralism would suggest that the perception of images on a screen is due to the summation of individual pixels. Following this approach, combined depth would be perceived as the combination of stereo depth and pictorial depth. Hence neural activities during the perception of combined depth should correspond with neural activity during the perception of pictorial and stereo depth.

Gestalt theory. The Gestalt theory suggests that the brain organizes and analyzes sensory information to form a Gestalt, or a whole new percept (Koffka, 1935). This resulting Gestalt could be entirely different from the raw sensory information, as seen through the impression of a white triangle in Kanizsa's triangle (Figure 3). According to this theory, pictorial and stereo depth would combine to form a whole new percept which would behave differently from pictorial or stereo depth percepts, and hence induce different neural activities.

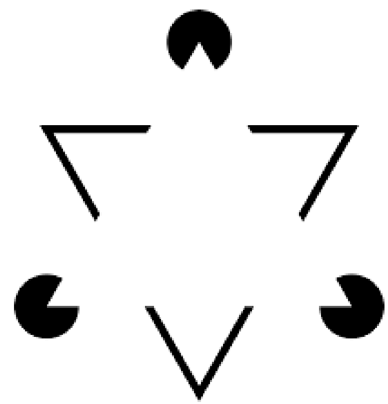

Figure 3. Kanizsa's Triangle as a demonstration of Gestalt.

It should be noted that Gestalt theory is considered more descriptive than analytical due to its absence of quantitative data. Nonetheless, it is a commonly used theory for explaining percepts (Hawkins, Houpt, Eidels & Townsend, 2016; Jäkel, Singh, Wichmann & Herzoge, 2016). I propose that the idea of a Gestalt could be examined quantitatively through looking at combined depth perception.

Existing literature suggests that experience plays a role in stereo and pictorial depth perception (H. Bulthoff, I. Bulthoff & Sinha, 1998; Sinha & Poggio, 1996), meaning both percepts behave like a Gestalt – they are formed through organizing and analysing sensory information. It is likely that the resulting percept from combining these Gestalt-like percepts would also behave like one, integrating both pictorial and stereo depth.

Based on the literature, this study would take a Gestalt approach to analyse the combined perception of depth. This would be achieved through the comparison of different neural activities in stereo, pictorial and combined depth perception.

Hypotheses

H_0: Combined depth perception will yield EEG signals which are equivalent to the sum of stereo and pictorial depth components.

H_1: Combined depth perception will yield EEG signals which differ from the sum of stereo and pictorial depth perception.

Objectives

1. Create stimuli to induce neural activity for pictorial, stereo and combined depth
2. Locate and determine neural correlates for stereo depth and pictorial depth
3. Analyze neural activity in combined depth perception

Methods

I reached out to Dr. Pawan Sinha from MIT after doing independent research on depth perception, geometric optics and basic neuroscience. After disclosing my ideas for a study on combined depth perception, he suggested that I use the EEG system in his lab (Sinha Lab at MIT) and a steady-state visually evoked potential (SSVEP) paradigm to quantify the percepts of depth. I then proceeded to create the objectives, rationale and procedures of the study independently. Before working with human subjects, I acquired my CITI certification on human research online. I was also added to an approved IRB protocol in order to operate an EEG on human subjects. Under the supervision of my mentor and Dr. Wasifa Jamal, a postdoctoral associate of the lab, I independently created all the stimuli and taught myself MATLAB to code the presentation program. Upon my arrival at the lab, I learned how to operate an electrolyte EEG cap system. I ran my study on eight volunteers and ensured that they understood the procedures and had given appropriate consent. With help from Dr. Jamal, I pre-processed the EEG data with a band pass filter and ran fast Fourier transformation (FFT) on the data to convert the data into the frequency domain. I also developed a novel method to quantify frequency peaks and plot them on a brain heat map. I independently ran statistical analysis and drew conclusions from the data. The entire study was fully administered by me with minimal assistance from my mentor and members of the lab.

Subjects

Subjects (n=8) aged between 17 and 32 participated in this study. All eight subjects have corrected-to-normal visual acuity, and a sample stimulus was presented to ensure they were able to view stereoscopic images in red and green anaglyphs. After going over the experimental procedures and consent material, informed consent was obtained from each of the subjects. Parental consent was also obtained for minors under the age of 18.

EEG Paradigm

The Steady-State Visually Evoked Potential (SSVEP) paradigm was used in this study. An SSVEP paradigm involves presenting a flickering stimuli at a set frequency, then locating EEG electrodes which responded at this specific frequency. This paradigm is known to have a relatively high signal-to-noise ratio amongst known EEG paradigms (Norcia, Appelbaum, Ales, Cottereau & Rossion, 2015), which makes it suitable for this current study as a means to detect neural activities.

Stimuli

I designed and created nine pairs of stimuli using Adobe Photoshop CC 2017 and presented them on a 16x12in 2007FPb Dell monitor with 1600x1200px resolution and a refresh rate of 60Hz. All pairs were programmed to flicker every 400ms (2.5Hz) for 20 seconds using Psychtoolbox for MATLAB, with consideration of the latency of evoked responses to depth from prior research (Hou et al., 2006; Liu, 2013). A conventional flickering checkerboard that alters at the same frequency was used to detect the peak frequency (Norcia et al., 2015). The other eight stimuli were equally split into two groups of four – stimuli which involved *detecting* depth and stimuli with constantly *alternating* amounts of depth.

Stereoscopic stimuli. Stereoscopic stimuli in this study utilizes binocular disparity and red-green anaglyphs. Since red and green are opposite colors, green light would appear black and red light would appear white through red lenses, and vice versa. Using this property of light and a pair of red-green glasses, two different images could be projected into the left and right eye. All stimuli in this study are created in grayscale due to the usage of anaglyphs. Dots are also added to the stimuli to enhance disparity information. Since electronics would interfere with EEG acquisition, this technique is optimal as it can induce binocular disparity without using electronics such as head mounted displays.

Using my knowledge of geometry and optics, I developed a formula (see Equation 1) to calculate the distance between left and right images in order to produce a stereoscopic stimulus (Figure 4, overleaf):

$$S = \frac{|D_2|}{D_1 + D_2} \times IPD \qquad (1)$$

The direction the left and right images have to shift is dependent on whether the apparent image is in front of or behind the screen. If the image is behind the screen, the red image is shifted towards the right and the green image is shifted to the left by S/2, and vice versa. In my experiment, the viewing distance (D_1) would be 60cm, the distance from stimulus to screen (D_2) was 5cm and an interpupillary distance of 6.3cm was used, which is the average adult interpupillary distance (Dodgson, 2004). These measurements were used to create the stimuli. Red and green images were created using the RGB color filters in Adobe Photoshop. The two images were perspective warped according to the angle of convergence of the eyes.

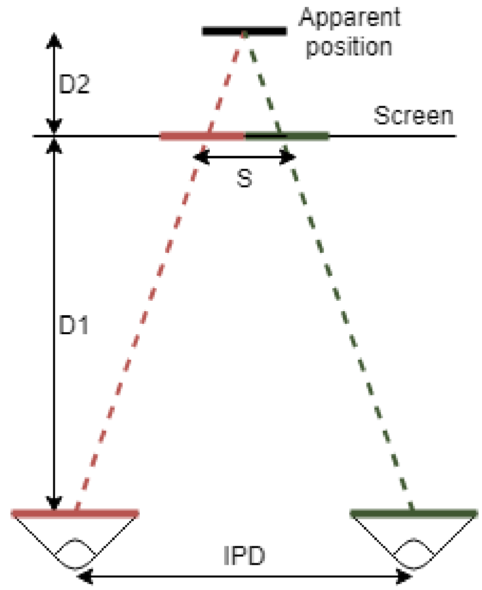

Figure 4. Geometry of stereoscopic viewing through anaglyphs. Assume an observer with interpupillary distance of IPD and D_1 away from the screen. An image with apparent position of D_2 away from the screen would require two different projections S apart from each other on the screen. If the image is in front of the screen , D_2 would be negative and the shift would be in the opposite direction.

Detection of depth. Stimuli for detection of depth flickered between pictures which had depth information and no depth information. Stimuli with pictorial depth utilized shading as a pictorial depth cue by using the drop shadow function in Photoshop. Five images of a rectangle that subtended a visual angle of 7.6 x11.4 degrees were created. The five images were: *control condition* which only contained the rectangle; *pictorial condition* which had a drop shadow behind the rectangle; *stereo condition* in which the rectangle was viewed with disparity that made it appear 5cm in front of the screen; *combined condition* in which the rectangle had both a from the *pictorial condition* and disparity from the *stereo condition*; *conflicting*, which was the combination of the *pictorial condition* and disparity which made the rectangle appear 5cm behind the screen. Since SSVEP uses flickering pictures to detect neural activity corresponding to specific percepts of depth, the five images were paired up to create four conditions: *control-stereo* (Figure 5A, overleaf), *control-pictorial* (Figure 5B, overleaf), *control-combined* (Figure 5C, overleaf) and *control-conflicting* (Figure 5D, overleaf).

Alternation of depth. Stimuli for alternation of depth flickered between two pictures with had different amounts of depth. Stimuli with pictorial depth utilized perspective as a pictorial depth cue by placing a perspective grid in the background. Eight images were created, four of which had a larger rectangle subtending a visual angle of 10.9 x 16.2 degrees and the other four had a smaller rectangle of 3.27 x 4.9 degrees, maintaining the same aspect ratio between rectangles. The larger rectangle was viewed with disparity that presents it 4 cm behind the screen, while the smaller rectangle had disparity that presents it 12 cm behind the screen during stereo and combined conditions. These eight images were grouped into four pairs, each pair included one larger and one smaller rectangle. The *control condition* had the rectangles on a gray background and no disparity (Figure 6E); the *pictorial condition* had the rectangles on a perspective grid, so the rectangle would appear to move back and forth (Figure 6F); the *stereo condition* had disparity information presented with the rectangles on a gray background (Figure 6G); the *combined condition* had both disparity and a perspective grid (Figure 6H).

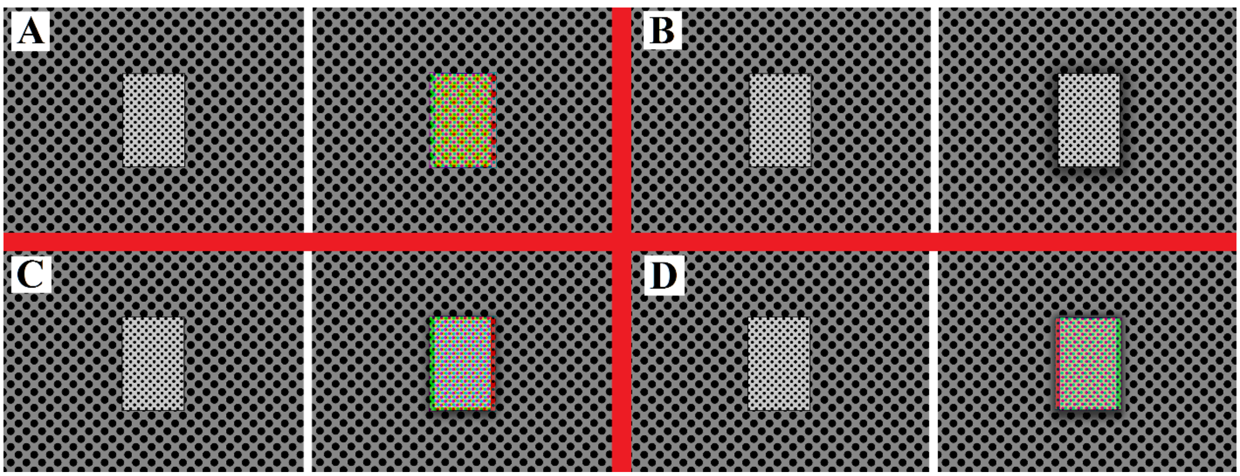

Figure 5. All four detecting condition stimuli. A) Control-stereo, which induces the detection of stereo depth. B) Control-pictorial, which induces the detection of pictorial depth. C) Control-combined, which induces the detection of combined depth. D) Control-conflicting, which induces the detection of conflicted depth.

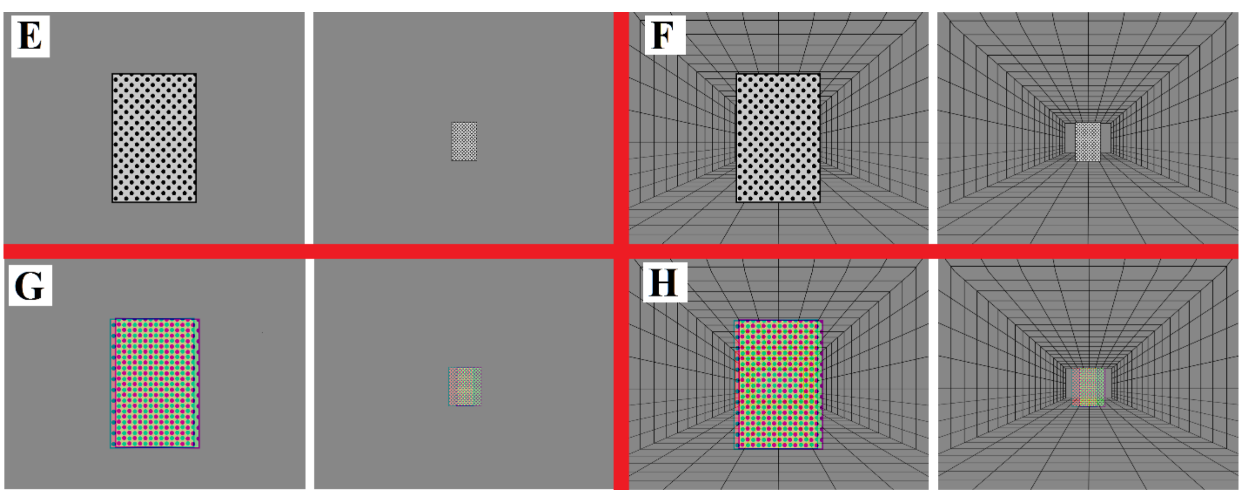

Figure 6. All four alternating condition stimuli. E) Control, which induces activity related to the alternation of size. F) Pictorial, which induces the perception of alternating pictorial depth. G) Stereo, which induces the perception of alternating stereo depth. H) Combined, which induces the perception of alternating combined depth.

Procedure

Subjects were seated 60 cm away from a computer screen inside a room with adequate lighting while wearing a pair of red-green glasses and an EGI HydroCel GSN Plus EEG net. Beforehand, the net was soaked into an electrolyte solution, composed of 5 g potassium chloride, 5 mL baby shampoo and 1 L of distilled water. The midpoint of the subject's scalp was marked by the intersection of the lines from anion to nasion and from left ear to right ear. The Cz (reference) electrode of the net was placed on this midpoint and the other 127 electrodes were positioned according to Cz (Figure 7).

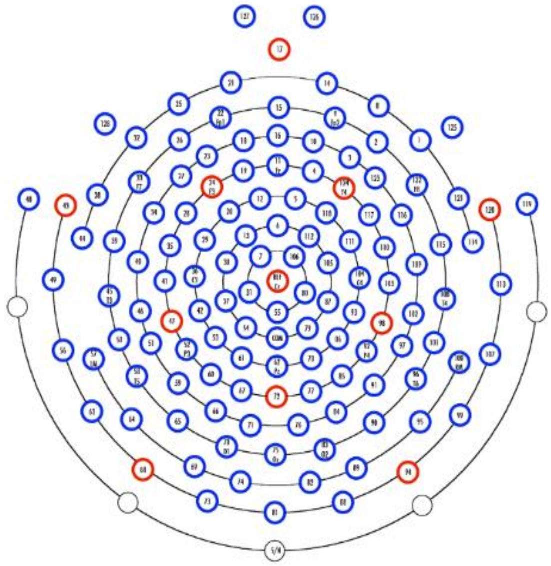

Figure 7. Channel map for EGI HydroCel GSN Plus.

The subjects were then asked to look at the center of the screen for all nine conditions – *checkerboard, alternating pictorial, alternating control, alternating stereo, alternating combined, control-pictorial, control-stereo, control-combined* and *control conflicting* - each lasting 20 seconds. A fixation cross was presented for 1 second between conditions to mark the center of the screen and facilitate binocular fusion when viewing stereoscopic stimuli. All EEG information was collected through an EGI Net Amps 300 amplifier and recorded with NetStation.

Data Analysis

Pre-processing. Each subject's EEG recording was first pre-processed with a band pass filter of 0.5 – 40 Hz, filtering high and low frequency noises, then it was split into nine 20 second epochs and fast Fourier transformation (FFT) was run on each epoch. This transformation breaks down a wave into its frequency components, plotting the magnitude of each component on a power spectrum (Figure 8, overleaf).

Although the presentation was programmed to flicker at a rate of 2.5 Hz, most electrodes in the occipital to parietal regions had a response at 2.4 Hz during the checkerboard condition (Figure 8, overleaf). In order to confirm the flickering rate, I recorded the screen presentation rate with a high speed camera and confirmed the flickering rate to be 2.4 Hz.

Plotting magnitude. Due to the usage of a high density EEG net, heat maps were most suitable and efficient in pinpointing activated electrodes. However, due to the small magnitude ($\mu V/Hz^{0.5}$) of frequency responses, any noise or artifacts could cause noticeable differences in the magnitude of neural activity between electrodes. One main source of artifacts were the non-identical neural activities in different regions of the brain, which made it unreliable to plot magnitudes on a heat map. For example, at electrode 93, the response peak for 2.4 Hz was around 1.1 $\mu V/Hz^{0.5}$, while the average magnitude of theta waves (4-8 Hz) was around 0.3 $\mu V/Hz^{0.5}$; the response peak was three times as large as the average activity of the electrode (Figure 8). In electrode 81 of the same subject, the response peak was around 1.8 $\mu V/Hz^{0.5}$, which was almost two times larger in magnitude than the peak in electrode 93. However, the average activity of electrode 81 seemed to be around 0.4 $\mu V/Hz^{0.5}$, which made the response peak three times as large as the average activity (Figure 9).

If the magnitude of response peaks were to be used for the analysis, the response peak in electrode 81 would be two times larger than the peak in electrode 93. However, if the average activity of the electrode were to be taken into account, both the response peaks are around three times larger than their respective average electrode activities. This implies that they are equal in response strength, as the larger magnitude of the response peak in electrode 93 would be a result of stronger overall neural activity rather than a stronger response. In order to quantify the idea of a peak and take the average activity strength of the electrode into account, I proposed and implemented a variable called the *relative peak strength* (RPS).

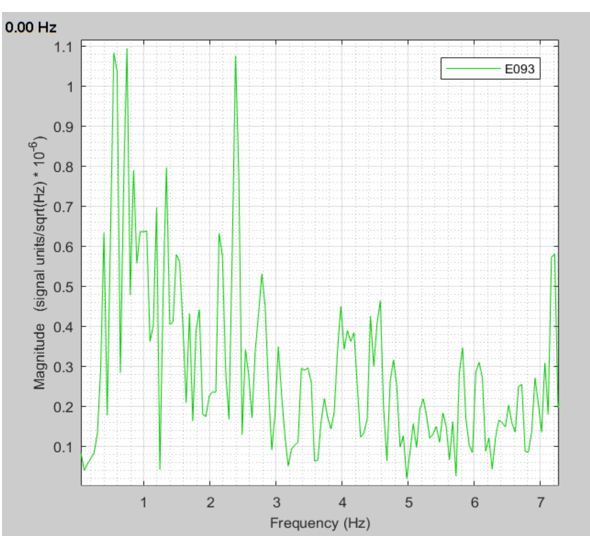

Figure 8. FFT power spectrum for checkerboard condition at electrode 93. (Occipital parietal)

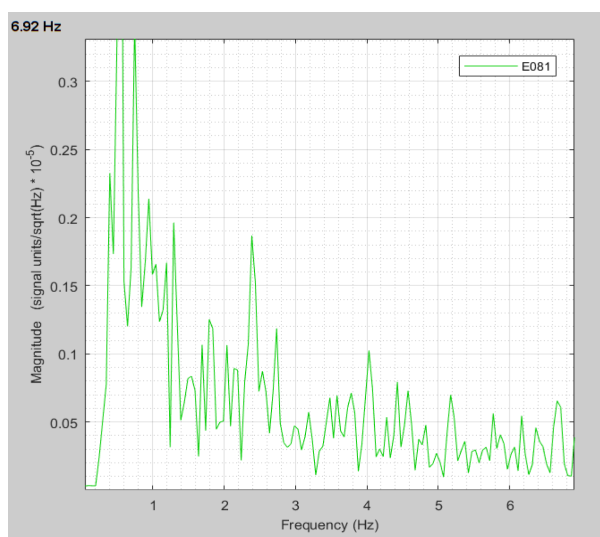

Figure 9. FFT power spectrum for checkerboard condition at electrode 81. (Occipital parietal)

Relative peak strength. The relative peak strength (see Equation 2) is a measure of the difference in magnitude of a target frequency (V_{target}) and the average magnitude of the successive frequencies (V_{avg}) relative to the average magnitude. This measurement is similar to the signal-to-noise ratio (SNR), but it has more specific targets for comparison. The scale starts at 0 and anything above 1 qualified as a peak.

$$RPS = \frac{V_{target} - V_{avg}}{V_{avg}} \tag{2}$$

I programmed an algorithm in MATLAB to calculate the 2.4 Hz RPS for each electrode relative to the average magnitude of frequencies from 3-8 Hz, a range that includes all frequencies of theta waves and is near the target frequency. The algorithm would then plot these values on a heat map to help visualize the data and look for regions of interest.

Statistical analysis. In order to locate neural correlates for pictorial and stereo depth, a heat map was made by subtracting the control RPS values from the experimental conditions. Regions of interests (ROI) on the heat maps were chosen through visual inspection. Student's t-tests were run between the average RPS in regions of interest under alternating control condition and experimental conditions. Statistical tests were not run between detection conditions due to the lack of a control scenario, though regions of interest were mentioned.

Next, a theoretical combined depth condition was created based on the activated regions from pictorial and stereo depth. This condition was made on the assumption that activated regions from pictorial and stereo depth perception would be the only activated regions when the two are combined. If a region was activated in both conditions, the higher RPS value was used. This theoretical condition would be used to model the simultaneous perception of stereo and pictorial depth based on the structuralistic approach.

Finally, the perception of combined depth was analysed through visual inspection. In addition, two tailed t-tests were run between the overall RPS of experimental combined scenarios and stereo, pictorial and theoretical conditions for both alternation and detection of depth to compare the amount of neural activity.

Only the posterior of the head was considered, since it contains most regions of the brain that were responsible for visual perception. Alpha was set at 0.05 for all t-tests.

Results

Results of alternating and detecting individual depth were investigated to look for neural correlates for stereo or pictorial depth. Then, the alternation and detection of combined depth were observed. Finally, statistical tests were conducted on observed trends.

Neural Correlates

Alternation of depth. RPS values from the control was subtracted from pictorial and stereo depth conditions to take into account neural activity due to the alternation of size of the stimuli. These new values were plotted on a heat map to visualize regions of the brain that were activated by the alternation of depth (Figure 10, Figure 11). After visual inspection of the RPS heat maps, Student's t-tests were run between the control and experimental conditions to determine whether the RPS values were significantly different from the control condition (Table 1). Results showed no significant difference in left temporal (p=0.155), occipital (p=0.231) or right parietal-temporal (p=0.3885) regions for pictorial depth. There was also no significant difference in left occipital (p=0.4504), right occipital (p=0.3807), right temporal (p=0.3968) and left temporal (p=0.2702) regions for stereo depth.

Table 1.
Results of t-tests on region of interest.

Depth	ROI	Electrodes	p	Color
Pictorial	L.temporal	46, 50, 57	0.155	Red
	Occipital	81	0.231	Green
	R.parietal-temporal	91, 92	0.3885	Purple
Stereo	L. temporal	57	0.2702	Red
	R.temporal	96	0.3968	Yellow
	L.occipital	73	0.4504	Green
	R.occipital	82	0.3807	Purple

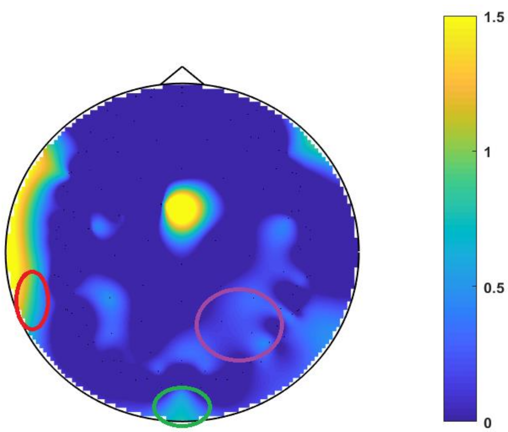

Figure 10. RPS heat map for alternation of **pictorial** depth (control subtracted).

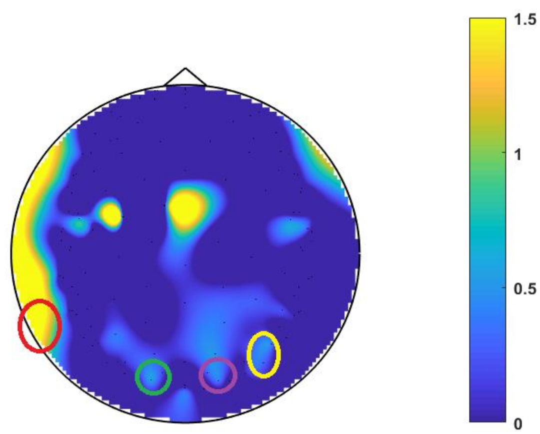

Figure 11. RPS heat map for alternation of **stereo** depth (control subtracted).

Detection of depth. Due to the inclusion of the control in the stimuli, no statistical tests were run on the detection of depth, as there was no control condition *per se***.** The RPS heat map for pictorial depth detection (Figure 12) shows activity extending from the occipital lobe (blue circle) to the right temporal lobe (red circle) and left temporal lobe (purple circle). For stereo depth, the right temporal lobe (purple circle) was observed to respond to the stimuli, while weaker activity was observed to extend form the occipital lobe up to the parietal lobe (blue circle) and right parietal-temporal lobe (red circle) (Figure 13).

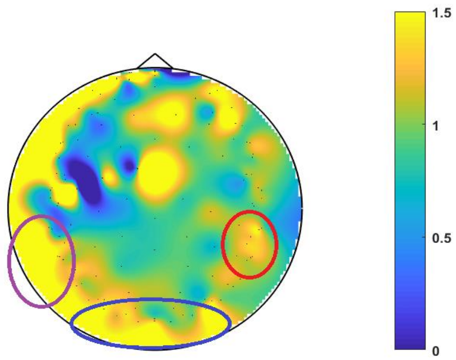

Figure 12. RPS heat map for detection of **pictorial** depth.

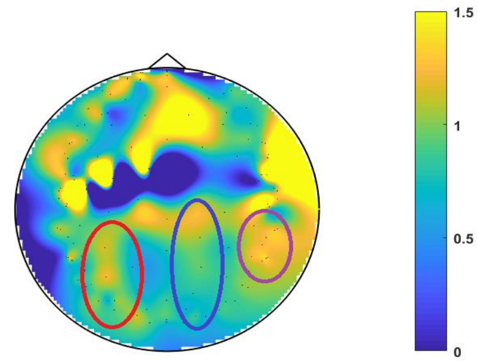

Figure 13. RPS heat map for detection of **stereo** depth.

Combined Depth Perception

Visual inspection. Figure 14 shows activation from occipital to parietal regions (red) during the alternation of combined depth; Figure 15 shows activations from occipital regions extending into parietal (red) and temporal (blue) regions during the detection of combined depth.

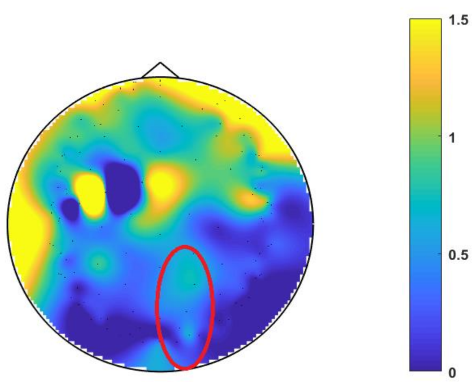

Figure 14. RPS heat map for **alternation** of combined depth.

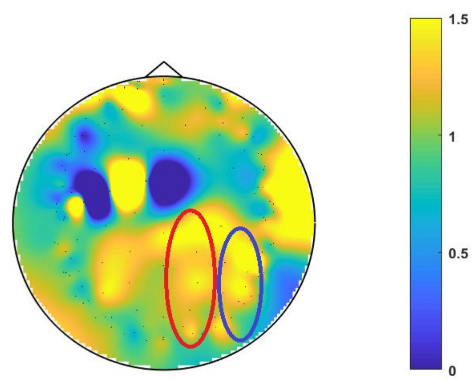

Figure 15. RPS heat map for **detection** of combined depth.

Condition comparisons. The mean RPS for all posterior electrodes for each scenario was computed to quantify and compare the neural activity in each scenario (Table 2, overleaf). Two-tailed t-tests were run to compare pairs of conditions in Table 2 (overleaf), only visually observed trends were tested. Comparisons involving the control-conflicting scenario yielded no significant results.

Table 3 shows the results of t-tests on observed trends. The mean RPS in alternation of combined depth was observed to be smaller than the alternation of pictorial (p=0.1065), stereo (p=0.1143) and significantly smaller (p=0.003) than theoretical conditions. The mean RPS in detection of combined depth was not significantly different than the detection of pictorial (p=0.4884) or stereo (p=0.3901) and was observed to be smaller than the theoretical condition (p=0.136). The mean RPS of alternation conditions was significantly greater than detection conditions (p=0.0187).

Table 2.
Mean RPS table.

Condition	Mean RPS
AltP	0.6583
AltS	0.6607
AltCOM	0.2733
AltTheo	1.1957
DetP	0.9455
DetS	1.0776
DetCOM	1.0973
DetTheo	1.7004

Table 3.
Results of two-tailed t-tests comparing mean RPS values. (Significant results highlighted)

Condition	p
AltP>AltCOM	0.1065
AltS>AltCOM	0.1143
AltTheo>AltCOM	0.003
DetP<DetCOM	0.4884
DetS<DetCOM	0.3901
DetTheo>DetCOM	0.136
Alt<Det	0.0249

Discussion

Objectives

Stimulus creation. Utilizing fundamental geometrical optics, stereoscopic stimuli were successfully made. All subjects were able to perceive stereo depth within the stimulus. My derivation of the equation for disparity on a screen will facilitate future experiments that require precise stereoscopic stimuli.

Neural correlates. This study successfully utilized a high density EEG net and the SSVEP paradigm to create an algorithm which locates regions of the brain that respond to a stimulus, despite the poor spatial resolution that EEG systems are known for (Burle et al., 2015). The lack of significant results in this study was likely to be due to the small sample size (n=8) and a low SNR in EEG data. It is worth noting that the regions of interest were still largely consistent with other studies (Hou et al., 2006; Liu et al., 2013).

Due to the large number of activated regions, ROI is not the optimal method to analyse the data; instead, the data was analysed by observing streams of activation instead.

Two-streams hypothesis. According to the two-streams hypothesis, visual information tends to travel along two different streams – the dorsal stream and ventral stream (Goodale & Milner, 1992). The dorsal stream runs from the occipital lobe to the parietal lobe; the ventral stream runs from the occipital lobe to the temporal lobe. Visual observations of the results suggest neural activity related to depth perception followed these streams. Specifically, neural activity was observed along the dorsal stream during the alternation of stereo depth (Figure 11) and along the ventral stream during the alternation (Figure 10) and detection of pictorial depth (Figure 12). The detection of stereo depth had activation in both streams (Figure 13).

Mishkin, Ungerleider & Macko (1983) suggested that the dorsal stream is responsible for processing spatial information, meaning both stereo and pictorial depth should be processed in it. Other studies suggested that the dorsal stream processes sensory information that guides action, while the ventral stream is responsible for information related to perceptual tasks, such as recognizing (Goodale & Milner, 1992; Hebart & Hesselmann, 2012). In this study, pictorial and stereo depth neural activity followed different streams, which conflicts with the explanation offered by Mishkin et al. (1983).

Stereo depth was observed to induce activation in dorsal areas. Since stereo depth is the most reliable type of depth perceived in real life and is believed to guide visuomotor tasks (Levi, et al., 2015), the fact that stereoscopic visual information is processed through the dorsal stream would seem reasonable, as the dorsal stream is responsible for action-related information. On the other hand, pictorial depth was observed to be processed in the ventral stream. One possible reason is that since pictorial depth only provides relative depth information, it only serves as a guideline towards spatial information – making pictorial depth responsible solely for the reconstruction and comparison of depth, not the quantification of it.

Despite the lack of significant results to support this claim, observations suggest that the perception of stereo and pictorial depth are processed in different neural streams, and hence each serves a different purpose in the processing of spatial information.

Combined depth perception. Due to lack of significance of ROI in neural correlate analysis, more general approaches were taken in analyzing combined depth perception, such as stream observation and overall RPS.

Rejection of H_0. The null hypothesis took a structuralistic approach towards combined depth perception, hypothesizing that the combined perception of depth would be solely the sum of pictorial and stereo depth. Were this to be the case, the EEG would reflect neural activity similar to the theoretical condition, which models the simultaneous perception of stereo and pictorial depth.

The results of two-tailed t-tests show that the experimental combined condition had significantly less activity than the theoretical during alternation of depth ($p=0.003$). During detecting depth, the experimental condition was observed to have less activity than the theoretical, despite the lack of significant result ($p=0.136$). Another observation was that ventral stream activity was not seen in the alternation of combined depth, despite being in the alternation of pictorial depth.

These results make a strong point that the perception of combined depth is not solely the sum of the stereo and pictorial depth, but rather the result of integrating and re-analyzing both stereo and pictorial depth information.

Acceptance of H_1. The working hypothesis suggests that perceiving combined depth would yield different neural activity than during the perception of stereo and pictorial depth. Based on the statistical and observed difference between the theoretical and experimental combined depth conditions, specifically the absence of ventral stream activation in combined depth perception, it can be said that combined depth acts like a Gestalt and is not a mere sum of stereo and pictorial depth perception.

Streams. Activity in the dorsal stream was observed during the alternation of combined depth and stereo depth, while activity in both streams was observed during the detection of combined depth and stereo depth. When having a common perceptual task (alternation or detection), the neural activities during combined depth perception were observed to resemble those of stereo depth. This suggests stereo depth plays a prominent role in the percept of combined depth. The fact that neural activity during combined depth perception does not resemble pictorial depth also implies that there were interactions between the dorsal and ventral streams.

Perceptual interactions. Although the results were not significant, interesting trends were seen in the two-tailed t-test results. The combined perception of depth had less activity than pictorial ($p=0.1065$) and stereo ($p=0.1143$) during alternation, but similar activity than pictorial ($p=0.4884$) and stereo depth ($p=0.3901$) during detection. The lack of significance was likely due to the low signal-to-noise ratios of EEG analysis and the low number of subjects.

Based on observed similarities between stream activities in stereo and combined depth in the previous section, stereo depth would be the predominant type of depth perceived in combined depth. Following this idea, pictorial depth information would be the have a supplementary role in combined depth perception, causing different neural activities in stereo and combined depth.

Pictorial depth in alternation. Both alternation of combined depth and stereo depth induced neural activity in the dorsal stream, however combined depth had smaller RPS values despite being the combination of two depth percepts. One possible explanation for this is that, since the alternation of depth is a series of continuous comparisons of depth, the relative depth information in pictorial depth acts as a supplement that facilitates the perception of depth, reducing the intensity of brain activity needed to perceive the alternation of depth.

Pictorial depth in detection. The detection of stereo depth and combined depth both induced neural activity in dorsal and ventral streams. However, the mean RPS for combined depth is similar to both stereo and pictorial depth. Note that the detection stimuli set had the control rectangle as a part of the stimuli, so subjects would have had to quantify depth without a reference. One possible explanation for the similar RPS values is that the brain is unable to utilize relative depth information from drop shadows to supplement stereo depth in quantifying depth. Instead the brain disregards pictorial depth and relies solely on stereo depth, since it provides quantitative depth information. This explains the similar neural activity in combined depth and stereo depth.

Alternation vs. Detection. Results showed that the detection of depth had significantly larger RPS than the alternation of depth ($p=0.0249$). This implies that it is easier for the brain to compare depth than detect depth, which supports the claim that perceiving alternation of depth is a series of comparisons, while detecting depth requires analyzing depth information with no reference point.

Stream interactions. Since pictorial depth and stereo depth are supposedly processed in different visual streams, it is reasonable that these streams would interact with each other in a real life scenario (Goodale, 2014) like combined depth perception, but the interaction between the dorsal and ventral stream varies depending on the condition (Milner, 2017).

In a recent study, Van Dromme, Premereur, Verhoef, Vanduffel & Janssen (2016) suggested that the dorsal stream seemed to have a guiding role over the ventral stream during stereo depth perception in monkeys. However, it is important to note that their stimuli had texture gradient as a pictorial depth cue, so their observations were actually of *combined depth perception*. Their results are consistent with my study, supporting the guiding role of the dorsal stream (stereo depth), which is the predominant type of depth perceived in combined depth. However, my study suggests that the interaction between streams is task-dependent, meaning this relationship does not always hold true, for example, pictorial depth (ventral stream) could have a supplementary role in combined depth perception, thereby lowering neural activity.

Applications

RPS paradigm. By combining the use of a high density EEG net and the SSVEP paradigm, EEG could be used as a means to locate and quantify neural activity with high spatial resolution. When combined with the RPS variable, this method could be used to verify hypotheses for stereo depth perception, observe multimodal interactions or observing stream activity.

Cognitive implications. This study found that alternating and detection stimuli could induce different behaviors, which is important for future research that utilizes the SSVEP paradigm. This study is the first study to look into unimodal perceptual interactions using an EEG, focusing on the dynamics of perceptual interactions. This study supports the two-stream hypothesis and provides insight of stream activities and interactions during depth perception. This study also quantitatively demonstrated the idea of Gestalt perception through combined depth perception.

Real life applications. The findings in this study are important to keep in mind when creating displays that incorporate both stereo and pictorial depth, such as 3D movies or head mounted displays for virtual reality (VR). Since large amounts of neural activity was shown to be a cause of viewer discomfort when viewing stereoscopic displays (Malik et al., 2015), viewing material should be carefully created in order to keep neural activity to a minimum. Based on the discussion of this study, the detection of depth induces more neural activity than the comparison of depth due to the inability to utilize pictorial depth information. Cai et al. (2017) suggested that it is the generation of stereovision that induces visual fatigue, not the sustaining of it, which is similar to the findings of this study since the sustaining of stereovision only requires a series of comparisons. Combining the findings of these studies, I suggest that visual fatigue in stereoscopic displays could be reduced by avoiding images with zero disparity, which would allow the brain to constantly sustain stereovision.

Other studies have demonstrated that simulated blur, a pictorial depth cue, can facilitate binocular fusion and ease visual discomfort. This supports the claim that pictorial cues could have supplementary roles of in combined depth perception (Banks, 2015; Johnson, Parnell, Kim, Saunter, Love & Banks, 2016; Maiello, Chessa, Solari & Bex, 2014; Vinnikov, Allison & Fernandes, 2016).

The perception of depth plays a major role in the immersiveness of stereoscopic displays, since the perception of stereo depth introduces realness in stereoscopic displays (Bowman & McMahan, 2007). By understanding depth perception, the notion of immersiveness and realness could be explored in a more quantitative and detailed manner.

Conclusion

Heretofore unreported in the literature, this study successfully utilized a steady-state visually evoked potential (SSVEP) paradigm to quantify neural activity during combined depth perception by implementing a novel variable. Ventral stream activity was observed during pictorial depth perception and dorsal stream activity during stereo depth perception, suggesting that the two percepts play different roles in analyzing spatial information. This study is the first to quantitatively look into the combined perception of depth, the combination of stereo and pictorial depth perception, which was demonstrated to behave like a Gestalt. In this Gestalt of visual depth, stereo depth was suggested to play a predominant role; pictorial depth had different roles, either supplementary or playing no role at all, depending on the task. This highlights the radically different interactions between the dorsal and ventral streams.

The association between depth percepts and stream activity is an interesting connection that requires further conformational research and studies, as it may lead to further clarification on the functional connectivity of the brain and the processing of visual information. By investigating depth perception, we can decode the process of perceiving depth and realness, allowing us to understand how we reconstruct the visual world we perceive.

Acknowledgements

I would like to thank Dr. Pawan Sinha for his continuous support and giving me the opportunity to run this study in his lab. I would also like to thank Dr. Wasifa Jamal and members of the Sinha lab for supporting and guiding me through my study. Finally, I would like to thank Mrs Stephanie Greenwald, Mr James Gulick, Ms. Megan Salomone and Dr. Caroline Matthew for encouraging me and helping to prepare me for my study.

References

Akay, A. (2009). A brain electrophysiological correlate of depth perception. Neurosciences 2009; Vol. 14 (2).

Backus, B. T., Fleet, D. J., Parker, A. J. & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. Journal of Neurophysiology. 2001 October 4; 86 (4): 2054-2068.

Banks, M. (2015). The Importance of Focus Cues in Stereo 3D Displays. SID Symposium Digest of Technical Papers, 46. doi:10.1002/sdtp.10465.

Bowman, D. A. & McMahan R. P. (2007). Virtual reality: How much immersion is enough? Computer, volume 40, no 7, pp 36-43. July 2007. doi:10.1109/MC.2007.257

Bulthoff, I., Bulthoff, H. & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. Nature Neuroscience, 1, 254–257.

Bulthoff, H. & Mallot, H. (1988). Integration of depth modules: stereo and shading. J Opt Soc Am A. 1988 Oct;5(10):1749-58.

Burle, B., Spieser, L., Rogers, C. L., Casini, L., Hasbroucq, T. & Vidal, F. (2015). Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. International Journal of Psychophysiology, 97(3), pp. 210-220.

Cai, T., Zhu, H., Xu, J., Wu, S., Li, X. & He, S. (2017). Human cortical neural correlates of visual fatigue during binocular depth perception: An fNIRS study. PLoS ONE. 12(2): e0172426. 2017 Feb 16. doi:10.1371/journal.pone.0172426.

Dodgson, N. (2004). Variation and extrema of human interpupillary distance. Proceedings of SPIE, 5291, 36-46.

Gillmeister, H. & Eimer, M. (2007). Tactile enhancement of auditory detection and perceived loudness. Brain Research. Volume 1160, 30 July 2007, Pages 58-68.

Gogel, W. C., & Dasilva, J A. (1987). Familiar size and the theory of off-sized perceptions. Perception & Psychophysics, 41, 318–328.

Goodale M. & Milner A. (1992). Separate visual pathways for perception and action. Trends Neurosci 15:20-25.

Goodale, M.A. (2014). How (and why) the visual control of action differs from visual perception. Proc Biol Sci. 2014 Jun 22; 281(1785): 20140337. doi:10.1098/rspb.2014.0337

Hawkins, R., Houpt, J., Eidels, A. & Townsend, J. (2016). Can two dots form a Gestalt? Measuring emergent features with the capacity coefficient. Vision Research. Volume 126, September 2016, Pages 19-33.

Hebart, M., & Hesselmann, G. (2012). What Visual Information Is Processed in the Human Dorsal Stream? Journal of Neuroscience, 13 June 2012, 32 (24) 8107-8109. doi:10.1523/jneurosci.1462-12.2012

Hidaka, S. & Ide, M. (2015). Sound can suppress visual perception. Scientific Reports 5, Article number: 10483. doi:10.1038/srep10483.

Hogendoorn, H. (2015). From sensation to perception: Using multivariate classification of visual illusions to identify neural correlates of conscious awareness in space and time. Perception 44: 71–78. doi:10.1068/p7832.

Holmes, N., & Spence, C. (2006). Multisensory integration: Space, time, & superadditivity. Current Biology, 15. 2005, pp. 762-764

Hou, C., Pettet, M., Vildavski, V. & Norcia, A. (2006). Neural correlates of shape-from-shading. Vision Research 46, Issues 6–7, March 2006, Pages 1080-1090.

Ide, M. & Hidaka, S. (2013). Tactile stimulation can suppress visual perception. Scientific Reports 3:3453 (2013).

Jäkel, F., Singh, M., Wichmann, F. & Herzoge, M. (2016). An overview of quantitative approaches in Gestalt perception. Vision Res. 2016 Sep, 126, 3-8.

Johnson, P., Parnell, J., Kim, J., Saunter, C., Love, G. & Banks, M. (2016). Dynamic lens and monovision 3D displays to improve viewer comfort. Opt. Express 24, 11808-11827.

Johnston, E., Cumming, B. & Parker, A. (1993). Integration of depth modules: stereopsis and texture. Vision Res.1993 Mar-Apr; 33(5-6):813-26.

Koffka, K. (1935). Principles of Gestalt psychology. London, UK: Routledge.

Levi, D., Knill D. & Bavelier D. (2015). Stereopsis and amblyopia: A mini-review. Vision Research, 114 (2015), pp. 17–30.

Liu, B., Meng, X., Wu, G. & Dang, J. (2013). Correlation between three-dimensional visual depth and N2 component: Evidence from event-related potential study. Neuroscience, Volume 237, pp 161–169.

Maiello, G., Chessa M., Solari F., & Bex P.J. (2014). Simulated disparity and peripheral blur interact during binocular fusion. Journal of Vision, 14(8):13, 1-14.

Malik, A. S., Khairuddin, R., Amin, H. U., Smith, M. L., Kamel, N., Abdullah, J. M., Shim, S. (2015). EEG based evaluation of stereoscopic 3D displays for viewer discomfort. BioMedical Engineering Online, 14, 21.

Mishkin, M., Ungerleider, L., Macko, K. (1983) Object vision and spatial vision: two cortical pathways. Trends Neurosci 6:414-417.

Milner, A. D. (2017). How do the two visual streams interact with each other? Exp Brain Res. 235(5): 1297–1308. Published online 2017 Mar 2. doi:10.1007/s00221-017-4917-4.

Norcia, A., Appelbaum, L., Ales, J., Cottereau, B. & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. Journal of Vision, May 2015, Vol.15(4). doi:10.1167/15.6.4.

Parker, A., Smith, J. & Krug, K. (2016). Neural architectures for stereo vision. Philos Trans R Soc Lond B Biol Sci. 2016 Jun 19; 371(1697).

Rosemann, S., Wefel, I., Elis, V. & Fahle, M. (2016). Audio–visual interaction in visual motion detection: Synchrony versus Asynchrony. J Optom. 2017 Feb 23.

Schiller, P., Slocum, W., Jao, B. & Weiner V. (2011). The integration of disparity, shading and motion parallax cues for depth perception in humans and monkeys. Brain Res. 2011 Mar 4; 1377: 67–77.

Sekuler, R., Sekuler A. & Lau R. (1997). Sound alters visual motion perception. Nature 385, 308 (1997). doi:10.1038/385308a0.

Shams, L., Kamitani, Y. & Shimojo, S. (2000). Visual illusion induced by sound. Cognitive Brain Research, 14, 147-152.

Sinha P. & Poggio T. (1996). Role of learning in three-dimensional form perception. Nature, 384, 460-463.

Van Dromme, I.C., Premereur, E., Verhoef, B.E., Vanduffel, W. & Janssen, P. (2016). Posterior parietal cortex drives inferotemporal activations during three-dimensional object vision. PLoS Biol. 14(4):e1002445. 2016 Apr 15. doi:10.1371/journal.pbio.1002445

Vinnikov, M., Allison, R. S., Fernandes, S. (2016). Impact of depth of field simulation on visual fatigue: Who are impacted? and how? International Journal of Human-Computer Studies, 91 (2016), 37-51.

Wheatstone, C. (1838). On some remarkable, and hitherto unobserved, phenomena of binocular vision. Philosophical Transactions of the Royal Society of London, 33 (1838), pp. 371–394.