Evaluating adaptable and adaptive user interfaces: lessons learned from the development of the AVANTI Web browser

Constantine Stephanidis, Alexandros Paramythis, Michael Sfyrakis

Institute of Computer Science
Foundation for Research and Technology - Hellas
Science and Technology Park of Crete
Heraklion, Crete, GR-71110, Greece
{cs, alpar, sfyrakis}@ics.forth.gr

Abstract. One of the critical issues in the development of adaptable and adaptive user interfaces concerns the lack of appropriate evaluation methods and techniques. Existing approaches cannot be used to assess the way and extent to which the adaptation facilities of the interface affect interaction qualities such as accessibility and usability. This paper reports on the particular approach taken within the development of the AVANTI Web browser for the assessment of the adaptation characteristics of the user interface. Based on the practical experience gained, a number of requirements have emerged towards the development of generic methods and techniques for evaluating adaptation-capable user interfaces.

1. Introduction and Background

The AVANTI project[1], which was concluded in August 1998, aimed to address the interaction requirements of individuals with diverse abilities, skills, requirements and preferences (including disabled and elderly people), using Web-based multimedia applications and services. Detailed descriptions of the AVANTI system and its main components can be found in [Fink, Kobsa & Nill, 1997; Stephanidis et. al., 1998; Fink, Kobsa & Nill, 1999].

The AVANTI Web browser acts as the front end to the AVANTI system, and aims to satisfy diverse end user abilities, requirements and preferences, as well as different usage contexts. The design and development of the AVANTI Web browser followed the Unified User Interface Development approach [Stephanidis, Savidis & Akoumianakis, 1997], which led to the construction of a unified browser interface, capable of adapting itself to suit the requirements of different user categories: able-bodied, people with light or severe motor disabilities, and blind people [Stephanidis et al., 1998].

Unified User Interfaces employ adaptation techniques to automatically tailor themselves to different sets of user and usage context characteristics [Stephanidis, Savidis & Akoumianakis, 1997]. Specifically, the design phase involves the construction of a *polymorphic task hierarchy*,

[1] Part of the R&D work reported in this paper has been carried out in the context of the ACTS AC042 AVANTI project "Adaptive and Adaptable Interactions to Multimedia Telecommunications Applications" (September 1995 - August 1998), partially funded by the European Commission (DG XIII). The AVANTI consortium comprised: ALCATEL Siette (Italy) - Prime contractor; CNR-IROE (Italy); ICS-FORTH (Greece); GMD (Germany); University of Sienna (Italy); MA Systems (UK); MATHEMA (Italy); VTT (Finland); ECG (Italy); University of Linz (Austria); TELECOM ITALIA (Italy); EUROGICIEL (France); TECO (Italy); Studio ADR (Italy).

within which different tasks may have alternative instantiations in the user interface (called *instantiation styles*, or simply *styles*). The task decomposition thus proceeds in a polymorphic fashion, defining alternative styles and task hierarchies, according to requirements and preferences of different user categories. In other words, different styles define alternative ways in which a specific task can be realised.

Following Unified User Interface Design, the resulting single design artefact may have multiple instantiations during initiation of interaction (adaptability), in order to ensure accessibility for a wide range of users. Moreover, each interface instance is continuously enhanced at run-time (adaptivity), in order to provide high-quality of interaction to all potential users. In AVANTI, both adaptability and adaptivity were supported through a rule-based framework, which necessitated the transformation of the design rationale as this was captured in the polymorphic task hierarchy, into corresponding adaptation rules [Stephanidis et al., 1998]. A detailed description of the Unified User Interface Design method in general, and polymorphic task hierarchies in particular, can be found in [Savidis et al., 1997].

One of the key problems in the development of self-adapting user interfaces today is the inadequacy of available evaluation methods and techniques to be used for the evaluation of adaptable and adaptive interfaces. Specifically, existing evaluation methods are appropriate for assessing "static" user interfaces, but not the way and extent to which the dynamic adaptation facilities of the interface affect interaction qualities, such as accessibility, usability, acceptability, etc. Although there have been several attempts in the past to construct both objective and subjective expert- and user-based evaluation methods in the are of interface adaptation (e.g., [Totterdell & Boyle, 1990; Grüniger & van Treeck, 1993; Höök, 1997]), the lack of understanding of the dynamic dimensions of adaptive user interfaces (as well as of the differences introduced by alternative approaches to achieving and "driving" adaptive behaviour), compromises the applicability of solutions that have been suggested to date. The main deficiency of most of the aforementioned approaches is that they fail to identify the adaptation-oriented characteristics of the user interface which have detrimental, or, adversely, beneficial effects on interaction.

Due to these shortcomings, the approach taken in the evaluation of the adaptable and adaptive user interface of the AVANTI Web browser has been the introduction of a two-fold assessment process, which involved:

(a) Iterative, expert-based assessment cycles in the design of appropriate interaction styles, the definition of adaptation rules, and the development of the decision mechanism for materialising the required adaptation behaviour; expert-based assessment has been intended to compensate for the lack of appropriate evaluation techniques for adaptation-capable user interfaces and the lack of empirical evidence upon which to base the design of adaptations.

(b) End-user based evaluation activities (using questionnaires, observations and interviews), intended to assess the overall usability and accessibility of the user interface.

2. Expert evaluation

Expert evaluation activities within the development of the AVANTI Web browser aimed to employ accumulated knowledge and experience in the areas of user interface design, usability, and assistive technology, for: (a) the design of alternative interaction styles that cater for the

different user and usage-context requirements, as well as (b) the design of appropriate adaptation behaviour to be built in the resulting interface.

2.1 Evaluating the design of interaction styles

Early evaluation activities were intended to assess the appropriateness of the designed interaction styles for the specific interaction context and the particular user characteristics for which they were intended. Particular emphasis was placed in the evaluation of the accessibility features provided by the designed interaction patterns to the target disabled user categories (i.e., blind and motor-impaired).

The experts reviewed each interaction style separately, based on established accessibility and usability guidelines and heuristics (e.g., [Vanderheiden & Kaine-Krolak, 1995; ISO, 1997; Microsoft, 1997; Story, 1998]), and were asked to identify potential accessibility, usability, or other problems of each style, as well as to propose possible improvements in the design, based on their experience. The outcome of these inspection activities was collected and analysed, and the results were fed back into the design process, where they have led to three types of intervention to designed interaction styles: (a) re-design of styles, based on identified problems, or on contributed ideas towards their enhancement; (b) elimination of redundant (due to the similarity in the characteristics of the end-users or the usage contexts they were intended to cater for) interaction styles; (c) introduction of new interaction styles, to cover user characteristics and contexts of use that were not addressed adequately by existing styles.

2.2 Validating the adaptation rules

The development of the adaptation rules took place in two steps. Firstly, the rules were defined by a group of experts, through several iterations following each task-decomposition phase, as well as each stage in the definition / selection of alternative interaction styles. Secondly, a process was defined, to assess the design of adaptations by validating the resulting adaptation rules. A detailed description of the adaptations rules is available in [Stephanidis et al., 1997], while a summarising account can be found in [Stephanidis et al., 1998].

The validation of the adaptation rules has itself taken place in three consecutive phases: evaluation of the rules by experts; verification of the adaptation mechanism on a per-rule basis; and, verification of the adaptation mechanism across sets of rules. The results of the experts' assessment phase have led to four types of intervention to adaptation rules: (a) elimination of rules that were deemed inappropriate, or not sufficiently supported; (b) introduction of new rules (based on the recommendations of the experts); (c) modification of the rules' triggering conditions (e.g., adding, or removing a particular user characteristic from the description of the triggering situation); (d) modification in the rules' decisions (e.g., addressing a particular situation only through guidance, instead of through guidance and extensive interim feedback).

The validation of adaptation rules was followed by the verification of the adaptation mechanism on a per-rule basis, and verification of the system's adaptation behaviour across sets of rules. The mechanism used for valuating rules and carrying out the respective adaptation decisions was tested for consistent behaviour, by: (i) testing that the triggering conditions for each individual rule led to the desired (adaptation) behaviour on the part of the user interface; (ii) testing sets of rules in combination, to assess the degree to which they affect each other from a functional, as well as from the user's point of view. The former procedure (i.e., testing rules individually) was

performed by examining the defined rules one-by-one and verifying system behaviour, when the activation parameters were set, or changed. A "wizard of oz" technique was used to simulate the functionality of the User Modelling Server[2]. The later procedure (i.e., testing combinations of rules), was performed through the development of representative scenaria, where multiple activation parameters were set or changed simultaneously.

The verification procedure resulted in the identification of conflicts in the activation of specific styles and inappropriate activation of certain rules in specific tasks. The main problem arose from the redundant activation of styles under certain conditions. The outcomes of the validation procedure initiated specific modifications in the pre-defined rules, as well as the adaptation mechanism itself.

3. User-based evaluation

Formal usability evaluation studies of the AVANTI Web browser have been carried out in the context of the experimental and field evaluations of the AVANTI system in the three user trial sites in Kuusamo, Siena and Rome [Andreadis et. al., 1998]. These experiments evaluated the overall usability of the AVANTI information systems following a common evaluation framework. The trials were performed on distributed heterogeneous network environments supporting different access points, including: public information kiosks, home computers, travel agency offices, and laboratory sites. The subjects that took part in the experiments included citizens and tourists at the trial sites, as well as travel agency staff (in the case of the Siena information system). In terms of physical abilities, subjects were drawn from all three categories supported by the project, i.e., able-bodied, blind and motor-impaired. In total, there were 175 subjects in all three experiments, exposed to more than one instance of the user interface, sometimes through iterative evaluation sessions.

The usability goals set up by the common evaluation framework, and assessed during the experiments were: learnability, efficiency of use, memorability, errors, satisfaction, user attitude, adaptability and adaptivity. The experiments adopted a task-based evaluation approach, utilising both qualitative and quantitative evaluation methods. The tools which were utilised include observations and interviews (qualitative) as well as attitude scale questionnaires and log-files (quantitative). The functionalities of the AVANTI Web Browser as well as the supported adaptability and adaptivity features were addressed in the observation sessions, the interviews and the subjective evaluation (attitude scale) questionnaires.

The results of the evaluation were encouraging, both in terms of user acceptance of the characteristics of the interface, and in terms of the fulfilment of the initial goals that led to the employment of adaptation techniques in the user interface. In particular, adaptability addressing accessibility issues for the various end-user groups proved quite successful, as each user category conceived the interface as having been specifically developed to cater for their particular requirements. The results were similar for the non-disability related categories in which users were classified (e.g., according to their computer expertise).

Adaptivity was assessed to a lesser degree than adaptability during the evaluation, mainly due to the following reasons: (a) adaptivity requires that interactive sessions are rather lengthy, so that

---

[2] For this purpose, a software module that simulates the functioning of a user model server has been developed and was used for "manually" generating the dynamic user situations.

adequate information about the user and the context of use is collected, before any practically useful inferences can be made, and (b) existing user interface evaluation techniques do not offer themselves for the evaluation of dynamically changing, non-deterministic (from a user's perspective) systems. As far as the first of the above issues is concerned, the typical duration of the interaction sessions performed during the experimental activities was not adequate for the extraction of dependable inferences that dynamic adaptation could be based on. As a result, users were aware of only a minimal set of adaptive features in the interface; however, their reaction to the features they did observe was positive. As far as the availability of empirical methods and techniques for the evaluation of adaptivity is concerned, it has already been argued that existing knowledge in the area of user interface evaluation is inadequate for the derivation of appropriate techniques and instruments to measure the effects of adaptive system behaviour on interaction.

4. Conclusions

As already argued, there do not exist today evaluation methods and techniques that adequately address the assessment of adaptable and adaptive user interfaces, and their evaluation must, therefore, be carefully planned and conducted on a case-by-case basis. Specifically, evaluation of adaptation-capable user interfaces should aim to identify those aspects of the interface that have beneficial / detrimental effects on the accessibility and interaction quality offered by the interface for different categories of users and in different contexts of use. Two coarse evaluation dimensions can be derived from the above goal. The first concerns the appropriateness of the different instantiation styles for the purpose they were developed. This entails the assessment of the styles themselves as individual interactive artefacts and as components of the overall interface, as well as the assessment of the design rationale / decision logic that activates (or deactivates) these styles, based on user and usage characteristics. The second dimension concerns the evaluation of the dynamic adaptation (adaptivity) in the interface. This is in fact the most difficult part to evaluate, as there are multiple factors that determine the various qualities of the interface. For example, an adaptation may be conceived as entirely dissatisfactory by the user if: (a) the adaptation logic itself is flawed; (b) the "triggers" of the adaptation were wrongly inferred by the user modelling component; (c) the adaptation was not "timely" (e.g., it came "too late" from the user's perspective); or (d) the adaptation policy is not satisfactory (e.g., because the user is not given enough control over it).

To counterbalance the inherent difficulties in evaluating dynamic adaptation in the interface, evaluators should plan the evaluation process carefully from the early design phases, and should actually base the evaluation plan on the overall design process. Thus, evaluation should not be restricted to summative activities; rather, it should proceed in parallel to the design of the user interface and should strive to identify deficiencies and possible problems as early as possible, informing and guiding the development process. The evaluation activities of the AVANTI Web browser can be considered as preliminary steps towards generic methods and techniques for the evaluation of adaptation-capable user interfaces. However, a lot more research and practical experience are required in this direction, before we can derive valuable results that will be reusable across application domains, user categories and contexts of use.

5. References

[Andreadis et. al., 1998] Andreadis, A., Giannetti, L., Marchigiani, E., Rizzo, A., Schiatti, E., Tiberio, M., Pentila, M., Perala, J., Leikas, J., Suihko, T., Emiliani, P.L., Bini, A., Nill, A., Sabbione, A.,

Sfyrakis, M., Stary, C., & Totter, A. Global evaluation of the experiments. ACTS AC042 AVANTI Project, Deliverable DE030, August 1998.

[Fink, Kobsa & Nill, 1997] Fink, J., Kobsa, A., & Nill, A. Adaptable and Adaptive Information Access for All Users, Including the Disabled and Elderly. In A. Jameson, C. Paris, & C. Tasso (Eds.), *Proceedings of the 6th International Conference on User Modelling (UM '97)*, Sardinia, Italy (pp. 171-173). New York: Springer-Verlag, 1997.

[Fink, Kobsa & Nill, 1999] Fink, J., Kobsa, A., & Nill, A.: Adaptable and Adaptive Information Provision for All Users, Including Disabled and Elderly People. *New Review of Hypermedia and Multimedia*, in press.

[Grüniger & van Treeck, 1993] Grüniger, C., & van Treeck, W. Contributions of a Social Science Based Evaluation for Adaptive Design Projects. In M. Schneider-Hufschmidt, T. Kühme, & U. Malinowski (Eds.), *Adaptive User Interfaces: Principles and Practice* (pp. 319-330). Amsterdam, The Netherlands: Elsevier Science Publishers, 1993.

[Höök, 1997] Höök, K. Evaluating the Usability of an Adaptive Hypermedia System. In *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI '97)*, Orlando, USA (pp. 179-186). New York: ACM Press, 1997.

[ISO, 1997] International Standards Organisation. *Ergonomic Requirements for office work with visual display terminals, Part 11: Guidance on Usability (ISO 9241-11)*. Draft International Standard (DIS). Geneva, Switzerland: International Standards Organisation.

[Microsoft, 1997] Microsoft Corporation. *The Microsoft Windows Guidelines for Accessible Software Design*. [On-line]. Available: http://www.microsoft.com/enable/dev/guidelines/ software.htm

[Savidis et al., 1997] Savidis, A., Paramythis, A., Akoumianakis, D., Stephanidis, C. Designing user-adapted interfaces: the unified design method for transformable interactions. In *Proceedings of the ACM Conference on Designing Interactive Systems: Processes, Methods and Techniques - DIS '97* (pp. 323-334), Amsterdam, The Netherlands, 18-20 August 1997.

[Stephanidis et al., 1997] Stephanidis, C., Karagiannidis, C., Paramythis, A., Sfyrakis, M., Stergiou, A., Savidis, A., Schreck, J., Fink, J. (February 1997) Report on Rules. ACTS AC042 AVANTI Project, Deliverable DE012, 82 pages.

[Stephanidis, Savidis & Akoumianakis, 1997] Stephanidis, C., Savidis, A., and Akoumianiakis, D. Unified Interface Development: Tools for Constructing Accessible and Usable User Interfaces. Tutorial 13, HCI International '97 Conference, San Fransisco, USA, August 1997. Available on-line at: http://www/proj/at-hci/html/tutorials.html.

[Stephanidis et al., 1998] Stephanidis, C., Paramythis, A., Sfyrakis, M., Stergiou, A., Maou, N., Leventis, A., Paparoulis, G., and Karagiannidis, C. Adaptable and Adaptive User Interfaces for Disabled Users in AVANTI Project. 5th International Conference on Intelligence in Services and Networks (IS&N '98), "Technology for Ubiquitous Telecommunications Services", Antwerp, Belgium, 25-28 May 1998.

[Story, 1998] Maximising Usability: The Principles of Universal Design. *Assistive Technology Journal*, 10 (1), 4-12.

[Totterdell & Boyle, 1990] Totterdell, P., & Boyle, E. The Evaluation of Adaptive Systems. In D. Browne, P. Totterdell, & M. Norman (Eds.), *Adaptive User Interfaces* (pp. 161-194). London: Academic Press, 1990.

[Vanderheiden & Kaine-Krolak, 1995] Vanderheiden, G.C. & Kaine-Krolak, M. Access to Current and Next-Generation Information Systems by People with Disabilities [On-line]. Available at: http://www.tracecenter.org/docs/access_info_sys/full_doc.htm