PSYCHOMETRIC ASSESSMENT OF STEREOSCOPIC HEAD-MOUNTED DISPLAYS

Logan Williams[1], Charles Lloyd[2], James Gaska[1], Charles Bullock[1], and Marc Winterbottom[1]

[1]OBVA Laboratory, USAF School of Aerospace Medicine, Wright-Patterson AFB, OH

[2]Visual Performance LLC, Ellisville, MO

ABSTRACT

In this work, which is part of the U.S. Air Force School of Aerospace Medicine Operational Based Vision Assessment program, a high resolution stereoscopic head-mounted display (HMD), X-Plane image generators, and an OptiTrack head-tracking system were used to render an immersive three-dimensional constructive environment. The purpose of this effort was to quantify the impact of aircrew vision on an operationally relevant rotary wing call-to-landing task to research the applicability of U.S. Air Force Flying Class III depth perception standards. Prior to performing this research, an evaluation was carried out to determine whether our simulated environment could support eye-limited stereoscopic disparity. This paper details the psychometric validation of the stereoscopic rendering of a virtual environment using game-based simulation software in a high resolution HMD. The minimum perceived stereo threshold capabilities of this system are also quantified, including its applicability to simulated tasks requiring precise depth discrimination. This work will provide an example validation method for future stereoscopic virtual immersive environments applicable to both research and training.

INTRODUCTION

In 2015, the U.S. Air Force (USAF) School of Aerospace Medicine Operational Based Vision Assessment (OBVA) program implemented a unique three-dimensional (3D) immersive environment to evaluate the applicability of USAF depth perception standards for non-pilot aircrew responsible for clearing aircraft for landing (Flying Class III (FCIII) aircrew with scanner duty). X-Plane was selected, primarily due to its expansive software development community and relatively low cost, to render an immersive 3D constructive environment. An SA Photonics SA-62 head-mounted display (HMD) with head-tracking was selected as the display device to enable a continuous field of regard and to enable the use of motion parallax as a cue to depth, which would normally be available in the natural environment. The use of an HMD, head-tracking, and X-Plane enabled a simulated, operationally relevant MH-60 call-to-landing task, as shown in Figure 1. Prior to implementation of the primary research scenario, this system was validated to ensure eye-limited stereoscopic performance was indeed achievable to meet the requirements of this study.

Figure 1. X-Plane rendering of the simulated call-to-landing task.

This paper details the validation of the 3D virtual environment used to simulate the MH-60 call-to-landing task, including stereoscopic rendering using game-based simulation software, hardware/software antialiasing, and image warping/correction within the HMD. The minimum perceived stereo threshold capabilities of this system are also quantified, with comparison to clinical screening methods. A more detailed description of the apparatus is described in a previous paper [1].

BACKGROUND

Although detailed discussion of the FCIII depth perception research scenario is not the specific goal of this paper, a brief description will illustrate the motivation for the psychometric assessment described herein. Under various operational MH-60 landing conditions, the flight engineer

(FCIII aircrew) is responsible for calling out distance to ground to estimate altitude and avoid obstacles, such as the relatively aggressive landing illustrated in Figure 1. The flight engineer typically leans far outside the aircraft (side window) to make estimations of relative distance between various ground obstacles and aircraft components (e.g., landing gear). The objective of this research was to determine the importance of stereoscopic vision in performing accurate calls to landing. The simulated system, therefore, must be able to replicate the visual cues used by operational aircrew. In particular, the stereoscopic display must be capable of providing eye-limited stereo acuity.

HMD CONFIGURATION

The primary stereoscopic display system consisted of an SA-62 binocular HMD, composed of two 1920x1200 opaque organic light emitting diode displays with 100% overlap, 55°/65° horizontal/diagonal field-of-view (FOV), 4-meter virtual image distance (i.e., 0.25-diopter focal distance), and was capable of motion blur reduction (variable pixel hold time). This HMD was selected in part because it is one of the highest resolution, largest FOV HMDs currently available. Previous work (Lloyd, 2012) suggested that a relatively high resolution was needed to adequately display stereo imagery. Figure 2 shows the HMD with a custom 3D printed rigid body constellation containing five IR reflectors that is mounted to the HMD for use with an OptiTrack IR tracking system.

The native optics of the SA-62 HMD produce significant pincushion distortion; therefore, the Brown-Conrady model for radially symmetric image warping [2] was implemented to minimize optical distortion while adding negligible latency. The model parameters were determined empirically by subjective evaluation of the final image geometry. Two slightly different sets of Brown-Conrady parameters were evaluated to determine the effect on stereo perception of subtle variations in distortion correction. In this case, subtly altering the warp parameters had no measureable effect on stereoscopic threshold or task performance. Therefore, the warp parameters yielding the "best" rectilinear image (based on subjective assessment) were selected.

Figure 2. SA-62 HMD. Photo provided by SA Photonics, used with permission.

STEREOSCOPIC EVALUATION

The utility of any display system for stereo perception research is dependent upon the minimum stereo threshold that may be accurately represented, which ideally will be less than or equal to the minimum stereo threshold of the observer. Such a system is said to be "eye-limited" rather than "display-limited" in stereo acuity. According to USAF aeromedical policy [3], good stereo acuity is defined as 25 arcsec or better. This is consistent with previous research showing that some individuals have stereo acuity as good as about 5 arcsec [4]. However, some individuals may lack stereo vision (stereo blind), or may have very poor stereo acuity (stereo deficient), with thresholds in the 100s of arc seconds. Thus, an eye-limiting level of stereo acuity, for the purposes of this evaluation, is assumed to be around 5-10 arcsec, such that it will not limit even the best observers.

Unfortunately, analytical determination of the absolute stereo threshold of a stereoscopic display system is not straightforward and often depends upon several factors that are unknown, unquantifiable, or reconfigurable by the user. In addition to relatively straightforward parameters such as pixel pitch and horizontal parallax, the stereo threshold of the system can be affected by various factors that are more difficult to quantify, such as antialiasing or pixel interpolation methods, and image warping, as well as the effect of human perception, practice/learning, and actual scene content.

It is important to note that stereo acuity is not determined, or limited, solely by the pixel pitch of the display. Based simply upon a 1-pixel disparity between left/right images, the stereo threshold would simply be equal to the pixel pitch of the display, which is 1.72 arcmin for the SA-62 HMD. However, human stereo perception tends to employ an integrating effect over the extended range of the target(s), not a single pixel, and stereo thresholds far below the pixel pitch are routinely realized [5, 6]. Additionally, rendering effects that smooth edges (e.g., antialiasing) are expected to improve the stereo threshold, while effects that discard or collapse pixels (e.g., image warping) are expected to degrade the stereo threshold [7, 8]. Additionally, the features of the stimulus, such as size/length of each element and separation between elements of the stimulus, will affect the best achievable stereo acuity [9]. Therefore, it becomes necessary to quantify the stereo threshold of the display system in the "as-built" configuration using psychometric measures. Using observers with good stereo acuity, it is then possible to determine if the system is either display or eye-limited for the intended stereo application.
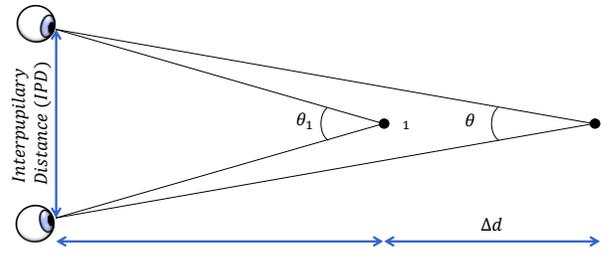
WHAT IS STEREO ACUITY?

Stereo acuity is a measure of an individual's ability to use binocular disparity alone to discriminate the difference in relative depth between two objects placed along substantially identical lines of sight. Therefore, stereo acuity is only one of many cues that make up "depth perception." Other depth cues include known/relative size, geometric perspective, motion parallax, texture gradients, occlusion, etc. The minimum stereo acuity required to discriminate the relative depth between two objects with longitudinal spacing Δd, at a distance d, is given by the minimum angle of binocular disparity $\Delta \theta$:

$$\Delta \theta (arcsec) = 3600 \cdot \tan^{-1} \frac{IPD \cdot \Delta d}{d^2} \qquad (1)$$

It should be noted that this calculation assumes the geometry of Figure 3 holds true, namely, that the objects at points P_1 and P_2 lie on the same longitudinal axis along the line of sight. For many binocular stimuli, this is not strictly true. Often points P_1 and P_2 exhibit some small transverse deviation from the line of sight, such that one object is adjacent to, rather than in line with, the other. Such variations alter the calculation of binocular disparity and greatly complicate its relationship to stereo acuity. Such non-coaxial geometry increases the eccentricity of the stimuli, when viewed by a human observer, by placing the

retinal images farther from the fovea (which subtends ~1°), thus degrading the measured stereo threshold, as shown by Westheimer & McKee [9]. It has also been shown that stereo acuity measurements may be strongly affected by learning on the part of the observer, with well-practiced observers showing substantial improvement over novice observers for identical measurement tasks, especially with electronic displays [10, 11].


Figure 3. Geometry for calculating binocular disparity.

EVALUATION METHODS

Comparison to OBVA Stereo acuity Test

Three observers with good stereo acuity participated in an experiment to establish the stereo acuity threshold of the HMD in a simulated environment. The baseline "clinical" stereo acuity threshold of each observer was measured using the high-fidelity OBVA stereo test battery, which isolates stereo cues by employing constant-size concentric ring stimuli on a flat-panel electronic display (ASUS VG278). The OBVA stereo test battery is similar to the Freiburg Stereo acuity Test [7], albeit using constant size concentric ring stimuli similar to the well-known Titmus/Randot graded circle test. The concentric ring stimulus allows the OBVA stereo test to closely represent the ideal axial geometry, as illustrated in Figure 4.
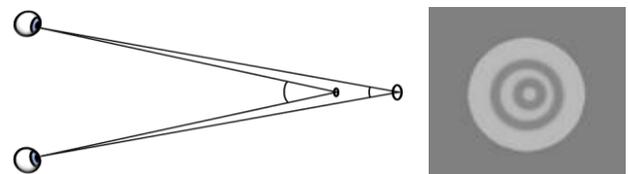

Figure 4. Geometry (left) of the OBVA stereo test battery with concentric ring stimuli (right).

To verify this test is repeatable in the stereo HMD, the OBVA stereo test was repeated in the HMD for three observers with excellent stereo acuity. For both the OBVA stereo test, and the HMD stereo test, the well-accepted Ψ

(psi) method was adopted to estimate a psychometric function using a simple two-alternative forced-choice experiment for stereo acuity [12, 13]. For this test the observer is simply asked to repeatedly discriminate whether the inner circle is in front of, or behind, the larger outer circle. The stimulus adapts to the subject's responses, becoming more or less difficult to discriminate based upon previous answers, until the subject's stereo threshold is determined.

A typical Ψ estimate of the psychometric function is shown in Figure 5, as well as the manner in which the stimulus is altered by the Ψ algorithm between each trial.
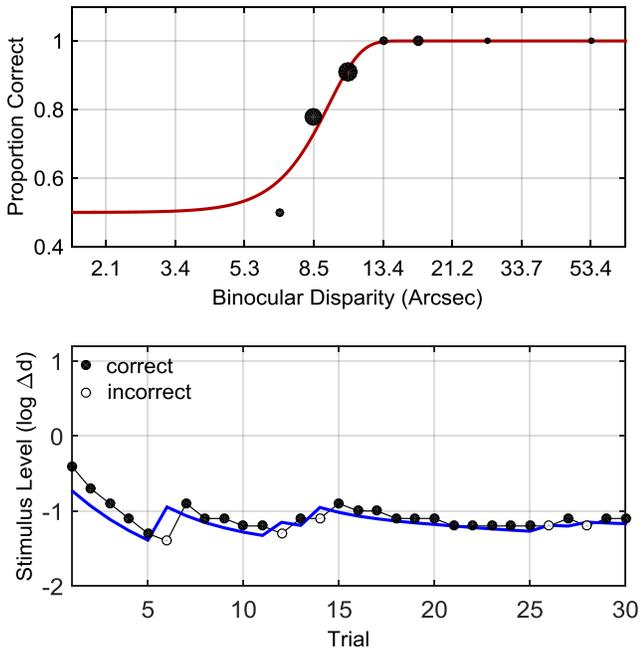
Figure 5. Typical estimate of the psychometric function given by the Ψ method (top) and convergence action of the adaptive algorithm (bottom).

The red curve is the maximum likelihood fit of a Weibull psychometric function that relates binocular disparity to the proportion of correct responses. The blue curve is the Ψ algorithm's current estimate of the observer's threshold and is updated with each correct or incorrect response. The stimulus intensity units are given as $\log(\Delta d)$. Note that the Ψ algorithm tends to converge rather quickly. Typically, at $d = 10$ m, observers with excellent stereo acuity and a typical IPD of 65 mm are able to discriminate real-world differences between objects of about $\Delta d \sim 3.7$ cm ($\Delta\theta \sim 5$

arcsec). Therefore, it is desirable that the stereoscopic display system be able to replicate this performance within the virtual environment using realistic stimuli.

If the HMD is indeed eye-limited in stereo acuity, the OBVA stereo test should yield nearly identical results in both conditions, which is indeed the case, as shown in Table 1. These results provided evidence that the simulation environment developed for this research was capable of providing eye-limited binocular disparity.

Table 1. Comparison of OBVA Stereo Test Battery on Flat Panel and HMD Devices for Three Observers (Red, Blue, and Green)

Observer	Red	Blue	Green
Subject IPD	66 mm	65 mm	6 3mm
Stereo acuity (flat panel)	3.7"±1.7"	6.0"±1.4"	3.0"±1.2"
Stereo acuity (HMD)	5.0"±1.6"	5.1"±1.7"	6.2"±1.4"

Simulated vs. Real-World Targets

There are several differences between the measurement of stereo acuity using the OBVA test battery and similar measurements performed in a rendered scene. First, the OBVA test battery presents stimuli of a constant size, regardless of depth, in an effort to isolate only the stereo cue. Second, because the scene is not rendered in perspective, there is no need to account for an individual observer's interpupillary distance (IPD). Third, the OBVA test battery makes specific use of antialiasing algorithms to achieve sub-pixel shifts in the concentric rings to present a specific stereo stimulus. However, in the rendered scene the level of antialiasing may be set for primarily aesthetic or computational performance reasons, without regard to specific stereo cues, which may affect the achievable stereo acuity limit in a realistically rendered scene. Therefore, it becomes necessary to test the stereo acuity limit of the HMD in the as-built configuration, including the intended virtual environment and rendering settings.

To quantify the stereo threshold of the display in a realistic simulated environment, experimental stimuli were created that could be placed in a virtual environment to simulate the additional size and geometric perspective cues, which will always be present in a realistic scene. The experimental stimuli consisted of two static cross-shaped objects placed $d = 10$ m from the observer (within the virtual environment), with one object slightly closer than the other, as shown in

Figure 6. The geometry of this configuration is more representative of real-world object locations, although it departs further from the ideal coaxial object alignment by placing the target objects adjacent to one another, separated by 3° (center to center), as shown in Figure 6.
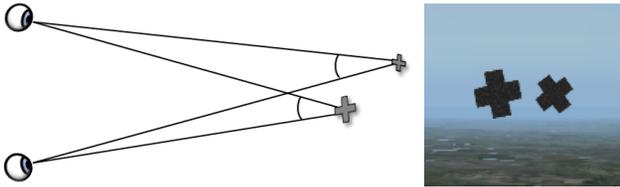

Figure 6. Cross-shaped test targets in X-Plane environment (left). Geometry of the targets used to determine stereo thresholds (right). Note that the relative sizes reveal which object is closer in the absence of stereo cues.

The subject is forced to choose the closer of the two, and the discrepancy Δd between objects is altered by the Ψ algorithm after each trial, based upon the observer's previous response. Each cross is randomly rotated after each trial. The Ψ method was implemented with 30 trials per block, with at least 10 blocks per antialiasing condition, across five available antialiasing conditions, under both stereo and non-stereo rendering conditions, with identical Brown-Conrady image warping implemented across all trials (to correct native HMD pincushion distortion). Stereo images were rendered using each observer's measured IPD, whereas non-stereo images were rendered with IPD = 0, which produces identical centered images for each eye. Three observers with good stereo acuity participated in the experiment, with results detailed in both Table 2 and Figure 7.

Table 2. Comparison of Stereo Threshold Results at Various Antialiasing Multiples

Observer	Red	Blue	Green
IPD	66mm	65mm	63mm
Clinical Stereoacuity	3.7"	6.0"	3.0"
Stereoacuity at 0x	8.0"	14.3"	8.7"
Stereoacuity at 2x	7.6"	12.2"	9.8"
Stereoacuity at 4x	7.3"	11.2"	10.4"
Stereoacuity at 8x	6.8"	11.8"	7.6"
Stereoacuity at 16x	6.2"	10.3"	8.9"
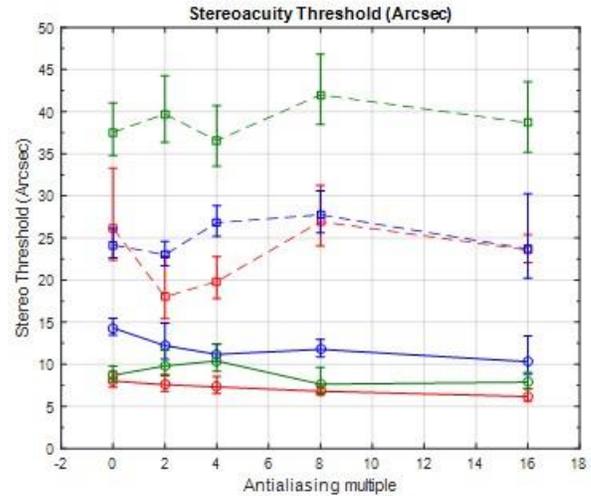
Figure 7. Stereo/depth thresholds of three observers (red, blue, green) at various antialiasing multiples, with (solid) and without (dashed) stereo enabled.

Inspection of Figure 7 reveals several interesting characteristics of the "as-built" HMD system. The first is that this system generally preserves the stereo acuity of the observer (Figure 7, solid lines). Second, the effect of antialiasing on stereo threshold in this system is relatively minor, although it trends somewhat toward better stereo threshold with increased antialiasing (Figure 7, solid lines). However, it should be noted that all observers judged the antialiasing levels of 4x and greater to be the most comfortable for stereo viewing. Low, or no, antialiasing (0x, 2x) produced noticeable "jaggies" at the edges of each object, which were not well correlated between the left and right eye images, resulting in a "shimmering" effect at the edges. Although this had minimal effect on performance, it was judged to be distracting by all observers.

The observer with the best stereo acuity was able to resolve stereo imagery within ~6 arcsec (at 16x antialiasing), which, although still near the limit of human performance, falls somewhat less than the clinical measure. This slight discrepancy is likely due to the differences in stimuli between the two methods. Rather than spatially co-located concentric circles, the simulated stimuli ("cross" objects) are spatially separated, extended objects (as opposed to spatially localized, thin lines), with 3° of separation from center to center, and ~1° to ~1.3° between nearest edges, depending on the angle of rotation. Thus, the spatial separation between the stimuli is larger in comparison to the stereo acuity circles test. This will have the effect of

degrading the measured stereo acuity thresholds, as previously discussed. Although each observer was considered to be well practiced at this task, the effects of learning may also account for some variability in the data.

When stereo rendering was disabled (Figure 7, dashed lines), the stereo threshold of these observers increased to a range of approximately 20" to 40". However, this is likely due to the nature of the rendered stimuli. The simulated stimuli are representative of "real-world" depth scenes, in which the stereo cue is combined with several other depth cues to provide an overall perception of depth. The most notable of these additional cues is object size. For depth inequalities of $\Delta d > 15$ cm at d=10 m, the relative size of the two objects becomes a dominant depth cue, allowing these observers to accurately judge depth with stereo rendering disabled, based solely upon non-stereo cues, to within a 20" to 40" range. This is illustrated by Figure 6, in which the object on the left appears larger (thus closer), even in the absence of stereo cues (i.e., Figure 6 is a 2D image).

DISCUSSION

Based on three different evaluations, we conclude that the X-Plane/SA-62 simulation environment developed for depth perception standards research is capable of providing eye-limited stereo imagery. First, for the simulated cross targets, the fact that individual differences are generally preserved within the stereo cuing range (i.e., below ~20", prior to size cue dominance), rather than converging to some common performance level, is the primary figure of merit in determining the eye-limiting stereo performance of the system. If the system were display-limited, individual differences in stereo threshold would collapse to some common minimum threshold at some or all antialiasing conditions. Since this does not occur, this system can be said to be "eye-limited" with regard to stereo acuity. Second, depth thresholds were comparable between simulated and real-world cross-shaped targets. And third, stereo acuity thresholds were very similar for a laboratory stereo acuity test in comparison to the same test implemented using the HMD.

RESEARCH & TRAINING APPLICATIONS

In this work, which specifically pertains to depth perception research, it was necessary to verify the HMD exhibited eye-limited stereoscopic performance prior to conducting further studies of the effect of stereo vision on operational performance. However, this work may also provide an example verification method/framework using psychometric methods for evaluation of stereoscopic virtual immersive environments applicable to either research or training. Examples of training systems where stereoscopic displays could be important include aerial refueling, helicopter landing, and paratrooper training. Similar verification procedures might also be relevant to applications such as remote/laproscopic surgery. Such stereoscopic HMDs and verification processes can be adapted to a wide range of simulation systems to fulfill a broad spectrum of research and training requirements where accurate stereo representation proves to be a necessary requirement. Although the authors consider this HMD to be "eye-limited" with respect to stereo, it should also be noted that a stereoscopic display that provides less than "eye-limited" stereo performance may still be adequate for a particular task, so long as it can provide a "useful" amount of stereo cuing.

CONCLUSION

This research has demonstrated a psychometric method for quantifying the stereoscopic rendering limit of binocular displays, particularly HMDs, using the "as-built" configuration. Such psychometric methods become necessary when the calculation of stereo threshold becomes difficult or impossible in practical application, as is often the case with simulated virtual environments. Additionally, it has been shown that the use of high-resolution HMDs can provide a fully immersive stereoscopic scene without significantly limiting the stereo acuity of even the keenest users. The display system described herein permits stereo resolution dramatically lower than the display pixel pitch while providing a useful (eye-limited) degree of stereo to enable depth perception research.

ACKNOWLEDGMENTS

REFERENCES

[1]   L. A. Williams, C. T. Bullock, M. D. Winterbottom, J. P. Gaska, S. C. Hadley, and C. J. Lloyd, "3D immersive environment using X-Plane for depth perception research," in *Proceedings of the Interservice/Industry Training, Simulation and Education Conference 2015*, paper no. 15261, pp. 1–10, 2015.

[2]   D. C. Brown, "Decentering distortion of lenses," *Photogramm. Eng.*, vol. 32, no. 3, pp. 444–462, 1966.

[3]   U.S. Air Force, *Air Force Waiver Guide*. Wright-Patterson AFB, OH: U.S. Air Force School of Aerospace Medicine; 2016.

[4]   S. P. McKee, "The spatial requirements for fine stereoacuity," *Vision Res.*, vol. 23, no. 2, pp. 191–198, 1983.

[5]   J. A. Ferwerda and D. P. Greenberg, "A psychophysical approach to assessing the quality of antialiased images," *IEEE Comput. Graph. Appl.,* vol. 8, no. 5, pp. 85-95, 1988.

[6]   S. P. McKee, L. Welch, D. G. Taylor, and S. F. Bowne, "Finding the common bond: stereoacuity and the other hyperacuities," *Vision Res.,* vol. 30, no. 6, pp. 879-891, 1990.

[7]   M. Bach, C. Schmitt, M. Kromeier, and G. Kommerell, "The Freiburg Stereoacuity Test: automatic measurement of stereo threshold," *Graefes Arch. Clin. Exp. Ophthalmol.*, vol. 239, no. 8, pp. 562–566, 2001.

[8]   C. J. Lloyd, "Effects of spatial resolution and antialiasing on stereoacuity and comfort," in *AIAA Model. Simul. Technol. Conf.*, pp. 1–13, 2012.

[9]   G. Westheimer and S. P. McKee, "Stereogram design for testing local stereopsis," *Invest. Ophthalmol. Vis. Sci.*, vol. 19, no. 7, pp. 802-809, 1980.

[10]   M. Fahle and S. Henke-Fahle, "Interobserver variance in perceptual performance and learning," *Invest. Ophthalmol. Vis. Sci.*, vol. 37, no. 5, pp. 869-877, 1996.

[11]   S. P. McKee and D. G. Taylor, "The precision of binocular and monocular depth judgments in natural settings," *J. Vis.*, vol. 10, no. 10, p. 5, 2010.

[12]   F. A. Kingdom and N. Prins, *Psychophysics: A Practical Introduction*. London: Academic Press, 2010.

[13]   Z. Lu and B. Dosher, *Visual Psychophysics: From Laboratory to Theory.* Cambridge, MA: The MIT Press, 2013.