https://doi.org/10.2352/CIC.2023.31.1.6


# The perceptibility of color differences in continuous transitions

*Ján Morovič, Peter Morovič, HP Inc.*

## Abstract

*Whether two stimuli appear to be of different colors depends on a host of factors, ranging from the observer, via viewing conditions to content and context. Previously, studies have explored just noticeable difference thresholds for uniform colors viewed with or without spatial separation, for complex images and for fine features like lines in architectural drawings. An important case that has not been characterized to date is that of continuous color transitions, such as those obtained when selecting two colors and generating a sequence of intermediate colors between them. Such transitions are often part of natural scenes (e.g., sunsets, the sky, curved surfaces, soft shadows, etc.) and are also commonly used in visual design, including for backgrounds and various graphical elements. Where the just noticeable difference lies in this case will be explored here by way of a small-scale, pilot experiment, conducted in an uncontrolled, on-line way. Its results suggest a threshold in the region of 0.5 to 0.8 ΔE2000 for the few stimuli evaluated in the pilot experiment reported here and indicate a behavior that is in the region of viewing solid colors without a gap. A pilot verification with complex images also showed thresholds with a comparable range.*

## Introduction

The perceptibility of differences between color stimuli is not a property of only the stimuli in question but of where and how they are present and of how and by whom they are viewed. Beyond obvious factors like viewing distance and intensity of ambient illumination (in the case of surface color), where extremes can immediately be recognized as leading to very different experiences, a host of more subtle considerations are at play too.

One way to understand the differences is to ask what level of color difference, e.g., as computed using CIE ΔE2000 corresponds to a just noticeable difference (JND) under different conditions. Note that this is a much more narrowly posed question than one about how magnitudes of color difference compare under such conditions, which would also need to deal with the additivity failure of color differences over longer distances in color space [1].

The obvious starting point here are the conditions based on which that ΔE metric was derived, which is the viewing of color stimuli that are uniform, take up a substantial proportion of an observer's field of view, that are presented on a mid-gray background and that have a clear spatial separation between them. Under these conditions ΔE2000 is designed to result in a value of 1 for color pairs that are just noticeably different from each other [2]. Even just removing the separation between such a pair of stimuli results in a change of the JND threshold to 0.3 ΔE00 for colors that differ in lightness, while keeping it unchanged for color differences in only chroma and hue [3]. Over a broad sampling of color space this leads to a threshold of around 0.8 ΔE00, which has also been found in the context of tooth-colored restorative material [4].

Going in the other direction of adding complexity to the stimuli by using photographic images instead of solid patches then leads to the JND threshold going up to around 2 ΔE00 for the 95th percentile of per-pixel differences [5]. Finally, reducing the angular subtense even just of one of a pair of simple stimuli, e.g., when comparing thick and a thin colored lines then leads to the threshold increasing significantly to around 7 ΔE00 [6].

The question that the present paper will address is that of perceptibility thresholds in color transitions, where the intent is for change to be gradual and continuous between two colors and for no edges to appear in-between. What then is the level of color difference that mustn't be exceeded for that experience of continuity to hold? A small-scale, on-line pilot experiment will therefore be introduced next, and its results will be analyzed to provide a first approximation of the color difference perceptibility threshold in color transitions. The images evaluated in the experiment will then also be analyzed using a model of visual difference perception. Finally, another small-scale, on-line pilot experiment is conducted using photographic images instead of synthetic ones that also contain a significant proportion of transitions, and its results are analyzed as well.

## Experimental setup and analysis

The on-line experiment conducted to explore the perceptibility of color differences in continuous transitions that will be introduced here follows the same approach as previous work on perceptibility of color differences between thin lines [6]. It sacrifices control over stimulus, viewing conditions, experimental procedure, and observer characteristics in favor of a much larger number of participants that would be feasible in a laboratory setting. This is also facilitated by keeping the experimental procedure as simple and quick to complete by observers, and it has meant collecting a minimum of data rather than imposing an extended burden on participants. This clearly introduces limitations and means that an analysis of contributing factors is often impossible, but it leads to a first sketch that can then be followed up under controlled conditions.

The stimuli in this experiment were three sets of color transition images between black and white, red and cyan in turn, defined in 8-bit sRGB space. The choice of 8 bits for representing the individual stages of a transition imposes constraints on the smallest steps available at different locations in color space and means that a transition that is intended to have a certain, fixed step size will instead have a distribution of color differences. In a laboratory setting 16-bit precision would lead to more uniform transition steps, but it would make participation in an on-line experiment much more restrictive.

Each of the transitions here was generated by starting from black (sRGB=[0,0,0]) and evaluating unit sRGB increments towards a choice of target color in terms of its ΔE2000 versus the previous step in the transition. If the

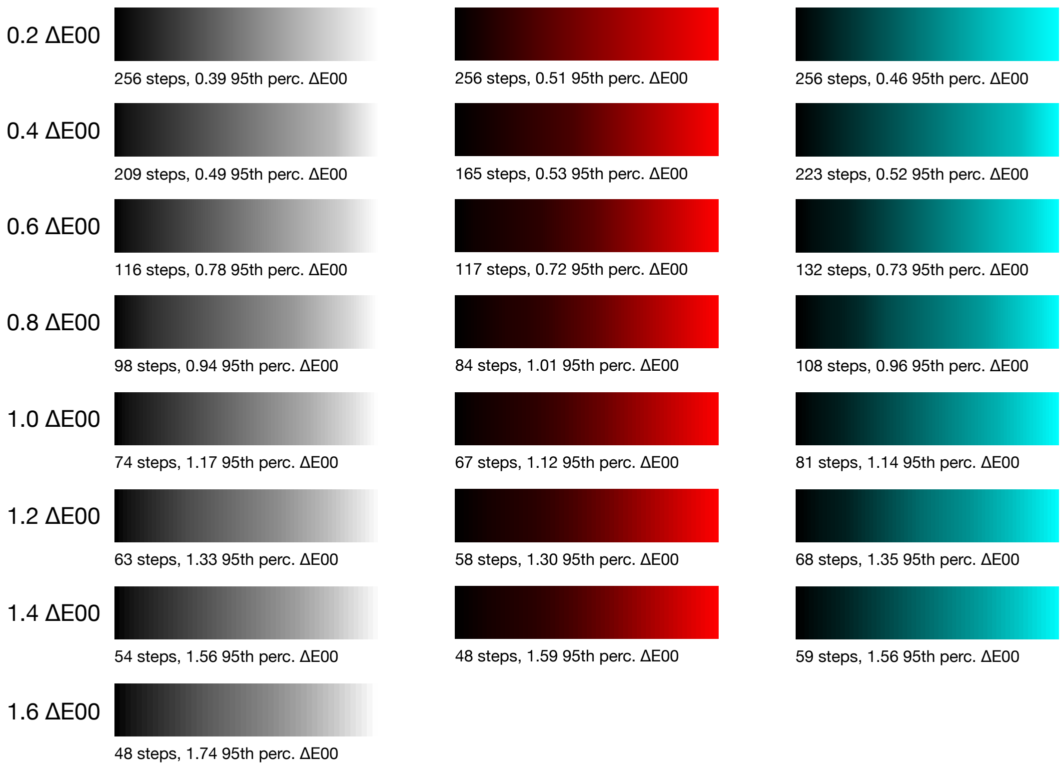


*Figure 1. Transition images showing number of steps at each target ΔE2000 threshold level (left column) and the 95th percentile of the ΔE2000 steps (under corresponding transition image).*

current increment exceeded the chosen threshold, it and the previous increment were compared to the threshold and the sRGB values closer to the threshold were added to the transition. The process was repeated until the target color was reached. This resulted in varying numbers of steps between black and each of the three target colors, which were then used to generate transition images of fixed 1000x200 pixel resolution using pixel replication. Fig. 1 shows the 22 transition images used in the experiment where each image indicates the target ΔE2000 threshold used to generate it and the number of transition colors that resulted in. Since the images were restricted to 8 bits, the resulting transitions had a distribution of color differences, of which the 95th percentile is also shown in Fig. 1.

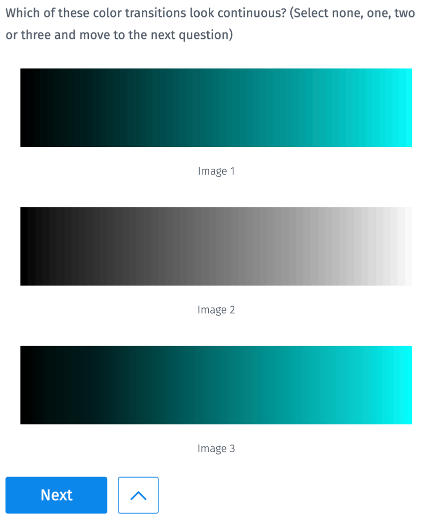


*Figure 2. Stimulus presentation and instructions*

33 stimuli were shown to observers, consisting of the 22 images from Fig. 1 with 11 of them repeated (2, 3, 7, 8, 11, 12, 14, 15, 17, 18 and 20, numbered sequentially by column from top left as shown in Fig. 1), which allows for quantifying intra-observer repeatability. Sets of three stimuli were shown to each observer in a fixed, randomized order and they were asked: "Which of these color transitions look continuous? (Select none, one, two or three and move to the next question)", as shown in Fig. 2. Showing 3 stimuli at the time meant that the context in which each set of 3 judgments was made shifted from set to set. However, even showing all 33 together would have meant shifting contexts since taking in all stimuli in one go is practically infeasible for an observer and they would still be considering a stimulus in its more restricted neighborhood. Choosing 3 at a time also made the task simpler for the observer, which was important for maximizing participation. The experiment was made available using the questionpro.com platform.

Image display and viewing conditions were entirely uncontrolled and unknown. Aspects like stimulus size; display size, brightness, contrast, gamut, or calibration state; viewing distance, angle, temporal duration; observer color vision status, adaptation state, age, sex, gender, or color evaluation experience were also unknown. This means that the results provide no basis for understanding the impact of these or many other experimental variables and only offer data about the cumulative and combined effect of the various and unknown conditions under which the individual responses were provided.

461 observers from 39 countries (Fig. 3) have completed the experiment, with a median time taken to do so of 3.1 minutes and an inter-quartile range of 2.3 to 4.2 minutes.

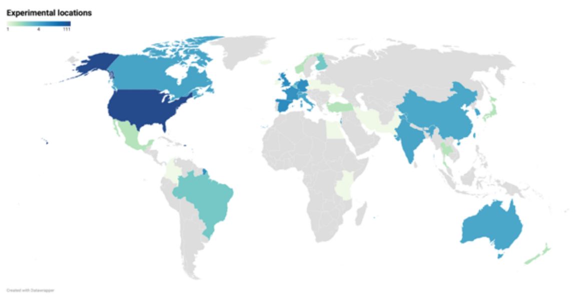


*Figure 3. Locations from which observers participated in on-line experiment.*

The raw data from this experiment were 461 binary, nominal, yes-no responses for each of the 33 stimuli for which observers were asked to select those whose transitions were continuous.

Intra–observer repeatability was quantified based on the 11 repeated stimuli. Counting how many of them were judged differently the first versus the second time resulted in an intra-observer repeatability error percentage. E.g., an error of 9% here would mean that only one of the 11 stimuli was judged differently on the two occasions.

The binary responses to each of the 22 stimuli (pooling the repetitions for 11 of them) allow for the computation of the percentage of times that a stimulus was judged to be continuous. E.g., an 80% value here would mean that 4/5 of the responses indicate continuity, while the other 1/5 judge it to exhibit some discontinuities.

Finally, a psychometric function (i.e., the logistic function) was fitted to the resulting percentages, augmented by an anchor of 100% detection at zero color difference. The ΔE2000 value at which the fitted function reaches 50% then identifies the perceptibility threshold, and therefore the value of the just noticeable difference between a transition being judged to be continuous and some, just-noticeable discontinuity appearing [8].

## Experimental results

### *Intra-observer repeatability*

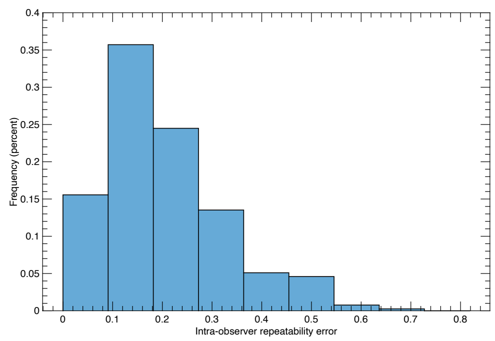


*Figure 4. Intra-observer repeatability error in judging transition continuity.*

Fig 4. Shows the percentage frequencies of various degrees of intra-observer repeatability error, where the median error across all 461 observations was 16%, and where over 15% of observers made no errors and 35% made only one error for the 11 repeated stimuli. This is a higher degree of intra-observer repeatability than was the case for the line color experiment [6], where median errors were at 24%.

### *Perceptibility thresholds*

The individual responses to the transition continuity experiment can then be used to compute continuity percentages for each of the 22 transitions and to plot them against the 95th percentiles of ΔE2000 color differences of adjacent colors within those transitions (Fig. 5). 95th percentiles were used since they have been found in previous studies to relate well to perceptibility of color differences between complex stimuli [5,11]. Since Euclidean distance in CAM16 UCS has recently been shown to be an even better match to psychophysical color difference data [9], the same analysis has also been carried out using that metric (Fig. 6). As can be seen, the 50% thresholds of continuity judgments are reached at different levels of color difference for the three colors used in this experiment, as shown in Tab. 1. It is interesting to note that the threshold is lower for the gray transition than for the cyan or red ones, even though the ΔEs used here have already been adjusted for the heightened sensitivity to color differences in neutrals versus other colors. The lower threshold here therefore indicates a possible additional effect for neutrals when it comes to transitions, on top of the already heightened sensitivity to them even for uniform colors that is already accounted for by ΔE2000 and ΔE UCS and the effect is similar to heightened sensitivity to differences in lightness seen for color differences between samples presented without spatial separation [3].

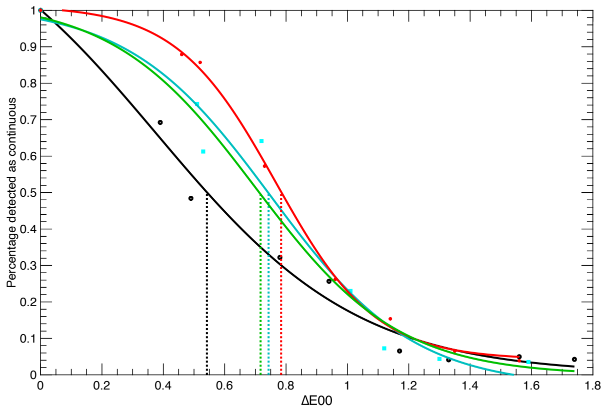


*Figure 5. Continuity perceptibility thresholds for gray, cyan, red and combined (green) transitions in terms of ΔE2000 95th percentiles.*

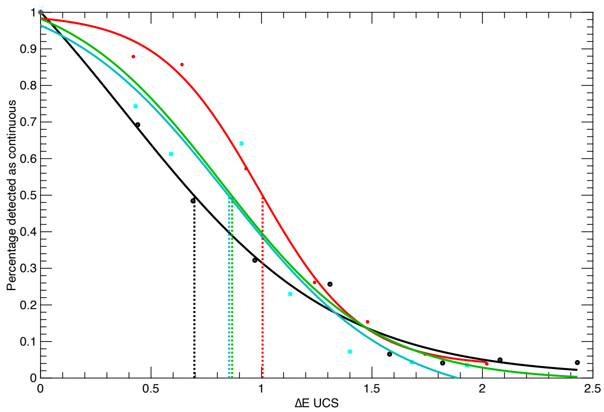


*Figure 6. Continuity perceptibility thresholds for gray, cyan, red and combined (green) transitions in terms of ΔE UCS 95th percentiles.*

**Table 1: continuity perceptibility thresholds**

| Transition color | ΔE2000 | ΔE UCS |
|---|---|---|
| Gray | 0.54 | 0.70 |
| Cyan | 0.74 | 0.85 |
| Red | 0.78 | 1.01 |
| Combined | 0.72 | 0.86 |

### *Cyan "inversion"*

While the continuity perceptibility percentages obtained from this uncontrolled experiment follow a continuous relationship relative to the color difference statistics of the corresponding transition images, there is a curious "inversion" in the cyan data. There, the third transition, with a $95^{th}$ percentile ΔE2000 of 0.73, is slightly less often judged to be continuous than the fourth one, whose corresponding ΔE2000 statistic is 0.96, while both the corresponding red and gray transitions around these ΔE2000 values have distinctly different perceptibility percentages and ones that have a strong inverse relationship with ΔE2000.

This may have pointed to an error in assigning the correct images to the corresponding experimental questions, but a verification confirmed that no mistake was made. Since the sequence and grouping of stimuli was fixed for all observers, the result may instead suggest some form of simultaneous contrast effect, since the third cyan transition (shown twice during the experiment) and the fourth one (shown once) were always viewed alongside one other, fixed transition (Fig. 7). It turns out that the randomization resulted in both transitions being shown alongside others that were much less continuous than either of these. In particular, the coarsest cyan transition (with a $95^{th}$ percentile ΔE2000 of 1.56) was shown alongside both. This suggests some form of criterion shift from set of 3 to set of 3, , where the perceived difference between the smaller ΔE2000 steps of both transitions may have been suppressed since they were viewed alongside a much coarser third transition. This context-dependence hypothesis could be tested in a future, targeted experiment.

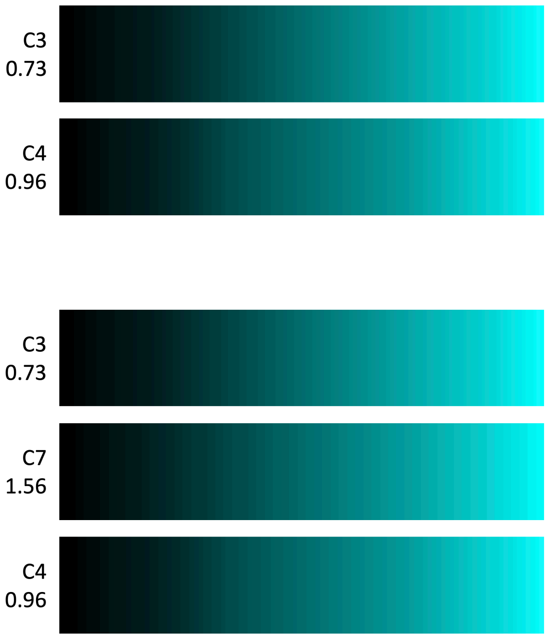


*Figure 7. Cyan transitions 3 and 4 (top) and alongside the coarsest cyan transition (bottom).*

## Evaluation using HDR-VDP

Given the color transition images that were evaluated in the on-line experiment reported here, it is possible to evaluate them not only in terms of their individual ΔE2000 statistics but also to apply models that predict the discriminability of differences. This, however, cannot be done directly, since such models take a pair of corresponding stimuli and make predictions about the likelihood of discriminability between them and also because they make predictions about all types of differences while the focus here is on the perception of continuity alone.

Nonetheless, such models can be applied to image pairs consisting of the transition with the lowest threshold and all the other transitions in turn, with the aim of identifying the transitions that are noticeably different from it. If the transition with the lowest threshold is perceived as continuous, as it indeed is for the stimuli used here (see first data points per color in Figs. 5 and 6), then finding the discriminability threshold from it can be a good approximation of the continuity perceptibility threshold.

What complicates matters is that the transition images here differ not only in their quantization thresholds, but also in their per-pixel colors, as a result of which discriminability models would also indicate increased probabilities for two images that are equally continuous but that differ in the rate of change along a color transition.

With these caveats out of the way, it is still meaningful to apply a model of discriminability to these images and consider how well its predictions relate to the psychophysical data collected here.

The model that will be used is Mantiuk's High-Dynamic-Range Visual-Difference-Predictor, HDR-VDP-3 [10, 11], which mimics the anatomy of the visual system and is tuned to predicting an extensive body of experimental data. The model accounts for features like intra-ocular light scatter, photoreceptor spectral sensitivity, luminance masking, neural noise, contrast sensitivity and contrast masking, when predicting the response to each of a pair of stimuli, before computing a detection probability image ($P_{map}$) in which each pixel represents the probability of detecting a difference.

Fig. 8 shows the $P_{maps}$ of detection probabilities between the nominally 0.2 ΔE2000 transition and the other transitions for each of the three colors, computed with HDR-VDP v3.0.7, 52.72 [pix/deg], (cpu). As can be seen, each of the images exhibits some regions of high discrimination probability and using a high percentile of the per-pixel detection probabilities, which would be consistent with previous image difference studies [5] and with the HDR-VDP paper [10], would lead to all images being judged as perceptibly different.

However, taking the median discrimination probability, which may be more robust to unintended color (as opposed to continuity) differences for the synthetic images evaluated here, results in a metric that correlates better with the data from the present experiment. Fig. 9 shows the result, including perceptibility thresholds computed using the logistic function. These indicate that median detection probabilities as low as 0.20 for grays, 0.30 for cyans and 0.47 for reds (with a combined threshold of 0.32) correspond to a just noticeable difference in terms of transition continuity.

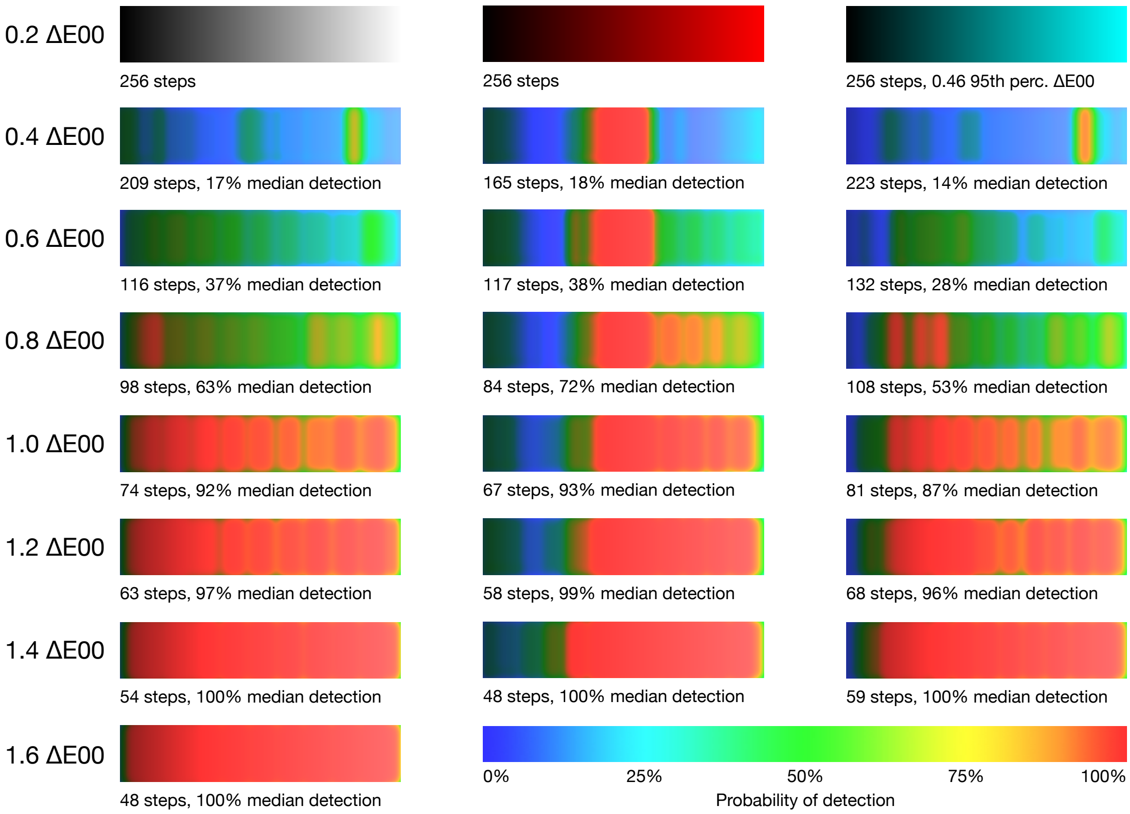


*Figure 8. $P_{map}$ difference images showing difference detection probability between the nominally 0.2 ΔE2000 transition (shown in the top row) and other transitions with higher ΔE2000 thresholds. Median probability values are shown below each difference image.*

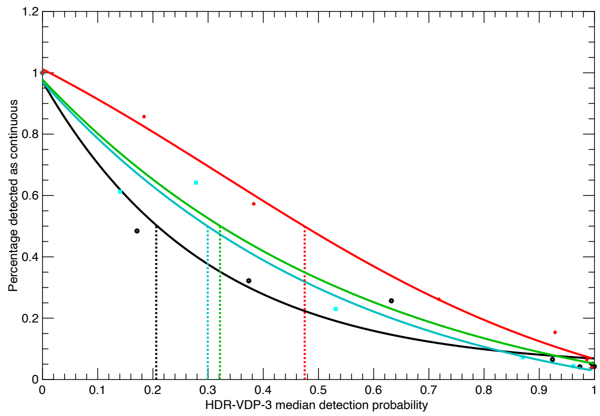


*Figure 9. Continuity perceptibility thresholds for gray, cyan, red and all transitions in terms of transition HDR-VDP-3 median detection probabilities.*

## Verification using complex images

Since the above results come from synthetic color transitions, the question of their applicability to complex images arises. To address it, a second pilot experiment was conducted in the same way as for the synthetic images described above (i.e., experimental design, on-line format, data analysis, etc.). Fig. 10 shows the three images used, each of which had prominent continuous transitions, alongside some of the nominal ΔE2000 threshold levels to which they were quantized and also showing the number of unique colors in each.

The images were generated by starting from originals where no quantization was performed and that were chosen for the continuity of their transitions. Each image then had its unique colors identified and ordered by frequency. A mapping from these original colors to a quantized set of target colors was then obtained by first building a color dictionary in the following way: The first entry was the most frequent original color. Subsequent entries were added by traversing the remaining original colors, checking against the target colors and either assigning the closest from among the current set of target colors, if there was at least one sub-threshold one, or adding the current entry to the target set. The resulting dictionary specified target colors for each original color such that no target color pairs below the chosen threshold were used and where frequency dictated traversal order.

**Table 2: Image properties in verification experiment (ΔEs are 95th percentiles of ΔE00 per-pixel differences and colors refer to unique colors in an image).**

| | Concert (242 colors) | | Sunset (20299 colors) | | Hands (47399 colors) | |
|---|---|---|---|---|---|---|
| **Target ΔE** | ΔE | Colors | ΔE | Colors | ΔE | Colors |
| 0.2 | 0.00 | 240 | 0.00 | 19949 | 0.20 | 42521 |
| 0.4 | 0.36 | 121 | 0.38 | 11319 | 0.38 | 25988 |
| 0.6 | 0.49 | 104 | 0.57 | 8638 | 0.57 | 16558 |
| 0.8 | 0.73 | 74 | 0.77 | 5913 | 0.75 | 10762 |
| 1.0 | 0.95 | 64 | 0.95 | 4056 | 0.94 | 7288 |
| 1.2 | 1.15 | 54 | 1.12 | 2889 | 1.11 | 5156 |
| 1.4 | 1.28 | 48 | 1.29 | 2152 | 1.27 | 3789 |
| 1.6 | 1.57 | 42 | | | | |

While the images used here contain large regions of continuous transitions, they also include areas of high spatial frequency. Therefore, a different quantification of the level of color difference in the quantized images was needed. Instead of being able to look at ΔE2000 statistics along a single color transition, the per-pixel color differences between the unquantized original and each of the quantizations resulting from different thresholds was used and is shown in Fig. 10 and Tab. 2.

One hundred and sixtyeight observers participated in this experiment with a median intra-observer repeatability error of 14%, indicating a similar degree of consistency with was the case for the synthetic transitions where it was 16%.

Fig. 11 shows the perceptibility thresholds for the three images based on continuity judgment proportions versus per-pixel ΔE2000 differences between original and variously-quantized versions. The threshold for the "concert" image was 0.73 ΔE2000, for the "sunset" image 0.60 and for the "hands" image 0.75. Pooling data from all three images resulted in a threshold of 0.69. This range of around 0.6 to 0.8 compares well with the thresholds from the synthetic transitions images which were in the 0.5 to 0.7 ΔE2000 range and could indicate a slightly lower sensitivity to differences, perhaps due to the more complex structure of the overall stimulus. Overall though, these ranges are a good first indication of the level of color resolution needed for images to appear to be continuous in their color transitions.

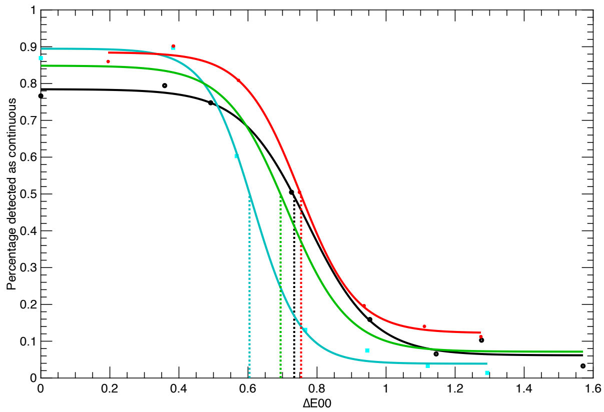


*Figure 11. Continuity perceptibility thresholds for concert (black), sunset (cyan), hands (red) and combined (green) images in terms of ΔE2000 95$^{th}$ percentiles of per-pixel differences.*

## Conclusions

Whether the difference between two stimuli will be perceived or not depends on a rich and complex variety of factors. Even when the focus is on color alone, as opposed to other visual attributes, context plays an important role, as is clear from a rich body of existing literature.

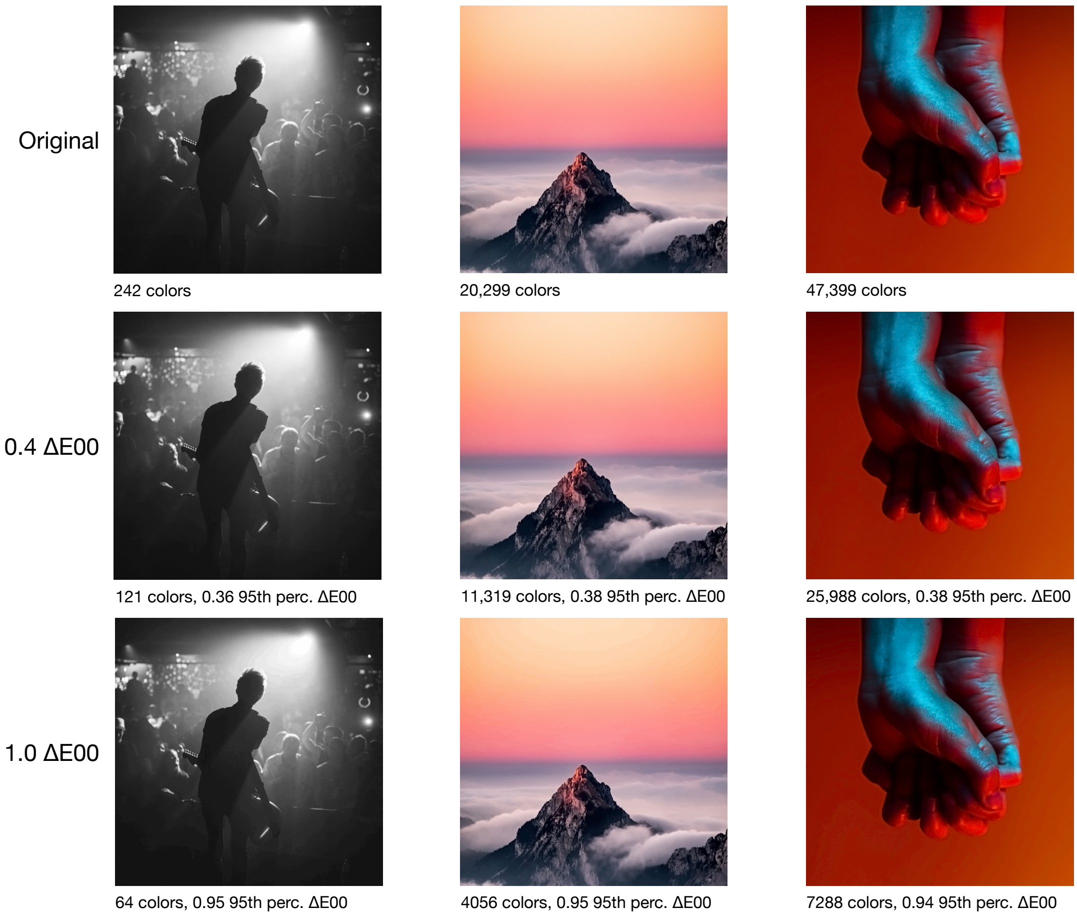


*Figure 10. Complex images showing number of unique colors at selected target ΔE2000 threshold levels (left column) and the 95$^{th}$ percentile of the per-pixel ΔE2000s between the original and quantized images (under corresponding original image). The "concert" image (left) is by Blake Carpenter, the "sunset" image (center) by Boris Baldinger and the "hands" image (right) by Ian Dooley, all freely available on Unsplash.com.*

The present paper then aims to provide a first sense of how color difference perceptibility thresholds are affected when the stimulus is not a uniform color but instead consists of a transition between two colors. It asks the question of what level of color difference between adjacent steps of such a transition needs to be exceeded for the sense of the transition's continuity to break.

The experiment reported here has many constraints, from the fact that it was performed under uncontrolled conditions, via the very limited number of color pairs between which transitions were constructed, to the limit imposed by generating 8-bit images so that these could be viewed by a broad on-line audience. While this means that it is not a good basis for generalizations, it nonetheless provides some first examples of what perceptibility thresholds may be like for color transition continuity. These point to lower thresholds that have been found in many other contexts to between around 0.5 and 0.7 ΔE2000 for the 95$^{th}$ percentile of transition color differences, which are most similar to the case of viewing color pairs without spatial separation. Such similarity makes good intuitive sense, since a transition can be thought of as a sequence of color pairs without spatial separation. To understand whether the two contexts behave in the same way would require further targeted experimentation.

An interesting artifact of the experiment having been conducted with a randomized but fixed stimulus sequence is the suggestion that there may be crispening-like mechanisms in terms of the perception of transition continuity, which too would merit further exploration, in addition to a more extensive and controlled study of the perception of transition continuity in the future.

Finally, a small pilot experiment was also conducted with complex images, which showed a similar range of continuity perceptibility thresholds in the 0.5 to 0.8 ΔE2000 range, making it a good first indication of the level of color resolution needed to render images that appear to have continuous color content.

## Acknowledgements

The authors would like to thank the 619 observers who took part in the on-line experiments for their generosity, the reviewers for their constructive suggestions and our colleagues at HP Inc., Andreu Gonzalez and Elizabeth Zapata, for their support.

## References

[1] L. D. Griffin, D. Mylonas, Categorical colour geometry. *PLoS One*, 14, e0216296. (2019).

[2] R. W. G. Hunt, M R. Pointer, *Measuring Colour*, 4th Ed., John Wiley & Sons, Ltd, Chichester, UK (2011) 57.

[3] F. Mirjalili, M. R. Luo, G. Cui, J. Morovic, Color-difference formula for evaluating color pairs with no separation: ΔENS, *Journal of the Optical Society of America A*, 36 (2019), 789-799.

[4] Paravina RD, Ghinea R, Herrera LJ, Bona AD, Igiel C, Linninger M, Sakai M, Takahashi H, Tashkandi E, Perez Mdel M. Color difference thresholds in dentistry. *J Esthet Restor Dent*. 2015 Mar-Apr;27 Suppl 1:S1-9. doi: 10.1111/jerd.12149.

[5] Uroz J., Morovič J. and Luo M. R. (2002) Perceptibility Thresholds of Colour Differences in Large Printed Images. In *Colour Image Science: Exploiting Digital Media*, MacDonald L. W. and Luo M. R. (eds.), John Wiley & Sons, 49–73.

[6] Morovič J., Morovič P. (2022) The Perceptibility of Color Differences between Thin Lines and its Application to Printing Imaging Pipelines, *30th IS&T Color and Imaging Conference*, 36 - 41, https://doi.org/10.2352/CIC.2022.30.1.08

[7] A. B. Watson, Probability summation over time, *Vision Research* 19 (1979), 515–522.

[8] P. G. Engeldrum, *Psychometric Scaling: A Toolkit for Imaging System Development*, Imcotek Press (2000).

[9] M. R. Luo, Q. Xu, M. Pointer, M. Melgosa, G. Cui, C. Li, K. Xiao, M. Huang, "A comprehensive test of colour-difference formulae and uniform colour spaces using available visual datasets," *Color Res Appl*, 48(3):267-282. (2023). doi:10.1002/col.22844.

[10] R. Mantiuk, K. J. Kim, A. G. Rempel, W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions." *ACM Transactions on Graphics*, 30(4):40:1–40:14 (2011). doi:10.1145/2010324.1964935.

[11] R. K. Mantiuk, D. Hammou, P. Hanji, HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content, (2023). https://arxiv.org/abs/2304.13625

## Author Biography

*Ján Morovič received his Ph.D. in color science from the University of Derby (UK) in 1998, where he then worked as a lecturer. Since 2003 he has been at Hewlett-Packard in Barcelona as a senior color scientist and later master technologist. He has also served as the director of CIE Division 8 on Image Technology and Wiley and Sons have published his 'Color Gamut Mapping' book. He is the author of over 100 papers and has filed 180+ US patents (142 granted).*

*Peter Morovič received his Ph.D. in computer science from the University of East Anglia (UK) in 2002 and holds a B.Sc. in theoretical computer science from Comenius University (Slovakia). He has been a senior color and imaging scientist at HP Inc. since 2007, has published 65+ scientific articles and has 180+ US patents filed (140 granted) to date. His interests include 2D/3D image processing, color vision, computational photography, computational geometry. His Erdős number is 4.*