Effects of aging and domain knowledge on usability in small screen devices for diabetes patients.

Andé Calero Valdez[1], Martina Ziefle[1], Andreas Horstmann[2], Daniel Herding[2], Ulrik Schroeder[2]
[1]Human Technology Centre (Humtec)
[2] Computer-supported learning research group, Department of Computer Science
RWTH Aachen University, Germany
Calero Valdez@humtec.rwth-aachen.de

Abstract. Mobile small screen technology increasingly penetrates health care and medical applications. However, usability research regarding the ease of using these devices and the acceptance did mostly address aspects of information and communication technologies and a young and healthy user group. Yet, the specific needs and wants of patients, which might use small screen devices for monitoring health states are not regarded sufficiently. This especially applies to usability barriers of older users, which are known to be very sensitive to suboptimal interfaces. The present study investigates impacts of ageing, technology expertise and domain knowledge on user interaction using the example of diabetes. The research followed a two-step procedure. First, software for the monitoring of diabetes had been developed and implemented on a simulated small screen device. In a second step, users' effectiveness and efficiency have been measured by considering the navigation performance of younger and older diabetes patients as well as healthy users. Results show that age and technology expertise have a big impact on usability of the device. Furthermore, impacts of user characteristics and success during the trial on acceptance of the device were surveyed and analyzed.

Keywords: Usability, Aging, Health care, Performance, Small screen device, Perceived Ease of Use, Perceived Usefulness, Technical experience, Domain knowledge

1 Introduction

An ever-increasing amount of technical devices with small screens and complex hierarchical menu systems surge into every day life. A simple press of a button on a mobile phone can connect the device to the Internet and add to the unknowing customers bill. Such handling errors on mobile phones are - though bothersome - a mere hassle compared to severe consequences of difficulty in using technical devices in a different context. Small-screen-device penetration in varying medical contexts is soaring for certain diseases. Individual disease related bio-physiological parameters are monitored electronically in order to regulate application of drugs and even enhance connectivity to nursing staff, physicians or family members. Slick usability is becoming the critical factor for acceptance, sustainability and competitive capacity of any mobile technical system, especially in regard of demographic changes, world wide increasing life expectancy and the resulting increase of older users. The share of population over the age of 65 already reached 20 percent in Germany in 2008 and is expected to swell to a 38 percent level in 2038 [1]. Similar forecasts apply to many western European countries [2]. Usage of electronic devices is also decreasingly voluntary because of either work or everyday life requirements (e.g. [3][4]). This impact will be even stronger concerning medical appliances of mobile devices. Since the increase of age related illnesses like diabetes accompanies both demographic change and sedentary lifestyle, medical care and age appropriate independent domestic care can only be economically realized through technical solutions (e.g. [5]).

Designing such solutions in a self-explanatory and usable way for heterogeneous user groups has not been realized to date [6][7][8]. Device development is still dominantly technical-oriented and criteria of usability and learnability are mostly applied subordinately, if at all. This is directly related to the development of these devices through computer scientists and engineers, and lack of harmonization with psychological and ergonomic knowledge of necessities, capabilities and cognitive structures of the end users.

This paper examines the usability of a small screen device for diabetes patients. In the following, first, the importance of diabetes as a main civilization disease is outlined, followed by the status of knowledge regarding the usability of small screen devices, in combination with the impact of the diverse user group, which is using these devices. The chapter closes with the research questions addressed by a usability experiment.

1.1 Diabetes and technology

Diabetes mellitus is a metabolism dysfunction, which affects about 8 million people (10% of the population) in Germany alone and for the year 2010 an increase of up to 10 million affected is expected. Diabetes and secondary disorder treatment already covers 20% of Germanys compulsory health insurance funds expenditure. Diabetes alone is expected to cause a hole of 40 billion Euros in Germanys health care budget in 2010 [9]. Diabetes predominantly causes a dysfunction of the blood glucose metabolism and is caused by different phenomena. The body produces either too little insulin or no insulin at all. In some cases the available insulin can no longer be effectively used by the body to regulate blood glucose levels. Generally two types of diabetes are differentiated.

Type-1-Diabetes

Diabetes mellitus type 1 occurs mostly in younger adults between 5 and 50 years, but can also occur later in life. Type-1-diabetes is an immune mediated disease but causes for its incidence are still unknown. Genetic factor, viral infections and environmental influences are expected to contribute to type-1-diabetes. The immune system of affected patients destroys the pancreas' beta cells, which then no longer produce insulin, causing hyperglycemia. The main symptom of type-1-diabetes is absolute insulin deficiency. Only 8-10% of all diabetes patients are type-1-diabetics [10].

Type-2-Diabetes

Diabetes mellitus type 2 occurs mostly after the 40th year of one's life but increasingly often occurs in children and younger adults today. Main causes for type-2-diabetes are obesity and lack of physical exercise. Typical secondary disorders encompass hypertonia (high blood pressure), micro- and macroangioma (mostly benign tumors) and arteriosclerosis. The main symptom of type-2-diabetes is body cell insulin resistance. This forces the pancreas into an overproduction of insulin (referred to as hyperinsulinism) to prevent hyperglycemia, which further raises insulin resistance in all body cells. This relative insulin deficiency can lead to absolute insulin deficiency when the pancreas' insulin production collapses. About 90% of all diabetes patients are type-2-diabetics [10].

Secondary disorders

Chronic high blood glucose levels cause long-term damage to the vascular and neural system. Over-frequently exposure to hyperglycemia manifests in diabetic heart disease, retinopathy (eye damage), nephropathy (kidney damage), and diabetic neuropathy (neural damage). These degenerative effects lastly cause blindness, renal failure, amputations and heart failure.

A persisting type-2-diabetes illness may also cause cognitive deficiencies, especially in patients older than 50 years executive functions and the neurocognitive processing speed are affected. Episodic memory, word flow and semantic memory though seem to be unaffected by type-2-diabetes [11].

Another frequent symptom of diabetes is hypoglycemia (insufficient blood glucose level). If a patient's glucose level drops too far (e.g. the patient administers too much insulin) diabetic coma occurs. This constitutes a case of emergency since the patient can no longer help himself and instant application of glucose becomes necessary.

Diabetes therapy

Goal of any diabetes therapy is a stable and healthy blood glucose level (<135mg/dl postprandial). This can be accomplished by multiple means. Oral anti-diabetic drugs can increase effectiveness of bodily insulin or decrease the rate of intestinal glucose reception. Depending on the severity of the diseases subcutaneous application of insulin is required. Insulin is mostly administered before or after meals, since eating has a big effect on blood glucose level.

Two types of insulin therapy are discerned. Conventional therapy (CT) follows a fixed injection plan and intensive conventional therapy (ICT) requires the patient to measure blood glucose level and inject insulin accordingly. ICT-patients can also be treated with an insulin pump - a device that constantly administers insulin over the day and that offers an interface to increase dosage after meals.

Type-2-diabetes can sometimes be treated by diet and physical exercise alone, since reduction of body weight can in some cases cause full remission [12].

Almost all therapy types require or can at least be assisted with mobile small screen devices, since monitoring, persisting and analyzing of blood glucose levels, insulin dosage and caloric intake increases therapy success, since their correlation and behavior can vary drastically between individuals.

1.2 The usability demands in small screen devices

It is a central claim that mobile devices are designed to be in line with users' specificity and diversity. However, the intelligent interface design of mobile devices, which meets the demands and abilities of especially older users, is an extremely sophisticated task. Aging itself represents a highly complex process. Not all users age in the same way, and the onset of aging processes as well as the consequences show considerable differences across humans. Design approaches should therefore take the user-perspective seriously. This includes that adults' behavior with current technical devices is carefully studied and also, that user abilities are identified, which affect the interaction with interactive computing devices.

The miniaturization of small screen devices may also contribute to usability shortcomings. Beyond handling and visibility problems, the restricted screen space allows only little information to be displayed at a time. By this, memory load is increased. In addition, orientation in the menu is complicated, because users do not experience how the menu might be "spatially" structured and how the functions are arranged [6][7][13][14][15][16]. In hierarchically structured menus disorientation occurs when complexity is high with respect to the depth and breadth of menu levels [13][17][18].

With respect to effects of users' age, the profound changes in sensory, physical, psychomotor and cognitive functioning over the life span are well known (e.g. [19]). These changes may account for older adults' lower performance when using technical devices. Furthermore, due to a different upbringing, older adults often have a lower technical understanding and are less experienced in computer usage. As a result, the majority of older adults possess limited computer knowledge, which may also account for differences in computer-based performance (e.g. [20][21][22]). However, it was found that age-related decreases could be compensated by expertise (e.g. [23]). Thus, performance of older adults can be just as good as that of younger adults when they can rely on elaborated domain-specific knowledge.

1.3 Questions addressed and logic of experiment

The present experimental study addresses two basic topics: the impact of aging and domain knowledge of diabetes on task performance on a small-screen diabetes living assistant. Therefore participants were selected from different age groups and screened for domain knowledge of diabetes. Additionally expertise in technology was surveyed.

Although this study was primarily designed exploratory, the following outcomes were expected:

- Younger users due to aging impacts on both cognitive and perceptual abilities outperform older users (e.g. [6][8][11][24]).
- Users with higher expertise in technology usage outperform users with lower expertise due to conceptual transfer of navigation user interfaces (e.g. cell phone navigation) (e.g. [14][19][24]
- Users with higher domain knowledge outperform users with lower domain knowledge due to improved understanding of tasks and higher appreciation of

purpose behind function of the user interface (e.g. a diabetes patient knows about bread unit calculation and it’s importance) (see [16][21][23]).

- Users with type-1-diabetes outperform users with type-2-diabetes due to the nature of those two illnesses and the coinciding difference in domain knowledge. Type-1-diabetes patients require more frequently and stricter regulation of blood glucose levels as this type of diabetes usually occurs earlier in patients’ lives. A good comprehension of the disease is critical for successful long-term treatment. Insulin medication is obligatory for this disease. Type-2-diabetes patients in contrast can be treated in many different ways. Some patients are only medicated with a single daily intake of an oral anti diabetic drug. Real time blood glucose regulation is often not as urgent for therapy success, since patients are mostly too old to experience long-term effects of the illness. Thus comprehension of the disease is not as exigent as in type-1-diabetes.

2 Method

The objective of the study was to understand influence of aging and domain knowledge on task performance on a small screen living assistant for diabetes patients and to gain knowledge of determining factors on navigation performance in small screen touch enabled devices. Since the current study claims to extend the earlier research, efforts were made to keep the method very similar to that used before. In this section the conceptual design and the procedure of the experiment are described.

2.1 Experimental Variables

In our study we considered five independent and five dependent variables. The first independent variable we examined is user age in order to measure influence on both task effectiveness (i.e. the amount of tasks solved correctly) and efficiency (the amount of time required to solve a task). Additionally we analyzed the influence of expertise with medical technology, overall technical expertise and in particular mobile phone navigation expertise on usability (as in EN ISO 9241-11, 1998) of the device. To measure impact of domain knowledge on effectiveness and efficiency knowledge of four key health parameters (blood glucose, HbA_{1c}, blood pressure, body fat percentage) were surveyed and aggregated as an independent variable.

As dependent variables five performances criteria were measured: success rate, total steps, detour steps, total time and time per step. Success rate is measured as the percentage of successfully performed task steps of each task. Effectiveness was not measured as a Boolean variable in order to account for users who were able to solve tasks mostly correct but missed a certain step to solve a task with 100% correctness. These users can still be viewed as effective, as not necessarily all steps are required in order to perform well enough for the device to be useful. Total steps are the amount of program interactions performed for a certain task. A user interaction is an interaction of the user, which changes the state of the program. Pressing on the non-interactive background or missing a button is not included in total steps. Many tasks could be solved in multiple ways allowing users to complete each task in differing amounts of total steps. Detour steps are all program interactions that do not account into solving the task at hand, such as navigation failures, accidently pressed buttons and unnecessary repeated input. Total time is the amount of time the users take to finish all tasks without account for reading time of task descriptions. Time per step is the average amount of time a user takes between to program interactions. A lower value represents a faster navigation pace but not necessarily a better navigation performance. Since total steps, detour steps, total time and time per step are also measured for unsuccessful tasks; data from these dependent variables must always be related to the success rate of the current task.

2.2 Participants

A total of twenty-three adults volunteered to take part in this study. In Figure 1, the age distribution of the sample is depicted. Among those, were seven young adults (2 males, 5 females) with a mean age of 27.4 years (SD = 2.6; range: 25 – 33 years), seven medium aged adults (2 males, 5 females) with a mean age of 51.3 years (SD = 8.3; range: 41 – 59 years) and nine older adults (3 males, 6 females) with a mean age of 67.9 years (SD = 7,8; range: 61 – 87 years). The younger participants were mostly university students of different academic fields (psychology, social science, engineering, medicine). Medium aged and older adults were reached by advertisement in local newspapers and through an exhibition on a local public diabetes convention and covered a broad range of professions and educational levels (e.g. administrative officers, secretaries, teachers, engineers, physicians).

Twelve participants were non-diabetic adults, who were mostly recruited through their social networks (3 males, 9 females, mean age = 44.8; SD = 18.4; range: 25 – 71 years). The eleven diabetic participants (4 males, 7 females; mean age = 56.7; SD = 16.8; range: 26 – 87 years) split up into a group of five participants diagnosed with Type-1-Diabetes (1 male, 4 females; mean age = 43.6; SD = 13.6; range: 26 – 64 years) and 6 participants suffering from Type-2-Diabetes (3 males, 3 females; mean age = 67.7 SD = 10.2; range: 59 – 87 years).

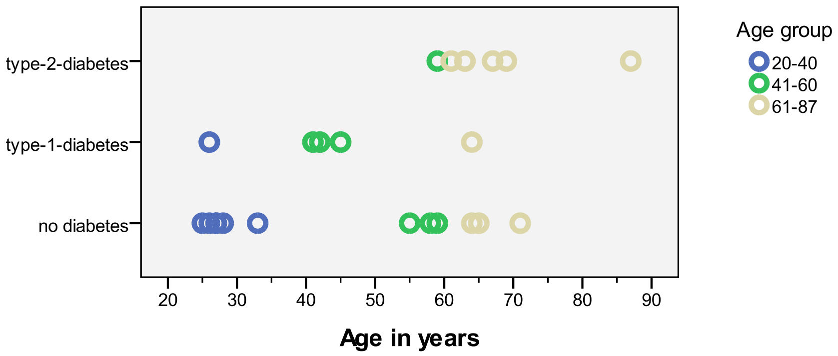

Fig. 1: age and health status distribution and age group allocation

Regarding the recruitment of older participants a prototypical ideal participant "diabetic but otherwise healthy senior" was aimed at. All medium aged and older adults participating were either active parts of the work force or otherwise mentally fit and not hampered by stronger age-related sensory and psychomotor limitations. All participants were novices to the small-screen device we developed.

In order to relate effects on usage performance to prior experience with modern technology or experience with medical technology, participants were asked about their experience with different nonmedical (mobile phone, computer, GPS navigation, digital camera, microwave oven, alarm clock, gaming console) and medical devices (blood glucose meter, hearing aid, blood pressure meter, heart rate monitor, in-house emergency call). Since the study was performed using a small screen device user experience with mobile phones was surveyed thoroughly as well. Measurement was applied to functions of mobile phones (calling, text messaging, address book, calendar, integrated camera, integrated radio, integrated GPS navigation, internet browser, games, alarm, email). Two types of measurements were applied to three different areas of technology. Perceived Ease of Use (PEU) (see [25][26][27]) and Usage Frequency (UF) were aggregated for the three categories of expertise (expertise with technology, expertise with medical technology, expertise with mobile phone menu navigation). Perceived Usefulness (PU) was only measured for mobile phone functions because usefulness in this study was concerned as an attribute of a function of device, rather than of a device it self. In addition to any technical experience, domain knowledge about diabetes was collected for all participants.

Technical expertise was surveyed by measuring the PEU and UF. Both PEU and UF were measured on a Six-Point Likert Scale. PEU was examined with questions like "How easy to use is for you..." (1 = very easy, 2 = easy, 3 = rather easy, 4 = rather hard, 5 = hard, 6 = very hard). UF was similarly examined with questions like "How often do you use a..." (1 = Daily, 2 = 2 - 3 times a week, 3 = once per week, 4 = 1 - 2 times a month, 5 = 1 - 2 times a year, 6 = never). Total expertise is calculated as the square root of the product of the mean of all PEU and the mean of all usage frequency (UF) in order to reflect a value that is also on a 6-point-Likert scale where 1 reflects a value with highest usage frequency and highest PEU and 6 represent the exact opposite. Intermediary values reflect both PEU and UF, with a tendency to rank equal values of PEU and UF higher than differing values. A person who uses a computer often, but finds it hard to use, scores lower (i.e. better), than a user that uses a computer not as often but ranks ease of use on a similar level. This effect is desired in order to account for misperception due to very extreme usage frequency (e.g. a person that only uses the computer to write letters once a year, but finds this task very easy). A person that finds a computer quite hard to use but uses it daily is expected to have a better computer expertise but a perception bias to find usage harder.

Expertise with medical technology was surveyed analogously to technical expertise. Devices that are used in medical context were used more frequently and perceived as easier to use (mean = 1.6; SD = 0,6; range = 1 – 3; N = 16) compared to normal technology (mean = 4.1, SD = 1.1, range 2 – 6, N = 23). Here, participants who had no experience with medical technology were not taken into account.

Mobile Phone expertise was also surveyed in the same manner (mean = 4.8; SD = 1.1; range = 2 – 6, N = 23) but additionally a total PU was measured (Six-Point Likert scale; 1 = very useful; 2 = useful; 3 = rather useful, 4 = rather not useful, 5 = not useful; 6 = not useful at all) for different functions of the mobile phone.

Domain Knowledge (mean = 4,0; SD = 1.7; range = 2 – 6, N = 23) was surveyed with a Five-point Likert scale ("How well do you know…"). Answers ranged from 1 = "very precise" to 4 = "not at all". Option 5 was labeled with "I don't know" and considers the item at hand to be completely out of the knowledge of the person contrasting to 4, where the person has heard about the measurement, but does not know about his own value for this measurement.

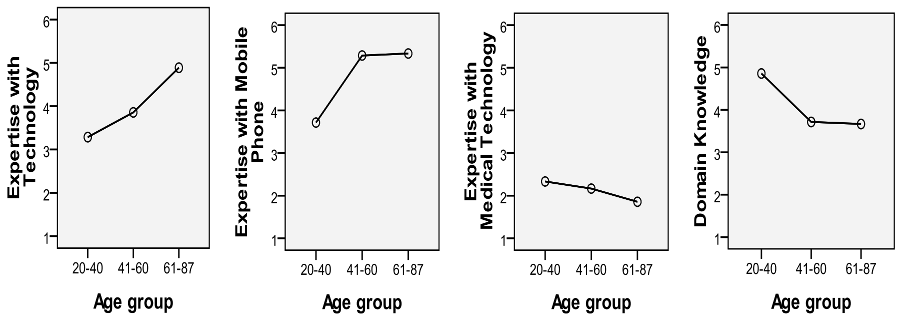

Fig.2: Means plot of Expertises over age groups

ANOVA-analysis of these four factors regarding to age shows, that technical as well as mobile-phone expertise are both correlated with ageing, where expertise with medical technology and domain knowledge are not (Table 1). This is expected since older users should be equally (if not more) prone to using medical technology as younger users. Especially in the case of diabetes domain knowledge should rather depend on the period of time being affected by the illness than on the numerical age.

Table 1: One Way-ANOVA table for mean differences of total technical expertise, total medical technical expertise, mobile phone expertise and domain knowledge regarding differences between the age groups

	SS	*MS*	*df*	*F*	*p*
Total technical expertise	10.65	5.33	2	6.20	< 0.01
Total mobile phone expertise	12.45	6.22	2	7.38	< 0.01
Total medical technology expertise	0.04	0.02	2	0.17	> 0.05
Domain knowledge	6.67	3.34	2	1.19	> 0.05

Post-hoc Bonferroni testing shows that significant differences only exist between young adult users and older adult users for technical expertise (p < 0,01). Medium aged adults are both more experienced than older users and less experienced than younger users, but this difference fails to reach significance level (p < 0,12).

Significant differences in mobile phone expertise only exist between young adult users and both medium (p < 0.05) and older aged users (p < 0.01).

2.3 Development of a small screen device for diabetes patients

Our goal is to develop a portable device that supports diabetes patients in their therapy and in their everyday lives. Before dealing with the hardware part of these devices, we wanted to concentrate on the usability of the software. For the user studies, we needed an application that can be run on standard hardware. Another constraint was given by our decision to use the Jacareto capture and replay toolkit (see [28][29][30]).

Jacareto can record the user interaction with a Java application, structure the captured data, and export it for statistical analysis. This saves time and effort during and after user studies, as it is no longer necessary to create, transcribe, and analyze video recordings. Unfortunately all existing diabetes applications that we found were not implemented in Java or were incompatible with Jacareto for other reasons. Therefore we implemented a new tool from scratch.

One advantage of Java applications is platform independence: you can run them on PC's, PDA's, cell phones, and other devices. However, Java ME (Java Mobile Edition, the Java version for portable devices) only covers a subset of the functionality of Java Standard Edition. As compatibility with small-screen devices such as cell phones was important to us, we had to meet some constraints during development. For example, we did not use Java 5 language features, and we used the Abstract Window Toolkit (AWT) instead of the more modern Swing. Furthermore we made sure that the user interface could be used without a keyboard, so that it can later be used on a touch screen-enabled mobile device.

Instead of creating a specialized application that is only useful for certain diabetes patients, we decided to include features that are required for the different types of the disease. On the first start, the user has to set up the application by entering his characteristic values (such as drugs that he has to use regularly for his therapy).

The most important feature in everyday use is the so-called diabetes diary (Diabetes-Tagebuch). Every time the patient measures or influences his blood sugar concentration, he is supposed to insert the data into his diary, using a wizard-based input mechanism. For instance, when the patient has measured that his blood sugar is low and has therefore eaten dextrose to raise it onto a normal level, he creates a new entry. He enters the time, the measured blood sugar, and the bread units' equivalent of the ingested dextrose (1 bread unit (BU) = 12 g of carbohydrates). In the wizard, he simply skips values that were not relevant for this entry, such as the bolus insulin. The entered data is then shown in a column of a table. The tabular representation is based on the layout of the paper-made diaries that are in common use in Germany (see fig. 3), and that a large part of our target user group is already familiar with. 61 % of diabetes patients in Germany are using a diary to record their values; 91 % of these are keeping their diaries on paper.

Another application feature that is inspired by a paper template is the health passport, or Gesundheits-Pass. After each quarterly examination, the doctor writes down the results into this booklet. Like the diabetes diary, the values are entered in a table. For example, there are table rows for the HbA_{1c} value, the blood pressure, and the body weight. We were unable to use this table representation in our tool because of screen size constraints. That is why we only show the values of one quarter of a year, while the paper version has columns for four quarters. Besides the actual values of the quarter, we offered the possibility to enter the desirable values that the doctor determined during the examinations. We added this feature because the health passport that was in use in 2008 had columns for desirable values. These columns were removed in the 2009 version of the Gesundheits-Pass.

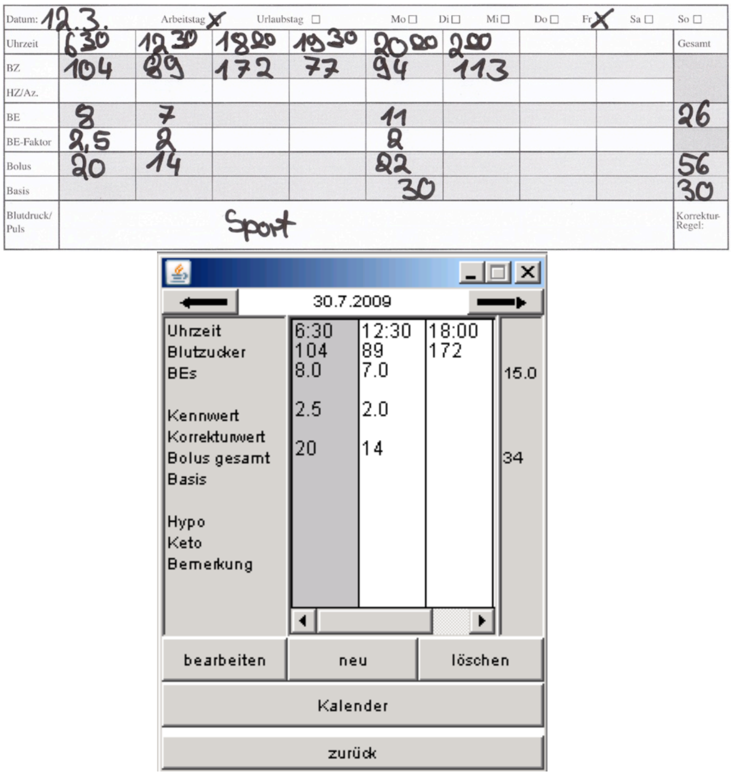

Fig. 3: The paper version of the diary (top) and the diary function of our application (bottom).

For a successful diabetes therapy, it is important to teach the patients a basic knowledge of the nutrient contents of groceries. Especially patients who inject insulin need to calculate their drug dosage on the basis of the food they consume. Most people use a scale to measure the weight of the food, and then look up the bread units (BU) per gram in a nutrition table.

After calculating the product of these values, they enter the result in their diabetes diary. We included an application feature that supports the user in looking up and calculating these values. He can choose a grocery from a predefined list, and then enter a weight or volume. The application then displays the bread units and kilocalories (see fig. 4). The user then repeats these steps with the remaining ingredients of his meal, and copies the resulting BU sum into his diary. To reduce the routine work of choosing the ingredients of regular dishes, the user can save a meal as a favorite, and reuse it later.

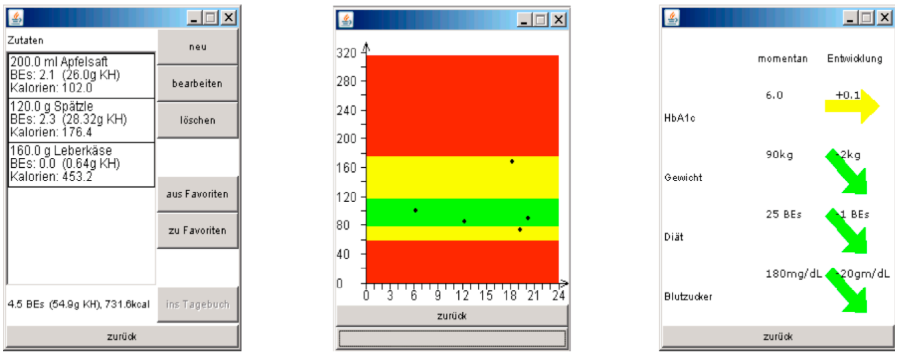

Fig. 4: The bread unit calculator (left), the plotter (center), and the screener (right).

There are two features that help the patient to keep track of the progress of his therapy. The first one is the so-called plotter, which shows the course of measured values in a history diagram (see fig. 3). We only implemented a functional diagram for a one-day overview of the blood sugar concentration. The remaining diagram types are static, which was sufficient for our research prototype. The other feature is called screener. It displays the latest entry of characteristic values such as body weight and blood sugar concentration, and compares it to the previous entry. Colored arrows visualize the tendency.

2.4 Experimental procedure

In order to test the research model and to determine the effects of domain knowledge and age variables on performance, an experimental setting with a simulated small-screen-device was conducted.

At the outset participants completed a paper-based questionnaire concerning demographical information (age, gender, educational achievement) and information about the familiarity with common technical devices, mobile phone and common medical technology (usage frequency as well the perceived ease of use). The assessing of demographic data was performed paper-based. It was of high importance that this questionnaire was realized prior to the simulation, as performance during the experiment could impact and bias self-assessment and thus expertise ratings.

After completing the survey participants were asked to perform a set of five tasks on the simulated device. Each task regarded a different main function of the device in order to create both a realistically setting and an interaction with various UI elements and uses cases for the device. All medical values that were to be entered into the device were predefined and given to the participant on a paper based task description to create equal preconditions for all participants. Task-information was printed on hardcopy and was available throughout. For all tasks a total time limit of 30 minutes was given implicitly (participants were told that the experiment should last for about 30 minutes). The fastest user completed all 5 tasks in about 6 minutes and all participants finished under 30 minutes.

After completion of the experimental tasks participants were asked to rate the perceived ease of use (PEU) and perceived usefulness (PU) of the used functions in the simulated device to assess the users acceptance of the device.

2.5 Small screen device simulation

The diabetes living assistant was simulated as a software solution one a PC running Windows XP connected to an Iiyama AX3819UT touch screen (15“ TFT-display, display resolution 1024x768 pixels). The simulated device spanned over 245x319 pixels (width = 7,27cm; height = 9,47cm) and was displayed in the center of the screen. The rest of the visible screen was covered with an opaque paper cutout to prevent any interaction with the operating system.

Participants were seated on a height-adjustable chair in a comfortable seating position. In order to control viewing conditions, participants were not allowed to choose viewing angle, viewing distance or inclination of the TFT-Monitor. If the participant required any corrective lenses, wearing those was obliged throughout the experiment. Lighting conditions were kept the same by choosing a room with no exterior lighting and a fixed interior lighting system.

2.6 Questionnaires

Perceived ease of use (PEU) and perceived usefulness (PU). Users’ technology acceptance was assessed by original items from the Technology Acceptance Model of Davis (TAM [25]). The perceived ease of use (PEU) implies ‘the extent to which a person believes that using a particular system would be free of effort’, and secondly, the perceived usefulness (PU) which is defined as ‘the extent to which a person believes that using a particular system would enhance his or her job performance’ [25]. The validity and reliability of TAM items had been proven by several empirical studies (e.g., [25][26][27]), and also showed satisfactory values in this study. The 5 presented PEU items had to be judged on a six-point Likert-scale ranging from 1 (very easy) to 6 (very hard). PU items were rated on a six point Likert scale ranging from 1 (very useful) to 6 (not useful at all). Smaller values would reflect a higher acceptance of the device.

Experimental tasks. Five Tasks were to be solved by the participants. In particular users were first asked to setup the “freshly unboxed” device and enter information about their current therapy (i.e. insulin type, dosage and schedule as well as total weekly calories) with given fake values. Secondly users were asked to fill out their health passport in order to complete the setup of the device. After completing the first two tasks participants should enter three blood glucose measurements along with dietary information and insulin dosage for three times of a the given day (morning, noon, afternoon) into the digital diary. Again all values were predefined. The fourth glucose measuring was preceded by a task in which the users had to calculate the bread units of a given meal using the BE-Calculator of the device. This value was then to be used as dietary information in the digital diary for the fourth measurement.

The last task required the user to simply view the daily blood glucose graph in the plotter of the device. All tasks were described in natural language but data for all input forms was given numerically.

In the following, examples of two task types are described:

- Example for 'digital diary'-task: *'After finishing configuration of your device, daily blood glucose measurements can be stored in the devices digital diary. Please enter the following measurement into the digital diary.*
  This morning 9:20am:
  Blood glucose level 123; consumed 3 bread units, no correction of insulin dosage; no basal-insulin dosage; no hypo- or ketoacidosis was measured'.
- Example for 'BE-Calculator-task: *'You are hungry and want to eat some fish sticks (200grams) and have a glass of apple juice (200ml). Please calculate the bread units for this meal using the BE-Calculator of the device.'*

3 Results

Results of this study were analyzed by one-way ANOVA, bivariate correlations, multivariate analysis of variance and univariate analysis of covariance and linear regression with a level of significance set at 5%.

The result section is designed as follows: first, we assess correlative relations and impact of individual factors (age, health status, domain knowledge, expertise with technology) on users' performance; second, a deeper analysis of aging effects on effectiveness and efficiency is conducted. At last effects of different factors on acceptance of the simulated device is presented.

3.1 Effects of age, domain knowledge and technical proficiency on performance

Relationship between factors and performance. To get a first insight into the data, correlations (Spearman rank analysis) between individual variables and performance measures were carried out (Table 2).

Table 2: Bivariate Correlations between age and user characteristics and performance

	Success Rate	*Total steps*	*Detour Steps*	*Total Time*	*Time per step*
Age	-0.664**	0.616**	0.472*	0.231	0.693**
Expertise with technology	-0.449*	0.330	0.244	0.476*	0.320
Expertise with med. technology	0.251	-0.266	-0.146	0.101	-0.342
Mobile Phone Expertise	-0.339	0.295	-0.006	0.393	0.301
Health Status	-0.179	0.342	0.181	-0.102	0.421
Domain knowledge	-0.53	-0.167	0.097	0.314	-0.244

**$p < 0.01$, *$p < 0.05$

Correlation analysis shows that only age and expertise with technology show a significant correlation with performance measures. Younger age is highly correlated with better effectiveness (r = -0.664) and efficiency. Younger users need less total steps (r = 0.616), make less navigation errors (r = 0.472) and have a faster navigation pace (r = 0.693). Expertise with technology is mostly correlated with effectiveness (r = -0.449) such that users with better expertise are more effective than users that are more inexperienced. This correlation does only affect one efficiency measurement significantly (i.e. total time r = 0.476), which also shows that higher expertise is related with better performance.

Apparently domain knowledge (r = 0.53) and health status (r = -0.179) seem to have an unexpected adverse effect on effectiveness, but further correlation analysis shows that age is highly correlated with health status, and health status highly correlated with domain knowledge (Table 3).

Table 3: Bivariate Correlations between age, health status and domain knowledge

	Age	*Health Status*	*Domain knowledge*
Age	1	0.509*	0.253
Health Status		1	-0.799**
Domain knowledge			1

**$p < 0.01$, *$p < 0.05$

To examine how domain knowledge and health status predict performance, two analyses of covariance (ANCOVA) were conducted using ‘domain knowledge’ and ‘health status’ as a covariate. The ANCOVA revealed no significant main effect for domain knowledge (F = 1.817; p > 0.05) with ‘domain knowledge as a covariate. Choosing ‘health status’ also reveals no significant main effect on effectiveness (F = 1.808, p > 0.05).

A means plot of effectiveness over age grouped by domain knowledge indicates that domain knowledge might have an effect on effectiveness, even though the significance level was not reached with the data at hand. Similar observations can be made for health status. Means for non-diabetic participants are lower than for type 2 diabetes participants, which are lower than means of the type 1 diabetes participants (fig. 5). The higher the domain knowledge the higher the success rate with the small screen device. In close relationship to the latter results we see that domain knowledge also advantages performance: Diabetes patients have higher success rates than healthy participants, and diabetes type 1 patients outperform diabetes type 2 patients. Again, this can be seen as an effect of domain knowledge, as diabetes type-1-patients cope with the disease for a much longer time compared to diabetes-type-2-patients. This tendency shows up in all age groups, however, is most pronounced in the oldest age group (61-67 years of age), corroborating that the “age-disadvantage” when using small screen devices may be relieved by domain knowledge.

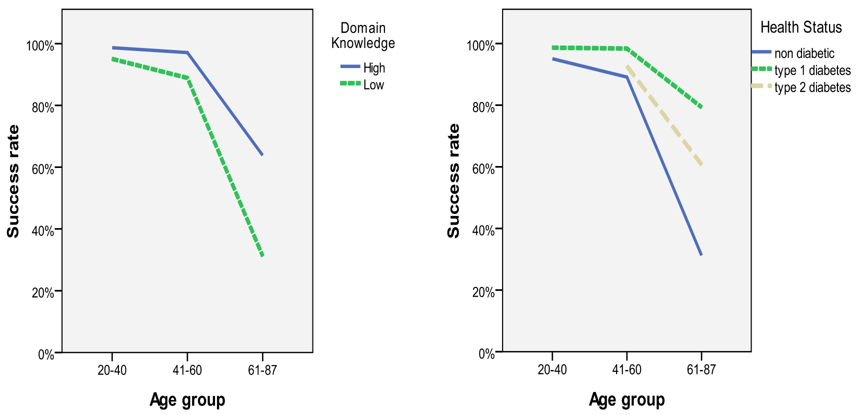

Fig. 5: Means plot: success rate over age group grouped by median split domain knowledge (left); success rate over age grouped by health status (right)

3.2 Effects of age on navigation performance

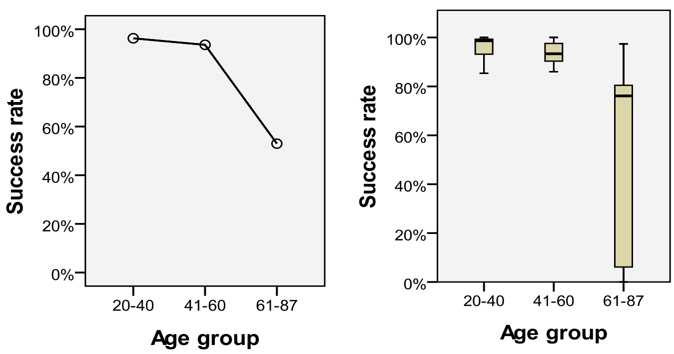

Fig 6: Means plot of success rate over age group, higher values indicate better effectiveness (left). Box plot of success rate over age (right)

Since total steps, detour steps and time per step are also measured for tasks, which had not been successfully solved in the time given, correction with success rate has to be performed. Corrected measurements are calculated by division of original measurement by success rate and indicated by "[c]". For instance participants that only completes half of the tasks successfully get a two-fold "penalty" on all efficiency measures. This can lead to overblown values, if very low success rates (e.g. 1%) are reached.

Table 4: One Way-ANOVA table for mean differences of success rate and total steps, detour steps, total time and time per step between age groups. All efficiency measures are corrected by success rate.

	SS	*MS*	*df*	*F*	*p*
Success rate	0.97	0.48	2	6.93	< 0.01
Total steps [c]	6475.41	3237.71	2	0.47	> 0.05
Detour steps [c]	1951.56	975.78	2	9.01	< 0.01
Total Time [c]	$5.02*10^{12}$	$2.51*10^{12}$	2	2.07	> 0.05
Time per step [c]	$6.89*10^{10}$	$3.45*10^{10}$	2	3.16	> 0.05

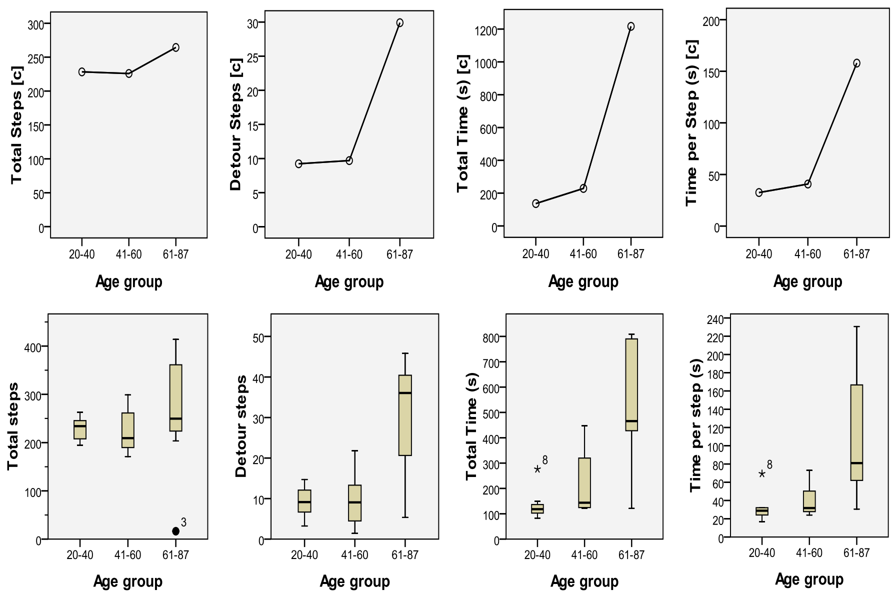

Fig 7: Means plots of efficiency measurements over age group (top row); from left to right: (1) total steps over age group, (2) detour steps over age group, (3) total time over age group and (4) navigation pace (time per step) over age group; Box plots of efficiency measurements over age group (bottom row)

Comparison of effectiveness shows that younger and medium users performed almost equally, and both outperformed older users. Younger users (Y) averaged a success rate of 96% (SD = 6.2%), similar to medium aged (M) users, who reached 94% (SD = 5.2%), while older users (O) only managed to reach a mean success rate of 53% (SD = 41.1%). The mean of corrected total steps shows almost no difference between age groups ($M_Y = 228.3$, $SD_Y = 25.3$; $M_M = 225.9$, $SD_M = 48.0$; $M_O = 264.3$, $SD_O = 133.8$). Older aged users made more than three times more navigational errors (detour steps: $M_O = 29.9$, $SD_O = 16.1$) during task performance than medium and younger users (detour steps: $M_M = 9.7$; $SD_M = 7.07$; $M_Y = 9.2$; $SD_Y = 4.1$). Older users also required more time to complete all tasks than medium or younger aged users (M_Y

= 136.4s, SD_Y = 65.3s; M_M = 228.8s, SD_M = 139.4s; M_O = 1216.5s, SD_O = 190.0s), which also reflects in a slower navigation pace. Older users trigger interactions almost 5 times slower than younger users, who are still almost 30% faster than medium aged users (M_Y = 32.5s, SD_Y = 17.1s; M_M = 40.7s, SD_M = 18.4s; M_O = 157.9s, SD_O = 179.1s).

Post-hoc Bonferroni testing shows that both younger and medium aged adults significantly ($p < 0.05$) outperform older adults in effectiveness (i.e. success rate). Younger and medium aged adults show now significant difference here. Apparently the only efficiency measure that shows significant mean differences is detour steps (corrected by success rate). Post-hoc Bonferroni testing showed similar effects for detour steps between groups as in success rate. Again younger and medium aged users outperform older users ($p < 0.05$), but fail to differ between each other significantly. All other efficiency measurements show differences betweens means that are not statistically significant.

3.2 Effects of navigational performance on acceptance

In order to understand how different factors (age, health status, expertise with technology, success rate in experiment) influence user acceptance of our simulated device non-parametrical analysis of correlations (Spearman's rho) was conducted (see Table 5). Both age and success rate show a significant correlation with acceptance of the device. Users that are more successful show higher acceptance of the device than unsuccessful users ($r = -0.507$). Increasing age also seems to lead to higher acceptance of the device ($r = -0.460$). Interestingly health status seems not to have any effect ($r = 0.027$) on acceptance at all ($p>0.05$).

Table 5: Bivariate Correlation between user characteristics, performance and acceptance

	Age	*DK*	*HS*	*TE*	*MTE*	*MBE*	*Success rate*
Acceptance	-0.460*	0.200	0.027	0.276	0.409	0.287	-0.507*

*$p < 0.05$

DK = Domain knowledge, HS = Health status, TE = Expertise with technology, MTE = Expertise with medical technology, MBE = Expertise with mobile phones

To examine how age and success rate affect acceptance, two analyses of covariance (ANCOVA) were conducted using 'age' and 'success rate' as a covariate. The ANCOVA analysis revealed no significant main effect for user age ($F = 3.502$; $p > 0.05$) with 'user age' as a covariate. Choosing 'success rate' as a covariate also reveals no significant main effect on acceptance ($F = 3.78$, $p > 0.05$).

Linear regression though contradicts this finding: both age and success rate explain 65.5% of the variance of acceptance of the simulated device, but success rate is a stronger predictor for acceptance ($\beta = -0.486$, $p < 0.05$) than user age ($\beta = 0.241$, $p > 0.05$). This suggests, that high performance in initial usage of a device might have a high impact on acceptance of a new device.

4 Discussion and conclusion

The present experimental study was conducted to provide deeper understanding of small-screen-device menu navigation performance in respect to age and domain knowledge in a medical context. A total of twenty-three participants accomplished five tasks designed for a diabetes living assistant. In order to analyze individual factors that may differentially affect user's performance, domain knowledge, expertise with technology, expertise with medical technology, expertise with cell phone navigation and their health status were surveyed and related to performance outcomes.

Basically it could be shown here that small screen devices do have a great potential in monitoring users' disease and therefore should be investigated in greater detail in the near future. The present study was one first step in this direction.

4.1 Impact of user characteristics on navigation performance

The study confirmed the large impact of user characteristics on small-screen-device menu navigation performance. The first influential factor found in the analyses was the user's age. Users technical expertise also showed positive influence on users effectiveness and efficiency. User age in particular stood out to be the best predictor of navigational performance. Especially users that are older than 61 years show drastically inferior navigational performance. They tend to make more navigational errors, require more time between each interaction and are less effective in solving the tasks at hand.

Domain knowledge and health status show no significant influence onto the measured performance criteria, but comparisons of means denote a correlation might exist. Thus, we can assume that the navigational performance is indeed facilitated if users show a high knowledge in both, computer experience and disease-related knowledge. The fact that we could not statistically confirm this on the significance level set is presumably due to the comparably small sample size. Future studies will therefore examine the relationship between the computer and disease-knowledge by enlarging the participant group.

4.2 Potential applications and limitations of this study

The findings underline earlier research regarding usability and aging. Further research is required to prove or increase understanding of influence of domain knowledge or diabetic status on user interaction, since findings of influence of age on performance are being studied. Further analysis of task related problems and identification of required neuropsychological characteristics for different tasks might lead into further input for further research. In this context, the comprehensibility of UI component labeling is of interest, as well as the investigation of the underlying mental model of device usage, which also could have impacted performance. Finally, individuals' coping styles should be incorporated into research scope.

However, the findings as promising as they are, also have to be looked at critically, especially as the participants here represented a kind of best-case scenario, which may not represent the whole group of ill and disease-limited patients.

1. Older non-diabetic users were recruited through social networks and are not representative in regard of total population. A best-case homogeneous user group might have led to skewed findings compared to different populations. Older users were all mentally fit, of relatively high education and mostly all of them had experience with computers.
2. All diabetic participants were highly interested in contributing to advancement of usability of diabetes small-screen-devices and thus highly motivated and to try out our prototype. Real life application cannot assume such perfect preconditions and must perfectly work even when the user is distressed, afraid or even in a case of emergency.
3. The software we used to simulate a living assistant was a prototype. Certain features that are not implemented yet, might have caused user distraction that would not have - or at least to a lesser extend - appeared in a finished retail product. Although this was not observed in the study during user interactions, perceptual distraction during navigation should be assumed. The UI-components of the device themselves were not all perfectly chosen and will have to be iteratively optimized in the next steps (see [31][32]).
4. Simulating a small-screen-device on a 15" display is a simplification of the situational context, since holding and handling a real small-screen device requires more cognitive and motor load (coordination of both hands). Therefore all performance measures are probably an overestimation of real life performance especially in regard to using fake values and not real user data. Users might be more concerned about using the device correctly and thus be more disturbed by unexpected behavior in a medical device. This sandbox operation might have led to a more carefree approach. This was reported by two older women (69 and 87 years old) who enjoyed trying out something new without fearing to break a device by accidental mishandling. Both agreed, that trying out the same device at home would have caused earlier abandonment of the device due to lower frustration thresholds. Although the opposite effect could also have occurred (i.e. giving up prematurely in order to not fail in front of a 'supervisor') no participant reported any such feelings after the test trial.

References

1. Federal Statistical Office Germany: Germanys population until 2050 – 11th coordinated population projection, Wiesbaden 2006, http://www.destatis.de/
2. Siebert, H. Economic Policy for Aging Societies, Berlin: Springer (2002)
3. Arning, K. & Ziefle, M. Understanding age differences in PDA acceptance and performance, in *Computers in Human Behaviour*, 23, pp. 2904-2927 (2007)
4. Arning, K. & Ziefle, M., Barriers of information access in Small-Screen-Device applications: The relevance of user characteristics for a transgenerational design, In C. Stephanidis & M. Pieper (Eds.)."*User Interfaces For All*": Universal Access in Ambient Intelligence Environments (pp. 117 - 136). Berlin, Heidelberg: Springer, LNCS 4397 (2007)

5. Leonhardt, S. „Personal Healthcare Devices“, in Satyen Mukherjee et al. (eds.), *„AmIware : Hardware Technology Drivers of Ambient Intelligence*“, Chapter 6.1, pp. 349-370, Springer, Dordrecht, NL (2006)
6. Ziefle, M. & Bay, S., How older adults meet complexity: aging effects on the usability of different mobile phones, In *Behaviour & Information Technology*, Vol. 24, No. 5, September 2005, 375 – 389 (2005)
7. Ziefle, M. & Bay, S., How To Overcome Disorientation in Mobile Phone Menus: A Comparison of Two Different Types of Navigation Aids, In *Human-Computer Interaction*, Vol. 21, pp.393-433 (2006)
8. Ziefle, M. & Bay, S., Transgenerational Designs in Mobile Technology, In *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, Vol. 1, pp. 122-141 (2008)
9. German Diabetes Union: Report on Public Health - Diabetes (2007), http://www.deutsche-diabetes-gesellschaft.de/
10. Tililil et al., Guideline of the German Diabetes Association for treatment of Diabetes melliutus type 2, p.177 (1998)
11. Yeung, Sophie E.; Fisher, Ashley L.; Dixon, Roger A., Exploring effects of type 2 diabetes on cognitive functioning in older adults, Neuropsychology. Vol 23(1), Jan 2009, 1-9 (2009)
12. Dixon et. al. Adjustable Gastric Banding and Conventional Therapy for Type 2 Diabetes, *The journal of the American Medical Association 2008,* 299(3), pp. 316--323. (2008)
13. Lin, D.M. Age differences in the performance of hypertext perusal. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting.* Santa Monica: Human Factors Society, pp. 211-215. (2001)
14. Ziefle, M., & Bay, S. Mental Models of a Cellular Phone Menu. Comparing Older and Younger Novice. In S. Brewster & M. Dunlop (eds.). Mobile Human Computer Interaction. Lecture Notes on Computer Science, LNCS 3160, pp. 25--37. Berlin, Germany: Springer (2004)
15. Ziefle, M., Aging and Mobile Displays: Challenges and requirements for age-sensitive electronic information designs. *Proceedings of the 9th International Conference on Work With Computer Systems, WWCS 2009*, Beijing, China. (2009)
16. Arning, K. & Ziefle, M., Assessing computer experience in older adults: Development and validation of a computer expertise questionnaire for older adults. *Behaviour and Information Technology,* 27 (1), 89-93 (2008)
17. Parush, A. & Yuviler-Gavish, T., Web navigation structures in cellular phones. The depth-breadth trade-off issue. *International Journal of Human-Computer Studies*, 60, 753 --770. (2004)
18. Westermann, S.J., Individual differences in the use of command line and menu computer interfaces. *International Journal of Human Computer Interaction*, 9, (1995), 183 --198. (1995)
19. Craik, F.I. and Salthouse, T.A., Handbook of Aging and Cognition. Hillsdale, N.J., Lawrence Erlbaum (1992)
20. Ziefle, M., The influence of user expertise and phone complexity on performance, ease of use and learnability of different mobile phones. Behaviour and Information Technology, 21(5), 303-311 (2002)
21. Downing, R.W.; Moore, J.L. & Brown, S.W. The effects and interaction of spatial visualization and domain expertise on information seeking. *Computers in Human Behaviour*, 21, 195-209 (2005)
22. Czaja, S. J. & Sharit, J. Age differences in the Performance of Computer-based work. Psychology and Aging, 8, 59-67 (1993)
23. Morrow, D.; Miller, L.S.; Ridolfo, H.; Kokayeff, N.; Chang, D.; Fischer, U. & Stine-Morrow, E. Expertise and aging in a pilot decision making task. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual meeting,* pp. 228--232 (2004)

24. Ziefle, M.; Schroeder, U.; Strenk, J. & Michel, T., How young and older users master the use of hyperlinks in small screen devices. *Proceedings of the SIGCHI conference on Human factors in computing systems 2007* (pp. 307--316). ACM (2007)
25. Davis F. D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319--340 (1989)
26. Adams, D. A., Nelson, R., Todd, P. A.: Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quarterly*, 16, 227--247 (1992)
27. Schajna, B.: Empirical evaluation of the revised technology acceptance model. Management Science, **42**, 85-92 (1996)
28. Schroeder, U. & Spannagel, C. The Role of Interaction Records in Active Learning Processes. In: P. Isaias, M. B. Nunes, & A. P. dos Reis (eds.), Proceedings of the IADIS Virtual Multi Conference on Computational Information Systems (2005), 99-104.
29. Spannagel, C., Gläser-Zikuda, M., & Schroeder, U. Application of Qualitative Content Analysis in User-Program Interaction Research. Forum: Qualitative Social Research, 6(2), http://www.qualitative-research.net/fqs-texte/2-05/05-2-29-e.htm (2006)
30. Jacareto Sourceforge Repository, http://jacareto.sourceforge.net/
31. Ziefle, M.: Spatial cues in small screen devices: Benefit or handicap? *In Proc. 12th IFIP TC13 Conference in Human-Computer Interaction (INTERACT)*. Lecture Notes on Computer Science, LNCS 5726, pp. 620-633. Berlin: Springer (2009)
32. Ziefle, M., Bay, S.: How to overcome disorientation in mobile phone menus: A comparison of two different types of navigation aids. Human Computer Interaction 21(4) 393--432 (2006)