Scenario-Based Evaluation of Perception of Picture Quality Failures in LCD Televisions

J. Keijzers[1], L. Scholten[1], Y. Lu[1], E. den Ouden[1,2]
[1]Sub Department of Business Process Design, Department of Industrial Design,
Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[2]Philips Applied Technologies, Industry Consulting,
High Tech Campus 5, 5656 AE Eindhoven, The Netherlands
J.Keijzers@tue.nl

Abstract
In innovative Consumer Electronics products, such as LCD televisions, consumers often perceive the product's malfunctioning differently than designers do. To support critical design decisions, it is therefore important to understand how consumers perceive potential product failures. This paper discusses the development of realistic failure scenarios related to the picture quality of an LCD television. The impact of television content as well as failure origin on the perception of failures is evaluated in an experiment. Advantages and drawbacks of using the scenarios to evaluate design decisions, and implications for further research are discussed.

Keywords:
Scenario, Product Design, User-Perceived Failure, Consumer Electronics

1 INTRODUCTION

Innovative Consumer Electronics (CE) products, such as LCD televisions, are becoming increasingly complex, both in terms of the embedded technologies (i.e. increasing software content, ambient technologies, and open systems) [1] as in terms of the number of functionalities provided [2]. For example, the LCD television of today can be used to access the Internet, watch digital photos, and connect to a personal computer to watch downloaded movie content. Furthermore, because of the globalization of demand, these products require a high level of connectivity with other products and other brands. Research by Den Ouden [3] has shown that companies report an increasing number of customer complaints and even product returns for such products. A closer view on these consumer complaints has shown that the number of 'No Fault Found' (NFF) failures in the CE industry has risen to a level of nearly 50% of the complaints in 2004 [4]. NFF failures are failures where the product performs according to the technical specifications but the product is still rejected because it does not meet the consumers' expectations [4]. Because consumers do not always understand how these complex products function [5], they often perceive the product's malfunctioning differently than designers do. Even when a product is still functioning according to technical specifications, a consumer might think otherwise. For example, configuration problems with the channel settings of an LCD television could be attributed by a consumer to the cable provider, the design of the television, bad instructions in the manual or even lack of knowledge of the user.

To prevent such product failures in the field, there is a need to incorporate more user focus in the new product development process [3]. Specifically, it is necessary for the designers to have more insight into how consumers perceive potential product failures, to support critical design decisions early in the product development process. Consumer tests can be used to measure consumer perception of product failures by letting consumers experience potential product failures and subsequently asking them to attribute the perceived failure to the perceived cause. However, reproducing realistic failures, which occur in actual product usage situations, in a controlled experiment, is often difficult, especially when resources and time available are limited. Alternatively, product failures can be presented to subjects by using different means, for example by showing simulated failures using video recordings [6] [7]. Yet, how to create realistic failures scenarios that need to be tested is still a question to be addressed. The goal of this paper is therefore to develop failure scenarios, related to picture quality problems of a high-end LCD television, which can be used to measure consumer perception of those failures.

The remainder of this paper is organised as follows. First, in section 2, relevant literature related to picture quality failures in LCD televisions and scenario-based product evaluation will be discussed. This section will conclude with the hypotheses addressed in this paper. In section 3, the development of the failure scenarios related to picture quality of an LCD television will be discussed. Subsequently, in section 4, the set-up of an experiment that was used to evaluate the developed scenarios, will be discussed. The results of this experiment will be discussed in section 5. Finally, in section 6, this paper will be concluded with a discussion and conclusions on the advantages and drawbacks of using the developed failure scenarios to evaluate design decisions.

2 LITERATURE REVIEW AND HYPTOHESES

2.1 Picture quality failures in LCD Televisions

LCD televisions are a typical example of the trends in the development of CE products discussed in section 1.

TV systems have an increasing number of features and the amount of software that is embedded in the TV is dramatically increasing [8]. Although the TV system has changed dramatically from a technical point of view, many consumers do not understand these changes. Consequently, as shown by De Visser [6], consumers perceive product failures and product failure severity of failures in TV systems differently than designers do. For this research, a choice is made to focus on developing failure scenarios related to picture quality of LCD televisions. Picture quality is chosen because it is one of the most important aspects of the quality of LCD televisions [6] and because it is easy to simulate picture quality related failures (see for example research by Puchihewa and Bailey on different artefacts in image and video systems [9]).

Regarding user-perceived failures in picture quality, among other things (see also De Visser [6]), two aspects are important when evaluating the use of failure scenarios. Firstly, it is important to investigate whether consumers perceive (i.e. notice) the simulated failures in picture quality by using the failure scenarios. Secondly, it is important to investigate the perceived failure impact. According to De Visser [6], failure impact can be defined as 'the percentage loss of functionality as a result of a failure' and is an important predictor of user-perceived failure severity.

2.2 Scenario-Based Evaluation of Product Failures

The use of scenarios in product design is widely supported. Scenarios can be used as a communication tool between designers, users, and stakeholders, they require less time and costs than using prototypes, and they provide designers flexibility [10] [11] [12]. Furthermore, scenarios can have many different views and forms [11] [13] and can be used for many different purposes throughout the design process [12]. Since product failures in CE products are difficult to reproduce in a real product for use in a controlled laboratory experiment, it is interesting to investigate whether product failure scenarios can be used instead.

In literature, several examples can be found of the use of scenarios to evaluate product failures [7] [14] or service failures [15]. However, how to create realistic failure scenarios of complex CE products that can be used to let consumers evaluate the failures as input for critical design decisions is currently not addressed. Research has shown that user-perceived failures are very dependent on both the characteristics of the failure and the user, as the context in which the failure occurs [6]. Consequently, the choice is made to focus this research on the influence of a scenario contextual variable (television content) and a product failure specific variable (failure origin) on user-perceived failures. These variables and related hypotheses will be subsequently discussed in section 2.3 and section 2.4.

2.3 Television Content

Earlier research by Ghinea and Thomas [16] has demonstrated the relation between the content of video clips and the level of user perception and understanding of the content of the video clip. Highly dynamic and information rich video clips have a negative impact on the users' understanding and information assimilation. This occurs because users have difficulty absorbing audio, visual, and textual information concurrently. The importance of audio, video, or textual information also varies between video clips with different content [7]. For example, a news-broadcast has different properties than a music video or a wild-life documentary. This research implies that there is a possible effect of television content on user-perceived failures in television picture quality. Therefore, one can hypothesize that television content that is so captivating to the user could result in the failure in picture quality being unnoticed. As a result, the following hypothesis was formulated:

Hypothesis 1: Captivating television content negatively influences the perception of a failure in picture quality of LCD televisions.

Besides the hypothesized influence of content on the perception of the presence of a failure, television content could also influence the perceived impact of the failure. In other words, one could hypothesize that, depending on personal preferences for television content, a failure occurring during an interesting news-broadcast or an exciting movie is perceived as having a higher impact than the same failure occurring during a broadcast of a political debate. This resulted in the following hypothesis:

Hypothesis 2: Captivating television content positively influences the level of perceived failure impact in picture quality of LCD televisions

2.4 Failure origin

According to technical experts, television picture quality can be influenced by internal problems in the TV (e.g. faults in the software) when processing or displaying the TV signal, as well as by external problems outside the TV such as bad weather or cable connection problems. Problems internal to the TV could, for example, result in ghosting or blocking artefacts (see [9] for more examples). Bad weather or cable connection problems could, for example, result in noise on the screen. Because consumers might attribute externally caused problems to the TV system or the other way around, it is important for product developers to gain insight into how consumers perceive these different types of failures in picture quality. Furthermore, because failure impact is an important predictor of user-perceived failure severity and subsequent consumer complaint behaviour, it is interesting to investigate whether failure origin influences perceived failure impact. Subsequently, the following hypotheses related to failure perception and perceived failure impact are developed:

Hypothesis 3: Failure origin influences the perception of a failure in picture quality of LCD televisions.

Hypothesis 4: Failure origin influences the level of perceived failure impact in picture quality of LCD televisions.

To test these hypotheses, failure scenarios with different television contents (captivating and less captivating) and different failure origins (internal or external to the TV) were developed. The design of these scenarios will be discussed in the following section.

3 SCENARIO DESIGN

3.1 Design Methodology

Many factors have to be taken into consideration when creating the required failure scenarios. In this section these factors and the methods used to incorporate those factors will be discussed. According to Rolland et al. [13], scenarios can be designed along four different views: purpose, content, form and lifecycle view. This framework was used as a guideline to define and develop the scenarios for this paper.

As previously mentioned, the purpose of developing failure scenarios here, is to let participants evaluate picture quality problems by viewing the failure scenarios which simulate experiencing the same failures in a realistic use situation. To achieve this, an iterative design process was used in which digital TV system experts (both picture quality experts and system testers) were actively involved in designing the scenarios and evaluating the quality and realism of the scenarios. The specific design choices and the resulting failure scenarios will be discussed in the following sections.

3.2 Selection of Scenario Form

A scenario can be made by using different media, all of which with their own advantages and disadvantages. Seawright and Sampson [17] compare written scenarios to video scenarios, in which they emphasize that in written scenarios there is an empirical experience but not a real (timely) experience. Also the interpretation variance is different because with a written scenario the user needs to interpret the meaning of words (e.g. 'fast', 'slow') and in a video scenario the user needs to interpret the simulation of the situation. On the other hand, Hamberg and De Ridder [18] make an argument for the use of still pictures instead of moving video to measure the perceptual image quality. The use of still pictures simplifies the stimulus material significantly, because stills do not have time variation of the scene content. In more current research by Pinson and Wolf [19] moving video scenarios are successfully used for the assessment of picture quality. This is because they use a continuous assessment method and again the timeframe is of important matter. Since timing and realism are of highest importance for experiencing picture quality failures in LCD televisions, it was decided to use a video scenario.

3.3 Design of Scenario Context

The main disadvantage of the use of any type of scenario for product failure evaluation is the detachment of the failure experience from its original context. Although consumer experiments in a laboratory setting have the same drawbacks, the context used in a failure scenario is important because such scenarios lack any type of user-product interaction. Consequently, the context of the scenario in which the picture quality failures were shown to the subjects was carefully designed. To simulate a realistic environment, a video recording of a living room environment was used in which someone acted as the user of an LCD television. Each failure scenario video had a duration of 90 seconds (30 seconds for the introduction, 30 seconds for television content without failure, and 30 seconds for television content with the implemented failure) and consisted of the following fragments:

- Introduction to the living room setting: family member enters the living room and switches on the TV.
- Interaction with the TV: family member switches through several TV channels and ends up at the channel with the implemented failure in picture quality. A fragment of this part is shown in Figure 1.
- Camera zooms in on the TV
- Video clip with failure scenario is shown on the TV in which after 30 seconds the implemented failure occurs.

To provide a frame of reference, a fragment of the wild-life video without an implemented failure in picture quality is shown in Figure 2.

Figure 1: Scenario context in living room.

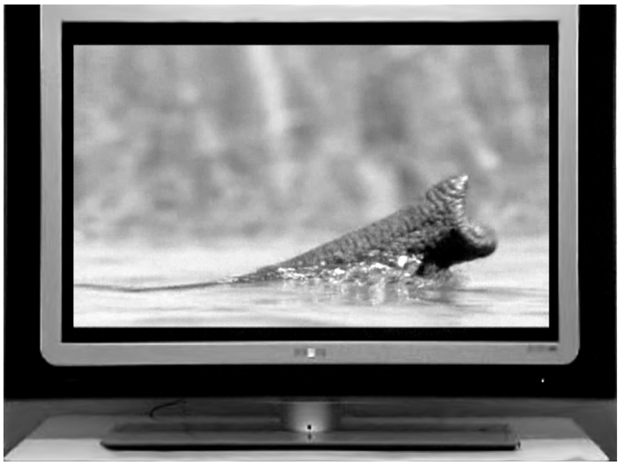

Figure 2: Fragment of scenario without a failure.

3.4 Selection of Scenario Content

To investigate the influence of TV program content, two video clips that differ on the level of captivating content were selected. The content description of both clips is shown in Table 1.

Each clip contained a meaningful segment of a TV program. To conduct an accurate study both contents needed to differ on only one factor: captivating. This meant that all the other factors, for example, movement, lightness, sound level, duration, and quality, needed to be kept as constant as possible. The term captivating TV content is personally linked. Each person will have their own taste and interest in television programs. However, the assumption in this study was made that an action

Scenario content	Description
Captivating content	Genre: action movie Content: fragment of 'Matrix Revolutions' with a motor and car chase scene
Less captivating content	Genre: wild-life documentary Content: fragment of 'Earth' documentary with walking and swimming elephants

Table 1: Failure scenario content description.

movie is experienced as more captivating for the selected test participants (see section 4.4) than a wild-life documentary. This assumption, however, needs to be validated by consumers when evaluating the developed scenarios.

3.5 Design of Failure Scenarios

To investigate the influence of failure origin, two different failures were selected in agreement with the digital TV systems experts. Two variations that were chosen to be integrated in the scenarios are:

- Failure which is most likely to be caused by (software) faults in the television: blocking artefacts on the TV screen.
- Failure which is most likely to be caused by a signal disturbance in the cable or a bad cable (connection): noise on the TV screen.

For both failure scenarios it is important to notice that the failure origin can differ depending on, among other things, the type of cable signal (analogue versus digital), the TV system configuration, the duration of the failure, and the appearance of the failure. Both failure scenarios used for this research were selected and designed to represent the difference in failure origin as discussed above. Because only failure origin was varied, all the other factors, for example start time, duration, severity; were kept as constant as possible.

Both the failure scenario with blocking artefacts and the failure scenario with noise were created with video editing software. Since videos are used as a medium, it was very important to use the highest video quality possible. On the one hand this ensured that the picture quality failures were clearly visible while on the other hand the implemented failures could not be attributed to bad video quality or the program used to display the video. Both scenarios were reviewed by the digital TV system experts on the quality, perceived failure impact and perceived failure origin to verify whether the scenarios were realistic. The scenarios were approved after adjustments made during two iterative cycles.

A fragment of the wild-life video with blocking artefacts and a fragment of the same video with noise are shown in Figure 3 and Figure 4 respectively. Both fragments are exactly the same fragment as the fragment without an implemented failure as shown in Figure 2. In the next section, the set-up of an experiment will be discussed that was used to test the hypotheses and to evaluate the use of failure scenarios.

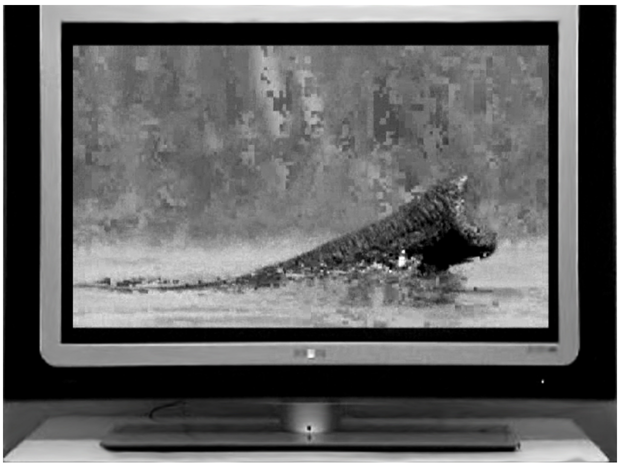

Figure 3: Fragment of scenario with blocking artefacts.

Figure 4: Fragment of scenario with noise.

4 EXPERIMENT METHODOLOGY

4.1 Overview

To investigate the influence of television content and failure origin on the perception of the failure, a 2 x 2 between-subjects experiment was set up. The subjects were asked to fill in a web-based survey in which the failure scenarios were implemented. To evaluate the validity of the failure scenario design, the questionnaire consisted of three parts. In the pre-experimental part, a video with either the action movie or wild-life documentary without a failure was shown to measure perceived quality. Subsequently, in the experimental part a video with the implemented failure and with the same content as the introduction video was shown to measure failure perception and perceived failure impact. Finally, in the post-experimental part several control variables and demographics were measured.

4.2 Experimental Variables

The independent variables were the content of the television program (wild-life-documentary versus action movie) and the origin of the failure (blocking artefacts versus noise).

As dependent variables, failure perception and perceived failure impact were measured. Failure perception was measured by a single yes/no question. Failure impact was measured on a 5-point scale (with 1 = not annoying and 5 = very annoying, adjusted from De Visser [6]).

Additionally, several control variables were measured to evaluate the realism of the scenario. To test the assumption that the action movie is more captivating than the wild-life documentary, interest in the content was measured on a 5-point scale (with 1 = interesting and 5 = not interesting). Furthermore, to evaluate the quality of the designed scenarios, the subjects were asked to rate the realism of the scenarios on a 5-point scale (with 1 = realistic and 5 = unrealistic). Finally, to evaluate the successful simulation of picture quality failures, subjects were asked to evaluate the picture quality of the LCD television for both the introduction and failure scenario on a 5-point scale (with 1 = very good and 5 = very bad).

4.3 Apparatus and Materials

The web-based questionnaires with inserted videos were set-up by using Limesurvey [20]. Each subject filled-in the questionnaire on a computer at the university campus. For each subject, a researcher was present to access the survey and select the appropriate settings (i.e. choosing the appropriate failure scenario to achieve equal sample sizes). No additional materials were provided and the subjects were not allowed to access the Internet to search for any additional information.

4.4 Participants

Since the goal of this research was to validate the design and use of failure scenarios, it was considered important to use a group of participants with homogeneous characteristics (similar to research by De Visser [6]). Therefore, the experiment was carried out with students of the faculty Industrial Design at Eindhoven University of Technology. They were chosen as participants because they are a group with homogeneous characteristics and because they are easily accessible. To further ensure that differences in perception of the picture quality failures would not be attributable to differences in personal characteristics among the participants, levels of familiarity (usage and ownership) and expertise [21] regarding televisions were measured. In total, 40 participants (27 males and 13 females) with a mean age of 22.0 years (SD = 2.42 years, range = 19 to 29 years) were randomly assigned to one of the experimental groups. Non-parametric Kruskal-Wallis tests [22] showed no significant differences between the personal characteristics of the participants in the four groups.

4.5 Procedure

At the beginning of the questionnaire, the participants were instructed that the experiment was set up to evaluate the quality of LCD televisions. The entire questionnaire was written in Dutch and consisted of the following parts:

- Introduction to the purpose of the questionnaire and general instructions
- Introduction scenario: video with no implemented failure. Additionally, information was provided on the capabilities of the LCD television displayed in the scenario and on the configuration of the whole set-up (i.e. type of cable signal and conditions under which the content on the TV was shown).
- Control questions on introduction scenario: measurement of interest, picture quality and failure perception
- Failure scenario: video with implemented failure
- Questions on failure scenario: measurement of picture quality, failure perception, failure impact and perceived realism
- Questions on familiarity and expertise regarding LCD televisions
- Questions on demographics

Filling-in the entire questionnaire took approximately 15 minutes. The entire procedure and questions were pre-tested in a pilot test and based on the comments several small adjustments were made to the formulation of questions and question order.

5 RESULTS

5.1 Overview of Results

In this section, an overview will be given of the results of this experiment and the validity of the scenarios will be discussed. The results for the most important measurements for the different scenarios are shown in Table 2 below. The measures did not meet the assumption of normality and equal variances between the different groups which are a prerequisite for being able to use parametric tests. In order to compare differences across the scenarios, one-way Kruskal-Wallis analyses were run. In order to compare differences between groups, separate pair-wise Mann-Whitney U tests [23] were used. Holm's sequential Bonferroni correction [24] was used to determine the corrected significance level when multiple comparisons were made. The level of significance was set at p = 5%. Results within the less restrictive 10% level are indicated as marginally significant.

Before testing the hypotheses, first the validity of the designed scenarios will be discussed. First of all, the results of a non-parametric Mann-Whitney U test confirmed that the action movie content is more interesting to the participants than the wild-life documentary content (p=0.012). Furthermore, from the results in Table 2 it can be seen that all of the scenarios are considered reasonably realistic, although the blocking artefacts scenario is considered less realistic than the noise scenario. Results of a Kruskal-Wallis test show there are no significant differences between the level of perceived realism for all the scenarios ($\chi2(2) = 2.83$, $p < 0.42$). Subsequently, the Wilcoxon signed-rank test [22] was used to test for a significant difference between the picture quality measurements of the introduction video with failure and the subsequent failure scenarios. The results show that the manipulation of the picture quality failures was successful (p=0.000). Finally, the failure impact measurements between all the scenarios were compared to test whether the severity of the failure was perceived as constant between the two failure scenarios within each content group. The results of a Kruskal-Wallis test show that the measurements of failure impact are significantly different across the scenarios ($\chi2(2) = 11.46$, $p < 0.01$). However, the separate Mann-Whitney U tests confirmed that, when comparing the noise and blocking artefacts scenario for the wild-life documentary (p=0.422) and when comparing the noise and blocking artefacts scenario for the action movie (p=0.112), the perceived failure impact is not significantly different. Overall, it can therefore be concluded that the manipulation of the failure scenarios was successful. In the next section the results of the hypotheses tests will be discussed.

5.2 Test of hypotheses

First, it can be concluded that the results of Mann-Whitney U tests show that there is no significant influence of television content (p=0.602) and failure origin (p=1.000) on failure perception. In other words, hypotheses 1 and 3 cannot be accepted. For all the

	Wild-life documentary			Action movie		
	No Failure	Noise	Blocking artefacts	No Failure	Noise	Blocking artefacts
Interest in scenario content	2.60	x	x	1.85	x	x
picture quality	2.20	4.60	3.90	2.20	4.0	4.20
failure perception [%]	x	90	90	x	80	80
failure impact	x	3.40	2.80	x	4.12	4.62
scenario realism	x	2.0	2.67	x	2.38	2.75

Table 2: Overview of mean scores for the experimental variables.

scenarios almost all the participants noticed a failure in picture quality. Although previous research discussed in section 2.3 showed that captivating content negatively influences attention when watching a video, the duration of the failure scenario videos might have been too short and too fragmented to fully capture the attention of the participant. Furthermore, both failures were made quite severe to counter for the use of videos (which are on a computer always smaller than a large LCD TV screen).

Secondly, the influence of television content and failure origin on perceived failure impact is tested. Mann-Whitney U tests confirmed that a picture quality failure in the action movie is perceived as having a higher failure impact than a picture quality failure in the wild-life documentary ($p=0.002$). Consequently, hypothesis 2 is accepted. A similar comparison on failure origin is less conclusive, since the result is only marginally significant ($p=0.088$). In other words, hypothesis 4 is rejected but the results indicate that there might be an influence of failure origin on failure impact. When comparing the scenarios one-by-one for perceived failure impact, only the comparison between the action movie blocking artefacts scenario on the one hand, and the wild-life documentary blocking artefacts scenario ($p=0.008$) and noise scenario ($p=0.008$) on the other hand, are significant after applying the Bonferroni correction. One possible explanation for these mixed results is the fact that blocking artefacts in fast moving video content cause more disturbance of the video and have different colour, contrast and detail variations than in slow moving content.

6 DISCUSSION AND CONCLUSIONS

To support critical design decisions in the development of complex CE products, it is important to understand how consumers perceive potential product failures. This study investigated the use of a scenario-based evaluation method which was used to test the influence of television content and failure origin on user-perceived failures in picture quality of an LCD television. Four different failure scenarios were designed which were implemented in a simulated living room context. A 2 x 2 between-subjects experiment with 40 participants was carried out to evaluate the design of the scenarios and to test the formulated hypotheses.

6.1 Use of Failure Scenarios to Measure User-Perceived Failures

In this study a different approach of the use of scenarios was investigated by specifically focusing on user-perceived product failures. Several control variables were used to measure the validity of the designed scenarios. Overall, based on the statistical results discussed in section 5.1, it can be concluded that the design of the scenario context and the implementation of the failures was successful. Although the blocking artefacts scenarios were perceived as less realistic than the noise scenarios, this may be explained by the fact that, in practice, noise on a TV screen is more common than blocking artefacts.

The advantages of using failure scenarios instead of trying to reproduce realistic product failures during actual product usage situations are not only the limited amount of resources and time required to design the scenarios and the possibility to use evaluation methodologies which easily allow to use larger sample sizes (e.g. survey instead of laboratory experiment). Similar as the argumentation by Carroll [11] in the context of scenario-based product design, the use of failure scenarios provides a significant larger degree of flexibility to explore different product failures which in actual product usage situations would be practically unfeasible.

However, the scenario-based approach for user-perceived product failure evaluation also suffers from drawbacks which require careful design and pilot tests but also limit the applicability in certain contexts. Firstly, the scenario-based approach is only useful for failure scenarios which do not interact with the form in which the scenario is shown. For example, video scenarios cannot be used to evaluate the user perception of damaged pixels or small disturbances in the sound quality since such failures could also be attributed to the video. Moreover, scenarios remove the failure from the context in which the failure occurs (e.g. no user-product interaction, different start time and duration of the failure etc.) This can only be partly adjusted for with a careful design of the scenario context. Therefore, the input from and evaluation by product designers, developers, and testers is of crucial importance for designing realistic product failure scenarios.

6.2 User-Perceived Failures in Television Picture Quality

Although the main goal of this paper was to evaluate the use of failure scenarios to evaluate user-perceived failures, several insights are gained about user-perceived picture quality failures. Firstly, the results of the experiment show that overall failure origin does not significantly influence the perception of a failure and the perception of failure impact. Secondly, the results did confirm that, for the student sample, failures in captivating content have a higher perceived impact than the same failures in less captivating content. To be importance for supporting product design decisions, among other things, the consumers' attribution of different picture quality failures to their perceived causes should be measured. Such insights could be used to support design decisions in the design of the user interface or the user manual.

6.3 Further Research

This study is part of a larger project to investigate the influence of failure and user characteristics on user-perceived failures in CE products (see also De Visser [6]). Although this explorative study was conducted with a narrow sample and limited sample size, the insights gained can be used to improve a scenario-based evaluation in future studies, to measure the perception and attribution of product failures. However, further research with a larger and different sample (i.e. more heterogeneous on, for example, age and product expertise [21]) and with different products is needed to validate the use of failure scenarios. Furthermore, more research is needed to account for the drawbacks of the scenario-based approach and, more specifically, to improve the design of the context in which the failure scenarios are implemented. For example one may evaluate the inclusion of user-product interaction in a scenario or enrich the scenario by using a more elaborate context description to place the failure into a relevant and realistic context.

7 ACKNOWLEDGEMENTS

This work has been carried out as part of the Trader project under the responsibility of the Embedded Systems Institute. This project is partially supported by the Netherlands Ministry of Economic Affairs under the BSIK program. We would like to thank Rob Golsteijn, Iulian Nitescu and Roland Mathijssen for their help with designing the failure scenarios. Furthermore, we would like to thank Jun Hu for his support with the survey software.

8 REFERENCES

[1] Siewiorek, D.P., Chillarge, R., Kalbarczyk, Z.T., 2004, Reflections on Industry Trends and Experimental Research in Dependability, IEEE Transactions on Dependable and Secure Computing 1(2): 109-127.

[2] Norman, D.A., 2007, Three Challenges for Design, Interactions January + February 2007: 46-47.

[3] Den Ouden, E., 2006, Development of a Design Analysis Model for Consumer Complaints, Ph.D. Thesis, Eindhoven University of Technology, The Netherlands.

[4] Brombacher, A.C., Sander, P.C., Sonnemans, P.J.M., Rouvroye, J.L., 2005, Managing Product Reliability in Business Processes 'Under Pressure', Reliability Engineering and System Safety 88(2): 137-146.

[5] Cooper, A., 1999, The Inmates are Running the Asylum: Why high-Tech Products Drive us Crazy and How to Restore the Sanity, Sams, Indianapolis.

[6] De Visser, I.M., 2008, Analyzing User Perceived Failure Severity in Consumer Electronics Products: Incorporating the User Perspective into the Development Process, Ph.D. thesis, Eindhoven University of Technology, The Netherlands.

[7] Jumisko-Pyykkö, S., Kumar, V., Korthonen, J., 2006, Unacceptability of Instantaneous Errors in Mobile Television: From Annoying Audio to Video, Proceeding of the MobileHCI 2006 Conference, Helsinki, Finland.

[8] Tekinerdogan, B., Sözer, H., Aksit, M., 2008, Software Architecture Reliability Analysis Using Failure Scenarios, Journal of Systems and Software 81(4): 558-575.

[9] Punchihewa, A., Bailey, D.G., 2002, Artefacts in Image and Video Systems: Classification and Mitigation, Proceedings Image and Vision Computing New Zealand, Auckland, New Zealand: 197-202.

[10] Carroll, J.M., 2000, Making Use: Scenario-Based Design of Human Computer Interactions, The MIT Press, Cambridge, Massachusetts.

[11] Carroll, J.M. (ed.), 1995, Scenario-Based Design: Envisioning Work and Technology in System Development, John Wiley & Sons Inc., New York.

[12] Anggreeni, I., Van der Voort, M.C., 2008, Classifying Scenarios in a Product Design Process: A Study Towards Semi-Automated Scenario Generation, Proceedings of the 18th CIRP Design Conference, Enschede, The Netherlands.

[13] Rolland, C., Achour, B., Cauvet, C., Ralyté, J., Sutcliffe, A., Maiden, N., Jarke, M., Haumer, P., Pohl, K., Dubois, E., Heymans, P., 1998, A Proposal for a Scenario Classification Framework, Requirements Engineering, 3: 23-47.

[14] Lancellottti, M.P., 2004, Technological Product Failure: The Consumer Coping Process, Ph.D. Thesis, University of Southern California, United States of America.

[15] Seawright, K.K., DeTienne, K.B., 2008, An Empirical Examination of Service Recovery Design, Marketing Intelligence & Planning 26(3): 252-274.

[16] Ghinea, G., Thomas, J.P., 1998, QoS Impact on User Perception and Understanding of Multimedia Video Clips, Proceedings of the Multimedia Conference 1998, Bristol, United Kingdom.

[17] Seawright, K.K., Sampson, S.E., 2007, A Video Method for Empirically Studying Wait-Perception Bias, Journal of Operations Management 25(5): 1055-1066.

[18] Hamberg, R., De Ridder, H., 1995, Continuous Assessment of Perceptual Image Quality, Journal of the Optical Society of America A: Optics, Image Science and Vision 12(12): 2573-2577.

[19] Pinson, M., Wolf, S., 2003, Comparing Subjective Video Quality Testing Methodologies, Proceedings of Visual Communications and Image Processing 2003, Lugano, Switzerland.

[20] Limesurvey v1.71, an open source survey application, www.limesurvey.org.

[21] Alba, J.W., Hutchinson, J.W., 1987, Dimensions of Consumer Expertise, Journal of Consumer Research 13(4): 411-454.

[22] Montgomery, D.C., Runger, G.C., 1999, Applied Statistics and Probability for Engineers, Second Edition, John Wiley & Sons Inc., New York.

[23] Mann, H.B., Whitney, D.R., 1947, On a Test of Whether one of two Random Variables is Stochastically Larger than the Other, Annals of Mathematical Statistics 18(1): 50-60.

[24] Howell, D.C., 2002, Statistical Methods for Psychology, Fifth Edition, Thompson Learning, Pacific Grove.