https://doi.org/10.2352/ISSN.2470-1173.2017.18.COLOR-036

Color Discrimination Threshold for Medical Test Devices

Nargess Hassani, Susan P. Farnand; Munsell Color Science Lab, Rochester Institute of technology; Rochester, NY, United States

Abstract

In many medical test designs, presence of a color spot can represent existence of a disease. This presence is usually verified by observation. For these tests, knowing the color discrimination threshold is necessary for modifying the indicator spots to have color differences above the threshold. In this work, a psychophysical experiment is used to determine the color difference threshold for a veterinarian test device from IDEXX company with blue-green spots. The study was conducted in two phases, In the first phase a preliminary investigation was conducted for the ideal situation, that is having perfect circular spots without any noise or non-uniformity. Method of constant stimuli was used to present designed test images to the observers. The results were analyzed using Probit analysis method. The second phase of the study was performed with objective of studying the effects of noise, imperfect spot shapes, presence of streak and presence of spot color gradient on the color difference threshold between the background and spot colors. The same experimental and analysis method was used in both phases. The results for the ideal situation showed an average discrimination threshold of 1.27 color difference (DE00) for overall data. For the realistic situation, the noisy appearance of the image and imperfectness of the shape of the spots did not affect the threshold when observers were expecting imperfect spots. However, the presence of streak and spot gradient increased the threshold.

Introduction

Color plays an important role in object recognition [1]. In medicine, the presence of a color spot can determine existence of a specific disease or health status. Pregnancy test kits are an example of these tests [2]. IDEXX is a company that produces veterinarian test devices that has a functionality similar to these pregnancy test kits. The device has a fabric matrix that includes indicator spots. When the test fluid is implied, in case of presence of specific diseases, color spots will appear in different locations of the matrix. The fluid also causes the color of the background to change. These test devices are used for veterinarian purposes and the results are obtained by visual observation of the designated result window of the device. Since the color appearance in the device matrix is caused by a chemical reaction, it usually has a noisy appearance. Therefore, in some cases, observers tend to get confused about the presence of spots in a test result.

Figure 1 *IDEXX SNAP test device (left) and a close view of a real test results window (right)*

Figure 1 shows the test device and a real result window sample. In the result window of the test device (Fig.1 on the left) we see that there are five spots. The spot on the upper left corner is a reference spot and is always present in the result window, representing that the device is working properly. Other four spots in other locations may or may not be present after implying the test fluid, depending on the presence or absence of certain diseases. As we see from figure 1, the real results are far from perfect. In many situations, there are color streaks in the background or in the edges of the matrix. The spots also are not usually perfectly clear and instead have a non-uniform appearance. These factors influence the process of visual observation and increase the difficulty of deciding on these test results. A color discrimination threshold must be determined for the indicator spots and the background. This threshold can define the minimum color difference of the spots and the background that is required for observers to confidently decide on the presence of the spot.

Color perception is a psychophysical phenomenon, which is usually a result of the interaction of reflected light from objects and eye. As a result, the response of the eye are important factors in color perception. There are different types of color deficiencies and these deficiencies are detectable. Pseudoisochromatic Plates and Farnsworth-Munsell 100 Hue Color Vision Test are two common color blindness tests. Pseudoisochromatic Plates (Ishihara test) includes plates with numbers or patterns made up of dots of random lightness. Depending on the color vision, observers see different numbers or patterns. Farnsworth- Munsell 100 Hue Color Vision Test has four sets of color chips and the observer should put them in a hue order. The mistakes that observer makes reveal the color vision deficiencies or strength of color discrimination [4].

A. Watson has developed a model of human visual sensitivity to contrast patterns [5]. This model is intended for display inspection and is optimized from five simple models by Watson [6]. These models are based on previous works done by Carney et al. [7][8]. Carney et al. worked on display specifications, psychophysical methods and stimulus definition [7] to gather psychophysical threshold data. The results were reported as a dataset (ModelFest data set) [8], which is available online. The stimulus shape that would be of interest for this experiment are black Gaussian blobs with white noises (for noisy stimuli). However, in neither of simple models or the optimized model by Watson, the threshold of the sample of interest is not well modeled. In addition, the samples from Watson et al. are in black and white and would not accurately predict the threshold for our study. Therefore, we were unable to use Watson et al results. However, the results of this study can be used to develop a similar model.

For this study, a psychophysical experiment was designed. The designed test images were presented on a display. Using a display to present the color limits us with displays ability to reproduce colors. Displays have a limited color gamut and are unable to reproduce all colors. When using a display in the experiment, it must be characterized. One method to characterize the display is the method described by Day et al [9]. In this method, a one-dimensional look up table (LUT) is defined for each channel to describe its optoelectronic transfer function. A three by four transform matrix that excludes black-level flare is also included. The matrix coefficients are estimated statistically by minimizing the average CIEDE2000 color difference for the colorimetric dataset of the display. Then throughout the optimization of the matrix coefficients the LUTs are recreated dynamically.

Method of Constant Stimuli was used for the psychophysical experiments. This is a common method for threshold experiments. In This method the procedure involves using the same set of stimuli repeatedly throughout the experiments in random presentation [10]. Several intensity levels of the stimuli are chosen around the threshold and are presented multiple times to the observers. The frequency of the perception of each stimulus is recorded. In the method of constant stimuli, observers are asked to respond "yes" if they perceive the stimulus and "no" if they do not perceive it. Then these data are analyzed using Probit analysis. Probit analysis is based on Probit transformation of the experimental results [11]. Probit function is the inverse of cumulative distribution function of normal distribution. Using Probit analysis, the sigmoid function of response transforms into a linear function. Probit of 5 (T50 point or threshold at 50%) represents the threshold. In this report the Method of Constant Stimuli was used to gather data for the designed images having test spots with different levels of color difference from the background.

Phase I

Experimental

In the first phase of the study, the images were designed to be a simple version of the result window of the device. The spots were designed to be perfectly circular and no noise was present in images. Figure 2 shows the design of the test images in phase one. There was a total of four spot locations, one in each central corner of the test images. The darker spot on the upper left corner was a reference spot and was present in all images. Other spaces were test locations that included test spots with a range of color differences from the background.

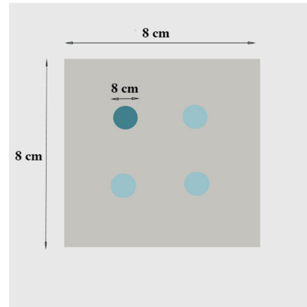

Figure 2 *The design of the images in the first phase of study*

Three backgrounds and six levels of color difference for each background were assigned. The test spots and the backgrounds were chosen regarding the IDEXX lab data set. The spots had the same CIE L* values as the reference spot. This was done to exclude effects of the lightness difference on the overall color difference. The color difference (CIEDE00) for the six levels of each background are given in the table 1.

Table 1 color differences for test levels (CIE DE2000)

Color difference	*Background 10%*	*Background 25%*	*Background 50%*
1	**1.36**	**0.6**	**1.34**
2	**2.04**	**1.18**	**1.77**
3	**2.69**	**1.75**	**1.99**
4	**3.42**	**2.29**	**2.205**
5	**4.13**	**3.34**	**2.84**
6	**5.14**	**4.34**	**3.42**

Method of Constant Stimuli was used to find the threshold of a positive spot. Using this method, observers would confirm existence or non-existence of a spot on the background of test images. The images were presented on a Dell Ultra Sharp 24 Ultra HD Monitor under the overhead lighting of D50 with and average CCT of 5111 K. The display was characterized using Day et al method [10]. Observes were seated in front of the display with a 30-cm distance while participating in the experiment. The images were presented in graphical user interface (GUI) of MATLAB in a random fashion and a mask was shown in between the images for 0.5 seconds. The GUI environment of the experiment is presented in Figure 3.

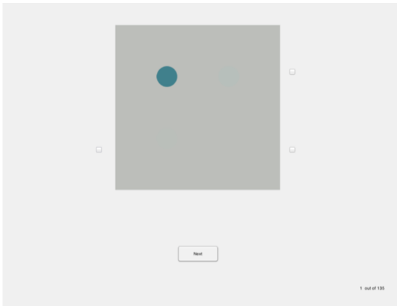

Figure 3. *GUI environment of the experiment*

If the observer saw a spot, they would check the box next to the spot, otherwise they would keep going. There were 135 images in total including 45 images for each of three backgrounds. For each background, there were 6, 12, 24 and 6 images having 0, 1, 2 and 3 test spots present, respectively.

In total, 20 observers participated in the experiment, including 6 females and 14 males with an average age of 20 and a median age of 16.

Before doing the experiment, observers would do the Ishihara color blindness test and one of the four sets of Farnsworth-Munsell 100 Hue Color Vision Test including the green-blue hues (Chips 43 to 63). All the observers correctly identified the numbers in the Ishihara plates and had fewer than five mistakes in the Farnsworth-Munsell test.

The analysis was done using Probit analysis method to find the T50 point for the color difference threshold. We assumed that the location of the test spot and number of spots present in the test image do not affect the results. Further analysis was done to validate these assumptions.

Results and Discussion

Probit analysis was used to find the T50 point (50% tolerance), which indicates the threshold [2]. Figure 4 shows the psychometric function for overall data. The line for Probit function was fitted using least square error method. The overall threshold value gained is 1.27 for all the backgrounds. The dots show the average response by observers and the dashed lines show the upper and lower 95% confidence bounds.

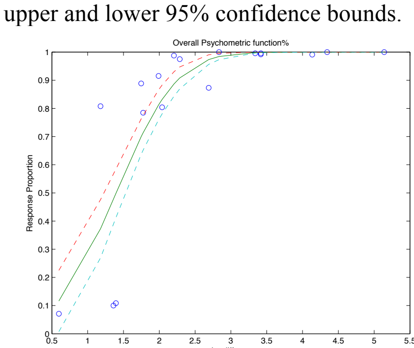

Figure 4. *Psychometric function for overall data*

The same analysis was done for each background. The obtained color difference threshold was 1.83, 1.01 and 1.56 for backgrounds 10%, 25% and 50% respectively.

Location dependence

The graphs below show the response ratio for each color difference level for locations 1, 2 and 3 in the test images.

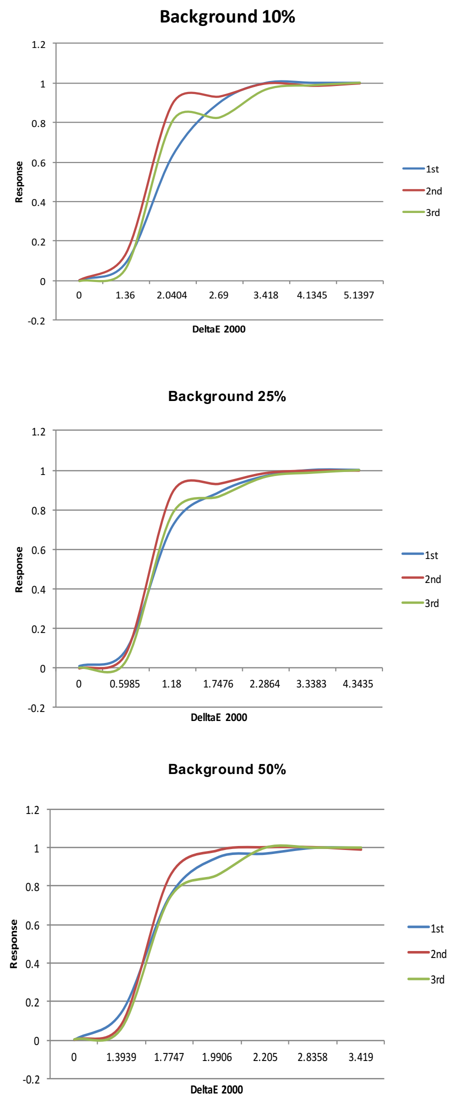

Figure 5. *Response ratio functions for background 10%, 25% and 50% for first (blue line), second (red line) and third spot (green line)*

In higher color differences the response difference for different locations is negligible. However, for smaller color differences the response difference for different locations is larger. For the first level of color difference location 2 has a higher response ratio for background 10%. Location 1 has higher response for background 25% and background 50%. The difference in locations between the one having the largest response and the one with the smallest is 7%, 5% and 8% for the first color difference levels for background 10%, background 25% and background 50%, respectively.

Number of test spots present in the image

The figure below shows the effect of number of test spots present in the image on noticing a spot.

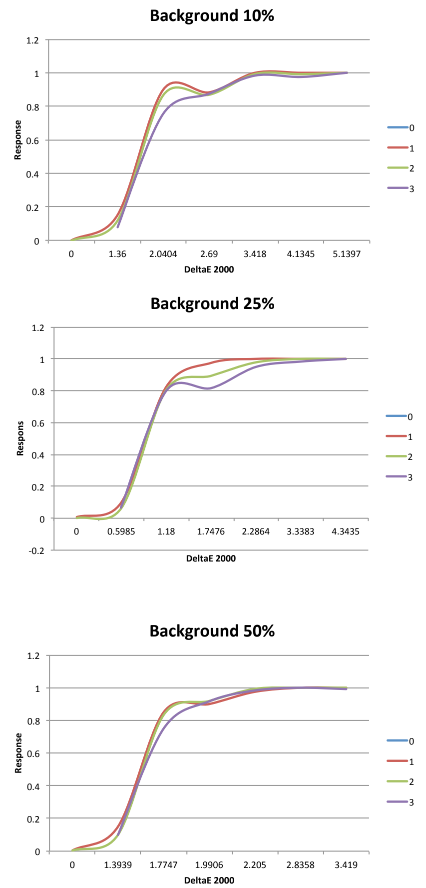

Figure 6. *Response ratio functions for background 10%, 25% and 50% for images with zero spots(blue line), one spot (red line), two spots (green line) and three spots (purple line)*

In all three cases, it is observed that, by increasing the number of test spots in the test image the response ratio decreases. For higher levels of color difference, the response difference is negligible. The difference for the first level of color difference is smaller than the difference for the second level. It is 7%, 3% and 5% for backgrounds 10%, 25% and %50 respectively. Therefore, the number of spots present in the image may affect the threshold. To further analyze this, probit analysis was used to determine the threshold for each background and for each type of test images including one, two or three test spots present. Table 2 shows the threshold values gained for each image type of each background. From table 2 we see that in most cases by increasing the number of test spots present, the threshold increases as well

Table 2. Threshold values gained for each image type of each background

Type of image	One spot	Two Spot	Three Spot
Background 10%	1.64	1.89	1.77
Background 25%	0.84	1	1.06
Background 50%	1.57	1.49	1.6

The difference between the minimum and maximum threshold values are 13.66%, 22.83% and 7.05% for background 10%, 25% and 50% respectively.

Phase II

Experimental

The experiment process of the second phase of the study was similar to the first phase. However, the test images were designed to better represent the result window of the real device. Figure 7 shows the schematic design of the images in the second phase.

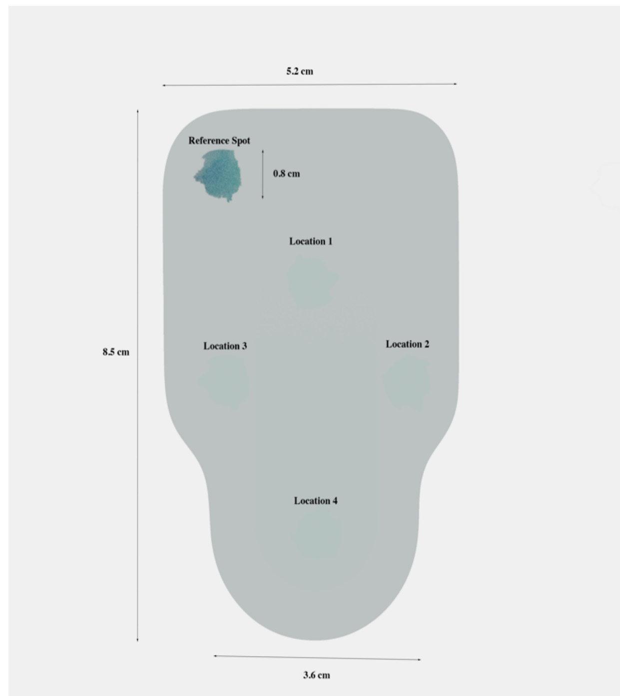

Figure 7. *The design of the images in the second phase of study.*

As it is seen in Figure 7, the shape of the background looks similar to the result window of the real device. The spots are not perfectly circular anymore and were designed to have different shapes. The images included a combination of random and fixed pattern noise (over the background and the spots). The reference spot is also not circular, does not have a uniform color and is noisy. This design was used for that observers do not expect a perfect spot in their observations. Two levels of the streak were designed and were present in some of the images. The streaks were designed as semi-ovals starting from lower of the spot at location 4 up to the location 1. The streaks were linear gradients of the starting color to the background color (in an upward direction). The starting color for the streaks had 1 and 1.5 DE00 color difference (similar color to test spots) for the streak level one and level two respectively. Also, in some images, the spots had a gradient of the color, starting from the assigned spot color up to the background color. This gradient was a linear gradient in a downward direction.

In total, 114 images were designed. Thirty images had no test spots, 24 images had one test spot, 36 images had two test spots and 24 images had three test spots present. Seventy-six images had streaks (38 images with level one streak and 38 images with level two streak), Forty-two images had spot gradient and 30 images had both streak and spot gradient.

Background 25% of the phase I study was chosen to represent the average color of the background Six levels of color difference for the spots and the background were assigned. These levels were 0.5, 1.0, 1.5, 2.0, 2.5 and 3 DE00. The colorimetric features were assigned after applying the noise and they represent average color of the measured area.

The spots had the same CIE L* values as the reference spot. The images were represented in graphical user interface (GUI) of MATLAB on the same display. Figure 9 show this GUI.

Figure 9. *Graphical user interface of the experiment*

Figure 5 shows the environment of the experiment designed in Graphical User Interface (GUI) with MATLAB. The procedure of the experiment was same as the first phase. For the spot located at location one, a checkbox was allocated on top of the image. An instruction image was shown to the observers in the beginning of the experiment. Observers were told to expect weak color spots and streaks and spot gradients in some images. They were told that images might include zero, one or multiple spots. Twenty observers participated in the experiment including 7 women and 13 men. Observers were mostly students with median age of 28 and an average age of 27.5. Before doing the experiment, observers would do the color blindness tests (same as phase one). All the observers were color normal.

Results and Discussion

The same analysis was done for phase II. Figure 10 shows the psychometric function of the overall data. The obtained threshold for overall data was 1.5 (DE00).

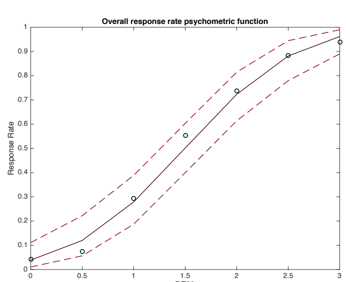

Figure 10. *The psychometric function of overall data.*

This threshold is for all the test images with or without streaks and spot gradients and including different number of spots present in the image. The dotted lines show the average response by observers and the dashed lines show the upper and lower 95% confidence bounds. We see that this threshold is higher than the threshold obtained for images with background 25% in phase I of the study (1.02 DE00). This shows that the imperfect appearance of the images increases the threshold. Figure 11 shows the psychometric function for images without any streaks or spot gradient.

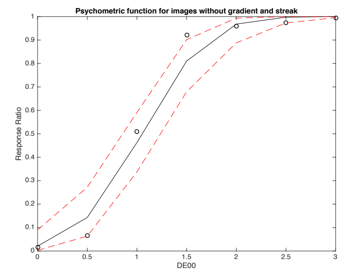

Figure 11. *The psychometric function for images without any streaks or spot gradients.*

Obtained threshold for images without any streaks or spot gradient was 1.05 DE00, which is very close to the threshold for the same background in phase I of the study (1.02 DE00). This shows that the effects of noise and imperfectness of the spot shapes on the threshold is negligible. However, in the experiment protocol, the observers were told to expect weak and non-perfect spots. This explanation in advance probably makes the observers more willing to respond positive to presence of a noisy and non-perfect spot. Next, the effect of streak on the threshold was analyzed. Figure 12 shows the psychometric function for images with streaks (both levels) and without any spot gradients.

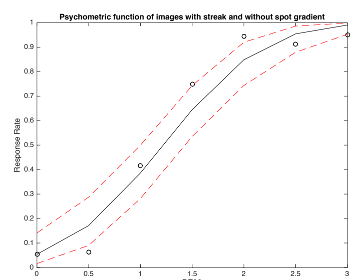

Figure 12. *The psychometric function for images with streak and without spot gradient.*

The obtained threshold for the images with streak (no spot gradient) was 1.22. This shows that existence of the streaks increases the threshold. Also the false alarm ratio (a false positive response by the observer when no spot is present) was higher for these images (0.05). However, the streak mostly affects the spot at location 4 and partly the spot at location 1. A separate analysis was done for spots at locations 1 and 4 in images with streak and no spot gradient. Figure 13 shows the psychometric function for these spots.

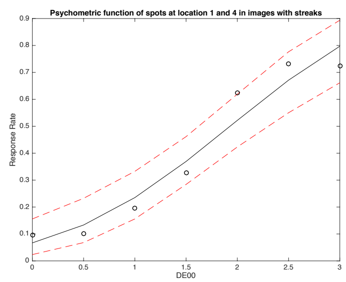

Figure 13. *The psychometric function for spots at locations 1 and 4 in images with streak and without spot gradient.*

From figure 13 we see that the false alarm ratio is higher for these spots (0.093). The highest response rate for color difference of 3 is about 0.7. These results imply that detecting a spot at these locations with streak present is a very hard task for the observer. The threshold for these spots is 1.93. This shows that the presence of the streak greatly affects the detectability of the spots at locations 1 and 4.

The effect of spot gradient was also analyzed. Figure 14 shows the psychometric function for images with spot gradient and without streak.

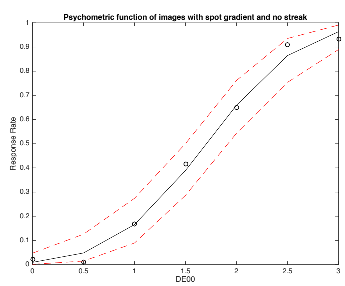

Figure 14. *The psychometric function for images with spot gradient and without streak.*

The obtained threshold for the images with spot gradient and without streak was 1.7. This result shows that the spot gradient also increases the threshold, making the detection of the spots even harder for the observers.

In the end, the worst situation for the image appearance was analyzed. The images that included both streak and spot gradient were analyzed to find the threshold for this situation. Figure 15 shows the psychometric function for these images.

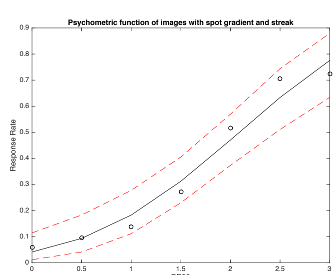

Figure 15. *The psychometric function for images with streak and spot gradient.*

From figure 12 we see that again the highest response is dropped to 0.7. The false alarm rate is 0.05 which is lower than that for the spots at locations 1 and 4 in the images with streak. The obtained threshold for these images was 2.09, which is the highest threshold obtained among other situations. This situation is the hardest situation in detecting the spots for the observers.

Conclusion

The study was performed in two phases. In the first phase, images were designed to reperesent the simplified version of the real images. Six color difference levels for each of the three different backgrounds were assigned. Using Probit analysis, the gained threshold value was 1.83, 1.01 and 1.56 for backgrounds 10%, 25% and 50% respectively. In order to verify the results, the data were analyzed further for location of spots and number of test spots present in the images. Analysis showed that the difference in the response ratio by location is negligible. However, the response ratio was affected by the number of spots present in the test image.

In the second phase the effects of the study, noise, imperfect spot shapes, presence of streak and presence of spot color gradient on was verified. For this purpose, new images were designed to better represent the result window of the real device. Twenty observers participated in the experiment and results were analyzed using Probit analysis method.

The results for images without any streak or spot gradient showed a similar threshold (1.05) to the obtained threshold for images with the same background color in phase one of the study (1.02). This shows that noisy appearance of the image and imperfectness of the spot shapes does not affect the detectability of the spots when the observers expect to see imperfect spots. The presence of the streak increases the threshold to 1.22. This effect is more noticeable for spots located at locations 1 and 4 with a threshold of 1.93. The impact of spot gradient on the threshold was also noticeable. The threshold for images containing spot gradient and not containing streak was 1.7. The worst situation for the image appearance was also analyzed. Images which included both streak and spot gradient had a threshold of 2.09. This threshold was highest among all situation.

Altogether, the noisy appearance of the image and imperfectness of the shape of the spots does not affect the threshold if the observers are informed about it. The presence of streak and spot gradient increases the threshold. The effect of the streak on threshold is more noticeable for spots at locations 1 and 4. The most difficult detection belongs to images that contain both streak and spot gradient with a threshold of 2.09.

Future work

In this work the ideal situation and more realistic situations were studied with observers screened for color vision deficiencies. One subject that might be of interest is to verify the impact of visual acuity on threshold detection. As the blurring that happens with low vision accuracy (for older people or people with low visual acuity) reduces the contrast between the spots and the background, detection of the spots might be more difficult. Also, the data of the two steps can be used to model the threshold objectively, similar to Watson et al [7]

Acknowledgements

This study was funded by IDEXX company. Authors would like to thank Rob Farrell and IDEXX lab researchers for providing experimental data. Authors also thank all the observers who participated in the experiment.

References

[1] I. Bramão, A. Reis, K. M. Petersson, and L. Faísca, "The role of color information on object recognition: A review and meta-analysis," *Acta Psychol. (Amst.)*, vol. 138, no. 1, pp. 244–253, Sep. 2011.

[2] S. D. Maddison, "Pregnancy testing method," US20100129935 A1, 27-May-2010.

[3] R. S. Berns, F. W. Billmeyer, and M. Saltzman, *Billmeyer and Saltzman's principles of color technology*. Wiley, 2000.

[4] A. Daxer, *Apparatus for testing color discrimination in the human visual system*. Google Patents, 1999.

[5] A. B. Watson, "Visual detection of spatial contrast patterns: Evaluation of five simple models.," *Opt. Express*, vol. 6, no. 1, pp. 12–33, 2000.

[6] A. B. Watson, "31.1: Invited Paper: The Spatial Standard Observer: A Human Vision Model for Display Inspection," in *SID Symposium Digest of Technical Papers*, 2006, vol. 37, pp. 1312–1315.

[7] T. Carney, S. A. Klein, C. W. Tyler, A. D. Silverstein, B. Beutter, D. Levi, A. B. Watson, A. J. Reeves, A. M. Norcia, C.-C. Chen, and others, "Development of an image/threshold database for designing and testing human vision models," in *Electronic Imaging'99*, 1999, pp. 542–551.

[8] T. Carney, C. W. Tyler, A. B. Watson, W. Makous, B. Beutter, C.-C. Chen, A. M. Norcia, and S. A. Klein, "Modelfest: Year one results and plans for future years," in *Electronic Imaging*, 2000, pp. 140–151.

[9] E. A. Day, L. Taplin, and R. S. Berns, "Colorimetric characterization of a computer-controlled liquid crystal display," *Color Res. Appl.*, vol. 29, no. 5, pp. 365–373, 2004.

[10] G. A. Gescheider, *Psychophysics: the fundamentals*. Psychology Press, 2013.

[11] D. J. Finney, *Probit analysis.* JSTOR, 1952.

Author Biography

Nargess Hassani *is currently a third-year Ph.D. student in Munsell Color Science Lab (MCSL) at Rochester Institute of Technology (RIT), under the advisement of Dr. Michael Murdoch. She previously worked in Disney Animation Studios as a technical graduate associate in summer of 2016. Nargess graduated from Amirkabir University of Technology (AUT) with a B.Sc. in polymer engineering/color science and technology. Her main research interests are color perception, augmented reality, color appearance, and animation.*

Susan Farnand *is a Visiting Assistant Professor in the Program of Color Science at the Rochester Institute of Technology. Her research interests include human vision and perception, color science, cultural heritage imaging and 3Dprinting. She received her BS in engineering from Cornell University, her Masters in Imaging Science and her PhD in Color Science from the Rochester Institute of Technology. She began her career at Eastman Kodak, designing and evaluating printer systems. She is Publications Vice President of the international Society of Imaging Science and Technology and serves as an Associate Editor for the Journal of Imaging Science and Technology. She participates in several Standards efforts including ISO TC 42 JWG26 Archival Imaging.*