A Video Prototyping Methodology for Evaluating Novel Interface Concepts in Cockpit Displays

Paul McKay

Stacey D. Scott

Jonathan M. Histon

Gerard L. Torenvliet

A VIDEO PROTOTYPING METHODOLOGY FOR EVALUATING NOVEL INTERFACE CONCEPTS IN COCKPIT DISPLAYS

Paul McKay, Stacey D. Scott, Jonathan M. Histon
Department of Systems Design Engineering, University of Waterloo
Waterloo, Ontario, Canada

Gerard L. Torenvliet
Esterline|CMC Electronics Inc.
Ottawa, Ontario, Canada

Modern cockpit displays contain a multitude of complex information sources. Integrating new interface concepts into an existing cockpit display to produce a high-fidelity prototype suitable for user testing can be extremely time-consuming. Discount prototyping methodologies are needed to enable user testing at earlier stages of the design cycle to ensure appropriate changes occur and high quality interfaces result. Video prototyping can provide a useful step between low-fidelity, static prototypes and higher-fidelity software prototypes. However, existing video prototyping methods are designed to elicit user feedback on design concepts. While user feedback is important to the adoption of aviation interfaces, it is also desirable to examine performance using more complex metrics, which have traditionally required the development of a fully interactive software prototype. We propose a new scenario-driven video prototyping methodology that allows designers to apply complex metrics during early-stage user evaluations.

Developing and assessing display concepts for complex, dynamic task environments, such as modern aircraft cockpits, can be both time consuming and resource intensive. The development of a prototype cockpit environment of sufficient fidelity for use in human participant evaluation can require as much as several months of effort from one or more software development experts. In an effort to reduce the work required to develop or modify testable dynamic interfaces, we examined the concept of video prototyping, which has been common in the human-computer interaction (HCI) domain for the past two decades (Mackay, Ratzer, & Janecek, 2000; Vertelney, 1989).

Video prototyping can provide a useful step between lower-fidelity static prototypes, such as paper or still image prototypes, and higher-fidelity software prototypes. However, use of these video prototypes has traditionally been limited to demonstrating interface concepts as a way of obtaining user feedback (Bardram, Bossen, Lykke-Olesen, Nielsen, & Madsen, 2002; Bardzell et al., 2006; Halskov & Nielsen, 2008; Mackay, et al., 2000; Muller, 1991; Tognazzini, 1994; Vertelney, 1989; Young & Greenlee, 1992). While obtaining user feedback is useful and important in the early stages of the user centered design process, the low fidelity and lack of interactivity in these non-software prototypes typically prevents them from being used in later stage performance-based interface evaluations. In the past, this has meant that examining the performance of a design concept based on complex concepts such as situation awareness required the development of a fully interactive software prototype. We propose that a new form of scenario-driven video prototyping, using video authoring techniques to show proposed display concepts layered atop existing interfaces, can provide several advantages over traditional prototyping techniques.

The primary advantage of the proposed technique is that it enables the development or modification of testable dynamic interfaces with reduced time and effort compared to traditional software prototyping. This reduction is possible because the work required is similar to adding special effects to films, meaning that it shifts the type of tools and expertise required from software development to traditionally lower-fidelity techniques like video editing and graphic design. Additional advantages associated with the proposed new technique include:

- the reduction in development time and cost enables user testing to be carried out earlier in the design process;
- the ease of creating overlays on an existing interface allows for rapid, low-cost user interface testing; and
- the scenario-driven methodology allows for performance testing based on complex concepts such as situation awareness.

In the following sections, we overview existing video prototyping techniques and identify limitations that led to the development of our proposed technique, describe the methodology for creating and evaluating a video prototype, and present a case study showing our use of this methodology to develop and evaluate a new display concept for the aviation context.

Video Prototyping

The use of video in the prototyping process began in the late 1980s with initial work primarily using video to record physical prototypes being manipulated by the designers to show their concept of use (Muller, 1991; Vertelney, 1989; Young & Greenlee, 1992). Further work using video as a prototyping tool has developed other ways of recording the design concepts, including using performers to show interaction with mock-ups of an interface (Tognazzini, 1994), using software rendering tools to generate 'virtual' video prototypes (Bardram, et al., 2002; Halskov & Nielsen, 2008), and using video game characters as virtual performers in prototype videos (Bardzell, et al., 2006). While all of these techniques are useful for communicating a design idea, they share two main limitations that prevent them from being used to test the utility of an interface design.

First, the scenarios on which these video prototypes are based are typically created by the designer to showcase their interface rather than on representative use cases. In some cases, such as Bardram (2002) and Halskov's (2008) studies of ubiquitous computing in the medical domain, scenarios capture an envisioned world and so may not be an accurate representation of the way the interface will function in an actual implementation. While these 'mock-up' scenarios can work very effectively as a method of demonstrating a design concept for the purpose of soliciting user feedback, they are not sufficiently realistic to be used for formal evaluations.

Second, many video prototypes are intended to function as storyboards, and so portray users interacting with an interface. As a consequence, the interfaces under design are typically not shown in detail, or are in detail for only a limited amount of time. Tognazzini (1994) talks specifically about some of the filmmaking techniques and directing decisions made in creating a video prototype to limit the amount of 'full-resolution' screen time needed. This works well for storyboarding a design concept as it limits the level of interface development required, but it is less useful for formal interface evaluation as it limits the level of detail available to be studied.

Methodology

To address the limitations with existing video prototypes, we developed a methodology for creating high resolution video prototypes based on realistic scenario data. The methodology has been developed for use at an intermediate stage of the user-centered design process (Preece, Rogers, & Sharp, 2002). The process can be applied to either new interfaces, or modifications of existing interfaces. It is assumed that several iterations of lower fidelity prototypes have been previously developed and evaluated to arrive at a relatively mature design concept. The key elements of the methodology being proposed can be divided into three phases.

Phase 1: Scenario Development and Data Collection

The first step is to develop an appropriate scenario and collect relevant data for use in creating a video prototype. The scenario should be a representative use case for the system being studied, so it will generally be necessary to consult with subject-matter experts to ensure that the scenario is realistic. There are many analysis tools available to assist in selecting and developing a representative scenario for evaluation, such as task analysis (Crandall, Klein, & Hoffman, 2006; Diaper & Stanton, 2003) or cognitive work analysis (Vicente, 1999). In the development and initial case study of this methodology, the analysis method we used was a form of operational sequence modeling (Chapanis, 1996).

Once a representative scenario is selected, data must be collected so that the scenario can be developed into a set of videos showing the proposed interface design in representative task conditions. The relevant data will vary depending on the scenario and the system being designed, but the primary component will be high-resolution video of the interface design showing the state of the display throughout the scenario. This video can be created using a screen-capture tool, such as Camtasia Studio[1], which records live interface graphics and user interactions during system usage. Other potentially relevant data to be recorded could include audio, interaction (cursor input, keystrokes, etc), or gaze-tracking information, among others.

[1] http://www.techsmith.com

Phase 2: Video Prototype Creation

Using the data recorded in the first phase of the video prototype methodology, the captured video can be synchronized with the other captured data (audio, interaction, etc) to produce a first set of control videos. These videos are then edited to produce a treatment set of videos showing the proposed interface design. This can be accomplished by using a standard video editing suite to create additional video channels with graphical overlays to show the modifications to the existing interface or the dynamic elements of a new interface.

Phase 3: Video Prototype Evaluation

The final phase of the methodology is the evaluation of the video prototype with representative users to examine the effects of the interface modification or new interface design. This evaluation takes a similar form to a typical usability evaluation with the exception that the participants are not able to interact with the prototype, and therefore need to be given tasks that are appropriate for the scenario context but still allow for assessment based on the desired metrics. Examples of such tasks could include: recording a log of scenario events, evaluating the performance of participants in the video prototype, or additional external tasks such as manual control.

Case Study – Supporting Collaboration in Modern Cockpits

The methodology described above has been used to evaluate an innovative interface design for advanced cockpits. New cockpit avionics architectures are emerging that use cursor control devices and keyboards for pilot interaction with individual and shared displays. This form of architecture has a number of advantages compared to a conventional glass cockpit, but brings some challenges as well. One of these challenges, resulting from the concentration of avionics controls into a keyboard and cursor control device, is crewmembers’ potential loss of peripheral awareness cues of each other’s actions. The design we developed in this case study aimed to restore some of this lost information by augmenting an existing interface design with information about operator usage history, including both input (keyboard or mouse) and visual (gaze) activity. To ensure that the design concept for visualizing this usage history information was sufficiently mature for performance testing, it was developed through several iterations of low fidelity prototypes before moving onto the video prototyping evaluation.

Phase 1: Capturing a Flight Scenario

An existing high-fidelity software cockpit interface prototype of an advanced two-pilot cockpit that enabled virtual flight simulation was used in phase 1 to develop a representative flight scenario and to collect data of in-flight cockpit display interactions. Three participants (two pilots and one air traffic controller) were recruited to act out the scenario. Display usage of the primary flight and navigation displays were recorded using the FRAPS[2] screen capture tool, while usage of the flight information display was captured using Camtasia Studio. A digital camcorder with lavalier microphones was used to record the prototype setup, including the radio and intercom conversation from the three participants. A gaze tracking system was used to capture the visual interaction data of the co-pilot, while the pilot’s visual interaction data were approximated based on a post-scenario interview, screen capture videos, and the wide angle video of the prototype setup.

Phase 2: Creating Control and Treatment Video Prototypes

In phase 2, the collected data were then used to create a control and a treatment set of video prototypes. The treatment videos showed visual traces of operator usage history by placing color-coded borders around the interface components that were viewed or edited by the pilots. The opacity of these borders were adjusted to indicate the recency of use (i.e., when an interface component was viewed or edited, the border for that component was set to full opacity, and would fade away over time when the component was not being used). The Adobe Premiere[3] video editing tool was used to create the video prototypes. Each prototype contained the interface sequences and interface usage history (treatment condition) of 30-minutes of flight scenario captured in Phase 1. Figure 1 (left) shows an un-treated snapshot from one of our screen capture videos, and Figure 1 (right) shows the same snapshot after editing.

---

[2] http://www.fraps.com
[3] http://www.adobe.com/products/premiere

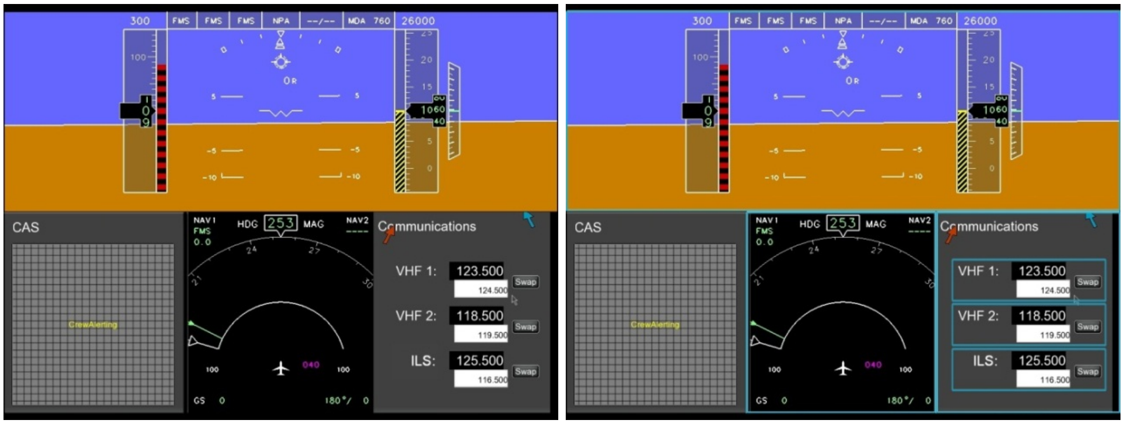

Figure 1. Un-treated screen capture snapshot (left) and proposed interface treatment (right).

Phase 3: Video Prototype Evaluation

In the final phase, a formal user evaluation of the developed video prototype was then conducted with representative users. An experimental display setup was assembled with a form similar to the prototype cockpit that was used as the basis for the design (Figure 2). Eleven trained pilots with a minimum of 15 flight hours participated in the study. A between-subjects experimental design was used, in which the participants watched the videos for either the control condition (five participants) or treatment condition (six participants).


Figure 2. Prototype cockpit used in video prototyping evaluation.

Participants were given two main tasks to perform while watching the scenario videos. Their first task was to take the role of an evaluator. This involved paying attention to the events of the scenario and the actions of the flight crew and, after the scenario, rating their individual and group performance. The second task involved completing a scenario log sheet by recording information about a variety of relevant flight information (such as radio frequency changes and ATC clearances), and the timing of flight events. The primary evaluation task was selected because its continuous cognitive aspect would make the secondary task challenging; instead of simply watching and listening for cues about information relevant to the second task, they needed to pay attention to the scenario events and integrate them into an overall understanding of the scenario and the performance of the flight crew.

Requiring participants to complete a log sheet provided a measure of how well they understood the scenario events and the actions of the flight crew (i.e., a basic measure of situation awareness). A measure of situation awareness was obtained by comparing each participant’s log sheet to a master log sheet that included all possible events. The reliability of the log sheet information depended on participants' cooperation with the data recording process (i.e., participants may understand more information than they record on the log sheet). Events were considered to have been recorded correctly if the participants recorded the correct information (e.g., a new radio frequency) at approximately the correct time (within one minute before or after the actual time).

After the scenario, participants completed a brief questionnaire that included three questions asking them to rate the crew’s performance and one question asking them to rate their confidence that their scenario log sheet captured all the relevant scenario information. Each question used a 7-point Likert-style rating scale. Finally, participants were interviewed using a semi-structured process that elicited additional details on three general topics: the post-scenario questionnaire, the information on the scenario log sheet, and the cockpit interface. Participants in the treatment condition were asked additional questions dealing specifically with the interface augmentation.

Case Study Results and Implications

The video evaluation provided unique and valuable insights into the strengths and weaknesses of the proposed interface treatment. The most interesting results arose from the participant debriefing interviews. Five of the six treatment condition participants reported using the usage history information to maintain awareness of important scenario information. One of these participants, who missed a radio frequency change while looking at a chart, noticed and recorded the change when he looked back up at the display; the interface treatment helped the participant gain awareness of the radio frequency change and, thus, supported their awareness of the situation.

The case study also highlighted an important challenge in performing the video evaluation: selection of the task for observers, and developing relevant quantitative measures of variables of interest. In the case study, the quantitative measure of participant awareness using the log sheet information did not show any statistically significant differences between the treatment and control groups; however, this was likely a result of two main limitations with the evaluation process. First, the use of a between-subjects design with a small sample size made it unlikely that any potential differences between groups would be detected. This limitation could be relatively easily addressed in future studies by increasing the sample size or by using a within-subjects experimental design.

The second limitation of the evaluation process was the use of the scenario log sheets as the primary measure of participant awareness. The self-reported nature of the log sheets made it difficult to ensure that performance was measured consistently across participants. More highly experienced pilots, for example, tended to report much less information, yet demonstrated a clear understanding of the scenario events during the post-scenario interviews. It is possible that experienced pilots decided to focus on watching and evaluating the actions of the flight crew for the purposes of the flight evaluator task, knowing that they could recall the flight event details from memory if needed, while participants with less experience were not as confident in their ability to evaluate the flight crew and instead focused on the log sheet task. Applying a standardized situation awareness measurement technique such as SAGAT (Endsley, 1990) may help reduce such variations in recorded awareness data in the future.

Discussion

The goals for the video prototyping methodology discussed in this paper were to allow user testing at a lower time and cost compared to conventional software prototyping techniques, and to allow this user testing to examine the performance of a design concept based on complex metrics such as situation awareness. The results from our initial use of the methodology indicate that it does have the potential to succeed at both of these goals. The use of video for prototyping allowed a single researcher with little or no software development experience to collect the necessary data, create a prototype, and conduct a performance evaluation with a similar level of effort as would have been required for a software expert to develop an interactive prototype. Additionally, the evaluation using the video prototypes generated results that demonstrated both the utility and the limitations of the proposed interface design concept for supporting awareness.

Methodology Considerations

In addition to addressing the limitations discussed above, several other considerations are relevant for future use of this video prototype evaluation method. Perhaps the most important consideration is the task participants are asked to perform during the evaluation. While the flight crew evaluator task worked well for our scenario as it approximated the task of monitoring a highly automated aircraft, such an evaluation task may not be suitable to other domains; the task must be tailored to the domain and scenario being studied.

Another consideration relates to the use of approximate data for the pilot’s point-of-gaze. It was initially unclear whether participants would easily notice a difference in activity between the pilot and co-pilot visual borders, possibly leading participants to distrust the interface treatment and begin to ignore it. However, of the six participants in the treatment condition, only one participant mentioned a difference, observing that the co-pilot’s eye movements seemed to dart around more than the pilot’s. Obtaining real gaze data for use in the prototype did have some benefits in that it made the process of prototyping the visual borders somewhat faster and ensured that they were accurate representations of the co-pilot’s eye activity, but these benefits came at a cost of the time required to set up and calibrate the gaze tracker and analyze the point-of-gaze data. We suspect that using only the approximate method would have greatly accelerated phase 1 with little cost to the realism of the video prototype.

Based on this successful use of approximate gaze tracking data, it is possible that creation of the video prototypes could be further accelerated by using approximate data for other aspects of the prototype. For example, in our evaluation, it is possible that phase 1 could have been accomplished using the automation of the existing cockpit prototype and simulating the input interaction and audio stream for the two pilots and air traffic controller. In this way, the video prototype could have been created by consulting a single expert pilot to confirm that the prototypes were realistic, instead of using three pilots to act out the scenario. It is also possible that the video prototype method could be used in cases where the interface design is not based on an existing prototype by creating a video based on a still image interface design and realistic approximations of a usage scenario.

Conclusion

This paper has presented a new discount prototyping methodology suitable for testing complex and dynamic interface concepts, such as advanced aviation interfaces, at fairly early stages in the design process. The proposed methodology uses a new form of video prototyping that adapts film-like special effects applications to digital video screen captures of existing interfaces. The proposed method enables designers of complex interfaces to begin performance testing of novel display concepts much earlier in the overall design process, and make appropriate modifications, before extensive and costly software development is needed.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Council of Canada (NSERC). We thank members of the University of Waterloo Human Factors Research Group for feedback on the methodology and case study interfaces. We also thank participants of the data collection and user evaluation phases of this research.

References

Bardram, J., Bossen, C., Lykke-Olesen, A., Nielsen, R., & Madsen, K. H. (2002). *Virtual video prototyping of pervasive healthcare systems*. Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques, London, England.

Bardzell, J., Bardzell, S., Briggs, C., Makice, K., Ryan, W., & Weldon, M. (2006). *Machinima prototyping: an approach to evaluation*. Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles, Oslo, Norway.

Chapanis, A. (1996). *Human Factors in Systems Engineering*. New York, New York, USA: John Wiley & Sons, Inc.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: a practitioner's guide to cognitive task analysis*. Cambridge, MA: The MIT Press.

Diaper, D., & Stanton, N. (Eds.). (2003). *The Handbook of Task Analysis for Human-Computer Interaction*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Endsley, M. R. (1990). A methodology for the measurement of situation awareness. *Situational Awareness in Aerospace Operations*, 1/1-1/9.

Halskov, K., & Nielsen, R. (2008). Virtual video prototyping. *Human-Computer Interaction, 21*(2), 199-233.

Mackay, W. E., Ratzer, A. V., & Janecek, P. (2000). *Video artifacts for design: bridging the Gap between abstraction and detail*. Proceedings of the conference on Designing interactive systems, New York City, New York, United States.

Muller, M. J. (1991). *PICTIVE- an exploration in participatory design*. Proceedings of the SIGCHI conference on Human factors in computing systems, New Orleans, Louisiana, United States.

Preece, J., Rogers, Y., & Sharp, H. (Eds.). (2002). *Interaction Design: Beyond Human-Computer Interaction*: John Wiley & Sons, Inc.

Tognazzini, B. (1994). *The "Starfire" video prototype project: a case history*. Proceedings of the SIGCHI conference on Human factors in computing systems, Boston, Massachusetts, United States.

Vertelney, L. (1989). Using video to prototype user interfaces. *SIGCHI Bulletin, 21*(2), 57-61.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work.* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Young, E., & Greenlee, R. (1992). *Participatory video prototyping*. Extended abstracts of the SIGCHI conference on Human factors in computing systems, Monterey, California.