Metric-based evaluation of graphical user interfaces: Model, method, and software support

Conference Paper · June 2013

DOI: 10.1145/2480296.2480331

1 author:

Mathieu Zen
Université catholique de Louvain
5 PUBLICATIONS **3** CITATIONS

SEE PROFILE

Metric-Based Evaluation of Graphical User Interfaces: Model, Method, and Software Support

Mathieu Zen
Université catholique de Louvain, Louvain School of Management,
Place des Doyens, 1 B-1348 Louvain-la-Neuve (Belgium) - mathieu.zen@uclouvain.be

ABSTRACT
Many factors contribute to ensuring User eXperience (UX) of Graphical User Interfaces, such as, but not limited to: usability, fun, engagement, subjective satisfaction. Aesthetics is a potential element that could also significantly contribute to this user experience. Although aesthetics have been extensively discussed, there is a need to rely on a sound, empirically validated methodology in order to properly evaluate how aesthetics could be measured, namely through metrics. Two main issues need to be addressed: the representativeness and the relevance of aesthetics metrics. In order to address these challenges, this paper introduces a methodology for metric-based evaluation of a graphical user interface of any type. This methodology is based on an underlying model that captures aesthetics aspects and related metrics, a method for computing them based on the underlying model, and software that supports enacting this method on any type of graphical user interface.

Author Keywords
Metrics, visual techniques, aesthetics, ergonomics.

ACM Classification Keywords
D.2.2 [**Software Engineering**]: Design tools and techniques – *User Interfaces*; H.5.2 [***Information Interfaces and Presentation***]: User Interfaces – *Graphical User Interfaces*.

General Terms
Algorithms; Design; Human Factors

INTRODUCTION
In the field of Human-Computer Interaction (HCI), a significant body of knowledge exists that addresses the quality of use of a Graphical User Interface (GUI) in general and its usability in particular [6]. Empirical studies are a major factor that could be used to assess the usability of a GUI. In particular, aesthetics plays an important role that could significantly affect the social acceptance of a GUI by end users [6, 13,16]. Although a complete thorough understanding of the role played by aesthetics in HCI remains to be studied, several references are useful to characterize its facets and to assess whether a GUI could be considered as aesthetic or not [16]. These methods are mostly manual and require a precise and rigorous application in order to guarantee its appropriate enactment, thus preventing inexperienced designers and developers from applying these methods. One straightforward way to assess the aesthetics of a GUI consists in computing metrics addressing visual aspects that are typically considered in aesthetics [3,14,15,17].

This paper discusses the main aspects of a work aimed at automating the GUI evaluation from an aesthetics perspective. A review of key papers that approach concepts related to GUI aesthetics is firstly delivered, including the use of visual techniques and the definition of aesthetics metrics. A methodology for conducting a metric-based evaluation of GUI is provided, based on the assumption that metrics could be considered as relevant estimators of aesthetics. This methodology consists of three pillars: a model that captures aspects relevant to a metric-based evaluation of a GUI, a structured method in order to compute these metrics based on the aforementioned model, and a software tool that provides designers and developers with some guidance on how to apply, enact this method on any GUI.

BACKGROUND AND RELATED WORK
Many methods for evaluating a GUI are used today, such as usability guides [4,12] and various software [5,8,11,17] are exploited to automate this process. With respect to metric-based evaluation, GUILayout++ [8] enables the designer to draw a GUI and to evaluate its layout based on BaLOReS [5], a suite of principles and metrics such as unused space, contents of interests, advertising space, website identity elements, and advertisement elements. Results are displayed quantitatively with percentage figures and qualitatively by a "smiley". These metrics, although properly defined in previous work, require substantive validation of their relevance and significance based on user studies [5].

Key references
Two main references were considered as initiators of our method: a set of 30 GUI visual techniques defined on a continuum between two extremes (e.g., regularity vs. irregularity) [18] and a set of 14 characteristics determining GUI aesthetics associated with their significance based on a formula providing a quantitative way of computing each technique:

- A *visual technique* [18] is considered as a commonly accepted visual principle regulating the visual elements of a GUI. It refers to a purely theoretical concept.
- An *aesthetics metric* [9,10] evaluates a visual technique via a mathematical formula for its quantification.

In this regard, we can exemplify the balance property. Considered as a visual technique, balance is a search for horizontal and vertical equilibrium between existing interface

EICS'13, June 24–27, 2013, London, United Kingdom.

objects [18]. Ngo *et al.* [9], guided by their goal to quantify aesthetics, introduce the optical weight in their definition of balance as the perception that one can have of an object weight. In a picture, the human eye perceives objects as more important than others. The metric of balance then results from the difference between the total weight of the various components of each part of the vertical and horizontal axes [9], a mathematical formula (Figure 1).

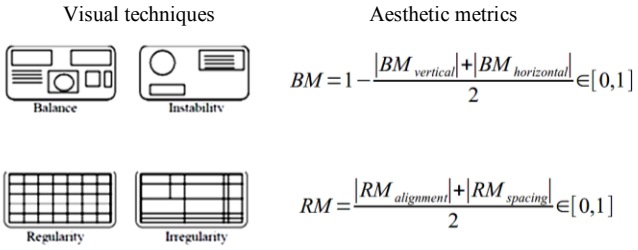

Figure 1. Two metrics associated with a visual technique: balance vs. instability, regularity vs. irregularity).

Many other GUI aesthetics properties could be defined similarly: for each principle (e.g., a visual technique or a general principle), a formalization of its rationale should be obtained that reflects a high value when considered positive by a human and a low value when considered negative. The numerical value of each formula should be correlated to the human estimation of its level.

CONTEXT AND MOTIVATIONS

GUI aesthetics have been extensively studied [1,2,6,7,9,13, 16], with some studies providing empirical evidence that aesthetics influence user experience [7,13]. Table 1 identifies four cases to be investigated where aesthetics could have an influence on user experience. A GUI that is considered both usable and aesthetic will probably be accepted positively, while a GUI that is both unusable and anesthetic will probably be rejected negatively [6]. A GUI that is usable and anesthetic or esthetic and unusable, poses some new challenges for evaluation methods.

Usable UI	Aesthetic UI	User eXperience
Yes	No	Good/Bad (?)
No	No	Bad
No	Yes	Good/Bad (?)
Yes	Yes	Good

Table 1. Cases of aesthetics influence on user experience.

RESEARCH SCOPE, GOALS AND METHODS

In this section, we first describe the scope of the research to be undertaken around the quality standard. Then, we define the research goals and finally propose a methodology to be used. We identify two main issues: the representativeness and the relevance of aesthetics metrics. On the one hand, representativeness explores the extent to which the result of a metric represents end-users opinions. On the other hand, relevance opens a larger question and addresses to what extent an evaluation based on one specific metric is relevant for a GUI in one specific domain, involving one typical task, in one determined environment and/or on one distinctive device. In this regard, an evaluation of GUI balance could be appropriate for a well-structured website where information content is the message to convey. On the contrary, it would be non-sense to assess balance of an artist's personal homepage since this visual technique is not on the side of contrast.

Scope of the research

Evaluating a GUI could be achieved by many methods, including standard compliance based on a standard related to quality, such as ISO9126 or ISO 25000 for software quality, or any other framework for quality management.

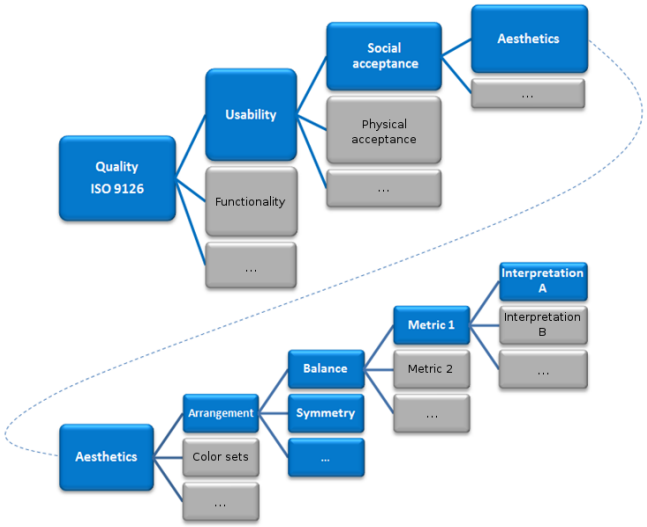

Figure 2. Conceptual map describing scope of research.

The ISO9126 standard for quality management is decomposed into six quality factors: functionality, reliability, usability, efficiency, maintainability and portability. *Usability* itself encompasses various sub-factors such as social acceptance and physical acceptance. One dimension that contributes to social acceptance is aesthetics. Aesthetics cannot be evaluated per se, since it involves multiple criteria that are inter-related: arrangement of objects on a display, used color sets and design of information support. Those criteria could be themselves recursively decomposed into sub-criteria until a measurable criteria is reached, for instance such as symmetry, balance, density, etc. For each property, one or several metrics may exist.

A *metric* is a quantitative measure linked to a known property. Finally, for each metric, we can find many *interpretations* (e.g. formula). This work is aimed at supporting computation of metrics through their interpretation of formula so as to investigate aesthetics in a computational way. In other words, we seek to point to elements influencing aesthetics through a backward tracking departing from the smallest item (metric formula) in order to establish a general model of aesthetics. The value of a metric must not necessarily be interpreted as good or bad. Its main goal is to describe or compare different interfaces and to generate new solutions.

Research Goals

Metrics used so far are partial and are used to characterize visual quality at a given time. By measuring these metrics, and others to define, we could establish a GUI quality model to estimate and monitor over time evolution of the aesthetics and usability of an interface, whether for a particular information system or for several linked information systems. This work therefore aims at:

- Defining the scope of research around the *quality norm* which is commonly used to characterize good practices in GUI design.
- Motivating, defining and formalizing its *underlying concepts*, such as usability, social acceptance, aesthetics, interpretation, recommendation, and formula.
- Defining a model based on metrics to characterize these concepts in different contexts of use.
- Designing and developing a method to estimate these metrics in various contexts of use.
- Validating the methodology on the basis of experimental studies.
- Developing a tool to proceed to automatic and semi-automatic metrics computations.

The research goal is to demonstrate relevance of using aesthetics metrics to evaluate interface design. Therefore, it is essential to conduct studies on concrete cases throughout the thesis. The following section presents a methodology adopted with this perspective in mind.

Methodology

In this section, we present a methodology to be used to conduct empirical studies under which results may help to answer the main research statement and, more specifically, the question of representativeness.

1. Formulate hypotheses to be empirically verified. Focus on what needs to be validated using statistical tests.
2. Select several concrete GUIs, e.g., webpages, application, based on their differences in terms of goals and, for each one, assess its visible aesthetic orientation in order to establish a corpus of interfaces to be analyzed.
3. Perform a mathematical formalization in order to transform visual techniques, mainly qualitative, into quantitative aesthetics metrics. Some metrics (balance, simplicity…) already have defined mathematical formulas [9], other metrics (e.g., alignment, grouping…) still need to be provided with formula.
4. Compute metrics values for each interface based on these formulas (Figure 3).
5. Conduct a survey among end-users to get reviews about the value of each visual technique for each considered interface.
6. Perform statistical tests to compare quantitative results obtained using defined formulas with user reviews provided by the survey.
7. Analyze the results with further discussions and try to draw relations between metrics results and user perceptions of a UI quality keeping in mind the main research statement.

Metric	Val.
Alignment	78%
Understatement	31%
Balance	35%
Symmetry	62%
Regularity	75%
Simplicity	35%

Economy	56%
Density	69%
Unity	41%
Grouping	80%
Consistency	89%
Predictability	88%
Continuity	52%

Figure 3. GUI evaluation method: selecting interface, designing interface grid, computing aesthetics metrics.

RESEARCH SITUATION AND DISSERTATION STATUS

The aforementioned methodology has already been applied to a first pilot experiment. The main finding is: it was not possible to legitimate aesthetics metrics as an exact representation of user reviews. However, we observed that it was possible to obtain a relative similarity; a similarity established on the rankings of metrics rather than on their raw values. Furthermore, we are currently working on the development of a software tool called **QUESTIM** (Quality Estimator Tool using Metrics) in order to support the auto-

matic or semi-automatic computation of metrics. It is a tool for designers requiring an objective evaluation feedback about their work. It is developed as a web service based on Google Web Toolkit, providing an evaluator of GUI quality using aesthetics metrics (Figure 4). The main goal of this tool is to load a website page or screenshot and proceed to automatic metrics computation or semi-automatic computation with objects superimposed by the user on the screenshot.

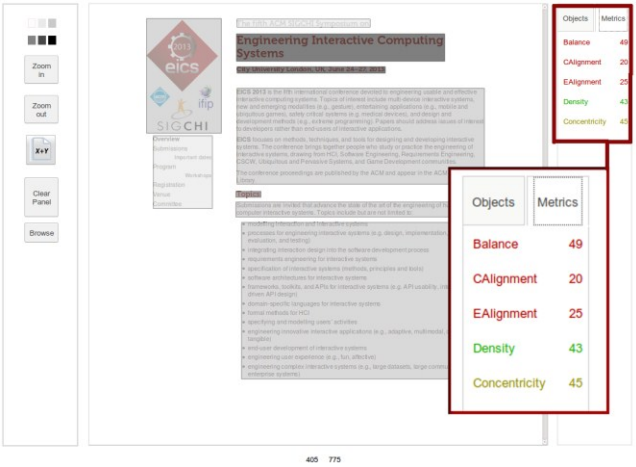

Figure 4. Semi-automatic metrics computation example on EICS2013 web page with QUESTIM.

We already have identified some lines of research to undertake or to further investigate:

- Some metric formulas may be optimized to better reflect human eye perceptions. Indeed, for some metrics, thresholds strongly influencing the outcome were defined empirically. These need to be validated.
- Regarding the automatic evaluation by metrics, we made a small step in providing a tool for drawing objects of an interface and get a quantitative result. It would be interesting to automate most of this evaluation process and propose a system of recognition of shapes and colors. The interest would be, on the one hand, to avoid manual drawing of objects and, on the other hand, the evaluation could be applied to any interface regardless of the language which is used.
- Finally, we developed a model based on five significant aesthetics metrics. The regression coefficient of this model was low (22%). A study on a larger scale (with a larger number of interfaces) could therefore be carried out to define a valid model of aesthetics.

EXPECTED CONTRIBUTIONS

Through this research project, we hope to provide interface designers with a tool using validated assessment items. The lifecycle of GUI usability design could therefore be directly improved. Indeed, designers or developers would be able to get a direct, detailed and objective feedback about their GUI design. In this way, we want to bring a real added value – validated and supported by the scientific world – to their work.

ACKNOWLEDGEMENTS

This work is partially supported by the QualIHM project (Tool-Supported Methodology for the Quality of Human-Computer Interfaces) of the Cwality research programme by Direction Générale DGO6 of Région Wallonne. A special thank goes to our colleagues Ugo Sangiorgi, Vivian Genaro Motti, and Suzanne Kieffer for comments on earlier version of this manuscript.

REFERENCES

[1] Birkhoff, G.D. *Aesthetic Measure*. Harvard University Press, Cambridge, 1933.

[2] Coyette, A., Kieffer, S., Vanderdonckt, J. Multi-Fidelity Prototyping of User Interfaces, *Proc. of INTERACT'2007*. Lecture Notes in Computer Science, Vol. 4662. Springer-Verlag, Berlin, 2007, pp. 149-162.

[3] Florins, M., Montero, F., Vanderdonckt, J., Michotte, B. Splitting Rules for Graceful Degradation of User Interfaces, *Proc. of ACM Conf. on Advanced Visual Interfaces AVI'2006*. ACM Press, New York, 2006, pp. 59-66

[4] Galitz, W.O. *The Essential Guide to User Interface Design*, 3rd *ed*. John Wiley & Sons, New York, 2007.

[5] González, S., Montero, F. and González, P. BaLOReS: a suite of principles and metrics for graphical user interface evaluation, *Proc. of INTERACCION '12*. ACM Press, New York, 2012, Article 9.

[6] Kurosu, M. and Kashimura, K. Apparent usability vs. Inherent usability, *Proc. CHI '95 Conf. Comp*. ACM Press, New York, 1995, pp. 292-293.

[7] Lavie, T. and Tractinsky, N. Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. *Int. Jour. of Human-Computer Studies*, 60(3), 2004, pp. 269-298.

[8] Montero, F. and López-Jaquero, V. Guilayout++: Supporting Prototype Creation and Quality Evaluation for Abstract User Interface Generation, *Proc. of the 1st Workshop on USer Interface eXtensible Markup Language* UsiXML'2010, Berlin, Thalès, 2010, pp. 39-44.

[9] Ngo, D.C.L., Teo, L.S. and Byrne, J.G. A Mathematical Theory of Interface Aesthetics. *Vis. Math.*, 4, 2000.

[10] Ngo, D.C.L., Teo, L.S. and Byrne, J.G. Modelling interface aesthetics. *Information Sciences*, Vol. 152, 2003, pp. 25–46.

[11] Mahajan, R., and Shneiderman, B. Visual and textual consistency checking tools for graphical user interfaces. IEEE Trans. on Software Engineering, 23(11), 1997, pp. 722-735.

[12] Mullet, K., and Sano, D. *Designing visual interfaces: Communication oriented techniques*. 1994.

[13] Sonderegger, A. and Sauer, J. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *App. Erg.*, 41 (3), 2010, pp. 403-410.

[14] Sears, A. AIDE: A step toward metric-based interface development tools, *Proc. of UIST'95*. ACM, pp. 101-110.

[15] Streveler, D.J. and Wasserman, A.I. Quantitative measures of the spatial properties of screen designs, *Proc. of INTERACT'84*. 1994.

[16] Tractinsky, N., Katz, A.S. and Ikar, D. What is beautiful is usable. *Interacting with Computers*, 13, 2000, pp. 127-145.

[17] Tullis, T.S. The formatting of alphanumeric displays: A review and analysis. *Human Factors*, 25(6), 1983, pp.657-682.

[18] Vanderdonckt, J. and Gillo, X. Visual techniques for traditional and multimedia layouts, *Proc. of ACM Conf. on Advanced Visual Interfaces AVI'94*, ACM, 1994, pp. 95-104.