Using Multiple iTV Displays for the Simultaneous Comparative Evaluation of Parallel Prototypes

Pierre-Alexandre Lapointe
User Experience Director
Yu Centrik
1654 Mont-Royal est
Montreal, QC, Canada
1 (514) 523-5862
palapointe@yucentrik.ca

Jay Vidyarthi
UX Designer / Research Coordinator
Yu Centrik
1654 Mont-Royal est
Montreal, QC, Canada
1 (514) 523-5862
jay@yucentrik.ca

Joelle Stemp
President / UX Specialist
Yu Centrik
1654 Mont-Royal est
Montreal, QC, Canada
1 (514) 523-5862
jstemp@yucentrik.ca

ABSTRACT
Best practices in iTV interface design can often be well represented in a wide range of design solutions. In an attempt to build and evaluate an optimized interactive TV interface under various constraints, we selected a combination of usability evaluation methodologies and used three high-definition television displays to obtain comparative user data on parallel prototypes simultaneously. The resulting data was instrumental in fueling subtle interaction decisions which led to a focused and optimized human-computer interface.

Categories and Subject Descriptors
D.2.2 [**Software Engineering**]: Design Tools and Techniques – *user interfaces.*
H.5.2 [**Information Interfaces / Presentation**]: User Interfaces – e*valuation/methodology*, *prototyping*, *user-centered design*

General Terms
Measurement, Design, Experimentation, Human Factors, Theory.

Keywords
iTV, User Experience, Usability Testing, Methodologies, Interface Evaluation, Human-Computer Interaction

1. INTRODUCTION
In technological design practice, scientific literature often plays the role of defining the constraints within which designers build effective human-computer interfaces. Inevitably, decisions concerning the effectiveness of contextual and subtle interaction elements which are beyond the scope of this literature must be agreed upon by project stakeholders. In 2009, brainstorms and discussion led to a collection of legitimate design concepts for the interactive TV interface of a major telecommunications provider in Canada. In such a situation, many have advocated the design of parallel prototypes to enable the direct comparison of different solutions [5, 6, 8, 10]; however, traditional qualitative and quantitative usability testing methods become laborious, repetitive, inefficient and redundant when used comparatively on multiple solutions to the same design problem [4]. As our project's time and budget were constrained, we needed to select an effective and efficient interface evaluation method to elicit user data on a number of subtly different design concepts.

2. METHODOLOGICAL SELECTION
Lacking an established usability methodology tailored to our situation, we explored the combination of related methods.

Nielsen [8] has described the benefits of designing and comparing parallel interaction modes for a single human-computer interface. The strength of this approach lies in combining the best elements of several design concepts, increasing the probability of effectiveness in the conglomerate design [6, 10]. In a similar vein, Buxton has been a proponent of comparative usability testing through the criticism of design processes which settle on a single concept too early [1]. Clearly our design concepts could be expressed through several parallel prototypes, but how could we evaluate them efficiently in a realistic usage context?

Our proposed interfaces all seemed to align with best practices; as mentioned, they differed in more subtle interaction modes. Since the resulting prototypes were all based on a similar framework, there was a clear value in considering comparative usability testing [11, 12]. However, this added even further constraint to the selection of an effective evaluation methodology: serial comparative evaluation would place a significant burden on participants' memory. Forcing participants to recall and discuss such nuanced differences would likely result in participants interacting with the interfaces under a significant cognitive load [7, 9], subtracting from the simulation of a realistic *lean-back* TV-watching experience [2, 3]. Considering this, we had no choice but to rule out these out-of-the-box methodologies; we needed a protocol which could generate firm evidence toward final design decisions within a realistic TV-watching context.

Knowing that the project budget could support 21 participants for an hour and a half each, we weighed the advantages and disadvantages of a few different methods. With a between-subjects design, we would be able to test each prototype in a unique group of 7 participants. However, with such low numbers our results would be fairly vague and inconclusive. On the other hand, a within-subjects design which presented all 21 participants with all 3 prototypes would create significantly more concrete data. Even so, in addition to the aforementioned problems with cognitive load, limited testing time would severely restrict the number of tasks and depth of discussion for each prototype. While searching for a better method, we hypothesized that the comparative evaluation of parallel prototypes might become effective through the use of corresponding parallel displays.

3. RESULTING METHODOLOGY

The resulting test procedure involved three high-definition television screens, each configured with interactive prototypes carefully designed to juxtapose subtle design elements. As a precursor to a retrospective cognitive walkthrough, the test moderator had all 21 users use parallel prototypes to perform 6 realistic TV-watching task scenarios (i.e. record a show using the EPG, choose and purchase a video-on-demand, etc.). After users commented openly, a forced choice comparison was used to determine subjective user preferences. Follow-up questions were also used to determine whether their least preferred interfaces had any specific elements which were superior to the others. Finally, a card sorting exercise was used as a manifestation of the client's desire to place an additional evaluative focus on categorization and labeling.

To build association and aid memory, tasks were initially performed with each interface on its respective screen. During the subsequent walkthrough, user attention was directed toward all three screens to enable the simultaneous consideration of multiple interfaces for each task. We hypothesized that users' familiarity with the scenarios from initial task completion combined with a simultaneous visual presentation of the prototypes would enable effortless and accurate discussion of subtle design differences.

Figure 1, The user's point of view; three high-definition screens are presented simultaneously and labeled 1, 2 and 3 to facilitate and encourage comparative discussion.

4. RESULTS AND CONCLUSION

The proposed methodology was effective in comparing and contrasting both individual features and the experiential whole of our parallel prototypes. For example, between proposed alternatives for the system's start-up interface, one model was clearly preferred. Yet in the case of video-on-demand, users identified advantages with each interface, leading to a hybrid design concept. It is also interesting to note that a wide range of user commentary and opinion significantly evolved the state of every single concept; no interface screen remained as-is.

Based on appropriately selected task scenarios and a unique testing infrastructure, we were able to compile user performance with subjective commentary and explicit preferences to draw concrete conclusions on subtle design issues. We attribute the richness of commentary elicited during the test to the fact that participants appeared to be comfortable with the approach; not only did they compare the prototype with ease, they did not seem to be intimidated by the three-screen interface display. When the testing phase was complete, we had conclusive results to fuel informed decisions on a large majority of our contentious design issues. These results have inspired a new prototype which will be subjected to a more standard usability test in 2010.

Every design project for every interface in every context has its own unique requirements. Identifying an appropriate methodology can be daunting. This case demonstrates how a creative and combinatory approach can help provide a tailored and focused methodology. Constraints inherent in our specific situation narrowed our approach to a few key methodologies. While keeping project objectives in mind, we were able to combine the most pertinent elements of relevant methodologies into an innovative hybrid approach which was effective in focusing our design efforts and moving the project forward.

5. ACKNOWLEDGEMENTS

The authors would like to thank our participants, the client organization and the EuroITV committee administration.

6. REFERENCES

[1] Buxton, B. 2007. *Sketching User Experiences: Getting the Design Right and the Right Design.* An Imprint of Elsevier. Morgan Kaufmann Publishers, Amsterdam, Netherlands.

[2] Chorianopoulous, K. 2007. Interactive TV Design That Blends Seamlessly with Everyday Life. In *Universal Access in Ambient Intelligence Environments*. Stephanidis, C. and Pieper, M., Eds. Springer, Berlin, Germany. 43-57.

[3] Griffiths, M. 2003. Internet Gambling: Issues, Concerns and Recommendations. *Cyberpsychology & Behaviour*. 6(6): 558-559

[4] Kostiainen, K., Uzun, E., Asokan, N. & Ginzboorg, P. 2007. *Framework for Comparative Usability Testing of Distributed Applications.* Nokia Research, Helsinki, Finland.

[5] Macbeth, S.A., Moroney, W.F., & Biers, D.W. 2000, Development and evaluation of symbols and icons: A comparison of the production and focus group methods, In *Proceedings of the IEA 2000/HFES 2000 Congress*, San Diego, CA, March 2000. 327-329.

[6] McGrew, J. 2001. Shortening the human computer interface design cycle: A parallel design process based on the genetic algorithm. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, Minneapolis, MN, October 2001. 603-606.

[7] Mullan, B. 1997. *Consuming Television*. Oxford, Blackwell Publishers.

[8] Nielsen, J. 1993. *Usability engineering*. An Imprint of Elsevier. Academic Press, San Diego, CA, 85-87.

[9] Olson, J. & Olson, G. 1990. The Growth of Cognitive Modeling in Human-Computer Interaction since GOMS. *Human-Computer Interaction*, Volume 5, 221-265.

[10] Ovaska, S. & Raiha, K.J., 1995. Parallel Design in the Classroom. In *Proceedings of the ACM CHI 95*, Denver, Colorado. ACM Press, New York, NY, 264-265. DOI= http://doi.acm.org/10.1145/223355.223666

[11] Rubin, J. 1994. *Handbook of Usability Testing.* John Wiley & Sons.

[12] Tullis, T. 2009. Tips for Usability Professionals in a Down Economy. *Journal of Usability Studies*, Volume 4, Issue 2, 60-69.