Techniques for User Evaluation

Lecture #12: User Evaluation

Joseph J. LaViola Jr.

Fall 2013

Usability Testing

- Not exact science (but we try!!)
- Want to evaluate users
  - performance
  - preference
  - feedback
- Goals
  - learn about individual UI techniques
  - learn about applications
  - learn about hardware

Basic Strategy

- "What do I want to learn?"
  - based on observations, theory, etc…
- Generate hypotheses (if applicable)
- Determine how to test the hypotheses
  - experimental setup and design
- Pilot studies
  - confirm study is sound
- Conduct study
- Analyze data
  - use statistics
- Report findings

Experimental Strategies

- Formative – gather feedback on evolving system, set of techniques, etc…
  - examine prototypes to refine system
  - improve UI techniques
- Summative – learn about system as a whole
  - does it do what it is designed to do
- Qualitative approaches
  - survey data, preference data, open ended questions
- Quantitative data
  - time to completion, error, number of clicks. etc…

Experimental Setup

- Want to make user comfortable
- Allow moderator to observe without getting in the way

Subject   Video Camera

Moderator

Experimental Design

- Difficult task
  - need to remove as much variability as possible
  - always want to err on the side of more data collection
  - art more than science
  - conditions (4 x 2, 2 x 2 x 2, etc…)
- Between subjects
  - subjects broken up into groups
  - each group gets one condition
  - requires more subjects
- Within subjects
  - every subject gets every condition
  - less subjects but have to deal with ordering effects
  - slightly harder to analyze
- Mixed
  - combines both between and within

Experimental Procedure

- How is the experiment carried out?
- Need to come up with plan for running subjects
- How does the experiment get administered?
- Need to ensure procedure is the same for all subjects

Pre- and Post-questionnaires

- Pre-questionnaire
  - Want to find about subject background
    - age, gender, handedness
    - particulars about experiment
      - experience with similar software
      - experience in particular area
- Post-questionnaire
  - valuable tool
  - used to gather qualitative data
  - used for qualitative data quantitatively
    - Lickert scale
 - open ended questions

Pilot Studies

- Run one or two subjects through experiment
- Why?
  - make sure experiment is sound
  - make last minute changes to design
  - convince yourself hypotheses make sense

Analyzing Data

- Look for trends, patterns, and statistical significance
- Understanding statistical tests and procedures is crucial
- Need to know
  - what kind of data (nominal, scale, ordinal)?
  - what tests to perform (T-Test, ANOVA, Friedman)?
  - what corrections to make (Bonferroni, Tukey)?
  - how to interpret results (α, confidence intervals, mean, median)?
- Statistical packages are your friend
  - SAS, SPSS, Matlab, etc…
- Sometimes there is no statistical test to apply

Example Experiment

- Not pen-UI related but techniques still apply
- Exploration of non-isomorphic rotation in VE
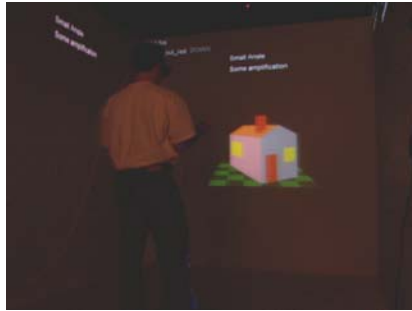
LaViola, J. and Katzourin, M. "An Exploration of Non-Isomorphic 3D Rotation in Surround Screen Virtual Environments", *Proceedings of the IEEE Symposium on 3D User Interfaces 2007*, 49-54, March 2007.

Example Experiment – Goals

- Further explore non-isomorphic rotation of virtual objects
- Systematic evaluation of different rotation amplifications
- Understand benefits of non-isomorphic in SSVE
  - head tracking
  - stereoscopic vision

Example Experiment -Design

- 16 subjects (13 male, 3 female)
- Conducted in Brown "Cave"
- Based on Poupyrev 2000 → Hinckley 1997 → Chen 1988
- 4 x 2 x 2 balanced, within-subjects design (16 conditions)
- Independent variables
  - amplification (1,2,3,4)
  - rotation amplitude (20-60, 70-180 degrees)
  - Error threshold (6, 18 degrees)
- Dependent variables
  - completion time
  - orientation error

Example Experiment – Procedure

- Task – rotate house from random to target orientation
- Pre-questionnaire
- 16 practice trials
- 16 sets of 10 trials each
- Ordering was randomized
- Post-questionnaire

Example Experiment –Results

- Repeated measures, three way ANOVA

Effect	Time	Error
S	$F_{3,13}$=3.26, p=0.056	$F_{3,13}$=4.8, p<0.05
T	$F_{1,15}$=13.66, p<0.05	$F_{1,15}$=22.96, p<0.05
A	$F_{1,15}$=55.46, p<0.05	$F_{1,15}$=0.001, p=0.98
S x T	$F_{3,13}$=0.29, p=0.83	$F_{3,13}$=1.575, p=0.243
S x A	$F_{3,13}$=0.87, p=0.523	$F_{3,13}$=0.562, p=0.649
T x A	$F_{1,15}$=5.03,p<0.05	$F_{1,15}$=0.573, p=0.46
S x T x A	$F_{3,13}$=0.73, p=0.55	$F_{3,13}$=0.97, p=0.436

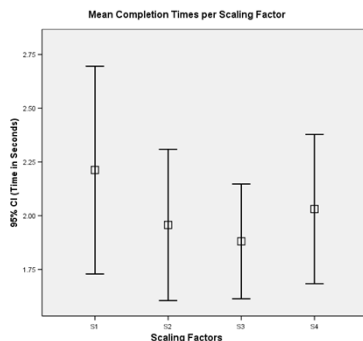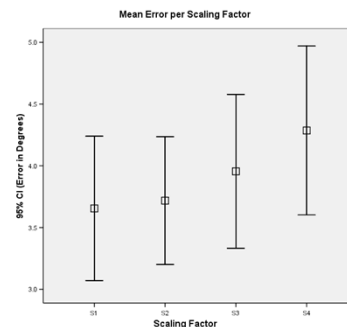**S = scaling factor   T = error threshold   A = angle**

Example Experiment – Results: Post Hoc Analysis

- Pairwise comparisons on scaling factor using Holm's sequential Bonferroni adjustment

Mean Completion Times per Scaling Factor

Mean Error per Scaling Factor

Significant differences between S1 and S2 and S1 and S3

Significant difference between S1 and S4

Example Experiment – Results: Subject Preferences

Subject Scaling Factor Preferences
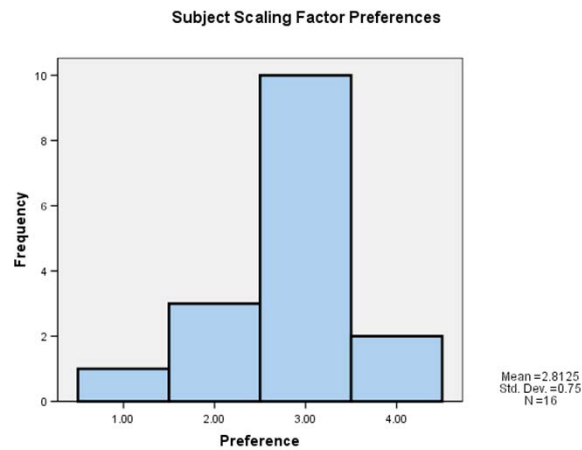
Example Experiment – Summary

- Subjects performed 11.5% faster with S2 and 15.0% faster with S3 with no statistically significant loss in accuracy
- Appears to be correlation between subject preferences and mean completion time
  - scaling factor of 3 is preferable amplification coefficent

Readings

- LaViola, J. "An Initial Evaluation of a Pen-Based Tool for Creating Dynamic Mathematical Illustrations", In the proceedings of the Eurographics Workshop on Sketch-Based Interfaces and Modeling 2006, 157-164, September 2006.
- Bragdon, A., Zeleznik, R., Williamson, B., Miller, T., and LaViola, J. "GestureBar: Improving the Approachability of Gesture-based Interfaces", Proceedings of ACM CHI 2009, 2269-2278, April 2009.
- LaViola, J., Leal, A., Miller, T., and Zeleznik, R. "Evaluation of Techniques for Visualizing Mathematical Expression Recognition Results", *Proceedings of Graphics Interface 2008*, 131-138, May 2008.