Guidelines for using color blending in data visualization

Sandra Gama
INESC-ID and Instituto Superior Técnico,
Universidade de Lisboa
Av. Rovisco Pais, 1
Lisboa, Portugal
sandra.gama@tecnico.ulisboa.pt

Daniel Gonçalves
INESC-ID and Instituto Superior Técnico,
Universidade de Lisboa
Av. Rovisco Pais, 1
Lisboa, Portugal
daniel.goncalves@inesc-id.pt

ABSTRACT
Visualization is a powerful way to convey data, showing potential for joining and interrelating data items. However, when dealing with large amounts of data, visually merging different classes of information poses several challenges. Color, however, due to its effectiveness for labeling and categorizing information, may be a solution to this shortcoming. Merging items with different colors may suggest mixing their original colors. This approach generates an immediately perceivable way to represent merged items. It also keeps context through the association of the mixed color to its original colors. We studied to which extent color blending provides users with the means to understand the provenience of data items by conducting two user studies using CIE-LCh, CMYK and HSV blending to ascertain (i) to which extent people are able to, given a particular color, understand its provenience, and (ii) the color model in which to perform color blending so that users find it intuitive. Results showed colors which are more suitable for blending so that users understand their provenience and indicated that the CIE-LCh model is more effective for representing color blending.

Categories and Subject Descriptors
H.5.2 [**Information interfaces and presentation**]: User Interfaces

General Terms
Human Factors

1. INTRODUCTION
We nowadays manipulate electronic devices on a daily basis, generating high volumes of data, which are it difficult to interrelate and analyze. One way to holistically represent information is through integrated visualization mechanisms. Visualization is a powerful way to convey information and alleviates cognitive load associated with data interpretation. Color, in particular, is effective for information coding, hence it is used to display data categories. This suggests the representation of complex data entities as visual objects, where colors naturally code their attributes [5]. It is often desirable to visually represent data entities which verify more than one property. As a result, it is relevant to find a solution for visualizations in which data items are joined and interrelated. That may be done through color blending. Nevertheless, a particular aspect remains to be studied that has not been particularly addressed by current research: to which extent people are able to, given a specific color (which results of the blending of two other colors), understand the blended color's provenience. Furthermore, the influence of color models over human perception regarding color blending may be important. We have studied both aspects through two user studies, which we will present on this paper. Results of the first study showed that people have some difficulties in understanding color blending. Additionally, results indicated that there are no statistically significant differences on user preferences on blending between CIE-LCh and CMYK models, but these perform much better than the HSV. Results have that the CIE-LCh model yielded significantly higher success rates than the CMYK model.

In this document we present our user study and analyze and discuss results, drawing guidelines for further work on color blending for information visualization.

2. STUDYING COLOR BLENDING
Considering the problem of merging data sets, we studied the potential of color to alleviate cognitive load associated with data interpretation. The main objectives of our study are: to understand whether users are able to understand the provenience of a given color resulting from the blending of original colors; and to find out the best color model for color blending. To do so, we performed two user studies. In the following sections we describe them in detail.

2.1 Designing the first color study
Color coding is excellent to display category information, however only between about five and ten codes [4, 3] are rapidly perceived. Considering the opponent process theory [3], we chose as main colors: (R: Red; G: Green; B: Blue; Y: Yellow). Since the CIE-LCh model represents a perceptual uniform color space [4], we performed color blending using this model. The first set of blended colors (set A) consists of the pairwise combinations of the main colors: RG, RB, RY, GY and BY. The second set (set B) consists of the triadic combinations: RRG, RGG, RBG, RYG, RRB, RBB, RYB, RYR, RYY, GBG, GBB, GYB, GYG, GYY, BYB and BYY.

Our 5-stage study started with a brief profiling questionnaire. On the second part, we performed a validated simplified 6-plate Ishihara test for color blindness detection [1]. Next, users were presented with sets A and B and were asked to pick exactly two colors which mix into that color from a palette. Different palettes were assigned to each set, derived from the way each color was obtained. For set A, the palette consisted of the four main colors, while the palette for set B consisted of both the main colors and the colors from set A. Colors were displayed over a white background and each blended color was displayed individually, to avoid visual artifacts. On the following step, we performed color blending in three different color spaces for all colors in sets A and B, and asked users to pick the most natural transition. We took into account HSV (a more realistic representation of additive color spaces), CMYK (of subtractive nature, which might be relevant since users may think in terms of ink mixtures) and CIE-LCh (taking perceptual uniformity into account) models. Lastly, a satisfaction questionnaire was performed where subjects were asked to rate, on a 5-point Likert scale (1 point = complete disagreement; 5 points = complete agreement), the following: (i) I found it easy to decide which pairs of colors resulted in a given color; (ii) I found it easy to decide the most natural blending option between two colors.

We created a web page with the questionnaire, using web-safe colors. We broadcasted it and kept it open for a week.

Out of 73 subjects, all (100%) participants have normal color vision. Success rates from performing blending perception tests for set A are as follows: $RG = 31.51\%$, $RB = 32.88\%$, $RY = 84.93\%$, $GB = 45.21\%$, $GY = 80.82\%$, $BY = 4.11\%$. Blending specific colors resulted in promising success rates: GYG (71.23%), GYY (71.23%), RYY (71.23%). These results are in line with set A results, showing promise in CIE-LCh blending for green and yellow, as well as red and yellow. Similarly, colors which did not blend well for set A propagated low success rates for set B, showing that subjects were not able to recognize the original colors that mix into the corresponding set A colors: RRG (0.64%), RBG (0.00%), BYB (0.00%), BYY (0.00%).

Regarding preference, in general terms, CMYK and CIE-LCh models perform betterthan the HSV model. However, there seems to be a preference for the CMYK model, with an average result of 46.30% against the other options (CIE-LCh=34.65%, HSV=20.47%, NA=5.99%). Despite lack of statistical evidence supporting the CMYK model (a Shapiro-Wilk test showed evidence against normality in some of the data $p_{HSV} = 0.01$, $p_{LCh} = 0.775$, $p_{CMYK} = 0.08$, $p_{NA} = 0.092$ and so we assumed a non-normal distribution and applied pairwise Wilcoxon tests to ascertain statistical significance, where $p_{CIE-LCh,CMYK}=0.535$). We performed a second study to analyze performance on CMYK model.

2.2 Follow up study
Since the goal of the second study was to understand whether the CMYK model yielded better performance than the CIE-LCh, we performed a follow-up study, similar to the first, except in two aspects: colors were blended using the CMYK color space for the third part of the study the color blending preference step did not exist on this version because we had already compared user preference on different color spaces.

Out of 44 participants, all had normal color vision. Success rates from performing color blending perception tests for set A are as follows: $RG = 18.18\%$, $RB = 75.00\%$, $RY = 90.91\%$, $GB = 59.09\%$, $GY = 75.00\%$, $BY = 11.36\%$. Compared to the CIE-LCh model, there is no statistical significant difference. Results for color set B ($\overline{X_{success}} = 15.91\%$, $\sigma = 18.46$) are lower than those for the CIE-LCh model (($\overline{X_{success}} = 28.23\%$, $\sigma = 27.01$)). The best success rate obtained was for RYY (64.64%), which is still quite low. In fact, the CIE-LCh model seems to yield better results ($avg_C IE - LCh_{(A,B)} = 50.34\%$, $stDev_C IE - LCh_{(A,B)} = 18.99\%$, $avg_C MYK_{(A,B)} = 30.97\%$, $stDevCMYK_{(A,B)} = 14.52\%$). In fact, a Mann-Whitney test showed these differences to be significative $U = 9.5$, $z = 2.31$, $p = 0.25$.

2.3 Discussion and design implications
We performed color blending in different models and found evidence showing the CIE-LCh model performs generally better than either CMYK or HSV models. Nevertheless, results are still not outstanding in general terms. Colors with either yellow and green or red and yellow components yield performances around 70%. These are the best combinations for blending two colors. Combining either red and blue or green and blue yields lower results, still above 30%. All other combinations (either of three different colors or containing colors which are opponents on the color wheel) yield extremely low results, showing these are impractical for representing data merge through color blending. In practical terms, if there are four different items to interrelate in a pairwise manner, a compromise may be to represent pairs of either orange and cyan or lime and purple. Nevertheless, due to relatively low average results, additional visual artifacts to guide users in data interpretation may be relevant.

3. CONCLUSIONS
Color is effective for conveying information, providing means for labeling and categorizing. We have studied whether color blending provides users with the means to understand the provenience of data items, providing guidelines on which model to use for color blending. Based on statistical evidence, we propose a set of colors to choose for depicting data merging through color blending.

4. ACKNOWLEDGMENTS

5. REFERENCES
[1] D. D. Alwis and C. Kon. A new way to use the ishihara test. *Journal of neurology*, 239(8):451–454, 1992.

[2] N. Gossett and C. Baoquan. Paint inspired color mixing and compositing for visualization. In *IEEE Symposium on Information Visualization*, pages 113–118, 2004.

[3] E. Hering. *Grundzuge der Lehre vom Lichtsinn*. Springer, 1920.

[4] G. Meyer and D. Greenberg. Perceptual color spaces for computer graphics. In *SIG-GRAPH 78*, pages 254–261, 1980.

[5] C. Ware. *Information Visualization: Perception for Design*. Elsevier, 2012.