Journal of Imaging Science and Technology® 55(6): 060504-1–060504-8, 2011.

Empirical Investigation of Display Quality

Ludovic Quintard and Mohamed-Chaker Larabi
XLIM-SIC UMR CNRS 6172, University of Poitiers, France
E-mail: chaker.larabi@univ-poitiers.fr

Ludovic Le Bigot
CerCa UMR CNRS 6234, University of Poitiers, France

Abstract. *This study examined how the display quality of liquid crystal display (LCD) devices is assessed. A subjective experiment was conducted to identify the factors involved in quality assessments and then to determine their respective contributions to these assessments. The results are expected to drive the development of an objective quality metric dedicated to color reproduction. The subjective data were analyzed by means of a multivariate analysis of variance and a correlation study. The results showed the main effect of display on all the factors that were investigated, namely, hue, contrast, saturation, naturalness, quality, and texture. Finally, among these perceptual features, contrast and hue have been demonstrated to be the most influential on the overall quality of a displayed image. © 2011 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2011.55.6.060504]

INTRODUCTION

Issues Related to Displays

Display devices are being increasingly used in everyday life. These devices can be television, computer, or cell phone displays, video game console screens, etc., and are based on different technologies, including liquid crystal displays (LCDs),[1–3] plasma displays, and organic light-emitting diodes.[4,5] A digital image observed via one of these devices can be modified according to its technology and intrinsic characteristics.[6,7] For instance, the use of an LCD improves the accuracy of target detection and may also reduce detection time. Although LCDs have several advantages over cathode ray tube displays,[8] the quality of color reproduction can vary from one device to another. For instance, as Engeldrum[9] argues, "quality is the integrated set of perceptions of the overall degree of excellence of the image. Within the framework of medical imaging, quality relates to the diagnostic capability of the image. In other applications, quality gets closer to the concept of beauty."

The aim of this study is to identify and subsequently investigate the correlation between reproduction quality and a particular display device. The first step is to find evidence demonstrating that the factors defined in the literature are indeed useful in assessing the quality of display devices. The second step is to highlight the most important factors for predicting the overall quality of such devices.

Received Mar. 2, 2011; accepted for publication Oct. 12, 2011; published online Dec. 21, 2011.
1062-3701/2011/55(6)/060504/8/$20.00.

How to Evaluate Display Quality?

One way of assessing a display device is to run subjective tests, regarded as a human-centric application. In this context, one device can be compared with a reference device, the goal being to harness human subjectivity to characterize the displays. This approach can be considered as the principal means of device assessment. Therefore, quality reflects the visual sensation of seeing an image through a given device.[10] The reference display is not always an absolute reference: Has it been calibrated/characterized? Is it the one with the widest gamut? It is not easy to answer these questions, as it is very difficult to achieve a perfect characterization or a gamut that is equal to the spectrum locus. Moreover, subjective experiments are very complicated, tedious, and time-consuming, and their design depends on whether the panel is made up of experts or novices. There is considerable interpersonal variability, which requires significant experimental rigor, although statistical tools make it possible to counterbalance these differences. One possible way to avoid these types of subjective testing constraints is to employ objective metrics to assess display quality.

An objective quality metric provides a measure based on a model that must be correlated to a subjective assessment.[11] Objective quality metrics do already exist and can be divided into three families: full reference, reduced reference, and no reference. "Reference" refers to the input signal that represents the standard or reference instrument. In the context of the quality of reproduction devices, the reference is the device that has the best fidelity/quality. The full-reference objective metric uses the original image in its algorithm. For instance, in the case of video encoders, the standard image is the uncompressed one. Examples of full-reference objective metrics are given in Ref. 12. The reduced-reference objective metric, on the other hand, uses only some of the image's properties in its algorithm.[13] These properties may be low-level features that can be computed on both sides of the workflow. This metric is suitable for systems involving transmission. Finally, the no-reference objective metric does not require the original image, but instead works by integrating some of the properties of the human visual system into its algorithm.[13–15] For instance, in the context of video transmission, it will obviously focus on artifacts, such as blurriness, blockiness, and ringing, generated by the codec's technology.[16] In the case of this metric, the visual features must have a mathematical representation and must be discriminant features for human perception.

To develop a no-reference objective metric, then, it is essential to know the importance of each quality factor and its correlation to the overall visual quality. So, the step aiming at defining the quality factors is crucial. Some of these factors come from the color field and have been first identified in the field of visual psychophysics[17] such as:

- **Brightness**: the representation of the visual sensation of light (weak, average, or high)
- **Hue**: the tonality of a color and its relationship to the dominant color name in the human language (red, green, blue, and yellow)
- **Saturation**: the intensity of a given hue. The color can be pure or desaturated (e.g., pink is a desaturated red).

Other features can be found in the literature, as well:

- **Naturalness**: the reflection of memory representation, which is often associated with visual quality.[18]
- **Contrast:** the relationship between the brightness of various objects of a scene.

Present Study

As mentioned previously, the aim of this study is to ascertain whether various different quality factors have an effect on the way humans assess color displays, and to find out which of them is most strongly correlated to overall quality. Color reproduction may vary from one display device to another,[7,8] and the factors investigated include contrast, hue, saturation, and texture (local contrast), parameters taken directly from psychophysics. To measure the influence of these individual features, we also examined two additional general features: overall quality and naturalness.[19,20] Yendrikhovskij,[19] for instance, has demonstrated the importance of some hues, notably those of skin, grass, and sky, in judging overall quality and lists the parameters corresponding to what are judged to be the most faithful reproductions.

Our current subjective study is performed using a wall of LCDs. The participants are asked to assess the images shown on these displays and to answer a detailed questionnaire. The methodology employed (participants, apparatus, display characterization, etc.) is described in the next section. The experimental results obtained by using various statistical tools are then presented in light of the perceptual features and displays assessed in the present study, and these results are discussed with regard to the experimental protocol.

METHODOLOGY

Participants

Thirty participants (15 men and 15 women) between the ages of 20 and 50 (mean, $M = 35.2$; and standard deviation, $SD = 8.90$) took part in the experiment. All the observers had either normal or corrected vision. As recommended by the International Telecommunication Union (ITU), their color vision was checked by means of the Ishihara test and their acuity by the Snellen test.[10] The participants received a compensation of €35.

Equipment and Test Material

Displays and Image Sets

Five widely available LCDs from various brands, named (A,B,C,D,E), were prepared for the assessment. Each display has a diagonal measurement of 38 cm. (4:3 aspect ratio) and a resolution of 1024×768. They have been uniformly characterized and calibrated using a spectroradiometer CS-1000 from Konika-Minolta. The test materials consisted of 12 still images representing portraits, food, and landscapes. The images were selected from the Kodak database[21] (alps02, bora04, burano20) on the Kodak CD (IMG0003, 0013, 0014, 0015, 0018, 0022) and from the *Laboratoire National de Metrologie et d'Essais* (LNE) database (woman, fruit, autumn). They had a resolution of either 768×512 (portrait) or 512×768 (landscape) and were uncompressed images (originals). Thumbnails of these images are given in Figure 1.

A wall of displays was constructed for the purposes of the experiment, as illustrated in Figure 2. They have been arranged in a Latin square between groups of observers to avoid any effect due to position. Observers were asked to evaluate each display separately, while recognizing the presence of the other displays.

The participants were seated on an adjustable chair at a distance of 120 cm from the LCD wall. Each display was calibrated in accordance with the ITU recommendations,[10] with a 6500 K white point of 80 cd/m^2. The displays were measured in terms of color reproducibility.

Figure 3 shows the ability of each display to reproduce colors, while Figure 4 shows the tone reproduction curve (TRC) of each display for the calibration described earlier.

The TRC shows the relationship between the input to the device and the output luminance. The assessment room was designed in accordance with ITU recommendations. The walls were gray, and the ambient light was approximately 60 l×, generated by D50 fluorescent tubes.

Subjective Quality Questionnaire

A six-item satisfaction questionnaire with a five-point ordinal scale (labeled from low = 1 to high = 5) was used. There were two types of items: four questions about perceptual features related to computational tools and two questions about general impression. The perceptual features were (1) image contrast, (2) image hue, (3) color saturation, and (4) image texture. The general impression items were (5) naturalness of the image and (6) overall quality of the image.

Procedure

Each participant was seated in the assessment room, and the questionnaire was explained to him or her. Additional information was provided on a separate sheet of paper, in which each question was described in detail. The French language descriptions of the quality factors were taken from standard definitions of the Commission Internationale de l'Eclairage

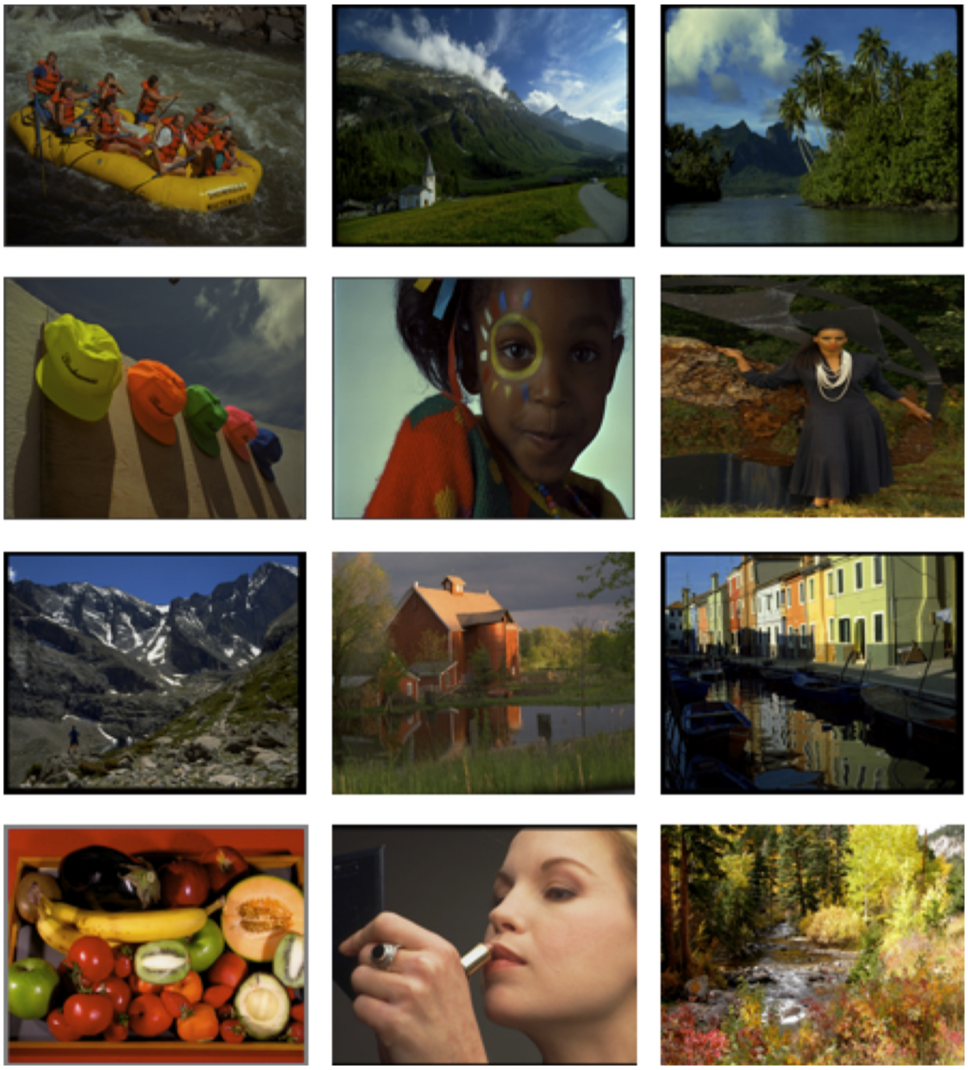

Figure 1. Thumbnails of the images used for the subjective experiments.

(CIE) or other organizations. Five displays were presented to the participant at the same time. To avoid any possible positional bias, there were five display configurations that were counterbalanced using a Latin square. Each participant was assigned to one of the five configurations and was shown 12 images presented in a random order on all five displays at the same time. The assignment was as follows: "The same image will be displayed on all of the screens at the same time. Please answer the different questions for each display/image pair. You do not have to compare them. You must answer the questions about the display/image pairs one by one. When you have finished answering all the questions about one particular image and the five displays, you can move on to the next image by hitting the spacebar."

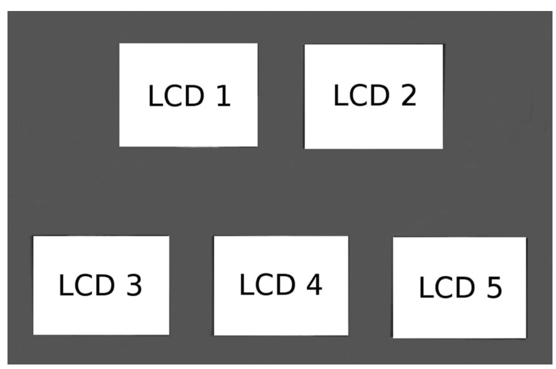

Figure 2. A wall of five LCDs prepared for the experiment.

Data Analysis

In this study, it is important to bear in mind that the data were generated by a repeated-measures experimental design, as every participant in the test answered the same question for a total of 12 images. Display was the independent variable. The six questions represented each dimension. The aim of the analysis was to test whether the means of the five displays were sampled from the same sampling distribution.

Multivariate Analysis of Variance

The multivariate analysis of variance (MANOVA) is an extension of the analysis of variance (ANOVA), where there

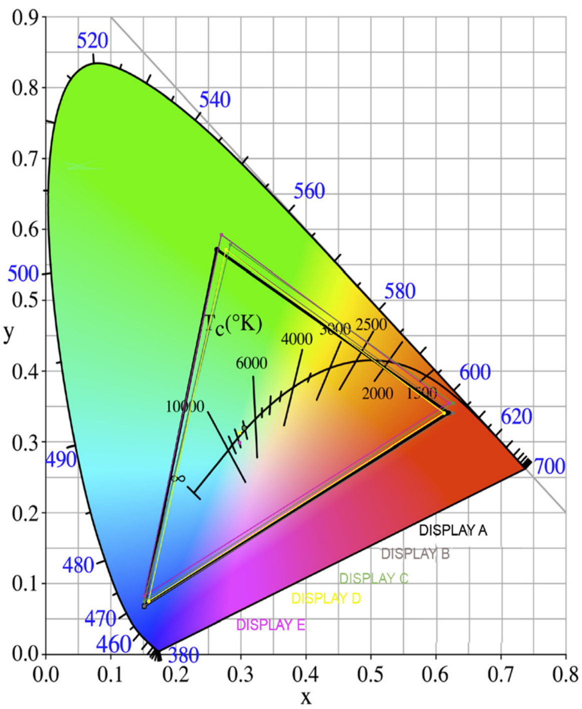

Figure 3. The gamut of the five LCDs.

is more than one dependent variable. The *t*-test is used to assess the difference between the means for two groups. The purpose of an ANOVA is to determine whether the means for two or more groups are taken from the same sampling distribution. The multivariate equivalent of the *t*-test is Hotelling's T-square distribution. Accordingly, the purpose of the MANOVA is to test whether the vectors of the means for two or more groups are sampled from the same sampling distribution.[22]

In this case, the five displays were the five groups and the analysis was conducted to find out whether the means of all the five groups were significantly different. In other words, the null hypothesis for each dimension was that there was no difference between the five displays. However, when a MANOVA is performed, it provides a different criterion, the most widely used one being Wilks' lambda criterion. In the multivariate setting with a combination of dependent variables, Wilks' lambda performs the same role as that of the F-test in a one-way ANOVA. Wilks' lambda is a direct measure of the proportion of variance in the combination of dependent variables, which is not accounted for by the independent variable (the grouping variable or factor). If a large proportion of the variance is accounted for by the independent variable, then this suggests that there is an effect of the grouping variable and that the groups have different mean values. Wilks' lambda can be converted (mathematically adjusted) to a value that has an approximate F distribution, which thus makes it easier to calculate the *p* value. Once the *p* value is known, we can either accept or reject the null hypothesis. If the *p* value is <0.05, we can reject the null hypothesis and accept the fact that the five displays did not come from the same sampling distribution. However, to find out whether just two or three of the displays came from the same sampling distribution, we need to perform a repeated contrast.

Repeated Contrast

The repeated contrast is basically a *t*-test comparison between two groups. With the MANOVA, it is possible to find out whether the vectors of the means of all the groups come from the same sampling distribution. However, with the repeated contrast, it is possible to find out whether the vectors of the two particular groups come from the same sampling distribution. To perform this test, the different group vectors have to be arranged in order from best to worst, after which paired *t*-tests must be performed.

Pearson's Correlation

Pearson's correlation coefficient is used for data on interval or ratio scales and is based on the concept of covariance. When X and Y samples are correlated, they can be said to covary; i.e., they vary in similar patterns. The product-moment r statistic is given by the following equation, in which n is the number of pairs of scores:

$$r = \frac{n\sum_{i=0}^{n} X_i Y_i - \left(\sum_{i=0}^{n} X_i\right)\left(\sum_{i=0}^{n} Y_i\right)}{\sqrt{\left(\left[n\sum_{i=0}^{n} X_i^2 - \left(\sum_{i=0}^{n} X_i\right)^2\right]\left[n\sum_{i=0}^{n} Y_i^2 - \left(\sum_{i=0}^{n} Y_i\right)^2\right]\right)}}. \tag{1}$$

The degree of freedom is $df = n - 2$.

Data Regression

As previously mentioned, this work aims at studying the role of five perceptual features in the quality judgment of a color reproduction device. From the data analysis, a quality formulation based on the features is defined as follows: where a_i are weighting factors associated with the perceptual features.

$$\begin{aligned}\text{Quality} = {} & a_1^* \text{ Contrast} + a_2^* \text{ Hue} + a_3^* \text{ Saturation} \\ & + a_4^* \text{ Texture} + a_5^* \text{ Naturalness}.\end{aligned} \tag{2}$$

The role of the regression, in our case, is to adjust the a_i parameters (Eq. (2)) of the mathematical model $f(x_j, a_i)$, where x_j are the perceptual features. In this study, a partial least square (PLS) regression is used. The parameter a_i was determined using the minimization of S as given in Eq. (3), where N is the size of the data (number of observation) obtained from the psychophysical experiments:

$$S = \sum_{N} [Y_i - f(x_i, a_i)]^2. \tag{3}$$

However, these data are obtained using an ordinal scale, whereas a continuous scale is needed for the regression step. So, the discrete scores are converted into a Z-score by

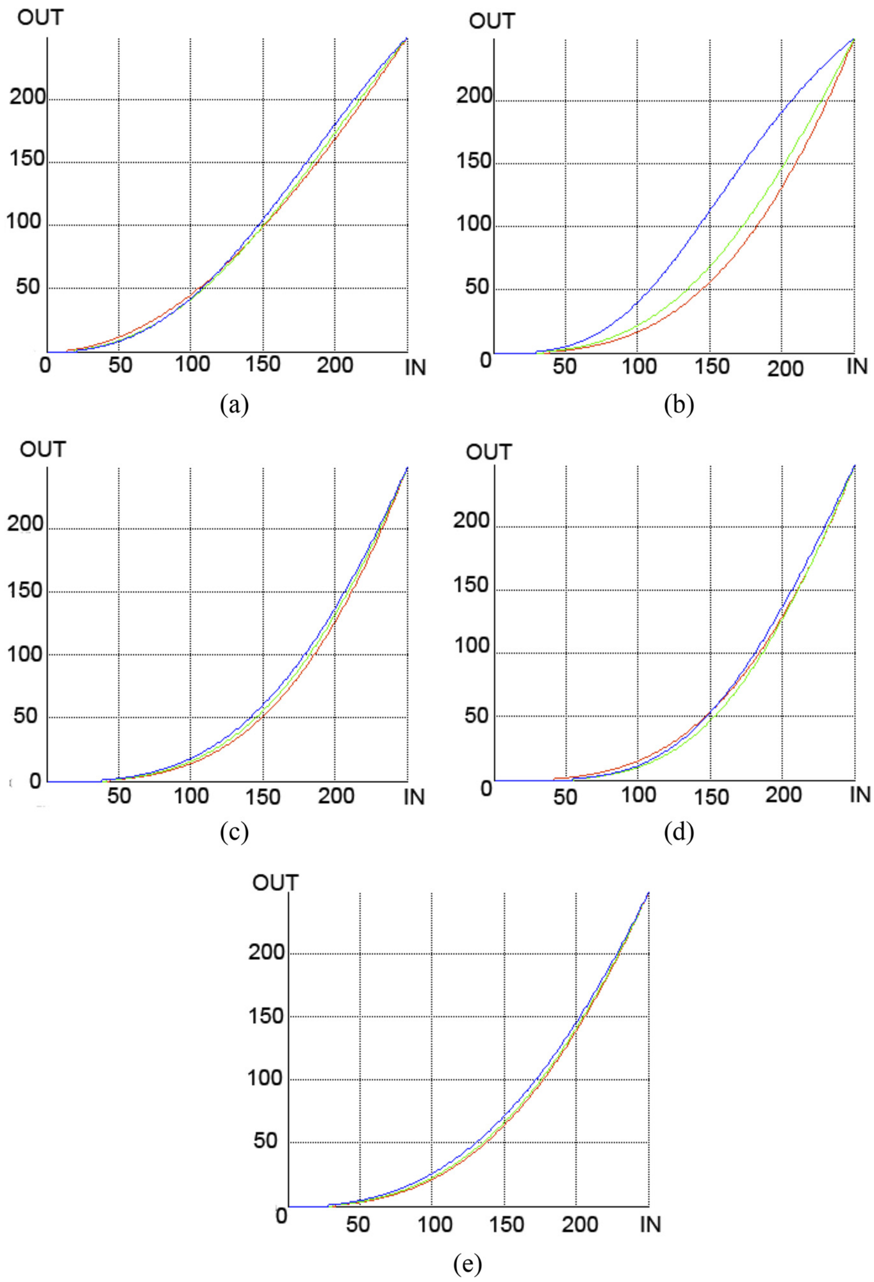

Figure 4. Tone reproduction curves of the displays.

using the mean and the standard deviation as follows, where μ and σ are the mean and the standard deviation of the dimension, respectively.

$$z = \frac{X - \mu}{\sigma}. \tag{4}$$

In previous literature, Choi et al. have proposed an interesting method for quantifying the important factors affecting the perceived quality.[23,24] In this work, six image appearance attributes have been selected: colorfulness, contrast, naturalness, visual information, sharpness, and image quality. There is an important difference between our proposed work and the one proposed in Refs. 22 and 23. In their case, each feature is first quantified and its variation is the generated following a nine-point scale. While our proposed approach is constructed with the aim of finding a formal relationship between the image quality and the perceptual features, The objective of Choi et al. is to determine the behavior of each feature on one display. Nevertheless, in the next stage of our work, the methods used by Choi et al. can be of particular interest.

EXPERIMENTAL RESULTS

A MANOVA was performed on the data from the satisfaction questionnaires. The display was treated as a within-participants factor (F1) and a within-images factor (F2) for each of the dimensions. A contrast analysis of repeated measures (t-test) was also performed for each dimension to

Table I. MANOVAs for dependent measures with F-ratio, degrees of freedom, mean square error, significance, and effect size with repeated measures ($>p<0.05$, $\gg p<0.01$, $\ggg p<0.001$).

	Wilks lambda	df1,df2	F-ratio	*p*-value	Partial η^2	Repeated contrast
Contrast						
Participant (F1)	0.349	4–26	12.14	<0.000	0.651	A,B≫C,D,E
Image (F2)	0.029	4–8	67.84	<0.000	0.971	A,B⋙C,D≫E
Hue						
Participant (F1)	0.364	4–26	11.35	<0.000	0.636	A,B>D,C>E
Image (F2)	0.052	4–8	36.766	<0.000	0.948	A,B≫D,C≫E
Saturation						
Participant (F1)	0.342	4–26	12.51	<0.000	0.658	A>B,D,C,E
Image (F2)	0.072	4–8	25.801	<0.000	0.928	A>B>D>C,E
Texture						
Participant (F1)	0.42	4–26	8.988	<0.000	0.58	A,B,D,C>E
Image (F2)	0.026	4–8	75.36	<0.000	0.974	A,B⋙D,C⋙E
Naturalness						
Participant (F1)	0.309	4–26	14.527	<0.000	0.691	B,A,D,C⋙E
Image (F2)	0.017	4–8	114.738	<0.000	0.983	B,A≫D,C⋙E
Quality						
Participant (F1)	0.319	4–26	13.883	<0.000	0.681	A,B≫D,C≫E
Image (F2)	0.011	4–8	177.191	<0.000	0.989	A,B⋙D,C≫E

obtain the significant difference in reproduction. Pearson's correlation coefficient was then performed to identify the factor(s) that made the greatest contribution to the assessment of overall quality.

The F1 and F2 analyses revealed the main effect of display in the different dimensions (see Tables I and II). For the contrast dimension, the participants perceived a significant difference between the displays. Furthermore, the overall statistics for F1 showed that the best display was A, followed by B, C, D, and E. The repeated-contrast analysis demonstrated that there was no significant difference between A and B, while C, D, and E differed significantly from A and B. For F2, the repeated-contrast analysis revealed a significant difference between E and D–C.

With regard to hue, participants perceived significant differences between the displays. The display statistics indicated that A was considered to be the best, followed by B, D, C, and E. The contrast analyses of repeated measures (t-test) showed that A and B, as well as C and D, were not as significantly different. However, A and B were significantly different from C and D for F1, and very significantly different for F2. E was regarded as the worst, and there was a very significant difference between it and C for F2, and a significant difference for F1.

With regard to saturation, the participants perceived significant differences between the displays. The statistics indicated that A was considered as the best, followed by B, D, C, and E. For F1, the contrast analyses showed that A and B differed significantly, while, B, C, D, and E did not. For F2, A, B, C, and D were estimated to be significantly different, but C and E were not found to be significantly different.

Concerning texture, the participants perceived significant differences between the displays. The statistics indicated that A was considered to be the best, followed by B, D, C, and E. The contrast analyses showed that A, B, C, and D were not regarded as significantly different, whereas E was found to be significantly different using F1. For F2, the t-test showed that A and B, as well as C and D, were not regarded as significantly different, although there was still a considerable difference between the two pairs. D and C were deemed to be highly different from E.

With regard to naturalness, the participants perceived significant differences between the displays. The statistics indicated that B was considered the best, followed by A, D, C, and E. The contrast analyses showed that for F1, A, B, C, and D were not estimated to be significantly different, although they did differ significantly from E. For F2, the analyses showed that A and B, as well as D and C, were not judged to be significantly different. However, B and A were significantly different from C and D. E was regarded to have a high level of significant difference from D and C.

In relation to quality, the participants once again perceived significant differences between the displays. A was considered to be the best, followed by B, D, C, and E. The contrast analyses showed that A and B, along with C and D, were not deemed to be significantly different. However, using F2, A and B were indeed considered to be significantly different from D and C, and E was considered to be significantly different from C.

Pearson's correlation matrix (see Table III) provided the different correlations between the overall quality and the individual quality factors. The correlation was significant for all dimensions.

All the dimensions had a high coefficient of correlation with quality, and hence, each dimension could theoretically be used to compute overall quality. Nevertheless, the lowest correlation coefficient was for texture and the highest correlation coefficient was for hue, thus making hue the most important quality factor.

Using the results given above, we applied the partial least squares regression in order to determine the contribution of each perceptual feature to the judgment of image quality. The first conclusions drawn from this statistical study indicated that the judgment made by the observers on naturalness was quite similar to that of overall quality. Moreover, the debriefing done after the experiments showed that naturalness was judged as a combination of hue and saturation. From this, we decided to ignore naturalness in the regression stage. We can effectively see that the weighting factors of hue and saturation have been increased in comparison to a regression including naturalness.

Table II. M = mean for the different dimensions assessed in the present study. SD1 = standard deviation for participant analysis; SD2 = standard deviation for image analysis.

		A		B		Display C		D		E	
Contrast M		**3.883**		**3.717**		**3.178**		**3.117**		**2.742**	
SD1	SD2	0.623	0.306	0.652	0.16	0.893	0.252	1.129	0.25	1.026	0.211
Hue M		**3.65**		**3.572**		**2.933**		**3.094**		**2.458**	
SD1	SD2	0.722	0.291	0.601	0.338	0.901	0.299	1.038	0.244	0.817	0.27
Saturation M		**3.661**		**3.331**		**2.719**		**3.028**		**2.458**	
SD1	SD2	0.638	0.286	0.733	0.29	0.837	0.288	1.117	0.185	0.833	0.341
Texture M		**3.856**		**3.736**		**3.283**		**3.319**		**2.861**	
SD1	SD2	0.566	0.342	0.655	0.153	0.769	0.137	0.95	0.24	0.859	0.214
Naturalness M		**3.703**		**3.728**		**3.15**		**3.247**		**2.447**	
SD1	SD2	0.617	0.304	0.55	0.187	0.831	0.195	1.022	0.25	0.863	0.259
Quality M		**3.808**		**3.733**		**3.056**		**3.167**		**2.528**	
SD1	SD2	0.646	0.308	0.512	0.308	0.921	0.198	1.102	0.228	0.866	0.277

As such, PLS regression results are given by Eq. (5). It shows clearly that hue is the most important feature followed by saturation, contrast, and texture.

$$\begin{aligned}\text{Quality} = {} & 0.1452^{*}\,\text{Contrast} + 0.5633^{*}\,\text{Hue} \\ & + 0.1646^{*}\,\text{Saturation} + 0.1136^{*}\,\text{Texture}. \qquad (5)\end{aligned}$$

This formulation of the quality function of the four perceptual features is an important step toward the implementation of a metric to better evaluate display quality.

Table III. Pearson's correlation coefficients between the different dimensions and the overall quality. ** ≤0.001.

	Contrast	Hue	Saturation	Texture	Naturalness
Quality	0.884**	0.943**	0.911**	0.871**	0.933**
Significance (2-tailed)	<0.001	<0.001	<0.001	<0.001	<0.001
N	150	150	150	150	150

DISCUSSION

This study was performed to investigate the correlation between different quality factors of color reproduction by LCD devices. The aim was to identify the quality factors used by the participants to assess a device, the hope of integrating the most important ones in an algorithm that would automatically assess the quality of a display. A number of different quality factors can be evaluated in this kind of study (e.g., naturalness), but some cannot be implemented because they do not have a formal or mathematical description. We, therefore, focused on hue, saturation, contrast, and texture. To determine the importance of each factor, a subjective assessment of five displays was created, having participants assess these different factors for 12 different images. The collected data were analyzed by means of (1) a MANOVA, which revealed the main effect of display for the different quality factors, (2) a contrast analysis of repeated measures, which yielded three classifications for the display, and (3) a Pearson's correlation coefficient, which highlighted the importance of hue in determining the color reproduction of a device.

The MANOVA analysis confirmed the importance of the quality factors used to assess color-reproduction devices,[17] such as LCDs. It also demonstrated how each technology reproduces the same information in a different way. Moreover, all the quality factors ranked the displays in the same order, ABDCE, with the exception of contrast, which ranked them as ABCDE. The repeated-measures contrast analysis divided the five displays into three categories of quality: (1) A and B, (2) C and D, and (3) E. This indicates that there was a significant difference between the three categories. This subdivision was confirmed by the analysis of the hue, contrast, texture, and naturalness factors, but not the saturation factor, with which there was no agreement.

Lastly, the analysis performed using Pearson's correlation coefficient indicated that all five factors were more or less correlated with the overall quality. The highest correlation was obtained for the hue factor, followed by naturalness and saturation.

The regression stage allowed for a formulation of the quality of a display as a weighted sum of the perceptual features addressed in the experiments.

CONCLUSION

In this study, we investigated the subjective quality assessment of displays to determine the importance of different factors in the human judgment of quality. Five displays were assessed in this subjective assessment and five selected factors were compared with the overall quality. Three statistical methods were used to determine the importance of each factor with regard to its contribution to quality: (1) The MANOVA, which was used to estimate the effect of a display on the different factors; (2) the contrast analysis of repeated measures, which served to define the categories of displays in relation to their factors; and (3) the Pearson's correlation coefficient studies, which revealed the appropriateness of each factor for the quality assessment.

The initial results demonstrated that all the quality factors used in this experiment, hue, saturation, contrast, texture, and naturalness, could discriminate between the displays in a qualitative way. This indicates that each of these factors could be used to assess display quality. However, Pearson's correlation coefficient showed that of all the different factors, hue, naturalness, and saturation correlated most closely with the overall quality. Nevertheless, the correlations of the other factors were high enough to estimate such quality (>87%). Lastly, the repeated-measures contrast study showed hue to be the most appropriate factor for estimating the quality of a display, as it defined the display categories in the same way as the overall quality factor.

Additional research was now needed to make the best use of the most discriminating factors in the construction of a no-reference objective metric dedicated to display assessment. The regression performed on the data has successfully defined the contribution of each factor in terms of quality judgment. The next research goal should be to separately implement these factors, while being able to both compute overall quality and maintain a high correlation with subjective scores.

REFERENCES

[1] C. Tombling and M. Tillin, "Innovations in LCD technology", Synth. Met. **122**, 209 (2001).

[2] S. Chang, "The TFT-LCD industry in Taiwan: Competitive advantages and future developments", Technol. Soc. **27**, 199 (2005).

[3] M. Katyama, "TFT-LCD technology", Thin Solid Films **341**, 140 (1999).

[4] O. Prache, "Active matrix molecular OLED microdisplays", Displays **22**, 49 (2001).

[5] T. Dobbertin, E. Becker, T. Benstem, G. Ginev, D. Heithcker, H.-H. Johannes, D. Metzdorf, H. Neuner, R. Parashkov, and W. Kowalsky, "OLED matrix displays: In-line process technology and fundamentals", Thin Solid Films **442**, 132 (2003).

[6] M. Chen and C. Lin, "Comparison of TFT-LCD and CRT on visual recognition and subjective preference", Ind. Ergon. **34**, 167 (2004).

[7] G. Sharma, "LCDs versus CRTs: Color-calibration and gamut considerations", Proc. IEEE **90**, 605 (2002).

[8] M. Menozzi, U. Npin, and H. Krueger, "CRT versus LCD: A pilot study on visual performance and suitability of two display technologies for use in office work", Displays **20**, 3 (1999).

[9] P. G. Engeldrum, *Psychometric Scaling* (Imcotek, Winchester, MA, 2000), pp. 94–98.

[10] ITU-R BT.500-11-2002, "Methodology for the subjective assessment of the quality of television pictures" (ITU, Geneva); www.itu.int, accessed November 15, 2011.

[11] VQEG TR-2000, "Final report from the video quality expert group on the validation of objective models of video quality assessment (VQEG)," www.vqeg.org, accessed November 15, 2011.

[12] W. Zhou, L. Ligang, and A. C. Bovik, "Video quality assessment based on structural distorsion measurement", Signal Process. Image Commun. **19**, 121 (2004).

[13] M. Nauge, M.-C. Larabi, and C. Fernandez, "A reduced-reference metric based on the interest points in color images", *Proc. PCS Conference* (IEICE, Tokyo, 2010), pp. 610–613.

[14] T. Brand and M. Queluz, "No-reference image quality assessment based on DCT domain statistics", Signal Process. **88**, 822 (2008).

[15] B. Bringier, N. Richard, M.-C. Larabi, and C. Fernandez-Maloigne, "No-reference perceptual quality assessment of colour image", *Proc. EURASIP EUSIPCO Conference* (Hindawi, Cairo, 2006), pp. 102–108.

[16] S. Winkler, *Digital Video Quality Vision Models and Metrics* (Wiley, Hoboken, NJ, 2005), p. 175.

[17] H.-C. Lee, *Introduction to Color Imaging Science* (Cambridge University Press, United Kingdom, 2005).

[18] T. J. W. M. Janssen and F. J. J. Blommaert, "Predicting the usefulness and naturalness of color reproductions", J. Imaging Sci. Technol. **44**, 93 (2000).

[19] S. N. Yendrikhovskij, "Color reproduction and the naturalness constraint", Ph.D. thesis (Technische Universiteit Eindhoven, The Netherlands, 1998).

[20] C. Boust, H. Chahine, and F. Viénot, "Color correction judgments of digital images by experts and naive observers", *Proc. IS&T's PICS Conference* (IS&T, Springfield, VA, 2003), pp. 4–9.

[21] For more information see Kodak PhotoCD database, ftp://ftp.kodak.com/www/images/pcd, accessed November 15, 2011.

[22] D. C. Howell, *Statistical Methods for Psychology* (Wadsworth, Florence, KY 2006).

[23] S. Y. Choi, M. R. Luo, M. R. Pointer, and P. Rhodes, "Investigation of large display color image appearance I: Important factors affecting perceived quality", J. Imaging Sci. Technol. **52**, 1 (2008).

[24] S. Y. Choi, M. R. Luo, M. R. Pointer, and P. Rhodes, "Predicting perceived colourfulness, contrast, naturalness and quality for colour images reproduced on a large display", *Proc. IS&T/SID's Sixteenth Color Imaging Conference* (IS&T, Springfield, VA, 2008), pp. 158–164.