CSE 165: 3D User Interaction

Lecture #17: Evaluation

Jürgen Schulze

Announcements

- Homework assignment #4 due Thursday, March 20th at 3pm in lab 260

CAPE

- Submit CAPE forms on-line in weeks 9+10
- Responses to all surveys are completely anonymous.
- Only a summary of results is provided to the CS department and the instructor.
- This summary is provided AFTER final grades have been posted.
- A minimum number of three evaluations must be submitted by students for summaries to made available.

Evaluation Metrics – System Performance

- System performance metrics
- Average frame rate (fps)
- Average latency / lag (milliseconds)
- Variability in frame rate / lag
- Network delay
- Distortion

- Only important for its effects on user performance / preference
  - frame rate affects presence
  - network delay affects collaboration

Evaluation Metrics – Task Performance

- Speed / efficiency
- Accuracy
- Domain-specific metrics
  - education: learning
  - training: spatial awareness
  - design: expressiveness

Evaluation Metrics – User Preference

- Ease of use / learning
- Presence
- User comfort
- Usually subjective (measured in questionnaires, interviews)

User Comfort

- Simulator sickness
  - Kennedy - Simulator Sickness Questionnaire (SSQ)
- Aftereffects of VE exposure
  - Stanney 1998: Aftereffects from virtual environment exposure: How long do they last?
- Arm/hand strain
- Eye strain

3D Usability Evaluation

Things to Consider

Formality of Evaluation

- Formal
  - independent & dependent variables
  - statistical analysis
  - strict adherence to a procedure
  - hold constant all other variables
  - usually done to compare multiple techniques or at the end of the design process
- Informal
  - looser procedure
  - often more qualitative
  - subject comments very important
  - looking for broad usability issues
  - usually done during the design process to inform redesign

What is Being Evaluated?

- Application:
  - Prototype - consider fidelity, scope, form
  - Complete working system
  - Controlled experiments are rare
- Interaction techniques / UI metaphors
  - Can still evaluate a prototype
  - More generic context of use
  - Formal experiments more often used
- Consider "Wizard of Oz" evaluation

Subjects / Participants

- How many people?
- What backgrounds?
  - technical vs. non-technical
  - expert vs. novice VE users
  - domain experts vs. general population
- What age range?
- Recruiting
  - flyers
  - email/listservs/newsgroups
  - psychology dept.
  - CS classes

Number of Evaluators

- Multiple evaluators often needed for 3DUI evaluations
- Roles
  - cable wrangler
  - software controller
  - note taker
  - timer
  - behavior observer
  - …

Procedure

- Welcome
- Informed consent
- Demographic/background questionnaire
- Pre-testing
- Familiarize with equipment
- Exploration time with interface
- Tasks
- Questionnaires / post-testing
- Interviews

Pilot Testing

- Pilot testing should be used to:
  - "debug" your procedure
  - identify variables that can be dropped from the experiment

Instructions

- How much to tell the subject about purpose of experiment?
- How much to tell the subject about how to use the interface?
- Always tell the subject what they should try to optimize in their behavior.
- If using think-aloud protocol, you will have to remind them many times.
- If using trackers, you will have to help users "learn" to move their heads, feet, and bodies – it doesn't come naturally to many people.
- Remind subjects you are NOT testing THEM, but the interface.

Formal Experiment Issues

- Choosing independent variables
- Choosing dependent variables
- Controlling (holding constant) other variables
- Within- vs. between-subjects design
- Counterbalancing order of conditions
- Full factorial or partial designs

Independent Variables

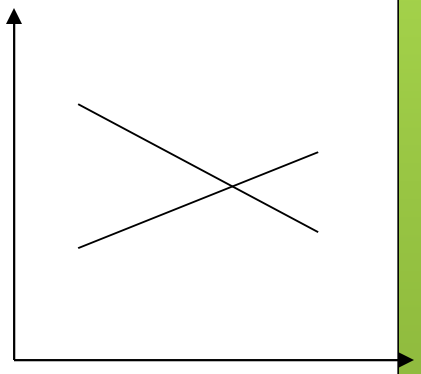
- Main variable of interest (e.g. interaction technique)
- Secondary variables
  - task characteristics
  - environment characteristics
  - system characteristics
  - user characteristics

Metrics (dependent variables)

- Task performance time
- Task errors
- User comfort (subjective ratings)
- Observations of behavior (e.g. strategies)
- Spoken subject comments (e.g. preferences)
- Surveys/questionnaires
- Interviews

Data Analysis

- Averages (means) of quantitative metrics
- Counts of errors, behaviors
- Correlate data to demographics
- Analysis of variance (ANOVA)
- Post Hoc analysis (t-tests)
- Visual analysis of trends (esp. learning)



- *Interactions between variables* are often important
- Expect high variance in 3DUI interaction studies

Analysis Tools

- SPSS, SAS, etc.
  - full statistical analysis packages
  - parametric and non-parametric tests
  - test correction mechanisms (e.g., Bonferroni)
- Excel
  - basic aggregation of data
  - Correlations
  - confidence intervals
  - graphs
- Matlab, Mathematica