Usability testing

Frank E. Ritter

17 april 2014

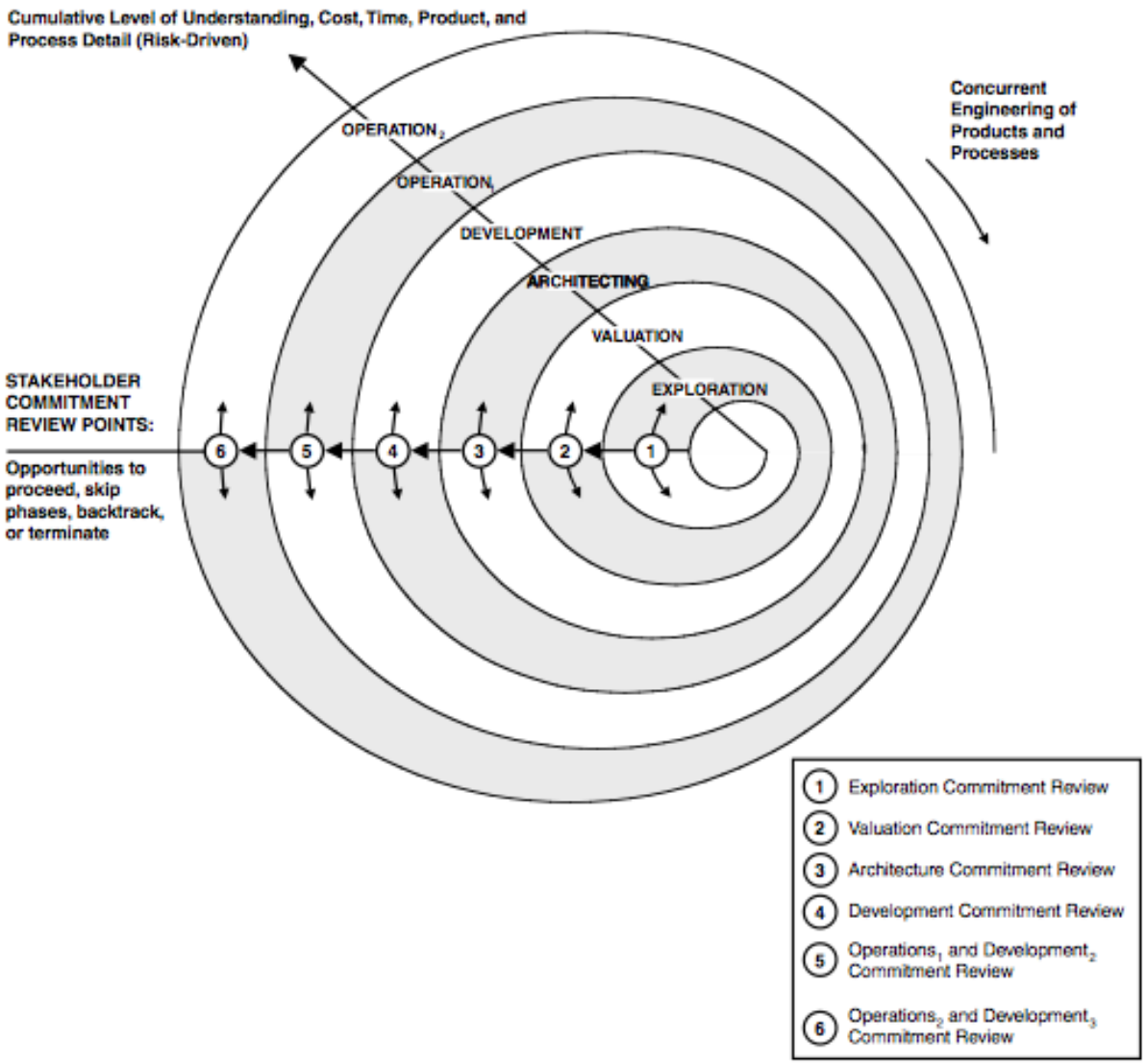
Cornwell, 2001

Why use user-testing?

- Demonstrate a weakness or strength of a design feature during the design process;
- Evaluate the adequacy of an overall design, or of particular design features;
- Where guidelines and principles do not always apply;
- Where guidelines and principles (and even task analysis) are not always persuasive (true, but sad, really);
- Where designers require feedback; and
- Because all people (including designers) make mistakes (i.e. user-testing can be used as a type of proof-reading).
- Because development process is buggered

Risk-driven Spiral Model

(Pew & Mavor, 2007; Boehm & Hansen, 2001)



Prerequisites

- Tasks/ task scenarios
- Example users

Note: not many prerequisites
but, can be expensive

Formative and summative evaluation

- **Formative, to help design, to IN form**
- **Summative, to SUM up**
  Is this interface up to standard?
  - **Time to do a task, time to learn a task**
  - **Error rates**
    - ***What is an error? Corrections you can find, blowback (other long term effects), on the other extreme, is harder***

Verbal Protocols

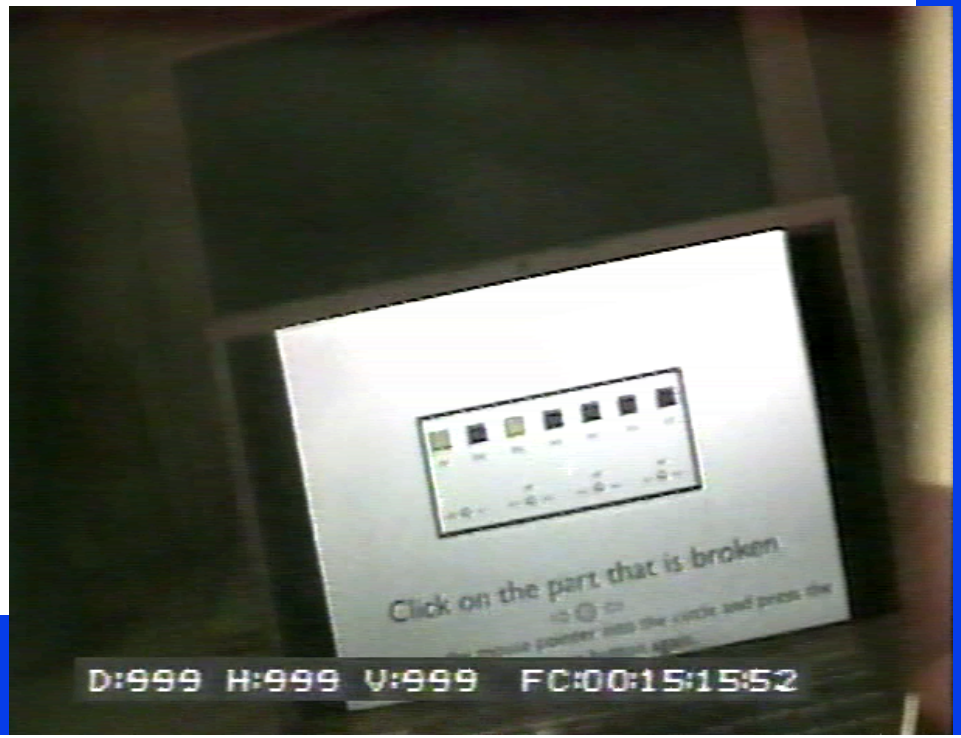
- Talk aloud while doing a task
- Ericsson and Simon (1983; 1984; 1990) provide a theory of when and why you can and cannot do this
- What information is in working memory
- NOT NOT: why someone thinks that they do a task
- Concurrent/retrospective

Visual Protocols

- Video of users
- May be unnatural
- Takes time to analyse
- Used in retrospective protocols
- Getting tools to analyse such things
- Dribble files, RUI, as cheap substitutes

Eye Movements

- Rarely used
- Expensive
- Used in yellow pages, web studies sometimes, cockpits



Patterns of use

- Put object in work environment: used in WTC robot work
- Can find disconnects between requests (we need style sheets) and use (Word users don't use styles until quite expert)

Surveys

- Can measure attitudes
- But have to be quite careful, maybe your item is the least disliked of a disliked category social effects of questions, social, political, religious, personal
- Again, a disconnect between behaviour (physical and verbal) and beliefs

Workload

- NASA TLX measure, how hard are you working?
- Some eye movements being used
- Heart rate and other heart measures
- Dual tasking, looking for degraded behaviour in a second task

Paper and pencil mockups

- Show evaluators mockups of the interface "Story boarding"
- Show to real users, other designers
- If done with computer or tools, Wizard of Oz
- Early, inexpensive, qualitative, broad brush

Prototyping

- Making something fast
- Looking at it, with any of the previous and later methods
- Can blend with WoO

- Real systems can also be used

Cooperative evaluation

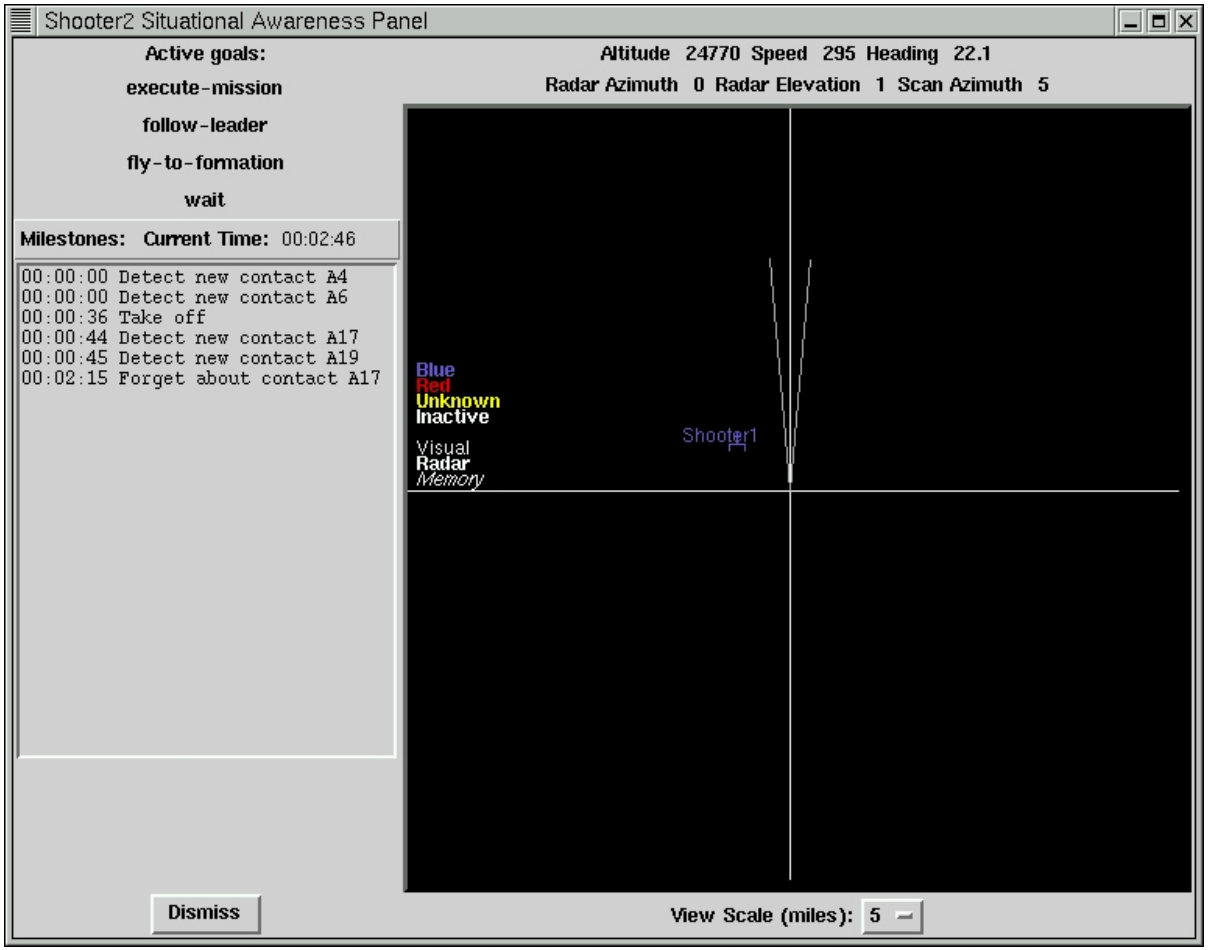
- So-called User-centered design
- Work with the users, include them on the design team

Ethics

- Treat subjects/participants, fairly, kindly, like you would like to be treated
- Anonymous data is the first step
- Don't waste their time
  - Efficient
  - useful
- IRB approval

Evaluation of an Interface

Avraamides, M., & Ritter, F. E. (2002). Using multidisciplinary expert evaluations to test and improve cognitive model interfaces. In ***Proceedings of the 11th Computer Generated Forces Conference,*** 553-562, 502-CGF-002. Orlando, FL: U. of Central Florida. [won a prize]



The subjects

1 Plan view/geographic information systems specialist
2 Graduate student in AI and cognitive modeling
3 Marine Major, specializing in logistics and infantry
4 Former software developer in Silicon Valley with Fortune 100 companies
5 Former merchant marine officer and expert on social and group processes
6 Navy fixed and rotary wing pilot. RWA instructor
7 Cognitive psychologist
8 Cognitive psychologist with some amateur flying experience
9-12 Former military pilot from BMH Associates

Users did tasks

- Some warmup tasks
- Some actual tasks we knew they should be able to do
- Followed by discussion of tasks and then wish lists

Method and analysis

- Video taped, both people and interface
- Analysed into a list of suggestions/problems
- And tasks that users do

Example problems

- 1, 2, 4, 6, 7, 8, 10, 12 Labels for agents in the Agent awareness display are not very clear Use different fonts, or font size, or icons to distinguish their status
- 10 Labels take much space in the Agent Awareness Display. Labels do not need to be visible all the time. Provide the user with the opportunity to hide them.
- 9 Label "ukn" is not needed The symbol used (square with a speed vector) is used in cockpits to represent unknown entities. Use the symbol of unknown and a different one for known.
- All subjects SAP resizes as more goals are added on the Active Goal panel Use scrolling windows for the Active Goal panel and the History panel so that a fixed number of goals/events is visible at any point, while the user can use the scroll bar to examine any goal or event
- 1, 4, 6, 10 The most recent goal/events are added on the bottom of the Active Goal/History panel. This is counterintuitive given that in goal-stack most recent items are added on the top. Reverse the order of events in the Active Goal and the History panels so that most recent events are added on the top. Do this along with adding scroll-bars so that users scroll down to view previous goals and history events.