

Cali Fidopiastis*

Optical Diagnostics and
Applications (ODA) Laboratory
University of Central Florida
Institute for Simulation and Training
and College of Optics and
Photonics
4000 Central Florida Boulevard
Orlando, FL 32816-2700

Christopher Fuhrman

Catherine Meyer

Optical Diagnostics and
Applications (ODA) Laboratory
College of Optics and Photonics
University of Central Florida
4000 Central Florida Boulevard
Orlando, FL 32816-2700

Jannick Rolland

Optical Diagnostics and
Applications (ODA) Laboratory
College of Optics and Photonics
Institute for Simulation and Training
University of Central Florida
4000 Central Florida Boulevard
Orlando, FL 32816-2700

*Correspondence to
cali@odalab.ucf.edu

Methodology for the Iterative Evaluation of Prototype Head-Mounted Displays in Virtual Environments: Visual Acuity Metrics

Abstract

Head-mounted display design is an iterative process. As such, a standardized user-centered assessment protocol of head-mounted performance during each phase of prototype development should be employed. In this paper, we first describe a methodology for assessing prototype head-mounted displays and virtual environments using visual performance metrics. We then present an application of the methodology using a prototype of a projection head-mounted display and the first module of our assessment: resolution visual acuity as a function of contrast. To evaluate the total system, we also used three different light levels and two different types of projection materials. Results from both studies indicate that the visual acuity metric resolution accurately identified reductions in user visual acuity caused by parameters of the projection display and those of the phase conjugate material. Results further support the need for benchmark metrics that allow comparison of prototype head-mounted performance through each stage of design.

I Introduction

According to Melzer and Moffit (1997), the human performance test and evaluation process of head-mounted displays (HMD) should be system oriented, user centered, and systematic. Applying user centered testing during the prototype phase of display development has the added benefit of reducing design errors, which in turn lends to reduced development costs. Performing systematic tests over the full design cycle allows for benchmarking, and thus a means of optimizing the system as a whole.

A prototype, whether foam or working model, can be thought of as an interactive version of the final product (Preece, Rogers, & Sharp, 2002). Users can test HMD attributes such as head strap adjusters for functionality and usability using a simple foam model. However, to further explore user capabilities and to utilize human performance data as an evaluative measure, the HMD must be a working model (Melzer & Moffit, 1997, p. 323). Given a working model, human performance data may be collected to either verify system parameters (e.g., HMD resolution) or to quantify user behavior (e.g., navigational ability).

There are several test batteries used to quantify human performance in vir-

tual environments. Both the virtual reality assessment battery (VRAB) and the augmented reality performance assessment battery (ARPAB) include more complex perceptual and psychomotor tasks such as object recognition and object manipulation to quantify how well the human interacts with the virtual environment (Lampton et al., 1994; Kirkley et al., 2002). On the other hand, before user metrics become meaningful, system designers must optimize VE technology to account for human visual performance (Davis, 1997; Kalawsky, 1993). Thus, as designers of HMD technology we propose a complementary HMD performance assessment battery that allows for optimizing prototype VE system parameters utilizing a human-in-the-loop paradigm (National Research Council, 1997; Rash & McLean, 1998).

Much like in the optimization of night vision goggles (Task, 1991), system developers are concerned with limits in HMD parameters such as display resolution (image quality), field of view (information quantity), and contrast (light intensity changes). Such parameters may be quantified by purely engineering means where the eye is replaced by a camera (Speck & Herz, 2000) or by human-in-the-loop methods. The advantage of the former is speed of measurement, while the latter is a more holistic evaluation of the system. Because of the similarity in display constraints, we propose to adapt psychophysical metrics used in optimizing night vision goggles such as “frequency seeing” curves to the assessment of prototype HMDs (National Academy of Sciences, 1980; Pinkus & Task, 1998). In the next section we describe how psychophysics can be applied toward optimizing HMDs as early as the prototype phase of design.

2 Psychophysics and HMD Testing

Classical threshold theory started in the 1800’s with the work of Gustave Fechner on quantifying the relationship between physical properties of the world such as changing levels of light illumination and human sensory perception. In this context, psychometric functions represent the relationship between the user’s ability to detect a stimulus and the changing levels of the

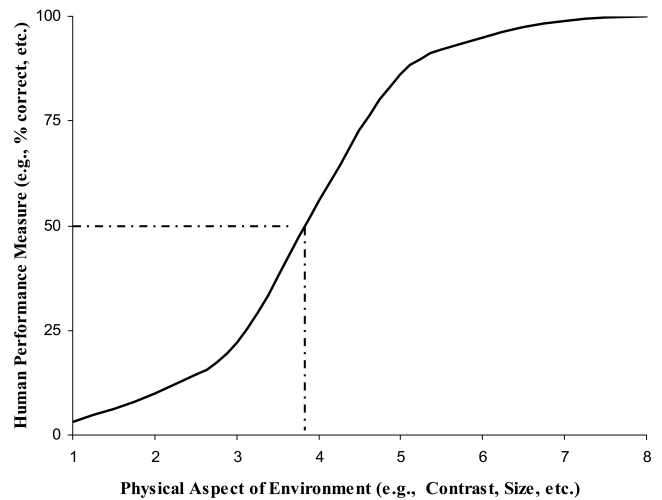


Figure 1. Sensory threshold example: A psychometric function reporting percent correct on a detection task versus one of the physical characteristics of the stimuli is shown. The sensory threshold along the x-axis is defined as the point corresponding to 50% correct.

stimulus parameter (e.g., lighting, size, contrast). Within the threshold theory framework, the sensory threshold calculated from the psychometric function measures the limits of sensory perception for a given sense organ. Although the existence of an absolute sensory threshold is under debate, the statistical concept does allow for the quantification of human sensitivity to physical stimuli (Bi & Ennis, 1998).

Figure 1 graphically illustrates the definition of the sensory threshold as the parameter level denoted on the abscissa where the observer positively detects the stimulus 50% of the time (Schiffman, 2000). Combined with the method of forced choice (e.g., the participant must select one of the available stimuli presented), the effects of guessing are minimized allowing for a more accurate estimation of the sensory threshold.

Because of its versatility and ease of use, the sensory threshold is frequently used as a metric for establishing the sensitivity of the human visual system. For example, the visual tests summarized in Table 1 measure the general perceptual phenomenon of visual acuity, which is defined as the ability to see fine details. Of the visual acuity tests, the Snellen test is the fastest to perform, and is therefore the most common. The Snellen test

Table 1. *Examples of Visual Acuity Tests*

Detection	Detect target in visual field	Examples
Vernier or localizing	Identify displacement of two lines in space	Manual vs Gaze pointing
Resolution (rVA)	Perceive the separation of two distinct elements in space (e.g., resolve gap size in the letter C)	Landolt C
Recognition	Name targets in space (e.g., identify letters)	Snellen
Dynamic	Locate moving targets in space	MultiCAD

Source: Shiffman, 2000, p. 99.

results are used most often for evaluating a driver's ability to read or recognize lettering on traffic signs. However, when evaluating visual perceptual issues regarding VE, such as the limited resolution of the HMD microdisplays, resolution visual acuity (rVA) which is the ability to resolve small visual angles, is a better metric (Howarth & Costello, 1997; Bullimore, Howarth, & Fulton, 1995). Thus, for HMD prototyping, choosing the most appropriate perceptual test is critical for predicting task-based human performance while wearing the HMD.

The further benefit of using psychophysical metrics is that we can create testable conditions that quantify the relationship between the physical parameters of the complete VE system (e.g., HMD limited resolution and 3D graphics) and perceptual performance (May & Badcock, 2002; Uttal & Gibb, 2001). Based on the knowledge of how the system is expected to perform, we can also make a priori hypotheses of how the user will perceptually perform (Meister, 1985). By choosing the most appropriate comparison variables for determining the relationship, we can more effectively evaluate and optimize the performance of the HMD within a VE system.

For example, the head mounted projection display (HMPD) used in the forthcoming experiments has a

52° field of view (FOV) and uses a 640 × 480 resolution VGA microdisplay (Hua, Ha, & Rolland, 2003). It may be thought of as a miniature projector mounted on the head with retroreflective material strategically placed in the environment (Hua, Girardot, Gao, & Rolland, 2000). Based on the HMPD angular resolution defined by display parameters of size, pixel number, FOV and the type of projection material, we predict a maximum angular resolution visual acuity (rVA) for the human observer wearing the HMPD to be 4.1 minutes of arc. The visual acuity metric necessary to quantify the effective resolution of the display is resolution visual acuity (rVA). Thus in this case, the Landolt C visual acuity test is the most appropriate test (Fidopiastis, Meyer, Furhman, & Rolland, 2003). For more advanced human performance tasks such as detection of targets in complex backgrounds, we measure rVA as a function of target contrast. This metric has shown to correlate well with these more advanced visual tasks.

Given a calibrated HMPD, differences between the system predictions and psychophysical measurements suggest that the user's rVA is either limited by properties of the HMPD or the properties of the projection material. Moreover, further tests could discern limitations attributed to each separate aspect of the system or the system as a whole. The end result is a basic, yet rigorous optimization procedure (Melzer & Moffit, 1997).

3 Human Vision and HMD Prototyping

The basic aspects of human vision relevant to developing a prototype HMD are the separate neural pathways mediating color, size/spatial frequency, and contrast (relative luminance changes). These neural based functional distinctions determine limits of the visual system, and thus the visual tests used to discern reduced performance across the respective visual channels (American Optometric Association, 2003). Receptive fields in the human retina are spatially tuned to respond best to specific sizes of stimuli in the visual scene. Visual acuity tests such as Landolt C test only the smaller receptive fields that are responsible for fine detail

Table 2. Table of Photometric Definitions and Their Respective Values

Quantity	Definition	Units
Luminance flux	The quantity of radiant flux that creates a visual perception	Lumen (lm)
Luminance intensity	The luminance flux emitted from a point source per unit solid angle	Candela (cd) or lumen · steradian ⁻¹ (lm · sr ⁻¹)
Illuminance	Luminance flux per surface area; mediates scotopic and photopic vision	Lux or lumen · meter ⁻² (lm · m ⁻²)
Luminance	The luminous flux propagated from a surface point in a particular direction; mediates brightness and contrast	Candela · meter ⁻² (cd · m ⁻²) or lumen · meter ⁻² · steradian ⁻¹ (lm · m ⁻² · sr ⁻¹)
Glare	A light source of higher luminance than anticipated by the visual system, which decreases visual performance	
Reflectance	The ratio of the luminance flux that strikes a surface to the luminance flux reflected off the same surface	

processing. Furthermore, the recommended contrast for optotypes used in these tests is .85, which corresponds to high contrast targets (National Academy of Sciences, 1980). While the contrast level of optotypes may be varied there are no standards defined for varying contrast as a function of stimulus size.

The contrast sensitivity function is a mapping of perceived contrast as a function of spatial frequency over all spatial frequency channels. Visual acuity resolution is the limit of this function. Future work will be expanded to investigate how the range of human spatial frequencies is filtered in the VE system. Currently, we focus on the sensory threshold tests for size (Landolt C) as a function of target contrast over the smaller frequency channels. Regardless of pathway, photometric variables are the input to the channels. Next, we define photometric variables typically used in HMD assessment.

4 Photometric Variables and HMD Testing

Photometric quantities are radiometric measures of light corrected for the spectral efficiency (range) of

the human eye. The relevant photometric quantities of light are summarized in Table 2. The rods and cones of the human retina perform optimally under different levels of illuminance (Lux) and luminance (cd/m²). The cones, where visual acuity is the best, mediate photopic vision or conditions of high levels of light and the rods are responsible for performance under low light or scotopic conditions. The interaction between the human visual system and both illuminance and luminance determines what we see and how we see it (Gilchrist, 1994).

Studies have shown that decreasing light illumination levels slows reaction time and decreases accuracy on different visual performance tasks (Murdoch, 1985). Under lighting conditions of low illumination, visual acuity is limited (e.g., reading in the dark) due to scotopic vision. Performance decrements may also be attributed to a decrease in luminance, or light emitted from an object (Boyce, 1981). Perceived contrast, as used in the experiment, was calculated using the luminance (L) differences between the target and background as defined by the equation:

$$\frac{L_{\text{Target}} - L_{\text{Background}}}{L_{\text{Background}}} = C_{\text{Perceived}} \quad (1)$$

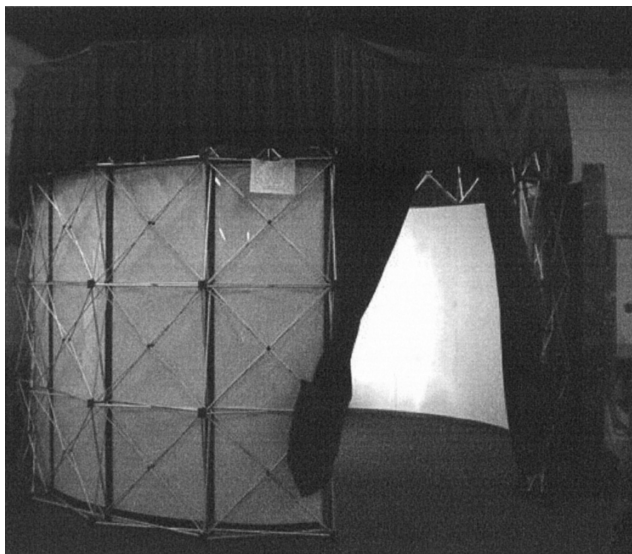


Figure 2. Augmented Reality Center (ARC) display.

For simulations, illuminance and luminance values can be estimated from a lighting compendium such as the Illuminating Engineering Society's publication, *Lighting Ready Reference*. Otherwise, the values must be measured using a light illuminance meter and a photometer prior to the experiment, respectively.

5 Methodology Application: Projection HMD

In our application, we are testing user rVA using a first generation HMPD custom designed at the ODA lab at the University of Central Florida. The rVA experiments were conducted within an augmented reality visualization center (ARC) shown in Figure 2. The ARC is a prototype multi-modal augmented reality system with 3D visual, 3D audio and haptic capabilities configured as a quasi-circle of 4.57 m (15 ft) diameter (Davis, Rolland, Hamza-Lup, & Ha, 2003; Rolland, Stanney, et al. 2004). The display consisted of a curved wall of retroreflective material, and a Linux-based PC. The projection optics and use of retroreflective material distinguishes this HMPD from other projective displays (e.g., CAVE) and conventional HMDs (Rolland, Biocca, et al. 2004).

The HMPD optics (lightweight, <6 g per eye) projected images to the user in a binocular 52° field of view (Hua, Ha, & Rolland, 2003). Off-the-shelf miniature LCDs within the HMPD had a VGA resolution of 640 × 480, and gave a predicted user rVA of 4.1 arc minutes.

5.1 Retroreflective Material

Currently, there are two types of retroreflective material used in the ARC display, Scotchlite¹ 3M Fabric Silver (Beaded) and Scotchlite 3M Film Silver (Cubed). The beaded material is an off-the-shelf material that acts through conjugate reflection (most reflected light is concentrated back in the direction of the incident ray). The cubed material was custom specified and utilizes total internal reflection (TIR), whereby the entire amount of incident light is reflected back in the same direction. Both materials are optimized for safety applications, not for imaging. However, an optical structure operating on TIR properties is expected to yield higher reflectance. The point-spread functions, which describe the surface brightness distribution of a point source imaged on each material confirm that the cubed material achieved not only higher levels of brightness across the image plane, but also higher resolution (Martins & Rolland, 2003; Rolland, Biocca, Hamza-Lup, Ha, & Martins, 2005). We thus hypothesize that the corner cubed material may yield higher rVA performance than the beaded material.

5.2 Test Parameters

For this study, we are evaluating the image resolution of the microdisplay as well as the performance of the retroreflective materials. In addition, we chose to test the VE system over different levels of illumination. We created a computerized version of a binocular modified Landolt C visual acuity test to measure the user's rVA across varying levels of target contrast.

1. Scotchlite is a registered trademark of the 3M Company.

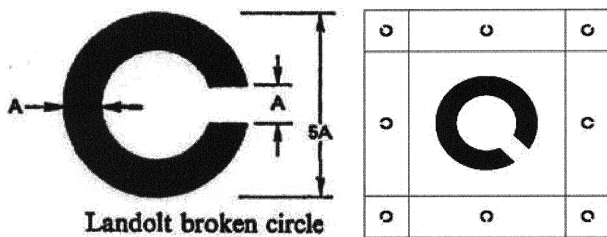


Figure 3. Standard Landolt C visual acuity test.

5.3 Landolt C Visual Acuity Test

The Landolt C visual acuity test is standard for measuring resolution acuity (National Academy of Sciences, 1980). The Landolt C is constructed on a 5×5 grid with the gap the width of 1 grid unit as shown in Figure 3. The gap is presented in one of eight positions: right, down, left, and up. From a visual perspective there are two drawbacks in using the Landolt C: uncorrected astigmatism may make some orientations of the gap easier to see, and guessing may show biased performance at low acuities, usually toward the right (Rabbetts, 1998, p. 32).

To control for these potential confounds, all participants were tested for resolution acuity using corrective lenses, if needed. Also, the participants had unlimited presentation time. Although guessing still occurred, there was no added pressure to perform under a time limit.

Typical stimuli for the Landolt C test, pictured on the right of Figure 3, are circular C's. Rendering a circular C requires the use of anti-aliasing techniques to remove the stair-stepping effect found in low bandwidth displays. Figure 4 shows typical Landolt C stimuli with aliasing and anti-aliasing effects present when rendering the curvature of the C.

To remove these effects, a square C was created that retained the 5:1 ratio between the C size and gap size. To further reduce aliasing and anti-aliasing effects the VA test was changed from an 8 AFC (see Figure 2) to a 4 AFC. Figure 5 shows the view of the square Landolt C test from the left and right eyes within the HMPD.

5.4 Lighting

Pilot studies showed that subjects performed at ceiling levels under ideal lighting conditions on both



Figure 4. Aliasing and anti-aliasing effect.

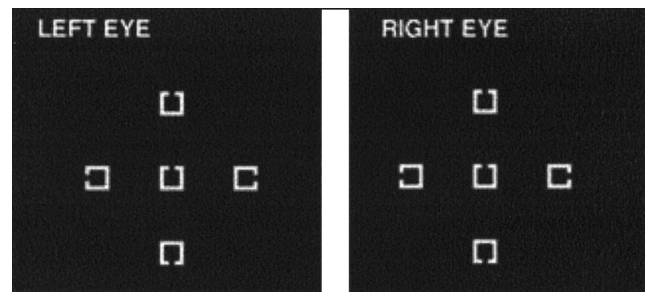


Figure 5. Modified Landolt C visual acuity test.

materials. The purpose of the lighting condition was to check for differences in material performance when lighting conditions changed. Because of limitations of the ARC display, only mesopic and scotopic light ranges were used. These light levels reported in Table 3 roughly correspond to *IES Lighting Ready Reference* (Kaufman & Christensen, 1989) values for lighting spaces for optimal user performance.

5.5 Probit Analysis

There are several statistical tests available to evaluate the sensory threshold. Probit, best PEST, and Quest tests have all been used to evaluate visual acuity type data (Finney, 1980, Leberman & Pentland, 1982; Watson, 1983). Adaptive methods such as best PEST and Quest use an adaptive procedure where the individual's response determines the next presentation of the stimuli. Our concern with this type of test was that cumulative errors would build over the large number of trials being conducted by a single subject (Kaernbach, 2001). Pinkus and Task

Table 3. *Illuminance Categories and Values*

Illuminance category	Illuminance value
Performance on tasks of high contrast or large size	200 to 500 Lux (18.58 to 46.45 footcandle)
Working spaces where occasional work is performed	100 to 200 Lux (9.29 to 18.58 footcandle)
Simple orientation for short temporary visits	50 to 100 Lux (4.64 to 9.29 footcandle)

Source: *IES Lighting Ready Reference*, 1989, p. 87.

(1998) provide some support for reliability issues using an adaptive method as a threshold estimator. We therefore chose a Probit analysis to determine user rVA thresholds.

In Probit analysis the percent correct responses are corrected for chance responding, then they are converted to z-scores or normal equivalent deviates (NED). The NED values are used as the dependent variable in a linear regression with gap size (visual acuity) as the independent variable (e.g., $NED = b_0 + b_1 \cdot VA$). We followed Probit analysis procedures outlined by Finney (1980) by using a maximum likelihood estimator to estimate users' rVA. Once obtained, the resulting predicted NED values were converted back into percents and evaluated against the threshold value to determine visual acuity. For a 4 AFC design, the sensory threshold is 62.5%.

6 Experiment 1: Method

6.1 Participants

Twelve participants (11 men and 1 woman, mean age = 28 years) performed the Landolt C Visual Acuity test under 3 different light levels (high, medium, and low) on both the computer and in augmented reality (e.g., using the HMPD). Each participant was either corrected for or had 20/20 vision. Glasses or contacts were worn during each part of the experiment.

6.2 Design and Procedure

Software ran on a computer system with Linux RedHat 7.2 OS and a dual processor graphics card.

For calibration purposes, the LCD monitor was a Dell 17 in. flat screen. Within the ARC, we replaced two of the Scotchlite 3M Fabric Silver (Beaded) panels with Scotchlite 3M Film Silver (Cubed) panels.

6.3 Landolt C Computer vs. Augmented Reality Version

Six different gap sizes representing the different levels of visual acuity in minutes of arc were presented. The stimuli were the same in each condition; however, the visual angle corresponding to each size (arc minutes) of the Landolt C changed from the computer to the augmented reality condition. Table 4 presents the change in visual acuity in both arc minutes and Snellen ratio equivalents for each test environment (i.e., computer or augmented reality). To accommodate for these differences, the participants were seated 1.85 m from the LCD monitor in the computer version and were 2 m from the retroreflective material in the augmented reality condition.

6.4 Lighting Setup

As shown in Figure 6, the lighting was suspended from the ceiling of the ARC to better control the effects of glare and extraneous light that may bounce off the floor onto the retroreflective material. Three 13 Watt 4-pin PI fluorescent lamps (780 lumens, warm 2700 K) were mounted on a 51.5 × 59.5 cm gator board. The board was attached to the ceiling in the center of the ARC as shown in Figure 6. A diffuser of sheer black material and length 0.72 meters was attached to the board.

Table 4. Visual Acuity Conversion from Arc Minutes to Snellen Ratio Values for Each Testing Environment (Computer and Augmented Reality)

Computer (arc minute)	Snellen conversion (ft and m)	Augmented reality (arc minute)	Snellen conversion (ft and m)
1	20/20 or 4/4	4.1	20/82 or 4/16.4
2	20/40 or 4/8	8.2	20/164 or 4/32.8
3	20/60 or 4/12	12.4	20/248 or 4/65.6
4	20/80 or 4/16	16.5	20/330 or 4/66
5	20/100 or 4/20	20.6	20/412 or 4/82.4
6	20/120 or 4/24	24.7	20/494 or 4/98.8

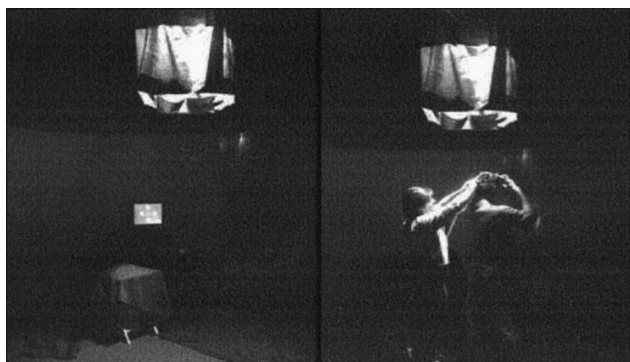


Figure 6. Lighting and setup.

The lighting levels were set using a Minolta T-10 illuminance meter. Illumination was set at three levels, high (225–209 Lux), medium (140–132 Lux), and low (61–53 Lux) to give an average 80 Lux difference between each light level.

6.5 Experimental Design

Participants were randomly placed in a retroreflective material condition (beaded or cubed). The three light levels were changed for each participant according to a partially counterbalanced design determined prior to the experiment. Within each test block, each Landolt C gap size was presented in a random order. As well, there were five random presentations of each direction (up, down, left, and

right) per level of visual acuity. Test blocks were run first with the computer version of the Landolt C test, and then within the HMPD. Participants completed one test block per condition for a total of six tests. There were a total of 720 responses per participant. Once the responses were converted to percent correct and then converted to NED values, a Probit analysis using light level, visual acuity, and gap direction as independent variables was completed.

6.6 Procedure

The initial lighting level was set for each participant prior to their entering the ARC. Participants were first seated 1.85 m from the computer monitor. The interpupillary distance (IPD) of each subject was measured using a pupillometer. This value was entered into the Landolt C software and the HMPD was adjusted. This initial setup allowed the participant to visually adapt to the lighting conditions within the ARC prior to the start of the experiment. Participants were given a keyboard and instructed to use the arrow keys to indicate the direction of the gap in the Landolt C stimuli.

Each time the participant key pressed a response, a beep would precede the appearance of the next stimulus. The test stimulus would remain on the screen until the participant indicated a response. Once the subject had completed the computer version of the VA test, they were asked to stand on a mark 2 m from the retro-

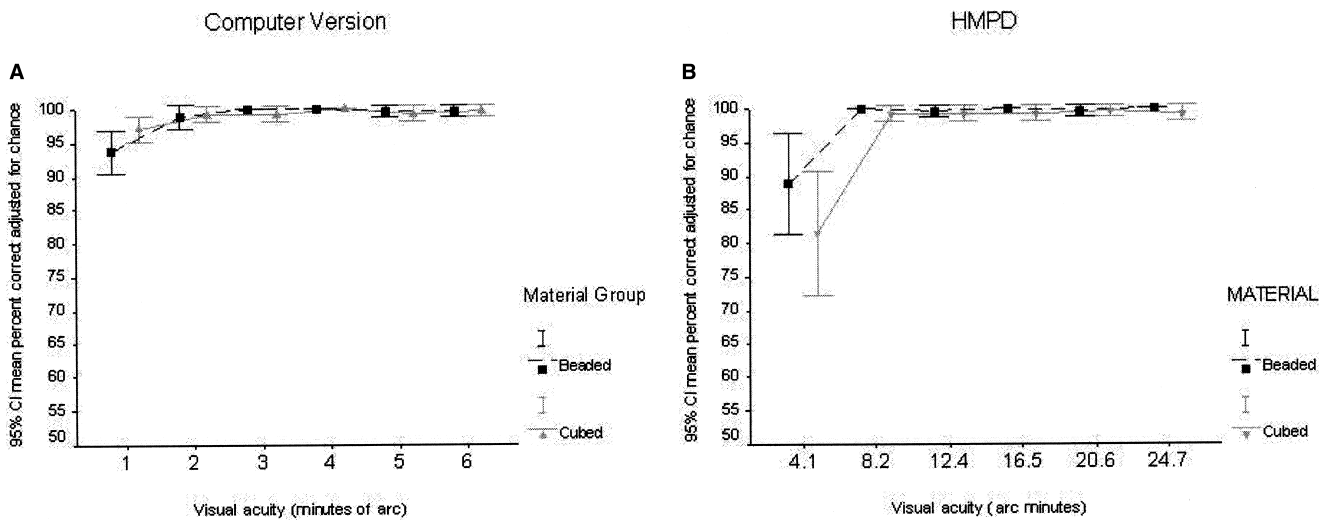


Figure 7. Probit analysis computer versus HMPD.

reflective material. The HMPD was adjusted for IPD and head size. Once the rVA test was completed within the HMPD, the subject was asked to sit again while the experimenter changed the light level. The procedure was repeated until data was collected for all three light levels.

6.7 Results Experiment 1

The results showed that participants scored well above threshold on the computer version of the Landolt C test ($M_{\text{Cubed}} = 98.64$, $SD = 3.22$; $M_{\text{Beaded}} = 99.07$, $SD = 1.96$). For visual acuity corresponding to 20/20 vision, the participants in each group performed well above the 62.5% threshold ($M_{\text{Cubed}} = 93.70$, $SD = 4.82$; $M_{\text{Beaded}} = 97.04$, $SD = 2.89$). Although performance was poorer in the HMPD for the smallest visual acuity (4.1 arc minute), the proportion of correct responses was still above the 62.5% threshold ($M_{\text{Cubed}} = 88.89$, $SD = 11.91$; $M_{\text{Beaded}} = 81.48$, $SD = 14.66$).

A between subjects ANOVA was used to analyze the percent correct responses adjusted for chance. The ANOVA showed that there was no significant effect of room lighting on rVA while testing in HMPD while viewing either type of retroreflective material $F(2, 144) = 4.39$, $p = .139$. There was a main effect of gap size.

6.8 Experiment 1 Conclusions

Results of the experiment confirmed that the visual acuity limit using the current see-through HMD system was 4.1 arc minutes. The microdisplay, and not the retroreflective material, imposes this VA limit. However, as we are able to use higher microdisplay resolutions, the retroreflective material may become the system limiting factor. The experiment also showed that given the constraint in lighting imposed by the HMPD, there is no benefit to visual acuity in changing light levels within a scotopic or mesopic range.

7 Experiment 2

7.1 Participants

Twelve participants (11 men and 1 woman, mean age = 28 years) performed the contrast version Landolt C visual acuity test under a constant low light level on both the computer and in augmented reality (i.e., using the HMPD). Participants were screened for contrast rVA above chance on the computer version of the test. Of the 12 participants only six participants went on to perform the complete experiment. Each participant was either corrected for or had 20/20 vision. Glasses or contacts were worn during each part of the experiment.

Table 5. Gray Level Values for Contrast with a Background Grayscale Value of 128

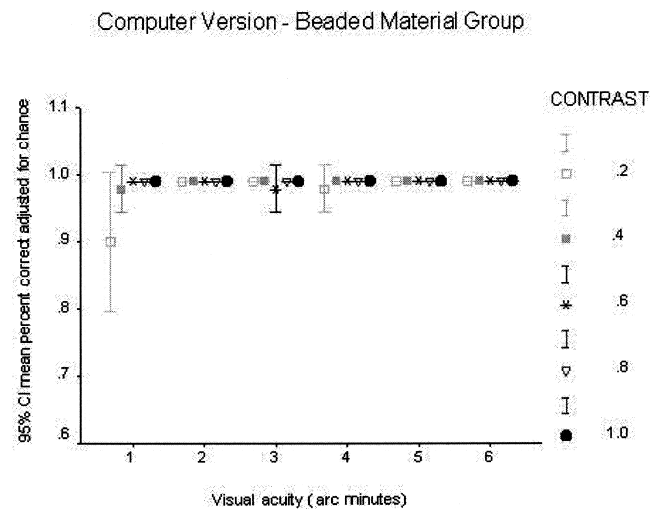
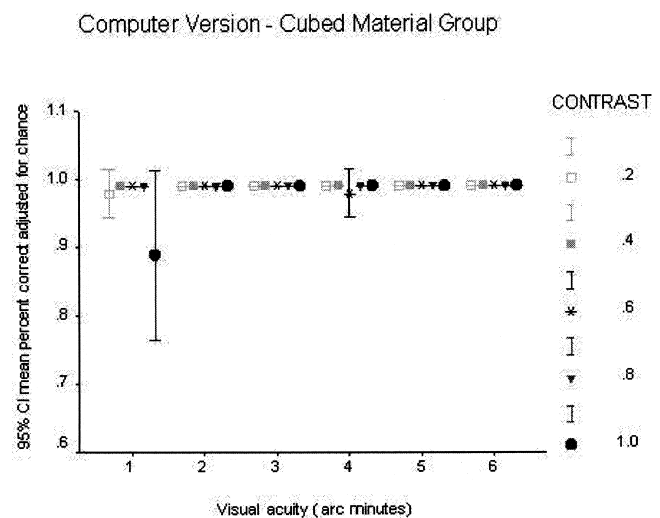
Byte value	Contrast
128	0
153	0.2
179	0.4
204	0.6
230	0.8
255	1

7.2 Design and Procedure

In this experiment we held the lighting constant to evaluate perceived contrast separately from changing illumination levels in the ARC. Luminance of the Landolt C was varied by adjusting the binary grayscale code within the computer program. The background of the display was set at a grayscale value of 128. From this value, target grayscale values giving contrasts of .2, .4, .6, .8, and 1 were computed. The corresponding target grayscale values are shown in Table 5.

7.3 Experimental Design

Participants performed the contrast rVA test over a two-day period. They were randomly placed in one of two retroreflective material conditions (beaded or cubed) on Day 1, and performed the contrast rVA test on the opposite material the following day. There were four random presentations of each direction (up, down, left, and right) per level of visual acuity per contrast level. Test blocks were run first with the computer version of the Landolt C test, and then within the HMPD. Participants completed one test block per condition for a total of six tests. There were a total of 480 responses per participant for each HMPD condition. Probit analysis using contrast, visual acuity, and gap direction as independent variables was completed. Additionally, a within subjects repeated measures ANOVA was performed on the percent correct responses adjusted for chance.

**Figure 8.** Computer version of test for beaded material group.**Figure 9.** Computer version of test for cubed material group.

7.4 Results

Figures 8 and 9 show the 95% confidence intervals for the means of the percent correct values adjusted for chance. The graphs show that there are no differences in rVA measured on the computer version of the Landolt C test over the five contrast levels and between the two days of testing. All percent correct adjusted for chance values were 90% or above for all conditions. The means for .2 contrast and gap size of 1 arc minute for each

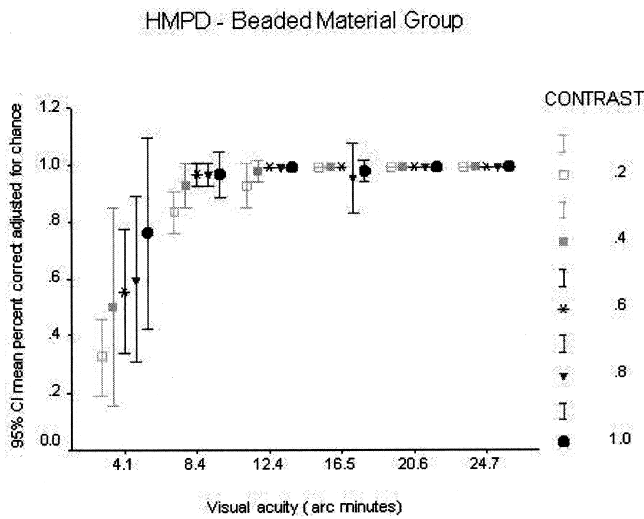


Figure 10. Computer version of test for beaded material group.

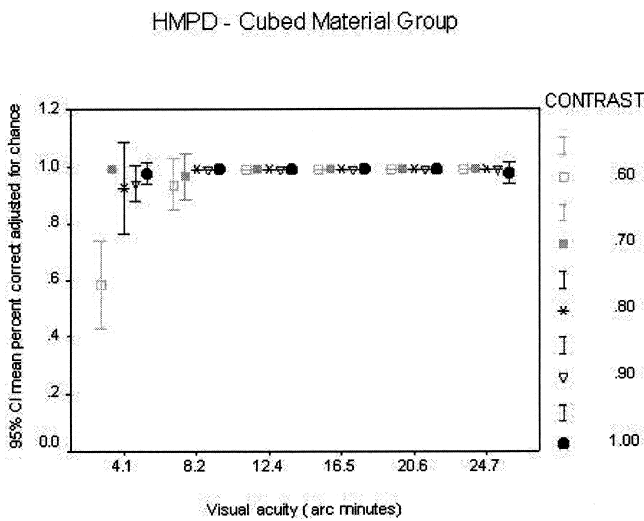


Figure 11. Computer version of test for cubed material group.

group were $M_{Cubed} = 97.8$, $SD = .027$ and $M_{Beaded} = 90.07$, $SD = .065$.

Results from the Landolt C test performed in the HMPD are presented in Figures 10 and 11. A within subjects ANOVA was performed on the percent correct values adjusted for chance. The results showed a two way interaction between the type of material tested on and the gap size $F(5, 25) = 4.718$, $p < .01$. The participants performing the contrast Landolt C test using the

Table 6. Predicted rVA from Probit Analysis

Material	Contrast	Predicted Visual Acuity	SD
Beaded	.2	8.9	1.71
Beaded	.4	6.8	2.12
Beaded	.6	6.1	2.25
Beaded	.8	5.5	2.12
Beaded	1	4.2	0.00
Cubed	.2	6.2	2.19
Cubed	.4	4.2	0.00
Cubed	.6	4.2	0.00
Cubed	.8	4.2	0.00
Cubed	1	4.2	0.00

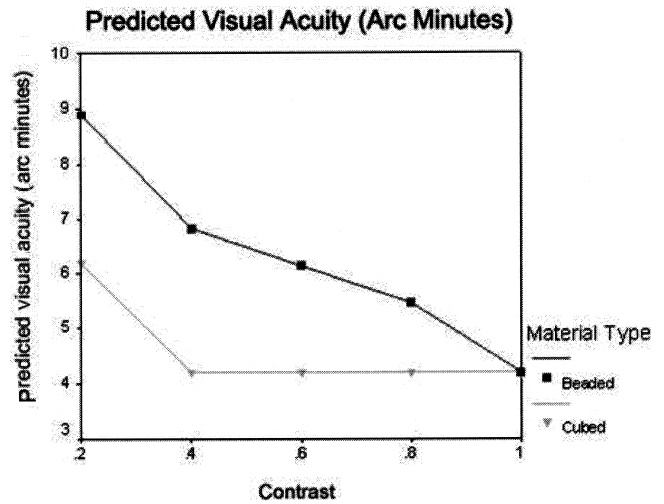


Figure 12. Predicted rVA threshold measured by the probit.

HMPD while viewing the beaded material performed worse than those who were viewed the cubed material for all levels of contrast at the gap size of 4.1.

The rVA thresholds predicted by the Probit analysis are shown in Table 6 and Figure 12. The predicted rVA while viewing the beaded and cubed material for a contrast of .2 was 8.9 arc minutes and 6.2 arc minutes, respectively. A within subjects ANOVA, with contrast and material as independent variables and the predicted rVA as the dependent measure, also showed a main effect for both contrast and material

$C_{\text{Contrast}}(4, 20) = 12.4$, $p < .01$ and $F_{\text{Material}}(1, 5) = 11.9$, $p < .02$, respectively. As expected, as contrast is lessened, rVA is reduced. The results also showed a difference in rVA performance between the two materials for the low contrast targets.

7.5 Experiment 2 Conclusions

The results from experiment 2 show that there is a difference in rVA threshold between the two retroreflective materials as predicted by the respective point-spread functions. The differences in rVA between the materials are only for the low contrast targets. These results confirm that multiple tests of system parameters are necessary for accurate assessment of the VE system.

8 Conclusion

In this paper, we quantified the rVA of a first prototype HMPD as part of assessment during prototype development. We investigated rVA for two different types of retroreflective materials, a commercially available material and a custom specified material, which we anticipated would yield higher performance than the beaded material. Results of two psychophysical experiments show that while the custom-specified material subjectively appeared higher resolution than the other material under all tests, for high contrast targets the rVA is the same for both materials and equal to 4.1 minutes of arc which is the limit imposed by the microdisplay. The results also show that for low contrast targets, the custom designed material is superior in performance. This experiment is critical to task-based performance where high contrast targets are not part of the task performed as would be the case for the detection of targets in complex backgrounds.

Acknowledgments

This research was supported by the Office of Naval Research Grants N000140210261 and N00014-02-1-0927 as well as a Link Foundation Fellowship given to the first author. We would like to thank Dr. Larry Hettinger for his discussions on Human

Factors and display assessment, Chris Stapleton of the Media Convergence Laboratory at the Institute for Simulation and Modeling for his assistance with the lighting configuration, and Drs. Ece and William Batchelder from the Cognitive Science Department at the University of Irvine for their helpful discussions on the statistical analysis. Finally, we thank 3M Corporation for the generous donation of the cubed retroreflective material.

References

- American Optometric Association. (2003). Science of vision. Retrieved May 1, 2003 from http://www.contrast-sensitivity.com/vision_science/.
- Bi, J., & Ennis, D. M. (1998). Sensory thresholds: Concepts and methods. *Journal of Sensory Studies*, 13, 133–148.
- Boyce, P. R. (1981). *Human factors in lighting*. New York: Macmillan Publishing Company.
- Bullimore, M. A., Howarth, P. A., & Fulton, J. (1995). Assessment of visual performance. In J. R. Wilson & E. N. Corlett (Eds.), *Evaluation of human work: A practical ergonomics methodology* (2nd Ed. pp. 804–839). Nottingham, UK: Taylor & Francis.
- Davis, E. T. (1997). Visual requirements in HMDs: What can we see and what do we need to see. In J. E. Melzer & K. Moffit, K. (Eds.), *Head mounted displays designing for the user* (pp. 207–251). New York: McGraw-Hill.
- Davis, L., Rolland J. P., Hamza-Lup, F., & Ha, Y. (2003, March). Enabling a continuum of virtual environment experiences. *IEEE Computer Graphics and Applications*, 23(2), 10–12.
- Fidopiastis, C. M., Meyer, C., Furfman, C., & Rolland, J. P. (2003, September). Quantitative assessment of visual acuity in projective head-mounted displays. In C. E. Rash & E. R. Colin (Eds.), *Proceedings of the SPIE Aerosense: Helmet and Head-Mounted Displays VIII: Technologies and Applications*, 5079, 399–406.
- Finney, D. J. (1980). *Probit Analysis* (3rd Ed.). Cambridge, UK: Cambridge University Press.
- Gilchrist, A. L. (1994). Absolute versus relative theories of lightness perception. In A. L. Gilchrist (Ed.), *Lightness, brightness, and transparency* (pp. 1–31). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Howarth, P. A., & Costello, P. J. (1997). The development of a visual test battery for virtual reality users. In S. A. Robertson (Ed.), *Contemporary ergonomics* (pp. 109–116). London: Taylor and Francis.

- Hua, H., Girardot, A., Gao, C., & Rolland, J. P. (2000). Engineering of head-mounted projective displays. *Applied Optics*, 39(22), 3814–3824.
- Hua, H., Ha, Y., & Rolland, J. P. (2003). Design of ultralight and compact projection lens. *Applied Optics*, 42(1), 97–107.
- Kaernbach, C. (2001). Slope bias of psychometric functions derived from adaptive data. *Perception & Psychophysics*, 63(8), 1389–1398.
- Kalawsky, R. S. (1993). *The science of virtual reality and virtual environments: A technical, scientific and engineering reference on virtual environments*. Wokingham, UK: Addison-Wesley.
- Kaufman, J. E., & Christensen, J. F. (1989). *IES Lighting Ready Reference*. New York: Illuminating Engineering Society of North America.
- Kirkley, S. E., Kirkley, J. R., Borland, S. C., Waite, T., Dumanoir, P., Garrity, P., et al. (2002, December). *Embedded training with mobile augmented reality*. Paper presented at the 23rd Army Science Conference, Orlando, FL.
- Lampton, D., Knerr, B., Goldberg, S., Bliss, J., Moshell, J., & Blatt, B. (1994). The Virtual Environment Performance Assessment Battery (VEPAB): Development and Evaluation. *Presence*, 3(2), 145–157.
- Lieberman, H. R., & Pentland, A. P. (1982). Computer technology microcomputer-based estimation of psychophysical thresholds: The best pest. *Behavior Research Methods & Instrumentation*, 14(1), 21–25.
- Martins, R., & Rolland, J. P. (2003). Diffraction properties of phase conjugate material. In C. E. Rash & E. R. Colin (Eds.), *Proceedings of the SPIE Aerosense: Helmet and Head-Mounted Displays VIII: Technologies and Applications 5079*, 277–283.
- May, J. G., & Badcock, D. R. (2002). Vision and virtual reality. In K. M. Stanney (Ed.), *Handbook of Virtual Environments Technology* (pp. 29–63). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Meister, D. (1985). *Behavioral analysis and measurement methods*. New York: Wiley.
- Melzer, J. E., & Moffitt, K. (1997). *Head mounted displays—Designing for the user*. New York: McGraw-Hill.
- Murdoch, J. B. (1985). *Illumination engineering—From Edison's lamp to the laser*. New York: Macmillan Publishing Company.
- National Academy of Sciences. (1980). Recommended standard procedures for clinical measurement and specification of visual acuity. *Advances in Ophthalmology*, 41, 103–148.
- National Research Council. (1997). *Tactical display for soldiers: Human factors considerations*. Washington, DC: National Academy Press.
- Pinkus, P., & Task, H. L. (1998). Measuring observers' visual acuity through night vision goggles. *ASC98-1884*. www.hecafrl.af.mil/publications/VisAcu.pdf.
- Preece, J., Rogers, Y., & Sharp, H. (2002). *Interaction design: Human-computer interaction*. New York: Wiley.
- Rabbetts, R. B. (1998). *Bennett & Rabbets' Clinical Visual Optics* (3rd ed.). Oxford, UK: Butterworth Heinemann.
- Rash, C. E., & McLean W. E. (1998). Optical performance helmet mounted displays: Design issues for rotary-wing aircraft. Retrieved May 1, 2003 from http://www.usaarl.army.mil/hmdbook/cp_0000_startup.htm.
- Rolland, J. P., Biocca, F., Hamza-Lup, F. G., Ha, Y., & Martins, R. (2005). Development of head-mounted projection displays for distributed, collaborative, augmented reality applications. *Presence: Teleoperators and Virtual Environments*, 14(5), 528–549.
- Rolland, J. P., Biocca, F., Hua, H., Ha, Y., Gao, C., & Harrysson, O. (2004). Teleportal augmented reality system: Integrating virtual objects, remote collaborators, and physical reality for distributed networked manufacturing. In K. Ong & A.Y.C. Nee (Eds.), *Virtual and augmented reality applications in manufacturing* (Ch. 11, pp. 179–200). London: Springer-Verlag.
- Rolland, J. P., Stanney, K., Goldiez, B., Daly, J., Martin, G., Moshell, M., & Washburn, D. (2004, July). Overview of research in augmented and virtual environments: RAVES. *Proceedings of International Conference on Cybernetics and Information Technologies, Systems and Applications (CITSA '04)*, 4, 19–24.
- Schiffman, H. R. (2000). *Sensation and perception: An integrated approach (5th ed.)*. New York: Wiley.
- Speck, R. P., & Herz, N. E. (2000, April). Impact of automatic calibration techniques on HMD life cycle costs and sustainable performance. Paper presented at *SPIE's 14th Annual International Symposium on Aerospace/Defense Sensing Simulation and Controls*, Orlando, FL.
- Task H. L. (1991). Optical and visual considerations in the specification and design of helmet-mounted displays. *Society for Information Display International Symposium Digest of Technical Papers*, 22, 297–300.
- Uttal, W. R., & Gibb, R. W. (2001). On the psychophysics of night vision goggles. In R. R. Hoffman & A. B. Markman (Eds.), *Interpreting remote sensing imagery: Human factors* (pp. 117–135). Boca Raton, FL: Lewis Publishers.
- Watson, A. B. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113–120.