AAL JP – project ASSAM

Building a comprehensive framework for the evaluation of mobility assistants for the elderly and impaired in ASSAM

Report R4.2

Roy S. Hessels

Abstract: HKU is the partner in the ASSAM project responsible for designing and managing the test setting. In this paper an overview is given of literature on evaluating projects of this magnitude, and conventions in evaluation metrics. From this we propose how measurements in the ASSAM project will take place, and what metrics to include. This paper serves as a reference and validation for the test interface and its corresponding paper version.

Author: Roy S. Hessels, HKU

Type: Report, Public

Date: 2013-09-03

Version history:

Date	Author	Content
2012-12-11	R.S. Hessels, HKU	First public version
2013-09-03	BKB, DFKI	Slight formatting
2013-11-19	Stahl, DFKI	Proof reading

Table of Contents

1. Introduction

The Assistants for Safe Mobility (ASSAM) project is a tri-national project funded jointly by the European Union and the national funding organizations of Germany, Spain and the Netherlands. The project aims to develop a new generation of modular mobility assistants for the elderly who experience a decline in physical and cognitive functions, whilst focusing particularly on promoting the autonomy of the end-user (Krieg-Brückner, Bothmer, Budelmann, Crombie, Guerín, et al., 2012). The project incorporates the development of several mobility assistants (i.e. walker, electric wheelchair and tricycle), which can all be fitted with several modular components assisting the end-user. These modules provide the end-user with help in case of emergencies, in maintaining balance, avoiding obstacles and finding directions, as well as allowing an easy transition between indoors and outdoors (Krieg-Brückner, Crombie, Gersdorf, Jüptner, Lawo, et al., 2012).

Evaluating the success of an undertaking with the scope of the ASSAM project poses a variety of obstacles to be overcome. The end-users involved in testing the mobility assistants come from three different countries, namely Germany, Spain and the Netherlands. This in itself already poses two distinct issues. Firstly, the end-users organizations in each of the three countries all possess different clienteles with distinct demographics on age, physical and mental disabilities, as well as the use of mobility assistants. Secondly, the evaluation of the newly developed mobility assistants will have to take place at three different locations across Europe. Furthermore, the ASSAM project will go through different iterations of testing, in which the setup of the modular devices will vary. In order for these obstacles to be overcome, an evaluation framework will have to be developed that is capable of testing at various locations, with varying user demographics, sensitive to all possible variations of the devices being tested.

ASSAM is not the first research & development project on autonomous mobility assistants (European Commission; LoPresti, Sharma, Simpson, & Mostowy, 2011; Röfer, Mandel, Lankenau, Gersdorf, & Frese, 2009; S. P. Levine, Bell, Jaros, Simpson, Koren, et al., 1999). Whilst providing valuable insights, the predecessors to the ASSAM project do not provide a comprehensive evaluation framework applicable to the broader scope of the ASSAM project. In this paper previous research on mobility metrics as well as research on user-centered design and usability evaluation methods will be evaluated and put in perspective with the ASSAM project. Furthermore, this paper will present the process that the Utrecht School of the Arts (HKU), responsible for developing the testing and evaluation in the ASSAM project, has gone through in developing a comprehensive evaluation framework capable of overcoming the obstacles that go with the tri-national research & development project that ASSAM is.

2. User-Centred Design

In the ASSAM project a user-centred (also known as participatory) design methodology will be applied to ensure participation of the end-users (Krieg-Brückner, Bothmer, et al., 2012). User-centred design originated in the 1980's and was popularized to the public with the publishing of Donald Norman's book *The Psychology of Everyday Things* (Norman, 1988). It is a broad concept, with the underlying idea that end-users are involved in the design process, and how the product eventually takes shape (Abras, Maloney-Krichmar, & Preece, 2004). It has been ~~more~~ further formalized by the International Organization of Standardization, originally in ISO 13407, but updated in 2010 to the new standard ISO 9241-210 (International Organization of Standardization, 2010).

In the ASSAM project all development decisions will be made with the goals and needs of end-users in mind. Consequently, the devices will be evaluated whether these goals and needs are met or not. Moreover, the evaluation of the devices in functional usability field trials, and prolonged use of the devices in everyday usability field trials, will partly determine whether the ASSAM project is successful or not (Krieg-Brückner, Bothmer, et al., 2012). For the evaluation to take place we need to be aware of what the end-users' goals and needs are, or the user requirements, and what situations they might face with the mobility assistants, or the use cases.

2.1 User requirements

To distil the user requirements from the tri-national clienteles, the ASSAM project ~~made~~ makes use of persona creation. Personas, a concept introduced by Alan Cooper in the 1990's (Cooper, 1999), are archetypal characters that describe behavioural patterns and goals. These personas use storytelling to provide valuable information and insights on what the user might need and what his or her goals are. While not being representations of just one user, personas represent specific groups of possible users, so that the body of personas is as extensive as needed be to describe all relevant behavioural patterns.

Personas for the ASSAM project were created by the representatives of each of the end-user organizations in Germany, Spain and the Netherlands. They were instructed by HKU to provide a set of personas representing their clientele based on their experiences with clients and their caregivers' information. All personas from the three end-user organizations were combined for a total of 11 personas. These were analysed for overlapping behavioural patterns, goals and particular mobility device needed, and combined into a total of 5 personas. These personas will provide the basis for the user requirements of the mobility assistants.

2.2 Use cases

To build a successful line of mobility assistants not only the needs and goals of the users are helpful, information on the situations these mobility assistants may encounter are also necessary. There are several different methods to gather this information, and although they might hold similar information, the means of gathering the information vary greatly. The first method is called 'use cases', which is somewhat deceptive as all the methods provide cases in which a device will be used. The 'use cases' method describes all interactions between an actor and a system, mostly following a specific set of conventions for doing so (Goodwin, 2009). The second method is called 'user stories', which comprises a few sentences of what an actor may encounter during the interaction with a system. Both of these methods do not incorporate the attitudes and feelings of the actor, something of great importance in user-centered design. There is, however, a method that does incorporate this, namely 'scenarios'. *"A scenario is a plausible description of the future based on a coherent set of assumptions"* (Goodwin, 2009). In a scenario the behavioural patterns of an archetype, in our case the above described personas, are described in relation to the use of a future solution.

As with the personas, the scenarios were provided by the end-user organizations based on the possible situations their clientele might face, totalling a number of 9. These scenarios were again analysed for overlapping patterns and distilled into a set of 5 scenarios, one for each persona using a mobility assistant in context. These scenarios provide the basis for the use cases; the situations a mobility assistant might be used in (i.e. not the 'use cases' method described above).

2.3 Usability evaluation

Now that the user requirements and use cases are known, the next step is to translate these into evaluation criteria. There is, however, no ~~one~~ single established method for doing so (Cresswell & Miller, 2000), and it is mostly done by ~~doing~~ qualitative analysis in the manner the evaluator sees ~~fit~~fitting. On the other hand, there are a few well established methods for usability evaluation (John & Kieras, 1996; Nielsen & Mack, 1994), though they do not apply to the analysis of personas and scenarios. They are particularly researched in terms of their predictive power and use without consulting end-users (Jacobson, 1999; John & Marks, 1997). Interestingly enough, these methods also suffer from inter-evaluator inconsistencies (Hertzum & Jacobson, 2003), which is a major criticism for doing qualitative analysis in user requirements analysis. Using personas and scenarios does, however, provide a rich basis of information and context to fall back on when design decisions have to be made. This much resembles the ‘thick description’ described by Clifford Geertz (1973).

While the established usability evaluation methods are not specifically applicable to the analysis of personas and scenarios, there are some aspects which can be applied in a similar manner. The usability evaluation method GOMS (Goals, Operators, Methods & Selection rules) is a method developed in the 1980’s (Card, Moran, & Newell, 1983). This method describes the interaction between users and an interface in terms of a human information processor. The goals are what a user wants to accomplish and operators are actions to accomplish this goal. A method is a group of different operators, and if different methods lead to the same goal, selection rules determine which method will be used. Lets consider this method in respect to the evaluation of user requirements and use cases. Say one of the use cases of the mobility assistants is that one can act when one gets lost. The user requirements in this case might be that there is a route home provided by the device, or that one can contact home. These translate directly into goals in the GOMS method: getting home, or contacting help. Now the operators are the steps the end-user takes in order to accomplish this task (i.e. turning on the navigation, selecting a route, choosing home as the destination). When the goal is not accomplished, the operators can be useful in describing where and when something went wrong. By using an analogy of the GOMS model, we can look at goals at different levels. A high-level goal for the end-user might be safety while travelling with their mobility assistant, whereas a low-level goal would be to maintain balance. Whenever a high-level goal is accomplished, the assumption can be made that the low-level goals required for this particular high-level goal were also met. If not, the approach would be to zoom in on the low-level goal with its corresponding methods and operators. By considering the user requirements and use cases as goals to be accomplished, and the failure of achieving a goal as a result of a selection of methods and the underlying operators, a standardized set of questions can be administered whenever this occurs.

Of course, there might be more to a failure in achieving a goal than a set of operators; there might be an environmental problem. In terms of the problem described above, an environmental issue in navigating home might be the lack of a GPS signal in the subway, which would lead to a failure to accomplish a goal regardless of the operators of the end-user. In the ASSAM setup, supervisors of test sessions are required to provide rich material (i.e. photo, video), which will suffice in pinpointing such problems.

3. Evaluation metrics

Designing for the elderly and disabled is a topic that has been given much attention over the past years. Research shows that development for the elderly and disabled takes quite some careful thought (Arning & Ziefle, 2009; Newell & Gregor, 2002), particularly on the development of interfaces (Holzinger, 2002; Holzinger, Searle, & Nischelwitzer, 2007; Kleinberger, Becker, Ras, Holzinger, & Müller, 2007). It would seem logical to get a comprehensive profile from each end-user participating in the testing and evaluation. Whereas the personas and scenarios describe the general needs, goals and attitudes of the end-users to be met in the design-process, this is not necessarily a portrayal of the users that are involved in the testing and evaluation process. The data from the testing sessions is highly dependent on the end-user in question. So in order to put the data into context we need to collect individual data on each participating in the testing and evaluation.

Urdiales and colleagues present an overview of the metrics used in performance evaluation of assisted wheelchair navigation, though most of these metrics can also be applied to other mobility assistants (Urdiales, Peula, Cortés, Barrué, Fernández-Espejo, et al., 2009). As not all metrics are applicable in the ASSAM test setting, the specific metrics used and their consideration will be mentioned.

3.1 User metrics

User metrics refer to the demographic profile of the end-user and all questionnaires where the user is central, not the interaction with the mobility assistant. This will be gathered prior to each testing session. First of all, the demographic profile including age, gender, level of education and location of residence will be collected. This will also include info on the mobility assistants and/or other assistive technology already in use.

Next, more specific information on mobility, activities of daily living, the nature of the end-user's disabilities as well as the intensity of the disabilities will be collected. There are several options to accomplish this. To establish a measure of mobility and independence in the Barthel Index of Activities of Daily Living will be administered (Mahoney & Barthel, 1965). This is a short questionnaire comprised of 10 questions, yielding a score range of 0 (completely dependent) to 20 (completely independent) (Collin, Wade, Davies, & Horne, 1988). For a measure of cognitive impairment or dementia there are also questionnaires available, such as the Mini-Mental State Examination (MMSE). However, as all ~~the~~ end-users involved in the testing and evaluation process are clients of one of the end-user organizations, it is quite possible that more extensive data on this is already present. This includes any information on intelligence scores of end-users. As the MMSE requires extra time during the test sessions, and all participants are deemed to be able to copy and view objects (something that is not the case for at least part of the clientele), we will rely on information provided by the end-user organizations.

Furthermore, the attitudes that the end-users ~~possess~~ have towards technology are something to consider also. This might well influence whether one accepts a new mobility assistant or not. Especially for the elderly, the attitudes of the end-user towards technology are a major reason for the rejection of technology assistants (Arning & Ziefle, 2009; Wilkowska & Ziefle, 2009). The elderly generally ~~possess~~ have less computer expertise and technical confidence, which correlate positively with perceived ease of use and perceived usefulness of assistive technology (Wilkowska & Ziefle, 2009). Perceived ease of use and perceived usefulness in turn are key components in technology acceptance (Davis, 1989; Venkatesh, Morris, Davis, & Davis, 2003). Getting the elderly to accept technology thus requires some extra effort on ease of use and usefulness of assistive technology, but also requires some form of measurement of attitudes towards technology during the evaluation. While there are questionnaires available for uncovering attitudes towards technology (such as the Matching Person & Technology scale or Psychosocial Impact of Assistive Devices Scale), implementing these in the ASSAM test sessions will greatly affect the temporal load on the end-users. We therefore will rely on the end-user organizations for this information and explicitly inquire about these attitudes, allowing more time for testing multiple scenarios.

3.2 Interaction metrics

Interaction metrics refer to all questionnaires and data about the interaction between the end-user and the mobility assistant. This is partly covered in the analysis of the user requirements and use cases; if a (sub-) goal is not achieved information on where it went wrong is collected. This in turn means there is a measure of interaction error, i.e. the number of times there was a conflict between the end-user and the mobility assistant.

Another, often used, interaction metric is the NASA Task-Load Index, developed over 20 years ago (NASA, 1986). The NASA-TLX is a subjective measure of mental workload, not uncommon in wheelchair studies (Sharma, Simpson, LoPresti, & Schmeler, 2012). The NASA-TLX consists of several subscales, such as physical effort, mental & sensory effort and frustration, which combine into a weighted scale of workload. The NASA-TLX will be administered in the ASSAM testing and evaluation as well, and hopefully show a decline in workload as the devices are longer in use, and also as functionality increases.

4. Integration into ASSAM test setting

With the details described of *what* is going to be tested ~~described~~, the question that now arises is *how* the testing and evaluation is going to take place. Due to the tri-national setup of the ASSAM project, it is impossible for one team to test and collect all the necessary data. Inevitable is the fact that testing will be done at multiple locations throughout Europe, and that the supervisors who will do the testing might not be the same every day. Furthermore, the sheer amount of expected data poses an interesting challenge for the manner of collecting data. In short, the system must be simple enough to train the supervisors over the internet (preferably by written text), location independent, and find a solution for the sheer amount of data while maintaining a secure handling of personal information.

The solution the HKU proposes is to develop an interface operable from a mobile touch screen device, such as a tablet. This interface will systematically guide the tester through gathering the demographic profile, specific configuration being tested, answers to questionnaires, and of course testing the user requirements. This system will securely store the data on the device, and securely transfer the data to the database when it's connected to a network. Again, the data will be securely stored, only accessible by the ASSAM-partners. From here the data can easily be used by the technical partners for the iterative process of development, as well as analyzed for trends in the quantitative data.

5. References

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-Centered Design. In W. Bainbridge (Ed.), *Encyclopedia of Human-Computer Interaction*: Thousand Oaks: Sage Publications.

Arning, K., & Ziefle, M. (2009). *Different Perspectives on Technology Acceptance: The Role of Technology Type and Age*. Paper presented at the Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society.

Card, S., Moran, T. P., & Newell, A. (1983). *The Psychology of Human Computer Interaction*: Lawrence Erlbaum Associates.

Collin, C., Wade, D. T., Davies, S., & Horne, V. (1988). The Barthel ADL Index: a reliability study. *Int Disabil Stud, 10*(2), 61-63.

Cooper, A. (1999). *The Inmates are Running the Asylum*: SAMS.

Cresswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative research. *Theory into practice, 39*(3).

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*(13), 319-337.

European Commission. SHARE-IT : Supported human autonomy for recovery and enhancement of cognitive and motor abilities using information technologies. Retrieved 10-10-2012, from http://cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=9076822

Geertz, C. (1973). *Thick Description: Toward an Interpretive Theory of Culture*.

Goodwin, K. (2009). *Designing for the digitale age: how to create human-centered products and services*. Indianapolis, Indiana: Wiley Publishing.

Hertzum, M., & Jacobson, N. E. (2003). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. *International Journal of Human-computer Interaction, 15*(1), 183-204.

Holzinger, A. (2002). User-Centered Interface Design for Disabled and Elderly People: First Experiences with Designing a Patient Communication System (PACOSY). In K. Miesenberger, J. Klaus & W. Zagler (Eds.), *Computers Helping People with Special Needs* (Vol. 2398, pp. 467-484). Heidelberg: Springer Berlin.

Holzinger, A., Searle, G., & Nischelwitzer, A. (2007). *On Some Aspects of Improving Mobile Applications for the Elderly.* Paper presented at the Universal access in human computer interaction: coping with diversity.

International Organization of Standardization. (2010). ISO 9241-210:2010, *Ergonomics of human-system interaction -- Part 210: Human-centred design for interactive systems*.

Jacobson, N. E. (1999). *Usability Evaluation Methods. The Reliability and Usage of Cognitive Walkthrough and Usability Test.* University of Copenhagen, Copenhagen.

John, B. E., & Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: Which technique? *ACM Transactions on Computer - Human Interaction*.

John, B. E., & Marks, S. J. (1997). Tracking the effectiveness of usability evaluation methods. *Behaviour & Information Technology*.

Kleinberger, T., Becker, M., Ras, E., Holzinger, A., & Müller, P. (2007). *Ambient Intelligence in Assisted Living: Enable Elderly People to Handle Future Interfaces.* Paper presented at the Universal access in human-computer interaction: ambient interaction.

Krieg-Brückner, B., Bothmer, H., Budelmann, C., Crombie, D., Guerín, A., Heindorf, A., Lifante, J., Martínez, A. B., Millet, S., & Velleman, E. (2012). *Assistance for Safe Mobility: the ASSAM Project.* Paper presented at the AAL-Forum 2012, Eindhoven.

Krieg-Brückner, B., Crombie, D., Gersdorf, B., Jüptner, A., Lawo, M., Mandel, C., Martínez, A. B., Röfer, T., & Stahl, C. (2012). *Challenges for Indoor and Outdoor Mobility Assistance.* Paper presented at the AAL Kongress, Berlin.

LoPresti, E. F., Sharma, V., Simpson, R. C., & Mostowy, L. C. (2011). Performance testing of collision-avoidance system for power wheelchairs. *J Rehabil Res Dev, 48*(5), 529-544.

Mahoney, F., & Barthel, D. (1965). Functional evaluation: the barthel index. *Maryland State Medical Journal, 14*, 56-61.

NASA. (1986). *Nasa Task Load Index (TLX) v. 1.0 Manual.*

Newell, A. F., & Gregor, P. (2002). Design for older and disabled people – where do we go from here? *Universal Access in the Information Society, 2*(1), 3-7.

Nielsen, J., & Mack, R. L. (1994). *Usability Inspection Methods*. New York: John Wiley.

Norman, D. (1988). *The design of everyday things*. New York: Doubleday.

Röfer, T., Mandel, C., Lankenau, A., Gersdorf, B., & Frese, U. (2009). 15 Years of Rolland. In B. Hoffmann, C. Lüth, T. Mossakowski & L. Schröder (Eds.), *Festschrift Dedicated to Bernd Krieg-Brückner on the Occasion of his 60th Birthday* (pp. 255-272): Sichere Kognitive Systeme, DFKI Bremen.

S. P. Levine, Bell, D. A., Jaros, L. A., Simpson, R. C., Koren, Y., & Borenstein, J. (1999). The NavChair Assistive Wheelchair Navigation System. *IEEE Trans Rehablil Eng, 7*(4), 443-451.

Sharma, V., Simpson, R. C., LoPresti, E. F., & Schmeler, M. (2012). Clinical evaluation of semiautonomous smart wheelchair architecture (Drive-Safe System) with visually impaired individuals. *Journal of Rehabilitation Research & Development, 49*(1), 35-50.

Urdiales, C., Peula, J. M., Cortés, U., Barrué, C., Fernández-Espejo, B., Annichiaricco, R., Sandoval, F., & Caltagirone, C. (2009). *A Metrics Review for Performance Evaluation on Assisted Wheelchair Navigation*. Paper presented at the 10th International Work-Conference on Artificial Neural Networks

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*(27).

Wilkowska, W., & Ziefle, M. (2009). *Which Factors Form Older Adults' Acceptance of Mobile Information and Communication Technologies?* Paper presented at the Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society.