

Investigating objective measures of web page aesthetics and usability

Helen C. Purchase

School of Computing Science,
University of Glasgow,
Glasgow, Scotland, G12 8QQ

helen.purchase@glasgow.ac.uk

John Hamer, Adrian Jamieson, Oran Ryan

Department of Computer Science,
University of Auckland,
Private Bag 92019, Auckland 1142,
New Zealand

j.hamer@cs.auckland.ac.nz
{[ajam063](mailto:ajam063@cs.auckland.ac.nz),[orya001](mailto:orya001@cs.auckland.ac.nz)}@aucklanduni.ac.nz

Abstract

As part of the ongoing debate about the role of aesthetic design of interfaces, this paper presents an aesthetic evaluation tool which quantifies the layout characteristics of a web page according to fourteen different metrics. By using the rich medium of web pages as our input, we have significantly extended the prior work done in this area which has typically focussed on simple interfaces. We report the results of an experiment to determine whether users' judgements of 'aesthetic appeal' and 'perceived usability' match the numeric metric results. We found that aesthetic appeal (but not perceived usability) was captured by a metric that considered the placement of all objects on the screen, and that the placement of images is a strong predictor of both aesthetic appeal and perceived usability. We suggest practical implications of this work for web page designers.

Keywords: aesthetics, layout, empirical study, perception.

1 Introduction

There is an increasing recognition of the role of visual aesthetics in interface design. Tractinsky (2004) relates interface aesthetics to three facets of architecture: strength, utility and beauty, saying that the first two facets have been a focus of system design for some time (in the form of system functionality and usability), and that the latter is becoming more important. Even Norman who is known for criticizing elegantly designed artefacts for their poor usability (Norman 2002) has acknowledged the importance of system beauty (Norman 2004).

Some evaluation studies have been performed on assessing the aesthetics of interfaces (e.g. Hartmann (2006), Hassenzahl (2004), Kuroso and Kashimura (1995)), and some theoretical work has been done on frameworks for the investigation of system aesthetics and on relevant and useful terminology (e.g. Hartmann and Suttcliffe (2005), Tractinsky (2004)).

When designing an interactive system, the set of functional requirements are specified in terms of inputs, processes and desired outputs. It is much more difficult to specify 'aesthetic' requirements in a rigorous and

quantifiable manner. Recent research has demonstrated that the aesthetic appearance of an interface is important, not just with respect to users' preferences (Pandir and Knight (2006)) and perception of usability (Kuroso and Kashimura (1995)), but with respect to their performance in visual search tasks (Salimun et al., 2010).

This paper brings together two aspects of visual aesthetics research: the means of measuring 'aesthetic appeal' in an objective manner (based on the work of Ngo et al. (e.g. Ngo, Teo and Byrne 2000)) and user ranking studies of the visual appearance of web pages (similar to work done by Pandir and Knight (2006)). This paper describes our adaptation and implementation of metric formulae for web pages, and our experiment to determine the extent to which these formulae relate to user perception. Our aim is to investigate whether the use of objective, measurable aesthetic formulae can help in producing aesthetically pleasing web pages.

2 Background

2.1 Visual aesthetics research

A limited, yet influential study by Kuroso and Kashimura (1995) provided initial data on the visual aesthetics of screen design in relation to both "apparent" and "inherent" usability. They used seven independent variables relating to interface design decisions for Automated Teller Machines (e.g. location of the main display, numerical sequence on the keypad), and found strong relationships between participants' judgments of usability and their aesthetic judgment. Tractinsky duplicated this experiment in Israel (1997), investigating whether Kuroso and Kashimura's results were affected by cultural or methodological bias. The former study's results were validated, with some cultural bias found.

Since these two influential studies, an increasing amount of research (both empirical and theoretical) has addressed the aesthetic design of interfaces. In most cases, the goal of the research is guidelines for the design of aesthetically pleasing interfaces.

Many studies have used web pages as their stimuli. De Angeli et al. (2006) conducted an experiment comparing two different interface styles (menu-based and interactive metaphors) on web sites with equivalent information. They found that participants' perception of the quality of the information was affected by the interaction style. Hartmann (2006) found that aesthetics affected perceptions of web page usability and content, but that the results were affected by users' backgrounds and tasks.

Knight and Pandir (2004) used existing web sites in relation to “pleasingness”, “complexity”, and “interestingness”, and found, in accordance with the theoretical approach of Berlyne (as discussed in Knight and Pandir 2004) which suggests that moderate arousal is most pleasurable, that the most pleasing web pages were not the most interesting nor the most complex. In their next study, Pandir and Knight (2006) demonstrated that when participants ranked web home pages, complexity was not a predictor of aesthetic pleasure. Parizotto-Ribero and Hammond (2004) used Gestalt theories as the basis of their screen layout experiment of five layout guidelines, presenting ‘good’ and ‘bad’ versions for participants to choose from; it is not clear whether their stimuli were abstract schematic diagrams or actual content-rich interfaces.

Some researchers have also attacked the daunting problem of defining “aesthetics” and proposing theoretical frameworks for aesthetic evaluation. Tractinsky’s model (2004) defines a process from design characteristics, through aesthetic processes and evaluation, to outcomes (e.g. attitudes, motivation etc.). Hartmann and Sutcliffe’s model (2005) links components of aesthetic judgment, identifying the relevant areas for potential study. Both groups of researchers acknowledge the importance of context and content in influencing aesthetic judgments, identify aspects of interaction design that could affect such judgments and are worthy of investigation, and describe the possible data that could be collected from such studies.

Lavie and Tractinsky (2004) provide an extensive list of aesthetic terminology to support subsequent experiments, and distinguish between “empirical studies of aesthetics” involving controlled studies with manipulation of visual variables, and an “exploratory approach” involving evaluation of existing stimuli. Most existing work falls into the latter category: while this has produced interesting results, without being able to relate the data back to quantitative or well-defined qualitative descriptions of the stimuli, the results remain descriptive rather than explanatory.

As part of all this research is a wide-ranging and continuing debate on definitions of beauty, aesthetics, goodness, usability, attractiveness etc., and the processes of perceiving beauty (visceral, behavioral and reflective) (Norman 2004).

2.2 Objective measurements of spatial layout

In this paper, we focus on the work done by David Ngo (e.g. Ngo et al. (2000); Ngo and Byrne (2001), Ngo (2001), Ngo, Teo and Byrne (2003)) who defined formulae for objectively quantifying different layout aspects of an interface. These formulae produce values between 0 and 1, each an indicator of the presence of an aesthetic feature of the interface (e.g. symmetry, balance).

Ngo uses the term ‘aesthetic’ for his metrics, even though in effect they characterise simply the placement of objects in 2D space. The traditional definition of ‘aesthetics’ with respect to the visual sense is much richer than simple object placement, encompassing the use of colour, texture and contrast. While we therefore prefer the term ‘visual layout’ for these measures, we use ‘aesthetic layout’ so as to remain consistent with Ngo’s work.

Ngo proposed four initial metrics (Ngo et al. 2000) (Balance, Equilibrium, Symmetry and Sequence), and conducted a pen-and-paper empirical study with graphic designers to validate them, showing that these metrics correlated with users’ perception of aesthetics. These findings were promising, as they demonstrated that the study of aesthetics could be translated to screen layout in terms of objective measures, and that measures could be designed so as to match perception.

Later, Ngo extended his measures to fourteen, with thirteen characteristics of an interface, and a linear combination of these characteristics to calculate the 14th metric ‘Order and Complexity’. He validated these measures (Ngo and Byrne 2001) and investigated whether they could be used to determine users’ acceptance of data entry screens. In the first experiment he asked seven designers to rank 57 screens; from these results he proposed a regression formula. He then proved that this formula could predict (within a small range) the rankings of new participants on different screens.

Ngo himself has questioned whether the different metrics should be weighted equally in the ‘Order and Complexity’ overall aesthetic calculation (Ngo, Teo, and Byrne 2000). Harrington et al. (2004) consider aesthetic layout for documents (rather than interfaces), and use a different set of measures for which they propose and justify a non-linear aggregation method.

This paper extends this use of Ngo’s measures in two significant ways: we use web pages, a much richer type of interface than any of those used in Ngo’s studies. In doing so, we adapt the use of the measures so as to appropriately deal with different types of visual object. While the main aim of our experiment is to validate the application of these metrics to web page design in terms of the perception of aesthetic appeal, we also consider perceived usability and the effect of colour.

2.3 The Metrics

The complete list of metric formulae can be found in Ngo and Byrne (2001); brief definitions from Ngo (2001) are in the appendix. Here we present three example metrics so as to demonstrate their purpose and application.

2.3.1 Balance

Balance is the distribution of optical weight in a picture: larger objects appear heavier than smaller ones. Good balance has equal weight of screen elements left and right, top and bottom. Fig 1(a) is a web page with poor balance, while Fig 1(b) has good balance.¹

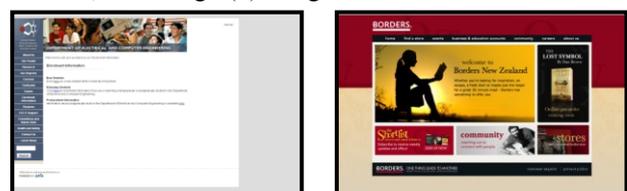


Fig 1: Two pages used in our empirical study: (a) with low balance value (L); (b) with high balance value (B).

¹ These, and all other web pages used as examples in this paper were used in our empirical study. They are identified by A-O, and are listed in the appendix.

The formula for the balance metric is:

$$BM = 1 - \frac{|BM_{vertical}| + |BM_{horizontal}|}{2} \in [0,1]$$

$BM_{vertical}$ and $BM_{horizontal}$ are, respectively, the vertical and horizontal balances with

$$BM_{vertical} = \frac{w_L - w_R}{\max(|w_L|, |w_R|)}$$

$$BM_{horizontal} = \frac{w_T - w_B}{\max(|w_T|, |w_B|)}$$

$$\text{with } w_j = \sum_i^{n_j} a_{ij} d_{ij} \quad j = L, R, T, B$$

where L, R, T, and B stand for left, right, top and bottom respectively; w_j is the total weight of side j ; 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; and n_j is the total number of objects on the side.

2.3.2 Simplicity

Simplicity is ‘directness and singleness of form’ (Ngo and Byrne, 2001). The metric involves counting the number of alignment points (the rows and columns on the screen that are used as starting positions for objects): high simplicity has few alignment points. Fig 2(a) is a page with low simplicity, while Fig 2(b) has high simplicity.

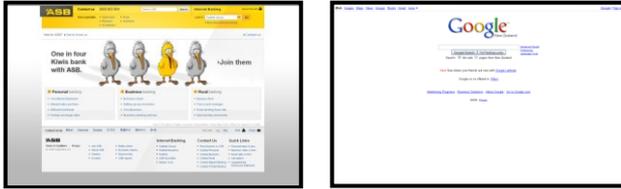


Fig 2: Two of the pages used in our empirical study: (a) page with low simplicity value (G); (b) page with high simplicity value (D).

The formula for simplicity is:

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

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

2.3.3 Unity

Unity is the extent to which elements are perceived together as a whole. The metric is based on the similarity of the size of the objects, and the space left between them. Fig 3(a) is a web page with low unity, while Fig 3(b) has high unity.

The formula for unity is:

$$UM = \frac{|UM_{form}| + |UM_{space}|}{2} \in [0,1]$$

UM_{form} is the extent to which the objects are related in size, with

$$UM_{form} = 1 - \frac{n_{size} - 1}{n}$$

and UM_{space} is a relative measure of the space between groups and that of the margins with

$$UM_{space} = 1 - \frac{a_{layout} - \sum_{i=1}^n a_i}{a_{frame} - \sum_{i=1}^n a_i}$$

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



Fig 3: Two of the pages used in our empirical study: (a) page with low unity value (I); (b) page with high unity value (A).

3 Implementation of the visual layout metrics for web pages

These metric definitions were intended for single screens of an interactive system (Ngo refers to them as ‘‘multi-screen interfaces’’), and are based on the positions of rectangular interface elements (for example, a window pane, or a button) on the screen. However, a browser is an example of a ‘‘multi-pane interface’’ (Ngo, Teo and Byrne, 2003), as different pages can be accessed from the same browser system. In this project, we confined our efforts to single web pages within the browser software, not considering the interface elements of the browser itself. And, by only using single-page web pages (i.e. those that fill the visual pane of a browser and do not need to be scrolled), Ngo’s multi-screen interface layout metrics could be directly applied: there is a single rectangular area within which visual elements are placed.

The calculation of the metrics required that a web page be taken as input to a program, and the numeric values for each of the metrics produced as output. To do this, the rectangular areas of all the component visual elements on the page were identified, as the metrics are solely based on the size and position of rectangular elements. The input could have been represented as either an image file (a screen dump of the web page) or as its source HTML code. In the former case, image processing algorithms would have been required to identify the elements. We chose rather to use the HTML code as we knew that it would clearly and unambiguously represent each of the visual elements, whereas we could not rely on the image processing approach being able to correctly distinguish the element edges.

We implemented a Firefox extension in JavaScript that scans the DOM of the currently loaded web page and calculates the values for 10 individual metrics as well as the composite ‘Order and Complexity’ metric. The remaining three metrics were difficult to unambiguously interpret and fully implement from the formulae given in Ngo’s papers. We do not think that the omission of these three invalidates our work, as the 11 implemented metrics cover a wide range of layout features.

A web page can be loaded, the extension executed via a menu option, and the values for these eleven metrics are then displayed to the user (as well as being stored in a local file). In implementing this extension, two different types of issues needed to be considered: system issues (relating to the architecture of the program: web site rendering, web standards, getting the information from a web page, etc.) and theory issues (relating to the process of adapting the Ngo measures for a web site, interpreting the formulae, etc.)

3.1 System Considerations

Despite the existence of browser ‘standards’, different browsers render the same page differently. In this project, the visual appearance of a page is important. If the same HTML document gives the same metric values, yet looks different in different browsers, then the extent to which these metric values truly represent the objective visual layout of the web page would be questioned. This problem being unsolvable, we confined ourselves to Mozilla Firefox as the only browser with which our implementation of these metrics would work, so as to at least control the variable rendering process over all the web pages we used in our experiment.

Firefox offers a well-supported framework for extension development: Firefox extensions are packaged enhancements that enable functionally not originally included. Firefox extensions are also easy to distribute: the files are packaged into an XPI file, which is simply a renamed archive. When a user downloads the file, Firefox will automatically install the extension, which can be used when the user restarts the browser. Firefox does not, however, fully comply with current rendering standards, so web sites could end up being rendered differently than their intention. For the purposes of this project, the intention of the web page designer does not matter: it is the actual visual appearance in the browser that is important.

The standards problem also extended to the use of HTML itself: very few web sites conform to the strict HTML standard (Beatty et al., 2008), which meant that the extension could not rely on web standards for conformity and meant that the extension had to take into account that the same type of content might be represented differently in different web sites.

3.2 Theory Considerations

Ngo defined his measures as a set of abstract mathematical formulae, the application of which he demonstrated in his papers using sketched embedded rectangles to represent screens and visual screen elements. The formulae refer to “screen components” or “visual objects” and although his definition of a

“component” is never made clear, his example interfaces imply that components are the common constructs of a desktop application’s interface: window panes and buttons. The few real examples in his paper are very simple screens comprising buttons, radio buttons and text entry fields.

We felt that this simple all-inclusive definition of a component was too narrow for the richness of web page visual elements, where much of the appearance of the page is the presentation of information, rather than the provision of interactive objects. Web sites consist of a large range of elements which can be broadly broken down into three categories: *text*, *images* and *controls*. *Text* and *images* are rectangular areas that contain text or a picture, while *controls* are elements similar to those used in Ngo’s examples: buttons, text-entry fields etc. These are typically represented in HTML in *form* elements.

We have categorised components according to their visual appearance (rather than their function), as befits an analysis of visual layout: links are considered text (rather than control) and images are always images (even if they respond to a mouse-click).

Many of the metrics refer to three types of screen space: the frame, the layout and the objects (Ngo and Byrne (2001)). The *object* definition can easily be carried over to mean the visual elements, and we interpreted *layout* and *frame* by analysing Ngo’s examples: the frame is the entire area of an interface, while the layout is the bounding box of all the visual elements.

3.3 The Implementation

The data used by the formula was collected by a DOM walkover using JavaScript. For each component, the following data was collected:

- Type: category of the component: text, image or control.
- X position: x co-ordinate of the top-left hand corner of the component.
- Y position: y co-ordinate of the top-left hand corner of the component.
- Width: the components’ width.
- Height: the components’ height.
- Area: area of the component.

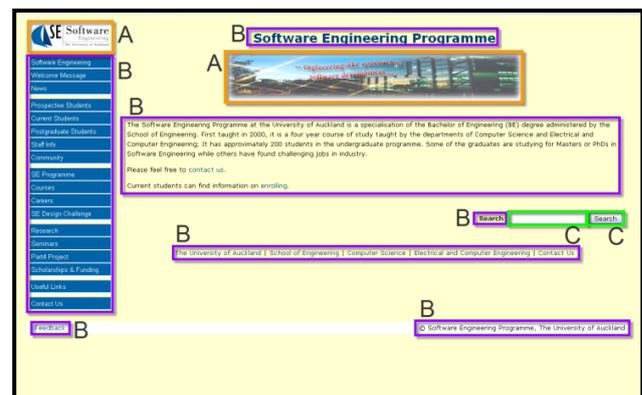


Fig 4: One of our experimental web pages (H), showing the visual components identified within the page as image (A), text (B) and control (C).

Figure 4 shows the components identified in one of our experimental web pages: A areas are identified as images, B as text, and C as controls.

The measures were taken directly from Ngo’s paper "Modelling interface aesthetics" (2003). The Firefox extension calculates 14 metric values for the page in its browser panel:

- Order and Complexity for all the components (the “overall” measure)
- Equilibrium, Density, Economy, Proportion, Cohesion, Balance, Sequence, Unity, Simplicity, Homogeneity for all components
- Order and Complexity for the set of the text components
- Order and Complexity for the set of the image components
- Order and Complexity for the set of the control components

Table 1 shows the 14 metric values for the web page shown in Fig 4.

	metric value (0-1)
Overall	0.3675
Equilibrium	0.9332
Density	0.2183
Economy	0.1963
Proportion	0.7394
Cohesion	0.6720
Balance	0.5171
Sequence	0.8999
Unity	0.1784
Simplicity	0.1993
Homogeneity	0.1218
Text	0.3603
Images	0.3874
Controls	0.2931

Table 1: The 14 metric values for web page in Fig 4.

3.4 Limitations and constraints

One of the main limitations of this approach is the inability to analyse components that exist within components (e.g. text wholly contained within an image, blank images included simply as space-fillers, large vertical spacing between paragraphs of text within one text component which give the visual appearance of two text components etc.) The original metrics do not consider this possibility, probably due to the simplicity of the examples to which were applied, where these cases did not occur. The JavaScript walkover of the DOM does not permit the visual content of any component to be identified (not even its colour). The image analysis option discussed above may have been able to allow for this more detailed analysis.

All the elements identified in the DOM analysis are considered to be rectangles, even though in the visual rendering of the web page, overlapping elements may

take the visual appearance of another shape. The metrics were defined to treat all elements as rectangles, so non-rectangular visual areas are inappropriate for the calculation of the metrics.

When viewing a web page, only the information in the browser panel can be seen at any one time, and scrolling is needed to reveal hidden parts of the page. It is reasonable, therefore, to consider (and measure) only that which can be seen at one time. For the purposes of this project, this meant confining our analysis to pages in which all components were visible on screen at one time. An obvious extension to our system would be the ability for the DOM to be only partially analysed, taking into account only those elements visible at one time.

4 Evaluation

4.1 Research questions

Using this implementation of these formulae for visual layout of web pages allowed us to address the following research questions in a user study:

- Q1. Do the metrics model users’ perception of ‘aesthetic appeal’?
- Q2. Do the metrics model users’ perception of ‘usability’?

While our two primary questions are focussed on the metrics, the data collected allowed us to investigate a complementary question:

- Q3. Is there a relationship between users’ perception of ‘aesthetic appeal’ and ‘usability’?

While the metrics do not in any way encode the use of colour, as colour is such a prominent feature of web pages, we conducted the same experiment twice, once in black and white, and once in colour. Thus all three questions above were investigated twice, and a further supplementary question could be addressed:

- Q4. Does colour affect users’ perceptions of ‘aesthetic appeal’ and ‘usability’?

4.2 The web pages

We chose a set of 15 different web pages for the study. As the metrics are designed for a single-screen layout, the content of all these web pages was all visible in one screen – none of them required scrolling. The variety of pages was chosen so as to cover a wide range of topics, and so that there were some that were text-heavy (e.g. the news item from the University of Auckland homepage (F)), image-heavy (e.g. the Borders home page (B)), control-heavy (e.g. the Seek homepage (K)) and a variety of all the components (e.g. Telecom (M), Gmail (O)).

Each web page was screen-grabbed once in black and white and again in colour, and printed in 1280 x 800 pixel resolution. All 15 web pages are listed and labelled A-O in the appendix, and are shown there or throughout this paper.

For each web page, we calculated the 14 metrics detailed in section 3.3 above and produced 14 ‘ideal’ rankings from 1-15 for the pages.

- overall ranking (Order and Complexity) over all components (1 ranking);
- a ranking for each of the ten individual metrics over all components (10 rankings);
- a ranking for each of text, image and control components (3 rankings).

For example, the ASB page (G, figure 2(a)) is ranked against the other 14 web pages as shown in table 2.

	rank (1-15)
Overall	7
Equilibrium	1
Density	12
Economy	15
Proportion	14
Cohesion	13
Balance	3
Sequence	15
Unity	3
Simplicity	15
Homogeneity	14
Text	7
Images	8
Controls	4

Table 2: The metric ranks for the ASB page (G).

In comparison with the other pages, therefore, the ABS page (G) is particularly good in equality, unity and balance, but not so good on simplicity, economy and sequence. It ranks middling with respect to the layout of its text and image components, and well with respect to the placement of its controls.

4.3 The participants

The 21 participants were friends, family and colleagues of the student experimenters. They spanned a varying level of education (from high school to tertiary education), the age range was from 18-80, there was approximate equal gender representation and a wide range of occupations was represented. No participants had any particular background in visual design.

4.4 Method

The 15 web pages were printed out on card, each in both black and white, and in colour.

First the black and white pictures were laid out on a table, and participants were asked to arrange them in a linear order from Most Aesthetically pleasing, to Least Aesthetically pleasing. No ties were permitted. After a random shuffle of the pictures, the participants were then asked to arrange them in linear order from Most Usable to Least Usable. Again, no ties were permitted.

These black and white images were removed from the participant, and the same two tasks were performed with the randomly arranged colour pictures of the web sites. Participants were not permitted to refer to their black and white rankings when ranking the colour pictures.

4.5 Data collection

For each participant, we collected four values for each web page:

- the rank given for its black and white aesthetic appeal (BWA)
- the rank given for its black and white perceived usability (BWU)
- the rank given for its colour aesthetic appeal (CA)
- the rank given for its colour perceived usability (CU)

4.6 Analysis

4.6.1 The appropriateness of the metric values

Here we consider the following questions:

- Q1. Do the metrics model users' perception of 'aesthetic appeal'?
- Q2. Do the metrics model users' perception of 'usability'?

We calculated the mean rank for each web page over all participants, with respect to the four measures of black and white aesthetics (BWA), black and white usability (BWU), colour aesthetics (CA), colour usability (CU) (Figs 5 and 6). In both figures, the pages are ordered from A to O along the x axis, with A being the web page ranked highest (rank 1) using the overall metric formula, and O being ranked lowest (rank 5). The bar chart representation (Fig. 5) allows us to see the extent to which the participants' ranking matched the overall metric ranking: an upward bottom-left to top-right trend would be expected. The line chart representation (Fig. 6) allows us to see which particular web pages stand out as having been ranked very differently from what the formulae would predict.

We performed bi-variate correlations between the mean rank values calculated over all participants and the ranks determined by the metrics, so as to investigate whether there was any relationship between them. Table 3 shows all the significant correlations found. There were no significant findings for Text, Controls, Equilibrium, Sequence, Unity and Homogeneity.

	BWA	CA	BWU	CU
Overall	0.71 (**)	0.66(**)	-	-
Images	0.72(**)	0.74(**)	0.52(*)	0.65(**)
Density	0.55(*)	-	-	-
Economy	-	-	0.52(*)	-
Proportion	0.58(*)	0.51(*)	-	0.55(*)
Cohesion	-	0.64(*)	0.53(*)	0.70(*)
Balance	0.62(*)	0.64(*)	-	-
Simplicity	-	-	0.55(*)	-

Table 3: Correlations between mean participant rankings and those determined by the objective metrics. () indicates significance at $p < 0.01$; (*) indicates significance at $p < 0.05$; - indicates no significance.**

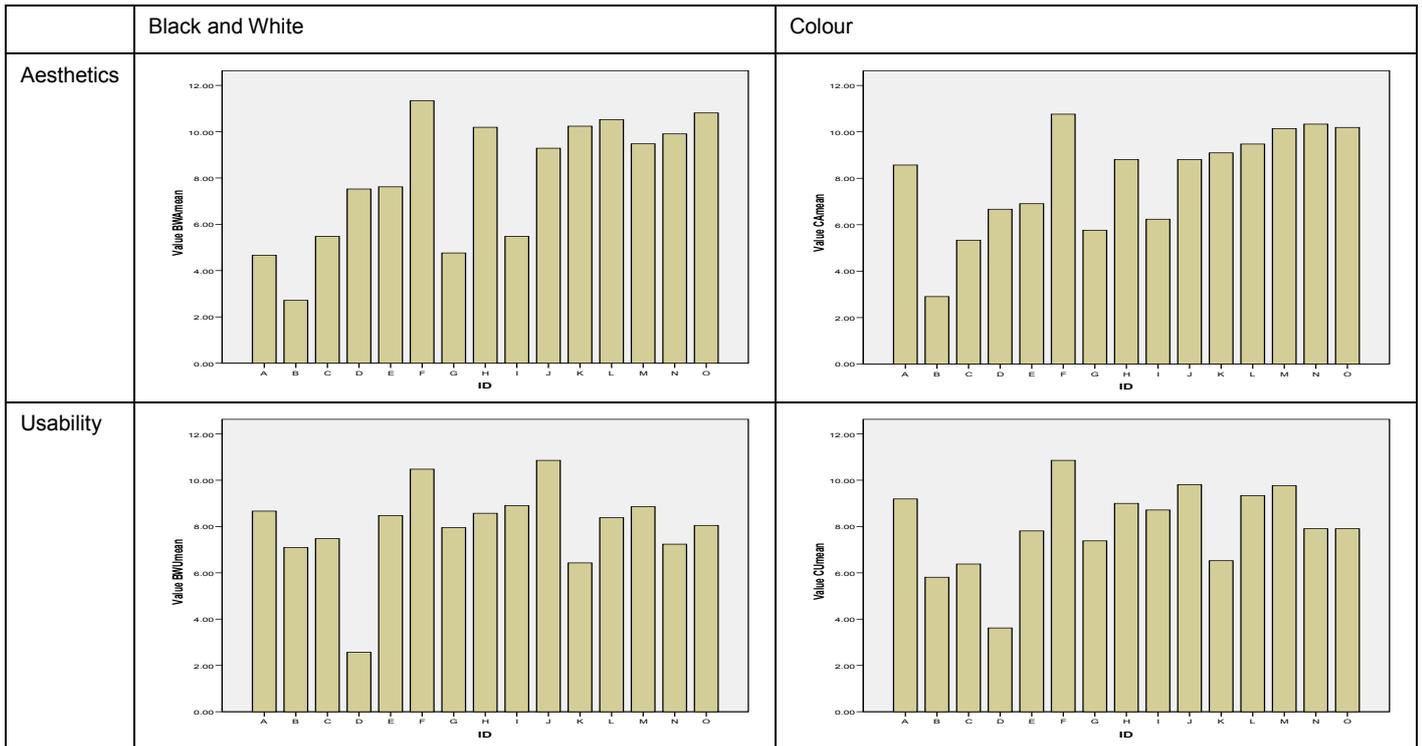


Fig 5: The mean ranks over all participants for all 15 web pages, according to perceived aesthetic appeal and usability, for black and white, and colour.

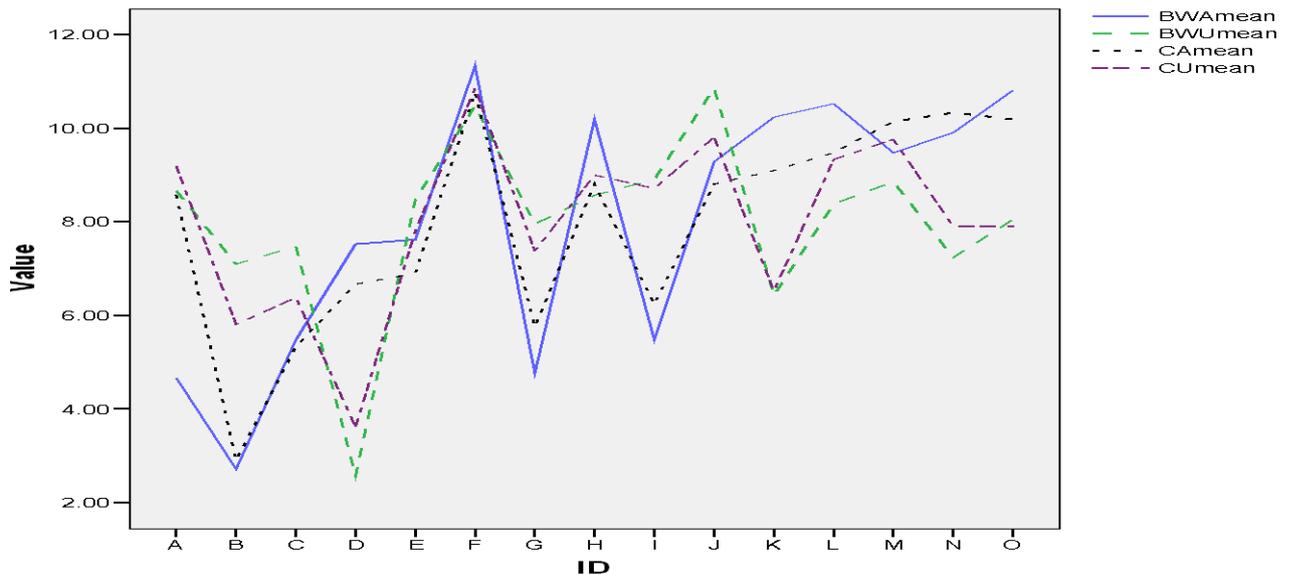


Fig 6: The mean values, over all participants, for each of the 15 web sites according to perceived aesthetic appeal and usability, for black and white, and colour.

4.6.2 Aesthetics and perceived usability

Here we consider the following question:

- Q3. Is there a relationship between users’ perception of ‘aesthetic appeal’ and ‘usability’?

We performed a bi-variate correlation analysis between the aesthetics and usability rankings, separately for black and white and colour. We found that only for colour was there a small but significant relationship: $r= 0.164$, $p=0.015$.

4.6.3 Colour

Here we consider the following question:

- Q4 Does colour affect users’ perceptions of ‘aesthetic appeal’ and ‘usability’?

We performed a bi-variate correlation analysis between the black and white and colour rankings, separately for aesthetics and usability. We found highly significant relationships for each: for aesthetics (0.880 , $p<0.001$) and usability (0.913 , $p<0.001$).

5 Discussion

5.1 Aesthetic Appeal

Aesthetic appeal is strongly captured by the overall metric value (Order and Complexity), for both black and white, and colour. Some single metrics capture aesthetic appeal on their own (Density, Proportion, Cohesion and Balance), with Proportion and Balance doing so for both black and white and colour.

Aesthetic appeal is very strongly captured by the placement of images on the screen, more so than when all components (including text and controls) are considered.

5.2 Perceived usability

The perception of usability is not captured by the overall metric, although when just the images are considered, there is a significant correlation. The cohesion metric on its own (over all components) gives a similar result to that of the images. This is surprising: it is not clear why perceived usability is affected by the fact that objects have similar aspect ratios. Simplicity, proportion and economy each individually capture perceived usability. Simplicity and economy only feature as significant results for black and white usability; we are surprised that these results did not carry over to colour usability.

5.3 Aesthetics, usability and colour

In our results aesthetic appeal does not match perceived usability; the only significant relationship was small. We are surprised at these findings, as they appear to contradict those of Kurosu and Kashimura (1995) and Tractinsky (1997), but speculate that this is because the web site images we used are much richer than their more simplistic interfaces.

Colour is not a dominant factor in judgement of either aesthetic appeal or perceived usability.

There was no significance for all text and all controls, indicating that these metrics, when applied only to text or controls, do not capture aesthetics or perceived usability.

5.4 Unusual individual results

Page F had a mean rank of approximately 11 for all four participant measures, even though its calculated overall rank is 6. This page is text heavy (see appendix 8.2). As we have discovered that the metrics capture aesthetic appeal and usability best with regard to the placement of images, this result does not come as a surprise. It may be that the equal weighting that we gave to text, images and control in the overall calculation is not appropriate.

The other page that stands out as giving obviously different ranking is A (Fig 3(b)), which is objectively ranked 1, yet had mean ranks of approximately 9 for BWU, CU and CA. This page has no obvious control elements, so it is unsurprising that the usability ranks are high, and its gray background appears to reduce its aesthetics appeal when it is presented in colour.

Both pages G (Fig. 2(a)) and I (Fig. 3(a)) produce lower mean rankings for both black and white and colour aesthetic appeal judgements than their objective ordering would predict: both of these pages are image heavy (while not conforming to constituent metrics like Balance, Simplicity and Unity).

5.5 Future work

Ngo himself queried whether the overall Order and Complexity metric should equally weight the constituent metrics (Ngo, Teo and Byrne, 2000): our results indicate that some metrics are better predictors than others, but we are not in a position with these results to suggest appropriate weightings. Additional empirical studies need to be performed. As part of this further study, implementing the remaining three metrics would be useful.

Our implementation is confined to web pages that fit on a single screen. A future implementation could provide metrics based on the current viewing window of a scrolling page. This would mean that the metric values would change as the user scrolls, and that a mechanism for integrating the values over all possible viewports would need to be derived.

In common with many visual aesthetics empirical studies, we asked our participants for their perception of the usability of the web pages. There is no guarantee that their perceived usability matches actual usability. Task-based experiments are required to determine whether these metrics can measure objectively the actual usability of a web site.

A more complex extension of this work would entail an automatic layout tool that proposes optimal positioning for a set of web page elements with respect to the metrics. An advanced end user tool could even allow the user to select their own aesthetic preferences, resulting in a personally optimised layout.

6 Conclusions

Aesthetic appeal and perceived usability are important because both encourage users' engagement in a site.

Our results show that these objectively calculated metrics can be useful in assessing the overall aesthetic appeal and perceived usability of a web page design even if only the images and an overall metric are used. Web designers can therefore make good use of these metrics as a design tool, allowing for interim quantitative feedback during the design stage.

The practical implications of this work are:

- Our tool gives quantitative and useful information to a web page designer on the aesthetic layout quality of a single-screen web page.
- This information could be used to ensure consistency and/or contrast in layout between web pages in a single web site.
- The tool could serve as a foundation for the 'critiquing tool' proposed by Ngo and Byrne (2001) whereby the program can make explicit suggestions about element placement during design.

Our work has shown that it is possible to quantify the aesthetic appeal of a web page. We suggest that there is a market for tools that embody such aesthetic metrics so as to encourage and guide useable and aesthetically appealing design.

7 Acknowledgements

We are grateful to the Department of Computer Science at the University of Auckland, which hosted a visit from the first author, and to all the experimental participants. Ethical approval for this experiment was given by the University of Auckland, 2009.

8 Appendices

8.1 The layout metrics

These brief definitions of Ngo's metrics are taken from Ngo (2001). The complete formulae can be found in Ngo and Byrne (2001).²

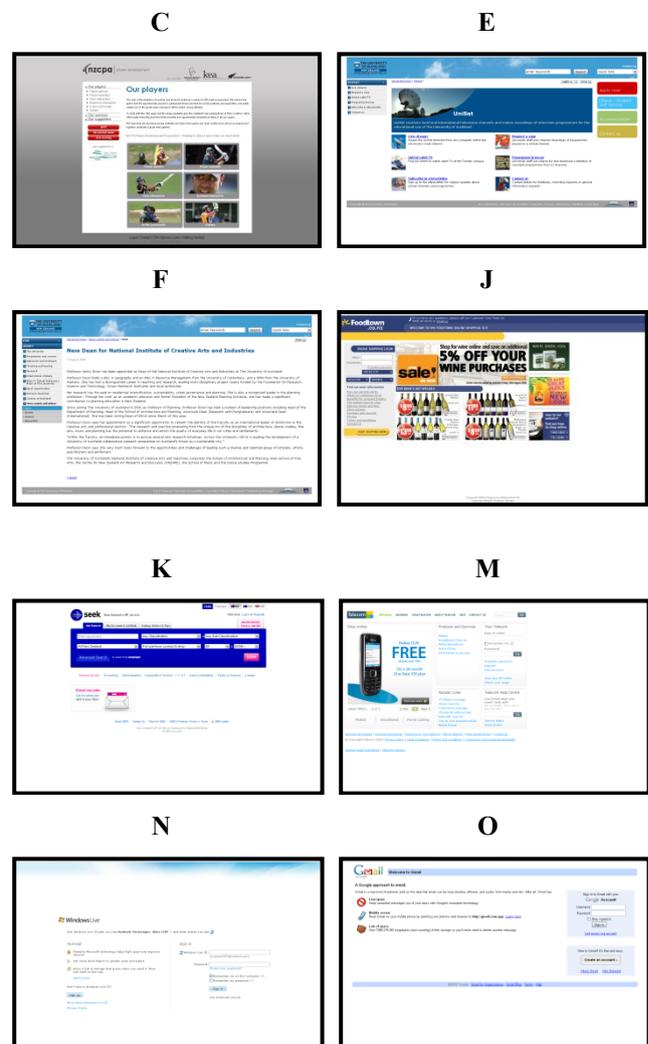
- Balance is computed as the difference between total weighting of components on each side of the horizontal and vertical axis.
- Equilibrium is computed as the difference between the centre of mass of the displayed components and the physical centre of the screen.
- Symmetry* is the extent to which the screen is symmetrical in three directions: vertical, horizontal and diagonal.
- Sequence is a measure of how information in a display is ordered in a hierarchy of perceptual prominence corresponding to the intended reading sequence.
- Cohesion is the extent to which the screen components have the same aspect ratio.
- Unity is the extent to which visual components on a single screen all belong together.
- Proportion is the comparative relationship of the dimensions of components to certain proportional shapes.
- Simplicity is the extent to which component parts are minimised and the relationships between the parts are simplified.
- Density is the extent to which the percentage of component areas on the entire screen is equal to the optimal level.
- Regularity* is the extent to which the alignment points are consistently spaced.
- Economy is the extent to which the components are similar in size.
- Homogeneity is a measure of how evenly the components are distributed among the quadrants.
- Rhythm* is the extent to which the components are systematically ordered.

8.2 Test Page stimuli

The test pages are listed below, in order of overall metric value, together with images of those sites not appearing elsewhere in this paper. All images were captured on 13/08/2009.

² * indicates those metrics not implemented for this particular project.

- A. <http://www.procricket.co.nz>
- B. <http://www.borders.co.nz>
- C. <http://www.procricket.co.nz/index.php?page=OurPlayers>
- D. <http://www.google.co.nz>
- E. <http://www.unisat.auckland.ac.nz/uoa/home/unisat-1>
- F. http://www.auckland.ac.nz/uoa/home/news/template/news_item.jsp?cid=197964
- G. <http://www.asb.co.nz>
- H. <http://www.se.auckland.ac.nz>
- I. <http://www.tematacheese.co.nz/>
- J. <http://www.foodtown.co.nz>
- K. <http://www.seek.co.nz>
- L. <http://web.ece.auckland.ac.nz/Jahia/pid/8>
- M. <http://www.hotmail.com/>
- N. <http://www.telecom.co.nz>
- O. <http://www.gmail.com>



9 References

Beatty P., Dick, S. and Miller J. (2008): Is html in a race to the bottom?: A large-scale survey and analysis of conformance to 6 w3c standards. *IEEE Internet Computing* **12(2)**: 76-80.

- Salimun, C., Purchase, H.C., Simmons, D.R. and Brewster, S. (2010): Preference ranking of screen layout principles. *Proceedings of the 24th BCS HCI Conference* (to appear).
- De Angeli, A., Sutcliffe, A. and Hartmann, J. (2006): Interaction, Usability and Aesthetics: What Influences Users' Preferences? *Proceedings of the Conference on Designing Interactive Systems*, University Park, Pennsylvania: ACM.
- Hartmann, J. and Sutcliffe, A. (2005): Aesthetic Judgment of Interactive Systems. *Proceedings of the BCS-HCI conference*, Edinburgh 2005.
- Hartmann, J. (2006): Assessing the Attractiveness of Interactive Systems. *CHI Doctoral Consortium, Montreal*.
- Hassenzahl, M. (2004): The interplay of beauty, goodness and usability in interactive products. *Human-Computer Interaction* **19(4)**: 319-349.
- Knight, J. and Pandir, M. (2004): An Experimental Aesthetics Approach to Evaluating Websites. *Proceedings of Aesthetic Approaches to Human-Computer Interaction Workshop*, Nordichi 2004.
- Kurosu, M. and Kashimura, K. (1995): Apparent usability vs inherent usability. *CHI'95 Conference Companion*: 292-293.
- Lavie, T. and Tractinsky, N. (2004): Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human Computer Studies* **60**: 269-298.
- Ngo, D., Samsudin, A. and Abdullah, R (2000): Aesthetic measures for assessing graphic screens. *Journal of Information Science and Engineering* **16(1)**: 97-116.
- Ngo, D. and Byrne, J. (2001): Application of an aesthetic evaluation model to data entry screens. *Computers in Human Behavior* **17(2)**: 149.
- Ngo, D. (2001): Measuring the aesthetic elements of screen designs. *Displays* **22(3)**: 73.
- Ngo, D., Teo, L. and Byrne, J. (2003): Modelling interface aesthetics. *Information Sciences* **152**: 25.
- Ngo, D., Teo, L. and Byrne, J. (2000): Formalising guidelines for the design of screen layouts. *Displays* **21(1)**: 3-15.
- Ngo, D. (2001): Measuring the aesthetic elements of screen designs. *Displays* **22(3)**: 73-78.
- Norman, D. (2002): *The Design of Everyday Things*. New York: Basic Books.
- Norman, D. (2004): Introduction to Special Section on "Beauty, Goodness, and Usability." *Human-Computer Interaction* **19(4)**: 311-318.
- Pandir, M. and Knight, J. (2006): Homepage aesthetics: The search for preference factors and the challenges of subjectivity. *Interacting with Computers* **18(6)**.
- Parizotto-Ribeiro, R. and Hammond, N. (2004): What is Aesthetics anyway? Investigating the user of the design principles. In *Proceedings of Aesthetic Approaches to Human-Computer Interaction Workshop*, Nordichi 2004.
- Tractinsky, N. (1997): Aesthetics and Apparent Usability: Empirically Assessing Cultural and Methodological Issues. *Proceedings of CHI*, Atlanta 1997.
- Tractinsky, N. (2004): Toward the Study of Aesthetics in Information Technology. *Proceedings of the 25th International Conference on Information Systems*, Washington DC, 2004.