Supporting Requirements Engineering for Medical Products – Early Consideration of User-Perceived Quality

Jörg Dörr
Fraunhofer IESE
Fraunhofer-Platz 1
67663 Kaiserslautern
+49 631 6800 1601
doerrj@iese.fraunhofer.de

Daniel Kerkow
Fraunhofer IESE
Fraunhofer-Platz 1
67663 Kaiserslautern
+49 631 6800 2154
kerkow@iese.fraunhofer.de

Dennis Landmann
Fraunhofer IESE
Fraunhofer-Platz 1
67663 Kaiserslautern
+49 631 6800 2100
landmann@iese.fraunhofer.de

ABSTRACT
The usability and, more generally, the overall user-perceived quality of medical devices is an important aspect, which is often insufficiently addressed in the corresponding system development activities. Fortunately, the development of new standards like IEC/DIN EN 60601-1-6 is strengthening the focus on usability / user acceptance issues. This paper argues for the need to consider usability and user acceptance issues in early system development phases like the requirements engineering phase. In this paper, an empirically validated new quality model for user satisfaction is described first. The importance of the quality aspects included in this quality model for the medical domain is outlined. Then, the new quality model is used to develop a systematic methodology called *Appraisal and Measurement of User Satisfaction (*AMUSE*)*, which allows using user acceptance information early in system development. The key activities of the AMUSE methodology and typical application scenarios are shown. Further on, the application of AMUSE, which was developed in close cooperation with Siemens Corporate Technology, is demonstrated in a real-world scenario at Siemens Audiologische Technik, a line of business of Siemens Medical Solutions. At the end, the first lessons learned from the application of the AMUSE methodology in this medical domain are discussed.

Categories and Subject Descriptors
K.6.3 [Software Management]: measurement, appraisal of user satisfaction, feature priorization

General Terms
Management, Human Factors

Keywords
User Satisfaction, Quality Measurement, Requirements Engineering, Requirements Prioritization, Product Innovation, Product Management.

1. INTRODUCTION
Medical devices are often used in critical, sometimes life-threatening situations. During emergencies, for example, decisions and actions have to be taken under extreme time pressure and in a stressful environment. The efficient, easy, flexible and especially safe handling of the medical device is an absolute must criterion for supporting physicians, nurses, and other users in successfully performing their job.

With the integration of software into medical devices, the flexibility to realize new and innovative features in a flexible way is immense compared to pure hardware realizations. Additionally, the graphical user interfaces that often come along with more software in a product may allow optimal support for the users of the devices in fulfilling their tasks. Beside the increased flexibility, however, poor usability seems to be a major challenge in the domain. A recent international survey performed by Fraunhofer IESE and associated research partners on the use of software engineering principles in medical device production [2, 3] shows that 96% of the respondents perceive system usability as a decisive quality aspect of their devices. Furthermore, 86% of the respondents perceived the assurance of high system usability as a major challenge, making this the number one challenge within the survey. One third of the medical device incidents reported to the U.S. Food and Drug Administration (FDA) are attributed to usage errors [11] but the exact number is estimated to be much higher. Obviously, there is an extreme need for methods that ensure high usability of medical devices.

Regulatory authorities address usability with several standards and regulations. IEC 60601 1-6:2004 [6], for example, is a collateral standard for the usability of medical devices and specifies requirements on how to analyze, design, and verify the usability of a medical device with respect to its overall safety. ISO 14971 [9] explicitly integrates usability considerations into the risk management process of a medical device, and IEC 62304 [8] addresses usability issues during software development. These standards provide valuable input on essential usability aspects that should be considered during the development of a medical device. However, the content of the standards is on an quite abstract level and it is hard for practitioners to decide how to integrate usability considerations into the (software) development process in a systematic fashion. From our point of view, the requirements engineering phase is a suitable starting point for integrating usability and first thoughts on user-perceived product quality (resulting in user acceptance) into the usual requirements engineering activities.

This paper presents empirically determined comprehensive quality dimensions of user-perceived product quality, which are

arranged in an innovative quality model. This model serves as a means for systematically integrating usability and user acceptance considerations into the development process early on. Using the quality dimensions of the quality model, an innovative approach that supports the early consideration of usability and user acceptance criteria in the development life-cycle, especially in the requirements engineering phase, was developed. This approach allows the evaluation of existing products, the development of prototypes with regard to their user acceptance, and strategic selection of decisive feature of a product. Use of this approach can lead to products with higher user acceptance and higher usability. Furthermore, this paper demonstrates example usages of our approach in a case study performed with a large German medical device manufacturer.

Section 2 addresses the role of standards for the development of medical products. Section 3 shows the new quality model with its quality dimensions and discusses the importance of the quality dimensions for the medical domain. Section 4 introduces the methodology applied for integrating considerations about user-perceived quality in the development process early on. Section 5 presents the results of an application of the model in the medical device domain and lessons learned. Finally, we summarize and conclude in Section 6.

2. THE ROLE OF STANDARDS

As introduced before, industrial standards for medical systems emphasize the effects these systems have on their users. This seems to be a rather traditional perspective, born in traditional medical research. There, the effect of pharmaceutical and non-pharmaceutical treatments on health and well-being are thoroughly analyzed scientifically before they are released to the market. Before the products/treatments find acceptance among practitioners or insurance companies, they have to prove their efficaciousness via controlled scientific studies.

Without a doubt, this tradition aims at keeping patients from risks and harms, but also at protecting practitioners and medical personnel from accidental flaws and existential risks to their profession. Increasing use of technology in the medical domain does not change this motivation. On the contrary: The dynamic possibilities of interactive systems provide even more motivation to anticipate potential threats during the development of these systems.

The effect of systems on users is understood differently by different standards and is generally not limited to patients as traditional users. Standards such as ISO9241 and ISO9126, which are not specific to the domain, generally address all stakeholders involved in the interactive system (product), as there are people who operate, people who get a treatment, people who maintain the system, and among others, people who use the results of the operation performed.

Both these standards address system qualities that affect the user and approach these from a user's perspective. ISO9241 does this by focusing on utilitarian properties such as effectiveness and performance, properties that are commonly subsumed under the term usability. ISO 9126 has a slightly larger-scaled definition, explicitly postulating an abstract level of quality: the so-called Quality in Use (QiU).

The recently published international standard IEC (DIN EN) 60601-1-6:2006 (VDE 0750-1-6) includes all those characteristics defining the "general requirements for basic safety and essential performance for medical electrical equipment". The fact that usability is treated in the amendment of a safety-related standard (IEC 60601-1) indicates the importance of this quality aspect in the medical device domain.

The focus of these standards clearly illustrates the importance of user–oriented medical systems engineering methods that ensure satisfaction, safety, or usability. Unfortunately, besides these standards, hardly any system engineering methodologies can be found that systematically integrate user acceptance issues early in the development lifecycle. In the following, we will present the AMUSE quality model and the AMUSE method, which were founded on well established standards and tailored to the specific needs of an industrial partner, a global player in the medical domain: Siemens MED.

3. USER-PERCEIVED QUALITY

In Section 1 and 2, we motivated the necessity for integrating the consideration of user-perceived quality early in system development. As a basis for considering user-perceived quality, a clear model of what user-perceived quality comprises is needed. In this chapter, we describe and define the concept of the user-perceived quality of interactive systems with a discussion about the role of each quality aspect for the medical domain. We present the AMUSE quality model which is derived by theoretical deduction from user acceptance models [1], based on well-established industrial quality standards [10].

Furthermore, the model was discussed and enriched in workshops with representatives from medical industry [4] and afterwards structured and validated using scientific and quantitative methods (cluster analysis, factor analysis, structural equation modeling) [4]. The AMUSE quality model presented in this section has been embedded into an engineering method (see Section 4), which has been successfully applied in the medical domain (see Section 5).

As we have published earlier [4], the starting point for the AMUSE quality model was a derivative of the Quality in Use (QiU) model suggested in [10]. According to this model, QiU consists of the quality aspects *safety*, *effectiveness*, *productivity*, and *satisfaction*. Theoretical foundations encouraged this initial set of quality aspects. Another important base of the AMUSE quality model is the Technology Acceptance Model (TAM). TAM refers to the overall evaluation of a product regarding attitude conformity and usage expectation from the user's perspective. Thus, TAM is based on theories that confront the expected *usefulness* of a technical product with the degree of difficulty in approaching this usefulness, namely the *ease of use*. In recent investigations [12], intrinsic aspects have also been being considered for TAM. Aligning the effectiveness of the QiU model with the usefulness of the TAM, the productivity of QiU with the ease of use of TAM, and the satisfaction of QiU with the intrinsic aspects of TAM extensions has substantiated the initial AMUSE quality model.

Nevertheless, the most important step for matching the needs of the medical domain was to gain a more elaborated understanding of what each quality aspect refers to in an industrial medical context. Elicitation workshops with Siemens MED led to a detailed instantiation of the above mentioned aspects. The AMUSE quality model derived in this manner covered the individual definitions of the quality aspects, which were given by the ISO QiU standard and TAM, as well as the addition of quality aspects that appeared to be relevant for users and process participants of the products developed by Siemens MED.

Although it is useful to have a very context-specific quality model within a project or company, it is important to build a model that can be applied in a broader context. Its dimensions ought to aim at an orthogonal structure and the definition of its elements should aspire to being shared inter-subjectively among practitioners. Supported by cluster and factor analysis, we reached this goal and designed a validated and generalized quality model. First structural equation analyses uncovered the weights (represented by the strength of the arrows in Figure 1) with which each quality aspect influences overall satisfaction.

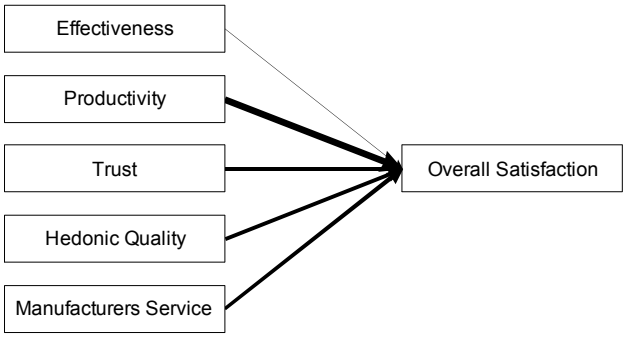


Figure 1: AMUSE quality model with preliminary regression weights (regressions weights correspond to line width).

In the following chapters, we will characterize the conceptual contents of each quality aspect as it is understood in the industrial medical context.

3.1 Effectiveness

Effectiveness describes the accuracy/correctness and completeness with which a planned activity and a planned result can be achieved by using the system (following ISO/IEC 9126 and DIN EN ISO 9000:2000).

Here, completeness and correctness are the main aspects. Completeness means that no additional manual tasks have to be added to an activity even under circumstances such as a worst-case scenario. Correctness addresses the error-free operation of the product and includes the robustness against many possible sources of errors (e.g., readability of serial numbers). Further, both completeness and correctness are required in unusual situations and environments, such as darkness, stressful, unpurified emergency situations, but also in situations where interfaces to other systems are rarely accessible (due to temporal or spatial availability). Effectiveness is also affected by the way tasks are completed, by providing flexible ways and degrees of freedom. Ensuring the correct and complete performance of a task reduces the possibility of unexpected or undesired usage of the medical device and consequently improves the overall safety of the product.

3.2 Productivity

Productivity describes the relation of means used by the users for executing an activity to the yielded results in a specified context (following ISO/IEC 9126) and addresses mainly economical and temporal aspects of the process in which the system is embedded.

Productivity does not only cover “ease of use” from the TAM model, which semantically rather meets the term “self efficacy” (the belief to be able to influence something with own efforts) than the term “usability”, but addresses more generally the effort a user has to spend to successfully perform a task.

Generally, bio-medical systems, especially diagnostic instruments, information systems, or productive systems, are embedded in larger processes and interacting with other activities. Such activities may be administrative decisions, waiting periods, ordering and logistics, storing, documentation, consulting patients, but also technical activities, such as installation, maintenance, or replacement of legacy systems,

Productivity in this sense does not solely have an impact on the “core usage process”, but also affects the whole socio-technical environment in which the medical product or system is embedded.

3.3 Trust

In terms of perceived qualities, we expect safety aspects as proposed in ISO 9126 [10] to be perceived with the meaning of trust. From our point of view, this term represents the aspects of safety and security from the user perspective in contrast to an internal product view induced by the two terms.

A high level of trust means that the user considers the risk of health and financial damages to humans, business, and environment caused by the system to be low.

Users, operators as well as patients, often rely on medical systems. So the most basic and important user-perceived quality aspect might be trust. Especially in the medical domain, trust is not always, but often related to harm to physical health. Preventing harm has a long tradition and high importance in the medical domain, so standards such as DIN 60601-1-6 require thorough prevention of any kind of risk to health and life.

But even those biomedical products that are not primarily a risk to health can influence trust: Many systems are more a threat to economical issues, such as the livelihood of a physician or the return of investment in the business processes of large health-related organizations.

Not considering the dependencies between various business processes, for instance those having impact on the management of supply materials, can bear the risk of a complete stop of servicing. Thus, high-qualitative medical products and systems will also prevent the kind of harm that will be felt by user-perceived trust towards the system, but also towards the operator and towards the manufacturer. Once a professional image is damaged, it can severely harm livelihood and business. High trust in the products used by practitioners is an important business factor in the medical domain.

3.4 Hedonic Quality

Hedonic quality [5] describes the perceived capability of the system to satisfy needs that exceed the pragmatic fulfillment of the primary functionality.

Subsumed under “satisfaction” in ISO 9126 and encapsulated within “intrinsic aspects” in TAM, hedonic quality means that a product, besides providing functionality, also has affective effects on the user and operator [6]. Hedonic quality means that a product satisfies the human needs for stimulation (i.e., personal growth, an increase in knowledge and skills) and identification (i.e., self-expression, professionalism).

Indeed, our analyses have proven the relevance of hedonic quality for the medical domain. Products must be visually appealing and provide the capability to experimentally explore their functions without risking any damage. Stimulation derived from the capability to perform innovative, demanding tasks is

welcome. The latter partly refers to the interplay of expertise, professionalism, and growing image and recognition among medical practitioners. This scales from the autonomous physician who builds up patients' loyalty in this way, towards management (head physician) in large health related organizations, where state-of-the-art methodology and high-tech systems demonstrate supremacy to the public.

3.5 Manufacturers' Service

Finally, another important quality aspect is the quality of service provided together with the product and how the manufacturer of the system is perceived.

Our analyses have shown a high impact on overall satisfaction if users got the impression that manufacturers had invested hard work into fulfilling users' expectations (about quality and lifespan) and surrounded the product with excellent services and support, such as efforts to facilitate the learning of a system for the users.

4. EARLY CONSIDERATION OF USER-PERCEIVED QUALITY IN THE SYSTEM ENGINEERING PROCESS

In product improvement or product invention, it is important to know which features to implement or improve in a future release. Usually, there are more features to select from than it is feasible and economically rewarding to implement. Especially in the medical domain, where innovation is a major success factor for a company, it is an essential but hard guess what accumulation of features will contribute the most to the users' satisfaction. One typical traditional approach is to develop different prototypes and identify a combination of features that evaluates the best against some quality measure. But developing different prototypes is an expensive approach in terms of money and time. Another problem is that the heuristics behind the selection of features is often an unstructured mixture of the product manager's experience, some direct feedback from the customers, issues that came up during training of the users or during deployment to the customer.

An advantageous approach that supports requirements engineers and product managers in finding the most promising features with regard to user satisfaction, and that can be used early on during system engineering is the AMUSE method. It supports the engineering of user satisfaction into products by systematically considering quality information given by end users during the requirements engineering phase. The goal of applying the AMUSE method is to contribute information about user satisfaction to the system development lifecycle in order to "build" user satisfaction into the product. The AMUSE method is clearly focused on getting information from users and assessing their satisfaction with the product. The quality aspects that are used to appraise and measure user satisfaction are directly derived from the AMUSE quality model described in Section 3. This usage of the quality model for measurement and appraisal is one of the key concepts for enabling the early consideration of user-perceived quality.

With the AMUSE method, we have developed a methodology offering three major activities that can be combined in various scenarios to analyze user satisfaction at an early stage of product development and measure user satisfaction on intermediate or final products (see Figure 2). One activity is the measurement of features that reveal current user satisfaction with a product on a fine-grained level, corresponding to the quality aspects from the AMUSE quality model (see Section 3). Another activity is the systematic appraisal of features with regard to the quality aspects from the quality model. The last activity is the prioritization of features, which uses the measurement results and appraisal results. In the following, we will describe each of these three activities.

4.1 Measurement

Measurement takes places if a product already exists and the users' satisfaction with the product is assessed. The goal of measurement is to assess how satisfied users are with the product by evaluating the perceived product quality. Therefore, the

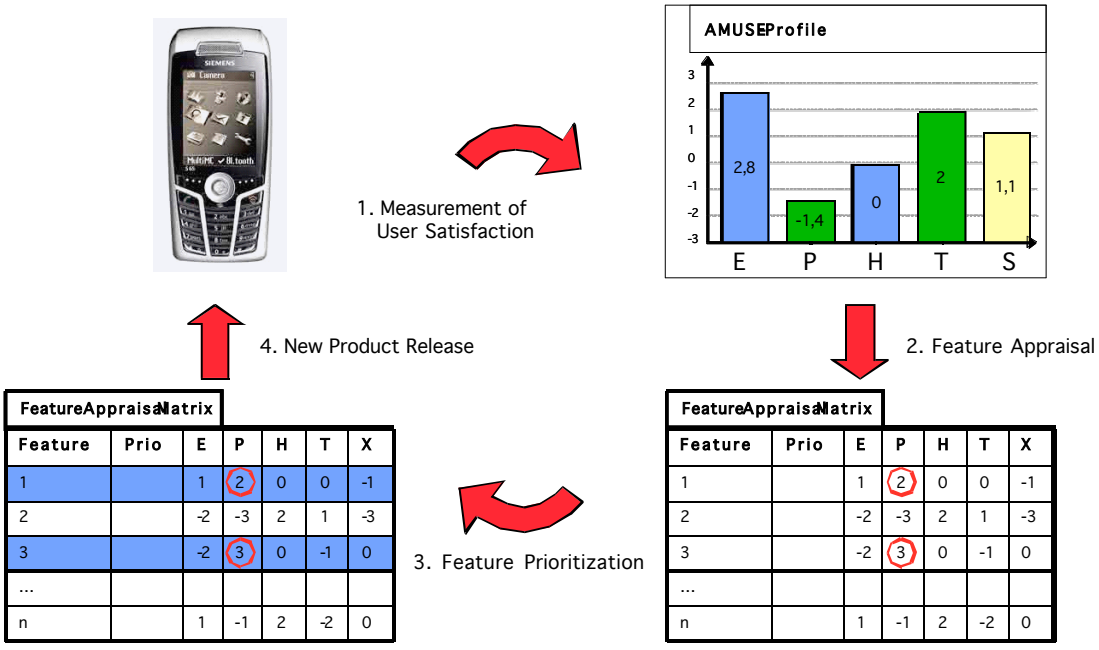
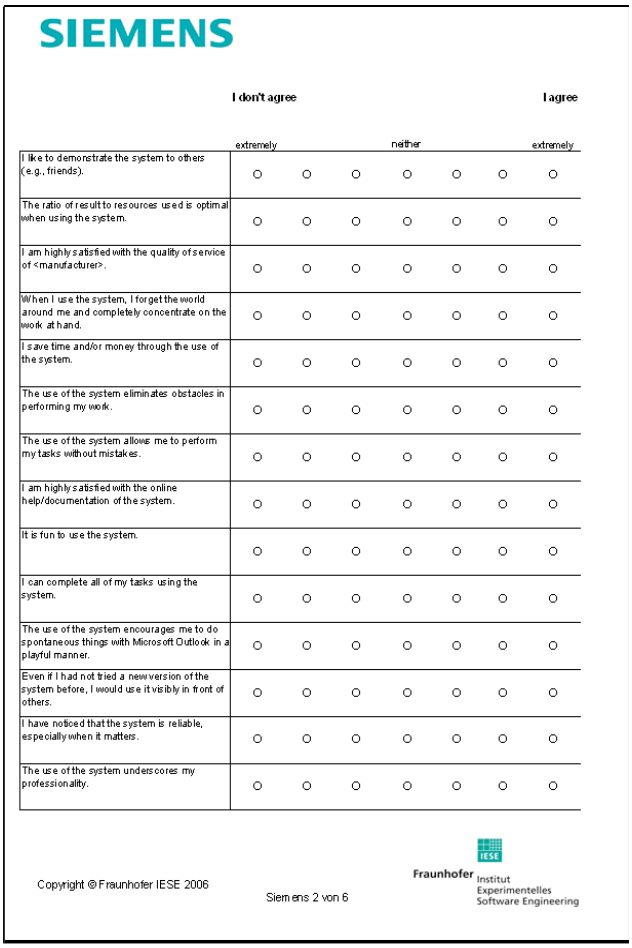


Figure 2. Typical application scenario of AMUSE: Systematic Product Enhancement

standardized and validated AMUSE questionnaire (see Figure 3) is given to a representative group of current product users. The results of the survey are aggregated into a so-called AMUSE profile. This profile represents the perceived product quality according to the AMUSE quality model (see Section 3). It gives a white-box view on which quality aspect of the product the users value as well or poorly developed. This indicates how well the users are satisfied with the product on a level that is fine-grained enough to apply this user satisfaction information early on in the system development process.



SIEMENS

	I don't agree extremely			neither			I agree extremely
I like to demonstrate the system to others (e.g., friends).	○	○	○	○	○	○	○
The ratio of result to resources used is optimal when using the system.	○	○	○	○	○	○	○
I am highly satisfied with the quality of service of <manufacturer>.	○	○	○	○	○	○	○
When I use the system, I forget the world around me and completely concentrate on the work at hand.	○	○	○	○	○	○	○
I save time and/or money through the use of the system.	○	○	○	○	○	○	○
The use of the system eliminates obstacles in performing my work.	○	○	○	○	○	○	○
The use of the system allows me to perform my tasks without mistakes.	○	○	○	○	○	○	○
I am highly satisfied with the online help/documentation of the system.	○	○	○	○	○	○	○
It is fun to use the system.	○	○	○	○	○	○	○
I can complete all of my tasks using the system.	○	○	○	○	○	○	○
The use of the system encourages me to do spontaneous things with Microsoft Outlook in a playful manner.	○	○	○	○	○	○	○
Even if I had not tried a new version of the system before, I would use it visibly in front of others.	○	○	○	○	○	○	○
I have noticed that the system is reliable, especially when it matters.	○	○	○	○	○	○	○
The use of the system underscores my professionality.	○	○	○	○	○	○	○

Figure 3. Excerpt from the AMUSE Questionnaire

The measurement results are a snapshot of the current user satisfaction. With measurements, user satisfaction can be monitored during the whole development. Measurement is preferably carried out during early development of a new product release. From the results, the goals for the feature prioritization step can be derived, e.g., features that improve the weakest quality aspect should have high priority.

The AMUSE questionnaire currently comprises 28 questions and has been validated in several studies. More details on the creation and validation of the AMUSE questionnaire and the underlying AMUSE quality model can be found in [4].

4.2 Appraisal

In this AMUSE activity, a list of features that are candidates for the next releases is appraised. The features are rated with regard to how well they will improve the particular quality aspect of the new product. We want to point out that during this appraisal, the contribution of a feature is rated not on an absolute basis, but relative to an existing product in use or a current situation without the support of a medical product. A tool-supported rating system is used for this activity instead of the AMUSE questionnaire, with the rating items being correlated to the items in the questionnaire via the AMUSE quality model (see Section 3). Typical examples of contributions of features to quality aspects are:

- A new dedicated function button on a medical device (functioning as a shortcut for a function selection) would increase productivity,
- 3D graphical representation of some data could increase hedonism,
- An added functionality rendering obsolete a manual task could increase effectiveness,
- A new encryption feature for sensitive data could increase trust (but might decrease productivity).

The output of the feature appraisal step is a feature appraisal matrix, where each feature has a value for each AMUSE quality aspect (see Figure 2).

This appraisal can be done early in the system development process, even before any prototype is set up. The reliability of and the effort for appraisal are scalable via the appraisal depth, the number of people involved in the appraisal, and the appraisal method chosen.

Appraisal depth refers to the level on which each feature is appraised: The most course-grained level (5 dimensions) is the level of quality aspects, the most fine-grained level is the level of low-level quality items (28 items) that specifically ask for the contribution of a feature under consideration to a specific quality. Quality aspects are defined by the quality model, but leave the interpretation of what is exactly asked for open to the user. The quality items basically correspond to the questions in the AMUSE questionnaire.

Answering all the questions is tedious and may not be applicable for all users of the questionnaire. During appraisal, the appraising person can choose for each feature and for each quality on what level he or she wants to answer a question according to how much guidance and support is needed. The answers are then automatically aggregated into the higher levels. That can greatly limit the effort for appraisal. In the end, each quality aspect can be quantified by a number representing its estimated influence on the perceived quality of the product. The relations between the estimates provide a hint on which aspect of the product should be emphasized during further development.

With regard to the number of people, a single appraisal tends to reflect the subjective perception of the expected user satisfaction by only one person. To further increase validity, objectivity, and reliability, one could make use of multiple single appraisals executed by several persons independently. The results would then be combined via aggregation mechanisms or discussion. Alternatively, group appraisals where several persons with different roles cooperatively appraise all features in a dedicated session can be used.

The last mechanism used for scaling the needed effort and reliability of the appraisal is the selection of the appraisal method. AMUSE basically offers two methods: estimation and counting. During estimation, the experts conducting the appraisal estimate the values for the quality dimensions directly on the feature. In contrast to this, the counting procedure uses predefined checklists that ask the experts to specify further information (e.g., what are

the tasks in which the user would use the feature) or ask the expert to calculate objective values (e.g., how much time would be saved by using the feature) for the various quality dimensions. By forcing the experts to think about this requested information, the reliability of the counted values can be improved, but the effort for this appraisal method is significantly higher.

4.3 Prioritization

AMUSE does not prescribe a specific priorization technique. To prioritize features in AMUSE, the information obtained from the measurement and appraisal activities can be incorporated into existing prioritization techniques, e.g., into prioritization according to Karlsson [6], Wiegers [13], or into an ABC Analysis. An AMUSE prioritization is usually a multi-dimensional prioritization that ranks features according to the predefined dimensions. These can be their prospective impacts on user satisfaction with the final product represented by the single AMUSE quality aspects or aggregated into one user satisfaction value. Further dimensions that are usually taken into account are implementation cost or the risks associated with the implementation of the feature.

The output of this step is the list of selected features that will then be incorporated into the next product release.

4.4 Combining the AMUSE Activities in Application Scenarios

Product managers and requirement engineers can have various objectives for wanting to obtain or use information about user satisfaction. The above-mentioned AMUSE activities can be used as stand-alone or be combined in several application scenarios, including:

- *Get snapshot*: The product manager obtains the AMUSE profiles (i.e. measurement results of user satisfaction) for one product. With this information he or she can understand in which quality aspects the product is perceived to be strong and weak.
- *Compare two or more products*: The product manager obtains the AMUSE profiles for at least two products (e.g., own and competitor product). By comparing these, product managers can get new insights into the difference between products and obtain information on the quality and quantity of that difference. This scenario can also be used to compare two product versions, for example before and after the introduction of new features.
- *Appraise and select features*: The product manager appraises the quality of single features in order to then prioritize and select the most beneficial features for the next implementation.
- *Product enhancement*: The product manager uses all three activities to continuously improve user satisfaction with a product by measuring and controlling the contribution of appraised single features to the quality of the whole product. One iteration of this scenario is depicted in Figure 2.

4.5 General Expected Benefits of Using AMUSE

As shown in the last paragraph, prospective appraisal and retrospective measurement can be employed in different ways – each separately in its own right, combined as pair, or in any way that seems reasonable. One single appraisal can be useful for identifying early in the development what features will probably satisfy future users the most. This valuable information will help management to make the right decisions for a successful future of the product. One single measurement can be useful for seeing how an already existing prototype or product really satisfies the users at a given time. The measurement of the qualities can show how the quality of the whole product is perceived and how satisfied the users are. A time series of measurements, i.e., done repeatedly with different versions during the product lifecycle, can show how quality has developed over time. Measurement can be employed in combination with appraisal to see if the development has actually enhanced the qualities of the product that were identified during appraisal. Prioritization helps to select the right features for the next product release in order to ensure high quality without neglecting other goals.

With AMUSE, the end-users play a significant role for the development process, more than with other methods. Based on the evaluation with AMUSE, product managers and project managers know how their users experience the product and its qualities. With the AMUSE method, they are supported in

- thoroughly understanding how well the user is satisfied with the product,
- taking care of user satisfaction early in the development process,
- identifying weak quality aspects to focus efforts in next releases,
- making a deliberate, strategic choice of which features to include in the next product or product version,
- reducing development efforts for features that have little influence on user satisfaction,
- justifying better to management and development why features were selected, and
- releasing products that satisfy their users.

5. APPLICATION IN THE MEDICAL DOMAIN

We applied the AMUSE method within Siemens Audiologische Technik (S.A.T.), a line of business of Siemens Medical Solutions. S.A.T. produces a variety of products in the area of hearing systems, including hearing aids, hardware and software systems for diagnostics, measurement and fitting of hearing aids, and supply chain software products.

5.1 Setup

In order to achieve a maximum of feedback on the validity and also on the utility of the AMUSE method, we joined a product development team responsible for the complete set of products mentioned above. The team consisted of one product manager and four engineers with different responsibilities, e.g., for requirements engineering, prototyping, or usability issues. All members had sound knowledge of the product and its customers on the German and American markets. The application for this product should provide an overview on the customers reception of the product. As only few customers exist for this medical product in Germany, the application here cannot be seen as full case study, but more as a subjective evaluation. However, we will use the term case study for simplicity reasons.

5.2 Scenarios Applied in the Medical Domain

The methodology was applied within several setups with different goals. Each of the scenarios had a specific utility for the software engineering and management process.

5.2.1 Scenario Get Snapshot

In a first application, we started a measurement campaign to evaluate the current user satisfaction regarding one specific product. The product was a scanning device for ear impressions, which the audiologist can use to scan a silicone impression taken from a patient's ear. The 3-dimensional scanning data is then sent to the hearing aid production facilities together with an audiological characterization of the patient.

The goal of the first study was to examine the user perception of these scanners in order to get an overview of the specific perceived strengths and weaknesses regarding user-perceived product quality. A measurement campaign was designed using the AMUSE questionnaire extended by questions concerning the installed system landscape. The questionnaire was completed by 8 customers in Germany. The evaluation revealed a positive perception with lacks in hedonic quality. As we designed the AMUSE questionnaire to be able to measure hedonic quality independently from a specific domain (i.e., the questionnaire can be used in a consumer product domain as well), the expectations were met, since the scanner is a productive tool not designed with a specific focus on hedonic quality.

5.2.2 Scenario Appraise and Select Features

The second case study of the method consisted of two appraisal workshops where new candidate features were estimated regarding their contribution user satisfaction. The first workshop was conducted mainly to increase familiarity with the appraisal procedures and to gather some initial feedback on the method. Therefore, we selected 15 features of similar abstraction levels (e.g., 3D display of scanned data) and compiled a feature appraisal matrix. The participants were then asked to count and estimate (see Section 4) the features and estimate their contribution to user satisfaction. After an initial level of confidence in the method was reached, a further workshop was held for estimating 25 features of a future product. The software was a new module for ordering customized spare parts for hearing aids. The individual results were finally averaged, and a group discussion was initiated. The features were then prioritized according to their contribution.

5.2.3 Scenario Product Enhancement

A more widespread scenario was the execution of a complete measurement and feedback cycle. A measurement campaign was designed consisting of an initial measurement of user satisfaction of a current product, appraisal of features to be integrated, prototypical implementation of the features, and the measurement of user satisfaction induced by the prototype. The initial measurements were done 2 weeks before appraisal of the features and presentation of the prototype to make sure the participant could not bias the results. The current results indicated a slight improvement in perceived quality.

5.3 Empirical Results and Benefits from Application

Independently from specific results, it can be stated that the introduction and adoption of the method could be achieved quickly. Due to the intuitive nature of the elements of AMUSE, all participants were able to perform key activities under supervision after a brief (half a day) coaching. Even if the first application could not produce strong reliability regarding statistical evaluation, the product team perceived the evaluation as quite useful and beneficial, since nearly all customers in Germany could be easily invoked in the study.

5.3.1 Measurement

The participants of the study were able to complete the questionnaire within 10 minutes. The measurement of the prototype revealed a slight increase in user satisfaction. Figure 4 shows a comparison of the results.

The users participating in the measurement were also able to provide a qualified feedback to most of the questions. However, some items of the questionnaire could not be answered by all participants, e.g., questions concerning product life-time or some aspects of hedonic quality. This was caused by our goal to design a domain-independent questionnaire applicable to a wide range of products. Therefore, some of the items in the questionnaire are not applicable to every context.

5.3.2 Product setup

For elicitating user satisfaction, it is essential to be aware of the user's product setup. The scoping of the measurement campaign is an essential step preceding the study. The interpretation of the data has to be done with regard to the user's context, which comprises the configuration of the product itself, the tool chain it is integrated with, and the work process it is used in.

In our case study, the first product measured was mainly used in two different setups: by audiologists in their practice and in manufacturing. Additionally, the integration depths into the work processes were different depending on which product configuration was used, e.g., with or without ordering software. Consequently, we were able to identify slight discrepancies in user satisfaction for all identified groups.

5.3.3 Raise of specification quality

One key benefit of applying AMUSE was an increase in the quality of the requirements specification. The in-deep examinations of features led to several additions to the requirements specification. The appraisal uncovered, e.g., under-specification caused by missing details or rationales. Especially effectiveness and productivity were hard to estimate in case of under-specification. Another effect was the exploration of "intrinsic" and "extrinsic" parts of the hedonic quality and trust of a feature. In our case study, the two aspects were difficult to estimate if the features were not explicitly designed to address those aspects, like a 3D display of scanned data. It turned out to be a benefit to be able to "normatively" specify a target level of quality for the implementation of the features, which will guide the developers. This information was included in the requirements specification, e.g., as an additional quality requirement: "Feature XY has to be designed to be impressive to the user". Through this effect, the appraisal activity can have a similar impact as a document review from a user satisfaction perspective.

5.3.4 Communication & prioritization

The permanent discussion about quality aspects of the product led to a deeper understanding of the rationale of a feature. Thus, the participants of the appraisal workshops were able to precisely derive prioritization goals from earlier measurements. In a productive environment, we expected a strong focus on

effectiveness and productivity, while hedonic quality was expected to be a "marginal" phenomenon. As the measurement then uncovered a lack of hedonic quality, the product team was capable of a) precisely selecting features contributing to hedonic quality, and b) communicating the rationale of their decision to their management effectively, justifying the need for features increasing hedonic aspects.

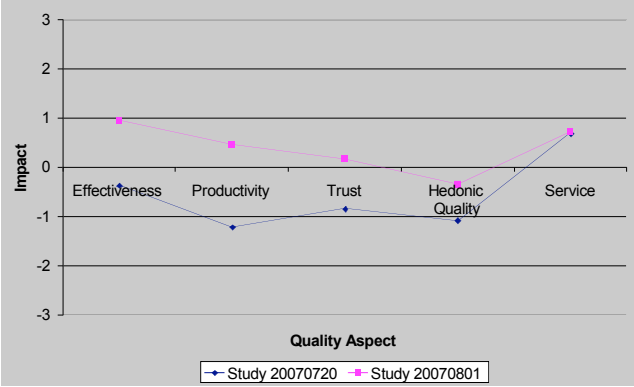


Figure 4: Comparison of user satisfaction regarding product and prototype

5.3.5 Discovery of key satisfiers

The appraisal also led to a fine granular distinction between user-oriented features (e.g., 3D display) and "must-features" necessary for backend operations. It was then possible to keep the perceived negative impact of "must-features" on user satisfaction as minimal as possible, while maximizing the visibility of key satisfiers contributing the most to user satisfaction. By introducing the key satisfiers found, our partner will be able to increase the usability of the system substantially as the new features will be closely aligned to the user's actual needs.

5.3.6 Usability improvement

The consequent orientation towards user-perceived satisfaction led to a higher accentuation of usability in general. The product team focused on user satisfaction, and therefore reasoned about different realization alternatives of features in order to maximize user satisfaction. Especially features with a high impact on the effectiveness of the software enable the user to perform new tasks, while features with a high impact on hedonic quality create high user satisfaction. The harmonization of features with user expectations thus automatically leads to higher usability.

5.3.7 Cost effectiveness

The costs of the application had a sound relation to the results generated by AMUSE. Design of the measurements (i.e., adding a short context description to the generic questionnaire), distribution of questionnaires, and evaluation of the questionnaires took less than one day. The completion of a questionnaire takes about 10 minutes and has to be done by the product users.

The highest effort was spent on feature appraisal, where we had two half-day workshops with three to four participants of the product team. The participants had to be familiar with the requirements specification and the user situation, which in our workshops was given (if not, additional effort would have been spent here). We recognized a significant increase in efficiency from the first workshop (5 features estimated in 2 hours) to the second (25 features in 2 hours), and also an increase in reliability by having a group decision instead of individual appraisals.

6. SUMMARY AND FUTURE WORK

This paper presents an innovative approach to considering usability and user satisfaction aspects early in the software development process. Through the definition of an explicit quality model for user satisfaction and its integration into a well-defined and scaleable process, usability and user-perceived quality become explicit to all stakeholders of a medical device. With the AMUSE method, the user perspectives are carefully considered right from the start of the development and user-perceived quality becomes an integral part of the complete development cycle. Hence, user satisfaction can be increased and errors caused by poor usability of the devices can be reduced. Especially defects that are caused by incomplete and underspecified requirements can be reduced by the application of AMUSE.

Besides these end product related quality improvements, the application of the method also shows benefits with respect to the overall development process. The white-box view on user satisfaction and the early involvement of the user's perspective on the medical device during the requirements phase is beneficial to the product team. With this knowledge, the road- mapping process is supported through measurement and visualization of user satisfaction and estimation of feature contributions. This provides more control over the development process to the product team and finally leads to products aligned with user expectations. Although the concepts and techniques of AMUSE were new to the product team, it was possible to generate results quickly.

The results of our case study are promising with respect to the improvement of user satisfaction and usability in the context of medical device production. In future work activities, we aim at the extension of the method to medical devices that are of higher safety criticality compared to those used in the initial studies. We are convinced that the principles of the AMUSE method can be applied to any type of device. We aim at a systematic investigation of the AMUSE method and the benefits that are achievable in future empirical studies.

7. ACKNOWLEDGMENTS

We are grateful to all S.A.T. members who have supported this project, and especially Eduard Kaiser, who enabled and supported the application of AMUSE in his development group. Thanks are also due to Klaus Beetz and Peter Amthor of Siemens Corporate Technology SE 1 for granting us great freedom in research and application. We would also like to thank Susanne Hartkopf, Kirstin Kohler, Markus Ciolkowski and Sonnhild Namingha of Fraunhofer IESE for their support.

8. ADDITIONAL AUTHORS

Additional authors: Christian Graf (University of Hamburg, email: graf@informatik.uni-hamburg.de), Christian Denger (Fraunhofer IESE, email: christian.denger@iese.fraunhofer.de), Anne Hoffmann (Siemens Audiologische Technik, email: anne.hoffmann@siemens.com).

9. REFERENCES

[1] Davis, F. D. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13, 3 (1989): 319-340.

[2] Denger, C., Feldmann, R., Höst, M., Lindholm, C. and Shull, F. A Snapshot of the State of Practice in Software

Development for Medical Devices, In *Proceedings of the 1st International Symposium on Empirical Software Engineering and Measurement*, to appear, 2007

[3] Denger, C., Feldmann, R., Höst, M., Lindholm, C. and Shull, F. *State of the Practice in Software Development for Medical Device Production*, Fraunhofer Institute for Experimental Software Engineering, Kaiserslautern, Germany. IESE Internal Report, 2007. (to obtain by request to the authors)

[4] Dörr, J., Hartkopf, S., Kerkow, D., Landmann, D. and Amthor, P. Built-in User Satisfaction: Feature Appraisal and Prioritization with AMUSE. In *Proceedings of the 15th IEEE International Requirements Engineering Conference (RE 2007)* (New Delhi, India, October 15-19th, 2007). IEEE, to appear.

[5] Hassenzahl, M., Platz, A., Burmester, M. & Lehner, K. (2000). Hedonic and ergonomic quality aspects determine a software's Appeal. *Proceedings of the CHI 2000 Conference on Human Factors in Computing*, The Hague, NL (pp. 201-208). New York: ACM.

[6] Karlsson, J. and Ryan, K. Prioritizing requirements using a cost-value approach. *IEEE Software*, 14 (1997): 67-74.

[7] IEC 60601-1-6 Medical electrical equipment Part 1-6: General requirements for safety - Collateral standard: Usability

[8] IEC 62304 Medical Device Software - Software Life Cycle Processes

[9] ISO 14971 Medical Devices - Application of risk management to medical devices. Revision of first edition (ISO 14971:2000) and ISO 14971:2000/Amd.1:2003, Geneva, 2005

[10] International Organization for Standardization, International Electrotechnical Commission. *International Standard ISO/IEC 9126. Information technology -- Software product evaluation -- Quality characteristics and guidelines for their use*. Geneva, 1991.

[11] Patterson, P. Fitting Human Factors in the Product Development Process. *Medical Device & Diagnostic Industry*, 28, 1 (2006): 124–133.

[12] Venkatesh, V. Creating Favorable User Perceptions: Exploring the Role of Intrinsic Motivation. *MIS Quarterly*, 23, 2 (1999): 239-260.

[13] Wiegers, K. E. *Software-Requirements*. Microsoft Press, 2005.