Techniques for User Evaluation

Lecture #12: User Evaluation
Joseph J. LaViola Jr.
Fall 2016

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

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 ($\alpha$, 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


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

<table>
<thead>
<tr>
<th>Effect</th>
<th>Time</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>$F_{3,13}=3.26$, $p=0.056$</td>
<td>$F_{3,13}=4.8$, $p&lt;0.05$</td>
</tr>
<tr>
<td>T</td>
<td>$F_{1,15}=13.66$, $p&lt;0.05$</td>
<td>$F_{1,15}=22.96$, $p&lt;0.05$</td>
</tr>
<tr>
<td>A</td>
<td>$F_{1,15}=55.46$, $p&lt;0.05$</td>
<td>$F_{1,15}=0.001$, $p=0.98$</td>
</tr>
<tr>
<td>S x T</td>
<td>$F_{3,13}=0.29$, $p=0.83$</td>
<td>$F_{3,13}=1.575$, $p=0.243$</td>
</tr>
<tr>
<td>S x A</td>
<td>$F_{3,13}=0.87$, $p=0.523$</td>
<td>$F_{3,13}=0.562$, $p=0.649$</td>
</tr>
<tr>
<td>T x A</td>
<td>$F_{1,15}=5.03$, $p&lt;0.05$</td>
<td>$F_{1,15}=0.573$, $p=0.46$</td>
</tr>
<tr>
<td>S x T x A</td>
<td>$F_{3,13}=0.73$, $p=0.55$</td>
<td>$F_{3,13}=0.97$, $p=0.436$</td>
</tr>
</tbody>
</table>

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

Significant differences between S1 and S2 and S1 and S3

Significant difference between S1 and S4
Example Experiment – Results: Subject Preferences

- 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 coefficient.
Understanding Visual Interfaces for the Next Generation of Dance-Based Rhythm Video Games

Emiko Charbonneau Andrew Miller
Chadwick Wingrave Joseph J. LaViola Jr.
University of Central Florida

Overview

- Problems with Current Dance Games
- RealDance Description
- Visual Interface problems with Dance Games
- Visual Interface Descriptions
- Experimental Design
- Results
- Conclusions
Interface Problems with Dance Games

- Among rhythm games, dance still doesn’t feel like dancing
- Full body interface games are now mainstream
- Initial Research Goal:
  - Create a video game that feels like dancing
  - Detect more specific movements
    - To teach better
    - To prevent cheating
  - Make fitness gaming more fun

RealDance Overview

- Dance Game Prototype
  - No buttons
  - No cameras
  - No wires
- Gesture Scoring
  - Impact
  - Impulse
  - Freeze

Real Dance (Charbonneau et al, 2009)
Visual Interface Trouble

- Icons scrolling along a path
- Goal to make as many different moves as possible
- But how to display it without being confusing?
  - Current rhythm games have 4-6 colored shapes
  - More specific icons get more confusing

Visual Interfaces in Video Games

- Surveyed 76 rhythm related games from about 10 years
- Current and previous rhythm game needs:
  - When to press button
  - What button to press
- 3DUI requires three things
  - When to move
  - Which body part to move
  - Where to move it to
Visual Interface: DDR and Rock Band

- Almost every rhythm, music and dance game uses a variation of this
  - Icons stream along path
  - A perpendicular line indicates when to press
  - Color, position and shape used to assist in deciding between actions
- Our first prototype as well!

Images from konami.com and rockband.com
Visual Interface: Elite Beat Agents

- A.K.A. Osu! Tatakae! Ouendan
- For Nintendo DS
- Uses touchscreen and stylus
- User taps the number circle when the outer circle shrinks to it
- For some notes they trace along a path
  - Only three other games with this UI
  - Image from Nintendo.com

Visual Interface: Beat Circles
Visual Interface: We Cheer

- Wii game using two Wiimotes as pompoms
- Player follows characters and arrow paths
- Timing is done by ghost image
- Color for different hands
  - Only two similar games
  - Image from Namco Bandai

Visual Interface: Motion Lines
Experimental Hypothesis

- Run a user study comparing three visual interfaces
- Users play RealDance with all of them
- Study their preferences and performance

- Our hypothesis: players will prefer Motion Lines or Beat circles over the Timeline interface, because the streaming icons must present too much information

Subjects and Apparatus

- Participants
  - 24 participants: 13 male, 11 female
  - Ages 18-29
  - 19 had no formal dance experience
  - 17 play video games > once a month
  - 14 familiar with Dance Dance Revolution

- Apparatus
  - Implemented in C# using XNA on a PC running Windows Vista
  - 50 inch Samsung HDTV, 1920 x 1080 resolution
Experimental Design

- Experiment takes place in an enclosed space
- Consent form, Pre-questionnaire, Icon sheet
- Suited up into Wiimote sleeves
  - One each wrist, one each ankle
- Experimental Task
- Post Technique Questionnaire
  - 16 questions, 4 open answer
- Post Questionnaire
  - 10 questions, 8 open answer

Experimental Task

- For each interface
  - Two practice sessions to Ghostbusters theme
  - Gameplay session to Thriller
    - RIP Michael Jackson 😞
- Scored based on timing if correct movement
  - Each move either 100, 75, 50, or 0
  - Compound moves scored per limb
  - Max score 6700
Results: Learning Effects

- Each participant received one of six arrangements
- Even though order was randomized, choreography was identical
- Repeated measures one way ANOVA
  - $F_{2,22} = 0.306, p = 0.738$
- No significant improvement from game play session order

Results: Score Analysis

- Participants performed better at spatial interfaces
- Holm’s sequential Bonferroni adjustment with three comparisons at $\alpha = 0.05$
  - ML > TL
    - ($t_{23} = -4.38, p < 0.0167$)
  - BC > TL
    - ($t_{23} = -3.26, p < 0.025$)
  - No significance between ML, BC
    - ($t_{23} = -1.20, p < 0.243$)
Results: Post Technique

- Easy to Follow?
  - BC > TL (Z = -2.69, p < 0.0167)
  - ML > TL (Z = -2.39, p < 0.025)

- Position of the icons confusing?
  - TL > BC (Z = -3.08, p < 0.0167)
  - ML > TL (Z = -2.38, p < 0.025)

- Score matched how you felt you did?
  - BC > ML (Z = -2.50, p < 0.0167)

Results: Post Questionnaire

- Only question 1 was found significant:
  - Which interface did you perform the best in? (Beat Circles)

- Worth noting that Timeline was least chosen interface for each question except for question 7:
  - Which did you like the least?

- Spatial nature of Motion Lines and Beat Circles may have divided choices
Discussion

- Timeline
  - Liked to see the approaching moves ahead of time
  - Still found it hard to know when to start moving
- Motion Lines
  - Much better sense of where to go, which body part to use
  - Repeated movements were harder to see
- Beat Circles
  - Icon position defined ending position, timing was easier
  - Overlapping circles made repeated movements confusing

Conclusion

- So far, the Timeline interface has worked well for rhythm dance games
- But as video game consoles explore 3D user interfaces, they can now create new gameplay experiences
  - Nintendo, Sony, and Microsoft all made interface announcements at E3 2009
- In our study spatially designed interfaces were easier and preferred overall
- Identified pros and cons for each design
Readings