An Approach to and Evaluations of Assisted Living Systems Using Ambient Intelligence for Emergency Monitoring and Prevention

Thomas Kleinberger[1], Andreas Jedlitschka[1],
Holger Storf[1], Silke Steinbach-Nordmann[1], and Stephan Prueckner[2]

[1] Fraunhofer Institute Experimental Software Engineering
Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
thomas.kleinberger@iese.fraunhofer.de,
andreas.jedlitschka@iese.fraunhofer.de,
holger.storf@iese.fraunhofer.de,
silke.steinbach-nordmann@iese.fraunhofer.de
[2] Department of Anaesthesiology and Emergency Medicine
Westpfalz-Klinikum GmbH, Hellmut-Hartert-Strasse 1, 67655 Kaiserslautern, Germany
stephan.prueckner@med.uni-muenchen.de

Abstract. Ambient Assisted Living (AAL) is currently one of the important research and development areas, where software engineering aspects play a significant role. The goal of AAL solutions is to apply ambient intelligence technologies to enable people with specific needs to continue to live in their preferred environments. This paper presents an approach and several evaluations for emergency monitoring applications. Experiments in a laboratory setting were performed to evaluate the accuracy of recognizing Activities of Daily Living (ADL). The results show that it is possible to detect ADLs with an accuracy of 92% on average. Hence, we conclude that it is possible to support elderly people in staying longer in their homes by autonomously detecting emergencies on the basis of ADL recognition.

Keywords: Ambient Assisted Living, Emergency Monitoring, Experiments.

1 Introduction

In most industrialized countries, demographical, structural, and societal trends tend towards an increase in the number of elderly people and single households, which has dramatic effects on public and private health care, emergency medical services, and the individuals themselves. The increasing average age of the total population, and the resulting rise in chronic diseases, will result in a dramatic increase in emergency situations and missions within the coming years. The practical problems arising in preclinical emergency medical treatment and the central role of context factors, such as social isolation, reduced mental capabilities, and the resulting need for help, were examined in an epidemiological study conducted in the framework of the EC-funded emergency monitoring and prevention project EMERGE [1]. Results show that

C. Stephanidis (Ed.): Universal Access in HCI, Part II, HCII 2009, LNCS 5615, pp. 199–208, 2009.

approx. 44% of Emergency Medical Services' (EMS) system resources are dedicated to patients older than 70 years of age.

In the long term, the consequence will be higher costs for such services or a decrease in service quality, or both. Assisted living solutions for elderly people using ambient intelligence technology can help to cope with this trend, by providing proactive and situation-aware assistance to sustain the autonomy of the elderly and by helping to limit cost increase, while concurrently providing advantages for the affected people by enhancing their quality of life. The goal is to enable elderly people to live longer in their preferred environments, to enhance the quality of their lives, and to reduce the costs for society and public health systems.

Today's commercially available products for emergency monitoring already use a broad range of modern technology, such as *Personal Emergency Response Systems* (PERS) (e.g., necklaces with emergency buttons, fall sensors in mobile phones with wireless notification of emergency services, vital data sensor tapes, etc.). However, they are mostly closed, stand-alone systems with a limited ability to describe the actual situation, often just too difficult for the elderly people to handle, and useless in emergencies. During our study in EMERGE, we observed that only 3% of the affected people in emergency situations over 65 years have a PERS at hand and, more importantly, the PERS was used only in 1.3% of the cases to report the incident. Reasons for this may include forgetting to wear the PERS or being unable to use it at all after an incident has happened. Still, assisted living systems promise a huge potential for disabled and elderly people.

In this paper, we present an approach as well as evaluation results in a near real-life environment (Ambient Assisted Living Apartment) for emergency monitoring applications and corresponding user interaction components resulting from two research projects: EMERGE and BelAmI [2]. The impact of the developments in both projects has been measured from several perspectives:

- the end users' point of view (elderly persons and their relatives),
- the professional point of view (physicians/medical caregivers),
- the research point of view (software and system engineers).

An overview of currently available assistive technology in elderly care and currently available products can be found in Miskelly [3], in a recent study on European Remote Patient Monitoring Markets by Frost & Sullivan [4], and in a study for the European Commission on ICT for independent living for elderly people [5]. Dishman [6] provides an introduction to indoor electronic home care systems for elderly people and their potential for aging societies. While the traditional geriatric approach usually consists of manually acquiring ADLs [7] via care personnel or requesting it from the individuals through questionnaires and use it to assess appropriate assistance/care programs for the respective persons [8] [9] [10], new approaches aim at automatic emergency detection/notification based on *Automatic Behavioral Monitoring Systems* (ABMS) [11]. ABMS monitor the environment and not the individuals themselves. They combine technical features of Smart Homes (with automation technology) and telemedicine (especially physiological process telemetry) [12] [13]. However, the most prominent state-of-the-art products available nowadays are mobile phones for elderly people that combine emergency call functions and user location tracking (e.g., Vitaphone [14]), call systems (e.g., Hausnotruf of Deutsches Rotes Kreuz [15]), or general activity monitoring devices (e.g., IST Vivago [16] or Sophia [17]). What is common to these currently commercially available solutions is that they fit into the affordable cost budget of an individual (typically with monthly fees below 50 € for equipment and services). Another commonality is that none of them makes use of behavior-based monitoring with the monitoring of ADLs.

The paper is structured as follows: The approach taken in EMERGE for recognizing ADLs and specific situations is briefly described in section 2. Afterwards, three evaluations for detecting ADLs are described and the results are explained in section 3. Section 4 then provides an interpretation of our initial evaluation results and threats to validity. Finally, we draw conclusions in section 5.

2 Description of the EMERGE System

The main goal of EMERGE is to support elderly people with emergency monitoring and prevention by using ambient, unobtrusive sensors and reasoning about arising emergency situations. To reach this goal, we aim to:

- Track the daily routines of elderly persons to explore behavior and activity patterns, and to detect deviations as early indicators for emergency situations.
- Anticipate emergency situations and try to prevent them by providing early assistance to the elderly person.
- Proactively assist in emergency cases by providing helpful instructions, communication means, and information about the current state of the patient.

The approach is to reason on unobtrusive and non-invasive sensor data. The solution is a combination of sensors, software components, IT platforms, and expert systems for situation recognition and decisions about the appropriate assistance [18].

An integrated approach for the assessment of the overall functional health status of a person needs to be taken. In the *Human Capability Model* (HCM), a functional description of several vital, mental, and psycho-social parameters over time is given [19]. Dependent on the reduced array of information that is available from a sensor environment, the process of fusing sensor data into a meaningful set of information will therefore often not result in a complete medical diagnosis, but will rather describe a situation as being critical or abnormal [20]. It was a key issue for the HCM to define a “normality model” for all these parameters and situations for each specific user. The model also defines end user communication and interaction facilities to allow the assessment of the situations severity and the adequate response. It focuses on two main areas for describing the functional health status: ADLs and vital parameters.

ADLs are one part of the functional health status and can be understood as single activities or a series of acts. Besides others, they are influenced by habits, by the time of the year, the day of the week, by culture, and by age. Activities like sleeping, going to the toilet, preparing meals, and general mobility have shown to be sensitive to indicating changes in the behavior [21] with relation to potentially arising emergencies (cf. fig. 1). ADLs follow certain individual stereotypes and the recognition of deviations or inconsistencies should allow drawing conclusions about a person’s health status.

Basic vital data concerning circulation, respiration, and consciousness, which can be further broken down into single parameters, such as pulse, heart rate, respiratory rate, oxygen saturation, blood pressure, etc. are the focus of telemedical applications. Up to now, this has usually been related to a more or less obtrusive use of sensors, power supply and wiring on the body. We limit vital data monitoring to the use of a wrist device, which provides continuous pulse measurement and discontinuous measurement of arterial blood pressure and body weight by stand-alone devices.

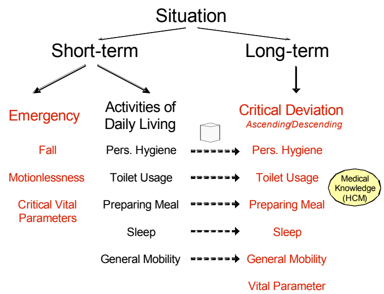

Fig. 1. Activities and situations

3 Evaluation of the ADL Detection

For planning the evaluations, we used the well-established GQM approach (Goal-Question-Metric) from the software engineering domain [22] and applied it to Assisted Living solutions.

In order to consider the stakeholder interests adequately, all stakeholders were already integrated into the requirements analysis and development process for the prototypes [1]. They participated in the requirements elicitation, the validation of the underlying theoretical models for the application, as well as in the tests and evaluations. The evaluation program aims at:

1. proving the validity of our application model,
2. enabling the developers to decide on how to optimize the solution, and
3. enabling end users/stakeholders to draw conclusions about the benefits (as well as the limitations) of adopting the solution in real life.

3.1 Hypotheses and Variables

The evaluation goal according to the GQM goal template was to analyze the *Activity Recognition Component* (ARC) for the purpose of evaluation with respect to its accuracy from the viewpoint of the researcher in the context of a controlled experiment in the assisted living apartment at the Fraunhofer IESE. For this purpose, we defined 3 hypotheses for the ADLs *toilet usage, personal hygiene*, and *preparation of meals*:

- H1: The accuracy of the ARC regarding the ADL *toilet usage* is higher than 80 % of all the performed scenarios.
- H2: The accuracy of the ARC regarding the ADL *personal hygiene* is higher than 80 % of all the performed scenarios.
- H3: The accuracy of the ARC regarding the ADL *preparation of meals* is higher than 80 % of all the performed scenarios.

To fix the expectations for the hypotheses, we held discussions with experts from medicine and geriatrics about what level of accuracy would be necessary to still be able to draw conclusions about long-term behavior deviations.

The ARC analyzes on-the-fly the continuous sensor data stream for characteristic patterns of ADLs. The three mentioned ADLs are the core activities for an ongoing long-term behavior monitoring. The accuracy of detecting these ADLs therefore is a very important indicator for proving the correct functionality of the ARC and a good quality parameter for long-term behavior deviation detection afterwards.

The independent variable in our experiment is the ARC under evaluation, respectively the rules implementing the activity model. The dependent variable is the accuracy of the ARC regarding the detection of the situation prescribed in a given scenario. The accuracy will be calculated by using the following formula:

$$\text{Accuracy} = \frac{\text{Number of all correctly recognized scenarios}}{\text{Number of all scenarios}} = \frac{\text{correct-positive} + \text{correct-negative}}{\text{overall number}}$$

3.2 Samples, Materials, and Environment

In a pre-test with an elderly person, we did not recognize any significant differences compared to a younger person in performing the evaluation task from a technological point of view. However, there might be time differences, which are not in the focus of this evaluation. Hence, for the recognition of ADLs, it does not matter whether the acting persons correspond to the addressed group. For playing out the defined scenarios, unbiased test persons were chosen on a voluntary basis. Their age was in the mid-twenties.

To define realistic scenarios, we followed a commonsense approach that was verified via a pilot study and an expert workshop. In the pilot study, a test person representing a part of the end user group (elderly woman, 73 years old, living independently in her apartment) performed the scenarios. Her activities were recorded and documented in a log. The results of the pilot study were compared with the initial scenarios. In a follow-up expert workshop, medical experts and social scientists defined alternative variations of the scenarios. To avoid an inherent bias, the scenarios were defined by experts who did not know the activity model. Scenario sets included positive and negative scenarios (e.g., *toilet usage / no toilet usage*). Some of the negative scenarios were intentionally very similar to positive scenarios. To assure statistical significance, we aimed to obtain a set of more than 50 scenarios. For that purpose, the positive and negative scenarios were multiplied without affecting the numerical proportion. Table 1 illustrates the final number of scenarios.

The set of scenarios was used to create scenario cards containing short scripts. These scripts describe the respective ADL alternative via step-by-step sub-activities. The scenario cards were applied during the experiment to assist the test persons performing the ADL. In the assisted living apartment, a wide range of sensors are used for detecting situations of interest. Figures 2 and 3 illustrate the individual sensor setup for the evaluation of the ADL detection.

Table 1. Overview of the scenario setting

ADL	Number of scenarios	Number of pos. scenarios	Number of neg. scenarios	Number of test persons
Toilet Usage	100	50	50	6
Personal Hygiene	54	33	21	5
Preparation of Meals	66	45	21	4

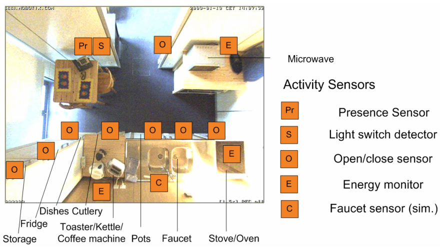

Fig. 2. Setup for evaluation of ADL *preparation of meals*

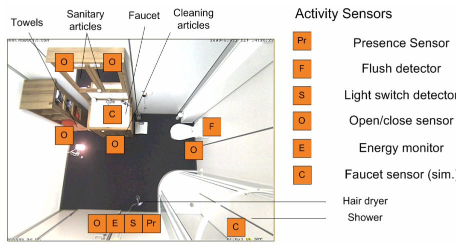

Fig. 3. Setup for evaluation of ADL *personal hygiene*

3.4 Experimental Design and Procedure

A controlled experimental design was chosen. The scenarios were to be performed by the test person(s) and the results, given by the ARC (experimental), were to be compared to the real facts as written on the scenario card (control).

The test persons received an introduction to the environment, i.e., they were told where they could find certain things that they would need to perform a specific task. All relevant sensors located in the areas of the evaluation were active together with their software components. The software system with services for collecting the sensor information and the ARC itself with the timeline GUI were started. To avoid faulty sensor information that might impact the recognition result, no person other than the test person was allowed to stay inside the apartment during the test phase.

The test person randomly picked one scenario card with the description of the scenario from a bowl. Then the test person performed the selected scenario by executing the steps written on the scenario card. Afterwards the scenario ID and the current time were noted and the card was removed from the scenario pool.

The sensor data was collected during the execution of the scenarios and stored by a recorder integrated into the software. In addition, we filmed the actors with a video camera while they performed the scenarios. This enabled us to assess the quality of the collected data, especially in “problematic” cases. The scenario ID, the result delivered by the ARC (ADL detected/not detected), and the time (start-time, end-time, duration) were documented in an Excel sheet.

3.5 Analysis

After performing and documenting the scenarios as described, the Excel sheet was extended with the type of the respective (real) scenario and the result (cf. figure 4).

EMERGE WP 8 Evaluation Date: 04.12.2008 Author: Silke Steinbach-Nordmann Slide 1
ADL Detection: Personal Hygiene & 12.01.2009

Count	Start-Time	End-Time	Duration	Scen. ID	Activity Aggregator Result	Real Scenario	Result
1	15:50:01	15:50:30	00:00:29	6	NPH	NPH	correct
2	16:03:10	16:04:00	00:00:50	24	NPH	NPH	correct
3	16:06:00	16:06:55	00:00:55	27	NPH	NPH	correct
4	16:08:20	16:09:30	00:01:10	41	PH	PH	correct
5	16:09:35	16:11:40	00:02:05	38	PH	PH	correct
6	16:13:50	16:14:20	00:00:30	11	NPH	NPH	correct
7	16:14:50	16:15:40	00:00:50	18	NPH	PH	not correct
8	16:16:10	16:17:40	00:01:30	51	NPH	PH	not correct
9	16:19:00	16:20:35	00:01:35	39	PH	PH	correct

Fig. 4. Extract of result-sheet

The results of all scenarios were transformed into confusion matrices. The accuracy was calculated using the formula in Section 3.1. Subsequently an χ^2-test was used for each ADL for verifying the independence of the classifier (p-level: 0.05).

Table 2. Evaluation results (confusion matrix) of ADLs

ARC result \ Scenario	*Toilet Usage*	*No Toilet Usage*	*n*	χ^2 test
Toilet Usage	50	2	100	p=0.00
No Toilet Usage	0	48		
ARC result \ Scenario	*Personal Hygiene*	*No Personal Hygiene*	*n*	χ^2 test
Personal Hygiene	25	1	54	p=0.00
No Personal Hygiene	8	20		
ARC result \ Scenario	*Preparation of Meals*	*No Preparation of Meals*	*n*	χ^2 test
Preparation of Meal	43	2	66	p=0.00
No Preparation of Meal	2	19		

The results for the ADL *toilet usage* are shown in table 2. We got 50 true positives (i.e., the ARC detected toilet usage and the scenario was toilet usage, too), 48 true negatives (i.e., the ARC detected no toilet usage and the scenario was no toilet usage, too), and two false positives (i.e., the ARC detected toilet usage but the scenario was not toilet usage). The formula for accuracy yields 98/100=98%. Interpreting the result of the test and the achieved accuracy, we accepted hypothesis H1.

The results for the ADL *personal hygiene* are shown in table 2. We got 25 true positives, 20 true negatives, one false positive, and eight false negatives. The formula for accuracy yields 45/54=83.3%. Interpreting the result of the χ^2 test and the achieved accuracy, we accepted hypothesis H2.

The results for the ADL *preparation of meals* are shown in table 2. We got 43 true positives, 19 true negatives, two false positives, and two false negatives. The formula for accuracy yields 62/66=94%. Interpreting the result of the χ^2 test and the achieved accuracy, we accepted hypothesis H3.

4 Interpretation

The analysis confirmed our hypotheses. The achieved results were much better than formulated in our hypotheses. Hence, it was shown that the environment of the assisted living apartment together with the ARC enables accurate detection of ADLs.

However, some problems were discovered, which are shown in table 3. The problematic tasks, i.e., false positive and false negative scenarios, are listed. The individual cause of failure could be established in retrospect by analyzing the video data and logged sensor events. The causes are presented in the last column of the table. Especially for personal hygiene, the accuracy could have been improved by increasing the common understanding of the definition of the ADLs. The people who defined the scenarios included washing their hands as part of the ADL *personal hygiene*, whereas the people who developed the activity model did not consider washing hands as part of this ADL.

Table 3. Problematic tasks

#	Scenario Type	Content	Cause of Failure
2	*No Toilet Usage*	Clean the toilet	Sensor sequence similar to positive scenario
3	*Personal Hygiene*	Shave	Malfunction of electric power sensor
3	*Personal Hygiene*	Wash hands	Not sufficient for Personal Hygiene from modeling point of view
1	*Personal Hygiene*	Brush teeth	Malfunction of sensor
1	*Personal Hygiene*	Have a shower	Wrong performance by test person (“too fast”)
1	*No Personal Hygiene*	Take cleaning cloth	Wrong performance by test person (“searching”)
2	*No Preparation of Meals*	Tidy up the kitchen	Sensor sequence similar to positive scenario
2	*Preparation of Meals*	Prepare frozen pizza	Performing without using dishes or cutlery

Threats that might have an impact on the validity of the results are discussed in the following.

Construct validity: Mono-operation bias, monomethod bias, and confounding constructs. We used several different scenarios for evaluating the accuracy of the ARC. Furthermore, we used video analysis to verify whether the tasks were performed in the “correct” way. Except for the laboratory setting, the tasks did not require any special knowledge or experience other than what any adult would have.

Internal validity: Instrumentation: The scenario cards were carefully designed and proof-read by several people for ambiguity of the task descriptions. Furthermore, we used video to check whether the tasks were performed according to the descriptions. Maturation: Due to the variation of the scenarios and the easy tasks (daily activities), we do not expect any learning effects. The experiments did not take longer than three minutes. Therefore, participants did not feel bored or got tired. Although they got an introduction to the environment, the participants might not have been familiar with the environment. However, because we did not measure their performance, this is not an issue. Selection: We used volunteers for the experiment. This is seen as a threat because they might be motivated more than the average population. However, because we did not measure their performance, this threat is not relevant for our experiments.

External validity: Although the setting was artificial and the participants did not belong to the addressed population, we claim that the results can be generalized. In a trial run with an elderly person, we did not observe differences in the performance of the evaluation tasks compared to a younger person. The rules of the ARC were not individually tailored to the participants' behavior and the apartment reflects a real habitat of an elderly person.

Conclusion validity: Reliability of measures and reliability of data processing: The measurement was conducted by the ARC. It worked identically for all participants and scenarios. Quality of data: The quality of the measured data can be considered as high, because the observations were recorded on video and checked against the correct performance of the scenarios. Statistical instruments were used to calculate the significance of the hypothesis.

5 Conclusions

We presented initial evaluation results from two research projects, EMERGE and BelAmI. In detail, we used an approach to recognize the ADL personal hygiene, toilet usage, and preparation of meals by analyzing characteristic patterns in a continuous sensor data stream. The approach was evaluated empirically by means of experiments in a near real-life environment of an assisted living apartment. The results from the experiments were quite positive. The interpretation of our evaluation results proved that it is possible to measure ADLs accurately enough for the purpose of detecting behavior deviations.

It can be concluded that we found a setup of a technological solution and controlled experiments in a near real-life environment that allows measuring ADLs unobtrusively with totally ambient sensors and nearly no restrictions on the normal behavior. To reach this objective, it turned out to be very useful to include all stakeholders very early in the requirements analysis and development process for the prototypes and especially in the setup of the experiments. This will now enable us to a) progress towards measuring long-term behavior deviations by putting the HCM into operation and perform experiments on long-term behavior deviation recognition and b) start real-life field trials and evaluations from the professional point of view and the end users' point of view.

Acknowledgements. This work was carried out in the projects EMERGE [1] and BelAmI [2].

References

1. EMERGE: Emergency Monitoring and Prevention, Specific Targeted Research Project (STREP) of the EC, Project No. 045056, http://www.emerge-project.eu
2. BelAmI: Bilateral German-Hungarian Collaboration on Ambient Intellgence Systems, funded by the German Federal Ministry of Education and Research (BMBF), the Fraunhofer-Gesellschaft, and the Ministry of Science, Education, Research, and Culture of Rhineland-Palatinate (MWWFK), http://www.belami-project.org
3. Miskelly, F.G.: Assistive technology in elderly care. Age and Ageing 30(6), 455–458 (2001), http://ageing.oxfordjournals.org/cgi/reprint/30/6/455
4. Frost, Sullivan: European Remote Patient Monitoring Markets, B519–B556 (2005)

5. ICT for independent living for elderly. A study on ICT enabled independent living for elderly for the European Commission created by VDI/VDE Innovation + Technik GmbH, Berlin, Germany, `http://www.ict-ile.eu/db/products`
6. Dishman, E.: Inventing Wellness Systems for Aging in Place. IEEE Computer 37(5), 34–41 (2004)
7. Katz, S., Ford, A.B., Moskowitz, R.W., Jackson, B.A., Jaffe, M.W.: Studies of Illness in the Aged: The Index of ADL: A Standardized Measure of Biological and Psychosocial Function. Journal of the American Medical Association 185(12), 914–919 (1963)
8. Lawton, M.P.: Scales to Measure Competence in Everyday Activities. Psychoparmacological Bulletin 24, 609–710 (1988)
9. Kane, R.A., Kane, R.L.: Assessing the Elderly – A Practical Guide to Measurement, pp. 25–67. Lexington Books, Lexington (1981)
10. Fillenbaum, G.: Multidimensional Functional Assessment – Overview. In: Mezey, M.D. (ed.) The Encyclopedia of Elder Care, pp. 438–440. Springer, New York (2001)
11. Glascock, A.P., Kutzik, D.M.: The Impact of Behavioral Monitoring Technology on the Provision of Health Care in the Home. Journal of Universal Computer Science 12(1), 59–79 (2006)
12. Kutzik, D.M., Glascock, A.P.: Monitoring Household Behavior to enhance Safety and Well-Being. In: Burdick, D., Kwon, S. (eds.) Gerontechnology – Research and Practice in Technology and Aging, pp. 132–144. Springer, New York (2004)
13. Glascock, A.P., Kutzik, D.M.: Automated Behavioral Monitoring. In: Keijer, U., Sandstrom, G. (eds.) Smart Homes and User Values, Eastern Mediterranean University, Faculty of Architecture, Gazimagusa, Mersin, Turkey, pp. 83–106 (2007)
14. Vitaphone, `http://www.vitaphone.de`
15. DRK Hausnotruf, `http://www.drk-hausnotruf.net`
16. IST Vivago, `http://www.istsec.fi`
17. Sophia, `http://www.sophia-nrw.de`
18. Storf, H., Becker, M.: A Multi-Agent-based Activity Recognition Approach for Ambient Assisted Living. In: 3rd Workshop Artificial Intelligence Techniques for Ambient Intelligence, European Conference on Artificial Intelligence, Patras, Greece (2008)
19. Nehmer, J., Karshmer, A., Becker, M., Lamm, R.: Living Assistance Systems: An Ambient Approach. In: Osterweil, L.J. (ed.) Proc. 28th Int. Conf Software Engineering, Shanghai, pp. 43–50. ACM Press, New York (2006)
20. Garsden, H., Basilakis, J., Celler, B.G., Huynh, K.: A home health monitoring system including intelligent reporting and alerts. In: IEEE Conf. Engineering in Medicine and Biology Society 2004, vol. (5), pp. 3151–3154 (2004)
21. Celler, B.G., et al.: An instrumentation system for the remote monitoring of changes in functional health status of the elderly at home. In: Proc. of the 16th Ann. Int. Conf. of IEEE Eng. Med. and Biol. Soc., vol. 2, pp. 908–909 (1994)
22. Basili, V.R., Caldiera, G., Rombach, H.D.: Goal Question Metric Paradigm. In: Marciniak, J.J. (ed.) Encyclopedia of Software Engineering, vol. 1, pp. 528–532. John Wiley & Sons, Chichester (2001)