Quantitative and Qualitative Methods for Human-Subject Experiments in Virtual and Augmented Reality

J. Edward Swan II, Mississippi State University, Organizer
<swan@acm.org>

Joseph L. Gabbard, Jr., Virginia Tech
<jgabbard@vt.edu>

Fully Day VR 2014 Tutorial Proposal

Abstract

This tutorial is for researchers and engineers, working in the field of Virtual and Augmented Reality, who wish to conduct user-based experiments and/or evaluations for assessing usability. We propose a full-day tutorial presenting both quantitative and qualitative approaches to conducting human-subject experiments. It will cover (1) the basic principles of experimental design and analysis, with an emphasis on human-subject experiments in AR (Swan), and (2) qualitative studies (e.g., formative evaluation methods) for assessing and improving AR user interfaces and user interaction along with lessons learned from conducting many user-based studies (Gabbard).

Swan, Gabbard, and other co-presenters have taught pre-cursor versions of this tutorial 11 previous times at IEEE Virtual Reality, IEEE Visualization, and ISMAR. This tutorial was most recently given at ISMAR 2012, where we included updated examples from our research and further expanded upon qualitative approaches for assessing usability and lessons learned from conducting studies. We both have current, active AR human-subject research projects, and if this tutorial is accepted to be presented at VR 2014, we will discuss some of these projects as case studies.

Motivation

In the early days of the IEEE VR community, most of the research and development effort was on just getting the technology to work at all. However, over the past 15 years or so, the performance of VR/AR technology has steadily improved to the point where many groups and companies are now developing interesting applications of VR/AR technology. Furthermore, over the past 5 years or so, the widespread adoption of smartphones means that millions of people carry a VR/AR-capable platform in their pockets. This is greatly increased the number of industrial concerns and startups developing VR/AR applications.

In order to assess how VR/AR applications are experienced by users, it is necessary to conduct human-subject experiments. We have both observed that the IEEE VR community has long valued human-subject experiments, and so we will not belabor that point here.

Certainly, professional gatherings such as IEEE VR are vary appropriate places to educate researchers and engineers on the benefits of human-subject experiments. Moreover, we believe a large portion of IEEE VR attendees did not formally study experimental design, data analysis, or formal usability methods as part of their degree programs, but instead are learning to do so "on

the job." Similarly, this tutorial will be taught by computer scientists who did not study these topics in school, but who have since learned the skills through many years of conducting usability evaluations and human-subject experiments.

Further, in recent years we have seen greater interest in qualitative approaches to assessing the quality of VR/AR user interfaces, leveraging approaches that aim to elicit higher level cognitive qualities of VR/AR user interface such as ease of use, intuitiveness, and insight (as opposed to measuring time-on-task or user error). By presenting both quantitative and qualitative approaches to experimentation, along with guidelines on when to use each approach, this tutorial will provide timely and much needed instruction for the IEEE VR community.

Target Audience

Who should attend: Researchers and engineers, working in the AR/VR field, who wish to either (1) conduct evaluation experiments with human subjects, and / or (2) gain a better understanding of the basic terminology of experimental design and analysis (e.g., the precise meaning of statements such as "$F(2,45) = 5.67$, $p = .023$"), and / or (3) are researching or developing AR/VR interfaces or whole applications that can benefit from qualitative user-based assessment (e.g., interfaces that are at or beyond prototyping phases and are readying for potential broader use).

Level of expertise: All Levels. (*Note*: This material is basic, and during previous times when this tutorial has been offered, we have been surprised to see attendees that we knew had backgrounds in experimental psychology or human-computer interaction. When we later asked them what value they felt they received from the tutorial, they indicated that although they had previously been exposed to the material, it had been years ago, and they still found the review helpful. We have therefore concluded that this material seems to be useful to attendees with multiple levels of expertise).

Outline of Tutorial

(1) *Experimental Design and Analysis* (Swan: morning). This session will introduce the basics of experimental design and analysis and will include illustrative case studies of actual human-subject experiments conducted by Swan and collaborators. It will introduce the basics of experimental design and analysis. Especially regarding experimental analysis, it will focus on the fundamental logic behind topics such as hypothesis testing and analysis of variance, while avoiding the complexities that come from considering individual statistical tests.

Topics include:

- generating empirically testable hypotheses
- experimental validity
- standard statistical designs
- independent and dependent variables
- experimental design and counterbalancing
- statistical tests
- gathering data
- describing data
- inferential statistics
- hypothesis testing
- power and effect size analysis
- analysis of variance (ANOVA)

- reporting statistical results
- lessons learned from 20 years of conducting human-subject experiments

(2) *Formative Usability Evaluation* (Gabbard: afternoon). This session introduces user-based formative evaluation methods aimed at iteratively improving MAR user interfaces. We will present details of how to prepare and conduct a formative usability evaluation and provide specific case studies to ground the discussion. We will also discuss how human-subjects experiments and formative evaluations can be used as separate methods, or as part of a larger usability engineering process — for example, illustrating how these evaluations can inform user interface design in absence of established design guidelines and metaphors. This is especially important given the novel aspect of many VR/AR user interfaces.

Topics include:
- scoping and planning formative usability evaluations
- developing representative user tasks
- conducting formative usability evaluations, with an emphasis on qualitative data collection (e.g., understanding *why* users find an VR/AR component intuitive or confusing)
- translating evaluation results to design recommendations
- key differences and synergies between formative usability evaluations and human-subjects experiments
- when to use formative usability evaluations as opposed to human-subjects experiments

(3) *Panel Discussion* (Swan and Gabbard: late afternoon). The last four times that we have taught this tutorial, towards the end of the day we have presented example experiments and evaluations as case studies. As we have done so, we have encouraged the audience to ask questions and/or give examples from their own experiments and evaluations, and this has naturally lead to a panel-type discussion, where we have engaged in a more general question-and-answer discussion. We have also found that, by the end of the day, everyone is tired of looking at slides and having a general discussion is refreshing and engaging. Often, these discussions continue over dinner.

Learning Objectives and Improvement Plans

Many tutorials survey current research topics; clearly these tutorials require updating as these research topics evolve. In contrast, this tutorial covers the broad topic of human-subject experimentation, while focusing on those aspects that the presenters have found to be most helpful when conducting experiments in AR. Therefore, the tutorial does not cover a specific research topic, but instead techniques that can be foundational to a variety of research topics. So although the tutorial has been updated and improved each time it has been given, the updates and improvements have primarily involved pedagogical considerations. For this submission, we have included approaches that provide higher-level cognitive assessment of AR user interfaces, such as intuitiveness and insight, as well as, an updated set of new lessons learned on, and considerations for planning and conducting human-subjects experiments.

In this tutorial, we have always sought audience interaction, and we have often been very successful at developing this interaction. This interaction — discussions with attendees — is the primary way that we have evaluated the success of this succession of tutorials. We have also altered what we have presented and our presentation style based on attendee feedback.

In fact, the long history of this tutorial is primarily driven by attendee feedback. Over the years we have both occasionally heard from attendees from years past who have told us that they found

the tutorial very helpful in their own research and application development. We have also been told, several times, that our tutorial was the best-attended at the conference; this has happened at both Virtual Reality and Visualization. This positive feedback is the primary reason why we are still teaching this tutorial — as any professor knows, positive student feedback, especially years later, is absolutely the best part of teaching.

Forms of Presentation and Tutorial History

Our presentation style is simple: Power Point slides presented in lecture style. The key to engaging the audience really comes from the example experiments and case studies, and the fact that many attendees are struggling to design human-subject experiments themselves, and often have questions.

If accepted, this will be the 12th time that this tutorial has been taught by Swan, Gabbard, and others. In particular, it was taught at the *IEEE Virtual Reality* conference in 2004, 2005, 2006, 2007, 2009, and 2012, at *IEEE Visualization* in 2006, 2007, 2008, and 2011, and at ISMAR in 2012.

After several years, we started writing down the number of attendees at various times during the tutorial. Upon looking back over these numbers, attendance at Virtual Reality has varied from an average of ~15 to ~25, with a peak of 37. At Visualization (a larger conference) attendance has varied from ~20 to ~30, with a peak of 53. Although the lecture style means that a large number can attend, the best discussions happen between 15 and 30; beyond about 30 attendees the social pressure involved in speaking up becomes larger and attendees don't say as much.

Publicizing Tutorials

To date, this tutorial has only been publicized by being listed on conference publicity materials, and through an abstract posted on the Advance Program. Because the material is general, we do not think that it will be necessary to use email lists or social media to reach potentially interested attendees.

Presenter Bios and Contact Information for Presenters

Dr. J. Edward Swan II is a Professor of Computer Science and Engineering, and an Adjunct Professor of Psychology, at Mississippi State University. He holds a B.S. (1988) degree in computer science from Auburn University and M.S. (1992) and Ph.D. (1997) degrees in computer science from Ohio State University, where he studied computer graphics and human-computer interaction. Before joining Mississippi State University in 2004, Dr. Swan spent seven years as a scientist at the Naval Research Laboratory in Washington, D.C. Dr. Swan's research has been broad-based, centering on the topics of augmented and virtual reality, perception, human-computer interaction, human factors, empirical methods, computer graphics, and visualization. Currently, Dr. Swan is studying perception in augmented and virtual reality, including depth and layout perception and depth presentation methods, as well as empirical techniques for evaluating and validating visualizations. His research has been funded by the National Science Foundation, the National Aeronautics and Space Administration, the Naval Research Laboratory, and the Office of Naval Research. Dr. Swan is a member of ACM, IEEE, the IEEE Computer Society, and ASEE.

email: swan@acm.org

phone: (662) 325-7057

Dr. Joseph L. Gabbard, Jr. is an Associate Professor of Industrial and Systems Engineering at Virginia Tech. He holds B.S. (1995), M.S. (1997), and Ph.D. (2008) degrees in computer science from Virginia Tech. He also holds a B.A. in Sociology (1993) from Virginia Tech. Dr. Gabbard's work centers on human-computer interaction; specifically usability engineering for novel user interfaces including (but not limited to), augmented reality, virtual environments, visualizations for life sciences, and multimodal interactive systems. Gabbard has been a pioneer in usability engineering with respect to applying to, and creating methods for, new interactive systems for more than 15 years. Currently, Dr. Gabbard is conducting a series of human-subjects experiments to better understand color perception in outdoor augmented reality. He also is applying formative usability evaluation methods to bioinformatics visualizations. His research has been funded by the National Science Foundation, the National Institutes of Health, the Naval Research Laboratory, and the Office of Naval Research.

email: jgabbard@vt.edu
phone: (540) 231-3559

Significant Joint Presenter Publications

Swan and Gabbard have been collaborating since 1997, and have many joint publications. A selected list includes:

- Joseph L. Gabbard, J. Edward Swan II, and Adam Zargar. Color Blending in Outdoor Optical See-through AR: The Effect of Real-world Backgrounds on User Interface Color. In *Proceedings of IEEE Virtual Reality 2013 Short Papers and Posters*, pp. 157–158, March 2013.
- Joseph L. Gabbard, J. Edward Swan II, "More Than Meets the Eye: An Engineering Study to Empirically Examine the Blending of Real and Virtual Color Spaces", Technical Papers, *Proceedings of IEEE Virtual Reality 2010*, Waltham, Massachusetts, USA, March 20-24.
- Joseph L. Gabbard, J. Edward Swan II, "Usability Engineering for Augmented Reality: Employing User-Based Studies to Inform Design," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 3, pp. 513-525, May/June, 2008.
- Joseph L. Gabbard, J. Edward Swan II, Deborah Hix, Si-Jung Kim, Greg Fitch, "Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications", Technical Papers, Proceedings of *IEEE Virtual Reality 2007*, Charlotte, North Carolina, USA, March 10-14, pages 35-42.
- Joseph L. Gabbard, J. Edward Swan II, Deborah Hix, "The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality", *Presence: Teleoperators & Virtual Environments*, Spring 2006, Vol. 15, No. 1, Pages 16-32.
- Mark A. Livingston, J. Edward Swan II, Joseph L. Gabbard, Tobias H. Höllerer, Deborah Hix, Simon J. Julier, Yohan Baillot, Dennis Brown, "Resolving Multiple Occluded Layers in Augmented Reality", *Technical Papers*, *The 2nd International Symposium on Mixed and Augmented Reality* (ISMAR '03), Tokyo, Japan, October 7–10, 2003, IEEE Computer Society, pages 56–65.

- J. Edward Swan II, Joseph L. Gabbard, Deborah Hix, Robert S. Schulman, Keun Pyo Kim, “A Comparative Study of User Performance in a Map-Based Virtual Environment”, Technical Papers, *IEEE Virtual Reality 2003*, 2003, pages 259–266
- Deborah Hix, J. Edward Swan II, Joseph L. Gabbard, Mike McGee, Jim Durbin, Tony King, *“*User-Centered Design and Evaluation of a Real-Time Battlefield Visualization Virtual Environment*”*, Proceedings *IEEE Virtual Reality ‘99*, IEEE Computer Society Press, 1999, pages 96–103.
- Joseph L. Gabbard, Deborah Hix, J. Edward Swan II, “User-Centered Design and Evaluation of Virtual Environments”, *IEEE Computer Graphics and Applications*, Volume 19, Number 6, November / December, 1999, pages 51–59.