Int. J. Appl. Math. Comput. Sci., 2001, Vol.11, No.2, 515–535

ANOTHER LOOK AT A MODEL FOR EVALUATING INTERFACE AESTHETICS

David Chek Ling NGO*, John G. BYRNE**

Gestalt psychologists promulgated the principles of visual organisation in the early twentieth century. These principles have been discussed and re-emphasised, and their importance and relevance to user interface design is understood. However, a limited number of systems represent and make adequate use of this knowledge in the form of a design tool that supports certain aspects of the user interface design process. The graphic design rules that these systems use are extremely rudimentary and often vastly oversimplified. Most of them have no concept of design basics such as visual balance or rhythm. In this paper, we attempt to synthesize the guidelines and empirical data related to the formatting of screen layouts into a well-defined model. Fourteen aesthetic characteristics have been selected for that purpose. The results of our exercise suggest that these characteristics are important to prospective viewers.

Keywords: screen design, interface aesthetics, aesthetic measures, aesthetic characteristics, multi-screen interfaces

1. Introduction

Gestalt psychologists promulgated the principles of visual organisation in the early twentieth century. These principles have been discussed and re-emphasised, and their importance and relevance to user interface design is understood. A survey (Mosier and Smith, 1995) was conducted of people who had received a report on guidelines for designing user interface software. The analysis of questionnaire responses indicates that design guidelines are generally considered useful, but there are significant problems in their practical application. For an effective application, generally stated guidelines must be translated into system-specific design rules. However, a limited number of systems represent and make adequate use of this knowledge in the form of a design tool that supports certain aspects of the user interface design process. Early examples of the tools that analyse alphanumeric screens are Tullis' Display Analysis Program (Tullis, 1988), and Streveler and Wasserman's system (Streveler and Wasserman, 1984). Tullis investigated the predicting user performance for static

* Multimedia University, Faculty of Information Technology, 63100 Cyberjaya, Malaysia, e-mail: `david.ngo@mmu.edu.my`.

** Trinity College, Department of Computer Science, Dublin 2, Ireland, e-mail: `john.byrne@cs.tcd.ie`.

alphanumeric displays. He explored the relationship between several metrics, including the overall density of the screen, local density of the screen, grouping of objects, and layout complexity and the time required for users to extract information from the display. He conducted several experiments and developed equations that could predict search times and preference ratings. Streveler and Wasserman described a system for quantitatively assessing screen formats that has many parallels with Tullis' model. Streveler and Wasserman describe three basic techniques for analysing screen formats: a 'boxing' analysis, a 'hot-spot' analysis, and an 'alignment' analysis.

There are two main problems. Firstly, the movement toward GUIs with richer fonts and layout possibilities has reduced interest in these metrics, but better analyses of layouts seem possible (Shneiderman, 1998). While it is conceivable to define a set of variables that characterise the key attributes of many alphanumeric display formats, such a task seems difficult for graphic displays because of their much greater complexity. Our model extends the general research objective to include multi-screen interfaces which are mostly found in one-time GUIs and wizards, as well as in multimedia applications. (With some modification, some of the techniques presented can also be used for other graphic screen types.) Secondly, the graphic design rules that these systems use are extremely rudimentary and often vastly oversimplified. Most of them have no concept of design basics such as visual balance or rhythm. While aesthetic guidelines exist to help designers to create attractive displays, the state of aesthetic theory is relatively primitive. That is, although rules of some kind exist, they have not been formally codified yet. We have introduced mathematical formulae for the aesthetic considerations that many (Berryman, 1990; Galitz, 1993; 1994; 1997; Mullet and Sano, 1995; Reilly and Roach, 1984; 1986; Zelanski and Fisher, 1995) present in much descriptive detail. The objective treatment of aesthetic issues adds much to the current body of work. In an earlier study (Ngo *et al.*, 2000), we presented seven formalised measures: symmetry, sequence, cohesion, regularity, homogeneity, rhythm, and order and complexity, and in this study, seven more are added: balance, equilibrium, unity, proportion, simplicity, density, and economy. We implemented what we felt to be a reasonable set of objectively defined characteristics for describing screen formats that should result in a measure that could accurately define important determinants of system acceptability. Section 2 presents a quantitative way of evaluating the seven new aspects of the screen design. Each of the measures is exemplified by a pair containing a good and a bad design. The model is summarised by a modification of order and complexity where each of the other thirteen measures is an argument of the function. Section 3 discusses a software tool that supports the calculations of the method. In Section 4 we consider an example presented in real screens in which empirical aesthetic data exist for those particular screens. Finally, the paper summaries and reviews the contributions.

2. Aesthetic Measures

Keep in mind that this discussion does not focus on the components on the screen, but on the perception of the structure created by the qualities such as spacing and borders. It is as if the screen is viewed through 'squinted eyes', causing the components themselves to become a blur. As a preliminary study, similar to our previous paper (Ngo *et al.*, 2000) on a different set of measures, we confine ourselves to examining only the dimension and position of rectangular regions in order to control content effects and to facilitate interpretation of the data analyses. The use of simple shapes permits an examination of viewer perceptions in a controlled environment in which two experimental variables, namely, the size and location, are carefully accounted for so that we can draw reasonable conclusions from the experiment's outcome. There are other variables such as the colour, tone and shape, which, we believe, are contributing to the overall aesthetic experience about which we wish to obtain information and for which, at this early stage, the examination is uncritical and constraining. Let us observe that the range of the measures is between 0 (worst) and 1 (best).

2.1. Measure of Balance

A balance can be defined as the distribution of the optical weight in a picture. The optical weight refers to the perception that some objects appear heavier than others. Larger objects are heavier, whereas small objects are lighter. The balance in screen design is achieved by providing an equal weight of screen elements, left and right, top and bottom. The balance is computed as the difference between the total weighting of components on each side of the horizontal and vertical axis and is given by

$$\mathrm{BM} = 1 - \frac{|\mathrm{BM}_{\mathrm{vertical}}| + |\mathrm{BM}_{\mathrm{horizontal}}|}{2} \in [0,1]. \tag{1}$$

Here $\mathrm{BM}_{\mathrm{vertical}}$ and $\mathrm{BM}_{\mathrm{horizontal}}$ are respectively the vertical and horizontal balances with

$$\mathrm{BM}_{\mathrm{vertical}} = \frac{w_L - w_R}{\max\big(|w_L|, |w_R|\big)}, \tag{2}$$

$$\mathrm{BM}_{\mathrm{horizontal}} = \frac{w_T - w_B}{\max\big(|w_T|, |w_B|\big)}, \tag{3}$$

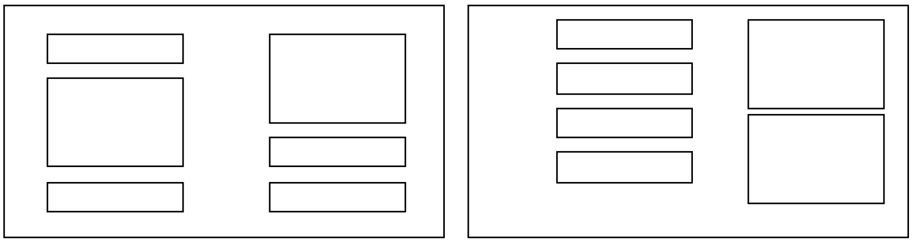
where

$$w_j = \sum_{i}^{n_j} a_{ij} d_{ij}, \quad j = L, R, T, B, \tag{4}$$

L, R, T and B stand for left, right, top and bottom, respectively, a_{ij} is the area of object i on side j, d_{ij} is the distance between the central lines of the object and the frame, n_j is the total number of objects on the side.

$\mathrm{BM}_{\mathrm{vertical}}$ is the normalised difference between the total weighting of objects on each side of the vertical axis. Better values are associated with smaller differences. Positive values indicate that there is more weight on the right side of the frame, while negative values mark the left side. $\mathrm{BM}_{\mathrm{vertical}}$ is zero when there is an equal weight of objects. $\mathrm{BM}_{\mathrm{horizontal}}$ is the normalised difference between the total weighting of objects on each side of the horizontal axis. Positive values indicate that the top half of the frame is heavier than the bottom half, while negative values signal that the bottom half is heavier. Figure 1 presents 'good' and 'bad' versions in a balance study.

In Fig. 1(a) the balance is achieved by providing an equal weight of screen elements, left and right, top and bottom. Figure 1(b) shows a layout in a visual imbalance (it looks as if it were about to topple over).



(a) (b)

Fig. 1. Two versions of screens in a balance study: (a) a good version, (b) a bad version.

2.2. Measure of Equilibrium

The equilibrium is a stabilisation, or a midway centre of suspension. The equilibrium on a screen is accomplished through centring the layout itself. The centre of the layout coincides with that of the frame. (There are minor deviations from this definition which we discuss in Section 5.) The equilibrium is computed as the difference between the mass centre of the displayed elements and the physical centre of the screen, and is given by

$$\mathrm{EM} = 1 - \frac{|\mathrm{EM}_x| + |\mathrm{EM}_y|}{2} \in [0, 1]. \tag{5}$$

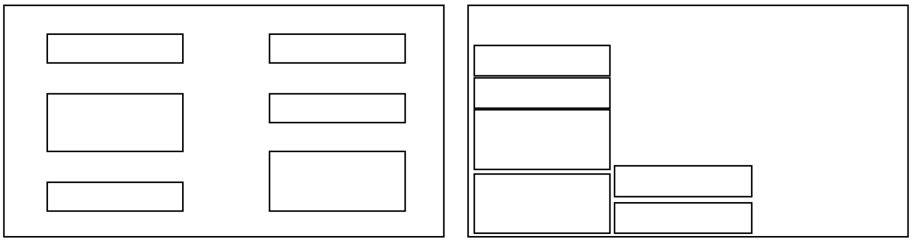
The equilibrium components along the x-(EM_x) and y-axes (EM_y) are respectively given by

$$\mathrm{EM}_x = \frac{2\sum_i^n a_i(x_i - x_c)}{b_{\mathrm{frame}} \sum_i^n a_i}, \tag{6}$$

$$\mathrm{EM}_y = \frac{2\sum_i^n a_i(y_i - y_c)}{h_{\mathrm{frame}} \sum_i^n a_i}, \tag{7}$$

where (x_i, y_i) and (x_c, y_c) are the co-ordinates of the centres of object i and the frame, respectively, a_i is the area of the object, b_{frame} and h_{frame} denote respectively the width and height of the frame, and n is the number of objects on the frame. Note that the maximum values of $|x_i - x_c|$ and $|y_i - y_c|$ are $b_{\mathrm{frame}}/2$ and $h_{\mathrm{frame}}/2$, respectively.

EM_x stands for the normalised x-coordinate of the centre of mass of the objects. Better (smaller) values are related to how closely the centre coincides with that of the frame. Positive values indicate that the centre is situated on the right-hand side of the frame, whereas negative values point out the left-hand side. EM_x is zero when the centre lies somewhere along the x-axis. Similarly, EM_y is the normalised y-coordinate of the centre of mass of the objects. Positive values indicate that the centre lies in the bottom half of the frame, while negative values involve the top half. EM_y is zero when the centre lies somewhere along the y-axis. Figure 2 presents 'good' and 'bad' versions in an equilibrium study. In Fig. 2(a) the equilibrium is achieved by centring the layout itself. The centre of the layout shown in Fig. 2(b) lies somewhat lower than the centre of the frame.



(a) (b)

Fig. 2. Two versions of screens in an equilibrium study: (a) a good version, (b) a bad version.

2.3. Measure of Unity

The unity reflects coherence, a totality of elements that constitute visually one piece. With unity, the elements seem to dovetail so completely that they are seen as one thing. Unity in screen design is achieved by using similar sizes and leaving less space between elements of a screen than at the margins. By definition, it is the extent to which the screen elements seem to belong together and is given by

$$\mathrm{UM} = \frac{|\mathrm{UM}_{\mathrm{form}}| + |\mathrm{UM}_{\mathrm{space}}|}{2} \in [0,1], \tag{8}$$

where $\mathrm{UM}_{\mathrm{form}}$ is the extent to which the objects are related in size with

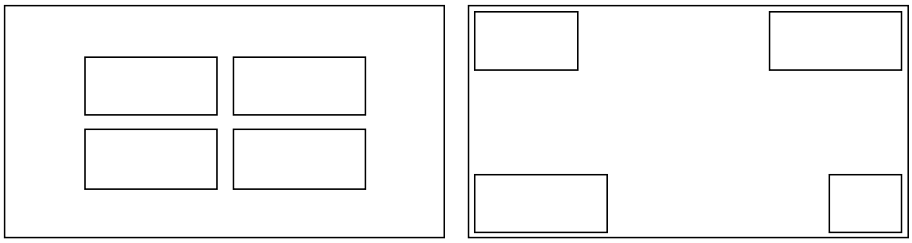
$$\mathrm{UM}_{\mathrm{form}} = 1 - \frac{n_{\mathrm{size}} - 1}{n} \tag{9}$$

and $\mathrm{UM}_{\mathrm{space}}$ is a relative measure of the space between groups and that of the margins with

$$\mathrm{UM}_{\mathrm{space}} = 1 - \frac{a_{\mathrm{layout}} - \sum_i^n a_i}{a_{\mathrm{frame}} - \sum_i^n a_i}. \tag{10}$$

Here a_i, a_{layout} and a_{frame} are the areas of object i, the layout and the frame, respectively, n_{size} stands for the number of sizes used, and n is the number of objects on the frame.

Better (higher) UM_{form} values are associated with fewer object sizes used. UM_{form} is 1 if the layout uses only one size. Better (higher) UM_{space} values are closely related to how tightly the screen is packed. UM_{space} is 1 when there is no space between objects and it is 0 when there is no space left at the margins. Figure 3 presents 'good' and 'bad' versions in a unity study. In Fig. 3(a) unity is achieved by using similar sizes and leaving less space between elements of the screen than at the margins. The elements are related in size and they are grouped together and surrounded by white space. The items in Fig. 3(b) look as if they were ready to move out from the screen.



(a) (b)

Fig. 3. Two versions of screens in a unity study: (a) a good version, (b) a bad version.

2.4. Measure of Proportion

Down through the ages, people and cultures have preferred proportional relationships. What constitutes beauty in one culture is not necessarily considered the same by another culture, but some proportional shapes have passed the test of time and are found in abundance today. Marcus (1992) describes five shapes as aesthetically pleasing: square (1:1), square root of two (1:1.414), golden rectangle (1:1.618), square root of three (1:1.732), and double square (1:2). In screen design, aesthetically pleasing proportions should be considered for major components of the screen, including windows and groups of data and text. The proportion, by definition, is the comparative relationship between the dimensions of the screen components and proportional shapes and is given by

$$PM = \frac{|PM_{object}| + |PM_{layout}|}{2} \in [0, 1]. \tag{11}$$

PM_{object} is the difference between the proportions of the objects and the closest proportional shapes described by Marcus with

$$PM_{object} = \frac{1}{n} \sum_{i}^{n} \left(1 - \frac{\min\left(|p_j - p_i|, j = \mathrm{sq, r2, gr, r3, ds} \right)}{0.5} \right), \tag{12}$$

$$p_i = \begin{cases} r_i & \text{if } r_i \leq 1, \\ 1/r_i & \text{otherwise,} \end{cases} \tag{13}$$

with

$$r_i = \frac{h_i}{b_i}, \tag{14}$$

where b_i and h_i are the width and height of object i, respectively. Note that the maximum value of $(p_j - p_i)$ is 0.5. $\mathrm{PM}_{\mathrm{layout}}$ is the difference between the proportions of the layout and the closest proportional shape with

$$\mathrm{PM}_{\mathrm{layout}} = 1 - \frac{\min\left(|p_j - p_{\mathrm{layout}}|, j = \mathrm{sq}, \mathrm{r2}, \mathrm{gr}, \mathrm{r3}, \mathrm{ds}\right)}{0.5}, \tag{15}$$

where

$$p_{\mathrm{layout}} = \begin{cases} r_{\mathrm{layout}} & \text{if } r \leq 1 \\ 1/r_{\mathrm{layout}} & \text{otherwise} \end{cases} \tag{16}$$

and

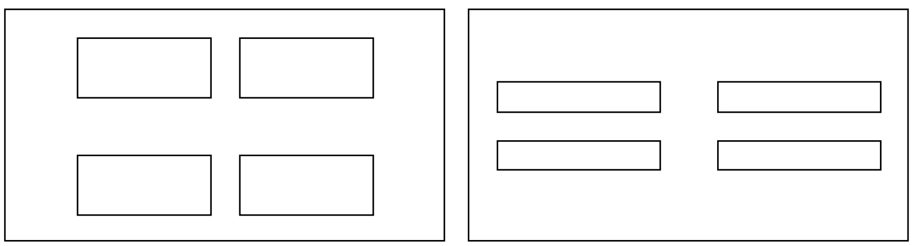
$$r_{\mathrm{layout}} = \frac{h_{\mathrm{layout}}}{b_{\mathrm{layout}}}, \tag{17}$$

b_{layout} and h_{layout} being the width and height of the layout, respectively. Note that the maximum value of $(p_j - p_{\mathrm{layout}})$ is 0.5, p_j being the proportion of shape j with

$$\left\{p_{\mathrm{sq}}, p_{\mathrm{r2}}, p_{\mathrm{gr}}, p_{\mathrm{r3}}, p_{\mathrm{ds}}\right\} = \left\{\frac{1}{1}, \frac{1}{1.414}, \frac{1}{1.618}, \frac{1}{1.732}, \frac{1}{2}\right\}, \tag{18}$$

where sq, r2, gr, r3 and ds stand for the square, square root of two, golden rectangle, square root of three, and double square, respectively.

Better (higher) $\mathrm{PM}_{\mathrm{object}}$ values are related to how proportionate the objects are. Better (higher) $\mathrm{PM}_{\mathrm{layout}}$ values are related to how well the layout is proportionate. Figure 4 presents 'good' and 'bad' versions in a proportion study. In Fig. 4(a) proportion is achieved by creating objects with aesthetically pleasing proportions. The items are close approximations to the proportional rectangles described by Marcus. These proportions cannot be recognised in the items in Fig. 4(b).



(a) (b)

Fig. 4. Two versions of screens in a proportion study: (a) a good version, (b) a bad version.

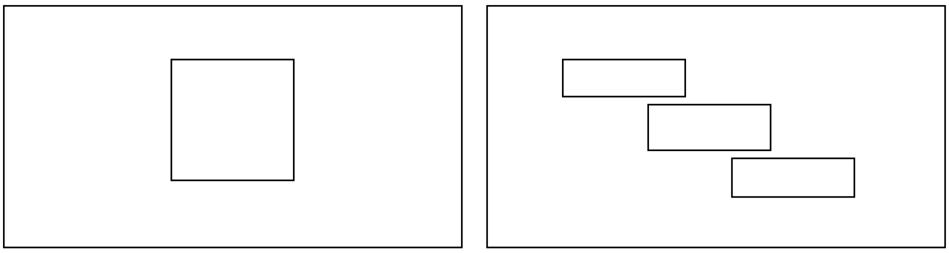
2.5. Measure of Simplicity

Simplicity is directness and singleness of form, a combination of elements that results in ease in comprehending the meaning of a pattern. The simplicity in screen design is achieved by optimising the number of elements on a screen and minimising the alignment points. Tullis (1984) derived a measure of screen complexity for text-based screens based on the work of Bonsiepe (1968) who proposed a method of measuring the complexity of typographically designed pages through the application of information theory. It involves counting the number of different rows or columns on the screen that are used as starting positions of alphanumeric data items. Information theory is then used to calculate the complexity of this arrangement of starting positions. Our method of calculation makes use of the formula

$$\mathrm{SMM} = \frac{3}{n_{\mathrm{vap}} + n_{\mathrm{hap}} + n} \in [0, 1], \tag{19}$$

where n_{vap} and n_{hap} are the numbers of vertical and horizontal alignment points, respectively, and n is the number of objects on the frame.

SMM is inversely proportional to the sum of the numbers of alignment points and screen objects: as the summation decreases, SMM tends to increase. Higher values are associated with a smaller number of objects used and related to how well the objects are aligned. SMM is 1 if there is only one object used. Figure 5 presents 'good' and 'bad' versions in a simplicity study. In Fig. 5(a) simplicity is achieved by minimising the number of elements on a screen and the alignment points. Figure 5(b) has a lower simplicity measure since it has more items and alignment points.



(a) (b)

Fig. 5. Two versions of screens in a simplicity study: (a) a good version, (b) a bad version.

2.6. Measure of Density

The density is the extent to which the screen is covered with objects. Density is achieved by restricting screen density levels to an optimal percentage. A measure of density, derived by Tullis, is the percentage of character positions on the entire frame containing data. Instead of looking at characters, our measure deals with objects through the formula

$$\mathrm{DM} = 1 - 2\left|0.5 - \frac{\sum_{i}^{n} a_i}{a_{\mathrm{frame}}}\right| \in [0,1], \tag{20}$$

where a_i and a_{frame} are the areas of object i and the frame, respectively, and n is the number of objects on the frame. Assume that the optimum screen density level for graphic screens is 50 percent.

DM is the extent to which the percentage of object areas on the entire frame is 50. Higher values are related to how closely the level is 50. DM is 1 when the density level is 50. Figure 6 presents 'good' and 'bad' versions in a density study. In Fig. 6(a) density is achieved by restricting screen density levels to an optimal percent. Figure 6(b) presents a cluttered, cramped layout.

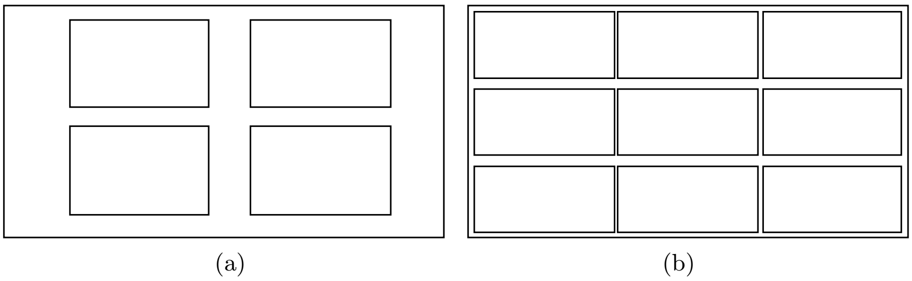


Fig. 6. Two versions of screens in a density study: (a) a good version, (b) a bad version.

2.7. Measure of Economy

The economy stands for careful and discreet use of display elements to get the message across in the simplest possible way. It is achieved by using as few sizes as possible. Economy, by definition, is a measure of how economical the screen is and is given by

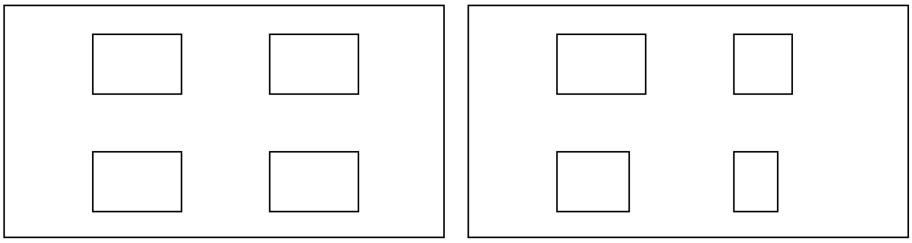
$$\mathrm{ECM} = \frac{1}{n_{\mathrm{size}}} \in [0,1], \tag{21}$$

where n_{size} is the number of sizes.

ECM is inversely proportional to the number of different object sizes: as their number increases, ECM tends to decrease. Higher values are associated with fewer sizes used. ECM is 1 if the number is 1. Figure 7 presents 'good' and 'bad' versions in an economy study. In Fig. 7(a) economy is achieved by using as few sizes as possible. Figure 7(b) has a lower economy measure since it uses more sizes.

2.8. Measure of Order and Complexity

This is a measure of order, which is written as an aggregate of the above measures for layout together with the other six measures described in (Ngo *et al.*, 2000). The



(a) (b)

Fig. 7. Two versions of screens in an economy study: (a) a good version, (b) a bad version.

opposite pole on the continuum is complexity. The scale created may also be considered as a scale of complexity, with an extreme complexity at one end and a minimal complexity (order) at the other. The general form of the measure is given by

$$\mathrm{OM} = g\{f_i(M_i)\} \in [0,1] \tag{22}$$

with

$$\begin{aligned}&\{M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, M_{10}, M_{11}, M_{12}, M_{13}\}\\&\quad = \{\mathrm{BM}, \mathrm{EM}, \mathrm{SYM}, \mathrm{SQM}, \mathrm{CM}, \mathrm{UM}, \mathrm{PM}, \mathrm{SMM}, \mathrm{DM}, \mathrm{RM}, \mathrm{ECM}, \mathrm{HM}, \mathrm{RHM}\},\end{aligned} \tag{23}$$

where $f_i(\cdot)$ is a function of M_i, which is functionally related to the measurable criteria which characterise g, and BM (balance) is given by (1), EM (equilibrium) by (5), SYM (symmetry) by (Ngo *et al.*, 2000, p.103), SQM (sequence) by (Ngo *et al.*, 2000, p.105), CM (cohesion) by (Ngo *et al.*, 2000, p.106), UM (unity) by (8), PM (proportion) by (11), SMM (simplicity) by (19), DM (density) by (20), RM (regularity) by (Ngo *et al.*, 2000, p.107), ECM (economy) by (21), HM (homogeneity) by (Ngo *et al.*, 2000, p.108), and RHM (rhythm) by (Ngo *et al.*, 2000, p.110).

One easy way of interpreting eqn. (22) would be to convert the computed values to z scores, and then simply add them together. Birkhoff (1933) proposed a specific method for describing polygons based on such a linear approach. It involves calculating the sum of four aesthetic measures: vertical symmetry, equilibrium, rotational symmetry, and polygon relation. Likewise, eqn. (22) can be written as the linear summation of the weighted measures with

$$g\{f_i(M_i)\} = \frac{1}{m}\sum_i^m \alpha_i f_i(M_i) = \frac{1}{13}\sum_i^{13} \alpha_i M_i \in [0,1], \quad 0 \le \alpha_i \le 1. \tag{24}$$

Each aesthetic measure M_i has its own weighting component α_i which is assumed to be constant. (Determining weights is one of the multi-dimensional optimisation problems that are application specific. A paper presenting a solution using objective-based evolutionary programming is currently under a review process.) Equation (24) is used to perform OM calculations in this paper and all the weighting components are set to 1, supposing that all these measures are equally important.

3. Aesthetic Measures in an Analysis Program

A program for measuring these fourteen screen format characteristics was written in the C programming language. The input to the program is a model example of the screen to be analysed, drawn over the original screen using a screen editor.

Exemplary 'good' and 'bad' screens are shown in Figs. 8 and 9, respectively, along with the corresponding outputs from the program. The output contains results of the fourteen measurements (balance, equilibrium, symmetry, sequence, cohesion, unity, proportion, simplicity, density, regularity, economy, homogeneity, rhythm, and order and complexity).

Having developed a program for measuring these characteristics of screen formats, the next task was to determine how these characteristics actually relate to the aesthetics of a screen. That was the purpose of the following exercise.

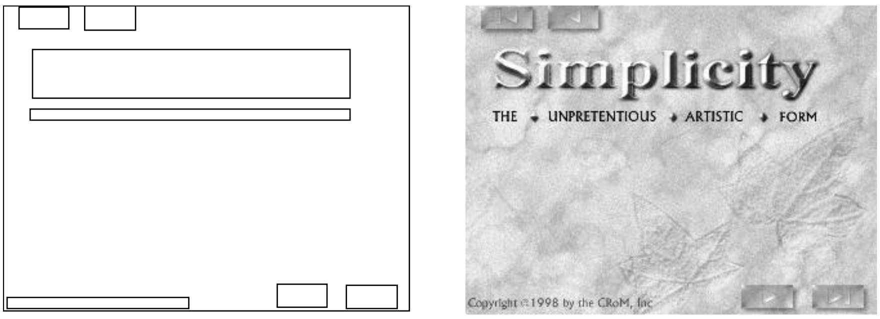


Fig. 8. Exemplary 'bad' screen.

Table 1. Output from the program for a 'bad' screen.

Measures	Values	Comments
M1 (BM)	0.35711	Unbalanced
M2 (EM)	0.80272	Stable
M3 (SYM)	0.45146	Asymmetrical
M4 (SQM)	0.50000	Random
M5 (CM)	0.67934	Cohesive
M6 (UM)	0.10784	Fragmented
M7 (PM)	0.73442	Proportionate
M8 (SMM)	0.14286	Complex
M9 (DM)	0.41532	Cramped
M10 (RM)	0.08333	Irregular
M11 (ECM)	0.14286	Intricate
M12 (HM)	0.00001	Uneven
M14 (RHM)	0.45306	Disorganised
OM	0.37464	Bad

Table 1. (continued)

— M1 —		
BMv:	0.634827	Low/Medium (vertically unbalanced: more weight on the left side)
BMh:	0.650946	Low/Medium (horizontally unbalanced: more weight on the top half)
BM:	0.357114	Low/Medium (unbalanced)
— M2 —		
EMx:	−0.10091	High (equal and frame x-coordinates)
EMy:	−0.29364	High (equal and frame y-coordinates)
EM:	0.802724	High (stable)
— M3 —		
SYMv:	0.645608	Low/Medium (vertically asymmetrical)
SYMh:	0.627489	Low/Medium (horizontally asymmetrical)
SYMr:	0.372511	Medium/High (radially symmetrical)
SYM:	0.451464	Medium (asymmetrical)
— M4 —		
SQM:	0.5	Medium (random)
— M5 —		
CMlo:	0.388273	Low/Medium (unequal object and layout aspect ratios)
CMfl:	0.970402	High (equal layout and frame aspect ratios)
CM:	0.679337	Medium/High (cohesive)
— M6 —		
UMform:	0.142857	Low (many object sizes)
UMspace:	0.072816	Low (less space left at the margins)
UM:	0.107837	Low (fragmented)
— M7 —		
PMo:	0.600173	Medium (aesthetically pleasing objects)
PMl:	0.868676	High (aesthetically pleasing layout)
PM:	0.734424	High (proportionate)
— M8 —		
SMM:	0.142857	Low (complex)
— M9 —		
DM:	0.415319	Medium (cramped)
— M10 —		
RMa:	0	Low (many alignment points)
RMs:	0.166667	Low (many distinct distances between alignment points)
RM:	0.083333	Low (irregular)
— M11 —		
ECM:	0.142857	Low (intricate)
— M12 —		
W:	1.10E+18	
Wmax:	1.93E+23	
HM:	5.68E−06	Low (uneven)
— M14 —		
RHMx:	0.585517	Medium (rhythmic horizontal arrangement variation)
RHMy:	0.553125	Medium (rhythmic vertical arrangement variation)
RHMa:	0.502181	Medium (disorganised dimension variation)
RHM:	0.453059	Medium (disorganised)

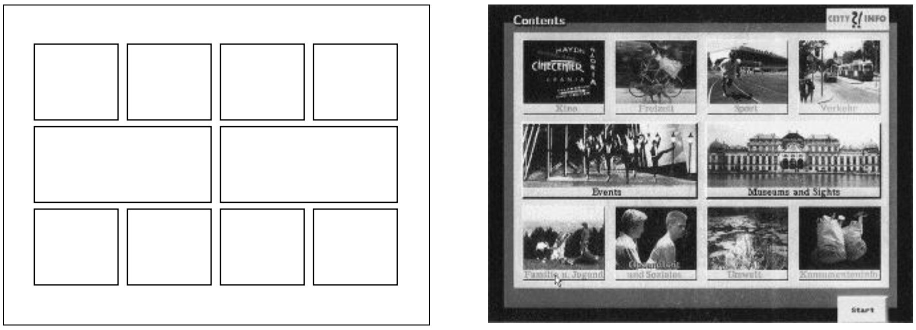


Fig. 9. Exemplary good screen.

Table 2. Output from the analysis program for a 'good' screen.

Measures	Values	Comments
M1 (BM)	0.9963	Balanced
M2 (EM)	1.0000	Stable
M3 (SYM)	0.9985	Symmetrical
M4 (SQM)	1.0000	Sequential
M5 (CM)	0.8012	Cohesive
M6 (UM)	0.8767	Unified
M7 (PM)	0.8686	Proportionate
M8 (SMM)	0.1765	Complex
M9 (DM)	0.8219	Spacious
M10 (RM)	0.7972	Regular
M11 (ECM)	0.5000	Intricate
M12 (HM)	1.0000	Homogeneous
M14 (RHM)	0.9984	Rhythmic
OM	0.8335	Good
— M1 —		
BMv:	0	High (vertically balanced)
BMh:	0.007494	High (horizontally balanced)
BM:	0.996253	High (balanced)
— M2 —		
EMx:	0	High (equal layout and frame x-coordinates)
EMy:	0	High (equal layout and frame y-coordinates)
EM:	1	High (stable)
— M3 —		
SYMv:	0	High (vertically symmetrical)
SYMh:	0.002251	High (horizontally symmetrical)
SYMr:	0.002251	High (radially symmetrical)
SYM:	0.998499	Medium (symmetrical)
— M4 —		
SQM:	1	High (sequential)

Table 2. (continued)

— M5 —		
CMlo:	0.716684	High (equal object and layout aspect ratios)
CMfl:	0.885645	High (equal layout and frame aspect ratios)
CM:	0.801165	High (cohesive)
— M6 —		
UMform:	0.9	High (few object sizes)
UMspace:	0.853359	High (less space between objects)
UM:	0.876679	High (unfied)
— M7 —		
PMo:	0.823134	High (aesthetically pleasing objects)
PMl:	0.91404	High (aesthetically pleasing layout)
PM:	0.868587	High (proportionate)
— M8 —		
SMM:	0.176471	Low (complex)
— M9 —		
DM:	0.821875	High (spacious)
— M10 —		
RMa:	0.65	Medium/High (few alignment points)
RMs:	0.944444	High (few distinct distances between alignment points)
RM:	0.797222	High (regular)
— M11 —		
ECM:	0.5	Medium (intricate)
— M12 —		
W:	2.31E+12	
Wmax:	2.31E+12	
HM:	1	High (homogeneous)
— M14 —		
RHMx:	0	High (rhythmic horizontal arrangement variation)
RHMy:	0.004796	High (rhythmic vertical arrangement variation)
RHMa:	0	High (rhythmic dimension variation)
RHM:	0.998401	High (rhythmic)

4. Validation of the Measures

Two experiments were conducted to evaluate the effectiveness of using the proposed measures to assess interfaces. Experiment 1 tested the validity of the measures, and Experiment 2 tested for potential social bias. The unique aspects of each experiment and its main results are described as follows.

4.1. Study Materials

Samples of all screens are shown in Figs. 10–15. Table 3 presents the element configurations of the screens and their aesthetic values calculated according to our formulae are summarised in Table 4. All values are in pixels. As indicated by the overall measure OM in Table 4, Figs. 10, 12 and 14 are rated as high, whereas Figs. 11, 13 and 15 as low. Among the layouts considered, Fig. 12 has the highest aesthetic value and Fig. 15 the lowest.

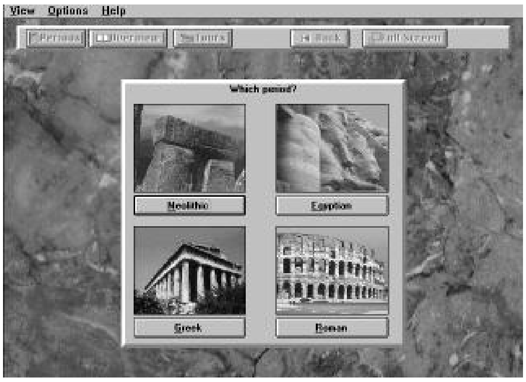


Fig. 10. Exploring Ancient Architecture, by Medio Multimedia, Inc.

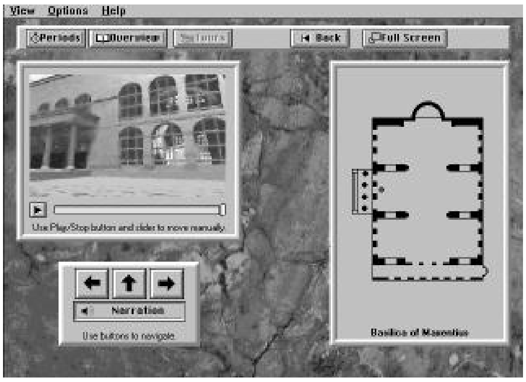


Fig. 11. Exploring Ancient Architecture, by Medio Multimedia, Inc.

Fig. 12. The main menu of the CITY-INFO kiosk.

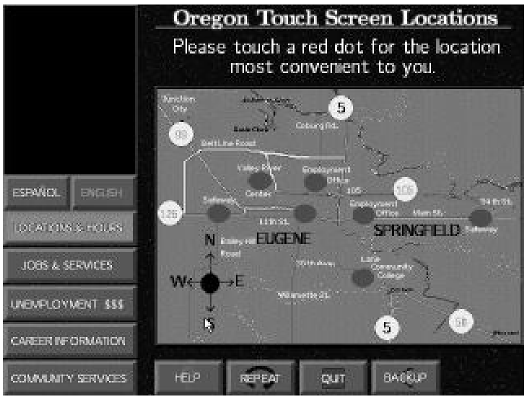


Fig. 13. A regional map showing Oregon Employment Division kiosk locations.

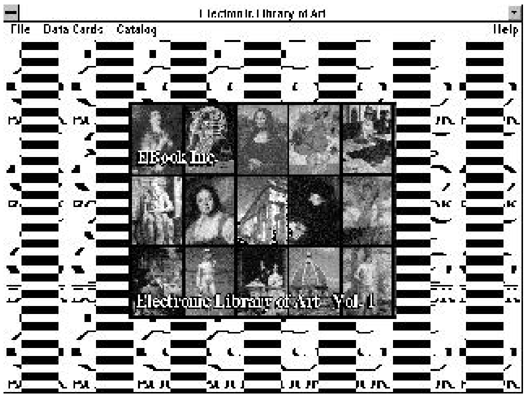


Fig. 14. Title screen from Ebook's A Survey of Western Art.

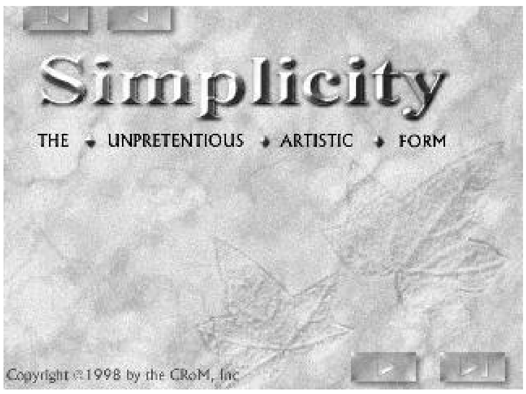


Fig. 15. A screen by CRoM, Inc.

Table 3. Summary of the layout properties.

Layout	Object	X	Y	Width	Height
Fig. 10	1	80	53	70	70
(319 × 221)	2	80	128	70	70
	3	168	53	70	70
	4	168	128	70	70
	5	6	5	306	16
Fig. 11	1	6	5	306	16
(319 × 221)	2	6	25	140	112
	3	32	150	88	52
	4	200	25	112	177
Fig. 12	1	23	29	64	58
(320 × 240)	2	93	29	64	58
	3	163	29	64	58
	4	233	29	64	58
	5	23	91	134	58
	6	163	91	134	58
	7	23	153	64	58
	8	93	153	64	58
	9	163	153	64	58
	10	233	153	64	58
Fig. 13	1	94	2	220	44
(320 × 240)	2	0	104	84	136
	3	94	50	226	160
	4	94	218	176	22
Fig. 14	1	159	81	62	86
(641 × 437)	2	224	81	62	86
	3	289	81	62	86
	4	353	81	62	86
	5	418	81	62	86
	6	159	169	62	86
	7	224	169	62	86
	8	289	169	62	86
	9	353	169	62	86
	10	418	169	62	86
	11	159	256	62	86
	12	224	256	62	86
	13	289	256	62	86
	14	353	256	62	86
	15	418	256	62	86
Fig. 15	1	46	70	499	76
(640 × 480)	2	537	438	81	36
	3	430	435	79	37
	4	26	2	79	36
	5	129	1	79	38
	6	43	163	502	18
	7	6	456	286	18

On a visual basis, the formulae clearly separate Figs. 10, 12 and 14 from Figs. 11, 13 and 15. The former group of screens shows strong aesthetic and organisational qualities. Figs. 10, 12 and 14 are aesthetically balanced with well-defined areas, multiple columns of graphics and with white space that is around the exterior screen margins. The other screens (Figs. 11, 13, 15) reveal three characteristics that are viewed negatively: they are fragmented (mean UM $= 0.193$), irregular (mean RM $= 0.278$), and intricate (mean ECM $= 0.214$). All the screens (10–15) are rather complex (mean SMM ratings $= 0.220$).

4.2. Experiment 1: Subjective Rating of Screens

Experiment 1 was conducted to provide an empirical validation of the measures. The participants were 180 undergraduate students in a Malaysian university. The subjects were members of a variety of information technology classes, who received credit for participating in the study. Although the subjects were members of information technology classes, they were not familiar with screen design concepts. The six design layouts were displayed on grey-scale transparencies in a large classroom, using an overhead screen projector. Each layout was displayed for about 20 seconds. During that time, the participants rated each layout on a low-medium-high scale regarding how nice it was. The median for the ratings was calculated to show the relative aesthetics of the six screens, according to the viewers (see Table 5).

In principle, the relative ratings by the viewers are consistent with those obtained using the proposed computational measures. Figures 10, 12 and 14 were rated as higher than Figs. 11, 13 and 15. To the viewers, the screens in the latter group were much more open, plainer and less interesting than those in the former group. Figures 10, 12 and 14 present three layouts that have high computed values and were rated high on aesthetics by the viewers. Figure 15, the lowest measured screen, was rated as low. Figures 11 and 13 with moderate OM were rated as medium. A relatively high relationship was obtained between these two data sets, which suggests that our model is somewhat related to the viewers' perceptions of aesthetics.

Table 4. Computation of aesthetic values for six layouts.

	Fig. 10	Fig. 11	Fig. 12	Fig. 13	Fig. 14	Fig. 15
BM	0.87412	0.63771	0.99625	0.64216	0.89884	0.35711
EM	0.99368	0.98150	1.00000	0.98134	0.98488	0.80272
SYM	0.66871	0.57828	0.99850	0.44697	0.74258	0.45146
SQM	1.00000	0.50000	1.00000	0.25000	1.00000	0.50000
CM	0.71578	0.74103	0.80116	0.72686	0.71232	0.67934
UM	0.52435	0.32515	0.87668	0.14543	0.99050	0.10784
PM	0.89779	0.83207	0.86859	0.79443	0.88041	0.73442
SMM	0.27273	0.30000	0.17647	0.30000	0.13043	0.14286
DM	0.69493	0.72407	0.82188	0.40792	0.57105	0.41532
RM	0.51250	0.37500	0.79722	0.37500	0.81310	0.08333
ECM	0.50000	0.25000	0.50000	0.25000	1.00000	0.14286
HM	0.69444	0.00463	1.00000	0.00714	0.00340	0.00001
RHM	0.66176	0.54721	0.99840	0.46527	0.73516	0.45306
OM	0.69314	0.52282	0.83347	0.44558	0.72790	0.37464

Table 5. Medians of subjective ratings from Experiment 1.

Layout	Fig. 10	Fig. 11	Fig. 12	Fig. 13	Fig. 14	Fig. 15
Exp. 1's Score (Median)	High	Medium	High	Medium	High	Low

4.3. Experiment 2: Testing for Social Bias

Experiment 2 was designed to test the robustness of the results in Experiment 1 to social differences. A different group of twenty six subjects, including nine GUI designers, six industrial designers, eight lab engineers and three secretaries, participated in the experiment. The procedure used was identical to the one used before: The six design layouts were displayed on grey-scale transparencies in a large classroom, using an overhead screen projector. Each layout was displayed for about 20 seconds. During that time, participants rated each layout on a low-medium-high scale regarding how nice it was.

The median for the ratings was calculated to show the relative aesthetics of the five screens, according to the viewers (see Table 6). The results resemble those of Experiment 1, weakening the alternative explanation that the relationships between viewer judgements from Experiment 1 and computed values from the analysis program are primarily the result of an in-group bias. Thus, Experiment 2 lends further support to the overall strong correlation between perceived and measured aesthetics of the interface.

Table 6. Medians of subjective ratings from Experiment 2.

Layout	Fig. 10	Fig. 11	Fig. 12	Fig. 13	Fig. 14	Fig. 15
Score (Median)	High	Medium	High	Medium	High	Low

5. Conclusions and Future Work

In this article, we have presented a computational theory of evaluating interface aesthetics. The results of our informal study confirmed the effectiveness of the model, but at the same time they also suggested some improvements to enhance its usability. We can increase the scope to include the colour, tone and shape of objects in balance, for example. A designer can control some elements of composition to achieve a balance. For instance, the colour is visually heavier than black and white; big things are visually heavier than little things; black is visually heavier than white; irregular shapes are visually heavier than regular shapes. By controlling the colour, size, tone and shape of objects in a design, one distributes the visual weight and thus influences the balance. Much of the descriptive work has already been done (Simpson, 1998; Tjalve, 1979; Zelanski and Fisher, 1995). Secondly, a layout is in equilibrium when its centre corresponds approximately to the centre of the frame. Practically speaking, there are, however, minor deviations from this definition. Owing to the visual gravitational pull, the balancing centre of the layout will lie somewhat higher than the centre of the frame, thereby compensating for the greater weight in the upper half of the area. But such discrepancies are small.

There are many interesting research topics involving the computation and use of our formulae. First, experiments must be conducted to provide an additional empirical validation of the formulae and conventions. It should be emphasized that we had to make two assumptions: (1) that the interaction between the selected characteristics is linear, and (2) that all these characteristics are equally important. Future research should focus on investigating the interplay between the selected characteristics which, in contrast to our original assertion, may be nonlinear. Additional research is also necessary to evaluate the effects of different weighting strategies. (Weighting deals with the problem that we care about some characteristics more than about others.) The characteristics that are common to the feeling which gives one an aesthetic experience should not be limited to the few described here, and more appropriate ordering principles with the corresponding design conventions must be found if this approach is to be improved.

Acknowledgement

The research reported here was supported by the IRPA under Grant No. 191/9636/0012. The authors wish to express their gratitude to Boon Leong Law for his help in conducting the experiments and analysing the data.

References

Berryman G. (Rev. Ed.)(1990): *Notes on Graphic Design and Visual Communication.* — Menlo Park: Crisp.

Birkhoff G.D. (1933): *Aesthetic Measure.* — Cambridge: Harvard University Press.

Bonsiepe G. (1968): *A method of quantifying order in typographic design.* — J. Typograph. Res., Vol.2, No.3, pp.203–220.

Galitz W.O. (1993): *User Interface Screen Design.* — Boston: QED.

Galitz W.O. (1994): *It's Time to Clean Your Windows: Designing GUIs That Work.* — New York: Wiley-QED.

Galitz W.O. (1997): *The Essential Guide to User Interface Design.* — New York: Wiley.

Marcus A. (1992): *Graphic Design for Electronic Documents and User Interfaces.* — New York: ACM Press.

Mosier J.N. and Smith S.L. (1995): *Application of guidelines for designing user interface software*, In: Human Factors Perspectives on Human-Computer Interaction: Selections from Proceedings of Human Factors and Ergonomics Society Annual Meetings, 1983–1994 (G. Perlman, G.K. Green and M.S. Wogalter, Eds.). — Santa Monica: HFES.

Mullet K. and Sano D. (1995): *Designing Visual Interfaces: Communication Oriented Techniques.* — Mountain View: SunSoft Press.

Ngo D.C.L., Teo L.S. and Byrne J.G. (2000): *A method for determining the properties of multi-screen interfaces.* — Int. J. Appl. Math. Comput. Sci., Vol.10, No.2, pp.101–115.

Reilly S.S. and Roach J.W. (1984): *Improved visual design for graphics display.* — Comput. Graph. Applicns., Vol.4, No.2, pp.42–51.

Reilly S.S. and Roach J.W. (1986): *Designing human/computer interfaces: A comparison of human factors and graphic arts principles.* — Educat. Technol., Vol.26, No.1, pp.36–40.

Shneiderman B. (1998): *Designing the User Interface: Strategies for Effective Human-Computer Interaction (3rd Ed.)*. — Reading: Addison-Wesley.

Simpson J.W. (1998): *Creating Meaning Through Art*. — Englewood Cliffs: Prentice Hall.

Streveler D.J. and Wasserman A.I. (1984): *Quantitative Measures of the Spatial Properties of Screen Designs*. — Proc. 1st IFIP Human-Computer Interaction Conf. *INTERACT'84*, London, U.K., pp.1.125–1.133.

Tjalve E. (1979): *A Short Course in Industrial Design*. — Boston: Newne-Butterworths.

Tullis T.S. (1984): *Predicting the Usability of Alphanumeric Displays*. — Ph.D. Thesis, Rice University, Kansas.

Tullis T.S. (1988): *Screen design*, In: Handbook of Human-Computer Interaction (M. Helander, Ed.). — Amsterdam: Elsevier.

Zelanski P. and Fisher M.P. (1995): *Design Principles and Problems (2nd E d.)*. — Fort Worth: Harcourt Brace College Publishers.

Received: 14 August 2000
Revised: 13 December 2000