The Effect of Feedback on a Color Selection Interface

Sarah A. Douglas
Ted Kirkpatrick
Department of Computer Science
University of Oregon
Eugene, OR, USA 97403
douglas@cs.uoregon.edu, ted@cs.uoregon.edu
(541) 346-3974

Abstract

Computer interfaces for color selection consist of a visible screen representation, an input method, and the underlying conceptual organization of the color model. We report a two-way factorial, between-subject variable experiment which tested the effect of high and low visual feedback interfaces on speed and accuracy of color matching for RGB and HSV color models. The only significant effect was accuracy due to feedback type. Qualitative analysis of the individual matches suggests that feedback assisted users most in their final refinements to the match.

Keywords: Color selection, feedback, color models, direct manipulation.

Introduction

Visual feedback is an identifying characteristic of modern direct manipulation interfaces. Ben Shneiderman's classic definition of direct manipulation [4] includes "continuous display of the object of interest" and "rapid, incremental, reversible operations whose impact on the object of interest is immediately visible." Direct manipulation color selection tools are commonly used for color selection in computer graphics systems. These tools provide some visual representation of a *color model*: an organization of the range of displayable colors into a three-dimensional space. The visual representation is intended to both orient the user within the color model and provide a means for directly navigating within that model.

The close relationship between the color model and its representation on the screen makes it difficult to determine the effects of each one. If users can select colors with one system more quickly than another, is it because of the underlying color model or the style in which that model is represented on the screen?

We performed an experiment which separated the effects of the color model from the interface which presented it to the users. In a prior paper [1], we reported that the color model had no effect on the speed and accuracy of color selection. By contrast, visual feedback had a significant effect. This paper extends that work to consider the mechanics of the influence of visual feedback on the color selection process.

Previous Research

There is surprisingly little empirical data comparing color models, and even less comparing several interfaces to the same model. Schwarz, Cowan, and Beatty [3] performed the major experiment to date, referred to in the remainder of this paper as "the Schwarz study". This is the current reference work for empirical data about color models. In this study, five different color models were compared: LAB, HSV, Opponent, YIQ, and RGB. Each color model was implemented with two input styles. The experimental design was five by two factorial, with color model and input style as the between-subjects variables.

While the experiment studied two styles of interfaces, these styles differed only in input method. In each case, the screen representation was identical. For each match, a "target" color was displayed in a rectangle on the screen and the subject used the interface to set the color of a second "controlled" rectangle to match the first one as closely as possible. The interfaces differed in the input method. A puck on a graphics tablet controlled the three parameters using either horizontal motion or a combination of horizontal and vertical motion.

The authors analyzed the time required by the subjects to match, the accuracy of the final match, relationship of color model axes to accuracy in various perceptual attributes of the target, and time to reach a given level of accuracy, and learning. Accuracy was measured in color distance units (cdus) of the LAB color space. Of the many results reported in the paper, the following are most relevant to this discussion:

1. Subjects matched significantly more quickly with the RGB and Opponent models, with HSV 20 percent slower. On the other hand, HSV was more accurate than the other two.

2. Matches were characterized as having two distinct phases: an initial *convergence phase* where the subject rapidly approached the general region of the target color, and a *refinement phase* where the subject fine-tuned the match. The effect of the color model was most pronounced in the convergence phase as the subjects moved to within 25 cdus of the target. The RGB model allowed subjects to reach this threshold more quickly than the HSV model.

While the experiment was done with due rigor, there were aspects of the experimental task which may limit the extension of the results to contemporary interfaces. The experiment was published in 1985, at a time when graphical user interface design was in its infancy. In particular, their interface used a digitizing tablet and puck with little visual feedback. We located the following limitations in their interface:

1. *No visual feedback of the location of the current color in the color space.* The subjects were given no indication of the dimensions of each parameter of the space. Furthermore, they were given no indication of how close they were to the boundaries of each axis. The controlled color simply stopped changing without warning when the subject moved the puck outside the range of one of the coordinates. They refer to this process as "clipping".

2. *No visual feedback of the effects of each parameter.* While the subjects were initially instructed in the effects of each parameter, they had no visible representation of what the parameters did. It is difficult to infer the effect of a parameter from simply moving the puck and observing the results, because the kind of change in the controlled color varies depending upon what color is currently displayed. For example, in the HSV model the saturation parameter will vary the amount of red if the controlled color has a red hue, but will increase the amount of blue if the controlled color is in that hue. Similar interactions between the parameters occur in the other color models.

In summary, the Schwarz study raises questions about the effect that feedback in the user interface has on use of color models.

Experimental Study

The experiment used in this study considered color model and feedback level as separate factors. We tested the effect of high and low feedback interfaces on speed and accuracy of color matching for the RGB and HSV models. The following summarizes the relevant points for the analysis of the feedback. Complete details of the experiment may be found in [1].

Experimental Design

The experimental design was two-way factorial, with color models and interfaces as the two between-subject variables, for a total of four subject groups. Twelve subjects were in each group, randomly assigned to a particular color model x interface group.

Our concern with the Schwarz interfaces was that they lack important kinds of feedback. To address our first concern about feedback and representation of the color space, we displayed the location of the current color within the color model. Each of the three parameters of the model was represented by a slider, and an arrow indicated the current value of the associated parameter within its total range. The user controlled the current color either by dragging the arrow along the slider with the mouse or by clicking directly on some row of the slider. Both of these methods were explicitly demonstrated to the user during the instructional phase.

Our second concern was that there was no visual feedback of the *effects* of each parameter. To compare the influence of such a representation, we used two different styles of sliders. One interface, called "position-only", (Color Plate 1) gave no indication of what each parameter did: the interior of the sliders was gray at all times. The second interface, called "position+effect", filled each slider with a range of colors. Each pixel row on the slider displayed the color that the controlled rectangle would take if the arrow were moved to that row. A user could look at the slider and know what effect it would have if it were moved to any point. Color Plate 2 shows the "position+effect" interface style for the HSV model.

Colors

We selected thirty colors. Six of these were from the original Schwarz experiment. The remaining twenty-four were taken from the MacBeth ColorChecker chart

[2], a standard reference chart for tests of color rendition. Twelve of these are representative of colors commonly found in natural and office environments (flesh tones, sky blue, and common office colors), six are the additive and subtractive color primaries, and the final six are an achromatic ramp from black to white.

Our data collection program logged the current slider positions and the value of the controlled color every tenth of a second. In particular, the final reading of each match indicated the total time and distance between the controlled and target colors.

Experimental Task
Subjects adjusted a controlled color until it matched a target color. Each subject used a single color model x interface combination to match the same sequence of thirty different colors. All interaction with the program was done using the standard Macintosh mouse to move the sliders.

The experimenter demonstrated how to manipulate the sliders using the mouse. To replicate a situation of use similar to what most users typically encounter, no abstract explanation of the color model was given to subjects. They were not explicitly told what the three parameters of their color model were, nor were the parameters named on the screen. In the instructions, subjects were asked to learn how the different sliders affected the color during the course of the experiment. We used Schwarz' wording to describe how closely the subjects should try to match the target color: "Continue to refine the match until you think they are the same color or until it becomes *extremely* difficult to get the colors any closer to each other."

Results
No significant effects were found due to color model ($F (1, 44) = .233$, $p = .632$) and there was no interaction effect between color model and feedback ($F (1, 44) = .077$, $p = .783$). The level of feedback had a significant effect on accuracy but not time. The results of the analysis of variance of the different feedback styles are given in Table 1.

The number of subjects in our study gave us enough statistical power to detect large differences between the conditions. With an 80 % power level we could detect a difference of 19.5 seconds, which is 33 percent of the mean time to match for all subjects. While this level of analysis leaves out smaller effects, it has the benefit that any effect detected is large enough to be both statistically robust and present in situations of actual use outside the laboratory.

Table 1: Times and accuracies for feedback types.

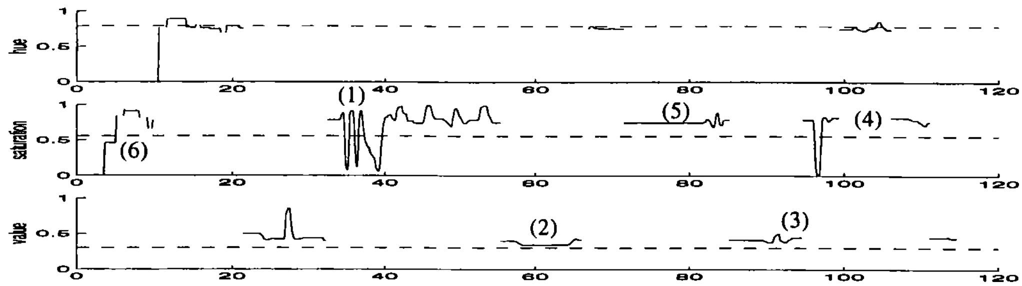
Feedback Type	Mean Sec.	Std. Dev.	Mean LAB cdus[1].	Std. Dev.
pos.-only	60.9	23.2	8.23	2.89
pos.+effect	54.8	22.7	6.07	2.74
% change	not. sig.		26%[2]	

[1]The cdu data were positively skewed. For analysis of variance, the natural log of the distances was used.
[2]Significant: $F(1, 44) = 9.321$, $p=.0038$.

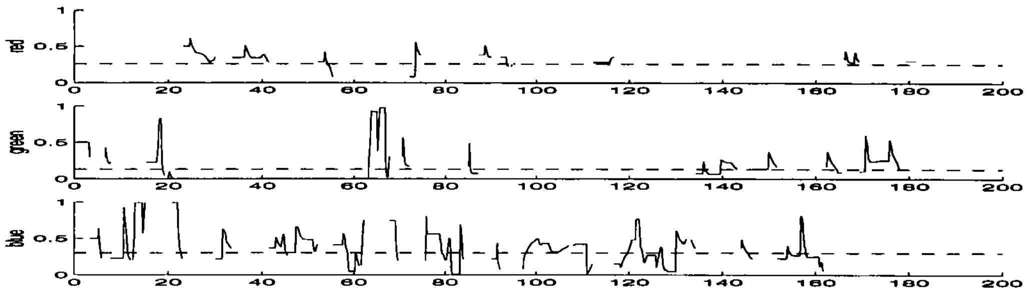
Discussion
Our experiment found that feedback had a significant effect on accuracy but not time. This indicates that the subjects using the "position+effect" interface actually performed better: they matched more accurately in about the same amount of time, rather than simply selecting a different tradeoff of speed and accuracy. These results agree with a similar experiment by Wells [5], who found that sliders similar to our "position+effect" interface allowed her subjects to match more accurately and slightly faster.

By contrast with the large effect of feedback, color model had no effect on time and accuracy. The data suggest that it makes little difference which color model is presented to the user for color selection, *provided the location of the current color within the color space is represented.* Users adapt quickly to whichever model they are given and the model may not make a difference.

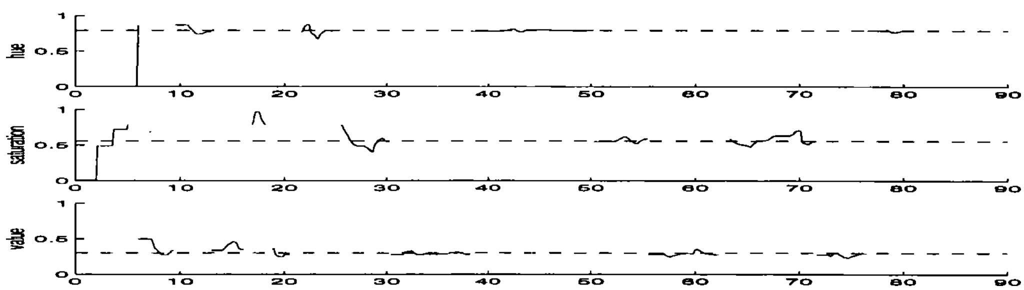
A Qualitative Analysis of the Impact of Feedback on Accuracy
The analysis of variance demonstrated that better feedback improved the accuracy of the subjects, but could not describe the mechanisms underlying this improvement. To understand how feedback improved performance, consider a change to a slider setting as a move through the three-dimensional color space. The sequence of settings in a given color match describe a path from the starting gray to the final color. We looked at the detailed record of the paths the subjects took. Due to the enormous amount of data in this detailed record, we focused our work on a single color. We picked the twenty-seventh color in the series, the purple from the MacBeth chart, with moderately low luminance (CIE XYZ values of 0.059, 0.040, and 0.102). This color was a good test case, as it was one



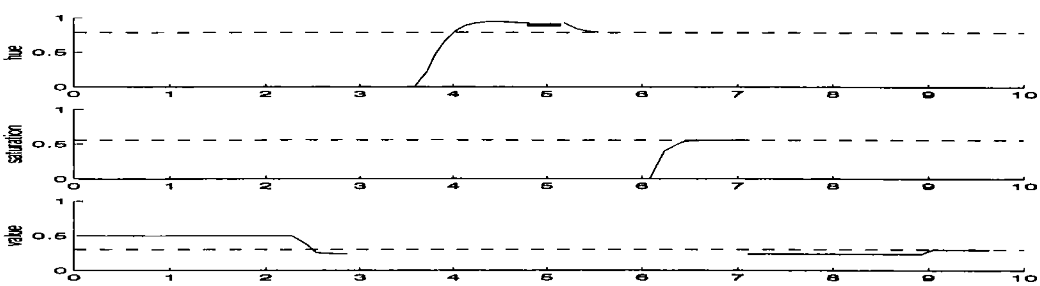
(a) Subject 36 (HSV, low-feedback)
Time: 99 seconds (rank 36), final distance: 22.23 cdus (rank 46)



(b) Subject 42 (RGB, high-feedback)
Time: 181 seconds (rank 48), final distance: 47.1 cdus (rank 48)



(c) Subject 3 (HSV, high-feedback)
Time: 80 seconds (rank 32), final distance: .952 cdus (rank 4)



(d) Subject 6 (HSV, low-feedback)
Time: 9 seconds (rank 1), final distance: .424 cdus (rank 1)

Figure 1 (a)-(d): Slider traces of the 27th trial for four subjects

Sample Traces

We analyzed traces for several subjects, looking for the above features. Two of the least accurate subjects and two of the most accurate subjects for this color are described below:

Subject 42 (Figure 1b) did not even complete the match within the three minute limit. The accuracy at the end of the trial was the worst for all subjects. The trace is characterized by a tremendous amount of scanning, continuing undiminished into the third minute of the trial. There are twenty-five target crossings and twenty-four clicks—over twenty of which moved the color *further* from the target. In this case, the clicking was clearly random. This subject never settled in to a region close to the target color.

Subject 36 (Figure 1a) had a significant amount of scanning but only nine target crossings. There were fifteen clicks in the first twenty seconds, then the subject used dragging for the rest of the trial. This subject got within about twenty cdus of the target color in fifty seconds but could never get any closer in the remaining seventy seconds.

Subject 3 (Figure 1c) had the fourth-best accuracy for this color. This subject did much less scanning, had only three target crossings of any size, and was in the general vicinity of the target within thirty seconds. Four clicks were used in the first six seconds and then the subject used dragging for the rest of the trial. The last fifty seconds of the trial were spent honing the match to a very fine degree.

Subject 6 (Figure 1d) was the most accurate of all subjects in this trial. The performance is clearly expert: the three sliders are each moved once, coming very close to the target, and then the value slider is slightly adjusted to complete the match. This subject did no scanning and only crossed a target line once. It is interesting to observe the judicious way this subject used clicking. The moves of the hue and saturation sliders begin with clicks which move them part way towards the target level and then the slider is dragged the rest of the way to the target.

Representing the Traces in Terms of Their Features

These four examples suggest which features might most succinctly represent the important aspects of a slider trace. *Target crossings* and *scanning* seem highly correlated, so counting crossings is probably a good measure of scanning behavior. Crossings can be easily recognized by an algorithm, with higher crossing counts indicating less competence at color selection. The subjects all had similar proportions of moves which left a slider significantly further away from the target, so this feature may indicate little more than would be shown by target crossings. *Pauses* and *consecutive moves* of the same slider appear to be rare, and it is not clear what they indicate about subject performance. Finally, clicking must be interpreted with care, taking into account context. It would be especially difficult to come up with an algorithm for determining if a click indicated expertise or incompetence.

Cross-subject comparisons suggest the role that feedback might play in improving accuracy. We group the subjects into three categories: those who know too little to be helped by feedback, those who know enough to take advantage of it, and those who know so much that it adds no further value. Subject 42 (Figure 1b), who never matched the target color, falls in the first group: this subject apparently never became skilled enough to establish routine procedures for color selection. This subject clearly derived no benefit from the high-feedback condition. We conjecture that this subject would have performed equivalently poorly regardless of the level of feedback.

Subject 6 (Figure 1d) represents the other extreme, someone who learned the color model very well. This subject had low feedback, but it is hard to imagine how high feedback could have improved the performance on this trial.

Subjects 36 and 3 (Figures 1a and 1c, respectively) represent the most interesting group of subjects. Their traces are more directed than the wild guesses of Subject 42, yet not nearly so sure as those of Subject 6. Both get within a reasonable distance of the target within the first twenty to thirty seconds, but their behavior after that is quite different. Subject 36 can never get any closer to the target color. The wide scanning of the saturation slider and smaller scanning of the value slider clearly show that this subject does not know the model very well. The correct saturation value is attained nine times in the course of the trace, yet at the time of the final match the saturation slider is further from the target than it was one hundred seconds earlier.

Subject 3 is also scanning, yet the degree of scanning is damped down as the trial proceeds. In the period from twenty-three to thirty seconds, this subject reduces the saturation below the target level, realizes the error, and places the slider almost exactly on the correct value.

Compare this with the performance of Subject 36 in the comparable period, from twenty-seven to fifty-seven seconds. Starting at nearly the same point as Subject 3, Subject 36 spends twenty seconds gyrating the saturation slider, only to wind up with a value just slightly closer to the target.

This comparison suggests both the way that higher feedback improves accuracy and who benefits from it. Subject 36 and Subject 3 appear to have had similar levels of knowledge of the color model: a rough idea of how to get in the general area of a target but not the precise knowledge sufficient to make a close match. Feedback allows subjects such as these to get really close; too little feedback leaves them floundering. Feedback has little effect on subjects at the two extremes of understanding of the model, such as Subjects 42 and 6. Such subjects either don't know enough to take advantage of the feedback or know so much that feedback is unnecessary.

The above explanation fits well with the Schwarz analysis of a color match in terms of *convergence* and *refinement* phases. Schwarz reports that during the final portion of the refinement phase, when the subjects were within about 6 cdu of the target, they had difficulty getting any closer to it. The subjects appeared to randomly wander through the color space region close to the target. The region close to the target is precisely where the higher levels of feedback helped subjects such as Subject 3 in our study. Such feedback was not present in the Schwarz interface.

Conclusions

This experiment explored the relationship between the color model and the interface which presented it to the user. We considered the effect of differing levels of feedback in the user interface. The earlier Schwarz study found significant differences in time and accuracy between color models but had an interface which provided very little feedback to the user. We performed a similar experiment with one interface that visually represented the user's current location and a second interface which also provided an indication of the function of each color model parameter. With these interfaces we could not find a significant difference between the accuracy and speed of the RGB and HSV models, but did find that level of feedback had a significant effect on accuracy of the final match. A qualitative analysis of the way users navigated through the color space indicates that feedback helps users with limited knowledge of the model, allowing them to refine their match to a higher degree of accuracy. Users with very little knowledge or a lot of knowledge of the color model do not appear to be aided by increased feedback.

Future Work

Now that we have identified some important features of the user path data, we would like to code the entire set of data in terms of these features and quantitatively test our conjectures about the mechanics of feedback in color selection and the classes of users who make use of it. We are also interested in looking for several features which may provide clues to the higher-level strategies used to navigate through the color space. Finally, we would like to statistically compare the convergence phases of the subjects in the two feedback groups and determine if this phase of the match is where the feedback made the biggest difference.

Acknowledgments

The original version of the color matching program and the first color matching experiment were written and run by Don Hubbard. Professor Gary Meyer advised us on how to calibrate the monitor and on color model and measurement issues. The paper also benefited tremendously from the thoughtful comments of the reviewers.

References

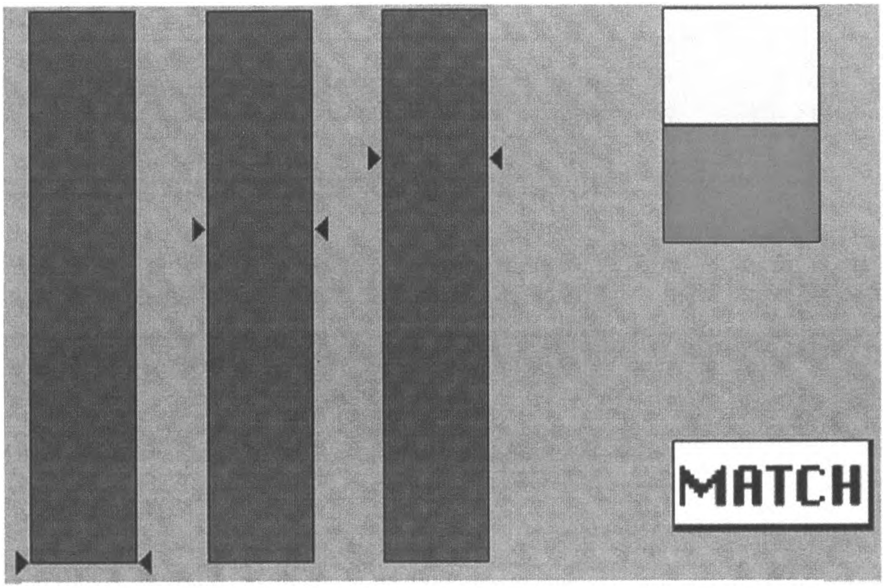
1. Douglas, S., and Kirkpatrick, T. Do color models really make a difference? *Human Factors in Computing Systems: CHI '96 Conference Proceedings*, (in press). ACM Press, 1996.

2. Macbeth. *ColorChecker chart* (1990 ed.). Munsell Color, Baltimore, MD, 1990.

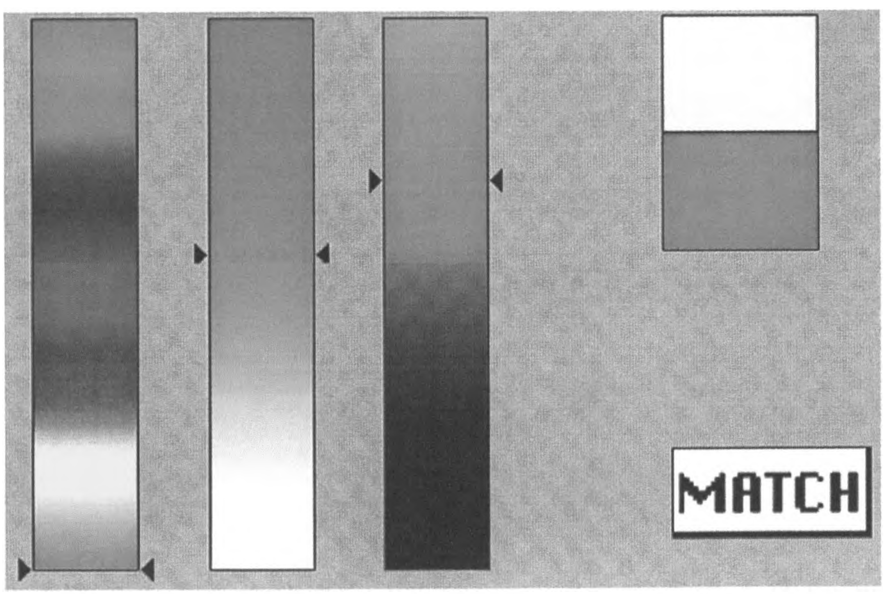
3. Schwarz, M., Cowan, W. and Beatty, J. An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models. *ACM Transactions on Graphics 6*, 2 (April 1987) 123-158.

4. Shneiderman, B. Direct manipulation: A step beyond programming languages. *IEEE Computer 16*, 8 (Aug. 1983) 57–69.

5. Wells, E. *A comparison of interactive color specification systems for human-computer interfaces.* M.S. Thesis in Visualization Sciences, Texas A and M University, 1994.



Douglas and Kirkpatrick, Color Plate 1: "Position-only" interface



Douglas and Kirkpatrick, Color Plate 2: "Position + Effect" interface