

1997

Layout complexity: does it measure usability?

Tim Comber
Southern Cross University

John R. Maltby
Southern Cross University

Publication details

Comber, T & Maltby, JR 1997, 'Layout complexity: does it measure usability?', in S Howard, J Hammond & G Lindgaard (eds), *Human-computer interaction: Interact '97, International Conference on Human-computer Interaction*, Sydney, NSW , 14-18 July, Chapman Hall, London, pp. 623-626.

ePublications@SCU is an electronic repository administered by Southern Cross University Library. Its goal is to capture and preserve the intellectual output of Southern Cross University authors and researchers, and to increase visibility and impact through open access to researchers around the world. For further information please contact epubs@scu.edu.au.

Layout Complexity: Does It Measure Usability?

Tim Comber,
John R. Maltby
Southern Cross University
PO Box 157
Lismore, NSW 2480
Australia.
tcomber@scu.edu.au
jmaltby@scu.edu.au

ABSTRACT This research investigates the validity of the layout complexity metric to GUI screen design. This metric offers a simple method to gauge the complexity of the visual design of a computer screen by classifying screen objects into classes based on common dimensions and positions. However, though it has been used by other interface researchers, it is not clear that it is a valid metric for GUIs. Initial work suggests that users prefer more complex layouts rather than the simple ones suggested by guidelines. The pilot study showed that screens midway between complex and simple were the most usable. The next experiment is at the data collection stage. It uses four screens differing in complexity and a Latin-squares design to avoid experimental bias and nuisance variables. Twenty-four human participants are randomly assigned to blocks. The final experiment will attempt to gain a wider range of users with a more complex task.

KEY WORDS Layout complexity, GUI, interface, usability.

1. RESEARCH THESIS

Despite the importance of screen displays to computer users (Shneiderman 1992), there are few empirical studies relating to modern, bit-mapped screens, (Galitz 1993) even though clearly most new computer systems use some form of GUI. This research examines the use of one *static measure* of layout quality; layout complexity and compares it to other measures eg white space, balance, and symmetry, to assess the usability of screen designs.

Static measures are defined as those metrics that do not require any knowledge of the task. They are independent of the functioning of the application. They

can be contrasted to *dynamic measures* such as Layout Appropriateness (Sears 1993) which require the designer to be able to identify the functions of screen objects. The chief attraction of static measures is that they can be incorporated into the design environment, eg see DON (Kim and Foley 1993), which automates part of the generation of a particular GUI and then lets an evaluation module compute the metrics (balance, symmetry, and the associated index). However, before layout complexity can be used in such a way it needs to be shown that it does measure usability. Earlier research has shown that it does affect users' attitudes to alphanumeric screens (Tullis 1988) but GUI's are more complex and visually oriented.

The layout complexity metric (LC) provides a measure of the horizontal and vertical alignment of objects and their positional alignment (Bonsiepe 1968). The layout complexity formula states that the complexity of a system C is given by

$$C = -N \sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

where:

$$N = \sum_{i=1}^n n_i$$

and

$$p_i = \frac{n_i}{n}$$

where:

- N = total number of objects (widths or heights, distance from top or side of page)
- n = number of classes (number of unique widths, heights or distances)
- n_i = number of objects in the i th class
- p_i = proportion of the i th class.

2. HYPOTHESIS:

Layout Complexity: There is a trade-off between usability (U) and complexity C with a relationship of the form $U = fC$ where U is a maximum for some intermediate value of C (Figure 1 and Figure 2).

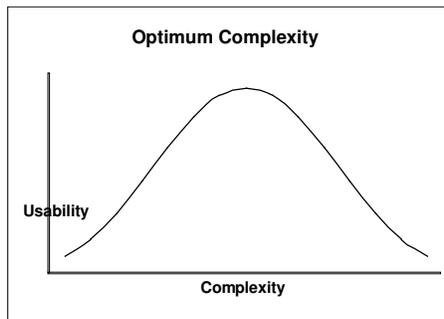


Figure 1: Relationship between complexity and usability

As the complexity figure becomes smaller, it becomes more difficult to distinguish different interface objects and the interface takes on an artificial regularity. On the

other hand, the interface becomes more predictable. At the other extreme as the interface approaches maximum complexity, it looks artificially irregular. What is more important, it becomes impossible for the designer to group objects with similar functions on the basis of size or position. However, the increase in complexity does mean that the user has more information and therefore more choice of operations

3. METHODOLOGY

My hypothesis implies a causal relationship between layout complexity and usability of computer interfaces. Therefore an experimental design is the most appropriate methodology. The advantages of the experimental method are well known (Simon 1969), (Robson 1973). A Latin-square design will be used for the main experiment in order to detect learning effects and control for individual differences.

In the first instance a survey was used to determine whether complexity theory could be applied to GUI design and if indeed it measured some aspect of design “quality” (Comber and Maltby 1994). This was followed by a pilot experiment (Comber and Maltby 1995) with layout complexity as the independent variable and usability as the dependent variable. An important characteristic of usability testing is that it should be carried out “by real users carrying out real tasks in a real technical, physical and organisational environment” (Bevan and McLeod 1994) p136. Therefore it was important to build an application that had real utility, one that could be perceived by the participants as being a real application rather than a contrived experimental one.

4. FINDINGS TO DATE

Initial work (Comber and Maltby 1994) focused on the use of the layout complexity metric (see (Bonsiepe 1968)(Tullis 1983; Tullis 1988; Tullis 1988)) applied to the opening screens of thirteen Windows 3.1 applications. It was found that the most complex screen was 66 times greater than that of the least complex screen. Screen dumps were made of these applications and presented to thirty people to determine whether Bonsiepe’s technique would provide a predictive

measure for users' ranking of different designs (Comber and Maltby 1994).

The subjects were asked to sort the screen prints from best design to worst design, with no ties. No attempt was made to define what was meant by "goodness" of design, this interpretation being left up to the individual. The expectation based upon the work of Tullis and Bonsiepe was that good layout design strives to be simple. This was not borne out by the results. A number of applications, including Microsoft Word and Excel, received rankings opposite to that expected. This suggests that users prefer more complex layouts. Having shown that layout complexity is both measurable for GUI's and that at least one aspect of usability, attitude, is affected by the metric, the next stage was to determine the metric's utility by building an application and measuring the effect of layout complexity on usability.

A pilot experiment (Comber and Maltby 1995) was designed to test three components of usability; effectiveness, learnability, and attitude. The pilot consisted of a simple application, running under Microsoft Windows, that calculated layout complexity for each design iteration. The screen with the lowest complexity score consisted of objects arranged in a neat grid with almost uniform sizes. The next two screens consisted of almost normal layouts and the final screen had every object with a different size and position. Seven experienced computer users volunteered to take part in the pilot study. It was found that the least and most complex screens were the least usable as expected (Table 1).

Usability	Screen 1	Screen 2	Screen 3	Screen 4
Complex	156	170	186	228
Error-free	36%	79%	86%	71%
Time	354	290	276	293
Rating	4	7	16	0

Table 1. Summary of results.

Table 1 summarises the results. The screens with a mid-range complexity, screens 2 and 3, rate better overall than the screens at either end of the complexity scale. However these results do need to be treated cautiously

because of the small number of subjects and the limited number of screens.

The results from this pilot showed differences in usability between screens differing in complexity. However, the application required further development to improve its data collection capabilities and to make it more automated.

5. PLANS FOR FURTHER WORK

The next stage involves a controlled experiment with a Latin-square design. With four different layouts, this will require 24 people. The application has been written to allow the experiment to proceed without the presence of the experimenter making it more cost-effective to administer. The application now collects details of layout appropriateness (Sears 1993) and white space as well as layout complexity. At the time of writing there remains the coding for the symmetry and balance metrics (Kim and Foley 1993).

A separate application saves the dimensions and positions of each control (widget) in the interface to a file to enable the reconstruction of the screen layouts by the software. This application also calculates the static metrics.

A simple Microsoft Windows application, *Launcher*, originally designed as an example application for a Visual Basic tutorial, forms the basis for the experiment. The application was originally designed to allow some simple file management routines for Windows 3.1; viewing files, adding file names to a list and renaming them with any string, deleting files etc. Subjects were asked to complete a simple task with each screen.

The software records data for each screen:

- time between each click,
- distance between clicks,
- errors made,
- requests for help (pressing F1).

The program presents an ethics disclaimer and allows the person to exit the experiment at any time in accordance with university ethics policy. Each person has the same task to perform and can practice the task before attempting the experiment. The program asks for the age group, sex, occupation, years spent computing, computer experience, familiarity with Microsoft Windows and the amount of time spent using a

computer per week. Each of the four screens is then presented in random order and the user is expected to complete the task on each screen. After each screen there is the opportunity to judge their opinion of the screen. At the conclusion the participants can review their choices and make comments.

Four different screen layouts were designed, each with a different complexity score. The screen with the lowest score consists of objects arranged in a neat grid with almost uniform sizes. The next two screens consist of almost normal layouts and the final screen has every object with a different size and position. Table 2 shows the complexity ratings for each of the four screens used in the experiment. The theoretical minimum was not achievable in VB, when using different objects, as some objects could not be resized to match other objects ie objects in VB have a fixed size relationship to other objects.

Complexity Scores for 16 Objects						
	Theory min.	Scr. 1	Scr. 2	Scr. 3	Scr. 4	Theory max.
C	64	133	156	178	212	248
%	0%	42%	49%	57%	78%	100%

Table 2: Complexity scores for 16 objects

The experiment is now in the data collection stage. The final experiment will attempt to gain a wider range of users with a more complex task.

6. REFERENCES

Bevan, N. and M. McLeod (1994). "Usability measurement in context." Behaviour & Information Technology **13**(1 & 2): 132-145.

Bonsiepe, G. A. (1968). "A method of quantifying order in typographic design." Journal of Typographic Research **2**: 203-220.

Comber, T. and J. R. Maltby (1994). A Formal Method for Evaluating GUI Screens. ACIS'94 - Doctoral Consortium, Melbourne, Australia, Dept. of Information Services, Monash.

Comber, T. and J. R. Maltby (1994). Screen complexity and user design preference in windows applications. OZCHI'94, Melbourne, Australia, CHISIG.

Comber, T. and J. R. Maltby (1995). Evaluating usability of screen designs with layout complexity. OZCHI'95, Wollongong, CHISIG.

Galitz, W. O. (1993). User-Interface Screen Design. Inc., P. O. Box 812070, Wellesley, MA 02181, Q. E. D. Information Sciences.

Kim, W. C. and J. D. Foley (1993). Providing high-level control and expert assistance in the user interface presentation design. Proceedings of the INTERCHI '93. S. Ashlund. Amsterdam, Netherlands, ACM: 430-437.

Robson, C. (1973). Experiment, Design and Statistics. Harmondsworth, Middlesex, Penguin.

Sears, A. (1993). "Layout appropriateness: A metric for evaluating user interface widget layout." IEEE Transactions on Software Engineering **19**(7): 707-718.

Shneiderman, B. (1992). Designing the User Interface: Strategies for Effective Human-Computer Interaction. Reading, MA, Addison-Wesley.

Simon, J. L. (1969). Basic Research Methods in Social Science. New York, Random House.

Tullis, T. S. (1983). "The formatting of alphanumeric displays: a review and analysis." Human Factors **25**(6): 557-582.

Tullis, T. S. (1988). Screen design. Handbook of Human-Computer Interaction. M. Helander, Elsevier Science Publishers B.V. (North-Holland).

Tullis, T. S. (1988). "A system for evaluating screen formats: Research and application." Advances in Human-Computer Interaction: 214-286.