

Perceptual Study of Real-World Colors on Ultra-WCG Displays

Farnaz Agahian¹, Sofie R. Herbeck^{1,2}, Dale Stoltzka¹

¹Samsung Display America Lab, San Jose, CA, USA, ²Rochester Institute of Technology, NY, USA

Abstract

This study builds upon our previous work, where we analyzed the range of real-world colors and identified images containing colors that exceed the boundaries of legacy color gamuts such as sRGB and DCI-P3, making them difficult for traditional displays to render accurately. In our current research, we conducted a series of visual experiments to evaluate perceptual differences and viewer preferences when such images are displayed on ultra-WCG displays compared to standard-gamut displays. Our findings indicate that observers could consistently distinguish between images shown on an ultra-WCG display and the same images calibrated to sRGB. The perceptual difference between DCI-P3 and ultra-WCG was notably smaller, resulting in lower detection rates that were more content-dependent. Overall, observers showed a strong preference for the ultra-WCG display, regardless of the viewing condition or the image content.

Background and objectives

Over the last several years, advancements in display technology have played a crucial role in improving visual fidelity, particularly through the development of wider color gamuts. A Wide Color Gamut (WCG), which exceeds the limitations of the standard sRGB color space, allows displays to render colors with greater accuracy and vibrancy—bringing digital content closer to what the human eye naturally perceives. This leap in color performance is often linked to higher image quality in both consumer and professional settings. Among the forefront technologies enabling this shift are OLED and QD-OLED displays, which support ultra-WCGs (referring to gamuts larger than DCI-P3), allowing for richer reds, cyans and deeper greens, and overall, more lifelike visuals. These capabilities are especially important for High Dynamic Range (HDR) content, which relies on a broader color range to achieve its intended visual impact. As a result, OLED and QD-OLED displays with ultra-WCGs are also widely regarded as top-tier choices for delivering premium HDR experiences.

An ultra-WCG display is particularly beneficial for applications such as photography and video production, where precise color representation is essential. Such displays are also beneficial for computer-generated content. By offering a broader range of colors, they enhance realism and visual appeal in applications like gaming and animation. This is especially important for creating believable characters, environments, and effects to make the content more convincing and enjoyable for users. In digital arts and 3D renderings like architectural visualizations and product designs, an ultra-WCG display allows artists and designers to create and view their work with greater color richness and depth. This results in more vibrant and nuanced images, enabling a higher degree of artistic expression and fidelity.

There are primarily two strategies to fully utilize a wider display color gamut:

1. Displaying content with colors that extend well beyond sRGB
2. Enhancing sRGB content by stretching and saturating colors.

Our current research focuses specifically on how users perceive and appreciate the broader color range of real-world scenes when they are shown on a QD-OLED display with an ultra-WCG. This study builds upon our previous work where we analyzed a collection of 1492 images from real-world scenes, identifying images with a range of colors that goes far beyond the gamuts of sRGB and even DCI-P3 [1]. To fully understand the benefits of ultra-WCG displays, a selection of these images was used to assess the perceptual significance of viewing them in a standard versus an ultra-wide gamut. A visual experiment was conducted to investigate the core research questions:

1. Do observers perceive differences between images rendered in sRGB or DCI-P3 and those presented in the QD-OLED display's ultra-wide color gamut?
2. When they see a difference, do they prefer one over the other?

We further explored whether the patterns identified in the primary questions differed across various lighting conditions or demographic factors such as age, gender, and cultural background.

Methods

The display used for the study was a Samsung S95B TV, which has a native gamut larger than DCI-P3 with the largest increases in the green/cyan and red areas (Figure 1).

Image Selection

From the findings of previous work [1], we selected a set of six images that exhibited the highest concentration of out-of-gamut pixels in the DCI-P3 color space. The out-of-gamut colors in these images spanned a variety of hues, including cyan, green, orange, and red. We observed two distinct distribution patterns for these colors. In some images, such as “Beads”, the out-of-gamut pixels were dispersed throughout the scene, whereas in others, like “Roses” they formed large, contiguous areas. This variation is critical for evaluating whether viewer perception and preference are influenced

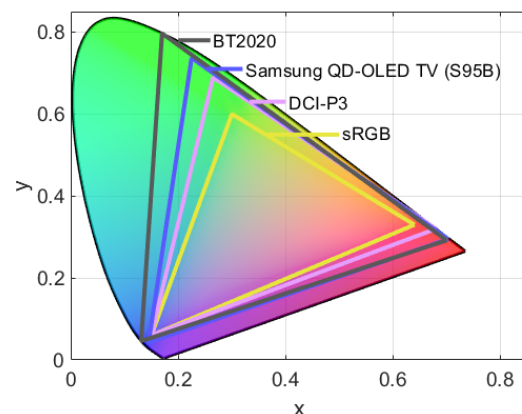


Figure 1. Comparison of multiple color gamuts shown in the CIE1931xy Chromaticity diagram.

by the spatial arrangement of the out-of-gamut pixels—whether they are concentrated on a large object or scattered across smaller ones. Our selection criteria also ensured a mix of natural and man-made objects to account for potential differences in user preference.

Image Processing and Calibration

To prepare the images for the visual experiment, we started with the original multispectral or hyperspectral scene data. These images capture reflectance information across multiple spectral bands, enabling accurate color calculations under D65 illumination, free from in-camera enhancements. To create sRGB and DCI-P3 versions, we calibrated the images by mapping the out-of-gamut colors to fit within the respective color spaces. The “Cups” image was cropped to direct observer focus to the area with gamut-dependent color changes. During the visual study, participants viewed the wide-gamut, sRGB-calibrated, and DCI-P3-calibrated versions of each image on a single QD-OLED display.

Experimental Setup

In the experiment, six images were evaluated across all three possible color gamut pairings: sRGB vs. DCI-P3, sRGB vs. TV (S95B), and DCI-P3 vs. TV (S95B). The order of presentation was randomized to minimize bias. During each trial, observers viewed a single image centered on a black background, alternating back and forth between two versions rendered in different color gamuts. For example, observers saw two versions of the same image on Samsung QD-OLED TV—one in its original color range (limited to TV native color gamut) and the other calibrated and mapped to fit within a commonly used legacy gamut, i.e. sRGB or DCI-P3. Some trials served as controls, where both images were identical (calibrated to the same gamut), allowing us to assess whether observers could still perceive a difference.

In each trial, participants were first asked, “Do you see a difference between these two versions?” If they answered “Yes,” they were then asked to rate the magnitude of the difference as small, medium, or large (see Table 1 for definitions). When a difference was perceived, observers were also asked which version they preferred—or whether they had no preference.

All these trials were repeated in two lighting conditions: an average ambient lighting condition, consisting of two floor lamps with 2700K bulbs placed on either side of the TV which produced an illumination level of 100 lux, and a dark condition. These combinations led to [6 images] x [4 gamut pairs, including a control] x [2 lighting conditions] = 48 trials. In total, 31 observers completed the experiment. The group was composed of 15 women and 16 men with ages ranging from mid-20s to mid-60s.

Table 1: Difference rating definitions.

| | |
|-------------------|--|
| Small difference | The colors are almost identical. I have to look very closely to see a difference. |
| Medium difference | I can definitely see the difference, but it doesn't change the whole image very much. The two images still have similar enough colors. |
| Large difference | The two images are clearly different, and I can see the difference right away, at a glance. |

To ensure a well-rounded evaluation, observers were selected from both technical and non-technical backgrounds. The technical observers had prior experience with color science and display-

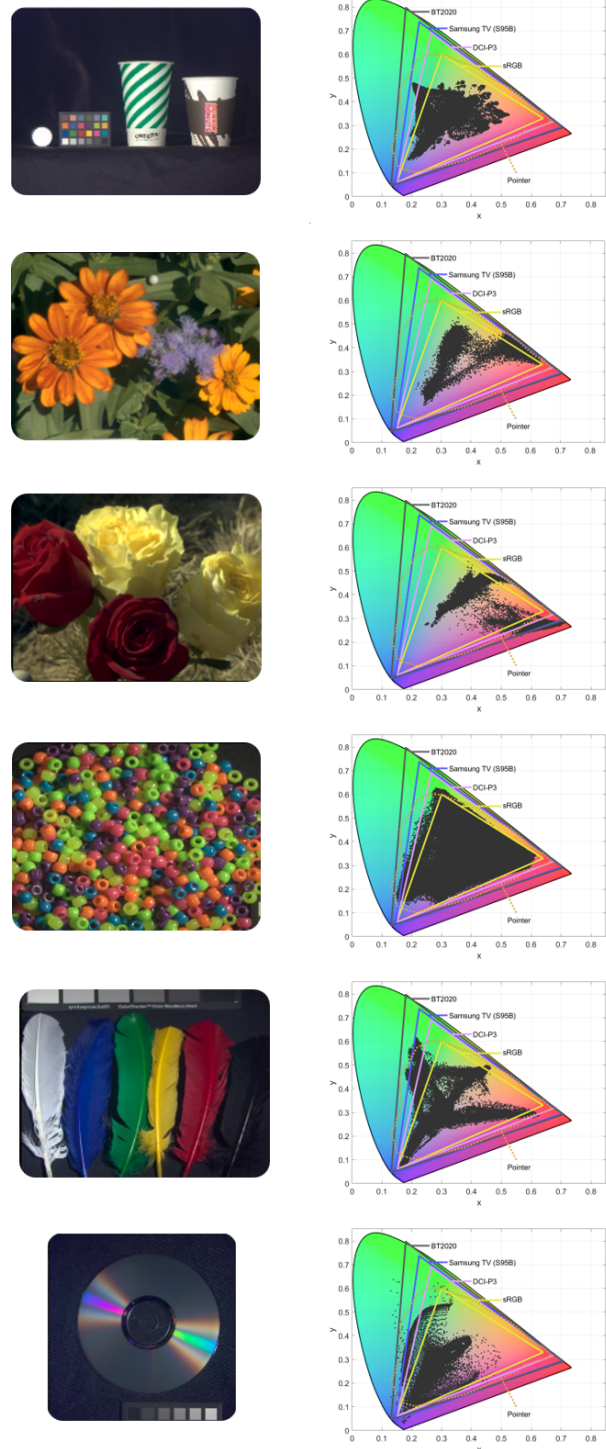


Figure 2. Stimuli & Pixel Distribution in the CIE 1931 xy chromaticity diagram. The images were sourced as follows: Beads, CD, and Feathers were sourced from [2]; Cups was sourced from [3]; Zinnias Flowers and Roses were sourced from [4].

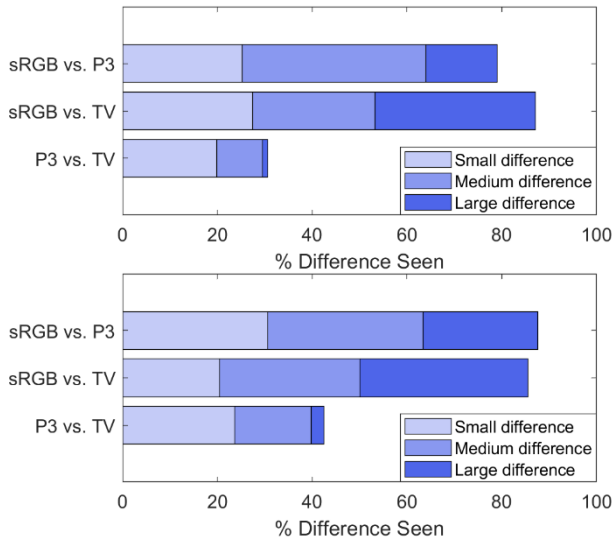


Figure 3. Overall difference ratings for both the average (top) and dark (bottom) lighting condition.

related projects, while the non-technical participants came from various departments outside the technical domain. This mix allowed for a broader range of perspectives during the visual assessment. Prior to the study, observers were screened for red-green forms of color vision deficiency (CVD) using the 14-plate Ishihara Color Test. One observer was known to be red-green colorblind, which was confirmed via his responses to the Ishihara test; another had normal responses but self-reported as having been previously diagnosed with protanomaly. Their results will be discussed in the following section.

Results & Discussion

Difference Ratings

As mentioned earlier, to evaluate whether observers could perceive differences among three versions of each image: (1) the original mapped to a QD-OLED TV with ultra-WCG, (2) sRGB-calibrated, and (3) DCI-P3-calibrated, we conducted a pairwise comparison experiment. In each pair, participants were asked whether they could perceive a difference between the two images (definitions are given in Table 1). The percentage of difference detected responses was recorded for each pairing, based on a total of 186 observations per gamut pair. The results show varying levels of detectability among the image pairs for both average and dark lighting conditions (see Figure 3 for a visualization):

a) Average condition

- sRGB vs. DCI-P3: Observers were able to detect a difference in 79% of the comparisons.
- sRGB vs. TV: A difference was detected in 87% of the cases.
- DCI-P3 vs. TV: A difference was detected in only 30% of the comparisons.

b) Dark condition

- sRGB vs. DCI-P3: Observers were able to detect a difference in 88% of the comparisons.
- sRGB vs. TV: A difference was detected in 85% of the cases.
- DCI-P3 vs. TV: A difference was detected in only 42.5% of the comparisons.

These findings suggest that, in a dark viewing condition, sRGB vs. DCI-P3, and sRGB vs. QD-OLED TV (with ultra-WCG), are perceptibly different to most observers with high discrimination rates above 85%. In contrast, the lower detection rate between images in DCI-P3 vs. QD-OLED TV indicates that observers often could not distinguish between them. The results also show that a dark viewing condition may heighten sensitivity to certain differences, particularly when distinctions are subtle (e.g., between DCI-P3 vs. TV). However, when the differences are more pronounced (as between sRGB vs. TV), lighting has little impact on observers' detection ability.

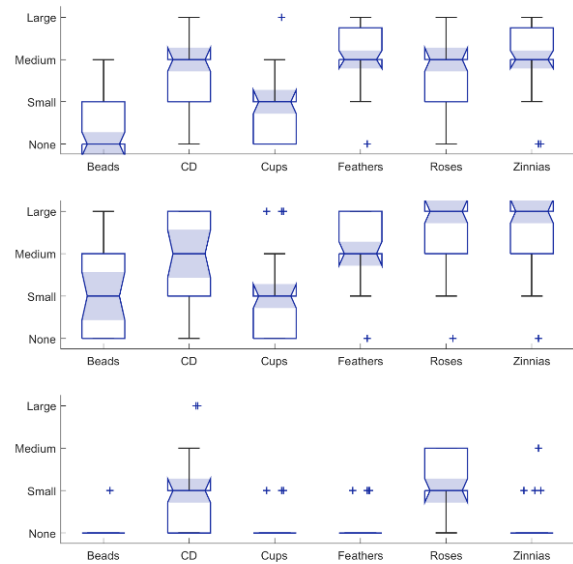


Figure 4. Box plots show the distribution of observers' gamut difference ratings in the "average" viewing condition across the six images used. From top to bottom: sRGB vs. DCI-P3, sRGB vs. TV, and DCI-P3 vs. TV.

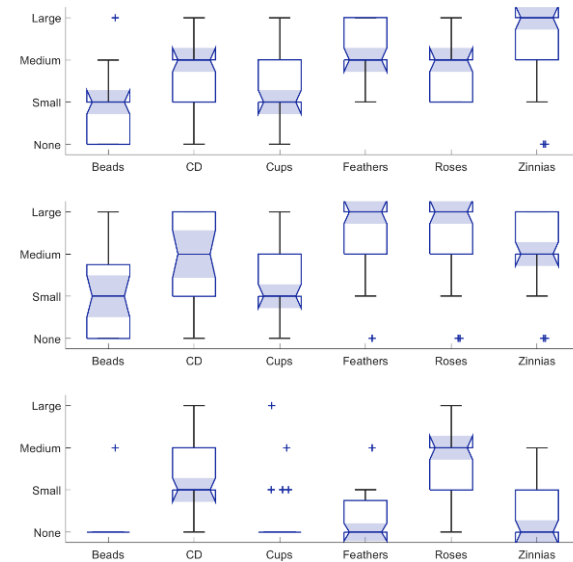


Figure 5. Box plots show the distribution of observers' gamut difference ratings in the "dark" viewing condition across the six images used. From top to bottom: sRGB vs. DCI-P3, sRGB vs. TV, and DCI-P3 vs. TV.

The ratings for each image are presented in the box plots in Figures 4 and 5, where the shaded regions highlight the median of each image's difference ratings. A lack of overlap between two images' shaded regions indicates a significant difference between their medians at the 5% confidence level.

The box plots show that, although changing the lighting condition from average to dark does not lead to a large overall difference in perceived differences, there are certain images where this change allows observers to detect differences that were not noticeable under average lighting. "Feather" and "Zinnias" are two prime examples.

When comparing the TV to the DCI-P3 gamut, the "CD" and "Rose" images elicited the largest reported color differences, particularly in dark viewing conditions. This suggests that the ultra-wide color gamut of the TV makes a noticeable difference in rendering specific hues, such as the cyan on the CD and the deep red on the rose petals. These differences were very salient to the observers.

Across all three comparison modes in both lighting conditions, "Feather," "Rose," and "Zinnias" consistently showed larger reported color differences than "Beads." This outcome can likely be attributed to the spatial distribution of out-of-gamut pixels. As detailed in the Methods section, the six selected images were chosen not only for their high percentage of pixels extending beyond the DCI-P3 gamut but also for the distinct ways in which these pixels were arranged. In images like "Beads," the out-of-gamut pixels are scattered throughout the scene, whereas in others, such as "Rose" and "Feather," they form a single, large patch. The concentration of color on a single object appears to be a factor that led observers to perceive a larger difference in these images across nearly all comparison modes. To gain a more complete understanding of the impact of an image's color distribution on viewer perception, further research is needed to isolate the influence of pixel distribution from that of hue dependency.

CVD Observers

The two CVD observers did not seem to find it more difficult to detect differences than non-CVD observers. There were slight variations between the two groups, such as the CVD observers tending to rate fewer differences as "Large" than non-CVD observers, and a lower detection rate for the DCI-P3 vs. TV comparisons (by approx. 17%). However, these differences were not deemed significant—especially given that only two out of 31 observers had some form of CVD (and further, these two did not have the same form as each other), it is difficult to draw conclusions about CVD observers from what might merely be individual variation. It is plausible to expect that CVD observers would not have a significantly different experience of color for this detection task: gamut differences primarily manifest in richness or vividness of the colors displayed, which are aspects of perception that are not largely affected for CVD individuals.

Preferences

The Bradley-Terry model [5] is a statistical approach used in this research to estimate the relative preference strengths of different color gamuts based on paired comparison data. In this study, three versions of an image mapped to three different color gamuts were compared in pairs, and observers indicated their preference.

In the average viewing condition (Figure 6), the preferences were fairly balanced: sRGB was preferred over DCI-P3 in 30.1% of trials, while DCI-P3 was preferred in 38.2%; sRGB and QD-OLED TV (S95B) were nearly tied, with sRGB preferred 38.2% of the time and TV 37.1%. TV also slightly outperformed DCI-P3, being

chosen in 12.9% of comparisons compared to P3's 10.8%. The Bradley-Terry model estimated that sRGB and TV had similar preference strengths, with DCI-P3 slightly lower.

In the dark viewing condition, the preferences were much more one-sided: sRGB was preferred over DCI-P3 in only 18.3% of trials, and over TV (S95B) in 12.4%, while DCI-P3 and TV were chosen over sRGB in 57.0% and 65.1% of cases, respectively. TV (S95B) also dominated DCI-P3, with preferences of 27.4% vs. 6.5%. The remaining percentage in each pairwise comparison corresponds to cases where observers expressed no preference, which were excluded from the Bradley-Terry estimation. The model translated these observed choice proportions into meaningful rankings: in the average viewing condition, sRGB and QD-OLED TV (S95B) had similar estimated strengths with DCI-P3 slightly lower, while in the dark viewing condition, QD-OLED TV (S95B) was strongly preferred, followed by DCI-P3, with sRGB ranked lowest.

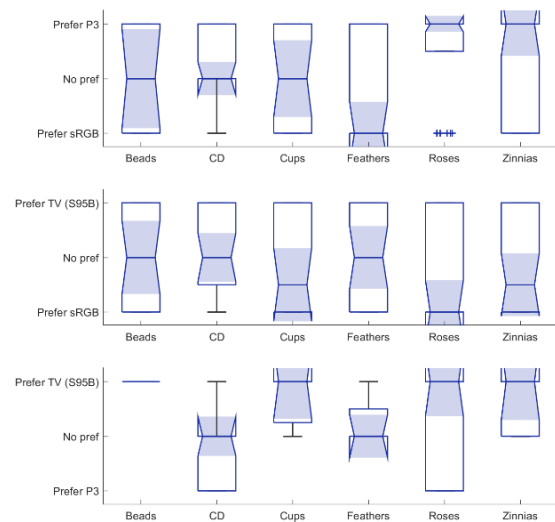


Figure 6. Box plots present the breakdown of observers' color gamut preferences in the "average" viewing condition across different images. From top to bottom: sRGB vs. DCI-P3, sRGB vs. TV, and DCI-P3 vs. TV.

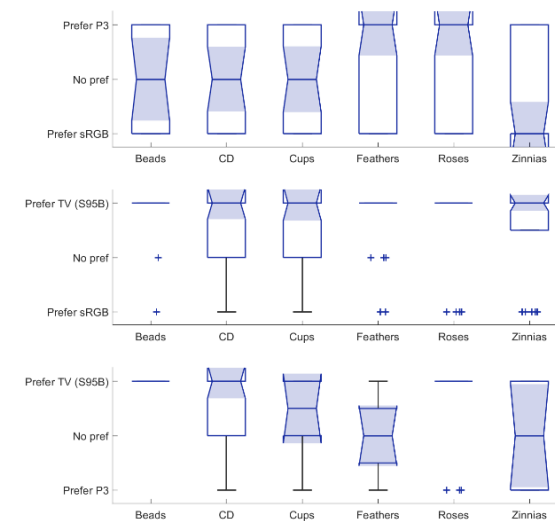


Figure 7. Box plots present the breakdown of observers' color gamut preferences in the "dark" viewing condition across different images. From top to bottom: sRGB vs. DCI-P3, sRGB vs. TV, and DCI-P3 vs. TV.

Figures 6 and 7 present the breakdown of observers' color gamut preferences across different images. No clear correlation was found between preferences for natural objects (e.g., flowers) and preferences for man-made objects (e.g., "Cups" and "CD"), suggesting that observers responded similarly to both categories.

In our analysis, we looked at how demographics like age and gender might have impacted the results. We didn't find a clear correlation or any significant patterns, so we chose not to include a detailed breakdown based on these in this paper.

CVD Observers

There was more of a difference between the two CVD observers' preferences and the non-CVD observers' preferences than there was between their difference detection abilities. The CVD observers responded "No preference" at a higher rate for the sRGB vs. DCI-P3 and sRGB vs. TV pairs than non-CVD observers, although they still preferred the wider gamut in each comparison a majority of the time (42% and 50%, respectively). For the DCI-P3 vs. TV comparisons, their preferences for DCI-P3 and for TV were equal (at 40%). This could be related to their slightly lower ability to detect a difference between these two gamuts, as a smaller perceptual difference could understandably lead to noisier or more ambiguous preferences. Regardless, as mentioned previously in the Difference Detection section, it is difficult to draw significant conclusions about CVD observers as the very small sample size means any differences observed may be due to individual variation rather than an actual population difference between CVD and non-CVD individuals.

Conclusion

Our research shows that while the sRGB/DCI-P3 color gamuts cover most real-world colors, ultra-WCG displays offer perceptible and meaningful improvements in reproducing certain hues, particularly deep reds and cyans. Building on our previous quantitative analysis [1], this subjective finding indicates that extending the color gamut beyond DCI-P3 is not only noticeable to viewers but also enhances the overall viewing experience. These insights are valuable for panel manufacturers and future color gamut standards, emphasizing the need to expand coverage in specific regions of color space where traditional gamuts fall short.

References

- [1] F. Agahian and D. Stoltzka, "From Scene to Display: A Quantitative Analysis of Real-World Color Gamut," in *SID Symposium Digest of Technical Papers*, San Jose, California, vol. 56, no. 1: pp. 895-898, 2025.
- [2] F. Yasuma, T. Mitsunaga, D. Iso, S.K. Nayar, "Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans Image Process.*, vol. 19, no. 9, pp. 2241-53, 2010.
- [3] I. Choi, D.S. Jeon, G. Nam, D. Gutierrez, M.H. Kim, "High quality hyperspectral reconstruction using a spectral prior," *ACM Trans Graph*, vol. 36, no. 6: pp. 1-13, 2017.
- [4] D.H. Foster, A. Reeves, "Colour constancy failures expected in colourful environments," *Proc Biol Sci.*, vol. 289, no. 1967, p. 20212483, 2022.
- [5] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. The method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324-345, 1952.

Author Biography

Farnaz Agahian is a Color Scientist at Samsung Display America Lab, focusing on the standardization of HDR displays. She holds a PhD in Color Science from Tehran Polytechnic and an MS in Computer Science from Simon Fraser University.

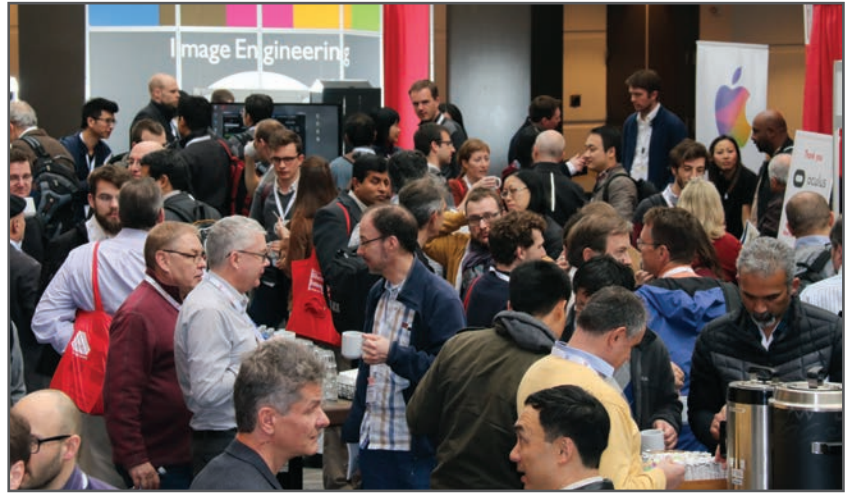
Sofie Herbeck received their B.A. in Computer Science & Theater and Performance Studies from the University of California, Berkeley, where they also worked in a vision science lab on the project that produced the novel color percept, 'olo.' They are currently a Color Science Ph.D. candidate in the Munsell Color Science Laboratory.

Dale Stoltzka is a senior principal engineer at the Samsung Display America Lab who leads the display standardization efforts and oversees the Color Laboratory. He has an MS in materials science from Cornell University.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



- **SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •**
- **INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •**

www.electronicimaging.org

