Visual Fidelity of Video Prototypes and User Feedback: A Case Study

B. Dhillon, P. Banach, R. Kocielnik, J.P Emparanza, I. Politis, A. Rączewska and P. Markopoulos
Eindhoven University of Technology
User-System Interaction, Industrial Design, P.O.Box 513
5600 MB Eindhoven, NL
{ b.k.dhillon[1], p.j.banach[2], r.d.kocielnik[3], j.peregrin.emparanza[4], i.politis[5], a.m.raczewska[6], p.markopoulos[7] }@tue.nl

This paper addresses the debate regarding the respective merits of high and low fidelity prototypes, in the domain of video prototyping. Video prototyping is a popular tool for interface designers. Despite this, there is practically no research reported to date examining the fidelity of the design representation that the video prototype should manifest. We report a case study where the same design concept was rendered on video in two formats with differing degree of visual fidelity: animated paper cut-outs (low visual fidelity) versus a video with real actors, edited to simulate computer output (high visual fidelity). A two-pronged comparative evaluation was carried out: a between-subjects questionnaire survey, consisting of AttrakDiff, open-ended questions completed by 99 participants, and semi-structured qualitative interviews with 9 participants. The results did not reveal any differences regarding the amount or quality of feedback one should expect from a low or a high fidelity video. These results lead us to suggest that the paper cut out animation is a valid prototype that should be explored more by interaction designers for obtaining early user feedback at low cost.

Video prototype, video prototype fidelity, feedback on fidelity, visual refinement

1. INTRODUCTION

Prototyping is a quintessential element of interaction design, particularly for soliciting feedback from user representatives in order to guide further design and development. Prototypes are working models or simulations of the product or design concept (Walker, Takayama and Landay, 2002). These models can vary substantially in the effort they require for their development, and their potential role in the design process.

Traditionally, prototypes have been distinguished in terms of the level of their fidelity, defined as the degree to which a prototype resembles the final product (Virzi, 1989). Low fidelity prototypes (e.g., paper prototypes) can look sketchy or incomplete and are often used in early stages of design (Rudd, Stern and Isensee, 1996). In contrast, higher fidelity is typically associated with interactive systems that have some of the intended interaction and relatively more detailed visual design of the final product. While the distinction of high and low fidelity is quite a long standing and widely held concept in the field of interaction design, several researchers have criticized it as being un-nuanced. For example, McCurdy et al (2006), in one of the most recent contributions to this debate, identify

five dimensions of prototype fidelity: visual refinement, breadth of functionality, depth of functionality, richness of interactivity and richness of data model. Visual refinement refers to the extent to which a prototype resembles the final product. Breadth of functionality addresses the number of features a prototype supports while depth refers to the level of detail worked out for each feature. The interactivity relates to the extent to which the interaction with the prototype is similar to the final product. Finally, the last dimension defines the extent to which the data used for the prototype is representative of the actual domain data. If a prototype scores low on one of these four dimensions, and if this is obvious to the user, one can speak of a low fidelity prototype. For example, hand-drawn sketches and wireframes usually lie towards the lower end of visual refinement spectrum (McCurdy et al., 2006).

An important advantage of low fidelity prototyping is that it usually takes lesser time and resources than high fidelity prototyping. This efficiency arguably prevents designers from getting too attached to the concept by limiting the investment of their time and effort (Rudd, et al 1996). Further, low fidelity prototypes provide more freedom to designers, as they are not limited by available technology (Wong,

1992). However, high fidelity prototypes may be more successful in simulating complex systems or systems requiring physical manipulations. Finally, high fidelity prototypes may be perceived as more "professional" or "attractive" than low fidelity ones, which is especially important when promoting a product or an idea to management or customers.

Several empirical studies have attempted to evaluate the differences between prototypes with different fidelity levels; for a recent overview, refer to Sauer, Franke and Ruettinger (2008). Overall, these studies have shown no major differences regarding the effectiveness of low and high fidelity prototypes as tools for identifying usability issues (Catani and Biers, 1998; Sauer and Sonderegger, 2009; Sefelin, Tscheligi and Giller, 2003; Virzi, Sokolov and Karis, 1996; Wiklund, Thurrott and Dumas, 1992). Only a few studies reported that using low fidelity prototypes was less effective when looking for usability issues (Sauer, Franke and Ruettinger, 2008). Related to this, Walker, Takayama, and Landay (2002) reported that although users gave more comments about high fidelity prototypes than about low fidelity ones, they did not find differences in the discovered number of usability issues. Such research findings can imply that low fidelity prototypes are valid tools for obtaining feedback from users, within some reasonable boundaries. Additionally, proclaimed benefits of one or the other form of prototype (e.g., see Rudd, Stem, and Isensee, 1996) have so far not been validated.

2. FIDELITY IN VIDEO PROTOTYPING

The present study used video as a prototyping medium. Video prototypes are powerful tools that can be used for design exploration, evaluation and presentation (Mackay, Ratzer and Janecek, 2000). The Starfire project is a classic example (Tognazzini, 1994) in which Sunsoft put forward a vision of a next-generation communication interface using a video prototype. Video prototyping has been growing in popularity. This development is, perhaps, a result of the increased availability of digital video cameras and simple, cheap, yet powerful video editing tools.

As discussed above, several studies have discussed advantages and disadvantages of low and high fidelity prototypes. However, as far as the authors are aware of, no studies to date have addressed fidelity differences in video prototyping. Thus, the potential effects of video prototype fidelity on the design process are unknown. It becomes quite relevant and necessary to provide some methodological guidance to potential creators of such prototypes.

The fidelity of video prototypes can be approached at two levels; it can pertain to the realism of the filming, e.g., using real actors versus cartoons, and it can pertain to the fidelity of the prototype of a device or product shown in the video. This paper focuses on the notion of realism of the video, i.e. higher realism corresponds to higher visual fidelity, and lower realism corresponds to low visual fidelity. The current case study aimed to examine the impact of visual fidelity of the video prototype on the amount and quality of feedback that viewers can provide during an evaluation session.

3. CASE STUDY

Two video prototypes of an interactive system, which encourages stair use, were created. The two videos differed in terms of the realism of the filming and consequently, in the visual refinement of the represented prototype; for brevity they are referred to below as the high and the low fidelity prototypes. Participants were surveyed regarding their reactions to these videos. An online questionnaire was used, allowing a comparison between groups of participants who had watched the high fidelity and the low fidelity video. In addition, qualitative interviews were conducted with a smaller group of participants. Firstly, the concept and the two videos prototypes are described, and then the study method and the results are presented.

3.1 Stepper's Club

The video prototypes used for the study presented "The Steppers" Club", an envisioned persuasive system for encouraging stair use in company buildings. Using RFID-technology, the Stepper"s Club System (SCS) would be able to recognize company employees and dynamically adjust the ambience (lighting and music) of the staircase according to their preferences. While climbing stairs, employees would earn points, which could be exchanged for snacks and coffee. SCS also consisted of public displays on every floor and a web application. The public display would show the ranking of a user in relation to other employees and the web application would provide a more analytical, detailed display of statistics, scores and ranking. SCS remained at conceptual level, and was never implemented.

3.1.1. Fidelities in practice

The two prototypes differed in terms of video tracks and had exactly the same audio track, which consisted of the background music and a voice-over explanation created using Ivona text-to-speech software (http://www.ivona.com/). The actors in the high fidelity video were researchers, having no professional training in acting. The video was filmed at a University Campus (Figure 1). Simple visual effects were applied to the video in post-processing (for example, adding color to the

staircase). Overall, it took approximately thirty-two person hours to prepare for, film and edit the high fidelity video.

Figure 1: *Screenshots from low fidelity (left) and high fidelity (right) versions of the video prototype.*

The low fidelity video was filmed on a tabletop using the same equipment as for the high fidelity video. The sketches depicted the scenes that were filmed in the high fidelity video. No digital animation was used. Two researchers were switching between the sketches during transition between the scenes and their hands were visible in the video. This was done intentionally to give a playful mood to the presentation, and speed up the filming by avoiding excessive stop-motion animation. The iMovie software was used for editing this video. It took approximately twelve person hours to prepare for, film and edit the low fidelity video.

3.2 Method

The study consisted of two parts: online and in-person. The online part included a questionnaire (explained later) assessing attractiveness and usability of the video, as well as, series of closed and open questions. During both parts, participants completed the same questionnaire. During the in-person part, participants were also interviewed following the structure of the open questions section of the online part of the study.

3.2.1. Participants
The online questionnaire was completed by 99 participants (59 males, 40 females). Their age ranged from 17 to 62 (mean=28.1 SD=8.7). Half of them watched the high fidelity video, the other half the low fidelity video. Similarly, 9 participants (4 males, 5 females) were recruited for the in-person study. Their age ranged from 25 to 40 (mean =28.08, SD=8.03). Overall, 108 participants (63 males, 45 females) took part in a study, with ages between 17 and 62.

3.2.2. Measures
To evaluate the attractiveness and usability of the concept, the AttrakDiff questionnaire (Hassenzahl and Monk, 2010) was used. This questionnaire measures attractiveness in terms of Pragmatic Quality (PQ) and Hedonic Quality (HQ). PQ is a measure of usability, whereas HQ measures the extent to which users identify with the product (HQ-I) and how stimulating they find it (HQ-S). Lastly, the questionnaire measures the overall attractiveness (ATT) of the product as perceived by the user.

To evaluate technology acceptance, an adaptation of the well-known UTUAT inventory (Venkatesh et al 2003) was used. Participants were asked to rate to what extent they would use the system (intent to use), how clear is the idea to them (understandability), how easy they think it is to use it (ease of use) and if they think the system could be built in near future (feasibility) on a five-point Likert scale. Finally, the questionnaire included five optional open questions, which concerned the participants˝ likes, and dislikes related to the system, as well as desirable additions or modifications to the system.

Hypotheses

H1: The high fidelity video will elicit a higher number of comments from the participants than the low fidelity one.

H2: The number of different issues discovered will differ between the two prototypes.

H3: The low fidelity video will invite viewers to provide more suggestions regarding additional functionalities.

H4: The appearance of the system in the high fidelity video prototype will be rated as more attractive than in the low fidelity one.

3.2.3. Interviews
Qualitative feedback was obtained via face-to-face interviews. During these interviews, participants were first shown the video and then asked to fill in the AttrakDiff questionnaire. These results were included in the quantitative analysis. Feedback related to feasibility, understandability and other aspects mentioned above was collected through a semi-structured interview. The intent of these questions was to gather qualitative data that might provide insights into results from the quantitative analysis.

4. RESULTS

4.1 AttrakDiff Data Analysis

Overall, data from 108 participants (99 online and 9 interviews) was collected for the AttrakDiff questionnaire. The analysis was performed using data from 105 participants, as three outliers were removed from the analysis. A total of 70 participants answered the open questions in the online questionnaire, thus in all, qualitative data from 79 participants were collected.

For all four dimensions of the AttrakDiff2 questionnaire, the seven corresponding bipolar

differential items (ranging from -3 to 3) were averaged and checked for internal consistency. All four dimensions had good internal consistency, with Cronbach"s alpha coefficients of (PQ) .68, (HQI) .72, (HQS) .82, and (ATT) .92.

To compare the scores of PQ, HQI, HQS, and ATT between the two videos, four independent samples t-tests were performed. A significant effect of fidelity was found on PQ [t(103)=-2.317, p=.022, η^2=.223] and on ATT [t(97.941)=-2.137, p=.035, η^2=.204]. For both dimensions, scores were higher in the low fidelity condition. On the other hand, HQI [t(103)=-1.220, p=.22] and HQS [t(101.204)=-.226, p=.82] did not show any statistical significance. All means and standard deviations are presented in "Table 1". Thus, the system in the low fidelity video prototype was rated both as more attractive (ATT) and more usable (PQ) than the high fidelity one.

Including both videos and all four scales together, mean scores ranged between 0.6 and 1.5 (on a scale where -3 is negative and 3 is positive). This implies that both systems were rated as rather neutral, with a slightly positive inclination. There are no clear differences between HQ and PQ, with both HQI and HQS, also rated as neutral.

Table 1: *Descriptive statistics for four AttrakDiff dimensions (scale of -3 to +3).*

Dimension	Condition	M	SD	N
Pragmatic Quality (PQ)*	*High fidelity*	.73	.74	54
	Low fidelity	1.08	.83	51
Hedonic Quality – Identity (HQI)	*High fidelity*	.53	.90	54
	Low fidelity	.74	.78	51
Hedonic Quality – Stimulation (HQS)	*High fidelity*	.88	1.00	54
	Low fidelity	.92	.83	51
Attractiveness (ATT)*	*High fidelity*	1.07	1.07	54
	Low fidelity	1.46	.80	51

4.2 Additional questions

Table 2: *Descriptive statistics for four additional questions in the online questionnaire (scale of 1 to 10).*

Dimension	Condition	M	SD	N
Intent to Use*	*High fidelity*	3.67	1.29	54
	Low fidelity	4.18	1.11	51
Understandability	*High fidelity*	4.70	.50	54
	Low fidelity	4.67	.82	51
Ease of Use	*High fidelity*	4.15	.89	54
	Low fidelity	4.41	.83	51
Feasibility	*High fidelity*	3.89	1.11	54
	Low fidelity	3.98	1.14	51

Additional questions asked in the online questionnaire referred to four aspects: intent to use, understandability, ease of use and feasibility. The responses to these questions were analyzed using an independent samples t-test. No significant differences were discovered for the perceived feasibility, ease of use and understandability of the concept of the system across two videos. However, ratings of intent to use the system [t(103)=-2.168, p=.032, η2=.209] were found to be significantly different between two videos. "Table 2" presents the results from analysis of these dimensions.

4.3 Qualitative data analysis

The analysis of qualitative data was conducted using Chi"s (1997) approach. Open coding of comments resulted in 20 categories, (freedom, aesthetics, sustainability, ambience, health, interface, points, acceptance, competition, goal, reward, social, technology, privacy, feasibility, fun, video style, efficiency, simplicity, general). A closed coding also categorized comments as positive, neutral and negative. This was done to determine the number of issues in the system (negative comments) in order to test H2. Additionally, comments were analyzed according to the aspects of the prototype they referred to (i.e. the video, underlying idea or system). Finally, there was a category named suggestion, assigned to the comments wherein the participants were suggesting any type of additional functionality that wasn"t existent in the presented system. This was done in order to test H3. The described process was applied to a total number of 441 comments coming from 70 users that completed the online survey, since not all users chose to provide comments in general. The comments for in-person interviews were not used for quantitative analysis as the mean number of comments for in-person interviews (34.1) was substantially higher than those from online survey (6.43) and resulted in a very high standard deviation. Two examples of the emergent coded categories are shown below:

- **Freedom** - participation in SCS is voluntary and not forced. Example: (I like) "that it doesn"t force you to take the stairs".
- **Aesthetics** - the visual appearance of the system. Example: (I would add) "A more attractive visualization of results."

Using these categories the comments were quantified and further analysed using an independent samples t-test to test the hypotheses stated previously (see Table 3). In this test, data from 68 participants (37 high fidelity, 31 low fidelity) was analyzed as two data points were removed as outliers. Data was found to be normal according to Shapiro-Wilk"s test for normality (p=.2). No significant difference was found between the two

videos; therefore, hypothesis H1 could not be accepted.

Table 3: Descriptive statistics for number of comments per video type.

Dimension	Condition	M	SD	N
Suggestions	High fidelity	1.73	1.68	37
	Low fidelity	1.26	1.44	31
Positive comments	High fidelity	1.95	1.68	37
	Low fidelity	2.39	2.67	31
Negative comments	High fidelity	2.27	2.05	37
	Low fidelity	2.29	1.99	31
Total comments	High fidelity	5.92	3.80	37
	Low fidelity	6.94	3.29	31

To compare the number of problems discovered with the concept, the problems belonging to category "negative" and concerning all particular aspects of the system (except for "general" and "video style") were counted and compared across the video types. Data from participants were compared using an independent samples t-test and no significant difference (p=.83) was discovered across the two categories. In conclusion, both, H2 and H3 were rejected.

5. DISCUSSION

H1 was rejected, i.e., no evidence could be found that the higher fidelity video would help elicit more comments than the low fidelity one (animated cut-outs). This finding contrasts with related literature which suggests that more comments would be received for a high fidelity video prototype as compared to a low fidelity video. This could be the effect of another factor not related directly to fidelity. The low fidelity video, as described earlier, may be perceived as unconventional and playful while the high fidelity one is not as fun to view, and looks amateurish despite the higher fidelity representation of the prototype and the higher realism achieved. As a result, as seen in "Table 3", the mean number of positive comments per participant was higher for the low fidelity prototype, while the mean number of neutral comments was slightly higher for the high fidelity video. Additionally, the mean number of negative comments was almost the same for both the videos.

The evidence reported here supports the contentions of Rudd, Stern, and Isensee (1996) that participants could have provided more positive comments in general for the low fidelity video. However, contrary to what is often assumed for lower fidelity videos, participants found the lower fidelity video to be more attractive. This might have

implications for a further nuance in the high and low fidelity debate pertaining to the video medium.

Even low fidelity video prototypes presented in a novel manner can be used to promote ideas with stakeholders, despite them being lower cost. If this is the case, well designed, attractive low fidelity videos could be recommended over higher cost productions.

H2 was rejected, as there were no significant results suggesting differences in the amount of comments relating to problems discovered. This result suggests that there is little added value in investing resources in a high fidelity video prototype, when the goal is to discover problems in the concept.

H3 was rejected, as there was no significant difference in users" suggestions about the concept. This was unexpected as, traditionally, low fidelity prototypes are supposed to „spark" the creativity of the user or to allow them more freedom with comments. Nevertheless, this finding also does not support the investment in producing high fidelity videos, when evaluation aims at eliciting creative feedback from users.

H4 was rejected, as the low fidelity video prototype was rated as more attractive. As mentioned before, the playful character of the video and the novel style could have resulted in this finding. Also, users expressed a greater willingness to use the system after watching the low fidelity video in comparison to the high fidelity video. This could be because they found the low fidelity video more attractive and simpler, as shown by the higher number of comments found for this video in the "simplicity" category. One could argue that this representation obfuscates the complexity of the interaction, something that is only too easy to do when making video prototypes (Tognazzini,1994). Admittedly, one could reasonably criticize the study design in that respondents could not have distinguished between the attractiveness of the filming (where the lower fidelity appeared superior) and the attractiveness of the system shown. Future studies will examine this issue more closely.

The videos made for this research showed that while less were resources used for low fidelity video prototype, low fidelity prototypes do not necessarily result in a different type of feedback. While revealing we recommend that this result is still treated with some caution. Other parameters like physical space limitations, available infrastructure, and skills of the team building the prototypes should be taken into account. In the present case, the fewer person hours required for the creation of the low fidelity video were possibly balanced by the proficiency of the creators, who were able to create sketchy but convincing visuals.

For a design team with limited sketching and illustration experience, shooting a video with actors

may be a cheaper and more efficient way to create a video prototype.

A general recommendation coming out of this study is that low fidelity video is more efficient and equally effective as a prototyping medium, though this may depend upon recruiting proficient sketchers and video directors for that purpose.

6. CONCLUSIONS AND FUTURE WORK

This study transposed the long standing debate regarding fidelity of prototypes, in the domain of video prototyping. Like much of the prior research, the experimental comparison of user feedback did not reveal any differences between the two levels of fidelity. Given that low fidelity still means lower cost, this study adds to the evidence supporting lower fidelity. This is particularly interesting as video prototyping could already be low–cost even in terms of producing higher fidelity representations (using real people and physical contexts). This remark also points at the numerous ways in which fidelity may vary in a video, for which the potential impact is yet unknown. For example, filming quality is a dimension specifically relevant to video prototypes, in all its possible manifestations. Originality, humour, directing and editing, dialogues, acting – all these aspects may impact the viewer perception of a concept visualized in the video. It would be unwieldy to attempt experimentally to evaluate all these options, however, future research could attempt to unpack the extent to which investing in filming quality can be justified. In this current case though, a smaller effort led to a more attractive result showing the value for good craft skills and creativity in the context of video prototyping.

REFERENCES

Catani, M.B. and Biers, D.W. (1998) 'Usability evaluation and prototype fidelity: Users and usability', in *Human Factors and Ergonomics Society Annual Meeting Proceedings*, Human Factors and Ergonomics Society.

Chi, M.T.H. (1997) 'Quantifying qualitative analyses of verbal data: A practical guide', *Journal of the Learning sciences*, vol. 6, no. 3, pp. 271--315, Available: 1050-8406.

Hassenzahl, M. and Monk, A. (2010) 'The Inference of Perceived Usability From Beauty', vol. 25, no. 3, pp. 235--260, Available: 0737-0024.

Mackay, W.E., Ratzer, A.V. and Janecek, P. (2000) 'Video artifacts for design: bridging the Gap between abstraction and', in *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, ACM.

McCurdy, M., Connors, C., Pyrzak, G., Kanefsky, B. and Vera, A. (2006) 'Breaking the fidelity barrier: an examination of our current characterization of prototypes and an example of a mixed-fidelity success', in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 06th edition, New York, NY, USA: ACM.

Rudd, J., Stern, K. and Isensee, S. (1996) 'Low vs. high-fidelity prototyping debate', *interactions*, vol. 3, no. 1, pp. 76--85, Available: 1072-5520.

Sauer, J., Franke, H. and Ruettinger, B. (2008) 'Designing interactive consumer products: Utility of paper prototypes and effectiveness of enhanced control labelling', *Applied Ergonomics*, vol. 39, no. 1, pp. 71--85, Available: 0003-6870.

Sauer, J. and Sonderegger, A. (2009) 'The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion', *Applied Ergonomics*, vol. 40, no. 4, pp. 670--677, Available: 0003-6870.

Sefelin, R., Tscheligi, M. and Giller, V. (2003) 'Paper prototyping-what is it good for?: a comparison of paper-and computer-based low-fidelity prototyping', in *CHI'03 extended abstracts on Human factors in computing systems*, ACM.

Tognazzini, B. (1994) 'The "Starfire" video prototype project: a case history', in *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, ACM.

Virzi, R., (1989) What can you learn from a low-fidelity prototype. Proceedings of the Human Factors Society, 33, 334-338.

Virzi, R., Sokolov, J.L., Karis, D. (1996) Usability problem identification using both low- and high-fidelity prototypes. In Proceedings CHI"96, ACM, New York, NY, USA, 236-243. Walker, M., Takayama, L. and Landay, J.A. (2002) 'High-Fidelity or Low-Fidelity, Paper or Computer Choosing Attributes When Testing Web Prototypes', in Human Factors and Ergonomics Society Annual Meeting Proceedings, 5th edition, Human Factors and Ergonomics Society.

Wiklund, M.E., Thurrott, C. and Dumas, J.S. (1992) 'Does the fidelity of software prototypes affect the perception of usability', in Human Factors and Ergonomics Society Annual Meeting Proceedings, 4th edition, Human Factors and Ergonomics Society.

Wong, Y.Y. (1992) 'Rough and ready prototypes: Lessons from graphic design', in Posters and short talks of the 1992 SIGCHI conference on Human factors in computing systems, ACM.