USER-CENTERED ANALYSIS OF FEEDBACK OPERATIONS FOR QUALITY IMPROVEMENT IN NEW PRODUCT DEVELOPMENT

Aylin Koca
Aarnout C. Brombacher

Business Process Design, Faculty of Industrial Design
Eindhoven University of Technology, The Netherlands

ABSTRACT
The context of operational processes has actively been changing since the advent of the digitization era. Current field feedback operations are not well suited to the dynamic and uncertain context of today's operational processes: They tend to serve the purpose of closing *rigid* control loops via measuring certain *prescribed* "key performance indicator"s to determine product quality. However, the modern context requires a more *flexible* extended approach to instigate the *exploration and identification* of new dominating context parameters and user-centered criteria, which drive today's global business success. In this paper, we advocate the latter approach: We use a conceptually- and empirically-grounded feedback analysis model we developed, in order to mine and discover patterns among the various feedback data sources compiled from real users' feedback, the results of which are intended to iteratively drive the corresponding refinements in operational processes at large. Furthermore, we present and discuss findings from two comparative industrial cases.

Keywords: user feedback, failures, quality improvement

INTRODUCTION
The context of operational processes has actively and dramatically been changing ever since the advent of the digitization era. As a consequence, today's businesses are prone to being both empowered by, but also victimized by, the accessibility of various technological and market opportunities. For instance, the merging of digital technologies allows for the development of very innovative, multifunctional and adaptive products that are capable of reaching wider markets in the global arena. However, especially the lean and agile operations in the development of such products also yield a pitfall: There are huge uncertainties, and little timely indicators, about the deployment of these products in the field, where there are too many new and unknown parameters (e.g., various users with changing expectations, networked devices, etc.). To avoid the pitfall, relevant and fast field feedback operations become more important than ever.

Current field feedback operations are not well suited to the dynamic and uncertain context of today's operational processes. They traditionally use "key performance indicator"s such as field return rates and repair of parts information. While these indicators are useful for sales and spare parts operations management, they do not reveal the necessary information for user-centered product quality assessment and improvement. Therefore, these indicators prove irrelevant to manage the growingly larger part of the field feedback about new high-end products, which is not due to broken hardware or software components, but due to broken expectations of users (Bly et al., 2006; Koca and Brombacher, 2008). Typically, such field reports get dealt with superficially on a per customer basis, and afterwards get piled up in a "No Fault Found" stack, without leading to any further action for product quality improvement. The increasing severity of this industrial problem about "soft reliability" management (Koca and Brombacher, 2008), calls for refined feedback operations oriented towards systematic capturing of *user-centered* feedback, rather than only *logistics-centered* feedback. Addressing this need will ultimately help in better diagnosis and treatment of each individual user feedback both at the (i) back-end "aftercare" activities, and at the (ii) front-end "design" activities during new product development processes.

In our ongoing collaboration with a global consumer electronics company, we identified a direct root cause for the described industrial problem as the incompatibility between the traditional approaches to feedback operations and the modern context of operational processes. More specifically, the traditional feedback operations tend to serve the purpose of closing *rigid* control loops via measuring certain *prescribed* "key performance indicator"s to determine product quality. However, the modern context requires a more *flexible* extended approach to instigate the *exploration and identification* of new dominating context parameters and user-centered criteria, which drive today's global business success. As a result, the incompatibility between the traditional operations approach and the modern context can be formulated as the difference between the currently deployed "measure-and-control" paradigm and the needed "explore(and model)-mine(and discover)-and refine(and control)" paradigm. In this paper, we advocate the latter paradigm: We use a conceptually- and empirically-grounded feedback analysis model we developed, in order to mine and discover patterns among the various data sources compiled from real users' feedback, the results of which are intended to iteratively drive the corresponding refinements in the feedback operations and in operational processes at large, to enable the desired outcome of a relevant feedback mechanism that is empowering both the users and the businesses. Furthermore, we present and discuss findings from two comparative industrial cases, to which we apply our approach.

The paper is organized as follows: The next section, 'Feedback Operations and Data Repositories', provides an overview of the operations context for various user feedback channels, and indicates possible feedback data repositories. Then the 'Feedback Lifecycle' section takes a failure-focused view (i.e., from failure origination during development activities, on to its revelation during use, and finally its utilization during improvement activities) of a generic feedback instance to explain how the product-use context should inform the product-development context for more effective quality improvements. This view is further developed, and augmented to include also positive user feedback, in presenting our 'user-centered feedback analysis and classification model'. In the section 'Industrial Case: Two Comparative Studies', two comparative studies on real feedback data collected from different repositories are described, together with the results of applying the advocated approach in this paper. The final section concludes the paper.

FEEDBACK OPERATIONS AND DATA REPOSITORIES

Feedback operations, and the feedback data repositories involved therein, are different in the case of products that have been released to the market and in the case of products that are still under development. Hereafter, we denote the former as field-feedback operations, and the latter as test-feedback operations. This section outlines both kinds of operations, as observed at our collaborator industrial site, a global consumer electronics company.

Field-Feedback

Field-feedback operations (Figure 1) typically get initiated by real users in the field, whose expectations from or perceptions of the product differ from those of the designers'. Depending on both the degree of disconfirmation the user experiences while using the product, and also other individualistic factors, the user may like to (i) hand in the product to a nearby authorized Service Center for repair; (ii) contact a designated Call Center for questions, complaints, and requests; or (iii) post questions, complaints, and general comments on official or unofficial user forums on the Internet. In these cases, the respective feedback information gets recorded in an electronic data repository in distinctive forms, later to be analyzed by the relevant members of the product development team, such as the "Knowledge Engineer", the "Product Manager" or the "Quality Manager" for product quality assessment and improvement.

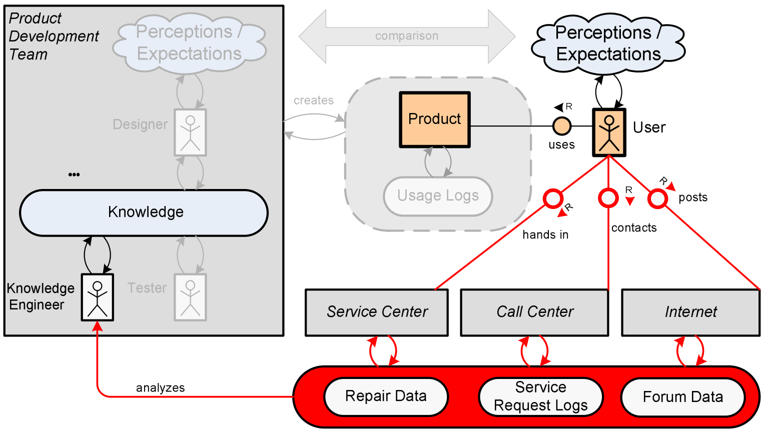

Figure 1 *Field-feedback operations*

The bookkeeping of the feedback data at Service Centers is customarily done by Service Center engineers, who record the defect and repair data of returned products into a central repository. While there exist extensive universal code lists to utilize for standardized registry format of such feedback data (e.g., IRIS: International Repair Information System), the primary purpose in this bookkeeping is two-fold in effect: First, to keep track record of defective product parts that have been repaired or replaced; and second, to claim costs from upper management, for those products treated under warranty. In (Koca et al., 2007a), we had outlined an illustrative case about how a Service Center functions, and the structure of the electronic repository deployed there.

At Call Centers, the feedback data are logged by call center agents as "service request"s from customers. Service requests may comprise phone calls or emails that are initiated by customers, and that get treated by call center agents (or get forwarded to a Service Center for hands-on repair, when necessary), sometimes iteratively over a period of time. While logging service requests into the Call Center's electronic repository, certain classification schemes are followed. The primary purpose of these classification schemes, and the overall logging of service requests by agents, is also two-fold in effect: First, for assessing and improving customer relationship management; and second, for identifying the *surfacing* "cost of non-quality" *of services* (and not getting down to that *of the product*), in order to cut down such costs. Two illustrative examples of currently deployed classification schemes are compared and contrasted in (Koca et al., 2007a), where their performance shortcomings are demonstrated.

In practice, the aftercare services of products as provided through Service or Call Centers are generally outsourced to domain-specialist organizations. Due to both the communications indirectness introduced via the contractor-subcontractor relationship (i.e., as opposed to direct in-house communications) and the driving incentive for such relationships to increase short-term contractual profits (e.g., responding to as many customers in as little time as possible), long-term business goals are often neglected. Consequently, systematic user feedback collection for the ultimate purpose of assessing and improving the user-perceived quality of products and related services cannot be achieved. This is discussed further in the "Feedback Lifecycle" section.

Internet feedback data repositories refer to both officially and unofficially maintained Web-based forums, where users share and exchange their questions, complaints, and general comments with their peers or with the manufacturer company's officials, by means of posting messages. Since these repositories are compiled by plain users who log their own feedback without having to conform to any codification or classification scheme; and due to the abundance of such repositories on the Web that enable sharing different modalities of user feedback about a product, the resulting data are often quite unstructured and dispersed. Nonetheless, as the Internet has been developing into "the ultimate communications medium" also among users of especially innovative, multifunctional and adaptive products; designers and developers are also starting to tap on this resource, although currently off the record.

Test-Feedback
Test-feedback operations (Figure 2) are initiated (i) when test participants (i.e., real trusted-users or potential users) provide *subjective* feedback about their use experience with the tested product; or (ii) when the tested product automatically, and hence *objectively*, logs certain use patterns of the user. Both kinds of processes tend to be restricted in either the numbers of test participants or tested products; or their purpose to test only specific aspects of the tested product and not the product as a whole. Therefore, there are usually no standards in collecting and analyzing the resulting data.

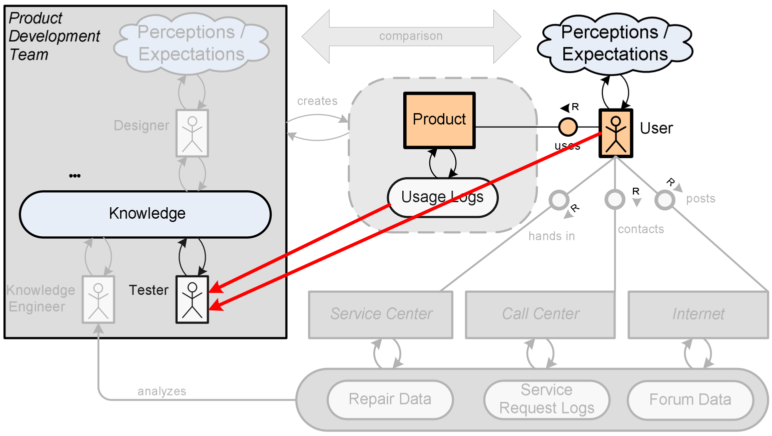

Figure 2 *Test-feedback operations*

Both the subjective and the objective data generated as the outcome of various user tests get analyzed by the "Tester", based on his specific interests in his field of expertise. Therefore, the results of such analyses are not relevant for all other stakeholders involved in product development, and also cannot be consistently scaled across projects (Gorlenko and Englefied, 2006; Koca and Brombacher, 2008).

An updated *unified user-centered* approach to gathering and analysing *all* feedback data, which addresses the presented shortcomings of current operations, is discussed in the next section.

FEEDBACK LIFECYCLE

Innovative, multifunctional and adaptive products of today are complex real systems. Such complex real systems, made *up* of other systems (cohesively integrated in their use context) and made *by* other systems (in their development context) -such as hardware, software or people-, evidently fail from time to time, prompting feedback operations. However, reducing the frequency and severity of failures is a major challenge for quality improvement as discussed previously. This challenge relates to (i) the difficulty of understanding and identifying individual "failures" in relation to their possible causes and consequences, i.e., the concepts of "faults" and "errors", which may be a subtle and iterative process; (ii) the uncertainty about actual system boundaries and specifications; and (iii) the involvement of a possibly inconsistent "judgmental system" that determines, from its viewpoint, a "failure". We now draw on these three challenge factors, in order, to later describe how they can be managed via appropriate feedback operations.

Firstly, based on the widely accepted definitions in (Avizienis et al., 2004) for the subtly-distinct concepts of "failure", "fault", and "error", we derive the following: A system *failure* occurs when the interaction with the system deviates from fulfilling the intended. An *error* is the state that is liable to lead to subsequent failure: an error affecting the interaction is an indication that a failure occurs or has occurred. The adjudged or hypothesized cause of an error is a *fault*. Note that a failure can be judged to have occurred when an error passes through the user interface and affects the interaction with the system. In the cases where the failure is significant, it may constitute a fault to the enclosing system. Thus the manifestation of failures, faults and errors follows a "fundamental chain": ... → failure → fault → error → failure → fault → ..., i.e., ... → event → cause → state →

event → cause → ... (Randell and Koutny, 2007). The difficulty of identifying this chain is not only due to the concepts involved, but also to the fact that this chain can flow from one system to (i) another system that it is interacting with; (ii) a system which it is a part of; and (iii) a system which it creates or sustains.

Secondly, the varieties of intricate failure chain flows among interconnected systems are not easy to trace, if at all possible, where complex systems with vague boundaries and specifications are involved. While systems are traditionally assumed to have complete, accurate and agreed specifications, this is not any more the case, due to the so-called "state explosion problem".

Third and lastly, the notion of a "judgmental system": It broadly captures the determining mechanism that concludes whether any particular activity or inactivity of a system in a given context constitutes or would constitute, from its viewpoint, a failure (Randell and Koutny, 2007). Note that judgmental system can be hardware or software, as well as people, e.g., users, developers. Accordingly, the judging activity may be clear-cut and automatic, or essentially subjective –though a degree of predictability is always needed to render the system designers' task possible. Thus, achieving predictability with subjective judgmental systems proposes an important challenge.

In order to address the three challenge factors in failure management for quality improvement, effective feedback systems and operations are essential. For that reason, in Figure 3 we show the ideal abstracted lifecycle of a failure triggered feedback (Koca et al., 2007b; British Standard 5760-8: 1998), and further motivate how user-centered feedback operations can be achieved. Note that while this generic feedback loop is applicable for all failure kinds, in this paper we focus only on "soft failures", the definition of which can be found in (Koca and Brombacher, 2008).

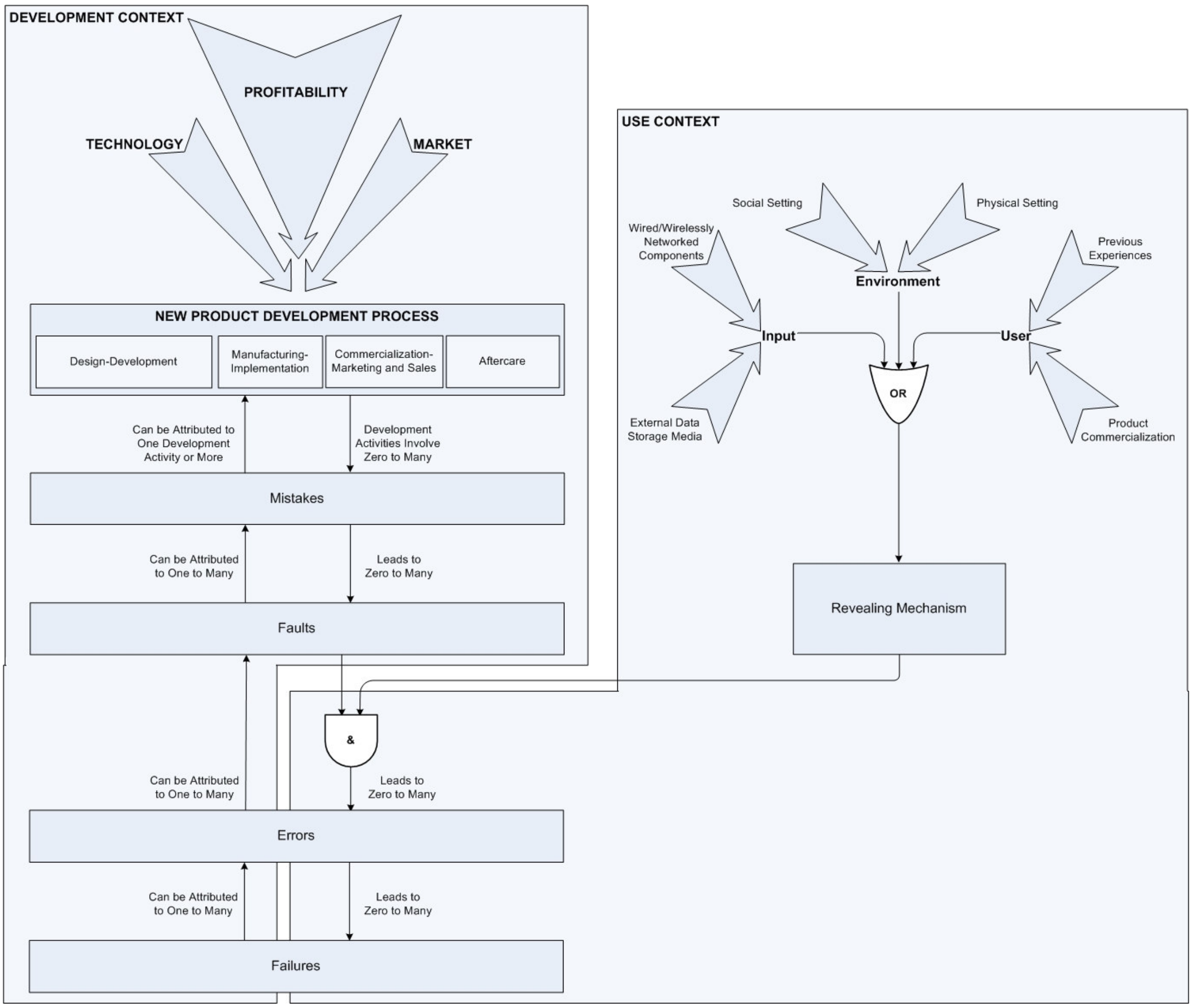

Figure 3 *Failure-focused feedback lifecycle, spanning use and development contexts*

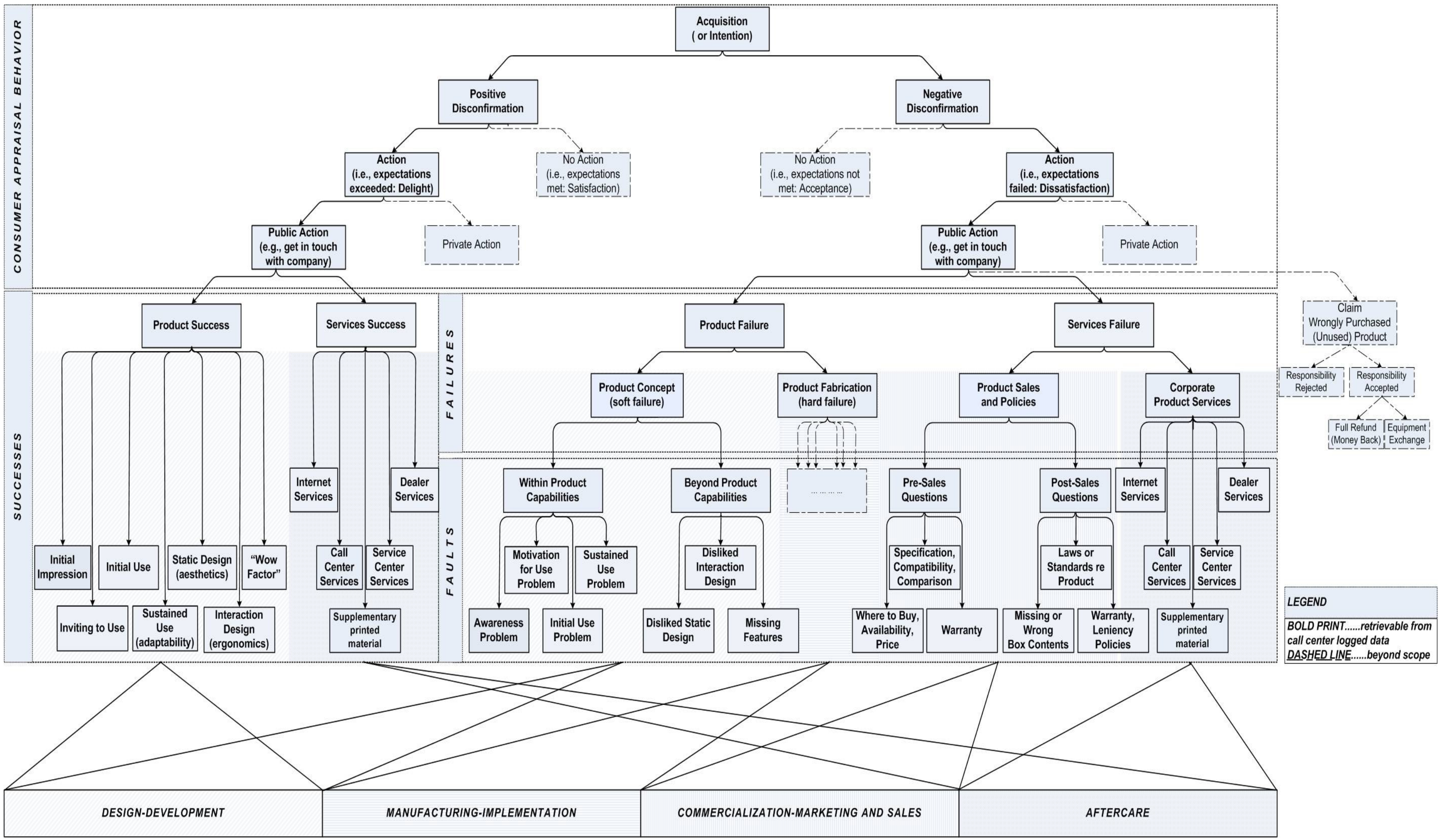

Figure 4 *User-centered feedback analysis and classification model*

Soft failures originate from *mistakes* during development activities in new product development processes (Figure 3), usually rising from lacking or improper user-centered design. These mistakes embed design-induced *faults* within the products under development, which may, depending on specific triggers within the use context (i.e., input, environment, user), then lead to *error* states during user-product interaction. The error state corresponds to an unexpected interaction state, and is typically driven by the tolerance level of the "judgmental system", i.e., the user. Determined by the degree of negative disconfirmation experienced (i.e., failing user expectations with respect to perceived product behavior), the user *may* conclude that, from his or her viewpoint, the product has failed, i.e., a soft system *failure* has occurred. At this point, to be able to effectively close the feedback loop, it is crucial to collect all relevant information about the use context (cf. Figure 3, right hand side block) and the expectations of the "judgmental system", i.e., the user.

To track the failure chain flow backwards (i.e., tracing the upward arrows in Figure 3) within the development context (cf. Figure 3, left hand side block), from the revelation of a failure during use, back to its originator development activities, it is necessary to render the unit of information being transferred (i.e., about the use context and user expectations) consistently correct, complete and meaningful. In view of that, we developed a user-centered feedback analysis and classification model (Figure 4), that not only aims at the appropriate specification of the information being transferred (i) from a manifested failure back to its root-cause faults (Koca and Brombacher, 2008); but also (ii) from a manifested success (i.e., positive disconfirmation of initially lower user expectations with respect to perceived product behavior) back to its relevant origins so as to identify the competitive edge of a product, especially during test-feedback operations. Note that the "faults" boxes in figures 3 and 4, and likewise the "failures" boxes are one and the same, and the "errors" box in Figure 3 corresponds to the right half of the "consumer appraisal behavior" box in Figure 4. The left half of the "consumer appraisal behavior" box in Figure 4 is for capturing the strong aspects of a product, and is actually a mirror image of the right half. Lastly, the bottom part of Figure 4 depicts the final links of the feedback loop, from e.g. faults to the related (abstracted) product development activities. As a result, the user-centered feedback model addresses the effects of the three challenge factors.

More detailed discussions regarding the conceptual- and empirical-grounds of our user-centered feedback analysis and classification model, and the intended way for its use, can be found in (Koca and Brombacher, 2008; Koca et al, 2007b). In the next section, we report the results of an empirical study, so as to illustrate the concepts and ideas of this paper using two comparative studies.

INDUSTRIAL CASE: TWO COMPARATIVE STUDIES

In the course of iteratively exploring and modeling soft failures with respect to all other failures involved in feedback operations, we conducted two studies in which we compared real feedback data from various repositories of two complex consumer electronic products. The main objectives of these studies were, by utilizing our feedback classification model in Figure 4, to compare and contrast (i) various feedback data repositories of one complex system and hence to identify the usefulness and suitability of each repository for quality improvement in the context of feedback operations; and (ii) the feedback data repositories of two similarly complex systems.

Data Sets

For each of the two complex consumer electronics products, called P_1 and P_2 hereafter, four feedback data sets, from year 2006 and 2007, were analyzed, each from either one of Service Centers (SC), Call Centers (CC), the Internet (I), or user Tests (T). The net usable distribution of feedback data over these sources can be seen in Table 1 and Table 2, respectively. Note that the initial data sets were much more populated before any filtering was done for useful and relevant content to be used with our model; and that Service Centers do not exist in either Table, since they were eliminated altogether due to the same filtering.

Classification Results and Discussion

All classifications were done manually by four raters for each data set, in the study for each product (i.e., eight raters in total), based on all the available information in the corresponding data repository. The results of the joint classifications are depicted in the respective Tables. Although not the exact same class distinctions from our model were used in the classifications of P_1 and P_2 data, the same patterns, as highlighted with bold print, were observed in both studies. That is, soft failures that can be resolved *within* the capabilities of the product by the user upon getting supporting guidance and instructions, is the most dominating category seen in Call Center repositories; whereas soft failures, which are due to users' higher or other expectations that are *beyond* what the product is capable of, is the most dominating category seen in both Internet and Test data repositories. The studies also revealed that Call Center repositories are currently the most populated ones and that they tend to have the richest, although not reliable, content. In contrast, Service Centers' repositories are currently the poorest in content, and hence not usable for soft reliability management as such.

Table 1 *Failure distributions over different data sources of P_1*

	Hard F.	**Soft F. within product**	**Soft F. beyond product**	**Services F.**
Call Center *(582)*	16%	**27%**	6%	**28%**
Internet *(135)*	1%	22%	**47%**	20%
Test *(36)*	1%	26%	**61%**	5%

Table 2 *Soft failure distributions over different data sources of P_2*

	Soft F. within product	**Soft F. beyond product**
Call Center *(832)*	**82%**	18%
Internet *(133)*	11%	**89%**
Test *(45)*	11%	**89%**

These explorative studies with our model, helped us mine and discover feedback data patterns manually, at a high level. Our ongoing research is on automating this process and also making it more detailed, by using automated mining tools to eventually reduce the duration of feedback processes, for quicker control and refinement in quality improvement. This is expected to shift the corrective fire-fighting attitude towards soft failures, to a more predictive attitude.

CONCLUSION

User feedback is invaluable information for achieving and maintaining user-expected quality of operational processes. Therefore, from the business perspective, it is essential to be able to systematically exploit feedback data to actively manage soft reliability of products, such that "No Fault Found" cases can be reduced, leading to reduced cost of non-quality of products and services raising from failed user expectations on technically sound products. To this end, we emphasized the incompatibility between today's feedback operations, and the changed requirements for an explorative rather than a prescriptive approach to deal with the many challenges of today's complex context of operational processes.

In this paper, we proposed the outcome of evolving know-how based on research in the diverse fields of operations management, new product management, systems and software engineering, dependable and secure computing, quality and reliability engineering, human-computer interaction, knowledge engineering and marketing.

ACKNOWLEDGEMENTS

This work is being carried out as part of the "Managing Soft Reliability in Strongly Innovative Product Creation Processes" project, sponsored by the Dutch Ministry of Economic Affairs under the IOP-IPCR program.

REFERENCES

Avizienis, A., Laprie, J.-C., Randell, B., Landwehr, C. (2004), "Basic Concepts and Taxonomy of Dependable and Secure Computing". In *IEEE Transactions on Dependable and Secure Computing*, Vol.1, No.1, pp. 11-33.

Bly, S., Schilit, B., McDonald, D.W., Rosario, B., Saint-Hilarie, Y. (2006), "Broken Expectations in the Digital Home". In *Proceedings of CHI Extended Abstracts on Human Factors in Computing Systems*, Quebec, Canada, ACM Press.

British Standard 5760-8: 1998 (1998), "Reliability of Systems, Equipment and Components. Guide to Assessment of Reliability of Systems Containing Software". BSI London, UK, pp. 5.

Gorlenko L., Englefied, P. (2006), "Usability Error Classification: Qualitative Data Analysis for UX Practitioners", In *Proceedings of CHI Extended Abstracts on Human Factors in Computing Systems*, Quebec, Canada, ACM Press.

IRIS coding, URL: http://www.eicta.org/index.php?id=37. Latest access: April 2008.

Koca, A., Brombacher, A.C. (2008), "Extracting "Broken Expectations" from Call Center Records: Why and How". In *Proceedings of CHI Extended Abstracts on Human Factors in Computing Systems*, Florence, Italy, ACM Press.

Koca, A., Schouwenaar, A.J.M., Brombacher, A.C. (2007a), "Field-Feedback in Innovative Product Development: A Comparison of Two Industrial Approaches". In *Proceedings of 14th International Product Development Management Conference*, Porto, Portugal.

Koca, A., Schouwenaar, A.J.M., Brombacher, A.C. (2007b), "Analysis of User-Centered Failure Mechanisms in New Product Development for Quality Improvement". In *Proceedings of 14th International Annual EurOMA Conference*, Ankara, Turkey.

Randell, B., Koutny, M. (2007), "Failures: Their Definition, Modelling and Analysis". In Jones, C.B., Liu, Z., Woodcock, J. (Eds.), *Lecture Notes in Computer Science 4711*, Springer, pp. 260-274.