A better test leads to a better understanding of reduced stereopsis in amblyopia

Sara Alarcon Carrillo

McGill Vision Research Unit, Department of Ophthalmology

Integrated Program in Neuroscience

McGill University, Montreal

August 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Neuroscience

TABLE OF CONTENTS

CONTRIBUTION OF AUTHORS

Experiment 1 and Experiment 3 will be part of the following manuscript:

Alarcon Carrillo, S., Baldwin, A.S., Yu, M., Zhou, J., Hess, R.F, Amblyopic stereo vision is efficient but noisy (in preparation)

The following is the contribution of individual authors to work presented within this thesis: RFH, ASB, and SAC contributed to the conception of the study. ASB designed stimuli, and SAC edited the stimuli. SAC performed the recruitment of participants in Montreal and Waterloo. MY helped with recruitment and translation for participants in Wenzhou. JZ helped fund data collection in Wenzhou. SAC collected, analyzed all data, and made figures for publishing. SAC wrote the first draft of the manuscript. RFH and ASB reviewed and revised the draft for submission.

ACKNOWLEDGMENTS

This thesis would not have been possible without the endless support from my supervisor, Dr. Robert Hess. Thank you for your guidance, great sense of humour, and all the learning opportunities you granted me throughout my time at the lab. A very special thank you to Dr. Alex Baldwin, working alongside you made this a tremendous learning experience. To Dr. Alexandre Reynaud, thank you for your continued participation and invaluable suggestions throughout my experiments. I would also like to thank Dr. Ben Thomson and Amy Chow at University of Waterloo for their vital help in participant recruitment. From Wenzhou Medical University, I want to thank Mao Yu and Yiya Chen for their patience, enthusiasm, and support during participant recruitment and testing. To Dr. Jiawei Zhou, thank you for your guidance and financial support during my time in China. Finally, I would like to thank my family and friends for all their encouragement and support throughout these two years.

ABSTRACT

People with amblyopia demonstrate a reduced ability to make depth judgements using stereopsis. Insensitive standard clinical stereopsis tests fail to detect residual amblyopic stereopsis, limiting our understanding of their deficit. This study presents a stereo test designed to detect poor stereoacuity. A disparity-defined target is presented within a random-dot display. Participants identify the location of the target while wearing 3D shutter-glasses. We used the equivalent noise method to evaluate the role of equivalent internal noise (signal to noise ratio of disparity signals) and processing efficiency (how efficiently noisy input is processed) in amblyopic stereopsis. We tested 32 amblyopic (4 strabismic, 10 mixed, amblyopic eye visual acuity between 20/200 and 20/14) and 17 control subjects (visual acuity above 20/20). Our tests detected evidence of stereopsis in nearly 70% of amblyopic participants. Mean amblyopic stereoacuity (131 arcsec) was significantly higher than that of controls (64 arcsec) ($t = 3.24$, $p < 0.05$). Using a linear amplifier model, we determined that amblyopic equivalent internal noise was significantly higher than that of controls (239 vs. 135 arcsec, $t = 3.34$, $p < 0.05$). There was no difference in processing efficiency across groups. Multiple linear regression determined 67% of the stereoacuity variance was predicted by processing efficiency and equivalent internal noise, with the latter being the strongest predictor. We find that a more sensitive test for amblyopic stereopsis provides a more accurate picture of the degree of stereo deficiency in amblyopia. This allows a better understanding of what is limiting stereopsis in amblyopia, namely reduced quality of the input entering stereo processing.

RÉSUMÉ

Les personnes souffrant d'amblyopie démontrent une capacité réduite à juger la profondeur en utilisant la stéréoscopie. Les tests stéréoscopiques cliniques existants ne sont pas suffisamment sensibles pour mesurer les fonctions stéréoscopiques résiduelles chez les amblyopes et limitent notre compréhension de ce déficit. Cette étude présente un test de stéréoscopie pour quantifier la stéréoacuité réduite. Une cible en profondeur définie par la disparité est présentée sur un champ de points aléatoires. Les participants identifient l'emplacement de la cible tout en portant des lunettes 3D actives. Nous avons utilisé la méthode de bruit équivalent afin d'évaluer le rôle de l'efficacité de traitement (efficacité avec laquelle un signal bruité est traité) et du bruit interne équivalent (rapport signal/bruit des signaux de disparité) dans la stéréoscopie amblyopique. Nous avons testé 32 amblyopes (4 strabiques, 18 aniso, acuité visuelle de l'œil amblyope supérieure entre 20/200 et 20/14) et 17 adultes témoins (acuité visuelle supérieure à 20/20). Notre test a détecté une stéréoacuité chez près de 70% des participants amblyopes. La stéréoacuité amblyopique moyenne (131 arcsec) était significativement plus élevée que celle des témoins (64 arcsec) ($t = 3,24$, $p < 0,05$). En utilisant un modèle d'amplificateur linéaire, nous avons déterminé que les amblyopes avaient un bruit interne équivalent plus élevé (239 arcsec) que les témoins (135 arcsec) ($t = 3,34$, $p < 0,05$). Il n'y avait aucune différence d'efficacité de traitement entre les groupes. Une régression linéaire multiple a déterminé que 67% de la variance de la stéréoacuité était prédite par l'efficacité de traitement et le bruit interne équivalent, ce dernier étant le prédicteur le plus puissant. Nous avons présenté un test de stéréoscopie suffisamment sensible pour quantifier la stéréoacuité chez les amblyopes et divisé la performance en deux facteurs contributifs. Globalement, la qualité réduite des signaux est un facteur limitant de la stéréoscopie amblyopique.

CHAPTER 1. INTRODUCTION AND BACKGROUND

1.1 Current views on binocular dysfunction in amblyopia

Amblyopia has an incidence rate of 0.2-6.2% in childhood (Mocanu & Horhat, 2018) and 3% in adulthood (Hess et al., 2010). It is a visual disorder that develops following an imbalanced binocular input during childhood and can lead to a permanent loss of visual acuity in the affected eye if untreated. The lack of balanced input from the two eyes can negatively impact the development of normal stereopsis (stereo) in these individuals (see for review, Levi et al., 2015). Stereo is a process in which the brain uses binocular disparity, the shift in the images projected onto the retina from each eye, to calculate the relative depth of objects. In clinical and laboratory settings, stereoacuity is used as a measure of stereoability and is defined as the smallest disparity (magnitude of depth) detectable by an individual. Poor stereo has been linked to gross motor deficits, such as a reduction in the accuracy of reaching and grasping motions (Buckley et al., 2010)). Additionally, it has been reported to lead to an inability to gauge and maintain an appropriate distance from others during social interactions (Smith et al., 2018).

There is a current debate on whether amblyopia causes the reduced binocularity (classical view) or whether the breakdown in binocularity results in amblyopia (more recent suggestion - Birch, 2013). Little is known about the origin of reduced binocularity. A contributing factor may be visual acuity, but it is unable to explain the phenomenon in isolation. It is known is that there is a correlation between stereoacuity and visual acuity in amblyopia that is only present in anisometropic amblyopia, characterized by a significant interocular difference in visual acuity (Levi et al., 2011). This relationship is not found in people with strabismic amblyopia, who experience a physical deviation of the affected and a generally greater deficit in stereopsis (Levi et al., 2015). Additional research is needed to understand the basis of the stereo deficit in

amblyopia. For example, why is it different in the two primary forms of the condition (i.e. anisometropic vs strabismic)? How might it bear on the etiology of the condition? Furthermore, given the recent surge in interest in binocular treatment approaches for amblyopia (Birch et al., 2015; Hess et al., 2010; Webber et al., 2016; Hess & Thompson, 2015), additional research is needed to understand the possibility for recovery of stereo vision in amblyopia.

1.2 Measuring stereopsis in the clinic

A significant obstacle in understanding the underlying factors affecting amblyopic stereopsis is our inability to accurately quantify baseline stereoacuity in amblyopes. Standard clinical tests are relatively insensitive and have a limited dynamic range. This results in participants whose stereoacuity falls outside the narrow testing range often being incorrectly labelled as "stereo-blind" (Zaroff et al., 2003). This term should only be used for people who are unable to detect the depth of any value using binocular disparity. Additional shortcomings render many clinical stereo tests unsuitable to accurately quantify the extent of stereo-stereoability, as well as provide a true measure of stereo-blindness (Chopin et al., 2019; Hess et al., 2019; O'Connor & Tidbury, 2018). Failure in a clinical stereo test, therefore, does not necessarily mean a lack of stereo. Instead, it could be due to the insufficiencies of the stimulus or the testing procedures themselves. A diagnosis of stereoblindness by an insensitive clinical test affords little information about the patient and limits the use of stereo as an endpoint measure in clinical trials. This is unfortunate, as this function would be particularly relevant for treatments primarily directed at restoring and improving binocular function.

In a 2019 paper, Chopin and colleagues reviewed twelve commercial tests available based on their ability to accurately identify poor stereoacuity. They highlighted the limitations of each

test and concluded that the majority of these tests fail to reveal the true ability of the patient and often misclassify true instances of stereoblindness. Test attributes, such as monocular and binocular non-stereo cues, allow participants lacking stereoacuity to pass the test and therefore overestimate the stereoacuity of participants (Chopin, Chan, et al., 2019). Conversely, other test attributes underestimate the stereoacuity. As previously mentioned, most tests have a limited and quantized disparity range, with few testing disparities above 1000 seconds of arc (arcsec). The result is a gross estimate of the subject's ability. Proper assessment of stereo-ability in amblyopic patients requires a test that addresses the limitations of common clinical tests (Chopin, Bavelier, et al., 2019).

1.3 Designing a test for amblyopic stereopsis

Taking these weaknesses into account, we designed a stereo test sensitive enough to quantify very poor stereoacuity in an amblyopic population. The test is based on a random-dot stereogram, a method that does not allow for any monocular cues during testing (Julesz, 1960). Participants are presented with four in-depth wedges defined by disparity. Three of the wedges are going in one direction of disparity, crossed (popping out of the screen) or uncrossed (sunken into the screen), while only one is presented in the opposite direction. The test employs a *modified* depth discrimination task to measure stereoacuity. Unlike traditional depth discrimination tasks, participants do not explicitly state the disparity direction of the target stimulus or identify the stimulus that appears "closer" to them. However, they must still be able to distinguish between crossed and uncrossed disparity to discriminate the "odd-one-out" shape with a different disparity polarity out of the four disparity-defined shapes. Notably, the test measures stereoacuity using a staircase procedure, allowing the test to target a disparity at which the participant can only just

perform the task. Testing for many trials enables us to calculate a measure of variability for each stereoacuity score. Lastly, the test displays an uneven number of crossed and uncrossed stimuli simultaneously on the screen. The literature has shown that individuals can have a selective insensitivity to one disparity direction (Richards, 1970, 1971; Van Ee & Richards, 2002). Therefore, we found it important to present both disparity directions concurrently in such a manner that would allow completion of the task even in the presence of such a selective deficit.

Amblyopic viewing is supported by the test in various ways. First, the test displays disparities from a continuous range up to 3500 arcsec, a ceiling higher than most clinical stereo tests. This disparity range was extended to 9450 arcsec with the use of a larger testing display. We predicted this modification would allow extremely reduced stereoacuity to be detectable. Second, large blurry dots are used in the stimuli as a loss of function at high spatial frequencies has been reported in amblyopia (Holopigian et al., 1986; Levi et al., 2015; Pelli et al., 2004). Lastly, to promote binocularity, the stimuli are aligned to compensate for any eye misalignment and are contrast balanced to reduce any effects of suppression (Levi et al., 2015; Li et al., 2011; Li et al., 2015; Mansouri et al., 2008; Ooi et al., 2013; Webber et al., 2018). Contrast balancing aims to equalize the eye inputs by showing a higher contrast image to the weaker amblyopic eye and a low contrast image to the stronger fellow eye. These test adaptations provide those with amblyopia with a better chance of seeing the in-depth stimulus and completing the task.

1.4 Equivalent Noise Method to understand sensitivity

A test that provides more accurate measures of stereo-ability in the amblyopic population also allows us to pursue more detailed quantitative questions concerning amblyopic stereo processing. Specifically, what limits amblyopic stereo processing? We investigated the role of two

factors using the equivalent noise method. This model allows thresholds to be partitioned into two components, equivalent internal noise and processing efficiency (Barlow, 1956). The former factor describes the quality of the input used by perceptual processes, while the latter details the ability of the system to process this noisy input (Steven C. Dakin et al., 2005; Park et al., 2017; Denis G. Pelli & Farell, 1999). The equivalent noise method was first used by Barlow in 1956 in relation to the quantal threshold required to see light (Barlow, 1956). It has since been used to study human sensitivity to various stimuli, including luminance (Cohn, 1976), contrast (Pelli, 1981), global form (Dakin, 2001), motion (Hess et al., 2006), contour integration (Baldwin et al., 2017), and binocular disparity (Wardle et al., 2012). The use of this model allows for a better understanding of why amblyopes exhibit reduced depth perception in terms of properties and limitations of the visual system (Barlow, 1956)

Previous studies have used this model to look at the role of equivalent internal noise and processing efficiency in normal stereo vision. Wardle et al. used the model to understand the role of these two factors in the reduced stereoacuity for targets in the periphery. They found equivalent internal noise increases with eccentricity while processing efficiency remained relatively constant. Therefore, increased equivalent internal noise was determined to be the limiting factor in peripheral stereoacuity (Wardle et al., 2012). In a previous study, we investigated how each of these two factors contributed to the variability in stereoacuity seen across the healthy adult population (Alarcon Carrillo et al., 2020). We determined that individual differences in processing efficiency and equivalent internal noise both play a role in determining stereoacuity in a healthy population. We used a relative disparity detection task, a modified version of a recently-developed stereoacuity test (Tittes et al., 2019; Webber et al., 2018), which served as the base stimulus design for the current study.

The equivalent noise model has also been used to understand the amblyopic visual system and its limitations. In 2004, Pelli and colleagues found that equivalent internal noise in amblyopic subjects is a factor of 1.4 that of controls in a letter identification task. Further, there was a tendency for equivalent internal noise to increase with the severity of amblyopia (Denis G. Pelli, Levi, et al., 2004). To our knoweldge, the equivalent noise method has not been used to explain amblyopic stereovision. This study investigates the following questions: i) what is the distribution of the stereo anomaly in amblyopia and how common is stereo-blindness in amblyopia? ii) is stereoacuity reduced in amblyopia because of a raised level of internal noise or reduced processing efficiency?

CHAPTER 2. METHODS

2.1 Participants

We tested thirty-two subjects with current or a history of amblyopia. Detailed information of amblyopic participants is found in Table 1. Testing was carried out at three locations: Montreal Canada (15 participants), Waterloo Canada (5 participants) and Wenzhou China (12 participants). Our sample included 4 strabismic, 18 anisometropic and 10 mixed amblyopic participants. Seventeen control subjects were tested, N5, N4, and N16 are authors. All controls had visual acuity that was normal or corrected-to-normal. We used the Freiburg Vision Test (FrACT v3.9.9a; (Bach, 1996) to measure visual acuity. The four-alternative Tumbling E task was used at a distance of 160 cm, making the best possible acuity measure to be 20/10. Acuities for the amblyopic eye ranged from 20/200 to 20/14. All controls had binocular visual acuity above 20/18. All participants provided written informed consent. All testing was performed in accordance with the Declaration of Helsinki and was approved by the Research Ethics Board of the McGill University Health Centre, Wenzhou Medical University and the University of Waterloo Office of Research Ethics.

Table 1. Clinical details of the amblyopic participants (1/3)

ID	Sex	Cyclopleigic Refractive Errors (OD/OS)	Amblyopia Type	Squint	LogMAR visual acuity* (OD/OS)	RPT stereoacuity (arcsec)	Testing location	History before the study
A1	M	-0.50 +3.25	Mixed	XT 6°	-0.14 0.20	N/A	Montreal	Detected at age 10.
A2	M	+1.50 +3.50	Mixed	ET 5°	-0.24 0.06	60	Montreal	Detected at age 4, patched 1 year at age 5.
A3	M	-1.0 +1.50	Mixed	ET 10°	-0.30 0.42	N/A	Montreal	No history of surgery.
A4	F	+0.25 -3.00	Mixed	ET 7°	0.04 0.32	N/A	Montreal	Detected at age ~ 7. Prescribed glasses at 13. No patching history
A5	F	+5.75 -0.5	Mixed	ET 4°	0.47 -0.04	N/A	Waterloo	No history of surgery. Minimal patching during childhood.
A6	F	-0.50 +1.25	Mixed	ET 8°	0.08 -0.01	N/A	Montreal	Detected at age ~12.
A7	M	plano +0.75	Strab	ET 8°	-0.10 1.00	n/A	Montreal	Detected at age 3. No history of patching or surgery.
A8	F	plano plano	Strab	ET 9°	-0.05 0.05	400	Montreal	Detected in childhood, strab surgery adulthood.
A9	M	-2.00 -0.25	Aniso	Ø	-0.18 -0.08	60	Montreal	Prescribed glasses at age 7. No history of patching.
A10	F	+1.75 +1.50	Strab	ET 5°	-0.10 -0.22	N/A	Montreal	Detected at age 5, no history of patching or surgery.
A11	F	plano/-0.25 x 180 +4.50/-1.50x060	Aniso	Ø	-0.28 0.20	N/A	Waterloo	Detected at age 4. Patched until age 8 for around 2-3hours/day.

"Aniso" denotes anisometropia, "mixed" denotes strabismus + anisometropia, "strab" denotes strabismus. RPT: Randot Preschool Test. * measured with the FrACT: Freiburg Visual Acuity Test. ET: esotropia, HT: hypertropia, XT: exotropia.

Table 1. Clinical details of the amblyopic participants (2/3)

ID	Sex	Cycloplegic Refractive Errors (OD/OS)	Amblyopia Type	Squint	LogMAR visual acuity* (OD/OS)	RPT stereoacuity (arcsec)	Testing location	History before the study
A12	F	+0.50 -1.00	Mixed	Ø	-0.23 -0.18	60	Waterloo	Strab surgery at age 10. Intermittent OS suppression - recovered amblyope.
A13	F	+0.25/-0.25x070 +1.50/-1.50x010	Aniso	Ø	-0.08 0.09	400	Montreal	Prescribed glasses at age 7. No history of treatment.
A14	F	0.5/-0.5x180 -2.75 -0.5x180	Aniso	Ø	0.04 0.09	200	Wenzhou	Detected and prescribed glasses at age 9. Intermittent patching (7 years), followed by intermittent bangerter filter (2.5 years) - recovered amblyope
A15	M	0.25/-1.25x70 -0.75/-0.75x180	Aniso	Ø	0.03 -0.18	40	Wenzhou	Detected and had laser surgery at age 18. Intermittent OS suppression.
A16	F	plano +2.25	Aniso	Ø	0.02 0.19	100	Wenzhou	Detected at age 10. No treatment
A17	F	+4.50/-2.00×150 plano/-0.25×150	Aniso	Ø	0.25 -0.17	N/A	Wenzhou	Detected at age 11. Occasional patch for 1 year.
A18	F	-6.00/-3.00×75 -6.00	Mixed	XT 10°	0.16 -0.12	40	Wenzhou	Detected at age 13. Occasional patch for 1 year.
A19	M	+6.50/-1.0×10 -2.50/-0.50×90	Aniso	Ø	0.55 -0.23	N/A	Wenzhou	Detected at 14 years old, wears glasses, no patching
A20	M	+3.50 +5.25	Aniso	Ø	-0.23 -0.08	400	Montreal	Detected at age 5. Minimal OS vision, was patched continuously for a year at age 6.
A21	M	+2.25 Plano/-0.25x90	Aniso	ET 1°	0.46 -0.26	N/A	Montreal	No history of surgery.
A22	M	-5.25 / -0.25x010 plano	Mixed	XT 3°	-0.16 0.96	N/A	Waterloo	Constant OS exotrope, no history of surgery.

"Aniso" denotes anisometropia, "mixed" denotes strabismus + anisometropia, "strab" denotes strabismus. RPT: Randot Preschool Test. * measured with the FrACT: Freiburg Visual Acuity Test. ET: esotropia, HT: hypertropia, XT: exotropia.

Table 1. Clinical details of the amblyopic participants (3/3)

ID	Sex	Cyclopleigic Refractive Errors (OD/OS)	Amblyopia Type	Squint	LogMAR visual acuity* (OD/OS)	RPT stereoacuity (arcsec)	Testing location	History before the study
A23	F	+1.00/ +0.75 x 105 +1.00/ +1.25 x 77	Aniso	N/A	-0.08 -0.21	100	Waterloo	Eye training when young.
A24	F	+6.00/+2.50 x 0 +3.00/-1.00 x 73	Mixed	ET 5°	0.74 -0.11	N/A	Montreal	Detected at age 12, patched for ~6 months.
A25	M	-0.50 +5.00/ -3.00 x180	Aniso	Ø	-0.26 0.49	N/A	Wenzhou	Detected at age 18. No treatment.
A26	M	-14.25/-1.00 x180 -7.75/-0.75 x10	Aniso	Ø	0.28 0.00	800	Wenzhou	Detected at age 18. No treatment.
A27	F	+4.50/-0.75 x 8 plano	Aniso	Ø	0.33 -0.04	N/A	Wenzhou	Detected at age 14. No treatment.
A28	M	-13.25/-2.00 x40 -1.50/-1.75 x7	Aniso	Ø	0.06 -0.22	N/A	Wenzhou	Detected and prescribed glasses at age 10. No patching treatment.
A29	M	+6.25/ -4.25 x40 -1.25	Aniso	Ø	0.82 -0.21	N/A	Wenzhou	Detected at age 5. Occasional patch half a year in childhood.
A30	F	+1.00 +4.00	Aniso	Ø	-0.19 0.80	N/A	Wenzhou	Detected at age 18. No treatment.
A31	M	-2.75 -1.25	Aniso	Ø	-0.03 0.11	100	Montreal	Detected during adulthood. No history of treatment.
A32	F	Plano/-0.75x120 -0.50/-0.50x180	Strab	XT 24°HT 2°	0.06 -0.15	400	Montreal	Patched during childhood. Ptosis surgery at age~7. No history of treatment.

"Aniso" denotes anisometropia, "mixed" denotes strabismus + anisometropia, "strab" denotes strabismus. RPT: Randot Preschool Test. * measured with the FrACT: Freiburg Visual Acuity Test. ET: esotropia, HT: hypertropia, XT: exotropia.

2.2 Apparatus

The stimulus was created on a Eurocom (Ottawa, Canada) P370EM Scorpius laptop with a 17.3" 120Hz display. For Experiments 1 and 3, the stimulus was displayed on the laptop and stereoscopic presentation of the stimulus was achieved with the NVIDIA 3D Vision 2 synchronised shutter glasses. The monitor resolution was 1920 × 1080 pixels with a mean luminance of 197 cd/m^2. The viewing distance for both Experiment 1 and 3 was 55 cm. Experiment 2 used two different setups due to equipment availability at the Waterloo and Montreal testing locations. For subjects A13 and A5, the task was run on the laptop above and dichoptic presentation was achieved using gamma-corrected LG three-dimensional (3D) cinema display (124 3 71 cm; Seoul, South Korea) viewed through the NVIDIA 3D Vision 2 synchronised shutter glasses. The screen luminance was 150 cd/m2 when the stimulus was viewed directly. Tested viewing distances were 60 cm and 90 cm. Subjects A31, A32, N5, N6, and N9 were run on MintBox Mini 2 with the stimulus displayed using an Optoma HD26 1080p 3D DLP Home Theater Projector. Stereoscopic presentation was achieved using the Top/Bottom 3D feature of the projector with viewing through Optoma ZD302 DLP Link Active Shutter 3D Glasses. Viewing distances were 73 cm and 110 cm. Testing was carried out in a dark room with the screen as the only light source. All displays were gamma-corrected with a VIEWPixx X-Rite i1Display Pro (VPixx Technologies, Saint-Bruno, Canada).

2.3 Stimulus displays

The stimulus on the testing laptop was made up of a 28.3 x 28.3 degrees of visual angle (deg) grey square with a 0.25 deg thick black and white border fusion border that encouraged convergence at the plane of the screen. A population of black and white dots populated a 28.0 deg

square region of the grey square. The dots were isotropic log-Gabors with a full width at half-height of 7.5 arcmin. They had a peak spatial frequency of 0.5 cycles/deg with a bandwidth of 2.4 octaves. The grid of dots had a mean dot-to-dot spacing of 26 arcmin. The position of the dots was jittered along the x and y dimensions with a random shift of up to ±11 arcmin. This shift was drawn from a uniform distribution. Disparity was introduced into the stimulus by shifting the horizontal position of the dots by an equal amount in opposite directions in each eye. A nasal shift in the dots was perceived as crossed disparity. A temporal shift in the dots was perceived as an uncrossed disparity. Subpixel interpolation was used to create disparity that could not be resolved in integer pixel shifts. Disparity resulted in a quarter-circle wedge with a radius of 11.8 deg. This disparity-defined shape was introduced to each quadrant in the dot field. Each of the four in-depth wedges was rotated accordingly to make a circle. Three of the wedges appear either in front of (crossed) or behind (uncrossed) the dot field at the fixation plane. The fourth wedge appears simultaneously with an equal magnitude of disparity magnitude but with opposite disparity direction. A red-green anaglyph example of the random-dot stimulus is found in Figure 1. The "odd" target wedge could be placed top-left, top-right, bottom-left or bottom-right of the display, with its quadrant location changing randomly at each trial. Significant horizontal shifts created blank 'voids' at the edges of the wedge stimuli. These voids were filled with dots at the fixation plane to prevent their use as a monocular cue. Stimuli were pregenerated with a quantized list of disparities (log-spaced from 0.125 to 4096 arcsec in steps of 9%). For Experiment 3, stimulus disparity noise was introduced by varying the disparity of the dots about their intended disparity. The disparity of each dot within a stimulus was drawn from a normal distribution of its intended disparity. This distribution had a mean set to the expected disparity. The standard deviation of distribution dictated the level of disparity noise. We tested six different standard deviations: 0, 32, 64, 128, 256, 512 arcsec.

Experiment 2 used the same stimulus code as Experiment 1 but used two different displays, a3D projector and a 3D LED TV. The physical dimensions of the stimulus were scaled according to the display method used. In Experiment 2, participants tested with the projector and with the TV were grouped according to the equivalent resulting dimensions of stimulus retinal image.

CHAPTER 3: PROCEDURE

We measured binocular and monocular visual acuity as well as stereoacuity using the clinical Randot Preschool Test (E. Birch et al., 2008) for each subject at the beginning of the study. We then calculated the contrast ratio for each amblyopic participant using a dichoptic global motion task (Gao et al., 2018; R. F. Hess et al., 2010; Knox et al., 2012; S. L. Li et al., 2015). This coherent motion discrimination task asks participants to identify the direction of coherent motion of a signal dot population presented to one eye, while the other eye sees a noise dot population with no common motion. The target contrast ratio is the one that equalizes the motion coherence threshold (the ratio of signal to noise dots required to detect coherent motion) of both eyes. This contrast ratio was applied to all stimuli in Experiment 1 and 3. A binocular alignment cross was used to correct physical deviations of the eye in strabismic participants during each run. The necessary vertical and horizontal shift for each participant was applied to all dichoptically viewed stimuli.

3.1 General: Odd One Out stereo task

A four-alternate forced-choice task was used to measure thresholds for discriminating the sole target wedge with the "odd-one-out" disparity direction from a group of four disparity-defined wedges. Participants used a keyboard to indicate which quadrant contained the target wedge. Each trial had auditory feedback with correct and incorrect responses denoted by two different beep

tones. The disparity direction of the target (crossed or uncrossed) was randomly selected on each trial. The remaining wedges were assigned an equal magnitude of disparity of opposite direction. Two randomly interleaved staircase routines for each target direction determined the magnitude of disparity in each trial. Each staircase started at 1000 arcsec. A correct response led to an increase in task difficulty, while an incorrect response made the task easier. The step size was a factor of two before the first reversal and was reduced to a factor of $\sqrt{2}$ thereafter. One staircase had a 1-up-1-down rule (converging at approximately 50% correct), and the other had a 1-up-2-down rule (converging at approximately 70% correct). The experiment terminated if all staircases had reached either 50 total trials or 8 reversals. A psychometric function was fit to the combined data between these two staircases to find the 55% threshold. Each run began with at least three training trials with an additional luminance contrast cue on the stimulus that showcased the shape of the wedges and the location of the target. Three training trials had to be answered correctly before the experiment began. The analysis did not include data from the training trials.

3.2 Experiment 1: Baseline amblyopic stereopsis

Experiment 1 used the laptop and tested a total of forty-nine participants (thirty-two amblyopic and seventeen controls) to obtain their threshold for discriminating the odd wedge with differing disparity direction. To make the task easier for amblyopic participants (compared to our previous study (Alarcon Carrillo et al., 2020), the stimulus was presented for a total of 2.75 seconds. The stimulus contrast ramped on and off the display modulated by a temporal envelope of a raised-cosine function. The contrast ramped from zero to 95% contrast over 225 ms. This contrast was maintained over a plateau of 2.25 seconds and returned to zero over 225 ms. Participants could only submit a response once the stimulus had disappeared entirely. This was done to ensure the same viewing time for all participants. Auditory feedback was given after each

trial. A session lasted 6-8 min and a second repetition was measured to assess the reliability of the results. A third repetition was performed for subjects with a large difference between their first two repetitions if time permitted. Analysis was carried out only using the two most recent scores. This was the case for 6 out of 49 participants.

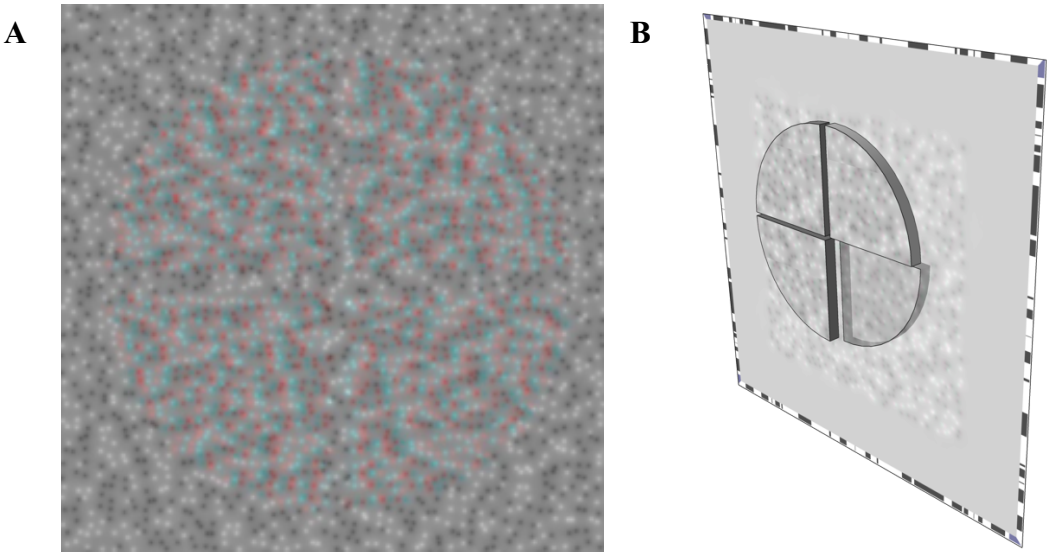

Figure 1. A) Example of the random dot area of the stimulus generated as an anaglyph image (red filter on the right eye for behind target). Depth can be viewed with red-green glasses. This stimulus has a target disparity of 512 arcsec with no disparity noise standard deviation. The target is located in the bottom right quadrant. In the test, stereoscopic presentation of the stimulus was achieved with the NVIDIA 3D Vision 2 synchronised shutter glasses. B) Schematic of the stimulus with the fusion frame.

3.3 Experiment 2: Stereopsis testing at a larger scale

Experiment 2 had the same paradigm as Experiment 1, but the test was displayed on a 3D projector or a 3D TV. Experiment 2 measured stereo thresholds in a subset of participants from Experiment 1 at two different viewing distances from the projector or TV screen. We tested four amblyopic and three control participants (A5, A23, A31, A32, N5, N6, N9) at the first viewing

distance, 73 cm on the projector and 60 cm on the TV. Six of these subjects (A5, A31, A32, N5, N6, N9) were also tested at a second viewing distance, 110 cm for the projector and 90 cm for the TV. Two repetitions were done at each viewing distance. This experimental setup scaled the retinal image of the stimulus and its disparity from those in Experiment 1 (highest tested disparity of 3496 arcsec) by a factor of 2.7 at the first viewing distance (9439 arcsec) and by a factor of 1.9 at the second (6642 arcsec).

3.4 Experiment 3: Equivalent noise analysis in amblyopic stereopsis

Experiment 3 presented stimuli on the same laptop as Experiment 1. This experiment used a noise-masking paradigm that measured a participant's thresholds at six standard deviations of external noise (0, 4, 32, 64, 128, 256, and 512 arcsec). We randomized the testing order for the noise standard deviations, and two repetitions were measured for each condition. The thresholds from Experiment 1 were included in Experiment 3 as the noiseless thresholds (0 arcsec external noise). Eighteen amblyopic participants and all seventeen controls from Experiment 1 were tested in Experiment 3.

CHAPTER 4: ANALYSIS

Data were fit with a logistic psychometric function using Palamedes (Prins & Kingdom, 2009) to obtain stereothresholds for discriminating a wedge with the odd-one-out disparity direction. The calculated threshold had a correct response probability of 55.2%, corresponding to a signal-to-noise ratio (d') of unity at the decision variable. Data from each repetition were fit separately. We used parametric bootstrapping to calculate the standard error for each threshold. For each psychometric function fit, 500 bootstrapped threshold estimates were generated. The first

repetition measured in Experiment 1 was used as the measure of stereoacuity for each participant. The second repetition in Experiment 1 was used in Experiment 3 for the equivalent noise model analysis. This was done to ensure independence between the noise masking factors and thresholds for analysis. There were subjects whose psychometric fit did not yield a stable threshold (standard error above 3 $\log_2$ units). These thresholds were deemed unreliable and were excluded from analysis. After this exclusion, participants who had only one reliable threshold measure in Experiment 1 (threshold from one repetition was eliminated) were included in analysis in Experiment 1 but were excluded from analysis in Experiment 3. This occurred for 10 amblyopic participants. Participants who had only one reliable threshold measure in Experiment 1 were included in analysis in Experiment 1 but excluded from analysis in Experiment 3. This occurred for 1 of 18 amblyopic participants and 1 of 17 control subjects that participated in Experiment 3. Finally, for subjects who were continuously unable to perceive the stimulus at the highest disparity, the task was terminated, and the subject was deemed unable to complete the task. This was the case for 4 amblyopic subjects.

In Experiment 2 the disparity magnitudes shown throughout the staircases first had to be scaled with the corresponding scaling factor at each viewing distance before fitting. We performed a likelihood ratio test in Palamedes (PAL_PFLR_ModelComparison function) to determine if data from Experiment 1 and Experiment 2 had the same threshold. This analysis determines if separate datasets from two different conditions (Experiment 1 and Experiment 2) are best fit by two psychometric functions or one psychometric function (Kingdom & Prins, 2016). The former reveals that the thresholds across the two conditions are significantly different, while the latter points to equal performance across conditions. The likelihood ratio test calculates the P-values of the comparisons by running Monte Carlo simulations (2000 samples). There were cases where a

P-value could not be calculated using this method due to unstable psychometric function fits. In these cases, we calculated approximate estimates of the P-values using the χ^2 survival function in the SciPy Python package (Oliphant, 2007). The likelihood ratio test was also used to compare performance at two different viewing distances on the larger display.

4.1 Equivalent noise model analysis

In Experiment 3, we fit the thresholds measured at each level of noise standard deviation with the linear amplifier model to calculate equivalent internal noise and processing efficiency. The model is the following

$$\text{threshold} = \frac{\sqrt{\sigma_{\text{external}}^2 + \sigma_{\text{eq.int.}}^2}}{\beta}, \quad (1)$$

where σ_{external} is the standard deviation of the disparity noise added to the stimulus, $\sigma_{\text{eq.int.}}$ is the equivalent internal noise in the visual system of the subject, and β is a measure of how efficiently the subject can integrate the noisy information to perform the task. We performed fitting using only the second repetition of noiseless thresholds from Experiment 1. For the remaining noise levels, we used the merged thresholds across both repetitions. Equivalent internal noise and processing efficiency are calculated according to the values of the parameters that result in the best fitting curve to the threshold vs. external noise data (plotted on log-log axes). Changes in efficiency result in a vertical translation of the curve. Changes in the equivalent internal noise cause a horizontal shift in the transition point between the flat and sloped region of the curve. We completed the fitting using the fminsearch function in MATLAB. This function minimizes the root-mean-square error between the model and the data on a $\log_2$-threshold axis. We used the samples generated in the parametric bootstrapping step of the psychometric function fitting to calculate the bootstrapped estimates of the standard error for the model parameters.

CHAPTER 5. RESULTS

5.1 Experiment 1

Thirty-two amblyopic subjects were tested with the Odd One Out stereo test and the Randot Preschool Test to measure stereothresholds. Eighteen amblyopic subjects had measurable stereothresholds with the Odd One Out test, while fourteen were unable to complete the test or had unreliable thresholds ($\log_2 SE > 3$). Amblyopes' stereo thresholds ranged from 54 to 564 arcsec, while controls ranged from 22 to 148 arcsec. The correlation of thresholds between the two task repetitions on the Odd One Out task can be seen in Figure 2. Both amblyopes ($R = 0.73$, $p < 0.01$) and controls ($R = 0.55$, $p < 0.05$) saw good test-retest reliability. We found a significant correlation between thresholds and the visual acuity of the amblyopic eye ($R = - 0.53$, $p < 0.025$).

14 of 32 amblyopic participants had measurable stereoacuity with the Randot Preschool test. All Randot Preschool test scores can be found in Table 1. Of the eighteen amblyopes that could not complete the Randot Preschool Test, six had measurable stereoacuity with the Odd One Out task. On average, the stereoacuity scores from the Randot Preschool were a factor of two higher than those from the Odd One Out task. 70% of subjects had lower stereothresholds with the Odd One Out task compared to their scores with the Randot Preschool test.

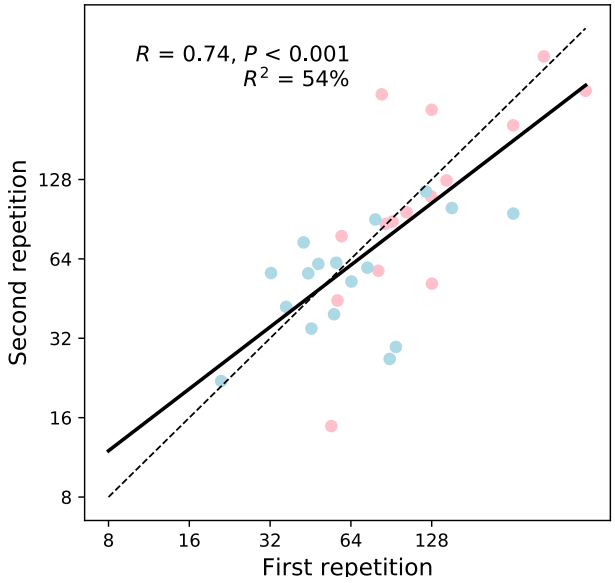

Figure 2. Test-retest correlation of the thresholds from the first and second repetition in Experiment 1. Control and amblyopic subjects are shown in blue and red dots respectively. The line of best fit (bolded) to the combined data across groups has an equation of $\log_2(y) = 0.78 \log_2(x) + 1.24$ with an R^2 of 54%. The dashed line is the unity line.

The distributions of stereothresholds measured with the Odd One Out task for amblyopic subjects and controls are shown in Figure 3. Mean amblyopic (M = 131 arcsec, SD = 1.0) and control controls (M = 62 arcsec, SD = 0.87) thresholds were significantly different (Welch's t-test in SciPy, t (32) = 3.24, p =0.003). We calculated the percentage of correct responses across all trials for participants whose psychometric fit did not yield a stable threshold in either repetition (n = 10). 95% confidence intervals of the performance measure were calculated for each participant and were used to determine if performance was significantly above chance (25%). The results can be found in figure 4. Three subjects (A11, A23, 25) without a stable threshold performed above chance and therefore display evidence of residual stereo-ability. The remaining subjects (A5, A7, A10, A21, A22, A24, A29) did not perform significantly above chance for the disparity range investigated.

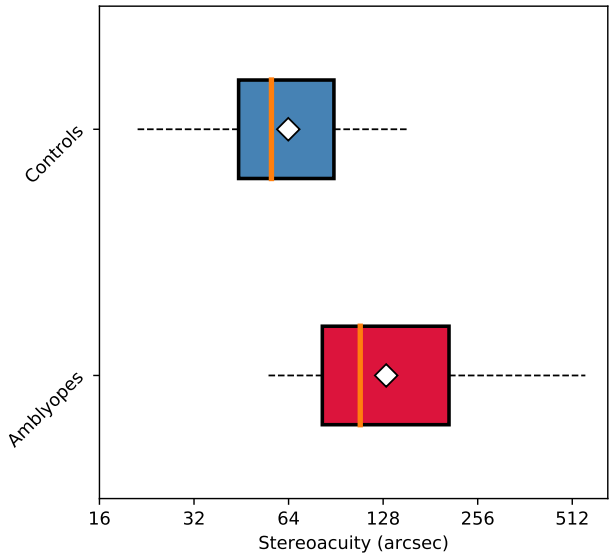

Figure 3. Boxplot distributions of stereothresholds measured in Experiment 1. Amblyopic data are shown in red (M = 131 arcsec, SD = 1.0, n =18) and controls are in blue (M = 64 arcsec, SD = 0.86, n =17). Dashed lines denote the range of thresholds. Median and mean thresholds are denoted by an orange line and white diamond respectively. The mean threshold was significantly different across groups (Welch's t-test in SciPy, t (32) = 3.24, p =0.003).

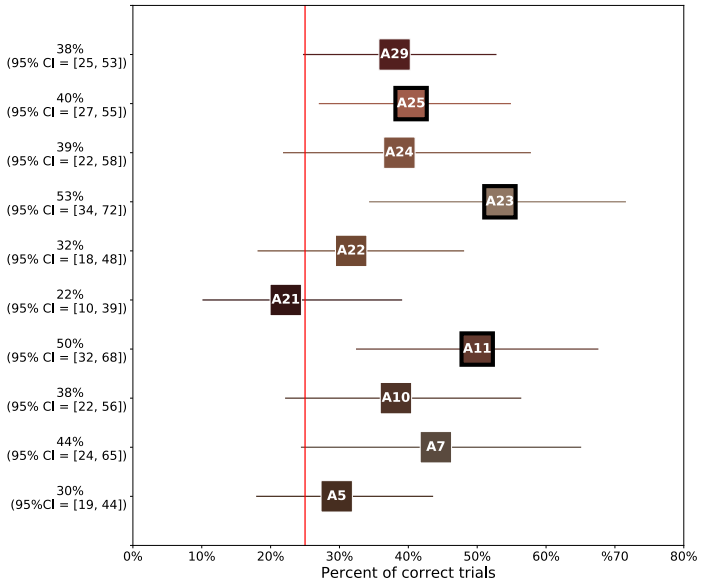

Figure 4. Measure of overall performance for subjects without a stable threshold in Experiment 1. Markers plot the percentage of correct responses from all trials for each subject. Error lines represent the 95% confidence intervals. The red vertical line is the chance level for the Odd One Out task (25%). Subjects with performance above chance are bolded.

5.2 Experiment 2

Experiment 2 measured stereoacuity using the Odd One Out task on a larger display. The first viewing distance granted access to disparities as high as 9430 arcsec. We tested four amblyopic subjects and three control subjects from Experiment 1. In Experiment 1, subjects A5 and A23 did not have a stable threshold and were excluded from analysis. Subjects A31 and A32 had a stable threshold and were included in the sample mean calculations. In Experiment 2, subjects A5, A31, A32, N5, N6, and N9 had measurable stereoacuity. Subject A23 remained unable to complete the task. Table 2 shows thresholds from Experiments 1 and 2, as well as the results from the likelihood ratio test comparing performance across experiments. Except for N6, all subjects performed equally across conditions.

ID	Experiment 1 thresholds (arcsec)	Experiment 2 thresholds (arcsec)	TLR	P
A5		2400 (SE = 0.59)		
A31	564 (SE = 0.92)	133 (SE = 0.21)	0.56	0.45*
A32	234 (SE = 1.1)	157 (SE = 0.34)	1.4	0.25
N5	46 (SE =0.026)	28 (SE = 0.68)	0.57	0.45*
N6	32 (SE = 0.23)	57(SE = 0.26)	5.4	**0.026**
N9	79 (SE = 0.55)	55(SE = 0.48)	0.98	0.31

Table 2. Experiment 1 and 2 stereothresholds for subjects tested in both experiments. Subject A23 could not complete either experiment. We performed the likelihood ratio test (TLR = transformed likelihood ratio) and compared data across the two experiments using a P-values are calculated from Monte Carlo simulations. We calculated P-values calculated by the χ^2 approximation in cases where Palamedes was unable to generate a P-value from simulations. An asterisk marks these subjects. Subjects with significant differences are bolded.

The stimulus in Experiment 2 (peak = 0.18 cycles/ degree) had a lower spatial frequency than Experiment 1 (peak = 0.5 cycles/degree). We wanted to investigate what effect this change in spatial frequency had on task performance. To do this, we compared performance for six subjects (A5, A31, A32, N5, N6, N9) across two viewing distances on the larger display. The first distance (detailed above) had a stimulus with a peak spatial frequency of 0.18 cycles/degree. The second distance displayed at 0.26 cycles/degree. Table 3 shows the likelihood ratio test used to compare performance across the two spatial frequencies. We found subject A5 and N6 performed significantly better when the test had a higher peak spatial frequency (farther viewing distance). All other participants performed equally across the two conditions.

Subject	Thresholds at 0.18 cycles/degree (arcsec)	Thresholds at 0.26 cycles/degree (arcsec)	TLR	P
A5	2400 (SE = 0.59)	132 (SE = 0.27)	10	**0.0013***
A31	133 (SE = 0.21)	68 (SE = 0.30)	4.3	0.066
A32	157 (SE = 0.34)	98 (SE = 0.33)	2.9	0.092
N5	28 (SE = 0.68)	22 (SE = 0.018)	0.11	0.74*
N6	57(SE = 0.26)	32 (SE = 0.20)	6.6	**0.012**
N9	55(SE = 0.48)	39 (SE = 0.40S)	0.83	0.36

Table 3. Stereothresholds measured in Experiment 2 at two stimulus peak spatial frequencies (PSF). We performed the likelihood ratio test (TLR = transformed likelihood ratio) and compared data across the two experiments using a P-values are calculated from Monte Carlo simulations. We calculated P-values calculated by the χ^2 approximation in cases where Palamedes was unable to generate a P-value from simulations. An asterisk marks these subjects. Subjects with significant differences are bolded.

5.3 Experiment 3

Nineteen amblyopic (8 mixed, 3 strabismic, 8 anisometropic) and seventeen control subjects from Experiment 1 were tested in Experiment 2. Subjects without a stable noiseless threshold in Experiment 1 were excluded from analysis. This was the case for three amblyopic subjects (A5, A7, A10). Subjects with only one repetition yielding a stable noiseless threshold in Experiment 1 (standard error below 3 $\log_2$ units) were excluded from analysis as parameters independent from their threshold performance could not be computed. This was the case for one amblyopic subject (A20) and one control subject (N10). Parameter values with a standard error higher than 3 $\log_2$ units were also excluded from analysis. This was the case for one amblyopic subject (A4). These criteria resulted in two strabismic, one anisometropic and two mixed subjects being excluded. Noise-masking functions for the remaining fourteen amblyopic subjects and sixteen control are presented in Figures 5 and 6, respectively. Thresholds are plotted as a function of the six tested levels of external noise standard deviation. All subjects demonstrated the expected noise-masking behaviour, with performance staying constant at lower external noise levels and increasing linearly with external noise beyond a critical value. The linear amplifier model was fit to the threshold data (using the fminsearch function in MATLAB) to find equivalent internal noise and processing efficiency for each participant. The average noise masking functions for both groups can be found in Figure 5A. On average, equivalent internal noise was significantly higher in amblyopic subjects (M = 239.37 arcsec, SD = 0.78) compared to controls (M = 134.6 arcsec, SD= 0.48): Welch's t-test in SciPy, $t(21) = 3.34$, $p = 0.003$. The two groups did not differ in processing efficiency ($p > 0.05$).

We wanted to determine the role of processing efficiency and equivalent internal noise in stereothresholds differences in amblyopia. We performed simple and multiple linear regression

(using the ordinary least squares function from the statsmodels Python package (Seabold & Perktold, 2010)) to analyse the relationship between Experiment 1 thresholds and the model parameters calculated in Experiment 3. There is no inherent correlation between the variables as Experiment 1 used thresholds from the first repetition and LAM fitting used thresholds from the second repetition. These two thresholds are significantly correlated (Figure 3). The analysis showed that stereothresholds were significantly correlated with equivalent internal noise in amblyopic participants (Figure 7A). A significant linear equation was found ($F(1,12) = 6.77$, $p = 0.023$) with a coefficient of determination $R^2 = 36\%$. Amblyopic stereothresholds can be predicted from equivalent internal noise by equation (1). Processing efficiency was not predictive of amblyopic thresholds ($p > 0.05$) (Figure 7B). We found a significant multiple linear regression equation ($F(2,11) = 10.96$, $p = 0.002$) with a coefficient of determination (R^2) was 67%. The relationship is described by equation (2). Both predictors were significant (equivalent internal noise: $t(13) = 3.98$, $p = 0.002$; efficiency: $t(13) = -3.171$, $p = 0.009$).

For control participants, processing efficiency was a significant predictor of thresholds (Figure 7D). The relationship is described by the significant ($F(1,14) = 6.86$, $p = 0.020$, $R^2 = 33\%$) line equation (3). Equivalent internal noise was not a significant predictor for thresholds ($p > 0.05$) (Figure 7C). We found a significant multiple linear regression equation ($F(2,13) = 24.50$, $p < 0.001$) with an R^2 of 79%, presented as equation (4).

5.3.1 Regression equations

(1) $\text{Log2 threshold} = 0.67 \times \log2\left(\sigma_{eq}\right) + 1.5.$

(2) $\text{Log2 threshold} = 1.62 + 0.79 \times \log2\left(\sigma_{eq}\right) - 0.82 \times \log2\left(\beta_{rel}\right)$

(3) $\text{Log2 threshold} = -0.80 \times \log2\left(\beta_{rel}\right) + 7.1.$

(4) $\text{Log2 threshold} = -1.89 + 1.34 \times \log2\left(\sigma_{eq}\right) - 1.20 \times \log2\left(\beta_{rel}\right).$

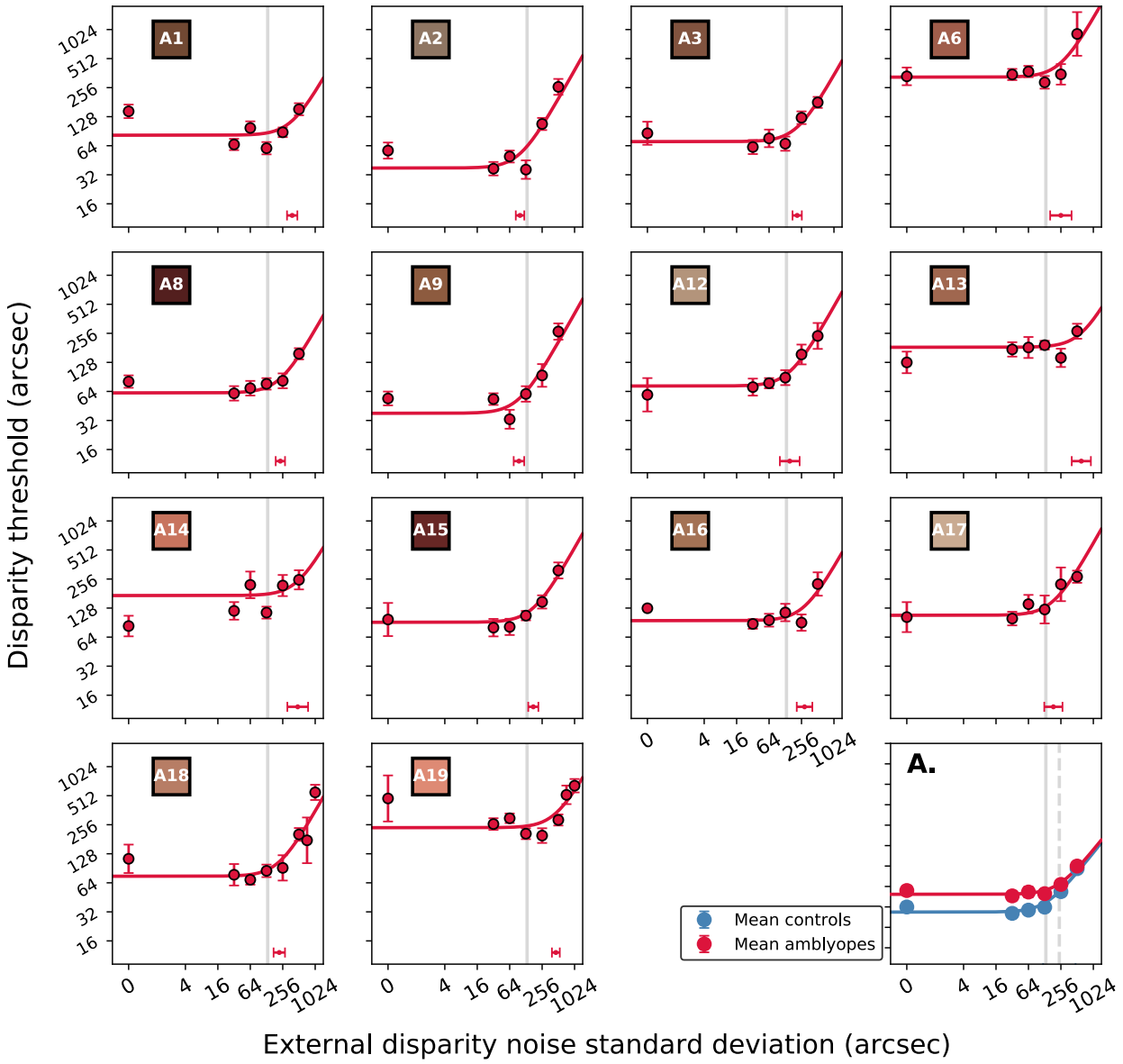

Figure 5. Noise masking functions for all amblyopic participants. Each panel includes the subject's number in a coloured square. Thresholds (d′ = 1) are plotted (with bootstrapped standard error) and fit with the linear amplifier model. The horizontal error bars show the equivalent internal noise calculated for each subject. The vertical grey line indicates the mean equivalent internal noise for controls (135 arcsec). A) Mean noise masking functions for controls (blue) and amblyopes (red). The dashed vertical line indicates the mean equivalent internal noise for amblyopes (239 arcsec).

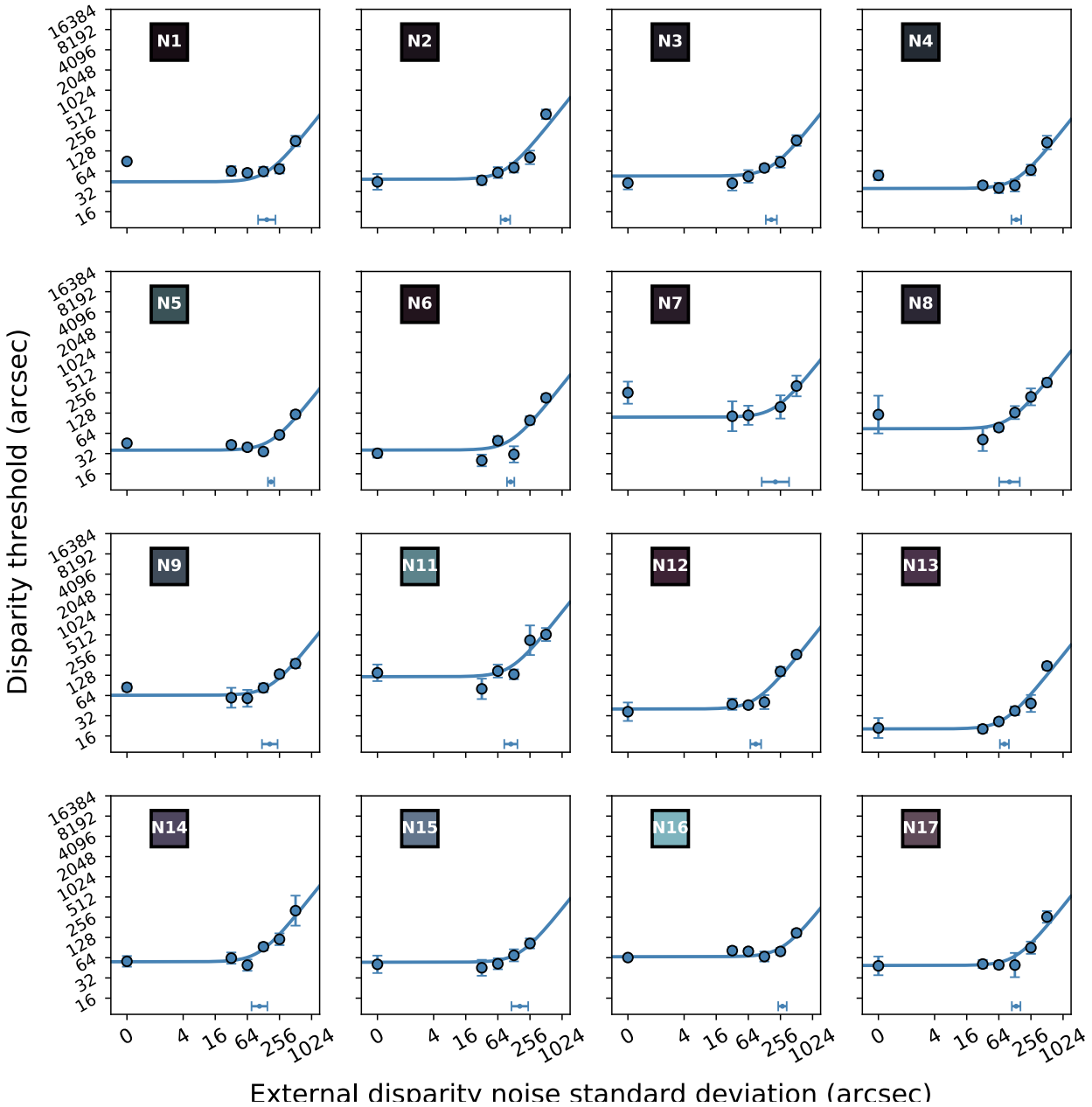

Figure 6. Noise masking functions for all control participants in Experiment 3. Each panel includes the subject's number in a coloured square. Thresholds (d′ = 1) are plotted (with bootstrapped standard error) and fit with the linear amplifier model. The horizontal error bars show the equivalent internal noise calculated for each subject.

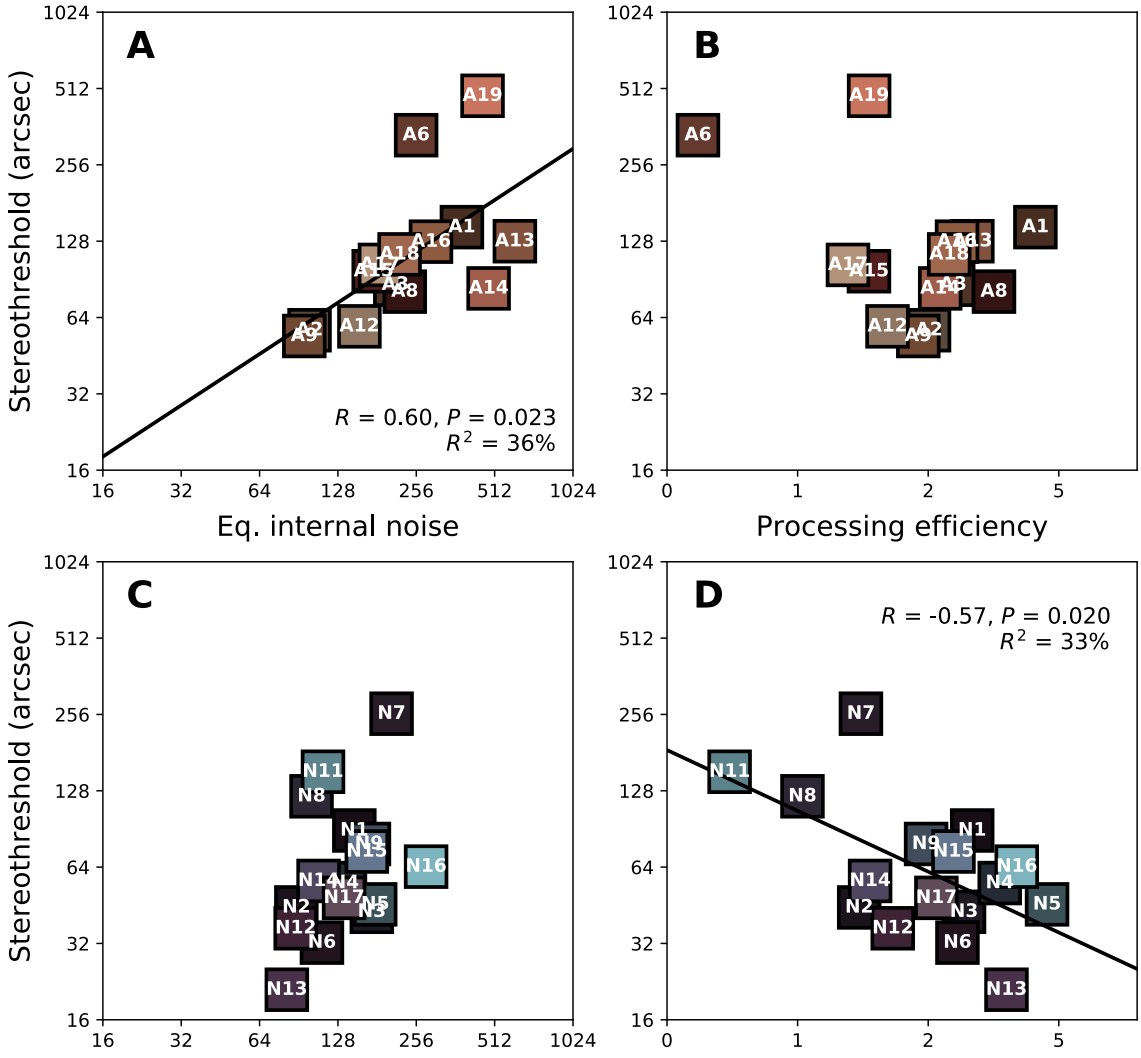

Figure 7. Correlations between parameters from Experiment 3 and stereothresholds from Experiment 1. The relationship between amblyopic thresholds and equivalent internal noise is shown in panel A and with processing efficiency in panel B . Panel C and D results for control subjects. The bolded line of best fit denotes a significant correlation.

CHAPTER 6: DISCUSSION

People with amblyopia are often reported to have a high incidence of diminished or absent stereopsis. However, the common clinical stereo tests used as a quick measure of stereoacuity are limited in their ability to quantify poor stereopsis. To better assess amblyopic stereo-ability, we developed a modified depth discrimination task designed to measure reduced stereopsis. We tested thirty-two amblyopic adults and were able to obtain stereoacuity for eighteen of these participants. The stereo-ability of six of these eighteen participants failed to be detected by the clinical Randot Preschool Test. This means that close to 20% of our amblyopic sample was erroneously classified as "stereoblind" by this standard clinical test. A critical attribute of our stereotest was the ability to test greater disparities than the standard clinical tests. The highest disparity displayed in our test was a factor of four higher than that of the clinical test. Only 11 patients out of our sample of 32 failed to show above chance performance at our original largest disparity (3496 secs).

Interestingly, three participants (A11, A23 and A25) performed above chance despite their data failing to yield a computable stereothreshold. This suggested that their stereo-ability that could be measured with an extended disparity range. When we tested disparities up to 9430 arcsec using a larger display, we were able to quantify the stereoacuity of an additional participant. Stereoacuity of this subject fell just outside our original disparity range. They reported perceiving depth and "layers of depth" (disparity noise) for the first time in a laboratory setting. Our findings demonstrate that residual stereoacuity in amblyopia is missed by standard clinical stereo tests but can be accessed with a more sensitive stereo test. In recent clinical trials (Holmes et al., 2016; Manh et al., 2018), up to 50% of amblyopic participants are classified as stereo-blind using standard clinical tests. This is one reason why stereopsis cannot be used as a sensitive endpoint

measure. In our study, only 31% of the amblyopic participants failed to demonstrate any stereo ability.

Access to baseline amblyopic stereoacuity allowed us to quantitatively assess their stereo-deficiency. We found that amblyopic stereoacuity was twice as poor as that of our healthy sample. Although this does demonstrate a stereo deficiency in amblyopia, this deficit was surprisingly mild. We found that there was considerable overlap between the control and amblyopic distributions (Figure 3). This implies that a more accurate and sensitive stereo test is required in order to detect both initial measures of stereoacuity as well as any improvements post-treatment. It is useful in this case to have an associated estimate of measurement variability. Current clinical stereo tests do not provide this as they only use single-trial testing at a few fixed disparities.

Using the equivalent noise method to calculate each factor, we found an increased level of equivalent internal noise in amblyopic participants compared to controls. Further, individual differences in stereoacuity were also found to be predicted by variations in equivalent internal noise. This limiting relationship between equivalent internal noise and task performance has been reported in previous studies on the amblyopic visual system, including contrast sensitivity (Kiorpes et al., 1999), position discrimination and detection (Dennis M. Levi, 2013; Dennis M. Levi & Klein, 2003), and global motion and orientation (Behzad Mansouri & Hess, 2006). There is evidence suggesting higher noise in the amblyopic system has a cortical origin (Kiorpes et al., 1999; Smith et al., 1997; Wang et al., 2017). Higher amblyopic noise may be related to the absence of appropriate binocular visual experiences (Dennis M. Levi et al., 2008) and the presence of binocular suppression (Wang et al., 2017) throughout development.

Our study found that processing efficiency does not play a significant role in amblyopic stereopsis. This was an unexpected finding as we had previously found that diminished stereoability in healthy adults was correlated with both an increase in internal noise and reduced processing efficiency (Alarcon Carrillo et al., 2020). Furthermore, studies have pointed to processing efficiency playing a role in other deficits seen in the amblyopic visual system. For example, plenty of literature demonstrates a decrease in processing efficiency contributes to reduced contrast sensitivity in amblyopia (Huang et al., 2007; Dennis M. Levi et al., 2008; Dennis M. Levi, 2013; Denis G. Pelli, Levin, et al., 2004). We, however, found that amblyopic subjects processed stereo information as efficiently as healthy participants. Amblyopic stereoacuity being limited by equivalent internal noise while being independent of processing efficiency variations mirrors the findings of stereoacuity in the periphery of healthy adults (Wardle et al., 2012). Our findings suggest that equivalent internal noise, and not processing efficiency, is a significant limiting factor in amblyopic stereopsis. We found no difference in stereoacuity or noise masking parameters across the different categories of amblyopia listed in table 1, although with such small numbers, our results are not definitive.

It was interesting to find that majority of amblyopic stereoacuity fell within the disparity range of the clinical stereo test. Moreover, some individuals able to perceive depth only with our test had stereothresholds low enough to be detectable by the clinical test. Therefore, the increased sensitivity of our stereo test cannot be solely attributed to an increase in disparity range. The Randot Preschool Test is composed of very sharp small dots, a stark difference to the large blurry dots found in our task. *Dmin* (smallest disparity detectable by an individual) has been shown to be dependent on the highest spatial frequency in a stimulus, improving with an increase in frequency in healthy adults (Hess et al., 2002). However, individuals with amblyopia are known to experience

binocular dysfunction at high spatial frequencies, such as reduced contrast sensitivity (Bradley & Freeman, 1981; Harrad & Hess, 1992; Levi et al., 1979) and diminished stereopsis (Holopigian et al., 1986; Levi et al., 2015). We applied a bandpass filter to our random dot stimulus to prevent an advantage for healthy participants. The importance of the spatial frequency composition of a stereo stimulus was highlighted by the performance of amblyopic subject A5 across conditions. This individual lacked stereoability when tested with a higher frequency stimulus (Experiment 1) but was able to perceive depth with lower spatial frequency stimulus (Experiment 2). Their stereoacuity showed significant improvement with an increase in the peak stimulus spatial frequency that remained within their bandpass. This suggests the amblyopic stereothresholds still demonstrate a dependency on the highest spatial frequency of a stimulus so long as it is within the bandpass of the system. Although our sample size was limited and further research is needed, these results highlight the importance of lowpass or bandpass filtered stereotests when measuring amblyopic stereopsis.

This study introduced a new design for a stereo test suitable for measuring stereoacuity in subjects who may test outside the normal range. We combined this with the equivalent noise method to further investigate the range of results we found. Importantly, we found that the absence of stereoacuity in amblyopia is not always due to neural deficits but rather due to limitations in standard tests. The stereoability in a larger part of the amblyopic population can be detected with an appropriate stereo test. Our findings highlight the importance of having a stimulus with a wide disparity range when measuring stereo ability in this population. Special consideration should also be placed on the spatial frequency component of the test, with lower spatial frequencies proving most beneficial. Obtaining accurate baseline stereoacuity provided a better idea of the extent of stereoblindness in amblyopia and allowed us to assess the underlying neural basis. We determined

that nosier signals reaching the stereo processing system, leading to lower input quality, is a limiting factor in amblyopic stereopsis.

6.1 Future directions

In future studies we would like to collect more stereoacuity data using the larger display. First, we would like to get a normative dataset using the projector to determine how results compare to measures taken with other clinical and laboratory stereo tests. Following this assessment, we would like to increase our sample size of amblyopic participants. Larger, more even samples of anisometropic, strabismic and mixed amblyopic subjects, would allow a proper assessment of differences between these groups. We would also want to run further testing on subjects of interest in our current cohort that we were unable to test due to external circumstances. Of particular interest are those who lacked stereoability under the conditions in Experiment 1. We want to determine if the stereoability of a larger percentage of "stereoblind" amblyopic participants is detectable with a larger disparity range, as was the case with the subject presented in the current study. Additional testing at different viewing distances while maintaining a constant stimulus size will also allow us to further investigate the effect of changing the spatial frequency in our bandpass filtered random dot stimulus.

Multiple amblyopic subjects claimed to have completed the task by perceiving an unidentifiable difference between the target and the remaining wedges unrelated to their direction of depth. Therefore, detecting stereoability in these individuals was possible due to our "modified" 4AFC depth discrimination task. This paradigm does not require explicit naming of the disparity direction of the target, but participants must still be able to discriminate the difference between the two percepts to determine the odd wedge out. Stereoblindness, originally defined as an inability

to discriminate between the two disparity directions, was first found and is most prevalent in studies using a traditional depth discrimination task. It may be possible that people with reduced or absent stereopsis still perceive crossed and uncrossed disparity, but not as "popping out" and "sunken in". This would cause a failure to complete a depth discrimination task, but not necessarily a deficit on our task. A logical next step to investigate this theory would be to test those amblyopic subjects using the Odd One Out task but instead ask them to state the disparity direction of the target. We predict this would lead to the rate of stereoblindness being more similar to the one found in the literature (Dorman & van Ee, 2017; Richards, 1970, 1971).

The last aim of this study was to examine the role of equivalent internal noise and processing efficiency in amblyopic stereovision. The noise-masking parameters were collected only from those amblyopic subjects who had measurable noiseless thresholds when tested on the laptop. This means that those with extremely reduced ability outside the range of this paradigm were excluded from the analysis. We would like to carry out a study using the Equivalent Noise Method on the projector in the future. This would allow access to measures of equivalent noise and processing efficiency in very poor performers. A study like this would further investigate the relationship between equivalent internal noise and amblyopic stereoacuity found in this study.

REFERENCES

Alarcon Carrillo, S., Baldwin, A. S., & Hess, R. F. (2020). Factors limiting sensitivity to binocular disparity in human vision: Evidence from a noise-masking approach. *Journal of Vision*, *20*(3), 9. https://doi.org/10.1167/jov.20.3.9

Bach, M. (1996). The Freiburg visual acuity test - automatic measurement of visual acuity. *Optometry and Vision Science*, *73*(1), 49–53. https://doi.org/10.1097/00006324-199601000-00008

Baldwin, A. S., Fu, M., Farivar, R., & Hess, R. F. (2017). The equivalent internal orientation and position noise for contour integration. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-13244-z

Barlow, H. B. (1956). Retinal Noise and Absolute Threshold. *Journal of the Optical Society of America*. https://doi.org/10.1364/JOSA.46.000634

Birch, E. E. (2013). Amblyopia and binocular vision. In *Progress in Retinal and Eye Research*. https://doi.org/10.1016/j.preteyeres.2012.11.001

Birch, E. E., Li, S. L., Jost, R. M., Morale, S. E., De La Cruz, A., Stager, D., Dao, L., & Stager, D. R. (2015). Binocular iPad treatment for amblyopia in preschool children. *Journal of AAPOS*. https://doi.org/10.1016/j.jaapos.2014.09.009

Birch, E., Williams, C., Drover, J., Fu, V., Cheng, C., Northstone, K., Courage, M., & Adams, R. (2008). Randot®Preschool Stereoacuity Test: Normative data and validity. *Journal of AAPOS*. https://doi.org/10.1016/j.jaapos.2007.06.003

Bradley, A., & Freeman, R. D. (1981). Contrast sensitivity in anisometropic amblyopia. *Investigative Ophthalmology and Visual Science*.

Buckley, J. G., Panesar, G. K., MacLellan, M. J., Pacey, I. E., & Barrett, B. T. (2010). Changes

to control of adaptive gait in individuals with long-standing reduced stereoacuity. *Investigative Ophthalmology and Visual Science*. https://doi.org/10.1167/iovs.09-3858

Chopin, A., Bavelier, D., & Levi, D. M. (2019). The prevalence and diagnosis of 'stereoblindness' in adults less than 60 years of age: a best evidence synthesis. In *Ophthalmic and Physiological Optics*. https://doi.org/10.1111/opo.12607

Chopin, A., Chan, S. W., Guellai, B., Bavelier, D., & Levi, D. M. (2019). Binocular non-stereoscopic cues can deceive clinical tests of stereopsis. *Scientific Reports*. https://doi.org/10.1038/s41598-019-42149-2

Cohn, T. E. (1976). Detectability of a luminance increment: effect of superimposed random luminance fluctuation. *Journal of the Optical Society of America A*, *66*(12), 1426–1428.

Dakin, S C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America A*, *18*(5), 1016–1026. https://doi.org/10.1364/JOSAA.18.001016

Dakin, Steven C., Mareschal, I., & Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Research*. https://doi.org/10.1016/j.visres.2005.07.037

Dorman, R., & van Ee, R. (2017). 50 Years of Stereoblindness: Reconciliation of a Continuum of Disparity Detectors With Blindness for Disparity in Near or Far Depth. *I-Perception*, *8*(6), 1–13. https://doi.org/10.1177/2041669517738542

Gao, T. Y., Guo, C. X., Babu, R. J., Black, J. M., Bobier, W. R., Chakraborty, A., Dai, S., Hess, R. F., Jenkins, M., Jiang, Y., Kearns, L. S., Kowal, L., Lam, C. S. Y., Pang, P. C. K., Parag, V., Pieri, R., Raveendren, R. N., South, J., Staffieri, S. E., … Boswell, S. J. (2018). Effectiveness of a binocular video game vs placebo video game for improving visual

functions in older children, teenagers, and adults with amblyopia: A randomized clinical trial. *JAMA Ophthalmology*. https://doi.org/10.1001/jamaophthalmol.2017.6090

Harrad, R. A., & Hess, R. F. (1992). Binocular integration of contrast information in amblyopia. *Vision Research*. https://doi.org/10.1016/0042-6989(92)90075-T

Hess, R. F., Mansouri, B., & Thompson, B. (2010). A new binocular approach to the treatment of Amblyopia in adults well beyond the critical period of visual development. *Restorative Neurology and Neuroscience*. https://doi.org/10.3233/RNN-2010-0550

Hess, Robert F., Hong Liu, C., & Wang, Y. Z. (2002). Luminance spatial scale and local stereo-sensitivity. *Vision Research*. https://doi.org/10.1016/S0042-6989(01)00285-1

Hess, Robert F., Mansouri, B., Dakin, S. C., & Allen, H. A. (2006). Integration of local motion is normal in amblyopia. *Journal of the Optical Society of America A*. https://doi.org/10.1364/josaa.23.000986

Hess, Robert F., & Thompson, B. (2015). Amblyopia and the binocular approach to its therapy. *Vision Research*. https://doi.org/10.1016/j.visres.2015.02.009

Hess, Robert F, Dillon, R., Ding, R., & Zhou, J. (2019). Deficient stereopsis in the normal population revisited: why current clinical stereo tests may not be adequate. *BioRxiv*, 585976. https://doi.org/10.1101/585976

Holmes, J. M., Manh, V. M., Lazar, E. L., Beck, R. W., Birch, E. E., Kraker, R. T., Crouch, E. R., Erzurum, S. A., Khuddus, N., Summers, A. I., & Wallace, D. K. (2016). Effect of a binocular ipad game vs part-time patching in children aged 5 to 12 years with amblyopia a randomized clinical trial. *JAMA Ophthalmology*. https://doi.org/10.1001/jamaophthalmol.2016.4262

Holopigian, K., Blake, R., & Greenwald, M. J. (1986). Selective losses in binocular vision in

anisometropic amblyopes. *Vision Research*. https://doi.org/10.1016/0042-6989(86)90010-6

Huang, C., Tao, L., Zhou, Y., & Lu, Z. L. (2007). Treated amblyopes remain deficient in spatial vision: A contrast sensitivity and external noise study. *Vision Research*. https://doi.org/10.1016/j.visres.2006.09.015

Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *The Bell System Technical Journal*, *39*(5), 1125–1162.

Kingdom, F. A. A., & Prins, N. (2016). Model Comparisons. In *Psychophysics: A Practical Introduction* (2nd ed., pp. 247–307). Academic Press.

Kiorpes, L., Tang, C., & Movshon, J. A. (1999). Factors limiting contrast sensitivity in experimentally amblyopic macaque monkeys. *Vision Research*. https://doi.org/10.1016/S0042-6989(99)00130-3

Knox, P. J., Simmers, A. J., Gray, L. S., & Cleary, M. (2012). An exploratory study: Prolonged periods of binocular stimulation can provide an effective treatment for childhood amblyopia. *Investigative Ophthalmology and Visual Science*. https://doi.org/10.1167/iovs.11-8219

Levi, D. M., Harweth, R. S., & Manny, R. E. (1979). Suprathreshold spatial frequency detection and binocular interaction in strabismic and anisometropic amblyopia. *Investigative Ophthalmology and Visual Science*.

Levi, Dennis M. (2013). Linking assumptions in amblyopia. In *Visual Neuroscience*. https://doi.org/10.1017/S0952523813000023

Levi, Dennis M., & Klein, S. A. (2003). Noise provides some new signals about the spatial vision of amblyopes. *Journal of Neuroscience*. https://doi.org/10.1523/jneurosci.23-07-02522.2003

Levi, Dennis M., Klein, S. A., & Chen, I. (2008). What limits performance in the amblyopic visual system: Seeing signals in noise with an amblyopic brain. *Journal of Vision*. https://doi.org/10.1167/8.4.1

Levi, Dennis M., Knill, D. C., & Bavelier, D. (2015). Stereopsis and amblyopia: A mini-review. In *Vision Research*. https://doi.org/10.1016/j.visres.2015.01.002

Levi, Dennis M., McKee, S. P., & Movshon, J. A. (2011). Visual deficits in anisometropia. *Vision Research*. https://doi.org/10.1016/j.visres.2010.09.029

Li, J., Thompson, B., Lam, C. S. Y., Deng, D., Chan, L. Y. L., Maehara, G., Woo, G. C., Yu, M., & Hess, R. F. (2011). The role of suppression in amblyopia. *Investigative Ophthalmology & Visual Science*. https://doi.org/10.1167/iovs.11-7233

Li, S. L., Reynaud, A., Hess, R. F., Wang, Y. Z., Jost, R. M., Morale, S. E., De La Cruz, A., Dao, L., Stager, D., & Birch, E. E. (2015). Dichoptic movie viewing treats childhood amblyopia. *Journal of AAPOS*. https://doi.org/10.1016/j.jaapos.2015.08.003

Manh, V. M., Holmes, J. M., Lazar, E. L., Kraker, R. T., Wallace, D. K., Kulp, M. T., Galvin, J. A., Shah, B. K., & Davis, P. L. (2018). A Randomized Trial of a Binocular iPad Game Versus Part-Time Patching in Children Aged 13 to 16 Years With Amblyopia. *American Journal of Ophthalmology*. https://doi.org/10.1016/j.ajo.2017.11.017

Mansouri, B., Thompson, B., & Hess, R. F. (2008). Measurement of suprathreshold binocular interactions in amblyopia. *Vision Research*. https://doi.org/10.1016/j.visres.2008.09.002

Mansouri, Behzad, & Hess, R. F. (2006). The global processing deficit in amblyopia involves noise segregation. *Vision Research*. https://doi.org/10.1016/j.visres.2006.07.017

Mocanu, V., & Horhat, R. (2018). Prevalence and Risk Factors of Amblyopia among Refractive Errors in an Eastern European Population. In *Medicina* (Vol. 54, Issue 1).

https://doi.org/10.3390/medicina54010006

O'Connor, A. R., & Tidbury, L. P. (2018). Stereopsis: are we assessing it in enough depth? In *Clinical and Experimental Optometry*. https://doi.org/10.1111/cxo.12655

Oliphant, T. E. (2007). *SciPy: Open source scientific tools for Python*. Computing in Science and Engineering. https://doi.org/10.1109/MCSE.2007.58

Ooi, T. L., Su, Y. R., Natale, D. M., & He, Z. J. (2013). A push-pull treatment for strengthening the "lazy eye" in amblyopia. In *Current Biology*. https://doi.org/10.1016/j.cub.2013.03.004

Park, W. J., Schauder, K. B., Zhang, R., Bennetto, L., & Tadin, D. (2017). High internal noise and poor external noise filtering characterize perception in autism spectrum disorder. *Scientific Reports*. https://doi.org/10.1038/s41598-017-17676-5

Pelli, Denis G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A*. https://doi.org/10.1364/josaa.16.000647

Pelli, Denis G., Levi, D. M., & Chung, S. T. L. (2004). Using visual noise to characterize amblyopic letter identification. *Journal of Vision*. https://doi.org/10.1167/4.10.6

Pelli, Denis G., Levin, D. M., & Chung, S. T. L. (2004). Using visual noise to characterize amblyopic letter identification. *Journal of Vision*. https://doi.org/10.1167/4.10.6

Pelli, Denis Guillermo. (1981). *Effects of Visual Noise*. Cambridge University.

Prins, N., & Kingdom, F. A. A. (2009). *Palamedes: Matlab routines for analyzing psychophysical data*. Http://Www.Palamedestoolbox.Org/.

Richards, W. (1970). Stereopsis and stereoblindness. *Experimental Brain Research*, *10*(4), 380–388. https://doi.org/10.1007/BF02324765

Richards, W. (1971). Anomalous Stereoscopic Depth Perception. *Journal of the Optical Society of America*, *61*(3), 410. https://doi.org/10.1364/JOSA.61.000410

Seabold, S., & Perktold, J. (2010). Statsmodels: econometric and statistical modeling with Python. *9th Python in Science Conference*. https://doi.org/10.1007/s00367-011-0258-7

Smith, E. L., Chino, Y. M., Ni, J., Cheng, H., Crawford, M. L. J., & Harwerth, R. S. (1997). Residual binocular interactions in the striate cortex of monkeys reared with abnormal binocular vision. *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1997.78.3.1353

Tittes, J., Baldwin, A. S., Hess, R. F., Cirina, L., Wenner, Y., Kuhli-Hattenbach, C., Ackermann, H., Kohnen, T., & Fronius, M. (2019). Assessment of stereovision with digital testing in adults and children with normal and impaired binocularity. *Vision Research*. https://doi.org/10.1016/j.visres.2019.07.006

Van Ee, R., & Richards, W. (2002). A planar and a volumetric test for stereoanomaly. *Perception*. https://doi.org/10.1068/p3303

Wang, Y., Zhang, B., Tao, X., Wensveen, J. M., Smith, E. L., & Chino, Y. M. (2017). Noisy spiking in visual area V2 of amblyopic monkeys. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3178-16.2016

Wardle, S. G., Bex, P. J., Cass, J., & Alais, D. (2012). Stereoacuity in the periphery is limited by internal noise. *Journal of Vision*, *12(6)*(12), 1–12. https://doi.org/10.1167/12.6.12

Webber, A. L., Schmid, K. L., Dowdall, R. J., Finlay, M. E., Larner, R. L., Lee, E., Oh, B. S., Skokidis, K. A., Baldwin, A. S., Reynaud, A., & Hess, R. F. (2018). Suppression rather than visual acuity loss limits threshold stereoacuity. *Investigative Ophthalmology & Visual Science*, *59*, 4115.

Webber, A. L., Wood, J. M., & Thompson, B. (2016). Fine motor skills of children with amblyopia improve following binocular treatment. *Investigative Ophthalmology and Visual Science*. https://doi.org/10.1167/iovs.16-19797

Zaroff, C. M., Knutelska, M., & Frumkes, T. E. (2003). Variation in stereoacuity: normative description, fixation disparity, and the roles of aging and gender. *Investigative Ophthalmology & Visual Science*, *44*(2), 891–900. https://doi.org/10.1167/iovs.02-0361