Available online at www.sciencedirect.com
ScienceDirect
Interacting with Computers 19 (2007) 342–356
Interacting with Computers
www.elsevier.com/locate/intcom

Towards an empirical method of efficiency testing of system parts: A methodological study

Willem-Paul Brinkman [a,*], Reinder Haakma [b], Don G. Bouwhuis [c]

[a] School of Information Systems, Computing and Mathematics, Brunel University, Uxbridge, Middlesex, UB8 3PH, United Kingdom
[b] Philips Research Laboratories Eindhoven, Prof. Holstlaaan 4, 5665 AA Eindhoven, The Netherlands
[c] Technische Universiteit Eindhoven, P.O. box 513, 5600 MB Eindhoven, The Netherlands

Received 7 December 2005; received in revised form 5 January 2007; accepted 19 January 2007
Available online 26 January 2007

Abstract

Current usability evaluation methods are essentially holistic in nature. However, engineers that apply a component-based software engineering approach might also be interested in understanding the usability of individual parts of an interactive system. This paper examines the efficiency dimension of usability by describing a method, which engineers can use to test, empirically and objectively, the physical interaction effort to operate components in a single device. The method looks at low-level events, such as button clicks, and attributes the physical effort associated with these interaction events to individual components in the system. This forms the basis for engineers to prioritise their improvement effort. The paper discusses face validity, content validity, criterion validity, and construct validity of the method. The discussion is set within the context of four usability tests, in which 40 users participated to evaluate the efficiency of four different versions of a mobile phone. The results of the study show that the method can provide a valid estimation of the physical interaction event effort users made when interacting with a specific part of a device.

Keywords: Efficiency; Usability testing; HCI methodology; Usability evaluation method; Log file analysis; Empirical method

1. Introduction

Evaluating the usability of a device on dimensions such as effectiveness, efficiency and satisfaction (ISO, 1998) has received considerable research attention in the past few decades. Researchers have come to realise that the acceptance of a system is significantly dependent on for example the ease with which people can operate a system (e.g. Davis, 1989; Davis and Venkatesh, 2000). Empirical usability evaluation methods such as user observations, questionnaires, and interviews are all tools that engineers can use to examine the usability of their system. Still, these methods do not provide quantitative data about the actual use of specific parts of the system, critical information when engineers apply a component-based software engineering (CBSE) approach. Because of the popularity of CBSE, it is important to have a suitable evaluation approach.

CBSE can be regarded as a response to the increase in the complexity of systems. As the complexity increases, design, development, and maintenance become more difficult. In response, software engineers have moved away from dealing with a system as a whole, and instead favour a more modularised or a component-based approach. The aim is to create autonomous components that hide the internal complexity from other components. This idea is considered as one of the major success factors behind object-oriented development; it reduces the complexity of large software projects and improves the maintainability and reliability of a system (Cox, 1990). In this approach, systems are not developed from scratch but are assembled

* Corresponding author. Tel.: +44 0 1895 266043; fax: +44 0 1895 251 686.
E-mail addresses: willem.brinkman@brunel.ac.uk (W.-P. Brinkman), reinder.haakma@philips.com (R. Haakma), d.g.bouwhuis@tue.nl (D.G. Bouwhuis).

0953-5438/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.intcom.2007.01.002

by using pre-produced parts (e.g. pop-menus, radio buttons, or more complex components such as a spell checker or an email component), which can be used in different applications. The promise of CBSE is reduced development cost and time, since ready-made and bespoke components can be used and re-used (Aykin, 1994).

CBSE and software development in general have also been studied in the context of interactive systems. For example, the IFIP's Working Group 2.7(13.4) (Gram and Cockton, 1996) has studied links between a set of user-perceivable properties of interactive systems, such as goal and task completeness, flexibility and robustness, with a set of software phenomena as seen from the software engineers' perspective, such as software architecture, tools, documents and code. They argue that CBSE can improve system modifiability and maintainability, which increases the system's lifetime and the ease of keeping it operational. Attempts have also been made (John et al., 2005) to develop usability-supporting architectural patterns, which address usability problems that arise because of software modularisation, such as responding to a user's cancellation command across a series of components. Furthermore, the compositional view has been used to explain and predict human–computer interaction. For example, Taylor (1988a) has proposed a layered interaction framework. He explained how users and components of a system interact across multiple layers. Several interaction mechanisms have been studied within this framework, such as a general protocol grammar (Taylor et al., 1999), diviplexing and multiplexing (Taylor and Waugh, 2000), communication synchronisation (Taylor, 1989), and layered feedback (Haakma, 1999). The framework has also been suggested (Haakma, 1998; Hilbert and Redmiles, 2000; Taylor, 1988b) as a framework for evaluating human–computer interaction, in other words, component-based usability evaluation. Usability can be considered a multi-dimensional construct (ISO, 1998). The evaluation method put forward in this paper, however, focuses only on a part of the efficiency dimension. It is therefore a first exploratory step towards the wider idea of component-based usability evaluation.

1.1. Usability and efficiency evaluation

Selection and customisation of components, when producing a new application, remains a key challenge in the CBSE approach. Applying this approach to interactive systems, gives efficiency evaluation a potential active role in the selection and customisation process of the components. Information about the efficiency of the different components in a new application would help to direct the software-engineers' attention towards components that decrease the overall efficiency of the application. Besides their ability to locate efficiency problems, the effectiveness of these methods also depends on their ability to accommodate a particular development approach, in this case CBSE. For example, simulation models such as GOMS (Card et al., 1983), or the Cognitive Walkthrough (Polson et al., 1992), that explain interaction from a cognitive model can be used when engineers start off by specifying the user interface and the user task. Without these specifications, but with a working prototype built from existing components, heuristic evaluation (e.g. Nielsen and Molich, 1990) or a user test (e.g. Rubin, 1994) would be effective. Heuristic evaluation, and new techniques, such as CASSM (Blandford et al., 2005), are analytical in nature. They do not directly analyse actual user behaviour or opinions. They are often employed to identify usability problems at an early stage of development when it is still relatively less expensive to make adjustments to the system. However, using off-the-shelf components shortens development time, making system adjustments less expensive. CBSE could therefore make the more time consuming empirical oriented techniques more attractive. These techniques such as questionnaires, user observations in the lab or in the field, ground their findings in actual applications usage and not in consolidated knowledge from previous findings with other devices. Interpreting the data and relating this back to design suggestions can however be a difficult step. For example, when it comes to observation-based evaluation, such as a qualitative oriented usability test, evaluators need to provide this link. They observe the users interacting with the system, write down the problem users encounter, but then the evaluator has to attribute these problems to parts of the system so engineers can try to solve them. This process has been criticised as being very subjective, in that different evaluators or entire teams of evaluators come up with completely different lists of problems when examining the same system (Molich et al., 2004), or even when examining the same observation tapes (Hertzum and Jacobsen, 2001). Evaluators might therefore benefit from additional quantitative information about the usability, or more specifically the efficiency with which users operate individual parts of a system. Existing quantitative indicators, however, measure usability on an overall level. For example, the average task completion might be 5.2 min, or the overall satisfaction, learnability or mental load score on a scale from 1 to 7, might be 4.3. How this relates to a design suggestion to improve a particular part of the system is unclear. Some usability questionnaires therefore include questions on specific parts of the system, such as font size, error messages and help facilities (e.g. Chin et al., 1988). Unfortunately, when it comes to behavioural measures, a link to the efficiency of a specific part of the system is currently lacking.

1.2. Component evaluation strategies

When evaluating components of a system, there are two basic strategies that can be applied: stepwise testing, and big bang testing (Broekman and Notenboom, 2003). Stepwise testing means that the test starts with a single or a limited number of components and is extended with other components each time the test results are satisfactory. If