# The Effect of Aesthetically Pleasing Composition on Visual Search Performance

**Carolyn Salimun[1], Helen C. Purchase[2], David R. Simmons[3], Stephen Brewster[4]**

[1,2,4]School of Computing Science,
University of Glasgow,
Glasgow, G12 8QQ, United Kingdom
{carolyn}{hcp}{stephen}@dcs.gla.ac.uk

[3]School of Psychology,
University of Glasgow,
Glasgow G12 8QB, United Kingdom
david.simmons@gla.ac.uk

## ABSTRACT

This paper presents the results of a study on the effect of the aesthetic layout properties of a computer interface on visual search performance. Search performance was measured at three levels of layout aesthetics: high, medium, and low. Two types of performance metric were recorded: response time and number of errors. Performance at the three levels of aesthetics was also compared between two search methods (with or without mouse pointing), and related to preference. The findings of the present study indicate that, regardless of search method used, response time (but not errors) was strongly affected by the aesthetics level. There is also a clear relationship between preference and performance when a composite measurement of aesthetics is used, although this does not seem to be due to the influence of individual aesthetic features. Further study is needed to identify other aesthetic factors that influence task performance, and to establish appropriate design guidelines.

## Author Keywords

Aesthetics, aesthetics measures, interface layout, task performance.

## ACM Classification Keywords

H5.2. User Interfaces: Screen design (e.g. text, graphics, color)

## INTRODUCTION

There is increasing evidence to support the role of visual aesthetics in interface design, since the remarkable discovery by Kurosu & Kashimura [11] of the strong correlation between interface aesthetics and perceived usability e.g. [8-9, 15, 17, 21]. Overall, this evidence suggests that interfaces with highly rated aesthetics (e.g. interfaces which are “pleasant” to look at) are preferred over relatively unaesthetic interfaces, and that these same high aesthetic interfaces are perceived as easier to use [17, 21] and might solve some usability problems[15]. While this evidence provides strong support for the argument that aesthetics is important in interface design, it is certainly not conclusive, as most of this evidence has focused on the effect of interface aesthetics on the “look and feel” of the application and not on the effect of interface aesthetics on actual task performance.

Although there has been some limited work investigating aesthetics and performance (e.g.[22]) this past work is restricted to investigating the effect of aesthetics on reducing task error. Other types of performance, such as efficiency, have not been investigated. A recent paper by Sonderegger and Sauer [18] investigated the effect of aesthetics on performance, and found better performance with aesthetically appealing design. However, their focus was on the product design and visual appearance of a mobile phone and not on more generic interface design of a typical software application.

Addressing these issues is important, given that, in general, the literature in HCI has largely neglected aesthetics (at least before the study by [11]) due to the belief that aesthetic interfaces might adversely affect usability [20]. Thus, empirical evidence on this issue will be a useful guide not only for determining what, how, and to what extent the aesthetics of interfaces influences task performance, but also to see whether aesthetics really matters when it comes to task performance or if it is just *icing on the cake* (something extra and not essential that is added to an already good situation or experience and that makes it even better [1]). Relying on subjective judgments to judge how interface aesthetics might affect errors in, and efficiency of, task performance [6, 16], is not convincing enough to support the notion that *attractive things work better* [15].

Cox [7] claims that mouse pointing is likely to aid interactive search. Hornof [10] reported that the layout design of the interface influences mouse movements. Is performance in visual search task influenced more by mouse movement than by the design of interface? This is an important relationship to investigate because the design of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Nordi*CHI, 2010*, October 16-20, 2010, Reykjavik, Iceland.
Copyright 2010 ACM ISBN: 978-1-60558-934-3...$5.00.

the interface will affect mouse movement, which in turn will affect the process of visual search. If the mouse movements are complex, then performance in the visual search will be reduced. If, when using a mouse to aid the visual search, the performance using a high aesthetic layout proves to be better than that with low aesthetic layout, this means that performance is more influenced by design than the use of a mouse.

To design an interface that is both visually attractive and optimizes performance is the main challenge for designers. Some designers might neglect aesthetic elements of the user interface due to the fear that it might degrade usability, and some might overload the interface with interface elements that make performance more difficult. It is obvious that creating a beautiful, effective and efficient interface is not an easy task. However, some existing guidelines on user interface design (e.g. guideline for textual displays, graphical display, ordering of data and content, color, navigation flow, and composition of the screen) might help this process.

In this study, we are focusing on *classical aesthetics/clarity* rather than *expressive aesthetics/richness*[12] and the aesthetics properties that we are interested in are concerned with the layout of the interface. We are concerned about the form and position of interface objects relative to other objects and their placement within a frame. The aesthetic layout of the interfaces were measured by using six mathematical formulae proposed by Ngo, *et al* [14]. Ngo *et al* developed fourteen mathematical formulae based on Gestalt principles to measure the layout aesthetics of graphic composition: *balance, equilibrium, symmetry, sequence, cohesion, unity, proportion, simplicity, density, regularity, economy, homogeneity, rhythm,* and *order and complexity*. The validity of these formulae has been tested by comparing the results obtained from the computed aesthetic value and subjective measures, in which it was found that there were high correlations between computed aesthetic value and the aesthetics ratings of human viewers[14]. The aesthetics of the interface was categorized into high, medium, and low with the range of 1 (best) and 0 (worst).

This study intended to investigate the effect of interface aesthetics on actual task performance rather than perceived usability. Two types of performance were recorded: response time, and the number of errors. The pattern of performance was also compared in terms of search method (with or without mouse pointing), and related to preference rankings.

## METHODOLOGY

Six formulae adapted from Ngo *et al*[14] were selected as a basis to measure the aesthetic level of the interface layout. These six formulae were selected from Ngo *et al*'s fourteen original formulae based on our analysis of his diagrams of each aesthetic, which revealed that most of the variability in interface layout could be captured by using just six of the formulae. They are (taken from [13]):

- *Cohesion:* Cohesion by definition is the extent to which screen components have the same aspect ratio.
- *Economy:* Economy is the extent to which the components are similar in size.
- *Regularity:* Regularity is the extent to which the alignment points are consistently spaced.
- *Sequence:* Sequence, by definition, is a measure of how information in a display is ordered in a hierarchy of perceptual prominence corresponding to the intended reading sequence.
- *Symmetry:* Symmetry, by definition, is the extent to which the screen is symmetrical in three dimensions: vertical, horizontal, and diagonal.
- *Unity:* Unity, by definition, is the extent to which visual components on a single screen all belong together.

More details on the six formulae are in the appendix.

The following hypotheses were tested in the experiments:

$H_1$: Response times in visual search tasks will increase with decreasing aesthetics level.

$H_2$: The number of errors in visual search tasks will increase with decreasing aesthetics level.

$H_3$: The use of mouse pointing in visual search tasks will produce a longer search time than without, but with the same dependence of search time on aesthetics level.

$H_4$: The use of mouse pointing in visual search tasks will produce fewer errors than without, but with the same dependence of error number on aesthetics level.

### Participant

Twenty two (11 male and 11 female) undergraduate and postgraduate students of University of Glasgow from a variety of backgrounds (e.g. Computer Science, Accountancy & Finance, Accounting and Statistics, Economics, Business and Management etc) participated in the experiment. All the participants were computer literate and used computers daily. The participants received no remuneration for their participation.

### Stimuli

The stimuli for this experiment were created using a JAVA program. The program randomly generated 90 stimuli, calculated the aesthetics value for each stimulus based on the average value of all the six aesthetics measures (*Cohesion, Economy, Regularity, Sequence, Symmetry,* and *Unity*), and categorized them as either high, medium, or low aesthetic level (table 1). The range of the aesthetics level is between 0 (worst) and 1 (best). Figure 1illustrates examples of layouts with different aesthetic values (high, medium, and low).

In the informal pre-pilot tests, participants described the stimuli with high aesthetic level as “orderly” or “tidy” and low aesthetics level as “disorderly” or “messy”.

Each stimulus consisted of 8 – 10 mixed inverted and upright triangles. The number of triangles in each layout was set to a maximum of 10. In the informal pre-pilot test we found that due to the high number of stimuli (180) the participants found the task too tiring when the numbers of triangles were more than 10. It is important to reduce participants’ fatigue effects as these could be confounded with low aesthetics.

The number of upright triangles on each stimulus ranged from 4 – 6. The triangles were drawn using black lines on a white background (figure 3) and were 5 - 25 mm in height and 50 - 25 mm in width. Since the main focus of this experiment was on the layout aesthetics, the colors were limited to black (color of the triangle line) and white (background) to avoid or minimize the effects of confounding factors.

The stimuli were presented to the participants using a custom-written JAVA program (the counting task), and on A4 paper (the preference task).

In the preference task, the stimuli were printed on two sheets of A4 paper. The first sheet showed three layouts, and the second sheet showed six layouts. The selection of the three layouts on the first sheet was based on the computed aesthetics value which categorized the layout as either high, medium, or low aesthetics (figure 1), while the selection of the six layouts (figure 2) on the second sheet was based on our subjective judgments that the placement of the interface objects on the interface were representative of high levels of specific aesthetics measures.

| Low | Medium | High |
|---|---|---|
| 0.0 < 0.5 | 0.5 ≤ 0.7 | 0.7 < 1.0 |

**Table 1. Aesthetics value range for each level of aesthetics**

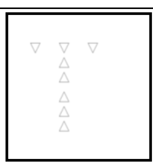
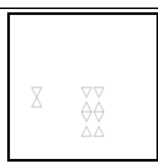
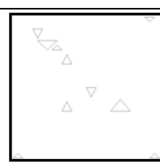
| | | |
|---|---|---|
| <br>(a) | <br>(b) | <br>(c) |
| Cohesion: 0.7778<br>Economy: 1.0<br>Regularity: 0.6116<br>Sequence: 0.75<br>Symmetry: 0.3067<br>Unity: 0.8665<br>Average : 0.7188 | Cohesion: 0.6897<br>Economy: 1.0<br>Regularity: 0. 7139<br>Sequence: 0.0<br>Symmetry: 0. 2386<br>Unity : 0. 9290<br>Average : 0. 5952 | Cohesion: 0. 8563<br>Economy: 0.25<br>Regularity: 0. 2972<br>Sequence : 0.5<br>Symmetry: 0. 6874<br>Unity: 0. 35<br>Average : 0. 4902 |

**Figure 1. Examples of *High* (a), *Medium* (b), and *Low*(c) aesthetics**

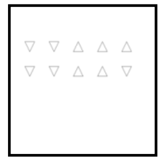
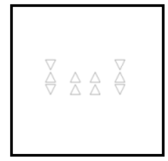
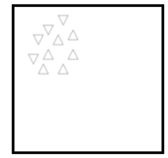
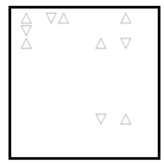
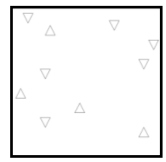
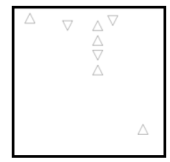
| | | |
|---|---|---|
| <br>(a) | <br>(b) | <br>(c) |
| Cohesion : 0. 3182<br>Economy : 1.0<br>**Regularity**: 0. 7194<br>Sequence : 1.0<br>Symmetry: 0. 2914<br>Unity : 0. 9238<br>Average : 0. 7088 | Cohesion : 0. 4375<br>Economy : 1.0<br>Regularity: 0. 6889<br>Sequence : 1.0<br>**Symmetry**: 0. 8514<br>Unity : 0. 9477<br>Average : 0. 8209 | Cohesion : 0. 8333<br>Economy : 1.0<br>Regularity: 0. 5139<br>Sequence : 1.0<br>Symmetry: 0. 5<br>**Unity** : 0. 9432<br>Average : 0. 798 |
| <br>(d) | <br>(e) | <br>(f) |
| Cohesion : 1.0<br>Economy : 1.0<br>Regularity: 0. 5333<br>**Sequence** : 0.75<br>Symmetry: 0. 3128<br>Unity : 0. 7364<br>Average : 0. 7221 | Cohesion : 0. 8793<br>**Economy** : 1.0<br>Regularity: 0. 2444<br>Sequence : 1.0<br>Symmetry: 0. 7148<br>Unity : 0. 5912<br>Average : 0. 7221 | **Cohesion** : 1.0<br>Economy : 1.0<br>Regularity: 0. 308<br>Sequence : 0.5<br>Symmetry: 0. 2814<br>Unity : 0. 6695<br>Average : 0. 6265 |

**Figure 2. Examples of interface with *Regularity*(a), *Symmetry*(b), Unity(c), *Sequence*(d), *Economy*(e), and *Cohesion*(f)**

**Figure 3. A screen shot of the experimental system, showing the response buttons in the top right-hand corner**

### Task

There were two tasks in the experiment:

a) Counting task – the participants were required to count the number of upright triangles rather than inverted triangles.
b) Preference task – the participants were required to order selected layouts based on *the most preferred* to *the least preferred* layouts.

### Procedure

At the beginning of the experimental session, the participants received written and verbal instructions, signed a consent form, and filled in a demographic questionnaire. Participants were then seated in front of a laptop (Screen size of 30 cm with resolution of 1024 x 768 pixels, viewed from approximately 50 cm).

A computer program, written in JAVA was used to present the stimuli, accept answers and measure response times. Before starting the experiment, participants were given a practice task. The purpose of the practice task was to ensure that participants were familiar and comfortable with the task before starting the experiment proper. The stimuli used in the practice task were randomly chosen from the 90 stimuli used in main experiment. The participants performed the practice tasks until they said they were ready to start the main experiment. Most participants did fewer than 10 practice trials.

The 90 different interfaces, which varied in aesthetic level (high, medium or low), were presented in different random orders in the practice task, in the experiment proper, and for each participant, to counter learning and ordering effects. There were three answer buttons labeled 4, 5, and 6. As the stimuli were generated randomly, there were 21 stimuli with the answer of 4, 38 stimuli with the answer of 5 and 31 stimuli with the answer of 6. Each screen was untimed. The next screen was automatically shown once the participants clicked on the answer buttons. This process continued until all 90 layouts were shown. The task took approximately 20 minutes.

There were two conditions in the experiment.

Condition 1: With mouse pointing
The participants were allowed to use a mouse to point to the triangles (but there was no effect of clicking). The use of mouse pointing can guide eye movements throughout the visual search task.

Condition 2: Without mouse pointing
The participants were not allowed to use the mouse to point to the triangles. They were only allowed to use the mouse to click on the answer button.

All participants were required to perform the task in both conditions. Participants were randomly assigned to perform either condition 1 first or condition 2. After finishing the first task (condition 1 or condition 2), the participants were given an opportunity to take a short break before continuing to perform the next task (condition 1 or condition 2, depending on which condition was completed first). The participants were given practice in each condition before the real experiment. Data from the practice task were not included in the analysis. Since the same stimuli were used in both conditions, the possibility that the participants would remember the answers while performing the task in the second condition might exist. However, this possibility was minimized by the randomized sequence of the stimuli in the two conditions.

After finishing both tasks in condition 1 and condition 2, the participants were shown two sheets of A4 paper. One page showed three stimuli with extreme aesthetic layouts; the other showed examples of high values for each of the six different aesthetics parameters. These preference stimuli were selected from the 90 stimuli used in the performance tasks. The participants were required to rank the layouts based on the most preferred to the least preferred.

### Variables

There are four variables of interest of this experiment.

a) Independent variables:
   - Aesthetics level (low, medium, high)
   - Visual search method (with or without mouse pointing)

b) Dependent variables:
   - Response Time
   - Errors

Previous literature [3-4, 7, 10] suggested that there was a significant effect of search method (with or without mouse pointing) used in a visual search task where the use of mouse pointing lengthens the response time and reduces the number of errors compared to without mouse pointing. It is important to investigate if there is a similarity of performance pattern on each level of aesthetics for both search methods. The similarity indicates that performance is affected by aesthetics level, while dissimilarity indicates that performance is not affected by aesthetics level but the search method.

## RESULT

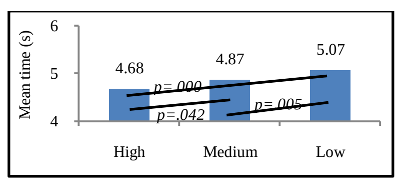
An ANOVA - General Linear Model with repeated measure analysis followed by multiple comparison tests with Bonferroni correction was used to analyze the data from the experiment. The assumption of Sphericity was tested and found to be valid for the overall performance data, but not valid for preference data associated with the stimuli used in the preference test. In this case, the data were corrected using the Greenhouse-Geisser correction.

### Result for aesthetics layout

*The effect of aesthetics level: time*

$H_1$: Response times in visual search tasks will increase with decreasing aesthetics level.

There was a significant main effect of aesthetics level for response time $F_{2,\ 42}$ =16.294, $p < 0.05$. The pairwise comparisons showed that all possible pairs were significantly different at $p < 0.05$, where response times for the high aesthetics level were significantly lower than those at medium and low levels (figure 4).




***The lines indicate where the significance was formed**

**Figure 4. The mean response time for each aesthetics level**

*The effect of aesthetics level: error*

$H_2$: The number of errors in visual search task will increase with decreasing aesthetics level.

There was no significant main effect of aesthetics level for errors $F_{2, 42} = 3.040$ $p > 0.05$ (figure 5).

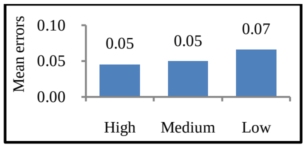



**Figure 5. The mean errors for each aesthetics level**

*The effect of answer value: time*

There was a significant main effect of answer value (the number of upright triangles) for response time $F_{(2, 42)} = 26.259$, $p < 0.05$. Response time for answer values of four was significantly higher compared to interfaces with answers of five and six. There was no significant difference between response times for interfaces with answer values between six and five (figure 6).

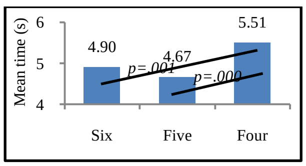



**Figure 6. The mean response time based on answer value**

Further analysis showed that there was a significant main effect of the aesthetics level for response time for all interfaces based on answer value:

a) Six answer value - $F_{2, 42} = 18.165$ $p=.000$
b) Five answer value - $F_{2, 42} = 4.422$ $p=.018$
c) Four answer value - $F_{2, 42} = 5.369$ $p=.008$

Overall, except for interfaces with answer values of five, response time was shorter for interfaces with high aesthetics and longer with low aesthetics (table 2).

| Answer Value | Aesthetics Level | Medium | Low |
|---|---|---|---|
| 6 | High : 4.40 | *p = 0.00* | *p = 0.00* |
| | Medium : 5.12 | | *p > 0.05* |
| | Low : 5.18 | | |
| 5 | High : 4.82 | *p = 0.024* | *p > 0.05* |
| | Medium : 4.57 | | *p > 0.05* |
| | Low : 4.61 | | |
| 4 | High : 5.22 | *p = 0.05* | *p = 0.022* |
| | Medium : 5.65 | | *p > 0.05* |
| | Low : 5.65 | | |

**Table 2. The mean response time based on answer value**

*The effect of answer value: error*

There was a significant main effect of the answer value for the number of errors $F_{(2, 42)} = 37.163$, $p < 0.05$. All possible pairs of the three answer value were also found to be significant (figure 7).

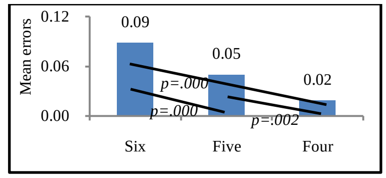



**Figure 7.The mean errors based on answer value**

Further analysis however showed that, the main effect of aesthetics level was significant for interfaces with answer values of six ($F_{(2, 42)} = 5.580$, $p < 0.05$), and not significant for interfaces with answer values of four and five (table 3).

| Answer Value | Aesthetics Level | Medium | Low |
|---|---|---|---|
| 6 | High : 0.05 | *p = 0.022* | *p = 0.005* |
| | Medium : 0.10 | | *p > 0.05* |
| | Low : 0.11 | | |
| 5 | High : 0.05 | *p > 0.05* | *p > 0.05* |
| | Medium : 0.04 | | *p > 0.05* |
| | Low : 0.06 | | |
| 4 | High : 0.01 | *p > 0.05* | *p > 0.05* |
| | Medium : 0.02 | | *p > 0.05* |
| | Low : 0.02 | | |

**Table 3. The mean errors based on answer value**

## Preference

The preference tasks were limited to the particular stimuli.

*Preference of extreme aesthetics level*

The Friedman analysis (figure 8) of the three interfaces (see figure 1), which were the subset of the 90 stimuli, showed that, preference of these interfaces were strongly affected by the aesthetics level ($\chi^2 = 26.273$, $df = 2$, $p = .000$). The higher the aesthetics level the more preferred the interface.

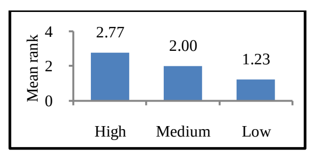



**Figure 8. Preference ranking of three stimuli**

*Performance of extreme aesthetics level: time*

Taking the performance measures for just the three stimuli (figure 1) used in the preference test, we found a significant difference in response time for the three stimuli $F_{2, 42} = 20.437$, $p = .000$ (figure 10). Pairwise comparisons showed that responses times for figure 1(a) (high aesthetics) were significantly different from those for figure 1(c) (low aesthetics), and that response times for figure 1(b) (medium aesthetics) were significantly different from those for figure 1(c). Response times for figures 1(a) and figure 1(b) were found not to be significantly different.

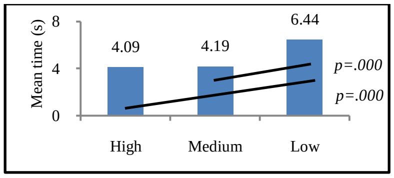



**Figure 9. Mean response time for three stimuli**

*Performance of extreme aesthetics level: errors*

Taking the performance measures for just the three stimuli (figure 1) used in the preference test, there was a significant difference in the number of errors for the three stimuli $F_{2, 42} = 10.059$, $p = .002$ (figure 10). Pairwise comparisons showed that errors for figure 1(a) (high aesthetics) were significantly different from those for figure 1(c) (low aesthetics) and errors for figure 1(b) (medium aesthetics) were significantly different from those for figure 1(c). Errors for figure 1(a) and figure 1(b) were found not to be significantly different.

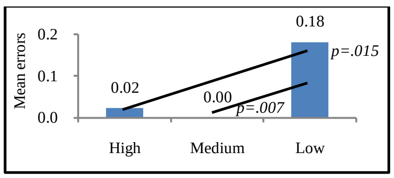



**Figure 10. Mean errors for three stimuli**

*Preference of six aesthetics measures*

The Friedman analysis (figure 11) of the six interfaces which were the subset of the 90 stimuli, showed that, figure 2(b) ranked as the most preferred, followed by figure 2(a), figure 2(c), figure 2(d), figure 2(f), and figure 2(e) ($\chi^2 = 57.974$, $df = 5$, $p = .000$).

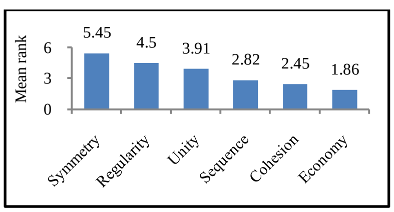



**Figure 11. Preference ranking of six stimuli**

*Performance of six aesthetics measures: time*

Taking the performance measures for just the six stimuli (figure 2) used in the preference test, there was no significant difference in response time of the six stimuli $F_{5, 105} = 1.482$, $p = .241$ (figure 12).

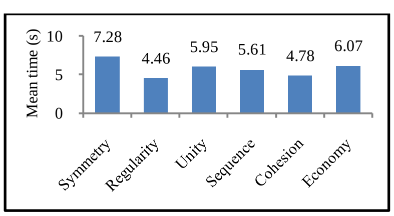



**Figure 12. Mean response time for six stimuli**

*Performance of six aesthetics measures: error*

Taking the performance measures for just the six stimuli (figure 2) used in the preference test, there was no significant difference in the number of errors of the six stimuli $F_{5, 105} = 2.739$, $p = .054$ (figure 13).

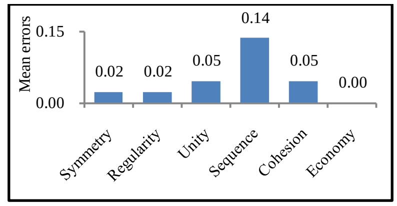



**Figure 13. Mean errors for six stimuli**

In the post-experiment informal interview in which we asked the participants why they ranked certain layout higher than the others, the participants used words and phrases like "orderly" or "tidy" or "neat", or "symmetrical" to describe the layouts which they preferred the most, and "disorderly" or "messy" or "too spread away" to describe the interfaces which they least preferred.

### Result of search method

*The effect of mouse pointing: time*

There was a significant main effect of search method for response time $F_{1, 21} = 5.663$, $p < 0.05$ (figure 14).

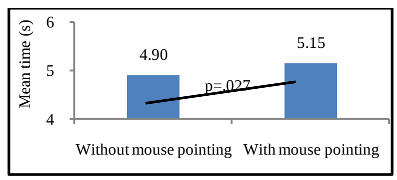



**Figure 14. The mean response time for search methods**

$H_3$: The use of mouse pointing in visual search tasks will produce a longer search time than without, but with the same dependence of search time on aesthetics level.

There was a significant main effect of aesthetics level for response time *without mouse pointing* $F_{1, 42} = 5.302$, $p = .009$ and *with mouse pointing* $F_{2, 42} = 8.184$, $p = .001$. Pairwise comparisons *without mouse pointing* showed that only the pair high and low aesthetics level was significantly different. Other pairs were found not to be significantly different. In *with mouse pointing* all possible pairs were significantly different except for the pair with medium and low aesthetics levels (table 4).

| | Aesthetics Level | Medium | Low |
|---|---|---|---|
| Without mouse pointing | High : 4.73 | *p > 0.05* | *p = 0.014* |
| | Medium : 4.89 | | *p > 0.05* |
| | Low : 5.09 | | |
| With mouse pointing | High : 4.90 | *p = 0.007* | *p = 0.003* |
| | Medium : 5.34 | | *p > 0.05* |
| | Low : 5.21 | | |

**Table 4. The mean response time with and without mouse pointing**

*The effect of without and with mouse pointing: errors*

There was no significant main effect of search method for errors $F_{1, 21} = .178$, $p = .677$ (figure 15).

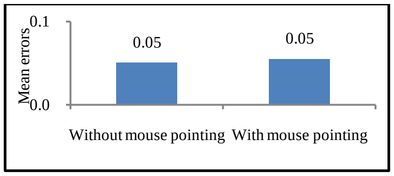



**Figure 15. The mean errors with and without mouse pointing**

$H_4$: The use of mouse pointing in visual search tasks will produce fewer errors than without, but with the same dependence of error number on aesthetics level.

There was no significant main effect of aesthetics level for errors *without mouse pointing* $F_{2, 42} = 2.245$, $p = .118$ and *with mouse pointing* $F_{2, 42} = 2.348$, $p = .108$ (table 5).

| | **Aesthetics Level** | Medium | Low |
|---|---|---|---|
| Without mouse pointing | High : 0.04 | $p > 0.05$ | $p > 0.05$ |
| | Medium : 0.05 | | $p > 0.05$ |
| | Low : 0.07 | | |
| With mouse pointing | High : 0.04 | $p > 0.05$ | $p > 0.05$ |
| | Medium : 0.06 | | $p > 0.05$ |
| | Low : 0.06 | | |

**Table 5. The mean errors of with and without mouse pointing**

## ANALYSIS

### Aesthetics and performance

*Time*

The significant main effect of aesthetics level for average response time indicated that participants' performance was strongly affected by the aesthetics level of the interface.

Further analysis showed significant differences between response times at each aesthetics level: the higher the aesthetics level the less time taken to complete the tasks, and the lower the aesthetics level, the more time spent to complete the task.

This result might be explained by referring to cognitive theory. Szabo & Kanuka [19] argue that good screen design leads to automatic control processing (the ability to do things without occupying the mind with the low-level details required, allowing it to become an automatic response pattern or habit [2]), thus less time is needed to complete the task. On the other hand, poor screen design leads to manual processing, thus more time is spent on completing the task. This result was also in accordance with Hornof [10] who noted that response time in visual search task depends on the organization or structure of the interface rather than just the number of targets.

*Error*

The lack of a significant dependence of error number on aesthetics level appears to be due to floor effects (figure 5). The total number of triangles seems to be so small that the participants were able to quickly identify the number of upright triangles, so they hardly made any errors. The overall error rate was 0.05.

This claim is made following our examination of the average number of errors on interfaces with four, five, and six upright triangles (regardless of aesthetics level), where the participants made more errors as the number of triangles to be counted increased (table5). We also note that the only significant dependence found for errors was with those for stimuli with six upright triangles (the higher the aesthetics level, the fewer errors occurred). There was no significant dependence of error on aesthetics level for stimuli with five and four upright triangles.

Thus, we believe the aesthetics level of the interface will become increasingly important as the task becomes more complicated (e.g. there are larger numbers of elements in the display and also in the target).

### Preferences

The result of the two preference tasks showed that, an interface was preferred when it looked symmetrical and orderly, and least preferred when it looked unsymmetrical and "messy".

The result of the first preference task, where we used three stimuli which were categorized as high, medium, and low aesthetics according to their computed aesthetics value, confirmed the robustness of the mathematical formulae proposed by Ngo, *et al* in measuring aesthetics of interface. Layouts with high aesthetic levels are most preferred followed by layouts with medium aesthetics levels, and then the low levels. This result was as expected based on previous studies reporting that aesthetics interfaces are more preferred than unaesthetic ones [5, 14]. These preferences are reflected in the task performance with these three stimuli, where we found that interfaces which ranked as the *most preferred* had the best performance with respect to response time (but not errors) compared to the *least preferred* interface.

The result of the second preference task, where we used six stimuli which were representative of high levels of specific aesthetics measures, showed a strong preference ranking of the six aesthetics measures. These preferences however are not reflected in the performance with these selected stimuli. Interfaces which ranked as the *most preferred* had worse response time performance compared to the *least preferred* interface.

Based on the results of these two preference tasks, we found a clear relationship between preference and performance for extreme examples of high, medium, and low aesthetics, but did not find a relationship between preference and performance for examples of high values of the individual aesthetics measures. We conclude that preference only relates to performance when a 'composite' measurement of aesthetics is used, rather than an individual measurement.

### Interface aesthetics and search methods

The lack of a significant effect of aesthetics level on the average number of errors for different search methods indicates that participants did not find any advantage of using mouse pointing. This outcome was not expected. We speculated that, in accordance with previous literature, participants would make fewer errors when they used mouse pointing compared to when they were solely reliant on eye movements to navigate the layout. Previous literature [3-4, 7, 10] indicates that mouse pointing significantly aids search by enabling the user to use the cursor to "mark" an object, while the eyes move elsewhere scanning for necessary information required for the tasks. The tagged object acts as a reference point and reduces the possibility of miscounts or recounts of previously identified objects, thus less errors are made. This robust finding however was not replicated in this study. There are two possible explanations for this finding.

The first explanation is that performance in visual search tasks is not affected by the search method. The second explanation is that the complexity of the layout was not high enough to produce an advantage of mouse pointing over unaided eye-movements. Previous studies ([7, 10]) suggested that mouse pointing significantly aids visual search when there are large numbers of distracters competing with the target objects. In this experiment there were only ten objects in total, including distracters (inverted triangles) and target objects (upright triangles). The minimum number of distracters was four and the maximum was six. Thus, the small number of distracters on each layout could be the answer why mouse pointing provides no advantage in this study.

The significant main effect of aesthetics level on response time, irrespective of search method, indicates that aesthetics level was the main determinant of response time. Further analysis revealed that mouse pointing produced a longer response time than without mouse pointing at each level of aesthetics. This result was in line with previous literature which indicates that response times with mouse pointing method were longer than with unaided eye-movements. Cox & Silva [7] stated that objects which are distracters in visual search are treated differently in mouse pointing and eye-movement strategies. When a user uses mouse pointing, all objects on display including distracters are treated as potential targets for action, thus it takes a longer time to complete the tasks, whereas without mouse pointing distracters are not treated as objects which require action, thus less time is spent on the task. Tipper, as cited in [7], explained that mouse use involves visuo motor processing and a consequent increase in processing time.

The most important aspect of the findings reported here, however, lies not in the finding that, overall, 1) performance in term of the number of errors was the same for both search methods, or 2) the use of mouse pointing lengthened task completion times compared to without the use of mouse pointing, since these findings have been frequently discussed in the literature. It is more important to look at the performance dependence with both search methods on aesthetics level. It was clearly shown that, regardless of search method used, performance in term of response time was better with high aesthetic interfaces than with low aesthetic interfaces. This is an interesting result not previously reported in the literature in HCI.

## CONCLUSION

The aim of this empirical study was to investigate the effect of aesthetic layout properties of a computer interface on actual task performance and to investigate any relationship between performance and preference. User performance was compared based on three levels of aesthetics of screen layout (high, medium, low) as specified by a previously published algorithm [14]. Two types of performance were recorded, response time and the number of errors.

The findings of the present study indicate that, regardless of search method used, user performance in term of response time in a visual search task was strongly affected by the aesthetics level of the interface, where the time performance improved as the level of aesthetics increased and the performance deteriorated as the aesthetics level decreased. The effect of aesthetics level on the number of errors however was not evident. We also found a clear relationship between preference and performance when a composite measurement of aesthetics was used. These performance data analyses however were limited to the particular stimuli that were used in the preference tests.

We have used triangles as our stimuli so as to remove any possible confounding factors (for example, content, shape, color etc.), while recognizing the limitations of doing so: typical interfaces contain richer and more varied objects. However, this controlled study that has focused on the layout of objects still provides us with useful information because the results obtained give information on the *layout* (the visual pattern) of the objects (based on the layout principles), not on their *content*. Further experiments may confirm whether the layout of content-rich objects produce similar results.

Further research is needed to identify other aesthetics factors that might influence performance, and to establish appropriate design guidelines that can assist task performance. The results suggest that designers should aim for highly structured layouts (e.g. consistent spacing between interface objects' alignment points both horizontally and vertically) and avoid unstructured layouts (e.g. inconsistent spacing and large spaces between interface objects). Put simply, a messy desktop not only looks bad, but also reduces the user's ability to complete tasks efficiently.

## REFERENCES

1 "Oxford Advanced Learner's Dictionary." Oxford University Press, 2010.

2 *Wikipedia*. http://en.wikipedia.org/wiki/Automaticity.

3 Arroyo, E., Selker, T., and Wei, W. Usability tool for analysis of web designs using mouse tracks. *Ext. Abstracts CHI 2006*, ACM (2006), 484-489.

4 Arroyo, E., Selker, T., and Wei, W. Exploring how mouse movements relate to eye movements on web search results pages. In *Proc. 30th Annual International ACM SIGIR Conference* (2007), 29-32.

5 Bauerly, M., and Liu, Y. Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *International Journal of Human-Computer Studies 64*,8 (2006), 670-682.

6 Chang, D., Dooley, L., and Tuovinen, J. E. Gestalt theory in visual screen design: a new look at an old subject. In *Proc. The Seventh world conference on computers in education conference on computers in education: Australian topics*,Australian Computer Society, Inc. (2002), 5-12.

7 Cox, A. L., and Silva, M. M. The role of mouse movements in interactive search. In *Proc. The 28th Annual Meeting of the Cognitive Science Society*(2006).

8 De-Angeli, A., Sutcliffe, A., and Hartmann, J. Interaction, usability and aesthetics: what influences users' preferences? In *Proc. Proceedings of the 6th conference on Designing Interactive systems*,ACM (2006), 271-280.

9 Hassenzahl, M. The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction 19*,4 (2004), 319 — 349.

10 Hornof, A. J. Visual search and mouse-pointing in labeled versus unlabeled two-dimensional visual hierarchies. *ACM Trans. Comput.-Hum. Interact. 8*,3 (2001), 171-197.

11 Kurosu, M., and Kashimura, K. Apparent usability vs. inherent usability: experimental analysis on the determinants of the apparent usability. In *Proc. Conference companion on Human factors in computing systems*,ACM (1995), 292-293.

12 Lavie, T., and Tractinsky, N. Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies 60*,3 (2004), 269-298.

13 Ngo, D. C. L. Measuring the aesthetic elements of screen designs. *Displays 22*,3 (2001), 73-78.

14 Ngo, D. C. L., Teo, L. S., and Byrne, J. G. Modelling interface aesthetics. *Information Sciences 152*(2003), 25-46.

15 Norman, D. Emotion & design: attractive things work better. *Interactions 9*,4 (2002), 36-42.

16 Reilly, S. S., and Roach, J. W. Improved visual design for graphics display. *Computer Graphics and Applications, IEEE 4*,2 (1984), 42-51.

17 Schenkman, B. N., and Jonsson, F. U. Aesthetics and preferences of web pages. *Behaviour & Information Technology 19*,5 (2000), 367-377.

18 Sonderegger, A., and Sauer, J. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Applied Ergonomics 41*,3 (2010), 403-410.

19 Szabo, M., and Kanuka, H. Effects of violating screen design principles of balance, unity, and focus on recall learning, study time, and completion rates. *Journal of Educational Multimedia and Hypermedia 8*,1 (1999), 23-42.

20 Tractinsky, N. Aesthetics and apparent usability: empirically assessing cultural and methodological issues. In *Proc. The SIGCHI conference on Human factors in computing systems*,ACM (1997), 115-122.

21 van der Heijden, H. Factors influencing the usage of websites: the case of a generic portal in The Netherlands. *Information & Management 40*,6 (2003), 541-549.

22 van Schaik, P., and Ling, J. Modelling user experience with web sites: Usability,hedonic value, beauty and goodness. *Interacting with Computers 20*(2008), 419–4.


**APPENDIX: Aesthetic layout formulae taken from [14]**

***Cohesion***

$$CM = \frac{|CM_{fl}| + |CM_{lo}|}{2} \in [0,1]$$

$$CM_{fl} = \begin{cases} c_{fl} & \text{if } c_{fl} \le 1 \\ \frac{1}{c_{fl}} & \text{otherwise} \end{cases}$$

with

$$c = \frac{h_{layout}/b_{layout}}{h_{frame}/b_{frame}}$$

where $b_{layout}$ and $h_{layout}$ and $b_{frame}$ and $h_{frame}$ are the widths and heights of the layout and the frame, respectively. $CM_{lo}$ is a relative measure of the ratios of the objects and layout with

$$CM_{lo} = \frac{\sum_i^n t_i}{n}$$

with

$$t_i = \begin{cases} c_i & \text{if } c_i \le 1 \\ \frac{1}{c_i} & \text{otherwise} \end{cases}$$

with

$$c = \frac{h_i/b_i}{h_{layout}/b_{layout}}$$

where $b_i$ and $h_i$ the width and height of object *i* and *n* is the number of objects on the frame.

***Economy***

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

where $n_{size}$ is the number of sizes.

***Regularity***

$$\mathrm{RM} = \frac{|\mathrm{RM_{alignment}}| + |\mathrm{RM_{spacing}}|}{2} \in [0,1]$$

$\mathrm{RM_{alignment}}$ is the extent to which the alignment points are minimized with

$$\mathrm{RM_{alignment}} = \begin{cases} 1 & \text{if } n = 1 \\ 1 - \dfrac{n_{vap} + n_{hap}}{2n} & \text{otherwise} \end{cases}$$

and $\mathrm{RM_{spacing}}$ is the extent to which the alignment points are consistently spaced with

$$\mathrm{RM_{spacing}} = \begin{cases} 1 & \text{if } n = 1 \\ 1 - \dfrac{n_{spacing} - 1}{2(n-1)} & \text{otherwise} \end{cases}$$

where $n_{vap}$ and $n_{hap}$ are the numbers of vertical and horizontal alignment points, $n_{spacing}$ is the number of distinct distances between column and row starting points and $n$ is the number of objects on the frame.

***Sequence***

$$\mathrm{SQM} = 1 - \frac{\sum_{j=\mathrm{UL,UR,LL,LR}} |q_j - v_j|}{8} \in [0,1]$$

with

$$\{q_{\mathrm{UL}}, q_{\mathrm{UR}}, q_{\mathrm{LL}}, q_{\mathrm{LR}}\} = \{4,3,2,1\}$$

$$v_j = \begin{cases} 4 & \text{if } w_j \text{ is the biggest in } w \\ 3 & \text{if } w_j \text{ is the 2nd biggest in } w \\ 2 & \text{if } w_j \text{ is the 3rd biggest in } w \\ 1 & \text{if } w_j \text{ is the smallest in } w \end{cases} \quad j = \mathrm{UL, UR, LL, LR}$$

with

$$w_j = \sum_i^{n_j} a_{ij} \quad j \equiv \mathrm{UL,UR,LL,IR}$$

$$w = \{w_{\mathrm{UL}}, w_{\mathrm{UR}}, w_{\mathrm{LL}}, w_{\mathrm{LR}}\}$$

where UL, UR, LL, and LR stand for upper-left, upper-right, lower-left, and lower-right, respectively and $a_{ij}$ is the area of object $i$ on quadrant $j$. Each quadrant is given a weighting in $q$.

***Symmetry***

$$\mathrm{SYM} = \frac{|\mathrm{SYM_{vertical}}| + |\mathrm{SYM_{horizontal}}| + |\mathrm{SYM_{radial}}|}{3} \in [0,1]$$

$$\mathrm{SYM_{vertical}} = \frac{\begin{array}{l} |X'_{\mathrm{UL}} - X'_{\mathrm{UR}}| + |X'_{\mathrm{LL}} - X'_{\mathrm{LR}}| + |Y'_{\mathrm{UL}} - Y'_{\mathrm{UR}}| + |Y'_{\mathrm{LL}} - Y'_{\mathrm{LR}}| + \\ |H'_{\mathrm{UL}} - H'_{\mathrm{UR}}| + |H'_{\mathrm{LL}} - H'_{\mathrm{LR}}| + |B'_{\mathrm{UL}} - B_{\mathrm{UR}}| + |B'_{\mathrm{LL}} - B'_{\mathrm{LR}}| + \\ |\Theta'_{\mathrm{UL}} - \Theta'_{\mathrm{UR}}| + |\Theta'_{\mathrm{LL}} - \Theta'_{\mathrm{LR}}| + |\Theta'_{\mathrm{UL}} - \Theta'_{\mathrm{UR}}| + |\Theta'_{\mathrm{LL}} - \Theta'_{\mathrm{LR}}| + \end{array}}{12}$$

$$\mathrm{SYM_{horizontal}} = \frac{\begin{array}{l} |X'_{\mathrm{UL}} - X'_{\mathrm{LL}}| + |X'_{\mathrm{UR}} - X'_{\mathrm{LR}}| + |Y'_{\mathrm{UL}} - Y'_{\mathrm{LL}}| + |Y'_{\mathrm{UR}} - Y'_{\mathrm{LR}}| + \\ |H'_{\mathrm{UL}} - H'_{\mathrm{LL}}| + |H'_{\mathrm{UR}} - H'_{\mathrm{LR}}| + |B'_{\mathrm{UL}} - B_{\mathrm{LL}}| + |B'_{\mathrm{UR}} - B'_{\mathrm{LR}}| + \\ |\Theta'_{\mathrm{UL}} - \Theta'_{\mathrm{LL}}| + |\Theta'_{\mathrm{UR}} - \Theta'_{\mathrm{LR}}| + |\Theta'_{\mathrm{UL}} - \Theta'_{\mathrm{LL}}| + |\Theta'_{\mathrm{UR}} - \Theta'_{\mathrm{LR}}| + \end{array}}{12}$$

$$\mathrm{SYM_{radial}} = \frac{\begin{array}{l} |X'_{\mathrm{UL}} - X'_{\mathrm{LR}}| + |X'_{\mathrm{UR}} - X'_{\mathrm{LL}}| + |Y'_{\mathrm{UL}} - Y'_{\mathrm{LR}}| + |Y'_{\mathrm{UR}} - Y'_{\mathrm{LL}}| + \\ |H'_{\mathrm{UL}} - H'_{\mathrm{LR}}| + |H'_{\mathrm{UR}} - H'_{\mathrm{LL}}| + |B'_{\mathrm{UL}} - B_{\mathrm{LR}}| + |B'_{\mathrm{UR}} - B'_{\mathrm{LL}}| + \\ |\Theta'_{\mathrm{UL}} - \Theta'_{\mathrm{LR}}| + |\Theta'_{\mathrm{UR}} - \Theta'_{\mathrm{LL}}| + |\Theta'_{\mathrm{UL}} - \Theta'_{\mathrm{LR}}| + |\Theta'_{\mathrm{UR}} - \Theta'_{\mathrm{LL}}| + \end{array}}{12}$$

$$\mathrm{SYM_{radial}} = \frac{\begin{array}{l} |X'_{\mathrm{UL}} - X'_{\mathrm{LR}}| + |X'_{\mathrm{UR}} - X'_{\mathrm{LL}}| + |Y'_{\mathrm{UL}} - Y'_{\mathrm{LR}}| + |Y'_{\mathrm{UR}} - Y'_{\mathrm{LL}}| + \\ |H'_{\mathrm{UL}} - H'_{\mathrm{LR}}| + |H'_{\mathrm{UR}} - H'_{\mathrm{LL}}| + |B'_{\mathrm{UL}} - B'_{\mathrm{LR}}| + |B'_{\mathrm{UR}} - B'_{\mathrm{LL}}| + \\ |\Theta'_{\mathrm{UL}} - \Theta'_{\mathrm{LR}}| + |\Theta'_{\mathrm{UR}} - \Theta'_{\mathrm{LL}}| + |R'_{\mathrm{UL}} - R'_{\mathrm{LR}}| + |R'_{\mathrm{UR}} - R'_{\mathrm{LL}}| \end{array}}{12}$$

$X', Y', H', B', \Theta'$, and $R'$ are respectively the normalized values of

$$X_j = \sum_i^{n_j} |x_{ij} - x_c| \quad j = \mathrm{UL, UR, LL, LR}$$

$$Y_j = \sum_i^{n_j} |y_{ij} - y_c| \quad j = \mathrm{UL, UR, LL, LR}$$

$$H_j = \sum_i^{n_j} h_{ij} \quad j = \mathrm{UL, UR, LL, LR}$$

$$B_j = \sum_i^{n_j} b_{ij} \quad j = \mathrm{UL, UR, LL, LR}$$

$$\Theta_j = \sum_i^{n_j} \left| \frac{y_{ij} - y_c}{x_{ij} - x_c} \right| \quad j = \mathrm{UL,UR,LL,IR}$$

$$R_j = \sum_i^{n_j} \sqrt{(x_{ij} - x_c)^2 + (y_{ij} - y_c)^2} \quad j = \mathrm{UL, UR, LL, LR}$$

where UL, UR, LL and LR stand for upper-left, upper-right, lower-left and lower-right, respectively $(x_{ij}, y_{ij})$ and $(x_c, y_c)$ are the co-ordinates of the centres of object $i$ on quadrant $j$ and the frame; $b_{ij}$ and $h_{ij}$ are the width and height of the object and $n_j$ is the total number of objects on the quadrant

***Unity***

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

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

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

and $\mathrm{UM_{space}}$ is a relative measurement, which means that the space left at the margins (the margin area of the screen) is related to the space between elements of the screen (the between-component area) with

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

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.