AAL in the Wild – Lessons Learned

Edith Maier and Guido Kempter

University of Applied Sciences, Hochschulstrasse 1, 6850 Dornbirn, Austria
{edith.maier,guido.kempter}@fhv.at

Abstract. In the EU-funded ALADIN project the prototype of an ambient assistive lighting system was subjected to a three-month test in private households of older people. Despite intensive usability testing in the development phase, field trials pose special challenges including ethical issues such as obtaining informed consent and the need for guidelines for interviewing old people. Besides, real-life settings give rise to particular distortion effects which have to be taken into account in the analysis of the results. Although the findings indicate any overall slight increase in people´s mental and physical fitness, they also suggest how the prototype can be improved in several respects. Above all it has been shown that packaging the technology with social support measures is essential to achieve higher user acceptance. Besides, the article discusses lessons learned related to the organization of user testing in real-life settings.

Keywords: AAL (ambient assisted living), ambient intelligence, lighting assistance, adaptive algorithm, field trial.

1 Introduction

Given a rapidly ageing population, the elderly have become a favorite target group for ICT developers. However, translating research findings into marketable solutions has turned out to be a big challenge for many who have developed smart devices, applications and systems that are meant to benefit older people. Market exploitation has been surprisingly slow despite the enormous economic opportunities due to demographic change. This may be due to reasons such as low market awareness and visibility, lack of sustainable business models, lack of standards, but also regulatory frameworks and public policies that hinder the uptake of age-related ICT-based products and services.

Since we believe that lack of user acceptance is a major barrier to market success of AAL solutions, we involved end-users throughout the duration of ALADIN, an EU-funded project aimed at developing an ambient adaptive lighting system. First of all, we carried out a detailed user requirements analysis with almost two hundred elderly people using in-depth interviews. We then conducted a thorough iterative testing process before the assistive lighting system was deployed in the households of elderly test persons. Nevertheless, the field trials in which we subjected the prototype to a three-month test in real-life settings proved to be a veritable endurance test, the experience of which we want to share with our colleagues.

The paper briefly describes the project aims as well as the prototype and its components. It then goes on to describe the different stages of usability testing in the

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2009, LNCS 5615, pp. 218–227, 2009.

course of its development. We discuss the special challenges we had to cope with during the field trials including the legal and ethical issues raised by usability testing "in the wild". When analyzing the results, various distortion effects related to real-life conditions such as time effects or habituation had to be taken into account. We briefly describe the results from the field tests and their implications for the redesign of the prototype. We then discuss the lessons learned which primarily concern the organization of user testing in real-life settings and the "packaging" of assistive technologies to increase user acceptance.

1.1 The ALADIN Prototype and its Components

The overall aim of the ALADIN project is to contribute to the growing body of knowledge about the effects of light on human health and wellbeing. So-called 'ambient lighting' with varying color temperature and brightness has been in use for some time. However, the user normally has no possibility to interact with the predefined control strategy (mostly defined by the time of the day) and the lighting solutions do not take into account individual differences [1]. Besides, whereas currently most cognitive assessment is done in a clinical setting [2], we use sensor-based monitoring combined with adaptive algorithms to assess people's level of functioning in a continuous way [3].

The fully developed ALADIN prototype consists of an Apple Mini Mac with the installed software booting up automatically, a flat screen television set with service and control interfaces, a USB driven infrared system connected to a normal remote control, a sensor glove with integrated Varioport device and bluetooth connection. A Luxmate bus box controls illumination panels on all four walls of the selected room as well as a ceiling lamp above the table where the test person spends most of the time. The lighting devices consist of luminaires of two colors that can be mixed by changing their intensity.

The computer and the flat screen are meant to replace the existing TV set and be integrated into the home entertainment system. Each testing prototype is a user-specific and flexible composition of the ALADIN components. In the field tests, the prototype was installed in the households of older adults for three months and subsequently removed without leaving any traces. As can be seen in Figure 1, the system comprises the following applications:

- Television (Fernsehen)
- Automatic lighting (Automatisches Licht)
- Manual lighting (Manuelles Licht)
- Exercises (Übungen)
- History (Rückschau)
- Advice & support (Wohlfühltipps)

Each of these applications works independently and has to be started deliberately by the user. "Automatic lighting" adapts the lighting to achieve an activating or relaxing effect. Because of the great diversity of situations, events or individuals, the direction of change or adaptation is not known beforehand. We have opted for genetic and simulated annealing algorithms to implement adaptive control. With "Manual lighting" a user can turn on and off different predefined lighting situations (e.g. reading light) as well as manually modify lighting situations. The application “Exercises” offers the user a variety of activating and relaxing exercises. “Advice & support” allows the user to browse through different recommendations aimed at healthy behavior. Finally, with “History” the user can view the results of the exercises he or she executed during the last five days. The “Advice” and “History” applications derive their input from the results of the exercises.

2 Usability Testing of AAL Applications

2.1 Usability Testing in the Development Phase

In line with our participatory and iterative development approach we carried out both heuristic inspections by experts and end-user tests in the development phase. Given our specific target group, i.e. older people with various impairments such as diminished vision, the user interface of ALADIN had to conform not only to general usability guidelines, but also to accessibility guidelines, more specifically WCAG 2.0, which were published as a W3C Recommendation on 11 December 2008. Since in the course of our user requirements analysis it emerged that a TV set could be found in virtually all households, it was decided to implement the adaptive lighting system on a computer with added TV functionality.

Interactive TV poses special challenges: watching TV is normally characterized by a “lean back” attitude whereas working on a computer is normally associated with a “lean forward” attitude. The human computer interface has been conceived with an active user in mind whereas TV systems are traditionally aimed at passive consumers. A TV screen is normally placed at a distance of a couple of meters from the viewer and has a lower resolution than a computer. When designing the screen layout we cannot use the full screen because TV sets vary in terms of the size used for presentation. The so-called “safe area” can be reckoned to be about 10% smaller than the full screen, or about. 576 x 460 pixels.

The first usability test was conducted early on in the project with a HTML dummy with 12 senior citizens between 65 and 84 years old to examine the screen design of the prototype controlled by a remote control. We assumed that our target users would be familiar with handling *remote controls*. Bernhaupt et al. [4], however, have observed that although people may still want the remote control as an input device, it is perceived as too complex and difficult to use by many. They would prefer a universal remote control with only one button which is integrated with a display to inform users about what they have to do next.

The test persons had no particular suggestions for navigation in addition to the well-accepted rules such as to always show the user’s own position within the software structure, to offer a return to the starting point and give immediate feedback after every user action. The majority of test persons preferred white text colour and blue background colour and were not interested in history information extending for more than seven days.

Mock-up lab testing. The first ALADIN prototype with TV and remote control was carried out with ten clients in a nursing home (65 - 94 years). Users liked the overall graphic design (font, colours, contrast) and found the GUI easy to navigate, provided that the number of navigation elements was kept to a miminum.

Several general observations could be derived from these early user tests:

- The majority of users prefer number keys to cursor keys.
- Users expect immediate confirmation or feedback from the system.
- The number of menu items to choose from should not exceed five or six items.
- (Semi-)technical terms (e.g. biofeedback, sensor belt) have to be replaced with simple ever-day terms.
- Users prefer an information structure with a very flat hierarchy

As far as sensor technology was concerned, the end-user tests led us to discard the chest belt originally envisaged for capturing biosignals. Test subjects found it too difficult to attach to their bodies because this required a degree of motoric dexterity and flexibility that even quite a few of the young assistants did not possess. As a result, the sensor expert in the consortium developed a biosensor glove in the run-up to the field trials. The change in sensor technology brought about a change of methodology: instead of measuring skin conductance response (SCR), peripheral pulse was measured by means of photoplethysmography (PPG) in the field trials.

2.2 Special Challenges Related to Field Trials

As in user testing in general, planning and organizing the field trials comprise preparing a test design, defining selection criteria for test persons and recruiting suitable test persons. When implementing a prototype in private households, however, particular ethical and legal issues arise. In our case, the test design had to take into account the constraints imposed by the high costs of the prototype and the limited time resources available on the part of the partners responsible for the field tests. We also had to take out insurance to cover any damages that might be incurred as a result of the installation of the lighting prototype.

Ethical Issues. An important concern is achieving a balance in the relationship between the demand for a better quality of life and the studies that aim to achieve this on the one hand, and the rights of the research participants on the other. In ALADIN, obtaining informed consent and the protection of personal data proved to be the most important issues and played a role at all stages of the research process, i.e. from the requirements analysis by means of interviews to user testing and dealing with research outcomes. *Policies regarding informed consent in usability testing* are developed by organizations on the basis of generally agreed principles concerning the treatment of human participants. Those principles plus additional ones are enumerated below. Seven of these principles are derived from the related discussion in Dumas and Redish [5]. There is one more principle added in regard to waivers.

1. Minimal risk. Usability testing should not expose participants to more than minimal risk. Though it is unlikely that a usability test will expose participants to physical harm, psychological or sociological risks do arise. If it is not possible to abide by the principle of minimal risk, then the usability professional should endeavour to eliminate the risk or consider not doing the test.

Dumas and Redish [5] citing the Federal Register state that minimal risk means that the probability and magnitude of harm or discomfort anticipated in the test are not greater, in and of themselves, than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.

2. Information. Informed consent implies information is supplied to participants. Based on the suggestions of Dumas and Redish, we supplied the following information: the procedures we would follow; the purpose of the test; any risks to the participant; the opportunity to ask questions; and, the opportunity to withdraw at any time.

3. Comprehension. The professional needs to ensure that each participant understands what is involved in the test. This must be done in a manner that is clear and unambiguous. It must also be done so as to completely cover the information on the form. The procedure for obtaining consent should not be rushed, nor made to seem unimportant. The procedure is about the participant making an informed choice to proceed with the test and therefore they need to be allowed opportunity for questions. Clearly one possible outcome of applying this principle is that the person involved may choose not to participate.

4. Voluntariness. Professional behavior influences participant involvement. Participants should not be rushed, nor should facilitators fidget while the participant reads the form. Coercion and undue influence should be absent when the person is asked to give their consent to participate in the test. Excessive pressure might come in a number of subtle ways that one needs to be cautious of.

5. Participants' rights. Countries vary as to their recognition of human rights. Even where there is general agreement, definitions of those rights and interpretations of how they apply vary. Participants should have is the right to be informed as to what their rights are. Karat [6] reviewed the codes of ethics of 30 national computer societies and found that they shared several major topic areas. The first on the list addressed the need to respect the rights of people involved with the technology. According to Dumas and Redish [5] the rights most relevant to usability testing include the right to leave the test without penalty, the right to have a break at any time, the right to privacy (such as not having their names used in reporting the results of the test), the right to be informed as to the purpose of the test and the right to know before the test what they will be doing.

7. Confidentiality. Confidentiality is different from the participant's right to privacy; it refers to how data about the participants will be stored. The ACS (2000) code stipulates that it is obligatory for members to preserve the confidentiality of others information. The ACM (2000) code has specific clauses on constraining access to certain types of data, and on organizational leadership to ensure confidentiality obligations are adhered to within organizations. In remote testing this can be extended to electronic data-logging over the internet.

8. Waivers. Permission needs to be obtained from participants to use materials such as questionnaires, audio and video recordings (and their transcripts). In many countries they have the right to refuse to give waivers. Participants should be given the option of having the data used for the purposes of the test, or of also having it used in a wider context. If the latter, then the consent form should state in what further ways the data will be used, so that an informed decision can be taken by the participant. Such permission should state the purposes for which the material will be used.

Informed consent is both a process and a formal record of the process. That formal record is typically a form, but may also be another type of recording, such as video. In our case we used a form that was signed by each participant.

Interviewing old people. We had to acknowledge that we were interviewing old people for this research and that taking part in an interview was a demanding experience for both interviewer and participant. Past research has suggested that such interviewing is both possible and highly rewarding. However, we must take into account that failing eyesight, hearing, mobility, cognitive impairment etc may affect interviews. Interviewers have to be prepared to take their time, repeat themselves, ask for clarifications, listen to repetitions, help them to orientate to the questions with reference points for reminders (e.g. anchoring events to dates). Interviewers should keep the interview very focused and watch out for participant fatigue. In our project, they were told, that it would be better to return at a later date than to conduct a lengthy and tiring interview.

What also needs to be taken into account are issues of power relationships. Older people might see researchers as high status compared to themselves and may easily defer to a researcher's opinions and ideas. Inequalities in power throughout the interview could stem from gender (male interviewer-female participant), age (younger interviewer-older participant), employment status, (employed interviewer-retired older person) and so on. We needed to think carefully about how such issues might have affected each interview and discussed this in our reflexive analysis. In situations where power inequalities were evident it was helpful to keep reassuring the participant that they were the experts and we were there to learn from their experiences. This helped to put them more in control of the interview, just as accepting a cup of tea allowed them to relocate the interview as a social situation which they were familiar with.

2.3 Possible Distortion Effects in Field Trials

In the analysis of the objective effects of the whole system the dependent variables, that is the outcome in terms of well-being and mental fitness are clearly defined by the survey instruments used and the data collected. Yet the definition of the independent input variables is much more complicated as we are not dealing with a laboratory test under controlled conditions but with a real life field test in which almost everything can happen and distort the assumed correlation between the application of the ALADIN components and the output. Besides a general Hawthorne effect, the fact that the situation of observing and surveying the test persons already has an effect on them, has to be taken into account (Adair 1984).

We can identify five sources of possible distortion and interference with the objective ALADIN effects:

Time effects: Three months of testing with alternating algorithms are sensitive to time effect because we cannot exclude that the outcome measured in one period could be a consequence of a factor present in a former period. Only the sequence of adaptive lighting algorithms (genetic and annealing algorithm) could be rotated to control sequence effects. This does not affect the evaluation of the system as a whole but it influences the weighting of the different components and periods.

Learning effects: Habituation and learning are in some way correlated to time effects and mainly affect measurements of mental fitness. Concerning the performance in the activation exercises it is well known that regular training leads to improvements. So we should expect a notable increase in performance even without further supporting factors.

Social support effects: Coaching and technical assistance are a vital part of a field test because we cannot expect the test persons to use the system without any help right from the start. We also know from the requirements analysis that any technology which replaces social contact would be rejected. Social support therefore has to be recognized as one of the main „disturbing" factors. Some control is possible through the personal diaries and the annotations of the assistants.

Particular events: Only in an isolated laboratory and under constant control can we exclude incidental distortions as they often happen in real life. In the analysis of the data we also have to consider meteorological phenomena as well as illness or social problems which we can only partly reconstruct from the collected data.

Way of use: A more interfering than distorting effect is closely correlated to the real life scenario. The test persons were free to use ALADIN as they wanted as long as they fulfilled some minimum criteria. This means that date, frequency and duration of usage of the different ALADIN components vary from test person to test person. To compensate for this, either complex modeling of the independent variables or a restriction to single case analysis are required. We chose the latter option.

3 Results and Conclusions

3.1 Summary of Results from Field Trials

The prototype was tested in twelve single households for a period of three months. At regular intervals, various mental and physical performance tests as well as life and sleep quality questionnaires were administered. We also organized two focus group meetings with our test persons to discuss any issues that might not be included in the questionnaires or log data.

The results indicate an overall increase in both mental and physical fitness. On the whole, test persons enjoyed the exercises and found the system easy to handle. However, they would like a bigger choice both with regard to the exercises and the music tunes that accompany the biofeedback sessions. Quality of life did not seem to improve significantly which may be due to the fact that most test persons started out from very high levels, leaving little space for any notable increase.

The people who might benefit most, e.g. the fragile, home-bound older elderly, were represented rather poorly in the test population. This is largely due to the selection criteria which stipulated that test persons must not suffer from any serous health conditions. The real life field testing of the ALADIN provided us with a vast amount of data concerning the factors of well-being and mental fitness in an ageing population and the potential effects of ambient lighting assistance. Even if the sample was small and therefore statistically not significant we gathered very valuable information for advancing the use of lighting for better ageing and improving well-being.

The conclusions which can be derived from the testing in real-life settings can be roughly divided into:

1. Ideas about how to improve the ALADIN prototype and its different components
2. Ideas about how we can improve the process of testing
3. Ideas about how to "package" AAL technology to enhance user acceptance.

3.2 Possible Improvements of the Prototype

A marketable version of ALADIN is certainly still some years away, but the field test helped us gather many proposals for improvement. Given the open and modular architecture, the prototype and its different components can be easily integrated into general building management systems. Different target groups might be addressed by different functions, modules or components. For example, we have learned that mental fitness training is an activity which appeals to the younger elderly. The automatic lighting adaptation on the other hand helps stabilize the circadian rhythm which is particularly welcome when someone is constrained to such a degree that outdoor activity and thus exposure to sunlight is no longer possible. To reach the fragile elderly, we would have to devise strategies to reach the relevant intermediaries such as health and care professionals, care organizations, building companies and housing associations. When it comes to the light installation we will have to develop a much cheaper, more stylish and less awesome solution. Besides, older people are very concerned about energy consumption which is why in a future redesign we may consider the use of LEDs. Some minor difficulties with usability and compatibility with home entertainment equipment could be avoided if ALADIN were connectable to any TV set as a plug-and-play device. In this case the remote control could be even simpler.

Finally, our results confirmed a basic insight of market research: Products and their features have to give a clearly visible benefit to the customer. In the case of ALADIN it was the automatic light adaptation which lacked this benefit since the adaptation occurred imperceptibly and at a subliminal level. Even if it may be beneficial in the long run, people missed an obvious immediate effect. This may be compensated by adding some features with visible effects such as a light timer to scare off burglars when absent, automatic switch-off as an energy saving measure, automatic switch-on by means of a movement sensor for more comfort, or an automatic navigation help for a safe nocturnal walk to the toilet.

3.3 Lessons Learned Regarding Usability Testing of AAL Applications

As is usually the case with such complex endeavors that consume lots of time and human resources, one would like to have another go to get everything right straight from the start. This is why it is so important to derive lessons learned from one´s experience. Even if you assume that you have prepared the field trials very well, cleared all the ethical issues, organized the logistics and written up detailed instructions, this does not ensure that they will run smoothly. In our case, matters were complicated by having three locations for conducting the field tests. On the whole, people prefer calling on a human person/expert rather than consult the manual or FAQs on the Web. Therefore, a need for central coordination throughout the field trials has emerged as essential for efficient implementation.

Another important lesson learned is that for measuring factors such as wellbeing, sleep quality and attitude to life the field trials should last for a whole year. This would also help neutralize the novelty effect in the beginning. In the focus group discussions, it also became clear that the test persons who used ALADIN in the winter months were overall more positive than the ones who used it in spring. Whilst we were aware of this, the high costs of the lighting system made this impossible. Having three systems run in parallel was the most we could afford.

The need for very precise instructions does not only apply to the test persons but also to the assistants and the people involved in installing the system. Although before the field tests detailed manuals for the organizations and the coaches involved in the field trials had been prepared, discussed and distributed to all the relevant parties, this still proved insufficient. Especially, we underestimated the need for technical assistance for installing the system. It might therefore be advisable to appoint a technical expert to handle the installation in all locations and who can be contacted when a problem arises.

3.4 Packaging Technology to Enhance User Acceptance

The findings especially from the qualitative data strongly suggest that technology can only complement but never replace interaction with humans and ideally should facilitate human contact. It is therefore necessary, to package assistive technology to achieve (higher) user acceptance. With ALADIN, we envisage the following measures:

Social support. Although the great majority of the elderly want to live independently at home for as long as possible, they nevertheless want to be embedded in a social network. In a follow-up project we would like to investigate how this type of support can best be delivered. One of our German partners, for instance, is planning to use SOPHIA, an information and communication platform developed by a housing foundation (Joseph-Stiftung) together with the university and clinic in Bamberg (see www.sophia-tv.de). Via a videophone this ‘virtual nursing home’ connects people to a large variety of health and care as well as social services in the region.

Advice on ageing-friendly housing. In the course of ALADIN we have acquired a great deal of knowledge about older people’s needs and preferences with regard to housing. Due to mobility constraints, many older people spend a large proportion of their time indoors, which makes optimising lighting so essential for their wellbeing. Lack of daylight may cause seasonal depression and sleep disorders due to irregular circadian rhythms. This can be compensated by longer exposure to light during winter times and higher illumination levels in general since with ageing people’s vision tends to deteriorate.

When installing lighting systems in the homes of the elderly, quantity, spectrum, timing, duration and spatial distribution are important characteristics to be considered. In addition, special age-related impairments such as impaired vision have to be taken into account. In the empirical research for ALADIN, the risk of falling down emerged as one of the most common worries among the elderly. The use of lighting for navigational purposes therefore would clearly respond to older people’s needs. A future lighting solution will also have to address safety and security concerns, e.g. as a protection against burglary or theft. Combining assistive lighting with a counseling service on ageing-friendly housing is an avenue we intend to pursue.

Fig. 1. ALADIN system as deployed in the field tests

Acknowledgments. We would like to thank the European Commission for funding this project as well as all our partners and test persons who have contributed.

References

1. Boyce, P., Mcibse, F.: Education: the key to the future of lighting practice. Lighting Research and Technology 38(4), 283–291 (2006)
2. Pollack, M.: Intelligent Technology for an Aging Population. AI Magazine 26(2), 9–24 (2005)
3. Maier, E., Kempter, G.: Increasing psycho-physiological wellbeing by means of an adaptive lighting system. In: Cunningham, P., Cunningham, M. (eds.) Expanding the Knowledge Economy - Issues, Applications, Case Studies and information on ALADIN website, pp. 529–536. IOS Press, Amsterdam (2007)
4. Bernhaupt, R., Obrist, M., Weiss, A., Beck, E., Tscheligi, M.: Trends in the Living Room and Beyond: Results from Ethnographic Studies Using Creative and Playful Probing. In: Cesar, P., Chorianopoulos, K., Jensen, J.F. (eds.) EuroITV 2007. LNCS, vol. 4471, pp. 146–155. Springer, Heidelberg (2007)
5. Dumas, J.S., Redish, J.C.: Handbook of Usability Testing. John Wiley & Sons Inc., USA (1994)
6. Karat, J.: Evolving the scope of user-centered design. Commun. ACM 40, 7 (1997)