

# Usability Engineering

## Empirical Evaluation

---

Matthias Rauterberg

HG 3.53

TU/e - ID - DI

[g.w.m.rauterberg@tue.nl](mailto:g.w.m.rauterberg@tue.nl)

Website:

<http://www.idemployee.id.tue.nl/g.w.m.rauterberg/>

## Content today

---

- Method overview
- Usability testing
- Thinking Aloud

# Evaluation with users

- Methods:
  - Observations (field or lab, final or prototype, controlled, less controlled)
  - Interviews and questionnaires
  - Focus Groups
  - Software logging

# Observations (1)

- What to observe?
  - Which activities? E.g.
    - Representative sample
    - Only new functionality/tasks
    - Global use
- Where? Lab, field study
- When? Sampling..
  - Learning phase, and/or expert use
  - Once, or more observations over time

## Observations (2)

- How?  
Non-obtrusive..
- What?  
Times, errors, ...
- Interacting with what?  
Final product,  
prototype

## Case: voice controlled audio-set

- Product: audio-system  
(CD, tape-deck, tuner)
- Speech input
- Question:
  - usability problems with audio-set
  - speech error handling
  - vocabulary issues
  - multiple-user issues



## Assignment: considerations

- Describe pros and cons:
  - doing observations at users' homes
  - observations in a laboratory (e.g. IPO living room)
  - Consider: access to people, time needed, data gathered, etc.



## Observation: considerations

- Observing users in lab
  - Giving them 10 tasks
  - Measure: ASR errors, time, 'confusion'
  - Controlled set-up
- Observe users at home
  - Study natural behaviour
  - Log their actions, using software
  - Better understanding of real issues

## Types of observations

---

- Participant observation:
  - researcher participates in activities, e.g. with or without others knowing
  - e.g. collaborative evaluation: prompted evaluation
- Non-participant observation:
  - non-intrusive
  - less chance of bias

## Observations in the field

---

- Evaluation of machine that manufactures CD's
- Observations in industry
- Contextual information uncovered
- Not all tasks performed when we were there (start-up, all sorts of alarms)!

## Observations on location



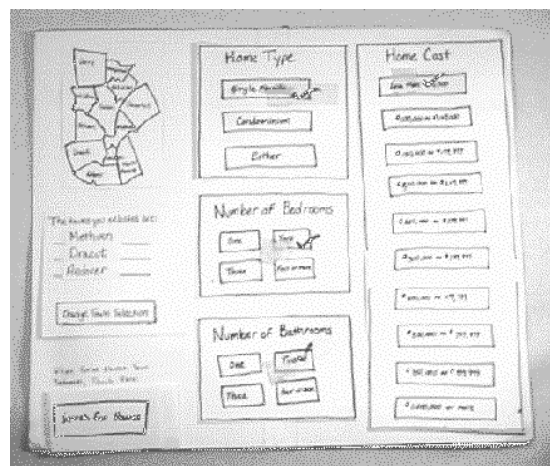
© M. Rauterberg, 2006

JFS-USI Primer-9

11/65

## Paper prototyping (1)

- Hand drawn
- Do tasks
- More inclined to comment
- Slow



© M. Rauterberg, 2006

JFS-USI Primer-9

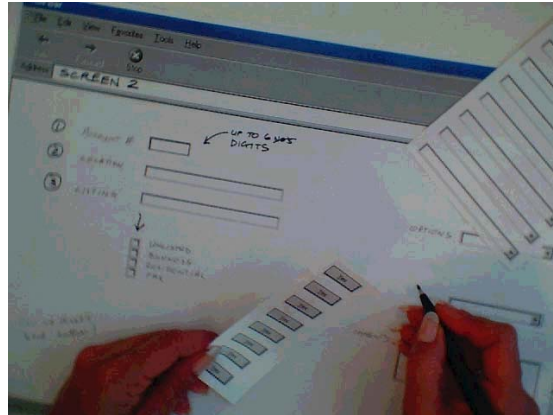
12/65

## Paper prototyping (2)

More final looking

More realistic (size, etc.)

Still slow



## Interviews

- Determine users' opinions
- Structured versus unstructured
- Open-ended versus closed questions
- One person versus groups
- Designers' versus users' world
  - contextual inquiry
  - ethnography

## Interview: structure

---

- Unstructured or in-depth interview:
  - interviewer develops an interview guide
  - questions are formulated within scope of guide
  - possibility to pursue interesting facts
- Structured interview:
  - Pre-determined set of questions
  - Uniform information, comparability of data

## Interview: advantages

---

- More appropriate in complex situations
- Collects in-depth information
- Information can be supplemented with impression of interviewees (non-verbal info)
- Questions can be explained



## Interview: disadvantages

- Time-consuming and expensive
- Quality of data depends on
  - quality of interaction
  - quality of interviewer
- Bigger chance of bias by interviewer

## Focus groups

Discussion of design  
issues  
In a group  
Reactions to each other  
  
Bias



## Focus groups: examples

- Early: discuss context of use of speech control for consumer products, to understand requirements for new product
- Late: show concrete product, and discuss possible uses, advantages and disadvantages

## Contextual enquiry

- Interviews in work context
- Work objects available
- Easier to address realistic issues
- Interruptions



## Contextual enquiry: example (1)

- Early in process: determine requirements for information sharing space
- Late: Evaluate how product is used in context of use

## Contextual enquiry: example (2)

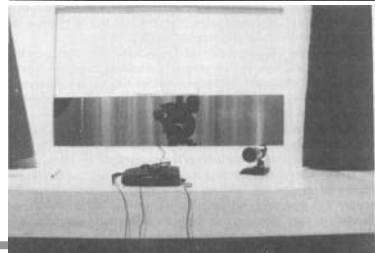
contextual information for info management space



## Private Camera Conversation



People talk about product  
More informal



© M. Rauterberg, 2006

JFS-USI Primer-9

23/65

## Questionnaires

- Aims: determine users' opinions
- Flexibility of data gathering more limited
- Open-ended versus closed questions
- In person, via mail, e-mail, the web
- Measure e.g.:
  - satisfaction
  - functionality issues

© M. Rauterberg, 2006

JFS-USI Primer-9

24/65

## Questionnaire: example (1)

---

- Voice controlled audio set
- Questionnaire (after use):
  - Ease of use
  - Perceived number of errors
  - Intention to buy
  - Open question: when (not) to be used
- Comparison: observation exercise

## Questionnaire: example (2)

---

- Early: information about user groups, activities, frequency of use, etc.
- Late: information about questions specific to new product, having been used, e.g. shavers: easy of use, smoothness after shaving, context of use

## Questionnaire: advantages

---

- It is less expensive than interviews
- It offers greater anonymity of subjects

## Questionnaire: disadvantages

---

- Limited to subjects that can read and write
- Usually low return rate (target: >30%)
- Self-selecting bias
- No possibility to explain questions
- No possibility for spontaneous response (have time to think over answer)
- Subject can read over all questions before answering
- It is possible to consult others

## Open question: pros and cons

---

- Provide in-depth information
- Analysis is more difficult
- Respondents can express themselves freely, but if they have trouble doing so information is lost
- Less possibility for investigator bias, but larger chance of interviewer bias

## Closed question: pros and cons

---

- The information lacks depth and variety
- More chance of investigator bias, because possible answers are limited
- Respondent may just tick one option without reflection
- Because answers are categorised, data is easy to analyse

## Questions sequence: >

- From wide to smaller 'scope':
- Start with general question, and slowly scope down to main issues
  - Helps respondent to build up understanding of area of interest
  - Provides opportunity for spontaneous comments, before pointers by more detailed questions

## Question sequence: <

- From detailed issue to wider scope:
- First ask detailed questions and then broaden questions to more global issues:
  - This allows respondent to have considered several detailed issues, before providing more general opinion



## Formulating questions

- Use simple language
- Use clear / unambiguous questions
  - Are you satisfied with your computer?
- Do not ask combined questions
  - How often en how long do you use your computer?
- Do not ask leading questions
  - Do you think it is a nice game?
- Do not ask questions that are based on presumptions
  - Which do you prefer TombRaider I or II?

## Cultural probes (1)

Questions in  
context

Over time

Personalised

Less control



## Cultural probes (2)



© M. Rauterberg, 2006

JFS-USI Primer-9

35/65

## Cultural probes (3)

- Probes handed out in user session
- Information returned over time
- Included:
  - Postcards with questions
  - Diary with camera
  - Maps with related questions
- Combination of user research and inspiration

© M. Rauterberg, 2006

JFS-USI Primer-9

36/65

## Usability Testing (1)

---

- Global idea:
  - Users interact with a product to determine usability problems
- Many variations possible
  - Number of users at the same time
  - Interact with mock-up or final system
  - Different ways of prompting for feedback (continuously, retrospectively)

## Usability Testing (2)

---

- Aim: controlled studies (hypothesis testing)
- Examples:
  - Comparison of two different sets of icons to be used at airports
    - research question: least amount of confusion
  - Comparing speech versus non-speech audio
    - research question: comparison, preference

## I: Planning a usability test

---

- Define usability goals
- Define user profile, and participants
- Selecting and presentation of tasks
- How to measure usability
- Preparing test materials
- Conducting a pilot test

## II: Conducting and analyzing the results of a usability test

---

- Interacting with the participants
- Conducting the test
- Tabulating and analyzing the data
- Recommending changes
- Communicating the results

## Making choices

---

- You can hardly test everything!
- Select the tasks/function set
- Select the users
- Select the usability focus
- The amount of training and preparation for participants

## Select tasks

---

- Possible argumentation:
  - Critical (safety, task failure)
  - New or modified
  - Frequently used, thus important
  - Infrequently used, thus might forget about
  - Under pressure/stress

## Task order

---

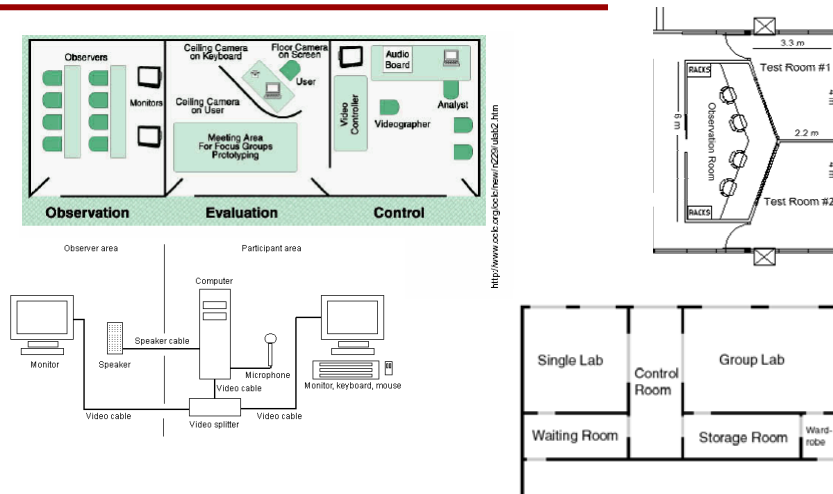
- Should be in logical order
- Important tasks should be earlier in the test
- Easy one first (“warm up”)
- Repeat tasks, to determine ease of learning, understanding

## Software logging

---

- Aim: determine (frequency of) use of functionality
- Logs user actions
- Can be automated
- Can be used to infer user errors
- No information about perceptions, or context / reasons for use

# Usability Labs: ground floors (1)

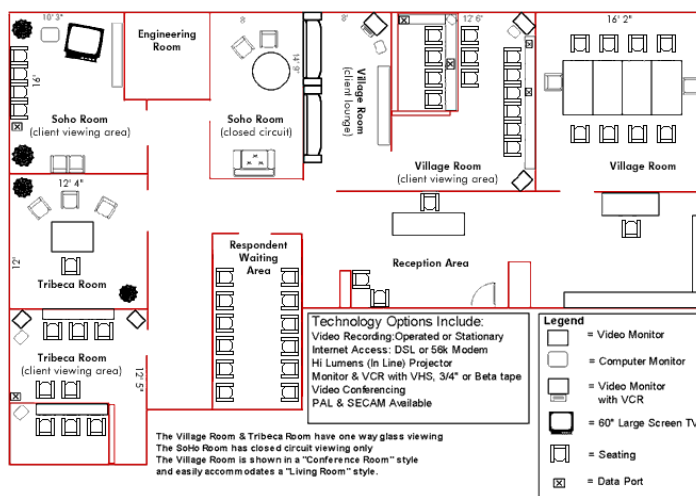


© M. Rauterberg, 2006

JFS-USI Primer-9

45/65

# Usability Labs: ground floors (2)



© M. Rauterberg, 2006

JFS-USI Primer-9

46/65

## Usability Labs: examples (1)



© M. Rauterberg, 2006

JFS-USI Primer-9

47/65

## Usability Labs: examples (2)

### ■ IPO living room:



© M. Rauterberg, 2006

JFS-USI Primer-9

48/65



# Common Industry Format (CIF)

This international technical specification is intended to be used by:

- usability professionals within supplier organizations to generate reports that can be used by customer organizations
- customer organizations to verify that a particular report conforms to this international technical specification
- human factors or other usability professionals in customer organizations who are evaluating both the technical merit of usability tests and the usability of the products
- other technical professionals and managers in the customer organization who are using the test results to make business decisions.

Reference: NIST (2001): [Common Industry Format](#)

## CIF: test subjects profile (cont.)

The characteristics shall be complete enough so that an essentially similar group of test subjects can be recruited.

Characteristics should be chosen to be relevant to the product's usability; they should allow a customer to determine how similar the test subjects were to the customers' user population.

EXAMPLE TABLE: The table below is an example; the characteristics that are shown are typical but might not necessarily cover every type of testing situation.

	Gender	Age	Education	Occupation / role	Professional Experience	Computer Experience	Product Experience
P1							

## CIF: task[s] (cont.)

Task A

User #	Unassisted Task Effectiveness [(%)Complete]	Assisted Task Effectiveness [(%)Complete]	Task Time (min)	...	Errors	Assists
1						
2						
N						
Mean						
Standard Deviation						
Min						
Max						

## CIF: summary (cont.)

Summary

User #	Total Unassisted Task Effectiveness [(%)Complete]	Total Assisted Task Effectiveness [(%)Complete]	Total Task Time (min)	...	Total Errors	Total Assists
1						
2						
N						
Mean						
Standard Deviation						
Min						
Max						

## Thinking Aloud (TA) approaches

---

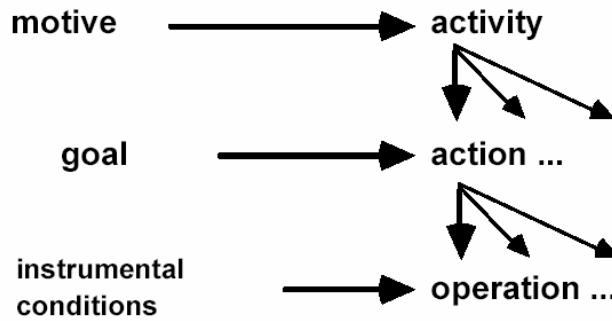
- Elicit user feedback during interaction with product
- Needs training
- Decide on how to prompt, to minimize disturbing natural user-product interaction

## Theoretical foundation

---

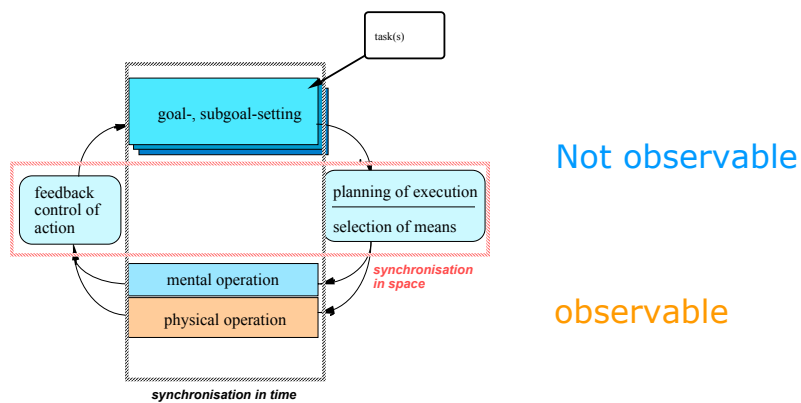
- Introspection is as old as psychology
- Purpose: get insight into cognitive processes (often: problem solving tasks)
- Subjects try to verbalise their own mental processes (during execution of tasks)

# Human Activity Hierarchy

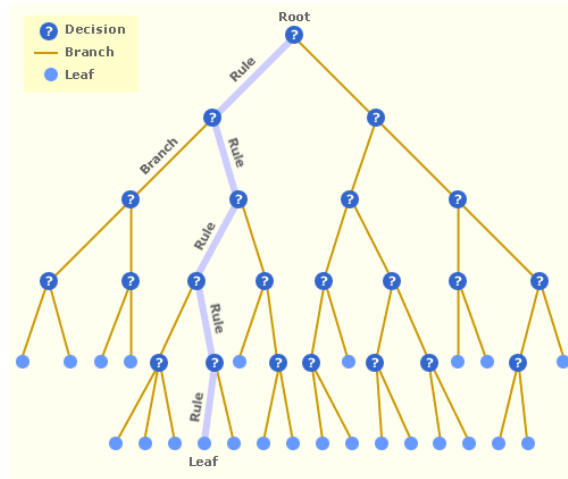


# Activity Theory: action cycle

## the complete action cycle



# Goal Hierarchy



© M. Rauterberg, 2006

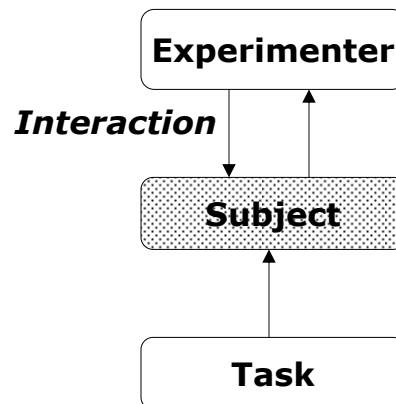
JFS-USI Primer-9

57/65

## Experimental setting: Protocol Analysis (PA)

Questions:

- do subjects have access to their mental processes?
- Can they verbalise them?



© M. Rauterberg, 2006

JFS-USI Primer-9

58/65

## Protocol Analysis - Verbal reports

[reference: Ericsson & Simon, 1984]

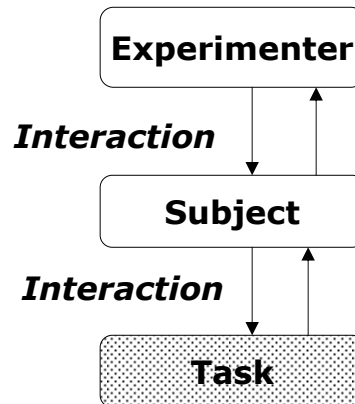
- **Level 1:** Verbalizations which do not require operations because they are already in the short-term memory (STM)  
(e.g. intermediate results of multiplying  $36 * 24$ )
- **Level 2:** Verbalizations which require one or more operations because they are not in verbal form in STM  
(e.g. verbalizing what is seen in pictures)

## Protocol Analysis (cont.)

- **Level 3:** Verbalizations which require cognitive effort beyond the task or verbalization.  
(e.g. filters ('only focus on that?'), interventions ('why did you do that?'), any form of interruption)
- Level 1 is hard data, but hard to obtain
- Level 2 is reliable (hard data)
- Level 3 is no data

## Think aloud for usability testing

- Emphasis not on understanding thought processes
- Assumption of sequential processes, possibly incorrect
- Too limited information



## Thinking Aloud (TA)

[reference: Ericsson & Simon, 1984]

- Preparation:
  - Explain the difference between explaining and thinking aloud
  - Emphasize the need for continuous verbalization
  - Practice! Give an example
- Remind people to think aloud (e.g. after 15-20 sec silence)
- Any other interaction between user and facilitator is prohibited!

## Alternative Communication Model (CM)

[reference: Boren and Ramey, 2000]

- Discrepancies between theory and practice
  - Reminders: “What do you think would happen if you click on that button?” (Too much info)
  - Interventions “Tell me what you like and what you don’t like”.
  - Only level 3 data => unreliable

## Alternative CM (cont.)

- Look for other theoretical foundation:
  - E.g. speech communication
- Basic ideas:
  - A listener (the facilitator) cannot remain passive, but has to be a dialogue partner
  - A dialogue partner is allowed to act both as a listener and a speaker
- Benefit: possibly more natural assumptions for role of facilitator



## References

---

- J. Dumas and J. Redish (1993) A practical guid to usability testing. Ablex Publ.
- A. Ericsson and H. Simon (1984) Protocol Analysis. MIT Press. (new revised edition 1993)
- Boren and Ramey (2000) Thinking Aloud: Reconciling Theory and Practice. IEEE TRANSACTIONS ON PROFESSIONAL COMMUNICATION, VOL. 43(3), 261-278 ([thinkaloud\[2000\].pdf](#))
- NIST (2001): Common Industry Format. ([Common-Industry-Format.zip](#))