Testing the Softproofing Paradigm II

A. Gatt, S. Westland, R. Bala[+] and Y. Ling[*]
Centre for Colour Design Technology, University of Leeds, Leeds LS2 9JT (UK)
[+]Xerox Corporation, Rochester, New York (USA)
[*] Henry Wellcome Building for Neuroecology, University of Newcastle-upon-Tyne (UK)
Corresponding author: A. Gatt (texag@leeds.ac.uk)

ABSTRACT

This article presents results from an experiment that is part of a larger series whose purpose was to evaluate the suitability of softproofing. In the present study, observers were prompted to assess whether two colorimetrically identical stimuli were generated by the same or different media. A complex viewing apparatus was specially designed for this study to ensure that no cognitive cues were visible to observers and that the surround conditions for viewing softcopies and hardcopies were in very close agreement. Overall, the general conclusion that can be drawn is that observers could reliably observe the media *difference* because of the intrinsic inaccuracies of the devices involved, but they could not *identify* the exact type of medium used to generate the stimuli. Although obtaining an exact colorimetric match is thus probably impossible for complex images, the type of medium used to generate stimuli does not influence their appearance.

1. INTRODUCTION

While softproofing is not a new concept to printers, the emergence of a global network has radically modified their production workflow. Although remote softproofing for content has matured to the point of becoming almost a standard in the production workflow[1], proofing for contract colour dropped behind due to diverse technical limitations. Despite advances in colour appearance models and colour management techniques, customers' final approval is still very subjective, and depends on their willingness to trust the softproof as an accurate representation of the appearance of the real print. The precision of the colour match that can be achieved is indeed subject to the accuracy of the colour reproduction media involved. Their colorimetric characterisation is furthermore complicated by the fundamentally different principles used to generate colour information, as simple measurement devices experience problems to evaluate prints and monitors in a consistent manner. In addition, these conceptual differences might account for a more fundamental difference in stimuli perception. Results[2] recently obtained at the late Colour and Imaging Institute challenged a very common practice in both research and industrial settings, i.e. the use of a softproof as a surrogate for the final print in judging colour quality. While studying the quality of diverse gamut mapping algorithms, it was found that their relative performance was very different when a softcopy simulation of the print was used instead of a real hardcopy, despite a careful characterisation procedure. This finding challenged a widely held belief that a colorimetric match is equivalent to a visual match under controlled viewing conditions. However, the validity of this statement is limited in practice, as the precision of a colorimetric match is subject to the accuracy of the devices involved. Obtaining a perfect match across a wide range of complex stimuli such as pictorial images is probably impossible. Practically, the more appropriate question is whether the medium used to generate stimuli has an influence on their appearance. A series of experiments[6] was therefore designed to address the fundamental question as to what effect the underlying media has on the appearance of visual stimuli. The first of them, presented here, investigates whether the fact that stimuli are displayed on different media alters observers' perception of them.

2. METHOD

For this set of experiments, the viewing conditions need to be equated as accurately as possible in order to maximize the apparent similarity of the two stimuli presented to observers. However, softcopies and hardcopies are generally observed under very different conditions. The solution adopted in this work was to place a screen between observers and stimuli that contains two apertures through which the stimuli were displayed. The size of the apertures was adjusted in order to hide anything that was not part of the stimuli themselves, such as the bezel of the screen. Both sides of the apparatus could be used to display either a hardcopy or a softcopy stimulus, so that the location of the two stimuli being compared could be changed at will. Figure 1 shows a schematic diagram of the experimental set-up that was employed.

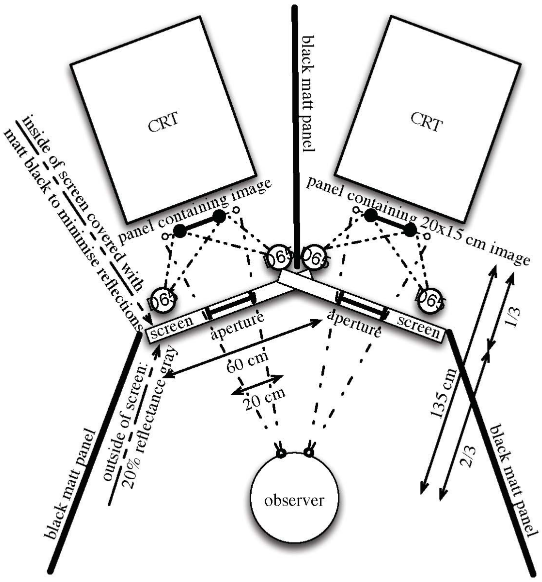

Figure 1: Schematic diagram of the viewing apparatus

Softcopies were displayed on two *Lacie electronblue IV* 19" CRT monitors. Hardcopy reproductions were made using a *Xerox Phaser 7300* printer on *Xerox*'s *Glossy Coated Paper* substrate. The overall performance of the cross-media reproduction framework was considered adequate (average: 3.1 ΔE^{*}_{ab}, maximum: 9.2 ΔE^{*}_{ab}) since it was only slightly worse than the accuracy of the printer. Given the current reproduction abilities of printers, it would be naive to expect to implement a reproduction framework where all errors would lie within the 3 ΔE^{*}_{ab} limit that is usually considered as the just-noticeable-difference observable for pictorial images[5].

Colorimetrically identical stimuli were generated using both types of media and presented simultaneously to observers. Their first task was to *discriminate* the types of media employed, i.e. reporting whether both stimuli were generated using the same medium or different ones. After a response was made, observers were prompted to rate their confidence in their answer on a five-point rating scale. Category *-2* corresponded to "not very confident" and Category *+2* corresponded to "very confident". Observers were then asked to perform their second task, i.e. *identifying* the type of medium used to generate both stimuli. The whole experiment was repeated three times to investigate observers' repeatability. Despite being one of the most commonly used metric to measure performance in the colour-imaging field, proportion correct *p(c)* is actually rarely adequate for this kind of task. A more appropriate performance index is the detection-theory metric *d'*. It has the desirable property of representing a uniform level of performance for any task, and is also independent of bias.

Twelve images covering a wide range of image content and colorimetric characteristics were selected. A small white border surrounding the stimulus itself was also included in order to steady the state of adaptation of observers. The colour gamut of both media being different, the gamuts of all experimental images were restricted to lie within the intersection of each medium's gamut, in order to make the comparison between those media meaningful. Seven women and six men participated in the experiment. They all completed the full Ishihara Test for Color Blindness successfully.

3. RESULTS

3.1 Media Discrimination

The sensitivity metric[3] *d'* can be simply obtained from the difference between the z-scores of the *hits* and *false alarms* rates. A more efficient way to eliminate observer bias is to obtain a full receiver-operator characteristics (ROC) curve[4], where different hit rates are plotted versus their corresponding false alarm rates. The parameters of the best-fitting ROC curve were estimated from the collected [*H*, *F*] pairs using a jackknifing technique involving maximum-likelihood estimation. The ROC curve obtained for the discrimination task is shown in Figure 2. The resulting *d'* value, displayed in Table 1, suggests that, on average, observers were able to reliably discriminate whether the two stimuli were generated by the same type of medium or not. A more intuitive impression of these data can be had from another measure of discriminability, *p(A),* which represents the proportion of the area of the entire [*H*, *F*] graph that lies beneath the ROC curve. Since 0.5 represents chance performance and 1.0 perfect discrimination, it confirms the conclusion that observers could reliably distinguish whether the two stimuli were generated by the same medium or not.

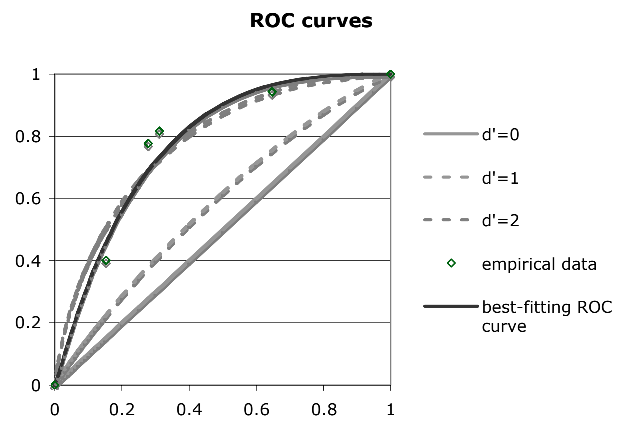

Figure 2: Media discrimination – ROC curves

Table 1: Media discrimination: sensitivity and bias metrics

d'	*p(c)*	*p(A)*	*C*
2.29 ± 0.10	0.77±0.02	0.78±0.02	0.14±0.02

3.2 Media recognition

The performance measure for the media-recognition task was much lower than for the discrimination task (Tab. 2). The corresponding *p(c)*, a metric equivalent to *p(A)*, although statistically significantly different from chance performance, nevertheless indicated that observers experienced difficulties to perform this task. Furthermore, one of the test images exhibited some spatial artifacts that clearly gave away the type of medium used. The data collected from this stimulus were therefore removed from the subsequent analysis. The results obtained from the ROC analysis are not presented here, as no additional conclusions emerge from them.

Table 2: Media recognition: sensitivity and bias metrics

d'	*p(c)*	*C*
1.01 ± 0.10	0.57±0.02	-0.14±0.02

3.3 Observers Repeatability

Table 3 displays the different performance metric *d'* obtained for the three different sessions. Although the accuracy metric obtained for the last session is significantly different from the two others, this result should not be interpreted as a lack of repeatability. Rather, it suggests that observers were in fact *learning* the specific task they were asked to perform, a phenomenon widely observed in the Signal Detection Theory field.

Table 3: Observers' repeatability: sensitivity and bias metrics for the three experimental sessions

	1st session	2nd session	3rd session
d'	2.08 ± 0.19	2.25 ± 0.19	2.44 ± 0.19
C	0.09 ± 0.04	0.05 ± 0.04	0.17 ± 0.04

3.4 Image Types Analysis for Media Discrimination Task

A one-way analysis of variance (ANOVA) was performed on the values of *d'* obtained for each of the 13 observers for both types of image. The average value for each image type can be seen from Table 4. There was no significant difference in the discrimination accuracy (p=0.67) for the two types of images. The level of difficulty of each task was therefore not influenced by the type of image used as stimuli.

Table 4: Sensitivity and bias metrics for different types of images

	Pictorial images	Business graphics
d'	2.05 ± 0.13	1.97 ± 0.16
c	0.22 ± 0.03	0.08 ± 0.03

4. DISCUSSION AND CONCLUSIONS

The results of the media-discrimination task show that observers on average could perceive quite reliably whether the two stimuli were generated by the same or different media. This was to be expected since the average precision of the colour-reproduction framework corresponded to the commonly accepted value for the just-noticeable-difference observable in pictorial images. Furthermore, the equivalent proportion correct value in this case was almost exactly halfway between chance performance and perfect accuracy, which corresponds to the most common location where the threshold is set according to the classical threshold theory. It could therefore be argued that this provides an indirect way to verify the location of the just-noticeable-difference for pictorial images.

Despite the relatively good accuracy achieved by observers, they often reported the difficulties they experienced to perform this task. The differences between the two types of stimuli, although noticeable, were close enough to create confusion. The amount of bias they showed towards "same" answers, as quantified by the *c* metric, reflect well this tendency. Additionally, the obtained ROC curve, besides providing an accurate measure of sensitivity, also allows an inference to be made about the decision strategy that observers adopted. Its shape, more precisely its symmetry with regards to the negative diagonal on the hits vs. false alarms diagram, indicated that the strategy observers chose was not optimal, emphasizing the trouble they were experiencing. This is consistent with previously obtained results, where it was concluded that the multi-dimensionality of colour perception forbids the use of an optimal strategy.

Observers thus experienced difficulties to produce judgements for the first task, but they manage on average to discriminate quite reliably between the two classes of stimuli. After completing the three sessions, observers were prompted to explain in more details how they perform both tasks. The majority of them reported that a difference in saturation between the two media was the main cue helping them to discriminate between both classes of stimuli. Their answers corroborate the analysis performed so far in the sense that the overall precision of colour reproduction was not high enough to go unnoticed by observers. However, their performance in the second task was much worse, although the theoretical level of difficulty of this part is least. Their performance at identifying the medium used was indeed just above chance performance. So, although they could perceive differences between the information generated by both types of medium, they could not utilise these differences to infer reliable information about the exact type of medium employed. The visual signal they received from both types of media had slightly different colorimetric characteristics, but was not significantly different depending on the medium used to generate it. It can be therefore concluded that, under experimental conditions similar to the present ones and subject to the device's precision, the type of medium used to generate a stimulus does not influence its appearance.

References

1. G. A. Bassinger and J. Marin, *GATF 2003 Soft Proofing Study* (GATF, 2003).
2. L. W. MacDonald, J. Morovic and K. Xiao, "A Topographic Gamut Mapping Algorithm", in Colour Imaging Science: Exploiting Digital Media, L. W. MacDonald and M. R. Luo (eds), John Wiley & Sons, 291-317 (2002).
3. G. A. Gescheider, *Psychophysics: The Fundamentals* (Lawrence Ehrlbaum Associates, 1997).
4. N. A Macmillan and C. D. Creelman, *Detection Theory: A User's Guide* (Lawrence Ehrlbaum Associates, 2005)
5. M. Stokes, M. D. Fairchild, and R. S. Berns, Colorimetrically Quantified Visual Tolerances for Pictorial Images, *ISCC/TAGA Comparison of Color Images Presented in Different Media*, vol. 2, 757–777.
6. A. Gatt, S. Westland and R. Bala, "Testing the softproofing paradigm" in *Proceedings of the 12th Color Imaging Conference* (2004), pp. 187-192.