Evaluating Software Testing Methods in an Active and Assisted Living Context

Stefan H. Ruscher, Josef Weinbub, Siegfried Selberherr

With the recent trend of older adults and individuals with handicaps employing smart technologies in their everyday lives, new challenges in usability are faced, as ageing leads to a varying individual decline of body functions. Yet, methods to assess usability of Active and Assisted Living solutions are not considered in the context of the ageing target audience. To tackle this shortcoming, available usability methods and their application in an Active and Assisted Living context have been studied, identifying essential shortcomings in test design and documentation. In order to support future developments, methods are discussed with regard to age-related disabilities of the test participants, outlining essential mental and motor functions for every method. Furthermore, recommendations for test design and implementation are presented in order to support usability engineers and researchers when working in an Active and Assisted Living context.

1 Introduction

The European Union is facing a drastic shift in age distribution among citizens towards an over-aged society [1]. Considering the technology generations described by Sackmann and Winkler [2], the so-called computer generation is currently aged between 34 and 51. Nevertheless, elder citizens, belonging to the entertainment appliance generation [3], who did not grow up with the internet technology and are 55 years or older, are getting drawn towards the usage of smart phones, tablets, social media, and even home automation technologies [4]. This trend is increasingly investigated throughout the last decade, due to the growing importance of the field. The scientific community summarizes research, design and development of computer systems and internet services for older adults and individuals with handicaps under the umbrella term Active and Assisted Living (AAL) [5].

One main problem with developing AAL solutions is the diversity of the inhomogeneous group of older adults. As aging is universal but highly individual [6], requirements for the usability of computer systems vary drastically between individuals, caused by different levels of (dis-)abilities. The International Organization for Standardization (ISO) offers a broad overview of potential individual decline in mental and motor skills within the ISO/IEC Guide 71:2014 [7]. Based on functional decline as well as limited understanding

1 TU Wien, Institute for Microelectronics stefan.ruscher@gmail.com, selberherr@iue.tuwien.ac.at
2 Christian Doppler Laboratory for HPTCAD, Institute for Microelectronics, TU Wien Gusshausstrasse 27-29/E360, 1040 Vienna, Austria, weinbub@iue.tuwien.ac.at
of software concepts, current research identified problems with specific methods in usability testing, when working with older adults [8]. Yet, there are no technical standards or scientific considerations on accessible usability testing available to overcome this issue. Based on this shortcoming, this work investigates common practices of software user testing in an AAL context. Initial research revolves around the definition of usability in literature, analyzing research findings and international standards (Section 2). Additionally, a collection of usability testing methods is conducted from meta studies on software testing in general. Transferring the knowledge presented into an AAL context, nine studies on testing AAL solutions are analyzed with respect to the used methodological approaches and overall scientific quality of the work (Section 3). This exploration reveals major shortcomings on the assessment of the applied test methods. In order to tackle this issue, recommendations on test design are provided (Section 4). The discussed methods are evaluated regarding accessibility in an AAL context with respect to the characteristic decline in abilities of the AAL target audience, as described in ISO/IEC Guide 71 [7].

2 Testing Software User Interfaces

2.1 Usability Criteria

The literature does not provide a consensus on the term usability. Usability, being a multidimensional construct determined by the task, the users, the product, and the environment [9], is most commonly described - in accordance with ISO 9241-11 [10] - as effectiveness, efficiency, and satisfaction, which can be reached by a specific user to achieve specific goals in a specific context. Nevertheless, Nielsen describes learnability, memorability, and error recovery as three additional attributes of usability [11], which are also part of ISO 9241-110 [12].

In 2003 Karoulis and Pombortsis stated, that they suspect a positive correlation between usability and learnability, however, without any study attempting to prove it [13]. This correlation was later analyzed by Lewis and Sauro identifying a moderate correlation between both [14]. Yet, they outline differential information provided by the two measures. These findings are confirmed by Borsci et al., who, therefore, suggest a disassociation of usability and learnability, in order to identify e.g., systems with high learnability but low usability [15].

Additional criteria for usability used across journal articles [9] and international standards [12] are controllability, helpfulness, adaptability, affect, conformity with expectations, ability of self-description, task suitability, and fault tolerance.

As all presented criteria address different aspects of usability, depending of the software purpose, none will be excluded from the further analysis. Yet it is noted, that the proper selection of criteria to be tested is essential, as it drives the overall test design and promotes transparency, as demonstrated by Jeng [9] and Ruscher et al. [8].
2.2 Software Usability Testing Methods

In 2005 Holzinger investigated common usability testing methods, analyzing them with respect to their advantages and disadvantages, also distinguishing between inspection and test methods [16]. Jeng performed a similar examination, exploring methods and criteria of various authors [9]. Due to different research fields and inclusion of various sources, only a partial overlap between the techniques identified by both authors can be found. As this work focuses on test methods for the primary AAL target audience, older adults and individuals with handicaps, Jeng’s findings are categorized into the two types described by Holzinger, justifying exclusion of inspection methods in later parts, as they were carried out by an expert or a team of experts, rather than by a group of actual users from the target audience [16]. For this work the test methods [16] considered are: card sorting, field observation, focus group, interview, questionnaire or survey, thinking aloud, as well as transaction log and log analysis.

3 AAL User Interface Testing

In order to get deeper insights into common practices on usability testing in an AAL context, a selection of nine representative manuscripts published within the last 23 years is reviewed and analyzed. The analysis focuses on the frequency of methods applied as well as on user groups involved in the tests, also investigating methodological changes with respect to the time at which the studies were carried out.

For almost all studies analyzed, a combination of at least two methods (presented in Section 2.2) can be found. Only Ziefle and Bay [17] rely on a single test method. Even though a wide range of publications is covered by the analyzed literature, no indication of time-related trends can be found. For example, transaction log methods are performed in older [18] and more recent studies [19] [20] alike. Thinking aloud [8] [20] [21] [22] [23] [24] and transaction logs [17] [18] [19] [20] are most commonly used in the investigated test sessions. Additionally, field observations are carried out, either by note taking [18] [20] or video recording [24]. Some of the sessions are accompanied by interviews and questionnaires before or afterwards [8] [19] [20] [21] [22] [23]. Focus groups on the other hand, are hardly performed during usability testing, being used only by Lin et al. [23]. Nevertheless, Strothotte et al. applied them in user requirement gathering [18]. Interestingly, nobody used the card sorting method.

Table 1 provides an overview of the literature analysis and displays the methods employed. While all publications considered within this section focus on testing usability with respect to age- or disability-related impairments, only two sources give detailed descriptions of the participants’ skills and limitations [18] [21]. In contrast to other research investigating visual user interfaces [8] [17] [19] [20] [22] [23] [24], these studies examine tools primarily supporting users with reduced sight. While the work by Strothotte et al. describes the severity of visual impairment of the participants [18], McGookin et al. describe the creation of an artificial visual impairment for healthy users with blindfolds [21].
Table 1. User interface testing methods applied by selected studies in the field of AAL.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Card Sorting</th>
<th>Field Observation</th>
<th>Focus Group</th>
<th>Interview</th>
<th>Question-</th>
<th>Thinking Aloud</th>
<th>Transaction Log</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strothotte et al. [18]</td>
<td>1996</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Ziefle and Bay [17]</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>McGookin et al. [21]</td>
<td>2008</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin et al. [23]</td>
<td>2009</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haggstrom et al. [19]</td>
<td>2011</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burns et al. [22]</td>
<td>2013</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruscher et al. [8]</td>
<td>2016</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As stated by ISO 9241-11, the specification of tasks is an essential part of usability test design [10]. While Burns et al. [22] and Ziefle and Bay [17] provide an overview of specific tasks to be completed by the subjects, other authors do not present all of them. However, this shortcoming causes non-transparent results. This lack in comprehensiveness becomes problematic in particular in the work of Xie and Pearson [20], who present their subjects’ success rates on a set of tasks, without outlining those.

A notable observation is that the researchers tend not to apply renowned test methods which were previously defined in the literature. Only Burns et al. [22] relate their test design to the Perceived Health Web Site Usability Questionnaire (PHWSUQ), a 12-item questionnaire patterned to evaluate health-related websites [25]. Nevertheless, they alter the method to better fit the needs of their own research and the group of participants [22]. Further explicit relations to literature published earlier are shown by Ostergren and Karras who follow the design guidelines of Redish et al. [26] for setting up their initial interface, but do not apply any established methods during test design or actual testing [24].

4 Discussion

Throughout the analysis carried out for this work, a twofold pattern was identified when considering studies on usability and AAL. On the one hand, there are research activities and meta studies, as performed by Jeng [9], dedicated to software usability testing, assessing methods independent from the systems under test in order to improve the test design process. On the other hand, literature presents AAL-driven software development
and testing, incorporating usability methods without any specific reasoning on the decisions made. Nevertheless, to the best of our knowledge, there are no studies carried out, which focus on a methodological approach for efficient and accessible usability testing for the main target audience of AAL solutions. The lack of suitable guidance on test design might cause unfitting test methods being employed, baring the risk of insufficient or even useless test result data, as well as of increased costs for test execution and result evaluation.

Motivated by this shortcoming, we investigated the applicability of test methods presented by Jeng [9] and Holzinger [16], considering mental abilities and motor skills required for participation, as derived from ISO/IEC Guide 71 [7]. Based on these findings, a set of recommendations for usability test designs in an AAL context was created in order to avoid the risk of insufficient result data and increased test costs.

4.1 Applicability of Methods

In order to establish a foundation for structured AAL test design, an approach based on the ISO/IEC Guide 71 [7] was chosen, as this guide provides a comprehensive overview on limitations of and age-related decline in the human body functions. Based on the outlined disabilities of the AAL target audience members presented in Guide 71, the essential physical and mental skills for specific test methods are determined. For consideration of cognitive skills, visual and auditive perception as well as the ability to produce and understand a language are selected. Furthermore, procedural and conceptual thinking are included. Motor skills covered by the analysis include abilities to move extremities and to produce speech.

Table 2 lists test methods explored in Section 2.2 in relation to abilities which might be impaired for older adults or individuals with handicaps. Specific mandatory skills are displayed for the respective test method. If one of these abilities is impaired for the test subjects, the respective method cannot be applied properly in software usability testing. In particular, Table 2 shows that especially the ability to produce and understand language is crucial for many test methods. Nevertheless, the modality used for communication, being either speech or a replacement, for example upper extremities and sign language, is not as important to the applicability of the methods, as demonstrated by Roberts and Fels [27]. Still, the means of communication should suite all parties involved in the testing itself as well as in the analysis of the results, in order to guarantee meaningful findings. Furthermore, it should be considered, that slower forms of communication prolong test durations, which influences performance of the participants and the cost of the test. In any case, it is suggested to assess the effect of switching methods with respect to the abilities of the participants and the number of subjects. In the case of Strothotte et al. [18], for instance, considering the creation of a questionnaire for visually impaired would result in more effort than interviewing the six test subjects. Additionally, the switch from digital to paper based tests might result in similar issues with visual impairment, as those affected cannot use screen readers, besides the conceptual comprehensiveness described by Ruscher et al. [8].
Table 2. Test methods and mandatory skills of the subject for successful execution. Disabilities in the respective motor or cognitive skills indicate an unfitting test method for the subject.

<table>
<thead>
<tr>
<th>Method</th>
<th>Auditive</th>
<th>Conceptual Thinking</th>
<th>Extremity Mobility</th>
<th>Language</th>
<th>Procedural Thinking</th>
<th>Speech</th>
<th>Visual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Card Sorting</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field Observation</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus Group</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Questionnaire</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thinking Aloud</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transaction Log</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The methods field observation and transaction log are accepted to be independent of specific disabilities of participants in general. Nevertheless, the system under test still has to properly address the needs of the user with respect to their individual limitations.

4.2 Recommendations for AAL-Driven Software Testing

Building on the previous analyses, this section provides an overview of relevant aspects when designing and performing usability tests in an AAL context. For all steps described, best practice examples from previously evaluated literature are provided. When designing a usability test for older adults or individuals with handicaps, the same disabilities should be considered as during the conception of the software itself. First, based on needs and skills of the target audience, a set of tasks should be defined, which aligns with the general problem solved by the software. The task description should be accessible and understandable to all parties involved: the study participants, the staff carrying out the study on site and the staff analyzing the results. Best practice examples on clear definitions of such tasks are the studies carried out by Burns et al. [22] and by Ziefle and Bay [17].

For every task, qualitative and quantitative measures should be introduced for usability criteria provided in Section 2.1, accompanied by test methods to collect data on these aspects. The methods should be selected according to Table 2. Ideally, those are based on previously established methods, as demonstrated by Jeng [9] in a general context and by Burns et al. [22] in an AAL setting. Furthermore, a detailed outline of the parameters measured, including pass-fail-criteria for the test, should be given as demonstrated in Ruscher et al. [8].

Relevant knowledge on the actual motor skills, the mental status, and the needs for assistance of every participant with respect to their tasks should be gathered before the test session. Infringement of personal and privacy rights nevertheless should be avoided,
where possible. An example of such an exploration can be found in the work of Strothotte et al. [18]. When multiple iterative test sessions are planned as suggested by Nielsen [11], after every iteration, the test design and tasks should be reflected with respect to imprecise formulation as well as required actions which contradict with the abilities and understanding of the subjects. Resulting changes in the test design should be outlined and reasoned as demonstrated by Xie and Pearson [20].

5 Conclusion

This research work investigates software usability testing in an AAL context. Major shortcomings in related studies were identified regarding the selection of test methods. Facing this problem, this work outlines essential skills of end users, required for a successfully participation in test sessions. Furthermore, a set of recommendations on considerations for usability test design in an AAL context is provided. This provides the foundation for a more structured, transparent, and cost-efficient testing of software and solutions for older adults and individuals with handicaps.

Acknowledgements

The financial support by the Austrian Federal Ministry of Science, Research and Economy and the National Foundation for Research, Technology and Development is gratefully acknowledged.
References


