Development and Evaluation of an Easy-to-Use Stereoscopic Ability Test to Assess the Individual Ability to Process Stereoscopic Media

Daniel Pietschmann[(✉)], Benny Liebold[(✉)], Peter Ohler, and Georg Valtin

Chemnitz University of Technology, Institute for Media Research, Chemnitz, Germany
{daniel.pietschmann,benny.liebold,peter.ohler, georg.valtin}@phil.tu-chemnitz.de

Abstract. With the rise of 3D cinema in recent years, 3D stereoscopic images have quickly conquered the entertainment industry. As a consequence, many scholars from different research disciplines study the effects of stereoscopy on user experience, task performance, or naturalism. However, parts of the population suffer from stereoblindness and are unable to process stereo images. For scientific studies, it is important to assess stereoblindness to avoid bias in the gathered data. Several clinical tests are available to measure deficiencies in stereo vision, but they often require special equipment and a trained investigator. We developed an easy to use and economic Stereoscopic Ability Test (SAT) that can be used directly within the intended experimental environment. Initial evaluation data for the test and guidelines for the test application are discussed.

Keywords: Stereoscopic vision · Psychology · Experimental · Diagnostics

1 Introduction

In recent years 3D technology has become publicly more and more accepted and by now almost every new TV-set is capable of 3D presentation. Research in stereoscopic virtual reality [1], video games [2, 3], and 3D-cinema [4] became more popular as well. On the other hand, not all users can enjoy the stereoscopic media: Some users are stereoblind, other users have healthy vision, but suffer from unacceptable side-effects such as nausea and headaches as part of motion- or cybersickness [2]. In this paper, we focus on users who suffer from stereoblindness. Reasons for not being able to process binocular depth information may stem from the loss of vision in one eye or medical disorder that prevent the eyes from processing binocular depth cues correctly (e.g. amblyopia, optic nerve hypoplasia, strabismus). While this group of users is still capable of perceiving monocular depth cues (e.g. relative size, motion parallax, accommodation, occlusion), they are incapable of directly perceiving an object's distance through eye convergence.

With reported numbers of the stereoblind population ranging from 2 to 12 percent [5, 6], this can have two serious effects on studies on stereoscopic media: First, stereoblind participants represent a strong bias in a study's sample. When stereoscopic 3D is manipulated in an experiment, they are unaffected by the manipulation, and whenever stereoscopic 3D is enabled in a virtual environment, they might be outperformed by the other participants due to a lack of spatial information. Second, using stereoscopic 3D can have benefits for user experience (UX). Thus, stereoblind participants might provide different UX ratings compared to participants with healthy vision. Most notably, participants may not be fully aware that their stereoscopic vision is impaired. Consequently, it is important to assess participants' ability to perceive stereoscopic images as part of a study. While there are many clinical tests for stereoblindness (e.g. cover tests, bagolini lenses, worth's four dots, measurement of accommodation and near point of convergence, etc.), they often require special equipment and a trained investigator to derive correct conclusions [7].

Additionally, while clinical tests assess general stereo vision for depth perception, stereoscopic media do not provide visual stimuli with real visual depth. Instead, the illusion of depth is created by introducing disparity between the images for both eyes on a fixed distance. Thus, the user has to adapt to the necessary convergence, while keeping accommodation constantly at the display. This is different from real depth perception, for which convergence and accommodation are highly correlated. Thus, to assess the ability to process depth cues in 3D media, the instrument should focus on the stereoscopic technology and its specific characteristics in addition to real stereo vision. Furthermore, from a methodological standpoint, assessing stereo vision within the same medium (i.e. display) that is used in a given study increases internal validity. Thus, our goal was to develop an easy-to-use test for stereoblindness that could be used with a wide range of stereoscopic displays.

2 Depth Perception and Stereoscopic Media

The function of human depth perception is the estimation of distances within our three-dimensional environment. Our eyes perceive two-dimensional images (one with each eye), but our cognitive system interprets all available stimuli to extract depth cues. The images on both retinas differ by the interpupilar distance (IPD) so that objects are perceived from two (slightly) different perspectives (i.e. disparity). Our cognitive system interprets both monocular and binocular spatial cues. Monocular cues include occlusion and relative sizes of objects, height in the visual field, linear and aerial perspective, texture gradients, and motion parallax [8]. Binocular cues encompass accommodation and vergence of muscles in the eyes and stereopsis [8].

Accommodation refers to the adaptation of the eye lenses to focus near or far objects. Vergence describes the rotation of the eyes towards each other (convergence) to focus near objects or away from each other (divergence) to focus far objects. Both processes are highly correlated and enable binocular vision. Due to physical constraints, binocular vision is usually limited to short distances of two to three meters, whereas we can interpret monocular depth information far beyond 30 meters and more [9].

2.1   Stereoscopic Media

Stereoscopy relies on the ability of the human perceptual apparatus to process separate images with each eye. Each eye is presented with a slightly different two-dimensional image, resulting in the illusion of binocular depth perception.

The first stereoscopic photographs were used in a Wheatstone's mirror stereoscope in the 19th century, and later in stereograms. The tradition of stereoscopic media is long, and features many different technologies to present stereo images [10], e.g. the color anaglyph system which was used in 3D movies in the early 20th century.

Stereoscopic images have been used in virtual reality systems since the early 1990s [11, 12]. In recent years, a new trend of stereoscopic media sparked in the movie industry with the release of movies like *Avatar* [13]. It quickly reached mainstream home entertainment like television and video games. Today, almost any new television set or video game console is capable of presenting stereoscopic media content. Currently, researchers are looking for new ways to minimize side effects caused by the technology [14, 15] and to maximize user experience with stereoscopic media, e.g. by including depth cues in cinematography [16] or video game design [17, 18]. In recent the past, new low-cost solutions (e.g. VR headsets such as Oculus Rift) were introduced to consumers, emphasizing the importance of further research on interaction design, positive and negative effects as well as future directions of stereoscopic media.

Currently, there are many different technologies to implement stereoscopy, divided into four main categories: (1) Active stereoscopic methods achieve the separation of both images through active shutters. These systems are usually glasses with built-in liquid crystal layers, which darken when voltage is applied, effectively blocking the view of an eye when an image is intended for the other eye and vice versa. (2) Passive systems do not use electrical components, and separate the images through color filters (e.g. anaglyph), prisms (e.g. chromadepth system) or polarization filters. (3) Head-mounted-displays do not require the separation of two superimposed images, as they can present two distinct images for each eye (e.g. on two small displays or side-by-side display). (4) Methods without special viewing aids are called autostereoscopic display technologies, as the stereo effect is achieved by viewing the images from a specific position (e.g. parallax barriers, volumetric displays, holography).

As stereoscopy relies on the illusion of depth cues, there are several key drawbacks like double images or simulation sickness [19]. When using HMDs, passive or active techniques, the images are presented on a flat surface with a simulated parallax. The parallax results in a convergence of the eyes behind the display pane. By manipulating this parallax, objects can be displayed in front or behind the display pane. This may result in a vergence-accommodation-conflict: When focusing objects with real binocular vision, accommodation and convergence are concordant with the estimation of the distance to the object. In stereoscopy, vergence and accommodation are different from each other, resulting in conflicting information about the distance estimation. This is especially true for scenes with very different fields of depth. Eye stain is induced by repeatedly changing focus from foreground to background objects. These problems can be minimized by limiting objects to only a small field of depth. If the display is farther away from the eyes, the distances between objects is also reduced, thus reducing the differences in the field of depth. Additionally, problems may occur if the images are

computed with the wrong IPD or with a curvature of the visible field, e.g. by different fields of view.

In practice, this means that due to physical limitations, stereoscopy is limited to a smaller area of depth perception than real stereo vision [20]. For long-term interactions, applications should even limit the use of depth cues to only a small portion of this area (Percival's Zone of Comfort [21]).

3 Development of a Stereoscopic Ability Test

The development of a test for the stereoscopic ability during media exposure faces three major challenges. First, the test needs to assess stereoscopic ability by isolating binocular visual cues. As long as users are able to infer depth information from monocular visual cues, for example from object size or texture resolution, the test has no diagnostic value. Second, the test needs to be compatible with various technological setups, may it be VR-environments, such as CAVE or HMD, TV-sets, or simple PC-based presentations. As a result, the test has to ensure that it produces similar results across all platforms. Furthermore, the created stimuli need to be easily supported by every platform. Third, the measure needs to account both for stereo blindness and for interindividual differences in familiarity with stereoscopic display technology. Thus, stimuli with appropriate difficulty levels need to be created.

We decided to develop the stimulus materials in a 3D editor to ensure a convenient manipulation of object parameters. The stimuli can be exported as two images (one for each eye), which should be feasible for most use cases. The images can either be directly displayed on the respective eye (HMD) or implemented into a 3D-movie format, which allows displaying the stimuli easily on modern TV-sets. Another option is to directly implement the 3D-objects into a rendering engine (e.g. CAVE-setups). We chose the 3D-editor *Blender 2.68* (Blender Foundation), because it is freely available and allows other researchers to change parameters, if necessary. The stereoscopic rendering within Blender is carried out by the Blender-Plugin *Stereoscopic Rendering in Blender 2.7*[1], which is also freely available. The plugin provides support for eye-based camera positions and automatically sets the near and far plane as well as the focal plane for stereoscopic presentations to guide the developer during the positioning of 3D elements in the scene.

Our goal was to develop a pool of different stimuli with varying difficulty that would undergo a pretest to select the most appropriate for the final test. As a starting point we decided to include four tasks into the test both to cover different aspects of stereoscopic vision during media use and to avoid boredom and learning effects during the testing procedure. Within each task either one (task 1) or three (tasks 2-4) grey squares were presented. Regardless of their positioning in the screen, each square was individually sized so that every square had the same size on the screen. We also did not
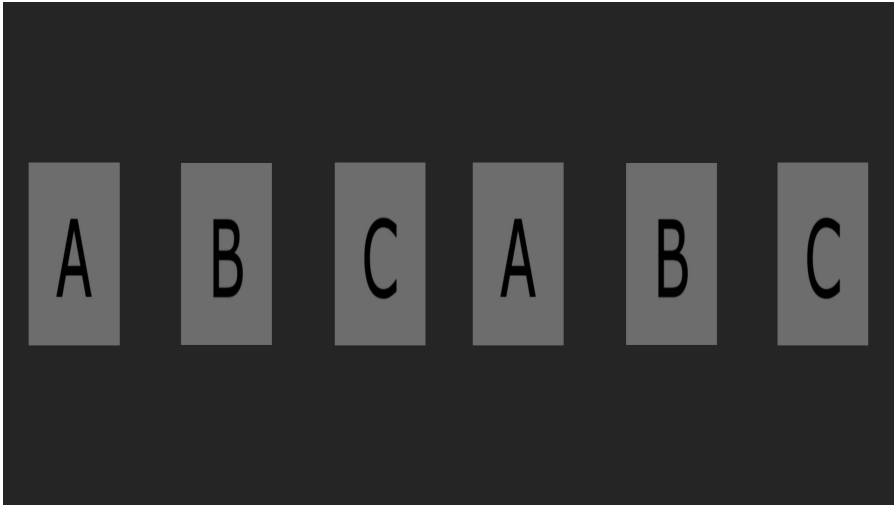
---

[1] http://www.noeol.de/s3d/

Fig. 1. In this example item (side-by-side presentation) of the stereoscopic ability Test (SAT), participants had to sort the squares front to back.

apply a complex texture to the squares, but colored them in a solid grey. This way it was assured that all monocular depth cues were removed from the squares. Because this procedure made focusing the squares nearly impossible (after all, they had no texture), we added a letter (A, B, C) to each square. The implemented tasks were the following (see also Fig. 1):

Task 1: Focal Plane Comparison. During task 1 participants had to decide whether a single square was in front of the screen or behind it. The task contained 53 items with varying depth. We expected the items near the focal plane to have the highest difficulty.

Task 2: Deviation Detection. In task 2 participants saw three squares, which were either all on the same plane or not. Half of the 56 items depicted a group of squares with identical depth. The remaining items depicted squares, which had a fixed distance between them and a varying distance from the observer. We expected items far away from the observer to be the most difficult, as the relative distance between the squares becomes smaller with increasing distance from the observer.

Task 3: Oddity Detection. During task 3 participants had to indicate, which of the displayed squares was different from the others. While two random squares were always on the same plane, one square was either behind or in front of that plane with varying distance. This task included 52 items. We expected smaller distances between the items to result in higher item difficulty.

Task 4: Sorting Task. In task 4 participants had to sort the squares from front to back. While the mutual distance between the squares was identical for every item, all squares varied in their distance to the observer. We again expected items farther away from the observer to be the most difficult. The task contained 44 items.

4 Empirical Evaluation

The evaluation of the created stimuli included two small-scale studies aiming to assess first estimations of psychometric properties of the scale.

4.1 Apparatus and Task

The stimuli were exported from Blender as two separate images for each eye. We then combined both images into side-by-side images with a resolution of 1920 × 1080 pixel using image-processing software. The stimuli were presented using side-by-side 3D-conversion on a 55" Full-HD TV-set with passive stereoscopy. The seating distance was standardized according to THX distance recommendations (*diagonal display size divided by 0.84*). Apart from the illumination from the TV-set, the room was completely dark.

All stimuli were presented using the software *E-Prime 2.0* (Psychology Software Tools), which allowed recording of user button presses for each stimulus. Each task began with an instruction followed by example items. The actual items were presented in a randomized order. Further, we took two measures to minimize eye-strain and exhaustion. First, every item was followed by a two seconds black screen. Second, we implemented a non-interruptible one-minute pause after task 2.

4.2 Procedure

Participants were recruited via mailing lists. After their arrival, they were explained that they were taking a test to assess their stereoscopic ability and their informed consent was gathered. We pointed out that the presentation of stereoscopic images could lead to symptoms of simulator sickness and that they could abort the experiment at any time, should they feel uncomfortable. They then took a seat in a comfortable cinema chair in front of the TV-set and put on the polarization glasses. We specifically instructed the participants, not to glimpse past the glasses.

4.3 Results of Study 1

In the first study, N = 8 participants (all normal sighted) answered all 215 stimuli. Our goal was to gather early estimates for the item characteristics of the developed stimuli. Thus, we conducted a preliminary item analysis to select suited stimuli. Most stimuli proved very easy and, thus, had only little diagnostic value for the developed test. As expected, item difficulty was highest, when the target stimulus was closest to the point of reference (i.e. the screen plane or other items). For each of the four tasks, we selected items of varying difficulty to cover a wide range of stereoscopic blindness and familiarity with stereoscopic displays. Because all items could be answered either correctly or wrong, an item difficulty approaching .50 in consecutive items from one task represents the threshold level at which, only half of the participants are able to provide the correct answer. Because this answering format also has a chance-level of
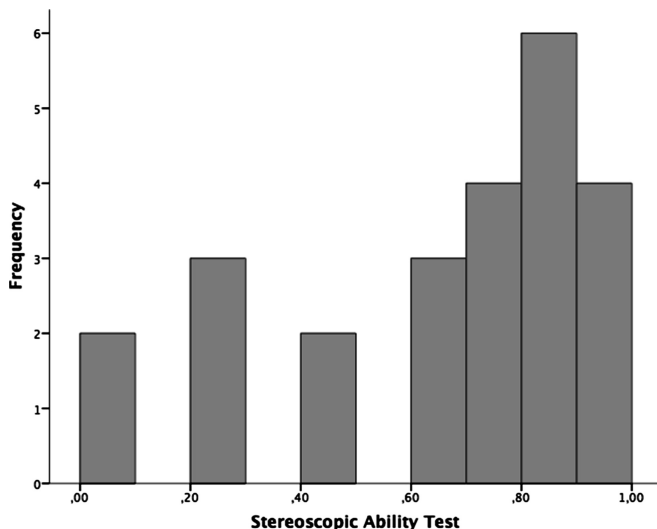
Fig. 2. Distribution of the raw values of the stereoscopic ability test

p = .50, items of this difficulty basically reflect that participants guessed the answer, because they were just about not able to give a definite answer. In most cases, these items only represented a small range of the manipulated parameters near the point of reference. We therefore created new items around these points, which were slightly more easy or difficult, to be able to select from a bigger pool of stimuli for the final test. The final sample consisted of 156 stimuli.

4.4 Results of Study 2

A sample of N = 26 student participants (mean age: 23 years; female: 16; 53.8% corrected to normal vision) had to complete all 156 stimuli to determine the final item selection for the Stereoscopic Ability Test (SAT). Items ranged from .38 to .88 regarding their difficulty. The final stimuli were selected according to their difficulty, variance, and variety of tasks in order to create a sensitive measure. From the initial set of items, 12 stimuli were selected to be included in the final test. The final scale had a mean score of $M = .66$ ($SD = .296$; see Fig. 2). Reliability analysis indicated a high internal consistency with $\alpha = .88$. An exploratory factor analysis supports the idea that the SAT has one underlying factor with $R^2 = .46$ and high factor loadings for every item. When only participants with scores higher than $M = .50$ are analyzed, the SAT sufficiently follows a normal distribution.

4.5 Discussion

Our goal was to develop an easy-to-use test to assess participants' stereo blindness and familiarity with stereoscopic display technology. The Stereoscopic Ability Test

(SAT) consists of 12 items that can be displayed on every stereoscopic display. The employed analyses suggest a high sensitivity to interindividual differences in stereoscopic ability and the high internal consistency suggests a reliable and one-dimensional measure.

The validity of the test needs to be addressed in a future study. Still, the current results give at least weak support for the discriminative power of test, as few participants score rather low in the test with the majority of the participants achieving medium to high scores. An open question remains, why some participants scored significantly lower than chance level ($p = .50$). Even stereo blind participants should score values of $M = .50$. Although our goal is not to develop a measure that could serve clinical purposes, the test needs at least to be compared with one clinical measure of stereo blindness.

5   Implementing the SAT for Research

The developed test items are freely available from the project website (http://www.medkom.tu-chemnitz.de/sat/). We provide the blender project files necessary to implement the stimuli in rendering engines as well as ready to use side-by-side images for simple use cases such as TV-sets. The website also features statistics of the test evaluations (including future studies), explanations for stimulus implementation as well as guidelines for test usage.

References

1. Baños, R.M., Botella, C., Alcañiz, M., Liaño, B.A., Guerrero, B., Rey, B.: Immersion and emotion: Their impact on the sense of presence. CyberPsychology Behav. **7**, 734–741 (2004)
2. Häkkinen, J., Pölönen, M., Takatalo, J., Nyman, G.: Simulator sickness in virtual display gaming: a comparison of stereoscopic and non-stereoscopic situations. In: 8th International Conference on Human Computer Interaction with Mobile Devices and Services Helsinki (2006)
3. Takatalo, J., Kawai, T., Kaistinen, J., Nyman, G., Hakkinen, J.: User experience in 3D stereoscopic games. Media Psychol. **14**, 387–414 (2011)
4. Sobieraj, S., Krämer, N.C., Engler, M., Siebert, M.: The influence of 3D screenings on presence and perceived entertainment. In: International Society for Presence Research Annual Conference (ISPR 2011), Edinburgh (2011)
5. Richards, W.: Stereopsis and stereoblindness. Exp. Brain Res. **10**, 380–388 (1970)
6. van Mierlo, C.M., Brenner, E., Smeets, J.B.: Better performance with two eyes than with one in stereo-blind subjects judgments of motion in depth. Vis. Res. **51**, 1249–1253 (2011)

7. Lee, J., McIntyre, A.: Clin. tests binocular vis. Eye **10**(Pt 2), 282–285 (1996)
8. Coren, S., Ward, L.M., Enns, J.D.: Sensation and Perception. Harcourt Brace College Publishers, New York (1994)
9. Cutting, E., Vishton, P.M.: Perceiving layout and knowing distances: the interaction of relative potency, and contextual use of different information about depth. In: Epstein, W., Rogers, S. (eds.) Perception of Space and Motion. Academic Press, San Diego, CA (1995)
10. King, R.D.: A brief history of stereoscopy. Wiley Interdisc. Rev. Comput. Stat. **5**, 334–340 (2013)
11. Biocca, F.: virtual reality technology: a tutorial. J. Commun. **42**, 23–72 (1992)
12. Steuer, J.: Defining virtual reality: dimensions determining telepresence. J. of Commun. **42**, 73–93 (1992)
13. Cameron, J.: Avatar. pp. 162 min. Twentieth Century Fox (2009)
14. Woods, A.J., MacKenzie, K.J., Watt, S.J., Holliman, N.S., Dodgson, N.A.: Eliminating accommodation-convergence conflicts in stereoscopic displays: can multiple-focal-plane displays elicit continuous and consistent vergence and accommodation responses? 7524, 752417-752417-752410 (2010)
15. Lambooij, M., Ijsselsteijn, W., Fortuin, M., Heynderickx, I.: Visual discomfort and visual fatigue of stereoscopic displays: a review. J. Imaging Sci. Technol. **53**, 1–14 (2009)
16. Liu, C.-W., Huang, T.-H., Chang, M.-H., Lee, K.-Y., Liang, C.-K., Chuang, Y.-Y.: 3D cinematography principles and their applications to stereoscopic media processing, 253 (2011)
17. Schild, J., LaViola, J.J., Masuch, M.: Altering gameplay behavior using stereoscopic 3D vision-based video game design. In: SIGCHI Conference on Human Factors in Computing Systems ACM CHI (2014)
18. Rivett, J., Holliman, N.S.: Stereoscopic game design and evaluation. In: Stereoscopic Displays and Applications XXIV. SPIE - The International Society for Optical Engineering (2013)
19. Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G.: Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. **3**, 203–220 (1993)
20. Ware, C., Gobrecht, C., Paton, M.: Dynamic adjustment of stereo display parameters. IEEE Trans. Sys. Man Cybernen. Part Sys. Hum. **28**, 56–65 (1998)
21. Hoffman, D.M., Girshick, A.R., Akeley, K., Banks, M.S.: Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. J. Vis. **8**, 33:31–33:30 (2008)