# The Benefits of Combining Paper- and Video- Based Prototypes for User Interface Evaluation

Hayet Hammami[†*], Fatoumata Camara[§], Gaëlle Calvary[*], Meriem Riahi[†] and Faouzi Moussa[†]
[*]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG
F38000 Grenoble France
Email: FirstName.LastName@univ-grenoble-alpes.fr
[†]Univ. of Tunis El Manar, Faculty of sciences of Tunis, LIPAH-LR11ES14
2092 Tunis Tunisia
Email: faouzimoussa@gmail.com, meriem.riahi2013@gmail.com
[§]HWR Berlin, Berlin Germany
Email: fatoumatag.camara@gmail.com

***Abstract*—The use of multiple User Interface (UI) designs for evaluation has been demonstrated beneficial for UI evaluation as it results in better feedback, both qualitatively and quantitatively. However, producing several designs is time-consuming. Moreover, the properties that the alternative UI must satisfy remain under-explored. The paper investigates the use of different prototype forms of the same design as support to evaluation instead of relying on alternative design solutions. We investigate two experimental conditions: (1) paper prototype first then video prototype, and (2) video prototype first then paper prototype. Results show that the combination of paper and video prototypes is well suited for UI evaluation, as feedback addresses all aspects of Human-Computer Interaction (HCI), namely, utility, usability, and aesthetics. When exposed to multiple prototypes, users develop an understanding of the functional core and of the interactive aspect of the system. The experiment outcomes indicate that, when evaluating the paper prototype first, then the video prototype, users tend to be more critical and provide more suggestions of improvements.**

***Keywords–UI evaluation; Feedback; Prototyping; Video prototyping; Comparative evaluation.***

## I. INTRODUCTION

The number of UI designs used for evaluation influences responsiveness of users as well as the amount and quality of feedback. Therefore, submitting different design examples could help UI testers see issues clearly, identify concrete steps for improvement, and integrate novel ideas [1].

Many research papers addressed the use of multiple design alternatives for UI evaluation (comparative evaluation) [1]–[5]. These alternatives consist of design variations at several levels of abstraction, such as syntactic and semantic levels. They can be designed by the same designer(s) [2] [4], or obtained via targeted research such as the visual aspects or the content [1].

Comparative evaluation increases the amount of comments (reviews and suggestions), gives rise to more and stronger criticisms, and facilitates comparative reasoning. Consequently, showing multiple design alternatives to users for UI evaluation represents *a way to get the right design*.

Previous work clearly highlights that comparative evaluation has great benefits. However, producing alternative design(s) can be time-consuming and difficult, particularly considering that existing literature does not address criteria that should be considered for the generation of the alternative design(s).

In this work, we investigate the use of different prototype forms as support to evaluation. Our aim is to determine whether the use of different forms could be as beneficial as different designs so that it could be considered as an alternative to multiple designs for evaluation. In this paper, we report an experimental evaluation in which we used both paper and video as prototype mediums for UI evaluation.

The remaining of this paper is organized as follows. Section II presents related work about system prototyping. Section III describes the experiment and the design elaboration. Section IV reports results and observations from the experiment and Section V concludes the paper and presents future work.

## II. MANY FACES OF PROTOTYPING

Prototypes can be of different levels of fidelity and can take different forms. The right level of fidelity and appropriate form of a prototype depends on the design stage and evaluation needs. A prototype can be of low-fidelity, medium-fidelity or high-fidelity with respect to the final UI [6]. According to Greenberg [7], "*The determining factor in prototype fidelity is the degree to which the prototype accurately represents the appearance and interaction of the product*".

Much of the often-cited literature emphasizes the use of low/medium fidelity prototypes [8] [9]. When assessing low fidelity prototypes, users feel more included in the design process and feel more free to criticize the design [10]. Oppositely, when evaluating high fidelity prototypes for the first time, users tend to focus on the details of the interface (e.g. color of icons, size of font) rather than on the overall structure [7]. Furthermore, it has been found that the usability data collected from low and high-fidelity levels are comparable [9] [10] [11].

Moreover, a prototype can take different forms, such as a sketch [8], a paper mock-up [4] or a tool to test [12]. Designers can also use presentation software like PowerPoint or Keynote [13], or videos [14] [15] in order to illustrate the dynamicity of interaction.

### A. Paper prototyping

Techniques such as paper prototyping excel at representing static visual properties [15]. They are very used during the

 ISBN: 978-1-61208-761-0 

design process due to their low cost and efficiency. They can perform the role of sketching in order to develop, explore, communicate and evaluate the designer's ideas [2] [8].

Paper prototyping presents a fast and easy way to communicate and to test initial ideas early and quickly, e.g., brainstorming sessions. Paper prototypes help in getting substantive user feedback, promote rapid iterative development [10], and allow for easy and inexpensive modifications. Moreover, paper prototypes can be very helpful for usability testing as mentioned in Tohidi et al.'s work: "*the interaction designer communicates with the user largely by means of an "interactive sketch", such as a paper prototype*" [8].

### B. Video prototyping

Video prototypes, on the other hand, are particularly well suited to assess interaction. Video prototyping is an established technique in HCI, which is typically used in early designs stages to enable software designers to evaluate the interaction prior to actual software implementation. Producing a video prototype is cheaper and less time-consuming in comparison to building a fully working prototype.

Video prototypes present great benefits [11] [14]–[17], for instance, they allow design exploration, evaluation and presentation [16] [17]. They also communicate and reflect the interaction design [15] and help capture and communicate the details of how users interact with software. Zwinderman et al. [18] compared the use of video prototypes and usability tests. They analyzed results regarding overall user experience (AttrackDiff), user acceptance (Unified Theory of Acceptance and Use of Technology (UTAUT)), and five "expectations" elaborated by the researchers in this work:

- Participants who use the product give more comments on the interface.
- Viewers of the video prototype make more comments on the context of use.
- Participants provide a similar number of comments as to when and where they use the application.
- Viewers of the video prototype suggest more new features.
- Participants who use the product suggest more improvements.

The conclusion of this study suggests that video prototypes help obtain feedback from users that is quite similar to that gathered when users test the final product.

In other works, researchers studied the impact of the visual fidelity-level of video prototypes on the amount and quality of feedback provided by users during evaluation. Dhilton et al. [11] compared feedback collected from using a low-fidelity video prototype (animated paper cut-outs) and feedback collected from using a high-fidelity video prototype (a video with real actors, edited to simulate computer output). The video prototype fidelity focused on the notion of realism of the video (Figure 1). Analysis considered attractiveness and usability of the concept (AttrackDiff), intent to use, understandability, ease of use and feasibility of the system presented (UTAUT). Additionally, during test sessions, users were asked to express likes and dislikes related to the system as well as suggestions for improvement. Dhilton et al.'s study concludes that the visual fidelity level of the video prototype does not affect the amount and quality of user's feedback during evaluation sessions.


Figure 1. Low fidelity (left) and high fidelity (right) versions of the video prototype used in [11].

Both paper and video prototypes foster discussion regarding interface and interaction with stakeholders. Video prototypes represent a powerful tool as they produce feedback that can be compared to feedback from other techniques that might be more costly. Both paper and video prototypes are cheap to produce and have been proven beneficial for HCI evaluation. However, to the best of our knowledge, the combined use of paper and video prototypes as support to HCI evaluation has not yet been investigated (particularly as alternative to using multiple design alternatives).

In this work, we use a paper prototype and a video prototype to assess a commercial Website. We investigate the benefits of using both prototype forms presented together in the same evaluation session for UI evaluation. The main research question tackled here is the following: "does using different prototype types as support to HCI evaluation could be as beneficial as relying on alternative UI designs?".

## III. EXPERIMENT

Our study focused on a commercial Web site for high-tech products. The case study was purely academic. The reason for this choice was that the e-commerce platform is widespread and familiar to many people, and thereby easy to explain.

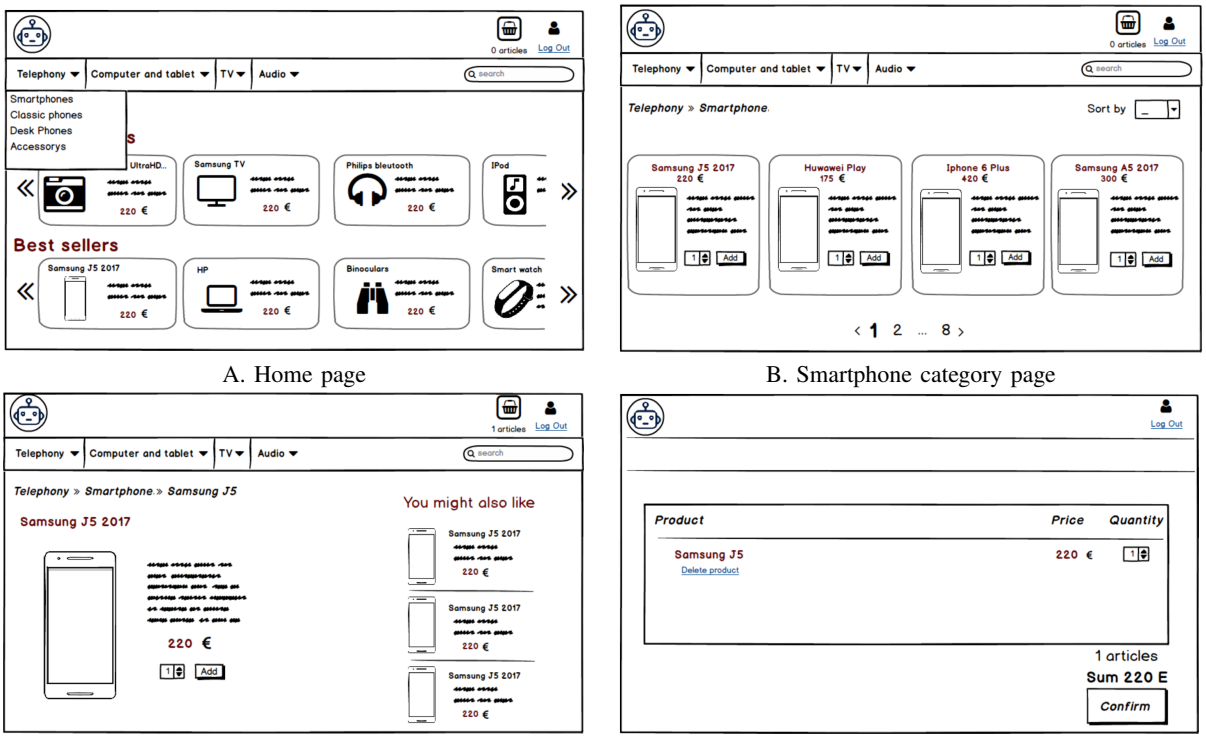
### A. Prototypes elaboration

The prototypes consisted of medium fidelity prototypes, designed using the Balsamiq Wireframes tool [19]. The UI illustrates essential features of the commercial Web site: menu, sub-menus, list of items on the home page, list of items for a specific category of products, details of a product, and content of the cart.

The paper prototype (Figure 2) consisted of four screenshots, illustrating four different pages of the Web site. Figure 2-A represents the home page of the Web site, Figure 2-B represents the page referring to the sub-category smartphones, Figure 2-C represents the page referring to the details of a selected product, and Figure 2-D represents the page referring to the cart. Using a tool like Balsamiq allowed us to assign hypertext links to different components of the interface in order to create functional widgets and navigate between the different pages.

The video prototype was then made by recording interaction through the same UI, using a video screenshot tool. There were no additional pages shown on the video. However, in some cases, the products displayed changed due to navigation in the list (Figure 3).

 ISBN: 978-1-61208-761-0 




A. Home page

B. Smartphone category page

C. Article details page

D. Cart page

Figure 2. Paper-based prototype presented to users

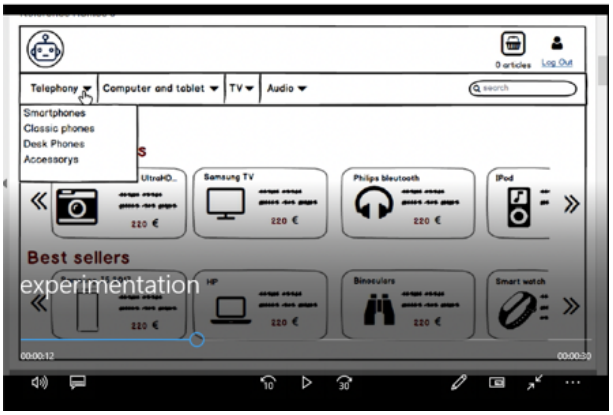



Figure 3. Video based prototype

## B. Scenario

The video is 42 seconds long and features a user choosing and purchasing a smartphone through the Web site. The scenario goes as follows: the user starts by navigating through the home page of the Web site, browsing the deals and the best sellers' lists, then, he/she browses the menu to explore the different categories and sub-categories presented, and selects the sub-category "smartphone". Next, we see the user being redirected to the smartphone sub-category page, from where he/she can choose a product from the list and add it to the cart. After consulting the cart, we see the user changing his/her mind about the product that he/she selected. He/she deletes it from the cart and goes back to the home page to choose another item.

The same person who designed the UI was the one who ran the scenario for the video.

## C. Participants

The technique used for our experiment is primarily a qualitative one, but which allows to collect quantitative data. With qualitative techniques, such as usability tests or interviews, the aim is to dig into topics (i.e., usability problems) while usually observing testers' reactions: as the name implies, the focus is more on quality rather than on quantity. Consequently, qualitative techniques require a low number of testers to get a fair overview over the addressed topics; for instance, as low as 5 participants for a usability tests [20]. Indeed, according to Nielson [20], "*with 5 users, we almost always get close to user testing's maximum benefit cost ratio*". However, it is important to mention that, to get statistically significant numbers, at least 20 participants should be included in the study. Folkler [21] suggest that 20 participants help obtain 95% of usability problems.

Our experiment involved 22 participants (11 women and 11 men, with a range of age between 22 and 58 years old). Participants included both HCI students and researchers as well as people with no knowledge at all in the field. The number of HCI students, females and males were equivalent for each group of participants.

Participants were asked about their online shopping frequency. 13 participants said that they often purchase products online, 7 participants said that they sometimes do online

 ISBN: 978-1-61208-761-0 

shopping, and 2 participants said that they rarely purchase products online.

### D. Protocol

We considered two experimental conditions: paper prototype, then video prototype and video prototype, then paper prototype. Participants were divided into two groups of eleven participants, and were asked to observe and critique both prototypes. Each group tested one experimental condition:

- **Experimental condition 1: paper prototype, then video prototype:** participants were first provided with the paper-based prototype and asked to observe the UI without enforcing any time limit. After looking at the prototype, the participants were asked to evaluate it. As a second step, we replaced the paper-based prototype with the video and asked the participant to watch it. The video could be watched up to three times at most. It was up to the participant to re-watch. However, it was not allowed to pause the video while watching. After watching the video, the participants were asked if they had something to add regarding their first evaluation.
- **Experimental condition 2: video prototype, then paper prototype:** participants were first presented with the video then with the paper-based prototype.

### E. Users feedback record

Participants were asked to provide feedback using their own words regarding three aspects: things they like, things they do not likee and improvements. To do so, they had to write their statements on magnetic post-its and place them as either 'like', 'dislike' or 'improvement' on a board (Figure 4).

Figure 4. Participant expressing his opinions

Most of the time, participants tended to explain what they are writing and expressed their thoughts orally. Observations were recorded by note-taking throughout the experiment. Furthermore, we recorded the evaluation sessions via a Dictaphone.

## IV. Results

This section presents the results of the experiment. We first classify users' feedback, then we discuss the impacts of the two types of prototyping on the UI evaluation.

### A. Categorization of user's feedback

We collected, counted, and classified users' statements in three categories with respect to the ones that were considered to collect feedback: likes (positive comments), dislikes (negative comments) and suggestions for improvement. This first categorization was inspired by the taxonomy elaborated in [2], and used in [4]. We used it to study the user's willingness to criticize during evaluation sessions. Then, we gathered statements inductively according to three criteria commonly considered in UI evaluation: utility, usability and aesthetics, in order to study users' feedback.

The total number of statements collected for both groups was approximately the same: 106 for the first group (paper then video) and 104 for the second group (video then paper).

As indicated in Table 1, in both experimental conditions, the number of statements provided with the prototype being presented at first was significantly higher than the number of statements collected with the second prototype. Indeed, during the first step of the evaluation, regardless whether the paper or the video prototype was presented first, the number of statements was roughly the same.

TABLE I. NUMBER OF STATEMENTS.

| | *Step 1* | *Step 2* | *Total number of statements* |
|---|---|---|---|
| Group 1 (Paper then video) | 77 | 29 | 106 |
| Group 2 (Video then paper) | 79 | 25 | 104 |

### B. Impact on user's willingness to criticize depending on the prototype presented and evaluation condition

In order to assess the impacts of presenting different forms of prototypes on participants' willingness to be critical, we compared the number of positive comments, negative comments and suggestions for improvement according to the order of prototypes presentation: paper then video (group 1) and video then paper (group 2).

Figure 5 shows how the total numbers of statements produced were classified. The first observation in Figure 5 is

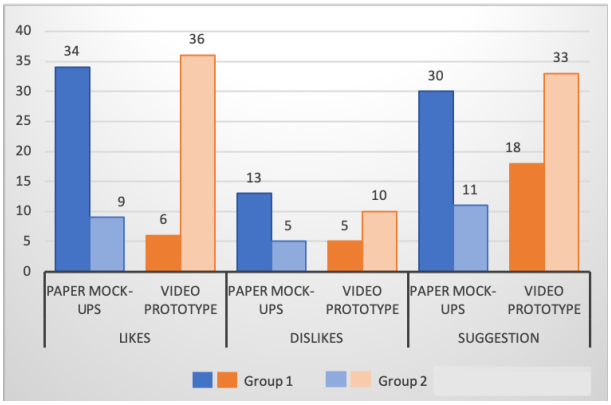



Figure 5. Numbers of positive comments (likes), negative comments (dislikes), and suggestions

that the number of likes and suggestions are higher than the negative comments for both groups. However, it is important to note that the number of dislikes added to the number

of suggestions is higher than the number of likes for each prototype for both groups.

The number of dislikes is significantly lower than the numbers of likes and suggestions. This result can be explained by the overlap between dislikes and suggestions. Indeed, during the experiment, participants often hesitated to choose between dislikes and suggestions for many comments.

Examples of likes included feedback about the access to best sellers on the welcome page and the simplicity of the UI. Examples of suggestions included feedback about integrating a more sophisticated search function and using vertical instead of horizontal scrolling. Examples of dislikes included feedback about the lack of the navigation menu in the cart page and the lack of the "add" button in the best deals list.

An important observation is that the feedback collected in the second step of the evaluation for both groups consisted mainly in suggestions and negative comments. Indeed, reevaluating the same UI design, but, presented through a different type of prototype pushed users to be more critical and to provide more insights about improvements.

Overall, Overall, regardless of what prototype was presented first, users started by expressing their likes and appreciations of the prototype in question. However, users in the first group provided more suggestions in the second step of the evaluation than the second group. Seeing the video prototype after the evaluation of the paper prototype pushed users to be more critical and to provide more suggestions for improvements.

Based on these results, we can say that the experiment highlighted the positive aspects of the design, while also gathering an even greater number of insights for design improvement.

### *C. Impact on users' feedback depending on the prototype presented and evaluation condition*

We qualitatively analyzed the data in order to relate participants' statements to utility, usability and aesthetics within each prototype and evaluation order. Feedback about utility addressed the system features, including comments such as deleting or adding a feature to the Web site. We considered usability as defined by the International Organization for Standardization (ISO): "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [22]. Feedback regarding usability included, for instance, the possibility to add a product directly from the welcome page; the absence of the navigation menu in the cart page; the possibility to add several products to the cart at the same time; or, the absence of an information message once a product is added or deleted from the cart. Feedback about aesthetics considered the graphic design (e.g., colors, icons design, fonts, widget choices, etc.) and included statements about the choice of the icon chosen as the Web site logo and the colors used.

Figure 6 shows how the total number of statements produced were classified.

Results show that the paper prototype produced the highest number of feedback related to utility whilst the video prototype produced the highest number of feedback related to usability.

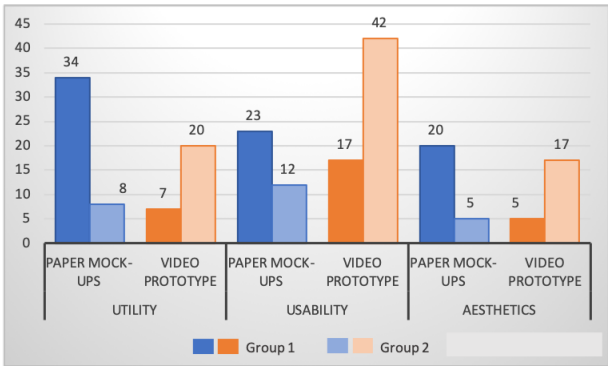



Figure 6. Numbers of statements according to utility, usability, and aesthetics.

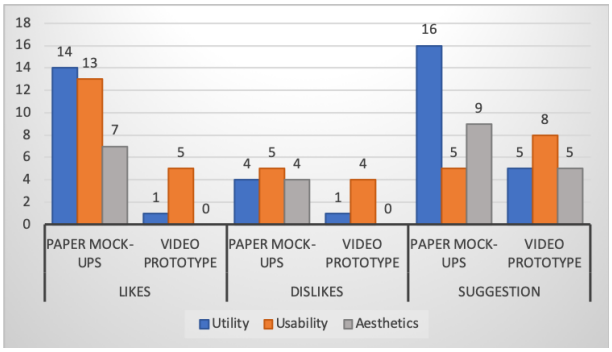



Figure 7. Numbers of statements according to utility, usability, and aesthetics for the first evaluation order (paper then video).

Finally, the numbers of feedback regarding aesthetics are roughly the same for the two prototypes over the two groups.

Furthermore, we summarize below the qualitative data providing indications according to the two mentioned classifications in order to compare the number of problems discovered and suggestions provided regarding each aspect (utility, usability and aesthetics) for each prototype for both groups.

Figures 7 and 8 show how the total numbers of statements produced were classified.

Results show that participants detected more problems and provided more suggestions in experiment condition 1 (paper

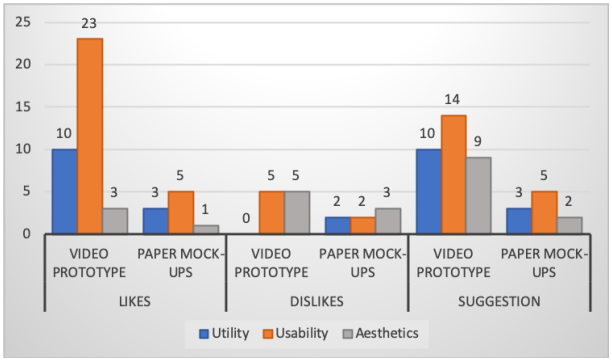



Figure 8. Numbers of statements according to utility, usability, and aesthetics for the second evaluation order (video then paper).

then video) than in experiment condition 2 (video the paper). For example, the number of likes regarding usability provided with the video prototype in the second group was similar to the number of dislikes and suggestions provided about this HCI aspect. However, in the first group, when comparing the number of likes with the number of dislikes and suggestions about utility, we find that the number of likes is considerably lower.

## V. CONCLUSION AND FUTURE WORK

This study explores a different approach to UI evaluation, which could guide practitioners towards getting the design right while minimizing the cost of UI design.

We conducted an experiment which consisted in using a medium-fidelity paper prototype and a medium-fidelity video prototype as support to evaluation and tested two experimental conditions. Users successively evaluated either a paper prototype then a video prototype, or a video prototype then a paper prototype.

The results indicate that (1) using paper prototypes allowed users to focus on features offered by the system ; actually, users took time to explore the system features, as such, they developed an understanding of the functional core and criticized mainly utility. (2) Using a video prototype allowed users to focus more on the interaction with the system, i.e., how a user can perform a specific task; as such, by seeing the system 'in action', they developed an understanding of the interactive aspects of the system and focused on usability. (3) Regarding aesthetics, feedback provided by users was comparable within both groups.

Moreover, it is important to note that the order of prototypes presentation to users does matter since evaluating the video prototype after evaluating the paper prototype incites users to be more critical and to provide more suggestions of improvements.

Overall, the results indicate that evaluating one prototype is not enough. Over both evaluation conditions, when seeing the UI in a different prototype form, users discovered problems and provided suggestions of improvements that did not occur to them when evaluating the first prototype provided.

A general recommendation coming out of this study is to supplement video prototypes with paper prototypes at first for UI evaluation, in order to increase the evaluation benefits.

In future work, we are interested in comparing evaluation based on different types of prototypes and evaluation based on alternative design solutions in order to identify the most beneficial approach in terms of relevant feedback at a minimum cost.

## REFERENCES


[1] H. B. Kang, G. Amoako, N. Sengupta, and S. P. Dow, "Paragon: An online gallery for enhancing design feedback with visual examples," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–13.

[2] M. Tohidi, W. Buxton, R. Baecker, and A. Sellen, "Getting the right design and the design right," in Proceedings of the SIGCHI conference on Human Factors in computing systems, 2006, pp. 1243–1252.

[3] S. Dow, J. Fortuna, D. Schwartz, B. Altringer, D. Schwartz, and S. Klemmer, "Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2011, pp. 2807–2816.

[4] H. Hammami, G. Calvary, M. Riahi, F. Moussa, and S. Bouzit, "Comparative evaluation? yes, but with which alternative ui?" Electronic Visualisation and the Arts (EVA 2017), 2017, pp. 1–7.

[5] R. S. Dicks, "Mis-usability: on the uses and misuses of usability testing," in Proceedings of the 20th annual international conference on Computer documentation, 2002, pp. 26–30.

[6] A. Coyette, S. Kieffer, and J. Vanderdonckt, "Multi-fidelity prototyping of user interfaces," in IFIP Conference on Human-Computer Interaction. Springer, 2007, pp. 150–164.

[7] S. Greenberg, "Prototyping for design and evaluation," November, vol. 30, 1998, p. 2004.

[8] M. Tohidi, W. Buxton, R. Baecker, and A. Sellen, "User sketches: a quick, inexpensive, and effective way to elicit more reflective user feedback," in Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, 2006, pp. 105–114.

[9] M. E. Wiklund, C. Thurrott, and J. S. Dumas, "Does the fidelity of software prototypes affect the perception of usability?" in Proceedings of the Human Factors Society Annual Meeting. SAGE PublicationsSage CA: Los Angeles, CA, 2016.

[10] M. Walker, L. Takayama, and J. A. Landay, "High-fidelity or low-fidelity, paper or computer? choosing attributes when testing web prototypes," in Proceedings of the human factors and ergonomics society annual meeting, vol. 46, no. 5. SAGE Publications Sage CA: Los Angeles, CA, 2002, pp. 661–665.

[11] B. Dhillon, P. Banach, R. Kocielnik, J. P. Emparanza, I. Politis, A. Rczewska, and P. Markopoulos, "Visual fidelity of video prototypes and user feedback: a case study," in Proceedings of HCI 2011 The 25th BCS Conference on Human Computer Interaction, 2011, pp. 139–144.

[12] H. Kim, C. Coutrix, and A. Roudaut, "Morphees+ studying everyday reconfigurable objects for the design and taxonomy of reconfigurable uis," in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–14.

[13] P. I. Khella. Use keynote and powerpoint to prototype web and mobile apps. [Online]. Available: https://keynotopia.com/ (Retrieved: July, 2019)

[14] W. E. Mackay, "Using video to support interaction design," DVD Tutorial, CHI, vol. 2, no. 5, 2002.

[15] G. Leiva and M. Beaudouin-Lafon, "Montage: A video prototyping system to reduce re-shooting and increase re-usability," in Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, 2018, pp. 675–682.

[16] W. E. Mackay, A. V. Ratzer, and P. Janecek, "Video artifacts for design: Bridging the gap between abstraction and detail," in Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques, 2000, pp. 72–82.

[17] W. E. Mackay, "Video prototyping: a technique for developing hypermedia systems," in CHI'88 Conference Companion Human Factors in Computing Systems, vol. 5. Citeseer, 1988, pp. 1–3.

[18] M. Zwinderman, R. Leenheer, A. Shirzad, N. Chupriyanov, G. Veugen, B. Zhang, and P. Markopoulos, "Using video prototypes for evaluating design concepts with users: a comparison to usability testing," in IFIP Conference on Human-Computer Interaction. Springer, 2013, pp. 774–781.

[19] Balsamiq wireframes. [Online]. Available: https://balsamiq.com/ (Retrieved: December, 2019)

[20] J. Nielsen, "How many test users in a usability study," Nielsen Norman Group, vol. 4, no. 06, 2012.

[21] L. Faulkner, "Beyond the five-user assumption: Benefits of increased sample sizes in usability testing," Behavior Research Methods, Instruments, & Computers, vol. 35, no. 3, 2003, pp. 379–383.

[22] ISO, "Ergonomics of human-system interaction–part 11: Usability: Definitions and concepts," 2018.


 ISBN: 978-1-61208-761-0 